

An Analysis of London House Prices

Harun Abdulqadir

20/04/2021

Introduction

London remains the most expensive region to live in the UK with house prices soaring reaching an average of £469,269 in February 2021. There are many reasons as to why house prices have soared, all of them nonetheless contributing to growth across average house prices and number of sales.

Despite reaching back all the way to 1995, the data set we will be observing today only accounts for house prices up until the year 2017. This, however is more than enough to work with to provide a detailed analysis. We will observe the different trends that have come to light in the 20 year period we are studying and ultimately aim to understand why London has been the subject to such a boom in recent years. Are certain areas disproportionately leading the way in the house price crisis? Are people buying less? Have house prices simply seen unsustainable growth?

A continued study will lead us to carrying out a linear regression on the data provided to us. In doing such, we will aim to underline the trajectory house prices will undergo. In doing this we hope to even draw conclusions of whether another bubble may be forming and be in risk of bursting as previously observed in the housing market. My analysis will follow a particular line of thinking and the following variables will be of specific interest:

- **Area:** The distinct boroughs of London.
- **Measure:** The metric used to measure the value.
- **Value:** The value corresponding to the 'measure' variable.

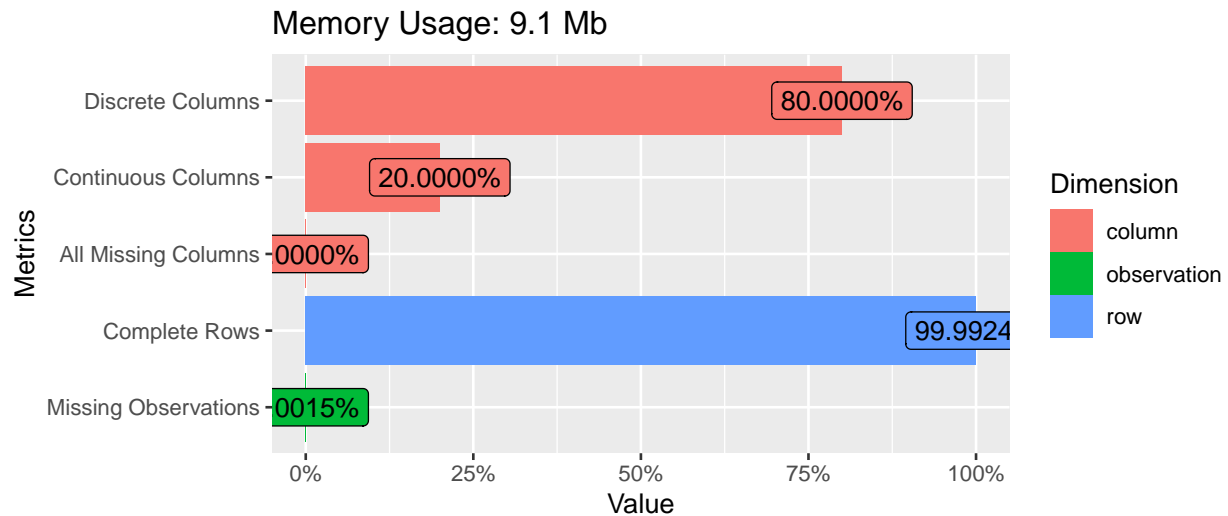
The Data

The dataset itself consists of the area code, respective an area of each borough in the first column. The second column represents the separate boroughs of London, followed by a code detailing the distinct areas in the borough. The next two are measures representing their value. The fourth column observes the method used to measure the value (mean, median, sales) and the fifth and final column observes the numeric value of their the previous measure.

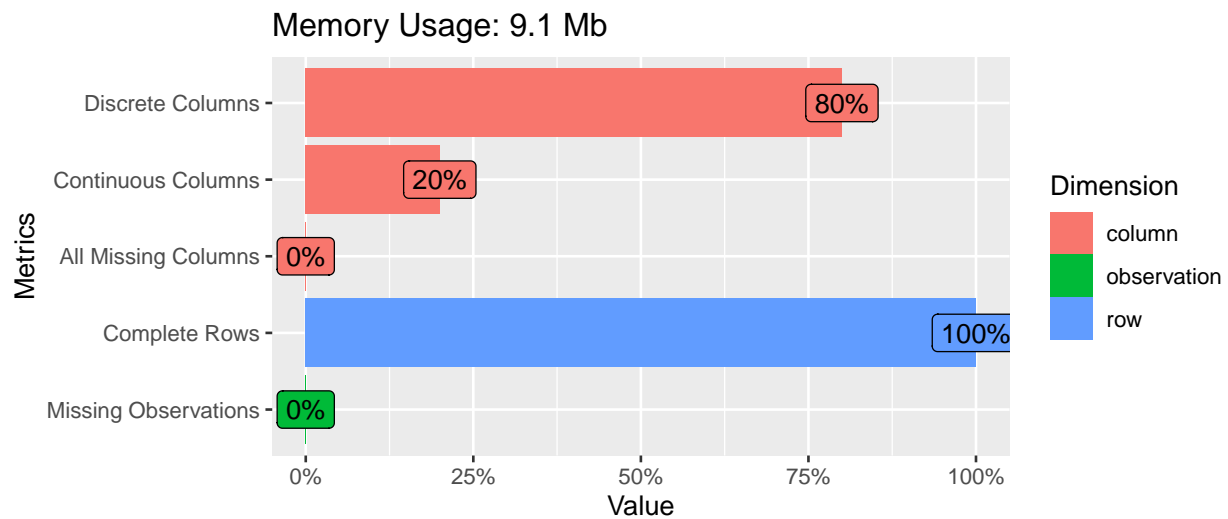
```
## [1] "Code"      "Area"      "Year"      "Measure" "Value"
```

The Data Continued

```
##      rows columns discrete_columns continuous_columns all_missing_columns
## 1 262441      5              4              1              0
## total_missing_values complete_rows total_observations memory_usage
## 1              0      262441      1312205      9578760
```



- The slight percentage in the missing observations metric demonstrates a need to tidy the data up

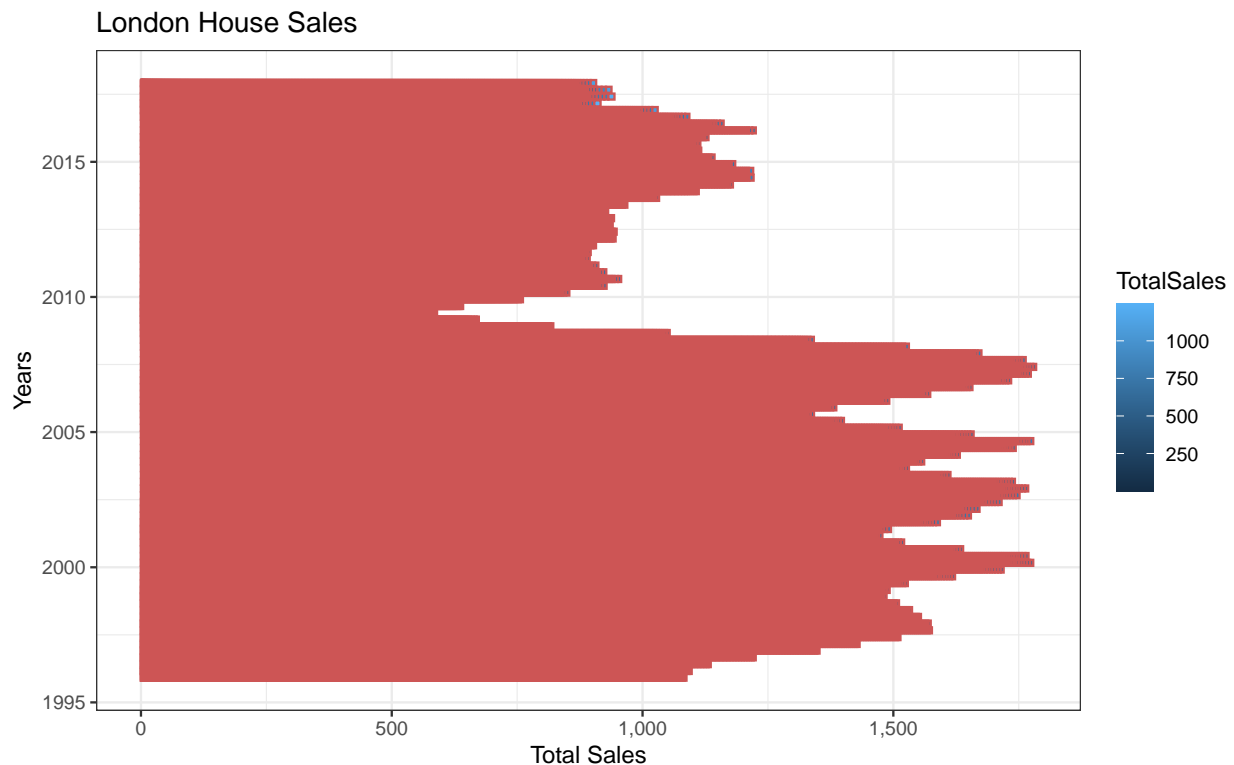


- As can be observed, the new dataset no longer contains the missing observations and the marginally incomplete rows. Cleaning these two important facets of the dataset ensure for a smooth and untroubled data analysis in which you can carry out tests and plot the data with a lot less hassle.

Median Sales and Prices in London

Average House Sales

- Having provided a brief overview of the data and cleaned it up a little, we can plot the housing by Sales and by Median price in London.
- Important to note is that the dataset contained three different ways of measuring the value, that being Sales, Median and Mean. In this section I will plot the Sales and the Median, leaving out the Mean average as the Median is preferred here. Reason being the Median does not depend on all the values in the dataset, consequently when some values are extreme the effect on the Median is smaller, creating a better measure of central tendency as opposed to the Mean.

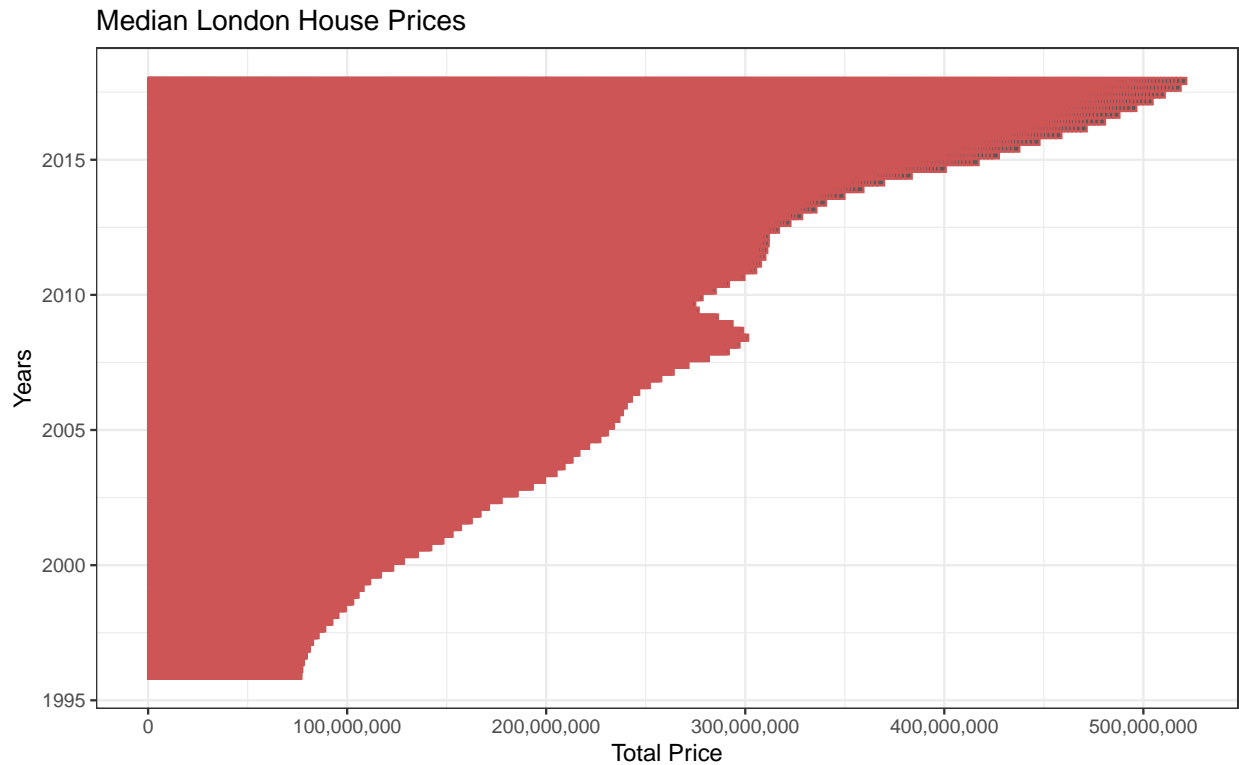


Inference

- Based on this graph, we can notice some interesting trends regarding house sales in London over the last twenty years. House sales generally fluctuated since 1995 but ultimately held a strong increase in sales.
- Following 2005 we begin to see a steep decline in house sales, primarily as a result of the housing bubble and the global financial crisis. It was perhaps no surprise that people were not trying to buy houses in a time of crisis.
- Post 2010 saw Housing sales take off again but never really reach the pre 2005 highs as people recovered and house prices simply outgrew Londoners and the average London Salaries.

Median House Price

- Average house prices in London have grown exponentially in recent decades, however the last decade has seen the housing market inflate to ridiculous levels.
- Since 1995 there has been a +695% change in the price for a detached house across London. Staggering growth.

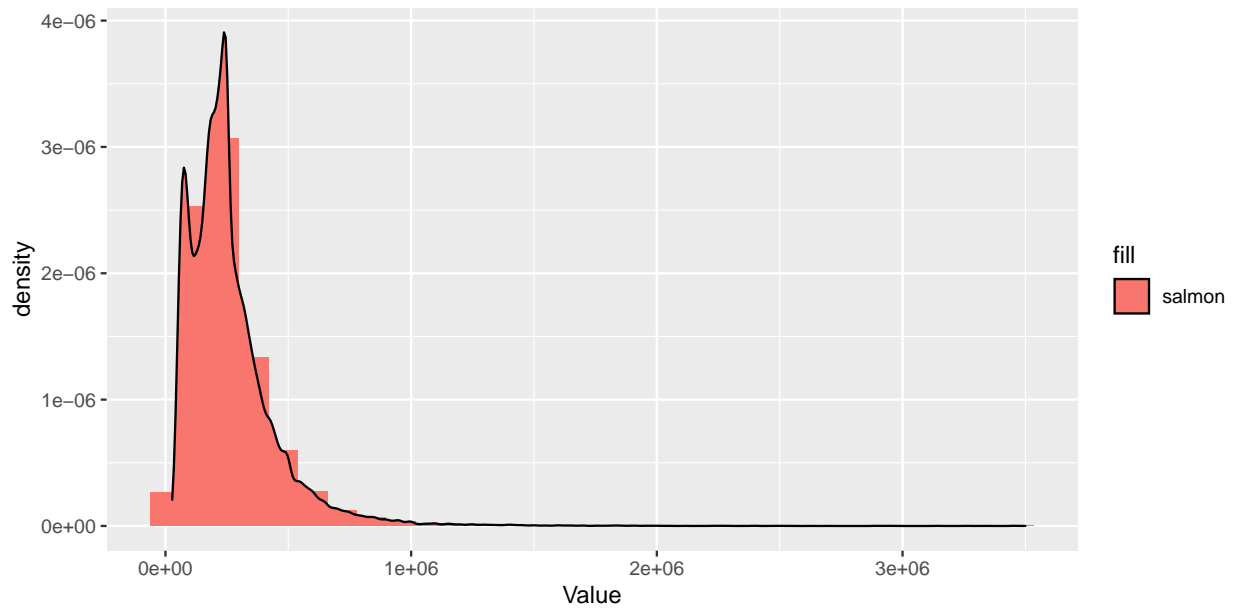


Inference

- After visualizing this, we can see that the median house price is around £250,000 as the average house price soars past £500,000 in 2017.
- London is however known for its typically affluent areas and like anywhere, there are also the impoverished are. As impressive as it seems, it must be interesting to see how exactly the distribution is shared.

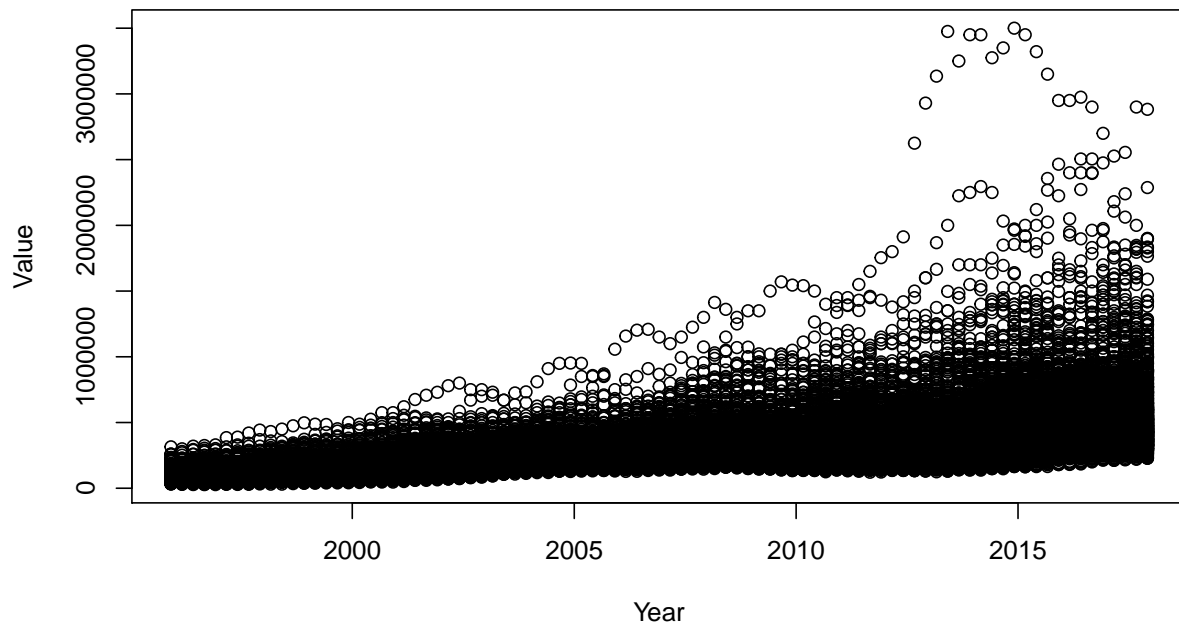
Distribution

- Below is a histogram combined with a density of the price of houses.



- Unsurprisingly, the distribution looks like it follows a power law. The easiest way to understand the power law distribution is to think of the famous 80:20 rule. In business this often translates to “80% of your revenue comes from 20% of your customers”. In the context of this dataset, you can make a very rough estimate that 80% of house prices are in the lowest 20% of the price range.

Scatterplot



- We can double down and further see with this simple scatterplot that majority of the house prices are in fact in lower 20% with some serious outlier areas in London driving those prices right up.

Summary

- In summary, we spotted a number of trends in surveying the data, mainly ones following before and after the global financial crisis in 2008. Since that juncture it has been a stark rise in house prices, one that is disproportionately in line with salaries and income in for the average person in London.
- This sharp, insurmountable rise in house prices has seemingly and almost expectedly discouraged house buyers in London. House sales have never really returned to their heights before 2008 and this is perhaps because of this reason, an unsustainable rise in house prices.
- We also suspected that some typically affluent areas were dramatically rising the average price, and although the median does well to calculate an average that minimizes skews in the data, it is always worth examining the outliers. As the histogram with the density estimate demonstrated, it was more than likely 20% of house prices were extraordinarily high and most of the data (80%) is collated from the lower 20%. The very basic scatterplot also helped visualise further as sometimes less is more.