Putting geography back into geodemographics

Martin Callingham

School of Geography, Birkbeck College, University of London
Malet Street, London WC1E 7HX
Tel: +44 (0)20 7631 6473 +44 (0) 20 7439 4377
Email: m.callingham@geography.bbk.ac.uk

1. Summary

This paper investigates for the first time the geographical properties of the Office of National Statistic's Output Area Classification system (OAC).

2. Background

Formal classifications of areas using demographics from the census have been developing over the last 50 years or so. The methods for creating these systems have now stabilised and usually involve the application of some form of cluster analysis to selected census variables (Webber et al, 1978). The output of this process is a smallish number of groups with an appropriate set of nominal variables to label them.

Considerable impetus was given to developing these systems when it was discovered that it was possible to estimate the average behaviour of people living in a place associated with a particular areal group (Bermingham et al, 1979). This had profound implications in terms of resource allocation and is the reason the term *geodemographic systems* is used to describe them rather than more conventional *areal classifications*.

OAC is a geodemographic system that has been created using the standard methods by the Office of National Statistics. Unlike most commercial systems, it is completely open. The method used is published and, just as importantly, the original database of variables used is freely downloadable: the system itself has no undue licence restrictions (Vickers et al, 2006).

As an example, the average household income in pounds sterling for each of the 52 sub-groups of OAC - OAC has a threefold hierarchy ranging from 7 super groups, 21 groups to 52 sub-groups - are plotted. The line of the x-axis is positioned to cut the ordinate at the average of the sub-groups' incomes, which, in this survey, was £26,000. The values were derived from a cross tabulation of data from a separate lifestyle survey of 998,731 respondents. Figure 1 shows the substantial variation in the modelled average household income of each of the OAC 52 sub-groups, which vary from just over £14,000 to just under £45,000.

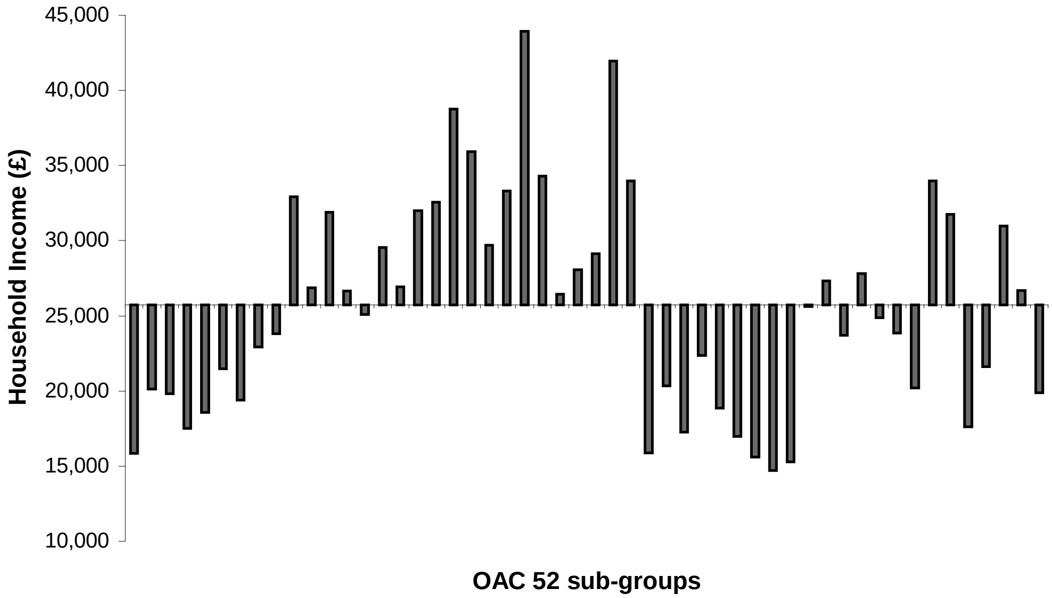

Figure 1. Mean household income for each OAC sub-group

3. The objective

Charts like this are well known and have been used for at least 25 years, see Harris et al. 2005. However, these applications have exclusively been in ascribing *people's* behaviour to the geodemographic codes. The question to be addressed in this paper is quite different and has not been examined before: are there any variations in the *geographical* properties that are associated with different geodemographic types? This question is of interest as OAC has been formed from demographic variables not geographic ones.

4. The findings

We find that:

- Codes spatially autocorrelate
- Their spatial dispersion across the UK varies by type
- They vary by their mean height above sea level
- They vary by coastal proximity
- The worker/resident ratio they contain varies by OAC sub-group

Within settlements, there is distinct bias in the locations of OAC codes with respect to:

- The population of the settlements they occupy
- Their proximity to the centre of the settlement
- And an east-west bias within their settlement

An example of one of these geographical properties, that of the variation in height above sea level of the OAC sub-groups, is shown below. This chart shows that the mean height of each OAC 52 group varies within a range of about 40 metres to 110 metres above sea level. For interest, the two clusters of sub-groups that are below the line (this being the mean elevation of all the OAC sub-groups – 68 m) are both biased towards the town centre.

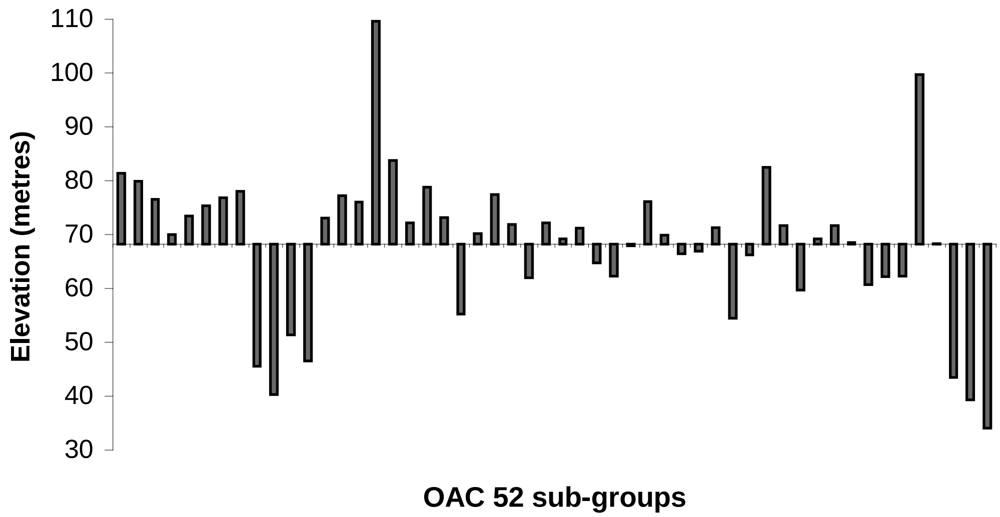

Figure 2. Mean height above sea level for each OAC sub-group

5. The method

The analysis was conducted on 218,038 OAC classified output areas of England, Wales and Scotland as follows:

1. Spatial autocorrelation was clearly shown by maps at the super group level.

2. Variations in spatial distribution were shown by analysis of the locations of the centre of gravity occupied by each sub-group again supported by maps.

3. The height effect was show by transforming a digital elevation database to each of the 218,038 output areas to give each a mean height from which the mean for each OAC sub-group was calculated.

4. Coastal proximity was measured by first coding coastal output areas of England and Wales as ‘coastal’ or ‘not’ (Scottish output area files helpfully come with a coast code). A 100m buffer was constructed around the coast and then intersected with the output area

boundary files. A ‘coastal index’ for each OAC sub-group was calculated by comparing the proportion of each OAC 52 sub-group in the coastal set to that present nationally.

5. The definition of settlements was taken from an ONS output area-to-urban area look-up table provided as part of the UK 2001 census, which defined 4570 settlements in England and Wales, and the equivalent file from the General Register Office of Scotland, which defined 587 settlements. Using these files, the population size for each settlement was obtained and from this, the mean settlement population for each sub-group easily calculated.

6. In addition, a population-weighted centroid was calculated for each settlement from its constituent output area centroids. Each output area (together with its associated OAC sub-group) could then be referenced to the centre of the settlement in which it was located and its crow-flying distance from the centre calculated by Pythagoras.

7. The spatial extent of each settlement was calculated by summing the areas of its component output areas and, by assuming that the settlement was circular, an equivalent diameter estimated. An indication as to the degree that this (circularity) assumption was true was taken to be the R-squared of the line separately put through the x, y coordinates of the output areas centroids of each of the 5157 settlements – an R-squared of zero was taken to indicate that the settlement was perfectly circular and of 1 that it was perfectly linear. The value of obtaining an equivalent diameter of the settlement is that the distance figures of the output areas from the centre could be indexed on the radius to give an indication of its relative position within the settlement.

8. However, as mean settlement populations varied across the sub-groups, the analyses for the east-west biases and the distance from the centre took place on the settlements with sizes most appropriate for them. Had this not been done, London, Glasgow and Birmingham, in particular, would have had a dominant affect on the findings.

9. The worker/resident ratio was calculated for each output area using the total resident counts from the main 2001 census data and from the little used workplace statistics file (where people have been relocated to their place of work).

6. A caveat

However, there is a fairly obvious reason why the effects shown so far may be an artefact arising from the set of variables used in the construction of OAC.

Amongst the 41 OAC variables, is population density, and it is possible that it is the presence of this non-demographic variable that is causing some or all of these associations.

We found that across the OAC 52 groups, there was a strong correlation with settlement size, very weak association with three of the properties and no association with the other three.

Given this, it is reasonable to conclude that the geographical associations reported earlier in this paper are only partly explained by population density.

75. Conclusion

It is clear that at least some locational properties are associated with areal types formed from demographic profiles. This is interesting as it implies that people of different *demography* are associated with areas of different *geography*. One might well argue that this association would be obvious, and has been claimed by social geographers for many years. However, this effect has been shown using modern methods and on a national scale. Perhaps of more interest than these particular findings, is the implication that OAC could be used as a general tool to investigate underlying geographical structures.

There are many other properties that could be examined, examples of which are: the nature of the *objects* to which these properties are applied, Unwin (2001); the influence of the *cluster hierarchy*; the role of the *cluster distance* variable as a measure of uncertainty, Fisher (2006); investigations into the *homogeneity* of the groups, Barr (2006) and the opportunity of further studying the *MAUP*. Doubtless there are more.

References

Barr, R. (2006), Private communication.

Bermingham, J., McDonald,C., Baker, K. (1979), *The Utility to Market Research of the Classification of Residential Neighbourhoods*. MRS 22nd Annual Conference, Brighton, March 21-23rd 1979.

Fisher, P. (2006), Private communication.

Harris, R., Sleight, P., Webber, R. (2005), *Geodemographics, GIS and Neighbourhood Targeting.* (Wiley: London).

Unwin, D. (2001), Private communication.

Vickers, D., Rees, P. (2006), *Introducing the Area Classification of Output Areas.* Population Trends 125.

Webber, R., Craig, J. (1978), *Socio-economic classifications of local authority areas.* Studies in Medical and Population Subjects **35**.Office of Population, Censuses and Surveys: London.

Office for National Statistics (2005), *Area classification for Output Areas*. [http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/default.asp].

Biography

Martin Callingham is a Visiting Professor at Birkbeck College, University of London where his research interests are areal classification and flow data. He has been involved with geodemographic systems since their inception and was a member of the ONS working party for the creation of their Output Area Classification