

Using Big Data to Measure the Demographic Changes in a Gentrifying Neighbourhood

Guy Lansley^{*1}, Wen Li^{†1} and Paul Longley^{‡2}

¹Department of Geography, UCL

January 7, 2019

Summary

This paper presents the use of a micro-level population dataset constructed from linked administrative and consumer data to model the local demographic changes that have occurred in a given area. The constituents of this database derive from 20 years of linked electoral registers and consumer datasets which when amalgamated represent the vast majority of the adult population. While the component data inputs only include dates, names and addresses, this paper shall demonstrate that a range of novel data linkage exercises can reveal local population trends that were otherwise unavailable from traditional data sources.

KEYWORDS: big data, population, population change, geodemographics, data linkage

1. Introduction

Like many cosmopolitan neighbourhoods of large cities around the world, Spitalfields in East London has experienced rapid social and demographic change through gentrification. However, the extent of these changes are largely unknown and pre-existing publicly available data on the population and internal migration patterns are restricted to datasets that are infrequently collected or unavailable at fine geographic scales. Thus, little is known about the origins of people who moved into the area as it gentrified, nor the destinations of those that left. However, data on where people live are routinely collected by government institutions and commercial organisations every year. Collectively, these data pose a considerable opportunity to improve our understandings of the population across space and time.

Gentrification refers to process through which the middle-class move into traditionally working class neighbourhoods altering the local social fabric (Smith and Williams, 1986). Indeed, Spitalfields has changed considerably over the past 20 years. The area is situated in close proximity to the City of London and hosts London's iconic Brick Lane, traditionally the heart of London's Bangladeshi community. As the area has regenerated its geodemographic composition is estimated to have changed considerably due to an influx of young professionals and students and the decline of the traditional Bangladeshi population (Kershen, 2005). Indeed, much academic literature has focused on the negative consequences of gentrification, mainly rising housing costs which can lead to the displacement of the pre-existing lower socio-economic population (Atkinson, 2000; Palen and London, 1984). However, these trends are inherently difficult to measure for small areas.

In light of this, this paper harnesses linked consumer and administrative data collected over a 20 year period to measure the geodemographic changes that have occurred in Spitalfields. Using an address-level linked database which is representative of the vast majority of the adult population, highly granular and annual estimations of geodemographic characteristics such as ethnicity and population turnover have been produced. In addition, the origins and destinations of internal migrants to and from Spitalfields have also been calculated in order to track possible ethnic displacement patterns.

* g.lansley@ucl.ac.uk

† wen.li@ucl.ac.uk

‡ p.longley@ucl.ac.uk

2. Data

This study considers an individual-level demographic database for the UK (1997-2016). It was constructed from the linkage of 20 years of public electoral registers with records from anonymised consumer sources to account for persons that have either opted out of the ‘edited register’ or were not eligible to vote (Lansley et al, 2018). The component datasets contained only the names of individuals and their addresses in each annual release. Following linkage, the ‘Linked Consumer Registers’ (LCR) provide a database which lists the adult residents at almost every active address in the UK for individual years.

While the LCR is limited in the raw variables it contains further analysis and novel data linkage exercises can reveal fascinating geodemographic trends at high resolutions. Some of these have briefly been described in **Table 1**.

Table 1 A summary of selected variables that can be derived from the Linked Consumer Registers

Variable	Description
Households	Adult household sizes and compositions. Surnames can indicate if multiple person households are families or not (Samuel et al, 2019)
Population turnover	The identification of the first and last instances of individuals and their household units at each address (Lansley et al, 2018).
Tenure	Through linking household turnover estimates to house sale data from England, Wales and Scotland it is possible identify properties that are not owner-occupied.
Gender and age	Gender and age group can both be inferred from a forenames database built from consumer data and birth certificate records (Lansley and Longley, 2016)
Ethnicity	Ethnic groups can be inferred from names through the Consumer Data Research Centre’s Ethnicity Estimator (Longley and Kandt, 2018).
Internal Migration	A novel matching algorithm has been developed to detect probable origin and destinations of internal migrants from the Linked Consumer Registers (Lansley and Li, 2018)

3. Results

An advantage of employing this dataset is its high granularity enables analysis to focus on very specific areas across a selected time period. Furthermore, longitudinal work on alternative datasets is often limited by the fact that small area units are subject to alterations over time. For this study we have considered only addresses that match postcodes within the Spitalfields and Banglatown ward as of 2018 (**Figure 1**).

Between 1997 and 2016 the Spitalfields area has experienced considerable population growth. The adult population has increased by over 40%. Furthermore, the new population are living for short durations on average. For instance, over 50% of households present in 2016 have moved in since 2012.

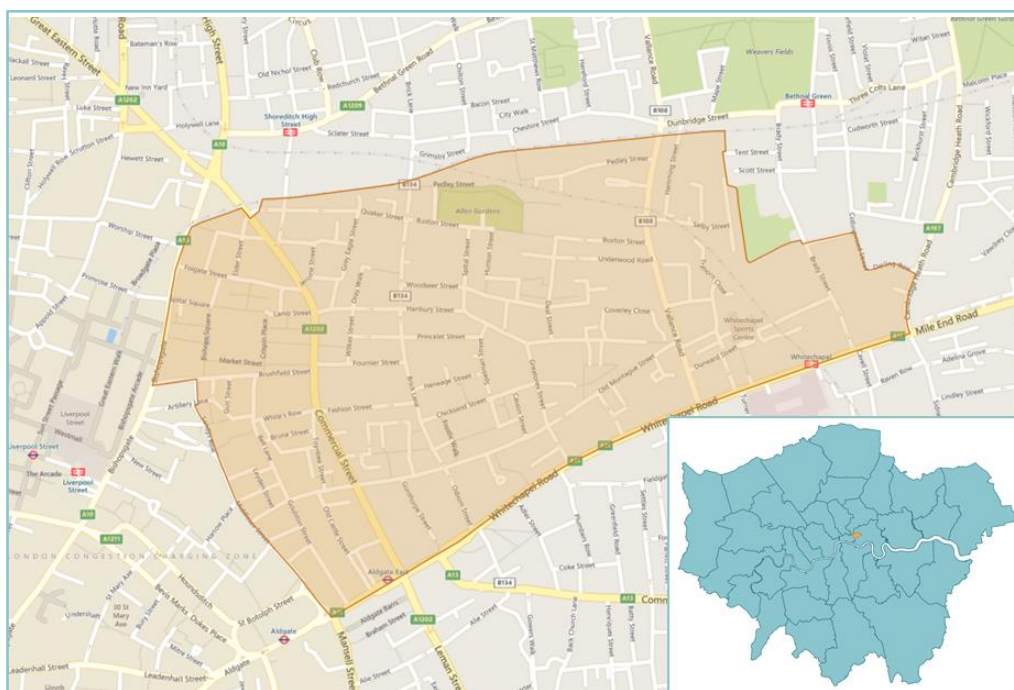


Figure 1 Map of the Spitalfields and Banglatown ward

During the 20 year period, the ethnic composition of the neighbourhood has also changed. The proportion of Bangladeshi's has almost halved, whilst the White British and White Other groups have grown (**Figure 2**). We also estimated that the proportion of males has been gradually increasing since 1997, this is probably indicative of the proximal financial district.

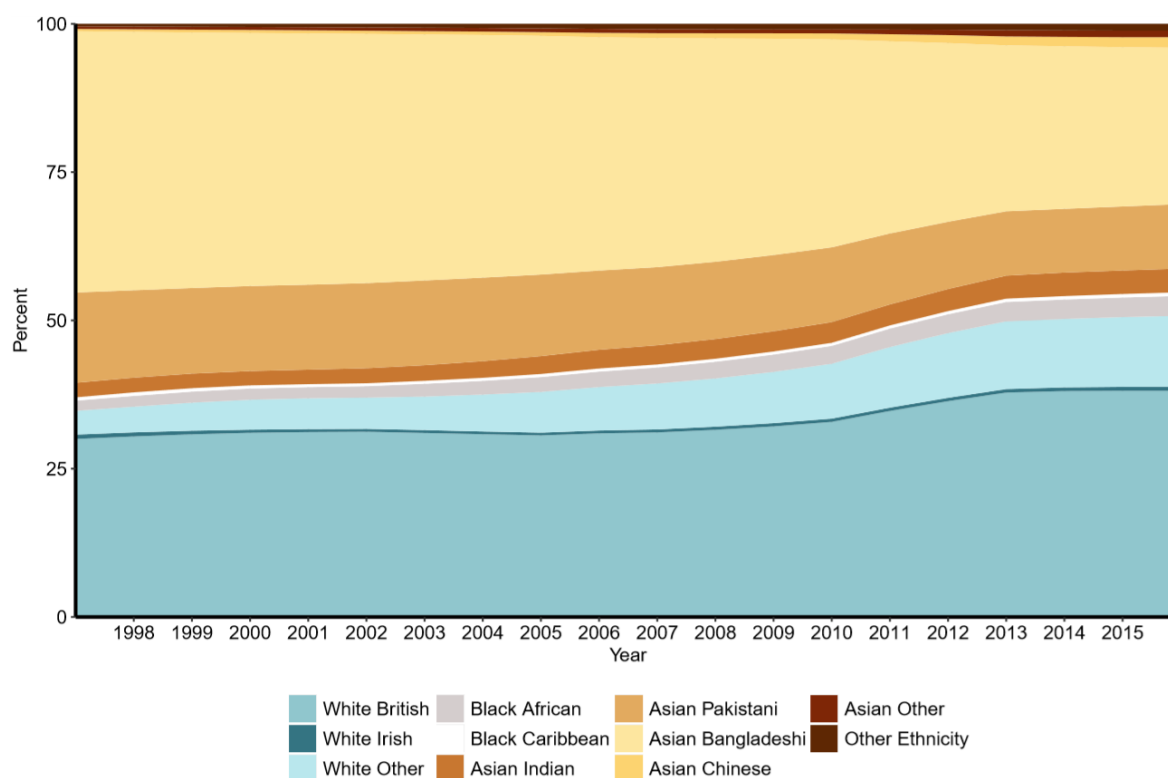


Figure 2 The changing ethnic structure of Spitalfields and Banglatown (1997-2016)

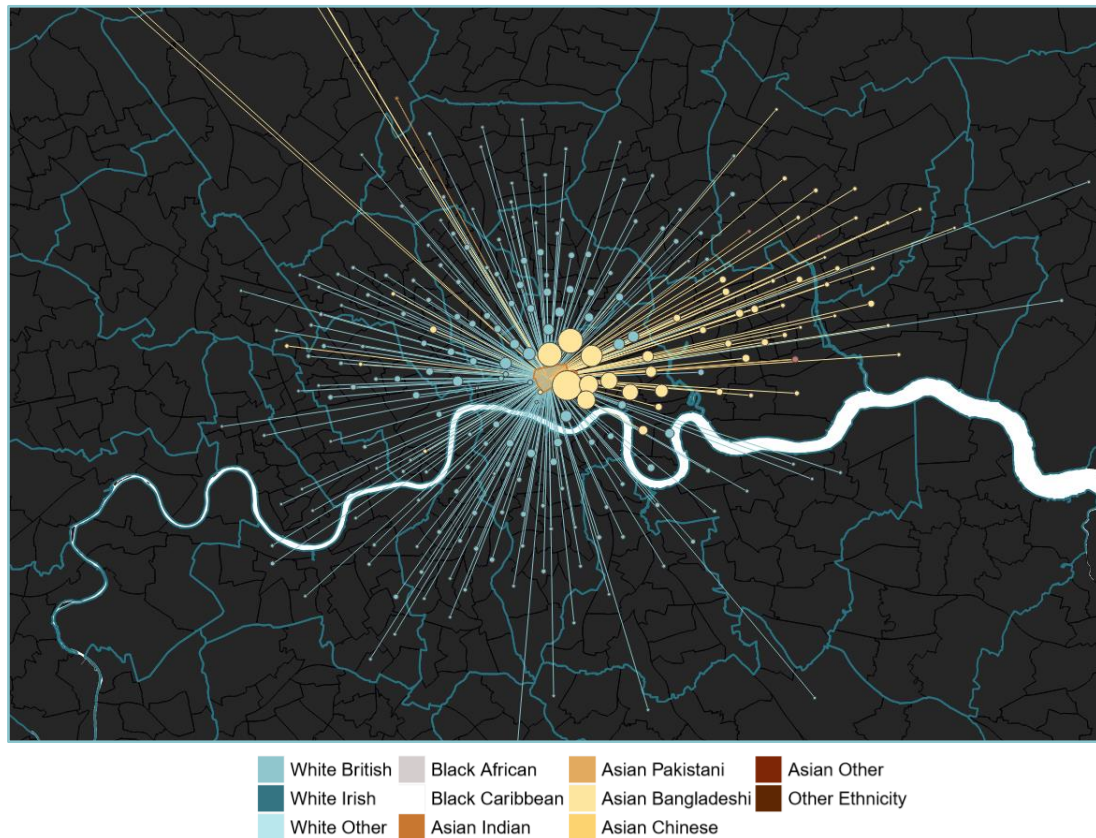


Figure 4 The destination wards of internal migrants from Spitalfields and Banglatown, coloured by the most common ethnic group within each flow. Only flows of 10 migrants or more are shown

4. Conclusions

This research has provided a demonstration of how more granular geodemographic insights can be gleaned from linking routinely administered big datasets which collectively represent the vast majority of the adult population. Overall, the Linked Consumer Registers provide an exciting opportunity to conduct demographic analysis at bespoke spatial and temporal scales more appropriate for particular research questions.

5. Acknowledgements

This work is funded by the UK ESRC Consumer Data Research Centre (CDRC) grant reference ES/L011840/1.

Biography

Guy Lansley is a Research Associate at the UK Consumer Data Research Centre and the Department of Geography at University College London (UCL). His research is primarily focused on harnessing geodemographic insights from big consumer datasets of unknown provenance.

Wen Li is a Data Scientist at the UK Consumer Data Research Centre and the Department of Geography at UCL. His main research focuses on data integration by applying methodologies from information retrieval and distributed computing.

Paul Longley is Professor of Geographic Information Science at University College London and director of the UK Consumer Data Research Centre at UCL. His publications include 14 books and more than 150 refereed journal articles and book chapters. He is a former co-editor of the journal

Environment and Planning B and a member of four other editorial boards.

References

- Atkinson, R. (2000). The hidden costs of gentrification: Displacement in central London. *Journal of housing and the built environment*, 15(4), pp.307-326.
- Kandt, J. and Longley, P.A. (2018). Ethnicity estimation using family naming practices. *PloS one*, 13(8), p.e0201774.
- Kershen, A.J. (2005) *Strangers, Aliens and Asians: Huguenots, Jews and Bangladeshis in Spitalfields 1660–2000*. New York: Routledge
- Lansley, G., Li, W. and Longley, P. (2018). Modelling Small Area Level Population Change from Administrative and Consumer Data. *Geographical Information Science Research UK (GISRUK) 2018*.
- Lansley, G. and Li, W. (2018). Consumer Registers as Spatial Data Infrastructure and their Use in Migration and Residential Mobility Research. In Longley, P.A., Singleton, A. and Cheshire, J. (Eds.) (2018). *Consumer Data Research* (pp. 15-27). UCL Press.
- Lansley, G. and Longley, P. (2016). Deriving age and gender from forenames for consumer analytics. *Journal of Retailing and Consumer Services*, 30, pp.271-278.
- Palen, J.J. and London, B. eds. (1984). *Gentrification, displacement, and neighborhood revitalization*. Suny Press.
- Samuel, A., Lansley, G. and Coulter, R. (2019). Estimating the Prevalence of Shared Accommodation across the UK from Big Data. *Geographical Information Science Research UK (GISRUK) 2019*.
- Smith, N. and Williams, P. (Eds) (1986) *Gentrification of the City*. London: Unwin Hyman.