

Exploratory Data Analysis

Mini Project 1: House prices and population growth(S470/670)

Tushar Kant Samantaray, Murali Krishna Parvataneni, Sai Ajay Vutukuri, Venkatesh Thota

Introduction

The project's objective is to understand how the housing prices have varied in the United State from 1975 to 2018 and to understand if the house prices are changing in relation to

1. House prices over time
2. Change in house prices for varying population density
3. Change in house prices for changing population

1. Housing Prices Over Time

We check whether the average house prices across the US have changed from the year 1975 to 2018. Since we know that the prices have increased in general over the year due to inflation, we need to transform the average price of each year taking inflation into consideration to get a fair comparison of prices over the years.

We use the data available from CPI, the Consumer Price Index to combine it with the average house prices of each year from 1975 to 2018 from the Freddie Mac House Price Index. We had considered the baseline year as 1983. To achieve this, we used the following transformation:

$$\text{Average House Price} = \frac{\text{Average House Price}}{\text{CPI Value}} * 100$$

We plot the normalized average house prices of each year, across the US, from 1975 to 2018, in **Fig. 1**.

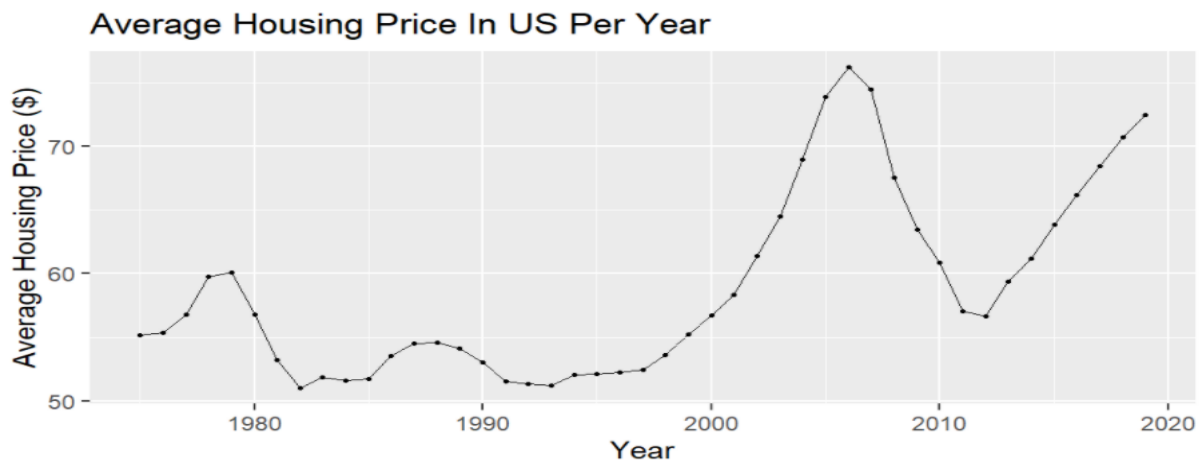


Fig. 1

From the plot in **Fig. 1**, we observe that the prices decreased from 1975 to 1997, after a small increase in between the late '80s to early '90s, but then the average house prices constantly increase at a huge rate till 2007. We see a major drop in the prices from 2007 to 2013 due to the housing crisis, and the prices seem to be increasing rapidly again since 2013.

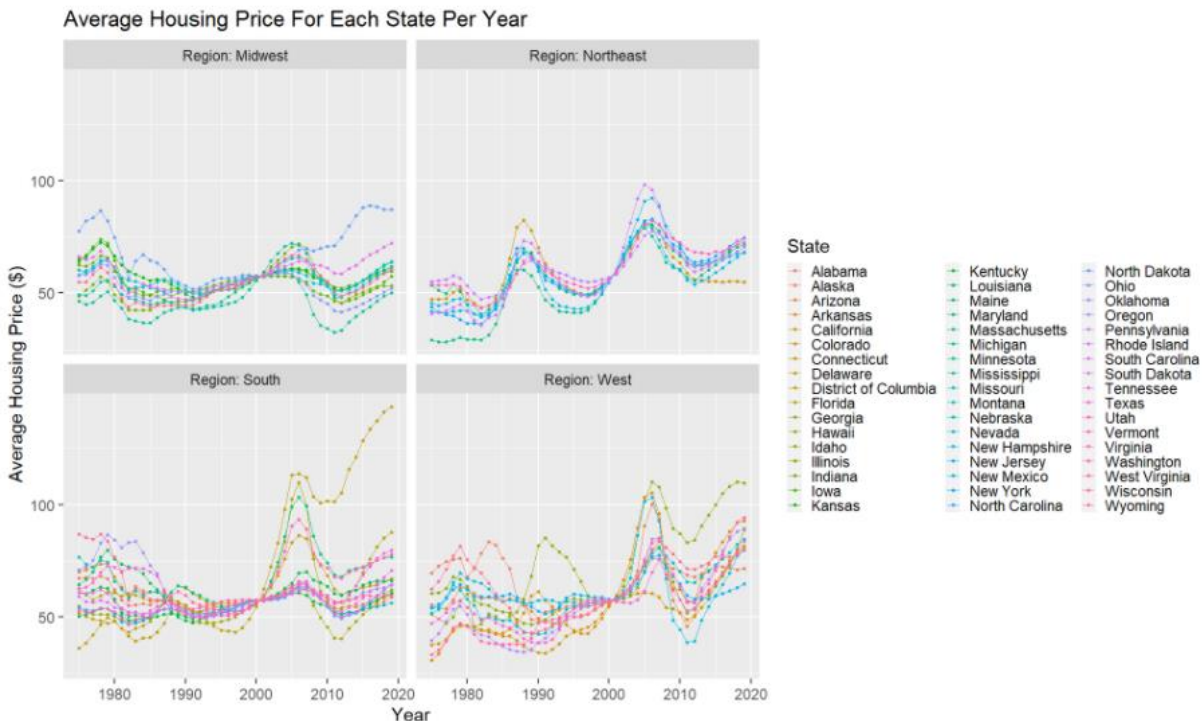


Fig. 2

In **Fig. 2**, we observe the change in the average prices in each state of the 4 regions in the US. To see that, we create a plot of average price change of each state faced by the region, since it is clearer than a single plot of all the states.

Fig. 2 shows the change in average house prices from the year 1975 to 2018 for each state in the US. We observe that the house prices have increased for most of the states, although the increase in a few states, such as the District of Columbia, California, Hawaii, and Washington having the most increases respectively. The District of Columbia has the biggest increase in housing prices over the year while the biggest decrease is seen in the state of West Virginia, followed by Arkansas.

Most of the states have followed the overall trend of change in average prices, with few exceptions such as North Dakota, which has a concave-shaped trend, and the District of Columbia which sees a surprisingly high increase in house prices.

We also observe that the changes in average housing prices vary between regions. The change is less in most Midwestern states and southern states in comparison to the northeastern and western states, with the District of Columbia being the most obvious exception.

Plotting the average price per year of each state by region gives us more clarity regarding the changes in prices per region and is shown in **Fig. 3**.

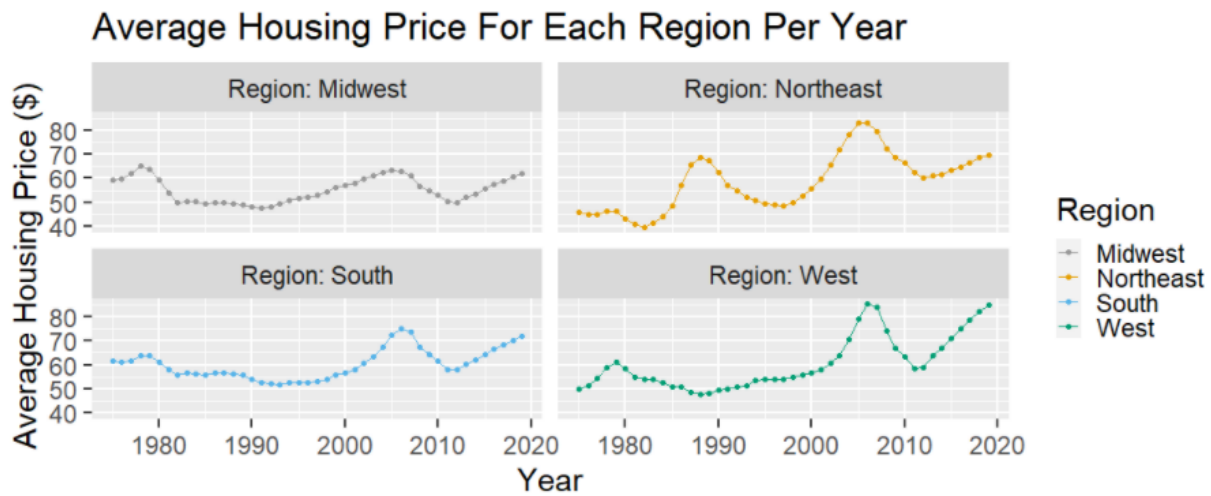


Fig. 3

From the above plot, it is evident that the average prices of each region have indeed increased. However, the West and the Northeast regions have increased more as compared to the South and Midwest regions, which have the lowest increase. Then price change trend of the West region is most close to the overall trend seen in **Fig. 1**.

2. Population density and changes in house prices

In the previous section, we see that the average housing prices across has indeed increased. Since it is a general understanding that house prices go up as the population density increases since it could mean that there is more employment or other opportunities that makes that region an attractive location for people to migrate or immigrate to. Let's see if the change in population density affects the change in average prices of housing.

The following plot in **Fig. 4** shows change in population density vs the percentage a change in house prices. To find the percentage change in average price housing from 1975 to 2018, we use the following formula:

$$\text{Percentage change in Average House Price} = \frac{\text{Sum(Average House Price}_{2018}) - \text{Sum(Average House Price}_{1975})}{\text{Sum(Average House Price}_{1975})} * 100$$

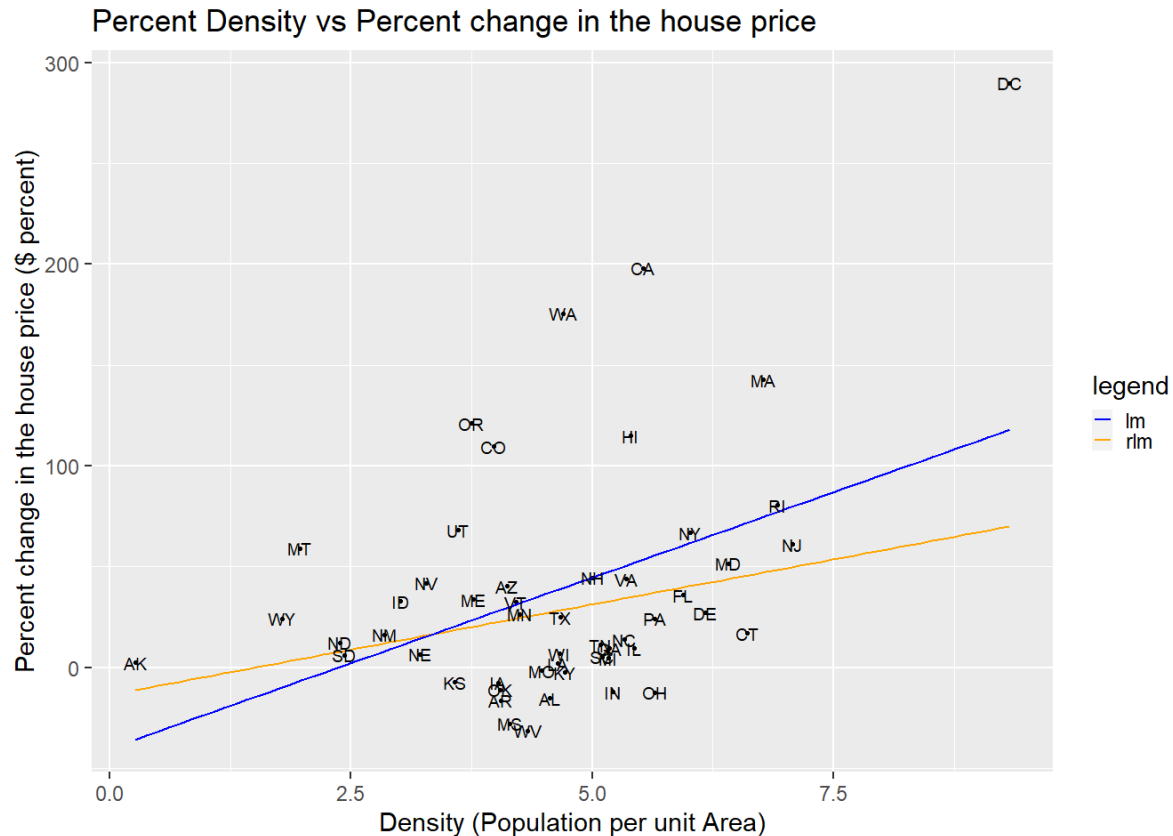


Fig. 4

From the above plot, we can notice that most of the states lie in the region where density change is in the range of 2.5 to 7.5, and the percentage in the house price is in the range of -50 to 100. District of Columbia is a clear and obvious outlier since it lies in the top right corner far away from other states, and D.C. is also not a state which follows different laws and trends in compared to states, and since Washington DC is the capital, it is reasonable that the prices are high. We have added an lm and an rlm trend line which shows us that the trend is heavily affected by outliers since there are about 4-5 states that should be below the average percentage change in house prices as the lm trend line is affected by the outlier. Further, the outlier might be significantly affecting the trend line and suggesting an increase in overall change in price with increasing density, which might not be true. Hence, we could remove the District of Columbia and create the plot again to get further clarity if the average price is increasing or not.

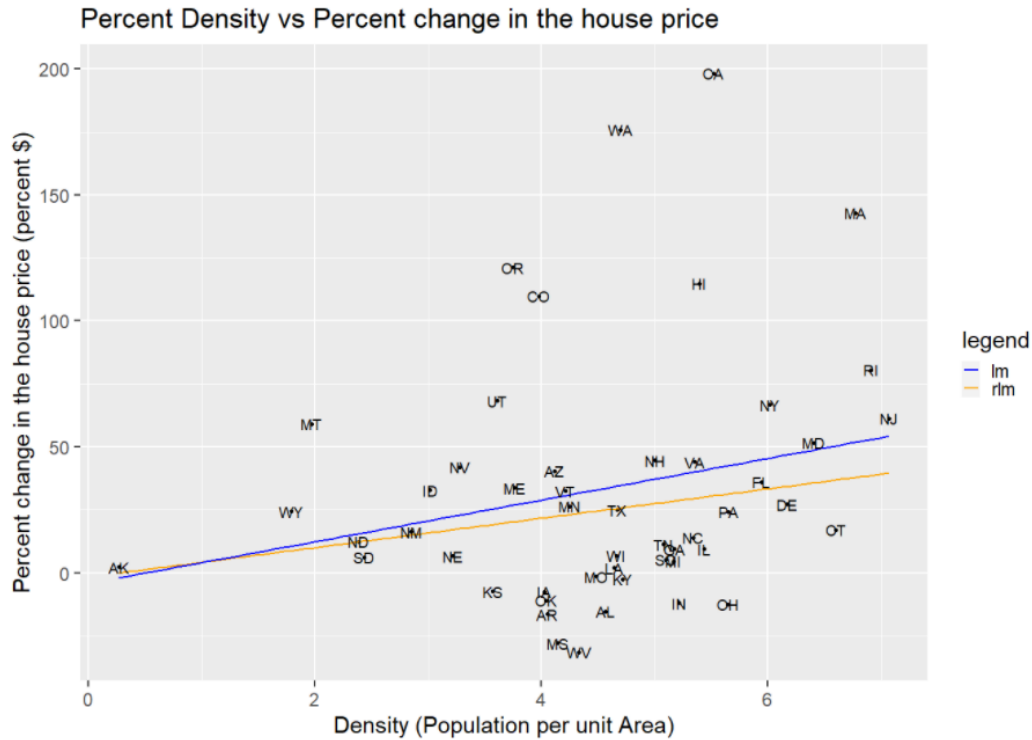


Fig. 5

After removing the outliers, we get the above plot shown in the above **Fig. 5**. Now, we can see that the increase seems to be weak, but as the density increases, average price of housing also increases, even without the effect of outliers.

Now, let's see if the same trend is seen in each region of the US. We could do so by creating a plot faceted by region as shown in the following **Fig. 6**.

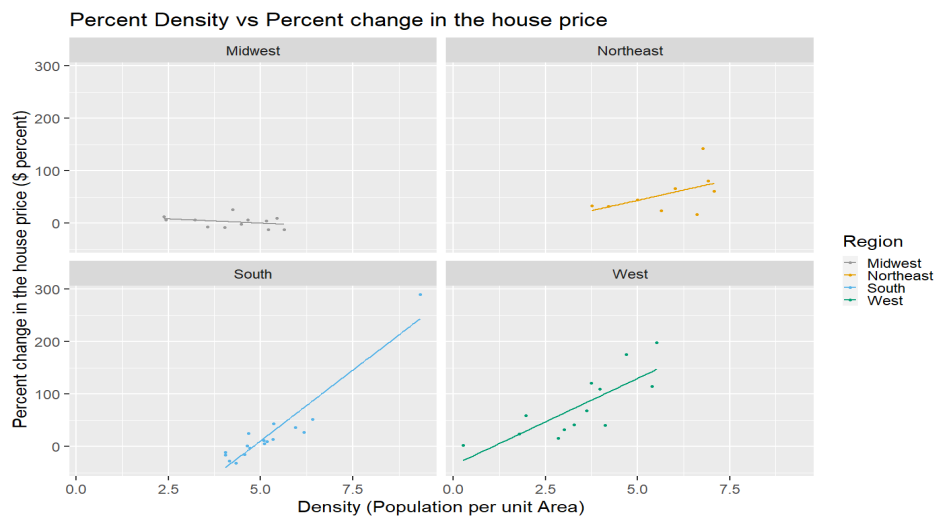


Fig. 6

From the above plot, we could infer that the average housing prices of the regions have increased except the Midwest region with increasing population density. The South region has the highest increase in average price, followed by the west, and then Northeast, and the average housing price of the Midwest region has rather decreased. It is noteworthy that the outlier District of Columbia is included in the above plot, and since D.C. has about 300% increase (from **Fig. 4**) in average housing prices with an increase in density, the South region seems to have the higher increase in average housing prices in comparison to the west.

3. Changes in population and changes in house price

We could now verify where the change in population influences change in price over the time periods 1990 to 2000, 2000 to 2010, and 2010 to the present.

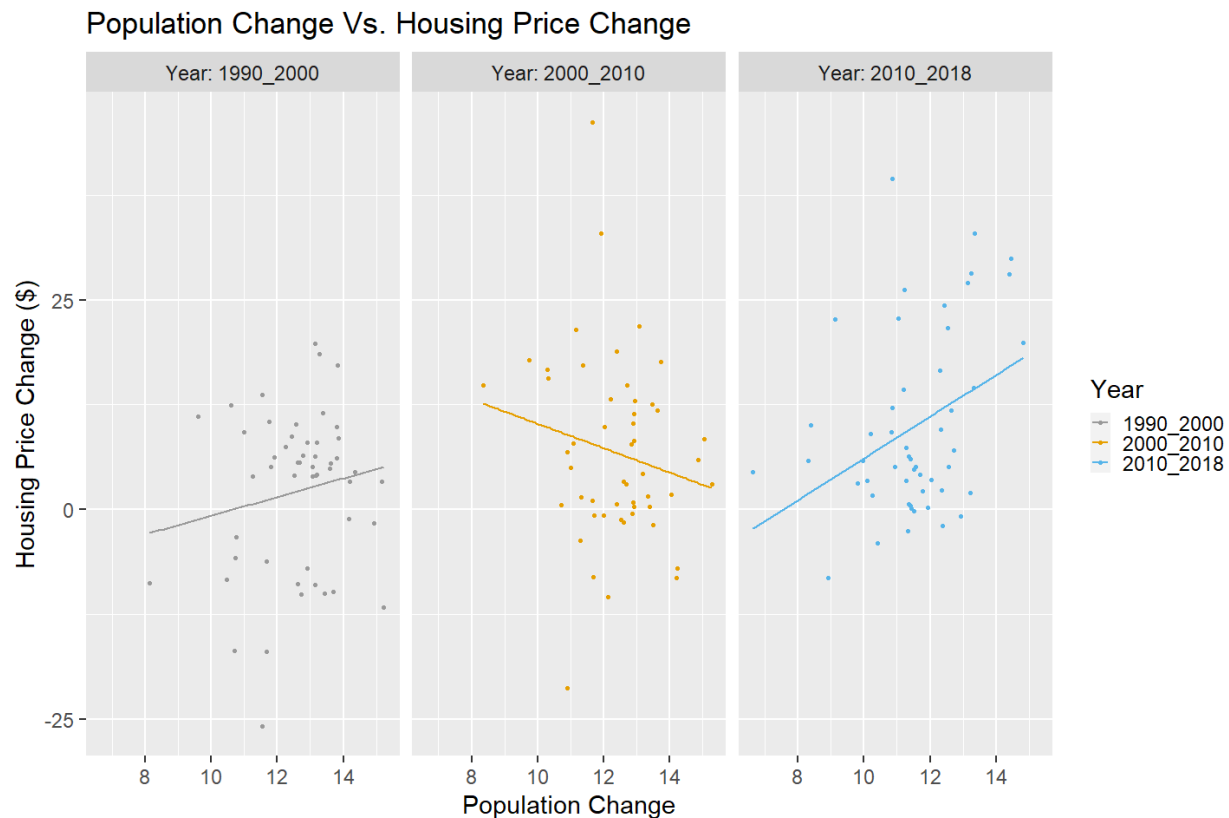


Fig. 7

From the above plot in Fig. 7, we notice that the population changes are directly proportional to the increase in the house price for the time periods 1990-2000 and 2010-2018, and surprisingly, we can observe that 2000-2010 shows an inverse proportional this might be because of the economic crash.

Further, we could see the change in the average housing prices with the change in population over the same periods by each region in the US. We achieve this by plotting the change in average house prices against the population faceted by region.

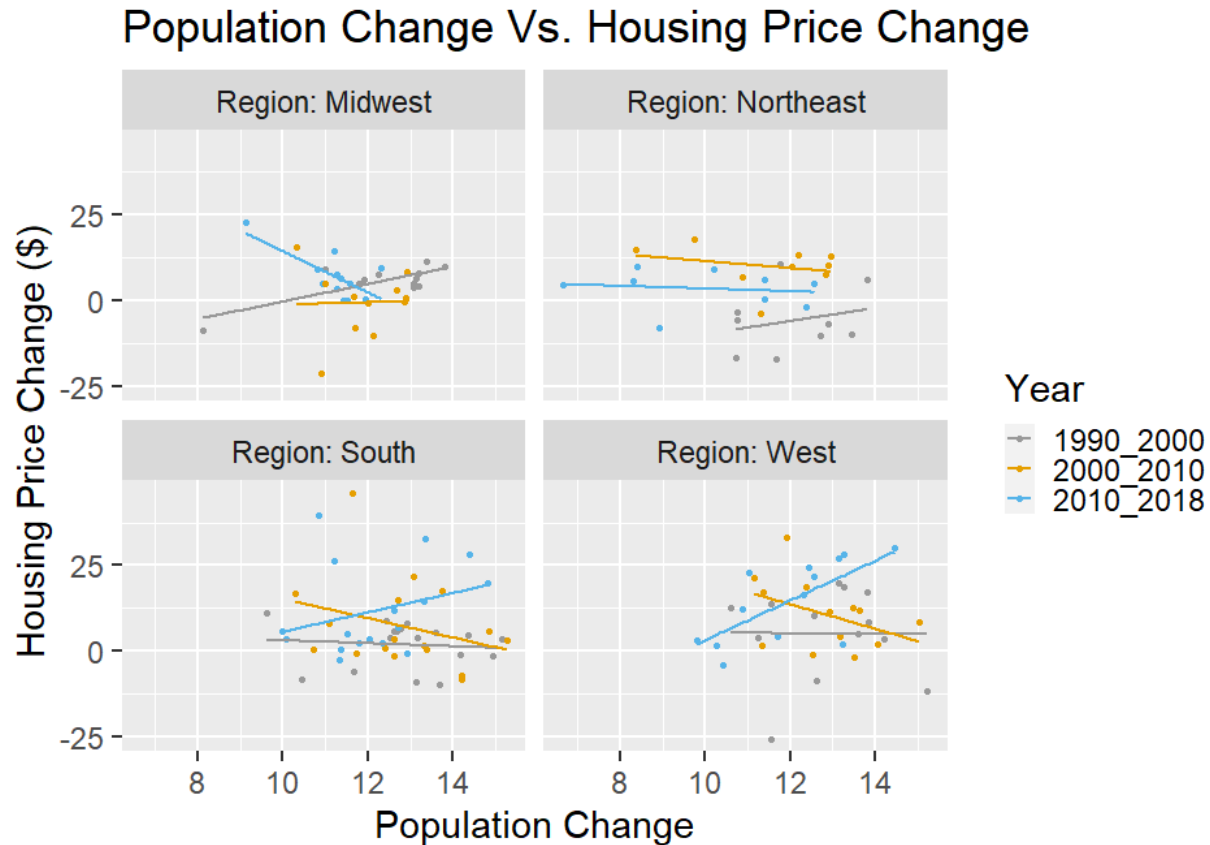


Fig. 8

From the above plot in **Fig. 8**, we see that the average price increased from 1990-2000 for the Midwest and remains almost unchanged from 2000-2010, while it dropped significantly for the 2010-2018 period. The housing price of the Northeast region only increased during 1990-2000 period and slightly decreased in the 2000-2010 and 2010-2018 periods. For the South region, the average price slightly decreased from the 1990-2000, decreased by a good margin from 2000-2010, and increased rapidly for the 2010-2018 period. Finally, in the West Region, the average prices almost stayed unchanged for the 1990-2000 period, dropped for 2000-2010 period, and increased hugely during the 2010-2018 period. From these inferences, we can say that there no set pattern between the population changes and housing price across regions over the years

Conclusion

Although there has been a significant rise and drop in the prices over the years, the average house prices of the US have increased over the years. The District of Columbia has seen the biggest increase in real house prices while West Virginia has the biggest decrease. Most states follow a similar increase and decrease trend except North Dakota which follows a concave-shaped trend, and the price increase in states in the Northeast and West Region saw a higher increase in prices in comparison to the remaining two regions.

Further, we saw that the changes in house prices are affected by the change in population density. Even when the outliers, District of Columbia were removed, the change in average prices still showed an increasing trend with an increase in population density. When compared amongst the region, except the Midwest region, the rest of the regions have increased average prices trend with increase in population density.

Then, we saw that with the change in population, the house prices stayed almost the same for the South and the West region, while it increased slightly for the Northeast region and increased for the Midwest. For the 2000-2010 period, except the Midwest, the rest of the regions saw a fall in house prices, with the West having the biggest fall. The average house prices increased the most for the West region the most for the 2010-2018 period, followed by the South, and almost stayed the same for the Northeast region and dropped hugely for the Midwest.

Though most of the observations seem reasonable, the huge fall in the house prices of the Midwestern region as the population increases for the 2010-2018 time does not seem ordinary, considering that the housing prices are constantly increasing throughout the US since 2013.