An analysis of the contribution of socio-economic measures to spatial variations in UK House Prices

Motivation

In 2017 housing in the UK is a major concern. Many younger people in their twenties and thirties spend an increasing amount of their income paying rent or on mortgage repayments (Citywire Money, 2017) and many get financial assistance from the 'bank of mum and dad' to place down a deposit on a property. (Financial Times, 2017). Rising house and rental prices outstrip the rise in real wages. Increasingly, young people opt to leave London. In the past two years, net emigration from London among 30-somethings has leapt by 25 per cent (Craw, 2016).

The many attractions of living in London, the capital of the UK, includes "the parks, the public transport, the schools; the restaurants, pubs, museums, and theatres; the general excitement, the sense that anything is possible." (Craw, 2016)

However, as house prices in London rise each year, increasingly more young people ask themselves, including this author, "Should I continue to live in London or move somewhere else?" (Crompton, 2015; Behr, 2015) The answer to this will be different for every person and depends on multiple and ever changing considerations.

This study seeks to glean insights from house price data collected by the land registry in 2016 and the most recent census data ("2011 Census data - Office for National Statistics"). The spatial relationship between house prices and various census variables, such as home ownership will be investigated. Additionally, spatial variations of house prices throughout the UK will be studied, with a focus on comparing the city of London with the rest of the country.

This paper seeks to ask:

- What socio economic variables are correlated with House prices?
- How does this correlation vary spatially?
- What type of people live in areas with higher house prices?
- Can spatial data inform a young adults decision on where best to live/work and buy a house?

Tasks and Approach:

Creating a data set from multiple sources

Before any analysis could take place in the investigation into the relationship between house prices and the UK population using a visual analytical methodology, it was necessary to a merge obtained UK house price data with its corresponding spatial attributes and census variables.

The House Price Data ("Price Paid Data - GOV.UK") is Price paid data for the year 2016 for each sale, that has the post code as its spatial attribute. However, the census data had local authority (LA) code as its spatial attribute.

Data Wrangle the House Price data.

To merge the House Price Data with the Census data, it was necessary for the data to share a common column. Therefore, post code column values within the house price data set were converted to its corresponding LA code. The pandas python package was chosen for its ease of use. This task was achieved through writing a function that used a postcode API("Postcodes.io"), reading and selecting the relevant LA code from a JSon object for each row and assigning this value to its corresponding post code. The House price data was then grouped and aggregated to achieve a mean house price for each local authority. This allowed for a meaningful visual representation mapped in a choropleth map.

Visually characterise spatial variation using cartographic techniques

For an initial investigation into Spatial variation in mean house prices, by Local Authority, a choropleth map was used with a gradient colour scheme that shows how house prices vary (in British Pounds) between LAs. LAs with more expensive mean house prices are shown in darker shades of green.

Local Authorities that contain urban areas have a higher population density than rural areas and therefore are smaller in size. A chloropleth map is useful in conveying human data because the map is changed from a conventional map to one where each area is weighted by the variable of interest to offer a 'corrected' visual saliency. The tmap package was used to map cartogram results on the map of the UK. In this case, a cartogram map shows the mean house price and is weighted by population in Figure 1 and Mean House Price in Figure 2. An alternative methodology to reduce giving too much visual saliency to large rural areas with low population densities is to plot the data by region see figure 3 and 4.

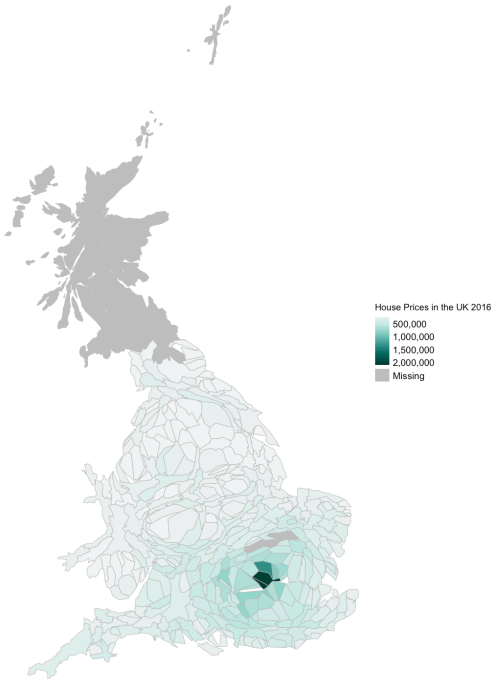

Figure 1.
Cartogram of house prices by local authority, weighted by population

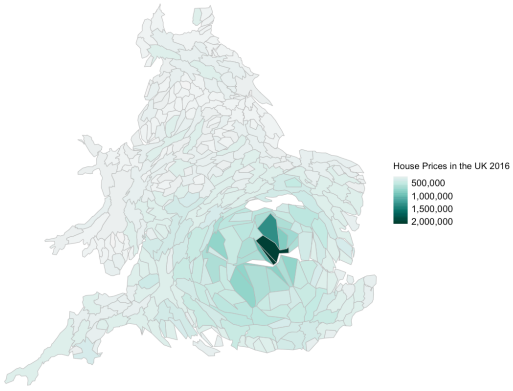

Figure 2.
Cartogram mean house price by local authority weighted by mean house price– it looks a lot more London centric.

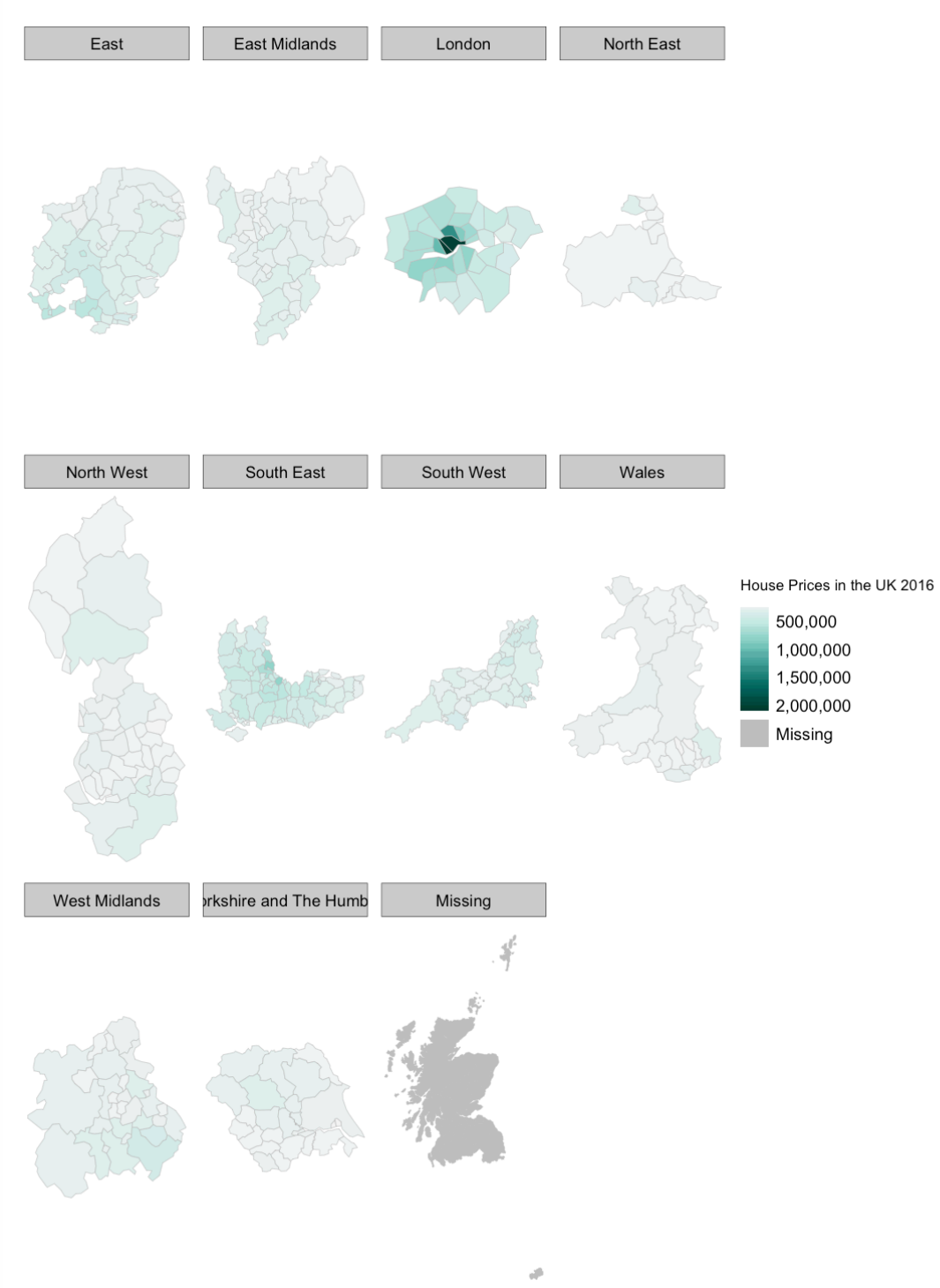

Figure 3
House Prices Regionally Distributed

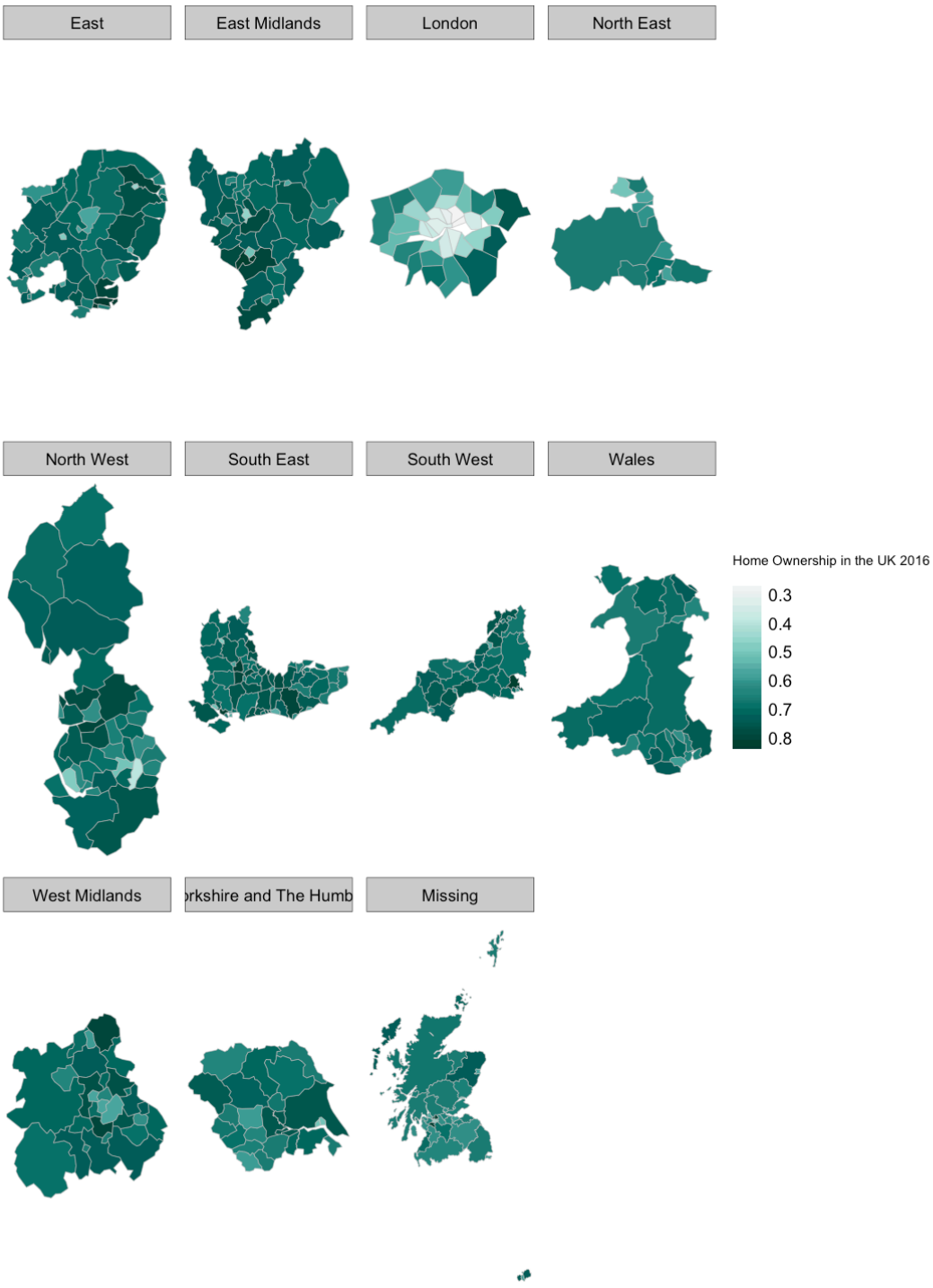

Figure 4
Home Ownership Percentages Regionally Distributed

Explore census variables and House prices correlation

Correlation Coefficients were calculated for census and house price variables in figure 5. This gave an indication as to which variables to explore further. Points in the scatterplot were coloured by region in figure 6, and coloured and sized by house price in figure 7.

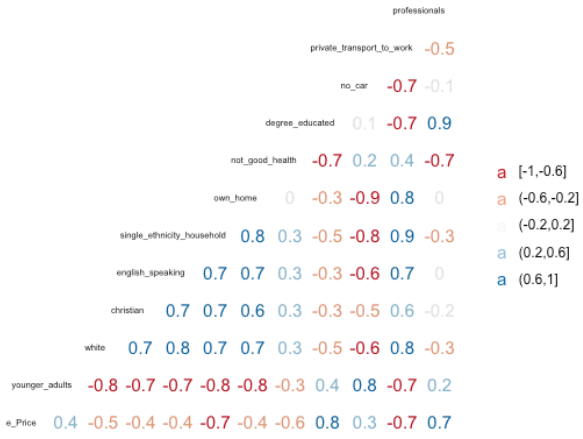

Figure 5
Correlation Coefficients

Evaluate explanatory variables using Variance Inflation Models (VIF) and build explanatory models using regression

A regression model is built to investigate any census variables that may have explanatory power in the value of house prices. In an effort to reduce collinearity, correlations were assessed using correlation coefficient matrices and Variance Inflation Factors (VIF). ("Variance inflation factor," 2017)

Explanatory variables were selected that were strongly correlated with mean house prices. These variables were examined for collinearity based on conceptual reasoning and VIF measures. A model was built based on minimising VIF values for the explanatory variables under consideration as well as maximising R squared.

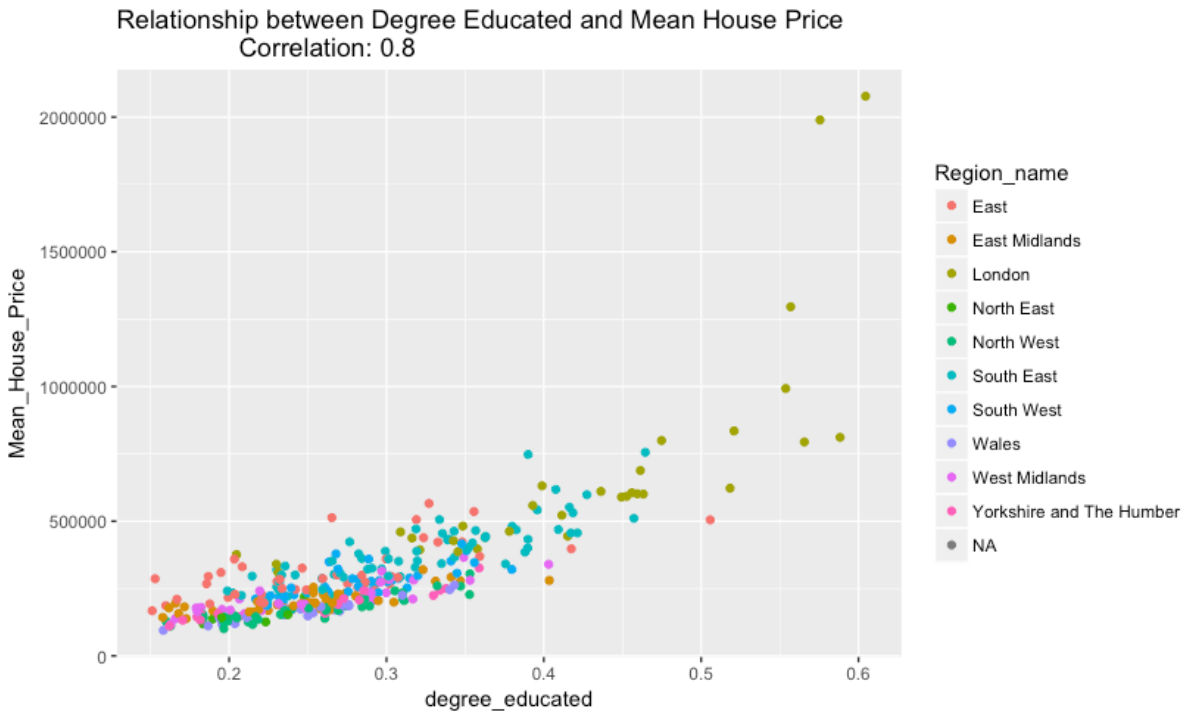
Relationship between Degree Educated and Mean House Price
Correlation: 0.8
Mean_House_Price
2000000
1500000
1000000
500000
0
0.2
0.3
0.4
0.5
0.6
degree_educated
Region_name
East
East Midlands
London
North East
North West
South East
South West
Wales
West Midlands
Yorkshire and The Humber
NA

Figure 6

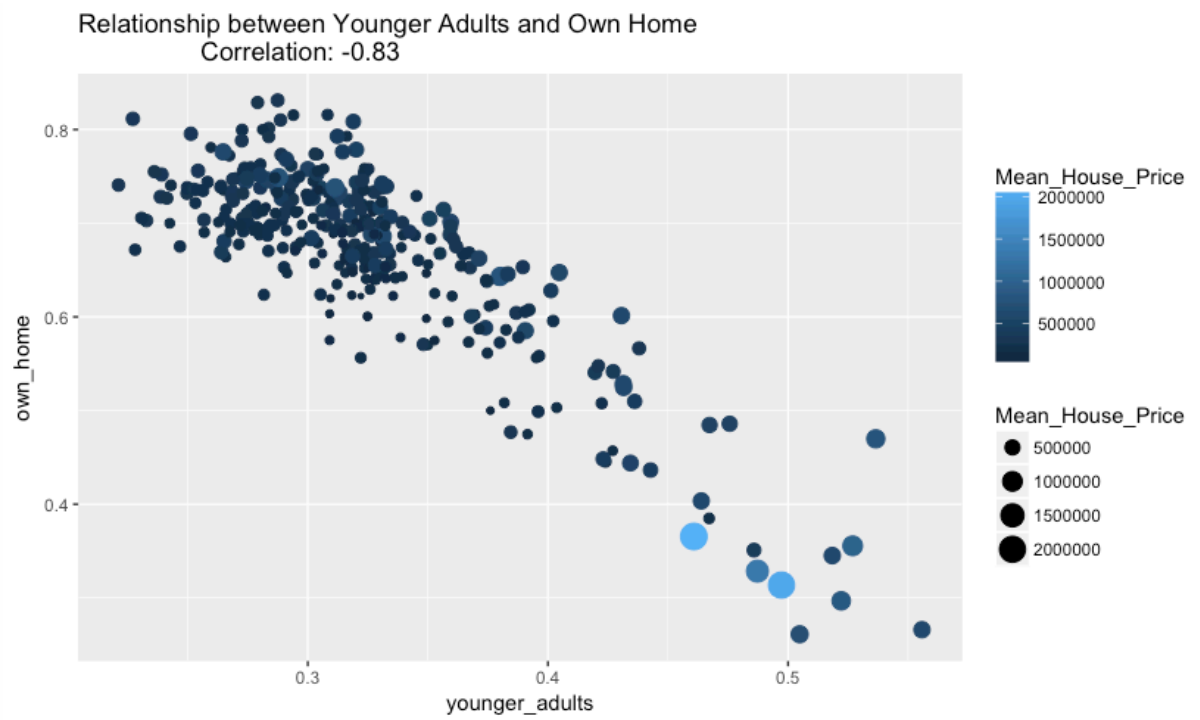

Figure 7

Analytical Steps

Spatial differences in House prices

By observing Figure 1, 2 and 3, a one can see that House Prices are significantly higher in or around London. This contrasts with the rest of the country. Unfortunately, at time of writing, house price data couldn't be obtained for Scotland. By comparing Figure 1 and 2 visually, one can see how much more housing wealth is concentrated in London relative to the population.

Spatial differences in home ownership

In Figure 4 one can see that home ownership is higher outside of London. Additionally, white spots in the north west indicate that home ownership is low also in major cities. This raised a question, to what extent is house prices correlated with home ownership. Maybe home ownership is low within major cities for other reasons.

Correlations between Census Variables and House Prices

In Figure 5, one can see the Pearson correlation coefficients across selected census variables within LAs and mean house prices within LAs. It's interesting to note that, home ownership is negatively correlated with house prices but only at a value of -0.4 which is not a strong correlation. On the other hand, variables of note that are positively strongly correlated with House Price are degree educated with a value of 0.8 and professionals 0.7 respectively.

Additionally, private transport to work and single ethnicity household were strongly negatively correlated with mean house price with a value of -0.7 and not good health with a value of -0.6.

It was postulated that there would be perhaps some collinearity between degree educated and professionals. This collinearity was tested using a VIF.

Explore Explanatory Variables on a Scatter Plot

In Figure 6 the relationship between degree educated and house prices, the most highly correlated variables were plotted on the scatter plot. Here one can see how London areas, represented with dark yellow dots are outliers in terms of degree educated and House prices.
Is there one that causes the other? Maybe London itself (a hidden attractiveness factor) attracts degree educated people which are more economically productive and therefore raise house prices, or maybe London already has high house prices which makes people more wealthy and therefore more likely to have the means to make it to higher education which causes a higher proportion of degree educated people. Or maybe it shows that a degree will increase one's chances of having a more valuable house if they move to London, since they will have higher salaries for the degree educated? However, a variable of interest to this paper is the role of age. How does one's age affect this interplay between Housing and degrees and jobs.
In Figure 7, the relationship between younger adults and home ownership is strongly negatively correlated. This is to be expected. However, of interest is the value of the house prices. The scatter plot circles are coloured and sized by House price variations. The lighter blue larger circles tend to be in areas where there are lower proportion of people who own a home and in areas with a higher proportion of younger people. One might expect in an area of lower home ownership and younger people, lower house prices prices. But in the case of London, there is lower home ownership, younger people and yet higher house prices. Why is this? Maybe buy to let investment has a role as well as a very complex interaction of factors. This paper will build a simple model to see if some of the census variables can predict house prices.

Building a model

A model, keeping Mean House Prices as a dependent variable, was built by examining VIF values and seeing which variables had significant explanatory power. This was done by removing and adding explanatory variables and recalculating R-squared. A R-squared value of 0.68 was achieved by using degree educated, own home, and not good health. By itself degree educated had R-squared value of 0.63, which was the highest of all the explanatory variables that, evaluated as the sole explanatory variable.

Findings

This paper sought to understand what socio economic variables are correlated with Mean House Prices and how do these variables vary spatially.

Figure 1 2 and 3 show how there are higher House prices in London and more wealth in London. Figure 4 shows how home ownership is low in London and other major cities. Figure 5 shows which socio economic variables are correlated with House prices. Notably younger adults were not strongly correlated with House Prices. However, degree educated and professionals were positively correlated with House Prices. Figure 6 shows this relationship between degree educated and House Prices on a scatter plot by region and reveals visually the

outliers of highly educated and high House Prices. This suggests again the uniqueness of London. Figure 7 shows that more younger adults are in areas of higher house prices and lower home ownership.

A linear model found that the best explanatory variables for House prices across LAs is degree education. Additionally, home ownership, and not good health also are negatively correlated with house prices and contribute to the calculation of R-squared.

Critical Reflection

It was interesting to see just how much housing wealth is centred in London. It shows the divide between London and the rest of the country. People in London are wealthier.
Does that mean young people should leave London or not?
Intuitively it makes sense that given that house prices are cheaper elsewhere maybe it makes sense to move out of London. Alternatively, house prices are correlated strongly with degrees and professionals. Maybe that means that it is better to stay and rather try to get a good job that is a professional job or requires a degree.

However, to answer the question, whether spatial data can inform a young adults decision on where best to live/work and buy a house, the author believes that more data is needed. Specifically, it would be useful to have more up to date census data. Additionally, time series data could help identify trends and causations. Further work could include looking at house prices over time to see what types of people benefit financially from living in areas of high house prices.

Further questions:

Is there evidence to suggest that house prices are high in areas where there are certain characteristics? Maybe it would be unreasonable for a young person to try to buy in an area in which house owners are much older and bought their house many years before at an affordable price. Maybe the structure of the UK housing market is such that in expensive areas, those who don't own a house must rent indefinitely as house prices rise?
If the housing market was made up of owners and renters, prima facie, one might expect home ownership to be mildly or even non-correlated to house price distribution across the UK.
However, as property in London is perceived as a safe asset, increasingly foreign investors buy property in London, pricing many people out of the housing market. This could explain the low home ownership rate.

Bibliography

2011 Census data - Office for National Statistics [WWW Document], n.d. URL https://www.ons.gov.uk/census/2011census/2011censusdata (accessed 1.4.17).
Behr, R., 2015. Goodbye London: why people are leaving the capital. The Guardian.
Craw, D.W., n.d. The great exodus: the net number of 30-somethings leaving London just leapt by a quarter [WWW Document]. CityMetric. URL

http://www.citymetric.com/politics/great-exodus-net-number-30-somethings-leaving-london-just-leapt-quarter-1959 (accessed 12.27.16).
Crompton, E., 2015. 13 reasons you should move out of London. Metro.
Postcodes.io [WWW Document], n.d. URL http://postcodes.io/ (accessed 1.4.17).
Price Paid Data - GOV.UK [WWW Document], n.d. URL https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads (accessed 1.4.17).
The rise of the 35-year mortgage [WWW Document], n.d. . Citywire Money. URL http://citywire.co.uk/money/the-rise-of-the-35-year-mortgage/a730398 (accessed 3.27.17).
Variance inflation factor, 2017. . Wikipedia.