Exploratory Spatial Data Analysis of House Prices in London

Introduction

Property prices in London are given repeated attention in newspaper articles in the UK, with a focus tending to be placed upon the fact they are much higher than other places, constantly growing, and largely unaffordable for the majority (e.g. Milligan, 2015; Bloomberg, 2017; Fraser, 2017; Gray, 2017; Scott, 2017). In addition to a distinct disparity between prices in the capital compared to those throughout the country, a pattern can be noticed within London. This causes property prices in some boroughs being significantly higher than others. Furthermore, it has been found that house prices in London can be up to 14.5 times the earnings of an average resident, resulting in many places being unaffordable (Fraser, 2017).

Consequently, this report aims to explore and model the spatial distribution of house prices across London and comment on the disparity between boroughs. Then combine the house price results with household income data to see if there is a correlation. These aims lead to the follow research questions:

1. Is there a spatial pattern for house prices throughout London?
2. Do any spatial patterns lead to house price disparities between boroughs?
3. Is there a spatial pattern within individual boroughs?
4. Is there a correlation between household income and house price?

Data

The data concerning property sales is from the HM Land Registry. Along with the price, the property postcode is included, meaning each sale can be combined with Ordnance Survey's Code Point Product and mapped. However, a number of limitations were observed.

Firstly, there were gaps influencing its completeness. For example, there were no property sales provided for the borough of Westminster and 448 points did not have postcode data. Initially, many more points were not being mapped. However, this was due to them not joining with the postcode table due to its inconsistent spacing (e.g. KT1

1AA and SE270QF), which was overcome by reformatting the postcodes to eliminate spaces.

Secondly, the points in the prices paid data are not exclusively houses, resulting in properties included with extremely high prices and the results to not be representative of only house prices. The range for Hillingdon was initially the highest amongst the boroughs, by a large degree (Figure 1). After creating histograms (Figure 2), it was found these results were skewed by one property, discovered to be the Sofitel hotel at Heathrow Terminal 5 sold for £330,000,000.

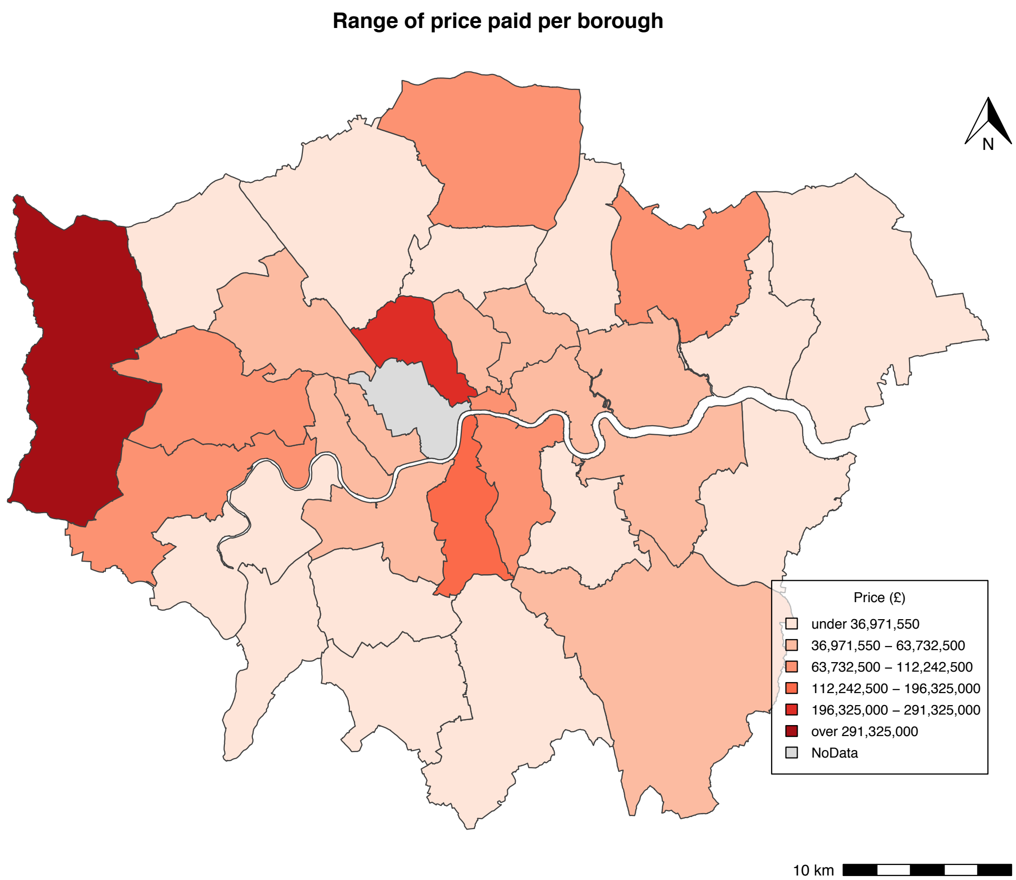


Figure 1: *The range for property price paid before the most expensive points were removed. Hillingdon can be seen on the far left as the borough with the largest range.*

Furthermore, average house prices by borough data was downloaded from the London data store and compared with the prices paid data (Appendix A). The average house price data excludes sales below £1,000 and above £20m. This revealed a significant difference in values, which may have been due to the prices paid data including properties other than houses. Consequently, any property with the sale price above £50m was given a N/A value.

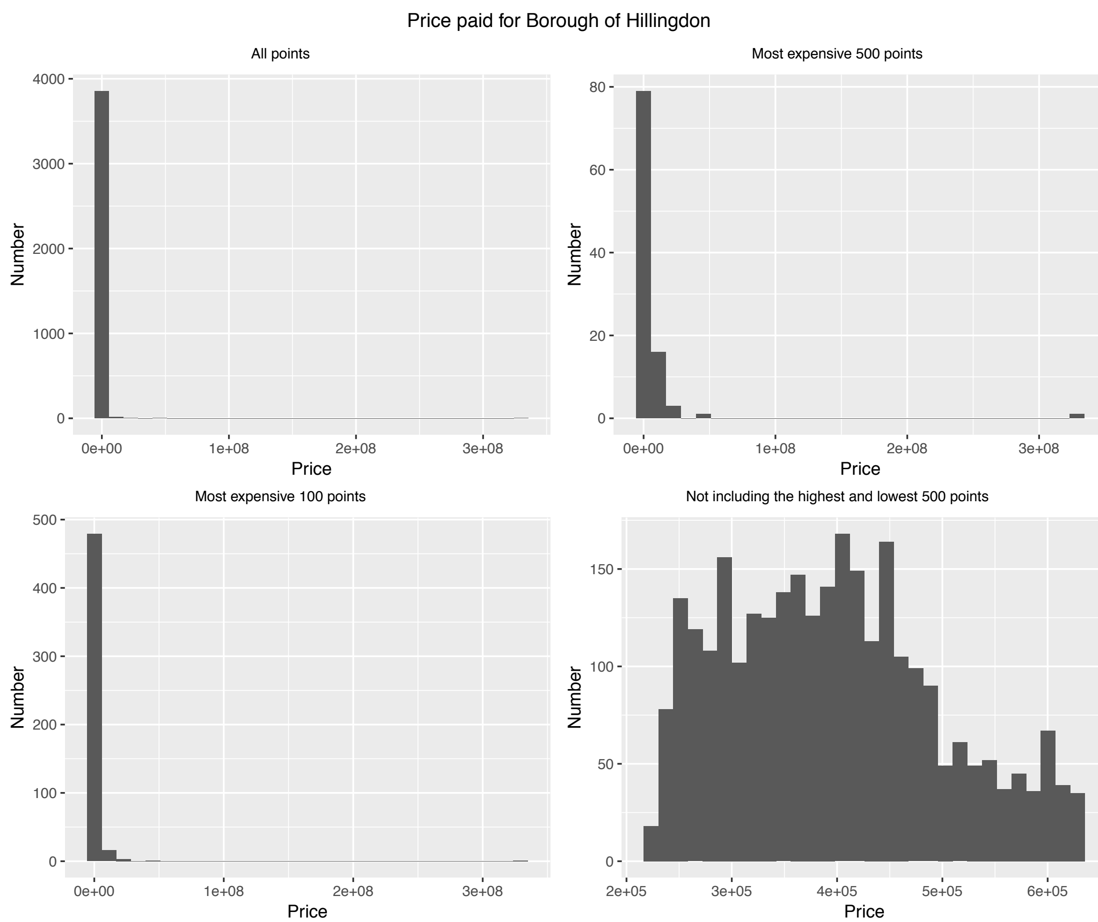
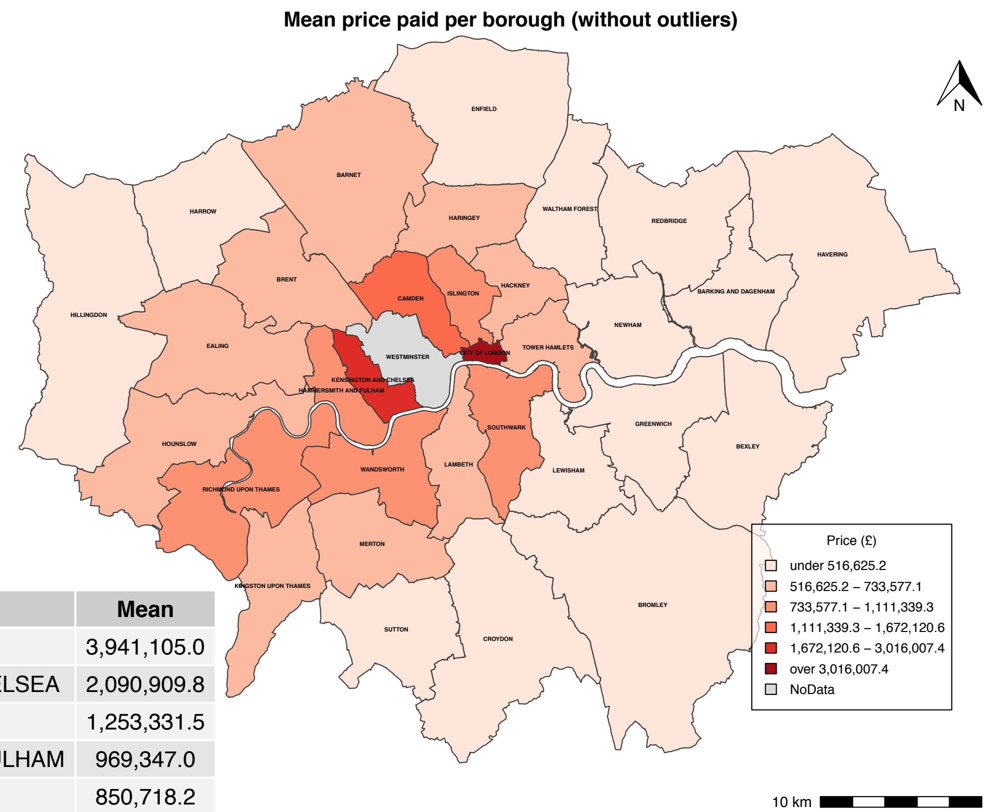


Figure 2: *Four histograms showing the distribution of points for the borough of Hillingdon. As seen in the first three histograms, there is an outlier result (the hotel at Heathrow terminal 5) that distorts the range. In the histogram on the bottom right, the highest and lowest 500 points were removed to see how the distribution would be affected. It can be seen that this data is more evenly distributed and therefore more representative of exclusively house sales.*

<u>Analysis</u>

Firstly, the mean price paid per borough was calculated and mapped to show the spatial distribution (Figure 3). This shows a general pattern whereby the mean value for each borough decreases the further out it is located from the centre of London. Interestingly, the five boroughs with the highest mean prices, which can be seen in the table (Figure 3), are clustered in the centre. Meanwhile, the three boroughs with the lowest mean prices paid, Barking and Dagenham, Bexley, and Havering, respectively, are clustered on the far east.

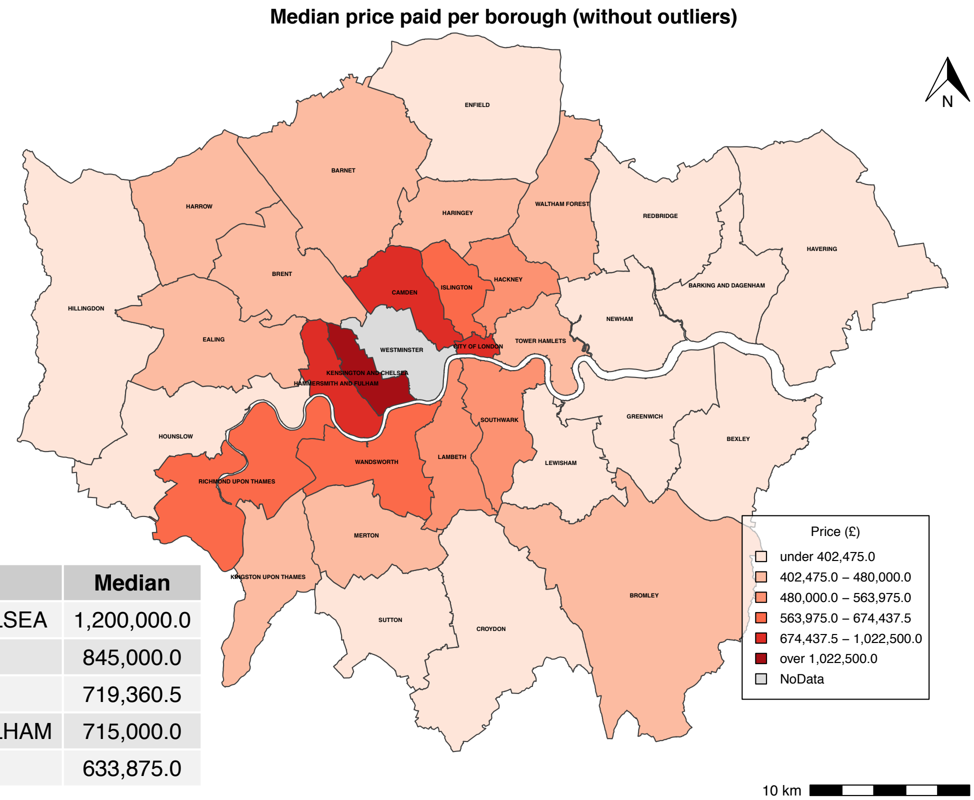


NAME	Mean
CITY OF LONDON	3,941,105.0
KENSINGTON AND CHELSEA	2,090,909.8
CAMDEN	1,253,331.5
HAMMERSMITH AND FULHAM	969,347.0
ISLINGTON	850,718.2

Figure 3: *A choropleth borough map of London showing the mean price paid per borough and a table showing the five highest mean prices.*

The median price per borough (Figure 4) shows a similar pattern, with the highest values concentrated towards the centre. Although in a different order, the highest three median values are the same as for mean: Kensington and Chelsea, City of London and Camden. Similarly, the lowest three median values are the same.

The median map shows a more even distribution of values within each price bracket than the mean map. Additionally, the highest median values are less extreme. This may be because the median value is more representative of the data in this case, as any exceptionally high or low values have little influence on the outcome. Despite manually removing the points with a price of over £50m, there are likely still high or low values present in the data, which may have influenced the mean but not the median.

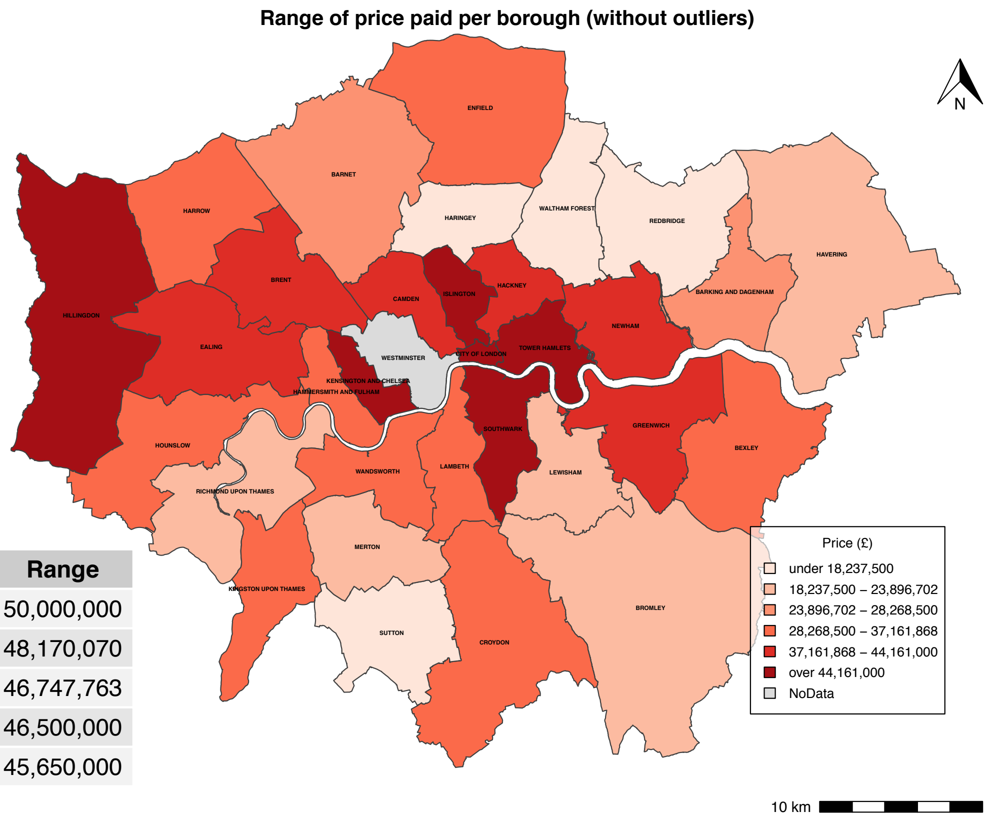


NAME	Median
KENSINGTON AND CHELSEA	1,200,000.0
CITY OF LONDON	845,000.0
CAMDEN	719,360.5
HAMMERSMITH AND FULHAM	715,000.0
WANDSWORTH	633,875.0

Figure 4: *A choropleth borough map of London showing the median price paid per borough and a table showing the five highest median prices.*

The range of price paid per borough was created to show the variation between the prices paid (Figure 5). There is a less obvious spatial pattern for the range, although the highest boroughs are concentrated towards the centre, with Hillingdon an exception. The three boroughs with the lowest range values, Haringey, Redbridge, and Waltham Forest, respectively, are located in the north.

The lowest range value, for the borough of Haringey, is £12,100,000. This means that for every borough within London there is a large disparity between the lowest and highest prices paid even after removing the highest, outlier points above £50m.



NAME	Range
TOWER HAMLETS	50,000,000
SOUTHWARK	48,170,070
CITY OF LONDON	46,747,763
ISLINGTON	46,500,000
HILLINGDON	45,650,000

Figure 5: *A choropleth borough map of London showing the range of price paid per borough and a table showing the five highest median prices.*

After joining the prices paid data with the postcode coordinates, each point was mapped and coloured by value (Figure 6). Additionally, the 500 most and 500 least expensive properties were mapped (Figure 7). Once again, this shows a spatial pattern similar to before, with high value properties clustered towards the centre of London and lower value properties dominating the outskirts.

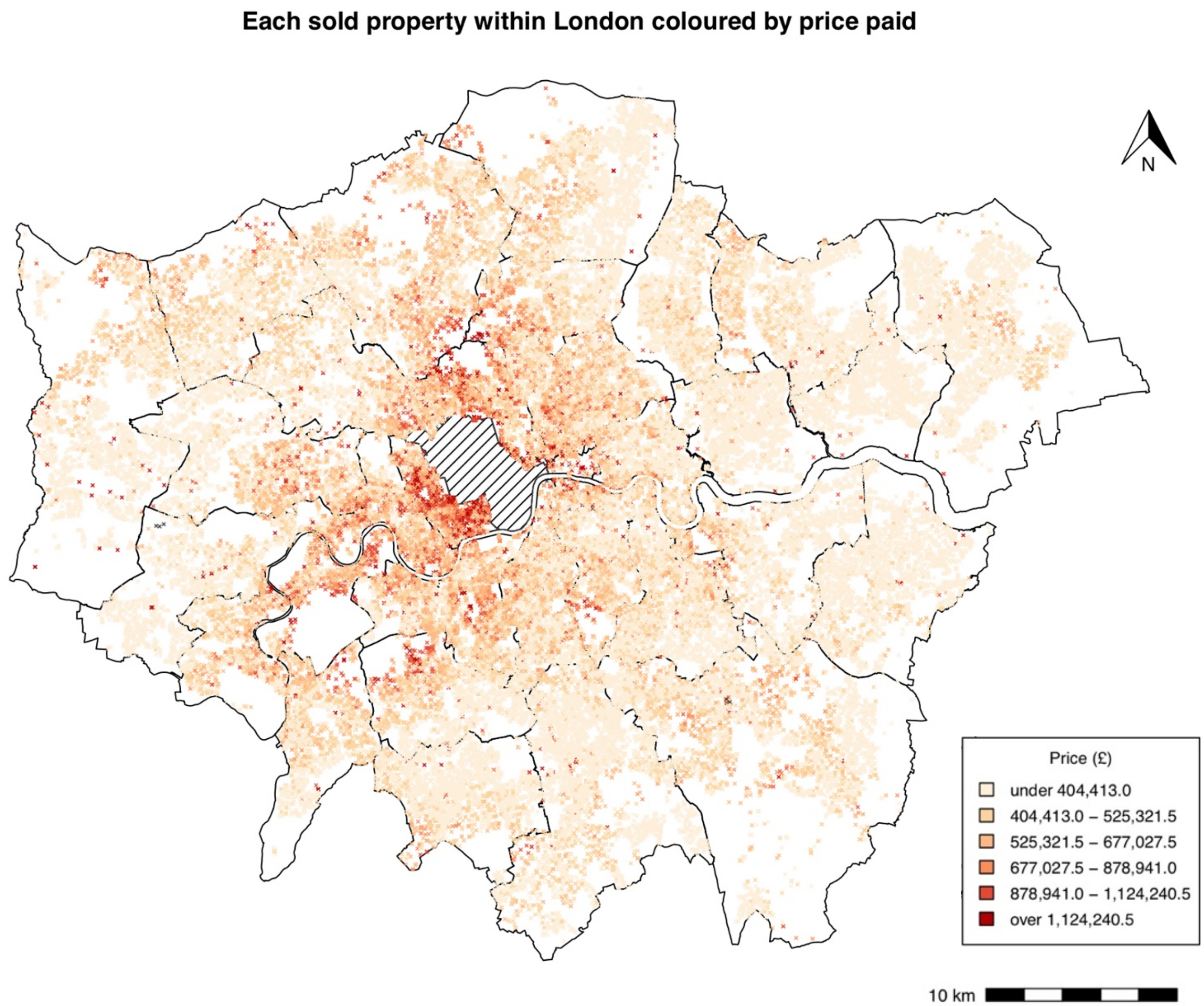


Figure 6: *Every property point plotted and coloured by price paid.*

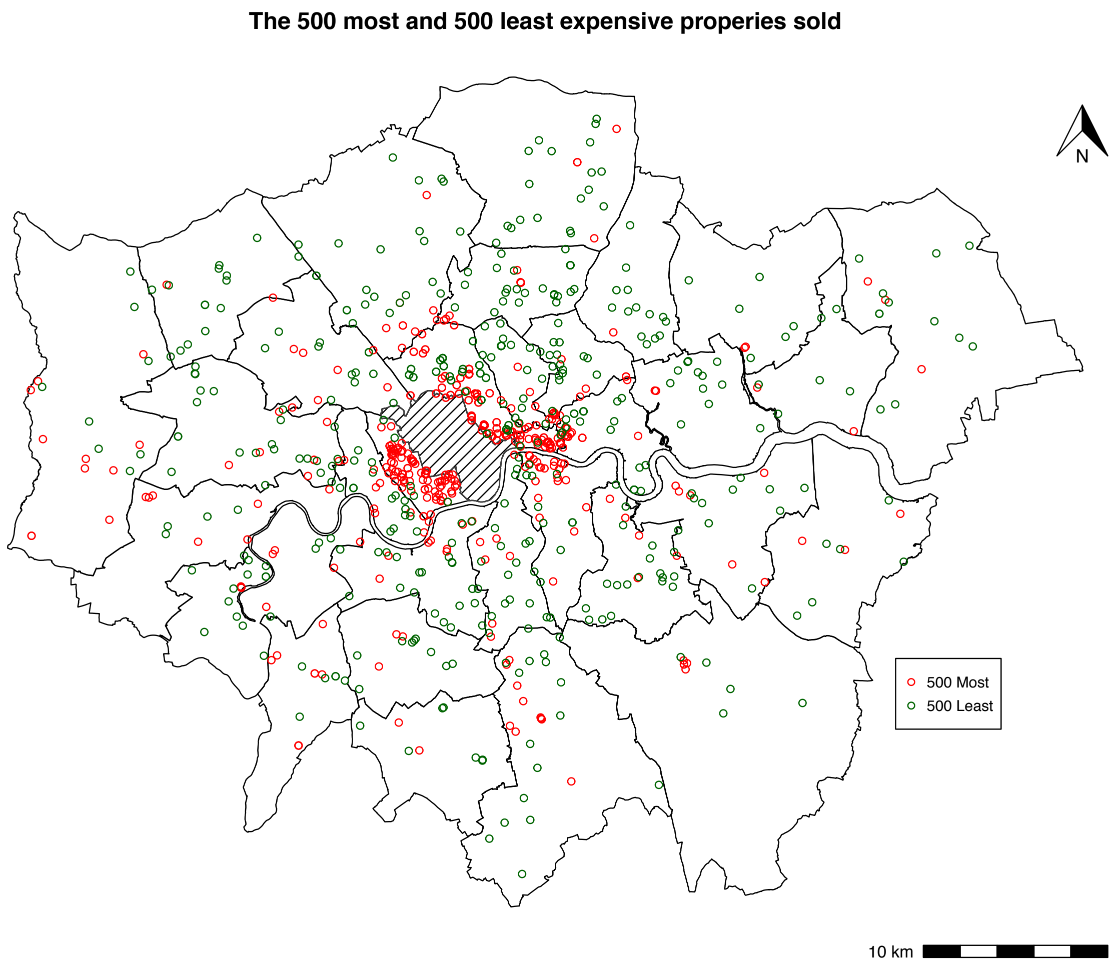


Figure 7: *The 500 most and 500 least expensive properties sold.*

In order to assess spatial patterns within a borough, the boroughs of Southwark and Enfield were selected to provide central and outskirts boroughs (Figure 8). For Southwark, there is a clustering of properties, many that are high price in the North, where Bermondsey is located, which is towards the centre of London.

Within Enfield, the majority of points are within the lower price brackets. The majority of the lowest values are towards the east and the higher prices towards the west.

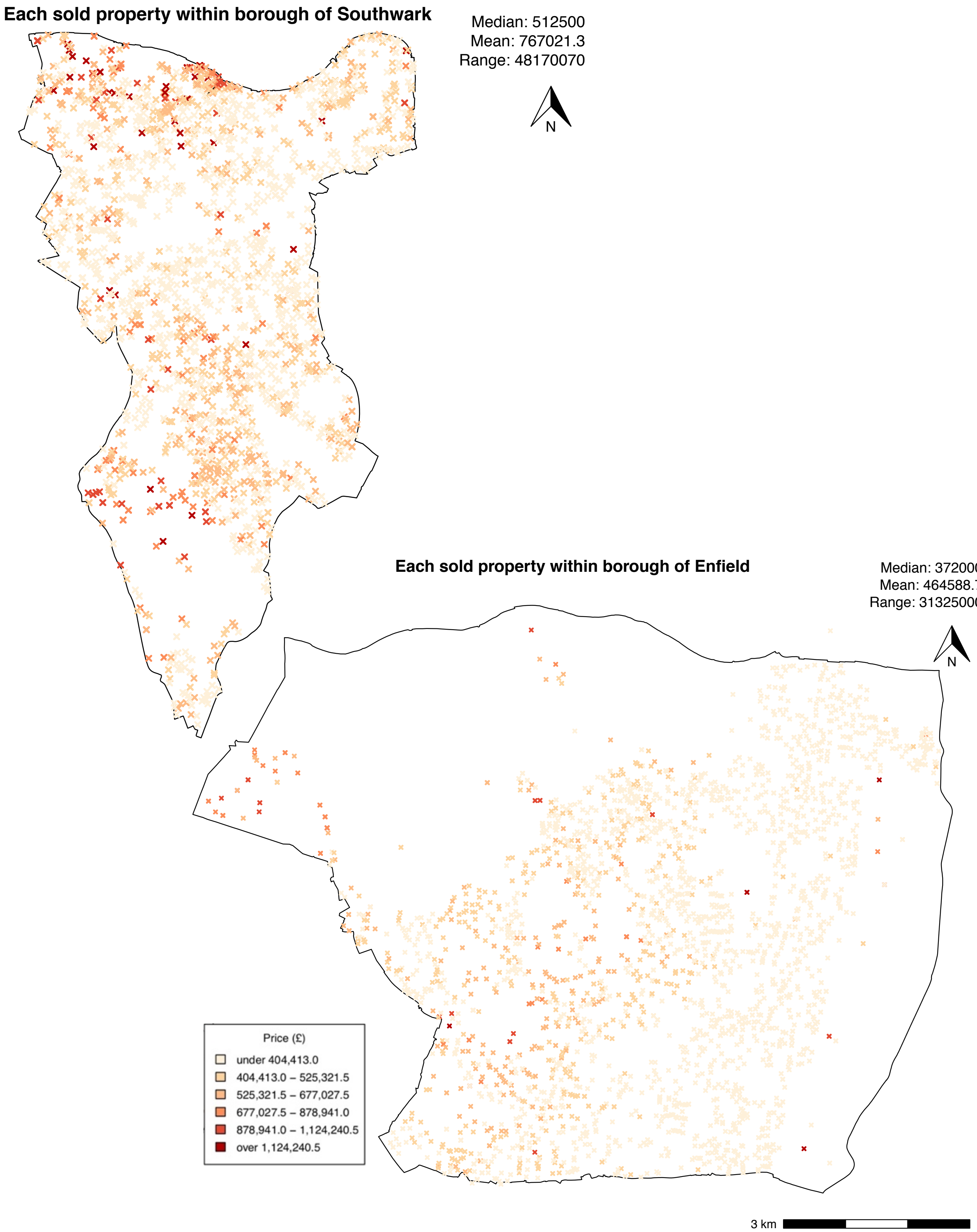


Figure 8: *The properties points for the boroughs of Southwark (top) and Enfield (bottom)*

Next, each borough's median household income was plotted against the median house price and the correlation calculated (Figure 9). The Pearson correlation was 0.78, demonstrating a strong positive relationship.

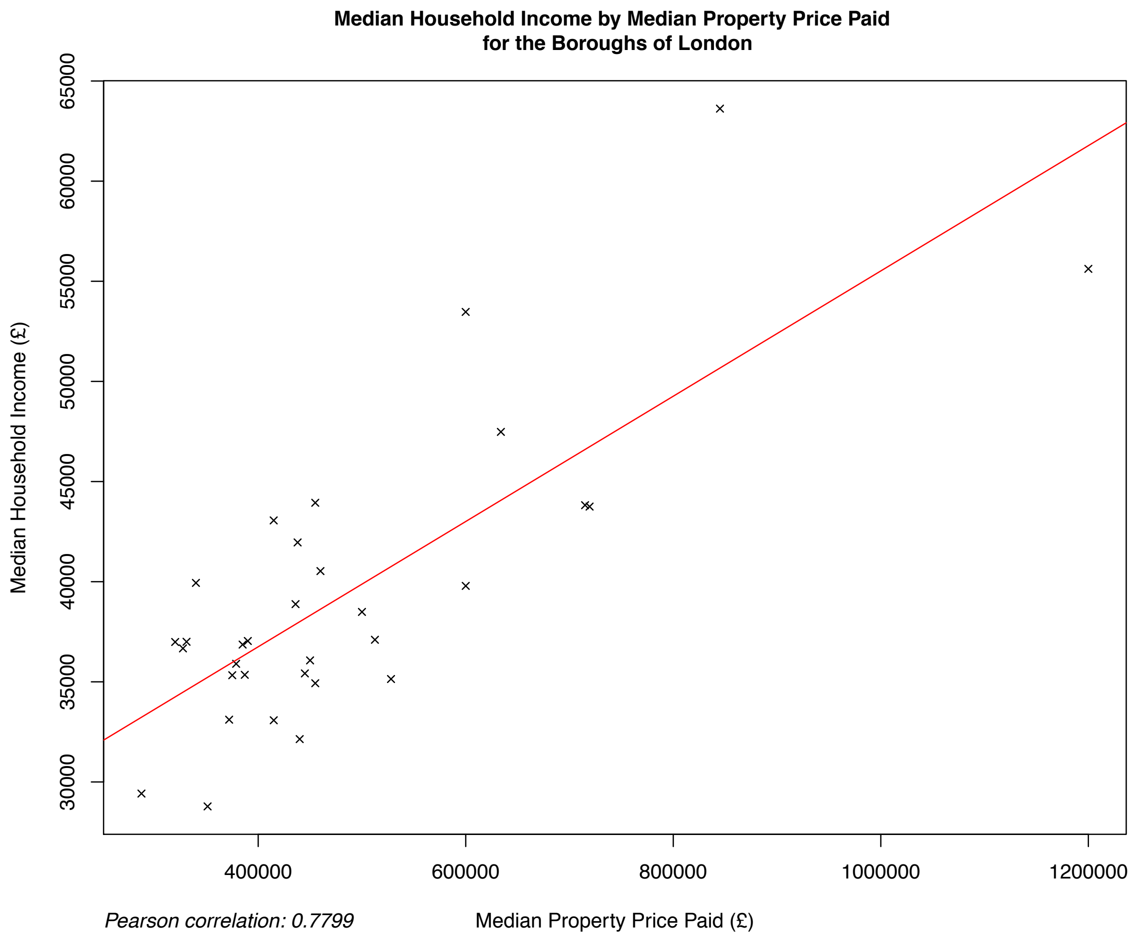


Figure 9: *A scatterplot showing the median household income against the median price paid for each borough of London.*

In order to gain further understanding, the residual values were calculated to show the difference between the predicted and observed values for median household income (Figure 10). The City of London and Richmond upon Thames are the boroughs with the most unexpected result, suggesting their median household incomes are higher than the median property price paid would expect.

Alternatively, Newham, Kensington and Chelsea, Brent, and Hackney have lower median household incomes expected. Spatially, a slight pattern is seen, with the lower median household incomes than expected boroughs, tending to be centrally located, compared to the majority of boroughs on the outskirts experiencing an expected or higher than expected median household income compared to median property price.

Residuals for Median Household Income by Median Property Price Paid for the Boroughs of London

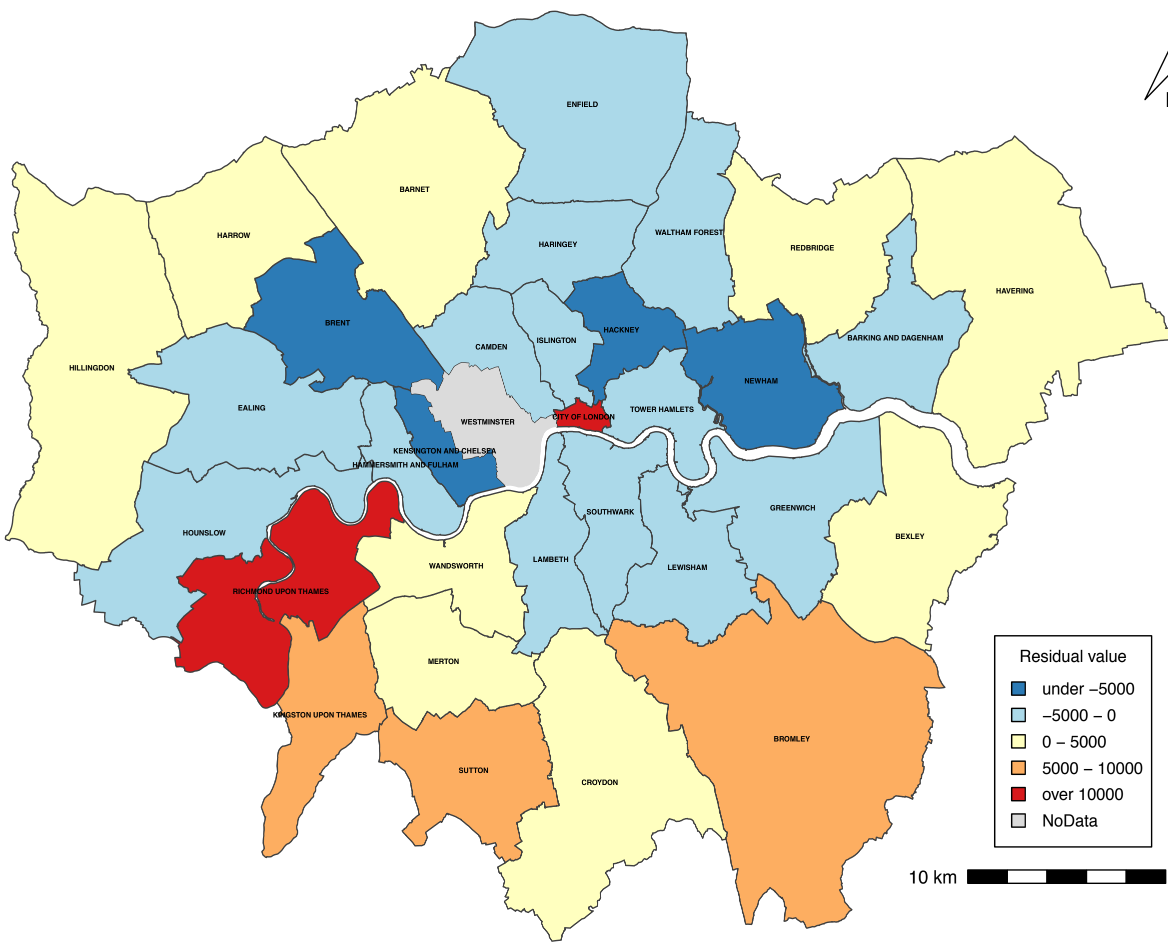


Figure 10: *A map showing the residual values for median household income by median property price. The red indicates boroughs where the actual values are larger than the model estimated. The blue indicates boroughs where the actual values are smaller than the model estimated.*

Discussion

The general spatial distributions of the mean, median, range, and plotted points follow a similar pattern, whereby higher values tended to be concentrated towards the centre of London and lower values on the outskirts. In other research, for London (Hamnett, 2003, 2009) and Chicago (McMillen, 2003) similar spatial distributions were found.

A simple geographical explanation for the spatial pattern is defined by Tobler's First Law of Geography: "everything is related to everything else but near things are more related than distant things". It also follows the Muth-Mills urban spatial structure, which predicts that variables, such as house prices, will decline with distance from the central business district (McMillen, 2003).

On a smaller scale, as shown by investigating individual boroughs, spatial patterns are also seen. Therefore, the variables influencing house prices throughout the city may be influential on a smaller scale. Davidson and Lees (2005) discuss the development of high-cost housing along the River Thames, which could be a variable influencing the high values in the north of Southwark and to some extent for the whole city.

Hamnett (2003) attributes high house prices in London's centre to the rapid growth of high-income professions and the very high earnings in the City of London. Earnings were also found to affect house prices in Spanish (de La Paz, 2003) and Canadian (Fortura & Kushner, 1986) cities. Therefore, the strong positive correlation found between income and house prices is well supported.

A possible explanation for the boroughs of Newham, Hackney, Tower Hamlets and Waltham Forest having a lower income than property prices would expect, is due to the Olympic park's location. It was estimated that the price increase for properties located in the four main host boroughs (those listed above) was about 3.3% after the London 2012 Olympics announcement (Kavetsos, 2011). Furthermore, Watt found Olympic-related displacement was felt by the residents of Newham due to the disparity caused between income and house prices (2013).

As emphasised by de La Paz (2003), there are many determinants of house prices within cities. Therefore, despite only considering household income in this report, it is vital to consider that there are other influential factors and not assume causation from the correlation. Future work may consider the influence of other factors, such as public transport, on house price spatial distribution or conduct further analysis into spatial patterns occurring on a borough scale. Additionally, this work could be replicated with a data set improving upon the limitations found, such as including data for Westminster.

References

Bloomberg (2017) ‘London House Prices Are Still Rising Almost Everywhere’, *Bloomberg.com*, [online] https://www.bloomberg.com/graphics/property-prices/london/ (21.12.17)

Davidson, M. and Lees, L. (2005) ‘New-Build “Gentrification” and London’s Riverside Renaissance’, *Environment and Planning A: Economy and Space*, 37(7), pp. 1165–1190.

Fortura, P. and Kushner, J. (1986) ‘Canadian Inter-City House Price Differentials’, *Real Estate Economics*, 14(4), pp. 525–536.

Fraser, I. (2017) ‘London house prices now a record 14.5x average earnings, as Barclays reveals the fastest-growing areas’, *The Telegraph*, [online] http://www.telegraph.co.uk/property/house-prices/london-house-prices-now-record-145x-average-earnings-barclays/ (30.12.17)

Gray, K. (2017) ‘What will happen to house prices in 2018?’, *Homes and Property*, [online] https://www.homesandproperty.co.uk/property-news/will-uk-house-prices-rise-in-2018-and-how-will-london-price-growth-compare-to-previous-years-a116236.html (30.12.17)

Hamnett, C. (2003) *Unequal City: London in the Global Arena*, Psychology Press.

Hamnett, C. (2009) ‘Spatially Displaced Demand and the Changing Geography of House Prices in London, 1995–2006’, *Housing Studies*, 24(3), pp. 301–320.

Kavetsos, G. (2011) ‘The Impact of the London Olympics Announcement on Property Prices’, *Urban Studies*, 49(7), pp. 1453–1470.

de La Paz, P. (2003) ‘Determinants of housing prices in Spanish cities’, *Journal of Property Investment & Finance; Bradford*, 21(2), pp. 109–135.

McMillen, D. P. (2003) ‘The return of centralization to Chicago: using repeat sales to identify changes in house price distance gradients’, *Regional Science and Urban Economics*, 33(3), pp. 287–304.

Milligan, B. (2015) ‘London house price gap “at record high”’, *BBC News*, 30th September, [online] http://www.bbc.co.uk/news/business-34391334 (30.12.17)

Scott, P. (2017) ‘A room in London now costs more than a whole house in the North East’, *The Telegraph*, [online] http://www.telegraph.co.uk/property/house-prices/mapped-places-building-extension-could-lose-money/ (30.12.17)

Watt, P. (2013) ‘“It’s not for us”’, *City*, 17(1), pp. 99–118.

Appendix A

A table showing the mean price paid compared to the average house price data that was downloaded from the London data store. This helped to validate the decision to remove the highest points from the price paid data as they were skewing the means created. Furthermore, they were likely not houses sold and rather much larger properties.

	Borough	Mean price paid 2016	Average house price 2016
1	BARKING AND DAGENHAM	314,072.6	274,835
2	BARNET	598,581.8	478,051
3	BEXLEY	362,537.2	312,076
4	BRENT	572,128.8	431,970
5	BROMLEY	533,292.7	423,557
6	CAMDEN	1,483,041.2	752,638
7	CITY OF LONDON	5,634,571.1	584,539
8	CROYDON	405,791.7	331,325
9	EALING	645,773.4	446,024
10	ENFIELD	483,113.1	371,791
11	GREENWICH	492,752.9	393,109
12	HACKNEY	717,750.6	515,050
13	HAMMERSMITH AND FULHAM	991,343.5	718,366
14	HARINGEY	601,507.8	449,162
15	HARROW	510,328.9	413,143
16	HAVERING	374,243.3	324,690
17	HILLINGDON	594,489.2	382,931
18	HOUNSLOW	594,871.5	395,683
19	ISLINGTON	850,718.2	616,442
20	KENSINGTON AND CHELSEA	2,112,903.2	1,243,237
21	KINGSTON UPON THAMES	604,726.1	471,030
22	LAMBETH	650,054.6	502,453
23	LEWISHAM	456,824.2	384,740
24	MERTON	601,931.1	467,941
25	NEWHAM	415,273.6	340,889
26	REDBRIDGE	455,767.3	373,756
27	RICHMOND UPON THAMES	822,355.3	611,014
28	SOUTHWARK	809,388.0	535,593
29	SUTTON	408,664.8	349,190
30	TOWER HAMLETS	592,324.7	459,567
31	WALTHAM FOREST	436,453.0	386,035
32	WANDSWORTH	827,128.6	635,689

Appendix B

Sections of the R code used to carry out the analysis in this research.

1 – the code used to import the files and work out the mean value for each borough.

```
#import the shape file
boroughShape <- readOGR(".", "London_Borough_Excluding_MHW")

#import the csv files data and reformat the postcodes
pricesPaid<-read.csv("PricePaidData.csv", header = T, sep = ",",
                     stringsAsFactors = F, check.names = T)
pricesPaid$Postcode<-str_replace_all(pricesPaid$Postcode, fixed(" "), "")

postcodeLookup<-read.csv("PostcodeLookup.csv", header = T, sep = ",",
                         stringsAsFactors = F, check.names = T)
postcodeLookup$Postcode<-str_replace_all(postcodeLookup$Postcode, fixed(" "), "")

#working out the mean for each borough
districtMean <- data.frame(aggregate(as.double(pricesPaid[, "Price"]),
                                     list(District=pricesPaid$District), mean))
districtMean[,"District"] <- as.character(districtMean[,"District"])
districtMean <- rename(districtMean, Mean=x)
```

2 – plotting and exporting the map for mean values

```
#accounting for Westminter NoData
westminster <- joinedAll[joinedAll[["NAME"]] == "WESTMINSTER", ]

#plotting and exporting the map for mean values
breaksMean <- classIntervals(pricesMean, n=6, style = "fisher") #setting the break value
my_colours <- brewer.pal(6, "Reds") #setting the colour scheme

pdf(file="Mean.pdf", width = 9.6, height = 8) #exporting as a PDF

prettymap(plot(boroughShape, col = my_colours[findInterval(pricesMean, breaksMean$brks,
                                                           all.inside = TRUE)],
               axes = FALSE, border = "#404040"), drawscale = FALSE,
          title = 'Mean price paid per borough') #plotting the map
prettymap(plot(westminster, col = "#dbdbdb", add = TRUE,
               border = "#404040"), drawscale = FALSE) #plotting Westminster (no data)
addnortharrow(pos = "topright", padin = c(0.15, 0.8), scale = 0.7,
              lwd = 1, border = "black", cols = c("white", "black"),
              text.col = "black")
addscalebar(plotunit = NULL, plotepsg = NULL, widthhint = 0.25,
            unitcategory = "metric", htin = 0.1, padin = c(0.15, 0.15),
            style = "bar", bar.cols = c("black", "white"), lwd = 1,
            linecol = "black", tick.cex = 0.7, labelpadin = 0.08, label.cex = 0.8,
            label.col = "black", pos = "bottomright")
breaks$brks <- round(breaksMean$brks, digits = 1) #reformatting the breaks value
breaks2 <- format(breaksMean$brks, big.mark=",", scientific=FALSE, trim = TRUE)
legend(x=549034.8, y=170656.5, legend = c(leglabs(breaks2), "NoData"),
       fill = c(my_colours,"#dbdbdb"), bty = "black", bg = "#FFFFFFA6",
       border = "black", cex = 0.75, title = "Price (£)")

dev.off()
```

3 – plotting and exporting the map showing the points for all of the price paid data

```
#merging the price paid with the postcodes
pricesPaid2 <- left_join(pricesPaid, postcodeLookup, by=c("Postcode"="Postcode"))

#Creating a new column for the colour of the point depending on price
pricesPaid2$Colour = "#0000008C"
pricesPaid2$Colour[pricesPaid2$Price<521810.8]="#fef0d9BF"
pricesPaid2$Colour[pricesPaid2$Price<763569.3 & pricesPaid2$Price>=521810.8]="#fdd49eBF"
pricesPaid2$Colour[pricesPaid2$Price<1237192.3 & pricesPaid2$Price>=763569.3]="#fdbb84BF"
pricesPaid2$Colour[pricesPaid2$Price<1797972.2 & pricesPaid2$Price>=1237192.3]="#fc8d59BF"
pricesPaid2$Colour[pricesPaid2$Price<3873737.1 & pricesPaid2$Price>=1797972.2]="#e34a33BF"
pricesPaid2$Colour[pricesPaid2$Price>=3873737.1]="#b30000BF"

pdf(file="Points plotted.pdf", width = 9.6, height = 8) #export to PDF

#plot the points of prices on a map
prettymap(plot(boroughShape), drawscale = FALSE,
          title = 'Each sold property within London coloured by price paid') #plot map
prettymap(plot(westminster,density=15, angle=45, add = TRUE, border = "#404040"),
          drawscale = FALSE) #plot Westminster (no data)
points(pricesPaid2$Eastings, pricesPaid2$Nothings, col = pricesPaid2$Colour,
       cex = .3, pch = 4) #overlay the points in the assigned colouur
addnortharrow(pos = "topright", padin = c(0.15, 0.8), scale = 0.7,
              lwd = 1, border = "black", cols = c("white", "black"),
              text.col = "black")
addscalebar(plotunit = NULL, plotepsg = NULL, widthhint = 0.25,
            unitcategory = "metric", htin = 0.1, padin = c(0.15, 0.15),
            style = "bar", bar.cols = c("black", "white"), lwd = 1,
            linecol = "black", tick.cex = 0.7, labelpadin = 0.08, label.cex = 0.8,
            label.col = "black", pos = "bottomright")
legend(x=549034.8, y=170656.5, legend = leglabs(breaks2),
       fill = c("#fef0d9", "#fdd49e", "#fdbb84", "#fc8d59",
                "#e34a33", "#b30000"), bty = "black", bg = "#FFFFFFA6",
       border = "black", cex = 0.75, title = "Price (£)")

dev.off()
```

4 – plotting and exporting the map showing the 500 most and 500 least expensive points

```
#assign variables for the points
pricesPaidReordered <- pricesPaid2[order(pricesPaid2$Price, decreasing = TRUE),]
first500 <- pricesPaidReordered[1:500,]         #500 most expensive
last500 <- pricesPaidReordered[120018:120518,]  #500 least expensive

pdf(file="MostAndLeast.pdf", width = 9.6, height = 8) #export to PDF

#plotting the most and least expensive points
prettymap(plot(boroughShape), drawscale = FALSE,
          title = 'The 500 most and 500 least expensive properies sold')
prettymap(plot(westminster,density=15, angle=45, add = TRUE, border = "#404040"),
          drawscale = FALSE)
prettymap(points(first500$Eastings, first500$Nothings, col = "red", cex = .8),
          drawscale = FALSE)
prettymap(points(last500$Eastings, last500$Nothings, col = "dark green", cex = .8),
          drawscale = FALSE)
addnortharrow(pos = "topright", padin = c(0.15, 0.8), scale = 0.7,
              lwd = 1, border = "black", cols = c("white", "black"),
              text.col = "black")
addscalebar(plotunit = NULL, plotepsg = NULL, widthhint = 0.25,
            unitcategory = "metric", htin = 0.1, padin = c(0.15, 0.15),
            style = "bar", bar.cols = c("black", "white"), lwd = 1,
            linecol = "black", tick.cex = 0.7, labelpadin = 0.08, label.cex = 0.8,
            label.col = "black", pos = "bottomright")
legend(x=551787.9, y=169345, legend = c("500 Most", "500 Least"),
       col = c("red","dark green"), bty = "black", bg = "#FFFFFFA6",
       border = "black", cex = 0.75, pch = c(1,1))

dev.off()
```

5 – generates residual values for median income and house price then plots and exports the values

```
#generates the residual values for the median income and house price
model <- lm(joinedProfiles2@data$Modelled.Household.median.income.estimates.2012.13
            ~ joinedAll2@data$Median)
resids<-residuals(model)
map.resids <- cbind(joinedProfiles2, resids)
names(map.resids)[13] <- "resids"
var <- map.resids@data$resids

breaks <- classIntervals(var, n = 5, style = "pretty") #set break values
my_colours <- rev(brewer.pal(5, "RdYlBu")) #set colour scheme

pdf(file="Residuals.pdf", width = 9.6, height = 8) #output to PDF

#plot map
prettymap(plot(map.resids, col = my_colours[findInterval(var, breaks$brks, all.inside = TRUE)],
     axes = FALSE, border = "#404040"), drawscale =FALSE,
     title = "Residuals for Median Household Income by Median Property Price Paid for the Boroughs of London")
prettymap(plot(westminster, col = "#dbdbdb", add = TRUE, border = rgb(0.8,0.8,0.8,0)),
          drawscale = FALSE)
addnortharrow(pos = "topright", padin = c(0.15, 0.8), scale = 0.7,
              lwd = 1, border = "black", cols = c("white", "black"),
              text.col = "black")
addscalebar(plotunit = NULL, plotepsg = NULL, widthhint = 0.25,
            unitcategory = "metric", htin = 0.1, padin = c(0.15, 0.15),
            style = "bar", bar.cols = c("black", "white"), lwd = 1,
            linecol = "black", tick.cex = 0.7, labelpadin = 0.08, label.cex = 0.8,
            label.col = "black", pos = "bottomright")
legend(x=553034.8, y=170656.5, legend = c(leglabs(breaks$brks), "NoData"),
       fill = c(my_colours,"#dbdbdb"), bty = "black", bg = "#FFFFFFA6",
       border = "black", cex = 0.75, title = "Residual value")
text(boroughCentroids, labels = boroughShape$NAME, cex=0.35, font = 2) #adds labels

dev.off()
```