

The internal structure of Greater London: a comparison of national and regional geodemographic models

Alex David Singleton¹ and Paul Longley²

Geodemographic classifications are categorical measures representing salient multidimensional population and built environment attributes of small areas. The UK Output Area Classification (OAC) is one such classification, created on behalf of the Office for National Statistics, and was built with an open methodology and entirely from 2011 Census variables. However, one criticism of national classifications such as OAC is that they do not adequately accommodate local or regional structures that diverge from national patterns. In this paper we explore this issue with respect to Greater London. We develop a London classification based upon the OAC methodology, and explore the extent to which these patterns are divergent from the national classification.

Key words London; geodemographics; clustering; regions; GIS

¹Department of Geography and Planning, University of Liverpool, Roxby Building, L69 7ZT
Email: alex.singleton@liverpool.ac.uk

²Department of Geography, UCL, Gower Street, London WC1E 6BT
Email: p.longley@ucl.ac.uk

Revised manuscript received 3 May 2015

Introduction

A geodemographic classification aims to provide a summary of salient socio-spatial characteristics of a small area zonal geography. Presented typically as a nested categorical typology, geodemographic classifications are designed to facilitate comparison between locations, for example, highlighting similarity in patterns of population structure between different parts of a country, or inferring the attitudes of a local population by coding of much more coarsely zoned national surveys. Classifications have been developed within multiple international jurisdictions (Singleton and Spielman 2014), including, but not limited to, Italy (Willis *et al.* 2010), Finland (Takala 2014), Japan (Asai and Yano 2001), Nigeria (Ojo *et al.*, 2013), the Philippines (Ojo *et al.* 2013), and the United States (Spielman and Thill 2008; Skupin and Esperbé 2011).

Indicator measures will typically be captured from a wide range of attributes about the characteristics and behaviours of populations, alongside attributes of the built environment; and will be drawn in different balances from both the public domain (e.g. open data) and private sector sources (e.g. consumer databases etc.). A geodemographic classification is compiled

through a process of cluster analysis, which is a computational technique that groups areas sharing the greatest overall similarity from within a complex of input attributes¹. As such, clusters are formed on the basis of social similarity alone, and are independent of location.

However, there are strong *a priori* reasons to anticipate that differences between regions will impede the utility of national classifications. Arranging areas into clusters optimised to represent the geography of a national extent may for example smooth away import regionally disaggregated local patterns. As such, a key motivation for creating our classification of the Greater London area independent of the rest of the UK arises out of the belief that there is something distinctively different about the geography of the UK's capital city. Such uniqueness can perhaps be illustrated most simply by comparing percentage scores for a number of 2011 Census variables selected to be illustrative of industry specialisation, economic and social diversity (see Table 1). Although these attributes might be argued as arbitrarily selected, differences such as those illustrated are also picked up in the wider literature across a range of perspectives, including but not limited to historic settlement geography

Table 1 Percentage of people with selected 2011 census attributes by regions in England and Wales

Region	Higher managerial*	Level 4 qualification and above	White	Industry finance or insurance industry
East	10.9	25.7	90.8	5.0
East Midlands	9.1	23.6	89.3	2.5
North East	7.5	22.2	95.3	2.8
North West	8.9	24.4	90.2	3.5
South East	12.6	29.9	90.7	4.5
South West	10.2	27.4	95.4	3.7
Wales	7.8	24.5	95.6	3.1
West Midlands	8.7	23.3	82.7	3.1
Yorkshire and The Humber	8.5	23.3	88.8	3.7
London	13.2	37.7	59.8	7.7

*All variables apart from 'higher managerial' are 2011 OAC inputs.

(Longley *et al.* 2011), evolution of urban structures (Masucci *et al.* 2014), economic resilience (Townsend and Champion 2014), concentration of graduates (Hoare and Corver 2010), foreign direct investment (Dimitropoulou *et al.* 2013) and employment specialisation (Faggian *et al.* 2013). Such drivers or outcomes of difference are perhaps less surprising within the context of London as the UK's only global city (Sassen 2001; Taylor 2010). However, we would also argue that such prior reasoning is also not necessarily unique to London, and might for example be similar to the ways in which one might hypothesise that the intrinsic nature of rurality in Scotland is also distinctive from other regions of the UK.

Such issues of regional and local difference have drawn the attention of various academics since the inception of geodemographic methods (Petersen *et al.* 2011; Reibel and Regelson 2011). Notable commentary on the first UK national geodemographic classification developed by Webber (1977) includes Openshaw *et al.* (1980) and response by Webber (1980). Although framed as a critique of geodemographic methods more generally, Openshaw *et al.* (1980) structure part of their discussion around performance differentials of a classification built for a local rather than national extent. However, Webber (1980) argued that classifications built for regional or local extents serve a different purpose to those built for a national geography. Furthermore, regional classifications lose some advantages of national classifications, such as augmentation with supplementary national survey data, or the inability to draw direct comparisons with locations that are not part of the classification.

Within the UK, the most widely used open geodemographic classification is the Office for National Statistics (ONS) Output Area Classification (OAC), which has been constructed from both the 2001 (Vickers and Rees 2007, 2009) and 2011 Census of the Population². The 2011 classification is mapped in Figure 1, and it can be observed that the socio-spatial

structure of London is dominated by a limited number of clusters at the most aggregate level of the classification. Such patterns reinforce our argument that there are clear reasons why a London-based OAC would be useful in practice, as when clustered as part of the national settlement system, 2011 OAC assigns the majority of output areas (OAs) in Greater London into a limited number of clusters, which understates the variegated nature of London's different neighbourhoods. Notwithstanding that all domains need to be represented in the data, this argues for separation of London from the rest of the UK.

In this paper we explore this assertion through the creation of a place-specific geodemographic classification for London, comparing this model to the UK 2011 OAC classification within the geographical extent of Greater London. The classification is created as a subset of a national typology, utilising the same data and methods of construction, however with a constrained geographic extent. We describe the output of this new classification and explore how the patterns represented are divergent from the national model, and second, assess relative performance within the regional context of London. In this work, we define the London region using the Greater London administrative boundary, although the research methods presented could be equally applicable for alternative functional definitions of London, or those of other regions.

Building a London Output Area Classification (LOAC)

In order to maintain comparability between the 2011 OAC and the London place-specific classification, LOAC, the 2011 OAC methods were replicated, with the geographical extent of inputs limited to Greater London. As such, prior to discussing the methodological decisions taken in construction of the LOAC, it is first pertinent to present an overview of how the UK 2011 OAC was constructed.

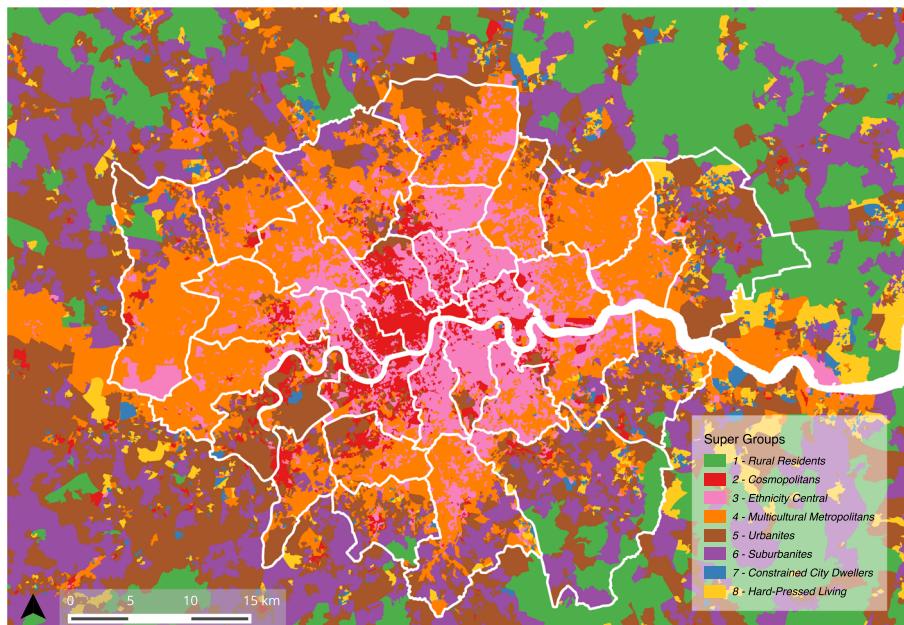


Figure 1 The 2011 OAC super groups in London

Source: contains National Statistics data © Crown copyright and database right 2015 and Ordnance Survey data © Crown copyright and database right 2015

The 2011 OAC was released in July 2014 by the ONS and was developed in partnership with University College London³, comprising data inputs sourced entirely from the UK 2011 Census of the Population. The 2011 OAC is an open geodemographic classification, and the full methodology used to create the classification has been published by the ONS⁴. Those design choices made by the classification builder were in part informed by an extensive national user consultation exercise, which concluded that broad comparability to the 2001 OAC was desired, and that methods used should ideally not represent a radical departure from the previous 2001 classification.

In brief, the construction methods for the 2011 OAC followed an initial process of variable selection, aiming to maintain similarity of input attributes with the 2001 OAC. Candidate variables were considered for their distribution, coverage between the different statistical jurisdictions of the UK and correlation. Sensitivity analysis was also implemented to examine which Census variables were the most important drivers of differentiation between areas. The purpose of this variable evaluation was threefold: to limit variables to those that varied the most between OAs, and thus offered the most discrimination potential; to limit where possible those attributes that had very strong correlation, and thus might overly influence a particular dimension in the output classification; and finally, to ensure a good spread of variables that fulfilled the aim for a

general-purpose classification. After these evaluation steps, a final set of 60 attributes were selected, comprising 58 percentage scores and two ratios. A summary of these input variables can be found in Table 2.

Table 2 2011 OAC input framework

Domain	Sub domain	Variable
Demographic	Age structure	Age bands
	Family structure	Married; children; dependent children
	Ethnicity	Ethnic groups; level of spoken English; EU and ascension EU
Housing	Composition	Density; communal establishments; student households; occupancy rating
	Type	Detached, semi-detached, terraced, flats
	Tenure	Socially rented; private rented; owned or shared ownership
Socio-economic	Health	Day-to-day activities limited a lot or a little; standardized illness ratio
	Employment	Unemployment; full time; part time
	Occupation	Occupation groups
	Education	Level highest qualification
	Mobility	Car ownership; private transport; public transport; active transport

The input data were transformed to make the attributes more normally distributed, and aiming to improve the efficiency of the cluster assignment (Harris *et al.* 2005). After evaluation of a number of different options, the 2011 OAC implemented an inverse hyperbolic sine transformation. Prior to clustering, the input data were also standardised onto a scale ranging from 1 to 0.

Although there are multiple algorithms that can be used to create a geodemographic classification (Adnan *et al.* 2010), for broad comparability to the 2001 OAC, *k*-means was implemented and a top-down clustering approach adopted. The *k*-means algorithm is initiated on the basis of *k* randomly allocated initial seeds. Areas are assigned to their closest seed, and then a new cluster centroid is calculated on the basis of these initial assignments. Areas are then reassigned a cluster if they are closer to a new centroid than the one currently allocated. This process continues iteratively until a no further reallocation occurs. The stochastic nature of *k*-means requires multiple runs, and for the 2011 OAC, clustering was optimised by 10 000 repeated runs, and then selection of those results with the lowest total within sum of squares (TWSS), measuring the overall compactness of the clusters. In a top-down classification, the most aggregate level is clustered first. These divisions are then used to split the input data, which are then clustered separately. This can occur multiple times to create a nested hierarchical structure. As with the 2001 OAC, a three-tier classification was selected for the 2011 OAC, and a range of *k* values tested for each division. These evaluations comprised exploration of both mapped patterns and also empirical measures of cluster fit. The final 2011 OAC comprised eight super groups, 26 groups and 76 sub groups. In collaboration with the ONS, names and descriptions were developed for each of the clusters, aiming to give an impression of distinctive and salient characteristics. The structure and names used in the 2011 OAC are shown in Table 3, alongside the UK and London frequency of OAs within each cluster. For example, within London there are far greater numbers of OA within '3. Ethnicity central', '4. Multicultural metropolitans', although some sub groups within these super groups do not follow these aggregate trends. Within the super group '2. Cosmopolitans', those sub groups within the group '2d. Aspiring and affluent' are also more overrepresented. Such differences give support to a place-specific London regional classification.

LOAC was created using the same inputs as the 2011 OAC, however with a geographic extent limited to London. The numerators and denominators that were used to create both the 2011 OAC and LOAC can be found in Appendix 1. These may not represent the most refined selection of attributes for London, but were retained because our priority was to create a classification that would be directly comparable to the 2011

OAC. Input attributes thus mirror the 2011 OAC for all variables except k035, which related to a standardised illness ratio calculation. The base for this variable in LOAC was London, whereas in the 2011 OAC this was the UK. After attribute calculation, the variables were normalised and standardised as described earlier using an inverse hyperbolic sine and range standardisation. As with the 2011 OAC, *k*-means was implemented as the clustering algorithm, and the classification was built from the top down. A scree plot was created to estimate an appropriate value of *k* for the most aggregate level of the classification (Figure 2). Such plots show the total within sum of squares values derived for classifications created with different values of *k*. The aim of such a plot is to select a value of *k* before a steady off in the rate of declining total within sum of squares; however, as discussed elsewhere (Singleton and Longley 2009), for geodemographic applications where the data have very high dimensionality, these are often not particularly prominent, and the value of *k* becomes a more qualitative choice (e.g. what would be useful to end users or comparable with other classifications) and also informed by a wider range of measures (e.g. cluster frequencies or mapped outputs). In this context, we aimed for a *k* value that would be similar to the 2011 OAC and maintaining reasonably even cluster sizes, and as such, selected *k* = 8, thus mirroring the 2011 OAC. Figure 2 also illustrates that there is a very moderate decrease in total within sum of squares after *k* = 9.

After creating the super group level of the classification, the input data were divided into separate datasets, and clustered to create the group level of the hierarchy. The most parsimonious divisions were found to be between two and four clusters. For each super group a range of different *k* values were tested and the characteristics of the output clusters explored. The final allocations of *k* values were those that created clusters with the most defined and differentiating characterisers, which were assessed through exploratory analysis. Although the initial intention was to create a three-tier classification to mirror the 2011 OAC, the decision was taken to terminate the clustering after two levels. The London data contained far fewer areas than the entire UK, and after a second level of hierarchy the differences emerging between any further divisions of the clusters were minimal, and not materially different enough from their parent group to warrant their own cluster assignment. As such, the final LOAC comprised eight super groups and 19 groups.

The socio-spatial structure of London

The allocations of OAs within each LOAC super group and group are shown in Table 4, with these locations shown visually for super groups in Figure 3. The distribution of OAs was reasonably even between both super groups and groups; however, moderately smaller

Table 3 UK and London OA frequency by the 2011 OAC

Super groups	Groups	Sub groups	London		UK (excluding London)	
			OAs (n)	OAs (%)	OAs (n)	OAs (%)
1. Rural residents	<i>1a. Farming communities</i>	1a1. Rural workers and families	0	0	2454	1.2
		1a2. Established farming communities	0	0	3086	1.5
		1a3. Agricultural communities	0	0	3043	1.5
		1a4. Older farming communities	1	0	1580	0.8
	<i>1b. Rural tenants</i>		1	0	10 163	5
		1b1. Rural life	0	0	5272	2.5
		1b2. Rural white-collar workers	10	0	4801	2.3
	<i>1c. Ageing rural dwellers</i>	1b3. Ageing rural flat tenants	0	0	3600	1.7
			10	0	13 673	6.5
		1c1. Rural employment and retirees	2	0	1028	0.5
		1c2. Renting rural retirement	2	0	1181	0.6
		1c3. Detached rural retirement	0	0	1240	0.6
2. Cosmopolitans	<i>2a. Students around campus</i>		4	0	3449	1.7
		2a1. Student communal living	28	0.1	439	0.2
		2a2. Student digs	1	0	573	0.3
	<i>2b. Inner-city students</i>	2a3. Students and professionals	15	0.1	1505	0.7
			44	0.2	2517	1.2
		2b1. Students and commuters	4	0	1445	0.7
	<i>2c. Comfortable cosmopolitans</i>	2b2. Multicultural student neighbourhoods	423	1.7	798	0.4
			427	1.7	2243	1.1
		2c1. Migrant families	61	0.2	1622	0.8
		2c2. Migrant commuters	9	0	1173	0.6
		2c3. Professional service cosmopolitans	21	0.1	1381	0.7
3. Ethnicity central	<i>3a. Ethnic family life</i>		91	0.3	4176	2.1
		3a1. Established renting families	1479	5.9	479	0.2
		3a2. Young families and students	1138	4.5	413	0.2
	<i>3b. Endeavouring ethnic mix</i>		2617	10.4	892	0.4
		3b1. Striving service workers	1356	5.4	26	0
		3b2. Bangladeshi mixed employment	652	2.6	8	0
	<i>3c. Ethnic dynamics</i>	3b3. Multi-ethnic professional service workers	874	3.5	47	0
			2882	11.5	81	0
		3c1. Constrained neighbourhoods	230	0.9	813	0.4
		3c2. Constrained commuters	7	0	536	0.3
			237	0.9	1349	0.7
4. Multicultural metropolitans	<i>3d. Aspirational techies</i>	3d1. New EU tech workers	1246	5	32	0
		3d2. Established Tech workers	957	3.8	137	0.1
		3d3. Old EU tech workers	1324	5.3	95	0
	<i>4a. Rented family living</i>		3527	14.1	264	0.1
		4a1. Social renting young families	370	1.5	4118	2
		4a2. Private renting new arrivals	91	0.4	3421	1.7
	<i>4b. Challenged Asian terraces</i>	4a3. Commuters with young families	2310	9.2	632	0.3
			2771	11.1	8171	4
		4b1. Asian terraces and flats	1778	7.1	1762	0.9
	<i>4c. Asian traits</i>	4b2. Pakistani communities	10	0	2596	1.3
			1788	7.1	4358	2.2

(Continues)

Table 3. (Continued)

Super groups	Groups	Sub groups	London		UK (excluding London)	
			OAs (n)	OAs (%)	OAs (n)	OAs (%)
5. Urbanites	5a. <i>Urban professionals and families</i>	4c3. Inner city ethnic mix	1403	5.6	886	0.4
			3674	14.6	2740	1.3
			8233	32.8	15 269	7.5
	5b. <i>Ageing urban living</i>	5a1. White professionals	91	0.4	7955	3.8
		5a2. Multi-ethnic professionals with families	1095	4.4	5283	2.5
		5a3. Families in terraces and flats	343	1.4	6357	3.1
			1529	6.2	19 595	9.4
		5b1. Delayed retirement	589	2.4	4372	2.1
		5b2. Communal retirement	111	0.4	4019	1.9
		5b3. Self-sufficient retirement	56	0.2	8426	4.1
6. Suburbanites	6a. <i>Suburban achievers</i>		756	3	16 817	8.1
			2285	9.2	36 412	17.5
		6a1. Indian tech achievers	550	2.2	2859	1.4
		6a2. Comfortable suburbia	1	0	4352	2.1
	6b. <i>Semi-detached suburbia</i>	6a3. Detached retirement living	29	0.1	7145	3.4
		6a4. Ageing in suburbia	39	0.2	4826	2.3
			619	2.5	19 182	9.2
		6b1. Multi-ethnic suburbia	436	1.7	2998	1.4
		6b2. White suburban communities	31	0.1	9829	4.7
		6b3. Semi-detached ageing	20	0.1	8158	3.9
7. Constrained city dwellers	7a. <i>Challenged diversity</i>	6b4. Older workers and retirement	35	0.1	5542	2.7
			522	2	26 527	12.7
			1141	4.5	45 709	21.9
	7b. <i>Constrained flat dwellers</i>	7a1. Transitional Eastern European neighbourhoods	1	0	2169	1
		7a2. Hampered aspiration	21	0.1	3193	1.5
		7a3. Multi-ethnic hardship	194	0.8	3811	1.8
			216	0.9	9173	4.3
	7c. <i>White communities</i>	7b1. Eastern European communities	3	0	1341	0.6
		7b2. Deprived neighbourhoods	4	0	1424	0.7
		7b3. Endeavouring flat dwellers	11	0	1969	1
			18	0	4734	2.3
	7d. <i>Ageing city dwellers</i>	7c1. Challenged transitionaries	9	0	3373	1.6
		7c2. Constrained young families	1	0	3224	1.6
		7c3. Outer city hardship	6	0	2895	1.4
			16	0	9492	4.6
8. Hard-pressed living	8a. <i>Industrious communities</i>	7d1. Ageing communities and families	23	0.1	1503	0.7
		7d2. Retired independent city dwellers	3	0	906	0.4
		7d3. Retired communal city dwellers	3	0	736	0.4
	8b. <i>Challenged terraced workers</i>	7d4. Retired city hardship	2	0	310	0.1
			31	0.1	3455	1.6
			281	1	26 854	12.8
		8a1. Industrious transitions	42	0.2	6406	3.1
	8c. <i>Hard-pressed ageing workers</i>	8a2. Industrious hardship	2	0	5024	2.4
			44	0.2	11 430	5.5
		8b1. Deprived blue-collar terraces	43	0.2	4246	2
		8b2. Hard-pressed rented terraces	6	0	4839	2.3
	8d. <i>Migration and churn</i>	8c1. Ageing industrious workers	8	0	5519	2.7
		8c2. Ageing rural industry workers	0	0	3209	1.5
		8c3. Renting hard-pressed workers	0	0	4053	2
			8	0	12 781	6.2
	8d. <i>Migration and churn</i>	8d1. Young hard-pressed families	2	0	3525	1.7
		8d2. Hard-pressed ethnic mix	148	0.6	4287	2.1
		8d3. Hard-pressed European settlers	0	0	2479	1.2
			150	0.6	10 291	5
			251	1	43 587	21

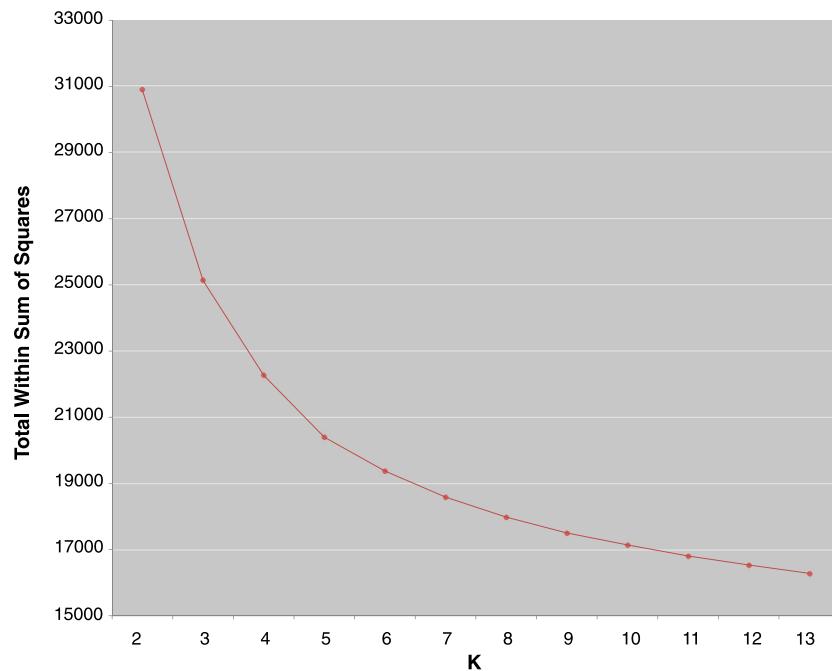


Figure 2 Scree plot on LOAC input data

Table 4 Output area frequency within LOAC super groups and groups

Super group	Group	OA frequency
A	A1	1929
	A2	1312
	3241	
	B1	1409
B	B2	604
	B3	1125
	3138	
C	C1	996
	C2	786
	C3	479
	C4	652
2913		
D	D1	1128
	D2	1248
2376		
E	E1	1864
	E2	1659
3523		
F	F1	1536
	F2	1688
3224		
G	G1	1774
	G2	1935
3709		
H	H1	1296
	H2	1633
2929		

clusters emerged within super group C, accounted for by the requirement for a slightly larger number of groups. The spatial distribution of the super group clusters illustrated good separation, and the emergence of clusters with concentric ring structure, representing both central areas and more peripheral zones radiating out of the core. Two clusters also represented more diffuse patterns, infilling between the other clusters.

The final stage in building the classification was to assign short names and descriptions to each of the clusters making up the typology. There are multiple approaches to this task; however, our preferred method was to create a large table, with rows representing clusters, and columns for each of the input variables to the classification. Cells were filled with statistics representing the deviation from the national mean, and enabled the characteristics of clusters to be compared (see Appendix 2). The super group labels and descriptions were created by the authors, although the Greater London Authority, who sponsored this work were consulted and approved or adjusted as deemed necessary. The LOAC hierarchy of super groups and groups are presented in Table 5, and mapped at the group level in Figure 4. The super group descriptions are presented in the remainder of this section, although, these are also expanded with group descriptions on the supporting website⁵.

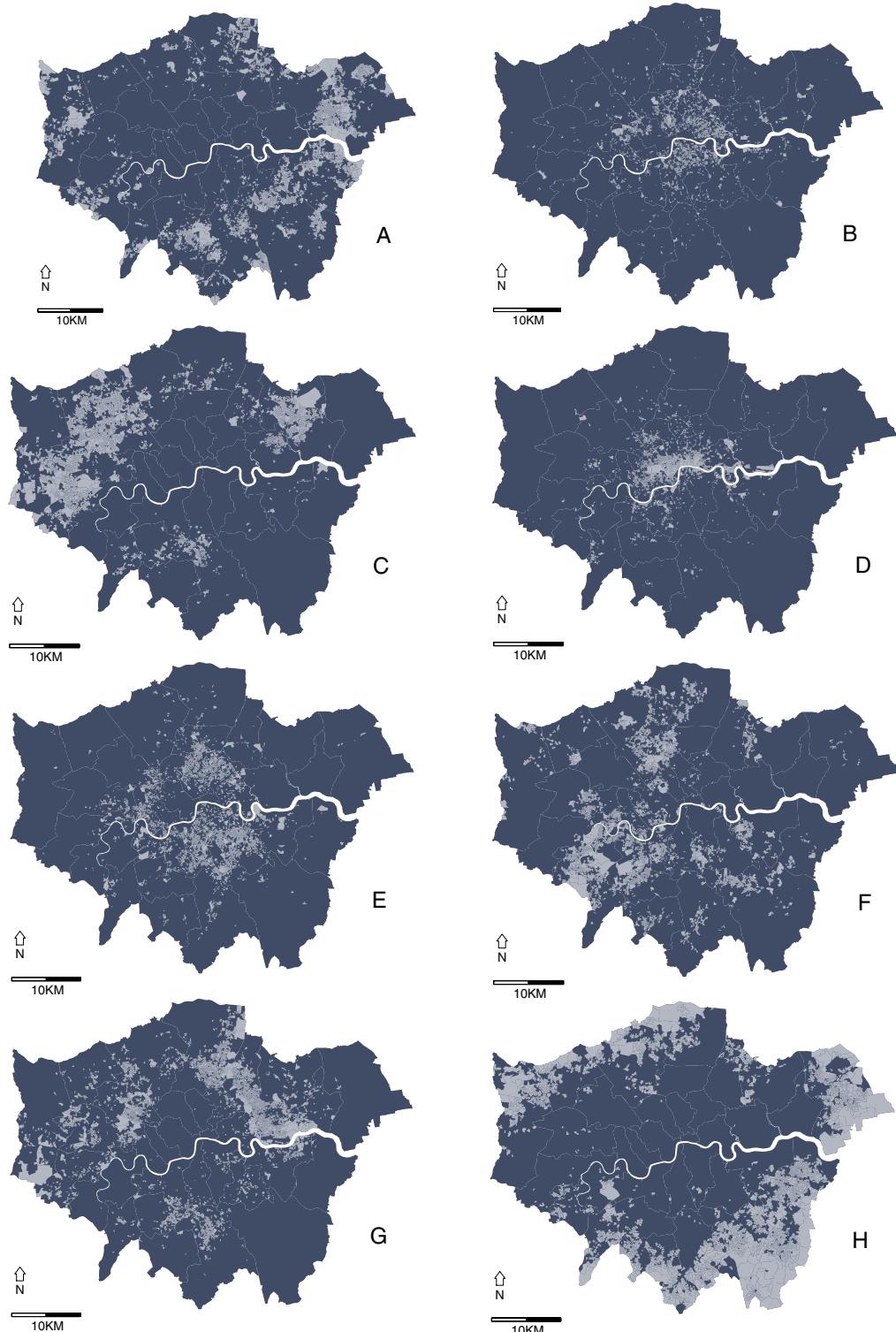


Figure 3 Distribution of the eight LOAC super groups

Source: contains National Statistics data © Crown copyright and database right 2015 and Ordnance Survey data © Crown copyright and database right 2015

Table 5 LOAC super group and group labels

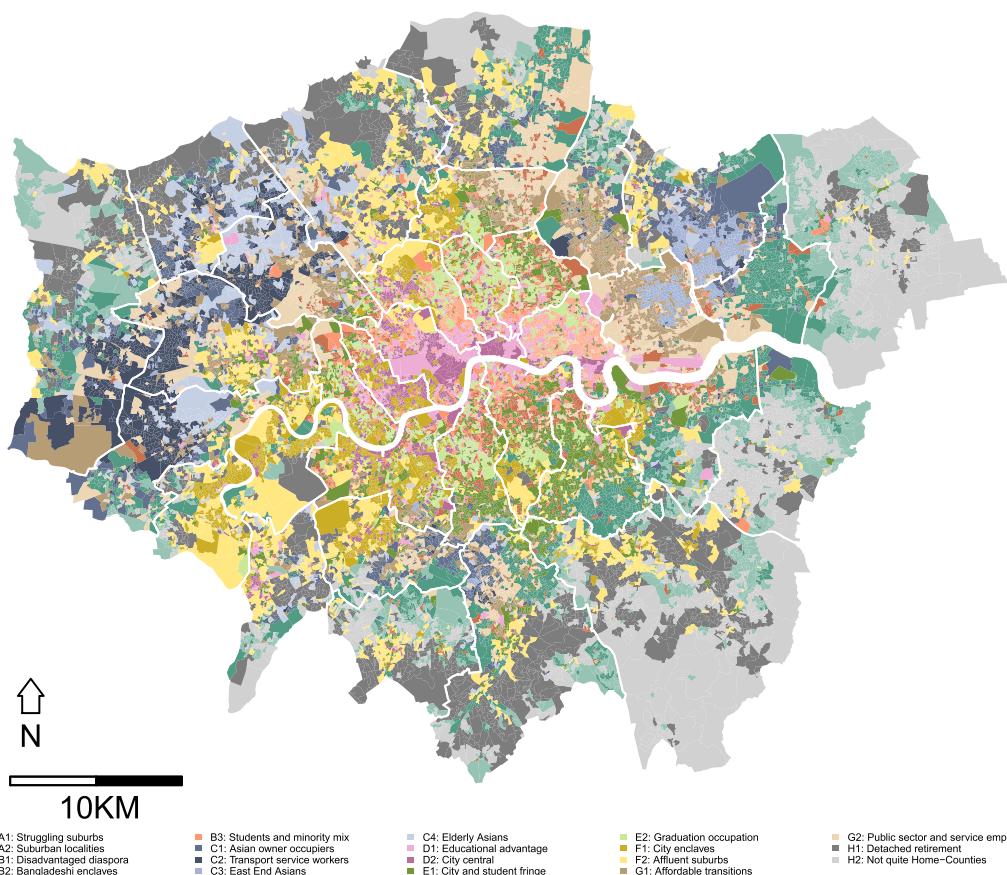
Super group	Group
A. Intermediate lifestyles	A1. Struggling suburbs A2. Suburban localities
B. High-density and high-rise flats	B1. Disadvantaged diaspora B2. Bangladeshi enclaves B3. Students and minority mix
C. Settled Asians	C1. Asian owner occupiers C2. Transport service workers C3. East End Asians C4. Elderly Asians
D. Urban elites	D1. Educational advantage D2. City central
E. City vibe	E1. City and student fringe E2. Graduation occupation
F. London life-cycle	F1. City enclaves F2. Affluent suburbs
G. Multi-ethnic suburbs	G1. Affordable transitions G2. Public sector and service employees
H. Ageing city fringe	H1. Detached retirement H2. Not quite home counties

A. Intermediate lifestyles

Although exhibiting no over-arching discriminating characteristics, households in this super group are predominantly in later stages in the life-cycle, white and born in the UK. Rather few households have dependent children and most live in single family terraced or semi-detached properties. Households that rent their properties do so in the social rented sector. Employment levels are average for London, and are split between full- and part-time working in a range of intermediate occupations. Levels of highest qualifications are below the London average.

B. High-density and high-rise flats

Concentrations of this super group are found in densely populated areas of flats. Many families have children of school age and many residents are of Bangladeshi origins, along with high numbers of households describing themselves as black or as coming from mixed or other ethnic groups. There is a high incidence of households

**Figure 4 LOAC groups with borough boundaries**

Source: contains National Statistics data © Crown copyright and database right 2015 and Ordnance Survey data © Crown copyright and database right 2015

in which the main spoken language is not English. Levels of highest qualifications are below the London average, although some residents are full-time students living in shared accommodation. Levels of unemployment and part-time working are above the London average, while many of those in employment work in administration, or in accommodation and food services industries.

C. Settled Asians

Occupants of this super group typically live in traditional single-family houses, above average numbers of which are owner occupied. Although drawn from the full age range of London residents, the main language spoken in many households is not English. Those in employment have occupations drawn from a wide range of non-professional sectors. Members of this super group identify themselves with their Asian origins, although many are second or subsequent generation British residents.

D. Urban elites

This super group comprises many young professionals working in the science, technology, finance and insurance sectors. Additionally, large numbers of students rent rooms in centrally located communal establishments. Most others rent privately owned flats. Residents are disproportionately drawn from pre-2001 EU countries, and there is also high representation of households from Chinese, Arab and other minority backgrounds.

E. City vibe

There are many young, single professionals in this super group, mostly found within Zone 2 of the London travel

network. Compared with the London average, few individuals originate from the Indian sub-continent, but mixed ethnic groups are well represented, as are migrants from pre-2001 EU countries. A large number of households comprise full-time students in shared or communal establishments. Individuals rent within the private sector, are well qualified and are employed in a range of professional, scientific and technical occupations.

F. London life-cycle

Predominantly white in ethnic composition (including individuals from other pre-2001 EU countries), these households cover the full family life-cycle age spectrum – meaning that, overall, fewer households than the London average include dependent children or students. Residents are highly qualified, employment rates are high and employment is concentrated in the technical, scientific, finance, insurance and real estate industries.

G. Multi-ethnic suburbs

Members of this super group are drawn from a wide range of non-white ethnic groups and white groups are less represented than average for London. Citizens of countries that joined the EU post 2001 are well represented. Many households have young children or children of school age, and the over 65s are not much in evidence. There is above average incidence of family housing in overcrowded terraces, much of it rented within the social housing sector. Levels of unemployment are high, and those who are employed are mainly found working in blue-collar occupations.

Table 6 OA frequency and percentage by LOAC and 2011 OAC in London

	1. Rural residents	2. Cosmopolitans	3. Ethnicity central	4. Multicultural metropolitans	5. Urbanites	6. Suburbanites	7. Constrained city dwellers	8. Hard-pressed living
A. Intermediate lifestyles	0 0.0	0 0.0	327 10.1	2293 70.7	206 6.4	0 0.0	256 7.9	159 4.9
B. High-density and high-rise flats	0 0.0	4 0.1	3128 99.7	0 0.0	0 0.0	0 0.0	6 0.2	0 0.0
C. Settled Asians	0 0.0	7 0.2	19 0.7	2887 99.1	0 0.0	0 0.0	0 0.0	0 0.0
D. Urban elites	0 0.0	1973 83.0	403 17.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
E. City vibe	0 0.0	411 11.7	3024 85.8	87 2.5	0 0.0	0 0.0	1 0.0	0 0.0
F. London life-cycle	0 0.0	1185 36.8	159 4.9	998 31.0	868 26.9	0 0.0	14 0.4	0 0.0
G. Multi-ethnic suburbs	0 0.0	4 0.1	2203 59.4	1502 40.5	0 0.0	0 0.0	0 0.0	0 0.0
H. Ageing city fringe	15 0.5	0 0.0	0 0.0	466 15.9	1211 41.3	1141 39.0	4 0.1	92 3.1

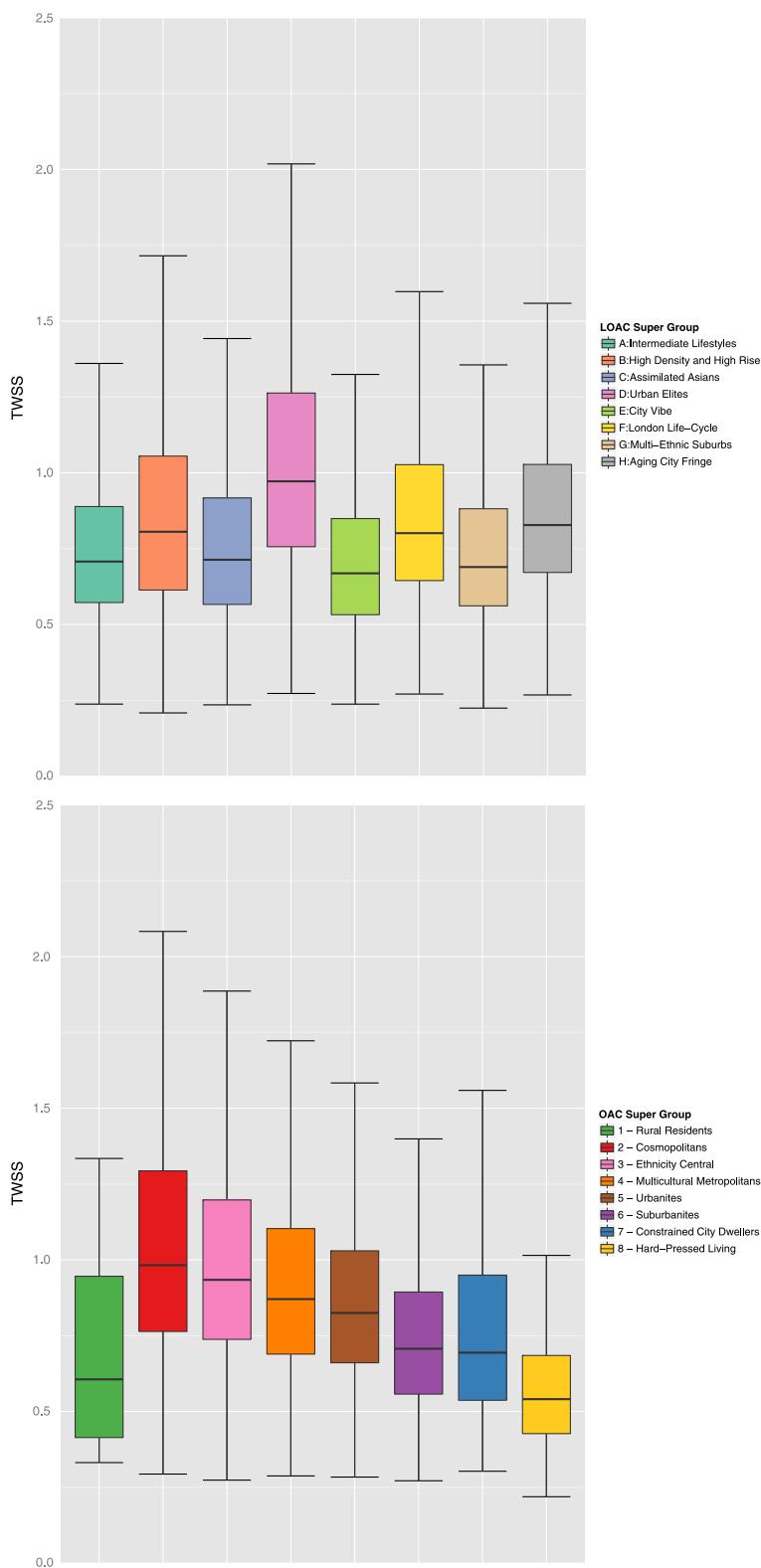


Figure 5 (a) Total within sum of squares by LOAC super group. (b) Total within sum of squares by OAC super group

H. Ageing city fringe

Many of the residents in this super group are over 45, and many are above state pensionable age. There are high levels of marriage and established white residents are very much in evidence. By contrast, representation of ethnic minorities and EU migrants is very low, relative to the London average. Much of the dwelling stock comprises semi-detached and detached houses, occupied at low residential densities. Levels of qualifications are low, as might be expected for these age cohorts. Private vehicle ownership is high, with some households possessing two or more vehicles. Levels of unemployment are very low. Employment is drawn from a range of sectors and this is the only super group in which agriculture is an important source of jobs.

Comparing London OAC and LOAC geodemographic structure

A cross tabulation between the LOAC and 2011 OAC super groups is presented in Table 6. This shows, for example, that LOAC super group 'A. Intermediate lifestyles' corresponds most closely to '4. Multicultural metropolitans'; and similarly, LOAC super groups 'B. High-density and high-rise flats' and 'C. Settled Asians' with '3. Ethnicity central' and '4. Multicultural metropolitans' respectively. Such patterns highlight how although ethnicity might be a defining feature of these areas in OAC, within London, these attributes are far less pronounced. As might be expected, most LOAC

clusters offer greater differentiation than the 2011 OAC: for example, OAs assigned to the LOAC super groups 'D. Urban elites' and 'E. City vibe' correspond to those within the former 2011 OAC classes '2. Cosmopolitans' and '3. Ethnicity central'. Both further highlight the different ethnic composition of London relative to the UK. OAs assigned to the LOAC super group 'G. Multi-ethnic suburbs' are represented in significant part by the 2011 OAC super group '3. Ethnicity central', despite such areas within London being more peripheral to the core. The two LOAC super groups that share little correspondence with the 2011 OAC are 'F. London life-cycle' and 'H. Ageing city fringe', and represent groups that are more distinctive to London than the rest of the UK. Finally, the 2011 OAC super group '1. Rural residents' is predominantly absent from within London.

A total within sum of squares statistic can be calculated for each OA in London and used to compare the fit of the national OAC versus LOAC. These scores are calculated by comparing each attribute to their cluster centroid, which are squared to remove the sign of the value, and then summed for all attributed within each OA. Lower scores identify areas where the attribute values for the OA are closer to the assigned cluster mean, and as such, are well summarised by the cluster description. Direct comparison was possible only at the super group level of both classifications, where LOAC and OAC share the same number of clusters. The output of this analysis is summarised by box plots of the OAC and LOAC clusters, shown in Figure 5(a)

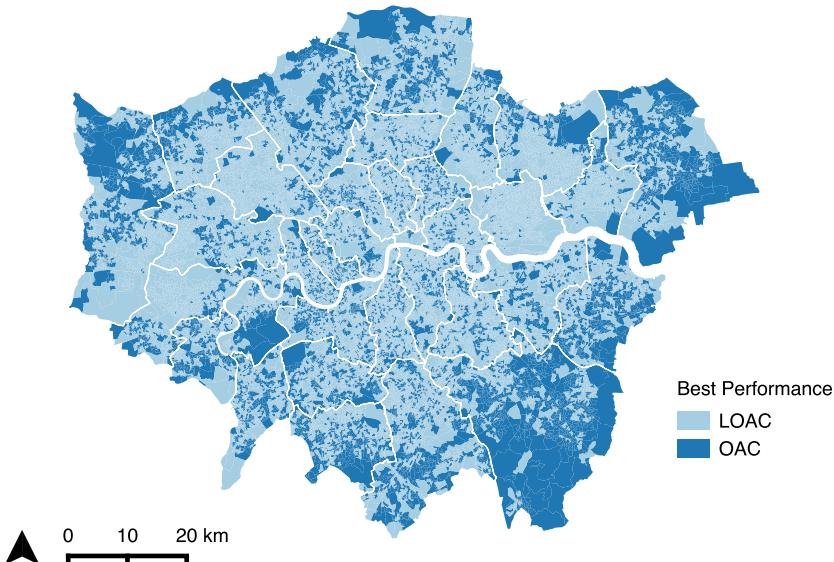


Figure 6 Best OAC or LOAC performance by OA in London

Source: contains National Statistics data © Crown copyright and database right 2015 and Ordnance Survey data © Crown copyright and database right 2015

and (b) respectively. Clusters are separated by coloured boxes that represent the interquartile ranges, the thick central line identifies the median cluster value, and the whiskers extend to the minimum and maximum values (with outliers suppressed from display).

An overarching and encouraging pattern is that LOAC super group clusters have lower scores, and as such, clusters that are more representative of London OA attributes than were observed in OAC. Furthermore, LOAC appears more uniform in the scores between clusters, with greater variance observed in the OAC assignments. The least successful cluster in LOAC is 'D. Urban elites', an outcome which may in part arise because these areas are more fine grained than the OA scale, for example, particular apartment complexes or streets within an OA. Such patterns would result in greater within-OA heterogeneity; however, signature characteristics (e.g. particular employment types) may be those factors that lead to this cluster being differentiated. Figure 5(b) shows the more successful OAC super group clusters in London to include '8. Hard-pressed living' and '1. Rural residents'; however, as shown in Table 6, both pertain to very small numbers of OAs.

The cluster level analysis can be expanded to examine the geography of relative performance by calculating the difference between the OAC and LOAC total within the cluster sum of square scores for each OA. A map of these results is presented in Figure 6, which shows the areas where either the OAC or LOAC classification performed best, in terms of the total within-cluster sum of squares criterion. Although there are OAs scattered across London where the OAC assigned clusters outperformed those of the LOAC, there are evident concentrations of OAs for which OAC performed best towards the outer fringes of London. The socio-spatial structure of these peripheral, lower density and more rural areas are likely more akin to those in the UK as a whole, and thus, more similar to the patterns identified in the national OAC. As such, it could be argued that a more optimal classification for London may be possible by considering how the London regional extent is delineated; perhaps, for example, treating inner London separately.

Discussion and conclusions

This paper was motivated by the hypothesis that a geodemographic classification optimised for a national extent would lose predictive ability within regions where the internal socio-spatial structure deviates significantly from national patterns. This paper has presented a methodological extension to the 2011 OAC, creating a classification for the Greater London region (the LOAC). The LOAC methodological approach mirrored that of the 2011 OAC; however, inputs to the classification were constrained within the Greater London

boundary, and through cluster analysis, aimed to develop a typology that created a better fit to the internal regional rather than national socio-spatial structure.

The output classification comprised eight super groups and 19 hierarchical groups; however, unlike the 2011 OAC, did not adopt a third tier of hierarchy, as it was found that this offered little extra in terms of discriminating features beyond the group level. This is perhaps not unsurprising given the reduced dimensionality of the data input as a result of the constrained regional extent. As is typical in the construction of a geodemographic classification, rates for the input attributes were considered within the output clusters, enabling names and descriptions to be compiled, which was conducted in collaboration with the Greater London Authority.

Comparison between LOAC and the 2011 OAC showed that the internal structure of the London region differed in terms of a number of dimensions. Areas that might represent high ethnic diversity for the UK, can be nearer average if considered within the context of the London region, and additionally, the suburbanisation of ethnic minorities is also more marked. Furthermore, central core areas of London were shown to deviate from national patterns, with a more disaggregate series of clusters both reflecting particular characteristics of the built environment and also being less marked by concentrations of specific ethnic groups. The LOAC performed better than OAC for the majority of inner London, albeit less so towards the suburban fringes of the Greater London Authority boundary.

This work has highlighted that in the context of London there is significant merit to consideration of geodemographic structure at a regional level, and that national models can smooth away key characteristics of internal socio-spatial structure. However, it could be argued that a disadvantage of the regional approach is that such classifications may lose those practical benefits associated with classifications created for the national extent: for example, adding descriptive detail found within national survey estimates or comparative opportunities. However, as the growth of regional data stores (e.g. <http://data.london.gov.uk/>) continues, it is likely that such limitations will be overcome as further small area descriptor data become available. Furthermore, such developments also offer the potential to develop non-census-based classifications for regions, which additionally will offer the benefit of being updatable between each Census.

A number of challenges and directions for future research emerge from our findings. First, there is a need to address how we can balance the additional descriptive power of regional geodemographics relative to the loss of national comparability? Functional regional disaggregation of national extents (Folch and Spielman 2014), and compilation of geodemographics pertaining to their internal structure requires new thinking about how such models could be reassembled to enable

cross-comparison. Geographical weighting of geodemographic inputs across multiple scales may offer some potential in this regard. Second, we also recognise that London is not typical of the rest of the UK, and as such, the internal structures of other regions may not be that divergent to national patterns; however, such considerations also warrant further investigation. Finally, the classifications that were presented in this paper were

all constructed from 2011 Census data, which in this context provides a useful base for comparison between national and regional classifications. However, over time such Census-based classifications become outdated, and there is an argument for open regional classifications to be constructed from non-census sources which can be more prevalently updated. However, as with Census attribute availability between the countries

Appendix A: Variable specification

Variable	Denominator	Variable description	Numerators
k001	KS102EW0001	Age 0 to 4	KS102EW0002
k002	KS102EW0001	Age 5 to 14	KS102EW0003, KS102EW0004, KS102EW0005
k003	KS102EW0001	Age 25 to 44	KS102EW0010, KS102EW0011
k004	KS102EW0001	Age 45 to 64	KS102EW0012, KS102EW0013
k005	KS102EW0001	Age 65 to 89	KS102EW0014, KS102EW0015, KS102EW0016
k006	KS102EW0001	Age 90 and over	KS102EW0017
k007		Density (number of persons per hectare)	KS101EW0008
k008	KS101EW0001	Lives in a communal establishment	KS101EW0005
k009	KS103EW0001	Single	KS103EW0002
k010	KS103EW0001	Married or in a registered same-sex civil partnership	KS103EW0003, KS103EW0004
k011	KS103EW0001	Divorced or separated	KS103EW0005, KS103EW0006
k012	KS201EW0001	White	KS201EW0002, KS201EW0003, KS201EW0004, KS201EW0005
k013	KS201EW0001	Mixed/multiple ethnic group	KS201EW0006, KS201EW0007, KS201EW0008, KS201EW0009
k014	KS201EW0001	Asian/Asian British: Indian	KS201EW0010
k015	KS201EW0001	Asian/Asian British: Pakistani	KS201EW0011
k016	KS201EW0001	Asian/Asian British: Bangladeshi	KS201EW0012
k017	KS201EW0001	Asian/Asian British: Chinese and other	KS201EW0013, KS201EW0014
k018	KS201EW0001	Black/African/Caribbean/Black British	KS201EW0015, KS201EW0016, KS201EW0017
k019	KS201EW0001	Arab or other ethnic groups	KS201EW0018, KS201EW0019
k020	KS204EW0001	United Kingdom and Ireland	KS204EW0002, KS204EW0003, KS204EW0004, KS204EW0005, KS204EW0006, KS204EW0007
k021	KS204EW0001	Other EU: Member countries in March 2001	KS204EW0008
k022	KS204EW0001	Other EU: Accession countries April 2001 to March 2011	KS204EW0009
k023	QS205EW0001	Main language is not English and cannot speak English well or at all	QS205EW0005, QS205EW0006
k024	KS105EW0001	No children household	KS105EW0005, KS105EW0008
k025	KS105EW0001	Non-dependent children household	KS105EW0007, KS105EW0010, KS105EW0012
k026	KS105EW0001	Full-time student household	KS105EW0014
k027	KS401EW0005	Whole house or bungalow: detached	KS401EW0008
k028	KS401EW0005	Whole house or bungalow: semi-detached	KS401EW0009
k029	KS401EW0005	Whole house or bungalow: terrace and end-terrace	KS401EW0010
k030	KS401EW0005	flats	KS401EW0011, KS401EW0012, KS401EW0013
k031	KS402EW0001	Owned or shared ownership	KS402EW0002, KS402EW0003, KS402EW0004
k032	KS402EW0001	Social rented	KS402EW0005, KS402EW0006
k033	KS402EW0001	Private rented	KS402EW0007, KS402EW0008

(Continues)

Appendix A: (Continued)

Variable	Denominator	Variable description	Numerators
k034	QS408EW0001	Occupancy room rating -1 or less	QS408EW0005, QS408EW0006
k035		Day-to-day activities limited a lot or a little standardised Illness ratio	KS301EW0002, KS301EW0003, KS301EW0005, KS301EW0006
k036	KS301EW0001	Provides unpaid care	KS301EW0014, KS301EW0015, KS301EW0016
k037	KS501EW0001	Highest level of qualification: Level 1, Level 2 or apprenticeship	KS501EW0003, KS501EW0004, KS501EW0005
k038	KS501EW0001	Highest level of qualification: Level 3 qualifications	KS501EW0006
k039	KS501EW0001	Highest level of qualification: Level 4 qualifications and above	KS501EW0007
k040	KS501EW0001	Schoolchildren and full-time students: age 16 and over	KS501EW0009, KS501EW0010
k041	KS404EW0001	2 or more cars or vans in household	KS404EW0004, KS404EW0005, KS404EW0006
k042	QS701EW0001	Public transport	QS701EW0003, QS701EW0004, QS701EW0005, QS701EW0006
k043	QS701EW0001	Private transport	QS701EW0007, QS701EW0008, QS701EW0009
k044	QS701EW0001	On foot, bicycle or other	QS701EW0010, QS701EW0011, QS701EW0012
k045	KS601EW0001	Unemployed	KS601EW0005
k046	KS604EW0001	Part-time	KS604EW0002, KS604EW0003
k047	KS604EW0001	Full-time	KS604EW0004, KS604EW0005
k048	KS605EW0001	Agriculture, forestry and fishing	KS605EW0002
k049	KS605EW0001	Mining, quarrying and construction	KS605EW0003, KS605EW0007
k050	KS605EW0001	Manufacturing	KS605EW0004
k051	KS605EW0001	Energy, water and air conditioning supply	KS605EW0005, KS605EW0006
k052	KS605EW0001	Wholesale and retail trade; repair of motor vehicles and motor cycles	KS605EW0008
k053	KS605EW0001	Transport and storage	KS605EW0009
k054	KS605EW0001	Accommodation and food service activities	KS605EW0010
k055	KS605EW0001	Information and communication and professional, scientific and technical activities	KS605EW0011, KS605EW0014
k056	KS605EW0001	Financial, insurance and real estate activities	KS605EW0012, KS605EW0013
k057	KS605EW0001	Administrative and support service activities	KS605EW0015
k058	KS605EW0001	Public administration and defence; compulsory social security	KS605EW0016
k059	KS605EW0001	Education	KS605EW0017
k060	KS605EW0001	Human health and social work activities	KS605EW0018

Appendix B: Input attribute means for LOAC

Variable description	A	B	C	D	E	F	G	H	A1	A2	B1	B2	B3	C1	C2	C3	C4	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	mean
Age 0-4	7.8	7.1	5.2	6.7	7.4	8.7	5.3	8.2	7.0	9.5	8.2	6.8	6.9	7.4	8.3	6.2	4.9	5.4	7.4	5.8	8.0	6.8	8.7	5.1	5.5	7.1		
Age 5-14	13.5	13.5	12.5	4.4	8.5	9.8	13.0	12.0	14.3	12.4	15.1	14.7	10.8	12.8	12.3	13.9	11.4	4.3	4.6	9.3	7.7	9.8	9.8	11.0	14.9	12.3	11.8	11.1
Age 25 to 44	29.1	35.2	32.0	54.4	43.9	37.1	36.3	22.8	29.7	28.3	32.8	35.3	38.0	30.4	34.1	34.0	30.4	55.4	53.5	42.0	46.1	39.4	35.1	41.0	32.0	21.3	24.0	36.0
Age 45 to 64	23.3	18.7	22.4	15.5	19.3	23.2	18.8	29.1	22.7	24.1	19.5	15.7	19.4	24.0	21.2	18.9	24.0	13.0	17.8	19.4	19.2	23.0	23.4	17.6	20.0	29.8	28.6	21.3
Age 65 to 89	12.5	8.9	11.3	6.9	8.7	12.6	8.1	18.2	10.9	14.8	8.5	7.8	9.9	12.0	10.0	8.2	13.9	4.6	8.9	8.7	8.6	11.3	13.8	7.0	9.0	18.9	17.6	10.8
Age 90 and over	0.7	0.4	0.5	0.4	0.4	0.8	0.4	0.8	0.5	0.9	0.4	0.3	0.5	0.3	0.5	0.2	0.5	0.5	0.4	0.4	0.6	1.0	0.3	0.4	0.9	0.7	0.5	0.5
Persons per hectare (D)	74.3	237.8	91.2	224.2	133.1	84.7	119.1	46.7	81.2	64.3	230.4	264.6	232.5	82.8	91.8	145.8	63.1	263.2	188.9	118.9	149.0	102.3	68.6	127.7	111.2	37.5	53.9	124.0
Lives in a communal establishment	0.5	0.9	0.5	2.4	1.1	1.0	0.7	0.7	0.4	0.6	0.4	1.1	1.5	0.3	0.5	0.5	0.8	4.2	0.7	1.4	0.8	0.5	1.5	0.8	0.5	1.0	0.4	0.9
Married or in a registered same-sex civil partnership	39.0	51.2	35.1	59.7	56.1	39.5	45.3	27.6	40.3	37.2	48.6	50.2	55.2	35.1	37.5	33.4	64.8	55.1	53.5	59.0	40.5	38.6	46.4	44.4	25.5	29.2	44.1	39.8
Divorced or separated	40.6	28.8	51.3	29.3	29.1	45.7	37.3	57.7	40.2	41.2	29.2	33.7	25.7	50.5	51.1	50.5	53.3	25.1	33.0	30.6	27.3	46.4	45.0	38.7	36.0	61.4	54.7	39.8
White	13.5	14.9	8.2	8.6	10.9	9.6	12.8	7.8	13.4	13.7	17.1	11.1	14.2	8.9	8.5	7.5	7.4	8.3	8.9	11.6	10.1	8.9	10.2	11.1	14.3	6.3	9.0	11.0
Mixed/multiple ethnic group	71.0	44.4	36.8	70.2	66.4	80.8	44.1	82.8	63.3	82.3	40.3	37.4	53.5	47.7	27.4	20.0	43.8	62.3	77.3	61.5	72.0	85.6	76.5	45.1	43.1	76.9	87.5	61.6
Asian/Asian British:	4.7	6.6	3.8	4.9	6.4	4.2	6.0	2.6	5.5	3.4	7.2	4.5	7.0	4.5	3.3	3.1	3.8	5.4	4.4	6.9	5.9	4.1	4.3	5.5	6.5	3.0	2.4	5.0
Indian	2.7	1.9	22.1	4.5	2.2	3.7	6.4	5.2	3.0	2.2	1.9	2.0	2.0	13.6	30.8	24.4	22.9	5.5	3.5	2.8	1.6	2.0	5.2	7.9	4.9	8.6	2.5	5.9
Asian/Asian British:	1.2	1.1	8.2	1.1	0.9	0.9	4.9	1.0	1.5	0.7	1.1	1.1	1.0	5.7	7.9	18.5	4.8	1.4	0.7	1.1	0.6	0.7	1.2	6.7	3.2	1.6	0.6	2.4
Pakistani	0.9	8.5	2.8	1.1	1.5	0.3	3.6	0.4	1.2	0.5	2.0	34.1	2.8	0.8	9.9	0.9	1.9	0.5	1.0	2.0	0.3	0.4	3.9	3.3	0.4	0.4	2.4	
Asian/Asian British:	3.9	5.6	11.8	8.3	4.4	4.4	8.0	3.7	4.7	2.8	5.4	4.7	6.3	10.4	14.5	10.5	11.5	10.4	6.4	5.0	3.8	3.2	5.5	8.8	7.2	4.9	2.6	6.2
Bangladeshi	13.9	26.4	10.4	4.9	15.1	3.4	22.6	3.0	18.5	7.2	36.8	12.5	20.9	12.2	10.6	10.4	7.2	7.2	2.9	18.5	11.2	2.5	4.2	17.7	27.1	2.8	3.2	13.0
Chinese and other Black/African/Caribbean/Black British	1.7	5.4	4.1	5.0	3.1	2.3	4.5	1.2	2.2	0.9	5.3	3.6	6.7	3.5	4.7	3.3	5.1	5.9	4.1	3.3	2.9	1.7	2.8	4.4	4.6	1.8	0.8	3.4
Arab or other ethnic groups	79.4	58.1	53.2	51.9	66.7	72.3	54.9	85.7	74.6	86.4	58.3	58.0	60.9	45.2	47.9	54.9	48.5	55.1	65.0	68.5	72.5	72.1	50.2	59.2	80.6	89.8	65.5	
United Kingdom and Ireland	1.7	4.8	2.5	11.5	6.0	5.3	3.2	1.6	2.0	1.3	3.7	4.2	6.5	2.8	2.3	1.8	2.9	10.8	12.1	5.1	7.0	6.7	4.0	3.8	2.7	2.1	1.2	4.4
Other EU: Accession countries April 2001 to March 2011	3.1	3.4	6.2	4.1	4.1	3.4	8.4	1.7	3.8	2.0	3.4	2.5	3.8	6.5	6.5	4.4	6.7	5.2	3.2	5.5	2.5	2.3	4.4	11.1	5.9	2.1	1.3	4.4
Main language is not English and cannot speak English well or at all	2.1	6.4	6.9	2.3	2.9	1.5	7.1	0.9	2.7	1.1	6.1	9.7	5.0	5.1	9.2	8.8	5.5	3.1	1.5	3.5	2.3	0.9	1.9	7.6	6.6	1.2	0.7	3.8

No children household	11.7	7.9	9.9	21.3	16.1	18.9	9.6	16.1	10.7	13.1	6.2	7.9	10.0	11.4	8.3	7.3	11.3	19.9	22.5	14.7	17.6	20.5	17.4	11.7	7.7	16.0	16.3	13.7	
Non-dependent children household	12.1	8.3	14.2	2.8	6.5	7.0	9.6	14.3	12.3	11.9	8.9	8.5	7.4	14.6	14.6	12.2	14.8	2.6	3.0	7.0	5.9	6.1	7.8	11.2	14.4	14.2	9.4		
Full-time student household	0.2	1.1	0.7	1.9	0.9	0.4	0.8	0.1	0.2	0.1	0.4	2.2	1.4	0.4	0.6	1.5	0.6	3.0	0.9	0.8	1.0	0.3	0.4	1.3	0.4	0.1	0.1	0.7	
Whole house or bungalow: detached	4.5	0.9	7.7	0.7	2.2	7.3	4.9	24.0	4.4	4.8	1.0	1.0	0.7	5.5	6.4	8.1	12.5	0.7	0.8	2.8	1.5	3.7	10.6	4.8	5.0	41.2	10.4	6.4	
Whole house or bungalow: semi-detached	27.3	1.6	36.7	1.6	6.8	18.1	13.3	50.3	24.2	31.9	2.0	1.7	1.0	32.0	36.2	16.3	59.4	1.2	1.9	8.7	4.7	14.0	21.8	11.8	14.7	46.4	53.5	19.3	
Whole house or bungalow: terrace and end-terrace	41.1	4.0	34.9	5.6	21.8	27.7	30.0	17.9	45.9	34.1	4.5	5.3	2.6	49.6	30.5	51.9	5.3	3.5	7.5	22.2	21.4	38.7	17.7	29.7	30.3	4.3	28.7	23.6	
Flats	26.9	93.5	20.6	92.0	69.1	46.8	51.7	7.7	25.5	29.1	92.4	91.9	95.6	12.9	26.8	23.7	22.8	94.6	89.8	66.2	72.4	43.5	49.8	53.7	49.9	8.2	7.3	50.6	
Owned or shared ownership	55.3	16.3	66.3	37.2	39.0	64.9	36.6	89.1	52.0	60.1	13.5	17.6	19.0	73.6	57.6	57.8	72.0	30.0	43.7	40.1	37.9	64.5	65.3	41.0	32.6	89.5	88.9	50.1	
Social rented	30.8	64.7	8.2	10.6	28.7	6.5	32.8	2.7	32.8	27.9	72.6	61.1	56.9	6.3	13.3	8.5	4.6	14.8	6.8	27.6	29.9	6.6	6.3	18.6	45.8	2.1	3.2	24.1	
Private rented	12.9	17.6	24.3	50.0	31.1	27.1	29.3	7.4	14.2	11.1	12.5	20.0	22.8	19.1	27.7	32.5	22.0	53.4	47.0	31.1	31.0	27.3	27.0	39.2	20.1	7.5	7.3	24.5	
Occupancy room rating – 1 or less	13.8	35.0	19.1	32.2	24.8	11.3	29.5	3.3	16.0	10.6	33.4	37.7	35.6	13.2	26.1	26.7	14.2	41.4	24.0	24.5	25.0	10.1	12.4	31.0	28.1	3.1	3.5	21.1	
Day-to-day activities limited a lot or a little	119.2	125.8	99.1	63.7	101.6	78.8	105.5	97.2	114.8	125.7	123.2	122.4	131.0	96.9	102.0	98.3	99.6	65.6	62.0	102.7	100.3	65.2	91.1	92.0	117.8	92.5	100.9	100.1	
standardised illness ratio																													
Provides unpaid care	9.7	8.0	9.6	5.0	7.3	8.0	7.9	11.3	9.4	10.0	7.7	8.8	7.8	9.6	9.4	9.4	10.2	4.6	5.4	7.4	7.3	7.3	8.5	7.2	8.5	11.6	11.0	8.4	
Highest level of qualification: Level 1, Level 2 or apprenticeship	33.5	25.2	25.8	11.3	18.6	18.2	25.8	33.0	32.3	35.3	28.5	23.5	22.0	27.9	22.0	27.9	25.2	24.7	24.0	12.2	10.5	21.1	15.8	14.2	21.7	22.9	28.5	36.7	24.2
Highest level of qualification: Level 3 qualifications	10.5	10.3	10.3	9.4	10.0	9.8	10.1	11.7	10.3	10.7	10.0	10.9	10.4	10.9	9.5	10.1	10.5	11.1	8.0	10.5	9.5	9.0	10.6	10.0	10.1	11.3	12.0	10.3	
Highest level of qualification: Level 4 qualifications and above	22.5	28.6	31.6	62.6	48.3	54.9	28.7	32.7	23.5	21.0	23.7	27.9	35.1	31.2	28.0	31.3	36.7	58.3	66.5	42.9	54.3	62.7	47.8	34.0	23.8	40.9	26.2	38.0	
Schoolchildren and full-time students: age 16 and over	8.6	13.2	12.2	11.3	9.4	6.6	12.9	7.0	9.7	6.9	12.5	15.9	12.7	10.5	12.7	16.8	10.7	15.9	7.2	9.9	8.8	5.8	7.3	13.2	12.7	7.6	6.4	10.1	
2 or more cars or vans in household	22.4	4.0	29.1	6.6	8.2	21.7	11.5	44.9	20.3	25.4	4.6	3.9	3.4	30.4	26.6	20.7	36.5	4.5	8.6	9.1	7.2	19.1	24.2	11.1	11.9	50.9	40.0	18.3	
Public transport	24.9	32.6	28.2	44.4	41.3	36.3	33.7	23.2	26.3	22.8	32.1	29.5	34.8	28.4	27.7	30.6	26.8	43.5	45.3	40.3	42.4	39.8	33.2	38.5	29.2	22.8	23.5	33.0	
Private transport	29.9	9.6	27.0	9.0	12.0	20.9	17.8	35.7	27.3	33.7	11.6	7.8	8.1	29.1	27.2	18.7	29.8	7.7	10.2	14.2	9.6	16.5	24.9	16.8	18.7	33.9	37.2	20.3	
On foot, bicycle or other	6.0	10.6	5.6	16.3	12.7	9.8	6.5	5.1	6.1	5.8	7.7	12.3	13.3	5.8	5.9	5.6	5.2	16.6	16.1	10.4	15.3	11.6	8.1	6.9	6.2	4.8	5.3	8.9	
Unemployed	6.0	8.4	4.9	3.4	5.3	3.0	7.0	2.8	6.5	5.3	9.5	8.3	7.1	4.6	5.4	6.0	4.2	4.0	2.8	5.7	4.7	2.5	3.4	6.2	7.7	2.6	3.0	5.2	
Part-time	28.2	32.2	29.4	15.1	21.6	21.3	30.9	28.6	28.9	27.2	34.6	35.9	27.1	28.1	28.8	35.2	27.7	17.1	13.3	23.8	19.1	19.3	23.1	28.7	32.9	29.1	28.1	26.2	
Full-time	71.8	67.8	70.6	84.9	78.4	78.7	69.1	71.4	71.1	72.8	65.4	72.9	71.9	71.2	64.8	72.3	82.9	86.7	76.2	80.9	80.7	76.9	71.3	67.1	70.9	71.9	73.8		
Agriculture, forestry and fishing	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	

(Continues)

Appendix B: (Continued)

Variable description	A	B	C	D	E	F	G	H	A1	A2	B1	B2	B3	C1	C2	C3	C4	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	London mean
Mining, quarrying and construction	9.6	5.0	8.2	3.2	4.5	5.1	8.3	9.3	9.0	10.5	6.0	3.6	4.5	8.6	8.3	7.0	8.4	3.3	3.0	5.4	3.6	3.9	6.2	8.8	7.9	7.8	10.5	6.7
Manufacturing	4.2	2.3	4.5	2.2	2.3	2.9	3.2	4.2	3.9	4.6	2.6	2.3	1.9	4.0	5.8	3.8	4.0	2.1	2.3	2.4	2.1	2.5	3.1	3.4	3.8	4.5	4.5	3.2
Energy, water and air conditioning supply	1.1	0.6	0.6	0.4	0.4	0.5	0.7	0.8	1.0	1.3	0.7	0.4	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.4	0.4	0.5	0.4	0.5	0.6	1.0	0.6	1.0
Wholesale and retail trade; repair of motor vehicles and motor cycles	15.3	13.9	17.8	8.6	10.3	9.7	15.5	13.4	15.3	15.4	15.0	14.3	12.4	16.3	19.0	19.9	17.1	9.4	7.9	11.2	9.4	8.2	11.1	14.6	16.3	13.2	13.6	13.2
Transport and storage	7.3	5.8	8.0	2.0	3.2	2.8	6.5	4.8	7.5	7.0	6.9	5.6	4.6	7.1	11.0	7.1	6.2	2.4	1.7	3.8	2.6	2.1	3.5	5.6	7.3	3.8	5.6	5.1
Accommodation and food service activities	5.0	10.5	6.8	5.4	6.4	3.6	9.2	3.0	5.5	4.3	9.8	13.6	9.7	5.5	8.8	8.1	5.5	7.3	3.8	7.4	5.2	3.0	4.0	10.1	8.4	2.9	3.0	6.3
Information and communication and professional, scientific and technical activities	9.8	13.7	12.2	28.6	24.2	25.5	10.9	15.3	9.5	10.2	9.6	14.8	18.1	12.6	10.0	10.8	15.3	27.0	30.1	20.2	28.6	29.7	21.6	12.8	9.2	18.1	13.0	17.2
Financial, insurance and real estate activities	6.7	6.8	6.8	19.9	9.4	13.3	5.5	11.1	6.2	7.5	5.1	8.1	8.3	7.2	4.8	7.4	8.0	16.7	22.8	7.7	11.2	15.9	11.0	5.9	5.2	12.2	10.3	9.6
Administrative and support service activities	6.2	7.7	6.2	4.3	5.5	4.6	7.9	4.3	6.5	5.7	9.4	5.9	6.7	5.6	7.7	6.4	5.4	4.6	4.1	6.1	4.9	4.3	4.8	7.9	7.8	4.1	4.5	5.9
Public administration and defence; compulsory social security	6.2	4.6	4.9	3.5	4.9	4.8	4.7	6.5	6.1	6.4	4.9	4.0	4.7	5.8	4.0	5.3	4.5	3.6	3.5	5.2	4.5	4.3	5.4	4.4	5.0	5.6	7.3	5.1
Education	9.9	9.5	8.8	6.9	10.3	10.5	9.2	11.3	10.1	9.6	9.2	10.8	9.2	10.0	6.5	9.4	9.3	7.4	6.4	10.8	9.8	9.6	11.4	9.0	9.4	11.4	11.2	9.6
Human health and social work activities	13.0	12.5	10.4	7.6	10.3	9.6	12.3	10.6	13.6	11.9	14.3	10.1	11.5	11.3	9.3	10.2	10.4	8.7	6.7	11.3	9.3	8.6	10.6	11.0	13.4	11.0	10.4	10.9

of the UK, a functional region would likely comprise multiple data stores, presenting inputs that would not necessarily be compatible or available for a full regional extent. We would argue that greater regional thinking is required when designing open data portals (Rae and Singleton 2015) to maximise the potential impact of these resources within regions.

Acknowledgements

This work was supported by ESRC grants ES/L014769/1; ES/L013800/1; ES/L011840/1 and ES/K004719/1 and EPSRC Grant EP/J004197/1, and was created in partnership with the Greater London Authority. We would also like to thank the editors and anonymous reviewers for useful commentary during the preparation of this manuscript.

Notes

References

- of the UK, a functional region would likely comprise multiple data stores, presenting inputs that would not necessarily be compatible or available for a full regional extent. We would argue that greater regional thinking is required when designing open data portals (Rae and Singleton 2015) to maximise the potential impact of these resources within regions.

Acknowledgements

This work was supported by ESRC grants ES/L014769/1; ES/L013800/1; ES/L011840/1 and ES/K004719/1 and EPSRC Grant EP/J004197/1, and was created in partnership with the Greater London Authority. We would also like to thank the editors and anonymous reviewers for useful commentary during the preparation of this manuscript.

Notes

 - 1 Further discussion on the development of geodemographics, and considerations of open versus closed methods of construction can be found elsewhere, so are not repeated here (see, for example, Singleton and Longley 2009a).
 - 2 www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/ns-2011-area-classifications/index.html
 - 3 www.opengeodemographics.com
 - 4 www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/ns-2011-area-classifications/index.html
 - 5 www.opengeodemographic.com

References

Adnan M, Longley P A, Singleton A D and Brunsdon C 2010 Towards real-time geodemographics: clustering algorithm performance for large multidimensional spatial databases: towards real-time geodemographics *Transactions in GIS* 14 283–97.

Asai Y and Yano K 2001 Construction of geodemographics based on the 1995 Population Census of Japan *Papers and Proceedings of the Geographic Information Systems Association* 10 279–84.

Dimitropoulou D, McCann P and Burke S P 2013 The determinants of the location of foreign direct investment in UK regions *Applied Economics* 45 3853–62.

Faggian A, Comunian R, Jewell S and Kelly U 2013 Bohemian graduates in the UK: disciplines and location determinants of creative careers *Regional Studies* 47 183–200.

Folch D and Spielman S E 2014 Identifying regions based on flexible user defined constraints *International Journal of Geographic Information Science* 28 164–84.

Harris R J, Sleight P and Webber R J 2005 Geodemographics GIS and Neighbourhood Targeting Wiley, London.

Hoare A and Corver M 2010 The regional geography of new young graduate labour in the UK *Regional Studies* 44 77–94.

Longley P A, Cheshire J A and Mateos P 2011 Creating a regional geography of Britain through the spatial analysis of surnames *Geoforum* 42 506–16.

Masucci A P, Stanilov K and Batty M 2014 Exploring the evolution of London's street network in the information space: a dual approach *Physical Review E* 89 012805.

Ojo A, Vickers D and Ballas D 2013 Creating a small scale area classification for understanding the economic social and housing characteristics of small geographical areas in the Philippines *Regional Science Policy & Practice* 5 1–24.

Openshaw S, Cullingford D and Gillard A 1980 A critique of the national classifications of OPCS/PRAG *Town Planning Review* 51 421.

Petersen J, Gibin M, Longley P A, Mateos P, Atkinson P and Ashby D I 2011 Geodemographics as a tool for targeting neighbourhoods in public health campaigns *Journal of Geographical Systems* 13 173–92.

Rae A and Singleton A D 2015 Putting big data in its place: a Regional Studies and Regional Science perspective *Regional Studies Regional Science* 2 1–5.

Reibel M and Regelson M 2011 Neighborhood racial and ethnic change: the time dimension in segregation *Urban Geography* 32 360–82.

Sassen S 2001 The Global City: New York London Tokyo 2nd ed Princeton University Press, Princeton, NJ.

Singleton A D and Longley P A 2009 Creating open source geodemographics – refining a national classification of census output areas for applications in higher education papers *Regional Science* 88 643–66.

Singleton A D and Spielman S 2014 The past present and future of geodemographic research in the United States and United Kingdom *Professional Geographer* 66 558–67.

Skupin A and Esperbé A 2011 An alternative map of the United States based on an n-dimensional model of geographic space *Journal of Visual Languages & Computing* 22 290–304.

Spielman S E and Thill J C 2008 Social area analysis data mining and GIS computers *Environment and Urban Systems* 32 110–22.

Taylor P J 2010 Specification of the World City Network *Geographical Analysis* 33 181–94.

Takala V 2014 Helsinki: a software sorted city?A case study of the geodemographics industry in Finland Fennia *International Journal of Geography* 192 140–53.

Townsend A and Champion T 2014 The impact of recession on city regions: the British experience 2008–2013 *Local Economy* 29 38–51.

Vickers D and Rees P 2007 Creating the UK national statistics 2001 output area classification *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 379–403.

Vickers D and Rees P 2009 Ground-truthing geodemographics *Applied Spatial Analysis and Policy* 4 3–21.

Webber R J 1977 An Introduction to the National Classification of Wards and Parishes (No 23) Planning Research Applications Group Technical Paper Centre for Environmental Studies, London.

Webber R J 1980 A response to the critique of the national classifications of OPCS/PRAG *The Town Planning Review* 51 440–50.

Willis I, Gibin M, Barros J and Webber R J 2010 Applying neighbourhood classification systems to natural hazards: a case study of Mt Vesuvius *Natural Hazards* 70 1–22.