Geodemographics: A Literature Review

Peter Prescott

Contents

List of Tables

List of Figures

1. Introduction

This paper offers a critical introduction to *geodemographic analysis*, reviewing the relevant literature so as to understand the current state of research, to identify unresolved questions and issues requiring further study, and to develop an agenda of priorities for this doctoral thesis.

We begin by setting our study within the context of the ubiquitous practice of locational inference. Next, we introduce geodemographics itself, offering a description and explanatory etymology. We then review the historical antecedents of geodemographic analysis in more detail: the pioneering visualizations of nineteenth-century urban thematic cartography; the rich investigative agenda of the early Chicago School's human ecology; and the shift to quantitative social area analysis and factorial ecology. From there, we examine the subsequent development and discussion of geodemographics within British academia, finding two principal components in the debate which continue to generate heated controversy: the ethical dimension of social surveillance on the one hand, and the ontological dimension of neighbourhood definition on the other. Consideration of the first leads us to appreciate the importance of free and open source geodemographic classifications, which have now become well-estalished; consideration of the second leads us to a survey of the intrinsic problems of the areal zone design, and a deeper consideration of neighbourhood dynamics and ontology. We conclude by summarizing the specific research goals resulting from our review of the literature.

2. The Ubiquity of Locational Inference

On meeting a person for the first time, it is common to ask *'Where are you from?'* (Myers, 2006). Knowing something about the place with which a person is associated reveals something about that person – such is the implicit and necessary logic underlying this question, and its commonplace conventionality reveals it to be a very widespread conviction. It may then turn out that the place is known directly to the one asking, in which case it may be considered in its unique specificity. More likely however, it will not: the contextual understanding gained from the answer must then be mediated by an attempt to fit the reply into some more generalized classificatory framework. This will enable the construction of analogies with other places of which the enquirer does have direct experience, and the conversation can progress. Such a process will more likely happen instinctively than reflectively, but it inevitably must happen – even if the (possibly subconscious) mental function to find an appropriate known category for the place in question returns an error.

Indeed, before the question 'Where are you from?' is verbally uttered, a provisional estimate of the answer will already have been made, on the basis of visual, auditory, and even olfactory clues. Skin tone, eye colour, facial features, and other physical characteristics reflect genetic variations optimized over millenia for particular environments (Bamshad & Olson, 2003). Accent and vocabulary mark the region in which the speaker learned to speak a language (Boland, 2010; Weatherhead et al., 2016). Clothing and other worn accessories differ according to what is culturally fashionable and commercially viable in different places (Crewe, 2017); so do perfumes, deodorants, and other noticeably fragrant grooming products (Havlíček & Roberts, 2013). The lingering smell of certain spices is suggestive of a diet shaped by a family background in a place where such ingredients were inexpensively available (Mallapragada, 2016).

But such qualitative interpretation of individual locational clues is inherently time-consuming (Cypress, 2019) and subjective (Kalu, 2019).

3. Geodemographics: The Analysis of People By Where They Live

Geodemographic analysis provides a quantitative approach to the "analysis of people by where they live" (Sleight, 1997; Harris et al., 2005; Webber & Burrows, 2018). By applying unsupervised machine learning to the demographic data associated with geographic areas, the complex multi-dimensional reality of human society can be reduced to a more manageable number of statistical types. Having been identified algorithmically, these *clusters* can then be described qualitatively (Tbl. 1) and presented visually (Fig. 1), creating products that have been used with great success in fields ranging from direct marketing (Evans, 1998), retail location selection (González-Benito & González-Benito, 2005), political campaigning (Robbin, 1980; Webber, 2006), and military recruitment (DeReu & Robbin, 1981), to social service resource allocation (Longley, 2005) : whether that service be in the field of health (Farr et al., 2008), education (Singleton & Longley, 2009), or policing (Ashby & Longley, 2005).

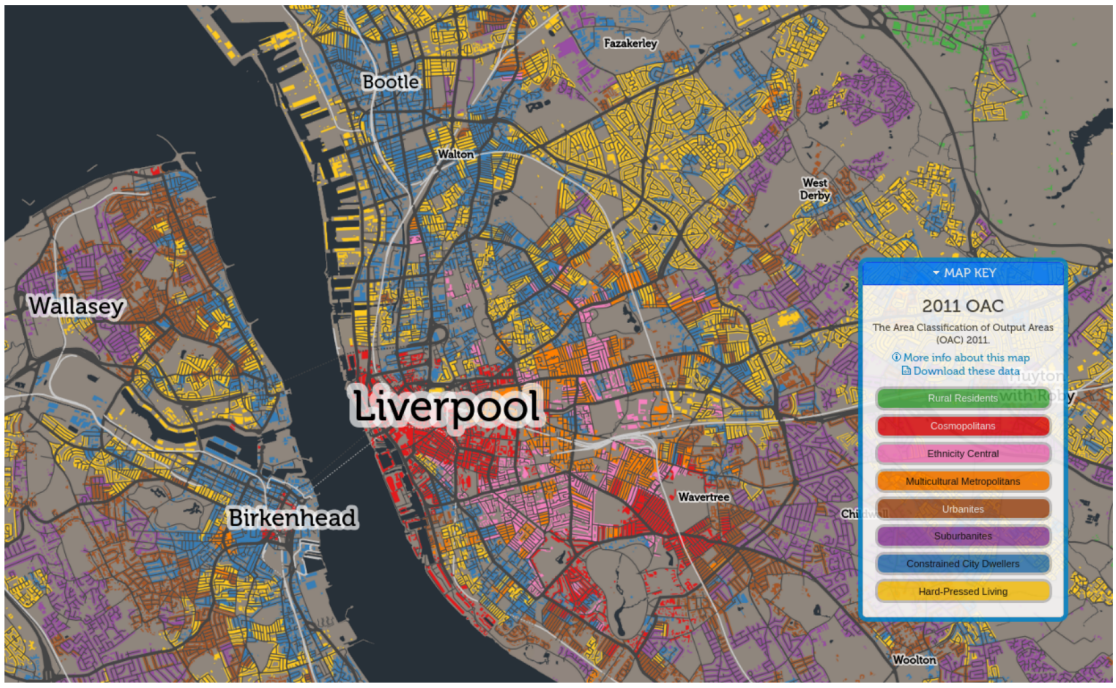

Figure 1: *An Example of Geodemographic Visualization: Choropleth Map showing 2011 Open Area Classification, after CDRC (2021)*

Table 1: *An Example of Geodemographic Cluster Description: Groups and Subgroups from the 2011 OAC, created by Gale et al. (2016)*

Supergroup	Group
Rural residents	Farming communities
	Rural tenants
	Ageing rural dwellers

—

Supergroup	Group
Cosmopolitans	Students around campus
	Inner city students
	Comfortable cosmopolitan
	Aspiring and affluent
Ethnicity central	Ethnic family life
	Endeavouring ethnic mix
	Ethnic dynamics
	Aspirational techies
Multicultural metropolitans	Rented family living
	Challenged Asian terraces
	Asian traits
Urbanites	Urban professionals and families
	Ageing urban living
Suburbanites	Suburban achievers
	Semi-detached suburbia
Constrained city dwellers	Challenged diversity
	Constrained flat dwellers
	White communities
	Ageing city dwellers
Hard-pressed living	Industrious communities
	Challenged terraced workers
	Hard-pressed ageing workers
	Migration and churn

The word 'geodemographics' was coined by Jonathan Robbin (1980) to describe the marketing tool his company had developed (Webber & Burrows, 2018, p. 94). By classifying American residential zip codes into groups with similar demographic characteristics, and then giving each group a memorable label and summary description, he had created a product designed to simplify the process of targeting prospective customers and selecting promising retail locations (DeReu & Robbin, 1981). Robbin's tool combined the latest in marketing theory with cutting-edge methods in quantitative geographic sociology.

Market segmentation (Smith, 1956) solved the problem of a complex market of heterogenous customers by dividing it into several sub-markets of homogenous customers. *Demographics* – that is, population attributes such as age, sex, income, and ethnicity – offer a straightforward way of applying this strategy. The idea of *psychographics* (Wells, 1975) is then to understand the psychology of a typical customer from a given market segment, so as to anticipate their needs, desires, and trigger points. Robbin added these techniques to the *social area analysis* (Shevky & Bell, 1955) he had been applying as a doctoral candidate at New York University (Ricercar, 2021),

assisting Edgar Borgatta in researching ways of classifying the social characteristics of American cities (Hadden & Borgatta, 1965).

While clustering census tracts provides a way of segmenting them into similar groups, the *ecological fallacy* (Robinson, 1950) means it does not follow that individuals within tracts with similar population demographics are necessarily similar at the individual level. The argument for the likely homogeneity of census tract units is made by reference to the account of the *neighbourhood* as a *natural area* given by Park (1925), who argued that the twin forces of homophily and social influence will tend to segregate an urban population into a "mosaic of social worlds" (Wirth, 1938). Park was the central figure of the *Chicago School* (Abbott, 2017a), whose distinctive *human ecology* combined the empirical investigation exemplified by Booth (1904) with the more theoretical sociology of Simmel (1908), with whom he had studied in Germany.

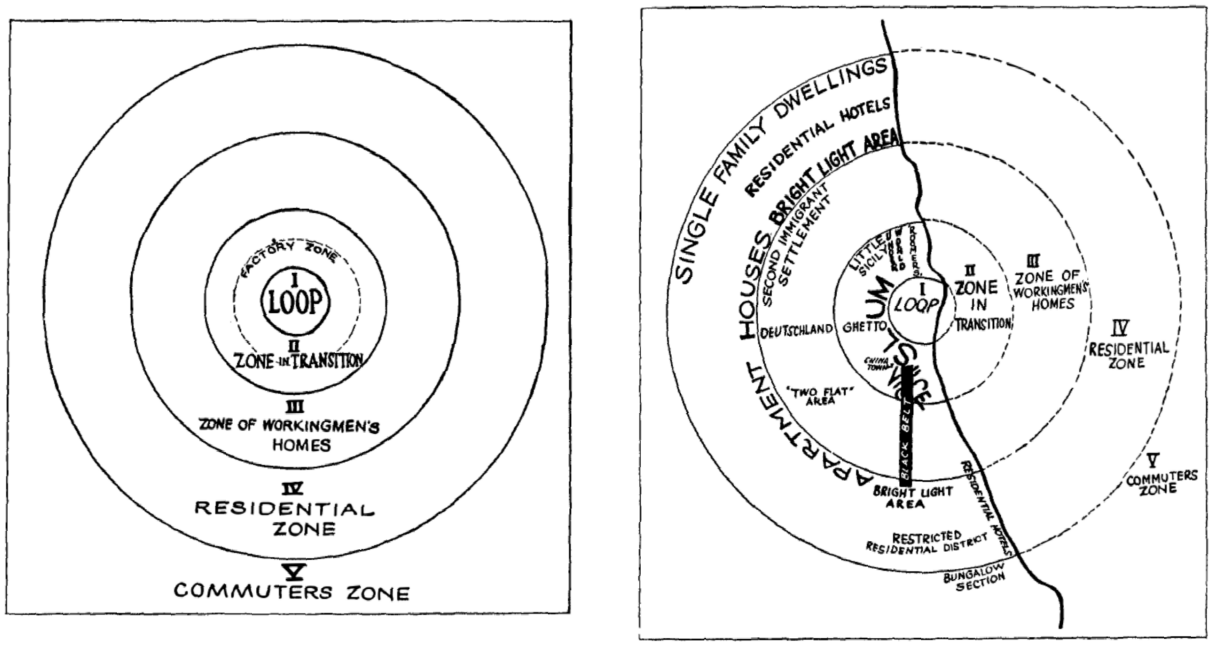

CHART I. The Growth of the City

CHART II. Urban Areas

Figure 2: *Concentric Zone Diagrams, after Burgess (1925)*

This account of geodemographic historiography is expressed in detail in the collaborative monographs of Harris et al. (2005) and Webber & Burrows (2018). Booth's urban poverty maps offer "the first example of applied geodemographics" (Harris et al., 2005, p. 30), the Chicago School's theory of urban *natural areas* (Fig. 2) provides the necessary "conceptual definition…[for] neighbourhood analysis" [p.39], the increasing availability of census data then stimulates the development of quantitative social area analysis and subsequent factorial ecology [pp.39-40], paving the way for the emergence of geodemographic products [p.55].

4. The Historical Roots of Geodemographics

Modern *demography*, the science of enumerating populations, began with Graunt (1662) and his realization that London's mortality bills and christening records could be used as data sources for an empirical analysis of London's population growth (Glass, 1963). Graunt's demographic

analysis was geographic, insofar as it dealt with data records from different parishes, and suggested the role of migration from rural parishes to London as a reason for the apparent ability of the metropolis to sustain a higher rate of deaths than births (Taeuber, 1941). Graunt's work inspired mathematicians such as Huygens, Leibniz, and Bernoulli in their development of classical probability theory (Kreager, 1993). But his originality was in the way he saw the socio-economic and political significance of numbering population (Bayatrizi, 2008).

The most straightforward way to establish the number of a population is simply to count it directly. Parliament debated the idea of a national census in 1753, but the bill did not pass; it was seen as an unprecedented intrusion into private matters, and some argued it was "subversive of the last remains of English liberty" (Robbins, 2010, p. 202). However, when the American colonies became independent, and established themselves as a republican democracy, a decennial census was made a constitutional requirement; a corallary of the need "to ensure the civic empowerment of each individual" (Sullivan, 2020). The first census of the USA was taken in 1790 (Anderson, 2015), and Europe followed suit shortly thereafter: Britain in 1801 (Glass, 1978; Levitan, 2011), and France soon afterwards (Cole, 2000).

The new era of (nineteenth-century) 'big data,' meant Fletcher (1847) could combine "the last Census of the Population; the Income Tax Returns; the Home Office Tables of Criminal Offenders; the latest Reports of the Poor Law Commissioners; and a Summary of Savings Banks" (p.194) to produce a Comparison of the different Districts of England and Wales, in respect to the Distribution of the Population, its Social Organization, its Education and Ignorance, its Providence and Improvidence, and its Crime, in which each county has been classified "on the evidence of the occupation abstract of 1841, with a further reference to their geographical contiguity, and the different origin of the great body of their population." Within each district, counties were also classed with regards to educational attainment. This sociodemographic analysis is then visualized by a map of England (and Wales) in which each county in coloured to show how it has been classified (Fig. 3). Two years later he expanded the work to included twelve shaded maps for each variable independently (Funkhouser, 1937), apparently motivated by a comment from Prince Albert (Wainer, 2012).

At the smaller scale of the neighbourhood, a Church of England minister in Liverpool called Abraham Hume (Vaughan, 2018, p. 62) organized a detailed door-to-door survey of the social condition of his parishioners (Hume, 1850, 1858a), which he presented on a large coloured map (Hume, 1858b), showing simultaneously poverty, crime, health, church and school data on a vertical plan of the city (Fig. 4). Poverty is shown by street, according to whether the street is classed as 'wholly pauper' or partially so, as assessed by "the relieving officers within the Borough" (p.21).

Hume is aware that a more precise analysis would be made possible "if we could compare *houses, families, or individuals* relieved" (p.22) Still, when Tobias (1974) applied modern mathematical methods to Hume's data (examining p-value scores, chi-squared tests and correlation coefficients) he found that "Hume's impressionistic descriptions have been shown to be consistent with the data he collected but doubtless never analysed in this way" [p.229]. Hume is also interested in patterns of residential mobility and their effects on neighbourhood change; in particular, the way that congregational Nonconformist churches (in contrast to the territorial Established Church), tend to follow the paths trod by the middle classes from the centre of a town to the suburbs (Pickering, 1972, p. 42; Wilcox, 2014, p. 12).

Although the claim is widespread that Charles Booth's maps are the first example of geode-

Figure 3: *England's Manufacturing Districts, after Fletcher (1847)*

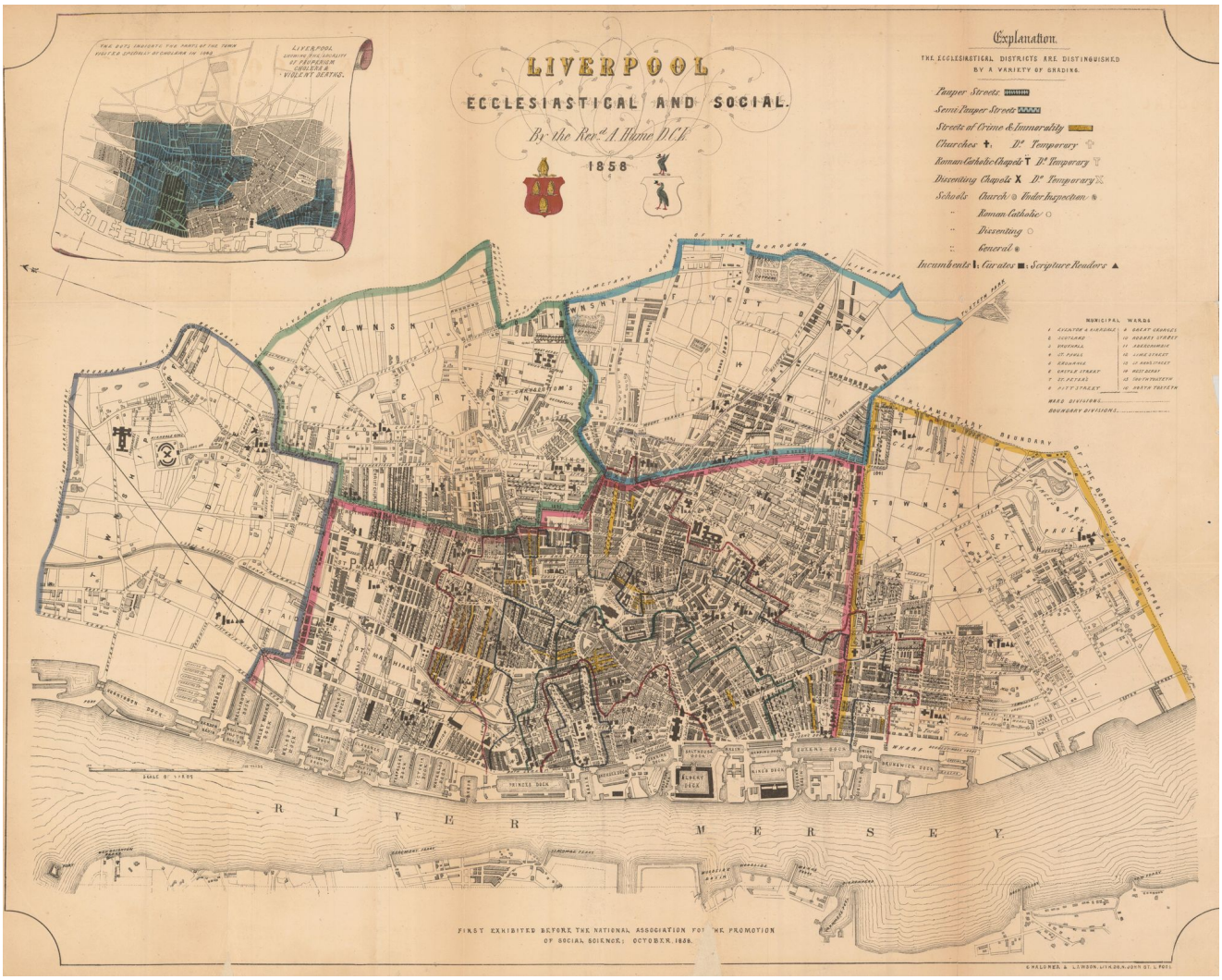

Figure 4: Ecclesiastical and Social Map of Liverpool *after Hume (1858b)*

mographic visualization (Harris et al., 2005, p. 30; Burrows & Gane, 2006, p. 794; Dalton & Thatcher, 2015, p. 3; Webber & Burrows, 2018, p. 31), we have seen that typological neighbourhood mapping in fact preceded Booth by more than half a century (Robinson, 1982, p. 184; Vaughan, 2018, p. 62). To state this takes nothing away from Booth's magisterial work, which we can now consider in more detail.

Charles Booth was a wealthy Liverpool shipowner and enthusiastic member of the Royal Statistical Society, which he served for a time as president (Booth, 1893). He was also an earnest Unitarian, and felt a sense of personal responsibility for the wretched conditions of poor labourers (Collet, 1945). In 1887, Booth began work on a "proposed inquiry into the condition and occupations of the people of London" (Booth, 1887), motivated perhaps by a "desire to prove the Socialists wrong in their estimate of the extent of poverty in London" (Pfautz, 1967, p. 21). Harris et al. (2005, p. 30) follow Pfautz (1967, p. 21) and Simey & Simey (1960, p. 69) in accepting the claim of Hyndman (1911, p. 303) that he was a causal influence; however O'Day & Englander (1992, pp. 30–31) disagree, arguing that althoguh Hyndman's conclusions may have disturbed Booth, "they certainly did not provide the main impetus for Booth's survey." Over the next seventeen years, Booth and his research team filled 450 notebooks with information about the capital's living and working conditions, which were published in three printed editions, the final running to seventeen volumes (Donnelly, 2002). Funding for the project came from the profits of Booth's own commercial success (Donnelly, 2002).

Booth and his team drew on extensive interviews with the house-to-house visitors employed by the School Board (Booth, 1887), which they combined with anonymized data from the 1891 Census (Booth, 1893). This information then allowed each street block (Vaughan & Geddes, 2009) to be allocated to one of his eight levels of deprivation. These in turn were colour-coded according to their relative level of poverty, with darker colours representing the poorer residential areas, although the precise correspondence between class and colour turns out on close inspection to be somewhat ambiguous (Kimball, 2006).

Booth expressed his doubts that he could make his figures "as luminous and picturesque to any other eye as they are to mine" [Booth (1904), p.63), but for some he certainly succeeded. The map (Fig. 5) received intense media attention when it was first put on public display (Bales, 1999), and his methodology influenced other reform-minded social researchers, such as Addams (1895), Du Bois (1899), Rowntree (1902), and Marr (1904).

On the other hand, Booth was sidelined by the academic institutionalization of sociology, marked in Britain by the appointment of a first Professor of Sociology in 1904 at the London School of Economics (Topalov, 1993). When compared with the strongly theory-driven work of continental sociologists such as Durkheim and Weber, Booth's "uncontrolled empiricism" has been judged to lack "clear objectives and ideas" and "sociological imagination" (Westergaard, 1969). Half a century later, Simey & Simey (1960) attempted to position Booth as the founding father of British sociology, but this was arguably driven primarily by a desire to locate a respectable pedigree for their subject (O'Day & Englander, 1992). The claim that Booth invented the concept of 'the poverty line' (Simey & Simey, 1960, p. 88) also fails under scrutiny; Gillie (1996) shows that it must rather have come directly from the London School Board, whose home-visitation workers supplied Booth's project with information about families' living conditions, and which was required by the implications of the 1870 Education Act to develop a criterion of poverty.

Booth's statistical enthusiasm also lacked mathematical expertise (Selvin, 1976). The physicist Yule (1895) was provoked "to severely criticize the [statistical] methods on which he bases his

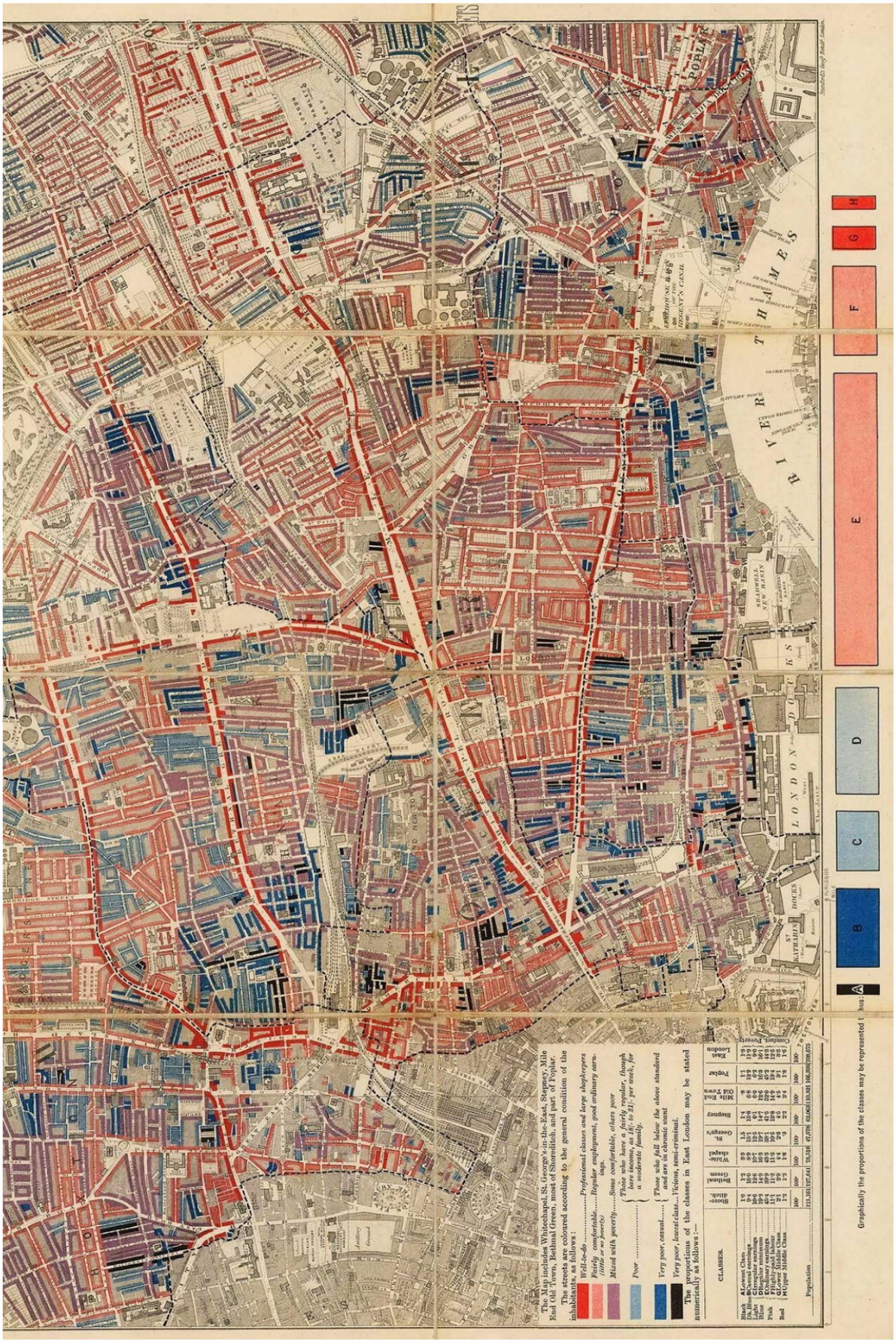

Figure 5: *Close-up from* Descriptive Map of London Poverty *after Booth (1889)*

results" (p.608), showing repeatedly (Yule gives sixteen specific examples) that Booth claims his figures show no definite relation, when in fact a strong numerical correlation is present.

It is also worth noting explicitly that while Booth's research was multivariate in its input (Davies, 1978), its output was not a true multidimensional classification of different *types* of residential neighbourhood, but rather a unidimensional index of poverty separated into quantiles. In that respect it is more similar to the modern Index of Multiple Deprivation (Deas et al., 2003) than a modern geodemographic classification. This meant that Booth's analysis was ill-equipped to deal with communities differentiated by factors other than his economic index, and it is thus unsurprising that he struggled to fit Jewish immigrants into his schema (Englander, 1989).

However, though Booth may have been sidelined by the development of academic sociology in Britain, his style of detailed social survey had somewhat more influence on its development in America, particularly through the Chicago School.

The phrase 'Chicago School' implies an institution, a group of people, a body of work, a set of ideas – and a period of time during which those entities were, if not isomorphic, then at least functionally equivalent. The reality of such an equivalence has been critically deconstructed by Harvey (1987) and Cortese (1995). In fact, the phrase seems to have first been used by Bernard (1930) as a way of dismissing the 'Park Chicago School' for being in his view too much a personal following of Robert Park, rather than having "science as the central interest" (p.133), and being inclined "to regard case studies as superior to statistical studies" (p.127). Nevertheless, Abbott (2017b) argues, after a comprehensive historiographical survey, that the enduring interest in and quantity of writing on the 'Chicago School' at least "suggests that there was, in fact, something real behind the phrase."

The institution, the University of Chicago's Department of Sociology was established in 1892. It was the first separate university department focused on the subject, at a time when American universities were transitioning to being primarily focused on research rather than merely teaching (Turner, 1988). It was initially headed by Albion Small, who founded the *American Journal of Sociology*, and in his opening paper (1895) is at pains to distinguish the scientific understanding "of associated human activites as a whole" which the Journal hopes to promote, from the "palliatives" of social reformers.

Though founded by Albion Small, it is Robert Park who is generally credited with shaping the distinctive urban emphasis of the University of Chicago's Department of Sociology. Park was already fifty years old when he joined the faculty of Chicago in 1914, having worked previously as a journalist (Lindner, 1996), and as an assistant to the Black American spokesman Booker T. Washington (Fenton, 1981). A year later Park (1915) published the first version of an article that would later be reprinted as the first chapter of a collective volume by members of the Chicago School defining their vision for research into *The City* (Park et al., 1925).

Park (1925) begins by considering 'The City Plan,' but in his account urban planning turns out to be largely a backdrop for more specifically social processes. Although a city plan may impose a certain order upon the built environment of the city, Park is primarily interested in "the inevitable processes" of personal tastes, vocational and economic interests, which tend to segregate the population of a city (cf. Schelling, 1969). Thus over time "every section and quarter of the city takes on something of the character and qualities of its inhabitants," transforming "what was at first a mere geographical expression into a neighborhood" (p.6).

He then turns to 'The Neighborhood' specifically, that "simplest and most elementary form"

of urban society, based on "proximity and neighborly contact" (p.7). While for some areas, "the easy means of communication and transportation" of modern urban life "tend to destroy the permanancy and intimacy of the neighborhood," it can equally be the case, particularly "where there is racial prejudice," that population segregation can intensify the "intimacies and solidarity" of the local neighbourhood group (p.9).

The volume includes E. Burgess's famous Concentric Zone diagram (Fig. 2), an essay by R. McKenzie explaining the Chicago School's distinctive 'The Ecological Approach,' and a comprehensive bibliography (credited to Louis Wirth) (in which Booth's *Life and Labour…* is highlighted as "Especially interesting… for its description of the natural areas of [London]"). But more than half of the chapters come from Park, and it is clear that the questions of his 'Suggestions' for investigating modern urban society drive the book. Indeed, a century on, they remain fresh and stimulating; Karvonen (2020) calls it "essential reading… because it establishes a research agenda that continues to be an inspiration."

While Burgess' Concentric Zone model has remained the dominant visual image of the Chicago School's urban research, closer investigation shows that this diagram, based on the theory of the German agricultural economist Von Thunen (1826), actually had limited substantive influence, as they moved towards more empirically grounded ecological maps and census tract maps (Owens, 2012) – for example, see Fig. 6.

The American Census Bureau first reported statistics by tract for the 1920 census (Foley, 1953), beginning with ten cities and expanding by the 1950s to seventy-two. The first census tract maps were produced (for Chicago) by Burgess & Newcomb (1931), and a few years later (for St Louis) by Fletcher et al. (1935). Knaap et al. (2019) note that census data meant that "social processes, which are difficult to observe [using social survey methods], could be treated as latent variables and modeled… easily."

Shevky & Bell (1955) offered a method of classifying a census tract by reducing the attributes of its census data into a simplied expression of three factors, 'social rank,' 'urbanization,' and 'segregation,' which they suggested both retained the important details of the data, and corresponded to the essential nature of contemporary society. Each datapoint could then be visualized by a small circle in a two-dimensional scatter-plot, with the attributes along the x and y axes corresponding to the two more significant factors, and the circle's colour the third. The datapoints can then be divided up according to their position, and since their positional proximity is a function of their statistical similarity, census tracts with datapoints in the same segment can be considered as being of the same type. This was done first for Los Angeles (Shevky & Williams, 1949), and then San Franciso (Bell, 1953, Fig. 7), before being advocated more broadly as a method of analyis with general applicability to the social analysis of American cities (Shevky & Bell, 1955).

Tryon (1968) achieved a more sophisticated way of grouping census tract datapoints, showing how his *cluster analysis* (Tryon, 1939) could identify 'clusters' of similar datapoints, avoiding the arbitrariness of simply segmenting the attribute space by intervals. Instead these could be detected by an iterative computational algorithm, made available in reproducible FORTRAN code (Tryon & Bailey, 1966).

But initially it was the other aspect of the methodology of Shevky and his associates that saw more widespread adoption, the *factor analysis* of the principal components of a social data matrix, and the *factorial ecology* (Rees, 1971) which sought to draw insight into social processes from the

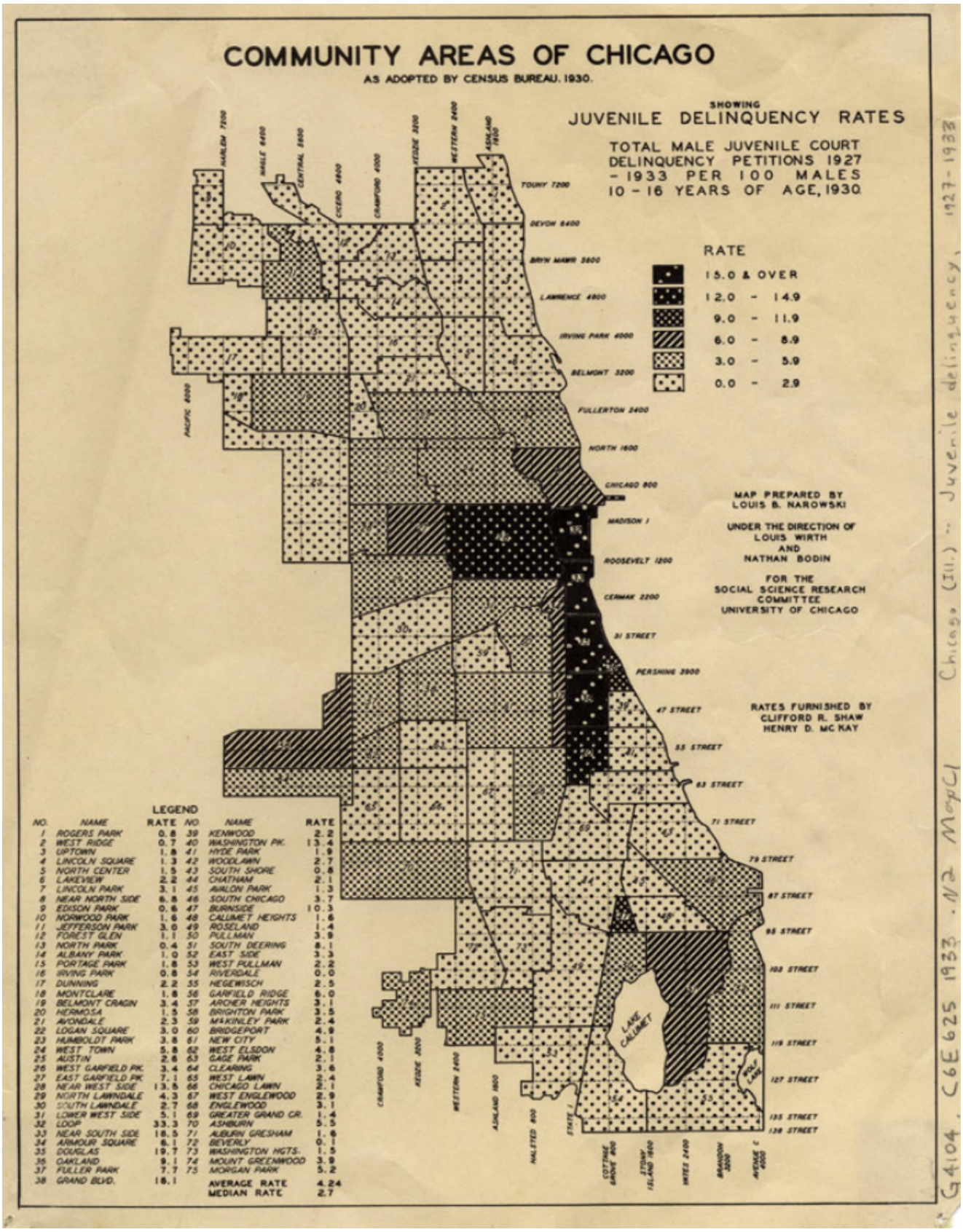

Figure 6: *Choropleth Map showing Chicago Delinquency Rates in 1930*

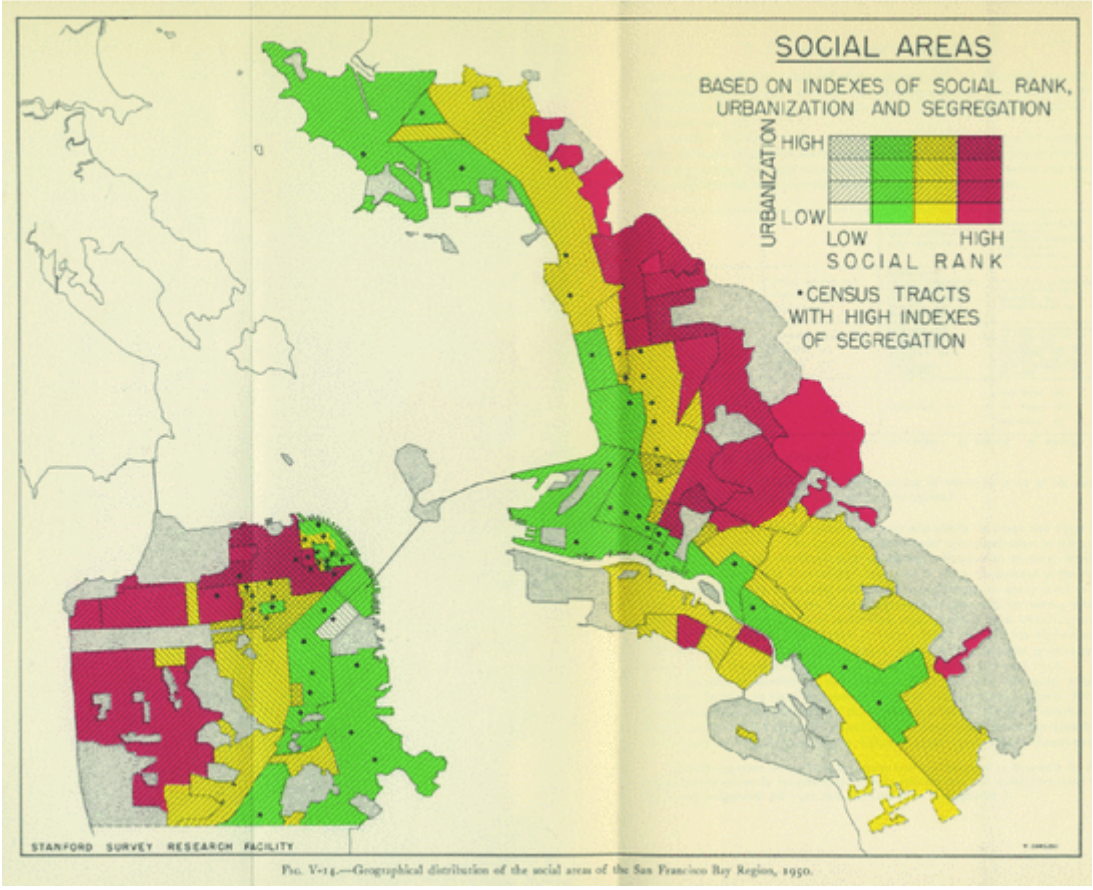

Figure 7: Social Area Analysis *of the San Francisco Bay Region, after Shevky & Bell (1955)*

application of such algebraic analysis to the quantitative data describing residential populations. Such studies of American cities include those of Boston (Sweetser, 1965), Newark (Janson, 1974), Seattle (Schmid, 1960), and San Francisco (Tryon, 1955); and the technique was also tested in the contexts of other cultures, in Calcutta (Berry & Rees, 1969), Cairo (Abu-Lughod, 1969), Canberra (Jones, 1965), Helsinki (Sweetser, 1965), and Sunderland (Robson & Robson, 1969).

Timms (1971),p.55 notes that "the results of a factor analysis vary not only with the nature of the data input and the particular type of factor analytic technique employed, but also with the theoretical predilections of the investigators," but is nevertheless impressed that for all that the residential differentiation of the majority of cities seems to be a reflection of underlying factors similar to those identified by Shevky.

5. The Development of Geodemographics and Ongoing Debates

Having reviewed the historical antecedents of geodemographic analysis we now survey its continued development. While Robbin coined the term 'geodemographics' and successfully turned it into a profitable commercial product, his direct impact on the academic understanding of neighbourhoods was quite limited. He dropped out of his PhD without completing it, after his supervisor left to take up a post at another university (Ricercar, 2021), and it has subsequently been the case in North America that, as a method for studying neighbourhood dynamics and effects, "many academic social scientists ignore geodemographics" (Reibel, 2011, p. 310).

In Britain however, the situation is quite different, largely because of the different circum-

stances of its separate development by Richard Webber. Independently from Robbin, Webber was also applying cluster analysis to census data; first, for the purpose of helping the Liverpool City Council identify priority areas for social service provision (1975), and then on a national scale (1978). Although Webber did go on to work in the marketing industry, the initial context of his research for a public authority meant that his work was published openly, rather than being a proprietary secret. There has then been a continued tradition of free, national, *open* (Singleton et al., 2016) geodemographic typologies of Britain, produced using the data from the 1981 (Charlton et al., 1985), 1991 (Blake & Openshaw, 1994), 2001 (Vickers & Rees, 2007), and 2011 (Gale, 2014; Gale et al., 2016) censuses.

This has allowed robust academic debate about the validity and utility of geodemographic typologies. Openshaw & Gillard (1978) showed the instability of clustered classifications by demonstrating their sensitivity to subjective decisions at multiple points in their construction, including "the selection of variables, the choice of algorithms and methods, and various data management operations" (p.101), and concluded that such classifications in general, and Webber's 1975 study in particular, "should not be used until they can be replicated at the individual level" (p.118). Openshaw et al. (1980) repeated the warning "for all users to be aware of the practical limitations of [Webber's subsequent national] Classifications… [as] they are unsuitable or many of the applications that have been suggested" (p.438). Webber apparently was apparently unaware of the first critique, but the second quickly provoked a thorough rebuttal, in which Webber (1980) concluded that his critics were not "in touch with either public policy or the commercial world" (p.449). Presumably, the point was well taken, as Openshaw (1985) then applied census-data cluster classification to rural areas, complaining of the resistance of government departments to such methods, and of their "preference for old fashioned pre-computer age techniques" (p.286).

Openshaw (1997) then found himself in another notable controversy concerning geodemographics, this time offering a defense, when he offered a widely-ranging response to the various criticisms of Geographic Information Systems gathered by Pickles (1995). The point which stands out as having the most continued validity in his discussion is his suggestion that "data protection legislation" can mitigate some of the dangers of unfettered technological surveillance, foreseeing the need for the sort of data protection legislation now established by statutes such as GDPR (2016).

In our contemporary context, Gilbert (2021) offers another helpful response to the suggestion that gathering and profiting from people's data is a sort of exploitative 'data colonialism' (Thatcher et al., 2016), in which an intrinsically valuable resource is being unfairly extracted. Rather than being 'the new oil' (Spijker, 2014), Gilbert suggests that a better metaphor for big data would in fact be 'the new manure': "a mundane by-product of life" [p.36], which, like manure that is processed into fertiliser, only has economic value because there are businesses that have invested in processing it into something useful.

We cannot in this paper attempt a comprehensive evaluation of Gilbert's attempt to defend big data analytics and develop a positive account of *digital legitimacy* (Greene & Gilbert, 2020). For our present purposes it will suffice to note that on the one hand, a positive argument can be made for it, while on the other hand, to whatever extent the negative assessment is considered valid, the development of free and open alternatives would seem to be a necessary strategy of resistance (Swanlund & Schuurman, 2019).

As well as these two debates, there have also been other less contentious developments in the field of geodemographic analysis, particular since the realization that geodemographic products

could assist in efficient public service delivery led to "a renaissance" (Longley, 2005) of research in the area. Cluster analysis is of course used not only in geography and the social sciences, but in fields ranging from biology (Gönen & Margolin, 2014; Bustamam et al., 2017) to psychology (Steinley, 2008) to finance (Rinn et al., 2015) – any subject that involves large quantities of high-dimensional data.

Methods for grouping geoographic areas have therefore progressed significantly since the crude interval partition used by Shevky & Bell (1955). The iterative *k-means algorithm* (Steinley, 2006) seems consistently to be a default choice of algorithmic technique for geodemographic classification, from Jonathan Robbin's first geodemographic products (DeReu & Robbin, 1981) to the Open Area Classification from the 2011 Census (Gale et al., 2016). In fact it has been one of the mostly widely used clustering algorithms over the last fifty years, because of its "ease of implementation, simplicity, efficiency and empirical success" (Jain, 2010, p. 653). The algorithm begins by randomly selecting k cluster centroids, and then proceeds iteratively by allocating each datapoint to whichever cluster centroid is nearest (for some given metric function in attribute space, most usually the Euclidean metric), and then shifting the position of the cluster centroid so that it it located at the mean position of all those datapoints allocated to its cluster, continuing until a stable solution has been reached. In theory, one can define the best k-means partition of a dataset as the one which minimizes the total distance between each point and its cluster centre, but considering every possible partition is unfeasible for large datasets, and so in practice the solution is dependent on the initial random seed – as well as on the choice of how many clusters the algorithm should find, on the choice of what metric should define proximity within the attribute space, and of course on the choice of algorithm itself.

Attempts have been made to address some of the drawbacks of the k-means algorithm. Pelleg & Moore (2000) showed how the optimal number of clusters could be automatically estimated, using a criterion such as the Akaike Information Criterion or the Bayesian Information Criterion. Arthur & Vassilvitskii (2007) suggested an improvement to the initial seeding of cluster centroids, which they called *k-means++*. Bahmani et al. (2012) implemented a parallel version of the k-means algorithm able to take advantage of the efficiencies of distributed computing. Liu et al. (2019) demonstrated that using *principal component analysis* to reduce the variables from the 2011 Census helped them achieve a lower (and better) within-cluster sum-of-squared-distance score compared to the official 2011 Output Area Classification.

Other clustering strategies have also been implemented in geodemographic analysis. Heumann et al. (2020) used *affinity propogation* to cluster the zip codes of the contiguous United States into eleven categories, each characterized by a cluster exemplar; thus achieving a quantitative way of identifying the 'typical American community' or 'Middletown,' an idea first developed by Lynd & Lynd (1929). Spielman & Thill (2008) used the *self-organizing map* introduced by Kohonen (1990) to analyze census tracts in New York City. Cheshire et al. (2011) suggested that *consensus clustering*, using multiple algorithms in combination, offered a way of overcoming the arbitrary nature of cluster results dependent on choice of algorithm. Feng & Flowerdew (1999) first introduced the use of *fuzzy* geodemographic classification, in which an area is not assigned unambiguously to a single cluster, but rather is assigned scores reflecting how well it fits in each cluster. Grekousis & Thomas (2012) compared the more common fuzzy *c-means algorithm* with the *Gustafson-Kessel* algorithm, and concluded that for low values of c, the latter was superior.

While there is nothing specifically geographic about cluster analysis, there have been several attempts to include geographical context in building geodemographic classifications (Alexiou &

Singleton, 2015). Singleton et al. (2012) integrated a geodemographic typology into a spatial interaction framework to model the flow of students into higher education. Grekousis (2021) presented a spatially-aware fuzzy geodemographic algorithm that updated the cluster membership scores with reference only to datapoints within a certain neighbourhood range. Wolf (2021) explored how *spectral clustering* could blend geodemographic classification techniques with regionalization.

As well as being basically aspatial, geodemographic classifications have also been essentially atemporal, merely providing a simplified snapshot of data captured at some static point – in stark contrast to typical *demography*, where "time, in various guises, is integral to the conceptual and methodological apparatus [of the subject]" (Bell, 2015). But in this area too, there have been exceptions. Gale & Longley (2013) used mid-year population estimates and housing stock data to develop three temporal uncertainty indicators for gauging neighbourhood stability. Singleton et al. (2016) explored the stability of geodemographic cluster assignments over an intercensal period by classifying simultaneously data from the 2001 and 2011 censuses.

Singleton & Longley (2009) suggested that the usefulness of general-purpose geodemographic products was approaching an "imminent demise." That has not meant an end of geodemographic analysis however, but rather the opposite, as the 'avalanche' (Miller, 2010) of social data (Arribas-Bel, 2014) has given researchers many more sources of sociospatial quantitative information than were previously available. Whereas previous analyses were largely restricted to the decennial census, recent research has applied geodemographic techniques to a wide variety of data sources ranging from New York taxi GPS data (Liu et al., 2021) to historical individual census data from over a century ago (Lan & Longley, 2021). Although some of this newly available data is open, much of it is proprietary and can not be released because of confidentiality requirements; however Singleton & Longley (2019) demonstrate how it is possible to create the necessary *data infrastructure* to do geodemographic analysis on such data in a secure environment, thus enabling the creation of insightful products that can then be safely shared more widely.

A different approach to creating geodemographic products is that suggested by Adnan et al. (2010), who explored the possibility of an application that might enable the creation of real-time 'on-the-fly' geodemographic classifications. Adnan et al found that "a major problem is the speed at which the clustering algorithm can be used to create robust partitions of datasets into homogeneous groups" (p.285), but in the decade or so since, there have been various advances that might make such a thing easier, particularly in the area of parallel computing (Adnan et al., 2014), with frameworks such as Zaharia et al. (2016).

Having shown that it is not everywhere the case that academics ignore geodemographics, we return to the question of its theoretical rootedness. As an example of unsupervised machine learning, geodemographic classifications are said to be 'theory-free,' in that they are not based on explicitly defined hypotheses that can be verified with clear statistical tests (Singleton & Spielman, 2014). In a certain, limited sense, this is true, but there is certainly much that a consideration of geodemographics can offer social theorists. If geodemographics is seen as a sociological research device, then one might consider 'the spatialization of class' (Parker et al., 2007); if it is seen as a marketing tool, then one might consider how field-capital theory explains consumer segmentation (Tapp & Warren, 2010); if it is seen as a way of understanding residential segmentation, then one would need to consider the social causes and spatial patterns of housing differentiation (Timms, 1971) – why do people choose to move (or not) to certain areas? (Rossi, 1955); and why do different neighbourhoods develop the way they do? (Jacobs, 1961).

But as intimately connected to the subject as they may seem, these are arguably extraneous to

the essential nature of geodemographic classification *per se*. On the other hand, the fundamental ontological question of *what* it is that geodemographic analysis classifies, is unavoidably intrinsic to the subject, and yet remains unresolved.

6. Defining Neighbourhoods: Problems and Possibilities

One possible response to the question would be to suggest that in fact the problem of neighbourhood definition is nothing more than a particular instance of the the more general *Modifiable Areal Unit Problem* (MAUP), described with typical clarity by Openshaw (1983), but in fact identified fifty years previously by Gehlke & Biehl (1934). The problem is a profound one for quantitative analysis involving spatial data, for it observes that the same basic dataset can yield quite different statistical results depending on the specific ways that its data has been aggregated. The effect is found not only when data is gathered at different scales, but even when it is aggregated at the same scale with differing boundaries. A contemporary example is shown in Fig. 8, in which Singleton & Cheshire (2021) demonstrate how, depending on the size of the population of the areal unit used for analysis, COVID-19 rates can appear "as low as 295 per 100,000 people or as high as 736 per 100,000."

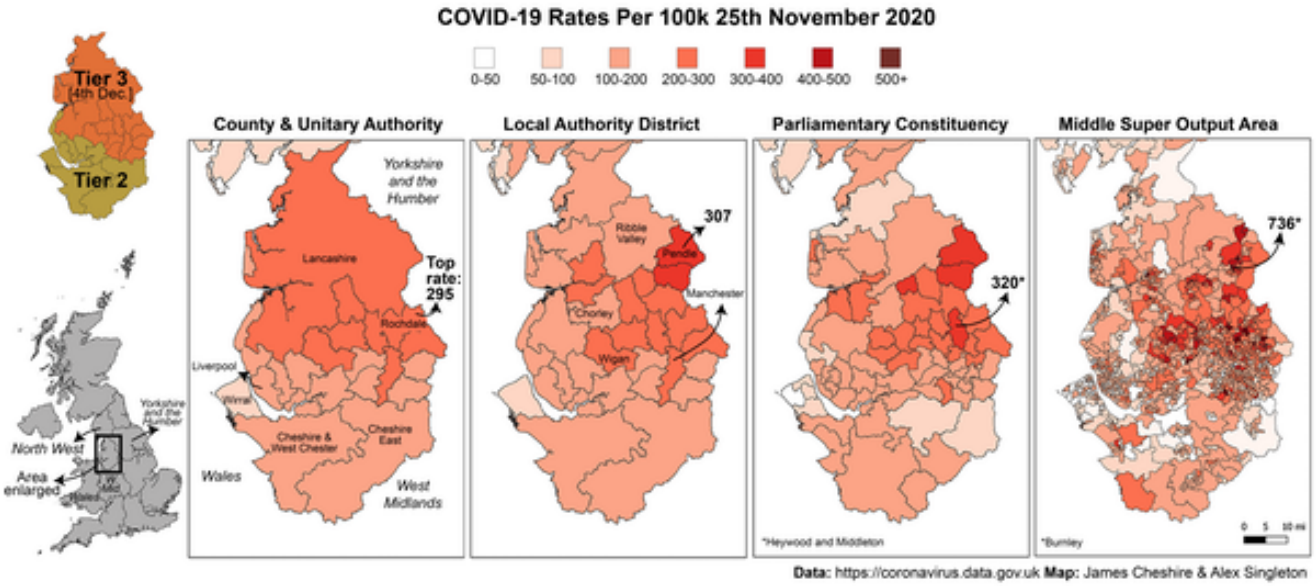

Figure 8: *Demonstrating the Modifiable Areal Unit Problem with COVID-19 Rates, after Singleton & Cheshire (2021)*

A slightly different challenge to quantitative spatial analysis is the *Uncertain Geographic Context Problem* (UGCoP) described by Kwan (2012). The MAUP is a problem of how to aggregate individual datapoints into collective units. It observes that different ways of aggregating spatially located datapoints into broader areal units may lead to different statistical conclusions, and therefore warns against treating any particular areal unit as authoritative. The UGCoP, however, points out that even when only considering a single social datapoint, that point is associated with a human individual who will have experienced exposure to relevant contextual influences in a variety of geographical contexts not limited to the point at which they live, or perhaps were interviewed, the details of which will in general be unknown to the researcher. But

rather than merely confounding the issue further, Kwan frames the problem in such a way as to suggest that there is in fact some "true causally relevant geographic context" (p.959).

For Kwan, this suggests a turn "from location to movement, from place to mobility, and from space to space-time" (p.966), and she suggests "using GPS data to delineate activity spaces" (p.965). This is all very well, but in heeding her suggestion we might perhaps stray too far from our own topic of geodemographic ontology. But nevertheless, the suggestion that for every spatial effect, there must be *some* true causally relevant geographic context, rekindles the hope that even in considering a general typology of residential neighbourhoods, such a context might be found.

What then are the causally relevant contextual elements that make a neighbourhood a neighbourhood? We find ourselves returning to the observation made by Park (1925, p. 1), that "there are forces at work… within the limits of any natural area of human habitation… which tend to bring about an orderly and typical grouping of its population and institutions." Does Park's claim still hold? And if so, with which "forces" should we be primarily concerned? Is it "the economic organization of the city… based on the division of labour" (p.2)? Or rather "racial, cultural and vocational interests" (p.11)? Or "the breaking down of local attachments and the weakening of the restraints and inhibitions of the primary [family] group, under the influence of the urban environment" (p.25)? Or it is the economic expansion of the city, and the accompanying "tendency of each inner zone to extend its area by the invasion of the next outer zone" (Burgess, 1925, p. 51), Fig. 2?

The simplest possibility is that suggested by Morphet (1993), who assumes that if census enumeration districts (EDs) were to constitute a 'natural area,' they must be "homogeneous in their social composition." He investigated whether ED boundaries correspond to significant social boundaries, providing a figure to illustrate the spectrum of possibilities Fig. 9: as well as the obvious cases where the ED boundaries and social geography either differ completely or agree entirely, he also considers the possibility that ED boundaries could coincide with some meaningful social boundaries but that within the ED there might be a dissimilar diversity of social geographic areas. And it seems to him a self-evident truth that "[t]he only significant ED boundaries are those to be found between *different but homogeneous* EDs" (p.1274; emphasis original). Morphet encounters difficulties due to the fact that his data is only available at the ED level, and so it is impossible (without some supplementary source of information) to say whether and where there are boundaries between areas of social homogeneity that do not coincide with ED boundaries. The approach he takes is to consider variables whch plausibly "might exhibit a 'natural' geography of variation at or around the ED scale" (p.1271), but even when he relaxes his definition of homogeneity by defining it negatively rather than positively, or reducing the threshold from 100% to 95%, he finds that "[t]he number of homogeneous EDs is … surprisingly small" (p.1272); "only in terms of tenure type was there a significant degree of homogeneity in the EDs of Newcastle upon Tyne in 1981" (p.1273). It is therefore no subsequent surprise that he finds a negligible proportion ("out of many thousand ED boundaries in the 1981 Census map of Newcastle, only two") to constitute significant social boundaries.

Morphet is certainly right to draw attention to the fact that census enumeration districts do not necessarily constitute socially meaningful geographic units. Indeed, it used to be the case that a key consideration governing the definition of census tract boundaries was that census enumerators should have an equal and manageable workload (Denham, 1980). The lack of correspondence between administrative areal boundaries and socially recognizable neighbourhood units as caused a number of writers to complain hyperbolically of the "tyranny" of "imposed zoning systems"

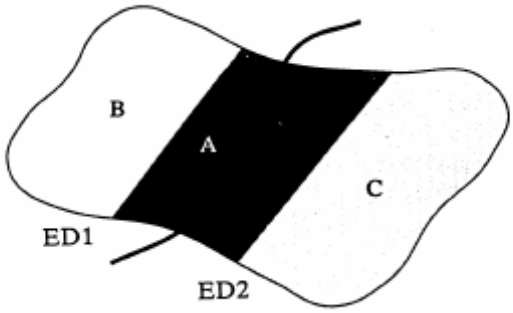

Figure 1. Enumeration districts (EDs) are dissimilar with respect to social geography (A, B, and C), but location of boundary is insignificant.

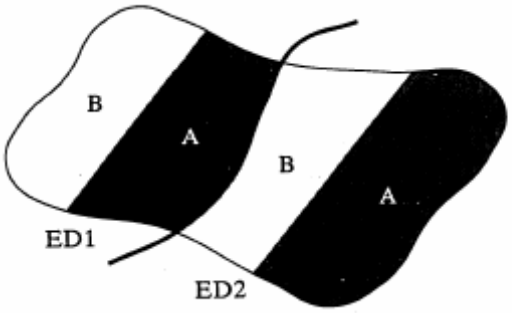

Figure 2. Boundary of enumeration districts (EDs) is significant at sub-ED scale with respect to social geography (A and B); at ED scale the EDs are similar and hence the boundary is not meaningful.

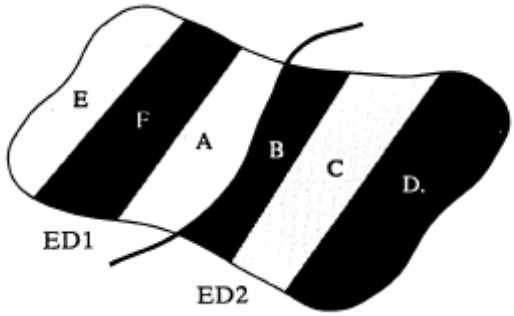

Figure 3. Enumeration districts (EDs) are dissimilar in terms of social geography (A–F), but the pattern of variation is again at sub-ED scale and thus the location of boundary is not significant.

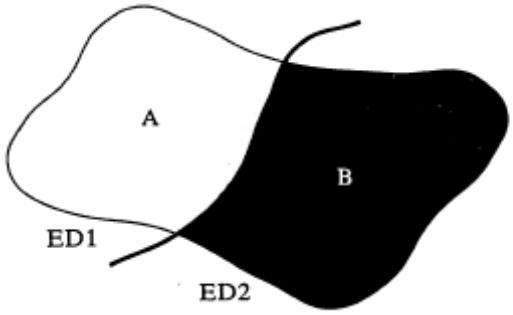

Figure 4. Boundary of enumeration district is significant with respect to social geography (A and B).

Figure 9: *Consideration of how census tract boundaries might relate to social natural areas, after Morphet (1993)*

(Openshaw & Rao, 1995), "census geography" (Sperling, 2012) and even "neighbourhood" itself (Petrović et al., 2020).

We should note though that in Britain at least the geographic boundaries of census output data have since 2001 been decoupled from the enumeration district boundaries of census data collection (Martin, 1998, 2000; Martin et al., 2001; Rees & Martin, 2002), making use of the flexibility afforded by the digitalization of data record-keeping and geographic boundary definition fig. **??**. Official census geography has to navigate the tension of several competing priorities: confidentiality requires that the census output areas cannot be smaller than an appropriate threshold; census output areas are also required by law to nest within statutary boundaries; and those who make use of census data differ as to whether they would prefer areal units that align with well-known postcode geographies, with administrative areas, with a grid, or – for the sake of temporal consistency – with whatever output units were used previously (Duke-Williams & Rees, 1998, p. 580). In Britain, "due to residential clustering at the small area level, [unit] postcodes also tend to exhibit reasonably strong internal homogeneity of the socioeconomic and built environments" (Cockings et al., 2013, p. 1405), and so the unit postcode has been used as the basic spatial unit which is then formed into output area zones by randomly allocating adjacent units into zones and then iteratively optimizing for a selection of target outcomes, using an automated zoning procedure which was first suggested by Openshaw (1977). As well as consistent population size and regular shape, those target outcomes also include social homogeneity (Martin, 2002, p. 45).

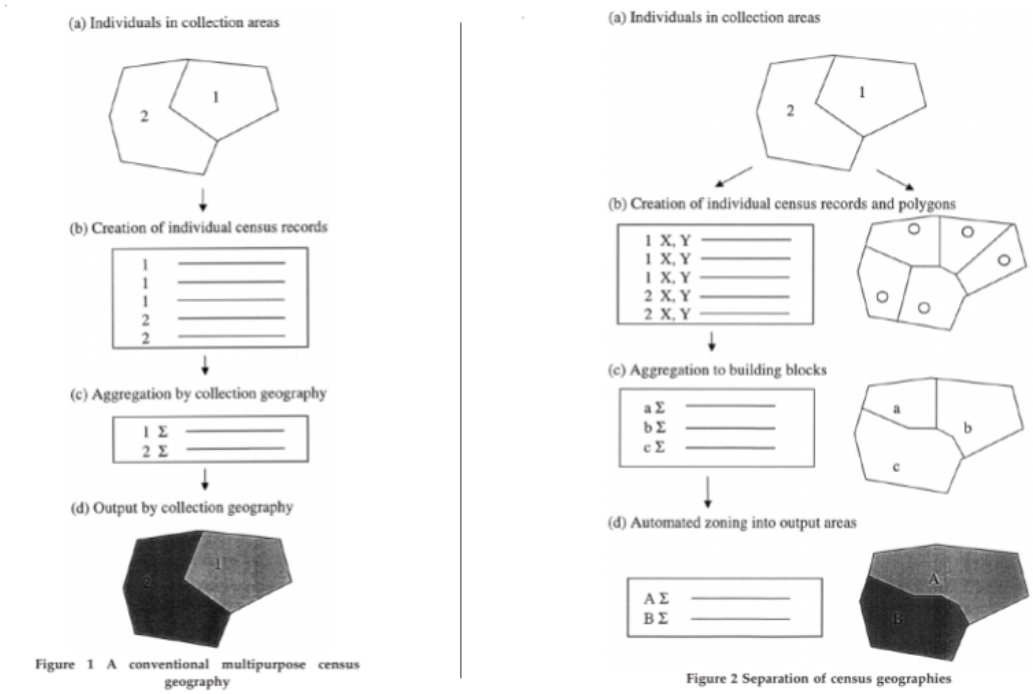

Figure 10: *In Britain since 2001, there has been a separation of the census geographies use for data collection and output (Martin, 2000)*

However, we might question whether a meaningful geographic social 'natural area' must be homogeneous? Statistically, it is obviously true that if data is released in (maximally) homogeneous units, then less information will be lost. But a quantity can very easily be 'maximal' while still being very small. It is also true that the proverbial observation that 'birds of a feather flock together' (McPherson et al., 2001) has often been used to offer an intuitive justification for

why geodemographic classification should be useful (Harris et al., 2001; Longley, 2012; Leventhal, 2016). But the easy rhyming familiarity of the phrase conceals as much as it reveals. If indeed people living in the same place are more likely to be similar, or as Tobler's "first law of geography" (Miller, 2004) might put it, "near things are more related than distant things" (Tobler, 1970), is the operative causal process one of *homophily*, in that those already similar prefer to be nearer to each other, or of *contagion*, in that those already near become more similar (Shalizi & Thomas, 2011)? Timms (1971) also makes a helpful distinction, noting that in his view "the criterion of homogeneity is not that all the people inhabiting a given area should be the same, but that the probability of their being of a particular characteris should be alike in all parts of the area" (p.42).

An impressive case is made by Galster (2019) that "to understand the causes and effects of neighborhoods one must embed them in a framework in which four spatial levels–metropolitan, local jurisdiction, neighbourhood, and individual–are interconnected in mutually causal ways" (Fig. 11). At the individual level, we have the mobility and investment behaviours of individual households, dwelling owners and developers. These both influence and are influenced by the demographic, social, economic and physical characteristics of the household's surrounding neighbourhood. Simultaneously, these neighbourhood characteristics are engaged in circular interaction with public and private service providers operating at a broader geographic scale. And all of this occurs within the context of the regional housing market.

Support for Galster's multilevel suggestion is found in the recent work of Alessandretti et al. (2020) on human mobility. Previous empirical studies of human mobility (Brockmann et al., 2006; Gonzalez et al., 2008; Song et al., 2010) based on massive datasets show that the distribution of travelling distances decays as a power law, which resembles the cumulative distribution of scale-free random walks, in apparent contradiction of our intuitive geographic sense that human behaviour is constrained by the boundaries of different sorts of place at a variety of nested spatial scales: certainly rooms, buildings, cities, and countries, and perhaps too one or more spatial containers corresponding to the somewhat elusive 'neighbourhood.' Alessandretti et al first show that "mixtures of normal (or lognormal) distributions with different variances can generate power laws" (p.402), and then demonstrate that their nested container model "provides a better description of mobility compared with other state-of-the-art models (p.405).

Galster credits his account of a multilevel neighbourhood model to the inspiration of Suttles (1972), whose "groundbreaking observation that people are cognitive of four distinct spatial levels of neighbourhood" (p.39). At Suttles' time of writing, the 'natural area' concept of Park, Burgess, and the interwar Chicago School, had fallen thoroughly out of fashion. Alihan (1938) had concluded that although "the ecological school [was] one of the most definite and influential schools in American sociology" (p.xi), "the concept 'natural area,' so fundamental to human ecology, has not as yet been consistently defined and logically classified… [and] no amount of empirical investigation can rectify the inconsistencies inherent in the theoretical statements pertaining to it" (p.240). Refusing to accept Alihan's damning verdict, Suttles (1972, p. 21) attempted "to resurrect the concept… and show that it may still be usefully applied to urban areas," noting the need to consider both the "physical structure of the city," and "the cognitive map which residents have."

Drawing on the suggestion that there are analogies between human social behaviour and the ideas of animal *territoriality* developed by zoologists such as Lipitz (1969) and Morris (1967), Suttles (1972) suggests that neighbourhood community "is best conceived of as a pyramid of progressively more inclusive groupings" (p.45), and identifies four relevant levels of analysis: the

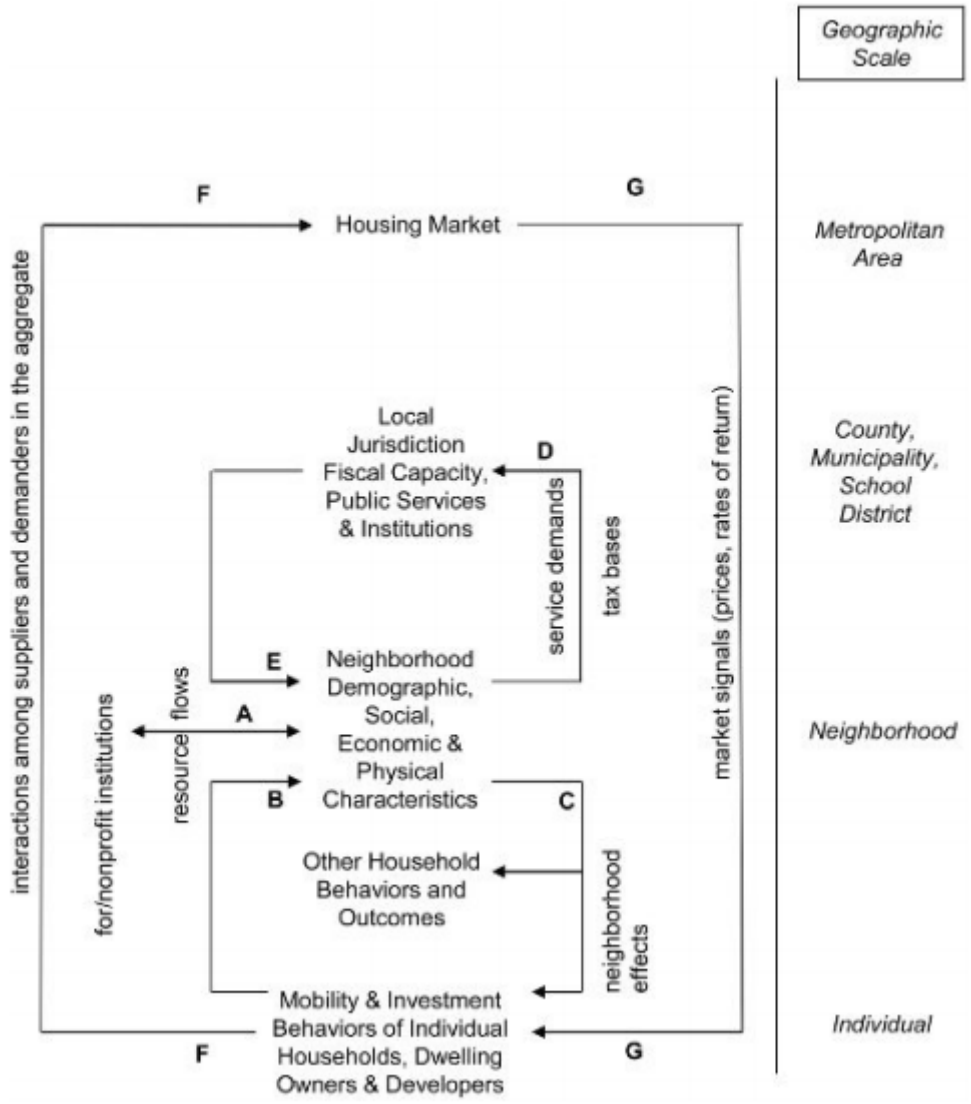

Figure 11: *Holistic, multilevel, circular causation model of neighbourhoods, after Galster (2019)*

'face-block,' the 'defended neighbourhood,' the 'community of limited liability,' and the 'expanded community of limited liability.' At the smallest level, the *face-block* "is the smallest discrete areal unit other than the household which [residents] can point to" (p.56). Suttles takes for granted that his readers will understand what a 'face-block' is, but confuses matters somewhat by introducing the concept together with the loose local network of acquaintances selected "because they are known from shared conditions of residence" (p.55). In an otherwise excellent review of the literature, Chaskin (1997) incorrectly identifies Suttles' definitions of 'local network' and 'face-block,' and suggests that a face-block has no precise residential identification. But in fact Suttles notes that unlike the loose network which is "unlikely to have any sharp boundaries" (p.55), the face-block is notable specifically for having an areal basis so clear that parents are able to use it for instructing their children (p.56). For an explicitly articulated definition, we must turn to Grannis (2009, p. 31), who explains (consistently with Suttles' usage) that "the face block includes all of the dwellings that front on the same street and are situated between the first cross streets, of any type, encountered in both directions away from the respondent's house."

The face-block is of particular interest, because it offers a unit of analysis that is primary from both perspectives necessary to a robust neighbourhood ontology, both those of physical structure and of cognitive mapping. In the last decade the explosion of ubiquitous urban data (Arribas-Bel, 2014) has catalyzed significant advances in the morphological analysis of urban physical structure, with the studies of Barthelemy (2017), Louf & Barthelemy (2014), Schirmer & Axhausen (2016), Boeing (2019) (2020), and Fleischmann et al. (2020) of particular note. But with regard to the latter point of cognitive social maps, although the essential ideas have been well-established since the studies of Gould & White (1974) and Lynch (1960), there remains more work to be done in integrating these concepts into large-scale analyses that take advantage of the detailed data now available. The attempt of Lai et al. (2020) to profile urban places based on geotagged Twitter data for London suggests one possible direction of enquiry. But if we can show more generally that there are strong theoretical reasons for the significance of the face-block, then we can use the analytic tools already developed for morphological analysis and claim them for more social investigation as well.

At a larger level, it is well-established that the structural features of major roads, railway lines and rivers – Jacobs (1961) refers to them as *border vacuums* – are also perceived as social boundaries. Burgess (1925) acknowledges that his simplified theory of urban economic expansion is complicated "by the lake front, the Chicago River, railroad lines, [and] historical factors in the location of industry" (p.52), and these complications are shown on Chart II of his well-known Concentric Zone Diagram (Fig. 2). But Grannis (2009) demonstrates that not only are neighbourhoods defined by the way that urban areas are *divided* by major roads (and railways, rivers, and so on), but that for the households within the same set of boundaries to be accessible to each other, they also need to be *connected* by safe, walkable pedestrian streets – that is, by contiguous residential face-blocks.

Grannis roots his argument in a simple account of how neighbouring relationships necessarily develop along a natural scale of relational availability (Tbl. 2). At the lowest level (0), we have the situation where there is simply no availability at all – and thus there is no neighbourly relationship. The most basic level (1) in actually being neighbours is geographical availability, for "proximity is essential to the very definition of neighbouring" (p.19). The next level (2) is achieved when passive contact takes place, as neighbours "unintentionally encounter each other on a regular basis." The relationship can then develop to involve intentional contact (level 3) and mutual trust

(level 4).

Table 2: *Levels of Relational Availability, (after Grannis, 2009)*

Level	Relational Availability
0	No availability
1	Geographic availability
2	Passive contact
3	Intentional contact
4	Mutual trust

He then suggests that these individual neighbourly relations concatenate to form networks corresponding to the relevant relational stages (pp.37-47). In particular, what becomes apparent is that to "transcend the network of geographic availability … is logically impossible" (p.40). He thus concludes that "the maximal concatenation of contiguous face blocks… represents the maximal consolidation of individual residents' potential contact with each other" (p.42).

We can supplement Grannis' theory of how contiguous walkable face-block networks necessarily bound neighbourhood networks, with some of the insights of Jacobs (1961) about how a city's streets need to serve the vital social purposes of creating a natural place for public contact (pp.72-96), and of providing the "eyes upon the streets" (p.45) necessary to induce the social restraint which makes for public safety (pp.37-71).

7. Conclusion: The Emerging Research Agenda

More could doubtless be said, but our survey of the literature has already revealed enough to establish a firm foundation for fruitful further research. Geodemographic analysis has been proven over the last forty years to be a tool that can effectively be applied to any field of activity interested in engaging real human communities within the local geographical spaces that provide the context for their everyday lives. However, in spite of much scholarly debate over various aspects of geodemographics, the issue of neighbourhood ontology on which the whole practice rests has received surprisingly little focussed attention. In considering it explicitly we have found that rather than being some intractably arcane and abstract issue, there have in fact been a number of writers who have suggested that the most basic spatial unit of geographical neighbourhood can be found in the *face-block* – but this suggestion has not yet been brought into contact with geodemographic analysis. This may be because previously there were not availability the necessary data or computational techniques to operationalize such a concept. But in our contemporary context of abundant data availability and rapidly improving open-source computational tools, this should now be possible.

The first goal emerging from our literature review must therefore be to operationalize the neighbourhood face-block unit as a basic geographic unit for the whole of Britain, developing whatever methods might be necessary to obtain such a unit from open data. In doing this, it will be worth exploring also whether it is possible to unambiguously operationalize the concept of neighbourhood networks of connected and contiguous residential face-blocks.

If these face-blocks, and perhaps also their connected networks, are to not merely be basic geographic units but actual *geodemographic* units, then they must be associated with demographic

data relating to their resident populations. The second goal of our research must therefore be to develop a method of interpolating data from the output areas for which it is made available to the face-block units in which we are interested. In order to establish the success of our methods, we must also find some way of testing the extent to which such a transformation is accurate or not.

If these first two research goals are achieved, then the natural third goal would be at least to produce an open geodemographic classification using the theoretically-grounded neighbourhood face-block as its fundamental unit. Rather than simply producing a static product, it might be more fruitful to develop a tool (whether a web application or a software package) to enable others to perform geodemographic analysis using the methods of neighbourhood definition and data interpolation that have been developed.

So, having reviewed the literature we find ourselves with three clear research goals. If these three aims can be fulfilled, then I believe I will have successfully contributed something to the field of geodemographic analysis and to the understanding of neighbourhood ecology.

Bibliography

Abbott, A. (2017a). *Department and Discipline: Chicago Sociology at One Hundred*. University of Chicago Press.

Abbott, A. (2017b). The Historiography of the Chicago School. In *Department and Discipline: Chicago Sociology at One Hundred* (pp. 4–33). University of Chicago Press.

Abu-Lughod, J. L. (1969). Testing the theory of social area analysis: The ecology of Cairo, Egypt. *American Sociological Review*, 198–212.

Addams, J. (Ed.). (1895). *Hull-House Maps and Papers: A Presentation of Nationalities and Wages in a Congested District of Chicago, Together with Comments and Essays on Problems Growing Out of the Social Conditions*. T.Y. Crowell.

Adnan, M., Longley, P. A., Singleton, A. D., & Brunsdon, C. (2010). Towards real-time geodemographics: Clustering algorithm performance for large multidimensional spatial databases. *Transactions in GIS*, *14*(3), 283–297.

Adnan, M., Longley, P. A., Singleton, A. D., & Turton, I. (2014). Parallel computing in geography. In *GeoComputation* (Vol. 2, pp. 49–67). CRC Press Boca Raton, Florida.

Alessandretti, L., Aslak, U., & Lehmann, S. (2020). The scales of human mobility. *Nature*, *587*(7834), 402–407.

Alexiou, A., & Singleton, A. (2015). The Role of Geographical Context in Building Geodemographic Classifications. In *http://dx.doi.org/10.6084/m9.figshare.1491375* (pp. 40–45). Leeds.

Alihan, M. A. (1938). *Social Ecology*. Columbia University Press.

Anderson, M. J. (2015). *The American Census: A Social History, Second Edition*. Yale University Press.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, *49*, 45–53. https://doi.org/10.1016/j.apgeog.2013.09.012

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The Advantages of Careful Seeding, 11.

Ashby, D. I., & Longley, P. A. (2005). Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS*, *9*(1), 53–72.

Bahmani, B., Moseley, B., Vattani, A., Kumar, R., & Vassilvitskii, S. (2012). Scalable k-means++. *arXiv Preprint arXiv:1203.6402*.

Bales, K. (1999). Popular Reactions to Sociological Research: The Case of Charles Booth. *Sociology-the Journal of The British Sociological Association*, *33*(1), 153–168. https://doi.org/10.1177/S0038038599000085

Bamshad, M. J., & Olson, S. E. (2003). Does 'Race' Exist? *Scientific American*, *289*(6), 78–85.

Barthelemy, M. (2017). From paths to blocks: New measures for street patterns. *Environment and Planning B: Urban Analytics and City Science*, *44*(2), 256–271. https://doi.org/10.1177/0265813515599982

Bayatrizi, Z. (2008). From Fate to Risk: The Quantification of Mortality in Early Modern Statistics. *Theory, Culture & Society*, *25*(1), 121–143. https://doi.org/10.1177/0263276407085160

Bell, M. (2015). Demography, time and space. *Journal of Population Research*, *32*(3/4), 173–186.

Bell, W. (1953). The Social Areas of the San Francisco Bay Region. *American Sociological Review*, *18*(1), 39–47. https://doi.org/10.2307/2087847

Bernard, L. L. (1930). Schools of sociology. *The Southwestern Political and Social Science Quarterly*, *11*(2), 117–134.

Berry, B. J. L., & Rees, P. H. (1969). The Factorial Ecology of Calcutta. *American Journal of Sociology*, *74*(5), 445–491. https://doi.org/10.1086/224681

Blake, M., & Openshaw, S. (1994). GB Profiles: A User Guide.

Boeing, G. (2019). Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science*, *4*(1), 67. https://doi.org/10.1007/s41109-019-0189-1

Boeing, G. (2020). The right tools for the job: The case for spatial science tool-building. *Transactions in GIS*, *n/a*(n/a). https://doi.org/10.1111/tgis.12678

Boland, P. (2010). Sonic Geography, Place and Race in the Formation of Local Identity: Liverpool and Scousers. *Geografiska Annaler: Series B, Human Geography*, *92*(1), 1–22. https://doi.org/10.1111/j.1468-0467.2010.00330.x

Booth, C. (1887). The Inhabitants of Tower Hamlets (School Board Division), Their Condition and Occupations. *Journal of the Royal Statistical Society*, *50*(2), 326–401. https://doi.org/10.2307/2340191

Booth, C. (1889). *Labour and Life of the People ...* Williams and Norgate.

Booth, C. (1893). Life and Labour of the People in London: First Results of An Inquiry Based on the 1891 Census. Opening Address of Charles Booth, Esq., President of the Royal Statistical Society. Session 1893-94. *Journal of the Royal Statistical Society*, *56*(4), 557–593. https://doi.org/10.2307/2979431

Booth, C. (1904). *Life and Labour of the People in London*. London ; New York: Macmillan.

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, *439*(7075), 462–465. https://doi.org/10.1038/nature04292

Burgess, E. W. (1925). The growth of the city. In *The City* (pp. 37–44). Chicago: University of Chicago Press.

Burgess, E. W., & Newcomb, C. S. (1931). *Census data of the city of Chicago, 1920 EW Burgess, CS Newcomb - 1931 - University of Chicago Press*. Chicago: University of Chicago Press.

Burrows, R., & Gane, N. (2006). Geodemographics, Software and Class. *Sociology-the Journal of The British Sociological Association*, *40*(5), 793–812. https://doi.org/10.1177/0038038506067507

Bustamam, A., Tasman, H., Yuniarti, N., Frisca, & Mursidah, I. (2017). Application of k-means clustering algorithm in grouping the DNA sequences of hepatitis B virus (HBV). *AIP Conference Proceedings*, *1862*(1), 030134. https://doi.org/10.1063/1.4991238

CDRC. (2021). CDRC Mapmaker: Output Area Classification. https://mapmaker.cdrc.ac.uk/#/output-area-classification?h=2&lon=-0.1699&lat=52.6515&zoom=6.31.

Charlton, M., Openshaw, S., & Wymer, C. (1985). Some new classifications of census enumeration districts in Britain: A poor mans ACORN. *Journal of Economic and Social Measurement*, *13*(1), 69–96.

Chaskin, R. J. (1997). Perspectives on neighborhood and community: A review of the literature. *Social Service Review*, *71*(4), 521–547.

Cheshire, J., Adnan, M., & Gale, C. (2011). The use of consensus clustering in geodemographics. *Proceedings of GIS Research UK, Portsmouth*.

Cockings, S., Harfoot, A., Martin, D., & Hornby, D. (2013). Getting the Foundations Right: Spatial Building Blocks for Official Population Statistics. *Environment and Planning A: Economy and Space*, *45*(6), 1403–1420. https://doi.org/10.1068/a45276

Cole, J. (2000). *The Power of Large Numbers: Population, Politics, and Gender in Nineteenth-Century France*. Cornell University Press.

Collet, C. E. (1945). Charles Booth, The Dension Club and H. Llewellyn Smith. *Journal of the Royal Statistical Society*, *108*(3/4), 482–485.

Cortese, A. J. (1995). The rise, hegemony, and decline of the Chicago School of Sociology, 1892. *The Social Science Journal*, *32*(3), 235–254.

Crewe, L. (2017). *The Geographies of Fashion: Consumption, Space, and Value*. Bloomsbury Publishing.

Cypress, B. S. (2019). Qualitative research: Challenges and dilemmas. *Dimensions of Critical Care Nursing*, *38*(5), 264–270.

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

Davies, W. K. D. (1978). Charles Booth and the Measurement of Urban Social Character. *Area*, *10*(4), 290–296.

Deas, I., Robson, B., Wong, C., & Bradford, M. (2003). Measuring Neighbourhood Deprivation: A Critique of the Index of Multiple Deprivation. *Environment and Planning C: Government and Policy*, *21*(6), 883–903. https://doi.org/10.1068/c0240

Denham, C. (1980). The geography of the Census 1971 and 1981. *Population Trends*, *19*, 6–12.

DeReu, J. A., & Robbin, J. E. (1981). *Application of geodemographics to the Army recruiting problem*. CLARITAS CORP ARLINGTON VA.

Donnelly, S. (2002). Mapping rich and poor: Putting Charles Booth's Enquiry into London life and labour online. *The Cartographic Journal*, *39*(1), 81–85.

Du Bois, W. E. B. (1899). *The Philadelphia Negro*. Philadelphia: University of Pennsylvania Press.

Duke-Williams, O., & Rees, P. (1998). Can Census Offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. *International Journal of Geographical Information Science*, *12*(6), 579–605. https://doi.org/10.1080/136588198241680

Englander, D. (1989). Booth's Jews: The Presentation of Jews and Judaism in "Life and Labour of the People in London". *Victorian Studies*, *32*(4), 551–571.

Evans, M. (1998). From 1086 and 1984: Direct Marketing into the Millennium. *Marketing Intelligence & Planning*, *16*(1), 56–67.

Farr, M., Wardlaw, J., & Jones, C. (2008). Tackling health inequalities using geodemographics: A social marketing approach. *International Journal of Market Research*, *50*(4), 449–467.

Feng, Z., & Flowerdew, R. (1999). The use of fuzzy classification to improve geodemographic targeting. In *Innovations in GIS 6* (pp. 149–161). CRC Press.

Fenton, S. (1981). Robert Park: His life and his sociological imagination (a review article). *Journal of Ethnic and Migration Studies*, *9*(2), 294–301. https://doi.org/10.1080/1369183X.1981.9975698

Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2020). Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Computers, Environment and Urban Systems*, *80*, 101441.

Fletcher, J. (1847). Moral and Educational Statistics of England and Wales. *Journal of the Statistical Society of London*, *10*(3), 193–233. https://doi.org/10.2307/2337698

Fletcher, R., Hornback, H., & Queen, S. (1935). *Social Statistics of St. Louis by Census Tracts*. St Louis: Washington University Press.

Foley, D. L. (1953). Census Tracts and Urban Research. *Journal of the American Statistical Association*, *48*(264), 733–742. https://doi.org/10.2307/2281068

Funkhouser, H. G. (1937). Historical Development of the Graphical Representation of Statistical Data. *Osiris*, *3*, 269–404.

Gale, C. G. (2014). *Creating an open geodemographic classification using the UK Census of the Population* (Doctoral). UCL (University College London).

Gale, C. G., & Longley, P. A. (2013). Temporal uncertainty in a small area open geodemographic classification. *Transactions in GIS*, *17*(4), 563–588.

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

Galster, G. C. (2019). *Making Our Neighborhoods, Making Our Selves*. University of Chicago Press.

GDPR. (2016). General Data Protection Regulation (GDPR) Compliance Guidelines.

Gehlke, C. E., & Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, *29*(185A), 169–170.

Gilbert, S. (2021). *Good Data: An Optimist's Guide to Our Digital Future*. Welbeck Publishing Group Limited.

Gillie, A. (1996). The Origin of the Poverty Line. *The Economic History Review*, *49*(4), 715–730. https://doi.org/10.2307/2597970

Glass, D. (1963). John Graunt and his Natural and political observations.

Glass, D. V. (1978). *Numbering the People: The Eighteenth-Century Population Controversy and the Development of Census and Vital Statistics in Britain*. Gordon & Cremonesi.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

González-Benito, Ó., & González-Benito, J. (2005). The role of geodemographic segmentation in retail location strategy. *International Journal of Market Research*, *47*(3).

Gould, P., & White, R. (1974). *Mental Maps*. Penguin.

Gönen, M., & Margolin, A. A. (2014). Localized data fusion for kernel k-means clustering with application to cancer biology. *Advances in Neural Information Processing Systems*, *27*, 1305–1313.

Grannis, R. (2009). *From the Ground Up: Translating Geography into Community through Neighbor Networks*. Princeton University Press.

Graunt, J. (1662). *Natural and political observations mentioned in a following index, and made upon the bills of mortality*. London.

Greene, A. R., & Gilbert, S. J. (2020). More Data, More Power? Towards a Theory of Digital Legitimacy. *Towards a Theory of Digital Legitimacy (October 9, 2020)*.

Grekousis, G. (2021). Local fuzzy geographically weighted clustering: A new method for geodemographic segmentation. *International Journal of Geographical Information Science*, *35*(1), 152–174.

Grekousis, G., & Thomas, H. (2012). Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and GustafsonKessel methods. *Applied Geography*, *34*, 125–136. https://doi.org/10.1016/j.apgeog.2011.11.004

Hadden, J. K., & Borgatta, E. F. (1965). *American Cities: Their Social Characteristics*. Rand McNally.

Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting*. West Sussex, England ; Hoboken, N.J: Wiley.

Harris, R., Voas, D., & Williamson, P. (2001). The Diversity of Diversity: Is There Still a Place for Small Area Classifications?[with Response]. *Area*, *33*(3), 329–336.

Harvey, L. (1987). The nature of 'Schools' in the sociology of knowledge: The case of the 'Chicago School.' *The Sociological Review*, *35*(2), 245–278.

Havlíček, J., & Roberts, S. C. (2013). The Perfume-Body Odour Complex: An Insightful Model for CultureGene Coevolution? In M. L. East & M. Dehnhard (Eds.), *Chemical Signals in Vertebrates 12* (pp. 185–195). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-5927-9_14

Heumann, B. W., Liesch, M. E., Bogen, N. R., Meier, R. A., & Graziano, M. (2020). The contiguous United States in eleven zip codes: Identifying and mapping socio-economic census data clusters and exemplars using affinity propagation. *Journal of Maps*, *16*(1), 57–67. https://doi.org/10.1080/17445647.2020.1736193

Hume, A. (1850). *Missions at Home; or a Clergyman's account of a portion of the Town of Liverpool*. London: J. & F. Rivington.

Hume, A. (1858a). *Condition of Liverpool: Religious and Social; Including Notices of the State of Education, Morals, Pauperism, and Crime*. T. Brakell.

Hume, A. (1858b). Liverpool, ecclesiastical and social.

Hyndman, H. M. (1911). *The record of an adventurous life*. Macmillan.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. Vintage Books.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Janson, C.-G. (1974). The spatial structure of Newark, New Jersey. *Environment and Planning A*, *6*(3), 273–290.

Jones, F. L. (1965). A social profile of Canberra, 1961. *The Australian and New Zealand Journal of Sociology*, *1*(2), 107–120.

Kalu, M. E. (2019). How does "subjective I" influence a qualitative research question, theoretical approach and methodologies? *Global Journal of Pure and Applied Sciences*, *25*(1), 97–101.

Karvonen, A. (2020). Book review: The city by robert e. Park and ernest w. Burgess. *LSE Review of Books*.

Kimball, M. A. (2006). London through rose-colored graphics: Visual rhetoric and information graphic design in Charles Booth's maps of London poverty. *Journal of Technical Writing and Communication*, *36*(4), 353–381.

Knaap, E., Wolf, L. J., Rey, S. J., Kang, W., & Han, S. (2019). *The Dynamics of Urban Neighborhoods: A Survey of Approaches for Modeling Socio-Spatial Structure* (Preprint). SocArXiv. https://doi.org/10.31235/osf.io/3frcz

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

Kreager, P. (1993). Histories of Demography: A Review Article. *Population Studies*, *47*(3), 519–539. https://doi.org/10.1080/0032472031000147286

Kwan, M.-P. (2012). How GIS can help address the uncertain geographic context problem in social science research. *Annals of GIS*, *18*(4), 245–255. https://doi.org/10.1080/19475683.2012.727867

Lai, J., Lansley, G., Haworth, J., & Cheng, T. (2020). A name-led approach to profile urban places based on geotagged Twitter data. *Transactions in GIS*, *24*(4), 858–879.

Lan, T., & Longley, P. A. (2021). Urban Morphology and Residential Differentiation across Great Britain, 1881. *Annals of the American Association of Geographers*, *111*(6), 1796–1815. https://doi.org/10.1080/24694452.2020.1859982

Leventhal, B. (2016). Birds of a feather still flock together: The continuing relevance of geodemographics. *Applied Marketing Analytics*, *2*(1), 52–56.

Levitan, K. (2011). *A Cultural History of the British Census: Envisioning the Multitude in the Nineteenth Century*. Springer.

Lindner, R. (1996). *The Reportage of Urban Culture: Robert Park and the Chicago School*. Cambridge University Press.

Lipitz, M. L. (1969). The territorial imperative. *Public Administration Review*, *29*(4), 384–386.

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

Liu, Y., Singleton, A., Arribas-bel, D., & Chen, M. (2021). Identifying and understanding road-constrained areas of interest (AOIs) through spatiotemporal taxi GPS data: A case study in New York City. *Computers, Environment and Urban Systems*, *86*, 101592. https://doi.org/10.1016/j.compenvurbsys.2020.101592

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

Longley, P. A. (2012). Geodemographics and the practices of geographic information science. *International Journal of Geographical Information Science*, *26*(12), 2227–2237.

Louf, R., & Barthelemy, M. (2014). A typology of street patterns. *Journal of The Royal Society Interface*, *11*(101), 20140924.

Lynch, K. (1960). *The Image of the City*. MIT Press.

Lynd, R. S., & Lynd, H. M. (1929). Middletown; a study in contemporary American culture.

Mallapragada, M. (2016). Curry as code: Food, race, and technology. *Global Asian American Popular Cultures*, 263.

Marr, T. R. (1904). *Housing Conditions in Manchester & Salford: A Report Prepared for the Citizens' Association for the Improvement of the Unwholesome Dwellings and Surroundings of the People, with the Aid of the Executive Committee*. Sherratt and Hughes, University Press.

Martin, D. (1998). Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographical Information Science*, *12*(7), 673–685. https://doi.org/10.1080/136588198241590

Martin, D. (2000). Towards the geographies of the 2001 UK Census of Population. *Transactions of the Institute of British Geographers*, *25*(3), 321–332.

Martin, D. (2002). Output areas for 2001. In P. Rees, D. J. Martin, & P. Williamson (Eds.), *The Census Data System* (pp. 27–36). Wiley.

Martin, D., Nolan, A., & Tranmer, M. (2001). The Application of Zone-Design Methodology in the 2001 UK Census. *Environment and Planning A: Economy and Space*, *33*(11), 1949–1962. https://doi.org/10.1068/a3497

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*, 415–444.

Miller, H. J. (2004). Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*, *94*(2), 284–289. https://doi.org/10.1111/j.1467-8306.2004.09402005.x

Miller, H. J. (2010). The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, *50*(1), 181–201.

Morphet, C. S. (1993). The Mapping of Small-Area Census Data: A Consideration of the Role of Enumeration District Boundaries. *Environment and Planning A: Economy and Space*, *25*(9), 1267–1277. https://doi.org/10.1068/a251267

Morris, D. (1967). *The naked ape: A study of the human animal*. New York, Delta Books.

Myers, G. (2006). 'Where are you from?': Identifying place. *Journal of Sociolinguistics*, *10*(3), 320–343.

O'Day, R., & Englander, D. (1992). *Mr. Charles Booth's Inquiry: Life and Labour of the People in London Reconsidered*. A&C Black.

Openshaw, S. (1977). A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling. *Transactions of the Institute of British Geographers*, *2*(4), 459–472. https://doi.org/10.2307/622300

Openshaw, S. (1983). *The Modifiable Areal Unit Problem*. Geo Books.

Openshaw, S. (1985). Rural Area Classification Using Census Data. *Geographia Polonica*, (51), 285–299.

Openshaw, S. (1997). The truth about Ground Truth. *Transactions in GIS*, *2*(1), 7–24. https://doi.org/10.1111/j.1467-9671.1997.tb00002.x

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

Openshaw, S., & Gillard, A. A. (1978). On the stability of a spatial classification of census enumeration district. In P. W. J. Batey (Ed.), *Theory and Methods in Urban and Regional Analysis* (pp. 101–119). London: Pion.

Openshaw, S., & Rao, L. (1995). Algorithms for Reengineering 1991 Census Geography. *Environment and Planning A: Economy and Space*, *27*(3), 425–446. https://doi.org/10.1068/a270425

Owens, B. R. (2012). Mapping the City: Innovation and Continuity in the Chicago School of Sociology, 1920. *The American Sociologist*, *43*(3), 264–293. https://doi.org/10.1007/s12108-012-9160-7

Park, R. (1925). The City: Suggestions for the Investigation of Human Behaviour in the Urban Environment. In R. Park & E. W. Burgess (Eds.), *The City*. Chicago, IL, USA: University of Chicago Press.

Park, R. E. (1915). The City: Suggestions for the Investigation of Human Behavior in the City Environment. *American Journal of Sociology*, *20*(5), 577–612. https://doi.org/10.1086/212433

Park, R. E., Burgess, E. W., & McKenzie, R. D. (1925). *The City*. 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.

Pelleg, D., & Moore, A. W. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *Icml* (Vol. 1, pp. 727–734).

Petrović, A., Manley, D., & van Ham, M. (2020). Freedom from the tyranny of neighbourhood: Rethinking sociospatial context effects. *Progress in Human Geography*, *44*(6), 1103–1123. https://doi.org/10.1177/0309132519868767

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

Pickering, W. S. F. (1972). Abraham Hume (1814-1884). A Forgotten Pioneer in Religious Sociology. *Archives de Sociologie Des Religions*, *17*(33), 33–48.

Pickles, J. (Ed.). (1995). *Ground Truth: The Social Implications of Geographic Information Systems* (1 edition). New York: Guilford Press.

Rees, P. H. (1971). Factorial Ecology: An Extended Definition, Survey, and Critique of the Field. *Economic Geography*, *47*(sup1), 220–233. https://doi.org/10.2307/143205

Rees, P., & Martin, D. (2002). The debate about census geography. In P. Rees, D. J. Martin, & P. Williamson (Eds.), *The Census Data System* (pp. 27–36). Wiley.

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

Ricercar. (2021). Jonathan Robbin: Curriculum Vitae.

Rinn, P., Stepanov, Y., Peinke, J., Guhr, T., & Schäfer, R. (2015). Dynamics of quasi-stationary systems: Finance as an example. *EPL (Europhysics Letters)*, *110*(6), 68003. https://doi.org/10.1209/0295-5075/110/68003

Robbin, J. E. (1980). Geodemographics: The New Magic. *Campaigns and Elections*, *1*(Spring), 25–46.

Robbins, H. (2010). "We Are Seven" and the First British Census. *English Language Notes*, *48*(2), 201–213. https://doi.org/10.1215/00138282-48.2.201

Robinson, A. H. (1982). *Early thematic mapping in the history of cartography*. Chicago : University of Chicago Press.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–357. https://doi.org/10.2307/2087176

Robson, B. T., & Robson, P. (1969). *Urban analysis: A study of city structure with special reference to Sunderland*. CUP Archive.

Rossi, P. H. (1955). *Why families move: A study in the social psychology of urban residential mobility*. Free Press.

Rowntree, B. S. (1902). *Poverty: A Study of Town Life*. Macmillan & Company.

Schelling, T. C. (1969). Models of Segregation. *The American Economic Review*, *59*(2), 488–493.

Schirmer, P. M., & Axhausen, K. W. (2016). A multiscale classification of urban morphology. *Journal of Transport and Land Use*, *9*(1), 101–130.

Schmid, C. F. (1960). Urban crime areas: Part I. *American Sociological Review*, 527–542.

Selvin, H. C. (1976). Durkheim, Booth and Yule : The non-diffusion of an intellectual innovation. *European Journal of Sociology / Archives Européennes de Sociologie / Europäisches Archiv für Soziologie*, *17*(1), 39–51.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, *40*(2), 211–239. https://doi.org/10.1177/0049124111404820

Shevky, E., & Bell, W. (1955). *Social Area Analysis: Theory, illustrative application, and computational procedures*.

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

Simey, T. S., & Simey, M. B. (1960). *Charles Booth, social scientist*. London, Oxford U. P.

Simmel, G. (1908). Das Geheimnis und die geheime Gesellschaft. *Soziologie. Untersuchungen über Die Formen Der Vergesellschaftung*, 256–304.

Singleton, A., & Cheshire, J. (2021). How England's complicated political geography is confusing coronavirus rules. *The Conversation*. http://theconversation.com/how-englands-complicated-political-geography-is-confusing-coronavirus-rules-152036.

Singleton, A. D., & Longley, P. A. (2009). Creating open source geodemographics: Refining a national classification of census output areas for applications in higher education. *Papers in Regional Science*, *88*(3), 643–666.

Singleton, A. D., & Longley, P. A. (2019). Data infrastructure requirements for new geodemographic classifications: The example of London's workplace zones. *Applied Geography*, *109*, 102038.

Singleton, A. D., Spielman, S., & Brunsdon, C. (2016). Establishing a framework for Open Geographic Information science. *International Journal of Geographical Information Science*, *30*(8), 1507–1521. https://doi.org/10.1080/13658816.2015.1137579

Singleton, A. D., & 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.

Singleton, A. D., 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

Sleight, P. (1997). *Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business* (2nd Revised edition edition). Oxfordshire: NTC Publications.

Small, A. W. (1895). The era of sociology. *American Journal of Sociology*, *1*(1), 1–15.

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, *21*(1), 3–8.

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, *6*(10), 818–823.

Sperling, J. (2012). The tyranny of census geography: Small-area data and neighborhood statistics. *Cityscape*, 219–223.

Spielman, S. E., & Thill, J.-C. (2008). Social area analysis, data mining, and GIS. *Computers, Environment and Urban Systems*, *32*(2), 110–122.

Spijker, A. van 't. (2014). *The New Oil: Using Innovative Business Models to turn Data Into Profit*. Technics Publications.

Steinley, D. (2008). Stability analysis in K-means clustering. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 255–273.

Steinley, Douglas. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1–34. https://doi.org/10.1348/000711005X48266

Sullivan, T. A. (2020). Coming To Our Census: How Social Statistics Underpin Our Democracy (And Republic). *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.c871f9e0

Suttles, G. D. (1972). *The social construction of communities* (Vol. 728). Chicago: University of Chicago Press.

Swanlund, D., & Schuurman, N. (2019). Resisting geosurveillance: A survey of tactics and strategies for spatial privacy. *Progress in Human Geography*, *43*(4), 596–610. https://doi.org/10.1177/0309132518772661

Sweetser, F. L. (1965). *Ecological Factors in Helsinki and Boston*. University of Helsinki, Institute of sociology.

Taeuber, C. (1941). Rural-Urban Migration. *Agricultural History*, *15*(3), 151–160.

Tapp, A., & Warren, S. (2010). Field-capital theory and its implications for marketing. *European Journal of Marketing*.

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, *34*(6), 990–1006.

Timms, D. (1971). *The Urban Mosaic: Towards a Theory of Residential Differentiation*. CUP Archive.

Tobias, J. J. (1974). A Statistical Study of a Ninteenth-Century Criminal Area, *14*, 221.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(sup1), 234–240.

Topalov, C. (1993). The city as terra incognita: Charles Booth's poverty survey and the people of London, 1886. *Planning Perspectives*, *8*(4), 395–425. https://doi.org/10.1080/02665439308725782

Tryon, R. C. (1939). *Cluster Analysis*. Ann Arbor: Edwards Brothers.

Tryon, R. C. (1955). *Identification of social areas by cluster analysis: A general method with an application to the San Francisco bay area* (Vol. 8). University of California Press.

Tryon, R. C. (1968). Comparative Cluster Analysis Of Social Areas. *Multivariate Behavioral Research*, *3*(2), 213–232. https://doi.org/10.1207/s15327906mbr0302_6

Tryon, R. C., & Bailey, D. E. (1966). The BC Try Computer System of Cluster And Factor Analysis. *Multivariate Behavioral Research*, *1*(1), 95–111. https://doi.org/10.1207/s15327906mbr0101_6

Turner, J. H. (1988). The Mixed Legacy of the Chicago School of Sociology. *Sociological Perspectives*, *31*(3), 325–338. https://doi.org/10.2307/1389202

Vaughan, L. (2018). *Mapping Society : The Spatial Dimensions of Social Cartography*. UCL Press. https://doi.org/10.14324/111.9781787353053

Vaughan, L., & Geddes, I. (2009). Urban form and deprivation: A contemporary proxy for Charles Booth's analysis of poverty. *Radical Statistics*, (Autumn).

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), 379–403. https://doi.org/10.1111/j.1467-985X.2007.00466.x

Von Thunen, J. (1826). *Der Isolierte Staat*. Hamburg: Friedrich Perthes.

Wainer, H. (2012). Visual Revelations: Moral Statistics and the Thematic Maps of Joseph Fletcher. *CHANCE*, *25*(1), 43–46. https://doi.org/10.1080/09332480.2012.668467

Weatherhead, D., White, K. S., & Friedman, O. (2016). Where are you from? Preschoolers infer background from accent. *Journal of Experimental Child Psychology*, *143*, 171–178.

Webber, R. (1977). *An Introduction to the National Classification of Wards and Parishes*. London: Planning Research Applications Group.

Webber, R. (2006). How parties used segmentation in the 2005 British general election campaign. *Journal of Direct, Data and Digital Marketing Practice*, *7*(3), 239–252. https://doi.org/10.1057/palgrave.dddmp.4340529

Webber, R., & Burrows, R. (2018). *The Predictive Postcode: The geodemographic classification of British society*. Los Angeles: SAGE.

Webber, R. J. (1975). Liverpool Social Area Study 1971 Data: Final Report. Planning Research Applications Group.

Webber, R. J. (1978). Making the Most of the Census for Strategic Analysis. *The Town Planning Review*, *49*(3), 274–284.

Webber, R. J. (1980). A Response to the Critique of the OPCS/PRAG National Classifications. *The Town Planning Review*, *51*(4), 440–450.

Wells, W. D. (1975). Psychographics: A Critical Review. *Journal of Marketing Research*, *12*(2), 196–213. https://doi.org/10.2307/3150443

Westergaard, J. H. (1969). Review of Charles Booth on the City: Physical Pattern and Social Structure. *Social Research*, *36*(2), 324–327.

Wilcox, A. (2014). *The Church and the Slums: The Victorian Anglican Church and its Mission to Liverpool's Poor*. Cambridge Scholars Publishing.

Wirth, L. (1938). Urbanism as a Way of Life. *American Journal of Sociology*, *44*(1), 1–24.

Wolf, L. J. (2021). Spatiallyencouraged spectral clustering: A technique for blending map typologies and regionalization. *International Journal of Geographical Information Science*, *35*(11), 2356–2373.

Yule, G. U. (1895). On the Correlation of total Pauperism with Proportion of Out-Relief. *The Economic Journal*, *5*(20), 603–611. https://doi.org/10.2307/2956650

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., … Franklin, M. J. (2016). Apache spark: A unified engine for big data processing. *Communications of the ACM*, *59*(11), 56–65.