Exploring social dynamics: predictive geodemographics

Jennie Gray[*1], Lisa Buckner[†2] and Alexis Comber[‡3]

[1] Leeds Institute for Data Analytics (LIDA), University of Leeds, Leeds, LS2 9JT, UK
[2] School of Sociology and Social Policy, University of Leeds, Leeds, LS2 9JT, UK
[3] School of Geography, University of Leeds, Leeds, LS2 9JT, UK

January 16th, 2019

KEYWORDS: Geodemographics, spatiotemporal analysis, predictive analytics, spatial dynamics

Abstract

Geodemographics is the analysis of places or areas by the characteristics of the people that live there. It is underpinned by the notion that people residing within close proximity exhibit similar characteristics (Harris et al, 2005). Geodemographic data are typically area classifications or measures based on composite variables. They are typically used to segment populations into relatively homogenous groups. Famously the first such area classification is Booth's multivariate classification of the 1891 population Census variables including tenement size, occupancy, and number of servants, birth and death rates, and occupation (Booth, 1893). In the 1970's Webber developed the ward level classification for the UK known as ACORN (Webber, 1977) and spawned a number of parallel classifications. These early activities were targeted towards local authorities with an interest in understanding the distribution of people, housing, and social deprivation (Birkin and Clarke, 2009), with ACORN developed from work exploring inner-city deprivation in Liverpool, for example. Commercial geodemographic classification were to support spatially targeted marketing (e.g. Baker et al, 1997) with increased spatial resolution (Sleight, 2004). Recently open classifications have been developed such as the Output Area Classification (OAC) in the UK Vickers and Rees (2007) and geodemographic classifications are now a standard business tool supporting market analysis and decision-making (Leventhal, 2016). Geodemographic classifications have several limitations as highlighted by researchers (Gale and Longley, 2013; Harris and Feng, 2016). One of these is that they are typically developed using population census data (with other data). Census data are out date almost from the moment they are released (typically 2 or 3 years after the census period) and are collected infrequently: they are temporally static. These temporal limitations can mean that geodemographic data fail to capture the true nature of an area (Gale and Longley, 2013). Some work has been conducted to understand these dynamics and limitations. Singleton et al (2016) used both the 2001 and 2011 OAC to create a Temporal Output Area in order to analyse the stability of geodemographic classifications. The results indicated that 39% of OAs in 2011 were reassigned from their 2001 counterpart (Singleton et al, 2016), suggesting some level of geodemographic cluster instability, or geodemographic change. In reality, local areas are dynamic and may undergo changes in that are not captured by decadal censuses. Thus exploring the dynamics of geodemographics through the analysis of intercensal data may provide a critical understanding of the characteristics of geodemographic change, and the drivers of that change. This research analyses and explores the dynamics of geodemographics. Using a case study in Sheffield, it utilises open data from a number of sources, data providers and portals to examine the following questions:

1. What datasets are able to explain the dynamics of geodemographics?
2. Which societal or demographical processes therefore best reflect and explain the change within geodemographic clusters?
3. Are any cyclical patterns within the processes identified?

In answering these questions this research seeks to develop and inform a notion of *predictive*

[*] gyjhg@leeds.ac.uk
[†] L.J.Buckner@leeds.ac.uk
[‡] a.comber@leeds.ac.uk

geodemographics, the topic of my PhD. It seeks to understand how different social processes and dynamics as captured in different data, are able to efficiently explain geodemographic changes, thus potentially overcoming the issue of static data and temporal limitations of such classifications.

In this initial study, Temporal Output Area Classification (TOAC) clusters for Sheffield were analysed. Decennial changes (28.1% of OAs reassigned from 2001 to 2011 clusters) were not reflective of the annual TOAC reassignments (31.6% of OAs reassigned throughout 2001 to 2011). TOAC clusters are then analysed against dynamic data, that cover a period of 10 years and are updated at least biennially. This analysis takes a similar approach to that of Singleton et al (2016), but with the use of open data that covers the intercensal years in order to determine which data best explain the observed annual changes in TOAC cluster assignment over the period, thus the social processes they represent. Table 1 indicates the data used for analysis.

Table 1 Potential Dynamic Datasets for Analysis

Data	Source
Population Density	NOMIS
Unemployment Records	NOMIS
DWP Benefits Data (JSA, IB, etc).	NOMIS
Health Data	Gov.UK
Geodemographic Classification	Temporal Output Area Classification TOAC

Having data over periods of 10 years (2001-2011), allows demographic trajectories to be identified and to be validated by the second OAC, while also acting as a TOAC validation exercise. The outputs of this research will provide insights to aid the development and understanding of methods to support and generate Predictive Geodemographics. It suggests a number of areas of future work, some of which may be essential for my Predictive Geodemographics PhD, such as; how to select and handle data with greater temporal dynamics (e.g. real time data)? When threshold of change is sufficient to warrant a change in geodemographics class label? How should the changes in the characteristics of areas (i.e. their attributes in the database) and the associated impacts on statistical segmentation routines be handled? This research, and the wider questions it suggests have the potential to support research in a number of fields.

References

Baker K McDonald C and Bermingham J (1997). The utility to market research of the classification of residential neighbourhoods. *International Journal of Market Research*, 39(1).

Birkin M and Clarke G (2009) *Geodemographics*. International encyclopaedia of human geography. Oxford: Elsevier, 382-389.

Booth C (1893). Life and Labour of the People in London: First Results of an Inquiry Based on the 1891 Census. *Journal of the Royal Statistical Society*, 56(4), 557-593.

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

Harris R and Feng Y (2016). Putting the geography into geodemographics: using multilevel modelling to improve neighbourhood targeting - a case study of Asian pupils in London. *Journal of Marketing Analytics*, 4(2/3), 93-107.

Harris R Sleight P and Webber R (2005). *Geodemographics, GIS and neighbourhood targeting*. Chichester: John Wiley & Sons.

Singleton A Palvis M and Longley P (2016). The stability of geodemographic cluster assignments over an intercensal period. *Journal of Geographical Systems*, 18 (2), 97-123.

Sleight, P. 2004. *Targeting customers: how to use geodemographic and lifestyle data in your business*. Oxford: World Advertising Research Centre.

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*(2), 379-403.

Webber, R. 1977. *An Introduction to the National Classification of Wards and Parishes*. London.