***Business Intelligence Team id:***

***LTVIP2025TMID50687***

*visualizing housing market trends: an analysis of sale prices and features using tableau*

## Submitted by:

#### Rajesh Abothula

**CONTENTS**

[PART A. DATA SETS 2](#_bookmark0)

[PART B. DATA CLEANING 3](#_bookmark1)

[PART C. DATA VISUALIZATIONS 8](#_bookmark2)

[PART D. DASHBOARD 16](#_bookmark3)

[PART E. STORYTELLING 16](#_bookmark4)

# PART A. DATA SETS

Even though there are a multitude of sources that track and compile the real estate market information and statistics, for this project we have chosen to use the data sets provided by Zillow Real Estate Research given the vast amount and breadth of high quality data provided by them on the subject. Zillow Real Estate research is a division of the analytics group Zillow Inc. dedicated to analysis and research into various topics pertaining to the housing market. In particular, we will be utilizing Zillow Home Value Index (ZHVI) which tracks the median value of all homes in an area, regardless of whether the home was sold during the reporting period. We will use this information to gather insights into the performance of the U.S. real estate market across various geographic areas and observe trends over the past 20 years. Given that most of the data sets provide month over month time series data over the 20 year period, they consist of thousands of rows of data and 5-10 columns.

We are using the following 3 data sets available at the following link: <https://www.zillow.com/research/data/#median-home-value>

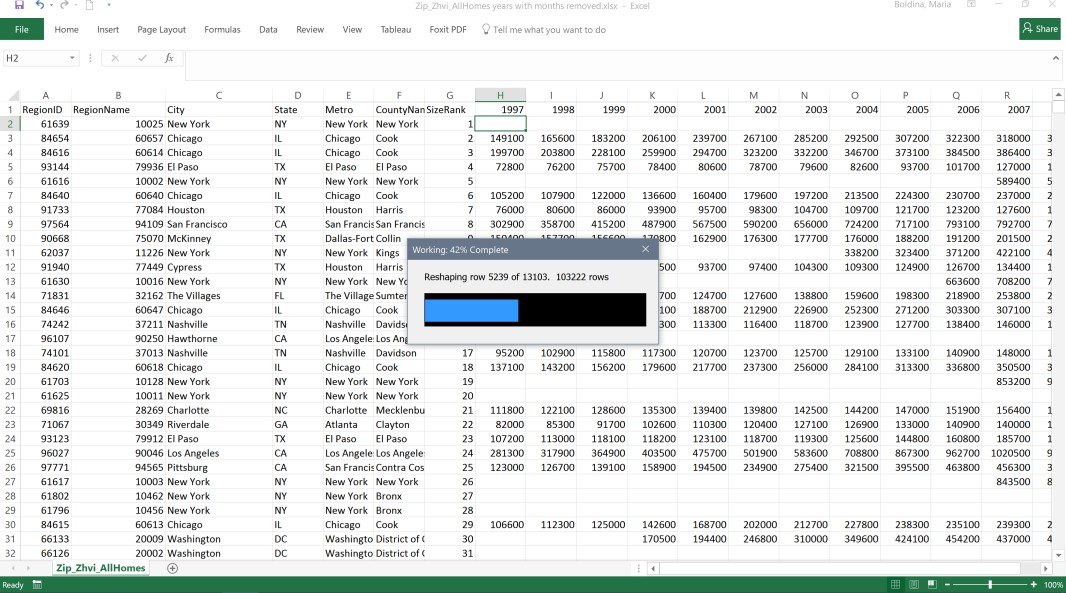
1. ZHVI Summary (dated January 2017) aggregated at the zip code level
2. ZHVI All Homes (SFR, Condo/Co-op) Time Series aggregated at the state level. This data set provides time series data for the period 1996- Jan 2017.
3. ZHVI aggregated at the United States and city level This data set is a time series data for the period 1996- Jan 2017.

# PART B. DATA CLEANING

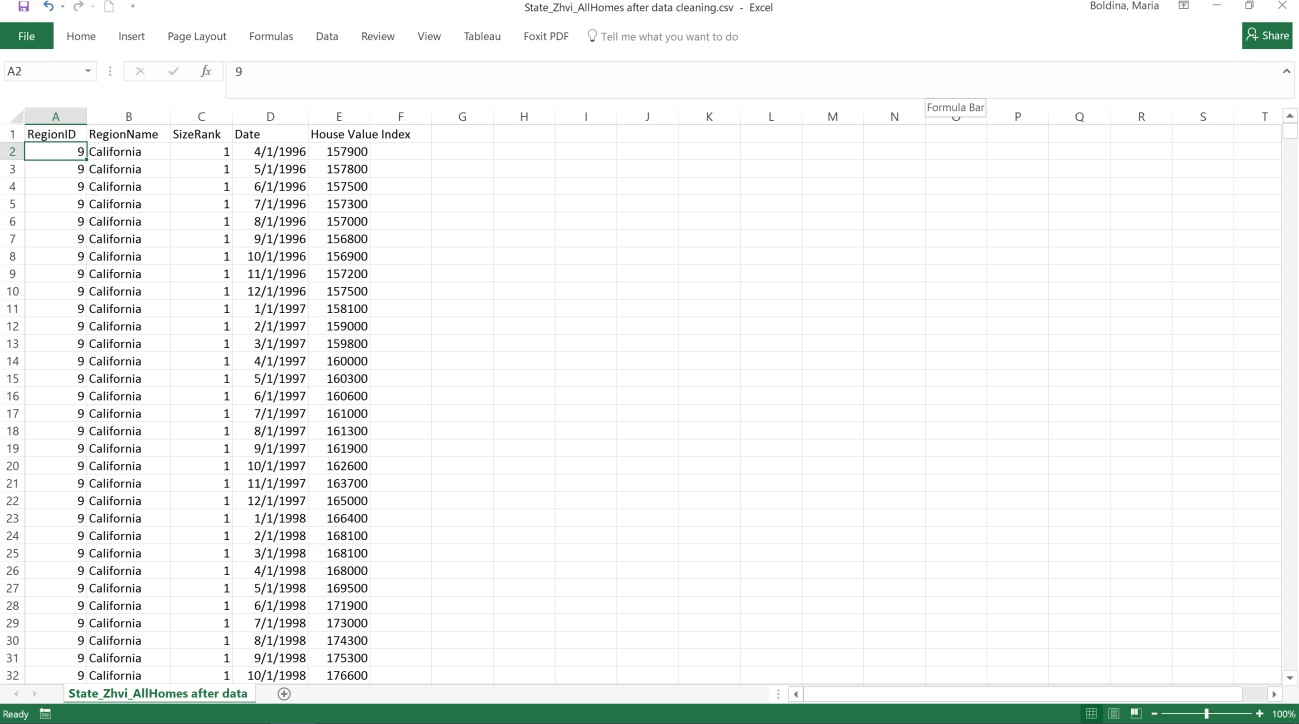
Overall, the data provided by Zillow was of high quality. There we no problems with data validity, accuracy, consistency and uniformity. For example, there were no duplicates, illegal values etc. as can be seen from screenshots below. However, there were some issues with data completeness and I had to make some data reshaping and date standardization given the time series presented by Zillow:

### Data reshaping Before:

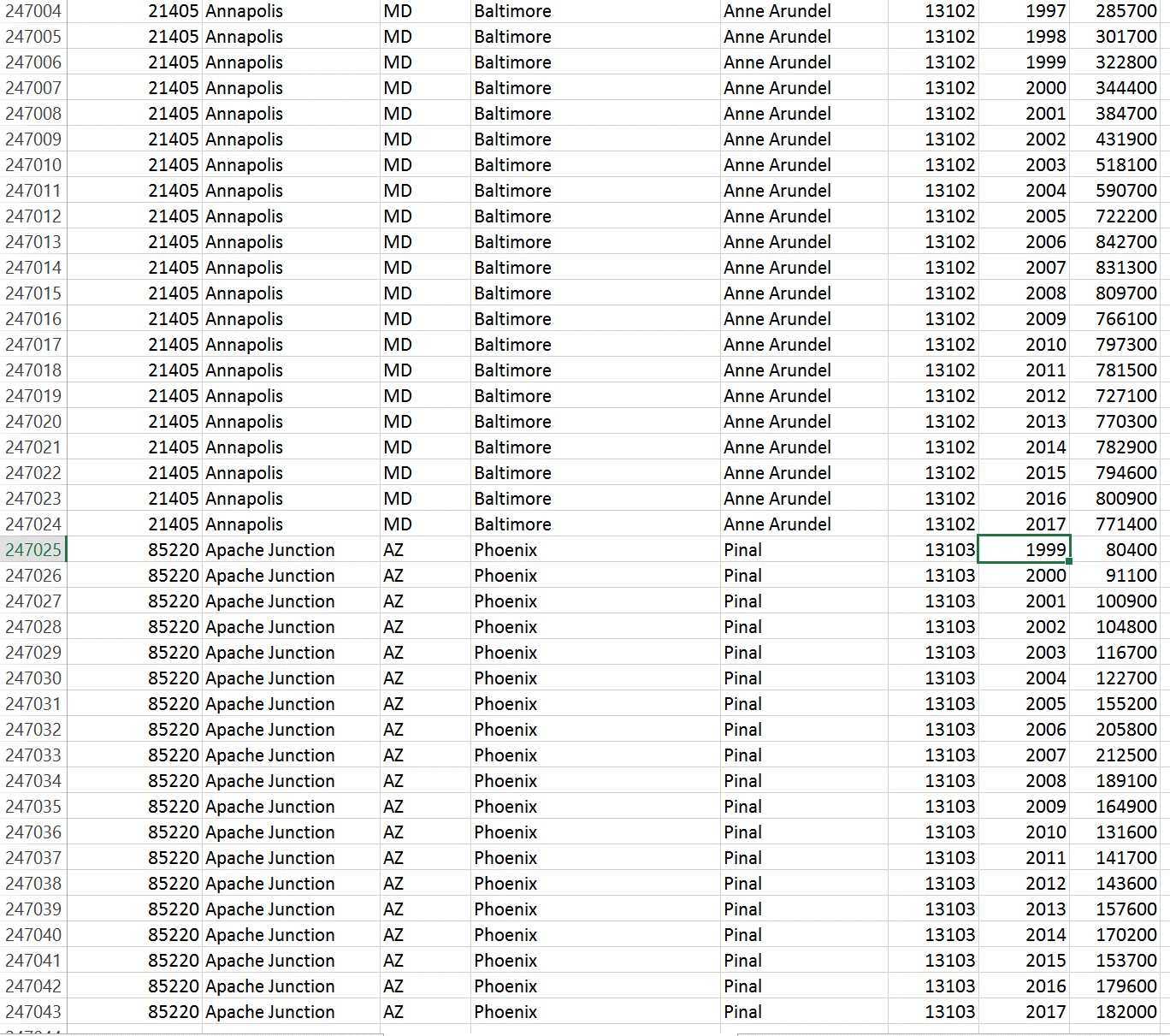
### Using Tableau Data Reshaper add-in for *Excel*



**After:** Data is tall and thin… the preferred type for Tableau.



### Missing Values:

****

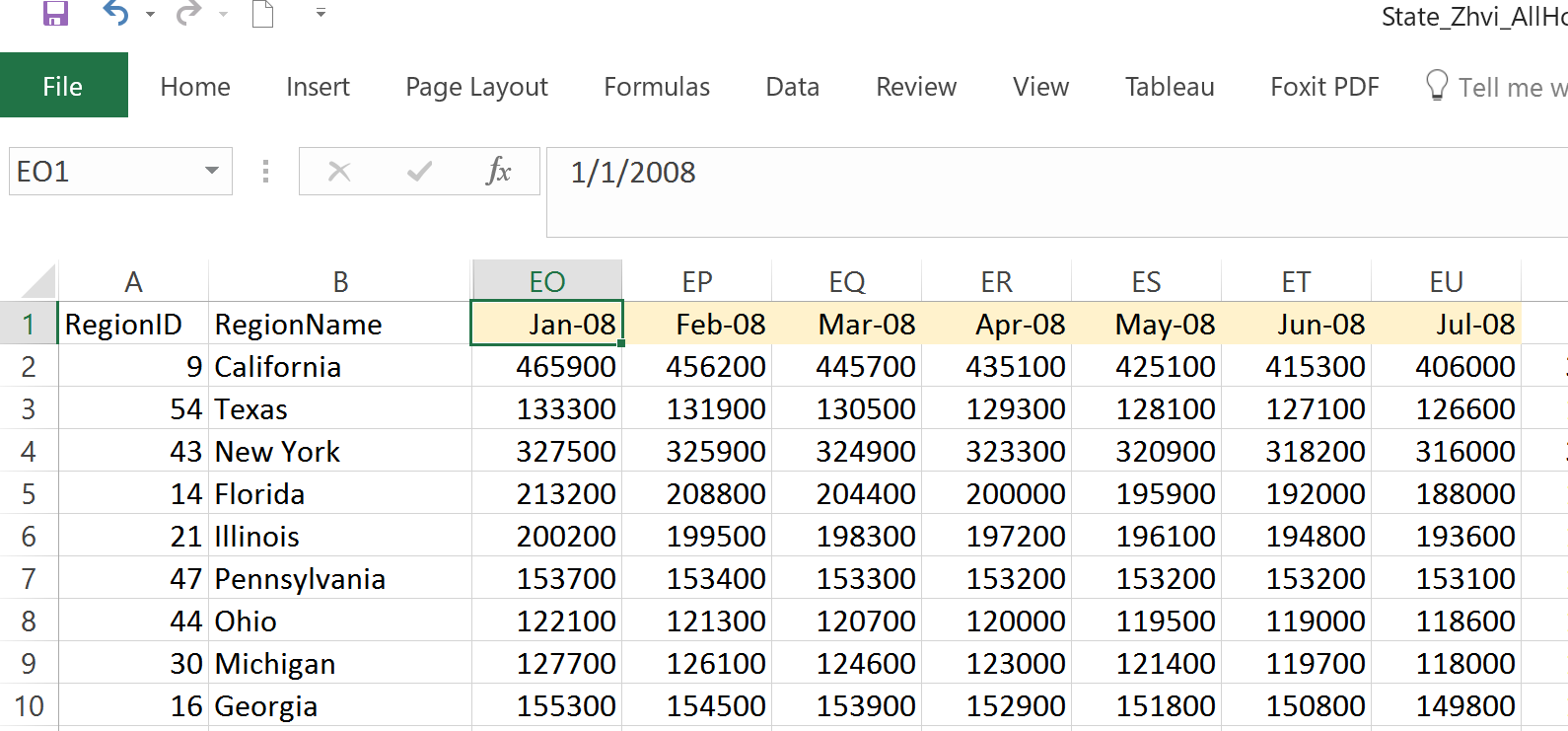
Please note that some data is null for certain zip codes in the years before the end of 2008 because such data was not being collected at the time. Instead of truncating the data to range only from 2008-2017 I decided to keep all the data because it could still be interesting to look at the time series for the zip codes that the information is available from 1997 - 2017. The null values do not affect the results for the Years 2008- 2017 year.

For example, in the screenshot above the ZHVI values for zip code 85220 Apache Junction are only available since 1999, but we can see that for zip 21405 Annapolis there are no missing values since 1997-2017.

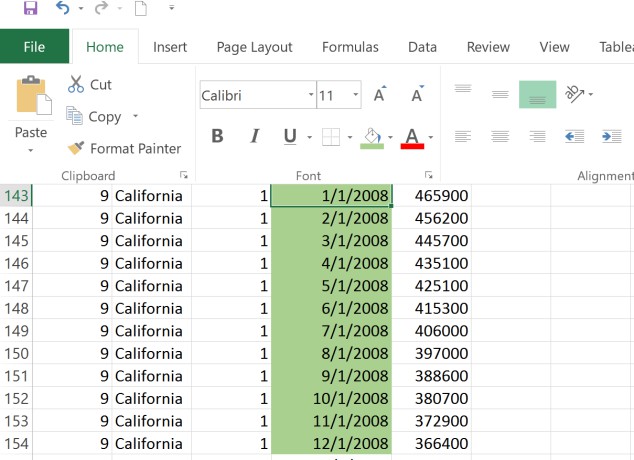
### Date Standardization

Date was standardized to be as MM/DD/YYYY or YYYY where monthly data is not provided, Original data was in Month-YYYY date format which is not ideal for Tableau and is not recognized as a date.

*Before:*

**

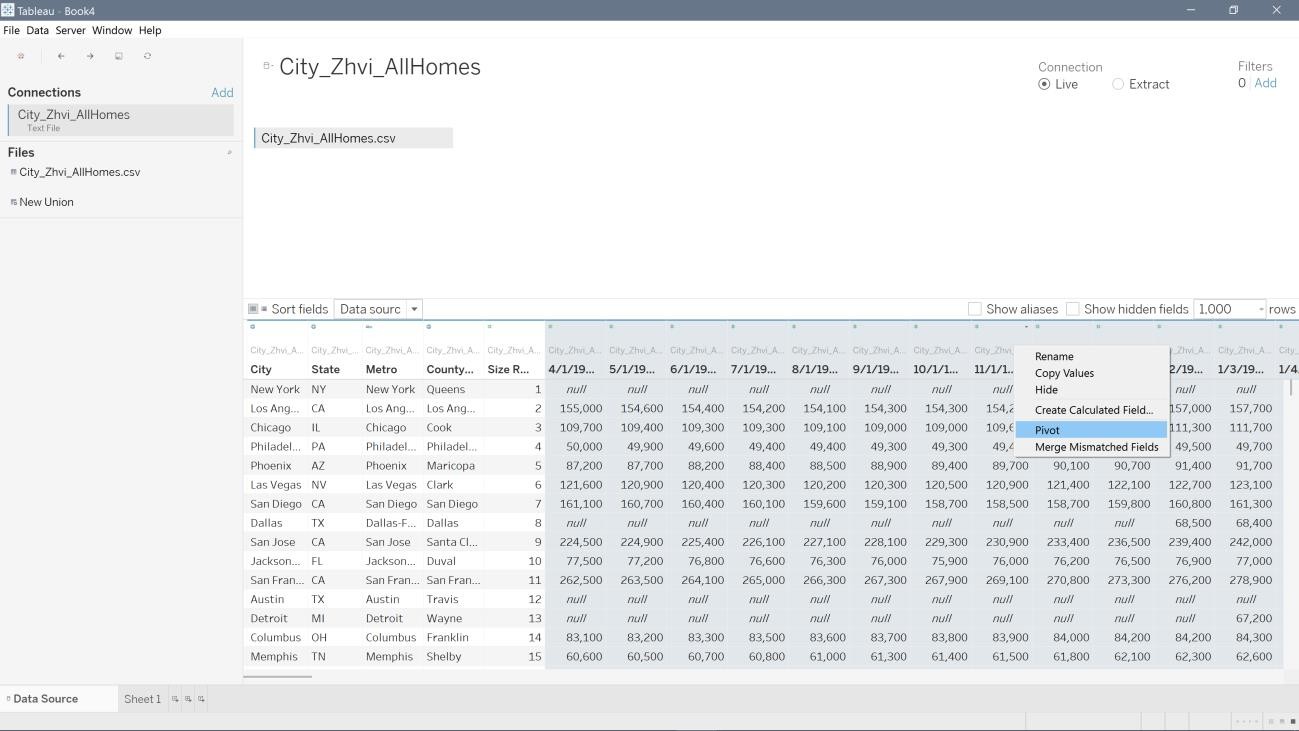
*After:*



### Removing extra columns and renaming columns

Data was also cleaned and reshaped in Tableau itself. Dates in time series were pivoted and unnecessary columns were hidden, pivoted columns were renamed and their geographic roles adjusted.

*Before:*

**

# PART C. DATA VISUALIZATIONS

### What is the current median house value in United States and how has it changed over the years?

****

As depicted in the line chart above the median house value in the United States s of January 2017 is $195,300. This that the house values have almost doubled since 1996. At its peak level the median house prices have reached $196,500 in May 2007. However, as evident from the chart, the US house value has lost almost 23% of its value due to the 2008 housing and economic crisis. The housing market bottomed out at $151,500 in December 2011. Now almost 10 years later, the house values are finally close to being at the same levels as their 2007 level peak (only 0.61% below).

*Categories used:*

* + *Reference line*
  + *Dates*

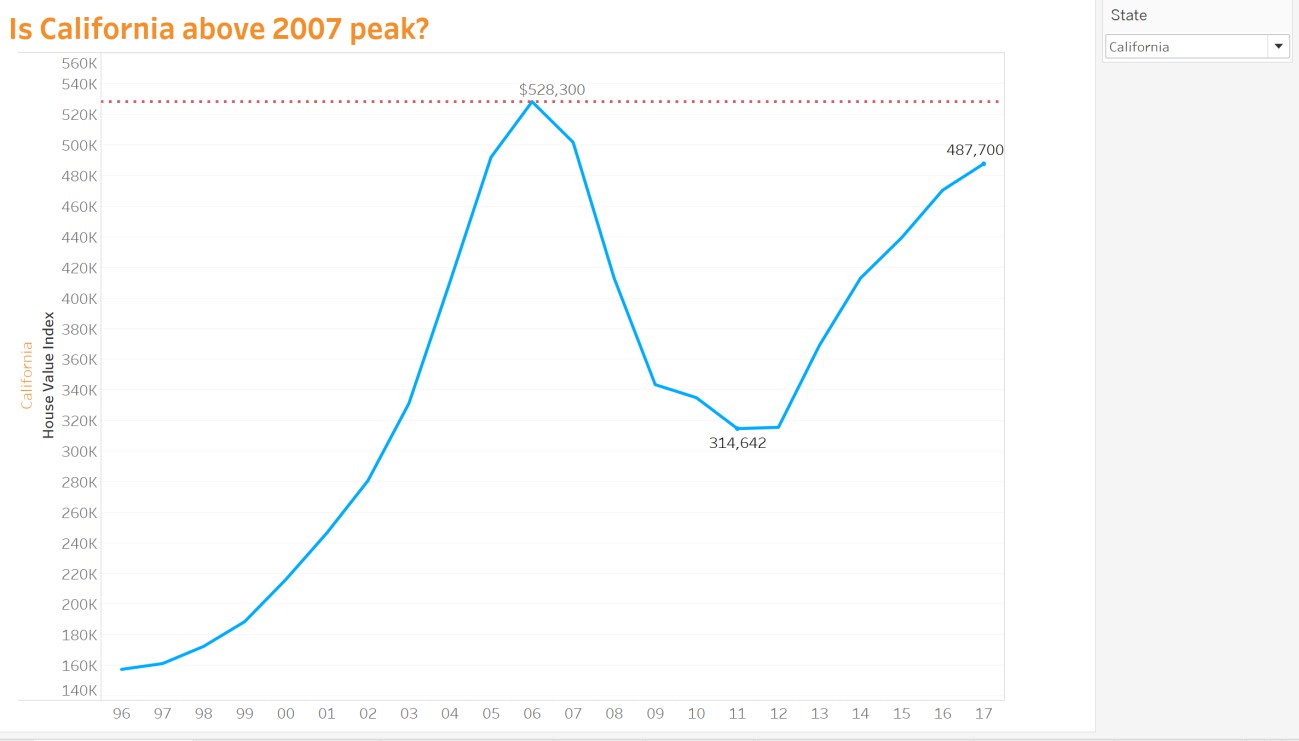
### Which States have the highest median home value in 2017?

As we can easily see on the above map the majority of states with the highest median house values in 2017 are located in the west of the United States. (Please note that although Hawaii and Alaska, are not immediately visible on this map, they could be easily seen by zooming out). The lowest house value is $102,400 (West Virginia) and the highest is $589,300 (Hawaii). We could also easily see the house values in any other year from 1996 to 2017 by selecting the appropriate value in the quick filter on the right.

*Categories used:*

* + *Geographic mapping*
  + *Dates*

### Have house prices in California surpassed their pre-2007 peak?

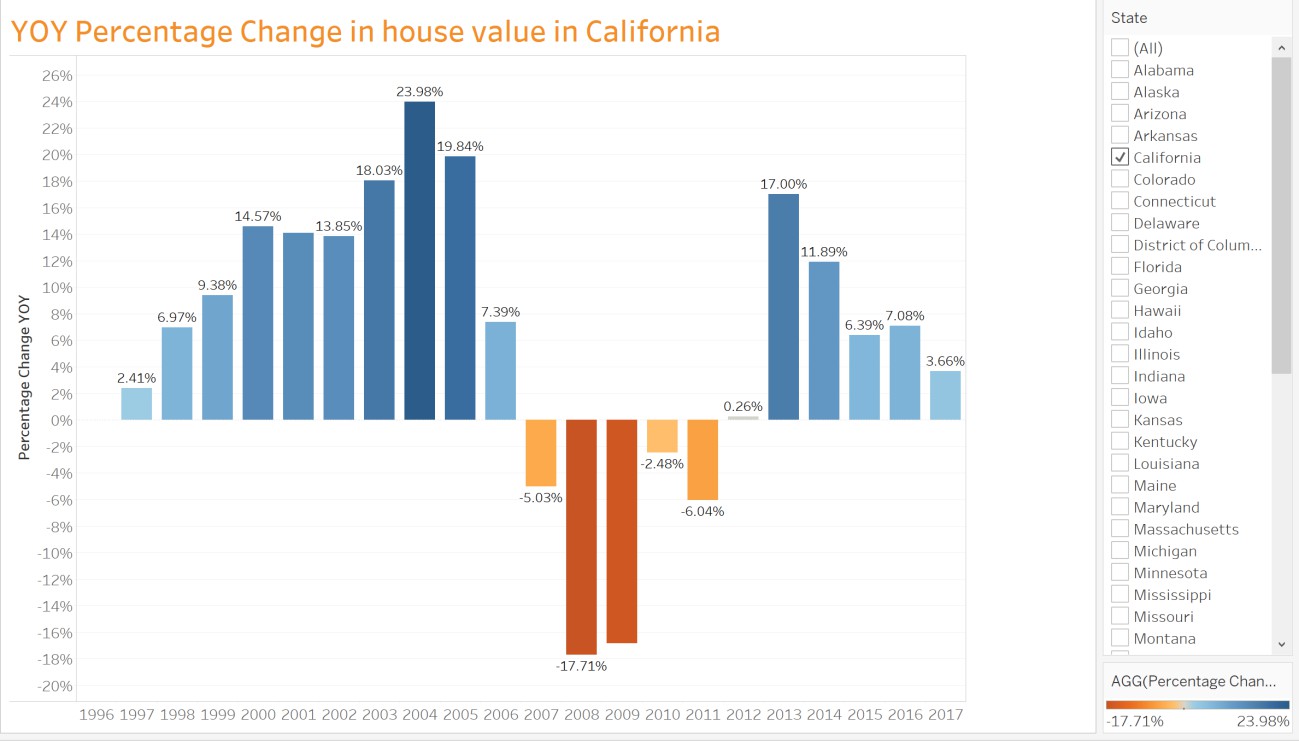
****

As of January 2017 median house value in California is $487,700. Therefore, interestingly enough California house prices still have not reached their 2006 peak of $528,300. Although they have regained most of the value lost since the 2011 bottom of $314,642. We can also do the same analysis for any of the other 49 states by using the quick filter on the right and selecting the appropriate state.

*Categories used*

* + *Reference line*
  + *Dates*

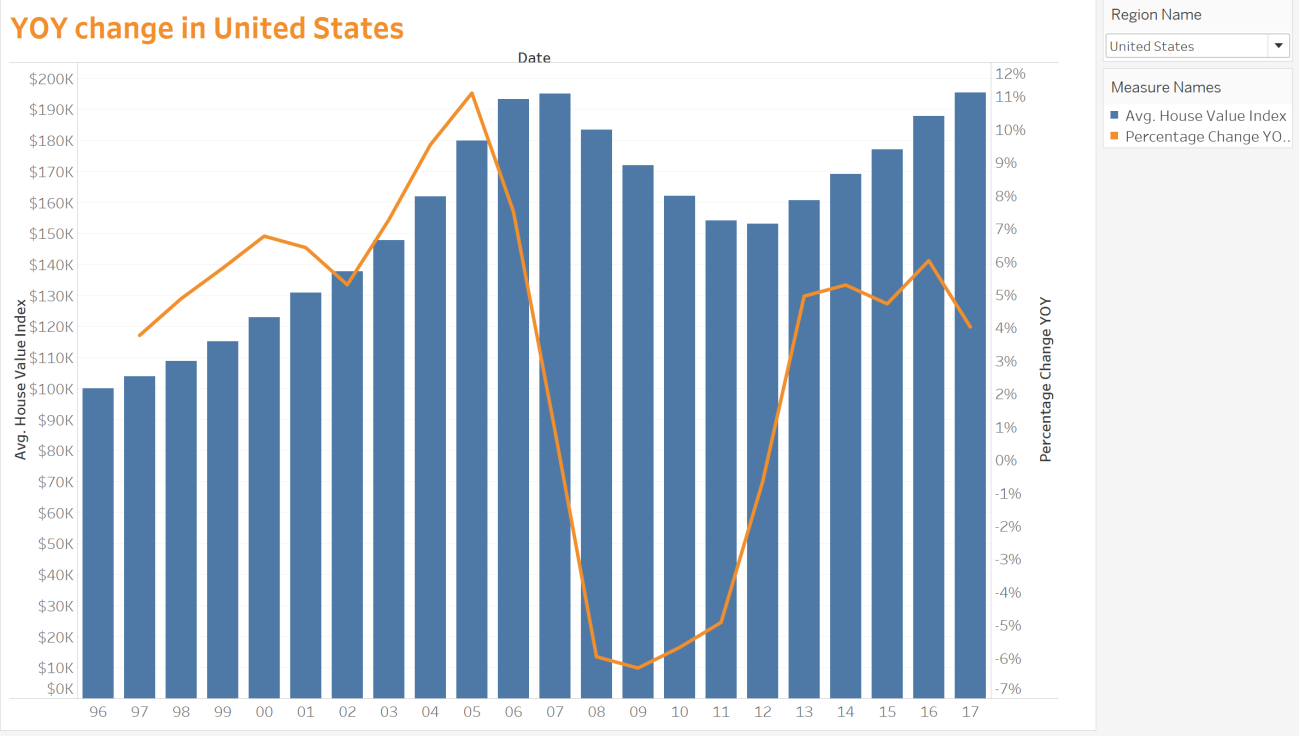
### What is the annual percentage change in home value appreciation in each state?

****

The bar chart above makes it easy to see the year over year percentage change in house values for any state in the United States. In the example above we can see that California house prices were hit pretty hard during the recession; house prices were falling each year since 2006 till the end of 2011. We can also observe that they decreased the most in 2008 (by 17.71%) and increased the most in 2004 (by 23.98%). In addition, we can see that that the house value has been increasing since the 2013 and in 2016 the house value increased by 3.66%.

*Categories used:*

* + *Calculated field (YOY Percentage Change)*
  + *Dates*

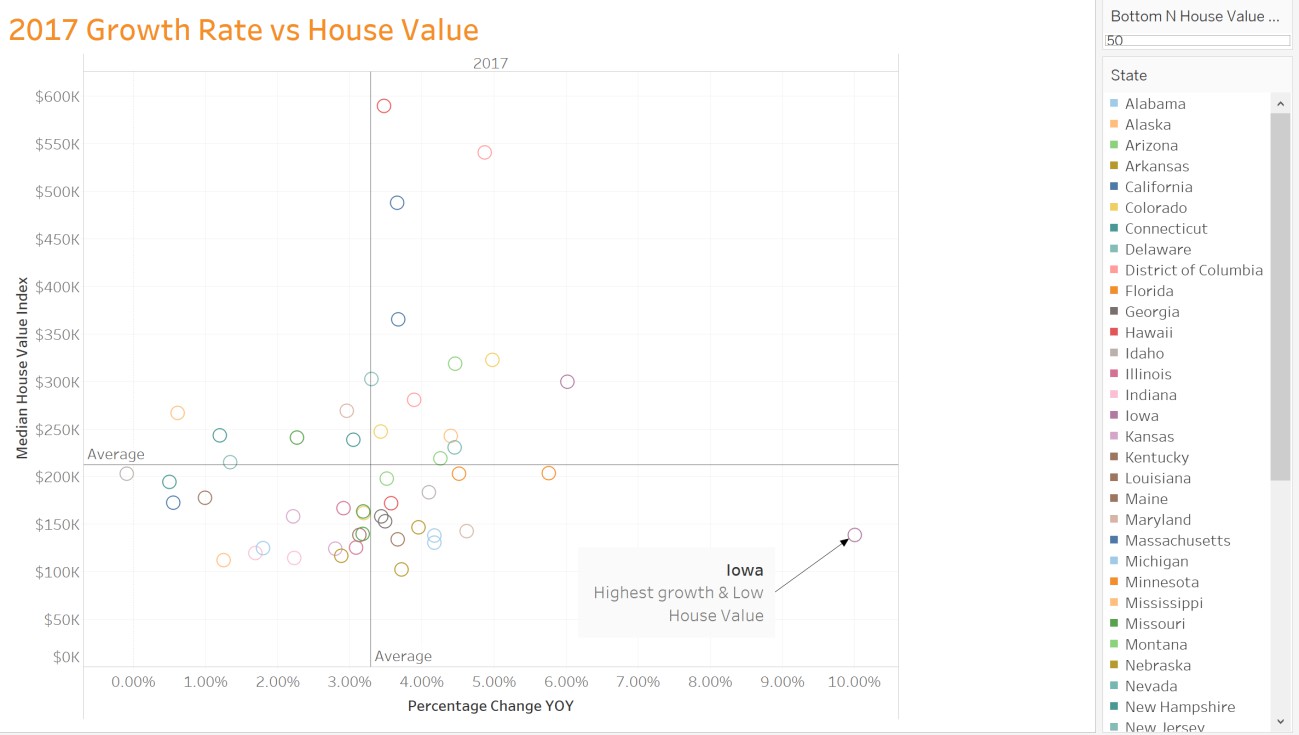


This chart is similar to the one above but due to the dual axis we can see both the average house value in any given year as depicted by the blue bar charts and the corresponding year over year percent change which is shown by orange line chart. This view can be shown for the United States in general or for any of the 50 States as desired.

*Categories used:*

* + *Calculated field (YOY Percentage Change)*
  + *Dates*
  + *Dual Axis*

### Which areas exhibit the fastest growth and have the lowest house prices?

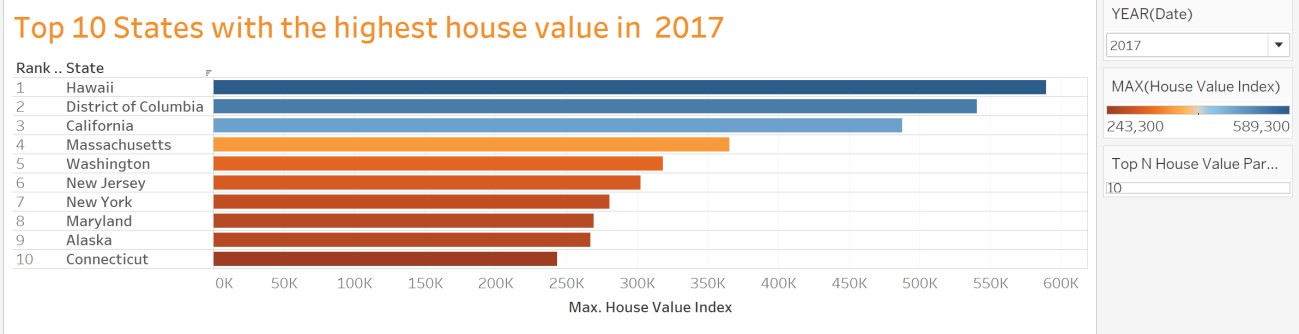
****

The scatter plot above makes it easy to see which states have the highest percentage of growth in 2017 as compared to 2016. For example, we can tell that the states represented by the circles in the bottom right quadrant have higher than average percentage of growth and lower than average house values. In particular, the state of Iowa stands out with the highest percentage of growth of 10.01% and relatively low house prices. So, it seems like it might be a good idea to buy a property in Iowa if the prices continue to increase at a similar rate. We could also use the Bottom N parameter to narrow down the properties with the lowest values, for example, bottom 10 states.

*Categories used:*

* + *Calculated field (YOY Percentage Change)*
  + *Scatter Plot*
  + *Parameter (Bottom N House Value)*

### What are the top 10 States with the highest house value index?

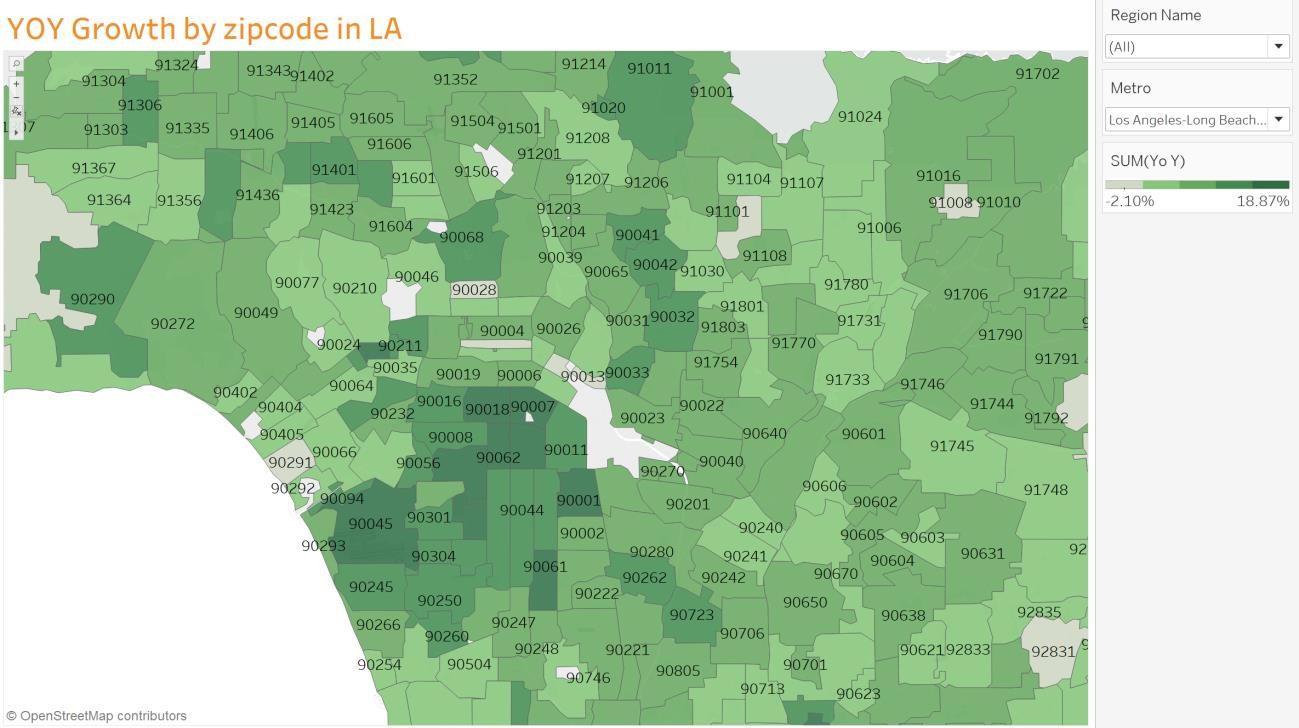
****

The bar chart above ranks the top 10 states with the highest house value index. As we can see Hawaii has the highest house value index of $589,300, followed by District of Columbia in the second place and California in third.

*Categories used:*

* + *Rank*
  + *Parameter (Top N House Value)*
  + *Calculated field (MAX)*

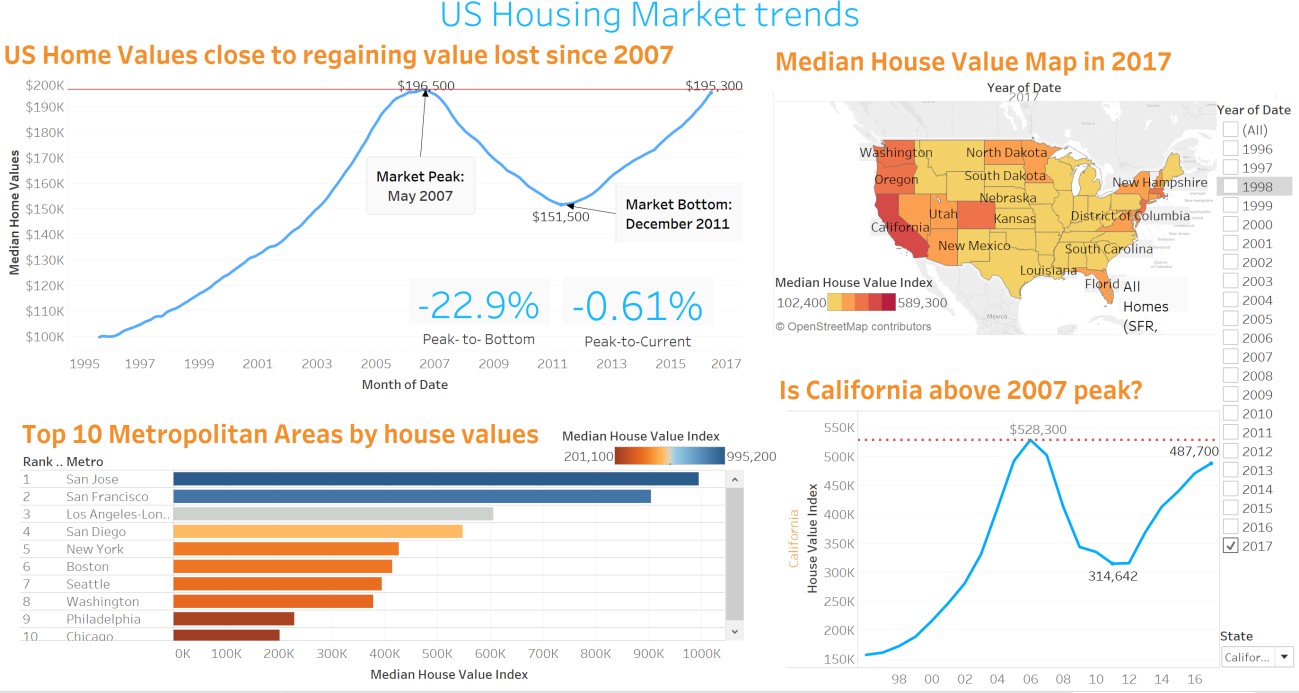
### In which zip codes in Greater Los Angeles area have the home prices appreciated the most since last year?

****

Looking at the map of the Greater Los Angeles area we can see that some neighborhoods have gained as much as 18.87% as compared to last year. Most of the zip codes in the area exhibit steady growth with very few exceptions as can be seen by mostly darker green areas on the map. *Categories used:*

* + *Geographic mapping*

# PART D. DASHBOARD

****

# PART E. STORYTELLING

The U.S. Housing Market has been a hot topic of research and discussions among scientists, economists and the general public, especially since the infamous housing bubble collapse in 2008 caused by the subprime mortgage industry collapse and subsequent economic recession.

The analysis provides an overview of the current situation of national and local real estate and rental markets as well as analysis of the U.S. Housing Market Trends from 1996 to 2017 using the rich capabilities of Tableau software to visualize the data and geo-mapping the data in particular, in order to be able to present the analysis in the most effective and easily comprehensible way. A particular attention in the analysis is be paid to the California and Greater Los Angeles real estate markets. In particular, we will be utilizing Zillow Home Value

Index (ZHVI) which tracks the median value of all homes in an area, regardless of whether the home was sold during the reporting period. For clarity and ease of understanding, we refer to ZHVI as simply home values or home value index in this report.

The most important question 10 years after the housing bubble burst remains whether the US housing market has recovered enough to surpass the levels reached at the peak of the market in May 2017.



Figure 1 Figure 1 shows that on a national level the median home value has reached $195,300, just 0.61% shy of their previous peak almost a decade ago. Home values have been growing at a steady pace since the market bottom of $151,500 was reached in December 2011. The housing market is red hot again. Interest rates of about 3.5% or less for 30-year, fixed-rate mortgages — not far off the all-time low of 3.31% in November 2012 — have helped fuel the gains2.

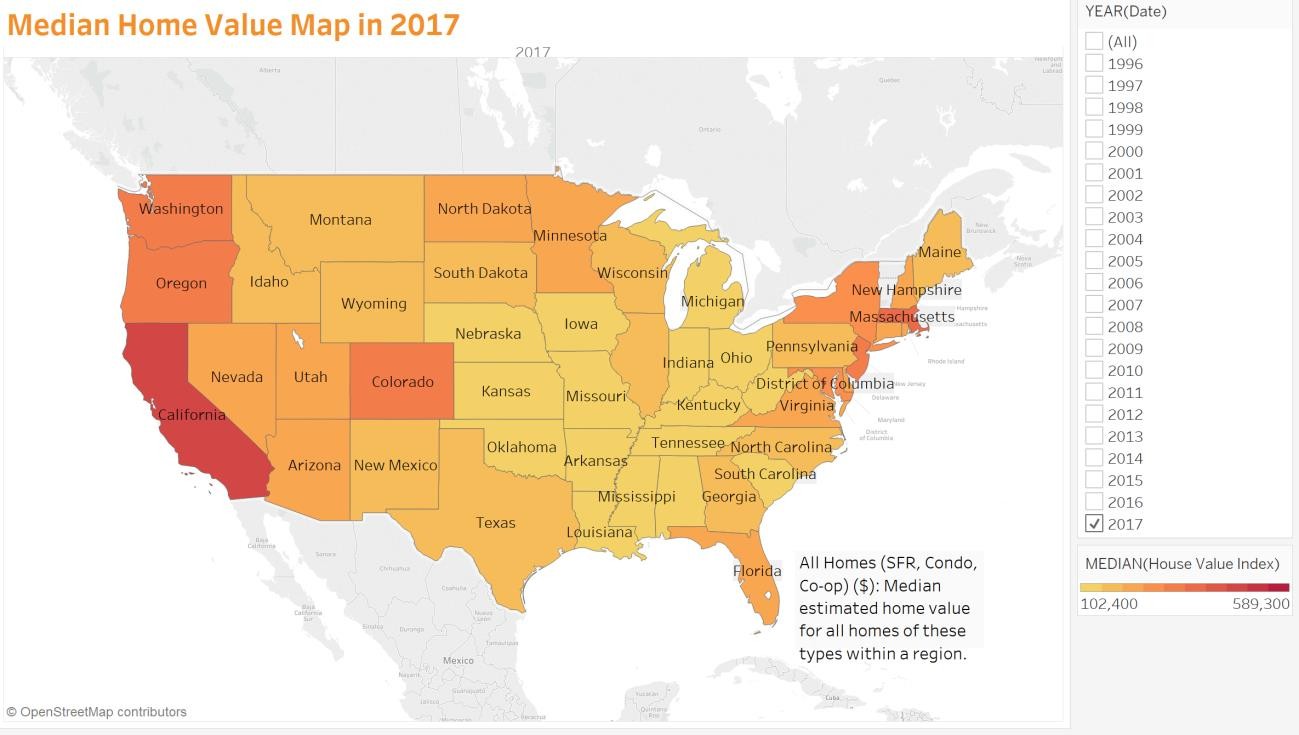


Figure 2



A look at the map in Figure 2 above shows that regionally, the highest home market values currently are in the West region of the United States and Hawaii. The lowest house value is

$102,400 in West Virginia and the highest is $589,300 in Hawaii.

By examining the figure 3 below we can see that California’s house prices were hit pretty hard during the recession. In fact, house prices were falling for 5 years straight from 2006 until the end of 2011. We can also observe that home value decreased the most in 2008 (by 17.71%) and increased the most in 2004 (by 23.98%) the rate of growth which hasn’t been surpassed since. However, home values have been appreciating steadily since 2013, although the growth slowed down somewhat to 3.66% year over year increase in January 2017.

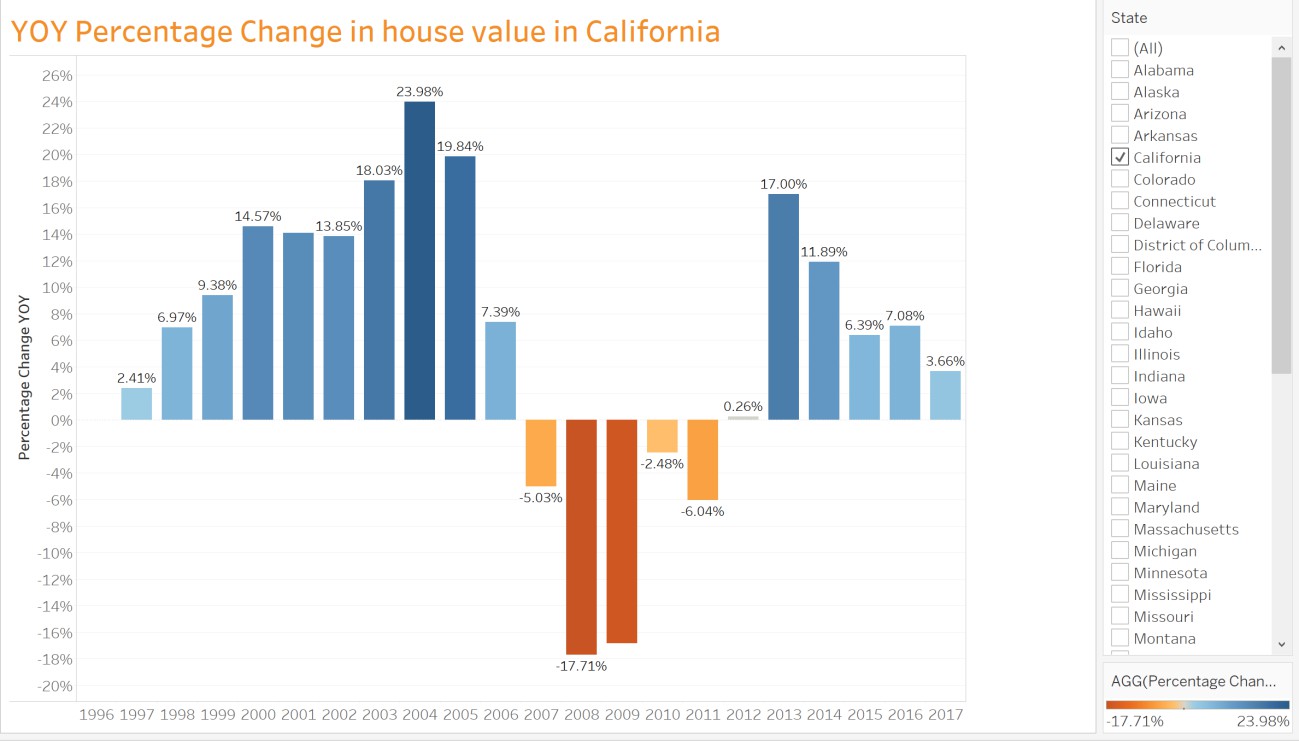


Figure 3 If we take a closer look at California in Figure 4 below we notice that as of January 2017 the median house value in California is $487,700. Therefore, interestingly enough, on the state level California house prices still have not reached their 2006 peak of $528,300. Although they have regained most of the value lost since the 2011 bottom of $314,642.

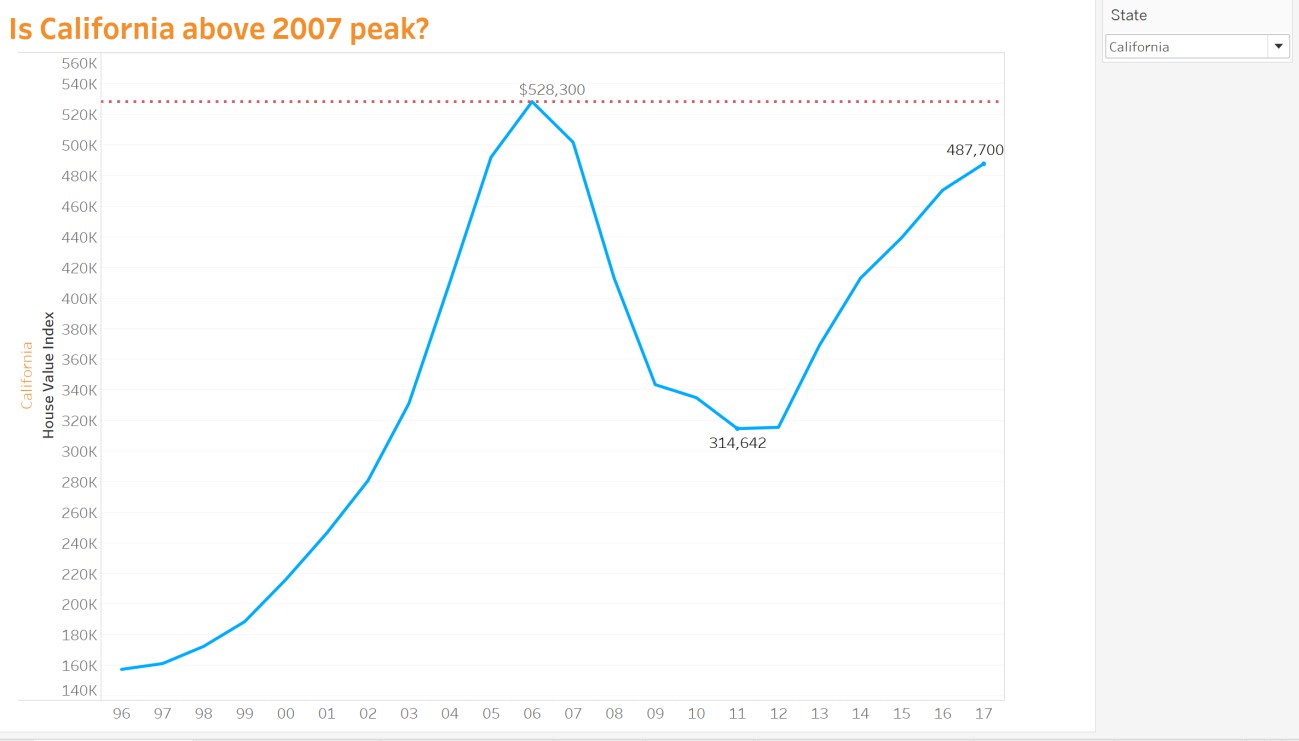


Figure 4 Looking at the map of the Greater Los Angeles area in Figure 5 below we can see that some neighborhoods have gained as much as 18.87% as compared to last year. Most of the zip codes in the area exhibit steady growth with very few exceptions as can be seen by mostly darker green areas on the map.

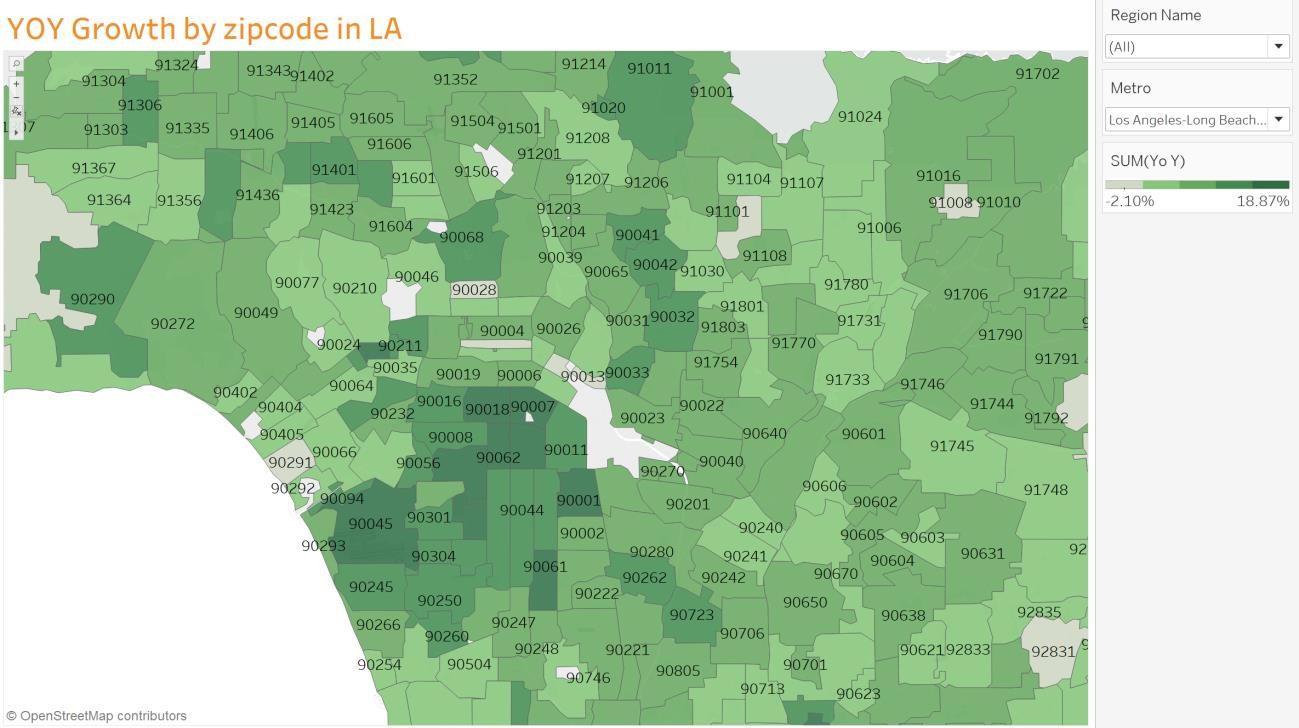


Figure 5

# REFERENCES

1. “Zillow Real Estate and Rental Data: Why We're Different.” *Zillow.com.* Zillow Real Estate Research. Web. 22 February 2017.
2. “Why home prices in Southern California keep climbing.” *LATimes.com*. Web. 14 July 2016.
3. “The US housing market recovery: The past is not prologue.” *Deloitte.UniversityPress.*

Web. 16 November 2016.