Neighbourhood profiling and classification for community safety

Peter James Shepherd

The University of Leeds
School of Geography

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

September 2006

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgments

The research recorded in this thesis was made possible thanks to the sponsorship of the Economic and Social Research Council and the Leeds Community Safety Partnership (ESRC CASE Award PTA-033-2002-00006). Many thanks must go to Andy Mills at Leeds Community Safety, for his commitment to this project and timely reminders for me to finish it! Recognition also needs to be given to all those that contributed data for the research, especially officers at Education Leeds, Leeds City Council, West Yorkshire Fire and Rescue Service, West Yorkshire Police, the Leeds Refugee and Asylum Support Service, Leeds Primary Care Trusts and Ordnance Survey.

The research was conducted within the friendly environs of the School of Geography, and many people have helped and encouraged me along the way. First and foremost I would like to thank John Stillwell and Graham Clarke for their patient supervision, and also Phil Rees and Adrian Bailey for their stewardship of the School during my stay. I should also like to thank Adam Crawford, School of Law, for finding me tasks which although distracting me from my thesis, broadened my knowledge and experience of community safety and enriched the PhD as a result.

On a more personal note, many thanks to Alison Heppenstall for showing me the PhD ropes and convincing me to ditch Microsoft Word in favour of LaTeX; and thanks to Dan Vickers for many discussions regarding classification. Special thanks also to Dave, Lou and Ruth for bike rides around the beautiful highways and byways of Yorkshire. Finally, I must say a big thank you to my parents, Keith and Brenda, and sister Liz, for their support of my various academic endeavors over the years.

Abstract

This thesis presents a new neighbourhood classification, the Leeds Classification for Community Safety (LCCS). This is used to demonstrate the usefulness of area classifications for providing area context information to crime analysis, and for identifying neighbourhoods with atypical crime profiles - given their neighbourhood type. The work can be seen as a development of the classifications produced by the Home Office for comparative performance purposes, but at a smaller, neighbourhood scale. There has been a recent trend among practitioners to use commercial geodemographic products for this task, but these tools are primarily designed for consumer segmentation applications and little is revealed about the way in which these classifications are constructed, or their ability to discriminate geographies of crime and disorder. The research presented in this thesis discusses critically both these issues .

The research draws upon academic and policy literature on the geography of crime, environmental criminology and community safety policy, and describes the types of task undertaken by community safety analysts. Existing knowledge about the causes and motivations for crime are used to select variables from new national and local sources. The final partition was created using the fuzzy c-means clustering technique, but alternative techniques were also employed and levels of agreement between the different results were measured. The design process also involved measuring the ability of different partitions to discriminate neighbourhood crime rates.

Numeric comparisons were made between the LCCS and existing general purpose classifications, and these show that the task-specific approach was better overall at discriminating crime rates. Practical applications of the LCCS are also demonstrated using recorded crime data for criminal damage and domestic burglary. Furthermore, variations in response to burglary target hardening are analysed using the LCCS, and the cost benefit to neighbourhoods of different types is shown. These practical demonstrations of the LCCS go to reinforce the assertion that area classification can be useful, practical tool to aid in the analysis and understanding of spatial patterns of crime and disorder.

Contents

Acknowledgments iii

Abstract v

Contents vi

List of Tables xi

List of Figures xv

1 Introduction 1
- 1.1 Area Classification for Community Safety 1
- 1.2 Aims and Objectives 3
  - 1.2.1 The Study Area 4
- 1.3 Thesis Structure 4
- 1.4 Outputs 6

2 Geography, Crime and Community Safety 7
- 2.1 Introduction 7
  - 2.1.1 The Development of a Geography of Crime 7
- 2.2 Crime and Criminology 8
  - 2.2.1 Criminological Theory 9
  - 2.2.2 Environmental Criminology 13
- 2.3 Policy Approaches 16
  - 2.3.1 Situational Crime Prevention 16
  - 2.3.2 Social and Communal Strategies 18
- 2.4 Community Safety 19
  - 2.4.1 Morgan Report 19
  - 2.4.2 Safer Cities 20
  - 2.4.3 Crime and Disorder Act 1998 21
  - 2.4.4 Safer Leeds Partnership 22
- 2.5 Information Strategies, Analysis and Evaluation 23
  - 2.5.1 Data Sharing 24
  - 2.5.2 The Partnership Business Model 26
  - 2.5.3 Triennial Audit and Strategy Setting 27
  - 2.5.4 Police and Justice Bill 2006 28

2.6 Concluding Remarks 29

3 Crime Mapping and Area Profiling 31

3.1 Introduction 31

3.2 Spatial Analysis of Crime 32

3.2.1 Spatial Statistics 32

3.2.2 'Hotspot' Mapping 34

3.2.3 Spatiotemporal Patterns 40

3.3 Moving 'Beyond Blobology' 43

3.3.1 Measuring Risk 43

3.3.2 Standardisation Techniques 45

3.4 Providing Area Context 47

3.4.1 Single Area Profiles 47

3.4.2 Geodemographic Profiling for Crime Analysis 49

3.5 Towards a New Neighbourhood Classification for Community Safety 51

3.5.1 Task-Specific Classification 51

3.5.2 Mutually Exclusive Membership 52

3.5.3 Geographical Scale 53

3.5.4 Homogeneity 55

3.5.5 White-Box Design 55

3.6 Concluding Remarks 56

4 Selecting Variables 59

4.1 Introduction 59

4.2 Garnering Opinion 59

4.2.1 National Community Safety Network 60

4.2.2 Leeds Statistics Scoping Event 60

4.2.3 Literature Review 62

4.2.4 Summary of Survey Responses 64

4.3 A New Neighbourhood Geography for Leeds 65

4.3.1 Re-zoning Leeds 65

4.3.2 Sub-Dividing Community Areas 66

4.4 Calculating Data Values for SCAs 67

4.4.1 Aggregation of Output Areas 67

4.4.2 Aggregation of Point-based Features 67

4.4.3 Aggregation of Area-based Features 68

4.4.4 Geographical Harmonisation 69

4.4.5 Surface Modelling 69

4.5 National Datasets 70

4.5.1 Census of Population 72

4.5.2 Selection of Census Variables 73

4.5.3 Cartographic and Land Use Data 79

4.5.4 Selection of MasterMap Variables 80

4.5.5 Business and Economic Data 86

4.5.6 Selection of Nomis and ABI Variables 87
4.6 Local Datasets 88
4.6.1 Leeds City Council Data 88
4.6.2 Police Force Data 90
4.6.3 Selection of Police Variables 97
4.6.4 Fire Service Data 97
4.6.5 Selection of Fire Service Variables 98
4.6.6 Health and Ambulance Data 98
4.6.7 Selection of Health Variables 100
4.7 Concluding Remarks 101

5 Classification Design **103**
5.1 Introduction 103
5.2 Cluster Analysis 104
5.2.1 A Typical Cluster Analysis 104
5.2.2 Clustering Methods 105
5.3 Variable Selection and Weighting 108
5.3.1 Mathematical Techniques for Determining Variable Weights 109
5.3.2 A Mixed-method Approach to Variable Weighting 110
5.3.3 Identifying Outliers 110
5.3.4 Correlation Between Variables 111
5.3.5 Descriptive Statistics 112
5.3.6 Principal Components Analysis 113
5.3.7 Removal and Reworking of Variables 113
5.3.8 Methods of Standardisation 115
5.4 Creating an Initial Partition 115
5.4.1 Finding an Optimum Value for k 115
5.4.2 Additional Crime Heuristics 119
5.4.3 Synthesising Findings 124
5.4.4 Improving Cluster Definition 124
5.4.5 k-means 125
5.4.6 Fuzzy c-means 126
5.4.7 Revisiting Variable Selection 129
5.4.8 OVW and HINoV 131
5.4.9 Final Partition Summary 133
5.4.10 Final Partition Information 134
5.4.11 Reductions in Variability 134
5.5 Concluding Remarks 138

6 Portraits and Classification Testing **139**
6.1 Introduction 139
6.2 Group Portraits 139
6.2.1 Visualising Group Profiles 141
6.2.2 Group 1 142

6.2.3 Group 2 144
6.2.4 Group 3 146
6.2.5 Group 4 148
6.2.6 Group 5 150
6.2.7 Group 6 152
6.2.8 Group 7 154
6.2.9 Group 8 156
6.2.10 Group 9 158
6.2.11 Group 10 160
6.2.12 Change in Crime Profiles Over Time 162
6.3 Comparisons with Other Classifications 162
6.3.1 The Gains Chart 163
6.3.2 Effect of Scale on Gains Analysis 164
6.3.3 Gains Comparisons by Crime Type 165
6.3.4 Disaggregating Gains 167
6.3.5 Comparisons of Reductions in Variability of Crime Rates 172
6.4 Correlation Between Crime and Membership Strength 174
6.5 Neighbourhood Isolation 175
6.5.1 Crime Contextualised Neighbourhood Isolation 177
6.6 Concluding Remarks 178

7 Exploratory Analysis of Criminal Damage **181**
7.1 Introduction 181
7.2 Crime Recording Issues 182
7.3 Exploratory Data Analysis 182
7.3.1 Introduction 182
7.3.2 Between-Group Comparisons of Criminal Damage 183
7.3.3 Within-Group Comparisons of Criminal Damage 184
7.4 Explaining Crime Patterns 196
7.4.1 Low Membership Strength Coefficients 196
7.4.2 Unique Neighbourhood Features 200
7.4.3 Geographical Circumstances 206
7.5 Concluding Remarks 208

8 Evaluating Policy: Burglary and BRIL **211**
8.1 Introduction 211
8.2 Counting Rules and Rate Calculations 211
8.3 Evaluation Rationale 212
8.4 Testing Classification Power 214
8.4.1 Difference in Burglary Rates 214
8.5 Geographies of Burglary Dwelling Offences 217
8.5.1 Supra-district Trends in Burglary Dwelling 218
8.5.2 Burglary Dwelling in Leeds 219
8.6 Other Burglary Dwelling Geographies 223

8.6.1 Stolen Property 223
8.6.2 Nominals 225
8.6.3 Repeat Victimisation 231
8.7 Impact of Burglary Reduction Initiatives 235
8.7.1 Burglary Reduction Initiative in Leeds (BRIL) 237
8.7.2 Types of Intervention 238
8.7.3 Links Between Crime Rates and Target Hardening Activity 238
8.7.4 Repeat Victimisation Evaluation 239
8.7.5 Impact of BRIL on Repeat Victimisation 242
8.7.6 Costs of not Reducing Repeat Victimisation 245
8.8 Concluding Remarks 246

9 Conclusions **251**
9.1 Summary of Research Achievements 251
9.1.1 Classifying Other Study Areas 257
9.1.2 Neighbourhood Dynamics 258
9.1.3 Interpretation Issues 259
9.2 Possibilities for Further Work 259
9.2.1 Existing Classification 260
9.2.2 New Classifications 260

A Maps of Sub-Community Area (SCA) Geography **263**

B Full List of Variables Used **267**

Bibliography **271**

List of Tables

3.1 Statistical profiles of disadvantaged areas with high and low levels of burglary, Merseyside (Hirschfield and Bowers, 1997a, page 112) 48

3.2 Membership of CDRP Family 4, 2001. 52

3.3 CDRPs most similar to Leeds, 2004. 53

4.1 Variables identified at NCSN Conference, Belfast, 2003. 61

4.2 Variables identified at Leeds Statistics Scoping Event, 2003. 62

4.3 Survey of neighbourhood variables used in community safety studies 64

4.4 Comparison of different census geographies in Leeds. 73

4.5 Classification variables selected, or derived, from the 2001 census. 74

4.6 Variables used in construction of Index of Ethnic Heterogeneity. 75

4.7 Variables identified from Ordnance Survey MasterMap. 80

4.8 Examples of footprint sizes for typical buildings. 83

4.9 Value of property stolen in cases of Theft From a Shop, ranked by total value of items stolen. 86

4.10 Variables identified from Nomis and the Annual Business Inquiry. 87

4.11 Variables identified from Leeds City Council departments and services. 88

4.12 Noise complaints received by Environmental Health Services. 90

4.13 Summary of domestic violence incidents in Leeds, 2001/02 93

4.14 Proportion of BCS estimate of all crime reported to the police and recorded by them (comparable subset of crimes), year to September 2003 (Dodd et al., 2004, from). . . 94

4.15 Percentage impact of NCRS on recorded crime in West Yorkshire, 2002/03 (Simmons and Dodd, 2003, adapted from) 95

4.16 Variables identified from West Yorkshire Police. 97

4.17 Variables identified from West Yorkshire Fire and Rescue Service. 98

4.18 Classification of ICD-10 codes into community safety super-classes 99

4.19 Primary and secondary addiction types for patients in treatment, 2001/02 to 2003/04. 101

4.20 Variables identified from Leeds PCTs. 101

5.1 Highly correlated variables 112

5.2 Problem variables identified by descriptive statistics 112

5.3 Problem variables identified by Principal Component Analysis. Eleven components (factors) used. 113

5.4 Subjective interpretation of the Silhouette Coefficient (SC), defined as the average silhouette width for the entire dataset (Kaufman and Rousseeuw, 1990, page 88) . . . 116

5.5 Cluster metrics for partitions where k=8 and k=10 117

5.6 Measures of central tendency and spread (values in bold represent the better outcome) 122

5.7 Level of agreement between neighbourhood classifications and a crime-only partition. 124

5.8 Cluster statistics for k-means solution generated with ClustanGraphics 125

5.9 Level of agreement between partitions created using k-means and the crime-only partition, $k = 10$. 126

5.10 Measures of central tendency and spread of the k-means $k = 10$ partition 126

5.11 Measures of central tendency and spread of the fuzzy c-means $k = 10$ hard partition . 127

5.12 Level of agreement between hard partition created using fuzzy c-means and the PAM $k = 10$ and crime-only partitions. 127

5.13 Strength coefficients for a selection of neighbourhoods that are not well fitted to any one cluster. 129

5.14 Reduction in variability of neighbourhood variables from fuzzy c-means (hard) and Townsend classifications 137

6.1 Main data sources used by classifications being compared to neighbourhood classification for community safety 163

6.2 Effect on rankings of penetration indices as a result of disaggregating *burglary elsewhere* offences, 2003/04. 171

6.3 Percentage reductions in variability for different crime types for different classifications 173

6.4 Reduction in variability of disaggregated crime variables from LCCS and Townsend classifications 174

6.5 Results of grouping observation - non-geographical (solution 1) and geographical (solution 2). 176

6.6 Spearman Rank correlations between crime types and level of crime-context isolation 178

7.1 Criminal damage rates (disaggregated) and partition diagnostics, including cluster strength coefficients (SC), for a selection of problematic neighbourhoods 193

7.2 Cohen's (Cohen, 1973) typology of vandalism 206

8.1 Pairs of neighbourhood groups where there was no significant difference between burglary rates in 2003/04, $\alpha = 0.05$. * denotes significant difference, $\alpha = 0.05$, in previous years. Source: WYP. 217

8.2 Breakdown of nominals and burglaries to which they have been connected, by group, for period 2001/02 to 2003/04. Source: WYP. 227

8.3 Breakdown of offences (percentages) for nominals linked to burglaries in different groups, for period 2001/02 to 2003/04. Source: WYP. 227

8.4 Comparison of percentages of offences by number of victimisations. Data marked n.a. is not available. Source: WYP and Home Office. 232

8.5 Trends in repeat victimisation in Leeds. Source: WYP. 232

8.6 BRIL funding from 2002/03 to 2004/05. Adapted from (Burden, 2005, page 17). . . . 243

8.7 Estimate of extra repeats that would have occurred had target hardening not taken place.246

B.1 Variables used during the design of the classification (not all of which were included in the final cluster analysis). Abbreviations (ABI=Annual Business Inquiry, LCC=Leeds City Council, PCT=Primary Care Trusts, WYFRS=West Yorkshire Fire and Rescue Service, WYP=West Yorkshire Police) 269

List of Figures

2.1 The crime triangle. Adapted from Clarke and Eck (2003) 15

3.1 Choropleth map of burglary dwelling, by neighbourhood (2003/04). 35
3.2 STAC standard deviational ellipses of Other Theft (1999-2002). 36
3.3 Kernel estimation of a point pattern. (based on Gatrell et al., 1996, page 260) 37
3.4 Single kernel density estimation of domestic burglary (1999-2002), τ=400m (fixed), normal kernel function. 38
3.5 Single kernel density estimation of domestic burglary (1999-2002), adaptive bandwidth (min. offences, 100), normal kernel function. 38
3.6 Dual kernel density estimation of domestic burglary (1999-2002), τ=500m, normal kernel function. 39
3.7 Density estimation comap of robbery incidents (2003/04) conditioned by day of week and hour of day. Number of incidents is shown in top right hand corner of each panel. 42
3.8 Standardised burglary rate by neighbourhood (excluding city centre), 2003/04. 46

4.1 Fitting the mental map of Chapeltown to the 2001 census output area geography . . . 68
4.2 Super Output Area Lower Level claimant count estimates based on disaggregation of postal sector-level counts. 70
4.3 Surface of public house density created with kernel density estimation (bandwidth=500m). 71
4.4 Pub density in sub-community areas standardised by population size. 71
4.5 Difference in statistical and geographic detail of different 2001 census outputs. Source: ONS. 72
4.6 Defensible space in Potternewton. Source: Ordnance Survey. 81
4.7 Differences in extent of multi-surface in contrasting residential areas. 82
4.8 Histograms of building footprint sizes. 82
4.9 Land use in SCA 1.01 (Adel). 84
4.10 Shop density in Pudsey town centre, overlayed with locations of shops identified in OS AddressPoint. 86
4.11 Numbers of asylum seekers arriving in Leeds per quarter, from Q1 2000 to Q2 2004. 91
4.12 Map showing actual locations of crimes originally recorded as having occurred at the Village Hotel, Otley Road, Leeds 96

5.1 Map of the number of variables for which sub-community areas (SCAs) have values over three standard deviations from the mean. 111
5.2 Average silhouette widths for different values of k 117
5.3 Comparison of partitions with respect to social renting 118

5.4 Comparison of partitions with respect to more affluent sub-urban neighbourhoods . . 118
5.5 Boxplots of all crime, arson and burglary rates for $k = 8$ and $k = 10$ partitions 120
5.6 Boxplots of criminal damage, other theft, vehicle crime and violent crime rates for $k = 8$ and $k = 10$ partitions 121
5.7 Geography of membership strength coefficients used to produce fuzzy c-means hard clustering. 128
5.8 Geography of silhouette widths in the PAM $k = 10$ partition. 128
5.9 Geography of membership strength coefficients for c-means group 4, the 'student' group. 129
5.10 Bar graph showing importance of individual variables to the definition of cluster 1 . . 130
5.11 Scree plot of HINoV TOPRI values. 132
5.12 Silhouette plot and group membership information of the final c-means $k = 10$ hard partition 134
5.13 Geography of group membership of the final c-means $k = 10$ hard partition 135

6.1 Geography and membership strength coefficients of neighbourhoods in group 1. . . . 142
6.2 Standardised scores for variables defining group 1. 143
6.3 Standardised scores for crime variables, for 2003/04, for group 1. 143
6.4 Change in group 1 standardised crime rates, 2002/03 to 2004/05. 143
6.5 Geography and membership strength coefficients of neighbourhoods in group 2. . . . 144
6.6 Standardised scores for variables defining group 2. 145
6.7 Standardised scores for crime variables, for 2003/04, for group 2. 145
6.8 Change in group 2 standardised crime rates, 2002/03 to 2004/05. 145
6.9 Geography and membership strength coefficients of neighbourhoods in group 3. . . . 146
6.10 Standardised scores for variables defining group 3. 147
6.11 Standardised scores for crime variables, for 2003/04, for group 3. 147
6.12 Change in group 3 standardised crime rates, 2002/03 to 2004/05. 147
6.13 Geography and membership strength coefficients of neighbourhoods in group 4. . . . 148
6.14 Standardised scores for variables defining group 4. 149
6.15 Standardised scores for crime variables, for 2003/04, for group 4. 149
6.16 Change in group 4 standardised crime rates, 2002/03 to 2004/05. 149
6.17 Geography and membership strength coefficients of neighbourhoods in group 5. . . . 150
6.18 Standardised scores for variables defining group 5. 151
6.19 Standardised scores for crime variables, for 2003/04, for group 5. 151
6.20 Change in group 5 standardised crime rates, 2002/03 to 2004/05. 151
6.21 Geography and membership strength coefficients of neighbourhoods in group 6. . . . 152
6.22 Standardised scores for variables defining group 6. 153
6.23 Standardised scores for crime variables, for 2003/04, for group 6. 153
6.24 Change in group 6 standardised crime rates, 2002/03 to 2004/05. 153
6.25 Geography and membership strength coefficients of neighbourhoods in group 7. . . . 154
6.26 Standardised scores for variables defining group 7. 155
6.27 Standardised scores for crime variables, for 2003/04, for group 7. 155
6.28 Change in group 7 standardised crime rates, 2002/03 to 2004/05. 155
6.29 Geography and membership strength coefficients of neighbourhoods in group 8. . . . 156

6.30 Standardised scores for variables defining group 8. 157
6.31 Standardised scores for crime variables, for 2003/04, for group 8. 157
6.32 Change in group 8 standardised crime rates, 2002/03 to 2004/05. 157
6.33 Geography and membership strength coefficients of neighbourhoods in group 9. . . . 158
6.34 Standardised scores for variables defining group 9. 159
6.35 Standardised scores for crime variables, for 2003/04, for group 9. 159
6.36 Change in group 9 standardised crime rates, 2002/03 to 2004/05. 159
6.37 Geography and membership strength coefficients of neighbourhoods in group 10. . . 160
6.38 Standardised scores for variables defining group 10. 161
6.39 Standardised scores for crime variables, for 2003/04, for group 10. 161
6.40 Change in group 10 standardised crime rates, 2002/03 to 2004/05. 161
6.41 Gains chart of total crime versus adult population. Data is taken from Ashby and Longley (2005) 164
6.42 Gains chart of violent crime versus adult population for different administrative geographies and different levels of a hierarchical classification. 165
6.43 Gains column graphs of violent crime versus adult population for different levels of the ONS hierarchical classification of output areas. (Columns are fixed width, not sized proportional to % population). 166
6.44 Gains charts for different crime types, as discriminated by different classifications. . 167
6.45 Gains chart of *other theft* using the IoD 2004 classification. 169
6.46 Gains charts for disaggregated crime types, as discriminated by the Leeds Classification for Community Safety (LCCS). 170
6.46 Histogram of burglary dwelling offences for different geographies. 172
6.47 Geographical grouping of the data shown in Table 6.5. 176
6.48 Measuring relative isolation in crime rate terms. 177

7.1 Local and national trends in Criminal Damage rates, 2000/01 to 2003/04. 183
7.2 Trends in criminal damage rates, 2000/01 to 2003/04, by cluster. 184
7.3 Anatomy of a boxplot. 186
7.4 Boxplot of Damage to Dwelling rates, by cluster (outlier neighbourhoods labeled). . 187
7.5 Boxplot of Damage to Motor Vehicle rates, by cluster (outlier neighbourhoods labeled).187
7.6 Typology of relationships between crime rates and trends, with respect to a cluster mean ($\bar{C}$). 188
7.7 Scatter plots of standardised Damage to Dwelling offence rates 2003/04 against change in rate between 2001/02 to 2003/04, standardised by group. 189
7.8 Scatter plots of standardised Damage to Motor Vehicle offence rates 2003/04 against change in rate between 2001/02 to 2003/04, standardised by group. 190
7.9 Map of Damage to a Dwelling rates, 2003/04, standardised by group. 194
7.10 Map of Damage to Motor Vehicle rates, 2003/04, standardised by group. 194
7.11 Map of change to Damage to Dwelling rates, 2001/02 to 2003/04, standardised by group. 195
7.12 Map of change to Damage to Motor Vehicle rates, 2001/02 to 2003/04, standardised by group. 195
7.13 Trends for major offence types, 2000/01 to 2003/04. 197

7.14 Mapping SCA 43.02 and its environs. 199
7.15 Map of Damage to Dwelling offence locations, 2003/04. 199
7.16 Plan by R.H.A. Livett, Housing Director, showing plans for Halton Moor housing estate, 1938. Source: Leodis photographic archive. 200
7.17 Trends for major offence types, 2000/01 to 2003/04. 201
7.18 Trends for major offence types, 2000/01 to 2003/04. 202
7.19 Comparison of values of variables in SCA 40.02 with distribution of values within cluster 7. 202
7.20 Trends for major offence types, 2000/01 to 2003/04. 203
7.21 Comparison of values of variables in SCA 70.01 with distribution of values within group 1. 204
7.22 Location of criminal damage offences in New Wortley, 2003/04. 204
7.23 Trends for major offence types, 2000/01 to 2003/04. 207
7.24 Comparison of values of variables in SCA 47.01 with distribution of values within cluster 7. 208

8.1 Comparison of burglary dwelling rates between members of CDRP 'Family 4'. Source: Home Office. 218
8.2 Kernel density hotspot map of burglary dwelling offences, 2003/04. Source: WYP. . 219
8.3 Map of burglary dwelling rates, 2003/04, standardised by group. Source: WYP. . . . 220
8.4 Maps of standardised burglary dwelling rates, 2003/04, for individual groups. Source: WYP. 221
8.5 Map of change in burglary dwelling, 2002/03 to 2003/04, standardised by group. Source: WYP. 222
8.6 The mean value of property stolen during burglary dwelling offences in 2003/04, by group. Source: WYP. 224
8.7 Frequency with which different types of property were stolen during burglary of dwellings during 2003/04, by group. Source: WYP. 226
8.8 Concentration of unique nominals living outside group 6 who are associated with burglaries inside group 6 (n=292). Source: WYP. 230
8.9 Concentration of unique nominals living outside group 2 who are associated with burglaries inside group 2 (n=142). Source: WYP. 230
8.10 Trends in burglary dwelling victimisation by group, 2000/01 to 2003/04. Source: WYP.234
8.11 Trends in burglary dwelling repeat victimisation rate by group, 2000/01 to 2003/04. Source: WYP. 235
8.12 The classic experimental design. 235
8.13 Variation in target hardening activity, by group, from April 2002 to March 2003. Source: WYP and the BRIL. 242
8.14 Variation in response to target hardening, by group, from April 2002 to March 2003. Source: WYP and the BRIL. 243

A.1 Original geography of 106 communities (generalised), derived from the Leeds City Council community mapping project. 264
A.2 Sub Community Area (SCA) geography (not generalised). 265

Chapter 1

Introduction

1.1 Area Classification for Community Safety

Central to the concept of community safety in the UK is the acknowledgment that reducing crime and the fear of crime must be a joint responsibility for all local authorities and agencies, and not just the police. Communities themselves, and the individuals and businesses therein also have a responsibility to help reduce crime, and the task of coordinating crime reduction activity has become the responsibility of local community safety partnerships. Such partnerships have existed in some parts of the country for a couple of decades or more, but for most people working within the community safety domain, the partnership approach now being used is that which is laid out in the Crime and Disorder Act 1998.

Thus, the delivery of the community safety agenda now requires that agencies such as the local authority, the fire service, the police and primary care trusts work more closely together to try and reduce crime. For some agencies this has always been a primary objective, but for others, such as the health services, crime reduction is a much lower priority for the organisation, although for services such as community health and treatment for drug addiction, the links with criminality are easier to appreciate.

The role of the data analyst has also been affected by partnership working. In the past, the police intelligence analysts would have concerned themselves with analysing policing problems using police data. The same would have been true for the health analysts, and those working within individual local authority departments. For example, police analysts may have analysed stolen vehicle recovery sites, but may not have linked up with fire service to look at the locations of car fires or taken into consideration the geography of abandoned vehicles reported to the local authority. To some extent this is still the case, but partnership working has also placed additional responsibility on analysts to think more widely about the consequences of their findings for other partner agencies. This may be achieved by either sharing findings more widely, or sharing raw data more widely and developing new skills and expertise with which to probe it.

This is the situation from which the thesis has emerged: a need to analyse data provided from a variety of community safety agencies and identify the extent to which locations may, or may not require interventions to reduce crime; and a need to consider the nature of these places and appreciate whether high or low crime rates might be expected given the neighbourhood context. It has been long understood that there are geographies of crime and that offenders and offences show marked tendencies to concentrate in particular types of location and these features are well documented (Herbert, 1982). Furthermore, it has also been shown that there is no uniformity in the spatial impacts of laws and justice (Harries and Brunn, 1978).

A range of techniques for identifying *crime hotspots* have been, and continue to be, developed by academics, sometimes specifically tailored to crime data (e.g. Bowers et al., 2004), but often applicable to other problems, such as epidemiology (e.g. Gatrell et al., 1996). Other academic studies demonstrate the usefulness of these techniques, and a point has been reached where quite sophisticated spatial analysis techniques have now been packaged within easy-to-use computer software (Levine, 2002; Anselin et al., 2006) and are being used by practitioners of community safety analysis.

Thus, the research task is no longer to prove the existence of spatial patterns but rather of examining their significance and of understanding the processes which produce them (Evans et al., 1992, page 2). More recently, dissatisfaction with the increasingly ritualistic use of hotspot techniques by many practitioners has led Laycock, for example, to call again for a move *beyond blobology* and more focus on context as well as increased recognition for crime science as a professional discipline. Thus, while the hotspot techniques are useful, there is now resurgence of interest among practitioners in understanding more about the social, economic, demographic and spatial context of crime hotspots. In practical terms, different contexts may affect decisions about the types of situational crime prevention that might be deployed. In other situations, understanding of the context of crime may help focus activity to address underlying causes of crime, such as drug addiction, alcohol misuse, housing abandonment and poverty. Furthermore, at around the same time, academics with an interest in area profiling and classification have begun to find a wider audience for their ideas, and demonstrations using commercial geodemographic products for crime analysis have begun to be published and presented more frequently (e.g. Ashby and Longley, 2005). As such, academic arguments for acknowledging the spatial context of crime are now widely accepted by the practitioner community, and research is now being scrutinised by practitioners for techniques, methods and tools with which to conduct the contextual analysis.

Crucial to the reception of these new ideas has been the ability of commercial geodemographic products to provide contextual information at a small-area scale. Classifications designed for the Home Office specifically to analyse differences in crime levels have been around for several years, but these have always remained at police force, local authority and police Basic Command Unit (BCU) scales, and have thus been of little use to practitioners wanting to reflect the new localism in government discourse by analysing the context to crime patterns at a neighbourhood or place-based community scale.

The aim of this thesis is to contribute to the debate around using neighbourhood-level area classifications for community safety analysis by researching and developing a task-specific classification for community safety at a neighbourhood scale for the local authority district of Leeds. As such, the work is a development of the ideas behind the classifications produced for the Home Office (Harper et al., 2002), and a challenge to the current preference among many for using commercial geodemographic products, the internal design of which little has been revealed.

That is not to say that revealing design decisions would necessarily prevent discontent about the veracity of a classification. For most practitioners, the principle use of the Home Office classifications has been to assess local crime reduction performance with that of peers. So emotive is the subject of performance within community safety partnerships, that any tool or metric used for monitoring is scrutinised in great detail - especially when comparative performance looks poor.

At present, the Home Office is not proposing to monitor performance at a neighbourhood scale, although that position might change. Thus, for the time being, the precedents of the classification designed for this thesis are tools for local-area description, such as The Profiler (Hirschfield and Bowers, 1997a). This tool and others like it produce area summaries from a wide range of data sets. The portraits that are produced are designed to be considered once crime hotspots have been identified, as an aid to finding explanations for particular crime problems. For example, what proportion of households are derelict, what proportion of the population are unemployed, what proportion of households are rented by students? Where more than one hotspot is found, likenesses can be sought in the respective area portraits. Similarly, if a number of areas are under scrutiny because of crime differences, area portraits can be generated to try and identify what it is that might be unique about the different locations.

Such area profiling first became practical with the publishing of small-area data from population censuses in an electronic form that could be manipulated by computers. Likewise, this thesis was conceived to coincide with the publication of the first results from the 2001 census of population. At the outset, it was also recognised that an increasing amount of data were being generated and maintained by different agencies within the local community safety partnership, and that this was probably under-exploited. The collation of locally-produced data sets for this research coincided with other local initiatives to study neighbourhood context and crime patterns. As such the research was able to contribute more widely to community safety analysis within the study area, as well as profit from increased interest and activity with regards to data sharing.

The principle purpose of researching neighbourhood classification, however, has been to demonstrate ways in which the grouping of similar neighbourhoods into distinct groups can help to analyse crime patterns and responses to crime prevention initiatives. The studies using commercial geodemographic classifications have tended to demonstrate analysis of between-group differences in crime, although it will be argued that findings are of limited utility. More interesting, it will be argued, are the possibilities of identifying different responses to initiatives by neighbourhood type, as hinted at in recent analysis of the effects of burglary reduction schemes (Hirschfield, 2004).

Analysis of crime patterns within-groups are less common in the existing literature, although to some extent this type of investigation is most interesting as it relies upon and tests the central notion behind geodemographics and area classification - that members of a given group will display similar characteristics and behaviours, to which, for the purposes of this thesis, might be added crime victimisation characteristics.

1.2 Aims and Objectives

As stated above, the main aim of this thesis is to contribute to the debate surrounding the usefulness of using area classifications for crime pattern analysis at a neighbourhood scale. This is achieved, in this instance, through the construction of a new neighbourhood classification for community safety for Leeds. Discussion and transparency regarding classification design decisions are an important part of the main aim and ensure that readers are able to scrutinise the methodology and thus make informed judgments about whether a similar approach might be useful for classifying different study areas. Practical examples that use the classification to analyse a number of local community safety datasets also help identify strengths and weaknesses of the approach. This, it is argued, has not always been paid sufficient attention by academic geographers, but is vital to knowledge transfer from academia to the police and local authority users of the research, itself an implicit aim of the ESRC CASE award that has made this research possible.

To achieve these aims, the following research objectives were pursued,

1. To review the criminology and community safety policy literature in order to better appreciate how geography can affect levels and patterns of crime, and why police and other community safety officers analyse crime patterns in the way that they do.

2. To review the geography literature to discover the techniques and approaches that have been developed to aid crime mapping and area profiling.

3. To assemble a database containing a wide variety of neighbourhood variables related to community safety issues.

4. To employ cluster analysis routines to create a partition of Leeds' neighbourhoods.

5. To use a range of visual aids and metrics to paint group portraits and thus turn the partition into a finished task-specific classification that can be compared with general purpose classifications.

6. To conduct a number of case studies in which the classification is used to analyse patterns of crime and crime reduction interventions.

1.2.1 The Study Area

The research has been funded by the Economic and Social Research Council in collaboration with the Leeds Community Safety Partnership (LCSP), since renamed the Safer Leeds partnership. Hence, the study area chosen for the research covers the Leeds Metropolitan District. A consequence of this is that the classification has to encompass neighbourhoods of a range of different types within the main urban core of Leeds city, as well as small towns and villages in the rural hinterland.

1.3 Thesis Structure

The thesis is organised into ten chapters, the sequence of which reflect the development of the research over time and the achievement of the objectives outlined above.

Chapter 2 begins by reviewing the academic and policy literature on criminology and community safety, and attempts to describe the context in which analysts find themselves considering geographies of crime. A number of criminological theories with distinct geographical aspects are discussed, as are some of the key ideas that have emerged from the development of a distinct environmental criminology. Policy approaches are reviewed since the major preoccupation at the Home Office in the 1980s with situational crime prevention, culminating in a discussion of the significance of the Crime and Disorder Act 1998, both in its original form and in light of the review conducted in 2005. The review is concluded with a summary of theoretical and policy issues likely to impact upon the research and discusses where the research is situated within academic debates.

Chapter 3 moves on to describe and discuss the relative merits and applications of a range of spatial analysis techniques that have been applied to crime problems. Gloria Laycock's 'beyond blobology' address is then introduced and used to discuss the different approaches adopted by geographers to take account of the importance of populations and neighbourhoods when analysing crime risks. The approaches range from the standardisation of crime counts by population size, through to existing efforts to employ geodemographic classifications. Drawing upon this existing research, arguments are made for design objectives for the new Leeds Classification for Community Safety (LCCS).

With an outline for the classification drawn, Chapter 4 discusses the variables selected to construct the classification. A number of different surveys were used to garner opinion about the suitability of different variables, and the results of these are presented. A description is given of different national and local data sources, and practical problems in gaining access to data sources are discussed. After a discussion of variable preparation techniques, the groups of variables are described in turn, with additional justifications for their inclusion. A description of the creation of the neighbourhood geography used for the classification is also given.

Chapter 5 begins by demonstrating how the variables that were selected in the previous chapter have been subjected to a number of different tests in order to weed out variables that would be likely to offer little in the way of discriminating power or were highly inter-correlated with other variables. Some theoretical background to cluster analysis is then presented, followed by a narrative of how different cluster analysis routines were tested and partitions interpreted and compared. This discussion culminates in a final classification with which to proceed with practical testing, although ways in which the classification production process might be improved are also discussed.

With a classification constructed, Chapter 6 'paints' group portraits by profiling the extent to which different variables contributed to the 'shape' of the group as a whole. In a similar vein, the uniqueness of the crime profiles for each of the groups are also discussed. Numerical tests are then applied to the new classification and alternative classifications in order to determine whether a task-specific classification approach can discriminate crime rates better than a general purpose classification.

Early on in the research, it was decided that to demonstrate clearly the capabilities and limitations of the classification it would be necessary to apply it in a number of case studies. The first of these is discussed in Chapter 7, and centres around the analysis of patterns of criminal damage. During this analysis, consideration is given as to whether aspects of the classification do appear to offer tentative explanations of differences in crime incidence, but equally, the possibilities of mis-classification and problems associated with the new neighbourhood geography are also considered.

Chapter 8 presents the second case study, an evaluation of a target hardening initiative for victims of domestic burglary. Here, the classification is used to analyse whether responses to the intervention differ by neighbourhood type. The nature of the problem precludes a within-group analysis, but the between-group analysis produces interesting findings with which to pose new hypotheses, and when combined with the results of a cost benefit analysis, suggests how this crime reduction initiative has benefited different types of neighbourhood and the communities therein.

The thesis concludes in Chapter 9 with a summary of the main points of the research and a discussion of the use of area classifications for crime pattern analysis at a neighbourhood level. Some limitations of the research are put forward and issues regarding the implementation of the classification within a CDRP are discussed. The chapter ends with a series of suggestions for future research and reflections on the approach adopted to the research.

1.4 Outputs

To date, the research have been presented at a number of academic and practitioner conferences, including:

Shepherd, P. (2003) Neighbourhood profiling. Workshop delivered at the National Community Safety Network Conference, Belfast, June 2003.

Shepherd, P., Stillwell, J. and Clarke, G. (2005) Identifying crime hotspots using a neighbourhood classification. Poster presentation to the 3rd National Crime Mapping Conference, London, April 2005.

Shepherd, P. (2005) Identifying patterns of crime: A neighbourhood classification approach. Presentation to the British Society of Criminology Conference, Leeds, July 2005.

Shepherd, P., Stillwell, J. and Clarke, G. (2005) Neighbourhood profiling and classification for community safety. Presentation to the RGS-IBG Annual International Conference 2005, London, August 2005.

An article based on the research has also been published:

Shepherd, P. and Stillwell, J. (2006) A neighbourhood classification for community safety. *The Yorkshire and Humber Regional Review* 15(2), pages 12-14.

In addition to producing an executive summary of the research for Safer Leeds, it is planned to produce a couple of articles for publication in academic journals. One of these will report on the production of the LCCS and the techniques used. The second will be more applied in nature and demonstrate the utility of using an area classification for crime pattern analysis.

Chapter 2

Geography, Crime and Community Safety

2.1 Introduction

The contemporary practical application of criminological theories to the analysis of crime and disorder in the UK has become rooted in theories which have supported the situational approach to crime prevention fashionable at the Home Office since the 1980s. Theoretical ideas are intertwined through increasingly routine and standardised forms of crime analysis being demanded, sometimes by statute, of police forces and local authority crime prevention units.

Hence, these theoretical and political environments shape much of the practical day-to-day analysis of crime and disorder. This thesis will argue that these analyses can be improved by giving a thorough acknowledgment to the importance of neighbourhood in studies of the geography of crime at a local level. For this reason, the following review begins by discussing theories from the field of criminology that are pertinent to the research. As appropriate, links are then made between theoretical work and the implementation of crime prevention and community safety policies, and some of the history and mechanics of (community safety) partnership working are reviewed to provide a context for the demands of government to share and analyse crime and disorder information locally. Finally, a couple of important, centrally directed, data sharing and community safety analysis projects are discussed to show how community safety policy implementation needs geographical analysis and area profiling, which will be discussed in Chapter 3.

2.1.1 The Development of a Geography of Crime

Neighbourhood profiling and classification for community safety is undertaken on the understanding that there is an interplay between crime, space and society. Criminology can provide theoretical ideas about how this interplay operates, and those interested in the geography of crime will often use empirical analysis of offenders, offences and the effects of crime to test different theoretical positions (Fyfe, 2000). The sub-discipline of the geography of crime is a relatively recent development that took shape at a similar time as environmental criminology emerged as a sub-discipline of criminology. Indeed, it has been pointed out that the mutual interests of these sub-disciplines has meant geographers have become directly involved in professional criminology (Lowman, 1986).

The spatial and ecological perspectives on crime developed in a number of different phases, over time. The nineteenth-century French cartographic school is often taken as a starting point, and Guerry and Quételet's work in the 1830s and 1840s showed that at the *département* level, crime was far from homogeneously distributed across the country. Developments included the 'ecological' work of the Chicago School of Sociology in the 1920s and 1930s (e.g. Shaw, 1938; Shaw and McKay, 1942), and the emergence of factor analysis (e.g. Lander, 1954) and social area analysis (e.g. Shevky and Bell, 1955) in the 1950s and 1960s.

Both social area analysis and factor analysis became possible because of the publication of new census data at small-area scales. Census data in the UK had been available for wards for many years but it was only in 1961 that data was published at the smaller enumeration district level. Shevky and Bell's social area analysis (Shevky and Bell, 1955) linked the changing urban social structure and residential patterns to economic development and urbanisation processes. They used a relatively small number of census variables to capture social rank, urbanisation and segregation. However, this drew criticism (Baldwin and Bottoms, 1976) because of the small number of variables and significantly because the variables chosen used *a priori* theorisation (Baldwin, 1979). Factor analysis, and related multivariate statistical methods, take in a much wider set of variables. As Robson (1969, page 58) argued:

> "Whereas the Shevky technique selects its constructs, and the variables which compose them, on the basis of possibly suspect theory, multivariate analysis selects its discriminating factors solely on the basis of the intercorrelations of the data itself - and a large body of data at that".

In the 1970s and 1980s came the development of environmental criminology (Mayhew et al., 1976; Brantingham and Brantingham, 1981) and studies of the geography of crime (e.g. Harries, 1974; Pyle et al., 1974; Herbert, 1976, 1982) as distinct research domains but neither were the exclusive preserve of criminologists or geographers. Since then, environmental criminology has tended to follow a technical path, using advances in spatial analysis made possible by increasingly sophisticated and accessible GIS techniques. Research into the geography of crime has in part followed the same course, with much research focusing on statistical (e.g. Herbert, 1994; Wong, 1997) and spatial analysis (e.g. Ceccato et al., 2002; Santiago et al., 2003) and cartographic research (e.g. Craglia et al., 2000). In addition to technical methodology development, however, there have also been developments of more critical and cultural themes. Fear of crime has received considerable attention (e.g. Smith, 1987; Pain, 2000), as has gender (e.g. Pain, 1991; Valentine, 1992; Pawson and Banks, 1993). The importance of wider social processes and forces, such as consumption (e.g. Hayward, 2004), housing and urban regeneration (e.g. Hancock, 2001) have been argued and issues particular to countryside and rural areas (e.g. Burgess, 1996; Yarwood, 2000) have also been investigated. Young peoples' experiences of, and involvement in, crime (e.g. Valentine, 1996; Collins and Kearns, 2001; Carr, 2003; Nayak, 2003) have also received an increasing amount of attention from geographers.

2.2 Crime and Criminology

Whilst acknowledging the conceptual and technical problems associated with defining crime, the definition used throughout this thesis is that crime is concerned with law-breaking behaviour. It should always be born in mind that the law is something that is constantly evolving and thus constantly redefining what constitutes a crime (Crawford, 1998). Yet, by defining crime in this way it can be treated as a discrete phenomenon and a separate and easily distinguishable (using officially recorded crime data) behavioural entity (Lowman, 1986). Accordingly, Home Office statistics show that during recording period 2004/05 (1st April to 31st March), there were 5.5 million crimes recorded in England and Wales, representing a 19% increase since 1996/97 but a 6% retreat from the highpoint of 5.9 million crimes in 2003/04. Within Leeds, the study area chosen for this research, the trend is similar. In 2004/05 there were 125,000 crimes recorded across the Leeds Metropolitan District, which represents a fall of 2.5% from the peak in 2003/04 but is still 28% higher than levels in 1997.

This type of numerical summary is a commonplace representation of crime levels, however, it could be argued that it belies certain complexities. For example, it is acknowledged that a great deal of crime is not officially recorded and that there is a social bias toward the recording of crimes perpetrated by the lower social classes. Crime can be found at every level of society and the more powerful the class, the greater the potential harm of its offending behaviour (Taylor et al., 1975). For this type of reason, it has been argued that greater attention be paid to social processes and social control (Lowman, 1986).

Also related to under-recording are incidents which, although most probably criminal, have not been drawn to the attention of the police. This non-reporting may be the result of victims not coming forward, perhaps because they feel the police are impotent or because the offence is seen as too trivial (Dodd et al., 2004). There can also be institutional reasons for non-reporting or non-recording, often bound up with issues to do with standardised reporting procedures and balances of proof. Arson is a good example of this problem. The fire service may feel able, based on their experience, to describe a fire as having been started deliberately even though there may be insufficient evidence for the police to feel confident enough to record the incident as arson. These particular difficulties in defining crime come from a rather traditional perspective of crime that is preoccupied with the criminal actor or criminal event.

Terms such as low-level nuisance, anti-social behaviour (ASB) and minor incivilities are used separately and also interchangeably throughout the thesis to refer to those problems which go unrecorded as crimes but which are believed to constitute a threat to community safety, and thus require analysis and monitoring. These types of behaviour can be identified using a number of different official sources including the local council, health authorities and the police, and were tested with a one day count across England and Wales in 2003 (Harradine et al., 2004). It could be argued that these non-criminal behaviours are equally important as law-breaking acts as they may be perceived to be equally serious and a possible precursor to criminal behaviour and neighbourhood decline (Skogan, 1990). Such 'anti-social' behaviour may also be the cause of public fear and anxiety - themselves problems which UK community safety policy has to address (Home Office, 1998a).

2.2.1 Criminological Theory

Against this background, the discipline of criminology searches to explain why crime occurs. Although still a relatively new discipline (Garland, 1988), criminological theories abound and sometimes compete and contradict with each other to explain the behaviour of criminals, the criminality of behaviour and the criminality of the state (Walklate, 1998). Some of the disagreement and lack of consensus within criminology is due to the broad range of subject areas represented - from psychology to sociology to biology - but then, as with geography, different and competing philosophical points of view are also represented. The importance of criminological theory to the research reported in this thesis is twofold. Firstly it can help identify aspects of neighbourhood that might benefit from being captured and included within a neighbourhood classification for community safety. Secondly, it can help provide explanations for results obtained from empirical studies using the neighbourhood classification that will be presented in later chapters.

Understanding Criminal Behaviour

Writing on the *behaviour of criminals* can be traced back to the eighteenth-century, and a classical criminology that presumed criminals made rational decisions. This criminology was underpinned by the assumption that individuals have free will and will seek to maximise pleasure and minimise pain. Beccaria's (1738-94) classical work on crime and punishment argued that there was a contractual relationship between the individual and the state which existed to prevent chaos. Accordingly, individuals gave up some of their liberties in the interest of the common good, or order. Which liberties had to be given up was the purpose of law, and punishment was meted out to those that broke the law. Importantly though, offenders were seen to have the same capacity for resisting offending behaviour as non-offenders.

The development of a number of important criminal justice systems and constitutions were founded on this classical criminology (Walklate, 1998), but as the Industrial Revolution saw rapid growth of cities and a deterioration in social conditions, so crime began to rise. Against this backdrop, the pleasure principle of crime motivation looked less convincing and people also began to question how children's criminal behaviour could be explained. The result was a shift to a positivist criminology and the commitment to the gathering of facts that caused crime, and particularly facts that supported the idea that individual biology, and not free will, was the cause of much crime.

Lambroso (1853-1909) is often cited as the father of positivist criminology and he in turn was influenced by the work of Charles Darwin. Lambroso used the concept of atavism - the recurrence of certain primitive characteristics that were present in an ancestor but had not occurred in intermediate generations - to explain why some people were effectively 'born criminal'. Such folk, it was claimed, could be identified by sloping foreheads, long arms and unusual ear size, for example, and in time, Lambroso developed a typology of criminals: those born criminal; the insane criminal; the occasional criminal; and the criminal of passion. For each of these types, their (criminal) behaviour was seen as a result of the abnormality which was outside their control, rather than the consequence of free will.

Lambrosian criminology left a profound legacy (Walklate, 1998), although today these ideas seem somewhat naive and are generally out of fashion. They did, however, inspire a great deal of science that focussed on biological and psychological factors as ways of explaining criminal behaviour. Furthermore, since the mapping of the human genome, the search for genetic causes of crime is receiving renewed attention (Wells, 1998).

The Criminality of Behaviour

In contrast to the behaviour of criminals, theories of the *criminality of behaviour* are concerned with factors external to the individual which might result in lawbreaking behaviour. Such factors might include aspects of neighbourhoods, lack of employment and prospects, and the emergence within societies of negative attitudes towards certain types of behaviour. These are important issues within community safety policy and often have distinct geographical traits. Thus, the central ideas of social disorganisation, strain theory, and labelling theory are considered in more detail.

Social Disorganisation

This concept has its origins in work undertaken by members of the Department of Sociology at the University of Chicago in the 1920s and 1930s. The Chicago School of Human Ecology, as it became known, attempted to identify the environmental factors associated with crime and to determine relationships between these factors. Specifically, they were interested in rapid social change at a neighbourhood level, and the term *ecology* was used because the foundations of the research drew heavily on work by plant ecologists.

The theory of human ecology - that there is a parallel between the distribution of plant life in nature and the organisation of human life societies - is attributed to Robert Park. His thesis was that the symbiotic interrelations between people in cities produced some sort of super-organism within which one could identify distinct natural areas (Park, 1952). Importantly, Park's second basic concept, also taken from plant ecology, was that the balance of nature in a given area could change. In human society terms, countries or cities could be invaded by, dominated by, and finally succeed to a new group. For example, dominant ethnicities in a natural area could change, and similarly, business and industry could take over a residential area.

Juvenile Delinquency in Chicago

From a criminological perspective, the most important development of this social disorganisation work was that carried out by Clifford Shaw on juvenile delinquency in Chicago. Shaw argued that the problem of juvenile delinquency was due to the young person's 'detachment from conventional groups' (Short Jr, 1969), which in turn, was exacerbated by social disorganisation. Shaw's neighbourhood studies were based upon Park's earlier work on human ecology and were designed to explore the extent to which illegal activity of delinquents was bound up with the environment in which they lived, and the changes over time therein.

Over the years of his studies, Shaw came to a number of conclusions that can be grouped into concerns relating to physical and economic status and population composition (Vold et al., 1998). Firstly, the neighbourhoods with the highest delinquency tended to be located within or adjacent to areas of industry or commerce; had the greatest number of condemned buildings; and had a decreasing population. In economic terms, the highest rates of delinquency were found in areas with low economic status and in areas with the highest infant mortality and tuberculosis. (Shaw and McKay (1969) concluded that economic conditions in themselves did not cause these problems). Finally, in population terms, areas of highest delinquency were associated with higher concentrations of immigrants and Afro-American heads of families. Shaw and McKay (ibid) also found that despite dramatic changes in ethnic mix over time, areas continued to suffer from high delinquency rates. Moreover, areas into which the older immigrants subsequently moved did not experience a subsequent rise in delinquency rates.

In addition to this research, Shaw published a number of 'life histories' of individual delinquents. Among the findings were that in areas with high juvenile delinquency, "the conventional traditions, neighbourhood institutions, public opinion, through which neighbourhoods usually effect a control over the behaviour of a child, were largely disintegrated" (Shaw, 1931, page 229). Additionally, the child often grew up "in a social world in which [delinquency] was an accepted and appropriate form of conduct" (Shaw, 1938, page 356), and delinquent activities began at an early age as part of street play (Shaw, 1931, 1930, 1938). In play activities, older boys tended to pass on neighbourhood traditions to younger boys. Thus, neighbourhoods may have been characterised by the same types of offences over a long period of time and the normal methods of official social control could not stop this process (ibid).

Shaw concluded that delinquency and other social problems were closely related to the process of invasion, dominance and succession that determined the growth patterns of cities (Morris, 1957). When an area of a city was 'invaded' by new residents, established social ties between residents were broken. In time, the area would be restored to a form of equilibrium, but before then the natural organisation of the area would be severely impaired. These 'interstitial areas', as Shaw called them, became afflicted with a variety of social problems that could be directly attributed to the rapid change in population. With the neighbourhood in transition, the residents no longer identified with it and cared less about its appearance and reputation. Neighbourliness decreased, and neighbourhood people were less able to control young people. The high turnover of children in local schools disrupted learning and discipline, and tensions between invading and retreating cultures were often the cause of gang conflicts between youths.

There have been a number of important criticisms of the work of Shaw and McKay, some of which are particular to the transferability of their ideas to a UK context. Firstly, there are clearly ecological fallacy problems in attributing characteristics to individuals based upon the characteristics of the neighbourhoods in which they live. Yet, to an extent, this problem affects many areal studies and as yet, most attempts to resolve it have failed (Johnston, 2000). It has also been argued that Shaw and McKay failed to take adequate account of external forces such as business and industry (Snodgrass, 1976) and local authority housing allocation policies (Baldwin and Bottoms, 1976) in perpetuating problems in socially disorganised neighbourhoods. Moreover, Shaw and McKay gave very little attention to delinquency found outside of the transitional zone, and particularly those other areas where business and industry were co-located (Snodgrass, 1976). In this vein more generally, there have been criticisms of the Burgess concentric zone model and in particular how it often needs considerable modification when being applied to British cities (Robson, 1969; Bottoms and Wiles, 2002). Most importantly, the work of Shaw and McKay was a major shift from the pathologising of the individual to the pathologising of a whole group. It was also implied that what these disorganised groups required to stabilise crime rates was more 'community'. Yet, this fails to take into account the strong sense of community sometimes found in deviant subcultures or on housing estates where deviance has become a locally accepted and unchallenged social norm (Baldwin and Bottoms, 1976).

Strain Theory

Merton was a sociologist who researched a diverse range of subjects, but his most important work was his social theory of deviance, which he called 'strain theory' (Merton, 1938). Merton theorised that deviant behaviour, including criminal behaviour, was caused by a societal structure that created the same goals for everyone while denying some people the means to achieve those goals. Thus, the poor, who have little access to good jobs, adequate education, and stable family structures, are still expected to strive for wealth, status and power. When they cannot achieve those goals they turn to deviant behaviour.

While Merton's theory was concerned with the individual, Albert Cohen (1955) was critical and argued that there would need to be a collective sub-cultural process where groups of people together tried to make sense of their world by a 'tentative, groping, advancing, backtracking, sounding-out process'. Crime and delinquency, Cohen argued, were one form of sub-cultural adaptation.

Cloward and Ohlin (1960) also worked on Merton's ideas but were concerned that Merton had suggested that strains would simply result in the breaking of the law and had not explained why the strains might result in one type of deviancy rather than another. They argued that delinquency was the result of the available illegitimate opportunities that would vary according to the individual's social and class position. Hence, while a middle-class deviant might turn to fraud or embezzlement, a lower-class deviant would be more likely to rob and steal (Crawford, 1998).

Labelling Theory

The previous collection of work on strain theory combined with another strand of research to come out of the Chicago School in the 1920s and 1930s - symbolic interactionism - to form the basis of a final group (for this review) of concepts on the criminality of behaviour, referred to as labelling theory.

Howard Becker is usually credited with defining labelling theory, although his work used ideas from George Herbert Mead on symbolic interactionism. At a theoretical level, symbolic interactionism concerns itself with the capacity people have to share understandings with one another. Of particular concern are the quality of interactions which take place between people, how those interactions are understood, and how they become refined and developed. For criminology, the concern is how behaviour comes to be understood as deviant and the role of shared norms and values in that process. In Becker's (1963) work, this is broken down into two strands of thought: a concern with how certain types of behaviour become labelled deviant, and a concern with the impact of labelling.

The main argument to the first strand is that society creates deviance by making rules or laws. The breaking of these laws constitutes deviance, and by the application of these laws people who deviate are labelled as outsiders. Hence, "deviance is not a quality of the act the person commits, but rather a consequence of the application by others of rules and sanctions to an offender" (Becker, 1963, page 9). The key point then, is that in understanding deviance, the onus is on the reaction to some behaviour, and not the behaviour itself.

2.2.2 Environmental Criminology

The sub-discipline of *environmental criminology* has been drawn upon by geographers and planners more than other types of criminology because it has traditionally been concerned with explaining the spatial distribution of offences and of offenders (Bottoms and Wiles, 2002). As such, the emphasis is moved away somewhat from the motivations and behaviours of the offender in order to study the crime events themselves (Brantingham and Brantingham, 1993). Some of the theories discussed above have very distinct spatial aspects to them and are claimed as important contributions to the development of this field. In addition to these, however, rational choice theory and routine activities theory have been developed to try and explain spatial patterns of offending and victimisation, and theses such as 'Broken Windows' (Wilson and Kelling, 1982) have also received much academic and political attention. Furthermore, with the emergence in recent decades of desktop geographical information systems (GIS), much research has concerned itself with the empirical testing of these theoretical underpinnings.

Designing Out Crime

Much of the development of environmental criminology can be attributed to interest and enthusiasm, and no small amount of criticism, provoked by work from C. Ray Jeffery (1971) and Oscar Newman

(1972). Newman argued that in residential areas there was a link between architectural design and crime rates that existed over and above other contributing factors. He suggested that the creation of building and estate designs least likely to promote or encourage criminality would act as a form of crime prevention. Central to his design ideas was the notion of 'defensible space', which he defined as a "model for residential environments which inhibits crime by creating the physical expression of a social fabric that defends itself" (Newman, 1972, page 3). Newman argued that architectural design could encourage a sense of territoriality among residents and foster feelings of responsibility for preserving a safe and well-maintained living environment. This was to be achieved by giving residents functional control over small spaces in and around their neighbourhood. These spaces need to be well demarcated so that it is clear which is public and which is private space. In so doing, outsiders are discouraged from entering and residents are encouraged to defend their territory. Housing layout and defensible spaces also need to be juxtaposed to enable surveillance by residents. Newman also argued for buildings to be designed in such as way that their architecture did not suggest vulnerability and avoided being seen as 'special' in any way that might lead them to become stigmatised.

Apart from obvious concerns over architectural determinism, much of the criticism of Newman's work has centered on his failure to take adequate account of social conditions. For example, the point has been made that not all communities have a social fabric active or strong enough to defend spaces created according to Newman's guidelines (Merry, 1981). Concerns have been voiced over the negative implications for pedestrian movement and the loss of informal social control that this can generate (Hillier and Hanson, 1984). Mawby (1977) also had a number of reservations, including Newman's failure to take adequate account of the importance of recidivism when looking at offence rates.

Interest in defensible space was rejuvenated somewhat when Alice Coleman developed Newman's work in the context of British housing estates (Coleman, 1985). Her 'design disadvantagement theory' put forward the idea that the likelihood of a crime being committed could be directly linked with the extent of design disadvantages of the places in the locale. She empirically tested the idea by counting design disadvantages in different locations and then comparing these with indicators of anti-social behaviour. Essentially, she found a positive correlation between poor design and high crime. The work had its dissenters, and Smith (1986) in particular criticised Coleman's quantitative measures of social breakdown. As with Newman, Coleman was also criticised for taking inadequate account of social processes, in particular the impact of housing allocation policies and stigma attached to different estates (Bottoms and Wiles, 1986).

Routine Activities Theory

While CPTED no longer enjoys quite the attention it used to - although it is still found in Home Office and local government policy - routine activities theory is one of the most widely cited theories used by environmental criminologists in the UK. With its origins in the work of Cohen and Felson (1979), it sets out to locate the behaviour of the individual within a wider social context. Crime, it is argued, is the product of three factors coming together at a particular place and time. These factors are a motivated offender, a potential victim, and the absence of a capable guardian. This last factor is more likely to mean the absence of a witness to a crime, be that a neighbour, friend or bystander. The guardian may also be an official figure such as a policewoman, warden or CCTV camera, but one of the strengths of routine activity is its ability to extend beyond the criminal justice system into a wide range of sites and settings.

Routine activity theory does not claim to explain the motivation for crime, nor does it explain why some individual behaviours are more likely to result in victimisation. Neither does it explain why some guardians might be more 'capable' than others. Nevertheless, it does recognise that there is a systematic patterning to crime, in agreement with some of the social ecology theories outlined in Section 2.2.1 (Walklate, 1998).

Since its conception, routine activities theory has also had a number of enhancements. For example, the concept of 'controllers' has been added for each of the three factors in the original theory. 'Handlers', it has been argued, influence the behaviour of the offender in such a way as to render offending less likely and may include persons such as parents or teachers (Felson, 1995). Meanwhile, managers control places where crime may occur. This control may be informal and could include persons such as pub landlords or store owners (Eck, 1995). As Felson argues, "crime opportunity is the least when targets are directly supervised by guardians; offenders, by handlers; and places, by managers" (Felson, 1995, page 55). This concept can be expressed in diagramatic form as the 'crime triangle' (Figure 2.1).

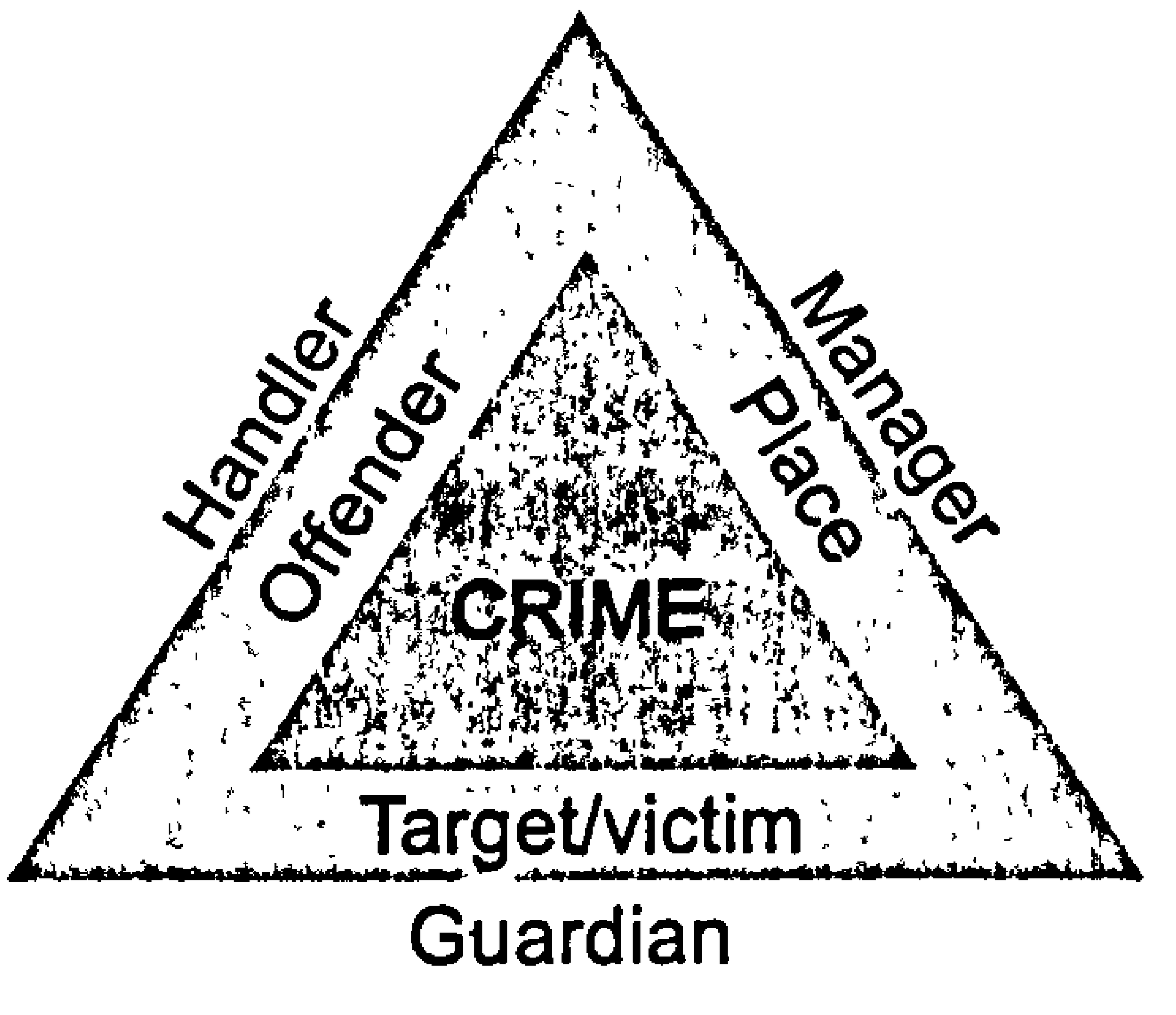


Figure 2.1: The crime triangle. Adapted from Clarke and Eck (2003)

Rational Choice Theory

Where routine activities theory acknowledges the importance of situational context, rational choice theory turns its attention back onto the motivations of the offender. Central to rational choice theory, is that individuals are driven by the motive of maximizing profit, whilst minimizing costs. Decisions to commit crimes are thus based on rational thought about the pay-off, while being constrained by the limits of time and the availability of information (Cornish and Clarke, 1986). The rational process of decision making accounts not only for the decision to offend but also for the time and place where the offence is committed.

One value of rational choice theory, at least for crimes that are not random or spontaneous, is its potential application in crime prevention. As Gibbons argues, "if many offenders, and predatory offenders in particular, weigh at least some of the potential risks against the gains they anticipate from law-breaking, criminal acts may often be deterred by making them riskier or harder to carry out" (Gibbons, 1994). As a result, if the target of the criminal can be made less vulnerable, there will be a greater chance that the criminal will choose another target. However, this only solves part of the range of sites and settings.

Routine activity theory does not claim to explain the motivation for crime, nor does it explain why some individual behaviours are more likely to result in victimisation. Neither does it explain why some guardians might be more 'capable' than others. Nevertheless, it does recognise that there is a systematic patterning to crime, in agreement with some of the social ecology theories outlined in Section 2.2.1 (Walklate, 1998).

Since its conception, routine activities theory has also had a number of enhancements. For example, the concept of 'controllers' has been added for each of the three factors in the original theory. 'Handlers', it has been argued, influence the behaviour of the offender in such a way as to render offending less likely and may include persons such as parents or teachers (Felson, 1995). Meanwhile, managers control places where crime may occur. This control may be informal and could include persons such as pub landlords or store owners (Eck, 1995). As Felson argues, "crime opportunity is the least when targets are directly supervised by guardians; offenders, by handlers; and places, by managers" (Felson, 1995, page 55). This concept can be expressed in diagramatic form as the 'crime triangle' (Figure 2.1).

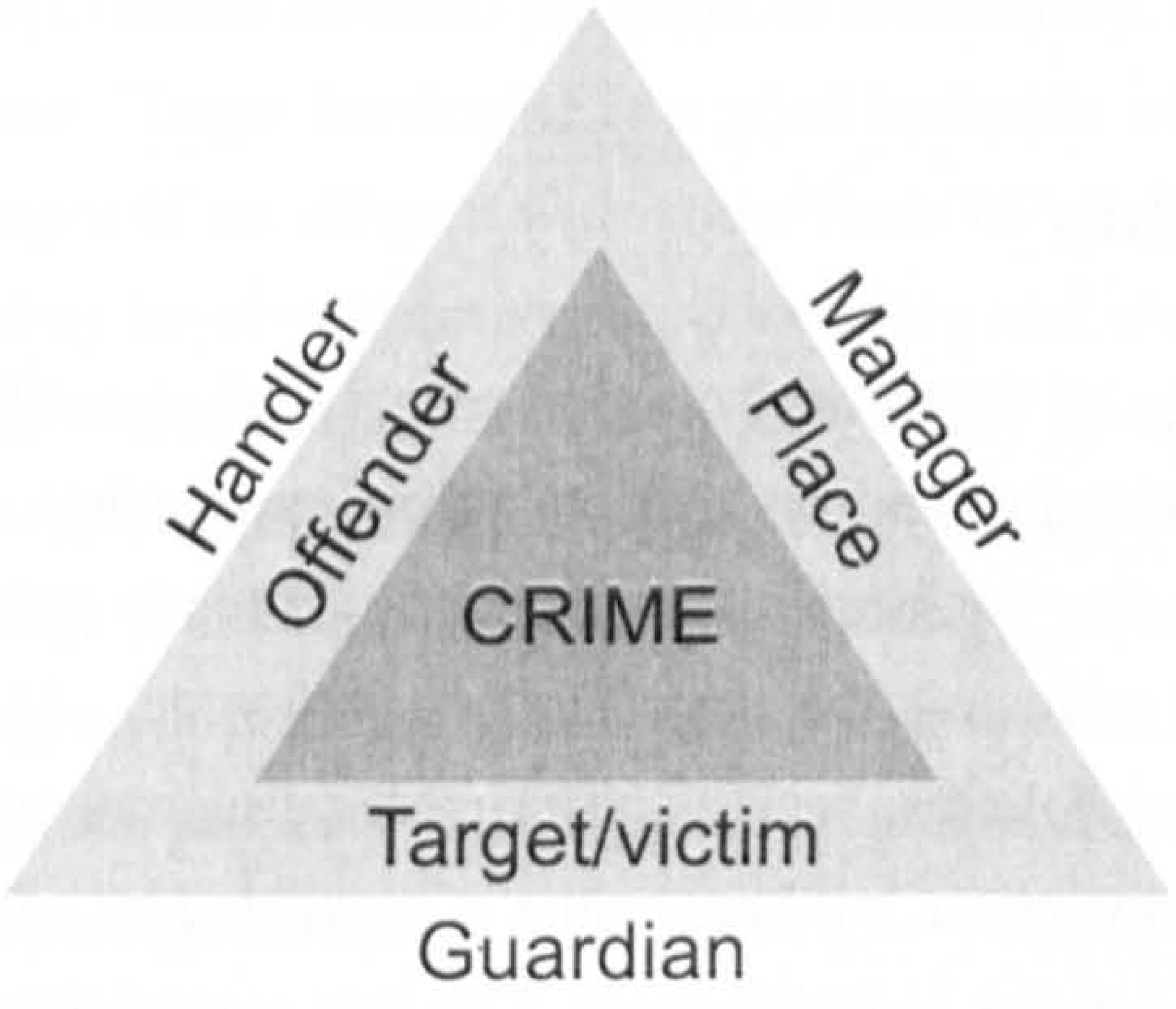


Figure 2.1: The crime triangle. Adapted from Clarke and Eck (2003)

Rational Choice Theory

Where routine activities theory acknowledges the importance of situational context, rational choice theory turns its attention back onto the motivations of the offender. Central to rational choice theory, is that individuals are driven by the motive of maximizing profit, whilst minimizing costs. Decisions to commit crimes are thus based on rational thought about the pay-off, while being constrained by the limits of time and the availability of information (Cornish and Clarke, 1986). The rational process of decision making accounts not only for the decision to offend but also for the time and place where the offence is committed.

One value of rational choice theory, at least for crimes that are not random or spontaneous, is its potential application in crime prevention. As Gibbons argues, "if many offenders, and predatory offenders in particular, weigh at least some of the potential risks against the gains they anticipate from law-breaking, criminal acts may often be deterred by making them riskier or harder to carry out" (Gibbons, 1994). As a result, if the target of the criminal can be made less vulnerable, there will be a greater chance that the criminal will choose another target. However, this only solves part of the *problem: the hardened target is given a reprieve, but the crime might be displaced to some other time and/or place.*

Some have insisted that this relegates rational choice theory to 'administrative criminology' (Young, 1986). Clarke, one of the main proponents of this approach, turned Young's argument around to make the point that much of criminology has been too absorbed with trying to explain peoples' motives to commit crime, and that "criminological theories have been little concerned with the situational determinants of crime" (Clarke, 1980, page 136). Theorising about psychological and sociological events has a role to play, but for practical crime prevention, "given that each event is in turn caused by others, at what point in the infinitely regressive chain should one stop in the search for effective points of intervention?" (Clarke, 1980, 137).

Another weakness of rational choice theory is its inability to account for both offending by young people and crimes which are committed as expressive or compulsive acts. At issue with child offending is whether the child is sufficiently developed enough to be capable of rational thought. Some have also argued that the development of rational thought will vary with age and circumstances (Valentine, 1996). Even for crimes committed by adults which might be deemed rational, it is contested whether the rationality exhibited is any different from the rationality assumed in most criminological theories (Akers, 1990).

2.3 Policy Approaches

Not surprisingly, the diversity of theoretical opinion reacts with prevailing political programmes to produce a range of crime reduction and prevention policy approaches. Over recent decades in the UK, situational crime prevention has been a very important force in policy and research, and it concentrates on controlling and reducing the opportunities for crime. At the same time, it has been recognised that broader social policies, such as city regeneration and education, can also have positive implications for crime prevention, but here crime is prevented by addressing causal factors in the development of criminal tendencies in individuals. The two approaches are not mutually exclusive, and in the UK they have developed alongside one another. Section 2.4 will show how the Crime and Disorder Act 1998 attempts to coordinate their delivery at a local level, but first, the situational and social policy approaches to crime prevention will be discussed further.

2.3.1 Situational Crime Prevention

Situational crime prevention has been a major area of research and policy at the Home Office since the beginning of the 1980s and has formed a large part of the Home Office's output. This work, directed and initially stimulated in large part by Ron Clarke and Patricia Mayhew (Clarke, 1980; Clarke and Mayhew, 1980), is often referred to as administrative criminology and emerged, according to Young (1986), as a result of the failure of academic and political endeavours to show that crime was caused by social conditions and that change to these conditions was needed to combat crime. During the 1980s, crime rose and as a result, the Home Office focussed on crime management rather than trying to understand and deal with all its causes. Indeed, much of the political appeal of the situational approach is that control measures are relatively easy and cheap to implement when compared to the "daunting psychological and socioeconomic circumstances that may have to be altered to reduce people's *motivation* to offend" (Shaftoe, 2004, page 80). This approach has drawn on rational choice theory, routine activities theory and theories of defensible space and design disadvantagement. Unlike the earlier work of the Chicago School, the situational approach looks at the nature of the relationship between crimes and places, rather than between crimes and offenders. Although a subtle difference, this allows consideration to also be given to those places which have a low offender density but high crime, for example, city centres and retail parks.

In general, the situational approach to crime prevention has been summarised (Crawford, 1998, page 66) as representing a shift towards:

- "the prioritisation of the *control* of crime, through practical yet limited policy-oriented measures;
- an emphasis on alterations to the *physical* environment;
- the significance of processes of *informal* social control; and
- the *offence* rather than the offender as the primary focus of attention".

The actual initiatives and interventions delivered under the banner of a situational crime prevention approach are usually designed to either increase the effort of criminality, increase the risk of detection, or reduce the rewards of crime. Target hardening is a good example of the first approach, with the aim being to surround the target of an offence with some form of physical protection. To protect a property from burglary this may involve improvements to doors and locks. Alternatively, to protect pedestrians from speeding motorists a situational approach might involve traffic calming measures such as speed bumps. The second approach, increasing risk of detection, might be accomplished by increasing surveillance capacity. This could be done by installing CCTV or by taking up ideas about defensible space and encouraging an increase in informal social control. Finally, strategies to reduce the rewards of crime could include property marking or the replacement of a coin-operated pay phone with a device only able to take swipe cards.

Although often technical in nature, situational prevention measures must have a social element to explain why the techniques work better on some occasions than others (Lea, 1992). Yet, usually, the social causes of crime are implicitly or explicitly overlooked and this has led some to suggest that situational crime prevention is an assault on the social causes of crime (O'Malley, 1992). Whether this extreme view is taken or not, the amount of attention afforded to situational approaches must have served to marginalise the importance of socio-economic factors to some degree (Crawford, 1998).

On a final note, one interesting and enduring piece of work to come from the administrative criminology of the Home Office has been the British Crime Survey (BCS). Begun in 1982, the BCS has been, and still is, a valuable tool for estimating the true extent of crime because it is victim-oriented and asks members of the public about their own experience of crime, whether ot not these were reported to the police. At the time of its inception, it was also hoped the BCS would dispel some of the myths and fears that had grown up around the scale of the crime problem and patterns of victimisation (Mayhew and Hough, 1988). It has certainly achieved this, and challenged conventional media images of those most likely to be personally victimised (Walklate, 1998). Furthermore, the BCS has also classified survey responses by the geodemographic profile (using CACI's ACORN categories) of the respondent. In so doing, this administrative tool does show that victimisation and perceptions of crime and fear do vary according to area categories. The profiles of some of these area categories are suggestive of the need for different types of crime prevention policy to address specific socio-economic and community groups, implicitly supporting, it could be argued, a role for more social and communal strategies for crime prevention.

2.3.2 Social and Communal Strategies

The explicit development of social crime prevention in England and Wales has lagged behind situational crime prevention (Crawford, 1998). Crime may be impacted upon by urban policy, education policy, housing policy, family policy, or employment policy, but the extent to which crime levels ought to be affected can be difficult to quantify. Evaluating the crime prevention benefits of some of these policies is also complicated by the amount of time it might take for any discernible effect to be detected, and on how multiple policies might combine to affect outcomes (Rhodes et al., 2005). Of course, the social policies may have more immediate benefits in terms of providing more jobs or more affordable housing, for example, and it could be argued it is unfair to criticise social policies with broad or non-crime specific aims if crime does not fall in the areas in which the policies are delivered.

Frequent subjects of social policies that have a crime prevention dimension are 'the failing school' and 'the dysfunctional family'. The school and the family are the two main institutions through which people form bonds with society, and it is argued that it is through social bonds that society encourages individual's to resist their natural weakness for criminal behaviour and conform to the rules (Hirschi, 1969). Furthermore, socialisation is claimed to be important in the fostering of self-control, the presence or absence of which will determine an individuals ability to resist crime (Gottfredson and Hirschi, 1990). As such, these control theories assume that under-socialisation produces crime.

Critics have argued that insufficient attention has been paid to the possible consequences of over-socialisation (Crawford, 1998). For example, Campbell (1996) argued that young men with a penchant for stealing cars were, in many ways, doing their best to conform with dominant male values by driving fast performance cars - the "quintessentially modern commodity with masculinised notions of power, flight and freedom" (Campbell, 1996, page 106). Nevertheless, social policies aimed at developing socialisation skills, ostensibly in young people, are many. Policies aimed at schools and education may attempt to deal with individual issues such as bullying, truancy and exclusions, while government schemes such as Sure Start aim to deliver a broad range of services for children and families, key among which aid social and emotional development.

In addition to control theories and arguments surrounding the importance of socialisation, theories such as strain theory can also be engaged to design social policies aimed at providing offenders or marginalised groups with a stake in conformity. Social policies can be designed to help young people "develop a sense of competence, a sense of usefulness, a sense of belonging, and a sense that they have power to affect their own destinies through conventional means" (Empey, 1977, cited in Crawford, 1998, page 107). It is the 'conventional means' here that are important for crime prevention, and again, 'stake in conformity' policies tend to be targeted towards young people. Unfortunately, a number of key projects adopting this approach, such as the Mobilisation for Youth Programme in the United States, were delivered at the same time as criminology was caught up in the causation crisis of rising crime rates but also improving employment, housing and educational conditions. This led to arguments that social policies aimed at reducing inequality had failed as crime prevention policies (Wilson, 1975). Communal policies for crime prevention differ by targeting whole (usually area-based) communities or neighbourhoods, rather than individuals identified as being at risk of offending. Community organisation approaches, such as the Chicago Area Projects developed by Shaw and McKay, often concern themselves with (re)establishing informal social control, and again, socialisation is seen as crucial, but this time at a communal, and not just individual level.

If the communities do not respond to policies designed to improve their organisation, there is also the option to help them defend themselves. In addition to the ideas about architecture and environmental layout mentioned earlier, the 'broken windows' thesis of Wilson and Kelling (1982) has chimed with communitarians such as Etzioni, whose *Spirit of Community* (1993) manifesto, was in turn highly influential on the British Labour Party leadership prior to its coming to power in 1997 (Crawford, 1996). Central to the 'broken windows' thesis are the importance of minor incivilities and the signals that these send out about the neighbourhood. Behaviour such as vandalism, grafitti, begging, drunkenness and disorder - a broken window left unrepaired - signify that no one cares and that the environment is uncontrolled and uncontrollable. With no signs that minor incivilities will be challenged, troublemakers can move in (the troublemakers are nearly always portrayed as outsiders, not locals) causing fearful residents to avoid confrontation and finally to leave the neighbourhood altogether. This in turn leads to less informal social control, which results in more crime, which leads to more fear, and so on. Importantly, it is the minor incivilities which are seen as the key target for crime prevention. Left unchecked these might 'tip' a neigbourhood into a spiral of decline. Hence, an amount of proactive order maintenance from the police and other local authorities ought to allow residents to reassert their own authority and control to better defend themselves against such a slide. Empirical studies have challenged the simplicity of the link between signs of incivility and crime or the fear of crime (Taylor, 1997). Moreover, it has been argued that the way in which communities perceive crime and other social problems will often be mediated by the political and social resources available to that community (Lewis and Salem, 1986).

2.4 Community Safety

Thus far, this review has considered a number of criminological theories that identify aspects of crime and its social, economic and environmental context that might be measured and included in neighbourhood classification for community safety. Some attention has also been given to the ways in which theories and theses can be used to suggest policy interventions to prevent crime. No mention has yet been made about the main actors in the delivery of crime prevention. Situational crime prevention measures developed in the UK have largely been seen as police business. By contrast, the delivery of crime prevention benefits from broader social policy has typically involved all the other Government departments, besides the Home Office.

This section will set out to describe how the two dominant crime prevention approaches - situational and social/communal - have come together to be directed by local partnerships charged with both delivery of crime prevention and also considering the wider social, economic and political contexts that give rise to some of the causes of crime. This involves activities and issues that have become encapsulated within the term 'community safety'.

2.4.1 Morgan Report

> "We see community safety as having both social and situational aspects, as being concerned with people, communities and organisations including families, victims and at-

risk groups, as well as attending to reduce particular types of crime and the fear of crime. Community safety should be seen as a legitimate concern of all in the local community" (Home Office, 1991).

The Morgan Report (Home Office, 1991), from which the quote above was taken, did much to introduce the concept of community safety in the UK. Commissioned by the Home Office, the Standing Conference on Crime Prevention, chaired by James Morgan, was asked to review developments in the partnership approach to crime prevention. This key idea - partnership working - had been outlined in the Home Office Circular (HOC) 8/1984 (Home Office, 1984) and was reinforced in HOC 44/1990, entitled *Crime Prevention - The Success of the Partnership Approach.*

The Morgan Report identified six elements crucial to multi-agency crime reduction work: structure, leadership, information sharing, identity, durability and resources. It was also recommended that at the middle level (local authority) of a proposed three-tier responsibility structure the following tasks should be tackled:

- formalise a joint commitment to crime prevention;
- formulate overall policy objectives for the strategy based on analysis of local problems;
- identify the responsibilities of different departments and organisations;
- assess the resource implications;
- monitor progress, partly by receiving progress reports about other activities; and
- produce an annual publication reporting on progress and setting targets for the following year.

Despite the fact that a clear and coherent structure of responsibilities had been defined by the Morgan Report, the Conservative Government of the day largely ignored it because of opposition to the idea that statutory responsibility for community safety be given over to local authorities (Crawford, 1998).

2.4.2 Safer Cities

Despite the cool response to the Morgan Report by Government, the partnership approach was an important aspect of the delivery of the Safer Cities Programme. This programme was announced by the Government in March 1988 as part of the Action for Cities initiative. Phase I ran from 1988 to 1993, and Phase II ran from 1993 to 1997. The stated aims were "a) to reduce crime, b) to lessen the fear of crime, and c) to create safer cities where economic enterprise and community life can flourish." (quoted in Tilley, 1992, page 1). Funding from Government came to a total of 22m over the seven years. A partnership approach was required, and these partnerships added to the resources by local fund raising and agency contributions, bringing the total to £40m (Audit Commission, 1999).

Phase II was targeted at 29 project cities, which included over 1,400 separate schemes. The cities targeted were those with high levels of crime and deprivation. Each project city consisted of a multi-agency steering committee responsible for developing a community safety strategy. In addition, three non-governmental organisations acted as managing agents. The role of these - NACRO, Crime Concern, and the Society of Voluntary Associates - was to provide expert advice to the project staff. Leeds was one of the cities to benefit from Phase II funding, and as such, the Leeds Safer Cities Programme was the predecessor of today's Safer Leeds community safety partnership.

Evidence of the success of Safer Cities Phase I has not been published widely, but a detailed evaluation of Phase II was produced for central government (Knox et al., 2000). One of the findings concerned the planning mechanism which required projects to audit crime in their area and develop a strategy implemented through annual action plans. It was felt that the mechanism itself had merit but that in practice it had been poorly executed. The crime audits were "generally fit for purpose", but the strategic planning was often "too unclear and process (rather than outcome) orientated" (ODPM, 2000). The evaluation seemed to accept, however, that the diverse nature of outputs from the many different schemes made it difficult to develop suitable indicators that would give a true sense of a scheme's performance.

2.4.3 Crime and Disorder Act 1998

One of the first major bills to be put forward by the New Labour Government aimed, amongst other things, to establish the partnership approach to community safety in statute. The Crime and Disorder Act 1998 (CDA 1998) (Great Britain, 1998) received Royal Assent on the 31st July 1998. The main purpose of the Act is to tackle crime and disorder and help create safe communities. To this end, the Act has a number of underlying themes (Home Office, 1998b).

- to deal with youth offending, with emphasis on checking this behaviour early;
- for the police and local authorities to work together with the whole community, in partnership, to cut crime; and
- for local authorities and other local organisations to consider the crime and disorder implications of their work.

For the purposes of this research, it is these latter two themes which are of particular interest. Within the Act, these take form in Sections 5, 6, 7, and 17. Section 5 places on local authorities and the police a joint responsibility for the formulation of crime and disorder reduction strategies in each district, borough or unitary local authority area in England and Wales. This continues the partnership working approach implemented in Safer Cities Phase II, and chimes with the Morgan Report recommendations. Section 7 is quite brief, and outlines the authority of the Home Secretary to ask for reports on community safety work done locally. Section 17 is more substantial, in that it states that all local authorities, special authorities, police authorities, national parks authorities and the Broads Authority should consider crime and disorder reduction while exercising all their duties. This is an acknowledgment that crime and disorder cuts across all these jurisdictions, and that to tackle the root causes of crime will require that everyone in local government consider the possibilities of reducing crime in their individual policies. Section 6 is perhaps the most important section for the purposes of the research, as it lays out the requirement for local authorities and the police to draw up and implement strategies for reducing crime and disorder in their area. To summarise, the responsible authorities must ensure that they,

- carry out an audit of levels and patterns of crime and disorder;
- prepare an analysis of the results of the audit;
- formulate a strategy, including objectives;

- set long-term and short-term targets for measuring the outcomes of the objectives; and
- monitor the effectiveness of the strategy, and make changes where necessary.

Sections 17 and 115 of the Act help facilitate the analysis and audit processes by enabling partner agencies to share information which hitherto may have been withheld to allay concerns, real or imagined, about data protection or human rights. Initially, the authorities bound by these Sections were the police and the local (district or unitary) authority. Collectively these organisations are referred to as the responsible authorities. Section 17 of the Police Reform Act 2003 required that police authorities and fire authorities also become responsible authorities, and since 30 April 2004 primary care trusts have also been added.

2.4.4 Safer Leeds Partnership

With the CDA 1998 enacted, every local district and unitary authority was faced with establishing a statutory Crime and Disorder Reduction Partnership (CDRP). Due to its involvement in Safer Cities, Leeds was already versed in establishing partnership working and the new statutory arrangement emerged as the *Leeds Community Safety Partnership*. With the merger of the Leeds Drug Action Team and the CDRP in 2005, the partnership was renamed *Safer Leeds*.

The Safer Leeds Executive members represent the responsible authorities and include:

- Children and Young People's Partnership;
- Leeds City Council;
- Leeds Strategic Partnership;
- National Offender Management Service/Probation;
- Primary Care Trusts;
- Safer Leeds Chair of the Board;
- Safer Leeds Partnership Manager; and
- West Yorkshire Police.

The membership of the partnership is wider still and includes community and voluntary organisations such as Community Action and Support Against Crime (CASAC), Leeds Tenants Federation, Leeds Victim Support and Leeds Voice. Members from the criminal justice system include the Crown Prosecution Service, Leeds Magistrate Court and the Youth Offending Service. Membership is also extended to the University of Leeds, Leeds Metropolitan University and Leeds Chamber of Commerce.

In accordance with its responsibilities under the CDA 1998, 2004/05 saw the partnership undertake its third triennial audit and strategy setting exercise . In many respects the current priority themes - acquisitive crime, anti-social behaviour, drugs, reassurance and violent crime (Safer Leeds, 2005) - look very similar to those in the two previous strategies (LCSP, 1999, 2002). In addition to these themes, the community safety strategies also identify cross cutting issues on which attention will be focussed. These *have* changed over time and may either be reflecting the extent to which problems have been solved and issues overcome, or be acknowledging changes in the political climate. For instance, the strategy for 2002-2005 identified a set of cross cutting issues almost exclusively concerned with community safety management - reassurance, communication, performance management, evaluation and partnership monitoring. Emphasis was clearly on the collection and dissemination of data and the setting up of organisational and technical frameworks to monitor performance and establish which policies and community safety initiatives were working and which were failing. Particular objectives pertinent to this research (which has run for the same period as the strategy) included ensuring targets were specific and measurable, improving gaps in information and data, comparing performance with similar CDRPs and establishing PhD study work on evaluation with the University of Leeds School of Geography (this thesis!).

As such, the 2002-2005 Strategy (LCSP, 2002) had all the hallmarks of 'new managerialism', or 'new public management' (Flynn, 1997), and displays the capacity to turn political and moral decisions into administrative and technical ones (Matthews and Pitts, 2001). There is also evidence of the encouragement of new forms of expertise in the manipulation of figures and possibly an implicit prioritising of initiatives whose outcomes are easily quantifiable (Loveday, 1999). It is not clear which of the cross-cutting issues emerged out of a genuine and original local need and which were dictated by the demands for performance information made by regional and central government. The latter might be expected in some cases, given arguments about the anxieties at the centre caused as a result of the devolvement of community safety policy to local authorities (Cooper and Lousada, 2000).

These issues are not at all unique to Leeds either. Instead, they reflect a wider preoccupation with performance assessment and audit (Power, 1997) that now permeates many public services. In politics, there were accusations by the Secretary for State for International Development, Clare Short, of 'control-freakery' on the part of New Labour, creating problems of excessive bureaucracy and centralised targets (HC Deb, 12 May 2003, vol.405 c.37). In a less impassioned statement, this time to the House of Commons Public Administration Committee investigation of the use of targets and performance measurement in public services, James Strachan, the Chairman of the Audit Commission observed that targets can become "a very distracting add-on and irritant ... [and] ... real obstacles for change". He went on to say that "a slavish devotion to the universal meeting of targets, many of which have not been set very intelligently, is a sure-fire way of not improving public services" (HC 62-vi, 2002-03).

2.5 Information Strategies, Analysis and Evaluation

The renewed thirst for data, information, performance measurements and indicators is not new and the technical pitfalls for the unwary were spelled out quite clearly back in 1960s in debated about the (then) 'new Philistinism' of economic accounting (Gross, 1965). Back then, the call went out for a wider set of quantitative variables to be used for measuring social structure and social performance, although it was also acknowledged that "quantitative information of any kind - no matter how picayune - tends to retard the circulation of qualitative information" (Gross, 1965, page 15).

Neither is the use of a broad set of social measurements to assess and direct policy necessarily immune from problems, especially when the policy is for such a wide-ranging concept as community safety. Amitai Etzioni outlined a number of 'dangers' in valid social measurement, foremost of which is the problem of fractional measurement. This, it was put, "concerns dysfunctions stemming from lack of coincidence between a social concept and its operational definition" (Etzioni and Lehman, 1967, page 1). Under this problem come issues such as over-reliance on single variables or single score indexes and the problematic tendency for people to measure the means rather than the goals of social policy. Finding examples of this last problem in community safety is very easy. Within Leeds, for example, the success of anti-social behaviour (ASB) strategy has been measured according to the number of ASB cases handled, the number of ASB Orders issued, and the number of Acceptable Behaviour Contracts entered into (the means). More of all of these is deemed a measure of success, and not the reduction in ASB incidents (the goal) reported. In defense of Safer Leeds, however, there may be times when such measurements are the lesser of a number of alternative evils. As Etzioni goes on to point out, the taking of goals into account "can itself have dysfunctional consequences if it leads to omission of other systemic considerations (Etzioni and Lehman, 1967, page 7). In the case of the multi-agency 'blitzs' on neighbourhoods with the worst crime and ASB problems, the use of a subsequent reduction in ASB reporting would be a poor choice of measurement. This is because one consequence of these actions has been an increase in ASB reporting, due, it is thought, to increased public confidence that the authorities are taking an interest (personal communication with Bev Yearwood, Killingbeck Divisional Community Safety Partnership Co-ordinator, 6th October 2004).

2.5.1 Data Sharing

To facilitate community safety analysis, there has been a genuine attempt, both nationally and locally, to collate data from a wider range of partners about a wider range of issues than used to be the case. This has been mindful, perhaps, of the problems of relying solely on a few crime rate indicators and recognises the diversity of community safety work that goes on besides policing.

Data Protection Action 1998

The CDA 1998 has acted as one enabling mechanism for this new community safety data sharing and has provided a legal obligation on CDRP partners to share information for the purposes of investigating and preventing crime. This is important, because the Data Protection Act 1998 (DPA 1998) is not explicit about how personal information may be used other than to say the use must be lawful. Within the context of this thesis, it is likely that most data will not include personal data, either because it simply is not recorded or because it will have been filtered out by data providers mindful of their data protection responsibilities. There are, however, special exemptions within Section 33 of the DPA 1998 that pertain to the use of personal data for research. Essentially, Section 33 allows personal data to be held indefinitely, which is not the usual case. Furthermore, providing it is for research, the data may be used for a different purpose than that for which it was originally intended. Lastly, and most conveniently, the data controller (researcher) is exempt from Section 7 of the DPA 1998 and thus is not obliged to notify the data subject(s) about the use of their information unless publication of research findings would enable an individual to be identified. Along with the rights afforded by Section 33 come responsibilities. In short, the data controller must ensure that data are not processed to support measures or decisions with respect to particular individuals and that data are not processed in such a way that substantial damage or distress is, or is likely to be, caused to any data subject.

Technological Gatekeepers

The greater use of increasingly complex and sophisticated computerised information systems has also made data exchange much easier. Yet, practical problems do arise and there are situations when individuals are reluctant to provide information, for a variety of reasons. Some theoretical discussions about information control, particularly as a power resource, may not specifically relate to computer data and may indeed predate the microprocessor. Nevertheless, a brief review of some of the issues surrounding data, control and power may be useful.

One such issue is the role of 'technological gatekeepers', as defined by Allen and Cohen during their research on the flows of technical information in a variety of research and development laboratories (Allen and Cohen, 1969). These gatekeepers were well connected with information and ideas circulating both within their own organisation and outside via professional friends and research literature. Within the context of the wider organisation or a project team - and it seems reasonable to draw parallels with a community safety partnership - the role of these people, it was argued, was critical. It has been pointed out that Allen and Cohen rather imply that technical gatekeepers would use their position for the general good. Yet, there also exists an alternative possibility that a gatekeeper might seek to use their position to improve their own power status or thwart the aggrandisements of others (Pettigrew, 1972).

Whether technological gatekeepers can be identified so easily in todays typically distributed computer (network) environment is not so clear. People with specific data responsibilities still exist, indeed are flourishing in community safety partnerships, yet within the large organisation they are often peripheral and dealing with with specific local issues rather than global ones. The decision making can appear to be decentralised, yet when disparate information starts to be collated and synthesised by the centre, whether for research or monitoring by management, it can create again a sense of centralisation of power and control (Robey, 1981).

In this vein, it might appear that technological gatekeepers - those people on the periphery with the day to day responsibility for collecting and storing data - could become marginalised as management information and decision support systems become more common. Furthermore, the central collation and presentation of management information may play down peripheral expertise based on non-quantitative data sources which do not fit easily within a performance framework (Bloomfield and Coombs, 1992). It follows that centralised data may begin to paint a quite different picture from that which it was originally intended to produce and outsiders, such as academic researchers, may need to take care whose data definition they work with.

A children's needs study in Sheffield grouped their problems in collecting datasets as: communication difficulties; diverging organisational priority; resource constraints; data protection issues; technical difficulties; and data quality (Signoretta and Craglia, 2002). Some of the difficulties were due to poor briefing of technical gatekeepers by their managers and different interpretations of the DPA 1998. Some partners in the project were also only able to provide partial or incomplete datasets which could then not be accommodated within the research design.

Within the crime reduction and public protection field, similar difficulties have been experienced. Synthesising a number of Home Office and police reports, Bellamy et al. (2005) highlight the under-developed state of information and communications and the failure of nationally strategic computer projects such as those for the National Probation Service and magistrates courts. The unreliability of the Phoenix offender database on the Police National Computer has also drawn criticism - especially following the trial of Ian Huntley in 2003 for the Soham murders. And yet again, the point was made that there is widespread uncertainty about the law relating to data sharing.

In both of these examples, the problems can in part be seen to have arisen because of the tensions between policy stances on data sharing and privacy. On the one hand the Government is promoting the kinds of risk assessment, data matching and social sorting techniques that are often associated with surveillance while, on the hand, the Government is committed to developing data protection mechanisms to protect privacy and human rights (6 et al., 2005).

2.5.2 The Partnership Business Model

In practice, data sharing agreements have been established in many partnerships across England and Wales and these are used for a wide variety of data analysis purposes. For those CDRPs still requiring guidance, a Partnership Business Model (PBM) was conceived by the Home Office as another way of supporting partnerships in the delivery of crime and disorder reduction agendas. The PBM is an approach to problem solving and partnership working designed to enable better decision making, and deliberately borrows from the National Intelligence Model (NIM) developed for police forces in England and Wales (NCIS, 2000). The NIM received governmental approval in the *National Policing Plan, 2003-2006* (Home Office, 2002), which in turn has become enforceable under the Police Reform Act 2002. The NIM is a model for policing, and thus concerns operational matters that would not normally be of concern to non-police agencies (whether statutory CDRP partners or otherwise). Nevertheless, it is fairly straightforward to see how aspects of the NIM could be applied to the types of community safety work carried out within local authorities. In particular, the NIM establishes a place for the spatial analysis of crime patterns, demographic and social trends, and results of tactics and crime prevention measures. It is unlikely that non-police CDRP partnership agencies would adopt the fortnightly Tactical Assessments described within the NIM for use at BCU level, but some local authority based community safety teams are starting to adopt the twice-yearly Strategic Assessments approach.

As plans for the PBM have developed, the implementation emphasis has come to rest on the building upon of existing systems rather than replace existing systems with a 'one size fits all' solution. These systems would ensure (Beaney, 2003):

- an agreed data quality standard (analytical tools package);
- minimum standards for problem solving;
- a secure and standardised process for using and sharing information (data sharing and storage facilities);
- a way for CDRPs to assess the latest information on crime reduction (good practice database); and
- a systematic method for analysing and reducing certain types of crime and disorder (crime reduction schemas and problem solving methods).

It was anticipated that if the PBM could be successfully piloted then it would be promoted by the Home Office as an example of best practice for all CDRPs to follow. Middlesborough became the principal pilot site, although Leeds also received money from the Home Office to develop some IT infrastructure aspects of the PBM. At some point during 2004, the promotion of the PBM by the Home Office changed tack, and the PBM project received a new title - Partnership Support. There is scant information in the public domain about this change of direction, but it appears that the new onus is on providing separate tools and guidance that can be used singly, rather than promoting a sophisticated framework for all eventualities.

Some of the concepts of the PBM have been encompassed in a Leeds project of the same name that has been developing IT infrastructure to support crime pattern analysis and initiative evaluation. This, and its surrogate parent application - Leeds Statistics - are discussed in the next chapter. The Partnership and Neighbourhood Data Access (PANDA) project at Wakefield and Greater Manchester's GMAC project (GMAC, 2005) are similar examples of data access portals and associated processes designed to facilitate local PBMs.

2.5.3 Triennial Audit and Strategy Setting

If CDRPs do implement the principles of the PBM, as the Home Office have urged (Home Office, 2004b, para 5.18), then one of the largest of periodic assessments for a CDRP - the statutory triennial crime audit and subsequent strategy review - ought to be made easier. Embedded within the Crime and Disorder Act 1998, the audit should (adapted from Fox and McManus, 2001):

- review the levels and patterns of crime and disorder in the area taking due account of the knowledge and experience of persons in that area;
- analyse the results of that review; and
- be published locally.

Elsewhere, the audit has been described as being broken down into four aspects: data collection, consultation, policy review and expert input (Mills and Pearson, 2000). The warning, however, is that "there is an especial danger of getting hung up over the first two of these, and paying insufficient attention to the latter two" (Mills and Pearson, 2000, page192). Whichever way it is described, the triennial crime audit is a large exercise in data collection that often struggles to get beyond data summary and description.

Yet, there is ample guidance on what audits can and should include in the way of analysis. Leaving aside surveys, the analysis of crime data should look for patterns in space and time and use descriptive or inferential statistics to form a picture of the nature and scale of crime in particular areas. Profiles should also be generated to describe the typical characteristics of offenders and victims and measures employed to prevent or detect crime during the period being audited ought to be evaluated (Fox and McManus, 2001). Three separate guidance documents have been produced by the Home Office (Hough and Tilley, 1998; Home Office, 1998a, 2004a) specifically to provide guidance on conducting crime audits and developing strategies. In addition, generic Home Office documents (Read and Oldfield, 1995) and publications from the Jill Dando Institute of Crime Science (Clarke and Eck, 2003) have sought to raise awareness among the practitioner community of crime analysis techniques and how to employ them. Taken together, the guidance documents are supportive of the use of GIS as a means of harmonising datasets whose boundaries are not coterminous. Emphasis is also placed on the map as a means of making area statistics more readily accessible and engaging. It is recommended that the GIS is used to help provide social, economic and demographic context to crime patterns and these patterns themselves are identified using hotspot mapping techniques.

Only Read (1995) offers much in the way of critique on analysis techniques, pointing out that in the past, police forces have tended to be seduced by the appeal of GIS. Part of the seduction may be due to GIS developers and users heeding Brunsdon's advice that "it is essential that the computer systems ... 'hides' the internal statistical analyses, and presents results in terms of output more meaningful to police managers or crime pattern analysts", who, "although experts in the analysis of crime patterns, may not have the training to directly apply these statistical methods, or interpret their output" (Brunsdon, 1989). Although as Read suggests, "while the latter is probably true, police officers are likely to remain wary of a predictive system whose logic is 'hidden' from them unless, and until, the accuracy of the algorithms that generate these patterns is proved operationally" (Read and Oldfield, 1995, page 32).

On classification and standardisation, Hough (1998) briefly mentions standardising crime and disorder rates by geographical area in order to better appreciate the extent to which an area's rate is different from that across the wider geographic area. Furthermore, it is suggested, but not explicitly drawn from the last point, that there is utility in identifying areas with locally high rates for further analysis. This idea of beginning with a 'broad and shallow' analysis followed by lines of enquiry that are more 'narrow and deep' is central to the guidance presented in (Home Office, 2004a), although here the use of socio-demographic data overlayed onto crime patterns is seen as something to be undertaken at the 'narrow and deep' stage and not as a preliminary means of identifying atypical crime patterns.

2.5.4 Police and Justice Bill 2006

Finally, the Government published a Police and Justice Bill early in 2006, which was in part shaped by the results of a review of the Crime and Disorder Act 1998 (CDA). In a written communication (3rd November 2005) to the Minister of State for Crime, Security and Communities, Hazel Blears, the Head of Community Safety and Local Government Unit, David Truscott, outlines a number of proposals for the Safer Communities Bill which would develop the PBM and change the nature of the crime audit and strategy setting process. The proposals can be summarised thus:

- **Introduction of a partnership framework for delivery based on NIM principles, that would be compulsory.** This would involve undertaking regular strategic intelligence assessments informed by community consultation and engagement. Priorities emerging from these assessments would include action plans at neighbourhood level.

- **Lift the requirement in legislation for CDRPs to undertake three year audits and strategies of crime, anti-social behaviour and substance misuse.** This is in recognition of the considerable demand that the triennial audit can place on CDRP analytical resource. The suggestion is that a partnership framework based on NIM principles (i.e. a PBM) would enable routine analysis of real time data, negating the need for large stand alone audits every three years.

- **The production of annual rolling three year plans, based on CDRP strategic assessments and community consultation.**

- **The strengthening of Section 155 of the CDA to place a duty on all CDRP responsible authorities to share depersonalised aggregate data.** this would include the extension of

Section 155 to Primary Care Trusts and Fire and Rescue Authorities, who although responsible authorities, are not explicitly covered by Section 115 at present.

- **The provision of information to local people on progress on delivering their community safety priorities.** Although not stated, it is presumed that because strategic assessment would identify priorities at a neighbourhood level then the information of progress towards these priorities would be provided at the same scale.

All of these proposals are evident in some form in the resulting Police and Justice Act that was given Royal assent in the autumn of 2006. At the time of writing (December 2006), guidance on how to implement these changes is being produced by the Home Office.

2.6 Concluding Remarks

This chapter has shown that there is a wide body of literature within both geography and criminology that can be drawn upon to inform this research. In a sense, a neighbourhood classification approach can be viewed as a continuation of the attention paid by geographers to multivariate analysis back in the 1950s and 1960s. That work was sparked by the production of a new set of small area census data, rather as this research is seeking to exploit new data published in the 2001 Census. Given the developments within environmental criminology and the geography of crime, the research is probably more in tune with the technically oriented methodological developments of the former discipline. Although, the more critical and reflective research produced by geographers over the last couple of decades also provides important insights into crime and its social context.

Despite the number of arguments against reliance on officially recorded crime data, the research is being sponsored by agencies for whom crime control is the principal preoccupation. Thus, for the purposes of this research, crime will be defined as law-breaking behaviour. Where necessary, assumptions will have to be stated about the consequences of this decision for the interpretation of experimental results. When interpreting theoretical ideas, the research will also have to be mindful that studies produced overseas may not always transfer easily to the UK context and its particular social and political history of urban (re)development. Some theories can also appear to be of little significance given the dominance of the situational approach to crime prevention in the UK. Chainey and Ratcliffe, for example, appear to take the Felson and Clarke (1998) view that criminological theory has long seemed irrelevant to those dealing with offenders 'in the real world'. They suggest (of such theories) that, while interesting, "they are often difficult to convert into a practical crime reduction strategy from a policing or practitioner perspective. After all, it is difficult to tell just by looking at people walking past on the street how divorced they are from their means to achieve their life goals and who therefore might be a potential offender "(Chainey and Ratcliffe, 2005, page 80). Such a view, it is argued, displays both a failure to appreciate both the theory and the role of social policies being delivered, albeit often only in part, with crime prevention and community safety as a stated aim.

The review has also reported that although technical in nature, situational prevention measures must have a social element to explain why interventions work better on some occasions than others (Lea, 1992). This would seem an interesting hypothesis to test using a classification which can discriminate between neighbourhoods of different types. A number of the criminological theories discussed also suggest possibilities for variables to use in the classification. For example, control theories point to the school and family as possible sources of proxy variables for levels of socialisation, possibilities that are discussed further in Chapter 4.

There are also warnings of the danger of fractional measurement, both from the academic literature (Etzioni and Lehman, 1967), and Government attempts to evaluate complex policy initiatives (ODPM, 2000). Community safety partnership working certainly involves many different agencies and types of crime prevention work, and the data that the Crime and Disorder Act 1998 will enable partners to provide for this research will present a range of technical and interpretive problems. That said, as partners themselves, the Leeds universities do have a responsibility to help find solutions for these technical problems.

Needless to say, not all academic points of view are in agreement on how best to react to the demands of issues such as target setting and performance assessments. The neighbourhood classification approach could be criticised for perpetuating problems inherent with 'new managerialism' (Flynn, 1997), but the next chapter will show that there is still scope for improving spatial analysis of crime and community safety initiatives. Furthermore, as central government funding for community safety becomes more competitive, it will be more important than ever to be able to produce analysis that can highlight genuine need and demonstrate fairly the effectiveness of interventions in different neighbourhood settings.

Usefully, the importance of spatial analysis is acknowledged both implicitly and explicitly in the community safety policy guidance material. Traditionally, the triennial crime audit has provided the impetus for steps forward in data collection and analysis methodology, while the Partnership Business Model looks set to define more tightly the role for crime pattern analysis in general, and also spatial analysis in particular, as community safety policy for England and Wales develops.

Chapter 3

Crime Mapping and Area Profiling

3.1 Introduction

Given the various statutory demands and the need for intelligence and analysis to target community safety action identified in Chapter 2, academic geographers and criminologists have responded to the problems of spatial crime pattern analysis in a number of ways. Firstly, there has been, and continues to be, development of methodologies and techniques with an emphasis on technical sophistication and novelty. In addition, there are other studies which are more applied in nature whose main aims tend towards the testing of a hypothesis on the spatial distribution of crime. Lastly, there are studies, such as those which aim to evaluate social policy and crime prevention measures using a range of techniques, which develop an understanding of the crime patterns but make no special attempt to judge the efficacy of the particular technique(s) adopted. This is not a strict typology, and some research studies include more than one, or all, of these aspects.

In needs to be acknowledged, that there is also research into the spatial analysis of crime being undertaken outside academia, in police forces and in local and national government departments. Most of this work is applied, rather than developmental, and the extent of activity can be difficult to judge as much of this work goes unpublished. Nevertheless, examples of research from this domain will be reviewed within this chapter, serving to emphasise the point that in time, moves to professionalise crime analysis, as promoted in the 2004 White Paper *Building Communities, Beating Crime* (Home Office, 2004b), may begin to blur some of the current distinctions between academic and practitioner-led research.

The chapter begins with a brief review of numerical and hotspot detection techniques that have been applied to crime data. Some of these techniques have been developed specifically with crime analysis applications in mind, while others have been borrowed from elsewhere. The discussion will then move on to consider the implications of the call made by Gloria Laycock, Director of the Jill Dando Institute of Crime Science, for those engaged in crime mapping research and practice to move 'beyond blobology'. In turn, this provides a useful way to introduce and review research into the provision of area context for crime analysis and mapping. Finally, the chapter revisits the aims and objectives of this research and makes the claim for the research and development of a new neighbourhood classification for community safety.

3.2 Spatial Analysis of Crime

It has been clear for a very long time that there are geographies of crime, as offenders and offences often show a tendency to concentrate in particular sections of geographical space (Evans et al., 1992). The features of these spaces are well documented (see Herbert, 1982), although according to the mechanisms of routine activities theory, for example, some potential victims and targets will never experience a crime. Vulnerable targets still require a motivated offender and, in some instances, the lack of a capable guardian, before they become victimised.

This, it is argued, outlines a paradox for those engaged in analysis for community safety - the intervention will nearly always be too late. Risky targets can be identified *a priori*, but there are usually insufficient resources available to protect them all. Thus, resources are aimed at the most needy targets, and this relies on identifying such locations by analysing where the most number of crimes have occurred historically - retrospective mapping (Groff and La Vigne, 2002). For this reason, predictive crime mapping is becoming an increasingly important research topic for environmental criminologists and crime geographers (for example Groff and La Vigne, 2002; Corcoran et al., 2003; Bowers et al., 2004; Johnson and Bowers, 2004), but for the purposes of this thesis, the interest is still with identifying spatial patterns in existing crime records. Throughout, the term 'hotspot' will be used to describe those locations in space where crimes cluster together. Spaces where crimes are conspicuous by their absence may be referred to as 'coldspots'. In the identification of both hotspots and coldspots, the temporal dimension may also be taken into consideration, but in practice this is less common.

3.2.1 Spatial Statistics

The simpler numeric and statistical techniques that are employed to analyse crime patterns are those that measure global, first-order properties of the spatial distribution of objects. Typically, these techniques measure spatial distribution or spatial autocorrelation. Hence, analyses are able to demonstrate the central location of groups of crimes, or suggest whether crimes are likely to be spatially clustered. Second-order properties, by contrast, refer to sub-regional or neighbourhood patterns within the whole distribution. For the most part, all these techniques take individual crime incidents and their point locations as input, rather than aggregations of crimes within zones.

Measures of Spatial Distribution

Examples of centrographic statistics include the mean centre, median centre, standard distance and standard deviational ellipse. Details of the mathematical derivations can be found elsewhere (e.g. Hammond and McCullagh, 1974; Shaw and Wheeler, 1994; Robinson, 1998; Levine, 2002). The *mean centre* is simply the point in space represented by the average x coordinate and average y coordinate of all the points in the study. Its use has been explored for the profiling of serial offenders (Paulsen, 2004), but ultimately it suffers because it is easily influenced by outliers and need not actually represent a crime site, or indeed a location within the study area at all. The *median centre*, identified using the median x and median y coordinates, also suffers from these problems, but is less sensitive to outliers. The *standard distance* measures the degree of dispersion of the distribution, rather in the same way as standard deviation measures dispersal of values about the mean. It can be combined with the *mean centre* to provide a non-directional indication of the distribution of the data about the *mean centre*, and has been used to show levels of dispersion between different crime types

(Chainey et al., 2002).

The directional problem inherent with the *standard distance/mean centre* combination can be solved by employing the more computationally complex *standard deviational ellipse* (SDE). The ellipse is centered on the mean centre, with its long axis in the direction of maximum dispersion and its short axis tangential to this (Ebdon, 1985). The definition of the ellipse only really makes sense when added to a map.

Measures of Spatial Autocorrelation

Measures of spatial description that employ the concept of spatial autocorrelation begin to address the problem of identifying clustering of crime incidents, albeit still at the global level. Spatial autocorrelation is the presence of a spatial pattern in a mapped variable due to geographical proximity. Moran's I statistic (Moran, 1950) is one of the oldest indicators of spatial autocorrelation and is found in various analyses of crime (e.g. Mencken and Barnett, 1999; Zhu et al., 2004). Geary's C statistic (Geary, 1954) is similar to Moran's I, but instead of comparing deviations from the mean it compares the difference in the intensities themselves. For brevity, just Moran's I will be described.

Moran's I was originally applied to zones which have continuous variables associated with them, such as the rate of a particular type of crime. For any continuous variable, a mean can be calculated and the deviation of any one observation from that mean can also be calculated. The Moran's I statistic then compares the value of the variable at any one location with the value at all other contiguous locations (Ebdon, 1985). The technique employs a contiguity matrix that effectively weights zones in a binary manner. Adjacent zones receive a weight of 1 and non-adjacent zones receive a weight of 0. Following work by Cliff and Ord (1973), some implementations of Moran's I use variable weighting according to distance, opening up the opportunity to compare point patterns in addition to zonal patterns. It is also possible to test the significance of the Moran I coefficient in a number of ways (see Levine, 2002).

Anselin's Local Moran

In contrast to the global statistic produced by the standard Moran I test, Anselin's Local Moran (LMoran) statistic (Anselin, 1995) applies the Moran I to individual zones, allowing them to be identified as similar or different to patterns in their local neighbourhoods. Interpretation of the Local Moran test requires a conceptual shift in the way hotspots are considered, as typically, techniques for hotspot detection concentrate on identifying zones or areas in or about which there is a high density of crime. The Local Moran statistic, by contrast, is not a measure of density, but a measure of (dis)similarity. High positive standardised scores indicate clustering of *similar values* (either high or low), whilst *high negative* standardised scores indicate clustering of *dissimilar values*. The higher the standardised score, the more the observation is similar (positive) or dissimilar (negative) to its neighbours. In other words, the Local Moran statistic is a good indicator of both hotspots and coldspots, highlighting those zones that stand out as being different from their surrounding neighbourhoods. To differentiate between a hotspot and a coldspot, the user must look at the underlying intensity values of the zone and its neighbours.

Other local indicators of spatial association (LISA statistics) include Local Geary's C and the Gi and Gi* statistics, which along with the Local Moran technique have been implemented in the GeoDa software (see Anselin et al., 2006) developed by Luc Anselin, itself gaining increasing recognition from authorities responsible for promoting crime mapping (e.g. Cameron and Leitner, 2005). Citing their own work (Ratcliffe and McCullagh, 1999; Chainey et al., 2002; Eck et al., 2005), Chainey and Ratcliffe (2005) argue that the Gi and Gi* statistics (Getis and Ord, 1996) have received the most attention from crime mappers, as the hotspots are defined in a conventional way (areas where local averages are significantly different from global average). Yet, there are also a growing number of studies that utilise the Local Moran statistic (e.g. Mencken and Barnett, 1999; Messner et al., 1999; Chakravorty and Pelfrey, 2000; Murray et al., 2001; Cahill, 2005).

3.2.2 'Hotspot' Mapping

The Local Moran statistic can be visualised either in scatter-plot form or displayed on a more conventional map. This is important because it serves as a reminder that visualisation is an essential component of LISA techniques. Although computational methods can search large volumes of data for a specific type of pattern very quickly with mechanical accuracy and consistency, they can have very limited ability in adapting to various data sets and interpreting complex patterns. In contrast, the human eye can pick out complex patterns very quickly, attaching meaning to patterns and generating hypotheses for further analysis (Peuquet, 2002), although variations in perceptions will be influenced by map design (Sadahiro, 1997). The cognitive and intellectual process of decoding symbols, making sense of features and relationships portrayed on a map has been discussed in depth by MacEarchen (1995) and Muehrcke and Muehrcke (1997).

Some of the different types of mapping technique employed to map crime are discussed below. A number of these techniques, plus many others, have been implemented in a piece of crime mapping software called CrimeStat (Levine, 2002). Developed by Ned Levine for the National Institute of Justice (NIJ), USA, CrimeStat has been used for a number of the examples and illustrations below. All other maps have been produced using ESRI's ArcGIS and SpatialAnalyst software.

Choropleth Mapping

The most common technique for identifying crime hotspots is to produce a choropleth map. When the geography is such that the population at risk is the same for every zone then counts alone may be mapped. This is rarely the case however, so it usual to standardise the counts by dividing them by the population of the zone to which they refer. The geography used to define zones may be administrative (such as wards, census output areas or police beats), or be defined by a regularised geography (such as quadrats). Whichever geography is used, an important objective is to select an interval scale that will enable the reader of the map to identify easily those zones with the same or similar numbers of offences or rates.

Figure 3.1 shows the rate of burglary dwelling across a group of Leeds neighbourhoods. Those neighbourhoods with high rates which cluster in the south-west of the map are coincident with the main area of student housing, whilst to the north-west of the map are more affluent suburbs with lower rates. One problem with choropleth mapping is that zone size plays an important role in how the map is read, with large zones tending to draw the most attention. For geographies where the zone size is designed to capture equal population irrespective of how the large the zone has to be (in area terms), there is a danger that attention becomes drawn to areas of least interest. In addition, the geography might not make any logical sense in crime or any other terms, for example when zones capture areas at both extremes of the crime risk spectrum, such as a ward which contains both rural land and deprived urban areas. The abruptness of changes in shading at boundaries between zones can also be

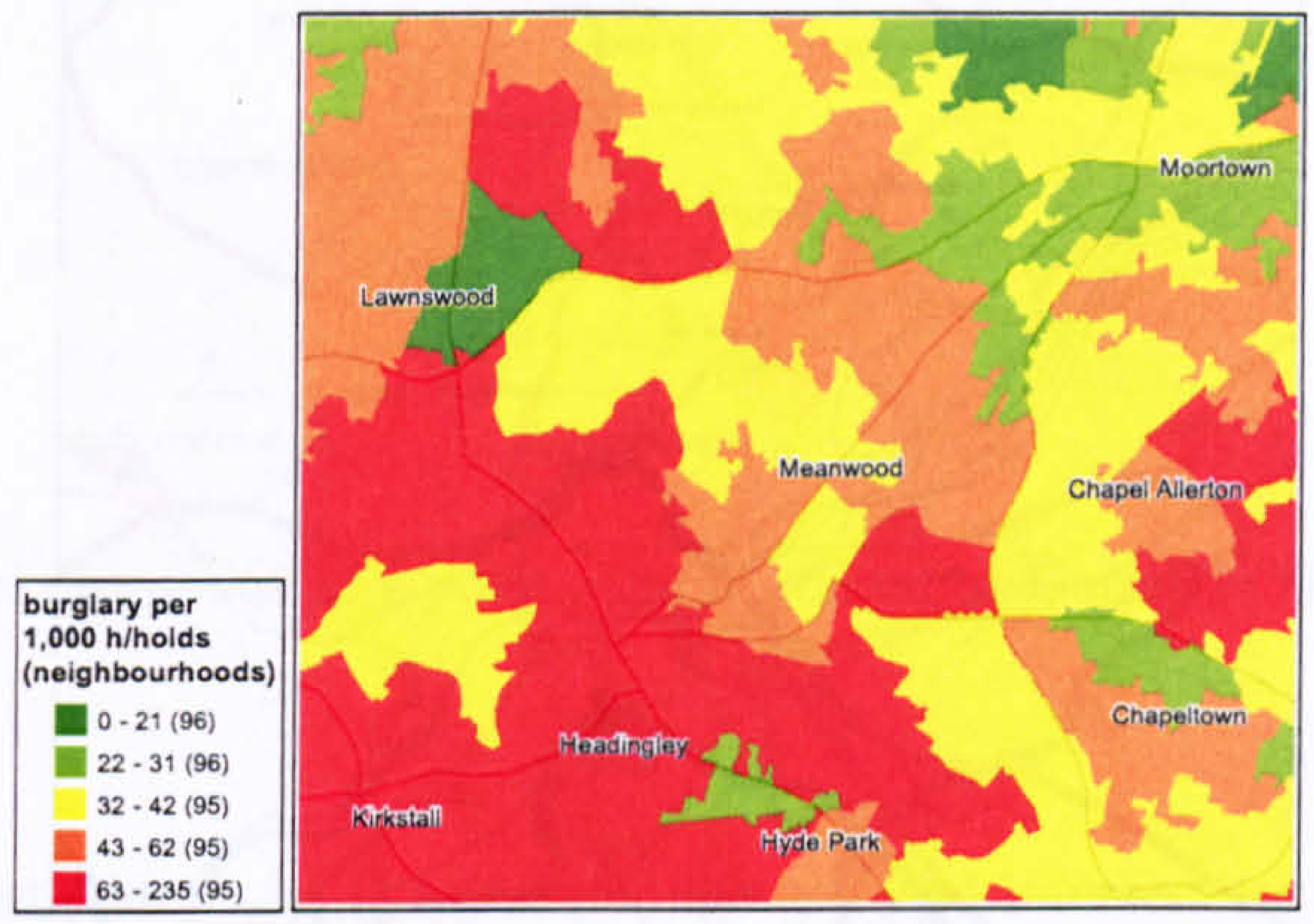


Figure 3.1: Choropleth map of burglary dwelling, by neighbourhood (2003/04).

misleading, but more importantly, the reliability of a choropleth map depends on the homogeneity of its zones. For larger administrative geographies, zones tend to be more heterogeneous and often the level of heterogeneity will vary across the map region. Despite these problems, choropleth mapping with census or other administrative boundaries for which population at risk data are available does have the advantage of being able to map crime rates, thus reflecting, to an extent, the level of crime risk in an area.

Spatial and Temporal Analysis of Crime (STAC)

Developed in 1989, STAC is a crime hotspot analysis tool developed by the Illinois Criminal Justice Information Authority (ICJIA) and has been used extensively, with uptake being highest in the USA. In the UK, STAC has been used in a number of studies (e.g. Hirschfield et al., 2001; Bowers and Hirschfield, 1999; Craglia et al., 2000). For the purposes of this chapter, the STAC approach as implemented in *CrimeStat* will be described and demonstrated.

STAC employs a scanning technique to hotspot identification, and begins the process of identifying clusters amongst point data by overlaying the study area with a 20 x 20 lattice (either triangular or regular). In turn, a search circle is centered at each intersect node on the lattice. The radius of the search circle is specified by the user, and then *CrimeStat* STAC multiplies this by 1.414 to create circles that overlap (presumably only if the search radius is less than or equal to the distance between lattice nodes). *CrimeStat* STAC then counts the number of incidents falling within each circle.

The x, y coordinates of the nodes that have at least two incidents in the search radius are recorded, along with the number of incidents found for each node. These results are ranked, and the 25 nodes with the highest number of incidents are stored. A check is then made to see if any of the points stored fall within more than one search circle. If they do, then incident points for the search circles are combined to form a superset, called a *Hot Cluster*, more than one of which may be found.

Using the coordinates of each incident point, the program then calculates a best-fitting standard deviational ellipse (SDE) for each *Hot Cluster*. This may not contain every *Hot Cluster* point, and could, theoretically, include points which are not in the *Hot Cluster*. The SDEs can then be drawn on a map to represent the 'hotspots' for the particular crime type being investigated.

Figure 3.2 shows the standard deviational ellipses calculated by *CrimeStat* using the STAC routine for *Other Theft* in Leeds. There is no relationship between the size of the ellipse and the amount

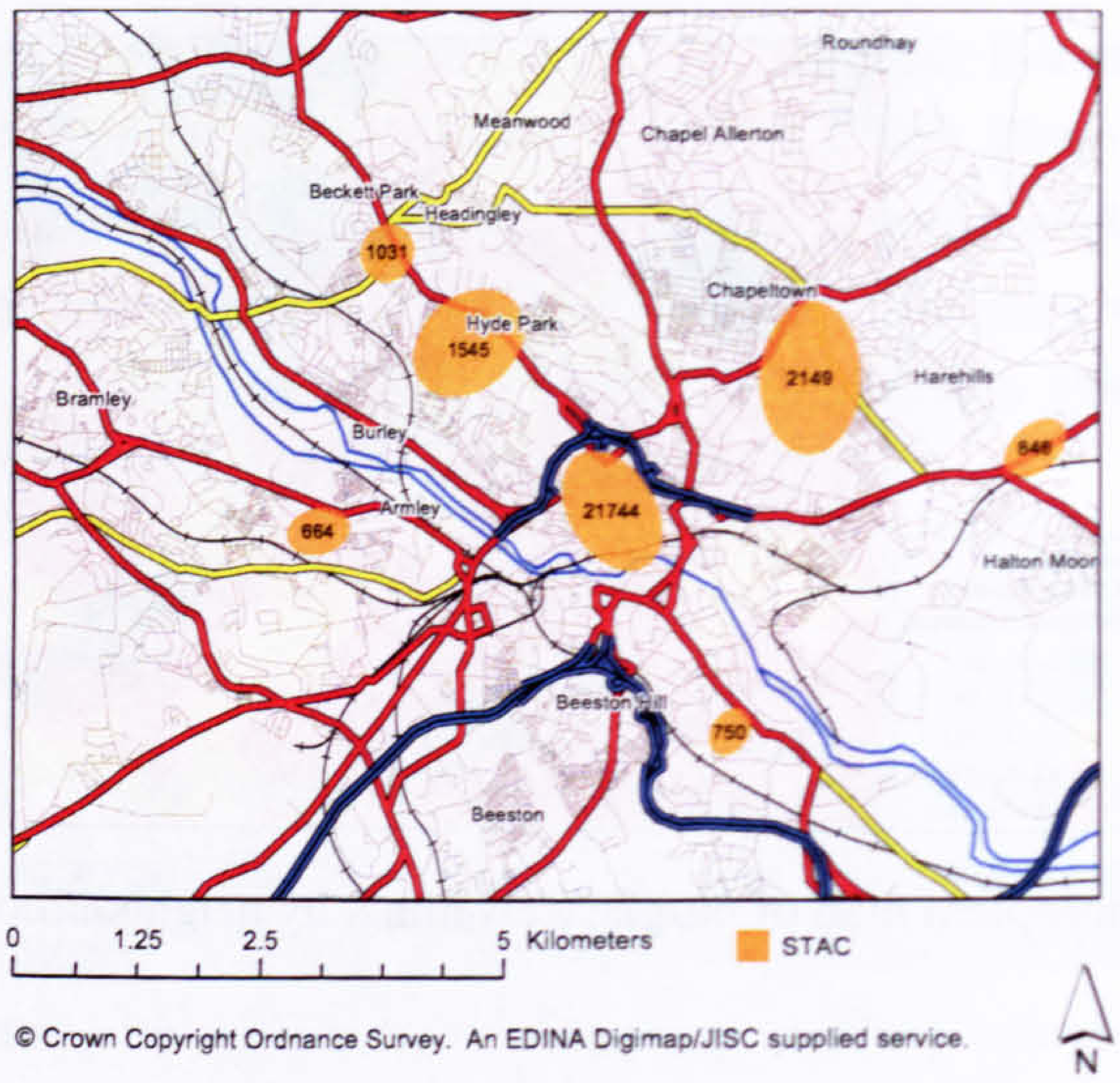


Figure 3.2: STAC standard deviational ellipses of Other Theft (1999-2002).

of crime contained within it. For this reason, a count of the crimes contained by the ellipse has been added in the form of a label, although ellipses could also be shaded in the manner of zones in a choropleth map. *Other theft* includes offences such as shoplifting, stealing things from people without the use of aggression or threat (which would be a robbery), theft from an employer and theft from an ATM. STAC shows that the city centre is the largest hotspot (in volume terms) for these types of offence. Other clusters appear in Harehills, Hyde Park and Headingley. Smaller clusters have been detected around the Killingbeck Retail Park and the Penny Hill (retail) Centre.

One advantage of STAC has over the kernel density estimation techniques discussed next is that the map reader is presented with far less information, thus enabling important features to be identified quickly and easily. It could also be argued that the use of labels to represent intensity values also means that areas with much lower intensities (but which have still been identified as hotspots) are less likely to be overlooked. Among the limitations of STAC is the fact that the distribution of offences within the ellipse may not be uniform (equivalent to the zone homogeneity problem with choropleth mapping). In addition, STAC cannot accurately represent clusters with irregular shapes. Neither is it possible to reflect the underlying risk of crime by taking into account population or household density.

Kernel Density Estimation

Kernel estimation (or kernel density) techniques represent more recent advances in point pattern analysis. In *simple* estimation techniques, such as STAC, the intensity of crimes or other point events are calculated by counting the number of points that fall within a moving window of fixed size centered on a number of locations arranged on a fine grid superimposed over the study region. As the window is moved, it overlaps parts of areas covered by previous windows. In this way it proves a smoother density surface than one built from counting points in discrete quadrats. *Kernel* estimation develops the simple estimation idea by replacing the window with a three-dimensional kernel function which "weights events within its sphere of influence according to their distance from the point at which the intensity is being estimated" (Gatrell et al., 1996, page 259).

The concept is displayed graphically in Figure 3.3.The study region R is overlayed with a fine

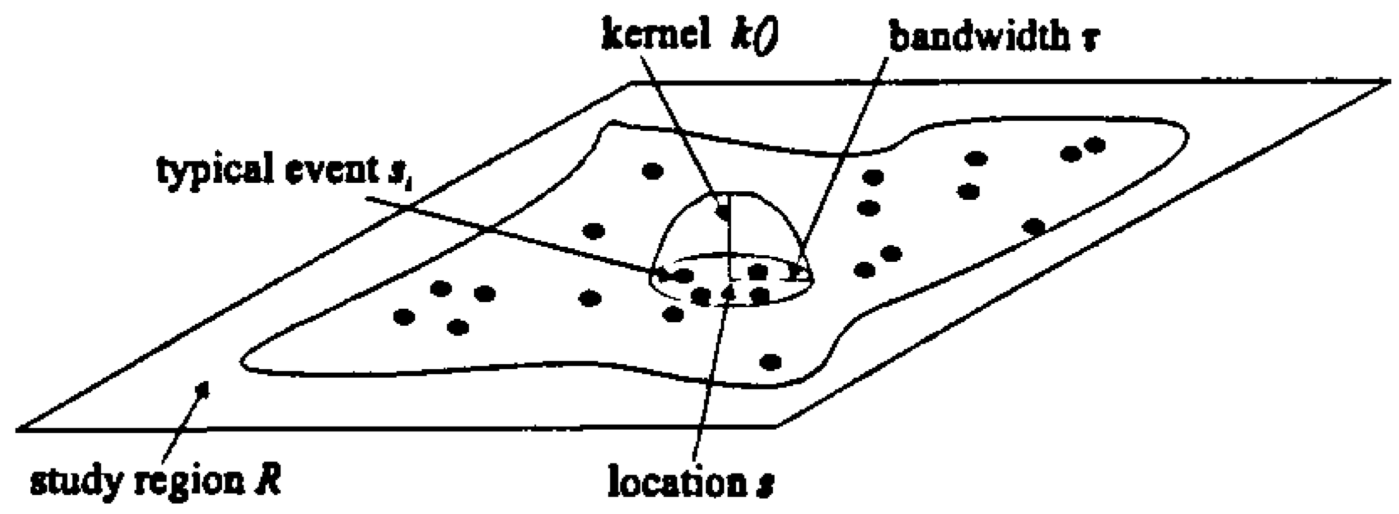


Figure 3.3: Kernel estimation of a point pattern. (based on Gatrell et al., 1996, page 260)

grid, and an intensity estimate $\hat{\lambda}_\tau(s)$ is calculated at each grid intersect (or centroid) point using a kernel function $K()$. The region of influence for each intensity calculation is defined by a circle of radius τ (also known as the *bandwidth*) centered on s. The distance from the centre of the circle to each point s_i is measured and acts as a weighting for the intensity estimate at s. The grid of intensity values can then be mapped as a surface using the contouring algorithms that come built into most modern GIS. The kernel function K may take a number of different forms (for examples see Burt and Barber, 1996; Levine, 2002; Gatrell et al., 1996; Bailey and Gatrell, 1995) but the most common is the normal distribution function.

In *CrimeStat*, the user can choose which kernel function to use. In both *CrimeStat* and *ESRI Spatial Analyst*, the user must also specify the bandwidth, τ. One option is to use a fixed bandwidth over the entire study area. Choosing a large value of τ will produce a greater degree of smoothing, while a small value of τ will produce a more 'peaky' surface. This 'peaky' surface may be better at detecting very localised clusters but if these are very numerous then the utility of the technique to focus attention on areas in an exploratory sense is somewhat diminished. The smooth surface, by contrast, may be better at giving a general impression of the extent to which crimes cluster, but taken to extremes the technique will have no discriminating ability at all. Moreover, as a result of the smoothing process, areas where no crimes have occurred may find themselves being highlighted otherwise. For example, Figure 3.6 shows a density surface for burglary dwelling spreading over areas such as motorways, canals and railway lines - locations where such offences ought not to occur.

Within the existing literature, there are different views about how best to determine an appropriate bandwidth, with a number of different methods being proposed (e.g. Bailey and Gatrell, 1995; Williamson et al., 1998; Brimicombe, 2004). Disappointingly, in some respects, ESRI's *Spatial Analyst* continues to base its default bandwidth on the size of the study area and does not implement any of the numerical techniques proposed elsewhere. It is argued that this may deprive crime analysis practitioners who rely solely on ESRI's tools the chance to experiment easily with alternatives and engage more with academic research. Alternatively, it could also be argued that it forces users to appreciate Bailey and Gatrell's view that "the value of kernel density estimation is that one can experiment with different values [of the bandwidth], exploring the surface ... using different degrees of smoothing in order to look at variation in [the surface] at different scales"(Bailey and Gatrell, 1995, pages 86-87).

CrimeStat also relies on user experimentation, or *a priori* analysis, to determine a suitable bandwidth, although it does offer the opportunity to employ an adaptive bandwidth technique. This acknowledges that points (in this case crime points) will often tend to be sparse in areas where human activity is similarly thin on the ground (rural areas for example). In response, adaptive bandwidth algorithms increase bandwidth in areas where point density is low, and decrease bandwidth in areas where density is high (in densely populated areas, for example).

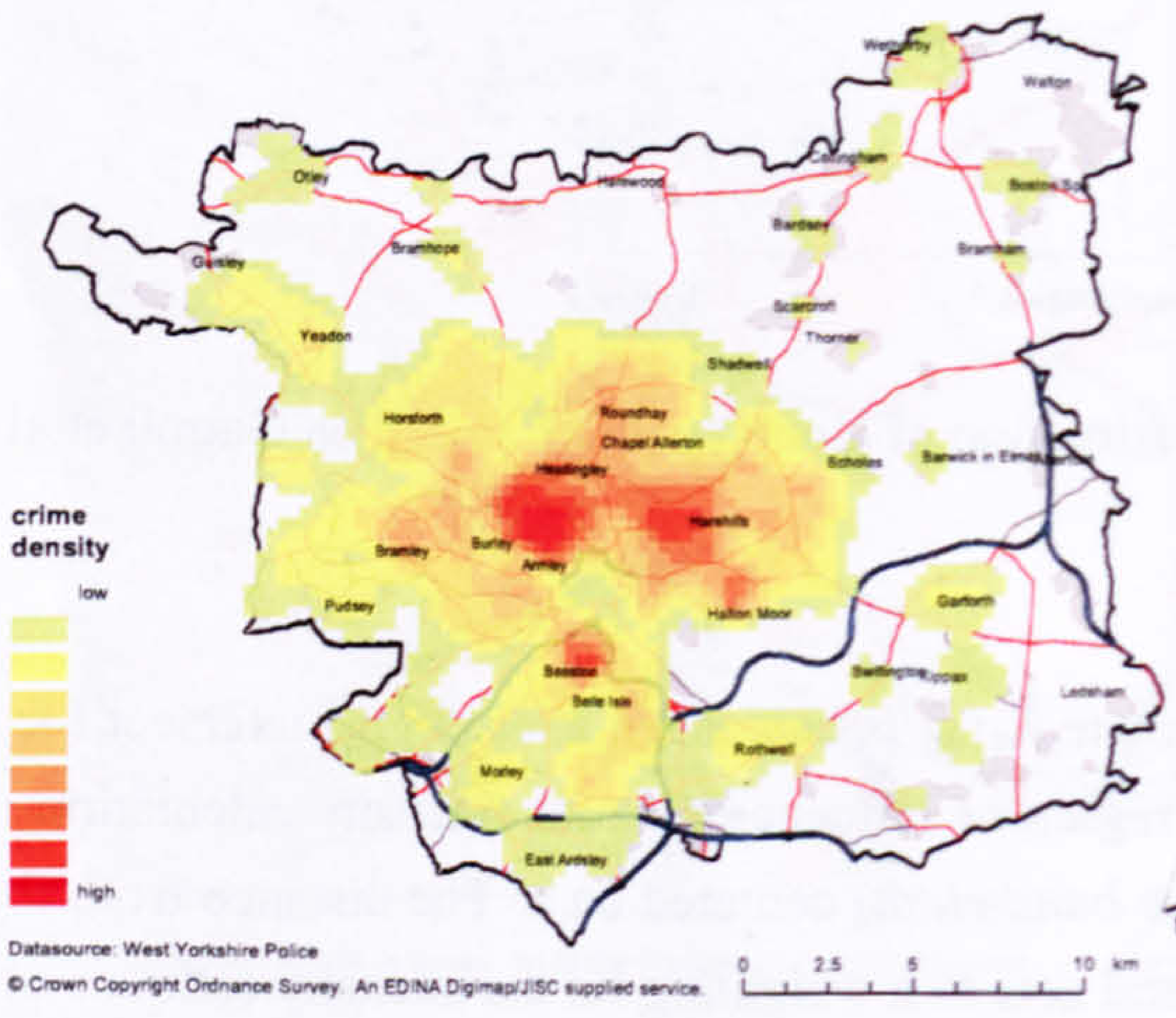


Figure 3.4: Single kernel density estimation of domestic burglary (1999-2002), τ=400m (fixed), normal kernel function.

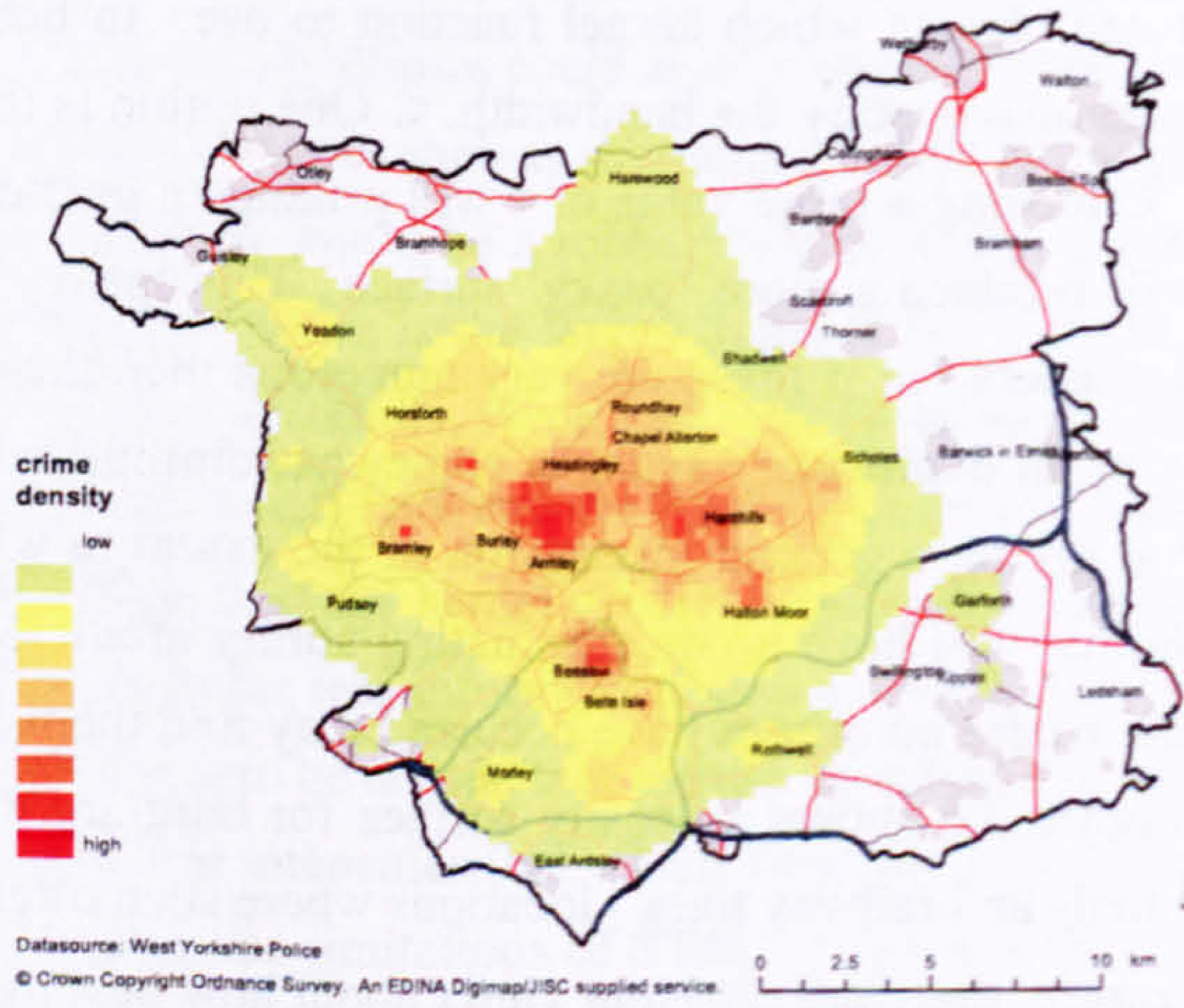


Figure 3.5: Single kernel density estimation of domestic burglary (1999-2002), adaptive bandwidth (min. offences, 100), normal kernel function.

A number of adaptive bandwidth routines have been proposed (e.g. Bailey and Gatrell, 1995; Levine, 2002) but the effect of the adaptive bandwidth estimation technique implemented in CrimeStat can be seen by comparing Figures 3.4 and ??. The results of the fixed bandwidth kernel estimation have, by and large, pinned down domestic burglary to within settlement boundaries, albeit with only moderate definition between some small areas. As well as the main city core, small towns and villages are picked up as having some burglary. The adaptive bandwidth kernel estimation, by contrast, gives better definition inside the city core, but a rather more confusing picture of burglary in rural areas, with large shaded areas covering what is ostensibly agricultural land. This second effect could probably be ameliorated by varying the minimum number of offences parameter on which the adaptive algorithm depends, but the example perhaps serves as a reminder that although a technique might be mathematically more complex, interpreting the output does not necessarily become any easier and the demands on the practitioner to choose appropriate parameters increase.

One last technical issue is how to incorporate the population at risk. So far, although the kernel density hotspot maps have been visually arresting, it has not been clear whether the hottest spots represent high or low risk, given that the geography of housing density, for example, is not even across the district. One solution is to employ dual-kernel estimation. This method has been shown to be a powerful exploratory tool for epidemiology applications, using one density estimate to represent disease cases, and another to represent a healthy control (Bithell, 1990; Kelsall and Diggle, 1995). Transferring the concept to crime analysis, one kernel density estimate can be calculated for the crime point data, and another kernel density estimate can be calculated for the underlying 'at risk' population.

Different combinations of the two density surfaces will produce different types of result, and can be chosen to reflect the needs of the analysis at hand. For example, Kelsall and Diggle (1995) use a log ratio of densities to help mute the effect of very high density estimates. If a heavily spatially skewed distribution is not such a problem, then a simple ratio of densities could be used instead. Other possibilities exist (Levine, 2002), but in Figure 3.6, a simple ratio of densities divides the kernel estimate for domestic burglaries by the kernel estimate for households (as recorded in the 1991 census) to produce an estimate of domestic burglary risk.

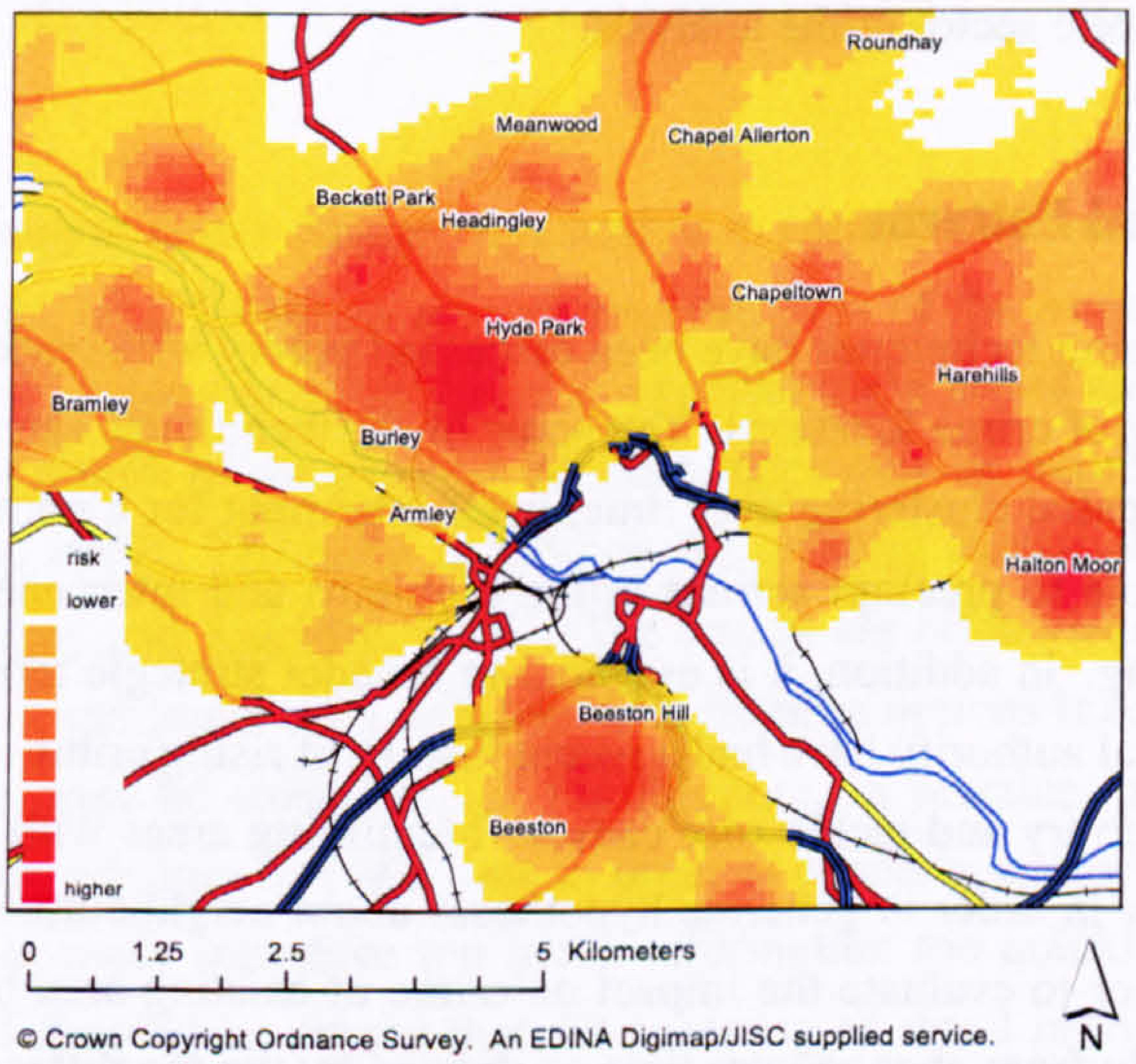


Figure 3.6: Dual kernel density estimation of domestic burglary (1999-2002), τ=500m, normal kernel function.

Although some advantages and disadvantages of kernel density mapping have been mentioned, a number of issues still need to be highlighted. Of the advantages, arguably the most important, is the ability of the resulting maps to render hotspots according to the true underlying spatial distribution of the data, and not to be constrained by a regular geometry such as the ellipse, or forced to conform to an administrative geography. It could also be argued, as hinted earlier, that some of the uncertainty of the technique - in that there is little agreement what is an optimum set of parameters (bandwidth, kernel function, etc.) - can also help foster an exploratory approach and encourage users to engage more with their data and tools. Some have sought to remove this uncertainty and organise an appropriate methodology, for example, to determine class intervals for shading the density surfaces (Chainey et al., 2002). Admittedly, the selection of class intervals can have a great bearing on how the map is read, but again it could be argued that a technique sensitive to the aims of exploratory data analysis, such as the use of boxplots to guide class interval choices (Brimicombe, 2004), might be more appropriate.

The main disadvantage of kernel density mapping for crime analysis (parameter selection aside) is that it is still difficult to adequately reflect the population at risk, despite the possibilities of dual-surface techniques. In part this is due to the extra number of parameter choices (for the population surface) that have be made, but more fundamentally has do with the difficulty in accurately locating and scaling the population at risk in the first instance. Dwelling-related offences do not pose too many problems in this regard, but when the victim or target is mobile, such as a person or an automobile, the problems are considerable and solutions are likely to either have to make fairly bold assumptions or else use considerable resources to collect and/or model specific at-risk populations. This approach was demonstrated by Chainey and Desyllus (Chainey and Desyllus, 2004) in their analysis of street crime risk. This set of complaints are not particular to kernel density mapping; they can equally be leveled at all crime mapping techniques that seek to express victimisation risk. Nevertheless, the kernel density map is arguably more susceptible to criticism than other techniques simply because the output can 'appear' significant and meaningful, not naive or vague. In this sense, the technique can become a victim of its own success, especially when relied upon in the increasingly routinized domain of police and public sector crime analysis.

3.2.3 Spatiotemporal Patterns

Thus far, all of the mapping techniques have been discussed purely with respect to their ability to visualise spatial clustering of crime locations. Yet, some of the techniques can also be used or adapted to visualise change in spatial clustering over time, itself important for a number of reasons. For example, it can be desirable to preempt serious crime problems and intervene early to prevent small problems from escalating. In addition, it is essential in broader strategic terms to understand which larger areas within a local authority have had a recent history of rising crime in order that longer term plans can be drawn up to try and tackle root causes. Identifying areas with falling crime over time is also of interest, either in order to generate hypotheses about neighbourhood dynamics that might be related to the trend, or to evaluate the impact on crime of existing area-based community safety interventions. Separately from change over time as defined by the the difference in crime rates over some period, it can also be useful to identify relationships between temporal and spatial hotspots. Such analysis might, for example, show how crime hotspots shift their location during different times of the day or week, or it might be used on data pertaining to interventions activity to analyse whether spatio-temporal clusters of community safety action are more effective than work carried out at random over space and time.

Change and Bi-variate Choropleth Maps

In situations where choropleth mapping is considered the most appropriate technique, two simple options exist for illustrating change over time. In both cases, change requires the consideration of a crime count or crime rate for two periods, for example two recording years. A simple calculation of percentage change (or percentage share change) can then be mapped and shaded, with zones experiencing the greatest percentage change, or the greatest percentage change in their share of crime overall - usually shaded darkest. Alternatively, the map can be designed to illustrate two variables, with the relationship between the counts/rates for two periods either displayed using shaded zones to map one variable, for example percentage change, and then overlaying this with proportional symbols displaying another variable, for example the most recent crime rate; or by using proportionately sized pie charts overlayed spatially onto some form of background map. Examples of these techniques are provided in Chainey and Ratcliffe (2005, page 234).

Density Surface Differencing

Another differencing technique can be employed with surfaces produced by kernel density estimation. Implemented within CrimeStat's dual-kernel density routines, the calculation of the relative difference in densities subtracts density values standardised by output cell area from one surface by those from a second surface (Levine, 2002). The result is a third surface which shows the extent to which density values have changed between time periods. As with all the kernel estimation mapping techniques, an advantage of using this approach is that resulting patterns are not restricted to fixed geometric shapes or administrative geographies. This can be especially helpful when comparing the results with bespoke geographical study areas or when overlaying onto geographies where population sparsity (city centres, for example) renders zone sizes too large for effective choropleth mapping (e.g. Crawford et al., 2004a).

Animation

An alternative way to analyse crime change over time is to consider where and to what extent crime levels vary, and/or cluster, on different months of the year, days of the week, or hours of the day. A chronological sequence of maps can be produced to examine these issues and if the number of maps becomes large the visualisation of the changes between maps may be improved by using animation (Dorling and Openshaw, 1992). The basic techniques for creating animation of crime maps are described elsewhere (Chainey and Ratcliffe, 2005). The results are often visually appealing, although often the time needed to experiment with different time framing options (necessary to create smooth transitions across space) may be considerable. Furthermore, in practice, paradoxically, it is often necessary to move between frames one at a time in order to properly appreciate the subtleties of the changes. Animation can easily introduce too much information too quickly for the map reader to interpret accurately, and it could be argued that the scanning of small individual frames, following Tufte's (1990) concept of 'small multiples', might be as effective while being less demanding of computing resources.

Comaps

Another extension of the 'small multiples' idea is the *coplot* (Cleveland, 1993), modified by Brunsdon (2001) to create the *comap*. Coplots show changes in the relationship between two variables conditioned by a third, and sometimes also a fourth variable. Within the current discussion of change in crime location at different times, the two variables being plotted would be the eastings and northings of the crime locations, while conditioning variables could be time of day and day of the week. Figure 3.7 shows such a comap for robbery offences in central Leeds, produced using the R software package (R Development Core Team, 2006) and a modified version of Chris Brunsdon's *comap* code (personal communication, Chris Brunsdon, 30.12.2005). The shingles (grey bars) identify the ranges for the given variables and would usually be overlapped in order to keep the number of observations in each panel approximately equal. This can be important for reducing misleading map artifacts caused by unusually high or low numbers of observations (Brunsdon, 2001). In this example, the given days represent the rows and the given hours the columns. Hence, the bottom-left panel (96 incidents) shows robberies on Sundays between midnight and 14:00 hours, while the panel to its right (62 incidents) represents robberies on Sundays between midday and approximately 20:00 hours.

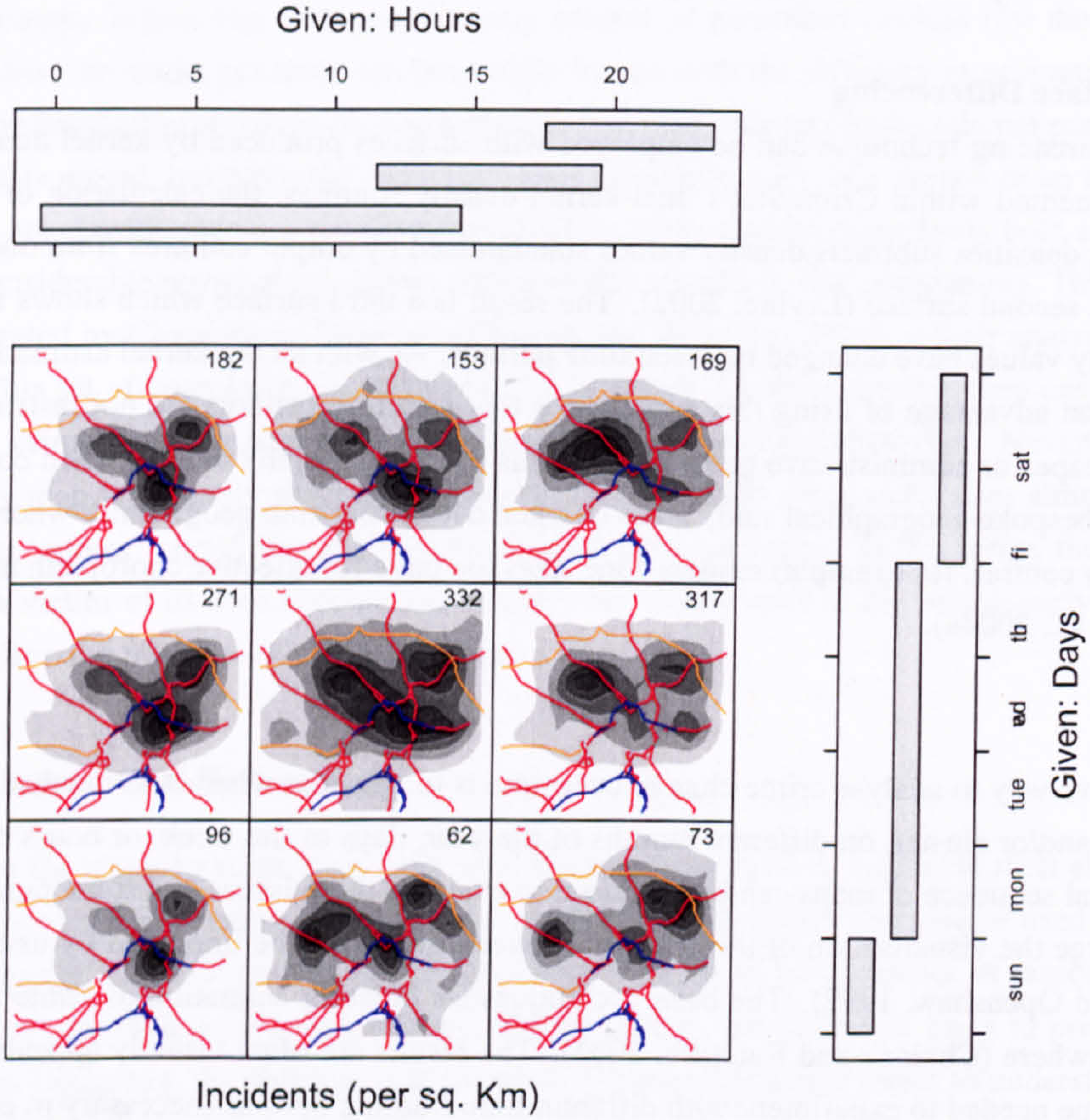


Figure 3.7: Density estimation comap of robbery incidents (2003/04) conditioned by day of week and hour of day. Number of incidents is shown in top right hand corner of each panel.

While the example in Figure 3.7 uses kernel density estimation to create the map, the same 'small multiples' idea can also be applied to condition choropleth maps (Carr et al., 2005).

Other more sophisticated techniques, such as SaTScan (Kuldorff, 2001; Kuldorff et al., 2005), are able to scan data with both spatial and temporal variables in order to search out clusters of records which coincide spatially and temporally. What is more, the calculation of Poisson generalised likelihood ratios enables the evaluation of statistical significance for each cluster, using Monte Carlo simulation. Chiefly still employed for epidemiological studies (e.g. Haining and Cliff, 2003; Kuldorff et al., 2005), SaTScan has started to be used by researchers interested in crime geographies (e.g. Ceccato and Haining, 2004; Zeng et al., 2004), although sometimes for its spatial scanning capabilities only. (There is not space for a worked example of SaTScan here, although a suggestion for a potential application with respect to this thesis is made in Chapter 9).

3.3 Moving 'Beyond Blobology'

Despite the relative maturity of most of the mapping techniques introduced thus far, at the time of the (first) National Crime Mapping Conference, held in London, 2003, there were still many intelligence analysts within police forces and local authorities who were either not using GIS at all, or else were using mapping software with limited functionality (the case in Leeds at that time). The conference was organised by the Jill Dando Institute of Crime Science (JDI), and it was perhaps a surprise to some that in her opening address, the director of the JDI and founder of the Home Office Police Research Group, Gloria Laycock, called for crime mappers and analysts to move 'beyond blobology' (Laycock, 2003).

Blobology was described as meaning the mapping of data geographically for no particular purpose. The criticism was that, although such mapping utilised advanced techniques and employed visually arresting styles, these types of map only facilitated a 'which' approach to problem solving - which neighbourhood has high crime?, which days of the week do crimes cluster at a particular location? The maps too often offered little scope to explain 'why' hotspots occur where and when they do. Cutting through Laycock's address were complaints - similar to those made by Etzioni (1967) - of ritualistic scientism. Moving 'beyond' blobology, maps still had a role but it needed to be understood that "the provisional identification of an area as a hot spot is a first step in deciding where to look for the factors which drive crime .. No-one should believe, ... that it represents more than the first step" (Home Office, 2005b). At the local level, crime mapping should, it was argued, be providing an early warning of emerging problems and be generating hypotheses.

Laycock acknowledged that problems needed to be overcome, such as improving data quality, building suitable computer hardware and software infrastructures, and engaging communities more in mapping public perceptions. Furthermore, research was needed to determine how best to use crime maps as early warning systems, how to avoid problems of data overload, and how best to interweave criminological and crime prevention theories into crime maps. The message at the end of the address was that these problems were not insurmountable, but there would need to be improvement in data quality and data access, and more multidisciplinary research.

At least this is one interpretation of Laycock's address. It will be shown below that some of the issues raised by the *Beyond Blobology* presentation had already been addressed and that sometimes obstacles identified by those in academia, or at least those at some distance from the day-to-day activities of practitioners, could be more perceived than real. Nevertheless, it can be argued that a number of 'blobology' problems are being perpetuated - even by those working at the JDI - and that the general tenor of the 'beyond blobology' argument provides a useful introduction to the ways in which geographers can provide area context to crime mapping and tackle the problem of determining 'why' hotspots occur where they do.

3.3.1 Measuring Risk

Most of the point mapping techniques discussed earlier reflect concentrations of offences, and the maps they produce may be adequate for the effective targeting of police resources to meet operational needs. However, at a strategic level (and within the wider context of community safety partnership working) it is also necessary to go beyond the raw counts of offences (Craglia et al., 2000). One reason for this is that maps based on raw counts of incidents or offences usually give no indication of relative risk. No consideration is given to household density, for example, and it would not be clear whether the number of offences at a particular location was unexpected or not.

A conventional method for measuring risk is to express crime as a number of offences per some standard number of population. In essence, a *crime rate* removes some of the confounding influences of variations in the underlying population between different areas and makes like-for-like (or at least similar-by-similar) comparisons possible. The crime locations may be zones defined by administrative boundaries, in which case it is usually fairly straightforward to estimate a population size. Alternatively, the locations may be defined by standard deviational ellipses or the peaks or troughs of a 3-dimensional surface, in which case the estimation of a population may be more problematic. Within community safety working practices, the denominators used most often are the residential population and the number of households. To keep the rate values to manageable proportions, the denominators are usually expressed per 1,000, 10,000 or 100,000 instances of the denominator variable. Thus burglary, a fairly widespread and frequent offence, may be expressed per 1,000 households, while rarer offences, such as some sexual crimes, may be quoted per 10,000 residential population.

Population Estimates

Accurate population and household figures are produced with each census for a wide range of geographies. In addition, models are used to produce official population and household mid-year estimates in some inter-censal years. The smallest spatial unit used by the Office of National Statistics (ONS) in the production of mid-year population estimates is the local authority. The same is true for the mid-year household estimates which are derived from the ONS population estimates by the Office of the Deputy Prime Minister.

For smaller areas, official mid-year population estimates for wards have been produced periodically by the Social Disadvantage Research Group at Oxford University, the mid-1998 estimates being the most recent example. Unofficial population and household estimates and projections are produced by some commercial geodemographic companies, such as Experian. These data are updated annually for postcode sector geography and in the past has been made available to the UK academic community through the Census Dissemination Unit at MIMAS, University of Manchester.

All of the population estimates described above have a national coverage and thus suffer from having to use data that is available and consistent across countries. For local analysis, alternative sources of information can provide a way to estimate small-area population, the Council Tax Register being a good example. Techniques to provide disaggregate population and household counts using the Postal Address File have also been suggested (Martin et al., 2000), as have techniques that sample population surface models (Coombes and Raybould, 2000; Martin, 2002; Harris and Chen, 2003).

With all of these different sources of denominator information, it is important to keep track of which figures or estimates are being used. This is especially the case when comparing rate-based statistics generated by different agencies, as each may have obtained population information from different sources. Problems with population estimates can also occur when administrative geographies are not clearly understood. For example, research into aspects of the drug misuse problem in Leeds revealed that the Department of Health had been publishing drug treatment rates for some local authorities that were almost twice the correct value. The error was due to confusion regarding the relationship between Drug Action Team (DAT) and local authority district boundaries, which in turn led to an underestimation of the DAT area populations (personal communication with Department of Health, March 2003).

The number and complexity of denominators is usually kept to a minimum to reduce the risk of confusion, but this too may have consequences. The Government's policy of Best Value, for example, dictates that authorities provide measurements of the effectiveness of their social and other policies. Within the domain of community safety, domestic burglary (BV 126) is measured in terms of offences recorded per 1,000 households, which makes sense. Yet, measuring the number of racial incidents recorded by an authority (BV 174) in terms of incidents per 100,000 population without taking into consideration the size of the ethnic population therein, could be argued to make much less sense.

Alternative Denominators

A variety of more appropriate denominators can be adopted for the calculation of rates for particular types of crime. The issue of racially motivated incidents has already been mentioned, and there is scope for denominator change for vehicle crime rates and possibly offences committed by juveniles. The aim in each case would be to develop a denominator that helped to reduce the confounding influence of variations in demographic and socio-economic composition and thus make small-area crime rate comparisons more meaningful.

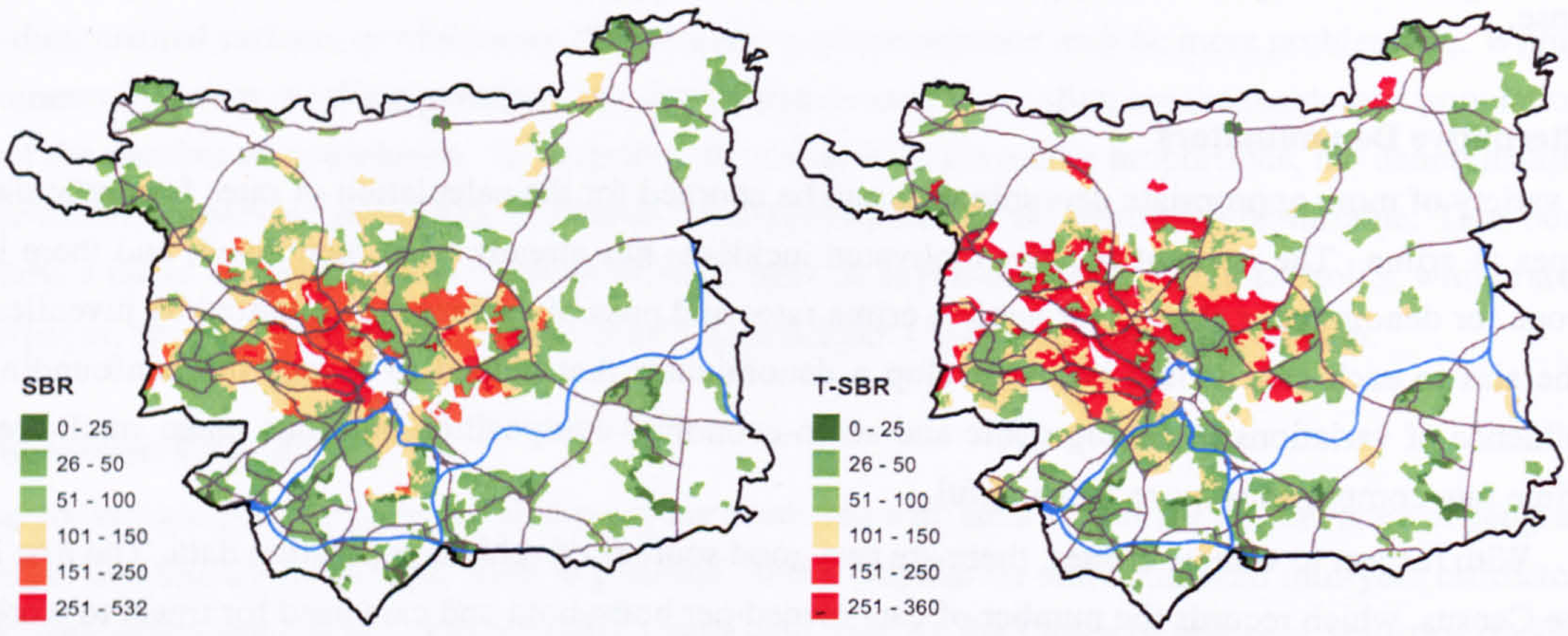
With respect to vehicle crimes, there are two good sources of vehicle population data. The first is the Census, which records the number of cars owned per household and cars used for travel to work, by workplace location. The second is the DVLA Vehicle Parc data, which contains information about vehicle registrations. The two sources complement each other in several ways. Census data are very good for determining how many cars there are in small areas but are only produced decennially and contain no detail about the type of vehicle. By contrast, the DVLA data contain more information about car type, age, make, etc. and are released annually, but are subject to problems related to place of registration of company cars (often headquarters locations) and vehicles sitting in showrooms and standing on garage forecourts.

In addition to using the number of cars as a denominator for vehicle related crime, it would be useful to be more specific about the subset of the vehicle population *most* at risk of theft. This could be tackled by using the Car Theft Index to determine which types of car are most at risk, and the DVLA data to determine how many such vehicles are registered in different areas. Vehicles over 8 years old are consistently identified in the Car Theft Index 2001 (Home Office, 2001a) as being the most likely to be stolen, reflecting, for example, the ease with which vehicles can actually be entered and hot-wired; the types of car that can be afforded in areas with high crime rates; and familiarity with the vehicles amongst the criminal community. Yet, the creation and manipulation of crime rate denominators in the ways mentioned above can become complicated. An alternative approach is to keep with the simple population and household denominators and demonstrate relative risk by standardising the rates.

3.3.2 Standardisation Techniques

Figure 3.8(a) shows the standardised burglary rate (SBR) by neighbourhood, for Leeds. The map is created by computing the district-wide household burglary rate (the number of burglaries divided by the number of households) and then for each neighbourhood, multiplying this rate by the number of households in the neighbourhood. This defines the expected number of burglaries for the neighbourhood, assuming that burglaries occur randomly across the district. The observed number of burglaries in the neighbourhood is then divided by the expected number, and this proportion is multiplied by

100. Thus, a neighbourhood with an SBR 200 would have a real burglary rate twice that expected given the population size, while a neighbourhood with an SBR of 50 greater would have a real burglary rate half that expected. For these and subsequent maps the neighbourhood geography has been clipped to developed land use areas (undeveloped and rural land being shown as white). This helps to reduce attention being drawn to large zones with very low population density and also helps the reader to orientate the map without recourse to extensive labeling.



(a) SBR (b) Townsend SBR

Figure 3.8: Standardised burglary rate by neighbourhood (excluding city centre), 2003/04.

In this example, the size of the neighbourhoods is quite large and one years worth of burglary data produces reasonably high counts of burglary per neighbourhood. Where extremely high or low SBRs based on on relatively small numbers of crimes are computed, it may be necessary to use techniques such as Bayesian adjustment to drive the SBRs towards the neutral value of 100 (Clayton and Kaldor, 1987).

Acknowledging that burglaries rarely do occur randomly across a local authority district such as Leeds, an alternative approach is to summarise the effect of a number of different confounding variables together using some general measure or classification of socio-economic circumstances (Jolley et al., 1992). Craglia et al. (2000) demonstrate such an approach by computing a Townsend index of material deprivation (Townsend et al., 1988) for each census enumeration district in their study area (Sheffield), then splitting it into quartiles before producing Townsend-class specific burglary rates. Similar approaches can be found in epidemiology, although deprivation scores may be subjected to cluster analysis routines in order to find more natural groupings than simple quantile intervals (Haining et al., 1994).

Using this technique, a simple deprivation classification based on unemployment, absence of a car, overcrowding and housing tenure can be created. Figure 3.8(b) shows burglary in Leeds standardised by Townsend quintiles, and there are a number of differences in the result compared to the SBR map (Figure 3.8(a)). The most obvious of these is the more diffuse pattern, with neighbourhoods on the northern fringes of the main urban core and some northern rural settlements showing higher than expected burglary rates. By and large, these are affluent neighbourhoods. They do not have very high burglary rates when compared with city centre neighbourhoods, for example, but the burglary rates are high when compared to other neighbourhoods with similar Townsend deprivation scores.

Craglia et al. were careful to question the appropriateness of such methods for identifying crime risk, and pointed out that "if not used with sensitivity, this method has a risk of mistaking cause and effect" (Craglia et al., 2000, page 728). Importantly though, they did not rule out a classification approach, and in line with the research aims, suggested that "from a policy perspective and in order to develop crime prevention strategies, it is relevant to try and identify those areas that need further study because they have a much higher (lower) rate than expected" (ibid, page 728).

3.4 Providing Area Context

Although standardisation techniques can identify areas which have atypically high or low crime rates, the patterns of crime do not allow the investigator to analyse, in depth, the relationships between levels of crime and the social and physical environment (Hirschfield and Bowers, 1997a). To address these shortcomings, techniques to provide area context must be developed, and this is usually done in one of two ways. The first is to use a suite of variables indicative of neighbourhood type to create data summaries of areas under investigation. The user of the *area profile* then interprets these data summaries, perhaps comparing local variable values against a global mean. The area profiles may be produced for pre-defined areas, or generated dynamically using a GIS comprising spatially integrated layers of information.

The second approach is to use an area classification. In this case, some appropriate variables are used to group together neighbourhoods into distinct and separate classes across the entire study area. The area portraits painted by these groupings can then be used to summarise neighbourhood characteristics. Some systems employ a combination of both techniques.

3.4.1 Single Area Profiles

An example of an area profiling tool that provided social, demographic and land-use information for the analysis of crime patterns was that developed during the course of a large, two-year study into the relationship between crime and disadvantage on Merseyside (Hirschfield et al., 1995). As well as containing crime and disorder data, 'the Profiler' database contained information on population and land use, guardianship, social disadvantage, social disorganisation, and neighbourhood profile (Hirschfield and Bowers, 1997a). Information was produced at enumeration district (ED) level and spatial intersection routines could produce profiles for single EDs; a bounded area comprising multiple EDs (e.g. a regeneration zone); a circular area about a point location; or a crime hotspot ellipse defined by a series of grid references. A modified version was also created to provide area profiles of areas neighbouring the study zone. Table 3.1 shows an example of area profiles produced for two disadvantaged areas in Merseyside, and is very similar to area profiles produced elsewhere (e.g. Berry and Jones, 1995).

Further to these academic studies, partnership working has greatly increased the sharing of data to support and enrich spatial crime analysis within local authorities and police forces. To manage the task, as was discussed in the Chapter 2, partnerships are developing Partnership Business Models (PBMs), information portals and planning support systems to summarise disparate variables at a neighbourhood level. Locally, the Leeds Statistics project (http://www.leeds-statistics.org/) provides area profiles for a range of geographies using local and national datasets. In the case of administrative geographies, these profiles are pre-defined, although it is also possible to produce queries for

	Disadvantaged and low crime	Disadvantaged and high crime	
	Walton/Bootle	Norris Green	*Merseyside*
Radius of area (m)	1,175	1,020	-
Residential population	7,694	7,693	1,400,275
% Land area non-residential	16.8	82.1[a]	68.2
% Land area residential	83.2[a]	17.9	31.8
Segregation no car	9.55	17.14	-
Segregation unemployed	8.59	10.71	-
% Recent migrants	4.5	9.5[a]	7.3
Index of ethnic heterogeneity	0.009	0.032	0.036
Index of social heterogeneity	0.864[a]	0.858[a]	0.830
% Risk of calling (burglary)	0.8	3.0[a]	1.65
% Risk of calling (all property)	4.5	13.8[a]	8.09
% Risk of calling (minor disorder)	5.3	10.9[a]	5.735

[a] indicates the value is higher than the Merseyside average

% Risk of calling = No. of calls to the police p.a. per 100 residents

Table 3.1: Statistical profiles of disadvantaged areas with high and low levels of burglary, Merseyside (Hirschfield and Bowers, 1997a, page 112)

customised zones. A community safety annexe to Leeds Statistics, with restricted access, summarises more detailed and sensitive crime and disorder data and also provides a small subset of the contextual data found on the parent website. Other examples of crime and community safety information sharing portals that provide area profiles include COMOS (Birmingham), Project Dragon (used in parts of Wales), the NERISS Project (North East Government Office Region), ComPaSS (Northamptonshire), SCaDIS (Surrey) and PANDA (Wakefield). (Some of these systems, plus others, have been evaluated for the Home Office by Chainey and Smith (2006)).

While it can be argued that such initiatives are commendable, for the most part, the single area profiles of the sort produced by these systems are deficient in one important respect. That is, it is rarely clear whether the data profile is describing a typical or an atypical neighbourhood. Table 3.1, for example, compares two zones with a global mean. The differences are clear in a table of this size, but if 7,700 is taken as the average population of a study zone, then how different are Walton/Bootle and Norris Green from the other 180 or so zones? Presuming other zones are similar (other council estates, for example), are the crime risks similar there as well? For planning or evaluating community safety initiatives, it is argued that answers to such questions can help considerably in the assessment of crime problems and community safety initiative outcomes.

Within Leeds, 33 area profiles, one for each ward, are very difficult to compare and contrast within the Leeds Statistics system. The readability problem is reduced if the geographical scale is increased to management area level (there are 10 management areas) but then the heterogeneity of the zones significantly reduces the usefulness of the profile. It could be argued that a geography similar in size to lower level super output areas might be an appropriate scale for small-area crime analysis, in which case a table of 477 zones becomes hard to construct, let alone read. Some alternative method of mining area profile data is required in such situations, a method that can identify homogeneous areas whose characteristics can be summarised succinctly - rather in the manner of the Townsend classes used to standardise burglary rates in Section 3.3.2, but reflective of a wider set of neighbourhood conditions. One such method is geodemographic profiling, also known as area classification.

3.4.2 Geodemographic Profiling for Crime Analysis

Technical aspects of geodemographic profiling are considered later, but to begin with a simple definition, geodemographics is the "analysis of people by where they live" (Sleight, 1997, page 16) founded on the notion that "birds of a feather flock together" (Flowerdew and Leventhal, 1998). More practically, geodemographic classifications typically comprise a number of different *classes* to which *objects* are assigned. The assignment is executed in a way that aims to group together similar objects into the same class whilst at the same time ensuring classes are distinct from one another. Objects may be neighbourhoods (usually residential), census output areas, postcodes, households or even individuals. Once objects have been assigned to classes, the within-class object characteristics are summarised in order to produce a portrait, or profile, for the class as a whole. Thus, use of a geodemographic classification will produce a description of a neighbourhood that is more general than the single area profiles described above.

The advantages of a classification approach to providing area context, however, are that other neighbourhoods with similar characteristics are grouped within the same class, and are thus easily identifiable. In this sense, the classification acts to stratify the population into distinct samples. Following Craglia *et al* (2000), this can enable atypically high crime rates to be identified by neighbourhood type, rather than just relying on analyses that identify high or low crime rates across the district as a whole. In some instances it may also become more straightforward to form general hypotheses about which types of neighbourhood suffer from which types, or mixes, of crime and disorder.

These possibilities for interesting analysis perhaps help to explain why it has been argued that since 1997, there has been a renaissance of interest within academia and government in the use of geodemographic profiling and area classification (Longley, 2005). Within government, broadly speaking, this has been driven by the pursuit of evidence-based policy. More specifically, there has been a need to set rational performance targets for public service delivery and a desire to improve efficiency of public services by targeting those most in need or at risk. Within academia, the establishment of centres such as the ESRC Centre for Neighbourhood Research shows a "recognition of the importance of local context and policy responsiveness to the needs of different local communities" (Longley, 2005, page 59). Furthermore, public service agencies now produce large quantities of quality digital data, which along with lifestyles data collected for commercial classification purposes are thus enabling the development of new digital data infrastructures (Longley and Harris, 1999).

Within the crime reduction and policing domain, geodemographic profiling has been applied in a number of academic studies. For example, the Super Profiles classification (Batey and Brown, 1994) has been used to explore links between crime and disadvantage in Merseyside (Bowers and Hirschfield, 1999), to demonstrate the utility of geodemographics for community safety analysis (Brown et al., 2000), and to evaluate crime and disorder on public transport systems (Newton, 2006). MOSAIC, a commercial geodemographics consumer segmentation tool from Experian, has been used to demonstrate the need to understand neighbourhoods when analysing crime data (Williamson, 2003); to aid the deployment of police resources (Ashby and Longley, 2005), to explore police performance assessment (Ashby, 2005), and to analyse patterns of youth offending and schools at risk (Hayden et al., 2005; Williamson et al., 2005). Practitioner case studies of crime prevention work or crime analysis using geodemographics have included a survey of sites and signs of crime and disorder (Kenwood, 2005), target analysis for reassurance policing (Scott, 2005), and an analysis of socio-demographic risk factors for domestic violence (Vincent, 2005).

All of the studies above have used an existing, general purpose geodemographic classification. By contrast, task-specific classifications have been constructed to study social stress and trauma (Harries, 1997), the spatial displacement of crime resulting from urban renewal programmes (Schumacher and Leitner, 2002), and the identification of crime hotspots (Shepherd et al., 2005). In addition, and as mentioned in Chapter 1, the production of a task-specific classification for comparing Crime and Disorder Reduction Partnerships (CDRPs) and police Basic Command Units (BCUs) has been created (Harper et al., 2001, 2002; Sheldon et al., 2002; Hall et al., 2003a,b) as has a classification of most similar police forces (Home Office, 2003a,b, 2004d).

Background to Geodemographic Profiling and Classification?

More generally, from the 1920s onwards, urban sociologists and geographers in academia have been interested in classification systems for developing urban theory. Already mentioned with respect to the development of criminological theories, Park and Burgess of the Chicago School of Sociology made early attempts to define natural areas in cities, and these persisted for some time in Chicago (see Rees, 1972). More methodologically rigorous attempts at within-city classification became more common with the development in the USA of census tables for the emerging tract geography, and with the social area analysis of Shevky and Bell (1955). In many instances, the rationale for much early academic work on area classifications was to confirm concepts and general principles about the internal spatial and social structure of cities. A different, more pragmatic approach has typically been adopted by urban planners (Batey and Brown, 1995), and the current renaissance in geodemographics amongst crime and community safety analysts is arguably more in tune with this perspective.

Broadly speaking, the concern of urban planners is still to summarise social and economic variation within cities, but to do so for practical purposes related to policy formulation and delivery. "In many cases the aim has been to develop a consistent and systematic approach to resource allocation, involving the definition of priority areas to receive favoured treatment" (Batey and Brown, 1995, page 77-78), a common theme in the studies cited above. There is also agreement between Batey and Brown's observation and recent applications, that the "area classification here ... is constructed using census variables, sometimes supplemented by other sources of data" (Batey and Brown, 1995, page 78), although this can also be used to subdivide the type of geodemographic profiling being undertaken by crime pattern analysts.

This division is between studies and applications that use pre-built classifications, usually of a general-purpose nature, and those that develop new, task-specific classifications to analyse specific crime issues. The distinction was introduced above by the order in which the examples were presented, and it can be argued that the distinction is important because of concerns that "it is ... doubtful whether satisfactory general purpose classifications can ever be devised" (Openshaw and Wymer, 1995, page 243). This view actually reflected concerns over the choice of variables used for general-purpose classifications, although others have argued that problems of within-class diversity inherent with some general-purpose classification support the argument that "special purpose classification systems based on variables of immediate interest would seem more useful that 'generic' products" (Voas and Williamson, 2001, page 73).

Other definitional issues remain. The field of geodemographics has been described as "a short-hand label for both the development and application of area typologies that have proved to be powerful discriminators of consumer behaviour and aids to 'market analysis'" (Brown, 1991, page 221). Indeed, it has been suggested that geodemographics and market analysis are virtually synonymous

(Birkin, 1995). This perhaps contradicts the idea that commercial geodemographic classifications used extensively for market analysis - such as MOSAIC, ACORN, CAMEO, or PRiZM - are general purpose tools. If geodemographics is synonymous with market analysis, then either the tools such as those mentioned are in fact task-specific classifications, that is, their task is consumer segmentation for marketing purposes, or else it might be more appropriate to describe genuinely general-purpose classifications - such as those produced by the Office for National Statistics (ONS) - as area classifications, or neighbourhood classifications if the scale is appropriate.

For the purposes of this research, the view will be taken that geodemographics is not synonymous with market analysis. Market analysis has, and will probably continue to be, the dominant application for geodemographics, but as has been discussed above, there are a growing number of uses for the technology being found in public service delivery. Instead, within this thesis, the terms geodemographics, area classification and neighbourhood classification will be used synonymously, with a distinction between classifications being made according to their intended specification - either as general-purpose or task-specific tools. A further distinction that can be made is whether the classifications are developed commercially or as tools to be given away for free. This is much less important, although commercial considerations can cause problems, as will be discussed below.

3.5 Towards a New Neighbourhood Classification for Community Safety

General weaknesses with the geodemographic approach include: the cross-sectional nature of systems and their usual inability to say much about change over time; the partial nature of systems that rely solely on census data (following Etzioni (1967)); and ecological fallacy problems encountered when inferences are made about individuals based on the classification of the wider area in which they live (Birkin, 1995). These and other issues are discussed in more detail below, with the objective of determining what type of classification the research ought to be attempting to produce.

3.5.1 Task-Specific Classification

As discussed above, classifications can be classed as either general-purpose or task-specific in nature. The obvious advantage of a general-purpose system is that it might be applied successfully to a number of different problems. The less obvious disadvantage is that neighbourhood characteristics useful for one application may obscure or cancel out the effects of neighbourhood characteristics that might benefit a different application. The task-specific response to this problem involves classifying using only those variables most applicable to the task in hand, and the discounting of variables that might mask or add 'noise' during the computation of the classification. When, for example, the Home Office set the terms of reference for the CDRP and BCU classifications mentioned earlier, the contractors were asked to collect the most applicable data for BCUs and CDRPs (Harper et al., 2002). Following this suggestion, theories used to explain patterns in the distribution of crimes - for example, routine activities theory - were used to identify variables, as was the Police Funding Formula and other literatures relevant to policing. The final number of variables used was just 20. A very similar approach was adopted by Harries (1997), identifying 24 variables representing social stress and violence, whilst, only 4 socio-economic variables were used to classify neighbourhoods in Schumacher's study (2002) of burglary risk and displacement.

It could be argued that the actual number of variables used to create a task-specific classification ought not to matter, it is the relevance of the variables to the study being undertaken that should be the determining factor in whether they are included or not. Yet, despite serious concerns with 'throwing everything into the pot' (Fowlkes et al., 1987; Kaufman and Rousseeuw, 1990; Gordon, 1999; Openshaw and Wymer, 1995), this appears to be the approach to variable selection adopted for some commercial general-purpose classifications. The result can be the use of several hundreds of variables, from common sources such as the census, but also from more diverse sources, such as shareholder registers, bankruptcy databases, land registry databases and consumption-based lifestyle databases.

Non-commercial general-purpose classifications, such as those published by ONS, may restrict themselves to using census variables only. Furthermore, constraints on variable selection may be imposed according to the geographical extent of the classification. For example, a UK classification that utilises the three censuses from England and Wales, Scotland and Northern Ireland - which do not all ask the same questions - must harmonize the variables down to a common set (for example, see Vickers et al., 2003). Furthermore, different numbers of variables may be used for different geographical scales, reflecting the likelihood that smaller areas are likely to be more homogeneous, and also reflecting the fact that at smaller scales distributions may become problematic, e.g. include a high proportion of zero counts (Vickers et al., 2005).

3.5.2 Mutually Exclusive Membership

As well as choosing a variable selection strategy, it is also necessary to identify the most appropriate model of cluster membership. So far, the only model described is one with a mutually exclusive membership policy that assigns objects to just one class. The advantages of this type of classification are that fixed and clear membership lists make them more intuitive and easy to understand. This is the most common type of classification membership model, and was chosen for the original Home Office CDRP Families classifications (Harper et al., 2001, 2002; Sheldon et al., 2002). Hence, local authorities were assigned to one of 13 'families' of like-CDRPs. Leeds was allocated to 'Family 4', along with 11 other local authorities (Table 3.2). As some of the variables used in the classification were published annually the classification process could be repeated and changes to family membership considered (Hall et al., 2003a,b), although during the life of the CDRP Families classification only one CDRP, Thanet, changed family.

CDRP Family 4	
Birmingham	Manchester
Bradford	Middlesbourough
City of Kingston upon Hull	Newcastle upon Tyne
Leeds	Nottingham
Leicester	Sheffield
Liverpool	Wolverhampton

Table 3.2: Membership of CDRP Family 4, 2001.

An alternative approach is to acknowledge that the mathematical techniques usually employed to create classifications may place an object in one class, while in numerical terms it is actually 'closer' to the centre of another class - making it an outlier within its class. A solution to this scenario is to group each object with some number of its 'nearest' or 'most similar' objects, having the effect of creating as many classes as there are objects. Although less intuitive to use, the groupings can turn out to be more compact, and thus more reliable, than mutually exclusive groupings. This was the approach adopted for the 'most similar forces' classification used by the Home Office to aid the assessment of police force performance (Home Office, 2003a,b, 2004d). It is also the most recent approach adopted by the Home Office to group together CDRPs for performance comparison (spreadsheet available at http://www.homeoffice.gov.uk/rds/pdfs04/comparisons04.xls). In this instance, each CDRP is grouped with the 'nearest' 14 other CDRPs, in terms of 20 variables selected by strength of correlation with crime rates. Table 3.3 shows those CDRPs computed as being most similar to Leeds. The membership is significantly different from the old Family 4, although how much of this is due to the new approach to group membership and how much is due to some different and more timely variables (e.g. from 2001 census) cannot be deduced, as details of the methodology have not been published. (Note: Some methodology information was circulated during the consultation prior to release of the new lists. The accuracy of this, however, must be questioned given that there were many changes to the memberships in the final classification published alongside the *Crime in England and Wales 2003/04* reports (Dodd et al., 2004).)

Nearest CDRPs (closest first)
Sheffield
Preston
Cardiff
Plymouth
Lincoln
Swansea
Bradford
Derby
Newcastle-upon-Tyne
Stoke-on-Trent
Northampton
Kirklees
Sunderland
York

Table 3.3: CDRPs most similar to Leeds, 2004.

A third option is to acknowledge that despite best attempts to create neat and compact clusters, some objects may defy clear-cut classification, and numerically may be similarly close to two or more class centres, making cluster membership much less obvious (Gordon, 1999). Fuzzy clustering techniques, based on the principles of (non-binary) fuzzy logic, address this problem by attributing each object with membership of every class and then assigning a membership 'strength' to indicate which class(es) an object 'belongs to most'. Fuzzy classifications have tended to find fewer applications outside of academia. They have also been criticised as not being of any great benefit when consideration is given to the rather nebulous nature of the classes themselves (Voas and Williamson, 2001).

3.5.3 Geographical Scale

Classifications based mainly on census data will have their zonal scale constrained by the available census geographies. Scaling up is rarely a problem, although cumulative errors resulting from the effects of disclosure control mechanisms may mean that tables published only at the smallest scales have to be aggregated with caution. In cases where a smaller geography is required - for example, to postcode, household or individual level - a reliable means of apportioning the census data must be found. There are a number of possibilities, including use of the Postal Address File or population surfaces (see Section 3.3.1), but then licensing conditions forbid attempts to identify individuals in the census data, so for the time being these possibilities remain academic.

An increasing amount of other types of data are becoming available at sub-census output area. Commercial sources include consumer product-guarantee returns, store loyalty cards and recorded travel behaviour (Longley and Harris, 1999), although the sample sizes for small areas may be very small and thus problematic (see discussion on the Target Group Index in Birkin, 1995, page 116). Local authorities are also making increasing use of household and individual level data, although this tends to be for specific, local projects, and this data have rarely been collated nationally. The English Index of Deprivation 2004 is a case where this been attempted, albeit to a super output area geography (lower level). Yet, even with the backing of the Office for the Deputy Prime Minister, the reliability of the data passed on to the Index designers was sometimes questionable (see discussion of the crime domain ODPM, 2004, page 39).

In addition to practical considerations, choice of scale ought also to depend on the application to which the classification is being used or designed. For market analysis, where individual households can be targeted using direct mail campaigns, a household-level classification might be the best solution. In urban planning scenarios, however, it would probably be unusual to discriminate to that degree, and indeed, even despite the availability of small-area data, much planning and performance monitoring for community safety policy in Leeds is still conducted using ward geography. It is argued that this geography really is too coarse for analysing crime patterns effectively, and smaller geographies such as output areas or super output areas should be considered. However, these geographies can still be criticised for not necessarily reflecting qualitative definitions of community or neighbourhood understood by the residents therein (c.f Lee and Campbell, 1997), or analytically definable social entities (Morphet, 1993; Martin, 1998). Moreover, whichever zonal system is chosen, the modifiable areal unit problem (Openshaw, 1984) is likely to apply.

Choosing a scale for the classification is problematic and care has to be taken not to choose an inappropriate scale for the application in hand. For example, the commercial consumer segmentation system Mosaic, produced by Experian, has been the classification of choice for a number of academic studies mentioned above, and is also being purchased and used by increasing numbers of police forces and local authorities. It is possible that this practitioner interest has in part been as a result of published and presented academic work, helped in part by the success of the *National Crime Mapping Conference* organised by the JDI. Yet, Mosaic classifies at the unit postcode level, each of which in Leeds would only capture an average of approximately 16 households. This is a very small area with which to sample crime unless multiple years and multiple crime types are aggregated together. Furthermore, the accuracy of crime geocoding is notoriously variable (Ratcliffe, 2004). One years worth of burglary dwelling data in Leeds, for example, would only produce approximately 0.5 of an offence per postcode per year, at 2005 levels. Thus, while between-class styles of crime analysis would be feasible, within-class crime analysis would be problematic, and perhaps even impossible, especially for the neighbourhood types with low crime levels. In the studies citied above, these limitations are not usually discussed.

In contrast to the very small-scale of postcode geography, problems can also occur when geodemographic classifications are based upon larger geographies. This is because the temptation to infer that the characteristics identified for an area also apply to individual households or persons therein - the ecological fallacy - has more serious consequences when the scale is large and the zones correspondingly less homogeneous.

3.5.4 Homogeneity

Homogeneity within a classification is important in two respects. Firstly, and linked to the problem of geographical scale, is within-zone homogeneity. Ecological fallacies incurred by the classification become less of a problem as zone homogeneity increases. Thus, the choice of a smaller scale is likely to be preferable, as this is more likely to group similar households and other neighbourhood attributes, following Tobler's first law of geography - that everything is related to everything else, but near things are more related than distant things (Tobler, 1970). Balancing within-zone homogeneity with the other scale issues mentioned above is one of a number of subjective classification design decisions that must be made.

The second homogeneity issue is within-class homogeneity. The benefit in achieving this is the greater chance of the class profiles being a reliable representation of all objects in the class, and greater reliability of between-class and within-class analyses. Homogeneity improves as the number of classes increases, up to a theoretical maximum when a class contains just one neighbourhood. Large numbers of classes may be awkward to work with however, and thus to an extent homogeneity must be balanced with usability issues and thus arises another subjective design decision. Fortunately, homogeneity can be expressed numerically and thus different classifications can be compared (see Voas and Williamson, 2001), lending some objectivity to the decision making.

Aside from scale, it might also be desirable to surrender some homogeneity in favour of greater spatial compactness of the classification geography. If this is the case, rules may be employed to merge classes in such a way as to minimise the fracturing, or alternatively, spatial clustering methods can be used to incorporate measures of spatial proximity into the classification routine itself. Again, which approach is adopted will depend on the uses to which classification is to be put. It has been argued that from a service delivery perspective, a classification with good spatial contiguity might be preferable, while compactness might not actually be necessary (Haining et al., 1994).

3.5.5 White-Box Design

Whichever design decisions are made with regards to the criteria discussed above, it is argued that these should be clearly documented and made available, eschewing the black-box approach and acknowledging that "understanding the strengths and weaknesses of a GIS providing evidence is *key* to informed management decisions" (Longley et al., 2005, page 416 (emphasis added)). In turn, Curry's (1995) concern regarding the complexity of technological systems would also be acknowledged, although it is argued that users almost always have to put 'faith' in classification systems at some stage, although with access to methodological decisions and measures of fitness for purpose, hopefully the move would be more of a 'step' than a 'leap'.

Discussions regarding the strengths and weakness of specific geodemographic classifications, as different from the approach in general, are almost always lacking in the crime analyses that have employed commercial geodemographic products. In large part this is because the originators of these products have to withhold such information in order not to surrender competitive advantage. However, the idea that scrutiny is not necessary because the technology is 'proven' (Ashby, 2005) is not accepted.

Some information about the classification design may be provided, such as indications of the extent to which characteristics used in the classification are over- or under-represented. Yet, there is rarely any indication, for example, of how typical a postcode area is of its class, nor how homogeneous the class is as a whole. Moreover, even the charting of class characteristics is only of marginal value, as no indication is given of the weightings applied to the different variables during the cluster analysis routines - themselves also not disclosed. Bowers and Hirschfield (1999) have made clear the point that when large numbers of variables are used in a classification process the class portraits may lack transparency and can give the impression of being 'black boxes', yet too few of the recent studies appear to be noting this.

With full disclosure of classification methodology, plus class portraits and fitness for purpose heuristics, users ought to be better able to make more informed judgments about any crime patterns the classification may reveal. For example, if a neighbourhood is highlighted as having an unusually high crime rate for its type, data pertaining to the robustness of the class as a whole, along with the quality of fit of that neighbourhood within its class, could be used to decide whether the outlier status conferred by the analysis was justified. Such heuristics are straightforward to generate and not difficult to communicate to users. Some have gone as far as to allow interactive clustering for data mining (Guo et al., 2003), recognising the importance of "allowing the human-computer interaction needed to effectively tease-out complex patterns" (Guo et al., 2003, page 229). It is argued that this level of sophistication might not be necessary or appropriate for the intended users of this research, yet it appears to offer further support for opening up the black-box in the ways that have been suggested.

3.6 Concluding Remarks

It is argued that geographers and geography have made, and continue to make, important contributions to crime pattern analysis. These contributions may either be thought of singly as the development of cartographic and area profiling techniques, although the combination of the two contributions, following Craglia et al. (2000), offers a third way of crime mapping to satisfy Laycock's (2003) call to move 'beyond blobology'. Accordingly, the development of classifications which offer greater sophistication than the simple dividing up of ranked Townsend deprivation scores would seem a natural next step to take, although at first glance it is perhaps curious that the recent interest in geodemographics for crime pattern analysis has largely overlooked this possibility.

However, the omission becomes easier to understand when some of the limitations of geodemographics, and some commercial systems in particular, begin to be considered. Products such as Mosaic and Acorn, for example, may come with sophisticated visualisation packages, but the postcode scale, black-box nature of the system, and design skew towards consumer segmentation applications raise too many questions to proceed with confidence with analysis along the lines of Craglia et al. (2000). By contrast, it is argued that the classification of BCUs and CDRPs by the Home Office provides a good, open model for the type of classification that the research is planning to create, with the caveat that the neighbourhood scale being proposed may make different demands on data availability (discussed in the next chapter).

Objections regarding the social implications of geodemographics and GIS have not been considered to any great degree, although some acknowledgment has been made to the arguments raised by Curry (1995; 1998) and Goss (1995) which were collectively presented with quite some impact in

Ground Truth (Pickles, 1995). This publication caused much reflection and no small amount of defense from those at the vanguard of GIS developments (see volume 30 of *Environment and Planning A*, 1998), although Sibley usefully argued that using an exploratory data analysis (EDA) approach, following Tukey (1977), quantitative analysis techniques of certain types "could encourage an open-minded approach to spatial ordering through emphasis on the continuous interplay or theory and data" (Sibley, 1998, page 243).

Collectively, EDA techniques are designed to enhance pattern recognition and to uncover data structure, and among their traits, simplicity of use is the most common (Sibley, 1990). Whether a geodemographic classification could be described as simple to use is uncertain. In using a classification to standardise crime rate maps and to identify within-class crime outliers, however, it does appear similar in some respects to the pre-computer EDA tools, in that it does encourage repeated reference to the original data and to patterns in the data, making it easier to identify interesting cases as the analysis proceeds. It is argued that a classification that operates as a black-box would have less to offer EDA, as residuals which it might uncover could not be investigated with the same rigour or freedom as a classification which allowed the residuals' goodness of fit to be scrutinised.

These issues pertaining to exploratory data analysis are tested in Chapters 8 and 9, and will become clearer when the final classification has been described. This leaves the questions prompted by the remainder of this chapter regarding what type of classification is likely to be most appropriate, at least in broad terms, at this stage. To begin with, as the arguments thus far have suggested, the most appropriate type of classification would appear to be one designed for the specific purpose of analysing crime data at a neighbourhood scale. There does not appear to be such a classification for the analysis of crime patterns in Leeds, so one must be created. Bearing in mind that the classification will need to be used by people with different levels of analysis experience, the more intuitive mutually exclusive approach to class membership would appear the best choice, although this could be reviewed once prototype classifications begin to reveal more about the goodness of fit and homogeneity of classes.

It is not intended to experiment with scale in this thesis. It is an interesting issue, but the opportunity cost would be less time to evaluate the classification against a range of real problems - something that critics have argued is most important, and often overlooked (Dubes and Jain, 1979; Rapkin and Luke, 1993). An initial motivation of the research was to deploy the 2001 census data, so this will have considerable bearing on the choice of scale, as will the availability and characteristics of other datasets. As these are discussed in more detail in the next chapter, the final decision on spatial scale will be deferred. However, the postcode geography adopted by many commercial classifications would cause problems for within-class analysis, so a larger scale than this will be sought.

Homogeneity within zones will be determined by the choice of classification geography, while a desire to undertake within-class crime pattern analysis will demand as much within-class homogeneity as can be achieved economically while at the same time having a classification that is not over-burdened in use by a large number of classes. By the same token, it is argued that while spatial contiguity may be desirable, it will not be pursued at the expense of within-class homogeneity. Finally, whatever the eventual form the classification takes, information on design decisions, goodness-of-fit tests and test against external criteria that are independent of the data being studied will be generated and presented. This is in order that the relative strengths and weaknesses of the classification can be scrutinised and patterns of crime analysed by all the best principles of exploratory (spatial) data analysis.

Chapter 4

Selecting Variables

4.1 Introduction

"Meaningful analysis of the residential differentiation of the urban population demands the availability of the right sort of data for the right sort of units" (Timms, 1971, page 39). This chapter discusses decisions regarding the selection of data used to derive the Leeds Classification for Community Safety (LCCS) and describes general features of the data sources from which the variables were drawn. Following this, a detailed description is given of the production of the new neighbourhood geography for Leeds, used for the classification.

From the outset, an objective of the research has been to identify data produced locally within Leeds at an individual or household level, particularly for inter-censal periods, as many of the datasets on which geodemographics has traditionally relied are published infrequently, at too coarse a geography, and do not capture short-term population dynamics (Longley and Harris, 1999). Nevertheless, attention has also been given to the possibilities of using local data extracted from large national datasets where characteristic completeness of coverage and scientific rigor make them attractive, plus in some cases there is no local equivalent source. Products such as Ordnance Survey MasterMap have been obtained under a special academic agreement in order to match the resources which are available within Leeds City Council. In the converse of this situation, datasets available via university agreements were not used unless they could also be identified as also being available within the Leeds local authorities. In this way, transfer of ownership of the classification would be more straightforward and the research might be more easily repeated by others.

4.2 Garnering Opinion

A decision was taken early on to make up for a lack of expert knowledge on crime and antisocial behaviour issues by consulting widely regarding the variables that ought to be included in a classification for community safety. Identifying data sources is inevitably interwoven with the identification of individual variables, and sometimes, the availability of a new data source sparked interest in the possibilities of the variables contained therein. On other occasions, the variable being sought is known and the problem is to identify where it might be recorded. The principle approach adopted in these consultations/surveys was to identify variables first and consider availability second. The following sections describe the results from three surveys that were undertaken.

4.2.1 National Community Safety Network

The National Community Safety Network (NCSN) is a practitioner-oriented organisation supporting those involved in promoting and delivering community safety policy in the UK. The author was invited to lead an interactive workshop at the NCSN Annual Conference in Belfast, June 25th-27th 2003, and this was used to gather opinion on the types of variables that were already being used in classifications such as the Home Office's CDRP 'Families' (Harper et al., 2002), and suggestions about what variables might be useful in a new neighbourhood-level classification.

The workshop consisted of approximately 70 delegates working in community safety within local authorities and police forces in the UK and Eire. Delegates were split into 8 groups, each focusing on one of a number of specific themes: health, social capital, built environment, education, social, demographic, economic and land use. Each focus group was asked to discuss the types of variable within their given domain that would be of value to the analysis of community safety issues. It was explained that ideas could include variables that might not be readily available, although undoubtedly the extent of people's exposure to existing data will have influenced their responses. During the discussion time the author visited each group and facilitated where necessary. Groups summarised their variable choices, and subsequently these were published on the NCSN website for the session attendees to share and consider. The complete list of variables suggested by this exercise is shown in Table 4.1.

Conspicuous by their absence are any variables that express fear of crime, even though this issue is prominent in definitions of community safety (Section 2.4). 'Satisfaction with area' is likely to be affected by fear of crime but will probably capture much else besides. Interestingly, a number of variables were suggested which capture something of left realist arguments about the causes of crime (Young, 1992). These include a number of *access* variables (e.g. access to community groups, access to community forums, and youth inclusion in community groups), *economic* variables (e.g. job creation schemes, levels of inward investment, and presence of regeneration investment); and *education* variables (e.g. class sizes, sport facilities, and after school provision).

4.2.2 Leeds Statistics Scoping Event

Although this particular event was not organised specifically for the benefit of this research it did serve to help identify local sources of community safety data. That main aim of the scoping event (19th November 2003, HOST, Chapeltown, Leeds) was to discuss ideas with LCSP colleagues for a proposed community safety annexe to the Leeds Statistics website (hereafter abbreviated to *Leeds Statistics*. The local delegates were invited on the basis that they themselves were officers responsible for producing and maintaining electronic data sets within their department or organisation.

During two breakout sessions, delegates were organised into groups and asked to identify variables, over which the group members themselves had ownership, that might be useful to other LCSP members when analysing community safety issues. The results of these discussions were then summarised and delegates were invited to suggest which variables should be pursued as a priority (Table 4.2).

In contrast to the NCSN Workshop, the Leeds Statistics event showed a greater concern with data which described how and where crime and disorder problems manifested themselves. Far less attention was paid to social and environmental conditions that might encourage crime, although to an extent this was possibly a reflection of the crime mapping tenor of the presentation. The NCSN work-

Health	Social Capital
Addiction treatment places	Access to community groups
Methadone prescriptions	Membership of community groups
Victims of domestic violence	Youth inclusion in community groups
Access to addiction services	Access to community forums
Addiction treatment outcomes	Participation in elections
Mental health problems	Local newspaper readership
Life expectancy	Neighbourliness
Teenage pregnancy	Satisfaction with area
	Size of local social network
Built Environment	**Education**
Void properties	Truancy levels
House type	Exclusions
Pedestrian/vehicle flow	Academic achievement
Natural boundaries	Ethnic diversity
Footpaths	Religious diversity
Development sites	Class sizes
Car parks	Sport facilities
Public transport networks	After school provision
Road traffic accidents	Free school meals
Street lighting	Pupil days lost to sickness
Location of CCTV	
Alleys and ginnels	
Defensible space	
Social	**Demographic**
Levels of debt/arrears	Lone parenthood
'Sense of community'	Single person households
Children 'at risk'	Young person households
Overcrowding	Elderly person households
Living conditions	
Housing tenure	
Economic	**Land Use**
Long-term unemployment	Open spaces, parkland
Youth unemployment	Public amenities
Average income	Retail areas
Means tested benefits	Private land
Number of small businesses	Refuse facilities
Employment type	Abandoned vehicles
Job vacancies	Allotments
Job creation schemes	Licensed premises
Inward investment	Leisure facilities
Insurance premiums	Entertainment Licenses
Regeneration investment	

Table 4.1: Variables identified at NCSN Conference, Belfast, 2003.

Variable
Police recorded crime
ASB data (police and ASBU incidents)
Youth offender data (YOT)
Education data (exclusions, truancy)
Fire data (arson, hoaxes)
Needle collections
Abandoned vehicle recoveries
CJIP (drugs treatment interventions)
Noise complaints
Social services ('at risk' populations)

Table 4.2: Variables identified at Leeds Statistics Scoping Event, 2003.

shop took a different starting point (in the CDRP Families classification), encouraging suggestions more in line with the idea of comparing the nature of places.

4.2.3 Literature Review

In addition to consulting with community safety practitioners, a review was made of documented community safety studies to see which variables investigators used most often in their models and explanations of patterns of crime and disorder. These studies were predominantly from the UK and the USA and the variables identified therein have many similarities with those identified in Table 4.1 and Table 4.2. A summary of the variables from the literature is shown in Table 4.3 using the same categories as previously.

Health Variables	Study
Low birth weight	(Chow, 1998)
Teenage pregnency	"
Infant mortality	"
Long-term limiting illness	(Berry and Jones, 1995; Fahmy et al., 2002)
Social Capital	
Collective efficacy	(Sampson and Raudenbush, 1999)
Levels of trust	(Kawachi et al., 1999)
Built Environment	
Presence of schools	(LaGrange, 1999)
Presence of malls	"
Void properties	(Taylor, 1993; Berry and Jones, 1995; Harries, 1997; Clark and Lab, 2000; Schumacher and Leitner, 2002)
Graffiti	(Taylor, 1993; Clark and Lab, 2000)
Litter	"
Bars and adult entertainment	(Hirschfield and Bowers, 1997a; Clark and Lab, 2000)
Poor area conditions	(Clark and Lab, 2000)
Poor housing condition	"
Window grills	"
Houses needing major repair	"
Road networks	(Harper et al., 2002)

Housing type	"
Motorway junctions	"
Education	
High school graduates	(Harries, 1997; Kawachi et al., 1999; Clark and Lab, 2000; Mosher, 2001)
Exclusions	(Welsh et al., 2000)
Attendance	"
Student turnover	"
College education	(Taylor, 1993)
High school dropouts	(Mosher, 2001)
No post-school qualifications	(Fahmy et al., 2002)
Social	
Housing tenure	(Williams, 1985; Kawachi et al., 1999; LaGrange, 1999; Harper et al., 2002; Schumacher and Leitner, 2002)
Homemakers	(LaGrange, 1999)
Social class	(Williams, 1985; Taylor, 1993; Mosher, 2001)
Overcrowding	(Harries, 1997; Welsh et al., 2000; Harper et al., 2002)
Small groups on street	(Taylor, 1993)
Males on street	"
No central heating	(Fahmy et al., 2002)
Segregation	(Hirschfield and Bowers, 1997a)
Social heterogeneity	(Hirschfield and Bowers, 1997b)
Super Profiles lifestyle	"
Index of Living Conditions	"
Townsend material deprivation	(Craglia et al., 2001)
Shared housing	(Hirschfield and Bowers, 1997a; Hirschfield et al., 2001)
Demographic	
Residential turnover	(Williams, 1985; Harries, 1997; Hirschfield and Bowers, 1997a; LaGrange, 1999; Sampson and Raudenbush, 1999; Welsh et al., 2000; Hirschfield et al., 2001; Harper et al., 2002)
Age	(Taylor, 1993; LaGrange, 1999; Mosher, 2001; Harper et al., 2002; MacDonald, 2002)
Population density	(Taylor, 1993; Hirschfield and Bowers, 1997a; LaGrange, 1999; Sampson and Raudenbush, 1999; Mosher, 2001; Harper et al., 2002; MacDonald, 2002)
Ethnicity	(Taylor, 1993; Harries, 1997; Hirschfield and Bowers, 1997b,a; Sampson and Raudenbush, 1999; Welsh et al., 2000; Harper et al., 2002)
Lone parenthood	(Harries, 1997; Hirschfield and Bowers, 1997a; Kawachi et al., 1999; Welsh et al., 2000; Mosher, 2001; Harper et al., 2002; MacDonald, 2002)
Divorced males	(MacDonald, 2002)
Day-time population	(Hirschfield and Bowers, 1997a; Harper et al., 2002)
Student households	(Harper et al., 2002)

Single households	"
Economic	
Households receiving benefits	(Harries, 1997; Chow, 1998; Clark and Lab, 2000; Harper et al., 2002)
Unemployment	(Williams, 1985; Berry and Jones, 1995; Harries, 1997; Hirschfield and Bowers, 1997b; Kawachi et al., 1999; La-Grange, 1999; Clark and Lab, 2000; Hirschfield et al., 2001; Mosher, 2001; Fahmy et al., 2002; Harper et al., 2002)
No car	(Berry and Jones, 1995; Hirschfield et al., 2001; Fahmy et al., 2002)
In poverty	(Taylor, 1993; Harries, 1997; Hirschfield and Bowers, 1997a; Kawachi et al., 1999; Mosher, 2001; Fahmy et al., 2002; Schumacher and Leitner, 2002)
Home value	(Taylor, 1993; Harries, 1997; Clark and Lab, 2000)
Rent	(Harries, 1997; Clark and Lab, 2000; Schumacher and Leitner, 2002)
Income	(Harries, 1997; Kawachi et al., 1999; Clark and Lab, 2000; Welsh et al., 2000; Mosher, 2001; MacDonald, 2002)
Income equality	(Kawachi et al., 1999)
Child poverty	(Fahmy et al., 2002)
Land Use	
Housing age	(Harries, 1997)
Housing density	(Taylor, 1993)
Abandoned cars	(Clark and Lab, 2000)
Homeless shelters	(Mosher, 2001)
Mixed land use	(Sampson and Raudenbush, 1999)
Residential land	(Hirschfield and Bowers, 1997a)
Offending/Policing	
Delinquency	(Chow, 1998)
Target hardened properties	(Berry and Jones, 1995)
Crime prevention activity	"
Police strength	(Mosher, 2001)
Police incidents	(Hirschfield and Bowers, 1997b)

Table 4.3: Survey of neighbourhood variables used in community safety studies

4.2.4 Summary of Survey Responses

There was quite a high level of agreement between the variables identified in the literature review and those suggested by participants at the NCSN workshop. There was a large bias towards quantifiable measures, especially those collated and reported upon by the census and government performance indicators. Qualitative measures were relatively few, reflecting, perhaps, the difficulty in obtaining them in general, and in large enough samples to reliably disaggregate to small-area level in particular.

The results of the Leeds Statistics scoping event were rather different in nature and reflected more of an interest in identifying hotspot areas according the criminal and antisocial behaviour experienced therein. This was probably in part caused by the presentation of project as being supplemental to an existing collection of general neighbourhood data - similar to that identified at the NCSN event and in the literature review. Yet it is also interesting to observe that practitioners were aware of the potential importance of the data they generated to others delivering services in neighbourhoods suffering from a range of problems.

On the basis of the surveys, the decision was taken to pursue data within seven broad categories; demography, economy, health, land use, minor incivilities (antisocial behaviour), social, and community cohesion. Education issues are divided between economy and minor incivilities (exclusions), while issues pertaining to the built environment are covered by land use.

4.3 A New Neighbourhood Geography for Leeds

At the same time as data started to be collected, consideration was also given to the geography to be used for classification. This decision could have been left until all the data were gathered, but data collection actually took a long time (due, in part, to delays with census publication) so a geography was identified early in order to allow preliminary experimentation with cluster analysis techniques.

From the outset of the research, it had been assumed that the geography for the new classification would be sub-ward level. Coterminosity with census output areas was also a desired trait, as this would make calculating data values for census-based variables more straightforward. At the time decisions needed to be taken, the ONS super output area geographies were still at the proposals stage, so although the lower layer of these looked as if it might be suitable, it was decided to build a new neighbourhood geography for Leeds, from census output areas (OAs).

4.3.1 Re-zoning Leeds

Whether any fixed-boundary definition of a local neighbourhood geography could be relied upon would be open to question (Canter, 1977; Moore, 1979). For example, Lee's (1968) work in Cambridge in the 1960's showed that even when neighbours from the same street were asked to define the boundary of their neighbourhood there was a great deal of variation in the responses. Children may also have distinct senses of distance and neighbourhood (Matthews, 1981, 1984b) and studies have shown gender differences to spatial perceptions in childhood (Matthews, 1984a) and adulthood (Everitt and Cadwallader, 1972). Perceptions of neighbourhood may be different for everyone, as they are shaped not only by physical features but also the spatial extents of peoples' social networks and routine activities.

Nevertheless, it transpired that staff at Leeds City Council (LCC) had undertaken a neighbourhood mapping exercise, although the results had been contested to the the extent that the results were not likely to be used by the council. The exercise had involved Leeds residents and council workers who were asked to help identify place-based communities as they were perceived by the people living in them and working with them. One hundred such *communities* were identified. It transpired that boundaries were frequently defined by prominent physical features, such as transport networks and other landscape features, as well as the historical consequences of the planning and development of residential areas in Leeds over the last couple of centuries.

The principal problem with the LCC community boundaries is that they are not coterminous with any other small-area administrative boundaries. To assign 2001 Census data to the new geography thus requires that some sort of fitting or apportionment exercise be undertaken with the community and OA boundaries. Areal apportionment of census data where community boundaries intersect OAs was rejected because of ecological fallacy problems with some of the large OA covering rural areas. Instead, it was decided to flex the community boundaries to conform to the OA geography. This would result in some loss of integrity of the original user defined community geography but no loss of census information. The fitting process involved three steps:

1. Using standard GIS intersection routines, each OA was assigned to a community if it fell wholly within a community boundary, or if over 50% of its area fell within the community boundary. New community boundaries were then drawn by grouping and merging the split OA polygons by community designation.

2. Each redrawn community boundary was then manually assessed and flexed according to two criteria. Firstly, some OAs were reassigned where community boundaries clashed awkwardly with natural boundary features such as major roads, rivers and railways. Secondly, some OAs were reassigned to improve the 'smoothness' of some very irregular new community boundaries. The two Horsforth communities was one example where this second form of flexing was exercised.

3. New community boundaries were then assessed for their size and the logic of their coverage. Under greatest scrutiny were large rural communities. This exercise was largely undertaken to reduce the problem of large areas on shaded maps being dominated by a relatively small proportion of the population. A number of the new communities were sub-divided. Decisions about where to make the divisions were based on topographical and communication features as well as coterminosity with parish boundaries. The number of communities rose from 100 to 106 (Figure A.1).

For the most part, the names assigned to communities by the original LCC work were maintained. Where some of the larger (in area terms) communities were split, new names were assigned according to the logic of the underlying settlement names. There were no major difficulties or contentious issues when deciding what these names should be.

There was one technical problem that could not be resolved by any of the stages described above. This involved a split OA in the region of Golden Acre Park (one part) and Leeds Bradford Airport (the second part). During the development of the 2001 Census OA geography it had been hoped that it would not be necessary to split OA polygons. However, other design rules had to be allowed to take precedence over the split polygon rule and this is what happened in this instance (personal communication with ONS: May, 2003). To expand, the rules over ensuring minimum OA population size; ensuring OAs could be aggregated to parish boundaries; and ensuring OAs did not straddle ward boundaries, won the day over the split OA polygon rule. As a result, OA 00DAFY0013 was assigned to Yeadon community, even though one part of the polygon is actually situated in Bramhope community. Both part-polygons are sparsely populated.

4.3.2 Sub-Dividing Community Areas

Although they proved useful for other purposes, the 106 community areas were considered to be too large to represent neighbourhoods. In part this was intuitive and influenced by a subjective reading of the geography given local knowledge of contrasting parts of Leeds. In addition, a figure of 1,500 people was being proposed by ONS for the Lower Layer SOA geography (which was being discussed as a being a neighbourhood geography) and this population was not too distant from a likely size for a neighbourhood suggested by Lee (1968). Thus, work was undertaken to subdivide the community areas into what have been termed Sub-Community Areas (SCAs). This subdivision was accomplished by using the ZDES3b software to aggregate the OAs within a community area to produce zones of approximately 1,500 people. The ZDES rezoning software is described in (Alvanides et al., 2002) and based upon earlier work by Openshaw (1977; 1978) and Alvanides (2000).

Once ZDES re-zoning had been completed, the SCA geography was checked manually for zones which were either very irregularly shaped or were straddling significant physical landscape features. The problem of very irregular shapes is made more difficult by the nature of the OA building blocks. These are defined by postal geography and thus have a tendency to follow street layouts. No attempt was made to individually name any of the final 479 SCAs although their identifiers were designed to reflect their parent community areas. Maps of the entire SCA geography occur throughout the remainder of the thesis, but for the area of Chapeltown, Figure 4.1 shows the transformation from the (a) community geography developed by LCC, the (b) flexing of the community boundary to the census OA geography, and (c) the sub-division into Sub-Community Areas. A map of the completed SCA geography is provided in Appendix A (Figure A.2).

4.4 Calculating Data Values for SCAs

Data values for each SCA were calculated for different variables using a variety of different methods.

4.4.1 Aggregation of Output Areas

A lookup table of census OAs to SCA was generated once the SCA geography had been finalised. A population size for each SCA was calculated by summing together the populations of each OA within a SCA. The same procedure was used to count households. Both total resident population and total households were used as denominators with which to standardise the census variables selected for the classification. Most of the census variables selected for the classification are treated in the same way. That is, individual OA counts are summed for each SCA, and are then divided by an SCA denominator count (usually either population or households) to produce a proportion.

The drawback of this approach is that it compounds the differences between the real variable values for each OA and the figures published in the census. Such differences are deliberately introduced into the census to limit the chances of disclosing information about individual persons or households. The effect of small cell adjustment, for example, has led to an estimation that wherever there are real counts of 0,1,2 or 3 the published census count will be 0 or 3 (Stillwell et al., 2005). The advice given by ONS is to limit the aggregation of OAs to larger geographies where at all possible. For the purposes of this research, however, there was, at the time data was being collected and prepared, no alternative.

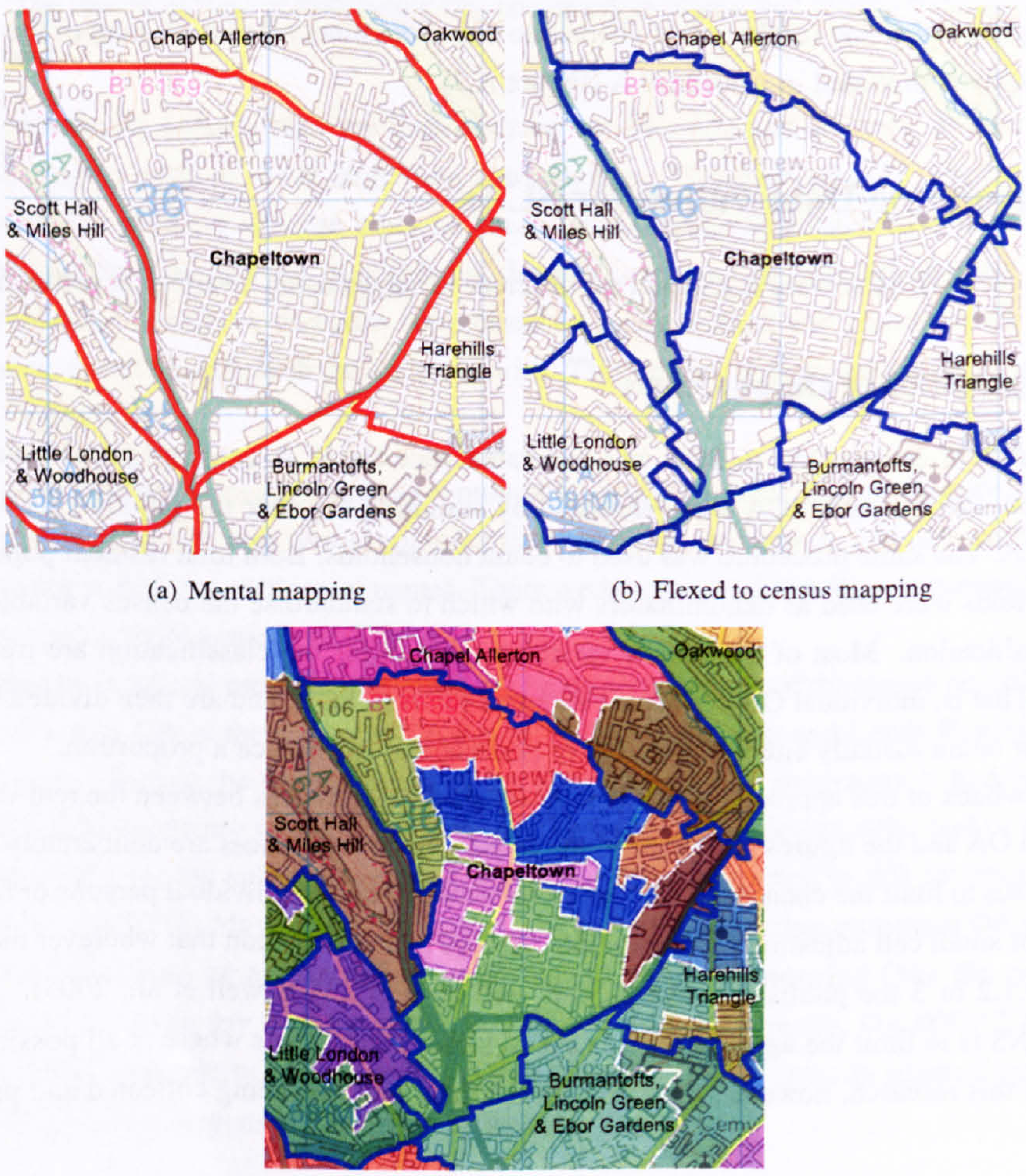
4.4.2 Aggregation of Point-based Features

Some of the variables selected for the classification reflect the numbers of some non-census artifact within an SCA, for example, the number of pupils achieving certain GCSE grades. In this and similar cases, individual counts have been ascribed to a point location reflecting the home address of the subject. Thus, data values for each SCA can be generated by counting the address locations that fall within each SCA boundary, using GIS point-in-polygon query routines.

Some of the data of this type was supplied already geocoded but in some cases geocoding had to be conducted using the Ordnance Survey Address Point database and Leeds City Council gazetteers.

4.4.3 Aggregation of Area-based Features

A variation on the aggregation of point-based features was the calculation of data values based upon the area they covered within a SCA, using the total area of the SCA as the denominator. The proportion of an SCA covered in natural land is one example. Again, a GIS was used to execute a spatial query to sum the individual areas within each SCA. However, it was first necessary to split the land parcels (if we continue with the natural land example) using the SCA boundaries. This ensured that large land parcels that spanned two or more SCAs would be correctly arealy apportioned among each of the intersecting SCAs.



(c) Sub-divided into SCAs

Figure 4.1: Fitting the mental map of Chapeltown to the 2001 census output area geography

From a computational perspective, the power of a standard PC was sufficient for processing the polygons relating to one, fairly large local authority. If data pertaining to a whole country or nation were to be processed, however, it is possible that extra computing power would be required.

4.4.4 Geographical Harmonisation

At the time data were being collected, some variables were only available at a postal sector level. Two approaches to harmonising this geography to the SCA geography were tested, area apportionment and the disaggregation to household level followed by a re-aggregation to SCA level.

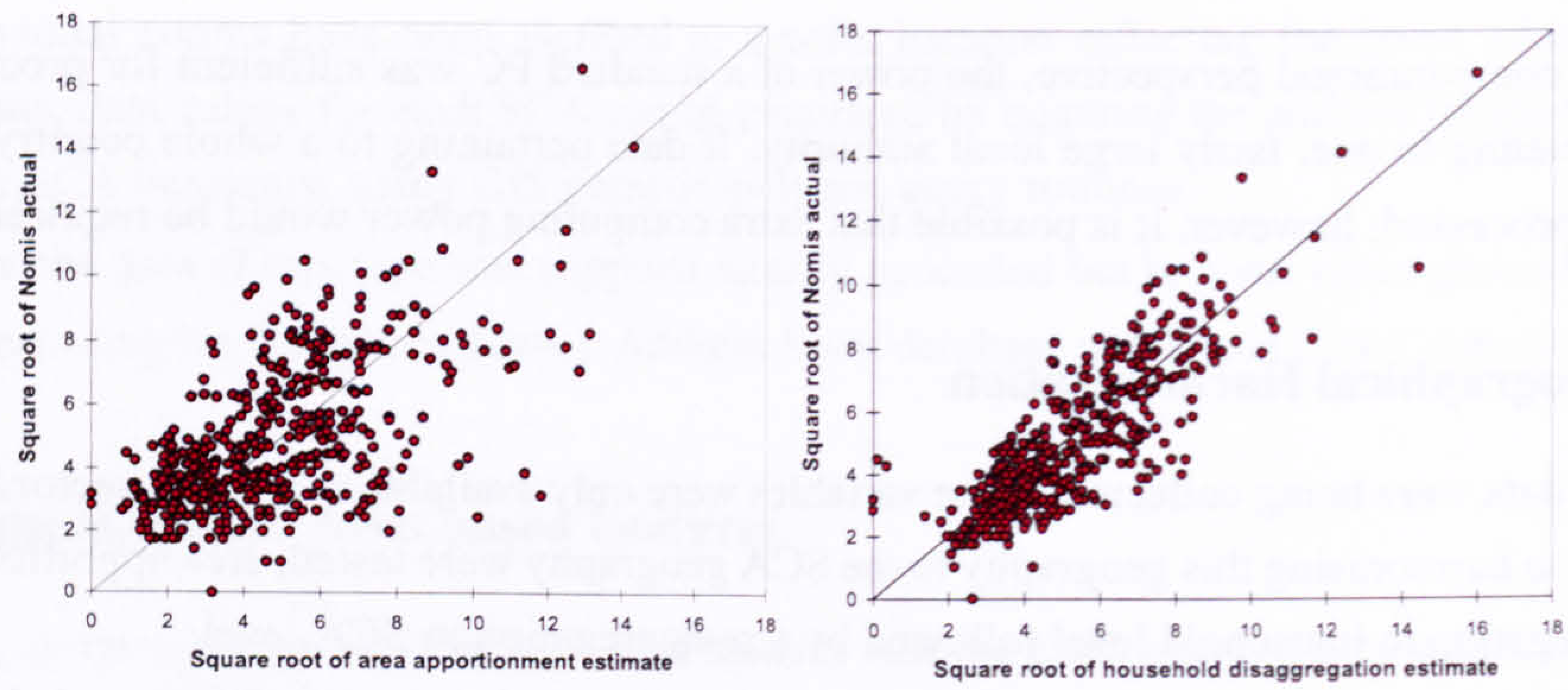
Area apportionment is conducted on the assumption that population (or whatever is being measured) will be evenly spread across each SCA. Using GIS, the postal geography is first split using the SCA boundaries. The data value for a postal sector is then shared out between the resulting fragments of the postal sector according the area of the fragment (as a proportion of the whole postal sector area). With the data thus arealy apportioned, the fragments can be grouped into whole SCAs, and the data value fragments summed. Although areal apportionment works as a harmonisation technique, the initial assumption would be unlikely to be true in real settlement patterns for geographies of this scale. Because of the completeness of coverage of the postal sector geography, some sectors (especially rural ones) will have large areas where there are no households.

The alternative of disaggregating to individual households is possibly less common, although it has been reported elsewhere (e.g. Simpson, 2002; Norman et al., 2003). The assumption this time, if measurements are for individual persons, is that each household will contain the same number of persons. Household location are plotted using data from a gazetteer such as the Ordnance Survey Address Point database, and fractions of the postal sector data value are then shared out equally among all the households within each postal sector. For example, if there are 127 benefit claimants in a postal sector of 1,000 households, then each household is assumed to have 0.127 of a claimant each. A GIS point in polygon query can then be executed and the sum of the data values for each household in a SCA calculated. The OS Address Point data are not ideal for this exercise, as discriminating between residential and non-residential addresses requires an amount of manual sifting and checking to produce reliable results.

To demonstrate the effect of the two apportionment methods a postal sector claimant count (available from Nomis) was disaggregated and then re-aggregated to the SOA lower layer geography, for Leeds. (This test is only possible because subsequent to initial data collection, Nomis have begun to publish claimant counts for the SOA lower layer geography). Using area apportionment, a Pearson Product-Moment correlation coefficient between the area apportioned estimates and the published SOA Lower Level data is 0.609 (P-value=0.000). Using the OS Address Point household data to harmonise the geographies, the Pearson Product-Moment correlation coefficient between the estimates and the published SOA Lower Level data is considerably stronger, at 0.847 (P-value=0.000). The difference in the fits can also be visualised using scatter plots (Figure 4.2), with the data in this instance transformed to lessen the impact of outlier values.

4.4.5 Surface Modelling

Some experimentation has been undertaken with the creation of surface models where the effects of neighbourhood features might have a more diffuse effect away from their immediate location. In terms of the kernel density estimation technique used, this follows previous research into the use of



(a) Area apportionment (b) Disaggregation to households

Figure 4.2: Super Output Area Lower Level claimant count estimates based on disaggregation of postal sector-level counts.

surface models to estimate populations (Martin et al., 2000; Martin, 2002) and identify neighbourhoods (Martin, 1998) and urban population densities (Harris and Chen, 2003).

Using public drinking as an example, a kernel density surface was created to reflect the spatial clustering of public houses in Leeds. When mapped (Figure 4.3), the high density locations reveal where public houses are more clustered and, therefore, where more offences aggravated by the effects of alcohol might be expected. Where the surface density is low, fewer offences might be expected. The important feature of the surface model is that the bandwidth and output cell size parameters used in its construction can be manipulated to create a 'smooth' surface, where the change in density values away from a high density location fall gradually, simulating a distance decay effect - with problems becoming less likely as the distance from the problem source increase. This hypothesis was tested, but in principle it could by using the detailed *modus operandi* text that is recorded along a crime, over a long enough period to generate a reliable sample.

To produce a single data value for each SCA, GIS grid functions were used to calculate the average value of the z-dimension in the area of the surface intersecting each SCA. The resulting distribution of pub density, standardised by the SCA populations is shown in Figure 4.4.

4.5 National Datasets

Typically, large, national datasets in the public domain have been the primary source of variables for geodemographic classifications. The decennial census of population has been used extensively to create area classifications (Rees et al., 2002a), due in part to its complete spatial coverage, comprehensive range of variables and tabulations and choice of geographical units. The use of detailed Ordnance Survey data in geodemographic classifications is much less common although the recent development of the National Land Use Database (NLUD) (ODPM, 2005) has highlighted certain possibilities. The desire to use detailed cartographic data for geodemographic classification has been expressed (Dugmore, 2004) but the costs of procuring the entire dataset and the computational resources to summarise it by different administrative geographies have been prohibitive. Other national datasets are available which deal specifically with economic or employment issues (e.g. Nomis, Annual Business Inquiry), although the exploitation of these may be restricted to government and research use

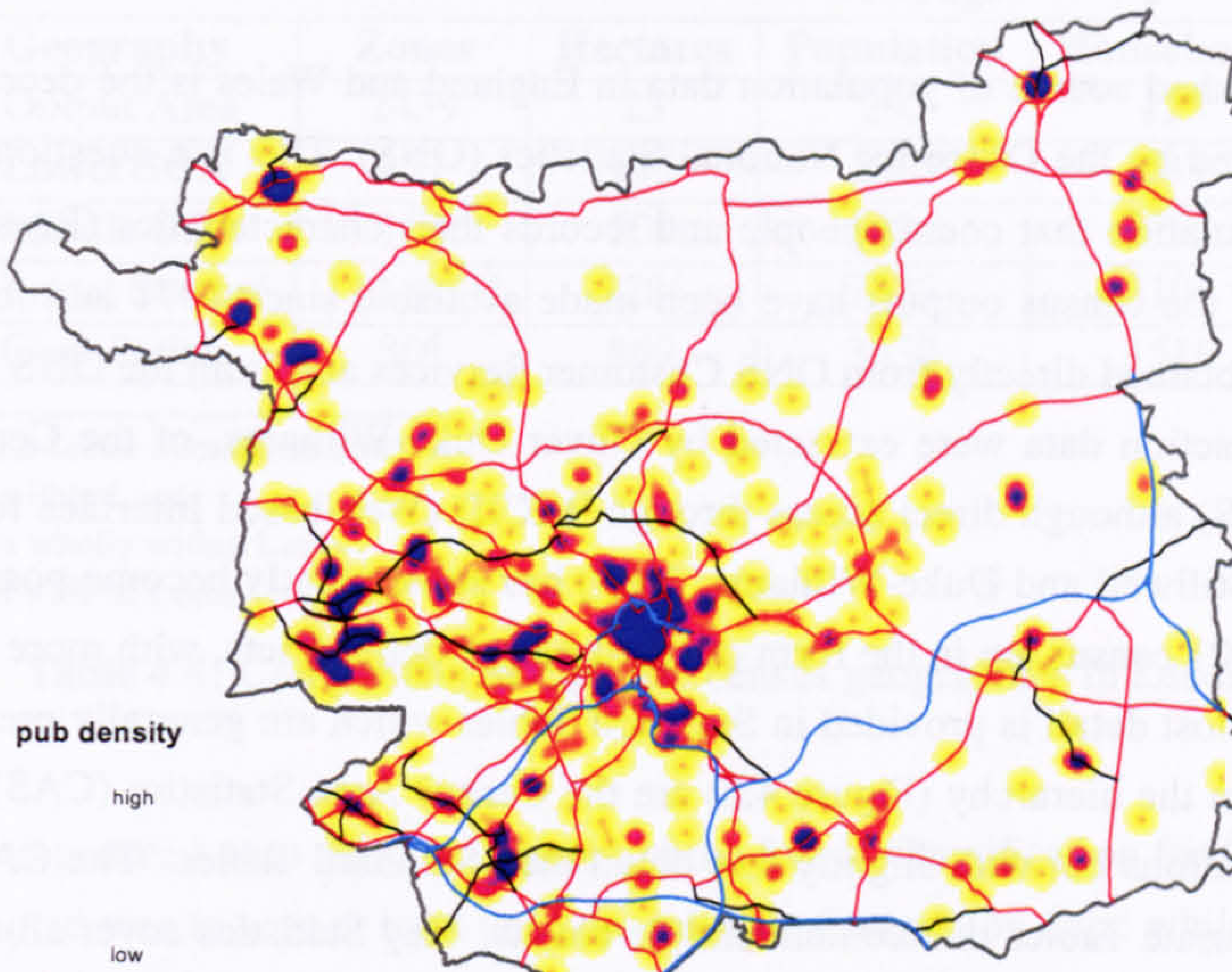


Figure 4.3: Surface of public house density created with kernel density estimation (bandwidth=500m).

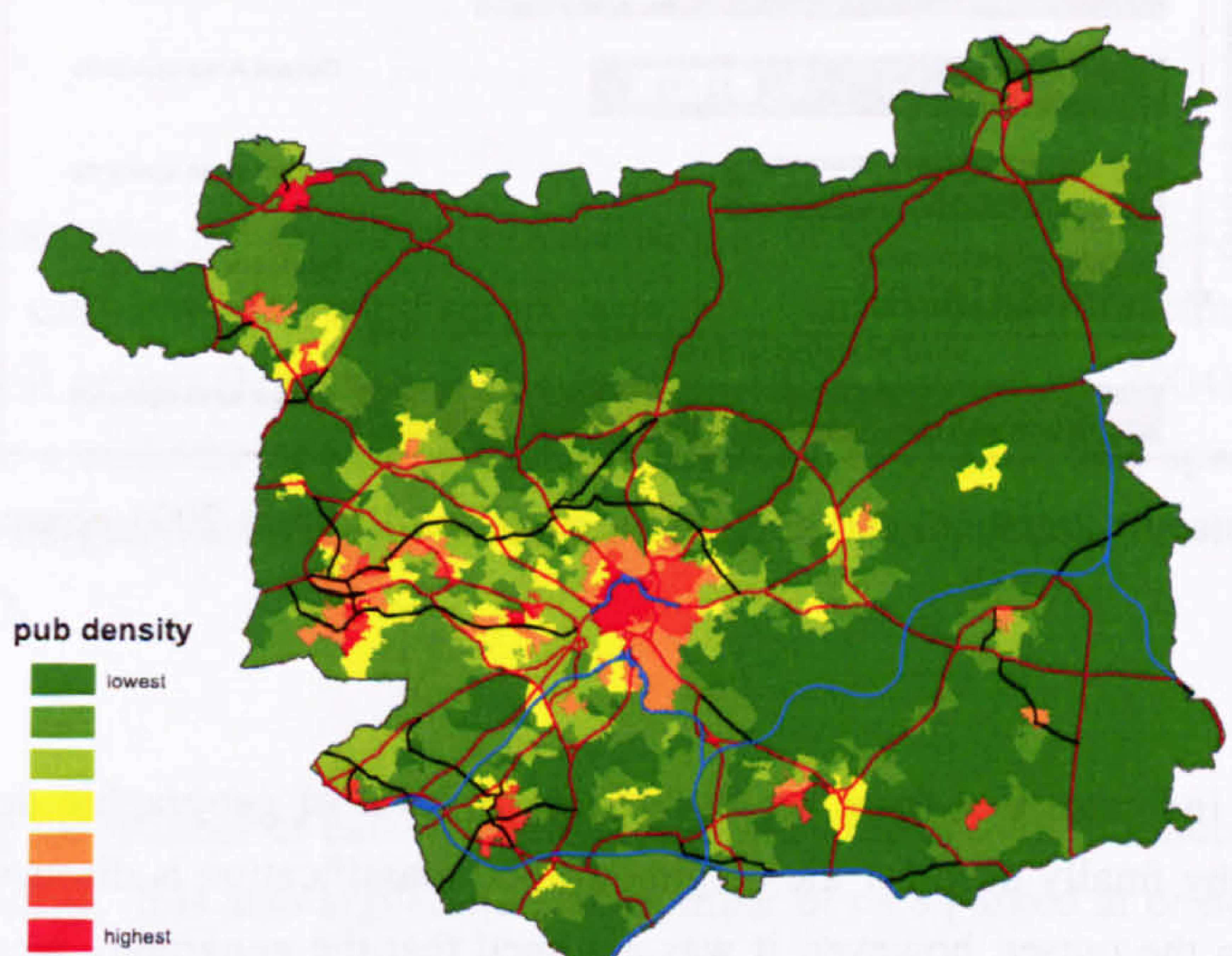


Figure 4.4: Pub density in sub-community areas standardised by population size.

only. Details of the national datasets used in the neighbourhood classification for community safety are discussed below.

4.5.1 Census of Population

The principal published source of population data in England and Wales is the decennial census of population conducted by the Office for National Statistics (ONS). This is a questionnaire survey of the whole UK population that counts people and records their characteristics (Rees et al., 2002b). Digital versions of the census outputs have been made available since 1971 and for this research, census data were obtained directly from ONS Customer Services and from the ONS National Statistics website. Interaction data were extracted by Oliver Duke-Williams, of the Census Interaction Data Service (CIDS) although direct access through WICID (Web-based Interface for Census Interaction Data) (see Stillwell and Duke-Williams, 2003) has subsequently become possible. The main results from the 2001 census are in the form of standard pre-defined sets, with more detail for larger populations. The most detail is provided in Standard Tables which are generally presented as cross-tabulations. Next in the hierarchy (Figure 4.5) are the Census Area Statistics (CAS) which are also mainly cross-tabulations but with slightly less detail than Standard Tables. The CAS also contains a number of Univariate Tables that contain simple counts. Key Statistics cover all the main census topics and are presented in an easily readable style for more general-purpose consumption. Specialist outputs include a postcode headcount of population and origin-destination matrices for migration and commutes between homes and workplaces.

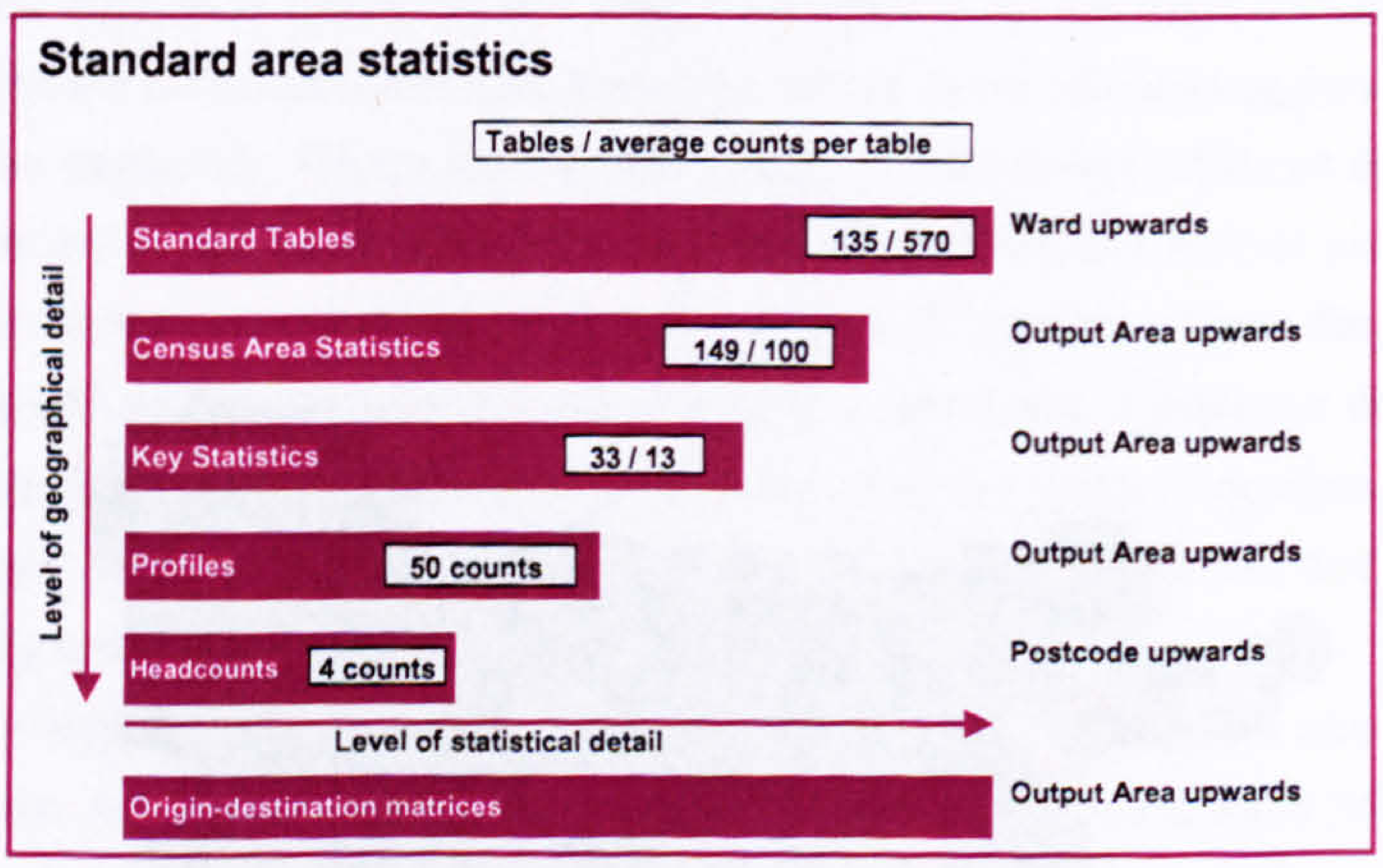


Figure 4.5: Difference in statistical and geographic detail of different 2001 census outputs. Source: ONS.

Figure 4.5 also indicates that there are differences in level of geographic detail between table types. The geography finally used for the neighbourhood classification is discussed in detail below (Section 4.3). From the outset, however, it was assumed that the geography would be at sub-ward level, meaning the census output geographies likely to be of most interest would be output areas, super output areas (lower and middle levels), postal sectors and civil parishes. Table 4.4 summarises the characteristics of these geographies in area, population and household terms for the Leeds Local Authority District.

Output Schedule

The details of exactly when 2001 census products were released is of little importance but the fact that there were significant delays in the release of some data did have a bearing on the time at which it was

			Average	
Geography	Zones	Hectares	Population	Households
Output Area	2439	23	293	123
Lower SOA	477	115	1503	633
Middle SOA	108	509	6624	2792
Postal Sector[a]	110 (82)[b]	328[c]	7323	3103
Civil Parish	30[d]	869	3632	1519

[a]using ONS postal sector boundaries 2003
[b]sectors wholly within Leeds
[c]based on sectors wholly within Leeds
[d]these only cover 47% of Leeds

Table 4.4: Comparison of different census geographies in Leeds.

thought prudent to proceed with the development of the Leeds Classification for Community Safety. The strategy adopted was to wait at least until the release of migration data, whilst the development of all three of the classifications (local authority, ward and OA levels) produced by (or for) ONS proceeded before any migration or workplace data became available, for example. The new local authority classification by Vickers *et al* (2003) is similarly affected.

4.5.2 Selection of Census Variables

A number of variables were selected from the 2001 census (Table 4.5) for possible inclusion in the LCCS. The selections reflect levels of agreement amongst the three surveys and other issues discussed below.

Age

In summary, age variables are included because people of certain age groups, most often those in their late teens and early twenties, are disproportionately likely to become victims of certain types of crime and to commit crimes (Hirschi and Gottfredson, 1983; Dodd et al., 2004). Age has also been demonstrated to be a correlate of fear of crime (Walklate, 1997) and some specific types of crime, such as distraction burglary, are very much targeted at certain age groups - the elderly in this instance (Lister et al., 2004).

Cars

Variables capturing the number of cars in a neighbourhood were chosen to reflect the risk of vehicle-related crime occurring. It is also argued that the number of cars parked at or near the workplace is a better indicator of risk during working hours, and these figures were estimated using the commute modality data included in the journey to work tables.

Ethnicity

Ethnicity variables are important when considering race-related victimisation and hate crime. Race and religion were also central to the report of the Community Cohesion Review Team, chaired by Ted Cantle, which investigated the consequences for national policy of the disturbances in Oldham, Burnley and Bradford during the summer, 2001. Yet, while Cantle argued for greater mixing, especially in schools, there are counter-arguments in social disorganisation theory around problems of compet-

Category	Id	Variable	Table
Demography	AG01	Population aged 0-4	
Demography	AG02	Population aged 5-14	
Demography	AG03	Population aged 15-24	
Demography	AG04	Population aged 25-44	
Demography	AG05	Population aged 45-64	
Demography	AG06	Population aged 65-84	
Demography	AG07	Population aged 85+	
Demography	ET01	Ethnic minorities: not white British or Irish	
Demography	ET02	Ethnic heterogeneity	
Demography	MG01	People who moved to area from outside UK	
Demography	PD01	Daytime to residential population ratio	
Economic	CA01	Cars at or near the workplace	
Economic	CA02	Cars at their residential address	
Economic	TE01	Households owned	
Economic	TE02	Households social rented	
Economic	TE03	Households private rented	
Economic	QU01	Population lowly qualified: no or 1	
Economic	QU02	Population highly qualified: 4/5	
Community cohesion	CC01	Proportion of working residents who work from, or less than 2km from, home	
Community cohesion	MG02	People who lived at same address one year prior to census day	
Land use	HD01	Household density	
Land use	HT01	Houses that are detached	
Land use	HT02	Houses that are semi-detached	
Land use	HT03	Houses that are terraced	
Land use	HT04	Houses that are flats: block	
Land use	HT05	Houses that are flats: shared house	
Land use	HT06	House type heterogeneity	
Land use	VO01	Households unoccupied	
Social	HH01	One person households	
Social	HH02	Lone parent households	
Social	SC01	People who are NS-SeC 1	
Social	SC02	People who are NS-SeC 2	
Social	SC03	People who are NS-SeC 3	
Social	SC04	People who are NS-SeC 4	
Social	SC05	People who are NS-SeC 5	
Social	SC06	People who are NS-SeC 6	
Social	SC07	People who are NS-SeC 7	
Social	SC08	People who are NS-SeC 8	
Social	SC09	People who are full-time students	

Table 4.5: Classification variables selected, or derived, from the 2001 census.

ing norms and values (Shaw and McKay, 1942), suggesting that neighbourhoods with a mixture of ethnicities may suffer tensions.

As well as selecting a non-white British ethnicity variable, a modelled variable, ET02, was constructed to reflect ethnic heterogeneity. The method used to calculate this was originally developed by Blau (1977) for calculating species diversity in biological systems. The Index of Heterogeneity has been used elsewhere for the profiling of neighbourhoods for crime pattern analysis (Hirschfield and Bowers, 1997a) and the same measures (by different names) have been used in studies of population (Brewer and Suchan, 2001) and ethnicity (Rees and Butt, 2003). The formula, following Hirschfield (1997a), is defined thus :

$$I = 1 - \sum Pi^2 \tag{4.1}$$

where Pi is the proportion of the population in group i. High values of the index, I, indicate a diverse ethnic mix while low values define an ethnically homogeneous area. Values of the index range from 0 to 1. The advantage of this approach over other more traditional measures of segregation (for a comprehensive review see Massey and Denton, 1988), for example, is that it allows consideration to be given to multiple ethnic groups, and not just bivariate comparisons such as white/non-white.

Five different ethnic groupings were used, drawing on the eleven different ethnic classes used in the 2001 census, details of which are summarised in Table 4.6. A additional variable, people who have moved to the area from outside the UK (MG01), was chosen to reflect the increased risk that first generation immigrants may face from race hatred and misunderstanding.

Grouping	Ethnicity
White	British
	Irish
White Other	Other White
Black	Caribbean
	African
	Other Black
	Mixed White and Caribbean
	Mixed White and African
Asian	Indian
	Pakistani
	Bangladeshi
	Other Asian
	Mixed White and Asian
Chinese and Others	Chinese
	Other Ethnic Group
	Other Mixed

Table 4.6: Variables used in construction of Index of Ethnic Heterogeneity.

Household Structure

Two variables reflecting aspects of household structure were selected. Lone parenthood, when combined with other sources of social and economic strain, may lead to single parents finding themselves with little option but to live in cheaper, often more deprived neighbourhoods. Here, the chances of becoming a victim of crime are higher and young people run an increased risk of being drawn into delinquent youth cultures through association with delinquent friends (Lipsey and Derzon, 1998) and increased likelihood of having to attend high delinquency rate schools (Graham, 1988). Research into the links between disrupted families and delinquency has a long history, but although separation and divorce can have disruptive consequences for children, the chances of this leading to an increase in offending have been shown to be similar to those for children living in intact high conflict families (Juby and Farrington, 2002). Single households may simply be at greater risk to crimes such as burglary due to the relative lack of guardianship compared to couple or family households (Tseloni et al., 2002).

Housing Type

Housing type has been demonstrated to be a factor in target selection by burglars, with first preference being for detached properties, followed by semi-detached, and then terraced housing (Bennett and Wright, 1984). However, depending on the neighbourhood, it could be argued that housing type may also be an indicator of wealth and high value goods more desirous of some types of burglar, although there are also affluent terraces and near-derelict detached houses to challenge these generalisations. High-rise living may be very secure where shared entrances are well maintained and/or supervised, but if overcome, burglars may be able to force entry into individual flats with little risk of being disturbed.

In response to the 'near repeat' hypothesis (Morgan, 2000), a modeled variable, HT06, was created to reflect homogeneity of housing type (e.g. not all terraced, or all semi's). The near repeat hypothesis argues that "areas with homogeneous housing are likely to experience higher rates of victim prevalence because offenders are not greatly advantaged by repeatedly breaking into the same house" (Townsley et al., 2003, page 618). The index was created using the same method employed for the ethnic heterogeneity index. Following the 'near repeat' hypothesis, more homogeneous neighbourhoods (lower index scores) may be at greater risk of near repeats because house-entering skills are likely to be more transferable to nearby properties. Conversely, neighbourhoods with a greater mixture of housing types (higher index scores) may repel those burglars for whom familiarity with property types is an important factor in target selection. To complement the homogeneity of housing type variable, a similar index was created to reflect homogeneity of house size (according to number of rooms). In combination, it is argued that these two indices ought to be better able to discriminate those neighbourhoods where housing type is fairly homogeneous but differences in style and design still exist.

Land Use

An additional modeled variable (PD01) which combines household populations and workplace populations was created to reflect mixed land use within neighbourhoods. A number of studies have demonstrated strong associations between mixed land-use, that is a mixture of residential and non-residential land uses within a neighbourhood, and social (and physical) disorder and crime (Sampson and Raudenbush, 1999; Wilcox et al., 2004). A number of propositions have also been made regarding mixed-use neighbourhoods and increased opportunities for offending and deviant behaviour (Stark, 1987). A variety of indicators of land use have been created using cartographic data (Section

4.5.3), but a proxy for mixed land use might also be undertaken using population census data alone by considering the daytime to residential population ratio, PR, calculated as,

$$PR = \frac{P_{wp.16-74} + P_{res.74+}}{P_{res} - P_{res.0-15}} \tag{4.2}$$

where $P_{wp.16-74}$ is the workplace population aged 16-74 recorded in Univariate Table UV37, $P_{res.74+}$ is the residential population aged 74 and over, P_{res} is the total residential population and $P_{res.0-15}$ is the residential population aged 0-15. Day time population comprises those people who live and work in the neighbourhood (or do not work) and those people who live outside the area and work inside the area. It excludes those people who live in the area but work outside the area. The population ratio could be refined further by utilising the Pupil Level Annual School Census (PLASC) to account for flows of pupils to schools, but this was not attempted. A more serious problem is the flow of population to retail and leisure outlets, but the decision was taken to try to take account of some of these land use issues using separate variables.

Community Cohesion

Community cohesion has been described as "the ongoing process of developing a community of shared values, shared challenges and equal opportunity ... based on a sense of trust, hope and reciprocity" (cited in Home Office, 2001b, page 69). Elsewhere, the concept has been reduced to "areas with relatively high levels of interaction between residents and a strong sense of community" (Hirschfield and Bowers, 1997b) suggesting strong similarities with the theme of social capital and the density of civic engagement in an area, based on trust between social actors, embedded in shared values and social networks (Coleman, 1988; Putnam, 1993).

Such concepts are important because it has been argued that social efficacy or social capital in a neighbourhood is one social factor that can differentiate areas which may appear similar demographically but vary in their levels of crime (Sampson et al., 1997). Thus, the more that neighbourhood residents are bonded though joint activities and face-to-face interactions, the more resistant the neighbourhood may become to predatory and violent crime as people will be prepared to intervene to prevent crime and impose shared behavioural norms.

Such assertions appear attractive and have prompted researchers to try to measure community cohesion in order to target police and community safety interventions where cohesion is least evident (Chainey, 2005). The danger in so doing, however, is that no account is taken of 'successful' suburbs which may have few of the features of community or neighbourliness (Baumgartner, 1988), or the poor neighbourhood with weak and inward looking networks which nevertheless offer strong support in adversity (Forrest and Kearns, 1999). Furthermore, within a neighbourhood, social cohesion may be high within sub-communities, for example religious groups, but if these micro-communities fail to mesh into an integrated whole then tensions may erupt, as was seen in a number of northern cities in the summer of 2001 (Home Office, 2001b).

On this evidence, the problems of measuring community cohesion may appear intractable. Attempts have been made and guides have been produced (e.g. Lochner et al., 1999), and there appears to be some level of agreement that variables such as membership of community groups, participation in local elections and readership of local newspapers are important. Unfortunately, reliable data on such issues within Leeds are either absent, partial in their coverage, or available at too coarse a spatial scale. One of the few exceptions is data regarding residential stability, captured in the census of population by variables pertaining to migration and commuting to work. Two variables were chosen for the LCCS: people who lived at the same address one year prior to census day (MG02), and the proportion of working residents who work from home or less than 2km from home. Respectively, it is argued that these variables may provide proxy measurements for residential stability and local ties to neighbourhood.

Students

A full-time students variable was selected to reflect historically high levels of victimisation to acquisitive crime experienced by this group within Leeds. Presence of large numbers of students within a neighbourhood is also likely to lead to greater social disorganisation, with student households tending to migrate annually. Resentment among longer-term residents is also a issue, and problems associated with littering, noise nuisance and alcohol-related loutish behaviour may also be more prevalent in neighbourhoods with high proportions of students. Within Leeds, for example, these problems have become serious enough for the local authority to impose restrictions on new student housing within the Headingley and Hyde Park neighbourhoods.

Student groups may also pose particular problems for the delivery of community safety policy. A number of evaluations of burglary reduction initiatives highlighted problems of maintaining effectiveness over time of target hardening and alley-gating measures in student areas - as a result of high residential turnover (Hirschfield, 2004).

Tenure

A synthesis of studies of crime on English housing estates over a period of years led Bottoms and Wiles to argue that the operation of the housing market was key to an understanding of offender-based residential crime careers (Bottoms and Wiles, 1986). Most of their focus was on social housing and the propensity of some estates to spiral into decay while others remained largely crime free. Tenure mix and the resulting social mix was also judged important by Bottom and Wiles, and more recent work into neighbourhood satisfaction (crime being one indicator of dissatisfaction) also demonstrates how being a member of a tenure minority in a given neighbourhood tends to make people more likely than usual (for their social class) to be dissatisfied with the area in which they live (Parkes et al., 2002).

Furthermore, barriers may be created by absentee or intransigent private landlords with regard to tenants providing themselves with adequate security and protection for their properties.

Qualifications

Low intelligence and attainment have predicted both juvenile and adult convictions (Farrington, 1992). Furthermore, low qualifications may be a barrier to finding stable, reasonably paid employment, especially in places where industries that have previously provided low-skilled, or at least low-academic skilled work, are in decline. Data from the census on qualifications is problematic in that the most commonly used cohort - those aged 16 to 74 - will have lived through different regulatory frameworks (e.g. people may have been able to leave school at 15) and through times when non-academic skills training, such as apprenticeships, were more common.

4.5.3 Cartographic and Land Use Data

The Ordnance Survey is a Government agency that has its roots in military mapping. Today, it produces a range of mapping products, with about 80% of turnover accounted for by electronic cartography. Electronic products include rasterised versions of the popular paper series, such as the 1:50,000 Landranger, and vector and/or object-based products such as Land.Line and MasterMap which can be manipulated within a GIS. The Joint Information System Committee (JISC) negotiates a license for UK Higher Education institutions to use a range of these electronic products. These are made available to users via the Digimap service hosted at the JISC-funded national data centre, EDINA.

In the past, small-area land cover and land use analysis would probably have been attempted using the line-based Land.Line cartography but would have been complicated by the need to use GIS enclosure routines to create region objects prior to sampling by small-areas. Fortunately, recent changes to the JISC license have made it possible for academic institutions to obtain samples of the MasterMap mapping layers. Within MasterMap, objects with area are represented as polygons which removes the need for a priori enclosure analysis, thus saving significant amounts of time, and making identification of area features more reliable.

Through the JISC license, a copy of the MasterMap Topography Layer was obtained for the Leeds Metropolitan Area specifically for the LCCS research. Leeds City Council and most other local authorities also license this data from Ordnance Survey under the local government Mapping Services Agreement. As such, although the use of MasterMap data for area classification is not common, data access should not present a problems for local authorities and CDRPs wanting to replicate this aspect of the research.

MasterMap Topography

The OS MasterMap Topography Layer is a large-scale digital database of the detailed surface features on the landscape. It is a highly accurate, flexible resource covering some 400 million man-made and natural features, from fields to pillar boxes, each with its own unique Topographic Identifier, or TOID. The data is classified in a number of ways, one of which is by *themes*. These themes are,

- roads, tracks and paths
- land
- buildings
- water
- rail
- height
- heritage
- structures
- administrative boundaries.

Within this hierarchy, individual themes are sub-classified in different ways. Within *land*, for example, objects are classified as being either *natural* or *man-made*. In addition, *land* may have a more descriptive label to discriminate between *rough ground*, *woodland* or *scrub*, for example.

Category	Id	Variable	Layer
Land use	DS01	Average area of multi surface (garden) per household (sq m)	Topography
Land use	FP02	Buildings of 29 to 150 sq.m.	Topography
Land use	FP03	Buildings of 150 to 1000 sq.m.	Topography
Land use	FP04	Buildings of 1000 to 3000 sq.m.	Topography
Land use	FP05	Buildings of 3000 to 8000 sq.m.	Topography
Land use	FP06	Buildings of 8000 + sq.m.	Topography
Economic	LU01	Public houses	Topography
Land use	NL01	Area covered in natural land	Topography
Land use	RA01	Area covered by railway tracks and land	Topography
Land use	RE01	Ratio of residential to non-residential delivery points	Address
Land use	RO01	Area covered in roads	Topography

Table 4.7: Variables identified from Ordnance Survey MasterMap.

Most of the topographic features that are of interest for this research are represented by polygons. For buildings this does not present many problems but roads need special treatment. In the OS Meridian2 dataset, roads are represented as vectors, so calculating network length within a given neighbourhood is straightforward. However, Meridian2 is not as comprehensive in detail as MasterMap, and some minor roads are omitted. MasterMap does include these missing minor roads but all roads are represented using multiple polygons. Thus, it becomes easy to calculate road area but not road length.

Address Layer

Another MasterMap layer obtained for the research (although not covered by the JISC license) was the Address Layer. At a national level, this contains information on more than 26 million addresses in the Royal Mail Postal Address File (PAF). Each address is referred to as a delivery point, and each of these records includes a TOID reference to enable the delivery point to be matched to a building outline in the Topography Layer.

Address fields within the Address Layer database pertaining to organisation or department name make it possible to estimate which delivery addresses are residential and which are non-residential. Some resource-intensive manual sifting is required to deal with properties which are owned by organisations but are in fact residential, e.g. university halls of residence. However, the result is the ability to estimate, for a given area, the balance of residential to non-residential delivery points, and, though linking with the Topography Layer, improve the accuracy of land use analysis based on building footprint sizes.

The Address Layer can also be used more conventionally to geocode local datasets used elsewhere in the research. This improves the resolution of geocoding compared to what is possible using the JISC licensed OS CodePoint product, which only provides geocoding at unit postcode level.

4.5.4 Selection of MasterMap Variables

Variables created from MasterMap data are shown in in Table 4.7.

Defensible Space

The theory of defensible space, first described by Newman (1973), is based on an understanding that features in the built environment - such as indicators of territory and surveillance opportunities - can defend an area against crime (see Section 2.2.2).

Defensible space *per se* is not defined within the MasterMap topography, yet areas defined as *multi-surface* have potential to be used as a proxy for this concept, although this is restricted to measuring private and not public defensible space. Types of area classified as *multi-surface* can be seen on the map of Potternewton (Figure 4.6). Typically, *multi-surface* is used to describe gardens or grounds of houses. It is not used for parkland (e.g. Potternewton Park), or large grassed areas of businesses or organisations (e.g. grounds of Chapel Allerton Hospital or Bracken Edge Primary School), although it does pick out grounds of communal properties such as flats (e.g. Newton Walk and Newton Park Court).

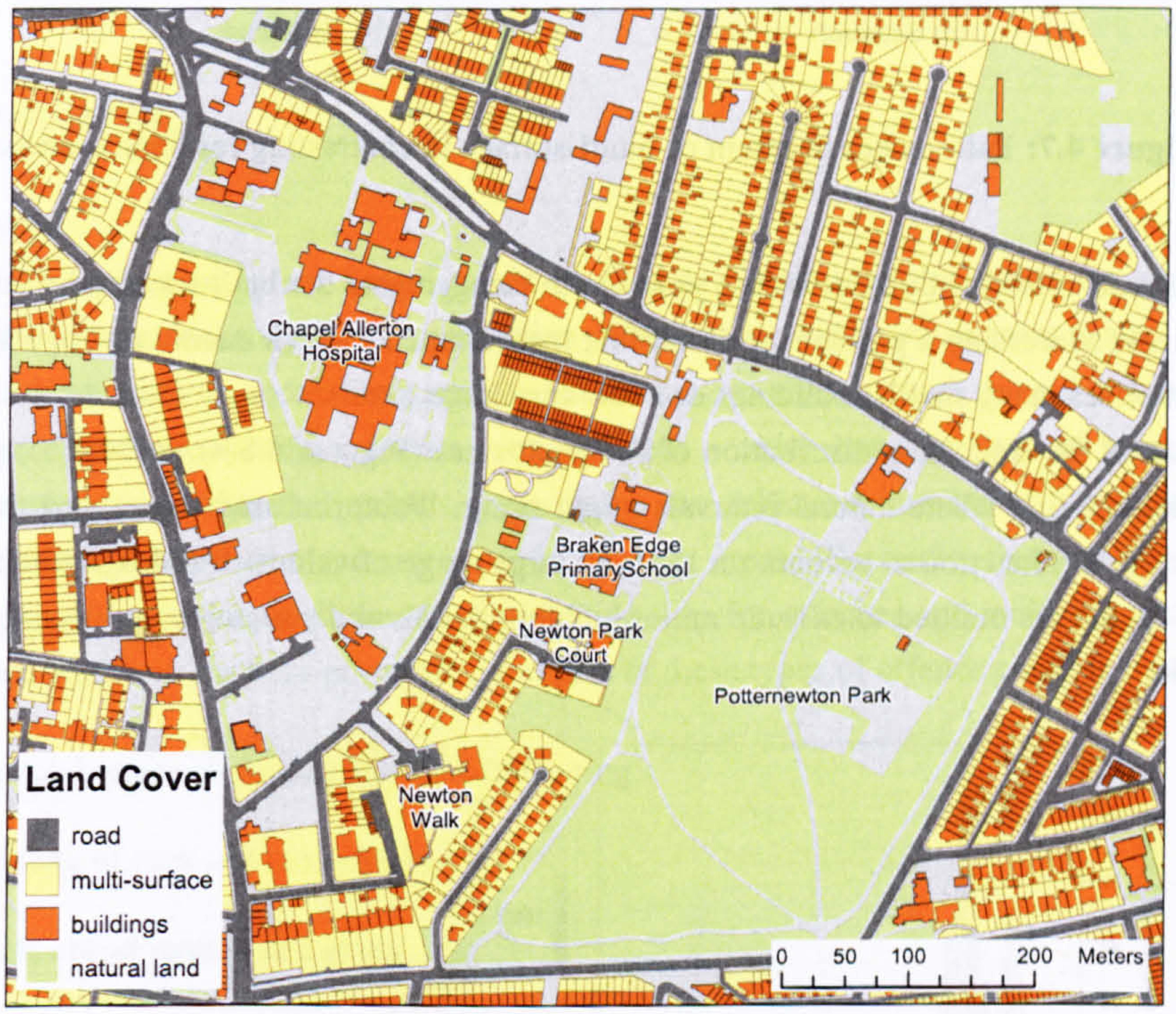
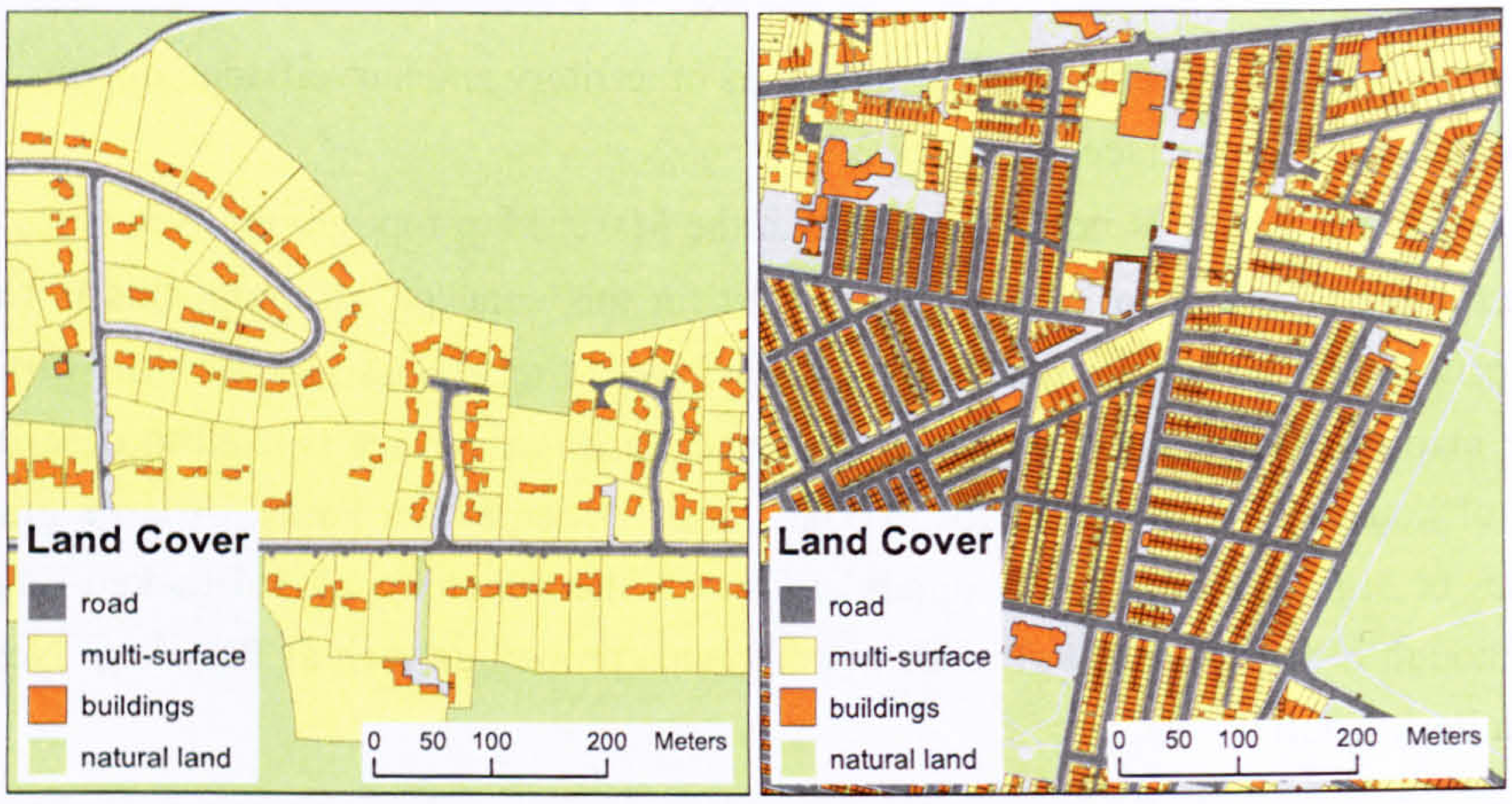


Figure 4.6: Defensible space in Potternewton. Source: Ordnance Survey.

In terms of private defensible space, *multi surface* can help differentiate areas such as Victorian terraces opening directly onto the pavement and areas where large private gardens are common. In the first scenario, the selection of burglary targets without attracting undue attention may be aided by being able to peer easily through front windows and by being able to use shared alleyways and backs to access the rear of properties. In the second scenario, the garden may act to clearly delineate private from public space, reducing opportunities for potential burglars to survey properties unnoticed. Examples of extreme values of this private defensible space are to be found in places such as Scarcroft (Figure 4.7(a)) and Hyde Park (Figure 4.7(b)). Both maps are at the same scale.

Building Footprints

Calculating the proportion of a neighbourhood that is covered by buildings provides an opportunity to measure the extent to which a neighbourhood is 'built up' (although elevation, of flats for example,

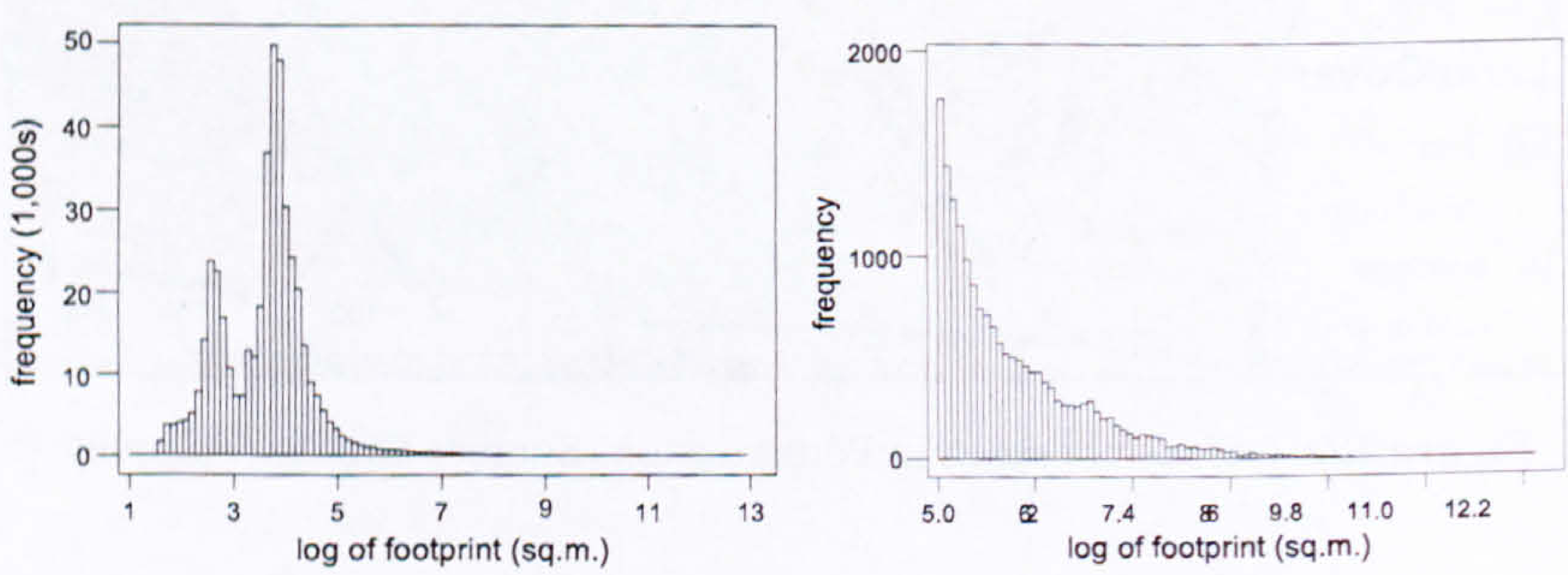


(a) Scarcroft (b) Hyde Park

Figure 4.7: Differences in extent of multi-surface in contrasting residential areas.

is not taken into consideration). It is also possible to disaggregate the built-up area by buildings of different sizes, addressing a problem with the first measurement, which cannot distinguish whether an area comprises many smaller buildings or a few very large ones.

Analysis of the frequency distribution of footprint sizes (Figures 4.8(a) and 4.8(b)) provides a basis for determining where natural intervals might occur. Footprint areas have been transformed using a natural log function to account for the very large range of values. Buildings with a footprint less than 5 sq.m. are omitted to exclude minor building details and structures up the size of a large garden shed.



(a) All buildings (b) Large buildings

Figure 4.8: Histograms of building footprint sizes.

Comparing the histograms with sample footprint sizes in Table 4.8 suggests the first peak probably represents single-garages and similarly sized structures. The major peak will be housing, ranging from small back-to-back terraced properties with a footprint little larger than a double garage, through to large detached properties. The frequency of large buildings then decays exponentially with building size, shown in more detail in 4.8(b).

The decision was taken to have one footprint interval to include residential housing (29-150 sq.m) and then a further four intervals to separate larger buildings, leaving open the possibility of collapsing two or more of these groupings into one should later analysis reveal the variables to have poor

Building	Area (sq.m.)	Ln(Area)
Single garage	14	2.64
Typical back-to-back terrace	29	3.37
Typical through terrace	40	3.7
Typical semi-detached house	46	3.83
Large detached house	132	4.88
School of Geography, Main Bldg.	716	6.57
Civic Hall	3520	8.16
Tesco, Roundhay	4348	8.37
British Library, Boston Spa	20399	9.92
Yeadon Airport Ind. Est. complex	106838	11.58

Table 4.8: Examples of footprint sizes for typical buildings.

discriminatory power.

Natural Land

Typically, the amount of 'green space' might be used to discriminate land used for agriculture or forestry, for example. Yet a significant proportion of land cover within urban areas is also 'green', including parkland, gardens, recreation grounds, allotments, woodland, rough ground and landscaping.

In a classification covering a district such as Leeds - which has a mixture of urban and rural areas - including a variable that reflects green space can achieve a number of objectives. Firstly, it can complement population sparsity in defining rural areas. Secondly, from an environmental criminology perspective, it might be able to identify sites for offences and incidents which are not linked to residential or business property. Examples of these types of offence could include,

- criminal damage to park furniture and planting
- joy-riding of cars and motorbikes
- vehicle abandonment and arson
- fly-tipping
- grassland fires

In addition, areas of green space in urban areas may increase the vulnerability of adjacent properties.

A third objective of including an urban green space variable is to reflect the possible positive amenity benefits provided by parks and recreation grounds. This tension and contradiction that exists between this positive appraisal of urban green space, and the fact that it can also act to increase opportunities for crime is problematic for area classification and local authority management. The specific meaning that the variable takes on in the classification may depend on its relationship with other variables.

Housing Density

Variables for housing type have already been identified and collected from the 2001 Census (see Section 4.5.2), but in addition it may also useful to consider household density. A basic indicator has been included which measures households per hectare (variable HD01), but expressing housing density in this way is insensitive to localised variations within the neighbourhood. For example, sub-community area (SCA) 1.01, which is in Adel (Figure 4.9), has 526 households. The area of SCA 1.01 is 2.631 sq. km., yet the area that contains the housing is just 0.266 sq. km., a factor of ten different.

An alternative method for expressing housing density is to use the total length of all roads in a neighbourhood as the denominator. Most households have vehicular access, and although roads do run through unpopulated areas, they are less common. Unfortunately, the MasterMap Topography Layer does not contain road centre-lines so road area has to be used as a proxy for road length. Thus, the total area of 'road or track' in SCA 1.01 is 54,750 sq. m., while the total area of 'road or track' in the area containing the housing is 35,874 sq. m., just a factor of 1.5 difference. It is argued that the difference between the denominator factors in this example suggest housing density by road area might be a better measure than housing density by land area.

Problems with this approach to housing density include the effect of roads with large surface areas, such as dual carriageways and motorways. Furthermore, there is no way of knowing whether a housing density by road network value is showing many houses tightly packed around a relatively small and compact road network or houses strung out in linear settlement pattern. Finally, areas with relatively small amounts of housing but large areas of other developed land (with their accompanying road network), such as industrial estates, would also introduce problems.

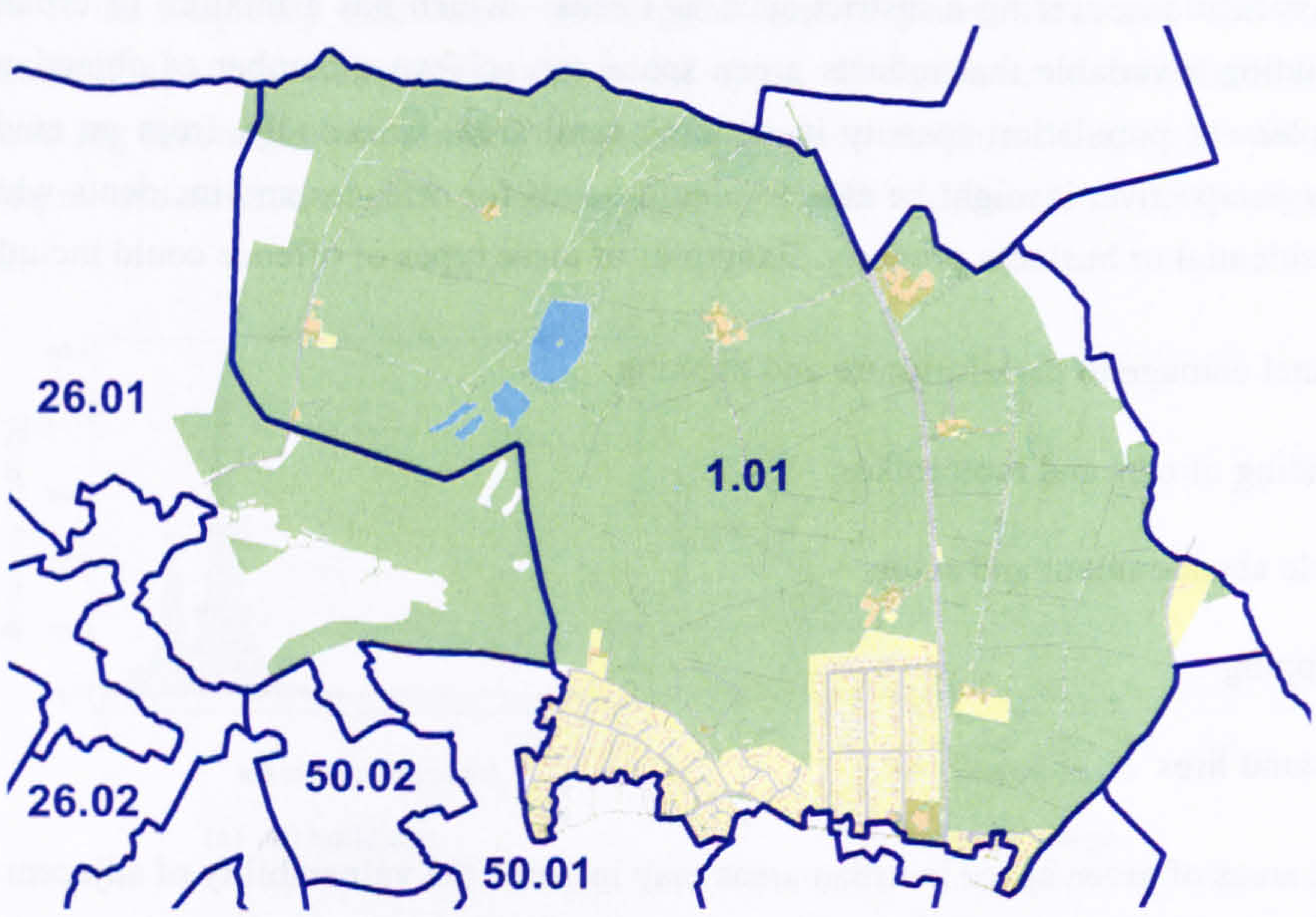


Figure 4.9: Land use in SCA 1.01 (Adel).

Mixed Land Use

It has been shown that neighbourhoods with more permeable boundaries have higher crime rates than those where residential homogeneity has a reducing impact on the through-flow of outsiders (Greenberg et al., 1982). Furthermore, as non-residential land use increases, residents' perception of control may decrease (Kurtz et al., 1998). These findings are in line with Stark's proposition that "mixed use increases familiarity with and easy access to places offering the opportunity for deviance" (Stark, 1987, page 899). Yet, beyond some obvious non-residential crime attractors and generators such as bars and schools, opinions differ about whether mixed land-use is necessarily a bad thing in a predominantly residential context (Taylor, 2002).

Whichever is the case regarding victimisation of residents in mixed-use neighbourhoods, information about non-residential urban land use is an important consideration when analysing crimes against businesses. To some extent, the census-derived variable PD01 (daytime to residential population ratio) is already indicative of which neighbourhoods have employers within them, but it might also be important to reflect mixed land use at a property level in order to express whether the non-residential land use is accounted for by a few or many different businesses and organisations. The ratio of residential to non-residential delivery points is thus calculated by exploiting the work described above (Section 4.5.3) to differentiate between residential and non-residential delivery points.

Public Houses

Alcohol may influence the social and cognitive processes that may lead to aggression and violent crime (Deehan, 1999). In addition, although alcohol may be neither a necessary nor sufficient cause of crime, it may nonetheless affect crime (Raistrick et al., 1999). Examples might include the use of alcohol to build up courage to commit an offence, or in more indirect and situation ways, such as frustration and violence that breaks out in many UK cities as a result of competition for transport home after licensed premises close on a Friday or Saturday night.

Thus, it may be useful to include a variable to reflect the density of public houses and bars in an area classification for community safety, although it may be that the densities which constitute a critical tipping point might only occur in a few locations. Whichever is the case, it is possible that problems are likely to lessen as distance from the pub, club or bar increases. A surface model was created using kernel density estimation in an attempt to simulate this decay effect.

Primary data was collected from Ordnance Survey MasterMap text labels of public houses. This method provides accurate point locations for 579 public houses. Establishments like the bars within Student Unions are not included. Night clubs are also missing. A better solution would be to use data directly from the local licensing agencies and then geocode the addresses of all licensed premises.

Retail Land Use

By definition, stealing from a shop is an offence constrained by the availability of retail outlets. Thus, a variable that can capture retail land use might help classify those neighbourhoods where we see high rates of *Theft from a Shop*. As with pub density, a surface modelling approach was adopted based on point locations of shops identified by organisation name in the OS MasterMap Address Layer. The data is more complete for the larger supermarkets and least complete for small independent businesses with names that do not reflect the nature of their activity.

Data provided by West Yorkshire Police on items stolen from shops (Table 4.9) shows that aside from the 'other' category, the most popular items for shoplifters are food, clothing, alcohol/cigarettes and cosmetics/toiletries. Thus, the importance of certain shop types was weighted according to the nature of goods sold - supermarkets being multiplied by a factor of 8, off-licenses by a factor of 4, and chemists by a factor of 2. An additional building footprint weighting was added to supermarkets to reflect the possibility that larger stores would suffer from more thefts.

Kernel density estimation was used to generate the surface (Figure 4.10). The compound effect of the different weighting measures was tested against a surface created without weights (i.e. all shops were weighted equally). Pearson product-moment correlation coefficients show the weighted surface

goods	stolen	offences	% of total stolen
other	£93,925	1275	18.5
clothing	£85,997	727	17.0
jewelry and watches	£56,311	96	11.1
food items	£39,232	1306	7.7
perfume and toiletries	£26,224	347	5.2
wines and spirits	£24,777	601	4.9
spectacles	£11,627	46	2.3
mobile phone	£11,406	49	2.2

Table 4.9: Value of property stolen in cases of Theft From a Shop, ranked by total value of items stolen.

is slightly better correlated to Theft from a Shop (0.392, P=.000) and Other Theft (0.353, P=.000) (which contains shop and other types of theft) than a non-weighted surface (0.351 and 0.341, P=.000). The reliability of these calculations will be affected by bias in the underreporting of problems by retail outlets to the police, with some stores reporting many incidents to the police and others none, according to the value of the loss and decisions about whether to pursue an insurance claim to recover the loss (Shury et al., 2005).

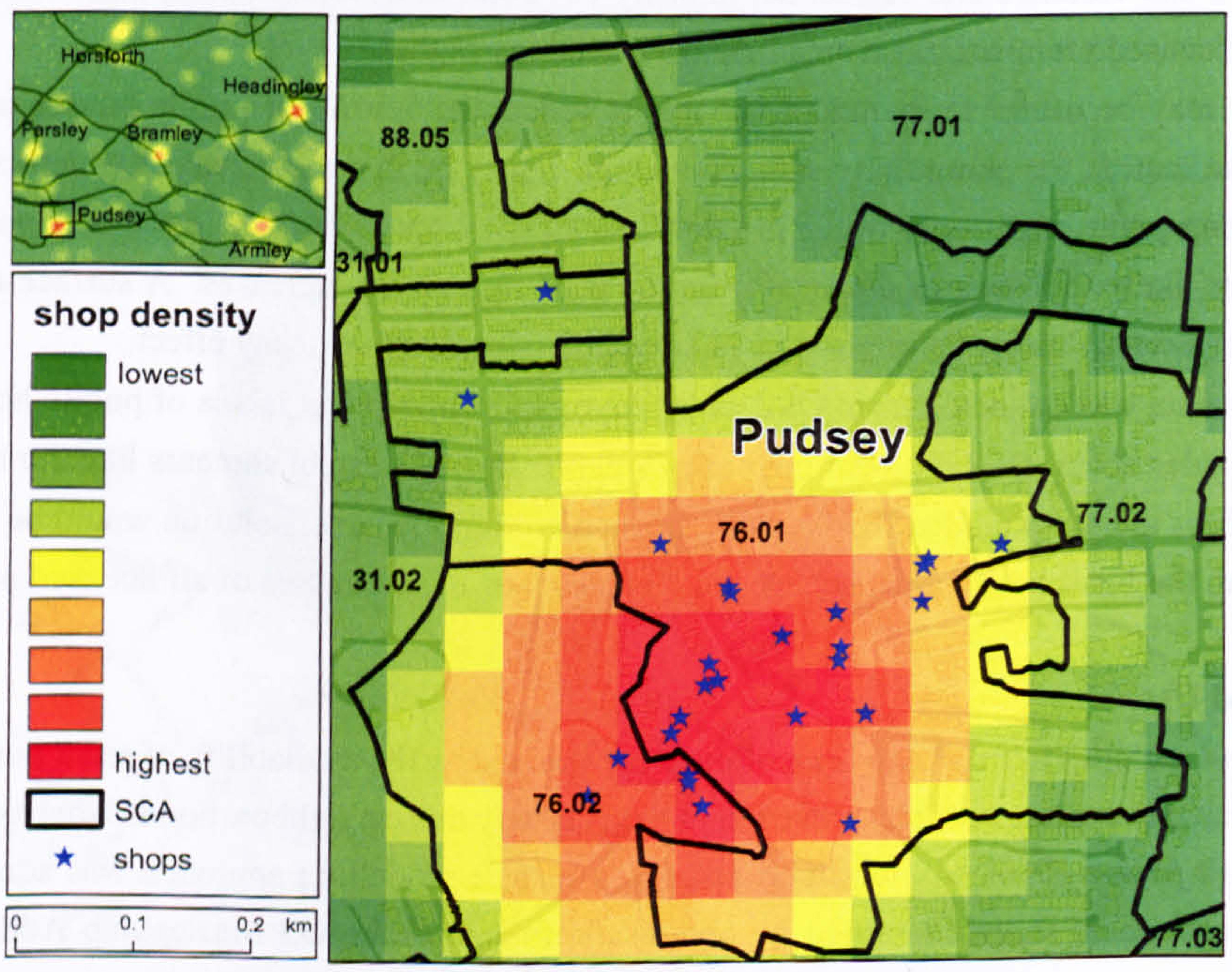


Figure 4.10: Shop density in Pudsey town centre, overlayed with locations of shops identified in OS AddressPoint.

4.5.5 Business and Economic Data

Well before the planned release of workplace data from the 2001 census, ONS were making it known that disclosure-related concerns might mean some variables and/or whole tables would be suppressed in the final outputs. The fears were subsequently confirmed and OA level tables such as the proposed UV77 Industry (Workplace Population) were never released. This had implications for identifying

Category	Id	Variable	Source
Economic	CL01	Population who are claiming employment related benefits in April 2004	Nomis
Economic	CL02	Change in the claimant count rates from April 1997 to April 2004	Nomis
Economic	JB01	Difference in workplace jobs from 1998 to 2002	ABI

Table 4.10: Variables identified from Nomis and the Annual Business Inquiry.

the scale and location of different economies in Leeds, particularly the nighttime economy of bars, pubs and clubs. For this reason, access to the Annual Business Inquiry (ABI) was secured via the ONS official labour market statistics portal, Nomis. Apart from offering an alternative to the census for some data types, the ABI also has the capacity for producing much more timely data.

The ABI, as its name suggests, is conducted every twelve months and contacts all UK businesses that are registered for Value Added Tax (VAT) and/or Pay As You Earn (PAYE). Each business is classified by type using the Standard Industrial Classification (SIC). Data are aggregated and published using a range of geographies, the smallest of which was, until recently, the postal sector. (Since the construction of the classification data has been published for the Lower Level Super OA geography).

Also available through Nomis is a more timely source for unemployment, or more accurately, claimant count data. The 2001 census definition of unemployment is consistent with the International Labour Organisation (ILO) definition, whereas the claimant count is derived as a by-product of the Department for Work and Pensions administrative system and covers the number of people who are claiming unemployment-related benefits (Job Seekers Allowance and National Insurance credits) at Jobcentre Plus local offices.

Nomis claimant data are published monthly down to Lower Level Super OA and postal sector levels. Data at Super OA level only goes back to October 2004 while data at postal sector level goes back to June 1983. The length of this time series affords the opportunity to explore the dynamics of the claimant count and identify neighbourhoods where the unemployment situation has improved or worsened. Seasonally adjusted counts are also available.

4.5.6 Selection of Nomis and ABI Variables

Numerous studies have shown substantial associations between unemployment and crime (Rutter et al., 1998). A number of these show that antisocial behaviour in childhood predicts unemployment in adult life, but it is obvious that most unemployment does not arise from antisocial behaviour (Petersen and Mortimer, 1994). Instead, what are of interest are the effects that unemployment, or the threat of unemployment (Ferrie et al., 1995), might have on an individual's propensity to commit crime or behave antisocially. Most obviously, job loss is likely to lead to a marked drop in income which in turn might lead to an increase in committing crimes for material gain - something suggested by a longitudinal study of young men living in London (Farrington et al., 1986). Work may also be an important source of social interaction, providing people with a sense of personal worth, and unemployment has been shown to be associated with a deterioration in mental health and a loss of self-esteem and may reduce a person's bond to society and to its values (Sampson and Laub, 1993). Life stresses such as unemployment may also accentuate pre-existing behavioural tendencies (Caspi and Moffitt, 1993), and have been shown to contribute to the problems encapsulated within

Category	Id	Variable	Department
Demography	AS01	Population who are asylum seekers (summer 2004)	Leeds Refugee and Asylum Support Service
Minor incivilities	ED01	Days lost per pupil to fixed term exclusions	Education Leeds
Minor incivilities	ED02	Pupils receiving fixed-term exclusions	Education Leeds
Minor incivilities	ED03	Pupils permanently excluded	Education Leeds
Social	ED04	GCSE students only getting 0,1 or 2 passes	Education Leeds
Social	ED05	Average number of GCSE passes A* to C	Education Leeds
Health	MI02	Discarded needles collected	Environmental Services
Minor incivilities	MI03	Noise complaints	Environmental Health
Minor incivilities	MI04	Abandoned cars recovered	Highways

Table 4.11: Variables identified from Leeds City Council departments and services.

strain theory, affecting, for example, the likelihood of men admitting to spousal violence, although this particular behaviour may also be mediated by age (being more problematic for younger men) (Howell and Pugliesi, 1988).

4.6 Local Datasets

It was an objective of the research to negotiate access to locally produced data, particularly that produced by various departments within Leeds City Council and their community safety partners. Opportunities arose to work more closely with community safety partners on a couple of large analysis projects. Although strictly outside the scope of the research, these consulting roles had the effect of conferring *insider* status on the author. This new status, along with opportunities to demonstrate competence in data handling and analysis generated trust and improved access to information and key personnel considerably.

4.6.1 Leeds City Council Data

The Council as a whole is a responsible authority within the terms of the Crime and Disorder Act 1998. Accordingly, all departments are required to consider community safety issues during the planning and delivery of services, although this is more pertinent for some departments than others. The departments and units whose data were used to prepare variables (Table 4.11) for the classification are discussed below. The Anti-Social Behaviour Unit is not included because, although the nature of the data they generate is very appropriate, the information systems designed to record the information had not been up and running long enough to produce a suitable number of records. Moreover, no data were obtained from the Arms Length Management Organisations (ALMOs) responsible for managing the Council's housing stock. Information relating to voids, dereliction, demolition and anti-social behaviour is maintained by the ALMOs, but at the time data were being collated, data sharing agreements between the ALMOs and Leeds Community Safety were at a very early stage of development.

Environmental Services

Among the services offered by the Environmental Services department is the safe removal of needles discarded by drug misusers. Although the needle finds can be mapped there are interpretation issues to consider. Firstly, some jobs to individual properties, particularly voids, yield hundreds of needles. These may have been discarded by one person or a small group and thus may not be indicative of a serious problem in terms of numbers of misusers. Furthermore, from a harm reduction perspective, the presence of many discarded needles at a single location may be evidence of safe needle use (i.e. not sharing or re-using). Also, when a single needle is recovered, there is no way of knowing whether it has been used once or many times, or by one or many users.

Education Leeds

Education Leeds are a non-for-profit company wholly owned by Leeds City Council and are responsible for providing all education support services that relate to children and young people of statutory school age in Leeds. Typically, the types of education information of interest to community safety are academic performance of pupils and records of problematic behaviour. The Key Stage and GCSE performance indicators can help to identify neighbourhoods where young people may be becoming disaffected with school. Similarly, data on exclusions may indicate the extent to which neighbourhoods are having to contend with problems associated with pupils identified as having a poor behaviour record.

It is still the norm for published performance indicators to be school based, but there are many benefits to studying the geography of education at an individual pupil level based upon home address. In Leeds, this is especially important, as the legacy of the permeable school catchment boundaries means that a sizeable proportion of pupils do not live in the immediate neighbourhood to their school.

A special data sharing agreement was entered into for the purposes of the research. One year of data pertaining to GCSE performance, exclusions and the Pupil Level Annual School Census (PLASC) - all at individual pupil level - was obtained. The geocoding of all the data was done within Education Leeds. The PLASC data contains information about the pupils' home location, their gender and ethnicity, their school and their educational stage. GCSE performance data contains the pupils' home location and an indication of how many GCSEs were passed and how many passes were at grade C or above, for pupils of the appropriate age. Information about exclusions contains the pupils' home address and records permanent and fixed-term exclusions separately. For the latter type of exclusion, the length of the time the pupil was excluded, in days, is also recorded.

Environmental Health Services

Environment Health deal with a range of problems, and among these is another type of anti-social behaviour - noise nuisance. Complaints recorded in the dataset may turn out to be unfounded, other problems may be resolved and some problems may require formal proceedings to be undertaken. Two years worth of data (2002 and 2003) were obtained as a result of work undertaken for the Leeds Statistics project. Table 4.12 indicates the different ways in which cases are classified, and their frequency. Geocoding of the dataset was undertaken using postcode information included with each record.

Highways Services

Issues of road safety are not always prominent in community safety strategies so data relating to road traffic accidents was not pursued for the research, although it is recorded by the police using the Department for Transport STATS19 procedures. By contrast, data collected by Highways Services regarding abandoned vehicles is frequently included in crime audits and strategies and was obtained

	Frequency	
Complaint Type	**2002**	**2003**
Total number	4073	4761
Formal action taken	175	214
Unfounded	84	57
Music/parties	2389	2854
Alarms	344	423
Other (inc. dogs)	1340	1484

Table 4.12: Noise complaints received by Environmental Health Services.

as part of the Leeds Statistics project. Abandoned vehicles can represent an environmental pollution hazard and may also be a target for criminal damage, especially arson. The abandonment of a vehicle is also a criminal offence, whether it is the result of a theft or a case of an owner trying to avoid complying with the regulations on motor vehicle disposal.

One years worth of data (2003) recording the recovery of abandoned vehicles was obtained. This contains information pertaining to 1,782 jobs for the contractor responsible for the work. As the data were primarily intended for the Leeds Statistics project, the geocoding of jobs was done at a post-code level. Discussions with the staff on the abandoned vehicles team revealed that non-addressable locations (e.g. canals, waste ground, woods) have been assigned the postcode of a nearby location. However, moves were being taken to ensure that future bidders for the recovery contract used GPS technology to generate precise coordinates of job locations.

Leeds Refugee and Asylum Support Service

The Leeds Refugee and Asylum Support Service provides access to housing and gives support to people seeking asylum in the UK. Data relating to the place of residence (used for geocoding), country of origin, length of time in the UK and family status were provided for 2004 Leeds Crime Audit. Separate permission was negotiated to use this data for purposes of this research. Since the year 2000, Leeds has been home to 1,640 applicants and 1,045 of their immediate family members (Figure 4.11) and this has posed problems for service providers. Mistrust in some neighbourhoods has been high and asylum seekers have made convenient scapegoats when some local people have looked to explain social and economic decline in their area. As such, refugees and asylum seekers might be at greater risk of becoming victims of racially motivated crimes or ASB.

4.6.2 Police Force Data

England and Wales is covered by 43 regional police forces, with Leeds coming under the responsibility of West Yorkshire Police. British Transport Police also have responsibilities at certain locations in Leeds, and other non-geographic forces (e.g. The National Crime Squad) are also active for specific types of crime. While there are nationally set standards for police procedures and crime recording, the computer systems that provide information management vary from force to force. Thus, some of the discussion below only applies to the situation in West Yorkshire. Problems may be more severe in some forces, or non-existent in others.

Recorded Incidents

Incident data are recorded on the Incident Based Information System (IBIS) Command and Control

System. The details recorded include the callers own details, the nature of the incident and its location. The incident record is then transmitted to the open incident list at the appropriate dispatch centre where it is prioritised for action. IBIS presents the operator with a list of available resources, and enables the operator to enter details of when resources arrived at, and subsequently left, the incident. Details reflecting the outcome of a dispatch are also recorded, for example 'all quiet on police arrival'. If no criminal offence is thought to have been committed then none is recorded, although the incident records are stored and can be queried later by WYP and other analysts.

The nature of an incident is initially recorded by the operator based on the description given by the caller. When a call is finalised (e.g. after an officer has attended the incident), a final definitive code for the incident, which may be different, is recorded. Prior to April 2005, WYP were using a hierarchical incident classification that in 2003/04 was used to generate 6,961 *different* final codes at the lowest level. Since the introduction of the National Standard for Incident Recording (NSIR), adopted by WYP on 1st April 2005, incidents are now coded using a two level hierarchical classification with 7 classes at the top level and a more manageable 135 classes of incident at the lowest level. Qualifier codes are also available to enable operators to describe more of the nature of an incident (e.g. if an incident is thought to be hate-related, then there are separate qualifiers for race, religion/faith, gender, homophobic, transphobic, disability and age).

The spatial information recorded in IBIS is not always reliable or complete and geocoding is not quality checked in any way. As such, individual level spatial analysis has to undertaken with care and aggregations become more reliable as the scale of the geography increases. A number of subsets of this data were obtained for the research. These cover ASB, disorder incidents and traffic incidents and each represent aggregations of records based upon final codes.

Recorded Crimes

If an offence is known or (sometimes) suspected of having been committed, then a crime record is created on the Criminal Investigation System (CIS) database. For certain types of offence, the decision to record a crime is straightforward, in the case of report of a burglary, for example. Other types of incident may be less clear cut, for example crimes that arise from incidents of domestic violence, street disorder, racist insults and minor violence. The Policing Standards Unit (PSU) recently

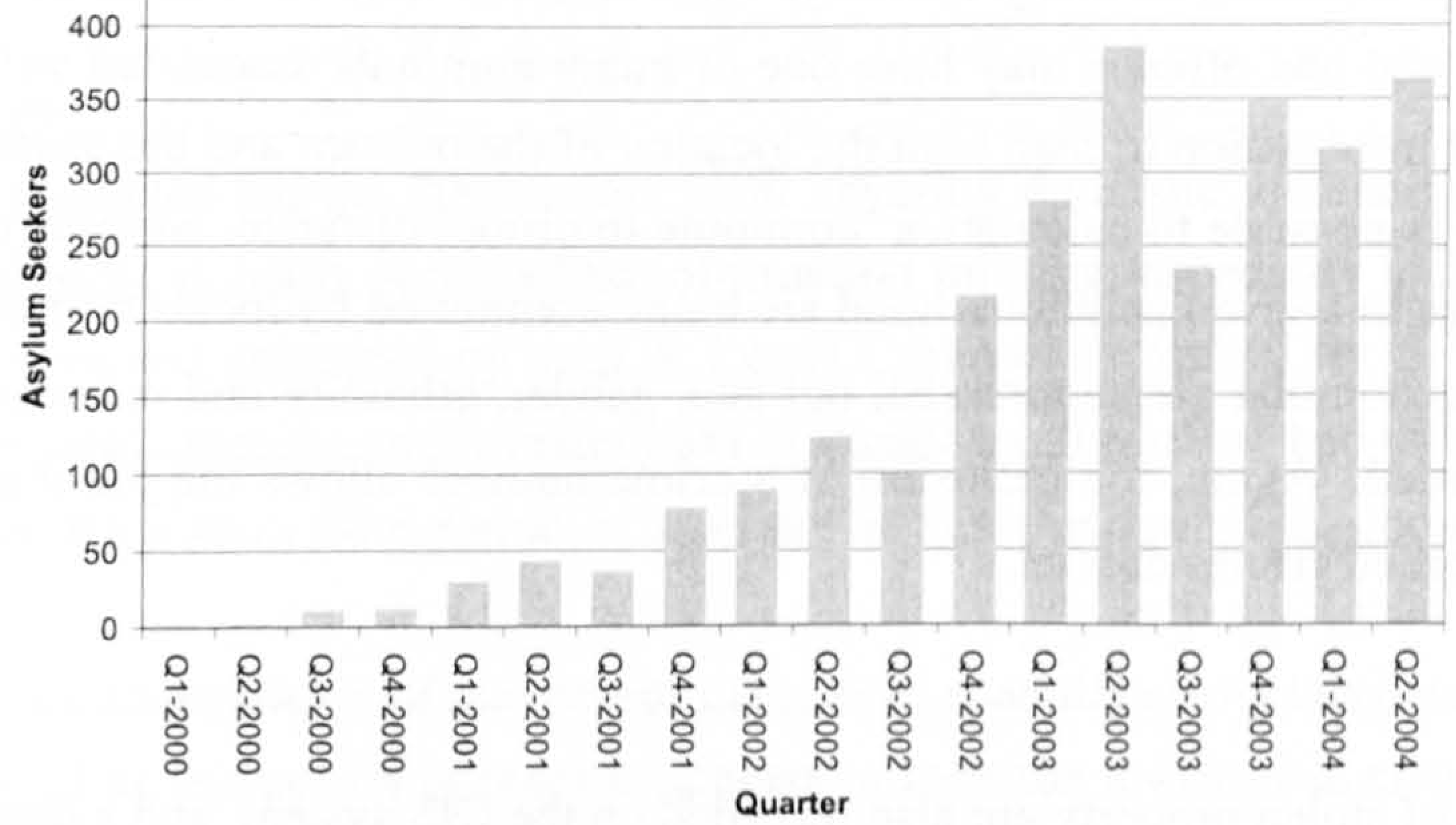


Figure 4.11: Numbers of asylum seekers arriving in Leeds per quarter, from Q1 2000 to Q2 2004.

audited all police forces and compared their IBIS and CIS-type logs. This was undertaken to assess how effectively police forces are implementing the National Crime Recording Standard (NCRS) (see below). West Yorkshire Police performed well in some areas, but badly when recording crimes for the incidents just mentioned, and overall were awarded an 'amber' rating (The Beat, 2003).

CIS data extracts were obtained for this research and contain basic details about the nature and location of the crime, along with a unique crime number. Age, gender and ethnicity of the victim are also recorded, along with the same data for the offender, if known. Data for five years, from 2000/01 to 2004/05 were provided by WYP. Each offence is given a 100m resolution Ordnance Survey grid reference generated from address information. Non-addressable locations are geocoded using the nearest postal address to the scene of the crime, and other geocoding issues are discussed in more detail below.

Modus Operandi

Information about how an offence was carried out is also entered on the CIS record. For this research, *modus operandi* data were obtained as a separate extract and only for years 2002/03 and 2003/04. The inclusion of a crime number allows the MO information to be joined with its crime record. MO information for crimes such as burglaries may include the method by which the burglar gained entry (e.g. climb drainpipe), the entry point (e.g. window - upper floor front), any particular means used (e.g. property - removed in bin liner) to gain entry, what was taken (e.g. jewelry - gents) and whether the burglar left any trademarks (e.g. shower used). Although the classifications used to describe different aspects of the MO are comprehensive, discussions with senior intelligence analysts reveal that they are not used consistently.

Nominals

This dataset, covering offences recorded for 2001/02 to 2003/04, contains information about known individuals who WYP suspect were engaged in or connected to recorded crimes. In the case of proven detections, the link between the nominal and the offence is clear. Yet, nominals may also be connected with offences if they are suspects, known accomplices or in some other way likely to have been involved. In each of these cases, the possibility exists that the nominal had nothing to do with crime at all.

A single record in the dataset matches one nominal with one offence. The relationship between these two variables however, is many-to-many. One nominal may have been associated with one or many offences and one offence may have one or many nominals associated with it. The dataset contains sufficient information to map both the location of the offence and the usual residence of the nominal. Thus, it is possible to calculate a 'commute to crime' distance. Similarly, it is possible to test whether crimes in a given neighbourhood are being committed by locals or outsiders. The name and address of the nominal are suppressed, but age, gender, ethnicity and reference number of the nominal are recorded. Again, the inclusion of a crime number allows the joining of the nominals dataset to the recorded crimes dataset.

Stolen Property

Types and values of stolen property are also recorded on the CIS system, and a dataset was obtained for the year 2003/04. All crimes, except theft of motor vehicles, are contained within the dataset. In 2003/04, £29 million of property was stolen in Leeds and the dataset enables analysis of specialised

		Number	%
Total incidents		18325	
Repeat incidents		6294	34.3
Victim	Female	15906	86.8
	Male	2352	12.8
Suspect	Female	2048	11.2
	Male	16328	89.1
Children present		6286	34.3
Incident involved	Alcohol	7748	42.3
	Drugs	492	2.7
	Both	158	0.9

Table 4.13: Summary of domestic violence incidents in Leeds, 2001/02

geographies of theft, and also helps to estimate the size of local economies for stolen property.

Hate Crime

Originally obtained for the purposes of the 2004 crime audit, this dataset represents offences or incidents which were deemed to have a hate element to them. The data covers years 2001/02 to 2003/04. Coding within the data allows hate to be disaggregated into racial and homophobic types. Special permission was sought from WYP to use this dataset for the research and this was granted on the understanding that it would be aggregated to a neighbourhood geography. In its raw state, the data do not include the names of victims or their house number. The postcode is provided for geocoding purposes, although the omission of location information from a large number of records means that only 69% of cases could be geocoded. Other information about the victim includes gender, ethnicity, asylum status and whether they are a council tenant. The nature of the incident is also recorded and this is coded using terms such as abuse, assault, graffiti, etc. The suspect's ethnicity is recorded if known, and a crime number is provided in cases where the incident represents a notifiable offence.

Numbers of hate crimes have remained fairly constant over the three year period, with around 1,100 racially motivated cases each year and between 27 and 97 homophobic cases. The most common form of incident is abuse (25.4%), followed by assault (14.8%); 51.6% of victims were Asian, 23.4% were white and 15.4% were black. Men were only slightly more likely to be victims of homophobic hate crime than women.

Domestic Violence

The first dataset presented for the 2004 crime audit covering domestic violence was of little value for small-area analysis, as many records had information missing, especially location information. A subsequent dataset was produced by staff at Leeds Community Safety for a 7 month period in 2003/04. Much manual cleaning and investigation of location information had to be undertaken, but the result, although for a short time period, is spatially accurate and provides complete coverage for Leeds.

Table 4.13 shows a breakdown of the records in the first, spatially erroneous dataset. Data are only provided for 2001/02 as this period of records appears to have the fewest data errors. Unfortunately, the level of detail in this table is missing from the later 7 month dataset. It might have been possible to join the two datasets by creating a compound key based on attributes such as date and time, but this was not attempted.

BCS Crime	percent
Not reported to the police	58
Reported to the police but not recorded	11
Reported and recorded	31

Table 4.14: Proportion of BCS estimate of all crime reported to the police and recorded by them (comparable subset of crimes), year to September 2003 (Dodd et al., 2004, from).

Reporting and Recording Crime

For a variety of reasons, not all crimes are reported to the police and despite national standards and guidelines on recording practices, not all the crime that is reported is subsequently recorded by the police. Many factors affect levels of crime under-reporting and the British Crime Survey (BCS) is the most reliable way to estimate the size and nature of this problem. The BCS measures the amount of crime in England and Wales by asking people about crimes they have experienced in the last year. The survey is conducted annually, with approximately 40,000 interviews of people aged 16 or over. It is still possible that respondents will not recall all their experiences of crime, but the assumption is that they will. Table 4.14 summarises the scale of the under-reporting and under-recording problems for a comparable subset of crimes in England and Wales. The important fact to draw from this information is that despite much attention being focussed on producing accurately geocoded sets of crime records, it is likely that these capture less than a third of all crimes committed, assuming the problem of under-reporting and under-recording in Leeds follows national trends.

Furthermore, reporting varies by the type of offence. Most likely to be reported are thefts of vehicles (95%) and burglaries where there was a loss (78%), for reasons related to the pursuit of insurance claims. At the opposite end of the scale, vandalism and common assult (with no injury) are only reported 31% and 30% of the time respectively. In 72% of instances where a crime was not reported, the main reasons given were that it was either too trivial or the police would or could not do anyting about it.

The main reason why the police would not record a crime that has been reported is that they deem the report could not be substantiated. In all other instances, there are national guidelines to ensure that forces across England and Wales record crime in the same way. The most important recent set of guidelines are defined within the National Crime Recording Standard (NCRS), introduced in West Yorkshire in April 2002. The NCRS aims to establish a more victim-oriented approach to crime recording and its effect is discussed in detail by Simmons et al. (2003). The impact of the NCRS has been to inflate some crime rates and deflate others. At the time of writing the impacts are thought to have stabilised although rises in the recording of violent crime are still being attributed to the NCRS and a greater awareness of violence issues. This point is given support from the BCS, which in 2003/04 showed violent crime had remained stable compared to the previous period in contrast to an increase of percentage points in the percentage of reported violent incidents that were subsequently recorded (Dodd et al., 2004).

Finally, the impact of the NCRS on recording rates have not been uniform across all police forces. Table 4.15 shows the estimated impact of the NCRS on recorded crime in West Yorkshire in the first year of its use. The relatively low impact on burglary and robbery offences has been attributed to an earlier policy at WYP to record onto the CIS all such incidents recorded on IBIS. Throughout this research no attempt has been made to correct for the impact of the NCRS or under-reporting,

Crime type	Estimate of impact
Violence against person	47%
Burglary dwelling	1%
Burglary other	9%
All burglary	3%
Robbery	-9%
Vehicle theft	24%
Other theft	61%
All theft	40%
Criminal damage	27%
Total crime	25%

Table 4.15: Percentage impact of NCRS on recorded crime in West Yorkshire, 2002/03 (Simmons and Dodd, 2003, adapted from)

although theoretically it might be possible to model under-reporting by matching the characteristics of respondents in the BCS to the average characteristics of populations in neighbourhoods.

Geocoding

As well as highlighting the general problems associated with using address-based data (Gatrell, 1989, see), some authors have been at pain to consider and test the importance of geocoding match rates for crime data in particular (Ratcliffe, 2004). For this research the problems inherent with the geocoding of WYP recorded crime data were apparent from an early stage but a paucity of geocoding tools meant that only small improvements could be made. These improvements centered around the appropriateness of different strategies for assigning a point location to a postcode. While WYP's geocoder snaps postcodes to a point on a 100m lattice it was found that a better strategy might be to use a product such as CodePoint (from Ordnance Survey), which assigns the postcode to a delivery point closest to the geographic centre of all the delivery points for a postcode. Essentially though, this just had the effect of shuffling points around in space, rather than improving their accuracy to any great degree.

A more serious problem became apparent during small-area analyses of crime patterns, by neighbourhood in Leeds. This problem was not with the geocoding of postcodes as such, but involved the way in which postcodes appeared to be used. Figure 4.12 summarises the problem in map form. The problem came to light when a point map of offences in Headingley revealed a concentration of crimes at the Village Hotel and leisure club. On closer inspection however, only about 30% of the 156 crimes had actually occured at this location and the remainder had actually ocurred over quite a wide area. The source of the problem was the use of one postcode, LS16 5PR. This postcode only has one delivery point - the Village Hotel, but was being assigned erroneously to many other crime records in the Headingley area. Discussions with police analysts at WYP revealed that this problem was not uncommon and was the product of poor data entry practice, brought about, in part at least, by pressure to complete data entry quickly. The suggestion was that this postcode most likely appeared at the top of a list of possible postcodes for the street name and was chosen because it was easier to select this item than scroll down and identify the correct delivery point.

Without a delivery point gazetteer, the problem appeared intractable. However, interviews with senior intelligence analysts at WYP had revealed that manual correction of these errors was undertaken routinely (every day) outside of the CIS in a software system called CPA (Crime Patterns Analysis), yet the results of these efforts were never fed back into the CIS - the source for crime records most often used by community safety analysts outside of WYP. This manual checking, undertaken by divisional analysts, does not cover every crime type, but it does cover all the main volume crime groups. In addition to correcting problems with addressable locations, analysts also manually adjust the map coordinates when crimes have occurred at non-addressable locations.

The problems of cleaning recorded crime data for the Leeds Statistics project gave the geocoding problem more prominence, and for the 2004 crime audit it was decided to make the effort to gather together all the divisional corrections to crime locations made in CPA and join them back to records from the CIS. This work was undertaken by the author as part of the 2004 crime audit. It was calculated that 83.5% of CIS crime records, covering a four year period, had their eastings and northings improved, or at least changed, by this technique. As an afterword to the Village Hotel example, the linking of the CPA records back to the CIS data solved all of the coordinate problems!

Other cleansing measures included removing crimes in the dataset that actually occurred outside of the Leeds boundary and the flagging of crimes that occurred within police stations and prisons. This last task is important when considering patterns of crime in neighbourhoods that house these types of establishment. While the crimes do take place and are serious for the officers concerned, these offences happen within a secure, not public, environment and it is argued should not be used to describe the general characteristics of the surrounding neighbourhood.

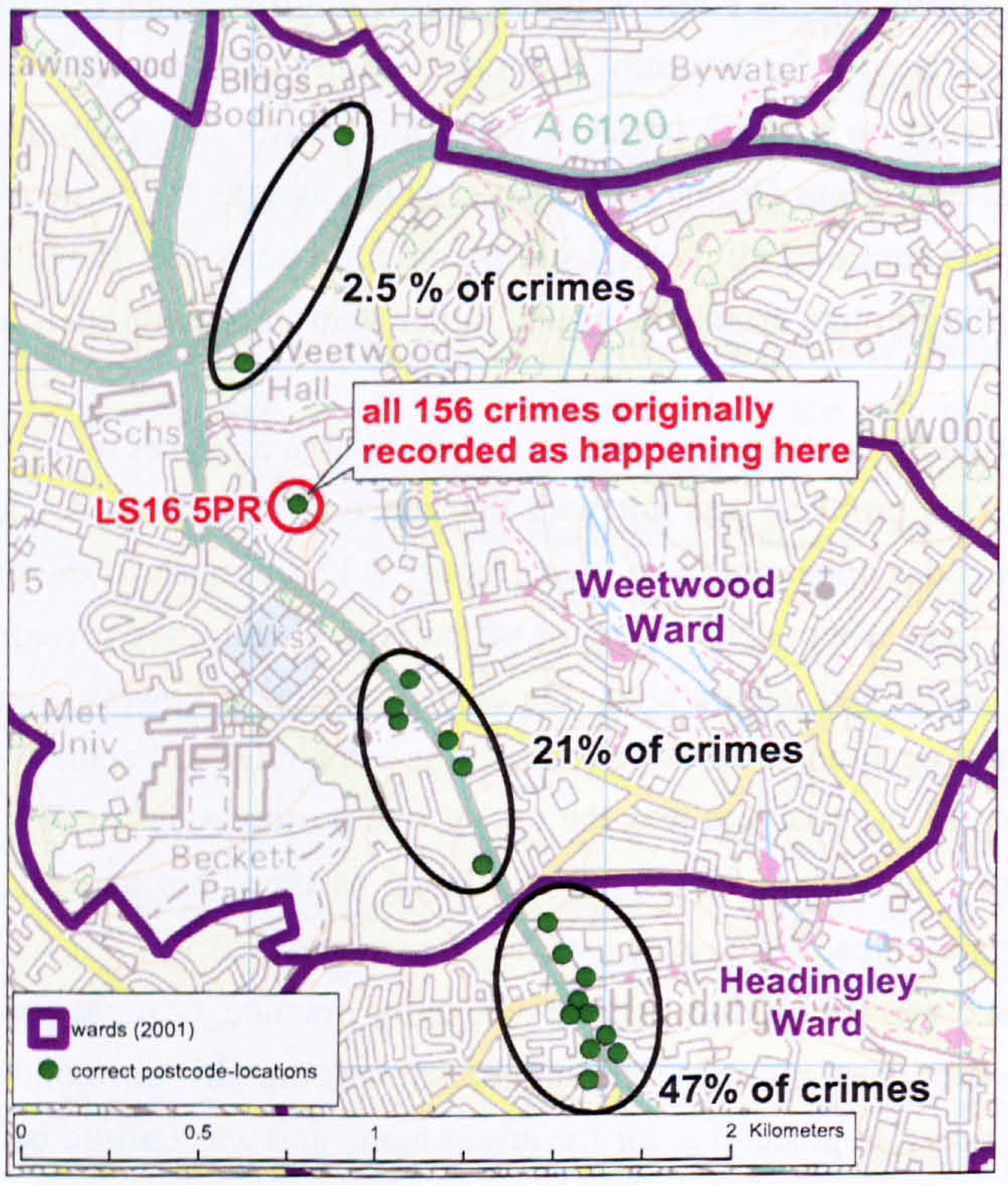


Figure 4.12: Map showing actual locations of crimes originally recorded as having occurred at the Village Hotel, Otley Road, Leeds

Category	Id	Variable	Database
Minor incivilities	MI05	Disorder incidents	IBIS
Minor incivilities	MI06	Traffic incidents	IBIS
Minor incivilities	MI07	Anti-social behaviour incidents	IBIS

Table 4.16: Variables identified from West Yorkshire Police.

4.6.3 Selection of Police Variables

Following the approach adopted for the CDRP 'families' (Harper et al., 2002) classification, it is argued that crime variables should be kept as dependent and not be included within a neighbourhood classification for community safety. It would not be usual to mix dependent and independent variables in other types of multi-variate analysis, and is it possible that at best, a classification that did so would represent a unique descriptive tool. Such a classification could not, however, be used with any confidence for between-group or within-group analysis of crime differences as all results would, to some extent, be self-fulfilled prophecies.

Nevertheless, there may be a place for other types of police data within the classification, particularly data that might indicate levels of non-criminal disorder and minor incivilities. This is because early analysis of the National Crime victimization Survey (USA) led some to suggest that urban conditions (in the form of visible signs of disorder), and not just crime, were troublesome and could inspire residents' concern for safety (Garofalo and Laub, 1978). Furthermore, the 'broken windows' thesis (Wilson and Kelling, 1982) places a lot of emphasis on the effects of disorder, and these ideas were extended by Skogan (1990) to consider the effect disorder can have on informal social control, community morale and effects on the housing market. The effect of signs of incivility on fear of crime has also been demonstrated (Lewis and Salem, 1986).

After consideration, three indicators of disorder were created from the WYP IBIS incident data (Table 4.16). Each category of incident represents a collection of different Final Codes in the incident data. The collections were determined by WYP and were aggregated to neighbourhood-level using aggregates of incidents geocoded to postcode locations.

4.6.4 Fire Service Data

Firefighting and rescue services within Leeds are provided by West Yorkshire Fire and Rescue Service (WYF&RS). The fire services are responsible authorities within the Crime and Disorder Act 1998 and data sharing between WYF&RS and other LCSP partners is well developed. WFF&RS fire data are important when attempting to analyse patterns of arson, and although the common public perception might be of firefighters dealing with accidental fires, in Leeds in 2003/04, 85.8% of the 10,233 calls attended by an appliance were for fires that had been started deliberately. This number of deliberate fires was 78% higher than the number of arson offences recorded by WYP for the same period.

The reasons for the size of this discrepancy vary but for vehicle arson, for example, it is the way that crime is recorded by the police that is significant. For example, if a vehicle is stolen, abandoned and then set alight by the thief, then the offence is recorded a theft, and the subsequent arson is not recorded. In this type of situation, arson will only be recorded if it took place separate from the theft, i.e. someone else started the fire after the car had been abandoned by the thief. Yet, in discussions with senior fire fighters, it has been admitted that police dis-interest is another factor contributing to under-recording. While in the past the fire service would have made a call to the police to attend

Category	Id	Variable
Minor incivilities	FI01	Deliberate secondary fires
Minor incivilities	FI02	Fire hoax calls attended

Table 4.17: Variables identified from West Yorkshire Fire and Rescue Service.

certain types of less serious fire, the lack of response by WYP has led to a situation where fire officers no longer bother to the make the call - being concerned that they would be wasting their time.

Fires are classified into primary and secondary incidents. Primary fires cover vehicle fires and the burning of dwellings, schools and other non-derelict premises. Examples of secondary fires include the burning of refuse and refuse containers (e.g. 'wheelie' bins), grass fires and fires in derelict buildings and derelict vehicles. Whether a fire was started deliberately or accidentally is a decision made by the fire officer attending the scene. The burden of proof necessary to deem a fire as deliberate is often lower than that demanded by the police in cases of arson.

The data supplied by WYF&RS for this research cover the three years between 2001/02 and 2003/04. Fields in the data include date and time of the incident, the cause of the fire, premises type, location information and an easting and northing at either 1m or 100m resolution. The reason for this mix of geocoding resolutions reflects lack of precise location information in many of the records, especially where fires occurred at non-addressable locations.

In addition to the actual fire event data, there also exists a dataset of hoax calls, of which there were 655 in 2003. These are the cases when an appliance is dispatched to a non-existent fire. Other hoax calls are detected as such by the 999 operators but the calls are terminated and appliances are not dispatched. An analysis of the target locations for these calls shows that 13% were targeted at hospitals (including a significant number of mental health establishments) and 9% were against schools.

4.6.5 Selection of Fire Service Variables

Two fire variables were selected for inclusion in the design of the classification (Table 4.17). In discussions, senior fire officers expressed the opinion that deliberate secondary fire starting was a good indicator of anti-social behaviour. Furthermore, the sites of many of these types of incidents - such as derelict buildings and vacant lots - are not captured well by other neighbourhood variables.

4.6.6 Health and Ambulance Data

The addition, in 2004, of Primary Care Trusts (PCTs) to the Crime and Disorder Act 1998 (section 5(1)) list of responsible authorities means there is now a statutory responsibility for PCTs to share data, and work more closely with other community safety partners. Traditionally, the principle common concern of PCTs and CDRPs has been the commissioning of services for the treatment of drug misusers, but at the time data were being sought for the research the PCTs still had reservations about releasing patient data for community safety analysis. Instead, hospital episodes data provided for the 2004 Crime Audit and spanning the three years from 2001/02 to 2004/04 were analysed.

Each hospital episode record contains information about the patient, their home address and a diagnosis of their (sometimes multiple) health problems using 4-digit ICD-10 codes (International Statistical Classification of Diseases and Related Health Problems). To identify possible uses of the hospital episodes data a number of data summaries were generated. To begin with, the 12,427 different 4-digit ICD-10 diagnosis codes were considered for inclusion in a subset of codes that might be useful from a community safety perspective. Eventually, 1,146 diagnosis codes were included in this subset and each was then classified into one of 14 community safety 'super-classes'. These community safety diagnosis codes were then matched with the diagnosis codes related to each hospital episode in the dataset. Using the data from 2003/04, the episodes were then queried to determine how many fell into each of the 14 community safety super-classes. The list of the 14 super-classes and an indication of the ICD-10 codes within them is given below (Table 4.18), along with the frequency of episodes in 2003/04.

Super-Class	Frequency 2003/04	Example of types of disease of health problem
Alcohol	3132	Mental and behavioural disorders due to use of alcohol Findings of drugs and other substances not normally found in blood - alcohol Accidental poisoning by and exposure to alcohol Evidence of alcohol involved determined by level of intoxification
Drugs	2013	Mental and behavioural disorders due to use opiods, cannabinoids, cocaine, volatile solvents, etc. Poisoning by narcotics and psychodysleptics (hallucinogens) Findings of drugs and other substances not normally found in blood - narcotics
Assaults	969	Assault by bodily force, blunt object, sharp object, etc. Hit struck kicked twisted bitten/scratched by another person
Road Traffic Accidents	846	(numerous)
Firearms	28	Handgun, rifle, shotgun and larger firearm discharge Assault by handgun, rifle, shotgun and larger firearm discharge Handgun, rifle shotgun and larger firearm discharge undetermined intent
Fireworks	9	Discharge of firework
Aggression	7	Symptoms and signs involving emotional state - physical violence
Sexual Assault	5	Sexual assault by bodily force Problems related to negative life events in childhood - alleged sexual abuse
Neglect	2	Maltreatment syndromes Neglect and abandonment
Abuse	0	Maltreatment syndromes - physical, sexual, psychological abuse
Police/legal	0	Legal intervention involving firearm discharge Legal intervention involving gas
Pyromania	0	Pathological fire-setting [pyromania]
Kleptomania	0	Pathological stealing [kleptomania]
Paedophilia	0	Disorders of sexual preference - paedophilia

Table 4.18: Classification of ICD-10 codes into community safety super-classes

Although more dialogue with PCT clinicians needs to be developed over the medium to long-term, an attempt has been made below to summarise some of the genuine and perceived problems associated with using this type of data for community safety.

- The full range of diagnosis codes pertinent to community safety are probably not being used fully or consistently by clinicians. This is not a fault of the data but probably a feature of the priorities of primary care, i.e. to treat patients, not interrogate them.
- Patients may not want to disclose how an injury was sustained, especially if it was sustained as a result of criminal or disorderly behaviour (either as the victim or offender).
- Intoxification with alcohol or other alcohol related problems may not have contributed to the patient's condition and may well pose no threat to the patient or the community.
- There are some (not many) duplicate records, especially where a single hospital visit may have involved a number of different procedures, possibly for different problems. Follow up visits to hospital for one injury also pose a problem. No reliable method for removing these types of duplicate record could be determined, so this analysis still includes them.
- Selecting a subset of diagnosis codes that are pertinent to community safety has been done by researchers who are not clinically trained. If this dataset is to be used more often in the future then it is recommended that the choosing of appropriate diagnosis codes is done in collaboration with clinical staff.
- Although the postcode matching rate is higher than for many other datasets used in the 2004 Crime Audit, some problems need to be resolved to ensure reliable and complete geocoding.

Some drug addiction treatment was obtained from the Leeds Addiction Unit, again as part of the 2004 Crime Audit process. To some extent the data are more a reflection of local agencies' capacity to treat people rather than an accurate indication of the number of people that may have needed treatment. Furthermore, the dataset does not contain information about all drugs users, only those that are referred, or self-refer, for treatment. The data that were obtained contained records of people who had been through treatment at some time during 2003/04.

Unfortunately, inconsistencies with the way in which patient home postcodes were recorded meant that geocoding could only be executed reliably for postal sector geography and even then, only 80.7% of cases, or 2289 records could be properly matched. There were also problems with the completeness of the dataset but manual cleansing created reliable records of the primary and secondary addiction types, the patient's age, gender and ethnicity and whether the patient has been referred to a treatment agency in the past. The ratio of male to female patients was 2.5 to 1 and approximately 60% of patients were between 22 and 34 years old. The ethnicity breakdown of the patient population was very similar to the population structure in Leeds as a whole, with whites just slightly over represented and Asians slightly underrepresented. Table 4.19 shows a breakdown of the primary and secondary addiction types. It has to be remembered that this breakdown is based on a snapshot of patients at one moment in time. It is not possible to know for certain whether the relatively low numbers of people in treatment for crack adiction (or any other drug) is evidence of the scale of the crack problem or the ability of users to manage their addiction without recourse to treatment agencies.

4.6.7 Selection of Health Variables

Two variables were selected for inclusion in the classification (Table 4.20). Both are experimental in the sense that this type of data has not been used previously for community safety analysis within

	Importance of Addiction			
	Primary		Secondary	
Drug Type	Patients	%	Patients	%
Heroin	2385	84.2	11	0.4
Cannabis	165	5.8	34	1.2
Amphetamine	57	2.0	6	0.2
Cocaine	42	1.5	37	1.3
Crack	35	1.2	83	2.9
Methadone	35	1.2	0	0
Solvents	17	0.6	1	0.03
MDMA	10	0.4	6	0.2
Alcohol	4	3.0	22	0.8
Other	84	0.1	16	0.6

Table 4.19: Primary and secondary addiction types for patients in treatment, 2001/02 to 2003/04.

Category	Id	Variable	Database
Health	AL01	Hospital episodes involving alcohol	Hospital Episodes
Health	DR01	Hospital episodes involving narcotics	Hospital Episodes

Table 4.20: Variables identified from Leeds PCTs.

Leeds, and more research in collaboration with clinicians needs to be undertaken to assess the suitability and accuracy of these indicators. The narcotics variable (DR01) is probably the more reliable of the two, as all narcotics use is against the law, even if it is not an aggravating factor in other types of offence. It could be argued that the presence of alcohol on the patient record might be more significant if alongside other diagnosis codes such as those pertaining to assaults or woundings, but analysis suggests these latter codes are being underused.

4.7 Concluding Remarks

The survey of practitioners regarding variable selection provided confirmation of the perceived usefulness of some of the indicators of community safety identified from the survey of existing literature, as well as suggesting original possibilities. In addition, local data sharing and analysis projects provided access to variables that were being used for community safety analysis in Leeds for the first time. In most cases, variables were chosen to reflect directly aspects of neighbourhoods that the criminology literature suggest might be important. On occasions, however, new and interesting data would be sourced and arguments for inclusion sought retrospectively.

The national census still provides a large number of variables for the classification, although attempts were made to ensure that alternatives were used where possible, even if modelling techniques were required to harmonise geographies. In future it might be worth trying to explore the potential of the council tax register, housing databases and local land gazetteers to provide alternatives to decennial census variables. For this research there was insufficient time or scope for negotiation to explore these possibilities.

The use of information from the PLASC and pupil level information of school exclusions provided more opportunities for capturing information about young people than is possible with the national census and it is possible that future analysis of truancy data might yield useful variables. By contrast to education, the health variables capturing alcohol and narcotics-related hospital episodes need to be developed further with clinicians to ensure that issues with data coding reliability are properly understood and that the diagnosis codes used do accurately reflect problems relevant to community safety. Furthermore, this last issue raises further questions about the scope of community safety and whether variables should be included that reflect geographies of accidental incidents, such as accidental fires and road traffic collisions.

Overall, identifying local data sources presented the greatest number of problems and provides the most scope for further work. Some of the problems were linked with the inability of agencies to provide suitable electronic records, while at other times there were problems with misunderstandings or conflicts with regards to data protection and confidentiality. Access to data improved once involvement with other CDRP projects conferred a degree of insider status, yet, it is argued that this is perhaps symptomatic of data sharing problems in general given that even large projects such as the Community Safety Annexe to Leeds Statistics have failed to be properly developed beyond their initial brief and have not been maintained or regularly updated. A complete list of the variables described in this Chapter is provided in Appendix B.

Chapter 5

Classification Design

5.1 Introduction

The principle aim when designing a classification is to group together objects that are similar while at the same time trying to ensure that the individual groups are distinct from one another (Cormack, 1971). For an area-based classification such as the LCCS, the objects are neighbourhoods and the result of the grouping exercise is referred to as a partition. Thus, with a set of neighbourhood variables identified and values for the different Sub-Community Areas calculated, the next stage in the research is to use these data to create a partition of the neighbourhoods in Leeds. The final classification becomes complete once the partition has been interpreted and the groups given some sort 'profile' or 'portrait', a process that is discussed in Chapter 6.

If a relatively few neighbourhoods were being defined by just a few variables, then a partition might be created intuitively, by hand. Indeed, classifying and grouping objects is a process that humans develop from a young age, and classification and taxonomies of one sort or another can be found in many aspects of our daily lives (Everitt et al., 2001). Yet for this research, a large number of neighbourhoods are being considered, and a fairly large number of data variables have been gathered together. As a consequence, the problem lends itself to group of numerical techniques for creating partitions, referred to as clustering methods. Each neighbourhood object is represented mathematically by a point in a multi-dimensional space, where each dimension reflects a different variable. The values of these variables along each of the dimensions fixes the location of each object (Anderberg, 1973).

The Chapter begins with a review of some clustering methods and considers some of their individual strengths and weaknesses. Consideration is then given to variable selection and weighting issues, ensuring that the variables identified in the previous chapter are likely to contribute positively to a partition and not produce unwanted 'noise' or mask the importance of other variables. Experiments using different clustering methods are then presented along with a discussion comparing the resulting partitions.

Henceforth, cluster analysis is used to describe the collection of linked, but distinct steps that begin with identifying data through to the implementation of a final classification - in effect, the work discussed across Chapters 4, 5 and 6. The term clustering methods is limited to describing the numerical routines for creating groupings of object, which is just one step in the production of a classification.

5.2 Cluster Analysis

"Cluster analysis is the art of finding groups in data" (Kaufman and Rousseeuw, 1990, page 1), and is much more than simply processing a set of data using a numerical clustering algorithm (Milligan, 1996). A series of steps have to be taken to complete a successful analysis, and multiple decisions may have to be made at each stage, the impact of which may be difficult to predict (Milligan and Cooper, 1987). Inevitably, different decisions may be more or less suitable, according to the the purpose of the classification (Lorr, 1983). Furthermore, whatever the complexity of the mathematical techniques that may be used, it is the ability of the resulting classification to be useful and usable that will determine whether the cluster analysis has been successful (Davis, 1986).

Because multiple decisions have to be made, it is argued that it is helpful to have an understanding of the approach generally adopted to cluster analysis, as well as an appreciation of some of the specific numerical clustering methods that have been developed to help produce solutions. Several comprehensive and frequently cited reviews exist elsewhere (Kaufman and Rousseeuw, 1990; Everitt et al., 2001; Gordon, 1999; Lorr, 1983, for example), so what follows in this section is a summary of the general cluster analysis process, based upon Everitt's ((2001) reading of Milligan (1996), phrased to be applicable to the creation of a neighbourhood classification. This is followed by a discussion regarding the clustering methods and techniques used to design the LCCS.

5.2.1 A Typical Cluster Analysis

1. *Neighbourhoods to cluster.* For some applications, this may be a sample of a larger population. For the LCCS, however, complete coverage of the district is required, so all neighbourhoods are selected. Some neighbourhoods may contribute sufficient outlier values for variables that they may be better removed from the analysis and treated separately.

2. *Select neighbourhood variables.* Variables should only be included if there is a good reason to believe they will help define clusters. Variables should be removed if they might produce unwanted noise or mask other variables.

3. *Variable standardisation.* This is necessary for most clustering methods. Different techniques may be used, but range standardisation showed good results in simulations by Milligan and Cooper (1988) and might be preferable over techniques based upon standard deviations.

4. *Proximity measure.* Where continuous data are being used, the most commonly used proximity measure is *Euclidean distance,* where high values represent greater dissimilarity. A Pearson correlation is an alternative that is sometimes used, although in this case low values represent greater dissimilarity

5. *Clustering method.* Choice may be limited according to the type and/or amount of data being used. Whatever the case, the method should be chosen according to its ability to find the type of partition required of the classification. The generation of a number of partitions using different methods may provide helpful confirmatory evidence if the nature of the true partition is not known.

6. *Number of clusters.* This is one of the most difficult decisions to make. Where stopping rules suggest different numbers, the highest should be taken for safety. At the same time, the

classification must be usable if it is to be of value and end-user expectations may have to be considered. It is also possible that there are no clusters present.

7. *Replication, testing and interpretation.* Re-running the the analysis, perhaps using different clustering methods, to make sure the same or a similar solution is found on all occasions. Testing may involve mathematically analysing the extent to which the variance of variable values is reduced within clusters. Practical tests of the classification also need to be undertaken to help determine its usefulness, and this may require specialist knowledge of the application domain. Interpretation is critical if the groupings are to be properly understood.

5.2.2 Clustering Methods

There are many possible clustering methods from which to choose, although some represent only small modifications on existing methods. If methods designed for very specific applications (which may be used rarely) are discounted, then the most common methods can be grouped into two broad types: hierarchical methods and partitioning methods.

Hierarchical Methods

With hierarchical methods, the data are not grouped into a particular number of clusters in a single step. Instead, the classification consists of a series of partitions, ranging from a single cluster (containing all objects), to n clusters where each contains a single object. Constructing this series can be done using agglomerative methods (the most widely used) or divisive methods. Agglomerative methods operate top-down and start with each object apart; and in each step two clusters are merged until only one is left. Divisive methods start when all objects are together; and in each step a cluster is split up. In the simplest hierarchical methods, *single linkage*, the decision to merge (in agglomerative methods) is based upon the smallest distance between pairs of objects in different groups. For example (taken from Everitt et al., 2001), the following distance matrix, D_1, shows the distance between 5 objects to be clustered:

$$\mathbf{D_1} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0.0 & & & & \\ 2.0 & 0.0 & & & \\ 6.0 & 5.0 & 0.0 & & \\ 10.0 & 9.0 & 4.0 & 0.0 & \\ 9.0 & 8.0 & 5.0 & 3.0 & 0.0 \end{bmatrix}$$

The smallest non-zero distance is that between objects 1 and 2, so these are merged to form a two-member cluster. The distances between this new cluster and the other objects are then obtained as

$$d_{(12)3} = min[d_{13}, d_{23}] = d_{23} = 5.0$$
$$d_{(12)4} = min[d_{14}, d_{24}] = d_{24} = 9.0$$
$$d_{(12)5} = min[d_{15}, d_{25}] = d_{25} = 8.0$$

The new distance matrix, D_2, with the clustered (12) objects looks thus:

$$
\mathbf{D}_2 = \begin{array}{c} (12) \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0.0 & & & \\ 5.0 & 0.0 & & \\ 9.0 & 4.0 & 0.0 & \\ 8.0 & 5.0 & 3.0 & 0.0 \end{bmatrix}
$$

The next smallest distance would see the merging of objects 4 and 5, and so on. Once all steps have been completed, the decisions must be made about which of the multiple solutions bests fits the application.

In addition to single linkage methods there are a variety of alternatives, some with characteristics that produce quite distinct types of clusters. Everitt (2001) provides a guide to the different methods and their characteristics. Ward's method has often been used for classification of social data (ONS, 1999; ONS, 2003; Vickers et al., 2003, e.g.), for example, and of Ward's method, Everitt remarks, "tends to find same size, spherical clusters; sensitive to outliers" (Everitt et al., 2001, page 62).

These kinds of assessment of particular clustering methods can be used to inform decisions about the best approach for a particular application. Sensitivity to outliers might be one reason why a hierarchical method using Ward's algorithm might not be suitable for census data. Alternatively, hierarchical methods might be considered undesirable because, as Kauffman and Rousseeuw put it,"a hierarchical method suffers from the defect that it can never repair what was done in a previous step" (Kaufman and Rousseeuw, 1990, page 55). This refers to the problem that once objects or clusters have been joined to create new clusters, they can never be separated, regardless of whether successive steps might mean it would make sense to do so. This weakness, sensitivity to outliers,and the fact that the LCCS does not require a nested set (hierarchy) of classifications led to the decision not to pursue any experiments using hierarchical clustering methods.

k-means

Also known as *optimisation clustering* techniques or *iterative relocation* algorithms, partitioning methods assign cases to form a predetermined number of groups, the number usually being referred to as k.

The most popular of these methods, which include k-means, are based on the construction of central points. The objective of these methods is to minimise the within-cluster variability. This is done by moving objects from one cluster to another and testing the impact this has by measuring the sum of squared deviations within each cluster. Typically, the steps of one of these types of algorithm would be thus:

1. Create an initial partition of the objects into k non-empty subsets, perhaps by randon assignment.

2. Calculate the effect on the clustering criterion (for example, to minimise the Euclidean sum of squared deviations of cluster objects from the cluster mean) produced by moving each object from its own to another group.

3. Make the change which results in the greatest improvement according to the clustering criterion.

4. Repeat step 2 and 3 until no move of a single object leads to any improvement.

Despite widespread and continuing use, k-means has weaknesses which make it unsuitable for some applications. Firstly, it has been shown that k-means is particularly sensitive to outliers in the data (Kaufman and Rousseeuw, 1990). The order in which the objects are stored may also have a bearing on the final partition (ibid). K-means may also be problematic when clusters have widely differing sizes or convex shapes. Random seed selection can also mean that should a spurious object be chosen as a seed then no other objects will get clustered around it; leading to one or more clusters with just one or very few members. It is possible to solve this type of problem by supplying the k-means algorithm with initial seeds calculated as being likely to represent different and distinct groups in the data.

For this research, the ClustanGraphics sofware package, designed by Wishart (1987), was used to experiment with the k-means technique. ClustanGraphics provides the user with a choice of using random cluster seeds or using points identified *a priori* as being representative of some structure in the data. ClustanGraphics also allows a limited number of atypical, or outlier, objects to be placed in a separate 'residue' class during cluster analysis and considered separately (or just discarded) once the partition has been made. These features are not found in many software implementations of k-means.

Partitioning Around Medoids (PAM)

Partitioning Around Medoids (PAM) was developed by Kaufman and Rousseeuw (1990). The approach is designed to identify the k most representative objects, called medoids (or centrotypes in other literature) within the samples of objects. To form the k clusters, the non-selected objects are each associated with the medoid with which they have the greatest similarity, calculated by minimising the sum of the dissimilarities. The fact that the sum of dissimilarities is used instead of the sum of the dissimilarities squared (as with k-means) is important as it helps reduce the sensitivity of PAM to outliers in the data.

Details of the algorithm used by PAM to determine the optimum set of medoids can be found elsewhere (see Kaufman and Rousseeuw, 1990; Ng and Han, 2002). PAM can be expensive computationally for large datasets although a modification of PAM, called CLARA, gets around this problem by using a sample of the total number of objects. The benefits of using PAM are usually cited when comparisons are being made with k-means. Thus, the advantages of PAM are that it is less susceptible to the effects of outliers in the data and it is not affected by the order in which objects are presented. The nature of the algorithm also negates the problem faced by k-means of identifying initial cluster centres - since PAM's main aim is to find the most representative cluster centres, and group other objects about these.

A number of metrics and visualisation aids to support PAM have also been developed. The silhouette plot (Kaufman and Rousseeuw, 1990), for example, can be used to help determine the optimum number of clusters and the goodness of fit of the individual clusters. These tools and the PAM algorithm were experimented with using the *cluster* package available for the R statistical computing environment (R Development Core Team, 2006).

Fuzzy c-means

All of the clustering methods described so far have produced what is referred to as a 'hard clustering'. This means that clear-cut decisions are made about which cluster an object belongs to. However, this may sometimes be a considerable oversimplification of the structure in the data and there may well be objects for whom a reliable cluster membership is much less obvious (Gordon, 1999). Fuzzy clustering techniques are one way of handling these ambiguities.

Instead of a hard partition, fuzzy clustering produces a set of membership coefficients for each object to indicate the degree of belonging to each cluster. These coefficients are usually calculated so that they sum to 1. Thus, fuzzy partitions allow the user to see the degree to which an object belongs to each and every cluster. As will be shown later in the chapter, some neighbourhoods appear to defy being grouped in a consistent way in different experiments. The membership coefficients from a fuzzy partition can provide additional information to help understand this problem.

The authors of PAM also designed a fuzzy clustering method, called FANNY (Kaufman and Rousseeuw, 1990). In a number of ways, this technique is similar to the fuzzy c-means (also referred to as fuzzy k-means) developed by Bezdek (1987). The objective functions of each are identical except with respect to the squaring of the distances (fuzzy c-means does, FANNY, like PAM, does not). Their use of iterative relocation algorithms contrasts to earlier work on fuzzy techniques (Ruspini, 1969) that had links with aspects of multivariate mixture models.

For this research, practical considerations dictated the use of c-means rather than FANNY. The latter is implemented in R and S-Plus, but both versions of the contributed packages appear to have problems. Obtaining the original Fortran version from the authors was not an option, as this has a limit of 100 objects. The version of c-means used was that ported to the R environment (R Development Core Team, 2006) and included in the *e1071* package.

The c-means routine demands the setting of a number of parameters. Most of these choices are straight forward, but the setting of an appropriate 'fuzziness index', m, is less clear. A value of $m = 1$ produces a hard partition, i.e. for each object there is one coefficient of 1 and the rest are zero. As m is increased, so the size of the largest coefficient comes down in relation to the coefficients for the other clusters and thus the partition becomes 'more fuzzy'. Although some techniques have been designed to help find optimal values for k and m in fuzzy partitions (McBratney and Moore, 1985), many studies rely on experimentation to set m, and often decide on a value of $m = 2$ (Gordon, 1999, page 112).

5.3 Variable Selection and Weighting

With a number of clustering methods identified, attention can turn to reviewing in more detail the variables identified in the previous chapter. The danger of continuing to pursue the use of variables that have little discriminating power, or worse, which have no relationship with aspects of community safety, is that the cluster structure being sought could be clouded or destroyed. Some builders of classifications might suggest that a 'throw everything into the pot' approach is valid, but serious objections to this notion have been raised (Fowlkes et al., 1987; Kaufman and Rousseeuw, 1990; Gordon, 1999; Openshaw and Wymer, 1995; Everitt et al., 2001).

A number of studies have looked at the effects of masking (or random noise) variables on cluster recovery. Milligan (1980), for example, looked at eleven agglomerative hierarchical algorithms and four non-hierarchical centroid sorting procedures and simulated the effects of different types of error in data being partitioned. It was shown that the presence of random noise variables had by far the greatest adverse influence on the ability of all of the methods to recover the cluster structure. Moreover, the negative effect of random variables actually increased when using computed, non-random seeds for the cluster analysis. Similar results have been found in other studies (DeSarbo and Mahajan, 1984; Fowlkes et al., 1987, for example,).

5.3.1 Mathematical Techniques for Determining Variable Weights

A possible approach to the problem of masking or noise variables is to identify them prior to clustering using some mathematical technique. Once identified, the problematic variables can either be given a low weighting (presuming the clustering method method can cope with with variable weights) or given a weight of 0 - that is, removed from the analysis completely.

While there is no universal agreement about which is the best means of determining variable weighting (Everitt et al., 2001) a number of techniques have been proposed. Some care needs to be taken when assessing the results of such research. This is because there is a tendency for experiments to have used Monte Carlo simulations (see Milligan, 1989; Makarenkov and Legendre, 2001, for examples) based on synthetic test datasets which have may little in common with the types of data being used to define the neighbourhoods of Leeds. Reviews of the literature on variable weighting often begin with the work of DeSarbo et al. (1984) and their SYNCLUS programme. This solves for both variable weights and produces K-means clustering. Fowlkes et al. (1987) also proposed a method, called (in a subsequent paper) FGK, for selecting binary weights. In a later study, Gnanadesikan et al. (1995) compared the FGK procedure to De Soete's OVWTRE (1988) and to SYNCLUS. The analysis determined that the FGK forward selection procedure performed reasonably well compared to its competitors. However, subsequent to the FGK algorithm, Carmone, Kara and Maxwell (1999) proposed their own variable subset selection method based on Hubert and Arabie's (1985) adjusted (or corrected) Rand Index. Their method, called HINoV, was designed for partitioning using continuous variables and is described as a heuristic method based upon the Adjusted Rand Index. These authors conducted a series of Monte Carlo simulations using synthetic data with noise of various kinds added, including masking variables. The result indicated that variables selected using the HINoV procedure outperformed the all-variable cases in 70 out of 72 different runs.

Experiments with HINoV, which aims to identify noisy variables, as opposed to error-perturbed or outlier strewn variables, were conducted on the LCCS and are reported on later (Section 5.4.8), but the algorithm is described here,

1. Perform k-means cluster analysis for k clusters, one variable at a time, for each of the L variables. This yields an nxL matrix of cluster membership, where n is the number of objects.
2. Calculate Corrected Rand (Hubert and Arabie, 1985) statistics r_ij for all pairs of L variables and place these in an LxL matrix (PARIM), leaving out comparisons of a variable with itself.
3. Sum the rows of PARIM to give a total pairwise Corrected Rand index (TOPRI) for each variable.
4. Rank TOPRI values. Low values indicate potentially noisy variables which can be eliminated from further analysis. (The authors also recommend producing a scree plot of the ranked TOPRI values and identifying points of inflection, although the utility of this step was not explained.)

Using a similar technique, Makarenkov and Legendre (2001) explored how least-squares optimisation could be applied to the analysis of two-way data matrices in the context of ultrametric and additive tree construction as well as K-means partitioning, using their OVW software.

5.3.2 A Mixed-method Approach to Variable Weighting

For the classification of neighbourhoods in Leeds, tests on partitions using OVW and HINoV were undertaken and are discussed later (Section 5.4.8). However, in the first instance, it was decided to forgo these more sophisticated techniques and adopt a simpler approach to testing for variable fitness. Thus the aim was to find some middle ground between using every variable available (not desirable) and complex mathematical techniques such as those mentioned above (potentially difficult to apply and interpret).

Moreover, chroniclers of cluster analysis give support to the idea that the investigator ought to be making subjective decisions based on experience and understanding of the subject domain (Everitt et al., 2001; Kaufman and Rousseeuw, 1990; Lorr, 1983). These subjective decisions may be based upon personal experience and knowledge, or that of others (Openshaw and Wymer, 1995). However, there have also been warnings of hermeneutic problems (Sneath and Sokal, 1973) that remind us that selecting variables based on subjective judgments of what is important might simply reflect existing classifications of the data.

After consideration, it was decided to conduct a series of numerical and statistical tests to,

- identify outliers,
- identify strong correlation between variables,
- consider the data distributions for variables by measuring standard deviation, skew and kurtosis, and
- consider the extent to which variables figured strongly in principal components analysis.

The results of these tests are discussed below.

5.3.3 Identifying Outliers

As clustering methods may be sensitive to the presence of outliers in the data, tests can be made to identify locations where outliers are evident for several variables. Chow (1998) used a limit of three standard deviations from the mean to identify outliers which were excluded from the cluster analysis.

Using the limit of three standard deviations, Figure 5.1 identifies those neighbourhoods which contribute most to outliers in the data. Student neighbourhoods figure quite highly as well as areas close to city centre. Neighbourhoods where commerce and industry are the main types of land use also represent outliers for many variables. Overall, the main source of outliers is the city centre itself (SCA 23.02), which has values over three standard deviations from the mean for 19 of the 67 variables being considered. SCA 54.01, which covers the commercial zone to the south-east of the city centre and the mixed commerce/residential area of northern Hunslet might also be considered suitable for removal, having values over 3 standard deviations from the mean for 15 of the 67 variables. Sub-community area 18.02, which covers Lincoln Green and Sheepscar, has 17 variables with values over three standard deviations from the mean, but the nature of land use here - residential and primary heath care (St.James's Hospital) - is somewhat different from the previous two cases.

The city centre is widely regarded by Leeds Community Safety Partnership members as a unique place. It requires special policing and patrolling and accounts for a large share of total crime in Leeds. The specificities of the opportunities for offending are unlike most other parts of the city and

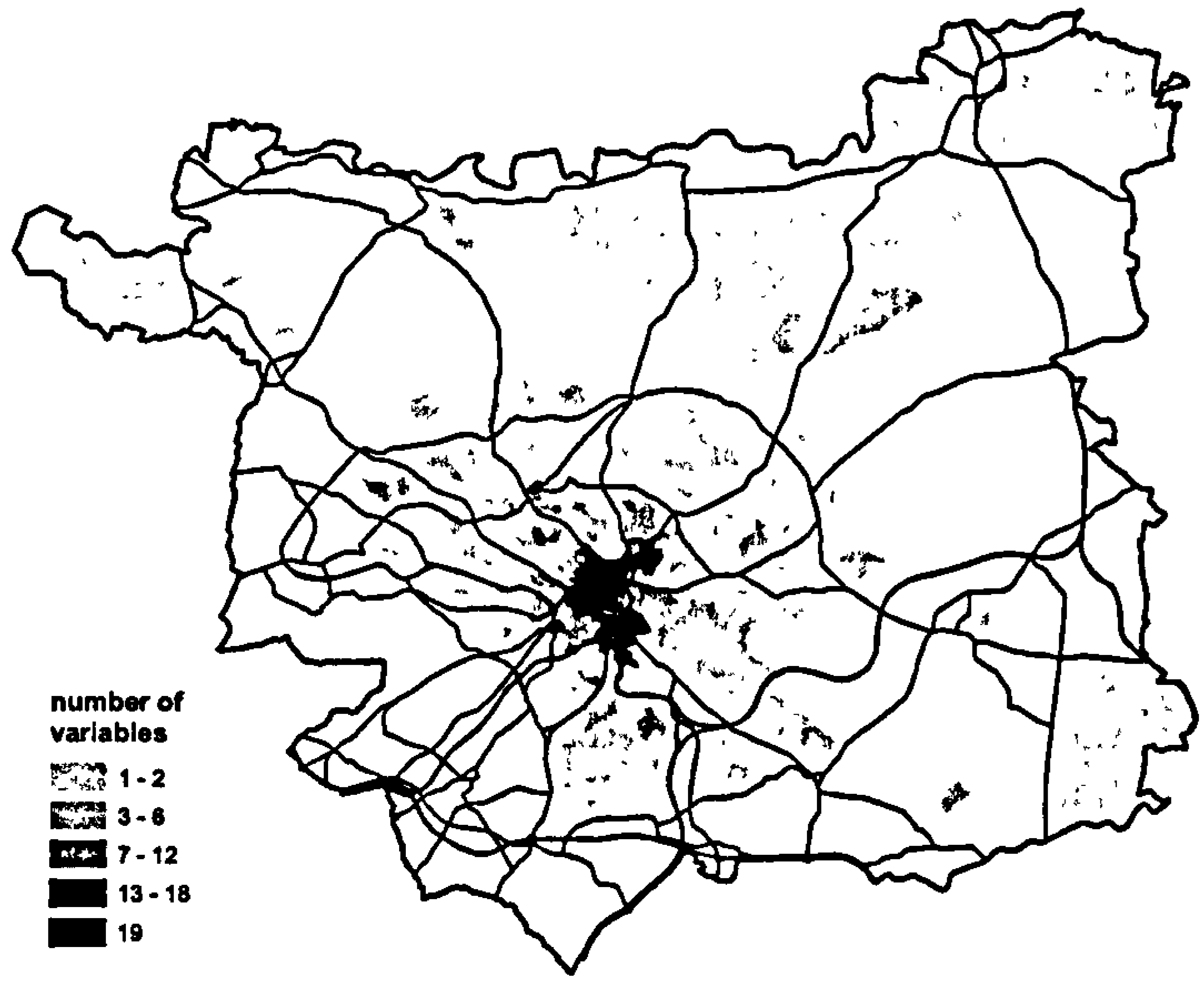


Figure 5.1: Map of the number of variables for which sub-community areas (SCAs) have values over three standard deviations from the mean.

the changing nature of the population throughout the day and week make estimating true levels of risk very problematic. Thus, at this early stage, it was decided to exclude the first two SCAs (23.02 and 54.01) mentioned above from the cluster analysis, in effect placing them in a cluster by themselves.

It was decided to leave Lincoln Green in the classification. This was partly because the land use here is subtly different from the other two neighbourhoods and because it falls within an area that has been subject of both Single Regeneration Budget (SRB) and Neighbourhood Renewal Fund (NRF) regeneration activity. As a consequence of the outlier exclusions, subsequent tests were conducted on 477 objects, and not the full 479 sub-community areas. Had this action not been taken at this juncture, all subsequent descriptive statistics and correlation coefficients would have been affected.

Another approach to the problem of outliers is to assign cases outside a specified distance from any cluster centroid to a special group. These can then be re-admitted to the analysis if, on a later pass, the cases comes back within the inclusion limit. Once the partition has been generated the cases in the outlier group can considered individually and assigned to clusters if desired, based on their distance from the cluster centroids. A number of software packages implement this technique, including ClustanGraphics (Wishart, 1987).

Even with the city-centre neighbourhoods remove to a 'residue group' outliers remain and clustering methods must be chosen with this in mind.

5.3.4 Correlation Between Variables

It is helpful to determine if a phenomenon is being measured more than once. This can happen when two or more variables are essentially measuring the same thing, for example, 'people drawing a state pension' and 'people over 65'. To include both variables in the cluster analysis effectively double-weights the importance of being retired. In a correlation matrix, these variables should show up as having a high positive correlation coefficient, enabling them to be identified and one variable removed from the dataset - unless the double weighting is desired. Similarly, some variable pairs

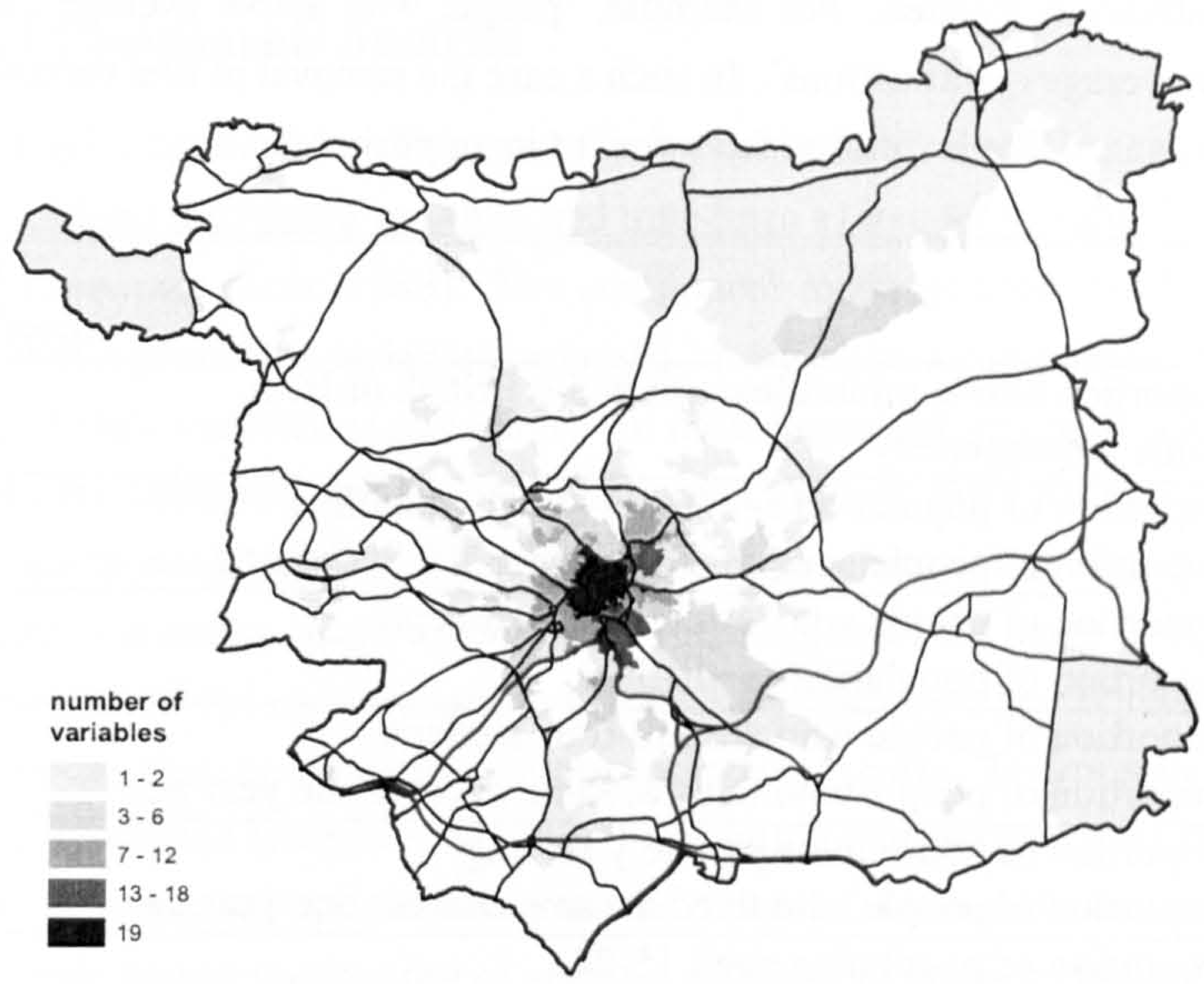


Figure 5.1: Map of the number of variables for which sub-community areas (SCAs) have values over three standard deviations from the mean.

the changing nature of the population throughout the day and week make estimating true levels of risk very problematic. Thus, at this early stage, it was decided to exclude the first two SCAs (23.02 and 54.01) mentioned above from the cluster analysis, in effect placing them in a cluster by themselves.

It was decided to leave Lincoln Green in the classification. This was partly because the land use here is subtly different from the other two neighbourhoods and because it falls within an area that has been subject of both Single Regeneration Budget (SRB) and Neighbourhood Renewal Fund (NRF) regeneration activity. As a consequence of the outlier exclusions, subsequent tests were conducted on 477 objects, and not the full 479 sub-community areas. Had this action not been taken at this juncture, all subsequent descriptive statistics and correlation coefficients would have been affected.

Another approach to the problem of outliers is to assign cases outside a specified distance from any cluster centroid to a special group. These can then be re-admitted to the analysis if, on a later pass, the cases comes back within the inclusion limit. Once the partition has been generated the cases in the outlier group can considered individually and assigned to clusters if desired, based on their distance from the cluster centroids. A number of software packages implement this technique, including ClustanGraphics (Wishart, 1987).

Even with the city-centre neighbourhoods remove to a 'residue group' outliers remain and clustering methods must be chosen with this in mind.

5.3.4 Correlation Between Variables

It is helpful to determine if a phenomenon is being measured more than once. This can happen when two or more variables are essentially measuring the same thing, for example, 'people drawing a state pension' and 'people over 65'. To include both variables in the cluster analysis effectively double-weights the importance of being retired. In a correlation matrix, these variables should show up as having a high positive correlation coefficient, enabling them to be identified and one variable removed from the dataset - unless the double weighting is desired. Similarly, some variable pairs are inherently negatively correlated. For example, 'people with above average qualifications' and 'people with below average qualifications'. In such a case the removal of one variable does not result in information loss as its value is inferred from that of its opposite.

Variables		Correlation coefficient ¿ .9
ET01	Proportion ethnic minorities: not white British or Irish	0.945
ET02	Ethnic heterogeneity	
QU01	Proportion of population aged 16-74 lowly qualified	0.943
SC67	Proportion of people aged 16-74 who are NS-SeC 6 or 7	
SC09	Proportion of people who are full-time students	0,980
AG03	Proportion of population aged 15-24	
SC09	Proportion of people who are full-time students	-0.954
MG02	Proportion of people who lived at same address one year ago	
TE03	Proportion of households privately renting	-0.900
MG02	Proportion of people who lived at same address one year ago	
AG03	Proportion of population aged 15-24	-0.960
MG02	Proportion of people who lived at same address one year ago	

Table 5.1: Highly correlated variables

On some occasions, a high positive (or negative) correlation between variables may not be a good justification for removal. For example, 'full-time students' and 'people aged 18-24' are usually highly correlated yet the former are an important community group in their own right and to remove 'people aged 18-24' might distort the age structure in those parts of the district which are not student areas (the majority of Leeds).

A correlation matrix of variables from Chapter 4 was calculated and examined, and a number of highly correlated variable pairs were identified (Table 5.1).

5.3.5 Descriptive Statistics

Variables which have peculiar frequency distributions or which are very kurtose should be examined (Openshaw and Wymer, 1995). Variables with low standard deviation may add little to cluster definition, and may indeed act to ameliorate the effect of more important variables. Conversely, variables with a high standard deviation may contain outliers and the frequency distribution of values needs to be considered by referring to measures of skew and kurtosis. A number of variables with potential problems were identified (Table 5.2).

		Problems with,		
	Variable	S.D.	Skew	Kurtosis
ED03	Proportion of pupils permanently excluded from school	•	•	•
JB01	Difference in workplace jobs from 1998 to 2002	•		
MI02	Needles collected per 1000 population	•	•	•
FI02	Fire hoax calls attended per 1000 population	•	•	•
ED04	Proportion of GCSE students only getting 0,1 or 2 passes	•		

Table 5.2: Problem variables identified by descriptive statistics

5.3.6 Principal Components Analysis

In some cluster analysis applications, Principal Components Analysis (PCA) may be encouraged to remove the effects of inter-correlated variables and transform a large set of variables down to a smaller set of principal components (Lorr, 1983). The component loadings produced by PCA indicate the amount of variance of a particular variable accounted for by a given component. Variables that have a large proportion of their variance explained by the most important principal components are more likely to exert a strong influence on the classification. Common rules for determining how many principal components to consider include selecting all components with an Eigenvalue of 1.0 or over (known as the Kaiser's criterion (Cattell, 1966)), or selecting however many components it takes to account for the first 90% of the variance within the dataset (Openshaw and Wymer, 1995).

Cluster analyses can then be conducted on the component scores. Interpreting partitions created in this way may be simplified (Everitt et al., 2001), assuming that the principal components can be interpreted clearly. However, components may not be directly interpretable and the orthonormalisation effect of PCA may distort the identity of specific variables of interest (Rapkin and Luke, 1993). Furthermore, the correlation coefficients used by PCA measure linear relationships and are sensitive to non-normality, which may make them unsuitable for use on neighbourhood-type data (Openshaw and Wymer, 1995).

On the basis of these arguments, it was decided not to conduct cluster analysis using principal component scores. However, PCA was used to assess the likely discriminatory power of each variable prior to clustering. Running a PCA that included all the variables produced 11 components with an Eigenvalue of over 1.0, which between them accounted for 76.4% of the variance in the data. A Varimax rotation was applied to the component loadings and variables that had very low loadings and which only featured in the less important components were identified (Table 5.3). Loadings less than 0.4 were discounted.

Variable	
JB01	Difference in workplace jobs from '98 to '02
HT06	Housing type heterogeneity
FI02	Fire hoax calls attended per 1000 population
MI02	Needles collected per 1000 population
LU01	Pub density by population
HT02	Proportion of houses that are semi-detached
ED04	Proportion of GCSE students with 0,1 or 2 passes
HT07	House size (number of rooms) heterogeneity

Table 5.3: Problem variables identified by Principal Component Analysis. Eleven components (factors) used.

5.3.7 Removal and Reworking of Variables

Based on the evidence from the tests above, the decision was taken to remove or rework a number of variables.

ET01 Proportion ethnic minorities: not white British or Irish Although there may be a higher risk of young black males being linked with the small but growing problem of gun crime, there is little else to recommend double weighting ethnicity in the classification. Areas with high ethnic population do suffer from high crime but this is due to coincident high levels of social and material deprivation, not skin colour. This variable was removed, leaving in the index of ethnic heterogeneity (ET02).

QU01 Proportion of population aged 16-74 lowly qualified In addition to being highly correlated to low social class this variable is also part of an opposite pair with QU02 (Proportion of population aged 16-64 highly qualified: 4/5), with a correlation coefficient of -0.848. This variable was removed, leaving in QU02.

SC09 Proportion of people who are full-time students Students suffer from high victimisation but the reasons are more likely to do with age, routine activities, naivety (on crime issues) and residence in poor, and poorly secured neighbourhoods, than the fact they are studying for degrees. Many of the student attributes mentioned are captured elsewhere, so this variable was removed.

JB01 Difference in workplace jobs from '98 to '02 Contemporary debates around crime reduction through area regeneration draw attention to the importance of improving employment opportunities for the long-term and young unemployed. However, while this variable captures changes in job opportunities to some extent, it does not reflect the nature of the jobs created or whether they are suitable for local needs, or even if there is a local demand. This variable was removed.

FI02 Fire hoax calls attended per 1000 population Hoax calls to the fire service are serious but not in sufficient numbers (n=658) to produce a frequency distribution likely to discriminate between neighbourhoods at this geographic scale.

MI02 Needles collected per 1000 population The frequency distribution of needle collection in neighbourhoods across Leeds is highly skewed by outliers. Furthermore, some of these outliers are aggregations from a wider area and represent needles collected in safe disposal bins situated in various premises and organisations around the city which the needle collection team empty periodically. This variable was removed.

ED04 Proportion of GCSE students with 0,1 or 2 passes Low PCA loadings and low standard deviation highlight problems with this variable. The cause of the problem is low numbers of pupils (n=443) falling into this category. The variable was removed. GCSE attainment is still captured by ED05 (Average number of GCSE passes A* to C).

ED03 Proportion of pupils permanently excluded from school This variable suffers from similar problems to the last, with only 143 cases of permanent exclusions across the 477 sub-community areas. This variable was removed.

HT07 House size (number of rooms) heterogeneity This variable was collected specifically to reflect areas which might be at greater risk of near-repeat burglaries. As such, it is attempting to measure the same as HT06 (Housing type heterogeneity). The HT07 variable had a factor of less than 0.4 in all of the principal components, and was removed.

Housing types These were highlighted as problematic by the PCA. The decision was taken to simplify their classification by merging some types and removing others. HT04 and HT05 (both types of flats) were merged, terraced housing and detached housing were left in and HT02

(Proportion of houses that are semi-detached) was removed. The closed number set nature of these variables means that no information is lost by the removal of HT02.

5.3.8 Methods of Standardisation

The final set of variables are based upon units whose scale vary markedly and it is therefore necessary to standardise variables in order that are weighted equally during the clustering process. However, it might not always be necessary to standardise variables, particularly if the original scales of measurement are theoretically meaningful (Hartigan, 1975).

It is more usual for cluster analyses to have used a standardisation technique based upon standard deviations from the mean, but for this research the decision was taken to range standardise variables between the values of 0.0 and 1.0, based upon the findings of a study of different techniques by Milligan and Cooper (1988). Range standardisation was also the method used by the Office for National Statistics 1991 classification of local authorities (Walace and Denham, 1996).

The range standardisation method is defined as:

$$R_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

where x_{max} is the maximum value of x, x_{min} is the minimum value of x and R is the range standardised value.

5.4 Creating an Initial Partition

The decision was taken to create an initial partition using the PAM clustering method. This was largely because of the support metrics and tools that have been designed to be used with it, and because PAM is less sensitive to outliers than other methods. The aim was to conduct experiments and explore different possible partitions and become familiar with the data and the different ways in which neighbourhoods might be grouped. Once confidence had been gained creating partitions using PAM, the experiment plan was to apply other clustering techniques to look for levels of agreement between the partitions, and thus provide extra confidence for the final partition from which the final neighbourhood classification would be constructed.

5.4.1 Finding an Optimum Value for k

An important consideration for all partitioning methods is to decide in advance how many groups to partition the objects in to. The difficulty of the task has led to the production of a number of formal statistical tests and modelling techniques to determine an optimum value for *k*. Some of these formal techniques are based upon specific assumptions that might not be applicable to some cluster analysis problems, and a comparison of tests has highlighted those with little validity (Milligan and Cooper, 1985).

Most methods for choosing an optimum number of groups, however, are informal (Everitt et al., 2001) and often involve the plotting of some criterion value against the number of groups for a range of possible values of *k*. The reliability of these inverse scree plot methods has been tested (Lathrop and Williams, 1987), and they were used by the Office for National Statistics to classify local and health authorities in Great Britain (ONS, 1999), and others (foe example Debenham, 2003; Vickers et al., 2003).

Due to their relative ease of use, the decision was taken to employ an inverse scree plot method and plot k against the average silhouette width of all objects for a number of different partitions. This silhouette width (or more correctly, coefficient) is calculated for each object and is the difference between the average dissimilarity between the object and all the other objects in its group and the average dissimilarity of the object to all the objects in the next nearest group (Kaufman and Rousseeuw, 1990). Points of inflection on the graph are then interpreted as values for k above which the partitions become less reliable.

When the silhouette width is close to 1 then the object is considered well classified. Silhouette widths near to -1 are considered misclassified and silhouette widths close to zero suggest the object could just as likely belong to the next nearest group. Silhouette widths can be averaged for each cluster and for the entire partition. Suggestions have been made about how to interpret different silhouette values, but these are just a guide (Table 5.4).

SC	Proposed Interpretation
0.71-1.00	A strong structure has been found
0.51-0.70	A reasonable structure has been found
0.26-0.50	The structure is weak and could be artificial
$\leq$ 0.25	No substantial structure has been found

Table 5.4: Subjective interpretation of the Silhouette Coefficient (SC), defined as the average silhouette width for the entire dataset (Kaufman and Rousseeuw, 1990, page 88)

An optimal value of k was sought within the range 5 to 25. While the best partition of the data may be obtained for a value of k outside this range, it is necessary to be pragmatic and give consideration to how the classification is likely to be received and used. In this case, it was considered that fewer than 5 clusters would be unlikely to convince users that the cluster analysis was a true reflection of neighbourhood differences in Leeds, and greater than 25 would be too cumbersome a solution to work with.

A range of partitions were produced and a graph of k against the average silhouette width for each partition was plotted (Figure 5.2). There is a degree of subjectivity in choosing an appropriate point of inflection on the graph but it was decided that values of $k = 8$ or $k = 10$ might be suitable candidates, although a case might also be made for selecting $k = 6$ or $k = 15$.

Thus, two partitions were created, one with 8 groups and one with 10 groups. The average silhouette widths of objects within the individual groups were then compared for each partition (Table 5.5), where n_j is the number of neighbourhoods in the group, *average sil* is the average silhouette coefficient and *medoid* is the representative neighbourhood around which all the other group neighbourhoods are clustered.

In the $k = 10$ partition there are 5 weaker groups where the average silhouette width is less than 0.1. In addition, the average silhouette width, 0.12, for all neighbourhoods in partition $k = 8$ is slightly better than for $k = 10$, 0.11. Examining the within-group average standardised scores of the variables that define the groups reveals that there is a measure of agreement between the two solutions. However, the addition of an extra two groups in the $k = 10$ partition allow for some types of neighbourhoods to be sub-divided further.

For example, Figure 5.3(a) and Figure 5.3(b) show the different ways in which neighbourhoods

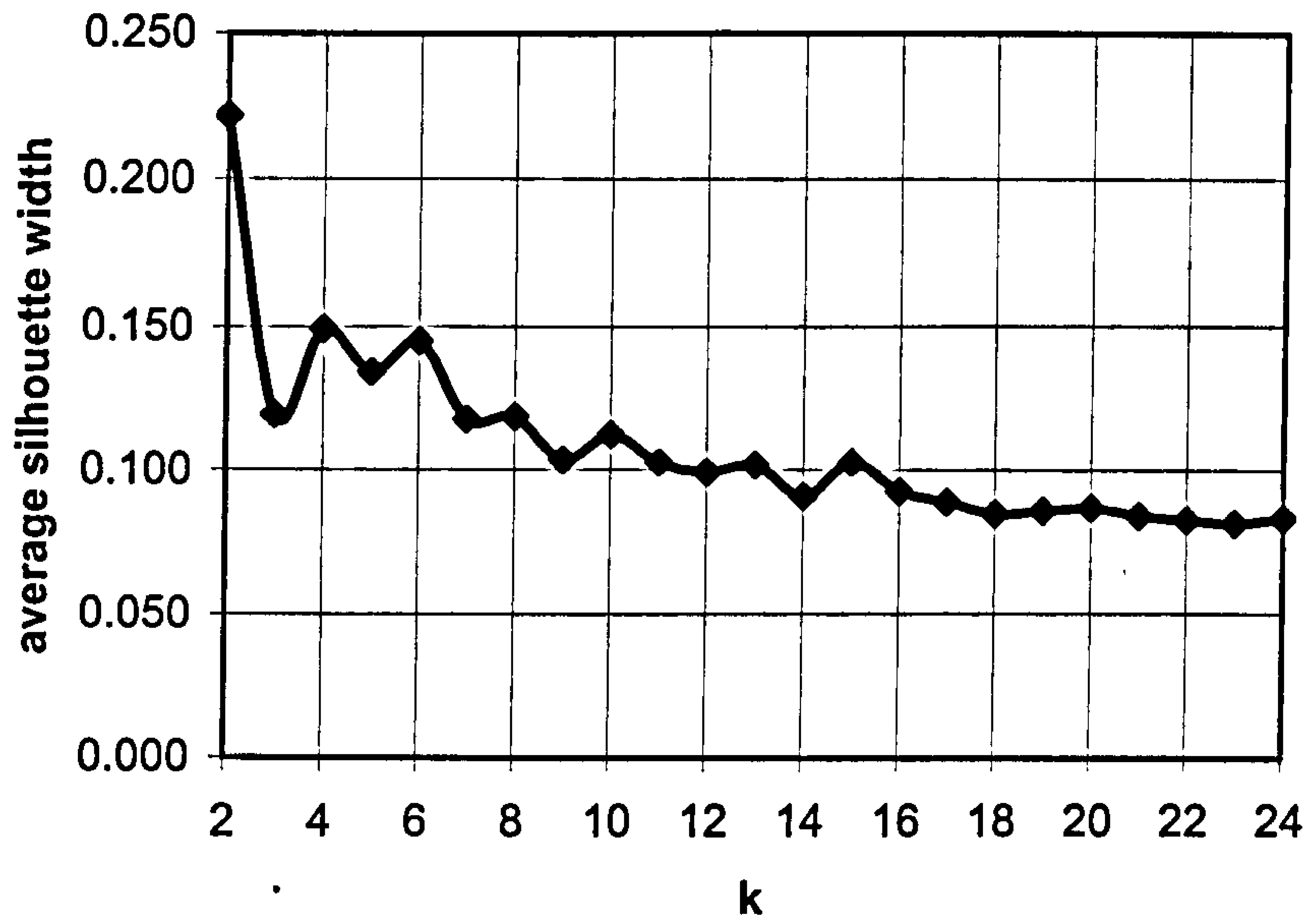


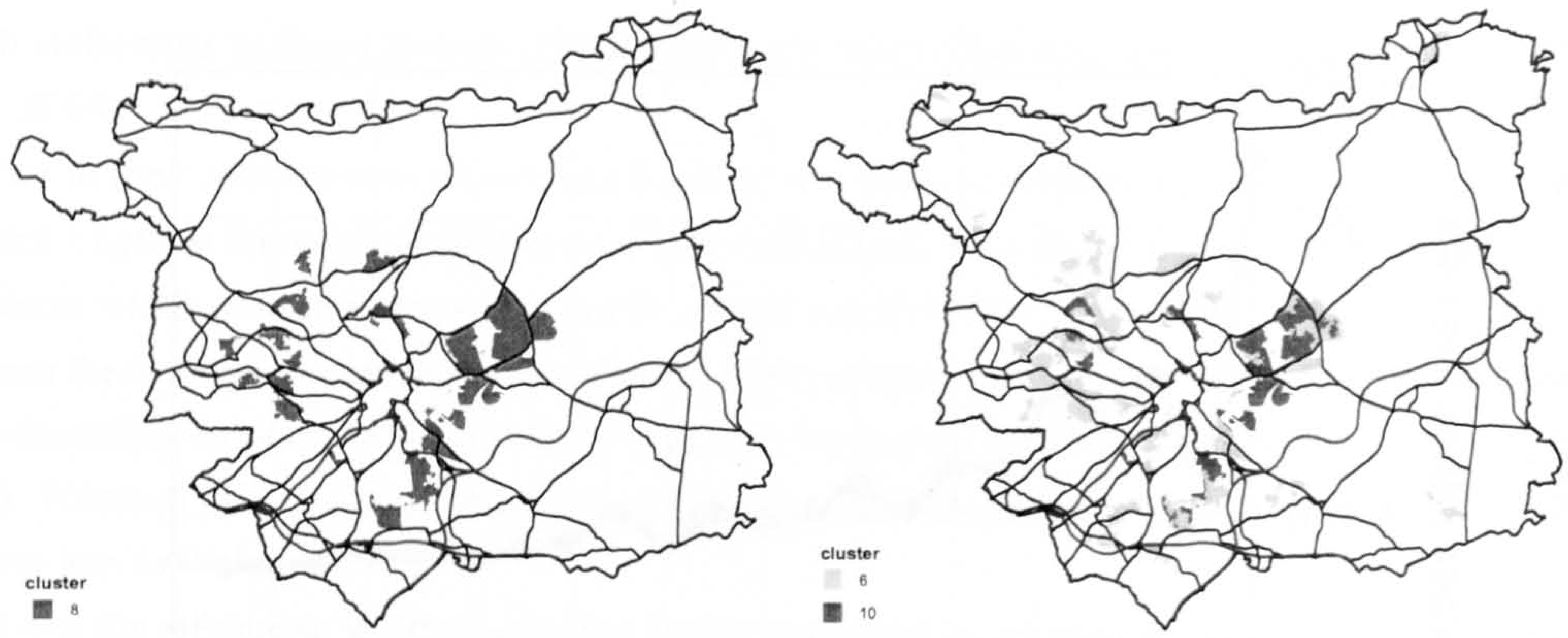
Figure 5.2: Average silhouette widths for different values of k

k=8				k=10			
cluster	n_j	average sil	medoid	cluster	n_j	average silhouette	medoid
1	43	.19	87.01	1	33	.27	7.02
2	25	.21	48.13	2	35	.085	71.02
3	74	.17	62.03	3	21	.23	48.12
4	130	.067	98.03	4	71	.14	62.03
5	79	.04	61.06	5	100	.073	98.03
6	44	-.035	5.06	6	72	.079	88.02
7	16	.29	45.02	7	57	.096	92.01
8	66	.24	72.02	8	37	-.033	5.06
				9	16	.29	45.02
				10	35	.15	86.03

Table 5.5: Cluster metrics for partitions where k=8 and k=10

with medium to high proportions of households renting from social landlords have been grouped. In the $k = 8$ partition, 66 such neighbourhoods are grouped together in group 8, and the relatively high silhouette width (.24) suggests the group is relatively well defined. By contrast, in the $k = 10$ partition, two groups (6 and 10) have captured these types of neighbourhood, and the total number of neighbourhoods has risen to 107. One positive aspect of the split is that it is now possible to discern those council estates which are most deprived, which for the most part are estates in the east of Leeds. While one negative effect of the $k = 10$ split is that the inclusion of extra neighbourhoods on the periphery of the less-deprived group (where social renting is also less important to cluster definition) has lowered the silhouette width of the groups.

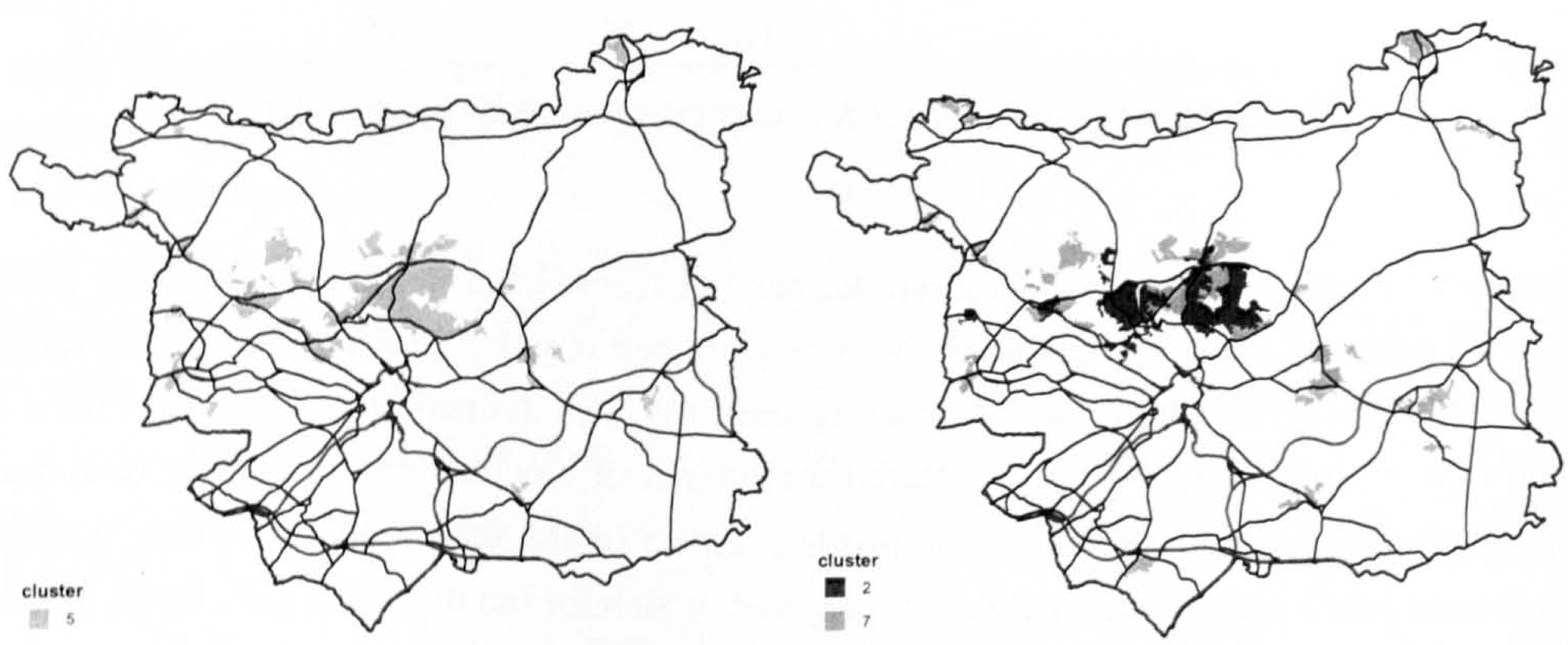
Furthermore, there is a difference in the way that the urban neighbourhoods with a higher proportion of people in the higher social classes have been classified (Figure 5.3(a) and Figure 5.3(b)). In the $k = 8$ partition, many of these (n=79) are grouped together in group 5. Analysis of the within-group average standardised scores of the variables shows that SC12 (professionals and managers), QU02



(a) k=8 (b) k=10

Figure 5.3: Comparison of partitions with respect to social renting

(higher qualifications), CA02 (car ownership) and TE01 (owner occupiers) are important in defining this group. In the $k = 10$ partition, however, the groupings are somewhat different. Instead of one there are now two groups (2 and 7), with a total of 92 neighbourhoods. Analysis of the standardised scores shows that group 2 is separating out those neighbourhoods where QU02 and SC12 are very much higher than average, while variables such as tenure and house type are much less important. Group 7, by contrast, is discriminating by tenure and house type, while qualifications and social class are less important. In geographical terms, this split has tended to separate neighbourhoods according to their distance form the city core, with group 7 being the more distant. Other variables that are prominent here are age and migration. Although the differences are not great, it appears that the young and the old are more prevalent in the neighbourhoods closer to the city core, and they are also migrating more. The greater than average incidence of flats in these neighbourhoods might be attracting both groups. In the group 7 neighbourhoods further away from the centre of Leeds, there are slightly more people in middle age, and they appear to migrate less than average.



(a) k=8 (b) k=10

Figure 5.4: Comparison of partitions with respect to more affluent sub-urban neighbourhoods

On all this evidence, it is still difficult to decide whether the partition would be better with 8 or 10 groups, although it has been suggested that where there is doubt, the higher figure should be taken

(Milligan, 1996).

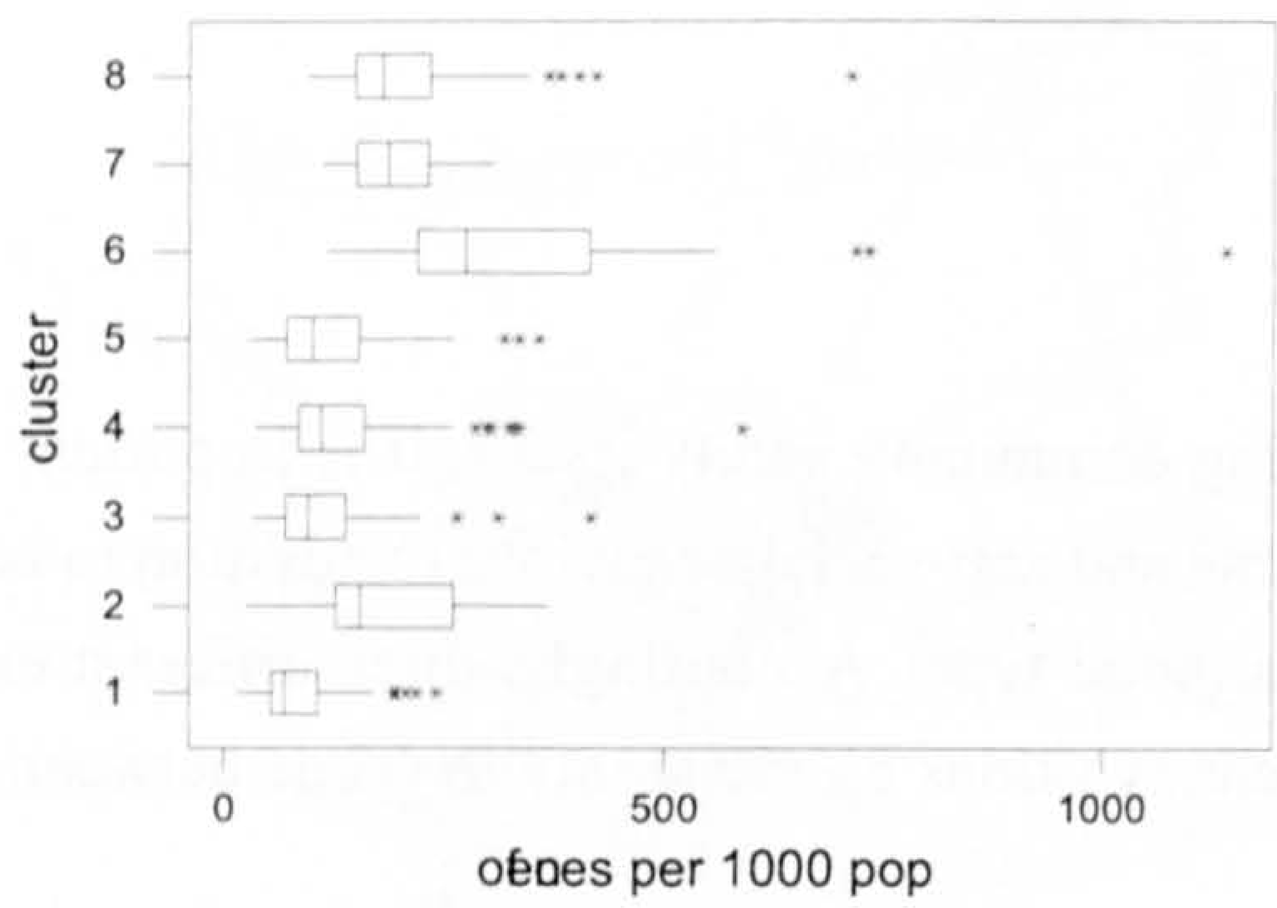
5.4.2 Additional Crime Heuristics

As the final classification is to be used for examining community safety issues, it is important that the partition should attempt to reflect patterns of crime and anti-social behaviour in addition to being a good reflection of general differences in neighbourhood type. Accordingly, three different crime heuristics were used to examine the effect of different partitions on variability in crime between and within groups.

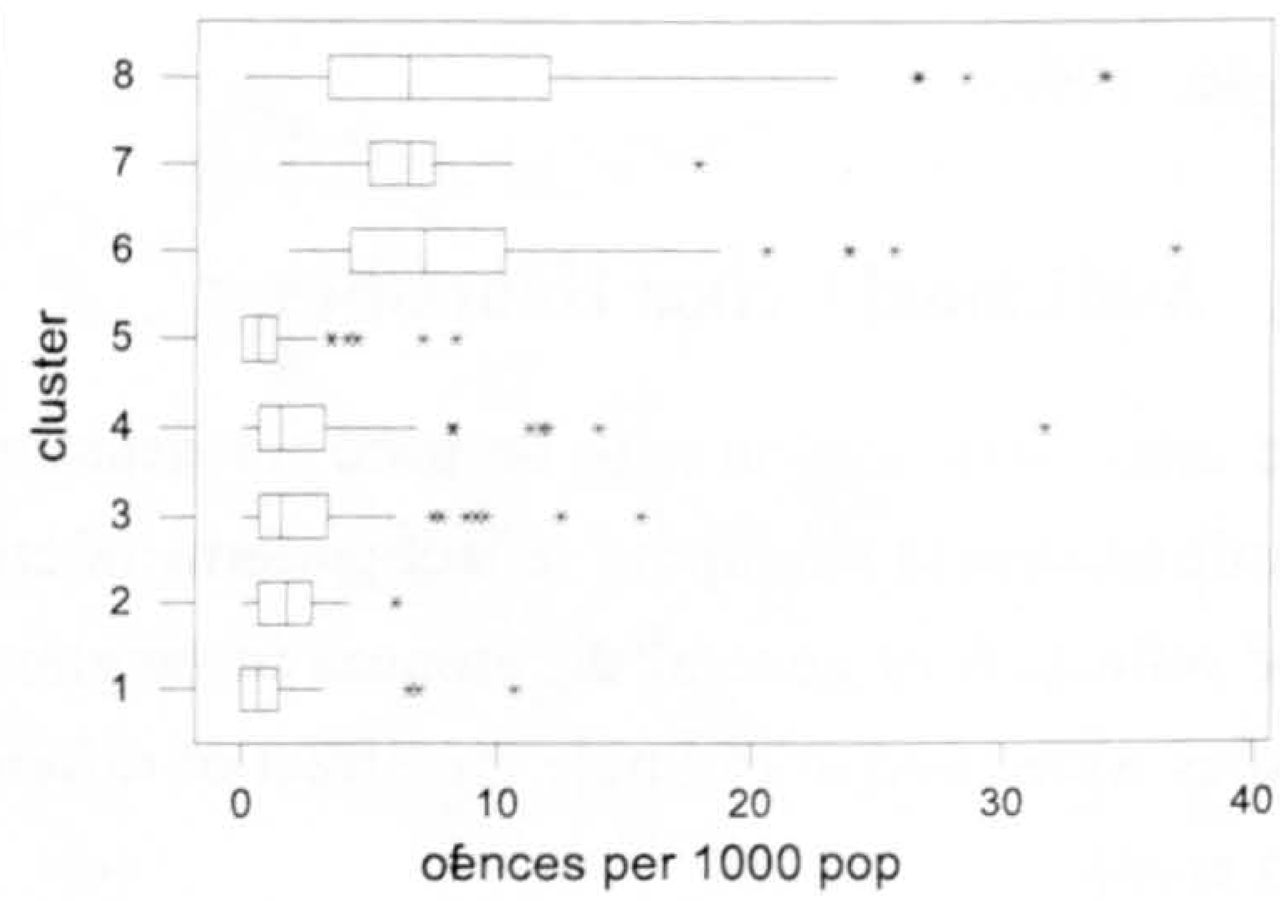
Firstly, it is possible to visually examine the variability of crime rates using box plots (Figure 5.5 and Figure 5.6). For each crime type under consideration, these show aspects of spread and central tendency of crimes rates for the neighbourhoods in each group. Following the general principle of cluster analysis - that groups should be internally homogeneous and distinct from one another - the range of crime rates in each group should ideally be low and the different median values of the groups should be spread. In addition, it is helpful if the number of outliers is minimised.

Thus, following on from the discussion of how the $k = 10$ partition has separated certain neighbourhood types, the boxplots begin to show whether this is useful for differentiating crime rates. Taking the social renting neighbourhoods again, Figure 5.5(c) shows that there is a general difference between crime rates in groups 6 and 10, with the more deprived group 10 suffering to a greater extent. At the extremes, the biggest difference between the two groups is for WYF&RS recorded primary arson (Figure 5.5(d)) and criminal damage (Figure 5.6(c)), which includes WYP recorded arson. By contrast, there is very little difference between the two groups with respect to vehicle crime (Figure 5.6(g)), burglary elsewhere (Figure 5.5(h)) and other theft (Figure 5.6(d)). There is some difference in violent crime rates, with group 10 having the higher values. Thus, the separation of social housing neighbourhoods into two groups in the $k = 10$ partition is useful for separating out those neighbourhoods where arson and criminal damage are a particular problem. From the opposite point of view however, the lack of discrimination for the high volume acquisitive crimes do not lend support for the $k = 10$ partition.

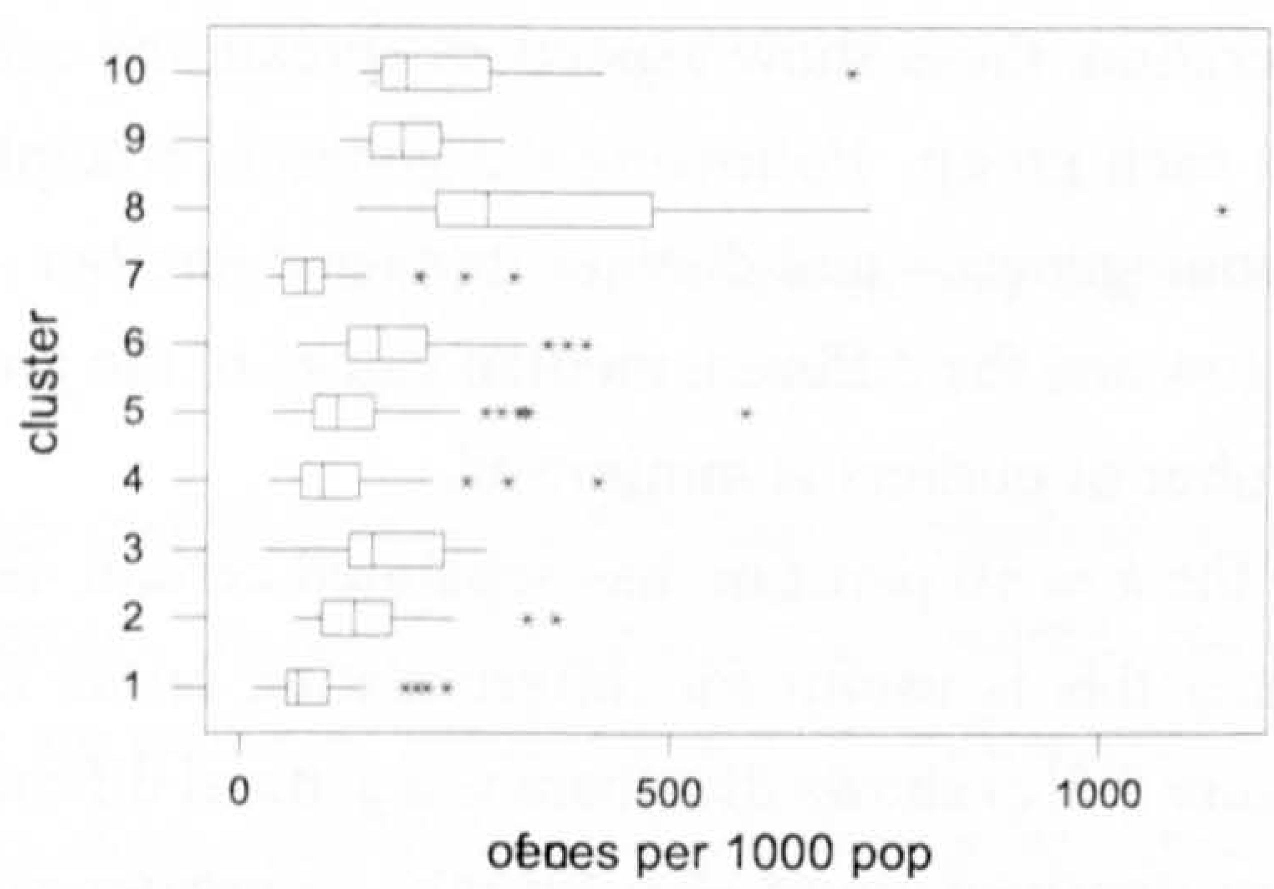
Considering the other major difference between the two partitions, a visual analysis of the boxplots reveals less marked changes in the discrimination of crime rates. The strongest support for the more central affluent suburbs (group 2) in the $k = 10$ partition comes from its ability to reflect higher levels of vehicle crime (Figure 5.6(g)) and burglary dwelling (Figure 5.5(g)) in these types of neighbourhood. By contrast, group 7, tends to group neighbourhoods with slightly higher rates of burglary elsewhere (Figure 5.5(h)). Interestingly, analysis of the group portraits suggests there is more mixed land use in group 2, which, if this were capturing business premises, might have been expected to reflect the higher risk of burglary elsewhere. What must be remembered, however, is that burglary elsewhere is not exclusive to non-residential properties, indeed, analysis of burglary elsewhere offences in 2003/04 reveals that for neighbourhoods in group 7 of the $k = 10$ partition, 616 of the 992 offences recorded (62%) were against garages, sheds, outhouses and other structures associated with residential properties.



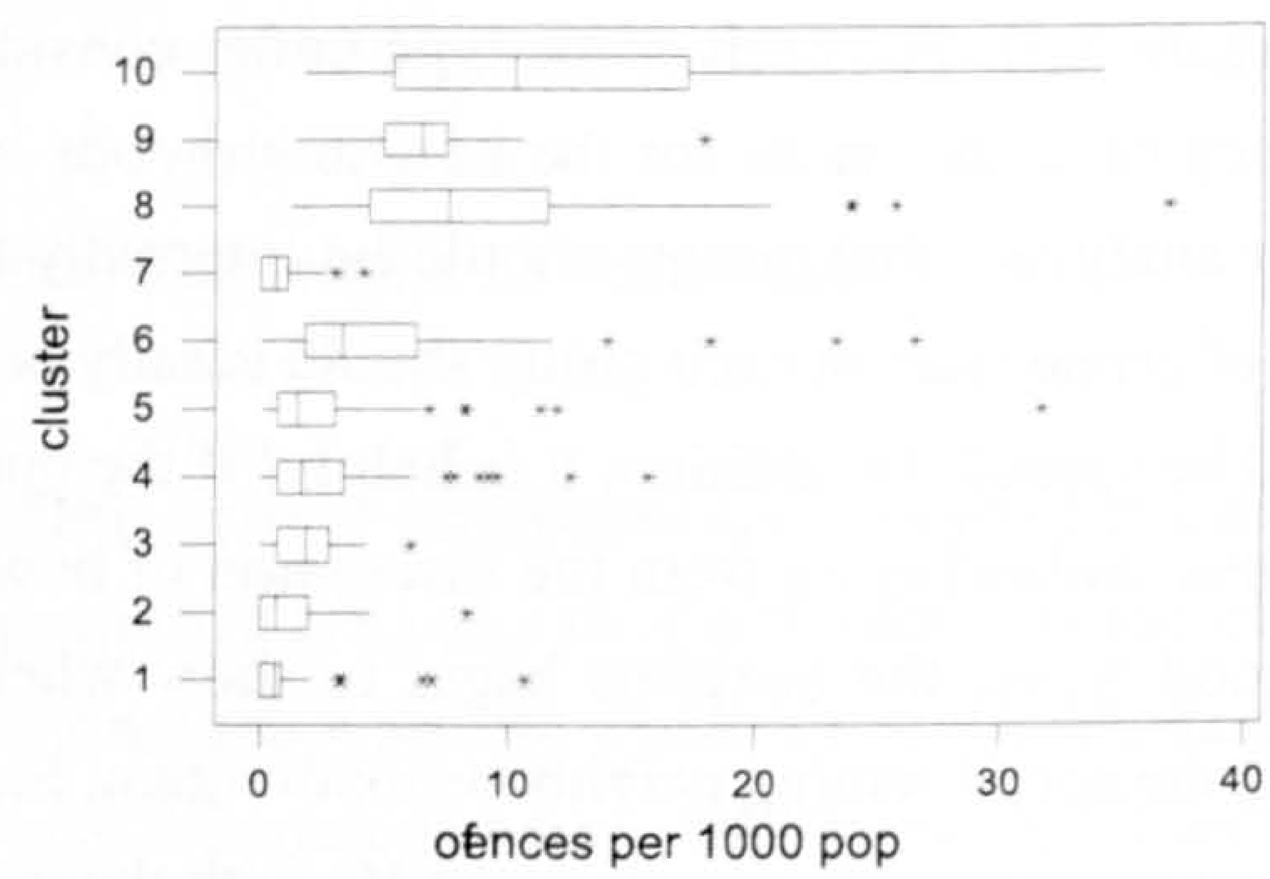
(a) all crime $k = 8$



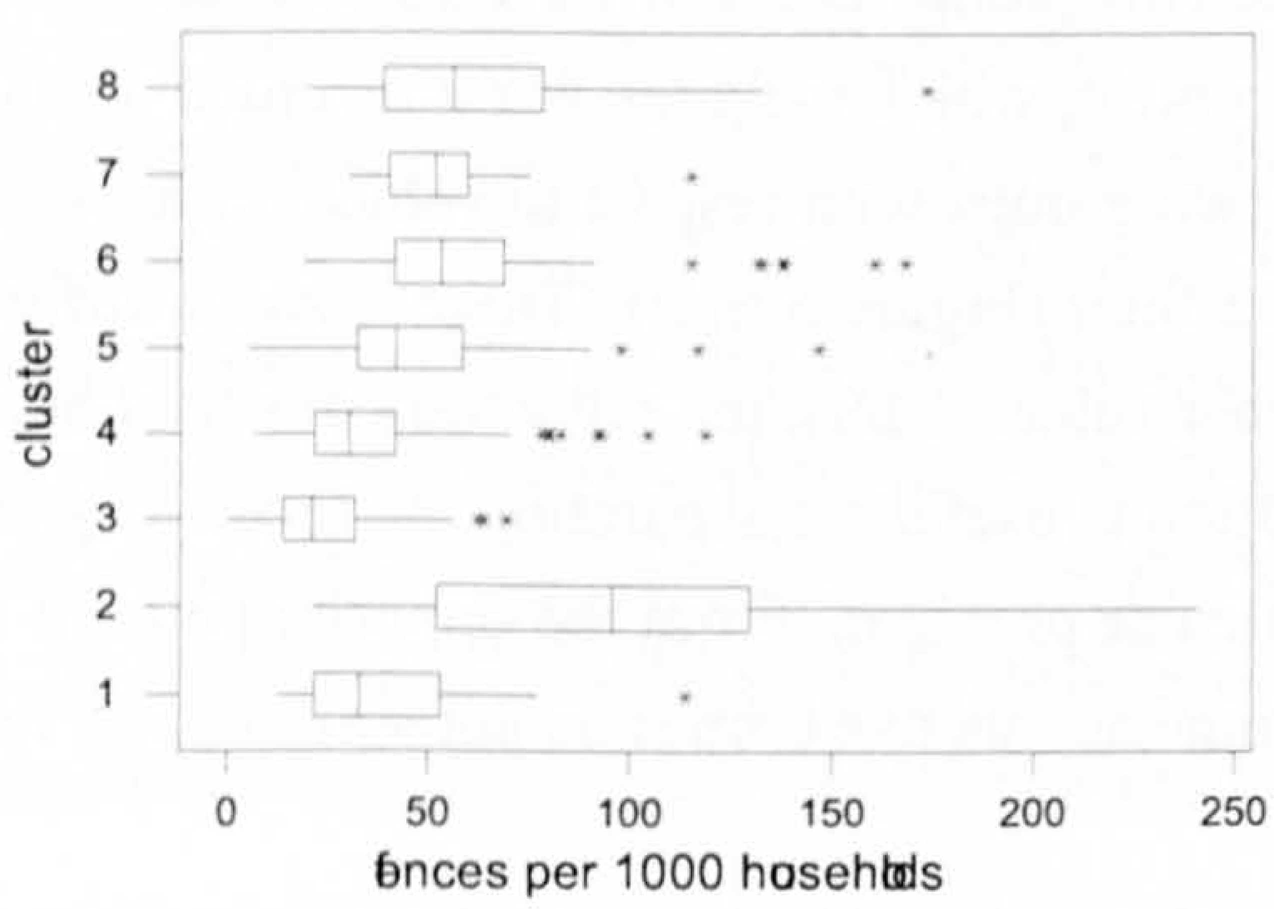
(b) arson $k = 8$



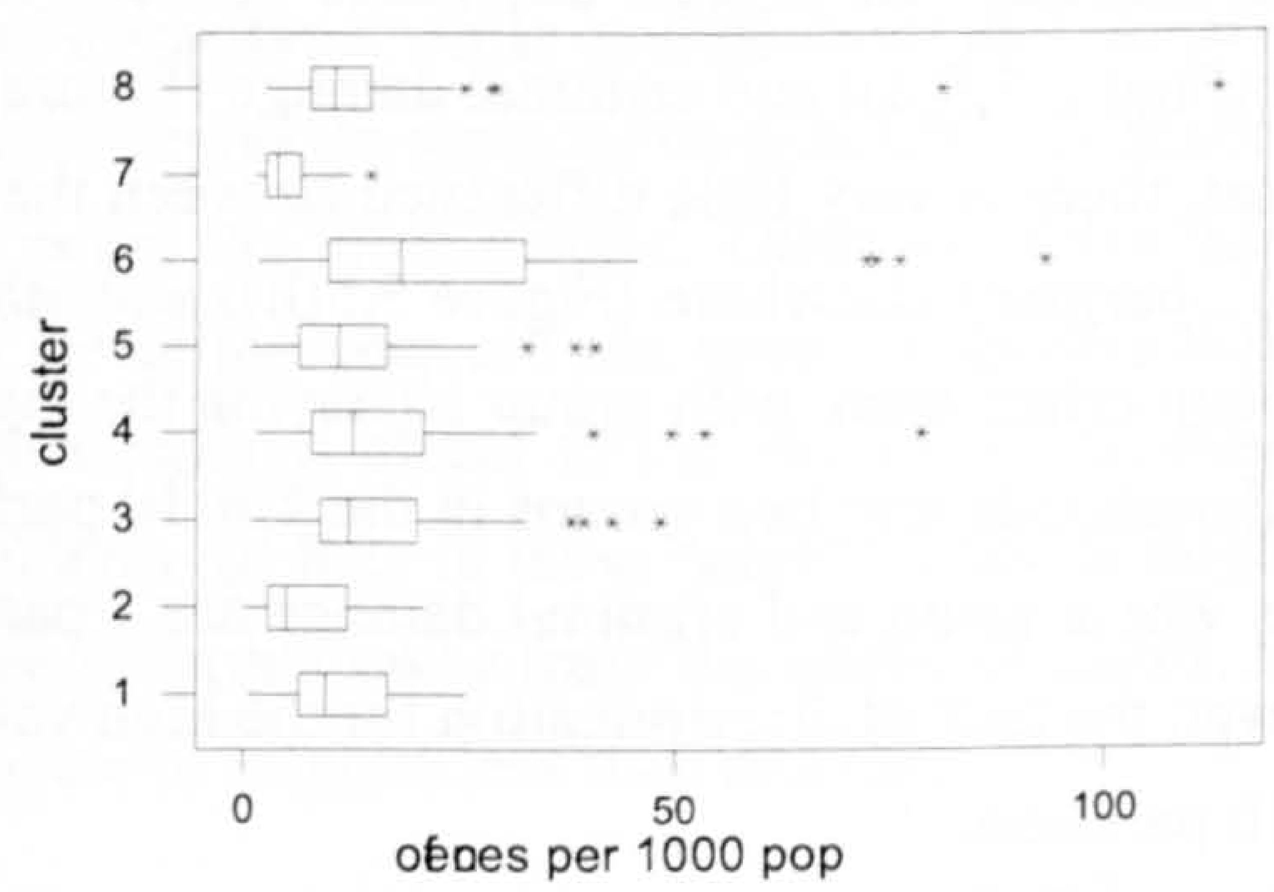
(c) all crime $k = 10$



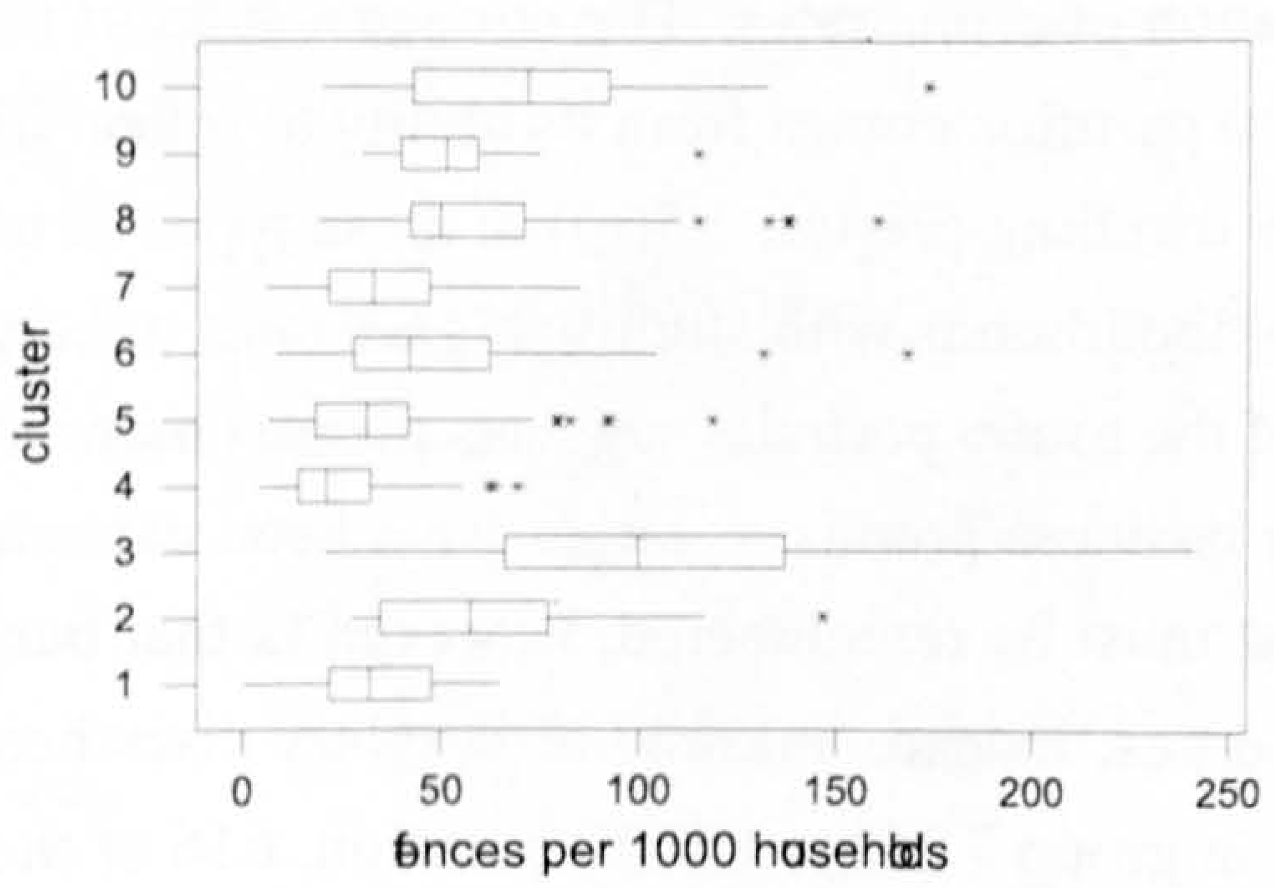
(d) arson $k = 10$



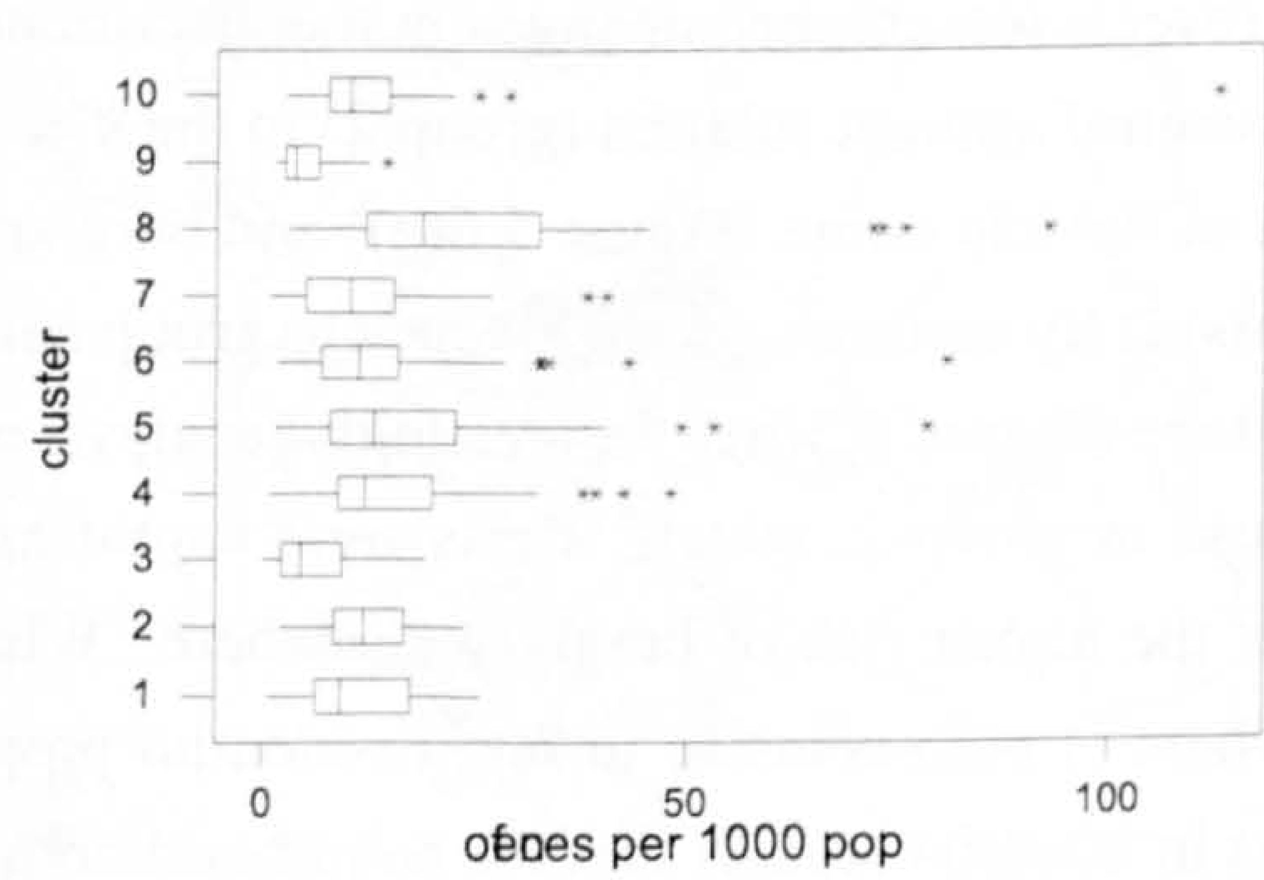
(e) burglary dwelling $k = 8$



(f) burglary elsewhere $k = 8$

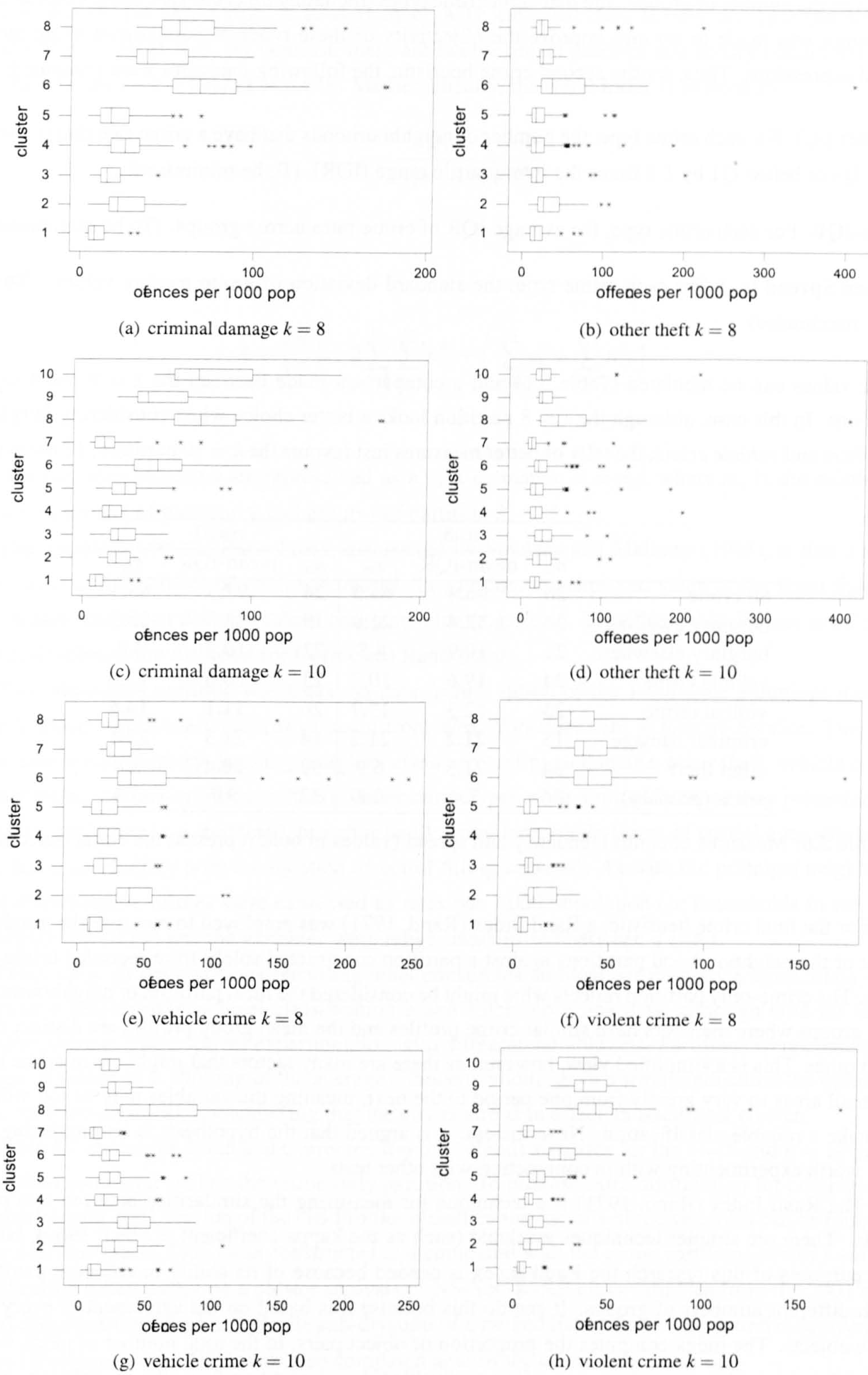


(g) burglary dwelling $k = 10$



(h) burglary elsewhere $k = 10$

Figure 5.5: Boxplots of all crime, arson and burglary rates for $k = 8$ and $k = 10$ partitions



(a) criminal damage $k = 8$

(b) other theft $k = 8$

(c) criminal damage $k = 10$

(d) other theft $k = 10$

(e) vehicle crime $k = 8$

(f) violent crime $k = 8$

(g) vehicle crime $k = 10$

(h) violent crime $k = 10$

Figure 5.6: Boxplots of criminal damage, other theft, vehicle crime and violent crime rates for $k = 8$ and $k = 10$ partitions

Given the number of groups, and that eight crime types (including *all crime*) are being considered, an attempt was made to try and improve the objectivity of these types of comparison using more formal expressions. Thus, for the second crime heuristic, the following measures were computed:

Outliers (n_o) For each crime type, the number of neighbourhoods that have a crime rate that is above Q3 or below Q1 by 1.5 times the interquartile range (IQR). (To be minimised).

Mean-IQR For each crime type, the average IQR of crime rates across groups. (To be minimised).

Median Spread (s_m) For each crime type, the standard deviation of group median values. (To be maximised).

These values can be tabulated (Table 5.6) and a comparison made between the $k = 8$ and $k = 10$ partitions. In this case, although the $k = 8$ partition looks a better choice when considering *burglary elsewhere* and *vehicle crime*, the tally of better measures just favours the $k = 10$ partition, 13 measures to 11 .

	n=8			n=10		
	n_o	mean-IQR	s_m	n_o	mean-IQR	s_m
all crime	26	**96.4**	**66.3**	**24**	97.3	65.3
burglary dwelling	24	**32.4**	22.9	**19**	33.8	**23.3**
burglary elsewhere	**21**	10.9	**4.5**	22	**10.3**	4.2
vehicle crime	**24**	**19.6**	10.2	25	20.6	**12.6**
violent crime	**23**	12.3	13.7	26	**11.1**	**14.6**
criminal damage	18	**21.2**	21.2	**14**	21.3	**22.1**
other theft	**34**	21.5	6.9	39	**20.4**	**8.8**
arson (primary)	36	**3.4**	2.8	**32**	3.7	**3.4**

Table 5.6: Measures of central tendency and spread (values in bold represent the better outcome)

For the final crime heuristic, a Rand Index (Rand, 1971) was employed to measure the goodness of fit of the neighbourhood partitions against a partition constructed solely from recorded crime rate data. The crime-only partition reflects what might be considered the ideal partition of neighbourhoods into groups where members have similar crime profiles and the mean group profiles are distinct from each other. This is a simplified view, however, as there are many factors that might cause crime rates in small areas to vary greatly from one period to the next, meaning the variables may be too volatile to make a reliable classification. Nevertheless, it is argued that the hypothesis is an interesting one and worth experimenting with in conjunction with other tests.

The Rand Index (Rand, 1971) is a technique for measuring the similarities between two partitions. There are simpler techniques available (such as the kappa coefficient (Cohen, 1960), but for the purposes of this research the Rand Index is needed because of its ability to compare partitions with different numbers of groups. It can do this because it is based on the agreement of every pair on n objects. The index computes the proportion of object pairs, to the total number of pairs, which are either (a) in the same group according to partition 1 and partition 2, or (b) in different groups according to partition 1 and partition 2. Values of the index range from 0 (no agreement between partitions) to 1 (complete agreement between partitions). To illustrate this, consider two partitions, Alpha and Beta. Each have 7 objects and the group memberships (in brackets) of the objects are Alpha[A(1),B(1),C(2),D(2),E(3),F(3),G(1)] and Beta[A(1),B(1),C(1),D(2),E(2),F(3),G(3)]. There are

21 possible unique object pairs (e.g. A-B, A-C, A-D, etc.), but there is only one case that satisfies condition (a), that is A-B. By contrast, there are twelve object pairings that satisfy condition (b), so the $RandIndex = (1+12)/21 = 0.619$. Mathematically, the Rand Index is defined as

$$I_R = \frac{2A}{n(n-1)} = \frac{A}{\binom{n}{2}} \tag{5.1}$$

where

$$A = \binom{n}{2} + 2\sum_{i=1}^{c_1}\sum_{j=1}^{c_2} n_{ij}^2 - \left(\sum_{i=1}^{c_1} n_{i.}^2 + \sum_{j=1}^{c_2} n_{.j}^2\right) \tag{5.2}$$

and the two classifications are represented as a c_1 x c_2 matrix $N = n_{ij}$, where n_{ij} is the number of objects in group i of partition 1 and group j of partition 2.

One problem with the Rand Index, discovered by Fowlkes and Mallows (1988), is that it tends to increase as the number of groups increases. In addition, the expected value of the Rand Index of two random partitions does not take a constant value (e.g. zero). To address this, Hubert and Arabie (1985) developed the Adjusted (or Corrected) Rand Index.

The crime-only partition was based on data from a subset of the 14 offence groupings used by West Yorkshire Police and records of arson from West Yorkshire Fire & Rescue Service. The variables used were: burglary dwelling, burglary elsewhere, criminal damage, other theft, vehicle crime, violent crime, and arson (primary fires). Other crime types were omitted because they presented too small a sample size (e.g. robbery), or suffer badly from undesirable forms of spatial autocorrelation (e.g. fraud and forgery with the location of petrol filling stations). As with the principal neighbourhood classification, crimes were expressed as rates per 1,000 population (or households in the case of burglary dwelling) and the variables were range standardised between 0 and 1.

Potentially, all the questions regarding what constitutes an ideal partition arise again when constructing a neighbourhood-level classification based solely on crime data, although with far fewer variables there is less scope for experimenting with different variable sets. Partitions were created for a range of values of k. Plotting of the average silhouette widths gave a strong indication the optimum value for k would be 10, remembering that the city centre is in effect an additional cluster.

Table 5.7 shows the Rand and Corrected Rand (C-Rand) statistics for the $k=8$ and $k=10$ partitions when compared against the crime-only partition. To provide extra information for comparison, an experimental 2001 version of the GB Profiler classification (details of the variables can be found in Blake and Openshaw, 1995) was constructed and compared with the crime partition, as was a simple classification based solely on a cluster analysis of NS-SeC social class variables from the 2001 census. A classification based on a decile sub-division of a ranked list of Townsend material deprivation scores (Townsend et al., 1988) was also compared against the crime-only partition.

Contrary to the findings from the previous set of heuristics, the $k=8$ partition appears to be a slightly better solution. The comparison of the task-specific classifications developed so far, against other classifications that might be (and sometimes have been) employed (for example Craglia et al., 2000) to analyse patterns of crime are also interesting. Despite initial confidence in the main hypothesis of the research - that a task-specific classification might be a better alternative to a general purpose classification - at this stage, and according to this heuristic, this does not appear to be the case to any great extent. The issue of the comparative performance of the final neighbourhood classification to alternatives is considered in more detail in Chapter 7.

	Pseudo-true crime partition	
	Rand	C-Rand
k=8	0.774	0.131
k=10	0.795	0.120
GB-Profiles (k=13)	*0.799*	*0.106*
NS-SeC (k=11)	*0.800*	*0.101*
Townsend deciles (k=10)	*0.804*	*0.081*

Table 5.7: Level of agreement between neighbourhood classifications and a crime-only partition.

5.4.3 Synthesising Findings

Considering a range of different cluster and crime heuristics helps to inform the decision about which partition, $k = 8$ or $k = 10$, to pursue further. Unfortunately, there is support for both partitions! Firstly, the way in which the $k = 10$ partition separates social housing estates into two groups is useful. There are many council tenanted properties in Leeds and the neighbourhoods in which they are situated receive a lot of management in various forms, including local housing officers and neighbourhood wardens. Assessing problems and policy responses in these areas is important and having the two groups might provide a useful tool for analysing crime patterns. Similarly, the subdivision of the more affluent residential suburbs in the $k = 10$ partition is useful for its ability to discriminate levels of property and vehicle crime - key priority areas for crime reduction strategy in Leeds. The second set of crime heuristics sways in favour of the $k = 10$ partition, albeit only by 13 better measures to 11, reducing to 12 to 9 if *all crime* is discounted. Furthermore, whether including or excluding the *all crime* measures, the $k = 10$ partition has slightly fewer outlier crime rates overall. Finally, although the third crime heuristic favours the $k = 8$ partition it is only by a small margin.

In conclusion, the decision was taken to proceed with the $k = 10$ partition, despite the slightly poorer results from the third crime heuristic.

5.4.4 Improving Cluster Definition

While this initial $k = 10$ partition is usable, the average silhouette widths for a number of clusters suggest it might be worth trying to improve the partition to produce more distinct and separate clusters. Additionally, even if the signs were otherwise it would be good practice to try and provide confirmatory evidence of structure in the data by applying alternative grouping techniques to the data. Thus, the next stage in the neighbourhood classification process was to re-evaluate the importance of different variables to partition structure and to look for levels of agreement between the PAM partition and partitions created using alternative clustering methods.

Two approaches to variable analysis were adopted. Firstly, graphs of z-scores for each variable in each cluster were examined to identify variables that might be contributing poorly to group differences. Secondly, by using the OVW (Optimal Variable Weighting) (Makarenkov and Legendre, 2001) and HINoV techniques applied to the $k = 10$ partition, an attempt was made to formally identify the relative importance of variables. The other cluster analysis techniques employed to look for partitions were k-means (as implemented in ClustanGraphics), and fuzzy c-means (1987).

5.4.5 k-means

To begin with, experiments were conducted to test the effect of different strategies for creating the initial partition. The first strategy, random assignment (MacQueen, 1967), involves putting every case into an initial cluster at random. The second strategy used the medoids from the $k = 10$ partition created by PAM to form the initial cluster centres around which the other objects would begin to be clustered. The third strategy was similar, in that it used a priori knowledge of data structure, but instead used the median values from each decile of neighbourhoods that had been ranked by Townsend material deprivation score. These different strategies made no difference with the dataset of Leeds neighbourhoods, and the final partitions all came out the same. Comparing the PAM $k = 10$ partition with the k-means solution produced a Rand Index of 0.880 and an Adjusted Rand Index of 0.439, suggesting a good deal more agreement than disagreement between the two partitions. Calculation of silhouette widths for the k-means partition reveal an overall average of 0.13. The detailed cluster statistics are given in Table 5.8. Overall, the results are slightly better than those for the PAM $k = 10$ partition.

	k=10	
cluster	n	average silhouette
1	50	.23
2	26	.025
3	44	.11
4	38	.083
5	51	.13
6	23	.20
7	71	.11
8	84	.18
9	71	.068
10	19	.26

Table 5.8: Cluster statistics for k-means solution generated with ClustanGraphics

Some experiments were also conducted with the outlier detection feature in ClustanGraphics. To effect this feature, a threshold has to be specified above which a case will be deemed an outlier and placed in the residue group. The impression given by Lorr (1983) is that, typically, cases which are dismissed to a residue group on one pass of the partitioning algorithm can be reintroduced if the group centres move back within range. This is also the case for ClustanGraphics (personal communication with David Wishart 11-Nov-2004). Using a maximum threshold of 0.04, a $k = 10$ partition was produced (using PAM medoids as seeds) that placed 19 cases in the residue group. However, as the aim is to classify all neighbourhoods in Leeds (except the two city centre neighbourhoods) each of these outliers had then to be manually assigned to the group whose centre was nearest. Comparison with the PAM partition produced slightly better Rand and Corrected Rand statistics than seen previously (0.895 and 0.466 respectively), suggesting some value in adopting this approach to outliers.

As a final experiment, the neighbourhood data were grouped using the k-means routine provided in the SPSS software package. The resulting $k = 10$ partition was unsatisfactory in that the size of group memberships were too extreme. The largest group contained 184 (38%) of the objects, and three other groups had small memberships of 2,4 and 13 respectively. The level of similarity between this and the PAM $k = 10$ partition was lower than seen using ClustanGraphics, with Rand and Adjusted Rand indeces of 0.820 and 0.381 respectively.

The level of agreement between the k-means partition and the crime-only partition was also measured (Table 5.9). The ClustanGraphics results of the Corrected-Rand tests were slightly worse than those seen for the PAM partition (Table 5.7) .

Partition	Rand	C-Rand
k-means	0.799	0.115
k-means with outlier detection	0.808	0.116

Table 5.9: Level of agreement between partitions created using k-means and the crime-only partition, $k = 10$.

The k-means and PAM partitions were then compared using the measures in the second crime heuristic, described above. The balance of better scores was in favour of the k-means partition, 13 to 11 (PAM results recorded previously in Table 5.6). This advantage drops to 11 to 10 when the *all crime* data are discounted.

	k-means $k = 10$		
	n_o		s_m
all crime	25	97.2	84.2
burglary dwelling	24	32.1	22.8
burglary elsewhere	19	10.9	6.3
vehicle crime	19	22.0	13.4
violent crime	16	12.3	14.3
criminal damage	18	19.8	20.5
other theft	30	21.3	12.2
arson (primary)	29	3.4	3.0

Table 5.10: Measures of central tendency and spread of the k-means $k = 10$ partition

Looking at the k-means groups revealed many similarities with the PAM partition, as indicated by the Adjusted Rand index score. The one important difference, however, is that unlike PAM, the k-means partition did not break up those neighbourhoods where there is a higher degree of social renting in the tenure structure. Instead, the extra group was used to further fragment neighbourhoods in the southern hinterland.

5.4.6 Fuzzy c-means

Although some studies have made suggestions about what 'fuzziness' index to use, tests were conducted with the Leeds neighbourhood dataset suggested that results were easier to interpret if the degree of fuzziness was kept fairly low. Most of the time, interest is only in the next two or three nearest group, and not all the k possible groups. A value of $m = 2$ produced too much fuzziness, so after trying out several alternatives, a value of $m = 1.4$ was chosen. Euclidean distances were used. In tests, the use of seeds based on the PAM $k = 10$ medoids yielded exactly the same partition as random selection of initial group centres, as was found with ClustanGraphics k-means. Table 5.12 shows the levels of agreement between the c-means hard partition and the PAM and crime-only partitions, although the very nature of a fuzzy partition makes comparison with other hard partitions somewhat meaningless.

	c-means $k = 10$ (hard)		
	n_o	IQR	s_m
all crime	23	100.4	86.3
burglary dwelling	19	32.4	23.9
burglary elsewhere	20	11.2	6.8
vehicle crime	17	21.7	14.7
violent crime	19	13.3	14.6
criminal damage	12	22.3	21.5
other theft	29	22.6	14.1
arson (primary)	27	4.1	3.1

Table 5.11: Measures of central tendency and spread of the fuzzy c-means $k = 10$ hard partition

	Rand	C-Rand
PAM k=10 partition	0.898	0.512
crime-only partition	0.808	0.134

Table 5.12: Level of agreement between hard partition created using fuzzy c-means and the PAM $k = 10$ and crime-only partitions.

Analysis of the hard partition group profiles produced by fuzzy c-means showed groups were very similar to those produced by PAM. The only significant difference was between the ways more affluent neighbourhoods were grouped. Both classifications make a split, but the PAM group 7 has higher proportions of ethnic minorities, people with higher qualifications and detached houses. In geographical terms, the split still follows a concentric ring settlement model, but whereas the PAM split was more evident in the northern suburbs the fuzzy c-means split (groups 6 and 9) is biased toward southern suburbs and towns.

The main usefulness of a fuzzy partition, however, is what the membership strength coefficients can reveal about neighbourhoods. Firstly, the membership strength coefficients can be used to highlight those parts of the c-means hard partition that are weakest (Figure 5.7), that is, hard allocations based on the lowest coefficients. When compared with the geography of silhouette widths from the PAM partition (Figure 5.8), there is some agreement about which neighbourhoods do not cluster well, which might have consequences when analysing crime and disorder patterns in these places. The author's home neighbourhood in Potternewton is one such area (also classified differently in the k-means partition), as are (possibly) parts of Moortown, Beeston and Sandford (Bramley).

A second way in which the c-means coefficients can be useful is to map the membership strength of every neighbourhood with respect to one group. Figure 5.9 shows this technique applied to c-means group 4, the 'student' group. The geography of the coefficients is much as would be expected for the core student neighbourhoods. Moreover, the map shows which of the student neighbourhoods are more 'studenty' than others. The map highlights that relatively high coefficients exist in more unexpected areas as well. Sometimes this may be due to the coincidence of student-like neighbourhood characteristics in non-student areas, such as residential churn and higher than usual levels of private renting.

Thirdly, following on from identifying neighbourhoods whose hard membership is based on a low coefficient, it is possible to see the extent to which neighbourhoods are sometimes 'pulled' with almost equal strength toward a number of different groups. Table 5.13 shows the strength coefficients for neighbourhoods with the lowest standard deviation between their 10 coefficients. The first two of

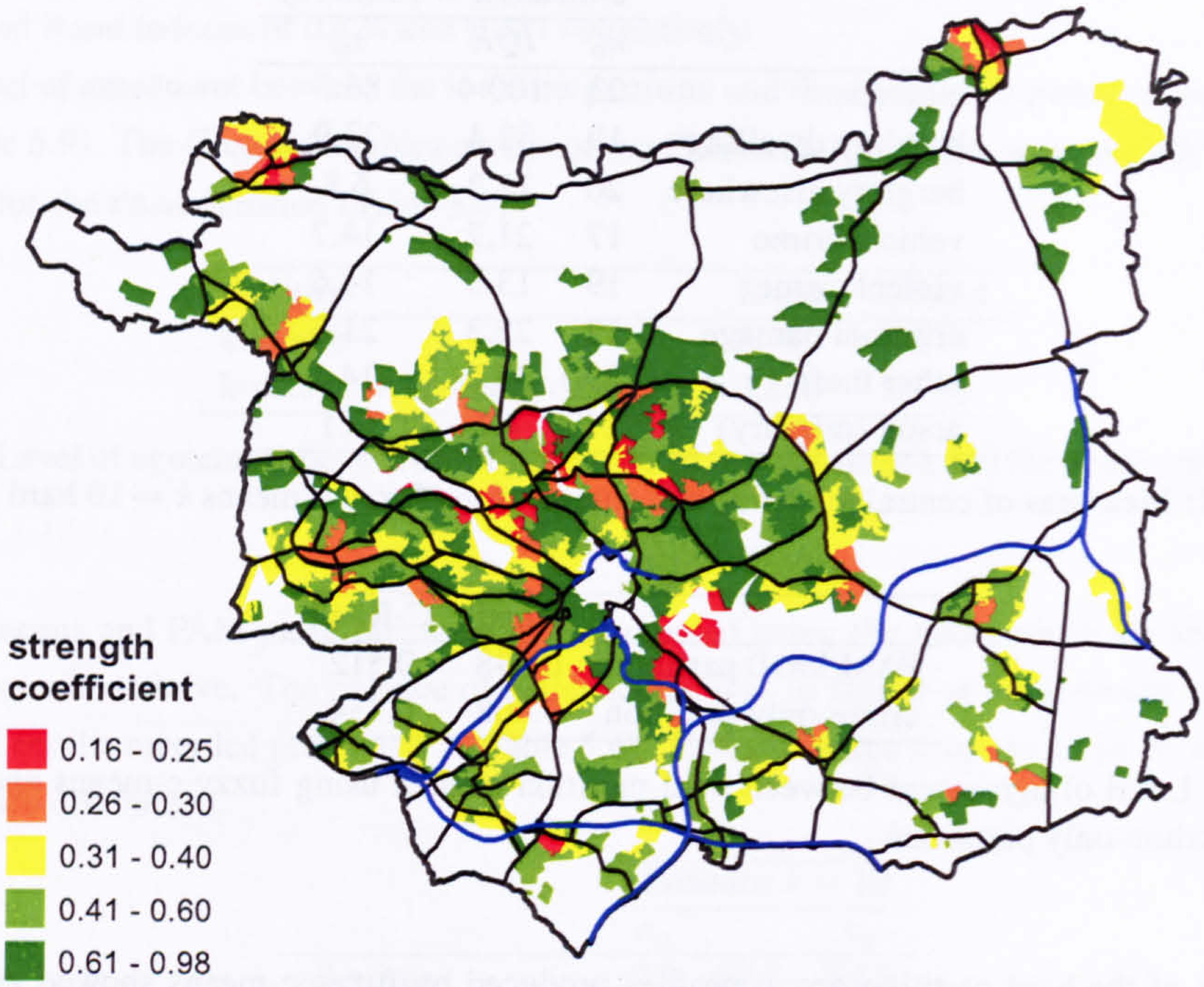


Figure 5.7: Geography of membership strength coefficients used to produce fuzzy c-means hard clustering.

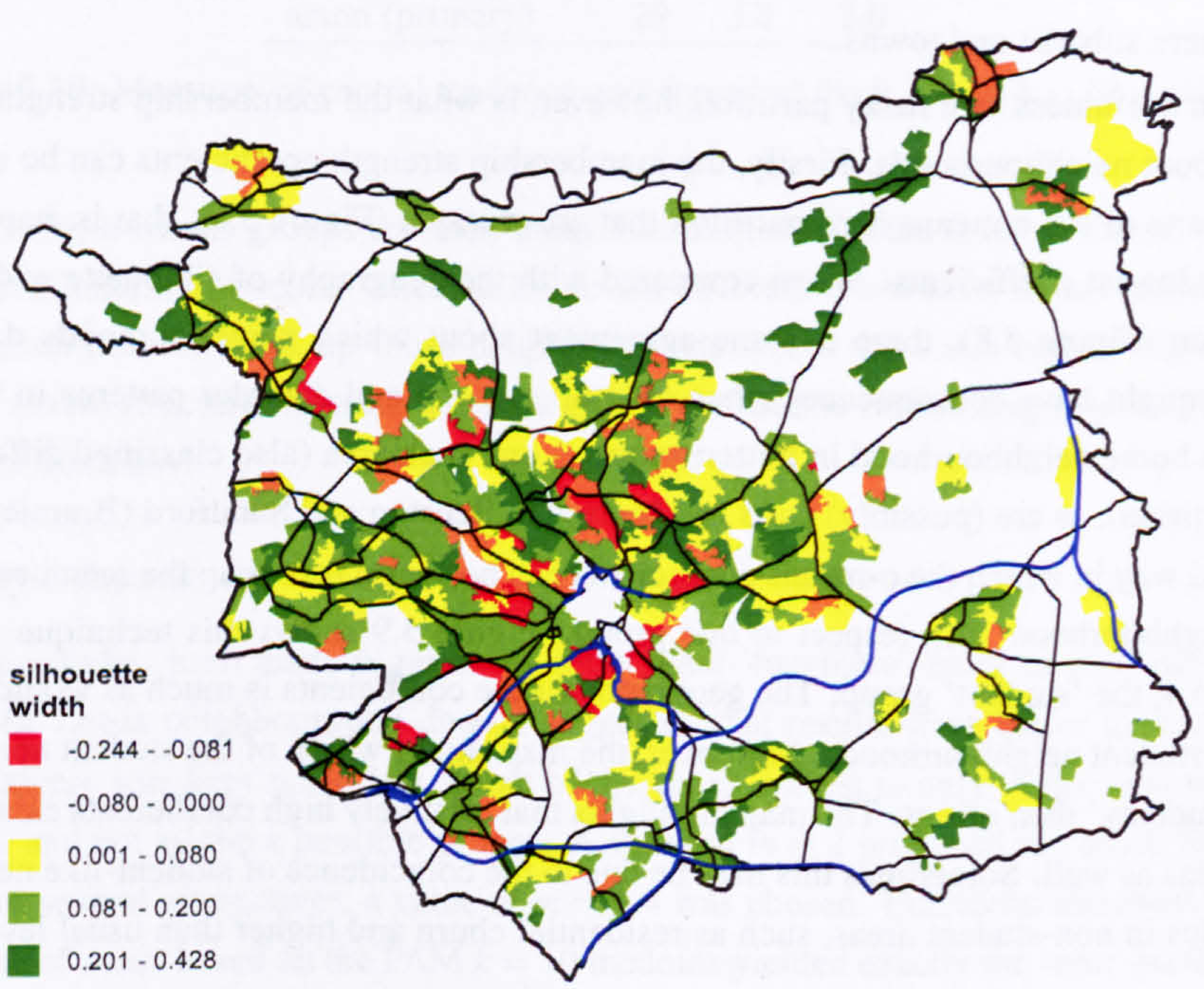


Figure 5.8: Geography of silhouette widths in the PAM $k = 10$ partition.

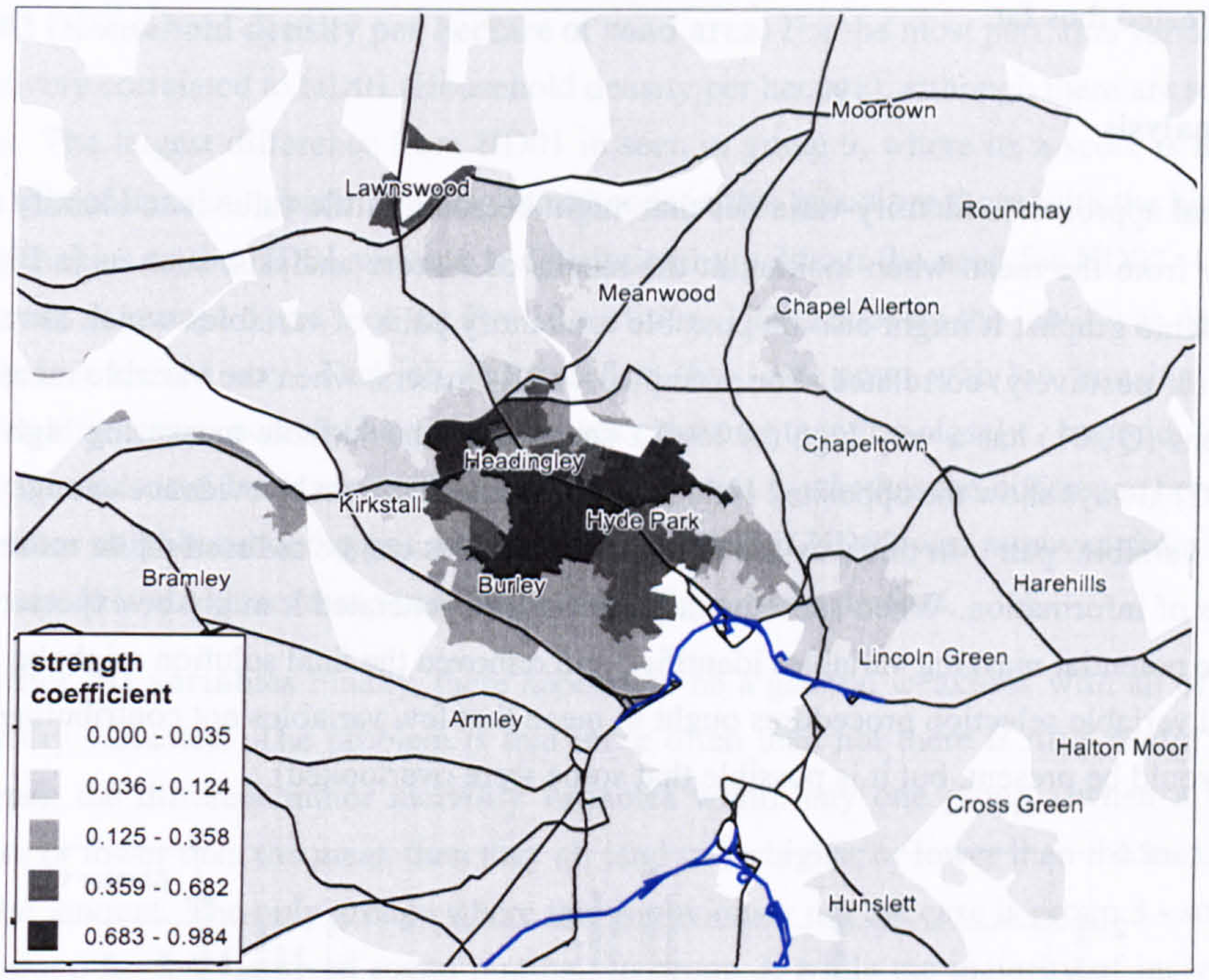


Figure 5.9: Geography of membership strength coefficients for c-means group 4, the 'student' group.

the neighbourhoods in the list were placed in the same type of group by the PAM $k = 10$ partition, but the strength coefficient of their, and other, group(s) shows that they might not be too out of place grouped with other types of neighbourhood.

SCA	c-means	PAM k=10 equiv.	PAM sil	Strength coefficients (group)				
82.01	4	same	-.03	.18 (4)	.13 (8)	.11 (5)	.11 (10)	.10 (9)
20.01	10	6	.14	.16 (10)	.16 (6)			
20.05	8	1	-.19	.16 (8)	.14 (6)	.14 (10)	.12 (3)	.11 (9)
16.02	4	6	.07	.22 (4)	.17 (10)	.13 (6)		
9.04	8	1	-.17	.20 (8)	.18 (10)	.11 (7)	.11 (3)	.10 (9)

Table 5.13: Strength coefficients for a selection of neighbourhoods that are not well fitted to any one cluster.

5.4.7 Revisiting Variable Selection

Through the use of a number of different cluster methods it is possible to begin to identify a suitable partition. Yet, further improvement might still be possible by revisiting the issue of variable selection. In particular, it might be feasible to improve cluster definition by the removal of variables that appear to be adding little of value and, conversely, accentuating the influence of variables which are important. This last task could be achieved by either increasing the weighting of existing variables or considering, with the benefit of hindsight, whether it might be appropriate to look again for new variables that capture a specific aspect of neighbourhood in Leeds that is not being reflected in the partitions created thus far.

Z-Score Analysis

One informal approach to identify variables that might be adding little value is to identify those that rarely stray from the mean when looking at the results of z-score analysis, such as in Figure 5.10. Using the same graphs, it might also be possible to identify pairs of variables which always appear negatively (or positively) correlated. For example, for all clusters, when the variable measuring low qualifications (QU01) has a very high (or low) z-score, does the variable measuring high qualifications (QU02) always show the opposite? If this is the case then it might be evidence enough to remove one of this variable 'pair'. In this way the number of variables can be reduced while minimising the overall loss of information. When subsequent partitions are generated it might be expected that with more of the potential masking variables identified and removed the final solution might be improved. (The initial variable selection procedures ought to mean that few variables not contributing to cluster structure would be present, but it is possible that some were overlooked).

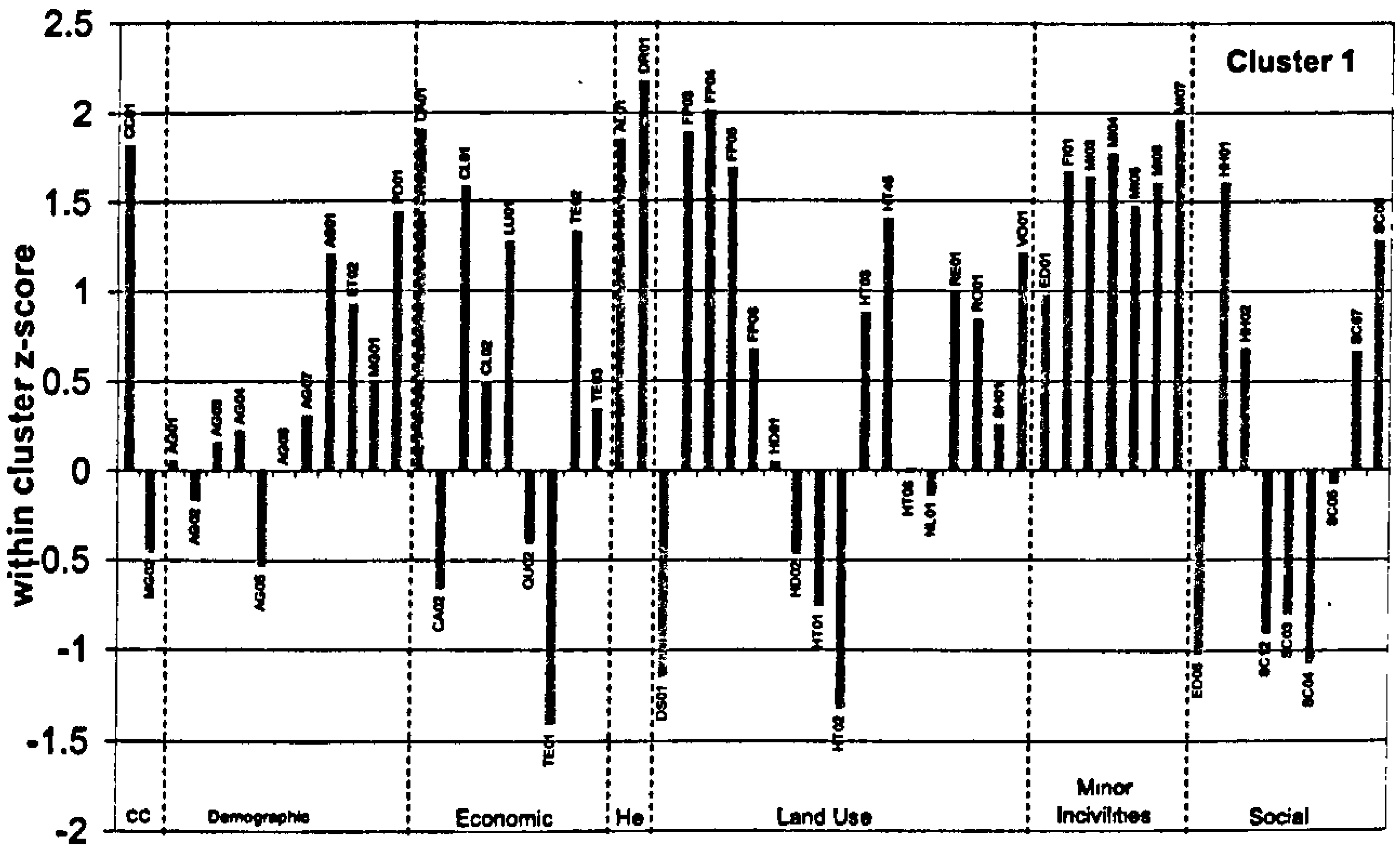


Figure 5.10: Bar graph showing importance of individual variables to the definition of cluster 1

Using this type of subjective assessment, a number of variables that might be masking the $k = 10$ PAM partition were identified:

- **RE01 (Ratio of residential to non-residential delivery points)** Although this variable is prominent in the profile of the deprived inner-city cluster (8), elsewhere its within-cluster mean rarely varies from the district mean. Cluster 8 has many variables with means significantly above the district mean, so the removal of RE01 ought not to have too detrimental an effect here.

- **PD01 (Ratio of daytime to residential population.)** Arguments for the removal of this variable are exactly the same as for RE01.

- **MI06 (Traffic incidents)** There are more general problems with the ways in which the minor incivilities variables are affecting the clusters, but on a visual inspection of cluster profiles it is MI06 that appears to vary least from the district mean from one cluster to another.

- **HD02 (Household density per hectare of road area)** For the most part, this variable appears positively correlated to HD01 (Household density per hectare), although there are some exceptions. The largest difference from HD01 in seen in group 9, where its z-score is much lower than that of its counterpart. The neighbourhoods in this group *are* those with the highest housing densities so the HD01 variable is discriminating without the need for HD02. Conversely, however, in groups 5 and 7 of the $k = 10$ partition, HD02 does do the job it was designed for, albeit in a limited way. That job was to reflect that some areas with low housing density (in simple area terms) actually have their houses clustered together closely - but with larger areas on non-residential land between the clusters. At issue is whether the difference between these and other neighbourhood types would be as a distinct if HD02 was removed from the cluster analysis. It is not clear from visual inspection of the group profiles.

- **all other MI variables** Finally, there appears to be a general weakness with all of the minor incivility variables. The problem is that more often than not there is little useful difference between the different minor incivility variables within any one group - when a variable is higher or lower than the mean then they all tend to be higher or lower than the mean and by a similar amount. The only groups where this is obviously not the case is group 3 - studentland, and group 6 - less deprived social renting. In group 3, while the majority of incivility types are below the district mean, MI03 (noise complaints) is appreciably higher. In group 6, noise complaints and disorder are both somewhat higher than the district mean (abandoned vehicle and ASB to a lesser extent) but MI06 (traffic incidents) is below the mean. At the outset, it was hoped there might be more such contrasts. It is not clear whether the cases outlined are sufficient justification to retain all these variables as they are. Possible alternatives would be to remove or merge some of the variables.

Thus, a partition was created with variables RE01, PD01, MI06, HD02 and MI05 (disorder incidents) removed. Using a scree plot of silhouette widths for range of k the optimum number of clusters changed to 8. For this partition, the overall silhouette width was 0.11, no improvement on either the original PAM $k = 10$ or c-means hard partition. Comparison with the crime-only partition yielded an Adjusted Rand index of 0.14, only a very slight improvement on previous partitions. Results of the second crime heuristic were marginally worse (10 to 13) than the original PAM $k = 10$ partition and worse (8 to 15) than the c-means hard partition. On this basis, the decision was taken not to remove any of these variables unless the final suite of tests provided stronger evidence for their removal.

5.4.8 OVW and HINoV

Section 5.3.1 introduced two mathematical techniques, OVW and HINoV, for calculating variable weightings. The decision was taken not to use these methods for the initial variable selection because of concern that they might not be appropriate for the Leeds neighbourhood dataset. A richer literature on the application of these techniques may have allayed fears, but such accounts could not be found. Furthermore, it could be argued that the complexity of the techniques might obfuscate a process that was deliberately designed at the outset to be kept clear and easy to follow. These concerns diminish somewhat, however, once a partition has been created and the emphasis has shifted to understanding which variables have been over and under-performing. Thus, it is argued that techniques like OVW and HINoV may be able to provide additional evidence to support an investigation into whether noisy/masking variables may still be present in the dataset, post-partitioning.

The OVM method was tested using an MS-DOS programme written by Makarenkov and Legendre (2001). OVW takes as input a rectangular object-variable matrix containing the measurements for the 477 objects (neighbourhoods) on the 56 variables, plus the group membership for each object. A Polak-Ribiere (Press et al., 1986) minimisation procedure is then used to produce and optimal set of weights that minimise the k-means objective function. Each weight is nonnegative and the sum of the weights is equal to 1.

The results of tests with OVW on Leeds neighbourhood data were disappointing. Despite an option within the OVW software to place a limit on the size of any one weight, the results gave almost all the weight (0.995383) to just one variable, *AG03 Proportion of the Population Aged 15 - 24*. The variables that took the remaining fractions appear to be appropriate, but with 37 variables having a weight of 0.000001 or less the results from OVW appeared of little practical value. The finding by Makarenkov and Legendre (2001) that OVW is not suitable for data that contains error-perturbed variables or outliers may explain some of the problem experienced here. Unfortunately, with the apparent failure of the one mechanism (maximum weight) that might have helped investigate the OVW result further, the decision was taken to abandon further attempts and move on to consider HINoV.

Macros for HINoV were obtained for the SAS System (SAS Institute Inc.) from the original authors (Professor Ali Kara, Pennsylvania State University, personal communication, November 25, 2004). To recap, unlike OVW, HINoV does not require that the object/variable matrix be supplied along with the results of a partition. It does, however, require the user to decide on a value for *k*, and this was set to 10. Figure 5.11 shows the TOPRI values calculated by HINoV and the graph can be used to identify variables with the lowest values. These, we might be led to believe, are noisy variables that ought to be removed. Thus, the most likely candidates for removal were FP06, AG07 and CC01. On reflection, *FP06 Largest building footprints* is perhaps worth merging with FP05, as buildings over 8,000 square metres are rare. AG07 (people aged 85+) is also a fairly small population and could be merged with AG06 (people aged 65-84). Variable *CC01 People who work close to home* is one of only two variables that were chosen to try and represent community cohesion, so it was thought inappropriate to remove it.

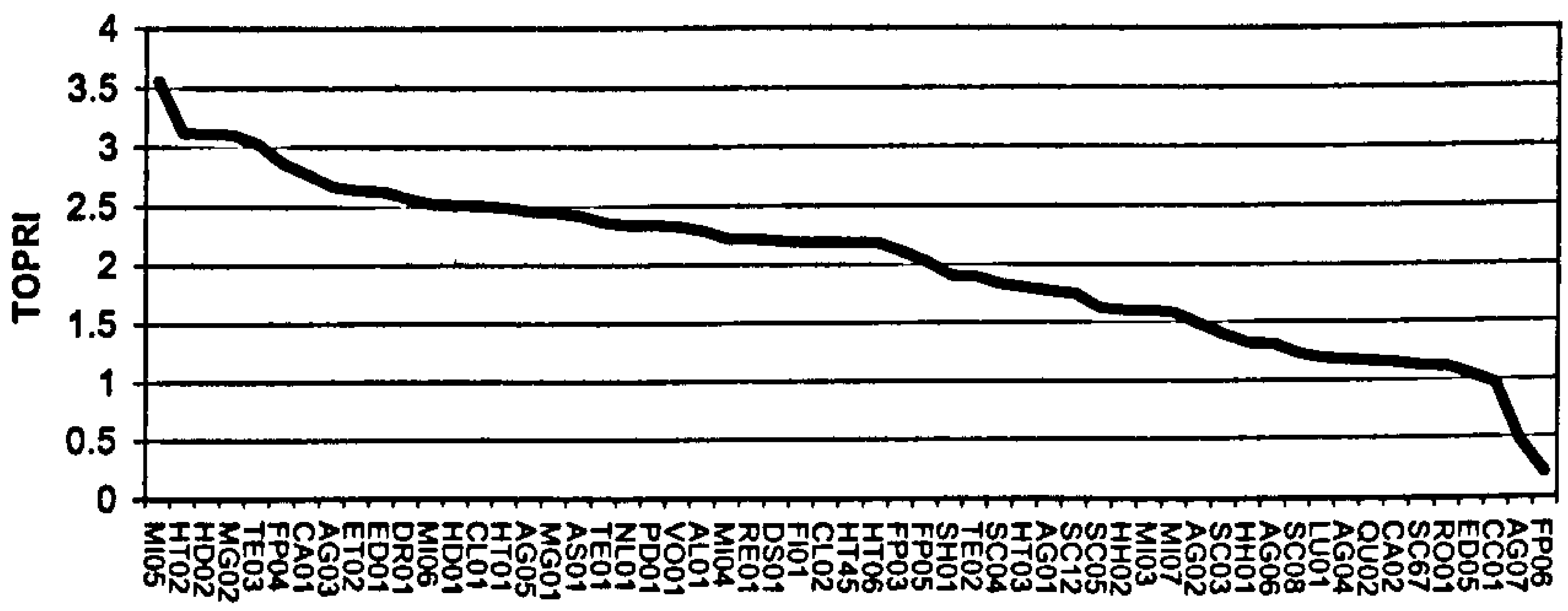


Figure 5.11: Scree plot of HINoV TOPRI values.

At this point, although HINoV certainly appears more useful than OVW, it is not without limitations. These have been pointed out by Brusco and Cradit (2001) in an exposition of their own variable-selection heuristic, VS-KM. They take issue with the subjective reading of scree plots, but the main limitations they point out are the likelihood of problems when variables are highly correlated and when there are more than one true cluster structure in the data. At this point in the research the second of these issues is not really a concern. In addition, the problem of high correlation between variables ought to have been mitigated to a large extent by earlier pre-cluster analysis variable selection procedures, but the problem cannot be ruled out altogether. Thus, as with many aspects of cluster analysis, care needs to be taken when interpreting the HINoV results. Retrospectively, it could be argued this lends further support to the strategy of keeping cluster analysis procedures clear and simple until structures in the data are better understood. Tests with VS-KM, a technique purported to improve upon HINoV, were not undertaken. The authors of the technique were contacted regarding access to the Fortran source code for the VS-KM algorithm, but no response was forthcoming.

A partition was created to accommodate the removal of the variables suggested by by the lowest TOPRI scores. The final result was closer to the c-means partition than the PAM10 partition when measured using the Rand Index test. There was no improvement over the c-means hard partition in either overall silhouette width (0.12) or agreement with the crime-only partition (Adjusted Rand .114). By the second set of crime heuristics the HINoV partition was also slightly worse than the PAM10 and c-means partitions. As a result, the variables identified by HINoV as being weak were not removed.

5.4.9 Final Partition Summary

At some point, a decision must be made about which partition to proceed with. Experiments with adding and removing small numbers of variables could be continued, but some consideration must be given to the law of diminishing returns. Furthermore, the next stage in the design of a final classification - utilising the partition to analyse real community safety issues - might generate better ideas about how a final partition could be improved than continued tinkering in an ad hoc manner. Thus, the decision was taken to proceed with the $k = 10$ fuzzy partition produced by c-means using the original 57 variables.

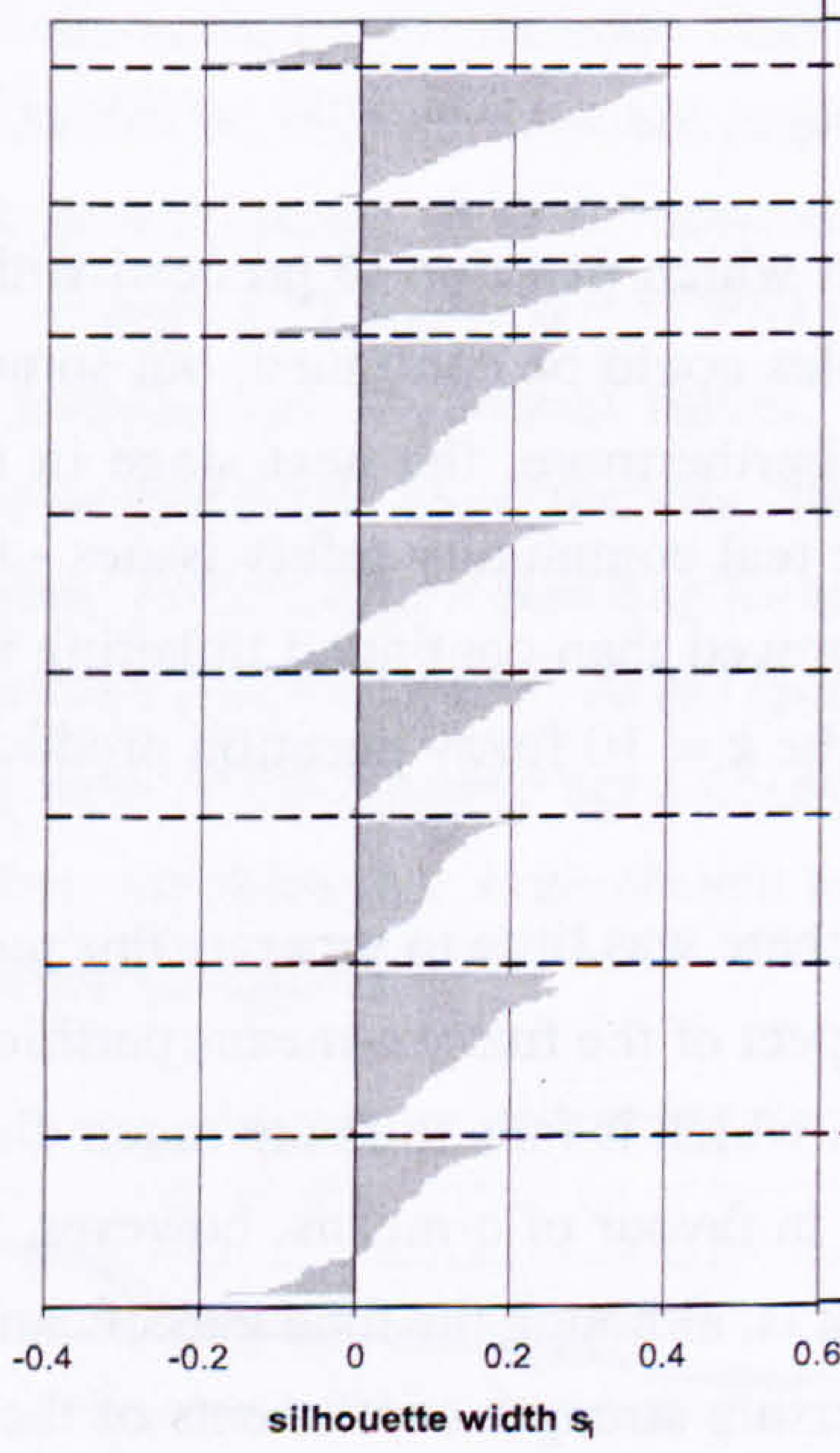
On balance of the various tests conducted, there was little to separate this partition with the original $k = 10$ partition produced by PAM. One aspect of the fuzzy c-means partition that is perhaps less desirable than the PAM partition is the way in which it fails to make much distinction between the more affluent suburbs in the north of the city. In favour of c-means, however, is the extra utility afforded by the fuzzy nature of the partition. That is, although the final classification assigns individual neighbourhoods to just one group, the membership strength coefficients of the fuzzy c-means solution provide a useful diagnostic tool with which to analyse further the nature of the assignment and possible alternative assignments to other groups. An important lesson from the various experiments described in this chapter it is that some types of neighbourhood, and some specific neighbourhoods themselves, are not easy to classify reliably. Such problems *are* tacitly accepted when embarking on a classification exercise, yet the fuzzy classification is more reflective of, and responsive to, such problems than hard partitioning techniques.

Further work could be undertaken to compare the fuzzy c-means approach with other fuzzy classifiers, such as FANNY (Kaufman and Rousseeuw, 1990). While fuzzy c-means minimizes the fuzzy sum of squares of clusters based on the Euclidean distance between an object and the centroid of its cluster, the FANNY algorithm minimizes the fuzzy equivalent of a sum of error functions for each cluster (Nicholls and Tudorancea, 2001). Advantages of the FANNY algorithm over the fuzzy c-means algorithm include its lower sensitivity to outliers and a better recognition of non-spherical clusters (Trauwaert, 1987). Some initial experiments were undertaken using the implementation of FANNY for the R statistical package. At the time when the experiments were conducted, however, the software routine had a limit of 100 cases, making it unsuitable for this research. Subsequently, a number of Kaufman and Rousseeuw routines have been updated for the R cluster package and the implementation of FANNY no longer has a case limit .

5.4.10 Final Partition Information

Information about the size and goodness of the final 10 groups, as measured by silhouette widths, is shown in Figure 5.12. The final partition has resulted in fairly equal sized groups and this ought to cause fewer problems for between-cluster comparisons of crime patterns. Group 3, with just 20 neighbourhoods, is rather small but its distinctiveness and strength of definition ought to mitigate for its lack of members.

Figure 5.13 shows the geography of hard membership of the different groups. The geography of the strength coefficients to create the hard partition have been shown previously (Figure 5.7) and are displayed again for individual groups in the portraits in the following chapter.



j	n_j	mean s_i	medoid SCA
1	25	.03	5.06
2	49	.19	7.02
3	20	.23	45.02
4	26	.19	48.12
5	68	.13	29.02
6	59	.08	66.05
7	52	.11	86.03
8	57	0.08	88.02
9	61	0.13	60.02
10	60	0.04	67.07

Figure 5.12: Silhouette plot and group membership information of the final c-means $k = 10$ hard partition

5.4.11 Reductions in Variability

Finally, as has been mentioned before, it is a central tenet of classification that groups ought to contain objects that are similar to one another (compact) but that groups as a whole should be distinct and distant from one another (separate) (Cormack, 1971). However, it has been pointed out that,"after geodemographic classification the differences between the various neighbourhood classes are frequently less than the differences within the neighbourhood classes" (Harris, 2001, page 329). Voas and Williamson (2001) demonstrated these problems using the GB Profiles and Super Profiles classifications and Harris (2001) showed that the problem was even greater when consideration was given to the household or individual-level diversity within small areas (census EDs).

Voas and Williamson use distance analysis to come to their conclusion. This involves determining the 'norm' for a given study area, calculated by finding the point in multi-dimensional space defined by the mean value of the variables defining each dimension. To analyse the differences within and between LCCS groups it is necessary to calculate the norm for each individual group separately. The mean distance between the group norms then provides a measure of group separation. This is different from the silhouette coefficient developed by Kaufman and Rousseeuw (1990) used elsewhere in this thesis, which, to recap, measures the difference between the average dissimilarity between a neighbourhood and all the other neighbourhoods in its group, and the average dissimilarity of the object to all the neighbourhoods in the next nearest group.

The mean distance between a group norm and the norms for each of its constituent neighbourhoods gives a measure of compactness for each group - which can themselves be averaged to provide a global measure of compactness. Where the number of zones in groups varies it is necessary to weight the calculations by the size of the group membership.

For the LCCS partition the distance was measured between every pair of neighbourhoods for each group. Using the mean distance for each group weighted by the number of SCAs in the group, the average within-group distance was calculated. This value was 67.7% of the average distance between any two neighbourhoods across Leeds as a whole (i.e. no segmentation). Thus, the global reduction of 32.3% in variability achieved by the LCCS is not large when considering the dataset of 57 variables as a whole. To provide a benchmark for this figure the calculations were repeated using a classification constructed from the deciles of a ranked list of Townsend deprivation scores calculated

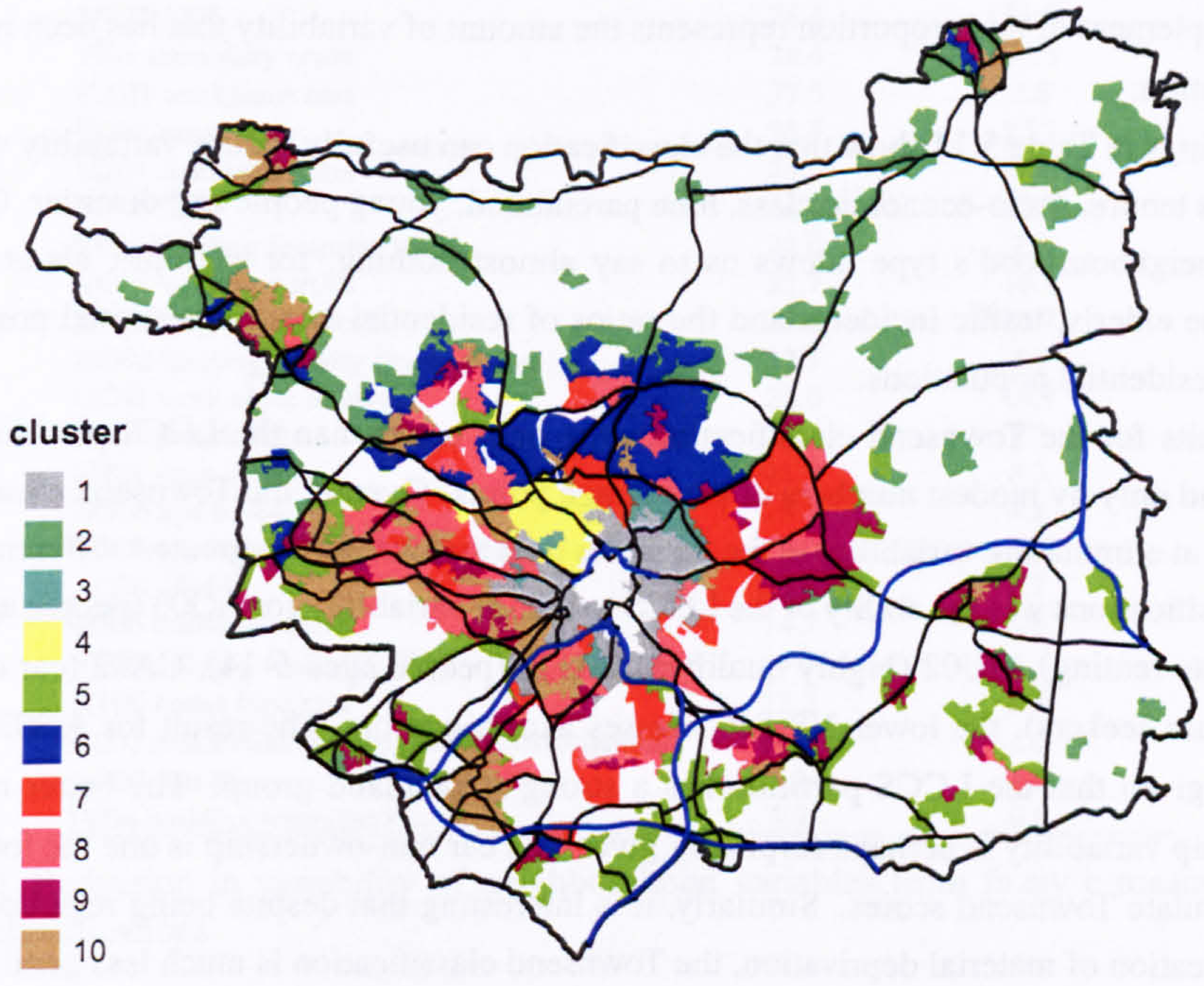


Figure 5.13: Geography of group membership of the final c-means $k = 10$ hard partition

for Leeds. For this deprivation segmentation the global reduction in variability is only 19%, a worse performance than the LCCS.

In the tests by Voas and Williamson, the global reduction in variability was 18% for GB Profiles and 16% for the Super Profiles classification. Given the relative diversities of the populations of Leeds and of England and Wales as a whole it is perhaps no surprise that reductions in variability of the national classifications are not as good as the LCCS. With a more homegeneous population in Leeds the classification does not have 'work' so hard. Then again, the Leeds classification only has 10 groups compared to 99 (in England and Wales) for GB Profiles and 128 for Super Profiles. Thus, it might be expected that the Leeds classification would have much more within-group diversity.

As well as considering distance in a multi-dimensional space, it is possible to examine differences along each dimension, or axis individually. Table 5.14 shows the average reduction in dispersion for each variable achieved by the LCCS partition. Again, the figures for the Townsend deprivation classification are included for comparison. The procedure for calculating the values in the tables is as follows:

1. Calculate the absolute difference between variable values for every pair of neighbourhoods within each group;

2. Average the values from (1) to produce a mean difference for each group;

3. Average the values from (2), weighted by the number of neighbourhoods in the group, to produce a mean within-group difference;

4. Divide the value from (3) by the overall mean absolute difference (i.e. between variable values for all pairs of neighbourhoods) to give the proportion of variability that remains after classification.

The complement of this proportion represents the amount of variability that has been removed by the classification.

The findings in Table 5.14 show that the classification can usefully reduce variability within variables such as tenure, socio-economic class, lone parenthood, young people and disorder. Conversely, knowing a neighbourhood's type allows us to say almost nothing, for example, about very large buildings, the elderly, traffic incidents and the ratios of residential to non-residential postcodes and daytime to residential populations.

The results for the Townsend classification were only better than the LCCS partition for three variables, and only by modest numbers of percentage points. Overall, the Townsend classification is much worse at eliminating variability in the variables than the LCCS. The greatest difference between the two classifications was the ability of the LCCS to reduce variability for AG03 (people aged 15-24), TE03 (private renting), QU02 (highly qualified), AG02 (people aged 5-14), CA02 (car ownership), AS01 (asylum seekers), the lower NS-SeC classes and migration. The result for AG03 is perhaps predictable given that the LCCS partition has a strong studentland group. The better reduction in car ownership variability is perhaps surprising given that car non-ownership is one the four variables used to calculate Townsend scores. Similarly, it is interesting that despite being regarded as a good general indication of material deprivation, the Townsend classification is much less good at reducing variability for the lower social class variables.

variable	reduction in variability (%) c-means	Townsend
TE01 owner occupier	57.1	73.2
TE02 social renting	55.1	48.6
AG03 age 15-24	53.9	9.0
SC67 NS-SeC 5/6	53.0	23.5
SC12 NS-SeC 1	52.5	43.2
HH02 lone parents	50.8	31.0
DS01 gardens	50.1	30.5
SC08 unemployed	5.0	40.1
TE03 private renting	46.1	9.2
MI05 disorder	44.7	30.3
AG05 age 45-64	44.2	30.7
HT01 detached houses	44.1	40.6
SC04 NS-SeC 3	43.1	32.9
QU02 highly qualified	42.1	13.4
SC05 NS-SeC 4	41.1	9.9
RO01 road density	39.9	15.8
ED05 GCSE performance	39.1	36.3
AS01 asylum seekers	38.4	8.6
HD01 housing density	38.4	12.8
MI03 noise complaints	38.3	34.5
CL01 unemployment benefit	37.9	28.0
DR01 drug episodes	34.6	25.0
SC03 NS-SeC 2	34.5	26.3
AG02 age 5-14	34.4	6.1
HT02 semi-detached houses	34.1	11.7
HT03 terraced houses	33.3	12.6
ET02 ethnic mix	32.8	15.1
AL01 alcohol episodes	32.8	28.2
MG02 not moved house	32.1	6.9
MG01 international in-migrant	32.1	6.9
NL01 natural land	32.1	8.0
CA02 resdidential cars	31.9	2.9
CL02 falling unemployment	30.5	23.7
FP03 building footprint 150-1000	29.9	7.9
AG01 age 0-4	28.9	3.8
MI07 ASB	28.4	21.6
FI01 secondary arson	28.4	16.3
CA01 workplace cars	27.5	2.8
ED01 exclusions	26.5	25.3
MI04 abandoned cars	25.3	20.4
HT45 flats	24.0	25.2
FP04 building footprint 1k-3k	24.0	2.8
VO01 void properties	22.7	16.1
HH01 single households	22.6	26.7
HD02 housing density (by road area)	22.3	4.7
CC01 work close to home	22.0	16.7
LU01 pub density	20.7	9.7
SH01 shop density	18.3	6.3
AG06 age 65-84	18.1	4.8
PD01 daytime/residential pop. ratio	15.2	-0.7
AG04 25-44	15.0	3.7
MI06 traffic incidents	14.7	8.8
FP05 building footprint 3k-8k	14.7	-1.4
HT06 house type mix	13.5	3.3
RE01 residential/non-residential postcode ratio	9.5	-1.3
AG07 age 85+	6.1	7.3
FP06 building footprint 8k+	3.7	-9.9

Table 5.14: Reduction in variability of neighbourhood variables from fuzzy c-means (hard) and Townsend classifications

What these findings also prompt is renewed uncertainty about whether certain variables are lending anything at all to the partition and indeed whether the continued use of some of these variables - e.g. AG07, RE01, HT06 - might actually be masking the importance of other variables. Voas and Williamson investigated this issue by creating a simple classification using just a handful of variables. This produced an average within-group distance that was almost identical to that achieved by GB Profiles or Super Profiles.

5.5 Concluding Remarks

This chapter began with reference to the argument that cluster analysis is the art of finding groups in data. Whether an art or a craft, it is argued that a balance needs to be struck between adhering strictly to the recommendations from formal mathematical techniques, and some amount of expert knowledge of the study area and the problem domain - in this case, crime. Furthermore, the needs of the end-user must be considered, and they may have less interest and/or faith in a overly complex formal solution, but are also likely to have more expert knowledge of the problem domain than the classification designer.

To reflect these issues, a number of alternative techniques were applied at each stage in the design of the LCCS. Results were then compared for levels of agreement, enabling cluster analysis to progress with more confidence that the final classification would be likely to be representative of genuine similarities between neighbourhoods.

The use of the R statistical software enabled the use of techniques that would have required considerable programming had all the analysis been conducted within a commercial software package, such as SPSS or Minitab. The open nature of the R project makes it a popular platform for developing and sharing quantitative techniques, and its ability to handle large data matrices makes it very suitable for cluster analysis projects.

The use of each of the three different clustering techniques helped to understand better how the neighbourhoods of Leeds were being grouped, but it is suggested that the fuzzy c-means technique offers the most information with which to analyse partition structure. Thus, the final partition selected for the LCCS was generated using fuzzy c-means, although it was a hard partition that was selected, and the fuzzy components were only used for testing, rather than than being used to create a true fuzzy classifier.

Finally, the reduction in variability tests show that the LCCS is better at reducing the variability in the data than a simple classification created from deprivation score deciles. Comparisons with other classifications are discussed in more detail in Chapter 6, but from a design perspective, the results presented here continue to raise questions about how many and what type of variables are actually needed to create a useful partition.

Chapter 6

Portraits and Classification Testing

6.1 Introduction

An important task after producing a classification of any sort is to ascertain what is unique about each group and what makes groups distinct from one another. For a general purpose classification, the creation of 'pen portraits' for each group might include a textual description of group characteristics and perhaps a graphical breakdown of the extent to which the mean values of group variables vary from the global mean. The description might also include reference to typical geographical areas where the group is prevalent and may be supplemented with photo montages designed to capture the typical traits of the neighbourhood architecture and the lives of the residents therein. While eye catching, these last devices are not provided below, instead charts help capture the average 'crime profile' of each group and maps of group neighbourhoods depict how strongly neighbourhoods belong to their group.

In addition to this portraiture, the chapter considers whether the Leeds Classification for Community Safety (LCCS) is of value. Since the outset, the hypothesis has been that the design of a task-specific classification for community safety would be at least as good as one of a number of existing alternative classifications, and most hopefully better! This is tested by comparing the LCCS with existing classifications with respect to dependent crime variables. A device borrowed from marketing analysis - the Gains Chart - is used first and this visualises the discriminating power of the different classifications. Some general features of this approach are also discussed and the effect of disaggregating crime categories is analysed.

Variability reduction tests described by Voas and Williamson (2001) in the previous chapter are then applied to some of the different classifications to see the extent to which they each reduce variability in crime rates across Leeds and thus help 'explain' patterns of crime.

6.2 Group Portraits

As well as being a tool for exploration and explanation, visualisation can be used to present and reveal geographic information within a GIS (Unwin and Hearnshaw, 1994). In a similar way, "geodemographic classifications and systems often incorporate extensive visualization tools such as multimedia guides that help the user to interpret the typologies and cluster labeling" (Harris et al., 2005, page 139). No multimedia guide has been developed for the LCCS but a number of graphical devices are presented below in order to provide users of the LCCS with a description of neighbourhood groups; a profile of crime recorded therein; and, the strength with which neighbourhoods belong to their group - along with group geography.

There are potential problems, however, with painting too rich a picture, and some of these stem from the ecological fallacies that many geodemographic classifications are inherently guilty of. Although the LCCS was created using a fuzzy classification technique, the final partition was a hard one based on mutual exclusivity. Much diversity may still exist within the individual groups and so clearly, the theoretical typical resident or household within a group might not represent all the individual residents or households in the group. It becomes important not to paint group portraits that imply a false uniformity.

In particular, it is important to be mindful of what Curry (1998) has termed the "the power of the visual". Is the reader of a chart or map of group geography likely to be less critical of the information being portrayed than if they were looking at the same data but in tabular form? Curry thinks this to be the case and his thesis produces evidence to demonstrate his claims.

There is also the risk of labeling groups using colloquial and highly general terms, such as 'Low Horizons' or 'White Van Culture' - to borrow from Experian's Mosaic. Or how about Charles Booth's 'Lowest class. Vicious semi-criminal', which on his late nineteenth poverty maps of London was coloured black? The problems are twofold. Firstly, there is the risk of slighting residents of areas thus labeled and also offending the sensibilities of public servants trying to provide and deliver services in these areas. Secondly, by attaching negatively charged labels to groups, there is a danger of leading classification users to look for problems in these neighbourhoods when there may be none. The converse is true too, of course - if a group is given a label with positive overtones then there may be a danger that problems that do exist in these neighbourhoods get overlooked, again - the opposite of what this research, in part is hoping to achieve.

It is doubtful whether a classification for community safety could, as a result of labeling, become a self-fulfilling prophecy in the way Goss (1995) describes with respect to retail. In that scenario, the marketers define identity in terms of goods most likely to be bought by people with certain lifestyles and then attempt to sell these goods to these people. If they (the marketers) are successful, then the purchasing habits of the lifestyle group come to conform more closely to the marketers' image of them and the lifestyle portrait becomes more real. There are parallels in criminology, however, and the term *deviance amplification* (Lemert, 1951) was coined to describe instances where an individual comes to believe society's judgement - inferred by a label - and so modifies his or her self-image to match that label. It is not envisaged that the LCCS will be published publicly so the danger of individual offenders responding to any LCCS labels seems unlikely. It is possible to imagine other scenarios however, such as a neighbourhood or group of neighbourhoods being repeatedly targeted for crime prevention work or a particular policing style because of assumptions drawn from geodemographics, as appears to be promoted by some (for example Ashby and Longley, 2005).

Given these considerations, the intention from the outset has been to avoid using labels for groups and simply refer to them by number (1 through 10). There have been temptations to do otherwise as the numbers are abstract and usually have to be embellished with textual labels in order to make their reference meaningful. Curry's (1998) complaints about maps are dealt with to some extent by including with the geography an indication of the fuzzy membership strength, using shading. The other column and bar charts could fall foul of the same problem but some consideration also has to be given to readability and conciseness.

6.2.1 Visualising Group Profiles

The choropleth maps (e.g. Figure 6.1) that appear in the following sections show group membership strength and have all been drawn using the same interval ranges. This allows the maps to be compared with one another more easily and further highlights the general fitness of groups.

The column charts that represent standardised scores for each of the neighbourhood variables are not drawn to the same scale. The variable values on the graphs (e.g. Figure 6.2) represent the average variable value within the group minus the average variable value for all 477 neighbourhoods, the result of which is divided by the standard deviation of all the variable values. Groups are considered to be most defined by variables with the highest and lowest standardised scores.

The bar charts of standardised scores for crime variables are created in the same way as above, but using the crime rates for neighbourhoods. The lighter-coloured bars represent disaggregated parts of major crime groups (darker bars).

Finally, the bar charts of change in standardised crime rates over time are designed to show whether groups exhibit different crime trends - for better, or worse - over recent recording periods.

6.2.2 Group 1

28 Neighbourhoods. Highest strength coefficient is .82, SCA 5.06 - Wortley.

According to the mean silhouette width (.0098), this group is not strongly defined. Yet, with neighbourhoods clustered tightly around the city centre there is a strong coherence to the geography of group membership (Figure 6.1). The only neighbourhood that does not conform to this pattern is Cottingley - a cluster of high-rise council flats and other low-cost housing west of Beeston. The group is characterised (Figure 6.2) by mixed land use, with large numbers of cars parked close to places of work during the daytime (CA01). Unemployment (CL01) is above average and the social class structure is biased toward the lower classes. Home ownership (TE01) is well below average and there are more than the usual proportions of terraced housing (HT03) and flats (HT45), of which a higher than usual percentage are void (VO01). The age profile is similar to the district mean but there is a greater concentration of people from ethnic minorities (ET02) and asylum seeker communities (AS01). People are more likely to work close to home (CC01), which along with higher than average material deprivation might explain relatively low levels of car ownership (CA02). There is a much higher than average ratio of daytime to residential population (PD01) and residential stability (MG02) is lower than average (second only to student neighbourhoods). Much higher than average levels of narcotics use (DR01) and minor incivilities (all types) are in evidence.

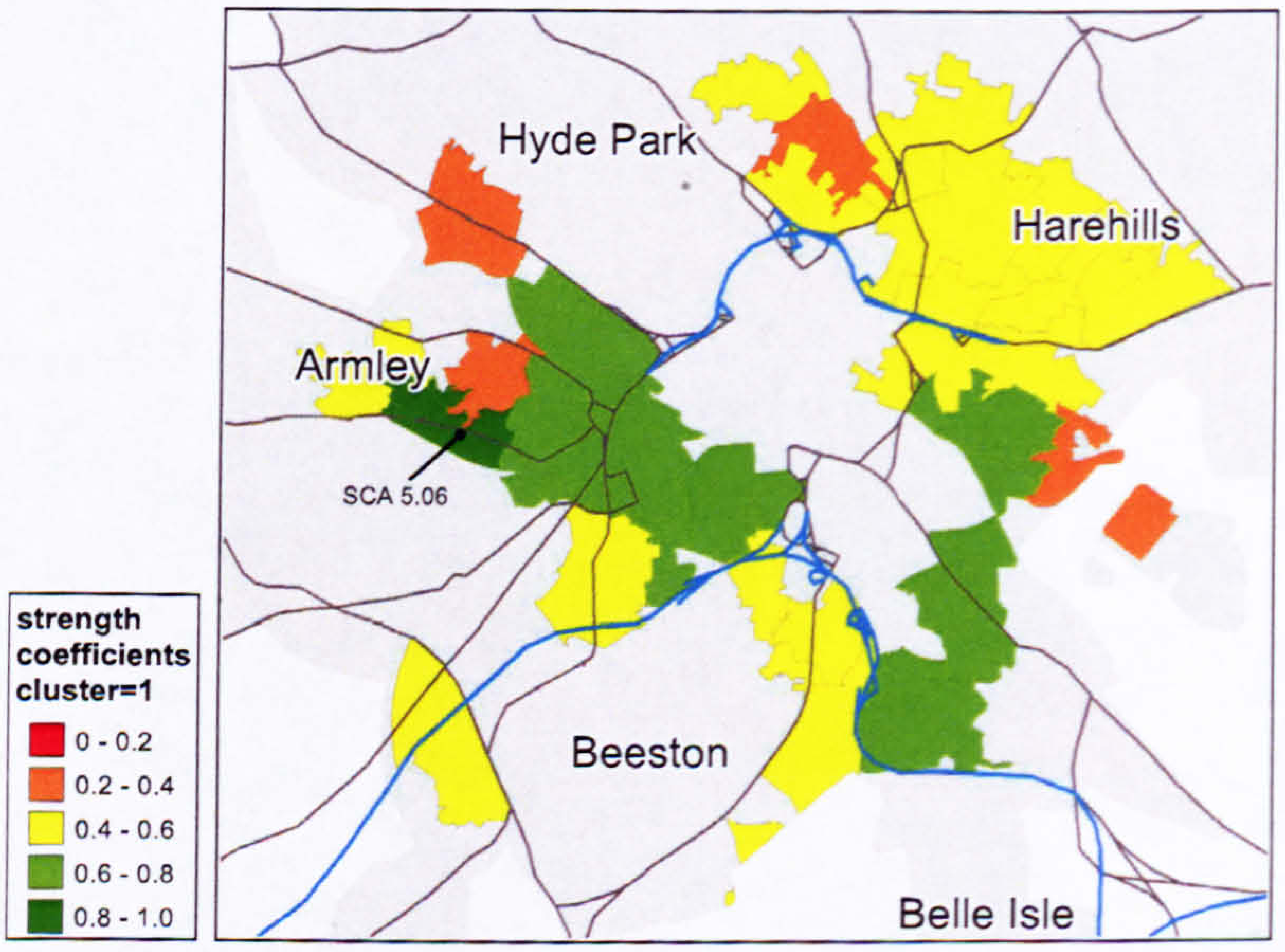


Figure 6.1: Geography and membership strength coefficients of neighbourhoods in group 1.

Of the aggregate crime rate variables, *burglary dwelling* is the closest to Leeds mean, while all other types are well over (Figure 6.3). Disaggregation of *burglary elsewhere* shows that burgling from residential out-houses is lower than average, which fits with the housing type profile for group 1. And although *theft from a shop* appears closer to the Leeds mean than other disaggregate types this is still a high standardised score when compared to other group crime profiles, as will be seen.

In 2004/05, most of the crime types are closer towards the Leeds mean, violent crime being the only exception (Figure 6.4). From a policy perspective, the falls in criminal damage and burglary dwelling look particularly interesting and may in part reflect increased levels of patrolling in these areas by police and neighbourhood wardens.

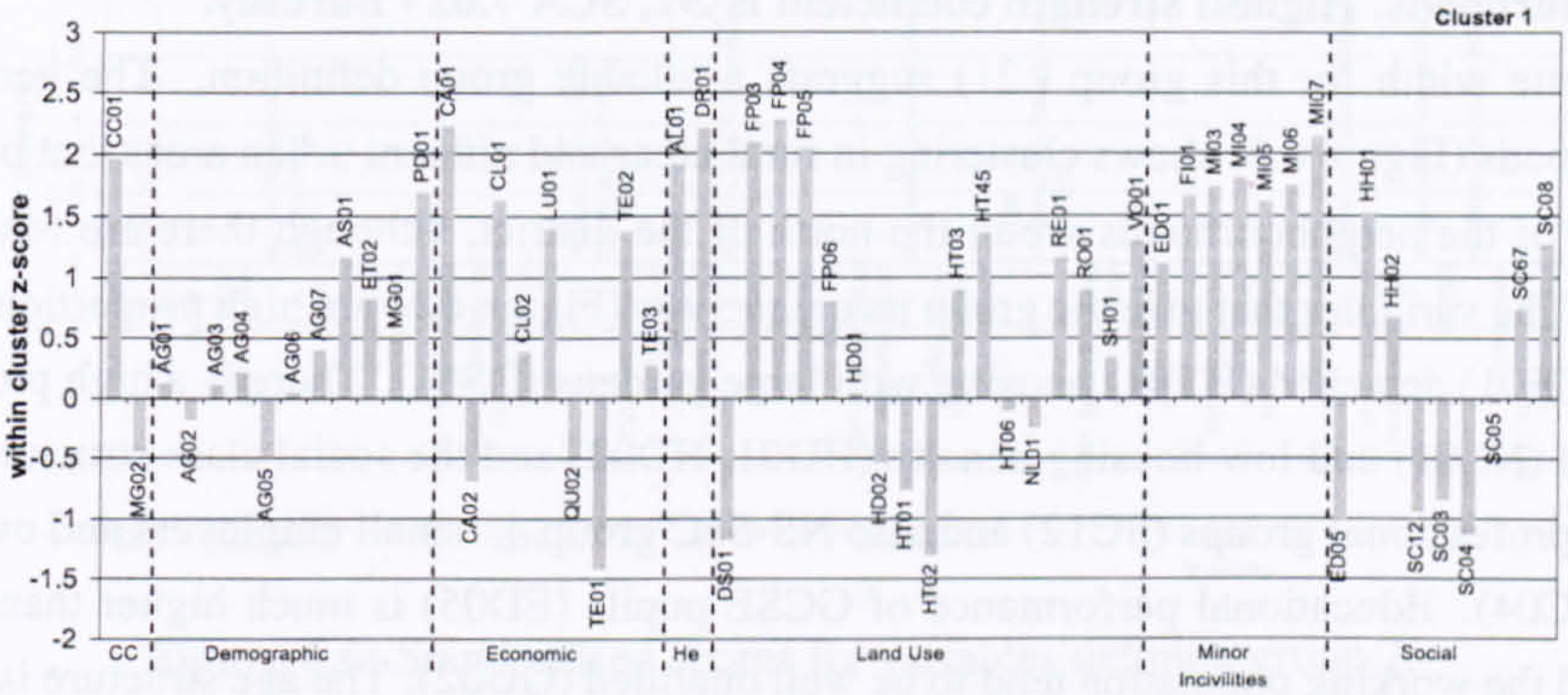


Figure 6.2: Standardised scores for variables defining group 1.

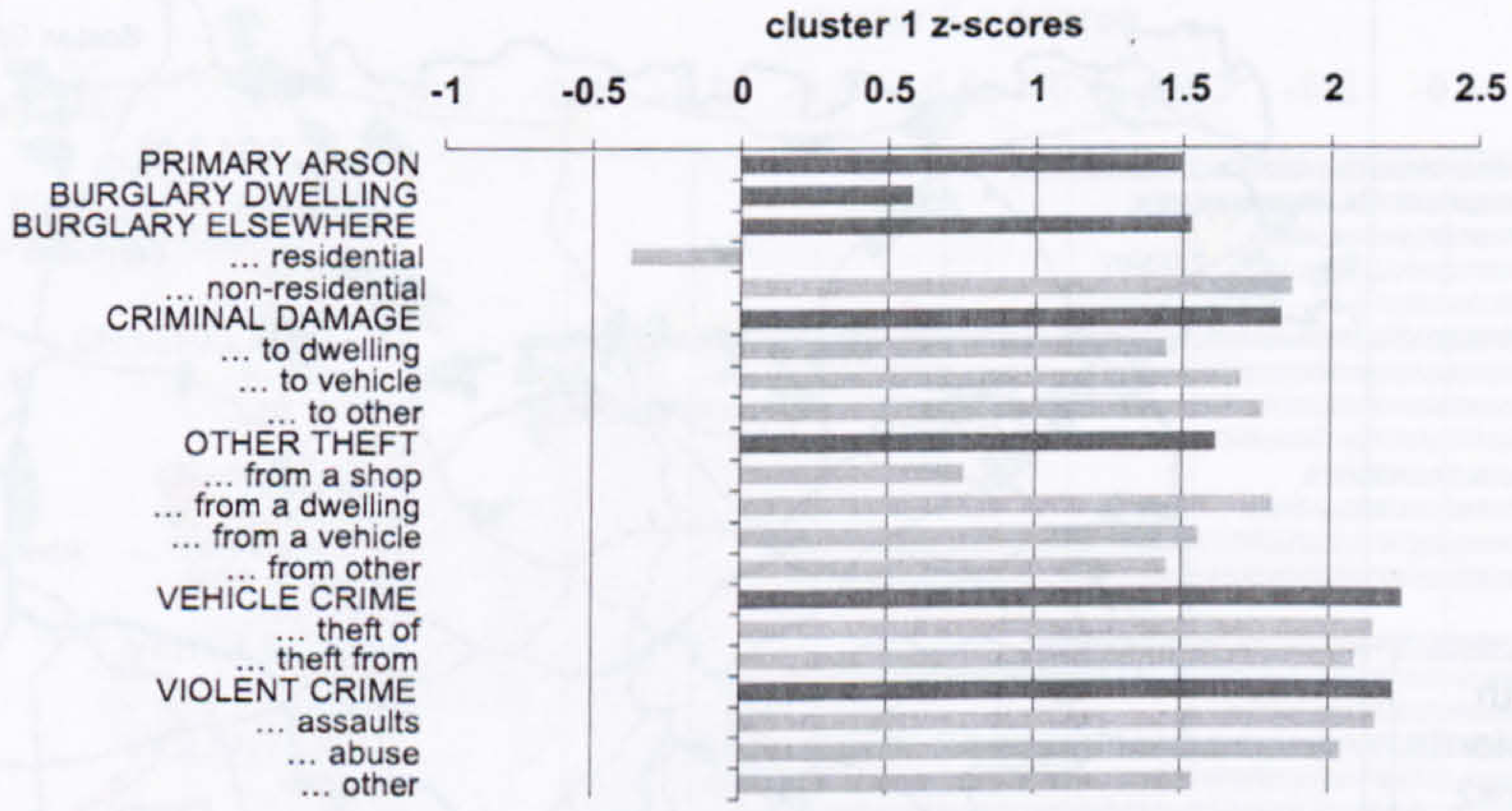


Figure 6.3: Standardised scores for crime variables, for 2003/04, for group 1.

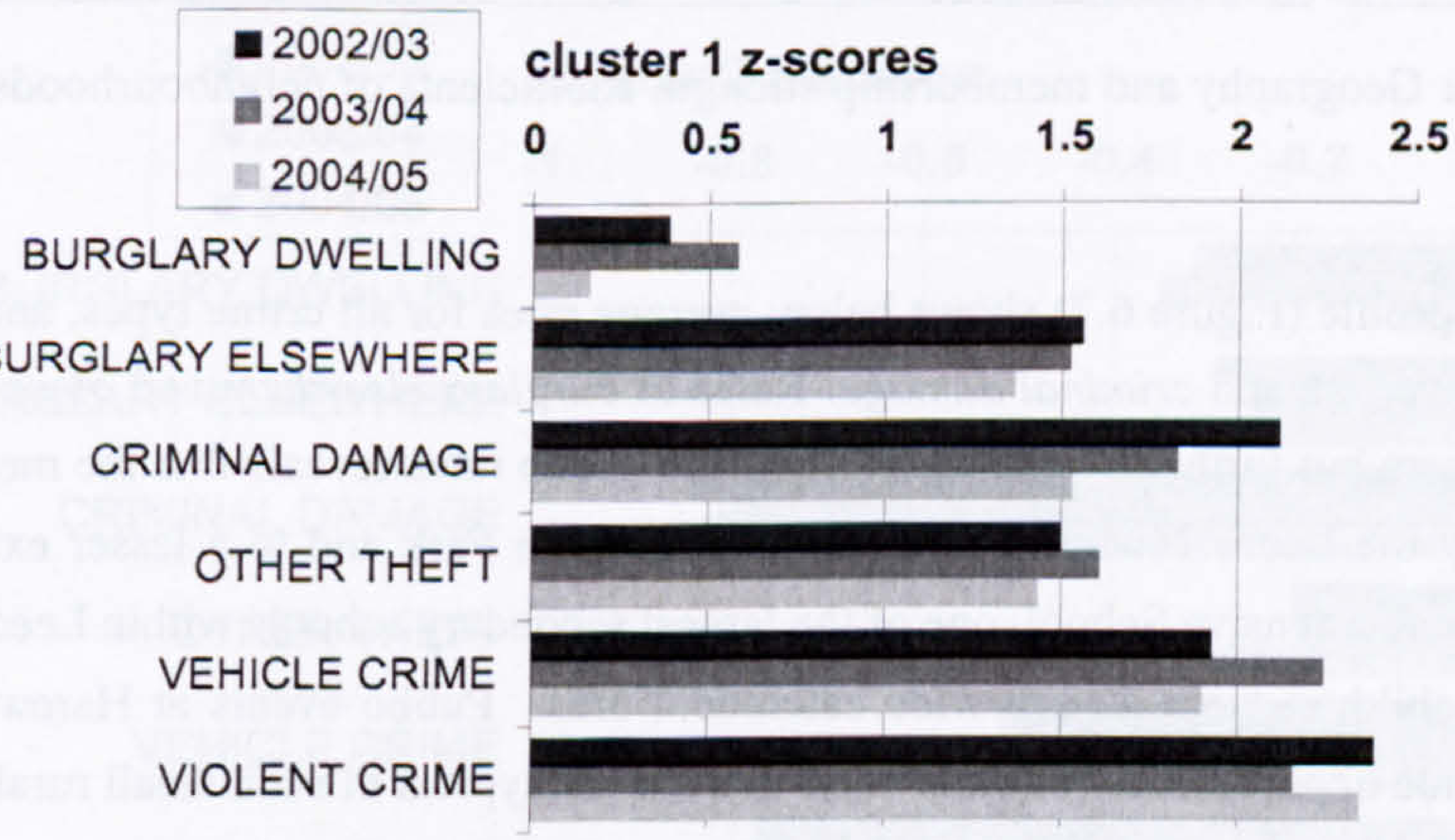


Figure 6.4: Change in group 1 standardised crime rates, 2002/03 to 2004/05.

6.2.3 Group 2

49 Neighbourhoods. Highest strength coefficient is .97, SCA 7.02 - Bardsey.

The silhouette width for this group (.21) suggests a reliable group definition. The geography of neighbourhoods (Figure 6.5) shows clustering in rural areas and affluent urban areas that border rural land. Most of the neighbourhoods are in the north of the district, although there are few in the far south-east. The variables that give the group its uniqueness (Figure 6.6) are high proportions of owner occupied (TE01) detached (HT01) housing with large gardens (DS01). There is a high proportion of natural land (NL01) and low housing density (HD01, HD02) and the social class structure is biased toward the professional groups (SC12) and also NS-SeC group 4 - small employers and own account workers (SC04). Educational performance of GCSE pupils (ED05) is much higher than the Leeds average and the working population tend to be well qualified (QU02). The age structure is somewhat biased toward people in middle age.

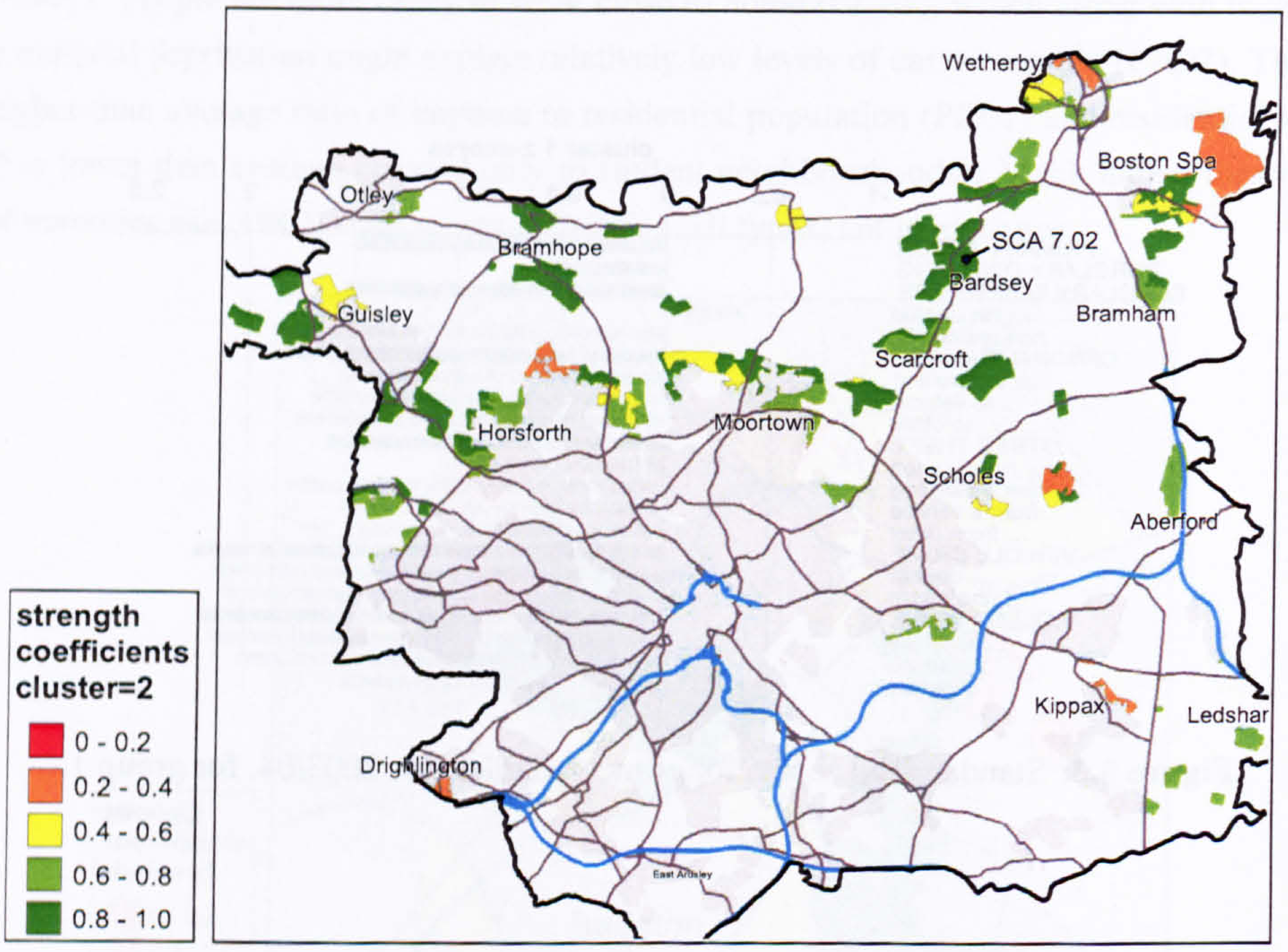


Figure 6.5: Geography and membership strength coefficients of neighbourhoods in group 2.

The crime profile (Figure 6.7) shows below average rates for all crime types, and especially low rates for *violent crime* and *criminal damage*. Rates of *burglary elsewhere* and *other theft* are closest to the Leeds norm but further investigation of within-group rates reveals that the mean theft rate has been biased by the Leeds Music Festival out at Bramham Park and to a lesser extent by thefts at Boston Spa Comprehensive School, one of the largest secondary schools within Leeds and providing education for children from a very wide catchment area. Public events at Harewood House also appear to provide opportunities for *other theft* that are not typical of such small rural settlements.

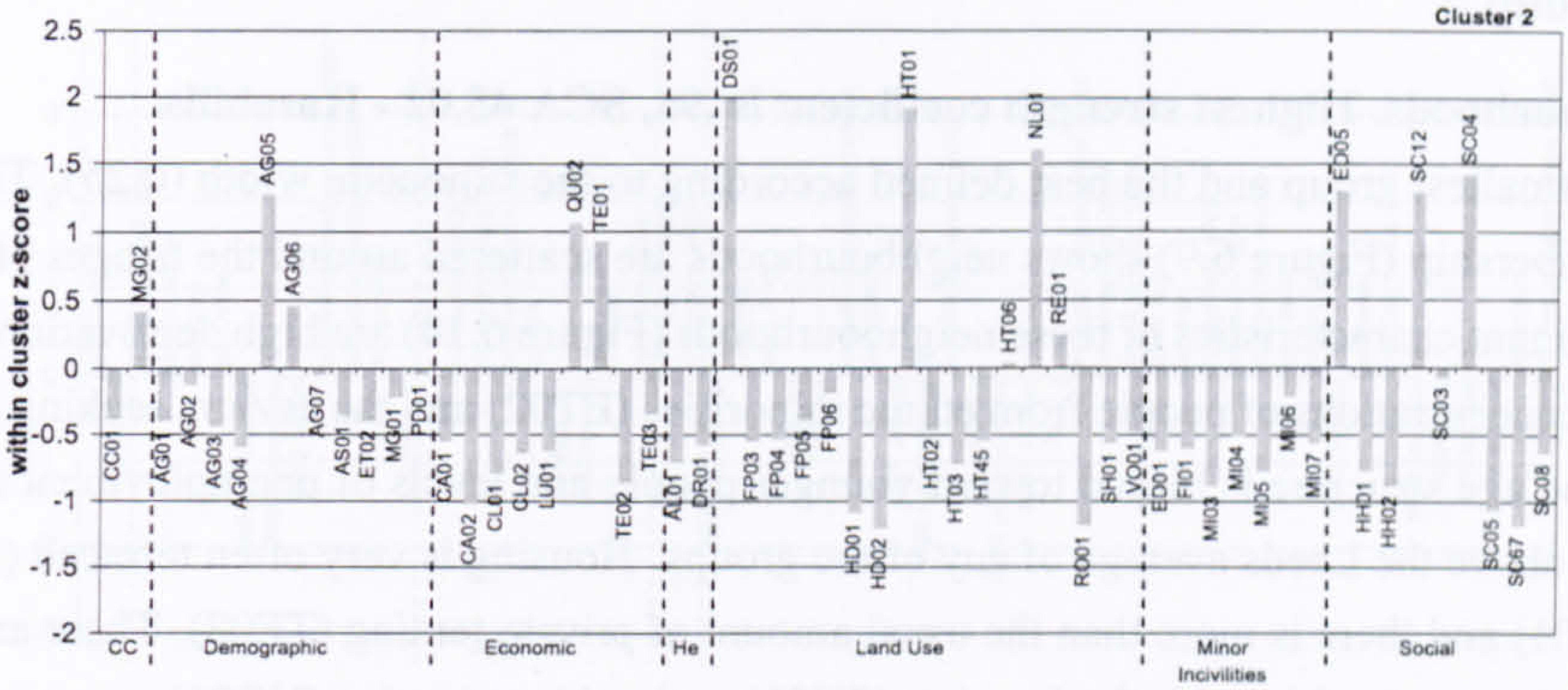


Figure 6.6: Standardised scores for variables defining group 2.

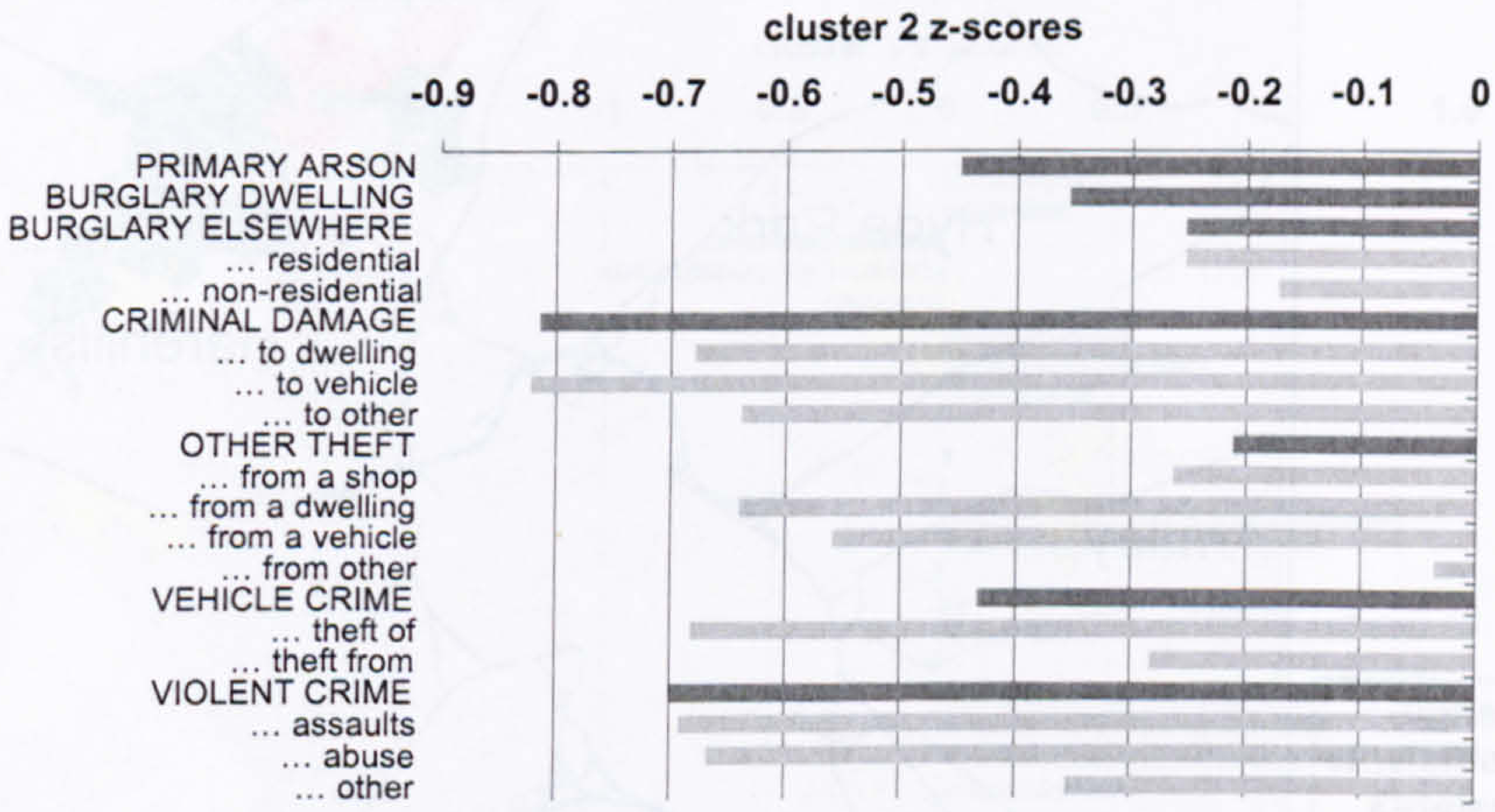


Figure 6.7: Standardised scores for crime variables, for 2003/04, for group 2.

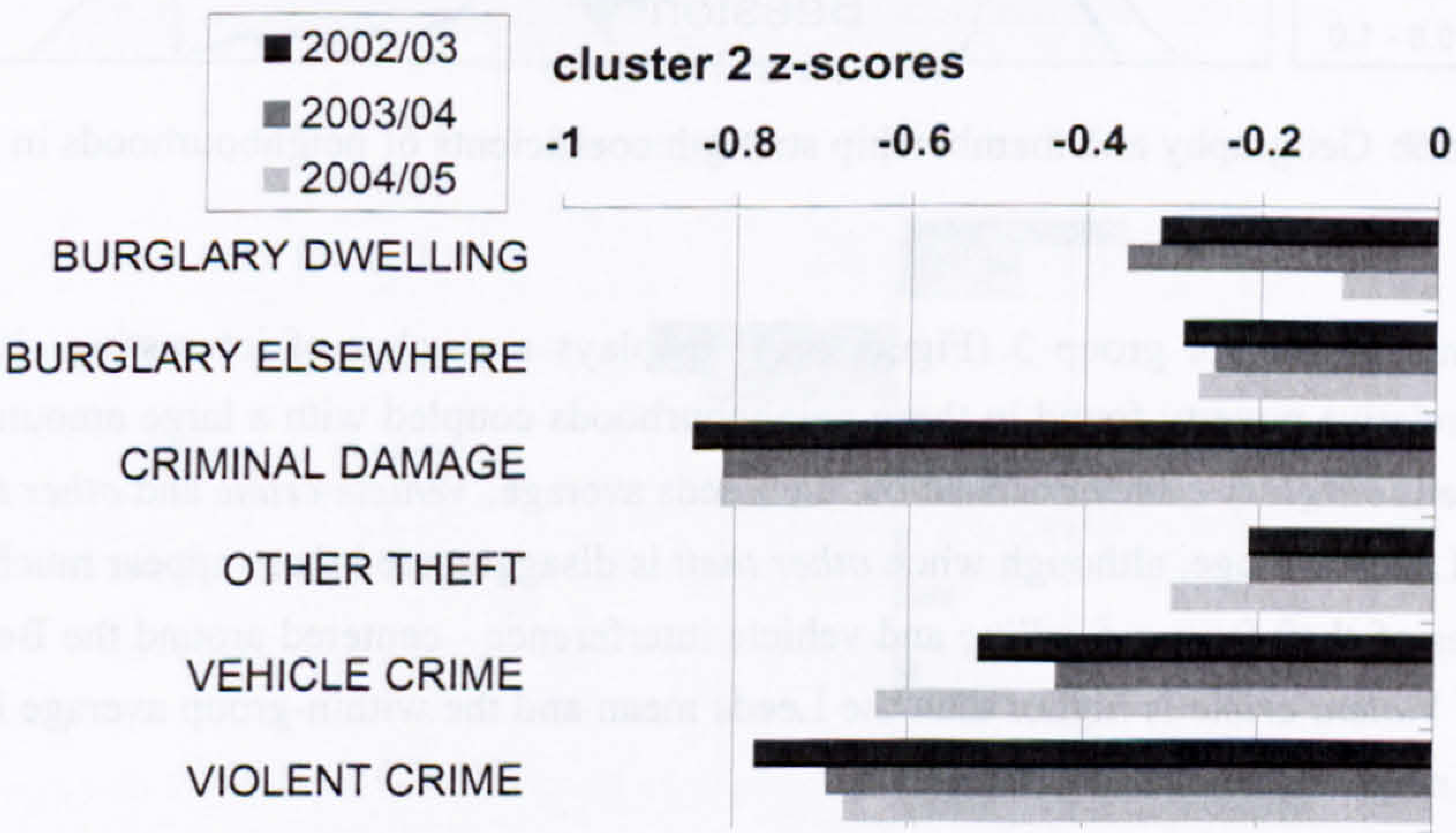


Figure 6.8: Change in group 2 standardised crime rates, 2002/03 to 2004/05.

6.2.4 Group 3

19 Neighbourhoods. Highest strength coefficient is .96, SCA 45.02 - Harehills.

This is the smallest group and the best defined according to the silhouette width (0.25). The geography of membership (Figure 6.9) shows neighbourhoods are scattered around the fringes of the group 1. The dominant characteristics of these neighbourhoods (Figure 6.10) are high deprivation and larger than usual concentrations of people from ethnic minorities (ET02) and the asylum seeking population (AS01). The age structure is biased toward younger people and levels of unemployment (CL01) are the highest above the Leeds average of any of the groups. Housing is very often terraced (HT03) and dense (HD01) and there is more than the usual amount of private renting (TE03). There are a greater number of shops mixed in with the housing (SH01) and void properties (VO01) are more common than usual. Minor incivilities are higher than average but not quite as high as neighbourhoods in group 1. Health problems related to alcohol (AL01) and narcotics (DR01) are more prevalent than usual but are not as severe as group 1.

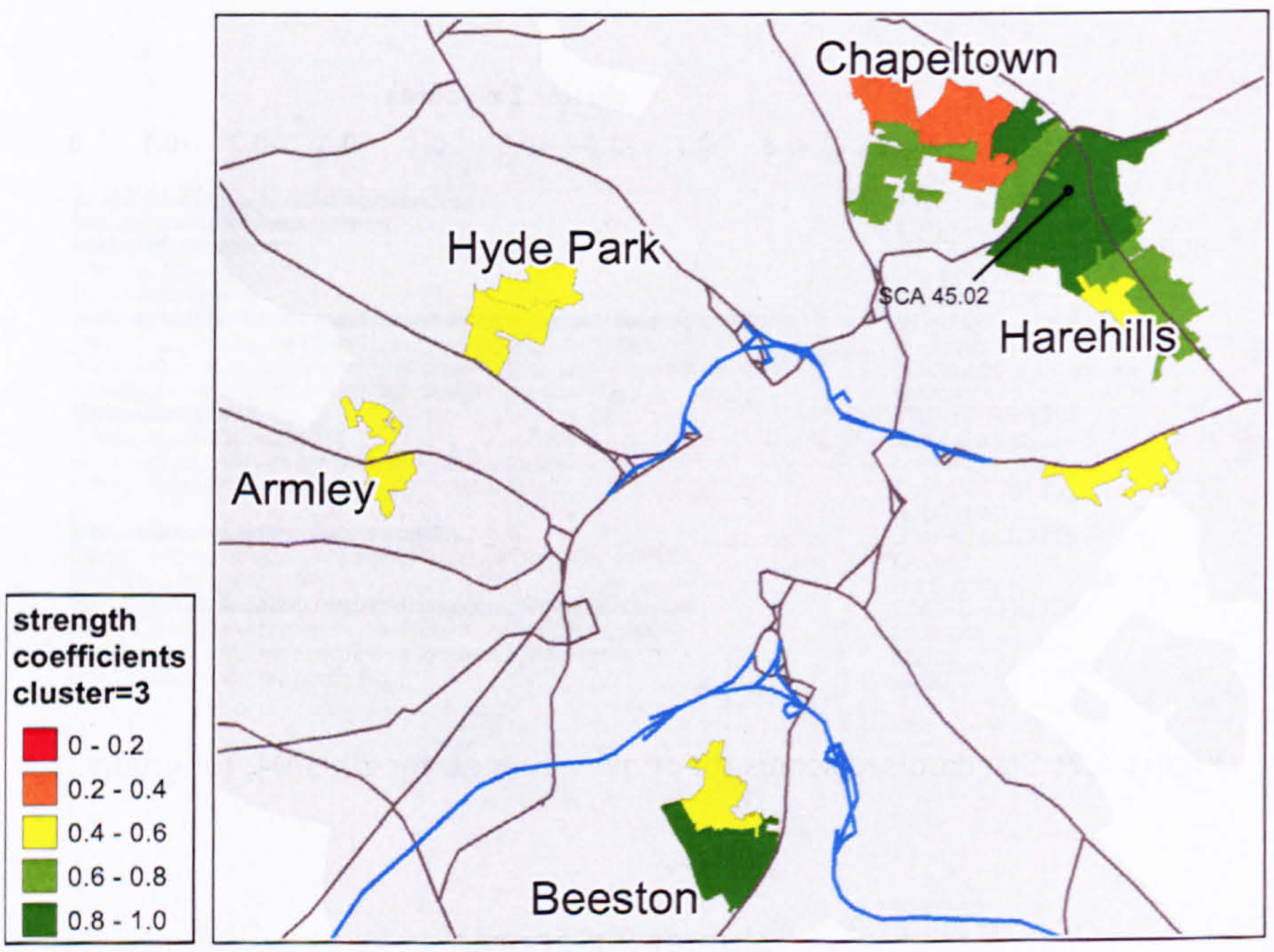


Figure 6.9: Geography and membership strength coefficients of neighbourhoods in group 3.

The crime profile for group 3 (Figure 6.11) displays a number of interesting characteristics. Firstly, the relative poverty found in these neighbourhoods coupled with a large amount of terraced housing means *burglary elsewhere* is below the Leeds average. *Vehicle crime* and *other theft* are both around the Leeds average, although when *other theft* is disaggregated there appear much higher than average rates of *theft from a dwelling* and vehicle interference - centered around the Beeston neighbourhoods. *Violent crime* is higher than the Leeds mean and the within-group average is inflated by a couple of neighbourhoods in Chapeltown.

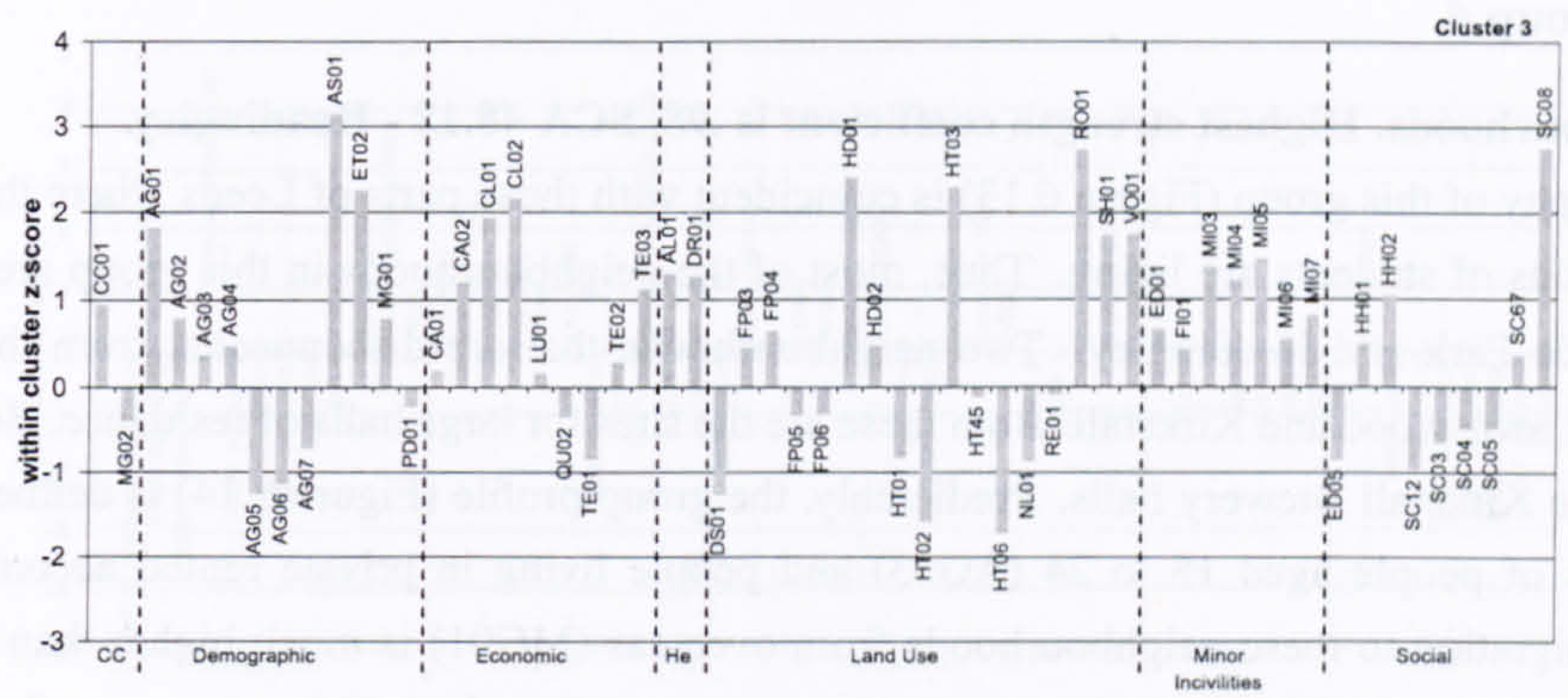


Figure 6.10: Standardised scores for variables defining group 3.

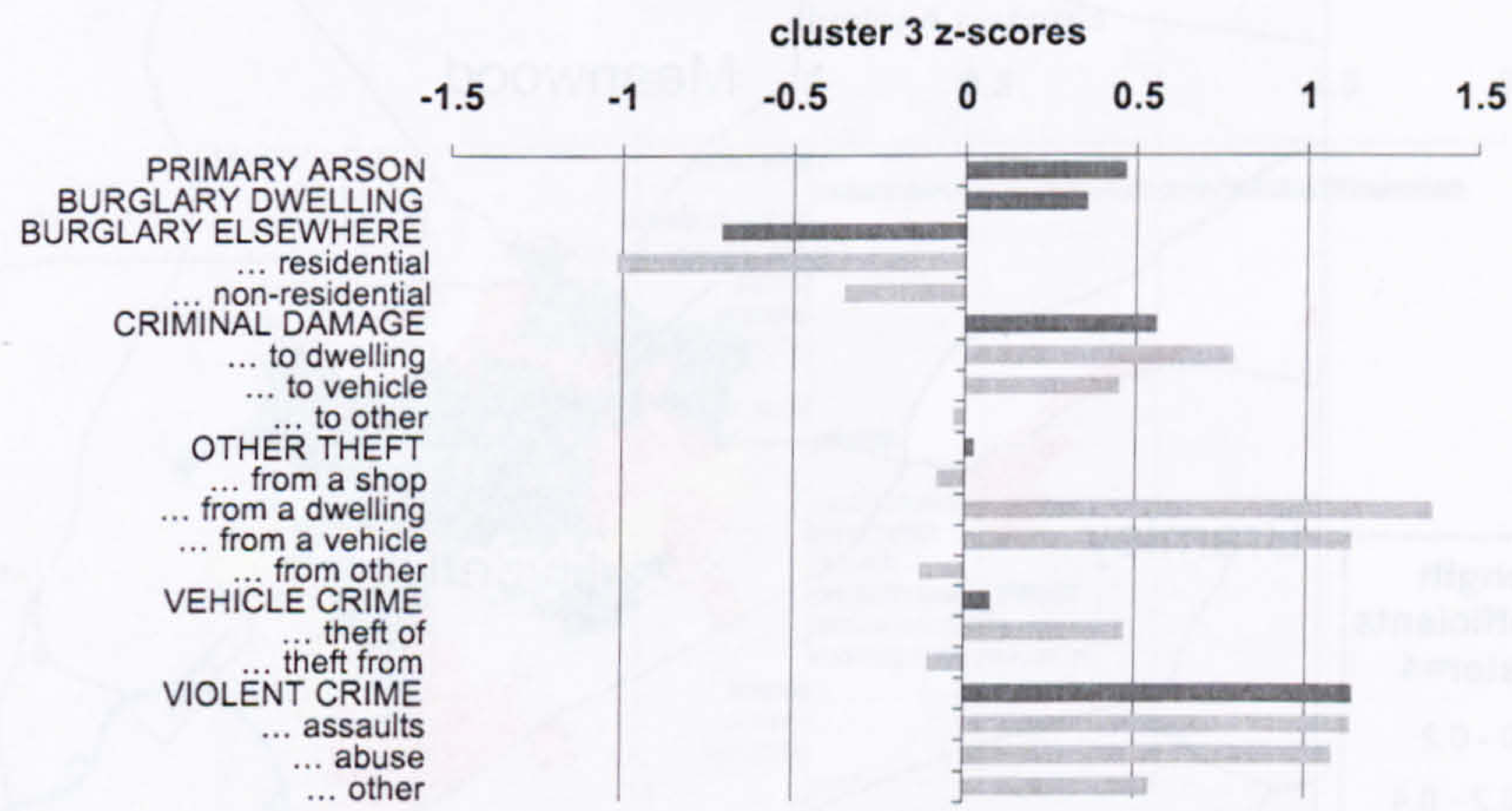


Figure 6.11: Standardised scores for crime variables, for 2003/04, for group 3.

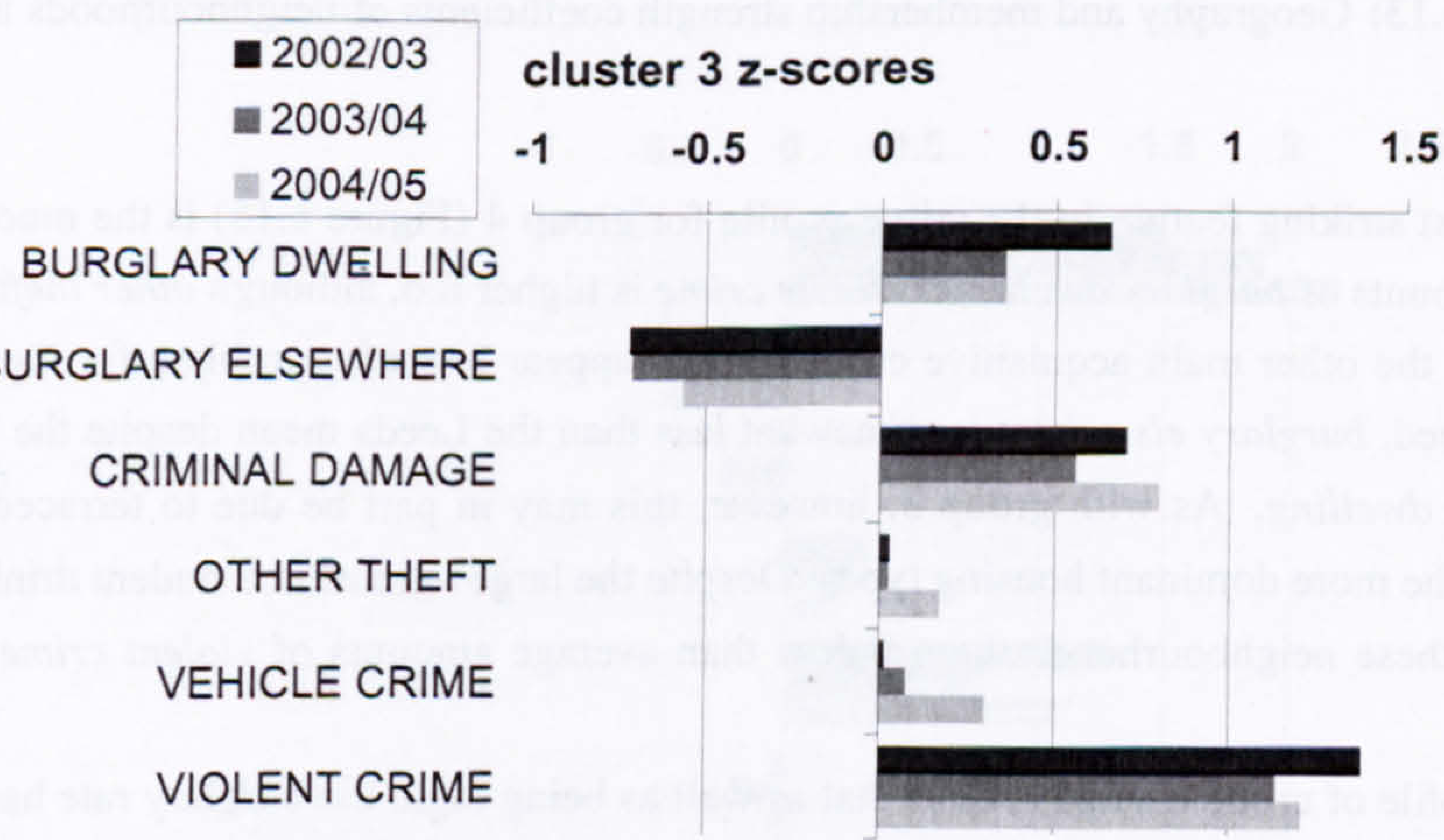


Figure 6.12: Change in group 3 standardised crime rates, 2002/03 to 2004/05.

6.2.5 Group 4

25 Neighbourhoods. Highest strength coefficient is .98, SCA 48.12 - Headingley.

The geography of this group (Figure 6.13) is coincident with those parts of Leeds where the greatest concentrations of students are living. Thus, most of the neighbourhoods in this group are focussed around Hyde Park and Headingley. Two neighbourhoods that are disconnected from the rest are situated in Lawnswood and Kirkstall. Both these are the sites for large halls of residence, Boddington Hall and the Kirkstall Brewery halls. Predictably, the group profile (Figure 6.14) is defined by high proportions of people aged 15 to 24 (AG03) and people living in private rented accommodation (TE03). Migration to these neighbourhoods from overseas (MG01) is much higher than usual and residential stability (MG02) is the lowest of all the groups. Minor incivilities are generally similar to the Leeds average, the only exception being noise complaints (MI03) - which are higher than usual.

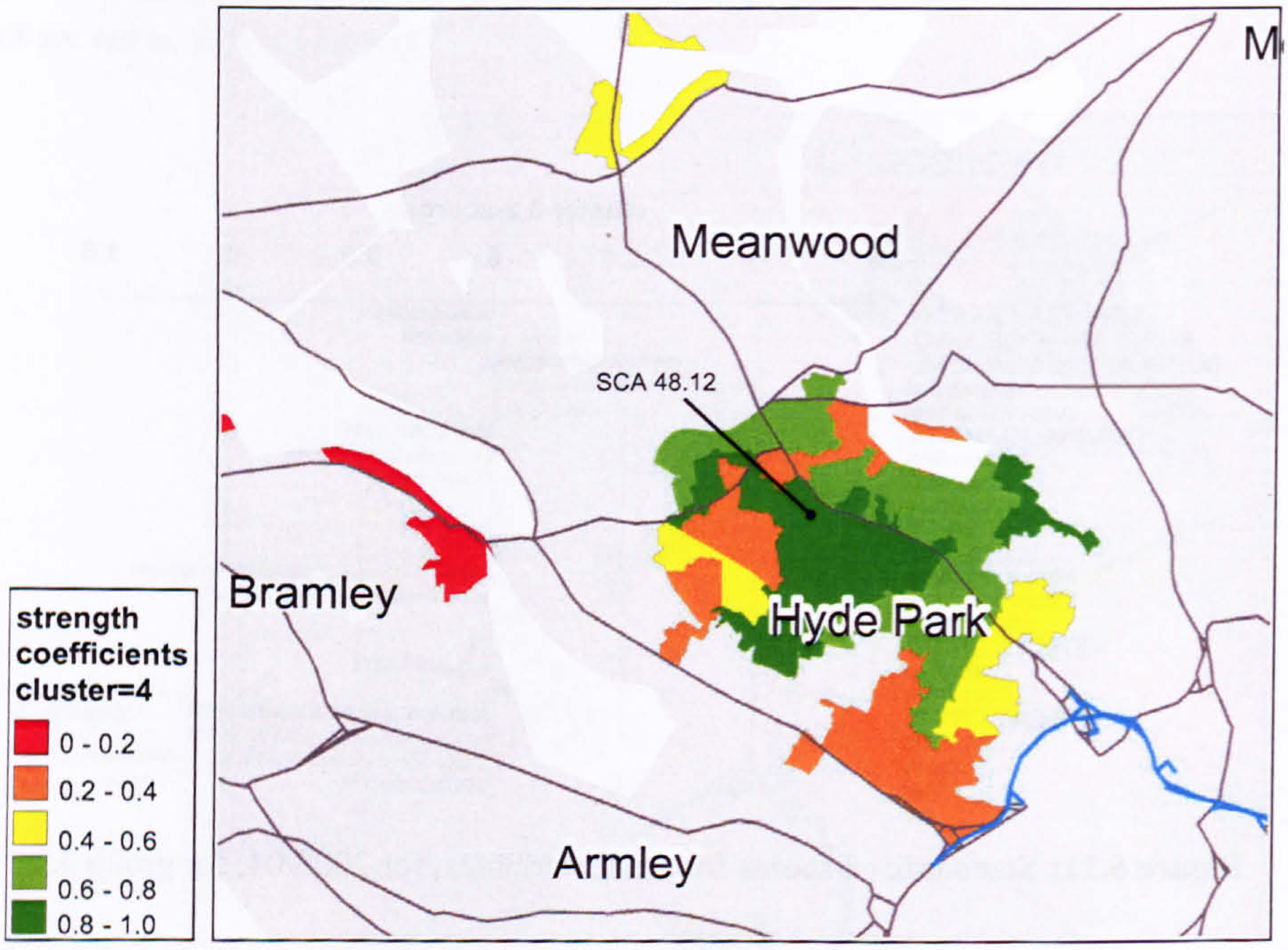


Figure 6.13: Geography and membership strength coefficients of neighbourhoods in group 4.

The most striking feature in the crime profile for group 4 (Figure 6.15) is the much higher than average amounts of *burglary dwelling*. *Vehicle crime* is higher too, although *other theft* and *burglary elsewhere* - the other main acquisitive crime types - appear less of a problem for these neighbourhoods. Indeed, *burglary elsewhere* is somewhat less than the Leeds mean despite the high amounts of *burglary dwelling*. As with group 3, however, this may in part be due to terraced housing and flats being the more dominant housing types. Despite the large amounts of student drinking that goes on around these neighbourhoods there is less than average amounts of *violent crime* and *criminal damage*.

The profile of crime change reveals that as well as being high, the burglary rate has been getting worse (Figure 6.16).

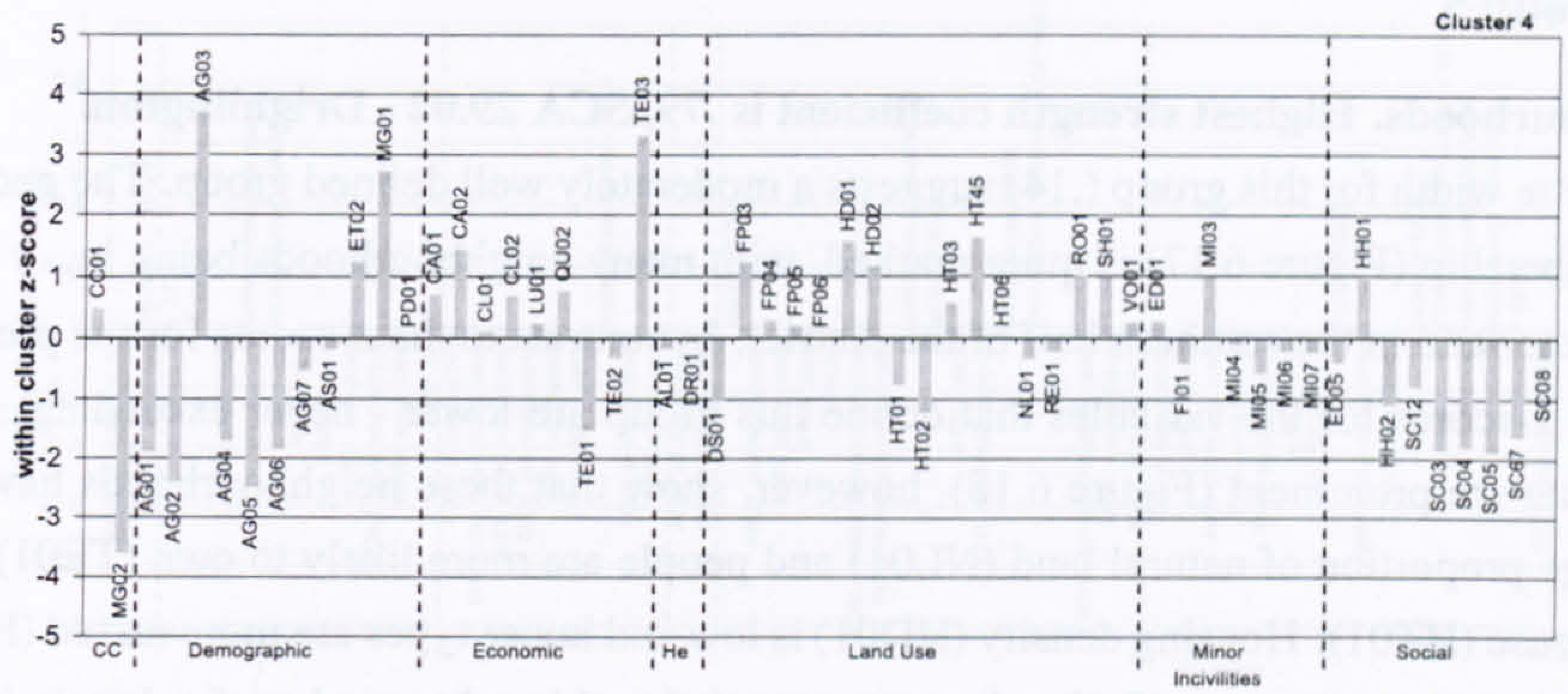


Figure 6.14: Standardised scores for variables defining group 4.

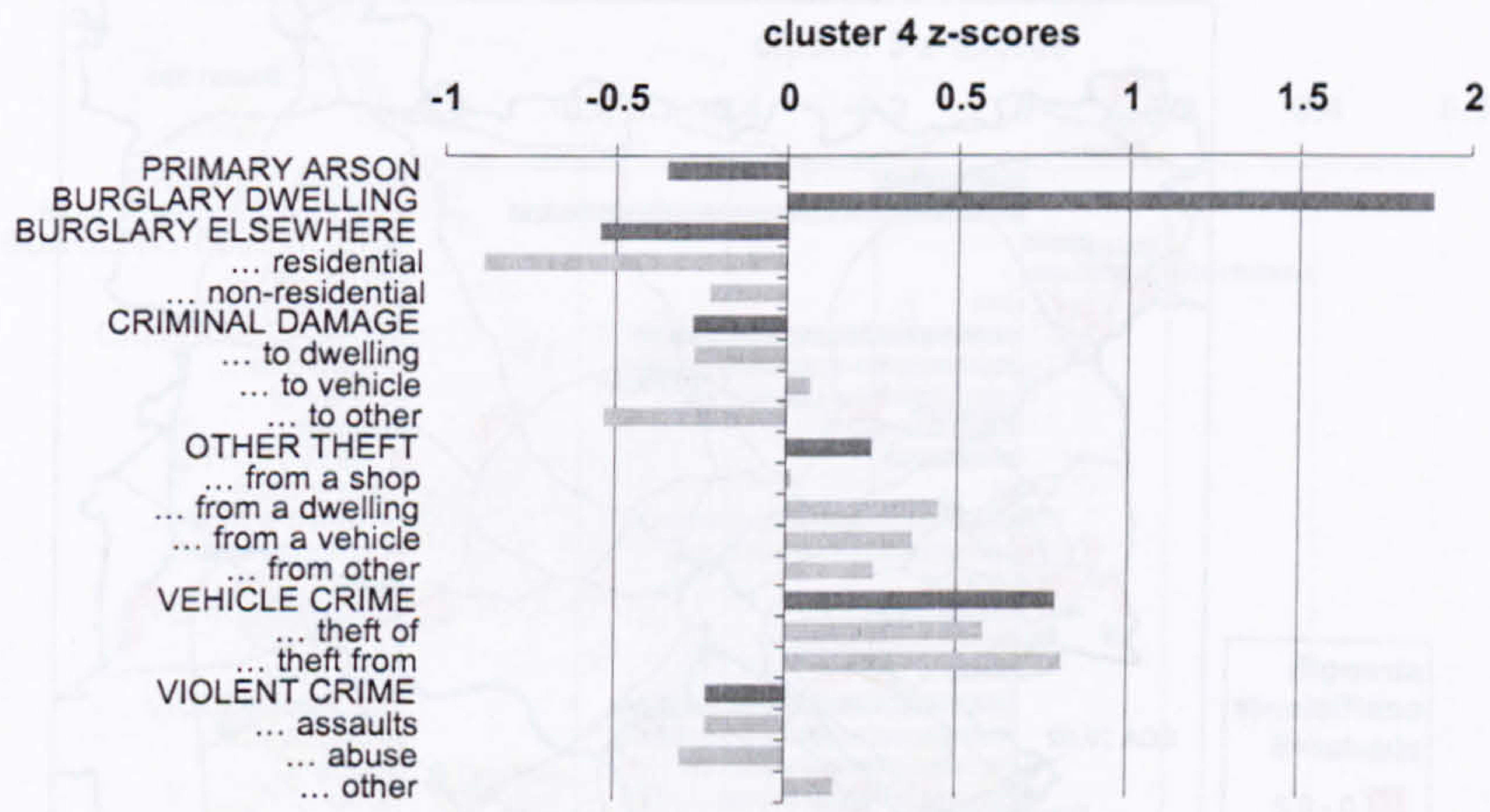


Figure 6.15: Standardised scores for crime variables, for 2003/04, for group 4.

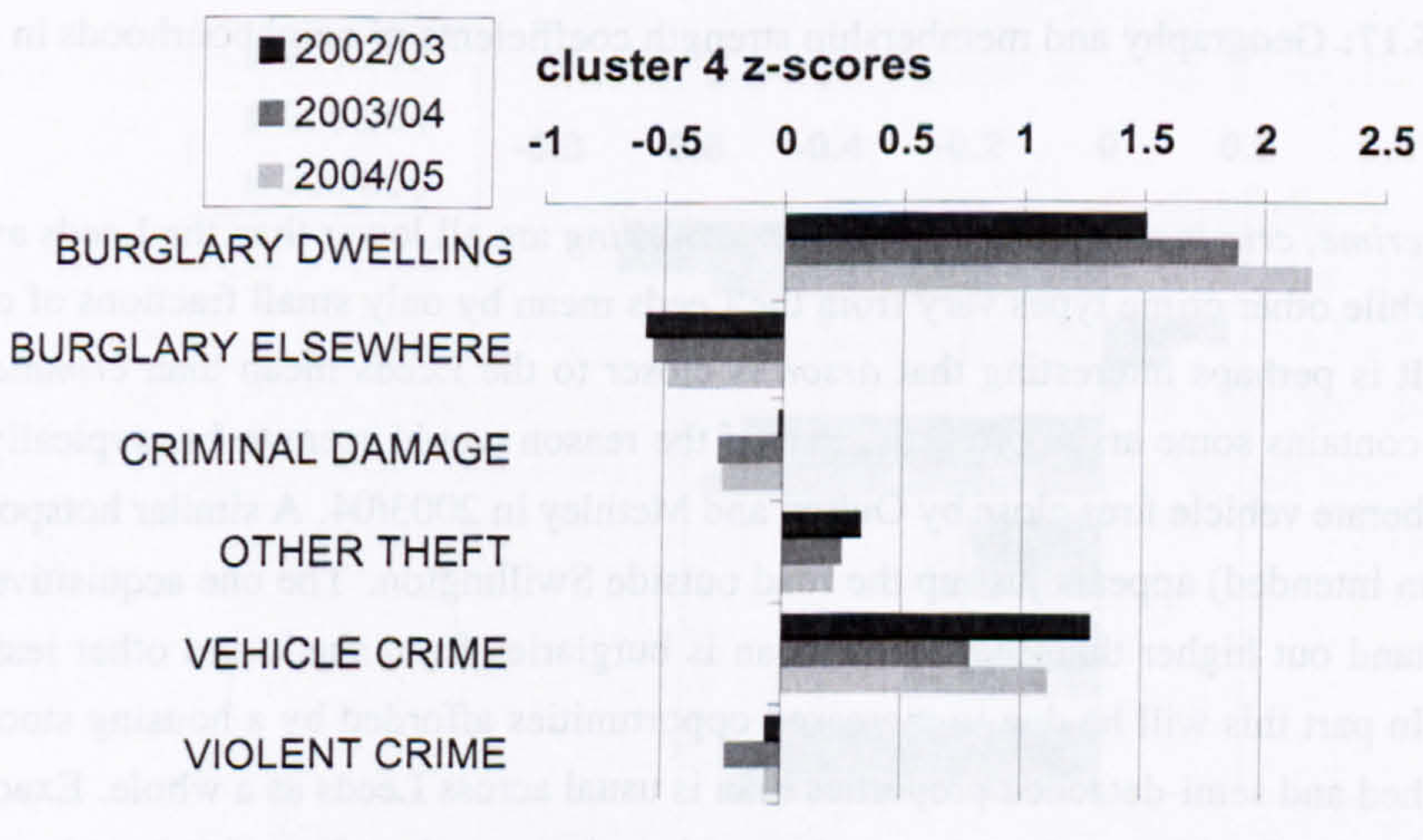


Figure 6.16: Change in group 4 standardised crime rates, 2002/03 to 2004/05.

6.2.7 Group 6

58 Neighbourhoods. Highest strength coefficient is .92, SCA 66.05 - Chapel Allerton.

The silhouette width for this group is below 1, suggesting a weak group. The membership geography (Figure 6.21) shows that these neighbourhoods are found in the suburbs of north and west Leeds. Variables of importance to the group profile (Figure 6.22) include a much more highly qualified population (QU02) than average, the best qualified of all the groups in fact. Related to this is a social class structure that is strongly biased toward people with professional and senior managerial careers (SC12). GCSE results (ED05) are good for pupils living in these neighbourhoods and people are more likely to own their own (TE01) semi-detached (HT02) home. The age structure is slightly older than average and their is a higher than average proportion of people from ethnic minorities (ET02). Unlike other groups however, these ethnic groups are not accompanied by a higher than average asylum seeking population. Minor incivilities are lower than average.

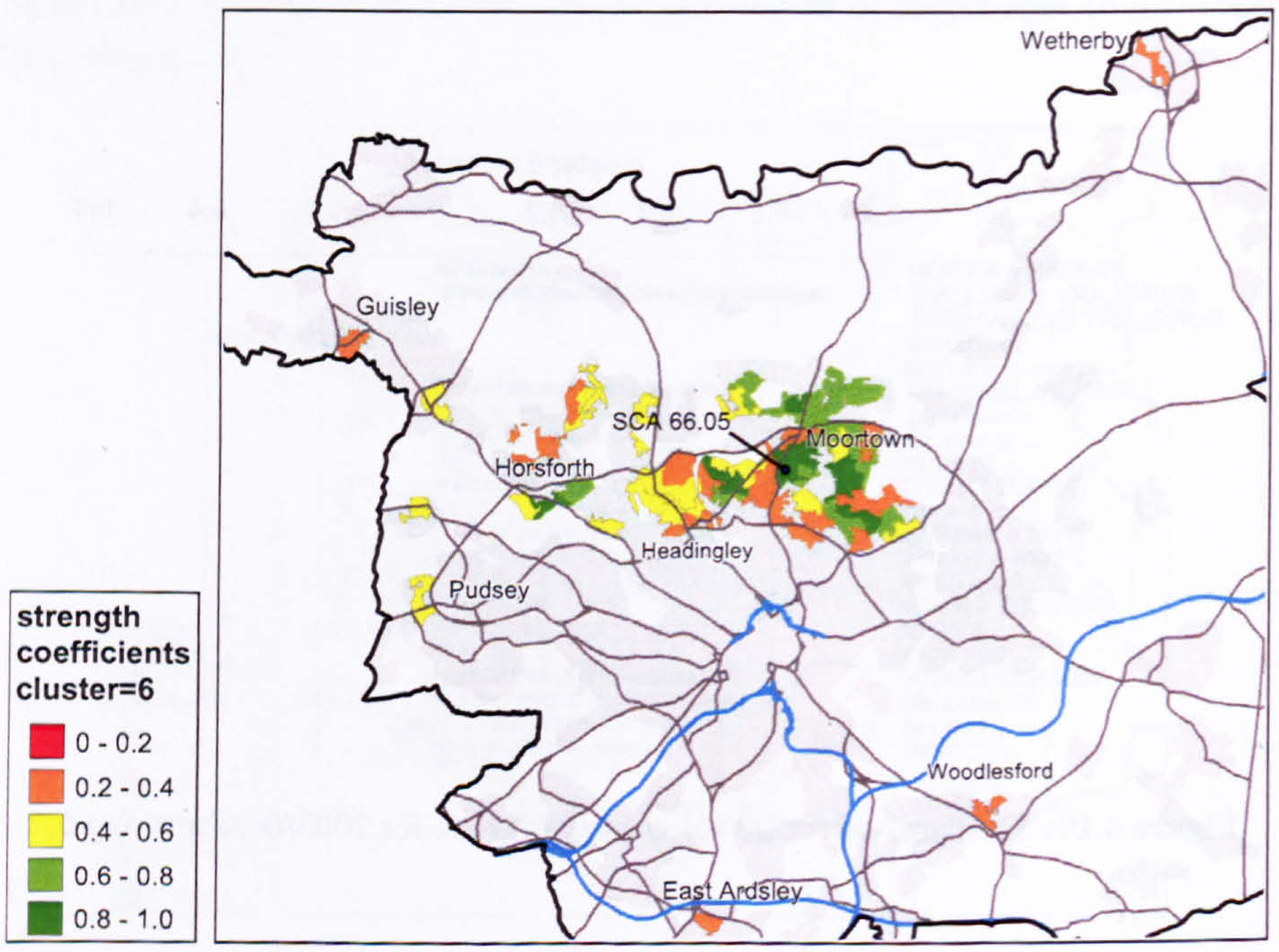


Figure 6.21: Geography and membership strength coefficients of neighbourhoods in group 6.

Of the aggregate crime types (in Figure 6.23) only *violent crime*, *criminal damage* and *arson* are well below the Leeds mean. Of the acquisitive crime types it is *burglary dwelling* and thefts from associated out buildings that are nearest to the district norm. Since 2002/03, burglary dwelling has fallen back to the district mean. At the same time, however, burglary elsewhere has been increasing (Figure 6.24).

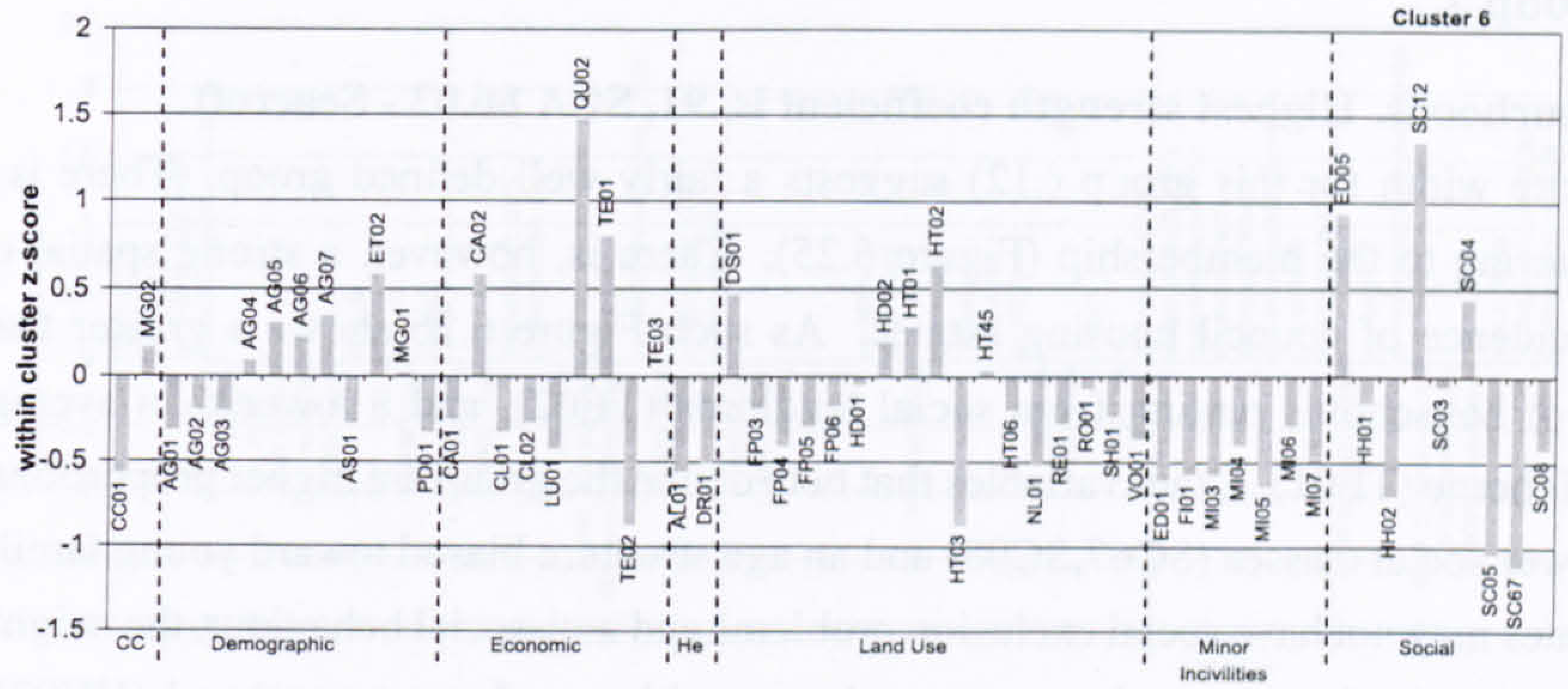


Figure 6.22: Standardised scores for variables defining group 6.

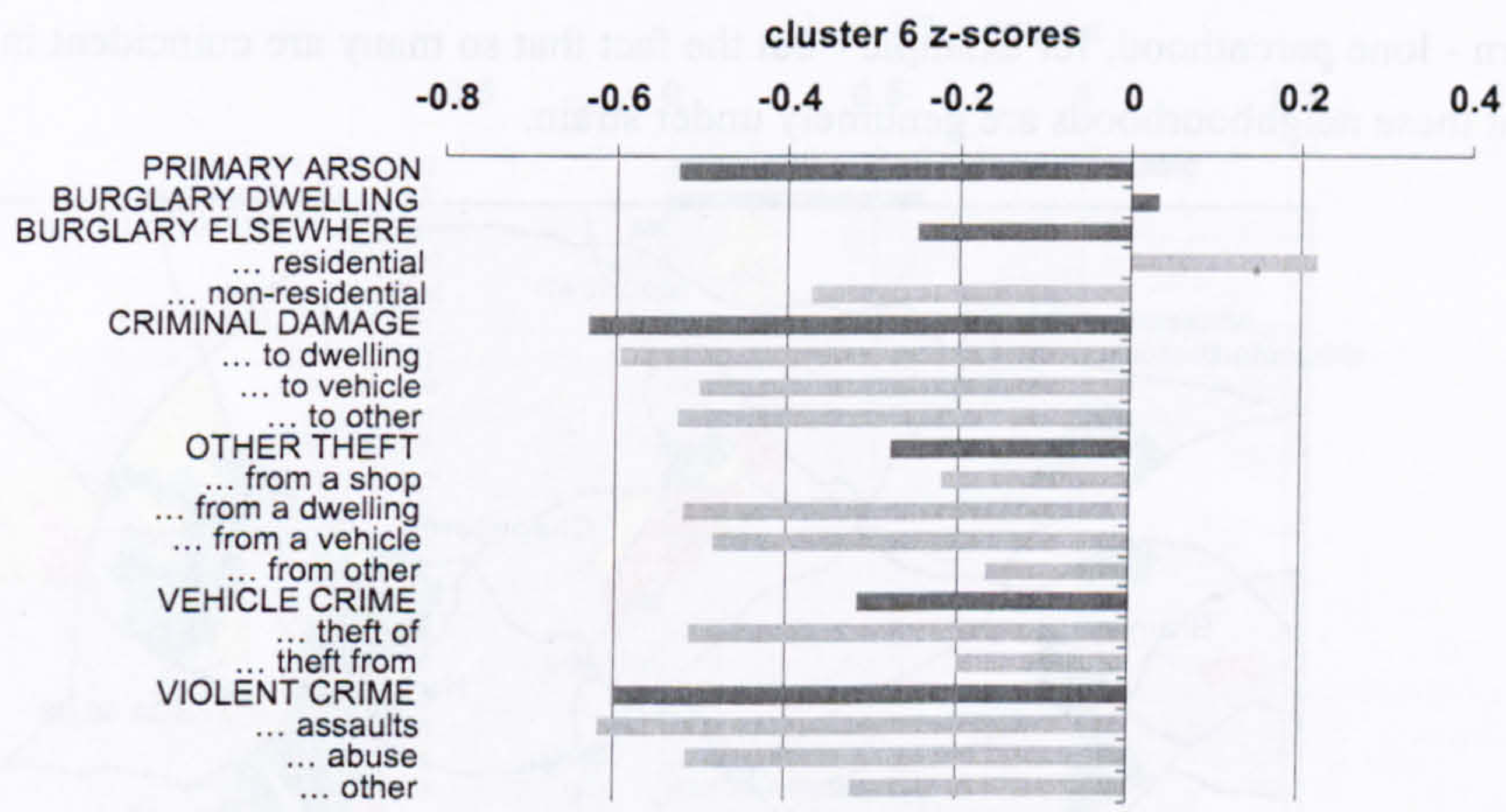


Figure 6.23: Standardised scores for crime variables, for 2003/04, for group 6.

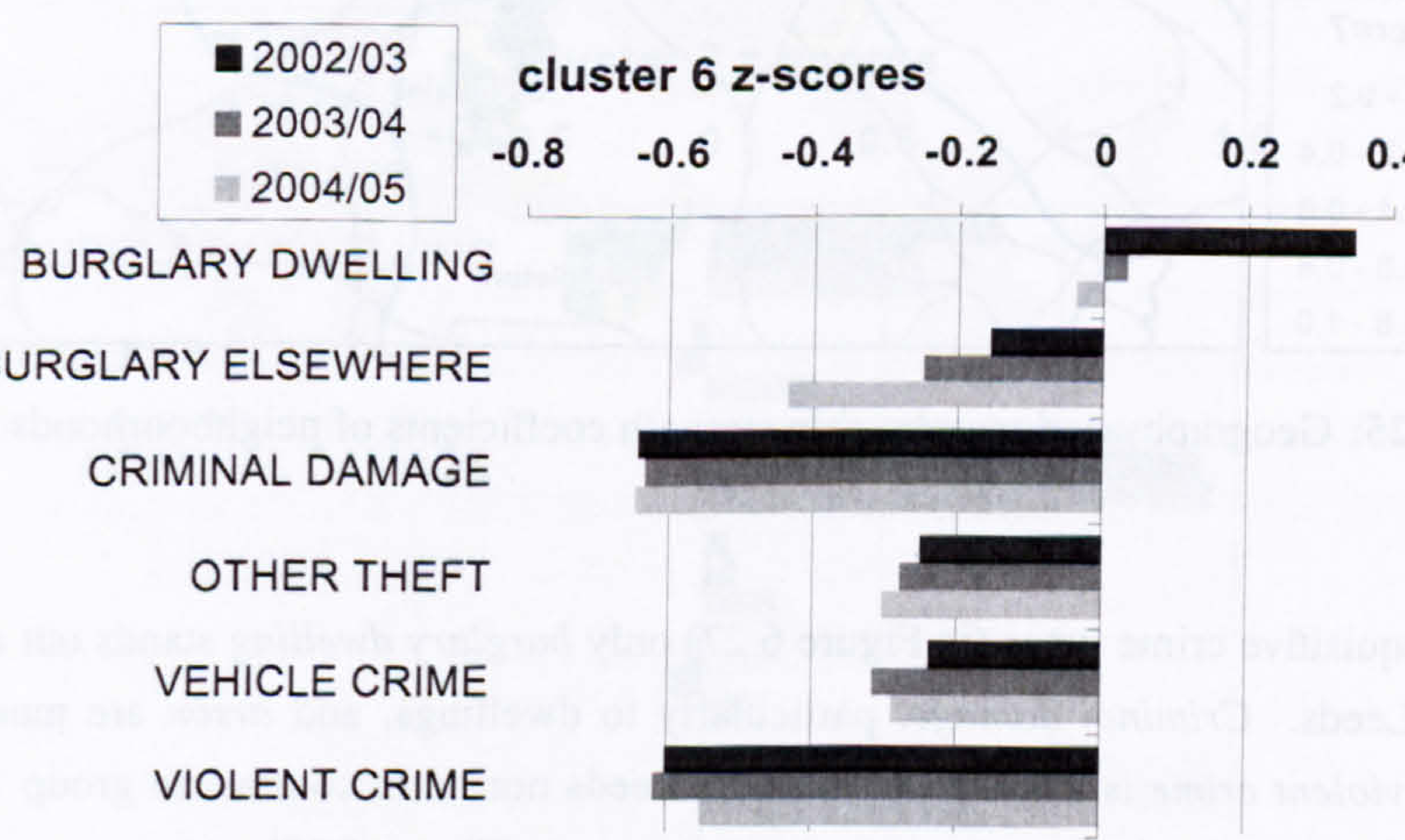


Figure 6.24: Change in group 6 standardised crime rates, 2002/03 to 2004/05.

6.2.8 Group 7

53 Neighbourhoods. Highest strength coefficient is .91, SCA 86.03 - Seacroft.

The silhouette width for this group (.12) suggests a fairly well defined group. There is no strong spatial clustering to the membership (Figure 6.25). There is, however, a strong spatial correlation with the incidence of council housing estates. As such Figure 6.26 shows a greater than average proportion of households renting from social landlords (TE02), and a lower than average number owning their home (TE01). Other variables that help define the group are higher proportions of people from the lower social classes (SC67,SC08) and an age structure biased toward young families. While council estates may not have social exclusion problems and anti-social behaviour, the neighbourhoods grouped together in this group do appear to have problems. Lone-parenthood (HH02) is higher than average, as is unemployment (CL01) and all classes of minor incivility. Also adding to the problems are poor GCSE performance (ED05) and exclusions from school (ED01), and the working population has lower levels of higher qualifications (QU02) than average. Void properties (VO01) are more common than usual and the proportion of people going to hospital for alcohol (AL01) or narcotics related incidents (DR01) is higher than usual. Singly, some of these issues might not be of great concern - lone parenthood, for example - but the fact that so many are coincident in this group suggests that these neighbourhoods are genuinely under strain.

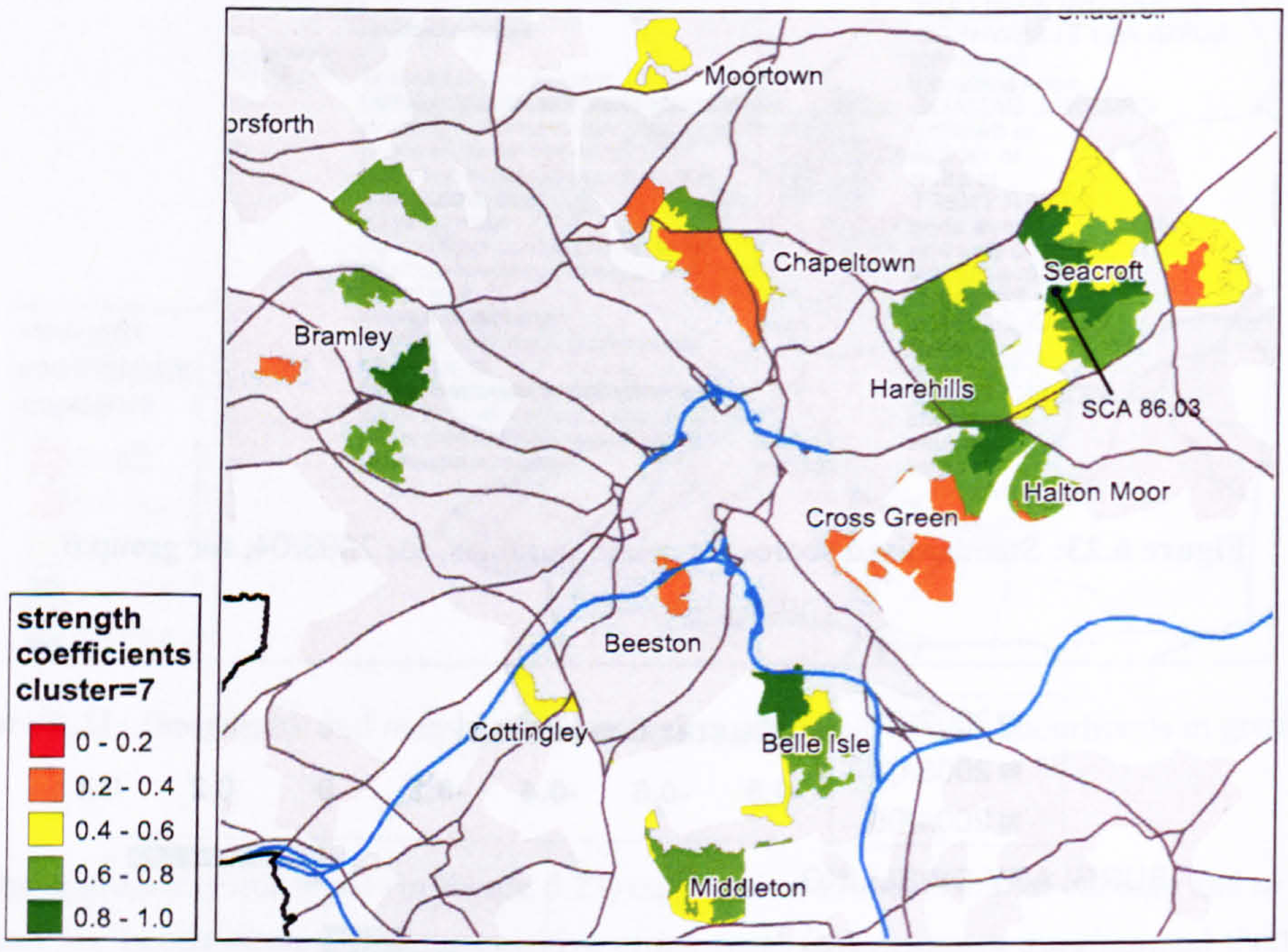


Figure 6.25: Geography and membership strength coefficients of neighbourhoods in group 7.

Of the acquisitive crime types (in Figure 6.27) only *burglary dwelling* stands out as being above average for Leeds. *Criminal damage*, particularly to dwellings, and *arson* are much higher than average and *violent crime* is also higher than the Leeds norm. In contrast to group 1, the criminal damage problem here has become worse over recent years (Figure 6.28).

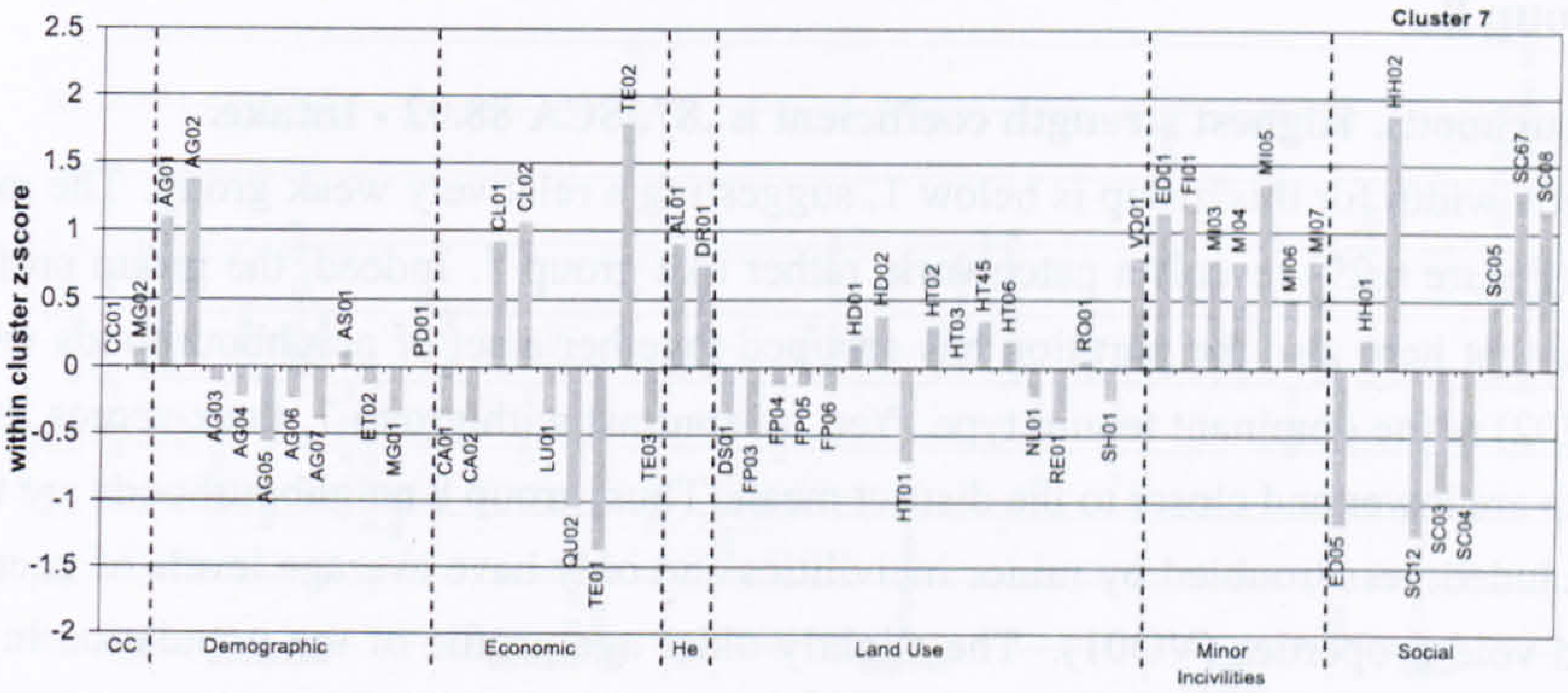


Figure 6.26: Standardised scores for variables defining group 7.

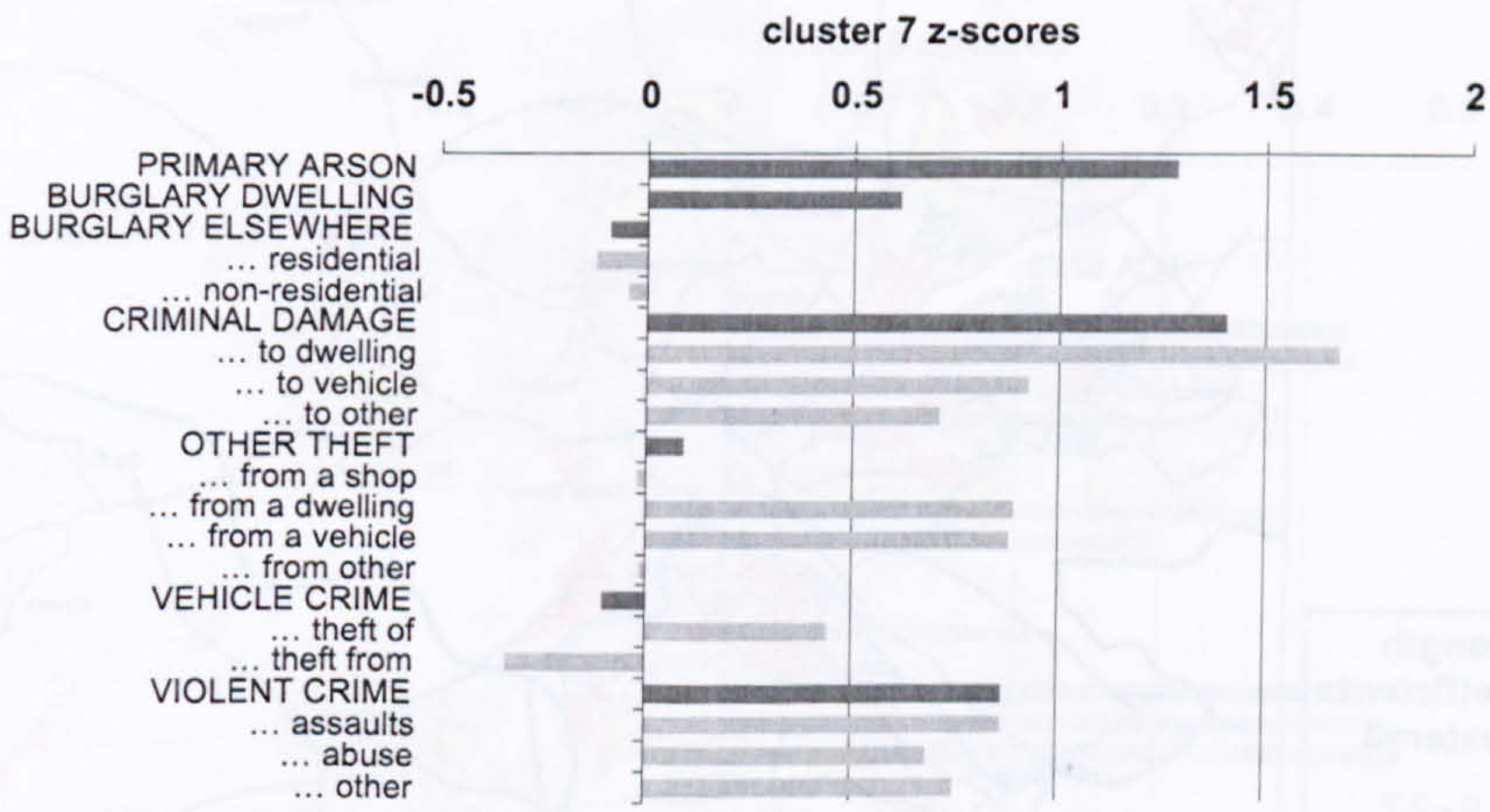


Figure 6.27: Standardised scores for crime variables, for 2003/04, for group 7.

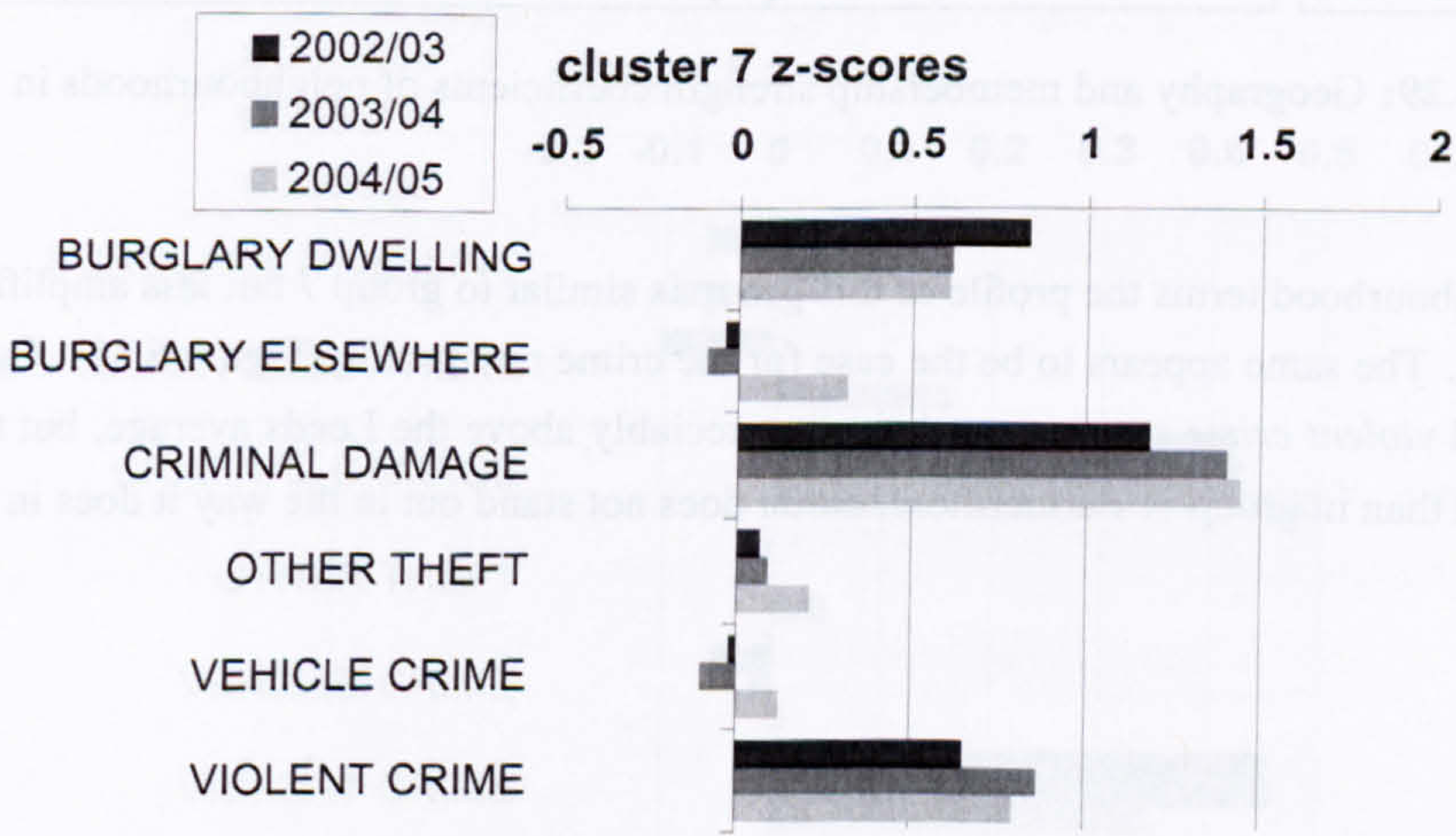


Figure 6.28: Change in group 7 standardised crime rates, 2002/03 to 2004/05.

6.2.9 Group 8

55 Neighbourhoods. Highest strength coefficient is .87, SCA 88.02 - Intake.

The silhouette width for this group is below 1, suggesting a relatively weak group. The membership geography (Figure 6.29) reveals a patchwork, rather like group 7. Indeed, the group profile (Figure 6.30) shows that here also the partition has grouped together a set of neighbourhoods where social renting (TE02) is the dominant tenure type. Yet, by contrast with group 7, the z-scores for variables in this group are lower and closer to the district mean. Thus, group 8 neighbourhoods are that bit less socially excluded, less troubled by minor incivilities and only have average levels of unemployment (CL01) and void properties (VO01). The slightly older age profile of the population in this group might explain some of the difference. If data were available (which it is not) it would be interesting to examine the relation between membership of this group and group 7 with the geography of right-to-buy take-up in Leeds.

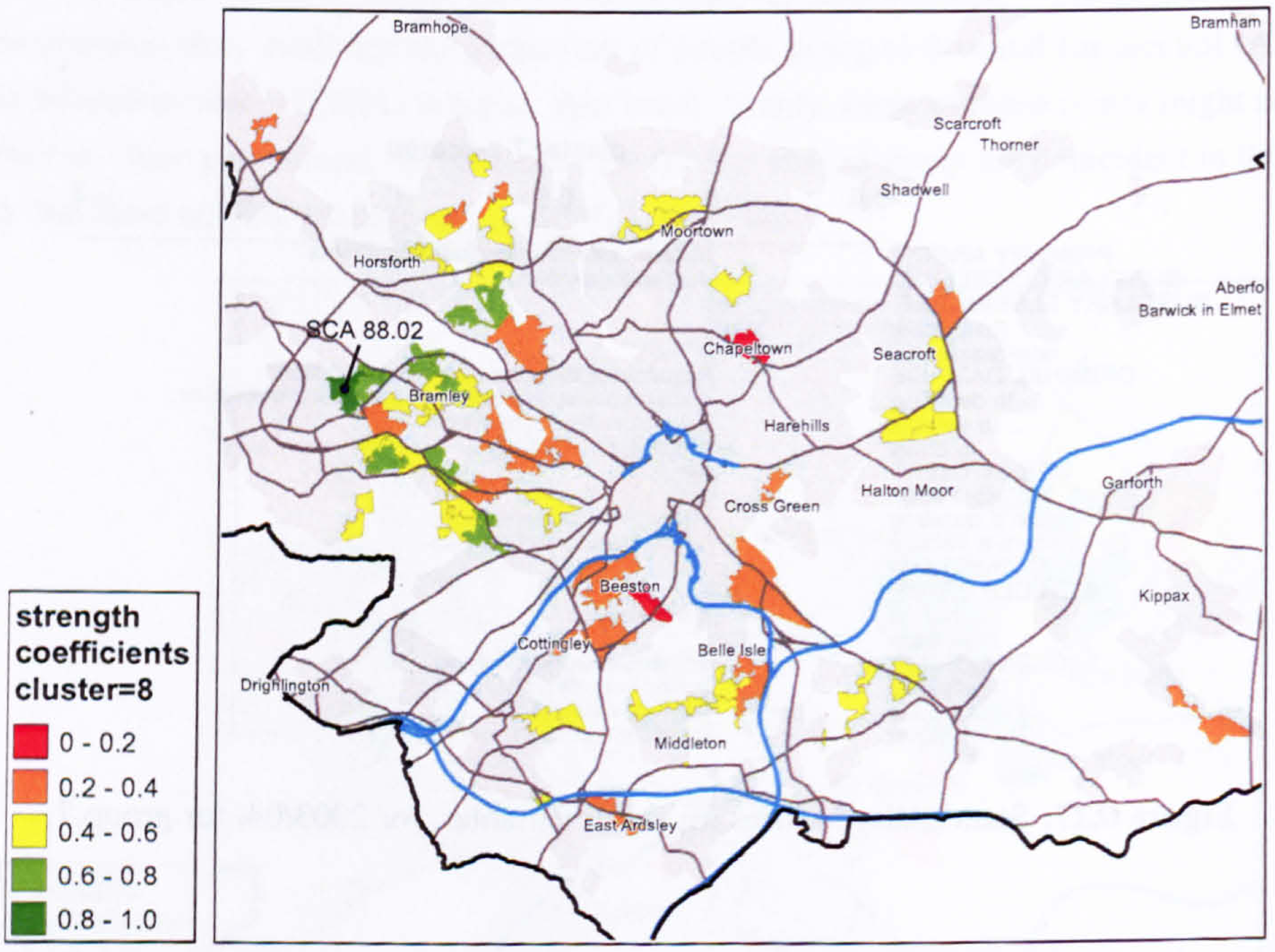


Figure 6.29: Geography and membership strength coefficients of neighbourhoods in group 8.

In neighbourhood terms the profile of this group is similar to group 7 but less amplified from the Leeds norm. The same appears to be the case for the crime rate profile (Figure 6.31). Only *criminal damage* and *violent crime* stand out as being appreciably above the Leeds average, but the extent is less marked than in group 7. Furthermore, *arson* does not stand out in the way it does in group 7.

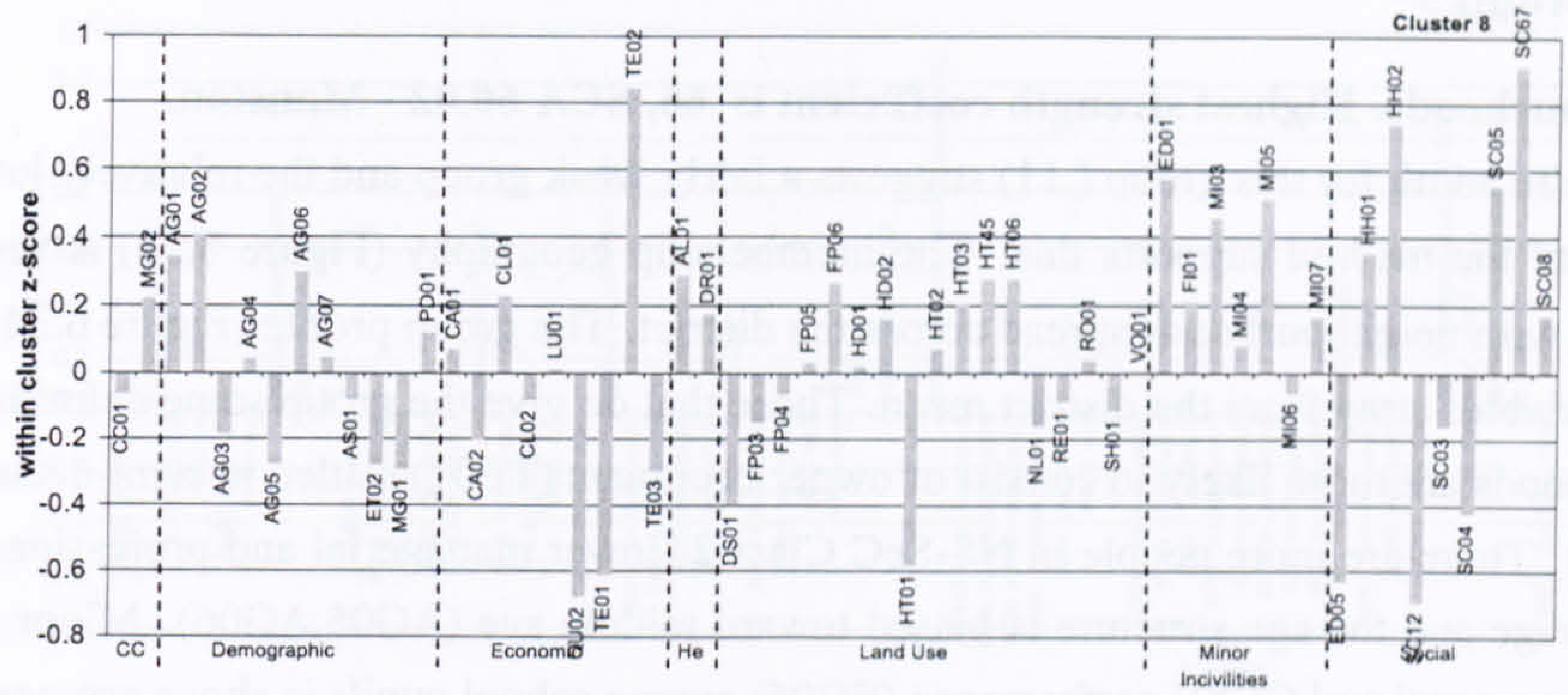


Figure 6.30: Standardised scores for variables defining group 8.

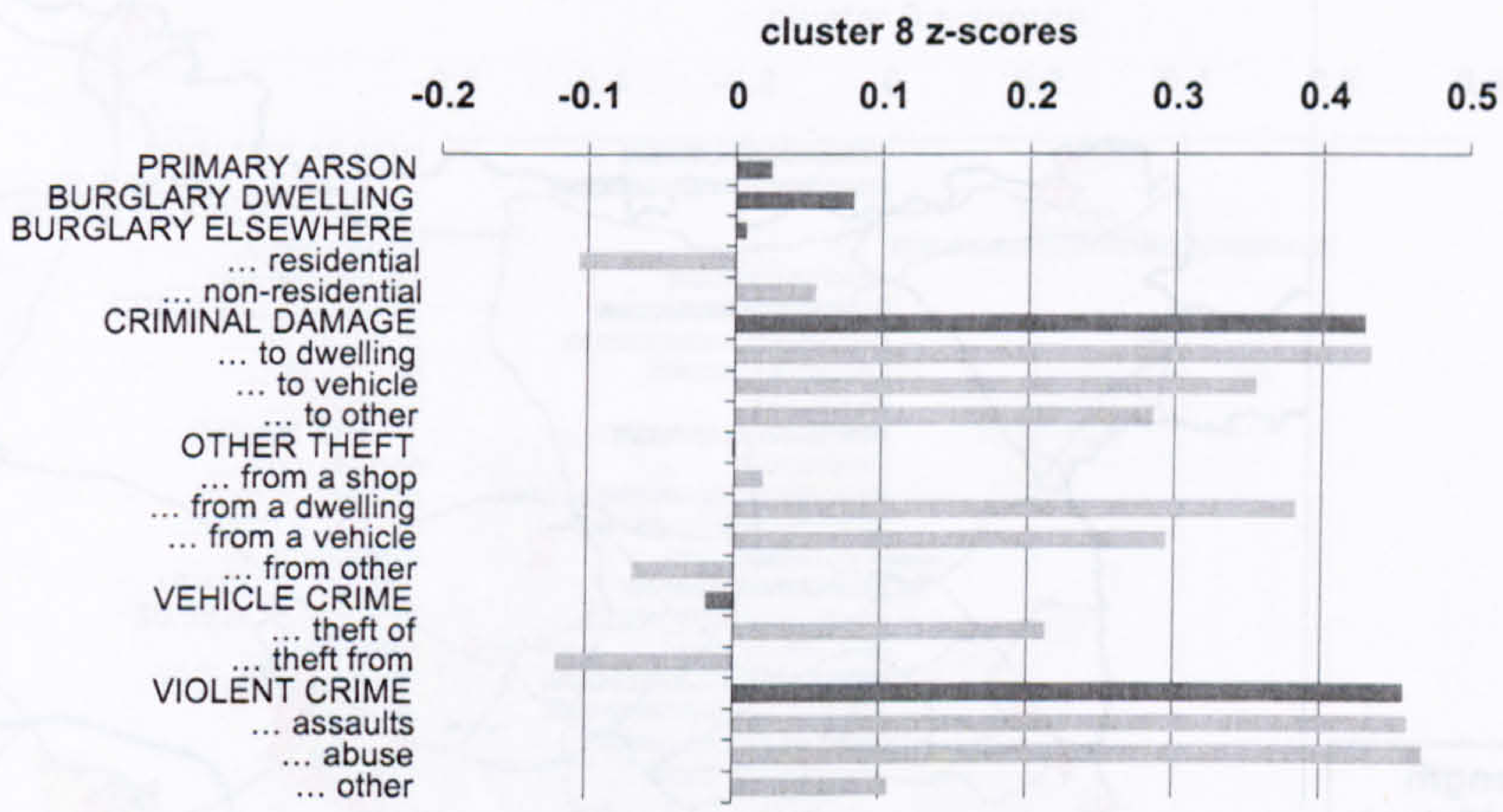


Figure 6.31: Standardised scores for crime variables, for 2003/04, for group 8.

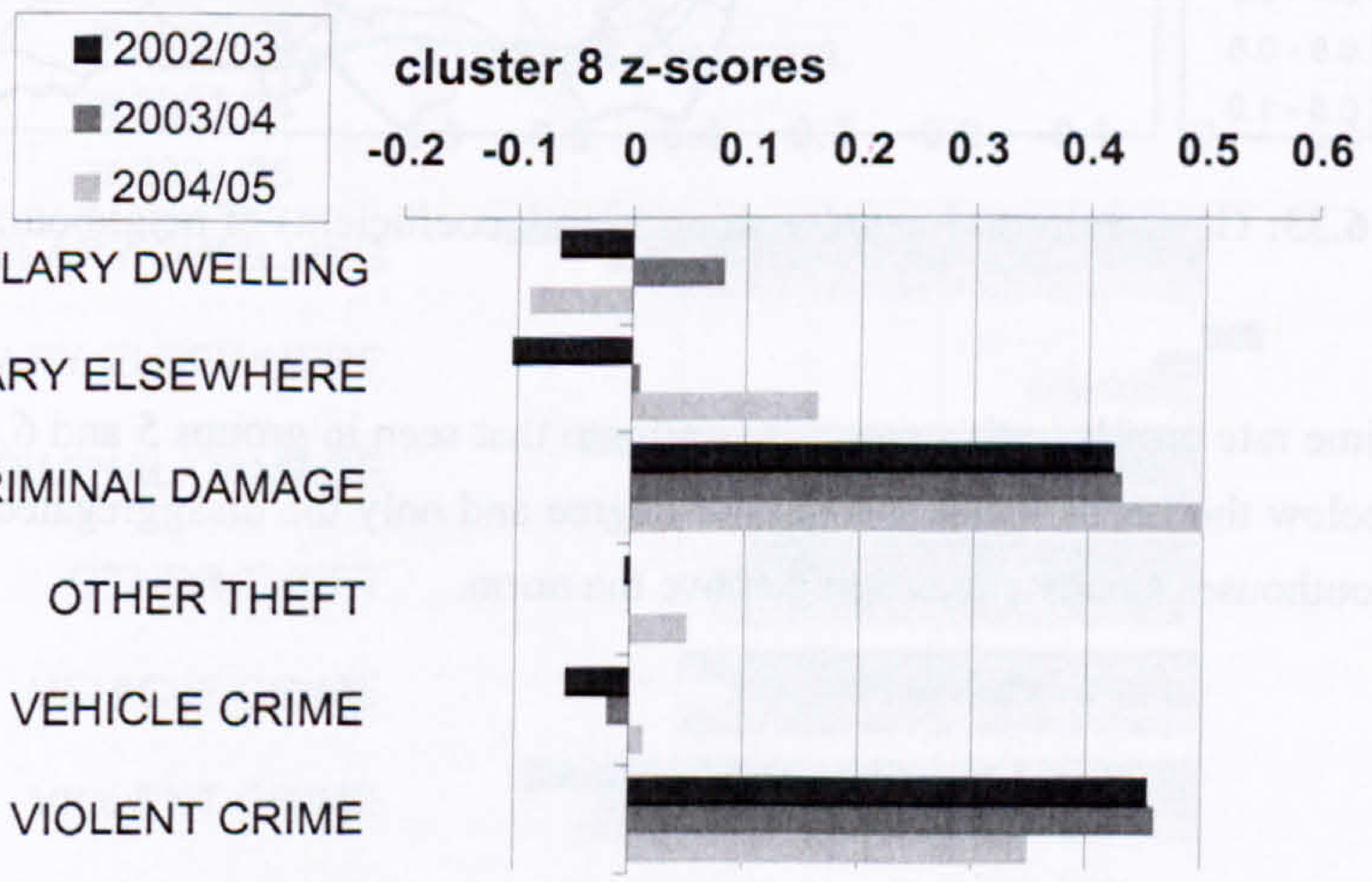


Figure 6.32: Change in group 8 standardised crime rates, 2002/03 to 2004/05.

6.2.10 Group 9

66 Neighbourhoods. Highest strength coefficient is .66, SCA 60.02 - Manston.

The silhouette width for this group (.11) suggests a fairly weak group and the relatively low strength coefficient of the medoid supports this. The membership geography (Figure 6.33) is very much a patchwork, with neighbourhoods spread across the district. The group profile (Figure 6.34) indicates that few variables stray from the district mean. Those that do give the group some definition suggest neighbourhoods are more likely to consist of owner occupiers (TE01) settled in semi-detached housing (HT02). There are more people in NS-SeC Class 2 (lower managerial and professional) (SC03) than is average and the age structure is biased toward middle age (AG05,AG06). Minor incivilities are lower than usual and GCSE performance (ED05) among school pupils is above average, although higher qualifications (QU02) are less common than usual. As has been mentioned before, there is a sense that this group complements group 6. Both have above average levels of owner occupation and a fairly affluent class structure. Group 6, however, comes across as being that bit more affluent and better educated than group 9, as well as being more ethnically diverse.

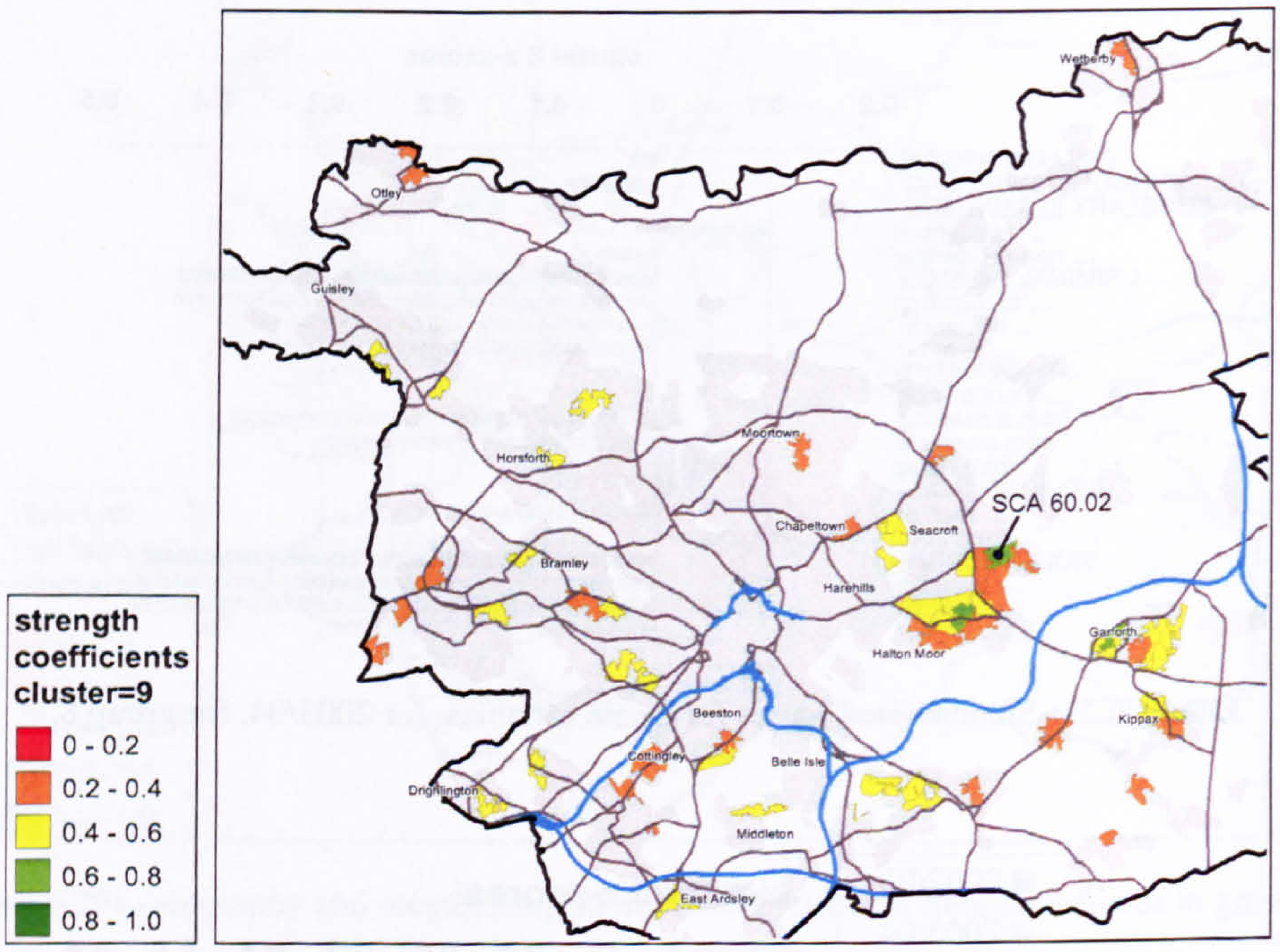


Figure 6.33: Geography and membership strength coefficients of neighbourhoods in group 9.

The crime rate profile in this group is similar to that seen in groups 5 and 6. All aggregate crime types are below the Leeds mean to a similar degree and only the disaggregated rates of burglary of sheds and outhouses stands out as being above the norm.

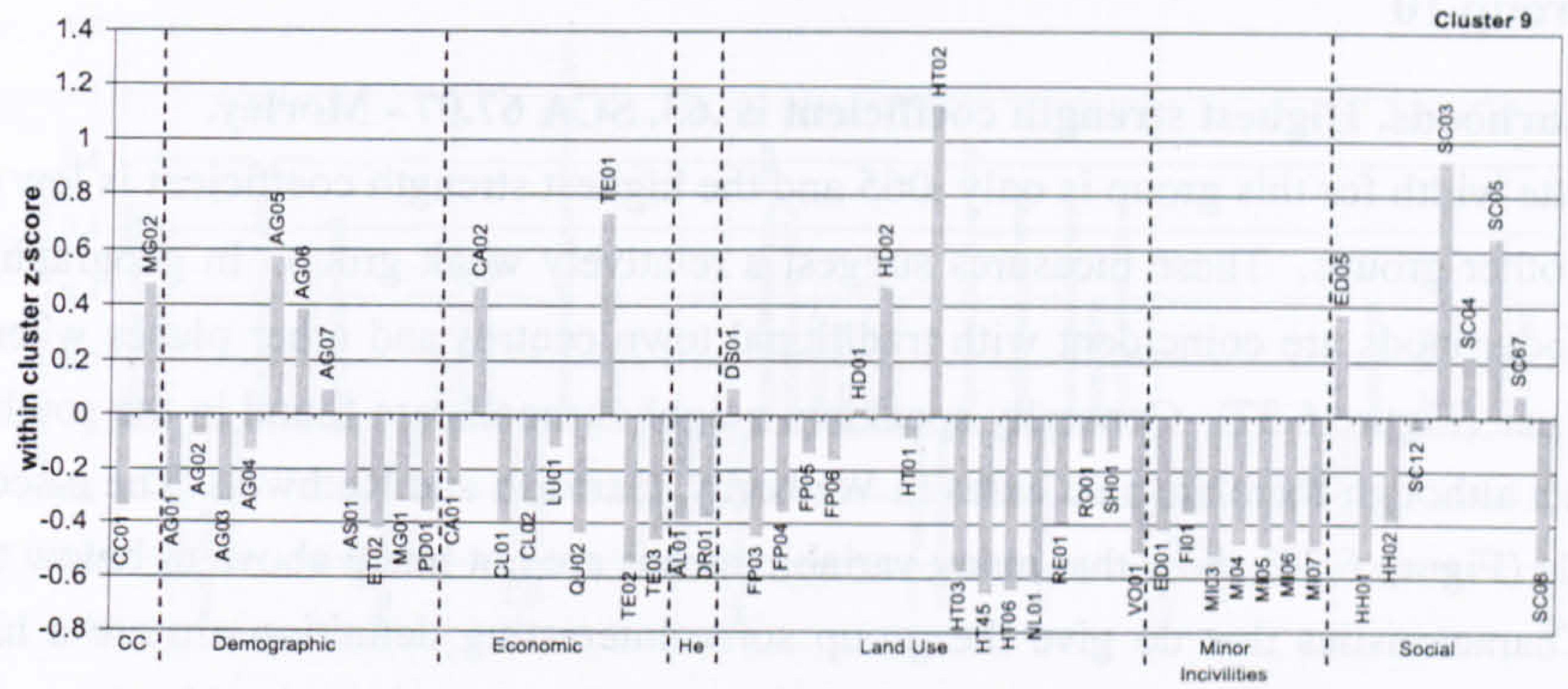


Figure 6.34: Standardised scores for variables defining group 9.

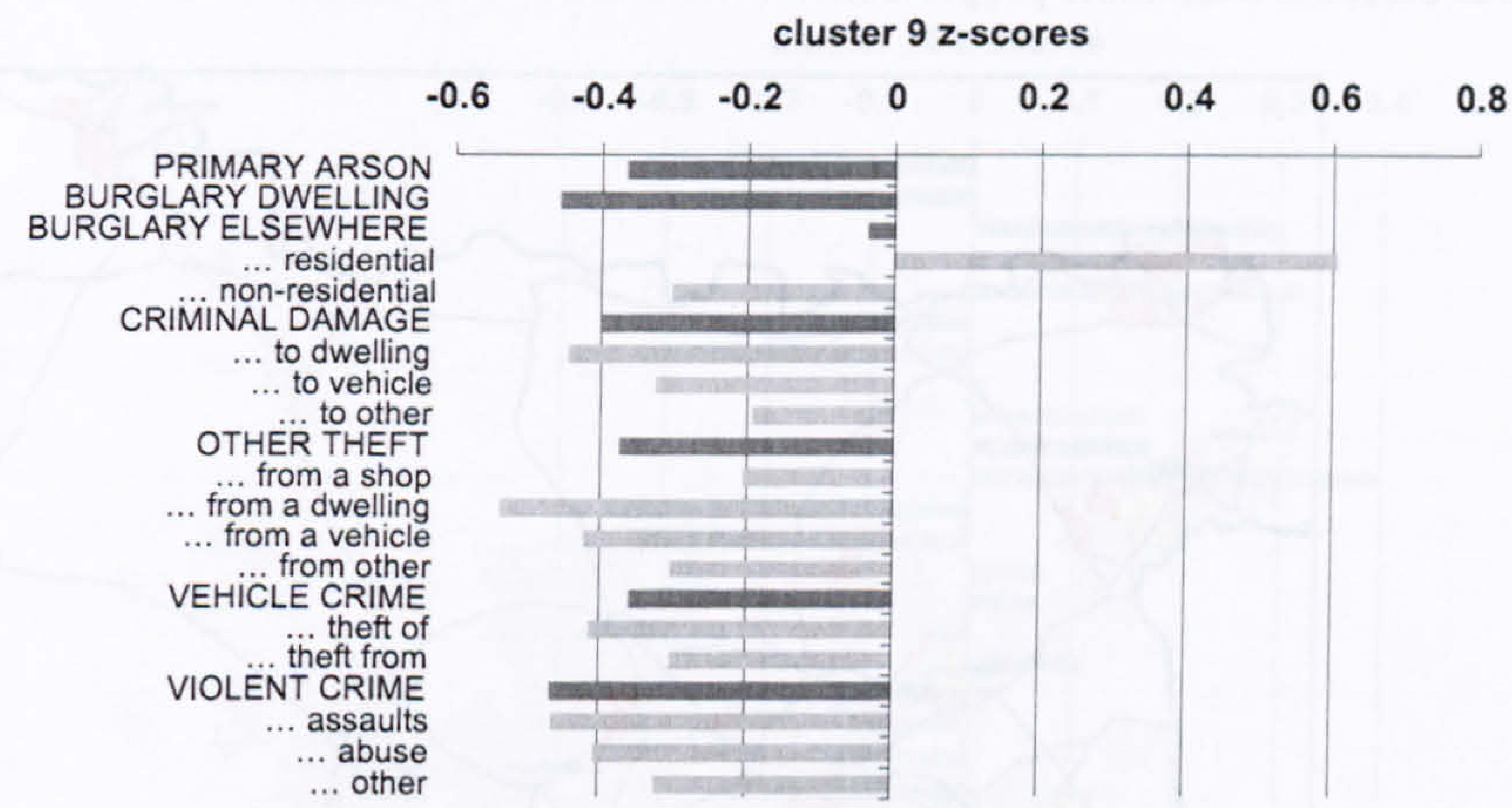


Figure 6.35: Standardised scores for crime variables, for 2003/04, for group 9.

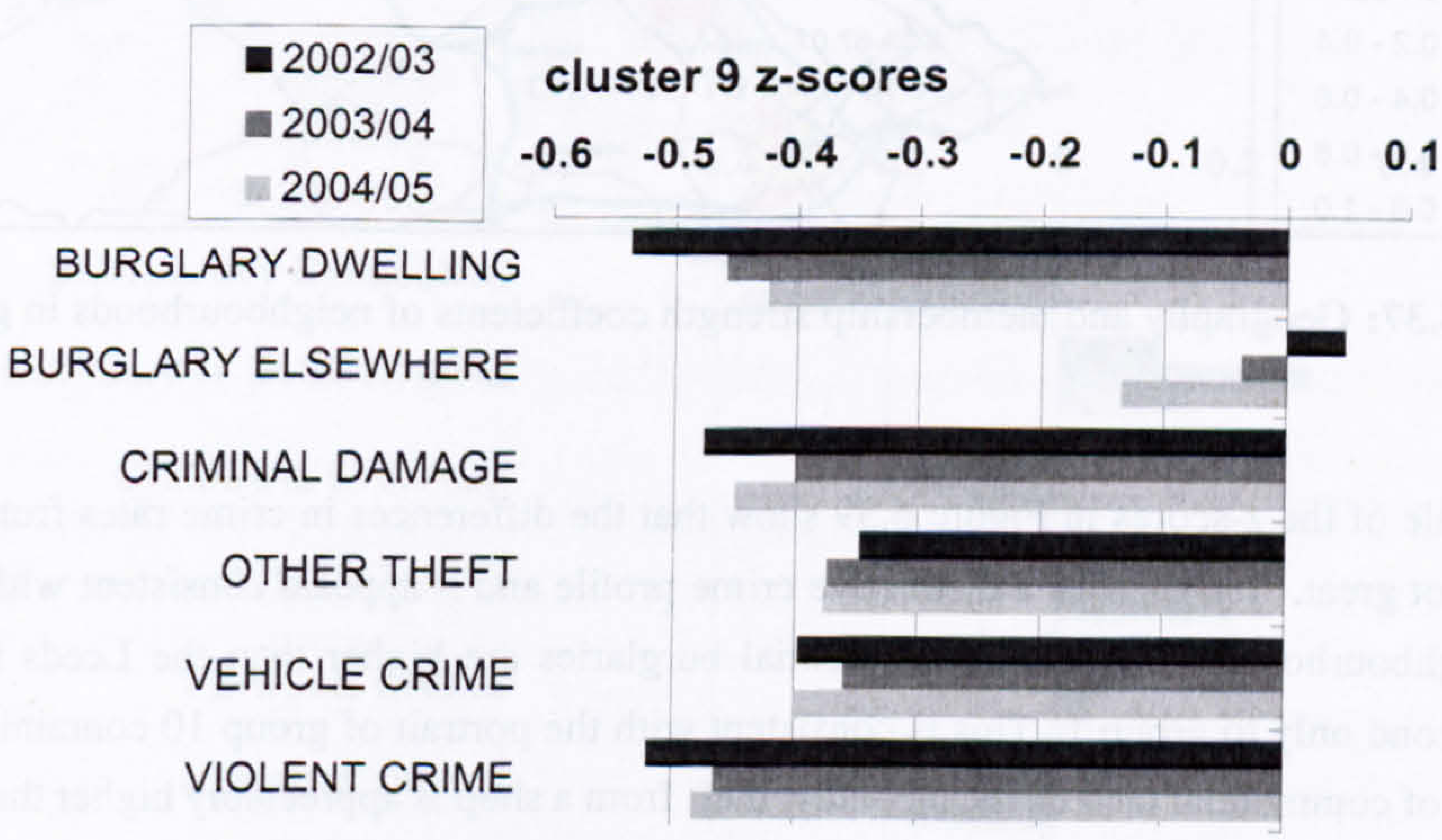


Figure 6.36: Change in group 9 standardised crime rates, 2002/03 to 2004/05.

6.2.11 Group 10

57 Neighbourhoods. Highest strength coefficient is .63, SCA 67.07 - Morley.

The silhouette width for this group is only .065 and the highest strength coefficient is low compared to those of other groups. These measures suggest a relatively weak group. In geographical terms these neighbourhoods are coincident with traditional town centres and other places where there is mixed land use (Figure 6.37). Generally speaking, neighbourhoods are found in the south and west of the district although there are also cases in Wetherby, Garforth and Rothwell. The z-scores of the group profile (Figure 6.38) show that many variable means are not much above or below the district average. Characteristics that do give the group some interesting definition are are a higher than average concentration of pubs (LU01) and shops (SH01). Also suggesting mixed land use are a higher than average number cars at the workplace (CA01), a slightly higher proportion of larger buildings and a higher than usual ratio of non-residential to residential postcodes (RE01). Residentially, there is a greater tendency for properties to be terraced (HT03) and owner occupied (TE01) although the variables for housing heterogeneity (HT06) suggests a greater than usual mix of house types. The age structure is biased toward older people and the social class structure is rather mixed.

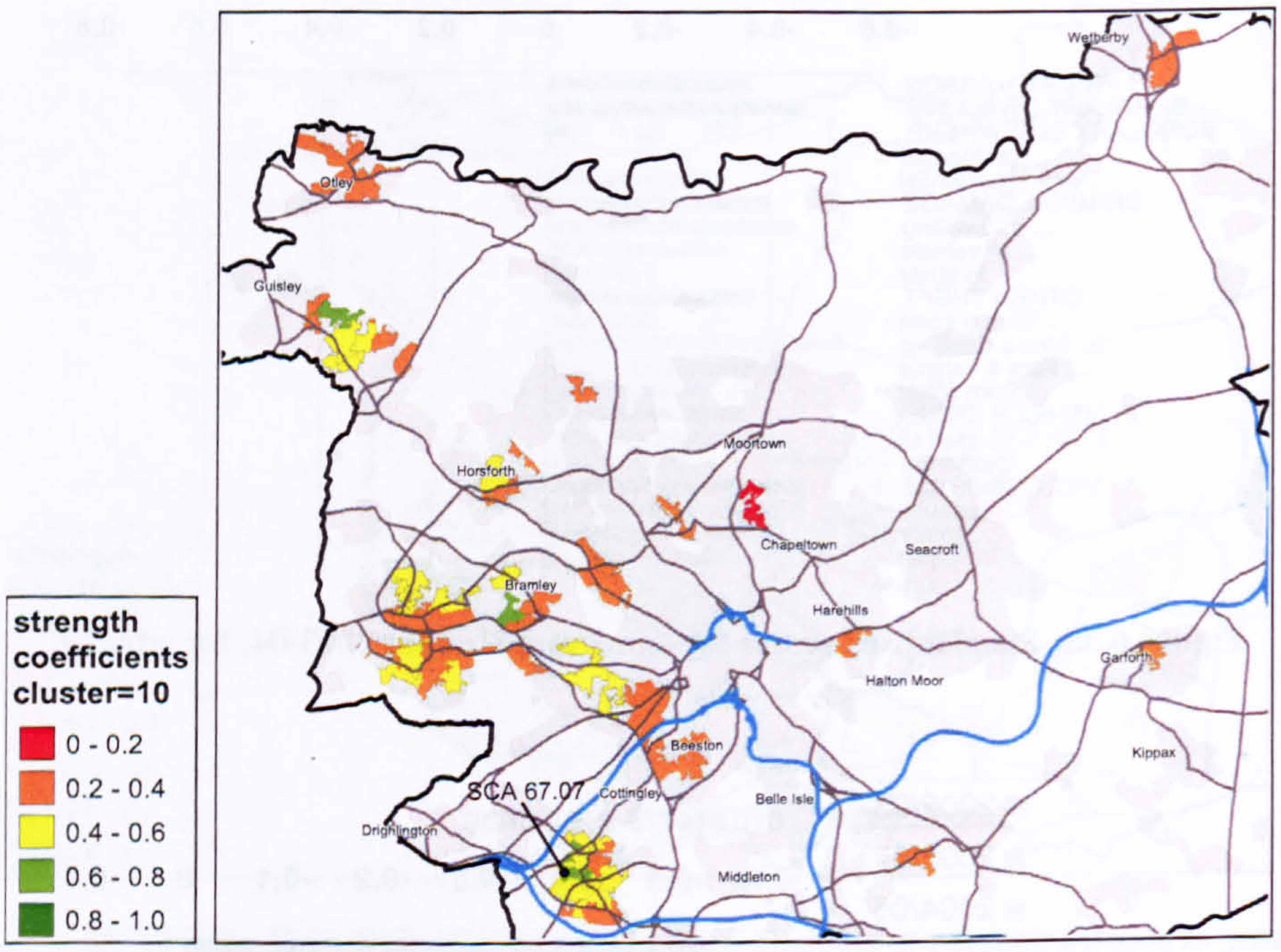


Figure 6.37: Geography and membership strength coefficients of neighbourhoods in group 10.

The scale of the z-scores in Figure 6.39 show that the differences in crime rates from the Leeds norm are not great. Yet, there is a distinctive crime profile and it appears consistent with the profile of the neighbourhoods. Firstly, non-residential burglaries are higher than the Leeds norm by an amount second only to group 1. This is consistent with the portrait of group 10 containing a greater proportion of commercial land uses. Secondly, theft from a shop is appreciably higher than the Leeds norm, again second only to group 1, and is again consistent with the commercial land use, especially retail.

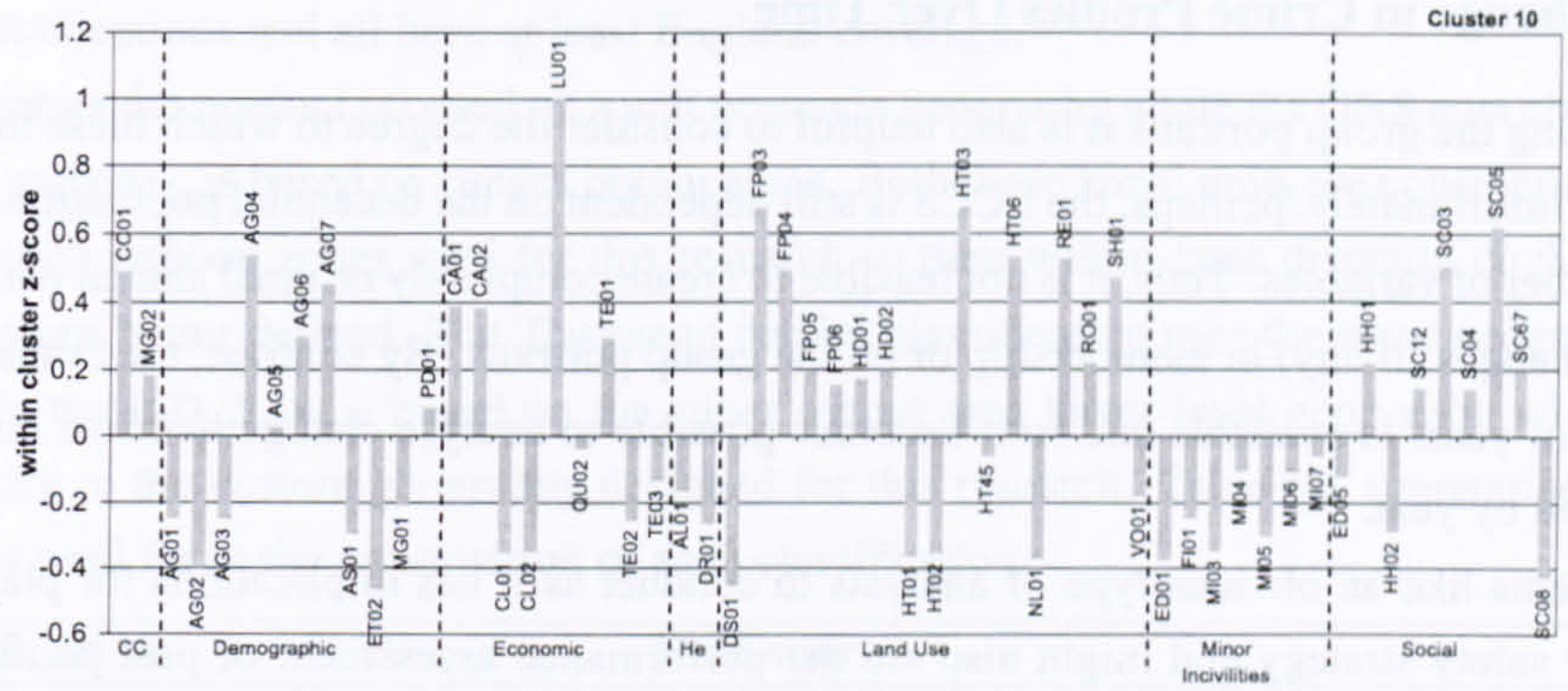


Figure 6.38: Standardised scores for variables defining group 10.

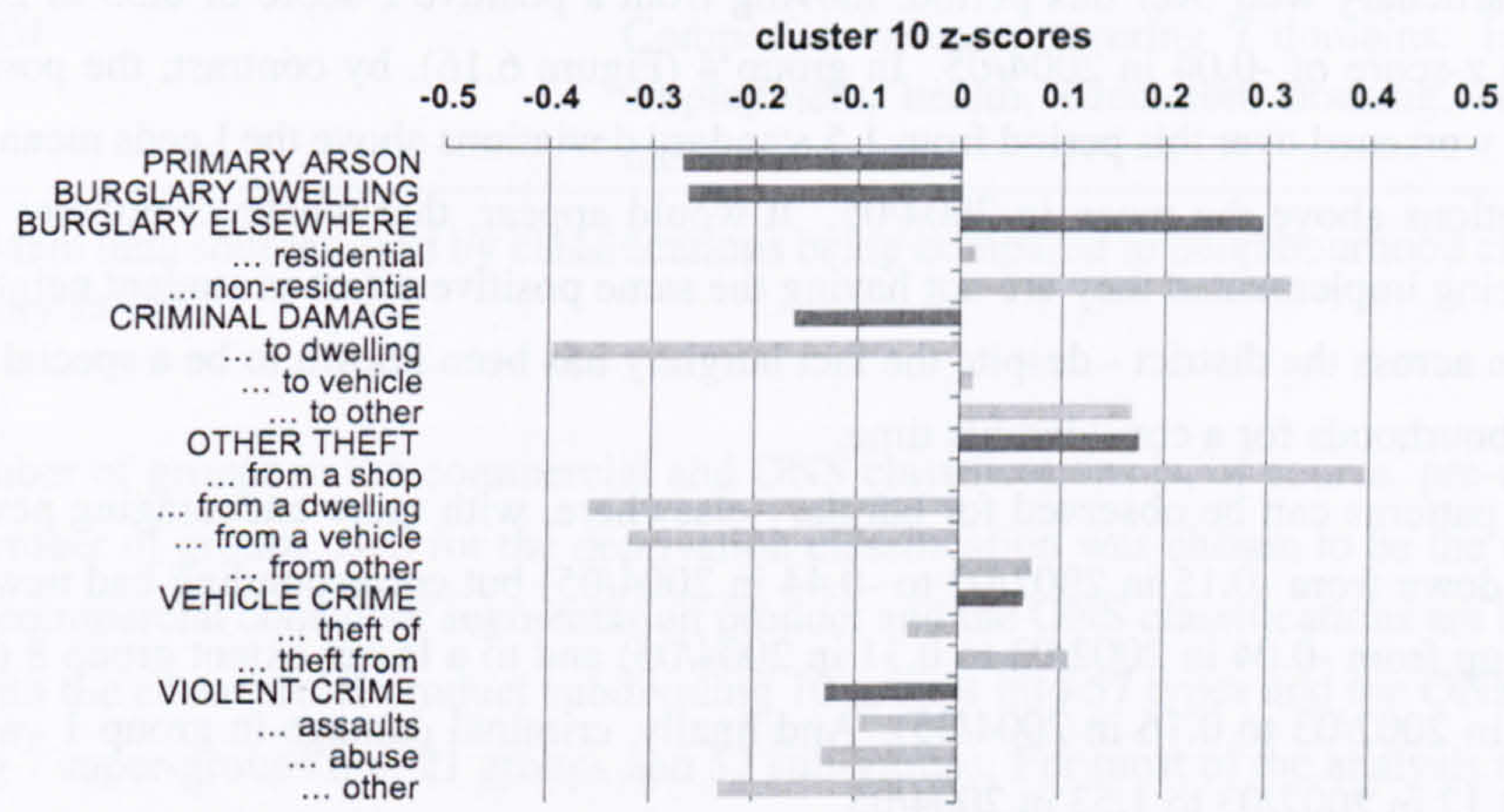


Figure 6.39: Standardised scores for crime variables, for 2003/04, for group 10.

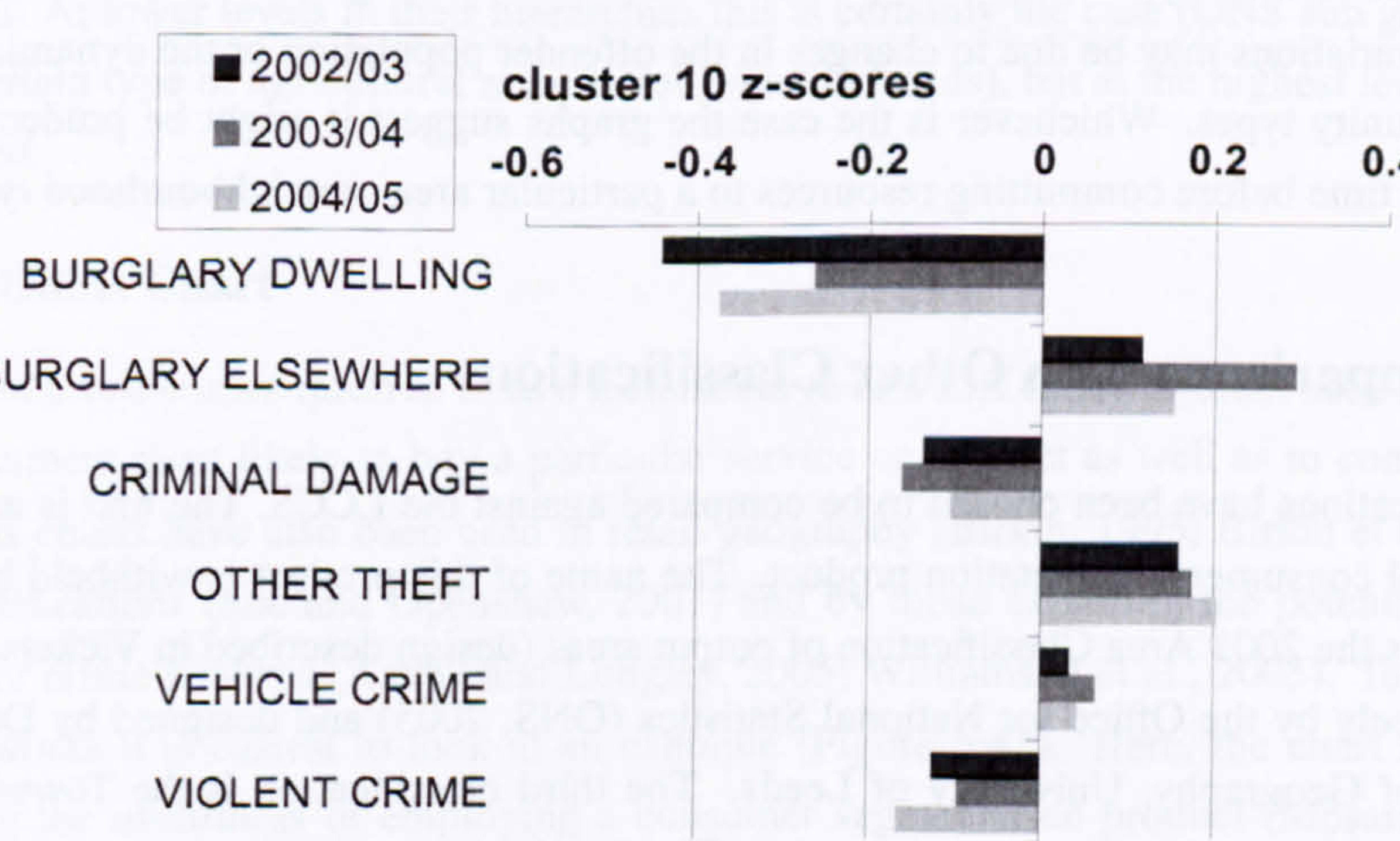


Figure 6.40: Change in group 10 standardised crime rates, 2002/03 to 2004/05.

6.2.12 Change in Crime Profiles Over Time

Before leaving the group portraits it is also helpful to consider the degree to which these may change over time. Unfortunately, perhaps, the LCCS is still dependent on the decennial population census for a large number of variables. Thus, it is not feasible to create completely original annual partitions and reflect on changes (if any) in membership or on the group portraits. By contrast, the crime data for a succession of years is available and so it becomes possible to analyse change in crime rate profiles, by group and by year.

This seems like an obvious type of analysis to conduct as it has implications for planning new community safety strategy and might also aid the performance assessment of past policies. When for example, the Leeds trend for a particular crime type is downward it would be useful to know whether the reductions have been felt equally across the city or whether some neighbourhood types have seen a disproportionately higher fall than others. Taking burglary dwelling as a case in point, between 2002/03 and 2004/05 there was been a drop of 42.1% in the incidence of this type of offence across Leeds. Scanning the crime profile histories reveals that group 6 neighbourhoods (Figure 6.24) benefitted particuary well over this period, moving from a positive z-score of 0.35 in 2002/03 to a below norm z-score of -0.04 in 2004/05. In group 4 (Figure 6.16), by contrast, the position of the group mean worsened over this period from 1.5 standard deviations above the Leeds mean in 2002/03 to 2.2 deviations above the mean in 2004/05. It would appear, that whatever policies or policing styles are being implemented they are not having the same positive effect in student neighbourhoods as elsewhere across the district - despite the fact burglary has been known to be a special problem in these neighbourhoods for a considerable time.

Similar patterns can be observed for burglary elsewhere, with more encouraging news in group 6 (z-scores down from -0.15 in 2002/03 to -0.44 in 2004/05) but corresponding bad news for group 7 (z-scores up from -0.04 in 2002/03 to 0.31 in 2004/05) and to a lesser extent group 8 (z-scores up from -0.10 in 2002/03 to 0.16 in 2004/05). And finally, criminal damage in group 1 swung from a z-score of 2.12 in 2002/03 to 1.53 in 2004/05.

Apart from the apparent trends described above, the graphs generally indicate that group crime profiles are not static over time. Some of the dynamics may be the result of area-targeted interventions or neighbourhood type-specific responses to interventions, negative and positive. In addition, some of the variations may be due to changes in the offender population or the dynamics of specific target/opportunity types. Whichever is the case the graphs suggest it might be prudent to consider changes over time before committing resources to a particular area or neighbourhood type.

6.3 Comparisons with Other Classifications

Four classifications have been chosen to be compared against the LCCS. The first is an example of a commercial consumer segmentation product. The name of this product is withheld by the author. The second is the 2001 Area Classification of output areas (design described in Vickers et al., 2005), published freely by the Office for National Statistics (ONS, 2005) and designed by Dan Vickers at the School of Geography, University of Leeds. The third classification is the Townsend material deprivation deciles used in the previous section. The fourth is a similar deprivation classification but this time constructed from scores in the *English Indices of Deprivation 2004* (IoD 2004) published by the Office of the Deputy Prime Minister (ODPM, 2004). All of these can be considered as general purpose classifications and all have at least England coverage.

The commercial product is based on a unit postcode geography while the ONS area classification, as its name suggests, is based on census output areas. Both these zonal units are considerably smaller than the neighbourhood zones used for this research so their within-zone diversity ought to be less and their groups better defined. The Townsend deciles classification uses the same geography as the LCCS while the IoD 2004 is based on the super output area lower-level geography which is very similar in size to the custom geography designed for this research. Table 6.1 summarises the main data sources used for in the construction of each classification.

Classification	Data sources
commercial consumer segmentation	2001 census, council tax band records, the electoral register, share ownership registers, companies house registers and consumer credit databases.
ONS classification of OAs	2001 censuses of England and Wales, Scotland and Northern Ireland.
Townsend material deprivation	4 variables from 2001 census: unemployment, overcrowding, no car and not owner-occupier.
IoD 2004	Composite measure covering 7 domains: income, employment, health, education, housing, environment and crime.

Table 6.1: Main data sources used by classifications being compared to neighbourhood classification for community safety

The number of groups in the commercial and ONS classifications is, of course, pre-determined, while the number of groups used for the deprivation classification was chosen to be the same as the LCCS. The commercial consumer segmentation product and the ONS classifications are hierarchical in nature, with the commercial product subdividing 10 groups into 57 types and the ONS classification dividing 7 super-groups into 21 groups and 52 sub-groups. For most of the analysis that follows the highest levels (10 and 7) are used. With fewer groups, the ONS classification would be expected to partition Leeds less well than a classification with 10 groups. Theoretically, as both of the classifications are national in coverage, it is also feasible that some of their groups might not be represented in Leeds at all. At lower levels in their hierarchies this is certainly the case (ONS sub group 3b2, for example, a certain type of agricultural area, is not found in Leeds), but at the highest level all groups are represented.

6.3.1 The Gains Chart

The Gains Chart, sometimes referred to as a Lorenz curve or a Lift Chart, is often used by marketers to target consumers most likely to buy a particular service or product as well as to compare market models. Gains charts have also been used in retail geography (Birkin, 1995; Birkin et al., 2002), to compare classifications (See and Openshaw, 2001) and by those exploring the potential of geodemographics for crime analysis (Ashby and Longley, 2005; Williamson et al., 2005). To see how the Gains Chart works it is easiest to look at an example (Figure 6.41). Here, the chart is concerned with analysing the usefulness of employing a consumer segmentation product (Mosaic) to explain the all-crime rate.

The diagonal line represents a situation where there are no underlying processes that might cause a disproportionately large number of crimes to be committed in areas with a disproportionately small

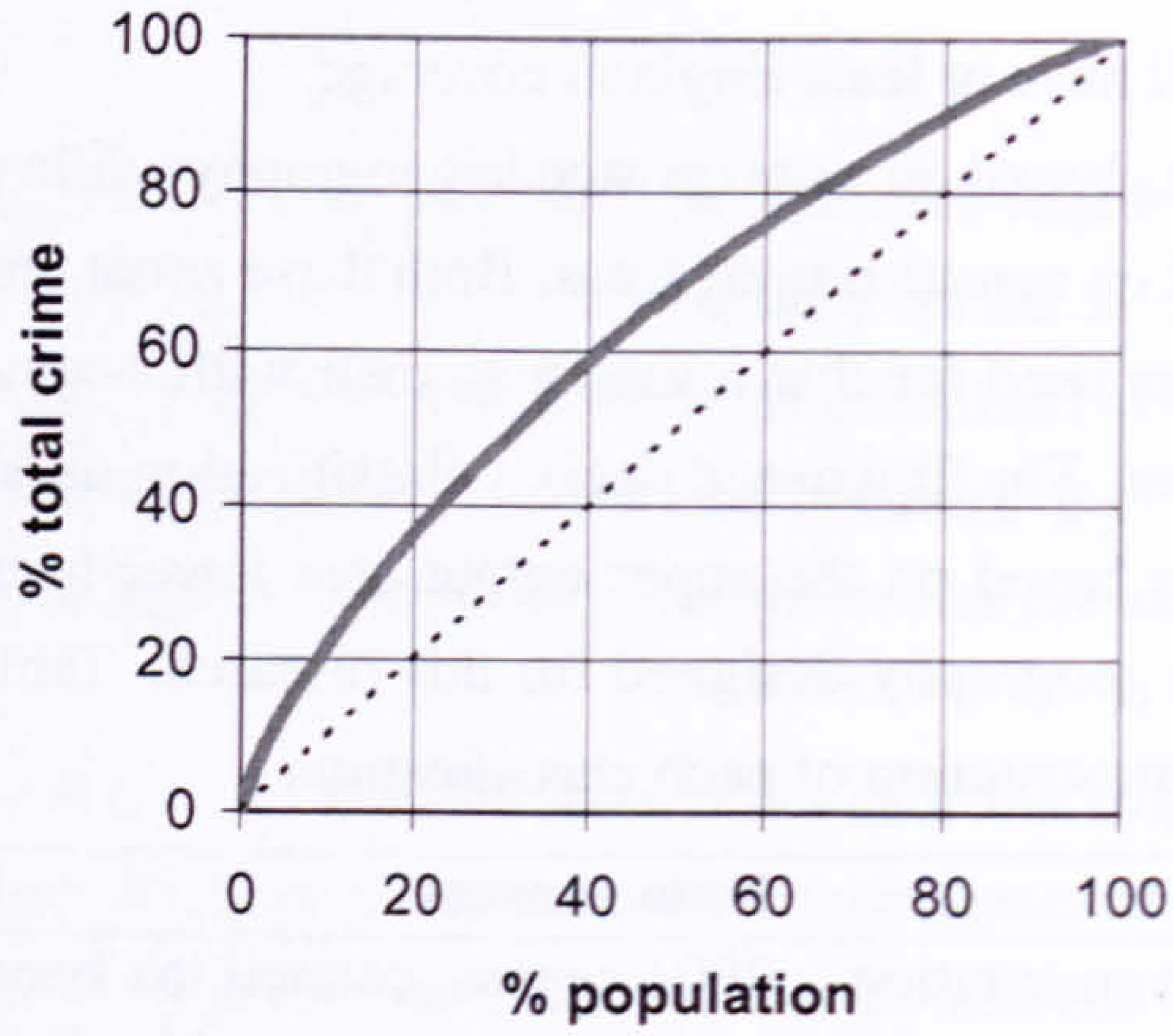


Figure 6.41: Gains chart of total crime versus adult population. Data is taken from Ashby and Longley (2005)

total population. The curve represents the level of discrimination being provided by the classification and can be used to make the observation that areas comprising 40% of the population suffer 60% of all crime, for example. The further the curve deviates from the diagonal the better the classification is at discriminating the variable on the y axis.

From a community safety policy perspective this type of technique can also be useful for identifying where limited resources should be deployed to effect the most benefit.

The technique for creating the Gains Chart can be described thus,

1. For each group in the classification, calculate the total population and the total amount of the variable of the y axis, in this case crime.

2. Divide the crime value by the population value and multiply by 100 to create a Penetration Index

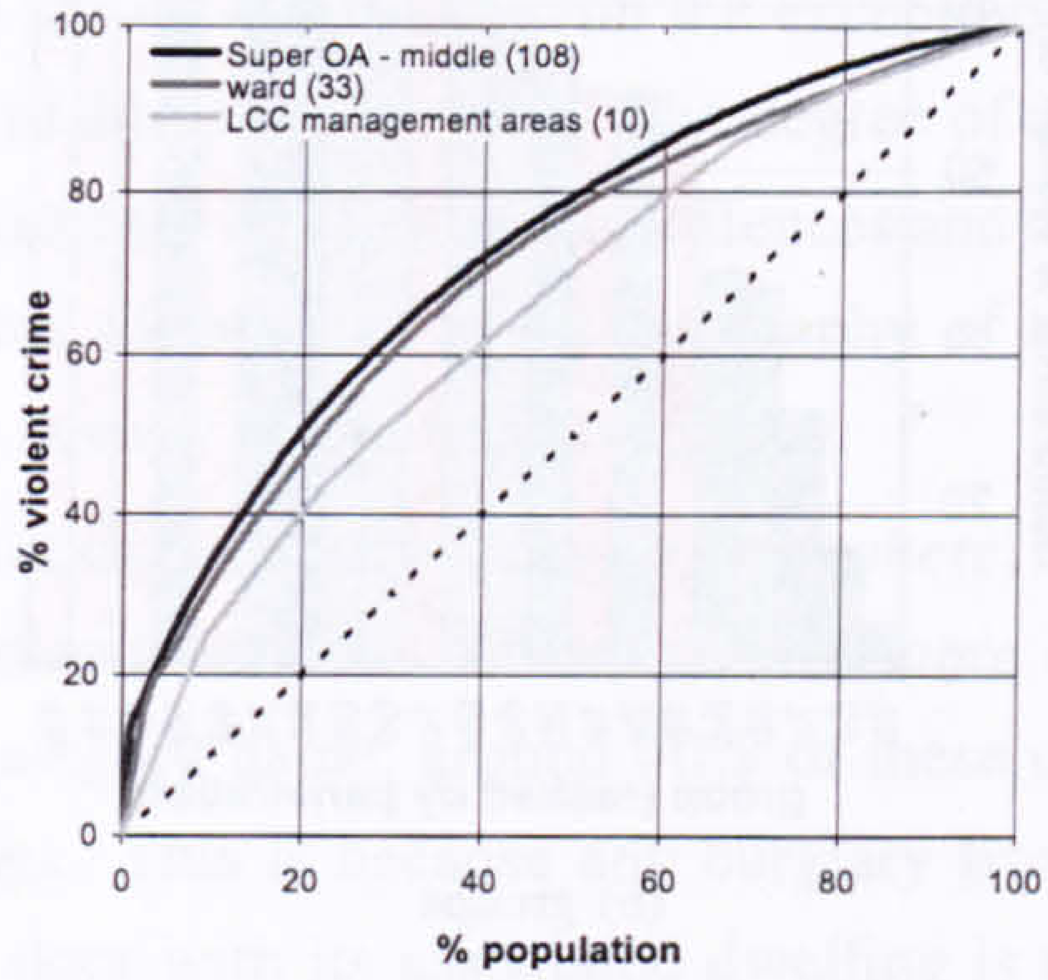
3. Order the groups by their penetration index for crime, highest value first.

4. Calculate cumulative percentages of population and crime, and then plot them.

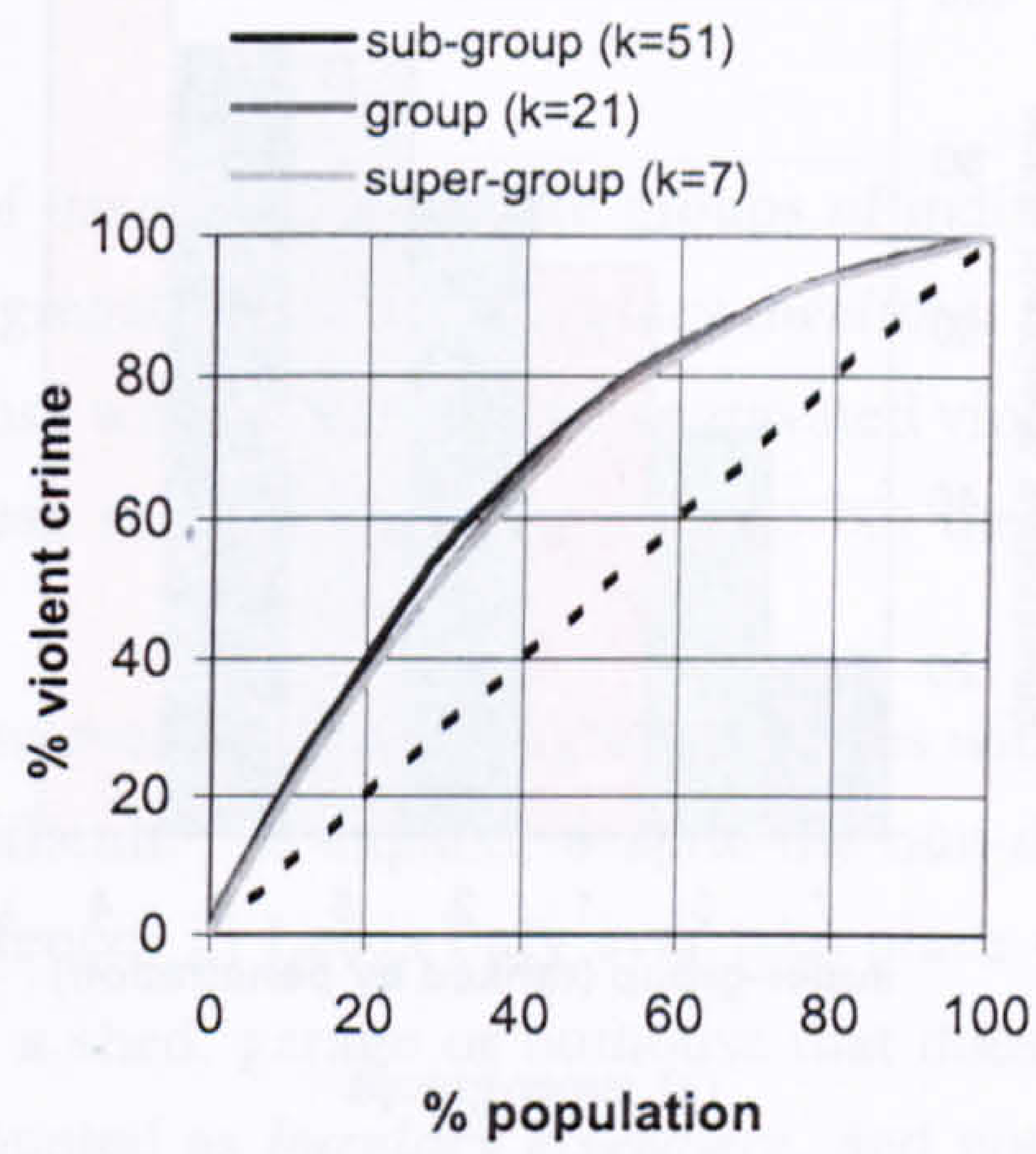
6.3.2 Effect of Scale on Gains Analysis

Before making the comparisons it is worth exploring a couple of general issues that can effect a gains analysis. The first of these is the impact that using classifications of different zone sizes can have on results and the second has to do with the handling of hierarchical classifications.

Figure 6.42(a) illustrates the first point by showing how existing administrative geographies segment Leeds in violent crime terms. In this example it is clear that even though administrative geographies are often criticised for being arbitrary they nevertheless do reflect the geography of crime in Leeds to some extent. As might be expected the performance in gains terms increases as zone size decreases. At the management area level there is more diversity within each area than at ward level and that if resources to tackle violent crime could only be targeted at, say, the population of three wards it would be better to identify these three wards with the highest penetration indices rather than picking the management area with the highest penetration index.



(a) administartive geographies



(b) hierarchical classification

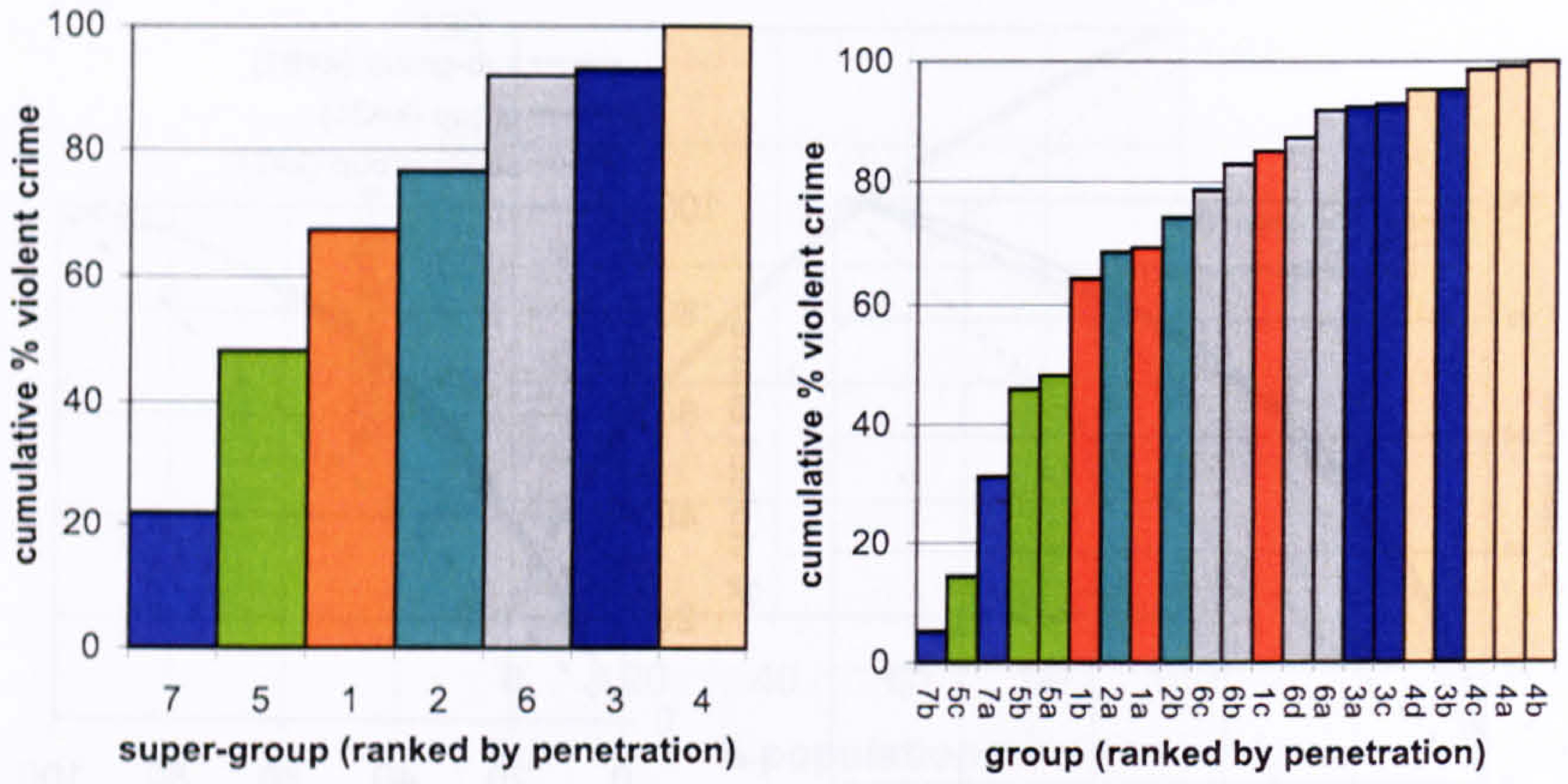
Figure 6.42: Gains chart of violent crime versus adult population for different administrative geographies and different levels of a hierarchical classification.

There is a smaller benefit by sub-dividing the wards down to Super OA lower level. There would be benefit in the sense that parts of wards with low crime rates would not be targeted unnecessarily (and the converse) but the net improvement is not large. If the local authority had resources to eradicate crime from a given percentage of the population the overall effect on the district crime rate would be little different if a ward or Super OA lower level geography was used to target the resources.

Hierarchical classifications also pose an interesting problem for gains analysis. One might expect that the increase in number of groups would improve a classifications' power to discriminate crime and that the gains curve would duly bulge out further from the diagonal, as was seen in Figure 6.42(a). However, with the administrative geographies the actual zone size used to sample the crime is changing and thus each segmentation is affected by aggregation effects to different degrees. In a hierarchical classification, however, the zone size stays the same and it is just the number of classes that varies. The aggregation effect remains constant and the curve stays looking the same at every level (Figure 6.42(b)). Penetration indices for sub-classes are effected by a reduction in group diversity which causes the super-class membership to fragment along the curve, in some cases. The effect can be seen if the gains curve is replaced by column graphs colour coded to reflect the class membership hierarchy (Figure 6.43). Figure 6.43(a) shows that super-group 1, for example, appears to have the greatest within-group diversity in violent crime rates. As the super-group is split into groups 1a, 1b and 1c (Figure 6.43(b) they 'spread out' the most along the gains curve although the net effect of this disaggregation is zero in terms of the overall gain.

6.3.3 Gains Comparisons by Crime Type

Prior to analysis, the coverage of each classification was manipulated to exclude the area of the city centre area not covered by the LCCS. Figure 6.44 shows the gains charts for the major recorded crime aggregations and deliberate arson. In addition to plots for each of the classifications under consideration a theoretical 'best case' plot has been added. This plot is created by treating each neighbourhood as its own group. The extent to which the real classifications vary from this 'best



(a) super-groups (b) groups

Figure 6.43: Gains column graphs of violent crime versus adult population for different levels of the ONS hierarchical classification of output areas. (Columns are fixed width, not sized proportional to % population).

case' serves as an indicator of the extent to which the classification designs have strayed from one of the main design criteria - to reflect neighbourhood crime profiles.

For all the classifications, discrimination appears best for *violent crime*, *criminal damage* and *arson* (some of which is also captured in criminal damage). The LCCS performs better than the consumer segmentation product and the ONS classification but there is little to separate it from the IoD 2004 deciles classification. For the remaining crime types - *burglary elsewhere*, *burglary dwelling*, *other theft* and *vehicle crime* - the gains performance is less good although the LCCS performs best - albeit by a slim margin.

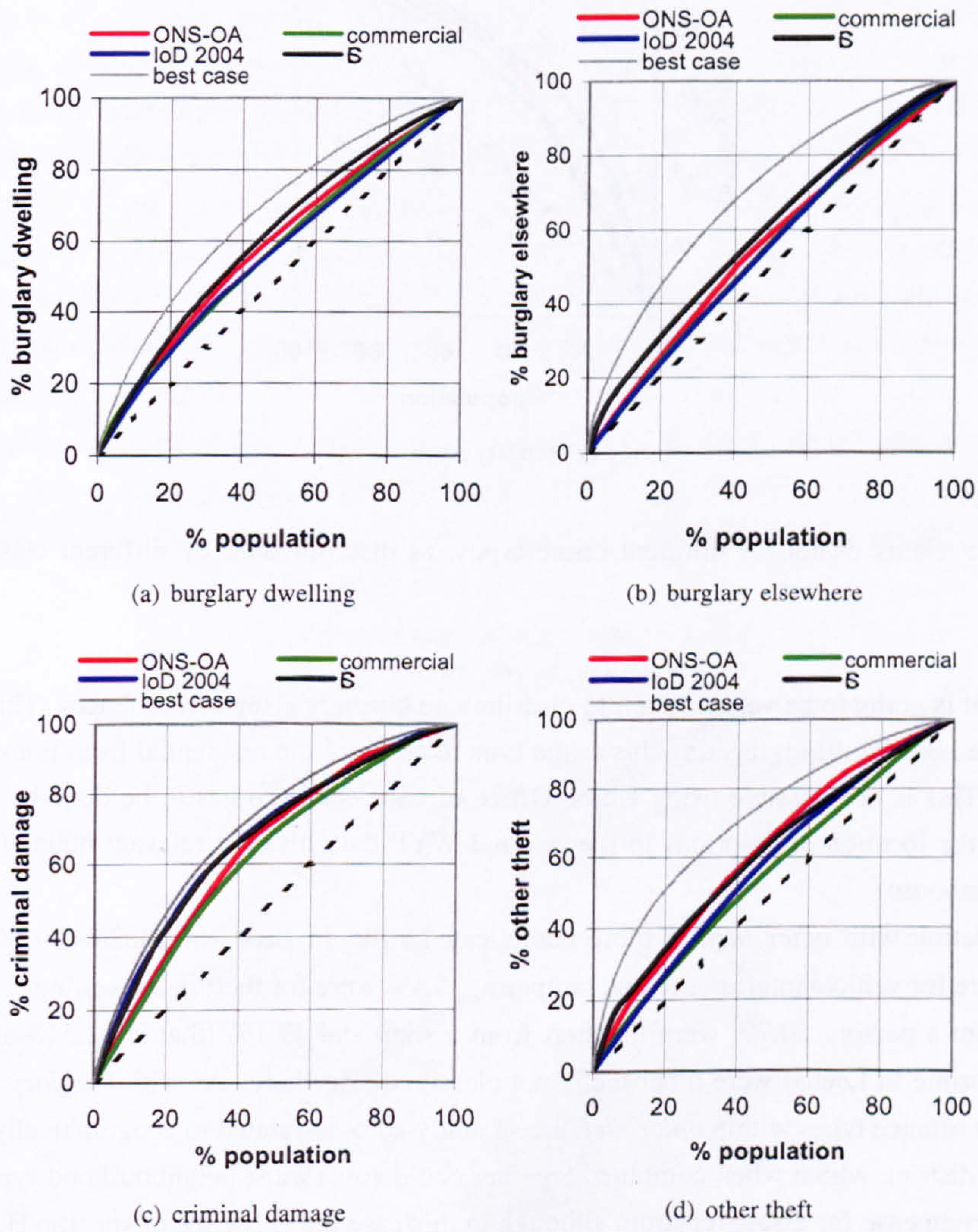
The largest differences between the best case segmentation of neighbourhoods and the LCCS are for *other theft*, *burglary elsewhere* and, to a lesser extent, *vehicle crime*. A close match to the best case has been achieved for *criminal damage* and *violent crime*. It is *primary arson* that has the steepest 'best case' curve of all crime types. The LCCS and IoD 2004 classifications match it quite well while the commercial classification discriminates this type of crime quite poorly.

The results would have looked quite different in some instances had the city centre been included as a *de facto* eleventh group; the LCCS and the deprivation partitions altered to 11 groups with the city centre included; and the ONS and commercial classifications left in their original form, that is with the city centre included. Figure 6.45 illustrates the point being made here by showing the effect of excluding just one lower level Super OA (covering the city centre) from the IoD 2004 deciles classification. The PI for the decile that includes this city centre OA (decile 7, where 10 is most deprived) leaps from 119.9 to 383.7 when considering *other theft* offences. To an extent this is a self fulfilling prophecy because the IoD 2004 contains a crime domain variable which captures some of the *other theft* being measured here. The extent is checked, however, by the fact that the *theft* crime grouping used in the IoD only accounts for 27% of the *other theft* recorded in the city centre SOA zone in 2003/04. Nevertheless, including the city centre (as defined by SOA E1011365) - which accounted for 12% of all crime in Leeds in 2003/04, but just 0.23% of the residential population - would probably flatter a gains analysis for any reasonable classification, which seems another good reason to exclude it from the analysis.

6.3.4 Disaggregating Gains

The results shown so far are, with the exception of *arson*, for aggregate groups of individual Home Office recordable offence types. The degree of aggregation varies. *Burglary dwelling*, for example, only disaggregates into aggravated offences and those where there was no aggravated violence. There may be some differences in the geography of these separate types of offence but the target - the dwellings - occupy a common geography.

For a crime type such as *burglary elsewhere*, however, there are geographic issues with the targets that make discrimination of these crimes more difficult. To expand, despite the non-dwelling tag given the category name, around 40% of these offences in Leeds only ever take place where there are dwellings. This is because any burglary from a shed, garage or outhouse that does not have a connecting door with its associated dwelling is counted as *burglary elsewhere*, and not a *burglary dwelling*. The other burglary elsewhere offences are mostly against shops, businesses, schools and the like and thus take place in non-, or less-residential areas. Thus, whichever way the district is



(a) burglary dwelling

(b) burglary elsewhere

(c) criminal damage

(d) other theft

Figure 6.44: Gains charts for different crime types, as discriminated by different classifications.

6.2.6 Group 5

67 Neighbourhoods. Highest strength coefficient is .79, SCA 29.02 - Drighlington.

The silhouette width for this group (.14) suggests a moderately well defined group. The geography of group membership (Figure 6.17) is quite marked, with many neighbourhoods being in, or adjoining, the rural hinterland in the southern half of the district. In contrast to the z-scores for the previous four groups, the z-scores for the variables that define this group are lower - never exceeding ± 1. Those variables that are prominent (Figure 6.18), however, show that these neighbourhoods have a higher than average proportion of natural land (NL01) and people are more likely to own (TE01) their own detached house (HT01). Housing density (HD01) is low and house types are more mixed (HT06) than is average. In contrast to group 2, the class structure is less biased toward professionals (SC12) and more biased toward the middle classes. There are average numbers of younger families, but people aged 25 to 64 (AG04,AG05) are over-represented. Also in contrast to group 2, GCSE pupils are only doing slightly better than average (ED05) and the qualification of people ages 16 to 74 (QU02) are below the district average.

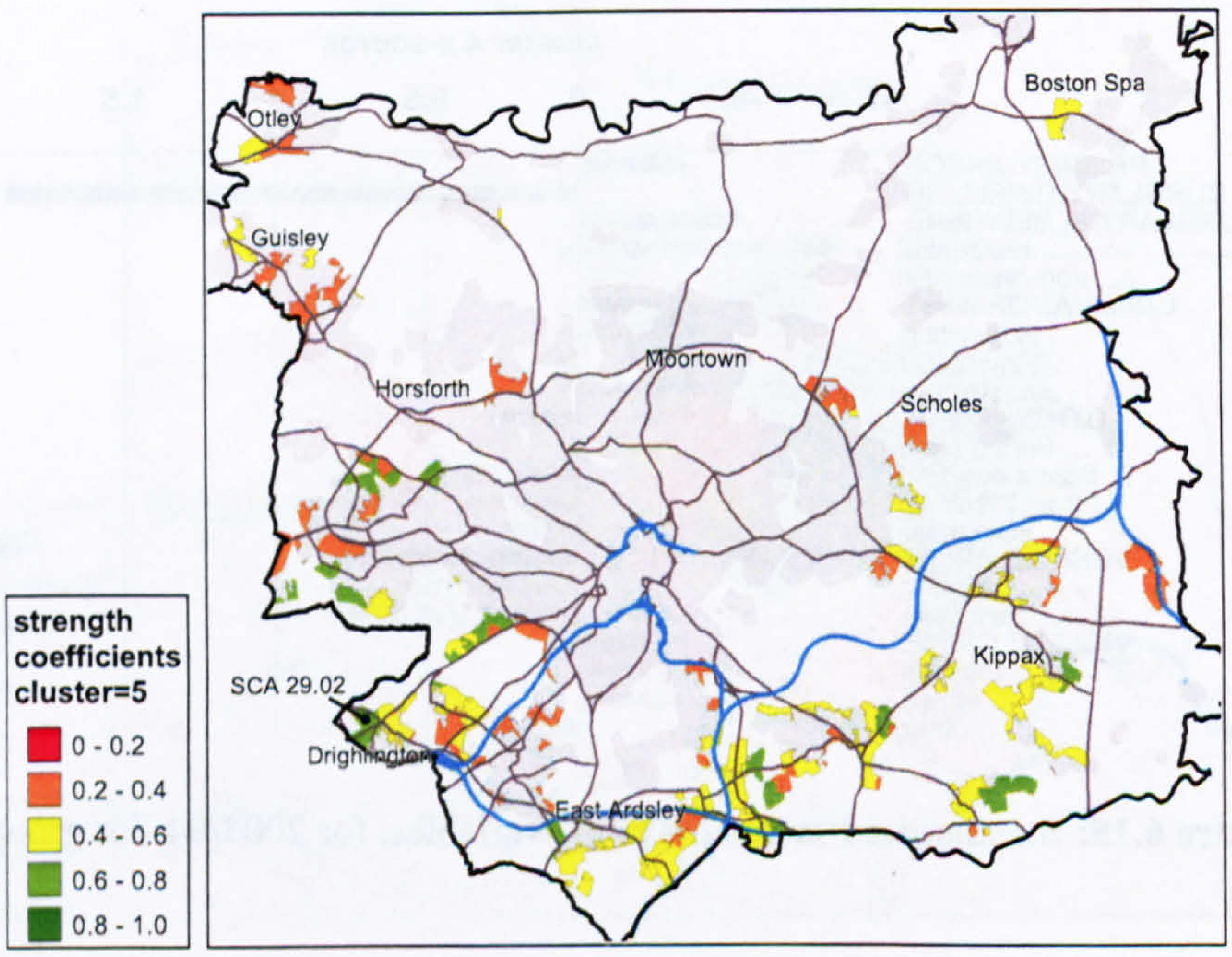


Figure 6.17: Geography and membership strength coefficients of neighbourhoods in group 5.

Violent crime, *criminal damage* and *burglary dwelling* are all lower than the Leeds average (Figure 6.19), while other crime types vary from the Leeds mean by only small fractions of one standard deviation. It is perhaps interesting that *arson* is closer to the Leeds mean than *criminal damage* - which also contains some arson offences. Part of the reason would seem to be atypically high numbers of deliberate vehicle fires close by Oulton and Methley in 2003/04. A similar hotspot for vehicle fires (no pun intended) appears just up the road outside Swillington. The one acquisitive crime type that does stand out higher than the Leeds mean is burglaries from sheds and other residential outbuildings. In part this will be due to increased opportunities afforded by a housing stock which has more detached and semi-detached properties than is usual across Leeds as a whole. Exactly why this type of theft is preferable to others is not clear, although it may be affected by a reduced likelihood of detection and less effort to gain entry.

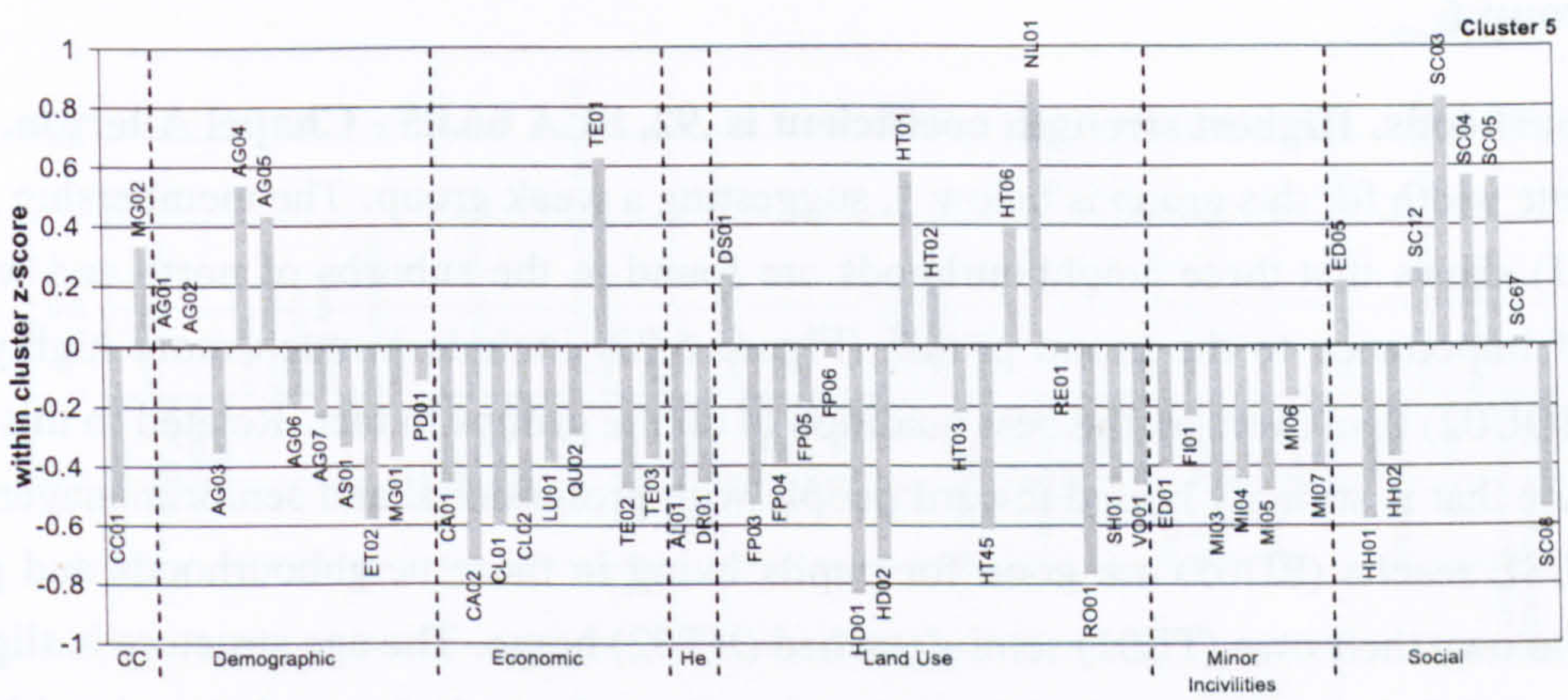


Figure 6.18: Standardised scores for variables defining group 5.

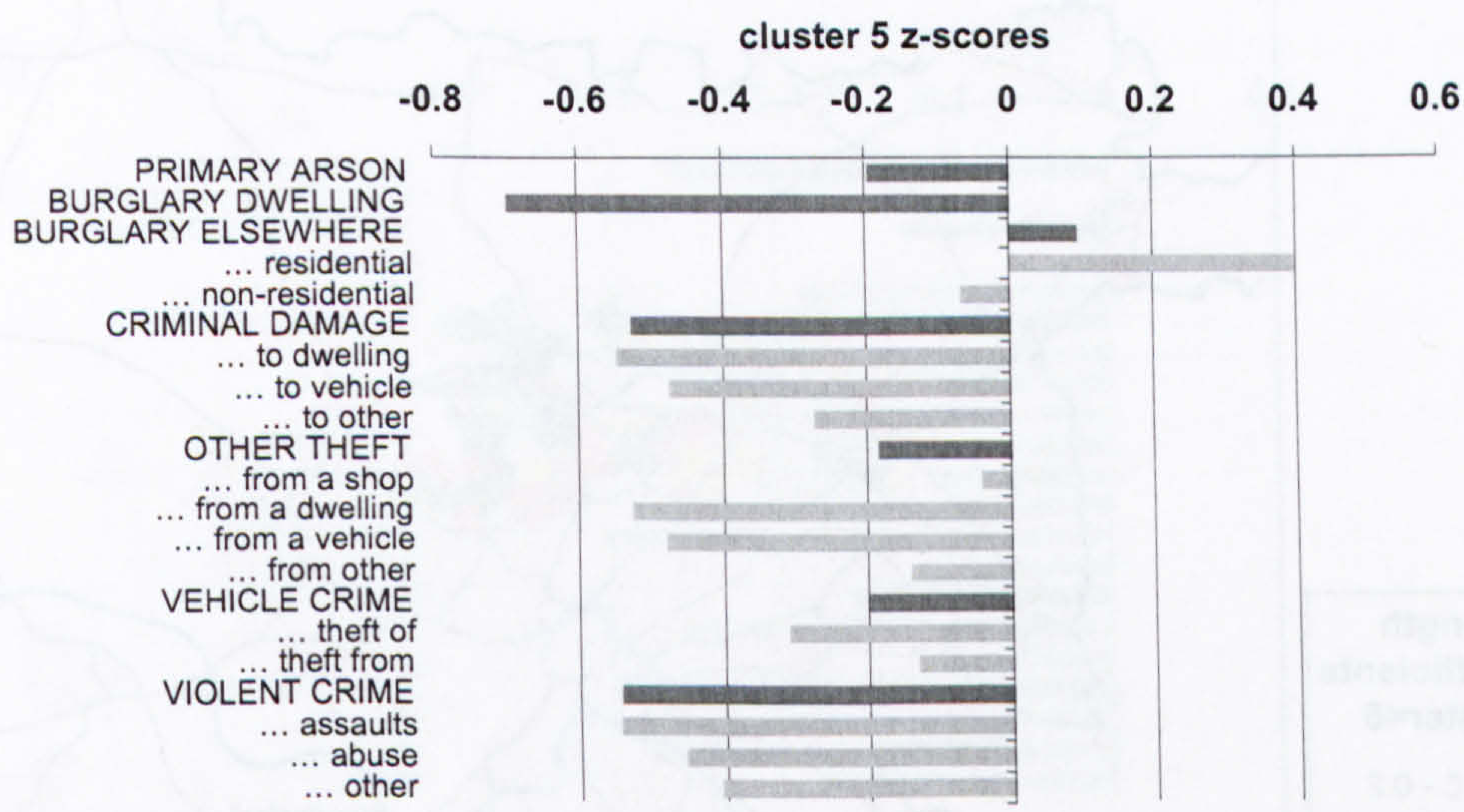


Figure 6.19: Standardised scores for crime variables, for 2003/04, for group 5.

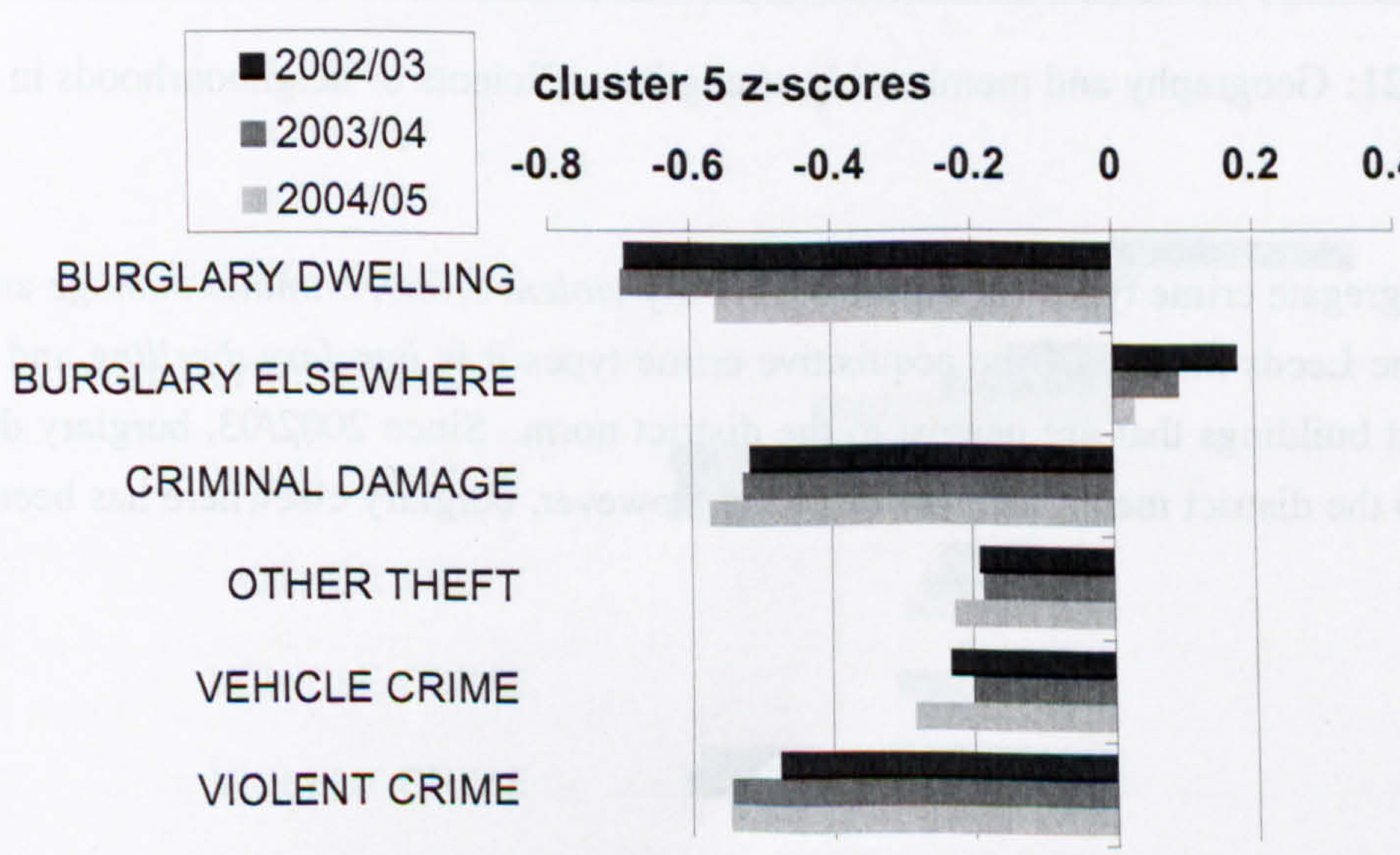
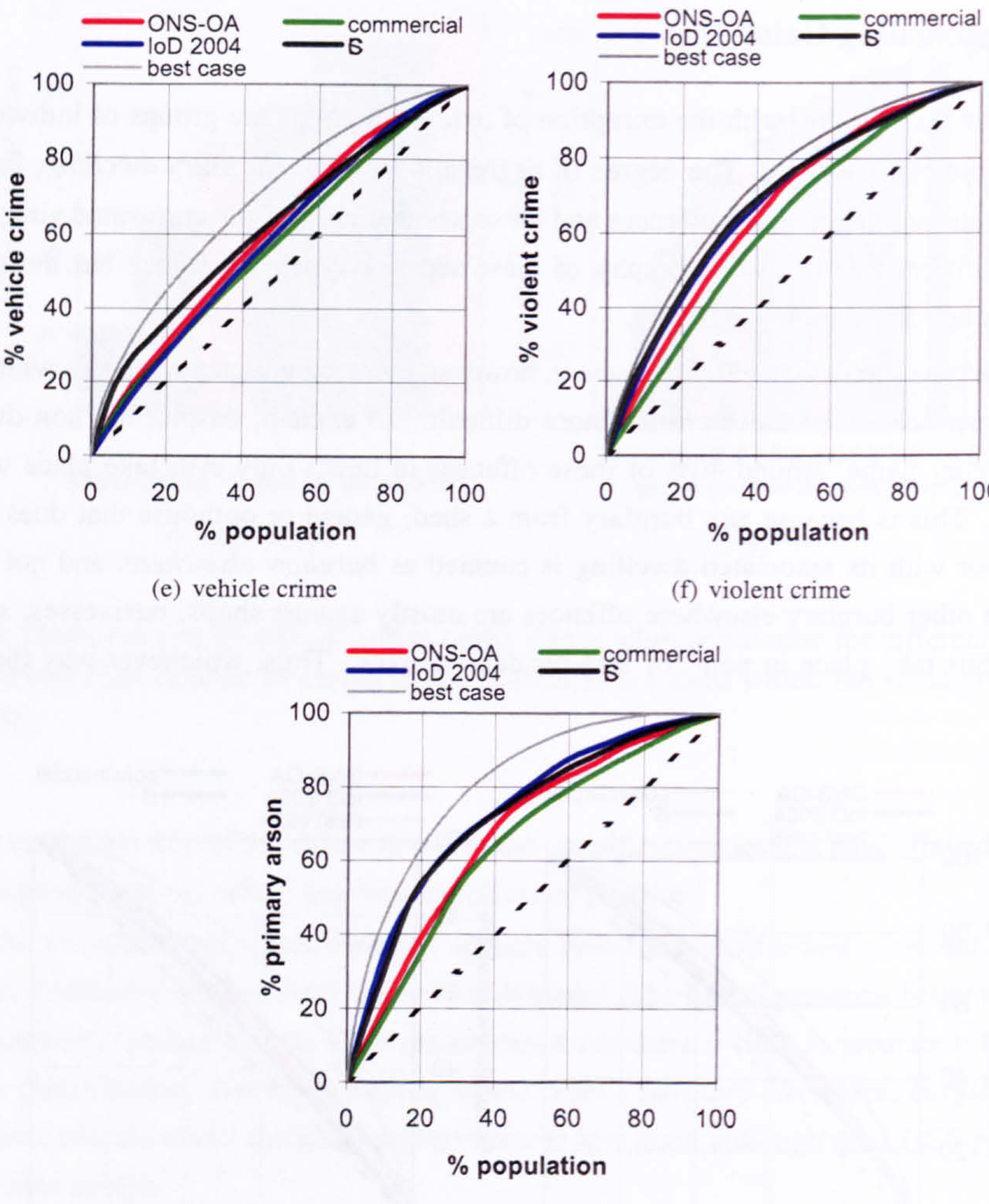


Figure 6.20: Change in group 5 standardised crime rates, 2002/03 to 2004/05.



(e) vehicle crime

(f) violent crime

(g) primary arson

Figure 6.44: Gains charts for different crime types, as discriminated by different classifications (cont.).

partitioned, it is going to be very difficult to discriminate burglary elsewhere offences. Thus, there is a strong case here for disaggregating this crime type to separate the residential from non-residential burglaries. This is not possible using Home Office offence codes so has to be done by fuzzy text searches of the location descriptions in the original WYP data files for relevant nouns (e.g. shed, garage, greenhouse).

The situation with *other theft* is more complicated still. In Leeds in 2003/04, 4.6% of these offences were for vehicle interference and tampering, 5.4% were for theft in a dwelling, 15.8% were for theft from a person, 20.7% were for theft from a shop and 47.1% (that is 12,518 offences, or 10% of all crime in Leeds) were other thefts not classified elsewhere. As with *burglary elsewhere*, the different offence types within *other theft* are spatially auto-correlated to geographically exclusive parts of the district, which when combined together could cover most neighbourhood types. Again, there is a strong case for disaggregation, although in this case the existence of specific Home Office offence codes make the task trivial.

Criminal damage can also be disaggregated to improve the discriminatory chances of the classification. The largest proportion of offences (37.9%) are for damage to a motor vehicle, which might occur almost anywhere. The next largest proportion (34.3%), however, are for damage to a dwelling, so the classification might be expected to reduce the variability of these offences better than criminal damages as a whole.

There is less point in disaggregating the other offence classes. The only exception would be for violent crime. Ignoring the temptation to not fix that which is least broken one can justify this disaggregation by the increasing amount of policy attention that this type of crime is attracting, both locally and nationally. Within Leeds, the increase between 2002/03 and 2003/04 was 42.6%. Within the Yorkshire and Humber GOR the increase was 39% for the same interval and across England and Wales the rise was 14.5%. Whilst most other headline crime types are falling (some dramatically), violent crime has risen sharply. The trend looks even worse if an earlier start period is used, but the difficulty with doing so is knowing how best to correct for the impact of the NCRS in April 2002. Indeed, the NCRS is still cited as being the cause of recent rises (Allen et al., 2005), although it is unlikely to account for all of Leeds' increase.

Also interwoven with the debate on counting rules is the issue of what actually constitutes a violent crime. It appeared that this question started to receive more Home Office attention as the violent crime rate began to rise quite sharply. Before that, however, others had been trying to give the issue greater recognition locally (Shepherd et al., 2004). The usual distinction that is made is between violent crimes that involve a physical attack and all other offences classified as *violent crime*. Shepherd et al. (2004) go further, and suggest that violent crimes be disaggregated into physical attacks, offences involving harassment, abuse or intimidation, and other offences such as resisting arrest or breaking the terms of an ASBO, which may have no injurious consequences (physical or psychological) whatsoever. As of April 2005, the Home Office released new counting rules for *violence against the person*.

To examine the impacts of the issues discussed above some of the crime classes have been disaggregated and the gains analysis repeated (Figure 6.46). The results of disaggregating *burglary elsewhere* are an improvement in the discrimination of non-residential offences but little difference in

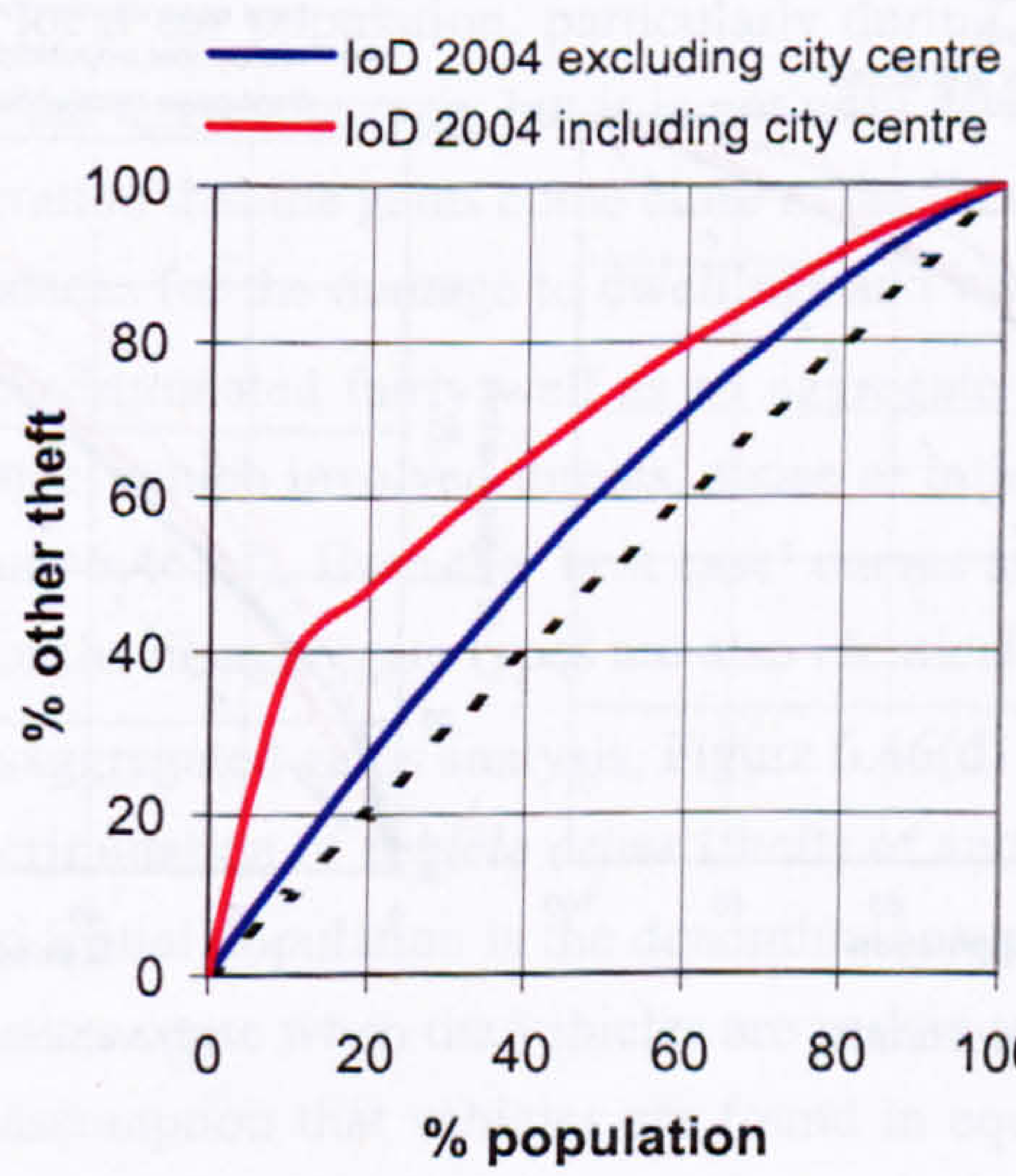
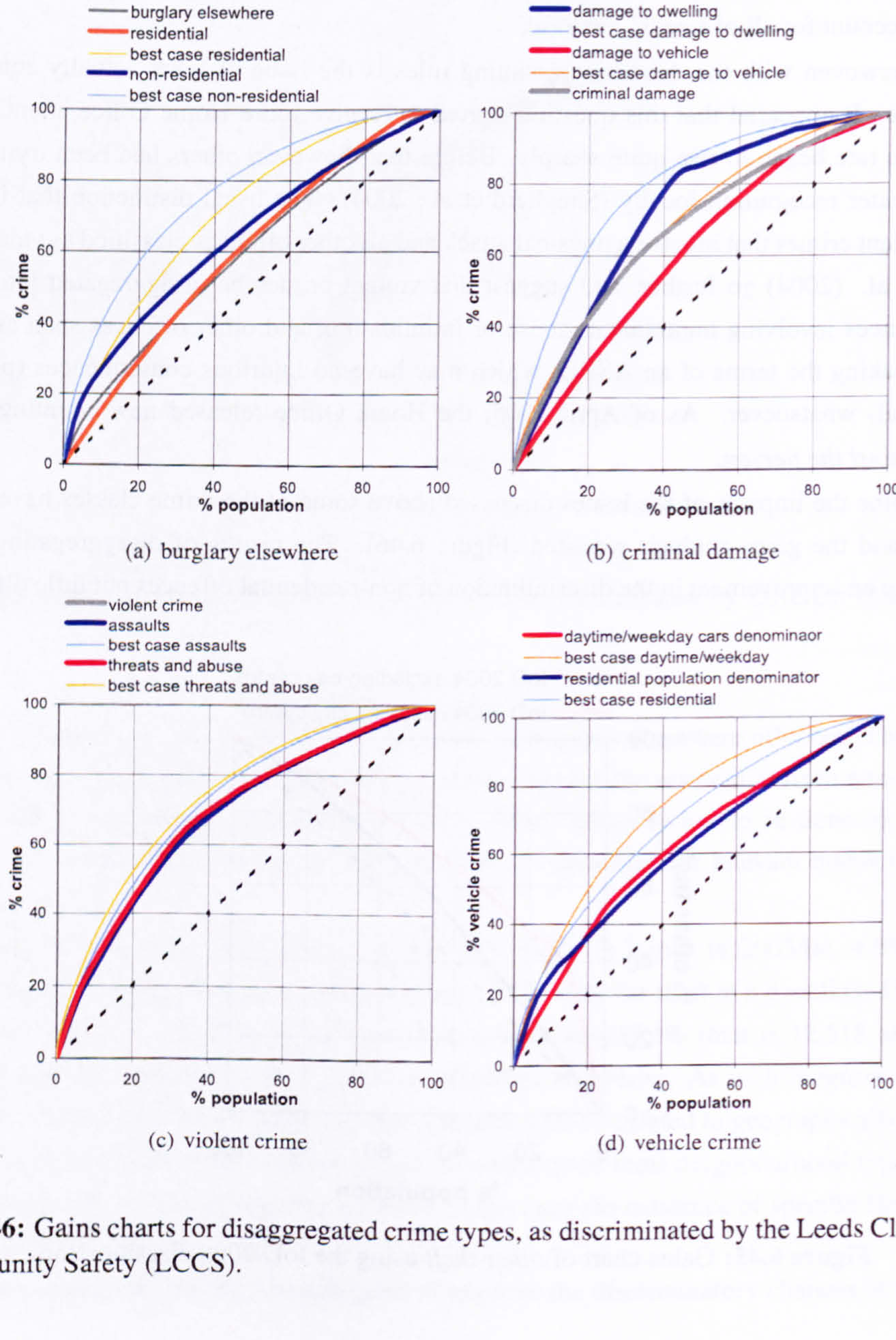


Figure 6.45: Gains chart of *other theft* using the IoD 2004 classification.

the extent to which residential burglary elsewhere is discriminated (Figure 6.46(a)). Both of these outcomes mirror shifts in the best case gains curves for the disaggregate crime types. Table 6.2 provides more information about how the Penetration Indices used to create the disaggregate gains curves vary by group. For the non-residential offences it is groups 1 and 10 that have the highest penetration. Respectively, these groups represent neighbourhoods adjacent to the city centre and neighbourhoods typified by a higher than usual proportion of land used for retail and other non-residential purposes. As the two groups with the most mixed of land use profiles it is encouraging that once *burglary elsewhere* has been disaggregated the classification can discriminate these offences more clearly.

The highest ranking penetration indices for residential offences go to groups 9, 5 and 6. Together, these groups contain 49.0% of the owner occupied housing in Leeds and 31.2% of people aged 16-74 belong to NS-SeC classes 1 or 2. It would seem these are the target neighbourhood types of choice when thieves set out to steal from sheds, garages and similar outbuildings. By contrast, the lowest



(a) burglary elsewhere (b) criminal damage

(c) violent crime (d) vehicle crime

Figure 6.46: Gains charts for disaggregated crime types, as discriminated by the Leeds Classification for Community Safety (LCCS).

Groups ranked by PI (descending)		
Burglary elsewhere	Non-residential	Residential
1	1	9
10	10	5
5	8	6
8	5	10
9	7	8
7	2	7
6	4	2
2	9	1
4	3	4
3	6	3

Table 6.2: Effect on rankings of penetration indices as a result of disaggregating *burglary elsewhere* offences, 2003/04.

ranked groups, 4 and 3, would appear to have fewer or less desireable targets of this type. This makes some sense when one considers that 86.6% of homes in the neighbourhoods contained within groups are either terraced houses or flats and thus much less likely to contain outbuildings. Furthermore, NS-SeC class 1 and 2 residents (who are more likely to have desireable goods in their garages) are much less in evidence - just 16.0% of the 16-74 years population.

Criminal damage offences were already being discriminated fairly well, but improvements can still be made by separating out the vehicle and dwelling related offences (Figure 6.46(b)). As might be expected, the *damage to vehicle* gains were somewhat lower than the gains for criminal damage as a whole and are consistent with the level of discrimination of vehicle thefts (Figure 6.46(d)). The highest gains for thefts, however, take in the studentland (group 4) and old town centres and retail (group 10) alongside central deprived (group 1) and ethically mixed (group 3) neighbourhoods. Whereas damage to vehicles, besides also being prevalent in groups 1 and 3, produces high penetration indeces in poorer council (group 7) and better council (group 8). These results would suggest that damage to vehicle is more likely to be linked with general levels of deprivation and social exclusion than the size of the local car population, particularly during the day. *Damage to dwelling* discrimination is better than the aggregate type, but it is not until 40% of the residential population have been taken into consideration that the gains come close to the 'best case' scenario. Interestingly, the ranking of penetration indeces for the damage to dwellings and vehicles is almost identical.

Violent crime was also discriminated fairly well as an aggregate group of offences. Disaggregating into assaults and offences which involved threats, abuse or intimidation appears to offer little benefit by this measure (Figure 6.46(c)). Even the 'best case' curves are almost identical. The group penetration index rankings for the disaggregate types are also identical.

To draw to an end this disaggregated gains analysis, Figure 6.46(d) shows an alternative approach for trying to improve the discrimination of *vehicle crime* (thefts of and thefts from) by re-examining the population. Although residential population is the denominator typically used to express *vehicle crime* rates this only really makes sense when the vehicles are parked at home - and that is only some of the time. Moreover, the assumption that vehicles are found in equal numbers per household is also rather dubious. There are similar problems with denominators for other crime types of course, but there is sufficient data available on vehicle populations to experiment cheaply with an alternative population model. What Figure 6.46(d) shows is the effect of using a daytime/weekday vehicle population as a denominator. This vehicle population has been described in detail in Chapter 5 but to recap is calculated by combining census data on car ownership with census data on journeys to work. Crime records also have to be sifted to remove offences not committed during daytime/weekday hours.

The 'best case' gains curve for the re-modelled vehicle crime is a little better than the 'best case' using a conventional residential population denominator. How much more of an improvement could be made with a more accurate vehicle model is unclear, but it would seem reasonable to expect some improvement if estimates of vehicle flows for retail, leisure and other users were available. Unfortunately, the re-modelling of the denominator appears to have made little improvement to the discrimination of vehicle crime by the task-specific classification. There is some shuffling in the PI ranks - with central deprived neighbourhoods (group 1) moving from first place to third, switching with ethnically mixed neighbourhoods (group 3).

6.3.5 Comparisons of Reductions in Variability of Crime Rates

While the gains charts give an indication of the ability of various classifications to explain crime they do so only at a between-group level. To better understand the performance of the classifications it is also necessary to consider within-group diversity and for this the Voas and Williamson (2001) tests can be employed again.

There is a problem, however, relating to the sampling of recorded crime at different geographies. Taking burglary dwelling as an example there were 13,833 offences recorded in 2003/04. Dividing this number into 477 LCCS neighbourhood zones produces a reasonably good distribution of values. Dividing the same offences by 2440 census output areas results in 6.3% of zones with zero values and a more skewed distribution than seen with the neighbourhood geography, but not so bad as to raise a lot of concern (although this is a subjective assessment). Dividing the 13,833 offences between 19,240 postcodes, however, results in 63.0% of zones being attributed a zero value and produces a very skewed distribution. Subjecting such a dataset to a set of calculations that measure difference between pairs of zones seems unwise. For this reason, the test of reduction in variability for the commercial consumer segmentation product has not been undertaken. The differences in the nature of the distribution of data values can be visualised using a histogram (Figure 6.46).

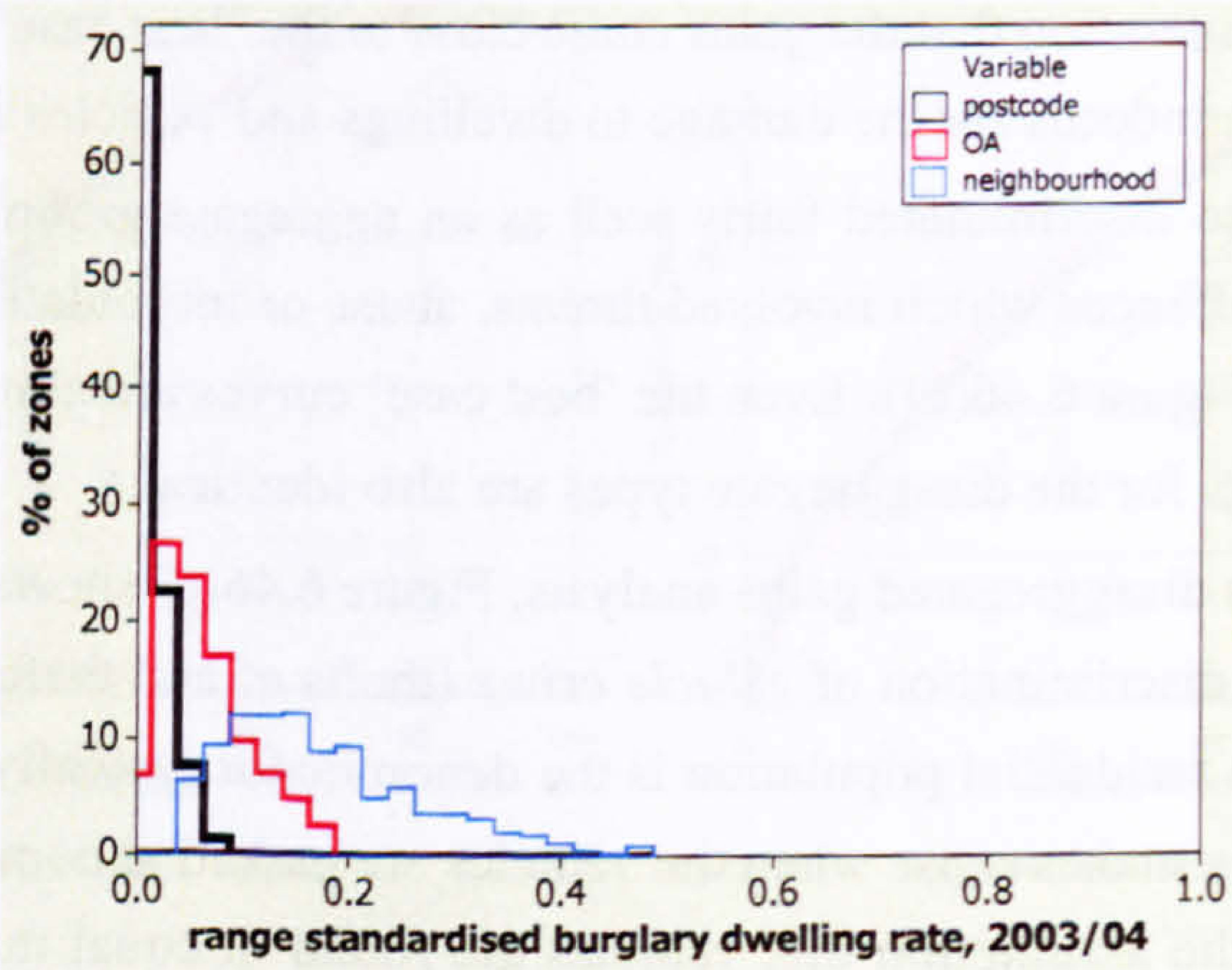


Figure 6.46: Histogram of burglary dwelling offences for different geographies.

The results of the tests are shown in Table 6.3. It needs to be remembered that within-group variability is likely to increase as the number of groups decreases. For this reason the 7 group partition of the ONS OA classification would not be expected to reduce variability to the same extent as the 10 group partitions.

Crime type	**LCCS** k=10	**IoD 2004** k=10	**Townsend** k=10	**OA Super** k=7	**OA Group** k=21
violent crime	40.6	40.2	33.7	15.3	17.3
criminal damage	39.7	42.1	31.9	15.3	18.1
primary arson	26.6	28.4	18.2	5.7	6.4
burglary dwelling	20.2	7.3	11.8	7.2	9.9
vehicle crime	17.7	8.1	9.9	6.7	8.0
other theft	10.9	4.7	5.3	3.7	4.7
burglary elsewhere	9.7	2.4	1.8	1.4	2.0
Overall reduction	**24.1**	**19.0**	**16.7**	**7.9**	**9.5**

Table 6.3: Percentage reductions in variability for different crime types for different classifications

The LCCS reduces variability most for *violent crime* (40.6%), *criminal damage* (39.7%) and *primary arson* (26.6%). These results are consistent with those from the gains analysis. Reductions in variability are less for the other crime types, with, on average, much of the variability seen across the district as a whole still present in the groups of the classification.

The results for the IoD 2004 deciles classification are slightly better for *criminal damage* and *primary arson* but overall the reduction in variability (19.0%) compares less favorably with that for the LCCS (24.1%). Part of the reason that the IoD classifcation does this well however (and this ought to be born in mind for the gains analysis results as well) is that one of the 7 domains that constitute the IoD is actually a crime domain, calculated from recorded crime data. Thus, to an extent, the fact that the IoD discrimnates between areas with different crime profiles is a self fulfilling prophecy.

The scale of the advantage can be determined by recalculating the IoD scores without the crime domain. To begin with, however, it is worth pointing out that each of the 7 domains in the IoD is weighted and that the crime domain weighting is only 9.3%. This is in common with the housing and environment domains; the highest weighted domains are income and employment, both weighted at 22.5%. Thus, the crime domain is not exerting a large influence on the IoD scores, so we wouldn't expect to see a very marked difference from an adjusted IoD deciles classification.

The overall IoD score is calculated by summing together exponentially transformed individual domain scores (X) weighted by the percentages mentioned above. The exponential transformation is of the form,

$$X = -23 * log(1 - R * (1 - exp(-100/23))) \tag{6.1}$$

where log denotes natural logarithm and exp the exponential or antilog transformation. R is the standardised domain score between 0 and 1, with 1 representing the most deprived neighbourhood (ODPM, 2004, page 151). If a domain is to be removed then the weights have to be adjusted proportionately to sum to 1.

The result of recalculating the IoD 2004 without the crime domain is an overall reduction drop from 19.0% to 17.8%. Violent crime drops to 39.3%, criminal damage drops to 39.4%, primary arson drops to 26.8%, burglary dwelling drops to 6.6%, vehicle crime drops to 6.2%, other theft drops to

4.0% and burglary elsewhere drops to 2.0%.

The overall reductions in variability achieved by the Townsend classification are smaller than either the LCCS or the IoD 2004 (in both its configurations). The overall reductions achieved by the ONS classification are noticeably less than the other segmentations of Leeds' population, but the different zone size makes the comparison unfair. It is clear, however, that the Group level classification, with its 21 groups, reduces the variability more than the Super-Group level, as would be expected.

The outcome of the tests for individual crime variables shows that some of the crime classes did benefit from disaggregation, as predicted (Table 6.4). For *burglary elsewhere* the disaggregation improved the reduction in variability. The fact that the results for both the disaggregated types (11.7% and 12.9%) were better than the original reduction (9.7%) demostrates the importance of recognising the different geographies embedded within crime counting rules. The Townsend classification also reduced the variability for burglaries from sheds and garages, but unlike the LCCS parition, did not improve on reducing the variability of burlaries against non-residential property.

	reduction in variability (%)	
variable	LCCS	Townsend
criminal damage to a dwelling	46.8	37.5
criminal damage - others	42.0	31.6
violent crime - assaults	38.3	31.9
other theft - in a dwelling	33.6	29.8
violent crime - threats, abuse	33.0	25.2
criminal damage - vehicle	23.2	22.0
violent crime - other	17.4	16.1
other theft - others	14.8	10.0
other theft - from a person	14.1	9.2
burglary elsewhere - others	12.9	1.5
burglary elsewhere - sheds, etc.	11.7	9.2
other theft - not classified elsewhere	7.1	0.5
other theft - vehicle	7.8	5.2
other theft - shoplifting	5.1	1.3

Table 6.4: Reduction in variability of disaggregated crime variables from LCCS and Townsend classifications

The results of disaggregating *other theft* were mixed, with a much better reduction in the variability of *other theft in a dwelling* but a worse result for *thefts from a shop*. The reduction of variability of *criminal damage* offences was better for dwelling and other targets, but worse when the target was a vehicle. The impact of disaggregating *violent crime* offences produced an opposite effect to that seen for *burglary elsewhere* in that the reductions in variability of individual parts were all less than for the aggregate whole! Nevertheless, it is useful to see that the classifications agree that they work best when trying to explain levels of physical violence. To close, it is worth restating that the disaggregation of *burglary elsewhere* and *violent crime* provide evidence that aggregation can have positive *and* negative effects.

6.4 Correlation Between Crime and Membership Strength

The fuzzy membership aspect of the LCCS design can be explored further with respect to the dependent crime variables. The analysis in the previous section has shown that there is within-group diversity in crime rates, varying by crime type, and tests can be undertaken to see if the scale of this diversity is in some way related to a neighbourhood's strength of membership to its group. If this were the case then it would be useful support for the decision to create a fuzzy classification. Furthermore, the effect could be mathematically controlled for when analysing within-group variability of crime and make the task of identifying truly atypical neighbourhoods that much more reliable.

To test whether there is a linear relationship between the group membership strength and variation from an average within-group crime rate a Spearman's rank correlation coefficient can be calculated. As with the previous tests, the data do not conform to normality so a parametric test, such as Pearson's Product Moment, is not appropriate. Although a number of non-parametric correlation tests exist Spearman's rank correlation is easy to compute and is frequently used by geographers (Shaw and Wheeler, 1994). Each pair of observations consists of two ranks. If the sequence of the two paired ranks is equal then the variables are perfectly positively correlated. The actual correlation coefficient is a measure of the degree to which the ranks differ for each observation and is expressed as:

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} \tag{6.2}$$

where D is the differences between ranks of corresponding values of X and Y, and N is the number of pairs of X and Y values (Shaw and Wheeler, 1994, page 186).

The Spearman's rank correlation coefficient where X is the membership strength coefficient and Y is the within-group standardised burglary dwelling rate (2003/04), for example, is 0.091, and is significant at the $\alpha = 0.05$ level. Thus, there is very little evidence of a linear relationship between the group membership strength and the extent to which the burglary dwelling rate varies from the within-group mean.

A similar very weak positive correlation exists for the burglary of sheds and garages (0.093, $\alpha = 0.05$), while there are very weak negative correlations for vehicle crime (-0.093), violent crime (-0.098) and other theft (-0.110), all where $\alpha = 0.05$.

6.5 Neighbourhood Isolation

Another criticism that can be leveled against geodemographic classifications is that the cluster analysis techniques they usually employ are inherently aspatial in nature (Harris et al., 2005, page 219). The classifications are geographical in the sense that they are based on data from different spatial zones and when mapped the classification usually displays some sort of non-random pattern among the zones. The classifications are also geographical in the sense that they are frequently used to analyse other geographical information, for example crime. Yet the data itself does not usually contain any information pertaining to its place in space and the cluster analysis is thus bereft of geographical input.

Harris et al. demonstrate the problem using a simple example (Harris et al., 2005, page 221) for which the data is reproduced below (Table 6.5). In addition, the spatial location of the observations is plotted (in a slightly amended form) in Figure 6.47.

The first solution groups the observations solely by their colour attribute and thus observation 2 is grouped along with observations 1,3,4 and 7. The second solution, however, also considers the geography of the observations as plotted in Figure 6.47. Harris et al (2005) suggest that for this example, considering the bigger picture, the difference between dark grey and white (group 1) or

ID	x	y	attribute	solution 1	solution 2
1	1	2	white	A	A
2	4.5	8	light grey	A	B
3	1.5	1.8	white	A	A
4	2	2.8	white	A	A
5	1.5	2.5	dark grey	B	A
6	4.5	8.8	dark grey	B	B
7	2	2	light grey	A	A
8	4.5	7.4	black	B	B
9	4	8	black	B	B

Table 6.5: Results of grouping observation - non-geographical (solution 1) and geographical (solution 2).

between light grey and black (group 2) is not always that great when considered in geographical context.

The point to this example is that it might sometimes be preferable to consider the attributes of a particular neighbourhood in the context of the neighbourhoods around it. Harris et al (2005) suggest that it might be preferable to include this type of contextual information as a variable in a geodemographic classification, and since 2003, Experian's MOSAIC classification has included such contextual variables to identify the isolation of social class and occupation groups.

To be clear, variables such as this are missing from the LCCS because by the time the idea came to the authors attention the die had been cast and the classification already produced. Unwittingly however, one contextual variable that did make it into the classification was proximity to public houses. As has been described previously, this variable was constructed by sampling a density surface over the public house point locations. In this way a neighbourhood may have a value that is high despite having no public houses within its own boundary - the value being 'borrowed' from the high density of public houses in adjacent or nearby neighbourhoods, according to the bandwidth of the density scanning window.

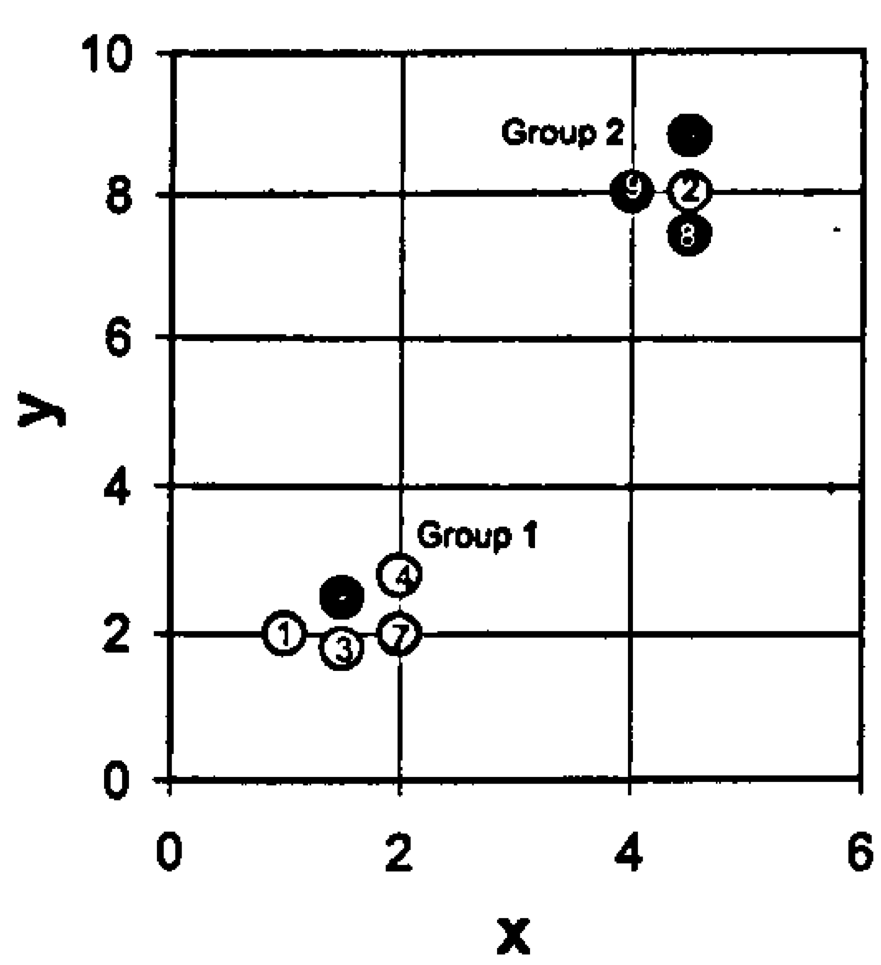


Figure 6.47: Geographical grouping of the data shown in Table 6.5.

6.5.1 Crime Contextualised Neighbourhood Isolation

All these discussions about contextualised variables would be more persuasive if some hard evidence could be found to confirm that being isolated in some way from group peers had a measurable effect on crime rates. That is, where there are cases where the patchwork effect of the group geography results in relatively affluent neighbourhoods being surrounded by poorer and higher crime neighbourhoods, and does this result in a higher than usual crime rate for the isolated neighbourhood? In the converse situation, when an isolated poorer neighbourhood has a lower crime rate than its group mean, is this due to the relatively low offending density in surrounding, more affluent, neighbourhoods?

To analyse this, each neighbourhood can be assigned the mean burglary rate for its group and this figure compared with the mean burglary rate for all of the adjacent neighbourhoods. Thus, in Figure 6.48, the cell at the centre has a burglary rate of 33.0 which can be compared with the average rate of all the adjacent cells, which is 18.75, a difference of 14.25.

Figure 6.48: Measuring relative isolation in crime rate terms.

In this case, the H_0 for the test is that there is no relationship between an atypical within-group burglary rate and the prevailing burglary rate in adjacent neighbourhoods (using data from 2003/04). Using all 477 observations, the Spearman's rank correlation coefficient is -0.274 and is significant at $\alpha = 0.01$. Thus we reject the null hypothesis and deduce that there is a weak negative relationship between a neighbourhood's within-group standardised burglary rate and the types of neighbourhood adjacent to it. The negative correlation suggests that usually low burglary neighbourhoods surrounded by usually high burglary neighbourhoods will have higher burglary rates than average for their group, while neighbourhoods with usually high burglary surrounded by neighbourhoods with usually low burglary will have lower burglary rates than average for their group.

If the sample is reduced in size to just include those neighbourhoods that are isolated from their peers by a minimum distance of 1.0km then the relationship produces a significant ($\alpha = 0.01$) correlation coefficient of -0.326 (n=170). If the minimum distance is increased to 1.5km then the strength of the relationship becomes stronger, with a significant ($\alpha = 0.01$) correlation coefficient of -0.478 (n=79). While reducing the sample to only include those neighbourhoods at least 2.0km from their nearest peer produces a significant ($\alpha = 0.01$) correlation coefficient of -0.466 (n=39).

The results of tests for other crime types are shown in Table 6.6. There are a few more statistically significant relationships that was the case when considering Euclidean isolation but for the most part relationships are very weak. Only primary arson produces a relationship approaching that for burglary dwelling, suggesting that for both these crime types the crime context in adjacent neighbourhoods does have some measurable bearing on crime rates for a neighbourhood isolated from its peers.

6.6 Concluding Remarks

The pen portraits of the different clusters help to reveal what it is that makes each group unique. Some groups are more unique than others and characteristically it is those neighbourhoods closest to the Leeds average that are among the most difficult to partition into distinct groups. Nevertheless, the pen portraits fulfill an important function especially when evidence from other statistical tests of group 'strength' warns caution. To illustrate this point consider group 1, which groups together neighbourhoods that are close by the city and includes some of Leeds' most deprived areas. The pattern of the extent to which the average values for individual variables deviate from the Leeds norm looks different to other groups, as does the crime profile. Yet the silhouette width (0.03) presented in the previous chapter suggests that on average these neighbourhoods would be as well classified if they were in their next nearest group. In multi-variate space this must be the case, but consideration of individual variables does provide evidence of group uniqueness. It is very difficult not to subjectively attach greater importance to some variables over others but then when the geography of group membership also appears unique and 'to make sense' the reliance of purely objective and (usually) mathematical metrics of partition fitness appears inadequate.

Comparing the LCCS with other classifications does appear to support the argument for a task-specific classification over a general purpose tool. In most tests it did as well as, or better than, its rivals. The margin by which it performed better however, was not as great as had been hoped at the outset. Clearly all the classifications are struggling to discriminate some types of crime and in these cases it may not be helpful to adopt a neighbourhood classification approach to crime pattern analysis. Neither is it certain whether modification to the design of the LCCS would improve the discrimination and/or reduction in variability. One possible line of enquiry that could be researched further would be to reduce the number of variables used in the cluster analysis (to reduce computation time) and use a stepwise approach to variable inclusion/exclusion and size of k to build up a set of partitions that reflected all the permutations and combinations of variables, and possibly some none-zero weights. Reduction in variability of crime rates by each partition could then be compared using the Voas and Williamson test and the areas between the curve and the diagonal in the gains analysis also calculated and compared.

Crime type	**all neighbourhoods**	**>1.0km**	**>1.5km**	**>2.0km**
Burglary dwelling	-0.274 **	-0.326 **	-0.478 **	-0.466 **
Burglary elsewhere	-0.107 *	ns	ns	ns
Sheds, etc.	+0.097 *	ns	ns	ns
Other	ns	-0.156 *	ns	ns
Vehicle crime	-0.194 **	ns	ns	ns
Violent crime	-0.170 **	ns	ns	ns
Criminal damage	-0.193 **	-0.154 *	ns	ns
Other theft	-0.108 **	ns	ns	ns
Arson (primary)	-0.188 **	-0.286 **	-0.311 **	ns

ns - not significant, ** $alpha = 0.01$, * $alpha = 0.05$

Table 6.6: Spearman Rank correlations between crime types and level of crime-context isolation

It is uncertain, however, if the computational expense of such an approach could be justified. Furthermore, the pressure to accept an optimal solution produced by such a method would be great, even if subsequent subjective appraisals of the final partition revealed problems. Such subjective appraisals of the classification in use are the focus of the remaining chapters. Reflecting again on Openshaw's words it is likely that thus far the research has been interpreting the observation that "a classification can only be deemed 'good' or 'poor' when it has been evaluated in terms of the specific purpose for which it is required" (Openshaw, 1983, page 245) to mean 'evaluate quantitatively', led on by the availability of large crime data sets which lend themselves to such an approach. The research has begun to develop by empirical study an understanding of both the strengths and weaknesses of the LCCS, but this understanding now needs to be reinforced by applying the LCCS to real crime pattern analysis problems.

Chapter 7

Exploratory Analysis of Criminal Damage

7.1 Introduction

With the classification constructed and portraits of the groups developed, the Leeds Classification for Community Safety (LCCS) can begin to be used as a lens through which to examine and analyse patterns of crime. From the outset, the hypothesis has been that neighbourhoods of the same type are likely to have similar crime rates and the previous chapter used mathematical tests to demonstrate the discriminating power of the classification for different crime types. By contrast, this chapter and the next test the usefulness of the classification in more practical terms, using recorded crime and interventions data to analyse crime patterns in Leeds. Particular attention is given to within-group variations in crime rates and the identification of neighbourhoods with unusually high or low crime rates compared to their group-peers.

Criminal damage is the crime type analysed here, partly because it accounts for a large proportion of offences in Leeds, but also because it may involve a range of offender motivations and environmental factors, and because politically it is becoming more prominent in debates around anti-social behaviour, such as the *Respect Agenda* (Home Office, 2006). Moreover, the positive results from the reduction of variability tests (Section 6.3.5) also suggest criminal damage may be explained quite well by the classification.

The earlier review of conventional hotspot mapping techniques demonstrated how crime patterns could be identified (Section 3.2). Here, an alternative methodology based upon exploratory data analysis (EDA) principles, and employing the neighbourhood classification, is demonstrated and discussed. This EDA approach to analysis is applied to four years of criminal damage data to identify neighbourhoods that stand out as being atypical. These neighbourhoods are then subjected to a secondary stage of analysis to look for possible explanations as to why the crime patterns stand out from those of their group peers. This secondary stage includes analysis of the group variable profiles, cluster membership strength, more detailed crime profiles and neighbourhood location and geography. Where plausible explanations or hypotheses can be identified, the implications of these for community safety policy, and/or the design of the classification, are discussed.

7.2 Crime Recording Issues

While reducing criminal damage has not been a specific priority either nationally or locally during recent years, these offences accounted for 19.5% of all recorded crime in Leeds and 20.3% across England and Wales in 2003/04. Furthermore, the British Crime Survey shows that this type of offence is only reported in 31% of cases recounted by survey respondents, compared to an average under-reporting of all crimes of 42% (Dodd et al., 2004).

The term criminal damage includes a range of offences including *Damage to a Motor Vehicle* (typically 40% of all criminal damage offences), *Damage to Dwelling* (34%), *Damage to Building Other Than a Dwelling* (11%), *Damage to Property - Not Specified* (10%), and *Arson*. The remainder of criminal damage only accounts for a small number of offences each year but include classes specifically designed to capture racially or religously aggrevated incidents, for example. For the analysis in this chapter, criminal damage is disaggregated into damage to dwellings and damage to to vehicles, partly because they represent the majority of offences, but also because content analysis of modus operandi text reveals that there is less ambiguity about these types of damage than others, particularly *Damage to Property - Not Specified.*

When looking at trends over time, it is important to consider that in England and Wales during 2002/03, the National Crime Recording Standard was estimated to have had the effect of raising recording of criminal damage offences by 9% over the previous year. Moreover, within West Yorkshire the Home Office estimated the impact had been much higher, at 25%. It is important to consider these effects when interpreting trends over the NCRS introduction period (NCRS began to be used across West Yorkshire in April 2002). Figure 7.1 has not been adjusted for the NCRS, so the slight rise, both in the national and local figures, between 2001/02 and 2002/03 is misleading when looking at the trend over all four years. If the NCRS had not been implemented, there would have been a sharp fall in the published criminal damage rates in 2002/03, followed by a slight rise in 2003/04. For the statutory crime audit conducted by the Leeds Community Safety Partnership in 2004, the decision was taken that given the NCRS problems, change over time would only be analysed between 2002/03 and 2003/04. The rationale for this decision was that the major impact of the NCRS ought to have worked itself out by this time, although there is evidence that for some crime types (especially violent crime) the NCRS effect is still being felt (Dodd et al., 2004).

For the current research, a different approach has been adopted, measuring change between 2000/01 and 2003/04. This has been done to reduce the problems of basing percentage change calculations on small numbers. The assumption, and it is one that could be challenged, is that the effect of the NCRS has been felt uniformly across all neighbourhoods in Leeds, thus the relative difference in levels of change between neighbourhoods ought to be the same with or without any correction for NCRS effects.

7.3 Exploratory Data Analysis

7.3.1 Introduction

Within that part of the human geographical discipline that has traditionally developed and applied quantitative methods, much of the focus of data analysis is confirmatory in nature. The aims are, among others, to specify models, estimate parameters and statistically test hypotheses. However, when, during the 1980s, positivist attitudes and quantitative geography came under uncreasing criticism from advocates of humanistic and behavioural geography, the case for using statistical techniques in a more exploratory manner started to gain currency. As with other social sciences, human geography began to pay more attention to the possibilities of exploratory data analysis (EDA) (Erickson and Nosanchuk, 1992; Sibley, 1990) following the innovative work of statistician John Tukey (Tukey, 1977). Indeed some social scientists began to argue that more attention should be paid to exploratory techniques rather than to continue to uncritically subject poorly understood data to complex forms of analysis (Besag and McNeil, 1976).

What EDA emphasises is the importance of examining available data using simple numerical description, such as measures of dispersion and centrality, in conjunction with imaginative graphical representation. Furthermore, the emphasis is usually not to try and prove hypotheses, but to formulate new hypotheses and generate new ideas for further investigation (Cox and Jones, 1981). Sibley's argument in favour of an EDA approach is that, "it encourages and facilitates repeated reference to the data and a cautious, sceptical attitude to theory ... the former is emphasised in the display of data in several alternative forms and the latter by focusing on anomalous or problematic cases" (1990, page 4). Importantly, focusing on cases shown to be residuals and outliers is central to the research being presented here.

In Section 3.2.2, some of the ways in which kernel density maps can be used to explore crime incident data were demonstrated, but by way of contrast, what follows is the demonstration of a suite of exploratory data analysis techniques that can be used in conjunction with the neighbourhood classification in order to analyse crime patterns. Some of these techniques are well documented in the EDA literature, others may be more unique but can be shown, nonetheless, to be very much in the spirit of exploratory data analysis in human geography.

7.3.2 Between-Group Comparisons of Criminal Damage

One of the simplest measures of central tendency, the mean, can be used to compare the rates and annual change in recorded crime between the different groups. The bar charts of damage to dwellings 7.2(a) and damage motor vehicles 7.2(b) show how the mean within-group crime offence rates have varied over the previous four recording years and also indicate the average scale of criminal damage problems in neighbourhoods of different types.

In scale terms, damage to dwellings is shown to be most incident in neighbourhoods belonging

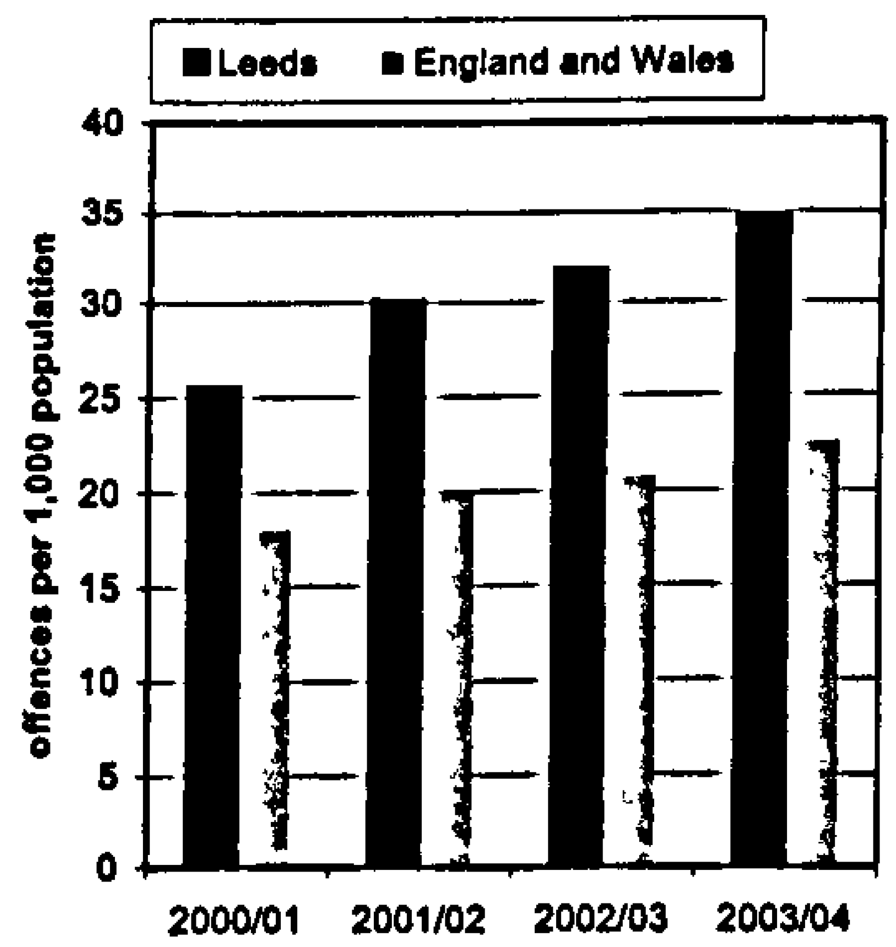
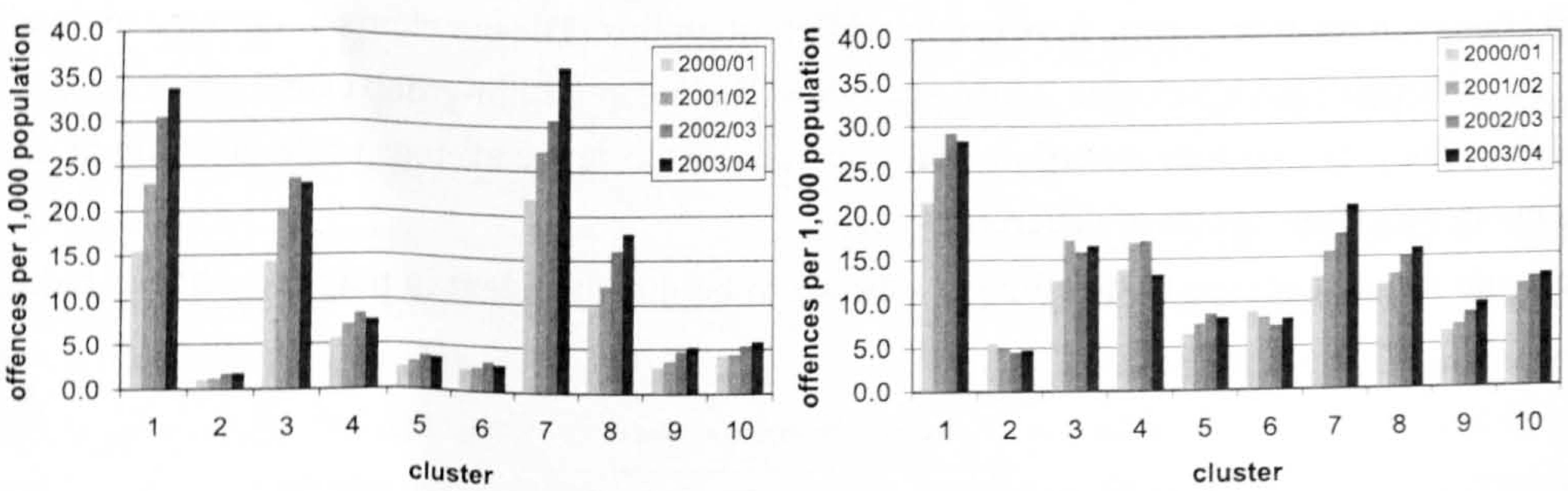


Figure 7.1: Local and national trends in Criminal Damage rates, 2000/01 to 2003/04.

to groups 1 and 3 and group 7, that is, inner-city, zone of transition neighbourhoods, including neighbourhoods with ethnic populations, and the poorer council housing estates. The less poor council housing estates in group 8 have more modest rates and most other groups have very low rates, particularly the northern rural and hinterland neighbourhoods (group 2). Among the variables that most define the profiles of the problematic groups are a number that might be pertinent to criminal damage problems, including higher levels of void properties (VO01), poorer school attainment (ED05) and higher levels of minor incivilities.

The relative differences in the scale of the damage to vehicle problem are shown to be more muted, with the classification discriminating less well (shown previously, in Chapter 7). Again, the inner-city neighbourhoods have the highest rates. The student neighbourhoods (group 4) show more prominently, perhaps associated with the day-time street-side car population that ebbs and flows daily during university term-time. The working hours car population variable (CA01) is not a major part of the group 4 profile but it does have the second highest positive z-score of all the groups, with only group 1 neighbourhoods having a higher value.

With respect to change over time, the general trend for both types of damage is an upward one, more so for damage to dwellings. Furthermore, it appears that for the most part, growth in the last period was less than seen in previous years. At this stage, any group that does not conform to these trends is noteworthy, as are groups which show different trends for the two types of offence. On this basis, it is perhaps firstly worth observing that in group 7, rates for both types of criminal damage rose more than elsewhere in 2003/04. Secondly, groups 1 and 8 have similar rates for the types of criminal damage, while groups 10, 9, 6, 5, 4 and 2 show more damage done to vehicles than dwellings. Thirdly, groups 7 and 3 show more damage done to dwellings than vehicles. Fourthly, what falls there have been in the cluster means are fairly modest. And fifthly, the most obvious fall - in damage to vehicles in cluster 4 between 2002/03 and 2003/04 - is not uniform within the group. The raw data reveals some group 4 neighbourhoods have seen a considerable fall, while a similar number have seen a rise, by similar orders of magnitude.



(a) Damage to Dwelling (b) Damage to Motor Vehicle

Figure 7.2: Trends in criminal damage rates, 2000/01 to 2003/04, by cluster.

7.3.3 Within-Group Comparisons of Criminal Damage

While some of the between-group differences discussed above are interesting, it may be that that a more useful application of the LCCS is to examine *within-group* differences in crime and disorder. It is argued that a shortcoming of the between-group analysis can be that it often produces findings that come of little surprise to community safety practitioners, and thus is of marginal extra benefit when looking for evidence with which to inform policy. In the practical day-to-day work of a typical partnership, more debate and interest may be stimulated by being able to demonstrate that some locations, though known to have problems, actually have very severe crime problems, and conversely, other locations thought to be problematic - because of other indicators of strain or their proximity to such places - actually have relatively few crime problems.

Factors that might affect why neighbourhoods have crime rates appreciably above or below the mean for their group could include,

1. the existence of a unique neighbourhood feature not found in the other neighbourhoods within the group. This feature may or may not be captured by the variables used to construct the classification;
2. a large 'distance' between the neighbourhood and the group mean, suggesting an alternative classification might be necessary;
3. proximity to other neighbourhoods from, or to, which crime is being displaced;
4. a crime prevention/reduction initiative that has not been delivered uniformly across all the neighbourhoods within the group; and,
5. chance.

No attempt will be made here to explore this last possibility, although it is deserving of attention. Instead, three techniques borrowed and adapted from the EDA discourse are employed to try and explore the first three issues. The influence of different geographies of crime prevention initiatives are considered in Chapter 9.

Boxplots

The first of the EDA techniques to be demonstrated is the boxplot, or box-and-whisker plot as it is sometimes known. This is a diagrammatic device that provides a summary of spread and asymmetry within a set of data. Referring to the example shown in Figure 7.3, the vertical line inside the box represents the median. Spread is indicated by the length of the box, with the beginning and end of the box representing the first and third quartiles respectively. Viewed together, the position of the median gives some indication of the amount of skew in the distribution. The whiskers emanating in either direction from the ends of the box are referred to as the upper and lower outside cut-offs, or the upper and lower extremes. The lower cut-off is usually defined as $Q1 - 1.5(Q3 - Q1)$, and the upper cut-off as $Q3 + 1.5(Q3 - Q1)$. Data points with values beyond these cut-offs, usually termed *outliers*, are represented with a symbol, usually a circle or star. However, in some software implementations of the boxplot, for example SPSS, outliers where the data value is greater than 3 times the Inter Quartile Range from Q1 or Q3 are referred to as *extremes*.

Boxplots can become especially useful when comparing a number of related samples, and this is the way in which they have been employed here. Sibley (1990) describes ways in which comparisons of spread might be made more revealing if the samples are standardised by first having a level (the median) removed and then having their spread standardised. Yet, for this example, it is suggested that it may be as useful to see the boxplots for the crime rates in different groups plotted using raw

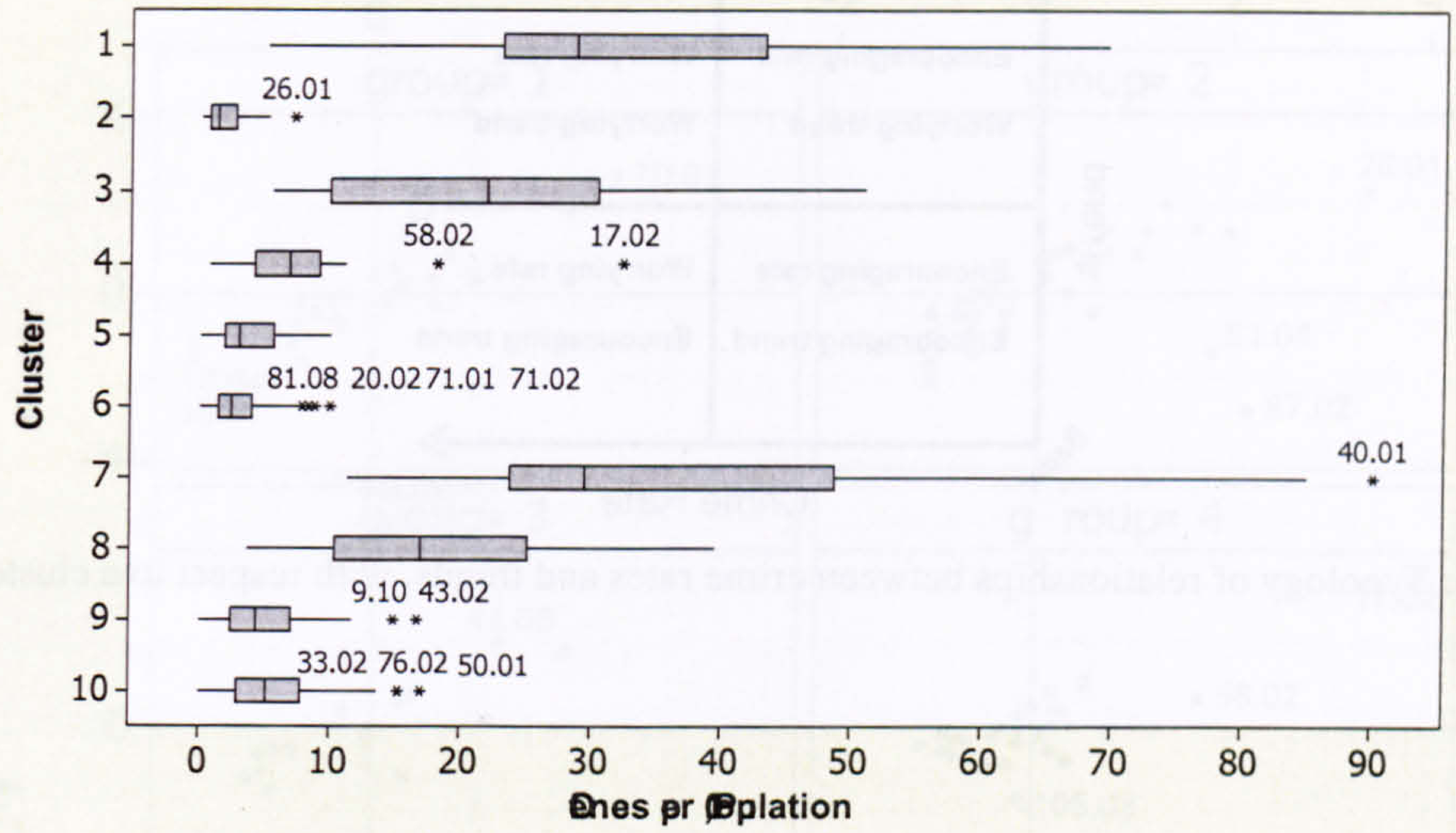


Figure 7.4: Boxplot of Damage to Dwelling rates, by cluster (outlier neighbourhoods labeled).

In this example, the data has been standardised by transforming values into z-scores. This can make it easier to see the extent to which a neighbourhood may be an outlier within the population as a whole, and not just within its own group. Figure 7.7 shows the resulting scatter plots for damage to dwellings, while Figure 7.8 shows the results for damage to motor vehicles. In both cases, the quadrants formed by the intersection of the axes are equivalent to the quadrants of the typology in Figure 7.6.

If the classification was discriminating ideally, all the points would be tightly bunched around the zero intersection of the x-axis (same crime rate) and a similar (but not necessarily zero) point on the y-axis (same amount of change in rate over time). This type of pattern would indicate that the classification was minimising the within-cluster variance in the crime rate while at the same time

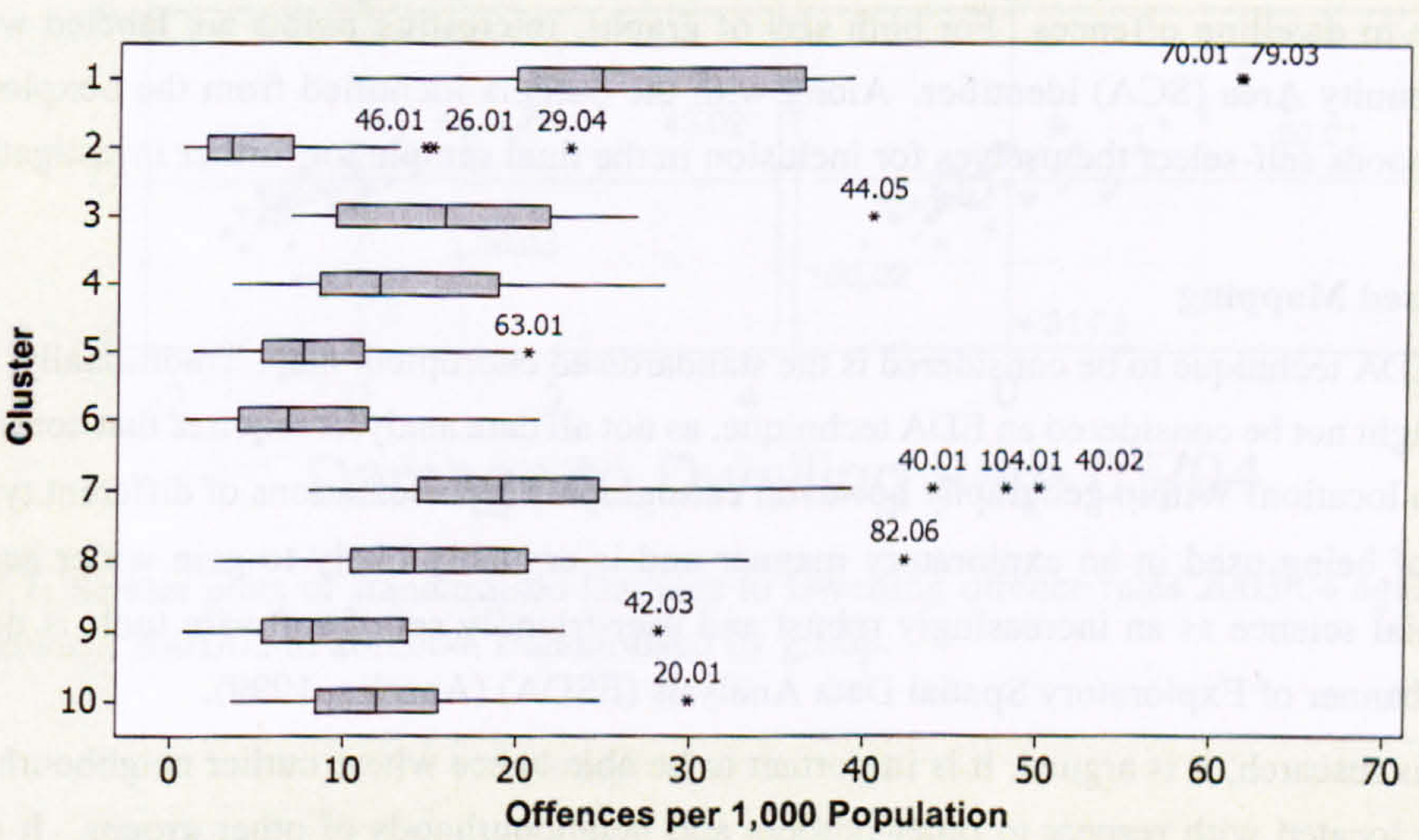


Figure 7.5: Boxplot of Damage to Motor Vehicle rates, by cluster (outlier neighbourhoods labeled).

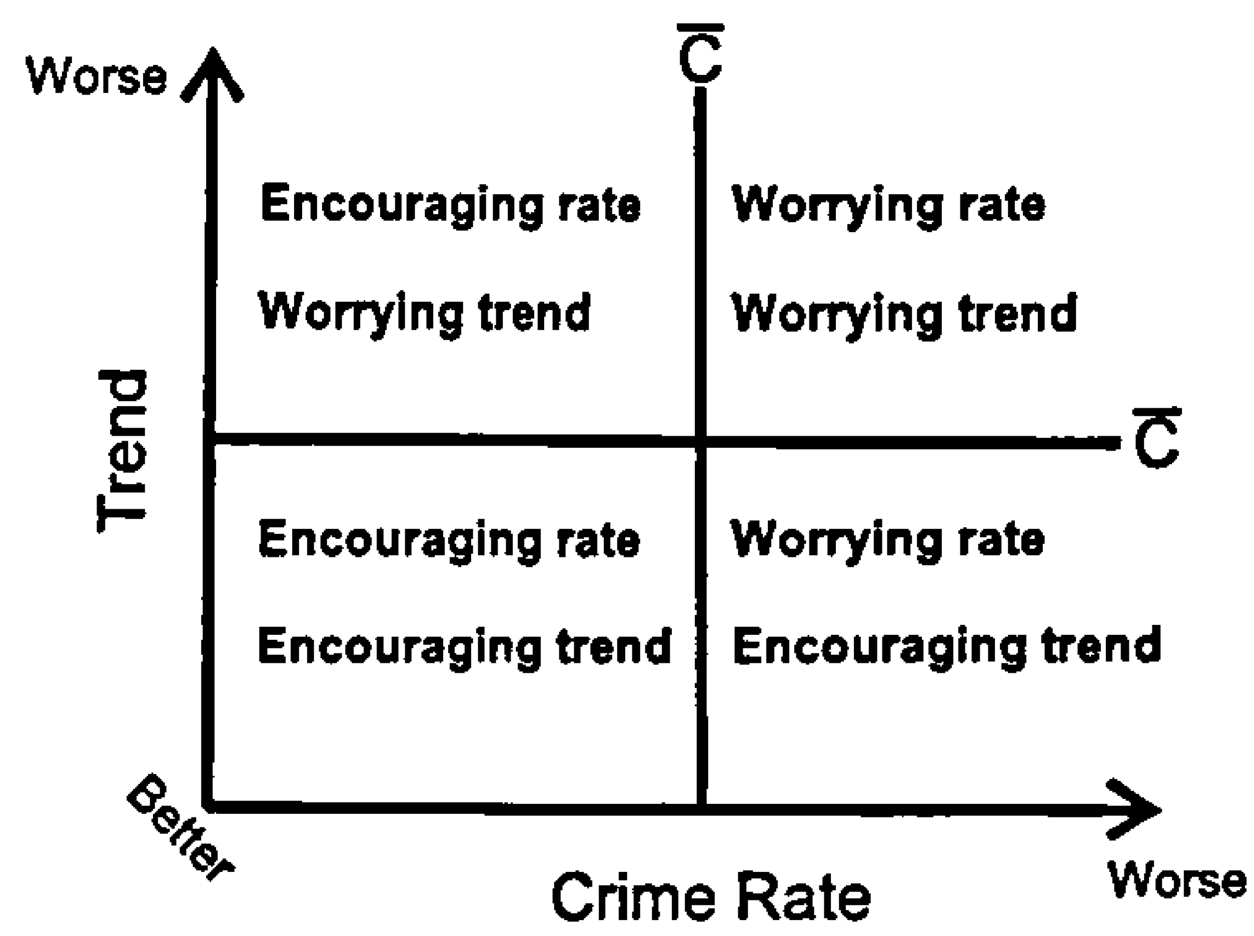


Figure 7.6: Typology of relationships between crime rates and trends, with respect to a cluster mean ($\bar{C}$).

indicating that neighbourhoods within a given group were also exhibiting similar trends, that is, their crime rates were generally all declining in recent years or all increasing in recent years. Such a situation might arise, for example, where the majority of neighbourhoods in a poorer group had been in receipt of substantial regeneration funding and benefited from the improvement programmes that followed. In these examples, the closest to this ideal pattern is found in the graph of damage to motor vehicles in group 2 - which Figure 7.5 confirms as having a compact form with respect to this particular crime rate.

In general terms, the scatter plots for damage to dwellings (Figure 7.7) show a fairly consistent pattern of neighbourhoods with high rates also having the worst trends, and neighbourhoods with low rates having the best trends. Where this is not the case, outliers tend to be in the 'worrying rate, encouraging trend' category. Examples of neighbourhoods with relatively low rates but which have still seen an appreciable rise over the last four years, are rare.

The scatter plots of damage to motor vehicles (Figure 7.8) are similar in nature to those for damage to dwellings, but there are more outliers outside the two principal quadrants and the points are less densely packed. Also, the range of standardised scores for the change in rates is less than for damage to dwelling offences. For both sets of graphs, interesting points are labeled with their Sub-Community Area (SCA) identifier. Along with the outliers identified from the boxplots, these neighbourhoods self-select themselves for inclusion in the final sample for further investigation.

Standardised Mapping

The final EDA technique to be considered is the standardised choropleth map. Traditionally, perhaps, the map might not be considered an EDA technique, as not all data analysis requires that consideration be given to location. Within geography however, cartographic representations of different types have a history of being used in an exploratory manner and is arguably likely to gain wider acceptance within social science as an increasingly robust and user-friendly set of software tools is developed under the banner of Exploratory Spatial Data Analysis (ESDA) (Anselin, 1999).

For this research, it is argued, it is important to be able to see where outlier neighbourhoods are physically located with respect to other outliers and neighbourhoods of other groups. It might be the case, for example, that a high-crime outlier neighbourhood surrounded by others with average crime values is identifying a neighbourhood that would be better off placed in another group where

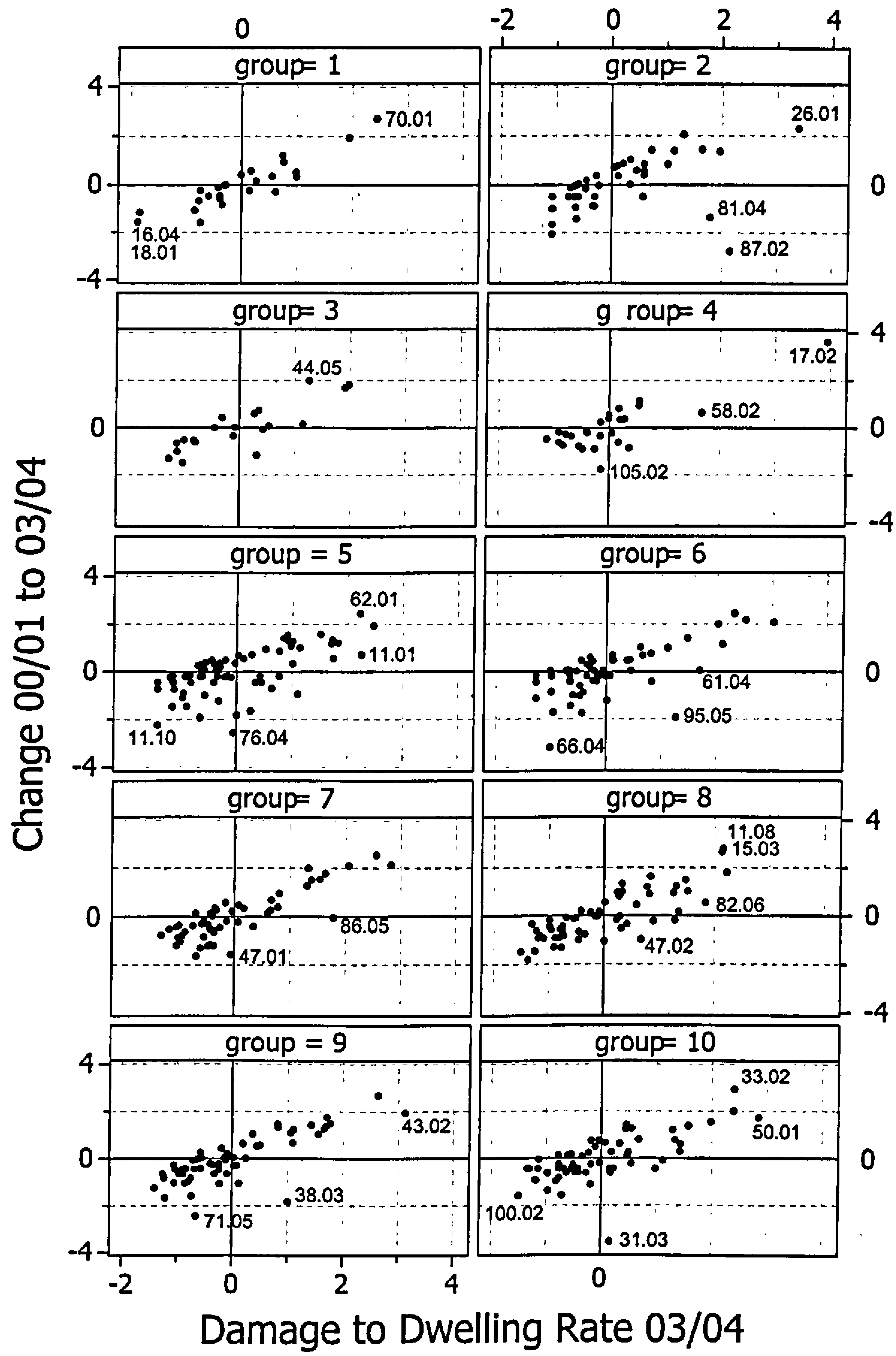


Figure 7.7: Scatter plots of standardised Damage to Dwelling offence rates 2003/04 against change in rate between 2001/02 to 2003/04, standardised by group.

its crime rate would no longer attract outlier status. Alternatively, an island outlier may have a high crime rate because crime problems from surrounding higher-crime neighbourhoods are diffusing in.

Identifying outlier locations can also be a useful check of classification integrity, affording the

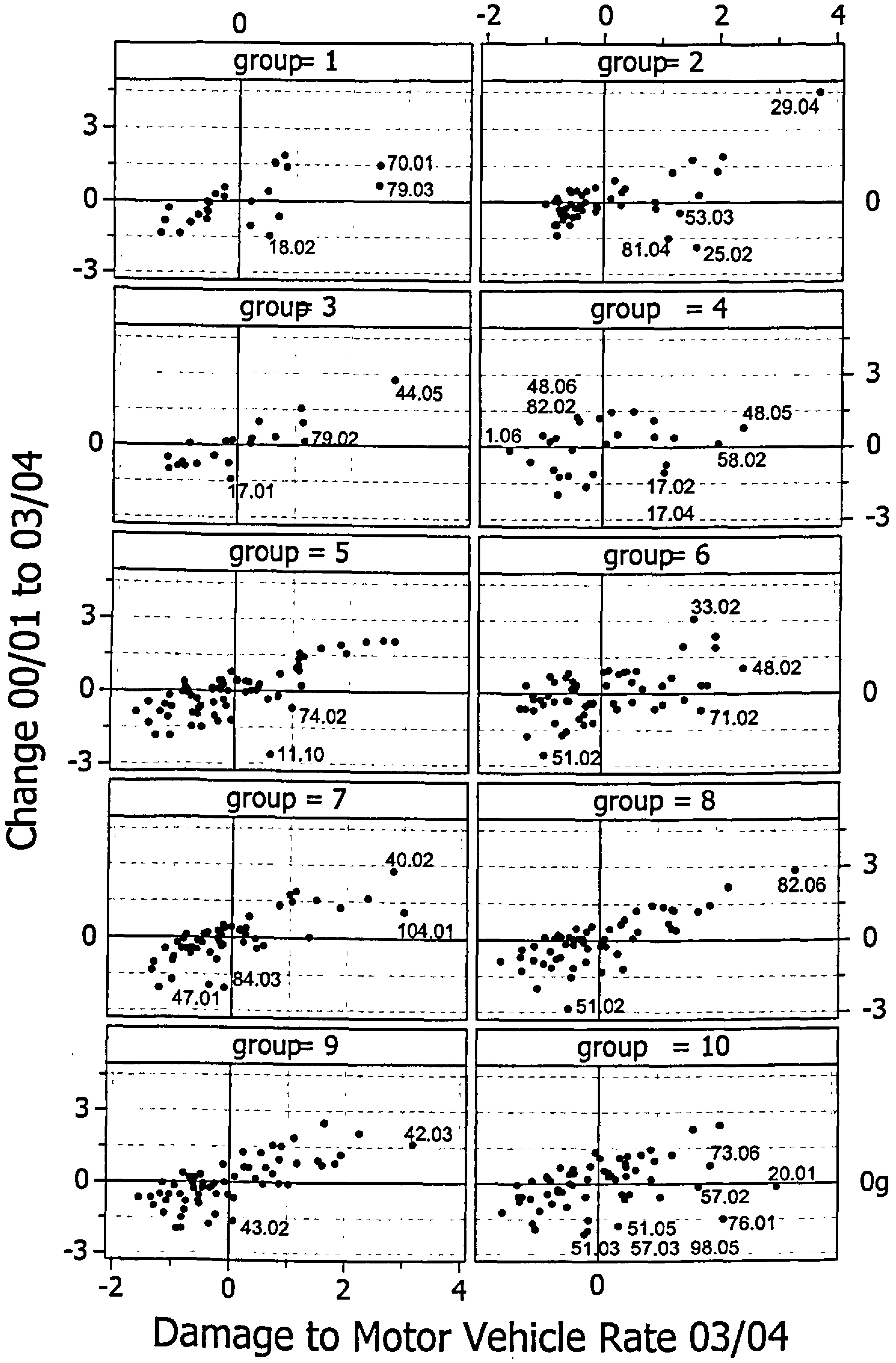


Figure 7.8: Scatter plots of standardised Damage to Motor Vehicle offence rates 2003/04 against change in rate between 2001/02 to 2003/04, standardised by group.

analyst a chance to detect neighbourhoods where the variables used for the classification fail to capture important local features and geographical circumstances. For example, neighbourhoods with large daytime populations but small residential populations, such as those that besides housing also contain retail parks or industrial estates, may suffer from inflated crime rates because of their smaller residential denominator population.

There are various ways in which crime rates within groups can be standardised prior to being mapped alongside one another. Using z-scores, (Craglia et al., 2000, following) has the advantage of highlighting only those areas that differ substantially from the mean but this may lead to bias toward neighbourhoods in poorly defined groups, as more variance in crime rates is likely therein. Another alternative is to range standardise each group to the same minimum and maximum values. In this case, every group has its highest and lowest crime rates represented by standardised scores of the same order of magnitude, and the potential for bias due to poor group definition is reduced. The disadvantage to this approach is that the map produces a rather broader focus than may be desired. On consideration, range standardisation is adopted here instead of a standard deviation based technique, principally because it complements the previous self-selection techniques which have the tendency to identify high extreme neighbourhood crime rates but not lower than expected neighbourhood crime rates.

To begin with, the range standardised crime rates per 1,000 population in 2003/04 are mapped. Figure 7.9 shows the geography for damage to a dwelling and Figure 7.10 shows the geography for damage to motor vehicles. Neighbourhoods whose high or low rates look particularly interesting, given their location, are identified and selected for the sample of neighbourhoods to be investigated in more detail. The same exercise is then repeated with range standardised maps of change over time (Figures 7.11 and 7.12).

A typical literal reading of the maps may go as follows:

- The neighbourhoods identified at *A* in Figure 7.9 stand out as an isolated cluster of high values in what is known from the classification geography to be a more affluent sector of the district.
- The neighbourhoods at location *B* appear interesting because high and low values appear spatially adjacent and are known to belong to the same group.
- The neighbourhood at *C* stands out because it represents the highest value in its group and because a high crime rate is at odds with the map reader's experience of the area.
- Meanwhile in Figure 7.10, the neighbourhoods at *D* are selected because this area is known to be an industrial zone with very little in the way of a residential population.
- The neighbourhood at *E* is singled out because a high crime rate is not something normally associated with this part of the district and it occurs in a group (10) of particular interest for its non-residential land use characteristics.
- The neighbourhood at *F* is selected because problems in rural areas are less common.
- Looking at change in damage to dwellings (Figure 7.11), neighbourhood *G* is selected because it is less usual to see crime rates change considerably for the worse in (semi)rural affluent neighbourhoods of group 2.
- Neighbourhoods highlighted at *I* are of interest because these sit within a block of fairly isolated council housing estates, surrounded by more affluent suburbs.

- The neighbourhood at *H*, is unusual in that it belongs to the more affluent north suburbs group (6) and is part of Chapel Allerton - an otherwise improving area which has experienced rapid growth in property values.

- And finally considering change in damage to vehicles (Figure 7.12), Pudsey town centre (location *K*) is of interest, especially since it belongs to the one group which captures such neighbourhoods reasonably well.

- Headingley and Becketts Park (location *J*) are of interest given the presence of the student community and the disproportionate risks they face from burglary and other types of theft.

- Location *L* is interesting because Lincoln Green and the adjacent St. James Hospital are situated in the poorest of groups and it is perhaps more usual to expect to find worrying trends trends here, not encouraging ones.

Thus, the map reading process involves a synthesis of the information portrayed on the Figures in question, but also draws on knowledge of the group geography and the experience of the analyst and their subjective understanding of particular neighbourhoods and crime locations. It could be argued that there are risks associated with allowing subjective judgments to enter into the analysis at this stage and that neighbourhoods should be singled out for closer examination based purely on their data values. In which case, the maps would be superfluous, even dangerous. Yet, the boxplots and scatter plots provide the more objective selection opportunities, and what mapping can help do is tap into practitioner experience and local intelligence to generate questions and hypotheses for further analysis.

Summary of Within-Group Analysis

The results from the standardised mapping exercise can be tallied along with the results from the boxplot analysis and scatter plot analysis. When the tallies are summed and ranked it is possible to begin to sift and sort the neighbourhoods into a final sample to take forward for further investigation. Reasons for not progressing with certain neighbourhoods, despite them having been highlighted by the EDA, could include small number problems affecting the usefulness of percentage change over time figures; small denominator populations which have adversely affected rates; and, relatively few tallies. In this analysis three or more tallies was the threshold used, with some neighbourhood with only two tallies added where the indicators were particularly strong. Given time, all neighbourhoods flagged by the EDA process could be investigated further, but in practice, the community safety analyst has to work quickly and efficiently and some amount of prioritisation has to be exercised. Table 7.1 lists those neighbourhoods that, through the three EDA processes, self-selected themselves for further analysis.

SCA	Pop.	Cluster (SC)		Disaggregated Criminal Damage Offences					
				2000/01			2003/04		
		Hard	Nearest	Dwel.	Veh.	Oth.	Dwel.	Veh.	Oth.
44.05	1711	3 (.68)	1 (.08)	11	25	17	70	69	31
70.01	2143	1 (.66)	7 (.09)	32	83	82	149	132	98
26.01	1675	2 (.36)	5 (.23)	5	9	7	12	25	8
82.06	1676	8 (.51)	7 (.29)	38	22	20	60	71	33
17.02	1640	4 (.32)	1 (.26)	21	40	18	53	33	15
40.02	1826	7 (.73)	8 (.14)	48	29	19	112	88	33
47.01	1503	7 (.65)	8 (.19)	68	39	56	53	26	17
43.02	1674	9 (.37)	5 (.27)	13	25	96	28	17	6
63.01	1852	5 (.30)	8 (.24)	6	13	1	15	38	30
71.01	1501	6 (.68)	5 (.08)	5	25	6	15	26	9
33.02	1503	10 (.27)	5 (.27)	4	7	2	23	15	6
25.02	1874	2 (.71)	5 (.13)	2	42	37	5	24	15

Table 7.1: Criminal damage rates (disaggregated) and partition diagnostics, including cluster strength coefficients (SC), for a selection of problematic neighbourhoods

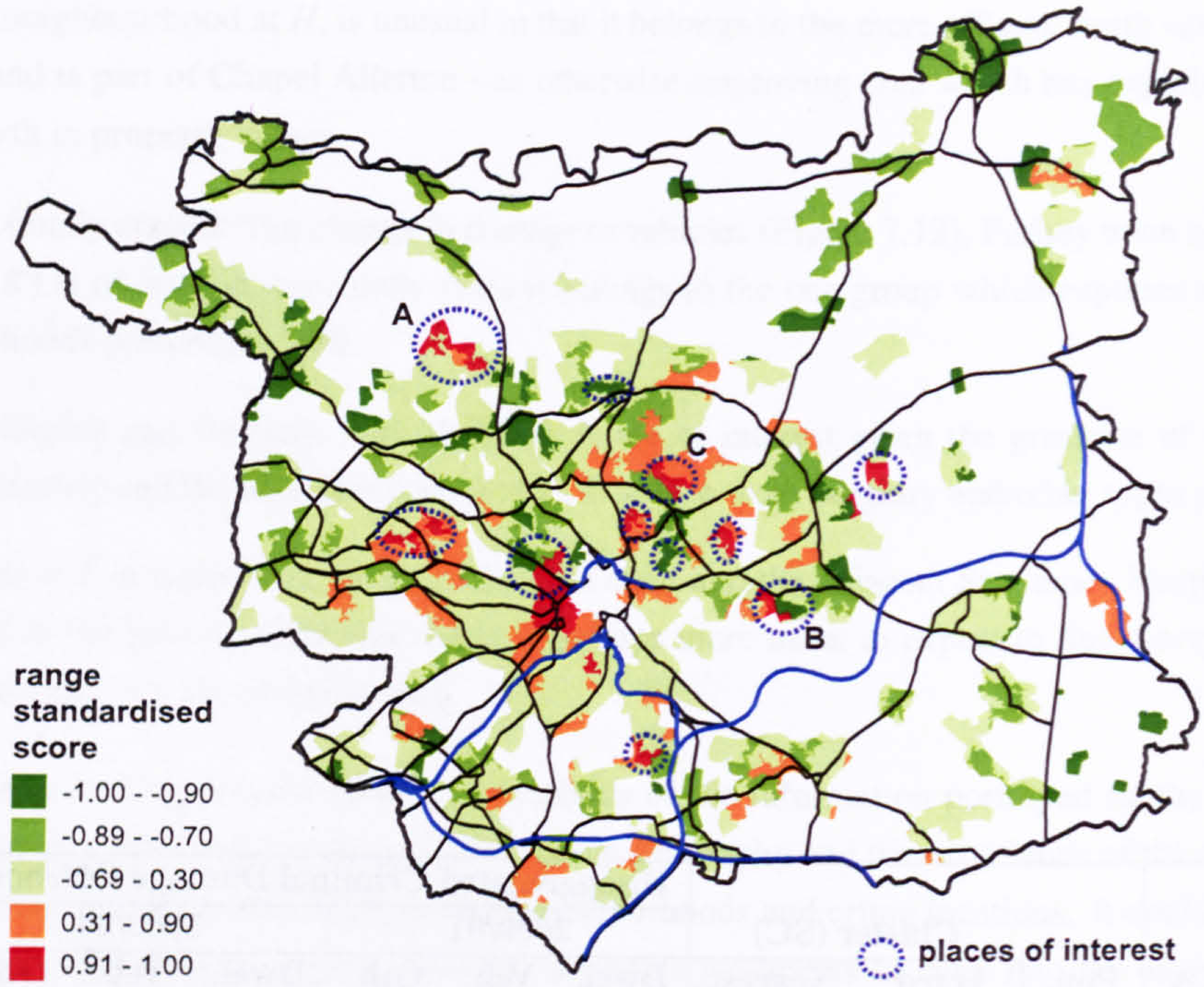


Figure 7.9: Map of Damage to a Dwelling rates, 2003/04, standardised by group.

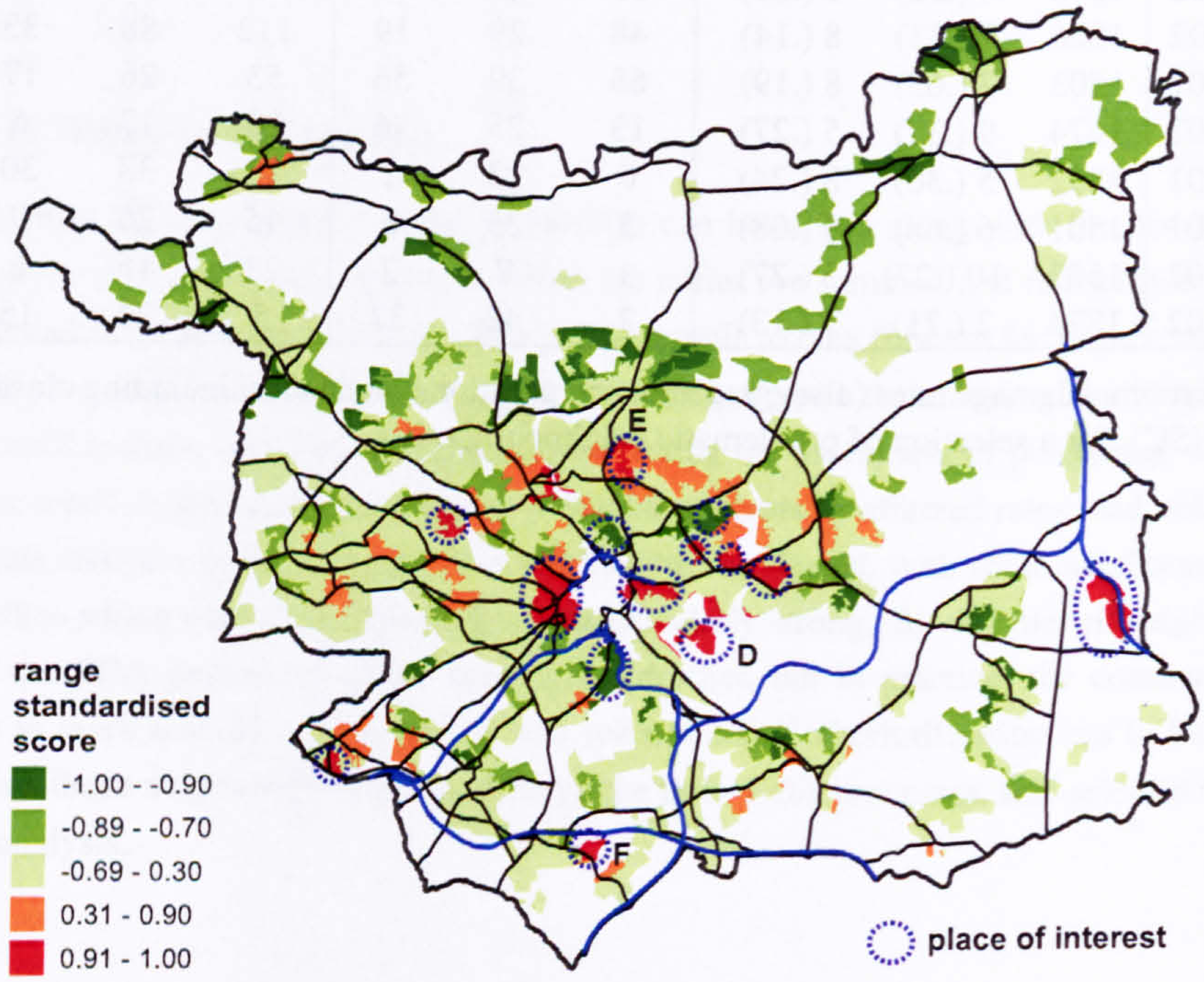


Figure 7.10: Map of Damage to Motor Vehicle rates, 2003/04, standardised by group.

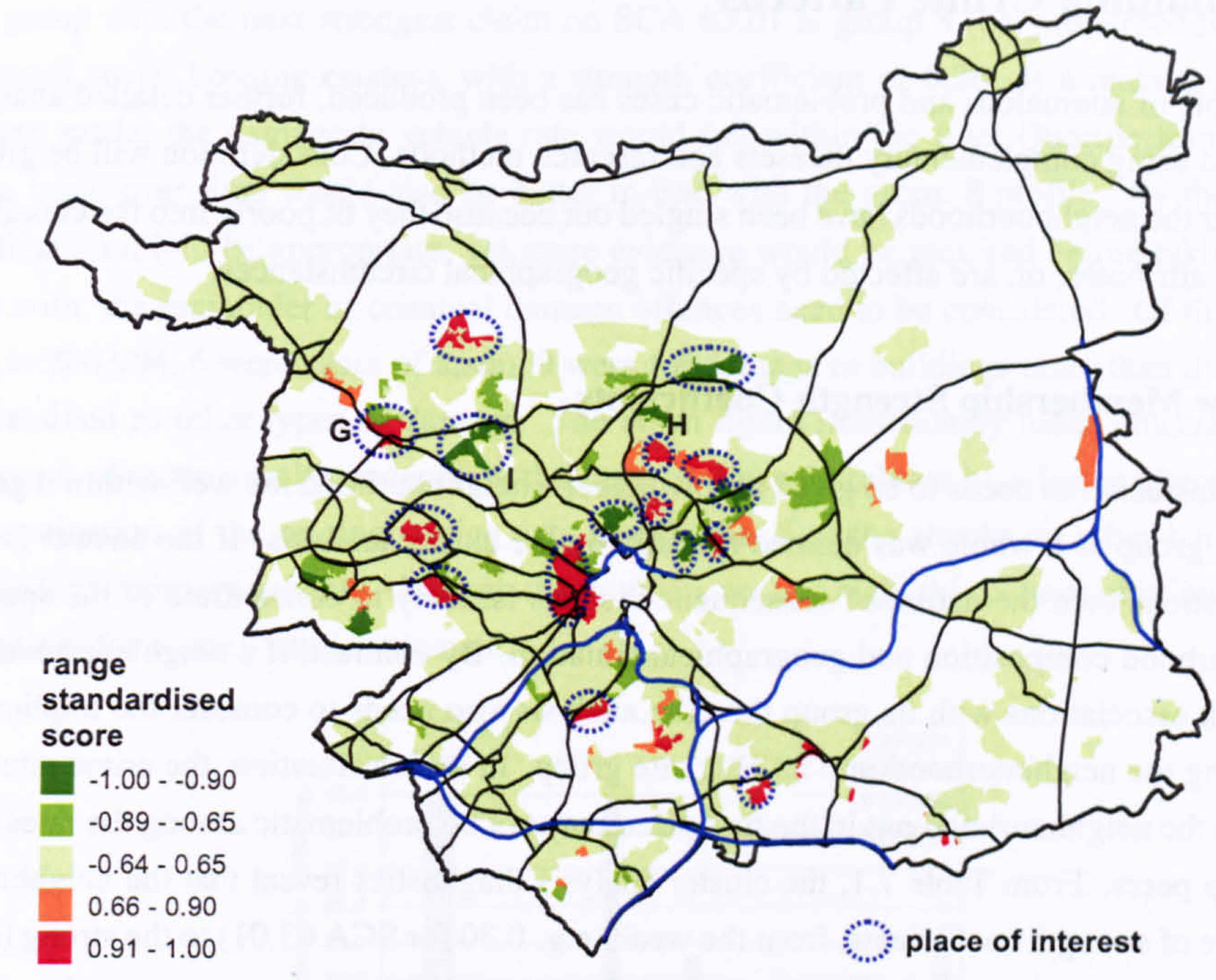


Figure 7.11: Map of change to Damage to Dwelling rates, 2001/02 to 2003/04, standardised by group.

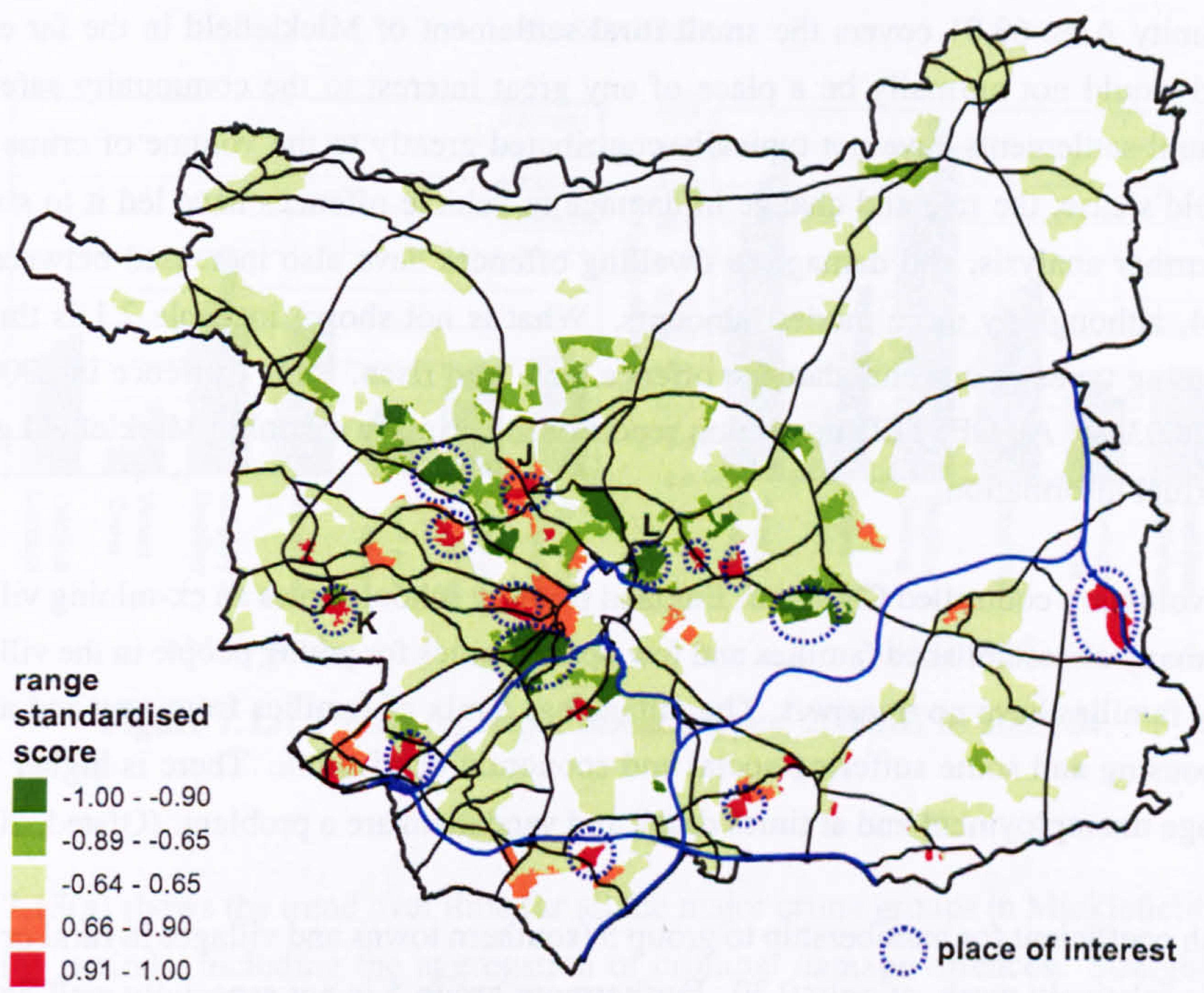


Figure 7.12: Map of change to Damage to Motor Vehicle rates, 2001/02 to 2003/04, standardised by group.

7.4 Explaining Crime Patterns

Once a sample of anomalous and problematic cases has been produced, further detailed analysis can be conducted using complementary datasets and research methods. Consideration will be given here as to whether the neighbourhoods have been singled out because they fit poorly into the classification; have unique attributes; or, are affected by specific geographical circumstances.

7.4.1 Low Membership Strength Coefficients

To begin, consideration needs to be given as to whether a neighbourhood fits well within it group and whether the group as a whole was defined robustly by the cluster analysis. If the answer is positive to both questions then the nature of subsequent analysis is likely to concentrate of the specificities of neighbourhood composition and geographical situation. By contrast, if a neighbourhood has relatively weak associations with its group then the analysis also needs to consider the implications of re-classifying the neighbourhood into an alternate group. In such a situation, the crime rate or trend that singled the neighbourhood out in the first place may not be problematic among the rates of a new set of group peers. From Table 7.1, the cluster analysis diagnostics reveal that the neighbourhoods have a range of strength coefficients, from the weak (e.g. 0.30 for SCA 63.01) to the strong (e.g. 0.73 for SCA 40.02). Below, two neighbourhoods with weak ties to their group are considered in more detail.

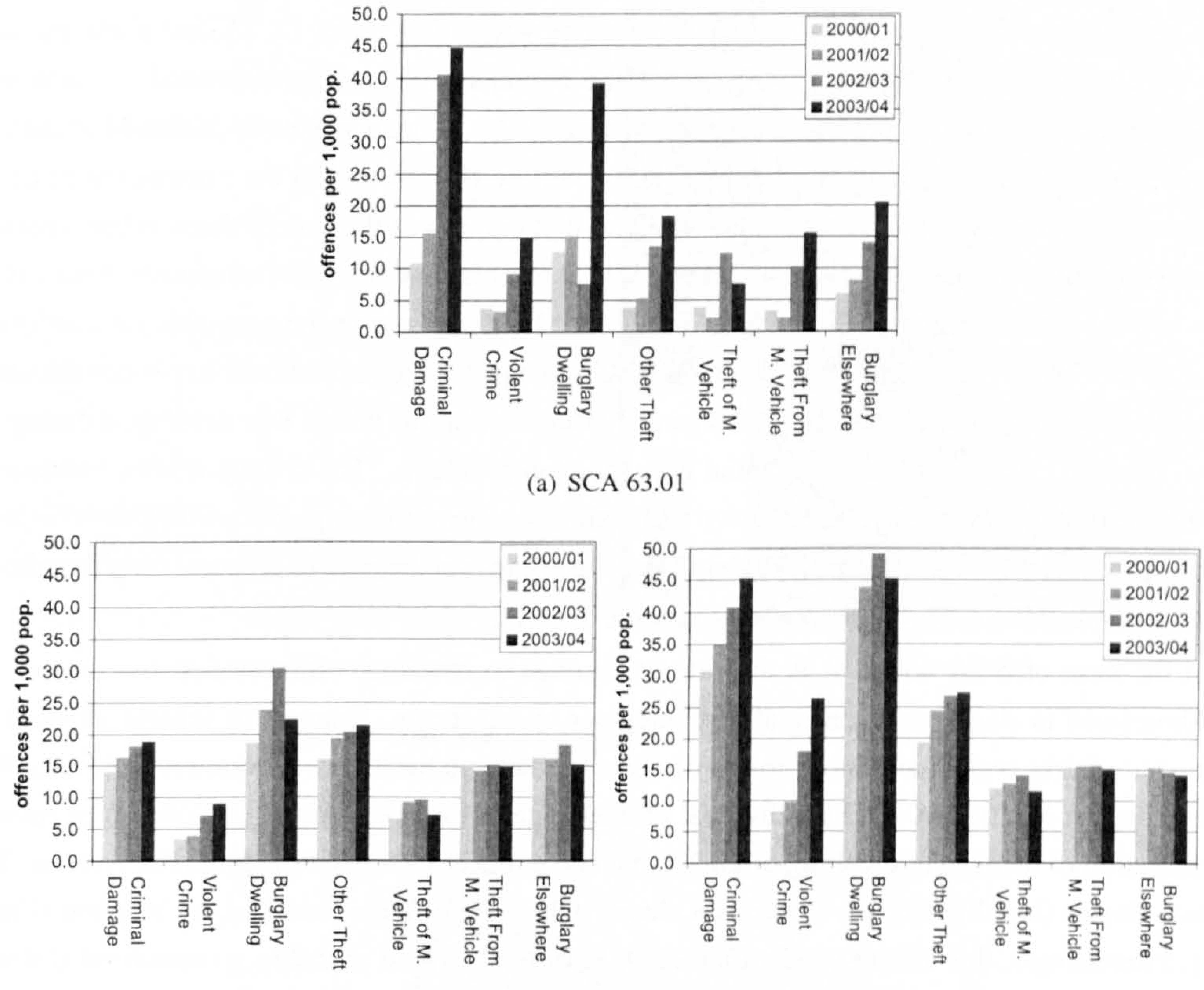
Micklefield

Sub-Community Area 63.01 covers the small rural settlement of Micklefield in the far east of the District, and would not normally be a place of any great interest to the community safety analyst given that rural settlements have not typically contributed greatly to the volume of crime in Leeds. In Micklefield's case, the rate and change in damage to vehicle offences have led it to stand out as worthy of further analysis, and damage to dwelling offences have also increased between 2000/01 and 2003/04, although by more modest amounts. What is not shown in Table 7.1 is that the sum of the remaining types of criminal damage offence have also risen, from 1 offence in 2000/01 to 30 offences in 2003/04. An OFSTED inspection report of the primary school in Micklefield gives some extra contextual information,

> This voluntary controlled Church of England primary school serves an ex-mining village with many old established families and few opportunities for young people in the village. Many families have no transport. The village has a mix of families from new and affluent housing and some suffering social and economic deprivation. There is higher than average unemployment and at times drugs and vandalism are a problem. (Ofsted, 2003)

The strength coefficient for membership to group 5 (southern towns and villages in rural or semi-rural locations) is relatively weak, at only 0.30. Furthermore group 5 is not especially well defined, with a silhouette width of 0.13, although in other ways it appears a reasonable grouping with a definite spatial pattern to its membership. What needs to be considered is whether SCA 63.01 would be better suited to belonging to a different group, or whether, when other factors have been considered, it may be better to leave it in group 5 and to look for alternate explanations of the higher crime levels.

The group with the next strongest claim on SCA 63.01 is group 8 (neighbourhoods typified by less-deprived social housing estates), with a strength coefficient of 0.26. If a reassignment to this group were made, the damage to vehicle rate would fall within the Inter Quartile Range of group 8 and the trend over time would also be better in-line with the group 8 profile. By these terms, a re-classification might be appropriate, yet more evidence would be required before taking this step. To begin with, the remainder of criminal damage offences need to be considered. Of these extra 30 offences in 2003/04, 6 were cases of arson, 9 were for damage to buildings other than dwellings and 15 are classified as other types of damage. The arson figure immediately looks unusual given that police records of arson usually undercount the scale of the real problem by a factor of approximately 7, and examination of the original police data reveals that 5 of the attacks on other buildings were against the local primary school. Although the numbers are not high, this pattern of criminal damage seems unusual for a small rural settlement.



(a) SCA 63.01

(b) Group 5

(c) Group 8

Figure 7.13: Trends for major offence types, 2000/01 to 2003/04.

Figure 7.13(a) shows the trend over time for all the major crime groups in Micklefield for the last four reporting periods, including the aggregation of criminal damage offences. Straight away it is clear that criminal damage has risen dramatically, and that burglary dwelling has also risen sharply in the last period. Most other offence types also saw large rises in percentage terms in the last period. Compared to average trends for group 5 (Figure 7.13(b)), Micklefield does appear to be different in a number of respects - particularly the size of recent increases for most offence types. Yet, when these patterns are compared to the trend profile for group 8 (Figure 7.13(c)), it is not clear that Micklefield would be better suited in this alternative group. Certainly, the recent rate of criminal damage and burglary dwelling in Micklefield, along with the more steeply rising trend for criminal damage, have more in common with group 8 than group 5. Yet, for the other offence types Micklefield appears to have more in common with the crime profile for group 5. To summarise, on this evidence SCA 63.01 should not be considered mis-classified. Instead, further investigation is required into the nature and make-up of the neighbourhood to try and explain the criminal damage (and probably other crime) patterns seen here.

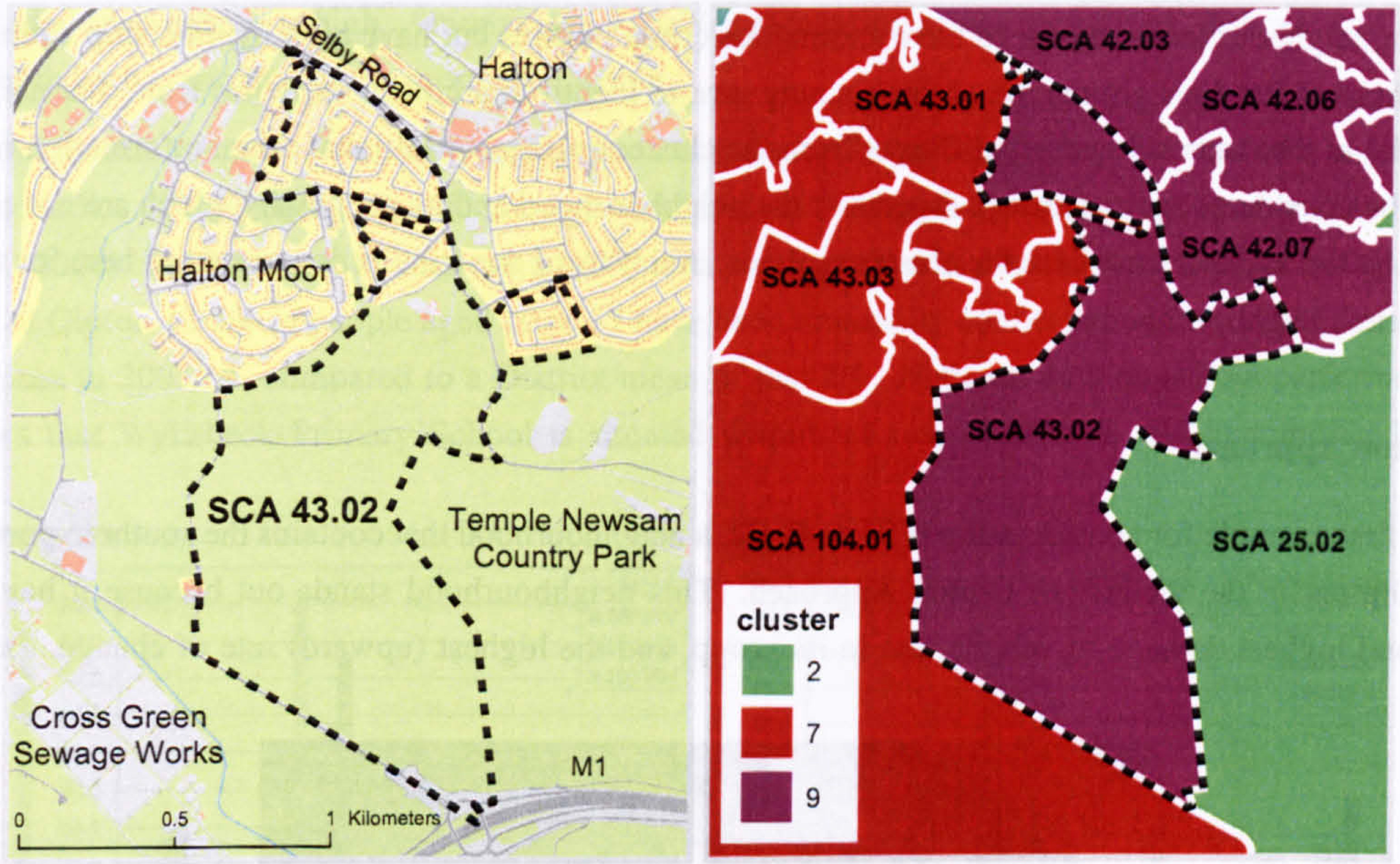
Halton Moor

A second example of the role strength coefficients can play in explaining anomalous crime patterns is provided by SCA 43.02, a part of Halton Moor. Unlike the previous example, this neighbourhood is not isolated. In addition, Halton Moor has a history of social and crime problems and has seen the delivery of many crime reduction policies, including wide-scale use of CCTV and plural policing (Crawford et al., 2004b). Figure 7.14(a) shows the boundary of the neighbourhood overlaid onto detailed cartographic data from the topography layer of the Ordnance Survey MasterMap dataset. This mapping reveals that the main area of interest will be the housing in the northern third of the neighbourhood, and that the farmland and parkland that comprises the rest of the neighbourhood is probably of only marginal interest. The mapping also reveals that the Sub Community Area (SCA) geography has carved through the middle of Halton Moor, separating the area along a northwest-southeast axis. With the benefit of hindsight, the SCA geography perhaps should have been manually flexed to accommodate the continuity of the neighbourhood (all of which was developed during the 1940s (Fowler, 1967)) - reducing potential classification problems. The current reality, however, is even more complex, as massively defensive road calming schemes have produced their own disjointing of the area. As well as land use considerations, the group membership of adjacent neighbourhoods needs to be considered, and this is shown in Figure 7.14(b).

In the case of SCA 43.02, it is the rate of damage to dwelling offences that has caused this neighbourhood to stand out as unusual for a member of this group. Damage to vehicle offences, by contrast are fairly typical and other forms of criminal damage have seen a dramatic fall from 96 in 2000/01 to just 6 in 2003/04. The individual offence data in the West Yorkshire Police files shows a concentration of damage to dwelling offences, and some repeat victimisation, in Kyffin Avenue. This offence-cluster (highlighted) then becomes clearer when the offence locations are mapped (Figure 7.15). Interestingly, this cluster of 12 offences (43% of total) would probably go unnoticed if it were part of neighbouring SCA 43.01 or SCA 43.03. The pattern of offences within the remainder of the neighbourhood is much more sparse, and has more in common with neighbouring (to the east) SCA 42.07 and SCA 42.07. The street layout of SCA 43.02, with the Kyffin Avenue area removed, also looks more like that seen in adjacent easterly neighbourhoods than it does the rest of Halton Moor, suggesting that the housing was built at a different period and perhaps for different groups of tenants. The suggestion then, is that damage to dwelling patterns might appear anomalous because the neighbourhood geography is improperly defined, rather than the neighbourhood being assigned to an alternative group. Some extra support for this idea could be read from the archived plans of the estate's development (Figure 7.16), which show that SCA 43.02 was conceived and developed seperately from the rest of Halton Moor. This argument is somewhat undermined, however, because of the ambiguity surrounding the development of Kyffin Avenue area - the street being shown on the

1938 plan, but not the housing along it.

However, it would not be sensible to change the neighbourhood geography on one piece of evidence alone, as other offence types may not conform to the same type of pattern. To investigate these, Figure 7.17(a) shows trends for the major offence groupings in SCA 43.02 over the last four recording periods. Firstly, the trend in criminal damage needs to be explained. In fact, the graph shows a corrected trend over the four years, as the NCRS has meant that post-April 2002, cases of arson following the taking of a motor vehicle without the owner's consent are no longer recorded (Halton Moor used to be one of the city's main hotspots for abandoned vehicles and car fires). This would account for some of the fall from 96 to 6 other types of criminal damage recorded between 2000/01



(a) Neighbourhood topography (b) Group membership

Figure 7.14: Mapping SCA 43.02 and its environs.

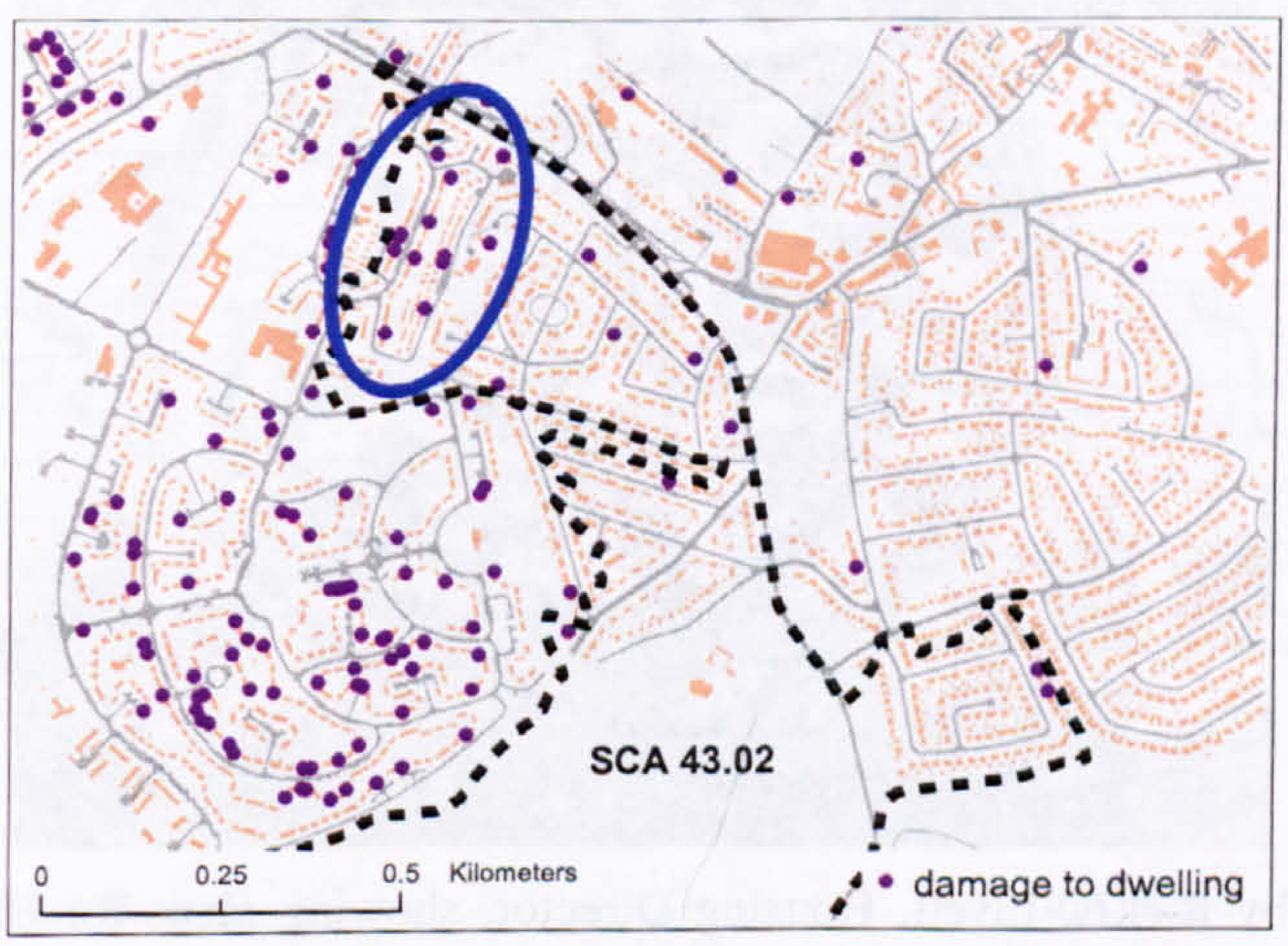


Figure 7.15: Map of Damage to Dwelling offence locations, 2003/04.

and 2003/04. This corrected pattern is more in line with the average for group 9 (Figure 7.17(c)), although the rate is higher.

Figure 7.17(b) shows the effect of removing the census output area that contains Kyffin Avenue from SCA 43.02. Although this lowers the damage to dwelling rate, most of the other crime rates change little. At the 14-offence classification level, only burglary dwelling really stands out as being very different from the group mean. This might not be enough to warrant a reclassification, but further analysis may be worthwhile to consider nearby burglary dwelling hotspots and possible displacement effects caused by the presence of CCTV in Halton Moor.

7.4.2 Unique Neighbourhood Features

When neighbourhoods appear to have anomalous crime profiles but have high membership strength coefficients for their groups the emphasis may turn to identifying cluster variables from that neighbourhood that have values very different from the cluster average. Additional crime information may also be considered, as can known aspects of the neighbourhood and its geography which are not covered by the variables used for the cluster analysis. Analysis of this kind is demonstrated here for two different neighbourhoods.

Gipton Approach

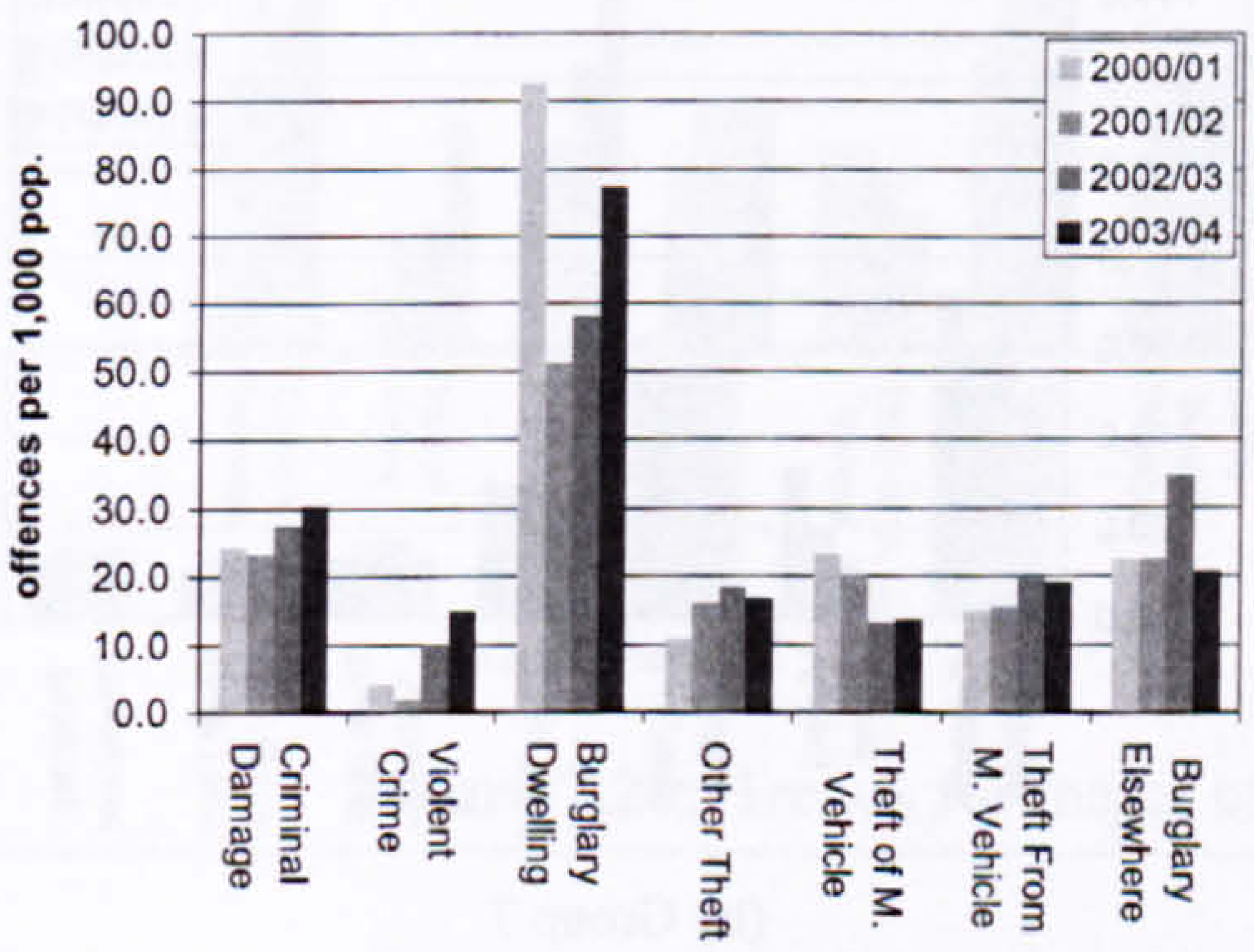
The first example for consideration is SCA 40.02, a neighbourhood that contains the southern portion of Gipton, in the vicinity of Gipton Approach. This neighbourhood stands out because it has the second highest damage to vehicle rate in its group, and the highest (upward) rate of change. Dam-

Figure 7.16: Plan by R.H.A. Livett, Housing Director, showing plans for Halton Moor housing estate, 1938. Source: Leodis photographic archive.

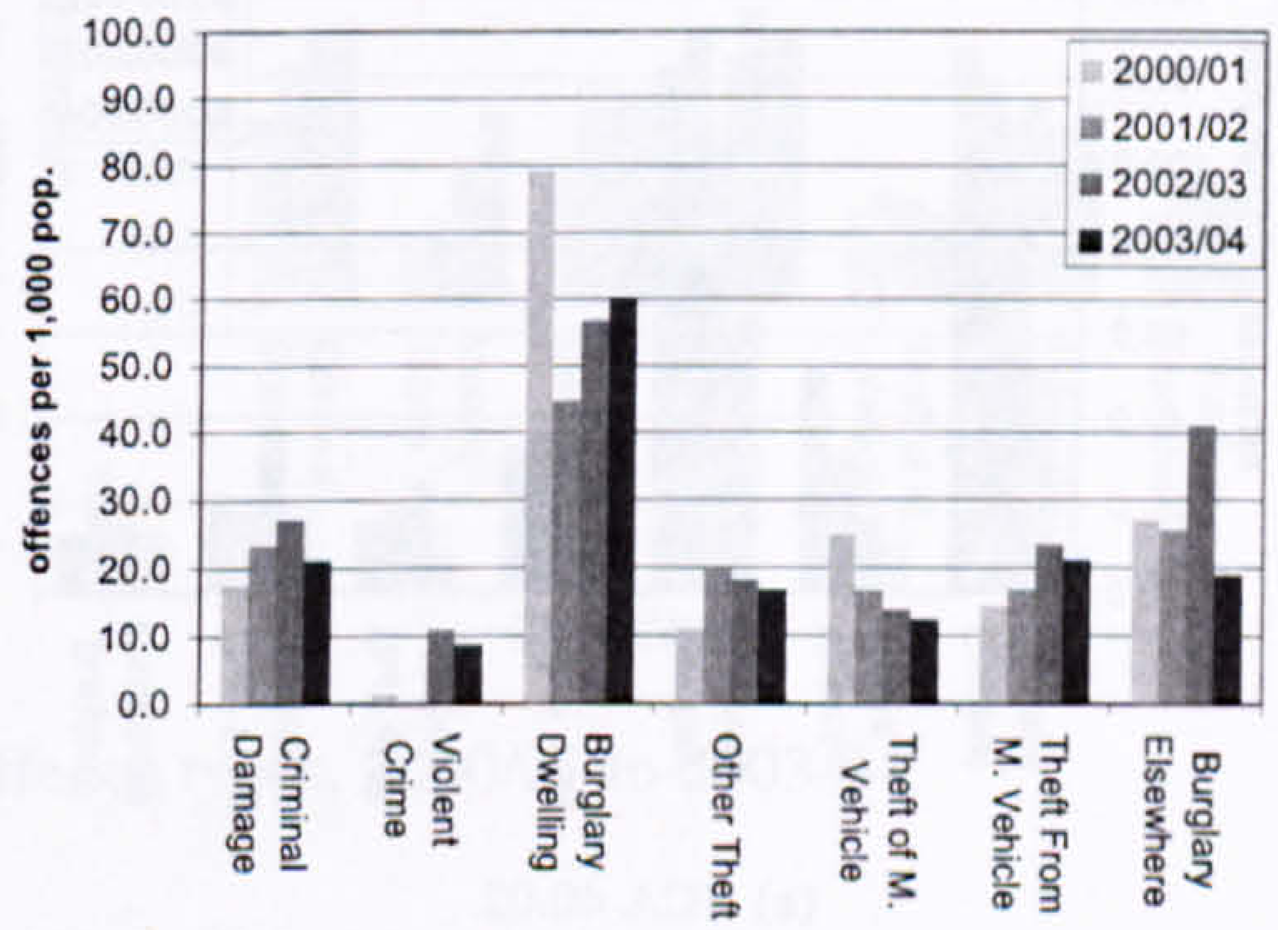
age to dwellings is also high, but is not considered especially problematic within group 7 (poorer social housing estates). The aggregated count of all criminal damage is high within the group (Figure 7.18(a)) and burglary dwelling rates are also much higher than is typical for this type of neighbourhood (Figure 7.18(b)). Violent crimes and thefts from vehicles are somewhat higher than average for the group, but the remaining crime types are about average. This mix of crime problem might suggest that a range of behavioural and/or economic needs are being satisfied.

If the values of the variables within the range found within group 7 are considered then a number stand out as being atypical. To begin with, the lower age groups are under-represented and there is a greater than usual proportion of households living in flats (while terraced houses are less common than usual). The proportion of properties that are void is high and the problem of children being excluded from school is high. Amongst the variables measuring minor incivilities, disorder and ASB are both high compared to the group as a whole.

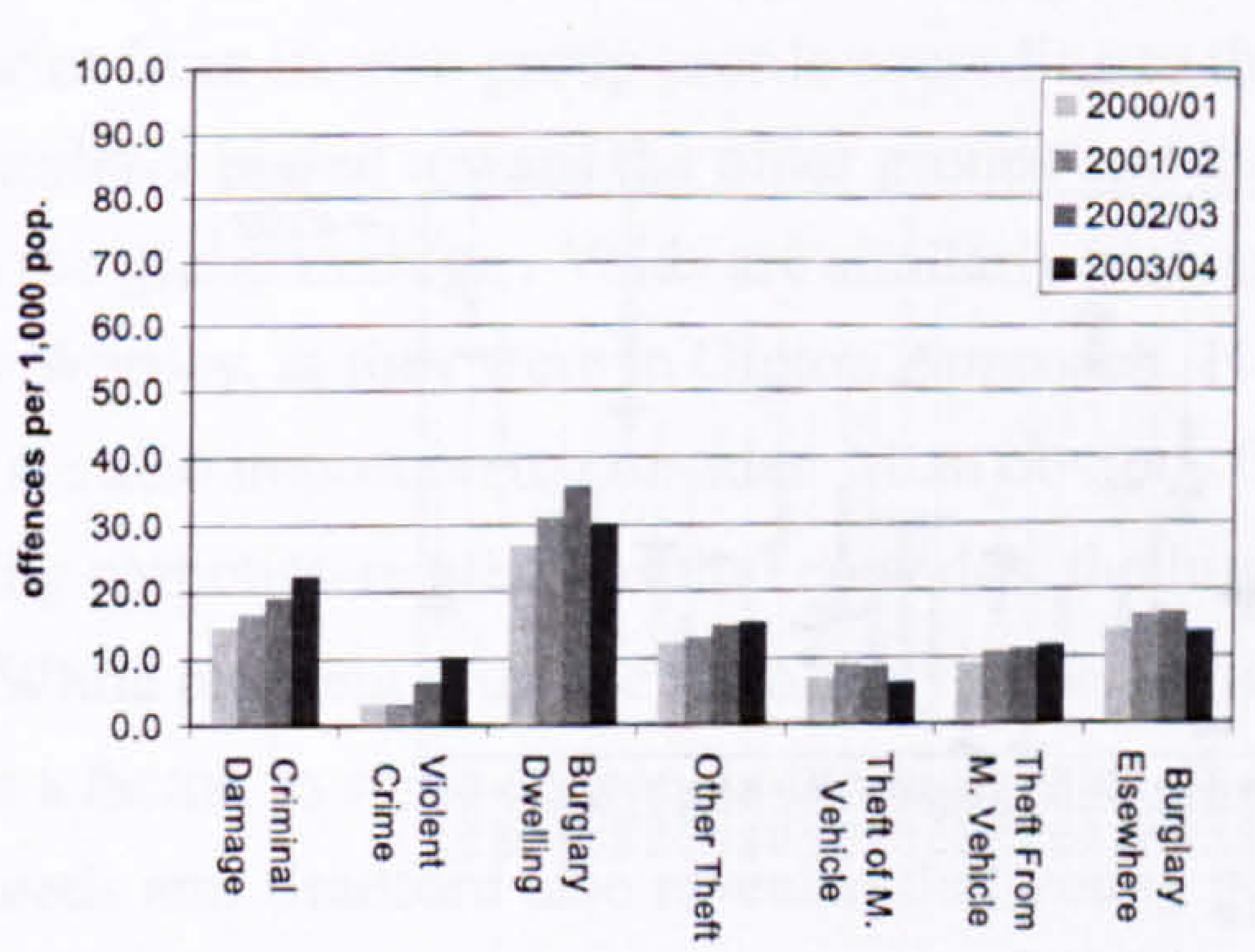
Additional sources of data and intelligence can enhance the picture further, for example, in Leeds-wide studies of offenders, criminal damage stands out as having the youngest age profile (Shepherd et al., 2004). Furthermore, by analysing WYP data on suspects it is possible to show that within this part of Gipton 12.5% of people aged 10 to 17 have been nominally connected with criminal damages offences in 2003/04, compared to a District mean of just 2%. National data on school performance shows that Wykebeck Primary School is situated within the neighbourhood and unauthorised ab-



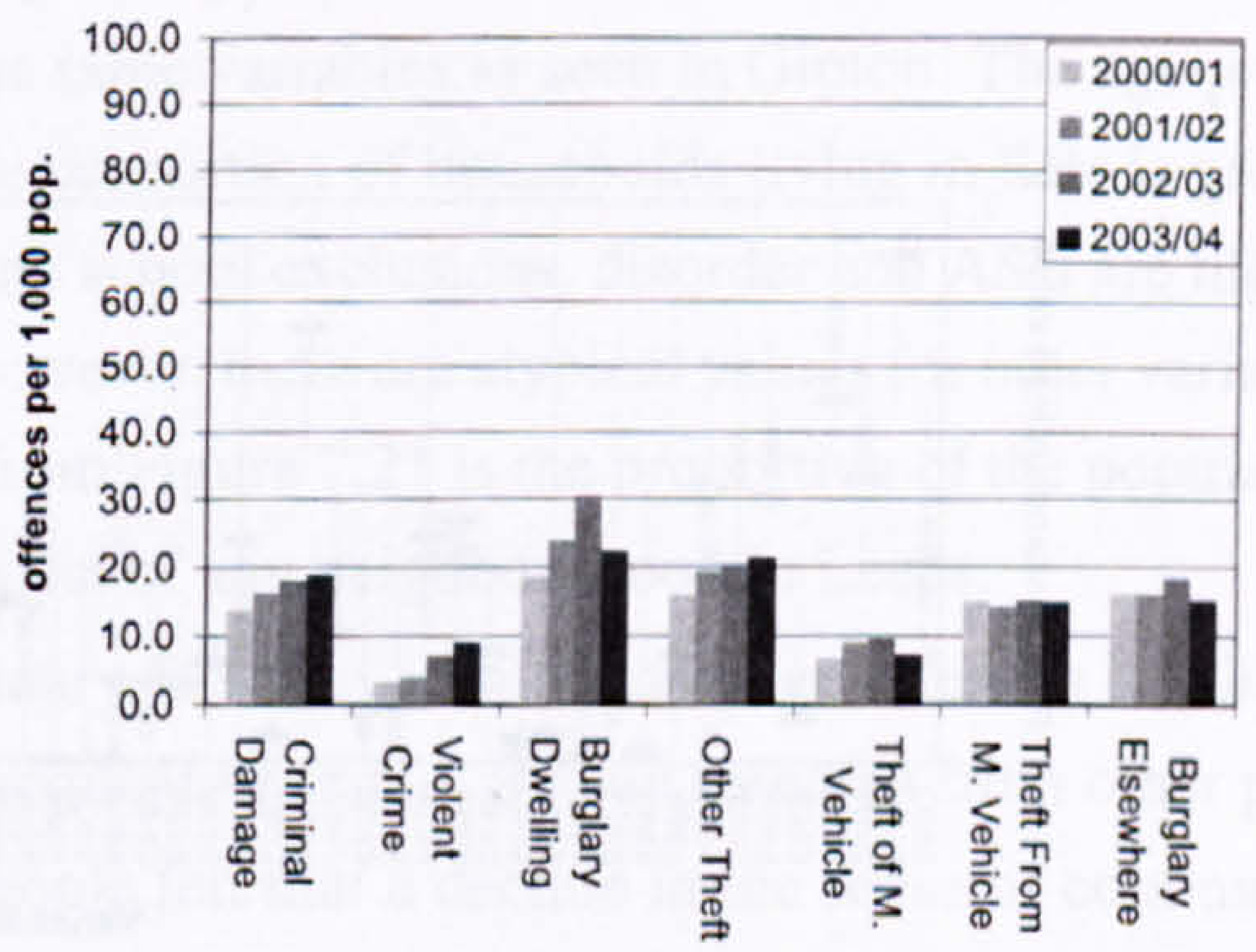
(a) SCA 43.02



(b) SCA 43.02 minus the census OA covering Kyffin Avenue



(c) Group 9

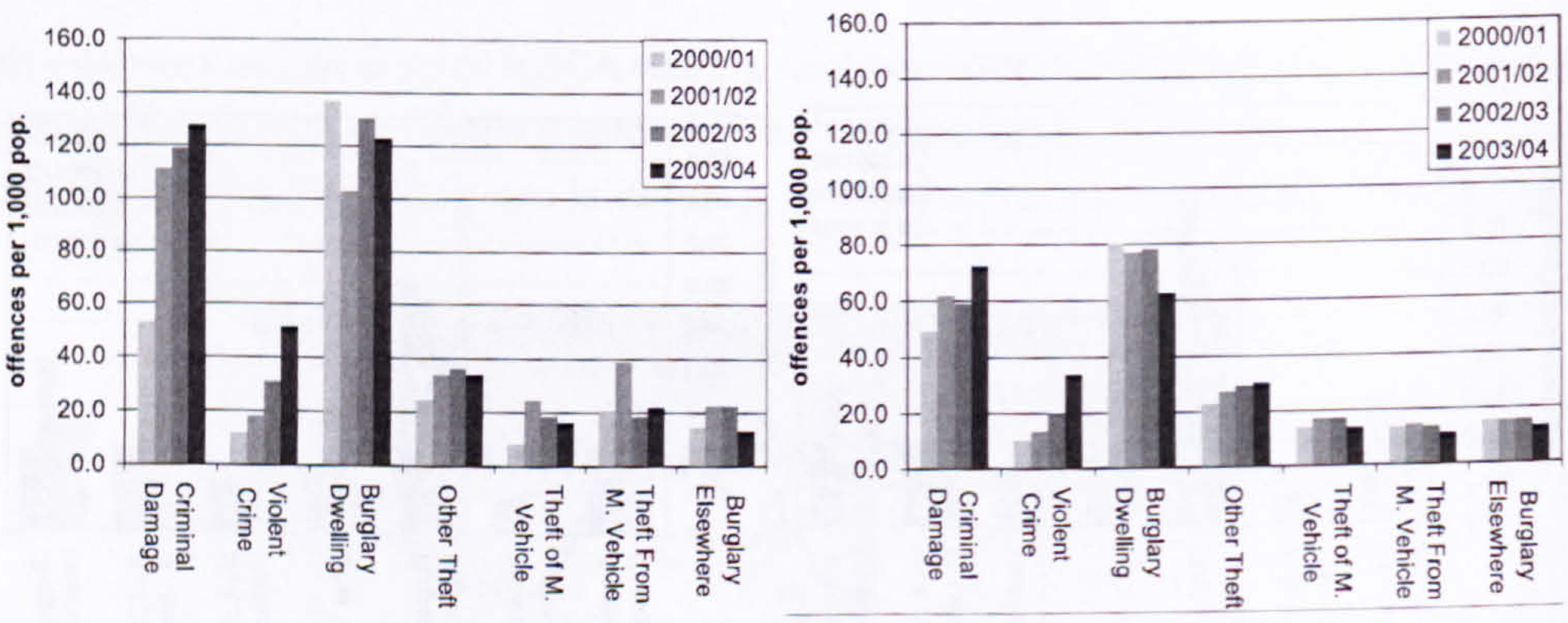


(d) Group 5

Figure 7.17: Trends for major offence types, 2000/01 to 2003/04.

senteeism is high at 2.3% (0.5% is the Leeds average) and performance at Key Stage 2 is amongst the lowest in the Leeds (Department for Education and Skills, 2006). While these children may not account for a large proportion of the criminal damage their apparent lack of success at school may not bode well for their future. For secondary education, analysis of the locally sourced Pupil Level Annual School Census (PLASC) reveals the highest proportion of pupils go to John Smeaton Community High School. Here, pupils are in a learning environment which saw the highest rate of unauthorised absence in Leeds in 2003 (6.67%) and only 85.04% attendence. Key Stage 3 results at John Smeaton in 2003 were also very low, with averages of only 27%, 31% and 39% for science, english and maths respectively (ibid).

A more detailed analysis of crime records indicates that 46% of the criminal damage (damage to dwellings and vehicles) in this Gipton neighbourhood occurs between 6.00 p.m. and 10.00 p.m. with monthly peaks occuring in August (coinciding with school holidays) and February. Levels of recorded domestic violence (28.5 offences per 1,000 pop.) are only slightly higher than the cluster mean (26). Besides the primary school, there are few services located in the neighbourhood. The Gipton station of the West Yorkshire Fire and Rescue Service is one exception, but fire officers give the impression that the station has the feeling of being under siege and appliances attending local calls will routinely deploy in pairs - one for the fire and one to provide protection from the threats posed by gangs of local youths.



(a) SCA 40.02 (b) Group 7

Figure 7.18: Trends for major offence types, 2000/01 to 2003/04.

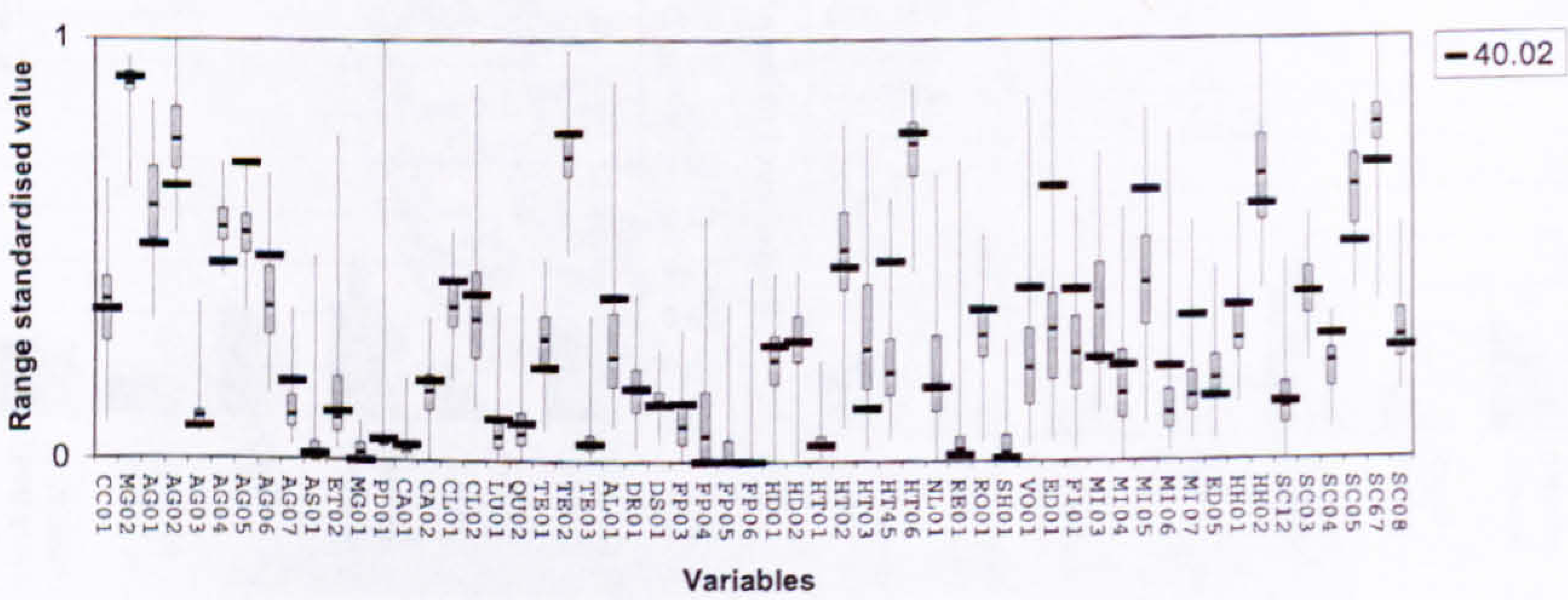
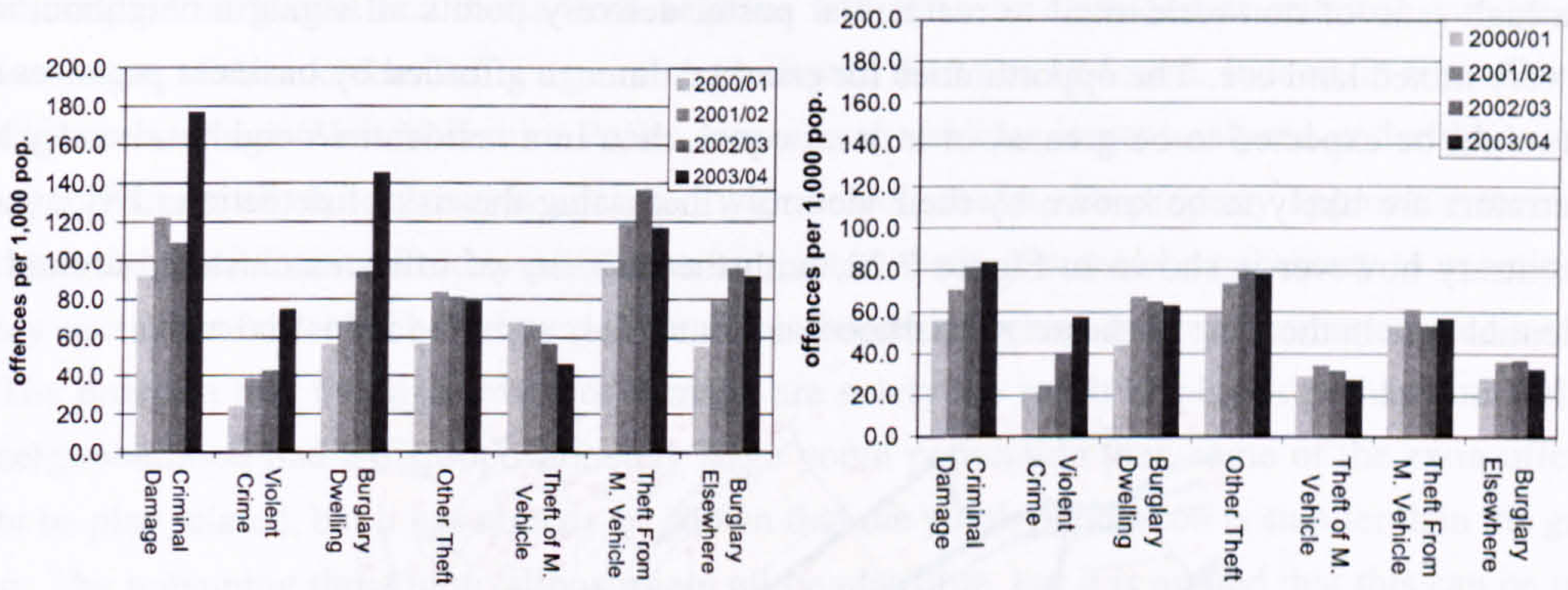


Figure 7.19: Comparison of values of variables in SCA 40.02 with distribution of values within cluster 7.

Considered together, these various indicators portray a neighbourhood under strain but it is not immediately clear why criminal damage should have the profile that it does. The lack of local services may extend to services and facilities for young people, and this is something that could be investigated. It is possible that motivations to vandalise property are connected with a need to exert a territorial claim on the area, and a stronger argument for such a hypothesis might be made if it could be shown people were targeting certain types of location but not others. An opportunity to consider such a situation occurs within the next neighbourhood selected for further analysis, New Wortley.

New Wortley

Like Gipton Approach, the reason New Wortley (SCA 70.01) stands out as atypical within its group (group 1, poor inner city neighbourhoods) is because its criminal damage rates are high. Specifically, criminal damage to both dwellings and vehicles was much higher than the group mean in 2003/04 and the rate of change is also much worse than the group mean for damage to dwelling. Many of the other major crime type rates (Figure 7.20(a)) are also appreciably higher than the group mean (Figure 7.20(b)).



(a) SCA 70.01 (b) Group 1

Figure 7.20: Trends for major offence types, 2000/01 to 2003/04.

While New Wortley is in a different group to Gipton Approach it is interesting to consider whether it varies from its own group profile according to the same variables as seen in Gipton. The age profile is similarly biased toward the older groups and the proportion of households living in flats is greater than the group average . Voids are similarly high and school exclusions, disorder and ASB are high in New Wortley, as they were in Gipton Approach. However, there are atypical values for other variables that are also important to consider. Most obvious from Figure 7.21 is the proportion of the population having narcotics-related hospital episodes, the highest of any neighbourhood in Leeds.

While problem drug use is unlikely to be a primary factor in criminal damage offences it is likely to be a factor, to some degree, in the high rates of acquisitive crime. Recent research from other parts of Leeds and Bradford also revealed that young people felt that a decline in the sense of community in their own neighbourhoods was connected with increases in the availability of drugs (Carr, 2003). Furthermore, this study group - itself socially excluded young offenders and truants - saw itself as disconnected with the communities formed by drug users. So while traditional conflicts over space in New Wortley might be playing out between young people and the authorities that construct and maintain space that exclude or subdue them, there may also be conflicts over space and identity between young people and local drug using groups.

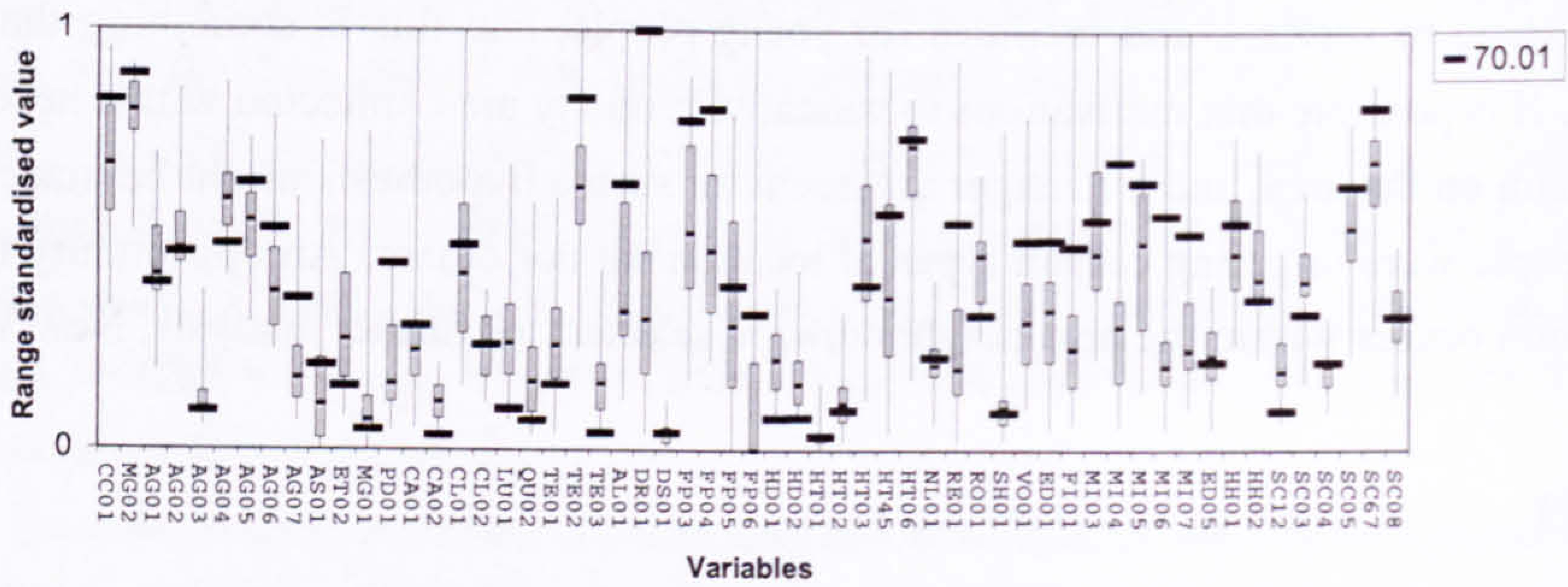


Figure 7.21: Comparison of values of variables in SCA 70.01 with distribution of values within group 1.

The variable profile also shows atypical values for several important land use variables. A high daytime to residential population ratio; higher than usual cars at the workplace; low housing density and a high ratio of non-residential to residential postal delivery points all signal a neighbourhood with very mixed land use. The opportunities for criminal damage afforded by business premises and plant might be expected to be greater, or at least equal, than in a residential neighbourhood where perpetrators are likely to be known by their victims - increasing the risk of detection. Evidence to the contrary however is shown in Figure 7.22, with the majority of offences clustered around the residential area in the west of the neighbourhood and much less so in the industrial area.

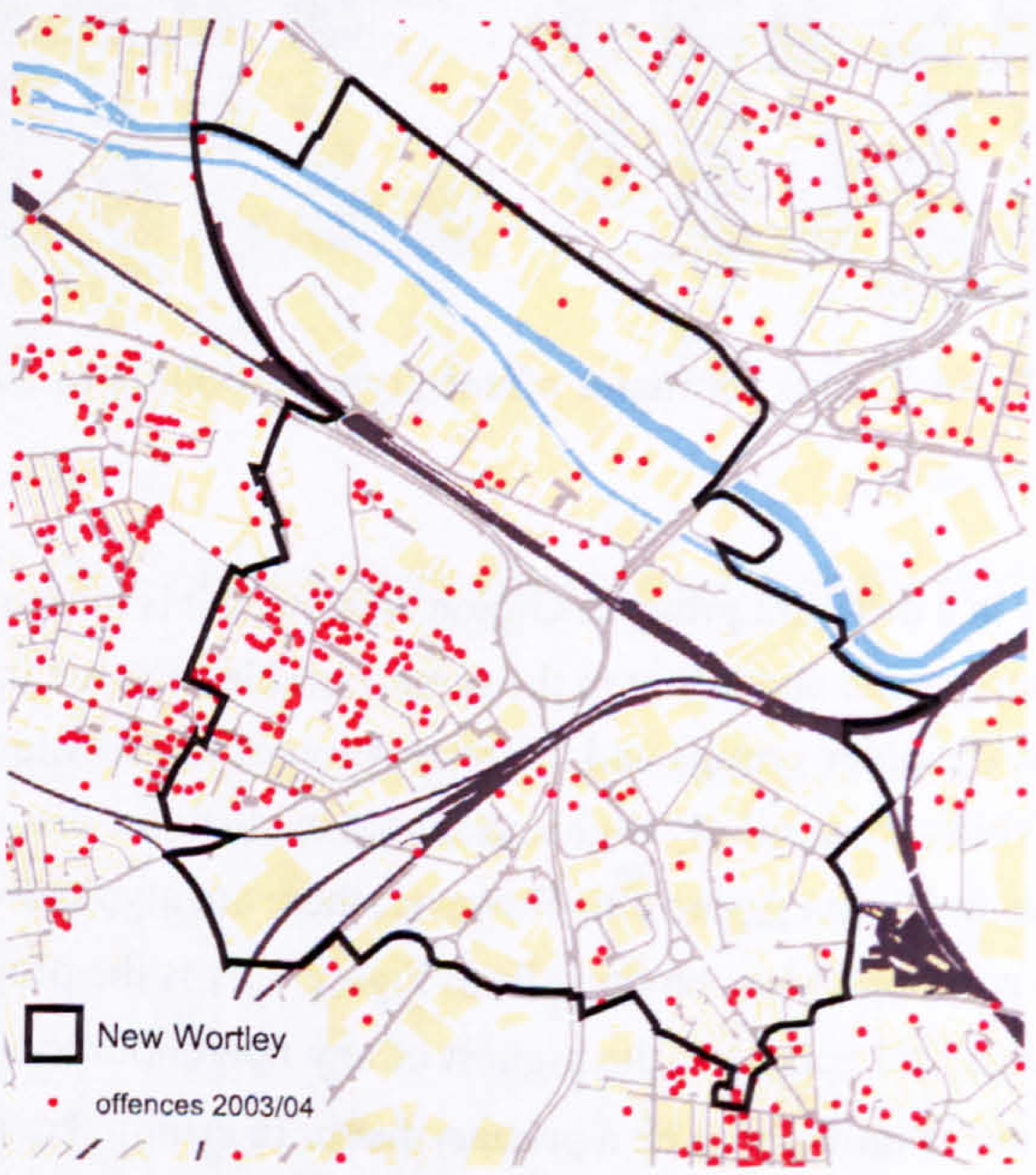


Figure 7.22: Location of criminal damage offences in New Wortley, 2003/04.

As in Gipton Approach, most criminal damage is caused in the evening, although the percentage (37%) happening between 6.00 p.m. and 10.00 p.m. is a little less than in Gipton. The same monthly peaks in August and February are also seen here. Within New Wortley there is a greater proportion of the population nominally associated with crimes although the proportion of juveniles is about half that seen in Gipton Approach. Interestingly though, analysis of the origin of nominals for all types of offence occuring in New Wortley reveals that this neighbourhood has the second highest proportion (28%) in the cluster coming from within, which is counter to the normal situation where most nominals come from (albeit not far) outide the neighbourhood. The PLASC data reveals that the majority of primary school pupils in New Wortley go to Castleton Primary School, which is an above average performer within Leeds (Department for Education and Skills, 2006). The secondary schools (West Leeds High and Wortley High) however are less good, with high levels of truancy, poor attendance and low Key Stage 3 results (ibid).

Although an exhaustive analysis of links between theoretical work and the patterns shown above is not possible here, a number of possible lines of enquiry could be followed, including Stanley Cohen's (1973) typology of the motivations for vandalism (Table 7.2). The prospect that criminal damage in Wortley is the result of acquisitive motivations is feasible given the high burglary rate and high theft from vehicle rate. The reason criminal damage and acquisitive crime types are often correlated is that attempted burglaries and thefts from vehicles will often be classified as criminal damage if no obvious signs of the incident being a genuine attempted theft can be found. Indeed, in some Division Intelligence Units at West Yorkshire Police, damage to vehicles is not analysed under the criminal damage group at all, but instead is analysed with vehicle thefts. To properly appreciate the extent to which the problems in Wortley are due to attempted thefts it would be necessary to examine the modus operandi text for each of the crime records, something which has not been attempted here.

The prospect that the high levels of damage are solely the result of play is perhaps unlikely. If the neighbourhood had a disproportionately large youth population then some of the extra offences might be play related, but it has already be shown that the youth population is smaller than the group mean. The remaining three motivations might all be plausible, but it is argued that this can be made clearer if to the typology are also added acts designed to signify a sense of belonging to a place or to reinforce boundaries and territory.

It was Phil Cohen that argued that, "in some inner-city neighbourhoods, the only way for young people to assert the dialectics of belonging is to efface or deface the official landscape in favour of their own landmarks of community which they create by these means" (Cohen, 1997, page 79). Furthermore, the point was made that this was the only way in which young people living within the regular and regulated environment typical of the large council estate could sometimes build their own defensible space. The following illustration makes the same point,

> Urbanisation is continually restricting the areas for their [the young] play and limiting what they may legitimately do. In multi-storey housing their play is hedged in by negatives - you mustn't play in the hall, chalk on the pavement, make a see-saw on that wall, cycle on this path. Moreover, the places where the child plays are becoming more exposed to public view. Parents contrast the relative privacy of their old back courts and 'our street' with the openness of the estate which has few defined spots to which their child has his own right. (Jephcott, 1971)

Further evidence of the way young people interact with their environment, and this time from contemporary Leeds, can be found in Carr's (2003) thesis. Study subjects from Bramley (a predominantly council run estate in west Leeds), were given a camera to record details of the spaces they frequented. These included the 'nav' (a local term used to describe the environs of the Leeds to Liverpool Canal) and a derelict property referred to as the 'haunted house'. Both of these sites were close-by the subjects' home estate but also separate and private, providing these and other young people with a place to congregate and claim their own.

Type	Description
Acquisitive	to acquire money or property, e.g. smashing coin boxes in telephone kiosks
Tactical	to gain attention for a particular cause, or sometimes to get arrested and detained
Vindictive	as an act of revenge for a real or imagined grievance; an outlet for anger
Play	often little malicious intent; sustained as a result of boisterous curiosity or competition
Malicious	may be a result of hatred; may also provide a sense of enjoyment and satisfaction

Table 7.2: Cohen's (Cohen, 1973) typology of vandalism

It is possible that the non-residential areas within and adjacent to New Wortley do not afford the same opportunities as the natural land north of Bramley and that some young people seek to claim space within their residential areas instead. It has been argued elsewhere that, in working-class cultures at least, 'you don't shit on your own doorstep' (Cohen, 1997), but perhaps when young people seek to mark out a local territory of their own that is what happens.

Consideration needs to be given to natural and man-made barriers that may hinder pedestrian travel from the residential to non-residential areas. Looking at the map of the area again (Figure 7.22), the railways very obviously carve up the neighbourhood into three distinct zones and main roads (A58 and A643) and the large traffic intersection in the centre of the neighbourhood provide further barriers. These formidable obstacles may also help to explain why criminal damage is confined to the residential area. If the housing were less self-contained and there more opportunities for privacy the need to efface the residential environment might be less.

7.4.3 Geographical Circumstances

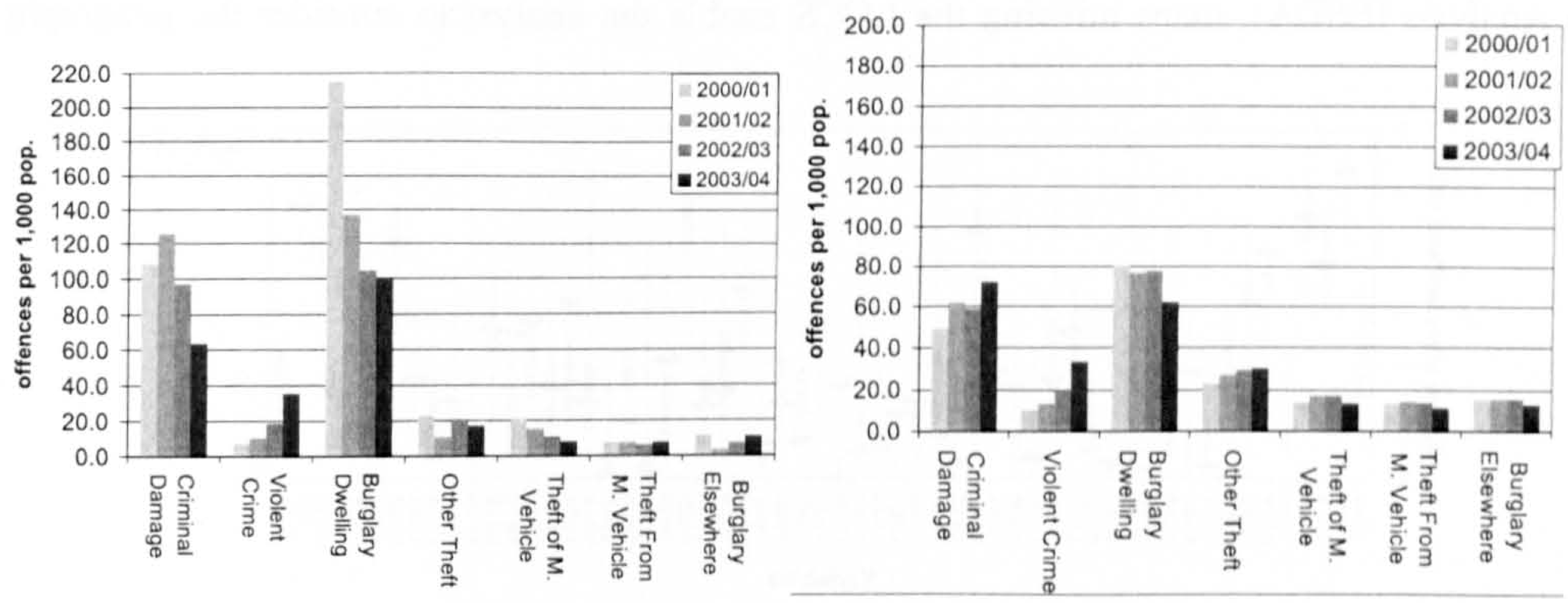
Hawksworth Wood

As a final example of how the classification can identify places of interest, consideration is given to a neighbourhood that has lower crime rates, or at least an encouraging trend, when compared to its cluster peers. SCA 47.01, in Hawksworth Wood, is such a place. As well as falls in damage to dwellings and vehicles Figure 7.23(a) shows a very substantial fall in burglary dwelling (slowing in the last period) and year-on-year falls in thefts of motor vehicles. By contrast, the cluster average for these offences (Figure 7.23(b)) has fallen much less. Burglary elsewhere and other theft might be expected to be lower in Hawksworth Wood given lower numbers of potential retail and business targets and this issue might also go some way to explaining why burglary dwelling is still higher than average.

Of the variables that define Hawksworth Wood (Figure 7.24), a number stand out as substantially different from the cluster mean. Firstly, the proportion of the population that are asylum seekers is almost five times higher than the average for group 7. Some of the other neighbourhoods in the group are not short of available housing so inquiries with the Leeds Refugee and Asylum Support Service might be made to see if Hawksworth Wood is considered a good location for asylum seekers for a particular reason. There is a greater than average proportion of people from the higher social classes and qualification levels are well above average. Despite this, the success in school is poor at GCSE level and exclusions are also worse than usual. Abandoned cars also appear to be a greater problem than elsewhere in the group, although analysis of fire service data reveals that deliberate car fires is down from 17 in 2003/03 to just 7 in 2003/04. Only 4.4% of juveniles appear in West Yorkshire Police's nominals database for criminal damage offences, less than Gipton Approach and New Wortley.

Two other issues that might be important for Hawksworth Wood both relate to its geographic location. Firstly, the neighbourhood is an island in the geography of its cluster, that is, the neighbourhood is isolated from its peers and is surrounded by neighbourhoods that are more affluent, and have lower-crime profiles. This, it could be argued, might reduce the likelihood of offenders migrating in to Hawksworth from nearby areas to commit crimes. Furthermore, SCA 47.02 (a member of group 8), which constitutes most of the rest of Hawksworth Wood, has also seen large downward trends in criminal damage and vehicle crime. The second issue involves school catchments at secondary level. In both Gipton Approach and New Wortley, the main secondary schools for local pupils were performing poorly. In Hawksworth Wood, by contrast, the three main schools that take pupils from the area (Lawnswood, Horsforth and Abbey Grange) are good to excellent for Leeds. Although the variable profile shows GCSE results were poor for Hawksworth Wood pupils in 2003, it might be that there are other, less quantifiable benefits to the children from Hawksworth Wood attending schools that are performing well overall.

Given that Hawksworth Wood self-selected itself because of change in crime rates over time it is it perhaps a weakness of the classification that it has been constructed with so few dynamic variables with which to make comparisons. Other data sources external to the classification do make statements about neighbourhood dynamics however, one such being NOMAD, Leeds City Council's Neighbourhood Orientated Model of Area Demand. The reason NOMAD was not itself incorporated into the classification is that it predominantly concerns itself with council housing stock, and so says nothing about large tracts of the city. In time this may change and recent editions of NOMAD have included more data from housing associations and other Registered Social Landlords (RSLs). Private house prices are included to provide context (but they do not affect the area demand scores), but only at a postcode sector geography - not coterminous with the SCA geography and too large to be harmonised reliably with the SCAs. At present, there are NOMAD datasets for 2001, 2002 and



(a) SCA 47.01 (b) Group 7

Figure 7.23: Trends for major offence types, 2000/01 to 2003/04.

2004, with a 2005 version in production at the time of this research. Although not a long timespan, interesting demand dynamics do appear in the dataset.

In the case of Hawksworth Wood (which for NOMAD purposes extends beyond SCA 47.01), a rating of 'On the Edge' has been given in all three years. The neighbourhood demand scores attributed by the model take into account a wide range of contextual variables including crime, education and benefits. By these overall scores Hawksworth Wood appears to have been getting worse, its score and ranking having risen period-on-period. Yet, looking at the individual variables for local authority (LA) housing, turnover in LA housing has fallen slightly from 13.8% to 13.0% between 2002 and 2004; applications per vacancy has risen from 13.5 to 17; empty homes as a percentage of LA stock has fallen from 3.41% to 1.0%. The only negative indicator is that the percentage of LA tenants in their homes for less than 2 years rose from 14.7% to 21.5%, although the 2004 figure needs to considered in light of the voids filled within the 2-year period and that the LA stock fell 5.0% from 703 homes.

7.5 Concluding Remarks

Overall, it is argued, the chapter has demonstrated the usefulness of the classification when applied in a manner following the principles of Exploratory Data Analysis (EDA). Support is also given to Sibley's argument that EDA "encourages and facilitates repeated references to the data and a cautious, sceptical attitude to theory" (Sibley, 1990, page 4). Furthermore, it is argued that some of the criticisms of geodemographics (e.g. Curry, 1995; Goss, 1995; Curry, 1998) might be avoided by probing rather than discarding residual neighbourhoods.

As an example of EDA, many aspects of the neighbourhood differences in criminal damage pose more questions than provide answers. The between-group comparisons were not explored at length, but they do prompt a review of the nature of the damage to vehicle data. The severity of the increases in damage to dwelling in the poorer groups is also highlighted by the analysis and has the potential to provide an initial guide for community safety practitioners who might be looking to target resources to tackle vandalism.

Arguably of more interest, however, is the extent to which problems vary within groups. A number of EDA techniques were demonstrated that can be used to identify residual neighbourhoods for further inspection. Following principles of the more recent development of Exploratory Spatial Data Analysis (ESDA), maps utilising the LCCS enable the analyst to consider the geography of

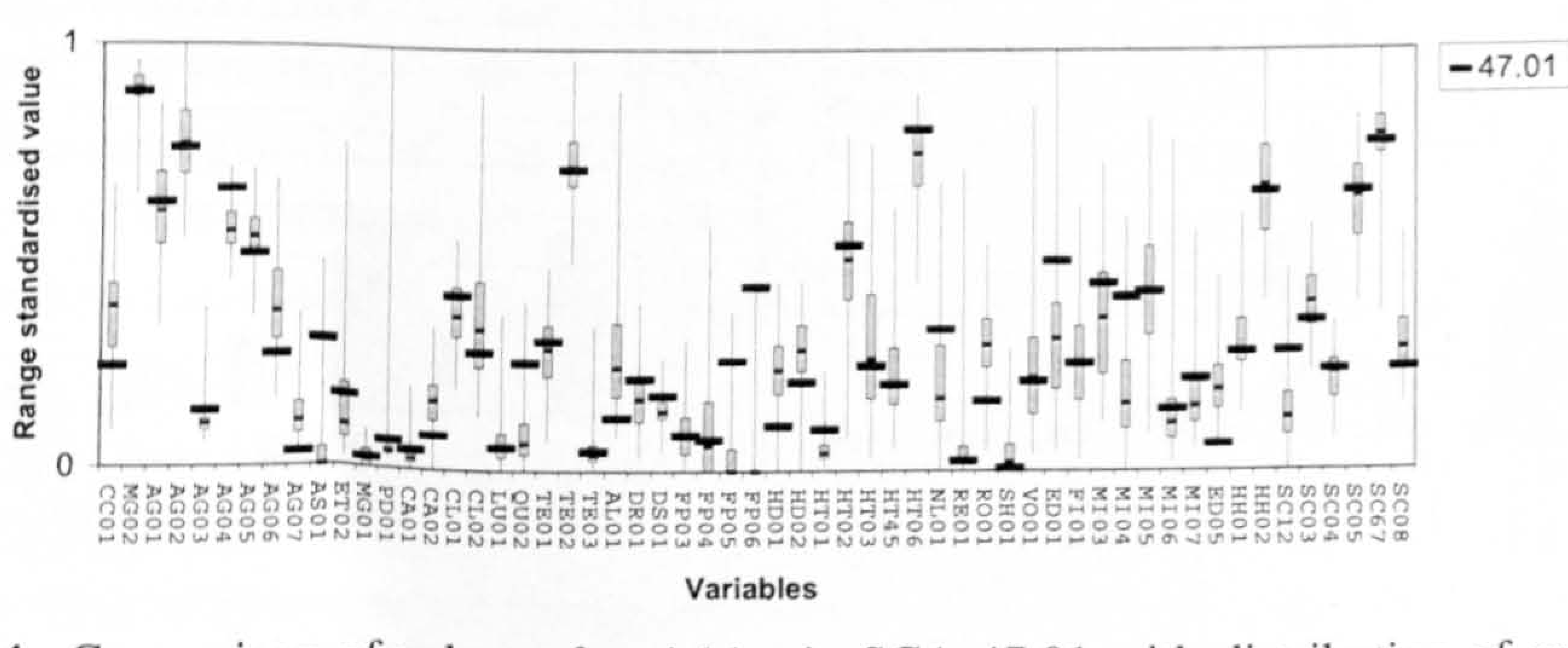


Figure 7.24: Comparison of values of variables in SCA 47.01 with distribution of values within cluster 7.

residuals in addition to their scale - as provided by the box plots and scatter plots.

With residuals identified, a range of lines of inquiry can be entered into in search of explanations for the atypical within-group crime rates. The analysis in this chapter demonstrates how to consider membership strength, unique neighbourhood features and geographical circumstances, and each of these can prompt new hypotheses that future research might test.

No attempt was made to investigate the possible effect of community safety initiatives on levels of criminal damage, principally due to a lack of suitable data about the location, duration and intensity of such projects. In the following chapter, however, the problem of analysing impacts is discussed, but with respect to levels of domestic burglary.

Chapter 8

Evaluating Policy: Burglary and BRIL

8.1 Introduction

While the previous case studies have tested the robustness of the classification in a variety of ways this final study investigates another important aim of the research; to establish whether variations in crime patterns in neighbourhoods of the same type could be used to assess the performance of a specific community safety initiative. The main hypothesis is that variations in patterns of crime across one group of neighbourhoods may be an indicator of the efficacy of a community safety intervention. Put another way, for a given number of similar neighbourhoods it might be reasonable to expect that variations in intensity and quality of delivery of an initiative would be reflected by variations in the crime level responses.

The Burglary Reduction Initiative in Leeds (BRIL) has been chosen to test these types of hypothesis. Within Leeds, reducing levels of domestic burglary has been one of the key aims of the LCSP in every community safety strategy since the Crime and Disorder Act 1998. To this end, BRIL has received considerable amounts of funding and has been able to reach out across the whole district to a large number of burglary victims, over a period of years.

The chapter begins by examining general trends in the burglary of residential properties (hereafter referred to as burglary dwelling, following Home Office terminology) and goes on to analyse various aspects of the geography of burglary dwelling in Leeds using the neighbourhood classification. With the scene set, the data from the BRIL is then introduced and an analysis of repeat victimisation is undertaken to explore the problems inherent in a study of this type and to determine whether the classification is able to make a contribution to the performance assessment of a complex community safety intervention.

8.2 Counting Rules and Rate Calculations

A 'burglary dwelling' is considered to have taken place if a house is entered without the occupier's consent, with the intention to steal. Derelict properties or properties that that have been disconnected from basic services are not considered as dwellings. A burglary is recorded irrespective of whether there has been a loss or not, and the Home Office recording rules have no codes to discriminate between these two situations, unlike the British Crime Survey, which does. Attempted burglaries are deemed to have occurred if damage to a door or window is more likely to have been done with the intention to burgle than to vandalise. A burglary of a shared house where the residents do not have separate lockable rooms is only counted as one offence - even though there may be several victims. Where residents do have separate lockable rooms then one offence is recorded per room broken into via the shared part of the property. It is sometimes suggested that this might inflate burglary dwelling rates in locations with high student populations, as this group often dwell in shared accommodation. Yet, evidence from UNIPOL, one of the main providers of student accommodation in Leeds, suggests that shared houses (not bedsits) with individually lockable rooms are the exception and not the norm (personal communication with UNIPOL, March 2003).

Burglary of outbuildings such as garages, sheds and green houses are classified as *burglary elsewhere* and not burglary dwelling. Likewise, burglaries of commercial and other non-dwellings are recorded as burglary elsewhere. The Home Office estimated that the introduction of the National Crime Recording Standard (NCRS) in April 2002 only inflated burglary dwelling across West Yorkshire by 1% in the following year (Simmons et al., 2003), compared to 24% in Essex, for example. The small scale of the NCRS effect was due to the existing West Yorkshire Police policy of recording every burglary dwelling offence once an incident had been logged. In Essex, the NCRS helped clarify the difference between attempted burglary and criminal damage, resulting in many more incidents being classified as the former.

It is worth noting that there is sometimes confusion when comparing burglary dwelling rates that have been calculated by different agencies within Leeds. The actual number of offences is not questioned - everyone accepts the counts provided by West Yorkshire Police. The (usually small) differences in burglary rates occur because different agencies use different sources to calculate the denominator, or at-risk population. Typically, this is the number of residential households. Household counts recorded by the decennial census are used most often but also in use are intercensal estimates and counts based on Council Tax records. Accordingly, calculations carried out for the 2004 Leeds Crime, Disorder and Drugs Audit (Leeds Community Safety, 2004) showed the burglary dwelling rate to be 44.5 offences per 1,000 households in 2003/04, based on the 2001 Census count of households (301,614), whereas the figure returned by Leeds City Council for purposes of Best Value Review (Leeds City Council, 2004) had the burglary dwelling rate at 43.5 offences per 1,000 households, based on an inflated household count for Leeds of 316,597, calculated from Council Tax records. The research presented here uses the 2001 Census household counts throughout.

8.3 Evaluation Rationale

The evaluation of the BRIL presented in this chapter is only partial, not comprehensive. This is partly because a full evaluation would be beyond the scope of this research, but also because such an evaluation has already been undertaken by the Policy Research Institute at Leeds Metropolitan University (Burden, 2005). There are some acknowledged omissions from Burden's (2005) work however, and the analysis in this chapter addresses some of these. The aim, therefore, is that the neighbourhood profiling and classification research will complement the existing evaluation, and thus further enhance our understanding of the BRIL and its outcomes.

Aside from some small technical problems, the existing evaluation could perhaps be criticised for being too concerned with internal validity and not considerate enough of external validity. The extent to which BRIL outcomes may vary in different settings, for different clients and at different times is given insufficient attention. As Crawford has argued, we should not only be concerned with 'what works', "but also with 'what works, for whom and under which conditions'?" (Crawford, 1998, page

197). Others have also spoken of this need to open up the "black box" of community crime prevention (Rosenbaum, 1988), and the analysis later in this chapter will try and uncover geographical variation in scheme delivery and outcomes.

Evidence to support the idea that neighbourhood type might influence BRIL outcomes can be found in a number of empirical studies. Maguire's (1982) examination of burglary in the Thames Valley is perhaps looking a bit dated now (do we still have coin operated gas and electricity meters?) but it discusses in some detail the differences in burglar behaviour between different types of neighbourhood. For example, Maguire found that burglary in local authority housing estates was more likely to be the work of locals, and these offenders were often known to the victim. Higher value properties on these estates were not victimised disproportionately and those offenders that were caught tended to be petty criminals rather than specialists. In the countryside, by contrast, apart from some offences being committed by local children, the burglars tended to 'professional' thieves with a preference for jewellery and silver. Detection rates for these rural offences were low and it was suspected that the burglars would hit a few targets in one area and then move on and not return for a considerable period. Patterns were harder to discern in middle-class Reading suburbs, but housing adjacent to poorer areas was more likely to be victimised. Silver and jewellery were the items of choice for burglars hitting affluent housing in urban areas and again these people tended to be burglary specialists from outside the area who were likely to have little or no involvement in other types of crime.

More evidence of the target selection tactics employed by specialist burglars was gleaned in a study involving the (incarcerated) burglars themselves (Bennett and Wright, 1984). When shown videotape of different types of neighbourhood the preference of the majority was for detached houses surrounded by cover (trees and bushes). Poorly maintained semi-detached council properties received very little interest. Similar results were found when a different sample of burglars were shown photographs of possible targets. In this case, the order of preference showed that detached houses were preferred most often, followed by semi-detached houses and then by terraced houses. The important thing to note is that the sample for this study comprised burglars who had been convicted for burglary. Between them they admitted to having committed a wide variety of other crimes but were deemed by Bennett and Wright to be 'burglary specialists', with the majority admitting to more than 50 burglaries in their lifetime.

These two studies seem to suggest that the Leeds neighbourhood classification might reveal some differences in burglar behaviour and the types of property being stolen, yet we must also be mindful of the possibility that there will be within-group variation as well. An influential study carried out in Sheffield in the late-1960s and early-1970s showed large differences between neighbourhoods of the same type, at least with regard to council housing estates (Baldwin and Bottoms, 1976; Bottoms and Xanthos, 1981). The latter account considered two such pre-war estates, one of which (council house high crime - CHH) had offender and offence rates three times higher than the other (council house low crime - CHL). Both populations were strikingly similar but the conclusion was that CHH had its own deviant subculture. It was suggested that attitudes consistent with criminality were being transmitted down the generations, and that obtaining goods by theft or from dubious sources had become something of a social norm. This norm was copied by children, rather than being taught explicitly, and the whole value system was sustained by strong family ties and high levels of social cohesion within the estate. The housing allocation policy was shown to reinforce this state of affairs, and "grading" and self-selection issues elsewhere have been shown to be important in determining why some estates seem to end up housing all the 'bad apples' (Wilson, 1963; Gill, 1977).

A small amount of more recent discussion about variation in burglary reduction scheme outcomes by neighbourhood type can be found in a group of studies that evaluated schemes funded under the Home Office's Reducing Burglary Initiative (1999-2002). Three comprehensive reports were produced by consortia that between them have evaluated 61 burglary reduction schemes.

All three consortia reported problems with delivering burglary reduction schemes in student neighbourhoods. In an alley-gating scheme in the south (Millie and Hough, 2004), keeping track of keys was difficult and gates were left propped open rendering useless the self-closing, self-locking mechanisms that had been provided to make it easier for local residents to use the gates. Elsewhere there were frequent problems in contacting and persuading often absentee landlords to invest in, or comply with, the desired dwelling security measures (Hope et al., 2004). In the north, it was a similar story, with delays occurring due to difficulties obtaining permission from absentee landlords and difficulty arranging appointments with students (Hirschfield, 2004).

Only Hirschfield gave further consideration to neighbourhood-type differences in scheme outcomes, although he was surprised by the results. Essentially, it appears as if the greatest impact upon burglary occurred in affluent areas and the four least successful schemes were in the most deprived areas. As he says, "the reasons for this are unclear, although, the more intractable problems experienced by residents in the most deprived communities, including lower levels of community organisation and cohesion, may have made successful outcomes more difficult to achieve" (Hirschfield, 2004, page 13).

8.4 Testing Classification Power

At the outset, it is important to recall that analysis in Chapter 7 showed that the neighbourhood classification only reduced the variability in burglary dwelling across Leeds (in 2003/04) by 20.2%. This is of some concern. More confidence in the results of analysis of burglary using the classification might be generated if more was known about differences in burglary dwelling patterns between and within groups, and this section explores these issues using confirmatory statistical techniques.

8.4.1 Difference in Burglary Rates

Crucially, what is needed is some confirmation that the burglary rates, and perhaps rates of change, are distinct between groups. Exploratory data analysis tools such as box plots of the different within-group distributions suggest to the eye that there are between-group differences, but it would be helpful if these could be backed up with more scientific measures.

A Kruskal-Wallis non-parametric hypothesis test was used for testing whether there is a significant difference between the neighbourhood groups with respect to burglary rates. A parametric hypothesis test might have been more powerful but the data fail to conform to the three principle requirements for parametric testing:

1. that the population from which the samples are taken is normally distributed;

2. that variances of the samples are equal; and

3. that observations are independent of one another.

The Kruskal-Wallis one-way analysis of variance by ranks (Kruskal and Wallis, 1952) has the advantage that it can compare three or more groups and is tolerant of different group sizes (Shaw and Wheeler, 1994). As the test operates on ordinal data, the burglary rates for all the neighbourhoods first have to be converted to ranks. The test uses the formula:

$$H_a = \left[\frac{12}{N(N+1)}\sum^{k}\frac{R^2}{n}\right] - 3(N+1) \tag{8.1}$$

Where R is the sum of the ranks in each group, N is the total number of observations, n is the number of observations in the group and k is the number of groups. One of the assumptions of the test is that the samples are taken from a continuous population, so if two or more neighbourhoods have the same burglary rate, and thus the same rank, a correction has to be made. This is done by dividing H_a by $1-\sum T/(N^3-N)$, where t is the number of tied ranks and $T=\sum t^3 - t$. Thus the statistic becomes,

$$H = \frac{H_a}{1-\sum T/(N^3-N)} \tag{8.2}$$

For burglary rate data from 2003/04, the number of observations $N = 477$, the number of groups $k = 10$ and the significance level $\alpha = 0.05$. The null hypothesis is:

$$\begin{aligned} H_0 \;=\;& \text{there is no difference between burglary rates in the neighbourhoods} \\ & \text{in the ten different neighbourhood groups} \end{aligned}$$

The degrees of freedom, $v = k-1 = 10-1 = 9$

$$H_a = \frac{12K}{N(N+1)} - 3(N+1) where \qquad K = \sum^{k}\frac{R^2}{n}$$

$$\begin{aligned} K &= \frac{R_1^2}{N_1}+\frac{R_2^2}{N_2}+\frac{R_3^2}{N_3}+\frac{R_4^2}{N_4}+\frac{R_5^2}{N_5}+\frac{R_6^2}{N_6}+\frac{R_7^2}{N_7}+\frac{R_8^2}{N_8}+\frac{R_9^2}{N_9}+\frac{R_{10}^2}{N_{10}} \\ K &= 30286587 \end{aligned}$$

$$\begin{aligned} H_a &= \frac{12(30286587)}{477(477+1)} - 3(477+1) \\ H_a &= 159.9889 \end{aligned}$$

There are 21 occasions where there is a tie between 2 burglary rates, i.e. $t = 2$. In each case, $T = (t^3 - t) = 6$, and so $\sum T = 126$.

$$
\begin{aligned}
H &= \frac{159.9889}{1-(126/(477^3-477))} \\
H &= 159.9891
\end{aligned}
$$

At the $\alpha = 0.05$ level, the critical value $\chi^2_{9,0.05} = 16.92$. This is less than the calculated value of 159.9 and therefore reject H_0 in favour of H_1: there is a difference between burglary rates in the neighbourhoods in the different neighbourhood groups.

This is an encouraging start, but unfortunately the result of the Kruskal-Wallis test does not give an indication as to which neigbourhood groups are most dissimilar or most alike in burglary dwelling rate terms. In order to analyse this detail, it is necessary to use a non-parametric test for two samples, such as the Mann-Whitney U test, for each pair of groups (that is 45 pairs for 10 groups). In common with the Kruskal-Wallis test, the Mann-Whitney (Mann and Whitney, 1947) is tolerant of unequal group sizes and operates on ordinal data, thus necessitating the conversion of the burglary rate values to ranks. In practical terms, the method for calculating the value of U depends on the sample size. For samples larger than 20, the distribution of U approximates to the normal so a z score can be calculated and used to determine whether to accept or reject the null hypothesis, rather than comparing U itself to a critical value from a lookup table. As the smallest group size is 20, the z score method is used here.

To begin with, the test statistic U is calculated thus:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{8.3}$$

where n_1 is the size of the smaller group, n_2 is the size of the larger group and R_1 is the sum of ranks in the smaller group. The z score can be calculated from U by:

$$\mu_U = \frac{n_1 n_2}{2}, \; \sigma_U = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}} \; and \; z = \frac{U-\mu_U}{\sigma_U} \tag{8.4}$$

At the $\alpha = 0.05$ level, the critical z value is 1.96. If the absolute value of the calculated z-score is greater than the critical value then H_0 is rejected. For the analysis of burglary rates in 2003/04, H_0 is:

H_0 = there is no difference between burglary rates in the neighbourhoods in the two groups

In the 45 different pairings of groups H_0 was accepted 9 times when comparisons were made using data from 2003/04 (Table 8.1). That is, there were 9 pairs of groups where there was no significant difference between burglary rates at the $\alpha = 0.05$ level. Group 3, which groups ethnic neighbourhoods, is shown not to have significantly different burglary rates from those in groups 1, 6, 7 and 8. This might suggest that group 3's most discriminating variable, ethnicity, is not important when considering burglary in Leeds. If similar results were obtained for other crime types then there might be value in downgrading the importance of ethnicity (perhaps by removing it altogether) in future versions of the classification. Another interesting result is that group 2, northern fringe suburbs and northern rural, is not different from groups 9 or 10 but was different from the other group with a considerable rural component, group 5.

For comparisons such as these, it is important to remember that the situation also changes over time. For those pairs of neighbourhoods where there was no significant difference in burglary dwelling in 2003/04, Table 8.1 also indicates the results of the same tests using data from previous years. A natural reaction to glancing at these statistics for the first time might be to question the usefulness of any classification when the crime situation is in such a state of flux. To an extent it is also a reminder of the need for timely neighbourhood data!

On the other hand, the ability to identify different crime trend trajectories between pairs of neighbourhoods is interesting in itself. Thinking again about the relationship between groups 1 and 3, for example, it appears that although the 2003/04 statistic cannot find them different, it is clear from the data in Table 8.1 that this has not aways been the case. Reference to Figure 8.10(a) shows that while burglary dwelling has fallen appreciably in group 3, it has failed to fall at the same rate in group 1. For now it looks as if their respective burglary rates have converged, but they may begin to diverge again if the trends continue in the same vein. The relationship between groups 6 and 8 is also interesting, as from 2000/01 to 2002/03 the burglary rates were significantly different - as would be expected given the very different profiles of these groups - and only a sharp drop in incidence in group 6 in 2003/04 has brought the two groups together.

	$\|z\|$ *score*			
Groups	**2003/04**	2002/03	2001/02	2000/01
1,3	**0.228**	0.754	1.736	2.364*
1,7	**0.163**	1.208	1.001	3.449*
2,9	**1.086**	1.164	0.541	1.320
2,10	**0.509**	0.283	1.209	1.523
3,6	**1.691**	0.744	2.988*	1.071
3,7	**0.729**	0.295	1.006	0.955
3,8	**1.528**	2.811*	4.589*	3.497*
6,8	**0.022**	3.438*	3.521*	3.366*
9,10	**1.913**	1.032	1.102	0.158

Table 8.1: Pairs of neighbourhood groups where there was no significant difference between burglary rates in 2003/04, $\alpha = 0.05$. * denotes significant difference, $\alpha = 0.05$, in previous years. Source: WYP.

It might be useful to develop a separate crime profiler to capture these dynamics, perhaps along the lines of the Leeds housing demand model, NOMAD. Certainly, to try and mix dynamic crime indicators in with much more stable and general socio-economic and demographic variables would be to lessen their impact somewhat. The presence of such crime dynamics adds justification for the decision not to include crime rate variables into the cluster analysis.

8.5 Geographies of Burglary Dwelling Offences

Now that it has been shown that there are statistically significant differences in burglary rates between groups, attention can shift to using the classification to identify interesting and useful geographies of burglary dwelling. While it is normal to concentrate on local patterns of offending, it is also important to looks at the current situation and past trends in Leeds alongside data from elsewhere.

8.5.1 Supra-district Trends in Burglary Dwelling

Across England and Wales between 2000/01 and 2003/04, the burglary dwelling rate remained fairly steady at between 18 and 20 offences per 1,000 households. Estimates by the Home Office suggested that the introduction of the NCRS in April 2002 would only have inflated recording by 3%, so it is possible that in 2002/03 there was actually a small drop in this type of crime. The British Crime Survey is usually able to provide useful supporting evidence for recording trends but unfortunately the BCS does not differentiate between burglary dwelling and burglary from garages, sheds and outhouses, which would be classed as burglary elsewhere.

As would be expected, at sub-England and Wales level there was greater variation in burglary dwelling rates over the same four year period. Figure 8.1 shows the trends in burglary dwelling within the Family 4 group of similar CDRPs (Harper et al., 2002). Differences in the impact of the NCRS between police forces (Simmons et al., 2003) makes it problematic to compare changes between 2001/02 and 2002/03 but changes between 2002/03 and 2003/04 ought to be comparable. Accordingly, Bradford, Leeds, Middlesborough and Sheffield all saw sizable drops in burglary dwelling rates, while places such as Birmingham and Nottingham saw rates stay the same.

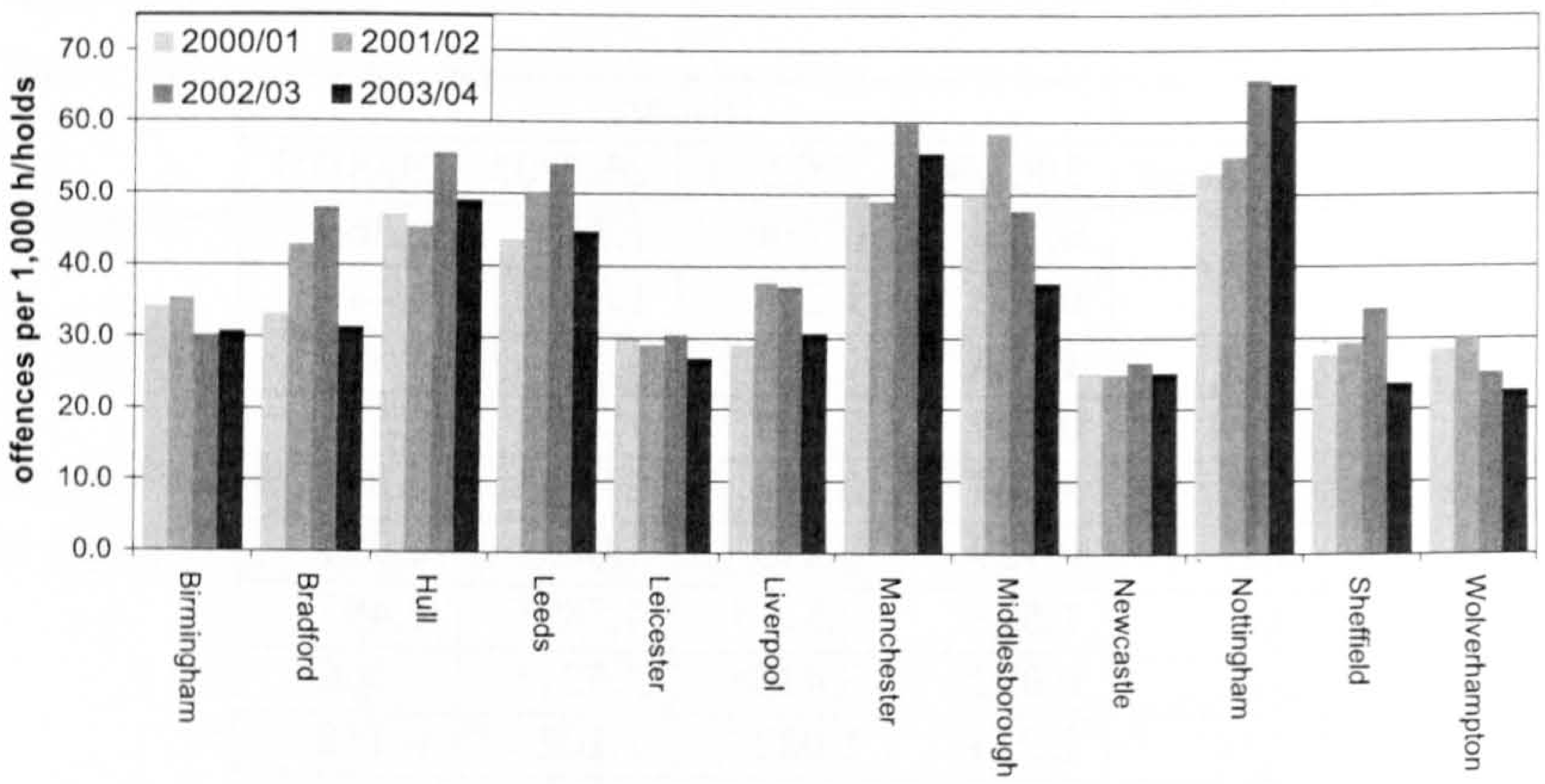


Figure 8.1: Comparison of burglary dwelling rates between members of CDRP 'Family 4'. Source: Home Office.

As has been mentioned previously, the intention of the Home Office was that the family groupings would "help local areas identify those in their family which have the lowest crime rates and, over time, are most successful at reducing crime, so that lessons can be learnt" (Harper et al., 2002, page 5). Leeds found itself looking to its peers in just such a way while undertaking the 2003 Best Value Review of community safety. The decision was made to base the review solely around the burglary dwelling problems in Leeds. As well as contacting and discussing issues with better performing family members, an analysis of census, employment, drugs, education and other datasets was undertaken to try and identify if there was anything about Leeds that made it particularly different from those family members with lower burglary rates (Shepherd, 2003). While some small (all they were ever likely to be) differences were identified, the analysis was really limited by not having access to small-area crime data for the other local authorities. Had such data been available it would have been possible to identify hotspots in other local authorities, focus on the conditions found therein and identify specific policy responses to those problems. This is very much what the current research project is aiming to do - to combine comprehensive local information with detailed crime data to investigate neighbourhood-level crime patterns within (and between) groupings of similar neighbourhoods.

8.5.2 Burglary Dwelling in Leeds

Within Leeds, conventional hotspot mapping using single kernel density estimation reveals those areas where burglaries are most concentrated (Figure 8.2). The student neighbourhoods of Hyde Park, Headingley and Woodhouse stand out most. Lesser hotspots can be identified in Armley, Bramley and Halton Moor. Generally speaking, the explanatory power of the map is poor. It is not clear whether the concentrations of offences are in neighbourhoods with high or low population density so it is not possible to determine whether the number of offences is unusual or not. A dual kernel density estimation technique would help account for population density effects. However, even maps that express crime rates tend to draw attention to only the worst affected neighbourhoods. There is no way of knowing whether there are neighbourhoods of a type similar to the hotspots which actually have much lower rates of burglary, or conversely, whether there are neighbourhoods that are of a type usually considered crime free where burglary is actually a problem, albeit perhaps not at the same intensity as in the high crime areas.

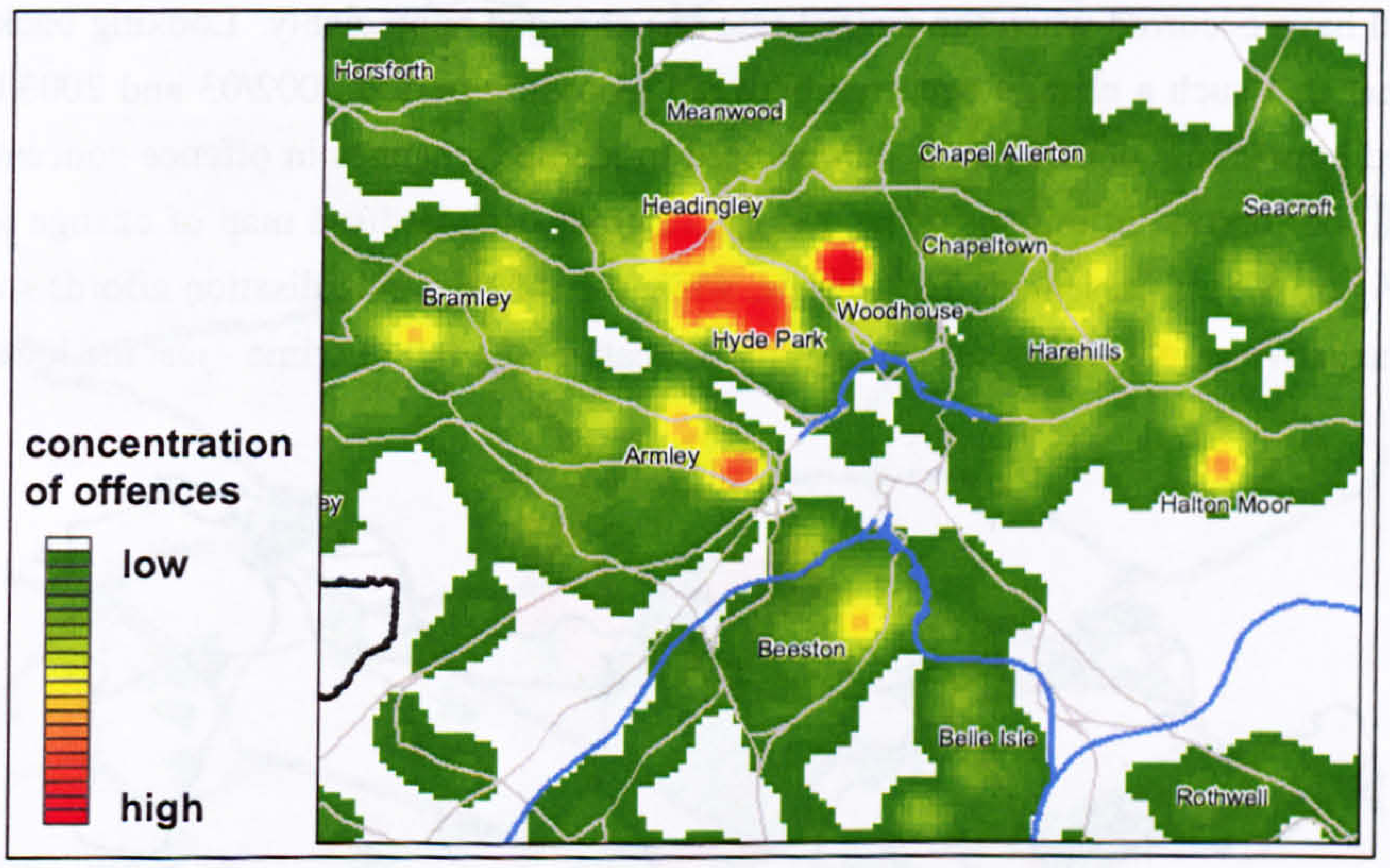


Figure 8.2: Kernel density hotspot map of burglary dwelling offences, 2003/04. Source: WYP.

The reader of the map is also bereft of information to help them understand the context of the crime patterns. To remedy this, use would need to be made of an open access tool such as ONS's Neighbourhood Statistics or a specialised tool designed along the lines of *The Profiler* (Hirschfield and Bowers, 1997a). The profile of an individual hotspot neighbourhood could then be ascertained and hypotheses formed about possible causes for a high volume of burglaries. It would still be very time consuming, however, to determine if there were similar neighbourhoods with fewer problems than might be expected or, whether there are problems for residents in conventionally middling and low crime areas. These are important issues for the planning and direction of burglary reduction policy and this is where a neighbourhood classification becomes useful.

Chapter 7 discussed how to range standardise crime rates by neighbourhood class and then present the results on a single choropleth map. One problem with this approach is that although the hottest and coldest-neighbourhoods within in each group are identifiable the group membership of each neighbourhood is not indicated. If the map is being studied within a GIS then this additional information can be accessed accurately and easily but if a hard copy of such a map (Figure 8.3) is being used, then the geography of hot and cold-neighbourhoods needs to be compared with a separate map of neighbourhood class membership (Figure 5.13). Another solution to this problem is to map the range standardised values for each group individually, producing ten separate figures. For brevity, examples for group 2 and group 4 are shown (Figure 8.4).

The difference in rates for neighbourhoods in group 2 show that burglary dwelling is most low in those more rurally isolated settlements and in Wetherby, while there are more problems for this type of neighbourhood in the northwest of the district, particularly in the area running from Leeds Bradford Airport, down through Calverley, and on to Bradford. Within group 4 there is also a distinct pattern to the burglary rates, with the highest in the group being found in Woodhouse and the St.Anne's area of Headingley. By contrast, problems in Hyde Park are relatively low as student neighbourhoods go, which is perhaps counter to what might be expected and is perhaps an indication of the success in disrupting offender behaviour in this area, with a consequent spatial displacement effect.

Although stability in the crime rate at a district level does not necessarily mean that individual neighbourhood rates will have stayed the same, it is more likely that significant neighbourhood changes will have occurred when the district rate has changed appreciably. Looking back at Figure 8.1, it is clear that such a change occurred between recording period 2002/03 and 2003/04. Chapter 8 showed how kernel density maps can be used to identify changes in offence concentration but this step will be forgone in this instance. Instead, a group-standardised map of change in burglary dwelling (Figure 8.5) is presented. To reiterate, while the range standardisation affords some neighbourhoods negative values this is not necessarily indicative of a fall in crime - just the location of the

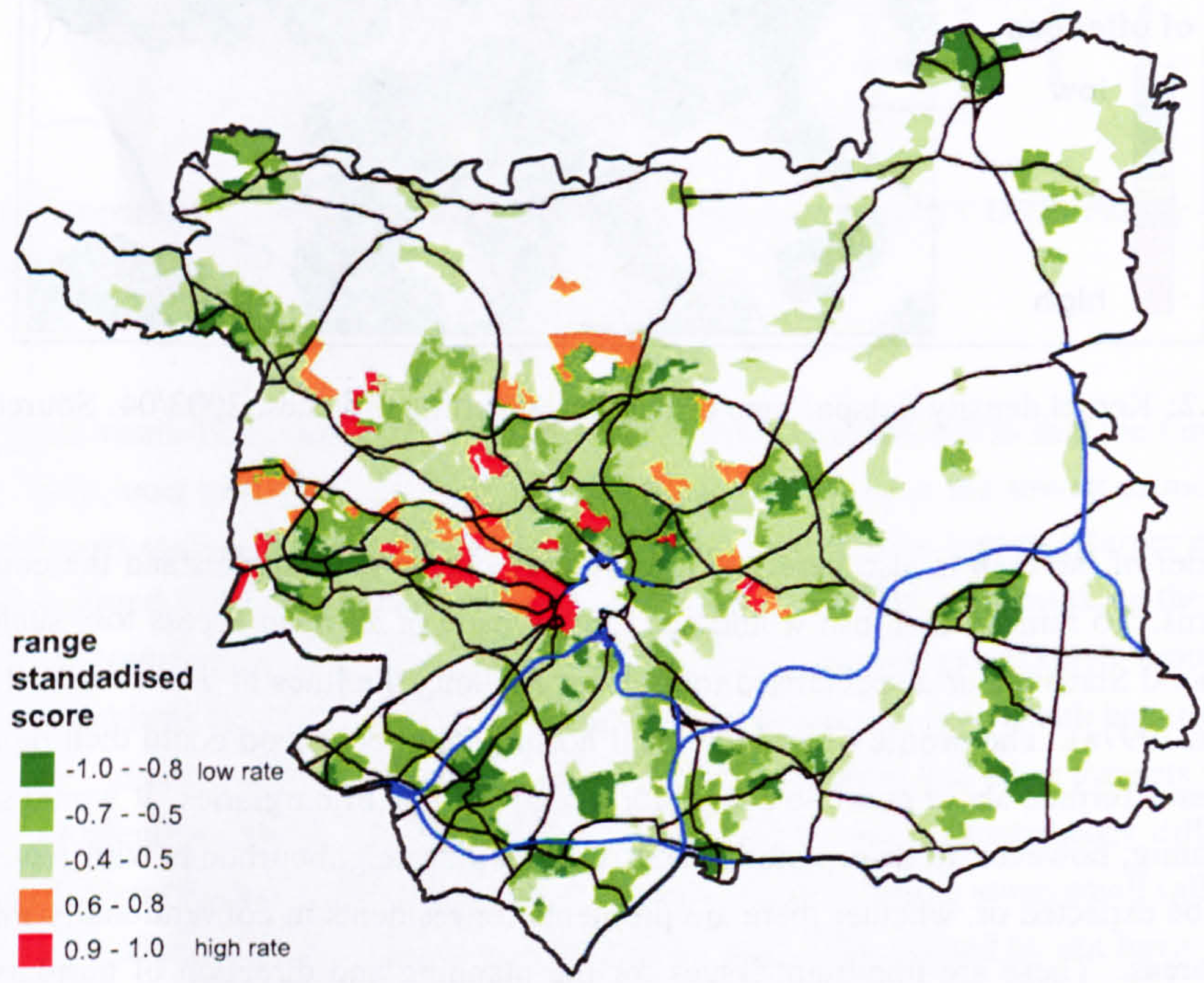
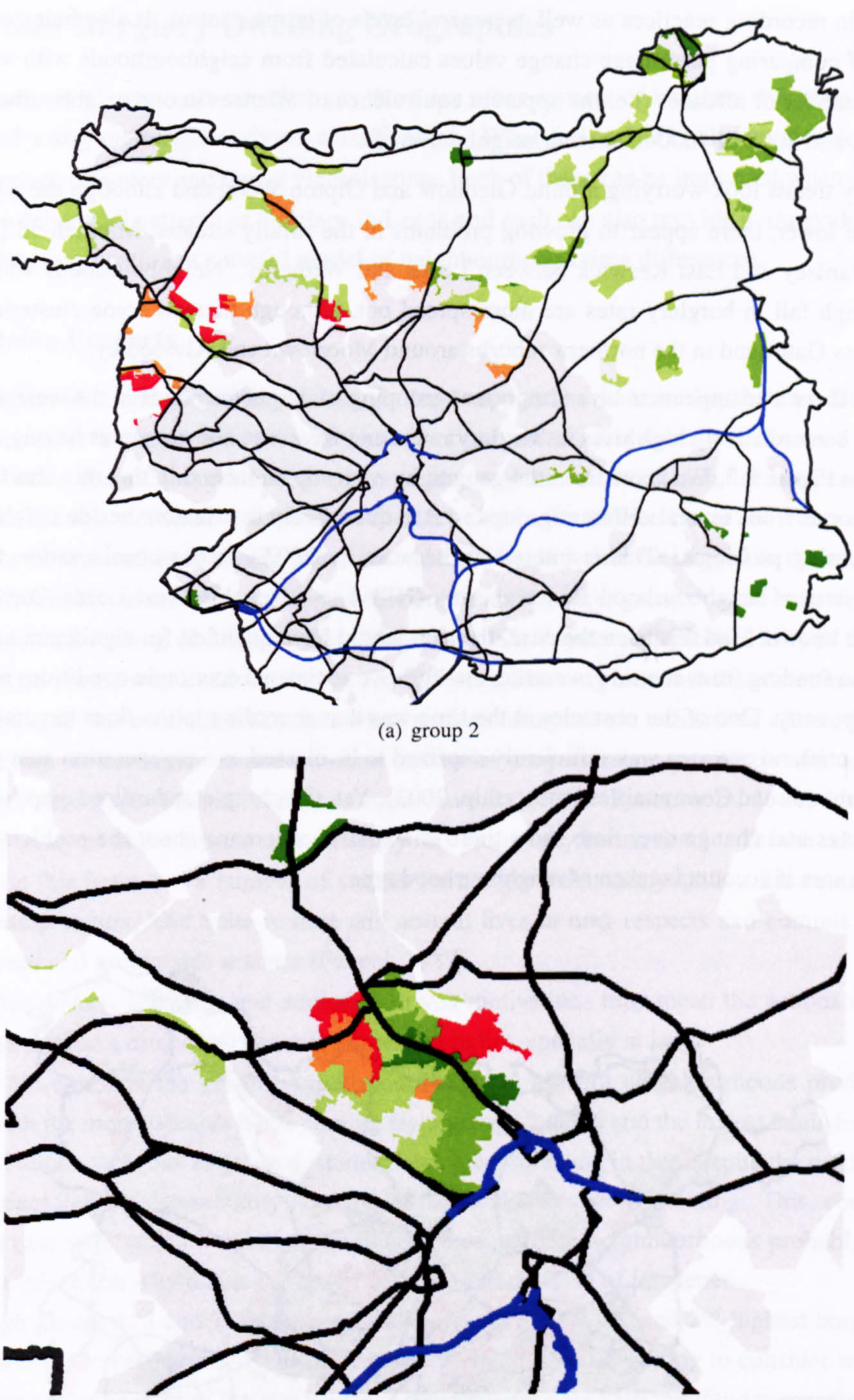


Figure 8.3: Map of burglary dwelling rates, 2003/04, standardised by group. Source: WYP.



(a) group 2

(b) group 4

Figure 8.4: Maps of standardised burglary dwelling rates, 2003/04, for individual groups. Source: WYP.

value in the within-group distribution. The change values themselves represent the extent to which a neighbourhood's share of offences within the district as a whole changed over time. Thus, if every neighbourhood was to see an identical *percentage* change the shares would stay the same and no change would be shown. The benefit of this method of calculation is that it helps to reduce the effect of changes in recording practices as well as general levels of crime control. It also helps reduce the problems of comparing percentage change values calculated from neighbourhoods with very small and larger number of offences (i.e. the apparent equivalence of offences in one neighbourhood rising from 2 to 4, and 2000 to 4000 in another neighbourhood).

Burglary trends look worrying around Gledhow and Gipton Wood and although the numbers of offences are lower, there appear to growing problems in the rurally situated Micklefield (in the far east) and Bardsey and East Keswick between Leeds and Wetherby. Neighbourhoods which saw a relatively high fall in burglary rates are more spread out, although there is some clustering in and around Cross Gates and in the northern suburbs around Moortown and Alwoodley.

Finally, there also appears to be an important grouping of neighbourhoods in the west of the city. These have seen relatively high rises in burglary rates, and have also showed up as having atypically high rates in Figure 8.3. We know from the geography of group membership that this area comprises neighbourhoods from several different groups so it is quite likely that factors beside neighbourhood type are causing problems. This is interesting because, in 2003, serious consideration was given to creating a new Neighbourhood Renewal Area (NRA) to cover this West Leeds Corridor, as it has become known. Had this been the case, the area would have qualified for significant amounts of regeneration funding from central government to improve social and economic conditions and reduce crime and poverty. One of the obstacles at the time was that according to the floor targets (of which burglary is one), no one area was sufficiently deprived to be classed as a regeneration area on its own (Neighbourhoods and Communities Partnership, 2003). Yet, the group-standardised maps of burglary dwelling rates and change over time show there is a spatial coherence about the problems across a fairly wide area if account is taken of neighbourhood type.

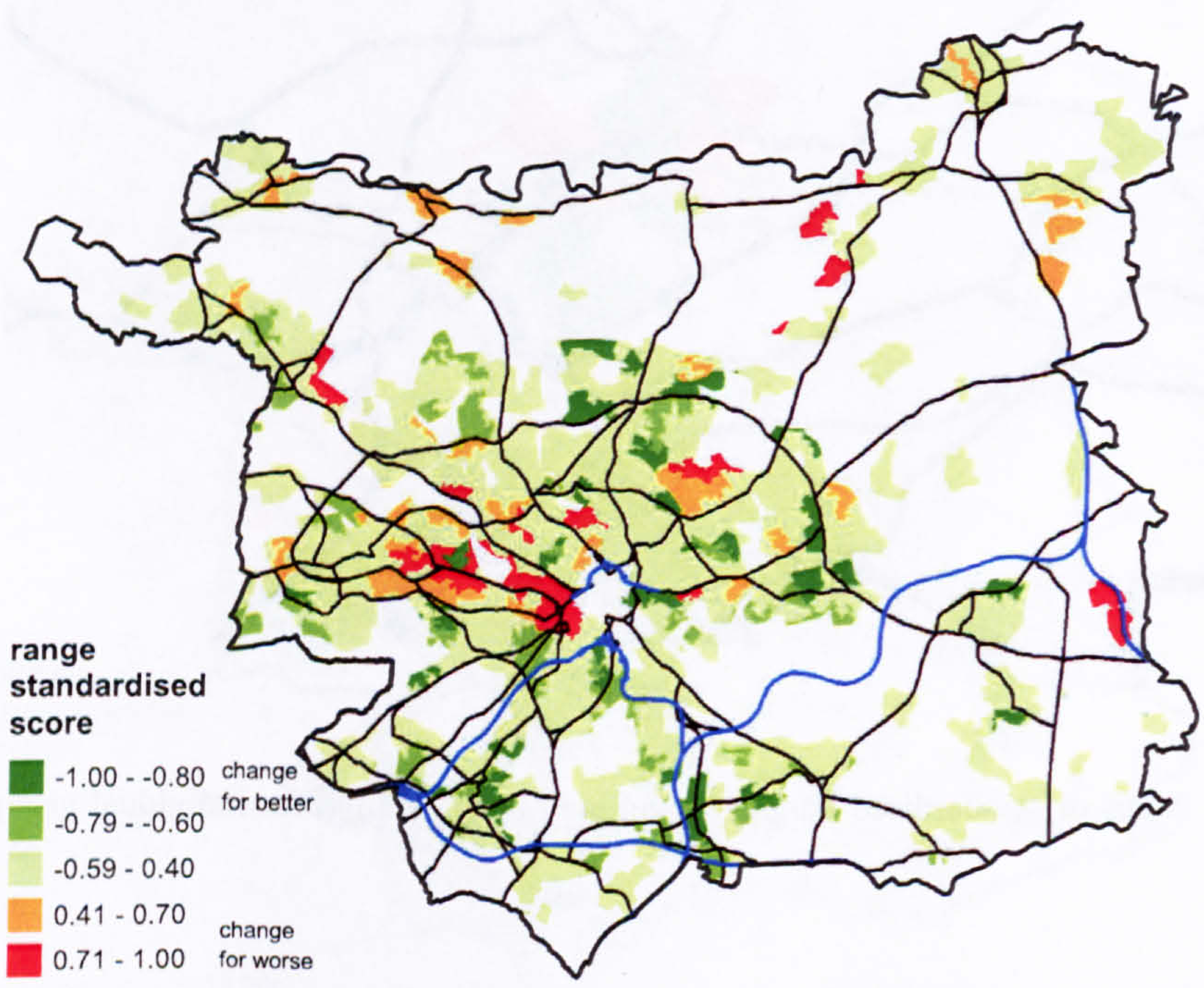


Figure 8.5: Map of change in burglary dwelling, 2002/03 to 2003/04, standardised by group. Source: WYP.

8.6 Other Burglary Dwelling Geographies

Besides analysis of rates and rate changes there are other aspects of burglary dwelling that can be investigated using the neighbourhood classification. What follow are analyses of stolen property, known burglary offenders and repeat victimisation. Each of these can be important when considering the reasons for spatial patterns of burglary in Leeds and each can also provide extra evidence for the case that the classification is a useful model of neighbourhood crime differences.

8.6.1 Stolen Property

Another aspect of burglary dwelling that might lend support for the case that the classification is an appropriate one is the type and value of goods that are stolen. A dataset was obtained from West Yorkshire Police that records the first five items stolen and the values of each, for burglaries reported in 2003/04. The value of items recovered is also included in the dataset, but is not analysed here.

It might be reasonable to expect that more valuable hauls would be bagged in the more affluent suburbs and that the values of goods from neighbourhoods in poorer groups would be lower. This had certainly been the finding of other studies (Maguire, 1982; Poyner and Webb, 1991) in the UK and there seems no reason why Leeds should be any different. The only caveat to this statement is that the previous research was conducted at a time when problem drug misuse was less of a issue (in the UK) than it is today. If the lifestyles and criminal careers of drug users are as chaotic and unstructured (the 'junkie' stereotype (Faupel, 1987)) as some have suggested (Rosenbaum, 1981, for example), then spatial analysis of the crimes committed by this group may be unreliable. Yet the scholarly jury is divided on this issue and a number of studies have argued that many addicts are actually skilled criminal entrepreneurs, lead quite routine and normal lives in may respects and commit crime in a well rehearsed and predictable manner (Faupel, 1987).

The often chaotic lifestyles and addiction-driven motivations may mean the actions of burglars who operate to fund a drug habit become harder to predict, spatially at least.

Figure 8.6 confirms the general hypothesis that more affluent neighbourhoods produce richer pickings, with the most valuable hauls coming from groups 2 and 6 and the lowest hauls from groups 1, 3 and 7. Neighbourhoods in group 4, studentland, are interesting in that despite the modern reality of high student debt this community do seem to have valuables worth stealing. This, coupled with the relative ease with which burglaries can be executed in these neighbourhoods probably provides part of the explanation why this is the group with the highest level of incidence.

Although groups 1, 3 and 7 (the poorest groups) have the 4^{th}, 3^{rd} and 2^{nd} highest burglary rates they also have the lowest returns in terms of property value. It is interesting to consider whether the high rates are a consequence of the low values. That is, do burglars have to offend more often because the value of their haul is typically low. And if this is the case, then at what point might burglars decide to seek better returns from more affluent neighbourhoods further afield? Similarly, what would be the consequences for these neighbourhoods should values of the items that are stolen generally become less. If the requirements of the burglar remained the same, then more offences would have to be committed, assuming it was preferable not to commute to other areas to offend. The consequent rise in burglary rates could also be compounded by an increase in the burglars' needs - a problem faced by some drug users, for example, as increased tolerance to a substance (especially heroin) can lead to the need for a higher intake to achieve the same 'high'.

Ideas and hypotheses such as these are difficult (or impossible) to test or substantiate using spatial models and statistics, but by looking at such data *through* the classification, attention is drawn to differences in offending patterns. These can serve as a reminder that some types of intervention (or inaction) might have unplanned consequences, both positive and negative.

More interesting patterns with consequences for crime prevention policy can be seen by looking at the specific types of property that are stolen. The graphs in Figure 8.7 each show the mean value of the item stolen, by group, along with the mean number of items stolen per offence. This second statistic is calculated by dividing the number of times an item is recorded as having been stolen by the total number of burglaries.

Looking at computers (Figure 8.7(a)) first, we do not see a lot of variation in the frequency with which these are taken, except in group 4 (studentland), where on average a computer is taken for every two burglaries. The value of these student computers is also high, being about equal to the value of computers taken from affluent rural and suburban neighbourhoods (groups 2 and 6). Correspondingly, the value of equipment taken from poorer groups is lower. The extent to which the frequency of computer theft in group 4 stands out might be a prompt to target students with special education campaigns on securing their equipment, perhaps even going so far as to sell property marking and locking kits in university shops.

The ubiquity of the mobile phone is highlighted in Figure 8.7(b), with little variation in the frequency with which they are stolen during a burglary, and little variation in their value. The situation with theft of cash is a little more varied, with the largest average amounts being taken from households in group 6 neighbourhoods. Bottom of the rankings in terms of frequency and value is group 4, the student group, as might be expected.

The taking of jewellery during a burglary, both in frequency and value terms, seems very much related to general levels of affluence. Jewellery is most likely to be stolen during burglaries of properties in the northern rural group (2), where it is also likely to be of the greatest value. Given the cultural significance of jewellery to people of South Asian origin, especially Hindus and Sikhs, it it perhaps surprising that jewellery hauls are so low from group 3. Then again, the relative poverty in these neighbourhoods may mean jewellery collections are very modest, or it may be that their investment

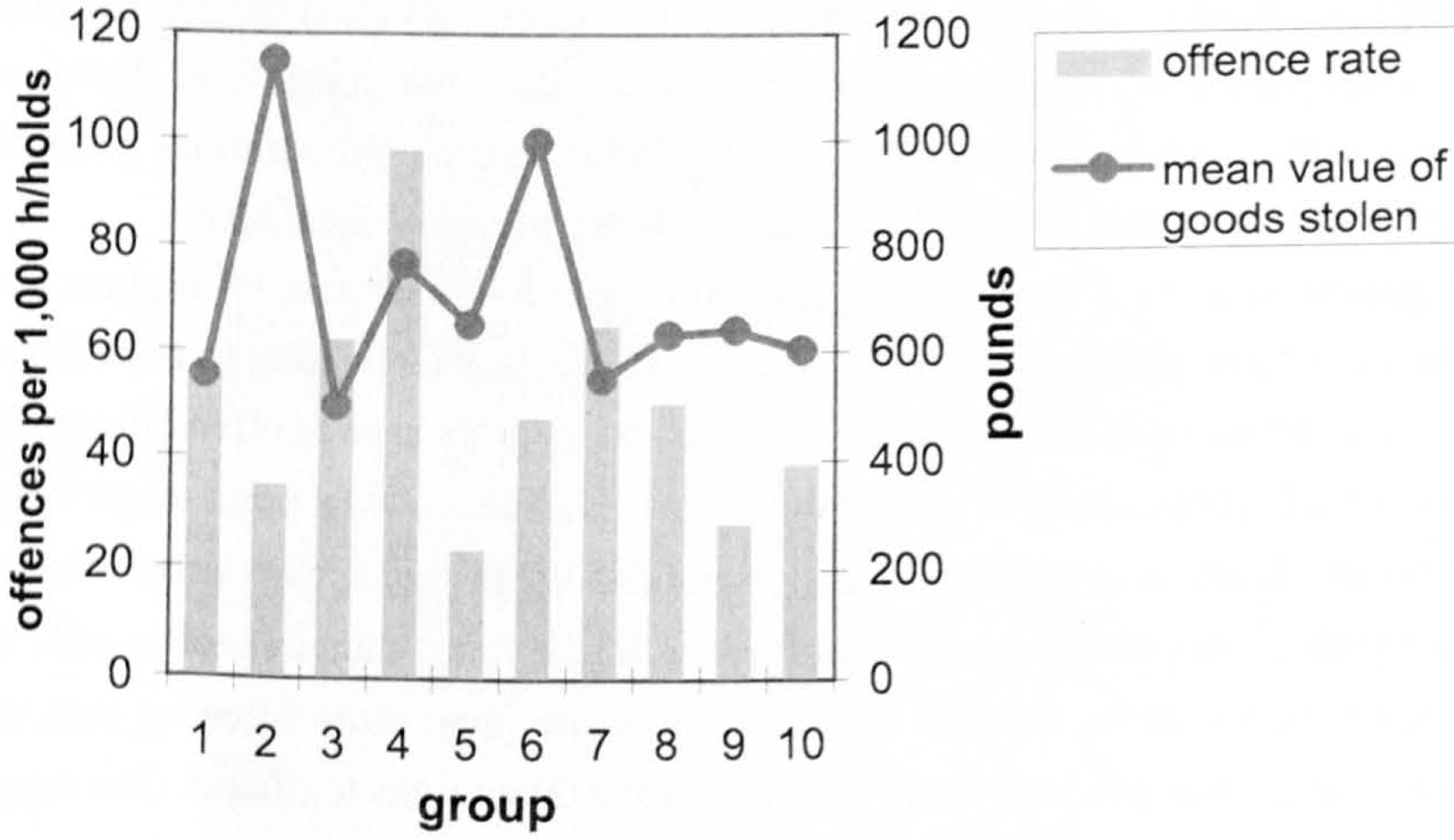


Figure 8.6: The mean value of property stolen during burglary dwelling offences in 2003/04, by group. Source: WYP.

and cultural importance drive the owners to take extra measures to secure these goods, either within their home or by depositing them with a bank. Furthermore, the Asian population in Leeds is largely Muslim, with Hindus and Sikhs only comprising 1.6% of the population, compared to Muslims who comprise 3.0% of the population.

Items on non-portable audio and visual equipment (Figure 8.7(e)) include televisions, hi-fi's, video and DVD players and recorders, etc. The average value of goods stolen does not vary a great deal, except for group 2, where values are typically 2 to 3 times higher than in other groups. In frequency terms these types of item are most likely to disappear from properties in neighbourhoods in group 7, the poorer council estates, followed by the other poorer groups (1,3 and 8), where this type of equipment might represent a significant proportion of the total value of goods in the property, and be very common. The situation with regard to portable items (Figure 8.7(f)) is somewhat different. Firstly, far fewer items are stolen during burglaries - their portable nature, however, does make them susceptible to other opportunities for theft. Secondly, the student neighbourhoods (group 4) lead the ranking in terms of frequency with which these items are stolen. Thirdly, there is not a lot of variability on the value of items stolen, and the second highest value occurring in the poorest group (1) is hard to explain.

In general, this breakdown of stolen property by group does seem to square with the profiles of the more prominent and distinct groups. This helps to further validate the authenticity of the classification and shows that the classification can distill interesting and policy-relevant information from a large community safety dataset.

8.6.2 Nominals

The details of the West Yorkshire Police 'nominals' database have been discussed in Chapter 4. In this section, the dataset is used to explore some of the spatial relationships between burglary offender and victim. To recap, the nominals dataset covers periods 2001/02 to 2003/04. Over this time there were 45,900 offences of burglary dwelling recorded, and of these, 4,349 (9.5%) have had one or more nominals associated with them. At group level, the proportions of offences with associated nominals varies from a low of 6.7% (218 offences) for burglaries in group 2, to 11.5% (561 offences) for burglaries in group 4. Admittedly, these samples are rather small, and they are likely to stay that way unless detection rates significantly improve. However, it is the only information of its sort that is available, and for an exploratory analysis of the sort below, it is defensible. Moreover, it is not without precedent (a similar approach was adopted for Maguire's study (1982)). It has not been possible to prepare the data to the same extent as Everson and Pease (2001), that is, the nominals dataset used here may include nominals who were arrested for an offence, but against whom further proceedings were not taken.

Table 8.2 breaks down nominal characteristics in a number of ways, beginning with a measure of the distribution of nominals across the neighbourhood groups. Also included are statistics which show in what percentage of burglaries with which a nominal has been associated (hereafter referred to as a burglary/nominal case) was the nominal living in the same neighbourhood (SCA) as their target and in what percentage of burglaries was the nominal living in the same group as the target. With this data it is possible to look for support for the colourful anecdote that burglars 'do not shit on their own doorstep' (Maguire, 1982, page 83), or maybe 'on their own type of doorstep'. Further support for this idea can be gleaned by looking at the mean distance between the target and nominals

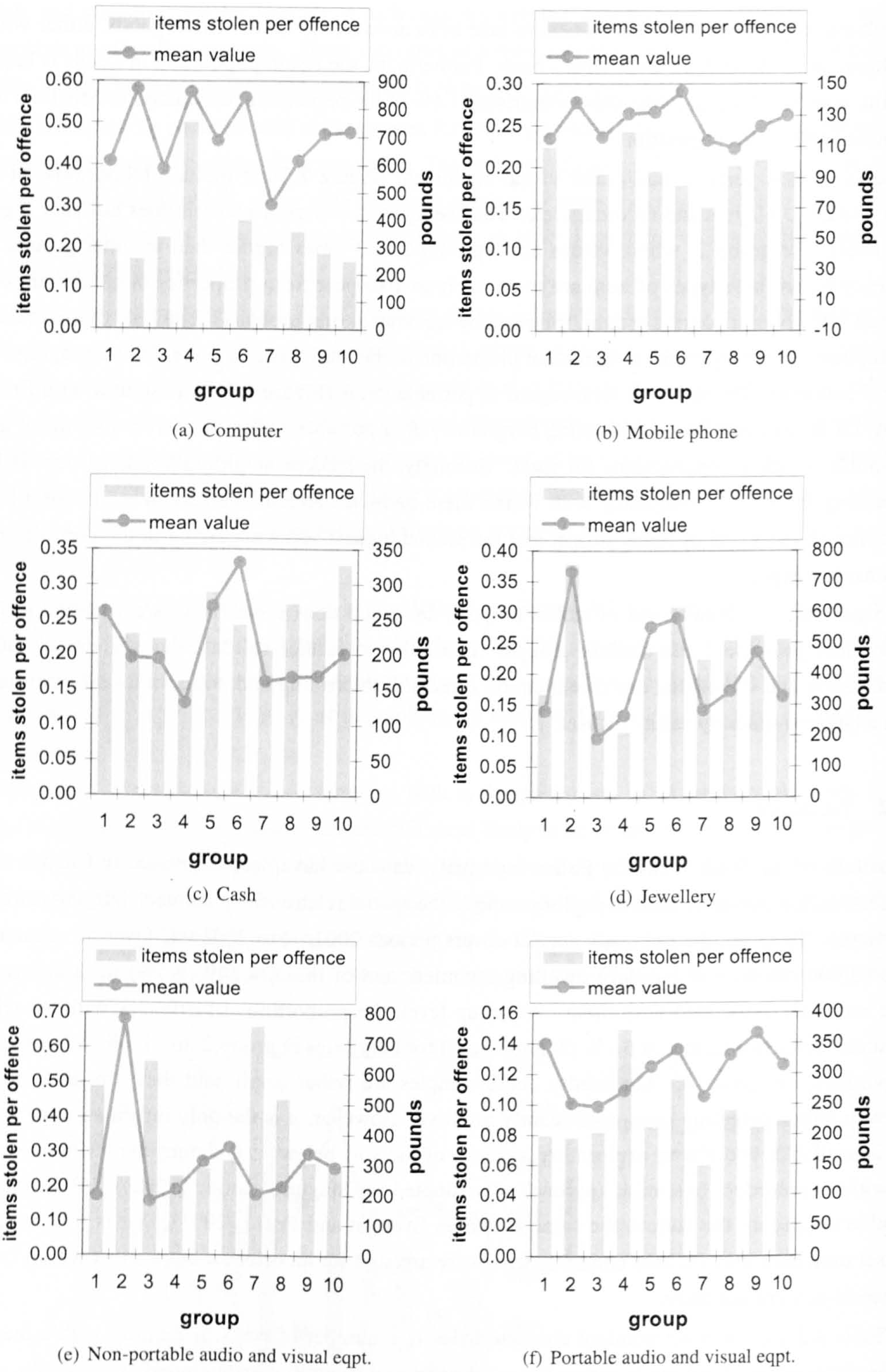


Figure 8.7: Frequency with which different types of property were stolen during burglary of dwellings during 2003/04, by group. Source: WYP.

	Group									
Nominals	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
mean offences / 1000 households	58.2	38.2	77.1	106.5	26.8	58.2	76.6	57.1	30.4	38.3
number of nominals resident	755	52	615	166	137	163	1419	648	249	419
% of total nominals	10.3	1.1	13.3	3.6	3.0	3.5	30.7	14.0	5.4	9.1
% offences where nominal resident in same SCA as target	19.6	6.0	21.4	3.6	9.3	1.1	37.1	20.7	5.5	9.4
% offences where nominal resident in same group as target	47.7	10.6	48.1	12.7	16.2	8.3	73.8	35.1	13.9	30.1
Mean commute (km) by offence	2.4	6.0	2.1	2.6	4.9	3.7	1.8	3.1	4.1	3.7
Median nominal age by offence	23	22	23	23	23	22	19	22	23	22

Table 8.2: Breakdown of nominals and burglaries to which they have been connected, by group, for period 2001/02 to 2003/04. Source: WYP.

	Group									
Offences	1	2	3	4	5	6	7	8	9	10
burglary dwelling	41.6	51.5	48.6	53.0	52.5	49.7	33.0	46.5	50.5	51.7
other acquisitive	29.2	29.2	23.7	25.9	26.2	24.5	33.7	26.3	27.9	29.2
criminal damage	5.5	2.3	4.5	3.5	4.1	4.0	7.2	6.3	3.5	3.6
drugs	3.0	2.8	3.1	3.1	2.0	3.2	3.5	3.4	2.6	1.9
other crime	6.5	6.8	4.7	4.1	6.3	7.3	6.2	4.8	8.1	6.6
violent crime	14.2	7.4	15.4	10.5	8.9	11.4	16.3	12.6	7.3	7.0

Table 8.3: Breakdown of offences (percentages) for nominals linked to burglaries in different groups, for period 2001/02 to 2003/04. Source: WYP.

home (mean commute). The median age of the nominals associated with offences in each group can provide further evidence of the type of people that are associated with burglaries. Alternative behavioural clues can also be found by considering the other types of offence that burglary nominals have been associated with (Table 8.3).

The most striking nominal profile is that for burglaries committed in neighbourhoods in group 7, the poorer social housing group. To begin with, 37.1% of nominals have been linked with burglaries in their own neigbourhood, the highest of any group by some margin. Moreover, at a group level, 73.8% of nominals associated with burglaries in these neighbourhoods came from similar neighbourhoods, i.e. neighbourhoods also in group 7. The local nature of buglary here is further emphasised by the lowest mean commute of just 1.8km. It has already been shown that the average value of goods stolen is low, so it seems unlikely that the nominals operating here are 'specialist' burglars. The median age of nominals is the lowest of any group (just 19 years) and burglaries of dwellings only account for 33.0% of all the offences to which the nominals have been linked, although 33.7% of their offences are for other forms of acquisitive crime. These facts would seem to back up the idea that the nominals operating in group 7 are similar in many respects to those identified by Maguire (1982) on council estates in Banbury. That was, local, often known to the victim, not professional burglars, but often petty offenders. It is perhaps only this last point that does not quite square with the picture from group 7, as 16.3% of offences to which these nominals were linked were violent (including violent crime, robbery, sexual assault and homicide).

Is is interesting to compare the nominals profile from group 7 with that for group 8. This group of neighbourhoods also includes many social housing estates but the group profiles from Chapter 6 reveal group 8 to have a slightly less deprived class structure and fewer social and economic problems. The percentage of nominals living in the same neighbourhood as their victim is certainly lower, at 20.7%. The figure for nominals living in the same group as their target is less reliable, as the group geography is more patchwork which would reduce opportunities for local within-group offending. However, the average commute distance does show that nominals are less local on average and they are also older. The offence breakdown is also different, with a greater propensity for specialising in burglary (46.5%), and slightly less other acquisitive crime (26.3%) and fewer violent offences (12.6%) between them.

The nominal profile for group 4, the student group, might be expected to be very distinct given it has the highest incidence of burglary, yet the offence breakdown for nominals active is not very different to those nominals identified as operating in group 8 and elsewhere. Conceivably, of course, some nominals will appear in all the profiles but where the number of nominals per group varies it ought to be safe to assume there are membership differences between the groups. Where group 4 does stand out as being different is in the percentage of nominals resident in the same neighbourhood as their victims (just 3.6%) and in the same group (just 12.7%). This would at least suggest that students are not preying on each other to any great extent. Surprisingly though, the relatively short mean commute of 2.0km would suggest nominals are mainly coming from close by and not from further afield as might be expected given the relatively rich pickings that are on offer.

The group which appears to suffer most from in-migration of nominals is group 6, the more affluent northern suburbs. Here only 1.1% of nominals live in the same neighbourhood as their victim, and only 8.3% live in a neighbourhood in the same group. Incidence here is fairly high (76.6 offences per 1,000 households). Something of the commute patterns to this group can be gleaned from mapping concentrations of nominals in-commuting to these neighbourhoods and overlaying the group 6 neighbourhood geography (Figure 8.8). Accepting that the nominals database does not provide information for all burglaries it nonetheless looks as if a significant portion of the group 6 burglary is being committed by nominals living in Harehills and Gipton. The standardised map of change in burglary incidence (Figure 8.5) also showed atyically large increases in this SE corner of group 6 and the adjacent concentration of nominals may be part of the explanation. There is also the issue of the nearby drugs market in Chapeltown. Although much reduced (according to police reports) as a result of intensive policing problem, drug users still associate this area more than others as the main place to buy crack and heroin in Leeds (COCA, 2004). Drug dealing areas can act as important anchor points for drug-dependent criminals (Rengert and Wasilchick, 1985) and as it is common to swap goods for drugs (Wiles and Costello, 2000), it might be expected that burglary close by drugs markets would be higher than in similar neighbourhoods elsewhere. Data from the Leeds Arrest Referral scheme from 2001/02 showed that about 20% of detainees agreeing to be interviewed ackowledged they resorted to burglary, among other crimes, to fund their addiction. A greater proportion (27%) resorted to shoplifitng (HEAT, 2002). In the analysis of crack habits produced for the 2004 Crime, Disorder and Drugs Audit, half the user respondents (n was only 16) said burglary was used to support their habit, and all of them said they resorted to shoplifting (COCA, 2004). Interestingly, results from the same report showed that 87% of drugs workers (police, youth workers, doctors, etc.) thought crime was used to fund drug abuse, but only 28% of users admitted this was the case. It is not clear whether the drugs workers are inflating the importance of crime or if users are unwilling to admit to it.

Finally, it is perhaps worth considering the other main neighbourhood type which Maguire (1982) had something distinctive to report on - the rural settlement. For Leeds this means examining nominals operating in groups 2 and 5, both of which have below average burglary dwelling rates. Both groups also have low proportions of nominals burgling in their own neighbourhood (6.0% for group 2 and 9.3% for group 6) and fairly low proportions of nominals coming from neighbourhoods of the same type (10.6% and 16.2% respectively). As would be expected, the mean commute is also higher than normal, all of which seems to signal that burglars tend to be outsiders. Exactly where they are sometime coming from, for group 2 at least, is shown in Figure 8.9. Nominals are again shown to cluster in Harehills and Gipton and from here they seem to roam quite far up the A58 corridor. A smaller concentration of nominals is highlighted in Holt Park and their range seems more restricted to local group 2 neigbourhoods.

There is also a reminder that Leeds is not an island and that many of the more affluent neighbourhoods in west Leeds will actually be in closer range of nominals resident in Bradford (particularly from north-western neighbourhoods such as Ravenscliffe and Greengates) than nominals from the centre of Leeds and further eastward. These nominals coming from outside Leeds are a problem for the analysis presented here as the areas in which they live have not been classified. The result is that the figures for the *% offences where nominal resident in same group as target* may be deflated, with between-group bias according to the group geography.

Maguire's work suggested that the rural communities tended to be preyed upon by specialist burglars in search of high value pickings such as jewellery and silverware. The previous analysis of stolen property shows that similar goods are sought in Leeds, especially in group 2, but an initial glance at Table 8.3 raises questions about the degree to which the nominals operating in these more rural groups are any more specialist that many of their urban counterparts. They appear less violent than some of those active in other groups, but other acquisitive crime types are still important to them. A further disaggregation of nominals into those that are resident in the group and those that are not shows that the group-locals are less specialist than the group-visitors and this is more noticeable in group 2 than in group 5. In neither group, however, does this have much impact upon the aggregate percentages, because the number of group-locals is relatively small.

This inability to find an obvious specialist burglar profile may be a failing of the classification or it may be that the nominals dataset is not up to the task. Then again, Maguire's stereotype of the specialist burglar was largely based upon supposition. At the time of his study, very few of the burglaries committed by this offender group were actually detected and suspects were identified as specialists according the sophistication of their MO and the nature of the goods stolen. With insufficient known individuals, it was not possible to suggest what percentage of these offenders' crimes were burglary and what percentage were for other offences. Although Bennett and Wright were working with incarcerated burglars they too did not quantify the offending histories of their specialists for crimes other than burglary. Instead, they just stated of their sample, "most admitted committing more burglaries than any other type of crime" which for the analysis presented above might mean any nominal whose offending profile included over 50% burglary dwelling. The bar was lowered even further - down to 43% - by Hearnden and Macgill (2004). Their study of burglar decision making does give an offence breakdown for their cohort of offenders and they claim their sample was made up of 'relatively experienced' burglars. In this study other types of theft accounted for 25% of offences.

Given the evidence from this last study, maybe some of the Leeds group-level nominal profiles do indicate that a specialist burglar population is operating. It is not possible to be sure.

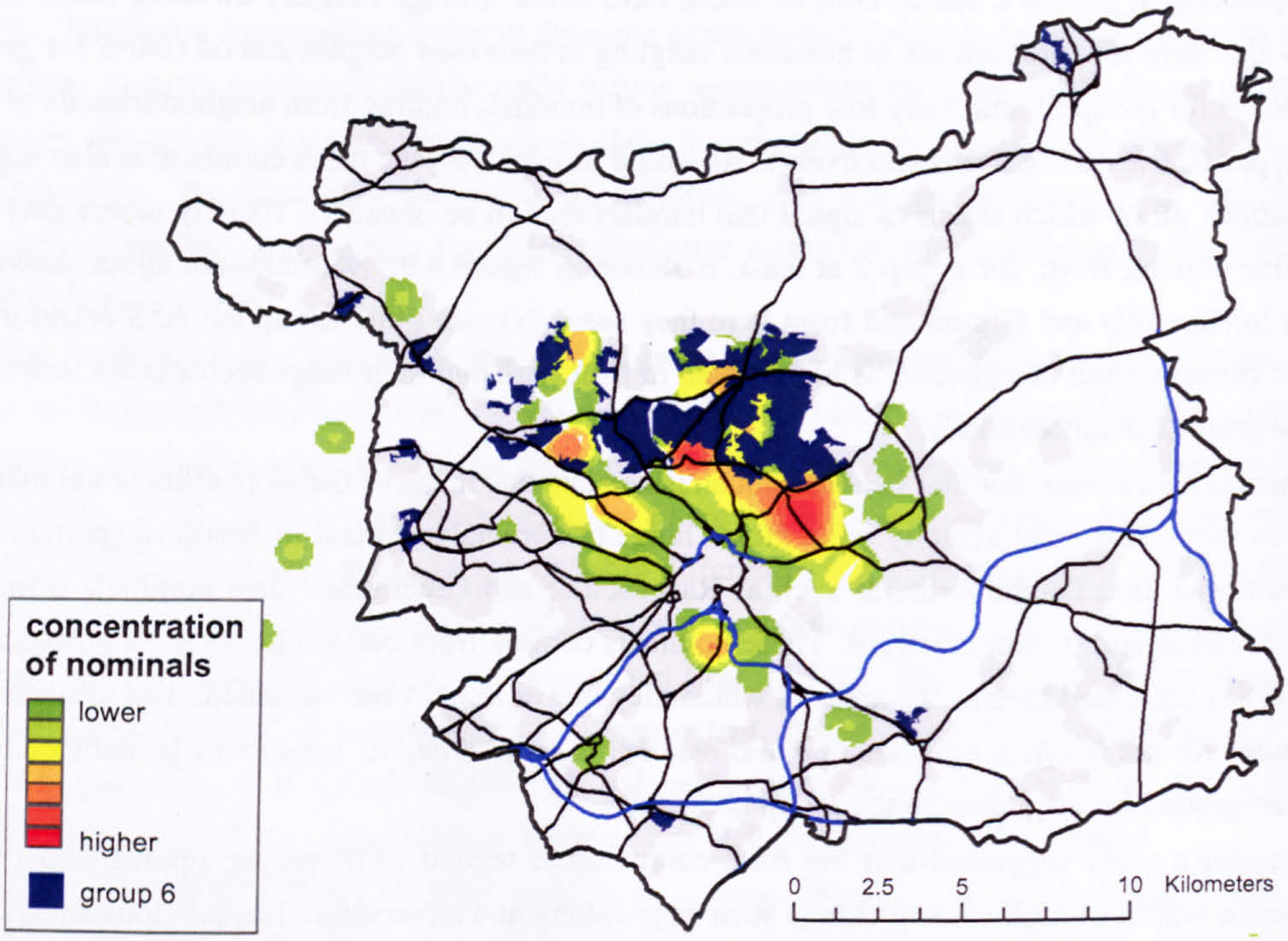


Figure 8.8: Concentration of unique nominals living outside group 6 who are associated with burglaries inside group 6 (n=292). Source: WYP.

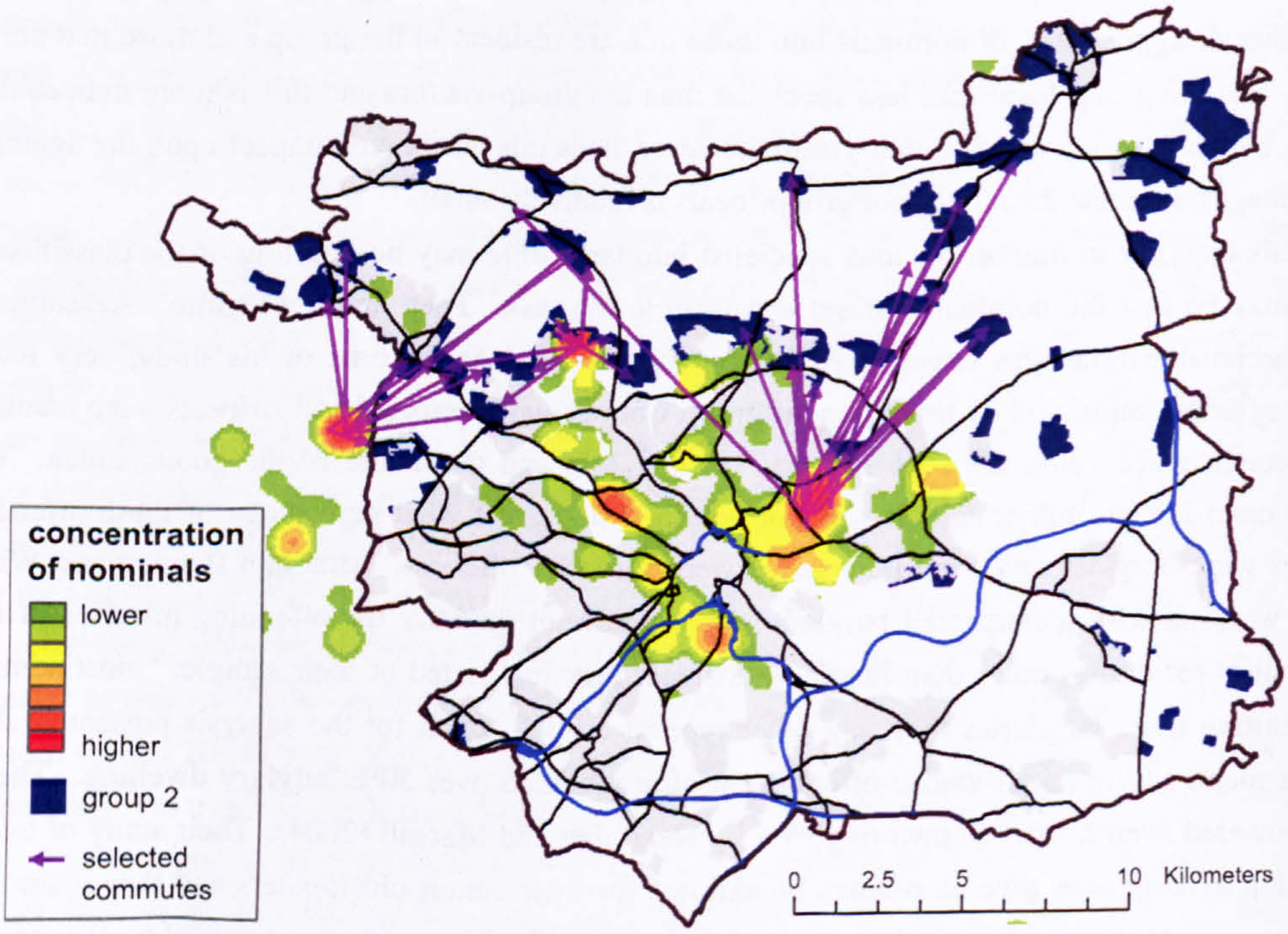


Figure 8.9: Concentration of unique nominals living outside group 2 who are associated with burglaries inside group 2 (n=142). Source: WYP.

8.6.3 Repeat Victimisation

Repeat victimisation is a term used to describe the recurrence of crime in the same places and/or against the same people. Expanding on this slightly, the Home Office state that repeat victimisation occurs "when the same person or place suffers from more than one incident over a specified period of time" (Bridgeman and Hobbs, 1997). While not usually given the same prominence as the headline crime rates (incidence) the importance of measures of repeat victimisation (prevalence and concentration) is self evident given some of the commonly cited assertions (Pease, 1998):

- victimisation is the best single predictor of victimisation;
- a major reason for repetition is that offenders take later advantage of opportunities which the first offence throws up; and
- those who repeatedly victimise the same target tend to be more established in crime careers than those who do not.

In the 1996 sweep of the British Crime Survey, 7% of burglary victims were victimised three or more times (Mirrlees-Black et al., 1996). Furthermore, attention grabbing figures such as *4% of people suffer 44% of crime* (Farrell and Pease, 1994), or *2% of the people who suffer most property crime (excluding vehicle) suffer 41% of all such crime in the BCS 1982-1992* (Pease, 1998) lend significant support to the assertion that victimisation is a very good predictor of subsequent victimisation.

Because of this importance, reducing repeat victimisation is a key policy in Leeds for reducing burglary dwelling incidence. The *Leeds Crime, Disorder and Drugs Audit 2004* (Leeds Community Safety, 2004, page 2) attempts to show the success of the policy by stating that in 2002/03 the percentage of burglary dwelling offences that were repeats was 14.5%, falling to 13.3% by August 2004. Futhermore, the *Audit* states that where properties had been target hardened under the BRIL initiative repeat victimisation was less than 1%, although the exact nature of this measurement is not made clear.

This section investigates whether the geography of repeat victimisation conforms in any way to the neighbourhood classification, but first some further analysis of district level trends is presented to provide more context than that provided by the crime audit. Presuming that the audit figures are correct, undertaking an initial analysis of this sort will also serve to verify the methodology employed here for identifying repeats.

The methodology is straightforward in principle. A burglary is deemed to have been a repeat if another burglary had occurred at the same property within the previous 12 months. As 4 years worth of recorded burglary dwelling crime data are available it is possible to calculate repeat victimisation statistics for periods 2001/02, 2002/03 and 2003/04. The analysis treats occasions where two or more burglaries occurred at one property on the same day as one burglary.

Difficulties in accurately identifying repeats arise because we have to rely on accurate and consistent recording of property addresses. These difficulties are greatest when working with flats, Homes in Multiple Occupation (HMOs) and university halls of residence, and an amount of manual data cleansing has to be undertaken to bring consistency to flat naming conventions, for example. Records with insufficient address information to identify a unique property were excluded from the analysis.

The final procedure employed to identifying repeats firstly involves ordering a table of all offences by address and date. A copy of this table is then made. Each record in each table (the original and the copy) is given a row number (integer), the sequencing of the row numbers for one table being offset by a value of 1. A simple SQL query then joins the tables by row number and identifies a repeat if the addresses are the same and the date in one table is within 12 months of the date in the other table.

To begin with, it is worth stating that headline grabbing statistics like those produced by Pease (Pease, 1998) can be misleading if their audience forgets that they refer to a basket of crime types which might befall a property or person. Such statistics represent summative victimisation. Focusing on burglary dwelling alone produces a very different set of statistics (Table 8.4). For both the BCS and Leeds, the data represent averages of the individual years within the time period. Mirrlees-Black's (1996) finding from the BCS in 1996 that 7% of burglary victims were victimised three or more times seems high compared to the average of 4.5% in Leeds between 2000/01 and 2003/04. Yet, if a more recent set of comparable results (2001/02 to 2003/04) from the BCS (Simmons and Dodd, 2003; Dodd et al., 2004) are considered then the mean national figure of 5.9% is closer to the 4.5% in Leeds.

	Burglary dwelling in Leeds 2000/01 to 2003/04		BCS Burglary 2000/01 to 2003/04	
Number of victimisations	Proportion of households	Proportion of events	Proportion of households	Proportion of events
0	95.6	0	96.6	0
1	4.0	82.2	n.a.	83.6
2	0.3	13.3	n.a.	10.6
3+	0.12	4.5	n.a.	5.9

Table 8.4: Comparison of percentages of offences by number of victimisations. Data marked n.a. is not available. Source: WYP and Home Office.

Although interesting, comparisons of this sort are of limited utility. What is required to build a picture of how burglary dwelling has changed in Leeds in recent years is a more detailed exposition of repeat victimisation. In particular, it is important to establish if there are relationships between the incidence, prevalence and concentration of burglary dwelling. This can be done at a district level to show the general nature of changes, while a repeat of the analysis at a neigbourhood class level should determine the extent to which positive and negative change has being experienced equitably across Leeds.

	Incidence	Prevalence	Concentration	% Repeats
Period	Offences per 1000 h/holds	% of h/holds victimised	Average number of times victims burgled	Using 12 month sliding window
2000/01	45.1	4.00	1.115	
2001/02	52.0	4.57	1.119	17.3
2002/03	54.3	4.86	1.107	15.4
2003/04	44.5	4.16	1.095	15.5

Table 8.5: Trends in repeat victimisation in Leeds. Source: WYP.

The district level picture (Table 8.5) shows that in 2000/01 45.1 households per 1,000 were being burgled. This represented 4.0% of all households and on average each of the burgled households was victimised 1.115 times. The following year, 2001/02, the incidence of burglary dwelling rose appreciably. These extra offences were accounted for by a large increase in the number of different properties being burgled, while the concentration stayed about the same. One reason for this might have been that the offender population grew. In such a situation the new offenders would not have been in a position to exploit the benefits of revisiting a previous target. An alternative explanation for the steady concentration rate is that victims' ability to improve the security of their homes - either by their own efforts or because of an intervention such as the BRIL - so burglars were less able to effect a repeat and had to target properties new to them.

In 2002/03 the incidence of burglary rose further. This time however, the prevalence had to rise even higher because the concentration rate fell from 1.119 to 1.107. In short, the size of the burglary problem overall had increased but victims were being repeat targeted less often. Finally, in 2003/04 the incidence of burglary dwelling fell to below the old 2000/01 level. Furthermore, concentration had fallen to a four year low of 1.095 burglaries per victim. This suggests, perhaps, that BRIL and general levels of public awareness of risk and risk reduction were having an impact. Prevalence, however, was slightly higher than the 2000/01 level showing that in 2003/04 the problem of burglary dwelling was being shared slightly more equitably across the district as whole.

Also included in Table 8.5 are figures that describe the proportion of offences that have been calculated to have been repeats. That is, they followed a previous burglary within 12 months. Although the trends look similar, this measure is different from concentration in that only the repeat offence is counted, the first offence in the repeat sequence is not part of the rate calculation. Because of this lag effect, it would be possible in a given period to have a concentration of 1.0, or parity, but a repeat victimisation percentage greater than 0%.

These trends in incidence, prevalence and concentration can be analysed further by disaggregating to individual groups (Figure 8.10). The same can be done with the repeat victimisation rate (Figure 8.11). At first, the sea of bars looks rather overwhelming, but on closer inspection a number of different types of pattern can be discerned.

The student group (4), for example, is the only group where on average the incidence rates were higher in 2003/04 than in 2002/03. At the same time concentration fell and prevalence rose. This suggests that although burglary rose, the problem was being shared out more equitably. It also indicates that although repeat victimisation is falling (also evident in Figure 8.11) it is insufficient to stop burglary from rising, suggesting that policy responses in these types of neighbourhood should either not be purely aimed at reducing repeat victimisation, or if they are to be, then the quality of the interventions needs to be improved.

A different type of situation can be seen in both group 2 and group 9. Here, although there are trends in incidence and prevalence levels, the concentration level is staying static. This would suggest that repeat victimisation is less of a feature of burglary dwelling in these neighbourhoods. There might be a number of reasons why this is the case, including the nature of the response by householders on being burgled. That is, they may be more motivated and better able, economically, to redouble their security and thus repel burglars who attempt to revisit. It might also be that burglars select their target in a different way in these types of neighbourhood.

Another type of pattern can be seen for group 3, the ethnically diverse neighbourhoods. Here, the scale of the drop in incidence and prevalence between 2002/03 and 2003/04 was not matched by the scale of drop in concentration, which was more modest. This might suggest that although a reduction repeat victimisation may be accounting for some of the drop in incidence there are other factors at work in these neighbourhoods also.

Finally, the changes between 2001/02 and 2002/03 in group 7, the poorer council estates, is worth scrutinising. In most groups, when there is a drop in concentration but an increase or no change in

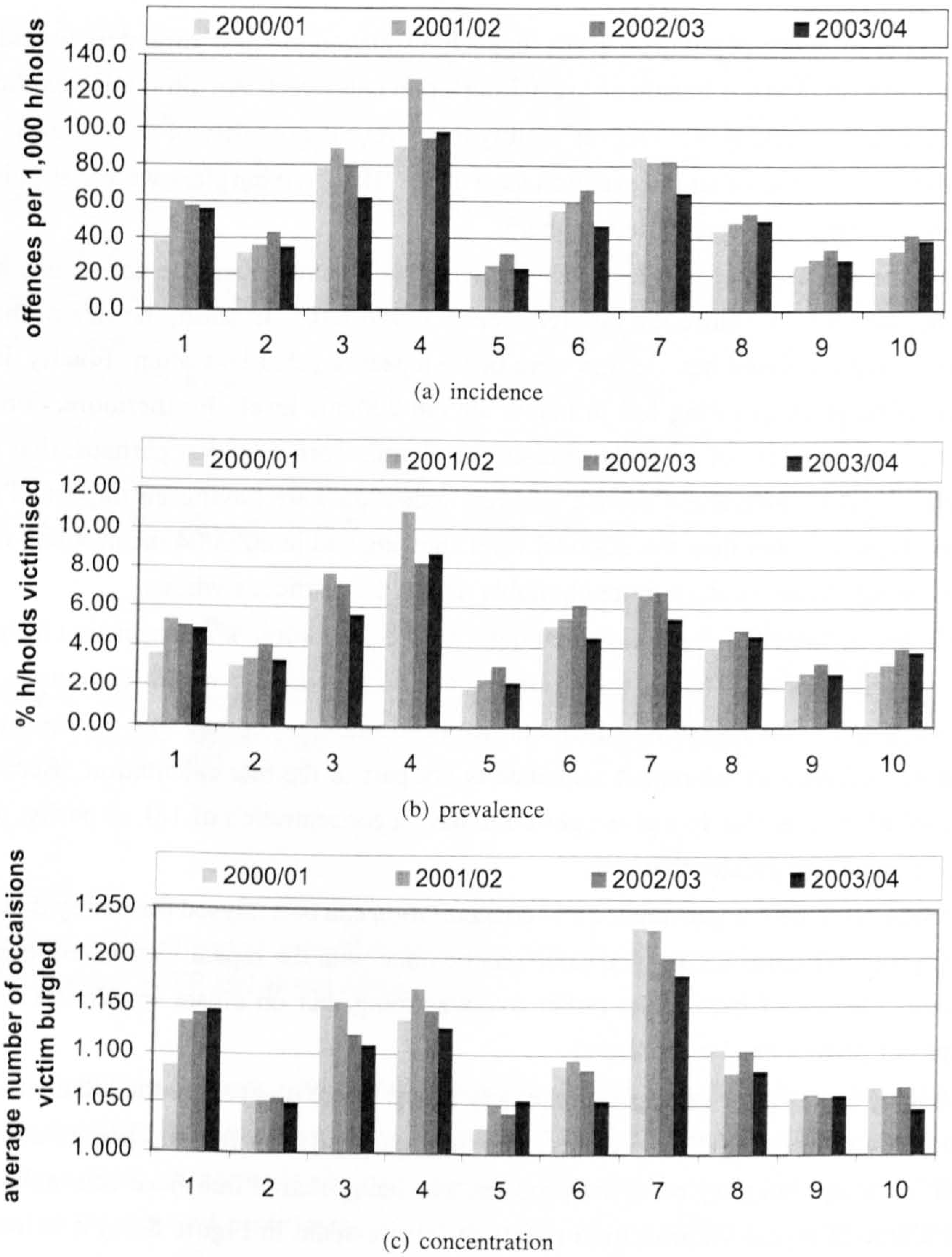


Figure 8.10: Trends in burglary dwelling victimisation by group, 2000/01 to 2003/04. Source: WYP.

incidence, a rise in prevalence is also observed. In group 7, however, this does to seem to be the case. Instead there appears to be no change in the number of offences, no change in the proportion of households being victimised, but people are being victimised less often? The reason for this becomes clear when consideration is given to statistics that reflect the dispersion of values among the neighbourhoods, rather than the mean value alone. For example, the standard deviation in the concentration of victimisation was 0.614 in 2001/02, while in 2002/03 there was a standard deviation of 0.551. As the steady prevalence already shows, the percentage of households being burgled has not changed. What has changed is the distribution of repeat victimisation levels at a household level, with fewer households suffering high numbers of repeats and the difference being born by lower-frequency repeat households rather than previously unvictimised properties.

8.7 Impact of Burglary Reduction Initiatives

It was an important original aim of the research to see whether unusual and atypical victimisation patterns of the type identified above might be explained by the location and intensity of community safety initiatives. If this were the case then it would support, for example, the continued use of the classification to match control and experiment neighbourhoods in outcome evaluations employing the classic experimental design.

To expand on this exemplar, this basic design of comparing pre-test (O_1) and post-test (O_2) conditions for an experiment group receiving the initiative (X) and a control group not receiving it has been a cornerstone of social policy evaluations and was famously described by Campbell (1963) with his OXO notation (Figure 8.12).

The theory of causation underpinning the design is that since the experimental and control groups are identical to begin with, the only difference between them is the application of the initiative and it is, therefore, only the initiative which can be responsible for the outcomes. The basic idea is that the causation itself cannot be observed. Instead, causation between initiative and outcome is inferred from the repeated succession of one such event by another. To work, therefore, every conceivable rival causal agent must be excluded from the experiment so that just the one causal link, the initiative, remains.

It is usual to acknowledge that controlling for all possible causal agents in studies of social programmes and initiatives is almost impossible. The social world is just too complex and dynamic and so investigators refer to their studies as 'quasi-experimental'. In the case of a burglary reduction initiative, for example, besides this intervention there might be other crime prevention schemes in progress, such as Neighbourhood Watch or heightened patrolling by police Community Support Officers. There may also be other, more general renewal and regeneration initiatives being delivered

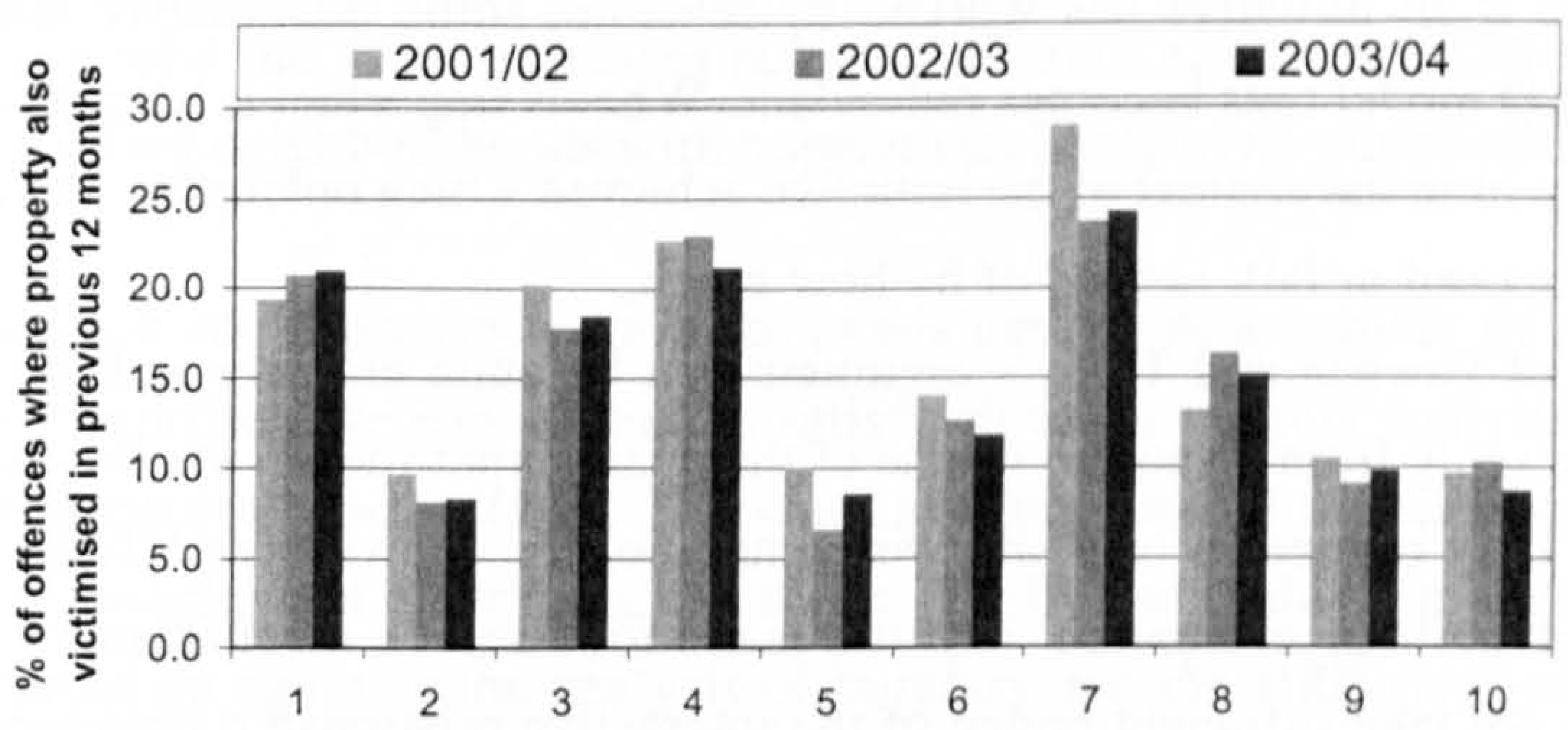


Figure 8.11: Trends in burglary dwelling repeat victimisation rate by group, 2000/01 to 2003/04. Source: WYP.

	Pre-test	Treatment	Post-test
Experimental group	O_1	X	O_2
Control group	O_1		O_2

Figure 8.12: The classic experimental design.

which may also reduce crime rates by their ability to change some of the underlying causes of crime, such as poverty or boredom. The victims are also independent actors and their behaviour will vary all the time. We could go on, but the problem should be clear: although the neighbourhood classification has been constructed using a wide variety of variables, there are a tremendous number of both general neighbourhood attributes and crime-specific attributes of neighbourhoods, their potential victims and offenders which are impossible or impractical to capture and include in the cluster analysis. The neighbourhood classification cannot control for everything in the way that the classic experimental design demands.

This is the crux of Pawson and Tilley's (1997) complaint against experimental and quasi-experimental design and the reason for their championing of a new *realistic evaluation* approach founded in scientific realist philosophy. They present a forceful argument for abandoning the OXO model of initiative evaluation to concentrate instead on a 'theory-driven' or 'explanatory evaluation' approach. Central is the identification and understanding of three key elements - namely *mechanisms*, *contexts* and *outcomes* - and their inter-relationships. They argue that "causal outcomes follow from the mechanisms acting in context" (Pawson and Tilley, 1997, page 58), and they define the individual elements thus (from Crawford, 1998, page204):

- *outcome*: the product which is caused by a mechanism acting under particular conditions
- *mechanism*: the intervention which allows the reaction
- *context*: the conditions which allow a mechanism to come into operation.

The various possible inter-relationships between each of these elements are identified and grounded in explicit theoretical assumptions. Importantly, 'realist' evaluators inspect outcomes "to discover if the conjectured mechanism/context theories are confirmed" (Pawson and Tilley, 1997, page 217), and not just to see if an initiative has worked by reaching some quantitative target. The control group from the OXO model thus becomes redundant. What is important is identifying exactly which mechanisms alive within the context of the initiative achieved which outcomes. Put another way, *how* did the initiative succeed or fail, instead of by how much.

The language of Pawson and Tilley's argument can be quite emotive and this has prompted at least one indignant reply from an author of one of the quasi-experimental studies they tore into (Bennett, 1996, see). They replied in turn and were unapologetic (Pawson and Tilley, 1996). Another criticism leveled at 'realistic evaluation' is that it is too focussed on mechanisms which are outcome oriented and does not take sufficient notice of the interactive relationship between mechanisms and contexts (Crawford and Jones, 1996). These interactive relationships can be better understood if due attention is paid to the initiative planning and implementation processes. Indeed, in community safety projects it has been suggested that these are as likely as anything to be the cause of an unsuccessful crime prevention programme (Hope and Murphy, 1983). As Crawford and Jones (1996) point out, 'processual mechanisms' such as 'multi-agency co-operation' and 'community consultation' may not in themselves be causal in a direct manner but may well influence the social context in which an initiative is being delivered, especially with regard to decision making, modes of communication and negotiation. Thus, "they may not cause a successful outcome but their absence may reduce the likelihood of attaining successful outcomes" (Crawford, 1998, page 208).

The actual decision to study the impact of a burglary reduction initiative using the neighbourhood classification was rather dictated by a paucity of good data pertaining to other types of initiative and to problems with coverage of other types of intervention. By good fortune, the administrators of the BRIL have been diligent in maintaining detailed records of their activities since their scheme began, so it is this initiative which is analysed below. Given all that has been introduced above, however, something ought to be said about where the analysis that follows might fit into one, or perhaps more, evaluation approaches.

To begin with, the analysis is not intended to be a comprehensive evaluation. In addition, this type of analysis was not foreseen at the outset as being something to be included in a BRIL evaluation and indeed is missing from the formal evaluation that has been conducted by Burden (2005). In general, the formal evaluation has the look of a piece of work conceived at the end (of a phase) of the BRIL project, rather than planned in advance of the commencement of the BRIL. Reflection by Crawford (1998) would suggest this is not uncommon.

As was mentioned at the beginning of this section, perhaps the classification could be used as a tool to aid the matching of control to experiment areas in a traditional OXO quasi-experimental design? It would, I believe, be a more efficient and rigorous method of achieving this than by laboriously poring over tables of census data for different neighbourhoods in search of areas that 'match'. Furthermore, the classification could be used in the context setting phase to help describe control and/or experiment neighbourhoods although the variables included would be insufficient in scope to overcome Pawson and Tilley's complaints against this type of exercise. The same argument could be levelled against the classification being used as the sole means of understanding context within a 'realistic evaluation'. It could perhaps be used as a point of reference for such a study, but Pawson and Tilley argue that context is something much more complex than that reflected in this particular classification.

As for evaluating process, the classification probably has least to offer here. Other than, perhaps, providing clues as to why the same processes achieve different effects in different types of neighbourhood - by identifying neighbourhoods with populations likely to be mistrustful of police contact, for example.

So the classification could perhaps contribute to evaluations of a number of different types but is unlikely to be a comprehensive enough tool to satisfy all of the context analysis needs of any one approach. This situation seems acceptable. The classification is a tool that can be used in a number of different ways to complement other tools in a range of different evaluation methodologies. These issues will be reflected on again as the analysis of burglary and the BRIL is discussed below. But first, it is worth looking at the BRIL project in more detail.

8.7.1 Burglary Reduction Initiative in Leeds (BRIL)

During the first phase of community safety strategy (1999-2002) in Leeds, there were a number of key achievements. These included the £2.5million Home Office funded Burglary Reduction in Leeds (BRIL) scheme and an additional £554,000 of Home Office funding to help residents at risk from distraction burglary (LCSP, 1999). Burglary reduction remained a priority during the second phase of community safety strategy (2002-2005), and burglary reduction initiatives such as BRIL and the Leeds Distraction Burglary Community Initiative have continued. Resources have also been allocated to reduce the risk from rogue traders, and to reminding people to take care to lock their doors and windows. Most recently, alley gating has been proposed, and a number of pilot sites in Leeds have been identified.

8.7.2 Types of Intervention

For the purposes of this analysis only the target hardening aspects of BRIL are being considered. Part of the original Home Office funding also went into investing in fast-track DNA testing in order to improve detection rates and speed up and secure more convictions. The target hardening aspect of the BRIL has been running since 2001 and is managed by a community organisation, CASAC (Community Action and Support Against Crime), based on Roundhay Road, Harehills.

Interventions are one of two types: reactive or proactive. A reactive intervention is offered when a victim referral is made by the police following the reporting of a burglary. If the victim lives within a regeneration area (SRB5 or NRA) then they may be eligible for free fitting of improved locks for doors and windows, and in some cases more sophisticated or robust forms of target hardening such as door and window grilles. The exact criteria for eligibility of this sort have varied over time according to the size each year of the contribution to BRIL operating costs from SRB5 and NRF budgets. Burglary victims who are not eligible for free work are offered the chance to go on a waiting list for free work should extra funding become available at a later date.

Proactive work has tended to be delivered in regeneration areas only. Sometimes this is in response to a perceived risk, which may or may not have been heightened by an increase in burglary in the locality generally. Some of this proactive work has been done on an ad-hoc basis but there is a growing trend to proactively harden a block of properties in the same neighbourhood at the same time. The thinking behind this strategy is to try and produce a neighbourhood effect. That is, by being conspicuous and appearing determined about improving a local area it is hoped potential burglars will be warned off the area entirely, and not just specific properties. This ought to be good news for the neighbourhood in question, possibly less good news for adjacent non-hardened neighbourhoods.

Target hardening is not compulsory. People can choose to refuse help, make their own arrangements or simply ignore attempts to be contacted.

8.7.3 Links Between Crime Rates and Target Hardening Activity

The first piece of analysis does not involve the classification at all and just considers the global impact of the initiative. Although a little simplistic, an initial hypothesis that can be tested is whether there is a linear relationship between the amount of BRIL target hardening undertaken over a period of time and the burglary rate in the same neighbourhood. The null hypothesis, H_0 being that there is no relationship between target hardening activity levels and burglary rates at a neighbourhood level. The aim of BRIL is to reduce burglary rates, so the scheme administrators would probably hope H_0 could be rejected.

The target hardening data is prepared by counting the total number of properties that have been in receipt of a target hardening intervention over a period from April 2002 to April 2004. A target hardening intervention is defined as a visit to a dwelling during which physical works were undertaken, such as the fitting of improved locks or the fitting of grilles. These counts are then divided by the total number of households in the neighbourhood to produce a proportion. Burglary dwelling rates for 2003/04 and 2005/05 are expressed per 1,000 households.

All of the distributions suffer from some skew, so a non-parametric bivariate correlation test is required. The decision was taken to use the Spearman Rank Correlation test, the workings of which have been described previously. The results show a significant ($\alpha = 0.01$), moderate positive linear relationship between the proportions of households target hardened and the burglary rates in both 2003/04 (coefficient is 0.554) and 2004/05 (coefficient is 0.430). That is, areas that have higher rates of target hardening also have higher rates of burglary. The result, a self fulfilling prophecy which rejects H_0, is to be expected. Indeed, it is a little surprising that the relationship is not stronger as reactive interventions, by definition, are undertaken in response to an actual offence and proactive interventions are targeted at high crime areas.

A more appropriate test would be to see if there is a linear relationship between the amount of BRIL target hardening over a period of time and *change* in burglary rates, using offence data from before BRIL compared to the most recent offence data. The null hypothesis, H_0 is that there is no relationship between target hardening activity levels and the change in burglary rates. Using data from all the 477 neighbourhoods a Spearman test produces results that show a significant ($\alpha = 0.05$) weak negative linear relationship, the coefficient being -0.309. H_0 is rejected, probably to the great relief of those that fund and manage BRIL! It appears that neighbourhoods which have higher target hardening rates see larger drops in burglary.

Potentially there are many and varied reasons why the relationship is not stronger. For example, there are a range of other deliberate interventions that are designed to either reduce crime in general, reduce the causes of crime, or reduce burglary in particular. None of these is controlled for in this test. It is also unsafe to assume that the offender population is always the same size and always operating in the same locations. Householders also have a will of their own, and their migration around the district and various attempts to secure their properties will all alter the likelihood of victimisation, quite apart from the BRIL efforts.

Moving on, the uncovering of global relationships is not the main concern of this research. What is of interest is the distinctiveness of relationships like those above within and between groups. The issue of distinctiveness can be explored by repeating the second (change) test above for each individual group. Put another way, we can analyse to what extent the relationship between intervention intensity and crime reduction varies by group. All the compounding factors that can not be easily controlled for become even more pertinent in this type of analysis so comparisons of results between different groups are tentative.

The results of Spearman Rank correlation tests on neighbourhoods within a single group are only significant for groups 2,6 and 10 - the most affluent groups. In each case the relationships are negative and weak: -0.309, -0.299 and -0.278 respectively. Perhaps it is possible that some of the confounding factors that might hinder the impact of target hardening are missing from these neighbourhoods, rather as Hirschfield (2004) suggested. Or maybe it is the presence of crime-reducing factors that the tests do not control for that is causing target hardening to have a better neighbourhood effect here than in poorer areas. Either way, given the principle mechanism of target hardening, it is questionable whether these are fair tests of the relationship between initiative activity and crime outcome.

8.7.4 Repeat Victimisation Evaluation

The problem is that although the general aim of BRIL is to reduce burglary, it tackles this by trying to prevent repeat victimisation and it could be argued that this can only be evaluated fairly by looking at burglary and target hardening at an individual household level. To be clear, in Pawson and Tilley's terms, what is under scrutiny here is the *mechanism* that involves physically improving the security of points of entry to individual dwellings. Not under consideration (although they ought to be at some point in time) are mechanisms that deliberately try to establish neighbourhood effects. Examples of these are explicit advertising of target hardening plans (see Laycock, 1992, for an example of the effect this had on a South Wales property marking scheme) or the implicit advertising of intent by proactive block hardening.

Unfortunately, from an evaluation perspective, but fortunately for people of Leeds, the number of properties being repeat victimised at a neighbourhood level in any one year is not large. Thus, as things stand to date (Summer 2005), BRIL has not been operating sufficiently long for there to be large enough samples of repeat victimisation available with which to conduct a within-group analysis of individual repeat victimisation. For this reason, it is only possible at the moment to evaluate the effect of target hardening on repeat victimisation between groups. Nevertheless, this might reveal interesting differences in the ways in which properties in neighbourhoods of a certain type respond to target hardening, with possible consequences for future BRIL delivery policy.

Controls

Typically, interventions such as BRIL are evaluated by analysing patterns of victimisation prior to the intervention period, both for the intervention site and a control site. The control is used to estimate what might have happened in the intervention site had the intervention not been introduced. Evaluating BRIL is made slightly awkward by the fact that the initiative has been deployed across the whole district, making it more difficult to establish a traditional type of control group.

What can be done is to exploit the individual nature of the study to create controls for each group from those properties that are known to have been burgled (from the WYP recorded crime data), but that have not received reactive target hardening, for one reason or another (which we can glean from the BRIL contacts database). The rather elegant result is a set of controls which come from the same neighbourhoods as the properties having their target hardening evaluated. With these control groups established, if households from the same group as the control that did choose to be target hardened turn out to suffer fewer repeats, then it would seem reasonable to infer that the BRIL has been a success. It is assumed that the target hardening of properties does not create a neighbourhood effect that would reduce the likelihood of other properties nearby a target hardened property from being victimised.

Definition of a Repeat

Following the recommendations of Clarke and Eck (2003), a 12 month sliding window is used to determine whether a burglary is considered a repeat visit. That is, if a household is burgled within 12 months of a previous burglary, then the most recent burglary is a repeat. As recorded crime data were only available up until 31st March 2004, only target hardening interventions prior to 31st March 2003 are considered. BRIL activity began in earnest in October 2001 which would allow for an 18 month period to be monitored for repeats. For clarity and ease of comparison, however, the start of the period to be analysed was set to April 2002.

Although the previously given definition of a repeat is rarely contested there is a good case for refining it in studies of this sort. The reason this is necessary is tied up with the actual nature of the burglaries themselves and what target hardening can and cannot be expected to achieve. For example, it would seem reasonable to expect the provision of improved door and window locks to deter burglars who attempt to enter a property by breaking through doors and windows. By contrast, if a repeat burglary is a sneak-in or deception burglary then it seems a little unfair to say the target hardening has failed.

It could, and perhaps should, be argued that target hardening activities should be delivered alongside educational efforts to raise awareness of risk and change the behaviour of victims, and that any repeat should thus be seen as a failing of intervention. At the other extreme, one statistic promoted by the LCSP (Leeds Community Safety, 2004) uses repeat victimisation figures where the modus operandi (MO) of the repeat involved exactly the same point of entry as was hardened. While producing flattering evaluation results, this definition of repeat victimisation of target hardened dwellings seems unrealistic. What is proposed instead is a definition of repeat victimisation that falls in the middle ground. Thus, sneak-ins (including climbing through open windows) and distraction burglaries are discounted, leaving only those offences where there was a forced entry with the repeat offence. The type of offence that triggered the target hardening is not considered.

Data Preparation

The same method as used previously was used again to identify the repeats in the police recorded data. This repeat dataset was then joined with the MO dataset to further refine the sample and remove repeats where the burglars sneaked or deceived their way in.

The task of joining the police recorded crime data with the BRIL contact database is rather more labour intensive. The interface to the latter has no built-in gazetteer to check the accuracy and consistent use of addresses and postcodes. The police recorded crime data are checked against a gazetteer, but the postcodes cannot be relied upon (explained elsewhere). Fortunately, records in the BRIL database pertaining to reactive work usually contain the crime number of the original offence, so the two datasets can be joined on this data. Remaining case records were joined using combinations of elements of the address information. Where reactive records in the BRIL database had no crime number this was inferred by comparing the date of the target hardening visit with the record of victimisation at the properties in question. In the reverse situation - a crime number is included but no visit date is recorded - a visit date was imputed by adding 7 days to the date the crime took place.

In some instances more than one target hardening visit was made to the same property. In these cases only the record of the last visit was used. There were also some cases where two, or sometimes more, properties in different locations were associated with the same crime number. Discussion with the BRIL administrators could not account for this with any degree of certainty, but it is possible that some of these records reflect a victim moving house and requesting help at their new address. In some cases one burglary at a flat resulted in several flats at the same location being target hardened, and there were also a few cases where a victims near neighbours had been target hardened.

To summarise, under consideration were target hardening interventions, both proactive and reactive for the period April 2002 to March 2003. The control groups were constructed from properties where there had been a burglary in the same period, but no target hardening intervention. Recorded crime data for the period April 2002 to March 2004 was used to find repeat victimisations within a 12 month sliding window, where the MO was a break-in. For the whole district, this sampling revealed 4,810 unique households where target hardening had been carried out, and 11,219 unique households which had been burgled but not subsequently target hardened (the control).

8.7.5 Impact of BRIL on Repeat Victimisation

It is worth pointing out that, at a group level, the intricacies of funding arrangements mean that the amount of target hardening undertaken varies from one group to another (Figure 8.13). Those neighbourhoods which intersect SRB5 or Neighbourhood Renewal areas will be the most likely to receive proactive interventions, and as householders in these areas also qualify for free reactive work, it might be expected that the take-up rates would be higher here than elsewhere.

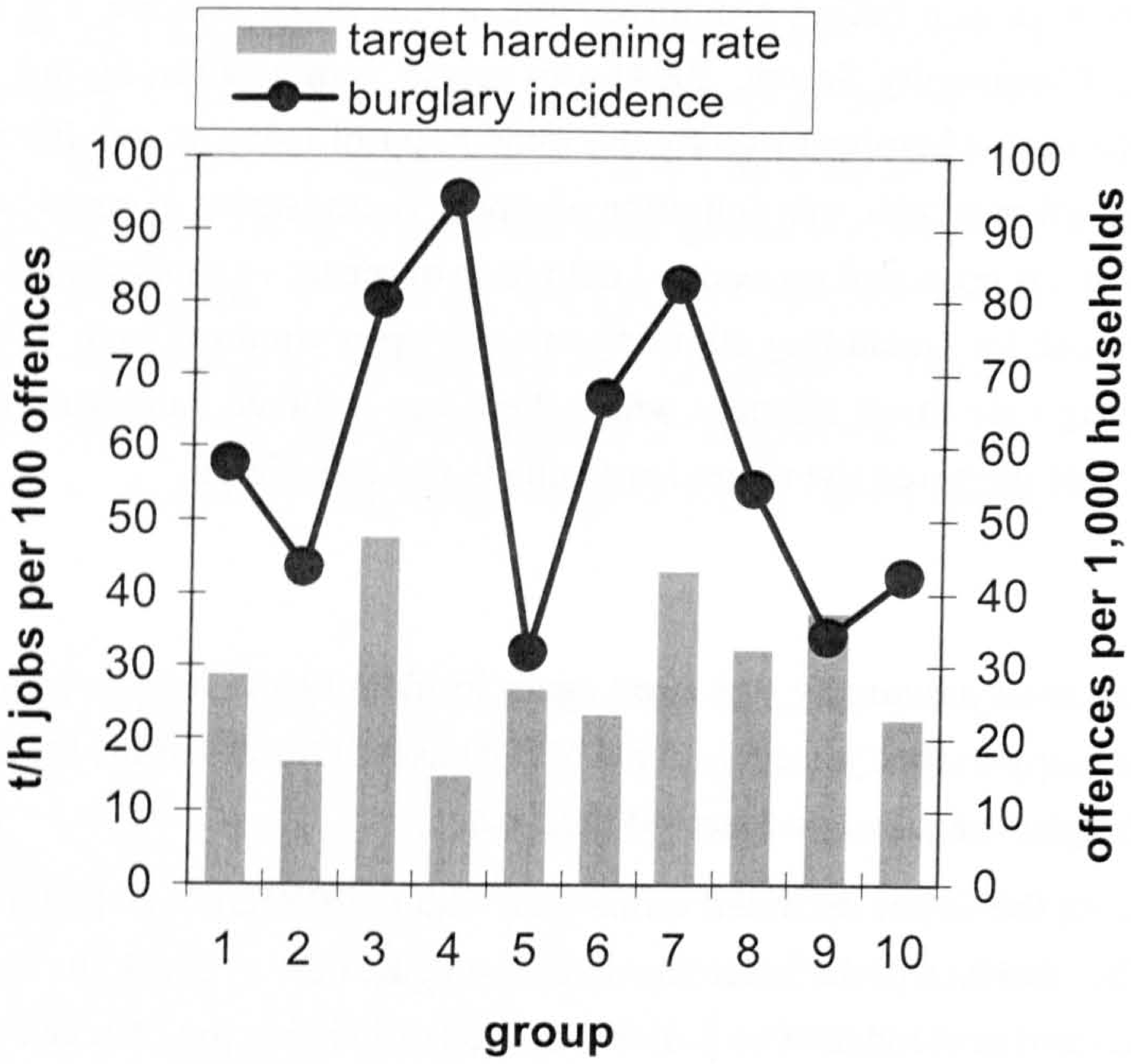


Figure 8.13: Variation in target hardening activity, by group, from April 2002 to March 2003. Source: WYP and the BRIL.

Most striking is the relatively low rate of target hardening activity in group 4, given that this group has the highest incidence of burglary dwelling. Approximately 36% of properties in group 4 intersect with SRB5 zones, but the high turnover of (student) residents and the high proportion of private rented accommodation are likely to be making it harder for BRIL to make inroads into this group. Of course, the data only pertains to activity up until the end of March 2003. If this is broken down and then compared to more recent activity it can be shown that in the six months up to the end of 2002/03, target hardening was carried out at 103 properties. In 2002/03, the number was 220 properties and in 2003/04 the number rose to 250 properties. The proportions for the different periods do not reveal that the relatively low level of activity in 2002/03 was compensated for to any great extent in 2003/04. (Note: the actual figures may be a little higher as Burden's (2005) analysis suggests that for properties in LS6 my analysis has undercounted target hardening by around 11% in 2003/04. My overall undercount for the same period is 3%, some of which will be explained by only counting one target hardening visit per property, when in fact some properties are target hardened over multiple visits).

The group which appears to have received a disproportionate amount of assistance is group 9, which groups together owner-occupier neighbourhoods which are a little less affluent than their group 7 counterparts. This group has the second lowest rate of burglary dwelling, but the third highest rate of target hardening. There is only a very small overlap between neighbourhoods in this group and SRB5 and NRA zones, so usually people would not be eligible for free services. Another unusual feature of target hardening in this group is the amount of proactive work that has been undertaken, again as there is almost no overlap with regeneration areas. It is possible that funding for the proactive work came from the central Partnership pot, or from the Street Crime Initiative funding that has been allocated to the BRIL budgets, but it is not possible to say for sure. In general, the complexities of funding of BRIL (Table 8.6) and the area-fenced nature of some, but not all, of the money makes it very difficult to argue whether areas have received more or less than their fair share of target hardening assistance.

Available Funds	**2002/03**	**2003/04**	**2004/05**
SRB5	365,686	279,377	229,427
NRF	2,000	331,000	0
Street Crime Initiative	60,000	59,300	86,700
Total External Funding	427,686	669,677	316,127
Local Funding	209,251	65,059	498,965
Total Budget	**£636,937**	**£734,736**	**£815,092**

Table 8.6: BRIL funding from 2002/03 to 2004/05. Adapted from (Burden, 2005, page 17).

Yet, regardless of where funding for target funding has come from, it is still possible to compare the effect of target hardening on different neighbourhood types. A calculation of repeat victimisation (repeat was a break-in) rates produces district-wide values of 9.5% for properties not target hardened, and 7.6% for properties that were target hardened. Figure 8.14 shows the variation from this district level, by group.

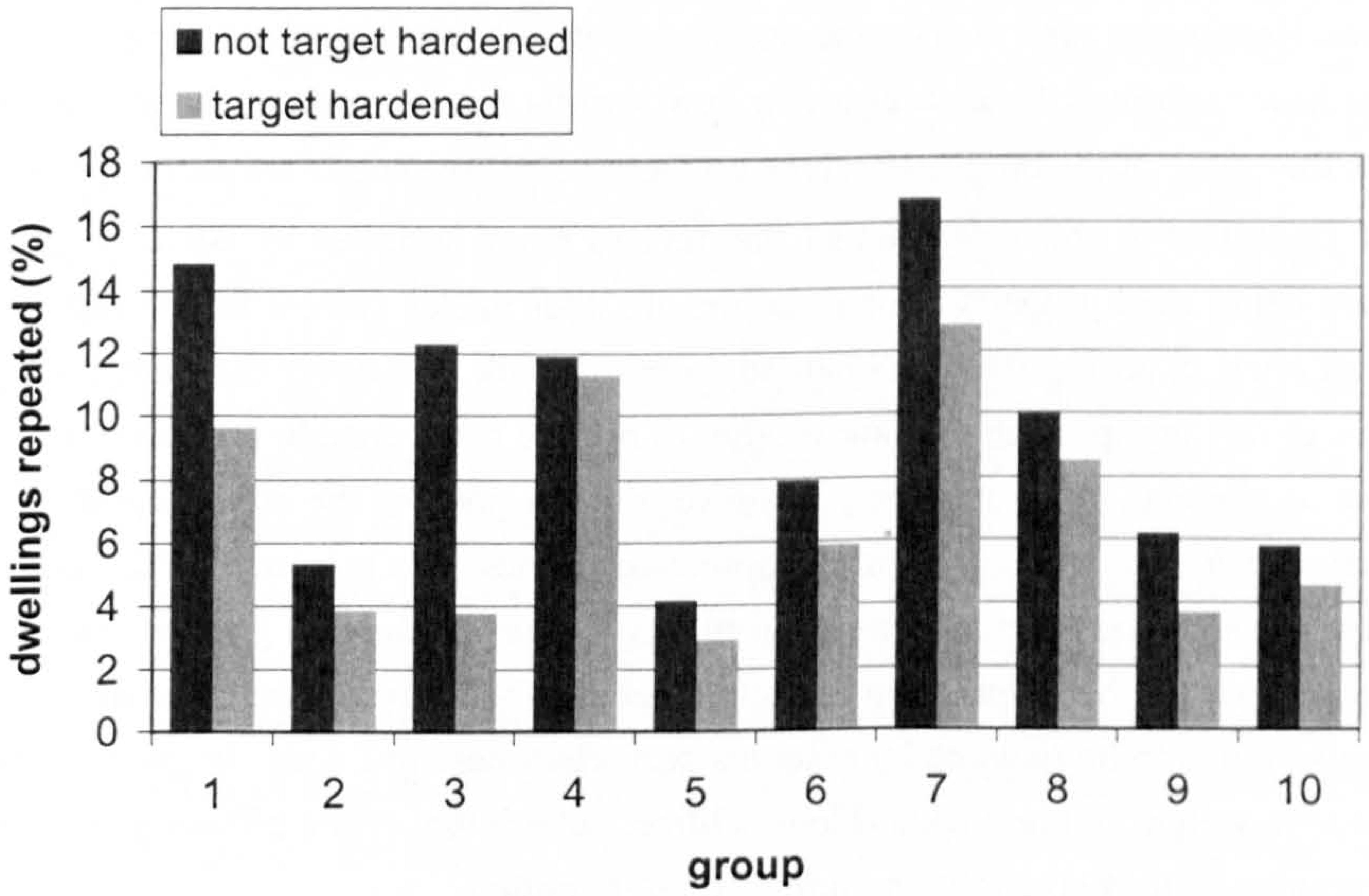


Figure 8.14: Variation in response to target hardening, by group, from April 2002 to March 2003. Source: WYP and the BRIL.

The variation in repeat victimisation by group has been discussed previously. What is of interest here is that it appears target hardening affects repeat victimisation to varying degrees according to neigbourhood type. In all groups the repeat rate is lower for properties that were target hardened.

The difference in response from group 3 (ethnically diverse) and group 7 (poorer council estates) makes an interesting first comparison. Although both refer about equal proportions of properties for target hardening assistance, the impact of the work appears much better for neighbourhoods in group 3. Interviews with BRIL administrators have revealed that levels of interest in BRIL are very high among people living in group 3 neighbourhoods. And this factor, perhaps supported by the fact that BRIL is administered by a charity (CASAC) that began life as a local black and Asian grassroots organisation, might mean people living in group 3 are more receptive and attentive of help and advice than those living in group 7 neighbourhoods. Other studies have certainly revealed that crime prevention advice for Asian communities (but not Afro-Caribbean) has been better received when it comes from community or religious organisations, and not the police (Thornton et al., 2003). Evidence of cultural variations in crime prevention uptake and outcome is much harder to find.

Alternatively, the differences may be due to spatial interrelationships between victims and offenders. Section 8.6.2 has shown that burglary appears to be a more local affair for nominals in group 7. It might be that the desire to offend locally in group 7 is strong enough to overcome the disincentive of discovering a target hardened property on a repeat visit. Or it could be that offenders here do not perceive that burgling a target hardened house as likely to increase the risk of detection. Whichever is the case, highlighting differences such as this might help BRIL administrators target certain types of neighbourhood for extra awareness campaigns. And if cultural issues are important, then consideration could perhaps be given to involving more local groups in the delivery of BRIL.

Group 4 neighbourhoods (studentland) have the highest burglary rates in the district. It might be a cause for concern therefore, to see that the impact of target hardening here appears to be the worst of all the groups. It is possible that the rewards to the burglar are great enough that extra efforts to overcome the target hardening are worthwhile. Furthermore, BRIL administrators have highlighted cases of intimidation being used by (presumably) burglars against BRIL workmen and local residents in the Hyde Park and Headingley area to dissuade residents from having target hardening work carried out. It is not clear how numerous these cases are or how serious are the types of threat, but nevertheless, it is evidence that some of the burglars preying on these neighbourhoods are particularly determined.

It is also important to consider whether the devices being supplied by BRIL in group 4 neighbourhoods are being used properly. For example, are door grilles (where fitted) kept locked or are they being left open, exposing the doors behind them to attack. Whatever the case, the high resident turnover rates in this group mean that knowledge of how to take sensible precautions is continually migrating out as students leave university or move to other parts of the city. Replacing them come new, often naive residents, who may neither appreciate or know of a previous BRIL visit, or be aware that their new property has been victimised in the past - putting them at greater risk. Certainly, on this evidence it appears that alternative measures need to be delivered in these neighbourhoods if repeat victimisation is to be reduced by amounts seen elsewhere in Leeds. In general, the results of this analysis lend weight to arguments (Home Office, 2004c) that crime affecting students demands a tailored response by those planning community safety policy.

In common with group 4, target hardening in groups 2 and 5 (northern and southern rural) produces only a small reduction in repeat victimisation. The difference here is that incidence is very much lower. It is possible that the types of offender that do target these neighbourhoods are not especially deterred by target hardening hardening to properties that are more likely to have been fairly secure to begin with. Whether evidence of low impact such as this should be used to scale down BRIL activity in these neighbourhoods is a difficult question, especially as Burden's (2005) evaluation found that 94% of people felt less afraid about crime after a BRIL visit, arguably a very worthy outcome by itself.

8.7.6 Costs of not Reducing Repeat Victimisation

To round off this analysis of the BRIL and its impact on repeat victimisation, the calculations produced in the previous sections have been used to answer the question: do the costs of target hardening outweigh the benefits? Burden's (2005) evaluation did consider the value of reassurance and reduction in fear of crime and also commented on quality of life issues and general benefits to society of projects such as the BRIL. Yet, the cost implications of target hardening were not evaluated to a conclusion, although the methodology was described and a recommendation made that this issue be researched further.

This issue of cost-benefit is often at the core of the 'what works' rhetoric of community safety policy in England and Wales at present. In a sense, attempting to estimate the cost-benefit of a scheme such as BRIL helps perpetuate this fascination when really the desire might be to be critical of it. Yet, while costs are not estimated, doubts may persist, and so there is an argument for undertaking a certain amount of quantitative analysis, if only so that debates about alternative evaluation approaches can then move forward. The distinctions between an intervention being effective, being cost effective and being cost-beneficial need to be outlined. An analysis of the first type would ask whether an intervention was effective at reducing burglary. A cost effectiveness analysis would seek to determine which is the cheapest means of reducing burglary, and a cost-benefit analysis would seek to determine whether the benefits from burglary reduction were greater than the costs of the intervention.

In general agreement with Burden (2005), the approach adopted has been to estimate the number of offences due to repeat victimisation that might have occurred had target hardening not been deployed where it was. Estimates of the number of extra offences are calculated for each individual group and then summed together. This total number of extra burglaries is then multiplied by the average whole life-cycle cost of a burglary incident as estimated by the Home Office.

These estimates of the economic and social costs of crime were published in 2000 (£2,300 per burglary dwelling) (Brand and Price, 2000) and revised for 2003/04 (£3,268 per burglary dwelling) (Home Office, 2005a). These costs, however, are based on the British Crime Survey, which includes counts of burglaries that may never have been reported to the police. Working with the 2003/04 data, which is when a greater proportion of the repeats occurred, the BCS estimated there were 943,000 burglary dwelling offences in England and Wales. Yet in the same period the number of offences recorded by the police was just 402,333, meaning 540,667 offences are likely to have gone unrecorded. Following Bowles and Pradiptyo (2004), to achieve a weighted average cost for recorded and unrecorded offences of £3,628, the average cost of a recorded burglary must be around £4,796 given that the costs of unrecorded burglary (anticipation costs plus consequential costs but excluding justice system costs) is around £2,131.

The total cost of the estimated extra offence across Leeds (based on the unit cost of £4,796) is then compared with the operating costs of BRIL for the same period. If the estimated costs of the extra offences are greater than the BRIL operating costs then BRIL would be considered to have been cost-beneficial. No attempt is made to quantify the cost benefits of reducing fear of crime, providing employment to BRIL staff, or being seen (by policy fund holders) to be doing something rather than nothing.

The estimate of extra offences is calculated by applying the repeat victimisation rate calculated for properties that were not hardened to the number of properties that were hardened. Then the actual number of repeat victimisations at properties that were hardened is subtracted to produce the estimated extra number of properties that would have been victimised had target hardening not been deployed.

Group	Actual repeats	Est. extra repeats	Est. cost (£)	Cost per household (£)
1	35	18.9	90,434	4.15
2	8	3.1	14,673	0.52
3	17	38.2	183,148	15.60
4	24	1.3	6,008	0.40
5	10	4.4	21,123	0.52
6	33	11.3	54,129	1.50
7	150	45.8	219,734	6.67
8	54	10.0	47,944	1.32
9	21	12.9	61,949	1.53
10	16	4.4	21,008	0.57
Totals		**150.2**	**720,150**	

Table 8.7: Estimate of extra repeats that would have occurred had target hardening not taken place.

The results of the calculations (Table 8.7) estimate that the economic and social cost that would have been incurred had target hardening not been deployed in 2002/03 would have been £720,150. This compares with BRIL operating costs of £636,937 in 2002/03. This suggests that BRIL was cost-beneficial during this period. The benefits of reducing repeat victimisation by target hardening outweigh the BRIL costs. Moreover, the costs of inaction would have been most keenly felt, on average, among the most deprived types of neighbourhood (groups 1,3 and 7), while those living in more affluent types of neighbourhood (groups 2,5 and 10) would have suffered the least financially. Thus, the delivery of BRIL during this period would also appear to have helped 'narrow the gap' between the most disadvantaged people and neighbourhoods and the rest of the city - a key long-term policy objective of Leeds City Council and its various partnerships (Leeds Initiative, 2004).

8.8 Concluding Remarks

The utility of the classification is largely predicated on its ability to reduce the overall variability of crime data and to discriminate between those groups where the penetration of community safety policy is likely to be highest and lowest. Findings from an earlier analysis of discriminatory power (Chapter 6) raised doubts about the classification's ability to discriminate burglary dwelling but confirmatory analysis in this chapter suggests that there are statistically significant differences between the group set as whole, although the extent of the differences between pairs of groups varies, with some pairs not being significantly different. Furthermore, this situation changes over time as crime goes up and down by differing degrees in different groups.

As has already been shown with criminal damage data, the mapping of group-standardised burglary rates and burglary rate changes highlights parts of Leeds that might not normally come to the analysts attention when reliance is placed solely on density mapping. Among local groups, problems may be understood but the analysis reported in this chapter shows that meaningful local patterns can also be found when studying the district as a whole if proper account is taken of neighbourhood type. That many of the patterns make common sense also lends support to the idea that classification is a valid model of neighbourhood type for Leeds. Yet, the extent to which groups stand out as having unique burglary characteristics does vary. Those with unique characteristics include the student group (4), the poorer of the social renting groups (7) and to some extent its slightly less deprived partner, group (8). The more rural groups (2 and 5) also stand out fairly well and there are sufficient differences between them to support the idea that the inclusion of two types of rural or urban/rural edge neighbourhood are justified. There are some interesting patterns for the more affluent northern suburbs group (6) and also for the ethnic group (3). This last group, however, serves as a reminder that it is easy to slip into a habit of ascribing all of a group's issues to just one, albeit important, variable, in this case ethnic heterogeneity. While there are well reported cultural issues with respect to police relations and race-related crime, the importance of ethnicity to more general patterns of victimisation and crime prevention response is less certain. Further research into cultural issues surrounding the delivery of, and response to, crime prevention policies is something to be recommended.

The analysis of stolen property, nominals and repeat victimisation are all penetrating in their own way and each contributes to the building up of a richer picture of how, when, where and why burglars burgle. The nominals analysis could almost certainly be improved with additional data from West Yorkshire Police. In particular, more detailed MO descriptions might be the only way to solve the riddle of the burglary specialist. The data to hand uses a classification system that is known not to be used reliably. Access to free text of the MO might be more useful and certainly has had to be resorted to for other Leeds community safety analysis (identifying alley-gating pilot sites).

The part-evaluation of the BRIL outcomes is hopefully a clear example of the utility of the classification as an aid to assessing the impact of crime prevention initiatives. It certainly seems to go some way to answering Crawford's call for evaluations that ask "what works, for whom and under which conditions?" (Crawford, 1998). In a fuller evaluation following a scientific realist approach the prior analyses of the burglary data would help in the identification and understanding of the different mechanisms and contexts of target hardening schemes. Although for the analysis presented here the main mechanism to understand was that concerning the fitting of improved door and window hardware in order to make it harder for burglars to effect a forced entry.

The assumption made at the outset of the BRIL analysis was that there would be no neighbourhood effects as a result of target hardening a property. Further research needs to be taken to establish whether this a reasonable assumption, or not. The administrators of BRIL certainly seem to believe that the target hardening of blocks of properties can have an effect greater than the sum of the parts, and they might well be right. Tests with a space-time permutation scan statistic (SaTScan) (Kuldorff, 2001; Kuldorff et al., 2005) have been able to identify statistically significant space-time clusters of target hardening activity, over a period of years. The same software has received a lot of attention from epidemiologists and health geographers and has been used to identify clusters of crimes in Sweden (Ceccato and Haining, 2004) and to identify space-time clusters of measles outbreaks in the USA (Haining and Cliff, 2003), for example. Further analysis of the target hardening space-time clusters could be undertaken to see if properties in these neighbourhoods did see any more benefit than properties in neighbourhoods where target hardening has been piecemeal. The reason for experimenting with SaTScan in the first place was to see if the geography of the space-time clusters it found would enable a between-group analysis of neighbourhood effects. The interim finding was that the target-hardening clusters were too frequently located across two or more types of neigbourhood - a matter made even worse if a buffer is to be used to measure wider neighbourhood effects - making such an analysis infeasible within the scope of the research as a whole, for now.

Given the question mark over the neighbourhood effect assumption, however, the between-group analysis has turned up some interesting results. The geography of area-targeted regeneration funding combined with the proportion of BRIL funding from these various sources each year makes it very difficult to say if BRIL activities have been fairly distributed geographically. Yet, it is possible to look at the between-group variation in programme delivery and wonder if some types of neighbourhood are perhaps receiving more or less attention than their burglary problems merit. It has also been possible to show that where target hardening has been applied there do appear to be between-group differences in the impact that it has had on repeat victimisation. Some tentative suggestions have been made to try and explain these differences and some of these seem to square with findings made by others who have evaluated similar schemes. Some of these issues, if they could be substantiated, might be of value to BRIL and other schemes when they come to determine how to make the delivery of their initiatives responsive to local needs and conditions. A last unintended side-effect of the BRIL evaluation design was the ability to estimate the likely number of burglaries that have been prevented as a result of target hardening. Burden (2005) has already suggested this was would be a way forward in estimating whether the savings as a result of BRIL outweigh the costs, and after some (justified) manipulation of the economic costs of burglary dwelling this has been deemed to be the case. Hopefully this is useful adjunct to complement the existing evaluation of BRIL. It might also be useful propaganda for the Safer Leeds partnership, although if the methodology is accepted as a valid one then opposite findings from cost-benefit analysis of subsequent years data (when it becomes available) would have to be accepted in turn.

Finally, one is continuously reminded of the way in which burglary victimisation and associated factors can vary quite markedly in space and time, and often within fairly short time frames. It seems unlikely that many of the variables used within the classification would be in such a similar state of flux, questioning whether the right variables have been chosen. One option would be to incorporate more dynamic variables and remove some of the the stolid measures, presuming, of course, that sources of more dynamic data could be unearthed. Yet one is constantly reminded from the literature and local experience that controlling for all causal factors is likely to be nigh on impossible. A radical redesign might not significantly improve the classification's ability to reflect the dynamics of the burglary problem and even if it could this might not suit analyses of other crime types.

Maybe there is no harm in having a conservative classification that reflects more fixed neighbourhood attributes. Indeed, if the classification was to incorporate more dynamic variables it would need to be re-calculated frequently which in turn might soak up unacceptable amounts of resource. Such a classification would also be more unwieldy in analyses with longer historic time frames as neighbourhoods may change their group membership over time. One would also have to rule in the possibility that evolutions of the classification would produce different numbers of groups and different group profiles.

This chapter suggests that most community safety analysis - as distinct from policing analysis - can probably be done with a classification of the sort that has been developed. That is, a classification that is fairly conservative in its aims; uses well understood variables; has only local, not national coverage; and incorporates as many locally produced datasets as possible - both to provide greater timeliness but also to embed a degree of ownership for different partner agencies. Such a classification might be re-calculated every two or three years, perhaps timed to coincide with the statutory triennial audit process.

In addition to this work, however, it would be interesting to explore the idea of designing a more dynamic model of crime and community safety, perhaps along the lines of the Leeds Neighbourhood Orientated Model of Area Demand (NOMAD). This would be updated annually, or perhaps even 6 monthly and be based purely on available crime, disorder and drugs data. Neighbourhoods would then be classified using ranks on a composite index and ascribed a label similar to those given to neighbourhoods where housing demand is high (Popular), where it is low (In Significant Decline) and where the situation is 'On the Edge'.

Chapter 9

Conclusions

The main aims and objectives of this thesis were detailed in Chapter 1 and Section 3.5, and referenced the existing literature on area profiling for crime analysis. This final chapter summarises and assesses the extent to which these aims have been met and also discusses possibilities for further work that might extend and improve the classification that has been developed. The limitations of the research are also considered, and some issues regarding the manner and style in which the research was undertaken are reflected upon.

It is also worth noting here that, prior to the completion of the thesis, the author moved away from academia and began working within the community safety practitioner domain for which this research is intended. This change of circumstance has given the author new, and sometimes different insights into the ways in which community safety is analysed, both geographically and otherwise, by local authorities, police and other agencies. Similarly, certain aspects of policy, both local and national, will be discussed below that were not considered in the earlier reviews (Chapter 2 and Chapter 3). This is not admitting that the initial reviews were not useful or correct, but that community safety policy direction from the Home Office and other Government departments is something that is continually evolving and changing, often much faster than the academic cycles for doctoral theses and journal article production.

9.1 Summary of Research Achievements

The main aim of the thesis was to contribute to the debate surrounding the usefulness of area classifications for crime pattern analysis at a neighbourhood scale. A number of objectives were set to achieve this aim and the extent to which these have been realised is now discussed.

The thesis began by reviewing the criminology and community safety policy literature in order to appreciate better the relationships between geography and levels and patterns of crime, and to understand why crime pattern analysis is conducted in the way that it is by both academics and practitioners. The academic literature was derived from a diverse range of sources, but most came either from those interested in the geography of crime or those criminologists interested in the effect that environments can have on criminal behaviour - a sub-discipline usually referred to as environmental criminology.

Reviewing this latter body of work did much to stimulate ideas about how different types of neighbourhood variable might influence crime levels, and thus informed decisions regarding the selection of data variables for a neighbourhood classification. Yet, the history of criminological theory also revealed a number of different ways in which the concept of crime itself could be defined, and an early epistemological choice had to be made about what type of knowledge a neighbourhood classification for Leeds would be based upon. It was argued that given the nature of the research sponsorship and end user expectations, it was necessary to consider crime as law-breaking behaviour, even though a number of interesting criticisms of this view have been made. Thus, the research might be criticised for not attempting to break free from the recent trend in England and Wales for administrative criminology and the new managerialist preoccupation with performance monitoring (Flynn, 1997).

Instead, the neighbourhood classification approach that was finally adopted can be viewed as a continuation of the attention paid by geographers to multivariate analysis back in the 1950s and 1960s. That early work was sparked by the production of a new set of small area census data, rather as this research has sought to exploit new data published from the 2001 Census. It is argued that there is a danger in relying heavily on data of this type, both because it can only fractionally (Etzioni and Lehman, 1967) capture aspects of behaviour which sections of the criminology literature suggest are so important, but also because its scientific veracity at the moment of collection may create a false sense of security that affects the way the data are then used to represent neighbourhoods for a decade or more. Some have rather weakly argued that crime pattern analysts should not be too preoccupied with behavioral criminology (e.g. Felson and Clarke, 1998; Chainey and Ratcliffe, 2005), but Lea's (1992) view was that even though technical in nature, situational prevention measures must have a social element to explain why interventions work better on some occasions than others (Lea, 1992) and an attempt to test this hypothesis was added to the research and was discussed in Chapter 8.

The importance of spatial analysis was acknowledged both implicitly and explicitly within the community safety policy literature. Traditionally, the triennial crime audit has provided the impetus for steps forward in data collection and analysis methodology, while it was argued that the Partnership Business Model looked set to define more tightly the role for crime pattern analysis in general, and also spatial analysis in particular, as community safety policy for England and Wales develops. This change in emphasis definitely looks set to gain royal ascent, with the Crime and Disorder Act review recommending that the triennial audit be replaced by six-monthly Strategic Intelligence Assessments (SIAs) being produced jointly between the police - for whom the SIA is an integral, multi-level part of the National Intelligence Model - and other CDRP agencies. This looks set to be accompanied by a set of minimum standards for the types of analysis that will be included within the joint-SIA and partnership working in general. The concern with respect to this research is that an overly prescriptive framework for partnership analysis will further over-stretch the resources that are available for more speculative and reflective pieces of research and analysis - such as that presented in this thesis.

The second objective of the research, was to review the geography literature to discover which techniques and approaches had been developed to aid crime mapping and area profiling. A number of cartographic techniques were reviewed before discussion turned towards the need for area profiling, citing the concerns of Gloria Laycock regarding the preoccupation by many analysts with seductive-looking, but limited hotspot maps, and the need, thus, to move 'beyond blobology' (Laycock, 2003). Craglia et al. (2000) have demonstrated a number of ways in which the combination of cartography with area classification might be used to address some of Gloria Laycock's concerns. Yet, despite accessible documentation of the Home Office classifications for CDRPs and BCUs, little evidence could be found of people taking forward crime analysis using these products, although the extent to which these classifications are being used by local and national government officers is often difficult to guage, as most work is not published. There is, however, evidence of an increasing amount of attention being paid by academic geographers - particularly those connected with the Centre for Advanced Spatial Analysis, University College London and Richard Webber - to the possibilities of using commercial consumer classifications to study crime patterns.

Used uncritically, the concerns voiced by Curry (1995; 1998) and Goss (1995) about the social implications of geodemographics might be well founded. Moreover, after reviewing a number of recent studies using commercial products such as Mosaic, it is argued that there is often insufficient discussion given to the problem of using a classification about whose design and construction very little is known. These factors influenced the decision to produce a neighbourhood classification for Leeds in keeping with the way in which the Home Office classifications for CDRPs and BCUs had been designed, that is, open and transparent about design and methodological decisions and restricted to the task of analysing crime patterns - in line with the arguments by others regarding the (un)suitability of general purpose tools for specific types of analysis (Openshaw and Wymer, 1995; Voas and Williamson, 2001).

Sibley's arguments regarding the use of an exploratory data analysis (EDA) approach to application of a technique such as geodemographics (Sibley, 1998) were also taken into consideration as a result of the review. Particular value was seen in the argument that by facilitating an open-minded approach to spatial ordering, there might be be benefits for the analysis of interplay between theory and data. This would appear to answer some of the concerns regarding the implications of geodemographics, and provide a rationale for continued development of geodemographics and area classifications. Evidence was also found of criticism of some projects for failing to evaluate the classification against a range of real problems (Dubes and Jain, 1979; Rapkin and Luke, 1993), strengthening the view that one objective of the thesis must be to apply the neighbourhood classification for Leeds to a number of real community safety analysis tasks.

Chapter 3 presented the results of the third objective of the research. This was to assemble a database containing a wide variety of neighbourhood variables related to community safety. Three surveys were undertaken to inform variable selection, one based on a review of academic studies of the geography of crime and two others which sought the opinions of community safety analysts and other officers. These consultations with people familiar with the concept of classification for community safety (in the form Home Office classifications) were particularly helpful. They not only provided suggestions of useful variables but also provided suggestions about how such data might be sourced. This was especially the case during the scoping event for the Leeds Statistics Project, and on reflection, the involvement with this project was key to the completion of the thesis as a whole.

Very often, there is very little space given over to discussing problems about obtaining appropriate data in the geography of crime studies that are presented in journals. Indeed, reference had to be made to information systems literature to find academic discussions on these issues. Typical problems include data simply not being collected, or misunderstandings surrounding the Data Protection Act or Human Rights Act leading to reluctance on the part of data holders to share information, or even advertise its existence. From the experiences gained during this research, much would appear to depend upon personal relationships with individuals, although widely supported projects such as Leeds Statistics can help greatly in bringing together disparate datasets for a common purpose. Insider status can also be conferred on outsiders working on a project such as Leeds Statistics, such as the author, making subsequent negotiations for access to data much easier.

Once data started to be collated it became apparent that many of the aspects of neighbourhoods identified as useful in the literature could be produced without too much difficulty or otherwise could be substituted with a proxy variable. On occasions, new and interesting data were also sourced without a clear aim as to how it might be used. Sometimes these datasets turned out to be unsuitable - either because of incompleteness or poor accuracy - but in some cases, arguments for inclusion of novel data sets were sought retrospectively. It should also be noted that a number of datasets used for the classification were being unearthed for community safety analysis in Leeds for the first time. This benefited the thesis and also helped those sponsoring the research to conduct analysis of their own that hitherto had not been possible.

The national census still provides a large number of variables for the classification, although attempts were made to ensure that alternatives were used where possible, even if modelling techniques were required to harmonise geographies. It is a regret, however, that more time was not available to identify local and more timely sources of data on population, benefits and housing structures. From experiences as a practitioner, the data does exist with which to conduct many types of useful analysis, but misunderstandings and lack of skilled data technicians can make its acquisition difficult. Similarly, if this research were to be repeated there might be merit in using the lower-layer super output area (SOA) geography created for England and Wales by ONS. This geography was published too late to be used for this thesis, but a number of useful datasets are now available for this and other SOA geographies, and awareness of the geography within local authorities is also increasing (probably in large part due to its use for the 2004 Index of Deprivation).

The timely release of census data was also a limiting factor, not because data had to be left out of the classification but because the decision was made to delay the creation of partitions until the census interaction data became available. Thus, an opportunity cost had to be born, principally by leaving less time to design and test the classification than would have been liked. An alternative would have been to adopt Vickers' approach (Vickers et al., 2005) and omit interaction and workplace from the classification and risk criticism that the classification might be in some way incomplete.

The use of so much census data limits the opportunity to refresh the classification at sub-decennial intervals. More will be said about Local Area Agreements (LAA) below (Section 9.2), but the focus that these are placing on performance data cutting across a wide range of local authority delivery areas might have the effect of improving data sharing within and between local partner agencies. Yet, this is likely to vary from one authority to another, and not all local authorities have had to produce an LAA as yet.

The use of information from the PLASC and pupil level information of school exclusions provided more opportunities for capturing information about young people than is possible with the national census. Furthermore, it is possible that future analysis of truancy data might yield useful variables. There might be a danger, on the other hand, of too much bias being given to youth-related issues simply because a lot of data is available.

By contrast to education, the health variables capturing alcohol and narcotics-related hospital episodes need to be developed further with clinicians to ensure that issues with coding reliability are properly understood and that the diagnosis codes used do accurately reflect problems relevant to community safety. Given the number of acquisitive crimes committed by problem drug users, other variables reflecting their geography and lifestyles might also benefit the classification. Unfortunately, such data are not easy to collate, sometimes because of tension between agencies with differing aims (e.g. prosecution by the police, treatment by the Drug Intervention Programme (DIP), and harm reduction by community health). The type of information routinely collated by the DIP is also a good example of how performance management culture has resulted in data being collected to reflect processes but not outcomes; that is, caseload process management, but nothing to do with recidivism.

In Chapter 4, the discussion moved on to the design process for the partition that forms the basis of the final classification. It was demonstrated that whether an art or a craft, a balance needs to be struck between adhering strictly to the recommendations from formal mathematical techniques and some amount of expert knowledge of the study area and the problem domain - in this case, crime. To reflect this, a number of alternative techniques were applied at each stage in the design of the LCCS, and the results were then compared for levels of agreement. More time could be spent experimenting with different approaches, and there is an extensive literature on cluster analysis, almost exclusively from outside of geography. That is not to say that geographers have not experimented with novel approaches to classification (e.g. Openshaw and Wymer, 1995; See and Openshaw, 2001), but despite a resurgence of interest in geodemographics (Longley, 2005), it is argued that there is a shortage of recent work by geographers to reflect the diversity of options within the discipline of cluster analysis and classification.

On reflection, the least satisfactory part of the design process was the design of the crime heuristics (Section 5.4.2). The choice of crime types could be changed to resolve problems caused by crime groups that aggregate too diverse a mix of offence types, and calculations of reductions in variability, following Voas and Williamson (2001), might be more appropriate than the measures deployed. The reason the variability test was not used is that its usefulness was not properly appreciated at the stage in the research when it would have been beneficial. Consideration also had to be given to the time taken to run tests of this sort against multiple partitions. With more time to develop programming skills for R, the variability test could be automated, making its use straightforward. The reduction in variability tests were used to show that the LCCS is better at reducing the variability in the neighbourhood data than a simple classification created from deprivation score deciles, yet the results continue to raise questions about how many and what type of variables are actually needed to create a useful partition.

With the partition created, Chapter 6 demonstrated the use of a range of visual aids and metrics with which to 'paint' group portraits and thus turn the partition into a final, task-specific classification that could be compared with other classifications. The pen portraits of the different groups helped to reveal what it is that makes each group unique, but it is clear that some groups are more unique than others. Characteristically, it is those neighbourhoods closest to the Leeds average that are among the most difficult to partition into distinct groups. How much of a problem this might cause is likely to depend on the nature of the analysis the classification is applied to. More often than not, policy attention is focussed on the most needy neighbourhoods and neighbourhoods with high crime - which often are the same places. Thus, in identifying these needy locations, the weaker classification of average neighbourhoods is likely to be unimportant. Yet, products such as the Index of Deprivation 2004 exist for this type of analysis, and using the LCCS in such a manner only makes sense if particular groups want to be highlighted, such as ethnic minorities or students. From the outset of the research, it had been envisaged that the classification would be used to identify high and low crime neighbourhoods *within-group*, but while uncertainty remains about the uniqueness and robustness of some groups, the results from this type of analysis will need to be scrutinised more carefully.

Comparing the LCCS with other classifications did support the argument for a task-specific classification over a general purpose tool, and this, it is argued, is a useful contribution to the debate about the limitations of general purpose classifications for crime analysis. In most tests, the LCCS did as well as, or better than, other classifications, although the margin by which it performed better was not as great as had been hoped for at the outset of the research.

It was clear that all the classifications struggled to discriminate some types of crime, and in these cases it may not be helpful to adopt a neighbourhood classification approach to crime pattern analysis. Neither is it certain whether modification to the design of the LCCS would improve the discrimination and/or reduction in variability, although a potential way to research this problem is suggested in Section 9.2. Nevertheless, the argument that "a classification can only be deemed 'good' or 'poor' when it has been evaluated in terms of the specific purpose for which it is required" (Openshaw, 1983, page 245) serves as a reminder that testing also needs to be supplemented by qualitative appraisals of case studies of genuine community safety analysis problems using the classification, and the remainder of the thesis concentrated on these.

The first case study to use the classification (Chapter 7) demonstrated its usefulness when applied in a manner following the principles of EDA. As such, many aspects of the neighbourhood differences in criminal damage tended to pose more questions and suggest hypotheses, rather than always providing explanations. The between-group comparisons were not explored at length, although the findings from this approach to using the classification might provide an initial guide for community safety practitioners looking to target resources in particular types of neighbourhoods.

Of more interest was the extent to which problems varied within groups. A number of EDA techniques were demonstrated to identify residual neighbourhoods for further inspection. Furthermore, following the principles of the more recent developments in Exploratory Spatial Data Analysis (ESDA), maps utilising the LCCS showed the geography of these residuals. The analysis then proceeded to consider membership strength, unique neighbourhood features and geographical circumstances, and each of these prompted new hypotheses that future research might test.

Sibley's argument that EDA "encourages and facilitates repeated references to the data and a cautious, sceptical attitude to theory" (Sibley, 1990, page 4) proved very convincing. Furthermore, it is argued that some of the criticisms of geodemographics (e.g. Curry, 1995; Goss, 1995; Curry, 1998) might be avoided by probing rather than discarding residual neighbourhoods. However, it is also argued that a balance must be struck between analysing residual neighbourhoods in depth and utilising the generalising power of the classification. However out of fashion generalisation may be in academic urban geography, some have argued it is the cornerstone to rational planning policy (Longley and Harris, 1999) for its ability to make predictions for which solutions can then be planned.

The second case study, presented in Chapter 8, is arguably a more balanced piece of analysis, although being restricted to between-group comparisons limited the extent to which residuals could be identified anyway. In this case study, of the geography of domestic burglary and the responses to target hardening interventions, the classification shows the usefulness of a particular intervention strategy in a way that is both academically interesting, but also useful for community safety managers as a component of a formal evaluation. Indeed, this chapter was specifically designed to complement an existing evaluation of the Burglary Reduction in Leeds (BRIL) scheme (by Burden, 2005).

The first conclusion to be reached was that there are statistically significant differences in burglary between the groups as whole, although the extent of the differences between pairs of groups varies, with some pairs not being significantly different. Furthermore, this situation appears to change over time as crime goes up and down by differing degrees in different groups. This dynamic nature of group attributes is discussed below (Section 9.1.2).

The mapping of group-standardised burglary rates and burglary rate changes also highlighted locations in Leeds that might not normally come to the attention of analysts when reliance is placed solely on kernel density hotspot mapping, for example. Among local officers, burglary problems may be understood, but the analysis demonstrated that meaningful local patterns can also be found when studying the district as a whole when account is taken of neighbourhood type.

It was possible to show that where target hardening had been applied, there appeared to be between-group differences in the impact on repeat victimisation. Some tentative suggestions were made to try and explain these differences and some of these seem to square with findings made by others who have evaluated similar schemes, particularly with respect to problems implementing burglary reduction schemes in student neighbourhoods. By showing that other types of neighbourhood also appear to respond poorly to target hardening, the LCCS helps bring to the fore questions about geographies of value for money. Politically, such arguments may be difficult, but then BRIL delivery is already biased by the ring-fenced nature of some of its funding streams.

Finishing on this issue of economics, an unforeseen side-effect of the BRIL evaluation design was the ability to estimate the likely number of burglaries that may have been prevented as a result of target hardening. Burden (2005) has already suggested this was would be a way forward in estimating whether the savings as a result of BRIL outweighed the costs, and after some re-calculation of the economic costs of burglary dwelling offences this was deemed to be the case. Further breakdown of cost savings by group also demonstrated that BRIL might have helped 'narrow the gap' between the most disadvantaged neighbourhoods and the rest of the city - a key long-term policy objective of Leeds City Council and its various partnerships (Leeds Initiative, 2004). Of all the analysis undertaken during the case studies, it is argued this cost-benefit analysis is the most interesting and valuable from a community safety policy perspective.

Considering all of the outcomes from the initial research objectives, it is argued that the main aim - *to contribute to the debate surrounding the usefulness of area classifications for crime pattern analysis at a neighbourhood scale* - has been achieved. In particular, the benefit of a task-specific approach to classification for community safety has been demonstrated, and it is argued that this, coupled with the white-box nature of the design, complements the research into crime analysis using commercial consumer segmentation classifications for crime analysis (e.g. Ashby and Longley, 2005; Ashby, 2005; Williamson et al., 2005). Whether the LCCS is the optimum grouping of neighbourhoods for the purposes of community safety is open to question and future work on designing heuristics and stopping rules might yield improved results. In practically demonstrating the LCCS, however, various features of crime and disorder in Leeds for the different groups lend confidence to the final partition. From a practitioner's perspective, the classification is an interesting and revealing tool for analysis, and this, it is argued, accords with the collaborative ethos of the sponsorship that made the research possible.

Some experiments were undertaken that were not included in the thesis and a number of suggestions for further work are considered in Section 9.2. Firstly though, some general conclusions are made about aspects of the classification design, and its potential uses.

9.1.1 Classifying Other Study Areas

Following on from the production of this thesis, an executive summary will be produced for the collaborative partner of the project, Safer Leeds, and it is anticipated that some parts of the work will be prepared for publication in academic journals. It is also planned to present the findings to the practitioner community, in the first instance at a regional level meeting of crime pattern analysts and

Home Office researchers. The main reason to publish the findings in this manner is to allow others to judge for themselves the merits of the approach, and also to provide sufficient guidance to anyone who might be interested in creating a similar classification for a different study area.

How well other local authorities, for example, might partition is an interesting prospect. Each is unique, and there are large variations in the internal geographies and sizes of populations. Evidence of this was provided by the unfortunate publication in May 2006 of the Urban Crime Rankings (Haldenby and Gibbs, 2006), by Reform, which failed to match denominator population geography with police geography, producing erroneous crime rates and very misleading results.

It seems possible that some local authority districts may classify for community safety more easily than others. There is no firm evidence for this notion, but observing how classifications with countrywide coverage group different authorities hints that a district such as Leeds - with a large continuous conurbation - might partition more easily than a district such as neighbouring Wakefield, for example, where the central city is quite small and five other towns make up the majority of the population. Indeed, considering how national classifications of small areas, such as that designed by Vickers (2005), partition different districts into different numbers of dominant neighbourhood types might be a recommended first step for anyone wanting to replicate the LCCS.

9.1.2 Neighbourhood Dynamics

It is hoped that others will experiment with producing classifications for other study areas, but a second general conclusion is that consideration should also be given to the way in which the dynamic aspects of neighbourhoods might be captured and analysed.

The original intention was that the variables used in the construction of the classification would comprise a mixture of variables capturing aspects of neighbourhoods that tend to remain fairly constant, e.g. ethnic mix, as well as more dynamic aspects, e.g. house price inflation or new business start-ups. On reflection, the number of dynamic variables finally used was small, resulting in a classification that might not change very much if the cluster analysis was repeated in a few years time. Indeed, those values derived from variables in the census would not change at all. This effect can also be observed in the original Home Office 'families' classifications of CDRPs, although here the much larger geographic scale makes group stability much more likely anyway.

Nevertheless, on a number of occasions during the thesis it has been shown the crime rates have changed significantly over time, and that the size, and sometimes direction, of the change has varied between groups. This raises the question of whether a classification for community safety ought actually to be more responsive to changes in crime trends and less a reflection of those aspects of neighbourhood that remain fairly constant over time.

Examples of such classifications, or more accurately, indices, include the Index of Deprivation (IoD) produced for the (former) Office of the Deputy Prime Minister, and a local index produced for strategic housing policy in Leeds, called the Neighbourhood Orientated Model of Area Demand (NOMAD). Both of these are designed to be re-created every year or so, and each attributes a score, or scores, to an area using a model that draws upon a number of dynamic variables. In NOMAD, for example, area demand for housing is calculated based upon variables such as the number of applicants per local authority (LA) housing vacancy, LA housing turnover, domestic burglary rates and average house prices. Furthermore, nine out of the ten variables are produced locally and can be refreshed annually.

Although indices such as these cannot 'classify' neighbourhoods in the same way that the LCCS does, what they do offer is the ability to analyse change over time much more easily. For example, according to the area demand score, NOMAD attributes each council housing estate with an adjective classification of either *area in significant decline, area on the edge, popular areas with specific problems* and *popular areas*. Users of NOMAD may analyse change in this classification over time and/or change in the scores and ranking positions. In this way it is possible to analyse neighbourhood dynamics, something that from a community safety policy perspective might be very desirable, and which if designed accordingly, might also be useful for some other policy areas.

9.1.3 Interpretation Issues

The final concluding remarks concern interpretation issues that arise when applying the LCCS to crime pattern analysis. Firstly, I would argue that the decision not to assign descriptive labels to groups was the right one. At times, this choice has led to some rather awkward descriptions of group differences, but I would hold that this is preferable to using permanent labels which, if they are to be 'catchy', run the risk of tending towards over-simplification and insincerity. A counter-argument could be that when the geographical extent of the classification is as localised as a local authority, there is more scope for choosing labels which are more locally sensitive and appropriate than some of those used by country-level classifications. In the end, the decision rests with individual classification designers whether to attach labels to groups or not, but if the classification is to be used for policy delivery, then persons choosing labels should expect some decisions to be challenged, not least by elected members for the areas concerned.

Secondly, there is the issue of ascribing explanations to crime problems in neighbourhoods based solely of group membership and/or prominent features of the group profile. For example, in the analysis of domestic burglary and the BRIL intervention, a number of the patterns in burglary were said to 'make common sense' - lending support to the notion that the classification is a valid model of neighbourhood type for Leeds. However, interpretations of this sort are problematic when based upon limited experience of the problem domain. Moreover, it is easier to fall into a trap when a group is defined strongly by a unique characteristic, with the findings pertaining to the group containing ethnically diverse populations (group 3) being one example. In such a situation, it might be tempting for the analyst to make inappropriate cause and effect assumptions based upon ethnicity rather than other potentially important issues, such as offender residence. More experienced practitioners may be less likely to make these mistakes, but where crime problems are poorly understood there might be a danger that the generalised nature of the classification leads toward a too-general interpretation of the problem. For this reason, it is argued that the LCCS, and indeed many other analytical tools - whether geographical or not - should not be used in isolation but compared with findings from alternative approaches to problem analysis, both quantitative and qualitative.

9.2 Possibilities for Further Work

Although a lot of time and effort has been put into developing and testing the LCCS, there remain further analysis tasks that could be conducted with the classification as it stands. In addition, there are a number of changes to the classification design that could be considered if the classification were to be refreshed for Leeds or created for the first time for another study area.

9.2.1 Existing Classification

Firstly, some initial experiments have already been conducted to examine what effect the isolation of a neighbourhood from its group peers may have on crime rates (Section 6.5). The results of tests on burglary dwelling produced some fairly strong inverse relationships, suggesting that if a neighbourhood from a higher-crime group is isolated among neighbourhoods from lower-crime groups the crime rate will often be lower than the higher-crime group mean. Results from tests for other crime types were either not statistically significant or else the relationships were weak. However, the potential usefulness of improving our understanding of isolation effects might warrant further work, paying close attention to the crime types chosen and the size of the recorded crime samples - perhaps by aggregating offence data from a number of years. As elsewhere, Leeds is planning some very large re-developments of existing housing estates with a view to creating a more heterogeneous tenure mix than currently exists. The relevance of tenure mix on crime is contested (e.g. Murie, 1997), but an empirical study based on Leeds data might be useful baselining for future evaluations of the regeneration of Little London and the £1billion east and south-east Leeds regeneration scheme, EASEL.

A second piece of further analysis that would have potential benefit to community safety policy developers would be to explore the hypothesis thrown up by the evaluation of BRIL (Chapter 8) that the effect of community safety initiatives varies by neighbourhood type. It has already been demonstrated for burglary dwelling, but further work needs to be undertaken to try and explain the reasons for the differences. The field of cultural criminology was not extensively reviewed for this thesis, but it might provide a starting point for an investigation into cultural variations in response to crime prevention.

It would also be helpful to look for similar variations with respect to other types of initiative. Take-up and activity levels among Neighbourhood Watch schemes might be worth investigating, as these schemes are very numerous and have good coverage across Leeds. Problems with studying Neighbourhood Watch, however, include a paucity of reliable data on activity within these groups, and a certain amount of scepticism among some community safety officers regarding the usefulness of Neighbourhood Watch. Another alternative could be to work with Victim Support to look at repeat victimisation following advice and support. The benefit of a study of this sort is that is is clear when a member of the public has received help (as with BRIL), and the Victim Support caseloads are recorded electronically. Problems that might be encountered could be with regard to concerns of victim privacy and also considering the extent to which advice can be effective given that the circumstances of some crimes mean repeat victimisation is likely to occur unless very considerable measures are taken by the victim - in cases of domestic abuse, for example.

9.2.2 New Classifications

A number of other pieces of further work are suggested for situations where classifications for new study areas are being considered, or if a refresh of the LCCS were to be conducted. Firstly, issues regarding the geography used for the LCCS have already been discussed, and my recommendation is to use the super output area (SOA) lower-level geography if neighbourhoods of a similar average size to those in the LCCS are required. The further work that this requires is in the routine geocoding of datasets to the SOA geography. Developments for national datasets are likely to be outside the scope of influence by local practitioners (although there is no harm in lobbying ONS), but local officers can routinise the geocoding of locally held datasets. This is not necessarily something that needs to be researched, it just needs doing, but there may be situations where research is needed to ensure the sample sizes produced by the geographic scale are appropriate and that disclosure control is given some consideration.

The only other aspect of neighbourhood boundaries that might be worthy of further work would be the design of a set of zones that discounted uninhabited space. This might mean that some isolated dwellings became omitted but that neighbourhoods that border open spaces are represented more accurately, particularly with respect to population density-related variables. It would be recommended that a settlement-clipped geography such as this only be attempted if sufficient data existed at a household or individual scale, as simple aggregation could then be used to create data values for neighbourhoods. This might not be a good idea given the current level of reliance on aggregate census data, although different methods of harmonising a settlement geography with the census output area geography could be tested further.

Research into the possibilities of new variables to capture community safety-related aspects of neighbourhoods is something that is likely to happen as a matter of course as local authorities record more and more information about neighbourhoods in electronic form. I am not convinced whether it would be useful to research the use of more dynamic variables for a classification as the LCCS, but as has been discussed above, such data might be worth collecting and/or modelling if a score-based index of crime and disorder problems were to be constructed.

Another factor affecting the choice of variables, however, is the scope of the definition of community safety. At present it is taken to reflect problems that might befall people due to the malicious acts of others, but in addition to this, all responsible CDRP authorities also respond to problems which are accidental in nature. The definitions of community safety in other countries (e.g. Australia) already reflect accidents and other problems that affect peoples' safety, such as wild fires. Thus, there may be value in researching the possibility of collecting data on accidental fires, road traffic and other accidents to see whether incidents occur in sufficient numbers and whether there are distinct geographies to these problems that might help in discriminating between neighbourhoods.

While the previous suggestion for further work would broaden the scope of the classification but keep it task-specific to community safety, the final suggestion is to research the creation of a more general-purpose classification. This might be appropriate as a complement to the development of a ranked community safety index, but it is not proposed to duplicate the efforts by the likes of ONS. Instead, it is suggested that a local classification could be constructed as general-purpose to the extent that it reflects priorities within a district's LAA (and/or voluntary Local Public Service Agreement (LPSA)).

An LAA is a three year agreement, based on local Sustainable Community Strategies, that sets out the priorities agreed between central government and the local authority and other key partners through Local Strategic Partnerships (LSPs). At present, each LAA is structured around four blocks (or policy fields): children and young people, safer and stronger communities, healthier communities and older people, and economic development and enterprise. Within each block are a number of intended outcomes, each of which have indicators attached for which baselines and annual targets have been calculated. For example, the Leeds LAA safer and stronger block has an outcome to reduce the number of people killed or seriously injured on the roads in Leeds. The indicator that is being used is the number of people killed as recorded by the police in the STATS19 data, itself a Best Value indicator.

Among the effects of the LAA are greater emphasis on partnership working to tackle a wide range of problems, not just community safety. Protectionism and vested interests are ameliorated to some extent by having pooled budgets, and some outcomes have stretched targets associated with them that, if met, attract a reward grant from central government. For community safety and other analysts working to support the delivery the LAA, gathering information and intelligence is key to helping deliver outcomes and monitor performance, particularly where the outcomes have stretch targets! Barriers to data are thus becoming easier to break down.

In conclusion, the LAA raises the possibility for the classifier of neighbourhoods to have access to sets of data that address a broad range of local priorities and that are increasingly well known by all local authority officers and partners. Establishing neighbourhood type according to LAA outcome indicators might help local authorities to appreciate better the geographies of neighbourhoods in need of extra resources to meet targets. Similarly, joint analysis using a classification of this sort might help to reduce the risk that agencies responsible for particular outcomes each identify their own hotspots independently, without consideration of how they might correspond with hotspots identified by others. Practical experiences of the author suggest that such joined-up analysis is perhaps rather hopeful for some local authorities, but a classification of this sort may be useful in the future, in which case further research might be appropriate now.

Appendix A

Maps of Sub-Community Area (SCA) Geography

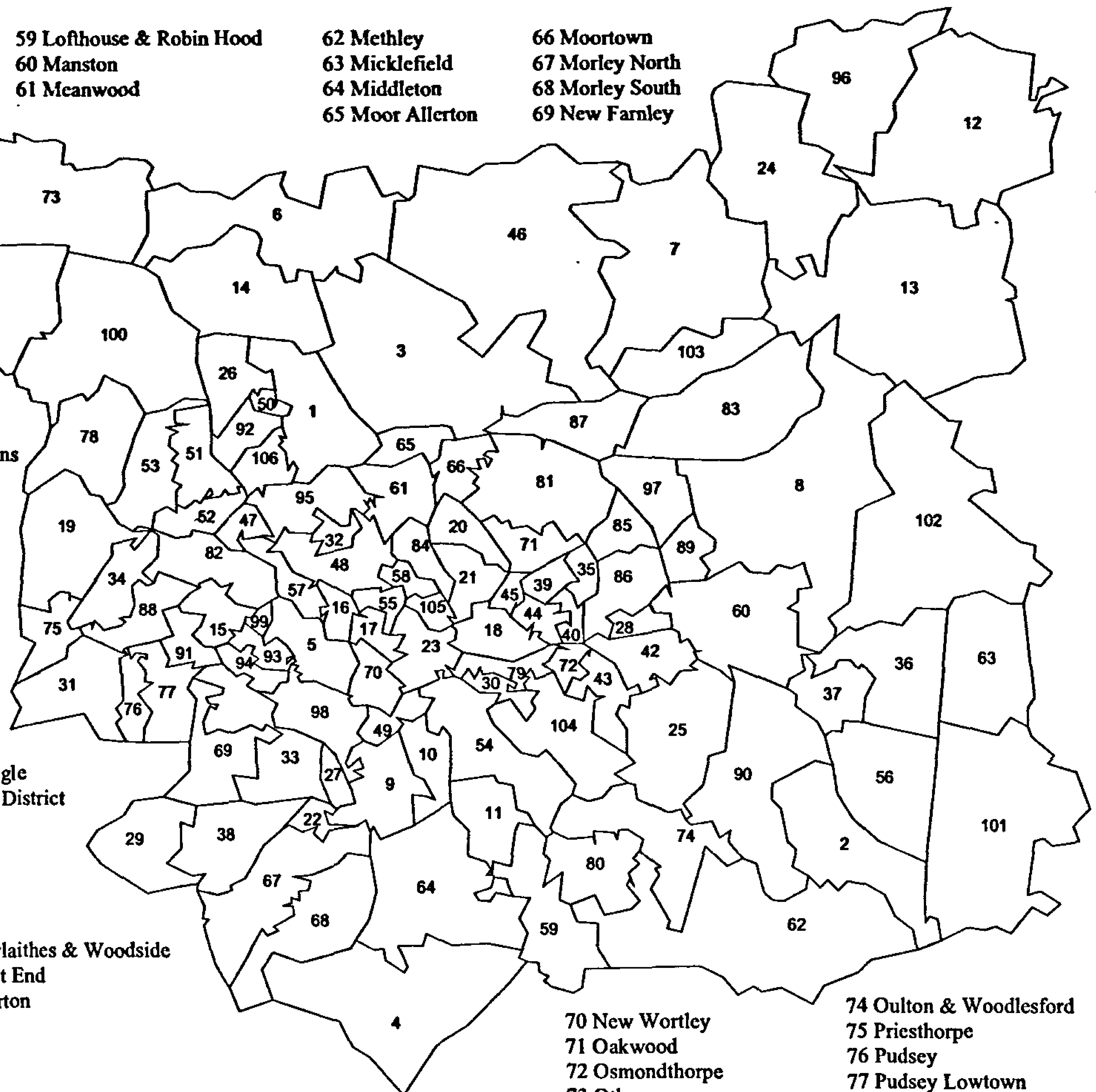


Figure A.1: Original geography of 106 communities (generalised), derived from the Leeds City Council community mapping project.

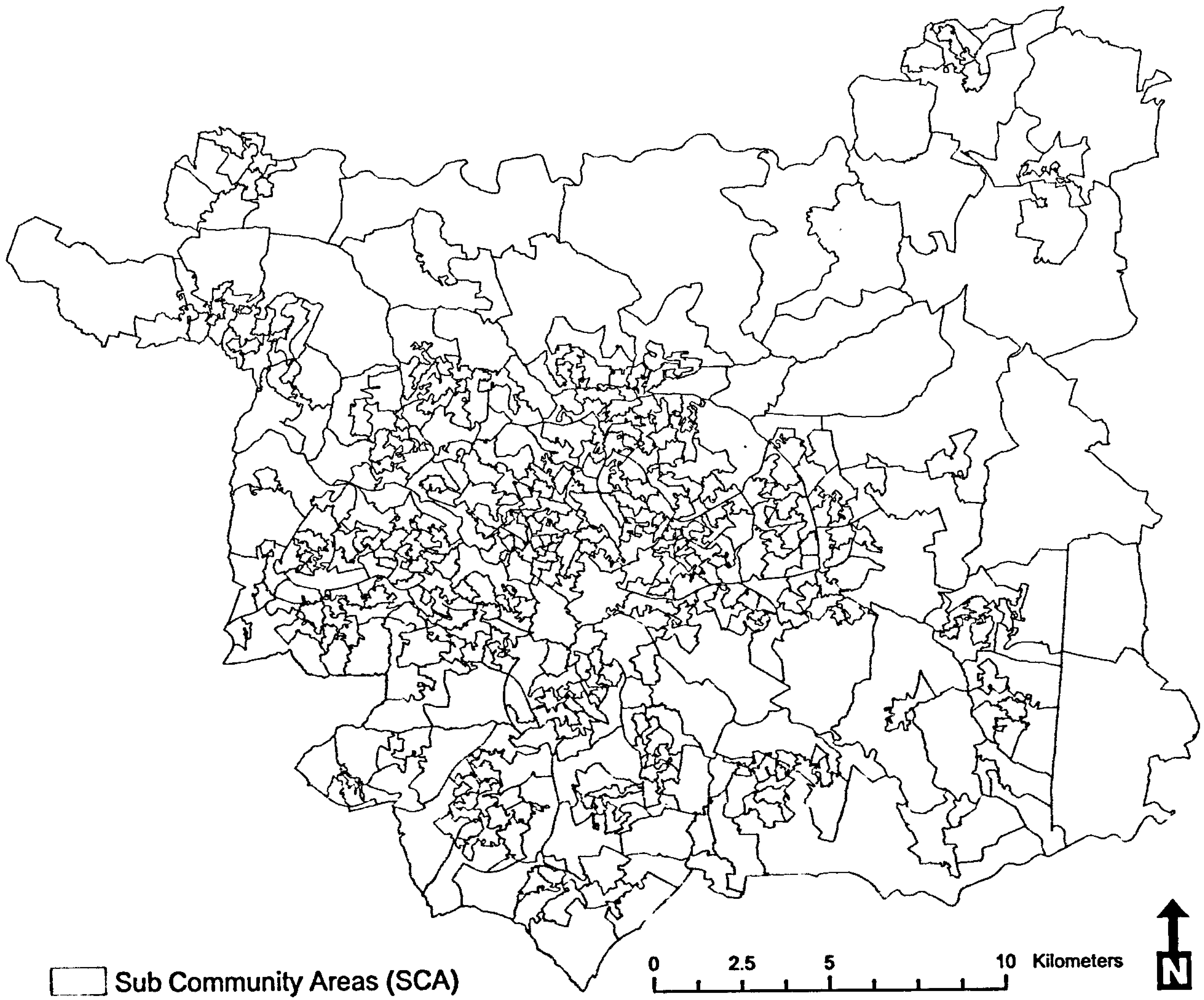


Figure A.2: Sub Community Area (SCA) geography (not generalised).

Appendix B

Full List of Variables Used

Category	Id	Variable	Source
Community cohesion	CC01	Proportion of working residents who work from, or less than 2km from, home	2001 Census
Community cohesion	MG02	People who lived at same address one year prior to census day	
Demography	AG01	Population aged 0-4	2001 Census
Demography	AG02	Population aged 5-14	2001 Census
Demography	AG03	Population aged 15-24	2001 Census
Demography	AG04	Population aged 25-44	2001 Census
Demography	AG05	Population aged 45-64	2001 Census
Demography	AG06	Population aged 65-84	2001 Census
Demography	AG07	Population aged 85+	2001 Census
Demography	AS01	Population who are asylum seekers (summer 2004)	LCC
Demography	ET01	Ethnic minorities: not white British or Irish	2001 Census
Demography	MG01	People who moved to area from outside UK	2001 Census
Demography	PD01	Daytime to residential population ratio	2001 Census
Demography	ET02	Ethnic heterogeneity	2001 Census
Economic	CA01	Cars at or near the workplace	2001 Census
Economic	CA02	Cars at their residential address	2001 Census
Economic	CL01	Population who are claiming employment related benefits in April 2004	Nomis
Economic	CL02	Change in the claimant count rates from April 1997 to April 2004	Nomis
Economic	LU01	Public houses	OS MasterMap
Economic	QU01	Population lowly qualified: no or 1	2001 Census
Economic	QU02	Population highly qualified: 4/5	2001 Census
Economic	TE01	Households owned	2001 Census
Economic	TE02	Households social rented	2001 Census
Economic	TE03	Households private rented	2001 Census
Economic	JB01	Difference in workplace jobs from 1998 to 2002	ABI

Health	AL01	Hospital episodes involving alcohol	PCT
Health	DR01	Hospital episodes involving narcotics	PCT
Health	MI02	Discarded needles collected	LCC
Land use	DS01	Average area of multi surface (garden) per household (sq m)	OS MasterMap
Land use	FP02	Buildings of 29 to 150 sq.m.	OS MasterMap
Land use	FP03	Buildings of 150 to 1000 sq.m.	OS MasterMap
Land use	FP04	Buildings of 1000 to 3000 sq.m.	OS MasterMap
Land use	FP05	Buildings of 3000 to 8000 sq.m.	OS MasterMap
Land use	FP06	Buildings of 8000 + sq.m.	OS MasterMap
Land use	HD01	Household density	2001 Census
Land use	HT01	Houses that are detached	2001 Census
Land use	HT02	Houses that are semi-detached	2001 Census
Land use	HT03	Houses that are terraced	2001 Census
Land use	HT04	Houses that are flats: block	2001 Census
Land use	HT05	Houses that are flats: shared house	2001 Census
Land use	HT06	House type heterogeneity	2001 Census
Land use	NL01	Area covered in natural land	OS MasterMap
Land use	RA01	Area covered by railway tracks and land	OS MasterMap
Land use	RE01	Ratio of residential to non-residential delivery points	OS MasterMap
Land use	RO01	Area covered in roads	OS MasterMap
Land use	VO01	Households unoccupied	2001 Census
Minor incivilities	ED01	Days lost per pupil to fixed term exclusions	LCC
Minor incivilities	ED02	Pupils receiving fixed-term exclusions	LCC
Minor incivilities	ED03	Pupils permanently excluded	LCC
Minor incivilities	FI01	Deliberate secondary fires	WYFRS
Minor incivilities	FI02	Fire hoax calls attended	WYFRS
Minor incivilities	MI03	Noise complaints	LCC
Minor incivilities	MI04	Abandoned cars recovered	LCC
Minor incivilities	MI05	Disorder incidents	WYP
Minor incivilities	MI06	Traffic incidents	WYP
Minor incivilities	MI07	Anti-social behaviour incidents	WYP
Social	ED04	GCSE students only getting 0,1 or 2 passes	LCC
Social	ED05	Average number of GCSE passes A* to C	LCC
Social	HH01	One person households	2001 Census
Social	HH02	Lone parent households	2001 Census
Social	SC01	People who are NS-SeC 1	2001 Census
Social	SC02	People who are NS-SeC 2	2001 Census
Social	SC03	People who are NS-SeC 3	2001 Census
Social	SC04	People who are NS-SeC 4	2001 Census
Social	SC05	People who are NS-SeC 5	2001 Census
Social	SC06	People who are NS-SeC 6	2001 Census
Social	SC07	People who are NS-SeC 7	2001 Census

Social	SC08	People who are NS-SeC 8	2001 Census
Social	SC09	People who are full-time students	2001 Census

Table B.1: Variables used during the design of the classification (not all of which were included in the final cluster analysis). Abbreviations (ABI=Annual Business Inquiry, LCC=Leeds City Council, PCT=Primary Care Trusts, WYFRS=West Yorkshire Fire and Rescue Service, WYP=West Yorkshire Police)

Bibliography

6, P., Raab, C., and Bellamy, C. (2005). Joined-up government and privacy in the United Kingdom: Managing tensions between data protection and social policy. Part I. *Public Administration*, 83(1):111–133.

Akers, R. (1990). Rational choice, deterrence, and social learning theory in criminology: the path not taken. *The Journal of Criminal Law and Criminology*, 81(3):653–676.

Allen, J., Dodd, T., and Salisbury, H. (2005). Crime in England and Wales: Quarterly Update to September 2004. Statistical Bulletin 03/05, Home Office, London.

Allen, T. and Cohen, S. (1969). Information flow in research and development laboratories. *Administrative Science Quarterly*, 14(1):12–19.

Alvanides, S. (2000). *Zone Design Methods for Application in Human Geography*. Unpublished PhD Thesis, School of Geography, University of Leeds, Leeds.

Alvanides, S., Openshaw, S., and Rees, P. (2002). Designing your own geographies. In Rees, P., Martin, D., and Williamson, P., (Eds.), *The Census Data System*. Wiley, Chichester.

Anderberg, M. (1973). *Cluster Analysis for Applications*. Academic Press, New York.

Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, 27(2):93–115.

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In Lonley, P., Goodchild, M., Maguire, D., and Rhind, D., (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*, pages 251–264. Wiley.

Anselin, L., Syabri, I., and Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, 38(1):5–22.

Ashby, D. (2005). Policing neighbourhoods: Exploring the geographies of crime, policing and performance assessment. *Policing and Society*, 15(4):413–447.

Ashby, D. and Longley, P. (2005). Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS*, 9(1):53–72.

Audit Commission (1999). *Best Assured: The Role of the Audit Commission in Best Value*. Audit Commission, London.

Bailey, T. and Gatrell, A. (1995). *Interactive Spatial Data Analysis*. Longman Scientific and Technical, Harlow.

Baldwin, J. (1979). Ecological and areal studies in Great Britain and the United States. In Morris, N. and Tonry, M., (Eds.), *Crime and Justice: An Annual Review of Research Vol. 1*, pages 29–66. University of Chicago Press, Chicago.

Baldwin, J. and Bottoms, A. (1976). *The Urban Criminal: A Study in Sheffield.* Tavistock, London.

Batey, P. and Brown, P. (1994). Design and construction of geodemographic targeting systems: what have we learnt in the last ten years? *Journal of Targeting, Measurement and Analysis for Marketing*, 3(2):105–115.

Batey, P. and Brown, P. (1995). From human ecology to customer targeting: the evolution of geodemographics. In Longley, P. and Clarke, G., (Eds.), *GIS for Business and Service Planning*, pages 77–103. GeoInformation International, Cambridge.

Baumgartner, M. (1988). *The Moral Order of a Suburb.* Oxford University Press, New York.

Beaney, A. (2003). Partnership Business Model (PBM). Presentation given as part of the Home Office Good Practice Seminars series, November 2003, London. Slides available online at http://www.crimereducction.co.uk/gppw02.ppt.

Becker, H. (1963). *The Outsiders.* Free Press, New York.

Bellamy, C., 6, P., and Raab, C. (2005). Joined-up government and privacy in the United Kingdom: Managing tensions between data protection and social policy. Part II. *Public Administration*, 83(2):393–415.

Bennett, T. (1996). What's new in evaluation? A note on the Pawson and Tilley article. *British Journal of Criminology*, 36(4):567–573.

Bennett, T. and Wright, R. (1984). *Burglars on Burglary.* Gower, Aldershot.

Berry, W. and Jones, H. (1995). Investigating spatial patterns of crime through police statistics, crime surveys and census profiles: findings from Dundee housing estates. *Scottish Geographical Magazine*, 111(2):76–82.

Besag, J. and McNeil, D. (1976). On the use of exploratory data analysis in human geography. *Advanced Applied Probability*, 8(4):652.

Bezdek, J. (1987). Some non-standard clustering algorithms. In Legendre, P. and Legendre, L., (Eds.), *Developments in Numerical Ecology*, pages 225–287, Berlin. Springer-Verlag. Proceedings of the NATO Advanced Workshop on Numerical Ecology, Brittany, France, June 1986.

Birkin, M. (1995). Customer targeting, geodemographics and lifestyle approaches. In Longley, P. and Clarke, G., (Eds.), *GIS for Business and Service Planning*, pages 104–149. GeoInformation International, Cambridge.

Birkin, M., Clarke, G., and Clarke, M. (2002). *Retail Geography and Intelligent Network Planning.* Wiley, Chichester.

Bithell, J. (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine*, 9:691–701.

Blake, M. and Openshaw, S. (1995). Selecting variables for small area classifications of 1991 UK Census Data. Working Paper No. 95/5, School of Geography, University of Leeds.

Blau, P. (1977). *Inequality and Heterogeneity: A Primitive Theory of Social Structure.* Free Press, New York.

Bloomfield, B. and Coombs, R. (1992). Information technology, control and power: The centralization and decentralization debate revisited. *Journal of Management Studies*, 29(4):459–484.

Bottoms, A. and Wiles, P. (1986). Housing tenure and residential community crime careers in Britain. *Crime and Justice*, 8:101–162.

Bottoms, A. and Wiles, P. (2002). Environmental criminology. In Maguire, M., Morgan, R., and Reiner, R., (Eds.), *The Oxford Handbook of Criminology.* The Oxford University Press, Oxford.

Bottoms, A. and Xanthos, P. (1981). Housing policy and crime in the british public sector. In Brantingham, P. and Brantingham, P., (Eds.), *Environmental Criminology*, pages 203–225. Waveland Press, Prospect Heights, Illinois.

Bowers, K. and Hirschfield, A. (1999). Exploring links between crime and disadvantage in north-west England: an analysis using geographical information systems. *International Journal of Geographical Information Science*, 13(2):159–184.

Bowers, K., Johnson, S., and Pease, K. (2004). Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*, 44(5):641–658.

Bowles, R. and Pradiptyo, R. (2004). Reducing burglary initiative: an analysis of costs, benefits and cost effectiveness. Home Office Online Report 43/04, Home Office, London.

Brand, S. and Price, R. (2000). The economic and social costs of crime. Home Office Research Study 217, Home Office Research, Development and Statistics Directorate, London.

Brantingham, P. and Brantingham, P., (Eds.) (1981). *Environmental Criminology.* Sage Publications, Beverly Hills.

Brantingham, P. and Brantingham, P. (1993). Nodes, paths and edges: considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, 13:3–28.

Brewer, C. and Suchan, T. (2001). Mapping Census 2000: The Geography of U.S. Diversity. U.S. Census Bureau, Census Special Reports Series CENSR/01-1, U.S. Government Printing Office, Washington, D.C.

Bridgeman, C. and Hobbs, L. (1997). Preventing repeat victimisation: the police officer's guide. , Home Office, London.

Brimicombe, A. (2004). On being more robust about 'hot spots'. Paper presented to the Seventh Annual International Crime Mapping Research Conference. March 31 - Arpil 3, Boston, MA. Paper available online at http://www.ojp.usdoj.gov/nij/maps/boston2004/papers.html.

Brown, P. (1991). Exploring geodemographics. In Masser, I. and Blakemore, M., (Eds.), *Handling Geographical Information*, pages 221–258. Longman, Harlow.

Brown, P., Hirschfield, A., and Batey, P. (2000). Adding value to census data: Public sector applications of the Super Profiles geodemographic typology. *Cities and Regions*, GIS Special Nov 2000:19–31.

Brunsdon, C. (1989). *Spatial Analysis Techniques Applied to Local Crime Patterns.* Unpublished PhD Thesis, University of Newcastle upon Tyne.

Brunsdon, C. (2001). The comap: exploring spatial pattern via conditional distributions. *Computers, Environment and Urban Systems*, 25:53–68.

Brusco, M. and Cradit, J. (2001). A variable-selection heuristic for k-means clustering. *Psychometrika*, 66(2):249–270.

Burden, T. (2005). Evaluation of Burglary Reduction Unit: BRIL Phase II Programme. , Policy Research Institute, Leeds Metropolitan University.

Burgess, J. (1996). Focusing on fear: The use of focus groups in a project for the Community Forest Unit, Countryside Commission. *Area*, 28(2):130–135.

Burt, J. and Barber, G. (1996). *Elementary Statistics for Geographers.* Guilford Press, New York.

Cahill, M. (2005). Accounting for Spatial Patterns of Crime in Statistical Analyses: An example from Portland, OR. Presentation to Eighth Annual Crime Mapping Research Conference, September 7 - 10, Savannah.

Cameron, J. and Leitner, M. (2005). Spatial analysis tools for identifying hotspots. In *Mapping Crime: Understanding Hotspots*, pages 35–63. National Institute of Justice, Washigton, D.C.

Campbell, B. (1996). Gender, crisis and community. In Kraemer, S. and Roberts, S., (Eds.), *The Politics of Attachment: Towards a Secure Society*, pages 102–109. Free Association Books, London.

Campbell, D. and Stanley, J. (1963). *Experimental and Quasi-Experimental Evaluations in Social Research.* Rand McNally, Chicago.

Canter, D. (1977). *The Psychology of Place.* The Architectural Press, London.

Carmone, F., Kara, A., and Maxwell, S. (1999). HINoV: A new model to improve market segment definition by identifying noisy variables. *Journal of Marketing Research*, 36:501–509.

Carr, D., White, D., and MacEachren, A. (2005). Conditioned choropleth maps and hypothesis generation. *Annals of the Association of American Geographers*, 95(1):32–53.

Carr, J. (2003). *The geographies of young people, crime and social exclusion.* Unpublished PhD Thesis, School of Geography, University of Leeds.

Caspi, A. and Moffitt, T. (1993). When do individual differences matter? A paradoxical theory of personality coherence. *Psychological Inquiry*, 4:247–271.

Cattell, R. (1966). *Hanbook of Multivariate Experimental Psychology.* Rand McNally, Chicago.

Ceccato, V. and Haining, R. (2004). Crime in border regions: The Scandanavian case of Öresund, 1998-2001. *Annals of the Association of American Geographers*, 94(4):807–826.

Ceccato, V., Haining, R., and Signoretta, P. (2002). Exploring offence statistics in Stockholm City using spatial analysis tools. *Annals of the Association of American Geographers*, 92(1):29–51.

Chainey, S. (2005). The police role in community cohesion: Using geographic information to identify vulnerable localities. Presentation given at 3rd National Crime Mapping Conference, London, April.

Chainey, S. and Desyllus, J. (2004). Measuring, identifying and analysing street crime risk. Presentation at the Second National Crime Mapping Conference, March 2004, London.

Chainey, S. and Ratcliffe, J. (2005). *GIS and Crime Mapping*. Wiley, Chichester.

Chainey, S., Reid, S., and Staurt, N. (2002). When is a hotspot a hotspot? A procedure for creating statically robust hotspot maps of crime. In Kinder, D., Higgs, G., and White, S., (Eds.), *Innovations in GIS 9*. Taylor and Francis, London.

Chainey, S. and Smith, C. (2006). Review of GIS-based information sharing systems. Home Office On-line Report 02/06. Available on-line at http://www.homeoffice.gov.uk/rds/pdfs06/rdsolr0206.pdf.

Chakravorty, S. and Pelfrey, W. (2000). Exploratory data analysis of crime patterns. In Goldsmith, V., McGuire, P., Mollenkopf, J., and Ross, T., (Eds.), *Analyzing Crime Patterns: Frontiers of Practice*. Sage, Thousand Oaks, CA.

Chow, J. (1998). Differentiating urban neighbourhoods: A multivariate structural model analysis. *Social Work Research*, 22(3):131–142.

Clark, R. and Lab, S. (2000). Community characteristics and in-school criminal victimization. *Journal of Criminal Justice*, 28:33–42.

Clarke, H., Sanders, D., Stewart, M., and Whitely, P. (2003). Britain (not) at the polls, 2001. *Political Science and Politics*, 36(1):59–64.

Clarke, R. (1980). Situational crime prevention: Theory and practice. *British Journal of Criminology*, 20(2):136–147.

Clarke, R. and Eck, J. (2003). *Become a Problem Solving Crime Analyst in 55 Small Steps*. Jill Dando Institute of Crime Science, London.

Clarke, R. and Mayhew, P., (Eds.) (1980). *Designing Out Crime*. HMSO, London.

Clayton, D. and Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681.

Cleveland, W. (1993). *Visualizing Data*. Hobart Press, New Jersey.

Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation*. Pion, London.

Cloward, R. and Ohlin, L. (1960). *Delinquency and Opportunity*. Free Press, New York.

COCA (2004). Crack in Leeds. Report prepared for leeds community safety partnership, Conference on Crack and Cocaine, London.

Cohen, A. (1955). *Delinquent Boys*. Free Press, New York.

Cohen, J. (1960). A coefficient of agreement for normal scales. *Educational and Psychological Measurement*, 20:37–46.

Cohen, L. and Felson, M. (1979). Social change and crime rate trends: a routine activity approach. *American Sociological Review*, 44(4):588–608.

Cohen, P. (1997). *Rethinking the Youth Question: Education, Labour and Cultural Studies.* Macmillan, Basingstoke.

Cohen, S. (1973). Property destruction: Motives and meanings. In Ward, C., (Ed.), *Vandalism*, pages 23–53. The Architectural Press, London.

Coleman, A. (1985). *Utopia on Trial.* Hilary Shipman, London.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement):95–120.

Collins, D. and Kearns, R. (2001). Under curfew and under siege? Legal geographies of young people. *Geoforum*, 32(3):389–403.

Coombes, M. and Raybould, S. (2000). Policy-relevant surfaced data on population distribution and characteristics. *Transactions in GIS*, 4(4):319–342.

Cooper, A. and Lousada, J. (2000). *The Meaning of Welfare*. Venture Press, London.

Corcoran, J., Wilson, I., and Ware, J. (2003). Predicting the geo-temporal variations of crime and disorder. *International Journal of Forecasting*, 19:623–634.

Cormack, R. (1971). A review of classification. *Journal of the Royal Statistical Society A*, 134:321–367.

Cornish, D. and Clarke, R. (1986). *The Reasoning Criminal: Rational Choice Perspectives on Offending*. Springer, New York.

Cox, N. and Jones, K. (1981). Exploratory data analysis. In Wigley, N. and Bennett, R., (Eds.), *Quantitative Geography*, pages 135–143. Edward Arnold, London.

Craglia, M., Haining, R., and Signoretta, P. (2001). Modelling high-intensity crime areas in English cities. *Urban Studies*, 38(11):1921–1941.

Craglia, M., Haining, R., and Wiles, P. (2000). A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies*, 37(4):711–729.

Crawford, A. (1996). The spirit of community: Rights, responsibilities and the communitarian agenda. *Journal of Law and Society*, 23(2):247–262.

Crawford, A. (1998). *Crime Prevention and Community Safety: Politics, Policies and Practices.* Longman, Harlow.

Crawford, A., Blackburn, S., Lister, S., and Shepherd, P. (2004a). *Patrolling with a Purpose: An Evaluation of Police Community Support Officers in Leeds and Bradford City Centres.* Centre for Criminal Justice Studies, University of Leeds, Leeds.

Crawford, A. and Jones, M. (1996). Kirkholt revisited: Some reflections on the transferability of crime prevention initiatives. *Howard Journal*, 35(1):21–39.

Crawford, A., Lister, S., Blackburn, S., and Burnett, J. (2004b). *Plural Policing: The Mixed Economy of Visible Security Patrols.* Nuffield Foundation, London.

Curry, M. (1995). GIS and the inevitability of ethical inconsistency. In Pickles, J., (Ed.), *Ground Truth*, pages 68–87. The Guilford Press, New York.

Curry, M. (1998). *Digital Places: Living with Geographic Information Technologies.* Routledge, London.

Davis, J. (1986). *Statistics and Data Analysis in Geology.* John Wiley and Sons, New York, 2nd edition.

De Soete, G. (1988). Ovwtre: A program for optimal variable weighting for ultrametric and additive tree fitting. *Journal of Classification*, 5:101–104.

Debenham, J. (2003). *Extending geodemographics: new small area classifications for Yorkshire and the Humber.* Unpublished PhD thesis, School of Geography, University of Leeds.

Deehan, A. (1999). Alcohol and Crime: Taking Stock. Crime Reduction Research Series No. 3, Home Office, London.

Department for Education and Skills (2006). School and college achievement and attainment tables. Online at http://www.dfes.gov.uk/performancetables/.

DeSarbo, W. and Mahajan, V. (1984). Constrained classification: The use of priori information in cluster analysis. *Psychometrika*, 49:187–215.

Dodd, T., Nicholas, S., Povey, D., and Walker, A. (2004). Crime in England and Wales 2003/2004. Home Office Statistical Bulletin 10/04, Home Office, London.

Dorling, D. and Openshaw, S. (1992). Using computer animation to visualize space-time patterns. *Environment and Planning B: Planning and Design*, 19(6):639–650.

Dubes, R. and Jain, A. (1979). Validity studies in clustering methodologies. *Pattern Recognition*, 11(4):235–254.

Dugmore, K. (2004). Accessing geodemographics and classifications for business applications. Paper presented at ESRC Transdisciplinary/Research Methods Seminar Series, UCL, 18-19 February 2004. Summary details retrieved 26 April 2005 from http://www.casa.ucl.ac.uk/lectures/ResearchMethods.html.

Ebdon, D. (1985). *Statistics in Geography.* Blackwell, Oxford.

Eck, J. (1995). A general model of the geography of illicit retail marketplaces. In Weisburd, D. and Eck, J., (Eds.), *Crime and Place*, volume 4 of *Crime Prevention Studies*, pages 67–93. Criminal Justice Press, Monsey, NY.

Eck, J., Chainey, S., Cameron, J., and Wilson, R. (2005). *Mapping Crime: Understanding Hotspots*. National Institute of Justice, Washington, D.C.

Empey, L. (1977). Crime prevention: The fugitive utopia. In Inciardi, J. and Siegel, H., (Eds.), *Crime: Emerging Issues*. Praeger, New York.

Erickson, B. and Nosanchuk, T. (1992). *Understanding Data*. Open University Press, Buckingham.

Etzioni, A. (1993). *The Spirit of Community*. Simon and Schuster, New York.

Etzioni, A. and Lehman, E. (1967). Some dangers in "valid" social measurement. *The Annals of the American Academy of Political and Social Science*, 373:1–15.

Evans, D., Fyfe, N., and Herbert, D., (Eds.) (1992). *Crime, Policing and Place: Essays in environmental criminology*. Routledge, London.

Everitt, B., Landau, S., and Lesse, M. (2001). *Cluster Analysis*. Arnold, London.

Everitt, J. and Cadwallader, M. (1972). The home area concept in urban analysis: the use of cognitive mapping and computer procedures as methodological tools. In Mitchell, W., (Ed.), *Environmental Design: Research and Practice*, pages sections 1.2.1–1.2.10. University of California, Los Angeles.

Everson, S. and Pease, K. (2001). *Crime against the same person and place: Detection opportunity and offender targeting*, volume 12 of *Crime Prevention Studies*, pages 199–220. Criminal Justice Press, Monsey, NY.

Fahmy, E., Gordon, D., and Cemlyn, S. (2002). Poverty and neighbourhood renewal in West Cornwall. Paper presented at the Social Policy Association Annual Conference, Nottingham, August.

Farrell, G. and Pease, K. (1994). Once bitten, twice bitten: Repeat victimisation and its implications for crime prevention. Crime prevention unit paper 46, Home Office, London.

Farrington, D. (1992). Explaining the beginning, progress and ending of antisocial bevaiour from birth to adulthood. In McCord, J., (Ed.), *Facts, Frameworks and Forecasts: Advances in Criminological Theory*, volume 3, pages 253–286. Transaction, New Brunswick, NJ.

Farrington, D., Gallagher, B., Morley, L., St. Ledger, R., and West, D. (1986). Unemployment, school leaving and crime. *British Journal of Criminology*, 26:335–356.

Faupel, C. (1987). Heroin use and criminal careers. *Qualitative Sociology*, 10(2):115–131.

Felson, M. (1995). Those who discourage crime. In Weisbur, D. and Eck, J., (Eds.), *Crime and Place*, volume 4 of *Crime Prevention Studies*, pages 53–66. Criminal Justice Press, Monsey, NY.

Felson, M. and Clarke, R. (1998). Opportunity makes the thief: Practical theory for crime prevention. Police Research Series Paper 98, Home Office, London.

Ferrie, J., shipley, M., Marmot, M., Stansfeld, S., and Davey-Smith, G. (1995). Health effects of anticipation of job change and non-employment: Longitudinal data from the Whitehall II study. *British Medical Journal*, 311:1264–1269.

Flowerdew, R. and Leventhal, B. (1998). Under the microscope. *New Perspectives*, 18:36–38.

Flynn, N. (1997). *Public Sector Management*. Prentice Hall, London, Third edition.

Forrest, R. and Kearns, A. (1999). *Joined-up Places? Social Cohesion and Neighbourhood Regeneration*. Joseph Rowntree Foundation, York.

Fowler, F. (1967). Urban renewal, 1918-1966. In Beresford, M. and Jones, G., (Eds.), *Leeds and its Region*, pages 175–185. British Association for the Advancement of Science (Leeds Local Executive Committee), Leeds.

Fowlkes, E., Gnanadesikan, R., and Kettenring, J. (1987). Variable selection in clustering and other contexts. In Mallows, C., (Ed.), *Design, Data, and Analysis: By Some Friends of Cuthbert Daniel*. Wiley, New York.

Fowlkes, E. and Mallows, C. (1988). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78:553–569.

Fox, C. and McManus, J. (2001). The NACRO Guide to Crime Audits. Community Safety Practice Briefing, NACRO, London.

Fyfe, N. (2000). Geography of crime. In Johnston, R., Gregory, D., Pratt, G., and Watts, M., (Eds.), *The dictionary of Human Geography*, pages 120–123. Blackwell, Oxford, Fourth edition.

Garland, D. (1988). British criminology before 1935. *British Journal of Criminology*, 28(2):1–17.

Garofalo, J. and Laub, J. (1978). The fear of crime: Broadening our perspective. *Victimology*, 3:242–253.

Gatrell, A. (1989). On the spatial representation and accuracy of address-based data in the United Kingdom. *International Journal of Geographical Information Systems*, 3(4):335–348.

Gatrell, A., Bailey, T., Diggle, P., and Rowlingson, B. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21:256–274.

Geary, R. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5:115–145.

Getis, A. and Ord, J. (1996). Local spatial statistics: an overview. In Longley, P. and Batty, M., (Eds.), *Spatial Analysis: Modelling in a GIS Environment*. GeoInformation International, Cambridge.

Gibbons, D. (1994). *Talking About Crime and Criminals*. Prentice-Hall, Englewood Cliffs, NJ.

Gill, O. (1977). *Luke Street: Housing Policy, Conflict and the Creation of a Delinquent Area*. Macmillan, London.

GMAC (2005). Greater Manchester Against Crime. Information available online at http://www.gmac.org.uk/.

Gnanadesikan, R., Kettering, J., and Tsao, S. (1995). Weighting and selection of variables for cluster analysis. *Journal of Classification*, 5:113–136.

Gordon, A. (1999). *Classification*. Chapman and Hall, Boca Raton.

Goss, J. (1995). Marketing the new marketing: the strategic discourse of geodemographic information systems. In Pickles, J., (Ed.), *Ground Truth: The Social Implications of Geographic Information Systems*, pages 130–170. The Guilford Press, New York.

Gottfredson, M. and Hirschi, T. (1990). *A General Theory of Crime*. Stanford University Press, Stanford, CA.

Graham, J. (1988). *Schools, disruptive behaviour and delinquency*. Her Majesty's Stationary Office, London.

Great Britain (1998). Crime and Disorder Act. , HMSO, London.

Greenberg, S., Rohe, W., and Williams, J. (1982). Safety in urban neighbourhoods: A comparison of physical characteristics and informal territorial control in high and low crime neighbourhoods. *Population and Environment*, 5(3):141–165.

Groff, E. and La Vigne, N. (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies*, 13:29–57.

Gross, B. (1965). The social State of the Union. *Trans-action*, 3(1):14–17.

Guo, D., Peuquet, D., and Gahegan, M. (2003). ICEAGE: Interactive clustering and exploration of large and high-dimensional geodata. *GeoInformatica*, 7(3):229–253.

Haining, R. and Cliff, A. (2003). Using a scan statistic to map the incidence of an infectious disease: Measles in the USA 1962-1995. In Toubiana, L. ans Viboud, C., Fhahault, A., and Valleron, A., (Eds.), *Geography and Health*, pages 177–194. Inserm, Paris.

Haining, R., Wise, S., and Blake, M. (1994). Constructing regions for small area analysis: material deprivation and colorectal cancer. *Journal of Public Health Medicine*, 16(4):429–438.

Haldenby, A. and Gibbs, B. (2006). Urban crime rankings. Revised version available on-line at http://www.reform.co.uk/filestore/pdf/UrbanJuly

Hall, R., Vakalopoulou, V., Brunsdon, C., Charlton, M., and Alvanides, S. (2003a). Maintaining basic command unit and crime and disorder partnership families for comparative purposes: 1 april 2002 results. , Home Office, London. Home Office Online Report 39/03. Available at http://www.homeoffice.gov.uk/rds/pdfs2/rdsolr3903.pdf.

Hall, R., Vakalopoulou, V., Brunsdon, C., Charlton, M., and Alvanides, S. (2003b). Maintaining basic command unit and crime and disorder partnership families for comparative purposes: 1 april 2003 results. , Home Office, London. Home Office Online Report 40/03. Available at http://www.homeoffice.gov.uk/rds/pdfs2/rdsolr4003.pdf.

Hammond, R. and McCullagh, P. (1974). *Quantitative Techniques in Geography.* Clarendon Press, Oxford.

Hancock, L. (2001). *Community, Crime and Disorder: Safety and Regeneration in Urban Neighbourhoods.* Palgrave, Basingstoke.

Harper, G., Williamson, I., Clarke, G., and See, L. (2002). Family Origins: Developing Groups of Crime and Disorder Reduction Partnerships and Police Basic Command Units for Comparative Purposes. , Police and Reducing Crime Unit, RDS, Home Office.

Harper, G., Williamson, I., See, L., Emmerson, K., and Clarke, G. (2001). Family Ties: Developing Basic Command Unit Families for Comparative Purposes. Briefing Note 4/01, Home Office, London.

Harradine, S., Kodz, J., Lemetti, F., and Jones, B. (2004). Defining and measuring anti-social behaviour. Development and Practice Report 26, Home Office, London.

Harries, K. (1974). *The Geography of Crime and Justice.* McGraw-Hill, New York.

Harries, K. (1997). Social stress and trauma: Synthesis and spatial analysis. *Social Science and Medicine*, 45(8):1251–1264.

Harries, K. and Brunn, S. (1978). *The Geography of Laws and Justice.* Praeger, New York.

Harris, R. (2001). On the diversity of diversity: is there still a place for small area classifications. *Area*, 33(3):329–336.

Harris, R. and Chen, Z. (2003). Giving dimension to point locations: urban density profiling using population surface models. *Computers, Environment and Urban Systems*, 29:115–132.

Harris, R., Sleight, P., and Weber, R. (2005). *Geodemographics, GIS and Neighbourhood Targeting.* Wiley, Chichester.

Hartigan, J. (1975). *Clustering Algorithms.* Wiley, New York.

Hayden, C., Williamson, T., and Webber, R. (2005). Schools, neighbourhoods and young offenders: can geodemographic analysis inform community programmes and policing. Paper presented at the British Society of Criminology Conference, 12-14 June, Leeds.

Hayward, K. (2004). *City Limits: Crime, Consumer Culture and the Urban Experience.* The GlassHouse Press, London.

Hearnden, I. and Magill, C. (2004). Decision-making by house burglars: Offenders perspectives. Home Office Research Findings 249, Home Office, London.

HEAT (2002). The HEAT Team 12 Month Report: 1st April 2001 - 31st March 2002. , HEAT Arrest Referral Team, Leeds.

Herbert, D. (1976). The study of delinquency areas: A geographical approach. *Transactions of the Institute of British Geographers*, 1:472–492.

Herbert, D. (1982). *The Geography of Urban Crime.* Longman, London.

Herbert, D. (1994). Neighborhood incivilities, crime, place and multilevel models. *Area*, 26(1):81–84.

Hillier, B. and Hanson, J. (1984). *The Social Logic of Space.* Cambridge University Press, Cambridge.

Hirschfield, A. (2004). The Impact of the Reducing Burglary Initiative in the North of England. Home Office Online Report 40/04, Home Office RDS, London.

Hirschfield, A. and Bowers, K. (1997a). The development of a social, demographic and land use profiler for areas of high crime. *British Journal of Criminology*, 37(1):103–120.

Hirschfield, A. and Bowers, K. (1997b). The effect of social cohesion on levels of recorded crime in disadvantaged areas. *Urban Studies*, 34(8):1275–1295.

Hirschfield, A., Bowers, K., and Brown, P. (1995). Exploring relationships between crime and disadvantage on Merseyside: An analysis using crime statistics, census data and geographical information systems. *european Journal on Criminal Policy and Research*, 3(3):93–112.

Hirschfield, A., Yarwood, D., and Bowers, K. (2001). Crime pattern analysis, spatial targeting and gis: The development of new approaches for use in evaluating community safety initiatives. In Clarke, G. and Madden, M., (Eds.), *Regional Science in Business*, pages 323–342. Spinger-Verlag, Berlin.

Hirschi, T. (1969). *Causes of Delinquency.* University of California Press, Berkeley, CA.

Hirschi, T. and Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology*, 89:552–584.

Home Office (1984). Crime prevention. Issued jointly with Department of Education and Science, Department of the Environment, Department of Health and Social Security and the Welsh Office., Home Office, London.

Home Office (1991). *Safer Communities: The Local Delivery of Crime Prevention Through the Partnership Approach.* Home Office, London.

Home Office (1998a). *Guidance on Statutory Crime and Disorder Partnerships.* Home Office, London.

Home Office (1998b). The Crime and Disorder Act: Introductory Guide. Available on-line at http://www.nationalarchives.gov.uk/ERO/records/ho415/1/cdact/index.htm.

Home Office (2001a). *Car Theft Index 2001.* Home Office, London.

Home Office (2001b). *Community Cohesion: A Report of the Independent Review Team Chaired by Ted Cantle.* Home Office, London.

Home Office (2002). *The National Policing Plan, 2003-2006.* Home Office, London.

Home Office (2003a). *Police Performance Monitoring 2001/02.* Home Office, London.

Home Office (2003b). *Police Performance Monitoring 2002/03.* Home Office, London.

Home Office (2004a). Audits and strategies toolkit. Available online at http://www.crimereduction.gov.uk/aud00.htm Accessed 14.11.2005.

Home Office (2004b). *Building Communities, Beating Crime: A better police service for the 21st century.* Home Office, London.

Home Office (2004c). Crimes against students: emerging lessons for reducing student victimisation. Home Office Development and Practice Report 21, Home Office, London.

Home Office (2004d). *Police Performance Monitoring 2003/04.* Home Office, London.

Home Office (2005a). The economic and social costs of crime against individuals and households 2003/04. Home Office Online Report 30/05, Home Office, London.

Home Office (2005b). Focus areas and hotspots toolkit. Available online at http://www.crimereduction.gov.uk/toolkits/fa00.htm.

Home Office (2006). *Respect Action Plan.* Respect Task Force, Home Office, London.

Hope, T., Bryan, J., Crawley, E., Russell, n., and Trickett, A. (2004). Strategic Development Projects in the Yorkshire and the Humber, East Midlands and Eastern Regions. Home Office Online Report 41/04, Home Office RDS, London.

Hope, T. and Murphy, D. (1983). Problems of implementing crime prevention: The experience of a demonstration project. *Howard Journal*, 22(1):38–50.

Hough, M. and Tilley, N. (1998). Auditing Crime and Disorder. Crime Detection and Prevention Series Paper 91, Home Office Police Research Group, London.

Howell, M. and Pugliesi, K. (1988). Husbands who harm: Predicting sousal violence by men. *Journal of Family Violence*, 3(1):15–27.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2:193–218.

Jeffery, C. (1971). *Crime Prevention Through Environmental Design.* Sage Publications, Beverly Hills.

Jephcott, P. (1971). *Homes in High Flats: Some of the Human Problems Involved in Multi-storey Housing.* Oliver Boyd, Edinburgh.

Johnson, S. and Bowers, K. (2004). The stability of space-time clusters of burglary. *British Journal of Criminology*, 44(1):55–65.

Johnston, R. (2000). Ecological fallacy. In Johnston, R., Gregory, D., Pratt, G., and Watts, M., (Eds.), *The Dictionary of Human Geography*, pages 190–191. Blackwell, Oxford, fourth edition edition.

Jolley, D., Jarman, B., and Elliott, P. (1992). Socio-economic confounding. In Elliott, P., Cuzick, J., English, D., and Stern, R., (Eds.), *Geographical and Environmental Epidemiology: Methods for Small Area Studies*, pages 115–124. Oxford University Press, Oxford.

Juby, H. and Farrington, D. (2002). Disentangling the link between disrupted families and delinquency. *British Journal of Criminology*, 41:22–40.

Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Kawachi, I., Kennedy, B., and Wilkinson, R. (1999). Crime: social disorganization and relative deprivation. *Social Science and Medicine*, 48:719–731.

Kelsall, J. and Diggle, P. (1995). Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine*, 14:2335–2342.

Kenwood, M. (2005). Partnership GIS delivering public reassurance and liveability through Civic Renewal. Presentation given to 3rd National Crime Mapping Conference, London, April.

Knox, J., Pemberton, A., and Wiles, P. (2000). *Partnerships in Community Safety: An Evaluation of Phase 2 of the Safer Cities Programme*. Department for Environment, Transport and the Regions, London.

Kruskal, W. and Wallis, W. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.

Kuldorff, M. (2001). Prospective time-periodic geographical disease surveilance using a scan statistic. *Journal of the Royal Statistical Society*, 164:61–72.

Kuldorff, M., Heffernan, R., Hartman, J., Assunção, R., and Mostashari, F. (2005). A space-time permuation scan statistic for disease outbreak detection. *PLoS Medicine*, 2(3):216–224. PLoS Medicine is a Public Library of Science Journal.

Kurtz, E., Koons, B., and Taylor, R. (1998). Land use, physical deterioration, resident-based control, and calls for service on urban streetblocks. *Justice Quarterly*, 15(1):121–149.

LaGrange, T. (1999). The impact of neighbourhoods, schools, and malls on the spatial distribution of property damage. *Journal of Research in Crime and Delinquency*, 36(4):393–422.

Lander, B. (1954). *Towards an Understanding of Juvenile Delinquency*. Columbia University Press, New York.

Lathrop, R. and Williams, J. (1987). The reliability of inverse scree tests for cluster analysis. *Educational and Psychological Measurement*, 47:953–959.

Laycock, G. (1992). Operation identification, or the power of publicity. In Clark, R., (Ed.), *Situational Crime Prevention: Successful Case Studies*, pages 230–238. Harrow and Heston, New York.

Laycock, G. (2003). Beyond Blobology. Paper presented at the First National Crime Mapping Conference, University of London, March.

LCSP (1999). *Leeds Community Safety Partnership Strategy 1999-2002*. Leeds Community Safety Partnership, Leeds.

LCSP (2002). *Leeds Community Safety Strategy 2002-2005*. Leeds Community Safety Partnership, Leeds.

Lea, J. (1992). The analysis of crime. In Young, J. and Matthews, R., (Eds.), *Rethinking Criminology: The Realist Debate*. Sage, London.

Lee, B. and Campbell, K. (1997). Common ground? urban neighbourhoods as survey respondents see them. *Social Science Quarterly*, 78(4):922–936.

Lee, T. (1968). Urban neighbourhood as a socio-spatial schema. *Human Relations*, 21:241–267.

Leeds City Council (2004). Council Plan 2004/05. , Leeds City Council, Leeds.

Leeds Community Safety (2004). Leeds Crime, Disorder and Drugs Audit 2004.

Leeds Initiative (2004). Vision for leeds 2004-2020. , The Leeds Initiative, Leeds.

Lemert, E. (1951). *Social Pathology*. McGraw-Hill, New York.

Levine, N. (2002). CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations. , The National Institute of Justice, Washington, D.C.

Lewis, D. and Salem, G. (1986). *Fear of Crime: Incivility and the Production of Social Problem*. Transaction Books, New York.

Lipsey, M. and Derzon, J. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In Loeber, R. and Farrington, D., (Eds.), *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*, pages 86–105. Sage, Thousand Oaks, CA.

Lister, S., Wall, D., and Bryan, J. (2004). Evaluation of the Leeds Distraction Burglary Initiative. , Home Office, London. Online Report 44/04. Available at http://www.homeoffice.gov.uk/rds/.

Lochner, K., Kawachi, I., and Kennedy, B. (1999). Social capital: a guide to its measurement. *Health and Place*, 5:259–270.

Longley, P. (2005). Geographic information systems: a renaissance of geodemographics for public service delivery. *Progress in Human Geography*, 29(1):57–63.

Longley, P., Goodchild, M., Maguire, D., and Rhind, D. (2005). *Geographic Information Systems and Science*. John Wiley Sons, Chichester, second edition edition.

Longley, P. and Harris, R. (1999). Towards a new digital data infrastructure for urban analysis and modelling. *Environment and Planning D: Planning and Design*, 26:855–878.

Lorr, M. (1983). *Cluster Analysis for Social Scientists*. Jossey-Bass, San Francisco.

Loveday, B. (1999). The impact of performance culture on criminal justice agencies in England and Wales. *International Journal of the Sociology of Law*, 27:351–377.

Lowman, J. (1986). Conceptual issues in the geography of crime: Toward a geography of social control. *Annals of the Association of American Geographers*, 76(1):81–94.

MacDonald, J. (2002). The effectiveness of community policing in reducing urban violence. *Crime and Delinquency*, 48(4):592–618.

MacEachren, A. (1995). *How Maps Work: Representation, Visualisation, and Design*. Guilford Press, New York.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Le Cam, L. and Neyman, J., (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, Berkeley, CA. University of California Press.

Maguire, M. (1982). *Burglary in a Dwelling*. Heinemann, London.

Makarenkov, V. and Legendre, P. (2001). Optimal variable weighting for ultrametric and addititive trees and k-means partioning: methods and software. *Journal of Classification*, 18:245–271.

Mann, H. and Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematics and Statistics*, 18:50–60.

Martin, D. (1998). Automatic neighbourhood identification from population surfaces. *Computers, Environment and Urban Systems*, 22(2):107–120.

Martin, D. (2002). Census population surfaces. In Rees, P., Martin, D., and Williamson, P., (Eds.), *The Census Data System*, pages 139–148. John Wiley and Sons, Chichester.

Martin, D., Tate, N., and Langford, M. (2000). Refining population surface models: experiments with Northern Ireland census data. *Transactions in GIS*, 4:343–360.

Massey, D. and Denton, N. (1988). The dimensions of residential segregation. *Social Forces*, 67:281–315.

Matthews, M. (1981). Children's perception of urban distance. *Area*, 13(4):333–343.

Matthews, M. (1984a). Cognitive mapping abilities of young boys and girls. *Geography*, 69:327–336.

Matthews, M. (1984b). Environmental cognition of young children: images of journey to school and home area. *Transactions if the Institute of British Geographers (New Series)*, 9:89–105.

Matthews, R. and Pitts, J. (2001). Introduction: beyond criminology. In Matthews, R. and Pitts, J., (Eds.), *Crime, Disorder and Community Safety*. Routledge, London.

Mawby, R. (1977). Defensible space: A theoretical and empirical appraisal. *Urban Studies*, 14:169–180.

Mayhew, P., Clarke, R., Sturman, A., and Hough, J. (1976). Crime as opportunity. Research study number 34, Home Office, London.

Mayhew, P. and Hough, M. (1988). The British Crime Survey: origins and impact. In Maguire, M. and Pointing, J., (Eds.), *Victims of Crime: A New Deal?*, pages 156–163. Open University Press, Milton Keynes.

McBratney, A. and Moore, A. (1985). Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 35:165–185.

Mencken, F. and Barnett, C. (1999). Murder, non-negligent manslaughter, and spatial autocorrelation in mid-south counties. *Journal of Quantitative Criminology*, 15(4):407–422.

Merry, S. (1981). Defensible space undefended. *Urban Affairs Quarterly*, 16(4):397–422.

Merton, R. (1938). Social structure and anomie. *American Sociological Review*, 3:672–682.

Messner, S., Anselin, L., Baller, R., Hawkins, D., Deane, G., and Tolnay, S. (1999). The spatial patterning of county homicide rates: An application of exploratory spatial data analysis. *Journal of Quantitative Criminology*, 15(4):423–450.

Millie, A. and Hough, M. (2004). Assessing the Impact of the Reducing Burglary Initiative in Southern England and Wales. Home Office Online Report 42/04, Home Office RDS, London.

Milligan, G. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45(3):325–342.

Milligan, G. (1989). A validation study of a variable weighting algorithm for cluster analysis. *Journal of Classification*, 6:53–71.

Milligan, G. (1996). Clustering validation: Results and implications for applied analyses. In Arabie, P., Hubert, L., and De Soete, G., (Eds.), *Clustering and Classification*, pages 341–375. World Scientific, Singapore.

Milligan, G. and Cooper, M. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179.

Milligan, G. and Cooper, M. (1987). Methodology review: clustering methods. *Applied Psychological Measurement*, 11:329–354.

Milligan, G. and Cooper, M. (1988). A study of standardisation of variables in cluster analysis. *Journal of Classification*, 5:181–204.

Mills, A. and Pearson, S. (2000). From audit to strategy: A practice review. In Ballintyne, S., Pease, K., and McLaren, V., (Eds.), *Secure Foundations: Key Issues in Crime Prevention, Crime Reduction and Community Safety*. IPPR, Southampton.

Mirrlees-Black, C., Mayhew, P., and Percy, A. (1996). *The 1996 British Crime Survey: England and Wales*. Home Office Research and Statistics Directorate, London.

Moore, G. (1979). Knowing about environmental knowing. *Environment and Behaviour*, 11:33–70.

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37:17–23.

Morgan, F. (2000). Repeat burglary in a Perth suburb: Indicator of short-term or long-term risk. In Farrell, G. and Pease, K., (Eds.), *Repeat Victimisation: Crime Prevention Studies, volume 12*. Criminal Justice Press, Monsey, NY.

Morphet, C. (1993). The mapping of small-area census data: A consideration of the role of enumeration district boundaries. *Environment and Planning A*, 25(9):1267–1277.

Morris, T. (1957). *The Criminal Area: A Study in Social Ecology*. Routledge and K.Paul, London.

Mosher, C. (2001). Predicting drug arrest rates: Conflict and social disorganization perspectives. *Crime and Delinquency*, 47(1):84–104.

Muehrcke, P. and Muehrcke, J. (1997). *Map Use: Reading, Analysis, Interpretation.* JP Publications, Madison, WI, Fourth edition.

Murie, A. (1997). Linking housing changes to crime. *Social Policy and Administration*, 31(5).

Murray, A. T., McGuffog, I., Western, J. S., and Mullins, P. (2001). Exploratory spatial data analysis techniques for examining urban crime. *British Journal of Criminology*, 41(2):309–329.

Nayak, A. (2003). 'Through children's eyes': childhood, place and the fear of crime. *Geoforum*, 34(3):303–315.

NCIS (2000). *The National Intelligence Model.* National Criminal Intelligence Service, London.

Neighbourhoods and Communities Partnership (2003). Briefing notes for the meeting of the [Leeds] Neighbourhoods and Communities Partnership, Thursday 30 January, 2003. Leeds University Business School, Leeds. Available online at http://www.leedsinitiative.org/initiativeDocuments/2005214_78005618.pdf. Accessed 23.08.2005.

Newman, O. (1972). *Defensible Space: People and Design in the Violent City.* Macmillan, New York.

Newton, A. (2006). Routing out the hot spots: Applying GIS and crime place principles to examine crime and disorder on public transport. In Wise, S. and Craglia, M., (Eds.), *Innovations in GIS 10*, London. Taylor and Francis.

Ng, R. and Han, J. (2002). Clarans: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering*, 14(5):1003–1016.

Nicholls, K. and Tudorancea, C. (2001). Application of fuzzy cluster analysis to Lake Simcoe crustacean zooplankton community structure. *Canadia Journal of Fisheries and Aquatic Sciences*, 58:231–240.

Norman, P., Rees, P., and Boyle, P. (2003). Achieving data compatibility over space and time: Creating consitent geographical zones. *International Journal of Population Geography*, 9(5):365–386.

ODPM (2000). *Regeneration Research Summary - Partnerships in Community Safety: An Evaluation of Phase 2 of the Safer Cities Programme.* Office of the Deputy Prime Minister, London.

ODPM (2004). *The English Indices of Deprivation 2004.* Office of the Deputy Prime Minister, London.

ODPM (2005). Generalised land use database statistics for england. , Office of the Deputy Prime Minister, London.

Ofsted (2003). Inspection Report: Micklefield Church of England Voluntary Controlled Primary School. Available online at http://www.ofsted.gov.uk/reports/107/107992.pdf.

O'Malley, P. (1992). Risk, power and crime prevention. *Economy and Society*, 21(3):252–275.

ONS (1999). *The ONS classification of local and health authorities of Great Britain: revised for authorities in 1999.* Office for National Statistics, London.

ONS (2003). Methods for national statistics 2001 area classification for local authorities. Accessed 28.1.2004.

ONS (2005). 2001 Area Classification for Output Areas. http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/default.asp.

Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers*, 2:459–472.

Openshaw, S. (1978). An optimal zoning approach to the study of spatially aggregated data. In Masser, I. and Brown, P., (Eds.), *Spatial Representation and Spatial Interaction*, pages 95–113. Martinus Nijhoff, Leiden.

Openshaw, S. (1983). Multivariate analysis of census data: the classification of areas. In Rhind, D., (Ed.), *A Census User's Handbook*, pages 243–263. Methuen, London.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Concepts and Techniques in Modern Geography. Geo Books, Norwirch.

Openshaw, S. and Wymer, C. (1995). Classifying and regionalizing census data. In Openshaw, S., (Ed.), *Census Users Handbook*. GeoInformation International, Cambridge.

Pain, R. (1991). Space, sexual violence and social control: integrating geographical and feminist analyses of womens fear of crime. *Progress in Human Geography*, 15(4):415–431.

Pain, R. (2000). Place, social relations and the fear of crime: a review. *Progress in Human Geography*, 24(3):365–387.

Park, R. (1952). *Human Communities*. The Free Press, Glencoe, Ill.

Parkes, A., Kearns, A., and Atkinson, R. (2002). What makes people dissatisfied with their neighbourhoods? *Urban Studies*, 39(13):2413–2438.

Paulsen, D. (2004). Geographic Profiling: Hype or Hope? Preliminary results into the accuracy of geographic profiling software. Paper presented at the 2nd UK Crime Mapping Conference, London, March.

Pawson, E. and Banks, G. (1993). Rape and fear in a New Zealand city. *Area*, 25(1):55–63.

Pawson, R. and Tilley, N. (1996). What's crucial in evaluation research: A reply to Bennett. *British Journal of Criminology*, 36(4):574–578.

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. Sage Publications, London.

Pease, K. (1998). Repeat victimisation: Taking stock. Crime Detection and Prevention Series Paper 90, Home Office, London.

Petersen, A. and Mortimer, J. (1994). *Youth Unemployment and Society*. Cambridge University Press, Cambridge.

Pettigrew, A. (1972). Information control as a power resource. *Sociology - the Journal of the British Sociological Association*, 6(2):187–204.

Peuquet, D. (2002). *Representations of Space and Time*. Guilford Press, New York.

Pickles, J., (Ed.) (1995). *Ground Truth*. The Guilford Press, New York.

Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford University Press, Oxford.

Poyner, B. and Webb, B. (1991). *Crime Free Housing*. Butterworth-Architecture, Oxford.

Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1986). *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Putnam, R. (1993). *Making Democracy Work*. Princeton University Press, Princeton, NJ.

Pyle, G., Hanten, E., Williams, P., Pearson, A., Doyle, J., and Kwofie, K. (1974). *The Spatial Dynamics of Crime*. University of Chicago, Department of Geography, Chicago.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Raistrick, D., Hodgson, R., and Ritson, B., (Eds.) (1999). *Tackling Alcohol Together: The Evidence Base for a UK Alcohol Policy*. Free Association Books, London.

Rand, W. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66:846–850.

Rapkin, B. and Luke, D. (1993). Cluster analysis in community research: Epistemology and practice. *American Journal of Community Psychology*, 21(2):247–277.

Ratcliffe, J. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Systems*, 18(1):61–72.

Ratcliffe, J. and McCullagh, M. (1999). Hotbeds of crime and the search for spatial accuracy. *Geographical Systems*, 1(4):385–398.

Read, T. and Oldfield, D. (1995). Local crime analysis. Crime Detection and Prevention Series Paper 65, Home Office Police Research Group, London.

Rees, P. (1972). Problems of classifying sub-areas within cities. In Berry, B. and Smith, K., (Eds.), *City Classification Handbook: Methods and Applications*, pages 265–330. Wiley, London.

Rees, P. and Butt, F. (2003). Ethnic change and diverity in England, 1981-2001. Paper presented at the 'RGS-IBG Annual Conference 2003', London, September 2003.

Rees, P., Denham, C., Charlton, J., Openshaw, S., Blake, M., and See, L. (2002a). ONS classifications and GB Profiles: Census typologies for researchers. In Rees, P., Martin, D., and Williamson, P., (Eds.), *The Census Data System*, pages 149–170. John Wiley Sons, Chichester.

Rees, P., Martin, D., and Williamson, P. (2002b). Census Data Resources in the United Kingdom. In Rees, P., Martin, D., and Williamson, P., (Eds.), *The Census Data System*. Wiley, Chichester.

Rengert, G. and Wasilchick, J. (1985). *Suburban Burglary: A Time and a Place for Everything.* Charles Thomas, Springfied, ILL.

Rhodes, J., Tyler, P., and Brennan, A. (2005). Assessing the effect of area based initiatives on local area outcomes: Some thoughts based on the national evaluation of the Single Regeneration Budget in England. *Urban Studies*, 42(11):1919–1946.

Robey, D. (1981). Computer information systems and organization structure. *Communications of the ACM*, 26(10):679–687.

Robinson, G. (1998). *Methods and Techniques in Human Geography.* Wiley, Chichester.

Robson, B. (1969). *Urban Analysis: A Study of City Structure.* Cambridge University Press, Cambridge.

Rosenbaum, D. (1988). Community crime prevention: A review and synthesis of the literature. *Justice Quarterly*, 5(3):323–393.

Rosenbaum, M. (1981). *Women on Heroin.* Rutgers University Press, New Brunswick, NJ.

Ruspini, E. (1969). A new approach to clustering. *Information and Control*, 15:22–32.

Rutter, M., Giller, H., and Hagell, A. (1998). *Antisocial Behaviour by Young People.* Cambridge University Press, Cambridge.

Sadahiro, Y. (1997). Cluster perception in the distribution of point objects. *Cartographica*, 34:49–61.

Safer Leeds (2005). *Strategy 2005-2008.* Safer Leeds, Leeds.

Sampson, R. and Laub, J. (1993). *Crime in the making: Pathways and turning points through life.* Harvard University Press, Cambridge, MA.

Sampson, R. and Raudenbush, S. (1999). Systematic social observation of public spaces: A look as disorder in urban neighborhoods. *American Journal of Sociology*, 105(3):603–651.

Sampson, R., Raudenbush, S., and Earls, F. (1997). Neighbourhoods and violent crime: a multilevel study of collective efficacy. *Science*, 277:918–924.

Santiago, A., Galster, G., and Pettit, K. (2003). Neighbourhood crime and scattered-site public housing. *Urban Studies*, 40(11):2147–2163.

Schumacher, B. and Leitner, M. (2002). Spatial crime displacement resulting from large-scale urban renewal programs in the city of Baltimore, MD: a GIS modelling approach. Paper presented at the IVth International Conference on GeoComputation, Fredericksburg, VA, USA, July.

Scott, R. (2005). Targeting resources using geodemographic profiling. Presentation given at 3rd National Crime Mapping Conference, London, April.

See, L. and Openshaw, S. (2001). Fuzzy geodemographic targeting. In Clarke, G. and Madden, M., (Eds.), *Regional Science in Business*, Advances in Spatial Science, pages 269–281. Springer, Berlin.

Shaftoe, H. (2004). Do symbolic barriers prevent crime and offer an increased sense of safety? A pilot study. *Planning Practice and Research*, 19(4):447–452.

Shaw, C. (1930). *The Jackroller*. University of Chicago Press, Chicago.

Shaw, C. (1931). *The Natural History of a Delinquent Career*. University of Chicago Press, Chicago.

Shaw, C. (1938). *Brothers in Crime*. University of Chicago Press, Chicago.

Shaw, C. and McKay, H. (1942). *Juvenile Delinquency and Urban Areas*. Chicago University Press, Chicago.

Shaw, C. and McKay, H. (1969). *Juvenile Delinquency and Urban Areas*. University of Chicago Press, Chicago.

Shaw, G. and Wheeler, D. (1994). *Statistical Techniques in Geographical Analysis*. David Fulton Publishers, London.

Sheldon, G., Hall, R., Brunsdon, C., Charlton, M., Alvanides, S., and Mostratos, N. (2002). Maintaining Basic Command Unit and Crime and Disorder Partnership Families for Comparative Purposes. Online report 06/02, Home Office, London. Available online at http://www.homeoffice.gov.uk/rds/pdfs2/basiccommand.pdf. Aceesed 14.1.2006.

Shepherd, P. (2003). Special Factors: Burglary in Leeds. Report for the Leeds Best Value Review of Community Safety, School of Geography, University of Leeds.

Shepherd, P., See, L., Kongmuang, C., and Clarke, G. (2004). An Analysis of Crime and Disorder in Leeds: 2000/01 to 2003/04. Report to the Leeds Community Safety Partnership. Unpublished.

Shepherd, P., Stillwell, J., and Clarke, G. (2005). Identifying crime hotspots using a neighbourhood classification. Poster presented at the 3rd National Crime Mapping Conference, London, April.

Shevky, E. and Bell, W. (1955). *Social Area Analysis: theory, Illustrative Application and Computational Procedures*. Stanford University Press, Stanford.

Short Jr, J. (1969). Introduction to the revised edition. In Shaw, C. and McKay, H., (Eds.), *Juvenile Delinquency and Urban Areas*. University of Chicago Press, Chicago.

Shury, J., Speed, M., Vivian, D., Kuechel, A., and Nicholas, S. (2005). Crime against retail and manufacturingl premises: findings from the 2002 Commercial Victimisation Survey. Online report 37/05, Home Office, London. Available from http://www.homeoffice.gov.uk/rds/pdfs05/rdsolr3705.pdf. Accessed 27.2.2006.

Sibley, D. (1990). *Spatial Applications of Exploratory Data Analysis*. CATMOG 49. Geo Books, Norwich.

Sibley, D. (1998). Sensations and spatial science: gratification and anxiety in the production of ordered landscapes. *Environment and Planning A*, 30:235–246.

Signoretta, P. and Craglia, M. (2002). Joined-up government in practice: A case study of Children's needs in Sheffield. *Local Government Studies*, 28(1):59–76.

Simmons, J. and Dodd, T. (2003). Crime in England and Wales 2002/2003. Home Office Statistical Bulletin 07/03, Home Office, London.

Simmons, J., Legg, C., and Hosking, R. (2003). National Crime Recording Standard (NCRS): an analysis of the impact on recorded crime. Part Two: Impact on individual police forces. , Home Office RDS, London. On-line Report 31/03. Available on-line at http://www.homeoffice.gov.uk/rds/pdfs2/rdsolr3203.pdf.

Simpson, L. (2002). Geography conversion tables: A framework for conversion of data between geographical units. *International Journal of Population Geography*, 8(1):69–82.

Skogan, W. (1990). *Disorder and Decline: Crime and the Spiral of Decay in American Cities.* Free Press, New York.

Sleight, P. (1997). *Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business.* World Advertising Research Centre, Henley-on-Thanes.

Smith, S. (1986). Utopia on trial: Visions and reality in planned housing. *Urban Studies*, 23:244–246.

Smith, S. (1987). Fear of crime: Beyond a geography of deviance. *Progress in Human Geography*, 11(1):1–23.

Sneath, P. and Sokal, R. (1973). *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* WH Freeman, San Francisco.

Snodgrass, J. (1976). Clifford R. Shaw and Henry D. McKay: Chicago criminologists. *British Journal of Criminology*, 16(1):1–19.

Stark, R. (1987). Deviant places: a theory of the ecology of crime. *Criminology*, 25(4):893–909.

Stillwell, J. and Duke-Williams, O. (2003). A new web-based interface to british census of population origin-destination statistics. *Environment and Planning A*, 35(1):113–132.

Stillwell, J., Duke-Williams, O., Feng, Z., and Boyle, P. (2005). Delivering census interaction data to the user: Data provision and software development. Working Paper 05/01, School of Geography, University of Leeds, Leeds.

Taylor, I., Walton, P., and Young, J. (1975). *Critical Criminology.* Routledge and Kegan Paul, Lodon.

Taylor, R. (1993). Community structural change and fear of crime. *Social Problems*, 40(3):374–395.

Taylor, R. (2002). Physical environment, crime, fear, and resident-based control. In Wilson, J. and Petersilia, J., (Eds.), *Crime: Public Policies for Crime Control*, pages 413–434. Institute for Contemporary Studies, Oakland, CA.

Taylor, R. B. (1997). Social order and disorder of street blocks and neighborhoods: Ecology, microecology, and the systemic model of social disorganization. *Journal of Research in Crime and Delinquency*, 34(1):113–155.

The Beat (2003). Action plan drafted for crime recording. *The Beat: The Newspaper of West Yorkshire Police*, 137:1.

Thornton, A., Hatton, C., Malone, C., Fryer, T., Walker, D., Cunningham, J., and Durrani, N. (2003). Distraction burglary among older adults and ethnic minority communities. Home Office Research Study 269, Home Office Research, Development and Statistics Directorate, London.

Tilley, N. (1992). Safer cities and community safety strategies. Crime Prevention Unit Series: Paper 38, Police Research Group, Home Office.

Timms, D. (1971). *The Urban Mosaic: Towards a Theory of Residential Differentiation.* Cambridge University Press, London.

Tobler, W. (1970). A computer movie. *Economic Geography*, 46:234–240.

Townsend, P., Phillimor, P., and Beatty, A. (1988). *Health and Deprivation: Inequality and the North.* Croom Helm, London.

Townsley, M., Homel, R., and Chaseling, J. (2003). Infectious burglaries: a test of the near repeat hypothesis. *British Journal of Criminology*, 43:615–633.

Trauwaert, E. (1987). L1 in fuzzy clustering. In Dodge, Y., (Ed.), *Statistical data analysis based on the L1 norm*, pages 417–426. Elsevier, Amsterdam.

Tseloni, A., Osborn, D., Trickett, A., and Pease, K. (2002). Modelling property crime using the British Crime Survey. *British Journal of Criminology*, 42:109–128.

Tufte, E. (1990). *Envisioning Information.* Graphics Press, Cheshire, Conneticut.

Tukey, J. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading, Mass.

Unwin, D. and Hearnshaw, H. (1994). *Visualization in Geographical Information Systems.* John Wiley and Sons, Chichester.

Valentine, G. (1992). Images of danger - womens sources of information about the spatial distribution of male violence. *Area*, 24(1):22–29.

Valentine, G. (1996). Angels and devils: moral landscapes of childhood. *Environment and Planning D: Society and Space*, 14:581–599.

Vickers, D., Rees, P., and Birkin, M. (2003). A new classification of UK local authorities using 2001 census key statistics. Working paper 03/03, School of Geography, University of Leeds, Leeds.

Vickers, D., Rees, P., and Birkin, M. (2005). Creating the National Classification of Census Output Areas: Data, Methods and Results. Working Paper 05/2, School of Geography, University of Leeds, Leeds.

Vincent, K. (2005). Mapping socio-demographic risk factors for domestic violence in Devon. Poster presented at the 3rd National Crime Mapping Conferences, London, April.

Voas, D. and Williamson, P. (2001). The diversity of diversity: a critique of geodemographic classification. *Area*, 33(1):63–76.

Vold, G., Bernard, T., and Snipes, J. (1998). *Theoretical Criminology.* Oxford University Press, New York.

Walace, M. and Denham, C. (1996). The ONS Classification of Local and Health Authorities of Great Britain. *Studies on Medical and Population Subjects*, 59.

Walklate, S. (1997). Risk and criminal victimization: A modernist dilemma? *British Journal of Criminology*, 37(1):35–45.

Walklate, S. (1998). *Understanding Criminology: Current Theoretical Debates.* Open University Press, Buckingham.

Wells, C. (1998). 'I blame the parents': Fitting new genes in old criminal laws. *The Modern Law Review*, 61(5):724–739.

Welsh, W., Stokes, R., and Greene, J. (2000). A macro-level model of school disorder. *Journal of Research in Crime and Delinquency*, 37(3):243–283.

Wilcox, P., Quisenberry, N., Cabera, D., and Jones, S. (2004). Busy places and broken windows? toward defining the role of physical structure and process in community crime models. *Sociological Quarterly*, 45(2):185–207.

Wiles, P. and Costello, A. (2000). The 'Road to Nowhere': the Evidence for Travelling Criminals. Home Office Research Study 207, Home Office RDS, London.

Williams, N. (1985). Crime patterns in Aberdeen. *Scottish Geographical Magazine*, 101(1):49–60.

Williamson, D., McLafferty, S., Goldsmith, V., McGuire, P., and Mollenkopf, J. (1998). Smoothing crime incident data: New methods for determining the bandwidth in kernel estimation. Available online at http://gis.esri.com/library/userconf/proc98/PROCEED/TO850/PAP829/P829.HTM. Accessed 11-6-2003.

Williamson, T. (2003). Crime data: Uses, misues and the need to understand neighbourhood context. Paper presented at 'Understanding Crime and the Neighbourhood', ESRC Centre for Neighbourhood Research, Bristol, June.

Williamson, T., Ashby, D., and Webber, R. (2005). Young offenders, schools and the neighbourhood: A new approach to data-analysis for community policing. *Journal of Community and Applied Social Psychology*, 15:203–228.

Wilson, J. (1975). *Thinking About Crime.* Vintage, New York.

Wilson, J. and Kelling, G. (1982). Broken windows: the police and neighbourhood safety. *The Atlantic Monthly*, March:29–37.

Wilson, R. (1963). Difficult housing estates. Tavistock Pamphlet No. 5, Tavistock Publications, London.

Wishart, D. (1987). *Clustan User Manual.* Computing Laboratory, University of St Andrews, 4th edition.

Wong, C. (1997). Crime risk in urban neighbourhoods: the use of insurance data to analyse changing spatial forms. *Area*, 29(3):228–240.

Yarwood, R. ad Gardner, G. (2000). Fear of crime, cultural threat and the countryside. *Area*, 32(4):403–411.

Young, J. (1986). The failure of criminology: the need for radical realism. In Matthews, R. and Young, J., (Eds.), *Confronting Crime*. Sage, London.

Young, J. (1992). Ten points of realism. In Young, J. and Matthews, R., (Eds.), *Rethinking Criminology: The Realist Debate*. Sage, London.

Zeng, D., Chang, W., and Chen, H. (2004). A comparative study of spatio-temporal hotspot analysis techniques in security informatics. In *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*, pages 106–111, Washington, D.C.

Zhu, L., Gorman, D., and Horel, S. (2004). Alcohol outlet density and violence: A geospatial analysis. *Alcohol and Alcoholism*, 39(4):369–375.