# GEODEMOGRAPHIC CLASSIFICATION SYSTEMS FOR THE DEVELOPING WORLD: THE CASE OF NIGERIA AND THE PHILIPPINES


Adegbola Adekunle Ojo

The University of Sheffield

Department of Geography


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

May 2010

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.

## Acknowledgements

This work has been a challenging and interesting experience for me and I sincerely wish to thank the good Lord for his guidance and protection throughout the period of my study at the University of Sheffield.

I am greatly indebted to my supervisors Dr Dimitris Ballas and Dr Dan Vickers for their advice, support and guidance. Their positive and unquantifiable contributions towards the success of this work are immeasurable. I sincerely appreciate the constant encouragements and strong belief you have shown in me. For me you have acted more like brothers than senior colleagues. I can ***never*** forget your sincere contributions. I am also grateful to Dr Jan Rigby for her support particularly at the early stages of the project.

I want to thank the University of Sheffield for providing part of the funding for this research work. I appreciate your kindness without which this thesis may not have been written. I also want to thank the National Bureau of Statistics, Nigeria and the National Statistics Office, Philippines for supplying data used for this project.

I am deeply grateful and heartily express my profound appreciation to the one I love, Funmilola Ojo. Funmi, what can I say? Your patience, understanding, love and tremendous unflinching support throughout my study programme. You are the best thing that has ever happened to me. Thank you so much for always being there. I am also indebted to our daughter Damilola Ojo who was born mid way through my research programme. Even as a child, you understood Daddy's tight schedule and acted maturely. You even went to one or two seminars with Daddy just to support him. You are my jewel.

To my wonderful parents Mr and Mrs G. O. Ojo, I owe you so much. Words can not express my feelings and sincere appreciation of your love and tireless sacrifices. I love you so much and I am deeply grateful. I also sincerely appreciate the strong encouragements from my brothers, Ayoola and Adedayo and my sister Tinuola.

I can not overemphasize the positive contributions of Chief and Chief (Mrs) A. A. Ashiru (my wife's parents). Thank you for your belief, patience, love, kindness and moral support. I am also deeply grateful to my wife's brothers Mr Kayode Ashiru, Mr Wole Ashiru and Mr Gbenga Ashiru and sister Ms. Kemi Ashiru for their encouragements throughout my study programme.

In addition, I can not but thank Deacon and Mrs B. Ikini and family for the genuine love you have shown towards me and my family throughout my programme. I am also grateful to my good friends Seun Olatoye-Ojo, Jide Biobaku, Bukunmi Omidiran, Tunji Badmus, Akin Adeniran, Bunmi Sobodu, Wole Osunneye, Seun Afuwape, Kayode Adelowokan, Akin Akingbogun, Ade Osinowo and a host of others too numerous to mention who have been very supportive with words of advice and encouragements from far and near. I am thankful to you all.

I must also specially thank Dr Pablo Mateos of the Department of Geography, University College London and his wife Brenda for their genuine continuous love and kindness towards me and my family. To my colleagues at the Department of Geography, University of Sheffield, you each deserve a hug for your friendship and all the extra-curricular activities we were all able to organise together.

Finally I would like to thank the friends I made at the Sheffield City Council – Rebecca Dixon and co. and at the Yorkshire and Humber Public Health Observatory. You all contributed in one way or the other to the success of this work. I am sincerely grateful to you all.

Adegbola A. Ojo
March, 2010

## Abstract

Since the emergence of modern day geodemographics, population geography has witnessed a renaissance in the area of policy related spatial analysis. These classifications group areas on the basis of similarity into cluster units which define their demographic and social characteristics. The methods used to create these systems combine geographic thought and theory with statistical manipulations of multivariate data.

There is currently paucity in the development and use of geodemographic systems in developing countries due to multiple related factors particularly data availability and access.

This thesis argues that the developing world has a lot to benefit from geodemographic systems especially in areas of public policy making. The aim of the thesis is to provide a comprehensive description of the decisions made during the creation of two area segmentation systems – one for Nigeria and the other for the Philippines. The systems are accompanied by detailed user guides with visuals and pen portraits

In addition to developing the two systems, examples of how the systems can be used to inform policy are presented in the areas of education, gender parity and maternal and child health care.

It is hoped that the creation of these systems will mark the beginning of a long awaited proliferation of geodemographic systems to developing countries.

# Table of Contents

## List of Figures

## List of Tables

## Abbreviations

| | |
|---|---|
| ACORN | A Classification of Residential Neighbourhoods |
| ADB | Asian Development Bank |
| ANC | Antenatal Care |
| ANOVA | Analysis of Variance |
| BIC | Bayesian Information Criterion |
| CBN | Central Bank of Nigeria |
| CES | Centre for Environmental Studies |
| CLI | Call-line Identification |
| CWIQ | Core Welfare Indicators Questionnaire |
| DA | Dissemination Areas |
| DHS | Demographic and Health Survey |
| DOTS | Directly Observed Treatment Short-course |
| EA | Enumeration Areas |
| ED | Enumeration District |
| FAO | Food and Agriculture Organisation |
| FCT | Federal Capital Territory |
| FRN | Federal Republic of Nigeria |
| GDP | Gross Domestic Product |
| GIS | Geographic Information Systems |
| GMCFS | Greater Manchester County Fire Service |
| HDI | Human Development Index |
| HDN | Human Development Network |
| HIV | Human Immunodeficiency Virus |
| AIDS | Acquired Immunodeficiency Syndrome |
| ICT | Information Communications Technology |
| IDT | International Development Targets |
| IDT | Index of Dissimilarity |
| ILO | International Labour Organization |
| IMF | International Monetary Fund |
| IPUMS | Integrated Public Use Microdata Series |
| IT | Information Technology |
| JSSC | Junior Secondary School Certificate |
| LGA | Local Government Area |
| MAUP | Modifiable Areal Unit Problem |
| MDG | Millennium Development Goal |
| MICS | Multiple Indicators Cluster Survey |
| MMR | Maternal Mortality Ratio |
| MTPDP | Medium Term Philippine Development Plan |
| NBS | National Bureau of Statistics |
| NEDA | National Economic Development Authority |
| NEEDS | National Economic Empowerment and Development Strategy |
| NPC | National Populations Commission |
| NSO | National Statistics Office |
| OA | Output Area |
| ODA | Official Development Assistance |
| OECD | Organisation for Economic Co-operation and Development |
| ONS | Office for National Statistics |
| PCA | Principal Components Analysis |

| | |
|---|---|
| PEARL | Urban Policy Evaluation and Analysis Research Laboratory |
| PPP | Purchasing Power Parity |
| RBM | Roll Back Malaria |
| RMS | Root Mean Square |
| SME | Small Scale Enterprise |
| TBA | Traditional Birth Attendants |
| TEEP | Third Elementary Education Project |
| UBE | Universal Basic Education |
| UDHR | Universal Declaration of Human Rights |
| UK | United Kingdom |
| UN | United Nations |
| UNDP | United Nations Development Programme |
| UNESCO | United Nations Educational Scientific and Cultural Organisation |
| UNFPA | United Nations Population Fund |
| UNICEF | United Nations Children Fund |
| URPERRL | Research and Policy Evaluation Regional Laboratory |
| USA | United States of America |
| USAID | United States Agency for International Development |
| WCED | World Commission on Environment and Development |
| WHO | World Health Organization |
| WPR | Western Pacific Region |
| WTO | World Trade Organisation |

# 1
# General Introduction

## 1.1 Background

*'For, indeed any city, however small, is in fact divided into two, one the city of the poor the other of the rich; these are at war with one another; and in either there are many smaller divisions, and you would be altogether beside the mark if you treated them as a single State. But if you deal with them as many, and give the wealth or power or persons of the one to the other, you will always have a great many friends and not many enemies'* (Plato et al., 1999 p. 137-138).

In the extract above, Plato was referring to one of the propositions of Socrates on the menace of segregation and inequality in ancient Greece. The policy solution he offered at the time was that it would be more appropriate to embark on a differentiating strategy rather than a one size fits all approach.

More than 2000 years after the teachings of Socrates, Charles Booth - a conservative social reformer in London began an inquiry aimed at debunking poverty related findings of the Socialist Federation (Pfautz, 1967). Booth's work was novel in many respects. First, it allowed for the first time the production to street level maps of social conditions of people living in London. The categorisations and colour coding of these maps took into consideration the idea that social heterogeneity occurs within spatial structures. Booth's approach to conducting the analysis served as a useful foundation to contemporary and modern day multivariate spatial analysis (Pfautz, 1967). This is where multiple authors converge and agree that the idea of geodemographics and social area segmentation originated from (e.g. see Harris et al., 2005; Vickers, 2006).

Following Booth's experiments, the concept of social area segmentation extended into the fields of urban sociology with multiple models evolving overtime within the Chicago School (Robson, 1971).

The idea of geodemographics is underpinned by the notion that people and localities which are close to one another have an inclination to exhibit similar characteristics (Birkin and Clarke, 2009; Harris et al., 2005). In the field of quantitative geography, such behaviour of interdependence of features in space is what Longley et al. (2001) described as spatial autocorrelation.

Well and beyond the notion that closely located people or places are alike (Tobler, 1970), geodemographic intricacies take the laws of spatial dependence a step further to institute that geographical distance is not necessarily a good criterion as distant features may embed characteristics which are very similar to each other (Vickers, 2007). In general, geodemographic principles are anchored on the idea that collections of spatially autocorrelated entities exist within space.

Since people live in places, we can then infer on this basis that not all areas have the same needs (Ballas et al., 2005). However, different areas within a defined geographical boundary to some extent approximate to a limited number of types (Miller and Han, 2001). Therefore knowing the types of areas that exist in space from the types of people living in them can be very useful in ascertaining the types and levels of needs that exist.

Geodemographic systems are created from a series of multivariate analytical techniques called cluster analysis (Harris et al., 2005). It makes sense to deploy these techniques on geographical entities of smaller spatial granularity to derive greater insight. This is one of the areas where the power of geodemographics has been exploited in most developed countries where such systems are available; its discriminatory ability to reveal what would otherwise be hidden at higher levels of geography to policy makers and researchers.

There is no need to recite here how much developed countries have benefited from uncovering local level disparities in the social makeup of their societies (Brown, 1989 and 1991; Brown et al., 1995, 1998 and 2000) and how this has helped to some extent re-shape what Dorling and Ballas (2008) described as a pear-shaped nature of poverty and inequality. Addressing the problem of social and spatial inequality at local geographical scales helps ensure that a substantial number of at-risk or in-need populations are reached and does not present imperfect impressions of a one-size fits all often given at higher levels of geography. Increasingly, geodemographic techniques are being adopted within public sector and academic research because of their ability to provide useful and effective summaries at local spatial scales.

The depth of social and economic disparity on multiple fronts within many developing countries is great (Fields, 1989; Nelson, 1998) and the gap between most developed and developing countries widened considerably until the beginning of the current millennium when the Millennium Development Goals (MDGs) were instituted (UN, 2000). Since the re-emergence of a collective global agenda aimed at addressing some of the challenges of developing world countries, the concept of inequality has re-emerged at the heart of programmes by numerous international agencies and indeed national governments. Unfortunately however, there still remains a failure particularly in developing countries to begin to tackle the problem at the local level where it is believed that the benefits of central government can reach the citizenry.

At a well-received talk titled '*Let my dataset change your mindset*' given by Professor Hans Rosling in the summer of 2009 to the U. S. State Department, he debunked myths about the developing world. Rosling (2009) noted and explained how much of the world's global policy makers are uninformed of global social trends. He painted a powerful picture of global convergence in his talk. Of greater importance however was the fact that he stressed the need for international agencies like the World Bank to move away from current approaches of social and spatial analysis. Rosling called for a discarding of simplistic extrapolation of data noting that if the scourge of inequality is to be addressed, then it '*has to be made a local issue*' (Rosling, 2009).

The work presented in this thesis addresses some of the issues raised by Rosling. Within-country inequalities present a more pressing challenge for policy makers. Rosling argues that policies should be shaped to cater for the groupings or clusters of people and areas which consciously or unconsciously exist within nations. This is one of the reasons geodemographics can serve as a veritable option for addressing the impact of local level inequalities and monitoring change.

At the commencement of this research, an inventory of countries which have had a taste of geodemographic techniques was created. It was found that no low income country has experienced the 'geodemographic revolution'. While commercially available systems are wide-spread in many developed countries, only one country – the United Kingdom has an open-source system (Vickers and Rees, 2007). The research presented in this thesis demonstrates how similar open-source geodemographic classification systems can be developed and used to address the needs of developing countries, by using Nigeria and the Philippines as case study regions.

Nigeria is currently Africa's most populous nation with over 140 million people (NBS, 2006a), multiple socio-cultural and linguistic groups numbering over 250 (Gordon, 2003) and enormous inequalities and social challenges (NPC, 2004a; Ogunbodede, 2006). The country's 36 states and 1 Federal Capital Territory (FCT) are split into 774 Local Government Areas (LGAs) where it is expected that the benefits and dividends of democratic governance can accrue to people at the grass roots. This spatial scale therefore forms the basis of Nigeria's first geodemographic experiment which will be discussed in this thesis.

The Philippines on the other hand is one of the most populated countries in the Asian continent. It ranks 12th globally and 7th in Asia behind China, India, Indonesia, Pakistan, Bangladesh and Japan (ADB, 2009). The current population of the country is about 88 million people. The dynamics of this population spreads across over 40,000 barangays. Barangays constitute the finest level of administrative geography in the country and are expected to provide residential accommodation to populations between 50 – 100 households. Data from the national census of year 2000 are used to explore the potential of geodemographics for the Philippines.

## 1.2 Aims and Objectives

This thesis is about exploring new ways in which multivariate spatially referenced information can be explored, simplified and mapped locally within developing countries. The overall aim of the thesis is to explore and demonstrate the policy relevance and potential of geodemographic classification systems for the developing world, focusing on Nigeria and the Philippines. In order to achieve this aim, a number of objectives were formulated:

**Objective 1:** Critically review relevant literature, data sources and methods in relation to the potential and policy relevance of geodemographics for developing countries.

**Objective 2**: Create small area geodemographic systems for Nigeria and the Philippines.

**Objective 3**: Develop value added outputs and user guides for both classification systems.

**Objective 4**: Test the effectiveness and usefulness of the segmentation systems in relation to policy issues such as the UN Millennium Development Goals.

**Objective 5**: Evaluate the success of the research project and determine future directions of further work.

The principal aim of this research project is ambitious and will form the first novel attempt at creating open source geodemographic systems for both continents – Africa and Asia. In view of this, it is expected that the research may present monumental challenges.

In order to achieve the overall aim therefore, the first objective will seek to establish the potential and usefulness of geodemographic approaches for developing countries. Doing this will not only serve as a contribution to the chapters of the thesis, but is considered essential when trying to convince the national authorities from which the potential datasets would be sourced.

Understanding the theoretical and methodological platforms upon which geodemographic systems are built is also paramount. Therefore part of the first objective will be to undertake a comprehensive review of various spatial statistical techniques relating to geodemographic analysis and establish their pros and cons.

The second objective will build on knowledge gained from the first objective to develop the first set of small area open source geodemographic systems for Nigeria and the Philippines.

Upon the successful assignment of small areas in both countries to their respective groupings and a critical discussion of the analytics, Objective 3 will focus on their usability. The usability of both geodemographic systems will be enhanced by embarking on the creation of value added products like users guides to summarise the detail of the work without statistical jargon. Such final products are what most non-technical policy makers are typically interested in.

Objective 4 will embark on a journey of testing the systems created for both countries by profiling datasets that were not included in the build process. The variables that will be analysed using various statistical techniques are linked to the Millennium Development Goals (MDGs).

The final objective will address the potential of future work that will ensue from the research project.

## 1.3 Thesis Structure

The various aims have been linked to one or more underlying objectives in section 1.2. The objectives themselves are major pieces of work which require a logical approach to achieving them. To this end, this thesis has been structured in a manner that is coherent with the stated objectives as shown in Table 1.1.

**Table 1.1: Thesis Outline**

| | Chapter Title | Objective |
|---|---|---|
| Chapter 2 | Underpinnings of Social Area Classification | 1 |
| Chapter 3 | The Millennium Development Goals and Their Nexus with Geodemography | 1 |
| Chapter 4 | A Framework for Building Classification Systems in Developing Countries | 1 |
| Chapter 5 | Classifying Nigeria's Local Government Area's | 2 & 3 |
| Chapter 6 | Geodemographic Segmentation for the Philippines | 2 & 3 |
| Chapter 7 | Investigating Primary Education in Nigeria and Philippines | 4 |
| Chapter 8 | Addressing Female Empowerment and Maternal Health Care | 4 |
| Chapter 9 | Conclusions: Breaking the Fallow Ground | 5 |

In chapter 2, a definition of geodemographics is provided. This is followed by a historical account of the origins of geodemographic thinking and its expansion and application particularly within developed countries where its origins lay. In addition to the strengths of area classification, the chapter also addresses some of the reasons why developing world countries have failed over the decades to imbibe the potential of geodemographics.

Chapter 3 titled *The Millennium Development Goals and Their Nexus with Geodemography* takes a deeper look into the developing world and tries to uncover modern day challenges. In particular, the chapter focuses on the 8 MDGs which have continued to generate cohesion of international debates and shape policies. The chapter addresses the three hierarchies of the MDG framework by discussing how they are measured particularly within a spatial context. In spite of the supposedly good intentions of the MDGs, the chapter argues that some of the approaches adopted by the UN and partner agencies are in-appropriate and outlines their spatial and non-spatial shortcomings. This is subsequently followed by a thorough discussion of what geodemographics can offer developing countries hoping to move closer to attaining the goals by 2015.

Having established the potential of geodemographics for developing countries, Chapter 4 offers a comprehensive review of quantitative and qualitative methods involved in creating a geodemographic segmentation system. The chapter weighs the different statistical approaches encapsulated by cluster analysis, evaluates different types of clustering algorithms, identifying the strengths and weaknesses of each and discusses some of the issues that should be taken into account when deciding on cluster numbers. In addition the chapter also provides an account on the process of visualising outputs from the analysis.

Chapter 5 builds on the knowledge gained from the second, third and fourth chapters to apply it in practice. The chapter provides a detailed outline of how a geodemographic segmentation system was developed for all the LGAs of Nigeria. The chapter accounts for the difficulty in securing datasets and goes on to justify the different decisions that were made in the process of developing the classification. The chapter also explains in detail how the outputs from the analysis were evaluated and provides added value outputs in form of maps and a user's guide for the new system.

Chapter 6 presents a thorough discussion of how a small area geodemographic classification system was developed for another developing country – the Philippines. The Philippines presented its own new challenges which were however mitigated on numerous fronts by the experience and knowledge gained from the procedures in chapter 5. Unlike the Nigerian classification system which is composed of only continuous variables, that of the Philippines combines both continuous and categorical variables and this necessitated a clustering algorithm different from the Nigerian system.

Chapter 7 puts both segmentation systems to the test by profiling datasets which were not included in the build process against it. Most of the variables analysed in this chapter are related to the MDGs which makes the analysis highly policy relevant. The chapter focuses on the evaluation of primary level education in both countries.

Chapter 8 addresses another topic which is of great importance to both countries. The focus of the chapter is women and children. Again this sort of analysis is also very novel due to the geographical scale of the analysis and the ability of to provide information of potential areas where policies should be targeted. Some examples of issues analysed include economic empowerment of women in Nigeria, educational attainment among women in the Philippines and child maternal health care.

The final chapter of the thesis evaluates the success of the project and summarises some of the key findings. The chapter also provides insight on potential areas of future research linked to geodemographics within developing countries.

In addition to the main body of the thesis, appendices showcasing more outputs from the work follow a detailed reference list. A compact disc containing user guides and spreadsheets with geodemographic coding and visuals for all the Nigerian Local Government Areas (LGAs) and Philippine Barangays also accompanies the thesis.

# 2

# Underpinnings of Social Area Classification

## 2.1 Introduction

The aim of this chapter is to provide an introduction to some of the historical, theoretical and practical underpinnings of area classification and geodemographics. While focusing on developed societies, the issues that will be examined will revolve around the reasoning behind early and contemporary approaches to clustering people and places into groups. The chapter is organized as follows: Section 2.2 discuses definitions for geodemographics, and considers the requirement for order within social, physical and economic contexts arguing that lack of an understanding or presence of such hierarchical structures often results in 'anarchy'. Section 2.3 explores the rise of area classifications from early times through to the present day. In section 2.4, the practical applications of geodemographics is provided and aligned along three spheres including the public sector, private sector and academia. Section 2.5 considers the imperfection of geodemographics by investigating associated problems relating to spatial, temporal and ethical concerns. Section 2.6 focuses on some key reasons why developing countries lack geodemographic systems. Section 2.7 draws the chapter to a conclusion having investigated theories, practical applications and the strengths and weaknesses of geodemographics.

## 2.2 Defining Geodemographics and Area Classifications

An array of terminologies has emerged over time to help describe the key issues concerned with the estimation of similarity. Much of the work that has been done within this area centres on cluster analysis with various disciplines aligning their terminologies in a way that does not overlap.

This work embraces the concepts and techniques of clustering but focuses on its use in the classification of areas within a socio-economic and demographic context. Simply put, social area classification is a way of segmenting geographical units into groups based on the socio-economic features of their residents (Vickers and Rees 2006).

Numerous schools of thought have emerged with different definitions of geodemographics. Sleight (1997) defines it as the study of people by where they live. There is no doubt that a significant part of the application area is centred on market area analysis (Longley et al., 2001). Brown (1991) particularly stressed that geodemographics concerns itself with area-based typologies and as a result proved to be adequate discriminators of the behaviour of consumers. While we can not contend that the commercial sector has embraced the theory and practical applications of geodemographical analysis in predicting consumer behaviour, it must be emphasized that geodemographic theory transcends beyond commercial classification systems.

Many concepts within the social sciences are characterized by broad definitions. The classification of areas which in simple terms links *social space* and *geographical space* (Burrows and Gane 2006) and tries to identify similarities and dissimilarities can also be characterized by subjective purviews. Voas and Williamson (2001) illustrated why it may be difficult to define geodemographics. If we are to claim a small geographical area is different from another, we would first have to clarify what we mean by 'different' because of the uniqueness of these areas if their inhabitants are examined using one quality at a time (Voas and Williamson 2001).

Much research within the sphere of geodemographics is centred on a view that people cluster together as a result of similar characteristics. This view was corroborated by Foley (1997, p. 6) who wrote as follows:

'*The theory of geodemographics is based on the fact that similar people tend to cluster together and households in the same postcode sector or enumeration district can be placed in the same category*' (Foley, 1997, p. 6).

The statement draws upon a sentence which was used to justify heuristic calculations in an urban growth simulation and eventually became known as the first law of geography. 'Everything is related to everything else, but near things are more related than distant things' (Tobler, 1970, p. 236). Vickers (2006) added to Tobler's law to fit geodemographics and coined the following statement:

'*People who live in the same neighbourhood are more similar than those who live in a different neighbourhood, but they may be just as similar to people in another neighbourhood in a different place*'. (Vickers, 2006, p. 16)

Geodemographic underpinnings assume that there is a linkage between the socio-economic landscape and the conglomeration and clustering of groups of behaviours.

### 2.2.1 The Need for Hierarchical Ordering

As human beings, we have different wants which makes us different and also helps to shape the way in which we live, work and recreate. An understanding of the complexities of human nature can also be indicative of the location of the neighbourhood of an individual, what it is like, the other type of residents and how they behave (Harris *et al.,* 2005).

There is a natural and unconscious tendency for the human mind to structure entities into groups. This provides a simplification and an understanding of the world around us. In essence, the capacity to assimilate quickly and transform information to knowledge is dependent on the simplicity of presentation. When objects are identified as different on the basis of their attributes, it provides a mass of too many different objects. If similarities are identified amongst these different entities, it sets in a law of order and more similar objects can be pulled together giving rise to groups of similar objects. In addition to minimizing confusion and enhancing understanding, the introduction of a law of order within a system based on the characteristics of the constituent units reveals hidden patterns and processes that exist within similar groups and between dissimilar groups (Brown *et al.,* 2000). The law of order which gives rise to a classification scheme is what Everitt *et al.* (2001) refer to as clustering.

## 2.3 Evolutionary Trends of Area Classifications

Many authors have given accounts on the origins of area classifications and geodemographics. Harris et al. (2005) gave a detailed account of the origins of geodemographics. This section summarises the trends in the development of the field.

### 2.3.1 How and Where it Started (Pre-1980's)

The origin of area classifications is generally credited to a survey conducted by social reformer Charles Booth in the 19th century. This was an unprecedented enquiry of the social and economic conditions of the people of London (Orford et al., 2002) which lead to an innovation. For the first time, large detail maps showing the social class of London at street geography were produced.

The rationale behind Booth's inquiry is described by Bales (1991) as a complex impetus driven by a blend of political and philosophical ideologies and a sense of social obligation. He conducted a survey which revealed that the incidence of poverty in London was far greater than he had imagined. From his results, he concluded that 30.7% of the city's population was actually below the poverty line (Simey and Simey, 1960).

Booth's first survey on poverty was aimed at showing that poverty incidence could be measured accurately (Orford et al., 2002). This, he expected would eventually influence the way policy was designed and ensure such policies met actual measured needs (Bales, 1991). Booth employed a team of researchers to help him. The crux of the data used was results from lengthy interviews with professionals who according to Orford et al. (2002) had expert knowledge and experience of working with those residing in London. Amongst these professionals, the most important for him were the school board visitors whom he strongly believed had detail knowledge of social conditions and poverty (O'Day and Englander, 1993). The information gathered from the school board visitors and notes he made during the survey were used to paint general socio-economic conditions in which people lived.

Developments in the Chicago School of Urban Sociology characterised the next phase of developments. The group at the University of Chicago constituted of urban sociologists who worked on a number of representations of social city structures (Robson, 1971). The concentric ring model was developed by Ernest Burges in 1925 and is patterned after Von Thunen's theory which relates to rural or agricultural land around a city or market centre (Akinmoladun, 1999). Other models developed by the school include the sector theory developed by Homer Hoyt in 1939 which is of the view that high rent residential neighbourhoods are instrumental in shaping the land use structure of a city (Akinmoladun, 1999) and the multiple nuclei model presented by Harris and Ullman in 1954 which is of the view that a number of separate nuclei actually shapes the land use pattern of a particular city.

Further developments in the United States were sparked by the publication of census small scale data typically for census tracts. This publication enabled a statistical methodology to be used for the first time to segment social areas in Los Angeles and San Francisco (Shevky and Williams, 1949; Shevky and Bell, 1995). More relevant research was conducted in America following the release of census data in 1960 covering more cities.

Another significant era in the development of geodemographic systems in the United States was between the late 1950's and early 1970's. The developments during this period were driven by Jonathan Robbin who is credited with pioneering contemporary, computer based geodemographic systems (Burrows and Gane, 2006). He combined theories from developments in the Chicago School with concepts within the sphere of factorial ecologies of positivist urban social sciences (Burrows and Gane, 2006). This enabled him to produce profiles of residential ZIP code areas and also formed much ground work for the Potential Rating Index for ZIP marketers (PRIZM). This first modern geodemographic system (Burrows and Gane, 2006) was based on an analysis of the clusters of data from the Census of the United States and consumer surveys.

Batey and Brown (1995) noted that in spite of the lead role set by Charles Booth in the United Kingdom, nothing much was done in the United Kingdom until the 1960s. Moser and Scott (1961) employed techniques within factor analysis and were able to aggregate 157 British cities and towns into 14 groups based on four factors: Social Class, Population change from 1931 to 1951, Population change from 1951 to 1958, and overcrowding. These four components were drawn out from an initial 57 variables and allowed them conclude on which cities were more alike than others. Vickers (2006) concluded that this research and the work conducted by Gittus (1964) in Merseyside and south-east Lancashire led to a resurgence in area classification in Britain.

While Robbin was adding a modern purview (developing PRIZM) to area classification in the USA, the work of Webber (1977) and Webber and Craig (1976 and 1978) also served as a platform for the development of modern day geodemographics (Burrows and Gane, 2006; Vickers, 2006; Harris et al., 2005). During the 1970s Webber who was working with the Centre for Environmental Studies (CES), London, commenced a development of national classifications at small scales (ward, parish and local authority levels) (Harris et al., 2005). With this development, clusters of Enumeration Districts (EDs) could be analysed comparatively with the national mean for certain census variables (Webber and Craig 1976 and 1978).

This development was significant because it had been previously difficult to embark on a country-wide comparison because of the different methodologies employed by the different studies which had been conducted (Vickers, 2006). The classification system was subsequently acquired by a company called CACI. CACI was established in 1975 and primarily concerned itself with providing Information Technology (I.T.) and marketing solutions for organizations in different sectors of the economy. CACI launched Webber's classification with a new name - *A Classification of Residential Neighbourhoods* (ACORN). A linkage of the classification system to the postcode geography provided significant discriminating information about consumer behaviours in different neighbourhoods for practitioners in the private sector (Harris et al., 2005). A continued development of versions of ACORN instituted the modern day geodemographic industry (Baker, 1997).

### 2.3.2 Geodemographics in the 1980's and Beyond

CACI remained the sole player within the UK geodemographics market between the very late 1970s and early 1980s, but competition from other emerging companies resulted in a gradual drift of some of CACI staff (Sleight, 2004; Harris et al., 2005). Developments in the industry in the mid-1980s were enhanced by the release of the 1981 Census. Webber moved from CACI to CCN and the first version of Mosaic was launched in 1986. While Webber (1977) used only census data, this period marked an era when non-census data was introduced into classification systems.

The increased use of geodemographic systems within the private sector had a number of underlying reasons. Of particular importance were significant developments which occurred within the sphere of retailing and marketing. Beaumont and Inglis (1989) identified a transition from an emphasis on 'mass marketing' to what they termed as 'niche marketing'.

The dominance of the industry by commercial segmentations seemed to have overshadowed academic interests in developing area classifications in the U.K. The first purely academic research oriented attempt is attributed to Marcus Blake and Stan Openshaw who developed a general purpose classification system using an unsupervised neural net technique (Blake and Openshaw, 1995). The methodology is based on Kohonen's self-organizing map and provides an avenue whereby the number of assumptions is reduced as much as possible and the sources of data uncertainty are incorporated (Openshaw, 1994a). The system however left a number of areas unclassified.

The most recent arrival in the U.K. industry is the Office for National Statistics (ONS) Classification of Output Areas (Vickers et al., 2005; Vickers and Rees, 2006). This is the first attempt with a published methodology aimed at creating a classification system at output area level. This is the smallest spatial scale at which the 2001 UK Census data were released. A total of 41 variables were selected from an original 129. These were categorized into five domains: demographic, household composition, housing, socio-economic and employment (Vickers, 2006). The Output Area classification employed an iterative relocation algorithm commonly referred to as K-means. The ONS classification is in three hierarchical groups. The first hierarchy is referred to as super-groups which consist of 7 clusters; the second consists of 21 clusters and is referred to as the groups while the third tier called sub-groups comprises 52 clusters (Vickers and Rees 2006). The classification can be accessed freely on the Office for National Statistics website.

The Canadian and U.S.A experience has been a little different. Their markets have been dominated by commercial segmentations which focus on market research activities.

Following the release of data from the Canadian Census of 2001 and the addition of a new geographical unit called Dissemination Areas (D.A.) (Environics Analytics, 2004) a team of researchers from a company called Environics Analytics Group (EA) developed what is one of the most recent geodemographic classifications in Canada- ***PRIZM$_{CE}$***.

PRIZM$_{CE}$ embeds selected variables from Canadian Census and lifestyle data described as behavioural and attitudinal variables derived from surveys. The argument for inclusion of these variables is that by adding these variables, the classification will work better in practice revealing trends not covered in the Census (Environics Analytics, 2004). The problem associated with this is that the databases from which these lifestyle data are adapted from cover only a fraction of spatial units from the national structure. In addition, the geography for sampling the behavioural and attitudinal variables is different from the geography of the census data (53,000 D.A's) it is claimed to supplement. This means data for other areas are interpolated or simulated and of course is prone to errors in the process. PRIZM$_{CE}$ was built at D.A. level as opposed to the nearly 800,000 residential postcodes of Canada due to non-availability of geodemographic data at such small scale. Just like other commercial companies in the U.K. and U.S.A, the methods used to develop PRIZM$_{CE}$ have not been made publicly available.

There is currently no knowledge of any general purpose system like the U.K. ONS classification of Output Areas built in Canada or the U.S.A. solely from national census data.

### 2.3.3 Geodemographics across the World

The spread of geodemographic techniques across the rest of the world has been relatively slow. Most of the companies that have capitalized on the techniques are commercial in nature aimed at market targeting. They add non-census data to available census statistics. Lifestyle data which is a term encompassing inventories collected based on sampled surveys of consumer choices and behaviours are integrated extensively with census data (Sleight 2004; Harris et al. 2005). The addition of such data to the classification systems helps fill the gap of trends not covered by the census according to Harris et al. (2005). Vickers et al. (2005) however argue that in terms of quality, such data are lacking in evaluation, are biased towards the affluent section of the society and are not comprehensive in their coverage.

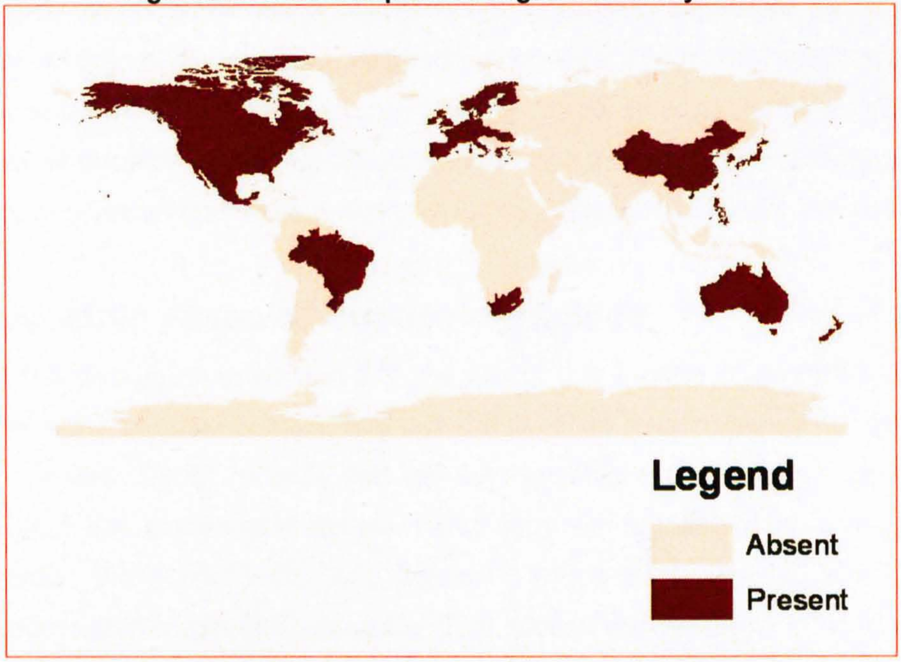
**Table 2.1: Overview of Major Geodemographic Classification Systems**

| Supplier | System | Website |
|---|---|---|
| AFD | Censation | www.afd.co.uk |
| Allegran | Gnuggets | www.allegran.com |
| Beacon Dodsworth | $P^2$ People & Places | www.beacon.dodsworth.co.uk |
| Business Geographics | Locale www.businessgeographics.co.uk | www.businessgeographics.co.uk |
| CACI | Health ACORN | www.caci.co.uk |
| CACI | ACORN | www.caci.co.uk |
| Claritas Acxiom | PRIZM | www.claritas.co.uk |
| ONS | Output Area Classification | www.statistics.gov.uk |
| The Clockworks | SONAR | www.theclockworks.co.uk |
| Acxiom | Personicx Geo | www.acxiom.co.uk |
| Euro Direct | CAMEO | www.eurodirect.co.uk |
| Experian | Mosaic | www.experian.co.uk |
| GeoBusiness | ATOMICube | www.geobusiness.co.uk |
| ISL | RESIDATA Lifetypes | www.isl-online.com |
| Streetwise | Likewise | www.streetwise-analytics.co.uk |

Source: Author's Research Findings

Based on the global penetration of commercial suppliers like Experian, Eurodirect and some others listed in Table 2.1 above, Figure 2.1 provides a global perspective to countries which have had a taste of the geodemographic experience.

**Figure 2.1: Global Imprint of Segmentation Systems**




Source: Author's Research findings

These commercial companies develop single classifications for profiling consumers from these countries in terms of their affluence and consumption lifestyles. Mosaic Global for instance covers over 16 countries and is developed on the notion that *'the world's cities share common patterns of residential segregation'* (Experian, 2005).The following premise summarises the methods used to develop the system.

*'Using local data from 16 countries and statistical methods, Experian has identified 10 distinct types of residential neighbourhood, each with a distinctive set of values, motivations and consumer preferences which can be found in each of the countries'* (Experian, 2005 p. 1).

This implies that one of the goals of developing a global system was to streamline the countries into homogenous groups based on similar characteristics. However, a key disadvantage of this approach is that it impairs the variations existing in each individual country. In addition, the use of similar names for identifying the neighbourhoods is not good practice. It narrows down the socio-economic characteristics of the countries.

Clearly, Figure 2.1 shows the absence of geodemographic technologies in many developing world countries. Such countries are faced with significant socio-economic challenges which make planning, resource allocation and policy making important issues for sustainable development especially at lower level geographies. Over time, some of these countries' statistical departments have accumulated volumes of national data relating to socio-economics and demography. These have been under-utilised.

## 2.4 Application Areas of Geodemographics

Following the discussion in section 2.3, the aim of this section is to review a series of application areas of some specific and general purpose geodemographic classification systems. Vickers (2006) records that geodemographic classifications are employed heavily within the commercial sphere particularly for marketing. In addition to the private sector, the public sector and academic arena have also benefited from the practical application of geodemographics. This section will therefore strive to present a seemingly balanced view of these three key areas of application.

### 2.4.1 Public Sector and Academic Research Applications

The exploitation of the discriminatory power of geodemographics within the public sector often sparks interests in issues related to resource targeting (Brown et al., 2000). The profiles derived from geodemographic typologies are increasingly playing a central role in ensuring effective deployment of resources by public services (Ashby and Longley, 2005). Some of the key areas where area classification typologies have played important roles include public health, community safety, education and politics. Area typologies have provided room for informed decision making within these sectors because of their capacity to provide simplified representations of the mass of detailed counts of data held in small area statistics (Brown et al., 2000).

The integration of GIS technology and geodemographics has greatly enhanced health-related applications. Staff at the Urban Policy Evaluation and Analysis Research Laboratory (PEARL) formerly Research and Policy Evaluation Regional Laboratory (URPERRL) were able to exploit these technologies significantly in planning health services (Brown et al., 1995; Hirchfield; et al., 1995; Todd et al., 1994). Brown et al. (2000) identified an approach to deploy health related resources. They suggested the production of a penetration ranking report in which for instance the target markets are ranked in terms of the rate of incidence of a particular health condition. Another classic example of application in the medical arena is the case where area typologies have been used to contrast sample areas from which cancer records were examined (Brown et al., 1998).

Geodemographic typologies have also been employed extensively in targeting resources for community safety and combating crime. Cases are drawn from Brown et al. (1999) and Ashby and Longley (2005). The needs of law enforcement agencies at local levels are greatly enhanced by revelations from geodemographic profiles, crime analysis and mapping. Bowers and Hirschfield (1999) have demonstrated linkages between profiles of areas of economic deprivation and crime. Brown et al. (1999) illustrated how area targeting can be a useful resource in identifying potential areas of concentration of fire risk in the Greater Manchester area in the UK. The research was commissioned by the Greater Manchester County Fire Service (GMCFS). The key premise underpinning their conclusions is summarised in the notion that knowledge of the ward and enumeration district of individual incidents can be used to establish the total number of similar incidents of a specific type of ward and more importantly can be used to derive rates of incidence with respect to the population at risk. They later recognized that a large proportion of fire incidents were concentrated in a relatively small number of wards many of which demonstrated similar socio-demographic characteristics (see Brown et al., 1999).

The third application area is drawn from examples from the UK political arena. Webber (2005) explored the effectiveness of segmentation systems employed by British political parties in the 2005 General election. Three areas of activities involved in planning for election campaigns were recognized. These include 'the briefing and management of national and regional news media to ensure favourable editorial coverage; the articulation of political messages through mass media such as newspapers and poster sites and through party political broadcasts; and the targeting of specific messages at individual electors at their home addresses' (Webber, 2005 p. 1 & 2). The challenge for many political parties however in implementing identified segmentation strategies is how to map the different segments based on demographics and lifestyles to onto key issues appearing in party manifestos.

This section has reviewed a number of applications of geodemographic typologies within research and public policy. Small area discriminators create room for value to be added to knowledge especially as derived from the census and other reliable surveys (Brown et al., 2000). A variety of data mining and exploratory analysis can be employed complementarily with geodemographics to improve performance in research and public policy making. The next section explores applications within the private sector where classifications appear to have been appreciated significantly.

## 2.4.2 Private Sector Uses

Commercial use of geodemographics has seen tremendous growth over the decades in the UK, USA and some other developed countries around the world. A contemporary approach to commercial segmentation systems has seen many of the companies which have developed these systems adding new sources of data to their classifications (Sleight, 2004). It has been argued that these data sources, 'lifestyles data in particular' lack quality assurance and are potent sources of bias against economically disadvantaged sections of the community (Vickers et al., 2005).

In section 2.3 it was recognised that in addition to increasing the quantity and quality of information used to plan strategically in retailing and marketing, the re-orientation from a mass market-centric approach to a niche market-centric approach (Beaumont and Inglis, 1989) acted as a catalyst in the growing development and use of commercial segmentation systems. Realisations that systematic and analytical methods could help reduce the costs associated with uncertainty in management of retail outlets. Commercial companies which currently dominate the geodemographics arena recognized the discriminatory power of small area typologies, as a means of differentiating areas which exhibit contrasting patterns of behavioural or other associated characteristics. Brown (1989) identified some key specialised areas of application of geodemographics in the commercial sector. He categorised these under four areas:

1. Retail site analysis and store location
2. Market research applications
3. Credit Rating and
4. Target Marketing

One key application area of geodemographics in retail is the use of the propensity of different lifestyle groups to consume particular products or services as an input when considering location allocation for new stores. When relating similar consumption rates to established profiles of prospective customers, it is possible to generate estimates of potential sales or consumption (Brown, 1989). Numerous approaches have been employed to answer related questions. The ideas that underpin the likelihood of a potential customer to visit a retail outlet are embedded in what is popularly referred to as spatial interaction models (Harris et al., 2005).

$$\boldsymbol{F}_{j(i)} \propto \mathbf{M}_j/(\boldsymbol{d}_{ij})^c \tag{2.1}$$

where:

*i* is the location of the consumer

*j* is the location of the retail outlet

$\mathbf{M}_j$ is the stores mass (the attraction a store has to offer to the consumer)

$\boldsymbol{F}_{j(i)}$ is the 'force of attraction exerted by the store at j in the consumer at i

c is a function which controls the distance.

This model has become very popular when embarking on a location allocation analysis. A note of caution is however sounded by Harris et al. (2005) regarding the unpredictable nature of consumer behaviour which may not be as lucid as the model assumes. In addition potential competition from other retail outlets will also impact on the behaviour of the model. This has given rise to calls from a number of authors (Fotheringham and O'Kelly, 1989; Fisher and Getis, 1999) to modify and adapt the model to fit some of these unpredictable circumstances.

A density surface is often seen as a suitable way of showing where existing customers are spread and also provides some knowledge of the concentration of such customers. However the need still often arises to devise novel approaches and market-specific discriminators as more information becomes available (Brown, 1999). An application of geodemographics in target marketing has been refined and explored significantly by commercial practitioners involved in address list processing. This systematic selection of names from address lists was what gave birth to 'Junk mail' (Brown, 1999). The messages disseminated via direct mail are expected to normally to generate greater level of attention; however Openshaw (1989) noted that response rates were very low. One of the key reasons that may account for this is information load. It has been suggested that 'efficiency may be improved by examining the variation in response rate between geodemographic system clusters' (Brown, 1999).

Another illustration of target marketing is drawn from a German example. Sliwinski (2002) showed how spatial point pattern analysis and kernel density estimation techniques based on geodemographic data can be used to establish description of target areas of potential customers within a GIS environment. He derived kernel estimates of 4089 customer records which were test data and generated socio-economic and demographic characteristics of customers by intersecting their spatial point patterns with survey data. This approach enabled him to model customer density patterns in a bid to target prospective new customers.

Other application areas include customer profiling and generation of loyalty schemes and club memberships. Callingham (2003) describes this as a social classification rather than a geographic classification. This view was shared by Burrows and Gane (2006). Applications in real estate and related industries are also demonstrated by Thrall (1998).The field of geodemographics has thrived tremendously in the commercial world because of its capacity to attribute market values to geographic areas (Callingham, 2003).

## 2.5 Shortcomings and Criticisms: All that Glitters is Not Gold

In the previous section, examples of successful and beneficial applications of geodemographics were examined. However in spite of this success especially in the commercial world, geodemographic classifications have attracted some measure of criticisms. Most of these centre on theoretical and methodological issues of the ecological fallacy and the modifiable areal unit problem (MAUP) and the temporal issues surrounding data. In addition, concerns have been raised on ethical implications (Burrows and Gane, 2006). This section focuses on a review of some of the criticisms leveled against geodemographics.

### 2.5.1 Scale and Aggregation

The relevance of scale when embarking on any form of spatial analysis is of immense importance. Scale is crucial not only when aggregating spatial data but also when using the derivatives of such spatially aggregated information. The earth's constituents and processes are naturally structured into hierarchies (Marceau, 1999). This helps to explain the increasing recognition of scale as a central concept in all levels data capture, analysis, modeling and visualisation.

The concept of scale and aggregation within the sphere of geodemographics is not particular to that field. This is explained by the confusion that still appears to dominate certain concepts in spite of the availability of abundant literature (Marceau, 1999). It is a widely held view that spatially aggregated datasets are generally imperfect by definition. There are always potential errors of some sort, missing values, distortions due to modeling the real world and deliberate adjustment of attribute and positional data due to privacy reasons (de Smith, 2006).

One reason for the definition of any census geography is to facilitate enumeration while electoral geographies are defined to ensure that the voting patterns and attitudes of the population are as fair as possible. There is therefore that tendency that these geographies which constitute the bedrock required for many of the neighbourhood and spatial analyses we do in geodemographics were designed without taking these potential analyses into cognizance (Harris et al., 2005). The questions that may arise then is that since these geographies were designed for some other purpose, what is the probability that socioeconomic data aggregated and modeled based on discrete phenomena reveal any meaningful or truthful dimension about the make-up of a derived classification?

When building geodemographic classifications, information can usually be taken from smaller scales to infer processes at larger scales. The same can be done with information at larger scales; they can be broken down into smaller constituent units. These, approaches respectively are what Jarvis (1996) terms as upscaling and downscaling. As a result of these activities, the geometric structure of data and their associated attributes (Marceau, 1999) may be distorted. The implications of this are better explained by the twin issues of the Modifiable Areal Unit Problem (MAUP) and the ecological fallacy.

## 2.5.2 The Ecological Fallacy

As indicated in the previous section, scale and aggregation often results in problems when dealing with grouped data. Apart from spatial scale and arrangement factors, another key issue when resolving spatial problems is that of representativeness. When modeling the real world, it may be practically impossible to secure every single data unit so many spatially related problems have resulted in generalisations and grouping of information. The question of homogeneity therefore arises. If grouped or generalised data are to be truly representative of the constituent units of a group, there must be a significant level of intra-member homogeneity within the group.

The ecological fallacy derives from a situation where members of a group are ascribed with the characteristics when only the overall characteristics of the group are known (de Smith, 2006). It is based on the probability that analysis based on area level data may give rise to conclusions seemingly different from unit level data (Steel and Holt, 1996). All forms of geographically oriented analyses experience the imprint of the ecological fallacy (Vickers, 2006). Steel and Holt (1996) have suggested that the key to analyzing data from grouped populations is the development of statistical models for the individuals, the groups and the interactions between them.

A study is caged by ecological fallacy if its results (based on aggregated zones or grouped data) are used to infer the characteristics of individual units which constitute the zone (Openshaw, 1984a). In theory, an internally homogeneous system will help mitigate its effects; however this is not the case with most geographical researches which focus on heterogeneous zoning systems.

### 2.5.3 The Modifiable Areal Unit Problem

When embarking on any research, the spatial unit of the enquiry is important. Many of these spatial units or objects are 'modifiable', implying that larger areal units are adopted in order to define a measure of spatial association between component units (Openshaw, 1984b).

To help understand what the Modifiable Areal Unit Problem (MAUP) is all about, Taylor et al. (2003) cited an illustration with the UK census data.

*'The UK census collects individual household level data and then aggregates up to a variety of larger zones, such as the Enumeration District, Ward or Local Authority. However, these zones determined for ease of enumeration may bear little resemblance to the social geography of the people they contain. Consequently, the analysis of such data in different zones, or levels, may alter the resulting pattern of aggregated observations'* (Taylor et al., 2003 p. 42)

The effects of the MAUP were examined in detail by Openshaw and Taylor (1977). They considered three experiments within varying statistical and spatial contexts. They explored data relating to the percentage of elderly voters and correlated these with Republican Party voters. The revelations were stunning. They found that if the 99 counties making up the state of Iowa were aggregated into larger and fewer districts, and a number of combinations of these districts were considered, the correlation values could range from +0.97 to -0.81. In addition, Fotheringham and Wong (1991) were also able to demonstrate that the MAUP has implications for multivariate statistical analysis. Vickers (2006) depicted the effect of data aggregation of socio-economic data within the context of the MAUP using arbitrary census areas. From his experiments, he came to a conclusion that there is no real solution to the problem; an inference which is often drawn by many researchers who have worked on the MAUP (see Gehlke and Biehl, 1934; Openshaw, 1984b; de Smith, 2006; Fotheringham and Wong, 1991).

Some propositions have been made to help combat the effects of the MAUP. The application of clustering techniques for grouping data was one of the first approaches. The aim is to ensure that the numbers and configuration of the clusters portray as much information as possible on the phenomena of interest. Openshaw (1984b) and Larson (1986) have criticised this methodology recognizing that zoning systems change with respect to the problem being investigated. Some useful suggestions have also been put forward by Jelinski and Wu (1996). They advocate for the use of individual non-modifiable entities to perform analysis. They however recognize the difficulty with this approach noting that it is not always possible to identify the individual entities of certain measures. They also admit that the approach may result in too much detail giving rise to more complex or impracticable analysis.

Another interesting proposition is to perform a sensitivity analysis to determine which variables are sensitive in terms of scale and configuration and the severity of such sensitivity (Jelinski and Wu, 1996). An approach which is similar to a sensitivity analysis is emphasis of spatial analysis on rates of change. Fotheringham (1989) has advocated this method in order to explore and explain the rate of change of variables and relationships of interest with respect to scale.

While some authors have regarded the MAUP in practice as 'not as big an issue as it might appear' (de Smith, 2006), some others (Openshaw, 1984; Openshaw and Taylor, 1981) have stressed the severity of the problem.

## 2.5.4 Ageing

In addition to the spatial criticisms of geodemographic systems discussed above, there is also a temporal shortcoming. The key data inputs for geodemographic classifications are derived from national censuses (Harris et al., 2005). Even in developed countries where censuses are conducted fairly regularly, it is usually within intervals of 10 years and results are normally published about two or three years after the conduct of the census. Developing countries like Nigeria may take longer periods before publishing results and conducts are not as regular due to political and financial reasons.

Vickers (2006) also recounts, using the UK example that about 12% of people relocate each year and 77% of people retain in their enumerated locations when results are published. He also highlighted the inevitable circumstances of deaths and births and the difficulty associated with updating census data between censuses. However spatial microsimulation techniques are increasingly being recognised as a useful means of dealing with the problem of updating data and predicting statistics for small areas (Ballas et al., 2004; Ballas et al. 2005). This can however only work where there is reliable source data as a base, ideally including good quality population survey microdata.

In developing countries, there are still neighbourhoods which are characterised as slums. As a result, many projects involving slum clearance and urban regeneration take place. George (1999) accounts for examples from Nigeria. Apart from urban or rural renewal activities, housing supply is usually also part of many government budgets on a yearly basis. These activities change the nature of the environment and in particular, the economic status of people who reside in such vicinities and it has been argued that it distorts the validity of census statistics for such areas as they become out of date.

Commercial classifications adopt numerous non-census data in their classifications. Many of these data are associated with the lifestyles of people which are reflected by economic factors and to some extent social trends. These trends change even more regularly making the need to update lifestyle data frequent.

The underlying implication of geodemographic systems is the ability to define the sort of neighbourhood an individual lives from the socio-economic profile of its residents. This implies that these typologies can also act as surrogates for the social patterns within society. Orford et al. (2002) re-examined Charles Booth's work over a century after the survey was conducted. They produced digitised maps from Booths poverty maps of London and were able to examine these within a GIS. In contrast with the 1991 census in the UK, their statistical enquiry revealed that absolute poverty which is defined by Ogunbodede (2006 p. 4) as *'subsistence living below minimum socially acceptable standards'* had decreased in London over the century, however in relative terms – a comparison of the lower segment with the upper segment of the population Ogunbodede (2006) - London had not changed much. Orford et al. (2002) were able to conclude that the poorest people in London still remained in the same neighbourhoods as the time of Booth's survey.

Living standards and the quality of life of people may improve or deteriorate. This will also mean outdating statistics used to predict the characteristics of these people. These changes are not rapid, and are often discrete in space. In most cases people change their neighbourhoods to fit their lifestyles. The extent of the effect of the temporal nature of socio-economic statistics on area classifications will depend on the size of the areal units in addition to the stability of the data inputs (Vickers 2006). Classifications based on large areal units and embedding relatively stable data inputs will be less susceptible to change over time.

## 2.5.5 Over and/or Under Representation

Some of the methodological approaches adopted by developers of commercial classifications are not in the public domain. Vickers (2006) also noted claims by some of these commercial companies about the quality of their products. He argued that by providing so much precise or inferred information on the clusters there is the tendency of 'mis-representing the classification and ignoring the diversity within the clusters'. It is important to provide relatively broad descriptions when commenting on clusters characteristics. The early work of Charles Booth reflected that there is within cluster variability's among members (Harris et al., 2005).

The problem of over representation is a common feature with numerous commercial classification systems. As a result of the lack of transparency and verification of the methodological approaches employed in producing these classifications, it is common practice for commercial vendors to claim that their systems are better than competitors. This often leads to overstating the capabilities of their products therefore causing potential users to apply them wrongly or become disappointed when the classification fails to meet their needs (Vickers 2006).

## 2.5.6 Social and Ethical Implications of an Over-coded Society

A key concept used in geodemographics is based on the idea that people are where they live. Commercial classifications are built on the notion that lifestyles congregate as a result, it is not uncommon to hear statements like *'we know you because we know where you live'*. These classifications embed information from a variety of sources notably credit reference information, customer loyalty schemes, lifestyle databases which are primarily used for target marketing (Curry, 1998).

There is increasing data protection, confidentiality and civil liberty concerns with individual level databases. The capacity of information communication technologies (ICT) to monitor processes and patterns is highly dependent on these databases and as a result, numerous ethical concerns have been raised. Geodemographic classifications are technologies which have emerged and are anchored on databases which some people consider as an infringement on privacy.

Graham (2004) considers geodemographics as an ICT approach for sorting social space. In his view, this will only provide enhanced services to those persons or neighbourhoods deemed as attractive while the less attractive ones will lag behind. The continuous development and provision of geodemographic profiles gives room for people to sort themselves out (Burrows and Gane 2006). Burrows et al. (2005) cite examples of websites in the U.S.A where people supply their socio-demographic preferences and provided with zip codes that match them. Examples like this indicate that using the power and results of geodemographic wrongly could lead to further segregation and polarization within the society.

It has also been argued that geodemographic classifications have caused bias and sentiments to be introduced by certain public and private service providers especially in the commercial sector. They introduce geodemographic filters in cases where everyone should be considered on a level playing ground. Burrows and Gane (2006) argued that this is a negative implication of social-space sorting. They wrote as follows:

*'geodemographic categories accessed via a postcode or something that can be linked to a postcode (such as a telephone number or an e-mail address) can be 'invisibly' used by socio-technical systems to 'softwaresort' places (and of course the people who live in them). This might happen in an increasing number of contexts such as: call-line identification (CLI) queuing systems in (increasingly ubiquitous) telephone call-centres; the determination of insurance premiums; credit ratings (see, for example, www.checkmyfile.com); and, possibly, selection procedures in higher education and for employment'* (Burrows and Gane, 2006 p. 805).

Vickers (2006) described this attitude as postcode persecution suggesting that the labeling of certain areas may contribute to treating all individuals associated with such areal unit being treated unfairly. Most of these practices are linked with commercial applications of geodemographics. Many *profit-centric* commercial organizations throw caution to the wind and exploit the discriminating power of geodemographics negatively. There are also no clear checks and balances in the industry to control the manner in which these companies use the classifications and in situations where government agencies try to streamline their activities, such policies often prove detrimental to well-meaning researchers.

## 2.6 Why Developing Countries Lack Segmentation Systems

Despite the identified benefits of geodemographic systems for making intelligent decision making and its potential for enhancing some of the targets of the MDGs at local level spatial geographies, numerous developing countries still lack these systems. At the vary basic level, some of the reasons for this (data related) may appear obvious, however some more pertinent issues like the misconception of what infrastructure really is has often been overlooked. The final sections of this chapter will discuss the factors that have contributed to the absence of these systems from third world countries.

### 2.6.1 A Misconception of What Constitutes Infrastructure

Due to the scourge of poverty and related endemic challenges, the focus of infrastructural development has been mainly directed at physical infrastructure like roads, water supply and electricity supply. There is no doubt that these things are vital however, this work argues that infrastructure transcends these physical developments.

The emphasis placed on physical infrastructure has all too often clouded the importance of building efficient data infrastructure in developing countries. Ironically, planning and policy decisions on the sitting of these physical infrastructures should ideally be based on evidence sourced from timely and informative datasets. Unfortunately, it appears to have been a case of putting the cart before the horse where policy makers make decisions based on sentiments and subsequently seek information to back up their policies.

Even in situations where a decision maker is transparent and committed to doing the right thing, the absence of the required raw materials and evidence base often weakens the decision-making process.

For instance, many public health authorities in developing countries only exist as physical structures. The lack of adequate data and information systems especially at detailed spatial granularity means that they are unable to adequately monitor health care and provide intervening strategies (Gething et al., 2006).

### 2.6.2 Access to Available Datasets

Another major challenge posed to geodemographics development in third world countries is the problem of access to available datasets. In some countries where data is available, they are often hoarded by the same sources from which they should be derived.

One of the contributory factors to the difficulty in accessing spatial statistics is the fact that many of these countries dwell on outdated legal frameworks for the release of information (See and Gibson, 2006). In some countries, the dissemination of digital or electronic data is frowned at. Often government agencies hide under the cover of disclosure controls.

In some other cases, the young establishment of democratic systems of governance does not encourage the freedom and access to information. For instance Nigeria has presently no legislation which assures its nationals access to public records and information. Since the presentation of the Freedom of Information Bill to the lower and upper houses of the national assembly in 1999 (FRN, 1999), it is yet to be signed into law in spite of the fact that the original version has been watered down significantly. This continued secrecy has greatly hampered the ease with which researchers can gain access to required information for unbiased analysis.

It is important for public office holders to recognise that publicly sourced information like censuses and surveys are a public good because members of the public are also stake holders in the creation of these datasets. Additionally making information available for the public good signifies that the government is transparent and ready to be accountable and responsible for its actions and policy decisions. It also allows an encouraging platform for constructive debate and discussions amongst academics and the public stakeholders.

### 2.6.3 Technical Expertise

Although not often admitted by developing country statistical bodies, many of them lack the technical expertise required for the creation, storage, manipulation and management of geographically referenced statistics (See and Gibson, 2006). Often, support is received from international donor agencies to provide training to members of staff. However what often results from this training is a brain drain to private sector companies due to in part to better working conditions and increased pay.

The problem of data accessibility is also linked to shortage in technical expertise. Many government agencies claim that they do not want to disclose confidential information. However the key problem is that they may lack the skills and techniques used for anonymising data.

Apart from the problem of keeping with the pace geostatistical methods, there is also the need for the expansion of training on how to use specialised software packages. Although there is evidence of a slowly narrowing digital divide (Weeks and Grant, 2003) the evidence of this problem is still persistent in developing countries. As such it is very common to find large sets of data infrastructure only available in analogue formats hence requiring significant time and cost resources for conversion.

## 2.7 Conclusions

The aggregation of entities into sets which are meaningful and simple in terms of their general attributes is an important feature of social scientific investigation. Area classifications therefore did not originate without key socio-economic implications. It is a field that emerged from conscious and unconscious efforts to understand the patterns and processes within geographical and social space with a view to solving problems.

Since the days of Charles Booth, and the emergence of modern day geodemographics, the sole aims of these classifications has been to create relatively homogeneous groups called clusters. Geodemographic classifications are therefore expected to provide simplicity within a complex system.

The fact that geodemographic typologies have proved very useful as important discriminators has seen the field exploited by numerous commercial organizations who now dominate the arena. These commercial classification systems have contributed to the growth and explosion of the industry and have also played significant roles in growing criticism of geodemographics.

The existing challenges in the field are both methodological and ethical. Methodological challenges can be ameliorated via the involvement of the academic community in research and encouraging the use of systems that make their methods publicly available and have had them rigorously verified. It will be difficult to draw commercial vendors into this due to competition therefore such classifications can not continue to be viewed and used as general purpose systems. This will be misleading.

In addition to an urgent need for geodemographic techniques to expand into developing economies like Nigeria and the Philippines, the future of geodemographic classifications lies in the ability to make them more sophisticated. Public and private databases form the potent source of elements used to build the classifications. The attributes of the people and neighbourhoods from which these databases have been built are not static therefore future systems need to *remain simple* yet become more dynamic and adaptive to change (Birkin and Clarke, 2009).

# 3

# The Millennium Development Goals and Their Nexus with Geodemography

## 3.1 Introduction

This chapter explores some of the benefits that can be gained in developing countries from using geodemographic techniques. The challenges of developing world countries are multi-faceted. Section 3.2 presents a review some of these challenges by introducing the Millennium Development Goals. In section 3.3 some shortcomings of the goals are reviewed while section 3.4 considers the potential of geodemographic tools for enhancing growth and development. Conclusive comments are provided in section 3.5.

## 3.2 Introducing the Millennium Development Goals

During the Second World War, world leaders came to an honest realisation that no country can be an island unto itself. They recognised the power of unity in confronting the challenge posed by the ongoing conflict and for the first time the sitting President of the United States coined the term 'United Nations' in January 1942 (UN, 2008a). It took another three years for the organisation to be formed officially following ratification of an agreed charter by five of the fifty-one original member countries. As such, the United Nations (UN) formally came into existence on 24$^{th}$ October 1945. To date, membership of the organisation has increased to 192 member states.

As set out in the charter which serves as a building block of the UN constitution (UN, 1945), the rationales of the organisation are:

1. To maintain international peace and security, and to that end: to take effective collective measures for the prevention and removal of threats to the peace, and for the suppression of acts of aggression or other breaches of the peace, and to bring about by peaceful means, and in conformity with the principles of justice and international law, adjustment or settlement of international disputes or situations which might lead to a breach of the peace;

2. To develop friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples, and to take other appropriate measures to strengthen universal peace;

3. To achieve international co-operation in solving international problems of an economic, social, cultural, or humanitarian character, and in promoting and encouraging respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion; and

4. To be a centre for harmonizing the actions of nations in the attainment of these common ends.

It is evident that major events of conflict and a long period of cold-war between some of the world's super powers dominate the character and content of the ideals of the organisation. It took another 55 years for the organisation to refocus its energy on some of the more endemic problems facing much of the world's population.

It is difficult to define or identify the universal basic needs as needs vary from person to person and country to country. To some priority needs relate to the enhancement of human capital while some others perceive the redistribution of welfare services and infrastructure as essential towards shaping quality of life.

The Human Development Index (HDI) combines measures of literacy, life expectancy, educational attainment, and Gross Domestic Product (GDP) per capita. The index provides a means for comparing and understanding the complex relationship between income and well-being and how it creates an array of choices and opportunities for human beings (UNDP, 1990). If the HDI is used as a proxy for assessing human capital; then the need to live a long healthy life; the requirement for education and the requirement for a decent standard of living may be viewed as essentials of life especially in third world countries. Other factors which have frequently influenced these important issues include crime and corruption, conflicts and indeed cooperation within the 'global village'.

Population growth is also an important challenge for many developing countries. The population of the world has exceeded the 6 billion mark (Bongaarts and Bulatao, 2000) even though there are predictions that there will be a decline in the growth rate of the world population in the second half of this century (Wolfgang et al., 2004). Most of the world's population currently reside in developing world countries (UN, 2009b) with Asia alone accounting for about 60% of the global total. Uncontrolled population growth tends to hamper the attempts of most governments to ensure stable economic growth. For instance increased population will require the labour market absorbing more people which may pose challenges to the rate at which jobs are created and the sustenance of public infrastructure.

At the wake of the current millennium, world leaders gathered at the headquarters of the UN to demonstrate their recognition and vision for a better world (UN, 2000). Much of the deliberations identified the need to reduce the scourge of poverty and hunger, improve education, health and opportunities for women, ameliorate environmental degradation and increase global mutual cooperation. These are important issues which culminate into the 8 Millennium Development Goals (MDG) listed below:

**Goal 1**: Eradicate extreme hunger and poverty
**Goal 2**: Achieve universal primary education
**Goal 3**: Promote gender equality and empower women
**Goal 4**: Reduce child mortality
**Goal 5**: Improve maternal health
**Goal 6**: Combat HIV/AIDS, malaria and other diseases
**Goal 7**: Ensure environmental sustainability
**Goal 8**: Develop a global partnership for development

While some of these challenges are common to both developing and developed countries, it can not be over-emphasised that less developed countries are most burdened with the depth and spread of the problems.

Due to the broad nature of the MDGs, a set of 18 related targets were identified to help facilitate the process of quantifying progress towards attaining the goals. The targets have since increased to 21. Additionally, each target can be assessed via a number of indicators which have increased from 48 to 60 (UN, 2003).

All ensuing analysis and reporting by the UN and complementary international agencies are often predicated on this framework of goals, targets and indicators. However in some countries, some non-standard indicators are used to prepare country reports. The following 9 sub-sections are organised by each MDG with discussions on how they are measured.

### 3.2.1 Eradicating Extreme Hunger and Poverty

Poverty is a multifaceted problem. Early attempts at describing this human condition focused mainly on income alone. However, the complex and dynamic nature of the phenomenon has since seen its definitions embrace concepts like insecurity, incapacitation and inequality (Grusky et al., 2006). The multidimensional purview of poverty suggests that attempts at reducing the scourge should extend beyond income poverty alone.

Poverty can appear in absolute and relative forms. Absolute poverty is a situation where people live below a minimum socially acceptable standard of living while relative poverty defines poverty in terms of a comparison of the relationship between the standards of living of different societies (Spicker, 1993; Dorling and Ballas, 2008). In 2005, it was estimated that about 1.4 billion people in third world countries live in extreme poverty. The recent food crises and subsequent increases in the price of food are to increase the world's poor by another 100 million people (UN, 2008b). These are figures which require a re-definition of some of the approaches used in identifying and reaching the poor.

**Table 3.1: Targets and Indicators for Goal 1**

| Targets | Indicators |
|---|---|
| **Target 1.A**: Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day | 1.1 Proportion of population below $1 (PPP) per day |
| | 1.2 Poverty gap ratio |
| | 1.3 Share of poorest quintile in national consumption |
| **Target 1.B**: Achieve full and productive employment and decent work for all, including women and young people | 1.4 Growth rate of GDP per person employed |
| | 1.5 Employment-to-population ratio |
| | 1.6 Proportion of employed people living below $1 (PPP) per day |
| | 1.7 Proportion of own-account and contributing family workers in total employment |
| **Target 1.C**: Halve, between 1990 and 2015, the proportion of people who suffer from hunger | 1.8 Prevalence of underweight children under-five years of age |
| | 1.9 Proportion of population below minimum level of dietary energy consumption |

Source: United Nations (2008c)

The goal of eradicating extreme hunger and poverty is ambitious. To this end, three targets listed in Table 3.1 have been adopted by the UN. Progress towards meeting each target is further understood by analysing the indicators listed in the table.

By 2015, it is expected that the number of people living in absolute conditions of poverty (less than $1 a day) would be halved based on a strategy encapsulated in the International Development Targets (IDT) of 1996 (IMF, OEDC, UN and World Bank, 2000). The 2008 progress report on the MDGs suggests that this feat is still within reach though requiring extra efforts especially in Asia (UN, 2008b). Current trends show that developing world countries and in particular low income economies are lagging behind and are unlikely to meet the set target.

Due to the uneven distribution of levels of poverty across countries and world regions (Chen and Ravallion, 2007), each country estimates absolute poverty by converting the 1$ per day figure to its local currency. This indicator is globally representative of the purchasing power parity (PPP) of each individual on a daily basis. Income and consumption measures are often used to compute the indicator (World Bank, 2008). To estimate average individual income, the consumption or income at household level is divided by the count of household residents. The indicator is computed only by the World Bank Development Research Group based on data sourced from national statistical departments and country offices of the Bank.

The poverty gap ratio measures the average distance separating the population from the assumed line of poverty. This distance is then expressed as a percentage of the poverty line. The indicator estimates what is called the poverty deficit. This deficit is a quantification of the resources required to translate people below the poverty line above it (Chen and Ravallion, 2002).

One challenge with computing the indicator however is the fact that it is dependent on the definition of the poverty line. The World Bank calculates the poverty gap ratio based on the agreed standard of $1 while the different national statistical offices compute the indicator based on their national poverty lines.

When national populations are segmented into 5 groups based on their income, the income that accrues to poorest group (which belong to the fifth quintile) is also assessed in a bid to meet target 1A. Information on income is usually very difficult to come by especially in developing countries hence the information that is often used is that of consumption rather than income. Another issue with this indicator is that while it provides information on the distribution of the poorest fifth of the national population, it is not necessarily indicative of the underlying characteristics of these people and the propensities of their spatial locations.

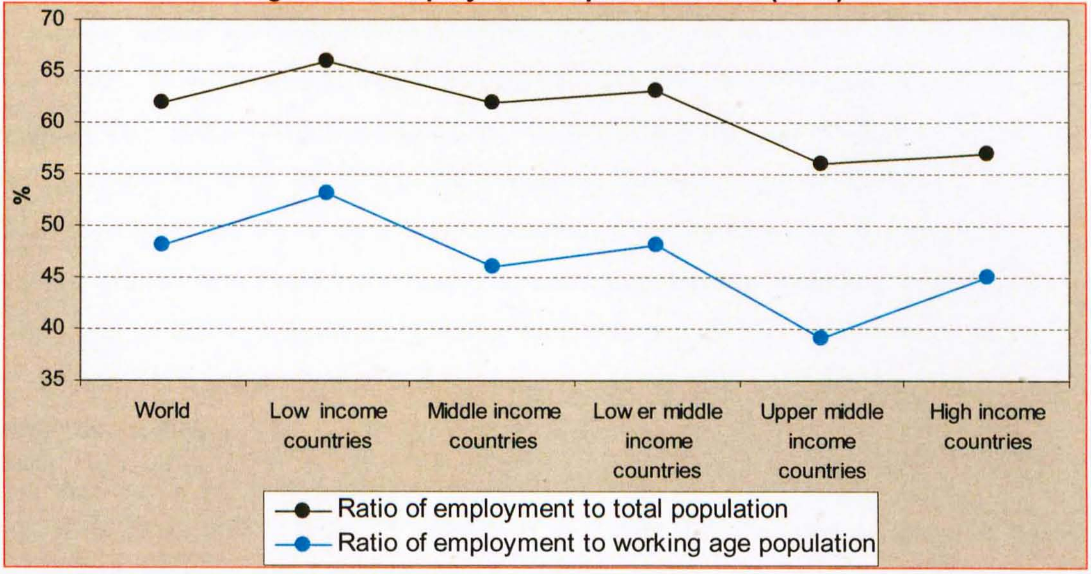
There is also an issue with the focus on the bottom fifth of the population. Why focus on this group alone? It may be more interesting to also understand the characteristic features of the other percentiles (i.e. the first to fourth quintile). Analysis of these groups of individuals may reveal previously unknown characteristics of these population groups that keep them above the fifth and poorest quintile.

Increased unemployment has also contributed significantly to poverty particularly in many third world countries (Pratap and Quintin, 2006). Employment (formal or informal) is traditionally the source of income for any human being. The second millennium development target (1B) bolsters this fact. It is interested in increasing productive employment especially amongst women and young people.

To meet this target, the first indicator assessed examines the productivity of labour. This is done by computing the ratio of output of employed persons or their Gross Domestic Product (GDP) to the number of employed persons (World Bank, 2008).

A statistic that has often been used to assess the ability of an economy to create jobs for its citizens is the ratio of its employed population to the employable or working age population. This statistic is also used as an indicator required for the second target.

**Figure 3.1: Employment Population Ratio (2007)**




Source: World Bank

Traditionally, the higher the value of this ratio, the better a country or region is (ILO, 2004). The trend shown in Figure 3.1 appears to put low income countries in an advantaged position. However, such interpretation may be misguided and block the path to understanding the truth. The reason for higher values in low income economies is essentially because of the nature of most of the jobs. Most jobs (about 60%) are concentrated in the informal sector. Such jobs often require more people and pay much less than jobs in the formal sector (Pratap and Quintin, 2006). Again, this is not unconnected with the problem of productivity. Productivity is higher in developed countries where they have embraced technology.

There is also an indicator that is often used by the UN monitoring agencies to relate poverty to employment. This is the proportion of people who are employed and living beneath the poverty line. This indicator is simply computed as a ratio of those employed and living beneath the poverty line to the number of employed people (World Bank, 2008). The fact that this indicator exists is an important reminder that by itself, employment alone does not transcend people from poverty to a better situation. The quality of employment matters to a significant extent. For instance an assessment of people who are underemployed could serve as a useful proxy for assessing this issue.

The fourth indicator required for meeting target 1B examines the susceptibility of those in different forms of employment. There is a particular focus on people who are self employed. These are groups who are often thought to be in local or global competition with larger firms. People in self employment jobs are in vulnerable situations largely because their remuneration is dependent on the profits made from the goods produced and services provided. In third world countries, these groups of people are important because they often make up small to medium scale enterprises. It has been argued that the growth and development of these enterprises are important to the economic and industrial base of developing countries (Hallberg, 2000). As a result microfinance institutions have arisen overtime to help sustain some of the efforts of people employed within this domain.

The third target (1C) which is aimed at reducing people who suffer from hunger concentrates on nutrition and dietary concerns. Child malnutrition is interlinked with health, poverty and sometimes education (Smith et al., 2000). Data used to compute this indicator is often derived from the Multiple Indicators Cluster Surveys (MICS) and national Demographic and Health Surveys (DHS). The agencies responsible for monitoring the indicator are the ministries of health of different countries, the United Nations Children Fund and the World Health Organisation.

The final indicator that is assessed within the framework of Goal1 and target 1C is the degree of undernourishment which is measured as the percentage of the population that is undernourished. Undernourishment is subjective in definition and can vary from country to country. Another problem with this indicator is that it is estimated by the Food and Agriculture Organisation (FAO) only at national level which are subsequently aggregated up to generate regional and global estimates. This makes it difficult to embark on an analysis of within country local level inequalities of this indicator.

### 3.2.2 Achieving Universal Primary Education

Education plays a pivotal role in the development of a society. The processes and stages involved in educating a human being ensure that he or she can align their natural aptitudes and potentials to derive optimum gains. It is therefore fundamental for the human mind to think aright. Education helps train the human mind to a point that it can that it can think clearly in the right direction and make confident decisions.

As far back as 1597, an English Philosopher and states man – Sir Francis Bacon coined the Latin phrase '***scientia potentia est***' which when paraphrased in English means '***Knowledge is power***' (Bacon, 2008). Over the centuries, this phrase has inspired many and helped unearth their true underlying potentials. An educated human being has confidence to make decisions for which they can assume responsibilities. Education acts as a source of light that lightens the path of men women and indeed children.

While it is desirable for children everywhere to receive education, it is recognized globally that at least a primary level education can help people write their own destinies. As a result, the second MDG; the challenge of ensuring that children everywhere have access to basic education should be considered fundamental to the completeness of their growth and development. It is a fundamental human right which is explicitly stated in the United Nations Universal Declaration of Human Rights (UDHR, 1948). Unfortunately this right has been seen especially in many developing world countries as a mere privilege.

The most recent UN MDG report notes that while slow progress is being made towards meeting the goal of the millennium declaration, about 72 million school aged children were denied the right to education in 2007 (UN, 2009a). According to the report, fifty percent of these kids have never stepped into a school class room. It is important to realize that since many of the MDGs are interlinked, there is a great chance that if the education related target is not met, it may have negative or undesirable consequences on other goals.

**Table 3.2: Targets and Indicators for Goal 2**

| Targets | Indicators |
|---|---|
| **Target 2.A**: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling | 2.1 Net enrolment ratio in primary education |
| | 2.2 Proportion of pupils starting grade 1 who reach last grade of primary |
| | 2.3 Literacy rate of 15-24 year-olds, women and men |

Source: United Nations (2008c)

Like the first goal, there is also a target and a set of three indicators for the task of achieving universal primary education. These are listed in Table 3.2 above.

To ensure that boys and girls everywhere are able to complete a full course of primary education, the first indicator that is measured and assessed is the net enrolment ratio in primary education. Net primary enrolment ratio is simply the ratio of the number of school aged children who are enrolled in primary school to the total population of school aged children (World Bank, 2008). These population groups are officially defined by the national education boards of each country because of differences in the definition of school age. Data on the indicator is therefore expected to be harnessed by the country ministry of education or derived from censuses and surveys. The ministries of education are also supported by the United Nations Educational Scientific and Cultural Organisation (UNESCO).

Beyond data on enrolment, data on primary school completion rates are also important. As a result the second indicator within the framework of MDG 2 aims at harnessing information on the survival rate of children who start primary level grade 1 and are expected to reach grade 5 (UN, 2003). This indicator was proposed by UNESCO statistics division and is principally derived from the MICS and DHS. It is important to monitor completion rates because not every child that enrols gets to the end of the primary school journey. However, one issue that is more important and less stressed is the co-related factors to high incompletion rates. Knowing these factors and nipping them in the bud will eventually increase completion rates.

The final indicator that is assessed within the education framework relates to the levels of literacy of persons aged between 15 and 24 years. These groups of people are very important because they make up the youths and indeed the future leaders of any country. The sustenance and continuous development of any country and its ability to remain globally competitive lies in the hands of those who are educated amongst this age group (UN, 2003). This indicator also serves the purpose of assessing how well the basic idea of universal primary education is entrenched into societies. The indicator is computed by dividing the number of educated persons within this age group by the total population aged 15 to 24.

### 3.2.3 Promoting Gender Equality and Empowering Women

It is pertinent to note that gender parity is very important for the attainment of any of the MDGs. Inequality in the opportunities that accrue to both men and women do not do any good, rather they skew the overall development of a country. For instance, countries with some of the largest representations of marginalised women also have high incidences of maternal mortality (WHO, 2005).

Female education is an important tool for economic development as women constitute the major backbone of the informal economy (World Bank, 2005). Recognising this importance, the target identified for meeting the third MDG seeks to eliminate gender inequality at all levels of education. The target of 2005 for the elimination of gender disparity at primary and secondary level education was missed requiring a re-aligning of the focus on 2015.

**Table 3.3: Targets and Indicators for Goal 3**

| Targets | Indicators |
|---|---|
| **Target 3.A**: Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015 | 3.1 Ratios of girls to boys in primary, secondary and tertiary education |
| | 3.2 Share of women in wage employment in the non-agricultural sector |
| | 3.3 Proportion of seats held by women in national parliament |

Source: United Nations (2008c)

The first indicator within the framework is the ratio of boys to girls in primary, secondary and tertiary education which is indicative of equality in educational opportunities. The ratio is calculated for all boys and girls regardless of their ages (UN, 2003). Data used for the computation of the indicator is usually sourced from country ministries of education or from censuses and surveys like MICS, DHS, Living Standard Surveys and the Core Welfare Indicators Questionnaire Surveys (CWIQ).

The second indicator in Table 3.3 is computed by expressing the share of women engaged in non-agricultural employment as a percentage of the total employment in the non-agricultural sector (UN, 2003). This information is important and indicative of the extent to which professional, industrial and services employment sector is open to female engagement in paid employment. The larger the proportion of women in non-agricultural employment, the better the opportunities, security and dignity women and their children are exposed to.

The promotion of equality in gender is also linked to monitoring the engagement of women in politics. This is vital because important decisions which have significant policy implications on the lives of women and children are made in the political spheres. In many scenarios, no matter the effort and commitment deployed by individuals or members of the public towards making developmental progress, the absence of political will can make such efforts futile. It is therefore very important for women to engage in politics and take positions especially in offices that deal directly with women and children affairs.

### 3.2.4 Reducing Child Mortality

Even though it appears fewer children under the age of five years are dying than ever before (UN, 2009b), the problem is still pronounced in the developing regions of the world. When the survival chance of a child is born in a third world country is compared to another born in a first world country, the child born in the third world country is 13 times more likely to die before their 5$^{th}$ birthday (UNDP, 2007).

The target set by the UN is to reduce the mortality rates of children less than five years old by two-thirds by 2015. The recent MDG report notes that little or no progress has been made in sub-Saharan Africa towards this target and recounts that in 2007 alone, one out of seven children died before age five (UN, 2009b). This is disheartening but more disturbing is the fact that to date, it is almost impossible to find a road map monitoring the trends of the scourge of child mortality at local levels in many of these third world countries.

**Table 3.4: Targets and Indicators for Goal 4**

| Target | Indicators |
|---|---|
| **Target 4.A**: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate | 4.1 Under-five mortality rate |
| | 4.2 Infant mortality rate |
| | 4.3 Proportion of 1 year-old children immunized against measles |

Source: United Nations (2008c)

In Table 3.4, the first indicator assessed is the under-five mortality rate. This rate represents the probability of a child born in a specific year dying before the age of five (UNICEF, 2000) and is expressed relative to 1000 live births. The motivation for the indicator is the need to assess and evaluate child survival rates especially in developing countries.

Ideally, data for this indicator should be sourced from the national statistics registration agencies covering up to 90% of the registration events. However due to the weakness of data gathering institutions in developing countries, direct and indirect estimates are often derived from census and survey datasets.

The second indicator in the framework for Goal 4 is the measurement of infant mortality. The UN definition of an infant is any child less than one year old (UNICEF, 2008). These population groups are highly vulnerable to diseases especially since their immune systems at that age would still be developing.

Infant mortality rate is quantified in a similar manner to under-five mortality and measurement of the indicator in third world countries is confronted by the same challenges earlier identified.

The third indicator in Table 3.4 seeks to measure the extent to which infants receive at least one dose of measles vaccine. Although UNICEF recorded that between 2000 and 2006 Africa reported a 91% drop in measles related child deaths (UNICEF, 2008), the disease is still a highly contagious viral disease that remains a leading cause of deaths among children. In 2007, it claimed the lives of 22 children every hour (WHO, 2008a).

The indicator is computed by relating the proportion of infants who received a dose of measles vaccine to children aged between 12 and 23 months. The WHO and UNICEF are responsible for gathering and compiling country level data for monitoring the indicator.

### 3.2.5 Improving Maternal Health

The importance of the economic contribution of women to the growth and development of many societies was mentioned briefly in section 3.2.3. Boserup et al. (2007) provide a detailed review of the role which women play in economic development tracing their contributions and challenges in rural and urban settings. When women are provided with adequate education the scourge of poverty and related challenges can be greatly ameliorated (Boserup, 2007). Not only does female education contribute to the reduction of poverty, but also and more significantly, women are able to take appropriate and timely measures for their survival and that of their children.

Maternal mortality is significantly linked with child delivery. Unfortunately, in many countries of the developing world where population is increasing at high rates (UN, 2009b), safe child delivery is often considered to be a privilege of those who are of great economic advantage (UN, 2009a).

**Table 3.5: Targets and Indicators for Goal 5**

| Targets | Indicators |
|---|---|
| **Target 5.A**: Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio | 5.1 Maternal mortality ratio |
| | 5.2 Proportion of births attended by skilled health personnel |
| **Target 5.B**: Achieve, by 2015, universal access to reproductive health | 5.3 Contraceptive prevalence rate |
| | 5.4 Adolescent birth rate |
| | 5.5 Antenatal care coverage (at least one visit and at least four visits) |
| | 5.6 Unmet need for family planning |

Source: United Nations (2008c)

Maternal mortality ratio which is the first indicator employed in a bid to reduce maternal mortality is the number of women who die from pregnancy related causes or the poor management of pregnancy (UN, 2003). This also includes women who die within 42 days as a result of the termination of a pregnancy. The ratio is computed for every 100, 000 live births.

Ideally, the computation of this indicator should be based on recorded maternal deaths and live births; however estimates are also used especially in developing countries where vital statistical registration systems do not operate effectively. Such estimates which are usually computed for higher level geographies are drawn from household surveys like the DHS and hospital studies.

While it is desirable to accurately estimate maternal mortality, it is very difficult due largely to inadequacies in registration data. As a result other proxy indicators are also assessed. The proportion of births attended to by skilled health personnel for instance is a pointer towards access to maternal health care and a proxy for assessing the exposure to health-related complications among women. The proportion of births attended to by skilled personnel is quantified as the percentage of deliveries in which trained health personnel provided the required supervision (UN, 2003). Skilled health personnel do not include Traditional Birth Attendants (TBA) even if they have received some form of training.

The second target within the maternal mortality frame work seeks to achieve universal access to reproductive health by 2015. To achieve this feat, four indicators (5.3 to 5.6) listed in Table 3.5 are monitored.

Contraceptive prevalence rate is measured for either women or their partners. It is quantified as the proportion that uses a contraceptive method at a point in time (WHO, 2009a). The key data sources for this indicator are the DHS and the MICS.

Adolescents are a very important demographic group women within this group (usually aged 15 to 19) make up about 11% of global births (WHO, 2008b). It is also on record that 95% of the births by these young girls take place in low and middle income countries. Because of the ages of girls in this category, they are vulnerable to health related complications. Not only are young mothers exposed to these problems, but so are their offspring (WHO, 2008).

The importance of antenatal care can not be overemphasized. The indicator measures the access and use of health care by women during pregnancy. So what really is antenatal care? The explanation given by the WHO (2009b) is that the this sort of health care constitutes

*'screening for health and socioeconomic conditions likely to increase the possibility of specific adverse pregnancy outcomes, providing therapeutic interventions known to be effective; and educating pregnant women about planning for safe birth, emergencies during pregnancy and how to deal with them'* (WHO, 2009b, p. 1).

Not many of these conditions are met in third world countries due in part to poor health care delivery systems and in some cases religious and cultural barriers (WHO, 2005).

It is not every woman that wants to get pregnant at every point in time. This is one of the key factors underpinning the measurement of unmet needs of women for family planning. In most cases and ideally, women should engage in family planning measures to prevent unwanted pregnancies. However a range of obstacles militate against women especially in third world countries. These obstacles have often been linked with the degree of access to health care (Ross and Winfrey, 2002).

### 3.2.6 Combating HIV/AIDS, Malaria and Other Diseases

The fact that three out of the eight MDGs are directly related to health underscores the importance of good health to development. The sixth goal is concerned with ameliorating the scourge of various infections and diseases notably, HIV/AIDS, malaria and tuberculosis. These are top killer diseases particularly associated with third world countries (WHO, 2009c).

**Table 3.6: Targets and Indicators for Goal 6**

| Targets | Indicators |
|---|---|
| **Target 6.A**: Have halted by 2015 and begun to reverse the spread of HIV/AIDS | 6.1 HIV prevalence among population aged 15-24 years |
| | 6.2 Condom use rate of the contraceptive prevalence rate |
| | 6.3 Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS |
| | 6.4 Ratio of school attendance of orphans to school attendance of non-orphans aged 10-14 years |
| **Target 6.B**: Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it | 6.5 Proportion of population with advanced HIV infection with access to antiretroviral drugs |
| **Target 6.C**: Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases | 6.6 Incidence and death rates associated with malaria |
| | 6.7 Proportion of children under 5 sleeping under insecticide-treated bed nets |
| | 6.8 Proportion of children under 5 with fever who are treated with appropriate anti-malarial drugs |
| | 6.9 Incidence, prevalence and death rates associated with tuberculosis |
| | 6.10 Proportion of tuberculosis cases detected and cured under directly observed treatment short course |

Source: United Nations (2008c)

The first target listed in Table 3.6 aims to halt the spread of HIV/AIDS by 2015. There is no doubt this is a challenging task especially within Sub-Saharan Africa where evidence from the recent UN report suggests that about 66% of people living with HIV are concentrated (UN, 2009a).

The first indicator refers to the percentage of pregnant women within the age bracket of 15 to 24 years who test positive to HIV (UN, 2003). The WHO and the Joint United Nations Programme on HIV/AIDS are charged with the responsibility of sourcing this indicator. An interesting feature of the indicator is that only anonymously sought blood samples are used. This is because providing anonymity helps increase participation in the testing and data collection process.

Condom use rate assesses the rate of use of contraception in women aged between 15 and 49 years. The indicator is computed by relating the proportion of women who use condoms or whose partners use condoms to all other women of the same age bracket identified. The women are ideally married or in a consensual union (UN, 2003).

The third indicator is the percentage of people aged 15 to 24 who have a detailed knowledge of the two principal ways in which HIV/AIDS can be prevented. These are those who know that using condoms and the limitation of sexual activity to one faithful and uninfected partner can help reduce the chances of an individual contracting the disease. Additionally, these persons are expected to know that it is possible for a healthy looking person to be living with the disease. This indicator assumes that this knowledge is probably the first major step in the fight against the scourge of the virus.

The impact of the HIV/AIDS epidemic on the lives of children is also assessed within the framework of Goal 6. Impact in this sense is computed by relating the proportion of children (under age 15 years) who have lost their mother, father or both parents to AIDS and are in school to the proportion of non orphans who are also in school (UN, 2003).

It is also hoped that by 2015, all persons living with the virus would have access to treatment. In light of this target, it is therefore imperative to continuously take stoke of the proportion of people infected by AIDS who have access to antiretroviral drugs. Unfortunately, within the framework of the UN, only people at an advanced stage of the disease are considered when calculating this indicator. This can prove a little unhelpful first because there is no clear definition for the word advanced and secondly, experience has shown that if treatment is commenced earlier, then the life of the individual involved may be prolonged (Kinloch, 1994).

The third target within the framework of Goal 6 focuses on other diseases like malaria and tuberculosis. The incidence of malaria refers to the frequency of development of malaria within the population year after year while death rates are computed as the number of deaths caused by the disease for every 100, 000 people (UN, 2003).

In addition to the incidence and death rates resulting from malaria, the use of insecticide treated nets amongst children under 5 years and the quality of treatment they receive are also assessed alongside data for appropriate anti- malarial treatment. An appropriate treatment in this context refers to those who have been treated with anti-malarial drugs.

Just like malaria, the incidence, prevalence (per 100, 000 people) and deaths linked with tuberculosis is also measured. The agencies responsible for harnessing these datasets are the ministries of health of the countries in collaboration with the WHO. On many occasions due to the absence of direct measures of prevalence, surveys like MICS and the DHS are utilised.

For tuberculosis, detection rates are also assessed. The detection rate is the percentage of estimated new infectious cases detected under the Directly Observed Treatment Short-course strategy (DOTS) (UN, 2003). DOTS is composed of five key components which include: Political commitment with increased and sustained financing; case detection through quality-assured bacteriology; standardized treatment, with supervision and patient support; an effective drug supply and management system; and monitoring and evaluation system, and impact measurement. In addition to the detection rates the rates of registered cases cured under DOTS are also used to monitor the disease and the target of halting its spread by 2015.

### 3.2.7 Ensuring Environmental Sustainability

The fundamentals of sustainable development are well documented in the Brundtland report of 1987. The report titled '*Our Common Future*' re-ignited the importance of discussions within the economic development-environment sustainability nexus. World leaders and policy makers were reminded that while it is desirable to make socio-economic progress, such progress must be anchored on the ideal that natural or environmental resources which support this development should not be depleted (WCED, 1987).

Within the context of sustainable development, three principal issues underpin the content of the report. First is the issue of environmental protection which is concerned with the preservation of ecological resources and encouraged the development of sustainable technologies. There is also the issue of encouraging economic growth which includes meeting food, energy and employment needs. This component is a bid to pacify the economists who often engage environmentalists in sometimes bitter arguments. The third element of the report is concerned with maintaining social equity and encouraging and ensuring that third world countries progress towards becoming developed nations themselves. These ideas are central to the seventh MDG as shown in Table 3.7.

**Table 3.7: Targets and Indicators for Goal 7**

| Targets | Indicators |
|---|---|
| **Target 7.A**: Integrate the principles of sustainable development into country policies and programmes and reverse the loss of environmental resources | 7.1 Proportion of land area covered by forest |
| | 7.2 CO2 emissions, total, per capita and per $1 GDP (PPP) |
| | 7.3 Consumption of ozone-depleting substances |
| | 7.4 Proportion of fish stocks within safe biological limits |
| | 7.5 Proportion of total water resources used |
| **Target 7.B**: Reduce biodiversity loss, achieving, by 2010, a significant reduction in the rate of loss | 7.6 Proportion of terrestrial and marine areas protected |
| | 7.7 Proportion of species threatened with extinction |
| **Target 7.C**: Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation | 7.8 Proportion of population using an improved drinking water source |
| | 7.9 Proportion of population using an improved sanitation facility |
| **Target 7.D**: By 2020, to have achieved a significant improvement in the lives of at least 100 million slum dwellers | 7.10 Proportion of urban population living in slums |

Source: United Nations (2008c)

The proportion of land area covered by forest relates forest areas to total land areas on a country by country basis and is sourced by the Food and Agriculture Organization (FAO). The second and third indicators which are ideally sourced from country energy ministries relate to carbon dioxide emissions and ozone layer depletion. A major limitation of these indicators is that many third world countries do not have adequate capacity for measuring them. Additionally, it has been identified that rather than embarking on inter-country comparisons, these indicators are better assessed globally as time-series data (UN, 2003).

The use of water resources and marine related stocks are also assessed within the framework of goal 7. The proportion of fish stocks within safe biological limits is monitored and reported by the FAO. This indicator is important in order to be able to address threats to the aquatic ecosystem but again data gathering mechanisms poses a critical challenge. The indicator on use of water resources is also

The two indicators related to target 7b help address threats to biodiversity loss. They transcend the aquatic domain to include the agricultural ecosystem. It is a long established fact that these systems which often constitute our ecological foot prints are important for sustaining our livelihoods (Wackernagel and Rees, 1996).

The final three indicators within the MDG 7 framework which deal with access to improved water supply, improved sanitation and the population living in slums are as much health and poverty related as they are environmental. Improved water resources can be subjective from country to country. However within the context of the MDG, these sources include piped water, public tap, borehole or pump, protected well, protected spring or rainwater (UN, 2003). Water sourced from vendors, truck and tanker deliveries, bottled water or unprotected wells and springs are classed as unhealthy sources. Improved sanitation facilities on the other hand include sewers or septic tanks, poor-flush latrines and simple pit or ventilated improved pit latrines which are not for public use.

### 3.2.8 Developing a Global Partnership for Development

The final MDG which seeks to strengthen inter-country partnerships and ties emphasizes more of the responsibilities that richer countries have towards their poorer counterparts. While it may be argued that the world realised the global mutual co-operation since former President of the United States Franklin Roosevelt coined the term United Nations in January, 1942 (UN, 2008a), much of the co-operation between rich and poor countries has been less on the economic front. A reason for this is that the events which initially led to partnerships between nations were militarily inclined. It has even been argued that in practice, the ideals of some global financial institutions like the IMF may not necessarily have implications which translate into positive outcomes on developing countries (Prasad et al., 2003).

**Table 3.8: Targets and Indicators for Goal 8**

| Targets | Indicators |
|---|---|
| **Target 8.A**: Develop further an open, rule-based, predictable, non-discriminatory trading and financial system Includes a commitment to good governance, development and poverty reduction - both nationally and internationally<br><br>**Target 8.B**: Address the special needs of the least developed countries Includes: tariff and quota free access for the least developed countries' exports; enhanced programme of debt relief for heavily indebted poor countries (HIPC) and cancellation of official bilateral debt; and more generous ODA for countries committed to poverty reduction<br><br>**Target 8.C**: Address the special needs of landlocked developing countries and small island developing States (through the Programme of Action for the Sustainable Development of Small Island Developing States and the outcome of the twenty-second special session of the General Assembly)<br><br>**Target 8.D**: Deal comprehensively with the debt problems of developing countries through national and international measures in order to make debt sustainable in the long term | Some of the indicators listed below are monitored separately for the least developed countries (LDCs), Africa, landlocked developing countries and small island developing States.<br><br>Official development assistance (ODA)<br>8.1 Net ODA, total and to the least developed countries, as percentage of OECD/DAC donors' gross national income<br>8.2 Proportion of total bilateral, sector-allocable ODA of OECD/DAC donors to basic social services (basic education, primary health care, nutrition, safe water and sanitation)<br>8.3 Proportion of bilateral official development assistance of OECD/DAC donors that is untied<br>8.4 ODA received in landlocked developing countries as a proportion of their gross national incomes<br>8.5 ODA received in small island developing States as a proportion of their gross national incomes<br><br>Market access<br>8.6 Proportion of total developed country imports (by value and excluding arms) from developing countries and least developed countries, admitted free of duty<br>8.7 Average tariffs imposed by developed countries on agricultural products and textiles and clothing from developing countries<br>8.8 Agricultural support estimate for OECD countries as a percentage of their gross domestic product<br>8.9 Proportion of ODA provided to help build trade capacity<br><br>Debt sustainability<br>8.10 Total number of countries that have reached their HIPC decision points and number that have reached their HIPC completion points (cumulative)<br>8.11 Debt relief committed under HIPC and MDRI Initiatives<br>8.12 Debt service as a percentage of exports of goods and services |
| **Target 8.E**: In cooperation with pharmaceutical companies, provide access to affordable essential drugs in developing countries | 8.13 Proportion of population with access to affordable essential drugs on a sustainable basis |
| **Target 8.F**: In cooperation with the private sector, make available the benefits of new technologies, especially information and communications | 8.14 Telephone lines per 100 population |
| | 8.15 Cellular subscribers per 100 population |
| | 8.16 Internet users per 100 population |

Source: United Nations (2008c)

The first set of indicators listed in Table 3.8 is aimed at addressing issues relating to aid provided to developing countries for the purpose of growth and development by richer nations. The ODA statistic listed in the table are computed by the Organisation for Economic Co-operation and Development (OECD) (UN, 2003; Fuhrer, 1996).

The fact that these indicators are sourced from developed countries means that it is not too difficult to access them. What is more difficult to assess is if and how these development aids translate to meaningful impact at local level geographies.

Market access indicators are interested in trade relations and barriers. The agency responsible for sourcing and computing indicators 8.6 and 8.7 is the World Trade Organisation (WTO) while the OECD is responsible for indicator 8.8. The proportion of ODA required to help build trade capacity is jointly handled by the WTO and OECD. A major limitation with these indicators is that they are available only at world geography level and their impacts at local level geography are difficult to assess.

The indicators relating to debt sustainability are important for developing countries because debt reduction is a vital ingredient for development (UN, 2003). For some countries, the amount of money required to service what is owed alone acts as a major hindrance for meeting their national financial targets and on many occasions there is need to streamline national budgets to include these large amounts (Ross and Harmsen, 2001). The World Bank and IMF are saddled with the responsibility of monitoring these set of indicators.

The indicator on the proportion of people with access to essential drugs on a sustainable basis seems a little out of place and would probably fit into one of the health related goals. Access within this context is defined as the availability of affordable drugs on a continuous basis at public or private health facilities and other drug outlets. Such outlets should ideally be within a one hour walking distance from a person's residence. There is a subjective definition for essential drugs. A drug is classed as essential if it satisfies the health care requirements of much of the population (UN, 2003). In principle, it means that essential drugs will vary from country to country and so will their costs and availability. This makes it difficult to really conduct an inter-country comparison of this indicator because we will not be comparing like-with like.

The final three indicators within the framework of goal 8 deal with accessibility and use of Information Communications Technologies (ICT's). It is no longer news that we are now in an information age where digital remote communication techniques are preferred to their traditional analog counterparts (Lallana and Uy, 2003). There is a massive digital divide between developed and developing countries. This divide is mostly attributed to cost related factors (Lallana and Uy, 2003). Dated policies and regulations also contribute to the slow diffusion of these developments. Unlike the use of internet technologies, the patronage and use of mobile phones has witnessed a rapid explosion in many developing world countries. One of the major factors contributing to this is the openness to privatisation within the sector.

### 3.2.9 Other Challenges of the Developing World

The eight MDGs alone do not completely address current global challenges especially those related to developing countries. In fact they probably do not address some of the more pertinent issues like greed and corruption; wars and conflicts; and weak institutional frameworks.

Corruption in some third world countries has mutilated many pro-poor development programmes. We have seen situations where loans granted by global financial corporations like the IMF or World Bank are embezzled by greedy leaders rather than been deployed towards investing in the common good of the citizens of their countries (Gray and Kaufmann, 1998).

Unequal distribution of income and opportunities has also continued to contribute to the entrenchment of poverty in civil society. In many third world countries where governance is often aligned along ethnic or religious lines, opportunities and wealth are often constricted in the hands of very few people. Helping the poor get out of poverty therefore requires understanding the roots of poverty and being able to systematically identify and target pockets of the poorest segments of the population.

Also linked with corrupt practices is the fact that many developing world countries have been plagued by instability in their systems of governance. The challenges poised by civil unrest often lead to conflicts and wars which eventually destabilise society and displace people from their homes. Typical examples are drawn from Iraq, Rwanda and Sudan. Such unwarranted population displacement aggravates the problem and the scourge of poverty. It often results in the erosion of the minimal assets or forms of livelihoods that have been amassed by the poor segments of community and replaces it with a lack of sense of dignity. The UN estimates that there are currently over 42 million people worldwide in displaced conditions (UN, 2008b) courtesy of the conflicts in Iraq and instability in Sudan.

What conflicts and displacement does is to incapacitate the productivity of people especially the core poor. It erodes their ability to transform their aptitudes (natural and/or learning-entrenched) into meeting and sustaining their basic needs.

Developing countries therefore need to strengthen their institutional frameworks rather than strengthening the bank accounts of the political elite.

## 3.3 Shortcomings of the MDGs and United Nations Approach

In section 3.2.9 some of the issues not addressed by the MDGs but which remain challenges to third world countries were discussed. The focus of the next four sections will be to address the limitations of the MDGs and the approach of the UN from a geographical perspective. Some of the issues discussed relate to the unclear definition of development which has often defined the activities of the UN; the failure to understand that differences in magnitudes and types of challenges should also mean that solutions are tailored to meet these differences in needs; the problem of comparing like with unlike; and most critical to this research, the failure to be able to drill down to finer levels of geography.

### 3.3.1 Vague Definition of Development Rather than Inequalities

Much of the work of the UN, the World Bank and other partner agencies tend to be centred on issues of development economics. As such interests are often directed at pursuing economic growth or an unclear development.

The UN and other partner organisations consistently, in different contexts, talk about a shared vision of development. The terminology is often used vaguely and sometimes interchangeably with other jargons. So what really does development stand for within the context of the MDGs?

All too often within the UN framework, the word development is synonymously viewed as economic growth. From a pragmatic point of view, development can not be easily measured; however it should relate to third world countries while growth would be of particular concern of developed countries. This is because economic growth concerns itself with the level of national productivity and is a lot less interested in the activities of the informal sector while development activities are defined by an attempt to improve the standards of living, self-esteem and the liberation of people from oppression (Brinkman, 1995).

There is too much interest in macro-level economic expansion which does not necessarily translate into better quality of lives for people. A major limitation with this is that it fails to address the more pertinent issues of relative inequalities of opportunities or circumstances of different social groups within member countries. By investigating the extent to which important factors like socio-demographics and life style related issues influence within-country disadvantaged groups, one can usefully identify hidden problems and develop policies to tackle such inequalities. Unfortunately, rather than the growth led agenda pursued by these international agencies, poorer countries can only develop by redirecting the focus on more equitable distribution of resources.

### 3.3.2 One Size Fits All Approach

A major limitation of the MDGs is the fact that many of the targets are rigid. The problem-solving approach often deployed especially at country level often fails to address the need for flexibility in the priorities of countries. For instance the MDG on education is particularly interested in a full course of primary level schooling but fails to address any issues on secondary and post-secondary education

Many of the non-flexible ideas often used by these development agencies tend to drive the activities of national governments (especially in the developing world) when they want to target their resources. Many times, they fail to address the fact that the types and magnitudes of needs of people vary. This would also mean that there would be variations in the way people respond to the strategies or resources targeted at them.

### 3.3.3 Inadequate Basis for Inter-country Comparison

It is common practice that many of the global reports fail to critically underscore the differences in definitions of some fundamental issues. A clear example is education. Different countries have different primary school age enrolment systems. One of the indicators examines the proportion of pupils starting grade 1 who reach last grade of primary school. If you compare a country where the expected age at last grade is 11 years to a country where the expected age is 12 years; you are probably much more likely to find that the country with an expected age of 11 years would do better than the latter taking into context the fact that people are probably going to drop-out or even die as they grow older.

Another issue with the inter-country comparison is that since different definitions of the indicators obtain in the different countries, some of the data gathering and aggregation techniques would also be slightly different. Unfortunately, most government statistical agencies especially in developing countries are not as transparent with these issues. Additionally proxy indicators are sometimes used and re-named to fit the UN required standards.

### 3.3.4 Focus on Higher Levels of Geography

The failure of the UN and partner agencies to use their influence as a veritable mechanism for driving the advancement of problems at the local level is a major limitation of their approach towards solving problems. It is also one of the major arguments and questions to which this research project provides a proposed solution.

Having reviewed different MDG country reports, it is my view that most of these reports only tend to highlight within-country disparities at regional levels or at best providing rural/urban indices on progress towards meeting the goals. There is often little or nothing mentioned about what goes on at local scales.

At higher levels of geographical aggregation, it is often a lot easier to uncover disparities and inequalities on profound policy issues like poverty and well-being, ill health and illiteracy. At the local scale, these issues are relatively difficult to track-down. However, it is at these local scales that the problems can quickly become endemic.

Neighbourhood and community ties are often stronger at the local scales of geography. It is at the local or neighbourhood scale that the populace feel the implications of policies designed for them. While the impacts of strategies and policies targeted towards reducing inequalities may not quickly become apparent at the local level, they often turn out on the long run to be more effective and sustaining. This is why there is a fundamental need for a radical shift in the way the UN and partner agencies evaluate progress towards meeting the MDGs. Country reports and analysis done at regional, state or country levels often cloud variations at the local levels.

## 3.4 The Potential of the Geodemographic Approach

A geodemographic approach is proposed and used in this work to show how some of the MDGs can be evaluated at local level geographies to generate more meaningful insight. It is however pertinent to note that some of the issues associated with the lack of MDG analysis and reporting at the local levels of geography in developing countries are also closely linked with the challenges militating against the development of geodemographic systems in these countries. These issues are discussed in greater detail in section 2.6 in the previous chapter. Geodemographic segmentations however offer alternative solutions on a number of fronts as detailed in sections 3.4.1 to 3.4.8.

### 3.4.1 Strengthening the Concept of National Identity

There are several commentaries and publications arguing that the population groups within many developing countries are too *different* (see for instance Montalvo and Reynal-Querol, 2005a; Montalvo and Reynal-Querol, 2005b). It is common practice to find discussions on polarisation within these countries often anchored on ethnic or religious issues. Unfortunately, debates like these on many occasions have further widened the chasm of misunderstandings that exist within the population groups in ethnically diverse developing countries.

The sad reality however is that the negative aspects of politics have often been used as an instrument by the political class to distract the attention of people from the real culprits of widening inequality. Contrary to the arguments of numerous commentators, the position of this work is that people living in the same country are not as different or divided as suggested by political opinions.

By geodemographically segmenting small administrative areas, it is possible to show inter and intra regional strengths of the similarity between residents of such areas. By doing this, we may be able to prove that while an individual may originate from the northern part of the country, some of their needs and lifestyles may be a lot more similar to someone who originates from the south or east.

This idea which focuses on solving problems by identifying inter-regional similarities is a means for strengthening national identity. We know that in practice, region level geographies in developing countries are actually aligned on ethnicity. To label an entire region as deprived for instance which is common practice with most international agencies may spark ethnic or religious unrest. For some developing countries, such misleading interpretations can fuel civil unrests or deepen the lack of trust between population groups based on their religion or ethnic origins. What is needed in most third world countries is a re-focus on issues that can bolster a sense of national interest.

## 3.4.2 Insight to Variations Within Sub-national Populations

One of the key limitations of spatial analysis done in most developing countries is the failure of generating insight of local level disparities in the characteristics of the populations. Since most developing countries often depend on aid from international agencies like the World Bank and IMF, the organisational cultures of some of these donor agencies sometimes influences the analytics performed in developing countries.

The macro-level approach to embarking on inquiry is often the norm of research even within some academic communities in third world countries. Most of the profound issues like poverty and well being are analysed and reported at regional or at best state levels of spatial aggregation. The patterns of inequality within sub-national populations are often overlooked.

A fundamental position of this research work is that it is at the local scale that much understanding and action is required in order to accelerate development in third world countries. Having said that however, it is relatively difficult to embark on these analyses and generate reliable results especially in scenarios where data availability and access are problematic.

Geodemographic segmentation systems present an option for the investigation of local level inequalities especially for the data-scarce countries of the developing world. A segmentation system developed with national or near national coverage data-set for small or local areas can be used to evaluate people and area types. By plugging a survey of respondents from different locations within a country into the area segmentation, it is possible to generate initial insight into the fundamental characteristics of respondents to survey. Results can then be extrapolated nationally on the basis of the assumption that people who are more similar in their geodemographic make-ups will most likely reside in the same locality and those are types will be spread across the country.

Many of the MDGs are assessed based on surveys that are collected nationally. However to date, local level country reports and analysis are almost impossible to come by. The geodemographic option therefore not only offers the technocrat an alternative way of investigating local level inequalities but it also affords an individual with a non-technical background the ability to quickly and easily assimilate patterns of population disparities at the local scale of geography.

## 3.4.3 Informed Decision Making and Resource Allocation

For developing countries, the judicious allocation of resources is important in many respects. First, these countries harbour the largest population groups of the world (UNFPA, 2007). Secondly, the economic resources required to sustain the teeming population is limited and foreign aid is often required as a supplement. Thirdly, because of political instabilities and tensions (usually linked to lack of trust amongst ethnic and religious groups) most governments are often under pressure to favour some group of people over others many times on the basis of their ethnicity or religion.

Unfortunately, over time this problem has been compounded by the menace of deliberate falsification of national statistics to favour different groups (Holz, 2002). Solving this problem would of course require political will. While geodemographics can not be used as a substitute for political will, one area where it's potential can however be exploited is in its ability to provide decision makers with much more holistic information.

Public sector agencies are sometimes too comfortable with anchoring their decisions on unidirectional analysis where they look at the traditional age-sex variables alone. In many instances however, a fusion of multiple indicators can have the ability to provide a more holistic picture and reveal something new. Geodemographic methods ensure the ability to synthesise multivariate information relevant for better decision making. A very useful feature of geodemographic systems is the underlying textual and graphical explanations that accompany the results of the analysis. These descriptions are often referred to as profiles and are used to summarise the predominant attributes of the population groups.

Geodemographic profiling helps elucidate (in qualitative terms) information inherent in complex quantitative analysis. It therefore means that when public sector datasets are linked with geodemographic typologies, decision makers are not only informed about the direct relationships of indicators but they are also provided with potential likely or unlikely solutions.

### 3.4.4 Targeting Interventions

The targeting of workable and preventive strategies is important for any government that seeks to reduce cost on the long run. The process of targeting interventions requires mechanisms for identifying special populations or sometimes vulnerable population groups. Identifying such groups requires unveiling their attributes and locating their distribution across geographical space.

For instance if it is identified within a state (A) that certain neighbourhood types have high incidence of a particular type of communicable disease; by evaluating the geodemographic features and lifestyles of the residential population within state (B), it may be possible to predict the likely hot-spots for such ailment. This may be useful in helping the health agencies deploy their scarce resources and social marketing and communication intelligently and forestalling future problems.

### 3.4.5 Monitoring Impacts of National Policies

It is not just enough to make policies for people; more importantly, it is vital for the effectiveness of such policies to be evaluated. Such monitoring and evaluation process allows for the generation of input-feed back mechanisms and helps enhance the effectiveness of policy making.

Geodemographic analysis allows users to benchmark population groups and characteristics. This means that the performance of a neighbourhood type can be evaluated relative to others and relative to a national performance benchmark.

Monitoring can be done by comparing the relative performances of the neighbourhood types based on the chosen indicator over time. The benefit of doing this can obviously only be realised where time-series data or longitudinal datasets are available.

### 3.4.6 Public Sector Social Marketing

In the world of business and commerce, it is fundamental for the right product or service to be communicated to the appropriate person for a sale to be made. If the product or service fails to reach the final consumer then we can safely assume that the cycle of production is incomplete.

The same way commercial organisations seek to market their products with the aim of profiting; public sector organisations are also interested in ensuring that their products and services like adequate health care, education and poverty alleviation programmes are directed at the appropriate target groups. The volume, form and spatial disaggregation of these public sector commodities are vital for the satisfaction of the consumers who are the public.

The idea of introducing strategic and dynamic targeting techniques which take cognizance of the types and levels of variations in the characteristics of the target population is what is known as social marketing (Andreasen, 1995). In the United Kingdom, geodemographics has been used extensively as a veritable tool for public sector social marketing especially in the health sector (Powell et al., 2007).

Geodemographics provides a unique option for the public sector because it enables service providers to know who their customers are, what they do and their attitudes. It can help uncover segments of the population that have been hard to reach in the past. It also has the benefit of allowing service providers to vary their communication channels and set differing objectives based on the dynamics of target population groups.

### 3.4.7 Academic Potential: Open Geodemographics

The idea of geodemographics has spread rapidly across a great number of developed countries as shown in Figure 2.2, section 2.3.3 in the previous chapter. The pattern of countries with geodemographic systems closely mirrors the pattern of World Bank high income economies. This may suggest that most providers of these segmentation systems have focused mainly on their commercial and marketing potentials. At present, only one country - the United Kingdom has a freely and publicly available segmentation system (Vickers, 2006).

Apart from expanding the use of geodemographic techniques within the public sector, openly created geodemographic systems which are ratified by the academic community can also increase the research potential of these techniques and as it is in the U.K. result in their introduction into the Geography university curriculums. This will especially benefit developing countries and lead to increased literature on the subject.

This research work provides a unique opportunity to investigate the potential of adding Nigeria and the Philippines to the global map of countries with open-source geodemographic systems.

## 3.5 Conclusions

This chapter has provided a thorough discussion on some of the key elements of the MDGs. The idea of the definition of development was revisited and the importance of the concept of inequality especially to developing nations was stressed.

MDG data gathering techniques, measurements and monitoring processes have also been discussed. In addition to in-country statistical agencies, quite a number of international development partner agencies are also involved in the data gathering process for MDGs. Each organisation is defined by its driving cultures and interests and as such variations exist in the methods for computing some of these indicators.

In addition to the problem of variation in methodologies, the chapter has also stressed the fact that different countries have different definitions for some of the indicators. This obviously makes it difficult to adequately carry out inter-country comparisons.

Additionally, the approach to reporting progress made towards the targets is devoid of local level enquiries. It has been stressed in this chapter that it is at these local geographical scales that meaningful analysis of inequality can be investigated. This therefore results in the issue of devising a means by which MDG indicators, mostly derived from surveys which sometimes do not have national coverage can be analysed to generate insight on a national basis and at the local scale.

The chapter has proposed the geodemographic option as a veritable means for bridging the gap between regional and local level investigation of progress towards meeting the MDGs. Not only does geodemographics address this problem with the simplicity often requested by less technical audience of policy makers; it also underscores the need for differentiating approaches to policy making as opposed to enforcing a one size fits all approach.

# 4

# A Framework for Building Classification Systems in Developing Countries

## 4.1 Introduction

This chapter aims at discussing the techniques and considerations involved in creating geodemographic classifications. The discussions here will therefore centre on an exhaustive list of procedures which combine to define cluster analysis.

There are a series of steps which make up cluster analysis. The key steps were outlined by Milligan (1996) as follows:

Step 1: Clustering elements (Objects to cluster, also known as operational taxonomic units)

Step 2: Clustering variables (Attributes of objects to be used)

Step 3: Variable standardization

Step 4: Measure of association (Proximity measure)

Step 5: Clustering method

Step 6: Number of clusters

Step 7: Interpretation, testing and replication

What these steps suggest is that the process of developing a social area segmentation system is not merely running data through a clustering algorithm but involves a number of important steps which require both quantitative and qualitative analysis.

The remainder of this section will therefore follow logically by discussing and elaborating on the general outline of Miligan. Section 4.2 will stress the relevance of purpose in creating classification systems. It will argue that the use to which a classification system is to be put can be reflective of the variables it encapsulates and the manipulations which these variables go through. In section 4.3, the theory of variable selection is introduced by examining some conceptual and qualitative issues. The writings in section 4.4 examine the geo-statistical techniques involved in choosing input variables. Section 4.5 explores methods for re-aligning the input variables into comparable scales and measurement units which is a fundamental pre-requisite of clustering. In section 4.6, the implications and procedure of assigning a value of importance to different variables is examined. Section 4.7 provides insight to clustering methods while section 4.8 moves further to examine some methods of choosing optimum cluster solutions. Prior to drawing a conclusion in section 4.10, section 4.9 examines methods for naming and describing the cluster solutions.

## 4.2 Purpose of the Classification

The aims and objectives of a research project define what is written in the final thesis. This same thinking applies to geodemographic classifications. It is common practice for vendors of commercial classifications to suggest that their system is the best for almost any purpose. However, this may not be true as Vickers (2006) suggests that there is a relationship between the purpose of a classification and the variables it encapsulates.

If a system is primarily targeted for the health industry for instance, a good proportion of the variables that will go into such a system will be useful proxies for measuring good or poor health. The inputs into commercial classifications emphasise lifestyle data which is reflective of affluence (Harris et al., 2005).

Geodemographic classifications have a varied number of application areas which have been discussed in the second chapter of this thesis. The wide range of application areas cut across different disciplines. Gordon (1999) suggests two key purposes of any classifications scheme. The first being *data simplification* and the second *prediction.*

It is helpful to have a clear purpose of the classification before progressing with analysis. This will be particularly helpful in deciding which variable domains should drive the classification. If a classification builder intends to apply weights to variables before clustering, emphasis may usually be on variables which define the purpose of the classification. This is because weights have the ability to reduce or increase the influence of particular variables on the classification (Harris et al., 2005). Gordon (1999) argues that this approach can be subjective and should be avoided if definitive information is absent about the relevance of the different variables.

By examining the different domains to which the variables used for a classification system belong and the proportional representations of these domains, it is possible to asses how *general* or *specific* the purpose of the classification is.

## 4.3 Prolonged Stability and Updatability of Inputs

In the previous section, the purpose of a classification has already been linked with the variables which it encapsulates. This section examines some more pertinent issues surrounding policy implications and the life-span of the input variables.

Variables that can sustain the longevity of the classification system are highly desirable. One key theoretical principle that should be considered during the choice of variables is the sensitivity of each variable. The rate at which the values of variables change over time can have implications on the reliability of the system (Vickers, 2006). Variables that will sustain the classification over its life course are those that do not have the potential of yielding to large changes across areas.

Census variables have a benefit of wide geographic coverage but face the problem of updatability. In Nigeria for example, the national Census should ideally hold every 10 years. However this has not been the case over time. Long years of political instability have contributed to inconsistency in the conduct of the census. The most recent census was in 2006. Prior to that, the census was in 1991.

This raises the importance and significance of national surveys. Many of the current widely used classifications in the developed world supplement census statistics with data derived from life-style surveys (Harris et al., 2005). The argument is that such surveys are conducted much more frequently, utilise the same geography as the census and ask more direct questions beyond population demographics and housing.

This argument therefore presents a valid option especially in developing countries where acquisition of government data inventories can present significant challenges when undertaking data-intensive research (Okonjo-Iweala, 2007). However when combining data from different sources, care must be taken to ensure that they are derived from comparable geography and that appropriate methods of interpolation or extrapolation is used (Goodchild et al. 1993; Fotheringham and Rogerson, 1993; Langford et al., 1991).

## 4.4 Variable Reduction Techniques

Although as discussed in Chapter 2, one of the geodemographic challenges faced in developing countries is that of securing data (Okonjo-Iweala, 2007); where such datasets are made available, it is important to decide on which variables will be appropriate for inclusion in the analysis.

In section 4.2, it was established that one of the general factors underpinning the creation of any segmentation system is the need to depict simplicity from what would otherwise be a complex set of information (Gordon, 1999). The process of accommodating simplicity therefore should commence from the stage of choice of variable inputs.

Where a substantial volume of variables are made available in the dataset provided for analysis, it is important to ensure that the most relevant variables are included in the build process (Harris et al., 2005). What this suggests is that in an ideal situation, not all variables made available would be appropriate inputs for the purpose of the classification system. Indeed the potency of the proposed classification system will often depend significantly on the underlying variables it encapsulates (Everitt et al., 2001). This is why appropriate care must be given to the procedure of variable selection.

Another important reason for ensuring that variables included in a segmentation system are carefully selected is because the proposed system is intended to reflect real-life social patterns as much as possible. A critical and rigorous approach to variable selection will ensure that the segmentation system is robust.

Different schools of thought exist on the proportion of variables to be included in the creation of segmentation systems. Some authors argue for fewer variables (Voas and Williamson, 2001a, 2001b) while some others particularly those who appear to be inclined towards commercial geodemographics appear to support the view that the more the variables included in a segmentation system, the more robust the system is likely to be (Harris et al., 2005).

For developing countries where classification systems are lacking, these arguments on the platform of *fewer* or *more* can be confusing for potential developers. The question that springs to mind is this: What is *few* and what is *more*? There is no academic literature anywhere that states the specific number of variables that should go into the build process of geodemographic systems. On this basis, it may be inappropriate to argue for fewer or more variables because there is no standard threshold of variables which can be included. Rather the focus of the variable selection should be based on evidence showing that such variables are adjudged to be relevant (conceptually appropriate) and analytically appropriate by adding something *distinct* to the classification system.

### 4.4.1 Principal Components Analysis

One of the key reasons why it is important to carefully examine a pool of variables and reduce them before further analysis is because of the possibility of multicollinearity within the dataset. Multicollinearity is a situation where two or more variables are highly correlated (Walker and Maddan, 2008). It is a feature that is quite common within socio-demographic variables particularly when such variables are conceptually deemed to belong to the same domain.

Multicollinearity is a feature that is not desirable for cluster analysis because of the introduction of unneeded or duplicated information into the analysis. A preliminary data preparatory analysis therefore usually is to determine which variables represent the most the principal dimensions of an original dataset. Principal Components Analysis (PCA) is a common technique which used to achieve this task. It is an analytical technique used to alter the relationship of correlation existing within a dataset by transforming a set of correlated variables to a smaller set of uncorrelated variables (Dunteman, 1989; Manly, 1994).

The technique of PCA works by analysing and explaining the variance within an observed dataset by linearly combining the original data. It is these linear relations that are termed the principal components of the data. What this means is that while PCA in cluster analysis can help mitigate the problem of redundancy (Everitt et al., 2001; Harris et al., 2005) the new data values which it creates (called composites) may not be easy to explain or interpret if included in the cluster analysis.

Another usefulness of PCA as a data reduction technique is its ability to help identify which variables are likely to have significant influence on the clustering process (Jolliffe, 2002). One of the outputs of the analysis is a weights table also called a components loading matrix which is indicative of the power that each variable exerts. Typically, for a variable, the higher the value of this weight, the better its variance can be explained by the corresponding principal component and consequently the greater the power it is likely to exert on the cluster analysis (Jolliffe, 2002).

If using PCA to determine which variables are likely to power the dataset, it is advisable to examine different components to see how variables behave in terms of the magnitude of their loading. Usually, variables with higher loading values are more likely to have greater influence on the dataset (OECD, 2008; Vickers, 2006). A trade-off however needs to be made between the different components to decide which one will inform the choice of variables that will most likely power the classification. Since the first principal component usually accounts for the greatest level of variation within the data (OECD, 2008) it is safe to use this component to determine which variables are likely to have the greatest influence on the classification.

## 4.4.2 Missing Values and Small Sample Sizes

It has already been noted in section 4.3 that apart from using census statistics which have nation-wide coverage, geodemographic classifications can often be supplemented with survey datasets. Although data from surveys particularly lifestyle surveys have the potential of supplying information not contained in national censuses (Harris et al., 2005), one of the shortcomings suffered by such datasets is that they often contain missing values for certain geographical areas. Where a variable contains too many missing values, it provides incomplete information which is not too helpful for the analysis.

Another important factor to consider when choosing input variables is the proportion of the population which they represent (Vickers, 2006). The fact that a variable is derived from the national census does not necessarily mean that it will represent a large percentage of the population. When a variable represents only a small proportion of the population there is the tendency for that variable to be volatile and change rapidly over time. Such a variable would not sustain the longevity of the classification (Vickers and Rees, 2007). Another problem that can result from variables with small sample proportions is that they provide little distinctive information for naming and profiling the clusters created from the analysis (Harris et al., 2005). A solution that can be considered for such variables is to merge them where they fall under the same domain and share a similar base.

### 4.4.3 Issues Relating to Skew

Prior to choosing a variable for inclusion in a clustering algorithm, it is important to consider the extent to which it is skewed. Skewness defines the extent to which a variable distribution is symmetrical about its mean (Crawshaw and Chambers, 2001). A variable is said to exhibit positive skew if its asymmetric tail (when charted) extends towards the positive values of the distribution while for a negative skew, the tail extends towards the negative values.

A normally distributed variable would produce a skewness value of zero which would ideally be desirable for variables to be included in the cluster analysis (Harris et al., 2005). However, in practice, this is rarely the case with spatially referenced socio-demographic datasets.

Skew can result from a number of factors. If a variable represents a small proportion of the population, most values will be concentrated around the lower end of a 0 to 100% scale. Skew can also be caused as a result of extreme values or outliers. An example of a variable that may demonstrate this feature in most developing countries is *population density*. It is common in most developing economies to find urban centres with disproportionately higher population concentrations than their rural counterparts as a result of mass rural-urban drift (George, 1999).

The problem with introducing highly skewed variables into the analysis is that they may obscure the rest of the dataset and create artificial or outlier clusters. It is therefore important to test variables for their skewness and avoid where possible the inclusion of highly skewed variables in the analysis. It is practically impossible to completely eliminate the presence of skew when creating classifications; however their effects can be minimized by transforming the data (Walker and Maddan, 2008). Methods of transformation are discussed in section 4.5.

### 4.4.4 Cross-correlations and Geographic Variation of Variables

Usually, a pool of independent variables is considered for inclusion in a cluster analysis. The fact that variables are independent of each other does not necessarily mean that certain association types do not exist between them. It is important to determine the types of relationship that exist between variables before including them in cluster analysis for a number of reasons.

Knowledge of the relationship existing between two or more variables is informative of the level of redundancy that may be introduced into classification (Walker and Maddan, 2008).The inclusion of two highly related variables in a clustering algorithm will often result in the repetition of the same information or population behaviour (Milligan and Stephen, 2003). This can give undue advantage to such behaviour and can also mask other important underlying characteristics existing within the population.

In mathematical terms, the level of association between two or more variables can be determined using the coefficient of correlation usually represented by the letter ***r*** (Crawshaw and Chambers, 2001). The correlation coefficient is a numerical value between -1 and 1 such that a value of 1 signifies a perfect positive correlation and -1 represents a perfect negative correlation. A value of 0 indicates no correlation.

Another important issue to consider in the variable selection process is the extent to which variables vary across geographical areas. A useful statistic for measuring the geographic variation of variables is the standard deviation. Variables with larger standard deviations will prove more useful than those with lower values (Harris et al., 2005). This is because they present better distinctions between areas. However it is important to consider the sample size of variables across areas especially when dealing with variables with low standard deviation values. Some of these variables can be merged and renamed.

### 4.4.5 Composite Variables

One of the ways of dealing with the problem of small sample sizes is to combine two or more variables into a single variable. Doing this results in what can be described as a composite variable (Vickers and Rees, 2007).

Apart from helping to deal with the problem of small sample sizes, creating composite variables also indirectly helps reduce the number of initial variables considered for inclusion into a cluster analysis.

When attempting to merge two or more variables, it is important first to ensure that the variables share the same denominator. However, it is also imperative to ensure that the merger makes sense. For instance if we consider different ways in which women in developing countries deliver their children, some means expose them to greater level of risk than others (UNICEF, 2009). All women (100%) would be described by the different means of child delivery. However when considering a merger of variables, it is more appropriate to merge variables that describe safer means of delivery together and vice-versa.

## 4.5 Transformation and Normalisation of Data

Following the decision on the input variables, the next step in the analysis is to prepare the data for classification. Data preparation in this context involves assembling the data into a database and deploying further analytics on them appropriate for inclusion into a clustering algorithm.

The next phase in the analysis is an exploration of the scales of the variables. Scale in this context refers to the unit of measurement like percentage or ratio. It is inappropriate to run a clustering algorithm on a dataset which consists of variables of different scales. The reason for this is that the nature of the scale may cause undue advantage to be given to certain variables while others suffer. Methods of normalisation help in re-scaling distributions (Berrar et al., 2003).

Prior to normalisation, variables may need to be transformed particularly when there is a need to contain the effects of skew (Harris et al., 2005). Different methods transformation and nprmalisation are discussed in the following sub-sections.

### 4.5.1 Transformation of Data

The most common type of skew inherent in spatially referenced socio-demographic datasets is the positive skew. The log transformation has been found to be particularly helpful in dealing with this problem (Vickers, 2006; OECD, 2008). A logarithm works by relating a base number to an exponent. The exponent to which a chosen base value must be raised in order to derive the original value of the variable is the logarithm. When this method is used on data, natural logarithms of the variable are used rather than the original values of the variable.

Other forms of transformation which are not as popular as the log transformation may also be considered. The square root transformation which involves taking the square root of all values in a distribution can also be considered (Crawshaw and Chambers, 2001). One thing to bear in mind with this method is that it is impossible to calculate the square root of a negative value. If a distribution has negative values, then a constant will need to be added to all values to move them above the zero (0) mark.

Reciprocal transformation involves dividing one (1) by the original value of a variable. The resultant value is an inverse of the original value of the variable. Reciprocal transformations have a characteristic of making very small numbers large and vice versa thereby reversing the order of the original data. This attribute may not be too helpful for a cluster analysis procedure.

### 4.5.2 Standard Normal Variate Score (Z-scores)

It is inappropriate to combine datasets that are measured in different scales because the resultant output will not be a true reflection of reality. Variables measured on a scale with a large range will be given undue advantage as opposed to those measured on a smaller range (Harris et al., 2005). By converting the entire datasets to z-scores, the problem can be greatly reduced.

$$Z_i = \frac{x_i - x_{(mean)}}{\sigma} \tag{4.1}$$

where

$Z_i$ is the standardised value of the variable for area i

$x_i$ is the original value of the variable for area i

$x_{(mean)}$ is the average value of the variable across all areas

$\sigma$ is the standard deviation of the variable across all areas

Equation 4.1 shows how z-scores are computed by dividing the difference between the original value of the variable and the mean across all geographic areas by the standard deviation of the variable across all areas. When a distribution is converted to z-scores, they have a new mean value of 0 and a standard deviation of 1.

### 4.5.3 The Range and Inter-decile Range Standardisation

Though not as popular as z-scores, range standardisation is another method for normalising data. It is also called re-scaling (OECD, 2008). This method was deployed in the creation of the United Kingdom (UK) Output Area (OA) classification (Vickers and Rees, 2007) and is given by equation 4.2.

$$S_i = \frac{x_i - x_{(min)}}{x_{(max)} - x_{(min)}} \tag{4.2}$$

where

$S_i$ is the standardised value of the variable for area i

$x_i$ is the original value of the variable for area i

$x_{(min)}$ is the minimum value of the variable across all areas

$x_{(max)}$ is the maximum value of the variable across all areas

Range standardised values usually range between a minimum value of 0 and a maximum value of 1. The method has the advantage of working relatively well with extreme values by shrinking all values within the 0 to 1 range.

A variant of the range standardisation is the inter-decile range standardisation. The difference between the first and ninth decile is known as inter-decile range (Kirch, 2008). The standardised values in this case are computed by dividing the difference between the original value of the variable and the median value of the distribution by the difference between the $90^{th}$ and $10^{th}$ percentiles.

## 4.6 Weighting of Variables

Adding a weight to a variable signifies the level of significance of such a variable (Gnanadesikan et al., 2005) to the classification system. This is precisely why this process is subjective. What one analyst considers important may not be important to another analyst. Even when a variable is considered important, the level of importance assigned to such a variable may vary from person to person.

Some of the factors that influence analysts in their decisions to weight variables include their knowledge of the country in question and their judgments in terms of how the variable may relate to policy. Public perception could also have its influence on the decision to assign weights to variables (OECD, 2008).

Applying a weight to a variable will ultimately influence the resultant output of the cluster analysis (Harris et al., 2005). Weighting results in a situation where the clusters from the analysis may be skewed in the direction of the variable with a higher weight (Gnanadesikan, 2005). The impact is even greater for entities (geographical areas) with higher values within such a variable. The effect of the weighted variable can place them in clusters that they would not belong to in reality.

Another reason why weights are deployed in cluster analysis is to compensate for the quality of the statistical dataset available for analysis (OECD, 2008). In this context, the judgment of which group of variables to weight will depend on the analyst. Whether the higher weights are assigned to variables with broader geographic coverage or vice-versa, care must be taken to ensure that variables which are deemed relevant for the analysis are not penalized for others.

In most studies that employed or discussed cluster analysis (Gordon, 1999; Everitt et al., 2001; Harris et al., 2005), authors do not say it is compulsory to apply weights to input variables before deploying the chosen clustering algorithm. Indeed, when embarking on the variable selection process, some form of unintentional weighting occurs. Choosing one variable over another means it is more important and potentially more relevant to the analysis.

If an analyst is not completely sure of the level of theoretical and statistical relevance of a variable then it is better not to weight such a variable. In developing countries where there is a paucity of geodemographic systems and where knowledge of the spatial patterns of potential clusters has not been experimented, it may be more appropriate to avoid weighting variables.

When working with equal weights however, it is very important to ensure that multi-collineraity is reduced to the barest minimum. If two or more collinear variables are included in the analysis, then the spatial dimension which they represent will be doubled are trebled as the case may be. If non-weighted variables can be exposed to this problem, then it can become even more problematic when correlated variables are weighted.

## 4.7 Clustering Methods and Algorithms

So far in this chapter, it has been shown how cluster analysis is not just a statistical model but a series of coherent analytical steps. The eventual aim in the process however is to be able to partition a set of objects (geographical areas in this case) into groups based on their defining characteristics.

There are numerous clustering methods and algorithms available. Among other things, the choice of a technique will often depend on the size of the dataset and the data types. Many of these methods have been described in detail by Gordon (1999) and Everitt et al. (2001). In general Harris et al. (2005) identified that clustering methods are broadly divided into two types:

(i) Hierarchical clustering methods and

(ii) Iterative relocation techniques

The next section will briefly discuss hierarchical techniques while sections 4.7.2 and 4.7.3 will focus on the K-means clustering and the TwoStep cluster analysis procedure which were used in this thesis.

### 4.7.1 Hierarchical Clustering Methods

Hierarchical clustering techniques do not require a specification of the number of clusters at the beginning of the analysis (Everitt et al., 2001). They are further sub-divided into agglomerative or divisive methods.

The process by which Agglomerative hierarchical clustering works was summarised by Johnson (1967) as follows:

(i) Start by assigning each item to a cluster
(ii) Find the closest pair of clusters and merge them into a single cluster
(iii) Compute distances between the new cluster and each of the old clusters
(iv) Repeat steps 2 and 3 until all items are clustered into a single cluster

The third step can be executed in various ways. This is what differentiates single-linkage from complete-linkage and average-linkage clustering (Everitt et al., 2001). The single-linkage method computes distance values between clusters on a nearest neighbour basis while the complete-linkage uses a furthest neighbour distance measure. Average-linkage on the other hand computes distances as the mean of all possible pairs of cases in the eventual cluster (Everitt et al., 2001). Ward's hierarchical clustering method is one of the most common methods of agglomerative clustering and it uses a squared Euclidean method to compute distances. This is discussed further in section 4.7.4.

The iterations deployed during the agglomerative technique ends up in a single cluster which is usually not the desired result. However also usually accompanying the output is a complete hierarchical tree or dendogram of multiple cluster solutions. This is the main advantage of this technique. These techniques have a computational weakness in coping with very large datasets (Everitt et al., 2001).

The divisive hierarchical clustering works in a reverse manner to the agglomerative technique. In this case, rather than make all objects clusters from the beginning, they are all assigned into a single large cluster and subsequently subdivided into smaller bits (Everitt et al., 2001). These methods are rarely used and like the agglomerative techniques, they are computationally demanding.

### 4.7.2 K-means Clustering Method

A major computational weakness of hierarchical clustering techniques is the fact that a distance matrix usually has to be computed between all pairs of clustering objects (Everitt et al., 2001). Doing this makes it difficult for these methods to be used on large datasets.

The K-means clustering method is an iterative relocation algorithm which does not require the computation of all possible distances between objects. Another major difference between K-means and hierarchical clustering techniques is that the number of clusters (k) has to be specified at the inception of the clustering process (Harris et al., 2005). This means that multiple iterations have to be run over and over again with different values of k and the final solutions compared to be able to decide on the most appropriate solution (Gordon, 1999).

K-means works by moving cluster centres and in the process testing for an improvement in the move. When an object is moved from one cluster to another, then the sum of the squared deviations of the cluster containing the object is calculated to see if the result is better than the previous result (Aldenderfer and Blashfield, 1984). Objects are assigned to the cluster to which they bring the greatest improvement. Every object is processed in this manner until no improvement occurs again.

Once all objects have been assigned to their appropriate cluster, the cluster centre is re-calculated based on the constituent objects of the cluster.

The fact that any of the seeds can be selected as initial cluster centres makes this method sensitive to the effect of outliers. If the initial cluster centre is defined by the presence of an outlier, the resulting final cluster may be comprised of a small number of objects (Gordon, 1999). Therefore if K-means is to be used, it is essential that the analyst examines variables thoroughly to ensure that outliers are removed as much as possible.

### 4.7.3 TwoStep Cluster Analysis

Some times the dataset to be clustered may be a combination of different data types. In chapter 6, categorical and continuous data were provided for the Philippines so a clustering algorithm that could combine both data types was used.

The TwoStep Cluster analysis procedure is an algorithm specifically designed to handle combinations of categorical and continuous datasets (Banerjee et al., 2004). The procedure gives the best results if categorical datasets appear to have a multinomial distribution and continuous variables display a normal distribution.

During the first step of this clustering procedure, preclusters are formed using a distance measure. The preclustering process does not require a pre-specification of the desired number of clusters. Every object is considered in relation to already formed clusters and based on the distance measure, it is decided if an object should start a new precluster or be assigned to an already formed cluster (Hellerstein and Stonebraker, 2005).

Once the preclusters have been formed, each of the clusters is considered as a single object. The second step of the procedure deploys a hierarchical algorithm on the preclusters (Hellerstein and Stonebraker, 2005). During this step, cognizance is taken of the number of preclusters formed. Large number of preclusters results in better solutions (Hellerstein and Stonebraker, 2005) but demand more computational power which in turn slows down the algorithm.

### 4.7.4 Similarity and Distance

The overall aim of creating a geodemographic classification system is to be able to assign objects into groups. In each group, it is essential that constituent elements are as similar to each other as possible on the basis of the dimensions of measurement.

How is similarity or dissimilarity measured? Since the analysis is a quantitative exercise, the extent to which objects are similar to each other is usually determined by a statistical measure of distance in taxonomic space (Harris et al, 2005).

Numerous measures of distance exist and are discussed extensively in literature with some excellent reviews by Gordon, (1999); Everitt et al., (2001) and OECD, (2008). The focus in this section will be on the methods considered for use in chapters 5 and 6 of this thesis.

The Euclidean distance measures the distance between two points in multidimensional space as a straight line (Harris et al., 2005). This distance measure is derived by extending the Pythagoras theorem and computing the square root of square differences between the coordinates of a pair of objects (OECD, 2008). Equation 4.3 illustrates how it is calculated.

$$D(x, y) = \left( \frac{\sum_{I=1}^{N_d} (x_i - y_i)^2}{N_d} \right)^{1/2} \qquad (4.3)$$

where

**D** is the Euclidean distance between points x and y

$\mathbf{N_d}$ is the number of dimensions

In the case of squared Euclidean distance, the same equation is used without taking the square root of the values. The fact that square roots are not computed makes this method of calculating distance faster than the Euclidean distance metric. Squared Euclidean distance is derived by using equation 4.4.

$$D(x, y) = \frac{\sum_{I=1}^{N_d} (x_i - y_i)^2}{N_d} \qquad (4.4)$$

where

**D** is the Euclidean distance between points x and y

$\mathbf{N_d}$ is the number of dimensions

The distance measures discussed above can handle continuous variables. If data is a mixture of categorical and continuous variables a different distance criterion (a likelihood ratio test) has to be used to handle the categorical variables. Likelihood ratio tests are based on probabilities and hypothesis (Hellerstein and Stonebraker, 2005).

The Schwarz Bayesian Information Criterion (BIC) given in equation 4.5 and used in chapter 6 is commonly applied in the field of econometrics. The BIC is a model selection principle used to determine which of ***m*** parametric models best represent a set of ***n*** data samples (Tritschler and Gopinath, 1999).

$$\mathbf{BIC} = -2 \log \text{likelihood} + \log(\boldsymbol{n}) \times \boldsymbol{m} \qquad (4.5)$$

where

**BIC** is the Bayesian Information Criterion

***n*** is the sample size

***m*** is the number of parameters in the model

It is a method that assesses different models with varying parameters in order to choose that which is the *best fit* to the observed dataset (Schwarz, 1978; Akaike et al., 1998). The term *best fit* within the context of clustering means that as objects are assigned to clusters a log likelihood ratio test is deployed with the aim of deriving the largest possible value of the likelihood ratio test.

The BIC has been used extensively within the subject of cluster analysis particularly for model-based clustering (Fraley and Raftery, 1998; Tritschler and Gopinath, 1999; Fraley and Raftery, 2002).

If a likelihood ratio test like the BIC is used, the higher the value of this ratio, the more objects within a particular cluster are similar. In the case of the Euclidean and squared Euclidean distances metrics, the smaller the computed value, the greater the similarity between objects.

When clustering spatial datasets involving categorical variables, BIC has been suggested and used by (Bacher et al., 2004, Zhao et al., 2008) for choosing the possible number of stable clusters present within the dataset. Essentially, by plotting the change in the BIC against the number of clusters, the analyst can observe to detect a point when there is an abrupt change in the BIC resulting in what Zhao et al. (2008) have called ‘a knee-point change’. A perceived ideal solution would be the point at which there is an increase in the BIC (Banerjee et al., 2004; Larose, 2006).

## 4.8 Choosing the Number of Clusters

One of the principal challenges encountered by analysts when embarking on cluster analysis is how to determine an appropriate number of cluster groups. In many developing countries where geodemographic classifications have not been previously created, it is even more challenging as there are no platforms against which results can be measured or assessed.

Methods used to investigate the probable number of clusters are slightly informal. Everitt et al. (2001) identified the possibility of plotting the value of the clustering criterion against the number of clusters and observe points of great change in the plot. Generally, the average distance of cases from their cluster centres can be plotted against the number of clusters.

Some times, evaluating the clustering criterion against the number of clusters may not clearly reveal the probable number of clusters as was the case in the Nigerian classification. Other issues that can be considered include the homogeneity of the cluster solutions. It may be practically impossible for each cluster to contain an equal number of objects but where the distribution of cases in each cluster is reasonably balanced, it is helpful when using the classification system to analyse further survey data particularly where the sampled population is not very large.

## 4.9 Naming and Profiling Cluster Subdivisions

Labelling a group with a name can be contentious. It is a very complex process requiring consideration of numerous issues. The names assigned to different clusters are expected to be as widely representative of the characteristics of most the people living in those areas as possible. This does not in any way imply that every single person within a cluster can be labelled that way. To some extent, diversity still exists within similarity (Voas and Williamson, 2001a). The names attached to clusters are only indicative of the predominant features of the areas in question.

By examining the composition of the different clusters, between and within cluster variation and the level of importance of each variable to the cluster, significantly important variables can be determined (Harris et al., 2005). This can serve as a guide when deciding the name to be assigned to a cluster.

Since geodemographic systems are linked with administrative geographic areas, it means they can be mapped. Visualising the spatial distribution of clusters can also help inform the name assigned to a cluster. For instance there may appear to be some bias of cluster for a particular geographic region.

Names should not in any way be offensive especially in multi-ethnic or multi-religious countries. Care has to be taken to ensure that the names chosen do not appear to stigmatise any section of the population.

Cluster profiles and pen portraits are textual and graphical descriptions which summarise the prevalent characteristics of each cluster. They have the benefit of elucidating (in qualitative terms) some of the information inherent in complex quantitative analysis. Profiles and pen portraits have been developed for both the Nigerian and Philippines classification systems and are contained in the enclosed compact disc accompanying this thesis.

## 4.10 Conclusions

The discussions in this chapter have shown that the process of making a geodemographic classification system is a series of coherent steps. At every step, very important decisions are made which influence the final solution from the analysis. Some decisions may be subjective and dependent on theoretical inclinations and will vary from one analyst to another; others require robust statistical evaluations backed by a host of literature.

Having a clear understanding of the purpose of a classification system is important. It may influence the choice of variable inputs that go into the analysis. In developing countries where spatial statistical datasets are difficult to come by (Okonjo-Iweala, 2007) making a clear argument for the purpose of a classification system and linking it to policy issues may be beneficial when sourcing for data from government statistical agencies.

It is important to test variables against each other before deciding to include them in the final clustering algorithm. In many scenarios, an analyst may find that all the original variables supplied for the analysis are not appropriate for inclusion due to factors explained in detail in section 4.4.

Except where it is the case that all variables made available to the analyst that they are all measured on the same scale e.g. percentages, it is important for variables to be transformed. Different transformation methods have their strengths and weaknesses as discussed in section 4.5 and some are more suited for certain data types than others.

In the literature of geodemographics, the application of weights to variables is also contentious and should be guided not just by expert knowledge of theories and variables as they relate to the country in question but also by a justified need to apply weights.

It was recounted in section 4.7 that clustering methods are split into two broad types: hierarchical clustering methods many of which are often more computational demanding and have the ability to cope with less data volumes and the iterative relocation techniques (Harris et al., 2005). The methods used in this project – K-means clustering technique and the TwoStep clustering procedure were reviewed.

Just like most other steps of the analysis, the choice of the final number of clusters is not formal although some suggested considerations have been reviewed in section 4.8. Once clusters have been chosen, it is important and helpful to profile the clusters and assign names to them so that they can be meaningful particularly to a less analytical minded audience.

# 5

# Classifying Nigeria's Local Government Areas

## 5.1 Introduction

The challenge of this chapter is to explain the series of decisions made during the creation of a classification system at the Local Government Area (LGA) geography in Nigeria. Section 5.2 provides a general description of the geography for which the classification is built. In section 5.3, the variables encapsulated by the classification are examined. This section also explains the methodology utilized for selecting variables. The writings in section 5.4 detail information on data manipulation including standardisation and clustering methods. Section 5.5 attempts to evaluate the outputs from the cluster analysis. Pen portraits and graphical visual details are also provided alongside detailed textual information. Section 5.6 provides conclusive comments.

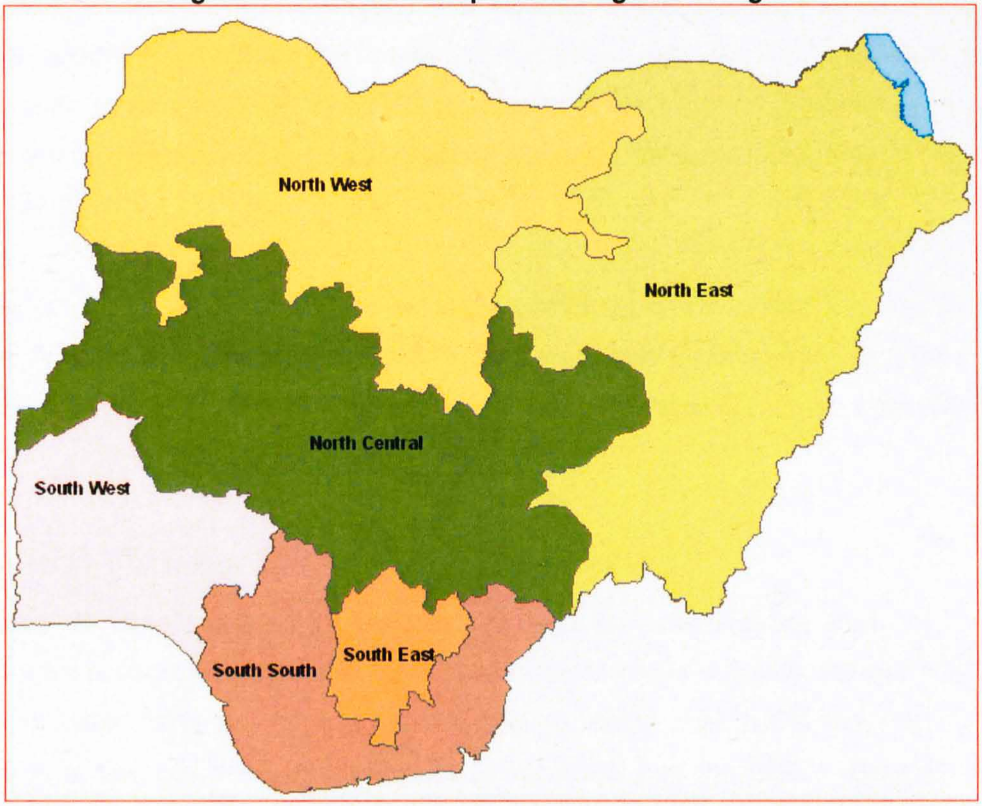
## 5.2 Demographics and Administrative Geography

With a population of about 140 million people, Nigeria is the most populous country in Africa (NBS, 2006a). The country also enjoys an average population growth rate of 2.4% on an annual basis (World Bank, 2007a). Between 1990 and 2005, an estimated 70.8% of the population lived on less than one United States Dollar a day according (World Bank, 2007b). The country is currently classified as a low income country.

Comprising an area totaling 356,669 square miles (Gordon, 2003) Nigeria has a rich base of human capital and natural resources. The country also ranks among the top ten oil exporters globally. Nigeria maintains a high profile economic and political status on the African continent. Indeed some commentators (Gordon, 2003) agree that several African countries have their economic stability hinged on the political and economic stability of Nigeria.

Nigeria is made up of more than 250 ethnic groups each with its local language (dialect) and multiple religions. Amongst these groups there are three dominant groups based essentially on their population. These are the Yoruba groups located mainly in the South-western corner of the country; Igbo groups can be found in the South East while the Hausa-Fulani ethnic groups concentrate in the Northern part of the country.

**Figure 5.1: The Six Geo-political Regions of Nigeria**




The geography of Nigeria is related to its ethnic makeup. Nigeria's top level of administrative geography comprises six geo-political zones shown in Figure 6.1 While the South West, South East and North West/North East are dominated by the three major ethnic groups, the North Central and South-South zones are composed of less populated ethnic groups.

Each geopolitical zone is further subdivided into states which are 36 in number. At the state-level geography, there is also a national capital city called Abuja which is located centrally in the North Central zone. The administrative boundaries for states are therefore technically 37 in number. Each state is administered by a state government with a governor as the head. States are further split into smaller geographies called Local Government Areas (LGAs). It has been argued that the LGA level is where most policy benefits can reach the masses at the grass root (Olowu and Ayo, 1985). Nigerian LGAs are 774 in number. Beyond LGAs are Enumeration Areas (EAs) which are traditionally used for census enumeration and total up to 662,000. Significant effort was made to try and secure data for the project at the EA level without success. After about ten months of consultation with the National Populations Commission (NPC) and the National Bureau of Statistics (NBS), the dataset used for the analysis was supplied by the NBS for all the 774 LGAs of the country.

## 5.3 Data Source and Input Variables

The NBS supplied data from the latest 2006 Census and the Core Welfare Indicators Questionnaire Survey (CWIQ). The NBS is an arm of the Nigerian government responsible for the provision of timely data and vital social statistics for various aspects of the country's economy.

The data provided by the NBS proved very useful as it comprises the same statistical resource employed by the government in planning and development. It has a national coverage of the entire 774 Local Government Areas and expands across a variety of topical domains.

### 5.3.1 Initial Variable Set

The choice of variables encapsulated in the Nigerian classification took into context a number of theoretical and quantitative issues. Dataset made available for the 774 LGA's in the key statistics table provided a total of 644 variables. The initial resource was made available in a disc but the alignments of the variables and the LGA's required significant manual work to prepare the data appropriately for further analysis. The resulting resource was a total of 498,454 data points (i.e. 644 × 774).

Clearly 644 variables would prove too much for inclusion in the classification. For one, some variables may be highly correlated with each other meaning they may be able to explain the same population characteristics as demonstrated in section 5.3.2. Secondly, it may prove very difficult to identify distinctive features of the LGA's after clustering which could make naming and profiling cluster groups hectic. It was therefore decided that the 644 variables would be reduced to a manageable size before progressing to other analysis.
As a starting point, from each of the 644 variables, a total of ten (10) domains were identified. These domains are listed below:

i. Agriculture
ii. Demographic
iii. Education
iv. Employment
v. Health
vi. Household Composition
vii. Household Infrastructure
viii. Housing
ix. Socio-economic
x. Women and Children

Each variable was assigned to one of the domains. This assignment painted an interesting picture of the spread of variables.

In making an initial selection of variables, certain principles were considered. First it was the intention of the exercise to try as much as possible to ensure a relatively representative number of variables across all the domains. Secondly, variables with multiple missing values were avoided. It was decided that since this was the first attempt at a national geodemographic system for Nigeria, artificial data generation should be completely avoided to prevent debatable distortions across taxonomic space.

The third principle which was appropriated in the initial selection relates to policy relevance and updatability of the variables. Variables which were deemed to provide useful proxies for assessing some of Nigeria's current policy programmes were considered. Such policy programmes include the National Economic Empowerment and Development Strategy (NEEDS) (NPC, 2004a) and the Universal Basic Education Project (UBE) (UNESCO, 2000).

**Table 5.1: Initial Set of Variables**

| Num | Table/Variable | Domain |
|---|---|---|
| 1 | % of households using agricultural inputs | Agriculture |
| 2 | % of households owning 1 cattle | Agriculture |
| 3 | % of households owning 2-10 cattle | Agriculture |
| 4 | % of households owning 11-20 cattle | Agriculture |
| 5 | % of households owning 21-50 cattle | Agriculture |
| 6 | % of households owning over 50 cattle | Agriculture |
| 7 | % of males aged 0-14 | Demographic |
| 8 | % of males aged 15-59 | Demographic |
| 9 | % of males aged over 60 | Demographic |
| 10 | % of females aged 0-14 | Demographic |
| 11 | % of females aged 15-59 | Demographic |
| 12 | % of females aged over 60 | Demographic |
| 13 | % of total population aged 0-14 | Demographic |
| 14 | % of total population aged 15-59 | Demographic |
| 15 | % of total population aged 60 and above | Demographic |
| 16 | Dependency ratio | Demographic |
| 17 | % of the total population never married | Demographic |
| 18 | % of the total population divorced | Demographic |
| 19 | % of the total population separated | Demographic |
| 20 | % of households with a member receiving a pension | Demographic |
| 21 | Population Density | Demographic |
| 22 | % of the total population uneducated | Education |
| 23 | % of the total population who completed primary education | Education |
| 24 | % of the total population who completed secondary education | Education |
| 25 | % of the total population with post secondary education | Education |
| 26 | % of the total population with head of household uneducated | Education |
| 27 | Adult literacy rate | Education |
| 28 | % of households which spend less than 15 minutes to nearest primary school | Education |
| 29 | % of households which spend less than 15-29 minutes to nearest primary school | Education |
| 30 | % of households which spend less than 15 minutes to nearest secondary school | Education |
| 31 | % of households which spend less than 15-29 minutes to nearest secondary school | Education |
| 32 | % of the total population in public sector employment | Employment |
| 33 | % of the total population in private formal employment | Employment |
| 34 | % of the total population in private informal employment | Employment |
| 35 | % of the total population self employed in agriculture | Employment |
| 36 | % of the total population self employed in other sector | Employment |
| 37 | % of economically active population | Employment |
| 38 | % of unemployed population | Employment |
| 39 | % of self employed population | Employment |
| 40 | % population employed in agriculture | Employment |
| 41 | % population employed in the fishing industry | Employment |

| | | |
|---|---|---|
| 42 | % population employed in the manufacturing sector | Employment |
| 43 | % population employed in the construction sector | Employment |
| 44 | % population employed in trade | Employment |
| 45 | % population employed in the transport sector | Employment |
| 46 | % population employed in public administration | Employment |
| 47 | % population employed in education | Employment |
| 48 | % population employed in health and social work | Employment |
| 49 | % population employed in the services sector | Employment |
| 50 | % of households which spend less than 15 minutes to health facility | Health |
| 51 | % of households which spend less than 15-29minutes to health facility | Health |
| 52 | % of households taking anti-malaria measures | Health |
| 53 | % of households with 1-2 persons | Household Composition |
| 54 | % of households with 3-4 persons | Household Composition |
| 55 | % of households with 5-6 persons | Household Composition |
| 56 | % of households with 7 persons and above | Household Composition |
| 57 | Mean household size | Household Composition |
| 58 | % of households owning a fixed line telephone | Household Infrastructure |
| 59 | % of households owning mobile phones | Household Infrastructure |
| 60 | % of households owning a personal computers | Household Infrastructure |
| 61 | % of households with access to safe source of drinking water | Household Infrastructure |
| 62 | % of households without toilet facility | Household Infrastructure |
| 63 | % of households with access to safe toilet sanitation | Household Infrastructure |
| 64 | % of households which use non-wood fuel for cooking | Household Infrastructure |
| 65 | % of households which use electricity for lighting | Household Infrastructure |
| 66 | % of households which spend less than 15 minutes to source of drinking water | Household Infrastructure |
| 67 | % of households which spend less than 15-29 minutes to source of drinking water | Household Infrastructure |
| 68 | % of households owning a generator | Household Infrastructure |
| 69 | % of occupancy by home ownership | Housing |
| 70 | % of occupancy by rent | Housing |
| 71 | % of subsidized occupancy status | Housing |
| 72 | % of free occupancy status | Housing |
| 73 | % of households built with mud/mud bricks | Housing |
| 74 | % of households built with stone | Housing |
| 75 | % of households built with burnt bricks | Housing |
| 76 | % of households built with cement/sandcrete | Housing |
| 77 | % of households built with wood/bamboo | Housing |
| 78 | % of households built with iron sheets | Housing |
| 79 | % of households built with cardboard | Housing |
| 80 | % of single room housing unit | Housing |
| 81 | % of flats | Housing |
| 82 | % of duplexes | Housing |
| 83 | % of whole buildings | Housing |
| 84 | % of households which seldom find it difficult to satisfy food needs | Socio-economic |
| 85 | % of households which sometimes find it difficult to satisfy food needs | Socio-economic |
| 86 | % of households which often find it difficult to satisfy food needs | Socio-economic |
| 87 | % of households which always find it difficult to satisfy food needs | Socio-economic |
| 88 | % of households which seldom find it difficult to pay school fees | Socio-economic |
| 89 | % of households which sometimes find it difficult to pay school fees | Socio-economic |
| 90 | % of households which often find it difficult to pay school fees | Socio-economic |
| 91 | % of households which always find it difficult to pay school fees | Socio-economic |
| 92 | % of households which seldom find it difficult to pay house rent | Socio-economic |
| 93 | % of households which sometimes find it difficult to pay house rent | Socio-economic |
| 94 | % of households which often find it difficult to pay house rent | Socio-economic |
| 95 | % of households which always find it difficult to pay house rent | Socio-economic |
| 96 | % of households which seldom find it difficult to pay utility bills | Socio-economic |
| 97 | % of households which sometimes find it difficult to pay utility bills | Socio-economic |
| 98 | % of households which often find it difficult to pay utility bills | Socio-economic |
| 99 | % of households which always find it difficult to pay utility bills | Socio-economic |
| 100 | % of households which seldom find it difficult to pay for health care | Socio-economic |
| 101 | % of households which sometimes find it difficult to pay for health care | Socio-economic |
| 102 | % of households which often find it difficult to pay for health care | Socio-economic |
| 103 | % of households which always find it difficult to pay for health care | Socio-economic |
| 104 | % of households owning vehicles | Socio-economic |
| 105 | % of households owning motorcycles | Socio-economic |
| 106 | % of households owning less than 1 ha of land | Socio-economic |
| 107 | % of households owning 1-1.99 ha of land | Socio-economic |
| 108 | % of households owning 2-3.99 ha of land | Socio-economic |
| 109 | % of households owning 4-5.99 ha of land | Socio-economic |
| 110 | % of households owning over 6 ha of land | Socio-economic |
| 111 | % of households which perceive crime level to be much worse | Socio-economic |
| 112 | % of households which perceive crime level to be worse | Socio-economic |
| 113 | % of households which perceive crime level to be the same | Socio-economic |
| 114 | % of households which perceive crime level to be better | Socio-economic |

| | | |
|---|---|---|
| 115 | % of households which perceive crime level to be much better | Socio-economic |
| 116 | % of households which spend 30-59 minutes to nearest food market | Socio-economic |
| 117 | % of households which spend over 60 minutes to nearest food market | Socio-economic |
| 118 | % of households which spend less than 15 minutes to nearest public transport | Socio-economic |
| 119 | % of households which spend 15-29 minutes to nearest public transport | Socio-economic |
| 120 | % of children under 18 years living with their mother only | Women & Children |
| 121 | % of children under 18 years living with their father only | Women & Children |
| 122 | % of children not in school due to teenage pregnancy | Women & Children |
| 123 | % of children not in school due to early marriage | Women & Children |
| 124 | % of vaccinated children | Women & Children |
| 125 | % of children under 5 not breastfed | Women & Children |

Having examined each of the 644 variables based on the principles explained above, an initial selection of 125 was made. Table 5.1 displays the 125 variables selected for further assessment. It is pertinent to say at this point that it is possible for different researchers to select a different list from the above as (Harris et al., 2005) identify that the process of selecting initial variables can be subjective.

## 5.3.2 Variable Reduction Methodology

The first step is to identify which variables would be used for the analysis (Miligan, 1996). In the creation of the Nigerian classification, the 125 variables were reduced in two phases.

The first set of variables was eliminated by analysing the variables on a domain-by-domain basis. For instance all ten (10) variables contained in the education domain were put to test against each other using quantitative and qualitative techniques to reduce them. This process (intra-domain variable reduction) resulted in the first set of reductions for each domain. The next step was to evaluate all variables which survived the first test together irrespective of their domains (i.e. inter-domain variable reduction).

Adopting this approach made the decision-making process easier. For instance, it would have been more tasking to examine a correlation matrix of 125 by 125 variables than it was to examine a matrix of 54 by 54 variables.

As an initial step in the reduction process, the proportion of the population for which each variable accounts was considered. It was considered necessary to avoid variables with small sample sizes. When a variable represents only a small proportion of the population there is the tendency for that variable to be volatile and change rapidly over time. Such a variable would not sustain the longevity of the classification.

**Table 5.2: Variables with the Smallest Sample Sizes**

| Variable | Rank |
|---|---|
| % of households built with cardboard | 1 |
| % of children not in school due to teenage pregnancy | 2 |
| % of households built with stone | 3 |
| % of children not in school due to early marriage | 4 |
| % of households built with iron sheets | 5 |
| % of duplexes | 6 |
| % of households built with wood/bamboo | 7 |
| % of subsidized occupancy status | 8 |
| % of children under 5 not breastfed | 9 |
| % population employed in the fishing industry | 10 |

Table 5.2 above shows the top ten variables exhibiting the smallest sample proportions across LGAs. The percentage of households that built their residences with cardboard represents has the highest rank. An initial look at this variable would make it appear as an interesting variable to include in the analysis but further examination revealed uninteresting patterns. Averagely, the variable represents only 0.48% of the population of each LGA which is approximately 88 people per LGA. These values are averages across all LGAs i.e. assuming all LGA's are represented. In the dataset, there are numerous unrepresented and under-represented zones characterized by the variable.

Another problem that can result from variables with small sample proportions is that they provide little distinctive information for naming and profiling the clusters created from analysis. A solution considered for some of the variables was to merge them (where they fall under the same domain and share a similar base). For instance the variables on children not in school due to early marriage and children not in school due to teenage pregnancy were merged as the shared the same base (100%) and fall under the same domain.

In addition to the examination of variables sample proportions, the skew exhibited by the variables was also considered. It is desirable to include normally distributed variables in the classification but this is often not the case with socio-demographic data. Positively skewed variables were particularly frowned at in the variable selection process. It occurs because of an accumulation of large values at the lower end of the distribution or where there are outliers or extreme values within the distribution.

**Table 5.3: Variables with the Largest Positive Skews**

| Variable | Skew |
|---|---|
| % of households which always find it difficult to pay house rent | 11.25 |
| % of households built with cardboard | 11.15 |
| % of children not in school due to teenage pregnancy | 8.87 |
| Population density | 8.09 |
| % of households built with stone | 7.41 |

The problem with most of these variables is that they identify small proportions of the population hence they concentrate at the lower end of the 0-100% scale. Vickers (2006) found out that variables that would work well within the classification are those that spread in their variation across geographic areas.

Consideration was also given to the relationships between the variables within the dataset. The inclusion of two highly related variables in a clustering algorithm will often result in the repetition of the same information or population behaviour. This can give undue advantage to such behaviour and can also mask other important underlying characteristics existing within the population. Repetition of related information is described as redundancy.

The co-efficient of correlation was used as the statistic for examining the relationship between variables. Correlations can be positive (where the values of the pair of variables increases or decreases in the same direction) or negative (where the value of a variable while its related variable decreases). Both high positive and negative correlations are not desirable within the dataset as they inform redundancy.

Some high correlations were observed within the dataset. In order to illustrate these different types of correlations a domain-by-domain data reduction process was first embarked upon. Table 5.4 shows the correlation matrix of the education domain.

The orange colours indicate high positive correlations while the blue colours identify high negative correlations. People who are uneducated have a high negative correlation with people completed secondary education. The reason for this is because these variables share the same denominator and an individual can only belong to one of the two categories.

**Table 5.4: Correlation Matrix of Education Variables**

| Variable | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | E10 |
|---|---|---|---|---|---|---|---|---|---|---|
| E1 | X | -0.77 | -0.82 | -0.65 | 0.95 | -0.77 | -0.01 | -0.19 | -0.25 | -0.31 |
| E2 | | X | 0.47 | 0.24 | -0.72 | 0.57 | -0.15 | 0.17 | 0.06 | 0.15 |
| E3 | | | X | 0.71 | -0.83 | 0.73 | 0.05 | 0.17 | 0.32 | 0.32 |
| E4 | | | | X | -0.67 | 0.62 | 0.19 | 0.10 | 0.44 | 0.34 |
| E5 | | | | | X | -0.78 | 0.01 | -0.21 | -0.27 | -0.30 |
| E6 | | | | | | X | 0.06 | 0.18 | 0.30 | 0.32 |
| E7 | | | | | | | X | -0.51 | 0.63 | 0.27 |
| E8 | | | | | | | | X | -0.27 | 0.27 |
| E9 | | | | | | | | | X | 0.22 |
| E10 | | | | | | | | | | X |

where

E1 = Uneducated Population

E2 = Population that completed primary school

E3 = Population that completed secondary school

E4 = Population with post-secondary education

E5 = Head of household uneducated

E6 = Adult literacy rate

E7 = Time to nearest primary school – less than 15 minutes

E8 = Time to nearest primary school – 15-29 minutes

E9 = Time to nearest secondary school – less than 15 minutes

E8 = Time to nearest secondary school – 15-29 minutes

Having examined the dataset on a domain-by-domain basis leading to the first set of data reductions and mergers, all the domains were also examined together. This allowed for the evaluation of inter-domain cross-correlations and the identification of another type of relationship between variables. Table 5.5 shows some variables within different domains which obviously do not share the same denominators but are related.

**Table 5.5: Relationship between Variables in Different Domains**

| Variable | Domain | Variable | Domain | Correlation | Redundancy (%) |
|---|---|---|---|---|---|
| Owns 2-10 cattle | A | Uneducated Population | E | 0.82 | 67 |
| Never married | D | Uneducated Population | E | -0.78 | 61 |
| Post-secondary education | E | Ownership of mobile phones | S | 0.78 | 60 |
| Renting a house | H | Difficulty in paying house rent-sometimes | S | 0.77 | 59 |
| Owns 2-10 cattle | A | Uneducated household head | E | 0.75 | 57 |
| Uneducated household head | E | Built with cement/sandcrete | HO | -0.75 | 56 |
| Post-secondary education | E | Non-wood fuel for cooking | HI | 0.74 | 55 |
| Private-formal employment | EM | Ownership of mobile phones | HI | 0.74 | 54 |
| Post-secondary education | E | Renting a house | HO | 0.72 | 52 |
| Age 0-14 | D | Completed secondary education | E | -0.71 | 51 |
| Never married | D | Uneducated household head | E | -0.71 | 51 |

where

A = Agriculture HC = Household Composition

D = Demographic HI = Household Infrastructure

E = Education HO = Housing

EM = Employment S = Socio-economic

H = Health WC = Women and Children

The relationships existing between the variables in Table 5.5 are due to the capability of one variable being able to explain the variation existing within the other. In other words, the characteristics inherent in one variable can also be caused by the presence of the other variable. People who own 2-10 cattle for instance have a high positive correlation of 0.82 with people who are uneducated. Such people are involved in cattle rearing. Some are nomads who are less involved with education. Indeed a good number of these people are not educated. The positive nature of the correlation indicates that there is a higher tendency for people who are not educated and own cattle to live within the same locality.

The percentage of the population aged 0-14 can inform 51% of the variability existing within the percentage of people who have completed secondary education. The difference here is that these variables are negatively correlated. To explain this, we can consider the system of education operated in Nigeria. The country operates a 6-3-3-4 system. Primary education begins at age six. A child has to spend at least six years in primary school before moving on to secondary school. Secondary education consists of junior secondary school, which is a minimum of three years, and senior secondary school, which is another three years. University education is a minimum of four years. If a child starts primary school at age six and spends (6+3+3) years before completing secondary school, the individual would be 18 years of age. So the average age expected for children who would be expected to complete secondary education is 18 years. This is why there is a negative relationship between the two variables even though one can explain a high percentage of the variance of the other. Another way to look at it is that the population within one of the variables cannot at the same time be within the other variable. They are mutually exclusive.

Following the examination of correlations, the manner in which variables vary across the geographical units (LGAs) was considered. It is important that variables spread well for them to improve the clustering algorithm. A useful statistic for measuring the geographic variation of variables is the standard deviation.

**Table 5.6: Variables with High Standard Deviations**

| Variable | Mean | SD |
|---|---|---|
| % of households built with cement/sandcrete | 40.14 | 33.22 |
| % of households built with mud/mud bricks | 54.45 | 31.80 |
| % of households using agricultural inputs | 40.70 | 30.74 |
| % of households with access to safe toilet sanitation | 50.68 | 28.96 |
| % of households owning motorcycles | 31.98 | 28.56 |
| % of households with access to safe source of drinking water | 45.63 | 28.56 |
| % of single room housing unit | 66.52 | 28.13 |
| % of whole buildings | 26.87 | 27.98 |
| % of the total population uneducated | 39.60 | 26.55 |
| % of households owning less than 1 ha of land | 26.42 | 26.09 |

In table 5.6, the percentage of people owning motorcycles combines a mean of 31.98 with a standard deviation of 28.56. We can deduce from this that two thirds of the values of the variable for LGA's lie between 3.42 and 60.55.

**Table 5.7: Variables with Low Standard Deviations**

| Variable | Mean | SD |
|---|---|---|
| Dependency ratio | 0.79 | 0.21 |
| % of households built with cardboard | 0.03 | 0.31 |
| % of the total population divorced | 0.48 | 0.66 |
| Mean household size | 4.97 | 1.12 |
| % of duplexes | 0.36 | 1.13 |
| % of the total population separated | 1.07 | 1.39 |
| % of households owning over 50 cattle | 0.64 | 1.42 |
| % population employed in health and social work | 1.53 | 1.63 |
| % of the total population in private formal employment | 1.06 | 1.72 |
| % of households owning a personal computers | 0.90 | 1.94 |

Variables with large standard deviations will prove more useful than those with lower values. This is because they present better distinctions between areas (Harris et al., 2005). However when dealing a variable with low standard deviation value, it is important to consider the sample size of the variable across the LGAs. Some of these variables can be merged with other variables and renamed if they share the same base and are within the same domain. A variable that was created in this manner is cattle ownership. The percentage of people owning over 50 cattle in table 5.6 shows little variation across the geographic areas with a standard deviation of 1.42. This variable was merged with three other variables within the same domain and sharing the same denominator. These include the percentage of people owning 2-10 cattle, 11-20 cattle and 21-50 cattle. By merging these variables, a new variable was created and called 'cattle ownership'. The standard deviation for the new variable is 22.19. This is a significant increase when compared with a value of 1.42.

One of the intentions of this classification system is the hope that it will prove useful for decision makers. As a result, the significance of the variables within the context of policy making and implementation in Nigeria was also considered. Priority was given to poverty and basic welfare of individuals. Indicators that measure issues connected to poverty directly or indirectly often motivate action from the local, state and national government. These issues were borne in mind when making a choice of variables for the classification. A theoretical underpinning in the course of the selection process was therefore to consider how well the variable could be linked with policy formulation issues especially as they relate to the Millennium Development Goals (MDGs).

Making decisions about the final choice of variables is an important activity. It may be practically impossible for two different classification developers to come up with a list of variables which are 100% the same. This is because along the course of the selection process, a number of subjective decisions are made. For instance, the variable on percentage of flats initially seemed to be very interesting. However, it was eventually excluded because of its behaviour during the clustering process because it masked some underlying distinctive features of the LGAs. The series of discussions reveal that the processes of trimming down the initial list for the Nigerian classification required detailed theoretical and analytical work. It is impracticable to describe every detail of the magnitude of work that was done in the reduction process. This section has therefore been able to reveal what could be considered a tip of the ice-berg.

**Figure 5.2: Variable Selection Methodologies**

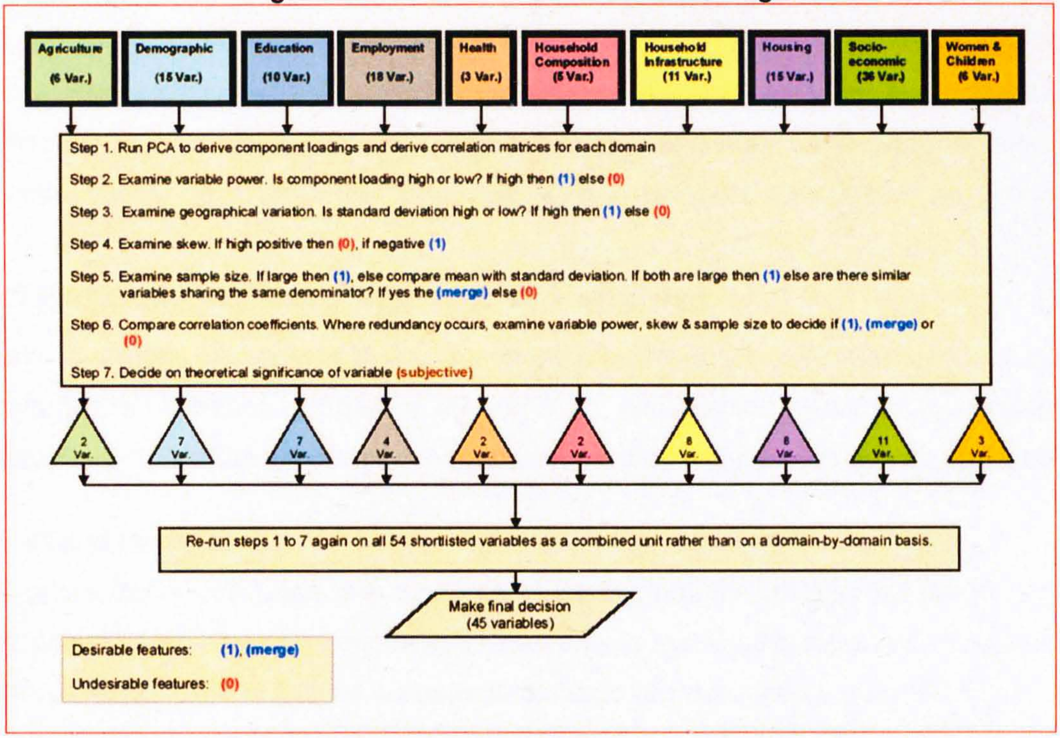



Figure 5.2 above is a flowchart depicting some of the methods adopted in the variable selection process. It is imperative to state that the chart is not static. In Appendix A1, a tabular illustration is given to elucidate some of the decisions made for each variable. Rejected variables have their cells coloured in orange. As discussed above, some variables were merged and in some cases given new names. These variables which are included have their cells coloured in yellow. Colour blue is used for variables which are also included without merger or change of name.

At the start of the exercise, there were 125 variables spread across ten domains. The socio-economic domain had the largest representation of 29% while the health domain represented just 2%, the least of all variables. The range between the highest and lowest domains in percentage was 27% (i.e. 29% - 2%). At the end of the first phase of analytical work, the socio-economic domain still characterized by a high 20% was greater than the agriculture, health and household composition domains (the least represented) by 16%. Following a new set of reductions, the household-infrastructure domain moved into the lead with 18%. Agriculture, health and household composition remained the least represented with 4%. By the end of the exercise, the range between the highest represented domain and the least represented was reduced by 48%.

The socio-economic domain showed the greatest variation in changing number of variables moving from an initial value of 36 variables to 11 and finally to 6 variables. Health domain had the least variation with the initial 3 variables reduced to 2. In summary, the exercise commenced with 125 variables which were reduced to 54. These were further reduced to 45 variables. The variables employed in the clustering process are described in section 5.3.3.

### 5.3.3 Definition of the Final List of Variables

Complex decisions were made during the variable selection process. The following sections describe the ten domains identified in section 5.3.1; explains the relevance of variables they encapsulate and elucidates some of the reasons for their inclusion, exclusion or merger.

***Agricultural Variables***

The agricultural sector plays a major role in its contribution to Nigeria's socio-economic development cannot be over-emphasized. Not only is agriculture important for food supply, it is also a sector which employs a major percentage of the country's population. Agriculture also has it's relevance to the manufacturing sector. A number of agricultural products like cocoa, rice, wheat, and rubber serve as raw materials for the production of some other products. In addition, agriculture is a key source of foreign exchange via the production of cash crops. If the first in list of MDGs (eradication of extreme hunger and poverty) is to be met in Nigeria, the agricultural sector has to be strengthened and made to compete well with other sectors especially the oil and gas sector.

Variables within this domain can also present interesting geographic information on urban/rural divide. In addition, it can be indicative of areas of food deserts, and to some extent malnutrition. It can also guide agricultural extension officers in making decisions.

The use agricultural inputs are an important agricultural indicator. It can help distinguish between areas where modern agricultural methods are currently practiced. The variable was included as it demonstrates a good variation across areas with large sample sizes.

Cattle ownership in Nigeria is usually associated with nomads. However, with increasing urbanisation, nomadic farming is on the decline. For the purpose of this classification, people who own single cattle were excluded as the variable demonstrates very little variation. In addition the variable does not contribute any theoretical information. Variables on people who own between two and over fifty cattle were merged to create a single variable called 'cattle ownership'. This helped increase the sample size and also their variation across geographic space. Cattle ownership is also indicative of households who rear domestic animals for commercial purposes.

***Demographic Variables***

Demographic studies explore issues relating to the size, structure, dispersement and development of human populations (Bloom et al., 1998). A lot of emphasis is therefore paid on understanding how age, sex distributions and marital status vary across geographical space. In general such knowledge is an important primary step in differentiating between various population types. Some of the MDGs and targets focus on specific demographic groups. For instance the goal of ensuring gender equality has a bias for female population groups.

Age is indicative of the behaviour of individuals during different periods in their life cycle. The variation of age across geographical areas can present a distinction between urban and rural characteristic of areas. Younger people would normally prefer residing in urban centres. The key reason for the migration and concentration of younger people in urban areas is the perception that more opportunities especially jobs exist within in these areas. Age can also help to shed light to health related issues particularly sexual health (House, 1990). Indeed numerous health-related studies often control for age by standardising (McLennan, 1997) before analysis.

This study employs three age groups. These include age 0-14, age 15-59 and age 60 and above. These age groups help to identify certain groups of population. Age 0-14 for instance can serve as a proxy for infants and young students. The maximum level of education expected for this age group is the Junior Secondary School Certificate (JSSC) so they are predominantly a dependent population group. Age group 15-59 distinguishes the group of population, which are predominantly economically active while the populations above 60 years constitute people who normally are recently retired and the aged.

The relationship between the economically active population and the economically inactive population is defined by the dependency ratio. In other words, the dependency ratio is the relationship ratio of the population aged 15-59 to the population aged 0-14 and over 60 years. The national figure is 0.8 indicating that less than less than one person is dependent on an economically–active person. The variable was not included, as it would have added redundant information with the age group 0-14 due to high correlation.

Marital information can be a very useful proxy in fertility studies. The geographical variation of marital status does have an implication on growth. Marriage can have a financial implication on individuals (Morgan, 1991). In addition divorce rates do have economic and social impacts (Morgan, 1991). In a country like Nigeria where divorce/separation rates are not popular because of religious and cultural perceptions, a discrimination of the prevalence of such demographic variable within certain neighbourhood types may reveal some other underlying phenomena.

The variable on unmarried population groups is not only indicative of people who are considered too young for marriage but also people who are of marital age but do not have the economic means to cope with it. Divorced and separated couples were merged into a single variable called separated couples. The variables contain similar information and also combine the same base.

Retirement often results in a change of lifestyle. There is a general trend in Nigeria for people to retire back to their hometowns after active service. In most cases these hometowns are located in the rural areas and have agriculture as the mainstay of their economy. Some people however do not return back to their original locality. The pensioners group was included because it is important for planning and development, as they constitute a major percentage of the economically inactive population.

Population density is defined as the number of people per square area of land. This variable is of significant importance as a distinguishing factor between urban and rural areas. Generally urban areas tend to attract significantly higher population while their rural counterparts suffer out-migration, which in turn reduces their population density. High population density can in turn have its impacts on housing and other social conditions like crime and public amenities. It is not uncommon to find slums in some of the world's largest cities (UN-Habitat, 2003). In Nigeria, densely populated areas are characteristically prone to high prevalence of poverty because of the pressure on infrastructure and high rate of unemployment.

***Education Variables***

Second on the list of the United Nations Millennium Development Goals (MDG) is the task of achieving universal education primary education by 2015. Education plays a major role in the measure of poverty prevalence. The incidence of poverty decreases with increasing level of education in Nigeria (NBS, 2006a). Information from education can also provide a better understanding on how to plan properly for investment in that sector.

Another unique reason why the education domain is considered important for a developing country like Nigeria is in its fight against ignorance. The country has in the past experienced long years of military rule. Following the transition to democratic governance in 1999, Nigeria joined the league of egalitarian societies. Democratic governance in Nigeria has been a learning process. Elections are sometimes plagued by widespread violent clashes initiated by politicians who use issues like religion, tribalism and money to lure less educated youths into committing these crimes. The argument here is that an educated individual is empowered, as knowledge is power. An educated person is more likely to exercise his/her democratic right without contemplating violence. Well educated people consider they have a bright future ahead of them and are less likely to take violent risks.

The percentage of people who lack any form of education was considered but later dropped as it was significantly positively skewed. In addition, it showed high correlations with the population of heads of households without education.

Two variables existed for education completion rates. These include those who had completed primary education and those who had completed secondary education. The populations who have completed secondary school ranks third in the list of variables that power the dataset with a loading of 0.88 in principal component 1 as opposed to a loading of 0.54 for those who completed primary education. Apart from this, those who have completed secondary education also add to the population of the economically active population that makes it even more useful. As a result, the percentage of the population indicating secondary school completion rates was chosen.

Post secondary education is no doubt an important tool for of manpower development. In addition it is a variable that is indicative of areas of potentially higher education. The variable passed the first phase of reduction tests. However in the second stage it showed very strong correlations with those renting a house and those who use non-wood fuel for cooking. It was therefore dropped.

It has been identified in the earlier section that education plays a major role in the measure of poverty. The role of the household head in Nigeria is very important. Traditionally, the head of the household is expected to contribute a very great proportion of resources to the well being of members. As a result, the income of the head of the household plays a major role in sustaining the family. Less educated household heads are less likely to generate high incomes (NBS, 2006a) and therefore households where the head of the household is not educated are more poverty prone.

Adult literacy is indicative of areas where older people still crave for knowledge. The variable demonstrates a high variation across areas and is included in the final variable list.

Access to primary school, is defined by the population spending a maximum of thirty minutes to the nearest primary school. In order to increase the sample size of this variable, the percentage population spending less than 15 minutes to the nearest primary school was merged with the percentage spending 15-30 minutes. Access to secondary school is also defined by the same parameters but in this case it is thirty minutes to the nearest secondary school. Access can help delineate areas where investment is needed in education. In addition it can reveal neighbourhoods where there are children of school age are not attending school due to certain underlying reasons. Access to primary school was included but access to secondary school was excluded due to strong correlation with access to health.

***Employment Variables***

Employment is usually the key source of income for various sections of the population. There is very stiff demand for employment in Nigeria with incomparable supply (NBS, 2006a). Unemployment is a phenomenon that aggravates poverty. Two principal sectors of employment exist in Nigeria. The public sector employs people who work mainly in government departments while the private sector is for non-government workers. Achieving full and productive employment has also been identified as a key target by the United Nations (U.N) if countries are to achieve the first MDG of poverty reduction and the elimination of extreme hunger.

Public and private sector employment variables were not included in the final list of selected variables because they demonstrate significantly low variation across areas. In addition they correlate strongly with some other sector-specific variables. For instance, private informal employment contains redundant information with the self-employed population.

Within the category of self-employed population are those who are self employed in agriculture and in other sectors. The self-employed population was used because of its greater variation across space. The variable self-employed in agriculture is too specific and shows very little variation across areas.

A person is classed as economically active when such an individual engages or makes attempt to engage in economically productive work. In other words, the economically active population forms the labour force (NBS, 2006a). This variable is a variant, which combines a proportion of the employed and unemployed population. The variable showed a reliable variation across areas and is also indicative of the country's manpower. It is included in the final list.

The unemployed population defines the percentage of economically active population that is out of work. This variable was dropped because of the little variation it shows across the LGAs and the fact that by including information on the employed population, we pseudo-include information about the unemployed.

Information about the exact type of employment was also considered for inclusion. A composite variable was formed by merging the population employed in the agricultural sector with the fishing sector, as both constitute agricultural work. Employment in trade showed a high correlation with employment in agriculture and was therefore excluded. Transport sector employment can be indicative of areas where there is a reasonable level of public transportation. Unlike Britain and most of the developed world, the private sector has a higher stake and involvement in public transportation in Nigeria.

Other variables like employment in the manufacturing, construction, public administration and social work showed very insignificant variation and were considered too specific in their definition and so were not included.

***Health Variables***

Good health is an important and desirable feature of all persons. Information on health can be useful in understanding the prevalence and incidence of morbidity within the population. Regional variations in the receipt of quality and timely health care are also a useful measure of deprivation. The importance of health is further demonstrated by the fact that three of the MDGs are health related.

The population that has access to health is defined as those who travel a maximum of thirty minutes to the nearest health-care facility. The variable was created by merging the variable defining the population spending less than 15 minutes to the nearest health facility with those spending between 15 and 29 minutes. The variable is included in the final list.

The combat against malaria is one of the priorities of the ministry of health in Nigeria. Malaria results in the deaths of about 20% of children on a yearly basis in Africa (UNICEF, 2004) with a good number of these being infants. The World Health Organization (WHO) and UNICEF to help mitigate the economic and developmental loss caused by malaria have initiated numerous programmes. A quote taken from a report published by UNICEF in 2004 on malaria, child death and poverty reads as follows:

*'It costs Africa US $10 billion to $12 billion every year in lost Gross Domestic Product – even though it could be controlled for a fraction of that'* (UNICEF, 2004 p. 2).

Taking preventive measures against malaria is therefore not only important for health but also has significant economic benefits. This variable is included in the final list.

***Household Composition Variables***

There are economic implications for households in terms of their composition. Larger households are usually able to generate substantial funds to meet some of their needs. The downside to this is that large households also contribute to over-crowding and pressure on facilities.

Where 1 to 2 people live in residential household informs sparsely populated neighbourhoods. The variable is included as it has a large variation across areas and representative sample sizes. Household sizes of 3 to 4 people, 5 to 6 people and over 7 people show high correlations with the mean household size and are not included

Mean household size also acts as a measure of density. Nigeria is the most populous country in Africa and the control of population growth has always been high on the agenda of successive governments. Mean household size can be related to reproductive health by discriminating between areas where family planning measures are used. The variable is included in the final list.

***Household Infrastructure Variables***

Household infrastructure serves as a proxy for measuring quality of life. Basic infrastructure relating to food, water and sanitation are primary to many Nigerians. Other secondary infrastructure like access to telecommunications also provides a means of understanding how sections of the population are diffusing innovation.

The spread of mobile phone technology in Nigeria has been unprecedented since the introduction of GSM technology. The growth started with urban centres and is moving at a fast pace towards rural areas. The advent of mobile phone use also resulted in significant economic boom like privatisation and the creation of jobs within the sector. Like mobile phones, infrastructure for fixed line phones has also been developed steadily, though not as rapid and prevalent as mobile phones.

Effective telecommunication is very important for business. It also saves travel trips and time, which has an indirect positive consequence on the environment.

The ownership of mobile phones has a better variation across geographical areas with better sample sizes as well. It is preferred to the ownership of fixed line phones. This variable is of particular importance for the 8$^{th}$ MDG which seeks to ensure the development of a global partnership for development.

Water is very important for sustaining life. Access to water is defined for people who spend less than 30 minutes to the nearest water source. Not all sources of water though are drinkable. Some sources are not as safe as others and can only be used for limited domestic or industrial use. An understanding of the variation of this variable also informs on the dispersal of population prone to water-borne diseases.

Another important environmental indicator identified in the list for the 7$^{th}$ MDG is sanitation. People with access to toilet facilities and those with access to safe sanitation facilities were included as they may provide useful information on public health. They both show good variation across the LGAs and have representative sample sizes.

Areas where wood is still used for cooking is expected to be less urban and experience high deforestation. The variable is included in the final list as it has a very high loading (0.82) in principal component 1. It is also well represented across the dataset with good sample sizes.

The energy sector in Nigeria has witnessed numerous challenges over the years leading to erratic power supply. The poor supply of electricity also contributes to the decline of small scale enterprises in the country. Ownership of power generating sets is therefore common practice in Nigeria. It was initially thought that this variable would show some reliability for inclusion in the classification but preliminary analysis showed it to be very highly skewed with very low variation within the distribution. The variable was therefore rejected.

### *Housing Variables*

Different forms of shelter or accommodation exist in Nigeria. It is important to be able to discriminate between these forms of residential unit as they help define identity. An understanding of housing can also reveal problems within the context of housing demand and supply. The variables explored within this domain pertain to occupancy mode, building materials and building type.

The percentage of people who own houses showed very good variation across the distribution and has reliable sample sizes. In combination with the percentage of people renting houses, we can get a good picture of housing demand. These two variables alongside the percentage of people in free accommodation were included. The percentage population in subsidized accommodation was however excluded due to low variation and very low sample sizes.

The nature of building materials is and important discriminator of regional identity in Nigeria. It can also be indicative of the degree of urbanisation. Building materials can also be a useful measure of deprivation when considered alongside household infrastructure. The problem with some of the variables within this category is that they show little variation.

Houses built with cement/sandcrete were merged with houses built with mud bricks because they show high correlations and share the same denominator. The new variable is called 'built with cement/mud bricks'. The percentage of houses built with burnt bricks was also included while houses built with bamboo, cardboard and iron sheets were excluded mainly due to very low variation across areas.

### *Socio-economic Variables*

The most difficult challenge faced by the Nigerian government is poverty reduction (Ogunbodede, 2006). Instability in the political system has led to inconsistency in pursuing various socio-economic policies. As a result sustainable economic development is lacking. This is also compounded by activities in the petroleum sector. Over the years since its first discovery in the 1970's, petroleum has caused attention to be diverted away from other sectors which used to be principal contributors to the economy in terms of job creation and sources of foreign exchange. Examples of the sectors which have witnessed steady decline include the agriculture and manufacturing sector.

Poverty breeds vices such as crime and corruption. The socio-economic welfare of the population is therefore of critical interest to any government of Nigeria. Geographical variations and inequalities also exist in terms of the level of socio-economic conditions of people. The variables explored within the category of socio-economics are discussed below.

This variable on difficulty in meeting basic needs represents people who reported that they always have difficulty in satisfying their food needs, paying school fees, paying house rent, paying utility bills and paying for health care. These groups of people are in dire need of aid.

The ownership of a private means of transportation is also a useful proxy for socio-economic status. The reason for this is simple. Basic needs come before secondary issues like car ownership so people who have extra to spend on ever-rising cost of fuel are on a higher socio-economic scale than those who can not. Alongside vehicle ownership motorcycle ownership was also included in the final list because of very high variation across areas

Land is an asset with value. The percentage of the population owning land over 6 hectares was included for its significant variation and the fact hat it is also a traditional measure of socio-economic status in Nigeria. Variables on 'land less than 6 hectares' were rejected. They showed little variation within the dataset.

Secured communities are very important to a lot of Nigerians and indeed to policy makers. The security condition in the country is linked with the attraction of foreign private investment. Insecurity does not encourage economic stability. People want to feel free, work hard and enjoy the fruits of their labour. Self perceived level of crime was considered in this category. The percentage of people who reported crime level as much better was merged with the percentage that reported that crime level is better to create a new variable called 'improved security'. The variables on people who described the situation as worse or stagnant were excluded as they are considered to have been pseudo included with people who describe the situation as better.

The variable on access to transportation was constructed by merging the percentage of people who spend less than 15 minutes to the nearest transportation with the percentage that spend 15-29 minutes to the nearest transportation. Time to nearest food market showed a strong relationship with time to nearest transportation so it was rejected from the final list.

***Women and Children Variables***

The rights and role of women in today's world has consistently being an issue of significance. On a global scale, the rights of women present an indicator for measuring well-being. While a few people may argue that concerns about the rights of women and children should be focused on societies where conflicts and religion have been used as platforms for undermining these rights, this may not be entirely true (IPS, 1998).

A key policy issue in Nigeria is the drive to reduce maternal and child deaths (NBS, 2006a). These two issues are also high on the UN agenda as reflected by the 4$^{th}$ MDG. Significant proportion of these deaths can be prevented by increasing investment in health services, female education and greater political commitment. Some of these decisions are hinged on better understanding of the geographical variations of indicators which prescribe information on women and children development. The variables considered under this topic are discussed below.

The variable on children living with single parents was created by merging the percentage of children living with their father only with the percentage of children living with their mother only. The variable which is included in final list can serve as an indicator of families that may be in need of some form of support.

Early marriage and teenage pregnancy are two important indicators particularly for young girls who are still of school age. These variables are very useful indicators for targeting and encouraging female education. In many scenarios, young girls with these features do not attend school because they are either married or pregnant. These two issues were examined critically to decide whether to merge them or not. It was decided that they should be merged because they share the same denominator and this would assist their statistical reliability. They were merged to create a new variable called 'Early marriage and teenage pregnancy'.

Child vaccination is important for the prevention of ailments which contribute to child mortality. In Nigeria, children are vaccinated against the risk of contracting yellow fever, polio, malaria, typhoid fever, meningitis and hepatitis to mention a few. However, religious beliefs and ignorance in terms of education causes some parents to exclude their children or wards from these important activities. Another cause of exclusions is poor information communication. This variable is considered important in theoretical terms and also behaves well statistically. It has a good variation, a negative skew and is representative of the population in terms of the sample size across areas. It is included in the final list.

The percentage of children under five years old who are not breastfed was initially considered for inclusion but with very little sample sizes and poor variation, high positive skew and a component loading of -0.06, the variable was rejected.

**Table 5.8: The Final List of 45 Variables**

| Variable | Domain |
|---|---|
| Use of agricultural inputs | Agriculture |
| Owns cattle | Agriculture |
| Age 0-14 | Demographic |
| Age 15-59 | Demographic |
| Age 60 and over | Demographic |
| Never married | Demographic |
| At least one pensioner | Demographic |
| Population density | Demographic |
| Separated couples | Demographic |
| Completed secondary education | Education |
| Head of household uneducated | Education |
| Adult literacy rate | Education |
| Access to primary school | Education |
| Economically active population | Employment |
| Self employed | Employment |
| Employment in the transport sector | Employment |
| Employment in agriculture | Employment |
| Taking anti malaria measures | Health |
| Access to health | Health |
| Household size 1-2 persons | Household composition |
| Mean household size | Household composition |
| Ownership of mobile phone | Household Infrastructure |
| Ownership of personal computer | Household Infrastructure |
| Access to water-safe source | Household Infrastructure |
| No toilet facility | Household Infrastructure |
| Safe toilet sanitation | Household Infrastructure |
| Non-wood fuel for cooking | Household Infrastructure |
| Lighting energy-mains electricity | Household Infrastructure |
| Access to water supply | Household Infrastructure |
| Own a house | Housing |
| Renting a house | Housing |
| Free accommodation | Housing |
| Built with burnt bricks | Housing |
| Single room | Housing |
| Duplex | Housing |
| Built with cement/mud brick | Housing |
| Vehicle ownership | Socio-economic |
| Motorcycle ownership | Socio-economic |
| Size of land-over 6 hectares | Socio-economic |
| Improved security | Socio-economic |
| Access to transportation | Socio-economic |
| Difficulty in meeting basic needs | Socio-economic |
| Children living with single parents | Women and Children |
| Early marriage and teenage pregnancy | Women and Children |
| Vaccinated | Women and Children |

The final list of variables is given in the table 5.8. These are variables that are further manipulated and prepared for the clustering algorithm.

## 5.4 Clustering Processes

Following the selection of the input variables, the next step was to prepare the data for classification. All the variables for the entire 774 LGA's were assembled into a single database and manually checked for any errors. Once this was done, the clustering process could commence.

### 5.4.1 Data Transformation and Standardisation

The next phase was to explore the scales in which the variables have been provided. Scale in this context refers to the unit of measurement like percentage or ratio. It is inappropriate to run a clustering algorithm on a dataset which consists of variables of different scales. The reason for this is that the nature of the scale may cause undue advantage to be given to certain variables while others suffer.

Clustering variables in different units will only result in incorrect results. Methods of normalisation help in re-scaling distributions. However, prior to any form of normalisation, it is good to transform the variables. Transformation methods are discussed extensively in section 4.5 in the previous chapter.

Vickers (2006) found that log transformation has the ability to cope well positive skew which showed a strong presence in the Nigerian dataset. In addition, it was a preferred transformation technique in this exercise because of its ability to retain the dimensions of the original dataset (Vickers, 2006).

The different methods of normalisation were explored with the data and z-score standardisation was adopted as the method for putting variables for the classification system on a common scale. The method has been used because of its ability to maintain a mean of zero for the standardised values and a standard deviation of one. With a mean of zero, distortions stemming form the central value of each variable can be avoided.

## 5.4.2 Clustering Algorithm

The taxonomic analysis of the LGA's seeks to identify homogenous groups within the dataset. This sort of analysis has never been conducted in Nigeria. This makes it a lot more difficult to place the results from this analysis on a yard-stick. Indeed how can we be sure there are clusters of LGA's in Nigeria?

A variety of cluster analysis methods have been explained in detail in section 4.7. Hierarchical methods available in were initially considered for the analysis; however the K-means clustering method was found to be less computer-intensive and was therefore adopted. In addition to being less computer intensive, the method generates a number of associated tables including the ANOVA table which proves useful for further evaluation of the clusters produced.

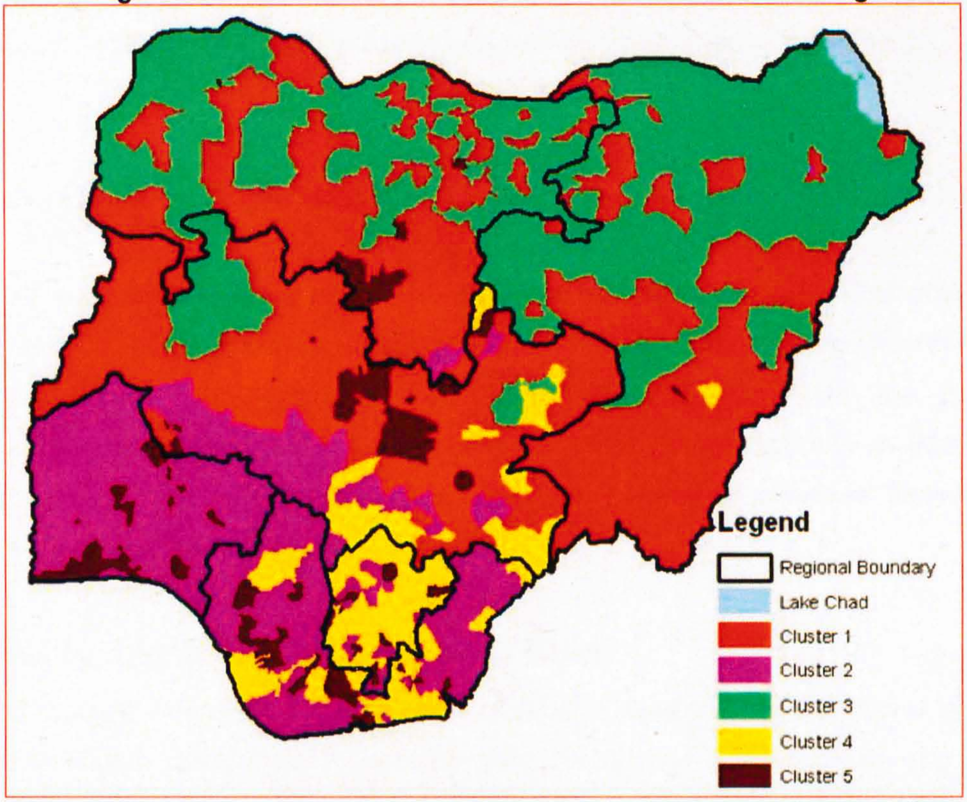
K-means works by allowing the user to specify the desired number of clusters in advance to running the algorithm. A technique therefore has to be adopted to identify the most suitable number of clusters. Section 5.4.3 sheds more light on how this was done.

Of paramount initial importance was the need to identify if all the variables work well within the algorithm without clouding differentiating features of the LGA's. Nigeria is known to comprise of six (6) geo-political regions. What were not known prior to these analyses were the inter-regional similarities and dissimilarities existing within the National system. If the algorithm produced clusters that were highly 'skewed' towards the regional divide (i.e. significantly portrayed what was already known), it would suggest two things:

1. The algorithm works for the Nigerian data
2. The data would have to be checked again and further manipulated before re-running the clustering algorithm.

The algorithm was therefore run for two to fifteen clusters on the log-transformed z-scores for each variable and LGA. It is important to mention that at this stage, there were 46 variables. Each of the 2 to 15 clusters was mapped for visualisation purposes. It was immediately evident that the algorithm was working but there was a problem. There was a very strong presence of the **North-South-Middle belt** divide in the results as depicted in Figure 5.3.

**Figure 5.3: North-South-Middle-belt Divide in Initial Clustering**




Between the two cluster solution and the seven cluster solution, this problem was very pronounced. From the eight cluster solution the problem was slightly ameliorated but at this stage, there were two major problems. First it would prove more difficult to visualise 8 clusters. More importantly however, the cluster memberships was significantly impaired (see table 5.9) with a high range of 24%. Cluster 4 with only three LGA's would represent less than 1% of the entire system. A good solution should maintain a relative balance in the proportional allocation of zones to clusters (Gordon, 1999; Everitt, 2001; Harris et al., 2005).

**Table 5.9: Cluster Membership Problem for the Eight Cluster Solution**

| Cluster | Number of LGA's | Percentage |
|---|---|---|
| 1 | 185 | 24 |
| 2 | 110 | 14 |
| 3 | 48 | 6 |
| 4 | 3 | 0 |
| 5 | 124 | 16 |
| 6 | 11 | 1 |
| 7 | 185 | 24 |
| 8 | 108 | 14 |

One key contributor to this problem was the presence of some strong outliers in the z-score used for the clustering algorithm. It was therefore decided that a range should be specified for the scores. Based on the frequency distribution of the z-scores, they were capped within a range of -3 and 3. So every value greater than 3 or less than -3 was rounded down/up to the benchmark.

The algorithms were re-run again for 2 to 15 clusters. The problem was minimized but still quite evident. At this point it was decided that (n-1) variables should be clustered.

where
n (number of variables) at this stage was 46

This meant each of the 46 variables was excluded for each 2 to 15 cluster solutions that were examined. This process was very tedious but proved helpful. After careful examination, it was discovered that by excluding the variable on 'Flats', the geo-political regional divide greatly diminished. It was at this stage that the difficult but inevitable decision to exclude the variable 'Flats' was made. The total number of variables included in the algorithm was therefore reduced to.

## 5.4.3 Deploying a Cluster Stopping Rule

Extensive thought was given to the idea of creating a hierarchy. It is common practice for many classification developers to develop their systems in a hierarchical manner. The choice in this research was to create a three tier hierarchy for Nigeria. This decision was made on a number of fronts. First, the choice of the hierarchy to use in any form of analysis is important. The top tier of the classification would be useful for visualisation but may not unearth the variations between states. This variation would be necessary for extrapolating surveys and analysis done at state level. For instance state governments may need to deploy resources across their LGAs and may want to use data available at state level only.

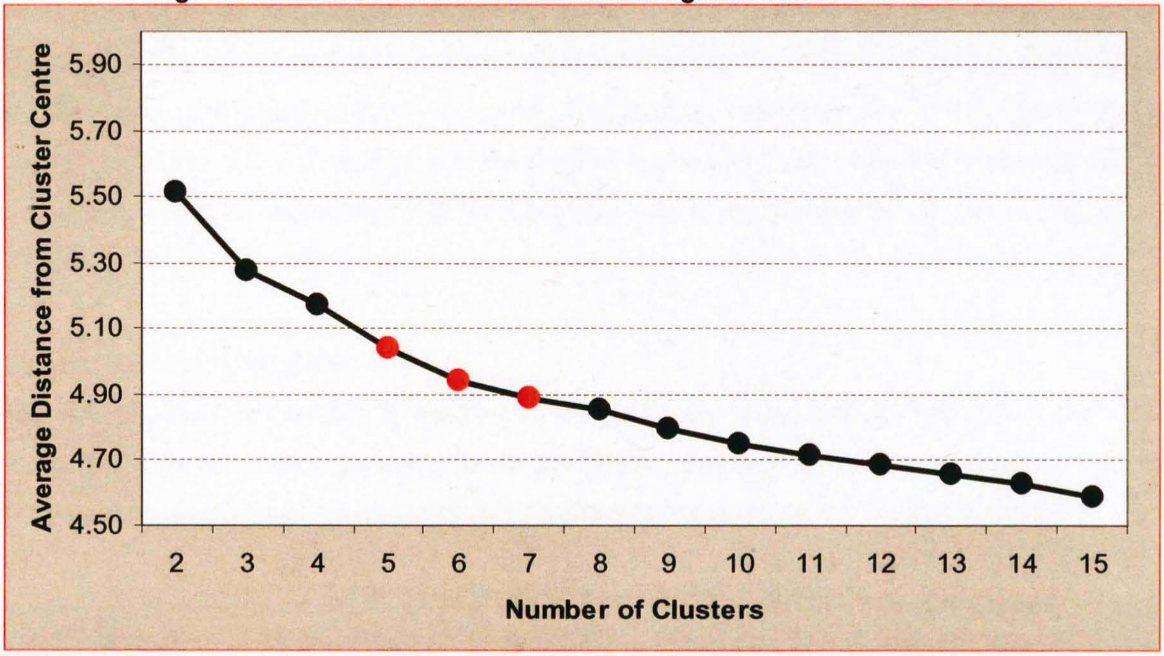
Before a hierarchy can be created, it is important to decide on the number of clusters at the first level of the hierarchy. If there had been previous work done in Nigeria (within this context), that may have served as a guide to choosing the number of clusters at the top level. The decision on the number of clusters constituting the top level of the classification was therefore made using some of the guiding principles from the work of Vickers (2006). These include:

1. The need for visualisation to be adequate
2. The need for cluster memberships to be relatively balanced
3. The need for suitable within geo-political region variation

Methods used to investigate the probable number of cluster at the top hierarchy number of clusters are slightly informal. Everitt et al. (2001) identified the possibility of plotting the value of the clustering criterion against the number of clusters and observe points of great change in the plot. Generally, the average distance of cases from their cluster centres can be plotted against the number of clusters.

In this exercise, the algorithm was run for 2 to 15 clusters. The average distance of each case from its cluster centre was computed and plotted against the number of clusters (see Vickers, 2006). This graph is depicted in figure 5.4.

**Figure 5.4: Distance from Cluster Centre against Number of Clusters**




A sharp increase in the average cluster centre would suggest the optimal solution for number of clusters at the top hierarchy. From the chart above, it is not easy to decipher where there is an abrupt change in the magnitude of the average distance from cluster centers.

The maps of the cluster solutions were examined at this point and it was observed that better discrimination between areas commenced from point 5 upwards and visualisation was more difficult after point 7. Again, from point 8, the range of cluster composition (i.e. number of zones in each cluster) increased greatly. The three points for clusters 5, 6 and 7 (in red) on the graph above are for indicative purposes. It was therefore decided that these three points would be put to test against each other.

### *Further Analysis of Distance to Cluster Centre*

An evaluation of the distance of each LGA from its cluster center was examined for clusters 5, 6 and 7. The less the distance a case is from its cluster centre, the better.

Each of the cluster solutions is positively skewed. The positive skew is significance that majority of the LGA's are at the lower distance categories. The larger the value of the positive skew, the better solution in terms of how close members are to the centre of their cluster. Both the 5 and 7 cluster solutions have a skew of 0.31 while the 6 cluster solution has a skew of 0.38 indicating that its members are more compact than the 5 and cluster solutions.

### *Cluster Membership Sizes*

The composition of clusters is desired to be relatively balanced. Where too many cases concentrate in too few clusters creates a skewed distribution. Cluster membership was assessed by examining the range of the distribution for each of the three solutions.

**Table 5.10: Assessing the Sizes of Clusters**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Five Cluster Solution** | | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | | | **Range** |
| | No. of LGA's | 208 | 110 | 128 | 120 | 208 | | | 98 |
| | % of LGA's | 27 | 14 | 17 | 16 | 27 | | | 13 |
| | | | | | | | | | |
| **Six Cluster Solution** | | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | | |
| | No. of LGA's | 181 | 166 | 114 | 126 | 82 | 105 | | 99 |
| | % of LGA's | 23 | 21 | 15 | 16 | 11 | 14 | | 12 |
| | | | | | | | | | |
| **Seven Cluster Solution** | | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 | |
| | No. of LGA's | 107 | 102 | 162 | 160 | 52 | 89 | 102 | 110 |
| | % of LGA's | 14 | 13 | 21 | 21 | 7 | 11 | 13 | 14 |

Again as can seen in Table 5.10, the seven clusters solution does not perform as well as the five and six cluster solutions.

### *Examining Predictive Power*

In order to assess the predictive power of each solution, seven variables (not included in the clustering process) were selected. The variables cut across a variety of topics and demonstrate varying values of the co-efficient of correlation when compared with the total population. The variables include: population in monogamous marriage, population in public sector employment, economically inactive population aged 15 to 24, uneducated population, household heads self employed in agriculture, population of children living in non nuclear households, children aged 0 to 4 vaccinated against measles.

The populations represented in each variable were aggregated by geodemographic cluster for each of the three solutions. Total populations by geodemographic clusters were also derived. The crude rates for each geodemographic cluster was calculated by dividing the total population by the population of the variable represented. For each LGA, the rate calculated for its geodemographic cluster was applied to its total population to derive a predicted value for that variable.

For each of the cluster solutions, the predicted values for the 7 variables were subtracted from the actual values. The error introduced was then quantified using a variant of Fisher and Langford's (1995) root mean square (RMS) error adapted by Gregory (2000).

$$E^{RMS} = \left[\frac{1}{m}\sum_{m}\left(\frac{y-y'}{y}\right)^2\right]^{1/2} \quad (5.1)$$

where

$E^{RMS}$ is the Root Mean Square error

m is the number of LGAs

y is the actual value of the variable

y' is the predicted value of the variable

**Table 5.11: Results from RMS Error Analysis**

| Variable Code | Variables | Correlation with Total Population | RMS for 5 Cluster Solution | RMS for 6 Clusters Solution | RMS for 7 Cluster Solution |
|---|---|---|---|---|---|
| Var 1 | Population in monogamous marriage | 0.84 | **0.51** | 0.52 | **0.51** |
| Var 2 | Population in public sector employment | 0.67 | 2.56 | 2.45 | **2.43** |
| Var 3 | Economically inactive population aged 15 to 24 | 0.88 | **3.24** | 3.25 | 3.38 |
| Var 4 | Uneducated population | 0.37 | 1.42 | **1.37** | **1.37** |
| Var 5 | Household heads self employed in agriculture | 0.19 | 5.31 | **5.23** | 5.55 |
| Var 6 | Population of children living in non nuclear households | 0.63 | 1.98 | **1.89** | **1.89** |
| Var 7 | Children aged 0 to 4 vaccinated against measles | 0.88 | 2.98 | **2.74** | 2.79 |

The results from the extensive analysis are provided in table 5.11. The five cluster solution appears to work with variables that are highly correlated with the total population. It performs well with two variables (Var 1 and Var 3).

The choice is clearly between the 6 and 7 cluster solutions. It is difficult to draw precise conclusions on these two solutions. However a closer observation shows that in both situations where the 7 cluster solution out-performs the 6 cluster solution, the RMS error is marginal. For Var 1, there is a marginal difference of 0.01 while Var 2 shows a marginal difference of 0.02.

Another interesting reflection pertains to Var 3. While the 5 cluster solution marginally out-performs the 6 cluster solution, the 6 cluster solution performs much better than the 7 cluster solution.

Based on these findings, one can conclude that both the 6 and 7 cluster solutions perform better than the 5 cluster solution, but the 6 cluster solution marginally out-performs the 7 cluster solution.

***Discriminatory Power***

Discriminatory power has been assessed by deriving the Gini coefficient for each of the seven variables listed in the section above across the three cluster solutions. The Gini coefficient is used to measure the degree of concentration of a variable within a distribution of its elements Brown (1994).

When used alongside the Lorenz curve, the Gini coefficient allows a graphical comparison of inequality. Values for the Gini coefficient range from 0, where there is perfect equality, and 1, where there is perfect inequality. The Gini represents an expression of the area located between the line of perfect equality and the Lorenz curve. Leventhal (1995) describes it as a method which can help mitigate the challenges posed by numerical methods of comparison of discriminatory power.

The Lorenz curves have been derived by calculating the index as described in equation (5.2). This index, an indicator of the propensity of the variable has been used to sort the percentage of each variable and base populations (total populations) in descending order (see table 5.12).

$$\text{Index} = \frac{n \div \sum_{1}^{k} n}{N \div \sum_{1}^{k} N} \times 100 \qquad (5.2)$$

where:

n = the count of people with a characteristic in geodemographic cluster k

K = the total number of geodemographic clusters

N = the count of people in geodemographic cluster k

**Table 5.12: A profile report re-ordered for the 7 Cluster Solution**

| Index | Cluster | Population of children living in non nuclear households (%) | Total Population (%) |
|---|---|---|---|
| 185 | 5 | 11.12 | 6.01 |
| 149 | 4 | 27.42 | 18.47 |
| 147 | 2 | 16.68 | 11.38 |
| 121 | 7 | 25.81 | 21.26 |
| 64 | 1 | 7.52 | 11.74 |
| 40 | 6 | 4.22 | 10.48 |
| 35 | 3 | 7.23 | 20.67 |

The profile and base percentage values were subsequently accumulated and used to derive the Gini co-efficient and Lorenz curves. For the Gini, the following formula suggested by Brown (1994) has been adopted:

$$G = 1 - \sum_{i=0}^{k-1} (y_{i+1} + y_i)(x_{i+1} - x_i) \quad (5.3)$$

where

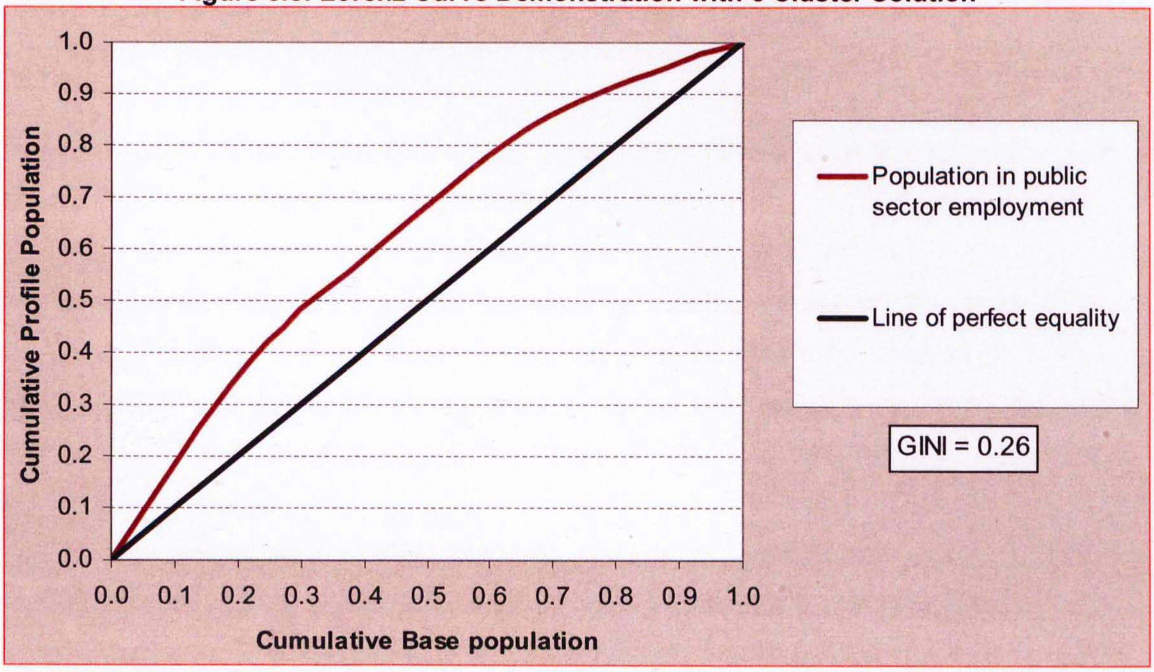
G is the value of the Gini coefficient

k is the number of data points for the profile and base populations

y is the profile population for a selected geodemographic cluster

x is the base population for the selected geodemographic cluster

**Figure 5.5: Lorenz Curve Demonstration with 6 Cluster Solution**




Since the population employed in the public sector are unevenly distributed (i.e. inequality exists), the curve shifts away from the diagonal line of perfect equality. The larger the area between these two lines, the better a system can uncover the differences within the population.

**Table 5.13: Comparison of the Discriminatory Power**

| Variable Code | Variables | Correlation with Total Population | GINI for 5 Cluster Solution | GINI for 6 Cluster Solution | GINI for 7 Cluster Solution |
|---|---|---|---|---|---|
| Var 1 | Population in monogamous marriage | 0.84 | 0.08 | 0.08 | 0.08 |
| Var 2 | Population in public sector employment | 0.67 | 0.25 | **0.26** | **0.26** |
| Var 3 | Economically inactive population aged 15 to 24 | 0.88 | **0.08** | 0.07 | **0.08** |
| Var 4 | Uneducated population | 0.37 | 0.31 | **0.34** | **0.34** |
| Var 5 | Household heads self employed in agriculture | 0.19 | 0.22 | 0.22 | 0.22 |
| Var 6 | Population of children living in non nuclear households | 0.63 | 0.26 | **0.28** | **0.28** |
| Var 7 | Children aged 0 to 4 vaccinated against measles | 0.88 | 0.10 | **0.11** | **0.11** |

The suggestion of Brown (1994) about the area between the Lorenz curve and the line of perfect equality is summarised in his writing:

'*Defined graphically, the Gini coefficient formally is measured as the area between the equality curve and the Lorenz curve, divided by the area under the equality curve*' (Brown, 1994 p. 1247).

From this suggestion it can be deduced that:

$$\mathbf{A} = \mathbf{G} \div 2 \quad (5.4)$$

where

***G*** is the value of the Gini coefficient

**A** is the area between the curve and line of perfect equality

Clearly a larger value of the Gini would equate to a larger value for the area inside the curve.

An assessment of table 5.13 reveals that with any variable, no one solution out-performs the other two. Solutions 6 and 7 again seem to perform better than solution 5. The seven cluster solution marginally out-performs the 6 cluster solution with Var 3 but has a tie with the five cluster solution. It is difficult to choose (in terms of discrimination) between the 6 and 7 cluster solutions.

After careful assessment of the results as discussed above, the 6 cluster solution was chosen. It outperforms both the 5 and 7 cluster solutions in virtually all the tests and performs just as well as the 7 cluster solution in the final test for discriminatory power.

The selection of a 6 cluster solution at the first hierarchy implied that the second and third hierarchies could be created. To create the second hierarchy, the top-down method used in the U.K. O.A. classification was used (see Vickers, 2006). This method ensures cluster groups in lower hierarchies maintain as much traits of their parent clusters. Each of the 6 cluster groups identified at the first level was clustered separately using K-means algorithm. At the second level, between 2 and 5 clusters were created for each cluster and evaluated in a similar manner as the first level. The process was tedious but ensured that the results are robust. The solutions resulting from the second level of analysis was a total of 23 clusters. The 23 cluster solutions were further manipulated in similar manner to create the third hierarchy of 57 cluster solutions.

## 5.5 Evaluating Deliverables

The outputs of the analysis can be communicated in a variety of ways. For one, since the system is linked with geography, LGA's in this case, then it can be mapped. In addition to this, textual descriptions can also be generated on the basis of the input variables about the different cluster groups. Statistical and other seemingly complex information about the system can also be communicated with charts and graphs. This section summarises how some of these information has been generated.

### 5.5.1 Adopting a Nomenclature

Labeling a group with a name can be subjective or contentious. It is a very complex process requiring consideration of numerous issues. The names are expected to be as widely representative of the characteristics of the people living in those areas as possible. This does not in any way imply that every single person within a cluster can be labeled that way. To some extent, diversity still exists within similarity (Voas and Williamson, 2001a). The names attached to clusters are only indicative of the predominant features of the areas in question.

In this exercise, some principles were considered prior to naming the clusters. First, it was thought that names should not in any way be offensive. In a multi-ethnic and multi-religious country like Nigeria, it can be easy for people to read diverse meanings to labels which are used to describe them. Care was therefore taken to ensure that the names chosen do not appear to stigmatise any section of the public. It was also decided that religious and ethnic languages should be avoided. This was difficult because the clusters reveal some underlying characteristics along these lines. However, history has it on record that polarisation and conflicts also align themselves on these issues (Gordon, 2003). Moreover, one of the major aims of this research is to identify similarities between different peoples and strengthen the concept of unity in diversity.

Table 5.14 provides a lit of the labels for the different clusters. The six clusters in the first hierarchy are referred to as **Super-groups** while the second level of 23 clusters is called **Groups.** The third hierarchy comprises 57 clusters called **Sub-groups**.

**Table 5.14: Cluster Labels for the Nigerian Geodemographic System**

| Super-group Labels | | Group Labels | | Sub-group Labels | | |
|---|---|---|---|---|---|---|
| 1 | Green Towns | 1.1 | Conventional Green Towns | 1.1.1 | 1.1.2 | 1.1.3 |
| | | 1.2 | Underprivileged Green Towns | 1.2.1 | 1.2.2 | |
| | | 1.3 | Flourishing Green Towns | 1.3.1 | 1.3.2 | 1.3.3 |
| | | 1.4 | Struggling Green Towns | 1.4.1 | 1.4.2 | 1.4.3 |
| 2 | Emerging Localities | 2.1 | Moderately Emerging Localities | 2.1.1 | 2.1.2 | 2.1.3 |
| | | 2.2 | Comfortable Emerging Localities | 2.2.1 | 2.2.2 | 2.2.3 |
| | | 2.3 | Transient Emerging Localities | 2.3.1 | 2.3.2 | 2.3.3 |
| 3 | Intermediate Territories | 3.1 | Constrained Intermediate Territories | 3.1.1 | 3.1.2 | |
| | | 3.2 | Well-to-do Intermediate Territories | 3.2.1 | 3.2.2 | 3.2.3 |
| | | 3.3 | Deprived Intermediate Territories | 3.3.1 | 3.3.2 | |
| | | 3.4 | Customary Intermediate Territories | 3.4.1 | 3.4.2 | |
| 4 | Diluted Societies | 4.1 | Thriving Diluted Societies | 4.1.1 | 4.1.2 | |
| | | 4.2 | Labouring Diluted Societies | 4.2.1 | 4.2.2 | 4.2.3 |
| | | 4.3 | Deprived Diluted Societies | 4.3.1 | 4.3.2 | |
| | | 4.4 | Modest Diluted Societies | 4.4.1 | 4.4.2 | 4.4.3 |
| 5 | Country Dwellings | 5.1 | Toiling Country Dwellings | 5.1.1 | 5.1.2 | |
| | | 5.2 | Deprived Country Dwellings | 5.2.1 | 5.2.2 | |
| | | 5.3 | Middle-class Country Dwellings | 5.3.1 | 5.3.2 | 5.3.3 |
| 6 | Urban Nodes | 6.1 | Prosperous Urban Nodes | 6.1.1 | 6.1.2 | |
| | | 6.2 | Disadvantaged Urban Nodes | 6.2.1 | 6.2.2 | |
| | | 6.3 | Average Urban Nodes | 6.3.1 | 6.3.2 | |
| | | 6.4 | Affluent Urban Nodes | 6.4.1 | 6.4.2 | 6.4.3 |
| | | 6.5 | Striving Urban Nodes | 6.5.1 | 6.5.2 | |

Closely linked with the procedure for developing pen portraits, the names attempt to depict some of the distinguishing characteristics of the areas. Distinguishing variables have been identified by looking at the resulting z-scores for the final cluster centres of each cluster. The global mean for each cluster was calculated and variables with values higher than the global mean stand out key defining variables. Details on how this was done are given in the next section.

### 5.5.2 Profiles and Pen Portraits

One of the very useful derivatives of this gedoemographic system is the underlying textual and graphical information that accompanies it. These textual descriptions often called pen portraits or profiles help summarise the prevailing general attributes of the cluster groups identified and are mostly devoid of statistical jargon. The pen portraits will be of particular interest especially to the less technical audience who will be making use of the system. Appendix A2 shows an example of the pen portrait for Underprivileged Green Towns. The enclosed compact disc contains detailed portraits for each super-group, group and sub-group.

Figures 5.6 to 5.11 show radar charts for the six super-groups. The charts provide profiles of the 45 variables which were used to create the classification system. The z-scores of each variable within the different clusters have been plotted and are compared with the national average distribution of each variable within each cluster; the average of z-scores being zero (0).

**Figure 5.6: Profile of Green Towns**

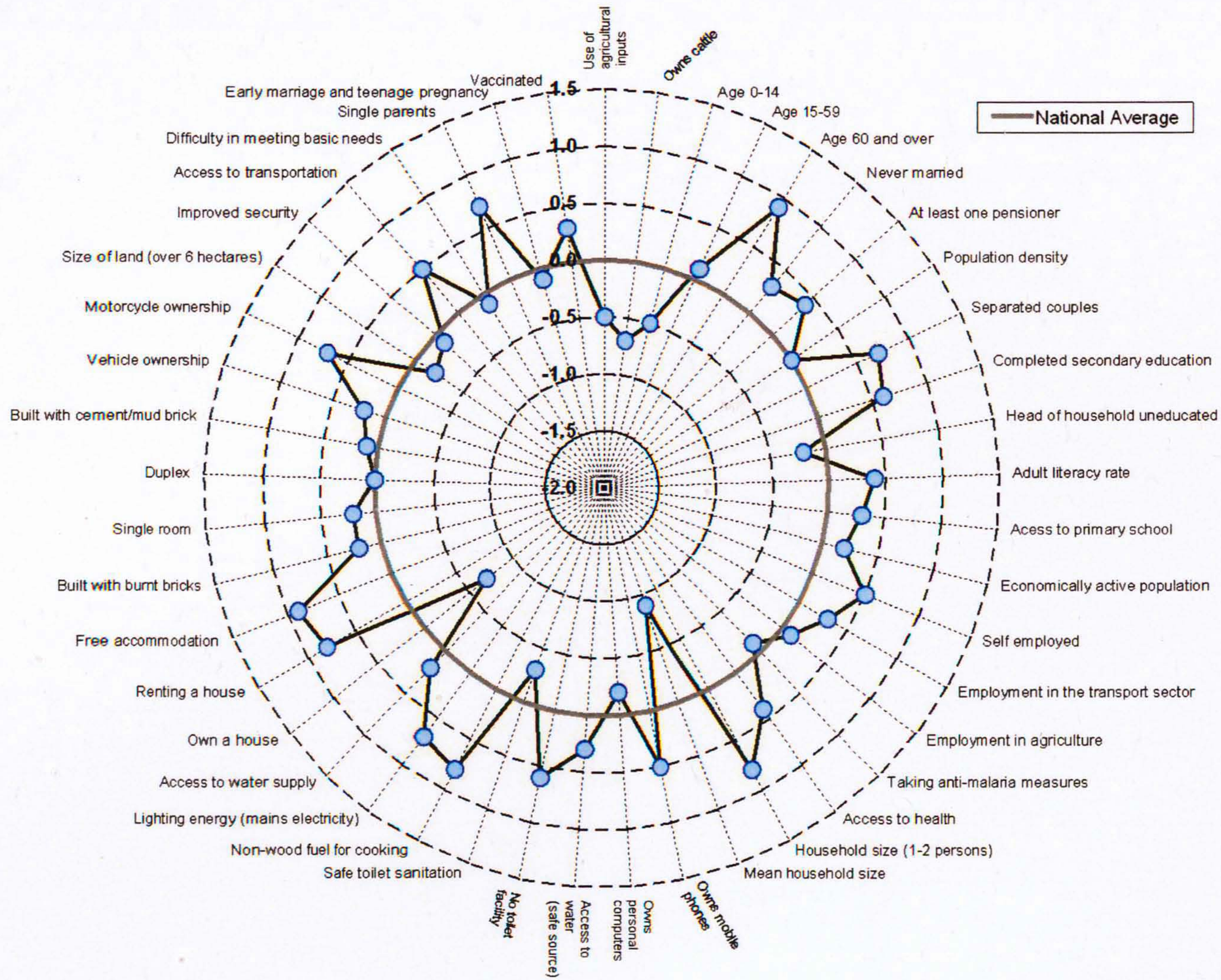


**Figure 5.7: Profile of Emerging Localities**

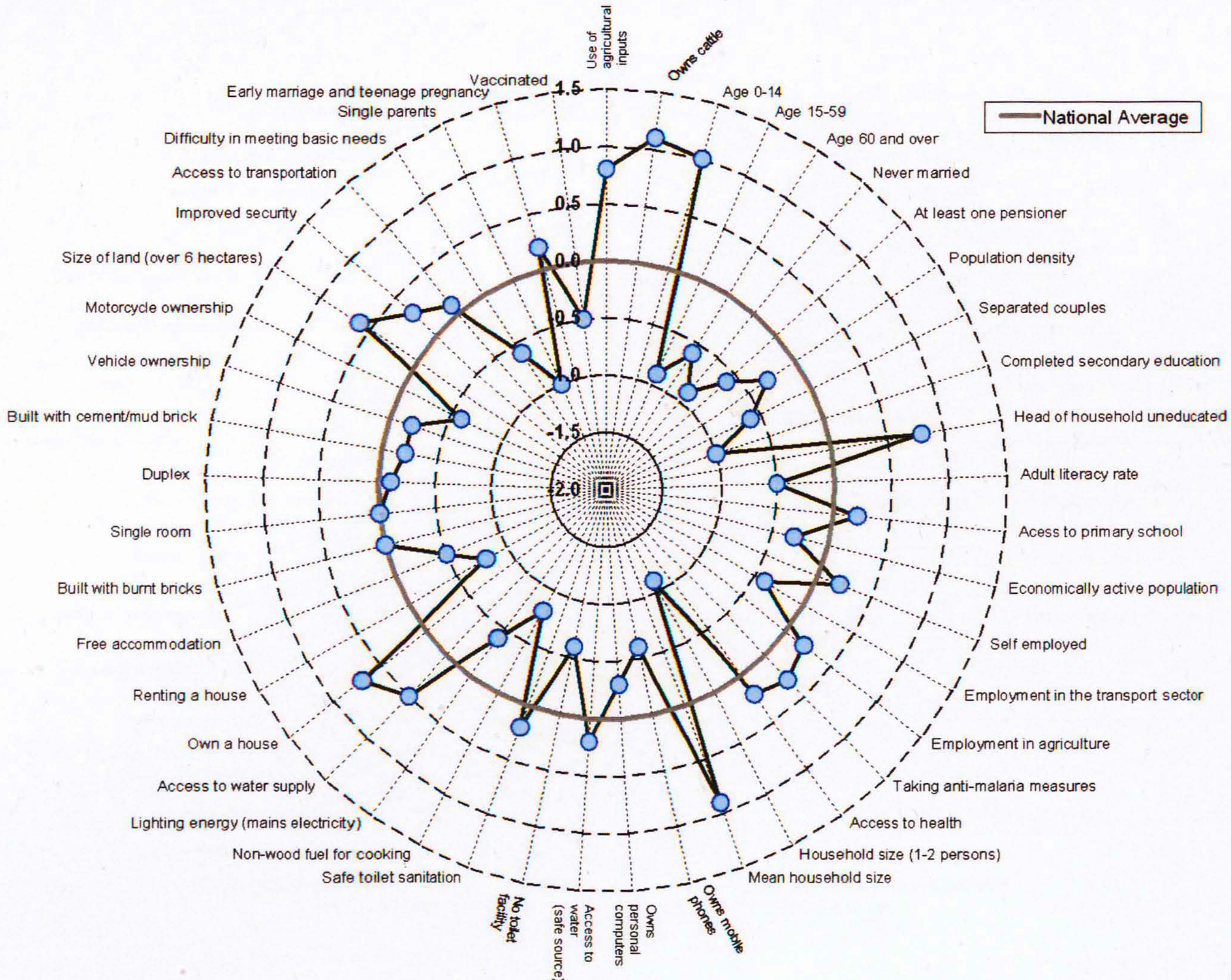


**Figure 5.8: Profile of Intermediate Territories**

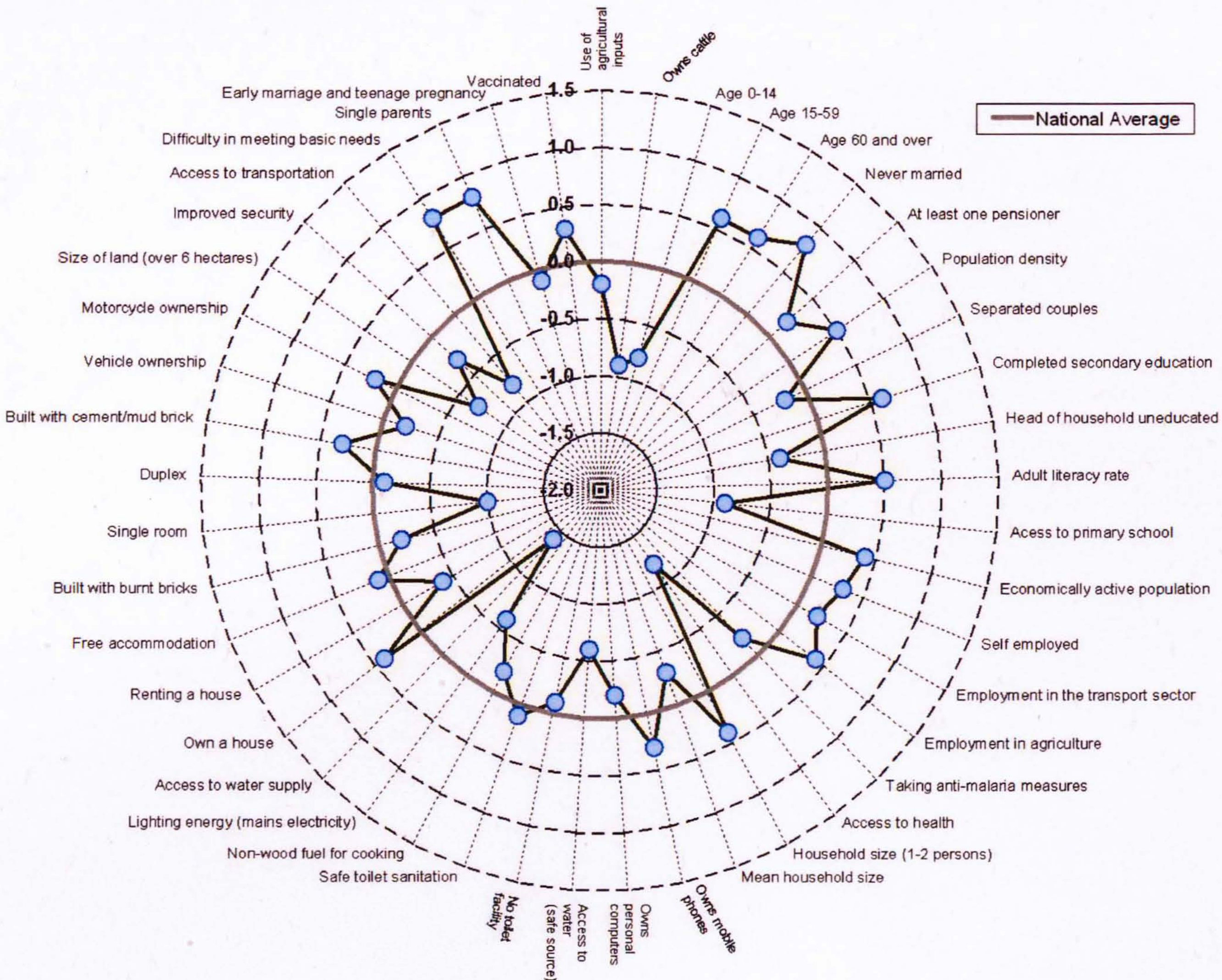


**Figure 5.9: Profile of Diluted Societies**

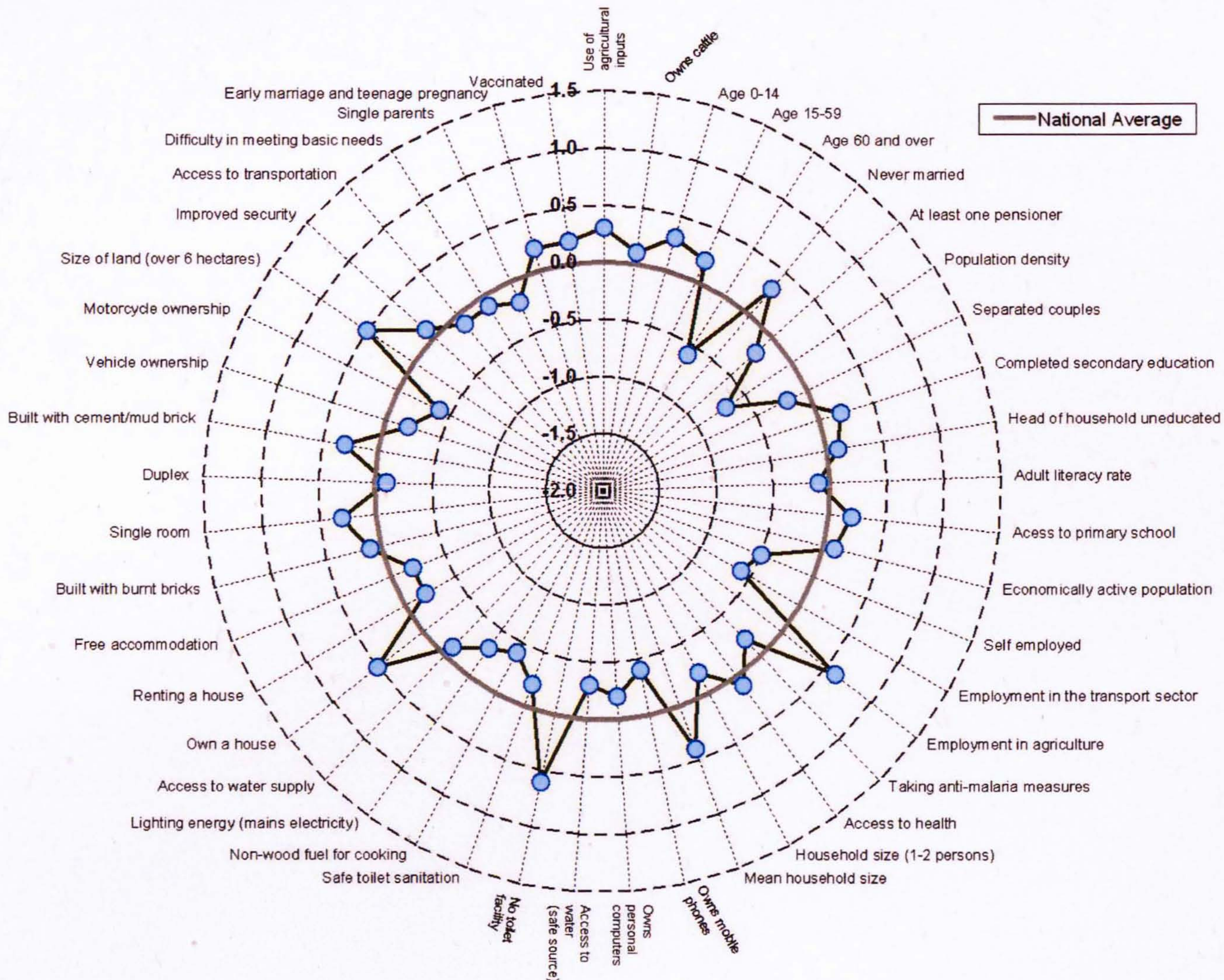


**Figure 5.10: Profile of Country Dwellings**

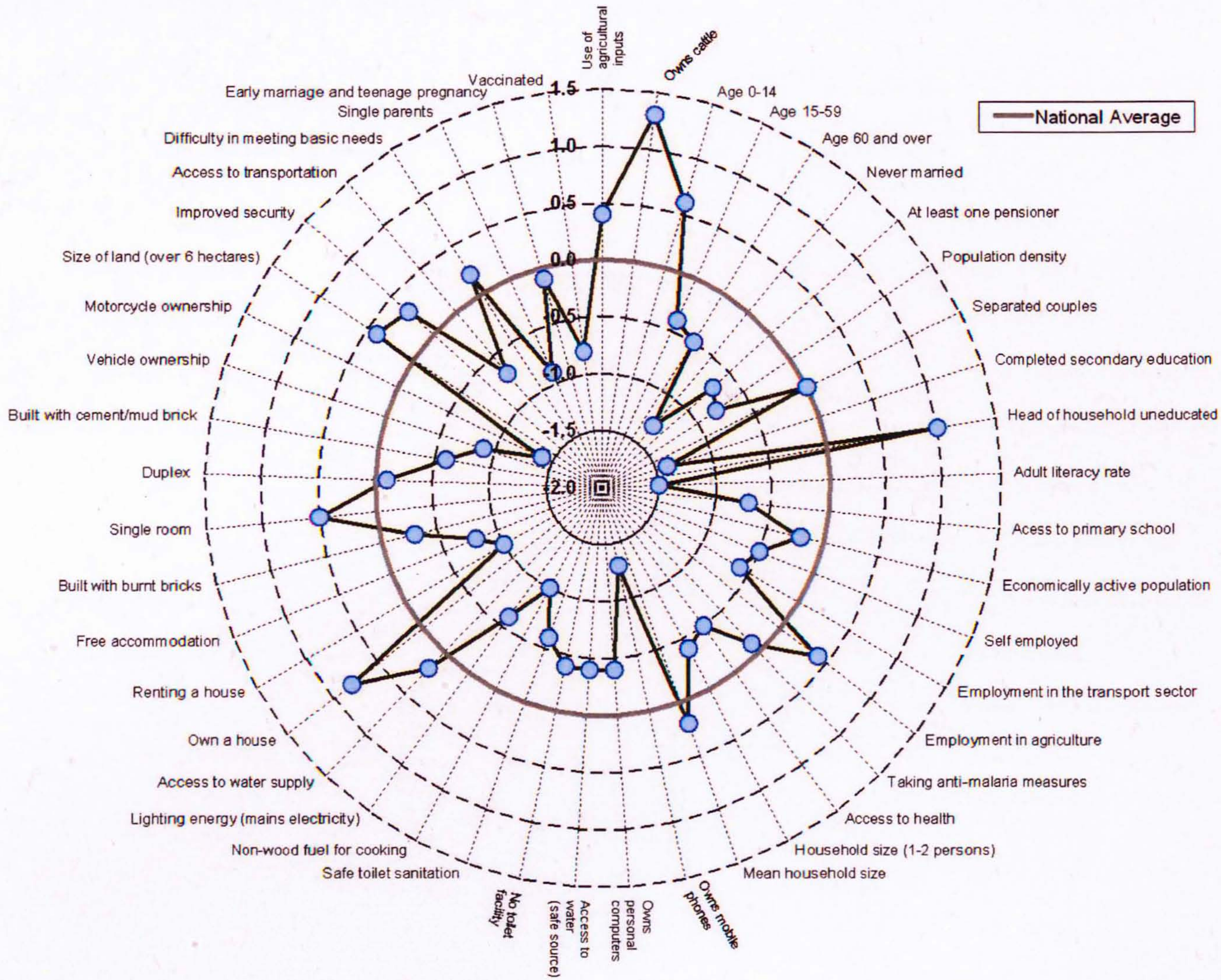
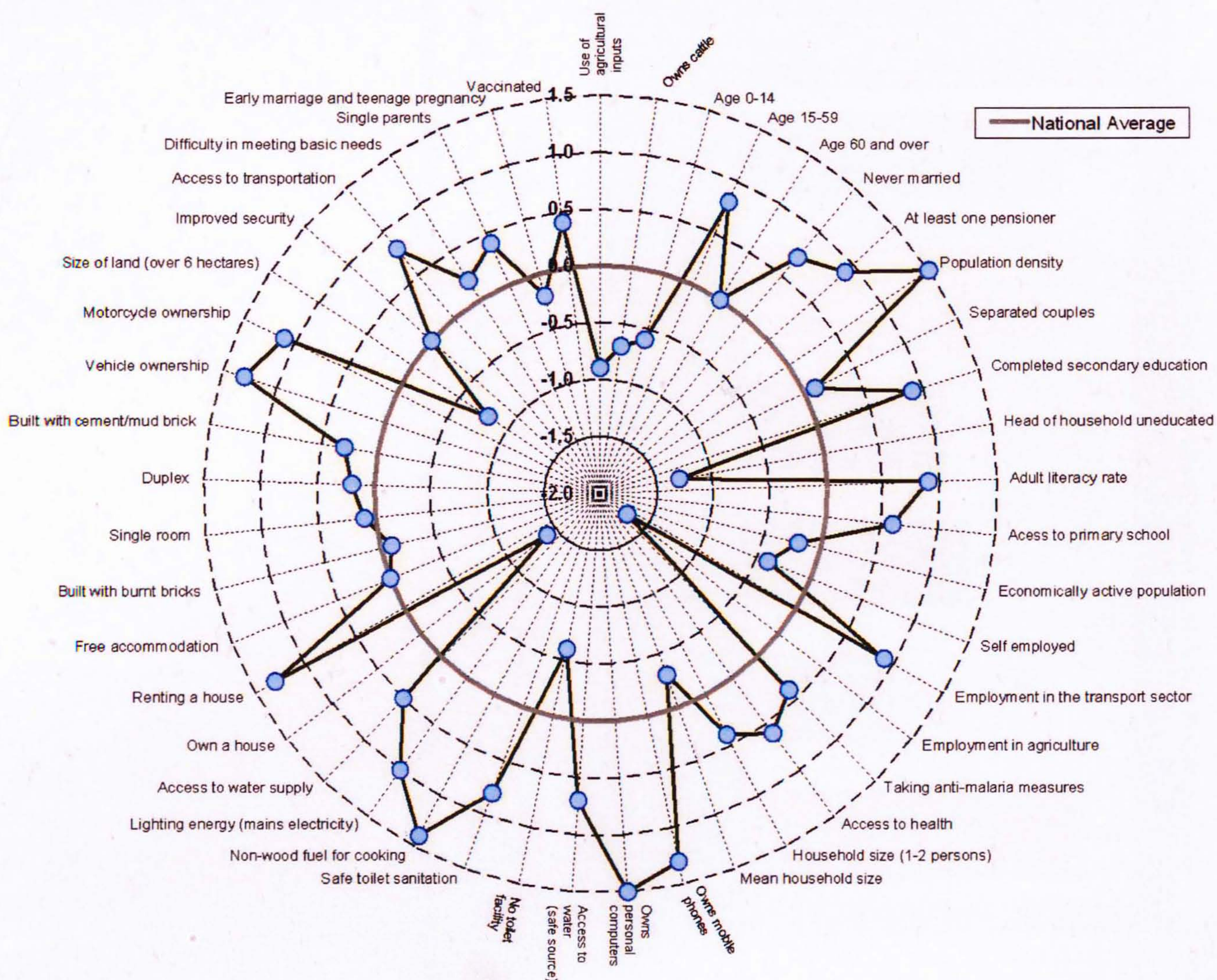


**Figure 5.11: Profile of Urban Nodes**



### 5.5.3 Visualisations

One of the strengths of the classification system is the fact that it is a geographical based system. Each cluster can be mapped to show the national pattern and distribution. Mapping the information provides additional insight for potential users. Appendix A3 and A4 show the visualisations of the super-groups and groups.

**Figure 5.12: National Population Distribution across the Super-groups**

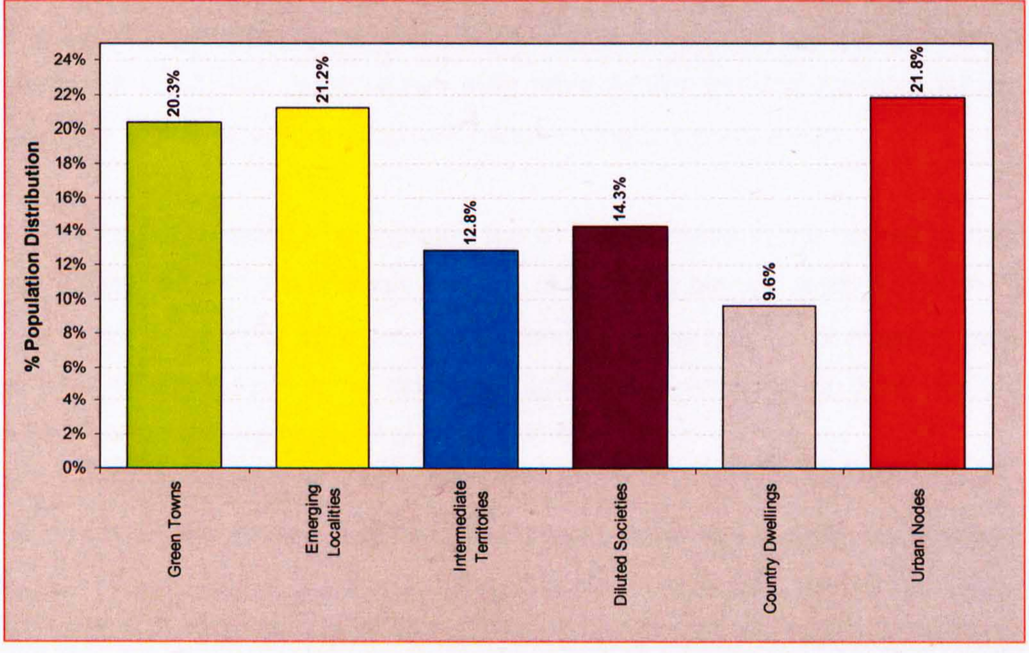



Figure 5.12 shows the national population distribution. From the chart we can easily understand the effects and possibly the direction of migration from the Country Dwellings which are often defined by lack of access to infrastructure to less rural areas like the Urban Nodes.

## 5.6 Conclusions

The discussions in this chapter has explained how against the odds and difficulties of securing datasets a segmentation system has been developed for Nigerian Local Government Areas by taking geo-statistical and current policy debates (particularly MDG related policies) into context.

At the commencement of this exercise, it was not known if clusters of population existed in Nigeria at Local government Area level. Indeed it was unclear if it was possible to prove that while residents within the same region may have similar geodemographic features, they may also have similar characteristics with people living in distant areas.

The chapter has succeeded in discussing the decisions made in the course of deciding on variables for inclusion in the analysis and the choice of methods deployed on the selected variables. While it has been stressed that the choice of the number of clusters especially at the top level of the classification can vary, light has been shed on how the number of clusters for this system was arrived at.

By itself however, the system that has been developed is not enough. It is not enough to provide jobs for the jobless or education for children or health care for the sick. There is the need for political will to do these things. Having said that, the new system can only be understood better if it is explored in detail and used to generate insight. It is very rare and uncommon to find local level national analysis done in Nigeria. In chapters 7 and 8, the Nigerian system will be used to demonstrate how this can be achieved.

# 6

# Geodemographic Segmentation for the Philippines

## 6.1 Introduction

Chapter six focuses on explaining and justifying the decisions made in the course of creating a geodemographic classification system for the Philippines. Sections 6.2 and 6.3 provide an overview of the social features of the country and the administrative geography of the Philippines. In Section 6.4 the demographic characteristics of the country are examined and information is provided on the sources of the datasets. Section 6.5 provides insight into the different datasets including their sources. From section 6.5 to 6.8, the methodologies deployed on the datasets resulting in the classification system are explained in detail while section 6.9 focuses on the nomenclature of clusters, mapping and profiling the classification. In section 6.10 a summary of the chapter is provided.

## 6.2 Natural Divisions and Administrative Geography

Philippines is comprised of 7,107 islands which are collectively known as Philippine archipelago (NSO, 2008). The total area of the country is nearly 300,000 kilometers square (NSO, 2008). About 1% (1,830 square kilometers) of this area is water.

The group of islands is naturally divided into four segments (Coursey, 2008). In the northern area, there is Luzon which is the largest amongst the four groups of islands. Luzon comprises areas like Palawan, Mindoro, Masbate and Marinduque.

The Visayas is another cluster of islands south of Luzon. This group is located in the central area of the country. Some of the important islands within the Visayas include Samar, Leyte, Bohol, Cebu, Negros and Panay (Coursey, 2008).

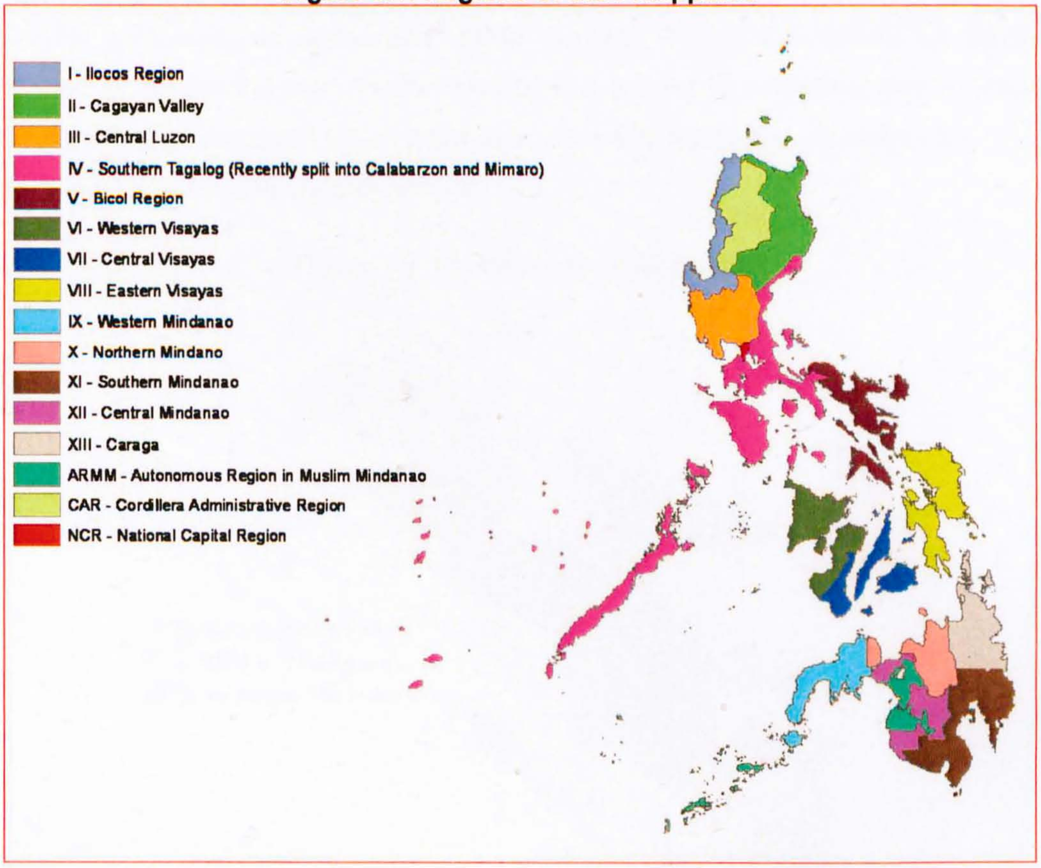
The second largest group of islands is located in the southernmost part country. This group is called Mindanao. Mindanao comprises Camiguin, Basilan, and a host of other islands including the Sulu archipelago.

The fourth group of islands is located in the south-western division and is called Palawan. The group comprises a cluster of islands extending in a narrow fashion and connecting islands in the Luzon group with the Visayas.

## 6.2.1 Regions, Provinces, Municipalities and Barangays

The administrative geography of the Philippines is comprised of four hierarchies. At the top level of the hierarchy are a set of regions totaling 17 in number. Each region is subdivided into a number of provinces. There are currently 81 provinces. At the third level of the delineation are a set of 1,494 municipalities (NSO, 2008). The finest level of geographical aggregation is called barangay which is a village or small district. Barangays are subdivisions of municipalities and sum up to 41,995.

**Figure 6.1: Regions of the Philippines**




Source: National Statistics Office, Manila

Regions provide government departments with bases to establish their offices.

Figure 6.1 shows the geographical distribution of the regions. The map shows 16 polygons which are contained in the GIS file supplied for this project by the National Statistics Office (NSO) in Manila. The Southern Tagalog region has recently been split into two regions called Calabarzon and Mimaro.

Based on the 2000 Census, the average population of each region was 4.7 million people. The Southern Tagalog region recorded 15% (the highest) of the national population. This suggests why it was split into two. It is followed by the National Capital Region which accounts for 13% of the national population. The other double digit percentage was recorded in Central Luzon which accounts for 11% of the national distribution. Cordillera Administrative Region accounts for only 2% of the national distribution which is the least amongst all regions. The range of population distribution within the regions is quite large at 10,428,435 people.

Provincial geography is embedded within regions. Figure 6.2 shows six provinces contained in the Bicol region. Each province has a level of autonomy and is governed by two arms of government which have executive and legislative powers (Wood, 2006). Each province is headed by a Governor.

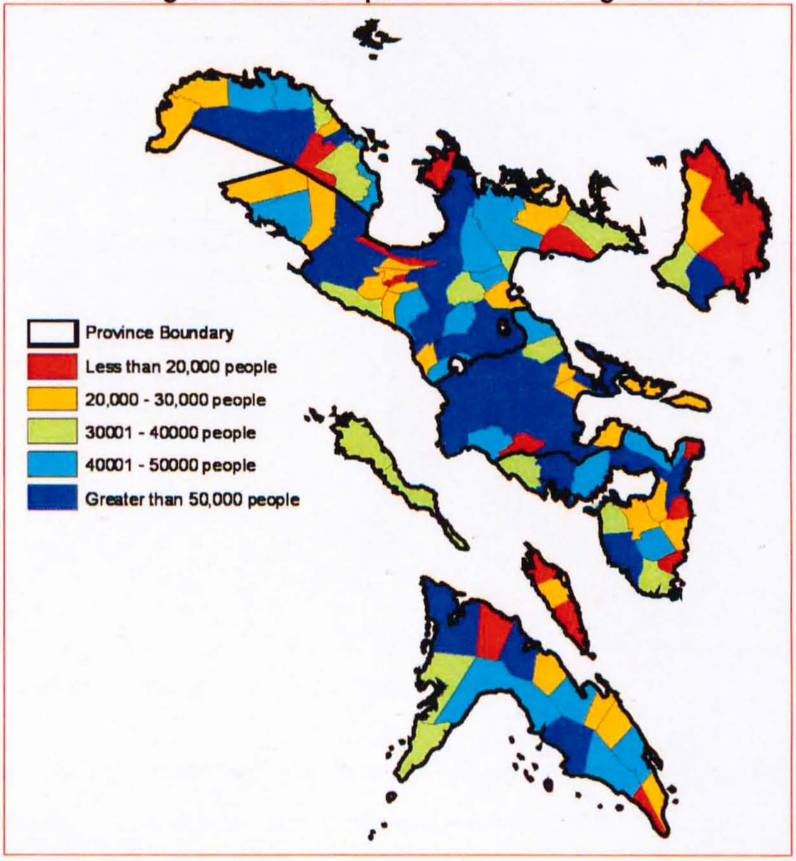
**Figure 6.2: Provinces in Bicol Region**


Source: National Statistics Office, Manila

The average population distribution within provinces based on the 2000 Census is 727,661. This figure represents a large disparity from the largest and least provincial population distributions. Pangasinan province (found in Ilocos region) has the highest population of 2,434,086 people while the least populated province is Batanes province (located in Cagayan Valley) with a population of only 16467 people. This gives a large range of over 2.4 million people.

Local level governance is conducted at municipality level in the Philippines. Municipalities are subdivisions of provinces and total 1,494 (NSO, 2008). Each municipality is governed by a Mayor who presides over executive issues (Woods, 2006).

**Figure 6.3: Municipalities in Bicol Region**

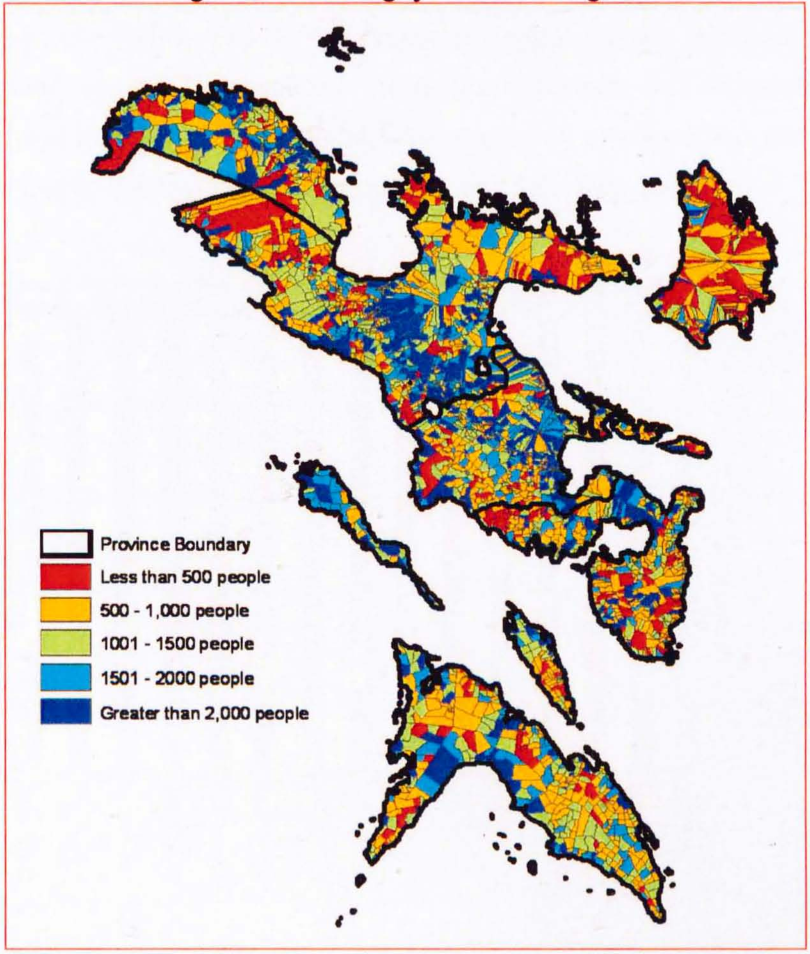



Source: National Statistics Office, Manila

The areas of municipalities vary significantly as demonstrated in Figure 6.3 and so do their populations. It is interesting to note that the average population of municipalities (47085 people) is greater than the population of some provinces. The municipality with the highest population is Quezon City with 2,173,831 people. Quezon City is located in the national capital region. Kalayaan is the municipality with the least population of only 223 people. The large range of population distribution within municipalities – over 2.1 million people is comparable to the range of population within provinces.

Barangays are the lowest level of geography for which digital boundary data are available. A barangay is a village or very small district. They are small communities expected to provide residential accommodation to populations of approximately 500 households (NSO, 2008). As a result social ties are thought to be tightly knitted at this administrative geography (Woods, 2006).

**Figure 6.4: Barangays in Bicol Region**

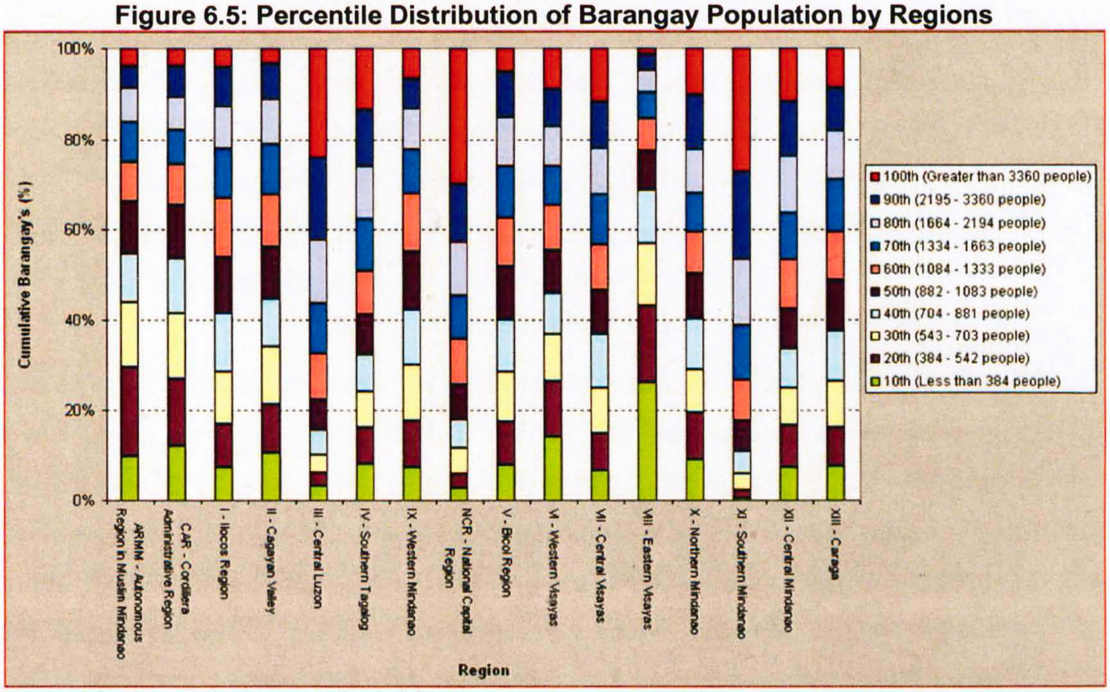



Source: National Statistics Office, Manila

The barangay is the geographical unit from which the census enumeration area (EA) is derived. Typically, barangays with projected households not exceeding 500 are considered as EAs (NSO, 2008). The average barangay population is 1,822 people. However, populations within barangays range massively from 10 people to a maximum of 188,419 people.

A large number of barangays in the National Capital Region have large concentrations of people. Figure 6.5 shows the percentile distribution of barangay population across the 16 regions. Although Southern Tagalog Region is characterised by the largest resident population, the population density within the National Capital Region exceeds it by a mile. The National Capital Region has a population density of 15,617 people per km$^2$. This figure is more than 61 times the national average while Southern Tagalog enjoys a density of 251 people per km$^2$ which is less than the national average value of 255 people per km$^2$ (NSO, 2001). The National Capital Region is comprised of many highly urbanised areas. The reasons for disproportionate concentrations of people within barangays in the National Capital Region are not unconnected with the general perception of better quality of life within this region (Schelzig, 2005).

**Figure 6.5: Percentile Distribution of Barangay Population by Regions**




Source: National Statistics Office, Manila

A reflection on the characteristics of Philippines administrative geography suggests as is the case with many countries of the world that administrative geographies are often delineated without taking future population growth and potential spatial analysis into consideration. Greater emphasis is often placed on natural boundaries or communal neighbourhoods. Results from an analysis reporting on barangays can have massive implications when an area comprising only 10 people is compared with another area comprising over 100,000 people.

Although urbanisation and migration are factors to consider when accounting for the disproportionate distribution of population within geographical areas (Kojima, 1996) future alterations to the administrative geography of the Philippines should make the patterns and distribution of population an important factor for consideration.

## 6.3 Social Aspects of the Study Area

The Philippines is one of the most populated countries in the Asian region. It ranks 12th globally and 7th in Asia behind China, India, Indonesia, Pakistan, Bangladesh and Japan (ADB, 2009).

The current inhabitants of the country are thought to have historical links with migrants from Indonesia, China and Malaysia (Coursey, 2008; Worcester, 2005). For about 300 years from 1556, the country was under Spanish Colonial rule (Woods, 2006). Spanish rule ended in 1896; however the collapse of the Spanish failed to result in independence. Rather a new era of colonial control by the United States commenced in 1898. The country gained her independence from the United States on the 4th of July 1946 (Woods, 2006; Coursey, 2008) but had its economy predominantly hinged to the economy of the United States.

The Philippines has more than 170 different languages (Gordon, 2005). Two of these (the English language and Filipino/Pilipino) are the national official languages. The country has a rich and interesting culture across various ethnic and linguistic groups. However, just like Nigeria, there are three major groups which constitute the greatest share of 50% of the entire population. These are the ***Tagalog's***, the ***Cebuano's*** and the ***Ilocano's*** (NSO, 2008). The Tagalog's make up 28% of the population, The Cebuano ethnic groups make up 13% while the Ilocano's make up 9%. Other groups include the Bisaya and Hiligaynon which constitute 8% each, the Bikol making up 6% of the population and the Waray total 3%. There are several other minority population groups which make up 24% of the population (NSO, 2008). These groups are regarded as very important in economic and political terms.

Many of the indigenous traditions of the Philippines have been influenced by American and Hispanic cultures largely due to the long periods of colonialism. Just like Nigeria, geography is correlated with the cultural distributions of Filipinos. The Ilocano's concentrate in the north, the Tagalog's are mostly found in the central plains of the country while the Visayan's are clustered in the central Islands (Worcester, 2005; Woods, 2006). Most of the other groups are scattered across the archipelago.

Since the country is an island archipelago, several aspects of its existence are dominated by aquatic life (Woods, 2006). About 1.5 million square kilometers of territorial water serves about 59 lakes and 132 major rivers which provide rich fishing resources. The country has a thriving tourism industry. Some of the key tourist destinations include Cebu, Bohol, Palawan and Boracay (Woods, 2006).

In spite of abundance in human and natural capital, the Philippines has an economy that lags behind many countries in Asia (Schelzig, 2005). In 2003, about 14% of the population was estimated to living in extreme poverty (NEDA and UNDP, 2007). Like Nigeria, many of the benefits these resources fail to trickle down to those in need largely due to inadequate targeting of policies, corrupt practices, emigration of human capital and political instability (ADB, 2007).

## 6.4 Demographics and Data Sources

The Philippines provided a new challenge different in many respects to Nigeria. However, knowledge and experience from the development of the Nigerian classification proved useful. On May 1, 2000, the Philippines conducted a national population and housing census. In most countries including the Philippines it takes at least two to three years to process the information from general censuses and make the results available in aggregated forms for varying levels of geography. As at the time of this research, the census conducted in the year 2000 proved to be the most comprehensive source of data for the project because of its national coverage.

According to NSO (2002a) projections that were based on the census, the population of the country would have doubled by the year 2030 increasing to over 140 million people. Interestingly, this is the current population of Nigeria. The average national population growth rate was 2.36%. Amongst the 16 regions, 5 had a growth rate above the national average (NSO, 2002a). The Autonomous Region in Muslim Mindanao registered the largest growth rate of 3.86%. Southern Tagalog was next in line with a growth rate of 3.72%. The other three regions with a growth rate above the national average in order of magnitude were Central Luzon, 3.20%; Central Visayas, 2.80% and Southern Mindanao 2.60% (NSO, 2002b-g; NSO, 2003a-i). Interestingly, the National Capital Region where Manila is located recorded the least growth rate of 1.06%.

The census showed a relative balance in the ratio of men to women. For every 100 women, there were 101.43 men. However half of the national population was under the age of 21 years (NSO, 2002a) and there were a greater number of males in the younger age categories, 0 to 19 years and amongst the 25 to 54 year age group. Every other age group was dominated by females (NSO, 2002a).

The National Capital Region recorded the highest median age of 24 years. Only five other regions had half of their population above age 21. These were Ilocos Region, Central Luzon, Southern Tagalog, Western Visayas and Central Visayas. Each of these regions had a median age of 22 years. Additionally, only two regions – the National Capital Region and the Autonomous Region in Muslim Mindanao had a ratio of fewer men to women (NSO, 2003h and 2003i). In both regions, there were 97 men for every 100 women.

Religion plays an important role in the culture and lifestyles of most Filipinos. The country is predominantly a Christian nation with over 95% of its population defined as Christians (NSO, 2003j). The majority of this group (81%) is Roman Catholic. With the exception of the Autonomous Region in Muslim Mindanao, every other region had over 50% of its population described as Roman Catholics (NSO, 2002a-g; NSO, 2003a-j). Five percent (5%) of the population are Muslims and 9 in 10 persons within the Autonomous Region in Muslim Mindanao.

It is not uncommon with many developing world countries that the national or regional levels of geography are typically the units of analysis. A good example is the very popular and useful Demographic and Health Surveys conducted in over 70 countries by Macro International and supported by the United States Agency for International Development (USAID) and the partner countries. While analysis conducted at higher levels of geographical aggregation is useful in gaining initial insight to general socio-demographic patterns (Longley et al., 2001), they conceal inequalities at sub-national geographies (Schelzig, 2005). A major argument of this research is that better understanding of inequities at sub-national geographies improves targeting and judicious allocation of scarce or limited resources which is often a prime desire of many governments in third world countries. The creation of a small area geodemographic classification for the Philippines can therefore explore the census in detail; create further datasets and products to which surveys can be coded to gain insight.

The datasets for the census were derived from two sources. The National Statistics Office (NSO) in Manila, Philippines provided GIS digital boundaries as well as census data for multiple geographical levels. However, most of the data provided at the finest administrative geography (Barangays) were nominal or categorical data describing the presence or absence of some facilities and services. The second source of data was derived from the Integrated Public Use Microdata Series-International (IPUMS). IPUMS is based at the University of Minnesota in Minneapolis. The data from IPUMS comprised a large database of 7,417,810 individual records forming 10% of the census and covering a variety of topics and indicators. However, the data was supplied at Municipality level.

## 6.5 Choosing the Inputs

As discussed above, the finest level of geography for which spatial statistics was made available was at Barangay level. Due to variations in the levels of aggregation of statistics supplied, it was initially thought that a multi-level set of classifications should be created at Municipality and Barangay level. However, further consideration was given to the different types of variables and the potential multivariate influence of each on the classification if they were combined.

Webber (2004a) found that by incorporating data from multiple levels of geography, the discriminatory power of geodemographic classification systems could be greatly enhanced. Additionally, geodemographics seeks not just to provide summary characteristics of residents of a geographical jurisdiction but the influences of the immediate and extended communities themselves on the individuals. Webber termed these as '*neighbourhood effects*'.

In this light, it was thought that by combining all the datasets available, and creating a classification system at Barangay level, not just the geodemographic features of people and places at the finest level of aggregation would be derived but a more comprehensive picture of the way in which Municipality level datasets (unavailable at Barangay level) contribute to the discriminatory ability of a Barangay level classification. This would bolster the ability of the classification system to better uncover varying types and levels of inequality.

It is important at this stage to mention that data was not made available for 14 Barangays. Correspondence with staff from the NSO revealed that these Barangays were either evacuated during the conduct of the census due to armed clashes between rebel forces and government troops or have been evacuated due to volcanic eruption at Mount Pinatubo. These 14 Barangays were therefore excluded from the analysis and form part of the unclassified clusters.

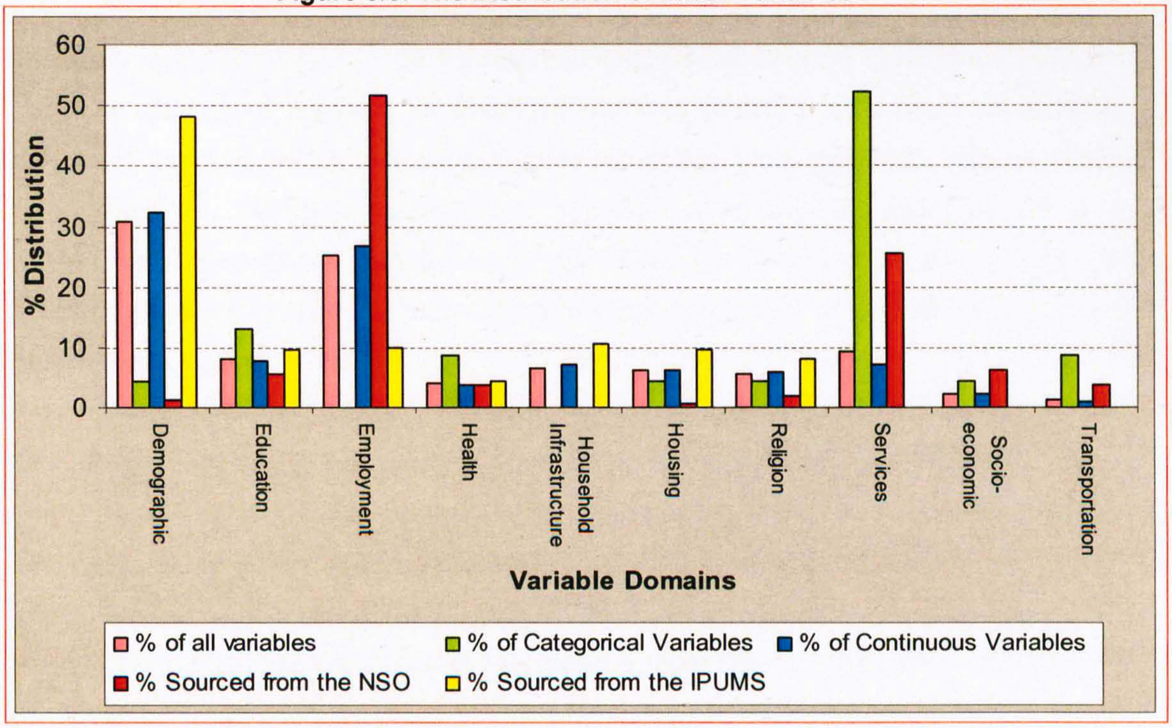
## 6.5.1 Evaluating Categorical and Continuous Variables

The statistical evaluation of the suitability of input variables is important when choosing variables. However, when attempting a classification system for any developing country, it is imperative for such a system to encapsulate variables which reflect key policy debates. This is likely to increase the usability and usefulness of the system because it will capture population patterns that will improve the targeting of such policies. Additionally, similar to the argument that different neighbourhood types have different types and levels of advantage or disadvantage (Harris et al., 2005) so do different countries define their priorities. For instance, obesity is quite high on the agenda of health priorities of the government of the United Kingdom (Zaninotto et al., 2006) but the focus in sub-Saharan Africa is malaria (RBM, 2008). In the Philippines, reproductive health and sex education form major health policy debates (Lakshminarayanan, 2003).

A total of 435 variables were available for initial consideration. For easy understanding, all the variables were aligned along 10 domains listed below. These domains were chosen based on experience from the classification of the Nigerian dataset. While they reflect the principal dimensions of the Philippine dataset, it is of course possible for some other researcher to come up with a set of different domains.

- Demographic
- Education
- Employment
- Health
- Household Infrastructure
- Housing
- Religion
- Services
- Socio-economic
- Transportation

**Figure 6.6: The Distribution of Initial Variables**




Source: NSO and IPUMS

From Figure 6.6, it is interesting to observe the variation across each domain of the data types (categorical or continuous) and the two sources of data. Of the 435 variables, 5% are categorical while 95% are continuous. A total of 161 variables accounting for 37% were sourced from the NSO while the remaining 63% were sourced from the IPUMS. Each variable was considered for potential inclusion in the classification subject to further evaluation.

The reduction of the initial variables and the selection of the final list were done in three phases. First it was decided that all 23 categorical variables should be analysed together as the statistical tests deployed on categorical data differ from continuous data. Secondly, all continuous datasets would be examined on a domain-by-domain basis and thirdly all continuous data surviving the domain-by-domain reduction would be evaluated together. This intra-domain and inter-domain analysis was deployed on the Nigerian dataset and it proved useful. At every stage in the quantitative analysis, the significance of each variable to key policy debates was also taken into cognizance. Additionally, care was taken to ensure that variables considered for inclusion would support the policy making for the Millennium Development Goals (MDGs) and sustain longevity of the classification.

## 6.5.2 Geographic Dispersion

The techniques used to analyse categorical datasets differ from continuous data and are somewhat more complex. In previous chapters (4 and 5), the importance of the spread of input variables across geographical space was stressed. Vickers (2006) noted that for adequate distinctions between areas, input variables need to demonstrate reasonably high levels of variation. In other words, the smaller the standard deviation value, the more clustered the distribution of the variable is across geographical space.

The standard deviation has been suggested by many authors (Bulmer, 1979; Urdan, 2005) as a useful statistic for measuring geographic dispersion of data. Standard deviation ($\sigma$) is defined as the square root of the variance (Crawshaw and Chambers, 2001). It provides a measure of the magnitude by which values tend to depart from the mean. This makes the derivation of the mean a focal point in calculating the standard deviation.

With continuous dataset the calculation of the standard deviation is more common and straight forward. For the categorical datasets, an *expectation* value was derived. Crawshaw and Chambers (2001) provide detailed explanation on how this can be done by calculating probabilities. The expectation is similar in interpretation to the arithmetic mean. According to them, it is '*the average value when the number of experiments increases indefinitely*' (Crawshaw and Chambers, 2001, p. 238).

**Table 6.1: Standard Deviations of Categorical Variables**

| variables | Code | Probability (Yes) | Probability (No) | Expectation E(X) or $\mu$ | $\mu^2$ | $E(X^2)$ | Var(X) | SD ($\sigma$) |
|---|---|---|---|---|---|---|---|---|
| Does the barangay have a street pattern or network of streets of at least 3 street roads? | T1 | 0.50 | 0.50 | 1.50 | 2.24 | 2.49 | 0.26 | 0.51 |
| Is Barangay part of town city proper or poblacion? | D1 | 0.45 | 0.55 | 1.55 | 2.41 | 2.66 | 0.25 | 0.50 |
| Is there a community waterworks system in the barangay? | S10 | 0.49 | 0.52 | 1.52 | 2.32 | 2.56 | 0.24 | 0.49 |
| Is there a puericulture center/health center in the barangay? | H2 | 0.62 | 0.38 | 1.39 | 1.93 | 2.16 | 0.23 | 0.48 |
| Is there a public plaza or park in the barangay? | S2 | 0.37 | 0.64 | 1.64 | 2.70 | 2.92 | 0.22 | 0.47 |
| Is there an elementary school in the barangay? | E1 | 0.74 | 0.26 | 1.27 | 1.60 | 1.79 | 0.19 | 0.43 |
| Is there a telephone in the barangay? | S7 | 0.28 | 0.73 | 1.73 | 3.00 | 3.18 | 0.19 | 0.43 |
| Is there electric power in the barangay? | S11 | 0.77 | 0.23 | 1.24 | 1.54 | 1.71 | 0.17 | 0.41 |
| Is the barangay accessible to the national highway? | T2 | 0.78 | 0.22 | 1.22 | 1.48 | 1.65 | 0.17 | 0.41 |
| Is there a cemetery in the barangay? | S3 | 0.22 | 0.78 | 1.79 | 3.19 | 3.35 | 0.16 | 0.40 |
| Is there a postal service in the barangay? | S9 | 0.20 | 0.81 | 1.81 | 3.28 | 3.42 | 0.14 | 0.38 |
| Is there a market or building with trading activities in the barangay? | SE1 | 0.18 | 0.82 | 1.82 | 3.33 | 3.47 | 0.14 | 0.38 |
| Is there a church/chapel or mosque in the barangay? | R1 | 0.83 | 0.17 | 1.18 | 1.38 | 1.52 | 0.14 | 0.37 |
| Is there a high school in the barangay? | E2 | 0.18 | 0.83 | 1.83 | 3.35 | 3.48 | 0.13 | 0.37 |
| Is there a newspaper circulation in the barangay? | S6 | 0.17 | 0.83 | 1.84 | 3.37 | 3.50 | 0.13 | 0.36 |
| Is there a barangay hall in the barangay? | S5 | 0.88 | 0.12 | 1.13 | 1.27 | 1.37 | 0.10 | 0.32 |
| Is there a housing project (government or private) in the barangay? | HO1 | 0.10 | 0.90 | 1.91 | 3.63 | 3.71 | 0.08 | 0.27 |
| Is there a town/city hall or provincial capital in the barangay? | S1 | 0.07 | 0.93 | 1.93 | 3.73 | 3.79 | 0.06 | 0.25 |
| Is there a telegraph in the barangay? | S8 | 0.08 | 0.93 | 1.93 | 3.74 | 3.79 | 0.05 | 0.22 |
| Is there a hospital in the barangay? | H1 | 0.06 | 0.95 | 1.95 | 3.81 | 3.85 | 0.03 | 0.19 |
| Is there a college/university in the barangay? | E3 | 0.04 | 0.96 | 1.97 | 3.87 | 3.89 | 0.03 | 0.16 |
| Is there a public library in the barangay? | S4 | 0.04 | 0.96 | 1.97 | 3.87 | 3.89 | 0.02 | 0.14 |

Table 6.1 shows the categorical variables arranged in ascending order of their variation across geographical space. The dispersion of continuous variables was determined using the standard deviation statistics as well. However, in this case the more common method of deriving the arithmetic mean was used. Table 6.2 shows some variables with very low geographic distribution.

**Table 6.2: Continuous Variables with the Lowest Standard Deviations**

| Variable | SD |
|---|---|
| Insurance and pension funding establishments | 0.02 |
| Age 91 | 0.02 |
| Manufacture of office, accounting and computing machinery establishments | 0.02 |
| Manufacture of coke, refined petroleum and other fuel products establishments | 0.02 |
| Five married couples | 0.02 |
| Manufacture of tobacco products establishments | 0.01 |
| Age 93 | 0.01 |
| Age 95 | 0.01 |
| Age 94 | 0.01 |
| Age 96 | 0.01 |
| Age 98 | 0.01 |
| Age 97 | 0.01 |
| Air transport establishments | 0.01 |
| Age 100 | 0.01 |
| Six married couples | 0.01 |

While age is very important especially in explaining life-stage activities, Table 6.2 above shows that single year age groups are less likely to provide much discrimination across geographical space. A solution to this problem was to create composites of the various age groups. The greater the value of the standard deviation, the better the variable is for inclusion in the clustering algorithm.

## 6.5.3 Measuring Association

One of the most common statistics used in the analysis of categorical data is the Pearsons chi-square ($\chi^2$) test (Johnson et al., 1994; Plackett, 1983). The chi-square is often used to test for goodness of fit between a theoretical and a frequency distribution. It is also used to investigate dependency between variables. However, two major short comings of the chi-square are that it does not give an indication of the direction of the association (i.e. positive or negative) between variables and is not indicative of the strength of the relationship.

A more useful measure of the strength of association in categorical data is to calculate the proportional difference between concordant (**P**) and discordant (**Q**) pairs (Agresti and Finlay, 2008) and derive a measure called the gamma using equation 6.1 below.

$$\gamma = (\mathbf{P}-\mathbf{Q}) \div (\mathbf{P}+\mathbf{Q}) \tag{6.1}$$

Gamma ($\gamma$) can take values ranging from -1 to +1. A value closer to +1 indicates a positive (concordant) association, while a value closer to -1 indicates a negative (discordant) relationship.

To derive the gamma, a crosstab query was run on the categorical variables to derive two-by-two matrices for each pair. Each of the two-by-two matrices was ordered by ***yes*** and ***no*** as shown in Table 6.3.

**Table 6.3: Crosstabulation of Two Categorical Variables**

| **Variables** | | **Is there an elementary school in barangay?** | |
|---|---|---|---|
| | | Yes | No |
| **Is there a health centre in barangay?** | Yes | 21141 | 4464 |
| | No | 9498 | 6347 |

In the example above, 50.4% (21141) of all barangays have both a health centre and an elementary school while 15.3% (6347) have none. The concordant (**P**) for the table was derived by multiplying 21141 by 6347 while the discordant (**Q**) was derived by multiplying 4464 by 9498. The ratio of the difference and sum of these two statistics gave the gamma. The strength and direction of the association between the different pairs of variables is shown in the matrix below.

**Table 6.4: Gamma Statistic Matrix for Categorical Datasets**

| | D1 | E1 | E2 | E3 | H1 | H2 | HO1 | R1 | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | SE1 | T1 | T2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D1 | | | | | | | | | | | | | | | | | | | | | | |
| E1 | -0.23 | | | | | | | | | | | | | | | | | | | | | |
| E2 | 0.32 | 0.53 | | | | | | | | | | | | | | | | | | | | |
| E3 | 0.63 | 0.22 | 0.91 | | | | | | | | | | | | | | | | | | | |
| H1 | 0.54 | 0.03 | 0.67 | 0.85 | | | | | | | | | | | | | | | | | | |
| H2 | 0.03 | 0.52 | 0.43 | 0.30 | 0.48 | | | | | | | | | | | | | | | | | |
| HO1 | 0.25 | 0.06 | 0.48 | 0.68 | 0.55 | 0.20 | | | | | | | | | | | | | | | | |
| R1 | -0.06 | 0.64 | 0.36 | 0.20 | 0.19 | 0.52 | 0.08 | | | | | | | | | | | | | | | |
| S1 | 0.78 | 0.01 | 0.67 | 0.72 | 0.75 | 0.32 | 0.36 | 0.21 | | | | | | | | | | | | | | |
| S2 | 0.10 | 0.24 | 0.37 | 0.39 | 0.34 | 0.28 | 0.24 | 0.40 | 0.58 | | | | | | | | | | | | | |
| S3 | 0.06 | 0.31 | 0.46 | 0.29 | 0.30 | 0.14 | 0.08 | 0.49 | 0.47 | 0.17 | | | | | | | | | | | | |
| S4 | 0.51 | 0.21 | 0.67 | 0.77 | 0.76 | 0.56 | 0.49 | 0.27 | 0.82 | 0.63 | 0.30 | | | | | | | | | | | |
| S5 | -0.02 | 0.46 | 0.32 | 0.12 | -0.36 | 0.61 | -0.04 | 0.36 | -0.09 | 0.38 | -0.17 | 0.19 | | | | | | | | | | |
| S6 | 0.42 | -0.23 | 0.38 | 0.66 | 0.59 | 0.22 | 0.55 | -0.16 | 0.45 | 0.07 | -0.16 | 0.61 | 0.20 | | | | | | | | | |
| S7 | 0.54 | -0.25 | 0.46 | 0.79 | 0.66 | 0.26 | 0.59 | -0.12 | 0.58 | 0.19 | -0.16 | 0.66 | 0.35 | 0.86 | | | | | | | | |
| S8 | 0.74 | -0.14 | 0.65 | 0.80 | 0.78 | 0.27 | 0.59 | 0.00 | 0.87 | 0.44 | 0.22 | 0.80 | -0.01 | 0.83 | 0.86 | | | | | | | |
| S9 | 0.43 | 0.00 | 0.54 | 0.67 | 0.65 | 0.39 | 0.46 | 0.06 | 0.76 | 0.26 | 0.09 | 0.72 | 0.25 | 0.87 | 0.77 | 0.94 | | | | | | |
| S10 | 0.32 | 0.00 | 0.34 | 0.54 | 0.44 | 0.24 | 0.36 | 0.16 | 0.42 | 0.18 | 0.10 | 0.53 | 0.26 | 0.55 | 0.55 | 0.63 | 0.51 | | | | | |
| S11 | 0.30 | 0.02 | 0.48 | 0.69 | 0.55 | 0.43 | 0.42 | 0.08 | 0.38 | 0.21 | -0.23 | 0.67 | 0.45 | 0.85 | 0.90 | 0.81 | 0.81 | 0.52 | | | | |
| SE1 | 0.38 | 0.20 | 0.63 | 0.62 | 0.63 | 0.42 | 0.34 | 0.35 | 0.73 | 0.39 | 0.48 | 0.64 | 0.14 | 0.33 | 0.39 | 0.66 | 0.50 | 0.41 | 0.19 | | | |
| T1 | 0.58 | 0.01 | 0.49 | 0.74 | 0.61 | 0.22 | 0.45 | 0.09 | 0.67 | 0.38 | 0.19 | 0.62 | 0.22 | 0.61 | 0.67 | 0.78 | 0.60 | 0.37 | 0.52 | 0.49 | | |
| T2 | 0.26 | 0.01 | 0.31 | 0.58 | 0.46 | 0.33 | 0.35 | 0.14 | 0.32 | 0.12 | -0.14 | 0.61 | 0.31 | 0.68 | 0.68 | 0.57 | 0.59 | 0.32 | 0.70 | 0.30 | 0.17 | |

where

D1 = Is barangay part of town city proper or poblacion?
E1 = Is there an elementary school in the barangay?
E2 = Is there a high school in the barangay?
E3 = Is there a college/university in the barangay?
H1 = Is there a hospital in the barangay?
H2 = Is there a puericulture center/health center in the barangay?
HO1 = Is there a housing project (government or private) in the barangay?
R1 = Is there a church/chapel or mosque in the barangay?
S1 = Is there a town/city hall or provincial capital in the barangay?
S2 = Is there a public plaza or park in the barangay?
S3 = Is there a cemetery in the barangay?
S4 = Is there a public library in the barangay?
S5 = Is there a barangay hall in the barangay?
S6 = Is there a newspaper circulation in the barangay?
S7 = Is there a telephone in the barangay?
S8 = Is there a telegraph in the barangay?
S9 = Is there a postal service in the barangay?
S10 = Is there a community waterworks system in the barangay?
S11 = Is there electric power in the barangay?
SE1 = Is there a market or building with trading activities in the barangay?
T1 = Does the barangay have a street pattern or network of streets of at least 3 street roads?
T2 = Is the barangay accessible to the national highway?

Generally, there is very weak positive association across the dataset. The red shadings indicate variables demonstrating high levels of positive correlation. For the purpose of this work Gamma values of +0.7 and above have been defined as high. For instance, there is a very strong positive relationship between the presence or absence of a telegraph and postal services. This positive relationship indicates that the presence of one service in a barangay may suggest the presence of the other and the absence of one may suggest the absence of the other. This may also suggest that users of telegraph services may prefer to live in a neighbourhood where there is also a postal service. Including multiple variables with strong levels of association can result in redundant information which can cloud other underlying characteristics.

The product moment correlation coefficient was used to explore strength and direction of relationships within the continuous datasets. As discussed in the previous chapter, strong correlations in any direction are generally undesirable for cluster analysis because they are indicative of redundant information.

**Table 6.5: Excerpt of Cross-correlations across Employment Domain**

| | EM1 | EM2 | EM3 | EM4 | EM5 | EM6 | EM7 | EM8 |
|---|---|---|---|---|---|---|---|---|
| EM1 | | | | | | | | |
| EM2 | 0.74 | | | | | | | |
| EM3 | 0.75 | 0.80 | | | | | | |
| EM4 | 0.76 | 0.83 | 0.88 | | | | | |
| EM5 | 0.58 | 0.60 | 0.67 | 0.68 | | | | |
| EM6 | -0.61 | -0.64 | -0.74 | -0.71 | -0.60 | | | |
| EM7 | 0.49 | 0.52 | 0.59 | 0.61 | 0.50 | -0.64 | | |
| EM8 | 0.57 | 0.63 | 0.72 | 0.70 | 0.56 | -0.74 | 0.76 | |

Where

EM1 = Legislators, managers and senior officials
EM2 = Professionals
EM3 = Technicians and associate professionals
EM4 = Clerks
EM5 = Service workers
EM6 = Agricultural and fishery workers
EM7 = Craft workers
EM8 = Plant and machine operators

Due to the mass of the dataset, each domain was initially examined independently. Some high correlations were found within the dataset. From Table 6.5 the presence of high level professionals can also be indicative of the presence of senior officials. Likewise, the presence of technicians can be indicative of the absence of agricultural and fishery workers.

Each pair of relationships was examined with care. It is easy to conclude that one of each highly correlated pairs of variables should be dropped. However, other characteristics of the variables could suggest otherwise.

### 6.5.4 The Analysis of Principal Components

Principal Components Analysis (PCA) was used to investigate which variables are likely to have the greatest influence on the classification. The components loading matrix for each principal component was examined.

The PCA conducted for both categorical and continuous variables. The objective of the analysis was to determine the variables which are likely to power the dataset. Although the components loading matrices of the first three components were examined, there had to be trade-off on which component would be used to determine the variables that power the dataset.

**Table 6.6: Loadings of the First Principal Component in Descending Order**

| Categorical Variable | Loading |
|---|---|
| Is there a postal service in the barangay? | 0.65 |
| Is there a telephone in the barangay? | 0.65 |
| Is there a telegraph in the barangay? | 0.62 |
| Is there a newspaper circulation in the barangay? | 0.58 |
| Does the barangay have a street pattern or network of streets of at least 3 street roads? | 0.54 |
| Is there a high school in the barangay? | 0.49 |
| Is there a town/city hall or provincial capital in the barangay? | 0.48 |
| Is there a market or building with trading activities in the barangay? | 0.44 |
| Is there a college/university in the barangay? | 0.43 |
| Is there electric power in the barangay? | 0.43 |
| Is there a community waterworks system in the barangay? | 0.43 |
| Is there a hospital in the barangay? | 0.43 |
| Is Barangay part of town city proper or poblacion? | 0.42 |
| Is there a public library in the barangay? | 0.41 |
| Is the barangay accessible to the national highway? | 0.35 |
| Is there a housing project (government or private) in the barangay? | 0.33 |
| Is there a puericulture center/health centre in the barangay? | 0.30 |
| Is there a public plaza or park in the barangay? | 0.29 |
| Is there a barangay hall in the barangay? | 0.14 |
| Is there a cemetery in the barangay? | 0.12 |
| Is there a church/chapel or mosque in the barangay? | 0.10 |
| Is there an elementary school in the barangay? | 0.04 |

The first principal component accounts for the greatest possible proportion of the variance of the variable set while the second accounts for the maximum remaining variance. Essentially later components explain less of the variation within the data (Dunteman, 1989; OECD, 2008). The first component was therefore used to determine which variables power the dataset.

Table 6.6 shows the results of the analysis conducted for the categorical variables. The presence or absences of a postal and telephone service are the two categorical indicators which are likely to power the dataset. This implies that 42% of the variance existing within the variable is explained by component 1.

Amongst continuous variables, population without access to piped water has the highest loading for component 1. Other highly important variables include population employed in public administration, free occupancy take-up, houses built with bamboo and retail trade establishments.

## 6.5.5 Coping with the Problem of Skew

It is desirable to include only normally distributed variables in the clustering algorithm (Vickers, 2006; Harris et al., 2005). However this is usually not the case with socio-demographic datasets. Amongst the continuous datasets, 80% of the variables are defined by a positive skew while 20% have a negative skew.

**Table 6.7: Variables with the largest positive Skew**

| Variable | Skew |
|---|---|
| Ancillary Unit establishments | 58.33 |
| Financial intermediation establishments | 50.1 |
| Integrated paper and paper products establishments | 45.59 |
| Other Manufacturing establishments | 42.42 |
| Recycling establishments | 38.62 |
| Philippines benevolent missionaries | 36.54 |
| Construction establishments | 33.49 |
| Forestry, logging and related service activities establishments | 31.58 |
| Electricity, gas, steam and hot water supply establishments | 30.57 |
| Electricity, Gas and Water Supply establishments | 29.51 |

Variables within the employment domain showed some of the largest positive skews. With the exception of Philippines benevolent missionaries (religion domain) every other indicator in Table 6.7 above is within the employment domain. One reason for this is the number of categories of employment. Clearly the more the categories, the more the data is split within these categories and the more the chance of zero values for some barangays. This could result when in the indicator concentrating within a limited number of areas. It could also result create the presence of outliers (Harris et al., 2005).

**Table 6.8: The Effect of Composite Variables on Positive Skew**

| Initial Variable | Skew | | Composite Variable | Skew |
|---|---|---|---|---|
| Age 3 | 0.64 | | Age 0 to 4 | 0.03 |
| Age 5 | 0.27 | | Age 5 to 9 | -0.18 |
| Age 18 | 3.12 | | Age 10 to 19 | -0.61 |
| Age 20 | 1.86 | | Age 20 to 44 | 0.16 |
| Age 50 | 1.42 | | Age 45 to 64 | -0.44 |
| Age 99 | 10.56 | | Age 65+ | 0.32 |

The problem of skew within the dataset was addressed by creating composite variables where the variables share the same denominator. For instance, some of the one year age variables especially in the older age categories demonstrated high levels of skew as high as 10.56. The one year age groups were combined into six age groups as shown in Table 6.8. Apart from increasing the standard deviations of the variables, the positive skews were also minimized. Another solution was to transform the dataset using a log transformation explained in section 6.7.

### 6.5.6 Sustaining the Longevity of the Classification

It is also important to ensure that variables included in the clustering algorithm sustain the life-course of the classification (Vickers and Rees, 2006). This is because it will be used to guide decision making processes some of which may have significant implications. All variables considered for inclusion in the Philippines classification were derived from the census which is decennial exercise. Though another census was conducted in 2007, it was not as comprehensive as the 2000 census and results from it have not been disaggregated to small area geographies.

Previous national census and survey questionnaires were secured to assess the nature and types of questions asked in order to gain an understanding of questions that were consistent. Embarking on this qualitative exercise was useful in making decisions on some of the variables to exclude. While it can be argued that this does not necessarily guarantee that future censuses will not change in the questions they ask, such change can not be expected to be very rapid. In such scenarios, proxy indicators can be used. The ability to update variables which appear not to change in space over time will sustain the life course of the classification.

Some users especially within the private sector would probably be interested in indicators defining levels of affluence which can be useful in helping them gain a competitive advantage (Birkin and Clarke, 1998). While there was no information on income just like the Nigerian classification, some other indicators on socio-economic status were included. Other proxy indicators especially under the employment domain can also be suggestive of levels of affluence. The beauty of geodemographics however is that users can profile their own pool of data against the system and still gain insight which can be useful for local or national planning. For instance if a private user is looking to expand their business, they can use the system to find where might be suitable and gain insight on profitability based on known locations of their current customers or clients (Harris et al., 2005)

### 6.5.7 Relevance to Policy Issues

The process of policy making is the means by which governments convert their political vision into actions (Lindblom and Woodhouse, 1992). Of particular policy relevance especially to developing countries is its potential application towards making better and relevant policies. While the lack of political will has often been blamed in many developing countries including the Philippines for slow growth (Montinola, 1999) some policy makers may be well intentioned but may be misguided.

The choice of variables included in the Philippines classification took cognizance of interest in current policy debates especially as they relate to the targets of the 8 MDGs. An example is drawn from the Medium Term Philippine Development Plan (MTPDP) which seeks to reduce poverty. It has been argued that policies focused on the dynamics of the population will be useful in achieving this feat (Schelzig, 2005). Clearly the scourge of poverty is multifaceted hence the fight against it requires a multi-dimensional approach which takes the types and levels of advantage or disadvantage of different groups into context.

The Philippines Department of Education also has quite a number of policy programmes in place. The Third Elementary Education Project (TEEP) for instance was aimed at curtailing the threat of illiteracy and speed up achievement and completion rates amongst school goers especially in poverty stricken areas (World Bank, 2007c). A key challenge for the future is the sustenance of some of the positive outcomes of project.

An argument of this research queries how the targeted provinces for the TEEP were chosen. In the view of the World Bank, poverty-plagued areas were the best to focus the primary education programme. However, just because an area is classed as poor does not make the residents educationally disadvantaged. Additionally, low rates of primary school completion may be linked to multiple factors like high teenage pregnancy rates.

## 6.6 The Final List of Variables

The selection of input variables was a lengthy and tasking process which required making complex decisions. Outside the framework of the statistical consideration, some decisions are subjective while others are based on experience and knowledge of the study area. The judgments made in the selection process were carefully considered and are by no means finite. A different classification developer for the Philippines may derive an entirely different list of indicators. The following subsections provide a summary of the final list of indicators.

***Demographic Variables***

Demographic indicators provide insight to the direct characteristics of the population. Knowledge of basic demographic characteristics typically serves as a foundational step to reaching the right people when communicating a message or providing a service.

The best information on urban or rural status was derived from a proxy variable on whether barangays are part of a city or poblacion. Poblacion is used to describe an area with town or city-like features. Population density is included as it provides information on the disproportionate concentration of people. Marital information and information on the head of the household were also included.

Age indicators were combined to reflect young children in pre-primary education (aged 0 to 5), older children in primary education (aged 6 to 10), adolescents in secondary and post secondary education (aged 11 to 20), young adults in their early career phases (aged 21 to 45), older adults nearing retirement (aged 45 to 65) and the elderly, many of whom are expected to be retired (aged 65+). Other demographic information included is the percentage of widows and information on language – specifically those who speak Filipino.

***Education Variables***

Variables on education were included to help reveal and understand the spatial patterns of educational attainment of the population. Education is also indicative of levels of affluence. The attainment of universal education at primary level is second on the list of 8 MDGs stressing the importance of the indicators within this domain for this work.

Information on the presence or absence of an elementary and high school was included. The presence or absence of either of these in an area may not necessarily imply the high attendance or completion rates in the area. It was therefore considered important to include information on the completion rates of basic primary education and the attainment of post secondary technical education.

### *Employment Variables*

Variables within this category inform the key sources of income for the population. There are also relationships between unemployment, poverty, crime and happiness (Gardner and Oswald, 2002). Indeed some studies have shown that overall well-being can significantly improve when an individual moves from a state of unemployment into employment (Clarke, 2003) and that employment-related major life events are particularly important determinants of subjective happiness (Ballas and Dorling, 2007). Another potentially useful information that employment can uncover is the spatial variation in the levels of educational attainment.

Different types of employment indicators were included ranging from information on senior professionals to transportation and agro-allied employments. The different employment indicators are important for understanding the different sources of income which can be related to the types and levels of poverty. Poverty reduction is number one of the list of MDGs.

The sizes of employment establishments were also considered but due to high levels of correlations between large and medium sized establishments only the small sized establishments were included. Small sized establishments also serve as pointers for micro-financing.

### *Health Variables*

There is increasing interest in the understanding and analysis of people's health by where they live (Wood et al., 2006; Webber, 2004b). For developing world countries like the Philippines, health and well - being are important policy issues and are typically high on the agenda of the government (Furtado, 2001).

As seen in Chapter 3, amongst the eight MDGs, three focus on health-related issues ranging from child health, to maternal health and HIV/AIDS and other diseases. This underscores the importance of this domain for this research project. However, just like the case of the Nigerian classification, the health domain represents some of the fewest variables. There is a challenge with securing reliable and timely health statistics in many developing world countries (Okonjo-Iweala, 2007). The selected variables include the presence or absence of a hospital or health centre and information on the less able population.

***Household Infrastructure Variables***

Quality of life and living standards can be assessed using essential household amenities. Many household infrastructural facilities are basic needs which drive other aspects of livability (Diallo and Wodon, 2007).

Information on access to electricity, water and cooking fuels were included as they are important essentials for today's modern lifestyles. Also included are television and telephone usage which could serve useful proxies for assessing the level of exposure to current information. The type of toilets used was also included as it can have links with healthy lifestyles.

***Housing Variables***

Numerous characteristics can be revealed from dwelling or housing unit data. Dwelling ownership, occupancy types and status, types of building materials and size of dwelling are informative for profiling affluence levels. Such information can also reveal information on the concentration of certain groups of people like single-person households or young couples. In addition, housing information can provide knowledge on the spatial variations of ageing apartments.

***Religion Variables***

Religion forms a major part of the socio-cultural dynamics of the people of Philippines (Woods, 2007). It can be indicative of cultural diversity which in turn provides pointers towards attitudinal and consumption patterns (Wu, 2004). Four indicators on religion have been included reflecting the principal dimensions of religion recognised in the country. These include Christianity which has been split to uncover areas dominated by Roman Catholics and other Christians. The other two religion indicators are Islam and Buddhism.

***Services Variables***

Public services infrastructures play essential roles in increasing employment opportunities and bolstering economic growth. In addition, they tend to attract population and location of firms (Martin and Rogers, 1995). Some services infrastructures also serve as centres for communal engagements which strengthen socio-cultural ties.

The services variables that were included are essentially in categorical form specifying the presence or absence of a town hall, public plaza, cemetery, telegraph service or postal service. Information is also included on the methods of waste disposal and the presence or absence of a waterworks system.

### *Socio-economic Variables*

Socio-economic variables are very useful proxies for measuring and understanding trends in poverty and well-being. One of the strongest surrogates for spatially targeting concentrations of poor people is information on income. However, this information can be very difficult to secure and is often subject to disclosure controls. Income information was not obtained for inclusion in the development of the Philippines classification.

The key socio-economic proxies are assessed using information on the presence or absence of a market, the number of manufacturing and auto repair establishments. The spread of restaurants and beauty parlours was also included.

### *Transportation Variables*

Transportation variables provide interesting insights about the development and expansion of settlements. When combined with some other indicators, they can reveal greater detail on the level of urbanization.

Transportation services are important in the Philippines because of the nature of the terrain. Two transportation indicators – the presence or absence of at least 3 street roads and accessibility to the highway were included.

Table 6.9: The Final List of 69 Variables

| Variable | Domain | | Variable | Domain |
|---|---|---|---|---|
| Part of city or poblacion | Demographic | | Electricity supplied | Household Infrastructure |
| Population density | Demographic | | No piped water | Household Infrastructure |
| No married couples | Demographic | | Cooking energy - liquid fuels | Household Infrastructure |
| Two or more married couples | Demographic | | Cooking energy - wood | Household Infrastructure |
| Stepchild of head | Demographic | | No telephone | Household Infrastructure |
| Age 0 to 5 | Demographic | | Television | Household Infrastructure |
| Age 6 to 10 | Demographic | | Water closet toilet | Household Infrastructure |
| Age 11 to 20 | Demographic | | Latrine | Household Infrastructure |
| Age 21 to 44 | Demographic | | Dwelling owned | Housing |
| Age 45 to 64 | Demographic | | Members squatting | Housing |
| Age 65+ | Demographic | | Free occupancy | Housing |
| Widowed | Demographic | | Land is occupied with consent | Housing |
| Speaks Filipino | Demographic | | Built with bamboo | Housing |
| Elementary school | Education | | Built with bricks, stone or concrete | Housing |
| High school | Education | | Built with mixed materials | Housing |
| Less than primary education | Education | | Iron and concrete roofing | Housing |
| Post secondary technical education | Education | | Buddhist | Religion |
| Professionals and Senior Officials | Employment | | Muslim | Religion |
| Agricultural and fishery workers | Employment | | Roman Catholic | Religion |
| Transportation and communications employment | Employment | | Other Christians | Religion |
| Public administration and defense employment | Employment | | Town hall | Services |
| Education employment | Employment | | Public plaza or park | Services |
| Health and social work employment | Employment | | Cemetery | Services |
| Overseas worker | Employment | | Barangay hall | Services |
| Real estate and business establishments | Employment | | Telegraph service | Services |
| Private establishments | Employment | | Postal service | Services |
| Co-operative establishments | Employment | | Waterworks system | Services |
| Small establishments | Employment | | Method of waste water disposal | Services |
| Media establishments | Employment | | Market | Socio-economic |
| Retail trade establishments | Employment | | Manufacturing establishments | Socio-economic |
| Banking institutions | Employment | | Auto repair shops | Socio-economic |
| Recreational establishments | Employment | | Restaurants and personal services | Socio-economic |
| Hospital | Health | | At least 3 street roads | Transportation |
| Health centre | Health | | Accessibility to the highway | Transportation |
| Disabled | Health | | | |

Table 6.9 shows the list of the 69 variables selected for further analysis and inclusion in the subsequent clustering algorithm. The process of variable selection entailed making complex decisions. It is almost impossible to narrate the entirety of the issues taken into consideration in the process.

The initial list comprised a total of 435 variables with the demographic and employment domains recording the highest proportions of 31% and 25% respectively. The other eight domains shared the rest of the 44% with the transportation and socio-economic domains accounting for low values of 1% and 2% respectively.

At the end of the first phase (intra-domain variable reduction/selection) of the selection process 81% of the initial variables were rejected reducing the number to 81 variables. At this stage, the proportion of all variables accounted for by the demographic and employment domains had reduced to 44%. The final phase of reductions took cognizance of the merged variables and inter-domain analysis. Another 15% of the variables were rejected. The final list comprises a total of 69 variables. The employment and demographic domains account for the largest proportions of 22% and 19% respectively. Three domains –household infrastructure, housing and services each account for 12%. The education, religion and socio-economic domains account for 6% each. Just like the Nigerian classification, health has a low representation of 4%. The least representation is demonstrated by the transportation domain which only accounts for 3% of all variables.

## 6.7 Segmenting the Barangays

The eventual selection of 69 variables meant the research could proceed to a new phase of multivariate analysis. All the selected inputs were assembled into a single database and manually checked for any errors. Discussions in the following sections would therefore focus on justifying the technical choices made to reduce the effects of skew (especially positive skew); to normalise the variables to the same scale; and to cluster the variables and evaluate the characteristics of cluster formations.

### 6.7.1 Transformation of Inputs

It has already been stressed that the type of variables desired for inclusion in the clustering algorithm are variables that present a bell shape or normal distribution (Harris et al., 2005). It is not uncommon with socio-demographic statistics that values sometimes concentrate at one end of the measurement scale. Many times it is often the variables which appear to be of seemingly unique interest that exhibit this attribute. A typical example is population density.

A number of transformation methods including log transformation, square root transformation, reciprocal transformation and square transformation were tested on the dataset. Each method was assessed on how much it reduced the positive skew within the data. Table 6.10 shows the results of the analysis on a number of variables across different domains.

**Table 6.10: Assessing Skew Before and After Transformation**

| Variables | Non Transformed | Log Transformation | Square Root Transformation | Reciprocal Transformation | Square Transformation |
|---|---|---|---|---|---|
| Less than primary education completed | 2.80 | 0.88 | 1.71 | 0.48 | 5.81 |
| Disabled population | 1.13 | -0.92 | 0.20 | 9.68 | 3.94 |
| Free occupancy | 1.11 | -0.37 | 0.42 | 6.18 | 2.58 |
| Muslims | 3.44 | 0.57 | 2.93 | 2.52 | 3.79 |
| Number of restaurants and beauty parlors | 2.38 | 2.02 | 1.70 | 2.71 | 2.96 |

After examining the results from the exercise, a logarithmic transformation was applied to reduce the effect of skew within the dataset. The method was used because of its ability to cope well with positive skew. Vickers (2006) also found out that this method worked well with UK Output Area census data.

### 6.7.2 Standardisation of Variables

Prior to proceeding with clustering any dataset, it is important to consider the different scales of measurement used for the input variables. This is because clustering variables measured in different units and between different scales will not yield the true picture of the areas (Vickers, 2006). Undue advantage will be given to variables measured within a larger margin as they would appear to have greater impact on the clustering algorithm.

The experience drawn from working with the Nigerian dataset was brought in to play when deciding on the method of standardisation. In the Nigerian classification, z-scores were used but they had to be capped within a range of +3 and -3 because of the effects of some extreme values (see also Vickers, 2006). Standard normal variate scores (z-scores) and range standardisation techniques were explored. After several tests - details of which are narrated in section 6.7.3, a z-score standardisation was adopted.

### 6.7.3 The Clustering Algorithm

Once the dataset to be clustered had been prepared, a choice of clustering algorithm had to be made. The size of the data meant that an algorithm that could handle the volume of data. The other issue to consider was the fact that the database was a combination of categorical and continuous datasets. Most clustering algorithms work well with continuous datasets. Detailed discussion on the different algorithms is contained in chapter 4, section 4.7.

An algorithm specifically designed to handle combinations of categorical and continuous datasets is the Two-Step Cluster analysis procedure (Banerjee et al., 2004). The procedure gives the best results if categorical datasets appear to have a multinomial distribution and continuous variables display a normal distribution. This is often not the case with socio-demographic datasets but by transforming the data with a logarithmic method, the effect of outliers has been greatly reduced.

All clustering algorithms group cases on the basis of similarity or dissimilarity. Similarity of cases within taxonomic space is measured by deriving a statistical quantification of distance (Everitt et al., 2001). Generally, similar cases have a closer distance. A likelihood ratio test has been recommended by Agresti (2007) as a reliable test for significance when dealing with categorical datasets. The log-likelihood ratio test was used to evaluate similarity within the datasets. It is based on probabilities and compares the maximised likelihoods of a null hypothesis to an alternative one. The larger the value of the statistic, the less the within-cluster variation and the more compact cases are unlike the euclidean distance used in the Nigerian classification which requires small values for increased within-cluster similarity.

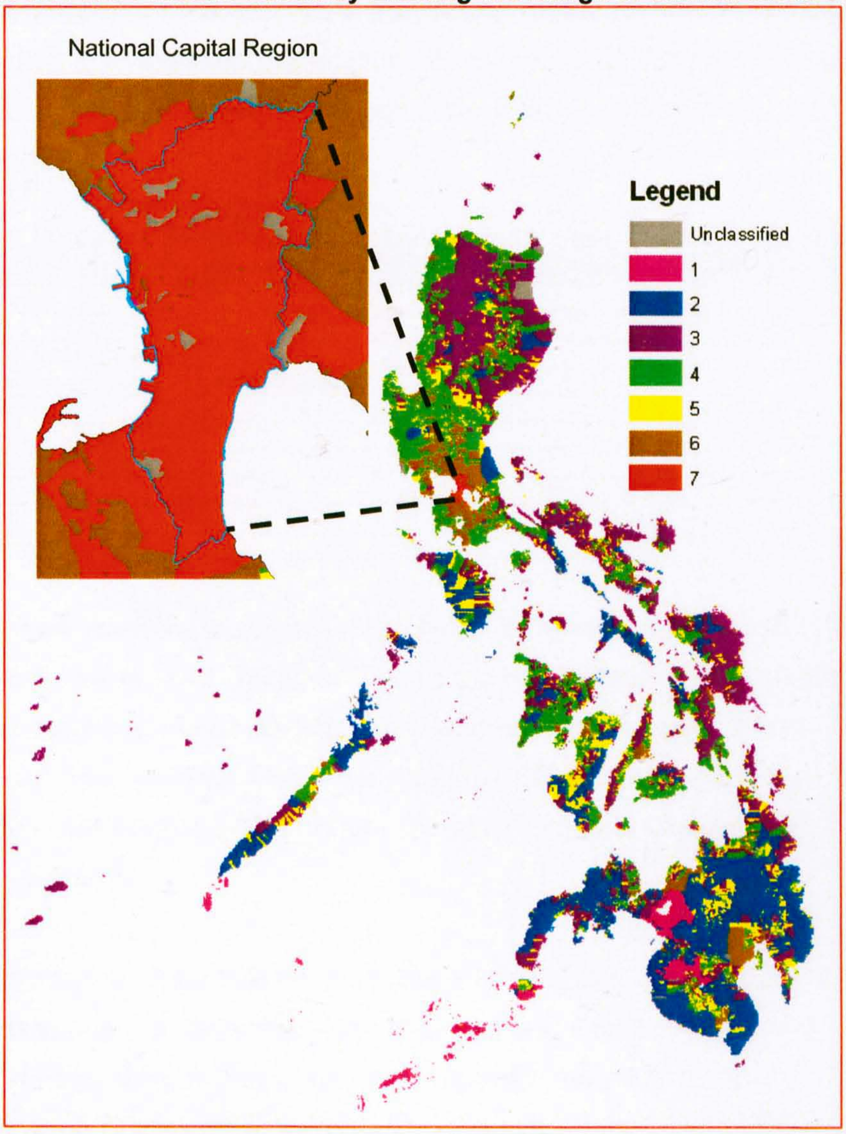
## 6.7.4 Problems with the Initial Clustering

The analysis was deployed with the hope of creating a hierarchy similar to the Nigerian classification. The top-down methodology employed by Vickers (2006) in creating the hierarchy of the U.K. output area classification was adopted because it ensures cluster groups in lower hierarchies maintain as much traits of their parent clusters.

When the algorithm was run on the database, the first observation was that the massive imbalance in the distribution of cases within clusters. Too few cases were concentrated in some clusters. This problem was caused by the categorical datasets and the way they were coded. In the file supplied by the NSO, '*yes*' answers were coded with 1; '*no*' answers were coded with 2; '*unknown*' answers were coded with 9. There were also codings of 0 which did not represent any response. These were the cases causing the skew in the solution. It was therefore decided that since these cases did not define anything, 0 values should be coded with 9 (unknown) and the algorithm was re-run.

A re-run of the algorithm presented two benefits. First it reduced the problem of relative imbalance in the number of cases within each cluster solution and it was able to uncover Barangays which had characteristics which were too different from other clusters. However there was a downside to the solution. It failed to uncover within-region variations. For instance, the entirety of the National Capital Region was put into a single cluster as shown in Figure 6.7.

**Figure 6.7: The Problem Posed by Limiting the Range of Standardised Values**




It was easy to observe that this solution did not fit the reality- especially given that the analysis is being conducted at a fine geographical scale. The algorithm was run several times on the data altering the number of clusters but this did not change the clumping of clusters.

## 6.7.5 Adopting a Revised Solution

At this point the effect of the method used to standardize the dataset was evaluated. While it was thought important to reduce the effect of outliers (Vickers, 2006) in the clustering algorithm hence the adoption of a range standardisation method, it appeared the effects of variables that could expose the diversity within areas (Voas and Williamson, 2001) were concealed. Range standardisation forces values between 0 and 1. In a sense, it would appear that imposing a limit implies some unintentional weighting is applied on the data such that larger values receive less weight. An easy way to test this is to examine the relationship in the dimensions of the original dataset and the standardised values. To demonstrate this, the spearman's rank correlation coefficient was computed.

**Table 6.11: Relationship between Standardised Values and the Original Dataset**

| Variable | Range Standardisation | Z-scores |
|---|---|---|
| Population density | 0.55 | 1 |
| Widowed | 0.99 | 1 |
| Agricultural and fishery workers | 0.82 | 1 |
| Real estate, renting and business establishments | 0.94 | 1 |
| Electricity supplied | 0.92 | 1 |
| Other Christians | 0.78 | 1 |
| Retail trade establishments | 0.85 | 1 |
| Age 11 to 20 | 0.90 | 1 |
| Latrine present | 0.86 | 1 |
| Members squatting | 0.75 | 1 |

The experiment revealed a perfect relationship between z-scores and the dimension existing in the original data. Table 6.11 sows the results for 10 of the variables. While it may appear there are marginally high correlations in some cases, it was discovered in the course of this research that such marginal differences could have far reaching effects in the clustering process. When the algorithm was deployed on the z-scores, the results improved.

Initial exploration of the solutions from the z-scores was acceptable. However, after carefully assessing the centroids of the final clusters, it was observed that centroids of population density were significantly higher than for the other variables. This is one of the shortcomings of the non limitation of z scores. It was therefore decided that before re-running the algorithm, the weight of population density be reduced and capped between +2 and -2 as suggested by (Vickers, 2006). This range was chosen after examining the frequency distribution of the z-scores for the variable.

## 6.8 Evaluating the Final Solution

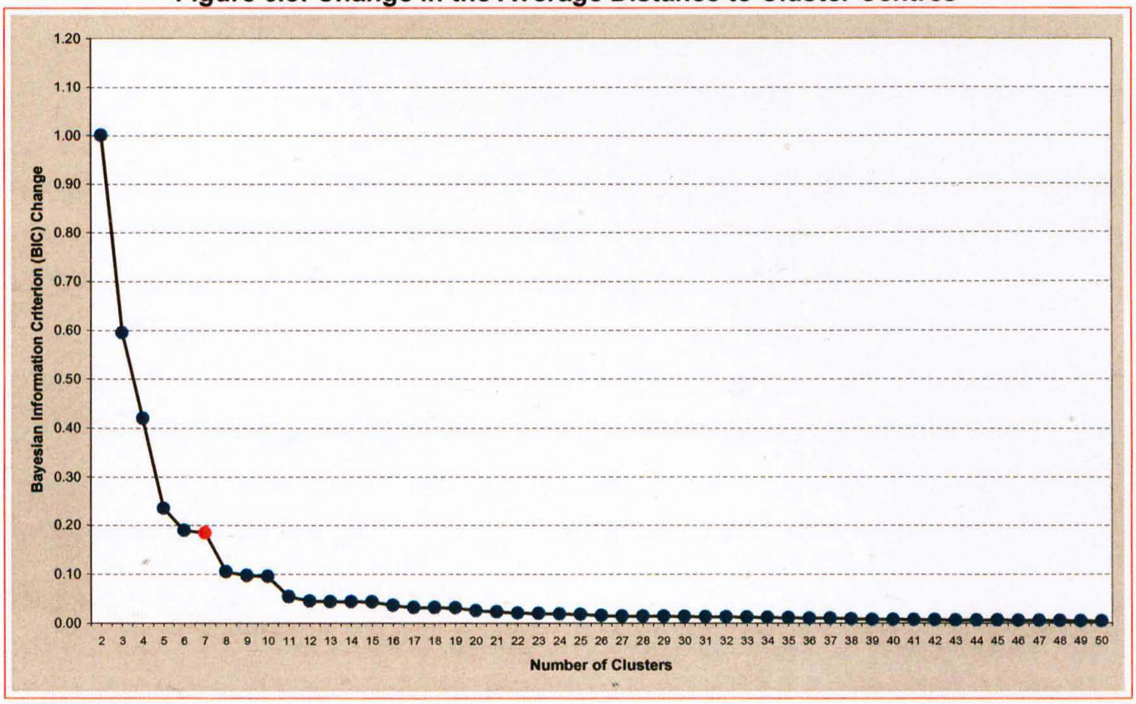
The experiments in the previous sections show that the solution to any cluster analysis exercise is partly hinged on the purpose of the classification (Everitt et. al, 2001). Additionally the results have also shown that while limiting the maximum values of standardised values was important in the Nigerian classification, this approach is only partially appropriate in the case of the Philippines.

In order to create the three tier hierarchy with the two-step clustering method, an acceptable number of clusters had to be derived at the topmost hierarchy. This again can be very daunting but one of the approaches suggested by Everitt et al. (2001) is to assess the clustering criterion against the number of clusters. Other issues to consider include the need for a balanced distribution of cases within clusters; a relatively balanced population distribution; an acceptable national distribution of clusters suitable for further analysis.

### 6.8.1 Clustering Criterion and the Number of Clusters

The clustering criterion selected for the exercise is the Schwarz Bayesian Information Criterion (BIC).

**Figure 6.8: Change in the Average Distance to Cluster Centres**



The change in BIC is basically the difference between log likelihood ratio statistic and the clustering parameters (Akaike et al., 1998). A perceived ideal solution would be the point at which there is an abrupt increase in the BIC (Banerjee et al., 2004; Larose, 2006). From Figure 6.8, this point is located at the 7 cluster solution after running the analysis for up to 50 clusters.

## 6.8.2 The Distribution of Cases and Population in Each Cluster

It is important for both the number of barangays assigned to each cluster to be reasonably balanced. This will ensure that the solution is suitable for further analysis (Vickers, 2006). For instance, it is expected that the solution will be used to explore datasets like surveys so that they can be extrapolated to derive a national picture of the phenomenon that is being explored. Usability of the classification will be increased if coded surveys are statistically representative in their distribution across the different clusters. Table 6.12 shows the extent to which clusters are sized in terms of Barangays and population.

**Table 6.12: Distribution of Barangays and population in Each Cluster**

| Cluster | Barangays (%) | Population (%) |
|---|---|---|
| Unclassified | 0.95 | 1.27 |
| 1 | 23.51 | 16.34 |
| 2 | 12.48 | 11.51 |
| 3 | 16.50 | 13.57 |
| 4 | 18.97 | 15.89 |
| 5 | 13.71 | 27.01 |
| 6 | 5.38 | 5.28 |
| 7 | 8.50 | 9.13 |

Based on the final solution, a total of 384 Barangays did not fit into any of the 7 clusters. These 384 Barangays were added to the 14 initially excluded due to lack of data to give 398 which constitute the unclassified group of cases.

Amongst the 7 clusters, the largest comprises 9862 Barangays while the least is made up of 2,258 Barangays giving a range of 7,604 Barangays. Population-wise, there is a range of 16.6 million people with the largest cluster comprising 20,650,941 people while the smallest has a population of 4,036,687 people.

It is also interesting to observe that the cluster with the largest number of Barangays is not the cluster with largest number of people. One factor accounting for that as is shown by exploring the characteristics of clusters in subsequent sections is the level of urbanisation of the cluster. The National Capital Region, Southern Tagalog and Central Luzon have the highest levels of urbanisation. Most Barangays within these three regions fall into cluster 5 as shown in Table 6.13 below. Cluster 5 as shown in Table 6.12 above also has the largest population. However urbanisation alone may not be the only factor accounting for the large population. It will be interesting to explore the penetration of family planning measures and religious factors within these areas. Some faith groups in the Philippines have been known to oppose some of these health measures (Ballweg, 1972).

## 6.8.3 Representative Distribution of Clusters at Regional Level

This analysis is being conducted at the finest level of administrative geography in the Philippines. One of the aims of the exercise is to be able to show that while at this scale, nearer Barangays are quite similar to each other in their geodemographic make up (Tobler, 1970), distant ones may also contain population groups with similar characteristics in spite of their non-contiguity (Vickers, 2006). Table 6.13 can be used to illustrate this position.

**Table 6.13: Percentage Distribution of Barangays by Regions and Clusters**

| Administrative Regions | Clusters | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Unclassified | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Autonomous Region in Muslim Mindan | 0.04 | 1.34 | 0.71 | 0.91 | 1.10 | 0.34 | 0.23 | 0.43 |
| Cordillera Administrative Region | 0.00 | 0.75 | 0.34 | 0.49 | 0.54 | 0.33 | 0.13 | 0.20 |
| Ilocos Region | 0.05 | 1.95 | 1.02 | 1.39 | 1.55 | 0.81 | 0.38 | 0.64 |
| Cagayan Valley | 0.07 | 1.41 | 0.72 | 0.97 | 1.02 | 0.53 | 0.27 | 0.53 |
| Central Luzon | 0.04 | 1.32 | 0.93 | 0.96 | 1.38 | 1.35 | 0.35 | 0.70 |
| Southern Tagalog | 0.22 | 2.82 | 1.61 | 2.08 | 2.47 | 2.31 | 0.75 | 1.13 |
| Western Mindanao | 0.05 | 1.31 | 0.66 | 0.88 | 1.07 | 0.44 | 0.22 | 0.46 |
| National Capital Region | 0.07 | 0.53 | 0.41 | 0.57 | 0.26 | 1.64 | 0.36 | 0.21 |
| Bicol Region | 0.08 | 1.94 | 1.10 | 1.31 | 1.66 | 1.04 | 0.47 | 0.73 |
| Western Visayas | 0.09 | 2.39 | 1.23 | 1.66 | 1.87 | 1.00 | 0.49 | 0.81 |
| Central Visayas | 0.02 | 1.70 | 0.86 | 1.30 | 1.32 | 0.94 | 0.37 | 0.61 |
| Eastern Visayas | 0.04 | 2.83 | 1.34 | 1.87 | 2.01 | 1.05 | 0.58 | 0.83 |
| Northern Mindanao | 0.02 | 0.81 | 0.43 | 0.57 | 0.67 | 0.63 | 0.20 | 0.30 |
| Southern Mindanao | 0.06 | 0.76 | 0.43 | 0.55 | 0.65 | 0.57 | 0.22 | 0.33 |
| Central Mindanao | 0.08 | 0.92 | 0.35 | 0.51 | 0.74 | 0.35 | 0.20 | 0.27 |
| Caraga | 0.01 | 0.75 | 0.35 | 0.48 | 0.65 | 0.38 | 0.16 | 0.32 |

Everything in the table sums up to 100%. An example of how the table can be interpreted is given below:

*In the Autonomous Region in Muslim Mindan, 1.34% of all national Barangays can be found in cluster 1.*

The table shows how robust the clusters are across all the regions. This will make it a lot easier to extrapolate data for further analysis and tailor policies or decisions to national levels.

After carefully deciding on the final methodology and evaluating the clusters at the top level of the classification, the hierarchy could be created. The unclassified cases were excluded from further analysis and the clustering algorithm was deployed on each of the 7 clusters at the top level. Issues described above were taken into consideration all the way and a second hierarchy of 24 clusters was created. The 24 clusters were also analysed further to create a third hierarchy of 66 clusters. The top hierarchy of 7 have been called **Super-groups**; the middle hierarchy of 24 have been called **Groups**; while the finest level of aggregation comprising 66 clusters have been called **Sub-groups**.

## 6.9 Cluster Profiles

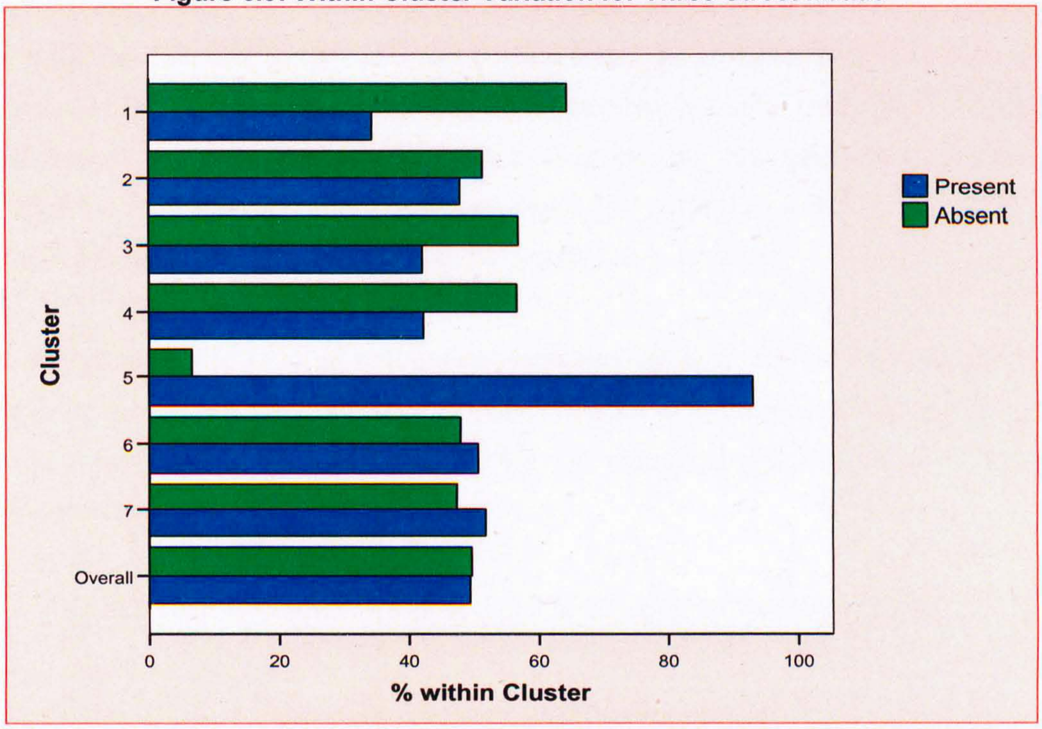
To gain an understanding of the geodemographic characteristics of each cluster, certain techniques were used to evaluate the outputs from the clustering analysis. It is very important to name the clusters such that they reflect the general characteristics of the people and the areas in which they live. To do this it is important to investigate the within and between cluster variations of variables across the clusters. It is also important to get an understanding of which variables drive the classification and in particular which are of greater importance for each cluster. By doing this, textual profiles can be more meaningful and the uniqueness of different clusters can be elucidated.

### 6.9.1 Between and Within Cluster Variation of Variables

An understanding of the distribution of variables within each cluster can be very helpful for profiling the clusters. Since this classification combines both categorical and continuous variables, the composition of clusters has been evaluated using different methods for both types of variables.

Categorical variables have been assessed by looking at the cross-tabulations of the percentage of the variables within each cluster. Figure 6.9 gives an indication of the percentage distribution of the presence or absence of '*at least three street roads*' in each cluster.

**Figure 6.9: Within Cluster Variation for Three Street Roads**




If we observe the pattern of distribution of the variable across the 7 clusters in relation to the overall pattern, we notice that there are variations. The overall distribution of the variable shows a fairly equal percentage distribution for the presence or absence of at least three street roads. Cluster 1 shows a higher incidence in the absence of the variable. In clusters 2, 3 and 4 there is also a higher incidence for absence for the variable but much less than that of cluster 1. Cluster 5 displays a disproportionately high presence of at least three street roads in comparison to all the other clusters. Both clusters 6 and 7 also have greater likelihoods of having at least three street roads though not as high as cluster 5.

The important lesson to draw from this analysis is the fact that the patterns displayed by the different clusters (in terms of absence or presence of the variable) differ from the overall pattern. This means the variable is an important one in the formation of clusters. In the case of continuous variables principal components analysis had already been used to examine which variables are likely to power the clustering process. Results showed that the household infrastructure variables would have great influence of the classification.

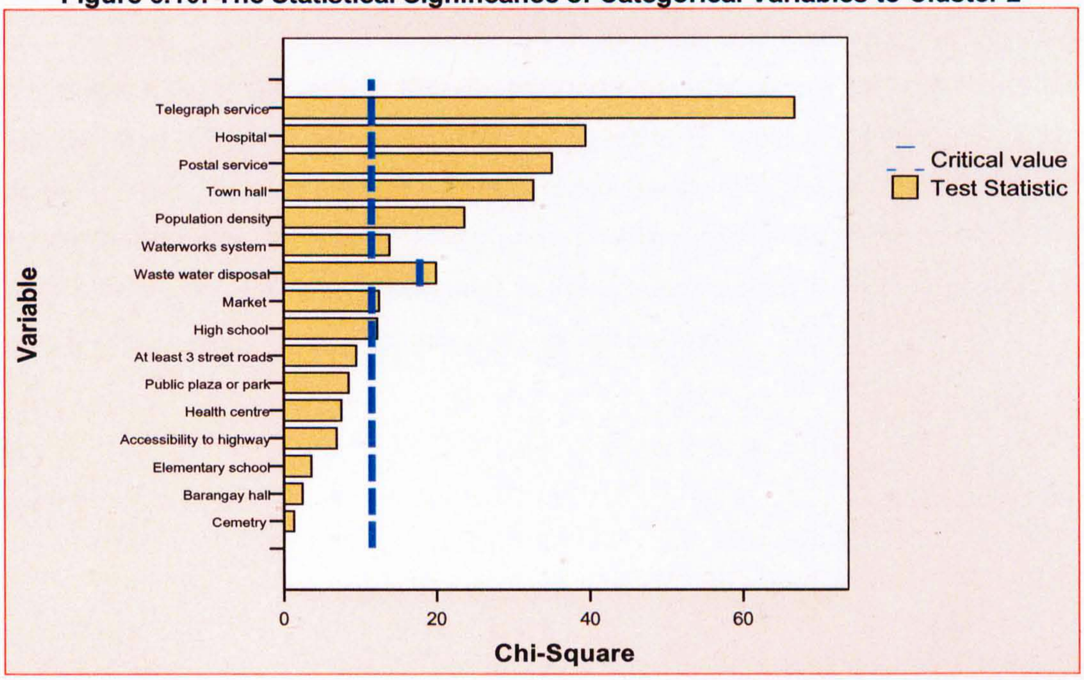
In addition to exploring the principal components loading matrix of these variables, a plot of the means for each cluster was analysed in relation to the 95% simultaneous confidence intervals for the cluster means.

## 6.9.2 Statistical Relevance of Variables

The analysis detailed in section 6.9.1 was used to uncover which variables are important in the formation of clusters. Another potentially useful piece of information is the statistical importance of each variable to each cluster. Just because a variable is a driver of the clustering algorithm does not make it important to every cluster. It may be statistically significant to one cluster and insignificant to another.

The statistical significance of categorical variables by clusters has been analysed by comparing the observed distribution of each variable to the expected distribution using the chi-square statistic. Figure 6.10 shows the statistical significance of categorical variables in the formation of cluster 2.

**Figure 6.10: The Statistical Significance of Categorical Variables to Cluster 2**




Amongst categorical variables, the presence or absence of a telegraph service is about the most important indicator for cluster 2 while the presence or absence of a cemetery is the least important indicator. The critical value line drawn in blue indicates the magnitude of importance above or below the cluster average. In the case of continuous variables, plots of a test statistic (t-statistic) have been used to compare variable means to the overall cluster means to ascertain their importance to their respective clusters.

While these analyses were painstaking, their importance in understanding the characteristics of the different clusters and the features that make them unique can not be overemphasised. The knowledge unearthed from the process also contributed to the development of pen portraits for each super-group and group.

### 6.9.3 Naming the Clusters

As noted in Chapter 5, the process of naming clusters can be a very tedious exercise. It is both an art and a science requiring the fusion of knowledge from different disciplines a high level of creativity. It also demands a good grasp of the meanings of different words. The naming process takes the general socio-demographic characteristics and geographical location of each cluster into consideration.

It can be very contentious to assign a label to a group of people. Different views will occur on what groups should be called. In multi-ethnic and multi-religious developing countries like the Philippines or Nigeria, labeling groups of people can often give rise to debates. It is important to try as much as possible to avoid stigmatising a group of people. In the United Kingdom, *political correctness* is typically the watchword in government quarters when embarking upon an exercise like this (Harris et al., 2005). Table 6.14 shows the names assigned to the 7 super-groups and 24 groups and the number of sub-groups into which each group has been split.

**Table 6.14: Labels for Clusters**

| Super-groups | Label | Groups | Label | Sub-groups | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | Agrarian Pockets | 1.1 | Evolving Agrarian Pockets | 1.1.1 | 1.1.2 | 1.1.3 | | |
| | | 1.2 | Underprivileged Agrarian Pockets | 1.2.1 | 1.2.2 | | | |
| | | 1.3 | Disadvantaged Agrarian Pockets | 1.3.1 | 1.3.2 | 1.3.3 | 1.3.4 | |
| | | 1.4 | Industrious Agrarian Pockets | 1.4.1 | 1.4.2 | 1.4.3 | | |
| 2 | Countrified Juniors | 2.1 | Struggling Countrified Juniors | 2.1.1 | 2.1.2 | | | |
| | | 2.2 | Typical Countrified Juniors | 2.2.1 | 2.2.2 | | | |
| | | 2.3 | Mixed Countrified Juniors | 2.3.1 | 2.3.2 | | | |
| 3 | Retiring Communities | 3.1 | Striving Retiring Communities | 3.1.1 | 3.1.2 | 3.1.3 | | |
| | | 3.2 | Enlightened Retiring Communities | 3.2.1 | 3.2.2 | 3.2.3 | | |
| | | 3.3 | Middle-class Retiring Communities | 3.3.1 | 3.3.2 | | | |
| | | 3.4 | Flourishing Retiring Communities | 3.4.1 | 3.4.2 | | | |
| | | 3.5 | Semirural Retiring Communities | 3.5.1 | 3.5.2 | | | |
| 4 | Enterprise Flux | 4.1 | Middle-aged Enterprise Flux | 4.1.1 | 4.1.2 | 4.1.3 | 4.1.4 | |
| | | 4.2 | Working-class Enterprise Flux | 4.2.1 | 4.2.2 | 4.2.3 | | |
| | | 4.3 | Diversified Enterprise Flux | 4.3.1 | 4.3.2 | | | |
| | | 4.4 | Young Enterprise Flux | 4.4.1 | 4.4.2 | 4.4.3 | 4.4.4 | |
| 5 | City-like Dwellers | 5.1 | Well-off City-like Dwellers | 5.1.1 | 5.1.2 | | | |
| | | 5.2 | Older City-like Dwellers | 5.2.1 | 5.2.2 | | | |
| | | 5.3 | Dependent City-like Dwellers | 5.3.1 | 5.3.2 | 5.3.3 | | |
| | | 5.4 | Serviced City-like Dwellers | 5.4.1 | 5.4.2 | 5.4.3 | | |
| 6 | Family Focused | 6.1 | Populated Family Focused | 6.1.1 | 6.1.2 | 6.1.3 | | |
| | | 6.2 | Sub-rural Family Focused | 6.2.1 | 6.2.2 | | | |
| 7 | Career-centric | 7.1 | Ambitious Career-centric | 7.1.1 | 7.1.2 | 7.1.3 | 7.1.4 | 7.1.5 |
| | | 7.2 | Established Career-centric | 7.2.1 | 7.2.2 | 7.2.3 | | |

Apart from the broad geodemographic characteristics of clusters and their geographic locations, the labels for each super-group were restricted to a maximum of two words to allow for easy memorisation. Groups were restricted to a maximum of three words.

While the labeling of clusters are a very important part of the exercise because this is often what captures the interest of users. It is important to state that a two-word label is not enough to summarise the multivariate characteristics of clusters. Labels should therefore be used with utmost care and greater importance should be paid on the evidence upon which the labels have been assigned. This evidence is defined by the variations in the importance and behaviours of the input variables as discussed in the next section.

## 6.9.4 Profiling and Pen Portraits

Cluster profiles simply refer to detailed representations of the characteristics of clusters. Many times charts are used to summarise these characteristics. For instance, it is typical to relate the importance of a variable to cluster to the overall importance of that variable.

Radar charts come in handy when trying to summaries such multivariate information. They are concise and relatively easy to interpret. Each of the 7 super-groups, 24 groups and 66 sub-groups has been profiled using the input variables as shown in Figures 6.11 to 6.17.

## Figure 6.11: Profile of Agrarian Pockets

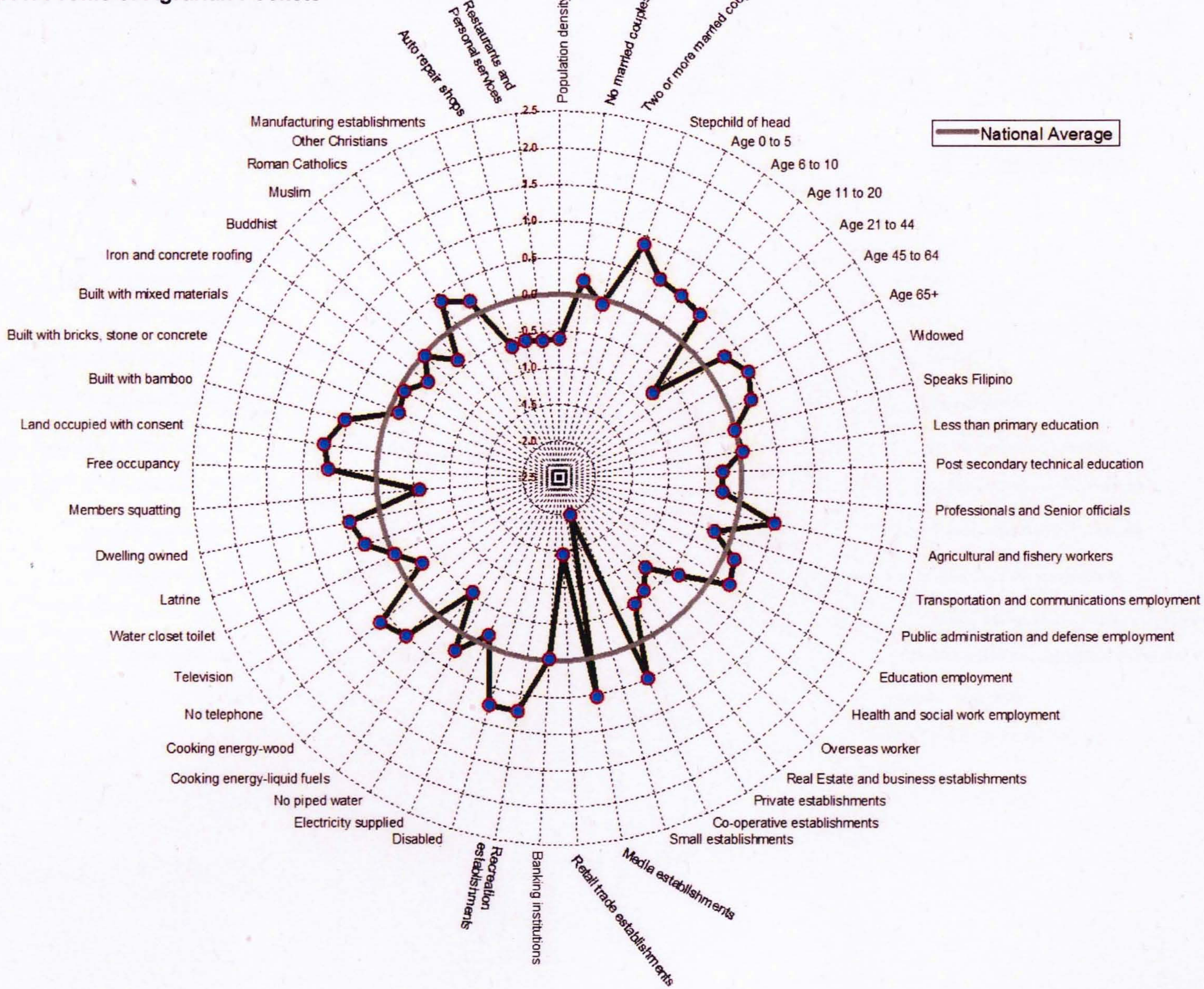


**Figure 6.12: Profile of Countrified Juniors**

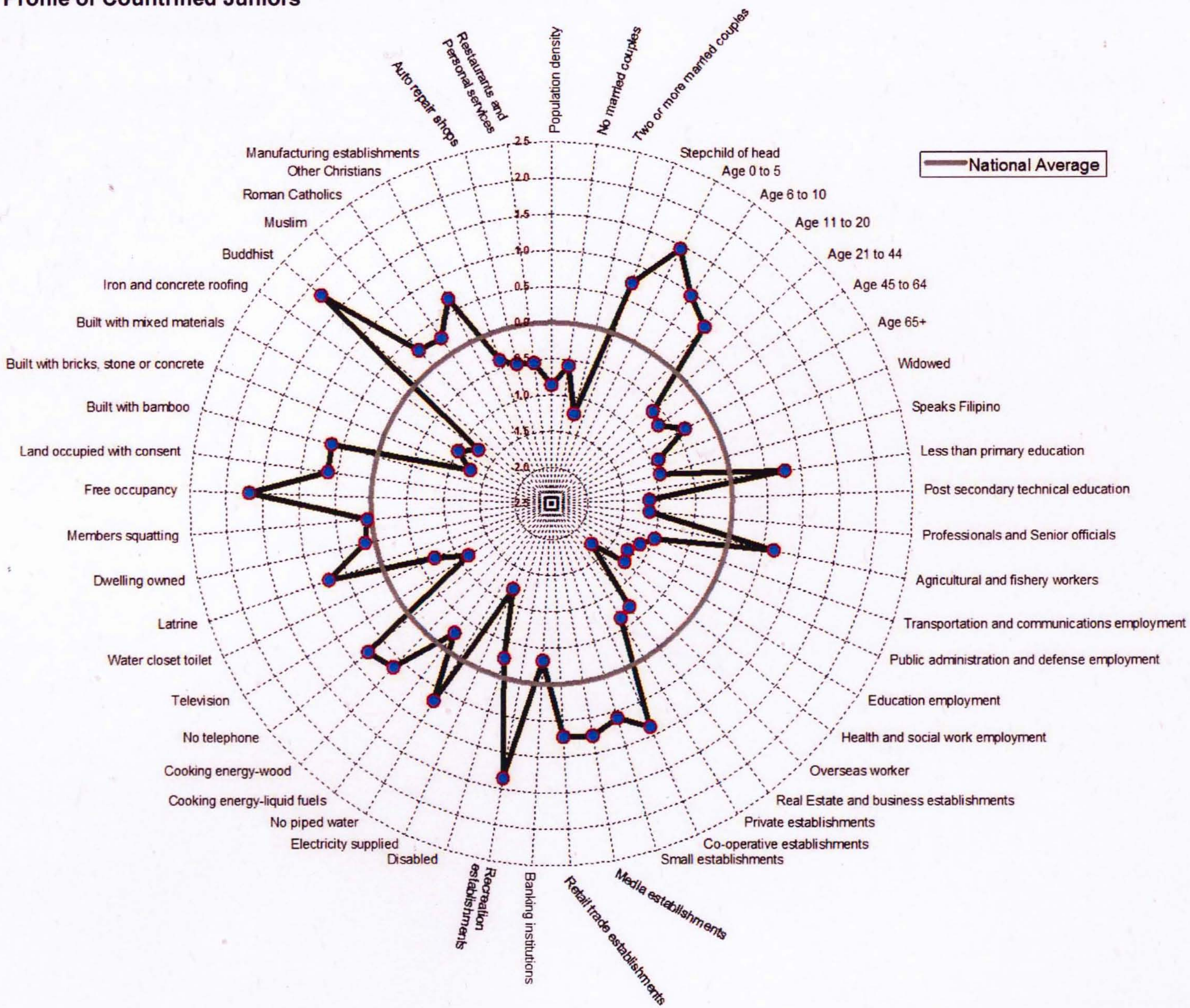
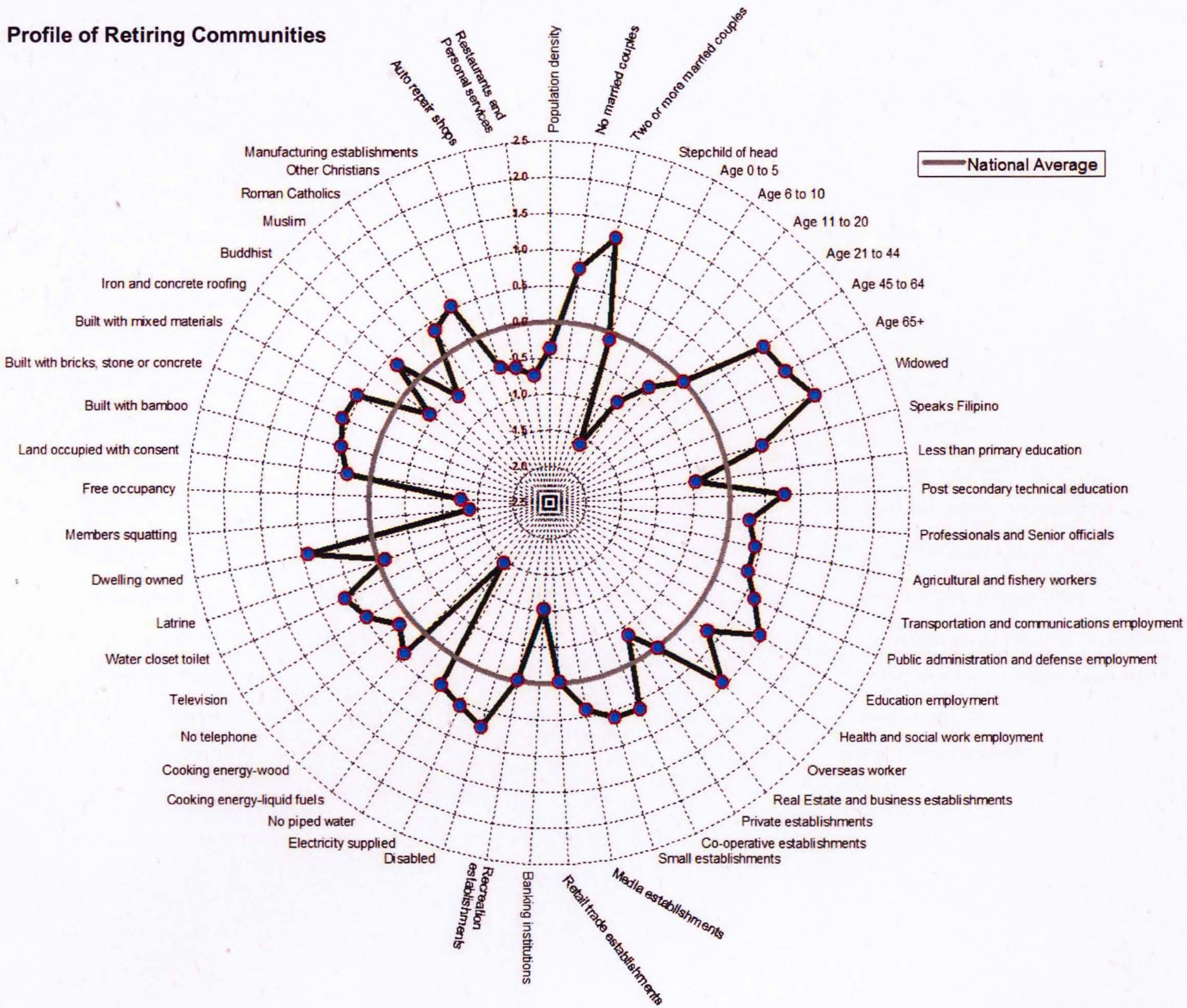


**Figure 6.13: Profile of Retiring Communities**



## Figure 6.14: Profile of Enterprise Flux

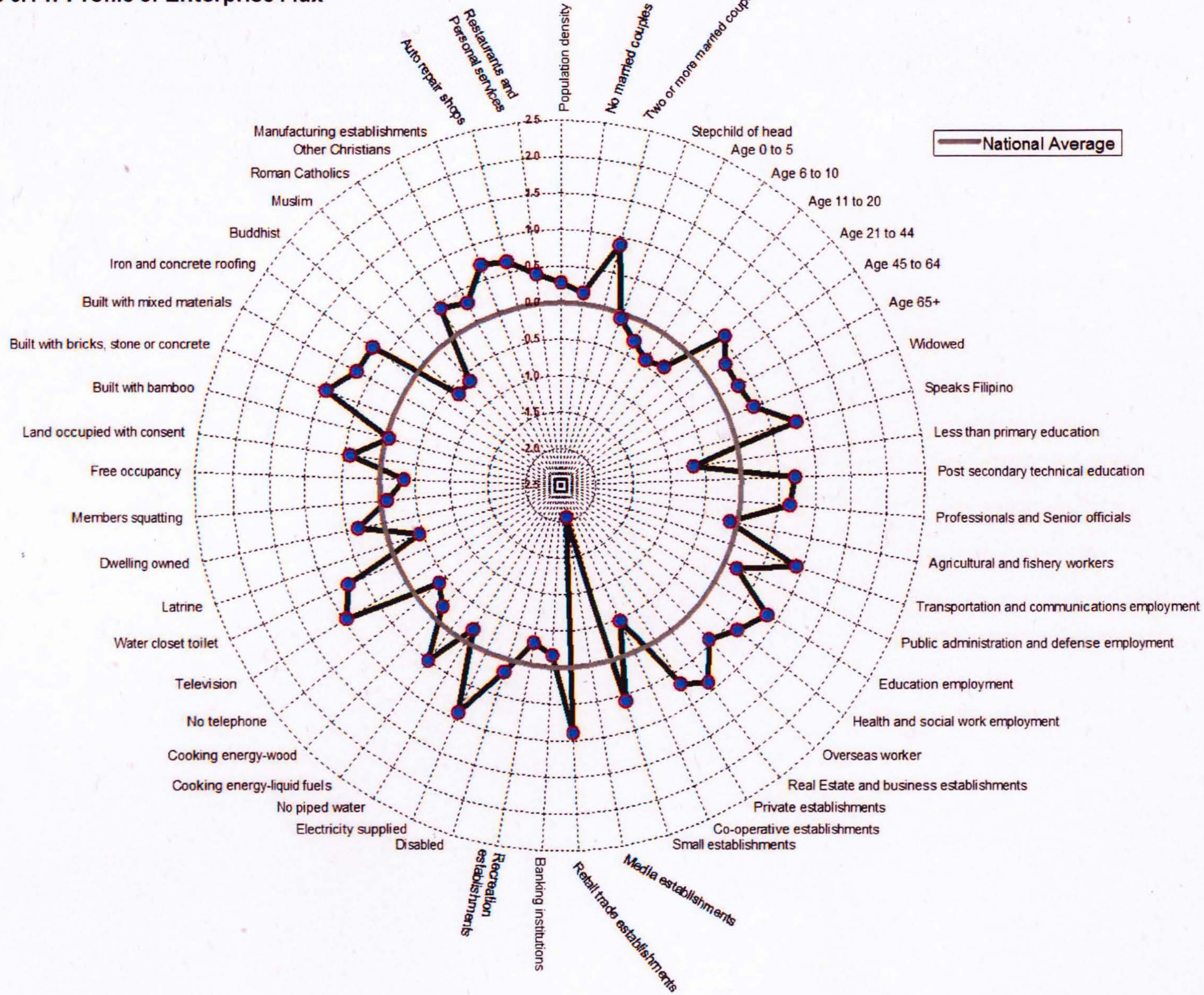


**Figure 6.15: Profile of City-like Dwellers**

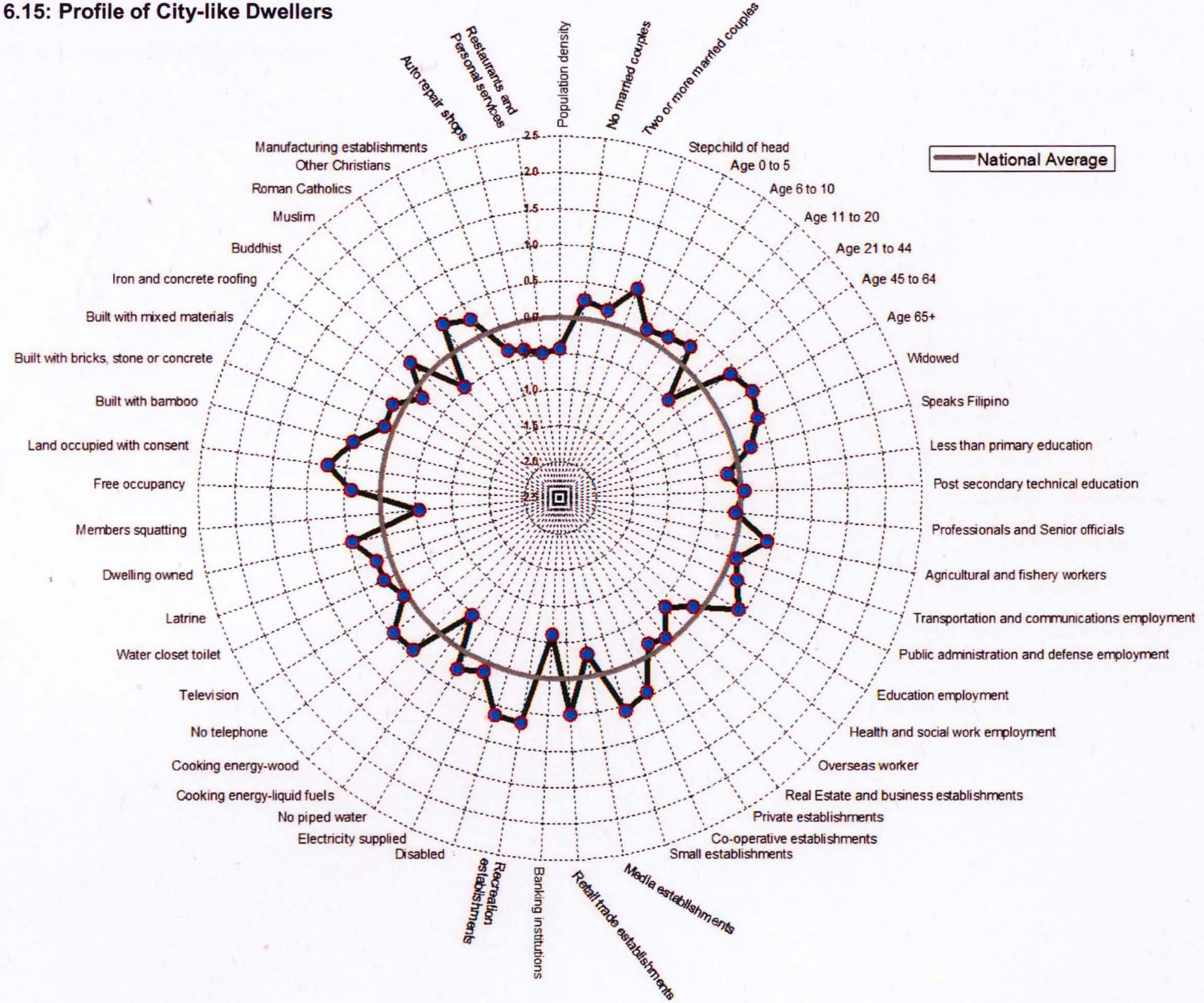
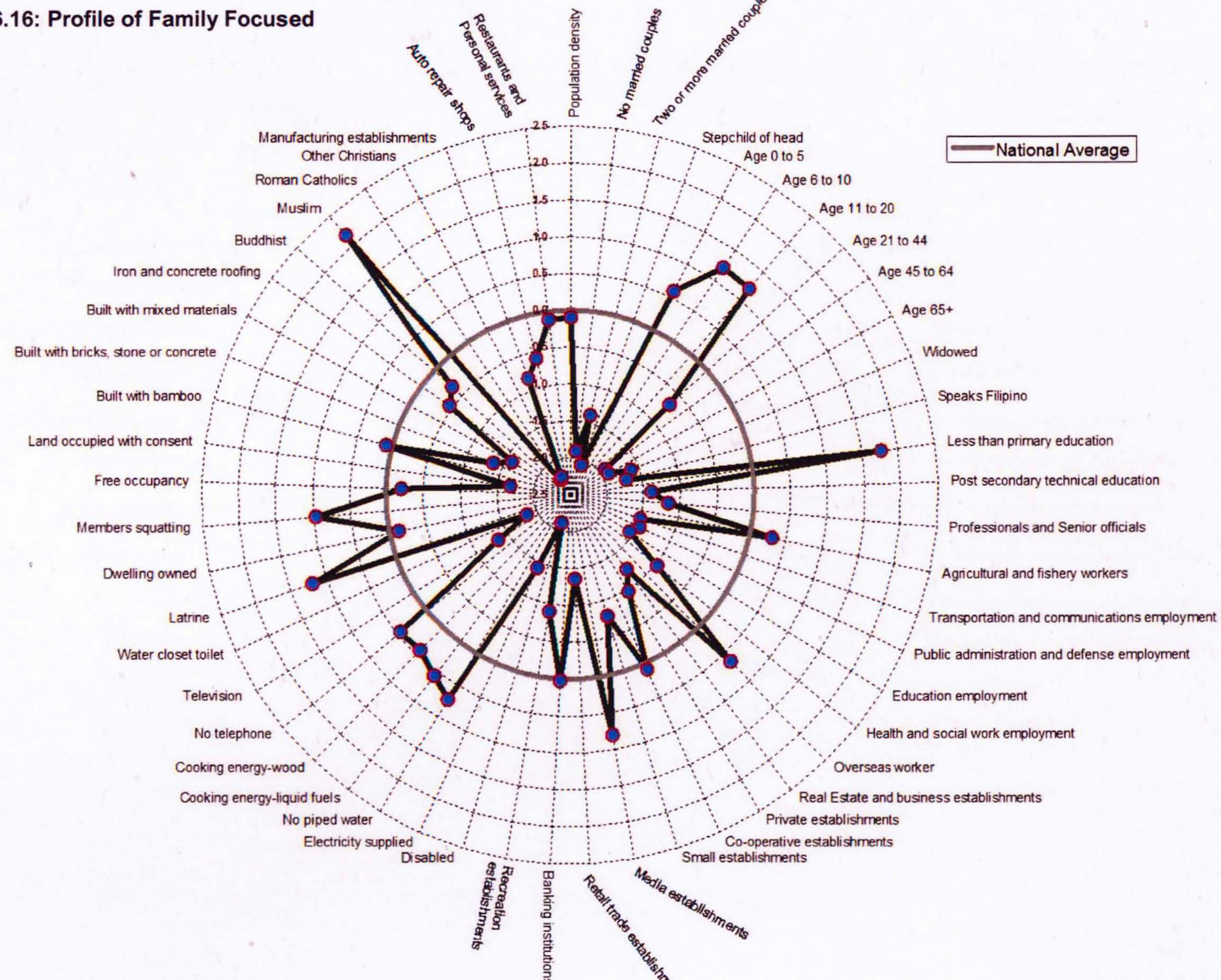
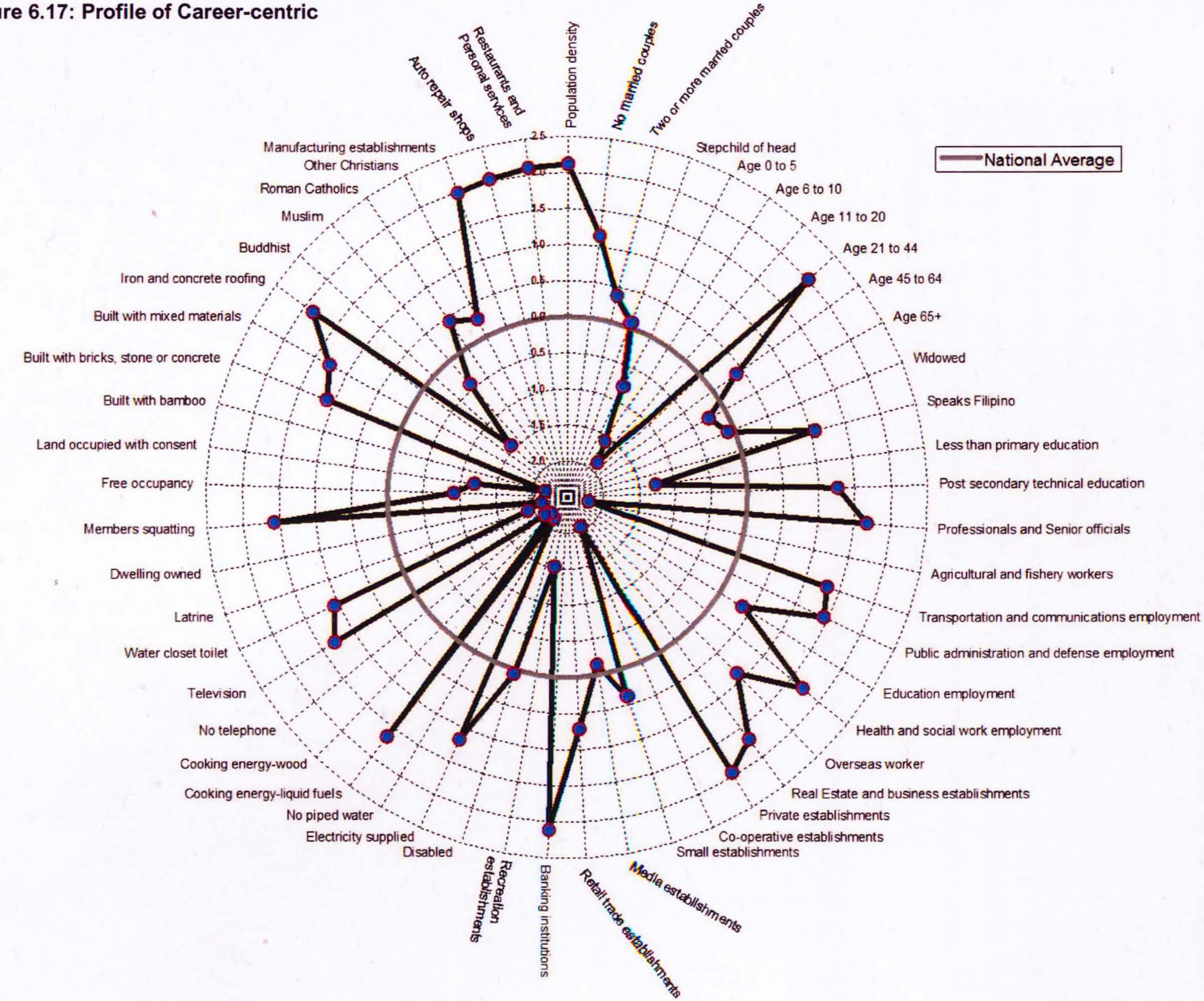


**Figure 6.16: Profile of Family Focused**



## Figure 6.17: Profile of Career-centric

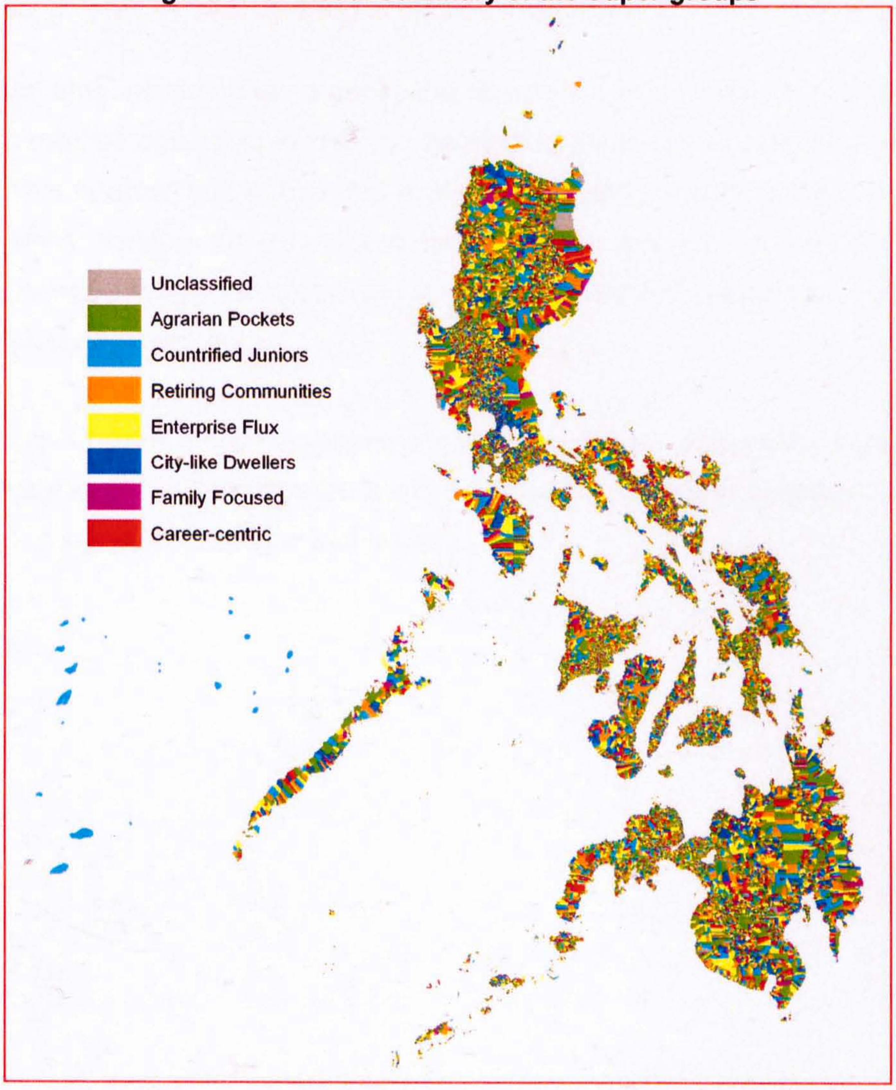


In addition to the radar charts, pen portraits which are comprehensive textual descriptions of the different cluster groups have also been created for each super-group and group. These descriptions have been created in such a way as to minimise the use of statistical jargon. In practice, it is this information that non technical audiences or policy makers are more interested in. The pen portraits for the Philippines classification are contained in the enclosed compact disc.

## 6.9.5 Mapping and Visualisation

The fact that the results from this analysis are linked with geography makes it possible to map it using Geographic Information Systems (GIS). Since the classification was created at the Barangay administrative level, this is also the finest scale at which the clusters and results of any analysis can be mapped.

**Figure 6.18: Visual Summary of the Super-groups**



The irregular shape of the Philippines makes it difficult to comprehend spatial patterns when information is mapped nationally. Dynamic visualisation like mapping upon a mashup surface like Google or Microsoft maps (Gibin et al., 2009) is one way of getting around this problem. For the purpose of this work, results have been mapped for each of the 16 regions and are shown in Appendix B.

## 6.10 Conclusions

In the Philippines, Barangays are the smallest administrative units of population aggregation. They therefore serve as an important basis for understanding local level inequalities when trying to understand key relating to national policy debates. Unfortunately it is very uncommon to find national analysis in the Phillipines conducted at this scale. The analysis discussed in this section is therefore very novel in many respects. A particularly innovative aspect of the work is the way in which categorical datasets were analysed and combined with the continuous datasets.

For the first time, census related geospatial datasets with national coverage has been used to create an open-source national geodemographic classification system for the country. This chapter has succeeded in discussing and explaining the key concepts used to justify variable selections and further cluster analysis. A total of sixty nine variables have been used to create the three tier hierarchical classification based on a two-step clustering algorithm.

Further outputs from the analysis are contained in the accompanying compact disc. The usefulness of the classification system will be put to test in chapters 7 and 8 by investigating some MDG related indicators.

# 7

# Investigating Primary Education in Nigeria and Philippines

## 7.1 Introduction

In chapter 3, the potential of geodemographics for policy related issues especially as they relate to the MDGs were reviewed. Chapters 5 and 6 provided discussions on how the Nigerian and Philippines classifications were developed. This chapter will illustrate the usefulness of geodemographics for policy by evaluating some of the Nigerian and Philippines MDG indicators and proxy indicators within the educational sector focusing specifically on primary level education.

Since Nigeria commenced evaluating and reporting progress towards meeting the targets of the MDGs, there has never been any national analysis or reporting at the Local Government Area (LGA) scale. The 2006 report, which is the third and most recent in the series of reports, was published in January, 2007 by the National Planning Commission (NPC). The report was based on data collected in 2004.

The most recent national report compiled for the Philippines monitoring progress towards the MDGs is a midterm report written in the year 2007. Again as is customary with the UN reports for other countries, considerable attention was paid to the regional level of geography. Even at this coarse level of geography, the report is honest to attest to the fact significant disparities in progress exist between regions (NEDA and UNDP, 2007). The magnitude of inequalities likely to exist at the local barangay scale can be expected to be more relevant for the purpose of targeting resources.

The analysis detailed in this chapter and the next provides the first opportunity of conducting national assessments of some of the indicators at more localised geographical scales using a geodemographic approach.

The remainder of this chapter is organised as follows: sections 7.2 and 7.3 provide some background information on primary education in both countries. In section 7.4 a brief description of the sources of data to be analysed is provided. Section 7.5 investigates enrolment and attendance issues in both countries while in section 7.6 the patterns of contentment with primary education in Nigeria are addressed. The same chapter addresses primary school performance in the Philippines. Conclusive remarks are provided in section 7.6.

## 7.2 Primary Education in Nigeria

It is one of the targets of the MDGs that by the year 2015 every boy and girl would be able to complete a full course of primary level education. To achieve this target, the indicators which are monitored include primary enrolment ratio, completion rates at primary level and literacy levels for males and females aged 15 to 24 years. These have already been discussed in detail in section 3.2.2 in the third chapter of this thesis.

It can be argued that for Nigeria, there are mixed feelings in the progress towards meeting the goal of achieving universal primary education (NPC, 2007). The MDG report for 2006 used data from 2004 and was analysed and reported at broad levels of geography (regional and state levels). According to the report, 84% of children attend school and the overall net enrolment of children in primary education increased from 81.1% in 2005 to 84.26% in 2005 (NPC, 2007). Additionally it is on record that the level of literacy also improved within the same time interval from 76.2% to 80.20%. Two key factors are thought to have contributed to the enhanced level of literacy. First a democratic political climate operating since 1999 and the Universal Basic Education (UBE) policy launched in the same year of democratic re-emergence (NPC, 2007).

A major challenge for these submissions is that they fail to uncover the patterns of local level variations in progress towards meeting the required targets. A contributory factor to this is that the interested international donor agencies themselves appear contented with reporting at the country or at best regional levels. This hides the inequalities within and between local areas (Dorling and Ballas, 2008) and can result in misplaced targeting of resources.

## 7.3 Primary Education in the Philippines

In the Philippines, education is highly recognised as a prime tool for economic development (HDN and UNDP, 2000). Between 1989 and 2003, literacy rates for people aged ten and above in the country have fluctuated between a range of 89.8% and 93.4% which would seem impressive for a developing country (NSO and Department of Education, 2009). These however are national rates which overshadow local level variations.

The significance of education is made vivid in the country's constitution in article 14, section 2 and sub-section 2 which recognises the right of children to be educated to at least high school (Constitutional Commission of Philippines, 1987). In spite of this stipulation, there are still many children who grow into adults without any form of education.

The second MDG focuses on ensuring that children everywhere receive quality primary level education to prepare them for challenges they may encounter in the future. Before children can be properly educated, it is important first for them to attend school. This means that barriers against school attendance should be understood and properly dealt with.

In addition to understanding the barriers working against the attendance of children at school, it may also be helpful to understand why certain areas do better than others in terms of school attendance. Knowing this can help education policy makers in the Philippines re-define their strategies towards increasing attendance rates.

While attendance at elementary school is important to ensure that the targets set within the MDG framework are met, one short-coming of the UN approach is that it fails to address problems relating to the quality of teaching and the ability of children to actually derive optimum benefit from their studies.

It is not every child that attends school that makes progress. Some have to repeat classes for various reasons while others eventually drop out of school. Although these two issues are not addressed within the MDG framework, they are important for measuring the overall performance and efficiency of educational systems (Catterall, 1985; Tyler and Lofstrom, 2009). The spatial pattern of these indicators and their geodemographic correlates are addressed in sections 7.5 and 7.6.

## 7.4 Data Sources

Many of the challenges of addressing local level inequalities within developing countries were discussed in chapter 2. Many developing countries including Nigeria and the Philippines lack adequate data monitoring and registration of vital statistics systems (NSO and ORC Macro, 2004). The availability and access to relevant datasets is therefore a major constraint to conducting spatial analyses in many of these countries. In spite of preliminary challenges and barriers to data sourcing experienced in the course conducting this research, variables (not included in the development of the classification systems) were eventually made available for both countries.

For Nigeria, the datasets analysed and profiled in this chapter and the next were sourced from the National Bureau of Statistics (NBS) Abuja – the government's national statistical agency. The NBS supplied statistics from its largest nation-wide wide survey in 2006 (NBS, 2006a) called the Core Welfare Indicators Questionnaire Survey (CWIQ).

The CWIQ is a World Bank assisted survey conducted in many developing countries aimed at collecting data useful in quantitatively profiling the well-being of residential population and deriving information for monitoring some of the MDG targets (NBS, 2006a). With a sample size of more than 77,000 households, every single LGA in Nigeria were contained in the survey.

For the Philippines, data was derived from the 2003 Demographic and Health Survey (DHS). The DHS is a nationally represented survey of almost 13, 000 households comprising more than 60, 000 individual responses which include men, women and children. All respondents to the survey were assigned a geographical geo-code which makes it possible to link the data to a GIS and subsequently to the geodemographic classification system created and discussed in chapter 6 for further spatial analysis.

According to the Philippines 2004 DHS report, the principal objective of the survey is to:

'*to provide up to-date information on population, family planning, and health to assist policymakers and program managers in evaluating and designing strategies for improving health and family planning services in the country*' (NSO and ORC Macro, 2004, p. 2).

The 2003 DHS which is the eight in a series of such surveys is a collaborative effort between the Philippines National Statistics Office (NSO) and the United States government. It was conducted between June and September 2003 with financial support from the United States Agency for International Development (USAID) and technical inputs from ORC Macro – a research and management consulting organisation based in Maryland, USA.

## 7.5 Enrolment and Non-attendance at Primary School

This section will focus on analysing two issues. Due to the data made available for this research, access and enrolment at primary level education will be investigated in Nigeria to determine their geodemographic correlates and spatial penetration patterns.

Section 7.5.3 then addresses the patterns of non-attendance at elementary school among children aged between 6 and 11 years.

### 7.5.1 Access to Primary Level Education in Nigeria

Primary education in Nigeria commences at age six and lasts for six years (NPC and ORC Macro, 2004). It is an important period in the lives of children as they receive their first introduction to formal education.

Within the context of this study, access to primary level education is defined for children aged six to eleven years who can reach their nearest primary school within thirty minutes. It has been shown that travel time can have its impacts on issues like enrolment not just in developing societies (King and Lilard, 1987) but even in developed countries (Singleton et al., 2007). For primary level education, most parents would normally be prepared to commute over shorter distances not just because of the travel time but because the greater the distance travelled the more expensive the cost. This can discourage low income families in particular.

In order to compute the rates of access to primary school by geodemographic types, it was essential to compare the total count of children of primary school age with access to primary school by geodemographic types with an appropriate base population. From the data available, the corresponding base population were children aged 15 years and below in each of the geodemographic types. By relating the rates of children with access to primary school by geodemographic types to the rates of children aged 15 and below by geodemographic types, indices were further computed for each geodemographic type as follows:

$$\text{Index} = \frac{n \div \sum_{1}^{k} n}{N \div \sum_{1}^{k} N} \times 100 \qquad (7.1)$$

where:

n = the count of people with a characteristic in geodemographic cluster k

K = the total number of geodemographic clusters

N = the count of people in geodemographic cluster k

The indices help reveal how far above or below the national mean (which is designated as a value of 100 here) a geodemographic typology is in terms of the variable (access to primary education in this case) is (Harris et al., 2005). An index value of 200 would therefore signify that the likelihood for a household within that geodemographic cluster to be in the lower class is double the national mean.

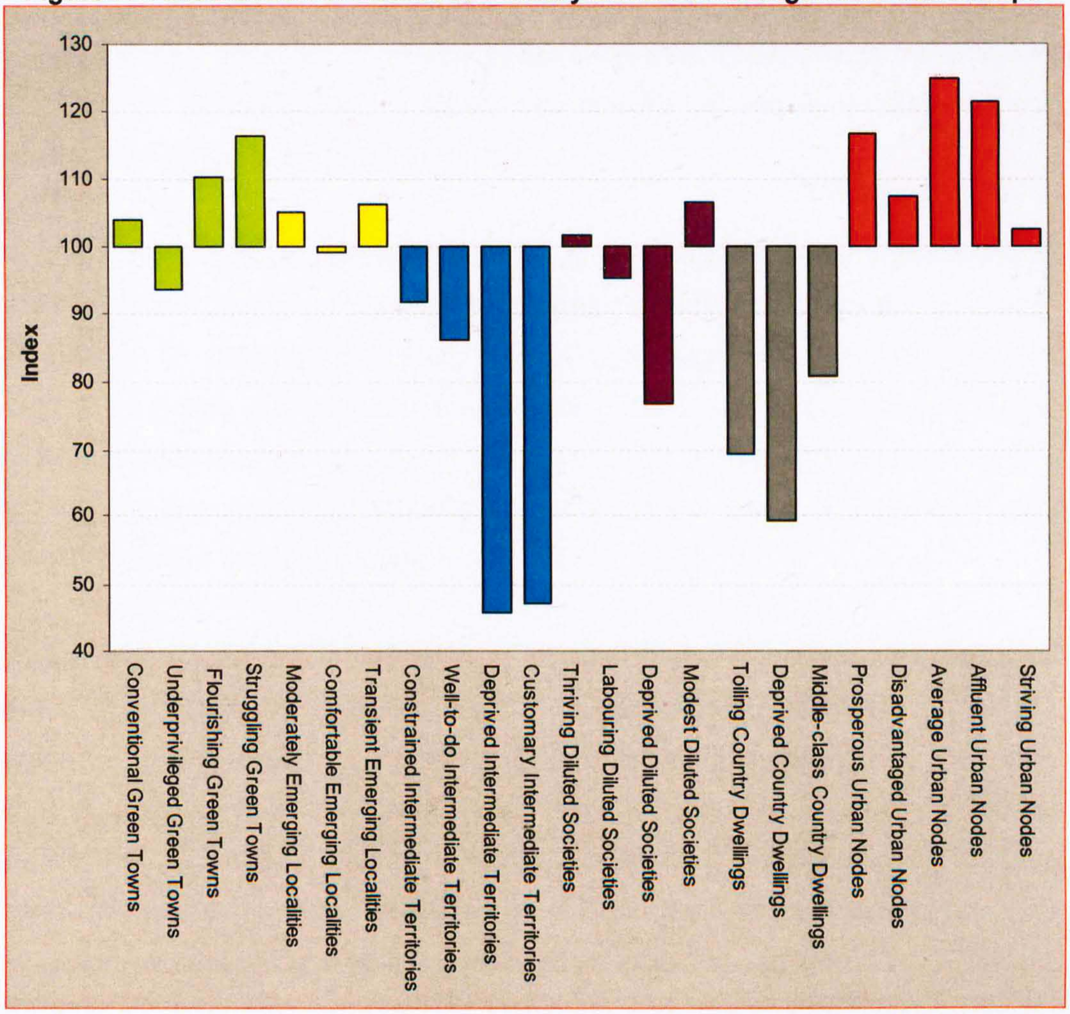
**Table 7.1: Penetration Report for Access to Primary Education in Nigeria**

| Super-groups | Population Share (%) | Rate (per 100 households) | Index |
|---|---|---|---|
| Green Towns | 20 | 81.23 | 108 |
| Emerging Localities | 27 | 78.59 | 104 |
| Intermediate Territories | 7 | 53.23 | 71 |
| Diluted Societies | 15 | 75.59 | 100 |
| Country Dwellings | 9 | 58.59 | 78 |
| Urban Nodes | 22 | 87.33 | 116 |

From the information contained in Table 7.1, it is evident that the largest national share of children who have the opportunity to spend less than thirty minutes to their nearest primary school are located in Emerging Localities. One reason accounting for this is the very large sizes of households within these areas typically averaging 6.1 persons. Additionally Emerging Localities are characterised by a disproportionate concentration of very young population. However, for the purpose of targeting in terms of access, Intermediate Territories and Country Dwellings demonstrate greater need.

Further disaggregation of the information at group level helps reveal which geodemographic typologies actually dominate in terms of access to primary education. It is evident from Figure 7.1 that access is closely associated with deprivation and urbanisation.

**Figure 7.1: Distribution of Access to Primary Education in Nigeria across Groups**




One limitation of the analysis is that it is not immediately clear if travel time is defined in terms of walking distance or vehicular travel. Although the information on means of travel was collected during the survey, the data was not made available for analysis. In this respect while one can not entirely rule out the distance factor, one can not also conclude that the areas that are likely to experience low rates of access do so due to distance. For instance primary schools within Intermediate Territories may just be as far from peoples residences as they are within Urban Nodes. However, the fact that the penetration rates of vehicle ownership and access to public transportation is significantly greater within Urban Nodes would mean that residents of Urban Nodes are likely to spend less time to their destinations.

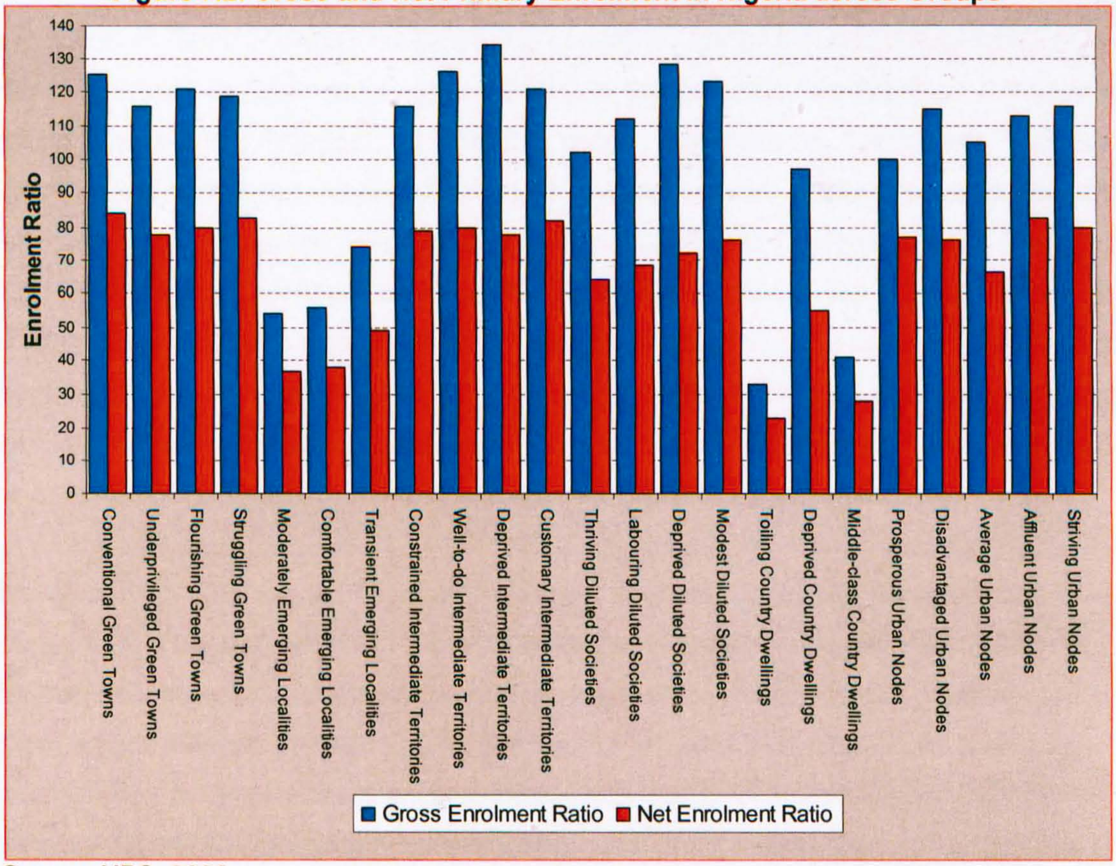
### 7.5.2 The Pattern of Primary School Enrolments in Nigeria

The first indicator within the framework of the second MDG pertains to the enrolment of primary school students. There is a particular bias for net primary enrolments because it relates to students of primary school age.

In this research, data was provided for both gross and net primary enrolments. Gross enrolment is the number of children of all ages currently in primary school divided by the number of children of primary school age. Alongside the UN required net enrolment, it is also very important to understand the pattern of gross enrolment for the purpose of budgeting and deploying resources. This is because when resources are deployed for use in primary schools, students are not asked if they are of primary school age before they can benefit.

Another reason why enrolment patterns are vital is that it can help education policy makers understand how to better target their campaign messages or strategies. For instance if certain areas have disproportionately high gross enrolments, it could mean that children within such neighbourhoods are under aged when they resume school. To mitigate the challenge, a campaign strategy highlighting the negatives of introducing children too quickly to formal education could be focused on such areas. On the other hand, high rates of gross enrolment could also mean that children do not start school early and as such, they are past the legitimate age by the time they complete their studies. To determine if and under aged or over aged specific campaign strategy is required, the individual ages of enrolled students would be required for the analysis. Individual age groups was however not included in the data supplied.

**Figure 7.2: Gross and Net Primary Enrolment in Nigeria across Groups**




Source: NBS, 2006

The chart shown in Figure 7.2, displays the average values for gross and net primary enrolment ratios across the groups of the Nigerian classification system. The trend for both types of enrolment is very similar with minor fluctuations which are discussed further.

For both gross and net enrolments, Emerging Localities and Country Dwellings have the least ratios. In general, primary school enrolment is least for Toiling Country Dwellings where literacy rates are generally low, household heads are under-represented within educational circles and separated couples have an above average presence.

In addition, there is important evidence within the pattern of gross enrolment. With the exception of Green Towns and Emerging Localities, the group with the greatest gross enrolment ratio for each super-group also has the greatest disadvantage in terms of social well-being. Amongst all groups within Intermediate Territories, Deprived Intermediate Territories have the greatest level of relative deprivation and also the largest average gross enrolment ratio. The same applies to Deprived Diluted Societies; Deprived Country Dwellings; and both Disadvantaged and striving Urban Nodes. This finding suggests a close association between deprivation and the enrolment of children outside the recommended school age bracket in Nigeria. Though not conclusive until verified with age statistics, it is probably likely that the contribution to the higher gross enrolments within these geodemographic types is by over-aged children.

Numerous studies on educational enrolment and attainment in Nigeria have been conducted at regional levels of geography and have often concluded that the Northern half of the country lags behind the rest of the country (NPC and ORC Macro, 2004). Many of the reasons that account for the lag are linked with socio-cultural and religious barriers. The findings in this research concur with that position (Emerging Societies and Country Dwellings are mostly in the north) and go the extra mile of being able to disaggregate enrolment down to the LGA scale. However, of more importance is the ability to uncover some more non conventional factors associated with some the neighbourhood types with the lowest net enrolment ratios.

For instance, Toiling Country Dwellings and Middle-class Country Dwellings have the lowest net enrolments with a proportion of more than 70% of children within these areas unlikely to be enrolled in school in both neighbourhood types. Outside the socio-religious factors that may contribute to their absence from school, most children within these areas are away from school due to the distance they have to travel and the fact that many are engaged in some form of child labour. Indeed Middle-class Country Dwellings have the largest national rates of children who are away from school due to child labour. Some times, too much focus on socio-religious barriers tend to overshadow some of the practical policy defining issues uncovered here.

### 7.5.3 Patterns of Non-attendance in the Philippines

Achieving universal education at the primary level of schooling starts with the attendance of children at school. In the Philippines and other developing countries, legislation makes it mandatory for parents and guardians to ensure that their children attend school. However, in many of these countries, the attendance of young children at school is still often treated as a choice rather than a right.

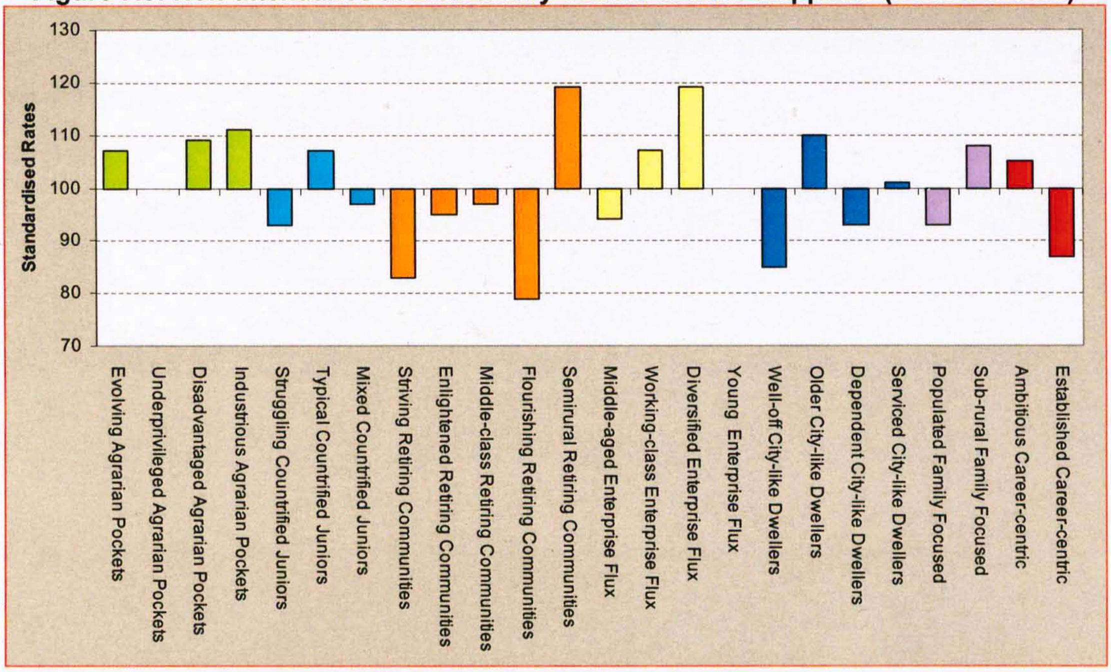
Elementary or primary school in the Philippines commences around age 6 and ends around the age of 11. A person who has elementary level education is presumed to have completed grade 6 at the primary level of schooling which is in line with the framework of the MDGs (NSO and ORC Macro, 2004).

Overall, there has been a steady decline in the rate at which young children have been enrolled for elementary schooling. It was only recently between the 2004/2005 and the 2005/2006 sessions that there was a marginal increase in the rate of enrolment. The increase was accounted for by a 13% rise in enrolment within private institutions. It is however noteworthy to mention that private institutions are typically more expensive than public or government owned institutions (Tsang, 2002). This automatically puts children from less deprived households at a disadvantage.

From the 2003 DHS, data for those aged between 6 and 11 years was extracted and those who had never attended school amongst them were profiled against the geodemographic system. Prior to this research, these datasets have only ever been assessed at the regional level of geography or by urban and rural areas (NSO and ORC Macro, 2004). However, this is not very helpful for targeting policies to local areas. By geodemographically profiling the data, greater insight can be derived and the results can be mapped locally.

First, the overall pattern of non attendance is assessed for both genders to see if there are significant variations. The results of the analysis are shown in Figure 7.3. The rates of overall non attendance have been standardised about a national mean of 100.

**Figure 7.3: Non-attendance at Elementary School in the Philippines (Both Genders)**




Clearly there are significant variations in the propensity of children living in different types of areas to attend elementary school. The highest rates of non attendance can be found within Semirural Retiring Communities which are characterised by large numbers of unmarried persons a very large presence of people aged over 45 years. The incidence of widowed persons is also one of the highest in the country for the neighbourhood type.

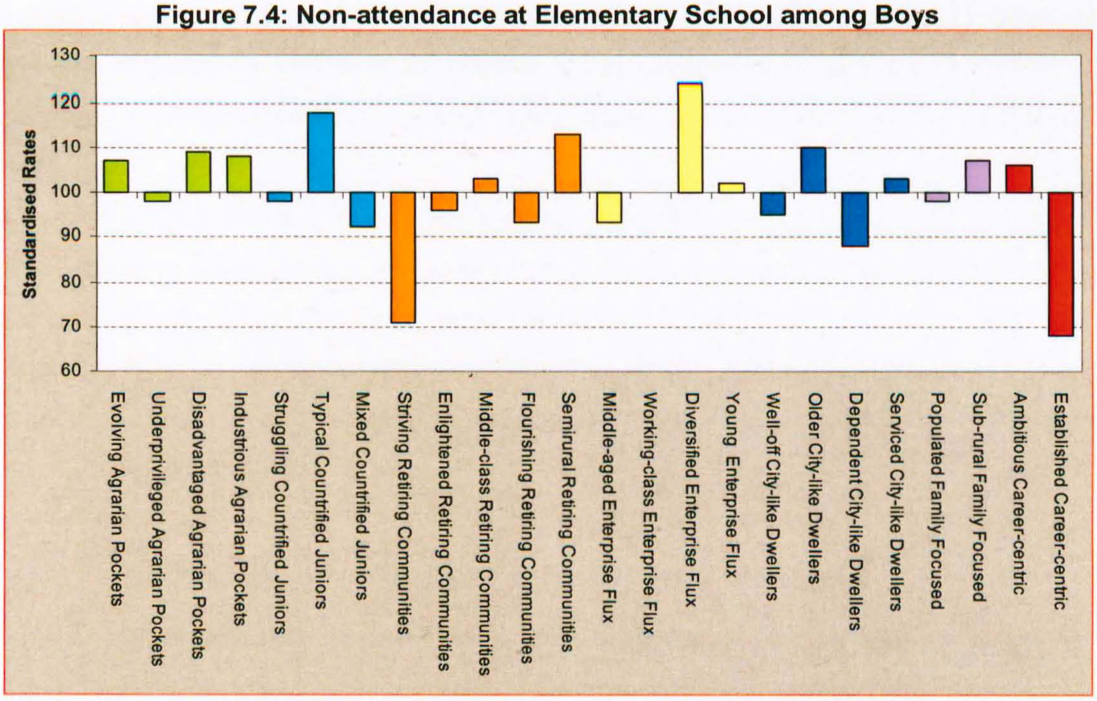
Surprisingly, there is also a large incidence of non-attendance within Diversified Enterprise Flux where although there are representative numbers of parents without a complete primary education. However when compared to the national distribution, more people within these neighbourhoods appear to have complete secondary level education. The neighbourhoods also have a large number of squatters whose families probably contribute to the presence of high non-attendance rates at elementary level schooling.

One of the short comings of analysing the data at higher levels of geography using regions is explained by the extract in the DHS report which reads as follows:

'*No major gender differences are observed for education. However, a significant difference is noted between urban and rural areas; the educational system favours residents of urban areas*' (NSO and ORC Macro, 2004 p. 11).

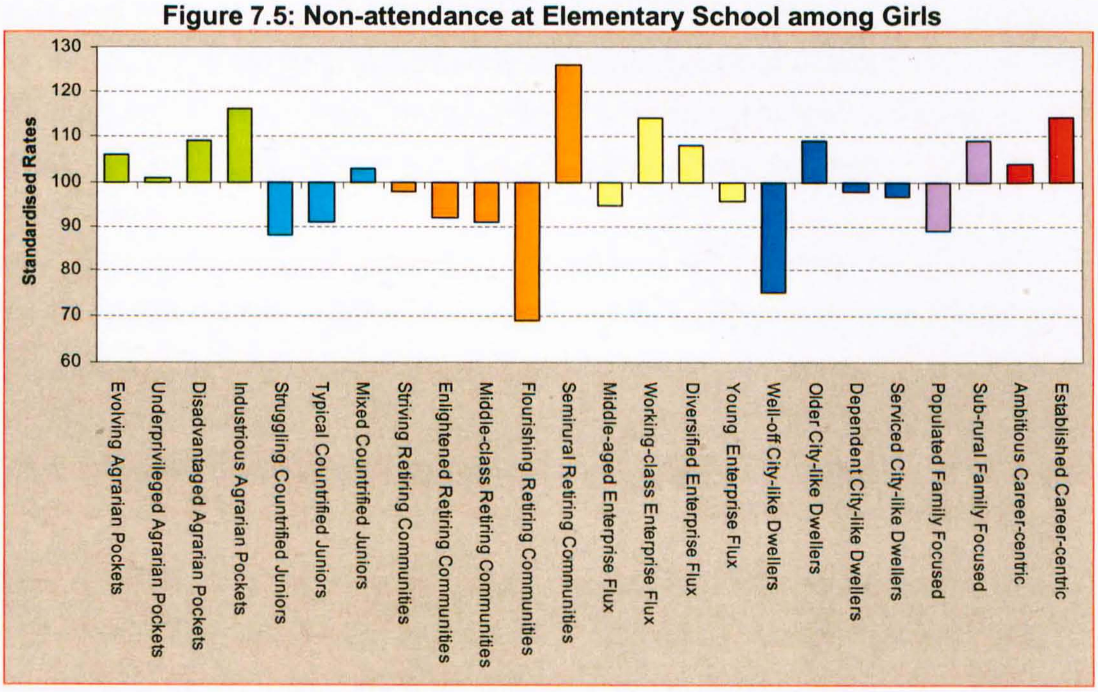
The statement demonstrates the weakness of evaluating data using only contiguous regions which do not necessarily encapsulate the characteristics of residential population.

**Figure 7.4: Non-attendance at Elementary School among Boys**




In Figures 7.4 and 7.5, the strength of the geodemographic system at uncovering inequality in non-attendance at elementary school level by gender is shown.

**Figure 7.5: Non-attendance at Elementary School among Girls**



If the two charts are related to the overall pattern of the distribution of non-attendance replicated in Figure 7.3, one can conveniently see how the geodemographic system stratifies inequality by gender. In Figure 7.3, both Semirural Retiring Communities and Diversified Enterprise Flux have high incidence of overall non-attendance. However when the data is disaggregated by gender, we find that the chance of boys not attending elementary school is far greater within Diversified Enterprise Flux while for girls, there is a disproportionate concentration within Semirural Retiring Communities.

Again this insight points to the need for policy makers to deploy more intelligent tactics when targeting areas for the purpose of improving attendance at elementary school. Non-attendance among girls seems to be concentrated within rural areas with older population and a strong presence of widows while amongst boys, areas where completion rates are not as high though relatively less rural with a mix of middle-aged and older people.

## 7.6 Contentment with and Performance at Primary School

Section 7.5 addressed issues on access and enrolment in Nigeria while non attendance at primary school was investigated for the Philippines. While it is important to encourage parents and guardians to ensure that their children attend school, it is also important to be able to assess how comfortable children are with the quality of teaching services they receive.

Apart from the quality of teaching, some other factors like the availability of services and the supply of books may influence the contentment of children with the formal education they receive. Good performance at elementary school is desired by every parent who enrols their children. However, this is not always the case.

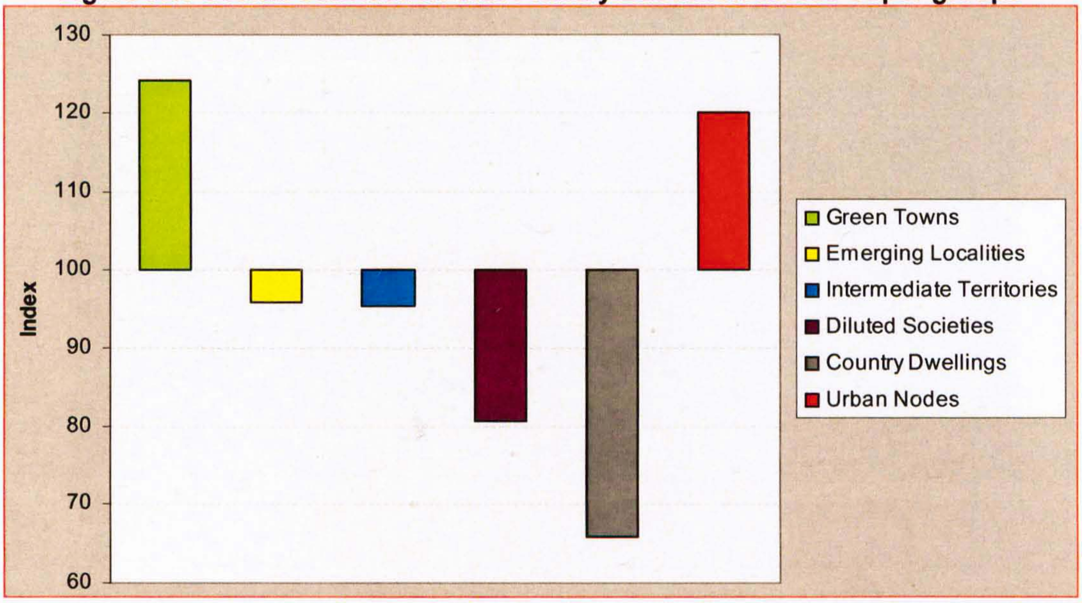
In this section, the level of satisfaction and reasons for dissatisfaction with education among primary school children is assessed while performance is assessed in the Philippines on the framework of progress, repeat and drop-out rates.

### 7.6.1 Satisfaction with Primary Education in Nigeria

The geodemographic profiling of satisfaction surveys can provide useful insight for policy directions and targeting campaigns. In scenarios where disposable resources are scarce, it can help policy makers narrow their decisions and campaign initiatives appropriately. By doing this, value for money can be assured.

In the CWIQ survey data supplied by the NBS, one of the questions directed at primary school students focused on their contentment with education. The results shown in Figure 7.6 are from analysis conducted using Equation 7.1 discussed in section 7.5.1. From the chart, the greatest level of contentment is enjoyed by residents of Green Towns and Urban Nodes where there is also a significantly higher level of adult literacy than other geodemographic typologies.

**Figure 7.6: Overall Satisfaction with Primary Education across Super-groups**




Parents of a satisfied student are more likely to recommend school attendance to other parents and vice versa. This may stimulate interest in higher rates of enrolment within these areas. A widening of the satisfaction gap may therefore contribute to increased inequality in the pattern of school attendance or enrolment. Additionally, areas of higher satisfaction with primary education are also probably likely to attract younger families although this may not necessarily be true for Urban Nodes where younger couples are often more interested in establishing their careers.

Not all areas with high rates of overall satisfaction are contented with all aspects of schooling. By focusing only on the overall rates of satisfaction, there is the possibility that the contributory factors to the levels of dissatisfaction may be concealed. This position is illustrated with the results presented in Table 7.2 which compares the overall rates of satisfaction with primary education with the key determinants of dissatisfaction at all levels of schooling. The analysis was done at the group level of the classification system and all rates have been standardised to a national mean of 100 to help uncover the varying magnitudes of inequality.

**Table 7.2: Dissatisfaction Indices for Different Factors by Groups**

| Groups | Satisfaction with Primary Education | Reasons for Dissatisfaction at Primary and Secondary Levels of Schooling | | | | |
|---|---|---|---|---|---|---|
| | | Supply of Books | Quality of Teaching | Lack of Teachers | Lack of Facilities | High Fees |
| Conventional Green Towns | 127 | 62 | 95 | 128 | 85 | 89 |
| Underprivileged Green Towns | 92 | 93 | 116 | 155 | 78 | 126 |
| Flourishing Green Towns | 122 | 54 | 84 | 86 | 77 | 140 |
| Struggling Green Towns | 140 | 75 | 50 | 66 | 83 | 141 |
| Moderately Emerging Localities | 88 | 150 | 127 | 116 | 114 | 18 |
| Comfortable Emerging Localities | 108 | 139 | 137 | 131 | 90 | 19 |
| Transient Emerging Localities | 93 | 145 | 128 | 101 | 119 | 60 |
| Constrained Intermediate | 75 | 94 | 124 | 160 | 124 | 88 |
| Well-to-do Intermediate Territories | 106 | 70 | 123 | 115 | 79 | 174 |
| Deprived Intermediate Territories | 109 | 52 | 67 | 155 | 87 | 122 |
| Customary Intermediate Territories | 76 | 147 | 65 | 75 | 78 | 211 |
| Thriving Diluted Societies | 99 | 108 | 86 | 83 | 128 | 97 |
| Labouring Diluted Societies | 76 | 100 | 107 | 135 | 131 | 68 |
| Deprived Diluted Societies | 61 | 85 | 48 | 112 | 134 | 91 |
| Modest Diluted Societies | 80 | 85 | 91 | 117 | 116 | 124 |
| Toiling Country Dwellings | 53 | 160 | 113 | 129 | 112 | 26 |
| Deprived Country Dwellings | 34 | 64 | 118 | 89 | 172 | 35 |
| Middle-class Country Dwellings | 74 | 131 | 151 | 129 | 145 | 26 |
| Prosperous Urban Nodes | 116 | 21 | 44 | 36 | 66 | 171 |
| Disadvantaged Urban Nodes | 108 | 61 | 68 | 54 | 85 | 165 |
| Average Urban Nodes | 98 | 167 | 76 | 66 | 78 | 78 |
| Affluent Urban Nodes | 134 | 48 | 74 | 45 | 78 | 178 |
| Striving Urban Nodes | 117 | 65 | 73 | 53 | 68 | 172 |

The maps in Figures 7.7 to 7.11 are also used to illustrate the national spatial distribution of different reasons affecting the level of satisfaction of primary and secondary school students in Nigeria.

Figure 7.7: School Students Dissatisfied with the Supply of Books in Nigeria

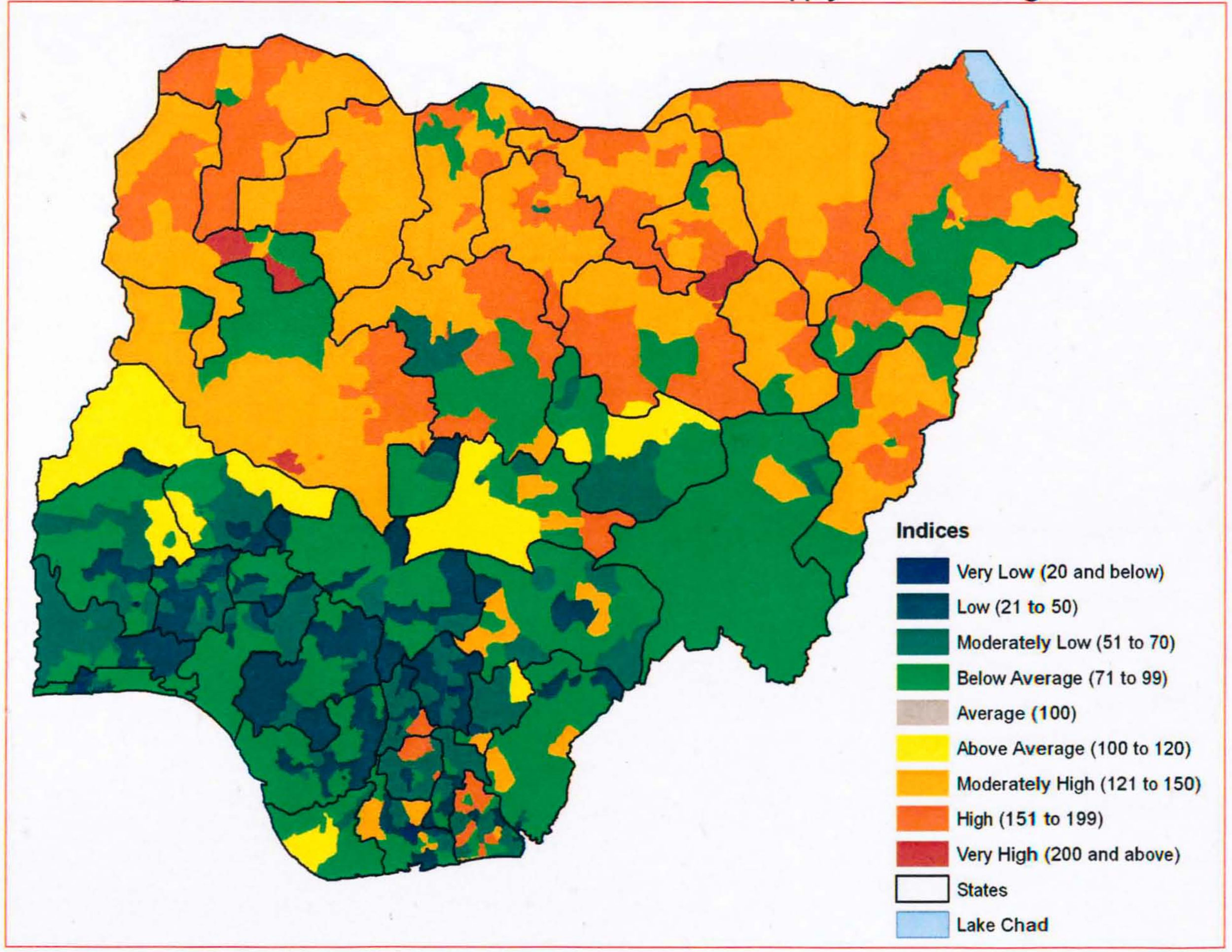


**Figure 7.8: School Students Dissatisfied with the Quality of Teaching in Nigeria**

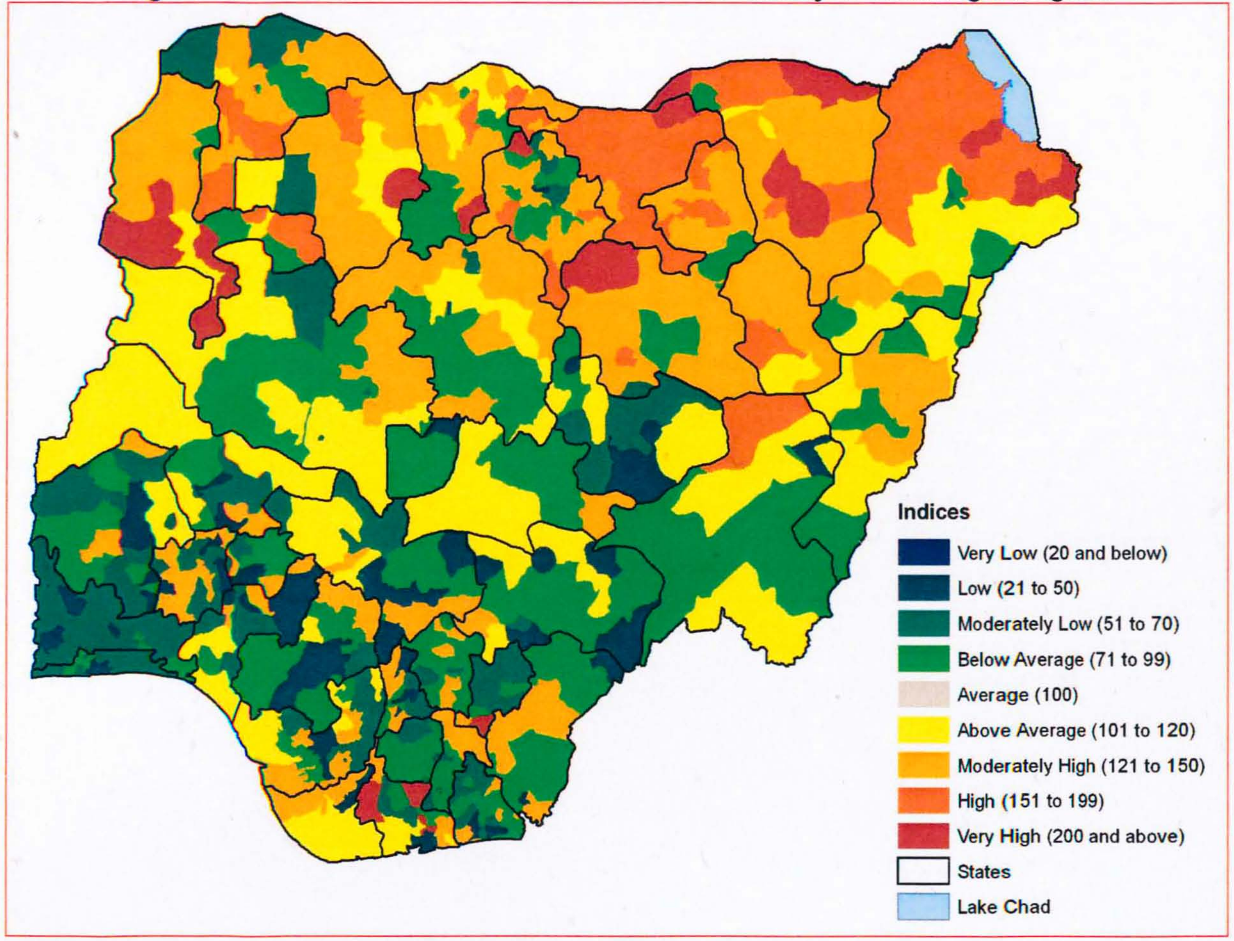


**Figure 7.9: School Students Dissatisfied Due to Lack of Teachers in Nigeria**

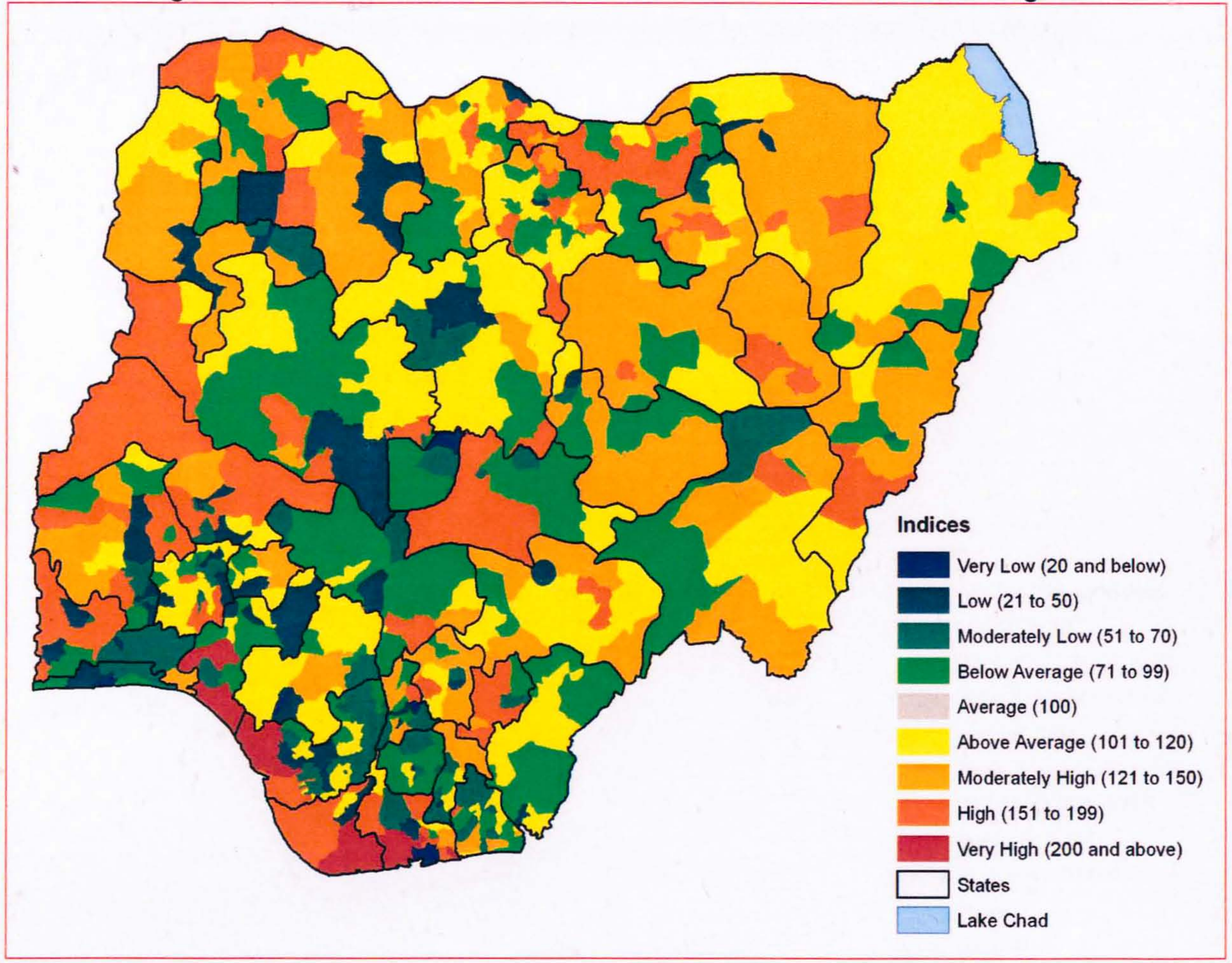


**Figure 7.10: School Students Dissatisfied Due to Lack of Facilities in Nigeria**

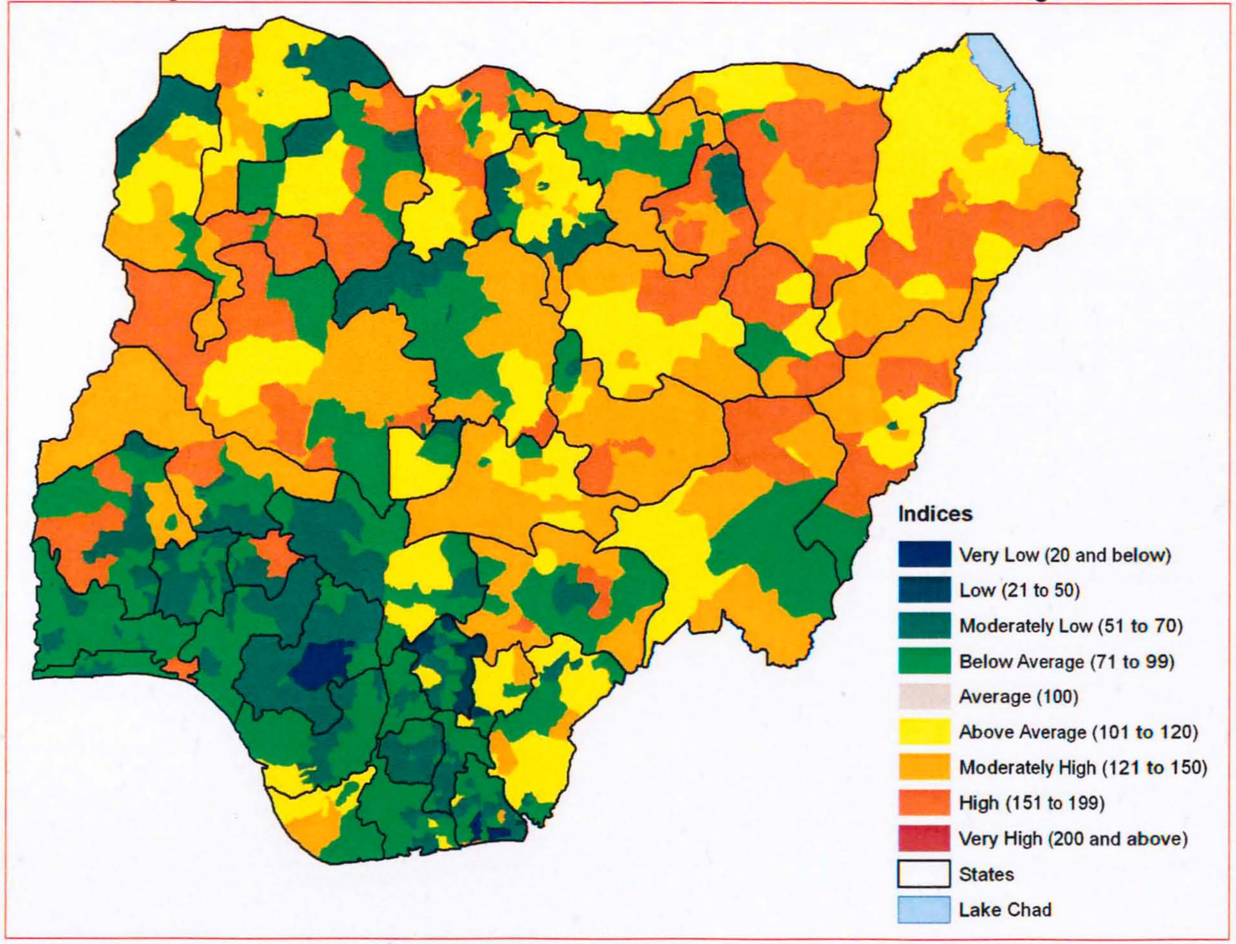
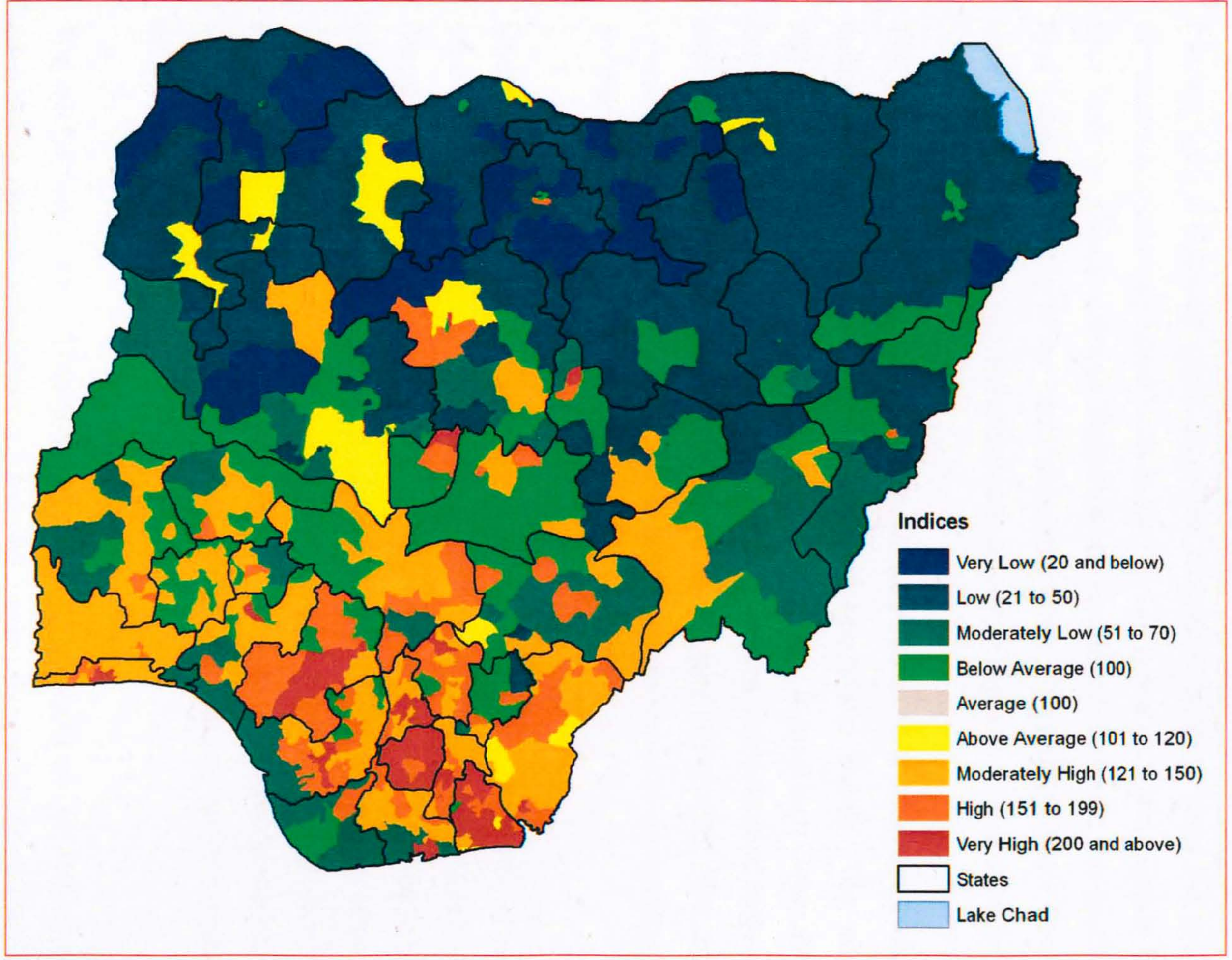


**Figure 7.11: School Students Dissatisfied with High Fees in Nigeria**



First let us examine the area types where the rate of satisfaction is above average. With the exception of Deprived Intermediate Territories, Well-to-do Intermediate Territories and Comfortable Emerging Localities, all other neighbourhood types with a satisfaction rate above the national mean only have one key factor of dissatisfaction at a rate above average. This to some extent helps explain why primary school students of these areas have relatively higher levels of satisfaction.

Primary school students within Deprived Country Dwellings have the greatest rates of dissatisfaction; the key factor contributing to this is the lack of adequate facilities. They also have an above average incidence of discontentment with the quality of teaching they receive. On the other hand, students within Struggling Green Towns where the level of satisfaction is highest are mostly unhappy with the cost of schooling.

The supply of books is a key problem for students in neighbourhoods where there are substantial representations of uneducated household heads. The engagement of parents in agricultural or transportation employment is also high within these areas. As can be seen from Figure 7.7, the bulk of areas where students tend to be very unhappy with the supply of books lie within the northern axis of the country. The map suggests a north-south divide in the rates at which the supply of books is a problem among school children.

Children who are dissatisfied with the quality of teaching are disproportionately concentrated within Emerging Localities and Country Dwellings. Middle-class Country Dwellings where polygamous marriage is common and literacy rates are relatively low demonstrate the largest incidence of discontentment with the quality of teaching. Figure 7.8 reveals that the quality of teaching is more of a problem in north-western states like Borno, Yobe, Gombe, Bauchi and Jigawa. Pockets of areas in states like Bayelsa, Rivers, Ebonyi and Osun all in the southern part of the country are also characterised by this problem.

The spatial distribution of the problem of lack of teachers shown in Figure 7.9 does not necessarily concentrate within a particular region. However, from the map, it can be deduced that the greatest need for teachers is within the oil-rich Niger Delta area of Nigeria. This finding is synonymous with a quote credited to the Governor of Rivers State – Mr Chibuike Rotimi Amaechi during a presentation at Chatham House, London in February, 2009. According to him:

*'Rivers State is the biggest in the Niger Delta, and it receives the most money from the Federal Account. Still, when I came into office the education sector had collapsed and there were no teachers. I declared an emergency in education.*

*I believe that the richest group should fund the foundations of a community, and so the state government took over control of primary education from the local government councils. There are now 250 primary schools in Rivers State. My target is to build 750 in total. There are also 2000 teachers, but we need to hire more*' (Chatham House, 2009 p. 6).

The dominance of the problem of lack of teachers spreads across groups within Diluted Societies, Intermediate Territories and Emerging Localities. Within Constrained Intermediate Territories, access to public transportation is poor and engagement in private informal employment is quite high. School children within these areas have the greatest level of dissatisfaction with the lack of teachers.

The level of dissatisfaction with the cost of school fees perhaps presents one of the most interesting revelations. It is evident from the table that children who reside within relatively affluent areas are most likely to be unhappy with the school fees they have to pay. In 2004, the Universal Basic Education (UBE) Act was passed by the Nigerian National Assembly to encourage ***free*** and compulsory education at primary and junior secondary levels (NPC, 2007). In practice however, education in Nigeria is not free as there are increasing numbers of privately owned institutions springing up mainly within relatively affluent neighbourhoods (Olubor, 2009). These institutions can charge whatever they want as school fees because of the weak government institutions. In addition, the UBE Act is not necessarily binding on them. From the map shown in Figure 7.11, the dissatisfaction with cost shows the strongest evidence of a north-south divide when compared with the other factors. School children within many of the southern states are mostly unhappy with cost of school fees.

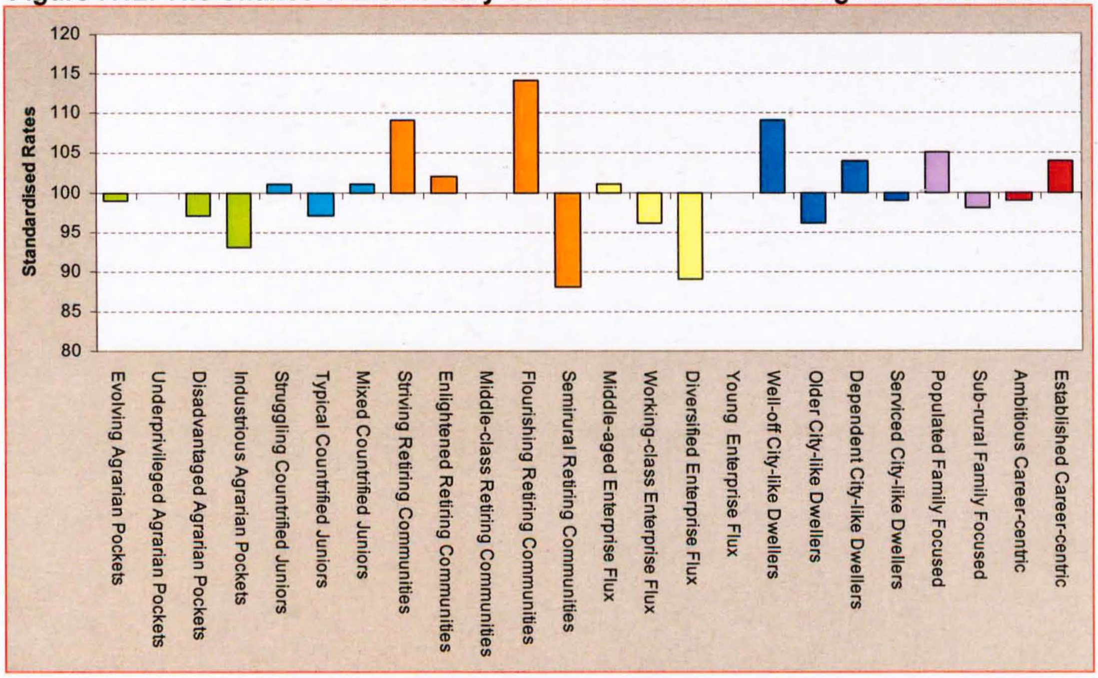
However, it is important to observe an area of the Niger Delta on the map. Evidence from Figure 7.9 indicates that school children within the same axis require more teachers. Further investigation and evidence provided by the map shown in Figure 7.11 suggests school fees may not be much of a problem to these children when compared to the national distribution. From the account of Governor Amaechi above, it appears some Niger Delta states are prepared to subsidise education which translates to the fact that teachers would also be relatively well paid (Chatham House, 2009). So what keeps teachers away from the area? Some of the issues that need to be addressed lie at the heart of improved security and community safety and the provision of other basic infrastructure for enhanced quality of life (Chatham House, 2009).

### 7.6.2 Patterns of Progress, Repeat and Drop-out Rates in the Philippines

Apart from ensuring children attend elementary school; it is also imperative that their progress is monitored. Although the MDGs fail to address them, these issues are considered pertinent for this research. Knowledge of academic performance at elementary school level can help policy makers make strategic decisions to tailor support for children. For instance, Shin (2007) found that beyond the financial background of children in Korea, peer relationships influenced their social behaviours which in turn have effects on their overall performance.

To investigate performance among elementary school children, children who are of elementary school age and who were recorded to have made progress in their academic work were profiled against the segmentation system.

**Figure 7.12: The Chance of Elementary School Children Advancing with Their Studies**



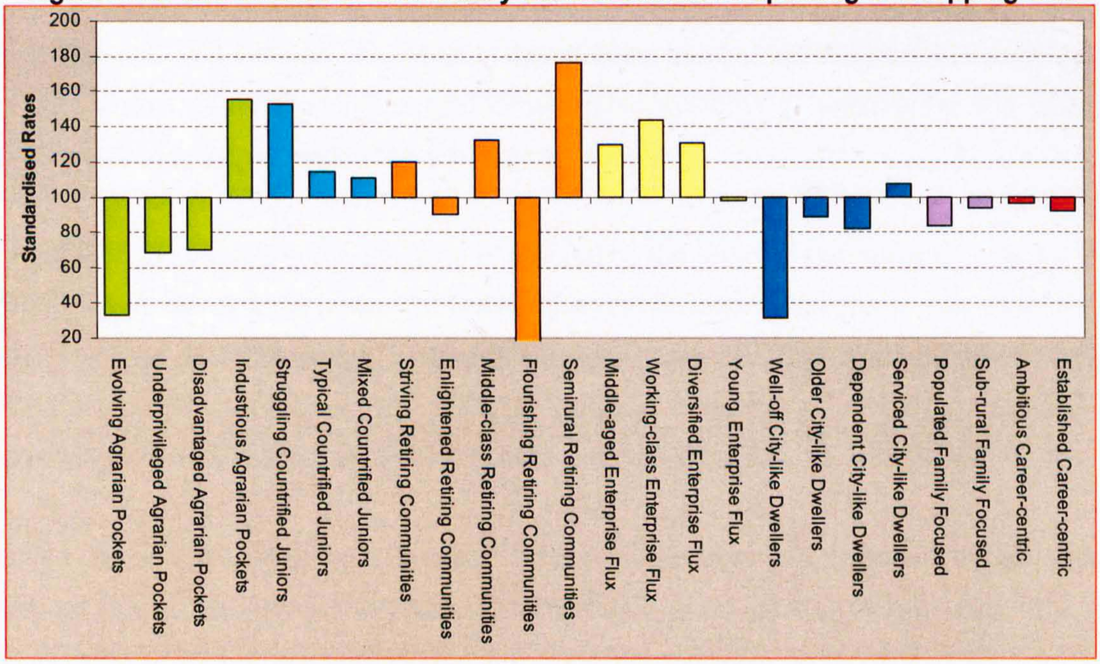

Evidence from results of the analysis reveals first that relative affluence has a positive impact on the chances of a child making progress with their academic work. Additionally, those children who live in neighbourhoods where older people appear to dominate are potentially more likely to advance forward.

Children resident in group 3.4 – Flourishing Retiring Communities demonstrate the highest propensity to advance with their elementary education. The neighbourhood type is characterised by a strong presence of people aged over 45 years and a very high incidence of widowed persons. Flourishing Retiring Communities have the greatest national incidence of female headed households which buttresses the strength of association between women and achievement at elementary school level. Commonly spoken languages within these areas are Aklanon and Tagalog.

Another revelation from Figure 7.12 is that in areas where attendance rates are relatively low, competition is not as high and as such the rates of achievement may not be as high. This is true for Semirural Retiring Communities and Diversified Enterprise Flux where school attendance rates are the lowest in the country.

Repeat and drop-out rates were also profiled to gain insight. Repeat rates are defined for children who were in a given grade during the previous school year and remain in the grade during the next school year. Drop-out rates on the other hand are computed for students who were in a given grade in the previous school year and at the time of the survey were out of school.

**Figure 7.13: The Chance of Elementary School Children Repeating or Dropping Out**




Semirural Retiring Communities have the highest rates for children who are likely to repeat or drop out of school at the elementary level. It is followed by Industrious Agrarian Pockets where also unmarried persons are numerous and there is a large presence of widowed population.

Incidentally, these are also the neighbourhood types where girls are highly unlikely to attend elementary school. What this suggests is that beyond the cultural and economic barriers that may inhibit girls from attending elementary school, it is also important to pay attention to the effects of school performance. High repeat rates which may sometimes contribute to children dropping out of school may serve as a discouragement for other parents or guardians sending their children to school particularly if they have to pay.

What may be necessary in these areas could be a need to strengthen after-school services; consider the possibility of devolving one-to-one teaching services to students; and encourage home visits by teachers or educational mentors. There should also be a strong focus on single parents, aged population groups and widows to ensure that they are supported to enhance primary level performance amongst their wards.

## 7.7 Conclusions

The research and findings described in this chapter has helped illustrate the usefulness of the Nigerian and Philippines geodemographic systems in uncovering the relative patterns of inequalities within the educational sector.

In addition to evaluating the patterns of inequality, the various examples reported in the chapter also serve as a means of testing the classification systems to see how well it discriminates for differences amongst the population groups. Much of the findings conform to previous discoveries at higher levels of geography but goes a step further revealing local level inequalities and associated geodemographic correlates.

When cross-referenced with primary school enrolment, the Nigerian classification reveals that in general, primary school enrolment is least for areas where literacy rates are generally low, household heads are under-represented within educational circles and separated couples have an above average presence.

In section 7.5.3, the analysis also debunks claims in the 2003 DHS report that no major gender differences occurred in the attendance at elementary school. The reason for that claim is because a regional level of analysis was conducted. By geodemographically profiling the data at a localised scale it was shown that significant neighbourhood differences occur which should help guide policy initiatives.

Another interesting finding is that in areas where attendance rates are low, competition is not stiff. This in turn affects the level of academic achievement of students residing in these neighbourhoods. Semirural Retiring Communities and Diversified Enterprise Flux are typical examples of these neighbourhood types.

There is a strong agreement between both countries that religious beliefs influence the rates at which parents enrol their children in school (Nguyen, 2007; NPC and ORC Macro, 2004). However the results discussed in section 7.6 prove that beyond religious inclinations, attention should also be focused on issues of contentment and performance. Significant variations and spatial differences exist at local scales in both countries requiring differentiated targeting strategies.

The findings presented in this chapter prove that geodemographic tools are not just sophisticated geo-statistical tools but are useful and important for **simplifying and addressing** policy challenges of the MDGs and other associated initiatives in Nigeria and the Philippines.

Due to the weakness of data registration systems, many of these policy initiatives are highly dependent on the type of surveys analysed in this chapter (UN, 2003). However just like the MDG reports, analysis and reporting of many national policy initiatives in both countries has often been limited to higher levels of geographical aggregation. This makes it difficult to disburse financial and material resources intelligently to ensure that at-risk or in-need population groups are reached (Deichmann, 1999). In this respect, one can say that many of these surveys have been underutilised.

The geodemographic analysis of Nigerian and Philippines primary educational data makes it possible to extrapolate the issues associated with these variables to local areas and present national pictures which were previously unknown.

Doing this shows that at the local level of governance, areas have varying levels of advantage or disadvantage which also closely matches their geodemographic correlates. Knowledge of these geodemographic correlates therefore increases the ability of policy makers to differentiate their strategies towards meeting the yearning needs of the people.

# 8

# Addressing Female Empowerment and Maternal Health Care

## 8.1 Introduction

Chapter seven presented a thorough description of how the Nigerian and Philippines systems can be used to address educational policy issues at the primary level.

Using data from the 2006 Core Welfare Indicators Questionnaire Survey (CWIQ) for Nigeria and the Demographic and Health Survey (DHS) for the Philippines, attention in this chapter shifts towards women. The analyses reported here provide insight into gender parity issues and maternal related health care in both countries.

Section 8.2 investigates the economic empowerment of women in Nigeria by investigating issues relating to the ownership of landed property and income. In section 8.3, the gender disparity in the Philippines is addressed by focusing on the level of educational attainment among women. Sections 8.4 and 8.5 focus on child and maternal related health care in both countries while section 8.6 provides some conclusive remarks.

## 8.2 The Empowerment of Women in Nigeria

One of the key concerns of the MDGs is to reduce the discrepancy in inequality between men and women in developing countries (UN, 2003). There is particular interest in the reduction of educational barriers against women, the involvement of women in non-agricultural employment and their engagement in governance. These are however not the only areas of great challenge for the international community. There is growing evidence of widening global inequality in domestic abuse and domestic violence against women and young girls (Nasir and Hyder, 2003). In times past, violence against people of female gender has been linked to the imbalance in perceptions of supremacy between men and women (UN, 1993).

Gender disparity is an important issue for Nigeria. The empowerment of women can greatly enhance the capacity of attaining several of the other goals. For instance, well educated women are more likely to take health and maternal issues more seriously and will be well informed (NPC and ORC Macro, 2004). This may in turn translate to the maintenance of relatively good health for their children and help minimise infant and maternal mortality.

In most developing countries including Nigeria, it has been argued that the semi-formal economy can help expand the middle class and reduce the scourge of poverty particularly in rural areas (Saito et al., 1994). In Nigeria, the involvement of women in the growth of cottage industries and small scale enterprises (SMEs) for instance can not be overemphasised (Kitching and Woldie, 2004). Not too long ago, the Central Bank of Nigeria (CBN) put together a policy on micro-finance to help nurture the growth of SMEs and consequently provide for the economic empowerment of those who may otherwise not be able to secure financial assistance from larger banks (CBN, 2005). This policy which could have been used an instrument for empowering women and closing the economic gap between them and their male counterparts is however devoid of any specific strategy for women.

In the absence of the specific MDG indicators for promoting gender equality outlined in Table 3.3 of chapter 3, the analyses reported here will focus on proxy indicators made available by the NBS relating to the economic empowerment of women. These include land and home ownership and the contribution of females to household income.

## 8.2.1 The Ownership of Landed Property among Women

Landed properties are key economic assets in Nigeria (Bello, 2007) and oftentimes serve as collateral for loans or some other important applications. Due to the fiscal importance of land and homes in Nigeria, there have been multiple scenarios of inter and intra-communal violence even after the promulgation of the Nigerian land use decree of 1978 (Mayowa, 2001). Since the ratification of the 1978 decree, authority over land in every state has been vested in the state governor. Due to the prolonged years of military administration in the country, the process of allocating land ownership among the citizenry has been erratic and sometimes non transparent (Koehn, 1983) and has contributed to socio-economic and spatial disparity in land ownership in the country (Williams, 1992).

Ashiru (2007) noted that during the settlement of divorces for instance, women are subjected to discrimination when allocating landed property. Again the problem is significantly linked with the perception of superiority among men. Most studies do not cover the entire country and so don't provide opportunities for comparison and benchmarking the extent of the problem. Rather they focus on selected areas or regions and fail to account for the magnitudes of neighbourhood effects that exist within areas. By explaining the neighbourhood parameters that contribute to the disparity in the ownership of landed property among women at LGA level, policy makers would better informed on how to construct and channel their messages for visible impact.

To profile the penetration pattern of ownership of landed properties among women the indices equation given in section 7.5.1 of chapter seven is used. All households headed a female who owns a home or pieces of land are compared to all other households in the country to determine the rates of ownership of landed property among women in Nigeria. The rates have been standardised about a national mean of 100.

**Table 8.1: Penetration of the Ownership of Landed Property among Women in Nigeria**

| Super-groups | Household Share (%) | | Rate (per 100 households) | | Index | |
|---|---|---|---|---|---|---|
| | Land | Home | Land | Home | Land | Home |
| Green Towns | 38 | 31 | 8 | 8 | 154 | 128 |
| Emerging Localities | 2 | 3 | 1 | 1 | 14 | 18 |
| Intermediate Territories | 35 | 35 | 14 | 17 | 262 | 259 |
| Diluted Societies | 10 | 10 | 4 | 5 | 80 | 75 |
| Country Dwellings | 2 | 3 | 1 | 2 | 23 | 30 |
| Urban Nodes | 12 | 18 | 3 | 5 | 52 | 79 |

The information contained in Table 8.1 reveals strong evidence of low rates of both land and home ownership among women in the country. Out of every 100 households in most areas, there are less than 10 where females own either a land or a home. In both Emerging Localities and Country Dwellings, there is less than half a chance of the national mean for women to own a landed property. Even within areas with city-like characteristics where one would expect that literacy rates should influence perceptions on these issues, the indices reflect a lower than expected penetration among women.

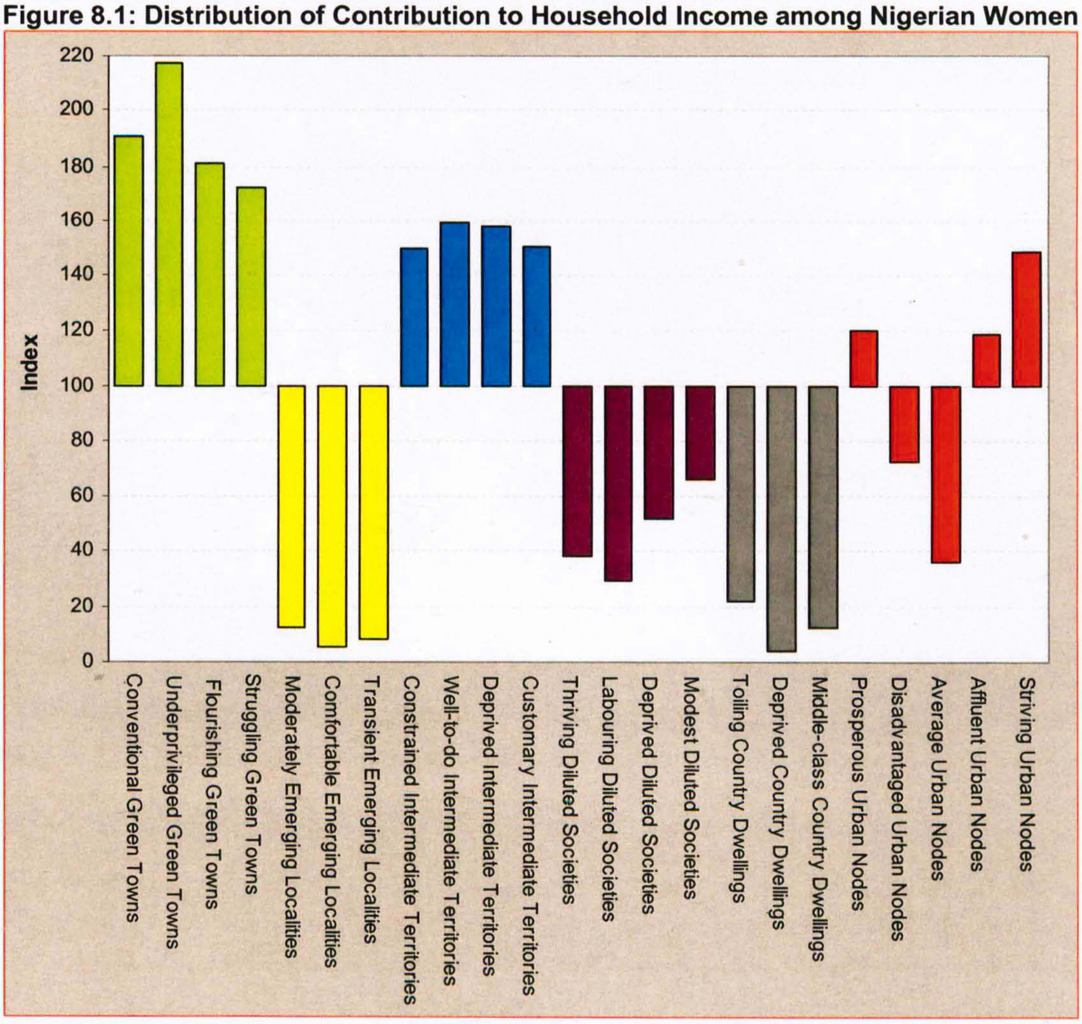
In both scenarios (land and home ownership) Intermediate Territories which are largely concentrated in the South Eastern part of the country record a very significant incidence of home ownership. Intermediate Territories are generally characterised by middle aged and older adults. Many have completed their secondary education and adult literacy rates are above the national average within these areas. Monogamous marriage is also significantly higher than the national mean within these areas.

However, there is a disproportionate concentration of single parent households and a very high number of people find it difficult to meet their day to day needs. The pattern above suggests that women within Intermediate Territories have a chance of more than double the national average to own a landed property and that greater effort should be directed towards educating and encouraging those within Emerging Localities and Country Dwellings where polygamy is common.

## 8.2.2 The Contribution of Females to Household Income

The capacity of a woman to contribute meaningfully to the overall income of her household can be a useful measure of the economic value of that woman. When women partake in the financing of the home, it means that they are regarded as stake holders and can also influence decision making within their families. The home should form the epicentre for the process of involvement of women in national issues. Women who are encouraged to amass economic power and contribute to their homes are more likely to grow in confidence to the level that they believe that they can also engage in political governance with a view to contributing to the larger society.

**Figure 8.1: Distribution of Contribution to Household Income among Nigerian Women**

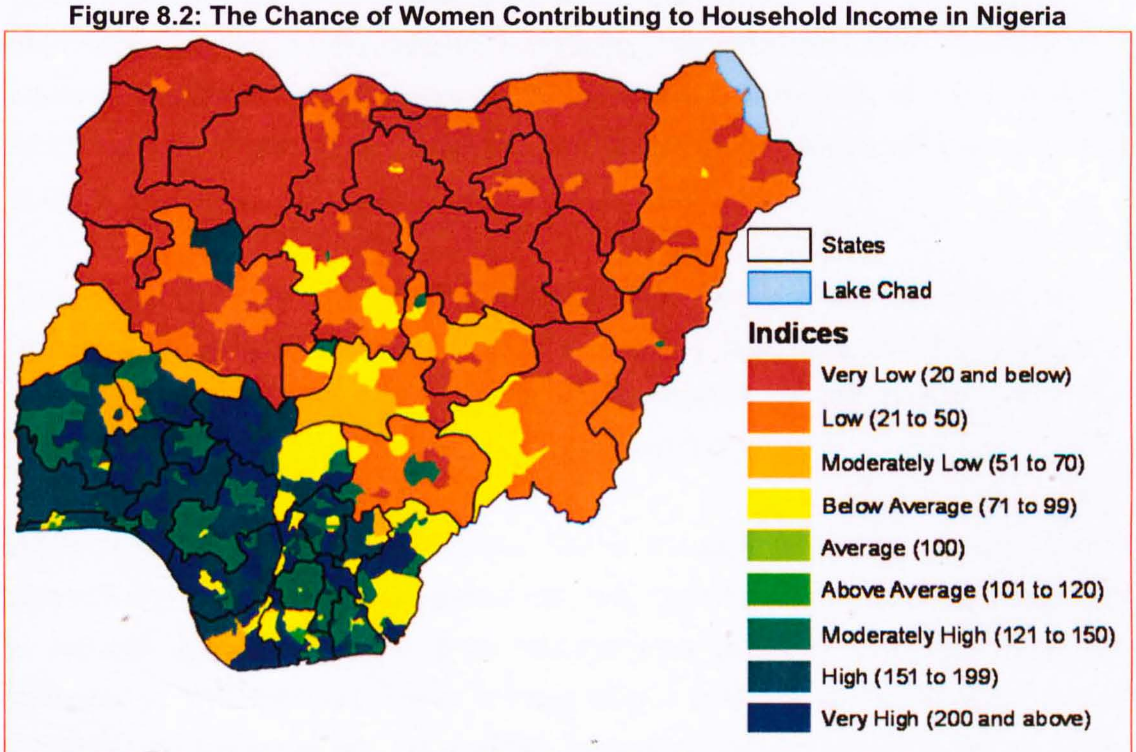


Initial analysis of the data provided revealed a slightly higher national rate for the contribution of women to household income than for the ownership of landed property. About 1 in 10 households have women contributing to the income of their homes in comparison with rates of 5 and 6 per 100 for land and home ownership respectively.

In spite of the slightly higher rate of contribution to household income among women, massive inequalities do exist within community types as shown in Figure 8.1. There are very low probabilities for women to engage in income financing within homes located in Emerging Localities and Country Dwellings; a pattern replicated by their performance with the ownership of landed properties.

Women in households located within Green Towns have the highest indices for contributing to family income. Surprisingly, Underprivileged Green Towns prove to be the areas where women contribute the most to household revenue. There are large concentrations of widowed persons within these areas and a lot of unmarried people with a high proportion of single parents. These areas also have a very high concentration of separated couples. Underprivileged Green Towns are also characterised mainly by households of 1 to 2 persons.

**Figure 8.2: The Chance of Women Contributing to Household Income in Nigeria**



The spatial distribution of households where women contribute significantly to revenue shown in Figure 8.2 reflects a clear north south divide. Women living up north have a lower chance of being economically empowered than their southern counterparts who are relatively better educated (NPC and ORC Macro, 2004).

## 8.3 Educational Attainment among Women in the Philippines

As far back as the early 1980's, evidenced based research has shown that the education of women in the Philippines contributes immensely to the country's economic growth (Crawford and Sidener, 1982).

The educational attainment of women is also likely to impact significantly on the health and livelihoods of their children (Caldwell, 1981). In the Philippines, Becker et al. (1993) found that educational level of women greatly influenced their health choices and behaviours particularly as it relates to maternal and child health. Their findings also provide evidence of an urban-rural dichotomy in varying impacts of female educational attainment with rural areas appearing to be hit the greater.

In a bid to improve access of women to education and mitigate the challenge of gender equity especially within the educational sector, the government of the Philippines in 1987 introduced within the national constitution, an affirmation of the equality of all citizens irrespective of their gender (Constitutional Commission of the Philippines, 1987). In spite of this legal declaration, considerable inequalities still exist within the country.

There is strong evidence of the impact of traditional beliefs which discriminate against citizens of the female gender especially in the country side (ADB, 2004). Even in urban centres, female political and economic empowerment is still minimal essentially because of the disparity in the educational attainment of women.

As is common practice with the other topical issues investigated, the problems of illiteracy and educational attainment in the Philippines has never been investigated on a national scale beyond the large administrative areas or from an urban rural perspective. This approach makes it more difficult to target resources and campaign strategies to local areas with the greatest potential of accruing benefit to the population in need.

In the next two sections, a geodemographic approach will be deployed to investigate the depth of illiteracy among women in the country at the barangay scale and to uncover at the same scale variations in the patterns of educational attainment among women.

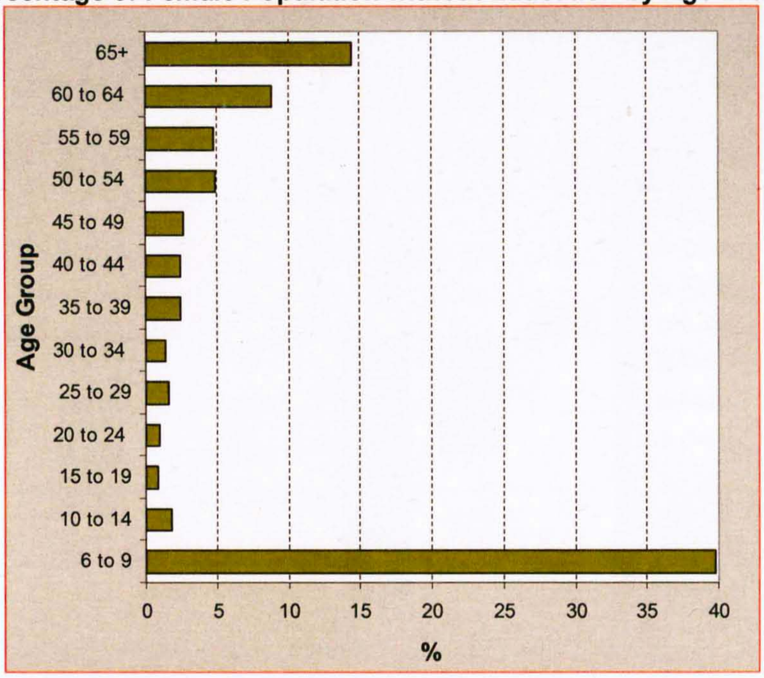
### 8.3.1 Depth of Illiteracy among Women

In the third chapter of this thesis, the third MDG which is concerned with promoting gender equality was discussed in detail. Female education has already been identified as an important weapon for closing the gap between male and females particularly in socio-economic terms.

Of the three indicators assessed by the UN for MDG 3, the first one is concerned with educational attainment among women. In particular, the indicator is concerned with the ratio of girls to boys at different levels of schooling. The processes involved in computing the indicator is discussed in section 3.2.3. In the absence of this indicator, a proxy indicator is used to geodemographically assess literacy levels among women in the country.

Results from initial analysis of the DHS data revealed that among females aged 6 and above, 7% had no formal education (NSO and ORC Macro, 2004). The majority of these women, about 9.6% were found to be concentrated within rural areas while 5.3% of women residing in the urban areas had no formal education. Figure 8.3 shows females without formal education by age groups.

**Figure 8.3: Percentage of Female Population without Education by Age in the Philippines**



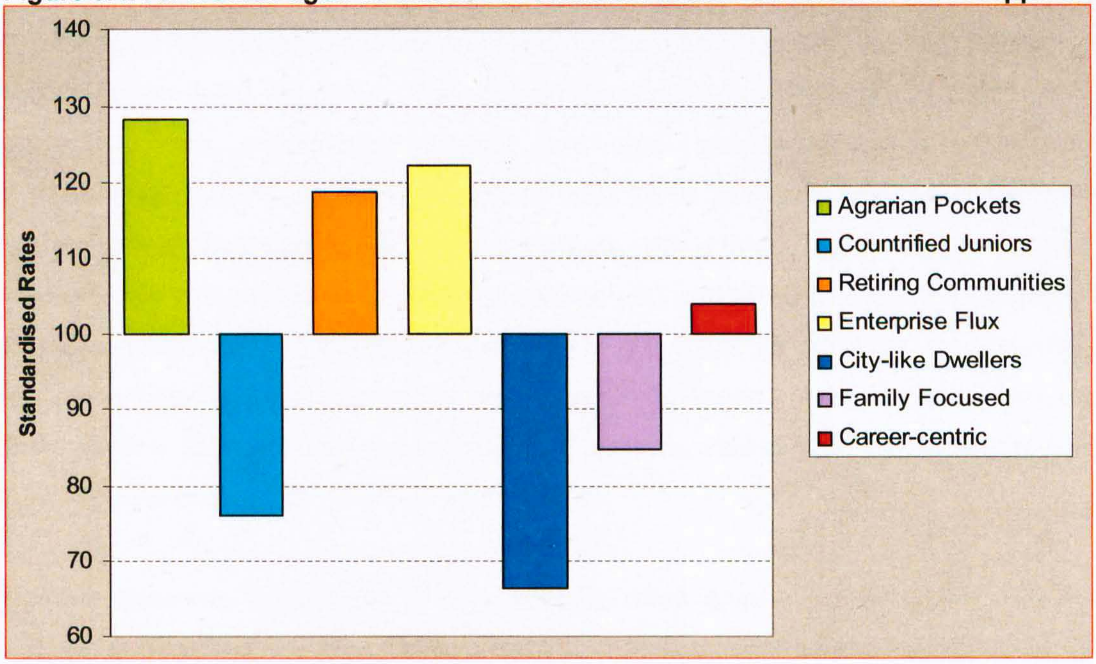

Source: NSO and ORC Macro, 2004

From the chart above, there is a rather large share of females without any formal education amongst the very young population aged between 6 to 9 years. In spite of the fact that the legitimate age for commencing elementary schooling is 6 years, almost half of the female population interviewed between the age of 6 and 9 were not literate. However, by the time girls get to age 10, there is a dramatic reduction of about 38% in the percentage of non literate females suggesting that most girls do not start school on time or at the right age.

Again as with other indicators, the information has only been disaggregated and analysed at the regional level which means it is impossible to embark on localised targeting of resources. By analysing the data with the geodemographic system, one can unearth neighbourhood effects relating to illiteracy among women and the information can be mapped locally with the aim of deploying a social campaign aimed at raising the awareness of the importance of education for members of the female gender.

From age 15, it is expected that primary and secondary level education should have been attained in the Philippines. As such, women aged 15 years and above were profiled to investigate the depth of illiteracy in the country. Figure 8.13 shows the standardised rates for women without any formal education.

**Figure 8.4: All Women aged 15 and above Without Formal Education in the Philippines**




Clear evidence from the chart above suggests that efforts aimed at improving gender parity within the sphere educating women should be focused in neighbourhood types that are rural where residents are predominantly engaged in agriculture and fisheries and the public sector. Dependency ratio is also quite high within these areas. Additionally such areas have a representative concentration of people in need of disability related care.

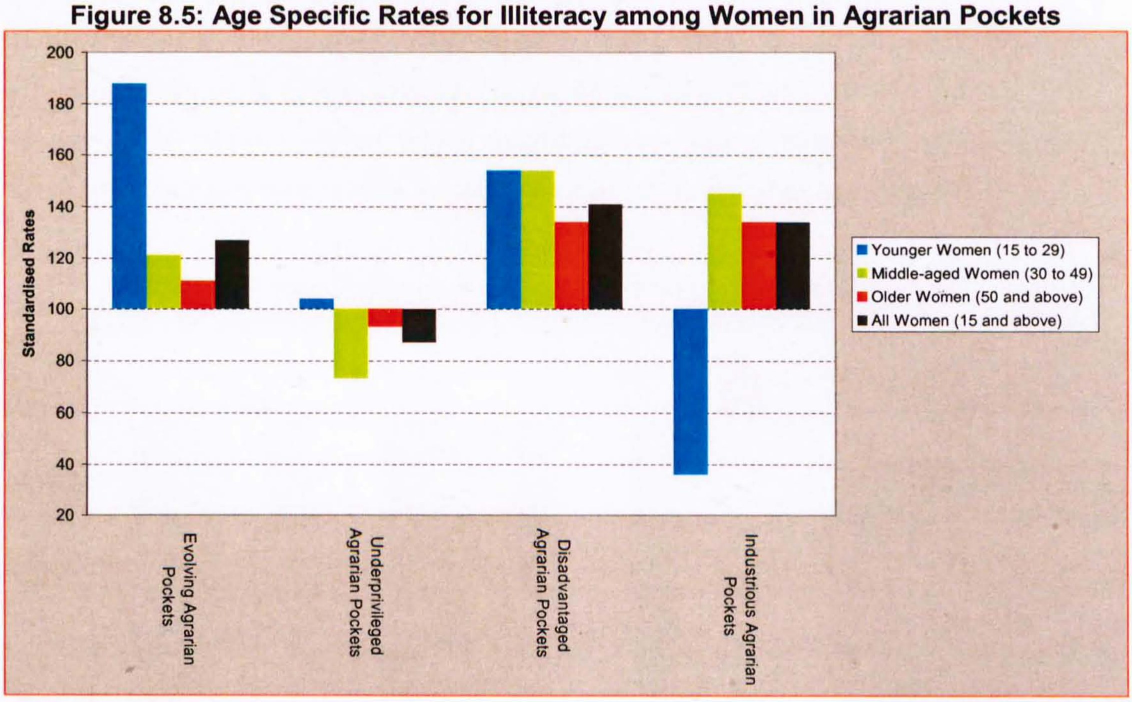
## 8.3.2 Age Standardised Variations in Illiteracy among Women

As the age of respondents to the DHS is available, it is possible to derive age-specific rates of the incidence of illiteracy by neighbourhood types. Knowing this may serve as a way of providing insight into the trend of literacy performance among women within different neighbourhood types. Comparing literacy levels among women by age will not only shed light into the magnitudes of inequality existing between neighbourhood types but also help shed some light on the effectiveness of area based initiatives for women at different life courses.

The analyses are conducted at the group level of the classification hierarchy. All women aged 15 and above were analysed as they are expected to have attained at least a primary level education. The women are subdivided into three groups defining younger women, middle-aged women and older women. Younger women are constituted by three five year age cohorts from 15 to 29 years who often are at the early stages of their careers or nursing mothers with infants. Middle-aged women on the hand combine women who are aged between 30 and 49 years who are expected to be relatively established career-wise or mostly guardians for teenage children. Older women are aged over 50 and define women who are nearing retirement or retired from active service. Many women within this group are also widows some of whom help look after their grand children.

First we examine the pattern of age specific rates among groups within Agrarian Pockets. In Figure 8.4 in the previous section, it has already been shown that for this super-group, the overall rate for illiteracy among all over 15 females is the highest for the whole country. When the rates are further disaggregated as shown in Figure 8.5, some interesting revelations are derived.

**Figure 8.5: Age Specific Rates for Illiteracy among Women in Agrarian Pockets**




Within Evolving Agrarian Pockets, women at all the different life stages have above average illiteracy rates. It is however most pronounced for younger women which should be a more serious course for concern. These neighbourhoods have concentrations at almost double the national average rate for younger women.

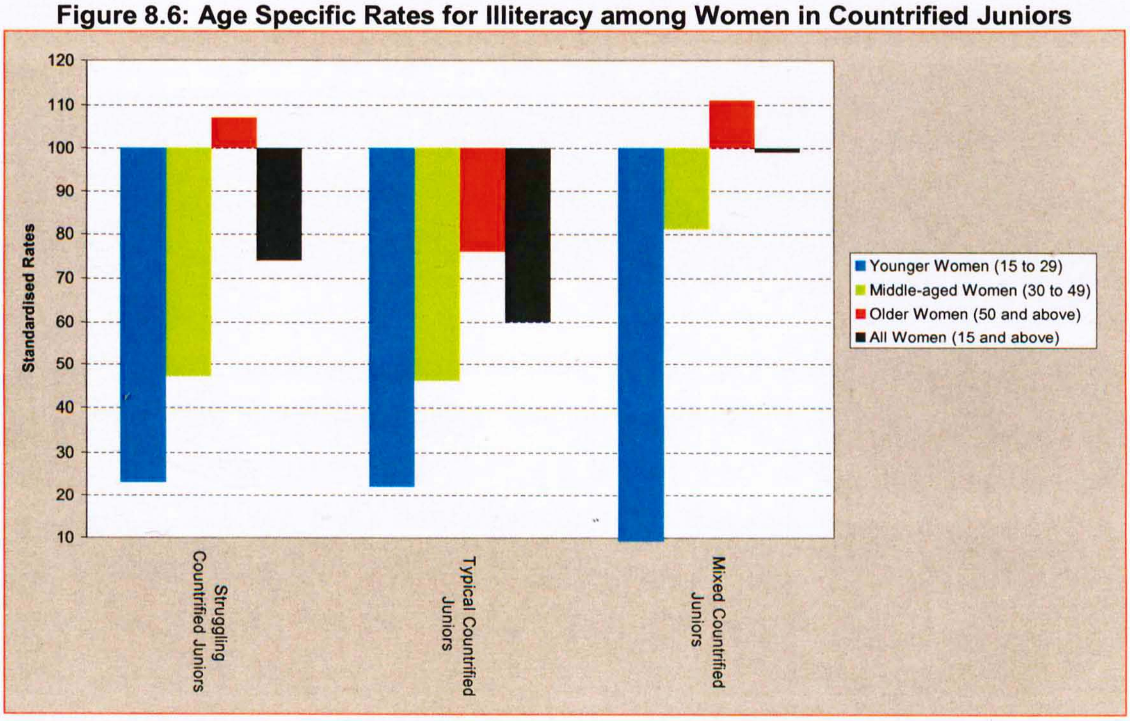
In general, despite the fact that a representative number of residents (male and female) lack any form of education within Underprivileged Agrarian Pockets, the overall concentration of all women above age 15 who are educationally disadvantaged is below the national mean although again there is just a slightly above average concentration of illiteracy rates among younger women

Evidence from Figure 8.5 proves that Disadvantaged Agrarian Pockets contribute the greatest to the high overall rate of illiteracy among women within Agrarian Pockets particularly due to high rates of illiteracy among younger and middle-aged women. In spite of this fact, the rate of illiteracy among younger women within this subgroup is still not as high as that of Evolving Agrarian Pockets.

Although the Industrious Agrarian Pockets have the second highest rates of illiteracy among all the women groups, it is particularly concentrated amongst middle-aged and older women. It is gratifying to note that the younger generation of women are better educated with relatively below average rates of illiteracy within the neighbourhood type.

One final look at the trends of the female groups in Figure 8.5 shows that the trend for older women is most closely associated with the trend for all women aged over 15. What this tells us is that if a national area based educational initiative was targeted at all women at random without taking neighbourhood age differences into consideration, such an initiative would mostly benefit the older women within Agrarian Pockets.

**Figure 8.6: Age Specific Rates for Illiteracy among Women in Countrified Juniors**

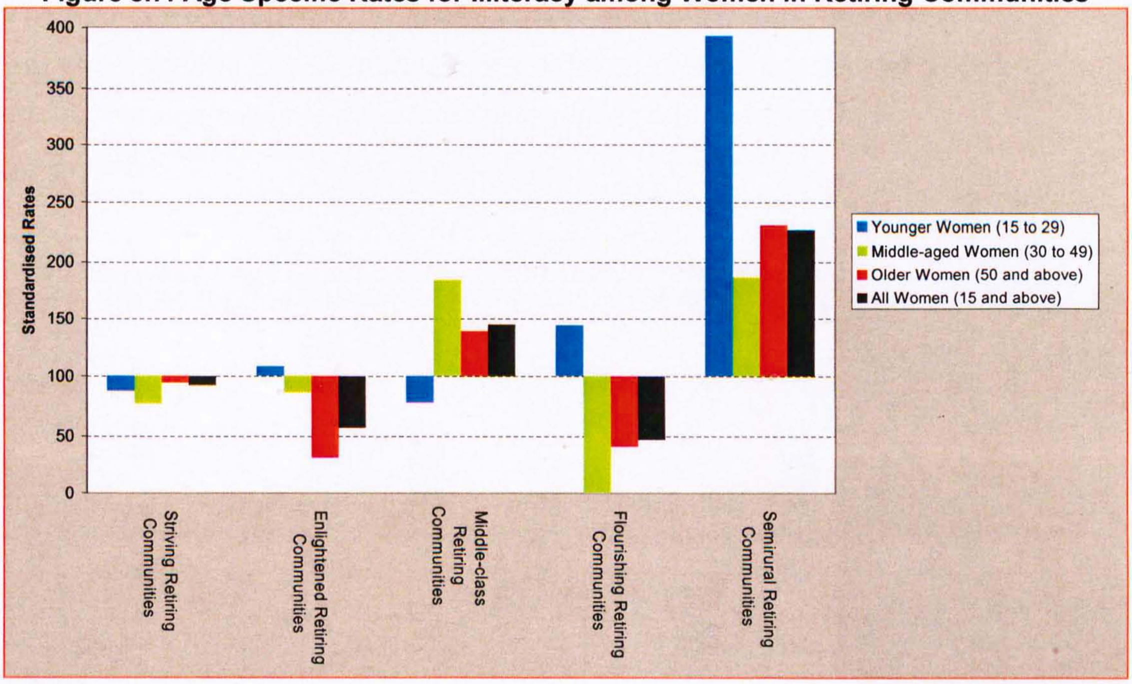


In Figure 8.6, the variations for Countrified Juniors are illustrated and reflect significant differences from those of Agrarian Pockets just described. In general, the overall rates of illiteracy among the three groups are beneath the national mean all be it at different magnitudes. The overall level of illiteracy among all women is greatest among Mixed Countrified Juniors where female household heads are more dominant than their male counterparts. However a more cheering revelation is the fact that the younger generation of women within these areas have the best prospects of education.

Again as can be determined from the chart, any national area based initiative targeted at all women aged 15 years and above would be most suited towards middle-aged and older women within Countrified Juniors.

Retiring Communities present an even more interesting pattern in the variation of the depth of illiteracy during the life courses of women by neighbourhood types.

**Figure 8.7: Age Specific Rates for Illiteracy among Women in Retiring Communities**



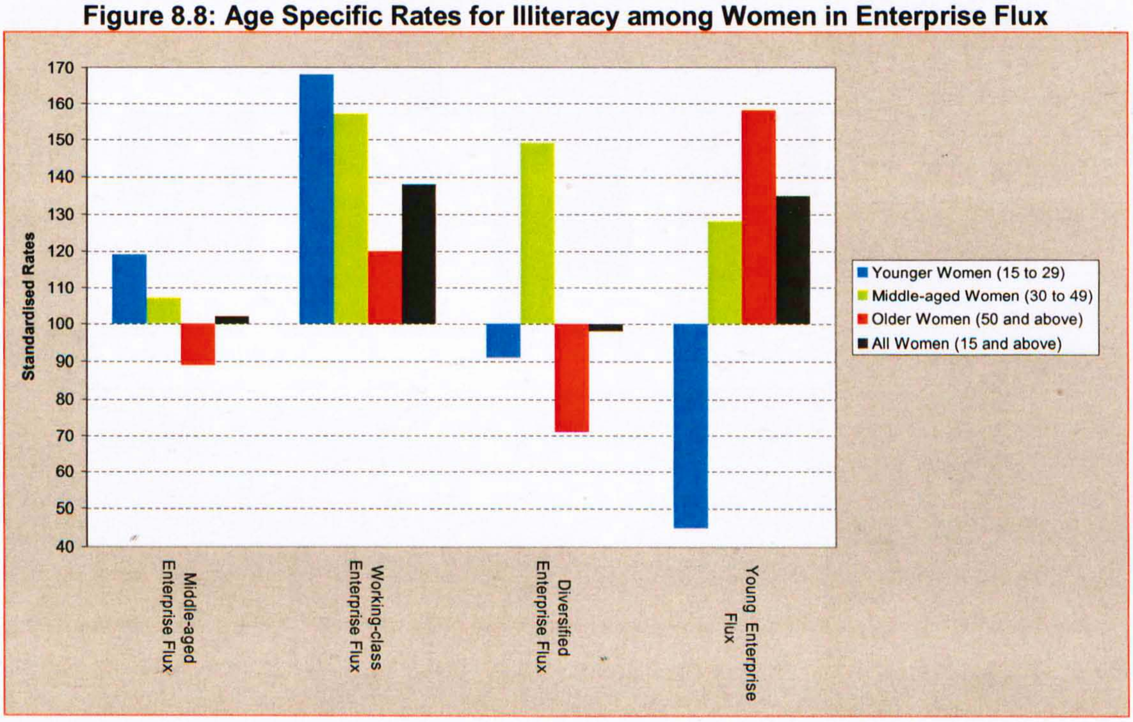

From Figure 8.7, we observe for the first time a group where all the different groupings of women have significantly above average incidence of illiteracy. Not only do Semirural Retiring Communities have the largest rates of illiteracy for all the age groups, it is also into these neighbourhood types that younger illiterate females are mostly concentrated. The rate of illiteracy among women within this age group is almost four times the national mean distribution. The most dominant population shares within this group can be found in Southern Tagalog, Western Visayas and Central Visayas regions.

In spite of the fact that secondary school completion rates (among males and females combined) within Flourishing Retiring Communities are moderately high when compared to the national mean and people also go on to study higher education at above average levels, there is evidence that the depth of illiteracy among younger females is still significantly higher than the national average distribution within the neighbourhood type.

The pattern for the overall illiteracy rates for all women above is most similar to that of women in the older age groups which provides grounds to suppose that national area based initiatives aimed at reducing female illiteracy which fail to take cognizance of neighbourhood differences in age would work for the older women within Retiring Communities. If the intention is to target younger women, then neighbourhood inequalities by life stage matters rather than just randomly deploying a campaign at all women.

In Figure 8.8, age specific inequalities in the level of illiteracy by neighbourhood types are even starker. The variability in the patterns of the magnitude of female illiteracy among the age categories are more pronounced for Enterprise Flux than for any other super-group.

**Figure 8.8: Age Specific Rates for Illiteracy among Women in Enterprise Flux**

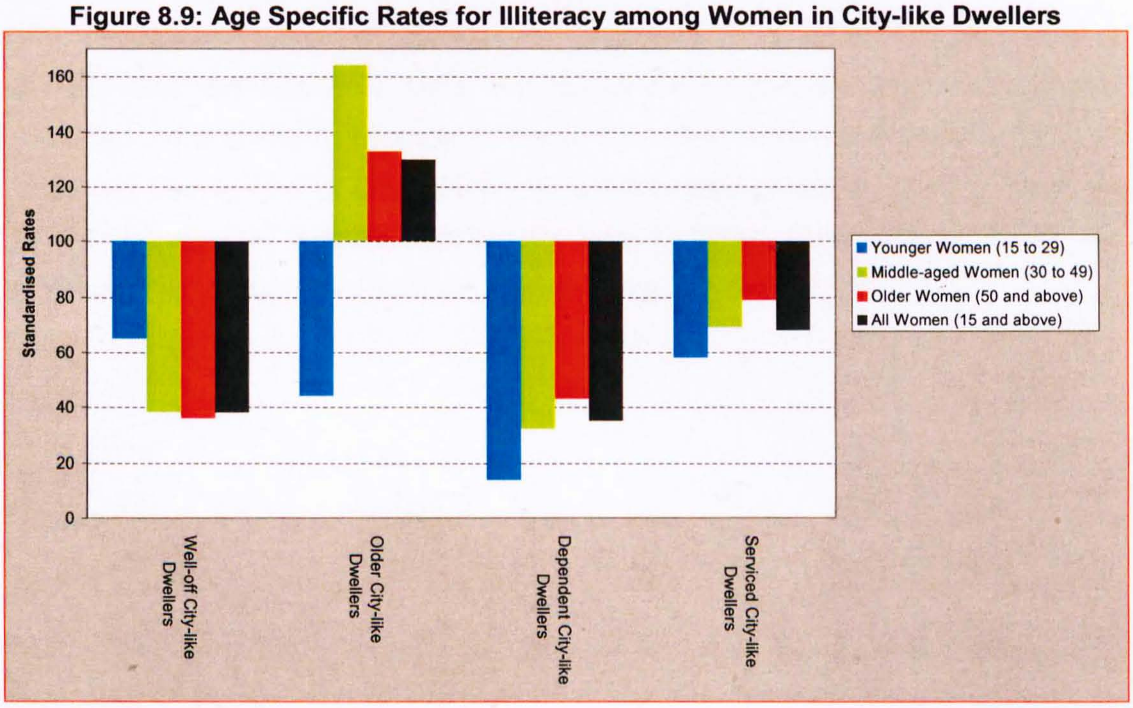


Although when all women are considered, Young Enterprise Flux has an above average incidence of female illiteracy due to the concentration less educated females particularly in the older age groups, the neighbourhood type has one of the lowest national rates for illiterate young females. Amongst all the groups within Enterprise Flux, only this group has the least population density and generally above average incidence of young children typically aged 0 to 20 years.

Again as with the other charts reviewed above, older women within Enterprise Flux are more likely to benefit the most from a random campaign targeted at all women to reduce illiteracy if neighbourhood inequalities in age are not given appropriate consideration.

In the previous section, Figure 8.4 revealed that when all women aged over 15 are considered residents of areas categorised as City-like Dwellers have the least chance of being without formal education. This is not unsurprising as these areas are mostly modernised with general access to public facilities. However as shown in Figure 8.9, significant variations do exist on the basis of age among women at the subgroup level.

**Figure 8.9: Age Specific Rates for Illiteracy among Women in City-like Dwellers**



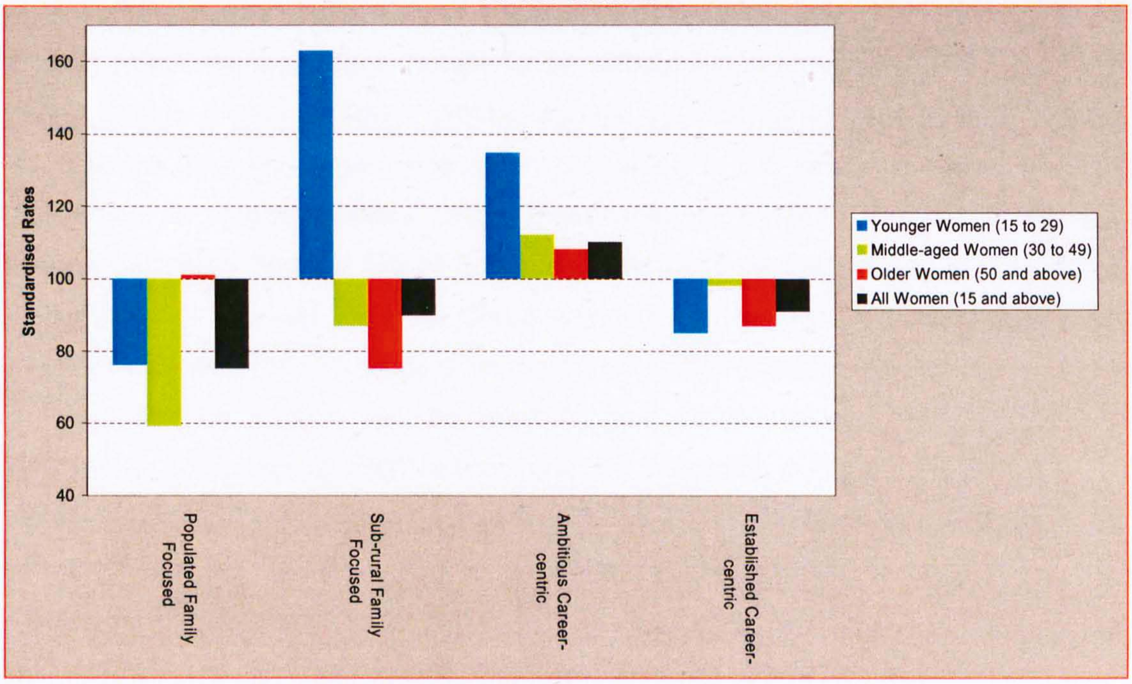
Amongst all City-like Dwellers, the Older City-like Dwellers exhibit the largest overall rates for women who are educationally disadvantaged. There is a major contribution of middle-aged and older women to the level of illiteracy among women. These areas are characterised by high proportions of people in older age categories. Many residents are 65 years and older. There is also a high representation of widowed population within these areas. Households of two or more married couples are just above average when compared to the national distribution. Commonly spoken languages include Hiligaynon, Aklanon, Maranao and Karay-a.

In general, and across the four groups within City-like Dwellers, the rate at which younger women are excluded from education is below the national average. With the exception of the Well-off City-like Dwellers, younger women within the other three groups have lower illiteracy rates than the other two categories of women.

The pattern of illiteracy rates among all women across the four groups suggests that if all women within City-like Dwellers were targeted randomly without taking geodemographic differences in age into consideration, such an effort would most likely favour women in the older age categories.

Groups within the Family Focused and Career-centric geodemographic clusters are combined in the analysis presented in Figure 8.10. Both Sub-rural Family Focused and Ambitious Career-centric neighbourhoods exhibit rates at above average levels for younger uneducated females. Additionally, only Ambitious Career-centric areas have all three groups of women above the national mean.

**Figure 8.10: Age Specific Rates for Illiteracy among Women in Family Focused and Career-centric**




Areas defined as Ambitious Career-centric have one of greatest incidence of population density in the country. There are also vey many unmarried persons in these areas. In terms of age distribution, there is a very great concentration of middle-aged persons between 21 and 44 years. The presence of all other age groupings is less than the national mean. There are also very many households which are headed by females. These are much more than male headed households.

Evidence from the trends in the chart provides grounds to conclude that a national initiative targeted at women randomly without taking geodemographic and age variations into consideration would suit middle-aged women within family Focused and Career-centric neighbourhoods.

From the DHS survey, the Administrative Region in Muslim Mindanao (ARMM) recorded the largest proportion of women without formal education. About 23% of women within the region are thought to be without formal education (NSO and ORC Macro, 2004). When ARMM is profiled against the Philippines segmentation system, we discover that the largest population proportion in the region equating to about 10.7% live in neighbourhoods called Disadvantaged Agrarian Pockets. If we also retrace our steps back to Figure 8.5 above, we observe that among all the groups within Agrarian Pockets, only the Disadvantaged Agrarian Pockets have significantly high rates for all three groupings of women. This relationship provides evidence of how the classification system can be used to gain insight beyond regional levels of geography. Four maps of illiteracy are shown for the region in Figures 8.11 to 8.14. The maps (the first of their types to be produced in the Philippines) take into cognizance the effects of age standardised rates by geodemographics and are mapped at the most local scale of geography for the region.

**Figure 8.11: All Women (15 years and above) Without Formal Education in the ARRM**

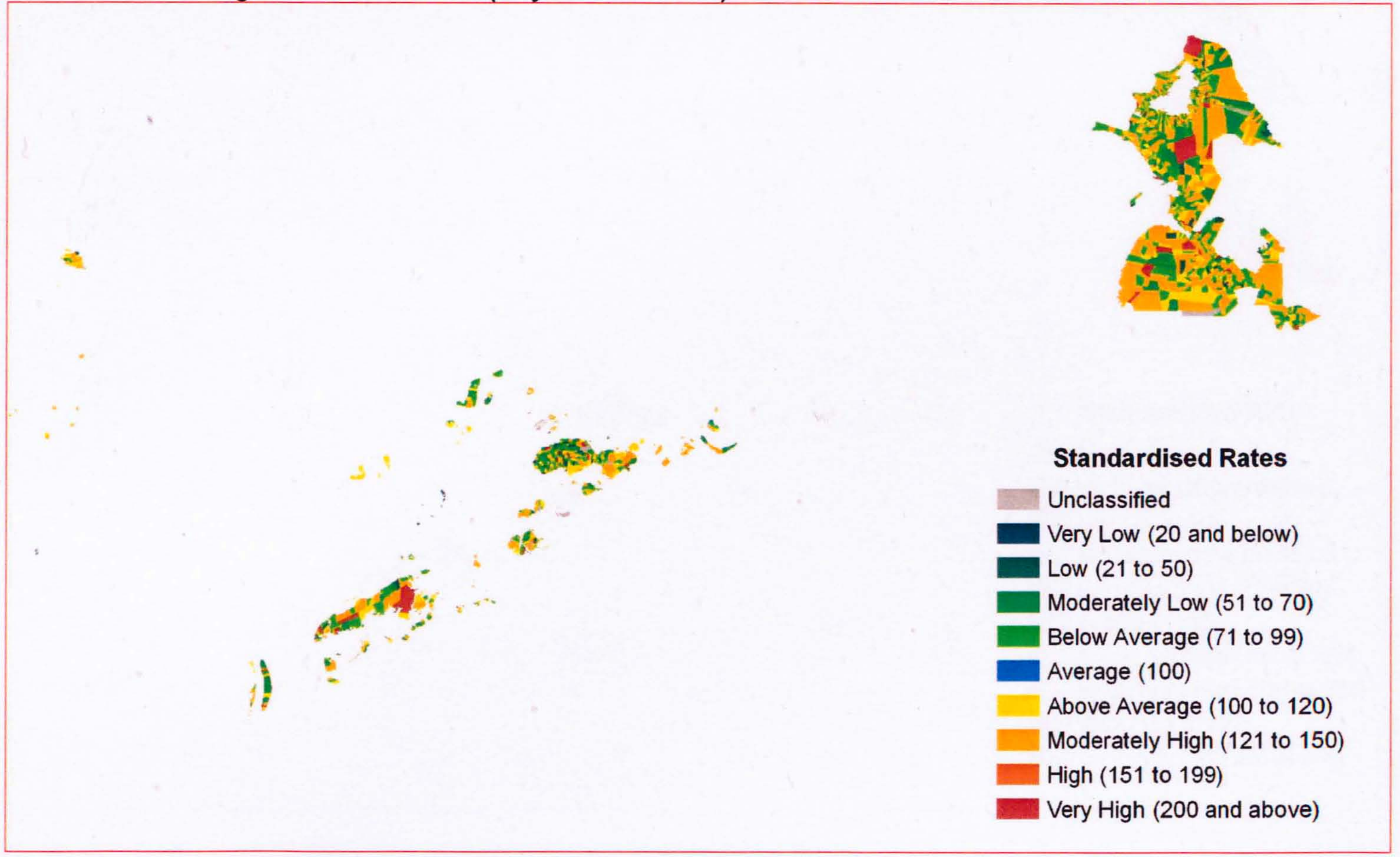


**Figure 8.12: Younger Women (15 to 29 years) Without Formal Education in the ARRM**

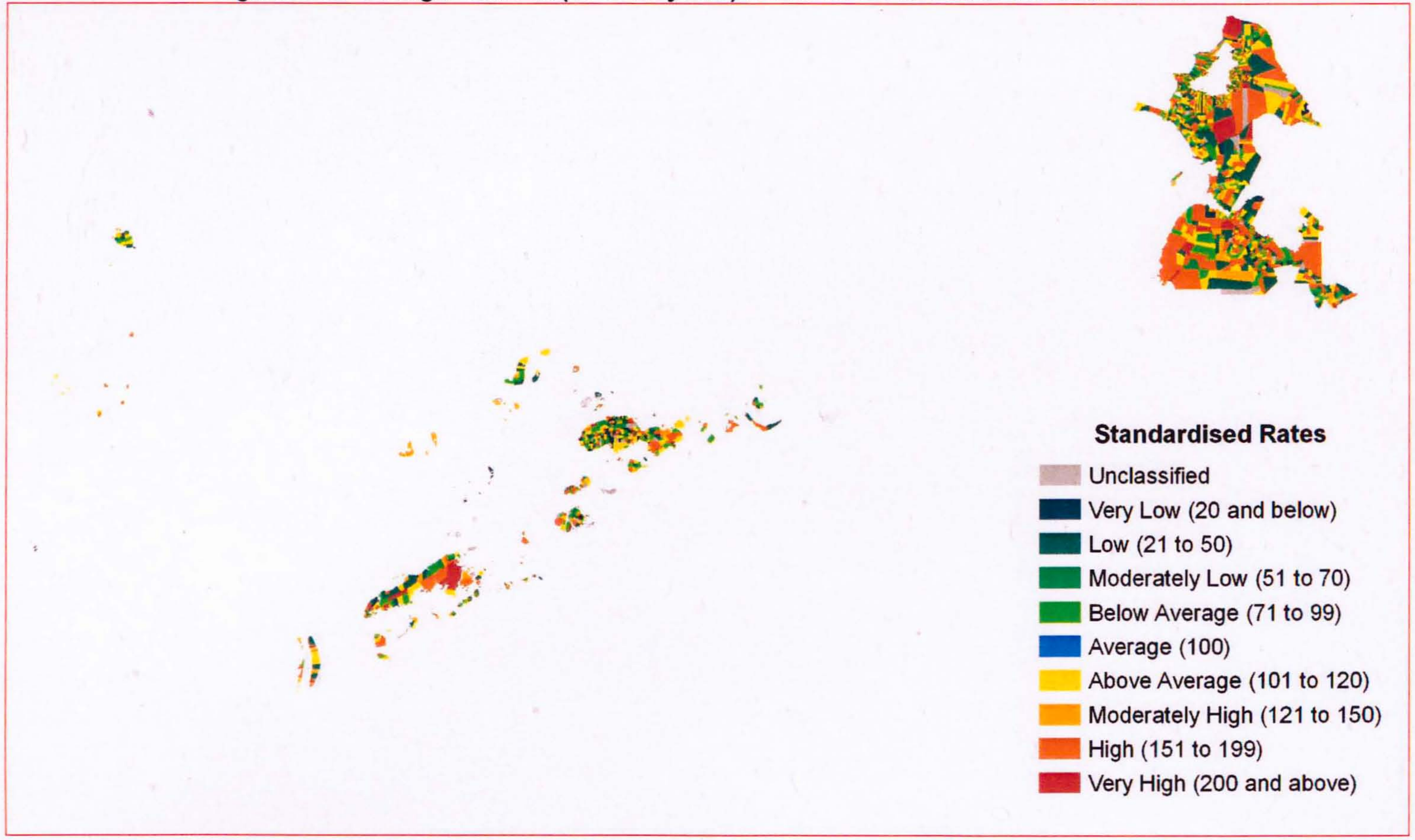


**Figure 8.13: Middle-aged Women (30 to 49 years) Without Formal Education in the ARRM**

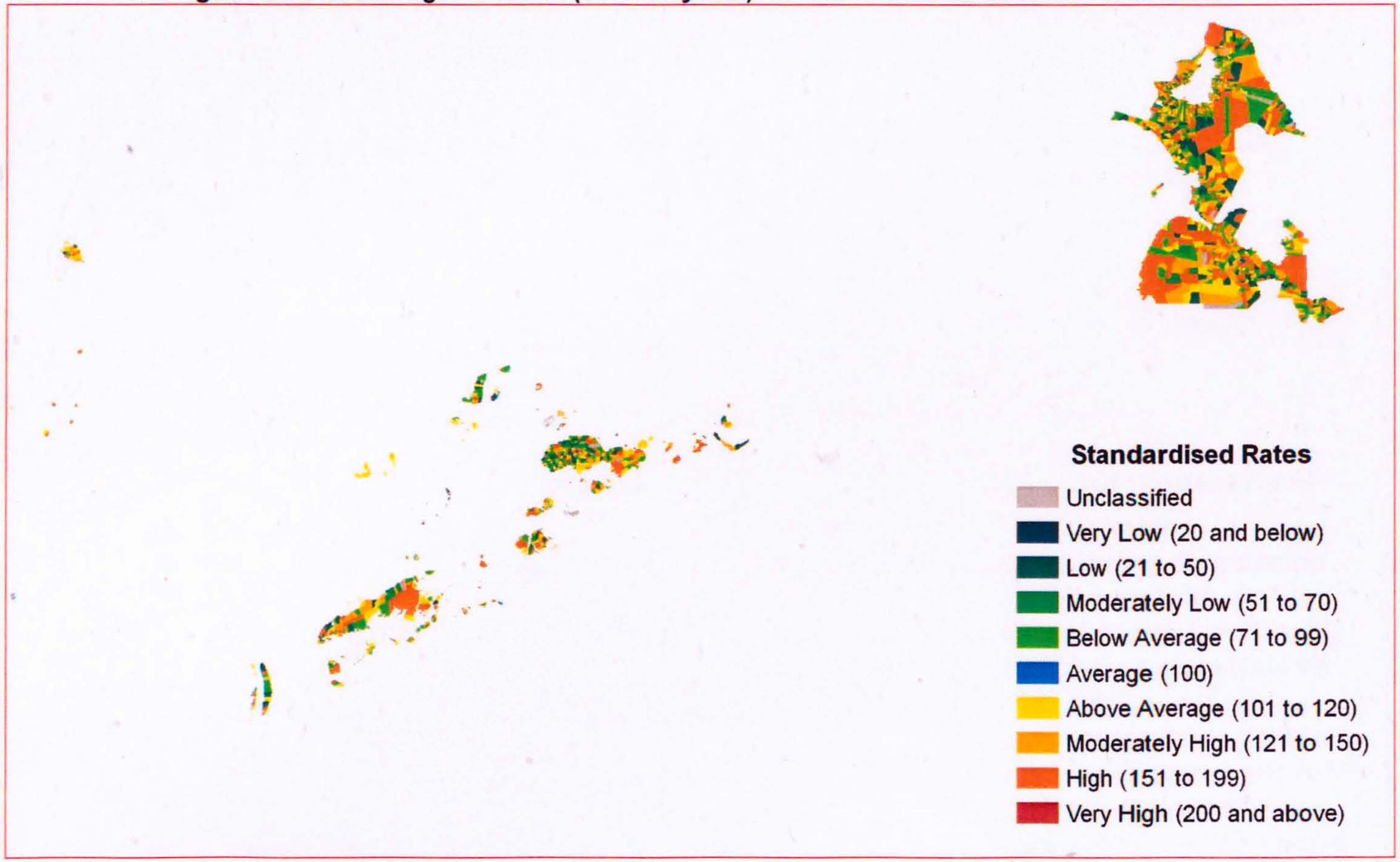
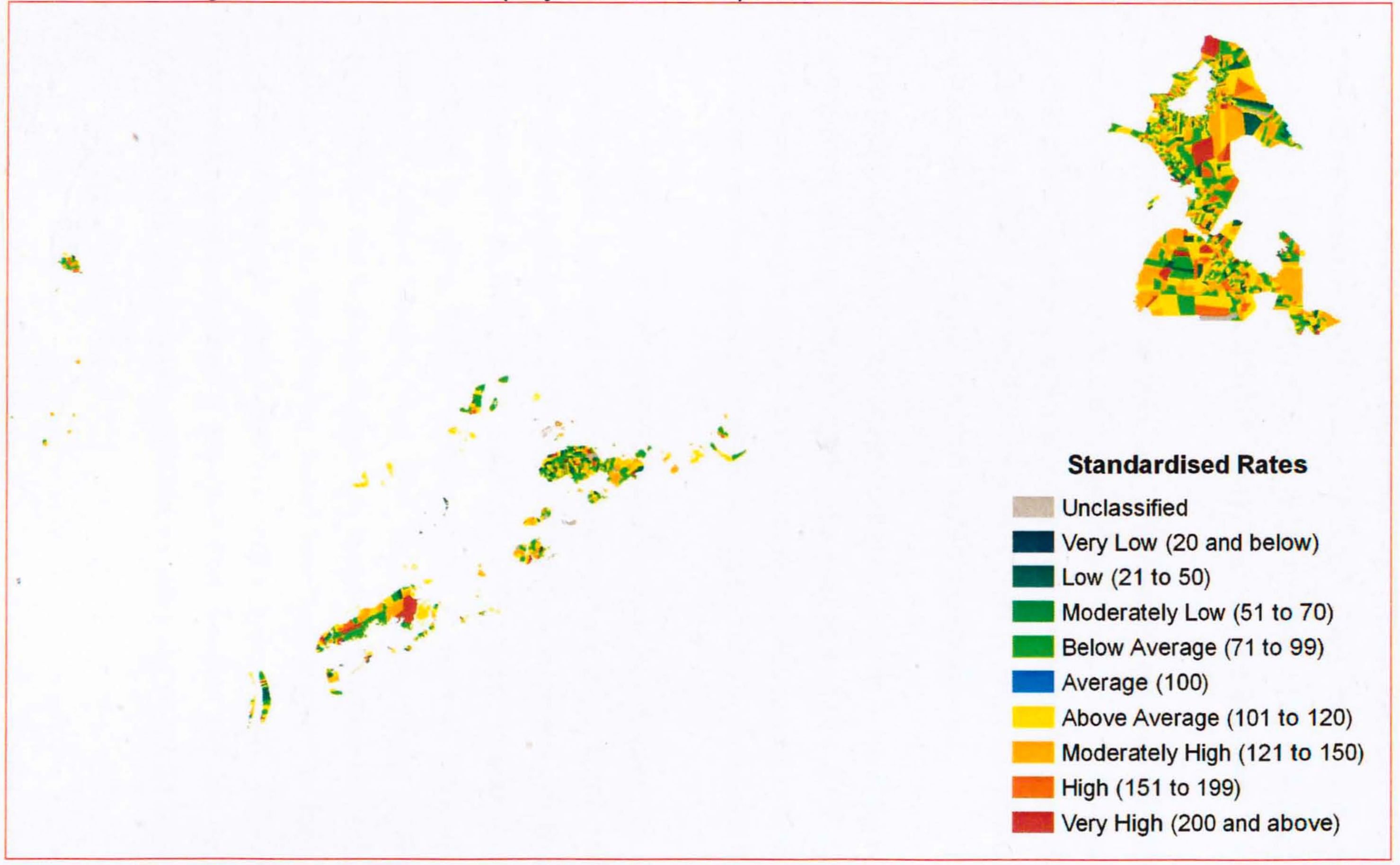


**Figure 8.14: Older Women (50 years and above) Without Formal Education in the ARRM**



## 8.4 An Insight into Maternal Related Mortality in Nigeria

All over the world, maternal related deaths are a cause for concern. The effect of the death of women due birth related complications can greatly affect the future of their children and in turn have an impact on the society at large. The greatest concern for these deaths however lies within the heart of the developing world where more than 99% of all maternal deaths occur (UNICEF, 2009). Worse still, about 84% of maternal deaths are concentrated within arguably the two least developed regions of the world (Sub-Saharan Africa and South Asia). This massive inequality is alarming.

Amongst all the MDG targets, the objective of reducing child and maternal mortality ratio in Nigeria presents the stiffest challenge to policy makers (NPC, 2007). On a global scale, Nigeria is only second to India in the ranking of maternal mortality rates (UNICEF, 2009). According to the most recent Nigerian MDG report, about one in thirteen women die from childbirth related complications.

The process of tracking maternal mortality in Nigeria is severely weakened by the weak institutional data gathering frameworks available (NPC, 2007; Okonjo-Iweala, 2007). This implies that the actual rate of maternal related deaths may be higher than the one in thirteen women quoted in the MDG report. This is a key cause for concern.

There continues to be growing evidence that the disparity in maternal mortality in Nigeria varies from rural to urban areas (Umeora et al., 2005; WHO, 2009d); from one region to another (NPC, 2007; WHO, 2009d); and indeed by direct and indirect causes (Lindroos, 2004; NPC, 2007; UNICEF, 2009) spear headed by haemorrhage which accounts for about 30% of complications. In spite of these evidences, the rate of maternal related deaths continues to increase. Many strategic decisions have accompanied the findings of previous research; however the lack of political will and a national policy on reproductive health has been blamed for the continued increase in the rate of maternal related mortality in the country (NPC, 2004a). This includes non-responsiveness on the part of the institutions saddled with the task of co-ordination and collating timely and accurate statistics on reproductive health among Nigerian women.

However, a more serious question remains unanswered. It is a known fact that maternal mortality is a multi-faceted problem without a solitary resolution (Lindroos, 2004). The great magnitude of the problem of maternal mortality in Nigeria does compare with the scarce resources that accrue to policy makers for tackling the problem. The implication of this is that intelligent mechanisms need to be deployed and adopted to ensure that these resources are targeted to derive the greatest benefit. In essence, there is need to better understand at more localised scales the patterns of the problems and issues associated with maternal health in Nigeria. Despite this seemingly vivid need, there has been no known published work understudying the problem even at the LGA scale let alone a more localised scale. Within sections 8.4.1 to 8.4.4, a geodemographic approach will be used to investigate inequalities in the receipt of prenatal care and the types of assistance received during child delivery with a view to modelling the risk of maternal mortality at LGA level.

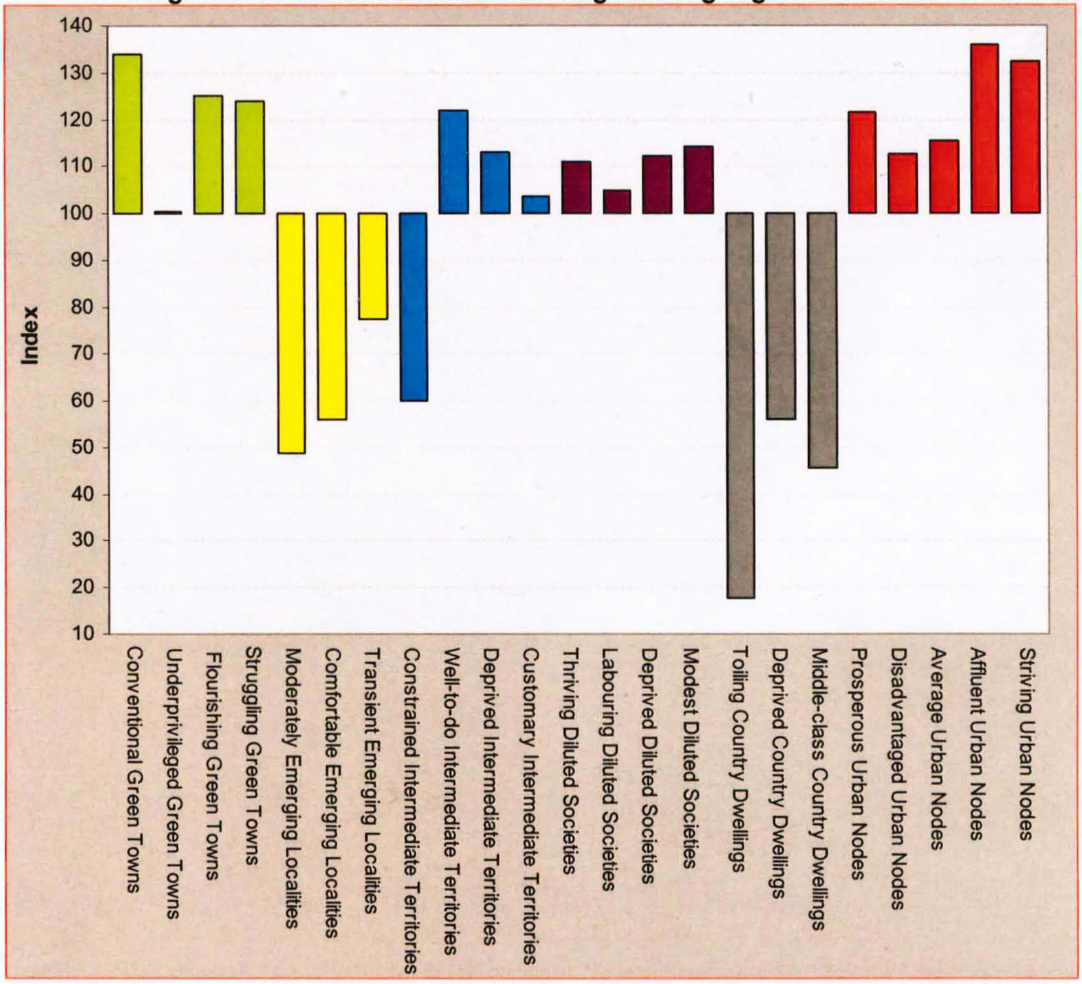
### 8.4.1 Patterning the Receipt of Antenatal Care

Antenatal care (ANC) refers to the screening of health and socioeconomic conditions likely to instigate adverse outcomes of pregnancy among women (WHO, 2009b). Effective antenatal care is therefore important for reducing the risk of complications during child birth.

In Nigeria about 15% of women in urban areas do not receive any form of ANC and about 46% of their rural counterparts fail to attend ANC services (NPC, 2007). These urban and rural differences are significant and expected in a country where there is massive disparity in education and well-being between urban and rural dwellers. What would be more significant for the purpose of targeting resources however are better picture of the dynamics at the LGA scale and the geodemographic correlates of the receipt or non-receipt of antenatal care among women.

Figure 8.15 shows indices for the penetration pattern of the attendance of women at ANC services in 2006. There is a suggestion of evidence of the urban-rural dichotomy reported in previous publications. However, it is pertinent to note that a significant level of rural characteristics (particularly the involvement in agro-allied employment) can also be found within Green Towns, Intermediate Territories and Diluted Societies yet there are some groups within these super-group clusters with above average levels of attendance at ANC. This explains why it is important to go beyond explaining the problem in terms of contiguous or higher levels of geographical regions but rather to examine inequality in terms of neighbourhood types.

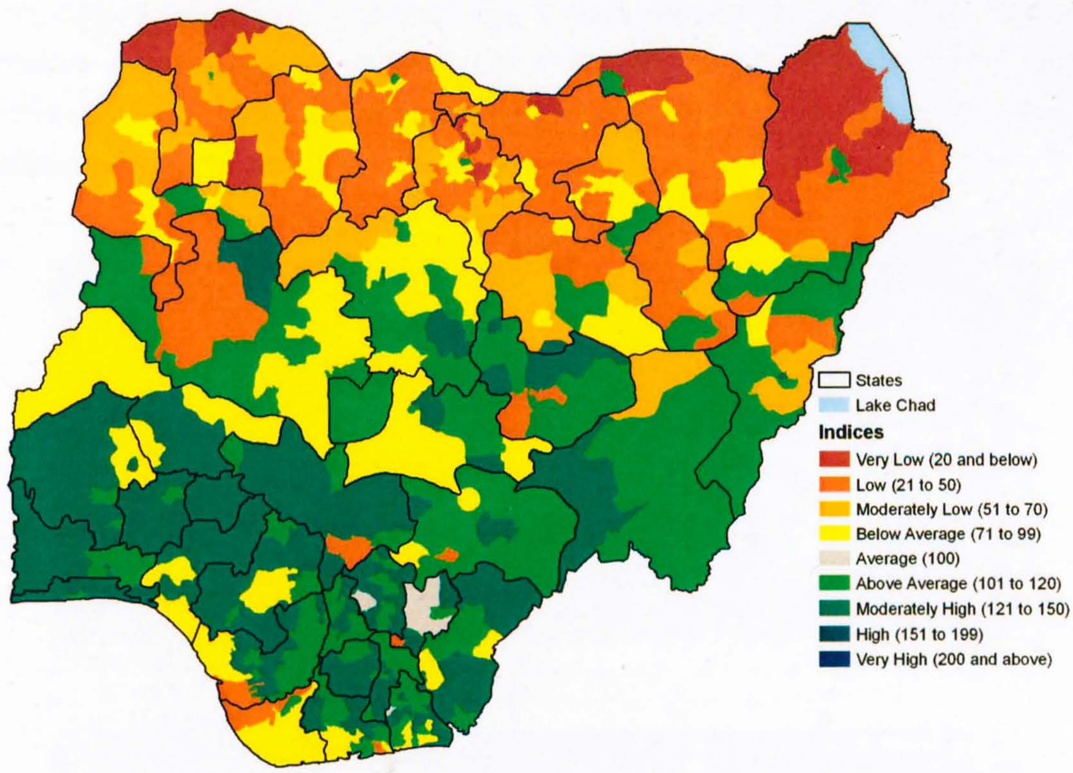
**Figure 8.15: Antenatal Care Patronage among Nigerian Women**




Women who are least likely to undertake ANC live in areas where literacy rates are very much below the national average and there is a large presence of uneducated household heads. Interestingly, there is an above average presence of people who are economically active particularly in agriculture.

Residents in these areas have below average access to health facilities while overall dissatisfaction is much higher. Several households spend more than one hour to their nearest health facility. The absence of drugs and cost are key factors contribution to dissatisfaction with health services. Community health centres are sparingly used by residents. However, the patronage of traditional healers is greatest amongst residents than for all other neighbourhood types in the country. What is also very striking about these areas is that vaccination rates for infants are significantly low.

**Figure 8.16: The Chance of a Woman Patronising ANC in Nigeria**




The national spatial pattern of the profile of attendance at ANC among women shown in Figure 8.16 shows the lowest penetration amongst numerous LGAs in the north and pockets of LGAs within the southern areas in states like Edo, Bayeslsa and Rivers.

## 8.4.2 Births Assisted by Skilled Personnel

Having investigated the penetration pattern of attendance at ANC, another maternal mortality risk factor that can be evaluated is the inequality in the types of workforce that help women during child birth. Women who receive assistance from people who are adequately trained to provide birth assistance services would be less exposed to the chance of morbidity from complications.

Who are those defined as skilled personnel? Within the context of the UN MDGs, skilled medical personnel who can assist during child delivery include doctors, nurses and midwives (UN, 2003). This also conforms to the definition of skilled personnel in Nigeria. Unfortunately only about one third of children are delivered by these professionals in Nigeria.

To evaluate inequalities in the types of assistance received by women during child birth, data for infants aged between 0 and 4 years was provided by the NBS. Table 8.2 provides a summary of the indices for all children delivered with the help of a trained professional and other columns showing the variations in likelihood of receiving assistance from specific types of professionals.

**Table 8.2: Indices for Deliveries Assisted by Skilled Personnel by Groups**

| Groups | Children Delivered by Skilled Professionals | Specific Assistance Received During Child Delivery | | |
|---|---|---|---|---|
| | | Children Delivered by Doctors | Children Delivered by Nurses | Children Delivered by Midwives |
| Conventional Green Towns | 179 | 177 | 196 | 130 |
| Underprivileged Green Towns | 105 | 63 | 112 | 115 |
| Flourishing Green Towns | 168 | 105 | 196 | 134 |
| Struggling Green Towns | 177 | 168 | 197 | 126 |
| Moderately Emerging Localities | 22 | 17 | 16 | 43 |
| Comfortable Emerging Localities | 26 | 30 | 24 | 25 |
| Transient Emerging Localities | 35 | 25 | 30 | 54 |
| Constrained Intermediate Territories | 52 | 9 | 70 | 31 |
| Well-to-do Intermediate Territories | 178 | 145 | 167 | 239 |
| Deprived Intermediate Territories | 133 | 54 | 149 | 148 |
| Customary Intermediate Territories | 147 | 230 | 76 | 290 |
| Thriving Diluted Societies | 95 | 65 | 89 | 134 |
| Labouring Diluted Societies | 78 | 47 | 89 | 70 |
| Deprived Diluted Societies | 61 | 27 | 65 | 75 |
| Modest Diluted Societies | 94 | 61 | 118 | 48 |
| Toiling Country Dwellings | 15 | 19 | 10 | 24 |
| Deprived Country Dwellings | 34 | 41 | 34 | 31 |
| Middle-class Country Dwellings | 19 | 13 | 14 | 37 |
| Prosperous Urban Nodes | 196 | 431 | 147 | 157 |
| Disadvantaged Urban Nodes | 150 | 157 | 147 | 151 |
| Average Urban Nodes | 95 | 76 | 106 | 78 |
| Affluent Urban Nodes | 199 | 306 | 185 | 158 |
| Striving Urban Nodes | 194 | 155 | 203 | 199 |

In general, women within Urban Nodes (excluding Average Urban Nodes), Green Towns and Intermediate Territories (excluding Constrained Intermediate Territories) are greatly disposed to patronising trained health personnel for child birth. It is pertinent to note also that these clusters also have the largest penetration rates of literates.

Of greater significance is the fact that neighbourhood types have different levels of indices for different medical professionals. For instance while Underprivileged Green Towns demonstrate a slightly above average level of patronage of trained professionals, women within these areas receive greater assistance from midwives and nurses and are greatly unlikely to secure assistance from a doctor during child birth.

Women who live within Prosperous and Affluent Urban Nodes are four and three times more likely to receive medical attention from a medical doctor during child birth while within all Emerging Localities and Country Dwellings, women have less than half the chance of the national mean to receive help from a doctor.

Of all neighbourhood types, women within Constrained Intermediate Territories are most unlikely to receive medical attention from a doctor. This is not surprising as these areas have quite a large number of people without access to health services and the overall dissatisfaction with services is significantly high. The group has a large representation of people who spend more than one hour to their nearest health facility. People are particularly dissatisfied with the lack of trained professionals and the fact that they often receive unsuccessful medical treatments. Many people consult traditional healers, pharmacists or chemists. Another great danger posed to women within these areas is that early marriages and teenage conception rates are quite high and it has been found that teenage mothers are in greater disposition to complications during child birth (Smith and Pell, 2001).

### 8.4.3 Births Assisted by Non-skilled Attendants

To further evaluate risk factors that may contribute to increased exposure to maternal mortality it is pertinent to also examine which women take up assistance from people who are not professionally accredited to help with child delivery. Non-skilled personnel include trained and untrained traditional birth attendants (UN, 2003).

While some women receive some form of assistance during child birth, there are those who do not get any form of help. Such women are much more in harms way should there be any complications. Amazingly, an alarming 17% of child deliveries are done without any form of assistance.

**Table 8.3: Indices for Deliveries Assisted by Non-skilled Personnel by Groups**

| Groups | Children Delivered by Non-skilled Personnel | Specific Assistance Received During Child Delivery | | |
|---|---|---|---|---|
| | | Children Delivered by Trained Traditional Birth Attendants | Children Delivered by Untrained Traditional Birth Attendants | Self Deliveries |
| Conventional Green Towns | 39 | 67 | 24 | 62 |
| Underprivileged Green Towns | 98 | 137 | 123 | 27 |
| Flourishing Green Towns | 46 | 104 | 46 | 25 |
| Struggling Green Towns | 41 | 52 | 42 | 34 |
| Moderately Emerging Localities | 166 | 146 | 178 | 145 |
| Comfortable Emerging Localities | 160 | 70 | 162 | 187 |
| Transient Emerging Localities | 152 | 99 | 133 | 213 |
| Constrained Intermediate Territories | 141 | 337 | 163 | 21 |
| Well-to-do Intermediate Territories | 42 | 107 | 45 | 11 |
| Deprived Intermediate Territories | 80 | 132 | 93 | 31 |
| Customary Intermediate Territories | 64 | 163 | 74 | 6 |
| Thriving Diluted Societies | 106 | 117 | 108 | 98 |
| Labouring Diluted Societies | 111 | 136 | 124 | 74 |
| Deprived Diluted Societies | 86 | 31 | 88 | 102 |
| Modest Diluted Societies | 101 | 74 | 100 | 113 |
| Toiling Country Dwellings | 173 | 113 | 168 | 206 |
| Deprived Country Dwellings | 122 | 365 | 93 | 105 |
| Middle-class Country Dwellings | 160 | 76 | 167 | 172 |
| Prosperous Urban Nodes | 25 | 87 | 14 | 30 |
| Disadvantaged Urban Nodes | 65 | 73 | 67 | 57 |
| Average Urban Nodes | 105 | 109 | 75 | 172 |
| Affluent Urban Nodes | 25 | 55 | 19 | 29 |
| Striving Urban Nodes | 26 | 54 | 31 | 7 |

Table 8.3 shows the penetration rates of the use of non-skilled personnel among women by geodemographic types. Although there is generally a below average chance of using an unskilled person within these areas, Underprivileged Green Towns show above average levels of use of trained and untrained traditional birth attendants and surprisingly women within the Flourishing Green Towns also record a slightly above average incidence in their use of trained traditional birth attendants.

Women within Emerging Localities generally receive help from unskilled personnel but also have significantly high tendencies to undertake self deliveries. A similar trend is observed within Country Dwellings

Amongst Intermediate Territories, the greatest use of unskilled personnel is demonstrated by women within Constrained Intermediate Territories where quite a large number of people find it difficult to meet their basic daily needs and a substantial concentration of households find it difficult paying for their health care.

Surprisingly, the least disposition of women towards unskilled medical personnel within Diluted Societies is not within Deprived Diluted Societies. One of the factors that may account for this is the presence of an above average representation of unmarried people within these areas which may mean conception rates are lower. However within these areas, the level at which women engage in self deliveries is greater than the national mean.

As expected, women within Urban Nodes present low indices for the patronage of unskilled medical personal. The only exception to this is the Average Urban Nodes which has the greatest representation of young children amongst all Urban Nodes and a strong presence of people in polygamous marriages. An above average number of children within these areas are delivered at home and self deliveries are very common. Additionally infants within these areas have a less than average chance of receiving vaccinations.

If we also consider the pattern presented within Striving Urban Nodes, we notice that the incidence of patronage of unskilled personnel is low particularly for women who will undertake self deliveries. A major contributory factor to this is the demographic representation of people within these areas. Amongst all groups within Urban Nodes, Striving Urban Nodes have the greatest representation of older population over 60 years. They also have an above average representation of middle aged people and a substantial number of pensioner households. There are high numbers of single parents and a representative number of separated couples. There is also a rather large representation of widows. Indeed this is the only group within Urban Nodes with a higher than average incidence of widowed persons.

### 8.4.4 Inequality and Maternal Mortality Risk

In addition to the analyses presented in the earlier sections and to further uncover the pattern and level of inequality in the types of assistance received by neighbourhood types, a statistic called the index of dissimilarity (ID) was computed using equation 7.3 for each of the types of assistance received by women. The index is a standard measure of social and spatial segregation often used to measure how isolated or integrated people groups or social patterns are within geographical space (Rodrigue et al., 2009).

$$\mathbf{ID} = 0.5 \sum_{i=1}^{N} \left| \mathbf{X}_i - \mathbf{Y}_i \right| \qquad (8.1)$$

where:

**ID** is the Index of Dissimilarity

N is the number of geodemographic clusters

$\mathbf{X}_i$ is the target population group for cluster i

$\mathbf{Y}_i$ is the base population group for cluster i

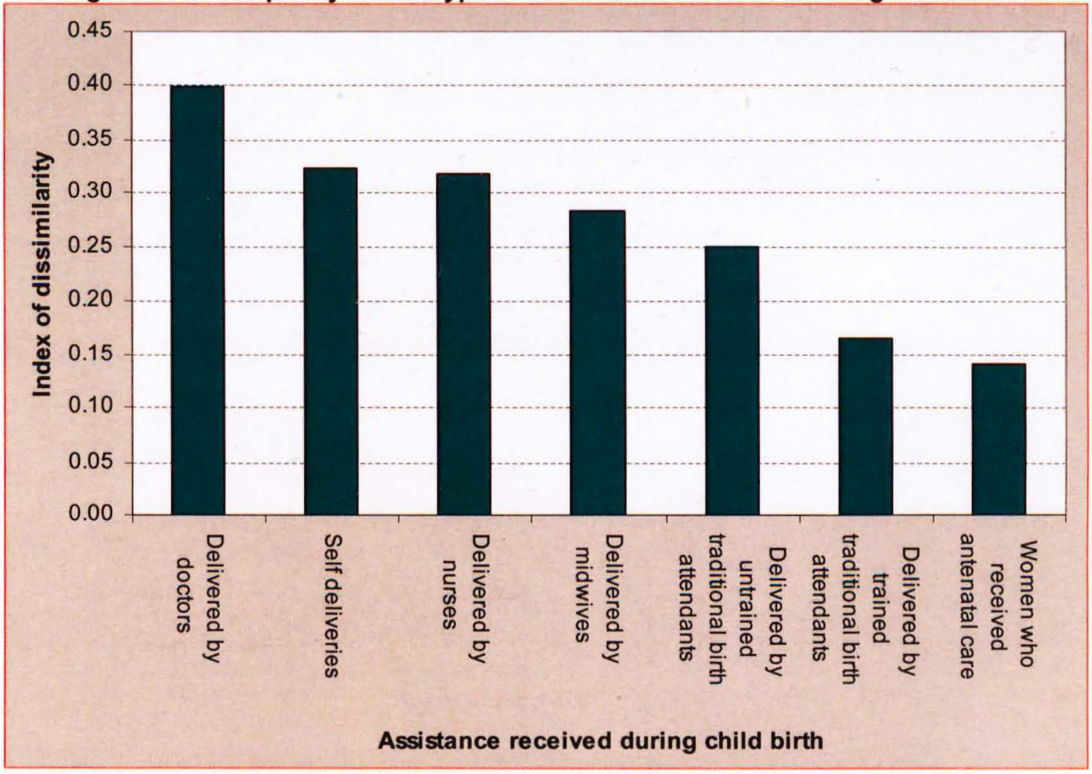
$i$ =1 indicates the start value for the summed index

Where the index is expressed as a percentage, it ranges from 0 to 100. Where it is not expressed as a percentage, the values are between 0 and 1. A value of 0 indicates that there is no segregation while a value of 1 or 100 means maximum inequality between areas.

In this research, a range of values between 0 and 1 have been adopted for the index. The ID for mothers who receive assistance from skilled personnel is 0.31 while for those who are catered for by unskilled personnel have an ID of 0.24. The finding implies that there is greater unevenness in the distribution of mothers who receive medical assistance from skilled people.

In spite of the relatively low level of inequality among women who receive some form of antenatal care, significant differences exist in the types of assistance received by women during child birth.

**Figure 8.17: Inequality in the Type of Assistance Received During Child Birth**




In Figure 8.17, the magnitude of inequality existing between various types of assistance received by women is displayed. At the extreme end of the inequality scale are the best and worst possible forms of assistance. This hints that women in these two categories are highly concentrated within their locality types. For instance when we consider the pattern of inequality among women who receive assistance from doctors, almost half of the country's potential mothers will have to relocate from their residences to other areas for a state of national equilibrium to be attained.

The trend of the other types of assistance shown in the chart establishes the position that the better the form of medical assistance received, the greater the inequality among women.

The standard way of measuring the risk of maternal mortality is to divide the number of maternal births by the total number of live births and multiply by 100, 000 (UN, 2003). Securing accurate data from vital registrations on medically certified deaths and their causes is therefore important for determining the level of exposure to maternal related deaths. As this information was not supplied, a different method was used to estimate the risk of mortality among women in Nigeria especially as it relates to birth complications.

The severity of exposure to maternal mortality was derived by creating a composite score from the probabilities of each of the risk hazards (access to prenatal care and types of assistance received during child birth). Each risk hazard was standardised between a range of 0 and 1 such that 1 represents the highest risk and 0 the lowest risk. The composite risk score is derived from the summation of the standardised values of each hazard.

**Figure 8.18: Exposure to Risk of Maternal Mortality in Nigeria**

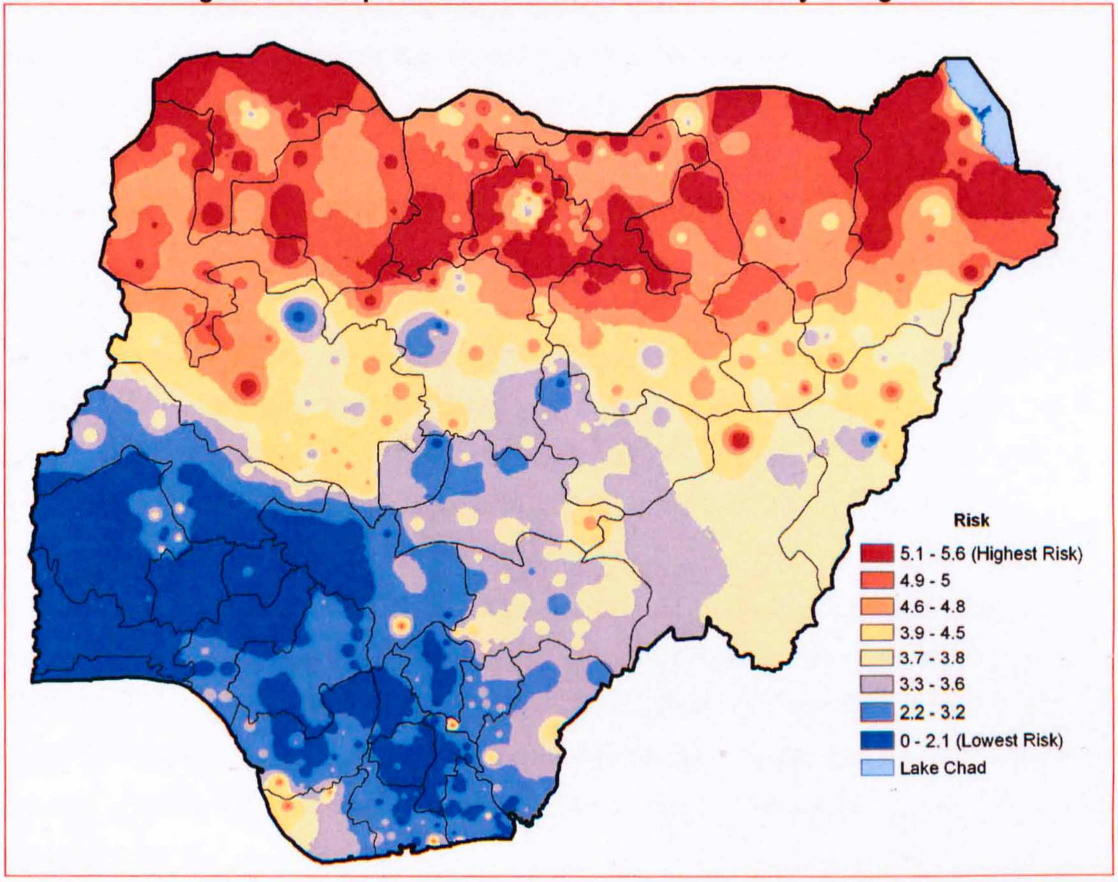



Figure 8.18 shows the mapped result of the risk model at LGA level. Evidence from the map reveals that when we consider the probability of women accessing prenatal care and the types of assistance they receive during child birth, a lot more women in the northern part of Nigeria are exposed to the risk of maternal deaths. States like Sokoto, Zamfara, Katsina, Kano, Jigawa and Borno reflect relatively high concentration of risk.

In the Abuja, the capital city, the risk of exposure ranges from medium to low. Most areas within the middle belt also have relatively average levels of exposure to risk.

The southern states in general are the least problematic although pockets of medium to high levels of risk can be found in some of these areas particularly in the south east and south south.

## 8.5 Child and Maternal Health Care in the Philippines

In the most recent Philippines MDG report published in 2007, there is an acknowledgement that unless there is a stepping up of efforts and a possibly a change in strategy, the chance of meeting the 2015 target of 52 deaths per 100,000 live births in maternal mortality ratio (MMR) may not be met.

In 2006, the MMR was put at 162 deaths per 100,000 live births (NEDA and UNDP, 2007). There appears to have been a steady national decline from 209 deaths per 100,000 in 1993. However in the recent report published by the WHO, the MMR was put at 230 deaths per 100, 000 live births (WHO, 2009d). The current rate is significantly high; much higher than the regional average for the WHO Western Pacific Region (WPR) which is currently 82 deaths per 100,000 live births. This increased rate calls for a lot of momentum to improve preventive targeting mechanisms in the country.

Apart from the depth of maternal deaths, under-5 mortality is another source of cause for alarm in the country. It is estimated that there are about 28 deaths among children less than 5 years for every 1000 live births in the country (WHO, 2009d). This figure is above the regional average of 22 deaths per 1000 live births for the WPR.

What is perhaps more surprising about the rate of under-5 mortality is the fact that many of the key causes of death are actually preventable. Using data from 1981 to 1991 for Cebu province, Wagstaff (2001) found that not only was under-5 mortality most prevalent amongst the poorest segments of the society, but that 52% of deaths among children resulted from communicable diseases like measles, diarrhoea, fever and pneumonia.

Quite a substantial number of communicable diseases affecting children can be mitigated by ensuring that infants receive adequate doses of vaccination to ensure that they develop potent immunity against such ailments (Salisbury et al., 2006).

In the next two sections of this chapter, healthcare amongst women and children will be evaluated to help understand the patterns of inequality. Section 8.5.1 investigates inequity in the patterns of vaccination among children while 8.5.2 focuses on the issue of pre and postnatal care for women.

### 8.5.1 The Vaccination of Children

In section 3.2.4 of the third chapter, it was established that the rate at which young children under the age of five die in developing world countries is up thirteen times that of their developed world counterparts (UNDP, 2007). When children are born, it takes a while for their immune systems to develop. Indeed it has been suggested that at birth, the immune system of and infant is immature (IMAC, 2006). Vaccinations are important for inducing immunity in infants against pathogenic diseases.

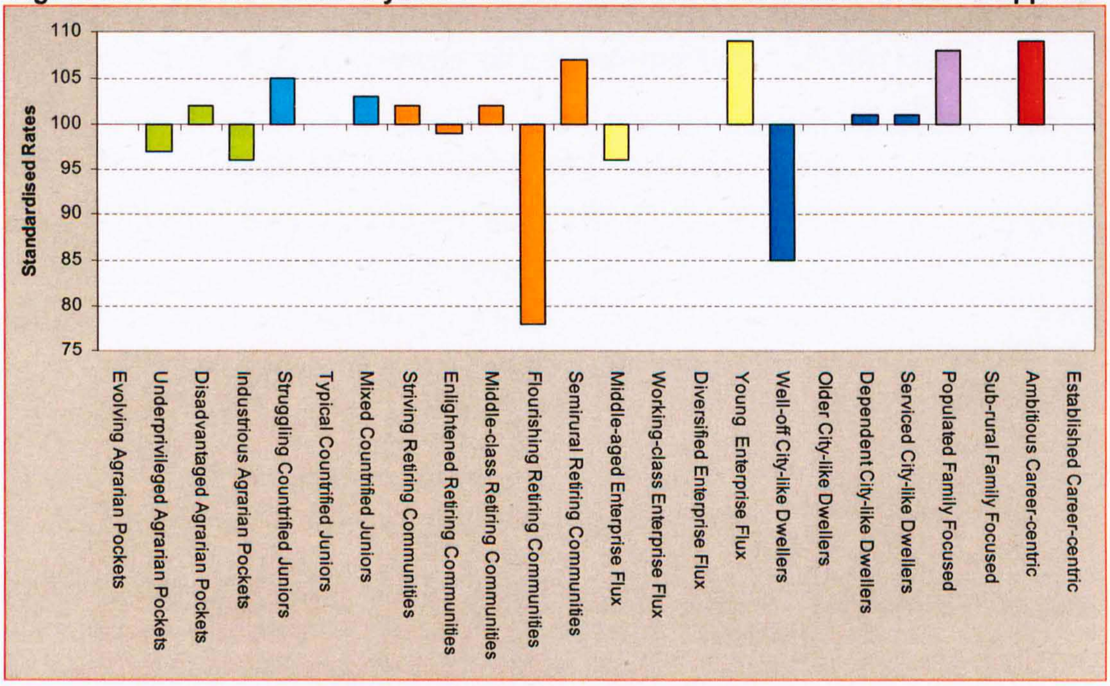
Initial analysis of the 2003 DHS at the regional level revealed that the immunization of infants vary by region with about 81% in Western Visayas to 44 percent in the Administrative Region in Muslim Mindanao (NSO and ORC Macro, 2004). However regional level insight does not necessarily provide a means to target resources and campaigns intelligently at the local level. By plugging the data into the Philippines geodemographic system, not only are we able to identify any geodemographic correlates associated with immunisation or non immunisation among children, the results can also be disaggregated locally at barangay scale for targeting purposes.

**Table 8.4: Profile of Children under 5 years who did not Receive Vaccination**

| Super-groups | Population Share (%) | Rate (per 100 persons) | Standardised Rates |
|---|---|---|---|
| Agrarian Pockets | 19.75 | 69.20 | 98 |
| Countrified Juniors | 12.15 | 71.38 | 101 |
| Retiring Communities | 12.82 | 70.28 | 99 |
| Enterprise Flux | 19.88 | 70.12 | 99 |
| City-like Dwellers | 19.69 | 68.34 | 97 |
| Family Focused | 5.36 | 73.07 | 103 |
| Career-centric | 10.35 | 72.76 | 103 |

Table 8.4 reveals a geodemographic profile at super-group level of children who did not receive any form of vaccination. What is noticeable is the fact that the standardised rates for the seven neighbourhood types are very closely related to the national mean of 100. At this level of the hierarchy, both Family Focused and Career-centric neighbourhoods have the least population shares. However, they also jointly have the highest national rates of non-vaccination among infants.

**Figure 8.19: Children under 5 years who did not Receive Vaccination in the Philippines**




Evidence from the analysis presented in Figure 8.19 reveals that children who do not receive vaccination are not only mostly concentrated in neighbourhoods dominated by young children (Young Enterprise Flux and Populated Family Focused) but also disadvantaged rural communities like Semirural Retiring Communities and Struggling Countrified Juniors.

What is perhaps surprising is the rather high rate of non-vaccination experienced by children residing in localities classed as Ambitious Career-centric. These are neighbourhoods characterised by vey many unmarried persons. In terms of age distribution, there is a great concentration of middle-aged persons between 21 and 44 years. The presence of all other age groupings is less than the national mean. There are also very many households which are headed by females.

While a representative proportion of residents are professionals, many are at the beginning of their careers. This is possibly one of the factors that have contributed to the high rates of non-vaccination of children in these areas. While parents and guardians of these children may be well informed on the significance and implications of not attending immunisation, they probably put the development of their careers ahead of their children or sometimes do not remember appointments.

In general, school information systems, radio and television jingles are most appropriate for communicating the importance of vaccination to residents of the neighbourhoods with above average rates of non-vaccination of children.

### 8.5.2 The Receipt of Prenatal and Postnatal Care

In the 2007 Philippines MDG report, there was an observation that there has been a slowing down of the decline in the number of maternal deaths per 100, 000 live births. The implication of this is that the 2015 target of 52 deaths in the maternal mortality ratio (MMR) remains threatened (NSO and ORC Macro, 2004).

Access of women to adequate health care before and after child birth is very important for ameliorating risk of complications related to child birth. The recommendation from the Philippines Department of Health is that all women should be exposed to at least four visits of antenatal care. However, three in 10 women do not receive this recommended care (NSO and ORC Macro, 2004).

As information on the receipt of antenatal or postnatal care was not included as variables for the development of the segmentation system, it is possible to geodemographically profile and evaluate the likelihood that women in particular neighbourhoods will receive prenatal and postnatal care in the Philippines. The resulting information can then be mapped at barangay level for the purpose of targeting areas where women are less susceptible to health care before and after child birth.

The analysis is done at the group level. First rates of non attendance for pre and post natal care are computed for all women aged between 15 and 49 years and then age specific rates are analysed for younger women (aged 15 to 29 years) and middle aged women (aged 30 to 49 years). The rates have been standardised about a national mean value of 100 to determine which neighbourhood types have women who have above average tendencies not to receive antenatal and post natal health care. Table 8.5 shows results from the analysis.

**Table 8.5: The Receipt of Antenatal and Postnatal Care among Women in Philippines**

| | All women aged 15 to 49 | Younger women aged 15 to 29 | Middle aged women aged 30 to 49 |
|---|---|---|---|
| **Index of Dissimilarity** | **0.013** | **0.012** | **0.010** |
| | | | |
| Evolving Agrarian Pockets | 96 | 99 | 96 |
| Underprivileged Agrarian Pockets | 100 | 97 | 100 |
| Disadvantaged Agrarian Pockets | 98 | 98 | 97 |
| Industrious Agrarian Pockets | 94 | 97 | 99 |
| Struggling Countrified Juniors | 106 | 104 | 104 |
| Typical Countrified Juniors | 98 | 98 | 99 |
| Mixed Countrified Juniors | 98 | 106 | 104 |
| Striving Retiring Communities | 100 | 105 | 100 |
| Enlightened Retiring Communities | 100 | 100 | 97 |
| Middle-class Retiring Communities | 95 | 95 | 101 |
| Flourishing Retiring Communities | 101 | 103 | 98 |
| Semirural Retiring Communities | 104 | 99 | 100 |
| Middle-aged Enterprise Flux | 100 | 103 | 103 |
| Working-class Enterprise Flux | 96 | 96 | 97 |
| Diversified Enterprise Flux | 98 | 102 | 102 |
| Young Enterprise Flux | 102 | 102 | 104 |
| Well-off City-like Dwellers | 105 | 100 | 101 |
| Older City-like Dwellers | 100 | 101 | 101 |
| Dependent City-like Dwellers | 103 | 100 | 103 |
| Serviced City-like Dwellers | 105 | 102 | 101 |
| Populated Family Focused | 106 | 99 | 100 |
| Sub-rural Family Focused | 98 | 100 | 95 |
| Ambitious Career-centric | 100 | 98 | 99 |
| Established Career-centric | 96 | 95 | 98 |

In addition to the standardised rates shown in the Table above, a measure of inequality across the three variables (all women aged 15 to 49, women aged 15 to 29 and women aged 30 to 49) has also been computed. The index of dissimilarity which was described in section 8.4.4 was adopted as the statistic for measuring inequality. In this research, the index can assume a value between a range of 0 and 1. A value of 0 indicates that there is no segregation while a value of 1 means maximum inequality between areas.

Although overall inequality among all women (0.013) is just marginally greater when compared to the age-specific statistics, results from the analysis reveal that in general across the 24 groups, inequality is quite low. Indeed, it is much lower than the Nigerian case.

The greatest concentration of the target group when all women are combined can be found in Struggling Countrified Juniors where there is a dominant presence of young children and a large proportion of residents have less than primary education; and Populated Family Focused which also have similar characteristics. In both neighbourhood types, health centres are commonly used to provide health care.

Younger women who do not receive any form of antennal and post natal care are more concentrated than their middle aged counterparts. Contributing the most the concentration are women within Mixed Countrified Juniors where female headed households are much more common than male headed ones; and women within Striving Retiring Communities where there is a very strong presence of unmarried persons and a very large presence of widows.

Amongst middle aged women, both Struggling and Mixed Countrified Juniors again top the list of neighbourhood types where women fail to attend pre-natal and post natal care. Young Enterprise Flux also has a significant presence of this characteristic. Incomplete primary and secondary school attainments within these areas are just above average. Additionally, this is the only group where the general use of all different health service providers is below the national mean.

## 8.6 Conclusions

In chapters five and six, a thorough explanation of the decisions made in the process of creating the segmentation systems of Nigeria and the Philippines was given.

The analyses and discussions presented in this chapter not only serve as an opportunity for evaluating local level variations of several indicators linked with gender parity in both countries, but also as a framework for testing the efficiency and discriminatory power of the classification systems.

The variables profiled against the classification systems were not included in the creation of the geodemographic systems. The analyses demonstrates how the system can be used to account for differences in the features of different women and children and different areas by splitting nationally sought survey datasets against the clusters.

What has also been established in this chapter is that *one size fits all* policy approaches need to be reviewed as people groupings in both countries mean that advantages and disadvantages with respect to local level residential neighbourhoods vary by those neighbourhoods.

In Nigeria women are generally more likely to partake in the contribution to household income than they are likely to own landed property. However, stark inequalities do exist in the pattern of contribution to income. The chance of a woman contributing to income within the home is greater for women who live in areas where widows are disproportionately represented, single parents and separated couples dominate and households are not of large sizes usually of 1 to 2 people.

The issue of gender parity in educational attainment among women in the Philippines was addressed in section 8.3. Significant variations exist by neighbourhood type. Not only do illiterate women generally concentrate in certain neighbourhood types but age-specific variations also exist at the neighbourhood level. As has been explained, these variations should be taken into consideration when targeting policy initiatives to ensure efficiency.

The risk of maternal mortality in Nigeria is multi-faceted and shown in the analysis presented in section 8.4.4. Women vary in their susceptibility to different risk factors. However a near north-south divide exists in the risk of exposure to the problem.

Apart from deprived communities with large concentrations of young children, efforts at strengthening infant vaccination in the Philippines should also focus on neighbourhoods with young parents who are particularly trying to develop their careers. In addition to strategically targeting pro-poor policies, social marketing campaigns and dissemination of information can be further aided by using the geodemographic system. There are significant proportional variations in the susceptibility of different neighbourhood types to receive information via various sources. Therefore deploying the communication strategy via the most appropriate means will also help improve the chances of accepting the information. This can be particularly helpful when trying to raise awareness on the importance of attending pre-natal and post-natal care. Educated women living in career-centric neighbourhoods are less likely to respond to focus groups than their counterparts in more rural settings.

The analysis and reporting of indicators at the national level often create the notion that situations within the country are homogeneous. For too long data analysis and reporting have been restricted to higher levels of administrative geography in Nigeria and the Philippines (NSO and ORC Macro 2004; NPC and ORC Macro, 2004; NPC, 2007; NEDA and UNDP, 2007). This makes it difficult to uncover stark differences that exist at the local scale of geography in order to target policy.

The geodemographic analysis of Nigerian and Philippines survey data on maternal health care presented in this chapter has helped reveal that significant variations do exist at the local scale of geography in both countries. In order to helps us summarise how effective and beneficial the analysis presented in this chapter is for policy makers in Nigeria and the Philippines , an extract from the work of Deichmann (1999) on the importance of identifying target groups at the local level of residential geography is provided. According to him:

'*In designing intervention schemes and allocating subsidies, resources will be used more effectively, if the most needy groups can be reached effectively. This reduces the leakage of transfer payments to non-poor persons (Type I error), and it minimizes the risk that a poor person will be missed by a poverty alleviation program (Type II error). Detailed information on the location of target groups is thus often a key ingredient for effective projects*' (Deichmann, 1999 p. 3).

The datasets profiled in this chapter are very important when making national policy decisions for Nigeria and the Philippines (NBS, 2006a; NSO and ORC Macro 2004). However prior to the research reported in this thesis, these datasets had not been analysed to reveal inequalities at the LGA scale for Nigerian and at the Barangay scale for the Philippines. The findings of the analyses show that with geodemographics, intelligent local level targeting can be done. This has helped address the issues raised by Deichmann.

In addition to identifying priority areas within both countries, some of the geographic and demographic factors that influence maternal health care and the empowerment of women have also been revealed. This makes it possible to differentiate policy propositions by area types thereby resolving the problem of one-size fits all approaches.

# 9

# Conclusions: Breaking the Fallow Ground

## 9.1 Introduction

This chapter summarises the findings of the thesis. The summary of research findings are presented in section 9.2 by summarising each of the previous eight chapters. In section 9.3, the key contributions to knowledge are highlighted. Section 9.4 provides some of the lessons learnt in the research exercise. In section 9.5, some of the challenges and limitations of the study are presented while section 9.6 provides a review of potential for future work. A final concluding statement is provided in section 9.7.

## 9.2 Summary of Research Findings

Each of the objectives outlined in Table 1.1 of the first chapter were aimed towards achieving the overall aim of the thesis which was to create relatively small area general purpose geodemographic classification systems for Nigeria and the Philippines and further illustrate the usefulness and potential of the classifications.

In order to achieve the big picture, the first objective was to review some of the historical events that surrounded the emergence of modern day geodemographics. In the second chapter, the concept of geodemographics was defined with examples. Additionally, evolutionary trends of social area classifications were reviewed by addressing developments before the 1980's and beyond this time. It was established that although geodemographic ideas emerged from the research arena, their expansion and use have to a large extent been appreciated in the commercial world.

Chapter two also helped address the first objective of the thesis which was to investigate the uses and proliferation of segmentation systems across the globe. The chapter reviewed current application areas within the public sector, private sector and academia. Qualitative evaluations discussed within this chapter showed that there is paucity in the development and application of geodemographic techniques within developing countries.

Multiple factors contribute to the paucity in the proliferation of geodemographic systems within developing countries. Most of these factors centre on the availability of the required datasets. Even in scenarios where the datasets are available, access to them is often restricted by the authorities that have collected them (Bishop et al., 2000). In many cases, the legal frameworks used to protect the datasets are outdated.

In addition to the problem of data availability and access, technical competence has also contributed to the inhibition of geodemographic developments in these countries (Bishop et al., 2000). Many government departments often lack the required level of technical expertise to embark on advanced spatial analysis.

Some of the modern day challenges experienced in developing countries were also addressed as part of the first objective. The framework of the United Nations (UN) Millennium Development Goals (MDGs) was used to discuss these problems. Chapter three introduced the eight MDGs and critically reviewed their strengths and weaknesses. Specific shortcomings of the MDGs addressed include the focus on a vague concept of development rather than within-country inequalities. This often leads to the adoption of one-size fits all approaches by member countries. Data collection and analytical techniques also make inter-country comparisons inadequate with the MDGs. One of the most serious defects of the MDGs is that the UN approach often restricts spatial analysts to conducting their analysis at higher levels of geographical aggregation where within-country inequalities are hidden (Dorling and Ballas, 2008).

Having addressed the general challenges of developing countries at the beginning of the third chapter, a comprehensive discussion of the benefits of geodemographic options was provided at the latter part of the chapter.

It was established that geodemographics can help developing countries strengthen their concepts of identity by proving to them that not only do people who reside within the same localities have similar needs (Tobler, 1970), their needs and characteristics may just be similar to those of people within distant neighbourhoods (Vickers and Rees, 2007).

The potential of geodemographics at addressing sub national policy problems was also discussed in the third chapter. It was established that geodemographics can help developing countries better understand the inequalities existing within residential population groups at the local level if they are created at these scales.

Geodemographics also offers developing countries the option of better planning, targeting of resources, targeting of interventions and monitoring the impacts of national policies at the local level. In addition to these policy related benefits, the academic potential of open geodemographics for developing countries was also addressed in the final section of chapter three.

Having established that geodemographic segmentation systems offer developing countries multiple benefits, a review of some methods of multivariate and cluster analysis was provided in chapter four.

Part of objective two was addressed in Chapter 5. The chapter provided a detailed discussion of how the first geodemographic classification system was developed for Nigeria at Local Government Area (LGA) scale. The chapter provided a comprehensive review of the variables that were included in the classification system and explanations of how they were selected. A total of forty-five variables were selected for inclusion in the development of the Nigerian classification system. The processes that followed the selection of variables and the deployment of a k-means clustering method resulted in the creation of 6 cluster groups at the top hierarchy of the classification. The top-down approach led to the creation of another 23 clusters at the second hierarchy and a final group of 57 clusters at the third hierarchy.

The knowledge gained from Chapter 5 proved useful for Chapter 6 which also addressed the second objective. The chapter presented a discussion of the decisions made during the creation of the classification of the Barangays of Philippines. The Philippines analysis presented new challenges as categorical and continuous datasets were combined to create the classification system. A different clustering technique – the TwoStep clustering procedure was used for the Philippines analysis which resulted in the creation of 7 clusters at the top hierarchy, 24 clusters at the second hierarchy and 66 clusters at the third hierarchy.

After creating the classification systems for both countries, value was added to them by providing profiles, visualisations and comprehensive user guides contained in the compact disc accompanying this thesis. This helped resolve objective three.

In Chapters 7 and 8, the two classification systems were put to test. Chapter 7 focused on addressing problems related to primary level education in both Nigeria and the Philippines. The analyses and discussions helped show how geodemographics can be used to address educational policy issues and deploy scarce resources.

Chapter 8 addressed the problem of gender parity in both countries. For Nigeria, variables relating to landed property and the contribution to household income were profiled while in the case of the Philippines age-specific differences in educational attainment among women were discussed.

Another important topic that was addressed in Chapter 8 is maternal and child related health care. Geodemographics was used to evaluate the risk factors relating to child delivery in Nigeria while for the Philippines it was shown that local level differences exist by neighbourhood types in the rates at which children are vaccinated.

## 9.3 Contributions to Knowledge

This section highlights some of the areas in which this research work contributes to knowledge within the ambit of spatial analysis.

***Geodemographics for the Developing World***

In the second and third chapters, it was shown that there is a paucity of geodemographic classification systems in developing countries. This research work builds on the work of Vickers (2006) and presents the first development, evaluation and application of open-source geodemographic systems for developing countries. The systems discussed here are the first for Africa and Asia.

***Segmentation Analysis for the Millennium Development Goals (MDGs)***

This is also the first time that the MDGs have been assessed using geodemographic spatial analytics. This form of assessment presents an opportunity to radically change the way of analysing, profiling and reporting progress on the MDGs. Segmentation analysis presents an important option for decision makers who wish to understand the anatomy of inequalities in progress towards these MDGs at local scales of geography. It presents policy makers within this arena with an opportunity to understand and monitor social and spatial change within the framework of the MDGs for small areas.

***Methodological Contributions to knowledge***

In chapter 6, some methodological contributions to knowledge are presented. Virtually all geodemographic systems in developed countries (summarised in Table 2.1 on page 16 of chapter 2) encapsulate continuous datasets. Data challenges presented in the Philippines required the combination of categorical and continuous datasets. This research work provides the first set of spatial analyses leading to the creation of a geodemographic classification system from categorical and continuous data.

One of the important findings in the course of the analysis reported in section 6.7.4 of Chapter 6 is that when dealing with mixed variables, range standrdisation can present problems for the clustering algorithm. Indeed Bacher et al. (2004) seemed to suggest this when they wrote as follows:

'*Simulation studies suggest, that z-standardization is ineffective (for a summary of simulation results; see Everitt et al. 2001: 51). Better results are reported for standardization to unit range (ibidem). However, standardization to unit range is problematic for mixed type attribute*' (Bacher et al., 2004, p. 5).

The research presented in the sixth chapter has tested the hypothesis captured in the statement above using spatial datasets and found it to hold.

## 9.4 Lessons Learnt from the Study

In the course of conducting this research and writing this thesis, multiple lessons have been learnt. These lessons are further expanded in this section.

***Challenges and Benefits of Having Two Classifications***

The research has been conducted across two case study regions - Nigeria and the Philippines. Doing this has presented both challenges and benefits. While both countries are relatively comparable in population size and overall levels of economic development, differences exist between them in terms of social, cultural and governing systems. It was therefore challenging trying to balance knowledge on these issues for both countries when interpreting the classification systems and writing the thesis.

A key benefit of working with both developing countries however was in the transfer of knowledge both from sourcing data and in the creation of the classification systems. Data was first sourced from the Nigerian suppliers because the thesis was initially designed to focus on Nigeria alone. The Nigerian agencies did not respond to letters and emails until contact was made with them over the phone.

The process of generating data sharing agreements was also slightly cumbersome and it took about to eight months after the first initial contact before the data was eventually released.

The experience gained from engaging with the Nigerian officials paid off when the decision was made to include the Philippines as another case study region. Direct contact was made with them over the phone to demonstrate the seriousness and importance of the research before further contacts were made by email. The Philippines datasets were released about three months after the first communication with the National Statistics Office (NSO) in Manila.

***Need for Clarity of Purpose***

The purpose of any classification system is very important and should be as clear as possible. Purpose is also often reflected in the nature and types of input variables. For developing countries purpose is very important in driving across the message of the classification for potential users.

***General Purpose vs. Specific Purpose Classifications***

Geodemographic systems are generally divided into *general* and *specific* purpose systems. For specific purpose systems, the classification system developer will usually be interested in particular niches like health, crime or education. The particular niche often determines the mix of input variables. A health-specific classification system would therefore encapsulate a lot more health variables than variables from any other domain. An example of such a system is Health Acorn (CACI, 2006).

For developing countries where there is a paucity of geodemographics, the recommendation from experiences gained in this thesis is that the focus should be on developing general purpose classification systems in the first instance. General purpose classifications would provide researchers with initial robust knowledge of the pattern of areal clusters present in the country. Additionally, the development of specific purpose classifications requires a large pool of initial variables for the niche of interest. From the discussions in chapters two and three, it is obvious that this may be very challenging as data problems are very common in developing countries. It is usually easier to secure a mix of variables spread across a wide variety of domains than a large concentration of variables within a particular domain of interest.

It is however important to stress that bespoke geodemographic classifications have their value. One of the downsides of general purpose classifications is that they may conceal hidden diversity (Voas and Williamson, 2001a). Such within-area dissimilarity may be exhumed by specific-purpose classifications.

Apart from the possibility of uncovering greater level of diversity within areas, another major implication of creating bespoke classifications is that the spatial patterns of cluster groups are likely to be different from the results of the general purpose classifications developed in this research.

***Relevance of Classifications to Policy Debates and Usability***

Another important lesson learnt in this research work is the place of policy relevance in developing countries when developing geodemographic systems. One of the problems discussed in chapter three is the difficulty in accessing public data from government sources. It is highly important for a classification developer to be able to demonstrate the potential benefits of the intended research for policy making. When data providers in developing countries know they stand to gain a lot (particularly in terms of easing the complexity of their work and process of decision making) they may be encouraged to released data.

Policy relevance also plays a very important role when considering the usability of the classification system. That a system is open-source does not entirely guarantee high level of usage. However, a geodemographic system that has the potential of tackling key policy debates in a country is more likely to attract the interest of policy makers.

***Understanding Organisational Cultures***

Another important lesson learnt from this exercise is how to relate diplomatically with government agencies in developing countries when making data requests. A significant chunk of academic research conducted within spatial analysis for developing countries often focuses on small geographical areas within the country. Organizational attitudes in developing countries are influenced by cultural characteristics (Hofstede, 2001). It is important when approaching these organisations for researchers to understand these organisational characteristics. The studies reported in this research prove that some of the barriers against access to data can be broken down if the researcher learns to engage in a manner that is *acceptable* within the country's organisational cultures.

## 9.5 Limitations of the Study

The previous section discussed the success of the research project. To a large extent, the aims and objectives detailed in Chapter 1 were achieved. In spite of this success, the research project is confronted by a number of limitations which are discussed in this section.

***Scale Issues: Ecological Fallacy and the Modifiable Areal Unit Problem (MAUP)***

Due to the datasets available, the Nigerian classification was created at the LGA level. The population of Nigeria is approximately 140 million people and there are 774 LGAs. This gives an approximate population of 180,000 per LGA; although this varies from area to area. This large population is bound to give rise to problems associated with the ecological fallacy which is a situation where members of a group are ascribed with the characteristics when only the overall characteristics of the group are known.

The Philippines classification also suffers from a similar problem because it is reliant on the units of analysis (Barangays) which are modifiable in terms of their population which range from as low as 10 people to over 100, 000 people.

The Nigerian and Philippines systems are built at very different administrative geographies. The Philippines administrative geography comparable to Nigerian LGAs would be Municipalities which are 1,494 in total (NSO, 2008).

When developing area segmentation systems, it is possible to take information from finer scales of geography to infer processes at larger scales and vice versa. These, approaches are what Jarvis (1996) terms as the process of upscaling and downscaling.

Since the Philippines census data are available at the barangay scale (finer than municipalities) data can be aggregated up to municipality level. This opens up an opportunity to develop a new classification system at municipality level and evaluate its results relative to the systems already developed and described in this thesis.

A system developed for the Philippines at the municipality scale will provide the opportunity to embark on robust comparative analysis with the Nigerian LGA geodemographic system.

Another potentially useful research project that may ensue from developing a Philippines classification at municipality scale is the possibility of investigating the effects of scale on inequality in the Philippines. For instance, the same set of indicators (like those analysed in chapters 7 and 8) can be profiled against the barangay classification and the municipality classification to see if there are differences or similarities in spatial patterns of results or the geodemographic correlates associated with the profiled indicators.

***Partial National Statistics***

Another limitation of the Nigerian classification system which is unavoidable is a data related problem. Some data from the census had to be combined with data from the Core Welfare Indicators Questionnaire (CWIQ). Although the CWIQ is a national survey, the data still suffers from holes which to some extent have been complemented by the census data and the fact that both datasets are for 2006.

***The Problem of Timelines***

The Philippines system benefits from the fact that it uses wholly census data which have a national coverage. However, the timeliness of the classification system may be questionable. The census data used is for 2001 – nine years ago. The static nature of the system and the time of data collection mean that some changes to the current geodemographic make-up of these areas may have occurred. However it is important to mention that without large scale urban or rural development or natural disasters, change in the make-up of social areas is unlikely to be very significant (Orford et al., 2002).

***The Pros and Cons of Crowd Sourcing Data***

The difficulty is accessing government statistics is a key limitation to some of the analyses presented particularly in chapters 7 and 8. In the future, developing countries may need to consider crowd-sourced data for spatial analysis. The process of crowd sourcing data entails making an open call to members of the public to engage in some form of consultation exercise. Such consultations may be accompanied by forms of compensation to encourage participation. While there are some benefits to doing this, it is also important to sound a note of caution.

The proliferation and acceptance of Information and Communication Technologies (ICTs) like the internet and mobile phones to developing countries (Gunasekarana and Harmantzis, 2007) provide a unique platform upon which public consultations can be anchored. Social spaces like social networking sites (Facebook and Twitter) also provide important avenues for capturing spatial data on policy relevant debates (Singleton and Longley, 2009).

Most surveys used by the UN and their partner agencies are done on a face-to-face basis. Sourcing data on the ground is expensive and often results in smaller sample sizes. The strongest benefit of crowd-sourcing data for developing countries is that it can help increase sample sizes and can be cost effective.

In spite of the rapid proliferation of ICT to developing countries, there are still relatively large numbers of people without access to these technologies. For instance as at 2009, the penetration of 3G mobile handsets for Africa and the Middle East was 7% (Morgan Stanley, 2009). In the same year, Asia (excluding Japan) also recorded a penetration of 7%.

The penetration rates for internet access in 2010 were also quite low within developing regions of the world. Africa had 9% penetration with Nigeria alone accounting for about 40% of users within the continent. In the same year, Asia recorded about 20% penetration in access to the internet. Together, China and Japan accounted for about 63% of the total population with access to the internet in Asia (MMG, 2010).

It is predicted that with increasing innovation diffusion the penetration of 3G mobile phones across Africa and the Middle East should rise to about 35% in 2014 while in Asia (excluding Japan) the rate is expected to rise to about 37% (Morgan Stanley, 2009). Current statistics however indicate that relatively small proportions of populations within developing countries have access to ICTs which are essential for sourcing information en-mass from the public.

In order to truly take advantage of crowd sourcing, one needs a large diverse crowd. Ideally, contributors to crowd sourced data should be widely spread across geographical areas and socio-economic groups in order to avoid skewed responses. This requirement presents its challenges in developing countries. The digital divide between most urban and rural dwellings is remarkably high in the developing world (Bjørn and Stein, 2007). This makes it difficult to acquire spatially balanced information using ICTs. This problem is a key is another reason why crowd-sourced information may not appear favourable in some quarters for spatial analysis.

A third problem that can arise from crowd-sourcing data is that noise may be introduced into the data often because of multiple responses. It is common to adopt the use of incentives when deploying consultation exercise via ICT platforms. Incentives are used to encourage respondents partake in the data collection exercise. In many cases, responses to crowd-sourced consultations are unsupervised. Sometimes, over-zealous respondents may want provide multiple responses which can skew the overall result. This can be very common where incentives like entries into raffle draws accompany the consultation. In order to increase their chances of winning, some respondents may provide numerous answers with varying identities.

To ameliorate this problem, a smart-computing technique like Internet Protocol (IP) address monitoring can be used to discourage multiple entries. The downside to this however is that it restricts responses to surveys to one entry per computer. This invariably reduces the sample size for a consultation exercise. In developing countries where most households lack ownership of personal computers and access to the internet is mostly via internet cafes Bjørn and Stein, (2007), the use of IP address monitoring may negatively influence sample sizes.

One more issue that can present a challenge for the use of crow-sourced data in developing countries is that internet speed may discourage participation. Broad-band internet provision is not very common in most developing countries (Bjørn and Stein, 2007). Where broad-band is not used, internet browsing speed is greatly limited. Also linked to this is the issue of mode of access to the internet. Internet users in developing countries (who mostly use cafes) have to spend substantial amounts of monies. As such once they are on the internet with only a pre-defined time-frame, what they want is ***value for money***. For many, providing answers to a survey may be considered secondary when compared to several other direct benefits like local and international networking and investment opportunities for which the internet can be used.

In most developing countries, the degree of current challenges facing most crowd-sourcing methods appears to be great. Urban-rural divide in the proliferation ICTs in these countries can mean rural dwellers may be under-represented in consultations. Additionally, certain age-groups (the very young and very old) may not utilise these ICTs whether they have access or not.

Face-to-face surveys cannot be discarded in developing countries at the present stage of their social, economic and technological development. There is still a great need for investment in technology for development informatics in these countries. Based on the arguments presented in this section, it would appear that the option of crowd-sourced data as a sole means for engaging in public consultations in developing countries at the moment may not yield robust and representative statistics across all social groups within developing societies.

If crowd-sourcing is to be used at all, it is advisable to complement the data with **extensive** face-to-face consultations. Additionally, it is recommended that consultations deployed via ICTs should not be too lengthy. Respondents may be easily put off if the surveys are lengthy. For rural areas, one way in which crowd-sourcing of information can be experimented is to have data collection nodes with on-the-ground volunteers using mobile internet technologies to source information.

## 9.6 Potential for Further Work

The research reported in this thesis has opened up many opportunities for further work in the future. This section will highlight some of the directions for which future research can ensue from this research.

***Developing Classifications at Finer Geographies in Nigeria***

It has been shown in this research how a classification system has been developed for Nigeria at the LGA scale. The examples of analysis detailed chapters 7 and 8 also help illustrate the usefulness of the classification system for policy making. There are other smaller scales of geography in Nigeria (Enumeration Areas and Postcode) at which the methodology deployed on the LGA data can be used.

### *The Philippines 2010 Census*

In 2010, a Census of population and housing was conducted for the Philippines. Although results from the census are not yet published, it may provide a possible research avenue for developing a new segmentation system.

The new Philippines census also presents a research avenue for investigation what Vickers (2010) described as 'changing social geology'. Such a project may be used to explain how small areas in the Philippines have changed over the period between 2001 and 2010 in terms of their social and economic identities.

### *Further Work on the MDGs and Multilevel Modeling*

Only a few of the MDG related variables have been evaluated in this work. There are several other MDG indicators as detailed in chapter 3 which can be further investigated.

In the future, it will be important to place emphasis on the statistical reliability of analysis conducted with geodemographic systems. Multilevel modeling techniques can be used to achieve this. Multilevel modeling is a technique embedding hierarchical linear models used to establish the variations of model parameters at varying levels (Goldstein, 2003). The technique has been used in educational research where for example the levels might be pupil, class, school and district.

In Chapters 7 and 8 for instance, quite a lot of topics were evaluated and indicators profiled against the geodemographic systems to identify geodemographic groups for which certain characteristics are under-represented or over-represented. In the future, multilevel-modeling can be used to investigate the statistical significance of the trends observed in data profiled against geodemographic systems. Harris et al. (2007) have recently demonstrated how geodemographics can be combined with multilevel modeling to increase the statistical reliability of outcomes.

### *Expanding Open-source Geodemographics to Other Developing Countries*

Only three countries now have open-source geodemographic systems. Joining the United Kingdom now are Nigeria and the Philippines. There are numerous other developing countries that can benefit from geodemographics. This provides an important avenue for research and development. At the end of this PhD programme, applications will be made to secure further research grants to expand the techniques to other developing countries particularly within the Asian, African and South American continent.

## 9.7 Concluding Statement

It is fair to say that the results from the analyses in this project confirm that amidst challenges surrounding availability and access to relevant data developing world countries can also benefit from geodemographic methods.

Apart from contributing to the literature on open geodemographics, the creation of the Nigerian and Philippines systems will be useful in solving numerous unanswered policy related questions. It will spark interest from the academic community and public sector.

Strengthening the concept of national identity has constantly been a challenge for successive governments. This research has shown as suggested by Vickers and Rees (2007) that even though closely situated neighbourhoods are alike, they are just as similar in their geodemographic make up to distant neighbourhoods. This also suggests that there will be commonalities in their wants and ways of life.

**Figure 9.1: Home Page of the Nigerian Geodemographic Classification System**

Some outputs from the analysis leading to the development of the Nigerian classification system can be accessed from www.nigerianlgaclassification.com.

**Figure 9.2 Dynamic Maps from the Nigerian Website**

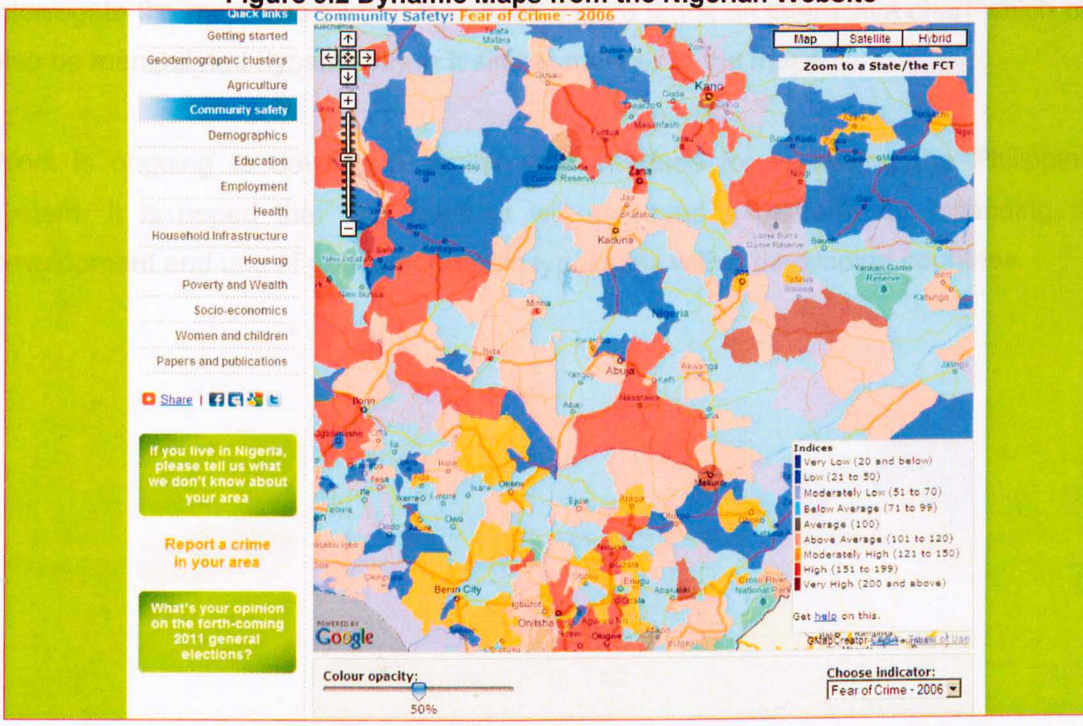
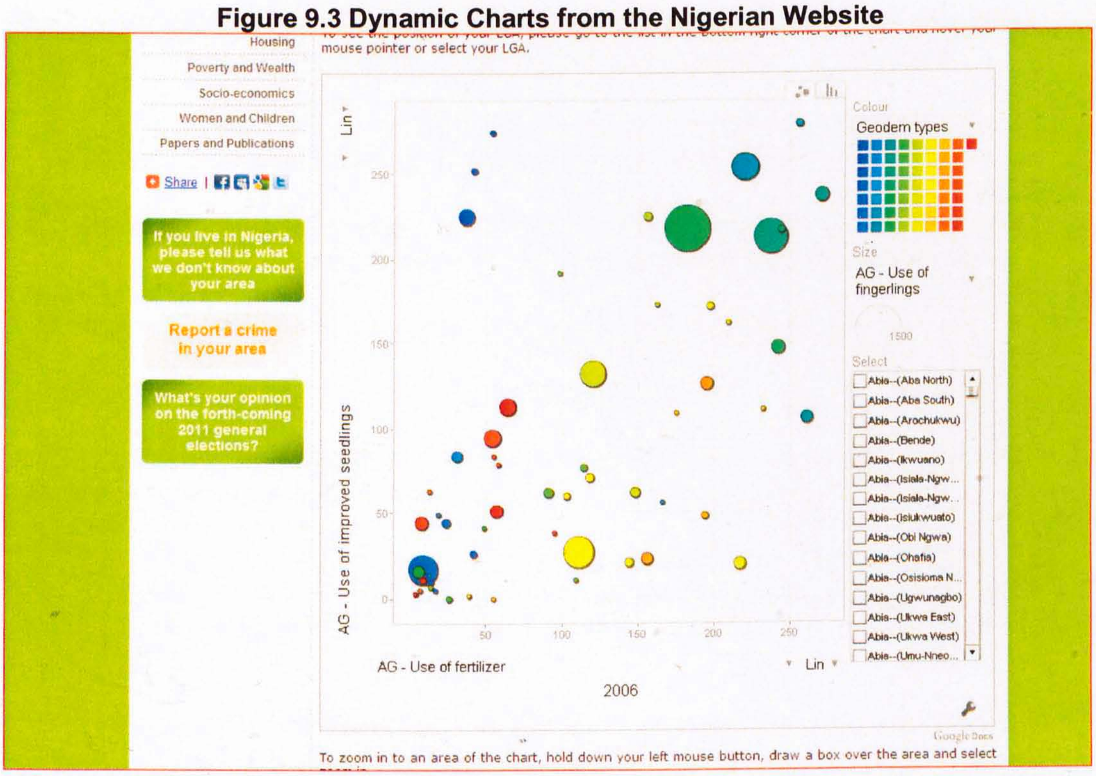



**Figure 9.3 Dynamic Charts from the Nigerian Website**



Figures 9.1 to 9.3 show some screen shots from the Nigerian website. Figure 9.1 shows the home page for accessing the Nigerian system. Figure 9.2 is a map visualization overlain on Google base map which allows users to interactively interrogate the map. The third screen shot- Figure 9.3 is an interactive chart which can also be manipulated dynamically in a similar manner to the maps.

Work is ongoing to develop a similar web interface for accessing the Philippines system. It is hoped that this exercise will serve as a platform for extending the development and use of geodemographic typologies within developing countries.

# References

ADB (2004): *Country Gender Assessment: Philippines*, Asian Development Bank, Southeast Asia Department, Manila.

ADB (2007): Philippines: Critical Development Constraints, [Online] http://www.adb.org/Documents/Books/Philippines-Critical-Dev-Constraints/highlights-phil-critical-dev-constraints.pdf, accessed 29/08/2008.

ADB (2009): *Key Indicators for Asia and the Pacific 2009*, [Online] http://www.adb.org/Documents/Books/Key_Indicators/2009/default.asp, accessed 15/11/2009.

Agresti, A. (2007): *An Introduction to Categorical Data Analysis,* 2nd Edition, John Wiley and Sons.

Agresti, A. and Finlay, B. (2008): *Statistical Methods for the Social Sciences*, 4th Edition, Prentice Hall.

Akaike, H., Parzen, E., Tanabe, K, Kitagawa, G. (1998): *Selected Papers of Hirotugu Akaike*, Springer.

Akinmoladun, O. I. (1999): *Glossary of Words and Terminologies Used in the Field of Environmental Studies*, Olak Publishing Company, Lagos.

Aldenderfer, M. S. and Blashfield, R. K. (1984): *Cluster Analysis*, London, Sage.

Andreasen, A. R. (1995): *Marketing Social Change: Changing Behavior to Promote Health, Social Development, and the Environment*, San Francisco: Jossey-Bass.

Ashby, D. I. and Longley P. A. (2005): Geocomputation, Geodemographics and Resource Allocation for Local Policing, *Transactions in GIS*, **9**, 53-72.

Bacher, J., Wenzig, K. and Vogler, M. (2004): SPSS Twostep Cluster: A first evaluation. (Arbeits- und Diskussionpapiere. 2, 2) Erlange—Nurnberg, University of Friedrich—Alexander, [Online] http://www.statisticalinnovations.com/products/twostep.pdf, accessed 10/06/2010.

Bacon, F. (2008): *Essays of Francis Bacon*, LeClue22.

Baker, K. (1977): The Utility to Market Research of ACORN - Foreword, *Journal of the Market Research Society*, **39**, 53-54.

Bales, K. (1991): Charles Booth's Survey of Life and Labour of the People in London 1889-1903, In Orford, S., Dorling, D., Mitchell, R., Shaw, M., Smith, G. D., Life and Death of the People of London: A Historical GIS of Charles Booth's Inquiry, *Health and Place* **8**, 25-35

Ballas, D., Kingston, R., Stillwell, J., and Jin J. (2004): *Building a Spatial Microsimulation Decision Support System*, Conference on Geographic Information Science, Heraklion, Greece.

Ballas D., Rossiter D., Thomas B., Clarke G. and Dorling D. (2005): *Geography Matters. Simulating the Local Impacts of National Social Policies*, United Kingdom, Joseph Rowntree Foundation.

Ballas, D. and Dorling, D. (2007): Measuring the Impact of Major Life Events upon Happiness, *International Journal of Epidemiology*, **36**, 1244-1252.

Ballweg, J. A. (1972): Sinfulness of Family Limitation: Interpretation vs. Practice in the Philippines, *Sociological Analysis*, **33**, 110-116.

Banerjee, S., Carlin. B. P., Gelfand, A. E. (2004): *Hierarchical Modeling and Analysis for Spatial Data*, CRC Press.

Batey, P. W. J. and Brown P. J. B. (1995): From Human Ecology to Consumer Targeting: The Evolution of Geodemographics, In Longley, P. and Clarke G. P. Eds., *GIS for Business and Service Planning*, Cambridge, GeoInformation International.

Beaumont, J. R. and Inglis, K. (1989): Geodemographics in Practice: Developments in Britain and Europe, *Environment and Planning A*, **21**, 587-604.

Becker, S., Peters, D. H., Gray, R. H., Gultiano, C., and Black, R. E. (1993): The Determinants of Use of Maternal Child Health Services in Metro Cebu, the Philippines, *Health Transition Review*, **3**, 77-91.

Berrar, D. P., Dubitzky, W., Granzow, M. (2003): A Practical Approach to Microarray Data Analysis, Springer.

Birkin M., and Clarke, G. (1998): GIS, Geodemographics, and Spatial Modeling in the U.K. Financial Service Industry, *Journal of Housing Research*, **9**, 87-111.

Birkin M and Clarke G (2009) Geodemographics, in Kitchin R, Thrift N, (Eds) International Encyclopaedia of Human Geography, Elsevier, 382-389.

Bishop, I. D., Escobar, F. J., Karuppannan, S., Suwarnarat, K., Williamson, I. P., Yates, P. M. and Yaqub, H. D. (2000): Spatial Data Infrastructures for Cities in Developing Countries: Lessons from the Bangkok Experience, *Cities*, **17**, 85-96.

Bjørn, F. and Stein, K. (2007): *A Rural-urban Digital Divide? Regional Aspects of Internet Use in Tanzania*, Proceedings of the 9th International Conference on Social Implications of Computers in Developing Countries, São Paulo, Brazil.

Blake, M. and Openshaw, S. (1995): *Selecting Variables for Small Area Classification of 1991 UK Census Data*, Working Paper 95/2, School of Geography, University of Leeds, [Online] http://www.geog.leeds.ac.uk/papers/95-2/, accessed 07/11/2006.

Bloom, D. E., Sachs, J. D., Collier, P., Udry, C. (1998): Geography, Demography and Economic Growth in Africa, *Economic Activity*, **2**.

Bongaarts, J. and Bulatao, R. A. (2000): *Beyond Six Billion: Forecasting the World's Population*, National Academies Press. Washington, D.C.

Boserup, E., Kanji, N., Tan, S. F., and Toulmin, C. (2007): *Woman's Role in Economic Development*, Earthscan.

Bowers, K. and Hirschfield, A. (1999): Exploring Links between Crime and Disadvantage in North-West England: An Analysis Using Geographical Information Systems, *International Journal for Geographical Information Science*, **13**, 159-184.

Brinkman, R. (1995): Economic Growth Versus Economic Development: Toward a Conceptual Clarification, Economic Issues, **29**

Brown, M. C. (1994): Using Gini-style Indices to Evaluate the Spatial Patterns of Health Practitioners: Theoretical considerations and an Application Based on Alberta Data, *Social Science Medicine,* **38**, 1243-1256.

Brown, P. J. (1991): *Exploring Geodemographics*, In Handling Geographical Information, Ed. Masser, I. and Blakemore, M. J., 221–258. London, Longman.

Brown, P. J. B. (1989): *Geodemographics: A Review of Recent Developments and Emerging Issues*, Liverpool, URPERRL.

Brown, P. J. B., Batey, P. W. J., Hirschfield, A. F. G. and Marsden, J. (1995): *Analyzing Spatial Patterns of Disease: Some Issues in the Mapping of Incidence Data of Relatively Rare Conditions*, In M.J.C. de Lepper, H.J. Scholten and R.M. Stern Eds., The Added Value of Geographical Information Systems in Public and Environmental Health, 145-163. Dordrecht: Kluwer.

Brown, P. J. B., Hirschfield, A. F. G. and Batey, P. W. J. (2000): *Adding Value to Census Data: Public Sector Applications of Super Profiles Geodemographic Typology*, Working Paper 56, URPERRL, Department of Civic Design, University of Liverpool.

Brown, P. J. B., Hirschfield, A. F. G., Merrall, S., Bowers, K. J., and Marsden, J. (1999): *Targeting Community Safety Projects: The Use of Geodemographics and GIS in the Identification of Priority Areas for Action*, 39th Congress of the European Regional Science Association, Dublin.

Brown, P. J. B., McCulloch, P. G., Williams, E. M. I. and Youngson, J. (1998): *Control Sample Design using a Geodemographic Discriminator: An Application of Super Profiles*, Working Paper 45, URPERRL, Department of Civic Design, University of Liverpool.

Bulmer, M. G. (1979): *Principles of Statistics*, Dover publications.

Burrows, R., Ellison, N. and Woods, B. (2005): *Neighbourhoods on the Net: Internet-Based Neighbourhood Information Systems and their Consequences*, Bristol: Policy Press.

Burrows, R. and Gane, N. (2006): Geodemographics, Software and Class, *Sociology*, **40**, 793-812, London.

CACI Limited (2006): Health Acorn User Guide, London.

Caldwell, J. C. (1981): Maternal Education as a Factor in Child Mortality, *World Health Forum*, **2**, 75-78.

Callingham, M. (2003): *Current Commercial Sector Use of Geodemographics and the Implications for the ONS Area Classification Systems*, [E-mail Correspondence to Vickers, D.] 14/10/03.

Catterall, J. (1985): *On the Social Costs of Dropping Out of Schools* (Report No. SEPI-86-3). Stanford, Stanford University, Center for Educational Research.

Chatham House (2009): State Government Solutions to the Niger Delta Troubles: Key Challenges, Needs and Strategies, [Online] http://www.chathamhouse.org.uk/files/13602_260209nigerdelta.pdf, accessed 13/12/2009.

Chen, S. and Ravallion, M. (2002): *How Did the World's Poorest Fare in the 1990s?* Working Paper No. 2409, pp.1-5. World Bank, Washington, D.C. [Online] http://www.worldbank.org/research/povmonitor/publications.htm, accessed 10/05/2009.

Chen, S. and Ravallion, M. (2007): *Absolute Poverty Measures for the Developing World, 1981-2004*, World Bank Development Research Group. [Online] http://mdgs.un.org/unsd/mdg/Resources/Attach/Capacity/Chen%20&%20Ravallion%20(2007).pdf, accessed 10/05/2009.

Clark, A. E. (2003): Unemployment as a Social Norm: Psychological Evidence from Panel Data, *Journal of Labor Economics*, University of Chicago Press, **21**: 289-322.

Constitutional Commission of the Philippines (1987): 1987 Constitution of the Republic of the Philippines [Online] http://www.thecorpusjuris.com/laws/constitutions/8-philippineconstitutions/70-1987-constitution.html accessed 25/09/2009.

Coursey, O. W. (2008): *History and Geography of the Philippine Islands*, BiblioBazaar.

Crawford, L and Sidener, N. L. (1982): *Women's Formal Education and Economic Growth: The Case of the Philippines*, Paper presented at the National Conference on the Third World 5th, Omaha, NE, October, 1982.

Crawshaw, J and Chambers, J. (2001): *A Concise Course in Advanced Level Statistics with Worked Examples*, 4th Ed. Cheltenham, Nelson Thornes.

Curry, M. (1998): *Digital Places: Living with Geographic Information Technologies, Routledge*, London.

De Smith, M. J. (2006): Handbook of GIS Analysis.

Deichmann, U. (1999): Geographic Aspects of Inequality and Poverty, [Online] http://siteresources.worldbank.org/INTPGI/Resources/Pro-Poor-Growth/5319_povmap.pdf, accessed 13/08/2009.

Diallo, A. B, Wodon, Q. (2007): Demographic Transition Towards Smaller Household Sizes and Basic Infrastructure Needs in Developing Countries, *Economics Bulletin*, **11**: 1-11.

Dikhanov, Y. (2005): *Human Development Report 2005: Trends in Global Income Distribution, 1970-2000, and Scenarios for 2015*, UNDP.

Dorling, D. and Ballas D., (2008): Spatial Divisions of Poverty and Wealth, in Ridge, T, Wright, S (Eds), *Understanding Poverty, Wealth and Inequality: Policies and Prospects*, Policy Press, Bristol.

Dunteman, G. H. (1989): *Principal Components Analysis*, Sage Publications, Inc., USA.

Environmental Analytics (2004): *PRIZM$_{CE}$ Methodology*, [Online] http://www.tetrad.com/pcensus/can/environics/prizmce.html, accessed 16/1/2007.

Everitt, B.S. (1993): *Cluster Analysis*, 3rd Ed., London, Edward Arnold.

Everitt, B. S., Landau, S. and Leese M. (2001): *Cluster Analysis*, 4th Ed. London, Arnold.

Experian (2005): *Mosaic Global: E-Handbook*, [Online] http://www.appliedgeographic.com/Mosaic%20Global%20E-Handbook.pdf, accessed 12/1/2007.

Federal Republic of Nigeria Official Gazette (1999): *Freedom of Information Bill, 1999*, The Federal Government Press, Lagos [Online] http://www.foicoalition.org/PDF%20Files/FOI%20Official%20Gazette.pdf, accessed 19/04/2009.

Fields, G. (1989): Changes in Poverty and Inequality in Developing Countries, *World Bank Research Observer*, **4**, 167-185.

Fisher, M. and Getis, A. (1999): New Advances in Spatial Interaction Theory, *Papers in Regional Science*, **78**, 117-118.

Fisher, P. F. and Langford, M. (1995): Modeling the Errors in Areal Interpolation between Zonal Systems by Monte Carlo Simulation, *Environment and Planning A* 27(2), 211-224.

Foley, T. (1997): Business Minded, *New Perspectives*, **6**, p. 6

Fotheringham, A. S. (1989): Scale-independent spatial analysis, In *Accuracy of Spatial Databases*, Eds. Goodchild, M., and Gopal, S., Taylor and Francis, 221-228.

Fotheringham, A. and O'Kelly, M. (1989): *Spatial Interaction Models: Formulation and Applications*, Kluwer, Dordrecht.

Fotheringham, A. S. and Rogerson, P. A. (1993): GIS and Spatial Analytic Problems, *International Journal of Geographical Information Systems*, **7**, 3 - 19.

Fotheringham, A. S. and Wong, D. W. S. (1991): The Modifiable Areal Unit Problem in Statistical Analysis, *Environment and Planning A*, **23**, 1025-1044.

Fraley, C. and Raftery, A. E. (1998a): How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis, The computer Journal (8) 578-588.

Fraley, C. and Raftery, A. E. (2002): Model-Based Clustering, Discriminant Analysis, and Density Estimation, Journal of the American Statistical Association (458) 611-631.

Furtado X. (2001): Decentralization and Public Health in the Philippines, *Development*, **44**: 108-116.

Gardner, J. and Andrew, O. (2002): *Does Money Buy Happiness? A Longitudinal Study Using Data on Windfalls*, Royal Economic Society Annual Conference 2002 81, Royal Economic Society.

Gehlke, C. E. and Biehl, K. (1934): Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material, *Journal of the American Statistical Association Supplement*, **29**, 169-170.

George, C. K. (1999): *Basic Principles and Methods of Urban and Regional Planning*, 1st Ed., Lagos, Libro-Gem.

Gething, P. W., Noor, A. M., Gikandi, P. W., Ogara, E. A. A., Hay, S. I., Nixon, M. S., Snow, R. W., and Atkinson, P. M. (2006): *Improving Imperfect Data from Health Management Information Systems in Africa Using Space–Time Geostatistics*, PLoS Medicine **3**, [Online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1470663, accessed 12/05/2009.

Gibin, M., Mateos, P., Atkinson, P., Petersen, J. (2009): Google Maps Mash-ups for Local Public Health Service Planning, in Geertman, S. and Stillwell, J. (eds.) *Planning Support Systems: Best practices and New Methods*, Springer.

Gittus, E. (1964): The Structure of Urban Areas: A New Approach, *Town and Planning Review*, **35**, 5-20.

Gnanadesikan, R., Kettenring, J. R and Tsao, S. L. (2005): Weighting and Selection of Variables for Cluster Analysis, *Journal of Classification*, **12**, 113-136.

Goldstein, H. (2003): Multilevel Statistical Models. London: Hodder Arnold.

Goodchild, M. F., Anselin, L., and Diechmann, U. (1993): A Framework for Areal Interpolation of Socio-economic Data, *Environment and Planning A*, **25**, 383 – 397.

Gordon, A. (2003): *Nigeria's Diverse Peoples: A Reference Source Book*, ABC-CLIO, Santa Barbara.

Gordon, R. G., Jr. (ed.), (2005): *Ethnologue: Languages of the World*, 15th Edition. Dallas.

Gordon, A. D. (1999): *Classification*. 2nd Ed., London Chapman and Hall.

Graham, S. (2004): The Software Sorted City: Rethinking the Digital Divide, In Burrows, R. and Gane, N., Geodemographics, Software and Class, *Sociology*, **40**, 793-812, London, BSA Publications.

Gray, C. W., Kaufmann, D. (1998): Corruption and Development, *Finance and Development*, March, 7- 10 [Online] http://www.derechoasaber.org/documentos/pdf0057.pdf, accessed 18/08/2008.

Gregory, I. N. (2000): An Evaluation of the Accuracy of the Areal Interpolation of Data for the Analysis of Long-Term Change in England and Wales [Online] http://www.geocomputation.org/2000/GC045/Gc045.htm . Accessed 5/3/07.

Grusky, D. B., Kanbur, S. M. R., Sen, A. K. (2006): *Poverty and Inequality*, Stanford University Press.

Gunasekarana, V. and Harmantzis, F. C. (2007): Emerging Wireless Technologies for Developing Countries, *Technology in Society*, **29**, 23-42.

Hallberg, K. (2000): *A Market-oriented Strategy for Small and Medium Scale Enterprises*, World Bank.

Harris, R., Johnston, R., Burgess, S. (2007): Neighborhoods, Ethnicity and School Choice: Developing a Statistical Framework for Geodemographic Analysis, *Population Research and Policy Review*, **26**, 553-579.

Harris, R., Sleight, P. and Webber, R. (2005): *Geodemographics, GIS and Neighbourhood Targeting*, London, Wiley.

Hofstede, G. (2001): *Culture's Consequences: Comparing Values, Behaviours, Institutions, and Organizations across Nations*, Sage Publications.

HDN and UNDP (2000): Philippine Human Development Report 2000, Human Development Network and The United Nations Development Programme, [Online] http://78.136.31.142/en/reports/nationalreports/asiathepacific/philippines/name,2789,en.html accessed 21/20/2009.

Hellerstein, J. M. and Stonebraker, M. (2005): Readings in Database Systems, 4th Ed., Cambridge, Massachusetts Institute of Technology Press.

Hirschfield, A.F.G., Bowers, K.J. and Brown, P.J.B. (1995): Exploring Relations between Crime and Disadvantage on Merseyside, *European Journal on Criminal Policy and Research*, **3**, 93-112.

Holz, C. A. (2002): Institutional Constraints on the Quality of Statistics in China, *China Information*, **16**, 25-67.

House, J. S., Kessler, A. R., Herzog, R. (1990): Age, Socioeconomic Status and Health, *The Milbank Quarterly*, **68**, 383-411.

IMAC (2006): *The Infant Immune System and Immunisation, Immunisation Advisory Centre*, Auckland, [Online] http://www.immune.org.nz/site_resources/Professionals/Vaccinology/The_infant_immune_system_and_immunisation.pdf, accessed 28/11/2009.

International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, World Bank Group (2000): *A Better World for All: Progress towards the International Development Goals*, [Online] http://www.paris21.org/betterworld/pdf/bwa_e.pdf, accessed 10/11/2008.

Inter Press Service (1998): Women: Reservations Grow over UN Women's Treaty. [Online] http://web.archive.org/web/20040423160533/http://www.oneworld.org/ips2/mar98/unwomen.html. Accessed 11/10/07.

International Labour Office (2004): *Key Indicators of the Labour Market*, 3rd Edition, Routledge.

Jarvis, P. G. (1995): Scaling Processes and Problems, *Plant, Cell, and Environment*, **18**, 1079-1089.

Jelinski, D. E. and Wu, J. (1996): The Modifiable Areal Unit Problem and Implications for Landscape Ecology, *Landscape Ecology*, **11**, 129-140.

Jonhson, N. L., Kotz, S., Balakrishnan, N. (1994): *Continuous Univariate Distributions, Second Ed., Vol. 1,* John Willey and Sons.

Johnson, S. C. (1967): Hierarchical Clustering Schemes, *Psychometrika*, **2**, 241-254.

Johnston, R.J., Rossiter, D.J. and Pattie, C.J. (2005). Disproportionality and bias in US Presidential elections: how geography helped Bush defeat Gore but couldn't help Kerry beat Bush. *Political Geography*, **24**, 952-968.

Jolliffe, I. T. (2002): *Principal Components Analysis*, New York, Springer.

King, E. M. and Lillard, M. A. (1987): Education Policy and Schooling Attainment in Malaysia and the Philippines, *Economics of Education Review*, **6**, 167-181.

Kinloch S., Hoen B., Cooper D., Hirschel B., Lazzarin A., Mathiesen L., de Witt S., Antunes R., Luthy J., Von Overbeck J. (1994): *Treatment of Primary HIV Infection with Zidovudine*, International Conference on AIDS, Switzerland.

Kirch, W. (2008): *Encyclopedia of Public Health*, Springer.

Kojima, R. (1996): Population Migration and Urbanisation in Developing Countries, *Development Economies*, **34**: 349-369.

Lakshminarayanan, R. (2003): Decentralisation and its Implications for Reproductive Health: The Philippines Experience, *Reproductive Health Matters*, **11**: 96-107.

Lallana, E. C. and UY, M. (2003): The Information Age, [Online] http://www.apdip.net/publications/iespprimers/eprimer-infoage.pdf, accessed 20/02/2009.

Langford, M., Maguire, D., Unwin, D. J. (1991): The Areal Interpolation Problem: Estimating Population Using Remote Sensing in a GIS Framework, In Masser, I. and Blakemore, M. (Eds.), *Handling Geographical Information: Methodology and Potential Applications*, New York: Longman, 55-77.

Larose, D. T. (2006): Data Mining Methods and Models, Wiley

Larson, R. C. (1986): The Invalidity of Modifiable Areal Unit Randomization, *The Professional Geographer*, **38**, 369-374.

Leventhal, B. (1995): Evaluation of Geodemographic Classifications, *Journal of Targeting, Measurement and Analysis for Marketing*, 1-11.

Lindblom, C. E., Woodhouse, E. J. (1992): The Policy Making Process, 3rd Edition, Prentice Hall.

Lindroos, A. (2004): *Antenatal Care and Maternal Mortality in Nigeria*, [Online] http://www.uku.fi/kansy/eng/antenal_care_nigeria.pdf, accessed 23/07/2009.

Longley, P. A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. (2001): *Geographic Information Systems and Science*, U.S.A, John Wiley and Sons.

Manly, B. (1994): *Multivariate Statistical Methods*, London, Chapman and Hall.

Marceau, D. J. (1999): The Scale Issue in Social and Natural Sciences, *The Canadian Journal of Remote Sensing*, **25**, 347-356.

Martin, P. and Rogers, C. A. (1995): Industrial Location and Public Infrastructure, *Journal of International Economics*, **39**: 335-351.

McLennan, W. (1997): Mental Health and Wellbeing: Profile of Adults, Australia, Australian Bureau of Statistics.

Miller, H. J. and Han, J. (2001): *Geographic Data Mining and Knowledge Discovery*, Taylor and Francis, New York.

Milligan, G. W. (1996): Clustering Validation: Results and Implications for Applied Analyses, In Arabie, P., Hubert, L. J. and De Soete, G. Eds., *Clustering and Classification*, Singapore, World Scientific.

Milligan, G. W. and Stephen, C. H. (2003): Clustering and Classification Methods, In *Handbook of Psychology*, Irvin, B. W., Schinka, J. A., Velicer, W. F. (Eds), John Wiley and Sons.

Miniwatts Marketing Group (2010): *Internet World Stats: Usage and Population Statistics* [Online] http://www.internetworldstats.com/stats.htm, accessed 25/07/2010.

Montalvo, J. G. and Reynal-Querol, M. (2005a): Ethnic Polarization, Potential Conflict, and Civil Wars, American Economic Review, **95**, 796-813.

Montalvo, J. G. and Reynal-Querol, M. (2005b): Ethnic Diversity and Economic Development, Journal of Development Economics, **76**, 293-323.

Montinola, G. R. (1999): Parties and Accountability in the Philippines, Journal of Democracy, **10**: 126-140.

Morgan, L. A. (1991): After Marriage Ends: Economic Consequences for Midlife Women, Sage Publications.

Morgan Stanley (2009): *The Mobile Internet Report Setup*, [Online] http://www.morganstanley.com/institutional/techresearch/pdfs/2SETUP_12142009_RI.pdf, accessed 25/07/2010.

Moser, C. A. and Scott, W. (1961): British Towns: A Statistical Study of their Social and Economic Differences, Edinburgh, Oliver and Boyd Ltd.

NBS (2006a): *Core Welfare Indicators Questionnaire Survey, Final Statistical Report, Federal Republic of Nigeria*, National Bureau of Statistics.

Nelson, J. M. (1998): *Poverty Inequality and Conflict in Developing Countries*, Rockefeller Brothers Fund, Inc [Online] http://www.rbf.org/usr_doc/Proverty,_Inequality_-Conflict_in_Developing_Countries.pdf, accessed 27/07/2009.

NEDA and UNDP (2007): Philippines Midterm Progress Report on the Millennium Development Goals, [Online] http://www.neda.gov.ph/econreports_dbs/MDGs/midterm/01-96%20UNDP_final.pdf, accessed 15/01/09.

Nguyen, P. I. (2007): Religious Practice and its Effects to Educational Attainment in Southeast Asia, Paper presented at the Annual Meeting of the American Sociological Association, New York, [Online] http://www.allacademic.com/meta/p183995_index.html, accessed 30/01/2010.

NPC (2004a): *Meeting Everyone's Needs: National Economic Empowerment Development Strategy*, National Planning Commission, Nigeria.

NPC (2007): *Nigeria Millennium Development Goals: 2006 Report*, National Planning Commission, Abuja.

NPC and ORC Macro (2004): *Nigeria Demographic and Health Survey 2003*, Calverton, Maryland.

NSO (2001): Final Counts by Region, Province, Municipality and Barangay, *2000 Census of Population and Housing - National Reports*, [Online] http://www.census.gov.ph/census2000/index.html, accessed 12/01/09.

NSO (2002a): Population Expected to Reach 100 million Filipinos in 14 years, *2000 Census of Population and Housing - National Reports*, [Online] http://www.census.gov.ph/data/pressrelease/2002/pr02178tx.html, accessed 12/01/09.

NSO (2002b): Cordillera Administrative Region: The least Populous Region in the Philippines, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2002/pr02199tx.html, accessed 12/01/09.

NSO (2002c): Bicol Region: One in Two Households Used Wood for Cooking, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2002/pr02198tx.html, accessed 12/01/09.

NSO (2002d): Western Mindanao: Over Half of the Total Households Used Kerosene/gas for Lighting, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2002/pr02197tx.html, accessed 12/01/09.

NSO (2002e): Southern Mindanao: Ninety Percent of the Population were Literate, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2002/pr02188tx.html, accessed 12/01/09.

NSO (2002f): Central Mindanao: Literacy Rate at 87 Percent, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2002/pr02196tx.html, accessed 12/01/09.

NSO (2002g): Caraga Region: Three in Every Ten Households Owned Agricultural Land, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2002/pr02190tx.html, accessed 12/01/09.

NSO (2003a): Ilocos Region: To reach Five Millionth Mark in Nine Years, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0319tx.html, accessed 12/01/09.

NSO (2003b): Central Luzon: Third Biggest Region, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0314tx.html, accessed 12/01/09.

NSO (2003c): Southern Tagalog: Biggest Region in the Philippines, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0303tx.html, accessed 12/01/09.

NSO (2003d): Eight percent of the total populations were from Western Visayas, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0304tx.html, accessed 12/01/09.

NSO (2003e): Central Visayas: Three in Every Five households had Electricity, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0302tx.html, accessed 12/01/09.

NSO (2003f): Eastern Visayas: Population to Increase by 149 Persons per Day, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0313tx.html, accessed 12/01/09.

NSO (2003g): Northern Mindanao: Half of the Population were below 20 Years, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0311tx.html, accessed 12/01/09.

NSO (2003h): Autonomous Region in Muslim Mindanao: Nine in Every Ten Persons were Muslims, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0301tx.html, accessed 12/01/09.

NSO (2003i): National Capital Region: Close to 10 Million Persons, *2000 Census of Population and Housing – Reports by Region*, [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0312tx.html, accessed 12/01/09.

NSO (2003j): 2000 Census of Population and Housing [Online] http://www.census.gov.ph/data/pressrelease/2003/pr0323tx.html, accessed 10/09/08.

NSO (2008): The Philippines in Figures 2008, NSO [Online] http://www.census.gov.ph/data/publications/PIF2008.pdf, accessed 02/11/08.

NSO and Department of Education (2009): Simple Literacy of the Population 10 Years Old and Over 1989, 1994, 2000 and 2003 [Online] http://www.nscb.gov.ph/secstat/d_educ.asp accessed 20/10/2009.

NSO and ORC Macro (2004): *Philippine Demographic and Health Survey 2003*, Calverton, Maryland.

O'Day, R. and Englander, D. (1993): Mr Charles Booth's Inquiry: Life and Labour of the People in London Reconsidered, In Orford, S., Dorling, D., Mitchell, R., Shaw, M., Smith, G. D., Life and Death of the People of London: A Historical GIS of Charles Booth's Inquiry, *Health and Place*, **8**, 25-35.

OECD (2008): *Handbook on Constructing Composite Indicators: Methodology and User Guide*, Organisation for Economic Co-operation and Development, OECD Publishing, Paris.

Ogunbodede, E. F. (2006): *Developing Geospatial Information for Poverty Reduction: Lessons and Challenges from Nigeria's 2006 Census*, GSDI-9 Conference Proceedings, Santiago, Chile.

Okonjo-Iweala, N. (2007): Improving Health Statistics in Africa, Brookings Institution [Online] http://www.brookings.edu/articles/2007/1103_health_okonjo-iweala.aspx, accessed 10/6/2008.

Olubor, R. O. (2009): Private Cost Analysis of Pre-School Education in Nigerian Private Schools, *Journal of Social Sciences*, **19**, 141-148.

Openshaw, S. (1977): Optimal Zoning Systems for Spatial Interaction Models, *Environment and Planning A*, 169-184.

Openshaw, S. (1984a): Ecological Fallacies and the Analysis of Areal Census Data, *Environment and Planning A*, **16**, 17-31.

Openshaw, S. (1984b): The Modifiable Areal Unit Problem: *Concepts and Techniques in Modern Geography,* **38**, Norwich, Geo Books.

Openshaw, S. (1989): Making Geodemographics More Sophisticated, *Journal of the Market Research Society*, **31**, 111-131.

Orford, S., Dorling, D., Mitchell, R., Shaw, M., Smith, G. D. (2002): Life and Death of the People of London: A Historical GIS of Charles Booth's Inquiry, *Health and Place*, **8**, 25-35.

Pfautz, H. W. (1967): *Sociologist of the City*. In C Booth, On the City: Physical Pattern and Social Structure (selected writings). London, University of Chicago Press.

Plackett, R.L. (1983): Karl Pearson and the Chi-Squared Test, *International Statistical Review* **51**: 59–72.

Plato, De Botton, A., Knight, M. J., Jowett, B. (1999): *The Essential Plato*, Quality Paperback Book Club.

Powell, J., Tapp, A., Sparks, E. (2007): Social Marketing in Action - Geodemographics, Alcoholic Liver Disease and Heavy Episodic Drinking in Great Britain, *International Journal of Non Profit and Voluntary Sector Marketing*, **12**, 177 – 187.

Prasad, E., Rogoff, S. W., Kose, M. A. (2003): Effects of Financial Globalization on Developing Countries: Some Empirical Evidence, IMF.

Pratap, S. and Quintin, E. (2006): *The Informal Sector in Developing Countries: Output, Assets and Employment*, UNU-WIDER [Online] http://www.wider.unu.edu/publications/working-papers/research-papers/2006/en_GB/rp2006-130/, accessed 20/03/09.

RBM (2008): *Key Facts, Figures and Strategies: The Global Malaria Action Plan*, Roll Back Malaria Partnership.

Robson, B. T. (1971): *Urban Analysis: A Study of City Structure*, Cambridge, Cambridge University Press.

Rodrigue, J-P., Comtois, C., Slack, B. (2009): *The Geography of Transport Systems*, Second Edition, Routledge, New York.

Rosling, H. (2009): *Let My Dataset Change Your Mindset*, [Online] http://www.ted.com/talks/hans_rosling_at_state.html, accessed 30/10/2009.

Ross, D. C. and Harmsen, R. T. (2001): *Official Financing for Developing Countries*, International Monetary Fund.

Ross, J. A. and Winfrey, W. L. (2002): Unmet Need for Contraception in the Developing World and the Former Soviet Union: An Updated Estimate, *International Family Planning Perspectives*, **28**

Salisbury, D., Ramsay, M., Noakes, K. (2006): *Immunisation against Infectious Disease*, The Stationery Office under License from the Department of Health, London.

See, L. and Gibson, P. (2006): *Using Geodemographics and GIS for Sustainable Development,* In Campagna, M., GIS for Sustainable Development, Taylor and Francis, USA.

Schelzig, K. (2005): *Poverty in the Philippines: Income, Assets, and Access*, Asian Development Bank.

Schwarz, G. (1978): Estimating the Dimension of a Model, Annals of Statistics, **6**, 461-464.

Shevky, E. and Bell, W. (1955): *Social Area Analysis: Theory, Illustrative Application and Computational Procedures*, Stamford, Stamford University Press.

Shevky, E. and Williams, M. (1949): *The Social Areas of Los Angeles: Analysis and Typology*, Berkeley, University of California Press.

Shin, Y. (2007): Peer Relationships, Social Behaviours, Academic Performance and Loneliness in Korean Primary School Children, *School Psychology International*, **28**, 220-236.

Siliwinski, A. (2002): Spatial Point Pattern Analysis for Targeting Prospective New Customers: Bringing GIS Functionality into Direct Marketing, *Journal of Geographic Information and Decision Analysis*, **6**, 31-48.

Simey, T. and Simey, M. (1960): *Charles Booth. Social Scientist*, Oxford University Press, London.

Singleton, A., Davidson-Burnett, G., and Longley, P. (2007): *University Market Area Analysis for Widening Participation*, Centre for Education in the Built Environment, Cardiff.

Singleton, A. D. and Longley, P. A. (2009): Geodemographics, Visualisation, and Social Networks in Applied Geography, *Applied Geography*, **29**, 289-298.

Sleight, P. (1997): *Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business*, NTC Publications, Henley-on-Thames.

Sleight, P. (2004): *Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business*, Henley-on-Thames, World Advertising Research Centre.

Smith, G. C. S. and Pell, J. P. (2001): Teenage pregnancy and Risk of Adverse Perinatal Outcomes Associated with First and Second Births: Population Based Retrospective Cohort Study, *British Medical Journal*, **323**, 1-5.

Smith, L. C., Haddad, L. J. (2000): *Overcoming Child Malnutrition in Developing Countries: Past Achievements and Future Choices*, International Food Policy Research Institute.

Spicker, P. (1993): *Poverty and Social Security: Concepts and Principles*, Taylor and Francis.

Steel, D. G. and Holt, D. (1996): Analyzing and Adjusting Aggregation Effects: The Ecological Fallacy Revisited, *International Statistical Review*, **64**, 39-60.

Taylor, C., Gorard, S. and Fitz, J. (2003): The Modifiable Areal unit Problem: Segregation between Schools and Levels of Analysis, *International Journal of Social Research Methodology*, **6**, 41-60.

Thrall, G.I. (1998): GIS Applications in Real Estate and Related Industries, *Journal of Housing Research*, **9**, 33-59.

Tobler, W. R. (1970): A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*. **46**, 234-240.

Todd, P., Bundred, P., Clarke, J. R. E., Brown, P. J. B. and Forbes, H. (1994): *GIS in Health Care Planning: Locating Cancer Treatment Centres*, Working Paper 41, URPERRL, Department of Civic Design, University of Liverpool.

Tritschler, A. and Gopinath, R. (1999): *Improved Speaker Segmentation and Segments Clustering Using the Bayesian Information Criterion*, Proceedings of Eurospeech, 678-682.

Tsang, M. C. (2002): *Comparing the Costs of Public and Private Schools in Developing Countries*. In Levin, H. and McEwan, P (eds.) 2002 Yearbook of the American Education Finance Association.

Tyler, J. H. and Lofstrom, M. (2009): Finishing High School: Alternative Pathways and Dropout Recovery, *America's High Schools*, **19**, 77-103.

Umeora, O. U., Esike, C. O., and Egwuatu. V. E. (2005): Maternal Mortality in Rural Nigeria, *International Journal of Gynaecology and Obstetrics*, **88**, 321-322.

UN (1945): *Charter of the United Nations, June 26, 1945, 59 Stat. 1031, T.S. 993, 3 Bevans 1153, Entered into Force Oct. 24, 1945*, [Online] http://www1.umn.edu/humanrts/instree/aunchart.htm, accessed 25/02/2008.

UN (1948): *Universal Declaration of Human Rights*, [Online] http://www.hrea.org/index.php?base_id=104&language_id=1&erc_doc_id=445&category_id=24&category_type=3&group=, accessed 10/06/09.

UN (2000): *Resolution Adopted by the General Assembly: United Nations Millennium Declaration*, [Online] http://www.un.org/millennium/declaration/ares552e.pdf, accessed 17/05/07.

UN (2003): *Indicators for Monitoring the Millennium Development Goals: Definitions, Rationale, Concepts and Sources*, United Nations, New York.

UN (2006): *The Millennium Development Goals Report*, United Nations, New York.

UN (2007): *The United Nations Development Agenda: Development for All*, United Nations, New York.

United Nations (2008a): *The United Nations Today*, United Nations Department of Public Information.

UN (2008b): *The Millennium Development Goals Report*, United Nations, New York.

UN (2008c): Official List of MDG Indicators, [Online] http://mdgs.un.org/unsd/mdg/Resources/Attach/Indicators/OfficialList2008.pdf, accessed 13/07/09

UN (2009a): *The Millennium Development Goals Report*, UN, New York.

UN (2009b): *World Population Prospects: The 2008 Revision*, [Online] http://www.un.org/esa/population/publications/wpp2008/wpp2008_highlights.pdf, accessed 20/4/09.

UNCHS (2003): *Slums of the World: The Face of Urban Poverty in the New Millennium*, United Nations Human Settlements Programme, *UN-HABITAT*.

UNDP (1990): *Human Development Report 1990*, [Online] http://hdr.undp.org/en/reports/global/hdr1990/chapters/, accessed 20/2/08.

UNDP (2007): MDG Monitor, [Online] http://www.mdgmonitor.org/index.cfm, accessed 15/05/2009.

UNESCO (2000): *UNESCO/Nigeria Co-operation for Universal Basic Education*, United Nations Educational, Scientific and Cultural Organization.

UNFPA (2007): State of world population 2007. Unleashing the Potential of Urban Growth, United Nations Population Fund, New York.

UNICEF (2000): *Monitoring Progress toward the Goals of the World Summit for Children*, United Nations Children's Fund, New York.

UNICEF (2004): Malaria: A major cause of child death and poverty in Africa, United Nations Children Fund, [Online] http://www.unicef.org/publications/index_19019.html. Accessed 9/6/2007.

UNICEF (2008): The State of the World's Children 2009: Maternal and Newborn Health, United Nations Children's Fund, New York.

UNICEF (2009): *The State of the World's Children 2009: Maternal and Newborn Health*, United Nations Children's Fund, New York. [Online] http://www.unicef.org/sowc09/report/report.php, accessed 21/02/2009.

Urdan, T. C. (2005): Statistics in Plain English, 2nd Edition, Lawrence Erlbaum.

Vickers, D., Rees, P. and Birkin, M. (2005): *Creating the National Classification of Census Output Areas: Data, Methods and Results*, Working Paper 05/3, School of Geography, University of Leeds, [Online] http://www.geog.leeds.ac.uk/papers/05-03/, accessed 04/11/2006.

Vickers, D. W. (2006): *Multi-level Integrated Classifications Based on the 2001 Census*, Unpublished PhD Thesis. School of Geography, University of Leeds.

Vickers, D. and Rees, P. (2006): Introducing the Area Classification of Output Areas, *Population Trends*, **125**, 15-29.

Vickers, D.W. and Rees, P. H. (2007): Creating the National Statistics 2001 Output Area Classification. *Journal of the Royal Statistical Society*, Series A, **170**, 379–403.

Vickers, D. (2010): England's Changing Social Geology, in Stillwell, J. & Norman, P. Eds, Spatial and Social Disparities, Springer.

Voas, D. and Williamson, P. (2001a): The Diversity of Diversity: A Critique of Geodemographic Classification, *Area*, **33**, 63-76.

Voas, D. and Williamson, P. (2001b): The Diversity of Diversity: A Critique of Geodemographic Classification, *Area*, **33**, 335-336.

Wackernagel, M. and Rees, W. (1996): *Our Ecological Footprint: Reducing Human Impact on the Earth*, New Society Publishers, Gabriola Island.

Wagstaff, A. (2001): What Do Poor Children Die From? Some Evidence from Cebu, The Philippines, The World Bank and University of Sussex [Online] http://siteresources.worldbank.org/INTPAH/Resources/Publications/Recent-Papers/9218_poorchildren.pdf accessed 28/11/2009.

Walker, J. T., Maddan, S. (2008): Statistics in Criminology and Criminal Justice: Analysis and Interpretation, Jones and Bartlett.

Webber, R. (1977): *An Introduction to the National Classification of Wards and Parishes*, Planning Research Applications Group Technical Paper No. 23, London, Centre for Environmental Studies.

Webber, R. (2004a): *The Relative Power of Geodermographics Vis a Vis Person and Household Demographic Variables as Discriminators of Consumer Behaviour* [Online] http://www.casa.ucl.ac.uk/working_papers/paper84.pdfl, accessed 02/06/06.

Webber, R. (2004b): *Neighbourhood Inequalities in the Patterns of Admissions and Their Application to the Targeting of Health Promotion Campaigns*, [Online] http://www.casa.ucl.ac.uk/working_papers/paper90.pdf, accessed 23/2/2007.

Webber, R. (2005): *The 2005 British General Election Campaign: How effectively did the parties apply segmentation to their Direct Marketing strategies*? Paper Presented at the Joint seminar in ESRC/NERC/EPSRC Transdisciplinary Series and ESRC Research Methods Seminar Series, Center for Advanced Spatial Analysis, London.

Webber, R. and Craig, J. (1976): Which Local Authorities are alike? *Population Trends*, **5**, 13-19.

Webber, R. and Craig, J. (1978): Socioeconomic Classifications of Local Authority Areas, *Studies in Medical and Population Subjects*, **35**, London, OPCS.
Weeks, A. D., Grant, M. (2003): Closing the Digital Divide, *British Medical Journal*, **326**, 1035-1035.

Wolfgang, L., Warren, C. S., Sergei, S. (2004): *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*, Earthscan.

Wood J., Hennell T., Jones A., Hooper J., Tocque K., Bellis M.A. (2006): *Where Wealth means Health: Illustrating Inequality in the North West.* North West Public Health Observatory. ISBN: 1-902051-77-7.

Woods, D. L. (2006): *The Philippines: A Global Studies Handbook ABC-CLIO*, California.

Worcester, D. C. (2005): *The Philippine Islands and their People*, Adamant Media Corporation.

World Commission on Environment and Development (1987): *Our Common Future*, [Online] http://worldinbalance.net/agreements/1987-brundtland.php, accessed 14/03/2009.

WHO (2005): *The World Health Report: Make Every Mother and Child Count*, [Online] http://www.who.int/whr/2005/whr2005_en.pdf, accessed 15/05/2009.

WHO (2008a): Measles, Fact Sheet Number 286, [Online] http://www.who.int/mediacentre/factsheets/fs286/en/index.html, accessed 12/04/2009.

WHO (2008b): *Why is Giving Special Attention to Adolescents Important for Achieving Millennium Development Goal 5?* [Online] http://www.who.int/making_pregnancy_safer/events/2008/mdg5/adolescent_preg.pdf, accessed 01/08/2009.

WHO (2009a): *Contraceptive Prevalence Rate*, [Online] http://www.who.int/whosis/indicators/compendium/2008/3pcf/en/index.html, accessed 1/06/2009.

WHO (2009b): *Antenatal Care Coverage*, [Online] http://www.who.int/whosis/indicators/compendium/2008/3acf/en/index.html, accessed 03/07/2009.

WHO (2009c): *Major Causes of Death*: A Primer, [Online] http://www.who.int/mediacentre/factsheets/fs310/en/index2.html, accessed 09/06/2009.

WHO (2009d): *World Health Statistics 2009*, World Health Organisation, Geneva [Online] http://www.who.int/whosis/whostat/2009/en/index.html, accessed 13/10/2009.

World Bank (2005): *World Development Report*, World Bank, Washington DC.

World Bank (2007a). *Independent Evaluation Group Approach Paper Nigeria: Country Assistance Evaluation* [Online] http://lnweb18.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/3EE7F2E5A37A9BEC85257321007A7697/$file/nigeria_cae_approach_paper.pdf. Accessed 7/05/2008.

World Bank (2007b). *World Development Indicators 2007*, CD-ROM, Washington, D. C.

World Bank (2007c): Philippines Third Elementary Education Project, [Online] http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2007/05/04/000020953_20070504144634/Rendered/PDF/ICR204.pdf, accessed 17/01/09.

World Bank (2008): *World Development Indicators*, World Bank, Washington DC.

Wu, Y. (2004): *Understanding International Food Consumption Patterns*, [Online] http://www.biz.uwa.edu.au/home/research/discussionworking_papers/economics/2004?f=150995, accessed 15/12/08.

Zaninotto, P., Wardle, H., Stamatakis, E., Mindell, J., Head, J. (2006): *Forecasting Obesity to 2010*, United Kingdom Department of Health, [Online] http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_4138630, accessed 10/09/08.

Zhao, Q., Hautamaki, V. and Fanti, P (2008): *Knee Point Detection in BIC for Detecting the Number of Clusters*, [Online] http://cs.joensuu.fi/~zhao/PAPER/BIC_ACIVS.pdf, accessed 03/05/2010.

# Appendix A

# Further Results from the Nigerian Analysis

## Appendix A1: Decisions Made for the Nigerian Variables Considered

| Domain | Variable | Decision | Reason | New name |
|---|---|---|---|---|
| A1 | % of households using agricultural inputs | Include | Satisfies all tests and can help differentiate between areas where modern methods of agriculture are used. | |
| A2 | % of households owning 1 cattle | Exclude | Sample size is too small and no variation across areas. | |
| A3 | % of households owning 2-10 cattle | Merge with A4, A5 and A6 | Need to increase sample size. | Owns cattle |
| A4 | % of households owning 11-20 cattle | Merge with A3, A5 and A6 | High correlation and need to increase sample size. | Owns cattle |
| A5 | % of households owning 21-50 cattle | Merge with A3,A4 and A6 | High correlation and need to increase sample size. | Owns cattle |
| A6 | % of households owning over 50 cattle | Merge with A3, A4 and A5 | Need to increase sample size. | Owns cattle |
| D1 | % of males aged 0-14 | Exclude | Redundant information in D7 and D10. | |
| D2 | % of males aged 15-59 | Exclude | Redundant information in D8 and D10. | |
| D3 | % of males aged over 60 | Exclude | High Positive skew and redundant information in D6 and D9. | |
| D4 | % of females aged 0-14 | Exclude | Redundant information in D7, D8 and D10. | |
| D5 | % of females aged 15-59 | Exclude | Redundant information in D1, D7, D8 and D10. | |
| D6 | % of females aged over 60 | Exclude | Redundant information in D3 and D9. | |
| D7 | % of total population aged 0-14 | Include | Large sample size and informs about students and economically inactive. | |
| D8 | % of total population aged 15-59 | Include | Large sample size and informs about economically active. | |
| D9 | % of total population aged 60 and above | Include | Large sample size and information about the aged. | |
| D10 | Dependency ratio | Exclude | Redundant information in D7 and D8. | |
| D11 | % of the total population never married | Include | High variation, large sample size and indicative of areas of low marriages which can also inform male income. | |
| D12 | % of the total population divorced | Merge with D13 | Similar information and the same base. | Separated couple |
| D13 | % of the total population separated | Merge with D12 | Similar information and the same base. | Separated couple |
| D14 | % of households with a member receiving a pension | Include | Indicative of people receiving a pension. | |
| D15 | Population Density | Include | urban-rural indicator. | |
| E1 | % of the total population uneducated | Include | High variation, large sample size and indicative of areas of low literacy levels. | |
| E2 | % of the total population who completed primary education | Exclude | Redundant information in E1. | |
| E3 | % of the total population who completed secondary education | Include | Large sample size. Variable exerts strong power based on principal component 1.Important indicator. | |
| E4 | % of the total population with post secondary education | Exclude | Strong correllation with E3, S22, HI7 and HO2 | |
| E5 | % of the total population with head of household uneducated | Include | High variation and large sample size. Good indicator | |

| | | | | |
|---|---|---|---|---|
| E6 | Adult literacy rate | Include | Indicative of areas where older people still crave for knowledge | |
| E7 | % of households which spend less than 15 minutes to nearest primary school | Merge with E8 | Defines access | Access to primary school |
| E8 | % of households which spend less than 15-29 minutes to nearest primary school | Merge with E7 | Defines access | Access to primary school |
| E9 | % of households which spend less than 15 minutes to nearest secondary school | Exclude | Strong correlation with access to health and suggests the same information as completed secondary school | Access to secondary school |
| E10 | % of households which spend less than 15-29 minutes to nearest secondary school | Exclude | Strong corellation with access to health and suggests the same information as completed secondary school | Access to secondary school |
| EM1 | % of the total population in public sector employment | Exclude | Low variation between areas | |
| EM2 | % of the total population in private formal employment | Exclude | Low variation between areas | |
| EM3 | % of the total population in private informal employment | Exclude | Redundant information in E8 | |
| EM4 | % of the total population self employed in agriculture | Exclude | Redundant information in EM9 | |
| EM5 | % of the total population self employed in other sector | Exclude | Vague information | |
| EM6 | % of economically active population | Include | Indicative of man-power | |
| EM7 | % of unemployed population | Exclude | Low variation between areas | |
| EM8 | % of self employed population | Include | Good variation between areas | |
| EM9 | % population employed in agriculture | Merge with EM10 | Similar information and the same base | Employment in agriculture |
| EM10 | % population employed in the fishing industry | Merge with EM9 | Need to increase sample size | Employment in agriculture |
| EM11 | % population employed in the manufacturing sector | Exclude | Too specific and poor distinction between areas | |
| EM12 | % population employed in the construction sector | Exclude | Too specific and poor distinction between areas | |
| EM13 | % population employed in trade | Exclude | Redundant information in EM9 | |
| EM14 | % population employed in the transport sector | Include | Indicative of level of transportation | |
| EM15 | % population employed in public administration | Exclude | Too specific and poor distinction between areas | |
| EM16 | % population employed in education | Exclude | High positive skew | |
| EM17 | % population employed in health and social work | Exclude | Too specific and poor distinction between areas | |
| EM18 | % population employed in the services sector | Exclude | Too specific and poor distinction between areas | |
| H1 | % of households which spend less than 15 minutes to health facility | Merge with H2 | Defines access | Access to health |

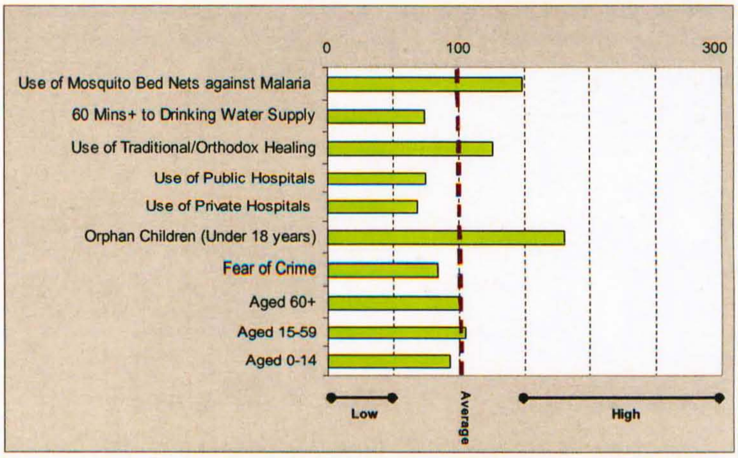
| | | | | |
|---|---|---|---|---|
| H2 | % of households which spend less than 15-29minutes to health facility | Merge with H1 | Defines access | Access to health |
| H3 | % of households taking anti-malaria measures | Include | Good variation between areas and important indicator for the fight against malaria | |
| HC1 | % of households with 1-2 persons | Exclude | Redundant information in HC5 | |
| HC2 | % of households with 3-4 persons | Exclude | Poor variation between areas | |
| HC3 | % of households with 5-6 persons | Exclude | Poor variation between areas | |
| HC4 | % of households with 7 persons and above | Exclude | Redundant information in HC5 | |
| HC5 | Mean household size | Include | An indicator of the extent of crowding | |
| HI1 | % of households owning a fixed line telephone | Exclude | Poor variation between areas, small sample size and high positive skew | |
| HI2 | % of households owning mobile phones | Include | Indicative of access to telecommunications | |
| HI3 | % of households owning a personal computers | Include | useful indicator of the e-society | |
| HI4 | % of households with access to safe source of drinking water | Include | Important basic amenity indicator | |
| HI5 | % of households without toilet facility | Include | An indicator which can inform health conditions | |
| HI6 | % of households with access to safe toilet sanitation | Include | Differentiates between access to sanitation and access to safe sanitation | |
| HI7 | % of households which use non-wood fuel for cooking | Include | High variation across dataset. Variable exerts strong power based on principal component 1 | |
| HI8 | % of households which use electricity for lighting | Include | Indicative of provision of electricity supply | |
| HI9 | % of households which spend less than 15 minutes to source of drinking water | Merge with HI10 | Defines access | Access to water supply |
| HI10 | % of households which spend less than 15-29 minutes to source of drinking water | Merge with HI9 | Defines access | Access to water supply |
| HI11 | % of households owning a generator | Exclude | Poor variation between areas | |
| HO1 | % of occupancy by home ownership | Include | Useful indicator for measuring housing demand | |
| HO2 | % of occupancy by rent | Include | Indicative of poorer individuals | |
| HO3 | % of subsidized occupancystatus | Exclude | Poor distinction between areas and very small sample sizes | |
| HO4 | % of free occupancy status | Include | Large variation between areas | |
| HO5 | % of households built with mud/mud bricks | Merge with HO8 | High level of correlation | |
| HO6 | % of households built with stone | Exclude | Poor distinction between areas and very small sample sizes | |
| HO7 | % of households built with burnt bricks | Include | Large variation between areas | |
| HO8 | % of households built with cement/sandcrete | Merge with HO5 | High level of correlation | Built with cement/mud |
| HO9 | % of households built with wood/bamboo | Exclude | Poor distinction between areas and very small sample sizes | |

| | | | | |
|---|---|---|---|---|
| HO10 | % of households built with iron sheets | Exclude | Poor distinction between areas and very small sample sizes | |
| HO11 | % of households built with cardboard | Exclude | Poor distinction between areas and very small sample sizes | |
| HO12 | % of single room housing unit | Include | Good distinction between areas and indicative of areas where there are unmarried individuals | |
| HO13 | % of flats | Exclude | Creating a regional divide and masking underlying discriminatory features of other clusters | |
| HO14 | % of duplexes | Include | Describes a building type | |
| HO15 | % of whole buildings | Exclude | Redundant information in HO12 | |
| S1 | % of households which seldom find it difficult to satisfy food needs | Exclude | Information covered in S4 | |
| S2 | % of households which sometimes find it difficult to satisfy food needs | Exclude | Information covered in S5 | |
| S3 | % of households which often find it difficult to satisfy food needs | Exclude | Information covered in S6 | |
| S4 | % of households which always find it difficult to satisfy food needs | Merge with S8, S12, S16 and S20 | High levels of correlation and indicative of chronic socio-economic situations | Difficulty in meeting basic needs |
| S5 | % of households which seldom find it difficult to pay school fees | Exclude | Information covered in S8 | |
| S6 | % of households which sometimes find it difficult to pay school fees | Exclude | Information covered in S9 | |
| S7 | % of households which often find it difficult to pay school fees | Exclude | Information covered in S10 | |
| S8 | % of households which always find it difficult to pay school fees | Merge with S4, S12, S16 and S20 | High levels of correlation and indicative of chronic socio-economic situations | Difficulty in meeting basic needs |
| S9 | % of households which seldom find it difficult to pay house rent | Exclude | Information covered in S12 | |
| S10 | % of households which sometimes find it difficult to pay house rent | Exclude | Information covered in S13 | |
| S11 | % of households which often find it difficult to pay house rent | Exclude | Information covered in S14 | |
| S12 | % of households which always find it difficult to pay house rent | Merge with S4, S8, S16 and S20 | High levels of correlation and indicative of chronic socio-economic situations | Difficulty in meeting basic needs |
| S13 | % of households which seldom find it difficult to pay utility bills | Exclude | Information covered in S16 | |
| S14 | % of households which sometimes find it difficult to pay utility bills | Exclude | Information covered in S17 | |
| S15 | % of households which often find it difficult to pay utility bills | Exclude | Information covered in S18 | |

| | | | | |
|---|---|---|---|---|
| S16 | % of households which always find it difficult to pay utility bills | Merge with S4, S8, S12 and S20 | High levels of correlation and indicative of chronic socio-economic situations | Difficulty in meeting basic needs |
| S17 | % of households which seldom find it difficult to pay for health care | Exclude | Information covered in S20 | |
| S18 | % of households which sometimes find it difficult to pay for health care | Exclude | Information covered in S21 | |
| S19 | % of households which often find it difficult to pay for health care | Exclude | Information covered in S22 | |
| S20 | % of households which always find it difficult to pay for health care | Merge with S4, S8, S12 and S16 | High levels of correlation and indicative of chronic socio-economic situations | Difficulty in meeting basic needs |
| S21 | % of households owning vehicles | Include | Good indicator of wealth | |
| S22 | % of households owning motorcycles | Include | In Nigeria, this variable is indicative of a type of public transport employment | |
| S23 | % of households owning less than 1 ha of land | Exclude | Low variation across areas | |
| S24 | % of households owning 1-1.99 ha of land | Exclude | Low variation across areas | |
| S25 | % of households owning 2-3.99 ha of land | Exclude | Low variation across areas | |
| S26 | % of households owning 4-5.99 ha of land | Exclude | Low variation across areas | |
| S27 | % of households owning over 6 ha of land | Include | Good indicator of wealth. Has large variation as well | |
| S28 | % of households which perceive crime level to be much worse | Exclude | Pseudo defined by S31 and S32 | |
| S29 | % of households which perceive crime level to be worse | Exclude | Pseudo defined by S31 and S33 | |
| S30 | % of households which perceive crime level to be the same | Exclude | Pseudo defined by S31 and S34 | |
| S31 | % of households which perceive crime level to be better | Merge with S32 | Indicative of improved security | Improved security |
| S32 | % of households which perceive crime level to be much better | Merge with S31 | Indicative of improved security | Improved security |
| S33 | % of households which spend 30-59 minutes to nearest food market | Exclude | Highly correlated with other access variables especially transportation | |
| S34 | % of households which spend over 60 minutes to nearest food market | Exclude | Highly correlated with other access variables especially transportation | |
| S35 | % of households which spend less than 15 minutes to nearest public transport | Merge with S36 | Defines access | Access to transportation |
| S36 | % of households which spend 15-29 minutes to nearest public transport | Merge with S35 | Defines access | Access to transportation |
| WC1 | % of children under 18 years | Merge with | An indicator of families that may be in need | Children living |

| | | | | |
|---|---|---|---|---|
| | living with their mother only | WC2 | of support | with single parents |
| WC2 | % of children under 18 years living with their father only | Merge with WC4 | An indicator of families that may be in need of support | Children living with single parents |
| WC3 | % of children not in school due to teenage pregnancy | Merge with WC1 | Need to increase sample size and variation | Early marriage and teenage pregnancy |
| WC4 | % of children not in school due to early marriage living with their father only | Merge with WC3 | Indicative of areas where female education is discouraged | Early marriage and teenage pregnancy |
| WC5 | % of vaccinated children | Include | A useful indicator of health and mortality in children | |
| WC6 | % of children under 5 not breastfed | Exclude | Small sample size with little variation across areas | |

## Appendix A2: Pen Portrait for Underprivileged Green Towns




**3.5% of national population**
**4.0% of national households**

Underprivileged Green Towns can be found in 7 states. These are Akwa-Ibom, Bayelsa, Cross River, Delta, Edo, Ondo and Rivers state. Delta state has more areas classified as Underprivileged Green Towns but more significantly, Rivers state has the greatest population (24.75% of its population) defined by this Group.

The average household size is 4.1 persons. Population density is also below the national benchmark at an average of 393 people per $Km^2$.

This Group comprises a total of 29 LGAs.

Underprivileged Green Towns are made up of very large number of households that use fingerlings, hooks and nets for their agricultural livelihoods. Indeed well over nine times the national average number of households makes use of hooks and nets. A substantial number of households secure their agricultural inputs from co-operative societies.

There are large concentrations of widowed persons within these areas and a lot of unmarried people with a high proportion of single parents. These areas also have a very high concentration of separated couples. Just like Conventional Green Towns, Underprivileged Green Towns are characterised mainly by households of 1 to 2 persons. There is also an above average representation of households of 3 to 4 persons in size.

The patterns of the highest levels of education reflect that secondary education is quite dominant within these areas. However, it is also note-worthy to mention that large numbers of people also have primary education as their highest level of education. Although secondary school completion and adult literacy rates are relatively high, access to primary schools is below the national average. A lot of students are unhappy with the lack of teachers and the high fees they have to pay. As a result a high proportion of school aged children are not in school due to these costs.

The engagement in public sector employment within these areas is just above the national average with a relatively high number of people considering themselves as under-employed. There are also a considerably high number of people in the unemployed category. Indeed amongst all Green Towns, the Underprivileged Green Towns have the largest proportion of unemployed persons.

With below average levels of access to health facilities, the dissatisfaction with health services is also significantly high within these areas. The absence of trained medical professionals and unsuccessful medical treatments constitute the major reasons for dissatisfaction. People are also generally unhappy with the lack of drugs and cost of treatment which they consider too high. A very large number of people use community health centres, traditional healers and in some cases pharmacists or chemists. The use of bed nets and anti-malaria drugs as measures against malaria is a characteristic feature of residents of these areas. Teenage pregnancy rates are high.

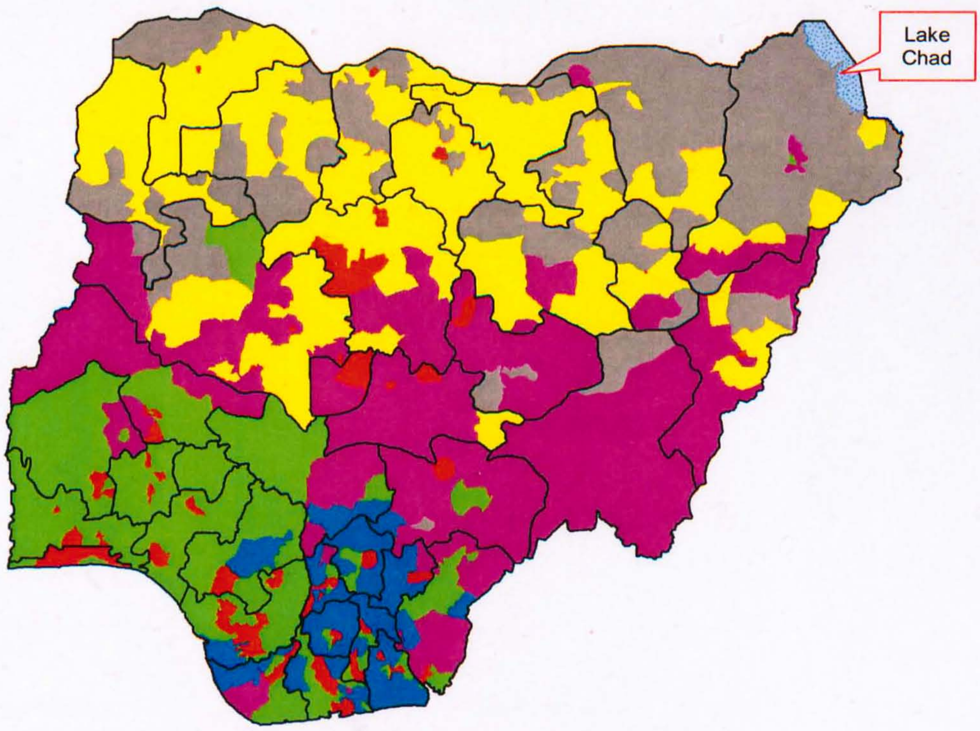
A substantial number of households tend to have access to untreated pipe-borne water but very many also derive their drinking water from unprotected wells, rain, rivers and ponds. Pails and buckets are the major types of toilet facilities and kerosene is the most common form of cooking fuel.

Households built with iron roofing sheets are common within Underprivileged Green Towns. Many people are in rented accommodations or leasehold but the concentration of people not paying any rent is much higher than the national average. Households without any form of occupancy document are many.

Amongst all Groups within Green Towns, this Group comprises people who find it difficult meeting basic day to day need. They find it most difficult paying for school fees and health care. Pensioners are not paid regularly and people tend to spend over one hour to their nearest food market. Vehicle ownership is low but ownership of motorcycle is quite high. A good number of people are economically active.

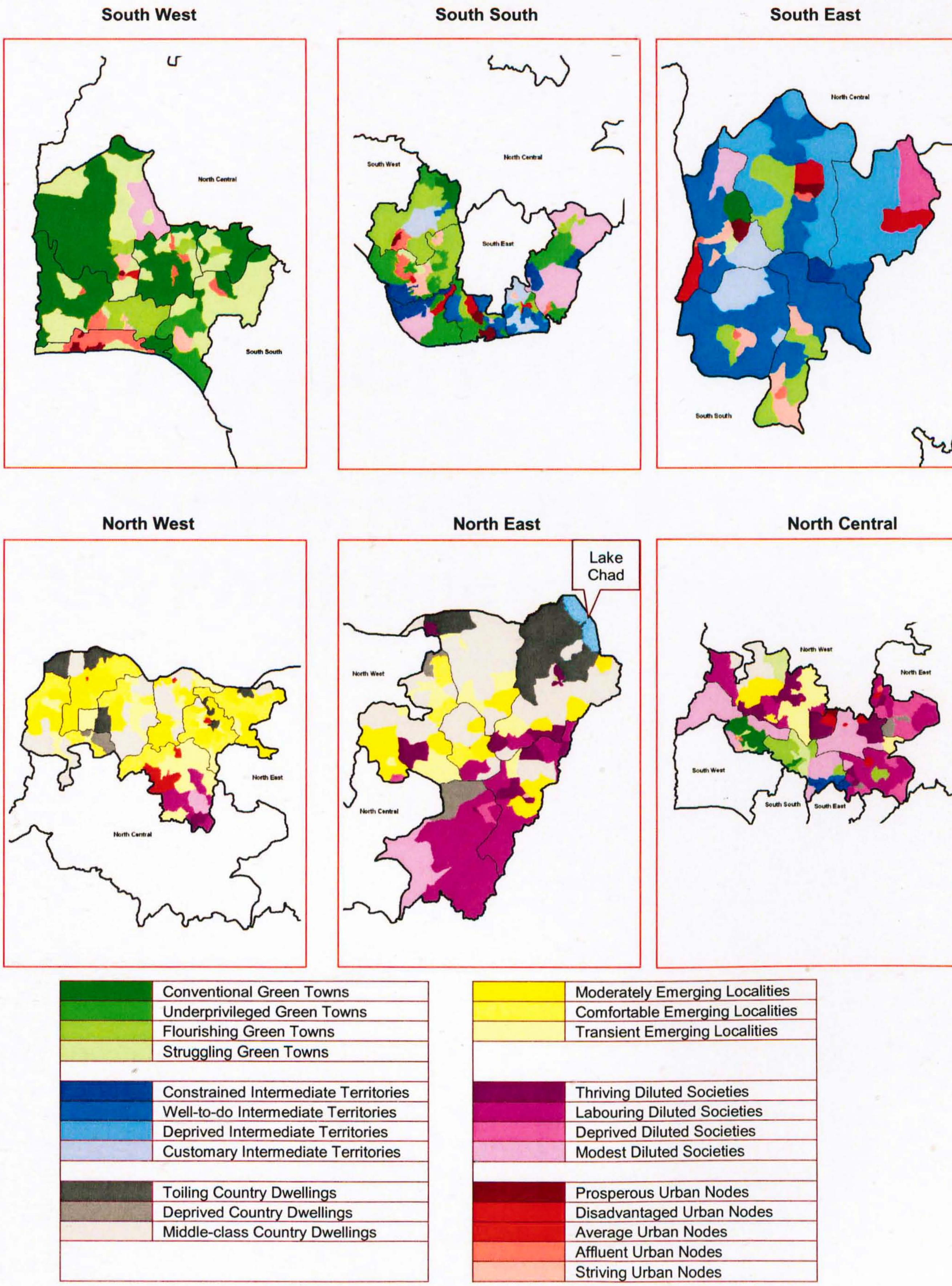
There is a slightly above average representation of women who receive some form of pre-natal care. Some children are delivered in hospitals by nurses or midwives but a majority of children are delivered by trained and untrained traditional birth attendants. Infant vaccinations are not high but just a little above the national average.

## Appendix A3: Visualising the Six Super-groups




| | |
|---|---|
| | Green Towns |
| | Emerging Localities |
| | Intermediate Territories |
| | Diluted Societies |
| | Country Dwellings |
| | Urban Nodes |

## Appendix A4: Visualising the Twenty-three Groups



# Appendix B

## Further Results from the Philippines Analysis

## Appendix B1: Decisions Made for the Philippines Variables Considered

| Variable Code | Short Name | In, Out or Merged | New Name | Reason |
|---|---|---|---|---|
| D1 | Part of city or poblacion? | | | Urban-rural indicator |
| D2 | Population density | | | Indicative of population disbursement across areas |
| D3 | No married couples | | | Very good variance across space. The living arrangements and demands are also important. |
| D4 | One married couple | | | High negative correlation with D129 |
| D5 | Two married couples | | **Two or more married couples** | These variables were merged to increase their sample sizes. In the absence of information on the sizes of households, this variable serves as a proxy indicator. |
| D6 | Three married couples | | | |
| D7 | Four married couples | | | |
| D8 | Five married couples | | | |
| D9 | Six married couples | | | |
| D10 | Head of household | | | Vague information with very low variation. |
| D11 | Spouse of household head | | | Vague information with very low variation. |
| D12 | Child of household head | | | Vague information with very low variation. |
| D13 | Stepchild of household head | | | Variable is relevant to marital harmony. |
| D14 | Grandchild of household head | | | Strong positive correlation with D5, D16 and D129. |
| D15 | Parent of household head | | | Low power with low variance |
| D16 | Child-in-law of household head | | | Strong correlation with D5, D6 and D14. |
| D17 | Sibling of household head | | | Vague information with low power and low variation. |
| D18 | Aunt or uncle of household head | | | Very low power and high positive skew. |
| D19 | Nephew or niece of household head | | | Very low power and high positive skew. |
| D20 | Other relative of household head | | | Very low power and high positive skew. Adds no new information. |
| D21 | Not related to household head | | | High skew. |
| D22 | Age 0 | | **Age 0 to 4** | Indicative of infants and younger children and relevant to MDG 4. |
| D23 | Age 1 | | | |
| D24 | Age 2 | | | |
| D25 | Age 3 | | | |
| D26 | Age 4 | | | |
| D27 | Age 5 | | **Age 5 to 9** | Indicative of older children and also relevant for assessing MDG 2 and MDG4 |
| D28 | Age 6 | | | |
| D29 | Age 7 | | | |
| D30 | Age 8 | | | |
| D31 | Age 9 | | | |
| D32 | Age 10 | | | Secondary school aged children and teenagers. |
| D33 | Age 11 | | | |

| | | | | |
|---|---|---|---|---|
| D34 | Age 12 | | **Age 10 to 19** | |
| D35 | Age 13 | | | |
| D36 | Age 14 | | | |
| D37 | Age 15 | | | |
| D38 | Age 16 | | | |
| D39 | Age 17 | | | |
| D40 | Age 18 | | | |
| D41 | Age 19 | | | |
| D42 | Age 20 | | **Age 20 to 44** | Younger adults. Important for monitoring people who are at the early and middle stages of their careers. It also provides information on younger families |
| D43 | Age 21 | | | |
| D44 | Age 22 | | | |
| D45 | Age 23 | | | |
| D46 | Age 24 | | | |
| D47 | Age 25 | | | |
| D48 | Age 26 | | | |
| D49 | Age 27 | | | |
| D50 | Age 28 | | | |
| D51 | Age 29 | | | |
| D52 | Age 30 | | | |
| D53 | Age 31 | | | |
| D54 | Age 32 | | | |
| D55 | Age 33 | | | |
| D56 | Age 34 | | | |
| D57 | Age 35 | | | |
| D58 | Age 36 | | | |
| D59 | Age 37 | | | |
| D60 | Age 38 | | | |
| D61 | Age 39 | | | |
| D62 | Age 40 | | | |
| D63 | Age 41 | | | |
| D64 | Age 42 | | | |
| D65 | Age 43 | | | |
| D66 | Age 44 | | | |
| D67 | Age 45 | | **Age 45 to 64** | This variable is an important measure for understanding people nearing retirement age. |
| D68 | Age 46 | | | |
| D69 | Age 47 | | | |
| D70 | Age 48 | | | |
| D71 | Age 49 | | | |
| D72 | Age 50 | | | |
| D73 | Age 51 | | | |
| D74 | Age 52 | | | |
| D75 | Age 53 | | | |
| D76 | Age 54 | | | |
| D77 | Age 55 | | | |
| D78 | Age 56 | | | |
| D79 | Age 57 | | | |
| D80 | Age 58 | | | |
| D81 | Age 59 | | | |
| D82 | Age 60 | | | |
| D83 | Age 61 | | | |
| D84 | Age 62 | | | |

| | | | | |
|---|---|---|---|---|
| D85 | Age 63 | | | |
| D86 | Age 64 | | | |
| D87 | Age 65 | | Age 65+ | This group of people are essential for planning as they constitute a major chunk of economically inactive population. |
| D88 | Age 66 | | | |
| D89 | Age 67 | | | |
| D90 | Age 68 | | | |
| D91 | Age 69 | | | |
| D92 | Age 70 | | | |
| D93 | Age 71 | | | |
| D94 | Age 72 | | | |
| D95 | Age 73 | | | |
| D96 | Age 74 | | | |
| D97 | Age 75 | | | |
| D98 | Age 76 | | | |
| D99 | Age 77 | | | |
| D100 | Age 78 | | | |
| D101 | Age 79 | | | |
| D102 | Age 80 | | | |
| D103 | Age 81 | | | |
| D104 | Age 82 | | | |
| D105 | Age 83 | | | |
| D106 | Age 84 | | | |
| D107 | Age 85 | | | |
| D108 | Age 86 | | | |
| D109 | Age 87 | | | |
| D110 | Age 88 | | | |
| D111 | Age 89 | | | |
| D112 | Age 90 | | | |
| D113 | Age 91 | | | |
| D114 | Age 92 | | | |
| D115 | Age 93 | | | |
| D116 | Age 94 | | | |
| D117 | Age 95 | | | |
| D118 | Age 96 | | | |
| D119 | Age 97 | | | |
| D120 | Age 98 | | | |
| D121 | Age 99 | | | |
| D122 | Age 100 | | | |
| D123 | Males | | | Vague information, very high negative skew |
| D124 | Females | | | Vague information, very high negative skew |
| D125 | Never married | | | Already included |
| D126 | Married | | | Already included as part of two or more married couples |
| D127 | Consensual union | | | Already included |
| D128 | Separated or divorced | | | Already included |
| D129 | Widowed | | | Very relevant to MDG 3. |
| D130 | Native born | | | Vague information and negative correlation with D131 |
| D131 | Foreign born | | | Very low power and high positive skew. |
| D132 | Unknown nativity | | | Vague information. |

| | | | | |
|---|---|---|---|---|
| D133 | Speaks Filipino | | | Relevant cultural information. |
| D134 | Does not speak Filipino | | | Negative correlation with D133 |
| E1 | Elementary school in Barangay? | | | Relevant to MDG 2. |
| E2 | High school in Barangay? | | | Also relevant to MDG 2 |
| E3 | College in Barangay? | | | Redundant information captured in E2 |
| E4 | Attending school | | | Very low power and negatively correlated with E5. |
| E5 | Not attending school | | | Very low power and negatively correlated with E4. |
| E6 | Illiterates | | | Negative correlation with E7, E11, E25 and E27. |
| E7 | Literates | | | Positive correlations with E11, E25 and E27. |
| E8 | Less than primary education completed | | | The indicator is important for monitoring the completion of basic education. |
| E9 | Some primary education completed | | | Strong negative correlation with E11, E12, E25, E27 and E29. |
| E10 | Primary-6years completed | | | Information already included |
| E11 | Secondary-general track completed | | | Strong correlation with HI21, HI23 and HI27. |
| E12 | Some college completed | | | Redundant information |
| E13 | Post secondary technical education | | | Information relevant to understanding the attainment of higher level education. |
| E14 | University completed | | | Redundant information |
| E15 | No year of schooling | | | Redundant information |
| E16 | 1 year of schooling | | | Redundant information |
| E17 | 2 year of schooling | | | Redundant information |
| E18 | 3 year of schooling | | | Redundant information |
| E19 | 4 year of schooling | | | Redundant information |
| E20 | 5 year of schooling | | | Redundant information |
| E21 | 6 year of schooling | | | Redundant information |
| E22 | 7 year of schooling | | | Redundant information |
| E23 | 8 year of schooling | | | Redundant information |
| E24 | 9 year of schooling | | | Redundant information |
| E25 | 10 year of schooling | | | Redundant information |
| E26 | 11 year of schooling | | | Redundant information |
| E27 | 12 year of schooling | | | Redundant information |
| E28 | 13 year of schooling | | | Redundant information |
| E29 | 14 year of schooling | | | Redundant information |
| E30 | Number of Barangays with an elementary school | | | Vague and redundant information |
| E31 | Number of Barangays with no elementary school | | | Vague and redundant information |
| E32 | Number of Barangays with a high school | | | Vague and redundant information |
| E33 | Number of Barangays with no high school | | | Vague and redundant information |
| E34 | Number of Barangays with a college/university | | | Vague and redundant information |
| E35 | Number of Barangays with no college/university | | | Vague and redundant information |
| EM1 | Legislators, managers and | | **Professionals and** | The indicator is important for |

| | | | | |
|---|---|---|---|---|
| | senior officials | | **senior officials** | understanding the concentration pattern of people in top employment cadres. |
| EM2 | Professionals | | | |
| EM3 | Technicians and associate professionals | | | Strong negative correlation with EM6 and EM11. |
| EM4 | Clerks | | | Strong positive correlation with EM1, EM2, EM3, EM8, EM16, EM17, EM18, EM19, EM21, EM23 and EM24. |
| EM5 | Service workers | | | Very low variation across areas. |
| EM6 | Agricultural and fishery workers | | | Employees within this sector are important for understanding employment in the informal sector. |
| EM7 | Craft workers | | | Strong positive correlation with EM8, EM13, EM15 and EM18. |
| EM8 | Plant and machine operators | | | Strong correlation with EM6 |
| EM9 | Elementary occupations | | | Vague definition |
| EM10 | Employed in the Armed forces | | | Redundant information |
| EM11 | Employed in the Agricultural industry | | | Redundant information |
| EM12 | Employed in the Mining industry | | | Low power and high positive skew |
| EM13 | Employed in the Manufacturing industry | | | Redundant information |
| EM14 | Employed in the Electricity gas and water industry | | | Very low variation across areas |
| EM15 | Employed in the Construction industry | | | Redundant information |
| EM16 | Employed in Wholesale and retail trade | | | Redundant information |
| EM17 | Employed in the Hotels and restaurants sector | | | Redundant information |
| EM18 | Employed in the Transportation and communications industry | | | The variable satisfies all statistical tests. |
| EM19 | Employed in the Financial services and insurance sectors | | | Strong positive correlation with EM1, EM2, EM3, EM17, EM18, EM21, EM23 and EM24. |
| EM20 | Employed in Public administration and defense | | | The public sector is a major employment hub especially for the middle class. |
| EM21 | Employed in Real estate and business services | | | Redundant information |
| EM22 | Employed in the Education sector | | | Demonstrates good variation across space. |
| EM23 | Employed in the Health and social work industry | | | This variable has a very strong influence on the entire dataset. |
| EM24 | Employed in Community and personal services | | | Redundant information |
| EM25 | Employed in private household services | | | Low power and positive skew. |
| EM26 | Overseas worker | | | Important for understanding and planning for emigration. |
| EM27 | Non overseas worker | | | Pseudo included with EM26 |
| EM28 | Number of establishments | | | Vague information |
| EM29 | Total employment | | | Vague information |
| EM30 | Agriculture, Hunting and | | | Redundant information |

| | | | | |
|---|---|---|---|---|
| | Forestry establishments | | | |
| EM31 | Fishing establishments | | | Redundant information |
| EM32 | Mining and Quarrying establishments | | | Redundant information |
| EM33 | Manufacturing establishments | | | Redundant information |
| EM34 | Electricity, Gas and Water Supply establishments | | | Redundant information |
| EM35 | Construction establishments | | | Redundant information |
| EM36 | Wholesale and Retail Trade, Repair of Motor Vehicles, Motorcycles and Personal and Household Goods establishments | | | Redundant information |
| EM37 | Hotels and Restaurants establishments | | | Redundant information |
| EM38 | Transport, Storage and Communication establishments | | | Redundant information |
| EM39 | Financial Intermediation establishments | | | Redundant information |
| EM40 | Real Estate, Renting and Business Activities establishments | | | Satisfies all relevant tests |
| EM41 | Education establishments | | | Redundant information |
| EM42 | Health and Social Work establishments | | | Redundant information |
| EM43 | Other Community, Social and Personal Service Activities establishments | | | Redundant information |
| EM44 | Single Proprietorship establishments | | | Redundant information |
| EM45 | Partnership establishments | | | Redundant information |
| EM46 | Government Corporation establishments | | | Redundant information |
| EM47 | Private Corporation establishments | | | Important variable for understanding the growth and development of the private sector |
| EM48 | Cooperative establishments | | | Cooperative establishments support small scale enterprises |
| EM49 | Others such as Foundation, NGOs, Association establishments | | | Vague information |
| EM50 | Single Establishment establishments | | | High skew with very low power. |
| EM51 | Branch establishments | | | Vague information |
| EM52 | Establishment and Main Office and with branch/elsewhere | | | Vague information |
| EM53 | Main Office Only establishments | | | Vague information |
| EM54 | Ancillary Unit establishments | | | Vague information |
| EM55 | Small establishments | | | Very good variation across areas. Also indicative of the dispersion of small scale entrepreneurs. |
| EM56 | Medium establishments | | | Large positive skew. |
| EM57 | Large establishments | | | Large positive skew. |
| EM58 | Growing of crops | | | Redundant information and very |

| | | | | |
|---|---|---|---|---|
| | establishments | | | large positive skew. |
| EM59 | Farming of animals establishments | | | Redundant information and very large positive skew. |
| EM60 | Agricultural and animal husbandry service activities, except veterinary activities establishments | | | Redundant information and very large positive skew. |
| EM61 | Hunting, trapping and game propagation including related service activities establishments | | | Redundant information and very large positive skew. |
| EM62 | Forestry, logging and related service activities establishments | | | Redundant information and very large positive skew. |
| EM63 | Metallic ore mining establishments | | | Redundant information and very large positive skew. |
| EM64 | Non-metallic mining and quarrying establishments | | | Redundant information and very large positive skew. |
| EM65 | Manufacture of food products and beverages establishments | | | Redundant information |
| EM66 | Manufacture of tobacco products establishments | | | Large positive skew and redundant information |
| EM67 | Manufacture of textiles establishments | | | Large positive skew and redundant information |
| EM68 | Manufacture of wearing apparel establishments | | | Large positive skew and redundant information |
| EM69 | Tanning and dressing of leather, manufacture luggage, handbags and footwear establishments | | | Large positive skew and redundant information |
| EM70 | Manufacture of wood and wood products establishments | | | Large positive skew and redundant information |
| EM71 | Integrated paper and paper products establishments | | | Large positive skew and redundant information |
| EM72 | Publishing, printing and reproduction of recorded media establishments | | | Satisfies all tests and demonstrates very good variation across areas. |
| EM73 | Manufacture of coke, refined petroleum and other fuel products establishments | | | Redundant information |
| EM74 | Manufacture of chemicals and chemical products establishments | | | Redundant information |
| EM75 | Manufacture of rubber and plastic products establishments | | | Redundant information |
| EM76 | Manufacture of other non-metallic mineral products establishments | | | Redundant information |
| EM77 | Manufacture of basic metals establishments | | | Redundant information |
| EM78 | Manufacture of fabricated metal products, except machinery and equipment establishments | | | Redundant information |
| EM79 | Manufacture of machinery and equipment, not | | | Redundant information |

| | | | | |
|---|---|---|---|---|
| | elsewhere classified establishments | | | |
| EM80 | Manufacture of office, accounting and computing machinery establishments | | | Redundant information |
| EM81 | Manufacture of electrical machinery and apparatus, not elsewhere classified establishments | | | Redundant information |
| EM82 | Manufacture of radio, television, equipments and apparatus establishments | | | Redundant information |
| EM83 | Manufacture of medical, precision and optical instruments, watches and clocks establishments | | | Redundant information |
| EM84 | Manufacture of motor vehicles, trailers and semi-trailers establishments | | | Redundant information |
| EM85 | Manufacture of other transport equipment establishments | | | Redundant information |
| EM86 | Manufacture and repair of furniture establishments | | | Redundant information |
| EM87 | Recycling establishments | | | Redundant information |
| EM88 | Other Manufacturing establishments | | | Redundant information |
| EM89 | Electricity, gas, steam and hot water supply establishments | | | Redundant information |
| EM90 | Water related establishments | | | Redundant information |
| EM91 | Automotive related establishments | | | Redundant information |
| EM92 | Wholesale trade establishments | | | Redundant information |
| EM93 | Retail trade establishments | | | Satisfies all relevant tests and relevant for measuring economic activity. |
| EM94 | Land transport; transport via pipelines establishments | | | Very low power and low variation. |
| EM95 | Water transport establishments | | | Very low power and low variation. |
| EM96 | Air transport establishments | | | Very low power and low variation. |
| EM97 | Auxiliary transport establishments | | | Very low power and low variation. |
| EM98 | Post and telecommunications establishments | | | Very low power and low variation. |
| EM99 | Banking institutions | | | Important economic indicator. |
| EM100 | Non-bank financial intermediation establishments | | | Redundant information |
| EM101 | Insurance and pension funding establishments | | | Redundant information |
| EM102 | Financial intermediation establishments | | | Redundant information |
| EM103 | Real estate establishments | | | Redundant information |

| | | | | |
|---|---|---|---|---|
| EM104 | Renting of machinery establishments | | | Redundant information |
| EM105 | Computer related establishments | | | Vague information |
| EM106 | Research and development establishments | | | Low power, weak variation and large positive skew. |
| EM107 | Miscellaneous business activities | | | Vague definition |
| EM108 | Sewage and refuse disposal establishments | | | Redundant information |
| EM109 | Recreational and cultural establishments | | | Very strong variation across areas. |
| EM110 | Other services establishments | | | Vague definition |
| H1 | Hospital | | | Relevant for understanding the relationship between socio-economic status and health service providers |
| H2 | Health centre | | | Relevant for understanding the relationship between socio-economic status and health service providers |
| H3 | Disabled | | | Important for planning for the less able population groups. |
| H4 | Not disabled | | | Redundant information. |
| H5 | Vision impaired | | | Redundant information. |
| H6 | Not vision impaired | | | Redundant information. |
| H7 | Hearing impaired | | | Redundant information. |
| H8 | Not hearing impaired | | | Redundant information. |
| H9 | Mute | | | Redundant information. |
| H10 | Not Mute | | | Redundant information. |
| H11 | Mental disability | | | Redundant information. |
| H12 | Not mentally disabled | | | Redundant information. |
| H13 | Psychologically disabled | | | Redundant information. |
| H14 | Not psychologically disabled | | | Redundant information. |
| H15 | Number of barangays with a hospital in the barangay | | | Redundant information. |
| H16 | Number of barangays with no hospital in the barangay | | | Redundant information. |
| H17 | Number of barangays with a health centre in the barangay | | | Redundant information. |
| H18 | Number of barangays with no health centre in the barangay | | | Redundant information. |
| HI1 | Electricity supplied | | | Indicative of welfare and the provision of fundamental infrastructure. |
| HI2 | No electricity | | | Redundant information |
| HI3 | Piped water exclusive to household | | | Strong positive correlation with HI13, HI14 and HI21. |
| HI4 | Piped water shared with other households | | | Strong negative correlation with HI5. |
| HI5 | No piped water | | | Important indicator for MDG 7 |
| HI6 | Cooking energy-electricity | | | Strong negative correlation with HI12. |

| | | | | |
|---|---|---|---|---|
| HI7 | Cooking energy-liquid fuels | | | Performs well across all tests. |
| HI8 | Cooking energy-liquefied petroleum gas | | | Strong positive correlation with H1and HI7 and strong negative correlation with HI5. |
| HI9 | Cooking energy-wood | | | Important indicator for MDG7 |
| HI10 | Cooking energy-coal | | | Low power and large skew. |
| HI11 | Cooking energy-other | | | Vague information. |
| HI12 | No telephone | | | Relevant variable for MDG 8. |
| HI13 | Telephone available | | | Redundant information |
| HI14 | Trash collected by a sanitation service | | | Redundant information |
| HI15 | Trash burned | | | Redundant information |
| HI16 | Trash buried | | | Redundant information |
| HI17 | Trash dumped in pit | | | Redundant information |
| HI18 | Trash fed to animals | | | Redundant information |
| HI19 | Trash composted | | | Redundant information |
| HI20 | Refrigerator unavailable | | | Redundant information |
| HI21 | Refrigerator available | | | Redundant information |
| HI22 | Television unavailable | | | Redundant information |
| HI23 | Television available | | | Important for socio-economic status. Satisfies all the tests and demonstrates very good variation across areas. |
| HI24 | Radio unavailable | | | Strong negative correlation with HI23 and HI25. |
| HI25 | Radio available | | | Strong positive correlation with HI23. |
| HI26 | No toilet | | | Redundant information |
| HI27 | Water closet toilet | | | Very relevant to MDG7 |
| HI28 | Latrine | | | Very relevant to MDG7 |
| HI29 | Other type of toilet | | | Vague information. |
| HO1 | Presence or absence of a housing project | | | Vague information and subject to rapid change. |
| HO2 | Dwelling owned | | | Very strong variation across space. |
| HO3 | Dwelling rented | | | Strong positive skew. |
| HO4 | Members squatting | | | Indicative of overcrowding. |
| HO5 | Free occupancy | | | Strong variation across most areas. |
| HO6 | Land is owned | | | Very weak power |
| HO7 | Land is leased | | | Very weak power. |
| HO8 | Land is occupied with consent | | | Important for development control. |
| HO9 | Land is occupied without consent | | | Redundant information |
| HO10 | Built with no walls | | | Very weak power. |
| HO11 | Built with improvised materials | | | Very weak power. |
| HO12 | Built with wood | | | Very weak power. |
| HO13 | Built with bamboo | | | Strong variation and power. |
| HO14 | Built with bricks, stone or concrete | | | Strong variation and power. |
| HO15 | Built with asbestos | | | Weak power and large skew. |
| HO16 | Built with galvanized iron | | | Redundant information. |
| HO17 | Built with glass | | | Large skew. |
| HO18 | Built with mixed materials | | | Good variation across areas. |
| HO19 | Built with other materials | | | Very low power and pseudo |

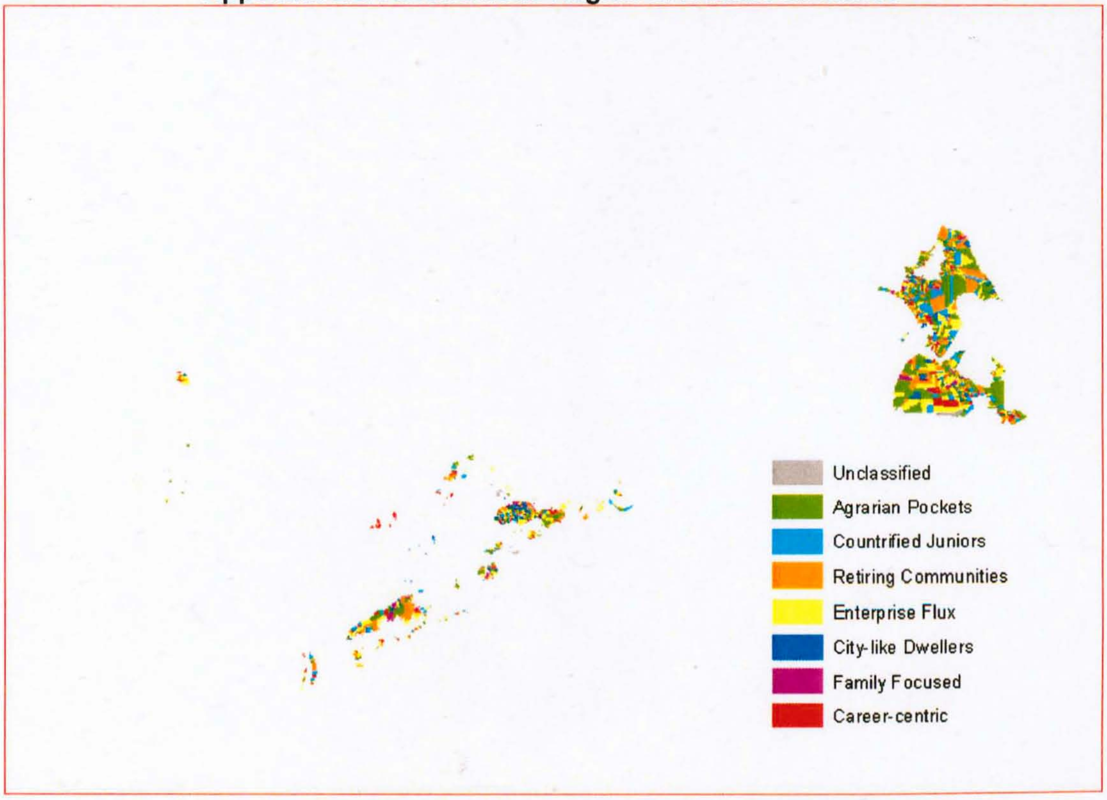
| | | | |
|---|---|---|---|
| | | | included with HO18 |
| HO20 | Cement roofing | | Redundant information |
| HO21 | Metal roofing | | Very low power and pseudo included with HO18 |
| HO22 | Wood roofing | | Very low power and pseudo included with HO18 |
| HO23 | Cogon roofing | | Very low power and pseudo included with HO18 |
| HO24 | Iron and concrete roofing | | Satisfies all the tests. |
| HO25 | Asbestos roofing | | Very high skew and low power. |
| HO26 | Scrap roofing | | |
| HO27 | Other roofing | | Vague information |
| R1 | Presence or absence of a church | | Redundant information. |
| R2 | No religion | | Large skew. |
| R3 | Buddhist | | Good geographical variation. |
| R4 | Muslim | | Very good variation across areas |
| R5 | Roman Catholic | | The most popular religious group. |
| R6 | Adventist | **Other Christians** | Some of the residential patterns and lifestyles of people in these religious groups differ greatly from the traditional Roman Catholics. |
| R7 | Assembly of God | | |
| R8 | Baptist | | |
| R9 | Church of Nazarene | | |
| R10 | Episcopalian | | |
| R11 | Jehovah's witness | | |
| R12 | Latter day saints | | |
| R13 | Lutheran | | |
| R14 | Methodist | | |
| R15 | Presbyterian | | |
| R16 | Aglipay | | |
| R17 | Bible Christian committees | | |
| R18 | Christian and missionary alliance | | |
| R19 | Foursquare gospel | | |
| R20 | God world mission | | |
| R21 | Iglesia Evangelista | | |
| R22 | Iglesia ni christo | | |
| R23 | Philippines benevolent missionaries | | |
| R24 | Number of barangays with a church/chapel or mosque in the barangay | | Redundant information |
| R25 | Number of barangays with no church/chapel or mosque in the barangay | | Redundant information |
| S1 | Town hall | | Important socio-cultural indicator. |
| S2 | Public plaza or park | | Relevant socio-economic indicator. |
| S3 | Cemetery | | Important socio-cultural indicator. |
| S4 | Public library | | Weak variation. |
| S5 | Barangay hall | | Strong variation across areas. |
| S6 | Newspaper circulation | | Redundant information. |
| S7 | Telephone service | | Redundant information. |
| S8 | Telegraph service | | Relevant to MDG 8. |
| S9 | Postal service | | Relevant to MDG 8. |
| S10 | Waterworks system | | Relevant to MDG 7. |

| | | | | |
|---|---|---|---|---|
| S11 | Electric power | | | Redundant information |
| S12 | Method of waste water disposal | | | Very relevant to MDG 7. |
| S13 | Number of recreational establishments in the barangay | | | Redundant information |
| S14 | Number of banking institutions, pawnshops and other financing/investment companies in the barangay | | | Redundant information |
| S15 | Number of barangays with town/city hall or provincial capitol in the barangay | | | Redundant information |
| S16 | Number of barangays with no town/city hall or provincial capitol in the barangay | | | Redundant information |
| S17 | Number of barangays with a cemetery in the barangay | | | Redundant information |
| S18 | Number of barangays with no cemetery in the barangay | | | Redundant information |
| S19 | Number of barangays with a public library in the barangay | | | Redundant information |
| S20 | Number of barangays with no public library in the barangay | | | Redundant information |
| S21 | Number of barangays with a barangay hall in the barangay | | | Redundant information |
| S22 | Number of barangays with no barangay hall in the barangay | | | Redundant information |
| S23 | Number of barangays with a housing project in the barangay | | | Redundant information |
| S24 | Number of barangays with no housing project in the barangay | | | Redundant information |
| S25 | Number of barangays with newspaper circulation in the barangay | | | Redundant information |
| S26 | Number of barangays with no newspaper circulation in the barangay | | | Redundant information |
| S27 | Number of barangays with a telephone in the barangay | | | Redundant information |
| S28 | Number of barangays with no telephone in the barangay | | | Redundant information |
| S29 | Number of barangays with a telegraph in the barangay | | | Redundant information |
| S30 | Number of barangays with no telegraph in the barangay | | | Redundant information |
| S31 | Number of barangays with a postal service in the barangay | | | Redundant information |

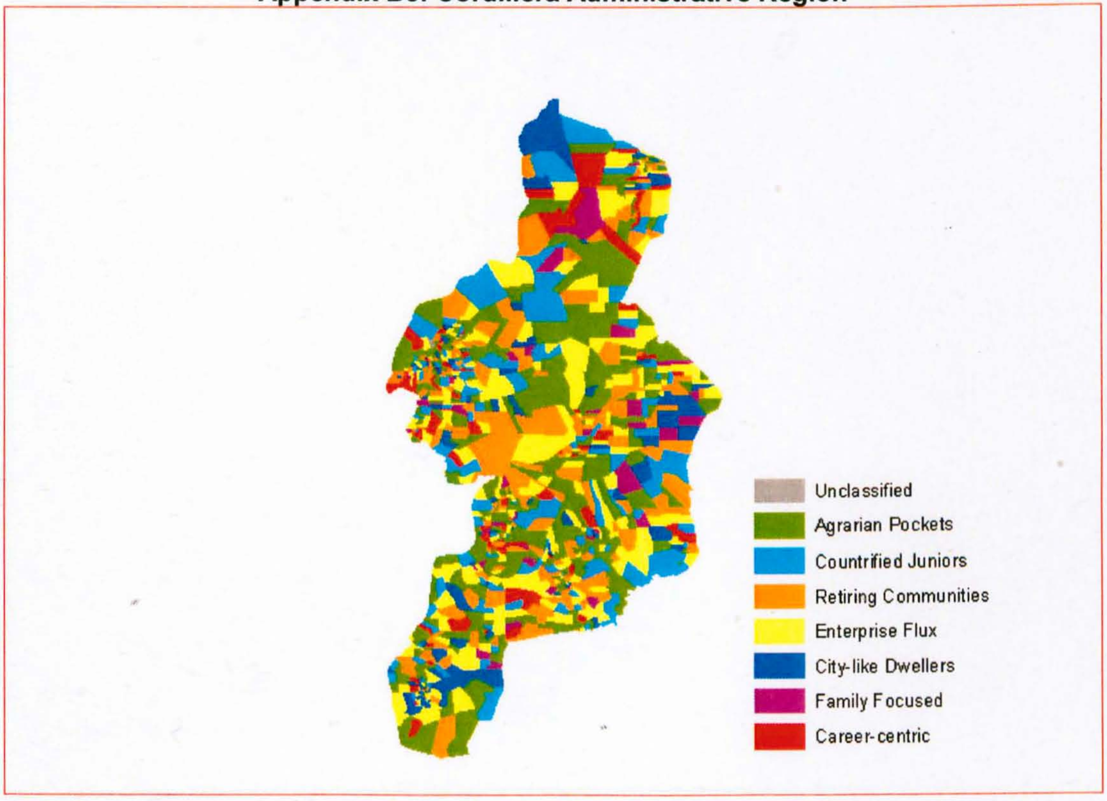
| | | | | |
|---|---|---|---|---|
| S32 | Number of barangays with no postal service in the barangay | | | Redundant information |
| S33 | Number of barangays with a community waterworks system in the barangay | | | Redundant information |
| S34 | Number of barangays with no community waterworks system in the barangay | | | Redundant information |
| S35 | Number of barangays with electric power in the barangay | | | Redundant information |
| S36 | Number of barangays with no electric power in the barangay | | | Redundant information |
| S37 | Number of barangays where most households dispose of waste water using septic tank | | | Redundant information |
| S38 | Number of barangays where most households dispose of waste water using open canal | | | Redundant information |
| S39 | Number of barangays where most households dispose of waste water using city/municipality sewerage | | | Redundant information |
| S40 | Number of barangays where most households dispose of waste water using water treatment facility | | | Redundant information |
| S41 | Number of barangays where most households dispose of waste water in the surroundings | | | Redundant information |
| SE1 | Market | | | Important socio-economic indicator. |
| SE2 | Number of wholesale stores, department stores, bazaars, sari-sari stores, gasoline stations and similar establishments in the barangay | | | Redundant information |
| SE3 | Manufacturing establishments | | | Satisfies all tests. |
| SE4 | Auto repair shops | | | Satisfies all tests. |
| SE5 | Restaurants and personal services | | | Very strong variation across areas |
| SE6 | Number of hotels, dormitories, and other lodging places | | | Redundant information |
| SE7 | Number of barangays with a public plaza or park in the barangay | | | Redundant information |
| SE8 | Number of barangays with no public plaza or park in the barangay | | | Redundant information |
| SE9 | Number of barangays with a market or building with trading activities in the | | | Redundant information |

| | | | | |
|---|---|---|---|---|
| | barangay | | | |
| SE10 | Number of barangays with no market or building with trading activities in the barangay | | | Redundant information |
| T1 | At least 3 street roads | | | Indicative of acceptable road infrastructure |
| T2 | Accessibility to the highway | | | Defines access |
| T3 | Number of barangays with a street pattern or network of streets of at least 3 street roads | | | Redundant information |
| T4 | Number of barangays with no street pattern or network of streets of at least 3 street roads | | | Redundant information |
| T5 | Number of barangays with access to the national highway | | | Redundant information |
| T6 | Number of barangays with no access to the national highway | | | Redundant information |

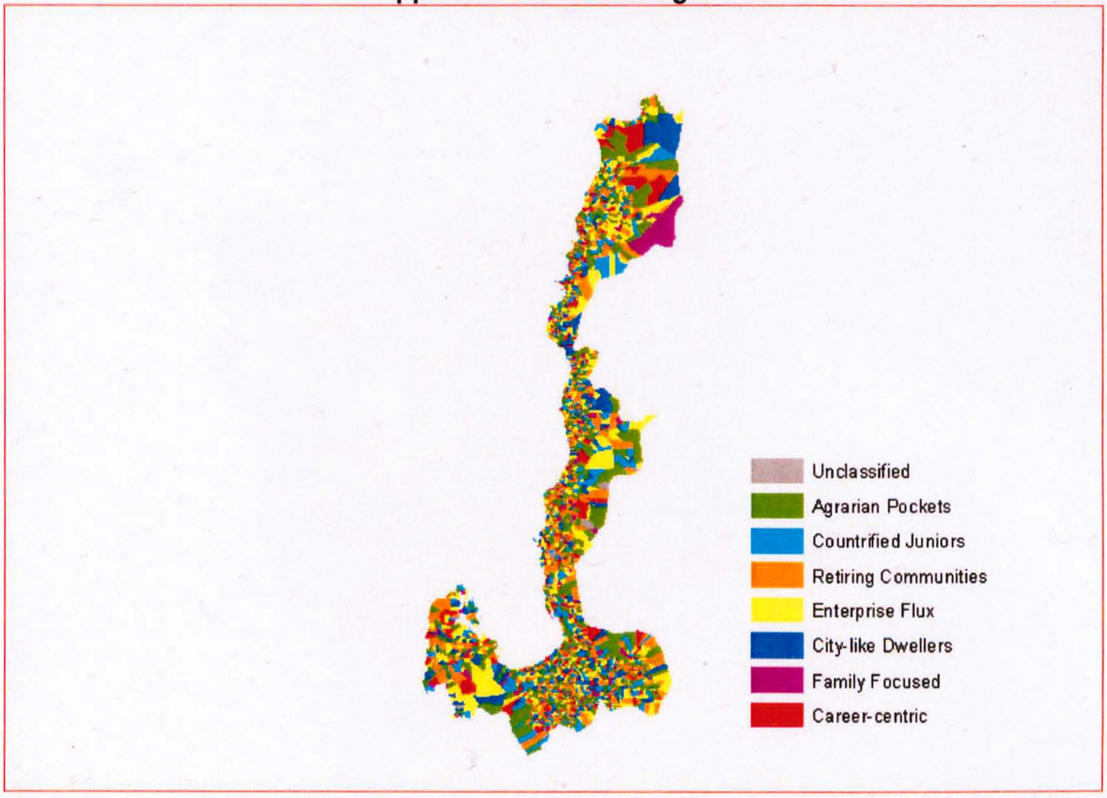
## Appendix B2: Autonomous Region in Muslim Mindanao



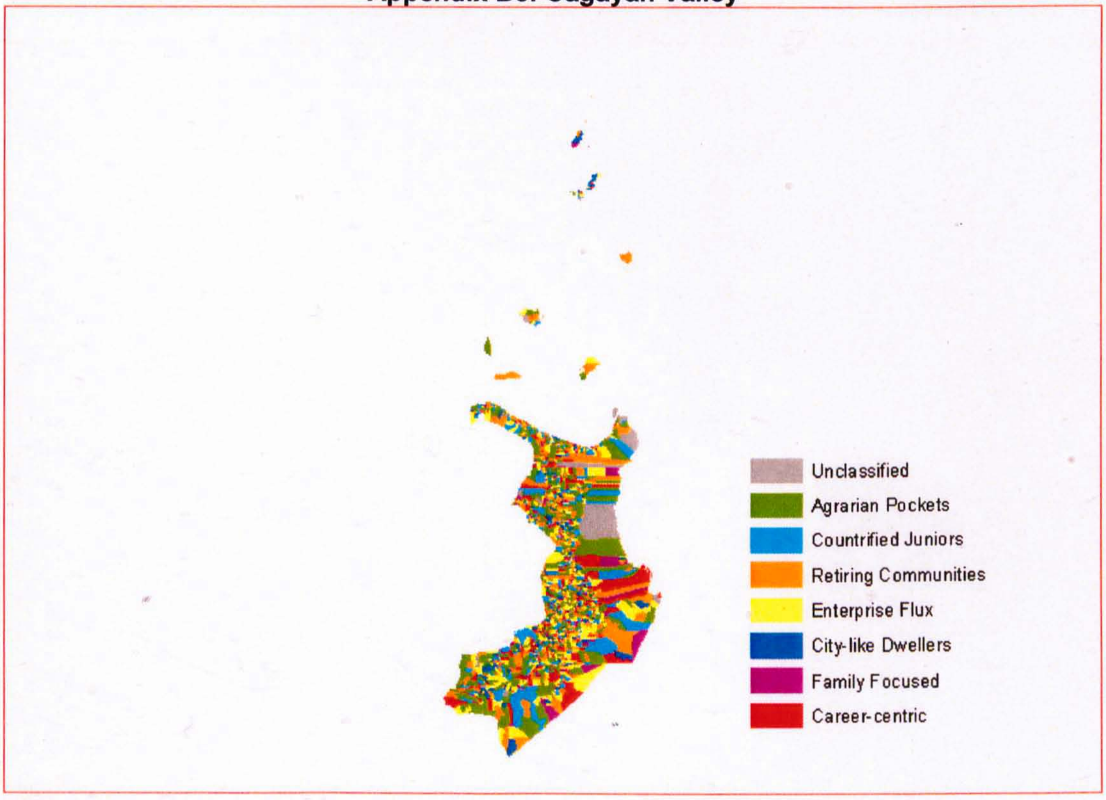

## Appendix B3: Cordillera Administrative Region



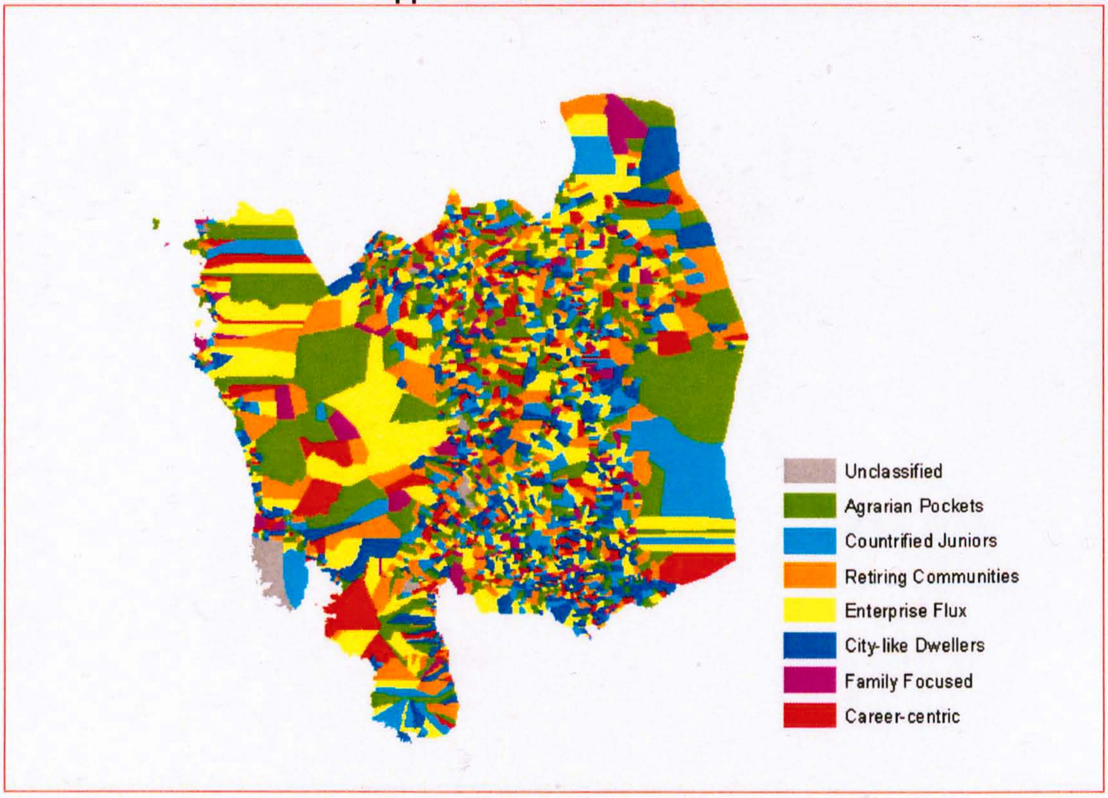
## Appendix B4: Ilocos Region



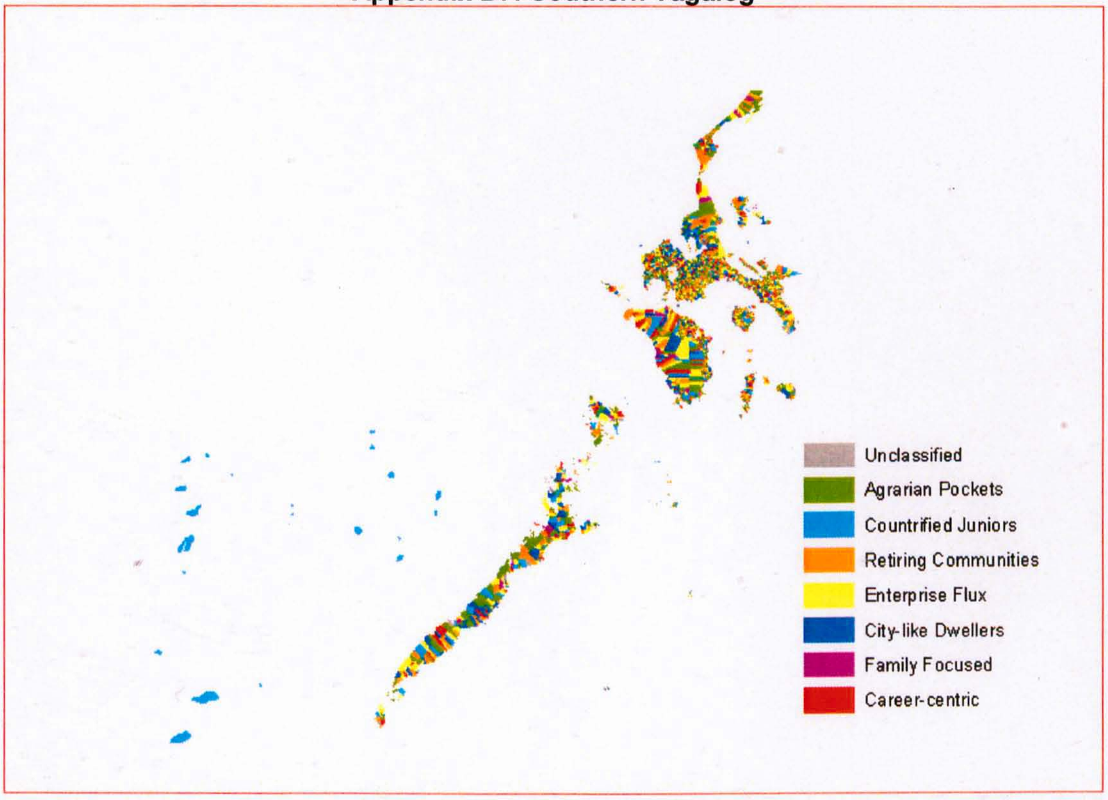

## Appendix B5: Cagayan Valley



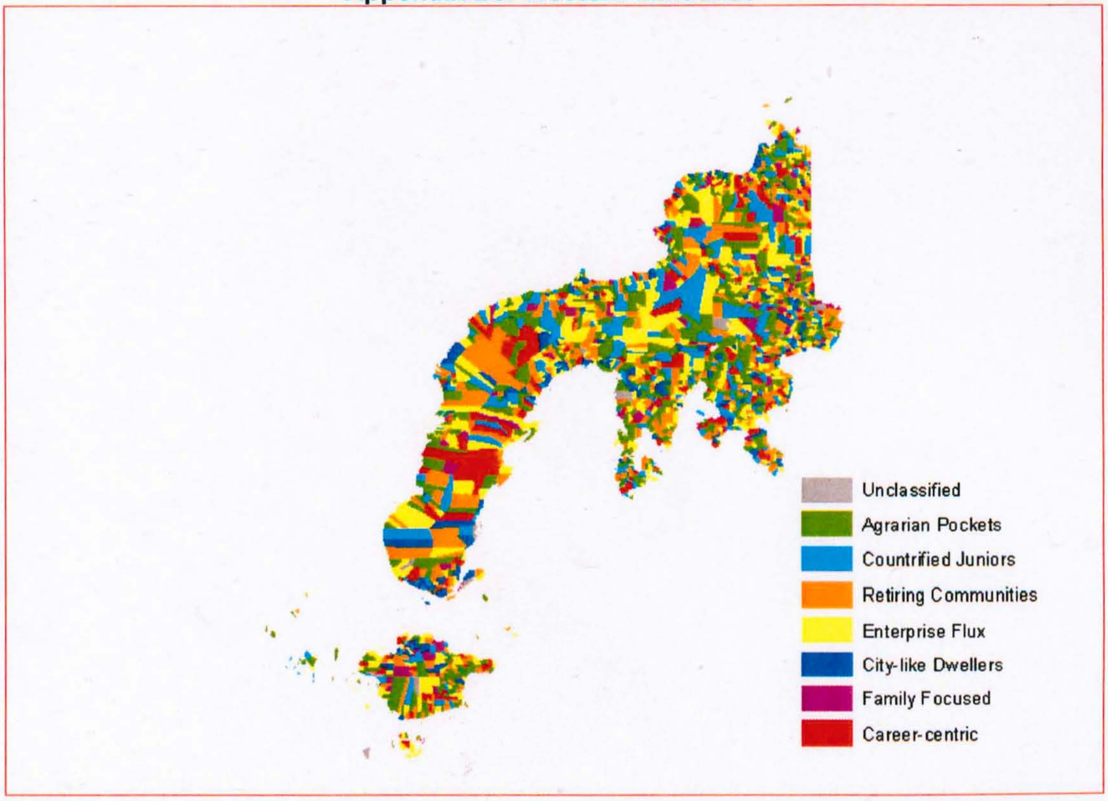
## Appendix B6: Central Luzon



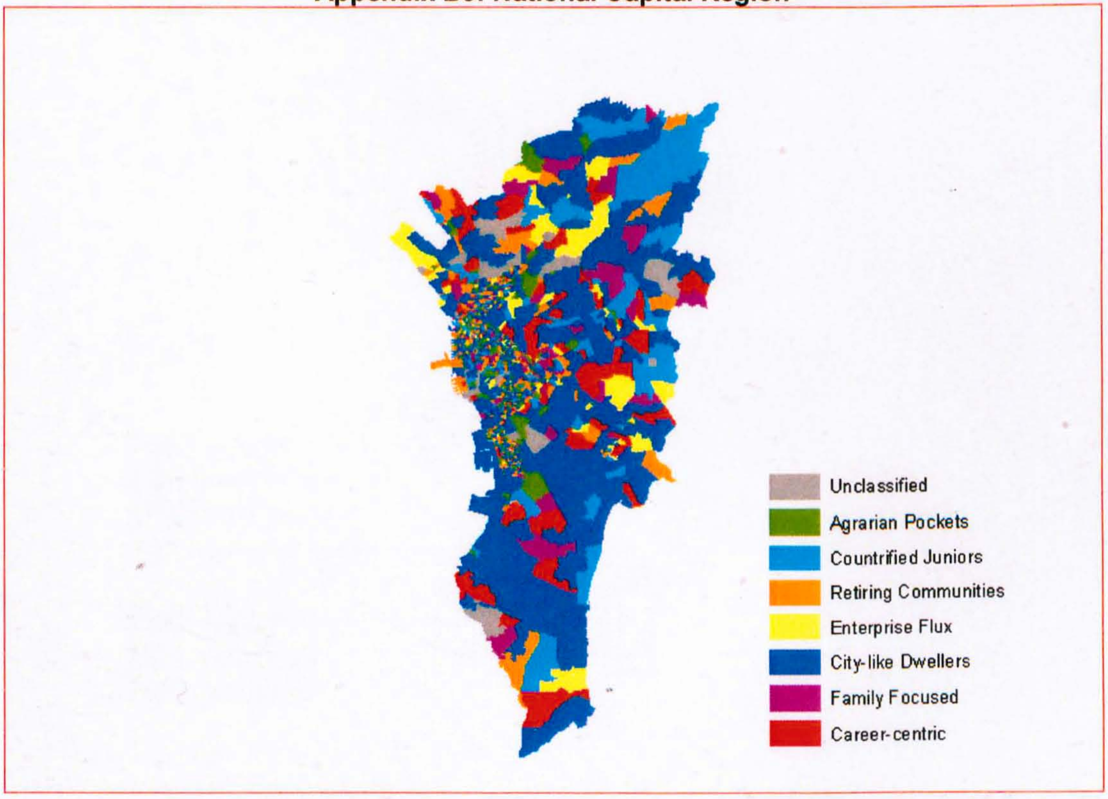

## Appendix B7: Southern Tagalog



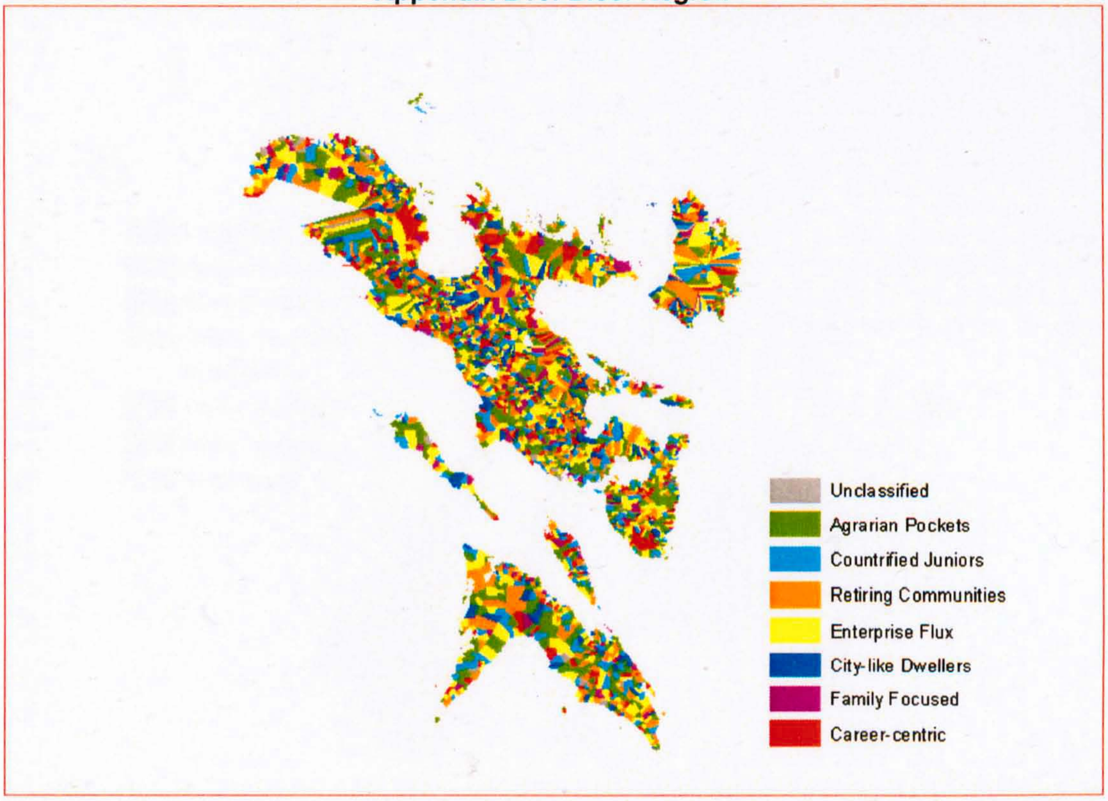
**Appendix B8: Western Mindanao**



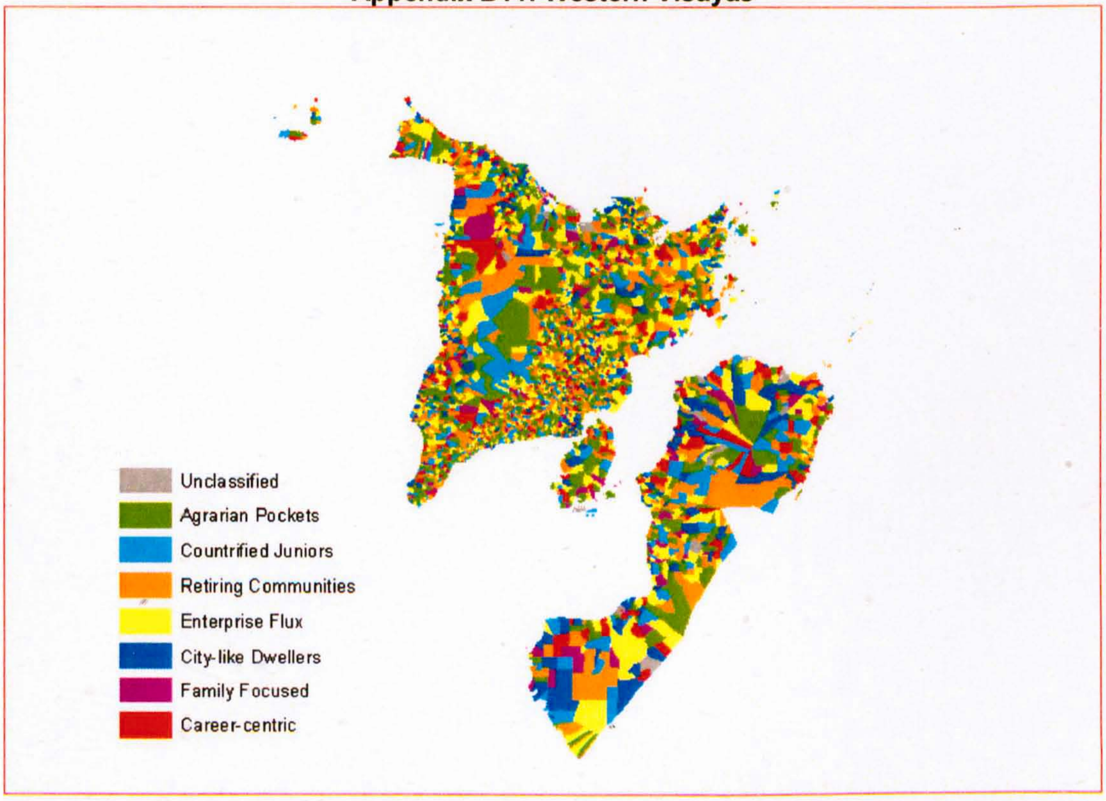

**Appendix B9: National Capital Region**



**Appendix B10: Bicol Region**

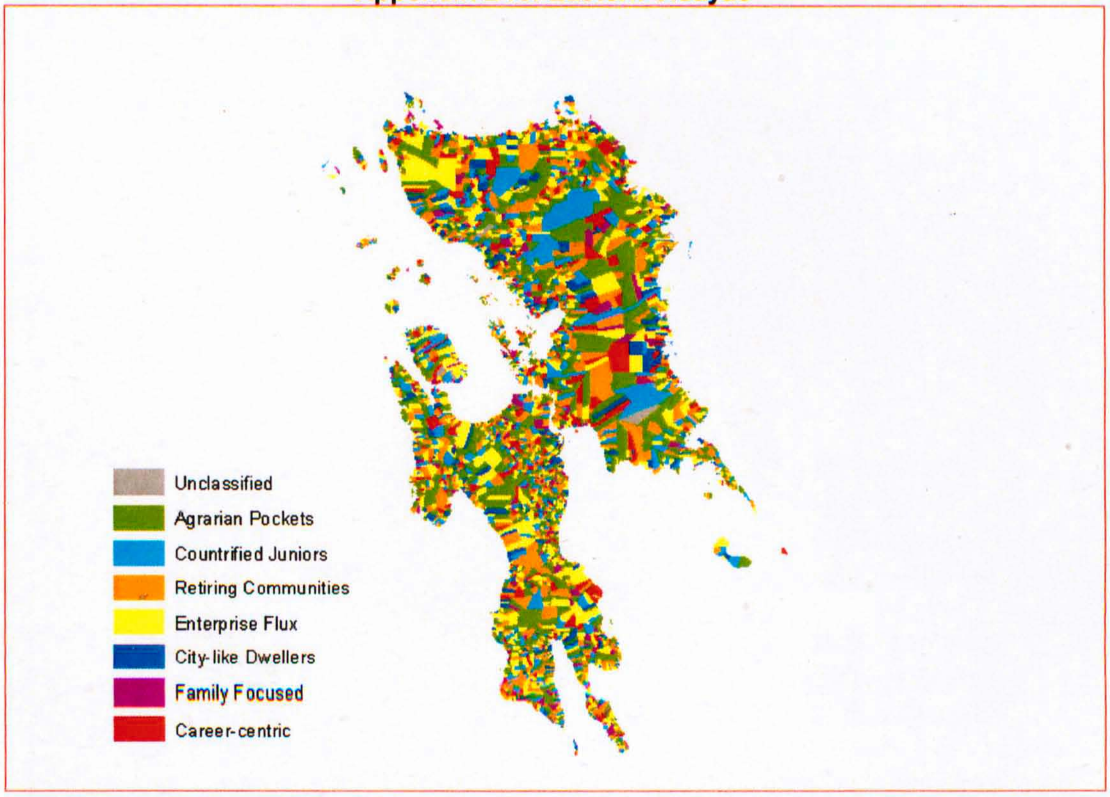



**Appendix B11: Western Visayas**

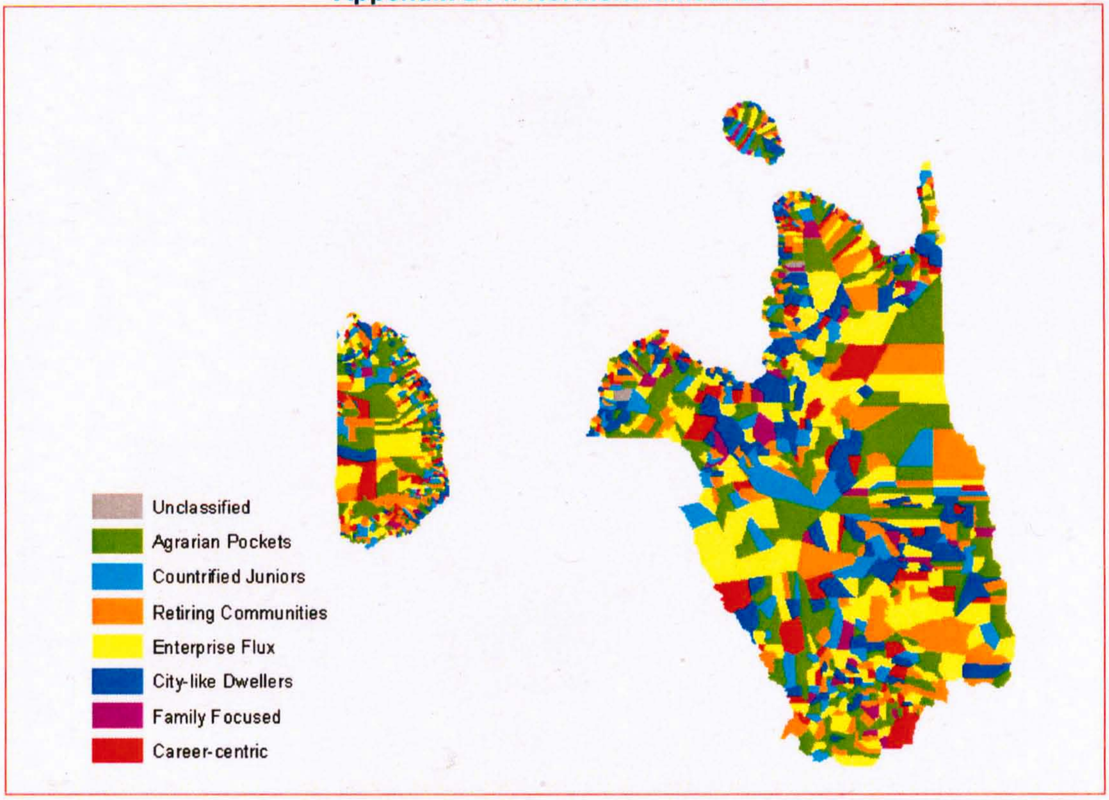


**Appendix B12: Central Visayas**

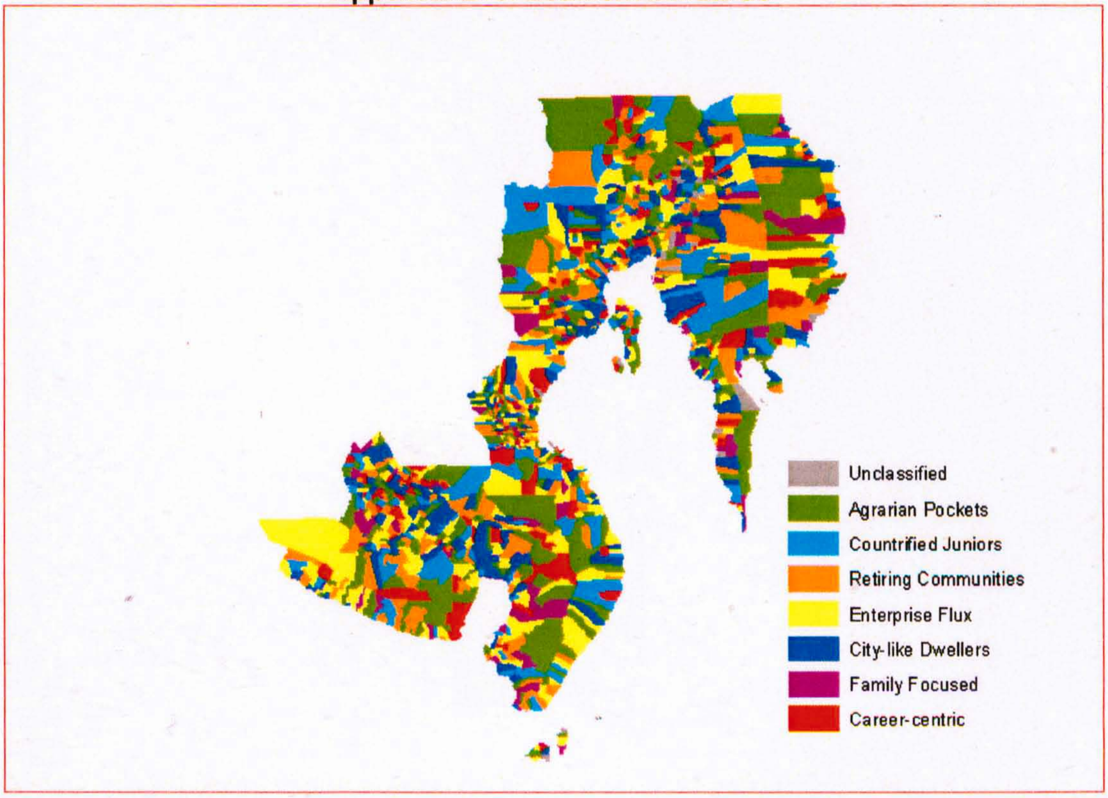

**Appendix B13: Eastern Visayas**



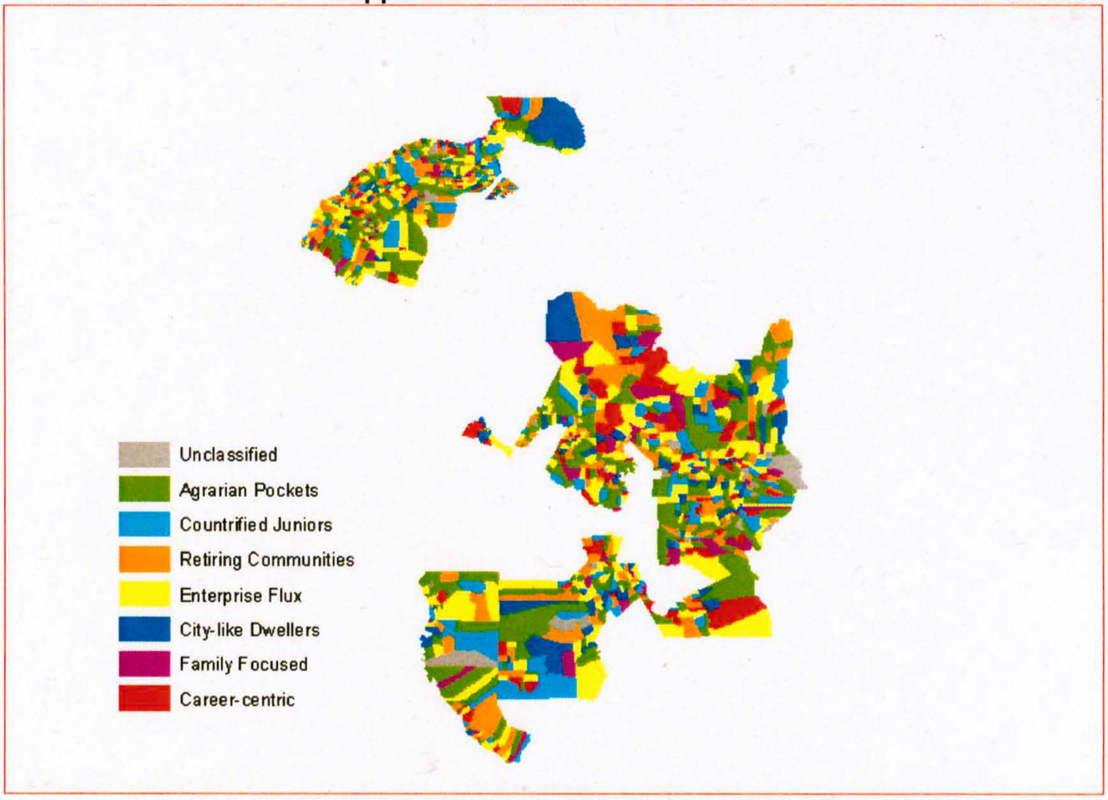
**Appendix B14: Northern Mindanao**




**Appendix B15: Southern Mindanao**



**Appendix B16: Central Mindanao**




**Appendix B17: Caraga**