Using linked consumer and administrative data to model demographic changes in London's city fringe

Justin van Dijk, Guy Lansley, and Paul A. Longley
Department of Geography, University College London

Abstract
Like many other cosmopolitan neighbourhoods around the world, several neighbourhoods in East London have experienced rapid social and demographic change through gentrification. This chapter harnesses linked consumer and administrative data collected over a 20-year period to measure the geodemographic changes that have occurred in three neighbourhoods in London's city fringe: Hoxton East and Shoreditch, Spitalfields and Banglatown, and Whitechapel. Using an address-level linked database, representative of the vast majority of the adult population in the United Kingdom, we produce highly granular estimations of geodemographic characteristics such as ethnicity and we characterise residential moves by their origins and destinations.

Keywords
consumer data, residential mobility, gentrification, demographic change, London's city fringe

Introduction
Like many cosmopolitan neighbourhoods around the world, several neighbourhoods in London's city fringe have experienced rapid social and demographic change through gentrification. Gentrification is a process of neighbourhood upgrade and transformation in which working-class neighbourhoods transition into middle-class neighbourhoods. Following an increase in rents as well as changing neighbourhood characteristics, gentrification typically goes hand in hand with the displacement of the poorer people living in those areas (Smith and Williams 1986, Atkinson 2000). In London, starting in the second half of the 1990s, areas in the ex-industrial city's fringe northeast of the City of London have been particularly affected (Hamnett 2003, 2009). Notwithstanding the urban regeneration of often run-down neighbourhoods, the most profound effect of gentrification is demographic change as a result of residential mobility. So far, however, much of the British literature on gentrification strongly draws on qualitative methods to describe demographic changes whilst the few studies that use quantitative methods heavily rely on data from the UK Census of Population (Reades *et al*. 2019).

In the United Kingdom, publicly available data on the population and internal migration patterns are restricted to datasets that are unavailable at fine geographic scales or infrequently collected. For example, the 2011 UK Census of Population estimated that around 11 per cent of the population changed house in the 12 months preceding the data collection. For later years, however, details on residential mobility are derived from Mid-Year Population estimates – which are likely to deviate further from the actual numbers in later years. Similarly, where the 2011 Census included details of the number of residential moves between Local Authority

Districts, the later District-level Mid-Year Population estimates only include the estimated total number of in-movers and out-movers (ONS 2019a, ONS 2019b). These datasets do not include the majority of short distance moves despite the fact that these are the most common residential moves (Lomax and Stillwell 2017).

In the absence of frequently updated data on the nature of residential moves and the circumstances of movers, this chapter uses a unique digital corpus of linked individual and household level consumer registers compiled by the UK Consumer Data Research Centre (Lansley *et al*. 2019). These so-called Linked Consumer Registers (LCRs) have been used to model residential mobility by assigning individuals vacating a property to their most probable destination (Van Dijk *et al*. 2020). Our highly disaggregate and scale-free analysis makes possible highly granular origin-destination analysis of residential mobility outcomes and allows us to relate demographic change with neighbourhood change. In light of this, this chapter harnesses these consumer and administrative data collected over a 20-year period to measure the demographic changes that have occurred in three neighbourhoods in London's city fringe that are well-known to have witnessed great change: Hoxton East and Shoreditch, Spitalfields and Banglatown, and Whitechapel (Figure 1).

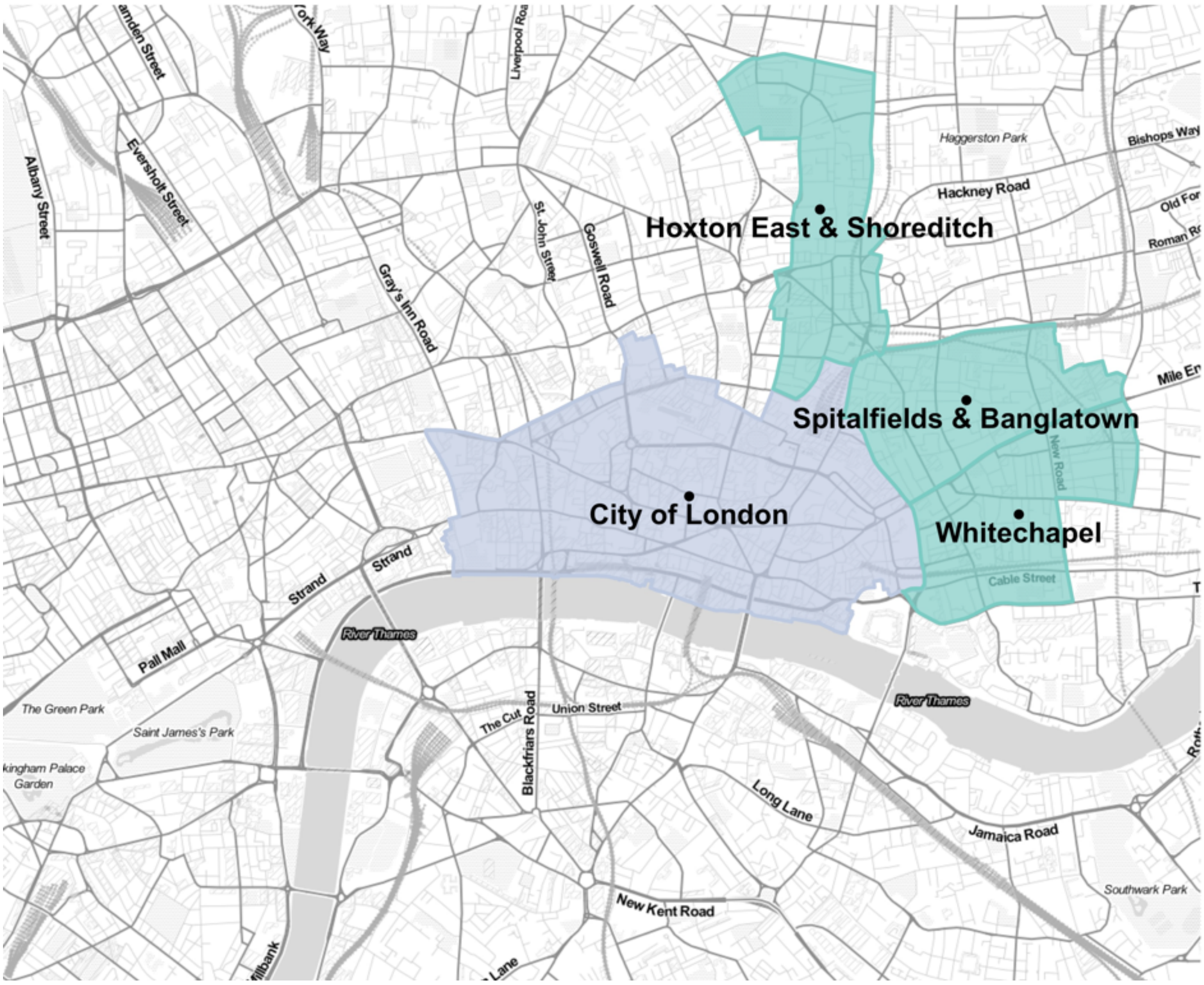

Figure 1. Location of the three selected neighbourhoods in London's city fringe.

Linked Consumer Registers and Migration Estimates

Traditional surveying methods are costly, and, around the world, longitudinal surveys have suffered from diminishing response rates (Bianchi and Biffignandi 2019). In contrast, new Big Data offer large volumes of data at greater spatial and temporal granularity to what is currently available via traditionally sourced social datasets. Many consumer or administrative datasets are already capturing huge shares of the population, although their provenance remain uncertain and no single dataset maintains completeness of coverage. In this chapter, an individual-level database is used that was constructed from the linkage and consolidation of 20-years of public electoral registers augmented with records from consumer source. These 'linked consumer registers' (LCRs) comprise a total of 143 million records of observed or imputed individual names and geo-referenced addresses in the United Kingdom for the period 1997 to 2016 (Lansley *et al.* 2019).

The LCRs achieve near-complete coverage in terms of counts, although quality of the data may vary due to variable speeds in data velocity across time, space and source provider. International migrants and young adults were found to be underrepresented, and it is also possible that many who live at multiple addresses might be duplicated (e.g. second homeowners). At the time of writing though, the LCR remains the most complete UK adult population register available for academic research. While the LCRs' raw data only contain a few variables (i.e. forename, surname, address, and first and last year the individual was recorded at the address), other variables can be estimated (Table 1). The LCRs have been used for a variety of studies on topics such as ethnic segregation (Lan *et al.* 2019) and, through indirect linkages to Historic Censuses of Population, intergenerational population change (Kandt *et al*. 2020). See Lansley *et al.* (2019) for a full discussion on the LCR data infrastructure, the novel methods used to link address information to Ordnance Survey's AddressBase Premium, and comparisons to official data sources such as the Mid-Year Population estimates and the 2011 Census of Population.

Table 1. Linked Consumer Registers and derivable information

Variable	**Description**
Households	Household size and composition as suggested by the surnames that are present on an address.
Population churn	Identification of the first and last member of a household at an address.
Tenure	Owner-occupied properties can be identified through linkage with databases on rental data and land registry property sale data.
Gender and age	Gender and age can be derived by forename analysis (see Lansley and Longley, 2016).
Ethnicity	Ethnicity can be estimated using forenames and surnames (see Kandt and Longley, 2018).

While the LCR data contained no information on movements, it is possible to take advantage of the granularity of the data and use names as discriminators of identity. A model was devised to link the last record of every adult at an address to other records that commence during the same year or thereafter. Adults were linked by names and, where full names were not unique, an algorithm was applied to allocate them to the most probable pairings by considering the total number of matched residents, distance between individual properties, time between observations and the possible occurrence of a housing chain in the Land Registry database. The Migration Model is estimated to have returned a significant share of the actual UK-wide internal migration moves for the period 1997 to 2016. Direct comparison with the 2011 Census estimates of residential moves of residents aged 16-years and older suggests that the Migration Model accounts for 45.3 per cent of all moves across the United Kingdom. A Pearson correlation coefficient of 0.97 between predicted inter-District moves and those recorded in the Census indicates that the level of under-recording is consistent between districts. The Pearson correlation coefficient remains stable at 0.86 ± 0.02 for post-Census years in which predictions are compared with ONS Mid-Year Population estimates (see Van Dijk *et al.* 2020 for a full discussion on the development of the Migration Model and comparisons to the 2011 Census of Population).

Demographic change in London's city fringe

The LCRs and the derived Migration Model allow for the analysis of residential mobility at a high temporal and spatial granularity. For instance, it is now possible to quantitatively analyse changing ethnic composition at the neighbourhood-level (electoral wards) by using the Ethnicity Estimator name-classification software (Kandt and Longley 2018, ee.cdrc.ac.uk) to assign individuals to their most probable ethnicity. This name-based classification tool builds on the experience of previous purely algorithmic classifications (e.g. Mateos *et al.* 2011) which through secure access to ONS Census data, making it possible to calibrate estimates using individual self-assignments of ethnicity. This procedure is robust, having a documented overall success rate of 88 per cent, but demonstrates the inherent uncertainty of estimating what is essentially a social construct (Mateos *et al*. 2009). For instance, individuals may not personally identify with an ethnic group on the basis of names determined by their forebears.

Figure 2 shows the share of population belonging to one of eleven ethnic groups differentiated between in the 2011 Census and. In all three neighbourhoods, there is an inflow of *White British* and *White Other* people. Particularly noticeable are the changes in Whitechapel and Spitalfields and Banglatown, where the Bangladeshi community has been declining rapidly a result of an influx of predominantly young professionals (Kershen 2005). In Hoxton East and Shoreditch, on the other hand, the effects are less pronounced, although it can be seen that the share of the *British Black African* and *British Black Caribbean* population has decreased over the years. Although not shown in Figure 2, it is worth noting that in neighbourhoods further to the East such as Bethnal Green, Stepney Green, and Shadwell, the share of *White British* and *White Other* is actually getting smaller as a result of an influx of other groups such as British Bangladeshis. This strongly suggests a displacement effect as a result of gentrification.

Not only the ethnic composition has changed, but also the population in the three neighbourhoods has grown significantly. Between 1997 and 2016, the number of adults living in Hoxton East and Shoreditch more than doubled with a population of 5,009 and 10,529 in 1997 and 2016, respectively. Similar trends can be seen in Whitechapel, which has seen an increase in population from 6,926 to 10,680, and Spitalfields and Banglatown, where the population grew from 6,461 to 9,813. Both the increase in population and the changing ethnic composition has likely been facilitated by loft conversions of old industrial buildings – a form of developer-led gentrification that particularly shaped these ex-industrial areas in the second half of the 1990s (Hamnett 2009).

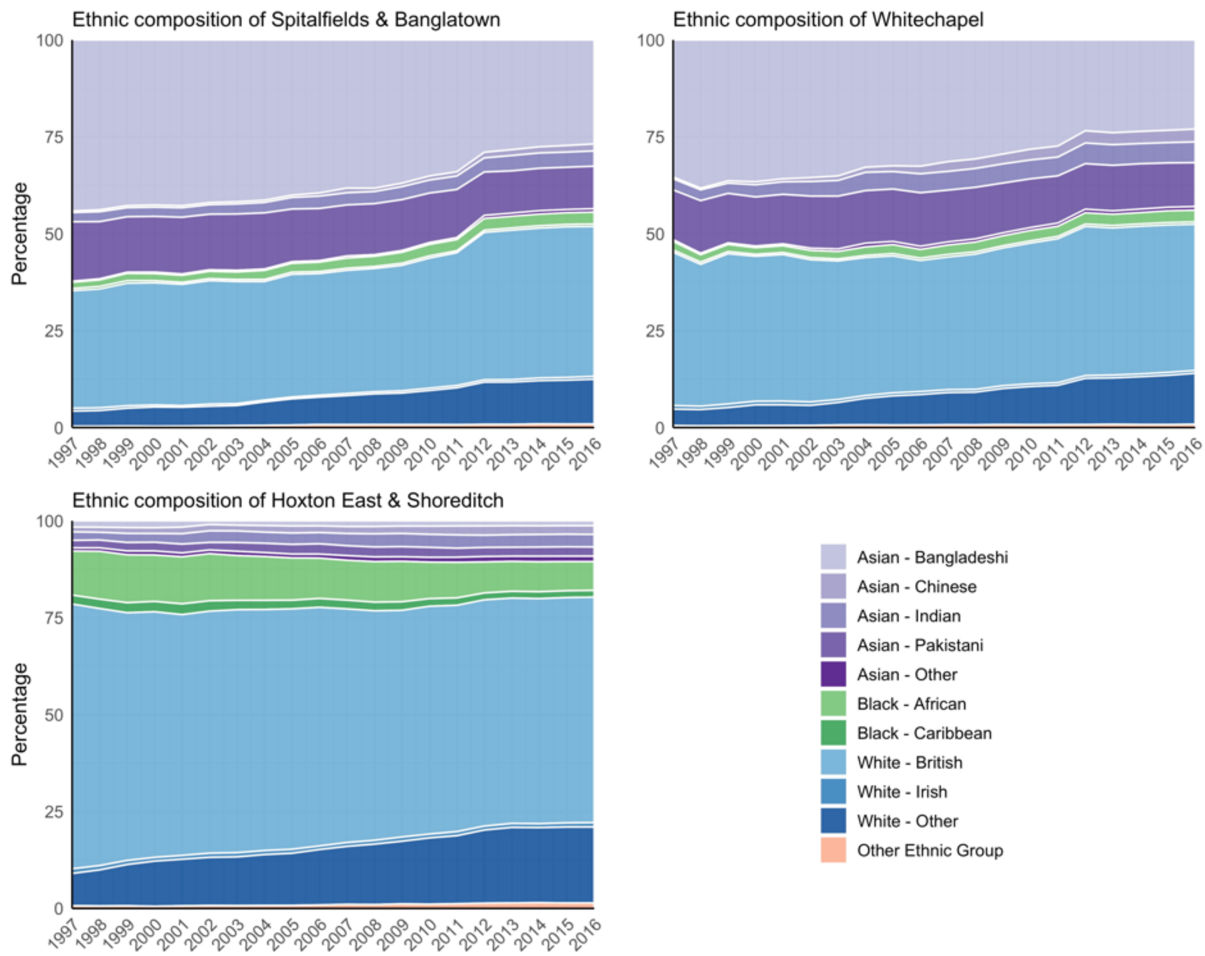

Figure 2. Changes in the ethnic composition of three neighbourhoods in London's city fringe: Spitalfields and Banglatown, Whitechapel, and Hoxton East and Shoreditch, 1997 – 2016.

Although the LCRs' Migration Model does not capture international moves, the migration estimates give a highly granular account of the internal residential mobility in the three neighbourhoods. Table 2 shows the total number of inward and outwards moves that are captured by the Migration Model across the entire period 20-year period for Spitalfields and Banglatown, Whitechapel, and Hoxton East and Shoreditch. In addition, the mean and median distances of these moves are shown. Interestingly, the mean of distance for inwards moves is overall higher than the mean of distance for outwards moves, suggesting that we can prudently infer that neighbourhoods in the city's fringe are more attractive to those who are new to the city.

Table 2. Mean and median distance of (metres) moves by inward moves and outward moves and count of total moves. Aggregates for 1997-2016.

		Mean	Median	Count
Spitalfields and Banglatown	*in-movers*	38,050	4,206	7,820
	out-movers	31,657	4,256	8,690
Whitechapel	*in-movers*	38,935	4,645	9,092
	out-movers	30,709	4,531	10,255
Hoxton East and Shoreditch	*in-movers*	36,556	5,436	7,314
	out-movers	31,342	4,940	7,153

Table 3 shows the proportion of moves that takes place within the same geographical area. Almost half of the moves, with Hoxton East and Shoreditch as an exemption, have taken place within the same London borough. Furthermore, although the numbers are rather similar, what is most striking is that the majority of the moves have taken place within the boundaries of the Greater London Authority. Again, this can suggest that these three neighbourhoods are slightly more attractive for people that are new to the city and use these neighbourhoods as an entry point. Residents who have been living in these areas for longer, or move on after their initial arrival in the area, are slightly more likely to move to other neighbourhoods within London

Table 3. Proportions of inward moves and outward moves coming from or going to the same Ward, Lower level Super Output Area (LSOA), Middle layer Super Output Area (MSOA), Local Authority District (LAD), or region. Aggregates for 1997-2016.

		Ward	LSOA	MSOA	LAD	Region
Spitalfields and Banglatown	*in-movers*	0.235	0.166	0.217	0.444	0.829
	out-movers	0.212	0.150	0.196	0.461	0.859
Whitechapel	*in-movers*	0.228	0.161	0.205	0.440	0.824
	out-movers	0.202	0.143	0.182	0.464	0.861
Hoxton East and Shoreditch	*in-movers*	0.199	0.156	0.198	0.350	0.831
	out-movers	0.204	0.160	0.201	0.378	0.847

The granularity of the Migration Model also allows residential moves to be related to other geodemographic classifications such as the 2019 English Index of Multiple Deprivation (IMD) scores. Table 4 categorises all inward moves and all outwards moves into the IMD quintiles of their origins or destinations. Moves within the same ward are excluded from this analysis. For example, the first quintile holds the proportions of in-movers that originate from areas that are least deprived. For out-movers, the first quintile holds the proportions of out-movers that move to areas that are considered least deprived. For all three neighbourhoods, the proportion of residents moving into the neighbourhood coming from an area that falls into the first three quintiles is smaller than the proportion of residents moving out of the neighbourhood to an area that falls into the first three quintiles. The opposite holds for the last two quintiles. It has long been known that one of the primary motives for a residential move is the desire of a household to improve their living conditions (Fielding 1992), and the proportions of moves between areas with different IMD quintiles seem to suggest that these neighbourhoods in London's city fringe fulfil a roll in this pattern of upward social mobility.

Table 4. Proportions of inward moves and outward moves by the 2019 English Index of Multiple Deprivation quintiles. Aggregates for 1997-2016. Quintile 1 being the least deprived quintile, quintile 5 being the most deprived quintile.

		Q1	**Q2**	**Q3**	**Q4**	**Q5**
Spitalfields and Banglatown	*in-movers*	0.230	0.366	0.176	0.123	0.090
	out-movers	0.259	0.395	0.155	0.107	0.068
Whitechapel	*in-movers*	0.246	0.342	0.170	0.129	0.095
	out-movers	0.262	0.364	0.159	0.121	0.080
Hoxton East and Shoreditch	*in-movers*	0.233	0.344	0.185	0.127	0.094
	out-movers	0.256	0.374	0.170	0.116	0.071

Conclusion and discussion

This chapter has demonstrated how new source of Big Data can provide granular insights into residential mobility, how individual-level address data can be linked to existing geodemographic classifications, and how quantitatively available data on residential moves can be used to make inferences about demographic changes and related to wider processes of gentrification. As such, the combination of the LCRs (Lansley *et al.* 2019) and the Migration Model (Van Dijk *et al.* 2020) comprises a unique, scale-free, population-wide resource that will be of strategic importance to a number of policy concerns, such as (a) the identification of the distance distribution of residential moves; (b) small area characterisation of the origins and destinations of residential moves using existing geodemographic classifications to better understand the linkages between residential and social mobility; (c) changing house price gradients that characterise moves between origins and destinations, and their implications for labour mobility; and (d) the interactions between all of the above and household structure, measured in terms of household size and composition in terms of ethnicity.

A final important question is whether the consumer Big Data used to create the LCRs and the Migration Model are sufficiently reliable to supplement or even replace less frequently collected data from the Census of Population. To date, comparative analysis suggests that our demographic and residential mobility statistics enable robust yet more frequently updated estimates than can be gleaned at small area level from official sources. In our own collaborative work with the Greater London Authority (GLA), user evaluation has suggested that deployment of these data for local policy formulation and analysis is very helpful but not without challenges. GLA evaluation (R Cameron 2020, personal communication, March 6) has suggested that ethnic group estimates for *British Asian* groups are reliable; the *White British* figures were considered to be overestimates, possibly as a result of corresponding underestimation of *White Irish* (where issues of self-assignment of members of this long-settled group) and *British Black Caribbean* (where similar naming conventions to the White British group create ambiguities, rendering the minority community hard to identify). This evaluation should also be tempered with the observation that names-based classification of ethnicity allows greater differentiation of groups that have distinctive naming practices but which are subsumed within somewhat unhelpful blanket categories in small area Census statistics – with the distinctiveness of names originating from different European nations, subsumed within the Census 'White Other' category being a prominent example. GLA evaluation suggests that geographical variability in population churn was captured very well over-all, albeit that estimates for several individual consistencies merited further investigation. Based upon this user evaluation and our own efforts to triangulate estimates with official sources, our interim view is that consumer Big Data offer an innovative and valuable approach to creating small area measures, consistent with disclosure control (cf. Harris and Longley 2002). Our own future research agenda will investigate issues of coverage of hard-to-reach groups.

Notes

1. Access to small-area aggregates and estimates from this project may be obtained upon successful application to the Consumer Data Research Centre Data Service (data.cdrc.ac.uk).

Acknowledgements

This research was funded by UK Engineering and Physical Sciences grant reference EP/M023583/1 (funder of the Urban Dynamics Lab: www.ucl.ac.uk/urban-dynamics-lab/) and UK Economic and Social Research Council grant ES/L011840/1 (funder of the Consumer Data Research Centre: www.cdrc.ac.uk/). We are grateful to Richard Cameron of the Greater London Authority for very helpful feedback on the usefulness of some of the work described here for local policy formulation and analysis.

References

Atkinson, R. 2000, The hidden cost of gentrification: Displacement in central London. *Journal of Housing and the Built Environment* 15, pp.307-326. doi: 10.1023/A:1010128901782

Bianchi, A. and Biffignandi, S. 2019. Social indicators to explain response in longitudinal studies. *Social Indicator Research* 141, pp.931-957, doi: 10.1007/s11205-018-1874-7

Fielding, A. J. 1992. Migration and social mobility: South East England as an escalator region. *Regional Studies*, 26:1, pp.1-15. doi: 10.1080/00343409212331346741

Hamnett, C. 2003. Gentrification and the middle-class remaking of inner London, 1961-2001. *Urban Studies* 40(12), pp.2401-2426. doi: 10.1080/0042098032000136138

Hamnett, C. 2009. City centre gentrification: Loft conversions in London's city fringe. *Urban Policy and Research* 27(3), pp.277-287. doi: 10.1080/08111140903132200

Harris, R. and Longley, P. A. 2002. Creating small area estimates of urban deprivation. *Environment and Planning A*, pp.1073-1093. doi: 10.1068/a34120

Kandt, J. and Longley, P. A. 2018. Ethnicity estimation using family naming practices. *PLoS ONE*, 13(8), p.e0201774. doi: 10.1371/journal.pone.0201774

Kandt, J., Van Dijk, J., and Longley, P. A. 2020. Family name origins and inter-generational demographic change in Great Britain. *Annals of the American Association of Geographers*, Advance Online Publication. doi: 10.1080/24694452.2020.1717328

Kershen, A. J. 2005. *Strangers, Aliens and Asians: Huguenots, Jews and Bangladeshis in Spitalfields 1660–2000*. New York: Routledge.

Lan, T., Kandt, J., and Longley, P. A. 2019. Geographic scales of residential segregation in English cities. *Urban Geography*. Advance Online Publication. doi: 10.1080/02723638.2019.1645554

Lansley, G. and Longley, P. A. 2016. Deriving age and gender from forenames for consumer analytics. *Journal of Retailing and Consumer Services*, 30, pp.271-278. doi: 1016/j.jretconser.2016.02.007

Lansley, G., Wen, L., and P. A. Longley. 2019. Creating a linked consumer register for granular demographic analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4), pp.1587-1605. doi: 10.1111/rssa.12476

Lomax, N. and J. Stillwell. 2017. *United Kingdom: Temporal change in internal migration.* In: Champion, T., Cooke, T, and I. Shuttleworth (Eds.). *Internal migration in the Developed World: Are we becoming less mobile?* London: Routledge, pp.120-146.

Mateos, P., Longley, P. A., and O'Sullivan, D. 2011. Ethnicity and population structure in personal naming networks. *PLoS ONE*, 6(9), p.e22943. doi:10.1371/ journal.pone.0022943

Mateos, P., Singleton, A., and P. A. Longley. 2009. Uncertainty in the analysis of ethnicity classifications: Issues of extent and aggregation of ethnic groups. *Journal of Ethnic and Migration Studies* 35 (9), pp.1437-1460. doi: 10.1080/1369183090312591

ONS. 2019a. Mid-year population estimates quality and methodology information. Online: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/annualmidyearpopulationestimatesqmi (Accessed 22.07.19).

ONS. 2019b. Local area migration indicators, UK. Online: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/migrationwithintheuk/datasets/localareamigrationindicatorsunitedkingdom (Accessed 10.12.19).

Reades, J., De Souza, J., and Hubbard, P. 2019. Understanding urban gentrification through machine learning. *Urban Studies* 56(5), pp.922-942. doi: 10.1177/0042098018789054

Smith, N. and Williams, P. (Eds.). 1986. *Gentrification in the city*. London: Unwin Hyman.

Van Dijk, J., Lansley, G., and P. A. Longley. 2020. Using linked consumer registers to estimate residential moves in the United Kingdom. Under review.