Residential Telephone Directories and Geodemographic Change in Britain

Nikki Tanu[*1], Maurizio Gibin [†1] and Paul Longley[‡1]

[1]Consumer Data Research Centre, Department of Geography, University College London,
Gower Street
London
WC1E 6BT

GISRUK 2023

Summary

Historical British telephone directories have recently availed themselves as disaggregate sources of population data for time periods not covered by other popular sources. Having devised a comprehensive processing pipeline to prepare these data for further analyses, this paper explores what they could uncover about British historical demography. For one, the differential uptake of telephones across space and time, owing to changes to its affordability and applications, reflects socioeconomic disparities. Being a data source that spans decades, it also lends its application to the study of intergenerational inequalities through the unit of family groups, as indicated by surnames that subscribers bear.

KEYWORDS: geodemographics, historical GIS, family names, telecommunications

1. Introduction

The recent digitisation and geocoding of historical telephone directories have provided researchers with the opportunity to re-examine the uptake of landlines in Britain quantitatively. However, more than merely document telephone ownership and the concomitant expansion of service networks, these directories also inadvertently serve snapshots of Britons from generations past, as a baseline capturing where they lived and what services they could afford. Where other major sources of recent historical information – like decennial censuses – are subject to disclosure rules and/or limited in frequency of data collection, digitised information in these directories can bridge an important gap and provide information at a granularity yet unseen for its temporal coverage. Moreover, beyond static profiling of places in Britain, telephone ownership at the household level could even afford the possibility of inferring patterns of intranational movement of its people. The potential for applications of the data is immense, although not without attendant uncertainties that must be acknowledged and, where possible, accounted for.

2. Data and Methodology

The dataset used in this research was obtained by the Consumer Data Research Centre (CDRC) in partnership with British Telecommunications (BT), comprising 1.6 million scans of individual pages from telephone directories spanning 1881 to 1984. In the interim, a processing pipeline has been developed to digitise these raw data (and undergoes continuous refinement) in which the following steps, described briefly, occur: text in the scanned images are first extracted using Optical Character Recognition. Then, the blocks of text are formatted into tables using Regular Expressions, with every

[*] n.tanu@ucl.ac.uk
[†] m.gibin@ucl.ac.uk
[‡] p.longley@ucl.ac.uk

row representing a unique telephone subscriber record. Finally, geographical coordinates are attached to records, similar to how Lan and Longley (2019) describe: this takes place through matching the address fields of each record against addresses in various databases, among which is the Ordnance Survey (OS) AddressBase.

Table 1 below provides a summary of the digitisation pipeline results for the earliest years, selected for their contemporaneity with decadal English censuses. Within each year, only the latest directory for London was used for analysis to mitigate double counting of records. Furthermore, because the directories contain both residential and commercial subscribers, a simple algorithm was written to differentiate the two categories; it becomes evident from **Table 1** that both the number of telephone subscribers and proportion of residential subscribers in London steadily increase with time.

Table 1 Overview of Record Numbers in London by Decade

Year	Latest Directory for London	Number of unique records	
		Total	Residential *(% of total)*
1881	Dec 1881	2,515	925 *(36.9%)*
1891	Oct 1891	9,776	3,883 *(39.7%)*
1901	Jul 1901	23,820	10,384 *(43.6%)*
1911	Jul 1911	105,826	70,894 *(67.0%)*
1921	Oct 1921	147,686	94,440 *(63.9%)*

A key initial objective of this paper is to generate a measure of comparison of telephone adoption rates across London. To this end, yearly data were analysed in conjunction with information from their respective contemporaneous census to give a Location Quotient for each registration district, shown below in Equation 1.

$$LQ_{rd} = \frac{\frac{N_{rd}}{H_{rd}}}{\frac{N_L}{H_L}} \tag{1}$$

where N_{rd} and N_L refer to the number of subscribers in a registration district and in London, and H_{rd} and H_L to the number of houses/households for the same geographical units, respectively

Changes in enumeration methods and in civil geographical boundaries meant that tallies of household counts were not readily available for all years of interest. Thus, for 1911 and 1921, they were instead estimated using figures from the preceding decade, proportional to the change in total population per district since then.

Another key objective of this research is to link the records of longstanding residential subscribers across years, so as to follow their life trajectories and understand their mobility across time. Ultimately, this ties in with the idea of a curated Linked Consumer Register, which aims to enable the accumulation of information on population structure at a disaggregated level (Lansley *et al.*, 2019) for a period for which other currently commonplace sources (including the census) do not cover: most of the 20th century.

Fuzzy string matching was used to achieve this, matching on two fields of information per subscriber – surname and address street name – and where additional address details (such as street number) were available, they were used for confirmation. This work is yet to complete, although the outcome of a trial on 1921 data is shown in **Table 2** below.

Table 2 Record Linkage Between 1911 and 1921, on a sample of 1,000 records from 1921

Matching Criteria	Proportion of Addresses (%)
Perfect Match (string distances = 0)	8.5%
Surname matched perfectly; address matched with string distance <=3	11.2%
Surname matched with string distance <=3; address matched perfectly	5.7%
Both surname and address matched with string distance <=3	23.5%

3. Discussion and Preliminary Results

3.1 Location Quotients

Fixed line telephony began as a technology prohibitively expensive to most of the population, which then grew to become more affordable and therefore widespread. **Figure 1** below depicts changes in the location quotient of telephone ownership among the registration districts between 1881 and 1921, from which we can draw preliminary conclusions about how spatial disparities of telephone ownership in London evolved. Districts are arranged in descending order of location quotients in 1921, with City of London and the Strand not having information displayed because their quotients are more than threefold higher than that of the next highest district (Kensington).

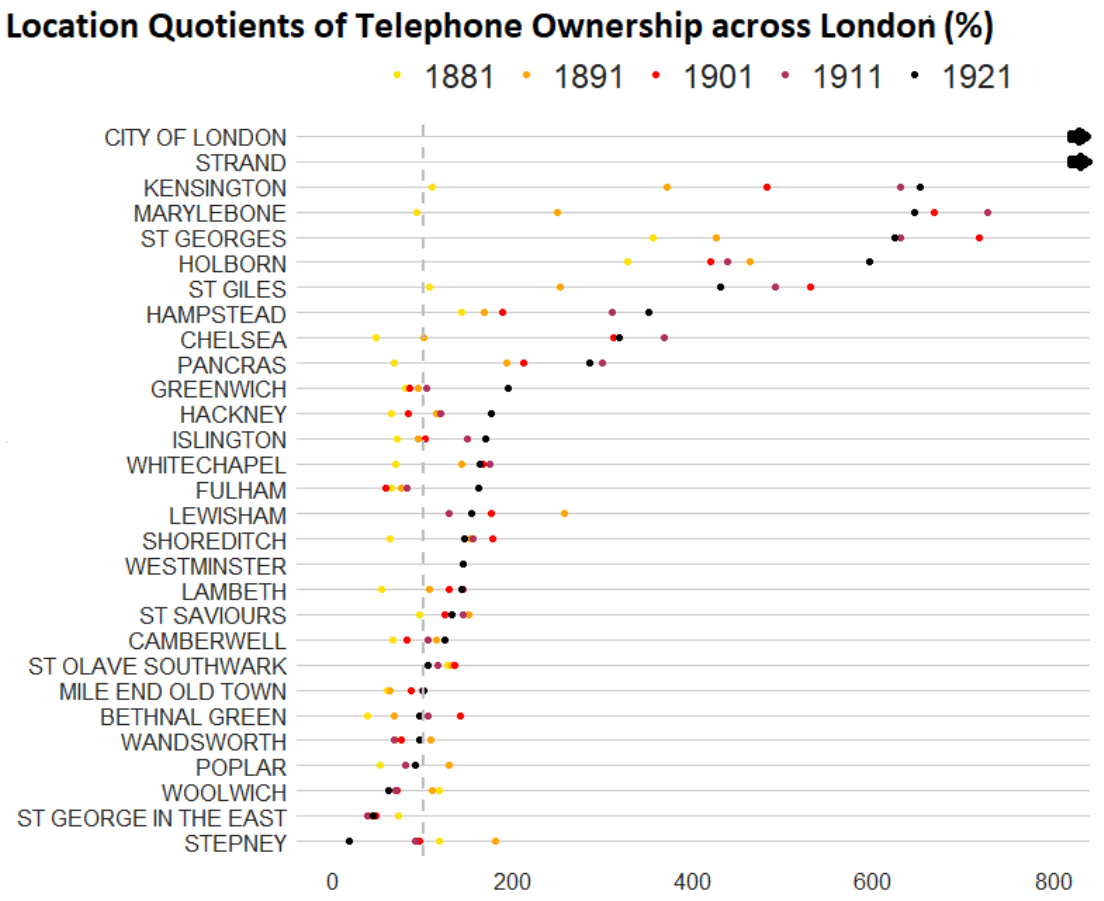

Figure 1 Change in the location quotients of telephone ownership across districts of London

(100 on the x-axis represents the average telephone adoption rate across London)

With few exceptions, the location quotients for a large majority of districts have increased from 1881 to 1921. At the top of **Figure 1**, a group of districts have evidently increased their telephone uptake rates by as much as 500% (Kensington and Marylebone). Conversely, it is worth noting that, in 1921, just 6 districts (from the bottom) remain below the London average in terms of telephone adoption; these districts, however, tended to be larger in population. This segregation in adoption patterns can also similarly be seen in **Figure 2**, which shows that the districts of highest location quotients begin in the centre of London in 1881 and continue to cluster in the city's northern half. Meanwhile, by 1921, districts with below-average telephone adoption rates remain almost exclusively in the East.

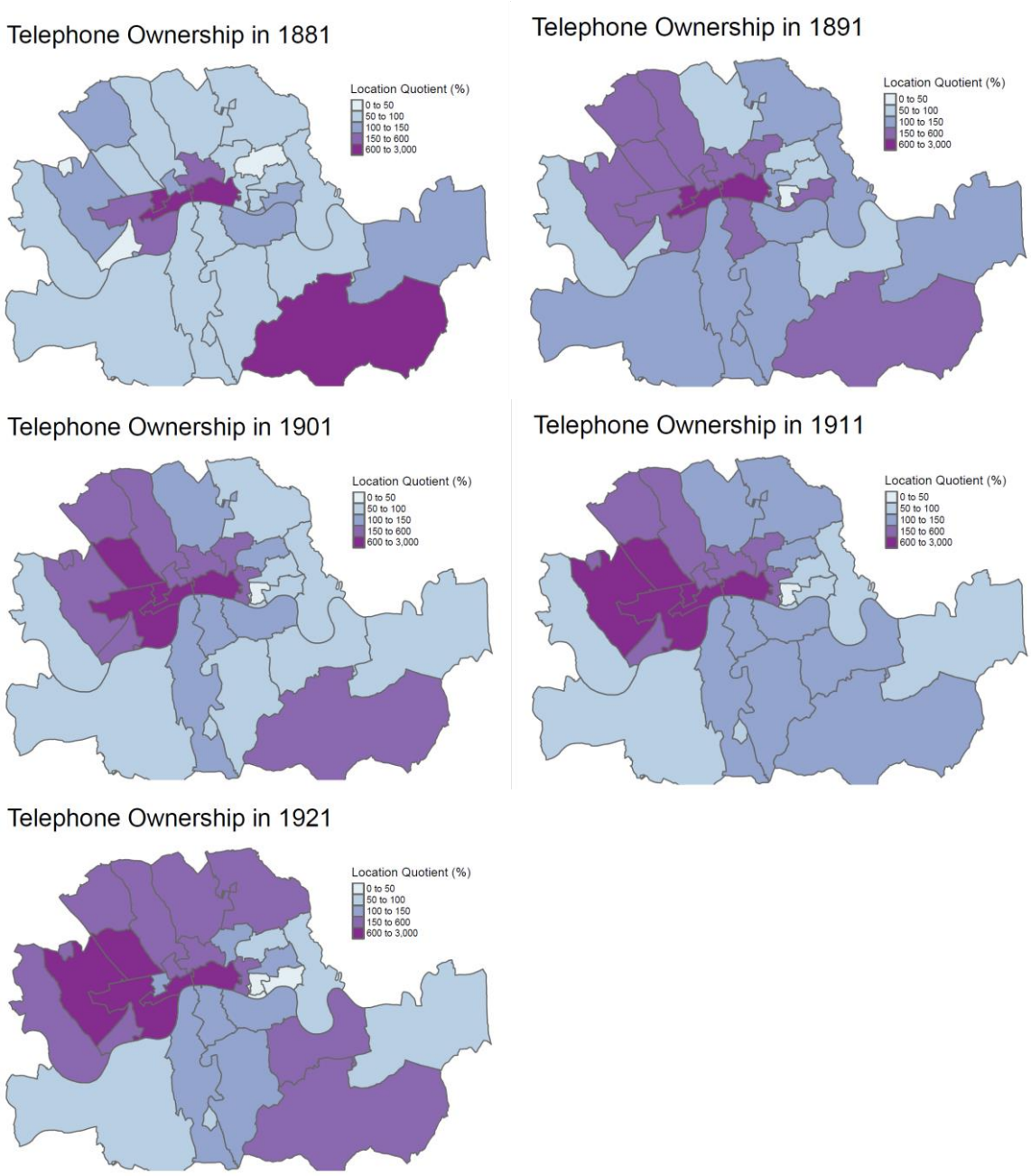

Figure 2 Maps of location quotients of telephone ownership across 5 different years

3.2 Representation of Family Groups

Another dimension of inequality in access to the telephone as an emerging technology pertains to social mobility. While data with which to evaluate social mobility in the early 20th century were scarce, surnames that telephone subscribers bear could be used as a rough indicator of geographical origin, whether they have migrated from their ancestors' lands and thereby the intergenerational inequalities that exist between peoples of different origins (Longley *et al.*, 2021). Since it was also earlier established that access to residential telephone subscriptions in the early years tended to correlate with wealth, it is a reasonable assumption that the most frequently reoccurring surnames in the early telephone directories may entail higher proportions of the wealthy. To examine this proposition, location quotients were re-calculated to include only subscribers bearing a surname within the top 10 of occurrence among all residential subscribers. **Figure 3** displays the results.

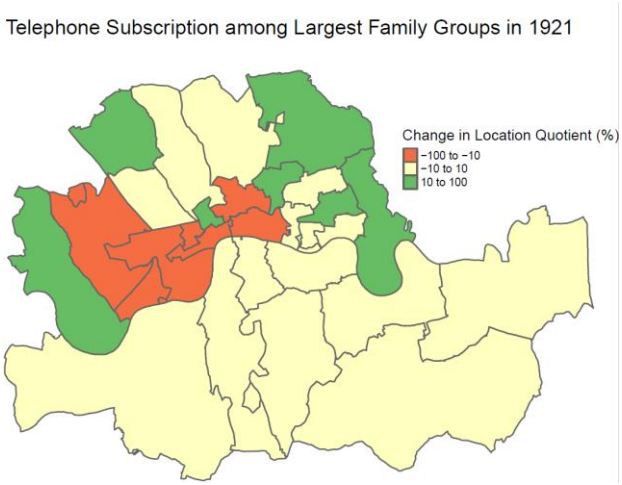

Figure 3 Change in location quotients when only subscribers bearing the 10 most frequently occurring surnames in the directory are considered

There appears to be overrepresentation of bearers of these surnames among subscribers in the outer districts of Northern London, while the adjacent inner districts tended to be underrepresented. Moreover, these districts correspond to areas of already higher location quotients in 1921 (**Figure 2**). In tandem with these observations, taking high rates of telephone adoption among a family group as a generalised marker of wealth, it could be assumed that the overrepresented districts contained wealthier residents.

4. Ongoing and Future Work

Following successful linkage of telephone subscriber records across years, this research aims to, through that, create a measure of individual mobility – of how long a subscriber has been resident at their current address. In turn, this enables the creation of aggregate metrics measuring average residency durations across geographical units that could be used vis-à-vis the existing location quotients. Certainly, this project also aims to extend the processing hitherto completed to directories in every decade up till the end of data coverage in 1984, as well as geographical coverage to other British cities that will serve to contrast London.

5. Acknowledgements

This work is funded by the UK ESRC Consumer Data Research Centre (CDRC) grant reference ES/L011840/1 and an ESRC Ph.D. studentship in partnership with BT Archives, which provided scanned image data under a collaborative agreement.

References

Lan T and Longley P (2019). Geo-Referencing and Mapping 1901 Census Addresses for England and Wales *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, 8(8), 320.

Lansley, G, Li, W & Longley, P A (2019) 'Creating a linked consumer register for granular demographic analysis', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182, 1587–1605.

Longley, P A, van Dijk, J & Lan, T (2021) 'The geography of intergenerational social mobility in Britain', *Nature Communications*, 12, 6050.

Biographies

Nikki Tanu is a PhD student at the UCL Department of Geography under the supervision of Professor Paul Longley and Professor James Cheshire.

Maurizio Gibin is a Senior Research Fellow at the UK Consumer Data Research Centre at UCL. His primary research interests are grouped around the capture, organisation, analysis and visualisation of geographically extensive datasets, including retail footfall, geodemographics, social media and historical archives.

Paul A. Longley is Professor of Geographic Information Science at UCL and Director of the UK Consumer Data Research Centre at UCL. His research focuses on the application of geographic information science, with a strong emphasis on the development and deployment of geo-temporal data infrastructures developed from Big and/or Open Data.