**EDA Mini Project 1**

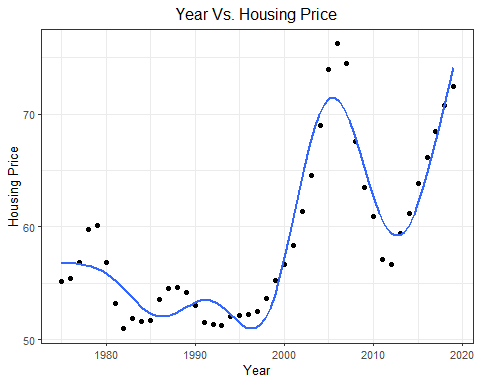
Team 3 (Ojaas H, Vijeet S, Meet V)

**Introduction:**

House prices and Population growth in the United States is a major topic which spans from the fields of census, economics to as far as determining the standard of living of the population in general. The main question to be answered here is whether the house prices have changed over the years and how is the change related to changes to population if any relation is found. This question is a complex question and needs to be answered in simple boiled down parts which can be assessed independently to get conclusions. The various sub questions are described in detail below.

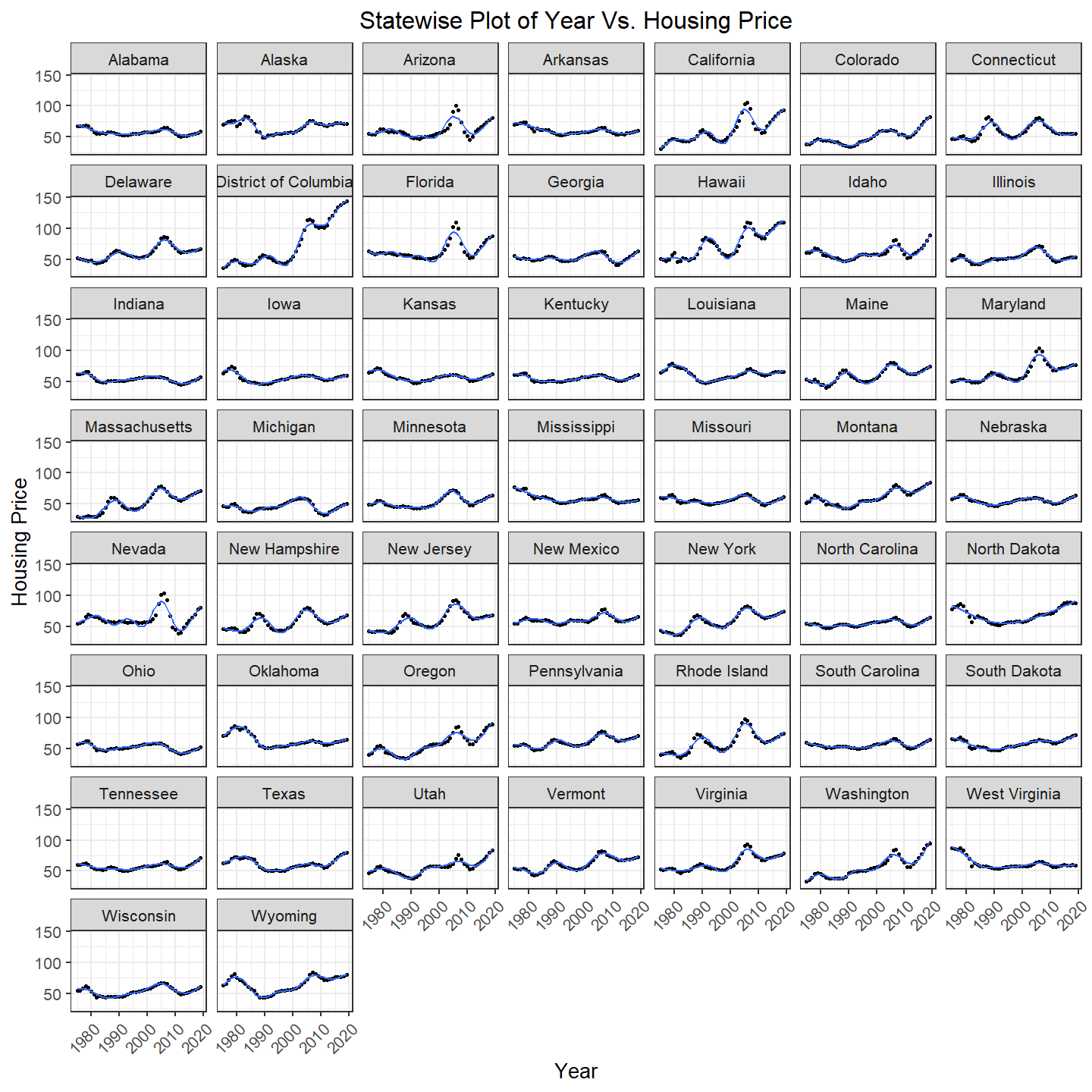
**House Prices Over Time:**

**Year vs House Price:**



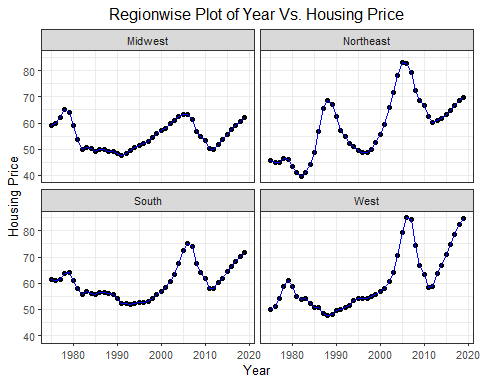
In this case, we have used a GAM model to fit the data. Gam is used to fit a non-parametric model to the data. GAM works very well when the factors to be compared is not particularly linearly related to each other. This can sometimes lead to a loss if interpretability, but in this case, we can easily interpret the change in house prices over the years. The graph clearly shows that there are fluctuations in Housing Prices over the time period. The prices for the houses over the years have been adjusted according to the CPI Index (inflation), hence the effect of market conditions and currency changes has been taken into consideration before plotting the graph. There is a steep increase in the trend for increase in house prices after 1995. The graph also shows that there are increased fluctuations in the 2000’s and forward. This effect can be attributed to the change in market dynamics and effects of globalization.

**Year vs House Prices faceted by States:**



From the graph we can very easily observe that the states show a varied trend for changes in house prices over the years, however the overall trend for house prices over the years remains approximately the same for all the states barring a few exceptions. The District of Columbia shows the highest local increase in the prices, while Nevada shows the maximum local decrease in prices over the years. The major outlier to this trend however is the state of West Virginia. In West Virginia, the house prices decrease over the years as compared to all other states where the prices increase over time.

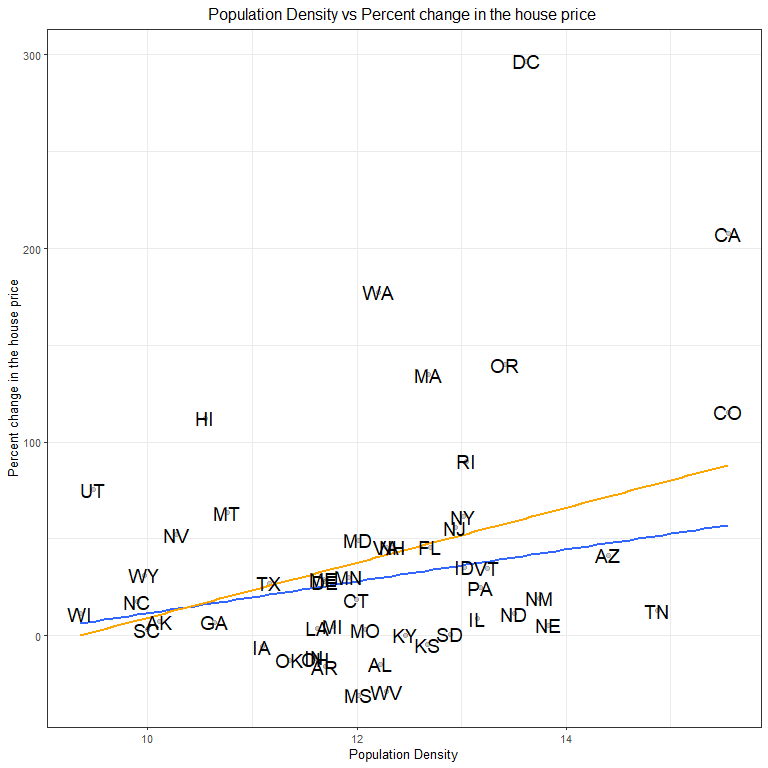
**Year vs House Prices faceted by Regions:**



The region wise faceted graph helps us to get a better insight of how the prices change over time in various sub parts of the country. In the later 2000’s and forward, all the regions show a sharp increase in house prices, apart from the Midwest region, where the house prices follow an approximately flat trend with not much increase or decrease. While we talk about the absolute difference in prices over the years, West region sees the highest increase followed by the Northeast region. Another interesting observation that comes to light here is that the Midwest and the South region have an identical trend with respect to increase in house prices over the years, but Midwest follows a mostly flat trend as compared to South region where prices of houses do increase substantially over time.

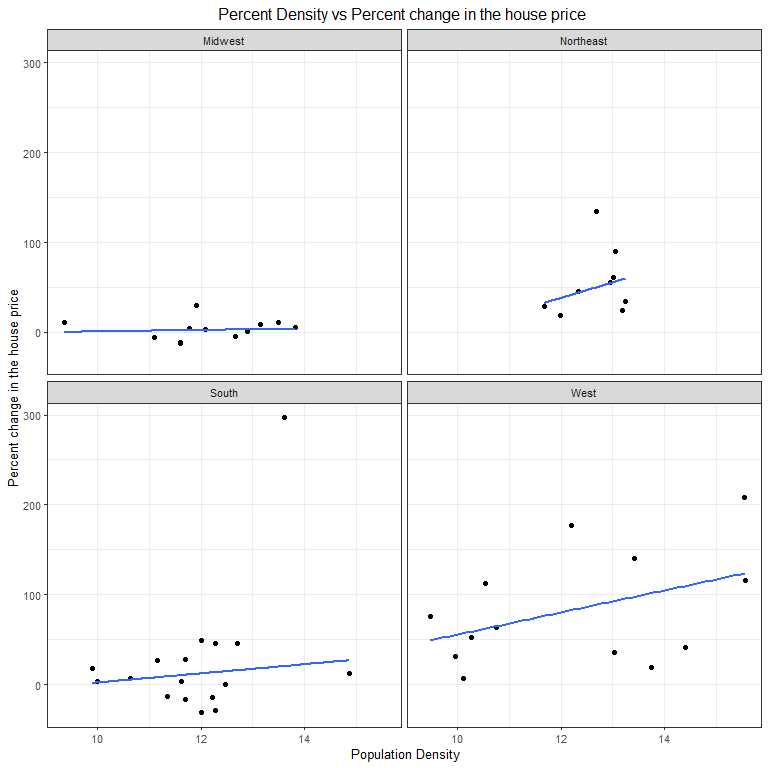
**Population Density and changes in House Prices:**

**Percent Density vs Percent Changes in House Prices:**



We have fitted a Robust Linear Model here in place of a linear model. The major property of RLM is that Robust linear model is used when we want to downweigh the effect of any outliers that affect the data. Here, we need to take the log of changes in population density (population density is calculated as [(endPopulation of interval-startPopulation of interval)/( startPopulation of interval) \* 100]). This is because the population change density in some states is very high as compared to some other smaller states. Hence, taking log for the same helps to scale them and fit them on a smaller axis. When we plot the graph of changes in population density with respect to changes in house prices, we see that there are two main outliers DC and CA. The houses prices in these two states has increased at a higher rate than any other place in the country. Hence, we see the slope of the fitted line is highly positive in that case. But when we drop those two states as outliers and observe the trend, we see that the overall change in house prices with respect to changes in density does not follow a highly positive trend. The removal of outliers changes the effects observed in the graph at a very high rate.

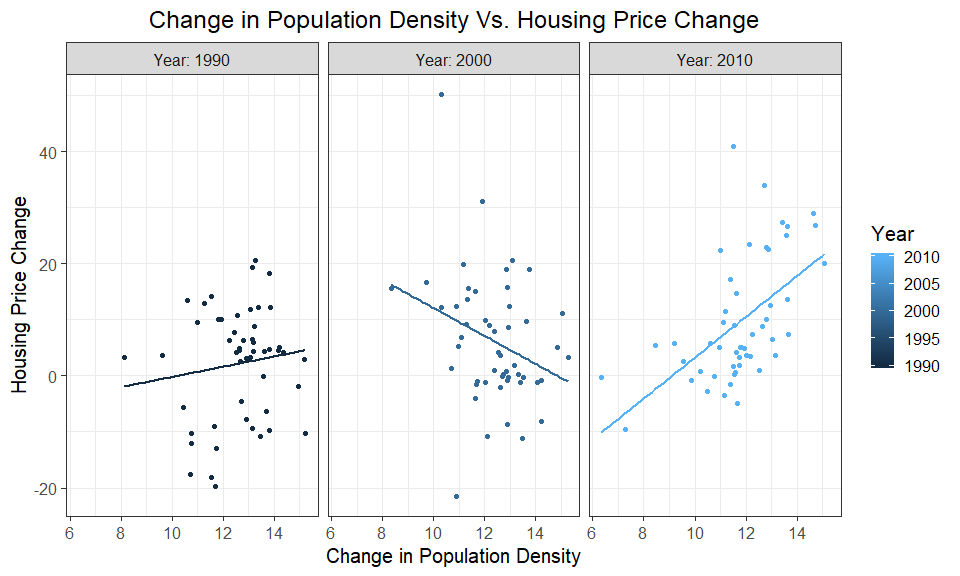
**Percent Density vs Percent Changes in House Prices faceted by Regions:**



The trend for the region wise graph remains approximately like the trend observed for changes in house prices over time. All the regions show substantial increase in house prices with increase in population density apart from Midwest which follows a flat trend. The south region shows similar characteristics to Midwest region but has higher increase in changes in house prices with changes in population density as compared to Midwest region.

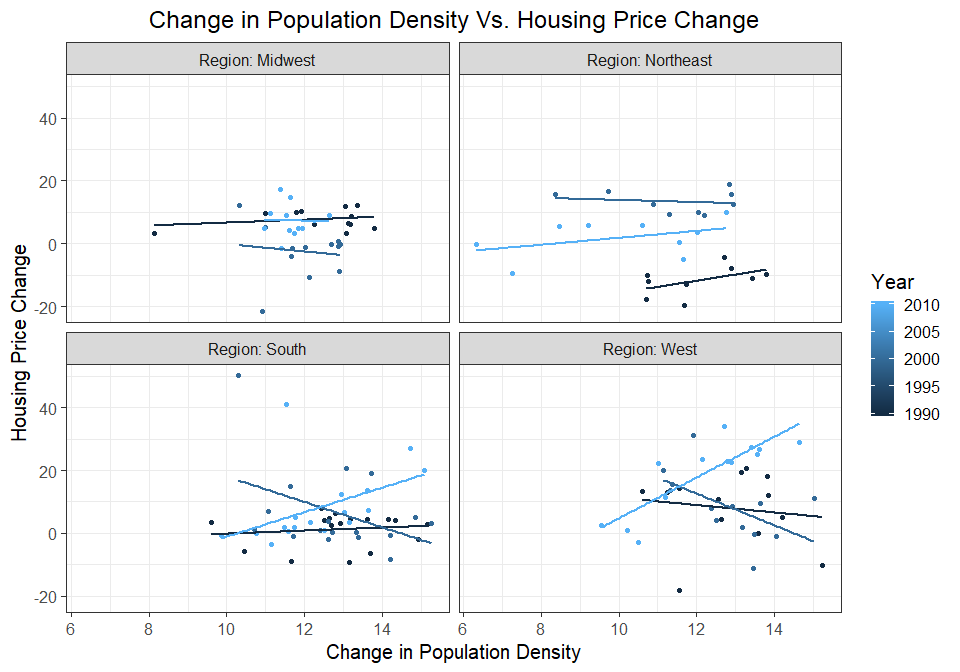
**Changes in Population and Changes in House Prices:**

**Changes in Population and Changes in House Prices faceted by Decade:**



The relation between changes in house prices and changes in population can be observed very clearly by plotting the graph which shows changes decade by decade. We have used a linear model to fit the data here, because there are no major outliers to the trends that can be spotted from the graph. When we observe the graph, we can see that the relation between the changes in population and house prices is almost flat with a very menial positive slope in the year 1990, while in the 2000’s, the trend is highly negative. But as we proceed to see the trend for the year 2010, the trend is again highly positive with a very steep slope.

**Changes in Population and Changes in House Prices faceted by Region:**



The region wise graph shows us a lot of similar looking trend over the decade factors for various years. When we see the graph for the years 2010 to 2018, we can see that there is an increase in house prices for all the regions, which is the same trend that was observed in the graph faceted by decade as well. When we see the trend for the years 2000 to 2010, we can see that there is a negative trend in house prices with respect to population for all the four regions. This trend is again the same as the one observed above in the decade facet graph. But the most interesting feature of this graph is the decade from 1990 to 2000, in this decade all the regions show a different trend with the changes.

**Conclusion:**

House Prices Over Time: We know for a fact that market dynamics and inflation affect the prices of goods over the years. Here we have taken the above factors into consideration by adjusting for the CPI rates. The graph shows that the house prices change positively over time. This is a very logical conclusion from the given data and the market trend that we know.

Population Density and changes in House Prices: The relationship between changes in population and changes in house prices is positive with a very small slope increase value. This is because the changes in population do not affect the changes in house prices on a very large scale, because house prices vary as an effect of inflation, market changes, currency changes in addition to changes in population. However, we can expect a positive change in house prices with respect to changes in population density and that can be observed from the graphs.

Changes in Population and Changes in House Prices: The changes in population with respect to changes in house prices is expected to display a positive trend as well, which can be easily confirmed from the graphs plotted above. However, we can see that the market condition (inflation in 1995) causes us to observe some weird and untold trends in the graph. Apart from that factor, all other parts of the graph show a trend of increase in house prices with increase in population.

From the above trends and graphs, we can say that there is a positive trend for changes in house prices over the years, while there is also a slightly positive trend for increase in house prices with increase in population (which is expected). We expect the house prices to vary positively with changes in population because changes in house prices in any country depends on the factors such as population, population density changes, market changes, inflation and changes in currency rates.