**Introduction:**

We have the House pricing index data collected in the United States from 1975 to 2016. It consists of 3-digit zip-codes 5-digit zip-codes, year, annual change percent, HPI when it was first recorded, HPI in 1990 and HPI in 2000.

The ZHVI summary file indicates the Zillow home value index. Zillow provides with the estimate of current market value for a home. The summary file consists of RegionID, RegionName, State and location information. It also provides zhvi, month on month, qoq, peak zhvi and time related information.

The data had some missing values which have been cleaned and then plotted.

We have worked on the analysis using Tableau and R Studio both so that we could develop a strong understanding of the data using the advantages of both platforms.

**Analysis:**

**1)File: HPI for Three-Digit ZIP Codes**

1) Annual Change vs Year

A close up of a map

Description generated with high confidence

We have plotted the sum of annual change percentage with respect to each year and shown the forecast till 2023. We can see zig-zag trends in annual change of HPI with high values in 1979 and 2005. The lowest value was in 2009 which started declining from 2007 which could be due to the great recession in United States. The blue dotted line represents the trend line and orange dotted line represents the forecast.

2) HPI, HPI in 1990, HPI in 2000 vs Year

A close up of a map

Description generated with high confidence

We have plotted sum of HPI, HPI in 1990, HPI in 2000 with respect to year and also shown the forecast till 2023. We can see the there is a gradual increase in the rates till 2007 after which there is a decline which can be seen by the orange line and then we can see the forecast till 2023. We can see from the data there is a gradual increase in the sum of HPI index rates.

3) HPI, HPI in 1990, HPI in 2000, Annual change vs Year

A close up of a map

Description generated with very high confidence

We have used to ggplot2 library to plot the above graphs. We can see the gradual increase in HPI every year with a drop in 2008 as seen previously. We have used the geom\_jitter and geom\_smooth methods from the ffplot2 library.

4) Longitude and Latitude vs HPI

A close up of a map

Description generated with high confidence

We can obtain the longitude and latitude information using the zip codes in the dataset. But after plotting it we realize that there could have been a data loss by using a 3-digit zip code as most of the data is plotted in Puerto Rico and there are 705 unknown values present.

**2)File: HPI for Five-Digit ZIP Codes**

1) HPI, Annual Change vs Year

A close up of a map

Description generated with high confidence

We have plotted the sum of annual change percentage and HPI with respect to each year and also shown the forecast till 2023. We can see a zig zag line with major decreases in 1981 and 2007. There is a huge drop in the annual change percentage from 2006 to 2009. The orange dotted line represents the forecast till 2023 and the blue dotted line represents trend line.

2) Annual change vs year boxplot

A picture containing sky

Description generated with very high confidence

We have plotted sum of annual change every year using boxplot. Boxplot is a great way to see outliers in the data and hence we can see the years in which there were drastic change in the data. We can see years 2008, 2009, 2010 and 2011 having a negative annual change rate in which the there was a drastic drop in annual change percentage. Years 2004 and 2005 had the greatest positive annual change percentages.

3) Longitude and Latitude vs HPI

A close up of a map

Description generated with high confidence

We have obtained the longitude and latitude information using the zip codes in the dataset. Unlike with the 3 digit zip code dataset we have obtained the results properly. We used the ggmap and zipcode library to generate the above results. We make use of the zipcode library to compare it with zip codes given in the dataset to receive the latitude and longitude information. We have then plotted it using ggmap library with colors corresponding to the House pricing index. We can see that high HPI values in California and New York states.

**3)File: ZHVI Summary all houses**

1) State vs Zhvi

**A picture containing sky, writing implement

Description generated with very high confidence**

We have plotted States with Zhvi values the color is based on the sum of size rank. We can easy make out the differences in zhvi values within different states using bar chart. We can see that California state has the highest Sum of Zhvi value followed by New York and Florida.

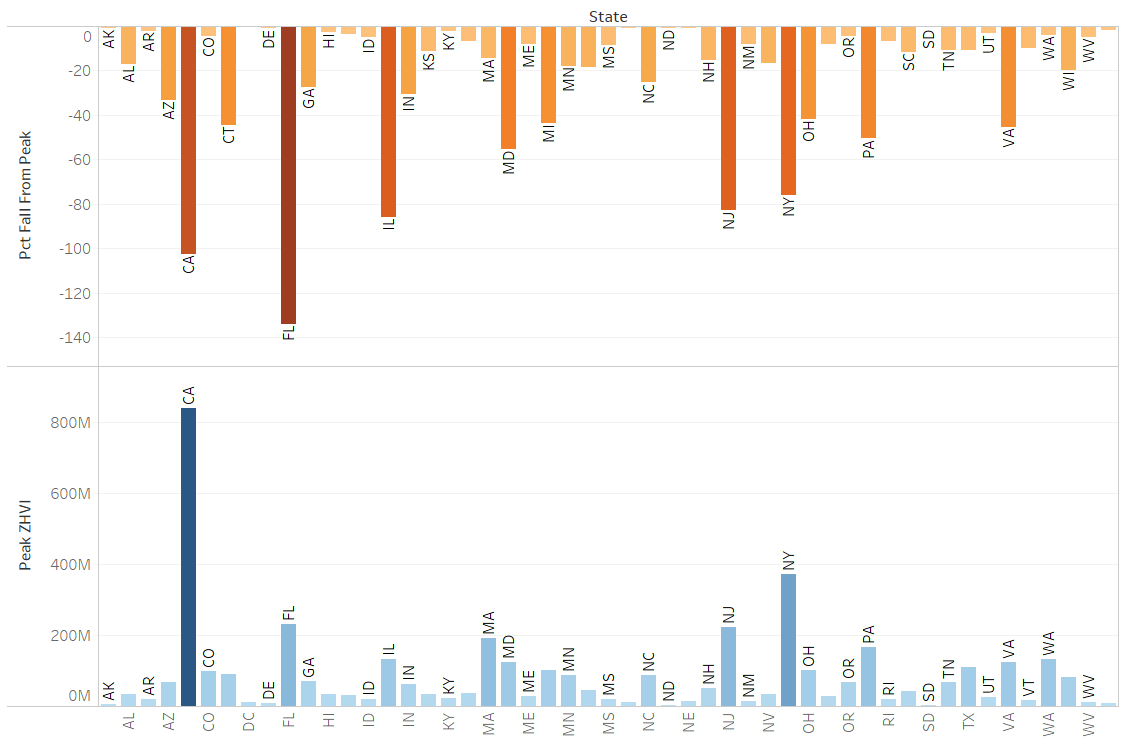
2) State vs MoM, QoQ, Yoy

A picture containing screenshot

Description generated with very high confidence

We have plotted States with Month on month, Quarter on quarter and year on year. We can see similar trends within different states using the bar chart. We can see that Virginia state has negative MoM and QoQ Values. We can also see that California state has the highest Mom, Qoq and Yoy values followed by New York and Florida.

3) State vs Peak Zhvi, Fall from peak



We have plotted States with Peak Zhvi and Fall from Peak. We have used barplot and plotted both attributes in such a way, so it is easy to see the contrast between the values. We can see that California and Florida has the highest fall from peak while California and New York have the highest Peak ZHVI.

4) Longitude and Latitude vs Peak Zhvi

A close up of a map

Description generated with high confidence

We have plotted the longitude, latitude with peak Zhvi. The longitude and latitude information is obtained using states data. We can see from the colors on the map that California, Florida and New York states have the highest sum of peak Zillow home value index.