UNIVERSITY OF LEEDS

Identifying and Predicting Neighbourhood Level Gentrification: A Data Primitive Approach

Jennie Helen Gray

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

The University of Leeds

Faculty of Earth and Environment

School of Geography

March 2023

Intellectual Property and Publications

The candidate confirms that the work submitted is their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Due to the context of the background, there may be some overlap between Chapter 2, and Chapter 4, where at the time of submission **Chapter 4** is an exact copy of the journal article published in the International Journal of Geo-Information:

> Jennie Gray, Lisa Buckner, and Alexis Comber (2021). "Extending geodemographics using data primitives: A review and a methodological proposal". In: *ISPRS International Journal of Geo-Information* 10.6. ISSN: 22209964. DOI: 10.3390/ijgi10060386

JG was responsible for literature searches, review, and authoring the paper. LB and AC aided with conceptualisation of the work. AC aided with methodology, software, and submission.

At the time of submission **Chapter 5** is an exact copy of the journal article published in the Applied Spatial Analysis and Policy journal:

> Jennie Gray, Lisa Buckner, and Alexis Comber (2023a). "Identifying Neighbourhood Change Using a Data Primitive Approach: the Example of Gentrification". In: *Applied Spatial Analysis and Policy*. ISSN: 1874-4621. DOI: 10.1007/s12061-023-09509-y. URL: https://doi.org/10.1007/s12061-023-09509-y

JG was responsible for data collection and collation, formal analysis, investigation, methodology, validation, and authoring the manuscript. LB and AC edited the final draft and supervised the work.

At time of submission **Chapter 6** is an exact copy of the journal article published in the Urban Science journal:

> Jennie Gray, Lisa Buckner, and Alexis Comber (2023b). "Predicting Gentrification in England: A Data Primitive Approach". In: *Urban Science* 7.2, p. 64. ISSN: 2413-8851. URL: https://www.mdpi.com/2413-8851/7/2/64

JG was responsible for formal analysis, investigation, methodology, and authoring the manuscript. LB and AC edited the final draft and supervised the work.

Signed: JHG

Authors Declaration

I, *Alexis Comber*, give permission for the following papers, co-authored with Jennie Gray, to appear in the PhD thesis:

- Jennie Gray, Lisa Buckner, and Alexis Comber (2021). "Extending geodemographics using data primitives: A review and a methodological proposal". In: *ISPRS International Journal of Geo-Information* 10.6. ISSN: 22209964. DOI: 10.3390/ijgi10060386
- Jennie Gray, Lisa Buckner, and Alexis Comber (2023a). "Identifying Neighbourhood Change Using a Data Primitive Approach: the Example of Gentrification". In: *Applied Spatial Analysis and Policy*. ISSN: 1874-4621. DOI: 10.1007/s12061-023-09509-y. URL: https://doi.org/10.1007/s12061-023-09509-y
- Jennie Gray, Lisa Buckner, and Alexis Comber (2023b). "Predicting Gentrification in England: A Data Primitive Approach". In: *Urban Science* 7.2, p. 64. ISSN: 2413-8851. URL: https://www.mdpi.com/2413-8851/7/2/64

Date: 17/03/2023

Signature: AJC

I, *Lisa Buckner*, give permission for the following papers, co-authored with Jennie Gray, to appear in the PhD thesis:

- Jennie Gray, Lisa Buckner, and Alexis Comber (2021). "Extending geodemographics using data primitives: A review and a methodological proposal". In: *ISPRS International Journal of Geo-Information* 10.6. ISSN: 22209964. DOI: 10.3390/ijgi10060386
- Jennie Gray, Lisa Buckner, and Alexis Comber (2023a). "Identifying Neighbourhood Change Using a Data Primitive Approach: the Example of Gentrification". In: *Applied Spatial Analysis and Policy*. ISSN: 1874-4621. DOI: 10.1007/s12061-023-09509-y. URL: https://doi.org/10.1007/s12061-023-09509-y
- Jennie Gray, Lisa Buckner, and Alexis Comber (2023b). "Predicting Gentrification in England: A Data Primitive Approach". In: *Urban Science* 7.2, p. 64. ISSN: 2413-8851. URL: https://www.mdpi.com/2413-8851/7/2/64

Date: 29/03/2023

Signature: LJB

Acknowledgements

First of all, I would like to thank my primary supervisor, Professor Lex Comber, for everything; his teaching, guidance, support, reassurance, and advice. He has helped me to develop my intellectual and analytical skills, and my overall ability to complete this research. Lex, your supervision throughout has been priceless, thank you.

My secondary supervisor, Lisa Buckner, has made this experience much more pleasant with her constant support throughout these years. She has offered me nothing but encouragement, words of support, and reassurance when my imposter syndrome hit hard. She is the role model every woman and girl needs in their life.

My CDT and PhD fellows, it has been a joy to experience this alongside you. And my girls, near and far, you're the best and I love you all.

My Mam has been there for me from my undergraduate, masters, to this Integrated PhD. Though she understands very little of what I do, she is proud, just like the rest of my immediate and extended family.

Adam, my love. For the past 12 years you have supported and encouraged me to become the woman that I am today. I hope to make you proud for the rest of our years. I love you always.

From Seed to Bloom: A Poetic Appreciation

I planted a seed over five years ago,
With hope and ambition for myself to grow,
And with the help of those around,
It flourished and bloomed, and grew unbound.

Like the sun that shines above,
Their encouragement filled me with love,
And like the rain that falls to earth,
Their guidance gave my work its worth.

Some nourished me with knowledge and care,
Others gave me room to breathe and share,
Together they helped me to flourish and thrive,
As I worked towards my goal and strived.

Now, like a flower in full bloom,
I look back and see how much I've grown,
And I know that without their aid,
I would not be where I stand today.

So, to my supervisors, friends, family, and partner,
I extend my heartfelt thanks forever,
For helping me reach my goal,
And supporting me with heart and soul.

Thank you for being the roots and stem,
The petals, fragrance, and nectar within,
That made it possible for me,
To become the best version of me.

Abstract

Identifying and analysing neighbourhood change is a critical task for urban planners and policy makers and is an active academic field. However, traditional approaches to neighbourhood change often rely on temporally static data and methods that reduce complex processes to one cluster label, or one score for example. This leads to a fragmented understanding of neighbourhood dynamics, on a temporal scale that does not align with the processes, resulting in the failure to capture their complex and multifaceted nature. These limitations highlight the importance of adopting new and innovative methods to provide more accurate and dynamic insights into neighbourhood dynamics. This research subsequently proposes a new approach, *data primitives*, and a methodological framework for their application. Data primitives are measurements of the fundamental components that capture the driving characteristics of clearly conceptualised neighbourhood processes. Their utility is explored in a regional analysis, identifying 123 cycles of gentrification and their respective temporal properties, which are exhaustively validated via Google Earth and Google Street View. This demonstrates the effectiveness of data primitives at capturing processes, and quantifying their changes over time, to provide a more comprehensive picture of neighbourhood change. These validated cycles of gentrification are used as a training dataset for training three machine learning algorithms for predicting gentrification in England. Three models were created to predict the presence of gentrification, the type of gentrification, and the temporal properties of the predicted types of gentrification in England. These predicted cycles of gentrification are explored, generating novel insights for the neighbourhood change and gentrification communities. Overall, the results of this research have important implications for urban planning and policy making, as they can provide a framework for informing decisions on where to invest resources and how to mitigate the potential negative effects of gentrification, in an appropriately scheduled timetable of interventions. They also provide a framework for uncovering novel insights into the complexities of neighbourhood processes, and their impacts upon neighbourhood change, thus developing upon knowledge in suitable academic fields.

Contents

List of Figures

List of Tables

List of Abbreviations

AGILE Association of Geograpic Information Laboratories in Europe.

AHAH Access to Healthy Assets & Hazards.

AONB Areas of Natural Beauty.

CDRC Consumer Data Research Centre.

CVA Change Vector Analysis.

DFT Department for Transport.

DWP Department for Work and Pensions.

EU European Union.

GBM Gradient Boosting Machine.

GISRUK Geographical Information Science Research U.K.

HMO Housing of Multiple Occupancy.

IMD Index of Multiple Deprivation.

LA Local Authority.

LCR Linked Consumer Registers.

LIME Local Interpretable Model-Agnostic Explanations.

LSOA Lower Super Output Area.

MAUP Modifiable Areal Unit Problem.

ML Machine Learning.

MSOA Middle Super Output Area.

NDVI Normalised Difference Vegetation Index.

NN Neural Network.

OA Output Area.

ONS Office for National Statistics.

OS Ordnance Survey.

PCA Principal Component Analysis.

RUC Rural Urban Classification.

SIC Standard Industrial Classification.

SOC Standard Occupation Classification.

VGI Volunteered Geographic Information.

XGBoost Extreme Gradient Boost.

Part I

Introduction

Chapter 1: Introducing the Case for Data Primitives

1.1 Research Rationale

Geodemographic classifications are models used for segmenting neighbourhoods into homogeneous groups. They have many uses including for targeted marketing (Leventhal, 2016), and improving the allocative efficiency of public services like policing (Ashby and Longley, 2005), education (Singleton and Longley, 2009), and the fire service (Corcoran et al., 2013; Taylor et al., 2016). Over recent years, they are also increasingly being used as a tool to analyse neighbourhood change over time (McLachlan and Norman, 2021). However, this research argues that geodemographic classifications are unsuitable for measuring neighbourhood change over time, because they are unable to accurately capture the true dynamics of an area. The static nature of the data (Reibel, 2011) and the cluster-based approach that classifications traditionally use masks many of the subtle changes that occur but are not large enough to result in a reassignment, but may still represent a relative change in the condition and quality of the overall cluster. Additionally, while commercial geodemographic systems claim to incorporate near real-time data into their clustering process, not all of the extra data is actually used, and that the truth is more complex since some may only use these attributes to augment group descriptions (Leventhal, 2016).

Moreover, neighbourhoods are dynamic and constantly changing. Their analysis via static cross-sectional analyses such as geodemographic classifications is a clear disjoint (McLachlan and Norman, 2021) between the temporal properties of the analysis mechanism and data, and the temporal properties of the neighbourhoods themselves.

Gentrification, for example, is a highly controversial and relatively rapid process that varies in form, driving characteristics, and subsequent periodicities. It can, however, be broadly described as the social uplift of an area via incomers who are typical of a higher socioeconomic status than the incumbent residents (Lees et al., 2010). This change typically comes at the cost (social, economic, cultural) of the incumbent residents often resulting in their displacement amongst other consequences (Davidson and Lees, 2010). Gentrification is consequently a highly politically charged term due to the negative consequences associated with the displacement of the working class (Cocola-Gant, 2019; Cooper et al., 2020), minority groups (Richardson et al., 2019), and those at the intersection of both (Huse, 2018). It can therefore also be considered a process of socio-spatial inequality (Hochstenbach and Musterd, 2021).

Gentrification is complex and multidimensional, and has a direct impact on neighbourhood character (Ilic et al., 2019). Studies of gentrification (and other neighbourhood processes) typically adopt a qualitative or quantitative approach. Qualitative research is often based on interviews of residents and case studies, providing rich contextual information to specific city-based studies (see Butler, 2007; Largent and Quimby, 2020). Quantitative gentrification studies on the other hand attempt to measure gentrification in some way using information typically captured in demographic data. Impacts upon neighbourhoods are typically measured through a selection choice of proxy variables, from sources like population censuses, but also beyond, such as housing stock, and economic data (Atkinson, 2000; Barton, 2016; Reades et al., 2019). Yet, such approaches to analysing gentrification over time can be characterised as 1) relying upon temporally static data from fixed intervals, and 2) information reductive methods. This suggests the need for methods able to analyse neighbourhood changes directly, at the temporal resolution of the neighbourhood change, and highlights a methodological gap in literature that could be filled by a data primitives approach.

Data primitives offer an alternative, extended approach for identifying, analysing, and predicting the spatial and temporal scope of neighbourhood change, with an emphasis on the processes driving the neighbourhood change. It is an approach borrowed from remote sensing to overcome issues in translating between different land-use classifications (Comber, 2008; Wadsworth et al., 2008). For neighbourhood research, they are instead designed to capture relative changes to and within neighbourhoods (small areas), through analysis of the most fundamental driving characteristics of the processes under investigation (Comber, 2008). Over a suitable temporal resolution, data primitives allow the state of a neighbourhood to be examined at different times, where such transitional changes in these neighbourhood states can identify and characterise a process of change, like gentrification, polarisation, urban decay, and urbanisation. Deeper analysis can uncover the periodicity of the processes, generating a novel understanding of complex neighbourhood processes on a smaller temporal scale than current methods for neighbourhood change analyses. These can then also provide the foundation for a model for predicting the future spatial and temporal extent of the identified neighbourhood process, on a national scale.

Although data primitives in remote sensing are tangible components in the form of spectral bands, electromagnetic radiation, and the geographic information of pixels, when translated into an applied sociodemographic and economic research context, data primitives are more abstract and conceptual. This is because in such sociological research, there is no truly "fundamental" unit of analysis. Thus, data primitives for neighbourhood change research require additional consideration, and lend themselves as a suitable conceptual framework for re-focussing analysis on to area change in place of the traditional static approaches, placing importance upon processes and the dynamics experienced.

Data primitives therefore afford opportunities for a methodological advancement to neighbourhood change studies, but place greater emphasis upon neighbourhood *processes*, as stated by Webber and Burrows (2018) and suggested by the works of McLachlan and Norman (2021). Data primitives are an approach that capture dynamic and relative change that can be identified, quantified, and analysed to generate novel understandings of neighbourhood processes. Such knowledge may be powerful for planners and local governments in the planning and preparation for their changing populations. For instance, understanding if a neighbourhood may gentrify, the likely type of gentrification, and when this cycle of change is expected to start, peak, and end, could enable the production of a framework and timeline of appropriate interventions for maximising benefits and reducing consequences. Data primitives could therefore be a valuable tool in public policy.

1.2 Aims and objectives

The overarching aim of this research is to examine the suitability of Data Primitives as a viable alternative and extended approach to geodemographic classification analysis, for analysing the spatial and temporal extent of neighbourhood change. A secondary aim is to explore the ability of the data primitive approach for predicting change via neighbourhood processes. To achieve this, several research questions are outlined in Section 1.2.1. To effectively satisfy these research questions, respective research objectives are proposed, alongside corresponding sections of thesis in which they are addressed (section 1.2.2).

1.2.1 Research Questions

- RQ1: To what extent can data primitives be conceptualised in reference to current literature, and explored as an alternative approach for analysing neighbourhood change?
- RQ2: How can the drivers of change associated with the neighbourhood process of gentrification, be conceptualised via data primitives, and be operationalised for the identification of neighbourhood change via gentrification?
- RQ3: To what extent can the conceptualised data primitives for gentrification be operationalised for identifying the temporal properties and manifestation of gentrification in a case study location?
- RQ4: How can machine learning be used to create predictive models for predicting the spatial extent of gentrification, and their associated temporal properties, and which model is the most suitable?
- RQ5: How do the results from the data primitive approach for measuring and predicting neighbourhood change via gentrification compare with other research? Can novel insights be gained?

The following research objectives assembled in Table 1.1 are designed to satisfy the research questions. Table 1.1 also shows where the respective objectives are evidenced within the thesis.

1.2.2 Objectives

Table 1.1: Thesis objectives

	Question	Chapter
1. Identify and suggest Data Primitives for neighbourhood processes to demonstrate their pertinence as an approach to analysing neighbourhood change via a number of neighbourhood processes	1	4
2. Examine the ability of Data Primitives for identifying and quantifying established cycles of gentrification and their relationships throughout space and time	2	5
3. Examine the ability of Data Primitives for identifying the temporal properties of established cycles of gentrification such that their process manifestation can be captured	3	5
4. Explore, analyse, and evaluate the extent to which Data Primitives can predict the spatial extent of established cycles of gentrification in England	4	6
5. Explore, analyse, and evaluate the extent to which Data Primitives can predict the temporal properties of established cycles of gentrification in England	4	6
6. Compare the results of the research and their insights gained with other gentrification studies to explore the originality and authenticity of the results	5	7

1.3 Thesis overview

This thesis follows an alternative format, *PhD by Publication*, summarised in Figure 1. Section I of the thesis consists of three chapters; Chapter I, Chapter 2, and Chapter 3. Chapter I is the Introduction, and establishes the need for such an approach in neighbourhood change studies, identifying the gap in literature in which they fill. Chapter 2, the Literature Review, discusses neighbourhoods and their dynamics in terms of their changes, and how such neighbourhood changes are typically measured. Geodemographic classifications and other traditional methods like indices are discussed in relation to measuring change, providing examples of their uses in gentrification studies. Together these methods are discussed regarding their limitations, and why another approach may be more suitable. Data primitives are then introduced, to which their potential benefits are explored. The chapter is then summarised with a problem statement. Chapter 3 includes descriptions of the methods used within each paper of this alternative thesis, providing greater detail than some paper counterparts on how the data primitives were created and used.

Section II of the thesis is formed of three papers fully published in peer reviewed journals Chapter 4, Chapter 5, and Chapter 6. Chapter 4 consists of the first paper, which is published via open access in the International Journal of Geo-Information, entitled *Extending Geodemographics Using Data Primitives: A Review and a Methodological Proposal.* It serves as the theoretical foundation and introduction of the data primitive approach to the wider quantitative urban geography research community, and includes a case study of how data primitives can be used. Chapter 5 consists of the second paper, entitled *Identifying Neighbourhood Change Using a Data Primitive Approach: the example of gentrification.* This paper is published via open access in the Applied Spatial Analysis and Policy Journal. It develops upon Chapter 4, extending research from a case study illustration, to include a full regional empirical analysis. It explores the establishment and manifestation of gentrification throughout the region, drawing insights for gentrification literature. Chapter 6 is the third and final paper, *Predicting Gentrification in England: A Data Primitive Approach*, and is also published via open access in the Urban Science journal. This paper develops upon Chapter 5, providing a national study for predicting the spatial and temporal resolution of different types of gentrification in England. Author CRediT statements for all papers can be found in Appendix A.

Section III is formed of Chapter 7, the Discussion, and Chapter 8, the Conclusion, and binds the three papers together as a whole. The Discussion summarises the results throughout each of the three papers and discusses them in relation to current literature. It discusses the implications of the main findings, the limitations of the research and the challenges encountered throughout, alongside some potential future work that could overcome the limitations and refine and improve the data primitive approach. Chapter 8 summarises the research and the major contributions of this thesis.

Finally, the appendix consists of three items. Appendix A consists of the CRediT author statements of each of the papers, stating author contributions. Appendix B consists of a large (64 page) table of the visual validation undertaken in Google Earth. Appendix C consists of a smaller table (10 page) of the visual validation undertaken in Google Street View. Appendix B and C were both based on results of Chapter 5, and undertaken for Chapter 6.

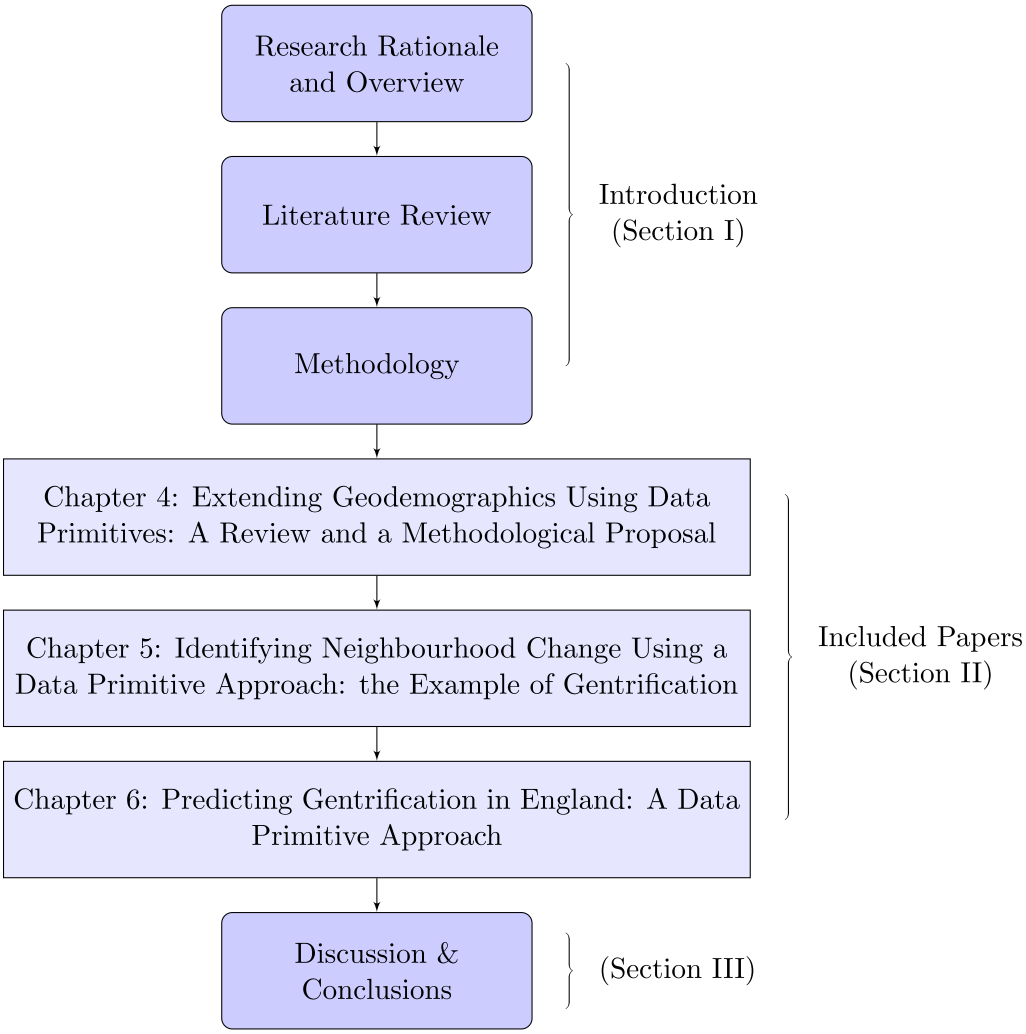


Figure 1.1: Thesis flow, including by publication structure

1.4 Thesis contribution

This research offers a significant contribution to the field of geodemographics and neighbourhood research; a new approach for analysing neighbourhood change over time, which does not rely on and is not restricted to the constraints of a cluster-based approach to change. This therefore enables neighbourhood change to be identified and analysed directly, instead of through the lens and constraints of a classification. This research therefore fills a gap in knowledge via the methodological advancement and creation of data primitives for analysing neighbourhood changes over time, via neighbourhood processes. Additionally, this thesis also offers some valuable research into the field of gentrification; when data primitives and other appropriate data (like change vectors) are correctly specified, the type of gentrification can not only be identified and predicted, but their manifestation via their periodicities can also be characterised. This provides valuable information about the spatial and temporal establishment and manifestation of different forms of gentrification.

The research has been presented at a series of events, including national and international conferences, colloquiums, and PhD Schools. These include the Geographic Information Research U.K. (GISRUK)- conference in Newcastle; the Association of Geographic Information Laboratories in Europe (AGILE) conference in Cyprus, and the GIMethods PhD colloquium, where precursors to this thesis were presented; and an AGILE PhD School in Estonia. Finally, outputs originating from this work have been disseminated via publication, with three publications in print at the time of submission, detailed below:

- Jennie Gray, Lisa Buckner, and Alexis Comber (2021). "Extending geodemographics using data primitives: A review and a methodological proposal". In: *ISPRS International Journal of Geo-Information* 10.6. ISSN: 22209964. DOI: 10.3390/ijgi10060386
- Jennie Gray, Lisa Buckner, and Alexis Comber (2023a). "Identifying Neighbourhood Change Using a Data Primitive Approach: the Example of Gentrification". In: *Applied Spatial Analysis and Policy*. ISSN: 1874-4621. DOI: 10.1007/s12061-023-09509-y. URL: https://doi.org/10.1007/s12061-023-09509-y
- Jennie Gray, Lisa Buckner, and Alexis Comber (2023b). "Predicting Gentrification in England: A Data Primitive Approach". In: *Urban Science* 7.2, p. 64. ISSN: 2413-8851. URL: https://www.mdpi.com/2413-8851/7/2/64

These papers are published within journals that seek to address issues of significance to the wider urban geography domain, and firmly situate the work within quantitative urban geography. This thesis and its associated publications provide a clear example which outlines the methodology and application of data primitives for identifying and predicting gentrification. Throughout this thesis, this research is contextualised via the wider literature, and evidenced using administrative and safeguarded data. The development of the predictive models are outlined in full detail, with results and limitations explored. The thesis clearly evaluates the performance of the data primitive approach for identifying gentrification, via extensive visual validation as seen in Appendix B and C, and the performance of the regional models used to predict gentrification and applied to a national case study. The latter of these remains unvalidated, but potential methods of validation and model refinements are explored within Section III.

Chapter 2: Literature Review

2.1 Introduction

This literature review is built around the assumption that the accumulation neighbourhood processes, like urban decline, urbanisation, and gentrification, which are driven by fundamental sociodemographic characteristics, result in neighbourhood change. These neighbourhood processes have different spatial and temporal extents due to the different accumulations and combinations of fundamental drivers, but each results in neighbourhood change. Yet the temporal mechanics of traditional approaches used to measure neighbourhood change limit the extent to which they can accurately capture *dynamic* neighbourhood change. Consequently, in this thesis, a new approach capable of capturing more dynamic and smaller neighbourhood changes (that of current and traditional methods cannot), is proposed.

Data primitives are conceptualised as a methodological advancement for neighbourhood change research. They aim to overcome the limitations of current approaches while advancing the understanding of neighbourhood processes. They are capable of generating deeper knowledge regarding their spatial and temporal extents, hence their manifestation throughout space and time. This literature review, therefore, grounds the data primitive approach within the wider academic field of neighbourhood change. It explores their relative advantages and disadvantages with respect to current literature and approaches to neighbourhood change analysis.

2.2 Neighbourhoods, Processes, and their Dynamics

2.2.1 Neighbourhoods

Neighbourhoods are where people live, they are where people have social interactions with their neighbours, where they feel comfortable and connected, and a place to have a sense of belonging. These feelings of community belonging translate into 'natural areas', where structural forces impact the composition of populations (Park, 1936; Hincks, 2015). Natural areas suggest homogenous characteristics, in terms of an area's demographic and socioeconomic characteristics such as their family structures, incomes, education levels, and ethnicity (Harris et al., 2005). This demonstrates Tobler's First Law of Geography "Everything is related to everything else, but near things are more related than distant things." (Tobler, 1970, p.234).

Administrative boundaries often represent neighbourhoods within research, or more increasingly, statistical boundaries. This is because they are easy to operationalise as they provide a standardised and consistent framework for collecting data. These administrative and statistical boundaries are used instead of the natural and social boundaries that residents often conceptualise as their neighbourhood, and the two may not align. This may be because the more practical statistical delineations (the conceived neighbourhoods) are invisible on the ground (Finney, 2013), differing from the perceived neighbourhood of residents (Coulton et al., 2001; Saar and Palang, 2009).

The Output Area is the smallest statistical building block, typically consisting of 300 people, or 120 households, on average. Output Areas (OA), and their agglomerated Lower Super Output Areas LSOA (LSOA) and Middle Super Output Areas (MSOA) (see Office for National Statistics, n.d.), are increasingly used as a proxy for neighbourhoods in recent neighbourhood research. These statistical boundaries were designed to be as socially homogenous as possible in relation to household dwelling and type, thus having more coherence to perceived neighbourhoods (Finney, 2013). These statistical boundaries are updated appropriately at each census according to demographic changes (Finney, 2013), to represent a consistent set of criteria, but may not represent actual changes to the neighbourhood.

Previous research has found that the LSOA is representative of neighbourhood effects and is a suitable spatial scale to conduct neighbourhood research (Van Ham and Manley, 2012). This research therefore adopts the statistical boundary of the LSOA spatial scale as the theoretical and practical basis for neighbourhoods; it represents a place bounded in space, where demography, social interactions, built environment and local politics and services can be represented through collated data (Galster, 2002).

2.2.2 Neighbourhood Processes

Neighbourhoods, regardless of how they are defined or delineated, are subject to dynamic processes that operate at different scales (Kearns and Parkinson, 2001). These processes contribute to the growth and decline of a neighbourhood and are crucial components of urban, or indeed any type of neighbourhood change (Buzar et al., 2005). These processes can be small or large in their spatial scale, but they function by driving changes in the geodemographic, socioeconomic, cultural, and even physical characteristics of a neighbourhood (Prouse et al., 2015). Thus, neighbourhood change can be attributed to the accumulation of the effects of multiple neighbourhood processes, which shift the general and overall characteristics of the residents and the area (Grigsby, 1986). Neighbourhood change is therefore a topic within urban geography regarding how residential areas transform over time (Prouse et al., 2015).

Larger processes act at larger spatial scales like regional, national, or even global levels, and they have a profound impact on neighbourhoods (Modai-Snir and Ham, 2019). For example, international migration is a macro process that impacts neighbourhoods, particularly regarding the spatial clustering of immigrants and their resultant levels of home ownership (Pamuk, 2004). The local pattern of demographic change is further driven by the age, sex, and educational attainment characteristics of those immigrants (Rees and Lomax, 2019). Some larger processes of change may also be the result of political decisions both nationally and internationally, like Brexit.

These larger scale processes may influence processes that occur on a smaller scale, from both the production and consumption side (see Crowe et al., 2020 for discussion on the production and consumption side of gentrification). The smaller scale processes act at a much more granular spatial scale, like the neighbourhood. It is these neighbourhood level processes that are the focus of this research, namely the most well-known and contested neighbourhood processes, gentrification.

2.2.3 Neighbourhood Change Research

Neighbourhood change is described as an accumulative process that can be examined from an ecological, sub-cultural, and political economy perspectives at different granularities (Temkin and Rohe, 1996). However, there are many facets to neighbourhood change. Lupton and Power (2004) review neighbourhood change literature and state that there are four main strands of work within neighbourhood change.

The first focusses on the community, with the aim of generating detailed understanding of particular neighbourhoods using qualitative methods and case study analyses (Lupton and Power, 2004). For example, Pinkster (2016) researched neighbourhood change via in-depth interviews in a working-class neighbourhood in Amsterdam and found a perceived process of neighbourhood decline. This was attributed to the respondents' strong sense of belonging to their changing neighbourhood, which led to their experience of disrupted daily life and a growing feeling of discontent with governing institutions and wider society.

The second derives from urban geography and sociology, with a more explicit focus on change. It has a more granular view of the neighbourhood and aims to generate understanding of how and why neighbourhoods change in relation to one another, by looking at patterns of residential settlement by race or class, and a number of other geodemographics (Lupton and Power, 2004).

The third relates to policy evaluation which focus upon neighbourhood-based policy interventions. Research within this strand typically focusses on disadvantaged neighbourhoods that have been subject to interventions and analyse the shorter-term impacts upon the area (Lupton and Power, 2004). For example, Kintrea (2007) investigated the impact of policies generated to improve social housing (known locally as 'council-built') neighbourhoods and found that they rarely have any discernible impact.

The final strand is 'neighbourhood effects research' which focusses on investigating the role of neighbourhood characteristics, like concentrated poverty, on the outcomes of residents (Lupton and Power, 2004). This domain's premise is based on the understanding that neighbourhood characteristics can have a significant effect on a resident's life chances over and above their individual characteristics (Van Ham et al., 2011). This domain therefore focusses upon the individual, as opposed to the neighbourhood, using quantitative techniques and large samples instead of case studies (Lupton and Power, 2004). However, one of its main challenges is identifying true causal effects (Durlauf, 2004), with many studies instead just highlighting correlations between outcomes and neighbourhood characteristics (Van Ham and Manley, 2009). Many of these strains of neighbourhood change typically fail to explicitly analyse neighbourhood change over time via longitudinal evidence, or multitemporal data (Lawless, 2012).

This research analyses neighbourhood change in respect to urban geography with an explicit focus on change at the small-area neighbourhood level. This research approaches neighbourhood change as an academic field within urban geography focussed upon the analysis of neighbourhood change over time, like the works of Grigsby (1986) for example. Traditionally, neighbourhood change analyses consisted of analyses of specific characteristics like income, race, social class, and ethnicity. However, the majority of neighbourhood change studies are now concerned with change in a more holistic sense, via a series of multivariate analyses which are explored in Section 2.4. Neighbourhood change has been attributed to the movement of people to adjust their housing consumption to meet their changing needs and preferences (Clark and Dieleman, 1996). These changing needs can arise due to or align with life events that impact household composition (Clark and Coulter, 2015), like parenthood and the need for a larger house to accommodate their growing family.

To summarise, neighbourhood processes and neighbourhood change arise from the accumulation of fundamental driving characteristics. As diverse types of people move, their diverse (accumulative) demographics, socioeconomics and cultural characteristics alter the driving forces of change. Thus, as the accumulation of different combinations of people move, the neighbourhood processes involved in producing the neighbourhood change differ. An example of a neighbourhood process that demonstrates neighbourhood change via the movement of certain types of people, is gentrification. Gentrification is one of the most widely studied processes throughout the world, to which this research contributes to.

2.3 Gentrification

Gentrification at its core is a process that changes the character of an area through newcomer residents, who are often of higher socioeconomic status than the incumbent residents, thus uplifting the overall neighbourhood status (Lees et al., 2008). This process over time changes the fundamental characteristics of the neighbourhood, such that those incumbent residents are either forced or decide to leave, resulting in displacement (Davidson and Lees, 2010). Numerous factors can cause displacement, economic causes include the increasing costs of rent, amenities, and other costs associated with the higher socioeconomic status newcomers (Atkinson, 2000). Cultural factors may include neighbourhood tensions that arise between the newcomers and the incumbents when the incumbents feel the loss of the connection to the neighbourhood they once had. Thus, displacement can have racial and class tensions (Richardson et al., 2019).

Subsequently, particular groups are more at risk of displacement, the working-class, private renters, older residents (aged 65 and over), unskilled workers, and those from ethnic minority groups (Sumka, 1979; Richardson et al., 2019).

Since the term's coining in 1964 to refer to the upgrading of the London neighbourhood via the 'invasion' of the middle class, replacing the working-class residents (Glass, 1964), the term has come to be applied to a wide range of variations of the phenomenon. This is because gentrification has mutated and intensified over time (Crowe et al., 2020), identified in many different forms in many different countries around the world (see Krase and DeSena, 2020 for example). These different forms of gentrification are sensitive to local contexts and temporalities; they relate to the different social and political structures in different cities, regions, and countries. Consequently, its form varies throughout the world and across a range of spatial (and temporal) scales (Lees et al., 2016; Crowe et al., 2020). See Table 2.1 on the next page for a table of some different types of gentrification, and their descriptions. Please note that this list is not exhaustive, and the processes may manifest differently outside of a U.K. perspective.

Table 2.1: Types of Gentrification and their Description.

Gentrification Type	**Description**
Super Gentrification	Gentrification that occurs in highly desirable, previously gentrified neighbourhoods, highlighted with the concentration of wealth and resources in major cities around the world.
Hyper Gentrification	Intense, rapid gentrification associated with high levels of investment. Often driven by millionaires and billionaires.
Marginal Gentrification	Often occurs on the periphery of established gentrified neighbourhoods in less desirable areas, in a slower and less intense manner. Can be driven by the spillover effects of investment and development in neighbouring areas.
State-led Gentrification	Gentrification actively facilitated by governmental agencies, policies and initiatives. Often associated with new-build gentrification.
Transit/Rail-induced Gentrification	Gentrification associated with the construction or expansion of public transit infrastructure, like railways. This makes neighbourhoods more accessible, connected, and attractive, which can lead to an influx of further investments. Properties closer to the transit link are more desirable, thus more expensive due to the "location premium".
Greenification	Gentrification that involves the introduction of greenspaces and sustainable environmental features. It includes the creation of parks, greenbelts, urban farms, and green transportation (bike lanes and pedestrian friendly streets), to promote urban sustainability and create more resilient and livable cities.
New Build Gentrification	Gentrification driven by the development of new housing estates. These can be located on brownfield land, on greenfield land, or via the displacement and demolition of existing housing estates. Can be driven by the state or private sector.

Table 2.1: Gentrification Types and their Description continued

Gentrification Type	Description
Private-led Gentrification	Gentrification facilitated by private developers, investors, and businesses. It typically involves the redevelopment of properties and businesses, resulting in up-scale housing and commerce, and increased rents.
Tourism Gentrification	Gentrification in previously affordable neighbourhoods to cater to affluent visitors, including new tourist accommodations (sometimes at the expense of the local housing market), as well as new shops, restaurants, and entertainment.
Retail Gentrification	Previously run-down or economically distressed urban neighbourhood is revitalised through the introduction of upscale retail, like boutiques, galleries and high-end restaurants. It transforms the local retail environment to meet the needs of the wealthier residents.
Student-Led Gentrification	Also known as Studentification, is associated with seasonal migration of students in university towns and cities. Housing tenure, commerce, and other urban amenities change to accommodate the student population, with more privately rented housing, bars and restaurants.
Rural Gentrification	Gentrification driven by the movement of residents from urban into rural areas, seeking a more peaceful, idyllic lifestyle. It includes the restoration of old homes and buildings, the development of new luxury homes, and vacation properties.

Valle (2021) has noted that researchers in the Global North overlook two important aspects that impact gentrification within the Global South, the roles of local political-economic forces, and the state. For example, gentrification in China has typically been a state-driven urban renewal and restructuring process intended to result in the class transformation of neighbourhoods (Guan and Cao, 2020). The process here is more akin to urban expansion and suburbanisation, and theories of gentrification are adapted to recognise local systems. Nevertheless, there are still some similar shared processes. Beijing has seen the construction of many parks in the past 10 years, which has led to green gentrification (Wu and Rowe, 2022), a process whereby the development of new urban greenspaces attract the upper-income residents, transforming the sociodemographic and socioeconomic characteristics of neighbourhoods (Nelson et al., 2010; Rigolon and Németh, 2020). However, there are many socio-spatial inequalities within this access to urban greenspace, which can compound the natural consequences of gentrification, like displacement (Wu, 2020).

The types of gentrification can also vary within specific localities. For example, the more affluent neighbourhoods of London like Barnsbury, and those on the periphery of the city and those in central London, are more susceptible to super-gentrification, due to the super wealthy professionals working in the city (Butler and Lees, 2006; Yee and Dennett, 2022). Beyond this, multi-millionaires and billionaires are driving hyper-gentrification in elite neighbourhoods in London (Lees, 2018). On the other hand, marginal gentrification, which is gentrification driven by the less privileged of the middle class (Mendes, 2013), is more likely in East London, where neighbourhoods are less affluent that those experiencing super gentrification (Mendes, 2013; Yee and Dennett, 2022). Marginal gentrification is closely associated with incumbent upgrading, for example, particularly in explaining gentrification in Amsterdam (Hochstenbach et al., 2015). However, incumbent upgrading is largely driven by the existing residents within a neighbourhood. This can be via an increase in their socioeconomic status over time, or the improvement of their own properties, and is therefore not usually associated with any displacement (Owens, 2012). Rural gentrification also occurs within specific localities, rural neighbourhoods, and is associated with lower levels of churn, but increased house prices via the restoration of old homes and buildings (Uysal and Sakarya, 2018)

Aside from local political systems, the form of gentrification can also relate to the type of development or commercial activity within an area. Examples include retail, tourism, transit and rail-induced gentrification. Retail-led gentrification is associated with the introduction of higher-end retail establishments, and the subsequent transformation of the commercial landscape. This contributes to the displacement of existing local businesses and services in which the working-class residents rely, and reshapes the socioeconomic dynamics of a neighbourhood (Hubbard, 2018). Tourism gentrification is whereby the attraction of visitors to areas not traditionally regarded as tourist spaces, increases the value of commercial and residential properties

(Cocola-Gant, 2018). Recent studies on tourism gentrification have focussed upon the impact of short-term rental properties, like those listed on Airbnb, and how they become a driving force in gentrification by re-spatialising tourism (Robertson et al., 2020). Transit-induced gentrification refers to the process by which the introduction, expansion, or improvement of transportation infrastructure, particularly public transit in lower-income neighbourhoods, stimulates increased property values, investment, and the subsequent displacement of lower socioeconomic status residents for the upper-class population (Bardaka et al., 2018; Delmelle, 2021).

Studentification is a type of gentrification associated with high concentrations of students in university towns and cities (Smith, 2019). It is a phenomenon observed throughout the world, and social changes include the seasonal in-migration of students and the displacement of settled residential populations. Such displacement can be associated with the conversion of owner-occupied family residences into private rented housing of multiple occupancy (HMO) for off-campus accommodation (Hubbard, 2008; Smith, 2019).

The type of gentrification can also be defined its initiator. For example, state-led gentrification is initiated by the government or government agencies with public funding in order to promote urban renewal and prevent decline (Shmaryahu-Yeshurun and Ben-Porat, 2021). Private-led gentrification is initiated by private investors, from developers to corporate landlords, and the transnational wealthy elites, for profit (Aalbers, 2019).

Gentrification can also be analysed from demand-side, and supply-side perspectives. Demand or consumption side gentrification focusses on the factors that drive an influx of higher socioeconomic residents into a neighbourhood, such as preferences for urban living, proximity to amenities, and the appeal of vibrant urban environments. These factors create a demand for housing in previously neglected or economically disadvantaged areas, leading to rising property values and changes in the socio-economic composition of the neighbourhood (Zapatka and Beck, 2021). Supply or production side gentrification, on the other hand, emphasizes the role of property developers, investors, and government policies in facilitating the physical redevelopment and revitalization of neighbourhoods - thus the state-led, and private-led gentrification, which is driven by profit-seeking development activities (Zapatka and Beck, 2021). This can involve new construction, renovation of existing properties, and the implementation of urban renewal projects that attract higher-income residents and businesses. Both demand- and supply-side factors interact and contribute to the process of gentrification, shaping the economic, social, and physical transformation of neighbourhoods over time (Zapatka and Beck, 2021).

There are connected themes throughout these types of gentrification in regard to their major characteristics. For example, some types are associated with a magnitude of change in house prices and incomes, like super and hyper gentrification (Yee and Dennett, 2022), retail gentrification (Lee, 2013), and rural gentrification (Stockdale, 2010), whilst others are associated with a magnitude of change in churn, like studentification (Sage et al., 2012), tourism and transit-induced gentrification (Chapple et al., 2017), and new-build gentrification (Davidson and Lees, 2010). Some are also associated with the displacement of lower-class and ethnic residents (Huse, 2018), like rural gentrification (Phillips et al., 2021), new-build gentrification (Davidson and Lees, 2010), and tourism gentrification (Cocola-Gant, 2018). These themes therefore suggest that four key variables can therefore capture the most primitive, essential components of gentrification; house price to represent the magnitude of change in house prices; professional occupation to represent the middle- to upper-class; neighbourhood churn to represent the magnitude of change in neighbourhood churn or turnover; and ethnicity to represent the displacement of the ethnic residents.

2.4 Measuring Neighbourhood Change

The Chicago School undertook some of the first quantitative studies of neighbourhood change in the 1920s, when they began analysing the sociological structure of cities and their neighbourhoods (Hincks, 2017). Shevky and Williams (1949) created indices of social processes to examine urban society; they looked at how cities changed over a period of 30 years, and developed the first indices to measure neighbourhood change, via the domains of urbanisation, social rank, and segregation. This set of works started an interest in area "dynamics" and neighbourhood analysis, and provided the basis for modern geodemographic classifications (Webber and Burrows, 2018).

Nowadays researchers use a number of different approaches for analysing neighbourhood change including index-based approaches, threshold-based approaches, and machine-learning approaches including clustering (Barton, 2016). Many are often used in tandem with another (Liu and O'Sullivan, 2016). They attempt to quantify neighbourhood changes in some way, and because of the multidimensionality of neighbourhood processes like gentrification, there is a growing consensus that changes across several indicators can measure gentrification without oversimplification (Anguelovski et al., 2018; Liu et al., 2019). This section will first explore the spatial scope of gentrification studies, and then go on to explore these methods in relation to neighbourhood change and gentrification.

Gentrification studies are usually limited within their spatial scope to specific localities. These can range from one or multiple cities, metropolitan areas, and even select neighbourhoods (Finio, 2022). Therefore, these studies are generally very location specific, with limited gentrification analyses observed beyond metropolitan areas, for example with spatial scopes of a region or country.

2.4.1 Threshold-Based Approaches

The threshold-based approach to analysing change is very simple and based on the premise that for a neighbourhood to change, it must cross a predetermined threshold and transition from one state to another (Reibel and Regelson, 2007; Hincks, 2017).

In gentrification studies, the most basic threshold used is often an initial *gentrifiable* criteria, neighbourhoods are identified as gentrifiable if they feature a particular characteristics at the beginning of the decade (Barton, 2016). The most simple is to mark areas as eligible to gentrify if they have specific characteristics above or below a threshold (Finio, 2022). Researchers can consider one or multiple characteristics which are often socioeconomic indicators, and typically determined via general rules of thumb. For example, a common threshold is a higher than metropolitan average within a specific indicator (Choi et al., 2018), many of which use median household income (Bhavsar et al., 2020). Cole et al. (2019), deemed neighbourhoods eligible to gentrify if more than 50% of residents fell into the low-income category, based on census data from 2000, whilst Gibbons (2019) and Gibbons and Barton (2016) consider neighbourhoods eligible to gentrify if their median household income was below that of their city. However, for Steinmetz-Wood et al. (2017) the eligibility criteria was different. They required a negative z-score in three measures (median household income, proportion of population with bachelors degree, average rent) and a positive z-score for proportion of population with low income, in the baseline year, relative to the metropolitan average. Freeman (2005) considered multiple characteristics, and marked areas as eligible to gentrify if they had a median income below the metropolitan average, and a level of housing construction below a 20-year metropolitan average. Though socioeconomic indicators are often used, data relating to the built environment is also considered. For example, Izenberg et al. (2018) used three criteria in their California study, median household income below median, at least 50% of the area considered "urbanised", and the proportion of building stock pre-1980 above the area median. This also displays a great example of cultural specificity in the characteristics of gentrification. A comparative example is that the housing stock associated with gentrification in London is traditionally associated with terraced housing (Hamnett, 2003).

Due to the simplicity of the threshold approach, thresholds can also be utilised at multiple points of the analysis, including as the eligibility criteria, and as a gentrification measure. When using a threshold to measure gentrification, researchers will typically first use their eligibility criteria to identify the areas likely to gentrify, and will then analyse those areas at the start and end of the study period. If the gentrification criteria is fulfilled by the end of the study period, like an increase in the proportion of higher socioeconomic status (Barton, 2016), the area is considered to have gentrified. Gentrification thresholds can also include a number conditions, Freeman (2005) considered a neighbourhood to have gentrified if it saw a greater increase in educational attainment than the metropolitan average, and an increase in real housing prices over the period. Gibbons (2019) used an increase in gross rent or median home value above the city median, and an increase in college-educated residents above the city median for a gentrification threshold. See Bhavsar et al. (2020) for a systematic review of gentrification papers regarding their eligibility and gentrification thresholds.

Thus, a neighbourhood is only considered to have gentrified if it passes a certain, subjective threshold between the (typically static) temporal boundaries of the study (Barton, 2016). Atkinson (2000) used this threshold strategy to identify neighbourhoods that experienced gentrification in London via those that experienced increases in the proportion of workers employed in a professional occupation between two time points, signifying an increase in the middle-class associated professions (see Galster, 2002; Liu et al., 2019).

A threshold-based method is a simple, convenient, and theoretically sound approach (Liu et al., 2019). Since they can be simple to initialise and use, different thresholds can be modelled, and the results compared. In this way, the best performing, or the most suitable threshold for the study can be identified (Preis et al., 2021). However, this does highlight the subjectivity of the threshold-based approach. The predetermined thresholds are often arbitrary figures and may not be relevant or generalisable to gentrification studies elsewhere in the nation or world

(Freeman, 2009).

Furthermore, due to the required minimum amount of neighbourhood change to be considered transitioned from one state to another, the threshold-based approach may not capture the full dynamics of neighbourhood change. When thresholds are applied to change analyses, the observations (neighbourhoods) must experience a level of change that surpasses the threshold in space and time (Comber and Wulder, 2019). This subsequently means that the smaller, yet potentially just as significant changes are not captured, which may give an indication into the type and magnitude of change yet to come (Zhu, 2017; Comber and Wulder, 2019). For example, an area might move from 6, to 1 standard deviations below the mean, but because it is still below the threshold, it would not be captured. There may be more suitable approaches to neighbourhood change detection that can capture these changes below the threshold and observe them as a signal of major change. This implication is also compounded by their temporal limitations: these are snapshots at specific points in time, some up to a decade apart. Consequently, the threshold approach to neighbourhood change is incapable of accurately capturing the smaller signals of change that occur at smaller temporal intervals than the temporal properties of the data used in analysis (Hincks, 2015; Hincks, 2017).

2.4.2 Index-Based Approaches

A composite index is a statistical method that groups information from multiple variables and summarises it into one single measure (Hawken and Munck, 2013). In neighbourhood based research, they are used to capture the state of a neighbourhood, for instance measuring deprivation, urban decay, or gentrification, though their usefulness depends upon their underlying weighting and aggregation (Zhou et al., 2007).

The Townsend (1987) and Carstairs and Morris (1989) indices are two of the most well-known composite indices, created to measure neighbourhood deprivation via the social and economic status of people or households in the neighbourhood. This approach selects several variables that capture and represent different facets of deprivation, which are then standardised and combined (with either equal or unequal weighting), to generate one index score – the sum of the un/weighted variables. For example, Townsend's index generated a "Townsend score" for the area, where a greater score implies a greater degree of deprivation. Areas can be ranked according to their scores to explore relative levels of deprivation. This can aid with the understanding of disadvantage to target resources for regeneration (Baing, 2009; Greig et al., 2010), exploring relationships between health outcomes, inequalities, and deprivation (Butler et al., 2013; Cabrera-Barona et al., 2015) and education (Higgs, 1999).

The Index of Multiple Deprivation (IMD) (Noble et al., 2006) is a well known composite index which uses multivariate data to measure deprivation in a range of domains (employment deprivation; health deprivation and disability; education, training, and skills deprivation; barriers to housing and services; living environment deprivation; and crime). Although it has numerous iterations from the 1990s to the most recent update in 2019, they are built with different methodologies across both time and space: the English, Welsh, and Scottish IMD are built differently, and are thus are not directly comparable (Payne and Abel, 2012). However, when directly comparable indices are created, they enable the characterisation and monitoring of neighbourhoods such that neighbourhood trends can be explored (Lupton and Power, 2004). For example, Norman (2010) extended the temporal application of indices and created directly comparable deprivation indices based on 1991 and 2001 Census data and the Townsend Index. He found that deprivation eased between the two decades, due to reductions in levels of unemployment. A more recent example is that of Airgood-Obrycki (2019), who developed a weighted composite index to analyse neighbourhood change from 1970-2010 in 100 U.S metropolitan areas. They were able to identify and categorise areas into one of four types to indicate whether an area had declined, remained stable or improved, finding socio-spatial differences regarding status and capability for improvement.

As an approach to measuring change, indices have become more sophisticated and specialised over time, due to critiques that they were not specific to certain localities, like rural areas (Clelland and Hill, 2019). They now also use a greater range of component indicators, at a smaller geographical unit, as new data sources have become available (Clelland and Hill, 2019). Due to their rather simple basis, composite indices are easily adapted to different topics and specific locations. They have been used to analyse climate change related environmental changes (Rincón, 2012), and to measure how healthy a neighbourhood is regarding their access to healthy assets and hazards (Daras et al., 2019). Due to their applicability, indices are a popular approach for measuring gentrification.

Index-based approaches to gentrification analyses often first identify areas that are ‘eligible’ to gentrify, typically based on the economic standing of an area in relation to the administrative average (Barton, 2016). This eligibility criteria is operationalised as a threshold based qualifier, as described in Section 2.4.1. Once the eligible neighbourhoods have been identified, researchers then employ a multivariate approach to measure the presence and extent of gentrification. Researchers will conceptualise the gentrification process in relation to their local settings, selecting a range of appropriate variables, before calculating a composite index at the start and end of the study period. These index scores will be compared, and neighbourhoods will subsequently be classified as having experienced gentrification or not (see Voorhees, 2014; Chapple and Zuk, 2016; Pegler et al., 2020 for examples).

The magnitude of the index score also gives an indication as to the amount of change the neighbourhood has experienced, thus neighbourhoods can be ranked according to their levels of change. For example, Kosta (2019) explored commercial gentrification in two Little Italies in New York, and found that his *food index* provided insight into different trajectories of neighbourhood change based on different patterns of consumption. Johnson et al. (2022) also explored gentrification in New York City, using principal component analysis on five key variables and smoothing scores through an autoregressive model. They found that their results were consistent with the general understanding of gentrification in New York and was partially validated via real home values. Yonto and Thill (2020) used three variables (occupation, education, median household income) to create an index to measure social status, census tracts with above average increases in the index were categorised as having gentrified. Another example is that of Holm and Schulz (2018) who analysed gentrification via their "GentriMap', which consists of two indices, a real estate index and a social index, designed to capture increases in real estate value and displacement-induced social upgrading respectively.

However, these approaches to change are also used in combination with another. For example, Atkinson (2000) used a threshold approach and index, he created an index of five weighted variables representing: the working class, unskilled labour, the elderly, lone parents, and the unemployed to explore . He found that gentrification was associated with above-average levels of losses of groups who have previously been associated with displacement.

Indices have been criticised for being ineffective at accurately capturing temporal change, due to their method of only providing static snapshots at points in time. So much so, that the index-based approach has been described as a hindrance to our collective understanding of neighbourhood change on a shorter-term basis (Hincks, 2015). Norman (2010) and Airgood-

Obrycki (2019) both use data of low temporal resolution of decadal timepoints only, with only two and five reference datapoints, respectively. This method smooths away shorter-term change patterns and neighbourhood dynamics. Consequently, little attention has been paid to realising the relationships between the mechanisms of change, and the transition of neighbourhoods between different states over time (Hincks, 2017).

2.4.3 Machine Learning Approaches

More recently, the field of neighbourhood analysis has developed models for predicting neighbourhood change with an interest in gentrification (Li, 2012). This is because other methods of quantitative analysis have been inadequate at capturing the complex spatial and temporal dynamics of neighbourhood change (Royall and Wortmann, 2015). These predictive models have been based on machine learning algorithms, where multiple variables operationalised as predictors are input into a model to predict the spatial extent of gentrification (Reades et al., 2019). They are often built with single classifiers or ensemble methods like Random Forests and often use census data for intercensal prediction (Hamnett, 2003). Although processes of gentrification are not uniform (Prouse et al., 2015), studies generally use demographic variables like the percentage of people from white ethnic groups, since gentrification often causes cultural displacement (Glaeser et al., 2018; Richardson et al., 2019).

Reades et al. (2019) provided the first and most comprehensive application of a regression-based machine learning model to predict future urban change (Thackway et al., 2021). They use Random Forests on 166 variables across domains including transport, housing, demographics, income, and wealth, from the 2001 and 2011 Census, to analyse existing patterns and processes of neighbourhood change. They go on to predict gentrification in London via the change within the socioeconomic status of a neighbourhood, and generated results of an adjusted r squared of 0.699. This result means that almost 70% of the variance within the target variable can be explained by the collection of predictor variables. It suggests that machine learning approaches to neighbourhood change analyses can model complex neighbourhood processes and predict future changes, with relatively good performance metrics. Palafox and Ortiz-Monasterio (2020) use data from four sources and Neural Networks (NN) to predict gentrification in Mexico City. The Local Interpretable Model-Agnostic Explanations (LIME) algorithm, is a model agnostic algorithm designed to increase the interoperability of black box ML algorithms (Ribeiro et al., 2016). LIME was used alongside NN to determine the importance of explanatory variables for each neighbourhood. Results found that although NN had reduced accuracy in identifying gentrification, the LIME algorithm aided with evaluating the process for specific neighbourhoods. Both studies, however, rely at least somewhat upon census data, with temporal intervals of 2001 and 2011 (Reades et al., 2019) and 2000, 2010, and 2016 (Palafox and Ortiz-Monasterio, 2020), which therefore limits their ability to analyse neighbourhood change via demographic drivers, which often occur on a smaller temporal scale than the decadal data of analysis (Thackway et al., 2021).

Thackway et al. (2021) forecasts future neighbourhood change in Sydney, in relation to gentrification and compare three different models, Random Forest, Gradient Boosting Machine (GBM), and Extreme Gradient Boost (XGBoost). They found that GBM performed the best, with a balanced accuracy of 74.7, and identified rings of gentrification expanding from the city centre. Balanced accuracy represents the mean of the sensitivity and specificity (true positives and true negatives respectively), and is used in an imbalanced dataset to provide a balanced view of the model's ability to correctly classify instances from different classes (Brodersen et al., 2010). Thackway et al. (2021) attempt to overcome the temporal issues of neighbourhood change research by introducing more non-census data alongside census data within their analysis. However, the temporal intervals of their data are unclear, so it is uncertain whether this study can accurately capture the more nuanced and local dynamics of gentrification in which it aims.

Machine learning based approaches therefore typically draw on census data to predict gentrification because it has always provided the best coverage. However, higher resolution data, and data from newer sources can also be easily utilised, which have the capability to highlight the more minor processes of change before the movement of people and businesses result in observable changes (Delmelle, 2022). For example, the language and phrasing written in residential rental advertisements can potentially be identified as an early indicator of gentrification (Delmelle, 2021). Geo-tagged tweets have also been found to augment the ability to identify processes of gentrification and displacement, finding that outsiders are more likely to visit neighbourhoods experiencing gentrification (Chapple et al., 2022). Twitter data can therefore utilise the dynamic and relational connections between people and places to aid the prediction of gentrification (Poorthuis et al., 2022). The introduction of Yelp data into predictive analyses found that gentrification is associated with an increase in the number of grocery stores, cafes, restaurants, and bars (Glaeser et al., 2018), whilst the addition of Airbnb data can help to quantify and track neighbourhood change via tourism, and the ensuing changes in housing affordability and demographics (Jain et al., 2021). This collection of research highlights the utility of both big data and machine learning as an appropriate approach for measuring neighbourhood change. These studies can generate unique results which the more traditional approaches to measuring neighbourhood change typically cannot.

Previous studies have however found that a threshold-based approach is more accurate at identifying gentrification than machine learning approaches (Liu et al., 2019), but others show that using more innovative approaches can improve the understanding of neighbourhood change more specifically (Palafox and Ortiz-Monasterio, 2020). Furthermore, it is often the case that several of these approaches are used in conjunction with one another. For example, Reades et al. (2019) created an index for socioeconomic status which they then used as the basis for their machine learning prediction. Additionally, gentrification studies can apply multiple thresholds by using a composite index compiled via the linear combination of socioeconomic variables (Hwang and Lin, 2016; Timberlake and Johns-Wolfe, 2017), or principal component analysis (Reades et al., 2019). These two different approaches can become inextricably linked, when a change in the index over time exceeds a threshold, gentrification occurs (Liu et al., 2019).

Most applications of machine learning in urban studies rely on supervised techniques (Grekousis, 2019), but Wang and Biljecki (2022) conducted a systematic review of 140 papers that used unsupervised learning in urban studies. They found that clustering is the most prominent unsupervised method and argue that unsupervised learning is key in learning spatial representations and the enhancement of spatial data (Wang and Biljecki, 2022). One final approach to neighbourhood change explored here is therefore more specifically cluster-based approaches.

2.4.4 Cluster-Based Approaches

The grouping of objects or observations is required in a wide range of domains from engineering, medical sciences, and the humanities and geography (Saxena et al., 2017). This grouping of objects or observations is achieved via a method called classification, in either supervised or unsupervised manners. Unsupervised classification sorts observations into groups, or clusters, by analysing the characteristics of the data and organising the observations into clusters based upon their similarity in a multidimensional feature space (Saxena et al., 2017). Clustering is the most established subcategory of unsupervised machine learning, with k-means being the most prominent one (Wang and Biljecki, 2022). k-means clustering is achieved by moving centroids and assigning observations closest to a given centroid into the same group, based upon the principle that observations within a cluster are more similar to one another, and are dissimilar to those in other clusters.

Geodemographic Classifications

Clustering in the domain of geodemographics and neighbourhood change uses a range of census variables, administrative data, and more recently some of the newer types of data (Big Data) (Singleton and Spielman, 2014), to assign observations into homogenous groups. They are typically underpinned with the k-means algorithm (see Hartigan and Wong, 1979), where n observations (neighbourhoods) are partitioned into k clusters (groups), where k is determined by the user, and with each observation belonging to the cluster centroid with the nearest mean in squared Euclidean distance. Here, the within-cluster variance is minimised, and the between-cluster variance is maximised, which makes the clusters as homogenous as possible (Abbas et al., 2009). Geodemographic classifications can also be nested with repeated clustering, and other algorithms, to create a hierarchical structure (Singleton and Longley, 2015).

Geodemographic classifications are valuable tools used within the private, public, and academic sectors, highlighted by their breadth of application. They are often used in the private sector for marketing, with applications including the targeted marketing of groups of people to promote product uptake (Sleight, 2004) and the identification of suitable locations for stores and warehouses (Leventhal, 2016). Geodemographic classifications also support academic research which can provide the basis for public policy (Harris et al., 2005). In doing so, they can suggest improvements for the allocative efficiency of public services (Longley, 2005) including:

- education (Singleton et al., 2012; Xiang et al., 2018);
- health, including the domains of:
  - cancer risk, screening and diagnosis(Bright et al., 2021; Nnoaham et al., 2010);
  - COVID risks (Grekousis et al., 2021; Grubesic et al., 2021);
  - dietary intake and access (James et al., 2021);
  - and long-term illness (Moon et al., 2019);

- road safety (Beecham and Lovelace, 2022);
- public transportation (Zhang et al., 2020; Liu and Cheng, 2020).

Geodemographic classifications have undergone numerous developmental stages throughout their history, which is explored in greater detail in Chapter 4. One developmental stage includes the addition of data beyond the traditional census sources to increase the temporal relevance of geodemographic classifications. For example Adnan et al. (2010) propose a real-time geodemographic classification that leverages live data sources, including transactional data, mobile device data, and social media activity. This approach theoretically captures and reflects dynamic consumer behaviour, preferences, and demographic patterns, thereby offering enhanced accuracy and timely insights. The development and adoption of real-time geodemographic classifications, however, require careful consideration of the computational requirements and suitable algorithms to handle large-scale and rapidly evolving datasets (Adnan et al., 2010). Commercial entities would be prime candidates for embracing real-time classifications, due to their abundant resources, but challenges remain in integrating real-time data sources, addressing privacy concerns, and developing robust algorithms capable of handling dynamic data (Adnan et al., 2010). These factors result in a lack of adoption of real-time geodemographic classifications by commercial entities.

Another development of geodemographic classifications relates to the increased spatial resolution of classifications. Initial demographic classifications were built at the ward level (Webber, 1977), but the creation of the Output Area (OA) after the 2001 Census led to the increased resolution of data, hence geodemographic classifications (Singleton and Spielman, 2014). Research has since gone further, with Burns et al. (2018) creating an individual level geodemographic classification. It was created under the assumption that an individual level classifications will reduce the impact of ecological fallacy and the subsequent generalisation of areas assigned to a cluster, resulting in the increased homogeneity of clusters. The geodemographic classification was created in much the same way as an area-based classification, but with individual level data. It was then joined to an individual population generated via microsimulation, synthesized to the OA. Next, individuals geographically referenced via OA, were then linked to the individual classification using common variables in order to assign a cluster code from the classification to each individual. However, in order for such a classification to be functional, it is aggregated back to an area (OA), highlighting the practical implications of such a classification. Additional concerns also relate to the ethics of using individual data, and that such high-resolution data may generate too much noise, making it difficult to pull out important patterns.

In more recent years, geodemographic classifications provide the basis for analysing change over time, based on the assumption that "having comparable geodemographic classifications over a period of time, will help demonstrate changes in socioeconomic and demographic structures" (McLachlan and Norman, 2021, p. 89). McLachlan and Norman (2021) created three directly comparable geodemographic classifications from three consecutive censuses. They achieved this by creating a geodemographic classification on the 2011 census data, and used these cluster centres as the starting point for the clustering of the 1991 and 2011 census data, where LSOAs were assigned to the nearest cluster centre, without any iterative cluster reassignment. Their results found that the majority of neighbourhoods were allocated into the same clusters over time, but they were able to characterise three types of change; socio-economic polarisation, the growth in the number of neighbourhoods with non-white residents, and the growth of neighbourhoods associated with urban areas. These neighbourhood changes allude to different neighbourhood processes, including spatial polarisation, a process that includes diversification, and a type of urbanisation. Webber and Burrows (2018) predicted that future geodemographic classifications will start to focus on the social changes of a neighbourhood, rather than the classification methodology, and McLachlan and Norman (2021) demonstrate the importance of neighbourhood processes in view of neighbourhood change. The current trend in research is therefore leaning towards attaining a deeper understanding of the processes of change.

Although McLachlan and Norman (2021) take a deeper consideration and approach to temporality than other similar geodemographic-based research, it is still conducted in a static manner, via three lots of cross-sectional decadal data, which leads to one of the most substantial limitations of geodemographics (and other approaches to measuring neighbourhood change), their lack of temporal concern. Therefore, although McLachlan and Norman (2021) identified neighbourhood changes on a larger temporal timeframe, they are not able to capture the more local and dynamic changes associated with changing neighbourhoods and the community-level neighbourhood processes that produce such changes. Thus, not only are the data static, but they may also not necessarily coincide with the periodicity of the neighbourhood changes, that is, the start, peak, decline, and end of the process (Comber and Wulder, 2019). Subsequently, this may be one explanation for why the majority of neighbourhoods remain stable over 30 years.

The geodemographic approach to analysing change over time, just like indices, threshold-based and machine learning approaches, typically compares neighbourhoods at only two points in time. Several studies have previously declared how such an approach neglects the underlying dynamic character of neighbourhoods (Hincks, 2015; Hincks, 2017) because various aspects of neighbourhoods change at different rates. Being sensitive to this within analyses is vital for understanding the nuances of neighbourhood change at and on smaller temporal scales, and for generating accurate neighbourhood trajectories (Reibel, 2011; Zwiers et al., 2017). Geodemographic classification also shares similar considerations to the threshold-based approach, in which smaller signals of change are not captured. However, both limitations are exacerbated with a classification approach to analysing change.

2.4.5 The Need for an Alternative

Each of the methods for measuring neighbourhood change explored throughout Section 2.4 has its merits. Their relatively simple methodologies mean they are easy to implement, and subsequent results are easily interpretable. They can be applied to many different domains to measure many different things. However, these quantitative approaches to analysing gentrification share some limitations.

Temporal Considerations

One major limitation relates to their approach, or more so lack of, to temporality. Neighbourhood dynamics have been positioned throughout this literature review as the core concern of neighbourhood change, and neighbourhood change research. However, all approaches described are constrained by their reliance upon and use of data with a temporal resolution that is not capable of accurate change detection concerning smaller, more dynamic neighbourhood change (Hincks, 2015; Hincks, 2017; Zhu, 2017; Comber and Wulder, 2019). Thus, the application of understanding processes of neighbourhood change to predict future trajectories of change remains limited (Thackway et al., 2021). This is because neighbourhoods are dynamic and constantly changing, yet approaches to analysing neighbourhood change are addressed via cross-sectional analyses (McLachlan and Norman, 2021), operationalised over periods of around a decade apart, using only two timeframes for analysis of a dynamic process (Reibel, 2011; Barton, 2016). Studies use data of a temporal resolution (decadal) that is unable to capture, understand, and predict neighbourhood dynamics that occur on a smaller temporal scale. This represents a clear disjoint between the temporal properties of the mechanism deployed to research the neighbourhoods and the temporal properties of the neighbourhoods themselves. This is represents the fundamental flaw to neighbourhood change analysis, attempting to understanding dynamics with static data (Reibel, 2011). Such an integral flaw to the methodological approach of measuring neighbourhood change has been described as a hindrance to our collective understanding of neighbourhood change on a shorter-term basis (Hincks, 2015).

This could be attributed to the lack of consideration of the synchronicity between the process phase and the temporal interval of data (Hincks, 2015; Comber and Wulder, 2019). Therefore, to generate more accurate presentations of neighbourhood change, there is a critical need for researchers to account for the smaller periodic population and neighbourhood changes. This requires imparting greater importance upon the synchronicity between the process phase in which they are measuring, and their measurement frequency of analysis (Comber and Wulder, 2019), and where this cannot be achieved due to data availability restrictions, data of a greater spatiotemporal resolution.

Methodological Considerations

Moreover, cluster-based approaches to measuring neighbourhood change over time (as described within Section 2.4) have an information reductive nature which limits their ability to capture dynamic changes, exacerbating the temporal limitations of the approaches.

For geodemographic classifications, this is their Boolean allocation of observations (neighbourhoods) to the cluster centroids they are nearest to in a multidimensional feature space. The purpose is to provide a convenient way of grouping observations such that each cluster is comprised of observations with similar characteristics (Hjørland and Pedersen, 2005), thus the parsimony is by design, but leads to further limitations. Firstly, the relative position of an observation within the cluster is masked; observations with different levels of 'belonging' are given the same cluster label. This, masks the inherent uncertainty associated with any hard classification (Fisher and Tate, 2015) and compounds the implication of information loss (Grekousis and Thomas, 2012). Secondly, it is possible that these clusters may just reflect delineations within the data as opposed to any natural grouping experienced in the real world (Everitt and Gill, 1993; Singleton and Longley, 2009). These limitations alone can impact the practicality of employing classification-based approaches, but they are again intensified when adopting geodemographic classifications for analysing neighbourhood change over time.

The notion of cluster-to-cluster change is reductive since a change in class or label is only recorded when it is statistically significant and surpasses a threshold in both space and time (Comber and Wulder, 2019) resulting in information loss. This information loss relates to their: i) relative change in multidimensional feature space between states over time, ii) uncertainty in class membership iii) condition and quality of the clusters, and iv), any smaller changes that represent signals of change yet to come. Chapter 4 *Extending Geodemographics Using Data Primitives: A Review and a Methodological Proposal* provides a more detailed example of how a cluster-based approach to change is reductive and leads to information loss.

Such information loss is also present within the threshold approach to change. All changes below the specified, and often arbitrary, threshold are not captured within the analysis. This could therefore also miss smaller changes that represent signals of change yet to come. Finio (2022) undertook a systematic review of the methods used to identify and quantify gentrification. They found that 82 papers (48% of papers they reviewed) used a discrete threshold without the use of any other method like indices, or a more complicated scheme, for measuring gentrification. Consequently, their capability of accurately capturing dynamic changes is restricted. Though the methodological complexity for measuring gentrification is increasing over time (Finio, 2022), basic threshold approaches without the use of other, more methodologically advanced approaches, remain inadequate at identifying and measuring accurate neighbourhood change.

Likewise, index-based approaches to change also present some information loss, due to their nature of reducing multivariate information to just one score. This reliance on one score conceals complexities, particularly when measuring change, since they only provide static snapshots at points in time (Hincks, 2015). Consequently, little attention, thus understanding, has been paid to realising the relationships between the mechanisms of change, and the transition of neighbourhoods between different states over time (Hincks, 2017). Furthermore, their pragmatic decisions in the weighting of the components are essentially arbitrary and represent a crude attempt in an otherwise 'sophisticated' approach to quantify deprivation, and indeed other types of neighbourhood change. This subjectivity can leave the methodology and subsequent results open to scrutiny (Clelland and Hill, 2019), particularly when specific indices have politically driven justifications (Deas et al., 2003). Finally, index scores are sensitive to the methods used within their creation, but regardless of the specific methods used (for example PCA, standardisation, min-max), the index remains subject to uncertainties which can lead to the fluctuation of scores and rankings (Dialga and Thi Hang Giang, 2017).

To summarise, since neighbourhood processes like gentrification are dynamic, the approach to analysing a neighbourhood's state over time at only two periods, with approaches that restrict their capability of accurately capturing neighbourhood change due to their methodological limitations which lead to information loss, is fundamentally flawed. They are not capable of accurate change detection, and cannot generate reliable insights into social research because vital information about the process such as the signals of change, and a deeper understanding of the dynamics of change is missing. Consequently, there is a need for a more suitable and flexible approach to neighbourhood change, which is sensitive to contextual, relational, and multi-scalar changes. This therefore highlights a methodological gap in the research, whereby methodological advancements within the field of neighbourhood change can increase the temporal relevance of approaches to neighbourhood change and subsequent policy (Compton et al., 2019), while supporting explicit analyses of social changes. Such a gap affords the discovery of deeper understandings of the processes driving the more nuanced and local changes in a neighbourhood (Adnan et al., 2010; Weiser and Frank, 2010), whilst also having the foundation to support improved prediction capabilities that can inform spatial planning, and other such appropriate policies (Christakos et al., 2002). One potential for this methodological advancement is data primitives.

2.5 Data Primitives

Neighbourhood changes arise from the accumulation of the effects of different neighbourhood-level processes; thus, an appropriate method is required to identify the different changes occurring and to examine which neighbourhood process is impacting neighbourhood dynamics. Data Primitives is an alternative, extended approach for identifying, analysing, quantifying, and predicting the spatial and temporal scope of neighbourhood change, with an emphasis on the processes driving the neighbourhood change. As an approach, data primitives enable the consistent measurements of and for the fundamental characteristics of the different processes driving these neighbourhood changes, which facilitates a deeper understanding of neighbourhood dynamics and provides a substantial basis for generating trajectories of future neighbourhood change.

This explicit consideration of processes driving the distribution of people is not a new concept. Early works in human ecology emphasised processes of change, and Webber and Burrows (2018) predicted the change in direction of geodemographic classifications away from methodology (and data) to the focus of social dynamics and their subsequent impacts upon the neighbourhood (changes) in which geodemographic classifications cluster upon. Furthermore, McLachlan and Norman (2021) suggested the presence of three longer-term processes of change (socio-economic polarisation, diversification, and urbanisation), and their impact on neighbourhood composition. Therefore, these studies suggest that an approach that places the neighbourhood processes at the centre of analyses regarding neighbourhood change would generate a deeper understanding of neighbourhood dynamics. Data primitives have the capability to generate novel understanding concerning the spatial organisation and temporal extent of neighbourhood processes, and their subsequent impact upon society.

Data primitives are an established approach in remote sensing, developed to overcome issues when translating between different land use classifications in remote sensing (Wadsworth et al., 2008) and have been used to link and separate land cover / land use semantics (Comber, 2008; Comber and Kuhn, 2018). Data primitives in remote sensing studies are derived from the physical properties of electromagnetic radiation (like radiance and reflectance), spectral bands (the wavelength of electromagnetic radiation), and geographic information (like elevation and slope), to capture and represent information like "naturalness', "vegetation height', "homogeneity of appearance', and "human activity' (Comber, 2008). These spectral, spatial, and contextual characteristics enable researchers to more accurately classify land into suitable land use and land cover classes.

The basic idea behind data primitives is that they represent the measurements or variables that capture the full range of fundamental dimensions or characteristics of the process process under investigation. They are the building blocks that underpin the concepts of the phenomenon (Comber, 2008). In remote sensing applications, these are tangible components, but with the adoption of the approach within the neighbourhood change domain, such building blocks are more abstract and conceptual, and subsequently requires additional consideration. For example, data primitives for neighbourhood change research may represent ethnic composition, levels of education, income inequality, house price, and so on.

Nevertheless, once the neighbourhood process has been conceptualised and appropriate data primitives have been identified to represent the sociodemographic components of the process, the approach can be adopted and temporally extended, to capture neighbourhood dynamics. This can be achieved by investigating the changes in an area's position within n-dimensional feature space composed of n data primitives, over time. These changes in position in multi-temporal feature space could be used to infer the changes in character experienced by a neighbourhood throughout time. In remote sensing works, Change Vector Analysis is often just in conjunction with data primitives, and analysis has shown that the angle and magnitude of such positional changes within the n-dimensional feature space can be used to infer the nature of change (Xu et al., 2019). Thus, the examination of the shifts via the data primitives and change vectors' angle and magnitude, which generate rich information, could therefore be used to derive neighbourhood dynamics, and quantify process cycles and potentially predict future states (Gray et al., 2021). This data primitive framework therefore reframes the quantitative analysis around change away from stable labels, to an approach where the data, tools, and purpose align.

There are however challenges in adopting a data primitive approach for neighbourhood research. The type of processes and urban dynamics that can be identified are dependent on the variables that are selected as primitives, as these need to capture the core drivers that characterize the neighbourhood process under investigation. Since these are more abstract and conceptual than remote sensing uses, and neighbourhood processes themselves are complex and multifaceted, there may be subjective choices in the selection of data primitives that may not always align with other researchers, particularly when measuring situational processes like gentrification. Furthermore, the shifts in an area's position in the feature space need to be filtered to determine potentially meaningful changes. Chapter 4 explores the theoretical foundation and methodological proposal of data primitives in greater detail, with a city-based illustration, whilst Chapter 5 explores the meaningful changes associated with gentrification.

Regardless of such challenges, data primitives, offer the opportunity for a methodological advancement to neighbourhood change, providing an approach that captures dynamic and relative change that can be analysed to generate novel understandings of neighbourhood change via the lens of neighbourhood processes. Data primitives have the capacity to separate defined neighbourhood processes, like gentrification, from the accumulative neighbourhood change (the overall transformation of a neighbourhood over time). This separation allows for the quantification of the process and its impacts upon the neighbourhood to be characterized, analysed, and predicted more directly. Such an approach can aid with unpacking the complexities of neighbourhood change.

2.6 Summary and Problem Statement

To summarise, neighbourhood processes continually alter neighbourhood characteristics (Prouse et al., 2015) such that the neighbourhood's overall character (social, socioeconomic, cultural, physical, and political) transforms over time. Researchers have already advocated for dynamic area models and approaches to analysing change, to account for and capture the dynamic processes that occur over different temporal scales (Reibel, 2011; Hincks, 2015; Barton, 2016). Yet, although neighbourhood research is growing increasingly more temporal (Chapple and Zuk, 2016), approaches to neighbourhood change research are typically constrained to cross-sectional analyses. Further, these cross-sectional analyses commonly occur at only two points in time (Reibel, 2011; Barton, 2016), which are incapable of capturing the neighbourhood dynamics that occur on a smaller temporal scale than that of their data collection. This is the central contradiction to neighbourhood analysis; the effort to explain change over time using small area data – at temporal intervals - that cannot measure change over time directly (Reibel, 2011). Therefore, the most notable gap in the literature, and integral to advancing neighbourhood research, is the need for a more suitable approach to sufficiently capture the neighbourhood dynamics that occur on a more granular temporal scale.

Data of higher spatiotemporal resolution offer greater insights into societal processes and improve analysis (Longley, 2012). Thus, the integration of multi-temporal data into neighbourhood research via data primitives, would not only improve the operationalisation of temporal thresholds for analysing change but generate much deeper insight into the processes of neighbourhood change in both their spatial and temporal aspects (Reibel, 2011; Longley, 2012; Comber and Wulder, 2019). The data primitive approach is one that aims not only aims to improve the identification, quantification, and prediction of neighbourhood change via the lens of neighbourhood processes, but one that aims to generate deeper understanding and improve explanations.

Chapter 3: Methodology

This Chapter outlines the data and methods used in this research. Constituent papers each contain the data and methods used in full detail, including any pre-processing, implementation and interpretation. By contrast this Chapter provides a comprehensive overview and justification for the data and methods used.

3.1 Scope of Study

3.1.1 Spatial Scope

There are multiple spatial scopes of study within this research. The first paper in Chapter 4, formally introduces data primitives in the context of neighbourhood change and includes a short data primitive example for Nottingham, a city in East Midlands, England. The second paper in Chapter 5, provides an in-depth analysis and widens the spatial scope from a city to a regional case study. It focusses upon South Yorkshire in Yorkshire and the Humber, North England. The third paper in Chapter 6, applies the validated dataset from the South Yorkshire data to an extend analysis of a national case study of gentrification in England.

Each scope relates to the purpose of the paper. The initial city-based study introduces the approach after a review of geodemographic classifications, offering an example of how it could be used in practice. It uses a total of nine data primitives that capture a range of change processes, seen in Table 3.2, to provide a broad overview of neighbourhood change in Nottingham between 2010 and 2016, a period between two censuses. This represents a period with only one census in it, so traditionally, researchers have no view of how the city might be changing during this time. The regional case study shows a full empirical analysis while grounding the approach for neighbourhood change. It conceptualises gentrification, identifying a set of specific, core data primitives for analysing the presence and temporal properties of gentrification.

These data are displayed in Table 3.1 and have extended temporal boundaries of between 2010 and 2019, representing 10 years of data, but still within the traditional decennial, census-to-census based analyses of change. The national study extends the analysis further and reinforces the suitability of the approach for analysing neighbourhood change, providing an English case study on gentrification also between 2010 and 2019. This study uses the core data primitives and a range of descriptive data, seen in Table 3.1, identified as useful in the regional analysis. These three papers together introduce, position, and emphasize the suitability of data primitives as a methodological advancement for analysing neighbourhood change over time, via the manifestation of neighbourhood processes.

3.1.2 Temporal Scope

There are two temporal scopes within this research, 2010 to 2016, and an extended scope of 2010 to 2019. There are 7 years of annual data within the initial analysis, and then 10 years of annual data within the following two pieces of work. These reflect the scope of the study, but also the availability of data at the time of analysis. Because the data primitive approach describes the state of an area in multivariate feature space, and compares these states of multiple time periods, the data is required at a sufficient spatiotemporal resolution. This research considers small-area (LSOA) - annual data as a sufficient resolution for neighbourhood change research. Such data requirements enable smaller changes that not previously captured - to be captured. The temporal boundaries of studies are therefore constrained to the availability of the required data of a suitable spatiotemporal resolution.

One of the main datasets used within this analysis is the Neighbourhood Mobility Index (representing neighbourhood churn – the in and out migration observed within a neighbourhood), created by the Consumer Data Research Centre (CDRC). At the time analysis was undertaken for Chapter 4, this dataset was only available up until 2016, hence the end temporal boundary of 2016. By the time the analysis was undertaken for Chapter 5 and Chapter 6, this dataset was updated to 2020. In the first iteration of the dataset, change was measured as a proportion of change from 1997. However, as seen in Table3.2, the measurement of this dataset was changed to the proportion of change in relation to 2020, with 2020 being the baseline thus no valid levels of change. Therefore, the end temporal boundary for these analyses was 2019. This does however demonstrate that the availability of data at the appropriate spatial and temporal resolution, (small area – annual) is likely to increase as time goes on.

The overall temporal boundaries of this research – a decade are therefore like other gentrification studies (Barton, 2016), but the annual temporal resolution of this research is greater than their decennial temporal resolutions. The greater temporal resolution enables the detection of not only intercensal changes, but also the smaller signals of change (Zhu, 2017; Comber and Wulder, 2019).

3.2 Data

Data are collected and collated from a range of mostly open sources and are detailed in Table 3.1. These reflect a range of socioeconomic data identified as the fundamental characteristics of gentrification, alongside a range of descriptive data including distances to a number of transportation links and greenspaces. Table 3.1 also details which data was used in each of the three papers, the data source, and also the original data source, indicating what the data was created from. Two components are used from the Access to Healthy Assets & Hazards Index, created by researchers at the CDRC. Both the Normalised Difference Vegetation Index and the distance to bluespaces are used within Chapter 6, both of which originate from different sources.

Table 3.1: Data Sources

Data	Chapter(s)	Source	Original Source
Population Estimates	4	www.ons.gov.uk	ONS
Population Density	4	Derived from above	ONS
Housing Affordability	4	www.ons.gov.uk	Census
Disability Living Allowance	4	stat-xplore.dwp.gov.uk	DWP
Jobseekers Allowance	4	stat-xplore.dwp.gov.uk	DWP
House Price	4 5 6	www.ons.gov.uk	ONS
Professional Occupation	4 5 6	www.nomisweb.co.uk	ONS
Residential Mobility Index (Neighbourhood Churn)	4 5 6	data.cdrc.ac.uk	Electoral Register
Black and Asian Ethnicities	4 5 6	data.cdrc.ac.uk	Electoral Register
White British Ethnicity	4	data.cdrc.ac.uk	Electoral Register
Rural-Urban Classification	5 6	geoportal.statistics.gov.uk	Census
OS Open Roads (Motorway Junctions)	5 6	https://os.datahub.os.uk	Ordnance Survey
OS Open Greenspace	5 6	os.datahub.os.uk	Ordnance Survey
National Public Transport Access Nodes	5 6	beta-naptan.dft.gov.uk	DFT
Normalised Difference Vegetation Index	6	data.cdrc.ac.uk	Sentinel
Bluespace	6	data.cdrc.ac.uk	Open Street Map

Table 3.2: Data Measurements and Resolution

Data	**Unit**	**Temporal Res.**	**Spatial Res.**
Population	Number	Annual 2010-2020	LSOA
Population Density	People per 1km^2	Annual 2010-2020	LSOA
Housing Affordability	Ratio of House Price to Residence-Based Earnings	Annual 2010-2020	LA
Disability Living Allowance	Proportion	Annual 2010-2020	LSOA
Jobseekers Allowance	Proportion	Annual 2010-2020	LSOA
House Price	Median House Price (£)	Annual 2010-2020	LSOA
Professional Occupation	Proportion	Annual 2010-2020	LSOA
Residential Mobility Index	Proportion of change in relation to 1997 (Ch. 4)	Annual 1997-2016	LSOA
	Proportion of change in relation to 2020 (Ch. 5 & 6)	Annual 2010-2020	
Black and Asian Ethnicities	Proportion	Annual 2010-2020	LSOA
White British Ethnicity	Proportion	Annual 2010-2020	LSOA
Rural Urban Classification	Factor - RUC Group Assignment	2011	LSOA
OS Open Roads (Motorway Junctions)	Point	November 2022	U.K.
OS Open Greenspace	Polygon - Greenspace Function & Area (km^2)	October 2021	U.K
National Public Transport Access Nodes	Point - Distance to Transport Nodes (m)	August 2022	U.K

One variable, Professional Occupation, was also derived from the Standard Industrial Classification (SIC) (2007) Occupation by Industry dataset. SIC is a method for classifying businesses by economic activity (Hughes et al., 2009). However, this variable was based upon subjective decisions regarding the makeup of the more professional industries, with Table 3.3 showing the industries included. However, this is not entirely ideal, as there are different levels of roles within each of the industries. Thus, there could be more "professional" roles within some of the industries considered as lower skilled, and likewise there may be lower, entry-level roles within the professional industries. The Standard Occupation Classification (SOC) is a system used to classify people according to their job, based on the level and specialisation of skill, ranging from “Managers, directors, and senior officials’, to “Elementary occupations’ (Statistics, 2023). This might have been a better classification to use over the SIC, but this data, but this was not available at the required spatiotemporal resolution. Furthermore, the SIC can also identify the creative industries, which are associated with driving neighbourhood change via gentrification (Behrens et al., 2022), which is not represented within the SOC.

Table 3.3: The Professional Occupation Variable

Professional Industries
Information and communication (J)
Financial and insurance activities (K)
Real estate activities (L)
Professional scientific and technical (M)
Administrative and support service (N)
Public admin and defense (O)
Education (P)
Human health and social work (Q)
Arts, entertainment and recreation (R)

3.2.1 Consumer Data Research Centre

Unfortunately, not all data required for this analysis was freely available. Two datasets for this research were obtained from the CDRC, a research centre that was established to increase engagement between academia, social sciences, and industry, by using consumer data for research purposes (Vij, 2016). The CDRC collaborate with a number of large data providers to provide a range of data in themes of population and mobility, retail futures, transport and movement, finance and economy, and digital, with a range of accessibility levels.

Both datasets obtained from the CDRC are derived from the Linked Consumer Registers (LCR), which link the open electoral register with consumer registers supplied by value-added resellers (Lansley et al., 2019). The LCR has over 885 million records, providing annual snapshots of the adult population on the open register from 1997 (Lan et al., 2022). It provides researchers with access to data and analysis ready products that would not usually be accessible to academics.

The Residential Mobility Index utilises over 32 million records of the LCR and the Unique Property Reference Number of households to determine origin-destination movement of individuals (Chen et al., 2021). As individuals register to vote using their home address, the LCR can track an individual’s residential movements by following names and updates to their addresses on the open register, including household formations, dissolutions, and relocation. Linked with data like social deprivation, the IMD, and energy performance certificates of households, the LCR allows researchers to analyse the changing composition of a neighbourhood, including the proportion of neighbourhood change related to a neighbourhood's in- and out-migration (neighbourhood churn). The Residential Mobility Index is used as the Residential Churn data primitive in Chapters 4, 5 and 6. As mentioned in Section 3.1.2, two iterations of this dataset have been used throughout this research. Version 1 was calculated as a measure of change *from* a 1997 baseline to 2016, whilst Version 2 was calculated as a measure of change compared *to* a 2020 baseline.

Though the Residential Mobility Index is open data, the Modelled Ethnicity Proportions are safeguarded data derived from the LCR, which can only be obtained after the approval of a mandatory application. The Modelled Ethnicity Proportions were created using the ethnicity estimator software, based upon work by Kandt and Longley (2018), where ethnicity is estimated using cluster analyses. It uses the forename and surname of individuals on the open register (LCR), which are used to generate an ethnicity estimate. Surnames have remained concentrated throughout the world, such that they can be used to characterise population origins (Kandt and Longley, 2018), while forenames are chosen according to prevailing local, cultural and temporal preferences (Finch, 2008). Thus, names can be indicative of ethnic background, gender, and age (Kandt and Longley, 2018). The ethnicity estimator examines the association between forenames and surnames in the LCR and self-ascribed ethnicity as per 2011 Census data via unsupervised learning, where names are assigned to clusters representative of their ethnic background (Kandt and Longley, 2018). At an aggregate level, the ethnicity estimator software enables the modelling of the ethnic composition of neighbourhoods, and their change throughout time, hence the Modelled Ethnicity Proportions. One variable was derived from this dataset, the percentage of people from Black and Asian ethnicities, of all ages.

The availability and quality of data products derived from the Linked Consumer Registers, are dependent on the open register, which raises concerns regarding potential limitations. Individuals have the option to remove themselves from the open register, potentially affecting the completeness of the LCR and consequently impacting the derived products' quality. In contrast, other data sources, particularly administrative sources, are expected to improve in resolution and availability as the prevalence of spatiotemporal data increases, thus data are likely to become more primitive.

The CDRC also created the Access to Healthy Assets & Hazards Index, a multi-dimensional index for measuring how "healthy" a neighbourhood is, considering access to: the retail environment, health services, the physical environment (including bluespaces and NDVI - which represents passive greenspace), and air quality. The NDVI and distance to closest bluespace variables were used as descriptive variables in Chapter 6.

3.3 The Data Primitive Approach

For the implementation of data primitives in neighbourhood change research, the data primitives first must be conceptualised and identified such that they capture the key drivers that characterise the process under investigation, in order to capture relative changes to and within neighbourhoods. The data underpinning them should be collected at a suitable spatiotemporal resolution, ideally small-area - annual, to enable such detailed temporal analyses. Once the process is conceptualised and the relevant data primitives are identified and collated, the neighbourhood can be described in terms of its position in n-dimensional feature space, hence the state of the neighbourhood. When examining neighbourhood change via gentrification as conceptualised in Chapter 5, appropriate data primitives and their respective directional changes are: house price (increase), professional occupations (increase), neighbourhood churn (increase), and Black and Asian ethnicities (decrease). House price increase serves as an indicator of rising property values, reflecting the economic transformation of neighbourhoods (Lees et al., 2010). The increase in professional occupation signifies shifts in the occupational composition of the population, highlighting changes in the socio-economic profile of gentrifying areas (Van Ham et al., 2020). Neighbourhood churn captures the turnover and displacement of residents, indicating the dynamic nature of gentrification (Yee and Dennett, 2022). Lastly, the decrease in black and Asian minorities highlights the demographic changes associated with gentrification (Huse, 2018).

An area's position in n-dimensional feature space composed of the data primitives, hence a gentrification space, is described at each annual or periodic interval, providing a new description of the state of the neighbourhood. Exploring the changes in state can quantify the amount of change experienced by the neighbourhood, and these transitions between states can indicate neighbourhood dynamics, hence the process phase or cycle of the process under investigation, which can then undergo further spatiotemporal analysis.

The data primitive approach to neighbourhood research on the whole, has two phases, the first seeks to quantify neighbourhood change, and the second seeks to describe the change based on a set of neighbourhood characteristics and their respective change vectors (Comber, 2008), as explored in Chapter 5. These neighbourhood dynamics can also provide the basis for predicting future states of neighbourhoods, and even entire cycles of the process (Gray et al., 2021).

3.4 Methods

All quantitative analyses were conducted using R in R Studio. The following sections cover each of the methods used within the respective papers, and their justifications.

This research takes a data primitive approach to analysing neighbourhood change via the lens of gentrification. There are a number of steps that are required in order to properly implement the approach. The first of which is a literature review, detailed in Section 3.4.1 to fully conceptualise the process and identify the core data primitives. Once the appropriate data has been collected and collated (Section 3.2), the first data primitive function (the process function), identifies areas experiencing changes associated with gentrification, and is described in Section 3.4.2. The second is the periodicity function, described in Section 3.4.3 and captures the temporal properties associated with each identified cycle of change. These functions also calculate change vectors. Change Vector Analysis is a separate, yet essential component of the data primitive approach that measures change as a vector between multi-temporal states, via the vector angle and magnitude, and is described in Section 3.4.4. Once these cycles of gentrification have been identified they can be explored, analysed, and predicted upon. They are thus used as a training dataset to predict the spatial and temporal extent of gentrification in England, investigated via three different ML algorithms, described in Section 3.4.6.

3.4.1 Literature Review

A literature review is a method in and of itself and is the building block of all academic research activities (Snyder, 2019). An effective review as a research method creates a firm foundation for advancing knowledge and facilitating theoretical development (Webster and Watson, 2002). The first paper of this research presented in Chapter 4 is based on a literature review of geodemographic classifications and their limitations, which gave way to the introduction of a new methodological approach to analyse neighbourhood change. Thus, Chapter 4 indeed does provide a foundation for advancing knowledge, by providing a theoretically sound grounding for the need for data primitives, given the limitations of geodemographic classifications. Chapter 4 and Section 2.6 details the data primitive approach, and the very first step is to determine the fundamental drivers of the specific neighbourhood processes of interest, via a literature review. Without a literature review to conceptualise the process and its data primitives, the data primitive approach will not be capable of measuring neighbourhood change via the desired neighbourhood process. The literature review is therefore an integral part of the approach.

Although literature reviews alone provide great foundations for development, Chapter 4 also consists of an illustrative example of the data primitive approach, to provide as proof of concept that such methodological advancement is feasible and practical. It uses a range of the methods described below including the pre-processing and change vector analysis to generate initial insights into the use of data primitives for analysing neighbourhood change in a city-based example of Nottingham, U.K.

Further, retaining the conceptual framework of gentrification, as determined via a literature review, helps focus the study on the essentials (or the fundamental driving characteristics) of the phenomenon (Liu et al., 2019).

3.4.2 Process Function

The first step for identifying changes associated with gentrification is to input data into the process function. Given access to the appropriate data, the function can be initialised to identify changes associated with a range of neighbourhood processes, like gentrification, urban decline, suburbanisation, and so on. The function requires that the data primitives and their respective directions of change be specified and is used within the research in Chapter 5 and 6.

Table 3.4 shows the algorithm of this process function. It seeks changes associated with the specified neighbourhood process between two given years, Time 1, and Time 2. When all of the specific changes are met between these two time periods, then the absolute sum of the z-scores is returned as the 'gentrification score', an index to represent the magnitude of change experienced by the neighbourhood. Thus, this function can identify neighbourhoods that have experienced changes associated with gentrification, and their magnitude of change between two time periods.

Table 3.4: Process Function Algorithm

Algorithm 1: Process - process_func() function

1. Define the function process_func with four parameters: time1, time2, varnames, and relations, where varnames are the data primitives, and relations are their respective directional changes.
2. Check that the lengths of the varnames and relations parameters are the same. If not, throw an error message.
3. Convert the relations parameter to lowercase.
4. Check that the relations parameter contains only the words "increase" or "decrease". If not, throw an error message.
5. Select the columns in time1 and time2 that match the variables listed in varnames.
6. Calculate the difference between the selected columns in time2 and time1 and assign it to t12.
7. Determine the number of terms in t12 and assign it to n_x.
8. Convert the relations parameter to logical expressions (">" for "increase" and "<" for "decrease") and assign it to logic_x.
9. Build a logical expression for each term in t12 using the corresponding element in logic_x and concatenate them using the "&" operator. The resulting expression is assigned to expression_x.
10. Trim the final "&" from expression_x.
11. Define a new dynamic function process_inc_dec_func that evaluates the logical expression in expression_x using the input vector x.
12. Apply the process_inc_dec_func function to each row of t12 and store the results in res.
13. Calculate the absolute sum of each row in t12 and multiply it by the corresponding value in res. This will return 0 for areas that have not experienced the process. The resulting vector is assigned to res_val.
14. Return res_val.

This function can be used in different ways, including inserting into a loop to iterate over each combination of time points within the data's temporal boundaries. For the 2010-2019 study period, this returns a total of 45 gentrification scores for 45 unique time periods (every one year, every two years, every three years and so on). This can seek the period(s) in which the gentrification process was present in a neighbourhood, by identifying the time periods that had a non-NA gentrification score. If several are returned, then the magnitude of the gentrification scores can give a potential indication into the gentrification cycle within that neighbourhood. Also, the total number of gentrification scores per neighbourhood can be summed, under the assumption that a neighbourhood with a greater gentrification count is more likely to experience gentrification than a neighbourhood with a lower gentrification count, or a 0 gentrification count. However, although this is operated in a more temporally considerate manner, the approach is still static, considering only one singular change between Time 1 and Time 2. Thus, another process, the periodicity function was created. This function is used within the research found in Chapter 5, and Chapter 6.

3.4.3 Periodicity Function

The periodicity function was designed and created to extend the functionality of the process function. Its purpose is to identify and analyse sequential changes associated with gentrification, given a starting year, within the multitemporal data. It is these sequential changes, a continuous type of change, which support the existence of the process of gentrification, opposed to the changes occurring in one period by chance. These captured sequential changes may therefore be reflective of an area's cycle of gentrification. The algorithm of this periodicity function is displayed in Table 3.5.

Table 3.5: Periodicity Function Algorithm

Algorithm 2: Periodicity - get_seq_and_peak() function

1. Define the function get_seq_and_peak() that takes in two arguments: x and i, where x is the dataframe and i is the starting year of interest.
2. Convert the input x into a vector using the unlist() function.
3. Get the indices of the attributescolumns that start with i, the starting year, using the which() function and starts vector.
4. Find the peak year and its value by doing the following:
   1. Get the sequence of value-to-value difference by subtracting the current value from the previous value using the diff() function, and adding a 0 to the start to make it the same length as the rest of the sequence.
   2. Find the index of the first negative difference using the which() function and subtract 1 from it to get the index of the peak year.
   3. Get the year of the peak by indexing into the ends vector using the peak index.
5. Find the start year of the change by doing the following:
6. Find the cumulative maximum of the sequence of interest using the cummax() function.
   1. Get the index of the first element in the cumulative maximum sequence that is greater than 0 using the which() function.
   2. Get the year of the start of the change by indexing into the starts vector using the start index.
   3. Find the duration to peak by subtracting the start year from the peak year.
7. Find the end year and the cumulative gentrification score by doing the following:
   1. Remove any missing values from the sequence of interest using the na.rm = TRUE argument in the sum() function.
   2. Find the cumulative gentrification score by summing the sequence of interest from the start year to the peak year using the sum() function.
   3. If there is a peak year, find the end year by doing the following:
      1. Create a vector of end years using the ends vector and indexing it from the peak index to the end of the sequence of interest.
      2. Find the index of the minimum value in the sequence of interest from the peak year to the end of the sequence using the which.min() function.
      3. Get the year of the end of the change by indexing into the end year vector using the end index.
      4. If there is no peak year, set the end year and the cumulative gentrification score to NA.
8. Return a vector of the start year, peak year, duration to peak, end year, and cumulative gentrification score using the c() function.

This function is used for every start year from 2010 up to 2016 (under the assumption that any starting in 2017 will not have a completed cycle by 2019), several overlapping cycles of gentrification for the same neighbourhood may be identified (for example 2010-2016, 2011-2016). In these cases, the cycles with the largest cumulative changes are selected, which most often also incorporates the cycle with the earliest starting year. However, there are potential improvements to this pragmatic decision, which are explored within the discussion in Section 7.5.2.

Given the multitemporal nature of data primitives, the periodicity function can identify several temporal properties of the identified cycle of gentrification. These can include the starting year of gentrification, the end year of gentrification, and the 'peak' year of gentrification (the year in which saw the greatest magnitude of gentrification associated changes). Further temporal properties can also be calculated, like the number of years it took for the cycle to reach its peak, the number of years it took for the cycle to end from its peak, and the overall duration of the gentrification cycle. These properties can be analysed to give an indication into the temporal manifestation thus periodicity of a process throughout space and time.

Once these unique cycles of gentrification have been identified, their temporal properties enable the manifestation (or periodicity, or process phase) to be identified and analysed throughout space and time. This exploration of neighbourhood processes and how they materialise throughout time is a new and novel research focus in both the neighbourhood change and gentrification communities. This function is used within the research found in Chapter 5 and Chapter 6.

3.4.4 Change Vector Analysis

The methodology of data primitives lends themselves to more accurate change detection than that of the current approaches to neighbourhood change explored in Section 2.4, particularly when used in conjunction with Change Vector Analysis. CVA also originates from the remote sensing community, and is used to determine land coverage changes from shifts in a pixel's position in multi-variate feature space of remote sensing image bands (Lindsay, 2012). Change Vector Analysis (CVA) is a robust and popular method of change detection, designed to explore changes via an interconnecting change vector. A change vector is the difference between two vectors in n-dimensional feature space, defined for two observations of the same geographical location (Lindsay, 2012; Tewkesbury et al., 2015). In remote sensing, these are corresponding pixels, but in neighbourhood change research, these would be the spatial units of observation.

The magnitude of change is the Euclidean distance (length) between vector endpoints, of the area's change in position within the n-dimensional feature space. Angle is the direction of the change event, thus referring to the multi-dimensional sector in which the change event occurs (Lindsay, 2012). Theoretically, the angle (direction) of change can discriminate between different types of change occurring, and is introduced as a component of the data primitive methodology within Chapter 4, or different drivers of change (Gray et al., 2023a). The magnitude can be useful for relative comparisons within and among those change types (Johnson and Kasischke, 1998).

Equation 3.1 shows how the measure of the Euclidean distance (magnitude) of change between vectors between two locations, x_1 and x_2 in a multivariate feature space, is calculated. Distance, **D**, is described as the square root of the row sums, of the differences between data primitives at Time 1 (x) and Time 2 (y):

$$D = \sqrt{x_1 - x_2}^2 \tag{3.1}$$

Equation 3.2 shows the angle of change. θ is calculated from the dot product of the vectors of x_1 and x_2 in the following way:

$$\theta = \cos^{-1}\left(\frac{x_1 x_2}{|x_1|.|x_2|}\right) \tag{3.2}$$

Where $|x_1|$ and $|x_2|$ are absolute values of the vectors.

In this way a CVA summarises change across the full dimensionality of the data and has been found to be robust to the nature and number dimensions in the feature space (Johnson and Kasischke, 1998). In neighbourhood analyses, CVA magnitude and direction can be extracted and explored alongside changes in neighbourhood primitives, as explored in Chapter 5. CVA can also be applied to the single time period that most strongly indicated the presence of gentrification, as explored in Chapter 6. Therefore, the versatility of the data primitive approach enables neighbourhood change analyses to be operationalised in a manner of ways, dependant upon the purpose of the analysis. This is beneficial over other temporally constrained and information reductive approaches, since CVA alongside data primitives, provide multidimensional features of the change opposed to one single measurement, or cluster label. This multidimensional change (with one change vector), describes the type of change, magnitude of change, direction of change in a multidimensional feature space, and once uncovered and investigated further, the speed of the change. This information alongside the data primitive changes and neighbourhood characteristics can afford many new, novel insights into the spatial and temporal dynamics of the process under investigation.

By using this approach, the authors are able to identify neighbourhoods experiencing gentrification based on multiple data primitives rather than relying on a single variable or clustering approach. This approach provides a more nuanced and comprehensive understanding of gentrification, as it captures the complex and interrelated changes that occur in neighbourhoods experiencing gentrification.

Overall, the use of the data primitive approach and CVA is a strength of the study, as it allows for a more comprehensive and accurate identification of gentrification and other forms of neighbourhood change.

3.4.5 Validation

Chapter 5 employs all of the previous methods to identify cycles of gentrification in South Yorkshire. Initially, seven of these identified cycles are validated via Google Earth and Google Street View in order to confirm whether the selected neighbourhoods had indeed experienced gentrification, and the confirmation of their temporal properties. This method of validation relies upon visual observations and local knowledge of the area, and is subsequently relatively subjective, even when consulting the data associated with the specific LSOA. Nevertheless, all 123 identified cycles of gentrification associated changes are validated, via images at the start, peak, and end of the process, and descriptions of the observed changes are generated.

Via the observed changes, local knowledge, data primitive changes, and neighbourhood characteristics, the likely gentrification type is determined. These range from replacement new-build gentrification to studentification, to rail-induced, to none. However, some of these gentrification types have small counts insufficient for training data for predictive models. A pragmatic decision was therefore made to assign each of the LSOAs to an agglomerated gentrification type of: 1) Residential Gentrification, 2) Rural Gentrification, and 3) Transport Gentrification. These represent the larger, more hierarchical gentrification types of the more specific types identified. Appendix B describes the observations made in Google Earth, and the assignments of each LSOA. It also highlights some of the ambiguity faced when determining the assignment, highlighting the subjective nature of the method of validation. Consequently, although great effort was made to assign the LSOA to the most appropriate gentrification type, if some others were to conduct the validation, they may argue for different assignments.

Appendix B is the most comprehensive output, providing descriptions for each 123 LSOAs. Appendix C, meanwhile, is less comprehensive as Google Street View is only used when observations are inconclusive in Google Earth validations (Appendix B), and is used as another layer of validation.

3.4.6 Machine Learning

As analyses develop throughout the papers, a typical machine learning approach is adopted in Chapter 6. Here, a number of ensemble models are trained using the data primitives, change vectors, and a range of descriptive variables, to fit models for predicting the spatial and temporal extent of gentrification in England. The extensively validated South Yorkshire data from Chapter 5 provides the training dataset upon which these models are trained. This enables a solid foundation before applying models to predict gentrification in England.

Ensemble methods aim to improve the predictive performance of statistical learning or models, and they do so by constructing linear combinations of some model fitting method, instead of a single classifier (Bühlmann, 2012). Thus, ensemble models are models that train multiple classifiers and combine their results to improve accuracy. They are robust to noise and variable interdependence (Dietterich, 2000; Zhou, 2012) Ensemble models can be bagged or boosted; bagging techniques work in parallel and create models independently, whilst boosting is sequential and creates classifiers considering the previous one (Friedman, 2002).

Several ensemble methods are compared in 6: two boosting algorithms, Gradient Boosting Machine GBM, and Extreme Gradient Boosting XGBoost, and one bootstrap aggregation algorithm (bagging: treebag). The caret package Kuhn:2008 in R was used to train, test, and validate the models. Models were fitted using a 70:30 train-test ratio, and parameters chosen via hyperparameter tuning, where appropriate. To measure the ability of the methods for predicting the spatial and temporal properties of gentrification, results were evaluated against their kappa values and subsequent confusion matrices and compared. The confusion matrix provides visualisation of model performance and generates greater insight into the types of errors within the classification via the sensitivity, and specificity values; true positives and true negatives respectively (Long et al., 2019). The model with the greatest overall performance for each response, was subsequently chosen and used to predict the respective presence, type, and temporal properties of gentrification in England.

The England predictions however remain unvalidated due to a number of reasons, which are explored in greater detail in Section 7.4.2.

Gradient Boosting Machine

Boosting algorithms assign weights to good classifiers and take a weighted average of their estimates, thus constructing an extremely accurate prediction from several roughly accurate predictions (Friedman, 2002). GBM are subsequently a family of powerful machine learning techniques that have shown success in a wide range of practical applications (Natekin and Knoll, 2013). GBM iteratively refines an initial model by examining the error within the previous model, improving upon weak learners until some accuracy (such as a loss (Kuhn and Johnson, 2013)) or iteration threshold is reached (Sagi and Rokach, 2018).

Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a reliable machine learning algorithm designed to scale up tree-boosting algorithms. It is a combination of gradient descent and boosting that can be used to build an optimal model for times series data (Chen and Guestrin, 2015). XGBoost is like GBM, but also includes regression penalties within the boosting equation, with the regularization controlling over-fitting and often generating better performing models (Chen and Guestrin, 2015). Consequently, XGBoost often outperforms other boosting algorithms including GBM in prediction. It also provides advantages including faster execution speeds, and enables parellelisation (Kadiyala and Kumar, 2018), with speeds up to 20x faster than GBM (Chen and Guestrin, 2015).

Bootstrap Aggregation

Bootstrap aggregating, commonly known as bagging, is an ensemble method designed for improving the stability and predictive performance of models (Bühlmann, 2012). Bagging is based on the concept of model averaging, it differs from boosting by training single models in parallel, rather than iteratively, and averages them to yield more accurate predictions (Lee et al., 2020). The conventional bagging algorithm consists of two steps 1) bootstrap sampling, which involves generating n different bootstrap training samples with replacement and training the algorithm on each bootstrapped algorithm separately; and 2) aggregation, where the predictions are aggregated at the end (Zhihao et al., 2019).

Bagging is employed as a technique because it improves the stability and accuracy of machine learning algorithms, while reducing variance and overfitting. Thus, bagging can improve the misclassification rate of the bagged classifier. It also holds an advantage over boosting algorithms in that bagging outperforms boosting in noisy datasets. However, bagging can also degrade bad classifiers, making model performance worse. Thus, bagging requires a good initial classifier to improve model performance (Zhihao et al., 2019).

3.5 Data and Methods Summary

This chapter provides a comprehensive overview of the methods and data used within this research to supplement the detail included directly in Chapters 4, 5, and 6. In Chapter 4, the literature review summarises the current literature regarding geodemographic classifications, neighbourhood processes, and the approaches typically used to measure neighbourhood change. This review found a gap in literature in which a new approach to neighbourhood change analyses was proposed, data primitives. Chapters 5 use data primitives to demonstrate their application and subsequent utility in generating insights regarding the spatiotemporal patterns of gentrification in South Yorkshire, whilst Chapter 6 use validated Chapter 5 results to predict the type and cycle of gentrification in England.

References

Aalbers, Manuel B (2019). "Introduction to the forum: From third to fifth-wave gentrification". In: *Tijdschrift voor economische en sociale geografie* 110.1, pp. 1–11. ISSN: 0040-747X.

Abbas, J, Carlin, H, Cunningham, A, Dedman, D, and McVey, D (2009). "Geodemographic segmentation". In: *York, UK: Association of Public Health Observatories [Technical Briefing 5]*.

Adnan, Muhammad, Longley, Paul A, Singleton, Alex D, and Brunsdon, Chris (2010). "Towards real-time geodemographics: Clustering algorithm performance for large multidimensional spatial databases". In: *Transactions in GIS* 14.3, pp. 283–297. ISSN: 1361-1682.

Airgood-Obrycki, Whitney (2019). "Suburban status and neighbourhood change". In: *Urban Studies* 56.14, pp. 2935–2952. DOI: 10.1177/0042098018811724. URL: https://journals.sagepub.com/doi/abs/10.1177/0042098018811724.

Anguelovski, Isabelle, Connolly, James J T, Masip, Laia, and Pearsall, Hamil (2018). "Assessing green gentrification in historically disenfranchised neighborhoods: a longitudinal and spatial analysis of Barcelona". In: *Urban Geography* 39.3, pp. 458–491. ISSN: 0272-3638.

Ashby, David I and Longley, Paul A (2005). "Geocomputation, geodemographics and resource allocation for local policing". In: *Transactions in GIS* 9.1, pp. 53–72. ISSN: 1361-1682.

Atkinson, Rowland (2000). "Measuring gentrification and displacement in Greater London". In: *Urban Studies* 37.1, pp. 149–165. ISSN: 0042-0980.

Baing, Andreas Schulze (2009). "Target-driven brownfield reuse - a benefit for deprived areas? A spatial analysis of brownfield reuse patterns in England's core city regions". In.

Bardaka, Eleni, Delgado, Michael S, and Florax, Raymond J G M (2018). "Causal identification of transit-induced gentrification and spatial spillover effects: The case of the Denver light rail". In: *Journal of Transport Geography* 71, pp. 15–31. ISSN: 0966-6923.

Barton, Michael (2016). "An exploration of the importance of the strategy used to identify gentrification". In: *Urban Studies* 53.1, pp. 92–111. ISSN: 0042-0980.

Beecham, Roger and Lovelace, Robin (2022). "A Framework for Inserting Visually Supported Inferences into Geographical Analysis Workflow: Application to Road Safety Research". In: *Geographical Analysis*. ISSN: 0016-7363.

Behrens, Kristian, Boualam, Brahim, Martin, Julien, and Mayneris, Florian (2022). "Gentrification and pioneer businesses". In: *Review of Economics and Statistics*, pp. 1–45. ISSN: 0034-6535.

Bhavsar, Nrupen A, Kumar, Manish, and Richman, Laura (2020). "Defining gentrification for epidemiologic research: A systematic review". In: *PloS one* 15.5, e0233361. ISSN: 1932-6203.

Bright, C J, Gildea, C, Lai, J, Elliss-Brookes, L, and Lyratzopoulos, G (2021). "Does geodemographic segmentation explain differences in route of cancer diagnosis above and beyond person-level sociodemographic variables?" In: *Journal of Public Health* 43.4, pp. 797–805. ISSN: 1741-3842.

Brodersen, K H, Ong, C S, Stephan, K E, and Buhmann, J M (2010). "The Balanced Accuracy and Its Posterior Distribution". In: *2010 20th International Conference on Pattern Recognition*, pp. 3121–3124. ISBN: 1051-4651. DOI: 10.1109/ICPR.2010.764.

Bühlmann, Peter (2012). *Bagging, boosting and ensemble methods*. Springer. ISBN: 3642215505.

Burns, Luke, See, Linda, Heppenstall, Alison, and Birkin, Mark (2018). “Developing an individual-level geodemographic classification”. In: *Applied Spatial Analysis and Policy* 11, pp. 417–437. ISSN: 1874-463X.

Butler, D C, Petterson, Stephen, Phillips, Robert L, and Bazemore, Andrew W (2013). “Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery”. In: *Health services research* 48 2 Pt 1, pp. 539–559.

Butler, Tim (2007). “For gentrification?” In: *Environment and Planning A* 39.1, pp. 162–181. ISSN: 0308-518X.

Butler, Tim and Lees, Loretta (2006). “Super-gentrification in Barnsbury, London: globalization and gentrifying global elites at the neighbourhood level”. In: *Transactions of the Institute of British Geographers* 31.4, pp. 467–487. ISSN: 0020-2754. DOI: https://doi.org/10.1111/j.1475-5661.2006.00220.x. URL: https://rgs-ibg.onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-5661.2006.00220.x.

Buzar, Stefan, Ogden, Philip E, and Hall, Ray (2005). “Households matter: the quiet demography of urban transformation”. In: *Progress in Human Geography* 29.4, pp. 413–436. ISSN: 0309-1325.

Cabrera-Barona, Pablo, Murphy, Thomas, Kienberger, Stefan, and Blaschke, Thomas (2015). “A multi-criteria spatial deprivation index to support health inequality analyses”. In: *International Journal of Health Geographics* 14.

Carstairs, Vera and Morris, Russell (1989). “Deprivation: explaining differences in mortality between Scotland and England and Wales”. In: *British Medical Journal* 299.6704, pp. 886–889. ISSN: 0959-8138.

Chapple, Karen, Loukaitou-Sideris, Anastasia, Gonzalez, Silvia R, Kadin, Dov, and Poirier, Joseph (2017). “Transit-oriented development & commercial gentrification: exploring the linkages”. In.

Chapple, Karen, Poorthuis, Ate, Zook, Matthew, and Phillips, Eva (2022). “Monitoring streets through tweets: Using user-generated geographic information to predict gentrification and displacement”. In: *Environment and Planning B: Urban Analytics and City Science* 49.2, pp. 704–721. ISSN: 2399-8083.

Chapple, Karen and Zuk, Miriam (2016). “Forewarned: The use of neighborhood early warning systems for gentrification and displacement”. In: *Cityscape* 18.3, pp. 109–130. ISSN: 1936-007X.

Chen, Meixu, Chi, Bin, Van Dijk, Justin, and Longley, P (2021). “The use of Linked Consumer Registers to understand social and residential mobility”. In: GIS Research UK.

Chen, Tianqi and Guestrin, Carlos (2015). “Xgboost: Reliable large-scale tree boosting system”. In: *Proceedings of the 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA*, pp. 13–17.

Choi, Myungshik, Van Zandt, Shannon, and Matarrita-Cascante, David (2018). “Can community land trusts slow gentrification?” In: *Journal of Urban Affairs* 40.3, pp. 394–411. ISSN: 0735-2166.

Christakos, George, Bogaert, Patrick, and Serre, Marc (2002). *Temporal GIS: advanced functions for field-based applications.* Springer Science & Business Media. ISBN: 3540414762.

Clark, W A William A V and Dieleman, Frans M (1996). *Households and housing: Choice and outcomes in the housing market.* Transaction Publishers. ISBN: 1412850509.

Clark, William A V and Coulter, Rory (2015). “Who wants to move? The role of neighbourhood change”. In: *Environment and Planning A* 47.12, pp. 2683–2709. ISSN: 0308-518X.

Clelland, David and Hill, Carol (2019). “Deprivation, policy and rurality: the limitations and applications of area-based deprivation indices in Scotland”. In: *Local Economy* 34.1, pp. 33–50. ISSN: 0269-0942.

Cocola-Gant, Agustín (2018). "Tourism gentrification". In: *Handbook of gentrification studies*. Edward Elgar Publishing.

— (2019). "Gentrification and displacement: urban inequality in cities of late capitalism". In: *Handbook of Urban Geography*. Edward Elgar Publishing, pp. 297–310.

Cole, Helen V S, Triguero-Mas, Margarita, Connolly, James J T, and Anguelovski, Isabelle (2019). "Determining the health benefits of green space: Does gentrification matter?" In: *Health & Place* 57, pp. 1–11. ISSN: 1353-8292.

Comber, A and Kuhn, W (2018). "Fuzzy difference and data primitives: a transparent approach for supporting different definitions of forest in the context of REDD+". In: *Geogr. Helv.* 73.2, pp. 151–163. ISSN: 2194-8798. DOI: 10.5194/gh-73-151-2018. URL: https://gh.copernicus.org/articles/73/151/2018/.

Comber, A J (2008). "The separation of land cover from land use using data primitives". In: *Journal of Land Use Science* 3.4, pp. 215–229. ISSN: 1747-423X. DOI: 10.1080/17474230802465173. URL: https://doi.org/10.1080/17474230802465173.

Comber, Alexis and Wulder, Michael (2019). *Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use.*

Compton, Mallory E, Luetjens, Joannah, and Hart, Paul't (2019). "Designing for policy success". In: *International Review of Public Policy* 1.1: 2, pp. 119–146. ISSN: 2706-6274.

Cooper, Adam Elliott, Hubbard, Phil, and Lees, Loretta (2020). "Sold out? The right-to-buy, gentrification and working-class displacements in London". In: *The Sociological Review* 68.6, pp. 1354–1369. ISSN: 0038-0261.

Corcoran, Jonathan, Higgs, Gary, and Anderson, Tessa (2013). "Examining the use of a geodemographic classification in an exploratory analysis of variations in fire incidence in South Wales, UK". In: *Fire Safety Journal* 62, pp. 37–48. ISSN: 0379-7112.

Coulton, Claudia, Korbin, Jill, Chan, Tsui, and Su, Marilyn (2001). "Mapping Residents' Perceptions of Neighborhood Boundaries: A Methodological Note". In: *American Journal of Community Psychology* 29, pp. 371–383. DOI: 10.1023/A:1010303419034.

Crowe, Adam, Shaphan Cox, and Davies, Amanda (2020). "Shifting Neighbourhood Dynamics and Everyday Experiences of Displacement in Kruezberg, Berlin". PhD thesis. Curtin University, p. 301.

Daras, Konstantinos, Green, Mark A, Davies, Alec, Barr, Benjamin, and Singleton, Alex (2019). "Open data on health-related neighbourhood features in Great Britain". In: *Scientific data* 6.1, p. 107. ISSN: 2052-4463.

Davidson, Mark and Lees, Loretta (2010). "New-build gentrification: its histories, trajectories, and critical geographies". In: *Population, space and place* 16.5, pp. 395–411. ISSN: 1544-8444.

Deas, Iain, Robson, Brian, Wong, Cecilia, and Bradford, Michael (2003). "Measuring neighbourhood deprivation: a critique of the Index of Multiple Deprivation". In: *Environment and Planning C: Government and Policy* 21.6, pp. 883–903. ISSN: 0263-774X.

Delmelle, Elizabeth C (2021). "Transit-induced gentrification and displacement: The state of the debate". In: *Advances in Transport Policy and Planning*. Vol. 8. Elsevier, pp. 173–190. ISBN: 2543-0009.

— (2022). "GIScience and neighborhood change: Toward an understanding of processes of change". In: *Transactions in GIS* 26.2, pp. 567–584. ISSN: 1361-1682.

Dialga, Issaka and Thi Hang Giang, Le (2017). "Highlighting methodological limitations in the steps of composite indicators construction". In: *Social indicators research* 131, pp. 441–465. ISSN: 0303-8300.

Dietterich, Thomas G (2000). "Ensemble methods in machine learning". In: *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1*. Springer, pp. 1–15. ISBN: 3540677046.

Durlauf, Steven N (2004). "Neighborhood effects". In: *Handbook of regional and urban economics* 4, pp. 2173–2242. ISSN: 1574-0080.

Everitt, J C and Gill, A M (1993). "The social geography of small towns". In: *The changing social geography of Canadian cities*, pp. 252–264.

Finch, Janet (2008). "Naming names: Kinship, individuality and personal names". In: *Sociology* 42.4, pp. 709–725. ISSN: 0038-0385.

Finio, Nicholas (2022). "Measurement and definition of gentrification in urban studies and planning". In: *Journal of Planning Literature* 37.2, pp. 249–264. ISSN: 0885-4122.

Finney, Nissa (2013). *Statistical boundaries and small area data: something worth saving?* https://citiesmcr.wordpress.com/. URL: https://citiesmcr.wordpress.com/2013/09/23/statistical-boundaries-and-small-area-data-something-worth-saving/.

Fisher, Peter and Tate, Nicholas J (2015). "Modelling class uncertainty in the geodemographic Output Area Classification". In: *Environment and Planning B: Planning and Design* 42.3, pp. 541–563. ISSN: 0265-8135.

Freeman, Lance (2005). "Displacement or succession? Residential mobility in gentrifying neighborhoods". In: *Urban Affairs Review* 40.4, pp. 463–491. ISSN: 1078-0874.

— (2009). "Neighbourhood diversity, metropolitan segregation and gentrification: What are the links in the US?" In: *Urban Studies* 46.10, pp. 2079–2101. ISSN: 0042-0980.

Friedman, Jerome H (2002). "Stochastic gradient boosting". In: *Computational statistics & data analysis* 38.4, pp. 367–378. ISSN: 0167-9473.

Galster, George C (2002). “Gentrification as diversification: Why Detroit needs it and how it can get it”. In: *JL Soc'y* 4, p. 29.

Gibbons, Joseph (2019). “Are gentrifying neighborhoods more stressful? A multilevel analysis of self-rated stress”. In: *SSM-population health* 7, p. 100358. ISSN: 2352-8273.

Gibbons, Joseph and Barton, Michael S (2016). “The association of minority self-rated health with black versus white gentrification”. In: *Journal of urban health* 93, pp. 909–922. ISSN: 1099-3460.

Glaeser, Edward L, Kim, Hyunjin, and Luca, Michael (2018). “Nowcasting gentrification: using yelp data to quantify neighborhood change”. In: *AEA Papers and Proceedings*. Vol. 108. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 77–82. ISBN: 2574-0768.

Glass, Ruth (1964). *London: aspects of change*. 3. MacGibbon & Kee.

Gray, Jennie, Buckner, Lisa, and Comber, Alexis (2021). “Extending geodemographics using data primitives: A review and a methodological proposal”. In: *ISPRS International Journal of Geo-Information* 10.6. ISSN: 22209964. DOI: 10.3390/ijgi10060386.

— (2023a). “Identifying Neighbourhood Change Using a Data Primitive Approach: the Example of Gentrification”. In: *Applied Spatial Analysis and Policy*. ISSN: 1874-4621. DOI: 10.1007/s12061-023-09509-y. URL: https://doi.org/10.1007/s12061-023-09509-y.

— (2023b). “Predicting Gentrification in England: A Data Primitive Approach”. In: *Urban Science* 7.2, p. 64. ISSN: 2413-8851. URL: https://www.mdpi.com/2413-8851/7/2/64.

Greig, Alastair, El-Haram, Mohamed, and Horner, Malcolm (2010). “Using deprivation indices in regeneration: Does the response match the diagnosis?” In: *Cities* 27.6, pp. 476–482. ISSN: 0264-2751.

Grekousis, George (2019). “Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis”. In: *Computers, Environment and Urban Systems* 74, pp. 244–256. ISSN: 0198-9715.

Grekousis, George and Thomas, Hatzichristos (2012). “Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and Gustafson–Kessel methods”. In: *Applied Geography* 34, pp. 125–136. ISSN: 0143-6228.

Grekousis, George, Wang, Ruoyu, and Liu, Ye (2021). “Mapping the geodemographics of racial, economic, health, and COVID-19 deaths inequalities in the conterminous US”. In: *Applied Geography* 135, p. 102558. ISSN: 0143-6228.

Grigsby, William (1986). “The Dynamics of Neighborhood Change and Decline”. In.

Grubesic, Tony H, Nelson, Jake R, Wallace, Danielle, Eason, John, Towers, Sherry, and Walker, Jason (2021). “Geodemographic insights on the COVID-19 pandemic in the State of Wisconsin and the role of risky facilities”. In: *GeoJournal*, pp. 1–23. ISSN: 1572-9893.

Guan, Haotian and Cao, Huhua (2020). “Gentrification in the Global South: new insights from Chinese Studies”. In: *Boletín de la Asociación de Geógrafos Españoles* 87, p. 2. ISSN: 0212-9426.

Hamnett, Chris (2003). “Gentrification and the middle-class remaking of inner London, 1961-2001”. In: *Urban Studies* 40.12, pp. 2401–2426. ISSN: 0042-0980.

Harris, Richard, Sleight, Peter, and Webber, Richard (2005). *Geodemographics, GIS and neighbourhood targeting.* Vol. 7. John Wiley and Sons. ISBN: 0470864133.

Hartigan, John A and Wong, Manchek A (1979). “Algorithm AS 136: A k-means clustering algorithm”. In: *Journal of the royal statistical society. series c (applied statistics)* 28.1, pp. 100–108. ISSN: 0035-9254.

Hawken, Angela and Munck, Gerardo L (2013). "Cross-national indices with gender-differentiated data: what do they measure? How valid are they?" In: *Social indicators research* 111.3, pp. 801–838. ISSN: 1573-0921.

Higgs, Gary (1999). "Investigating trends in rural health outcomes: a research agenda". In: *Geoforum* 30.3, pp. 203–221. ISSN: 0016-7185.

Hincks, Stephen (2015). "Neighbourhood change and deprivation in the Greater Manchester city-region". In: *Environment and Planning A* 47.2, pp. 430–449. ISSN: 0308-518X.

— (2017). "Deprived neighbourhoods in transition: Divergent pathways of change in the Greater Manchester city-region". In: *Urban Studies* 54, pp. 1038–1061.

Hjørland, Birger and Pedersen, Karsten Nissen (2005). "A substantive theory of classification for information retrieval". In: *Journal of documentation*. ISSN: 0022-0418.

Hochstenbach, Cody and Musterd, Sako (2021). "A regional geography of gentrification, displacement, and the suburbanisation of poverty: Towards an extended research agenda". In: *Area* 53.3, pp. 481–491. ISSN: 0004-0894.

Hochstenbach, Cody, Musterd, Sako, and Teernstra, Annalies (2015). "Gentrification in Amsterdam: Assessing the importance of context". In: *Population, space and place* 21.8, pp. 754–770. ISSN: 1544-8444.

Holm, Andrej and Schulz, Guido (2018). "GentriMap: A model for measuring gentrification and displacement". In: *Gentrification and Resistance: Researching Displacement Processes and Adaption Strategies*, pp. 251–277. ISSN: 3658203870.

Hubbard, Phil (2008). "Regulating the social impacts of studentification: a Loughborough case study". In: *Environment and Planning A* 40.2, pp. 323–341. ISSN: 0308-518X.

Hubbard, Phil (2018). “Retail gentrification”. In: *Handbook of gentrification studies*. Edward Elgar Publishing.

Hughes, John C, James, Gareth, Evans, Andrew, and Prestwood, Debra (2009). “Implementation of Standard Industrial Classification 2007: December 2009 update”. In: *Economic & Labour Market Review* 3.12, pp. 51–55. ISSN: 1751-8326. DOI: 10.1057/elmr.2009.205. URL: https://doi.org/10.1057/elmr.2009.205.

Huse, Tone (2018). “Gentrification and ethnicity”. In: *Handbook of gentrification studies*. Edward Elgar Publishing.

Hwang, Jackelyn and Lin, Jeffrey (2016). “What have we learned about the causes of recent gentrification?” In: *Cityscape* 18.3, pp. 9–26. ISSN: 1936-007X.

Ilic, Lazar, Sawada, Michael, and Zarzelli, Amaury (2019). “Deep mapping gentrification in a large Canadian city using deep learning and Google Street View”. In: *PloS one* 14.3, e0212814. ISSN: 1932-6203.

Izenberg, Jacob M, Mujahid, Mahasin S, and Yen, Irene H (2018). “Health in changing neighborhoods: A study of the relationship between gentrification and self-rated health in the state of California”. In: *Health & Place* 52, pp. 188–195. ISSN: 1353-8292.

Jain, Shomik, Proserpio, Davide, Quattrone, Giovanni, and Quercia, Daniele (2021). “Nowcasting gentrification using Airbnb data”. In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1, pp. 1–21. ISSN: 2573-0142.

James, William H M, Lomax, Nik, Birkin, Mark, and Collins, Lisa M (2021). “Geodemographic Patterns of Meat Expenditure in Great Britain”. In: *Applied Spatial Analysis and Policy* 14.3, pp. 563–590. ISSN: 1874-4621.

Johnson, Glen D, Checker, Melissa, Larson, Scott, and Kodali, Hanish (2022). "A small area index of gentrification, applied to New York City". In: *International Journal of Geographical Information Science* 36.1, pp. 137–157. ISSN: 1365-8816.

Johnson, R D and Kasischke, E S (1998). "Change vector analysis: A technique for the multispectral monitoring of land cover and condition". In: *International journal of remote sensing* 19.3, pp. 411–426. ISSN: 0143-1161.

Kadiyala, Akhil and Kumar, Ashok (2018). "Applications of python to evaluate the performance of decision tree-based boosting algorithms". In: *Environmental Progress & Sustainable Energy* 37.2, pp. 618–623. ISSN: 1944-7442.

Kandt, Jens and Longley, Paul A (2018). "Ethnicity estimation using family naming practices". In: *PloS one* 13.8, e0201774. ISSN: 1932-6203.

Kearns, Ade and Parkinson, Michael (2001). "The significance of neighbourhood". In: *Urban Studies* 38.12, pp. 2103–2110. ISSN: 0042-0980.

Kintrea, Keith (2007). "Policies and programmes for disadvantaged neighbourhoods: recent English experience". In: *Housing Studies* 22.2, pp. 261–282. ISSN: 0267-3037.

Kosta, Ervin B (2019). "Commercial gentrification indexes: Using business directories to map urban change at the street level". In: *City & Community* 18.4, pp. 1101–1122. ISSN: 1535-6841.

Krase, Jerome and DeSena, Judith N (2020). *Gentrification Around the World, Volume I: Gentrifiers and the Displaced*. Springer Nature. ISBN: 3030413373.

Kuhn, Max and Johnson, Kjell (2013). *Applied predictive modeling*. Vol. 26. Springer.

Lan, Tian, O'Brien, Oliver, Cheshire, James, Singleton, Alex, and Longley, Paul (2022). "From Data to Narratives: Scrutinising the Spatial Dimensions of Social and Cultural Phenomena

Through Lenses of Interactive Web Mapping". In: *Journal of Geovisualization and Spatial Analysis* 6.2, p. 22. ISSN: 2509-8810.

Lansley, Guy, Li, Wen, and Longley, Paul A (2019). "Creating a linked consumer register for granular demographic analysis". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182.4, pp. 1587–1605. ISSN: 1467-985X.

Largent, Avery and Quimby, Michelann (2020). "Gentrification, displacement, and perception of community among longtime residents of Austin, Texas: Implications from six case studies". In: *Journal of Integrated Social Sciences* 10.1, pp. 52–85.

Lawless, Paul (2012). "Can area-based regeneration programmes ever work? Evidence from England's New Deal for Communities Programme". In: *Policy Studies* 33.4, pp. 313–328. ISSN: 0144-2872.

Lee, Jae Kwang (2013). "Mega-retail-led regeneration and housing price". In: *disP-The Planning Review* 49.2, pp. 75–85. ISSN: 0251-3625.

Lee, Tae-Hwy, Ullah, Aman, and Wang, Ran (2020). "Bootstrap aggregating and random forest". In: *Macroeconomic forecasting in the era of big data: Theory and practice*, pp. 389–429. ISSN: 303031149X.

Lees, Loretta (2018). "Introduction: Towards a C21st global gentrification studies". In: *Handbook of gentrification studies.* Edward Elgar Publishing, pp. 1–10.

Lees, Loretta, Shin, Hyun Bang, and López-Morales, Ernesto (2016). *Planetary gentrification.* John Wiley & Sons. ISBN: 1509505881.

Lees, Loretta, Slater, Tom, and Wyly, Elvin K (2010). *The gentrification reader.* Vol. 1. Routledge London.

Lees, Loretta, Slater, Tom, and Wyly, Evelin (2008). *Gentrification.* Routledge.

Leventhal, Barry (2016). *Geodemographics for marketers: Using location analysis for research and marketing*. Kogan Page Publishers. ISBN: 0749473835.

Li, Han (2012). "Modeling gentrification on census tract level in Chicago from 1990 to 2000". PhD thesis.

Lindsay, John (2012). *Change Vector Analysis*. URL: https://jblindsay.github.io/ghrg/Whitebox/Help/ChangeVectorAnalysis.html.

Liu, Cheng, Deng, Yu, Song, Weixuan, Wu, Qiyan, and Gong, Jian (2019). "A comparison of the approaches for gentrification identification". In: *Cities* 95, p. 102482. ISSN: 0264-2751. DOI: https://doi.org/10.1016/j.cities.2019.102482. URL: https://www.sciencedirect.com/science/article/pii/S0264275118315440.

Liu, Cheng and O'Sullivan, David (2016). "An abstract model of gentrification as a spatially contagious succession process". In: *Computers, Environment and Urban Systems* 59, pp. 1–10. ISSN: 0198-9715.

Liu, Yunzhe and Cheng, Tao (2020). "Understanding public transit patterns with open geodemographics to facilitate public transport planning". In: *Transportmetrica A: Transport Science* 16.1, pp. 76–103. ISSN: 2324-9935.

Long, Jason S, Mistry, Bhakti, Haslam, Stuart M, and Barclay, Wendy S (2019). "Host and viral determinants of influenza A virus species specificity". In: *Nature Reviews Microbiology* 17.2, pp. 67–81. ISSN: 1740-1526.

Longley, Paul (2005). "Geographical information systems: A renaissance of geodemographics for public service delivery". In: *Progress in Human Geography* 29.1, pp. 57–63. ISSN: 0309-1325.

Longley, Paul A (2012). "Geodemographics and the practices of geographic information science". In: *International Journal of Geographical Information Science* 26.12, pp. 2227–2237. ISSN: 1365-8816.

Lupton, Ruth and Power, Anne (2004). "What we know about neighbourhood change: a literature review". In: *Centre for the Analysis of Social Exclusion* 27. ISSN: CASEreport 27. URL: http://eprints.lse.ac.uk/27357/1/What_we_know_about_neighbourhood_change_%28LSERO_version%29.pdf.

McLachlan, Gordon and Norman, Paul (2021). "Analysing socio-economic change using a time comparable geodemographic classification: England and Wales, 1991–2011". In: *Applied Spatial Analysis and Policy* 14.1, pp. 89–111. ISSN: 1874-4621.

Mendes, Luís (2013). "Marginal Gentrification as Emancipatory Practice: An Alternative to the Hegemonic Discourse of the Creative City?" In: *RCCS Annual Review. A selection from the Portuguese journal Revista Crítica de Ciências Sociais* 5. ISSN: 1647-3175.

Modai-Snir, Tal and Ham, Maarten van (2019). "Structural and exchange components in processes of neighbourhood change: A social mobility approach". In: *Applied Spatial Analysis and Policy* 12, pp. 423–443. ISSN: 1874-463X.

Moon, Graham, Twigg, Liz, Jones, Kelvyn, Aitken, Grant, and Taylor, Joanna (2019). "The utility of geodemographic indicators in small area estimates of limiting long-term illness". In: *Social Science & Medicine* 227, pp. 47–55. ISSN: 0277-9536.

Natekin, Alexey and Knoll, Alois (2013). "Gradient boosting machines, a tutorial". In: *Frontiers in neurorobotics* 7, p. 21. ISSN: 1662-5218.

Nelson, Peter B, Oberg, Alexander, and Nelson, Lise (2010). "Rural gentrification and linked migration in the United States". In: *Journal of Rural Studies* 26.4, pp. 343–352. ISSN: 0743-0167.

Nnoaham, Kelechi E, Frater, Alison, Roderick, Paul, Moon, Graham, and Halloran, Stephen (2010). "Do geodemographic typologies explain variations in uptake in colorectal cancer screening? An assessment using routine screening data in the south of England". In: *Journal of Public Health* 32.4, pp. 572–581. ISSN: 1741-3850.

Noble, Michael, Wright, Gemma, Smith, George, and Dibben, Chris (2006). "Measuring multiple deprivation at the small-area level". In: *Environment and Planning A* 38.1, pp. 169–185. ISSN: 0308-518X.

Norman, Paul (2010). "Identifying change over time in small area socio-economic deprivation". In: *Applied Spatial Analysis and Policy* 3.2, pp. 107–138. ISSN: 1874-4621.

Office for National Statistics (n.d.). *2011 Census: Population and Household Estimates for Small Areas in England and Wales, March 2011*. Ed. by National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23#:~:text=The%20average%20population%20of%20lower,983)%20was%20in%20Forest%20Heath..

Owens, Ann (2012). "Neighborhoods on the rise: A typology of neighborhoods experiencing socioeconomic ascent". In: *City & Community* 11.4, pp. 345–369. ISSN: 1535-6841.

Palafox, Leon and Ortiz-Monasterio, Pedro (2020). "Predicting gentrification in Mexico city using neural networks". In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–5. ISBN: 1728169267.

Pamuk, Ayse (2004). *Immigrant Clusters and Homeownership in Global Metropolises: Suburbanization Trends in San Francisco, Los Angeles, and New York*. Tech. rep.

Park, Robert E (1936). "Succession, an ecological concept". In: *American Sociological Review* 1.2, pp. 171–179. ISSN: 0003-1224.

Payne, Rupert A and Abel, Gary A (2012). "UK indices of multiple deprivation-a way to make comparisons across constituent countries easier". In: *Health Stat Q* 53.22, pp. 2015–2016.

Pegler, Claudia, Li, Hankan, and Pojani, Dorina (2020). "Gentrification in Australia's largest cities: A bird's-eye view". In: *Australian planner* 56.3, pp. 191–205. ISSN: 0729-3682.

Phillips, Martin, Smith, Darren, Brooking, Hannah, and Duer, Mara (2021). “Re-placing displacement in gentrification studies: Temporality and multi-dimensionality in rural gentrification displacement”. In: *Geoforum* 118, pp. 66–82. ISSN: 0016-7185.

Pinkster, Fenne M (2016). “Narratives of neighbourhood change and loss of belonging in an urban garden village”. In: *Social & Cultural Geography* 17.7, pp. 871–891. ISSN: 1464-9365.

Poorthuis, Ate, Shelton, Taylor, and Zook, Matthew (2022). “Changing neighborhoods, shifting connections: mapping relational geographies of gentrification using social media data”. In: *Urban Geography* 43.7, pp. 960–983. ISSN: 0272-3638.

Preis, Benjamin, Janakiraman, Aarthi, Bob, Alex, and Steil, Justin (2021). “Mapping gentrification and displacement pressure: An exploration of four distinct methodologies”. In: *Urban Studies* 58.2, pp. 405–424. ISSN: 0042-0980.

Prouse, Victoria, Grant, J, Ramos, Howard, and Radice, Martha (2015). *Assessing neighbourhood change: Gentrification and suburban decline in a mid-sized city*. School of Planning, Dalhousie University Halifax, NS, Canada.

Reades, Jonathan, De Souza, Jordan, and Hubbard, Phil (2019). “Understanding urban gentrification through machine learning”. In: *Urban Studies* 56.5, pp. 922–942. ISSN: 0042-0980.

Rees, Philip and Lomax, Nik (2019). “Ravenstein revisited: The analysis of migration, then and now”. In: *Comparative Population Studies* 44. ISSN: 1869-8999.

Reibel, Michael (2011). “Classification approaches in neighborhood research: Introduction and review”. In: *Urban Geography* 32.3, pp. 305–316. ISSN: 0272-3638.

Reibel, Michael and Regelson, Moira (2007). “Quantifying neighborhood racial and ethnic transition clusters in multiethnic cities”. In: *Urban Geography* 28.4, pp. 361–376. ISSN: 0272-3638.

Ribeiro, Marco Tulio, Singh, Sameer, and Guestrin, Carlos (2016). "" Why should i trust you?" Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144.

Richardson, Jason, Mitchell, Bruce, and Franco, Juan (2019). "Shifting neighborhoods: Gentrification and cultural displacement in American cities". In.

Rigolon, Alessandro and Németh, Jeremy (2020). "Green gentrification or 'just green enough': Do park location, size and function affect whether a place gentrifies or not?" In: *Urban Studies* 57.2, pp. 402–420. ISSN: 0042-0980.

Rincón, Francisco Andrés (2012). "An index for climate change: a multivariate time series approach". In: *Environmetrics* 23.7, pp. 617–622. ISSN: 1180-4009.

Robertson, Dustin, Oliver, Christopher, and Nost, Eric (2020). "Short-term rentals as digitally-mediated tourism gentrification: Impacts on housing in New Orleans". In: *Tourism Geographies*, pp. 1–24. ISSN: 1461-6688.

Royall, Emily and Wortmann, Thomas (2015). "Finding the state space of urban regeneration: modeling gentrification as a probabilistic process using k-means clustering and Markov models". In: *Proceedings of the 2015 14th International Conference on Computers in Urban Planning and Urban Management (CUPUM), Cambridge, MA, USA*, pp. 7–10.

Saar, Maarja and Palang, Hannes (2009). "The dimensions of place meanings". In: *Living reviews in landscape research* 3.3, pp. 5–24.

Sage, Joanna, Smith, Darren, and Hubbard, Phil (2012). "The rapidity of studentification and population change: There goes the (student) hood". In: *Population, space and place* 18.5, pp. 597–613. ISSN: 1544-8444.

Sagi, Omer and Rokach, Lior (2018). "Ensemble learning: A survey". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.4, e1249. ISSN: 1942-4787.

Saxena, Amit, Prasad, Mukesh, Gupta, Akshansh, Bharill, Neha, Patel, Om Prakash, Tiwari, Aruna, Er, Meng Joo, Ding, Weiping, and Lin, Chin-Teng (2017). "A review of clustering techniques and developments". In: *Neurocomputing* 267, pp. 664–681. ISSN: 0925-2312.

Shevky, Eshref and Williams, Marilyn (1949). *The social areas of Los Angeles: Analysis and typology*. Greenwood. ISBN: 0837156378.

Shmaryahu-Yeshurun, Yael and Ben-Porat, Guy (2021). "For the benefit of all? State-led gentrification in a contested city". In: *Urban Studies* 58.13, pp. 2605–2622. DOI: 10.1177/0042098020953077. URL: https://journals.sagepub.com/doi/abs/10.1177/0042098020953077.

Singleton, Alex David and Longley, Paul (2015). "The internal structure of Greater London: a comparison of national and regional geodemographic models". In: *Geo: Geography and Environment* 2.1, pp. 69–87. ISSN: 2054-4049.

Singleton, Alexander D and Longley, Paul A (2009). "Geodemographics, visualisation, and social networks in applied geography". In: *Applied Geography* 29.3, pp. 289–298. ISSN: 0143-6228.

Singleton, Alexander D and Spielman, Seth E (2014). "The past, present, and future of geodemographic research in the United States and United Kingdom". In: *The Professional Geographer* 66.4, pp. 558–567. ISSN: 0033-0124.

Singleton, Alexander D, Wilson, A G, and O'Brien, Oliver (2012). "Geodemographics and spatial interaction: an integrated model for higher education". In: *Journal of Geographical Systems* 14.2, pp. 223–241. ISSN: 1435-5949.

Sleight, Peter (2004). *Targeting customers: How to use geodemographic and lifestyle data in your business*. World Advertising Research Center Henley-on-Thames. ISBN: 1841161543.

Smith, Darren P (2019). "Studentification". In: *The Wiley Blackwell Encyclopedia of Urban and Regional Studies*, pp. 1–3.

Snyder, Hannah (2019). "Literature review as a research methodology: An overview and guidelines". In: *Journal of business research* 104, pp. 333–339. ISSN: 0148-2963.

Statistics, Office for National (2023). *SOC 2020: The current Standard Occupation Classification for the UK, published in three volumes.* URL: https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020.

Steinmetz-Wood, Madeleine, Wasfi, Rania, Parker, George, Bornstein, Lisa, Caron, Jean, and Kestens, Yan (2017). "Is gentrification all bad? Positive association between gentrification and individual's perceived neighborhood collective efficacy in Montreal, Canada". In: *International Journal of Health Geographics* 16, pp. 1–8.

Stockdale, Aileen (2010). "The diverse geographies of rural gentrification in Scotland". In: *Journal of Rural Studies* 26.1, pp. 31–40. ISSN: 0743-0167.

Sumka, Howard J (1979). "Neighborhood revitalization and displacement a review of the evidence". In: *Journal of the American Planning Association* 45.4, pp. 480–487. ISSN: 0194-4363.

Taylor, Mark, Higgins, Emma, Lisboa, Paulo, Jarman, Ian, and Hussain, Abir (2016). "Community fire prevention via population segmentation modelling". In: *Community Development Journal* 51.2, pp. 229–247. ISSN: 1468-2656.

Temkin, Kenneth and Rohe, William (1996). "Neighborhood change and urban policy". In: *Journal of Planning Education and Research* 15.3, pp. 159–170. ISSN: 0739-456X.

Tewkesbury, Andrew P, Comber, Alexis J, Tate, Nicholas J, Lamb, Alistair, and Fisher, Peter F (2015). "A critical synthesis of remotely sensed optical image change detection techniques". In: *Remote Sensing of Environment* 160, pp. 1–14. ISSN: 0034-4257.

Thackway, William, Ng, Matthew Kok Ming, Lee, Chyi Lin, and Pettit, Christopher (2021). "Building a predictive machine learning model of gentrification in Sydney". In.

Timberlake, Jeffrey M and Johns-Wolfe, Elaina (2017). “Neighborhood ethnoracial composition and gentrification in Chicago and New York, 1980 to 2010”. In: *Urban Affairs Review* 53.2, pp. 236–272. ISSN: 1078-0874.

Tobler, Waldo R (1970). “A computer movie simulating urban growth in the Detroit region”. In: *Economic Geography* 46.sup1, pp. 234–240. ISSN: 0013-0095.

Townsend, Peter (1987). “Deprivation”. In: *Journal of social policy* 16.2, pp. 125–146. ISSN: 1469-7823.

Uysal, Arzu Başaran and Sakarya, Ipek (2018). “Rural gentrification in the North Aegean countryside (Turkey)”. In: *ICONARP International Journal of Architecture and Planning* 6.1, pp. 99–125. ISSN: 2147-9380.

Valle, Melissa M (2021). “Globalizing the sociology of gentrification”. In: *City & Community* 20.1, pp. 59–70. ISSN: 1535-6841.

Van Ham, Maarten and Manley, David (2009). “The effect of neighbourhood housing tenure mix on labour market outcomes: a longitudinal investigation of neighbourhood effects”. In: *Journal of Economic Geography* 10.2, pp. 257–282. ISSN: 1468-2702.

— (2012). “Neighbourhood effects research at a crossroads. Ten challenges for future research introduction”. In: *Environment and Planning A* 44.12, pp. 2787–2793. ISSN: 0308-518X.

Van Ham, Maarten, Manley, David, Bailey, Nick, Simpson, Ludi, and Maclennan, Duncan (2011). “Neighbourhood effects research: New perspectives”. In: *Neighbourhood effects research: New perspectives*. Springer, pp. 1–21.

Van Ham, Maarten, Uesugi, Masaya, Tammaru, Tiit, Manley, David, and Janssen, Heleen (2020). “Changing occupational structures and residential segregation in New York, London and Tokyo”. In: *Nature human behaviour* 4.11, pp. 1124–1134. ISSN: 2397-3374.

Vij, Navta (2016). "Introducing the consumer data research centre (CDRC)". In: *Journal of Direct, Data and Digital Marketing Practice* 17, pp. 232–235. ISSN: 1746-0166.

Voorhees, Nathalie P (2014). *The Socioeconomic Change of Chicago's Community Areas (1970-2010).*

Wadsworth, Richard A, Comber, Alexis J, and Fisher, Peter F (2008). "Probabilistic Latent Semantic Analysis as a potential method for integrating spatial data concepts". In: ISSN: 3901716416.

Wang, Jing and Biljecki, Filip (2022). "Unsupervised machine learning in urban studies: A systematic review of applications". In: *Cities* 129, p. 103925. ISSN: 0264-2751.

Webber, R J (1977). "An introduction to the national classification of wards and parishes". In: *Planning Research Applications Group Technical Paper* 23.

Webber, Richard and Burrows, Roger (2018). *The predictive postcode: the geodemographic classification of British society.* Sage. ISBN: 1526448866.

Webster, Jane and Watson, Richard T (2002). "Analyzing the past to prepare for the future: Writing a literature review". In: *MIS quarterly*, pp. xiii–xxiii. ISSN: 0276-7783.

Weiser, Paul and Frank, Andrew U (2010). "Dynamic GIS–The final frontier". In: *Extended Abstract, GI-Forum, Salzburg.*

Wu, Longfeng (2020). "Socio-Spatial (In) Equality of Access to Urban Green Space: A Case in Beijing". PhD thesis.

Wu, Longfeng and Rowe, Peter G (2022). "Green space progress or paradox: identifying green space associated gentrification in Beijing". In: *Landscape and Urban Planning* 219, p. 104321. ISSN: 0169-2046. DOI: https://doi.org/10.1016/j.landurbplan.2021.104321. URL: https://www.sciencedirect.com/science/article/pii/S016920462100284X.

Xiang, Lili, Stillwell, John, Burns, Luke, Heppenstall, Alison, and Norman, Paul (2018). "A geodemographic classification of sub-districts to identify education inequality in Central Beijing". In: *Computers, Environment and Urban Systems* 70, pp. 59–70. ISSN: 0198-9715.

Xu, Hanqiu, Wang, Yifan, Guan, Huade, Shi, Tingting, and Hu, Xisheng (2019). "Detecting ecological changes with a remote sensing based ecological index (RSEI) produced time series and change vector analysis". In: *Remote Sensing* 11.20, p. 2345. ISSN: 2072-4292.

Yee, Joshua and Dennett, Adam (2022). "Stratifying and predicting patterns of neighbourhood change and gentrification: An urban analytics approach". In: *Transactions of the Institute of British Geographers* 47.3, pp. 770–790. ISSN: 0020-2754. DOI: https://doi.org/10.1111/tran.12522. URL: https://rgs-ibg.onlinelibrary.wiley.com/doi/abs/10.1111/tran.12522.

Yonto, Daniel and Thill, Jean-Claude (2020). "Gentrification in the US New South: Evidence from two types of African American communities in Charlotte". In: *Cities* 97, p. 102475. ISSN: 0264-2751.

Zapatka, Kasey and Beck, Brenden (2021). "Does demand lead supply? Gentrifiers and developers in the sequence of gentrification, New York City 2009–2016". In: *Urban Studies* 58.11, pp. 2348–2368. ISSN: 0042-0980.

Zhang, Yang, Aslam, Nilufer Sari, Lai, Juntao, and Cheng, Tao (2020). "You are how you travel: A multi-task learning framework for Geodemographic inference using transit smart card data". In: *Computers, Environment and Urban Systems* 83, p. 101517. ISSN: 0198-9715.

Zhihao, Peng, Fenglong, Yan, and Xucheng, Li (2019). "Comparison of the different sampling techniques for imbalanced classification problems in machine learning". In: *2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*. IEEE, pp. 431–434. ISBN: 1728121655.

Zhou, Peng, Ang, Beng Wah, and Poh, Kim Leng (2007). "A mathematical programming approach to constructing composite indicators". In: *Ecological economics* 62.2, pp. 291–297. ISSN: 0921-8009.

Zhou, Zhi-Hua (2012). *Ensemble methods: foundations and algorithms*. CRC press. ISBN: 1439830037.

Zhu, Zhe (2017). "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 130, pp. 370–384. ISSN: 0924-2716.

Zwiers, Merle, Kleinhans, Reinout, and Van Ham, Maarten (2017). "The Path-Dependency of Low-Income Neighbourhood Trajectories: An Approach for Analysing Neighbourhood Change". In: *Applied Spatial Analysis and Policy* 10.3, pp. 363–380. ISSN: 1874-4621. DOI: 10.1007/s12061-016-9189-z. URL: https://doi.org/10.1007/s12061-016-9189-z.

Part II

Included publications

Chapter 4: Extending Geodemographics Using Data Primitives: A Review and a Methodological Proposal

Overview

This chapter provides a review of geodemographic classifications, from their establishment to their current state. It highlights their major developmental steps throughout time, and discusses their strengths and limitations. It then goes on to propose a new methodological framework for analysing neighbourhood change, which uses *data primitives* to capture the more detailed and dynamic information about small-areas, which enable more nuanced analyses of sociodemographic trends. The proposed approach involves the creation of data primitives from a range of sources at a suitable spatiotemporal resolution, with the expectation that they represent the fundamental characteristics of the neighbourhood process of interest. These data primitives are associated with a directional change (increase or decrease), which when analysed temporally, can identify areas that fulfill those changes indicating where the process is active, and areas that do not fulfill those changes, indicating that the neighbourhood process was not present. A city-based illustration provides an example of how this approach can be used in practice. The article concludes with a discussion of the potential benefits and challenges of the proposed approach, as well as suggestions for future research. Overall, it argues that using data primitives can lead to more accurate and insightful neighbourhood analyses, and provides a roadmap for future research in this area.

International Journal of
Geo-Information

MDPI

Review

Extending Geodemographics Using Data Primitives: A Review and a Methodological Proposal

Jennie Gray [1], Lisa Buckner [2] and Alexis Comber [3,*]

1 Leeds Institute for Data Analytics, University of Leeds, Leeds LS2 9JT, UK; gyjhg@leeds.ac.uk
2 School of Sociology and Social Policy, University of Leeds, Leeds LS2 9JT, UK; l.j.buckner@leeds.ac.uk
3 School of Geography, University of Leeds, Leeds LS2 9JT, UK
* Correspondence: a.comber@leeds.ac.uk

Abstract: This paper reviews geodemographic classifications and developments in contemporary classifications. It develops a critique of current approaches and identifiea a number of key limitations. These include the problems associated with the geodemographic cluster label (few cluster members are typical or have the same properties as the cluster centre) and the failure of the static label to describe anything about the underlying neighbourhood processes and dynamics. To address these limitations, this paper proposed a data primitives approach. Data primitives are the fundamental dimensions or measurements that capture the processes of interest. They can be used to describe the current state of an area in a multivariate feature space, and states can be compared over multiple time periods for which data are available, through for example a change vector approach. In this way, emergent social processes, which may be too weak to result in a change in a cluster label, but are nonetheless important signals, can be captured. As states are updated (for example, as new data become available), inferences about different social processes can be made, as well as classification updates if required. State changes can also be used to determine neighbourhood trajectories and to predict or infer future states. A list of data primitives was suggested from a review of the mechanisms driving a number of neighbourhood-level social processes, with the aim of improving the wider understanding of the interaction of complex neighbourhood processes and their effects. A small case study was provided to illustrate the approach. In this way, the methods outlined in this paper suggest a more nuanced approach to geodemographic research, away from a focus on classifications and static data, towards approaches that capture the social dynamics experienced by neighbourhoods.

Keywords: urban dynamics; neighbourhood processes; state and change

Citation: Gray, J.; Buckner, L.; Comber, A. Extending Geodemographics Using Data Primitives: A Review and a Methodological Proposal. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 386. https://doi.org/10.3390/ijgi10060386

Academic Editors: Giuseppe Borruso, Ginevra Balletto, Beniamino Murgante, Carmelo Maria Torre, Hartmut Asche and Wolfgang Kainz

Received: 30 April 2021
Accepted: 1 June 2021
Published: 4 June 2021

1. Introduction

Geodemographic classifications provide a convenient method for grouping areas based on the similarity of their underlying characteristics and properties. They have been used to support applications in many different areas including transport [1], marketing [2,3], social inequalities [4], health [5], higher education uptake [6] and other domains concerned with understanding the varying spatial distribution of different types of people living in different types of areas [2]. However, the urban environment is increasingly characterised by rapid changes in neighbourhood (small area) character and composition. This paper reviews the development of geodemographic classifications, which seek to group and label neighbourhoods with similar characteristics, and their underpinning assumptions in the context of examining neighbourhood dynamics. It proposes a measurement framework for capturing neighbourhood character, composition and processes in order to address the limitations of geodemographic classifications in capturing neighbourhood dynamics.

Neighbourhoods tend to be spatially clustered with regard to their underlying socio-economic characteristics, and this provides the basis for geodemographic classifications [7]. These are models that segment areas into homogeneous, statistical clusters [8], with similar multivariate profiles, with the aim of providing a parsimonious approach to quantifying neighbourhood character in order to aid understanding and decision-making [9]. Clusters are given labels that reflect their multivariate properties, such as *multicultural metropolitans* and *rural residents*, and are typically accompanied by pen portraits to provide accessible summaries of typical cluster traits [10]. These are based on the multivariate properties of the cluster centre.

Although convenient, the cluster labels hide the inherent variation associated with any hard classification [11]: individual cluster members frequently have important differences in their multivariate distance to the cluster centre and in their value for any given variable. The result is that potentially important differentiating characteristics for any given area, as well as differences among areas in the same cluster are hidden. This is the same for any classification, but presents problems when the objective is to examine area temporal dynamics.

This paper reviews geodemographic classifications and identifies some of the major limitations in the context of examining area change before describing how a "data primitives" approach [12,13] could be used to both support geodemographics and to identify signals of neighbourhood change, before these changes result in a cluster label change. Such changes in area condition and quality provide important signals that can be used to infer different neighbourhood process and could be used to predict neighbourhood trajectories and future states.

2. Geodemographic Classifications

2.1. Evolution

Empirical research in the early 1900s established a number of principles about the socio-spatial structure of cities [7]. This included the idea of natural areas or geographical units of populations with homogeneous characteristics [14]. Though an extensive literature exists, the sequence of developments can be briefly summarised as follows: Charles Booth depicted spatial patterns in the distribution of social classes in the late 19th Century; the Chicago School devised a model of human ecology to explain patterns in neighbourhood racial and ethnic change [15]; Shevky and Williams [16] created indices of social processes to describe urban society, and social ecologists employed factor analysis on multivariate data for areal differentiation [17]. These developments emphasised the importance of understanding the processes driving neighbourhood character and how these varied in different locations, in order to understand socio-spatial structure and transitions [18]. They underpinned the conceptual and theoretical basis for the emergence of geodemographics in the 1970s [9,19], which coincided with a shift in empirical focus towards the analysis of cross-sectional, but temporally static patterns. At this, time theories of neighbourhood dynamics and process transitions over time such as racial change started to emerge [18], and a disconnect between such theories and the empirical focus was identified.

The first geodemographic classifications were developed in parallel. These included a social area analysis of Liverpool, which later evolved into ACORN (A Classification of Residential Neighbourhoods), a 36-cluster classification of 1971 U.K. census wards [20], and the Potential Rating Index of ZIP Markets (PRIZM), a 40-cluster classification of U.S. census tracts [21]. Singleton and Spielman [19] provided a comparative review of these. They were designed to manage high-dimensional census data to support local government's understanding of the distribution of people and social issues [22]. After an initial public sector focus [20], geodemographics became linked with commercial organisations, where most of the major advancements in the field have been made [23], with applications typically seeking to target consumers for marketing purposes. This is in contrast to the public sector, where geodemographic classifications are used as a policy tool for understanding social phenomena [24], such as health and education inequalities [4,25]. The open licensing of U.K. censuses resulted in the first open classification [26]. The activities at these times focused on describing areal differentiations, rather than advancing social or geodemographic theory and analysis [23].

2.2. Contemporary Classifications

Geodemographic classifications have undergone a series of developmental stages. The first of these integrated market research data to discriminate consumers [27]. Further extensions were initiated to overcome issues associated with the major data source, the population census, which included poor temporal resolution, and a lack of measures related to income, lifestyle and behaviour [23,28,29]. Commercial classifications were at the forefront of these developments, with a focus on improvement through the inclusion of additional data [18]. For example, CAMEO and Mosaic (U.K. and U.S.) now include information from the many new forms of data available including social media information, loyalty card schemes, mobile phone data, customer purchasing records, credit histories and house price, sales and rentals [19,30]. These data have greater temporal resolution (with annual updates, for example), but have been used to augment group descriptions rather than to support cluster reassignment [31].

2.3. Open, Closed and Hybrid Geodemographics

Commercial geodemographic classifications lack any external validation of their potentially subjective allocations related to algorithms, clusters, data inputs, weightings and transformations [32,33]. Some of these can be mitigated via data reduction and algorithm tuning, but these epistemological and semantic aspects are historically hidden from end-users [10]. This is important because classifications contain hidden, embedded assumptions and biases [34], as there is no objectively correct way to classify entities [35]. There are also ethical concerns since geodemographic classification provides the basis for discriminating consumers into target and non-target groups [36], and currently for determining citizen's credit worthiness. Some have argued that model development should thus be explicit, with explanations of how the clusters have been established and derived [35].

As a result, open geodemographic classifications such as the U.K.'s Output Area Classification (OAC) emerged. This uses publicly available census data [37], and the clustering process has been well publicised. The OAC is freely accessible and fully reproducible [33], and similar classifications have subsequently been developed, for example in Ireland [38] and the update to OAC [39].

Reflecting this drive towards transparency in research and data analysis [40,41], hybrid geodemographic classifications were developed. These take advantage of rich and timely data from openly accessible sources in a transparent manner [42], with subsequent developments reflecting the issues of data custodianship, resourcing and access regulations. For example, projects wishing to use these classifications and associated data have to be registered; researchers must be trained to work in data-secure environments and to access secure facilities to analyse the controlled data [42]. However, despite these developments, the proprietary nature of much of the additional data included in hybrid classifications can restrict their use by the wider research community [43].

2.4. Bespoke Classifications

A final evolution relates to purpose. Many commercial systems such as Mosaic (for example, in the U.K., U.S. and Romania) and CAMEO (for example, in Australia, Canada and Japan) are general-purpose and designed for use across different markets and applications [31]. They are available off-the-shelf, but lack specificity. Despite capturing a range of important area characteristics, they frequently lack inferential depth and capture neighbourhood processes of potential interest to differing degrees [38].

Bespoke classifications have been constructed to support specific applications [32], based on domain understanding and underpinned by data that capture the processes of interest [44]. This results in improved targeting and discrimination [31]. Commercial examples include Segmentos developed by EurekaFacts and the Green and Ethical classification developed by Call Credit (now TransUnion) of green behaviours, and research-led examples include classifications of digital inequality in the U.K. [45] and of mortality risk in Cyprus [46].

3. The Limitations of Geodemographic Classifications

Geodemographic classifications have a number of limitations. Most pertinent to the analysis of neighbourhood and area socio-economic processes are their temporally static nature [47], which precludes the analysis of neighbourhood dynamics [48], and the hard (or Boolean) allocation of areas to a single cluster—the one to which they are nearest in a multidimensional feature space. The classification of small areas in this way does not facilitate the analysis of anything other than very coarse or dramatic changes in neighbourhood composition [18], and given that they are usually constructed from decennial census data, any rapid sociodemographic changes may be missed.

3.1. Temporal Dynamics

Some research has examined the temporal nature of geodemographic clusters. Gale and Longley [49] constructed measures to identify areas susceptible to geodemographic change, and their results suggested the presence of several active neighbourhood processes that varied in extent, degree and the geodemographic classes to which they pertained. Singleton et al. [33] used the 2001 and 2011 OACs to create the Temporal OAC and found that 39 percent of areas were reassigned, suggesting a high degree of cluster instability and neighbourhood change. McLachlan and Norman [47] extended these analyses and used three population decadal censuses to examine area changes over time. However, although these studies embraced temporal dynamics, they assumed any local changes were captured first by decadal census data and second within the allocation of areas to clusters and labels. The often incorrect assumption of these classification-based approaches to change is that such temporally coarse data and the process of class allocation are able to adequately quantify area change processes over time [18]. In reality, many subtle, smaller, but nonetheless important changes in area condition and quality, which may occur over shorter time frames and may provide an earlier indicator of cluster change, are missed.

In this respect, geodemographic classifications fail to capture the impacts and cycles of social processes [50] and social change [51]. This is because the processes frequently operate over different spatial and temporal scales [52] to the serial and spatial properties of the data, and there may also be a lack of synchronicity between process phase and measurement frequency [53].

3.2. Hard Classification

The second major limitation is related to the nature of hard allocations of areas to classes. Classification assigns each area to the cluster to which it is closest in a multivariate feature space [44]. Clustering is, by design, a statistically parsimonious process, but results in the loss of potentially important information [54]. Consider two scenarios by way of illustration: (1) areas nearer to a single cluster centre are exemplar members, with all the typical characteristics of the cluster and very few characteristics of any other cluster; (2) areas near multiple cluster centres are allocated to the cluster they are closest to in the feature space, but contain characteristics that are typical of other clusters.

A further implication of hard allocation is the varying magnitude of area change needed for any cluster reassignment. Consider an area close to a single cluster centre in Case (1) above that has experienced large changes in some of its socio-economic properties (and associated variables). These would have to be much larger changes for reallocation into a new cluster than for Case (2) above, since the area in Case (1) is closer to the centre of the cluster's multivariate feature space region than for the area described by Case (2), which is at the cluster periphery.

The implication of this when considering area change and neighbourhood processes is that changes in class are only recorded when the change surpasses a threshold sufficient for the area to be nearest a different cluster centre [53], and this threshold varies for individual areas. Thus, much potentially useful information is ignored in classification or cluster-based approaches to change, despite such information being potential indicators of changes in area condition and quality and being indicative of emergent area-related processes [55].

Some of this could be handled by soft approaches to classification, which retain such information directly, for example as fuzzy memberships to multiple geodemographic classes [11] or as cluster probabilities, obviating the need to aggregate to a single label [56]. Soft classification approaches can capture changes in quality and condition that are not detected by hard classifications, but they can be complex to implement [1] due to the need to link the logic of soft classification change to the processes being investigated (for example, through Type II fuzzy sets [57], which require a different conceptualisation of change).

3.3. Summary

In summary, there are a number of considerations when seeking to examine area and neighbourhood change through geodemographics over time:

- Geodemographic classifications are temporally statistic and fail to capture the dynamic nature of many neighbourhoods;
- Classifications constructed on multiple decadal population censuses may not be sufficiently sensitive to the social processes experienced by neighbourhoods;
- The hard allocation of cluster labels masks the degree to which an individual area is a member of the class;
- When evaluated over time, clustering fails to capture any smaller signals of change or within-cluster changes.

Additionally, currently, data capturing many neighbourhood-related processes are routinely updated with greater frequency than previously [9]. For example, in the U.K., the government publishes annual data over small areas of mid-year population estimates, the number of people receiving different types of social security payment, planning applications (giving an indication of housing pressure), housing affordability and national insurance registrations indicating migration flows (and anecdotally, the biggest driver of neighbourhood change). This suggest there are opportunities for incorporating such data into models and workflows in order to support the analysis of social change and of the processes driving local changes [58,59], as well as to improve the capacity to predict area changes [60]. Together, these indicate the need for a different approach for analysing geodemographic neighbourhood change: data primitives are described in the next section.

4. Data Primitives

The challenge is how to address the limitations described above in order to advance geodemographics. Data availability is much enhanced due to the many new forms of data, as well as increased government reporting of intra-census information. For example, in the U.K., national and local governments publish population estimates, national insurance registrations of foreign nationals, social security registrations and planning applications at annual, quarterly or monthly frequencies, at relatively detailed spatial scales. These provide rich and freely available information about neighbourhoods and the processes they are experiencing. Anecdotally, the biggest driver of area change is related to national insurance registrations, which is available over Medium Super Output Areas (MSOAs) (around 5000 households), but the analysis of this can be finessed by examining social security benefits related to unemployment (Job Seekers' Allowance, Income Support, Housing Benefit, Employment Support Allowance), all of which are reported over Output Areas, which are nested within MSOAs. Other data are available that describe aspects related to public health (such as monthly hospital admissions and two-year aggregates of childhood obesity), as well as wider contextual socio-economic information such as annual changes in housing affordability (i.e., the ratio of house price to annual earnings). The frequency and free availability of these data support different methods for characterising neighbourhoods, ones that are able to examine the neighbourhood dynamics captured by such data. A potentially relevant alternative is to apply a data primitive approach [12,13].

4.1. Defining Data Primitives

Geodemographic classification changes arise from the accumulation of the effects of different neighbourhood-level processes. Identifying measures that capture key aspects of the processes driving these changes would allow neighbourhood dynamics to be captured, examined, analysed and predicted. This explicit consideration of processes that drive changes in the distribution of different socio-economic factors has the potential to support a deeper understanding of society and its spatial organisation, and thus urban structure, whilst also overcoming some of the critical limitations of geodemographic classifications.

Data primitives [12] offer a route to do this. They are the fundamental dimensions or measurements that capture the dynamics of the full range of processes associated with the domain under investigation. They are an extension of "approximation spaces" [61] and "quantified conceptual overlaps" [13]. They were developed in the land-use domain to overcome the difficulties in translating among classification systems and have been extended into the change dimensions [12,13,62]. They operate by identifying qualities or characteristics that different classes have in common (hence overlaps) and use extended set theory to determine class elements that are contained within other classes wholly or partially (hence, approximation spaces). In data primitives, the basic idea is to identify a set of dimensions or measurements that capture the full character of the domain of interest (e.g., land use or social processes), independent of the classification. Ideally, though not always possible, they should be unrelated and, if possible, orthogonal in terms of the characteristics (dimensions) they capture and explain, although recent work with data primitives has shown that orthogonality is less important in terms of discriminating power than first thought [62]. Therefore, in the geodemographic domain, the data primitives should describe components of the neighbourhood-level sociodemographic processes that define neighbourhood character and shape changes over time [63].

Data primitives, if correctly specified, provide a comprehensive foundation for quantifying the underlying processes driving neighbourhood characteristics and, unlike geodemographic classifications, are comparable through space and time. They allow the current "state" of an area to be quantified. They can also be used to quantify state transitions, indicating neighbourhood dynamics, and to predict changes in state. They enhance geodemographic classification approaches because they analyse geodemographic change directly via transitions and support predictive geodemographics.

The key issue with this approach, however, is which dimensions or data primitives to include within this multivariate feature space.

4.2. Data Primitives for Geodemographic Research

Data primitives for geodemographic research should capture the different attributes of the underlying neighbourhood and area social processes that drive change. By way of example, consider gentrification and displacement, two of the most studied neighbourhood processes. Gentrification was first defined by Glass [64], and though there is no singular globally accepted definition [65], some key indicators include the renovation of lower value, older properties by incomers of higher socioeconomic status [66], changes in economic, cultural, political and social characteristics [67], increases in house prices and incomes [68] with the influx of more highly educated residents [69] and increases in inequalities such as health disparities [70]. These suggest the need for measures of migration, education level, house prices and income, to capture changes in neighbourhood characteristics [71].

Displacement is a consequential process of gentrification. Working-class, blue-collar residents are typically displaced by middle-class, white-collar ones [72] because they cannot afford the increased costs of living [73]. In the short term, many original residents benefit from declining poverty and rising house values [74], but over time, working class households experience increased vulnerability, reduced security of tenure, reduced spending power, and reduced employment opportunities [75]. These processes are clearly linked and may occur concurrently, but given the right primitives, captured with the right time frame, such processes should be discernible.

In both cases, these processes are complex and multidimensional, but have a direct impact on neighbourhood character [76]. Within the domain of neighbourhood change, processes such as gentrification and displacement are typically measured through a subjective choice of proxy variables, sourced from demographic data such as population censuses and beyond [73,75,77,78]. The data primitive challenge is to determine which variables capture the mechanisms within the different processes. Table 1 describes an initial set of data primitives for a number of neighbourhood processes. The list of processes is not exhaustive, and they occur over different spatial and temporal extents [52]. Additionally, although the aim was to identify orthogonal measures as primitives, some degree of correlation is present in this initial list of variables.

Table 1. Neighbourhood processes, their characteristics and an initial set of potential data primitives.

Process	Characteristics	Data Primitives
Gentrification	Upward transition of neighbourhood by the influx of residents of higher income and education.	House price (increase) Education level (increase) Income inequality (increase) Migration churn (increase) Professional occupation (increase)
Rural flight	Rural-to-urban migration. Resulting from the industrialisation of agriculture. Exacerbated with the loss of rural services.	Low skilled occupation (decrease) Business vacancy rates (increase)
Urban sprawl	The unrestricted growth of urban areas with little regard for urban planning, generally on the urban fringe. Rapid expansion of the geographical extent of cities and towns.	Population density (increase) Business vacancy rates (decrease)
Displacement	Displacing low-income residents from gentrifying urban developments. Reduced security in tenure, employment opportunities and spending power.	Housing affordability (decrease) Low-skilled occupation (decrease) Income inequality (increase) Migration churn (increase)
Counter-urbanisation	Urban-to-rural migration. Can occur as a reaction to inner-city deprivation. In Europe, it involves de-concentration of one area to another that is beyond suburbanisation.	Income inequality (increase) Population density (decrease) Population flux (out) Unemployment (increase)
Suburbanisation	Urban-to-suburban migration. Can result in suburban sprawl, where low-density peripheral urban areas grow, as households and businesses move out of urban centres.	Population density (decrease) Population flux (out) Business vacancy rates (decrease)
White flight	Sudden or gradual large-scale migration of white people to more racially homogeneous suburban regions.	Ethnic minorities (increase) White ethnicity (decrease) Population density (decrease)
Urban decay	Downward transition of a neighbourhood, or parts of it, into disrepair by several interacting processes such as deindustrialisation and counter-urbanisation. Features increased poverty, fragmented families and low overall living standards and quality of life.	Unemployment (increase) Low-skilled occupation (decrease) Poor health (increase) Income inequality (increase) House price (decrease)
Deindustrialisation	The removal or reduction of industrial activity. Long-term decline in the output of manufactured goods or in employment in the manufacturing sector, shifting to the services sector.	Low skilled occupation (decrease) Unemployment (increase)
Municipal disinvestment	Urban planning process of abandonment, typically the poorest communities. Tends to fall along racial and class lines, perpetuating the cycle of poverty, since affluent individuals have greater social mobility.	Ethnic minorities (increase) Income inequality (increase)
Shrinking cities	Notable in the U.S. Dense cities experience notable population loss, often due to emigration. Cities that focus on one branch of economic growth are vulnerable.	Population density (decrease) Low-skilled occupation (decrease) Unemployment (increase)
Neighbourhood churn	The influx and outflux of residents such that the social character remains the same, but population turnover is high.	Population flux (in) Population flux (out)
International migration	The immigration of people from foreign countries. They tend to locate to the deprived inner-city where costs are lower and locate to established cultural neighbourhoods.	Population flux (in) Ethnic minorities (increase) Housing affordability (increase)

4.3. Analysing State and Change

Tracking an area's multivariate position through time allows state and changes in state to be identified, thereby capturing changes in the processes associated with a given neighbourhood. This is performed by examining current positions in the data primitive multidimensional feature space (state), shifts in feature space position over time (changes in state) and capturing these trajectories, for example, through a change vector approach [79]. Trajectories can be used to create predictive models to infer future feature space positions. The data primitives described in Table 1 suggest a multivariate data primitive feature space composed of the following area-level measurements:

- income inequality;
- occupation;
- unemployment;
- population density;
- population flux;
- ethnicity;
- housing affordability;
- house price;
- education;
- poor health;
- migration churn;
- business vacancy rates.

Inevitably, these data are of different types, and a number of questions remain at this stage. First, capturing data at appropriate spatial and temporal resolutions for each primitive is important, with some primitives having greater critical update constraints than others (population density and population flux, for example). Similarly, house price could be the average house price regardless of size (as is currently the case in the U.K.), price per square metre or even price per bedroom. Others will be harder to partition. What for example are the professions that should be included in "professional occupations" or "low-skilled occupations" under occupations? Should income inequality be defined according to the standard Gini coefficient measure or in a more relative manner? These are local application-level decisions, and any contributing data used to support or create a primitive can be retained for later changes in understanding or definition.

However, these measurements recorded at appropriate spatial and temporal resolutions allow the state of any area to be characterised at any given point in time, without resorting to simplistic and reductive geodemographic classification labels. In this context, the notion of change is different from current approaches that focus on changes in class label: here, change is quantified by determining differences in state at two different times. Under a data primitive approach, change is the shift in position in a multivariate feature space, removing the constraints of a cluster-to-cluster change. Change in a multivariate feature space of data primitives relates to differences in the *relative* position of areas over time, thereby capturing smaller, but potentially more locally relevant neighbourhood changes than with cluster analysis. Understanding the importance of shifts in multivariate feature space requires knowledge of the processes that are indicated by the shift and their likely trajectories.

This approach suggests that individual processes will be represented by vectors of change, rather than occupying specific regions such as geodemographic clusters. Vector approaches to change have long been used in remote sensing classifications [79], where changes in position are used to infer a new land cover class based on the magnitude and direction of the change vector [80].

In the change vector approach, the positions of each neighbourhood or area are determined in a multivariate feature space, and as new data become available, changes in position can be quantified using the change vector.

We suggest that such approaches could be used to infer area changes, both of the geodemographic class if that were required, but also to indicate the processes associated with the observed change, providing a more sensitive and nuanced approach to the analysis of temporal neighbourhood change. Thus, gentrification is a process that changes the relative position of an area in the dimensions of house price (increase), education level (increase), income inequality (increase) and internal migration churn (increase). Such changes can also be representative of displacement. However, the most defining difference is that displacement is also associated with changes in the relative position of in the dimensions of housing affordability (decrease).

Trajectories of change can be inferred through the analysis of changes in multivariate feature space position. These could be used to predict future states associated with specific processes, and an area's progression through a process can be examined, explained and predicted. The data primitive approach has the potential to support enhanced analyses for applications in the public sector that currently use geodemographic classifications, by providing timely, area-specific characterisations and trajectories of change, built from data routinely collected by local and central governments.

4.4. Case Study Illustration

To illustrate the data primitive approach, some initial data were gathered for Lower Super Output Areas (LSOAs) England and Wales. LSOAs containing around 1500 people were designed as part of a nested set of census reporting units [81] for the U.K. There are some 34,000 LSOAs in the U.K. Annual data for 2010 to 2016 for a number of the primitives were assembled from diverse sources for each LSOA: population density (people per 1 km^2), the proportion of the population who were white British, housing affordability, although this was at the local authority level, not the LSOA, average house price, population, the population receiving some form of disability living allowance, the proportion of households that have changed, the proportion of the working population in professional occupations and the proportion of the working population that were unemployed. The data sources and acronyms are listed in Table 2.

Table 2. The data used in the illustrative case study.

Acronym	Description	Source
POPD	Population density (people per 1 km^2)	Derived from census areas and population data
WBR	Proportion white British	From the Consumer Data Research Centre (CDRC) (see https://www.cdrc.ac.uk, accessed on 18 January 2021)
HAFF	Housing affordability	From the Office of National Statistics (ONS) (see https://www.ons.gov.uk, accessed on 3 November 2011)
HP	Average house price (in 1000 GBP)	From ONS (link above)
POP	Population total	From ONS (link above)
DLA	Proportion receiving disability living allowance	From StatXplore (see https://stat-xplore.dwp.gov.uk/, accessed on 10 April 2021)
CHN	Proportion of households that have changed	From the CDRC Residential Mobility Index (link above)
PROF	Proportion in professional occupations	From the ONS Standard Industrial Classification (link above)
UNEMP	Proportion unemployed	From StatXplore (as above)

Each variable for each year was transformed to z-scores (i.e., with a mean of zero and a standard deviation of one). The transformed data were used to calculate multivariate angles and distances for the period 2010 to 2016, by modifying the code in the rasterCVA function included as part of the RStoolbox R package [82]. Figure 1 shows these for the Nottingham area, and it was evident that the change (magnitude) was greater around the city centre, but that the nature of those changes as indicated by the direction of angle of the change vector was spatially clustered.

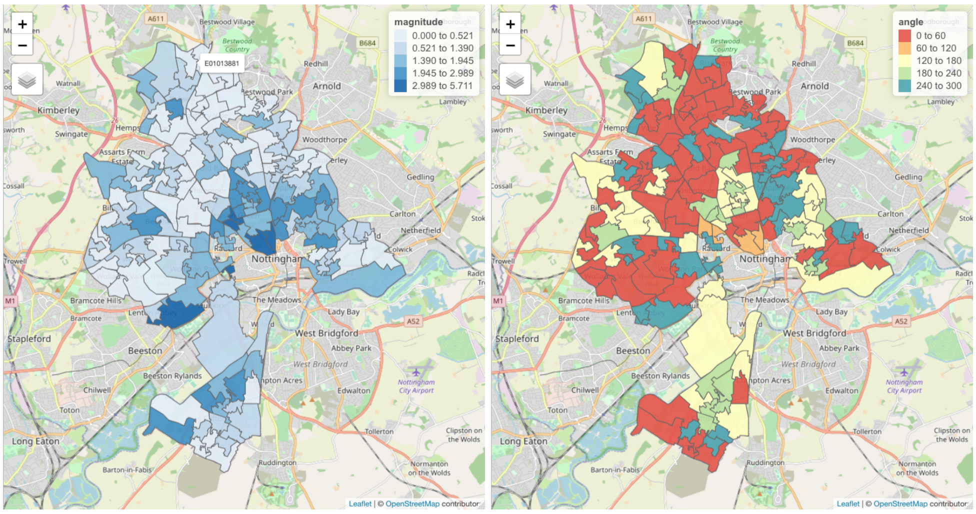


Figure 1. The results of the change vector analysis for LSOAs in the Nottingham area, both with an OpenStreetMap backdrop: (**left**) The magnitude of change. (**right**) The angle or direction of change.

It is perhaps more instructive to examine individual LSOAs. Four were selected, shown in Figure 2, to demonstrate how areas with seemingly similar changes (the vector magnitude) experience different processes, as shown by the vector angle. The rescaled values for 2010 were subtracted from the rescaled values in 2016 for the eight domains used to calculate the change vector, as shown in Table 3.

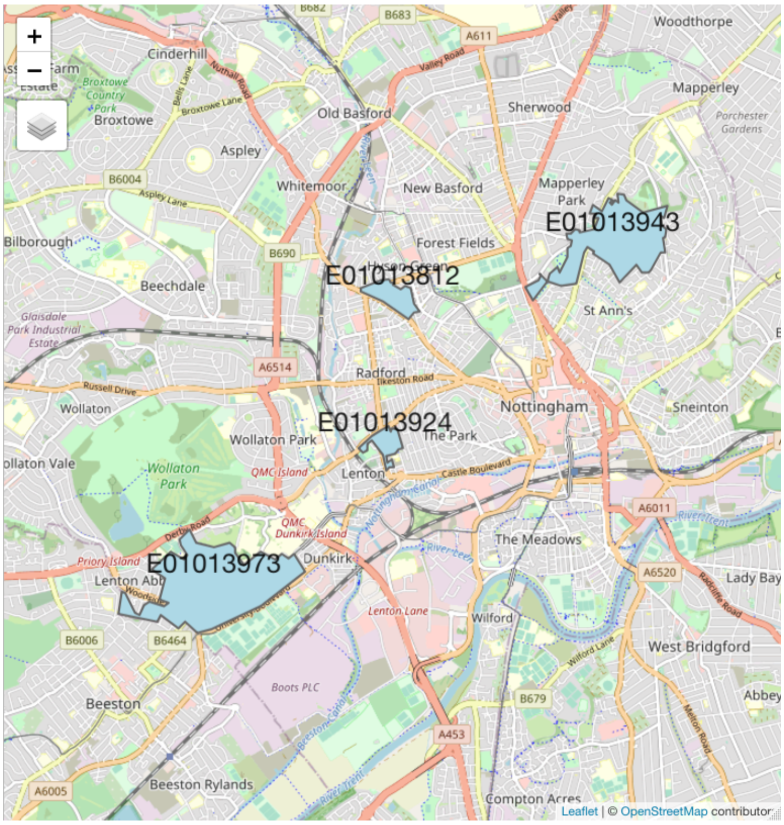


Figure 2. Four example LSOAs in Nottingham, with an OpenStreetMap backdrop.

Table 3. The changes in transformed values (2010 to 2016) for 4 LSOAs in Nottingham.

Primitive	E01013812	E01013924	E01013943	E01013973
angle	177.58	253.083	353.713	342.747
magnitude	4.101	3.935	2.692	5.711
POPD	0.028	−1.497	−0.027	−0.345
PROF	−0.666	−0.455	0.247	1.109
WBR	−0.284	0.071	−0.185	0.561
HAFF	0.294	0.294	0.294	0.294
HP	0.362	−0.266	0.585	0.211
POP	0.167	−1.335	−0.377	−5.411
DLA	−0.159	−1.586	−0.105	0.319
CHN	−0.292	−0.797	−2.178	−0.924
UNEMP	3.992	−2.819	1.352	0.766

We can see that the each LSOA experienced different net changes. E01013812 experienced large increases in unemployment (UNEMP changed by nearly four standard deviations) and reductions in the proportion of people in professional occupations (PROF). E01013943 experienced large in- and out-migrations since 2010 (CHN), as well as increases in unemployment (UNEMP). There were some similarities in these areas (for example, both experienced relative increases in house price (HP); both are relatively deprived areas; but one is in the process of starting to become gentrified (E01013943) with increases in professionals attracted by the proximity to the city centre and the less expensive housing stock; the other is still experiencing decline).

The other two areas were both dominated by students, but one was emerging as a student area (E01013924) and the other (E01013973) consolidating, as it already had a strong student presence. E01013924 grew into more of a student area over this period, potentially because of the relative decline in house prices (inexpensive properties near the university), and as residents moved out, unemployment (UNEMP) and population density (POPD) declined. E01013973 covered the university campus and surroundings, and the consolidation of this area as a student one (i.e., heavy studentification) was shown by the changes in households (CHN), population (POP) and population density (POPD).

These differences among the areas are illustrated Figures 3 and 4. Figure 3 shows the angles and magnitudes of change and the radar plots of the variable changes in Figure 4 indicate the origins of these. There is much more that could be done here (examining the annual shifts, exploring a greater number of primitives, etc.), but the purpose of the case study was to provide an illustration of what can be done and where it may lead, without either reducing all of this information into a geodemographic class or a composite indicator of some kind, both of which mask any subtly emerging processes.

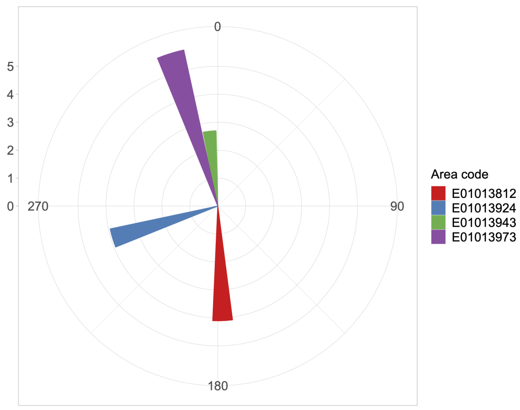


Figure 3. Polar plots of the magnitude and direction of the change vector for the 4 LSOAs in Nottingham.

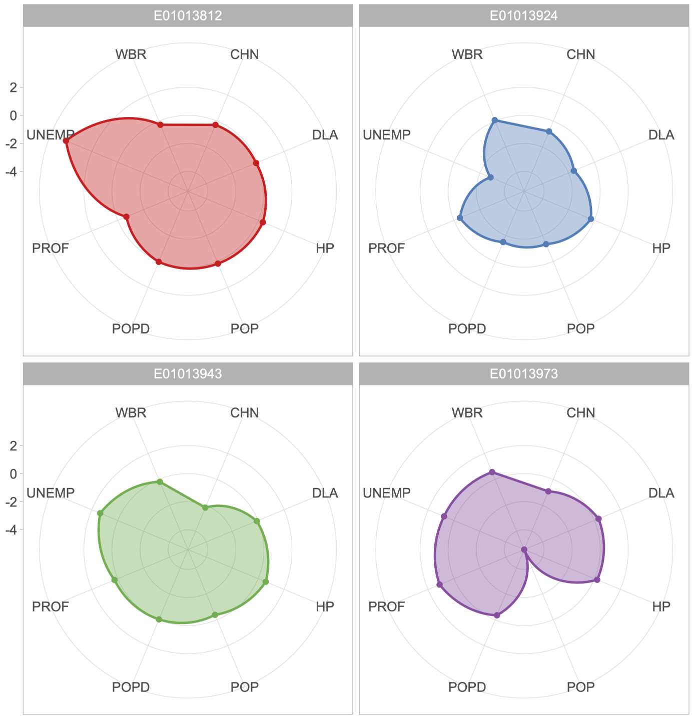


Figure 4. Radar plots of the changes from 2010 to 2016 in each of the 4 Nottingham LSOAs.

4.5. Problems Yet to Be Solved

A key issue in this approach is the availability of spatio-temporal data to underpin the primitives. Some countries suffer from data availability issues in terms of data existence, access and adequate spatial and temporal resolution. In the U.K., for example, many data are open, describing small area statistics, with annual or quarterly updates, as described above. This does not apply to many other countries, hindering the application of the data primitive approach. However, opportunities in these areas should arise as many new forms of data, from a variety of formal and informal platforms, become more widely available and accessible. A second issue is that the processes listed Table 1 operate at very different spatial scales such as rural flight, urban sprawl and counter-urbanisation. This requires some consideration of the scale at which the data are available and potentially the use of multi-level modelling approaches or similar to accommodate multiple process and data scales. Third, the list of processes and primitives we suggested in Table 1 is at this stage indicative. We are in the process of undertaking research to investigate the sensitivity of the data we have available to capture information about the dynamics of these processes. Future work will investigate and report on these issues.

5. Conclusions

Geodemographic classifications have developed considerably from their original foundations. They are heavily used in commerce, public policy and research, but have several limitations. These include a failure to capture neighbourhood dynamics [48,50] and the assumptions associated with the use of hard classifications, which although convenient, provide overly simplistic descriptions of neighbourhood character and require some threshold of change to be surpassed for a new class label to be assigned. The result is that subtle, but important changes in an area's condition from an accumulation of neighbourhood process may be missed. This paper proposes the adoption of a data primitive approach [12] arising from other strands of research examining geographic classifications. Such approaches have the potential to address these shortcomings and allow geodemographic research to take advantage of the many spatio-temporal data that are produced quarterly or annually over small areas, as well as the many new forms of data. In many ways, this approach operationalises the wider ideas behind the seminal work of Massey and Denton in 1988 [83] in their exploration of the dimensions of segregation by taking advantage of our data-rich era and extending into other area-level processes. Data primitives are the fundamental dimensions or measurements that capture the characteristics of the process under investigation. They use a multidimensional feature space to quantify the current state and changes in state. They can be used to create classifications if required, but critically, they support predictive geodemographics through the modelling and analysis of state trajectories. We suggested a set of primitives that could be used to characterise a range of social and economic processes experienced by neighbourhoods. These will allow the emergence of different neighbourhood-level processes to be quantified and enable geodemographic research to generate more nuanced outputs, thereby enhancing support for strategic planning of services to meet the demand and needs of changing populations.

Author Contributions: Conceptualisation, Jennie Gray, Lisa Buckner and Alexis Comber; writing—original draft preparation, Jennie Gray; writing—review and editing, Jennie Gray, Lisa Buckner and Alexis Comber. All authors read and agreed to the published version of the manuscript.

Funding: This research was funded by the ESRC Centre for Doctoral Training-Data Analytics and Society, Grant Number ES/P000401/1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Y.; Cheng, T. Understanding public transit patterns with open geodemographics to facilitate public transport planning. *Transp. Transp. Sci.* **2020**, *16*, 76–103. [CrossRef]
2. Harris, R.; Sleight, P.; Webber, R. *Geodemographics, GIS and Neighbourhood Targeting*; John Wiley & Sons: Hoboken, NJ, USA, 2005; Volume 8.
3. Mitchell, V.W.; McGoldrick, P.J. The role of geodemographics in segmenting and targeting consumer markets: A Delphi study. *Eur. J. Mark.* **1994**, *28*, 54–72. [CrossRef]
4. Xiang, L.; Stillwell, J.; Burns, L.; Heppenstall, A.; Norman, P. A geodemographic classification of sub-districts to identify education inequality in Central Beijing. *Comput. Environ. Urban Syst.* **2018**, *70*, 59–70. [CrossRef]
5. Petersen, J.; Gibin, M.; Longley, P.; Mateos, P.; Atkinson, P.; Ashby, D. Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *J. Geogr. Syst.* **2011**, *13*, 173–192. [CrossRef]
6. Singleton, A.D.; Wilson, A.; O'Brien, O. Geodemographics and spatial interaction: An integrated model for higher education. *J. Geogr. Syst.* **2012**, *14*, 223–241. [CrossRef]
7. Alexiou, A.; Singleton, A. Geodemographic analysis. *Geocomputation: A Practical Primer*; SAGE: London, UK, 2015; pp. 137–151.
8. Abbas, J.; Carlin, H.; Cunningham, A.; Dedman, D.; McVey, D. Technical Briefing 5: Geodemographic Segmentation. 2009. Available online: https://fingertips.phe.org.uk/profile/guidance (accessed on 4 April 2021)
9. Webber, R.; Burrows, R. *The Predictive Postcode: The Geodemographic Classification of British Society*; Sage: Thousand Oaks, CA, USA, 2018.
10. Singleton, A.D.; Longley, P.A. Creating open source geodemographics: Refining a national classification of census output areas for applications in higher education. *Pap. Reg. Sci.* **2009**, *88*, 643–666. [CrossRef]
11. Fisher, P.; Tate, N.J. Modelling class uncertainty in the geodemographic Output Area Classification. *Environ. Plan. Plan. Des.* **2015**, *42*, 541–563. [CrossRef]
12. Comber, A.J. The separation of land cover from land use using data primitives. *J. Land Use Sci.* **2008**, *3*, 215–229. [CrossRef]
13. Wadsworth, R.; Balzter, H.; Gerard, F.; George, C.; Comber, A.; Fisher, P. An environmental assessment of land cover and land use change in Central Siberia using quantified conceptual overlaps to reconcile inconsistent data sets. *J. Land Use Sci.* **2008**, *3*, 251–264. [CrossRef]
14. Herbert, D.; Thomas, C. *Cities in Space: City as Place*; Routledge: London, UK, 2013.
15. Park, R.E. The city: Suggestions for the investigation of human behavior in the city environment. *Am. J. Sociol.* **1915**, *20*, 577–612. [CrossRef]
16. Shevky, E.; Williams, M. *The Social Areas of Los Angeles*; University of California: Berkeley, CA, USA, 1949.
17. Rees, P.H. Factorial ecology: An extended definition, survey, and critique of the field. *Econ. Geogr.* **1971**, *47*, 220–233. [CrossRef]

18. Reibel, M.; Regelson, M. Neighborhood racial and ethnic change: The time dimension in segregation. *Urban Geogr.* **2011**, *32*, 360–382. [CrossRef]
19. Singleton, A.D.; Spielman, S.E. The past, present, and future of geodemographic research in the United States and United Kingdom. *Prof. Geogr.* **2014**, *66*, 558–567. [CrossRef] [PubMed]
20. Webber, R. An introduction to the national classification of wards and parishes. *Plan. Res. Appl. Group Tech. Pap.* **1977**, *23*.
21. Weis, M.J. *The Clustering of America*; Number 306.0973 W426c; Perennial Library: New York, NY, USA, 1989.
22. Birkin, M.; Clarke, G. Geodemographics. In *International Encyclopaedia of Human Geography*; Elsevier: Hoboken, NJ, USA, 2009.
23. Longley, P. Geographical information systems: A renaissance of geodemographics for public service delivery. *Prog. Hum. Geogr.* **2005**, *29*, 57–63. [CrossRef]
24. Sabater, A. Between flows and places: Using geodemographics to explore EU migration across neighbourhoods in Britain. *Eur. J. Popul.* **2015**, *31*, 207–230. [CrossRef]
25. Moon, G.; Twigg, L.; Jones, K.; Aitken, G.; Taylor, J. The utility of geodemographic indicators in small area estimates of limiting long-term illness. *Soc. Sci. Med.* **2019**, *227*, 47–55. [CrossRef]
26. Charlton, M.; Openshaw, S.; Wymer, C. Some new classifications of census enumeration districts in Britain: A poor mans ACORN. *J. Econ. Soc. Meas.* **1985**, *13*, 69–96.
27. Baker, K.; Bermingham, J.; McDonald, C. The utility to market research of the classification of residential neighbourhoods. *Mark. Res. Soc. J.* **1997**, *39*, 1–12. [CrossRef]
28. Howick, R. Building neighbourhood classifications-data sources and their geographic integration. In *ESRC Transdisciplinary/Research Methods Seminar Series*; UCL: London, UK, 2004; pp. 18–19.
29. Birkin, M. Customer targeting, geodemographics and lifestyle approaches. In *GIS for Business and Service Planning*; Longley, P.; Ed.; John Wiley: New York, NY, USA, 1995; pp. 104–149.
30. Longley, P.A.; Adnan, M. Geo-temporal Twitter demographics. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 369–389. [CrossRef]
31. Leventhal, B. *Geodemographics for Marketers: Using Location Analysis for Research and Marketing*; Kogan Page Publishers: London, UK, 2016.
32. Pratt, M.D.; Longley, P.A.; Cheshire, J.; Gale, C. Open Data Sources for Domain Specific Geodemographics. GISRUK Conference, 2013. Available online: https://www.geos.ed.ac.uk/~gisteac/proceedingsonline/GISRUK2013/ (accessed on 2 June 2021)
33. Singleton, A.; Pavlis, M.; Longley, P.A. The stability of geodemographic cluster assignments over an intercensal period. *J. Geogr. Syst.* **2016**, *18*, 97–123. [CrossRef]
34. Feinberg, M. Hidden bias to responsible bias: An approach to information systems based on Haraway's situated knowledges. *Inf. Res.* **2007**, *12*, 12–14.
35. Mai, J.E. Classification in a social world: Bias and trust. *J. Doc.* **2010**, *66*, 627–642. [CrossRef]
36. Burrows, R.; Gane, N. Geodemographics, software and class. *Sociology* **2006**, *40*, 793–812. [CrossRef]
37. Vickers, D.; Rees, P. Creating the UK National Statistics 2001 output area classification. *J. R. Stat. Soc. Ser.* **2007**, *170*, 379–403. [CrossRef]
38. Yazgi Walsh, B.; Brunsdon, C.; Charlton, M. Open Geodemographics: Classification of Small Areas, Ireland 2016. *Appl. Spat. Anal. Policy* **2021**, *14*, 51–79. [CrossRef]
39. Gale, C.G.; Singleton, A.D.; Bates, A.G.; Longley, P.A. Creating the 2011 area classification for output areas (2011 OAC). *J. Spat. Inf. Sci.* **2016**, *2016*, 1–27. [CrossRef]
40. Brunsdon, C.; Comber, A. Opening practice: Supporting reproducibility and critical spatial data science. *J. Geogr. Syst.* **2020**, 1–20. [CrossRef]
41. Brunsdon, C.; Comber, A. Big issues for big data: Challenges for critical spatial data analytics. *J. Spat. Inf. Sci.* **2020**, *2020*, 89–98.
42. Singleton, A.D.; Longley, P.A. Data infrastructure requirements for new geodemographic classifications: The example of London's workplace zones. *Appl. Geogr.* **2019**, *109*, 102038. [CrossRef]
43. Webber, R.J.; Butler, T.; Phillips, T. Adoption of geodemographic and ethno-cultural taxonomies for analysing Big Data. *Big Data Soc.* **2015**, *2*. [CrossRef]
44. Hjørland, B.; Pedersen, K.N. A substantive theory of classification for information retrieval. *J. Doc.* **2005**, *61*, 582–597. [CrossRef]
45. Singleton, A.; Alexiou, A.; Savani, R. Mapping the geodemographics of digital inequality in Great Britain: An integration of machine learning into small area estimation. *Comput. Environ. Urban Syst.* **2020**, *82*, 101486. [CrossRef]
46. Lamnisos, D.; Middleton, N.; Kyprianou, N.; Talias, M.A. Geodemographic Area Classification and association with mortality: An ecological study of small areas of Cyprus. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2927. [CrossRef] [PubMed]
47. McLachlan, G.; Norman, P. Analysing Socio-Economic Change Using a Time Comparable Geodemographic Classification: England and Wales, 1991–2011. *Appl. Spat. Anal. Policy* **2021**, *14*, 89–111. [CrossRef]
48. Longley, P.A. Geodemographics and the practices of geographic information science. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 2227–2237. [CrossRef]
49. Gale, C.G.; Longley, P.A. Temporal uncertainty in a small area open geodemographic classification. *Trans. GIS* **2013**, *17*, 563–588. [CrossRef]
50. Prouse, V.; Grant, J.L.; Ramos, H.; Radice, M. *Assessing Neighbourhood Change: Gentrification and suburban Decline in a Mid-Sized City*; School of Planning, Dalhousie University: Halifax, NS, Canada, 2015.
51. Batliwala, S. Measuring social change: Assumptions, myths and realities. *Alliance* **2006**, *11*, 12–14.

52. An, L.; Tsou, M.H.; Crook, S.E.; Chun, Y.; Spitzberg, B.; Gawron, J.M.; Gupta, D.K. Space–time analysis: Concepts, quantitative methods, and future directions. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 891–914. [CrossRef]
53. Comber, A.; Wulder, M. Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Trans. GIS* **2019**, *23*, 879–891. [CrossRef]
54. Grekousis, G.; Thomas, H. Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and Gustafson–Kessel methods. *Appl. Geogr.* **2012**, *34*, 125–136. [CrossRef]
55. Zhu, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 370–384. [CrossRef]
56. See, L.; Openshaw, S. Fuzzy geodemographic targeting. In *Regional Science in Business*; Springer: Berlin, Germany, 2001; pp. 269–281.
57. Fisher, P. What is Where? Type-2 Fuzzy Sets for Geographical Information [Research Frontier]. *IEEE Comput. Intell. Mag.* **2007**, *2*, 9–14. [CrossRef]
58. Adnan, M.; Longley, P.A.; Singleton, A.D.; Brunsdon, C. Towards real-time geodemographics: Clustering algorithm performance for large multidimensional spatial databases. *Trans. GIS* **2010**, *14*, 283–297. [CrossRef]
59. Weiser, P.; Frank, A.U. Dynamic GIS–The final frontier. In *Extended Abstract*; GI-Forum: Salzburg, Austria, 2010.
60. Christakos, G.; Bogaert, P.; Serre, M. *Temporal GIS: Advanced Functions for Field-Based Applications*; Springer Science & Business Media: Berlin, Germany, 2012.
61. Ahlqvist, O. A parameterized representation of uncertain conceptual spaces. *Trans. GIS* **2004**, *8*, 493–514. [CrossRef]
62. Comber, A.; Kuhn, W. Fuzzy difference and data primitives: A transparent approach for supporting different definitions of forest in the context of REDD+. *Geogr. Helv.* **2018**, *73*, 151–163. [CrossRef]
63. Buzar, S.; Ogden, P.E.; Hall, R. Households matter: The quiet demography of urban transformation. *Prog. Hum. Geogr.* **2005**, *29*, 413–436. [CrossRef]
64. Glass, R. *Introduction: Aspects of Change in Centre for Urban Studies*; Mac Gibbon: London, UK, 1964.
65. Lees, L.; Shin, H.B.; López-Morales, E. *Planetary Gentrification*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
66. Ley, D.; Yang, Q. Global gentrifications: Uneven development and displacement; and planetary gentrification. *AAG Rev. Books* **2017**, *5*, 112–115. [CrossRef]
67. Smith, N.; Williams, P. Alternatives to orthodoxy: Invitation to a debate. In *Gentrification of the City*; Routledge: London, UK, 1986; pp. 1–10.
68. Glaeser, E.L.; Kim, H.; Luca, M. Nowcasting gentrification: Using yelp data to quantify neighborhood change. *AEA Pap. Proc.* **2018**, *108*, 77–82. [CrossRef]
69. Chapple, K.; Zuk, M. Forewarned: The use of neighborhood early warning systems for gentrification and displacement. *Cityscape* **2016**, *18*, 109–130.
70. Gibbons, J.; Barton, M.; Brault, E. Evaluating gentrification's relation to neighborhood and city health. *PLoS ONE* **2018**, *13*, e0207432. [CrossRef]
71. Lees, L.; Slater, T.; Wyly, E.K. *The Gentrification Reader*; Routledge: London, UK, 2010; Volume 1.
72. Modai-Snir, T.; van Ham, M. Structural and exchange components in processes of neighbourhood change: A social mobility approach. *Appl. Spat. Anal. Policy* **2019**, *12*, 423–443. [CrossRef] [PubMed]
73. Atkinson, R. Measuring gentrification and displacement in Greater London. *Urban Stud.* **2000**, *37*, 149–165. [CrossRef]
74. Brummet, Q.; Reed, D. The Effects of Gentrification on the Well-Being and Opportunity of Original Resident Adults and Children 2019. FRB of Philadelphia Working Paper No. 19-30. Available online: http://dx.doi.org/10.21799/frbp.wp.2019.30 (accessed on 4 April 2021)
75. Atkinson, R.; Wulff, M.; Reynolds, M.; Spinney, A. Gentrification and displacement: The household impacts of neighbourhood change. *AHURI Final. Rep.* **2011**, *160*, 1–89.
76. Ilic, L.; Sawada, M.; Zarzelli, A. Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. *PLoS ONE* **2019**, *14*, e0212814. [CrossRef]
77. Barton, M. An exploration of the importance of the strategy used to identify gentrification. *Urban Stud.* **2016**, *53*, 92–111. [CrossRef]
78. Reades, J.; De Souza, J.; Hubbard, P. Understanding urban gentrification through machine learning. *Urban Stud.* **2019**, *56*, 922–942. [CrossRef]
79. Xu, R.; Lin, H.; Lü, Y.; Luo, Y.; Ren, Y.; Comber, A. A modified change vector approach for quantifying land cover change. *Remote. Sens.* **2018**, *10*, 1578. [CrossRef]
80. Tewkesbury, A.P.; Comber, A.J.; Tate, N.J.; Lamb, A.; Fisher, P.F. A critical synthesis of remotely sensed optical image change detection techniques. *Remote. Sens. Environ.* **2015**, *160*, 1–14. [CrossRef]
81. Martin, D. 2001 Census output areas: From concept to prototype. *Trends* **1998**, *94*, 19–24
82. Leutner, B.; Horning, N.; Schwalb-Willmann, J.; Hijmans, R. RStoolbox: Tools for remote sensing data analysis. R Package Version 0.1; 2017. Available online: https://www.researchgate.net/publication/312456069_RStoolbox_Tools_for_Remote_Sensing_Data_Analysis (accessed on 4 April 2021)
83. Massey, D.S.; Denton, N.A. The dimensions of residential segregation. *Soc. Forces* **1988**, *67*, 281–315. [CrossRef]

Chapter 5: Identifying Neighbourhood Change Using a Data Primitive Approach: the Example of Gentrification

Overview

This paper provides as a *proof of concept* for the data primitive approach.

This paper highlights the importance of understanding neighbourhood change at a small area level, and explores the use of a data primitive approach to identify neighbourhood change, specifically gentrification. They are used alongside Change Vector Analysis to quantify the direction and magnitude of change in the data, to identify areas of gentrification in a region in the United Kingdom. The results show that the data primitive approach is effective in identifying neighbourhood change and gentrification, as it can capture small-scale changes that may be missed by traditional methods. It explores the establishment and manifestation of the gentrification cycles, before it validates seven of the identified areas of gentrification, finding specific types of gentrification. Overall, the article demonstrates the potential of the data primitive approach and change vector analysis in measuring and understanding neighbourhood change, in the context of gentrification. The findings have important implications for urban planning and policy, with the capability of providing valuable insights for policy makers and researchers, highlighting the need to consider dynamic small area-level changes in the development of effective interventions to address gentrification and its consequences.

Applied Spatial Analysis and Policy
https://doi.org/10.1007/s12061-023-09509-y

Identifying Neighbourhood Change Using a Data Primitive Approach: the Example of Gentrification

Jennie Gray[1,2] · Lisa Buckner[3] · Alexis Comber[1,2]

Received: 11 March 2022 / Accepted: 16 February 2023

Abstract

Data primitives are the fundamental measurements or variables that capture the process under investigation. In this study annual data for small areas were collated and used to identify and characterise gentrification. Such data-driven approaches are possible because of the increased availability of data over small areas for fine spatial and temporal resolutions. They overcome limitations of traditional approaches to quantifying geodemographic change. This study uses annual data for 2010–2019 of House Price, Professional Occupation, Residential Mobility (in and out flows) and Ethnicity over small areas, Lower Super Output Areas (LSOAs). Areas of potential gentrification were identified from directional changes found in all of these variables, across combinations of start and end time periods. The initial set of areas were further processed and filtered to select robust gentrification cycles with minimum duration, and to determine start, peak and end years. Some 123 neighbourhoods in a regional case study area were found to have undergone some form of potential gentrification. These were examined further to characterise their spatial context and nature of the gentrification present, and specific types of gentrification were found to have specific periodicities. For example short-length durations (three to four years) were typically located in rural and suburban areas, associated with transit-induced cycles of gentrification, and greenification. Seven neighbourhoods were validated in detail, confirming the gentrification process and its type and their multivariate change vectors were examined. These showed that vector angle reflects the main data primitive driving the cycle of gentrification, which could aid with future prediction of gentrification cycles. A number of areas of further work are discussed.

Keywords Urban dynamics · Neighbourhood processes · Gentrification · State and change · Neighbourhood change · Data primitives

✉ Jennie Gray
gyjhg@leeds.ac.uk

1 School of Geography, University of Leeds, Leeds, UK

2 Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

3 School of Sociology and Social Policy, University of Leeds, Leeds, UK

Published online: 08 March 2023

Introduction

Geodemographic classifications summarise the socio-economic characteristics of areas and neighbourhoods. They are generated from statistical clustering of socio-economic data and provide an accessible shorthand of the characteristics of the people living within small areas. They are used by decision makers to target policy interventions and by commercial organisations to determine different market segments. Geodemographic classifications are also increasingly used to analyse neighbourhood and small area changes over time (for example, see Singleton et al., 2016; McLachlan & Norman, 2021), and to support cross-sectional studies to evaluate how social processes change over time (Reibel & Regelson, 2011). However, the geodemographic classifications and the data they are constructed from are not well-suited change analysis, for example of the temporal dynamics of social processes for two reasons. First, geodemographic classifications are typically constructed from decennial population census data (Leventhal, 2016), and change analysis that compare geodemographic class at different times requires those processes to manifest themselves within the time interval being considered (Reibel & Regelson, 2011). While some patterns of change may be captured, any findings will be dependent upon that temporal sampling frame. Consequently, matching the temporal sampling and intervals of analysis with the periodicity of the phenomenon being investigated is critically important (Comber & Wulder, 2019) but frequently overlooked in neighbourhood change analyses undertaken in this way. Second, there is an inherent limitation to the information content of statistical clusters like geodemographic classifications and their ability to capture socio-economic processes such as gentrification. This makes it difficult to capture neighbourhood dynamics through evaluation of cluster label change. In statistical clustering the typical properties of each class are defined in a multivariate feature space. Each observation is allocated to the cluster (class) to which it is nearest in this space. Small differences or changes in the socio-economic properties of each observation (for example in unemployment), although important in those areas, may not be sufficient to warrant a change in cluster label, due to the stability of other factors. As such, classification-based approaches to change analysis require multiple dramatic changes in socio-economic features for change to be identified (Reibel & Regelson, 2011).

An alternative to overcome the methodological limitations of geodemographic classifications in their ability to capture neighbourhood dynamics is to use a data primitives approach. Data primitives are the fundamental measurements that capture the processes under investigation (Comber, 2008; Wadsworth et al., 2008). Ideally, they are orthogonal, with each primitive defined to capture a dimension or property of the system or process. In this sense they are similar to Ahlqvist's conceptual spaces (Ahlqvist, 2004). Examining changes in data primitives has been proposed as a novel approach for capturing neighbourhood dynamics (Gray et al., 2021). In this approach the positions of neighbourhood areas in a multivariate feature space are evaluated at different times to identify the presence of neighbourhood change. Gray et al. (2021) identified the variables and the expected direction of change in those variables that would capture different neighbourhood processes such as Gentrification, Urban Decay, and Suburbanisation. A final novel component of the proposed approach is the inclusion of a change vector analysis (CVA) of the multitemporal data primitives. CVA was developed to determine change in land cover class by examining the magnitude and direction of change in a multivariate feature space composed of remote sensing image bands captured at two time periods (Bovolo & Bruzzone, 2007). Here it is used to explore the drivers of change for areas identified has having gentrified. Such data driven approaches to neighbourhood change detection and for capturing neighbourhood processes and dynamics are increasingly possible because of the greater availability and frequency of socio-economic data for small areas.

This paper uses the multitemporal data primitive approach outlined above to undertake an analysis of small area changes, in order to identify neighbourhoods undergoing gentrification. A thorough sweep of the data was undertaken identifying areas experiencing different types of gentrification, to differing degrees, at different rates, at different times, and driven by different processes.

Background: Data Primitives for Gentrification

Gentrification is a well-studied but controversial neighbourhood process (Lester & Hartley, 2014). It has an "elastic yet targeted" definition (Clark, 2005: 258) due to the different forms it takes and its association with varying political and social contexts (Shin et al., 2016). Examples of this variation include super gentrification (Lees, 2003), green gentrification (Gould & Lewis, 2016), rural gentrification (Smith et al., 2021) and new-build-gentrification (Davidson, 2018). However, it is almost always defined as the displacement of one type of incumbent resident by one of a new, typically higher, social class (Lees et al., 2008). The population being displaced are usually working-class (Paton, 2016), ethnic minorities, or the intersection of both (Huse, 2018; Richardson et al., 2019).

Quantitative analyses of gentrification (and other neighbourhood change research) have used a similar methodological framework to geodemographic change. Change is typically measured over two fixed data points using population census data, usually a decade apart (Reibel, 2011) and an index calculated from which the degree of gentrification is determined (see Johnson et al., 2021; Chapple & Zuk, 2016). However, as with classification, index-based approaches are information reductive. The various gentrification components are reduced to a single score which may fail to identify real changes when, for example, an increase in one component of the gentrification index occurs simultaneously with a decrease in another. Additionally, as gentrification can be rapid (Glass, 1964) analyses of decennial data may fail to capture the full dynamics of the process.

The application of the data primitive approach requires measures that capture the process of interest to be defined. Many UK-based gentrification studies consider gentrification a class-based phenomenon, entrenched in hierarchical society, whereby residents of a gentrifying neighbourhood are of a higher social status than before (Lees et al., 2010). This is frequently due to the in-migration of people who are more educated, from more professional occupations, than the current often lower or working-class resident population. This specific change in demographic character and attendant increase in income, is often used in gentrification studies to quantify the gentrification process (van Ham et al., 2020). However, there are other effects: house prices increase as do other costs, local services change to reflect the preferences of the new population (Lees et al., 2010). This prices out the incumbent working-class population and also prevents the in-migration of less affluent citizens. In many cases, the displaced population include ethnic minorities, who also tend to reside in lower-income neighbourhoods (Huse, 2018). Finally, as a result gentrifying neighbourhoods experience greater residential churn (in-and out-migration) than non-gentrifying ones (Yee & Dennett, 2020).

The above suggests a specific set of multitemporal data primitives to identify gentrifying neighbourhoods composed of:

- House price.
- Professional occupation.
- Residential mobility (i.e., the proportion of households that change, as a measure of in-and out-migration or neighbourhood churn).
- Ethnic composition (proportion white or non-white).

The next section describes how these data are analysed.

Methods and Analysis

Case Study and Data

This research uses annual data for 853 Lower Super Output Areas (LSOAs) in South Yorkshire, UK. South Yorkshire is a metropolitan county in the North of England, comprising four boroughs (local authorities – a unit of local government) each with an urban centre: Doncaster, Barnsley, Rotherham, and Sheffield (Fig. 1). The study area contains a range of landscapes. The west of the county includes part of the Peak District National Park, and there are many rural ex-mining communities in the central and eastern areas. There is a mixture of land uses, including industrial and brownfield land, and agriculture as well as built-up areas consisting of urban, large cities, and rural commuting towns. As in other gentrification studies, LSOAs were used as proxies for neighbourhoods. LSOAs have consistent populations of approximately 1,500 people or 500 homes (Cockings et al., 2011). They provide a degree of homogeneity for social analyses seeking to examine neighbourhood level effects (van Ham et al., 2012), and are robust units for examining neighbourhood level processes (Reades et al., 2019).

Annual data for four primitives were collated between 2010 and 2019 (Table 1). These were reported over LSOAs except for Professional Occupation which was reported over Middle Super Output Areas (MSOAs). MSOAs are composed of an average of five LSOAs and the MSOA data were interpolated to LSOAs using an area weighted interpolation approach. All datasets were open

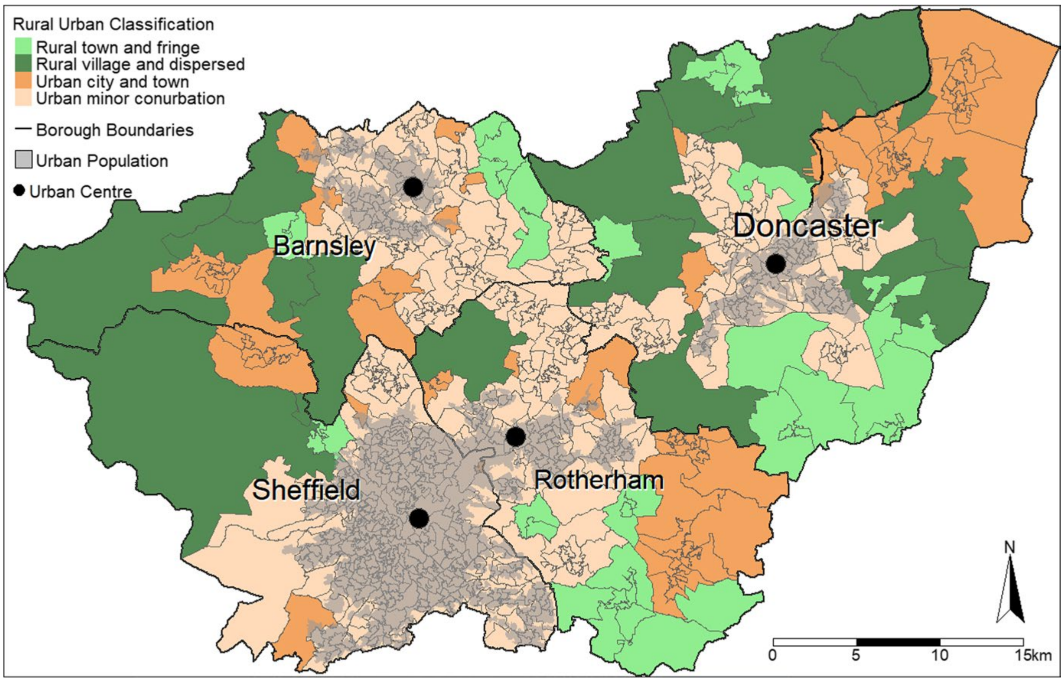


Fig. 1 A map of LSOAs in the study area shaded by the 2011 Rural Urban Classification (Bibby & Shepherd, 2004), the four boroughs (local authorities) and their urban centres

Table 1 The data primitives collated for each year, their spatial resolution and source

Data primitive	Trend	Unit	Source
House price	Increase	GB Pounds	UK government
Professional occupation	Increase	Count	UK government
Residential mobility (Churn)	Increase	Count	CDRC*
Black and Asian ethnicities	Decrease	Proportion	CDRC*

* see https://data.cdrc.ac.uk

source except the ethnicity dataset which was accessed via an application for use in this study. The data for each year were converted to percentages and transformed to z-scores with a mean of zero and a standard deviation of one.

In overview the approach taken was to evaluate LSOA change in each of the standardised primitives for 45 pairs of years, starting in 2010 and ending in 2019. The steps in this analysis were as follows:

1. The time intervals (i.e., the start and end years) were extracted where an increase or decrease (as specified in Table 1) of one standard deviation was found for all four of the primitives.
2. For gentrifying LSOAs, each time interval where gentrification was found, a gentrification score was calculated from the sum of the four absolute change values.

3. From these, the gentrification *cycle* was characterised by identifying the start and end years, the year of peak gentrification, the duration to the peak year, and the cumulative sum of the gentrification scores to the peak year.
4. Then some filtering was applied to identify *established* cycles of gentrification with the following characteristics:

   a) a minimum of two years to reach the peak year of gentrification to avoid identifying dubious neighbourhood changes.
   b) a peak gentrification score greater than one standard deviation, based on the assumption that gentrification scores below one standard deviation may not produce a clear, physical manifestation of the process on the ground (Ilic et al., 2019). This filter was also adopted by Reades et al., (2019), as standard deviations below one may represent noise within the dataset, rather than significant changes. This assumption also limits the potential impact of universal house price uplift, since only the more salient of changes are captured.
   c) A minimum cycle end date of 2014, with the assumption that gentrification can be rapid (Glass, 1964), an entire cycle is unlikely to conclude within three years.
   d) Where cycles are identified in several starting years, the sequence with the largest cumulative gentrification score to the peak year were retained, which typically coincided with an earlier start date. This was to ensure that overlaps likely to be part of the same cycle were avoided, for example cycles of 2010–2016 and 2011–2016.

Thus, the gentrification cycle conceptualised in this way captures sequences of years where gentrification increases, peaks, and then stabilises. This is perhaps best illustrated with an example. Table 2 shows the gentrification scores for one of the neighbourhoods. There are three potential gentrification cycles starting in 2010, 2011 and 2012. Of these only 2011–2016 has a score greater than one standard deviation in a sequence of increasing scores starting in 2011–2012 and ending in 2011–2019. Thus, for this neighbourhood gentrification starts in 2011, ends in 2019 (although this is the limit of the range of dates considered), peaks in 2016 and has a cumulative gentrification score of 4.801.

After this approach was applied, the data for seven LSOAs were explored using Google Earth and Google Street View to seek visual evidence of gentrification, and to determine the type of gentrification that had occurred. Finally, for each of these areas a CVA was undertaken as a tentative investigation of the extent to which CVA informs on the gentrification type. A change vector analysis generates measures of the Euclidian distance and the angle between two locations x_1 and x_2 in a multivariate feature space. Distance, D, is calculated as follows:

$$D = \sqrt{(x_1 - x_2)^2} \tag{1}$$

The angle between the points, θ, is calculated from the dot product of the vectors of x_1 and x_2 in the following way:

Table 2 An example of the gentrification scores for a single LSOA neighbourhood, for different time periods, with a score of zero indicating that gentrification was not found

	End Year								
Start Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
2010	0	0	0.729	0.918	0	1.153	0	0	0
2011	-	0.485	0.860	1.049	1.123	1.284	1.084	0.881	0.769
2012	-	-	0.375	0.564	0.638	0.799	0	0	0
2013	-	-	-	0	0	0	0	0	0
2014	-	-	-	-	0	0	0	0	0
2015	-	-	-	-	-	0	0	0	0
2016	-	-	-	-	-	-	0	0	0
2017	-	-	-	-	-	-	-	0	0
2018	-	-	-	-	-	-	-	-	0

$$\theta = \cos^{-1}\left(\frac{x_1 . x_2}{|x_1||x_2|}\right) \tag{2}$$

where $|x_1|$ and $|x_2|$ are absolute values of the vectors.

Results

The analysis was broken down into four parts: identification, temporal properties, manifestation, and validation.

Identification of Gentrification

Gentrification was identified in 123 LSOAs. Most of these were found within Sheffield (54) and Doncaster (41), with fewer in Rotherham (21) and Barnsley (7) (see Fig. 2). Of these 74% are within the Urban Population or Urban Minor Conurbation areas (see Fig. 1 for the distributions of these classes), 11% in Urban City and Towns, and the remaining 14% in Rural Areas. Taking a deeper look into each of the boroughs, Doncaster, Rotherham and Barnsley have similar spatial distributions with 58% of the gentrified areas located on the periphery of the

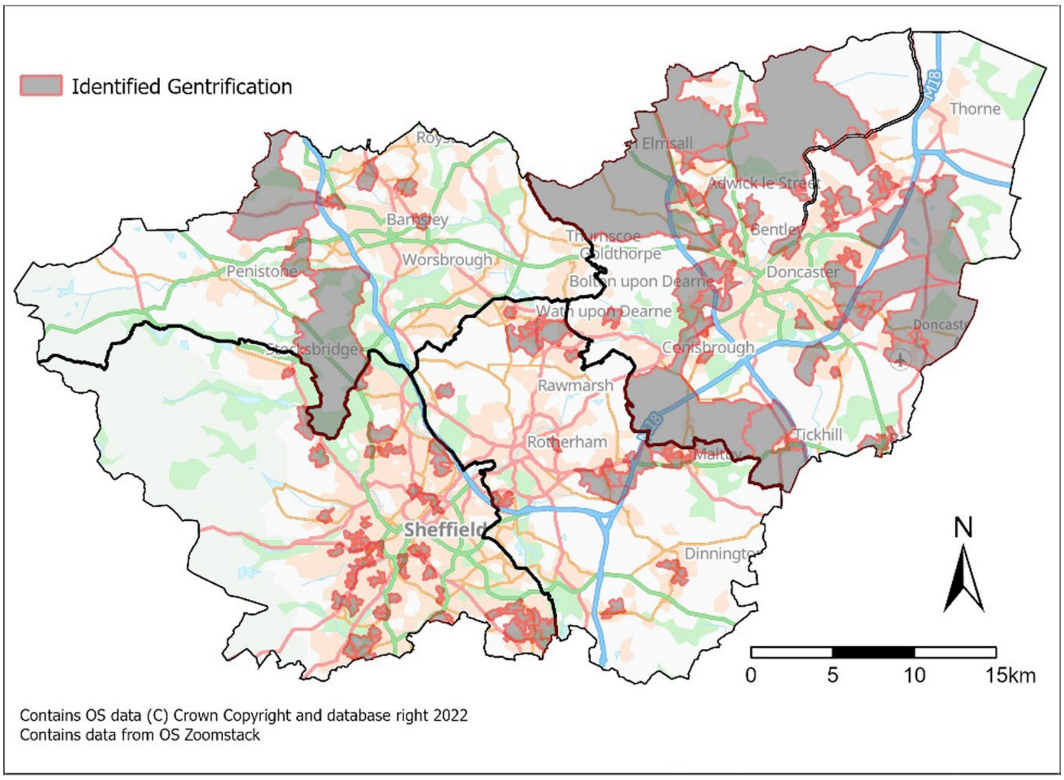


Fig. 2 The locations of neighbourhoods identified as having gentrified in the study area

urban conurbations, but close to suburban areas, urban parks or greenspaces. In contrast, 94% of the gentrification changes in Sheffield are within the main Urban Population area.

Table 3 tabulates the start, peak, and end years of the gentrification cycles. Most (64%) were found to start in 2010 and 2011, 36% between 2012 and 2016, with no start dates after 2016. There are three distinct gentrification end years, 2015, 2018, and 2019 which account for 70%, with no end years before 2014. Finally, there are distinct peak gentrification years in 2014, 2017 and 2018, each accounting for approximately 22% of the 123 areas.

The frequency of the start, end and duration of gentrification in the 123 LSOA neighbourhoods is sumamrised in Fig. 3. Gentrification was identified in 20 of the 45 time intervals and the highest frequencies were found in 2010–2015, 2010–2018, and 2011–2015. Visually two patterns stand out: the high frequencies of gentrification with short duration (5 years) and those with longer duraction (8 or 9 years), starting in 2010 or 2011.

Temporal Properties of Gentrification

The spatial distribution of the start, end, and peak years of the changes associated with gentrification, and their gentrification scores are shown in Fig. 4. It shows some differences between the boroughs in start of the gentrification cycle:

- In Doncaster, gentrification mostly starts before 2013 (83%) with many starting in 2010 and is located in the suburban towns around Doncaster.
- In Rotherham, gentrifying areas are around the edge of the borough, they start in 2010 with a cluster in 2014 to the centre and a cluster of later years to the north. The earliest years are in rural locations and later years are in more urban areas.
- In Sheffield the majority of the gentrification cycles (69%) start in 2010 and 2011 and are scattered throughout the area.
- Barnsley is different in that most gentrification cycles start after 2014.

Table 3 Counts of the start, peak, and end years of the 123 LSOA neighbourhoods identified as having gentrified

Years	Start	End	Peak
2010	44	0	0
2011	35	0	0
2012	19	0	0
2013	6	0	16
2014	6	8	28
2015	10	25	8
2016	3	13	17
2017	0	15	27
2018	0	24	27
2019	0	38	0

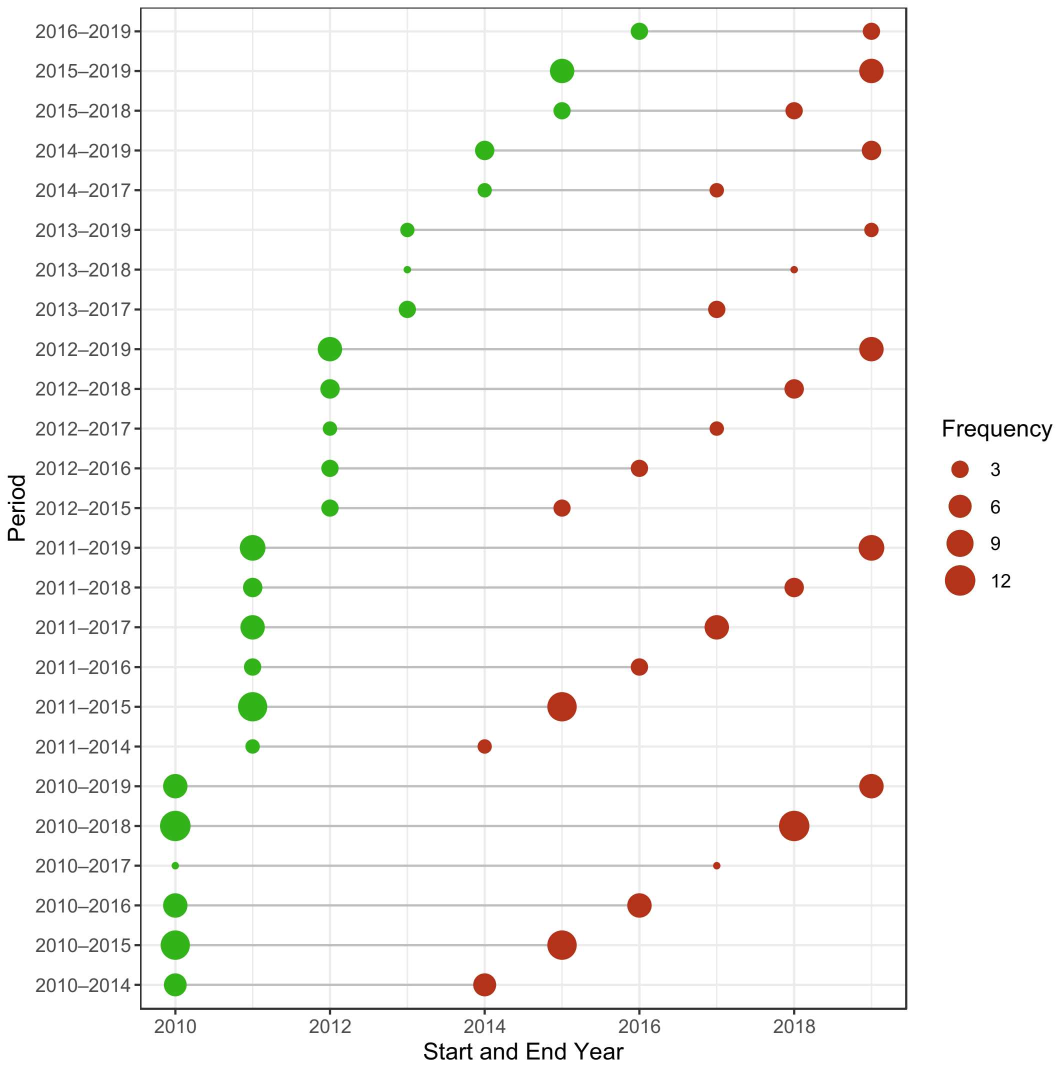


Fig. 3 The frequency of the start, end and duration of gentrification in the 123 LSOA neighbourhoods identified as having gentrified

However, a key observation throughout the study area is that gentrification is first established in one LSOA, with adjacent LSOAs following suit in subsequent years, with the exception of the south east of Sheffield. These starting neighbourhoods are frequently in suburban towns and villages, located within close proximity to transit links like motorway junctions, railway stations, and tram stops, or are associated with urban greenspaces and rural areas (see Fig. 5).

Most of the gentrification cycles have an end year of 2018 or 2019 and are found both within the urban conurbations and the surrounding towns, suggesting a long overall duration (see Manifestation of Gentrification section). Around 27% of areas have a gentrification end year of 2014 and 2015 and these are located in the edges of the borough, with the exception of those to the west of Sheffield close to the city centre. Throughout the region, end years are more clustered than

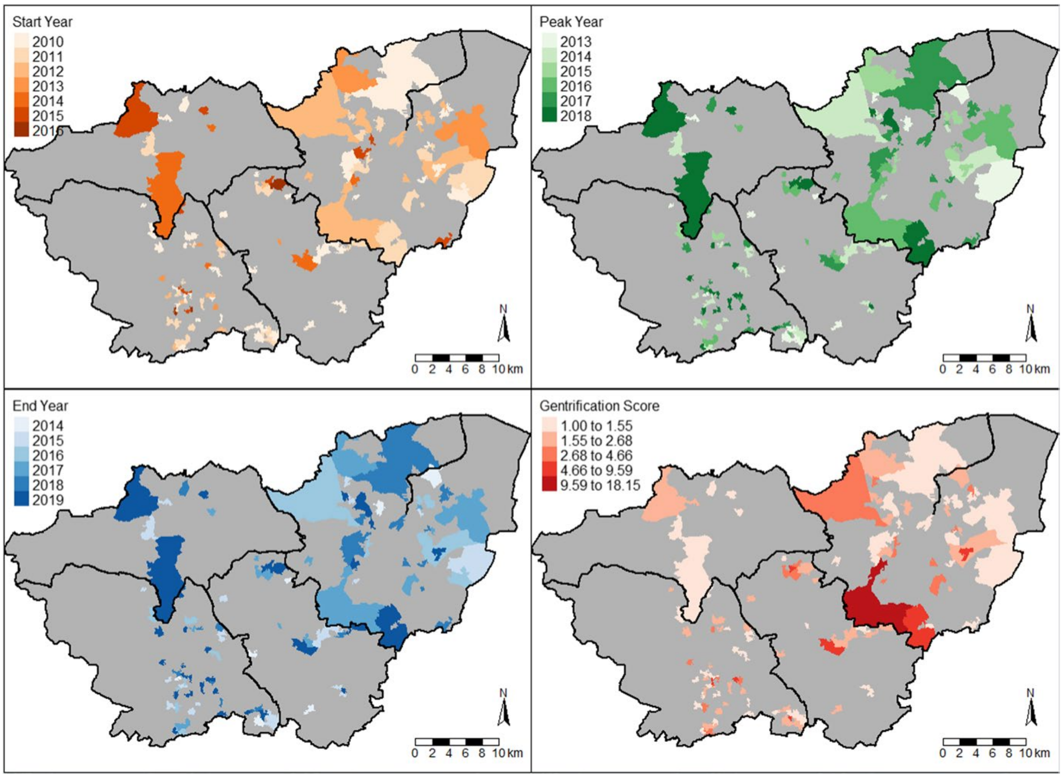


Fig. 4 The start, end, and peak year of gentrification, and the cumulative gentrification score for the LSOAs identified as having gentrified

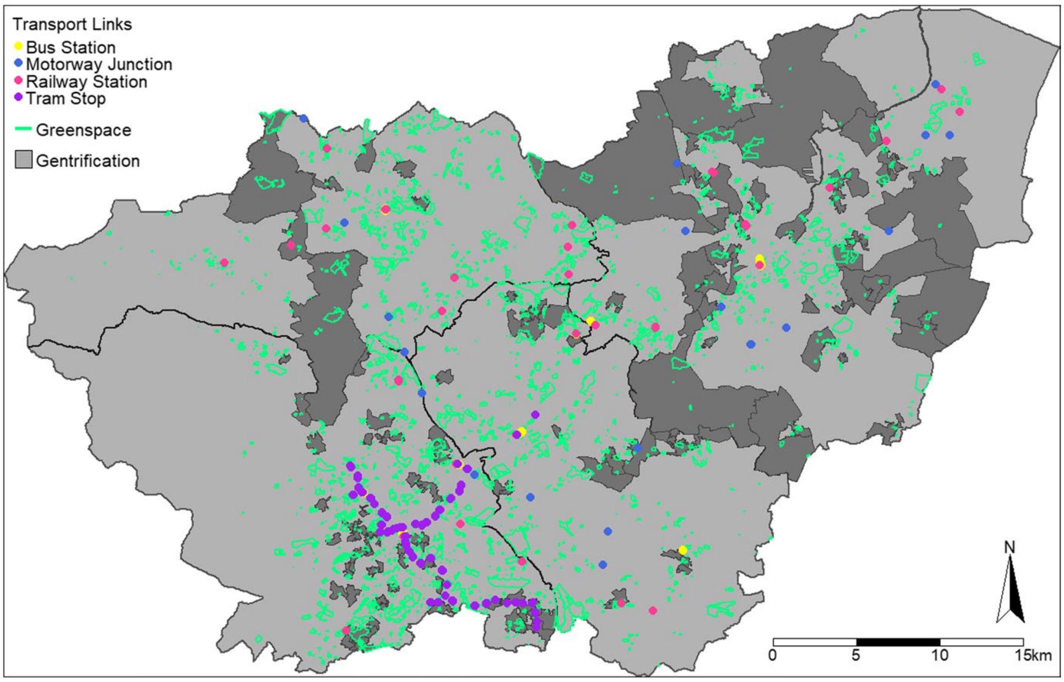


Fig. 5 Context for the case study areas: greenspaces and transit links and areas identified as having gentrified

start years, with many adjacent neighbourhoods experiencing the same end year. A similar trend is found in the gentrification peak years.

The gentrification scores capture the amount of change a gentrifying LSOA neighbourhood has experienced. The areas with the largest scores are located towards the south of Doncaster and many of the higher scores are located towards the outskirts of urban conurbations, with the exception of Sheffield. There are clusters of high gentrification scores throughout the region, with the smaller gentrification scores found in Barnsley and in rural locations.

Manifestation of Gentrification

Three metrics are used to explore the manifestation of gentrifiaction cycles: the years to peak gentrification, the years from peak to the end of the cycle, and the duration of the cycles. These are shown in Fig. 6.

Figure 6 shows that 42% of the gentrification associated areas have short periods of two or three years to the gentrification peak year in rural towns and villages, or within the urban conurbation but outside of the main urban centre. In the north of the Doncaster borough these are associated with areas described as Deprived in the 2011 Census data but are now changing, and towards the south of Doncaster in rural, more affluent areas. The areas experiencing four- or five-year periods to peak gentrification are found in less deprived areas within the urban conurbations, particularly in the west. These are rural and on the fringe of urban conurbations located within close proximity to transit links. Around a quarter (26%) of gentrifying areas have a long period to their peak of six to eight years and are associated with more deprived (rural) areas or suburban neighbourhoods with reduced access to transit links (Fig. 5).

The majority (76%) of the gentrifiying neighbourhoods reach the end of their cycle one year after their peak, with 17% in two years and 7% within three to six years. This potentially reflects the time it takes for genetrification to occur and the short 10 year date range of the data used in this study. However, longer peak to end times were found in a few areas to the southeast of Sheffield, the eastern border of Rotherham and the east of Doncaster.

The overall durations of gentrification associated changes are evently split, with 35%, 33%, and 32% for short, mid and long length durations, respectively. However, their spatial distributions vary. The longer durations (seven to nine years) are located in suburban towns and villages in Doncaster and Sheffield, and the more deprived rural areas. Shorter durations (three to four years) are found in rural areas and in the outskirts of urban conurbations. Mid-length durations (five to six years) are found in more urban areas than the short and long durations, and in deprived neighbourhoods within the urban conurbations.

Validating of Gentrification

Seven LSOA neighbourhood areas identified as having gentrified areas were chosen for an in-depth examination. These were selected to have a range of gentrification

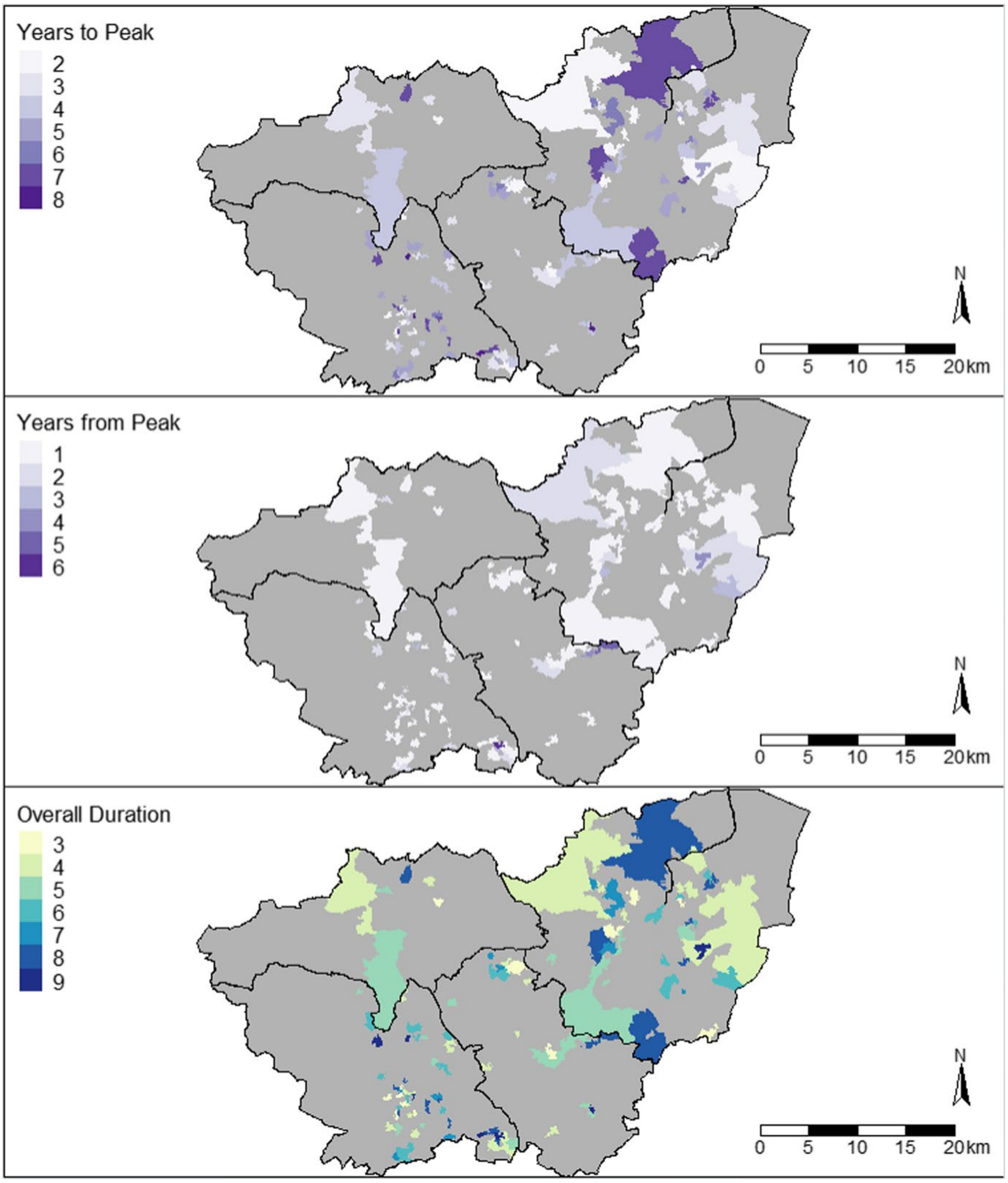


Fig. 6 The number of years from gentrification start to peak, number of years from peak to gentrification end, and gentrification duration

scores, periodicities (start, end, peak, duration) and in a range of different urban, sub-urban and rural/ village contexts. Three areas were selected in the west of Sheffield because of their contrast with the rest of the region and four were chosen randomly. The cycles of gentrification in these areas were examined using Google Street View and Google Earth as the gentrification process results in visible neighbourhood changes (Ilic et al., 2019). Descriptions and summaries of these are shown in Table 4, with descriptions added after examination.

Figure 7 shows examples of the three areas to the west of Sheffield, at the start and end of their gentrification, as close as the imagery allows. Sheffield is unique within

Table 4 Summaries of the seven LSOA neighbourhood areas selected for validation

LSOA code & location	Description	Start Year	Peak Year	End Year	Duration	Gentrification score
E01007860 Broomhall, Sheffield	Studentification in a diverse inner-city suburb, west of Sheffield city centre.	2010	2014	2016	6	5.424
E01007863 Endcliffe, Sheffield	Studentification in a wealthy suburban neighbourhood, south-west of Sheffield city centre.	2012	2014	2015	3	1.168
E01007935 Greystones, Sheffield	Studentification in a suburban neighbourhood, west of Sheffield city centre.	2010	2014	2016	6	1.933
E01007601 Branton, Doncaster	New-build gentrification in the affluent suburban village of Branton to the East of Doncaster.	2010	2015	2019	9	7.16
E01008131 Stannington, Sheffield	New-build gentrification in a suburb on the western edge Sheffield, previously an industrial brickworks.	2013	2018	2019	6	3.211
E01007704 Brinsworth, Rotherham	New-build gentrification on brownfield land, located in a village to the western border of Rotherham.	2010	2013	2014	4	1.600
E01007548 Edlington, Doncaster	New-build gentrification of an old council estate in the mining village of Edlington, to the South of Doncaster.	2010	2014	2015	5	3.676

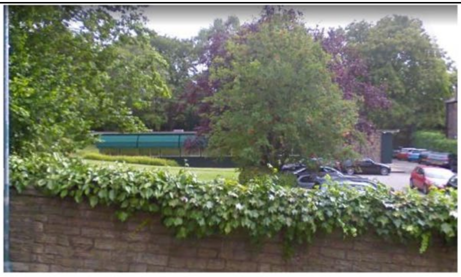
	Start Year	**End Year**
E01007860	 Gentrification started in 2011. Presence of a single story office building.	 Gentrification ended 2016, luxury student apartments completed in 2017.
E01007863	 Gentrification started in 2012. Presence of a large, detached residence.	 Gentrification ended in 2016. Change of use from residence into a bar and resturant.
E01007935	 Gentrification started in 2010 (imagery 2008). A row of terraced housing.	 Gentrification ends in 2016 (imagery from 2014). The exteriors have deteriorated, some houses for sale.

Fig. 7 Google Street View examples from the three selected neighbourhoods to the west of Sheffield City Centre at the start and end of their gentrification cycle, with their LSOA code

the study area. It is home to two universities, three campuses, and over 60,000 students (HESA, 2020), many residing within established student neighbourhoods. The first LSOA (E01007860) has Sheffield's largest gentrification score of 5.42, located within a diverse inner-city suburb and student area. A visible example of change is the demolition of single-story offices, replaced with luxury, purpose-built student accommodation. In the second area (E01007863), a large, detached residence is converted into a modern bar and restaurant, typical of changes in the neighbourhood. The third area (E01007935) shows the exterior deterioration of some houses, and the sale of others. The changes in these neighbourhoods are unique in the study area, due to the large student population, resulting in a different type of gentrification, studentification.

Google Earth was used to explore the nature of the gentrification in the four other LSOAs. In Branton, Doncaster (E01007601) and Stannington, Sheffield (E01008131) (Fig. 8) significant areas of new residential development on greenfield

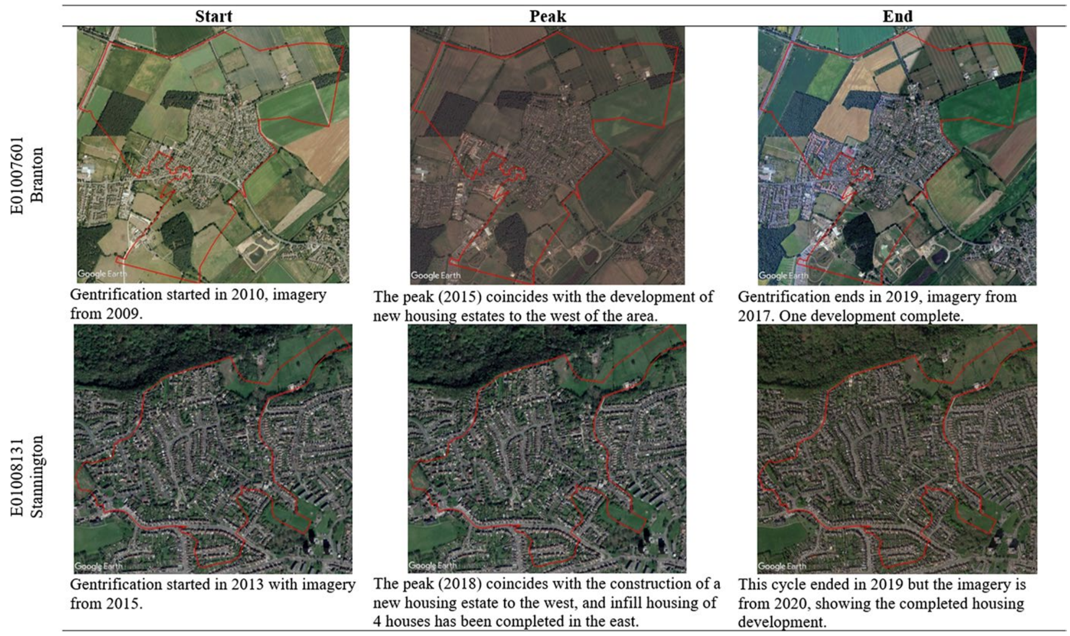


Fig. 8 Google Earth imagery of example LSOA neighbourhoods in Branton, Doncaster and Stannington, Sheffield at the start, peak, and end of their gentrification cycles, showing large residential developments

Fig. 9 Google Earth imagery of example LSOA neighbourhoods in Brinsworth, Rotherham and Edlington, Doncaster at the start, peak, and end of their gentrification cycles, showing large residential developments

sites were observed. Their peak gentrification year coincided with the completion of new housing estates in both cases. In Brinsworth, Rotherham (E01007704) and Edlington, Doncaster (E01007548) (Fig. 9), new brownfield (replacement) residential developments were found. In Brinsworth, an old industrial estate was demolished and replaced with new-build housing, whilst in Edlington a social housing estate of low-income, working-class residents, the Granby Estate, was demolished. A planning application for the demolition of 218 properties and replacement with 387 properties, 115 of which were to be affordable social housing, was approved in 2007. Due to funding problems only 64 properties were ultimately allocated to social housing (Goldsmith & Johnson, 2017). It is noteworthy that the Edlington example described a gentrification cycle of 2010–2015, but the housing development concluded in 2020. This indicates that only the first half of the cycle was correctly captured.

Figure 10 shows the changes in the standardised data primitives for further context. Here studentification (Fig. 10 bottom row) can be seen to be driven by relatively high amounts of Residential Mobility (Churn) and low changes in House Price, due to the out-migration the previous residents and the in-migration of students. The areas experiencing residential development (Fig. 10 top row) have two distinct patterns. E01008131 and E01007548 have high amounts of Residential Mobility (Churn), while E01007601 and E01007704 do not. All four areas have higher changes in House Price compared to the areas experiencing studentification, and E01007601 has a higher increase in Professional Occupations.

Finally, a CVA was explored for the seven LSOA neighbourhood areas selected for validation, with some surprising results. Figure 11 shows the angle and magnitude of change grouped by the two broad types of gentrification present in the sample areas. Initially, the angle was hypothesised to indicate the type of change processes (Gray et al., 2021). However, Fig. 11 shows that the angles for Residential Development driven gentrification differ and two of them are similar to Studentification driven gentrification. The origins of this were unpicked in the data and found to be because the angle actually indicates the driving data primitive, as illustrated in

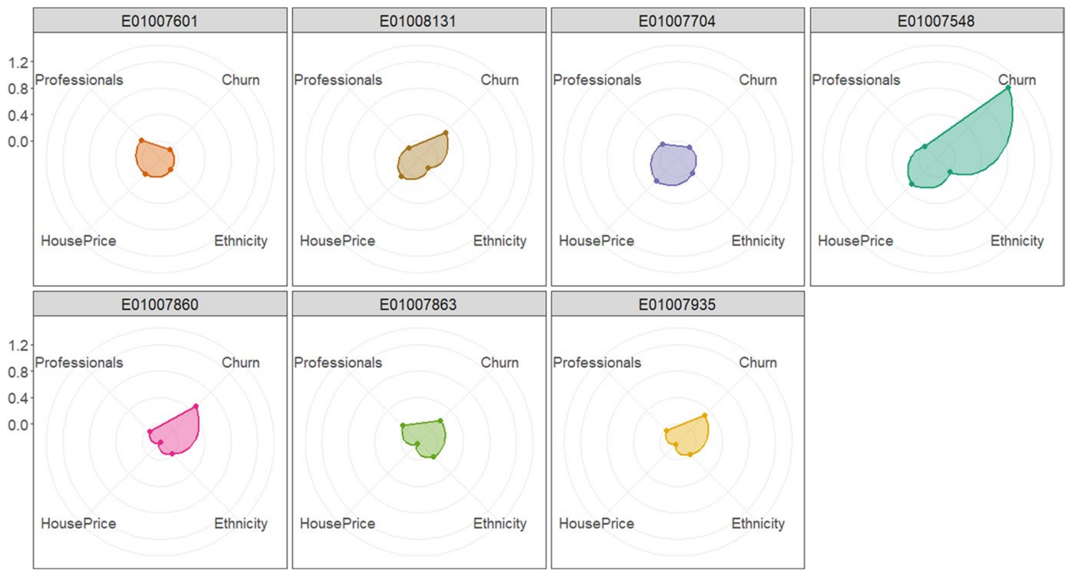


Fig. 10 The changes in standardised data primitives for each of the 7 LSOA neighbourhood areas selected for validation, with the residential development gentrification on the top row, and studentification on the bottom row

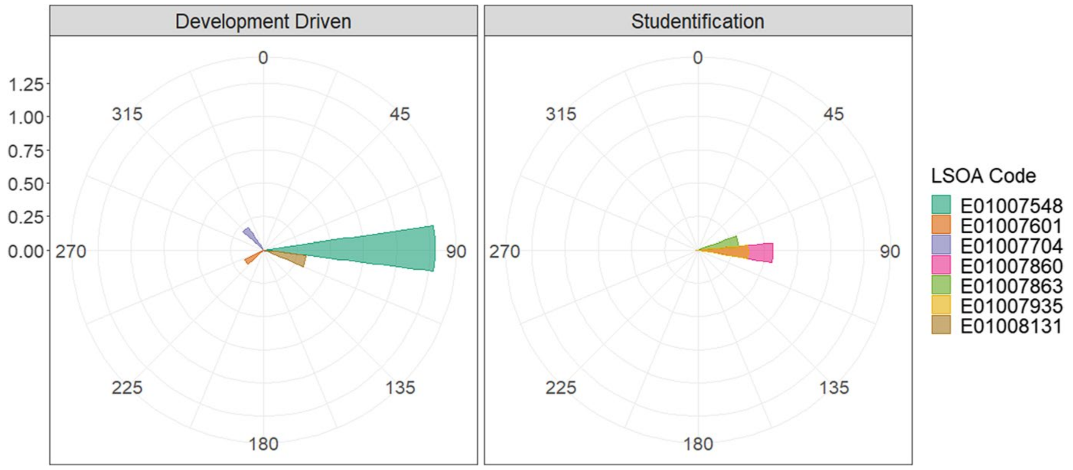


Fig. 11 The angle and magnitude of change for each case study, grouped by Residential Development and Studentification driven gentrification

Fig. 10. Here 90°±30° represents Residential Mobility (Churn), around 225° indicates Professional Occupations, and 315° represents House Price.

Discussion

The Establishment and Manifestation of Gentrification

This paper has implemented a data driven approach for exploring neighbourhood level processes, by identifying specific sets of neighbourhood area attributes, or data primitives, to describe specific processes. Here, annual data for four variables were examined in an attempt to quantify the temporal properties of gentrification processes for a small regional case study. Gentrification was conceptualised as being captured by sustained increases in House Price, the number of people in Professional Occupations, in- and out-migration (Residential Mobility or Churn), and decreases in people from Black and Asian ethnic groups. The analysis identified 123 (out of 853) areas that had experienced such changes. The properties and spatial context of these were examined and several types of gentrification were identified.

First, many areas identified as having gentrified were found to be located within the main urban population areas, close to the urban fringe (such as rural areas) or large urban greenspaces, reflecting a "green gentrification" process in which large greenspaces and parks serve as an anchor supporting gentrification (Pearsall & Eller, 2020). Within these areas, many of the early gentrifying neighbourhoods (those with early start dates), were found to be close to green or rural areas suggesting that they could be acting as a gentrification catalyst (Chen et al., 2021), with nearby areas gentrifying afterwards having more urban qualities. However, in some cases such patterns reflected new development expanding outside of the urban boundary. Such peri-urbanisation or rural areas is characterised by fragmented urban and rural characteristics (Saxena & Sharma, 2015) and is driven by urban spread, into previously undeveloped land near to urban centres (Webster & Muller, 2009).

Second, catalytic patterns were also observed near to transportation hubs, particularly railway stations in suburban towns around Doncaster, a major rail hub. Such rail-induced gentrification was also found in more rural communities, with cycles of gentrification found around rail stations (for example Silkstone Common in Barnsley and Kiveton in Rotherham). Proximity to motorway junctions, bus stations, and tram stops, are also associated with gentrification. For example, in the southeast fringe of Sheffield close to the city's tram route many gentrifying areas were found along the route, with cycles starting in the same two year period, reflecting other research on rail transit induced gentrification (Delmelle, 2021).

Third, short cycles of gentrification durations (three to four years) were found in rural and suburban areas with shorter periods to gentrification peaks (two to three years) and inevitably lower gentrification scores. These were associated with transit induced gentrification, as well as some greenification. This infers that cycles of gentrification associated with greenspaces and transit experience accelerated changes located rurally or on the outskirts of urban conurbations in suburban villages and towns. By contrast tram-induced gentrification to the southeast of Sheffield was found to have longer duration (eight years) but with both rapid and slow peaks. Thus, different types of transit-induced gentrification have different manifestations in this study area.

Fourth, mid-length gentrification durations (five to six years) were largely found within the urban conurbation or surrounding towns, especially in Doncaster. They were found in deprived neighbourhoods, but in the relatively less disadvantaged parts of the neighbourhood. Outside of these areas, mid-length durations were found to occur in rural areas with good transportation links, such as Doncaster Sheffield Airport (now closed). The mid-length gentrification areas with the greatest gentrification scores are within the wealthier neighbourhoods, such as those to the west of Sheffield, and the south of Doncaster. Thus, although mid-length gentrification occurs within urban and suburban deprived neighbourhoods, it is the relatively less disadvantaged parts of these neighbourhoods that experience uplift. These areas also experience gentrification to a lesser degree (have smaller gentrification scores), than gentrification cycles in the wealthier and least deprived neighbourhoods. This suggests that populations who are already relatively better off are benefitting from gentrification, potentially increasing inequalities, and deepening spatial polarisation (Modai-Snir & van-Ham, 2018).

Fifth, longer gentrification durations (seven to nine years) were linked with longer peaks (six to eight years) and slower changes. These areas were found in suburban and urban towns and villages, particularly in Doncaster and Sheffield and associated with more deprived neighbourhoods and ex-mining communities like Edlington, Armthorpe, and Hatfield (Doncaster); Maltby, Dinnington, and Wath-upon-Dearne (Rotherham); and Mapplewell (Barnsley). The gentrification scores associated with these areas are relatively high, but lower than the mid-length durations of gentrification.

Finally, most cycles of gentrification were found to end one (76%) or two (17%) years after their peak year, with the peak and the end years more clustered than start years. Many adjacent neighbourhoods experienced peak and the end years at the same time, suggesting that gentrification cycles have different velocities in different parts of the study area.

The Validation of Gentrification

Seven areas identified as having gentrified were examined using Google Street View and Google Earth. These areas fell into two groups: studentification and residential development driven gentrification.

Studentification is the concentration of higher education students in specific neighbourhoods in university towns (Smith, 2005). Purpose built student accommodation (PBSA) is frequently developed in close proximity to university campuses (Smith & Hubbard, 2014). The impacts of studentification include changes to commerce and services and other urban amenities (Moos et al., 2019). This is evident in Fig. 7, where large, detached houses are converted into amenities like bars and restaurants. There are other community impacts from studentification as students move into previously family houses (termed *housing of multiple occupation* - HMOs), which in England are often rows of terraced housing (Hubbard, 2009), and are subsequently not well maintained. This often leads to visible deterioration of the housing exteriors (Mosey, 2017), issues with residential parking, and impacts for other residents associated with student life. PBSA developments, with students flats and apartments, seek to overcome these issues (Hubbard, 2009): as students choose them over HMOs, the HMOs revert back to non-student occupation and are released back into the local housing stock (Stevenson & Askham, 2011). Typically, these then attract liberal intellectuals and retirees back into the neighbourhood (Bromley, 2006) due to enhanced local cultural facilities, restaurants, and other amenities, thus continuing the gentrification cycle. The gentrification identified in this study to the west of Sheffield city centre is different in this way to the gentrification identified in the rest of the study area due to the large student population and the impact of students in the local social geography (Moos et al., 2019).

Residential Development was a key driver of gentrification in many areas. Examination of the gentrification cycles start, end and peak years, as well visual investigation, showed that peaks of the process often coincided with the new residential developments. New houses are often built upon reclaimed industrial brownfield sites, or pre-existing residential land (Davidson & Lees, 2010). When this occurs no direct displacement of a population occurs, rather it is in the form of exclusionary displacement where the houses are priced such that the lower income groups are unable to access the property (Davidson & Lees, 2010). However, such developments can also occur in areas of old, large scale social housing (known as *council estates* in the UK). An example of this was found in Edlington, Doncaster (E01007548 in Table 4; Fig. 9). A large council estate was demolished and replaced with a larger, denser, development containing little affordable housing. Thus, two types of displacement were present: the initial direct displacement of working-class residents and the demolition of their properties, and the exclusionary displacement of lower income people through a very small amount of 'affordable' properties and the pricing of the remaining homes. This area has subsequently gentrified with an increase in the middle-class and a reduction in the working-class. Such gentrification, often state-led but completely or part-funded by corporate capital (Davidson & Lees, 2010), pushed the gentrification process further into and across lower-income neighbourhoods than classic gentrification would reach (Davidson & Lees, 2010). Both Edlington and Brinsworth (E01007704 in Table 4; Fig. 9) are working-class neighbourhoods which would not typically be candidates for gentrification.

The Data Primitive Approach

This research used a data primitive approach to identify 123 neighbourhoods suspected of having changed due to gentrification. This data driven approach quantified interannual changes of four selected variables over neighbourhoods represented by LSOAs. The variables and gentrification related changes are listed in "The Establishment and Manifestation of Gentrification" section above. Neighbourhoods and time periods for which significant changes were found in all four variables were further analysed to characterise the cycles of gentrification and their temporal properties (start, time to peak gentrification, end). The results were then filtered to determine established cycles of gentrification with a minimum of two years to peak gentrification, a cumulative peak gentrification score greater than one standard deviation, a minimum cycle end date of 2014 and where several cycles were found the sequence with the largest cumulative gentrification score was retained.

For the validated neighbourhoods, the end year generally coincided with the completion of large residential developments associated with gentrification. In one case the data suggested that the gentrification was complete in 2015 but the validation showed that did not occur until 2020. However, this area (Edlington) had the highest Residential Mobility (Churn) of all of the validated neighbourhoods, perhaps suggesting other changes not in Google Earth or the limitations of only 10 years of annual data. Other work had suggested that CVA angle could differentiate between types of gentrification (Gray et al., 2021) but here indicated the driver of gentrification (Figs. 10 and 11). This may be because of the small number (four) of primitives used in this study compared to the small case study introduced by Gray et al. (2021), with the result that here, the different gentrification types have overlapping characteristics: studentification is driven by Residential Mobility due to the in-and out-migration of students, as is Residential Development due to the displacement of incumbent resident, and the in-migration of the residents. Further work will explore this relationship between the number of primitives and the resolving power of CVA in order to unpack the potential of vector angle and magnitude for differentiating between different types of neighbourhood change. It may be that within cycle vectors (rather than a single overall CVA) may reveal insights about the different driving data primitives at different stages of the gentrification cycle such as displacement. Understanding these dynamics would inform the design and timing of interventions and provide valuable insights for planners.

This research has shown that the data primitive approach offers opportunities for geodemographic and related research into neighbourhood and dynamics. Synaptic (i.e., country-wide) socio-economic data over small areas are increasingly collected for a range of policy and planning purposes, for example to predict changes in education and health service demands with migratory flows. There are opportunities to refine the methods suggested here to ensure that complete cycles of gentrification are identified over longer periods, and to identify the early signals of changes in neighbourhood level processes before they manifest themselves fully. Examining data to capture emergent social processes, ones too weak to be picked up the filtering for changes greater than the one standard deviation threshold applied here, could provide early warnings of shifts in neighbourhood character and of processes that are not yet fully established, but are likely to develop into full cycles. This study identified multifaceted gentrification in a regional case study and further work will refine the choice of data primitives in order to support the more nuanced identification of different types of gentrification, as well as other types of neighbourhood process. There are also opportunities to extend this analysis from a regional study to a national study.

Finally, this research aimed to capture and analyse gentrification. It provides an indication of where neighbourhood changes associated with gentrification may occur and potential cycles of gentrification. But due to the complexity of the gentrification processes (Ilic et al., 2019) and the interconnectivity and overlap with other neighbourhood processes, similar processes may have been captured. For example, some neighbourhoods in affluent rural communities that experienced gentrification-like changes, may have experienced neighbourhood uplift. Similarly, other less affluent, more deprived urban communities with gentrification associated changes, may have experienced population churn but with in- and out-migration of populations with similar levels of socio-economic status.

Conclusion

This research uses a data driven approach to examine the spatial and temporal patterns of gentrification in the manner suggested by Gray et al. (2021). It used annual data for small areas (neighbourhoods) over a 10-year period to investigate changes associated with gentrification processes. A data primitive approach identifies the measurements that capture the full character of a process. Here, four variables encapsulating gentrification were selected and significant changes in all of these were used to infer gentrifying areas. Further analysis revealed the start, end, and peak years of gentrification. The results indicate that multifaceted gentrification was identified, including transit-induced gentrification, studentification, and also residential development driven gentrification on brownfield sites and housing stock replacement. Each gentrification type was found to be associated with specific spatial manifestations and periodicities (timings). Validation via online imagery and street views confirmed gentrification types.

The data primitive approach provides a basis for capturing the mechanics of gentrification within a multidimensional feature space. The methodology needs some refinement through the inclusion of additional variables to better distinguish between different types of gentrification and longer runs of data to capture full gentrification cycles. However, it offers a method for exploring neighbourhood level changes and provides a rich context to understanding how different processes manifest themselves in data. It overcomes the limitations of much previous research that examines change through analysis of data covering two points in time, often around a decade apart (Reibel, 2011). The nuanced results and area dynamics found within this research would not have been captured using these approaches.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jennie Gray. The first draft of the manuscript was written by Jennie Gray and all authors commented on versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was funded by the Economic and Social Research Council, grant number ES/P000401/1.

Declarations

Conflict of Interest No conflicts of interest identified.

References

Ahlqvist, O. (2004). A parameterized representation of uncertain conceptual spaces. *Transactions in GIS, 8*, 493–514.

Bibby, P., & Shepherd, J. (2004). *Developing a new classification of urban and rural areas for policy purposes–the methodology*. Defra.

Bromley, R. (2006). On and off campus: colleges and universities as local stakeholders. *Planning Practice & Research, 21*(1), 1–24.

Bovolo, F., & Bruzzone, L. (2007). A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. *IEEE Transactions on Geoscience and Remote Sensing, 45*, 218–236.

Chapple, K., & Zuk, M. (2016). Forewarned: the use of neighborhood early warning systems for gentrification and displacement. *Cityscape, 18*(3), 109–130.

Chen, Y., Xu, Z., Byrne, J., Xu, T., Wang, S., & Wu, J. (2021). Can smaller parks limit green gentrification? Insights from Hangzhou, China. *Urban Forestry & Urban Greening, 59*, p127009.

Clark, E. (2005) The order and simplicity of getrification: a political challenge, In: R. Atkinson and G. Bridge (Eds) Gentrifi cation in a global context: The New Urban Colonialism, pp. 256–264. London: Routledge

Cockings, S., Harfoot, A., Martin, D., & Hornby, D. (2011). Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales. *Environment and Planning A, 43*(10), 2399–2418.
Comber, A., & Wulder, M. (2019). *Considering spatiotemporal processes in big data analysis: insights from remote sensing of land cover and land use* (23, pp. 879–891). Wiley Online Library.
Comber, A. J. (2008). The separation of land cover from land use using data primitives. *Journal of Land Use Science, 3*(4), 215–229.
Davidson, M. (2018). New-build gentrification. *Handbook of gentrification studies*. Edward Elgar Publishing.
Davidson, M., & Lees, L. (2010). New-build gentrification: its histories, trajectories, and critical geographies. *Population Space and Place, 16*(5), 395–411.
Delmelle, E. C. (2021). Chapter six - transit-induced gentrification and displacement: the state of the debate. In R. H. M. Pereira, & G. Boisjoly (Eds.), *Advances in Transport Policy and Planning* (pp. 173–190). Academic.
Glass, R. (1964). *London: aspects of change*. MacGibbon & Kee.
Goldsmith, A., & Johnson, C. (2017). *The Granby Housing development in Edlington.* doncaster.moderngov.co.uk/. Accesed 20 Feb 2022
Gould, K., & Lewis, T. (2016). *Green gentrification: urban sustainability and the struggle for environmental justice*. Routledge.
Gray, J., Buckner, L., & Comber, A. (2021). Extending geodemographics using data primitives: a review and a methodological proposal. *ISPRS International Journal of Geo-Information, 10*(6), p386.
HESA (2020). Where do HE students study? Available online at: www.hesa.ac.uk/data-and-analysis.. Accessed 20 Feb 2022
Hubbard, P. (2009). Geographies of studentification and purpose-built student accommodation: leading separate lives? *Environment and Planning A, 41*, 1903–1923.
Huse, T. (2018). Gentrification and ethnicity. *Handbook of gentrification studies*. Edward Elgar Publishing.
Ilic, L., Sawada, M., & Zarzelli, A. (2019). Deep mapping gentrification in a large canadian city using deep learning and Google Street View. *PLoS One1, 14*(3), pe0212814.
Johnson, G. D., Checker, M., Larson, S., & Kodali, H. (2022). A small area index of gentrification, applied to New York City. *International Journal of Geographical Information Science, 36*(1), 137–157.
Lees, L. (2003). Super-gentrification: the case of Brooklyn Heights, New York City. *Urban Studies, 40*(12), 2487–2509.
Lees, L., Slater, T., & Wyly, E. (2008). *Gentrification*. Routledge.
Lees, L., Slater, T., & Wyly, E. K. (2010). *The gentrification reader.* Routledge.
Lester, T. W., & Hartley, D. A. (2014). The long term employment impacts of gentrification in the 1990s. *Regional Science and Urban Economics, 45*, 80–89.
Leventhal, B. (2016). Birds of a feather still flock together: the continuing relevance of geodemographics. *Applied Marketing Analytics, 2*(1), 52–56.
McLachlan, G., & Norman, P. (2021). Analysing socio-economic change using a time comparable geodemographic classification: England and Wales, 1991–2011. *Applied Spatial Analysis and Policy, 14*(1), 89–111.
Modai-Snir, T., & van Ham, M. (2018). Neighbourhood change and spatial polarization: the roles of increasing inequality and divergent urban development. *Cities, 82*, 108–118.
Moos, M., Revington, N., Wilkin, T., & Andrey, J. (2019). The knowledge economy city: gentrification, studentification and youthification, and their connections to universities. *Urban Studies, 56*(6), 1075–1092.
Mosey, M. (2017). Studentification: the impact on residents of an English city. GEOVERSE. Available online at: https://www.brookes.ac.uk/getmedia/52a18a39-a676-4282-b213-03057cd11d75/Studentification-MoseyM.pdf
Paton, K. (2016). *Gentrification: a working-class perspective*. Routledge.
Pearsall, H., & Eller, J. K. (2020). Locating the green space paradox: a study of gentrification and public green space accessibility in Philadelphia, Pennsylvania. *Landscape and Urban Planning, 195*, p103708.
Reades, J., De Souza, J., & Hubbard, P. (2019). Understanding urban gentrification through machine learning. *Urban Studies, 56*(5), 922–942.

Reibel, M. (2011). Classification approaches in neighborhood research: introduction and review. *Urban Geography, 32*(3), 305–316.
Reibel, M., & Regelson, M. (2011). Neighborhood racial and ethnic change: the time dimension in segregation. *Urban Geography, 32*, 360–382.
Richardson, J., Mitchell, B., & Franco, J. (2019). *Shifting neighborhoods: Gentrification and cultural displacement in American cities*. National Community Reinvestment Coalition. NCRC Research. Available at: https://www.researchgate.net/profile/Bruce-Mitchell-2/
Saxena, M., & Sharma, S. (2015). Periurban area: a review of problems and resolutions. *International Journal of Engineering Research & Technology, 4*(09), 2278 – 0181.
Shin, H. B., Lees, L., & López-Morales, E. (2016). Introduction: locating gentrification in the Global East. *Urban Studies, 53*(3), 455–470.
Singleton, A., Pavlis, M., & Longley, P. A. (2016). The stability of geodemographic cluster assignments over an intercensal period. *Journal of Geographical Systems, 18*(2), 97–123.
Smith, D. (2005). Patterns and processes of ‘studentification’ in Leeds. *The Regional Review, 12*, 14–16.
Smith, D. P., & Hubbard, P. (2014). The segregation of educated youth and dynamic geographies of studentification. *Area, 46*(1), 92–100.
Smith, D. P., Phillips, M., Culora, A., & Kinton, C. (2021). The mobilities and immobilities of rural gentrification: staying put or moving on? *Population Space and Place, 27*(7), pe2496.
Stevenson, R., & Askham, P. (2011). Purpose built student accommodation: changing face of student accommodation in Sheffield. *Sheffield Hallam University Built Environment Research Transactions, 3*(1), 6–16.
Van Ham, M., Manley, D., Bailey, N., Simpson, L., & Maclennan, D. (2012). Neighbourhood effects research: new perspectives. *Neighbourhood effects research: new perspectives* (pp. 1–21). Springer.
van Ham, M., Uesugi, M., Tammaru, T., Manley, D., & Janssen, H. (2020). Changing occupational structures and residential segregation in New York, London and Tokyo. *Nature Human Behaviour, 4*(11), 1124–1134.
Wadsworth, R., Balzter, H., Gerard, F., George, C., Comber, A., & Fisher, P. (2008). An environmental assessment of land cover and land use change in Central Siberia using quantified conceptual overlaps to reconcile inconsistent data sets. *Journal of Land Use Science, 3*(4), 251–264.
Webster, D., & Muller, L. (2009). Peri-urbanization: zones of rural-urban transition. *Human Settlement Development, 1*, 280–309.
Yee, J., & Dennett, A. (2022). Stratifying and predicting patterns of neighbourhood change and gentrification: An urban analytics approach. *Transactions of the Institute of British Geographers, 47*(3), 770–790.

Chapter 6: Predicting Gentrification in England: A Data Primitive Approach

Overview

This paper extends upon the work in Chapter 5, it starts by validating the entire 123 identified cycles of gentrification via the same manner, before assigning each LSOA to a specific gentrification type, or none, if no gentrification is observed. It then uses the data from Chapter 5, including the validated cycles of gentrification and a range of neighbourhood descriptives, to train three different algorithms (GBM, XGBoost, bagging) for creating three different predictive models. These aim to predict the presence of gentrification in England (binary response), the type of gentrification throughout England (multivariate response), and the temporal properties associated with the predicted gentrification types in England (start, peak, and end years). These predictions are explored, alongside their potential practical capabilities for local authorities and planners.

urban science

MDPI

Article

Predicting Gentrification in England: A Data Primitive Approach

Jennie Gray [1,2,3,*], Lisa Buckner [4] and Alexis Comber [2,3]

[1] Geographic Data Science Lab, University of Liverpool, Liverpool L69 3BX, UK
[2] Leeds Institute for Data Analytics, University of Leeds, Leeds LS2 9JT, UK; a.comber@leeds.ac.uk
[3] School of Geography, University of Leeds, Leeds LS2 9JT, UK
[4] School of Sociology and Social Policy, University of Leeds, Leeds LS2 9JT, UK; l.j.buckner@leeds.ac.uk
* Correspondence: jhgray@liverpool.ac.uk

Abstract: Geodemographic classifications are useful tools for segmenting populations and have many applications but are not suitable for measuring neighbourhood change over time. There is a need for an approach that uses data of a higher spatiotemporal resolution to capture the fundamental dimensions of processes driving local changes. Data primitives are measures that capture the fundamental drivers of neighbourhood processes and therefore offer a suitable route. In this article, three types of gentrification are conceptualised, and four key data primitives are applied to capture them in a case study region in Yorkshire, England. These areas are visually validated according to their temporal properties to confirm the presence of gentrification and are then assigned to a high-level gentrification type. Ensemble modelling is then used to predict the presence, type, and temporal properties of gentrification across the rest of England. The results show an alignment of the spatial extent of gentrification types with previous gentrification studies throughout the country but may have made an overprediction in London. The periodicities of (1) residential, (2) rural, and (3) transport-led gentrification also vary throughout the country, but regardless of type, gentrification in areas within close proximity to one another have differing velocities such that they peak and complete within similar times. These temporal findings offer new, more timely tools for authorities in devising schedules of interventions and for understanding the intricacies of neighbourhood change.

Keywords: data primitives; neighbourhood change; gentrification; urban geography; urban dynamics

Citation: Gray, J.; Buckner, L.; Comber, A. Predicting Gentrification in England: A Data Primitive Approach. *Urban Sci.* **2023**, *7*, 64. https://doi.org/10.3390/urbansci7020064

Received: 4 December 2022
Revised: 6 May 2023
Accepted: 10 May 2023
Published: 13 June 2023

1. Introduction

Geodemographic classifications are useful tools for segmenting areas into groups or classes based on the socio-economic characteristics of their populations and sometimes of the areas themselves. They support applications in domains that seek to understand the spatial distribution of different neighbourhood types and the people they contain [1]. Geodemographic classifications are frequently constructed from population census data, which precludes the analysis of neighbourhood dynamics [2], although they have been used to infer neighbourhood change over decadal timespans [3]. The problem with using such classifications to understand dynamics is that the processes of interest may operate over varying spatial and temporal scales [4] that may not be captured by a decennial population census. There are consequently obvious limitations to classification-based approaches to quantifying neighbourhood-level processes through class allocation with temporally coarse data and the process of class allocation [5]. These are compounded by the assumption of synchronicity between process phase and measurement frequency [6], which is likely to be unmet.

A related issue is that classification is Boolean and allocates areas to the class (statistical cluster) to which they are closest in a multivariate feature space. This limits analysis to only dramatic changes in neighbourhood composition [5] and prevents nuanced analysis of geodemographic change. For example, depending on an area's position within the feature space (i.e., near the cluster centre or edge), different magnitudes of change are required for class reallocation [6], with areas closer to the cluster edge requiring less change for reallocation than those near to the cluster centre. Additionally, within-cluster changes are not captured, even though they may indicate changes in cluster condition and quality or may be a signal of greater changes to come [7].

This paper adopts a data primitive approach to capture neighbourhood dynamics. The concept of data primitives [8] originated for land cover/land use mapping as a way of overcoming inconsistencies between different land use classifications in remote sensing [9] and has been used to link and separate land cover/land use semantics [10–12]. This paper extends the concept of data primitives into both the urban geography and temporal domains in an attempt to capture the neighbourhood process dynamics offered by data of a higher spatiotemporal resolution for a small area, thus capturing the nuances and dynamics of processes driving local changes [11].

In this study, interannual changes in four data primitives are examined to identify small areas that have been subject to gentrification, which are then manually validated. Using a national case study, three machine learning models are applied to selected annual data for small areas over a 10-year period that have been pre-processed in the same way as the training dataset. The aim is to predict the spatial distribution and timing of different types of gentrification nationally.

2. Background

2.1. Data Primitives

The absence of a dynamic element in geodemographic classification is a particular problem when dealing with change, such as occurs when an area undergoes gentrification. Conceptualising the data primitives—and the associated derived variables—as a kind of gentrification "space", this research draws on the data primitive approach to conceptualise gentrification as a change in the position of a small area within that data space over different time periods. In a neighbourhood analysis, these changes in position in a multi-variate feature space could be used to infer the changes in character experienced by a neighbourhood over time, and examining such shifts could be used to infer neighbourhood dynamics, to quantify process cycles, and to potentially predict future states [13]. This approach is, of course, dependent on the variables that are selected to identify and characterise the particular processes under investigation and the core drivers that characterise their changes. Further, the shifts in an area's position in the feature space must be filtered to determine potentially meaningful changes.

The data primitive approach is augmented with a change vector analysis (CVA) as a way to develop a clearer understanding of neighbourhood trajectories over time, as research in remote sensing change analyses have shown that the angle and magnitude of such positional changes can be used to infer the nature of the change [12]. CVA [14] originates from the remote sensing community and is used to determine land cover changes from shifts in a pixel's position in a multi-variate feature space of remote sensing image bands [15]. The magnitude of change is the Euclidean distance (length) between positions in the feature space, and the angle is the direction of the shift. Conceptually, the angle (direction) can help to discriminate between different types of change or different drivers of change [16], whilst the magnitude can be useful for comparisons within and among those change types [17].

A CVA generates measures of the Euclidian distance and the angle between two locations, x_1 and x_2, in a multivariate feature space. The distance, D, is calculated as follows:

$$D = \sqrt{(x_1 - x_2)^2} \quad (1)$$

The angle between the points, θ, is calculated from the dot product of the vectors of x_1 and x_2 in the following way:

$$\theta = \cos^{-1}\left(\frac{x_1.x_2}{|x_1||x_2|}\right) \quad (2)$$

where $|x_1|$ and $|x_2|$ are the absolute values of the vectors.

In this way, a CVA summarises a change across the full dimensionality of the data and has been found to be robust with respect to the nature and number of dimensions in the feature space [17]. In neighbourhood analyses, a CVA's magnitude and direction can be extracted and explored alongside changes in neighbourhood primitives. In this study, a CVA was applied to the single time period that most strongly indicated the presence of gentrification (see detail in the Section 3).

2.2. Gentrification

In UK-based studies, gentrification is often conceptualised as a class-based phenomenon: a product of a society rooted in a class-based hierarchy, whereby new residents of a gentrifying neighbourhood are of a higher social status than those in a prior time period [18]. It is driven by the in-migration of middle-class people who are more educated and more likely to be in professional occupations than the current (lower or working class) resident population. This increase in professional occupations is therefore often used to quantify the gentrification process [19]. There are also other effects: house prices increase, as do other costs, as a result of the changing nature of the local services reflecting the changing tastes of the new population [18]. This prices out the incumbent population while preventing the in-migration of lower- or working-class people. A further consequence of this situation is the ethnic "bleaching" of neighbourhoods as ethnic minorities, who tend to reside within lower-income neighbourhoods [20], are displaced. The consequence of this in- and out-migration is residential mobility or churn (the proportion of households that change) in gentrifying neighbourhoods, and it has recently been considered an important characteristic of gentrification [21].

While not necessarily exhaustive of the forms that gentrification might take—others [22,23] have noted super and green gentrification, for instance—these four data primitive domains, (1) professional occupation, (2) house price, (3) Black and Asian ethnicities, and (4) neighbourhood churn, should be sufficient to capture the changes associated with the fundamental drivers of gentrification in the UK.

3. Methods

To apply the data primitive approach, annual data covering these four key neighbourhood characteristics were collected, and machine learning models were trained on manually validated observations of gentrification.

3.1. Data

The data collected for Lower Super Output Areas (LSOAs) in England for the period 2010–2019 included the average house price, the proportion of people in professional occupations, the proportion of households that changed, and the proportion of the population that was Black and Asian. LSOAs are often used for neighbourhood-level analyses in the UK as they have a consistent population (~1500 people; ~500 houses) and have been found to be robust for analysing neighbourhood effects [24]. Table 1 summarises the attributes used as data primitives. These were collected from a range of open and safeguarded sources from which safeguarded data are only available via a successful application. Note that the professional occupation data are only available for Middle Super Output Areas (MSOAs), which have ~7500 people and ~2500 houses; this was spatially interpolated to LSOAs using area-weighted interpolation.

Two datasets were obtained from the Consumer Data Research Centre (CDRC) [25]. Modelled ethnicity proportions are safeguarded data for the *Black and Asian Ethnicities* data primitive, whilst the *Residential Mobility* primitive contains open data describing neighbourhood churn. Both datasets are products derived from the Linked Consumer Registers, which link the open electoral register with consumer registers supplied by value-added resellers [26]. The *Professional Occupation* data primitive was created by aggregating a selection of industries subjectively considered more "professional", as listed by the UK government. The data in the *House Price* primitive were similarly freely available from the UK government.

From these, a dataset of 60 attributes was derived for each LSOA neighbourhood observation in the following way:

1. The data primitives were rescaled using *z*-scores and for each pair of years, a change score was calculated from the sum of the absolute change in the four data primitive values (45 attributes).
2. The characteristics of potential gentrification cycles were determined by identifying the start and end years and duration, the year of peak gentrification, the start to peak and start to end durations, and the cumulative sum of the gentrification scores to the peak year. These were counted and then filtered where possible to identify established cycles of gentrification with the following characteristics: a minimum of 2 years to reach peak gentrification; a peak score >1 standard deviations, as in Reades et al. [27]; a cycle end date of 2014 or greater; and selection of the cycle with the largest cumulative gentrification score to the peak year (eight attributes).
3. From these start and end years, the change in each data primitive was determined, and the magnitude and direction from a CVA of these positions in a normalised multivariate feature space were calculated (seven attributes).
4. Finally, a set of descriptive variables was collated to aid in the separation of gentrification types. These described neighbourhood distances to transport links (railway station, tram stop, bus station, and motorway junction), the counts of the number of transport links within 1 mile, 2.5 miles, and 5 miles, the minimum distance to any transport, distances to blue space and green space, and the number of green space access points within 500 m. A neighbourhood rural/urban descriptor [28] was also extracted (15 additional attributes).

The final list of variables used can be found in the Supplementary Materials.

Table 1. The data primitives, their spatial resolution changes associated with gentrification, and the measurement unit.

Data Primitive	Resolution	Change	Unit
House Price	LSOA	Increase	GBP
Professional Occupation	MSOA	Increase	Proportion
Residential Mobility (Churn)	LSOA	Increase	Proportion
Black and Asian Ethnicities	LSOA	Decrease	Proportion

3.2. Ensemble Modelling

Ensemble learning refers to the combination of multiple models to enable a more robust prediction, often with greater predictive performance than single machine learning models [29]. Three ensemble models, the gradient boosting machine (GBM), extreme gradient boost (XGBoost) and bootstrap aggregation (or bagging) models, were trained and evaluated via their confusion matrices and sensitivity and specificity. GBM iteratively refines an initial model by examining the error within the previous model, improving upon weak learners until some accuracy or iteration threshold is reached [30]. XGBoost is like GBM but also includes regression penalties within the boosting equation, with regularization controlling overfitting and often generating better-performing models [31]. Bagging is based on the concept of model averaging; it differs from boosting by training single models in parallel, rather than iteratively, and averages them to yield more accurate predictions [32].

Several models were created to predict:

1. The presence of gentrification (binary: whether present or not, with responses of *None* or *Gentrification*);
2. The type of gentrification (with responses of *None*, *Residential*, *Rural*, and *Transport*);
3. The temporal properties associated with the predicted type of gentrification (start, peak, and end years).

The training dataset was split with a 70:30 train/test ratio using a bootstrap approach to ensure the response variable had the same distribution in the splits. Models for predicting the presence and type of gentrification were initialized with the neighbourhood characteristic variables, data primitives, change vectors, and the gentrification indicators over the 45 time periods throughout the study. The temporal properties were predicted with all the previous variables, the predicted gentrification type, and the additional temporal variables. The models were cross-validated with repeated k-fold cross validation and were hyperparameter-tuned to find the optimal parameters relevant to the specific model. Predictions were generated and evaluated against the test sample via model accuracy, kappa value, and confusion matrices. The best-performing models with respect to these metrics were chosen and then fit to the entire training set to create the final models for the prediction in England. The England dataset was created in the same way as the training dataset, using the same combination of variables. When predicting the temporal properties, the models were run as regressions and rounded to the nearest year. Prediction probabilities for the classifications (presence of gentrification; type of gentrification) were also retained, particularly for type since the characteristics of the types of gentrification can often overlap. The probabilities can provide an indication as to the likelihood that a neighbourhood will gentrify and the likelihood of the type of gentrification, highlighting confusion and where potential misclassification may occur.

3.3. Case Study and Training Data

This research is based on a case study of South Yorkshire, a metropolitan county in the north of England. It is a suitable training ground for developing a national model due to its variation in landscape, built-up areas, and subsequent mixes of land use and neighbourhood types. The west is distinguished by the Peak District National Park, and the region sits upon the Yorkshire Coalfield, which is home to many quarries, industrial areas, mines, and mining villages. There are urban and rural settlements, large cities, farming communities, and commuting towns by different modes. The case study therefore covers a range of neighbourhood types, though it is landlocked and not comprehensive in its coverage of neighbourhood types.

The training dataset consisted of 853 LSOAs. Change vectors, which were created via a function that included modified code from the rastercva function of the RStoolbox R package [33], a range of neighbourhood characteristics, and some previously calculated indicators of change. These indicators represented change in relation to each time period between 2010–2019 (every year, every two years, every three years, and so on), resulting in 45 unique time periods with indicators of change. Within the dataset, there were 123 LSOAs with an associated cycle of gentrification, all of which were visually validated via Google Earth and Google Street View [34], a method gaining in popularity (see [35–38] for example). According to a neighbourhood's data primitives, its characteristics, and visual observation, it was allocated to one of three broad gentrification types: residential, rural, or transport gentrification. Three of these 123 LSOAs were classified as none, due to a lack of visual evidence of gentrification and limited changes observed within the data; 60 were classified as residential, 20 were classified as rural, and 40 were classified as transport.

4. Results

To recap, a dataset of 79 attributes was derived, 60 of which were derived from the 4 data primitives, and 15 of which were taken from contextual features. These attributes were used to train three ensemble models for South Yorkshire, and the results were validated manually. The best-performing model was then retrained for England as a whole. Bivariate models were used to predict the presence of gentrification, multivariate models were used to predict the type of gentrification, and finally, regression models were used to predict the temporal properties of the predicted types of gentrification.

The first models were trained and fit to predict the presence of gentrification, with a binary response of gentrification or no gentrification. Table 2 shows that when fit on training data for South Yorkshire, bagging outperformed GBM and XGBoost, with accuracy and kappa values of 99.65 and 0.985, respectively. Two Type 1 errors were present, with 2 None LSOAs predicted as gentrification. This represents a sensitivity of 1 and a specificity of 0.997. The bagging model was then fit to predict gentrification in England, resulting in 4556 LSOAs, around 14% of the LSOAs in England, predicted to have experienced gentrification throughout the 2010–2019 study period.

Table 2. Model results for predicting binary gentrification in South Yorkshire.

Model	Accuracy (%)	Kappa
GBM	98.94	0.957
Tree Bagging	99.77	0.985
Linear XGBoost	99.30	0.971

Figure 1 shows that the results of the tree bagging model: neighbourhoods predicted to have gentrified are scattered throughout the country, from major cities such as London, Manchester, and Leeds to the more rural inlands between these major urban areas. See Figure 2 for a reference map of these built-up areas.

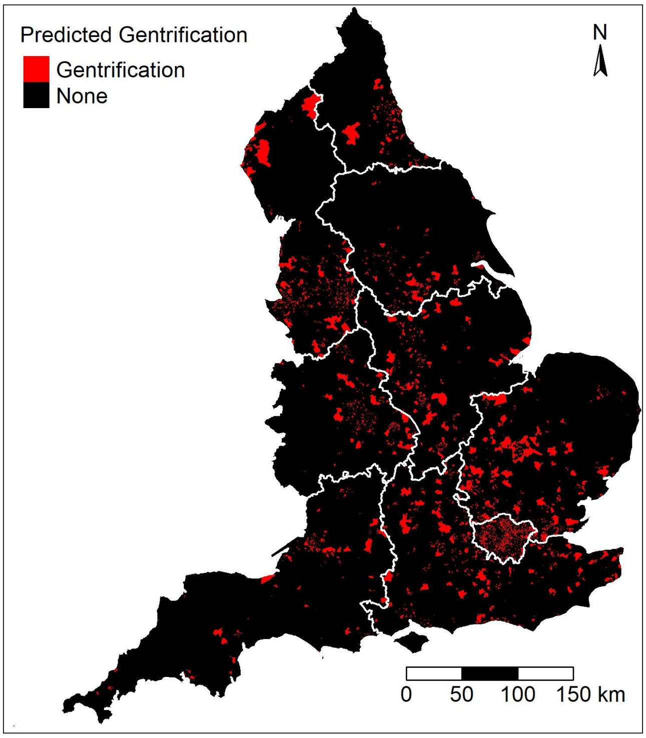


Figure 1. Probabilities for the binary prediction of the presence of gentrification in England.

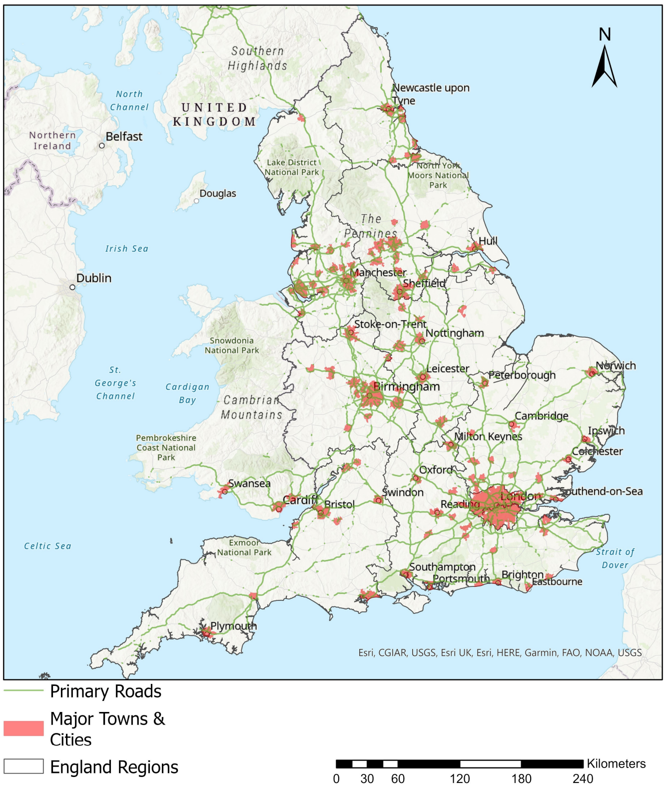


Figure 2. Reference map of England with major road networks and built-up areas.

The next models were the multivariate models, which were used to predict the type of gentrification, with responses of none, residential gentrification, rural gentrification, or transport gentrification. Table 3 shows that XGBoost outperformed GBM and bagging, with accuracy and kappa values of 98.59% and 0.945, respectively. Table 4 shows the confusion matrix, displaying the reference and predicted types of gentrification when applied to the whole of the training data. There was one misclassification for none, again a Type 1 error, which suggests that the non-gentrifying areas are sufficiently different from all types of gentrification in South Yorkshire but can confuse non-gentrifying with transport gentrification. Residential, rural, and transport gentrification all had Type 1 and Type 2 errors, with sensitivity values (true positives) of 0.95, 0.85, and 0.875, respectively. Though residential gentrification had the greatest sensitivity, it also had the most confusion and misclassification, with the lowest specificity value of 0.9917.

Table 3. Model results for predicting multivariate gentrification in South Yorkshire.

Model	Accuracy (%)	Kappa
GBM	98.01	0.922
Tree Bagging	98.48	0.941
Linear XGBoost	98.59	0.945

Figure 3 displays the probabilities of the different types of gentrification at the national level, displaying the presence of overlaps between residential and transport gentrification. Bardaka et al. [39] found that transit increases property values in neighbourhoods up to one mile from a station, which could explain some of the confusion between residential and transit gentrification. Figure 3d finally displays the overall predicted types of gentrification, a total of 4526 LSOAs, which is equivalent to 14% of the neighbourhoods in England.

Table 4. Confusion matrix of the GBM for predicting the type of gentrification on test data.

Reference Predicted	None	Residential	Rural	Transport
None	732	0	0	0
Residential	0	57	3	4
Rural	0	0	17	1
Transport	1	3	0	35

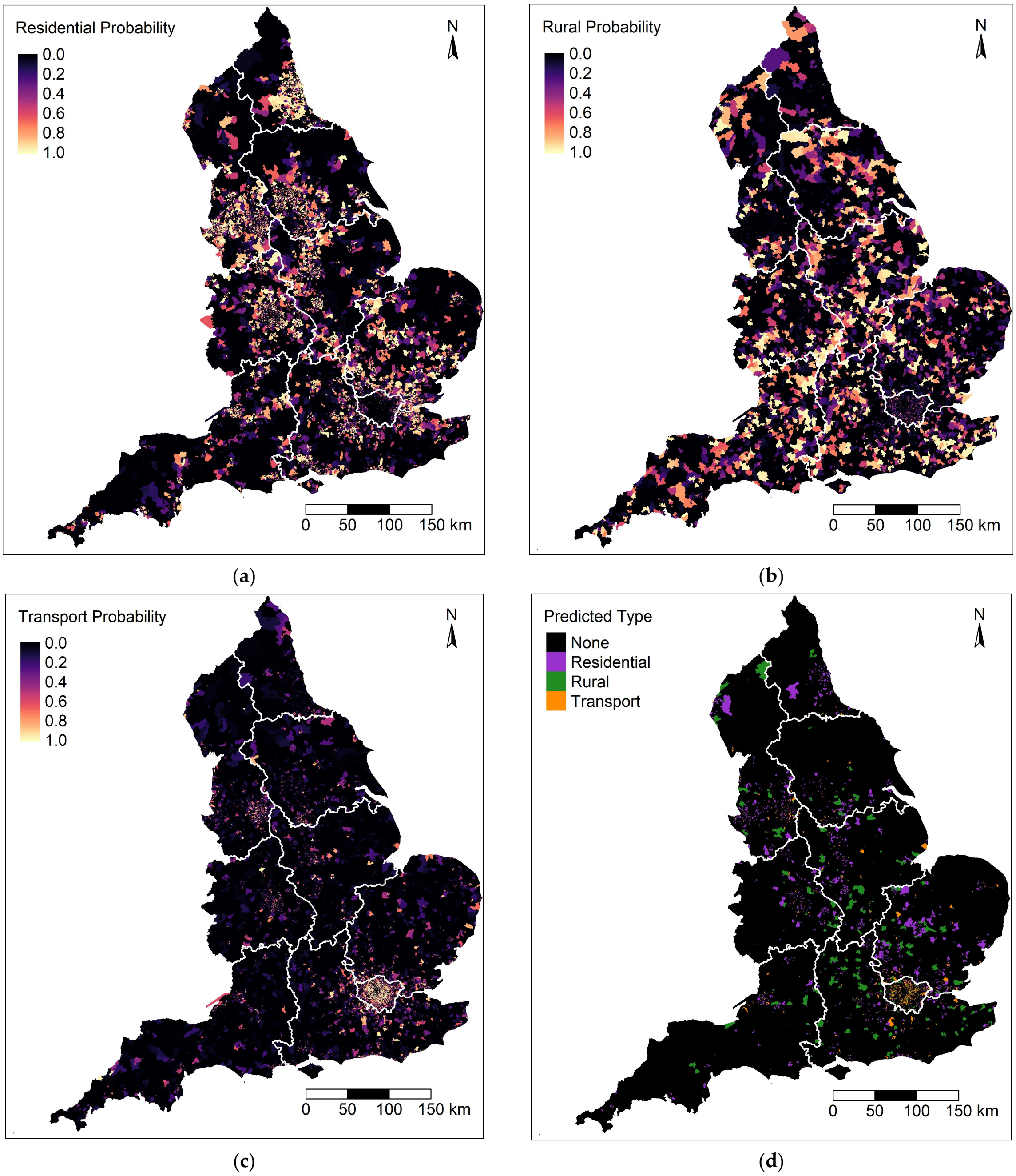


Figure 3. The probabilities of gentrification types: (**a**) residential gentrification probability; (**b**) rural gentrification probability; (**c**) transport gentrification probability; and (**d**) the overall predicted type of gentrification.

Residential gentrification (Figure 3a) was the most extensively predicted type of gentrification in England during the 2010–2019 study period and was predicted around major urban conurbations, including the outskirts of Greater London, Manchester, Newcastle, Birmingham, Nottingham, and Leeds. This supports previous research on gentrification within these cities: for example, gentrification in Newcastle was connected to development-driven (new-build) gentrification, a facet of residential gentrification [40]. State-led-replacement development-driven gentrification has also been experienced in Salford, Manchester, with negative impacts on those displaced [41].

The larger rural LSOAs distort the maps, but overall, rural gentrification (Figure 3b) is predicted with lower probabilities than residential gentrification. Rural gentrification in England between 2010–2019 occurred outside of major conurbations, often within proximity to national parks such as the North York Moors and Areas of Outstanding Natural Beauty. This highlights the pull of the rural idyl and supports previous research that explored rural gentrification in protected areas of England [42]. The residential and rural probability patterns are the inverse of one another.

Transport gentrification (Figure 3c) appears as the least likely type of gentrification and the most clustered; this is due to the densely populated LSOAs in which it was predicted. As is to be expected, transport gentrification was predicted around England's major transport hubs, such as London and Manchester. This supports previous research that found that the regeneration of a London Overground line catalysed gentrification [43]. Transport gentrification is also scattered in towns along major motorways running through the centre of England. Motorways contribute to suburbanization [44], which may facilitate gentrification in suburban neighbourhoods.

The final predicted gentrification types for England (Figure 3d) followed the highest probabilities for each gentrification type. Residential gentrification accounted for 54% (2454 LSOAs) and transport gentrification around 33% (1499 LSOAs), leaving rural gentrification with just under 13% (573 LSOAs).

The final models were run as regressions via XGBoost to predict the start, peak, and end years of the predicted gentrification cycles. These predictive models were applied to the LSOAs predicted with a gentrification type only (4526 LSOAs), opposed to the entire of England. Figure 4 shows the temporal predictions relating to the periodicity of gentrification: the start, peak, and end of gentrification in England. The gentrification start years were mostly predicted to be 2010 and 2011, but there were clusters with sequential starting years, mostly in the southern half of the country. The predicted peak years of gentrification indicate that clusters of LSOAs experiencing gentrification, regardless of their starting years, peaked at similar times, particularly in the south. Such clustering is also observed within the gentrification end years. This suggests that neighbouring localities of gentrification had varying velocities such that they peaked and completed their cycles at similar times. However, it does also show that although the model was applied to only those LSOAs that were predicted to gentrify, 141 LSOAs were consistently predicted without any temporal properties, suggesting no cycle of gentrification. However, the predicted zeros reflect areas where no temporal properties of the predicted gentrification were predicted.

The number of years taken to reach the peak of the process and the overall duration of the predicted gentrification in England were then calculated instead of being directly predicted. Table 5 shows the national averages of the duration, the number of years from the start to the peak, and the number of years from the peak to the end. Residential gentrification typically has a slower accumulation of change, taking longer to reach its peak before ending relatively swiftly, with the largest overall duration. On average, transport gentrification has similar manifestations to residential gentrification, with a more gradual accumulation of change, an accelerated peak to end, and a similar overall duration. Rural gentrification, however, has a more rapid accumulation of change, with a shorter start to peak duration before a relatively more gradual completion and a shorter average duration.

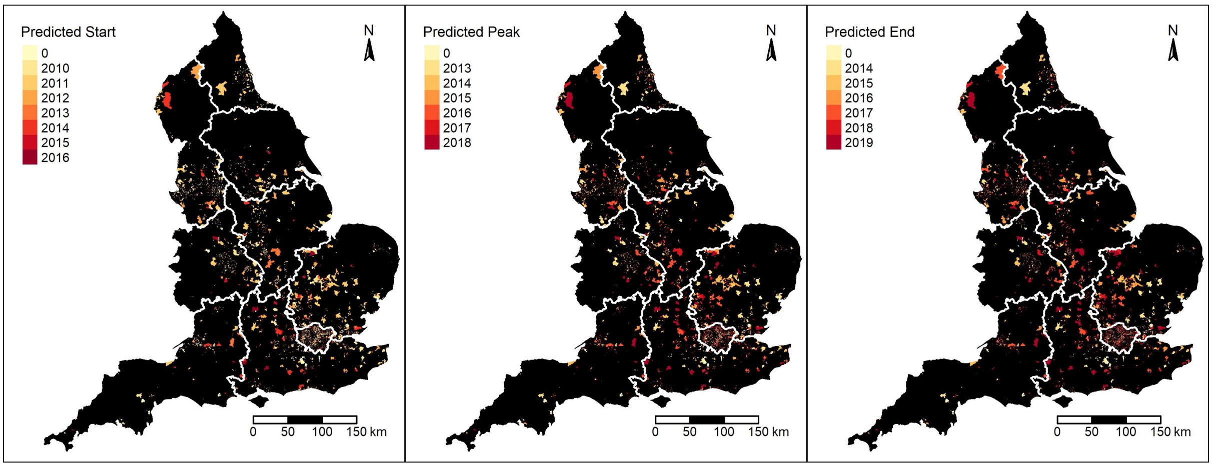


Figure 4. The predicted start, peak, and end years of gentrification in England.

Table 5. National averages of the duration, number of years from start to peak, and number of years from peak to end of the predicted gentrification in England.

Temporal Properties (Years)	Residential	Rural	Transport
Start to Peak	4.14	3.70	4.10
Peak to End	1.39	1.50	1.40
Duration	5.53	5.15	5.50

When observing these variables throughout space, there appear to be some more regional patterns, as shown in Figure 5, which demonstrate the duration of the predicted cycles of gentrification within England, faceted by region.

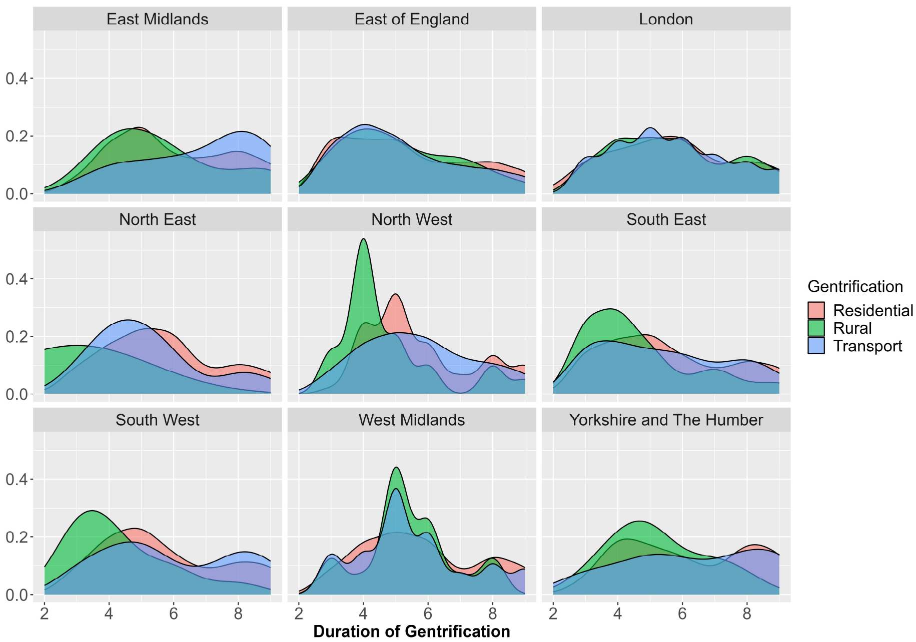


Figure 5. Density plots of the duration of predicted gentrification types by region.

The southeast and southwest had similar averages, with rural gentrification peaking at three years and residential and rural gentrification peaking at five and eight years, respectively. There is little differentiation between the different manifestations of gentrification types within the east of England and again within London, with each gentrification type having a wide range of durations. The rural gentrification of Yorkshire and the Humber was predicted to have mid-length cycles, peaking between three and four years, whilst its transport gentrification was predicted to have longer cycles of around eight years. This contrasts with the northwest, where rural gentrification had a considerable peak at four years, with transport gentrification peaking at five years and residential gentrification having a wider range of durations.

These results therefore suggest that cycles of gentrification are not consistent throughout the country, and they have regional patterns that could be explored in greater depth.

5. Discussion

This research demonstrates that the data primitive approach is a viable alternative to and advancement upon traditional approaches to analysing neighbourhood change. Gentrifying neighbourhoods, as well as different types of gentrifying neighbourhoods, can be distinguished through the use of data primitives at a resolution of years, not decades. Predictive models can distinguish between gentrifying and non-gentrifying areas with a kappa of 0.99 (99% accuracy) and between different types of gentrification with a kappa of 0.95 (98.6% accuracy). Thus, gentrifying and non-gentrifying neighbourhoods and different types of gentrifying neighbourhoods are markedly different within their neighbourhood characteristics and composition of data primitive changes over time in England.

Much of the gentrification predicted between 2010 and 2019 aligned with previous studies, such as the residential gentrification predicted in Newcastle [40] and Manchester [41], the London Overground line transport gentrification in London [43], and the rural gentrification in Areas of Outstanding Natural Beauty such as the Cotswolds [42]. When comparing the London-based results of this study to [27], there are overlaps in areas predicted as gentrifying, suggesting that some of the gentrification in London is likely to have been experienced between 2010 and 2019, a time period that they could, however, only speculate for in [27].

However, contrasting predictions were also observed for some areas (e.g., [27] predicts decline where these results predict gentrification), suggesting opportunities for further investigation: it could be that our selected training region of South Yorkshire is unsuitable for predicting changes across all of England, but it is also just as likely that the additional temporal resolution of our data yields more timely predictions than ones derived from the Census.

The confusion presented between these outputs and those within the initial misclassification on training data could suggest that further separation between the types of gentrification is needed to generate more accurate predictions. However, it could also be that the gentrification types were too broad, and that more specific types of gentrification would have provided better separation. Nevertheless, the conceptualisations provided within this study demonstrate the value of adopting a data primitive machine-learning-based approach to predicting process-associated neighbourhood change.

The binary and the multivariate predictive models generated generally consistent figures, with around 14% of neighbourhoods predicted to have experienced gentrification throughout the study period, which also aligns with the number of LSOAs identified as gentrifying in the case study region (14%).

This research also demonstrates that data primitives can predict the temporal properties of predicted gentrification, providing the power to suggest the process phase of gentrification. These results are novel to this approach, afforded by the spatiotemporal resolution of the data primitives. Results suggest that there is no singular pattern of periodicity for residential, rural, or transport gentrification throughout England. However, when observing the overall duration by gentrification type, rural gentrification has the shortest overall predictions on average and transport gentrification the longest. This could potentially be because rural neighbourhoods are less dense and require less change to make significant impacts and are thus completed more rapidly. Alternatively, their true start date may have been masked by the temporal boundary of the study, suggesting a synchronicity issue between the data and the phenomenon [6]. The length of transport gentrification could be explained by the investments that transportation brings [45] and their expanding catchments over time extending the length of the process [46].

Predictions of the peak and end of the gentrification cycle suggest that LSOAs experiencing gentrification within proximity to one another are likely to have differing velocities such that they peak and complete in similar time frames, aligning with the previous research [16]. A more in-depth exploration into the velocities of cycles via the interannual change vectors is warranted and is an interesting prospect of future work. However, presently, the greatest value of these novel process phase results is how they can be used. They offer great potential for planners and policy makers in developing a schedule of policy-based interventions, both to enhance the benefits of gentrification and to mitigate the consequences, such as displacement. This is because with a data primitive and machine learning approach, local authorities have the capability to predict whether a neighbourhood will gentrify, the type of gentrification they are likely to experience, and the process phase, and thus the sequence in which they will gentrify. This allows for the timely mitigation of consequential impacts on communities, such as by adopting community empowerment strategies to improve social cohesion in residential gentrification; enhancing tenant protections to reduce the polarisation associated with rural gentrification; and policy interventions for affordable housing to mitigate increased property prices in areas of transport gentrification around transport links [47]. Consequently, data primitives can provide local authorities with a tool for designing appropriate policy interventions at appropriate time periods to reduce the negative social, economic, and cultural impacts upon gentrifying neighbourhoods.

Limitations

There were 141 LSOAs with a predicted gentrification type (3%) that did not have any predicted temporal properties, suggesting no cycles identified and highlighting some level of confusion or misclassification between models. Thus, further explorations are required to generate more accurate predictions of the temporal properties. This could be achieved via a more explicit use of change vectors.

Neighbourhood characteristics and vectors of change were used alongside data primitives to predict three different types of gentrification in England: residential, rural, and transport. These gentrification types are not exhaustive, rather, they represent the aggregate validated types of gentrification identified in the training data region.

The visual validation of the detected gentrification in South Yorkshire and the assignment of LSOAs to a type of gentrification provided as a sound basis for the prediction of gentrification in England. However, it is an extremely time-consuming approach, and imagery is not always aligned with the years of interest [34]. Furthermore, it is also still an inherently subjective method of validation, with some difficulties in assigning LSOAs to just one type of gentrification for prediction. Nonetheless, this method validated 120 of the 123 identified LSOAs as gentrifying, representing an initial accuracy of 98% at capturing cycles of gentrification. Had the training data region been any larger, such method may not be viable without a larger team with more time and resources. Moreover, had a different region been selected, a different range of gentrification types may have been identified and consequently predicted for England via the validation.

Data primitives rely upon adequate spatiotemporal resolution data to generate dynamic insights into a process phase, but they are restricted within their temporal boundaries and are not yet capable of longer-term analyses. Thus, the universality of the approach is limited to those with suitable data representative of the fundamental drivers of neighbourhood processes. As the ubiquity of spatiotemporal data increases, some data, such as administrative data, are likely to increase in resolution and availability. However, as individuals become more aware of digital privacy, some will exercise their right of removal from the open register, which may impact the quality of products that rely on them, such as the CDRC data used within this research.

6. Conclusions and Future Work

There are several routes into areas of future work, some of which were described above. Change vectors were introduced as a component of the data primitive approach to represent an area's magnitude and direction of change in a multidimensional feature space. However, due to this paper's focus on prediction, they were not used to their full capacity: the deeper analysis of the change vectors, and their angles specifically, is a potential future area of work. Previous research has shown that the angle of change can reflect the type of change occurring [13] and consequently the drivers of gentrification [16]. Thus, a deeper analysis of interannual change vectors may generate deeper insight into the quantification of the process phase. Understanding the angles may also aid in improving the overall model precision and recall.

Finally, a more suitable predictive model may be one that explicitly considers spatiality, particularly when extending analyses to national studies. For example, the geographically weighted gradient boosting machine, which is built to improve the GBM via smoothing kernels to weight the loss function [48], may be an appropriate alternative. Nevertheless, this approach is novel in its way of generating a deeper understanding of the temporal manifestation of the different types of gentrification in England.

To conclude, neighbourhood change is dynamic and can often have a process phase that is shorter than the typical decennial intervals used in analyses, meaning that many cycles are missed. Our results show that data primitives can identify and distinguish gentrifying neighbourhoods from non-gentrifying and between different types of gentrification. Furthermore, the nature of data primitives enables the identification and prediction of the temporal properties of gentrification, providing the power to suggest the process phase of gentrification. Subsequently, such predictions can provide local authorities with the capability to schedule a timetable of appropriate policies and interventions to increase benefits and mitigate the consequences of specific types of gentrification. The distinct academic value of this approach is its ability to detect, analyse, and predict temporal properties of neighbourhood processes. More focused and specialised investigations into neighbourhood change via data primitives may therefore aid in the dissecting and understanding of the complexities of neighbourhood change.

Although the data primitive approach is in its infancy, it has started to highlight and unpack deeper understandings of the temporal properties of gentrification in England. It has created novel findings in an innovative manner, contributing both to the literature on gentrification and the neighbourhood change methodology. With further refinement, this approach has enormous potential for understanding the intricate spatiotemporal relationships between different types of neighbourhood processes and how they change throughout space and time.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/urbansci7020064/s1, Table S1: List of all variables used within the analysis, and their description. Figure S1: Interactive Map of Predicted Gentrification

Author Contributions: Conceptualization, J.G., L.B. and A.C.; methodology, J.G.; software, J.G. and A.C.; validation, J.G.; formal analysis, J.G.; investigation, J.G.; data curation, J.G.; writing—original draft preparation, J.G.; writing—review and editing, L.B. and A.C.; visualization, J.G.; supervision, L.B. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the ESRC Centre for Doctoral Training–Data Analytics and Society, grant number ES/P000401/1.

Data Availability Statement: Data are not available due to the use of safeguarded data that are integral to the study.

Acknowledgments: The data for this research were provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC 600-01, ES/L011840/1; ES/L011891/1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harris, R.; Sleight, P.; Webber, R. *Geodemographics, GIS and Neighbourhood Targeting*; John Wiley and Sons: Hoboken, NJ, USA, 2005; Volume 7.
2. Longley, P.A. Geodemographics and the practices of geographic information science. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 2227–2237. [CrossRef]
3. McLachlan, G.; Norman, P. Analysing Socio-Economic Change Using a Time Comparable Geodemographic Classification: England and Wales, 1991–2011. *Appl. Spat. Anal. Policy* **2020**, *14*, 89–111. [CrossRef]
4. An, L.; Tsou, M.-H.; Crook, S.E.S.; Chun, Y.; Spitzberg, B.; Gawron, J.M.; Gupta, D.K. Space–Time Analysis: Concepts, Quantitative Methods, and Future Directions. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 891–914. [CrossRef]
5. Reibel, M.; Regelson, M. Neighborhood Racial and Ethnic Change: The Time Dimension in Segregation. *Urban Geogr.* **2011**, *32*, 360–382. [CrossRef]
6. Comber, A.; Wulder, M. Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Trans. GIS* **2019**, *23*, 879–891. [CrossRef]
7. Zhu, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 370–384. [CrossRef]
8. Comber, A.J. The separation of land cover from land use using data primitives. *J. Land Use Sci.* **2008**, *3*, 215–229. [CrossRef]
9. Wadsworth, R.A.; Comber, A.J.; Fisher, P.F. Probabilistic Latent Semantic Analysis as a potential method for inte-grating spatial data concepts. In Proceedings of the Colloquium for Andrew U. Frank's 60th Birthday 2008, Vienna, Austria, 27 April 2008; Department of Geoinformation and Cartography: Vienna, Austria, 2008.
10. Adnan, M.; A Longley, P.; Singleton, A.D.; Brunsdon, C. Towards Real-Time Geodemographics: Clustering Algorithm Performance for Large Multidimensional Spatial Databases. *Trans. GIS* **2010**, *14*, 283–297. [CrossRef]
11. Comber, A.; Kuhn, W. Fuzzy difference and data primitives: A transparent approach for supporting different definitions of forest in the context of REDD+. *Geogr. Helv.* **2018**, *73*, 151–163. [CrossRef]
12. Xu, R.; Lin, H.; Lü, Y.; Luo, Y.; Ren, Y.; Comber, A. A Modified Change Vector Approach for Quantifying Land Cover Change. *Remote Sens.* **2018**, *10*, 1578. [CrossRef]
13. Gray, J.; Buckner, L.; Comber, A. Extending Geodemographics Using Data Primitives: A Review and a Methodological Proposal. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 386. [CrossRef]
14. Bovolo, F.; Bruzzone, L. A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 218–236. [CrossRef]
15. Lindsay, J. Change Vector Analysis. 2012. Available online: https://jblindsay.github.io/ (accessed on 25 February 2023).
16. Gray, J.; Buckner, L.; Comber, A. Identifying Neighbourhood Change Using a Data Primitive Approach: The Example of Gentrification. *Appl. Spat. Anal. Policy* **2023**, 1–25. [CrossRef]
17. Johnson, R.D.; Kasischke, E.S. Change vector analysis: A technique for the multispectral monitoring of land cover and condition. *Int. J. Remote Sens.* **1998**, *19*, 411–426. [CrossRef]
18. Lees, L.; Slater, T.; Wyly, E.K. *The Gentrification Reader*; Routledge: London, UK, 2010; Volume 1.
19. van Ham, M.; Uesugi, M.; Tammaru, T.; Manley, D.; Janssen, H. Changing occupational structures and residential segregation in New York, London and Tokyo. *Nat. Hum. Behav.* **2020**, *4*, 1124–1134. [CrossRef]
20. Huse, T. Gentrification and Ethnicity. In *Handbook of Gentrification Studies*; Edward Elgar Publishing: Cheltenham, UK, 2018; pp. 186–204.
21. Yee, J.; Dennett, A. Stratifying and predicting patterns of neighbourhood change and gentrification—An urban analytics approach. *Trans. Inst. Br. Geogr.* **2022**, *47*, 770–790. [CrossRef]
22. Lees, L. Super-gentrification: The case of Brooklyn Heights, New York City. *Urban Stud.* **2003**, *40*, 2487–2509. [CrossRef]
23. Gould, K.; Lewis, T. *Green Gentrification: Urban Sustainability and the Struggle for Environmental Justice*; Routledge: London, UK, 2016.
24. Cockings, S.; Harfoot, A.; Martin, D.; Hornby, D. Maintaining Existing Zoning Systems Using Automated Zone-Design Techniques: Methods for Creating the 2011 Census Output Geographies for England and Wales. *Environ. Plan. A Econ. Space* **2011**, *43*, 2399–2418. [CrossRef]
25. Vij, N. Introducing the Consumer Data Research Centre (CDRC). *J. Direct Data Digit. Mark. Pract.* **2016**, *17*, 232–235. [CrossRef]
26. Lansley, G.; Li, W.; Longley, P.A. Creating a linked consumer register for granular demographic analysis. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* **2019**, *182*, 1587–1605. [CrossRef]
27. Reades, J.; De Souza, J.; Hubbard, P. Understanding urban gentrification through machine learning. *Urban Stud.* **2018**, *56*, 922–942. [CrossRef]
28. Bibby, P.; Shepherd, J. *Developing a New Classification of Urban and Rural Areas for Policy Purposes—The Methodology*; Defra: London, UK, 2004.

29. Wu, H.; Levinson, D. The ensemble approach to forecasting: A review and synthesis. *Transp. Res. Part C Emerg. Technol.* **2021**, *132*, 103357. [CrossRef]
30. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
31. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016.
32. Lee, T.H.; Ullah, A.; Wang, R. Bootstrap aggregating and random forest. In *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*; Springer: Cham, Switzerland, 2020; pp. 389–429.
33. Leutner, B.; Horning, N.; Leutner, M.B. Package 'RStoolbox'. R Foundation for Statistical Computing, Version 0.1. 2017. Available online: https://cran.microsoft.com/snapshot/2017-09-17/web/packages/RStoolbox/index.html (accessed on 12 May 2023).
34. Ilic, L.; Sawada, M.; Zarzelli, A. Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. *PLoS ONE* **2019**, *14*, e0212814. [CrossRef]
35. Thackway, W.; Ng, M.; Lee, C.-L.; Pettit, C. Implementing a deep-learning model using Google street view to combine social and physical indicators of gentrification. *Comput. Environ. Urban Syst.* **2023**, *102*, 101970. [CrossRef]
36. Huang, T.; Dai, T.; Wang, Z.; Yoon, H.; Sheng, H.; Ng, A.Y.; Rajagopal, R.; Hwang, J. Detecting Neighborhood Gentrification at Scale via Street-level Visual Data. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022.
37. Ravuri, E.D. A Google Street View analysis of gentrification: A case study of one census tract in Northside, Cincinnati, USA. *Geojournal* **2021**, *87*, 3043–3063. [CrossRef]
38. Dickinson, S.T. *Exploring Green Gentrification in Established Urban Parks: A Study of Philadelphia's Neighborhood Parks*; Temple University: Philadelphia, PA, USA, 2022.
39. Bardaka, E.; Delgado, M.S.; Florax, R.J. Causal identification of transit-induced gentrification and spatial spillover effects: The case of the Denver light rail. *J. Transp. Geogr.* **2018**, *71*, 15–31. [CrossRef]
40. Cameron, S. Gentrification, housing redifferentiation and urban regeneration: 'going for growth' in Newcastle upon Tyne. *Urban Stud.* **2003**, *40*, 2367–2382. [CrossRef]
41. Hincks, S. Deprived neighbourhoods in transition: Divergent pathways of change in the Greater Manchester city-region. *Urban Stud.* **2016**, *54*, 1038–1061. [CrossRef]
42. Méténier, M. Lutte environnementale dans le parc national de Dartmoor: (re) définition d'un territoire de nature protégée par la dynamique conflictuelle. In *L'Espace Politique. Revue en Ligne de Géographie Politique et de Géopolitique*; Université de Reims Champagne-Ardenne: Reims, France, 2019.
43. Lagadic, M. Along the London Overground: Transport Improvements, Gentrification, and Symbolic Ownership along London's Trendiest Line. *City Community* **2019**, *18*, 1003–1027. [CrossRef]
44. Rocha, B.T.; Melo, P.C.; Afonso, N.; Silva, J.A. *Motorways, Urban Growth, and Suburbanisation: Evidence from Three Decades of Motorway Construction in Portugal*; Universidade de Lisboa: Lisbon, Portugal, 2021.
45. Chava, J.; Newman, P.; Tiwari, R. Gentrification in new-build and old-build transit-oriented developments: The case of Bengaluru. *Urban Res. Pract.* **2018**, *12*, 247–263. [CrossRef]
46. Lin, J.-J.; Yang, S.-H. Proximity to metro stations and commercial gentrification. *Transp. Policy* **2019**, *77*, 79–89. [CrossRef]
47. Ghaffari, L.; Klein, J.-L.; Baudin, W.A. Toward a socially acceptable gentrification: A review of strategies and practices against displacement. *Geogr. Compass* **2017**, *12*, e12355. [CrossRef]
48. Zhan, Y.; Luo, Y.; Deng, X.; Chen, H.; Grieneisen, M.L.; Shen, X.; Zhu, L.; Zhang, M. Spatiotemporal prediction of continuous daily PM2.5 concentrations across China using a spatially explicit machine learning algorithm. *Atmos. Environ.* **2017**, *155*, 129–139. [CrossRef]

Part III

Discussion and Conclusions

Chapter 7: Discussion

In this chapter, a research summary is provided in Section 7.1, containing an overview of the research questions, and the subsequent investigations conducted within the included papers of this thesis. The key findings are discussed in Section 7.2, before describing the advantages in Section 7.3. The limitations of this research are described in Section 7.4, detailing many of the problems encountered throughout the duration of this research. Finally, areas for future research are described in Section 7.5, mainly focussing around the development of the data primitive approach.

7.1 Research Summary

The overarching aim of this research was to establish data primitives as a viable alternative and extended approach for measuring neighbourhood change over time. The purpose of this was to overcome the temporal and methodological limitations associated with the current approaches to measuring neighbourhood change, as explored in Section 2.4.

This was for the purpose of analysing the spatial and temporal extent of neighbourhood change in greater spatiotemporal resolution, particularly in light of identifying, quantifying, analysing, and predicting neighbourhood processes. To achieve this, several research questions were devised, each with a respective research objective(s). In order to introduce and implement the data primitive approach with a clear focus, this research concentrated on the ubiquitous process of gentrification.

Data primitives were conceptualised and introduced to the field of urban geography via a paper in the International Journal of Geo-Information (Chapter 4), which reviews the history, developments, and limitations of geodemographic classifications specifically. Data primitives were then positioned as a methodological advancement to overcome these limitations, with an empirical case study to demonstrate their usage in analysing neighbourhood change. This paper provided the key foundations for introducing the data primitive approach with a solid framework. It provided the context and methodological structure for the empirical research in Chapter 5 and Chapter 6 to fulfil Research Questions 2, 3, and 4 (see Section 1.2.2). These research questions and respective objectives conceptualise and identify the spatial and temporal extent of gentrification (Research Question 2 and 3), before posing the investigation of predicting the spatial and temporal extent of gentrification in England (Research Questions 4).

Research was undertaken with data primitives using open and safeguarded data at the LSOA level, alongside a range of descriptive variables, for first identifying the spatial and temporal extent of gentrification in South Yorkshire. A subset of seven LSOAs identified as gentrifying, were then visually validated, identifying a range of different types of gentrification including two specific types of development driven gentrification (replacement new build gentrification and brownfield new build gentrification), transit-induced gentrification, and studentification (see Chapter 5). The spatialities and temporal properties of these types of gentrification were explored, gaining novel insights of the behaviours of gentrification, regarding their establishment and manifestation in a regional case study.

The next stage of the research included the visuall validation of the remaining 116 unvalidated cycles of gentrification. They were then categorised into aggregated gentrification types to reflect their major driving force or characteristic: residential gentrification, rural gentrification, and transport gentrification (see Appendix B to view aggregated categorisation). The purpose of this was to not only validate the previously identified gentrification, but to also provide as a validated training dataset for creating predictive models, which were trained and tuned on the training dataset and fit to predict the spatial extent of gentrification in England in binary and multivariate manners, and the temporal properties of the predicted areas of gentrification, thus reflecting the cycle of gentrification. Results indicate that there although there are average patterns for each type of gentrification nationally, that the manifestation of gentrification types vary throughout space at the regional level. Many novel insights were gained from this research, whilst others concur with current literature. The key findings of this research are detailed below.

7.2 Gentrification in South Yorkshire and England

There are several major results found throughout Chapters 5 and 6 that contribute to the fields of neighbourhood change and gentrification.These are beyond establishing a methodological advancement for neighbourhood change research. Interesting findings are grouped into subheadings regarding the establishment, manifestation, and periodicities of gentrification, the insights gained on gentrification, and finally the methodology of the approach.

7.2.1 The Establishment of Gentrification

This research focussed on a regional analysis of gentrification via data primitives, before applying machine learning upon the identified cycles of gentrification, to predict gentrification nationally. These both found interesting results regarding the establishment of gentrification, the way in which it emerges.

First, urban greenspaces and transportation links (particularly railway stations in Doncaster, South Yorkshire), were observed to serve as an anchor for gentrification, supporting the presence of green gentrification and transport gentrification. Green gentrification is a process whereby the presence of, and creation of urban greenspaces, parks, and other green infrastructure has the ability to induce gentrification (Pearsall and Eller, 2020). Likewise with transport gentrification, or transit gentrification, it is a process whereby proximity to existing, and newly developed, transit links induce gentrification (Delmelle, 2021).

These particular neighbourhood form and functions act as a catalyst of gentrification in neighbouring LSOAs. This catalytic pattern is observed throughout the case study region of South Yorkshire is also predicted throughout England. It results in clusters of neighbouring LSOAs with sequential starting years, which can appear as rings of gentrification in a town or village for example. This pattern is more apparent in the South of England, which could suggest that there are regional patterns to the sequential catalytic pattern of the establishment of gentrification. A deeper investigation could uncover the specifics of such relationships, whether it be space, gentrification type, or something else entirely.

7.2.2 The Manifestation of Gentrification

The manifestation of gentrification refers to the way in which the cycle of gentrification unfolds. For example, the velocity of neighbouring LSOAs that experience gentrification vary such that they reach the peak of their cycle (the year attributed to the most change), and complete their cycle (the attributed end year of the changes) within similar timeframes. Therefore, clusters of gentrification with sequential starting years culminate and complete their cycles at similar times. However, this research did not consider whether the identified cycles of gentrification in Chapter 5 were state-led (initiated by the government with public funding to promote urban renewal and prevent decline (Shmaryahu-Yeshurun and Ben-Porat, 2021), private-led (initiated by private investors, from developers to corporate landlords, and the transnational wealthy elites for profit (Aalbers, 2019)), or at the intersection of both (Zuk et al., 2018). Thus, it would be interesting to see whether this sequential pattern and velocity of gentrification similarly manifests in cycles of gentrification with these different initiators.

Additionally, the catalytic areas supporting greenification and transport gentrification in South Yorkshire, appear to be much faster in their peak and overall durations than other types of identified gentrification. This suggests that cycles of gentrification that catalyse gentrification in nearby areas experience a more intense manifestation of gentrification than the subsequential areas. It could also hints to relationships between the type of gentrification and their velocity. Such findings warranting further exploration of the relationship between gentrification type and velocity throughout space and time. This observation however may also be indicative of a significant redevelopment project that has brought about transformations in a particular area, which has potentially resulted in spillover effects in the adjacent neighbourhoods without them having experienced substantial new construction projects. See Bardaka et al. (2018) for an example of the spillover effects of transit-induced gentrification.

Notably however, the majority of validated cycles of gentrification, regardless of their gentrification type, were found to end one or two years after their peak year. This suggests that all cycles, short or long, rural or transport, fast or slow, wrap up with similar pace. However, those LSOAs with cycles of gentrification that had a longer peak to end duration may be of particular interest, in exploring the disparate neighbourhood characteristics.

7.2.3 The Periodicities of Gentrification

The periodicity of gentrification relates to the entire cycles of the process, the time period in which they are active. Results indicate that there are different periodic patterns according to the spatial scope of analysis. For example, results from Chapter 5 undertaken at a regional scope, show that the residential types of gentrification in urban locations have longer durations than the more rural and transport based types of gentrification, ranging from mid-length (5-6 years) to long-length (7-9 years) durations. When observing the duration of predicted gentrification types in England, residential gentrification also has the longest average duration, of 5.53 years, followed closely by transport gentrification with 5.5 years, and then rural gentrification at 5.15 years. Thus, these general patterns in a specific case study region in Chapter 5 and the patterns of the predicted gentrification in England in Chapter 6 align. However, results from Chapter 6 also show that there are regional differences throughout the country, that could benefit from further exploration.

7.2.4 The Insights of Gentrification

Some key findings relate more specifically to the characteristics of the process itself. For example, residential gentrification in South Yorkshire is polarised such that the relatively less disadvantaged areas of a neighbourhood experiencing gentrification, experience a greater uplift than those more disadvantaged areas, potentially leading to heightened spatial polarisation. Further, these cycles of gentrification in the more disadvantaged areas appear to have slower changes (6-8 years to peak), with longer overall durations (7-9 years) and moderate levels of change. These patterns were particularly evident in former coal-mining communities, which are associated with long-term economic decline, political disengagement, and low levels of trust (Abreu and Jones, 2021). Thus, gentrification in such neighbourhoods must be balanced such that sufficient improvements occur and provide benefit the original community, and limit consequences like displacement, and increasing social inequalities, and socio-spatial polarisation (Caragliu et al., 2011).

These slower, more moderate levels of change could potentially be associated with marginal gentrification, whereby gentrification is led by those less privileged than the typical, more privileged gentrifiers (Mendes, 2013). Thus, links between marginal gentrification, former mining communities, and socio-spatial polarisation could be investigated.

7.2.5 The Methodology

Beyond the development and implementation of the data primitive approach, the methods used offer some novel findings. CVA is an integral, yet currently under-explored component of the data primitive approach. Nevertheless, important results show that the angle of change, depending upon its implementation, can reflect the type of change occurring (see Chapter 4) and also the data primitive(s) driving the gentrification associated changes (see Chapter 5). This alludes to the rich information that the change vectors hold, and their ability to capture and predict dynamic neighbourhood change.

Although the validation of the predicted gentrification in Chapter 6 is currently not scalable to England (at least in the manner in which validation was conducted within the paper, via Google Earth and Street View), the general results of the predictions show an alignment to that of previous research. The predictive model for gentrification type predicted the presence of residential gentrification in major cities. Residential gentrification within the confines of the study encapsulated several more specific types of residential gentrification including replacement new-build (or development driven) gentrification (see Appendix B for gentrification assignments). For example, previous research has explored new-build gentrification in Newcastle (Cameron, 2003), state-led replacement new-build gentrification (Hincks, 2015).

7.2.6 Other Neighbourhood Processes too?

Many insights can seemingly be gained from such gentrification studies. However, gentrification is complex and multi-faceted, and in reality, there may be more than one type of gentrification active within an identified cycle of gentrification. This is highlighted in the confusion displayed in the models in Chapter 6 (the limitations of which are explored in greater detail in Section 7.4.1), and also the relative difficulty in determining the gentrification type in Appendix B. Additionally, the irony is not lost in that the data primitive approach was conceptualised and developed to overcome issues associated with the traditional methods of analysing neighbourhood change, including classification, whilst also assigning LSOAs experiencing gentrification, to one gentrification type only. However, the data primitive approach still affords detailed information regarding the cycle of gentrification.

Consequently, there is great utility in using the data primitive approach to explore specific types of gentrification in greater detail, and also different neighbourhood processes altogether.

A notable feature of the data primitive approach, is the way in which it is operationalised. Throughout this thesis, data primitives were conceptualised and specified for identifying, quantifying, and predicting gentrification. However, the design of data primitives for neighbourhood change research, is such that they can be specified to analyse *any* sufficiently conceptualised neighbourhood process.

Chapter 4 proposes a list of potential data primitives for different neighbourhood processes, highlighting their breadth of use, though these are not exhaustive nor resolute, rather suggestions. For example, urban decay is a process by which a neighbourhood experiences decline and deterioration, and transitions into a state of disrepair. It features increased poverty, fragmented families, low overall living standards and quality of life, the loss of services and the physical decay of buildings (Fertner et al., 2015; Andersen, 2019). Subsequently, urban decay could be identified with data primitives such as an increase in unemployment, poor health, income inequality, and a decrease in low-skilled occupation and house price (Gray et al., 2021). Another example is suburbanisation, a process of urban to suburban migration, which can result in suburban sprawl, where low-density peripheral urban areas grow, as households and businesses move out of urban centres (Champion, 2001). Such data primitives may therefore be a decrease in population density, an increase in business vacancy rates, and the out-flux of populations (Gray et al., 2021). However, these lists of data primitives are only proposed, they do not reflect comprehensive nor definitive combinations of the data primitives of the specific processes. The neighbourhood processes would still require careful conceptualisation prior to data primitive selection, according to the context of the city or region of analysis, to select the fundamental characteristics suitable to the situational characteristics of the study location.

7.3 Advantages of the Data Primitive Approach

Beyond the methodological improvements and the novel insights that data primitives can generate for neighbourhood change and gentrification, they have additional advantages.

First, data primitives are a flexible way of representing data, which allows for the development of customized metrics for measuring neighbourhood change. This approach can be adapted to different contexts, like the different neighbourhood processes as suggested in Section 7.2.6 and can be used to explore a range of research questions. Similarly, the data primitive approach enables the integration of diverse data sources, which can provide a more comprehensive picture of neighbourhood change over time. Although this research utilised only administrative data and safeguarded data from commercial sources, this approach can incorporate data from a range of other sources, including remote sensing data, and social media data, among others.

Second, the use of data primitives allows for greater transparency in the analysis of neighbourhood change. The approach makes it easier to identify the data sources and methods used in the analysis, which can increase the reproducibility of results and facilitate the comparison of findings across different studies. However, this may not always be the case, as this research is not fully reproducible due to the use of third party safeguarded data, which can only be accessed after a successful application.

Third, the granular spatiotemporal resolution and flexible nature of the data primitive approach makes it well-suited for informing policy decisions. The approach can help identify areas experiencing significant neighbourhood change, including gentrification, and can inform the development of targeted policies to address the impacts of such change on local communities, enabling the development of a schedule of appropriate interventions.

Overall, the data primitive approach has several benefits that make it a valuable tool for analyzing neighbourhood change over time. Its flexibility, spatiotemporal resolution, and data integration capabilities can provide a more comprehensive understanding of neighbourhood change, while its transparency and policy relevance make it well-suited for informing policy decisions.

7.4 Research Limitations

Although this research has generated significant insights for the development of neighbourhood change research and gentrification studies, there are limitations in relation to the technical specifications and application. This section explores some of these limitations, which would have been addressed, given the time.

7.4.1 Confusion and Misclassification

When using validated cycles of gentrification in South Yorkshire as the basis for training a predictive model, the in-test results displayed some confusion between types of gentrification, highlighted by several misclassifications. These were largely in respect to the sensitivity of residential gentrification, indicating the model's decreased ability to correctly identify residential gentrification. Confusion can arise due to a number of reasons, such as the differing methodologies and approaches to analysis that can yield conflicting results (Finio, 2022). Such conflicting results may lead to difficulties in the initial conceptualisation of the process, potentially not being explicit enough (Shin et al., 2016).

Neighbourhood processes are intricate phenomenons, resulting from the different spatial, social, institutional, and political landscapes in which they exist (Shin et al., 2016). The complexities of these processes and their diverse manifestations can result in differing academic insights and conflicting literature. This can be troublesome when defining and conceptualising the process under investigation prior to analysis, particularly when adequate identification is necessary (Cole et al., 2021). Although great consideration was made in the conceptualisation of gentrification for this research in Chapter 4, such complexities may have inadvertently influenced conceptualisation, and also led to the potential miscategorisation of validated gentrification in South Yorkshire (see Appendix B and C for all 123 validated cycles of gentrification and their subsequent categorisation). If this is the case, then such miscategorisation may have had implications on the prediction of gentrification, particularly so as they remain unvalidated.

However, confusion within this study's results could suggest that the conceptualisations of the gentrification types are okay, but further separation between the gentrification types are needed, and is explored in Section 7.5.2. Nevertheless, this limitation also returns back to an advantage of the data primitive approach. This is because the integral idea and theory behind data primitives (see Chapter 4), is that they are, and capture the most fundamental domains and characteristics of a particular process of change (Comber, 2008; Gray et al., 2021). With only four core data primitives, a number of different types of gentrification have been identified (see Appendix B), and predicted. This shows that regardless of the type of gentrification, so long as the fundamental components of the process are captured, the core, conceptualised, gentrification process will be identified. This outcome could go some way to validating the capability and efficiency of data primitives for analysing neighbourhood change via neighbourhood processes.

It is important to note, however, that since gentrification is a complex and multifaceted process influenced by a range of social, economic, and political factors, that any predictions about gentrification in England based on a single regional study should be considered tentative and would benefit from further research and analysis.

7.4.2 Validation

The validation of South Yorkshire's 123 identified cycles of gentrification was undertaken via Google Earth and Google Street View. This visual validation ensured a sound base for training a model for predicting gentrification in England. Validation via visual means is gaining in popularity (Hwang and Sampson, 2014; Ilic et al., 2019; Velastegui et al., 2019), with Bratchford (2020) appealing for the normalisation of aerial imagery within sociological research. However, it is an extremely time-consuming approach, and the imagery is not always suitably aligned for the years of interest (Ilic et al., 2019). For example, some neighbourhoods did not have Google Earth imagery temporally aligned with either their start, peak, or end years, or a combination of, and thus required visual validation for a second time via Google Street View. However, not every neighbourhood was completely 100% visually validated, but enough changes were observed make conclusions. Nonetheless, this subjective method validated 120 of the 123 identified LSOAs as gentrifying, representing an initial accuracy of 97% of the data primitive method at capturing cycles of gentrification. This shows that data primitives perform well at a regional level, evidenced through their accuracy and kappa values. However, although this method validated the gentrification in South Yorkshire, the results of the predictive models remain unvalidated, which is a considerable limitation of the data primitive approach as yet. This is because currently there is nothing to validate it against, no national gentrification dataset exists in which to validate these results with. However, there are other methods that could be used.

A potential route for the visual validation of these identified cycles of gentrification is via Volunteered Geographic Information (VGI). VGI is the outcome of the phenomenon of widespread engagement of private citizens, often with formal qualifications, in creating geographic information (Goodchild, 2007). The scientific research in which this can result in when contributing to a researcher-led project, is often referred to as citizen science (Gura, 2013). Citizen science is increasingly used in neighbourhood health studies, for example Barrie et al. (2019) trained senior citizens to audit public greenspaces, to explore the quality of neighbourhood greenspaces and their influences on ageing well. Citizen science offers a means of undertaking substantial, and laborious work in a more achievable and efficient manner (Gura, 2013). In order to achieve this for validating the data primitive approach, a website could be created with an interactive map of the identified cycles of gentrification, alongside their temporal properties. Users could then provide feedback on whether the neighbourhood has experienced changes associated with gentrification or not. If so, they would be asked further questions like, 1) if so, in what way, 2) when these changes became visually observable (likely around the time of the cycle peak), 3) when the neighbourhood change was complete, and 4) what they believe the main driver of the changes are, or what the gentrification is associated with. This would add rich information to these identified cycles of gentrification, and indeed, the predicted cycles of gentrification could be explored in the same way. This would therefore aid with the validation of both methods and help to externally validate the data primitive approach.

7.4.3 Temporal Boundaries and Scope

A limitation of the overall approach is that the temporal boundaries are constrained to the availability of adequate (annual - small-area) data, which for the conceptualised gentrification data primitives, was 2010-2019. This is because the data primitive approach relies on the availability of high-quality data sources, which can be challenging to obtain. This can also impact the universality of the implementation of the data primitive approach. In some cases, data may not be available for certain time periods, or at all depending upon the spatial scope of the study, which can limit the ability to accurately measure change over time. This consequently limits the focus of analysis to short-term analyses of neighbourhood change only. Although this approach offers detailed insights of the establishment and manifestation of neighbourhood processes and is novel at capturing neighbourhood dynamics on the short-term scale, which has previously been advised (Reibel and Regelson, 2007; Hincks, 2015; Barton, 2016; Hincks, 2017), it does not yet have the capability for longer-term analyses. Nevertheless, certain types of data, often relied upon in research of this type, will improve in their spatial and temporal resolution over time, increasing the availability of appropriate data. This will subsequently enable future longer-term neighbourhood change analyses with high spatiotemporal resolution. Such analyses will continue to expand the understanding of gentrification and its impact on neighbourhood change, throughout space and time.

7.4.4 In an Ideal World

This research project started with an external partner and a prior data sharing agreement. However, due to major internal changes with the external partner, this partnership never materialised, and neither did the data. This research subsequently required a major overhaul halfway through. Consequently, this research instead relied upon a range of open data, and some more safeguarded datasets from the Consumer Data Research Centre (CDRC). Subsequently, the temporal boundaries and the scope of the research were both constrained to the availability of the open and safeguarded data at the required spatiotemporal resolution. In an ideal world, the temporal boundaries of this research would have been a minimum of 20 years, with data going back to 2000. This would have enabled longer-term analyses to capture cycles of gentrification that have durations greater than 10 years, or those cycles that were not synchronised with the temporal boundaries (Comber and Wulder, 2019). However, Lower Super Output Areas were only introduced in 2004 for the 2001 U.K. censuses. Prior to this, small-areas were considerably larger, and administrative data was not routinely updated with sufficient temporal resolution, thus restricting data availability.

Ideally, data primitives would also be able to capture more than the sociodemographic foundational driving characteristics of neighbourhood processes. They would also represent more explicitly the natural and built environment because research has shown that the physical environment of a neighbourhood can place it at a higher risk of gentrification (Cole et al., 2021). For example, the greening of neighbourhoods via the transformation of vacant land into community gardens may encourage gentrification (Maantay and Maroko, 2018). Other studies present similar findings; proximity to community gardens and other types of greenspaces are associated with an increased likelihood of gentrification (Maantay and Maroko, 2018; Anguelovski et al., 2018). Thus, data primitives that incorporate the number, size, function, and access to greenspaces within a neighbourhood would be a great addition, particularly for the identification and analysis of green gentrification. The inclusion of other types of data, like remote sensing imagery where the greenness of neighbourhoods can be analysed (Franco and Macdonald, 2018) may also provide as useful data primitives. Thus, given the suitability of the quality, spatiotemporal availability, and theme of data, they would be included as extra data primitives for identifying specific types of gentrification.

Given the overhaul and redirection of the project halfway through, and the both the personal and professional implications of the coronavirus pandemic, this research was also severely constrained by time. Given the time, this research would have also gone beyond gentrification and explored more neighbourhood processes, like some of those that were conceptualised with data primitives within Chapter 4. It would then have explored the interrelation overall between the different neighbourhood processes, and the specific types of those neighbourhood processes, with the aim of generating some understanding around the complexities of neighbourhood change. Nevertheless, this is still possible, and is a potential piece of future work.

The overall scope of the research was limited to an achievable workload given the timeframe, leading to the development of an imperfect, yet competent and effective methodological advancement for neighbourhood research analysis. In an ideal world, frameworks to overcome all of these limitations would have been implemented, finishing the research with a “perfect” approach. As such, this would have formed the basis of an R package with all required functions for its implementation in other researcher’s work, allowing for the specification of other neighbourhood processes, as described above.

7.5 Future Work

Data primitives are established as an alternative approach to neighbourhood change analyses, but future work can be undertaken to refine and improve the approach and overcome some of the limitations highlighted above.

7.5.1 Exploring the Interrelationships between Gentrification, Space, and Time

First of all however, although this research initiated the exploration of gentrification throughout space and time, the main focus and aim of this research was to propose and establish a new approach for neighbourhood change studies. In this way, much deeper analyses into the relationships between gentrification, space, and time, are warranted. For example, those key findings in Section 7.2 each could provide the basis for specific studies, like investigating the velocity of clustered gentrification and non-clustered gentrification, and their gentrification types. Further, the predicted gentrification could undergo more detailed comparative analyses to existing studies. For example, the predicted residential gentrification in Manchester could be compared with a breadth of quantitative gentrification studies of Manchester, like the works of Hincks (2015) and Bratchford (2020). This could give an indication as to the validity of the predicted gentrification, whilst also exploring residential gentrification in Manchester in greater detail, particularly in regard to their temporal properties. Based on the results in Chapter 6 other potential comparative and detailed studies include:

- Transport gentrification in London, Manchester, and in commuter towns along major motorways, exploring both public and private transportation.
- Rural gentrification in AONB like the Cotswolds, and National Parks like the Lake District, and the Yorkshire Dales.

Furthermore, although novel insights into the temporal properties of different gentrification types have been generated, a deeper and more explicit analysis into the spatial distribution of the temporal patterns of the cycles (establishment, manifestation, and periodicity) of different types of gentrification are warranted. This is because research in Chapter 6 suggests that these patterns differ according to the spatial scale of the analysis, thus their relationship throughout space is meaningful and potentially informative for policy purposes

7.5.2 Refining the Data Primitive Approach

There is potential for the data primitive approach to be improved in a number of ways. One of the first limitations of the approach was highlighted in Chapter 5 when validating seven neighbourhood's gentrification cycles. Here, a gentrification cycle was captured for Edlington, Doncaster, which started in 2010 and ended in 2015. Validation showed that by 2017 (the most recent image available on Google Earth) the process was yet to finish, with the residential development only partially complete. From local knowledge, this neighbourhood's cycle of gentrification was not complete until 2020, five years after the end of the captured cycle. Thus, only half of the true cycle was completed. The reason for this may have been that the neighbourhood experienced a break in the sequence of gentrification type changes, and any 2017 onward changes thus representing the second half of the cycle, would have been filtered out by the analysis criteria.

This highlights that although the approach does work, the captured temporal properties may not be accurate, due to the pragmatic methodological choices like the filtering requirements, and that there are improvements to be made. Such improvements could include modifications to the sequential analysis function, in order to account for a one-year gap in the neighbourhood's data primitive changes, or the loosening of the identification criteria such that for one singular year, only three of the four data primitives are required, if pre and post this year, all four data primitive changes are fulfilled. This may ensure that the entire, completed cycles of gentrification are captured, thus also increasing the accuracy of the associated temporal properties of the identified gentrification. Subsequently, this may also then improve the predictive capabilities of models in predicting the start, peak, and end of gentrification in England, In turn, this may improve the insights to be gained surrounding the periodicities of different types of gentrification. This improvement may also provide more practical for local authority planners, who can plan a more accurate timetable for the services, provisions, and interventions for their changing population, to align the intervention with the mitigation of the likely potential implications of specific types of gentrification.

Another important refinement of the approach is to improve the separation of and between different types of gentrification, by adding extra data primitives into the process identification stage via the process functions. This refinement has been mentioned previously in Section 7.4.1. Theoretically, it would improve the implementation of the data primitive approach, making them more explicit in capturing a specific type of gentrification. This is opposed to the four core gentrification data primitives capturing numerous types of gentrification, and relying upon neighbourhood characteristics for distinction between them. In theory, this increased separation between gentrification types would limit the potential for confusion and misclassification, subsequently improving the accuracy and kappa values of models for predicting the type of gentrification in England.

7.5.3 Extending the Application of Data Primitives

Methods throughout this research have included steps to filter the cycles of gentrification identified via the two process functions, in order to analyse only the *established* processes. Those that were theoretically not likely to reflect entire, or complete cycles were filtered out, as to analyse and gain insights on complete gentrification cycles only. Future work could also focus on the identification of early signals of change. This could be achieved via looking at small windows of change in the years prior to the start of the 'established' cycles of gentrification that have already been analysed. The purpose of this would be to see whether there are any small changes that might have indicated this change was to come. By investigating neighbourhood change in this way, emergent social processes may be captured.

Another potential method to capture the early signals of change is via modifying the filtering assumptions placed upon the identified gentrification. Early signals of change could be identified by instead changing the assumptions to instead seek out emergent processes opposed to established processes. For example, a minimum duration of two years, and cumulative gentrification scores under one standard deviation of change to capture small, but consecutive changes associated with gentrification. This may subsequently identify processes that are not yet fully established but have made at least two sequential changes as determined by the process requirements and may likely to continue into a full cycle of gentrification.

7.5.4 Exploring Change Vector Analysis for Predicting Neighbourhood Change

Although change vector analysis was introduced as a method and used as descriptive variables within the models for predicting gentrification, they have had limited exploration within this research. Thus, future work could explore the relationships between the CVA angles of change and different types of gentrification in greater detail. Initial results from Chapter 4 suggested that the angle of change may reflect the type of change occurring, but results from Chapter 5 suggested that the angle of change may instead reflect the main data primitive(s) driving the gentrification. They were not explored in detail in Chapter 6, but instead used as a predictor of gentrification.

In CVA, change is recorded regardless of the magnitude. Throughout this research, those with a magnitude below one standard deviation were filtered out in order to investigate the more meaningful changes. However, without such filter, and a slight modification to the process function, future work could explore the smaller changes in an area that is not currently identified as having gentrified. This could generate insights into the natural levels of fluctuation a neighbourhood can experience, without it being associated with gentrification, or any other neighbourhood process. This may lead to the discovery of the *threshold* that pushes a neighbourhood from natural fluctuation to neighbourhood process. It could also be used to explore whether thresholds vary for different data primitives and how long these thresholds take to be satisfied.

Deeper analyses into the interannual change vectors as opposed to the overall change vector is warranted, and may generate deeper insight into the quantification of the process phase. For example, analysing a series of change vectors over time may uncover the different driving data primitives throughout the duration of the process, via the angle of change. It could therefore subsequently show the different stages of the process and the interaction of the data primitives. Additionally exploring how the magnitude of change varies throughout the process in relation to the angle may enable planners to understand at what point of the process they should expect to observe the greatest changes, and the consequences associated with this change, like displacement. This may enable planners to create a schedule of interventions when process phase (derived from data primitives and sequential analysis of change vectors) and its subsequent consequences are understood.

Additionally, the angular speed may also be investigated, which may reflect the velocity of change in a process, which may vary throughout the cycle, depending upon angle and magnitude. Understanding the angles may also aid with improving the confusion and misclassification observed within predicting the type of gentrification, improving the overall model precision and recall for specific gentrification types. The capabilities of CVA and data primitives for uncovering novel insights into neighbourhood change are untold.

7.5.5 User-Defined Analyses

Although these areas for potential work have focussed on neighbourhood change via gentrification, data primitives provide the capability of user-defined analyses. The approaches to measuring neighbourhood change explored in Section 2.4, are common approaches for good reason, because they are easily implemented and their results are easily interpretable. Thus, the data primitive approach could provide the foundations for these approaches, if the user wishes. For example, data primitives could support the development of a neighbourhood change index. This could be achieved by analysing several neighbourhood processes (gentrification, urban decline, suburbanisation) and creating an index representing the magnitude of change for each individual process. A composite index can then be calculated, reflecting the overall neighbourhood change experienced. This is achievable because the data primitives quantify the changes in state of an area over time.

7.5.6 Considering External Social and Political Changes

Other areas of future work include the consideration of wider social and political changes; the U.K. is undergoing huge societal change as a result of the Brexit referendum in 2016 in which 52% of voters (16 million) voted to leave the European Union (EU). The U.K. officially left the EU in 2020, and significant Brexit related impacts have already been recorded. These include its impact on international student figures (Amuedo-Dorantes and Romiti, 2021; Falkingham et al., 2021), NHS staffing and the subsequent workforce crisis (Majeed, 2017; Dolton et al., 2018;

Oliver, 2022), the agricultural workforce and subsequent crisis (Milbourne and Coulson, 2021; Korir et al., 2021), and hate crimes and associated civic tensions (Devine, 2021; Piatkowska and Stults, 2022; Williams, 2021).

Significant impacts of Brexit upon neighbourhood change should therefore also be expected, especially in regard to gentrification. This is because a decrease in Black and Asian ethnic proportions is one of the main data primitives of the gentrification process as conceptualised throughout this research (see Chapter 5 and Chapter 6). As a result of Brexit, a growth in particular non-white ethic proportions may be more prevalent in all neighbourhoods (based on the assumption that EU migrants are White), even in neighbourhoods that are not experiencing gentrification. However, data primitives work in tandem, and so gentrification will only be identified if all requirements are met. Yet, these increased levels in the growth of non-white ethnic proportions may skew the magnitude of change to which the neighbourhood has experienced as a direct result of gentrification. Consequently, it is likely that gentrification patterns observed within this research period of 2010-2019 may not be replicated in future study periods, particularly 2020 onwards. However, the novelty of data primitives is such that they also provide the capability of analysing and monitoring how Brexit has changed the dynamics of neighbourhood change at an adequate spatiotemporal resolution.

Another layer of complexity will be added to future studies however, with the changing intra and international residential moving behaviours as a result of the COVID-19 pandemic, and the subsequent changes to house prices (Bricongne et al., 2023). Current research has shown that people chose more rural and suburban neighbourhoods over urban neighbourhoods throughout the pandemic, particularly for the duration of lockdowns where daily life and movement outside was restricted, which was also reflected in house prices. During the COVID-19 crisis, house prices increased in rural regions, but London observed a continued decline (Bricongne et al., 2023). This may have a huge impact on the spatial distributions of rural gentrification throughout England in future studies, as well as their periodicities. Additionally, house prices have since increased at record levels across the country, but rising house prices have benefitted those in the middle of the wealth distribution the most (Blundell et al., 2022). Since an increase in house price is one core data primitive of gentrification, like with the potential impacts on gentrification as Brexit, such changes may skew the magnitude of the change the neighbourhood has experienced as a direct result of gentrification. COVID-19 impacts are likely to confound the Brexit impacts on gentrification, and the direct results of the two separate phenomenon may be inextricable. But, nevertheless, they can be monitored as a whole.

7.5.7 Considering Geography, Data, and Noise

This research was conducted at the LSOA geography, representing 1,500 people on average, since it is considered to be representative of neighbourhood effects (Finney, 2013). However, as the spatial resolution of the data is increased, the sample from which the data primitives are drawn, is reduced. These smaller samples can increase the sensitivity of these measures to extreme and anomalous values. An area of future work could therefore consider the impact of spatial resolution on the identified and predicted cycles of change. For example, does the Middle Super Output Area (MSOA) provide more "stable", less sensitive representations of neighbourhood change than the LSOA? For example Duque et al. (2018) have devised a statistical test - "S-maup" to find the maximum spatial aggregation that avoids the negative consequences of the Modifiable Areal Unit Problem (MAUP), which may indicate the most appropriate spatial scale at which to identify and analyse neighbourhood processes like gentrification.

Additionally, there is the potential for considering the combination of data primitives for exploring gentrification in the U.K. in future works. The selection of the four key variables - house price (increase), professional occupation (increase), neighbourhood churn (increase), and Black and Asian ethnicities (decrease) — to define gentrification demonstrates the strength of data primitives in capturing different types of this complex process. However, future research may consider the inclusion of residential rent data from sources like Zoopla which could further enhance the understanding of housing dynamics in gentrification processes, particularly from the supply side (Blasius et al., 2016). Additionally, experimenting with the ethnicity variable could provide valuable insights into the intersectionality of gentrification and its impact on different racial and ethnic groups. For example, altering the Black and Asian ethnicities to increase instead of decrease may capture cycles of gentrification associated with these ethnicities (Hwang, 2020). Alternatively, this primitive could be removed, and the differences within the identified cycles of gentrification could be explored. These may contribute to the refinement of the measurement and analysis of gentrification, thereby advancing the field's understanding of the complex urban phenomenon.

A consequence of the sensitivity of the algorithm (Table 3.4) is that it is vulnerable to noise in the source data, which raises concerns when coupled with the somewhat rigid methodology of the data primitive approach in identifying gentrification. The binary determination of "no gentrification" if any of the four measures fall on the wrong side of zero demonstrates the brittleness of the approach. This rigidity can potentially overlook subtler manifestations of gentrification or fail to capture the complexities inherent in the process, hence the potential for exploring the combinations of data primitives. These three issues highlight the need for a refinement of the data primitive approach, which has also been explored in Section 7.5.2.

7.5.8 Summary

To summarise, there are limitations and associated routes of future research regarding both the methodological improvements, and the applications of the data primitive approach, as stated above. Nevertheless, data primitives provide a novel approach to neighbourhood change that can lead to many unique and exciting results that contribute to the fields of neighbourhood change and gentrification, and beyond. Their design enables the identification of gentrification and other such neighbourhood processes at a spatiotemporal resolution that captures local dynamics. The analysis of these dynamics uncovers cycles of gentrification which when analysed through space and time reveal insights that could not be achieved with traditional approaches to neighbourhood change as explored in Section 2.4. Thus, data primitives, particularly when further refined, are an effective tool for identifying, quantifying, and predicting neighbourhood change via neighbourhood processes.

Chapter 8: Conclusion

The major contribution of this research is the methodological advancement of dynamic neighbourhood change research via the conceptualisation and establishment of the data primitive approach. The methods outlined throughout this thesis suggest a more nuanced approach to geodemographic and neighbourhood change research, away from a focus on classifications and other traditional methods, and static data, towards an approach that captures the social dynamics experienced by neighbourhoods.

This research has provided as a proof of concept and confirmation that data primitives can be conceptualised in reference to current literature and explored as an alternative approach for analysing neighbourhood change, particularly via neighbourhood processes (RQ1). Suitable conceptualisation of the data primitives via literature reviews enables them to be used for the identification of neighbourhood change, with this research focussing upon gentrification (RQ2). Although this research has highlighted the data primitive approach in relation to exploring gentrification, the approach can be easily adapted to capture a range of other neighbourhood processes like urban decay and suburbanisation.

Due to the temporal resolution of data primitives, sequential gentrification-associated changes can be captured. These sequential changes infer a cycle of gentrification, whereby their start, peak, and end years are identified, alongside the number of years the cycle took to reach its peak from start, the number of years to complete from its peak, and the overall duration (RQ3). These temporal properties give an indication of how gentrification manifests throughout space, described through the regional analysis in Chapter 5 (RQ3).

Identified cycles of gentrification in South Yorkshire were visually validated alongside insights from literature and the data primitive changes, resulting in the processes being assigned a type of gentrification; residential, rural, and transport. These then provided the basis for generating models to predict the spatial extent of different types of gentrification in England (RQ4). The regional model had excellent accuracy and kappa values, indicating minimal confusion, but misclassification still occurred. Additionally, the national model remains unvalidated.

To add to this, the identified temporal properties of cycles of gentrification in South Yorkshire provided as the basis for generating models to predict the temporal properties, which enable the exploration of the manifestation of gentrification via their cycles in England (RQ4). Such predictions would be incredibly useful for local authorities planning departments; they can generate a wealth of information regarding the neighbourhoods expected to gentrify (Chapter 6 RO1), the type of gentrification expected to be experienced (Chapter 6 RO2), and the timeline in which they are expected to gentrify (Chapter 6 RO3). Subsequently, a timetable of appropriate measures to both minimise disruption to the incumbent population, whilst generating benefits for incumbent and incomers can be devised.

Beyond the methodological advancement, this research contributes to the fields of neighbourhood change and gentrification. Traditional approaches to neighbourhood change research, which are static and cross-sectional, are not capable of measuring dynamic change over time. Data primitives can however generate some important insight into the spatial and temporal extent of gentrification within England that traditional approaches do not afford, like their associated manifestations and periodicities (RQ5).

References

Aalbers, Manuel B (2019). “Introduction to the forum: From third to fifth-wave gentrification”. In: *Tijdschrift voor economische en sociale geografie* 110.1, pp. 1–11. ISSN: 0040-747X.

Abreu, Maria and Jones, Calvin (2021). “The shadow of the Pithead: understanding social and political attitudes in former coal mining communities in the UK”. In: *Applied Geography* 131, p. 102448. ISSN: 0143-6228.

Amuedo-Dorantes, Catalina and Romiti, Agnese (2021). “International student applications in the United Kingdom after Brexit”. In.

Andersen, Hans Skifter (2019). *Urban sores: On the interaction between segregation, urban decay and deprived neighbourhoods.* Routledge. ISBN: 1351753711.

Anguelovski, Isabelle, Connolly, James J T, Masip, Laia, and Pearsall, Hamil (2018). “Assessing green gentrification in historically disenfranchised neighborhoods: a longitudinal and spatial analysis of Barcelona”. In: *Urban Geography* 39.3, pp. 458–491. ISSN: 0272-3638.

Bardaka, Eleni, Delgado, Michael S, and Florax, Raymond J G M (2018). “Causal identification of transit-induced gentrification and spatial spillover effects: The case of the Denver light rail”. In: *Journal of Transport Geography* 71, pp. 15–31. ISSN: 0966-6923.

Barrie, Helen, Soebarto, Veronica, Lange, Jarrod, Mc Corry-Breen, Fidelma, and Walker, Lauren (2019). “Using citizen science to explore neighbourhood influences on ageing well: Pilot project”. In: *Healthcare*. Vol. 7. 4. MDPI, p. 126. ISBN: 2227-9032.

Barton, Michael (2016). "An exploration of the importance of the strategy used to identify gentrification". In: *Urban Studies* 53.1, pp. 92–111. ISSN: 0042-0980.

Blasius, Jörg, Friedrichs, Jürgen, and Rühl, Heiko (2016). "Pioneers and gentrifiers in the process of gentrification". In: *International Journal of Housing Policy* 16.1, pp. 50–69. ISSN: 1949-1247.

Blundell, Richard, Costa Dias, Monica, Cribb, Jonathan, Joyce, Robert, Waters, Tom, Wernham, Thomas, and Xu, Xiaowei (2022). "Inequality and the COVID-19 Crisis in the United Kingdom". In: *Annual Review of Economics* 14, pp. 607–636. ISSN: 1941-1383.

Bratchford, Gary (2020). "Visualizing gentrification in Ancoats, Manchester: a multi-method approach to mapping change". In: *Gentrification around the World, Volume I: Gentrifiers and the Displaced*, pp. 141–174. ISSN: 3030413365.

Bricongne, Jean-Charles, Meunier, Baptiste, and Pouget, Sylvain (2023). "Web-scraping housing prices in real-time: The Covid-19 crisis in the UK". In: *Journal of Housing Economics* 59, p. 101906. ISSN: 1051-1377. DOI: https://doi.org/10.1016/j.jhe.2022.101906. URL: https://www.sciencedirect.com/science/article/pii/S105113772200078X.

Cameron, Stuart (2003). "Gentrification, housing redifferentiation and urban regeneration:'going for growth'in Newcastle upon Tyne". In: *Urban Studies* 40.12, pp. 2367–2382. ISSN: 0042-0980.

Caragliu, Andrea, Del Bo, Chiara, and Nijkamp, Peter (2011). "Smart cities in Europe". In: *Journal of urban technology* 18.2, pp. 65–82. ISSN: 1063-0732.

Champion, Tony (2001). "Urbanization, suburbanization, counterurbanization and reurbanization". In: *Handbook of urban studies* 160.1, pp. 143–161.

Cole, Helen V S, Mehdipanah, Roshanak, Gullón, Pedro, and Triguero-Mas, Margarita (2021). "Breaking Down and Building Up: Gentrification, Its drivers, and Urban Health Inequality".

In: *Current Environmental Health Reports* 8.2, pp. 157–166. ISSN: 2196-5412. DOI: 10.1007/s40572-021-00309-5. URL: https://doi.org/10.1007/s40572-021-00309-5.

Comber, A J (2008). "The separation of land cover from land use using data primitives". In: *Journal of Land Use Science* 3.4, pp. 215–229. ISSN: 1747-423X. DOI: 10.1080/17474230802465173. URL: https://doi.org/10.1080/17474230802465173.

Comber, Alexis and Wulder, Michael (2019). *Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use.*

Delmelle, Elizabeth C (2021). "Transit-induced gentrification and displacement: The state of the debate". In: *Advances in Transport Policy and Planning*. Vol. 8. Elsevier, pp. 173–190. ISBN: 2543-0009.

Devine, Daniel (2021). "Discrete events and hate crimes: The causal role of the Brexit referendum". In: *Social Science Quarterly* 102.1, pp. 374–386. ISSN: 0038-4941.

Dolton, Peter, Nguyen, David, Castellanos, Maria, and Rolfe, Heather (2018). "BREXIT and the Health & Social Care Workforce in the UK". In: *Report to the Cavendish Coalition, NIESR (forthcoming).*

Duque, Juan C, Laniado, Henry, and Polo, Adriano (2018). "S-maup: Statistical test to measure the sensitivity to the modifiable areal unit problem". In: *PloS one* 13.11, e0207377. ISSN: 1932-6203.

Falkingham, Jane, Giulietti, Corrado, Wahba, Jackline, and Wang, Chuhong (2021). "The impact of Brexit on international students' return intentions". In: *The Manchester School* 89.2, pp. 139–171. ISSN: 1463-6786.

Fertner, Christian, Groth, Niels Boje, Herslund, Lise, and Carstensen, Trine Agervig (2015). "Small towns resisting urban decay through residential attractiveness. Findings from Denmark". In: *Geografisk Tidsskrift-Danish Journal of Geography* 115.2, pp. 119–132. ISSN:

0016-7223. DOI: 10.1080/00167223.2015.1060863. URL: https://doi.org/10.1080/00167223.2015.1060863.

Finio, Nicholas (2022). "Measurement and definition of gentrification in urban studies and planning". In: *Journal of Planning Literature* 37.2, pp. 249–264. ISSN: 0885-4122.

Finney, Nissa (2013). *Statistical boundaries and small area data: something worth saving?* https://citiesmcr.wordpress.com/. URL: https://citiesmcr.wordpress.com/2013/09/23/statistical-boundaries-and-small-area-data-something-worth-saving/.

Franco, Sofia F and Macdonald, Jacob L (2018). "Measurement and valuation of urban greenness: Remote sensing and hedonic applications to Lisbon, Portugal". In: *Regional Science and Urban Economics* 72, pp. 156–180. ISSN: 0166-0462.

Goodchild, Michael F (2007). "Citizens as sensors: the world of volunteered geography". In: *GeoJournal* 69, pp. 211–221. ISSN: 0343-2521.

Gray, Jennie, Buckner, Lisa, and Comber, Alexis (2021). "Extending geodemographics using data primitives: A review and a methodological proposal". In: *ISPRS International Journal of Geo-Information* 10.6. ISSN: 22209964. DOI: 10.3390/ijgi10060386.

Gura, Trisha (2013). "Citizen science: amateur experts". In: *Nature* 496.7444, pp. 259–261. ISSN: 0028-0836.

Hincks, Stephen (2015). "Neighbourhood change and deprivation in the Greater Manchester city-region". In: *Environment and Planning A* 47.2, pp. 430–449. ISSN: 0308-518X.

— (2017). "Deprived neighbourhoods in transition: Divergent pathways of change in the Greater Manchester city-region". In: *Urban Studies* 54, pp. 1038–1061.

Hwang, Jackelyn (2020). "Gentrification without segregation? Race, immigration, and renewal in a diversifying city". In: *City & Community* 19.3, pp. 538–572. ISSN: 1535-6841.

Hwang, Jackelyn and Sampson, Robert J (2014). "Divergent pathways of gentrification: Racial inequality and the social order of renewal in Chicago neighborhoods". In: *American Sociological Review* 79.4, pp. 726–751. ISSN: 0003-1224.

Ilic, Lazar, Sawada, Michael, and Zarzelli, Amaury (2019). "Deep mapping gentrification in a large Canadian city using deep learning and Google Street View". In: *PloS one* 14.3, e0212814. ISSN: 1932-6203.

Korir, Lilian, Drake, Archie, Collison, Martin, Camacho-Villa, Tania Carolina, Sklar, Elizabeth, and Pearson, Simon (2021). "Current and emergent economic impacts of Covid-19 and Brexit on UK fresh produce and horticultural businesses". In: *arXiv preprint arXiv:2101.11551*.

Maantay, Juliana A and Maroko, Andrew R (2018). "Brownfields to greenfields: Environmental justice versus environmental gentrification". In: *International journal of environmental research and public health* 15.10, p. 2233. ISSN: 1660-4601.

Majeed, Azeem (2017). "Brexit will put further pressure on the NHS". In: *BMJ* 356. ISSN: 0959-8138.

Mendes, Luís (2013). "Marginal Gentrification as Emancipatory Practice: An Alternative to the Hegemonic Discourse of the Creative City?" In: *RCCS Annual Review. A selection from the Portuguese journal Revista Crítica de Ciências Sociais* 5. ISSN: 1647-3175.

Milbourne, Paul and Coulson, Helen (2021). "Migrant labour in the UK's post-Brexit agri-food system: ambiguities, contradictions and precarities". In: *Journal of Rural Studies* 86, pp. 430–439. ISSN: 0743-0167.

Oliver, David (2022). "David Oliver: Act on workforce gaps, or the NHS will never recover". In: *BMJ* 376. ISSN: 1756-1833.

Pearsall, Hamil and Eller, Jillian K (2020). "Locating the green space paradox: A study of gentrification and public green space accessibility in Philadelphia, Pennsylvania". In: *Landscape and Urban Planning* 195, p. 103708. ISSN: 0169-2046.

Piatkowska, Sylwia J and Stults, Brian J (2022). "Brexit, terrorist attacks, and hate crime: A longitudinal analysis". In: *Social Problems* 69.4, pp. 968–996. ISSN: 0037-7791.

Reibel, Michael and Regelson, Moira (2007). "Quantifying neighborhood racial and ethnic transition clusters in multiethnic cities". In: *Urban Geography* 28.4, pp. 361–376. ISSN: 0272-3638.

Shin, Hyun Bang, Lees, Loretta, and López-Morales, Ernesto (2016). "Introduction: Locating gentrification in the global east". In: *Urban Studies* 53.3, pp. 455–470. ISSN: 0042-0980.

Shmaryahu-Yeshurun, Yael and Ben-Porat, Guy (2021). "For the benefit of all? State-led gentrification in a contested city". In: *Urban Studies* 58.13, pp. 2605–2622. DOI: 10.1177/0042098020953077. URL: https://journals.sagepub.com/doi/abs/10.1177/0042098020953077.

Velastegui, Julia Desiree, Velastegui, Luis Alejandro, and Moyon, Manuel Fabián (2019). "Monitoring Land Use Change for Gentrification in Soho (London), Using Geographic Information Systems and Google Street View". In: *European Scientific Journa* 15.17.

Williams, Thomas (2021). "Mobilizing the Past: Germany and the Second World War in Debates on Brexit". In: *Revue LISA/LISA e-journal. Littératures, Histoire des Idées, Images, Sociétés du Monde Anglophone–Literature, History of Ideas, Images and Societies of the English-speaking World* 19.51. ISSN: 1762-6153.

Zuk, Miriam, Bierbaum, Ariel H, Chapple, Karen, Gorska, Karolina, and Loukaitou-Sideris, Anastasia (2018). "Gentrification, displacement, and the role of public investment". In: *Journal of Planning Literature* 33.1, pp. 31–44. ISSN: 0885-4122.

Appendices

Appendix A: Author CRediT statements

The CRediT (Contributor Roles Taxonomy) authors statements for the included publications

Chapter 4: Extending Geodemographics Using Data Primitives: A Review and a Methodological Proposal

Author CRediT statement

Jennie Gray: Conceptualization, Data Curation, Methodology, Validation, Formal analysis, Resources, Visualisation, Writing – original draft & Writing – review & editing, **Lisa Buckner**: Supervision, Writing – review & editing, **Alexis Comber**: Supervision, Conceptualization, Software, Writing – review & editing.

Chapter 5: Identifying Neighbourhood Change Using a Data Primitive Approach: the example of gentrification

Author CRediT statement

Jennie Gray: Conceptualization, Data Curation, Methodology, Software, Validation, Formal analysis, Visualisation, Writing – original draft & Writing – review & editing, **Lisa Buckner**: Supervision, Writing – review & editing, **Alexis Comber**: Supervision, Conceptualization, Software, Writing – review & editing.

Chapter 6: Predicting Gentrification in England: A Data Primitive Approach

Author CRediT statement

Jennie Gray: Conceptualization, Data Curation, Methodology, Software, Validation, Formal analysis, Resources, Visualisation, Writing – review & editing, **Lisa Buckner**: Supervision, Conceptualization, Writing – review & editing, **Alexis Comber**: Supervision, Conceptualization, Software, Writing – review & editing.

Appendix B: Validating Identified Gentrification with Google Earth

Appendix A: Validating 123 cycles of gentrification in South Yorkshire: Google Earth

Table A1: The criteria for the assignment of gentrification type, according to Google Earth and further validation through the Index of Multiple Deprivation

Gentrification Type	Including Facets of…	Criteria
Residential Gentrification	New Build	New build, or development driven, gentrification, typically occur in urban residential areas. It can include the development of new housing estates, on either brownfield land, or via the replacement of old housing estates. It can also include smaller scale developments like infill housing developments and residential upgrading.
	Endogenous	Endogenous gentrification is where the in-migration of residents into bordering neighbourhoods increases house prices in the original, poorer neighbourhood. Thus, no developments have been experienced within the bounds of the LSOA, but there has been significant development just beyond the LSOA.
	General	No real observable changes are identified in general gentrification, but the associated data primitive changes and change vectors for the LSOA are meaningful. This suggests that gentrification was observed during the identified period.
Rural Gentrification	Rural	Rural gentrification can include few observable changes. For example, it can include facets of residential upgrading, driven by the incomers of a higher socioeconomic status. Here, associated data primitives and change vectors suggest meaningful changes, which can be attributed to the rural location of the neighbourhoods.
	New Build	Rural new build gentrification is driven by new housing estates, on either brownfield land, or previously undeveloped land, in rural areas.
Transport Gentrification	Rail Induced	Rail-Induced gentrification refers to the gentrification associated with the proximity to railway stations. These neighbourhoods will have meaningful changes within their data primitives and change vectors.
	Transit Induced	Transit-Induced gentrification is the proximity to transportation links including tram stops, bus stations, and motorway junctions. These neighbourhoods will have meaningful changes within their data primitives and change vectors.
	Studentification	Studentification has many facets, like the development of purpose-built student accommodation. Studentification was placed into the Transport Gentrification due to the associated connectivity of student neighbourhoods, for the ease of movement to and from term time addresses (rail and motorway), and the ease of movement into university and other amenities (bus and tram).

None		No major developments were observed of any kind, and changes in data primitives were only slight. Thus, no gentrification was deemed to have occurred throughout this period.

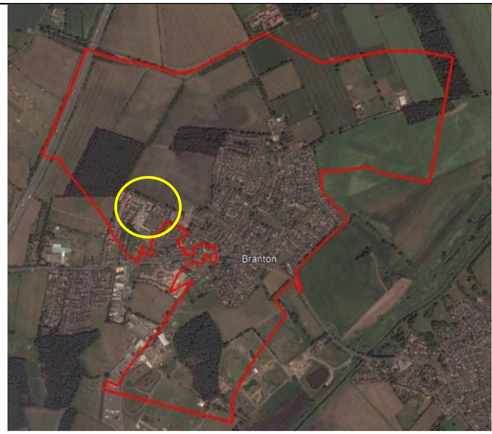
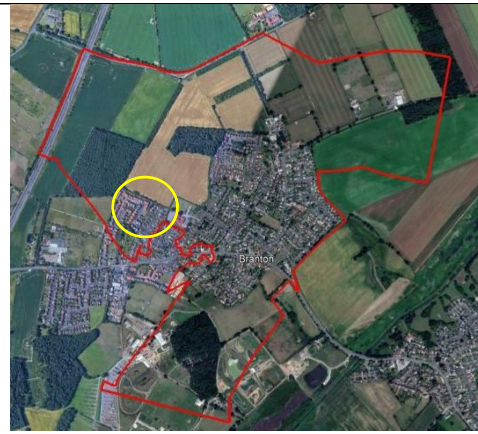
Although 10 different types of gentrification were identified within this validation exercise, they are assigned to one high-level type of gentrification, **Residential**, **Rural**, or **Transport**. Table A1 shows the criteria for the assignment. This is because some of the more specific identified gentrification types have limited counts, insufficient for training a predictive model. Table A2 shows the count of the final assignments.

Table A2: The frequencies of the identified gentrification types, and the gentrification types going forward into the analysis

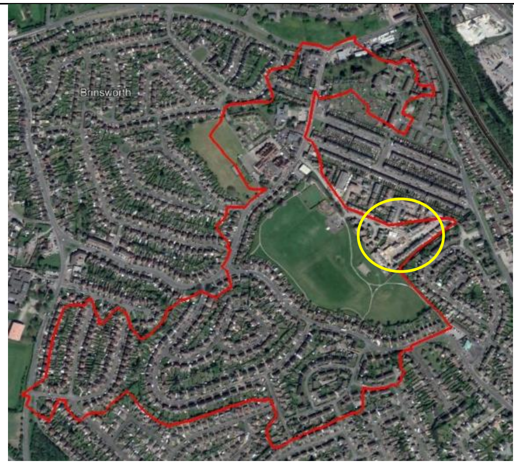
Gentrification Type	**Frequency**
Residential	60
Rural	20
Transport	40
None	3

Table A3 describes the changes observed in the validation, proposing a suitable type of gentrification. However, due to the ambiguity of some changes, multiple gentrification types may be suggested, with the final assignment being colour coded, in **bold.**

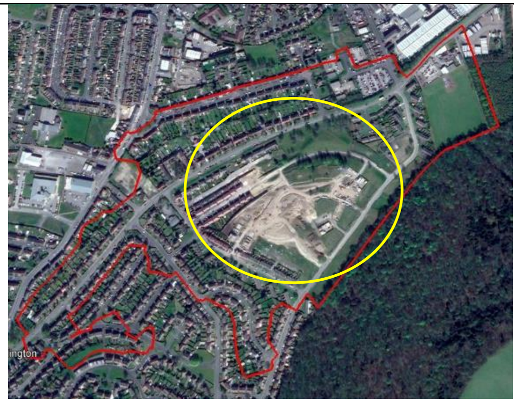
Table A3: The Google Earth validation of the 123 identified localities of gentrification in South Yorkshire

	Start	Peak	End
E01007601			
	Started in 2010, imagery from 2009.	2015	2019, imagery from 2017.
	The peak of gentrification in Branton coincides with the development of a new housing estate to the west of the LSOA. The latest image does not cover the full cycle of gentrification, but the housing estate is complete. This is a rural town, so rural new build gentrification. **Rural Gentrification.**		
E01008131			
	Started 2013, imagery from 2015.	2018	2019 Imagery, 2020
	The peak of gentrification here in Stannington coincides with the construction of a new housing estate to the very west of the LSOA. The last image is 2020 post gentrification, with the completed housing development. New build gentrification. **Residential Gentrification.**		

<table>
<tr><td rowspan="2">E01007704</td><td>

2010, imagery 2002.</td><td>

2013, imagery 2015</td><td>

2015, imagery 2016.</td></tr>
<tr><td colspan="3">Gentrification starts in Brinsworth in 2010. The first image shows an industrial estate to the east of the LSOA, beyond the field. By 2015 (closest image to the peak) a new housing estate is being build on the old industrial land. 2016, and the development is complete. New build gentrification – brownfield. Residential Gentrification</td></tr>
<tr><td rowspan="2">E01007548</td><td>
2010, imagery 2002</td><td>
2014.</td><td>
2015, imagery 2017.</td></tr>
<tr><td colspan="3">Gentrification starts in Edlington in 2010, but the first image is from 2002. It shows the existence of a low-income, working class estate. By the peak in 2014, the old estate has been demolished, and a new estate is being built. The cycle ended in 2015, but this does not coincide with the completion of the housing estate in 2020, nor the imagery from 2017. New build gentrification – replacement. Residential Gentrification.</td></tr>
</table>

E01007362	 2010, image 2009.	 2017, image 2016.	 2018.
	Gentrification starts in Mapplewell in 2010, by 2016 we can see the clearance and preperation of brownfield land ready for development, which by 2018 (the end of the cycle) the new housing development is half completed, but the completed section exists fully with the LSOA. New build gentrification – brownfield. **Residential Gentrification.**		
E01007327	 2015.	 2018.	 2019.
	Gentrification starts in Athersley South in 2015, and infill developments are in the midst of construction in two separate locations. By the peak in 2018, the small new build developments have been completed. New build gentrification. **Residential Gentrification.**		

E01007405

2014, image 2009.

2016.

2017, image 2018.

Gentrification starts in 2014, but the first image from 2009 shows an idustrial estate to the northeast of the LSOA. The peak in 2016 coincides with the development of a new build housing estate on this partial brownfield land. The captured gentrification cycle ended in 2017, but the estate was not completed until 2019. New build gentrification – brownfield. **Residential Gentrification.**

E01007387

2010, image 2009.

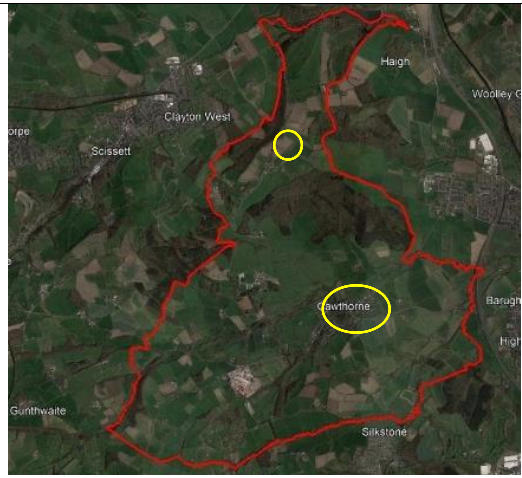
Gentrification peaks in 2013. No imagery available.

2015.

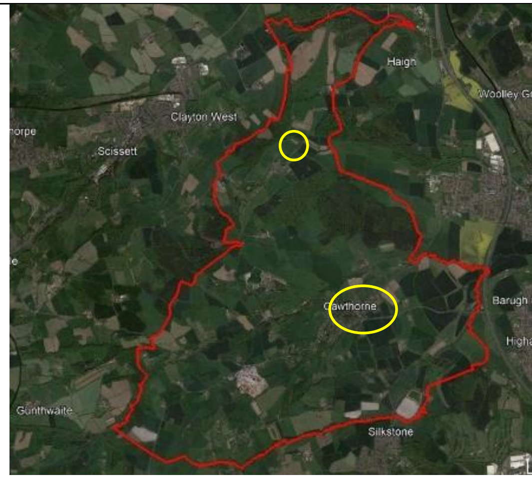
Gentrification starts in 2010 in Barugh Green. By the time the cycle ends in 2015, there has been some infill housing to the west of the LSOA, the demolition of semi-detached houses for a row of new-build terraced houses, and a new supermarket to the east of the LSOA, near the industrail site. Due to its “urban minor conurbation” classification, it is assigned as **Residential Gentrification.**

E01007418

2015



2018



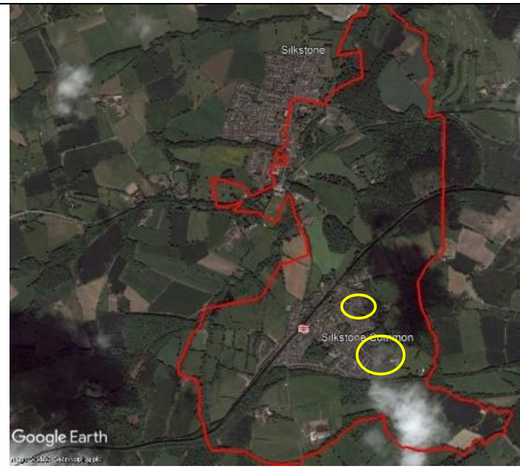
2019.

There are no clear signs of any specific type of development from this aerial imagery, until zoomed in. There were numerous residential upgrading projects that were undertaken throughout the LSOA, largely in Cawthorne, but also in High Hoyland, but some of which may have started prior to 2015. But, due to the location and setting of "rural village and dispersed" it has been assigned as **Rural Gentrification.**

E01007419

2011

Gentrification peaks in 2014. No imagery available.

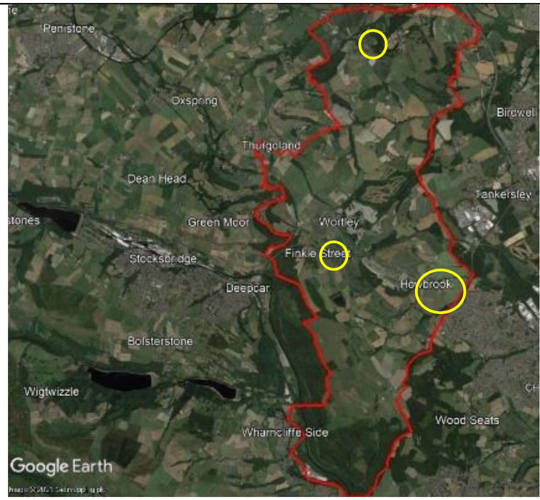


2015.

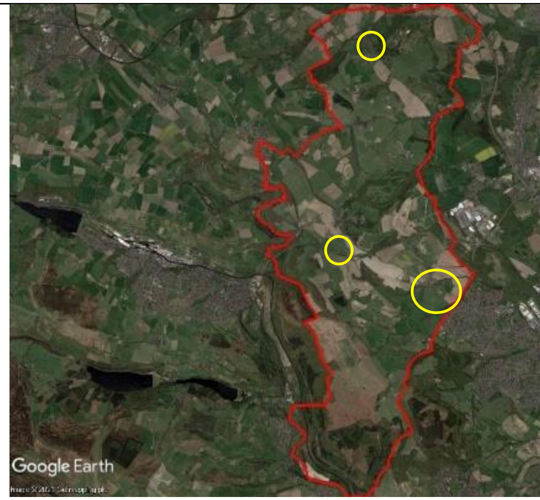
This LSOA consisting of Silkstone Common has several changes, including the demolition of a residential property, and its larger, new build replacement (x2) and further residential upgrading, and a new school building. However, it encompasses a rail station in a rural setting and could potentially be rail-induced gentrification, or rural gentrification. Assigned as **Transport Gentrification**, due to proximity of train station.

E01007421

2014, imagery 2009.



2018



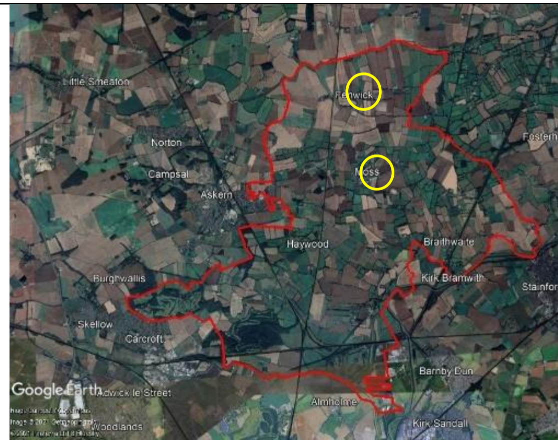
2019

There are no clear signs of development until zoomed in. There are three infill housing developments in Howbrook, that starts in 2016, 2017, and 2018 respectively, and an upgrading house. Infill is also identified in Finkle Street, upgrading in Hood Green, Due to these infill developments in "rural village and dispersed", it is assigned as **Rural Gentrification.**

E01007488

2010, imagery 2009.

2017. no imagery available.



2018, imagery 2019.

Again, there are no clear developments from this aerial imagery until zoomed in. There are infill housing developments in Fenwich and Moss. Due to its "rural village and dispersed" location, it is assigned as **Rural Gentrification.**

E01007490

2013.

2015.

2017.

Again, there are no clear developments from this aerial imagery until zoomed in. There is an infill housing development identified in both Burghwallis and Sutton, that coincides with the peak. Due to its rural vilalge and dispersed” location, it is assigned as **Rural Gentrification.**

E01007626

2010. Imagery 2009.

2013.

2014. imagery 2015.

This LSOA encompasses the western edge of Stainforth, and captures the development of a new build housing estate above the greyhound stadium, to the southeast corner of the LSOA. A much larger development is built in 2019 above this new build housing estate, but this is not captured by the analysis. However, this LSOA is within meters of a train station, thus such residential developments may be driven by access to transport infrastructure. This is therefore assigned as **Transport Gentrification.**

<table>
<tr>
<td rowspan="2">E01007469</td>
<td>

2011. Imagery 2009.</td>
<td>

2013.</td>
<td>

2015</td>
</tr>
<tr>
<td colspan="3">There are no clear developments in Skellow/Old Skellow from this aerial imagery, but its associated data suggests gentrification type changes. So, due to its rural setting, it is assigned as Rural Gentrification.</td>
</tr>
<tr>
<td rowspan="2">E01007555</td>
<td>

2010.</td>
<td>2017. no imagery available.</td>
<td>

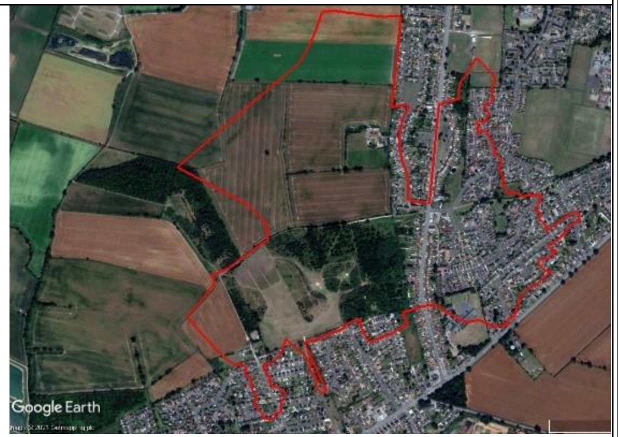
2018. Imagery from 2019.</td>
</tr>
<tr>
<td colspan="3">There are no clear changes from this aerial imagery in Hatfield/Dunscroft, but data suggests slight gentrification type changes. It is located within close proximity (just outside the bounds of the aerial image) to the Stainforth/Hatfield railstation. Therefore, it is potentially rail-induced gentrification.
Transport Gentrification.</td>
</tr>
</table>

<table>
<tr><td rowspan="3">E01007556</td><td>

2012. Imagery from 2009.</td><td>

2017</td><td>

2018. Imagery from 2019.</td></tr>
<tr><td colspan="3">There are few clear changes from this aerial imagery in Dunscroft, but there is evidence of some residential upgrading, and data suggests slight changes associated with gentrification. It is located within close proximity (just outside the bounds of the aerial image) to the Stainforth/Hatfield railstation. Therefore, it is potentially rail-induced gentrification. Transport Gentrification.</td></tr>
<tr></tr>
<tr><td rowspan="2">E01007562</td><td>

2010, image 2009.</td><td>2017. no imagery available.</td><td>

2018. Imagery from 2019.</td></tr>
<tr><td colspan="3">No clear changes from this aerial imagery, but this area’s IMD ranking increased (less deprivation) by 1179, so we can presume that neighbourhood changes occurred and gentrification was identified. Potentially rail-induced due to being located near to the hatfield/stainforth railstation. Transport Gentrification.</td></tr>
</table>

E01007621

2012. Imagery from 2009.

2015.

2016. Imagery form 2017.

This LSOA encompasses Kirk Sandall, a village that has its own railstation located just outside of the west of the LSOA. A new build housing estate is built, which may have been catalysed by the train station. **Transport Gentrification.**

E01007623

2011. Imagery from 2009.

2016. Iamgery from 2015.

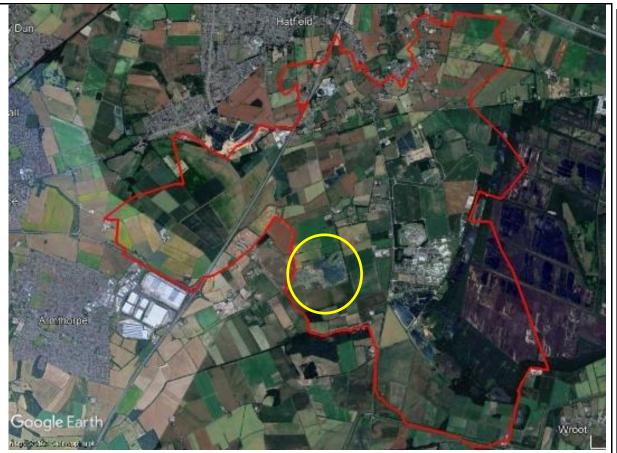
2017.

This LSOA encompasses the south of Kirk Sandall. Here, there are large changes in the centre of the LSOA, a block of flats have been demolished and replaced with a community hall, and an old church has been demolished and replaced with an assisted living residence. Furthermore, the peak also coincides with the development of a new build housing estate just outside the south of the LSOA. Again, due to its proximity to the railstation, it is assigned as **Transport Gentrification.**

E01007557

2013

2016. no imagery available.



2017

This is a rual LSOA. It encompasses several prisons; HMP Lindholme, HMP Hatfield Lakes, and HMP Moorland. To the east of the LSOA sees the development of Woodward Lakes and Lodges, a lakeside holiday home resort. Due to the nature and location of this change, it is assigned as **Rural Gentrification.**

E01007464

2011, imagery 2009.

2014

2015

Large changes are observed within between Woodlands and Adwick le Street; a new school – Outwood Academy – has been built, with the older school demolished, and an old police station was demolited and replaced with new build housing. New build housing – replacement. **Residential Gentrification.**

E01007465

2012. Imagery 2013.

2018

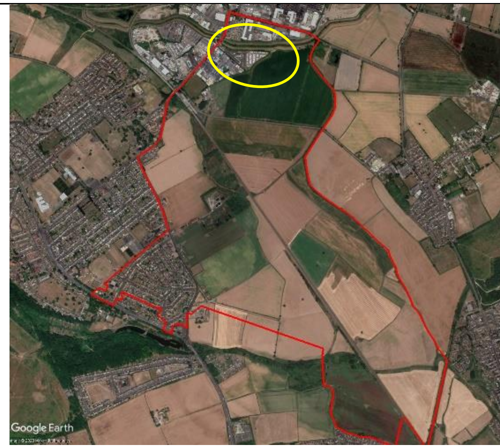
2019. Imagery 2020.

The development of a large new build housing estate was captured in Woodlands. It was half completed by the peak of the gentrification cycle in 2018, and fully completed by the end of the cycle in 2019. A smaller infill housing development was also completed to the east. New build gentrification. **Residential Gentrification.**

E01007504

2012. Imagery 2009

2018



2019. Imagery 2022.

Throughout the period of gentrification in Adwick, we can see the development of Adwick Business Park. The train station located within the LSOA may have catalysed such developments, by increasing. This is assigned as **Transport Gentrification.**

E01007506

2011. Imagery 2009.

2013

2014

Gentrification starts in the north of Bentley in 2011. By the peak in 2013, we can see the development of a new build housing estate. However, it was not completed by the end of the captured cycle in 2014. The development was completed in 2017. Thus, the entire cycle of gentrification in Bentley north was not captured. New build gentrification. **Residential Gentrification.**

E01007476

2012. Imagery 2009

2014.

2015.

The peak of gentrification in Edenthorpe coincides with the development of a new build housing estate at the very north, which has spread eastwerds by the end of the gentrification cycle. New build gentrification. **Residential Gentrification.**

E01007513

2015

2017

2019. Imagery 2022.

Gentrification started in Bentley South/Bentley Rise in 2015. By the peak in 2017, a new build housing estate is being built to the southwest of the LSOA, on brownfield land. It also has a train station, so could be rail-induced gentrification. New build gentrification – brownfield. **Residential Gentrification.**

E01007518

2011. Imagery 2009.

2017. New build housing to west of LSOA.

2018. Imagery 2022.

Gentrification starts in Scawthorpe in 2011. By the peak in 2017, an infill housing estate is being developed, which is completed by 2018. There is also the development of a new school and the demolition of the old school further south of the LSOA. **Residential Gentrification.**

<table>
<tr><td rowspan="3">E01007482</td><td></td><td></td><td></td></tr>
<tr><td>2010, image 2009.</td><td>2016. Imagery 2017.</td><td>2018. Imagery 2021.</td></tr>
<tr><td colspan="3">This cycle of gentrification captures infill housing being built under the hexagonal housing estate at the peak of the cycle, and a new supermarket built where an old community centre/library sat. Residential Gentrification.</td></tr>
<tr><td rowspan="3">E01007652</td><td></td><td></td><td></td></tr>
<tr><td>2012. Imagery 2009.</td><td>2017</td><td>2018 Imagery 2022.</td></tr>
<tr><td colspan="3">Gentrification started in Wheatley in 2012. A new housing estate is being developed to the centre of the LSOA, previously industrial land. New build gentrification - brownfield. Residential Gentrification.</td></tr>
</table>

E01007581

2015

2017

2018. Imagery 2022.

There are no clear observable devleopments or changes within these aerial images of rural Cusworth. The area’s IMD ranking increased (less deprivation) by 789 between 2015-2019, so we can presume that neighbourhood changes have occurred and gentrification was identified. Gentrification is therefore likely to have been rural, due to its setting. **Rural Gentrification.**

E01007483

2011. Imagery 2009.

2015

2016. Imagery 2017.

Gentrification starts in the south of Armthorpe in 2011. By the peak in 2015, we can see some development at the Southfield Primary School, which is completed by 2017. This development driven gentrification may be a facet of **Residential Gentrification.**

<table>
<tr><td rowspan="2">E01007484</td><td>
2011. Imagery 2009.</td><td>2016. no imagery available.</td><td>
2017. imagery 2021</td></tr>
<tr><td colspan="3">Just outside of the LSOA to the east, a new warehouse has been developed. Which has potentially improved job opportunities within the ex mining village of Armthorpe. There is also a small infill housing development to the west of the LSOA. There is also evidence of residential upgrading, with many houses installed solar panels on their roofs. This is assigned as Residential Gentrification.</td></tr>
<tr><td rowspan="2">E01007587</td><td>

2012. Iamge 2009</td><td>

2014</td><td>

2016. Image 2018.</td></tr>
<tr><td colspan="3">Within this captured locality of gentrification, when zoomed in, there is new build development in Clayton. Due to the “rual village and dispersed” lcoation, this is assigned as Rural Gentrification.</td></tr>
</table>

E01007585	 2011. Image 2009	 2015	 2016. Image 2017
	There are no clear changes until you zoom in. Here, there is a new housing estate built on an old industrial site, brownfield new-build gentrification. This is assigned as **Residential Gentrifiaction.**		
E01007617	 2010. Image 2009	 2017	 2018
	There are no clearly visible aerial changes in Sprotborough until zoomed in. Here, residential upgrading is identified, alongside a new children's playground. A little further south (in the same circle), a new farm is identified, but there are little other changes observed. Due to being within the "urban minor conurbation", this is assigned as **Residential Gentrification.**		

E01007619

2011. Image 2009

2016. Image 2017

2018, imagery 2022

Gentrification in Sprotborough started in 2011. This cycle of gentrification captured a new build developments in four separate developments, three of which all within very close proximity. Due to being within the "urban minor conurbation", this is assigned as **Residential Gentrification.**

E01007599

2011. Image 2009

2013. no imagery available.

2015

Gentrification starts in the rural area encompassing Blaxton in 2011. There are very few changes, but a building is demolished in Blaxton near the community hall, which is replaced with a small new housing estate. New build gentrification – rural. Due to the "rural village and dispersed" location, this is assigned **Rural Gentrification.**

<table>
<tr><td rowspan="3">E01007602</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>2010 . image 2009</td><td>2013 . image 2015</td><td>2016. Image 2017</td></tr>
<tr><td colspan="3">Gentrification started in the LSOA encompassing Finningley, and the Doncaster Sheffield (Robin Hood) Airport in 2010. By the peak in 2013, we can see expansion and further development of airport buildings to the northwest. Two large new build housing developments are also identified to the centre and east of the LSOA, the south of Finningley. Transport Gentrification.</td></tr>
<tr><td rowspan="3">E01007603</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>2012, image 2009</td><td>2014</td><td>2016, image 2017</td></tr>
<tr><td colspan="3">Gentrification started in the LSOA encompassing Old Cantley, the southern half of Branton, and a large portion of rural area in 2012, through to 2016. Thorughout this period, a small new-build housing development is identified in Branton. However, this was not completed by the end of the cycle in 2017, nor by the end of the last available image in 2017. Also, throughout this period saw significant expansion of Yorkshire Wildlife Park, a wildlife conservation and rehabilitation tourist attraction. Rural Gentrification.</td></tr>
</table>

E01007520

2012. Image 2009.

2017

2018. Image 2021.

Gentrification started in Bessecarr in 2012. Small changes are observed, including those of residential upgrading to houses. No other clearly visible aerial changes for the period of this gentrification cycle in Bessecarr. But changes in data suggest **Residential Gentrification.**

E01007605

2010. Image 2009

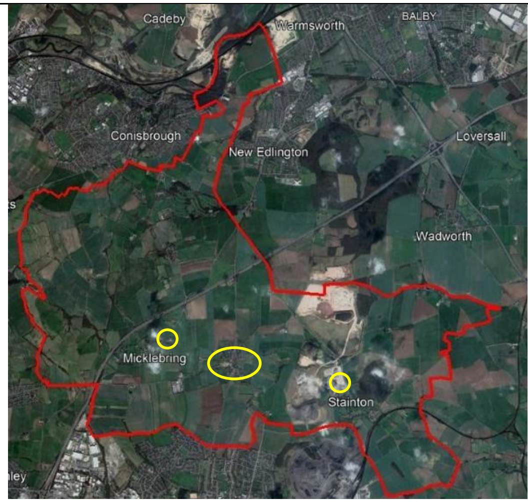
2017

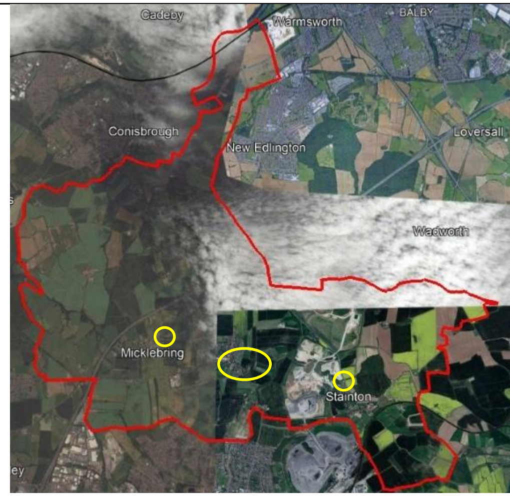
2018. Imagery 2022.

Gentrification starts in 2010 at Bessecarr Grange, infill new build developments are caputed in two locations to the western crop of the LSOA. **Residential Gentrification.**

E01007580	 2012. Image 2009	2016. no imagery available.	 2017
	Getrification starts in Mexborough in 2012. We can see the demolition (before 2012) of an old housing estate to the south of the LSOA. Thus, this cycle may have actually captured gentrification – the uplift of a neighbourhood - via the displacement of the low-income residents rather than via incomers. The town of Mexborough also has a railstation, just outside the bounds of the image. Thus, it could be **Transport Gentrification.**		
E01007552	 2014	2016. no imagery available	 2019. Image 2017
	No clearly visible changes in Warmsworth until the last available image in 2017 This are's IMD ranking between 2010 -2019 (cover entire cycle) increased (less gentrification) by 938. We can therefore presume neighbourhood change occurred, and gentrification was identified. Due to the location's proximity to a major motorway junction, this is lilkely **Transport Gentrification.**		

<table>
<tr>
<td rowspan="2">E01007613</td>
<td>

2012</td>
<td>

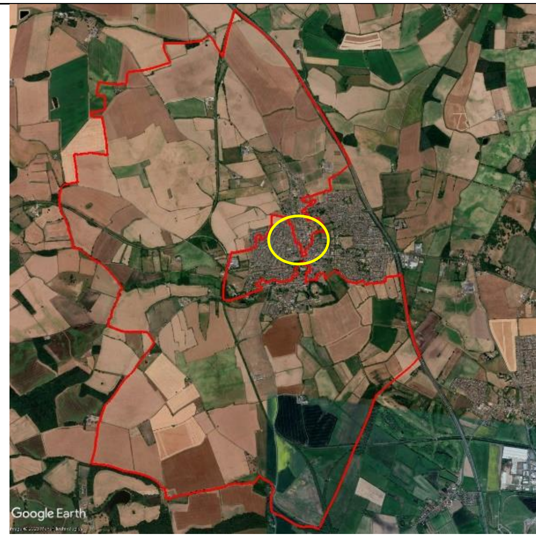
2016</td>
<td>

2017</td>
</tr>
<tr>
<td colspan="3">Changes are not easily visible from this aerial image due to its scale. However, when zoomed in, there are two new build development in Braithwell, the larger of which is built on land that was used as a garden centre. Infill housing is also observed in Stainton, and new-build housing is also seen in Micklebring. Due to its “rural and dispersed” location, it is assigned as Rural Gentrification.</td>
</tr>
<tr>
<td rowspan="2">E01007591</td>
<td>

2011. Image 2009</td>
<td>No imagery available. 2017</td>
<td>

2018</td>
</tr>
<tr>
<td colspan="3">It is hard to pick out major changes in Rossington, but there is evidence of residential upgrading. Many houses upgrade their houses with solar panels, also even an outdoor pool. Some infil housing just outside the LSOA is also observed. Residential Gentrification.</td>
</tr>
</table>

E01007593	 2011. Image 2009	2016. no imagery available.	 2017
	In New Rossington, brownfield land is prepared for a new build housing estate. The cycle end of 2017 does not align with the completion of the development, which is still underway. An industrial estate has been built outside of the LSOA. **Residential Gentrification.**		
E01007610	 2015	2017.no imagery available.	 2018. Image 2020
	We can see that a new build housing estate has been built just above the centre of Bawtry. There has also been some other kind of development (potentially commercial) just southwest of the new build devlepment. Due to its “rural town and fringe” location, it is assigned **Rural Gentrification.**		

E01007612	 2011. Image 2009	 2018. image 2016	 2019. Imagery 2022.
	Gentrification in Tickhill (North and South) captures a small new build development located within the outcrop to the northeast of the LSOA. Tickhill is also a very affluent village in the outskirts of Doncaster, located in a rural setting. **Rural Gentrification.**		
E01007615	 2011. Image 2009	 2018. Image 2016	 2019. Image 2022
	Gentrification started in the east of Tickhill in 2011, by 2016 several infill new build housing developments have been developed to the centre of the LSOA. Imagery from 2022 shows more residential developments to the north of the LSOA. However, due to the latest image being three years post gentrification, it is undertermined how much of this occurred before the 2019 end year. **Rural Gentrification.**		

<table>
<tr><td rowspan="3">E01007697</td><td></td><td></td><td></td></tr>
<tr><td>2010. Image 2009</td><td>2016</td><td>2017</td></tr>
<tr><td colspan="3">Gentrification starts in West Melton in 2010. In 2009, a school exists to the very west of the LSOA. By the peak in 2016, this school had been demolished and a new build housing estate was being developed in its place. By 2017 this was completed. Residential Gentrification.</td></tr>
<tr><td rowspan="3">E01007816</td><td></td><td></td><td></td></tr>
<tr><td>2016, image 2015.</td><td>2018</td><td>2019. Image 2020</td></tr>
<tr><td colspan="3">There are no clear changes within these aerial images in Wath upon Dearne (E), thus the IMD was consulted. The area’s IMD ranking decreased (more deprivation) by over 1000. So, since no changes are observed and the IMD ranking worsened, we can assumed that this area did not experience gentrification. None.</td></tr>
</table>

<table>
<tr>
<td rowspan="3">E01007817</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2012. Image 2009</td>
<td>2018</td>
<td>2019. image 2020</td>
</tr>
<tr>
<td colspan="3">Only very slight changes visible when zoomed in on Wath upon Dearne (SW). Two houses appear to upgrade throughout this cycle of gentrification, each with a large rear extension. However, there is also a new-build housing development just outside of the LSOA, which couls suggest endogenous gentrification. Residential Gentrification.</td>
</tr>
<tr>
<td rowspan="3">E01007819</td>
<td>
</td>
<td>2016. no imagery available</td>
<td>
</td>
</tr>
<tr>
<td>2011. Image 2009</td>
<td></td>
<td>2017</td>
</tr>
<tr>
<td colspan="3">There are a few small changes when zoomed in to the south of the residential area of Wath upon Dearne, including an infill housing development, whereas slightly further north we see the demolition of a residential building. Further north again, we see the demolition and replacement of another residential building. A further possibly residential building is in the midst of construction by the end of the captured cycle in 2017. Due to being classified as “urban minor conurbation”, this is assigned as Residential Gentrificaiton.</td>
</tr>
</table>

<table>
<tr><td rowspan="2">E01007821</td><td>
2015</td><td>
2018</td><td>
2019. Image 2020</td></tr>
<tr><td colspan="3">This LSOA encompasses the centre of Wath-upon-Dearne. There is a visible change to the northeast of the LSOA, however this just relates to land cover. There are new build housing developments outside of the LSOA to the northeast and the southwest, thus this may have supported the theory of endogenous gentrification. Residential Gentrification.</td></tr>
<tr><td rowspan="2">E01007798</td><td>

2010. Image 2009</td><td>2013. no imagery available</td><td>

2014</td></tr>
<tr><td colspan="3">Gentrification started here in Swinton in 2010 and ends in 2014. In 2014, a new build housing estate has begun development on partial brownfield land beside the River Don. This new build estate however was not completed until 2018, so not the entire cycle was captured. Further, just to outside the north of the LSOA is Swinton’s railstation. Thus, such new build gentrification may have been rail-induced. Transport Gentrification.</td></tr>
</table>

E01007808

2010. Image 2009

2013. no imgery available

2015

Gentrification starts in 2010, the image (2009) shows a new build housing estate in construction just outside of the LSOA. No other changes are observed from these aerial images of Thorpe Hesley. Further, this area's IMD ranking increased (less deprivation) by 3256, so we can presume that residential change has occurred and gentrification is correctly identified. Due to its rural fringe location, it is assigned **Rural Gentrification.**

E01007739

2013. Image 2014

2016. no imagery available

2017.

Gentrification started in East Dene in 2013. We can see a small new build development at the south, to let housing built onto the side of the local post office. There has also been a large residential upgrade further north. Its IMD ranking of 517 (2019) of 32844, shows that this area is considered within the top 10% of most deprived areas within England. **Residential Gentrification.**

E01007685

2010. Image 2009

2014

2015

Gentrification started in the LSOA encompassing Hellaby. There has been one commercial development near the centre of the LSOA, just above the motorway junction, and an infill housing development further east. The gentrification in this area is therefore either transit induced, due to the motorway junction in the centre of the LSOA, or residential. **Transport Gentrification.**

E01007754

2010. Image 2009

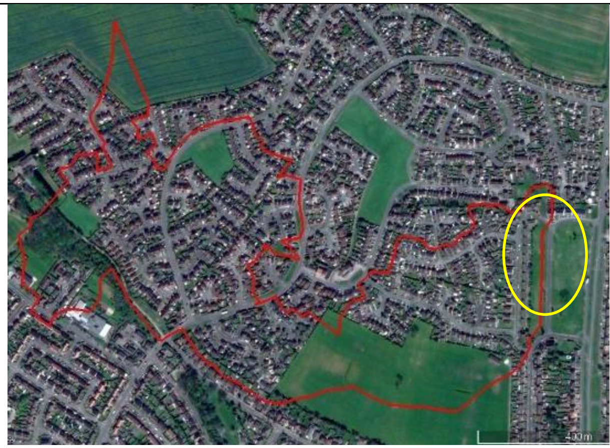
2014

2015

There are no clearly visible changes in the aerial imagery of the northwest of Maltby/Cliff Hills for the duration of gentrification. **Residential Gentrification.**

E01007756

2010. Image 2009



2014

2015

Gentrification in the north of Maltby (N) captures the demolition of five residential buildings to the very east of the LSOA, in which the estate had already started demolition. This entire estate (the rest beyond the LSOA) has since been replaced with a new housing estate and apartment block, which would be new-build gentrification - replacement. However, the cycle does not capture the period of the rebuild. Thus, this cycle may represent the 'uplift' of a neighbourhood via the displacement of the low-income. **Residential Gentrification.**

E01007762

2010. Image 2009

2014

2018. Image 2022.

This cycle captured the demolition and rebuild of a housing estate to the north of the LSOA in Maltby, a prime example of replacement new-build gentrification. Though this estate was not complete within the 2010-2018 period, Google Street View showed it was mostly completed by 2019. Further changes were observed like additional school buildings and a new football pitch. **Residential Gentrification.**

E01007757

2011. Image 2009

2014

2019. image 2016

Gentrification started in 2011 in the northeast of Maltby. By the peak in 2014, a building has been demolished – possibly part of the Maltby Leisure Centre - in the south of the LSOA. However again, the latest image in 2016, and so the last three years cannot be visually validated via Google Earth. The area's IMD ranking for this period increased (less deprivation) by 833, less than the 1000 criteria, however since this change is alongside observed changes, we can presume that neighbourhood changes have occurred, and gentrification is identified. **Residential Gentrification**.

E01007687

2014

2016. image 2015

2017

Gentrification in the north of Wickersley starts in 2014. We quickly see some new build housing being built to the west of the LSOA in 2015. By the end of this cycle in 2017, this development had been completed, and another to the east of the first development, is in construction, and completed in 2018. **Residential Gentrification.**

E01007690	 2014	 2017	 2019. image 2020
	Gentrification in the south of Wickersley starts in 2014, by the peak in 2017, a new build housing estate is underway at the northwest of the LSOA. There are also some small new build developments to the southeast of the LSOA. This cycle ends in 2019, but the image from 2020 shows yet another new build development in construction to the north of the eastern crop of the LSOA. **Residential Gentrification.**		
E01007705	 2010. Image 2009	2014. no image available	 2015
	No clearly visible changes from the aerial imagery for the gentrification cycle captured in the west of Brinsworth. But we can see outside the LSOA the completion of a new build housing development within the east of Brinsworth. Furthermore, during this period, the area's IMD ranking increased (less deprivation) by 3256, so we can presume that neighbourhood changes occurred and gentrification was identified. This therefore supports endogenous gentrification. **Residential Gentrification.**		

E01007792

2010, image 2008

2014

2015

Gentrification started in Dinnington (W) in 2010. In 2010 (the start, not pictured), some land is being prepared for infill housing, in which 2 of the 3 houses are built by the peak in 2014. The infill housing is completed by the end of the identified cycle in 2015. But, since this is the only new-build development, this locality is being assigned general gentrification, as per the criteria in Table A1. **Residential Gentrification**.

E01007659

2010

2014

2015

No visible changes observed from imagery in North Anston. The IMD ranking increased (less deprivation) by 1875 between 2010-2015, so we can presume that neighbourhood changes did occur and gentrification was identified. **Residential Gentrification.**

E01007786	 2010	 2018. Image 2015	2019. no image available
	Gentrification starts in the south of Dinnington/nort of North Anston in 2010. By 2015 we can see the development of a large supermarket to the north of the LSOA, in location where an older supermarket had been demolished. However, the latest image in 2015, and so the last 4 years cannot be visually validated via Google Earth. However, we can presume that commerce contributed to the gentrification. **Residential Gentrification.**		
E01007792	 2010. Image 2008	 2014	 2015
	Few visible changes from the aerial imagery for the gentrification cycle captured in the west of Dinnington, but include an infill development of 3 houses. This LSOA is located within a small town in a rural area, so **Residential Gentrification.**		

E01007747

2010

2013. no image available

2014

Gentrification starts in Kiveton Park in 2010. There are no visible changes within these areial images, however, just beneath the LSOA, a housing estate was developed, which the process began in 2014. Though Google Street View was used, this showed changes to the Kiveson highstreet, where some local stores were replaced with known brand stores (Go Local), a facet of commercial gentrification. **Transport Gentrification.**

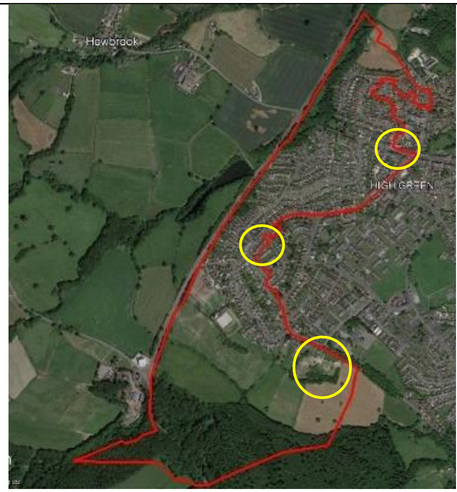
E01008146

2010. Image 2009

2013. no image available

2015

Gentrification started in Deepcar in 2010. By the end of the cycle in 2015, there is some development just on the periphery of the LSOA. When going into Google Street View, we can see that an old derelict building is being restored and redeveloped into new housing development. There is also evidence of residential upgrading, with the extension of multiple houses. Due to its rural-fringe location, it is assigned **Rural Gentrification.**

<table>
<tr><td rowspan="2">E01007899</td><td>
2015</td><td>
2017</td><td>
2019. Image 2020</td></tr>
<tr><td colspan="3">Gentrification in Highgreen started in 2015 and captured the nursery at the south of the residential area being disolved. Some infill new build housing has also been built toward the northeast, and centre of the LSOA. Though there are new-build developments, at this scale they alone are unlikely to have faceted gentrification. Residential Gentrification.</td></tr>
<tr><td rowspan="2">E01008137</td><td>
2010. Image 2009</td><td>
2015</td><td>
2016</td></tr>
<tr><td colspan="3">There is evidence of residential upgrading, via house extensions, and an infil development of a bungalow. This is the northermost part of the urban conurbation, so although the rural-fringe setting, it is assigned. Residenital Gentrification.</td></tr>
</table>

<table>
<tr><td rowspan="2">E01008056</td><td>

2013. Image 2009</td><td>

2017</td><td>

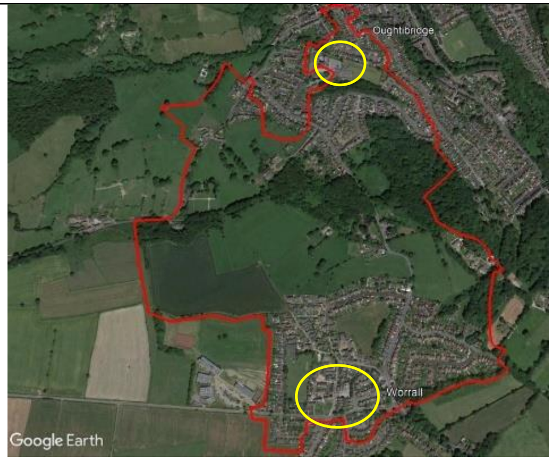
2018</td></tr>
<tr><td colspan="3">There is an infil development of a luxury semi-detached housing in Parson Cross, and residential upgrading via house extension. for the duration of its gentrification cycle. We can see a new build development being built just outside the LSOA, to the northwest.This may also therefore support endogenous gentrification. Residential Gentrification.</td></tr>
<tr><td rowspan="2">E01007856</td><td>

2012. Image 2009</td><td>

2017</td><td>

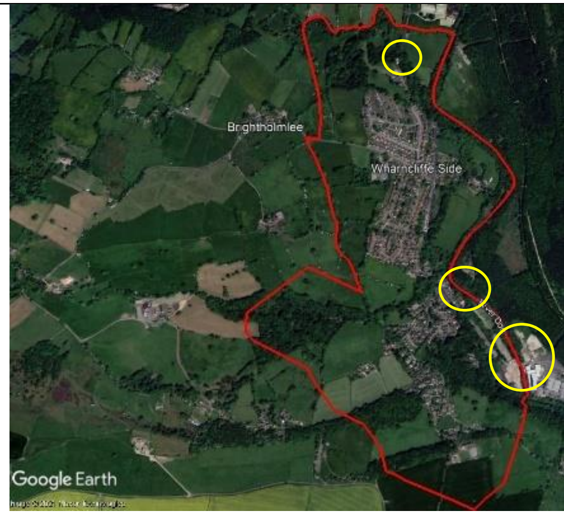
2018</td></tr>
<tr><td colspan="3">There are no clearly visible changes from this aerial imagery for Shiregreen, so the IMD was consulted. The area’s IMD ranking decreased (more deprivation) by over 2000. So we can assume that due to the lack of visible changes and the worsening IMD, that no gentrification occurred here. Transport Gentrification.</td></tr>
</table>

<table>
<tr>
<td rowspan="2">E01008126</td>
<td>

2010. Image 2009</td>
<td>

2017</td>
<td>

2019. Image 2020</td>
</tr>
<tr>
<td colspan="3">Gentrification starts in 2010. The neighbourhood to the south of the LSOA, Worral, has a new build development that has been completed by the peak in 2017. A school building has also been built within the neighbourhood to the north, Oughtibridge. Rural Gentrification.</td>
</tr>
<tr>
<td rowspan="2">E01008136</td>
<td>

2010. Image 2009</td>
<td>2015. no image available</td>
<td>

2016</td>
</tr>
<tr>
<td colspan="3">Gentrification starts in 2010, but the image from 2009 shows an industrial estate to the east of the LSOA, beside the River Don. By the end of this cycle in 2016, it has been demolished and the land is being prepared for development. A new build housing estate began development here in 2020, post the gentrification cycle and the temporal boundary of the data. Infill housing is observed further north in two separate locations. Rural Gentrification.</td>
</tr>
</table>

E01008085

2010. Image 2009

2018

2019. Image 2020

Gentrification starts in Birley Edge in 2010, and has one of the longest cycles. We can see by the peak in 2018 that the old school in the bottom of the LSOA has been demolished and replaced with a new school located just above it. There is also new brownfield land in the north. **Residential Gentrification.**

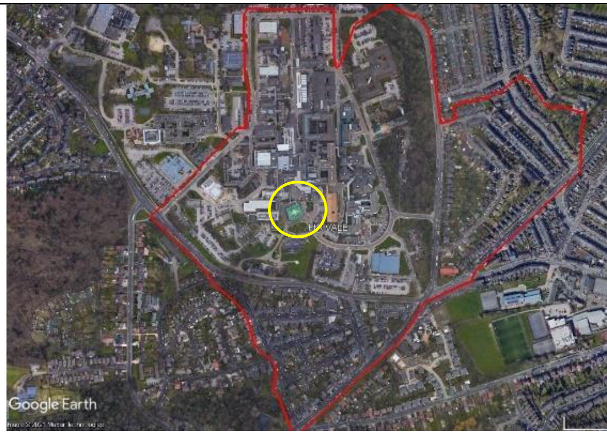
E01008117

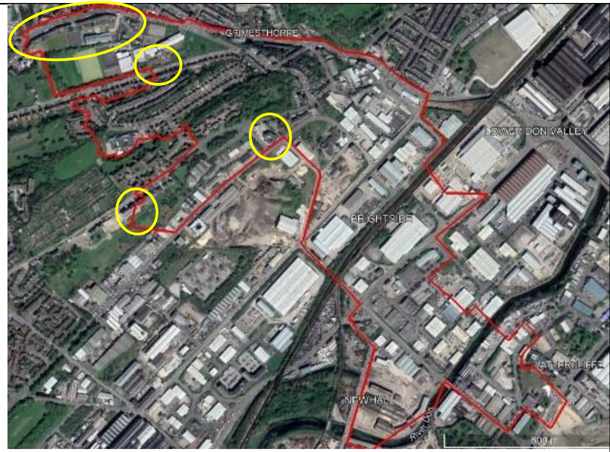
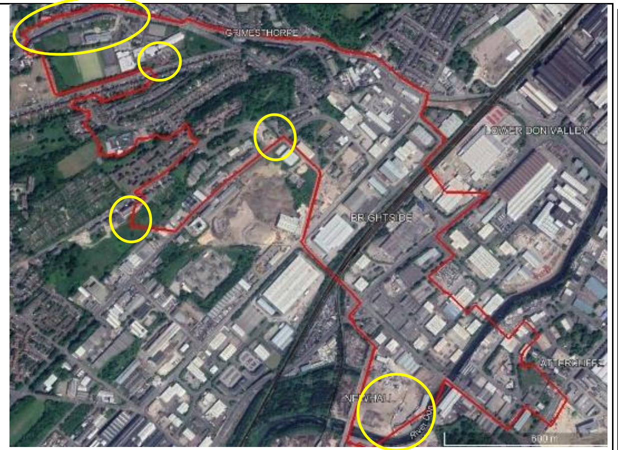
2011. Image 2009

2014. Image 2015

2017

Gentrification started in Parson Cross in 2011. Around the peak in 2015, a new build housing estate was being built in the top corner of the LSOA, upon brownfield land that once had an industrial estate. The estate was not completed until 2020, which does not align with the captured cycle of gentrification. **Residential Gentrification**.

<table>
<tr><td rowspan="3">E01007850</td><td></td><td>2014. no image available</td><td></td></tr>
<tr><td>2011</td><td></td><td>2015</td></tr>
<tr><td colspan="3">There is only one obserable change within these aerial images, the demolition of an industrial building to the very south of the LSOA. However, this change doesn't relate to gentrification, so the IMD was consulted. This area's IMD ranking decreased (more deprivation), by more than 1000. Thus, we can assume that no gentrification ocurred here. Transport Gentrification.</td></tr>
<tr><td rowspan="3">E01007942</td><td></td><td></td><td></td></tr>
<tr><td>2014. Image 2015</td><td>2018</td><td>2019. Image 2020</td></tr>
<tr><td colspan="3">Gentrification starts in Fir Vale in 2015, by the peak in 2018 a new hospital building had been built with a helepad, seen right in the centre of the LSOA. A hospital car park was also built further north. This area has many different hospital buildings and so gentrification here may be driven by the number of professionals working and residing here. This area is highly connected. Transport Gentrification.</td></tr>
</table>

<table>
<tr>
<td rowspan="3">E01007877</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2010. Image 2009</td>
<td>2014. Image 2015</td>
<td>2016</td>
</tr>
<tr>
<td colspan="3">Gentrification starts in Grimesthorpe and Brightside industrial estate in 2010. An entire street was demolished and replaced with a new school, which also saw the addition of a new building. Further developments include a mosque, a nursing home on the periphery, and the demolition of a large industrial unit. These are potentially driven by the connectivity of the area. Transport Gentrification.</td>
</tr>
<tr>
<td rowspan="3">E01008151</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2011. Image 2009</td>
<td>2017</td>
<td>2019. Image 2020</td>
</tr>
<tr>
<td colspan="3">Gentrification started in Walkley/Steel Bank in 2011. A number of infill housing developments are observed to the eastern outcrop, and to the south, and a commercial building was demolished in 2018, and replaced with residential lets further south. A commercial building was also demolished and replaced. Residential Gentrification.</td>
</tr>
</table>

E01008064

2015

2017

2018

Gentrification started in Crookes in 2015. By the peak in 2017, an old building had been demolished and replaced with new infill housing, there are no other observable changes. Though this area has student population, this part of the village is not as highly connected as the rest. **Residential Gentrification.**

E01007960

2010

2017

2018.Imagery 2020.

Gentrification started at the west of Crookes in 2010. Though there are no clearly visible changes assocatiated with gentrification within these aerial images. This area's IMD ranking decreased (less deprivation) by over 2000, thus we can assume that this area has experienced some level of gentrification. Again, due to the large student population is likely to be studentification. **Transport Gentrification.**

E01007962

2011. Image 2010

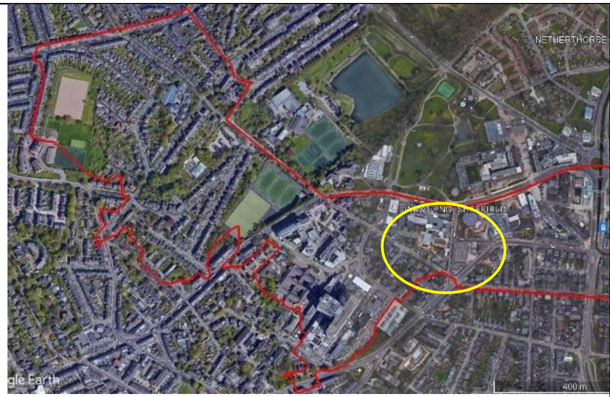
2017

2018

Gentrification started in the south of Crookes in 2011. Though there are limited changes within the LSOA, including a small infill housing development, and extension to a church. A new build housing estate has been developed on an old industrial site, just south of the LSOA. This could support endogenous gentrification. But since this area has a large student population and has high levels of conenctivity, it is assigned **Transport Gentrification.**

E01007866

2015

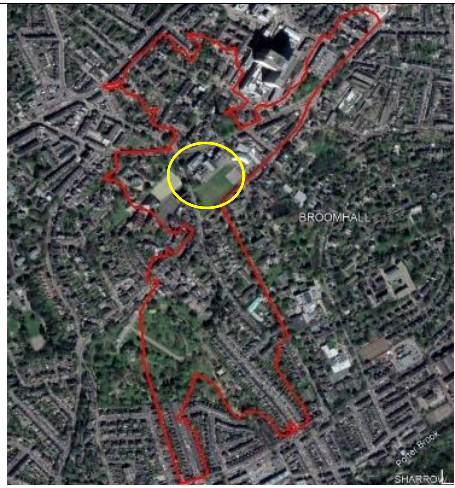
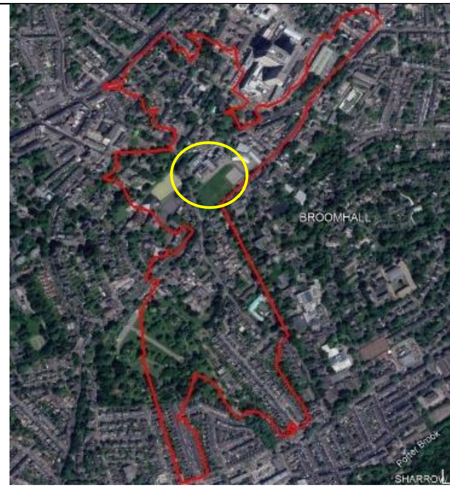


2018

2019. Iamge 2020

By the peak of gentrification in West End in 2018, a building had been demolished and replaced with a new multi-story hospital car park, another multi-story hospital car park and hospital building had also been built. Due to the presence of hospitals, gentrification may have been driven by the incoming professionals working at them. By 2019, a new hospital building was in construction. **Transport Gentrification.**

<table>
<tr><td rowspan="2">E01007868</td><td>
2011. Image 2010</td><td>2013. no image available</td><td>
2014. Image 2015</td></tr>
<tr><td colspan="3">Gentrification started in Broomhill in 2011, the largest visible change is the demolition of a university hall of residence, and the preperation of the land ready for the new build housing estate that is built post this captured gentrification cycle. Broomhill has a large student population, and is a relatively well connected neighbourhood, thus studentification. Transport Gentrification.</td></tr>
<tr><td rowspan="2">E01007954</td><td>
2012. Image 2010</td><td>2014. no image available</td><td>
2015</td></tr>
<tr><td colspan="3">There are few clearly visible changes in Crosspool for the duration of the gentrification cycle. One includes the (mid) construction of an infill residential development on Darwin Lane, there is also some residential upgrading (x2) to the west of the LSOA. Gentrification in Crosspool is therefore likely new-build gentrification. Residential Gentrification.</td></tr>
</table>

E01007860	 2010	 2014. Image 2015	 2016
	Gentrification started in Broomfield in 2010, and we can see in the upper half of the LSOA a large temporary building whilst a new school building is erected, and is completed by 2015. **Residential Gentrification.**		
E01007863	 2012. Image 2011	2014. no image available	 2015
	Gentrification was captured in Endcliffe from 2012 to 2015. There is a visible change in 2015 to the east of the LSOA, the development of a luxury apartment block that is completed in 2017. However, this LSOA has a very large student community, and so it could be expected that gentrification relates to the students – studentification. **Transport Gentrification.**		

<table>
<tr><td rowspan="2">E01007864</td><td>
2011</td><td>2014, no image available</td><td>
2015</td></tr>
<tr><td colspan="3">Gentrification started in Sharrow in 2011, however there are limited visible changes within these aerial images; the extension of a car park. This area's IMD ranking increased (less deprivation) by 3841, so we can assume that residential changes have occurred and gentrification was correctly identified. Sharrow is a neighbourhood with a large student population, and so gentrification could again be associated with students. Transport Gentrification.</td></tr>
<tr><td rowspan="2">E010332737</td><td>
2011</td><td>
2018</td><td>
2019. Image 2020</td></tr>
<tr><td colspan="3">Gentrification started in 2011, by the peak in 2018 we can see the development of a purpose built student accommodation (PBSA) – Royal Riverside. Such PBSA are facets of studentification. Transport Gentrification.</td></tr>
</table>

E01008097

2013. Image 2011

2018

2019. Image 2020

Gentrification started in this LSOA that encompasses Arbourthorne and Manor Top in 2013.By the peak in 2018 we can see the preperation of brownfield land for a new build housing estate, that we can see is only half completed by 2020. New build gentrification – brownfield. **Residential Gentrification.**

E01007935

2010

2014. Image 2015

2016

There are no easily visible changes within Endcliffe/Huters Bar, other than the update to the outdoor courts at the sports centre. This area's IMD ranking increased (less deprivation) by 3611, so we can assume that residential changes have occurred and gentrification was correctly identified. This is a neighbourhood with a large student population, and so gentrification assocaited with these students would be studentification. **Transport Gentrification.**

<table>
<tr><td rowspan="2">E01007936</td><td>
2011</td><td>
2018</td><td>
2019. Image 2020</td></tr>
<tr><td colspan="3">Gentrification started in Banner Cross in 2011.Observable changes include the development of some new build infill houses towards the southwest, a new school building to the northwest, and. A new mixed-use development was also build to the east. Residential Gentrification.</td></tr>
<tr><td rowspan="2">E01007937</td><td>
2016</td><td>
2018</td><td>
2019. Image 2020</td></tr>
<tr><td colspan="3">Gentrification started in Greystones in 2016. Two small infill developments are observed, to the norh and south of the LSOA. There are few other observable changes. Residential Gentrification.</td></tr>
</table>

E01008051	 2016	 2018	 2019. Image 2020
	Gentrification started in Nether Edge (East) in 2016, though there arent any visible changes within these aerial images. **Residential Gentrification.**		
E01008052	 2015	 2018	 2019. Image 2020
	There are no observed changes within these aerial images of Nether Edge (West). Additionally, data primitive and change vectors suggseted only limited changes associated with gentrification, thus this area is assigned as **None.**		

<table>
<tr><td rowspan="3">E01008099</td><td>
2011</td><td>
2018</td><td>
2019. Image 2020</td></tr>
<tr><td colspan="3">Gentrification started in Lowfield in 2011, but there are not many changes. An empty building to the south has been renovated, including the land behind it, and an industrial unit in the north has been demolished, but the land is still in use. Transport Gentrification.</td></tr>
<tr><td rowspan="3">E01007985</td><td>
2010</td><td>2014. no image available</td><td>
2015</td></tr>
<tr><td colspan="3">Gentrification in Norton Lees started in 2010. During this period, there have not been any changes assocaited with gentrification (work on the national grid plot to the south). This area’s IMD ranking increased (less deprivation) by 2650 between 2010-2015, thus we can assume that there have been changes within the neighbourhood, and gentrification did take place. Residential Gentrification</td></tr>
</table>

E01008002

2010

2017

2018

Gentrification started in Gleadless in 2010. This neighbourood runs parallel to the tram lines and has easy access to several tram stops, giving easy access to the city centre, and other transport links. This area could therefore be associated with transit induced gentrification – tram. An old pub was demolished, and replaced with a block of flats, and another block of semi-detached residences were built beside it. **Transport Gentrification.**

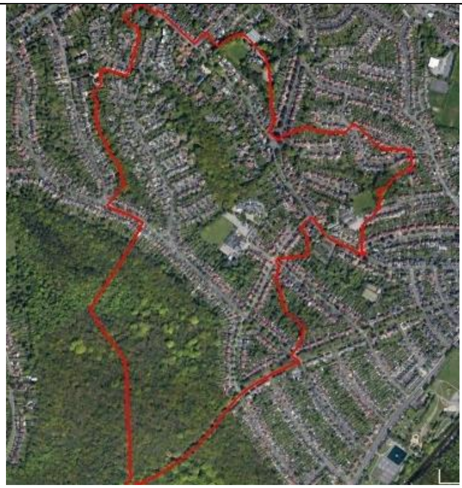
E01007927

2011

2014. no image available

2015

Gentrification started in Parkhead in 2011, by the end of the cycle, a few residential properties had been either replaced by a new build property, or majorly upgraded, with no other visible changes. **Residential Gentrification.**

<table>
<tr><td rowspan="2">E01007930</td><td>
2011</td><td>
2015</td><td>
2016</td></tr>
<tr><td colspan="3">Millhouses experienced gentrification from 2011 through to 2016. One change includes the demolition of a house which is replaced with a new, larger residential building. Millhouses is one of the least deprived LSOAs in England, and its IMD ranking between 2010-2015 increased from 31772 to 32333. Therefore we can assume that gentrification within this locality did occur. Residential Gentrification.</td></tr>
<tr><td rowspan="2">E01007828</td><td>

2011</td><td>2014. no image avaialble</td><td>

2015</td></tr>
<tr><td colspan="3">Gentrification started in Woodseats (North) in 2011. By the end of the captured process in 2015, we can see the demolition of an old social club in the top right, which was replaced by a new build appartment block. This was completed at the end of 2015, after the last Google Earth image. New build gentrification - brownfield. Residential Gentrification.</td></tr>
</table>

E01007831

2011

2014. no image available

2015

Gentrification started in Woodseats (South) in 2011. There are no visible changes within the aerial images. **Residential Gentrification.**

E01008033

2010

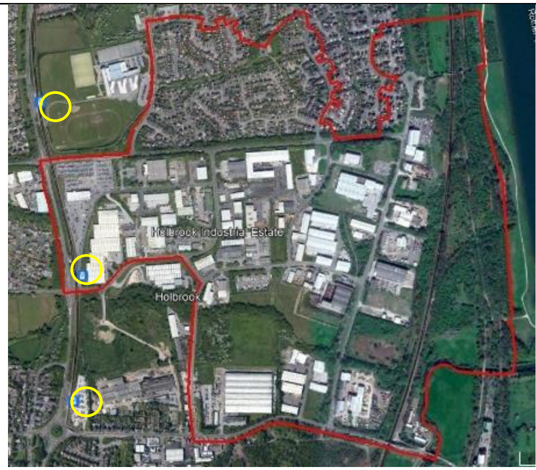
2017

2018

Gentrification started in Hackenthorpe (East) in 2010. There are few observable changes including the development of a petrol station for the asda supermarket, and something in the midst of construction to the north. There has also been some residential upgrading to the west of the LSOA. Thus, with these changes we can therefore assume that gentrification has occurred, and that it is likely general gentrification, or transit induced gentrification, due to the tram route running through the neighbourhood, and the three tram stops within close proximity. **Transport Gentrification.**

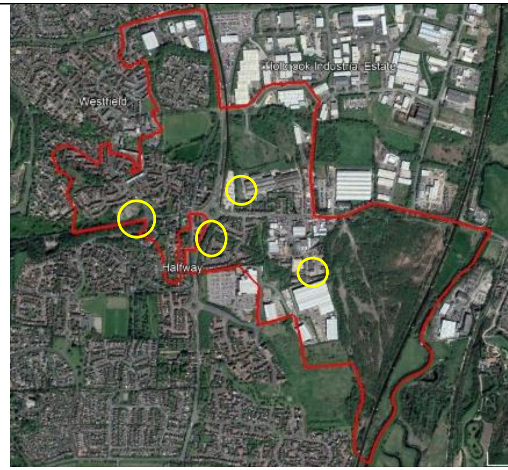
<table>
<tr>
<td rowspan="3">E01008038</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2010</td>
<td>2018</td>
<td>2019. Image 2020</td>
</tr>
<tr>
<td colspan="3">Gentrification started in Hackenthorpe (West) in 2010. There are slight changes observed within this LSOA, regarding the residential upgrading of a few houses throughout. This could suggest the in movement of wealthier residents who have the necessary funds for such developments. Though gentrification here is likely to be transit induced gentrification due to the access to the trams. Transport Gentrification.</td>
</tr>
<tr>
<td rowspan="3">E01007835</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2012. Iamge 2011</td>
<td>2017</td>
<td>2019. Iamge 2020</td>
</tr>
<tr>
<td colspan="3">Gentrification started in Charnock in 2012. There aren't any visible changes other than the land colour changes. Though this area's IMD ranking increased (less deprivation) by 2140 between 2010-2019, so we can assume that some residential changes have occurred, and gentrification was correctly identified. It is also within close proximity to a tram stop, thus it could also be Transport Gentrification.</td>
</tr>
</table>

<table>
<tr>
<td rowspan="2">E01008041</td>
<td>

2010</td>
<td>

2014</td>
<td>

2015</td>
</tr>
<tr>
<td colspan="3">Gentrification around the Holbrook Industrial Estate/Sothall started in 2010. IMD ranking increased (less deprivation) by 2174, so some residential changes have occurred and gentrification was identified. It has good access to the tram stops to the west, so potentially Transport Gentrification.</td>
</tr>
<tr>
<td rowspan="2">E01008019</td>
<td>

2011</td>
<td>2013. no image available</td>
<td>

2015</td>
</tr>
<tr>
<td colspan="3">Gentrification started in Halfway (North) and the bottom half of the Holbrook Industrial Estate in 2011. Visible changes include the demolition of a pub to the west, the preperation of some land for new-build housing after the end of this captured gentrification cycle, and the demolition of an industrial unit further east. However, it does have a tramstop in the centre (north) of the LSOA, and so it could be associated with Transport Gentrification.</td>
</tr>
</table>

<table>
<tr><td rowspan="2">E01008020</td><td>
2011</td><td>
2016</td><td>
2019</td></tr>
<tr><td colspan="3">Gentrification started in Waterthorpe/Westhorpe in 2011, though there are no changs within these aerial images for the duration of the captured gentrification cycle. However this area’s IMD ranking increased (less deprivation) by 1899 between 2010-2019, so we can assume that some residential changes have occurred and gentrification was correctly identified. Transport Gentrification.</td></tr>
<tr><td rowspan="2">E01008037</td><td>
2010</td><td>
2013. Image 2015</td><td>
2019. Image 2020</td></tr>
<tr><td colspan="3">Gentrification started in 2010. This LSOA encompasses half of the Crystal Peaks shopping centre, and several tram stops along the tram route that passes through,and also just outside the LSOA boundary . There are no visible changes throughout, but the IMD shows that the area’s ranking decreased (more deprivation) by 1930, so we can assume that no gentrification occurred within this period within this locality. None.</td></tr>
</table>

E01008039	 2011	2014. no image available	 2015
	Gentrification starts in Waterthorpe in 2011. There are no visible changes in these aerial images of gentrification captured in Waterthorpe, but the IMD shows that the area's ranking increased (less deprivation) by 4037, so we can assume that residential changes have occurred and gentrification was identified. This LSOA has a couple of tram stops with easy access into the city centre. **Transport Gentrification.**		
E01007824	 2011	 2016	 2017. Image 2018.
	Gentrification starts in Greenhill in 2011, but there are no visible aerial changes. The area's IMD ranking increased (less deprivation) by 1432 between 2010-2019, so we can assume that residential changes have occurred and gentrification was correctly identified. This LSOA is on the rural-urban fringe so is assigned as **Rural Gentrification.**		

E01007826

2011

2016

2017

This LSOA encompasses the majority of the large golf course, and Greenhill village. There are a couple changes including two new infill houses to the south, and a residential upgrade to the east. Potentially rural gentrification due to rural-fringe location. It also sits on the periphery of the Dore and Totle railstation, thus also potentially rail-induced gentrification. **Rural Gentrification.**

E01007917

2010

2014. Image 2013

2016

Gentrification started in Lower Bradway in 2010. Changes include the demolition of a residential building to the east, which was replaced in 2017, post the gentrification cycle, and the (mid construction) residential upgrade in the northwest. This LSOA also encompasses part of a golf course and lots of greenspace in a semi rural setting. Gentrification could therefore be associated with rural or green gentrification. The Dore and Totley railstation is just north the tip of the western tip of the LSOA, so could also be transit-induced gentrification. **Transport Gentrificaton.**

E01007922

2012. Image 2011

2018

2019. Image 2020

Gentrification started in Bradway in 2012. A new infill housing development was completed by 2020, and a new care home is in construction outside the east of the LSOA.. Bradway is a relatively highly connected area, thus **Transport Gentrification.**

E01008024

2010

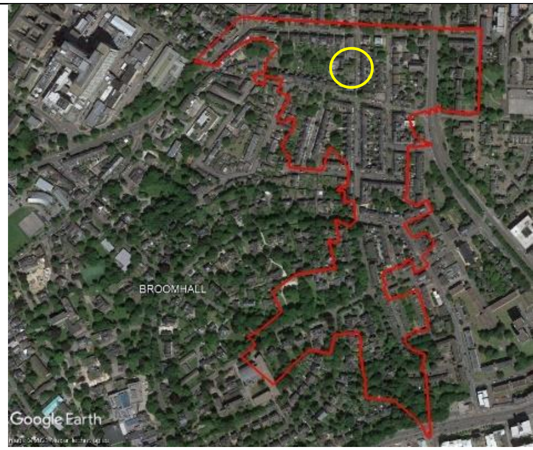
2013. no image available

2014. Image 2015

Gentrification in Mosborough started in 2010. There are no observable changes in these images. However, this is a large LSOA with a large proportion of greenspace in a rural-urban fringe setting. Gentrification may therefore be associated with rural or green gentrifcation. **Rural Gentrification.**

E01008115

2010



2018

2019. Image 2020

Gentrification started in Broomhall in 2010, through to 2019. In the north we can see infill development of four new buildings, but not entirely sure of their use. There is also a development mid construction in the south. Broomhall does however have a large student population, therefore gentrification here may be associated with students – studentification. **Transport Gentrification**.

E01007878

2012. Image 2010

2018

2019. Image 2020

Gentrification started in Wybourn in 2012. By the end of the cycle (image 2020), a new large new-build housing estate is in the midst of construction. However, since the cycle started much earlier, the new-build housing developments outside of the LSOA to the north may have impacted this neighbouring LSOA, supporting endogenous gentrification. New-build gentrification. **Residential Gentrification.**

E01007879

2012

2018

2019. Image 2020

Gentrification started in 2012. A couple of residential upgrades are observed within the LSOA. It is also located just east of Sheffield train station, which also has a tram platform too, thus gentrification could be transit(rail)-induced. **Transport Gentrification.**

E01007882

2010

2017

2018

Gentrification in Granville/Manor Top started in 2010. The Manor Lodge Primary School to the northeast has extended with a couple of additional buildings. There are no other changes within this period.Because this was the only observable change, the IMD was consulted. This showed that this area's IMD ranking between 2010-2015 increased (less deprivation) by 812, we can therefore presume that neighbourhood change did occur and gentrification was identified. **Residential Gentrification.**

Appendix C: Validating Identified Gentrification with Google Street View

Appendix B: Validating 123 Cycles of Gentrification: Google Street View

Table B1: The Google Street View validation of selected neighbourhoods' gentrification in South Yorkshire

<table>
<tr><th></th><th>Gentrification Start Year</th><th>Gentrification End Year</th></tr>
<tr><td rowspan="2">E01007860</td><td>
2011</td><td>
2016</td></tr>
<tr><td colspan="2">A single story timber office building was demolished and a luxury student apartments was built in its place.</td></tr>
<tr><td rowspan="2">E01007863</td><td>
2012</td><td>

2016</td></tr>
<tr><td colspan="2">A large, detached residence was turned into a bar a restaurant, a face of studentification in a student neighbourhood.</td></tr>
</table>

E01007935	 2010	 2016. Imagery 2014.
	This street consists of terraced housing, many of which are HMO student lets. During the cycle, the housing exteriors have deteriorated, and some houses for sale. This is a facet of studentification.	
E01007362	 2010. Imagery 2009.	 2017. Imagery 2016.
	The resturant and garage have been	

E01007327

2015. Imagery 2016.

2019. Imagery 2017.

During this cycle, the industrial estate in the background has been demolished. A further image from 2021 (outside of the captured cycle from data ending in 2019), shows the start of a new housing development. Thus, Brownfield new-build genrification.

E01007405

2014.

2017.

Throughout this cycle, several vacant commercial properties now house new shops. Some shop fronts have also been modernised, with a cleaner look.

E01007387

2010.

2015. Imagery 2016.

The empty plot has been bough and the land developed into shops and apartments.

E01007418

2015. Imagery 2016.

2019. Imagery 2020.

The previously dated looking resturant has updated its singage to a more modern style, and has also updated the windows to reflect the current trend of grey housing exteriors.

E01007419

2011.

2015. Imagery 2016.

The previously deteriorated pub exterioir has been renovated, a more modern and fresher design.

E01007421

2014. Imagery 2011.

2019. Imagery 2020.

The pub has undergone renovations, modernised signage, outdoor seating and seating under a gazebo. The outbuilding has also been demolished.

E01007488

2010. Imagery 2009.

2018. Imagery 2021.

The barn has been demolished and replaced with a brand new detached house with a private gated access.

E01007626

2010. Imagery 2009.

2014. Imagery 2016.

A full housing estate has been developed where the empty land once was (in the background of the first image). The shop on the right has a newer owner and a fresher look, and the previously vacant building to the left is now an indian takeaway.

E01007555

2010. Imagery 2009.

2018. Imagery 2016.

Vacant commercial plots in a row of shops in hatfield. This commercial street has undergone improvements. Vacant plots have been filled, businesses have been changed, more modern signages.

E01007562

2010.

2018. Imagery 2011.

Early on in the gentrification cycle, infil housing was being developed in the plot of what was previously a garden.

E01007621

2012.

2016.

A plot of land near an old church for sale, prior to the start of gentrification, in 2011. Throughout the gentrification cycle, a new housing development was being built, but it was not complete until 2021, so only first half of the cycle was captured.

E01007623

2011. Imagery 2009

2017. Imagery 2012.

Presence of a block of flats (in 2009 prior to the gentrification change). The block of flats have been demolished, and a new community church hall has been built in its place.

In the same LSOA, a church also exists behind this block of flats, which was demolished by the start of the gentrification period, and replaced by a an assisted living residence.

E01007465

2012. Imagery 2009.

2019. Imagery 2018.

During this gentrification cycle, the brownfield site has been developed into a large care home with public services including a café, beauty salon, and cinem, which was completed in 2014, half way through the cycle.

<table>
<tr>
<td rowspan="2">E01007506</td>
<td>

2011.</td>
<td>

2014. Imagery 2012.</td>
</tr>
<tr>
<td colspan="2">Throughout this gentrification cycle, the empty field was prepared for a new-build housing estate, which was completed in 2017, three years after the end of the captured cycle.</td>
</tr>
</table>