Computers, Environment and Urban Systems 125 (2026) 102396

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Review

Geodemographics and residential differentiation: A methodological review and future directions for learned representations of the social landscape

Alex Singleton[a,*], Seth E. Spielman[b]

[a] Department of Geography and Planning, University of Liverpool, UK
[b] Microsoft AI, 1650 Canyon Blvd., CO 80302, Boulder, USA

ARTICLE INFO

Keywords:
Geodemographics
Deep learning
Urban residential structure
Autoencoders
Spatial analysis
Computational geography

ABSTRACT

Residential differentiation reflects the complex patterns by which social groups distribute themselves across urban spaces, fundamentally shaping social, economic, and spatial structures. This paper reviews the methodological development of geodemographic classification, tracing its evolution from early social area analysis and factorial ecology through to contemporary approaches. We critically evaluate this lineage of methods for quantifying residential patterns, and identifying persistent limitations in capturing the non-linear complexities of contemporary urban environments. Building on this review, we explore potential future directions involving learned representations of the social landscape, which may offer alternatives to traditional linear dimensionality reduction techniques. Drawing on recent empirical work applying deep learning to geodemographic classification, we consider how such approaches might address identified limitations while acknowledging that their advantages over established methods remain context-dependent and require further empirical validation. We emphasise that any adoption of these techniques must prioritise transparency and interpretability. The paper concludes by outlining potential directions for future research, including how learned representations might be integrated within existing geodemographic workflows.

1. Introduction

In the fields of urban geography, planning, and sociology, there is a longstanding tradition of developing data-driven classification systems to identify patterns of residential differentiation. Such approaches aim to deepen our understanding of urban structure and those spatial processes that shape its evolution, combining theory with empirical observation. For the purposes of this discussion, we define residential differentiation, as referring simply to the spatial distribution of different demographic groups across urban areas. Residential differentiation emerges from a complex interplay of processes through which distinct groups become concentrated in areas of a city, often reflecting variations in socioeconomic status, ethnicity, race, family composition, lifestyle, or other demographic characteristics (Timms, 1971; van Ham et al., 2021; White, 1987).

While residential differentiation and segregation are closely related concepts, they differ in emphasis and analytical focus. Segregation often carries normative connotations about inequality and exclusion and is often associated with involuntary separation resulting from discrimination or structural constraints (Iceland, 2004; Musterd, 2005). However, such residential clustering also emerges through voluntary processes, as population groups may choose to live in neighbourhoods inhabited by those they perceive as similar; whether for reasons of cultural affinity, social networks, or access to ethnic institutions (Clark, 1992; Peach, 1996). Regardless of whether such patterns arise voluntarily or involuntarily, segregation can lead to negative outcomes including unequal access to resources, services, and opportunities (Williams & Collins, 2001). Research in this area focuses on explaining why certain spatial configurations of population groups exist in urban spaces, examining both the mechanisms of constraint and the role of residential preferences in producing these patterns. Residential differentiation, by contrast, serves as a more neutral descriptor of spatial patterns that focuses on describing the outcome of different groups living in different places without necessarily implying causality or making value judgments about the observed distributions.

For as long as cities have existed, differences in socioeconomic status have manifested geographically. However, as urban areas have grown increasingly complex, so too have the patterns of residential differentiation (Maloutas, 2012; Tammaru et al., 2019). Residential differentiation continues to be shaped by multiple factors including housing

* Corresponding author.
E-mail address: alex.singleton@liverpool.ac.uk (A. Singleton).

https://doi.org/10.1016/j.compenvurbsys.2025.102396
Received 23 June 2025; Received in revised form 16 December 2025; Accepted 18 December 2025
Available online 7 January 2026
0198-9715/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

affordability, employment accessibility, urban planning initiatives, and the legal, market, and social forces that drive segregation, both historical and contemporary. Extensive evidence demonstrates that these influences contribute to diverse outcomes, including the formation of affluent enclaves, economically disadvantaged areas and culturally or ethnically distinctive communities (Johnston et al., 2007; van Ham et al., 2021). Over time, such patterns of residential differentiation can profoundly impact social cohesion, economic vitality and the overall quality of life experienced by urban residents.

Despite advances in computational methods, many existing geodemographic models of residential differentiation struggle to capture the multidimensional and non-linear nature of contemporary residential patterns, particularly in rapidly changing urban contexts. This is perhaps most evident in the UK Output Area Classifications for the 2001, 2011 and 2021/22 Censuses where the extent of Greater London suffers much worse clustering outcomes than the rest of the UK (Singleton & Longley, 2015; Singleton & Longley, 2024). Traditional approaches, while valuable for their interpretability and extensive lineage, often rely on assumptions of linearity that may oversimplify the fluid and overlapping nature of urban residential patterns. For example, the relationship between educational attainment and residential location may vary substantially across income brackets: among high-income households, educational differences may have minimal effect on neighbourhood choice, whereas among middle-income households, the same differences may strongly predict sorting into distinct areas; a conditional relationship that linear methods may not adequately represent.

This review paper examines how quantitative methods of residential differentiation have evolved, with particular focus on geodemographics: which are classification systems that segment small area geography into clusters based on their demographic, built, socioeconomic and behavioural characteristics. The review juxtaposes a rich theoretical and historical perspective with emerging computational approaches to consider how machine learned representations of the social landscape might complement traditional geodemographic methods. Machine learned representations are numerical summaries of data produced by algorithms that automatically discover patterns and relationships from data, offering an alternative to traditional data reduction techniques such as Principal Component Analysis (PCA). We explore whether such approaches could help capture the increasingly complex patterns of residential differentiation in contemporary cities, while discussing how interpretability might be preserved. We proceed by reviewing quantitative representation approaches (Section 2) and tracing the historical development of geodemographic methods (Section 3). Following a critical evaluation of current methods' advances and limitations (Section 4), we explore potential future directions involving learned representations (Section 5) and outline a possible framework for integrating these approaches with traditional geodemographics (Section 6). The paper concludes with critical reflections on future directions for residential differentiation research and implications for urban theory and practice (Section 7).

2. Quantitative representation of urban residential patterns

In the process of distilling the complexities of the real world into quantifiable metrics of residential differentiation, critical choices are required regarding the specificity of appropriate models. Foundational concepts include the choice of representation, which includes considerations of how residential characteristics are encoded within the model whether as individual household-level data points, aggregated spatial units, or continuous surfaces, and which variables best capture meaningful differentiation patterns. The spatial scale frames both the resolution of an analysis (from individual property to census tracts to metropolitan regions) and the geographic extent of the study area, directly influencing the detection of segregation patterns through well-documented modifiable areal unit problem (Openshaw, 1984). Similarly, temporal scale determines whether models capture static cross-sectional patterns or dynamic processes of neighbourhood change, including decisions about measurement frequency (annual versus decennial), the treatment of temporal lags in segregation processes or the incorporation of path dependencies in residential sorting. The balance between process and outcome represents a fundamental tension between models that emphasise the mechanisms driving residential differentiation (household mobility decisions, housing market dynamics) versus those focused on measuring resulting spatial patterns (dissimilarity indices, exposure metrics). These conceptual decisions are not merely technical specifications but fundamentally shape how residential differentiation is understood, measured, and ultimately managed through policy interventions (Borrego-Díaz et al., 2012; Duckham et al., 2001; Fisher & Tate, 2015).

2.1. Challenges of generalisation and classification

Any classification system requires that categories are defined distinctly, avoiding overlap and ambiguity, thereby enhancing both clarity and practical utility. In residential differentiation analysis, this often involves classifying neighbourhoods or regions based on continuous socio-demographic measures, which requires establishing value boundaries and intervals to create meaningful categories (Reibel, 2011). For example, while some measures like Shannon entropy operate on continuous scales without inherent boundaries, researchers must define thresholds when categorising spatial units for comparative analysis. However, determining and operationalising these thresholds, essentially converting continuous scales into ordinal, interval, or categorical variables varies in difficulty. Some boundaries are well-established through convention or policy (such as age brackets or income thresholds relative to median values), while others require greater judgement, particularly when measures lack natural breakpoints or when definitions vary across contexts. Furthermore, any boundary delineation introduces edge effects: observations near category thresholds may be assigned to different classes with only minimal changes in their underlying values, potentially overstating differences between similar areas or obscuring meaningful distinctions (Openshaw, 1984).

Contemporary measures of residential differentiation must capture underlying spatial patterns and socio-economic structures that may not be absolute value of variables but manifest through complex, non-linear associations. For example, the relationship between education levels and residential choice may vary substantially across income brackets, or ethnic clustering patterns may intensify or dissipate at different spatial scales. Within late capitalist societies, technological and economic transformations have fundamentally altered traditional patterns of residential sorting. For instance, the widespread adoption of remote work technologies following 2020 has decoupled some employment locations from residential choice, enabling affluent professionals to relocate from urban centres to peripheral areas while maintaining metropolitan salaries, thereby disrupting established urban-suburban income gradients (Ramani et al., 2024). Additionally, the proliferation of short-term rental platforms like Airbnb has converted some residential housing stock into quasi-commercial properties in tourist-adjacent neighbourhoods, displacing long-term residents and altering demographic compositions (Cocola-Gant & Gago, 2021; Wachsmuth & Weisler, 2018). Furthermore, online property purchasing platforms that filter and rank neighbourhoods based on proprietary scoring systems (for schools, walkability, or safety) can amplify residential segregation by steering homebuyers towards algorithmically similar neighbourhoods, reinforcing existing spatial inequalities (Besbris & Faber, 2017; Boeing, 2020; Nadiyah, 2023).

Further complicating measurement is the distinction between directly observable variables and context-dependent constructs. Consider poverty measurement: while income can be measured directly in monetary terms, poverty represents a relative condition that varies by geographic context, household composition, and local cost of living (Ringen, 1988; Tomlinson et al., 2008). A household earning \$50 k annually might be considered impoverished in San Francisco but middle-class in rural Mississippi. Similarly, educational attainment, ostensibly a straightforward categorical variable, can become problematic when measuring migrant populations, as many countries in the Global North do not recognise university degrees awarded in Global South countries. This creates "brain waste" situations where highly educated migrants are registered as having only basic education due to non-recognition of their qualifications and consequently cannot obtain employment matching their actual skill level (Devos et al., 2025; Mattoo et al., 2008). In both examples, the same underlying reality (household resources, educational achievement) is represented differently depending on institutional and geographic context, introducing systematic measurement bias into residential differentiation analyses. Such context-dependency creates generalisation challenges when developing universal measures of residential differentiation across diverse urban contexts. Even earlier studies on residential differentiation questioned the spatial stationarity assumptions embedded in traditional ecological modelling approaches (Hughes & Carey, 1972). A more recent contribution has sought to address these challenges by developing geographically weighted composite measures, allowing coefficients and relationships to vary spatially and aiming to offer more nuanced assessments of local spatial structures (Lu et al., 2014).

2.2. Scale considerations

Spatial scale represents a critical consideration in residential differentiation analysis, encompassing multiple geographic abstractions from data collection zones to the range of spatial process variation (Goodchild, 2001; Oshan et al., 2022). The chosen scale fundamentally shapes observed patterns; small area analysis may highlight nuanced socioeconomic or racial variation, whereas analysis at broader scales can obscure this variation, producing homogenised representations of urban structure that mask underlying heterogeneity. Scale selection is guided by pragmatic constraints including data availability and analytical objectives, with the spatial extent of available data directly influencing both methodological choices and the development of context-specific models for characterising residential differentiation (Reibel, 2011; Singleton & Longley, 2015).

2.3. Outcome versus process

Residential differentiation research has evolved along two distinct analytical perspectives: one focusing on spatial outcomes and another on underlying processes. Outcome-based approaches excel at describing the spatial configuration of population groups at specific points in time: mapping where different socio-economic, ethnic, or demographic groups concentrate within cities. However, these approaches often fail to capture the mechanisms driving these patterns: why certain groups cluster in particular neighbourhoods, how migration flows reshape urban demographics, or what factors trigger neighbourhood transitions? (Li et al., 2022; Massey & Denton, 1988). Understanding these processes requires examination of the temporal dynamics and statistical associations that produce the observed spatial distributions.

Contemporary measures of residential differentiation predominantly describe single periods of time which contrast with some of the earliest models of urban form that emphasised processes, focusing on how social factors such as immigration shaped the evolution of urban form and the distribution of social groups within a city (Park et al., 1925). As such, contemporary models of residential differentiation do not necessarily improve an understanding of spatial processes behind neighbourhood change that lead to the observed patterns (Delmelle, 2022; Libório et al., 2022; Owens, 2012). Although, some models of residential differentiation have been operationalised to represent change in both specific urban settings or for national extents (Delmelle, 2015; Jung & Song, 2022; Kinahan, 2021; Silver & Silva, 2021; Singleton et al., 2016; Skupin & Hagelman, 2005). The validity of static measures of residential differentiation depends on the stability of the populations being classified. A neighbourhood classified as 'young professionals' illustrates this dependency. If residents of this type persist, the classification validly describes who lives there but becomes outdated as they age. If population turnover maintains the area's profile, the classification instead describes what kind of place it is, a transient zone through which successive cohorts move. Static measures cannot distinguish continuous population replacement from residential persistence, yet these carry fundamentally different implications. The former suggests stable demand for housing types and amenities, with the area functioning as a transitional zone likely to retain its character. The latter signals impending demographic transition as residents enter family-formation years, bringing shifts in service needs and housing demand that the static classification cannot anticipate. Such ambiguity limits the utility of cross-sectional classifications for anticipating neighbourhood trajectories or designing responsive policy interventions. Past work has explored how such changes can be detected for a national extent (Singleton et al., 2016) or the degree to which neighbourhoods age or are replenished by residents of similar characteristics (Wyszomierski et al., 2022).

Recent advances in residential mobility research has begun to bridge the traditional outcome-process divide by combining the interpretability of econometric frameworks with the predictive power of machine and deep learning. For example, Bostanara et al. (2024) used a survival analysis framework integrating tree-based ensemble methods to model the timing of household relocations, highlighting the importance of homeownership and accessibility as key predictors of residential change. Jin et al. (2023) employed an explainable deep learning approach to model individual relocation trajectories across the US, revealing path-dependent effects of prior mobility and localised neighbourhood context. Noferesti and Mirzahossein (2025) demonstrated the capacity of gradient boosting models to substantially outperform discrete choice models in predicting residential location decisions, effectively capturing complex, non-linear decision processes. At the neighbourhood scale, Silver and Silva (2021) combined geodemographic clustering with Markov modelling to simulate transition dynamics, offering new insights into path dependence and neighbourhood evolution. Such methodological innovations suggest further promising avenues for integrating process-oriented insights into geodemographic frameworks, extending past work in this area (Martin et al., 2018; Singleton et al., 2012).

3. Historical development of geodemographic methods for residential structure analysis

Geodemographic classification is the current dominant quantitative paradigm for mapping residential differentiation. In this section we trace the lineage of techniques, from early social surveys through ecological models to contemporary open and commercial systems, focusing on the rationale behind the methods, the types and limitations to the representations that they have presented.

3.1. Foundations: social surveys and urban ecology (1880s–1950s)

One of the earliest examples of systematic residential differentiation research is Charles Booth's analysis of neighbourhood-level poverty in London (1886–1903),[1] which produced a series of detailed maps combining survey data, census records, and ethnographic observation to classify small areas into poverty-based categories (Vaughan, 2018). This body of work aimed to highlight the consequences of unregulated

[1] The original study was published in two volumes - Life and Labour of the People, Vol. I (1889) and Labour and Life of the People, Vol. II (1891); including a series of maps. These can be viewed online - https://booth.lse.ac.uk/map

capitalism (Marr, 1904; Pfautz, 1967). Although influential, Booth's reliance on class rather than income-based definitions and use of moralistic labelling has subsequently attracted criticism (Spicker, 1990); and further related studies, such as Hubert Llewellyn Smith's "New Survey of London Life and Labour" (1928–35) have had concerns raised about their data quality and collection methods (Abernethy, 2017). In the United States, residential differentiation research emerged predominantly from the Chicago School's ecological tradition (1920s–1950s), where Park and colleagues applied biological principles to urban structure through their concentric ring model. This also explicitly acknowledged Booth's influence on this work (Park et al., 1925). Later models by Hoyt (1939) and Harris and Ullman (1945), expanded this approach through sector and multiple nuclei frameworks, though all faced criticism for their limited generalisability beyond their original contexts. These foundational studies established two enduring principles: that residential differentiation could be systematically observed and mapped, and that urban neighbourhoods constitute meaningful units of social analysis. However, both Booth's surveys and the Chicago School's ecological models relied primarily on qualitative judgement and theoretical abstraction rather than standardised quantitative methods, limiting their replicability across contexts.

3.2. Social area analysis and factorial ecology (1950s–1970s)

Post-Second World War studies in Los Angeles advanced an ecological tradition through Social Area Analysis which used detailed small-area census data to position neighbourhoods within a broader social context (Rees, 1971; Timms, 1970). Shevky and Williams (1949) and later Shevky and Bell (1955) developed these measures through indices of social rank, urbanisation, and segregation; although the approaches faced criticism for lacking generalisability (Ericksen, 1949; Hawley & Duncan, 1957). Nevertheless, Social Area Analysis demonstrated the utility of multivariate data for analysing residential patterns, directly influencing what became termed Factorial Ecology (Singleton & Spielman, 2014). Enabled by growing computational access, Factorial Ecology dominated residential differentiation research from the 1960s to 1970s and continues to be applied as a technique in some contemporary research (Al-Shawamreh & Farhan, 2018; Hirokane & Amemiya, 2023; Montosa Muñoz, 2023; Salvati et al., 2018). Like Social Area Analysis, factorial ecology identifies latent urban structures through combinations of variables, though it derives constructs empirically via factor analysis rather than from predetermined indices (Berry & Kasarda, 1977; Hale & Austin, 1997). However, despite its wide use, factorial ecology encountered persistent challenges that included: reduced clarity when mapping complex multidimensional data, and limited generalisability due to localised specificity and the inconsistency of city boundary definitions (Rees, 1971), the non-linear nature of geographic data (Johnston, 1977) and the sensitivity of outcomes to analytical choices, which raised concerns over reproducibility (Hughes & Carey, 1972; Hunter, 1972; Vies, 1978).

Social Area Analysis and Factorial Ecology made a critical methodological contribution by demonstrating that multivariate census data could be systematically reduced to interpretable dimensions of urban structure. Yet such approaches also highlighted persistent challenges including sensitivity to analytical choices, limited generalisability beyond their original contexts, and the assumption of linear relationships among variables that would constrain subsequent geodemographic methods.

3.3. The emergence of applied geodemographics (1970s–1980s)

During the 1970s, advances in computing and digitised administrative and census data significantly enhanced the potential for multivariate analysis of urban areas. A pivotal early example is the "Social Malaise in Liverpool" study (Amos, 1970), which was commissioned to guide social service resource allocation. This study, integrated data from multiple municipal departments and the 1966 Census, combining social and built environment variables, though it lacked sophisticated summary methods such as factor analysis (Cullingford et al., 1975). Following this study, the UK Department of Environment commissioned comparative studies for several inner urban areas in Birmingham, Liverpool, and London (Lambeth) in 1977 (Department of Environment, 1977). However, the Liverpool study stood out due to its collaboration with the Planning Research Applications Group (PRAG) at the Centre for Environmental Studies (CES) (Parliament. House of Commons, 1967). Two landmark studies on residential differentiation emerged from this collaboration. Cullingford et al. (1975) applied Principal Component Analysis (PCA) to summarise census and programme data into components that guided cluster analysis. Webber (1975) then refined this approach using 1971 Census data, aggregating enumeration districts into unified areas and creating hierarchically nested clusters, a structure that became a defining feature of what would be later termed geodemographic classification.

As part of work within CES, Webber's team extended their developed methods to other local authority areas and motivated by the goal of national standardisation, developed a nationwide enumeration district classification (Webber, 1979). This marked the final effort to produce a national residential differentiation model within the CES. In 1979 funding from the government was withdrawn, and the centre closed in the early 1980s. The financial stability of CES had been questioned towards the end of its operations, as by 1976 government funding accounted for over 80 % of its total revenue because of declining sources of other income (Parliament, 1976).

The Liverpool studies and subsequent national classifications established an operational template for residential differentiation that remains foundational to current geodemographic practice: variable selection, standardisation, linear dimensionality reduction, clustering and hierarchical labelling. This workflow has proved remarkably effective for policy targeting and commercial applications, yet its core methodological architecture, particularly where there is reliance on linear dimensionality reduction, has remained largely unchanged for five decades.

3.4. Commercialisation, critique and the open geodemographics movement (1980s–present)

The evolution of geodemographic residential differentiation approaches demonstrated a shift from narrowly focused models towards national-scale classifications designed to generalise across contexts (Batey & Brown, 2007; Batey, 2022; Webber, 1979). This however attracted criticism: Openshaw et al. (1980) questioned whether national classifications sufficiently represented local realities, while Webber (1980) defended their practicality and user acceptance. However, despite such challenge, it is worth noting that Openshaw and his wider team also subsequently developed national classifications (Blake & Openshaw, 1995; Charlton et al., 1985).

The commercialisation of geodemographic residential differentiation models emerged somewhat incidentally. In the late 1970s, integration of the CES classification with consumer survey data revealed variations in consumer behaviour linked to residential typologies, generating significant commercial interest (Webber & Burrows, 2018). Moving from CES in 1979, Webber joined CACI, which rebranded the CES classification as ACORN (A Classification of Residential Neighbourhoods), marking the beginning of commercial geodemographics. Comprehensive accounts of this commercial history are provided by Webber and Burrows (2018) and Harris et al. (2005) for the UK, and Singleton and Spielman (2014) and Weiss (1988, 2000) for North America.

Commercial success, particularly as an integral component of direct marketing during the 1980s and 90s, distanced geodemographics from its public-sector origins until renewed interest emerged in the mid-2000s, driven by the development of social marketing and as a tool for public service delivery improvements (Longley, 2005). Academic applications supported this renewed interest and expanded across health (Farr & Evans, 2005; Moon et al., 2019; Petersen et al., 2011), education (Butler et al., 2007; Hayden et al., 2006; Singleton, 2010; Singleton et al., 2012) and policing (Ashby, 2005; Williamson et al., 2006).

In parallel to renewed public use, critical GIS scholars also questioned geodemographics' societal implications more generally during this period, arguing that classifications perpetuated existing social structures by strongly associating identity with location (Goss, 1995, 2003; Pickles, 1995; Thrift & French, 2002). The introduction of 'open geodemographics', beginning with the 2001 Output Area Classification (OAC) developed collaboratively between academia and the Office for National Statistics (ONS) can be argued both as a response to such critique, as they aimed to build public trust through transparency, reproducibility, and openness; but also were a facilitator of broader public sector adoption of geodemographics. Open Geodemographics differentiate from most commercial offerings by being created entirely from open data sources and with published and reproducible methods, enabling free access and in formats that are fully open to scrutiny (Vickers & Rees, 2007). Following 2001 OAC, updates followed subsequent censuses (Gale et al., 2016; Wyszomierski et al., 2022), and similar open approaches have been extended internationally (Spielman & Singleton, 2015). Geodemographic models have also faced criticism for their limited theoretical engagement (Singleton & Longley, 2009), especially in commercial contexts. However, concurrent efforts have sought to reconcile geodemographics with social theory from both the commercial (Webber, 2007) and academic sectors (Burrows, 2008; Burrows & Gane, 2006; Parker et al., 2007; Uprichard et al., 2009), and more recently in unison (Webber & Burrows, 2018). Efforts have included drawing parallels between geodemographics to Bourdieu's concepts of habitus and field (Tapp & Warren, 2010) and linking classifications to phenomena like gentrification (Gray et al., 2023a; Lees, 2000; Somashekhar, 2021) or concepts such as habitus (Robson, 2003; Webber, 2007).

Today, the geodemographic industry operates through a bifurcated market structure. In the commercial sector, major providers including CACI (Acorn), Experian (Mosaic), and in North America, Claritas (PRIZM) and Esri (Tapestry) continue to dominate consumer segmentation markets across retail, financial services, utilities and telecommunications. These proprietary systems have evolved considerably from their census-dependent origins; and most no longer rely on census data alone for their core classification, instead drawing on continuously updated commercial and open data sources. Concurrently, the open geodemographics tradition has continued through successive iterations of the Output Area Classification, most recently updated following the 2021/22 Census (Wyszomierski et al., 2022). Public sector adoption remains substantial, with geodemographic classifications now embedded across health service planning, educational targeting, and local authority resource allocation. Yet despite this institutional maturation, fundamental questions persist about whether the core methodological approaches developed in an era of computational constraint and data scarcity remain optimal given contemporary analytical capabilities and data environments.

4. Advances and limitations of current geodemographic methods

Methodological advances in geodemographic classification have largely been driven by increased computational power, enabling clustering without reliance on sample-based methods or pre-aggregated zones, alongside an expanded range of input variables facilitated by the growth of the spatial data economy, although core analytical workflow has changed little since the 1970s. While commercial providers commonly adopt variations of standard methods, detailed descriptions remain limited in publicly accessible literature, with Chapter 6 of Harris et al. (2005) providing one of the few comprehensive reviews. Across open and commercial classification, variable selection remains heavily dependent on researcher judgement and data availability, with limited systematic frameworks for identifying optimal inputs (Otley et al., 2021) and normalisation and standardisation techniques. Most critically, when linear dimensionality reduction methods such as PCA are applied, these constrain the capacity to detect complex, non-linear relationships between socio-demographic characteristics. Similarly, while clustering algorithms have diversified, the dominant approaches of k-means and hierarchical clustering remain fundamentally limited in their ability to capture overlapping or fluid category memberships (Fisher & Tate, 2015).

The methodologies underpinning recent iterations of the UK's Output Area Classification (OAC) for 2001, 2011, and 2021 represent some of the most thoroughly documented approaches to residential differentiation. Nonetheless, their methods have remained largely unchanged, with incremental enhancements made to clustering techniques and data normalisation based on ground truthing (Vickers & Rees, 2007) and critical evaluations (Gale et al., 2016; Wyszomierski et al., 2022). Stability in method has often been justified by the proven utility and acceptance of previous classifications, a rationale similarly cited in regional classifications (Singleton & Longley, 2015). While this approach leverages established frameworks and their documented effectiveness, it may limit the exploration of potentially superior methods offering enhanced descriptive capability. Nonetheless, the geodemographic field has responded constructively to criticism (Knaap et al., 2024), fostering methodological innovations such as real-time classification using advanced computational platforms (Adnan, 2011), bespoke, application-specific models (Alexiou et al., 2016; Singleton et al., 2020; Yang et al., 2023), improved variable selection (Liu et al., 2019) and clustering methods (Brunsdon et al., 2018; De Sabbata & Liu, 2023; Spielman & Thill, 2008), increased transparency and reproducibility (Vickers & Rees, 2007), and public interfaces supporting user feedback (Longley & Singleton, 2009).

The emergence of detailed individual-level data from integrated consumer and synthetic databases has also driven interest in personal-level geodemographic classifications. Commercial systems such as Experian's UK Mosaic have evolved offerings from area-based to household-level classifications by linking consumer transaction records, credit data, and lifestyle surveys to individual addresses (Farr & Webber, 2001). Academic efforts have similarly explored individual-centred approaches, with Burns et al. (2018) developing classifications that assign demographic profiles directly to synthetic individuals rather than geographic units, while Tuccillo (2021) employed machine learning techniques to create person-level typologies from integrated administrative and survey data. However, these developments have attracted critical scrutiny. Dalton and Thatcher (2015) argue that the apparent precision of individual-level geodemographics represents "inflated granularity", a veneer of accuracy that obscures the inferential assumptions underlying such classifications. They also contend that personal-level profiling intensifies concerns about surveillance, privacy, and the potential for discriminatory outcomes, as individuals may face differential treatment in credit, insurance, or service provision based on algorithmically assigned characteristics they cannot see or contest. These critiques underscore the tension between the analytical appeal of finer-grained classification and the ethical responsibilities accompanying such granular profiling of populations.

These methodological advances have also not fully resolved concerns regarding an apparent bifurcation away from process-oriented models of residential differentiation (Knaap et al., 2024), despite recent efforts towards predictive geodemographics (Gray et al., 2023a, 2023b). Fundamentally, this debate hinges on whether geodemographics should intrinsically aim to articulate residential differentiation theories, or function primarily as an inductive framework facilitating external theory development. Examples of the latter approach have included integrating geodemographic clusters within multilevel models (Harris et al., 2007; Harris & Feng, 2016), spatial interaction models (Singleton et al., 2012), and microsimulation frameworks (Moon et al., 2019). Our perspective also aligns more closely with the view that geodemographic classifications should function primarily as inductive frameworks facilitating external theory development, rather than intrinsically aiming to articulate theories of residential differentiation themselves. Historically, attempts at developing endogenous, generalisable theoretical models have struggled to maintain applicability beyond their original geographical contexts. However, the increased availability of spatially referenced data and significant advancements in computational technologies present substantial opportunities for enhancing theory development through empirical generalisation. We argue that the limited methodological evolution of geodemographics since their inception has constrained their analytical potential, while their complex historical positioning across academic, public, and commercial domains has created ambiguity about their purpose and credibility. Together, these factors have arguably hindered broader acceptance within the social sciences and limited recognition of geodemographics as a valuable tool for urban research. In the subsequent section, we propose some new avenues for advancing models of residential differentiation.

5. Exploring potential directions: learned representations of data in geodemographics

5.1. Traditional geodemographic methodology

Creating geodemographic representations typically involves a two-stage process. In the first stage, input variables such as demographic, socioeconomic, built environment and behavioural data are selected based on both availability and theoretical relevance derived from ecological theories. These variables may have transformation applied and then are typically standardised using statistical techniques. In the second stage, clustering methods are applied to these variables to group geographic areas with similar characteristics. The most frequently employed clustering algorithms in geodemographic analysis are k-means, which partitions data into a predefined number of clusters based on centroids, and hierarchical clustering, which creates nested groups by progressively merging clusters.

5.2. Learned representations of data as a potential direction for geodemographics

When building a geodemographic classification, cluster analysis is highly sensitive to the structure of the input data; for example, using many correlated variables in a k-means analysis may bias the resulting clusters to disproportionately reflect the group of correlated variables. For this reason, it is common to reduce input dimensionality via careful variable selection, weighting or some other form of statistical data reduction. PCA has been widely applied as a linear transformation onto input data. However, where PCA identifies linear combinations of variables, this results in the loss of complex patterns such as conditional relationships between variables and non-linear associations. Newer developments within the machine learning literature, and specifically in the development of neural networks, have established alternate approaches that can identify compressed representations, sometimes called “embeddings”, that better capture non-linear interactions. Such methods provide an alternative to traditional data reduction methods like PCA, while acknowledging important caveats. Where PCA identifies linear projections that maximise variance, neural network architectures can learn nonlinear transformations that may better capture the intrinsic structure of certain datasets. In one such example, Hinton and Salakhutdinov (2006) demonstrated that deep autoencoder networks that compress inputs through a bottleneck layer before reconstructing them substantially outperform PCA in preserving data structure when reducing high-dimensional datasets to low-dimensional representations, particularly when underlying relationships are non-linear. Subsequent work has confirmed these advantages across diverse domains, showing that learned representations more effectively preserves local neighbourhood relationships and complex feature interactions than linear alternatives (Goodfellow et al., 2016; Wang et al., 2016). Within geographic applications, preliminary evidence suggests similar benefits for spatial data, with learned representations capturing place-based characteristics that linear methods fail to detect (De Sabbata & Liu, 2019; Yan et al., 2017). This could prove valuable for clustering when groups are separated by nonlinear boundaries, as PCA's linear constraint may fail to disentangle clusters that are clearly separable in the original space but become mixed when projected linearly. Additionally, PCA's variance-maximisation objective does not necessarily align with cluster separability; the directions of greatest variance may be orthogonal to the directions that best discriminate between clusters. Learned representations, through nonlinear encoding, could potentially bring similar points closer together in latent space while pushing dissimilar points apart, even when this structure is not accessible through linear projection.

However, it is essential to acknowledge that the superiority of non-linear methods is not guaranteed. Empirical comparisons across various domains have demonstrated that learned representations do not universally outperform PCA. Fournier and Aloise (2019, p. 211) found that k-NN classifiers achieved comparable accuracy on PCA and autoencoder projections when sufficient dimensions were retained, while “PCA computation time was two orders of magnitude faster than its neural network counterparts”. Their conclusions underscore that the potential advantages of learned representations depend heavily on the specific characteristics of the data and research objectives. Where relationships in the data are predominantly linear, the additional complexity of deep learning methods may offer little practical benefit while introducing interpretability challenges.

Furthermore, the flexibility of learned representations comes at the cost of losing PCA's interpretability and theoretical guarantees about variance preservation. The “black box” nature of neural networks presents particular challenges in policy-relevant contexts where transparency is essential (Liu et al., 2024). Other non-linear dimensionality reduction techniques may also warrant consideration. t-SNE and UMAP excel at preserving local neighbourhood structure in high-dimensional data. Although, while t-SNE creates visually interpretable clusters with strong local structure preservation, it becomes computationally prohibitive at scale and often distorts global relationships. UMAP mitigates these computational concerns while maintaining comparable local structure preservation. Parametric neural network approaches provide complementary advantages that address several limitations inherent in both linear methods like PCA and non-linear techniques such as t-SNE and UMAP, making them particularly suited to geodemographic applications requiring scalability, interpretability, and flexibility. Unlike t-SNE and UMAP, which require recomputing learned representations for new observations, neural networks learn a parametric encoder function that can efficiently map unseen data into the latent space. Architectures with decoder components also enable reconstruction back to the original feature space, potentially facilitating interpretation. Furthermore, such approaches naturally extend to multi-task learning frameworks where clustering objectives can be jointly optimised with reconstruction. The learned representations can be interrogated using gradient-based attribution methods, SHAP values, or activation maximisation to identify which input features drive specific latent dimensions, potentially providing mechanistic insight into the learned structure that purely algorithmic methods like t-SNE cannot offer (Liu et al., 2024).

The potential value of learned representations in geodemographic research therefore remains an empirical question requiring systematic comparison with established methods. We propose this direction not as a demonstrated improvement, but as a methodological hypothesis warranting investigation, given the theoretical possibility that residential differentiation involves complex non-linear interactions that current methods may inadequately capture.

5.3. Enriching input feature sets, and further considerations for data quality and bias mitigation

The integration of learned representation into geodemographic workflows would also enhance the potential for integration of more complex and heterogeneous datasets, potentially uncovering new insights about residential differentiation features and processes. Traditional variables used to build geodemographic indicators could potentially be augmented with many new sources of spatial and temporal data sources, including those which are unstructured (see Fig. 1).

However, given such potential to integrate wider sources of data, this also brings enhanced responsibility for ensuring data quality and representativeness. Without rigorous attention to such issues there is enhanced risk of perpetuating or amplifying systemic inequalities already embedded within datasets, reflecting broader critiques of data-driven urban analyses as articulated by Kitchin (2014). The inherent opacity and complexity of models used to generate learned representations exacerbates the "black box" problem, limiting interpretability which can be problematic in those contexts where a nuanced comprehension of underlying social dynamics is crucial, particularly when informing or needing to justify equitable and just policy interventions (Liu et al., 2024; Sieber & Haklay, 2015).

As such, we would argue that proactive measures would be essential to assess such issues in any application that might apply learned representations. Rigorous data audits should systematically examine sources, collection methods and historical contexts that could introduce bias. Techniques including demographic parity analyses (Feldman et al., 2014; Kamiran & Calders, 2012; Mehrabi et al., 2022), counterfactual fairness tests (Kusner et al., 2017) and bias impact assessments (Raji et al., 2020) could all help to quantify and mitigate unintended biases. Inclusive participatory processes involving diverse community members and local stakeholders could also further help ensure that data representativeness accurately reflects community realities. More generally, beyond data quality, we would also argue that prioritising model transparency and interpretability would be essential for any practical application within a geodemographic context. The integration of Explainable AI (XAI) methodologies is not merely beneficial in such context, but essential for responsible deployment in urban contexts. Recent work has begun to address this challenge: Liu et al. (2024) proposed combining graph convolutional networks with GNNExplainer for urban analytics applications, demonstrating how explainability methods can be integrated with spatial deep learning. Feature attribution techniques, saliency maps, and sensitivity analyses (Lundberg & Lee, 2017; Ribeiro et al., 2016) can also illustrate how deep learning models interpret urban space. We suggest that future research should prioritise the development of domain-specific interpretability tools that decompose learned representation outputs to reveal which spatial, social, and economic patterns drive latent representations. Such transparency would be crucial for building trust among policymakers and communities affected by geodemographic classifications.

6. Towards a future framework: considering the integration of learned representations with traditional geodemographics

This section outlines a potential framework that could integrate learned representation architectures within established geodemographic classification workflows. The framework remains speculative and would require empirical validation to determine whether the proposed advantages discussed in the previous section translate into meaningful improvements. As noted, existing comparative studies suggest that the advantages of non-linear dimensionality reduction are context-dependent and may not materialise in all geodemographic applications (Fournier & Aloise, 2019). The framework presented here should therefore be understood as a research agenda rather than a demonstrated improvement over existing approaches. Fig. 2 illustrates the key differences between traditional and proposed approaches.

This proposal differs from traditional methods in several interrelated respects. First, the capacity to identify relationships would expand from purely linear associations to encompass both linear and non-linear relationships, potentially enabling detection of complex interactions between variables, such as how the relationship between education and residential choice varies across income brackets, that linear methods

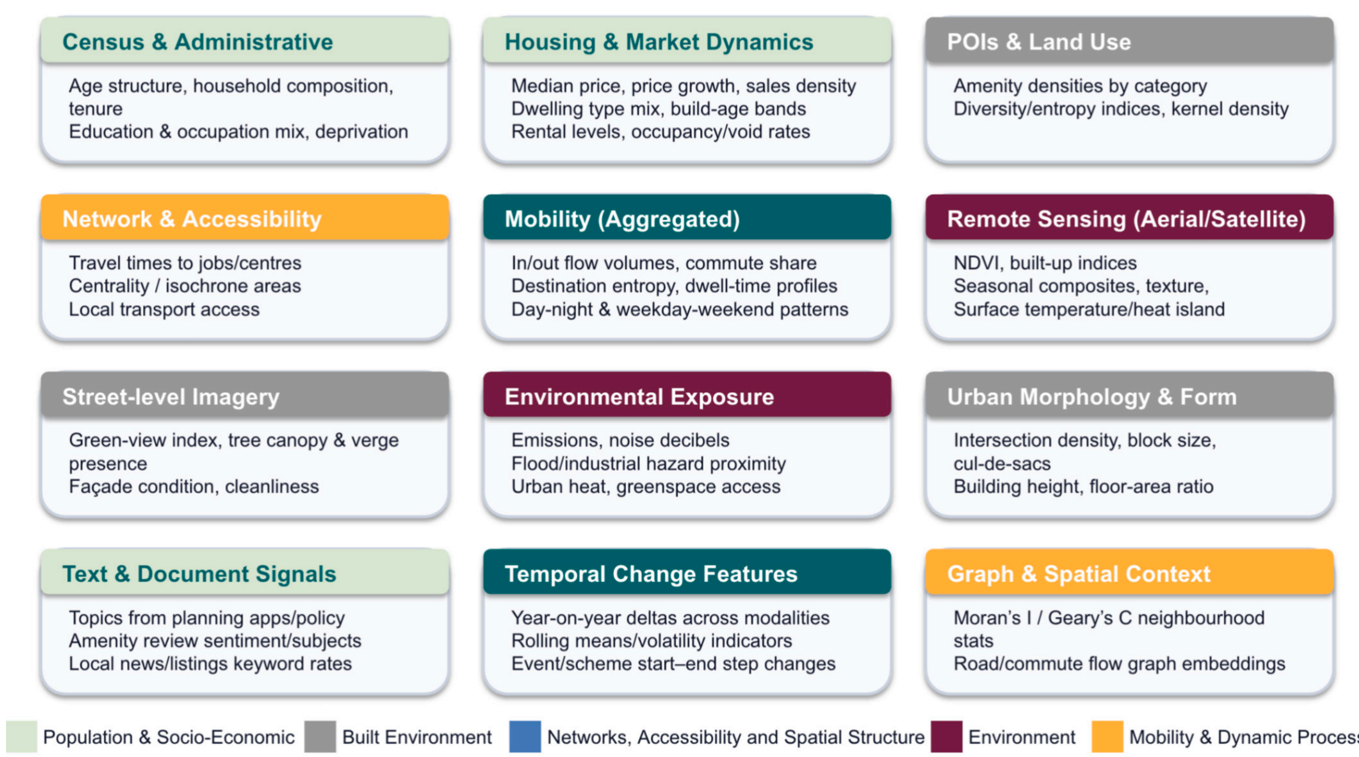

Fig. 1. Potential areas and example inputs to a learned representation of data for residential differentiation.

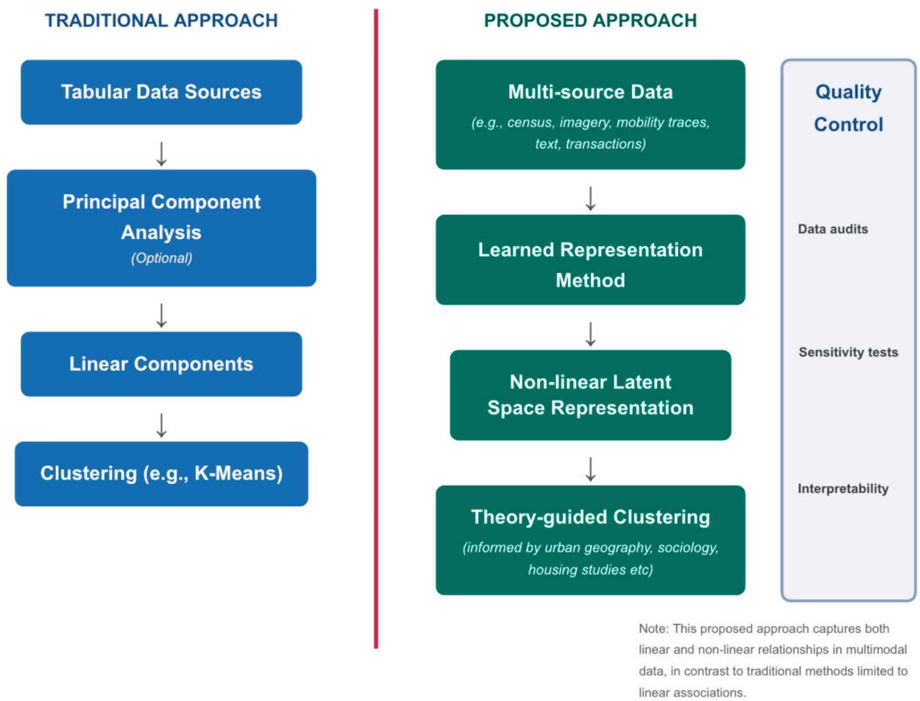

Fig. 2. A comparison of traditional and proposed approaches for developing geodemographic classifications. The traditional framework (left) employs linear dimensionality reduction techniques such as PCA, which constrain pattern detection to linear associations between variables. The proposed framework (right) substitutes learned representations, enabling identification of both linear and non-linear relationships in the input data while preserving compatibility with established clustering methods.

cannot capture. Second, data integration potential might increase substantially, allowing diverse and previously incompatible measures (for example, census tabulations, satellite imagery, mobility traces, and text-derived sentiment indicators) to coexist within a unified analytical framework. Third, interpretability mechanisms such as SHAP values, saliency maps, and activation maximisation could offer insights into how individual features contribute to emergent latent structures, though this would require careful development and validation.

Building upon this representational foundation, such a framework might also incorporate a theory-guided clustering stage that reintroduces conceptual reasoning into what has traditionally been a predominantly empirical process. This might involve using theoretical priors derived from urban geography, sociology, or housing studies to guide cluster granularity, assessing whether emergent segments reflect known urban structures (such as core-periphery gradients, tenure stratifications, or mobility catchments) or testing whether unexpected patterns challenge prevailing assumptions about neighbourhood differentiation. Whereas conventional clustering algorithms identify groups based solely on statistical similarity, a theory-guided approach would more closely align the clustering and interpretation of latent representations with established urban and social theories concerning spatial segregation, urban morphology, or residential mobility. In practice, this would involve iterative alignment between the empirical structure of the latent space and the conceptual expectations derived from theory, thereby aiming to ensure that the resulting classifications are not only statistically robust but also theoretically coherent and socially interpretable. Crucially, this does not imply imposing rigid theoretical templates on the data. Instead, we suggest that theory could function as a structured interpretive lens, shaping how clusters are evaluated, refined, and ultimately stabilised. The capacity of learned representation methods to capture non-linear relationships might be particularly valuable here: unlike PCA's linear projections, which may distort or obscure urban structures that do not align with linear axes of variation, they could potentially learn latent dimensions that more naturally correspond to theoretically meaningful constructs. Through this process, the analysis might become a dialogue between data and theory rather than statistical optimisation alone, positioning the resulting typologies as more conceptually anchored representations of urban structure.

Integral to this proposed framework would be rigorous quality control mechanisms. Data audits assessing representativeness and potential biases should precede model training, while sensitivity analyses would test the stability of learned representations across different architectural choices and input configurations. Interpretability techniques including feature attribution, saliency mapping, and counterfactual fairness tests would need to be embedded throughout the analytical pipeline rather than applied as post-hoc additions, ensuring that the framework maintains transparency despite the inherent complexity of non-linear models. Moreover, the inner workings of neural networks lack direct interpretability and we argue should also require the use of emergent eXplainable AI (XAI) methods. Such practical and interpretability challenges represent significant barriers that optimally would need to be addressed before such approaches could be widely adopted.

7. Critical reflections and future directions

7.1. Summary of contributions

This paper has advanced a proposition: that geodemographic representations of residential differentiation may benefit from integrating greater diversity of input data alongside learned representation methods, while maintaining the theoretical grounding and interpretability essential to urban research. The proposed framework presented in Section 6 preserves a neutral definition of residential differentiation established at the outset, as a descriptor of spatial patterns, while proposing expanded capacity to capture non-linear complexities increasingly characteristic of contemporary urban environments. The evolution from Booth's pioneering social surveys through the Chicago School's ecological models, Social Area Analysis, and factorial ecology to contemporary geodemographics represents a progressive refinement of methods for understanding urban residential structure. Each approach has been shaped by advances in data availability, computational capacity, and theoretical priorities, bringing new insights alongside inherent limitations. Our proposed framework represents a suggested continuation of this trajectory, specifically targeting the inability of traditional linear dimensionality reduction techniques that may be applied within geodemographic classification to capture complex, non-linear interactions identified as a limitation in Section 4. The Liverpool studies of the 1970s (Section 3.3) established the pragmatic utility of multivariate classification approaches for policy intervention. Our proposed framework seeks to maintain such operational utility while exploring methodological possibilities through the inclusion of non-linear dimensionality reduction, as illustrated in Fig. 2's comparison between traditional and proposed approaches.

The research agenda proposed by this framework represents further incremental development rather than wholesale replacement. Traditional clustering algorithms would be retained, preserving the interpretability and potential for nested categorical hierarchies that have proven valuable across decades of geodemographic application. What changes is the feature engineering stage: where we argue for the potential role of learned representations in data reduction to enable detection of non-linear patterns that linear methods cannot capture.

7.2. Critical reflections

Methodological sophistication brings commensurate responsibility. One longstanding tension in residential differentiation research concerns the balance between empirical pattern recognition and the deeper causal understanding valued in urban studies. Critics have argued that data-driven methods, particularly within "urban science," risk reductionism where complex social phenomena are abstracted into detached numerical patterns (Kitchin, 2020). While learned representation methods may reveal sophisticated patterns, their outputs are not inherently explanatory, and their opacity may hinder theoretical engagement or obscure harmful biases embedded within data (Sangers et al., 2022; Selbst et al., 2019). Thus, we do not advocate for uncritical adoption, but rather careful integration grounded in transparency and interpretability. The tools identified in Section 5.2, including SHAP values to quantify feature contributions, saliency maps to visualise spatial patterns, and activation maximisation to reveal archetypal inputs, all represent potential safeguards that would require further development and validation in geodemographic contexts. The objective would be to transform any application of a learned representation method from a black box into a more transparent analytical instrument, enabling researchers to interrogate how models are interpreting the attributes of geographic context.

Data quality considerations also remain foundational. Techniques such as demographic parity analyses, counterfactual fairness tests, and bias impact assessments, identified in Section 5.3 as relevant considerations should ideally be adapted and operationalised for geodemographic contexts in future work. Participatory methods, including feedback loops with local communities, could further ground model outputs and enable contestability when algorithmic classifications misrepresent lived reality (Acolin & Kim, 2024; Longley & Singleton, 2009). Public communication around model limitations and capabilities is vital for building AI literacy in urban governance (Fontes et al., 2024; Sanchez et al., 2025), ensuring that methodological developments do not outpace the democratic and ethical frameworks essential to their responsible use.

7.3. Future directions

Contemporary geodemographic models face ongoing challenges in capturing the complex, non-linear, and dynamic characteristics of urban residential patterns. We suggest that the proposed framework offers potential avenues for advancing both descriptive and analytical capabilities, though this remains to be demonstrated empirically. Key priorities for future research include: first, empirical testing of the proposed framework against traditional approaches using appropriate validation metrics; second, development of domain-specific interpretability tools tailored to geodemographic applications; third, creation of theory-grounded evaluation frameworks that assess alignment with established indicators of urban processes; and fourth, inclusive participatory approaches that ensure these methods serve society as comprehensively as they advance science. The framework's value will ultimately be measured not by technical sophistication alone, but by its capacity to generate insights that can meaningfully inform equitable urban policy.

Sources of Research Funding

This paper was funded by the Economic and Social Research Council (ESRC) under the grants ES/Z50273X/1 and ES/Z504464/1.

CRediT authorship contribution statement

Alex Singleton: Writing – review & editing, Writing – original draft. **Seth E. Spielman:** Writing – review & editing, Writing – original draft.

Data availability

No data was used for the research described in the article.

References

Abernethy, S. (2017). Deceptive data? The role of the investigators in the new survey of London life and labour 1928–32. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 50*(1), 1–15. https://doi.org/10.1080/01615440.2016.1182879

Acolin, A., & Kim, A. M. (2024). Teaching data justice: Algorithmic bias and critical spatial analysis in urban planning education. *Journal of Planning Education and Research, 44*(3), 991–1000. https://doi.org/10.1177/0739456X221116356

Adnan, M. (2011). *Towards real-time geodemographic information systems: Design, analysis and evaluation.* University College London.

Alexiou, A., Singleton, A., & Longley, P. A. (2016). A classification of multidimensional open data for urban morphology. *Built Environment, 42*(3), 382–395. https://doi.org/10.2148/benv.42.3.382

Al-Shawamreh, S., & Farhan, Y. (2018). GIS-based factorial ecology and social public space of the Twin City of Ramallah and Al-Bireh, West Bank, the Palestinian authority. *Journal of Geographic Information System, 10*(3). https://doi.org/10.4236/jgis.2018.103014. Article 3.

Amos, F. (1970). *Social malaise in Liverpool: Interim report on social problems and their distribution.* Liverpool Corporation.

Ashby, D. I. (2005). Policing Neighbourhoods: Exploring the geographies of crime, policing and performance assessment. *Policing and Society, 15*(4), 413. https://doi.org/10.1080/10439460500310079

Batey, P. (2022). Geodemographics and urban planning analysis: An historical review. In M. Kawano, K. Kourtit, P. Nijkamp, & Y. Higano (Eds.), *Theory and history in regional perspective: Essays in honor of Yasuhiro Sakai* (pp. 3–39). Springer Nature. https://doi.org/10.1007/978-981-16-6695-7_1.

Batey, P., & Brown, P. (2007). The spatial targeting of urban policy initiatives: A geodemographic assessment tool. *Environment and Planning A, 39*(11), 2774–2793. https://doi.org/10.1068/a38519

Berry, B. J. L., & Kasarda, J. D. (1977). *Contemporary urban ecology.* Macmillan.

Besbris, M., & Faber, J. W. (2017). Investigating the relationship between real estate agents, segregation, and house prices: Steering and upselling in New York state. *Sociological Forum, 32*(4), 850–873. https://doi.org/10.1111/socf.12378

Blake, M., & Openshaw, S. (1995). *Selecting variables for small area classifications of 1991 UK census data. 95/5.* School of Geography, University of Leeds Working Paper. https://www.geog.leeds.ac.uk/papers/95-2/.

Boeing, G. (2020). Online rental housing market representation and the digital reproduction of urban inequality. *Environment and Planning A: Economy and Space, 52*(2), 449–468. https://doi.org/10.1177/0308518X19869678

Borrego-Díaz, J., Chávez-González, A. M., Martín-Pérez, M. A., & Zamora-Aguilera, J. A. (2012). Semantic geodemography and urban interoperability. In J. M. Dodero, M. Palomo-Duarte, & P. Karampiperis (Eds.), *Metadata and semantics research* (pp. 1–12). Springer. https://doi.org/10.1007/978-3-642-35233-1_1.

Bostanara, M., Siripanich, A., Ghasri, M., & Hossein Rashidi, T. (2024). Sydney's residential relocation landscape: Machine learning and feature selection methods unpack the whys and whens. *Journal of Transport and Land Use, 17*(1), 369–399. https://doi.org/10.5198/jtlu.2024.2440

Brunsdon, C., Charlton, M., & Rigby, J. E. (2018). An open source geodemographic classification of small areas in the Republic of Ireland. *Applied Spatial Analysis and Policy, 11*(2), 183–204. https://doi.org/10.1007/s12061-016-9212-4

Burns, L., See, L., Heppenstall, A., & Birkin, M. (2018). Developing an individual-level geodemographic classification. *Applied Spatial Analysis and Policy, 11*(3), 417–437. https://doi.org/10.1007/s12061-017-9233-7

Burrows, R. (2008). Geodemographics and the construction of differentiated Neighbourhoods. In J. Flint, & D. Robinson (Eds.), *Community cohesion in crisis: New dimensions of diversity and difference* (pp. 219–237). The Policy Press.

Burrows, R., & Gane, N. (2006). Geodemographics, software and class. *Sociology, 40*(5), 793–812. https://doi.org/10.1177/0038038506067507

Butler, T., Hamnett, C., Ramsden, M., & Webber, R. (2007). The best, the worst and the average: Secondary school choice and education performance in East London. *Journal of Education Policy, 22*(1), 7–29.

Charlton, M. E., Openshaw, S., & Wymer, C. (1985). Some new classifications of census enumeration districts in Britain: A poor man's ACORN. *Journal of Economic and Social Measurement, 13*, 69–96.

Clark, W. A. V. (1992). Residential preferences and residential choices in a multiethnic context. *Demography, 29*(3), 451–466. https://doi.org/10.2307/2061828

Cocola-Gant, A., & Gago, A. (2021). Airbnb, buy-to-let investment and tourism-driven displacement: A case study in Lisbon. *Environment and Planning A: Economy and Space, 53*(7), 1671–1688. https://doi.org/10.1177/0308518X19869012

Cullingford, D., Flynn, P., & Webber, R. (1975). *Liverpool social area study: Interim report (1966 data)*. Planning Research Applications Group.

Dalton, C. M., & Thatcher, J. (2015). Inflated granularity: Spatial "big data" and geodemographics. *Big Data & Society, 2*(2), Article 2053951715601144. https://doi.org/10.1177/2053951715601144

De Sabbata, S., & Liu, P. (2019). Deep learning Geodemographics with autoencoders and geographic convolution. In *Proceedings of the 22nd AGILE Conference on Geographic Information Science, Limassol, Greece*.

De Sabbata, S., & Liu, P. (2023). A graph neural network framework for spatial geodemographic classification. *International Journal of Geographical Information Science, 37*(12), 2464–2486. https://doi.org/10.1080/13658816.2023.2254382

Delmelle, E. C. (2015). Five decades of neighborhood classifications and their transitions: A comparison of four US cities, 1970–2010. *Applied Geography, 57*, 1–11. https://doi.org/10.1016/j.apgeog.2014.12.002

Delmelle, E. C. (2022). GIScience and neighborhood change: Toward an understanding of processes of change. *Transactions in GIS, 26*(2), 567–584. https://doi.org/10.1111/tgis.12871

Department of Environment. (1977). *Inner area Studie: Liverpool, Birmingham and Lambeth. Summaries of consultants final reports*. HMSO.

Devos, L., Lippens, L., Lens, D., Rycx, F., Volral, M., & Baert, S. (2025). Labour market disadvantages of citizens with a migration background in Belgium: A systematic review. *De Economist, 173*(1), 121–175. https://doi.org/10.1007/s10645-024-09443-5

Duckham, M., Mason, K., Stell, J., & Worboys, M. (2001). A formal approach to imperfection in geographic information. *Computers, Environment and Urban Systems, 25*(1), 89–103. https://doi.org/10.1016/S0198-9715(00)00040-5

Ericksen, E. G. (1949). Review of review of the social areas of Los Angeles: Analysis and typology., by E. Shevky & M. Williams. *American Sociological Review, 14*(5), 699. https://doi.org/10.2307/2086667

Farr, M., & Evans, A. (2005). Identifying 'unknown diabetics' using geodemographics and social marketing. *Interactive Marketing, 7*(1), 47–58. https://doi.org/10.1057/palgrave.dddmp.4340504

Farr, M., & Webber, R. (2001). MOSAIC: From an area classification system to individual classification. *Journal of Targeting, Measurement and Analysis for Marketing, 10*(1), 55–65. https://doi.org/10.1057/palgrave.jt.5740033

Feldman, M., Friedler, S., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2014). Certifying and removing disparate impact (version 3). arXiv. Doi: 10.48550/ARXIV.1412.3756.

Fisher, P., & Tate, N. J. (2015). Modelling class uncertainty in the geodemographic output area classification. *Environment and Planning B: Planning and Design, 42*(3), 541–563. https://doi.org/10.1068/b130176p

Fontes, C., Al Haddad, C., Antoniou, C., & Lütge, C. (2024). Assessing the collective challenges of AI: An applied framework focused on people centeredness and public participation. *Urban Governance, 4*(4), 245–254. https://doi.org/10.1016/j.ugj.2024.12.003

Fournier, Q., & Aloise, D. (2019). Empirical comparison between autoencoders and traditional dimensionality reduction methods. In *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)* (pp. 211–214). https://doi.org/10.1109/AIKE.2019.00044

Gale, C. G., Singleton, A., Bates, A. G., & Longley, P. A. (2016). Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science, 12* (2016), 1–27. https://doi.org/10.5311/JOSIS.2016.12.232

Goodchild, M. F. (2001). Metrics of scale in remote sensing and GIS. *International Journal of Applied Earth Observation and Geoinformation, 3*(2), 114–120. https://doi.org/10.1016/S0303-2434(01)85002-9

Goodfellow, I., Courville, A., & Bengio, Y. (2016). *Deep learning*. The MIT Press.

Goss, J. (1995). Marketing the new marketing. The strategic discourse of geodemographic information systems. In J. Pickles (Ed.), *Ground truth* (pp. 130–170). Guildford Press.

Goss, J. (2003). The instrumental rationality of geodemographic systems. In D. Clarke (Ed.), *The consumption reader*. Routledge.

Gray, J., Buckner, L., & Comber, A. (2023a). Identifying neighbourhood change using a data primitive approach: The example of gentrification. *Applied Spatial Analysis and Policy, 16*(2), 897–921. https://doi.org/10.1007/s12061-023-09509-y

Gray, J., Buckner, L., & Comber, A. (2023b). Predicting gentrification in England: A data primitive approach. *Urban Science, 7*(2). https://doi.org/10.3390/urbansci7020064. Article 2.

Hale, R., & Austin, D. M. (1997). An exploratory factor model of social area analysis. *Sociological Spectrum, 17*(1), 115–125. https://doi.org/10.1080/02732173.1997.9982154

van Ham, M., Tammaru, T., Ubarevičienė, R., & Janssen, H. (2021). *Urban socio-economic segregation and income inequality: A global perspective*. Springer International Publishing AG.

Harris, C. D., & Ullman, E. L. (1945). The nature of cities. *The Annals of the American Academy of Political and Social Science, 242*, 7–17.

Harris, R., & Feng, Y. (2016). Putting the geography into geodemographics: Using multilevel modelling to improve neighbourhood targeting – A case study of Asian pupils in London. *Journal of Marketing Analytics, 4*(2), 93–107. https://doi.org/10.1057/s41270-016-0003-1

Harris, R., Johnston, R., & Burgess, S. (2007). Neighborhoods, ethnicity and school choice: Developing a statistical framework for geodemographic analysis. *Population Research and Policy Review, 26*(5), 553–579. https://doi.org/10.1007/s11113-007-9042-9

Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting*. Wiley.

Hawley, A. H., & Duncan, O. D. (1957). Social area analysis: A critical appraisal. *Land Economics, 33*(4), 337–345. https://doi.org/10.2307/3144311

Hayden, C., Williamson, T., & Webber, R. (2006). Schools, pupil behaviour and young offenders: Using postcode classification to target behaviour support and crime prevention programmes. *British Journal of Criminology, 47*(2), 293–310. https://doi.org/10.1093/bjc/azl053

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hirokane, S., & Amemiya, M. (2023). Describing the social changing pattern of Tsukuba Science City using social area analysis: Focusing on the decade from 2010 to 2020. *Reports of the City Planning Institute of Japan, 22*(2), 268–274. https://doi.org/10.11361/reportscpij.22.2_268

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. U. S. Government Printing Office.

Hughes, J. W., & Carey, G. W. (1972). Factorial ecologies: Oblique and orthogonal solutions. A case study of the New York SMSA. *Environment and Planning A: Economy and Space, 4*(2), 147–162. https://doi.org/10.1068/a040147

Hunter, A. A. (1972). Factorial ecology: A critique and some suggestions. *Demography, 9* (1), 107–117. https://doi.org/10.2307/2060548

Iceland, J. (2004). Beyond black and white. *Social Science Research, 33*(2), 248–271. https://doi.org/10.1016/S0049-089X(03)00056-5

Jin, C., Park, S., Ha, H. J., Lee, J., Kim, J., Hutchenreuther, J., & Nara, A. (2023). Predicting households' residential mobility trajectories with geographically localized interpretable model-agnostic explanation (GLIME). *International Journal of Geographical Information Science, 37*(12), 2597–2619. https://doi.org/10.1080/13658816.2023.2264921

Johnston, R. J. (1977). Principal components analysis and factor analysis in geographical research: Some problems and issues. *South African Geographical Journal, 59*(1), 30–44. https://doi.org/10.1080/03736245.1977.9713493

Johnston, R. J., Poulsen, M., & Forrest, J. (2007). The geography of ethnic residential segregation: A comparative study of five countries. *Annals of the Association of American Geographers, 97*(4), 713–738. https://doi.org/10.1111/j.1467-8306.2007.00579.x

Jung, P. H., & Song, J. (2022). Multivariate neighborhood trajectory analysis: An exploration of the functional data analysis approach. *Geographical Analysis, 54*(4), 789–819. https://doi.org/10.1111/gean.12298

Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems, 33*(1), 1–33. https://doi.org/10.1007/s10115-011-0463-8

Kinahan, K. L. (2021). Spatial–temporal neighborhood patterns in four legacy cities. *Urban Affairs Review, 57*(2), 492–526. https://doi.org/10.1177/1078087419897825

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

Kitchin, R. (2020). Urban science: Prospect and Critque. In K. S. Willis, & A. Aurigi (Eds.), *The Routledge companion to smart cities* (pp. 42–50). Routledge.

Knaap, E., Wolf, L., Rey, S. J., Kang, W., & Han, S. (2024). *The dynamics of urban neighborhoods: A survey of approaches for modeling socio-spatial structure.*. https://doi.org/10.31235/osf.io/3frcz

Kusner, M. J., Loftus, J. R., Russell, C., & Silva, R. (2017). Counterfactual fairness (version 3). arXiv. Doi: 10.48550/ARXIV.1703.06856.

Lees, L. (2000). A reappraisal of gentrification: Towards a 'geography of gentrification'. *Progress in Human Geography, 24*(3), 389–408. https://doi.org/10.1191/030913200701540483

Li, Q.-Q., Yue, Y., Gao, Q.-L., Zhong, C., & Barros, J. (2022). Towards a new paradigm for segregation measurement in an age of big data. *Urban Informatics, 1*(1), 5. https://doi.org/10.1007/s44212-022-00003-3

Libório, M. P., da Martinuci, O. S., Machado, A. M. C., de Lyrio, R. M., & Bernardes, P. (2022). Time–space analysis of multidimensional phenomena: A composite indicator of social exclusion through K-means. *Social Indicators Research, 159*(2), 569–591. https://doi.org/10.1007/s11205-021-02763-y

Liu, P., Zhang, Y., & Biljecki, F. (2024). Explainable spatially explicit geospatial artificial intelligence in urban analytics. *Environment and Planning B: Urban Analytics and City Science, 51*(5), 1104–1123. https://doi.org/10.1177/23998083231204689

Liu, Y., Singleton, A., & Arribas-Bel, D. (2019). A principal component analysis (PCA)-based framework for automated variable selection in geodemographic classification. *Geo-spatial Information Science, 22*(4), 251–264. https://doi.org/10.1080/10095020.2019.1621549

Longley, P. A. (2005). Geographical information systems: A renaissance of geodemographics for public service delivery. *Progress in Human Geography, 29*(1), 57–63. https://doi.org/10.1191/0309132505ph528pr

Longley, P. A., & Singleton, A. (2009). Classification through consultation: Public views of the geography of the E-society. *International Journal of Geographical Information Science, 23*(6), 737–763. https://doi.org/10.1080/13658810701704652

Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014). The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science, 17*(2), 85–101. https://doi.org/10.1080/10095020.2014.917453

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions (version 2). arXiv. Doi: 10.48550/ARXIV.1705.07874.

Maloutas, T. (2012). Contextual diversity in gentrification research. *Critical Sociology, 38* (1), 33–48. https://doi.org/10.1177/0896920510380950

Marr, T. R. (1904). *Housing conditions in Manchester & Salford*. Manchester University Press.

Martin, D., Gale, C., Cockings, S., & Harfoot, A. (2018). Origin-destination geodemographics for analysis of travel to work flows. *Computers, Environment and Urban Systems, 67*, 68–79. https://doi.org/10.1016/j.compenvurbsys.2017.09.002

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*(2), 281–315. https://doi.org/10.1093/sf/67.2.281

Mattoo, A., Neagu, I. C., & Özden, Ç. (2008). Brain waste? Educated immigrants in the US labor market. *Journal of Development Economics, 87*(2), 255–269. https://doi.org/10.1016/j.jdeveco.2007.05.001

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Montosa Muñoz, J. C. (2023). Application of the method of social areas and multivariate analysis in Peri-urban areas of Andalusia (Spain). *MOJ Ecology & Environmental Sciences, 8*(3). https://doi.org/10.15406/mojes.2023.08.00281

Moon, G., Twigg, L., Jones, K., Aitken, G., & Taylor, J. (2019). The utility of geodemographic indicators in small area estimates of limiting long-term illness. *Social Science & Medicine, 227*, 47–55. https://doi.org/10.1016/j.socscimed.2018.06.029

Musterd, S. (2005). Social and ethnic segregation in Europe: Levels, causes, and effects. *Journal of Urban Affairs, 27*(3), 331–348. https://doi.org/10.1111/j.0735-2166.2005.00239.x

Nadiyah, H. (2023). *A home for digital equity: algorithmic redlining and property technology*. https://doi.org/10.15779/Z38ZP3W18S

Noferesti, V., & Mirzahossein, H. (2025). Leveraging machine learning to predict residential location choice: A comparative analysis. *Results in Engineering, 25*, Article 104214. https://doi.org/10.1016/j.rineng.2025.104214

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A: Economy and Space, 16*(1), 17–31. https://doi.org/10.1068/a160017

Openshaw, S., Cullingford, D., & Gillard, A. (1980). A critique of the national classifications of OPCS/PRAG. *The Town Planning Review, 51*(4), 421–439.

Oshan, T. M., Wolf, L. J., Sachdeva, M., Bardin, S., & Fotheringham, A. S. (2022). A scoping review on the multiplicity of scale in spatial analysis. *Journal of Geographical Systems, 24*(3), 293–324. https://doi.org/10.1007/s10109-022-00384-8

Otley, A., Morris, M., Newing, A., & Birkin, M. (2021). Local and application-specific geodemographics for data-led urban decision making. *Sustainability, 13*(9). https://doi.org/10.3390/su13094873. Article 9.

Owens, A. (2012). Neighborhoods on the rise: A typology of neighborhoods experiencing socioeconomic ascent. *City & Community, 11*(4), 345–369. https://doi.org/10.1111/j.1540-6040.2012.01412.x

Park, R. E., Burgess, E. W., & McKenzie, R. D. (1925). *The City*. Chicago: University of Chicago Press.

Parker, S., Uprichard, E., & Burrows, R. (2007). Class places and place classes geodemographics and the spatialization of class. *Information, Communication & Society, 10*(6), 902–921. https://doi.org/10.1080/13691180701751122

Parliament. (1967). *House of Commons* (Vol. 741). Centre for Environmental Studies. Debated on Tuesday 21 February 1967 https://hansard.parliament.uk/Commons/1967-02-21/debates/5c1d7932-a11c-4a0f-98e6-c038678a6a03/CentreForEnvironmentalStudies?highlight=%22centre%20for%20environmental%20studies%22#contribution-40c0b6c6-e007-4184-a082-3fff4149adf3.

Parliament. (1976). *House of Commons* (Vol. 907). Centre For Environmental Studies. Debated on Wednesday 17 March 1976 https://hansard.parliament.uk/Commons/1976-03-17/debates/2b0e7c7c-ab8a-42ea-9f9f-fcaed748de93/CentreForEnvironmentalStudies?highlight=%22centre%20for%20environmental%20studies%22#contribution-758a4205-777c-435f-a957-f7d4e6d8b57a.

Peach, C. (1996). Good segregation, bad segregation. *Planning Perspectives, 11*(4), 379–398. https://doi.org/10.1080/026654396364817

Petersen, J., Gibin, M., Longley, P. A., Mateos, P., Atkinson, P., & Ashby, D. I. (2011). Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems, 13*(2), 173–192.

Pfautz, H. W. (1967). *Charles booth on the City: Physical pattern and social structure. Selected writings*. Chicago: University of Chicago Press.

Pickles, J. (Ed.). (1995). *Ground truth: The social implications of geographic information systems*. Guilford Press.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., … Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). https://doi.org/10.1145/3351095.3372873

Ramani, A., Alcedo, J., & Bloom, N. (2024). How working from home reshapes cities. *Proceedings of the National Academy of Sciences, 121*(45), Article e2408930121. https://doi.org/10.1073/pnas.2408930121

Rees, P. (1971). Problems of classifying subareas within cities. In B. J. L. Berry (Ed.), *City classification handbook: Methods and applications* (pp. 265–330). New York: Wiley-Interscience.

Reibel, M. (2011). Classification approaches in neighborhood research: Introduction and review. *Urban Geography, 32*(3), 305–316. https://doi.org/10.2747/0272-3638.32.3.305

Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 97–101). https://doi.org/10.18653/v1/N16-3020

Ringen, S. (1988). Direct and indirect measures of poverty. *Journal of Social Policy, 17*(3), 351–365. https://doi.org/10.1017/S0047279400016858

Robson, T. B. G. (2003). *London calling: The middle classes and the remaking of inner London* (1st ed.). Berg Publishers.

Salvati, L., Carlucci, M., & Serra, P. (2018). Unraveling latent dimensions of the urban Mosaic: A multi-criteria spatial approach to metropolitan transformations. *Environment and Planning A: Economy and Space, 50*(1), 93–110. https://doi.org/10.1177/0308518X17736313

Sanchez, T. W., Brenman, M., & Ye, X. (2025). The ethical concerns of artificial intelligence in urban planning. *Journal of the American Planning Association, 91*(2), 294–307. https://doi.org/10.1080/01944363.2024.2355305

Sangers, R., van Gemert, J., & van Cranenburgh, S. (2022). Explainability of deep learning models for urban space perception (version 1). *arXiv*. https://doi.org/10.48550/ARXIV.2208.13555

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). https://doi.org/10.1145/3287560.3287598

Shevky, E., & Bell, W. (1955). *Social area analysis*. Stanford University Press.

Shevky, E., & Williams, M. (1949). *The social areas of Los Angeles: Analysis and typology*. University of California Press.

Sieber, R. E., & Haklay, M. (2015). The epistemology(s) of volunteered geographic information: A critique. *Geo: Geography and Environment, 2*(2), 122–136. https://doi.org/10.1002/geo2.10

Silver, D., & Silva, T. H. (2021). A Markov model of urban evolution: Neighbourhood change as a complex process. *PLoS One, 16*(1), Article e0245357. https://doi.org/10.1371/journal.pone.0245357

Singleton, A. (2010). The geodemographics of educational progression and their implications for widening participation in higher education. *Environment and Planning A, 42*(11), 1560–2580. https://doi.org/10.1068/a42394

Singleton, A., & Longley, P. (2015). The internal structure of greater London: A comparison of national and regional geodemographic models. *Geo: Geography and Environment, 2*(1), 69–87. https://doi.org/10.1002/geo2.7

Singleton, A., & Spielman, S. E. (2014). The past, present, and future of geodemographic research in the United States and United Kingdom. *The Professional Geographer, 66* (4), 558–567. https://doi.org/10.1080/00330124.2013.848764

Singleton, A., Wilson, A. G., & O'Brien, O. (2012). Geodemographics and spatial interaction: An integrated model for higher education. *Journal of Geographical Systems, 14*(2), 223–241. https://doi.org/10.1007/s10109-010-0141-5

Singleton, A. D., Alexiou, A., & Savani, R. (2020). Mapping the geodemographics of digital inequality in Great Britain: An integration of machine learning into small area estimation. *Computers, Environment and Urban Systems, 82*, Article 101486. https://doi.org/10.1016/j.compenvurbsys.2020.101486

Singleton, A. D., & Longley, P. A. (2009). Geodemographics, visualisation, and social networks in applied geography. *Applied Geography, 29*(3), 289–298. https://doi.org/10.1016/j.apgeog.2008.10.006

Singleton, A. D., & Longley, P. A. (2024). Classifying and mapping residential structure through the London output area classification. *Environment and Planning B: Urban Analytics and City Science, 51*(5), 1153–1164. https://doi.org/10.1177/23998083241242913

Singleton, A. D., Pavlis, M., & Longley, P. (2016). The stability of geodemographic cluster assignments over an Intercensal period. *Journal of Geographical Systems, 18*(2), 97–123. https://doi.org/10.1007/s10109-016-0226-x

Skupin, A., & Hagelman, R. (2005). Visualizing demographic trajectories with self-organizing maps. *GeoInformatica, 9*(2), 159–179. https://doi.org/10.1007/s10707-005-6670-2

Somashekhar, M. (2021). Can we bring culture into the large-scale study of gentrification? Assessing the possibilities using geodemographic marketing data. *Urban Affairs Review, 57*(5), 1312–1342. https://doi.org/10.1177/1078087420921459

Spicker, P. (1990). Charles booth: The examination of poverty. *Social Policy & Administration, 24*(1), 21–38. https://doi.org/10.1111/j.1467-9515.1990.tb00322.x

Spielman, S., & Singleton, A. (2015). Studying neighborhoods using uncertain data from the American community survey: A contextual approach. *Annals of the Association of American Geographers, 105*(5), 1003–1025. https://doi.org/10.1080/00045608.2015.1052335

Spielman, S. E., & Thill, J.-C. (2008). Social area analysis, data mining, and GIS. *Computers, Environment and Urban Systems, 32*(2), 110–122. https://doi.org/10.1016/j.compenvurbsys.2007.11.004

Tammaru, T., Marcinczak, S., van Ham, M., & Musterd, S. (Eds.). (2019). *Socio-economic segregation in European capital cities: East meets west (first issued in paperback)*. Routledge.

Tapp, A., & Warren, S. (2010). Field-capital theory and its implications for marketing. *European Journal of Marketing, 44*(1/2), 200–222. https://doi.org/10.1108/03090561011008673

Thrift, N., & French, S. (2002). The automatic production of space. *Transactions of the Institute of British Geographers, 27*(3), 309–335. https://doi.org/10.1111/1475-5661.00057

Timms, D. W. G. (1970). Modernisation and the factorial ecology of the Cook Islands, Brisbane, and Auckland. *The Australian and New Zealand Journal of Sociology, 6*(2), 139–149. https://doi.org/10.1177/144078337000600206

Timms, D. W. G. (1971). *The urban Mosaic: Towards a theory of residential differentiation*. Cambridge University Press.

Tomlinson, M., Walker, R., & Williams, G. (2008). Measuring poverty in Britain as a multi-dimensional concept, 1991 to 2003. *Journal of Social Policy, 37*(4), 597–620. https://doi.org/10.1017/S0047279408002237

Tuccillo, J. (2021). *An individual-centered approach for geodemographic classification*. https://doi.org/10.25436/E2H59M

Uprichard, E., Burrows, R., & Parker, S. (2009). Geodemographic code and the production of space. *Environment and Planning A, 41*(12), 2823–2835. https://doi.org/10.1068/a41116

Vaughan, L. (2018). *Mapping society: The spatial dimensions of social cartography*. UCL Press.

Vickers, D., & Rees, P. (2007). Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*(2), 379. https://doi.org/10.1111/j.1467-985X.2007.00466.x

Vies, W. K. D. D. (1978). Alternative factorial solutions and urban social structure: A data analysis exploration of Calgary in 1971. *Canadian Geographies / Géographies Canadiennes, 22*(4), 273–297. https://doi.org/10.1111/j.1541-0064.1978.tb01523.x

Wachsmuth, D., & Weisler, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A: Economy and Space, 50*(6), 1147–1170. https://doi.org/10.1177/0308518X18778038

Wang, Y., Yao, H., & Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing, 184*, 232–242. https://doi.org/10.1016/j.neucom.2015.08.104

Webber, R. (1975). *Liverpool social area study, 1971 (no. PRAG technical paper 14)*. Studies, Centre for Environmental.

Webber, R. (1979). *Census enumeration districts: A socio-economic classification (no. Occasional paper 14)*. Office of Population Censuses and Surveys.

Webber, R. (1980). A response to the critique of the OPCS/PRAG national classifications. *The Town Planning Review, 51*(4), 440–450.

Webber, R. (2007). The metropolitan habitus: Its manifestations, locations, and consumption profiles. *Environment and Planning A, 39*(1), 182–207. https://doi.org/10.1068/a38478

Webber, R., & Burrows, R. (2018). *The predictive postcode: The geodemographic classification of British society* (1st ed.). SAGE.

Weiss, M. J. (1988). *The clustering of America*. Harper and Row.

Weiss, M. J. (2000). *The clustered world: How we live, what we buy, and what it all means about who we are*. Little, Brown and Company.

White, M. J. (1987). *American neighborhoods and residential differentiation (with National Committee for Research on the 1980 Census)*. Russell Sage Foundation for the National Committee for Research on the 1980 Census.

Williams, D. R., & Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports, 116*(5), 404–416. https://doi.org/10.1016/S0033-3549(04)50068-7

Williamson, T., Ashby, D. I., & Webber, R. (2006). Classifying neighbourhoods for reassurance policing. *Policing and Society, 16*, 189–218. https://doi.org/10.1080/10439460600662197

Wyszomierski, J., Longley, P. A., Singleton, A. D., Gale, C., & O'Brien, O. (2022). A neighbourhood output area classification from the 2021 and 2022 UK censuses. *The Geographical Journal*. https://doi.org/10.1111/geoj.12550

Yan, B., Janowicz, K., Mai, G., & Gao, S. (2017). From ITDL to Place2Vec: reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In *, 35. Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 1–35), 10.

Yang, Y., Dolega, L., & Darlington-Pollock, F. (2023). Ageing in place classification: Creating a geodemographic classification for the ageing population in England. *Applied Spatial Analysis and Policy, 16*(2), 583–623. https://doi.org/10.1007/s12061-022-09490-y