Mini Project Report

Atharva Mahesh Pore and Jagadeesh Chitturi

2023-10-18

## Prompt

A researcher for a thinktank wants to learn about how house prices in the U.S. have changed over the last few decades, and whether changes in prices are related to population in some way.She has taken an introductory statistics course using R, but that was a long time ago, so she is outsourcing the exploratory data analysis to YOU. The researcher’s major research question is: How have house prices in U.S. states changed over the last few decades, and are changes in prices related to population in some way?

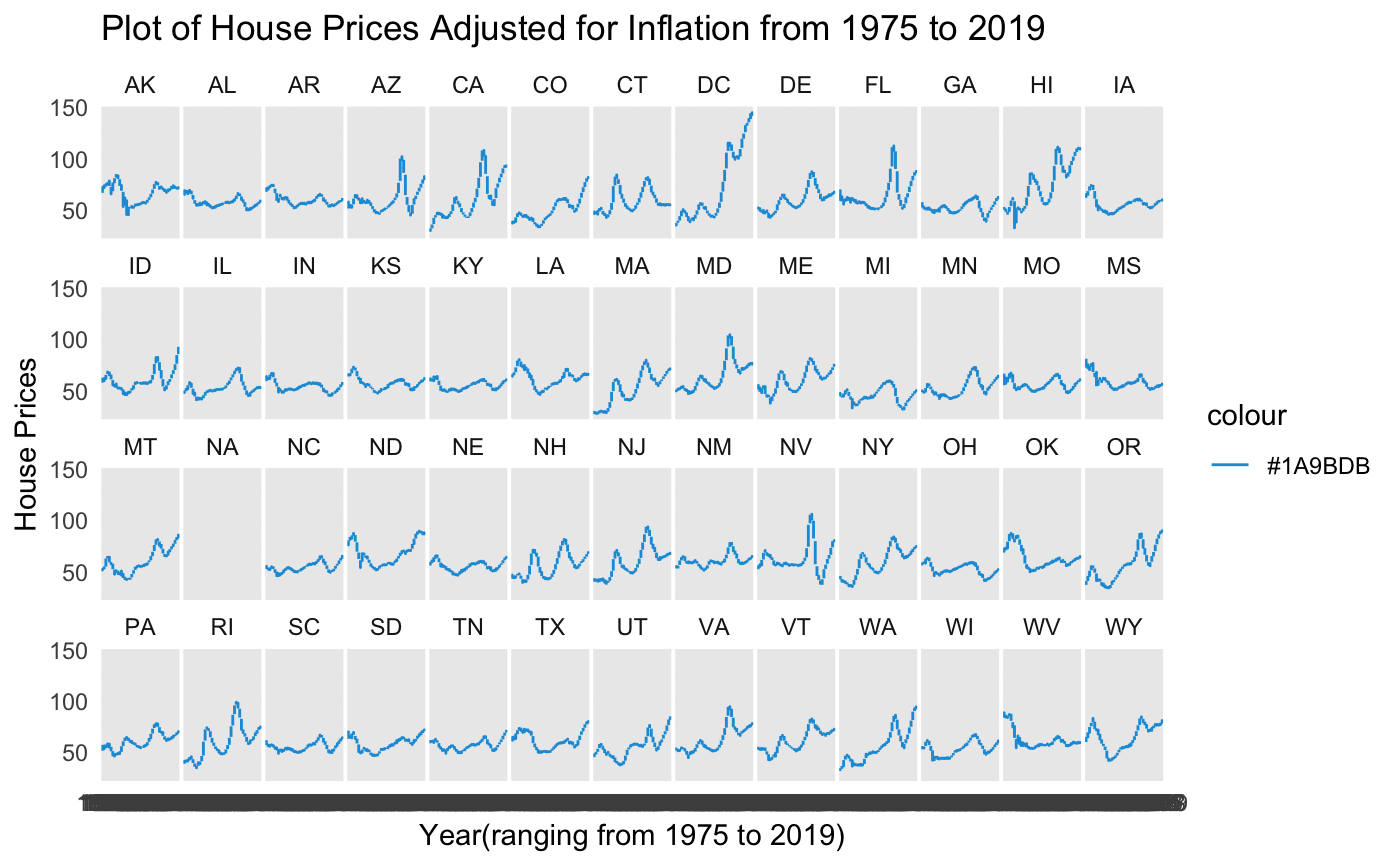
## Executive Summary

We explore and understand the relationship between house prices over the years and the effect that the population density have on them using the data primarily from <http://www.freddiemac.com/research/indices/house-price-index.html> and www.bls.gov/cpi/data.html. In short we understand that both time and population density have strong effects on the house prices(even with adjusted inflation). We can conclude that the house prices in some of the states increase by a considerable amount(from 1975 to 2019) whereas in some they remain constant(after adjusting for inflation) and even reduce in some. We can also conclude that the house prices have a slow but steady increase according to population density except for some of the outliers like Washington DC.

## Part 1

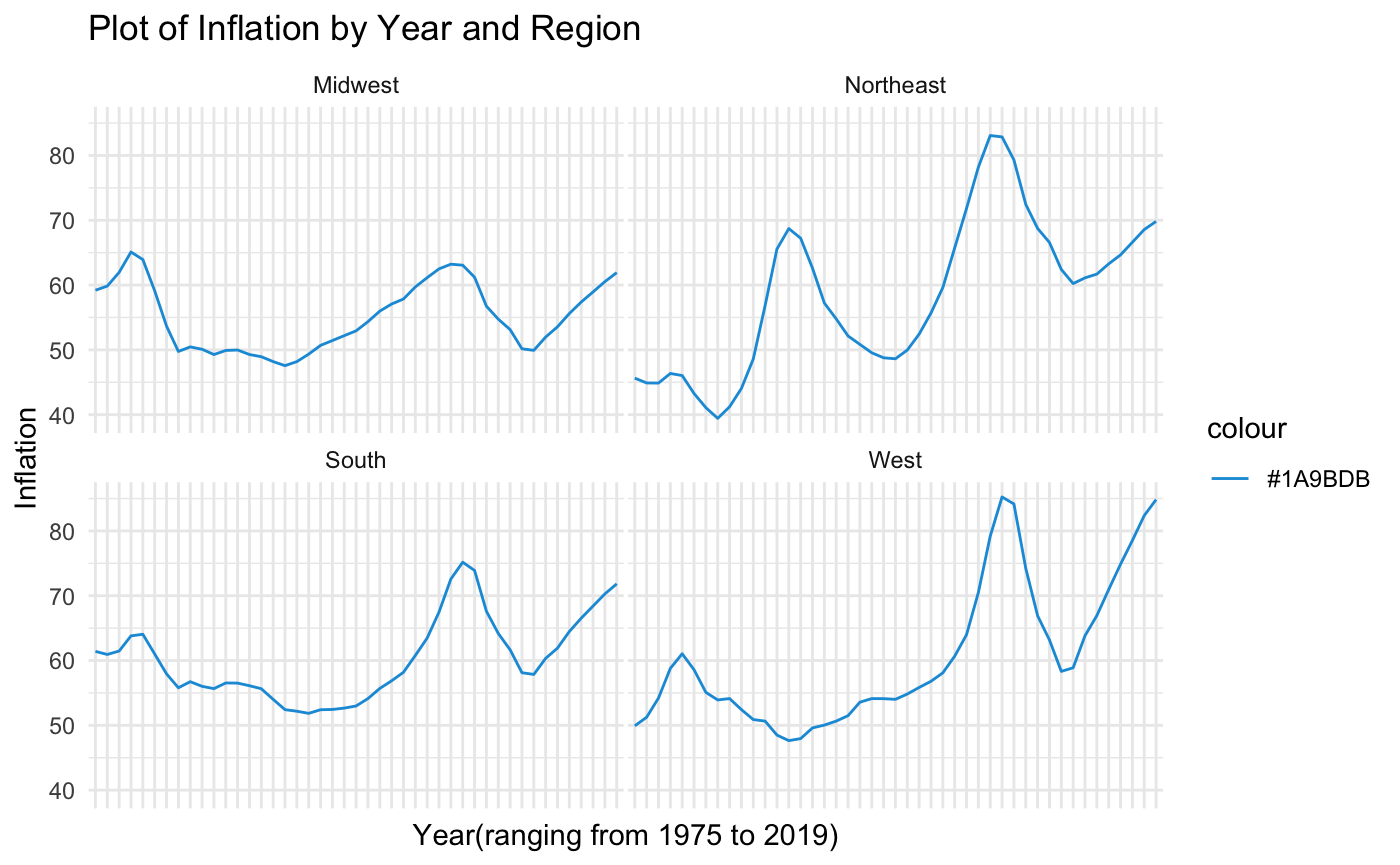
## House Prices adjusted by inflation from 1975-2019, statewise.

We begin with adjusting the house prices according to the inflation using the CPI dataset. This helps us in understanding by how how the prices actually vary by keeping inflation in mind. As we can see from the below plots, most of the states have increased their house prices over the years even with adjusted inflation. There are a few states such as the District of Columbia and Hawaii where the prices increased by a subsequent margin. Most of the states had there values increase but not as big of a margin as that of DC. A few of the states that reduced the values were Ohio, Indiana and West Virginia



We then use the data from the newer State code file, and then merge it into the main House Prices dataset to find the variation house prices in the four main regions of the United States, i.e. South, Northeast, Midwest and West.

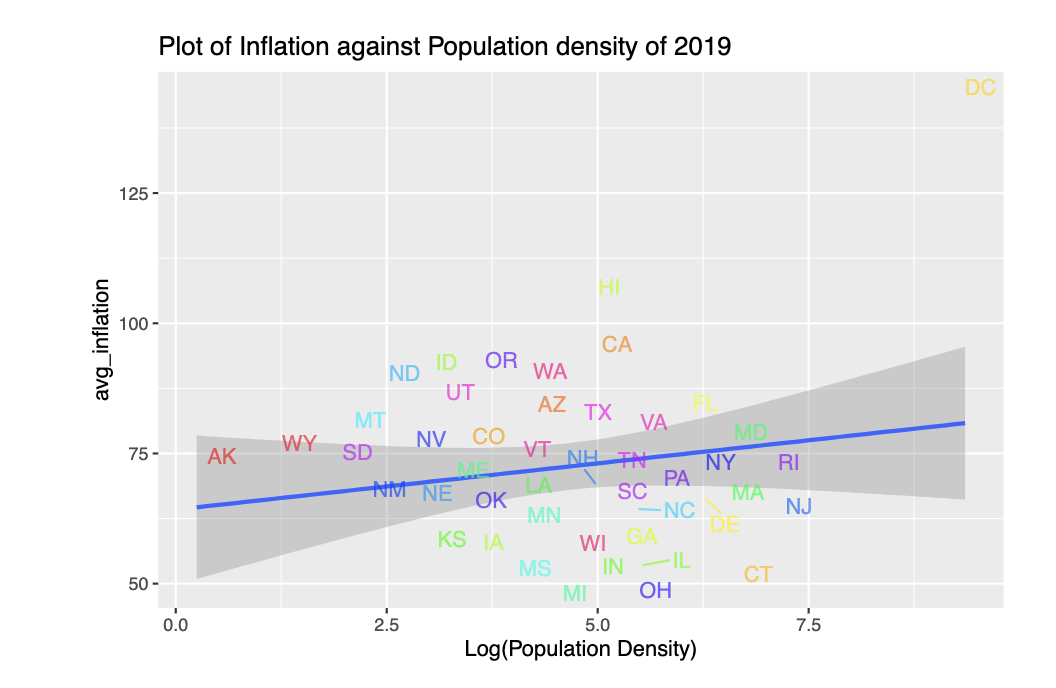
## House Prices adjusted by inflation from 1975-2019, regionwise in the US.



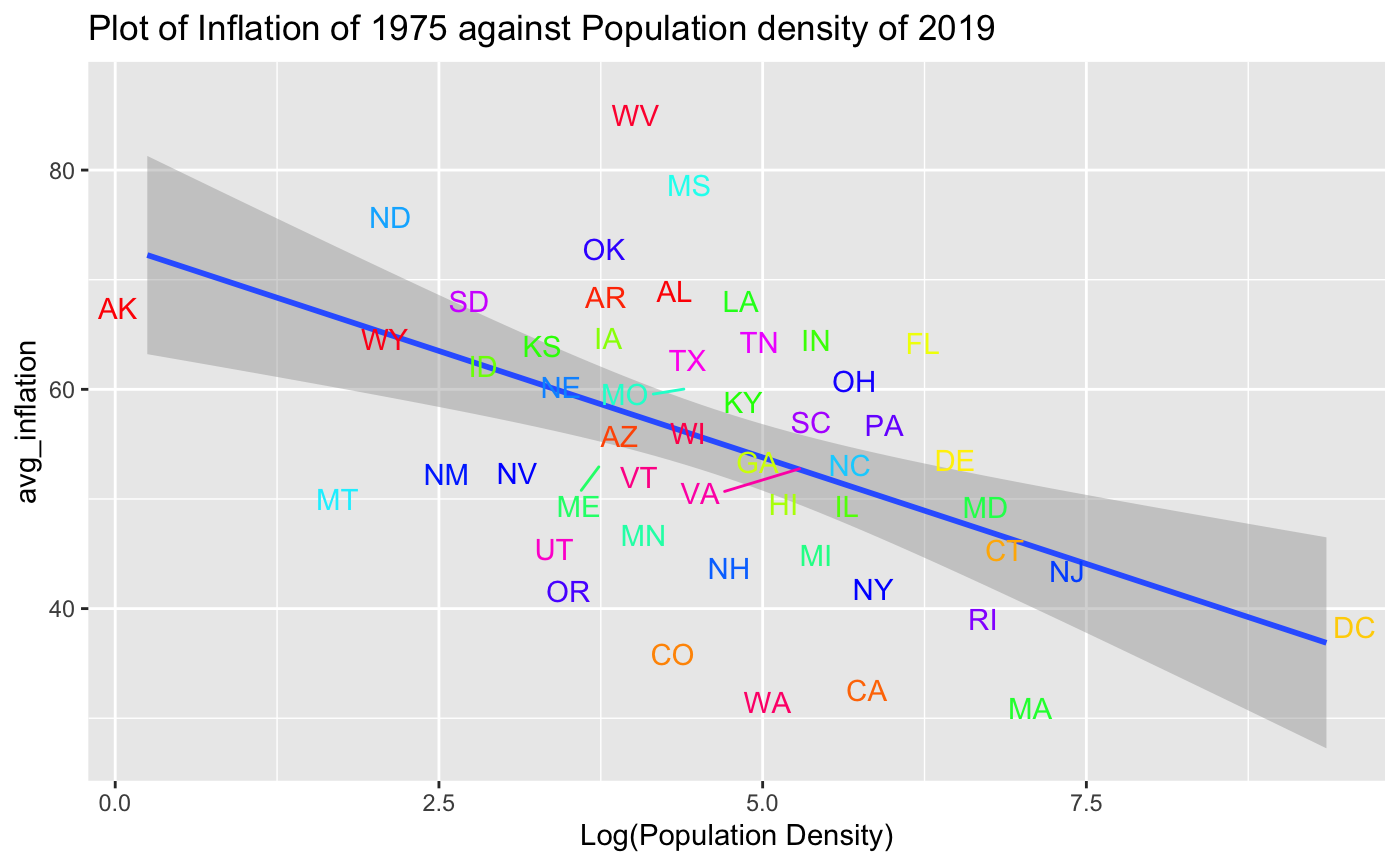
As we can see from the regionwise graphs drawn, the Northeast has the most increase in the house prices(adjusted for inflation) over the 44 years. The house prices in the west also increase by a sufficient amounnt. However, as seen from the above graphs, the house prices in the midwest fluctuate a lot and their starting and the ending point remains almost the same. The prices in the South also fluctuate a lot, but increase a lot in the final years of our study. Regarding the West region, it also goes on increasing till a certain year and then follows a hill climb curve, which means it fluctuates from high price to low price.

## Part 2

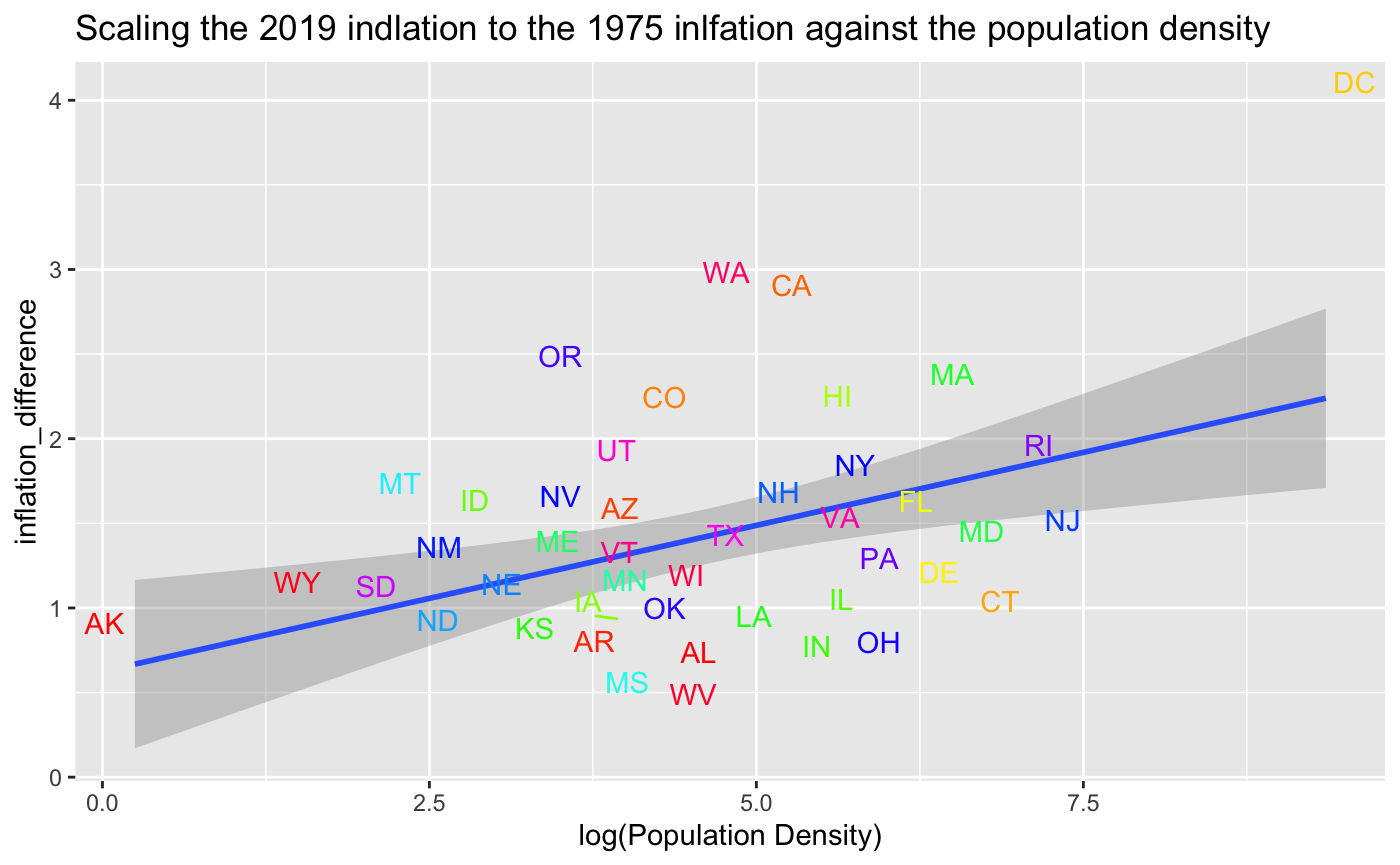
In this part we focus on the variations of house prices with changes in population density. We consider two specific years in this case: 1975 and 2019. We basically compare the house prices of 1975 against the population density of 2019 and the house prices of 2019 against the population density of 2019. We then scale these two house prices( take a divison ) and then plot them against the population density to get a deeper understanding of the variation of house prices against the population.



As we can see from the above graphs, the values have changed a lot from the the 1975 graphs. In the 2019 plot of the house prices against the log of the population density, we see some clear outliers such as the District of Columbia and the states of Ohio and Indiana. As seen from the question 1, we know the the house prices adjusted for inflation of DC increased by the most as compared to the other states. The linear model basically helps us identify the states which are outliers.

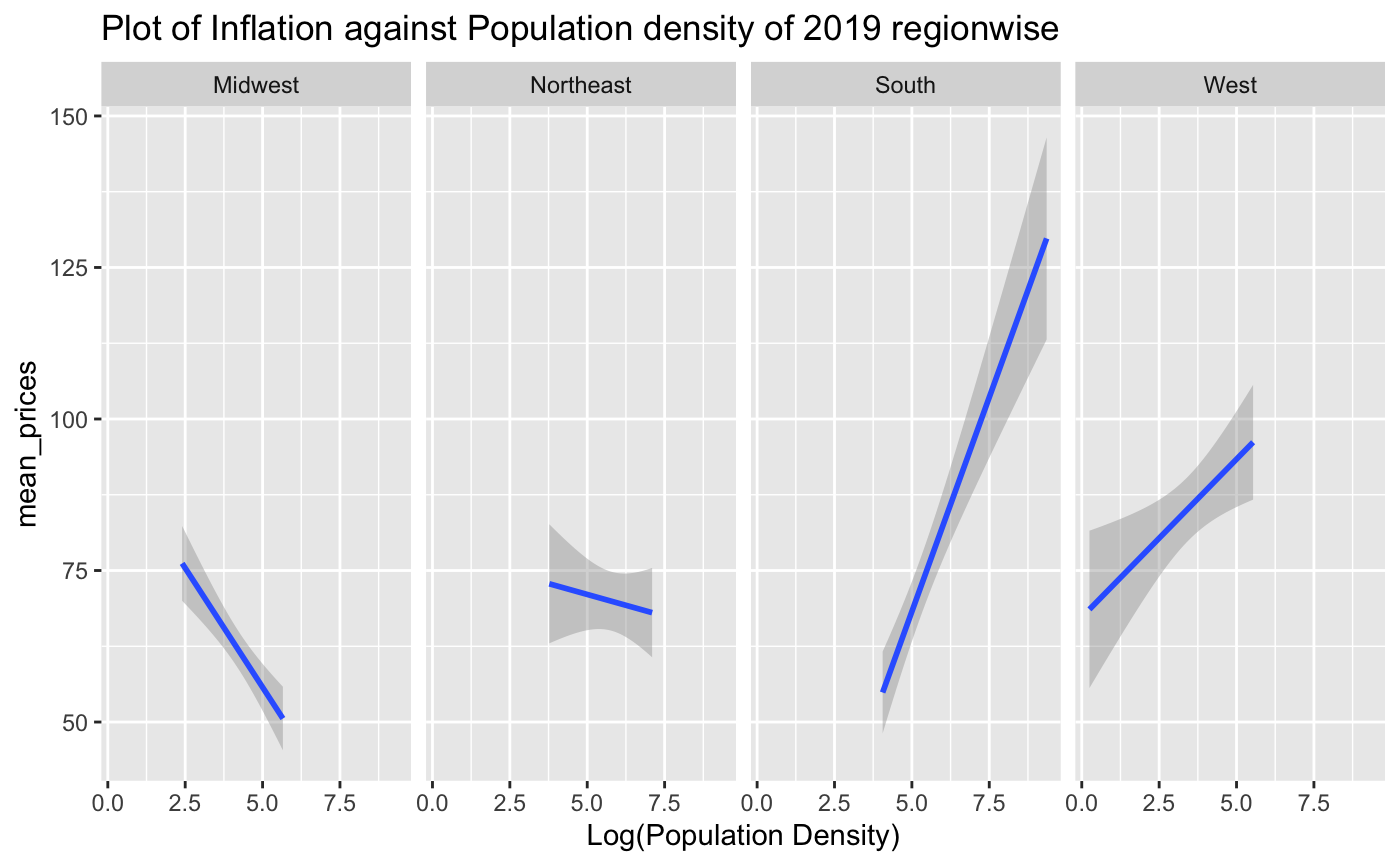


In the graph of the 1975 house prices against the log of the 2019 population density, the clear outliers are Massachusetts, West Virginia and the state of Washington. As seen from the question 1, these values are low at the start in 1975, which makes sense in this graph as they are outliers.

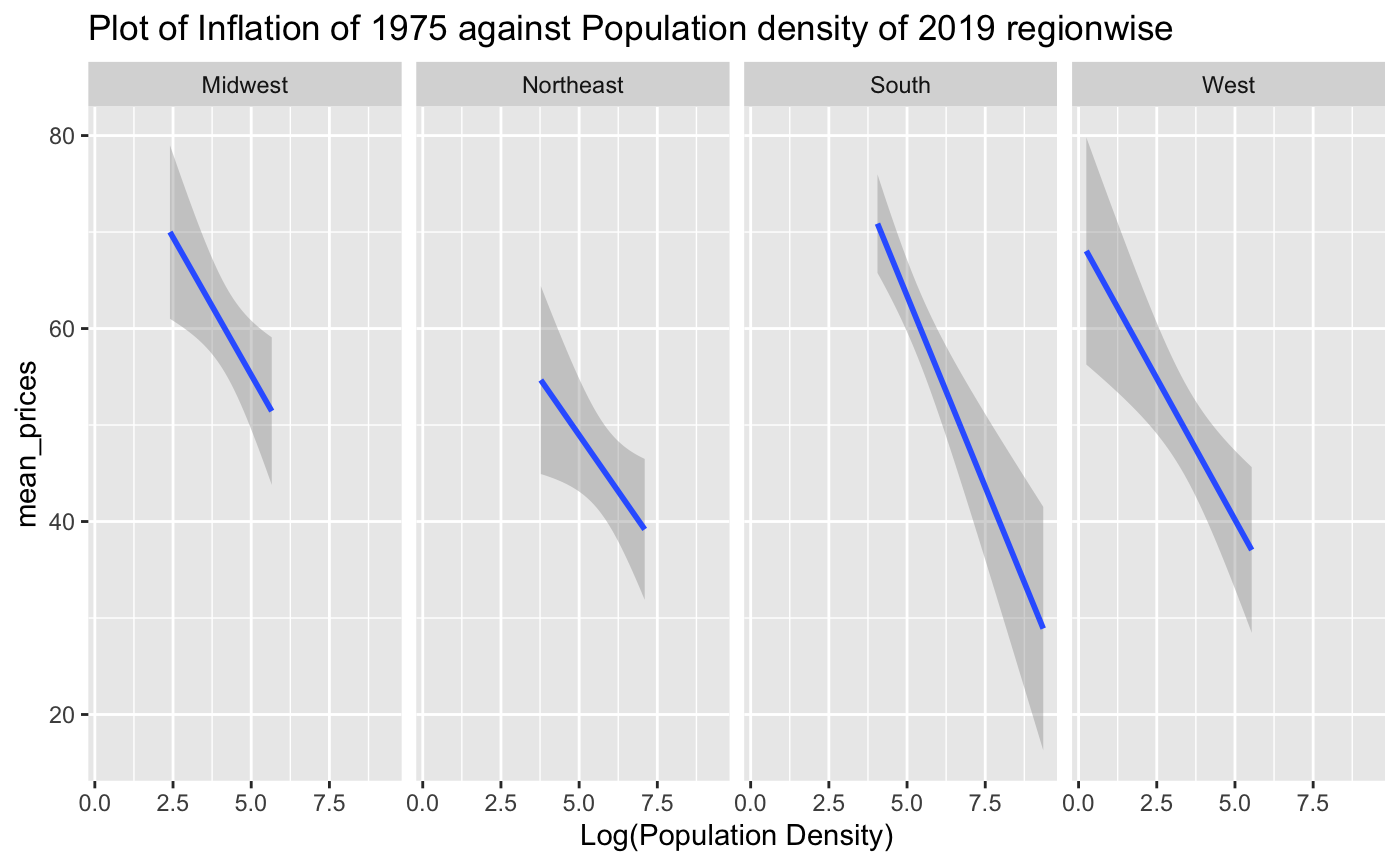


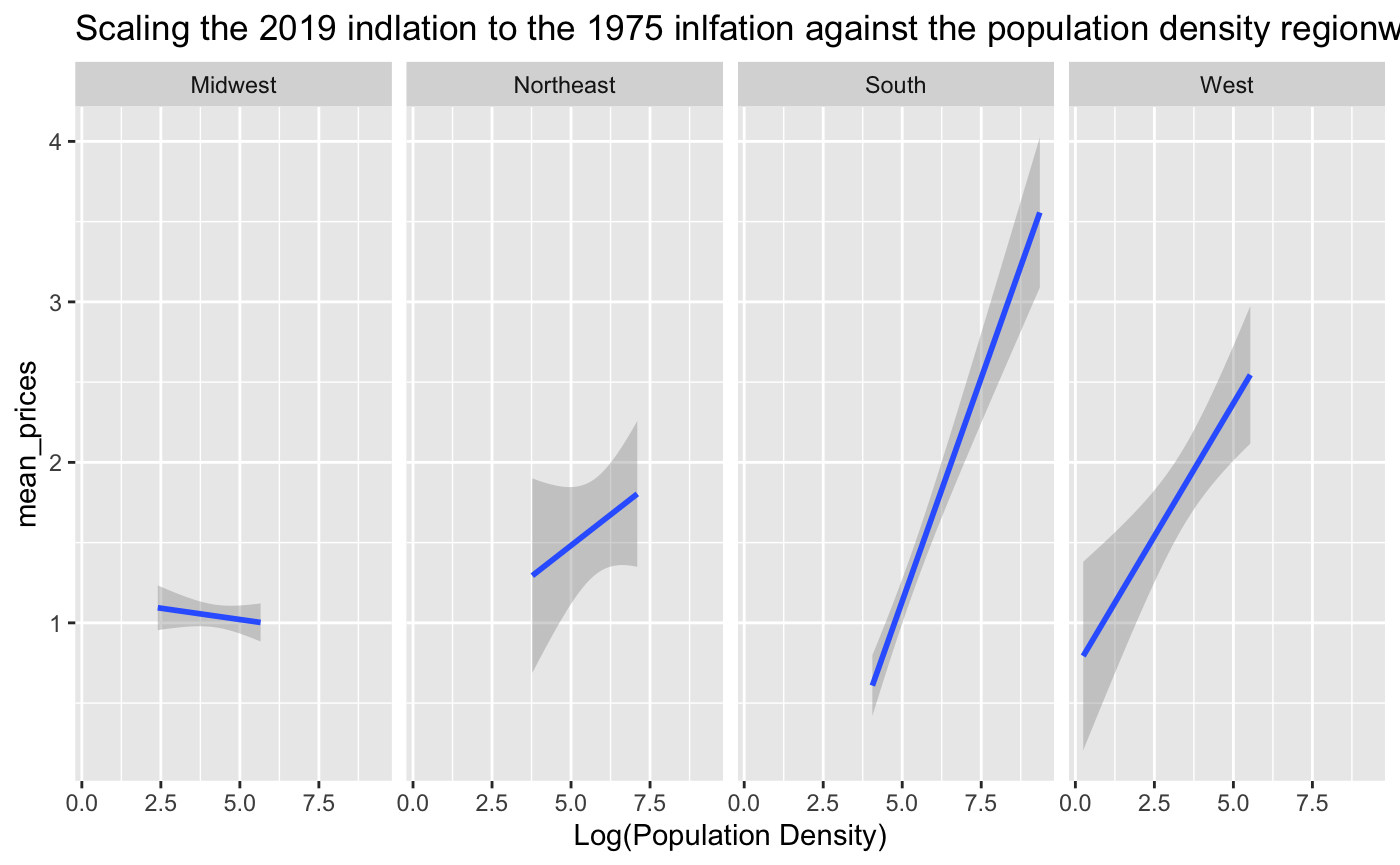
The above graph is a division of the 2019 values to the 1975 values to evaluate the differences between the two years. We can see in this graph and can reconfirm what we concluded in the first and the second graphs that the states of DC, washington and the California are outliers.

## Exploring the relationship between house prices and population density Regionwise



We can conclude some obvious things from the above faceted graphs. In the midwest(2019), the mean prices generally reduce with increase in the population density, th Northeast stays mostly constant. Both the South and the West have sharp increases in the house prices with increase in the density

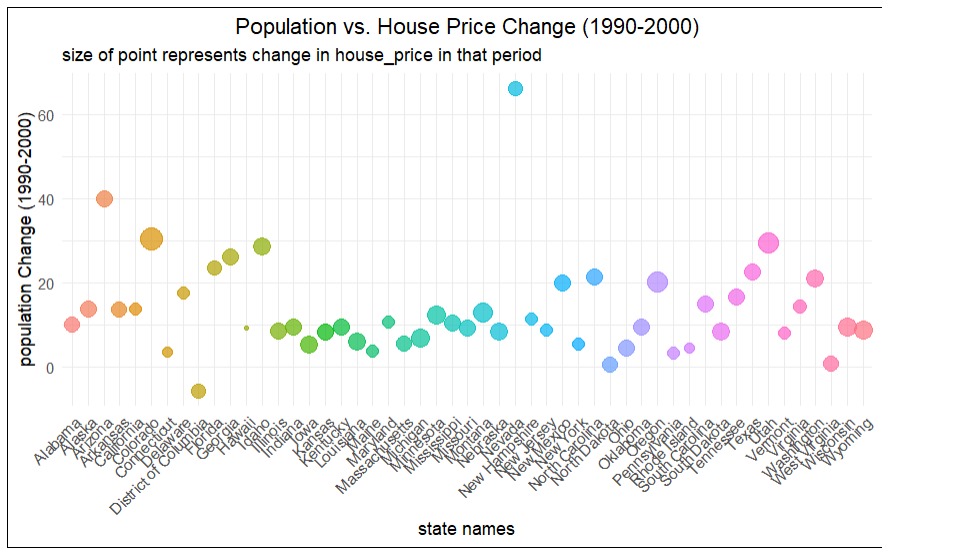


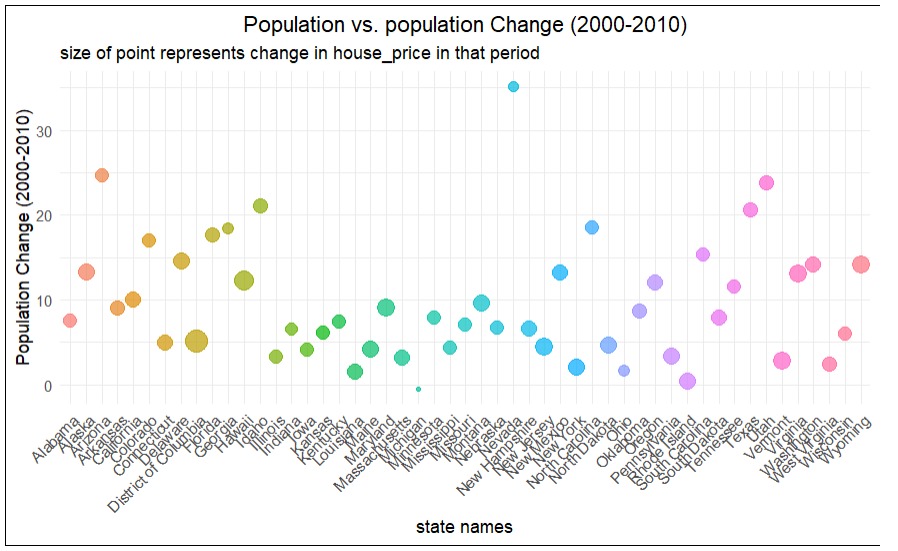
Similarly, in this plot we can see that for 1975, most of the house prices were decreasing with the increase in the population density for that particular year. 

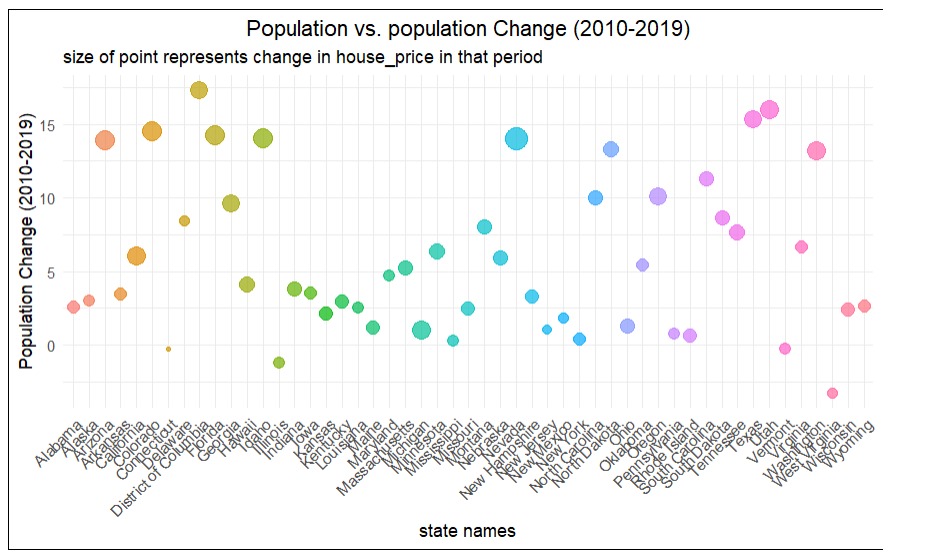
This is a scaled graph between the prices of 2019 proportionate to the prices in 1975, it basically shows us the gradual change in the house prices all over the years. As we can see, the midwest and the northeast regions barely have any considerable change in the house prices. However, the South and the West regions, mostly the south have a very steep increase in the house prices. This generally might be because of the Computer Industry boom that has taken place since the last 50 years in the Western States such as California and Washington(Seattle).

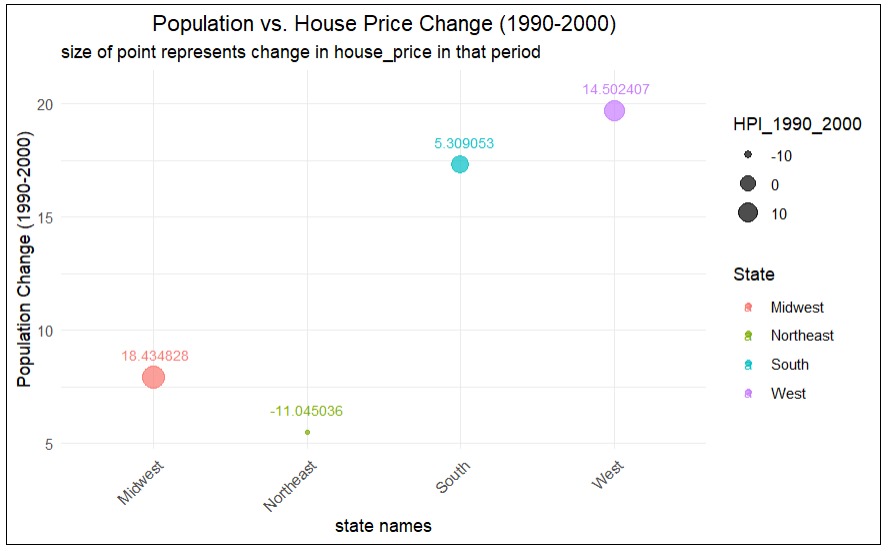
## Part 3

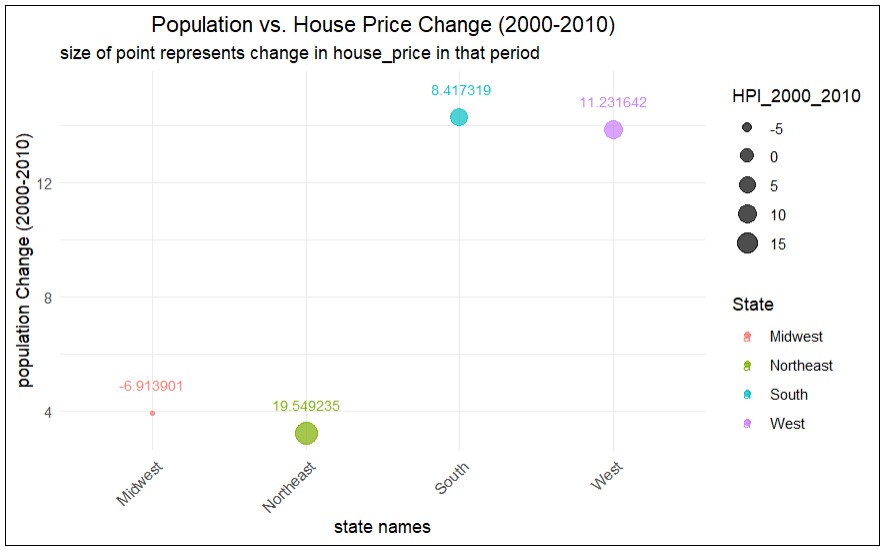
In this part, we explore the relationship between the changes in population over the different years, 1990, 2000, 2019 and 2019. In this problem, we have plotted scatterplots where the size of the dot signifies the change in house prices. We also analyse these relationships over the different regions as well as states.

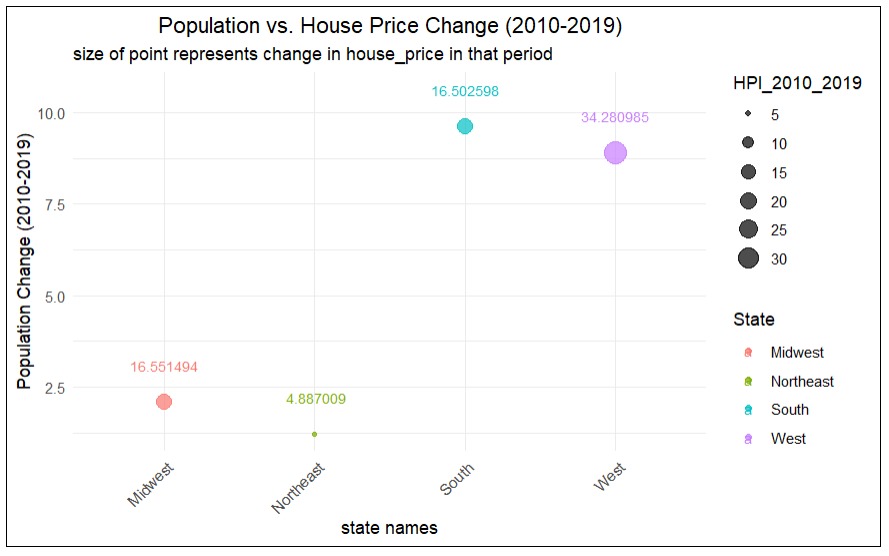












## State wise

For few states there is relation between population change and house price index. Nevada has more than 60% increase in population but house prices didn’t change much between 1990-2000. In 2000-2010 also the population change is very high ,but house price change is very negligible in 2010-2019 Nevada has competitive increase in population and the house price changes are also very high. it seems that for Nevada there is continuous increase in population while the the house price is not much affected in first two periods but increased a lot in 2010-2019 period. Colorado has highest change in house prices where the population change is with around 30% in 1990-2000 In 2000-2010 population increase is average with 15% and house price increase is also average. In 2010-2019 high change in population and house prices are also increased a lot. for Colorado itseems there is trend between the increase in house price and increse in population in all three time periods.

## Region wise

The West region has 20% increase in population in 1990-2000 and 14% increase in house price in 2000- 2010 around 14% increase in population and 11% increase in house price and in 2010-2019 9% increase in population and 34% increase in house prices. The house prices and population increased constantly in all three periods. based on Regions of states , considering northeast region it has low population increase in all three periods, but for house price the time period 2000-2010 has highest change in percentage. The west Region has good relationship between population change and house price change in all three periods of time

## Part 4

## Conclusion

As discussed in the above parts, the changes in the house prices vary according to the different states as well as regions in the United States. For some regions and states like Arizona and Utah (even if the the population change is considerable), we can conclude that the change in house prices is not that dependent on the population or the population density. However, for some states like Nevada and Michigan, the house prices change gradually as population increases and vice versa. But as proved by Part 2 as well as Part 3, some regions like the District of Columbia does follow any of the above relations. As the population or the density increases, the house prices increase and vice versa. But there is one relationship that both Part 2 and Part 3 can conclude. The states in the West region and the South region have a very steep increase in the house prices along with the population. There has been a very steady increase in the population of the cities in California, Washington(Seattle) and Texas because of the industry boom in these states. We can also assume that the tourism industry in Florida has increased in the last 50 years, giving birth to higher revenue, which in turn gives rise to people moving there for jobs and also increases the house prices.