Rectangular Hierarchical Cartograms for Socio-Economic Data

AIDAN SLINGSBY, JASON DYKES and JO WOOD

giCentre, Department of Information Science, City University, Northampton Square, London, EC1V 0HB, UK; a.slingsby@soi.city.ac.uk.

Abstract

We present rectangular hierarchical cartograms for mapping socio-economic data. These density-normalising cartograms size spatial units by population, increasing the ease with which data for densely populated areas can be visually resolved compared to more conventional cartographic projections. Their hierarchical nature enables the study of spatial granularity in spatial hierarchies, hierarchical categorical data and multivariate data through false hierarchies. They are space-filling representations that make efficient use of space and their rectangular nature (which aims to be as square as possible) improves the ability to compare the sizes (therefore population) of geographical units.

We demonstrate these cartograms by mapping the Office for National Statistics Output Area Classification (OAC) by unit postcode (1.52 million in Great Britain) through the postcode hierarchy, using these to explore spatial variation. We provide rich and detailed spatial summaries of socio-economic characteristics of population as types of treemap, exploring the effects of reconfiguring them to study spatial and non-spatial aspects of the OAC classification.

(Received 31st July 2009; Revised 15th March 2010; Accepted 18th March 2010)

ISSN 1744-5647
doi:10.4113/jom.2010.1090

1. Introduction

The geographical distribution of population classified according to the National Classification of Census Output Areas (OAC) is shown in Figure 1. OAC classifies population by Output Area (OA; spatial units used for UK census statistical output) according to social and economic population characteristics (Vickers and Rees, 2007). The image illustrates the difficulties associated with conveying social data on maps for areas whose population density vary markedly; i.e. it is difficult to resolve the OAC detail for densely-populated areas. OAs – in common with most other spatial units used for social data – vary in size depending on population. In conventional map projections, the size of any symbol has an inverse relationship with population; i.e. less map space (information space) is available for conveying information about large and dense population centres than for areas of low population density.

Population density-normalising cartograms are well established means dealing with this problem (Tobler, 2004). Spatial units are sized by population. This has the effect of enlarging densely populated areas at the expense of more sparsely populated areas, but it introduces spatial distortion into the cartography. Some cartograms preserve geographical shapes, resulting in non-tessellating and sometimes overlapping arrangements. Some maintain contiguity, resulting in necessary distortion of the map area shapes as is the case with Gastner Cartograms (Gastner and Newman, 2004). Others fix the shape of areas as rectangles (Raisz, 1934; Florisson et al., 2005; van Kreveld and Speckmann, 2007; Wood and Dykes, 2008), circles (Dorling, 1996) or hexagons (Shaw et al., 2008) for greater ease of size comparison, but contiguity between areas is not guaranteed.

Spatial units for mapping socio-economic data often nest in spatial hierarchies. Mapping data in this spatial hierarchy can give insights into the spatial granularity at which phenomena operate. In addition, many classification schemes are hierarchical, with classes that nest within less discriminating classes. This is true of OAC which contains a hierarchy of super-groups (used in Figure 1), groups and sub-groups as illustrated in Figure 2.

We present rectangular hierarchical cartograms as density-normalising cartograms for mapping national socio-economic data. We demonstrate this by presenting a detailed map of 1.52 million GB unit postcodes in their spatial hierarchy, sized by population and coloured by the OAC category that most closely characterises the population.

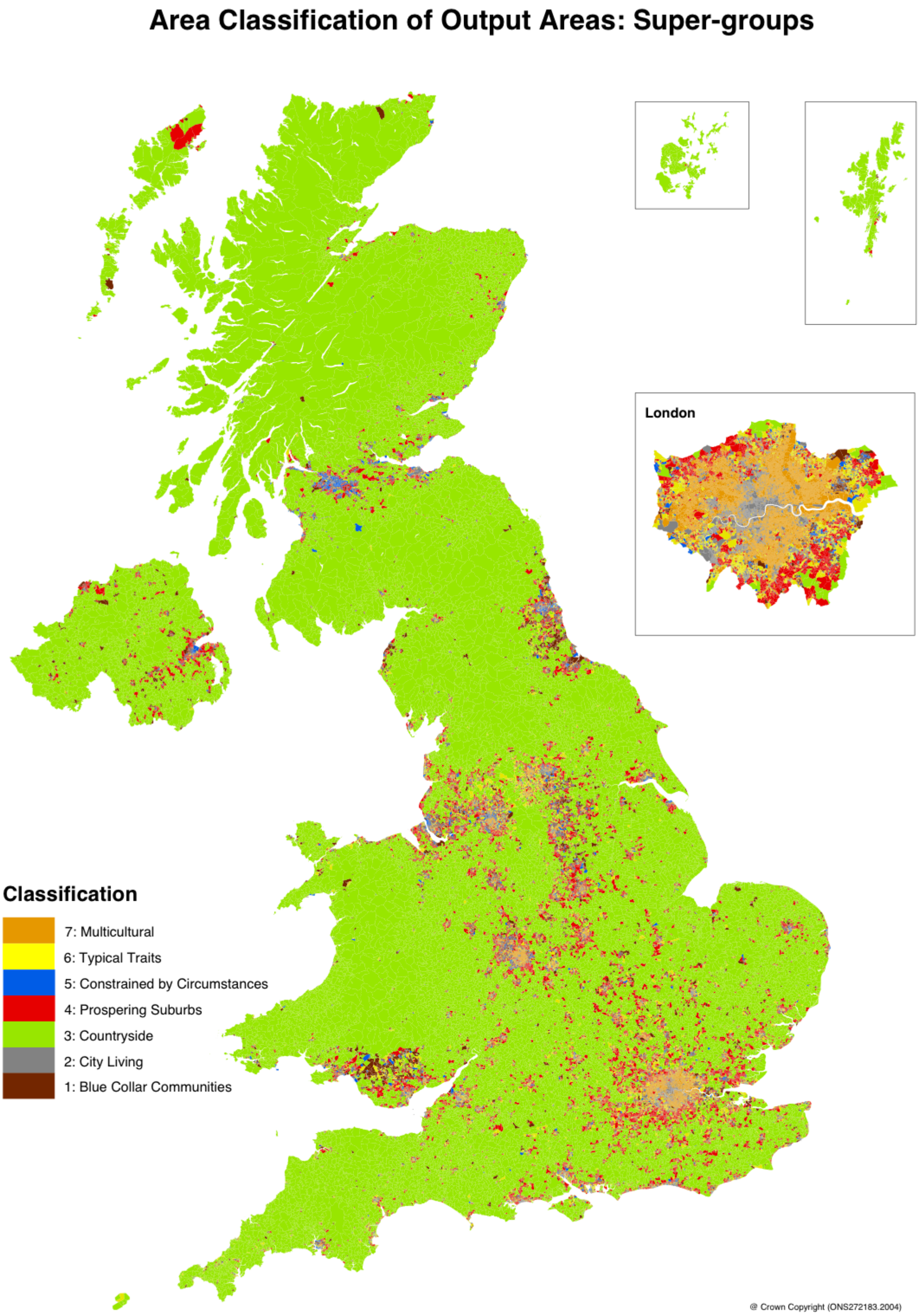

Figure 1. A map of Output Areas (OAs) coloured by their OAC category. It is difficult to resolve the detail in the most densely-populated areas. *Source:* *http://www.sasi.group.shef.ac.uk/area_classification/Cluster_Maps/super_grroups.pdf*

2. Methods

Socio-economic data are usually supplied in discrete spatial units that are often related to population density. 'Output Areas (OAs)' – containing an average of about 300 people in England and Wales and 125 people in Scotland – are the finest spatial units in which UK census statistics are disseminated. Postcodes are a hierarchical system for locating delivery points for the purpose of delivering post (Raper et al., 1992), however their familiarity to most individuals, their ubiquity and their relative high precision (typically 15 delivery points per postcode unit not associated with a large business) makes the postcode a useful base geography for many types of socio-economic data. UK postcodes contain a four-level hierarchy of area, district, sector and unit. For the postcode EC1V 0HB:

- EC: postcode area (125 in UK)
- EC1V: postcode district (3064 in UK)
- EC1V 0: postcode sector (11598 in UK)
- EC1V 0HB: postcode unit (1.78 million in UK)

OA boundaries were designed to incorporate postcode geography where possible (Martin, 2001) and, as such, can have census data mapped to them. In terms of spatial resolution, OAs lie between unit postcodes and postcode sectors. We use postcodes as our base geography because they are hierarchical, because their labels are concise and because they are widely recognised in the UK.

The National Classification of Census Output Areas classifies the population of OAs by social and economic characteristics based on 41 census variables (Vickers and Rees, 2007). OAC categories are hierarchical, containing 7 super-groups, 21 groups and 52 sub-groups as listed in Figure 2. Super-groups and groups have descriptive labels that characterise the statistical cluster (see http://www.sasi.group.shef.ac.uk/area_classification/ for more details). Discussion, evaluation and critique of geodemographics is outside the scope of this paper, but are well-reported elsewhere (e.g. Vickers and Rees, 2009; Singleton and Longley, 2009). However, we have found in other work that the cartographic framework we present offers considerable scope for visualising and exploring some of the uncertainties associated with such classification methods (Wood et al., 2010), particularly when combined with interactive exploratory interfaces (Slingsby et al., 2010).

All the figures produced for this paper use a common colour scheme. Super-groups have been allocated the closest hue from ColorBrewer's 'Set1' (Brewer et al., 2003) to those in Figure 1 from http://www.sasi.group.shef.ac.uk/area_classification/. The OAC group colours are produced by varying the lightness of these hues. These are further permuted (to a smaller degree) to obtain sub-group colours. Super-groups are easily distinguishable by their different hues. Although it is difficult to identify groups and almost impossible to identify sub-groups from the legend using colour alone, differences in lightness indicates heterogeneity within super-group. More sophisticated colour palettes for OAC that are more perceptually consistent have been produced by Wood et al. (2010).

Super-group	Group	Sub-groups
1 Blue Collar Communities	1A Terraced Blue Collar	1A1, 1A2, 1A3
	1B Younger Blue Collar	1B1, 1B2
	1C Older Blue Collar	1C1, 1C2, 1C3
2 City Living	2A Transient Communities	2A1, 2A2
	2B Settled in the City	2B1, 2B2
3 Countryside	3A Village Life	3A1, 3A2
	3B Agricultural	3B1, 3B2
	3C Accessible Countryside	3C1, 3C2
4 Prospering Suburbs	4A Prospering Younger Families	4A1, 4A2
	4B Prospering Older Families	4B1, 4B2, 4B3, 4B4
	4C Prospering Semis	4C1, 4C2, 4C3
	4D Thriving Suburbs	4D1, 4D2
5 Constrained by Circumstances	5A Senior Communities	5A1, 5A2
	5B Older Workers	5B1, 5B2, 5B3, 5B4
	5C Public Housing	5C1, 5C2, 5C3
6 Typical Traits	6A Settled Households	6A1, 6A1
	6B Least Divergent	6B1, 6B2, 6B3
	6C Young Families in Terraced Homes	6C1, 6C2
	6D Aspiring Households	6D1, 6D2
7 Multicultural	7A Asian Communities	7A1, 7A2, 7A3
	7B Afro-Caribbean Communities	7B1, 7B2

Figure 2. The hierarchy of OAC categories. Sub-groups do not have descriptive labels and are not distinguished by colour in this graphic (though we have done so in Figure 5). *Data source: http://www.sasi.group.shef.ac.uk/area_classification/Cluster_Names/PDF/cluster%20names.pdf*

OAC maps provide rich spatial summaries of multiple socio-economic characteristics of population. Several aspects of rectangular hierarchical cartograms make them well suited to this type of cartography. They are *space-filling*, making efficient use of information space – particularly useful where there are hundreds of thousands of elements to map. They use *consistently shaped elements* that are as square as they can be given their geographical arrangement. This enables effective size comparison which is particularly important where size is related to a property of the data (e.g. population). They are also hierarchical, enabling both spatial and categorical hierarchies to be depicted.

The algorithm to produce rectangular cartograms (Wood and Dykes, 2008) is fast – typically taking no more than a few seconds to produce cartograms from thousands of spatial units – enabling them to be generated on-the-fly in interactive applications (Slingsby et al., 2009). Gastner and Newman (2004)'s software takes "< 100 seconds" to produce a Gastner Cartogram of ∼2000 New York ZIP codes whereas we can produce rectangular cartograms of these data in a couple of seconds. Although there are certainly differences in the computer hardware used, the increased speed is largely due to the significantly fewer vertices of rectangles over polygons and the computationally-simpler algorithm. Rectangular cartograms of 1.52 million postcodes do however take significantly longer to produce (about four minutes including reading in the data and displaying) precluding on-the-fly generation in interactive applications. Where fewer spatial units are used or the layout is precomputed, interactive applications can be built based around these (Slingsby et al., 2009; 2010);

Figure 3 shows a rectangular hierarchical cartogram of all 1.52 million GB postcodes (England, Scotland and Wales) coloured by OAC sub-group. Figure 5 can be used as a legend for many of the figures in this paper and both figures form the basis of the map that accompanies this paper. Colours emphasise super-groups, but allow some distinction to be made between groups and sub-groups. Although there are too many sub-groups (52) to distinguish by colour alone, small differences in colour indicate that more than one sub-group is present. Postcode units are sized by the number of residential delivery points as a proxy for population, since head counts per unit postcode are only available for England and Wales. Appropriately, this results in business postcodes having zero size and thus not appearing in the map (as a result, the accompanying map contains 1.46 million of the 1.52 million GB postcodes). The most obvious effect of using a population density-normalising cartogram is that a much smaller proportion of the map area is identified as 'Countryside' (green) because this group relates to a small numbers of large areas with low population density. The spatial arrangement of postcode units retains the spatial structure of the national pattern – e.g. 'Countryside' is found between major conurbations. Spatial autocorrelaton (spatial homogeneity) is high for 'Multicultural' in inner London postcode areas (e.g. E, N, SE) and central Birmingham (B). Spatial autocorrelation is also high for 'Constrained by Circumstances' in Central Glasgow (G), Edinburgh (EH) and Newcastle (NE) all of which have 'Prospering suburbs' at their peripheries. The core-periphery structure of large cities such as Birmingham (B) can be distinguished. It is also interesting to study the spatial granularity of OAC classes by postcode area, though this is dependent on how space is discretised. For example, inner London postcode areas have high homogeneity at the *postcode area* level – expected as these are *parts* of a large city – whereas variation is at the *postcode sector* level in Glasgow (see square areas of homogeneity within the G area). Other areas such as Tunbridge Wells (TN) have a finer spatial resolution of variation.

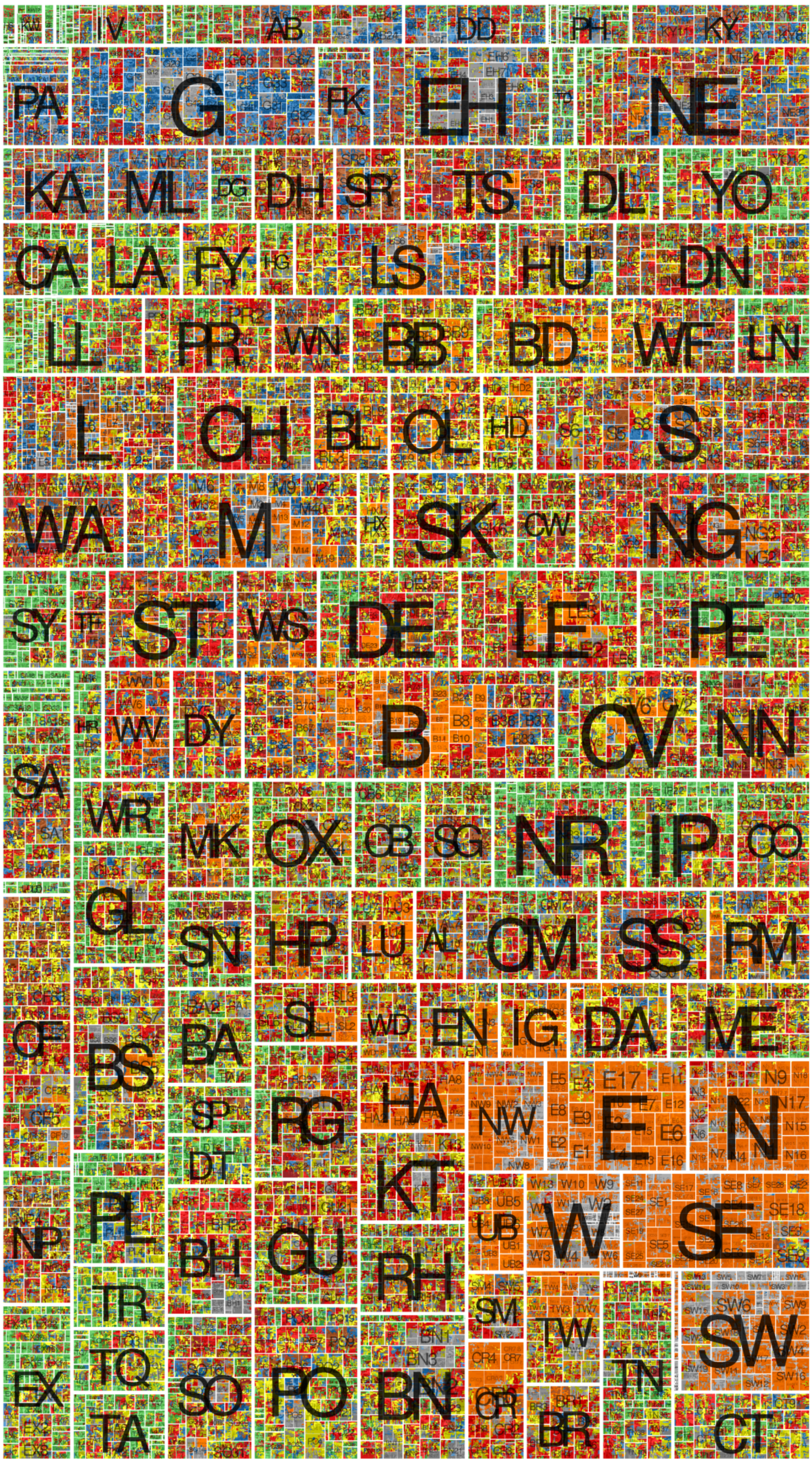

Figure 3. A hierarchical rectangular cartogram of 1.52 million GB postcodes (England, Scotland and Wales) coloured by OAC sub-group (see Figure 5) in a hierarchy of postcode area, postcode district, postcode sector and unit postcode. Each element is sized by the number of residential postal delivery points. See Figure 2 for colours. HiVE: `sHier(/,$pArea,$pDistrict,$pSector,$pUnit); sSize(/,$pop,$pop,$pop,$pop); sLayout(/,SP,SP,SP,SP).`

Pockets of population classified as 'Countryside' exist at postcode district and sector levels – in peripheral parts of Norwich (NR) and Ipswich (IP), northern parts in Swansea (SA) and eastern parts in Peterborough (PE).

The space filling-nature of this cartogram results in some considerable displacement of postcode areas from their original geographical positions. This is particularly true in the north as the landmass narrows and the population density becomes sparser. Wood and Dykes (2008) suggest the use of statistics, displacement vectors and colour to identify where such displacement exists and by how much. In Figure 4 we attempt to reduce such displacement by inserting 'dummy nodes' (white space) into the cartogram to try and maintain the shape of the landmass to assist user orientation. Whilst this reduces positional displacement with respect to the original geographical positions and assists in identification of the various parts of the country, it does result in a large amount of empty space – approximately a quarter of the information space. Cartographic design involves balancing such considerations. Our map emphasises the efficient use of information space, particularly important when producing such dense and information-rich graphics comprising over a million spatial units. We consider that the high degree of spatial structure preserved and the concise and familiar postcode labels provide an adequate frame of reference to users familiar with UK geography. Such a judgement is dependent on the map users and the nature of the data being depicted. In an interactive context, we have found that smooth transitions between different projections of data are a useful means to orient the map user (Slingsby et al., 2009).

Although our cartograms have concentrated on spatial hierarchies, other non-spatial hierarchies can be constructed. OAC itself consists of a three-level hierarchy (super-groups, groups, sub-groups). In Figure 5, *we do not* have spatial units in the hierarchy and elements *are not* arranged spatially. Instead, the hierarchy comprises entirely of the OAC class hierarchy (super-group, group and sub-group) and elements are arranged by size from top-left to bottom-right showing that the most populous super-group is 'Propersing Suburbs' which has approximately three times the population of 'City Living'.

Figure 5 is a squarified treemap (Bruls et al., 1999) in which elements have been arranged by size from the top left to the bottom right. Treemaps display hierarchical data by recursively subdividing rectangular space into rectangles using size, colour and order to convey characteristics of the data, with hierarchies represented using containment (Shneiderman, 1992). The treemap in Figure 5 shows the OAC hierarchy (Figure 2), in which categories are sized by population and ordered such that the categories with the most population ('Prospering suburbs'; 'Semis' within this; 4C2 within this) are in the top left and the categories with the smallest population ('City Living') are in the bottom right.

HiVE (Slingsby et al., 2009) is a language for describing hierarchical graphics. The hierarchy in the cartogram in Figure 3 can be expressed as `sHier(/,$pArea,$pDist-`

Figure 4. As Figure 3, but with “dummy nodes” inserted to help preserve the shape of the landmass.

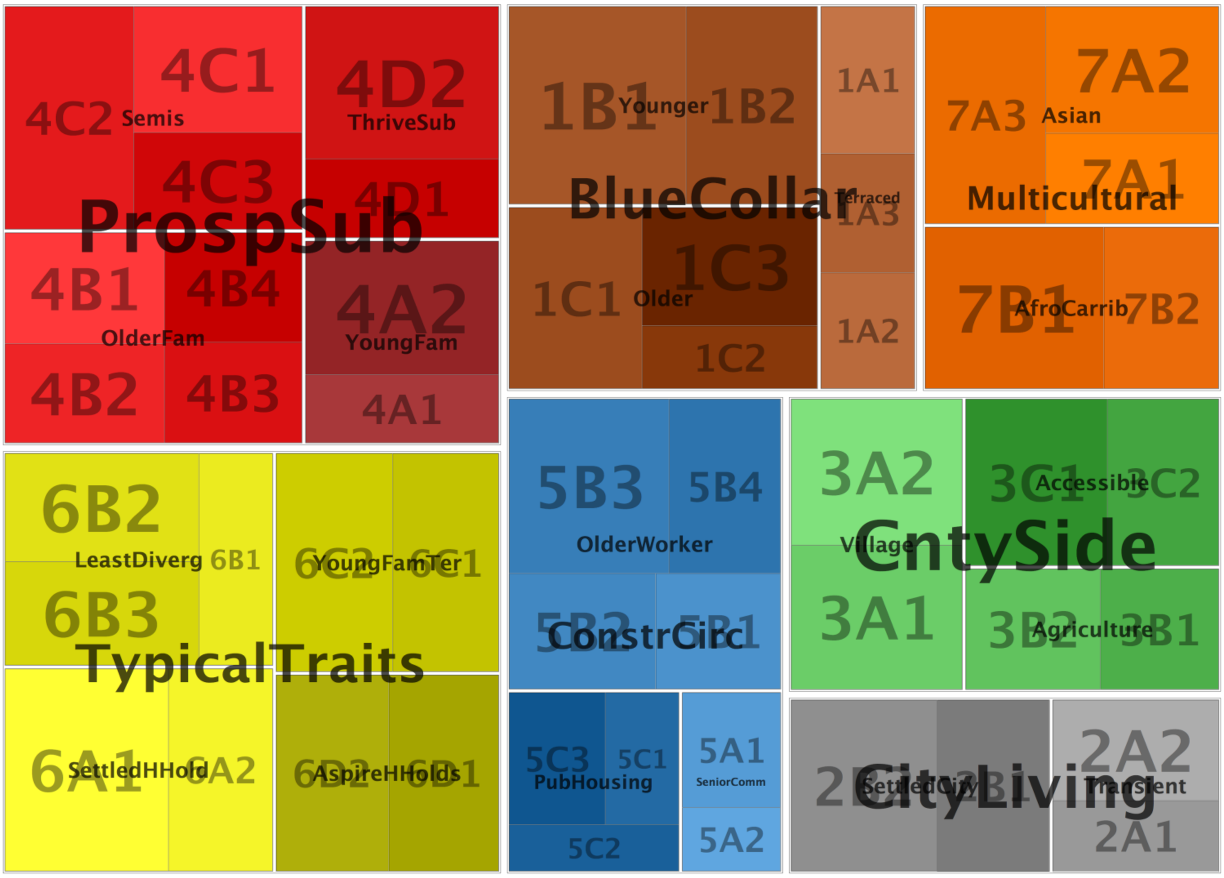

Figure 5. The three hierarchical levels of OAC category (Figure 2), sized and ordered (top left to bottom right) by population. HiVE: `sHier(/,$oacSuper,$oacGroup,$oacSub); sSize(/,$pop,$pop,$pop,$pop); sLayout(/,SQ,SQ,SQ,SQ)`

`rict,$pSector,$pUnit)` – where the variables (preceded with a `$`) refer to postcode area, district, sector and unit, respectively. `sSize(/,$pop,$pop,$pop,$pop)` indicates that element in all hierarchical levels are sized by population (`$pop`) and `sLayout(/,SP,SP,SP,SP)` indicates that each level is spatially arranged (`SP`). The hierarchy in the treemap in Figure 5 is `sHier(/,$oacSuper,$oacGroup,$oacSub)` – where the variables are OAC super-group, group and sub-group, respectively. `sSize(/,$pop,$pop,$pop,$pop)` indicates that size is related to population as before and `sLayout(/,SQ,SQ,SQ,SQ)` indicates that the 'squarified' layout (Bruls et al., 1999) has been applied, ordering elements by their size (population in this case) from the top left to the bottom right. Our hierarchical rectangular cartograms are treemaps in which elements are arranged spatially (Wood and Dykes, 2008).

Placing both spatial and non-spatial variables in the hierarchy gives interesting opportunities for studying granularity in both the spatial data and in the classification. Figure 6 shows the OAC breakdown of population within each postcode area through each of its three hierarchical levels. These themselves are arranged spatially; for example in the W postcode area, 'Multicultural' is *on average* in the west and 'City Living' is *on average* in the East. The problem with this is that it is impossible to see whether they are spatially concentrated or dispersed and where average geographical positions are similar, the relative positions of the corresponding rectangles may be rather arbitrary. This graphic does, however, allow us to detect relative proportions of OAC category by postcode more easily than in Figure 1 because they are spatially consolidated within the postcode area, rather than distributed across it. Since the hierarchy contains the full OAC hierarchy using containment, variation within super-groups and groups is apparent. For example there are more types of the 'City Living' OAC super-group in EH (Edinburgh) than in SW (in London).

Reordering the variables in the hierarchy helps draw out different patterns (Slingsby et al., 2009). In Figure 7, we have moved the OAC super-group variable to the base of the hierarchy. This spatially consolidates information about each super-group enabling us to compare variation by super-group more effectively – in this case, spatial variation by super-group. For each super-group, the population living in each postcode area is shown, broken down by group and sub-group. SE (part of London) is large in the 'multicultural' super-group, but its small size or absence in other super-groups indicates that a high proportion of the OAs in SE are classified as 'Multicultural'. However, Birmingham (B) is contained in all the OAC super-group areas of the treemap, indicating that the population in this particular postcode area is socio-economically diverse. Elements are spatially arranged, so 'Multicultural' is on average in the southeast and 'Countryside' tends to be northern, however, these groups are not necessarily spatially concentrated; i.e. southern postcode areas such as EX (Exeter) and CT (Canterbury) contain much 'Countryside'. In Figure 8, colour is used to represent absolute geographical position: green for the southeast, red for the southwest, purple for the northwest and blue for the northeast, based on a transformation of the perceptually-uniform CIELa*b* colourspace for the UK National Grid (Wood and Dykes, 2008). The wide range of colours for 'Countryside' indicates that although its average position is more northern than other super-groups, it contains populations from all parts of the country. 'City Living' and 'Multicultural' are well represented in the southeast (green) and poorly represented in the south west (red/yellow), whereas 'Blue Collar' and 'Constrained by Circumstances' are well represented in the north of the UK (purple and blue). This illustrates that colour can be an effective means to show absolute spatial position in graphics which do not do this through position.

3. Conclusions

It is difficult to visually resolve population data on conventional maps for areas with high variation in population density. Cartograms that normalise area by population

Figure 6. A cartogram of the three-tier OAC hierarchy by postcode area. Each element is spatially arranged. HiVE: `sHier(/,$pArea,$oacSuper,$oacGroup,$oacSub); sSize(/,$pop,$pop,$pop,$pop); sLayout(/,SP,SP,SP,SP)`

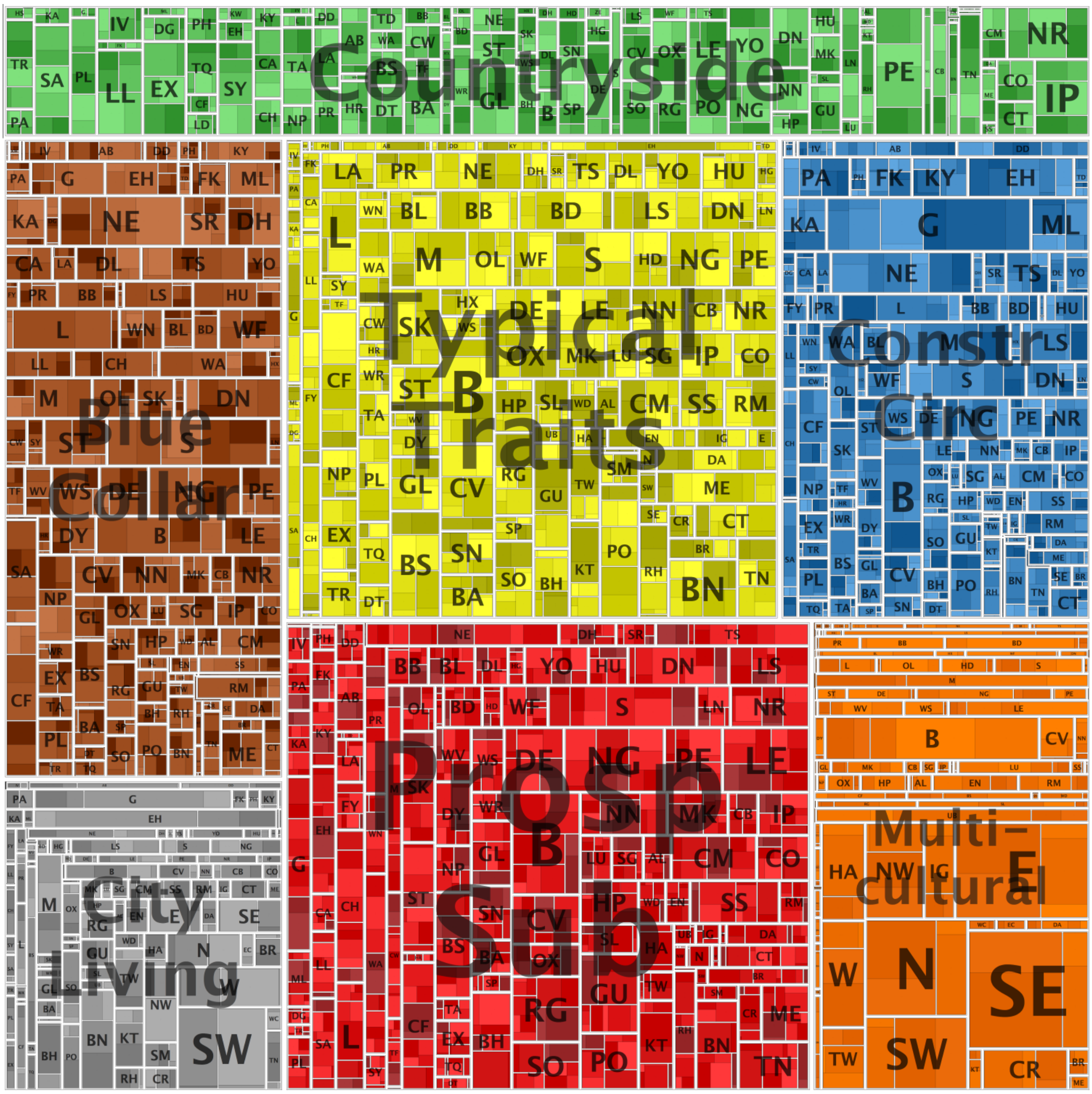

Figure 7. A cartogram showing the population of postcode areas by OAC super-group. Each postcode area contains the OAC group and sub-group and coloured by sub-group. All the elements are spatially arranged – the average location of 'Countryside' is north of the other super-groups' average locations and the postcode areas within each super-group are arranged spatially resulting in 7 rectangular maps of the UK. HiVE: `sHier(/,$oacSuper,$pArea,$oacGroup,$oacSub); sSize(/,$pop,$pop,$pop,$pop); sLayout(/,SP,SP,SP,SP)`

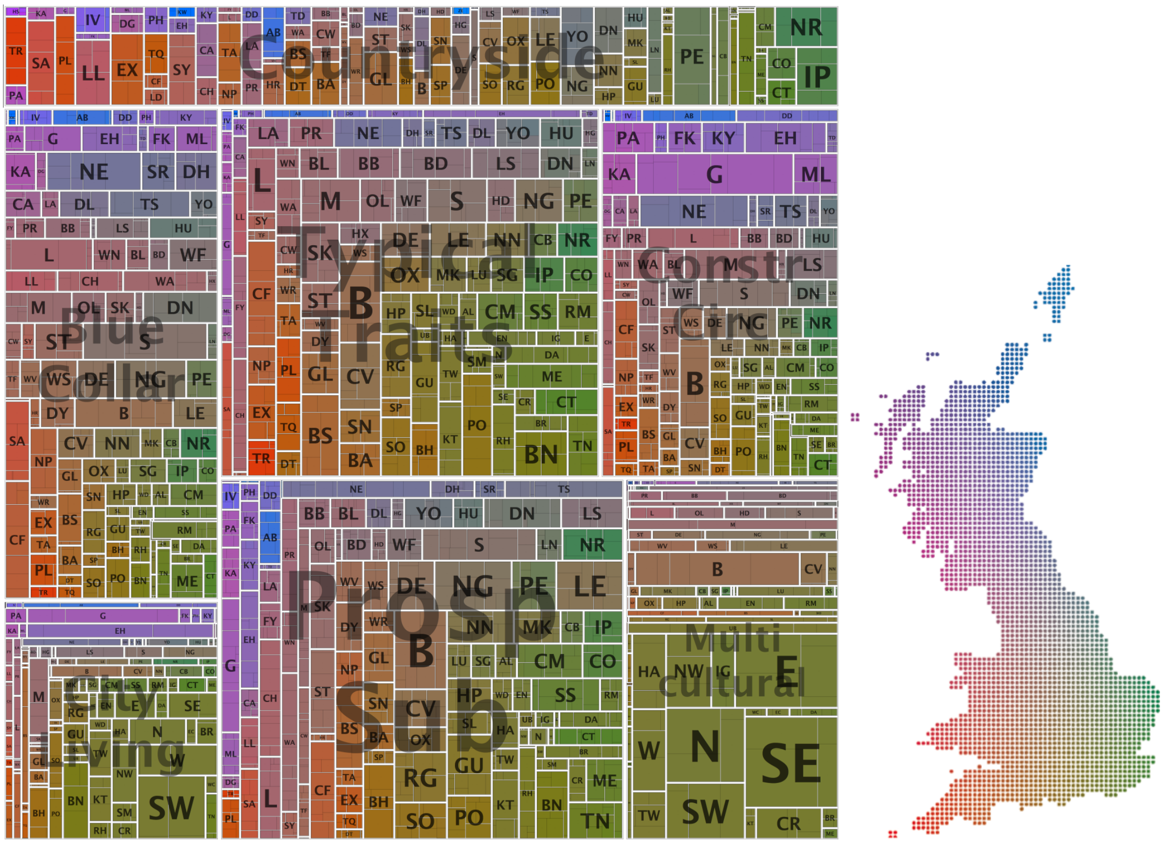

Figure 8. As figure 7, but coloured spatially according to the bivariate colours scheme shown on the right based on a transformation of the CIELa*b* colourspace for the UK National Grid (Wood and Dykes, 2008).

density can help address this by increasing the mapping space available for conveying data for areas of dense population. We show that the depiction of hierarchies is beneficial for social data: spatial hierarchies for studying space granularity in data variation and categorical hierarchies for hierarchical classification schemes. Interestingly we mix geography and attributes in our hierarchies and show how arranging geography and attributes in different ways results in maps that have different geographies and a variety of different and potentially useful characteristics.

We demonstrate the use of squarified rectangular hierarchical cartograms for showing the spatial distribution and granularity of population classified according to the National Classification of Census Output Areas (OAC). Postcode geography is used due to its concise labels, widespread familiarity and its hierarchical nature. We describe this variation using cartograms and treemaps, demonstrating the advantages offered by their properties. The fact that they are density normalising, allows variation in the data to be visually resolved irrespective of population density, by enlarging areas of high population density. The squarified and space-filling nature of the cartograms makes size comparison easier and makes efficient use of the information space. We consider efficient use of information space to be an important property of cartographic techniques, particularly when depicting rich and spatially-dense data. The rectangular shapes nest – as such, they can be used to show hierarchical data common in the social sciences. We show the effects of changing and reordering the hierarchy of variables to draw out different patterns, different arrangements, the use of colour to encode absolute space and the insertion of 'dummy nodes' to help preserve landmass shape to aid in interpretation. We encourage readers to experiment with these techniques for presenting and analysing social data, using the free *treeMappa* software (http://www.treemappa.com/).

Software

TreeMappa (http://www.treemappa.com/) was used to create the treemaps and hierarchical cartograms. The high quality PDF output of the map that accompanies this paper was produced programatically using Processing (http://www.processing.org/).

References

BREWER, C., HATCHARD, G. and HARROWER, M. (2003) ColorBrewer in Print: A Catalog of Color Schemes for Maps, Cartography and Geographic Information Science, 30, 5–32, doi:10.1559/152304003100010929.

BRULS, M., HUIZING, K. and VAN WIJK, J. J. (1999) Squarified treemaps, In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization, Springer Computer Science, pp. 33–42.

DORLING, D. (1996) Area Cartograms: Their Use and Creation, Concepts and Techniques in Modern Cartography (CATMOG), University of East Anglia.

FLORISSON, S., VAN KREVELD, M. and SPECKMANN, B. (2005) Rectangular cartograms: construction & animation, In Proceedings of the 21st annual symposium on Computational geometry, ACM, Pisa, Italy, pp. 372–373, doi:10.1145/1064092.1064152.

GASTNER, M. T. and NEWMAN, M. E. J. (2004) Diffusion-based method for producing density-equalizing maps, Proceedings of the National Academy of Sciences of USA, 101, 7499–7504, doi:10.1073/pnas.0400280101.

MARTIN, D. (2001) Geography for the 2001 Census in England and Wales, Office for National Statistics, London.
URL http://eprints.soton.ac.uk/16180/

RAISZ, E. (1934) The Rectangular Statistical Cartogram, Geographical Review, 24, 292–296, doi:0.2307/208794.

RAPER, J. F., RHIND, D. W. and SHEPHERD, J. W. (1992) Postcodes: The New Geography, Longman Science and Technology.

SHAW, M., SMITH, G. D., THOMAS, B. and DORLING, D. (2008) The Grim Reaper's Road Map: An Atlas of Mortality in Britain, Policy Press.

SHNEIDERMAN, B. (1992) Tree visualization with tree-maps: 2D space-filling approach, ACMTransactions on Graphics, 11, 92–99, doi:10.1145/102377.115768.

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 in Regional Science, 88, 643–666, doi:10.1111/j.1435-5957.2008.00197.x.

SLINGSBY, A., DYKES, J. and WOOD, J. (2009) Configuring Hierarchical Questions to Address Research Questions, IEEE Transactions on Visualization and Computer Graphics, 15, 977–984, doi:10.1109/TVCG.2009.128.

SLINGSBY, A., DYKES, J., WOOD, J. and RADBURN, R. (2010) OAC Explorer: Interactive exploration and comparison of multivariate socioeconomic population characteristics, Proceedings of GIS Research UK, 167–174.
URL http://gicentre.org/papers/slingsby_oacexplorer_2010.pdf

TOBLER, W. (2004) Thirty Five Years of Computer Cartograms Ground-truthing Geodemographics, Annals of the Association of American Geographers, 94, 58–73, doi:10.1111/j.1467-8306.2004.09401004.x.

VAN KREVELD, M. and SPECKMANN, B. (2007) On rectangular cartograms, Comput. Geom. Theory Appl., 37, 175–187, doi:10.1016/j.comgeo.2006.06.002.

VICKERS, D. and REES, P. (2007) Creating the National Classification of Census Output Areas: Data, methods and results, Journal of the Royal Statistical Society, Series A, doi:10.1111/j.1467-985X.2007.00466.x.

VICKERS, D. and REES, P. (2009) Ground-truthing Geodemographics, doi:10.1007/s12061-009-9037-5.

WOOD, J. and DYKES, J. (2008) Spatially Ordered Treemaps, Visualization and Computer Graphics, IEEE Transactions on Visualization and Computer Graphics, 14, 1348–1355, doi:10.1109/TVCG.2008.165.

WOOD, J., SLINGSBY, A. and DYKES, J. (2010) Layout and Colour Transformations for Visualising OAC Data, Proceedings of GIS Research UK, 455–462.
URL http://gicentre.org/papers/gisruk10/wood_layout_2010.pdf