Is there a relationship between local businesses and population displacement?

This report presents the result of an investigation on the topic of gentrification, population displacement and their relation to local business. The research question posed was whether there is a relationship between the change in the composition of local business in an area, and the displacement of parts of the population to other areas. The investigation was carried out using data about local business and population employment by occupation group in London boroughs, spanning seven years from 2010 to 2016.

1 Context

A topic that has now and again made the newspaper headlines across the world, gentrification is described as 'the restoration and upgrading of deteriorated urban property by middle class and affluent people, often resulting in displacement of lower-income people' (American Heritage Dictionary 2004, as cited in Lees et al. 2008)

Whether the issue is viewed from the perspective of local authorities willing to prevent population displacement – or property developers and investors, willing to capitalise on under-priced land – it is of great interest to be able to predict gentrification and its associated phenomena before they gain force in a certain area, by creating early warning systems for gentrification (Chapple and Zuk 2016). Because of that, research has been conducted in recent years on the associations between various measurements of gentrification and other factors on a local level. Most notably, Hristova et al. (2016) present a novel approach of mapping a socio-spatial network using data from location-based social media to measure the characteristics of each area in terms of how much . On the other hand,Steif et al. (n.d.) use machine learning and spatial analysis of the surroundings of an area to predict rising housing prices. Meanwhile, a similar analysis involving data about local businesses seems to not have yet been attempted. Therefore, it is a potentially valuable undertaking to quantify the temporal relationship between population displacement and the composition of local business, in order to get closer to understanding if the former can be possibly predicted by changes to the latter.

It is a challenging task to quantitatively define gentrification and its related phenomena, like population displacement. It seems that the notion of social class can help defining groups that are most at risk when it comes to displacement. Roxanne Connelly et al. (2016) describes the 'consensus that occupations are central to understanding social stratification' (p. 1). Although more socio-economic status indicators are often used in social research, like the NS-Sec (Office for National Statistics n.d.), estimates from the Annual Survey of Hours and Earnings (Office for National Statistics 2017) show a clear stratification in the median wages between the major occupation groups, and therefore the higher susceptibility of certain groups to rising housing prices. It seems, then, that changes to the occupation profile of an area's population can be a useful measure for displacement.

2 Data

Two separate datasets were used in this investigation. This section describes the sources of the data, the format in which the data was initially obtained, and some of the issues associated with the datasets. Section 3 (Methodology) talks about the performed data transformations necessary for the specific techniques used in the analysis.

Data about local businesses was acquired from The Local Data Company (Local Data Company n.d.). The company has been collecting data from across London and the UK for around a decade. For this work, an extract from the company's database was obtained containing records about all high street businesses in London since 2010, with the data exported in November 2017. The record for each business contains, among others, information about the location, name, and a series of hierarchical classifications according to the company's own system.

Second, data was obtained from the Greater London Authority's London Datastore on employment from 2004 to 2016 in London boroughs (Greater London Authority n.d.), broken down by the SOC2010 (Office for National Statistics 2010) major occupation group, and gender.The administrative area of the City of London, with its low population, was excluded from the analysis altogether as most of the data was missing from the dataset. Fig. 1 contains a series of overview charts displaying the percentage breakdown of occupation types for each borough, from 2010 to 2016. The figures have been rounded to the nearest hundred.

3 Methodology

A significance level $\alpha = 0.05$ was used for all hypothesis testing in this work.

The R programming language and its libraries were utilised for the processing and visualisation of data throughout the work.

In order to answer the research question of this work (is there a relationship in time between local businesses and displacement), first it was necessary to establish a method of interpreting the chosen datasets in the context of gentrification. With the ultimate goal in mind of performing time series regression with a single dependent variable, the dataset describing each borough as a set of nine occupation groups had to be mapped to a single variable. In order to achieve that, an exploratory analysis was carried out using the `pvclust` R package. The package performs hierarchical clustering and assigns p-values to the resulting clusters based on multiscale bootstrap resampling, which repeatedly performs the clustering task on different size samples taken from the input data, in the process calculating approximately unbiased p-values (Suzuki and Shimodaira 2006) for the hypothesis test where the null hypothesis is that a certain cluster does not exist (Suzuki and Shimodaira 2017). A distance matrix needs to be calculated between observations, after which a selected clustering algorithm is applied. The agglomerative clustering method proposed by Ward (1963), and euclidean distance between observations were chosen as parameters for `pvclust`. First, clustering on London Boroughs based on 2016 data was performed to understand how the areas are different from and similar to one another when it comes to occupation group percentage breakdowns. Next, the distance matrix calculated for the first clustering task was transposed in order to carry out the clustering of variables (Chavent et al. 2011) to discover whether there are certain occupation groups whose proportion in some areas is visibly different (lower or higher) than in others. Based on the results, the nine occupation groups were aggregated into two groups, the lower-earning of which was taken to be the variable whose negative changes should represent population displacement from the area in the context of this work.

Subsequently, the business data was prepared for analysis by aggregating the data up to borough level and calculating the borough totals of 41 business categories for each year. Partitioned by borough and year, together with the data on the lower-paid occupation group, this constituted a cross-sectional time series dataset, as explained by Fortin-Rittberger (2014) – albeit with an admittedly quite low T (number of points in time) equal to 7.

That number was further reduced by two procedures: most importantly, as the time series exhibit clear trends, they are therefore non-stationary, which can result in a spurious correlation (Fortin-Rittberger 2014). Both the dependent and independent variables are then transformed through differencing, that is calculating their differences between subsequent years.

Finally, as the aim was to measure whether the changes in local business precede the population displacement, the multivariate time series of predictor variables was lagged by one year (Hyndman and Athanasopoulos 2013). This way, the second part of the analysis was aimed at determining whether a change in the local business in one year is a good predictor for change in low earning population displacement in the next. This finally reduced the number of time points to only 5.

Ultimately, Pearson's correlation coefficients were calculated between the predictors and the dependent variable in order to assess the feasibility of further analysis of the data.

4 Results

Fig. 1 shows trends in the occupation profiles of selected boroughs, with a split into the nine groups.

The dendrograms in Fig. 2 and 3 present the results of hierarchical clustering of boroughs on occupations and of the inverse – occupations on boroughs. The AU values (in red) represent the confidence interval, and the red rectangles highlight statistically significant clusters (an AU confidence interval 95% is equivalent to significance level 0.05).

Finally, Table 1 shows the top fifteen Pearson correlation coefficients between change in local business and population displacement.

5 Discussion

The initial exploratory analysis shows some quite strong differences between boroughs in terms of the composition of their population, and analysis of the first dendrogram (Fig. 2) confirms that intuition, as a statistically significant cluster is highlighted of predominantly rich areas (e.g. Richmond and Kensington). An interesting cluster on the right could potentially represent the areas where currently gentrification occurs (Southwark, Tower Hamlets) or has started a long time ago (Islington). It is arguably in these, largely urban, areas where the low-earning population would be most in danger of being pushed out of by rising prices.

Next, occupation types are clustered into two clear groups, showing that these two groups tend to cluster into certain areas of the city. This result lead to the decision to define the lower income

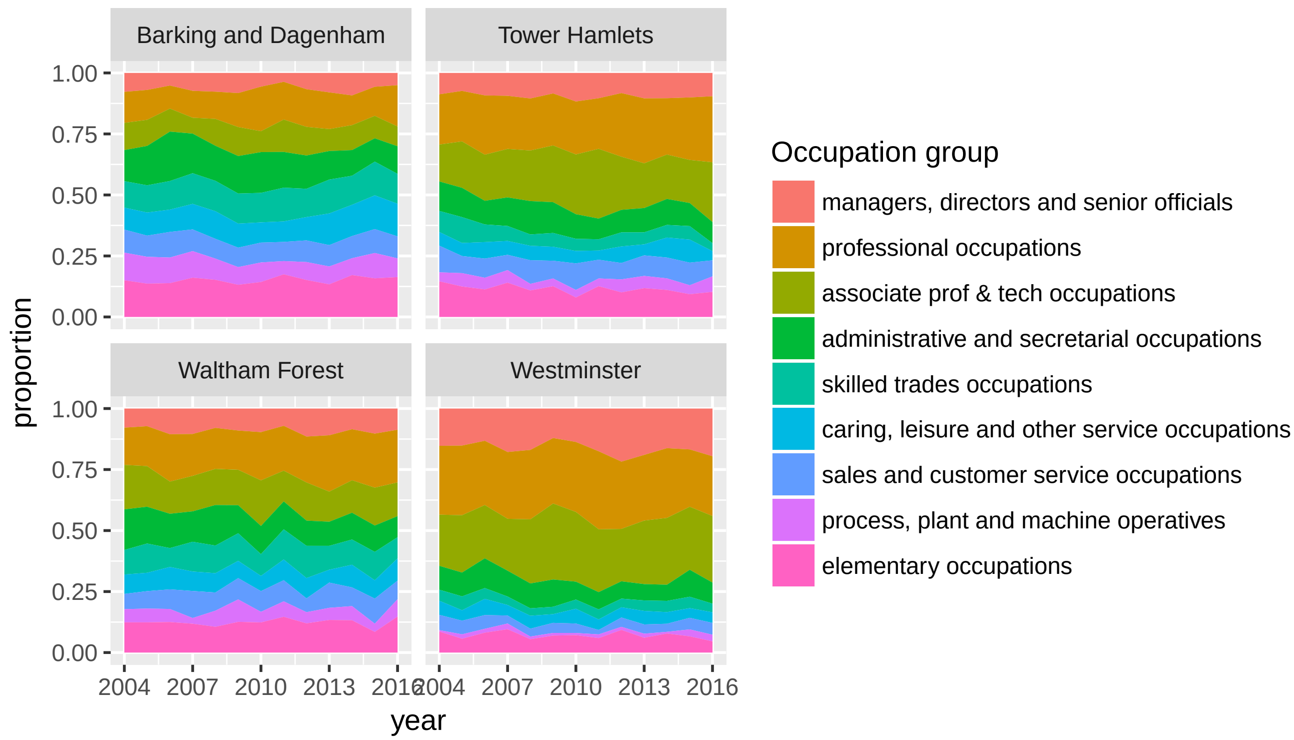

Figure 1: Occupation group percentage breakdown over time for selected boroughs

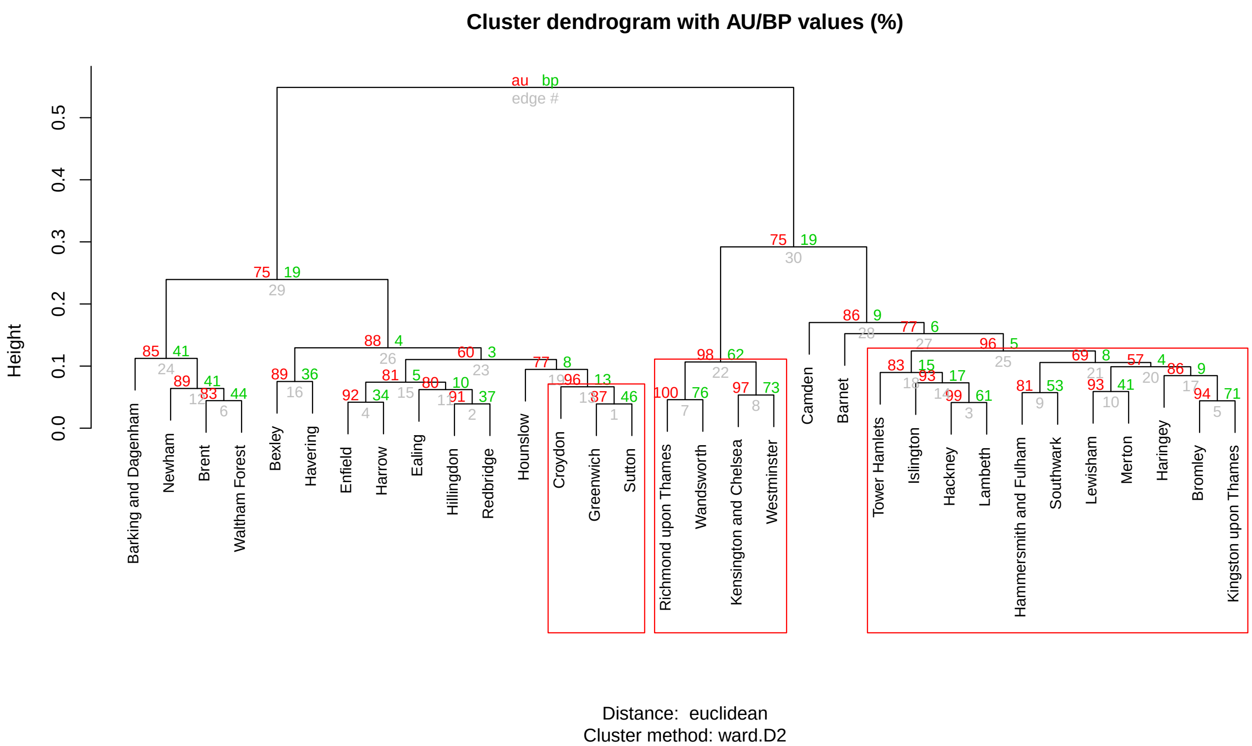

Figure 2: Result of hierarchical clustering of boroughs based on occupation types

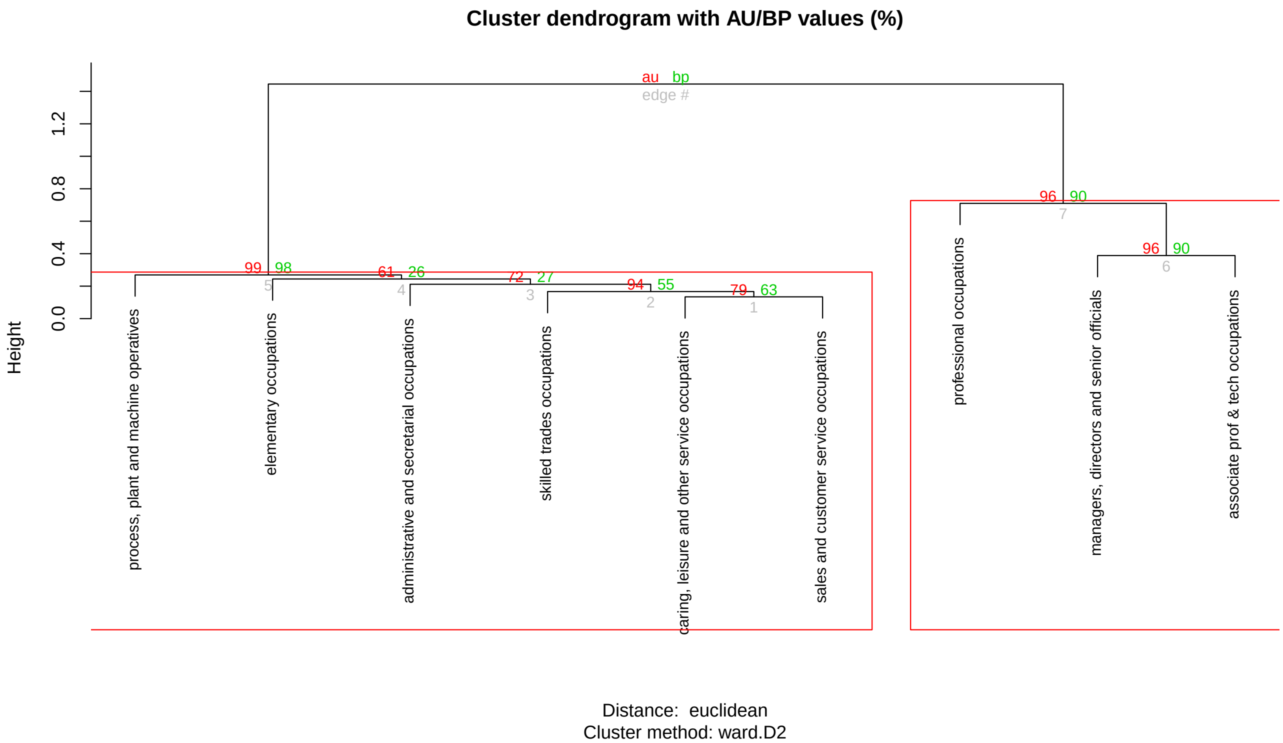

Figure 3: Result of hierarchical clustering of occupation types based on boroughs

Table 1: Top fifteen Pearson correlation coefficients between change in lower employment group and business categories

Independent variable	PC
Car & Motorbike Showrooms	0.2095019
Auto & Accessories	0.1679350
Petrol Filling Stations	-0.1485222
Household & Home	0.1231296
Discount & Surplus Stores	-0.1207302
Accommodation	0.1177125
Footwear	0.1167587
Entertainment	0.1162698
Estate Agents & Auctioneers	0.1151545
Locksmiths, Clothing Alterations & Shoe Repairs	0.1112151
Travel Agents & Tour Operators	0.1010843
Off Licences	-0.1007470
Electrical Goods & Home Entertainment	-0.0981370
Royal Mail Delivery Offices	-0.0842916
Restaurants	0.0799512

occupation group as consisting from all groups apart from the top three. The lower group change was calculated based on that.

The results of the Pearson correlation coefficient calculation show that there is some (weak) correlation between the change in lower income population and local business. Interestingly, businesses connected to cars had a positive correlation with the increase of lower income population, while the addition of discount and surplus stores was correlated with decreasing the lower income population (with a one year lag).

Finally, the discovered weak correlation with the explanatory variables does not in any way imply causality. Significant further research is needed in order to establish whether the data about local businesses could be used for prediction of gentrification. Ideally, data with better granularity and longer time coverage should be used for that purpose.

6 Conclusion

This work presented an analysis of the relationship between data on local businesses and population displacement, by applying techniques of hierarchical clustering with multiscale bootstrap resampling and preparation of time series data for analysis with a one year lag, followed by calculation of Pearson correlation coefficients. Interesting differences between London boroughs were discovered and the change in local business was discovered to be weakly correlated with some branches of business.

Overall, further regression analysis and testing for causality is needed to establish whether there is a strong relationship or even potential for prediction in case of local business and gentrification.

Word count: 1981

References

Chapple, K. and Zuk, M. (2016). Forewarned: The Use of Neighborhood Early Warning Systems for Gentrification and Displacement, *Cityscape* **18**(3): 109.

Chavent, M., Kuentz, V., Liquet, B. and Saracco, L. (2011). ClustOfVar: An R Package for the Clustering of Variables, *arXiv:1112.0295 [stat]* .

Fortin-Rittberger, J. (2014). Time-Series Cross-Section, *The SAGE Handbook of Regression Analysis and Causal Inference*, SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, pp. 387–408.

Greater London Authority (n.d.). London Datastore, https://data.london.gov.uk/dataset/employment-occupation-type-and-gender-borough.

Hristova, D., Williams, M. J., Musolesi, M., Panzarasa, P. and Mascolo, C. (2016). Measuring Urban Social Diversity Using Interconnected Geo-Social Networks, *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp. 21–30.

Hyndman, R. J. and Athanasopoulos, G. (2013). Forecasting: Principles and practice, OTexts, Melbourne, Australia.

Lees, L., Slater, T. and Wyly, E. K. (2008). *Gentrification*, Routledge.

Local Data Company (n.d.). Dynamic Location Intelligence, http://www.localdatacompany.com/.

Office for National Statistics (2010). *Standard Occupational Classification.*, Stationery Office, London. OCLC: 931106687.

Office for National Statistics (2012). Annual population survey (APS) QMI - Office for National Statistics, https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologie

Office for National Statistics (2017). Annual Survey of Hours and Earnings - Office for National Statistics, https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsu by-occupation.

Office for National Statistics (n.d.). The National Statistics Socio-economic classification (NS-SEC) - Office for National Statistics, https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioec

Roxanne Connelly, Vernon Gayle and Paul S. Lambert (2016). A Review of occupation-based social classifications for social survey research, *Methodological Innovations* **9**: 2059799116638003.

Steif, K., Mallach, A., Fichman, M. and Kassel, S. (n.d.). Predicting gentrification using longitudinal census data, http://urbanspatialanalysis.com/portfolio/predicting-gentrification-using-longitudinal-census-data/.

Suzuki, R. and Shimodaira, H. (2006). Pvclust: An R package for assessing the uncertainty in hierarchical clustering, *Bioinformatics* **22**(12): 1540–1542.

Suzuki, R. and Shimodaira, H. (2017). Pvclust: An R package for hierarchical clustering with p-values, http://stat.sys.i.kyoto-u.ac.jp/prog/pvclust/.

Ward, J. H. J. (1963). Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association* **58**(301): 236–244.