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

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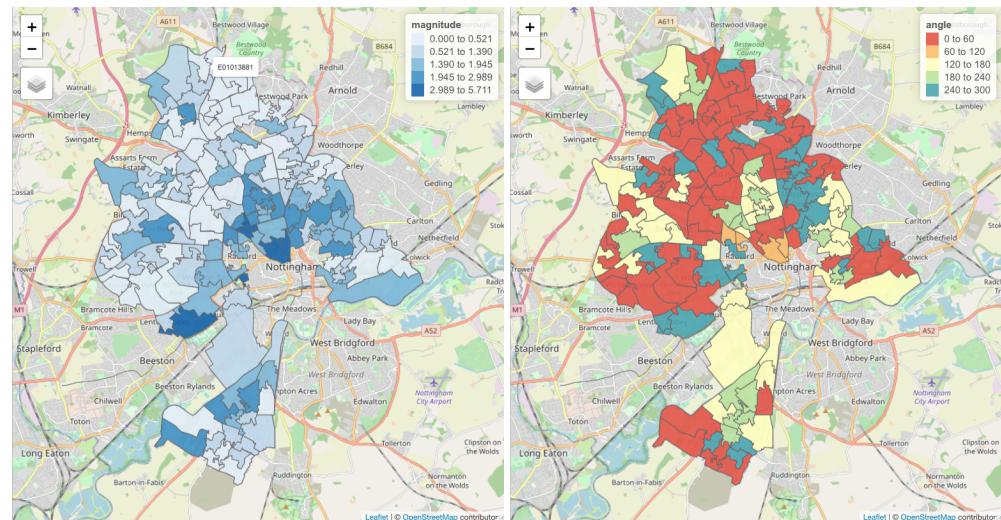
Figure 1. The results of the change vector analysis for LSOAs in the Nottingham area, both with an OpenStreetMap backdrop: (**left**) The magnitude of change. (**right**) The angle or direction of change.

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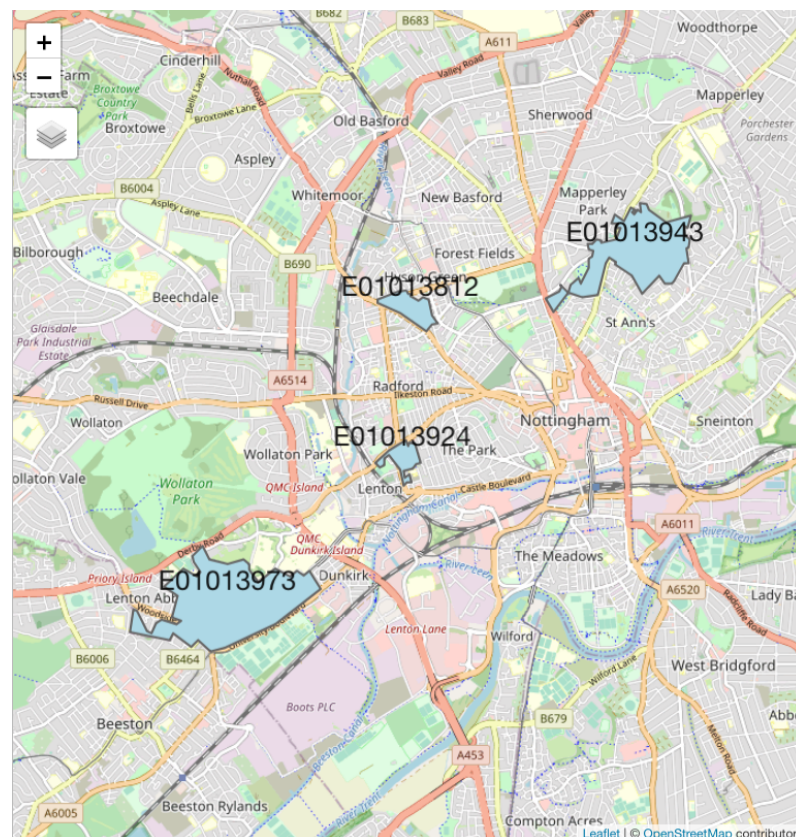
Figure 2. Four example LSOAs in Nottingham, with an OpenStreetMap backdrop.

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.
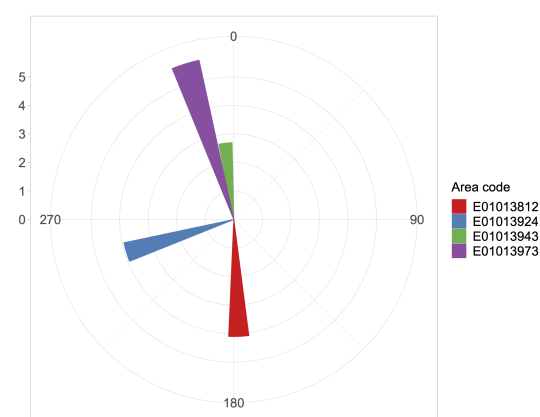
Figure 3. Polar plots of the magnitude and direction of the change vector for the 4 LSOAs in Nottingham.
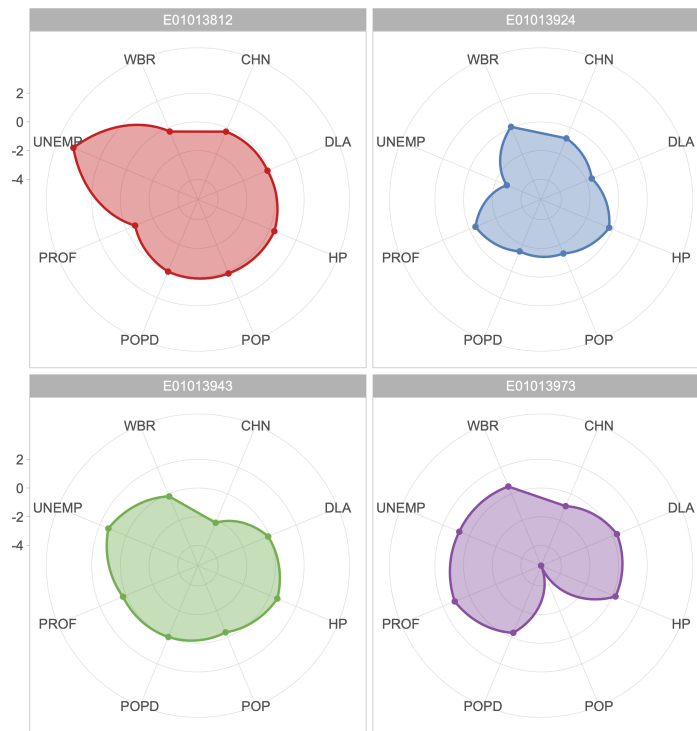
Figure 4. Radar plots of the changes from 2010 to 2016 in each of the 4 Nottingham LSOAs.

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.

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]