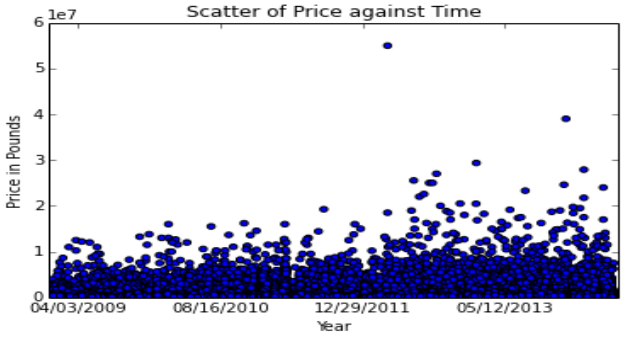
Daniel Jones, University of Cambridge,
dwj26@cam.ac.uk

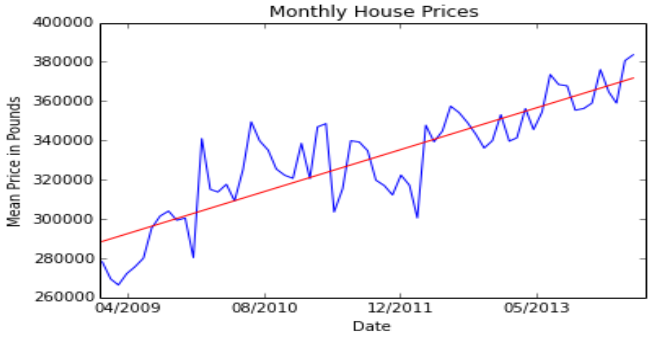
An Investigation into House Prices

(a)The data given was used to firstly create an investigation into the general statistical nature of house prices and the trends they tend to follow to see if anything interesting or striking could be found. Following this a more geographical approach was taken to see if the position of the houses and the local amenities affected the house prices. Figures are labelled from 1 - 10 in the order observed.

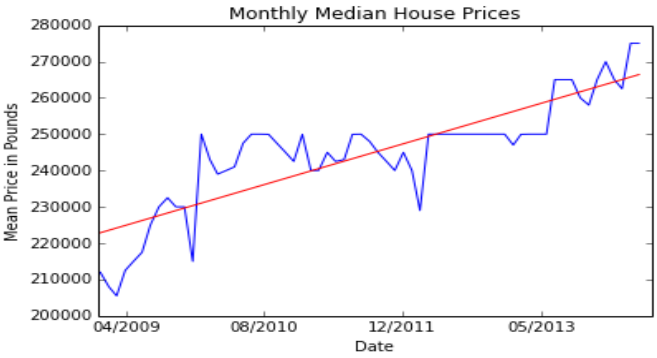
The General Investigation: Analysis

Figure 1 was simply a plot of all house sales against the date that they were made. The only real thing that I can see from this is that towards 2014 more outliers (houses far above the average value) were purchased. A further investigation of this would be required.



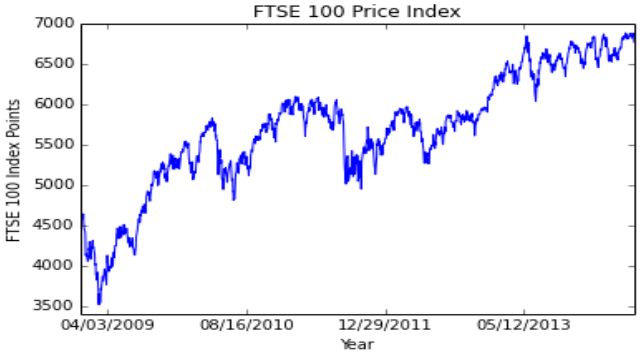


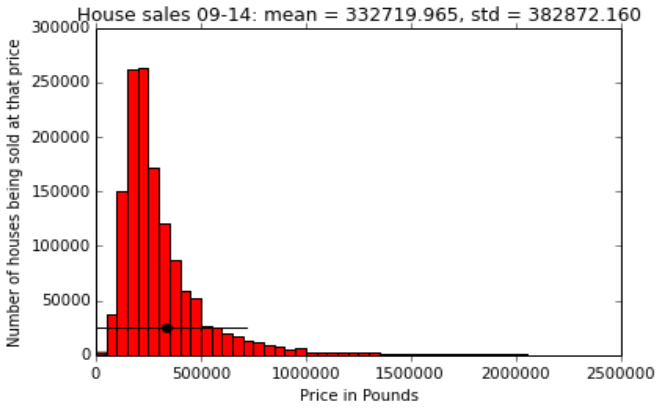
To investigate this further I sorted the data into months and then plotted a moving plot of the means, figure 2. Firstly, it can be seen that the mean was increasing along the scale of time. This may be due to the purchase of some very expensive houses (outliers) that were statistically insignificant so a plot of the median house prices was instead created, figure 3.



As can be seen the trend of increasing average house prices was still observed. It was thought that there may be some more data online that could predict the movement of the house price market. With a first thought looking at the general state of the economy, analysing data from the FTSE daily closing values for the same time period a new plot was produced, figure 4.

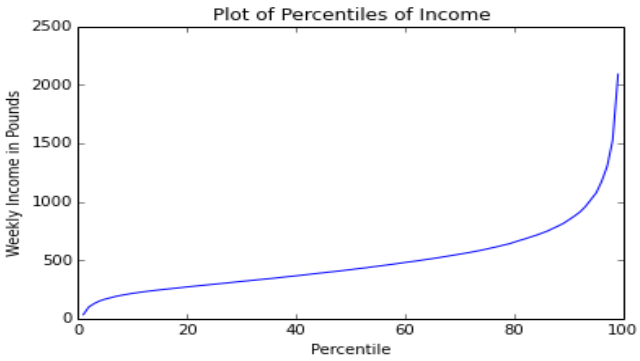
As can be see there is a matching trend between the two data sets. If this is a causal relationship this could be significant for investment in the housing market.





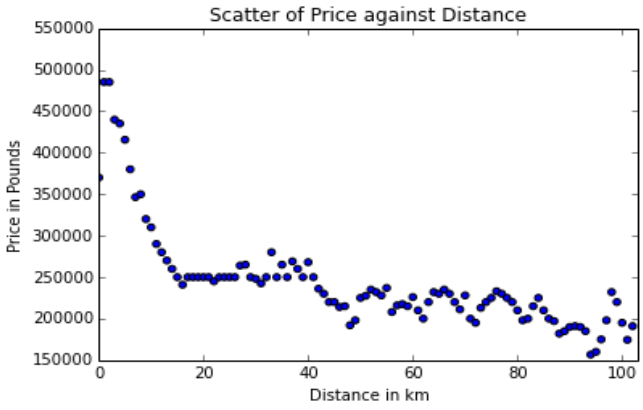
Daniel Jones, University of Cambridge,
dwj26@cam.ac.uk

Now a quick explanation of the statistical nature of house prices can provide insight into the pricing, and more specifically the sale of houses. As can be seen with the plot above (figure 5), the house market is clearly not Gaussian in nature, the plot has a long tail at the higher end of the market. This can easily be explained by open data relating to the percentiles of income (figure 6), as one reaches the top few percent of the country wealth starts to rise extremely quickly, allowing the purchase of such houses and giving rise to an extremely large tale. More importantly the amount of data outside 3 standard deviations is 1.30% with a Gaussian this should be 0.3% (but only 0.15% from the higher end of the Gaussian).

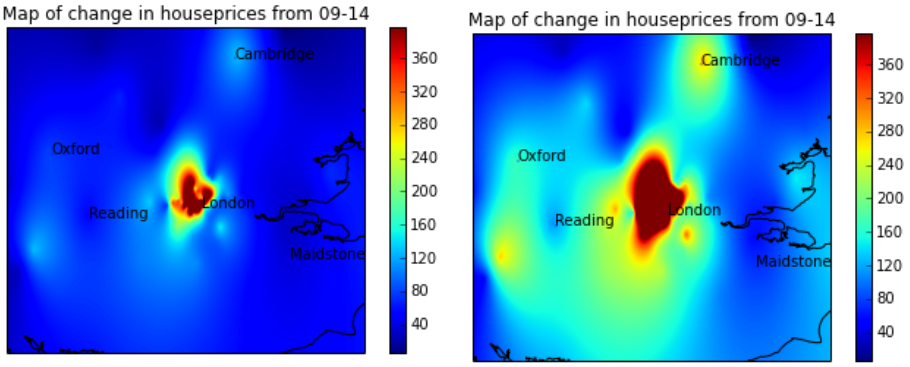


Creating a graph to check the how house prices varied depending on distance to the centre of London was up next. See fig 7

Next, I sorted the house sales into regions (using the first two letters of the postcode) and for each region found the gradient at which the house prices are increasing. Below are two plots, of differing resolutions (the resolutions are changed just by changing the maximum of the colour scale). See figure 8 and 9. These were created by interpolating between the regions and so any regions not covered were accounted for by the geographical interpolation of our data.



Finally, a quick look at intially the most easily accessible amenity in the data set, the railway stations. The data was sorted to group houses with the same number of railways within 1km together, and then average these house prices. The following pattern emerged. See figure 10.



These graphs are a selection of the most useful analysis taken out on the the dataset. The next part of the project will be discussing the implications of these in detail.