S-670 EXPLORATORY DATA ANALYSIS

GROUP 18 - HOUSE PRICES AND POPULATION GROWTH

INTRODUCTION

House prices in the U.S. have changed over the last few decades. A researcher wants to know whether changes in house prices are related to population in some way. Here we are trying to help her out by making plots and analyzing the relationships from the plots. We analyze using house price data, consumer price data and population data since 1975. The recent house price and consumer price data are from 2019, whereas the recent population data is from 2018. Let's see through the following analysis to know if population influences the house prices.

DATA

You will find the following data sources in the analysis below.

-> "State_and_US_SA.xls": This contains average house price (source: freddiemac.com) in each state since 1975, with index value fixed at 100 in December 2000 for each state. We use this data to get the average house price.

-> "cpi.csv": This contains how prices (source: bls.gov) in general have changed in the U.S. We use this data for adjusting for inflation.

-> "nst-est2018-alldata.csv": This contains the population estimate between the years 2000 and 2018. We use this to get the population estimate for 2018. The decennial year population data is obtained using census API key.

-> "land_area.csv": This contains the land area data for each state. We use this data to estimate population density.

-> "state_abbrevs.txt": This contains the names, two letter code and census region for each state. We use this to match HPI, CPI and population data of corresponding states.

1. House prices over time

House price index are adjusted to December 2000 values using consumer price index to get real HPI values. This standardizes house price index for inflation. The below figure shows trend in mean house price index over the years from 1975 till 2019 for the entire U.S.

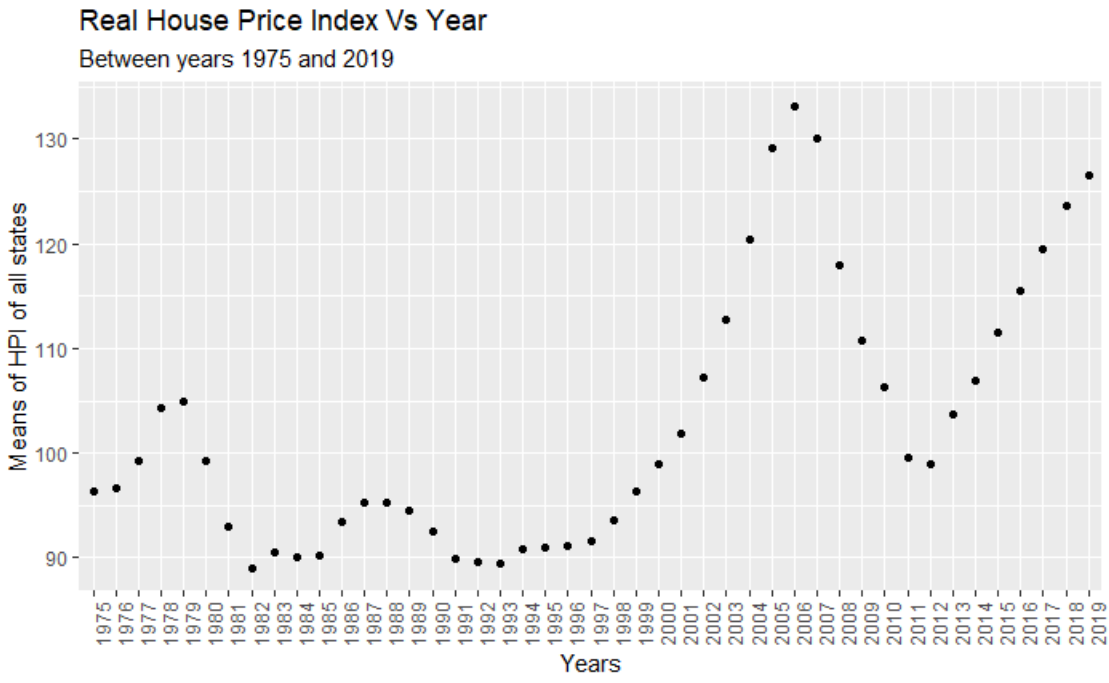

Fig 1

The trend is not monotonous since 1975. House prices have shown small peaks in 1979 and 1988. Later, prices have been continuously increasing until 2007, after which it drops till 2011 and then again increasing till 2019. The highest peak was in 2007. There is no static relationship that we could figure out from the above plot.

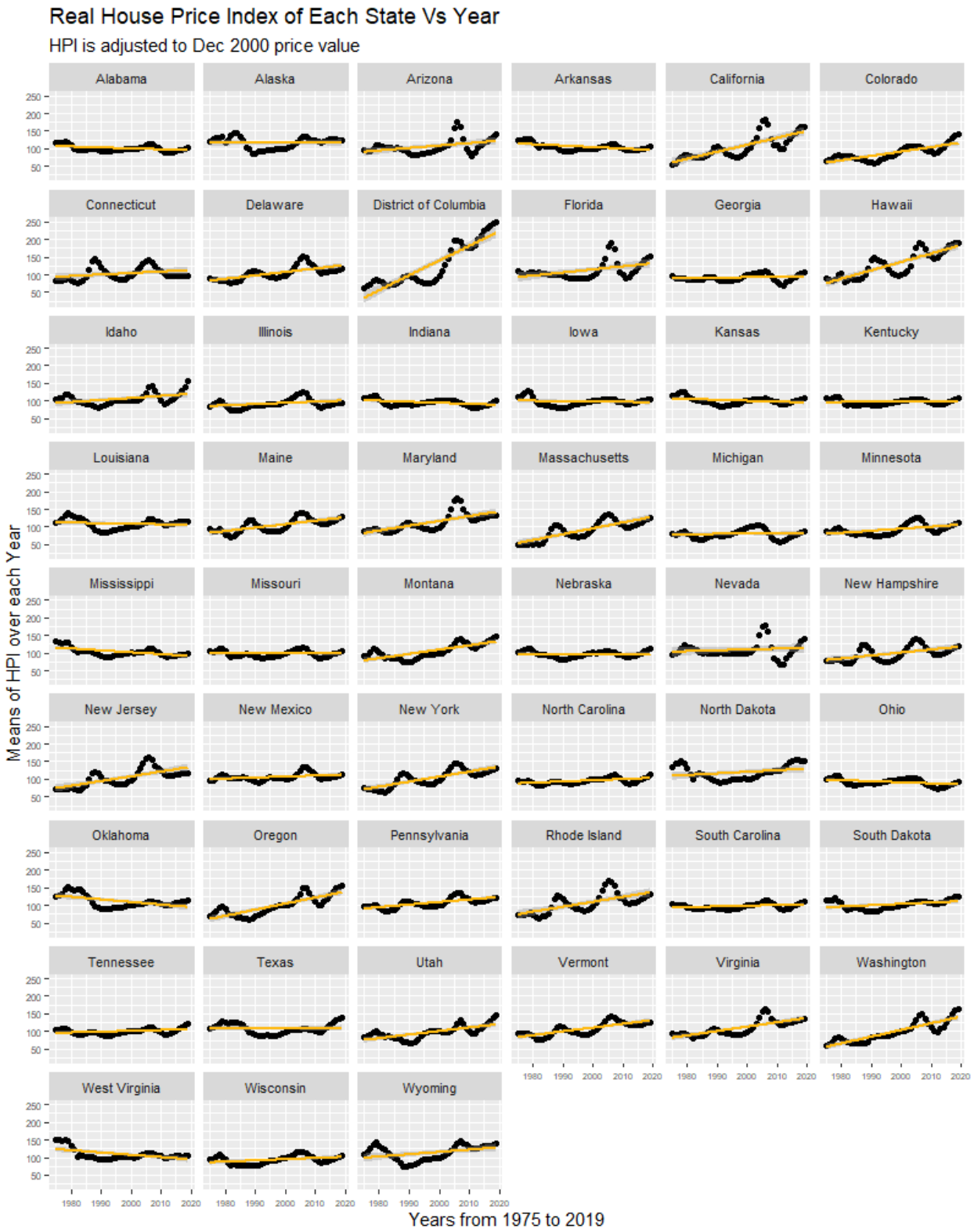

Fig 2

District of Columbia has the most increase in real house prices whereas West Virginia and Mississippi have got the most decrease in real house prices over time since 1975. In West Virginia and Mississippi house prices have fallen after mid 1980's and have continued to be constant since then. New York and New

Jersey look like they have similar trend in house prices. In states like Arizona, California, Florida, Maryland, Nevada, we can see some steep increase in the house price trend compared to other states.

Similarly, we can look at these price variations in regions, before plotting we average out HPI for all states in a region for every year.

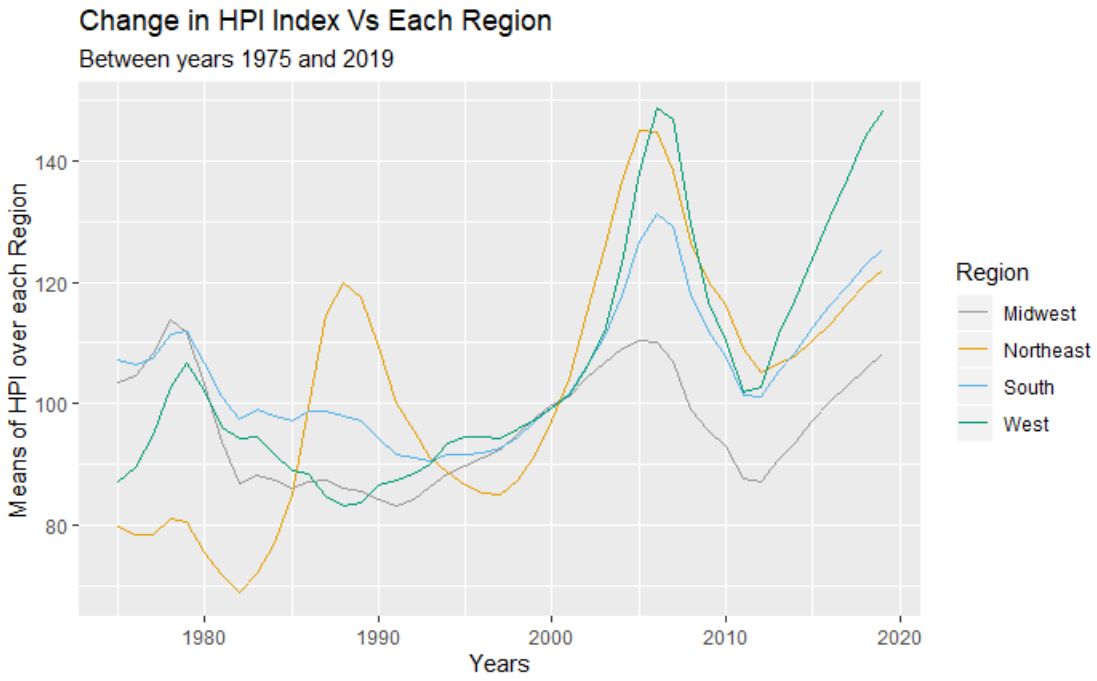

Fig 3

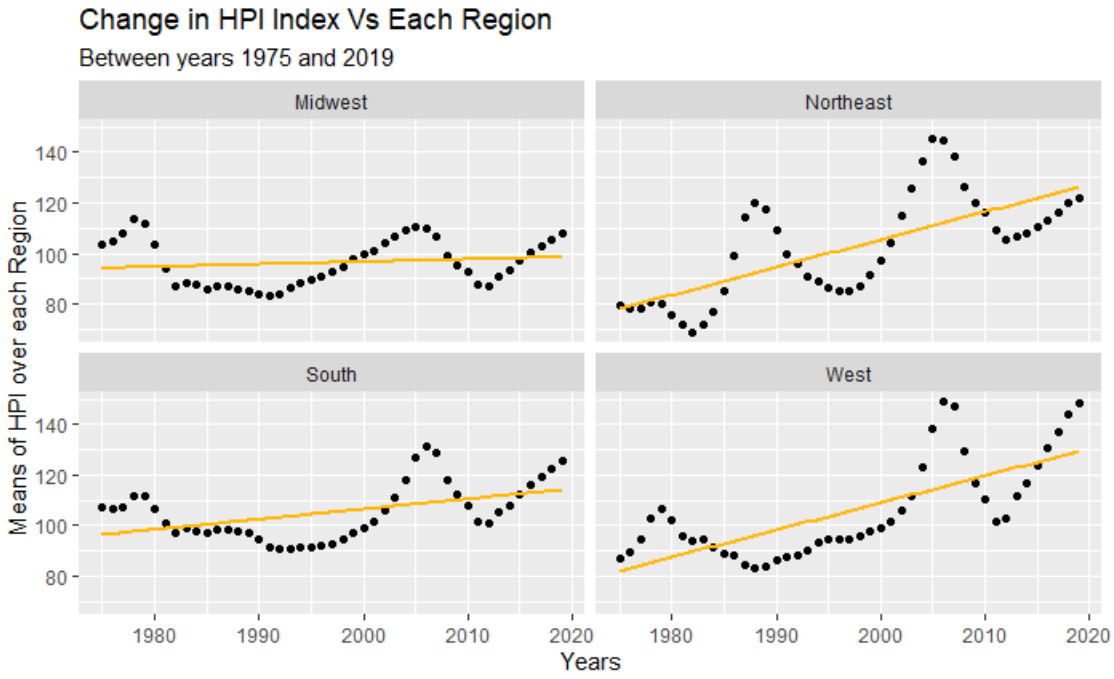

Fig 4

When we look at change in house price index for each region from 1975, West has seen a biggest increase over time from 1975. Except for Midwest region all the other regions have seen the peak of house prices between 2006-2008 period, but Midwest has seen its peak before 1980's. Mid-west region seems to have almost constant HPI over decades. We could observe a peak in prices in northeast region just before hitting 1990s.

2. Population density and changes in house prices

We will observe the changes in house price index with respect to the population density of the states and later go ahead with the region wise trend.

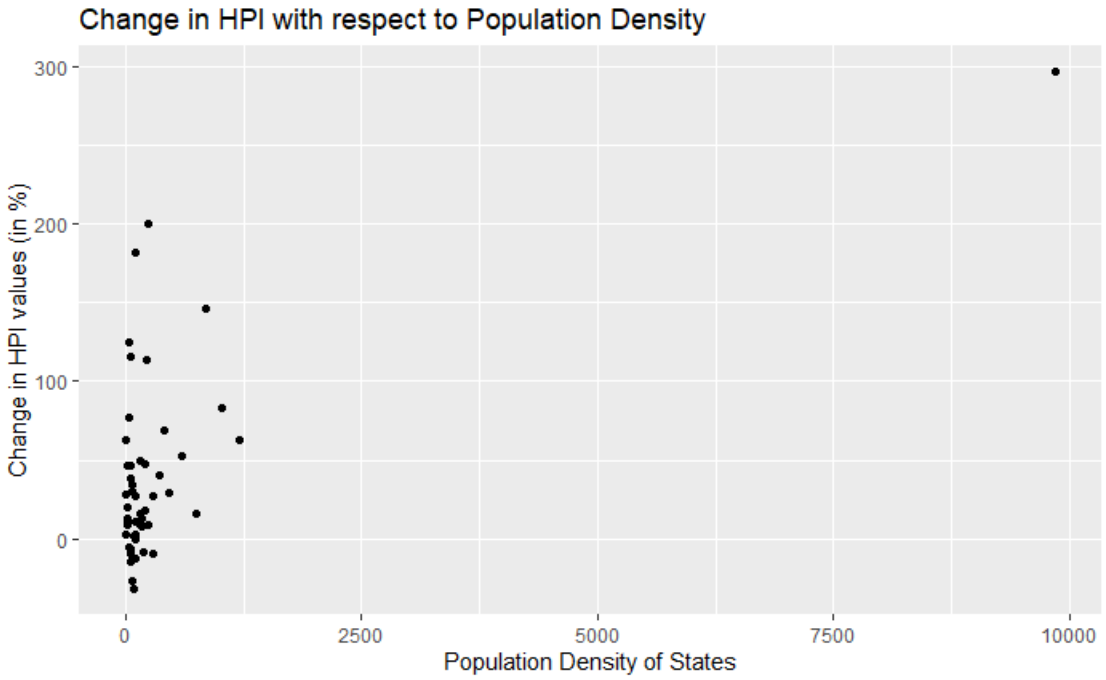

Fig 5

There is no clear understanding of the change in HPI values with respect to the population density due to the presence of outliers. We are unable to conclude anything from the above plot. District of Columbia has density: 9856.5 and change in HPI: 296.5837995 which is not as often as we observe in regular states. Being the capital may be the reason for the unusual value and we go ahead and eliminate it while comparing the overall states.

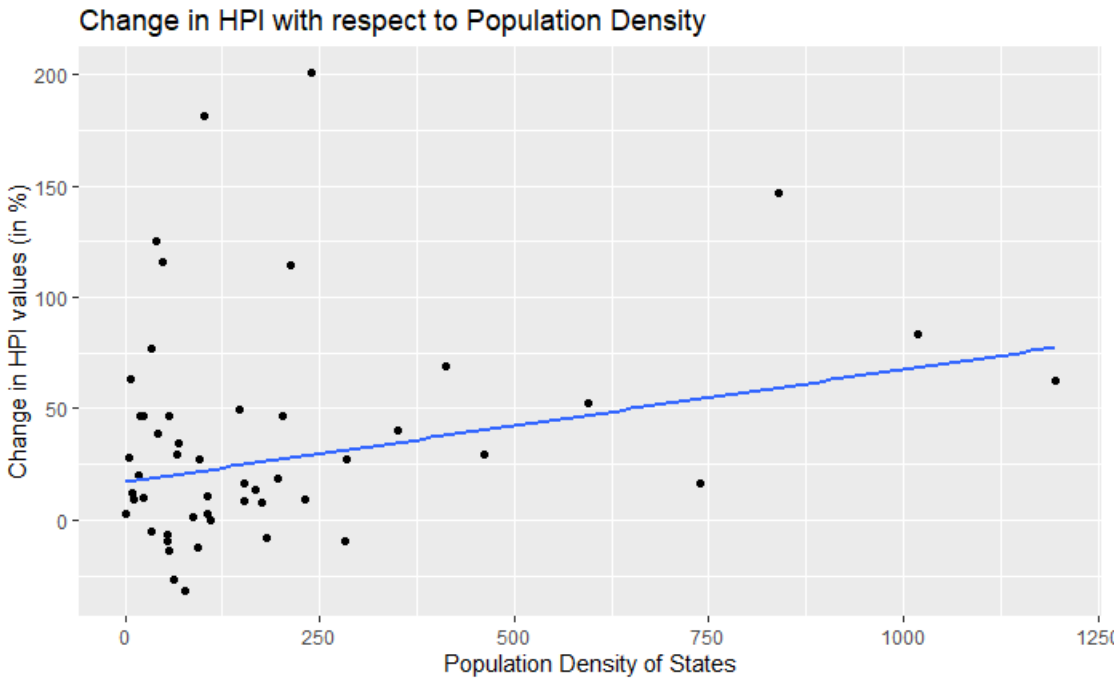

Fig 6

By removing the data points having density 1250 we are left with enough data points to observe the trend of change in house price index with respect to the population density. We can observe that the house price index has an increasing trend with respect to the density.

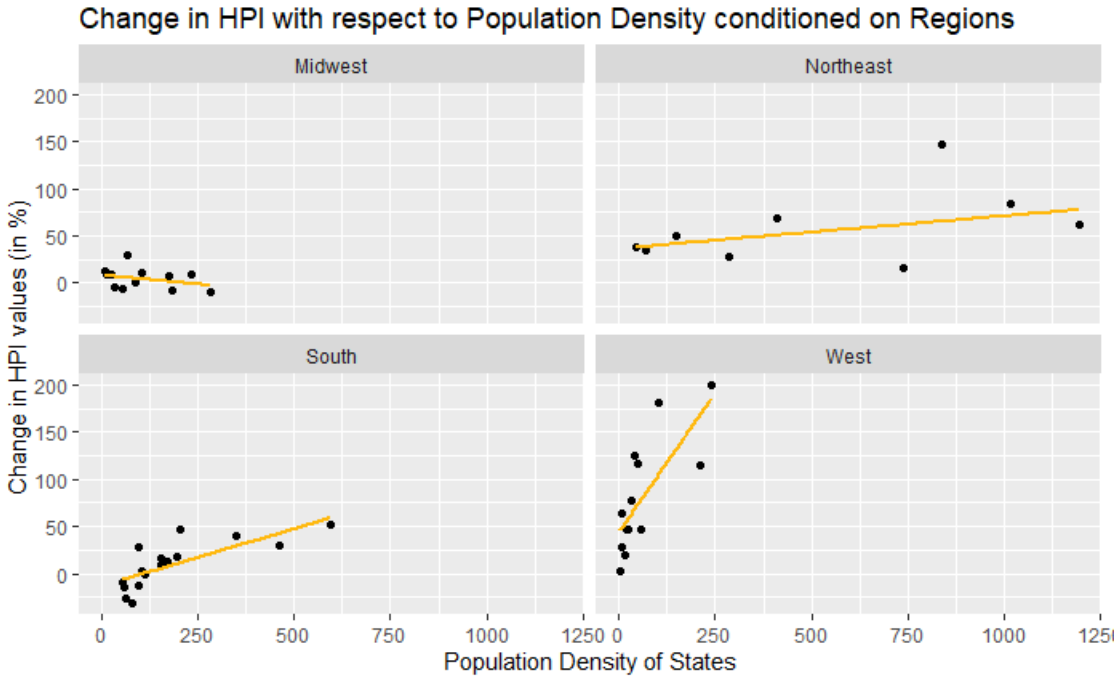

Fig 7

States in West region have seen the steep increase in HPI over time. The steep could be because of a few data points that are near 200. Northeast states have varied population density compared to any other region. Mid-west has very less change in HPI compared to any other regions. South states have an increasing trend in the HPI with density over time.

3. Changes in population and changes in house prices

We will observe how house price index is influenced with change in the population over the years across the states and regions.

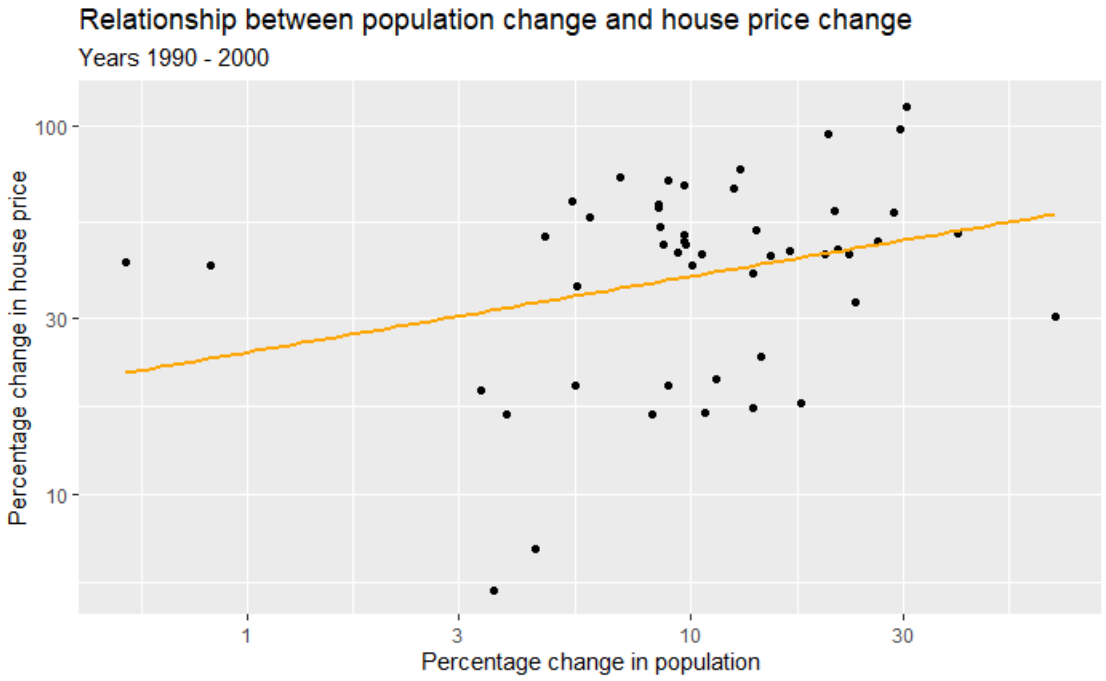

Fig 8

We can see from the above graph that the house price increases with increase in population of the states in the years from 1990 to 2000. This is a positive trend in the relationship between house price change and population change.

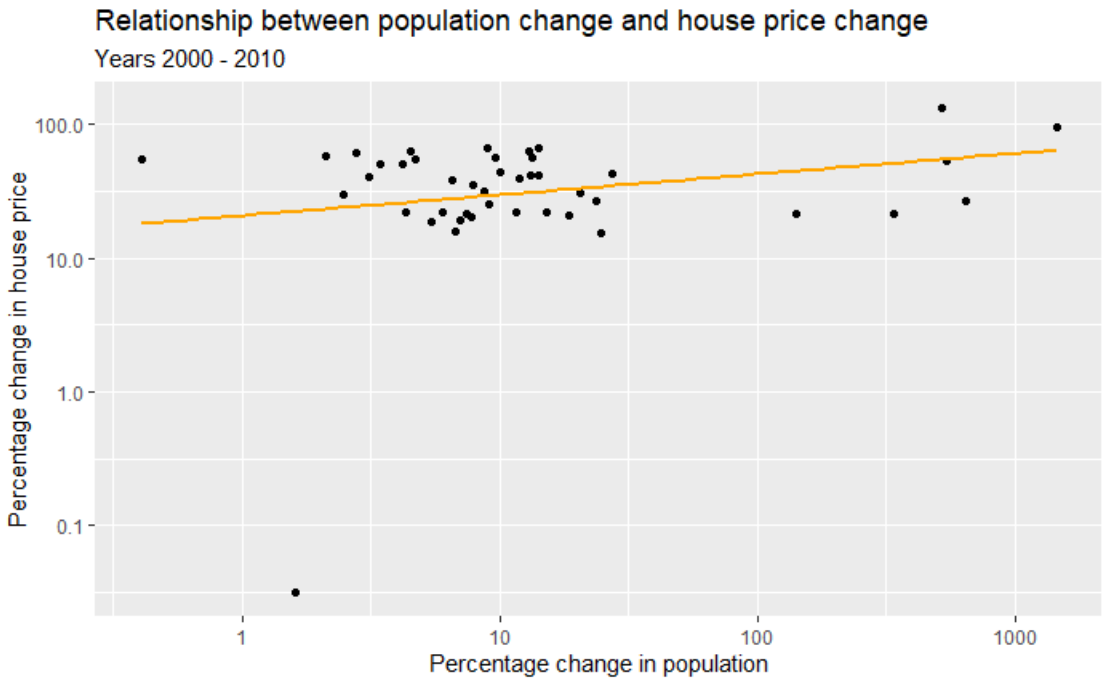

Fig 9

The above figure shows the relationship between house price change and the population change between years 2000 and 2010. Although the increase in the trend is not obvious like the one, we observed for years 1990 to 2000, there is still a linear relationship between population change and house price change. We can see there is a huge dip in the population change and house price change, which causes the loess curve to be pulled downwards when plotted. This dip in the change could be because of the recession that the country faced in 2009. Except for that, most of the data points are closer to the regression line and supports the linear relationship between the variables.

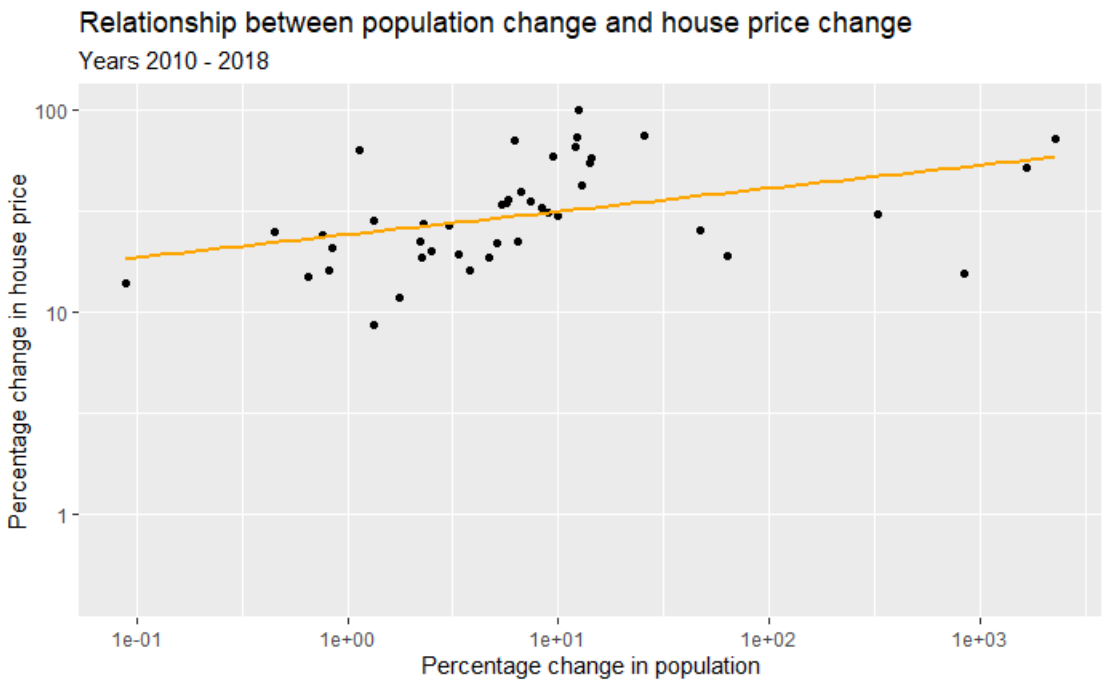

Fig 10

The relationship of house price changes and population changes between the years 2010 and 2018 is linear and positive. It is clearly seen that the increase in population caused increase in the house price and a decrease in population also decreased the house price in these years.

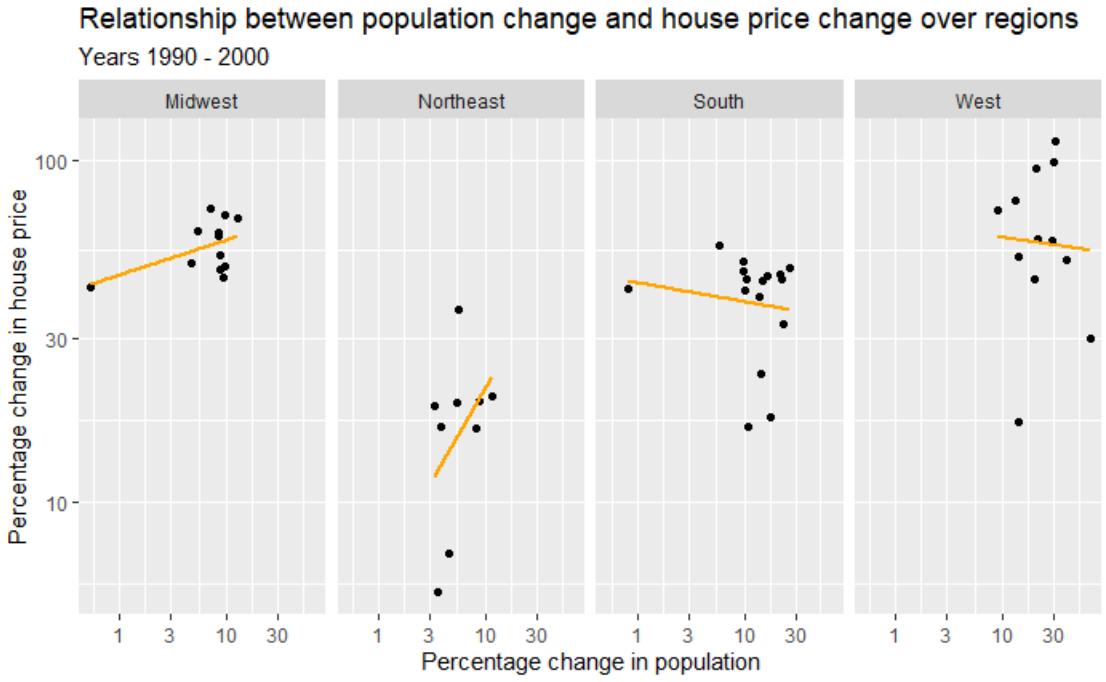

Fig 11

The above graph shows that the northeast region has seen the steep increase in the house price with respect to the population in the region change in the years 1990 to 2000. Mid-west states have positive trend in their house prices with respect to the population. South and west regions have a decrease trend in the house price. It could be because of population change in the south and west states.

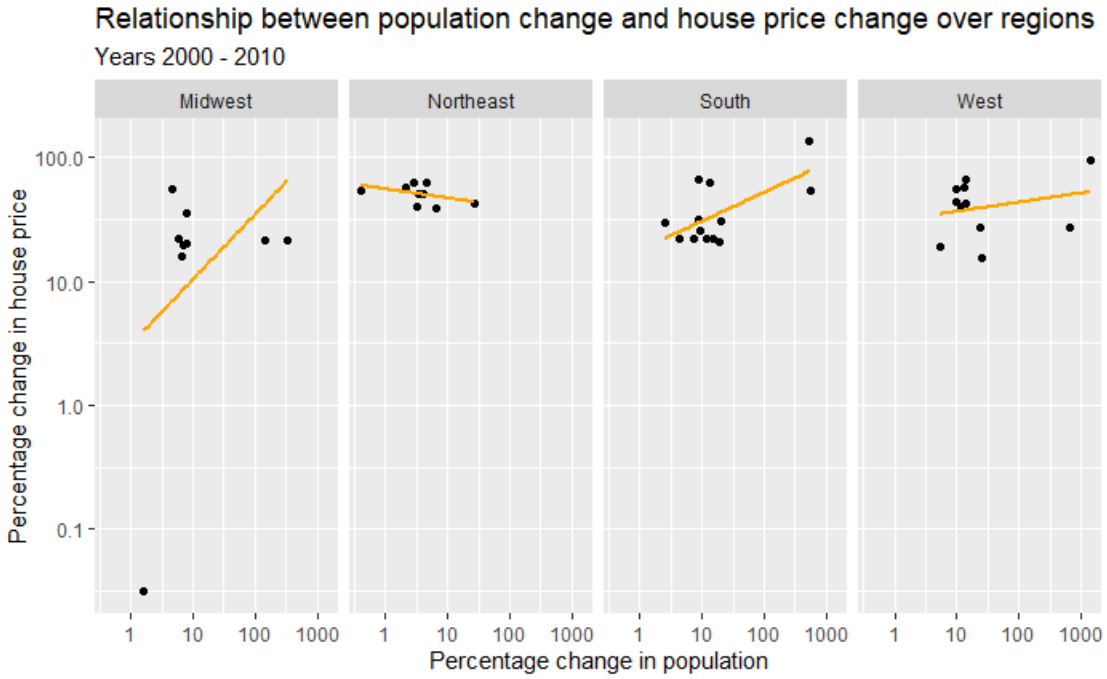

Fig 12

In the years 2000 and 2010, all regions have shown increase trend in the house price with respect to the population except for Northeast as there is a slight decrease trend in the region. Unlike the previous decade, this decade shows a steep increase in house prices over mid-west region.

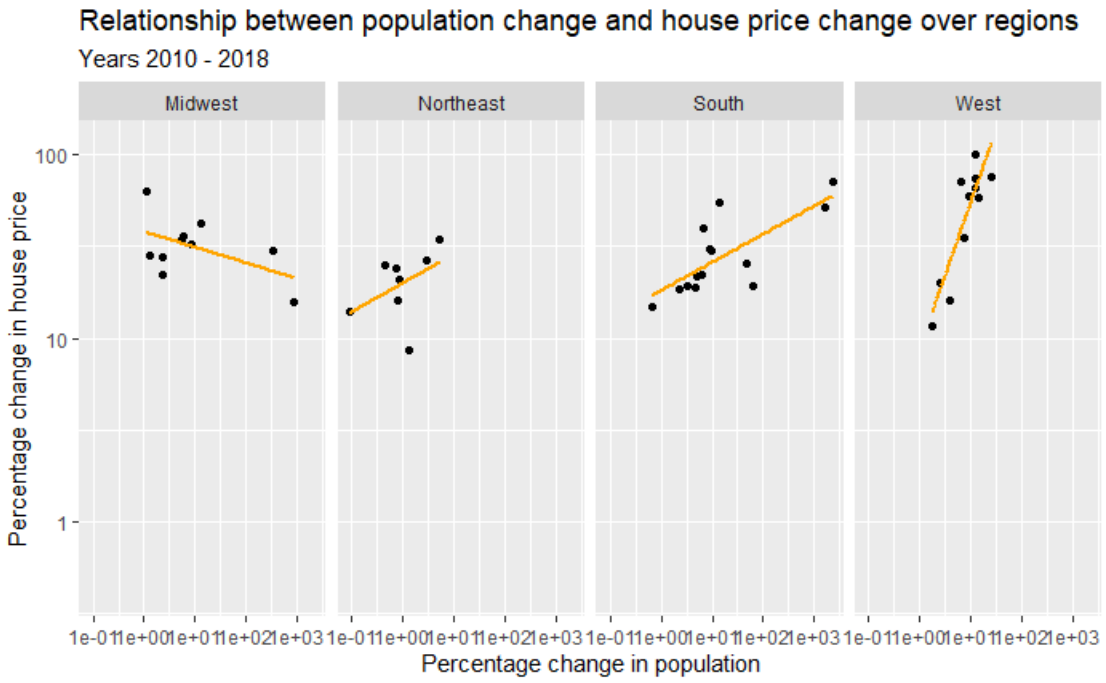

Fig 13

The decade after 2010 shows steep patterns in the relationship between house price and population density across the four regions. Except for mid-west region, other regions have seen huge increasing trend in the house prices with respect to the population.

CONCLUSION

There is no constant trend in house price changes across the country. Over the years, there has been ups and downs in the house prices. Although we could state that population density is one of the major factors influencing the house prices, we cannot conclude that it is the only factor which could affect the house prices. It is because, when we see region wise trend, mid-west region has positive trend in the relationship until before 2010 (fig 11 and 12). But the mid-west region shows a decreasing trend in the recent years (fig 13). We cannot attribute this decrease in the house price to only population change as there is less chance of population decreasing very quickly in less than a decade in a region. There could be some other factors which affect the house prices which we cannot conclude from the given data.