OAC Explorer: Interactive exploration and comparison of multivariate socioeconomic population characteristics

Aidan Slingsby[1], Jason Dykes[1], Jo Wood[1] and Robert Radburn[2]

[1]giCentre, Department of Information Science,
City University London, Northampton Square, London, EC1V 0HB, UK
Tel. +44 (0)20 7040 0180
{sbbb717 | jwo | jad7}@soi.city.ac.uk, http://gicentre.org/

[2]Leicestershire County Council
County Hall, Glenfield, Leicestershire, LE3 8RJ, UK
Robert.Radburn@leics.gov.uk

KEYWORDS: visualization, geodemographics, demography, oac, census

1. Introduction

National censuses are valuable sources of population data, but are too detailed their raw forms for most end-users. Geodemographic classifiers are derived data products that simplify these data by using statistical clustering to characterise local population-profiles in small areas. They are used widely in both the public and private sectors for population studies, sampling, marketing and service provision. OAC (Vickers and Rees, 2007) is an example of a geodemographic classifier that is based on the statistical clustering of 41 census variables (Table 2), categorising population by Output Area (OA; ~125 households; >200,000 in UK) into 7 super-groups and 21 groups[1] (Table 1). Considerable generalisation is involved as 7, 21 or 52[1] classes are produced from 41 continuous variables. Assigning OAs to single classes conceals the fact that most OAs are atypical cases sharing characteristics of many classes.

Local authorities are increasingly encouraged to use OAC to improve services (DCLG, 2009, p73). Leicestershire County Council (LCC) uses OAC for population profiling at different spatial scales (e.g. OAs, wards, postcode districts, postcode sectors) for service provision planning and for linking health, crime and other datasets.

Interactive visualization techniques for exploratory data analysis allow real-time selection and data filtering on demand in response to changing lines of enquiry. This "overview first, zoom, filter then details on demand" approach (Shneiderman, 1996) is the basis of many tools and environments (e.g. Dykes, 1998). LCC are interested in quantifying uncertainty in OAC classification, as this has implications for its utility, through visualization. OAC Explorer is designed to address the types of questions asked by LCC through a fast interactive visual interface epitomising this approach:

- Access to the census variables values used in the creation of OAC for all >200,000 OAs.
- Depicting uncertainty in the classification.
- Comparison with the typical cases of super-groups and groups.
- Comparison across different spatial scales.

---

[1] OAC's further 52 subgroups are not considered here.

Table 1. OAC super-groups and groups.

Super-group	Group
Blue Collar Communities	Terraced Blue Collar
	Younger Blue Collar
	Older Blue Collar
City Living	Transient Communities
	Settled in the City
Countryside	Village Life
	Agricultural
	Accessible Countryside
Prospering Suburbs	Prospering Younger Families
	Prospering Older Families
	Prospering Semis
	Thriving Suburbs
Constrained by Circumstances	Senior Communities
	Older Workers
	Public Housing
Typical Traits	Settled Households
	Least Divergent
	Young Families in Terraced Homes
	Aspiring Households
Multicultural	Asian Communities
	Afro-Caribbean Communities

2. Data

OAC is an open classifier whose creation methodology and census variables (Table 2) are freely available (SASI, 2009). Fuzzy membership is provided (SASI, 2009) as distances (similarity measures) from each of the super-group cluster centroids of each super-group for each OA. We normalise these distances to the maximum for each OA, inverting these to produce a measure of super-group (g) typicality (T_g):

$$T_g = 1 - \frac{d_g}{\max(d)} \tag{1}$$

(d_g = distance to cluster g, max(d) = distance to the furthest cluster and d_g=min(d) when closest cluster is used)

Scaled entropy (E; Fisher *et al*, 2004) is a measure of classification reliability, scaled between 0 (unreliable – where the OA is as similar to each super-group) and 1 (reliable, where the OA is typical of just one super-group):

$$E = \frac{-\sum_{i=7}^{i=1} d_i \times \log(d_i)}{-\log(1/7)} \tag{2}$$

(d_i =distance to super-group i)

Table 2. OAC's 41 census variables.

Age	0-4 **[v1]**, 5-14 **[v2]**, 25-44 **[v3]**, 45-64 **[v4]**, over 65 **[v5]**
Ethnicity	Indian/Pakistani/Bangladeshi **[v6]**, Black **[v7]**
Country of birth	Born outside UK **[v8]**
Population	Population density **[v9]**
Household	Separated/divorced **[v10]**
Household composition	Single non-pensioner **[v11]**, single pensioner **[v12]**, lone parent household **[v13]**, two adults no children **[v14]**, with non-dependents **[v15]**
Housing tenure	Public rent **[v16]**, private rent **[v17]**
Housing type	Terraced **[v18]**, detached housing **[v19]**, flats **[v20]**
Housing quality	No central heating **[v21]**, rooms per household **[v22]**, people per room **[v23]**
Education	HE qualification **[v24]**
Socio-economic	routine occupation **[v25]**
Travel	2+ car household **[v26]**, public transport to work **[v27]**, work from home **[v28]**
Health and care	Limiting Long Term Illness **[v29]**, provide unpaid care **[v30]**
Employment	Full-time students **[v31]**, unemployed **[v32]**, working part-time **[v33]**, economically inactive looking after family **[v34]**
Industrial sector	Agriculture/fishing **[v35]**, mining/quarrying/construction **[v36]**, manufacturing **[v37]**, hotel/catering **[v38]**, health/social work **[v39]**, financial intermediation **[v40]**, wholesale/retail trade employment **[v41]**

Figure 1. OAs in London, sized by population and coloured by super-group. Variable values displayed on parallel axes (below). The degree of typicality (T; equation 1) of an OA (indicated by the mouse pointer) to each super-group is shown in the barchart (top right).

3. Visualization techniques

Our new colour[2] scheme for OAC (Wood *et al*, 2010) uses 7 perceptually uniform hues for each super-group at constant lightness. In Table 1, we vary hue slightly for each group. It is not possible to

[2] Colour PDF version of this paper is available from http://gicentre.org/papers/slingsby_oacexplorer_2010.pdf

produce 21 distinguishable hues, so these colours are designed to reflect heterogeneity within super-groups rather than for class lookup. We have computed palettes for each of the 21 hues whereby lightness varies perceptually-linearly and is comparable across hues and use this in Figure 2 where lightness indicates typicality (T; equation 1; left) and entropy (E; equation 2; right).

Figure 1 contains a screenshot for the London area. The hierarchical rectangular cartogram (spatial treemap; Wood and Dykes, 2008) sizes OAs by their population within the postcode hierarchy. We use this representation over standard cartographic projections because it normalises by population density (population-dense areas have a greater map area so the detail is more easily resolvable), is space-filling (makes efficient use of space – important as there are >200,000 elements to map), retains a high degree of spatial structure (Slingsby *et al*, in press) and is already in use within LCC. The use of postcode geography places OAs in a widely-recognised spatial frame of reference allowing selection at different spatial levels − all OAs in the WC postcode area in this case (the strong blue colour indicates WC is selected). Each axis in the parallel coordinate display corresponds to a census variable (Table 2). Values for the OAs in WC are shown as dots with typical 'Multicultural' and 'Countryside' profiles as thicker lines (selected on the right). The thin red line in the parallel plot shows the census profile of the selected OA (indicated by the mouse pointer) and allows its relationship to other super-groups ('Countryside' in this case) to be seen. This interface allows national (typical cases for super-groups), regional (selected set of OAs; those in WC) and individual OAs to be selected and compared. Uncertainty in OA classification can be explored through colour lightness that indicates typicality (in Figure 2, left) and entropy (in Figure 2, right). The barchart for the selected OA shows proximity to all super-group centroids concurrently. Here, it is quite typical of 'Multicultural', but is also similar to other super-groups – particularly 'City Living' – and so has low entropy. This is interesting as many inner London OAs fall into these two classes. 'Typical Traits' show low entropies − the OAs of which form a tight cluster.

Figure 2. Screenshot excerpts that use *lightness* to show the typicality (T) of OAs to their closest super-group (left) and the scaled entropy (E) of classification (right).

Figure 3 shows the Leicester area. A 'Multicultural' core is distinguishable from the 'Prospering suburbs' and 'Countryside'. Box-plots of the 25th, 50th and 75th percentiles indicate variation in the census variables for 'Multicultural' OAs in the visible screen area. The middle two quartiles of v7, v20, v37 and v18 (Table 2) are significantly above the typical case for this super-group. Variables are sorted on the selected OA here enabling variables with atypical values to be identified.

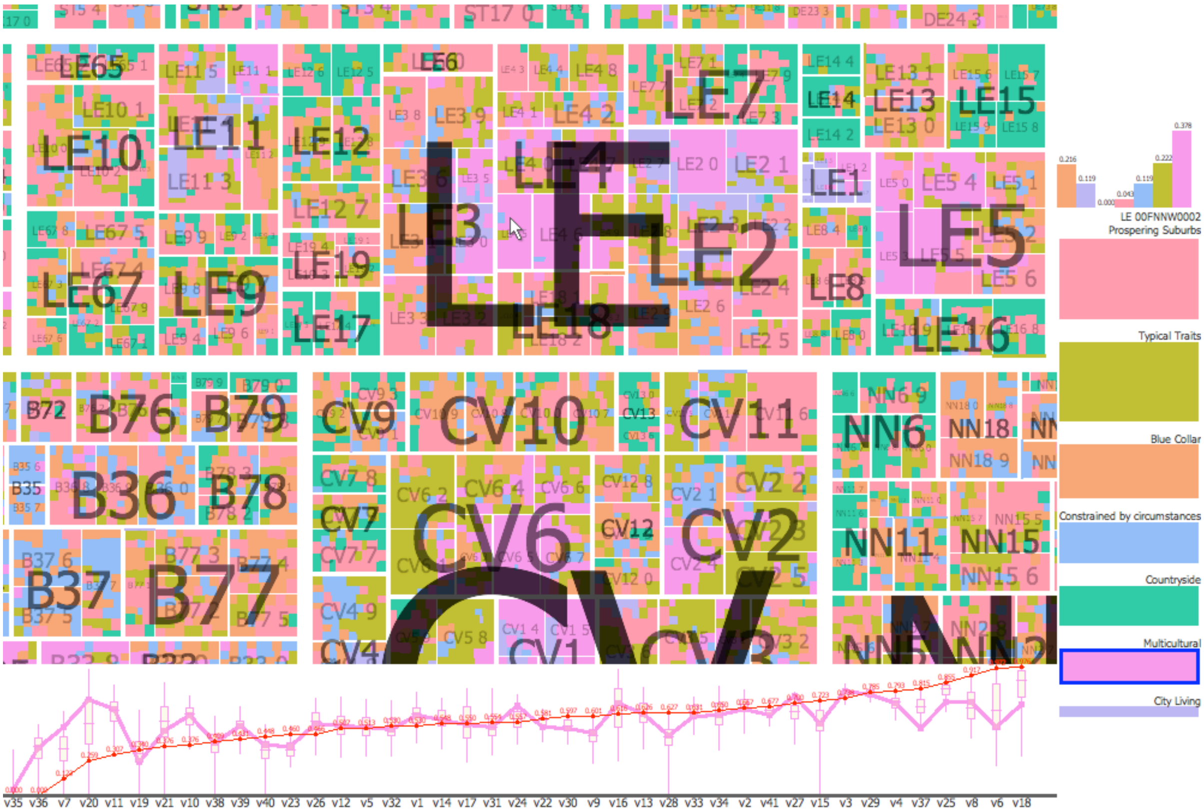

Figure 3. Screenshot for Leicester with Coventry to the south showing typical 'Multicultural' with box-plots summarising variation in the current field of view. The red line is the OA indicated with the mouse pointer, and variables are sorted by their magnitude on this line.

Figure 4 focuses on LE7. Variable values are shown as differences from the typical 'Prospering Suburbs' case. As expected, the 'Prospering Semis' subgroup deviates little from this (most notably, v19) but values for proportion of Indian/Pakistani/Bangladeshi (v6) vary significantly. This is very different to the national distribution (see Figure 5).

5. Conclusion and ongoing work

Geodeomographic classifiers help local authorities and others understand local populations through generalisation by categorising them into meaningful groups. We show that desktop computers are now powerful enough to provide rapid and interactive access to OAC alongside its original census variable values through which the effects of this generalization can be explored and evaluated. This visualization allows the reliability and uncertainty in classification to be determined at different spatial granularities and places. Ongoing work with practitioners in local government is developing and evaluating techniques for using visualization to meet their specific needs.

6. Acknowledgements

Thanks to Dan Vickers for providing the original OAC input variables. Aidan is funded by the Willis Research Network and Robert was supported by ESRC UPTAP User Fellowship RES-163-27-0017. The National Statistics Postcode Directory was obtained through UKBorders/Edina and other OAC-related data were from http://www.sasi.group.shef.ac.uk/area_classification/.

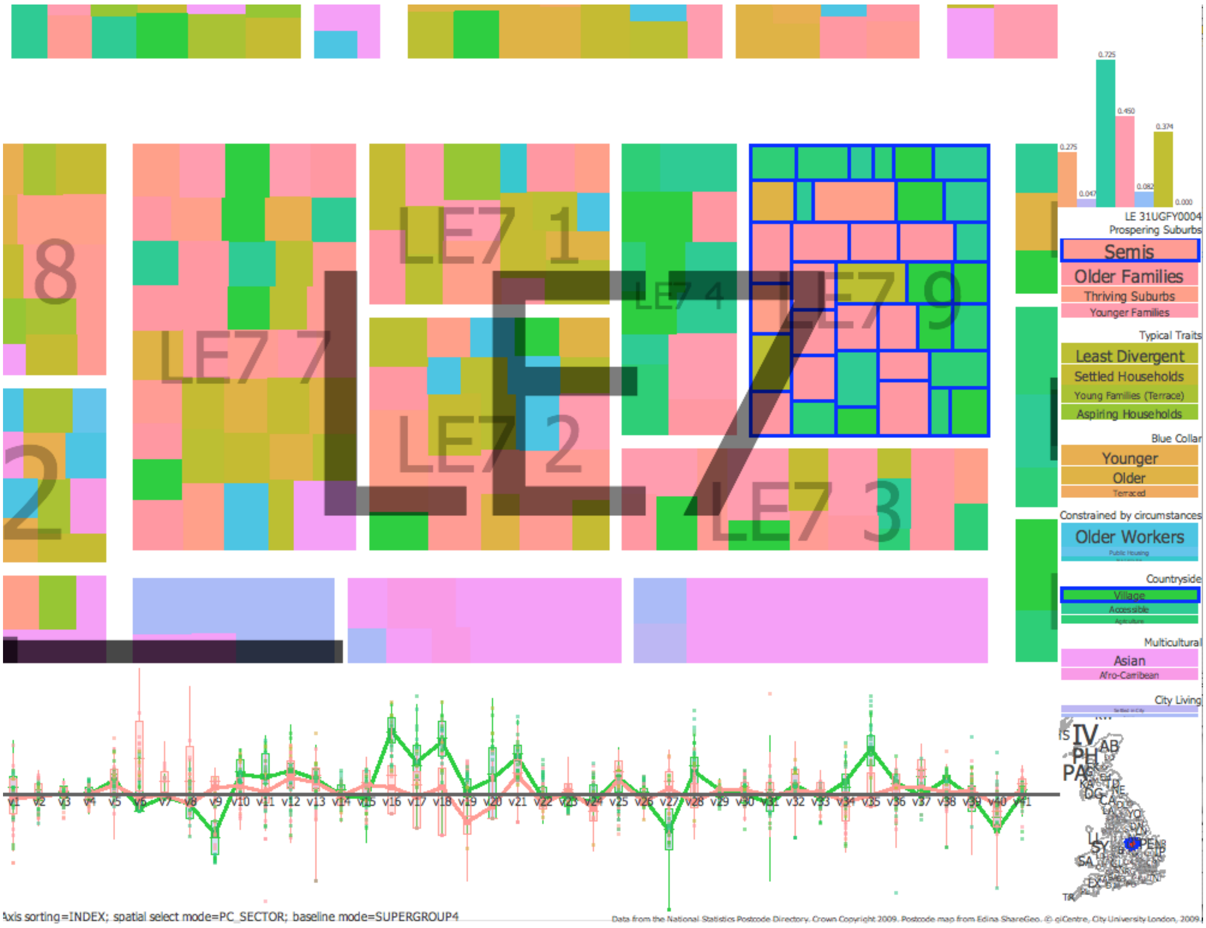

Figure 4. Screenshot focusing on LE7 with OAs of LE7 9 selected (its OAs are outlined in blue). The parallel plot shows deviation from 'Prospering Suburbs' super-group.

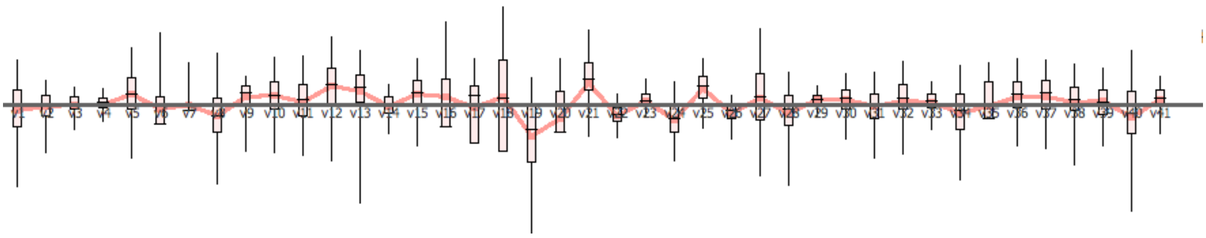

Figure 5. Box plot showing national variation in 'Prospering Semis'.

References

DCLG (UK Department of Communities and Local Governement) (2009). Supporting local information and research: Understanding demand and improving capacity. *http://www.communities.gov.uk/publications/communities/supportinglocalresearch*

Dykes, J., (1998), Cartographic Visualization: Exploratory Spatial Data Analysis with Local Indicators of Spatial Association using TCL/TK and CDV, *The Statistician*, **47**(3), 485-497.

Fisher, P., Wood, J. & Cheng, T., (2004). Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, **29**, 106-128.

SASI (Social and Spatial Inequalities Group, University of Sheffield) (2009). *The National Classification of Census Output Areas. http://www.sasi.group.shef.ac.uk/area_classification/*

Shneiderman, B., (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*. IEEE Computer Society, p. 336 http://portal.acm.org/citation.cfm?id=834354

Slingsby, A, Dykes, J. & Wood, J., (in press), Hierarchical Rectangular Cartogram of OAC by GB Unit Postcode. *Journal of Maps*.

Vickers, D. & Rees, P., (2007). Introducing the National Classification of Census Output Areas. *Population Trends*, **125**, 380-403.

Wood, J. & Dykes, J., (2008). Spatially Ordered Treemaps. *IEEE Transactions on Visualization and Computer Graphics*, **14**(6), 1348-1355.

Wood, J, Slingsby, A. & Dykes, J., (2010), *Layout and Colour Transformations for Visualising OAC Data*, GIS Research UK 2010 *[this volume].*

Biographies

Dr. Aidan Slingsby *is a Willis Research Fellow at the giCentre, City University London with research interests in designing, implementing and using geovisualization techniques for assessing data quality and variability and for visual data analysis.*

Dr. Jo Wood *is a Reader in geographic information at the giCentre, City University London with research interests in geovisualization, terrain modelling and object oriented programming for spatial sciences.*

Dr. Jason Dykes *is a Senior Lecturer at the giCentre, City University London undertaking applied and theoretical research in, around and between information visualization, interactive analytical cartography and human-centred design.*

Robert Radburn *is Research and Intelligence Team Leader at Leicestershire County Council and held an ESRC UPTAP research fellowship at the giCentre, City University London to develop capacity for visual exploratory analysis in local government.*