Applied Spatial Analysis and Policy (2023) 16:897–921
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 / Published online: 8 March 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

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

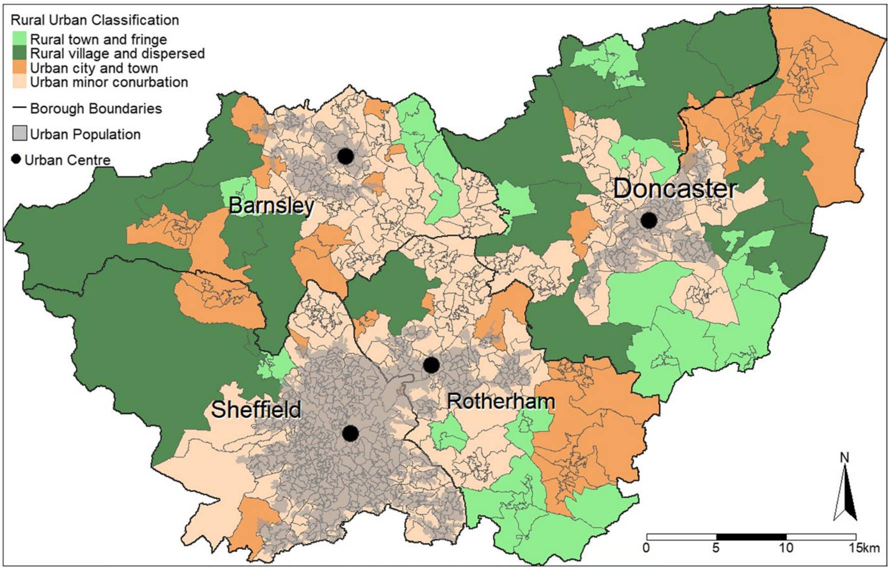

Fig. 1 A map of LSOAs in the study area shaded by the 2011 Rural Urban Classification (Bibby & Shepherd, 2004), the four boroughs (local authorities) and their urban centres

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

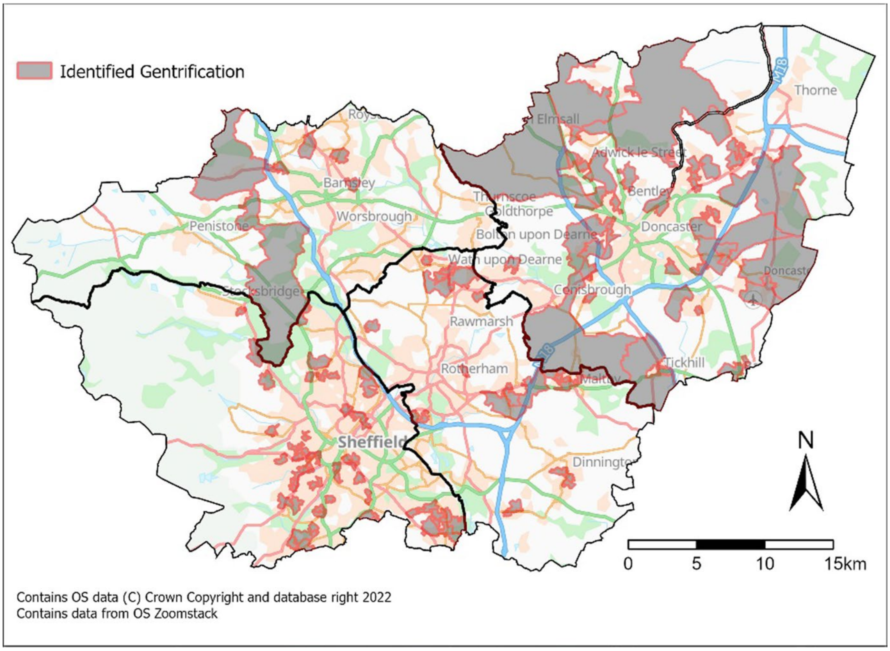

Fig. 2 The locations of neighbourhoods identified as having gentrified in the study area

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

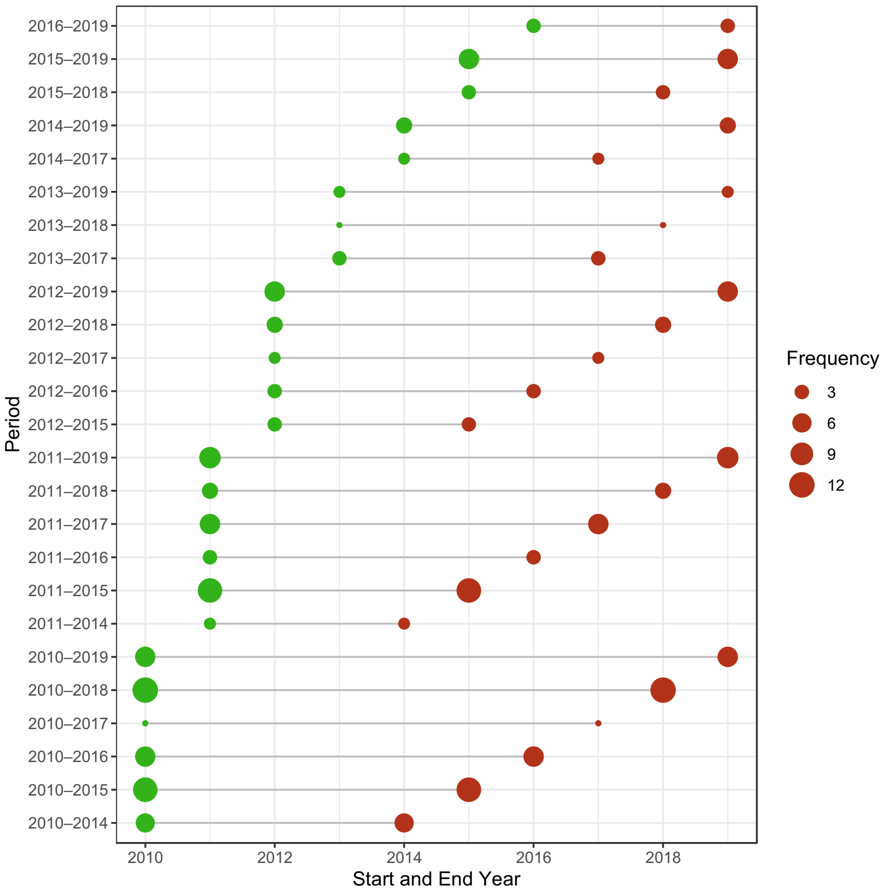

Fig. 3 The frequency of the start, end and duration of gentrification in the 123 LSOA neighbourhoods identified as having gentrified

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

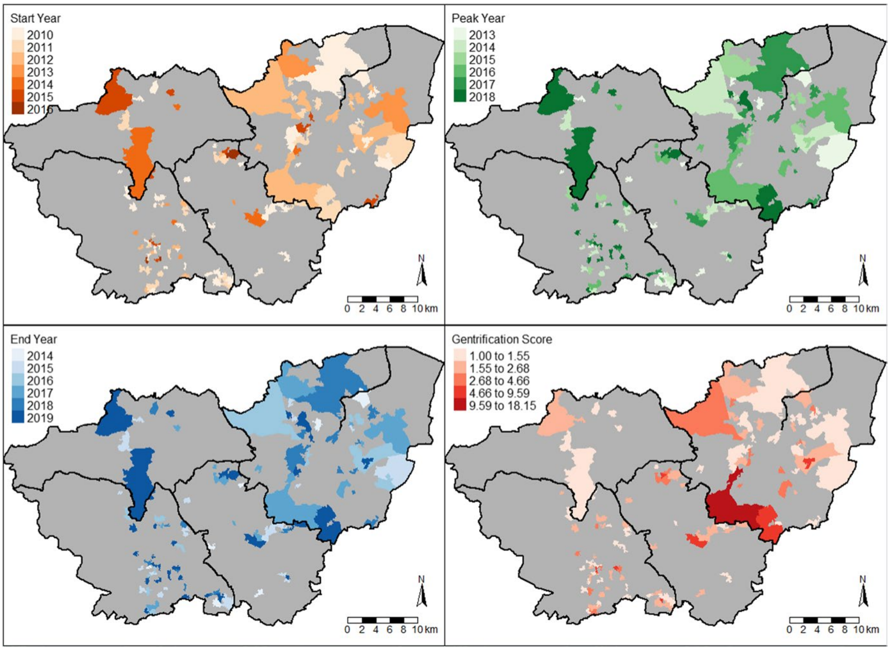

Fig. 4 The start, end, and peak year of gentrification, and the cumulative gentrification score for the LSOAs identified as having gentrified

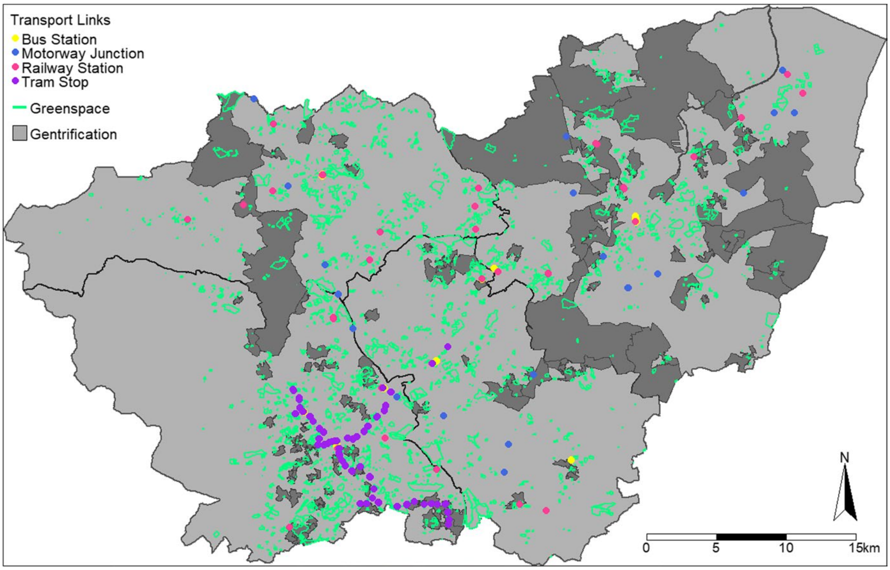

Fig. 5 Context for the case study areas: greenspaces and transit links and areas identified as having gentrified

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

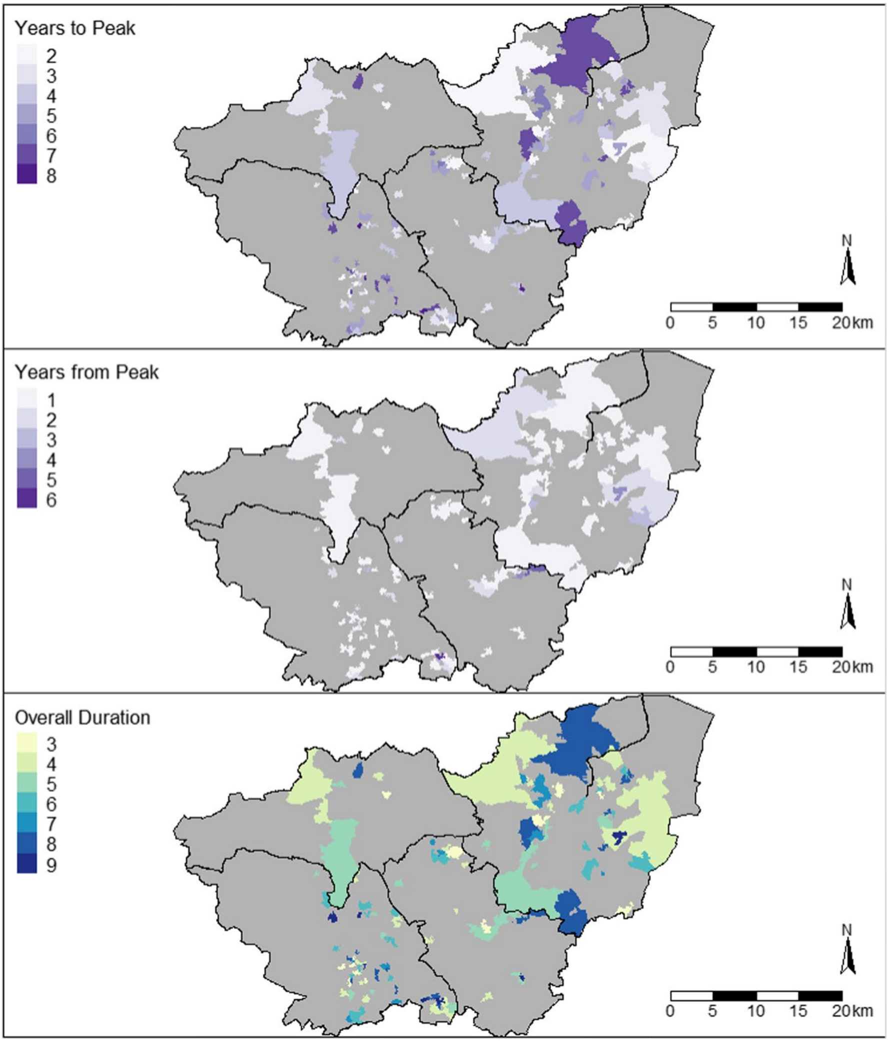

Fig. 6 The number of years from gentrification start to peak, number of years from peak to gentrification end, and gentrification duration

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	Gentri-fication 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

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

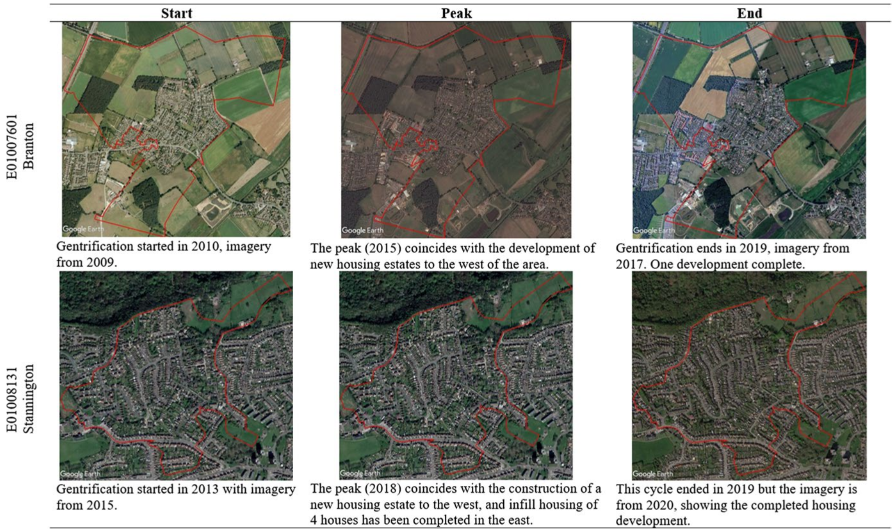

Fig. 8 Google Earth imagery of example LSOA neighbourhoods in Branton, Doncaster and Stannington, Sheffield at the start, peak, and end of their gentrification cycles, showing large residential developments

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

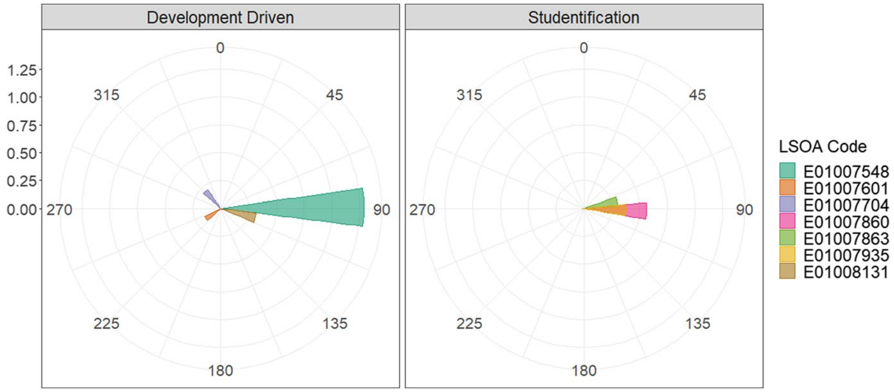

Fig. 11 The angle and magnitude of change for each case study, grouped by Residential Development and Studentification driven gentrification

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.