Business Intelligence

CIS 5270

U.S. Housing Market Trends Analysis

using Tableau

Submitted by:

Maria Boldina

CONTENTS

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

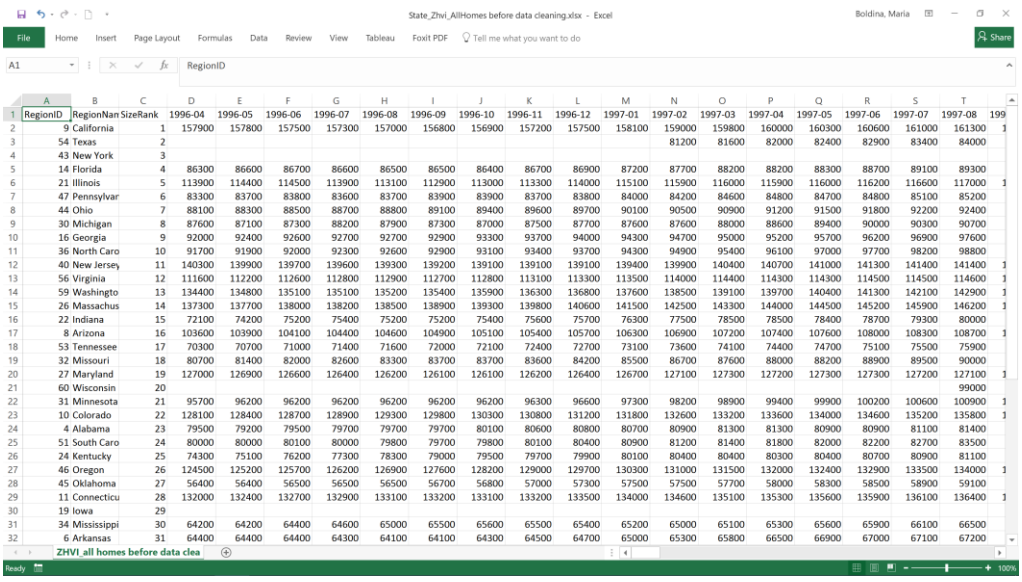
1) ZHVI Summary (dated January 2017) aggregated at the zip code level
2) ZHVI All Homes (SFR, Condo/Co-op) Time Series aggregated at the state level. This data set provides time series data for the period 1996- Jan 2017.
3) ZHVI aggregated at the United States and city level This data set is a time series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

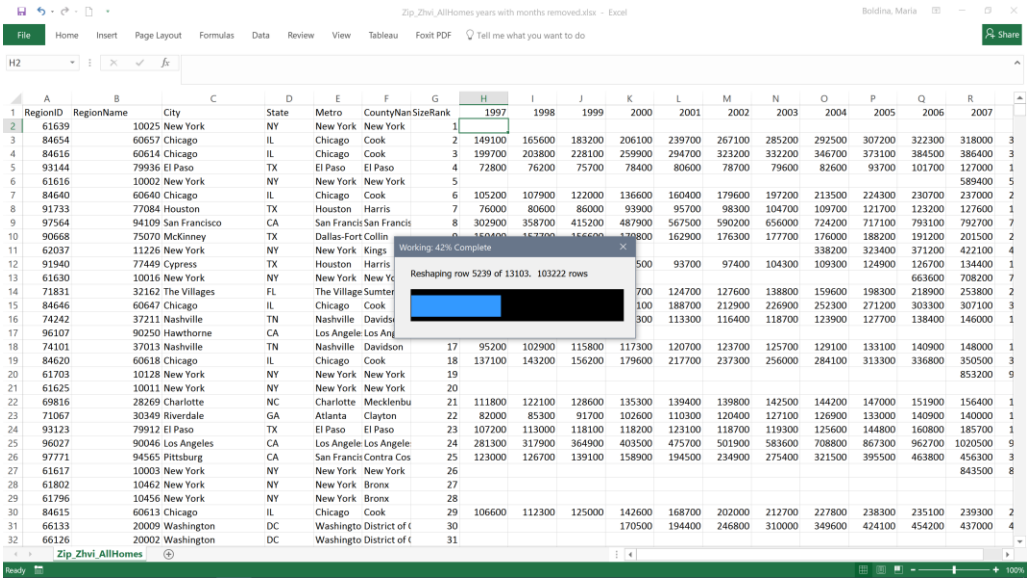
Data reshaping

Before:

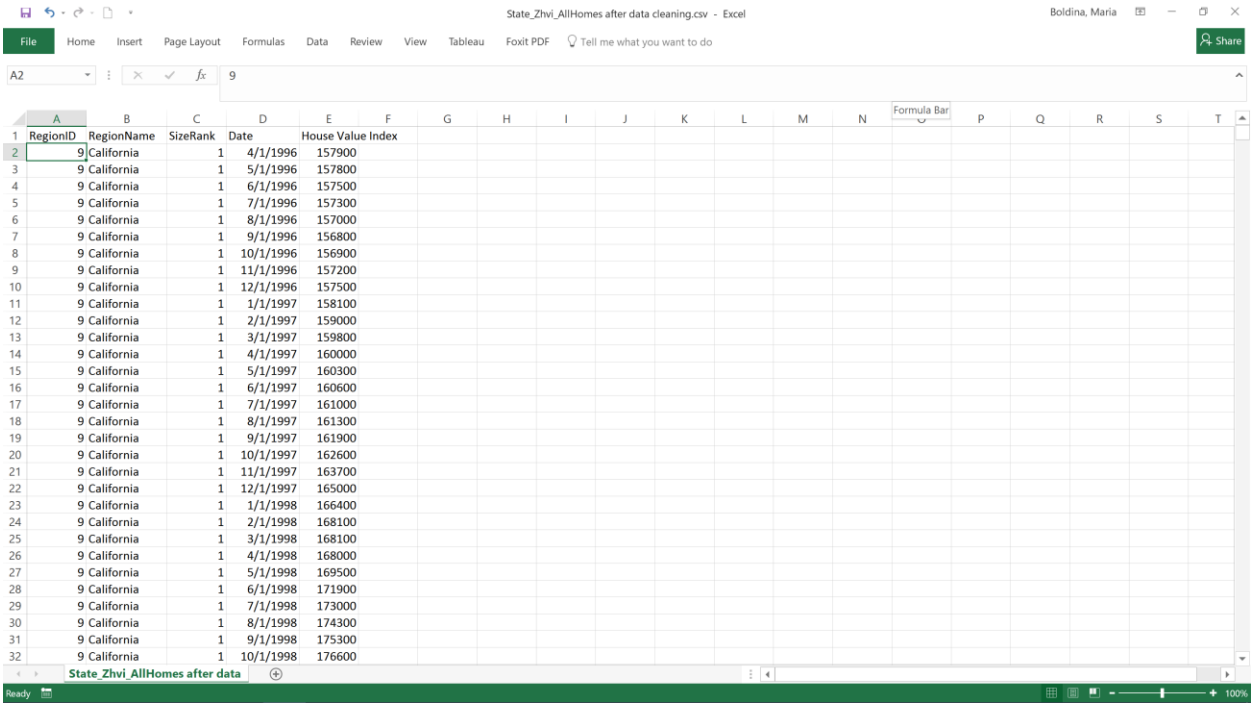


RegionID	RegionNam	SizeRank	1996-04	1996-05	1996-06	1996-07	1996-08	1996-09	1996-10	1996-11	1996-12	1997-01	1997-02	1997-03	1997-04	1997-05	1997-06	1997-07	1997-08	199
9	California	1	157900	157800	157500	157300	157000	156800	156900	157200	157500	158100	159000	159800	160000	160300	160600	161000	161300	1
54	Texas	2											81200	81600	82000	82400	82900	83400	84000	
43	New York	3																		
14	Florida	4	86300	86600	86700	86600	86500	86500	86400	86700	86900	87200	87700	88200	88200	88300	88700	89100	89300	
21	Illinois	5	113900	114400	114500	113900	113100	112900	113000	113300	114000	115100	115900	116000	115900	116000	116200	116600	117000	1
47	Pennsylvar	6	83300	83700	83800	83600	83700	83900	83900	83700	83800	84000	84200	84600	84800	84700	84800	85100	85200	
44	Ohio	7	88100	88300	88500	88700	88800	89100	89400	89600	89700	90100	90500	90900	91200	91500	91800	92200	92400	
30	Michigan	8	87600	87100	87300	88200	87900	87300	87000	87500	87700	87600	87600	88000	88600	89400	90000	90300	90700	
16	Georgia	9	92000	92400	92600	92700	92700	92900	93300	93700	94000	94300	94700	95000	95200	95700	96200	96900	97600	
36	North Caro	10	91700	91900	92000	92300	92600	92900	93100	93400	93700	94300	94900	95400	96100	97000	97700	98200	98800	
40	New Jersey	11	140300	139900	139700	139600	139300	139200	139100	139100	139100	139400	139900	140400	140700	141000	141300	141400	141400	1
56	Virginia	12	111600	112200	112600	112800	112900	112700	112800	113100	113300	113500	114000	114400	114300	114300	114500	114500	114600	1
59	Washingto	13	134400	134800	135100	135100	135200	135400	135900	136300	136800	137600	138500	139100	139700	140400	141300	142100	142900	1
26	Massachus	14	137300	137700	138000	138200	138500	138900	139300	139800	140600	141500	142500	143300	144000	144500	145200	145900	146200	1
22	Indiana	15	72100	74200	75200	75400	75200	75200	75400	75600	75700	76300	77500	78500	78500	78400	78700	79300	80000	
8	Arizona	16	103600	103900	104100	104400	104600	104900	105100	105400	105700	106300	106900	107200	107400	107600	108000	108300	108700	1
53	Tennessee	17	70300	70700	71000	71400	71600	72000	72100	72400	72700	73100	73600	74100	74400	74700	75100	75500	75900	
32	Missouri	18	80700	81400	82000	82600	83300	83700	83700	83600	84200	85500	86700	87600	88000	88200	88900	89500	90000	
27	Maryland	19	127000	126900	126600	126400	126200	126100	126100	126200	126400	126700	127100	127300	127200	127300	127300	127200	127100	1
60	Wisconsin	20																	99000	
31	Minnesota	21	95700	96200	96200	96200	96200	96200	96200	96300	96600	97300	98200	98900	99400	99900	100200	100600	100900	1
10	Colorado	22	128100	128400	128700	128900	129300	129800	130300	130800	131200	131800	132600	133200	133600	134000	134600	135200	135800	1
4	Alabama	23	79500	79200	79500	79700	79700	79700	80100	80600	80800	80700	80900	81300	81300	80900	80900	81100	81400	
51	South Caro	24	80000	80000	80100	80000	79800	79700	79800	80100	80400	80900	81200	81400	81800	82000	82200	82700	83500	
24	Kentucky	25	74300	75100	76200	77300	78300	79000	79500	79700	79900	80100	80400	80400	80300	80400	80700	80900	81100	
46	Oregon	26	124500	125200	125700	126200	126900	127600	128200	129000	129700	130300	131000	131500	132000	132400	132900	133500	134000	1
45	Oklahoma	27	56400	56400	56500	56500	56500	56700	56800	57000	57300	57500	57500	57700	58000	58300	58500	58900	59100	
11	Connecticu	28	132000	132400	132700	132900	133100	133200	133100	133200	133500	134000	134600	135100	135300	135600	135900	136100	136400	1
19	Iowa	29																		
34	Mississippi	30	64200	64200	64400	64600	65000	65500	65600	65500	65400	65200	65000	65100	65300	65600	65900	66100	66500	
6	Arkansas	31	64400	64400	64400	64300	64100	64100	64300	64500	64700	65000	65300	65800	66500	66900	67000	67100	67200	

Using Tableau Data Reshaper add-in for *Excel*



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



RegionID	RegionName	SizeRank	Date	House Value Index
9	California	1	4/1/1996	157900
9	California	1	5/1/1996	157800
9	California	1	6/1/1996	157500
9	California	1	7/1/1996	157300
9	California	1	8/1/1996	157000
9	California	1	9/1/1996	156800
9	California	1	10/1/1996	156900
9	California	1	11/1/1996	157200
9	California	1	12/1/1996	157500
9	California	1	1/1/1997	158100
9	California	1	2/1/1997	159000
9	California	1	3/1/1997	159800
9	California	1	4/1/1997	160000
9	California	1	5/1/1997	160300
9	California	1	6/1/1997	160600
9	California	1	7/1/1997	161000
9	California	1	8/1/1997	161300
9	California	1	9/1/1997	161900
9	California	1	10/1/1997	162600
9	California	1	11/1/1997	163700
9	California	1	12/1/1997	165000
9	California	1	1/1/1998	166400
9	California	1	2/1/1998	168100
9	California	1	3/1/1998	168100
9	California	1	4/1/1998	168000
9	California	1	5/1/1998	169500
9	California	1	6/1/1998	171900
9	California	1	7/1/1998	173000
9	California	1	8/1/1998	174300
9	California	1	9/1/1998	175300
9	California	1	10/1/1998	176600

Missing Values:

247004	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	1997	285700
247005	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	1998	301700
247006	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	1999	322800
247007	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2000	344400
247008	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2001	384700
247009	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2002	431900
247010	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2003	518100
247011	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2004	590700
247012	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2005	722200
247013	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2006	842700
247014	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2007	831300
247015	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2008	809700
247016	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2009	766100
247017	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2010	797300
247018	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2011	781500
247019	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2012	727100
247020	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2013	770300
247021	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2014	782900
247022	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2015	794600
247023	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2016	800900
247024	21405	Annapolis	MD	Baltimore	Anne Arundel	13102	2017	771400
247025	85220	Apache Junction	AZ	Phoenix	Pinal	13103	1999	80400
247026	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2000	91100
247027	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2001	100900
247028	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2002	104800
247029	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2003	116700
247030	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2004	122700
247031	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2005	155200
247032	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2006	205800
247033	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2007	212500
247034	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2008	189100
247035	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2009	164900
247036	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2010	131600
247037	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2011	141700
247038	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2012	143600
247039	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2013	157600
247040	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2014	170200
247041	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2015	153700
247042	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2016	179600
247043	85220	Apache Junction	AZ	Phoenix	Pinal	13103	2017	182000

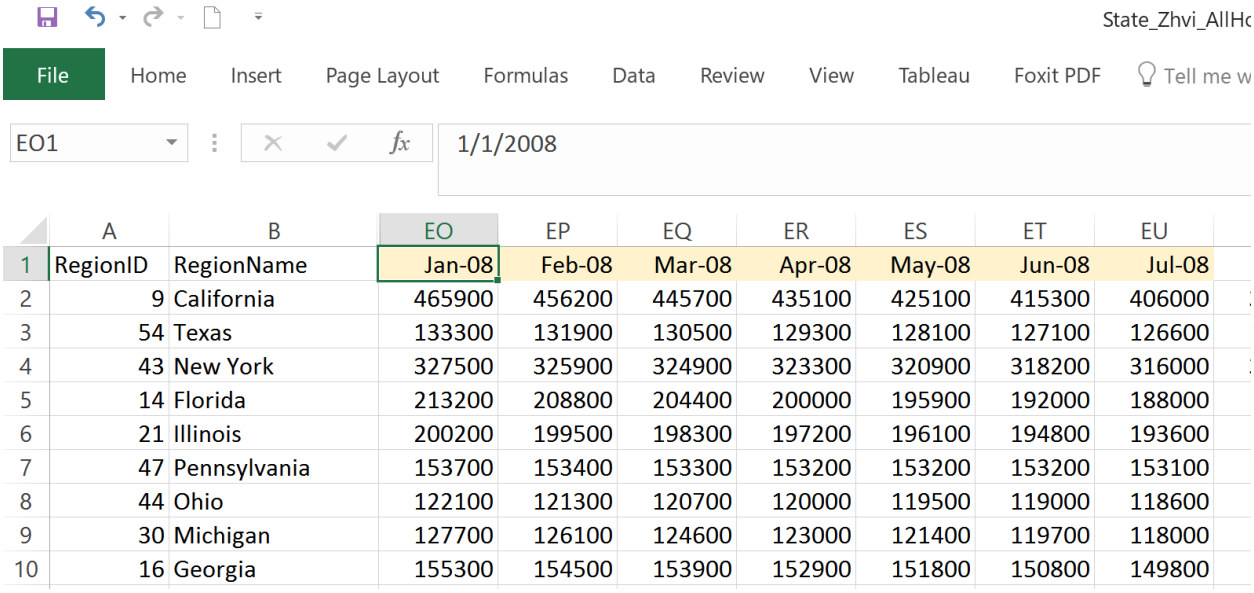
Please note that some data is null for certain zip codes in the years before the end of 2008 because such data was not being collected at the time. Instead of truncating the data to range only from 2008-2017 I decided to keep all the data because it could still be interesting to look at the time series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

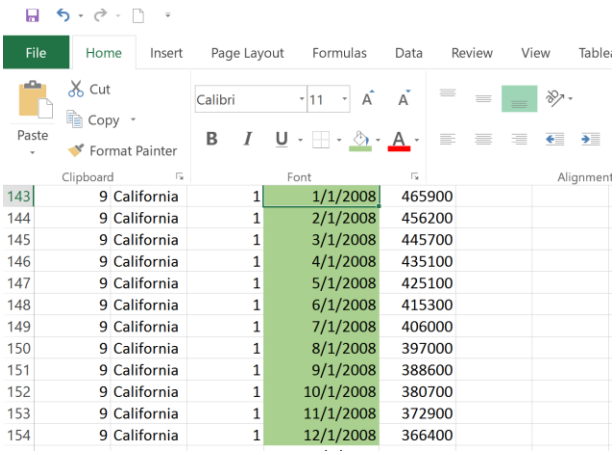
Date was standardized to be as MM/DD/YYYY or YYYY where monthly data is not provided, Original data was in Month-YYYY date format which is not ideal for Tableau and is not recognized as a date.

Before:



	A	B	EO	EP	EQ	ER	ES	ET	EU
1	RegionID	RegionName	Jan-08	Feb-08	Mar-08	Apr-08	May-08	Jun-08	Jul-08
2	9	California	465900	456200	445700	435100	425100	415300	406000
3	54	Texas	133300	131900	130500	129300	128100	127100	126600
4	43	New York	327500	325900	324900	323300	320900	318200	316000
5	14	Florida	213200	208800	204400	200000	195900	192000	188000
6	21	Illinois	200200	199500	198300	197200	196100	194800	193600
7	47	Pennsylvania	153700	153400	153300	153200	153200	153200	153100
8	44	Ohio	122100	121300	120700	120000	119500	119000	118600
9	30	Michigan	127700	126100	124600	123000	121400	119700	118000
10	16	Georgia	155300	154500	153900	152900	151800	150800	149800

After:

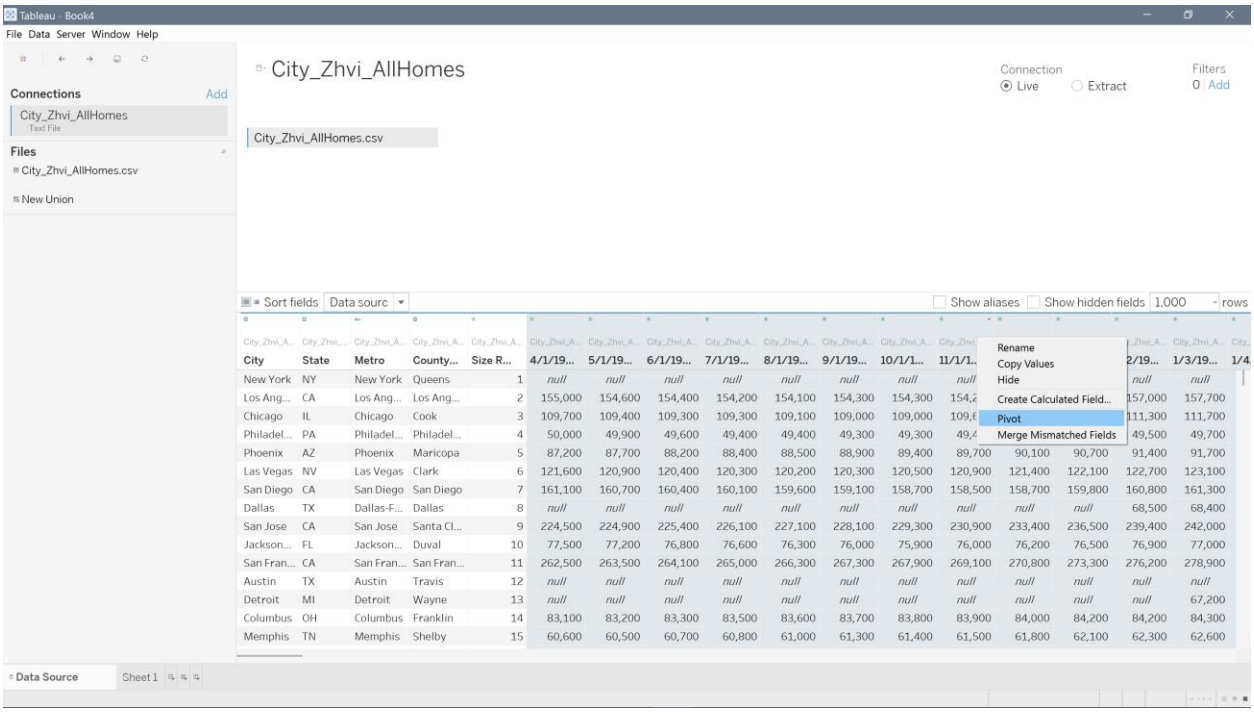


143	9	California	1	1/1/2008	465900
144	9	California	1	2/1/2008	456200
145	9	California	1	3/1/2008	445700
146	9	California	1	4/1/2008	435100
147	9	California	1	5/1/2008	425100
148	9	California	1	6/1/2008	415300
149	9	California	1	7/1/2008	406000
150	9	California	1	8/1/2008	397000
151	9	California	1	9/1/2008	388600
152	9	California	1	10/1/2008	380700
153	9	California	1	11/1/2008	372900
154	9	California	1	12/1/2008	366400

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.

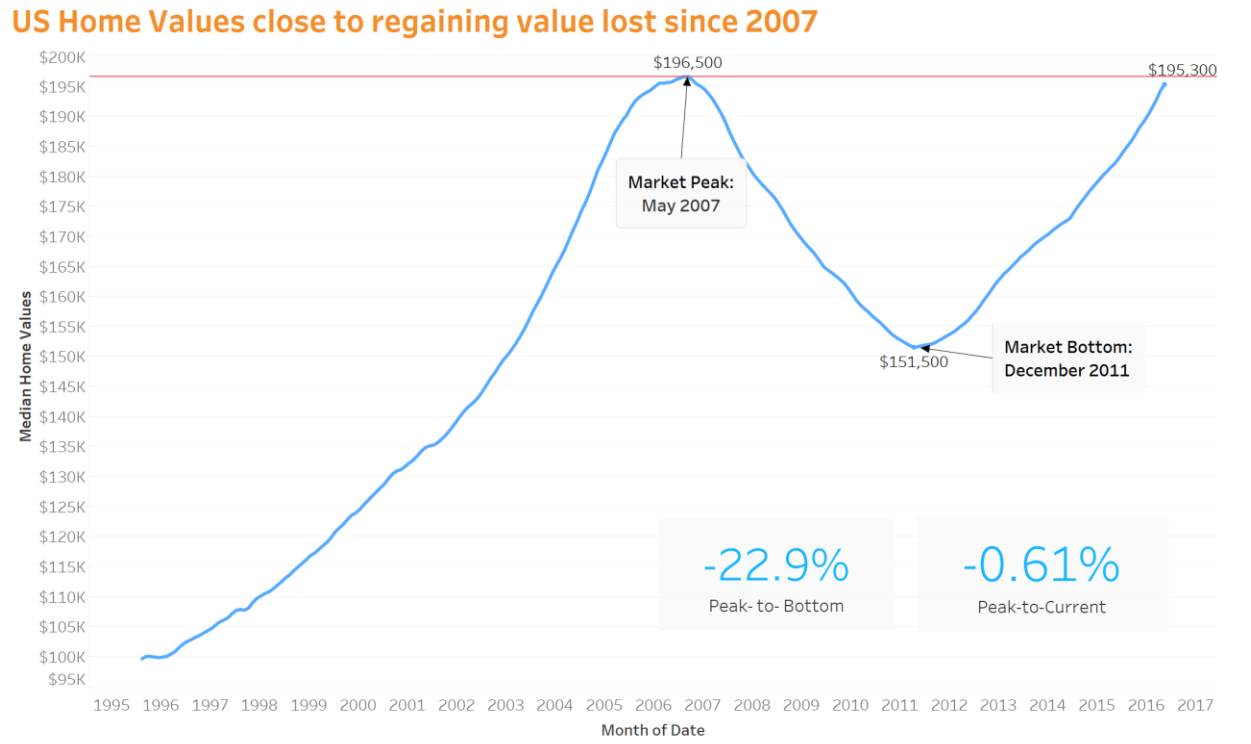
Before:



City	State	Metro	County...	Size R...	4/1/19...	5/1/19...	6/1/19...	7/1/19...	8/1/19...	9/1/19...	10/1/1...	11/1/1...	1/3/19...
New York	NY	New York	Queens	1	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
Los Ang...	CA	Los Ang...	Los Ang...	2	155,000	154,600	154,400	154,200	154,100	154,300	154,300	154,2	157,700
Chicago	IL	Chicago	Cook	3	109,700	109,400	109,300	109,300	109,100	109,000	109,000	109,(	111,700
Philadel...	PA	Philadel...	Philadel...	4	50,000	49,900	49,600	49,400	49,400	49,300	49,300	49,4	49,700
Phoenix	AZ	Phoenix	Maricopa	5	87,200	87,700	88,200	88,400	88,500	88,900	89,400	89,700	91,700
Las Vegas	NV	Las Vegas	Clark	6	121,600	120,900	120,400	120,300	120,200	120,300	120,500	120,900	123,100
San Diego	CA	San Diego	San Diego	7	161,100	160,700	160,400	160,100	159,600	159,100	158,700	158,500	161,300
Dallas	TX	Dallas-F...	Dallas	8	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	68,400
San Jose	CA	San Jose	Santa Cl...	9	224,500	224,900	225,400	226,100	227,100	228,100	229,300	230,900	242,000
Jackson...	FL	Jackson...	Duval	10	77,500	77,200	76,800	76,600	76,300	76,000	75,900	76,000	77,000
San Fran...	CA	San Fran...	San Fran...	11	262,500	263,500	264,100	265,000	266,300	267,300	267,900	269,100	278,900
Austin	TX	Austin	Travis	12	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
Detroit	MI	Detroit	Wayne	13	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	67,200
Columbus	OH	Columbus	Franklin	14	83,100	83,200	83,300	83,500	83,600	83,700	83,800	83,900	84,300
Memphis	TN	Memphis	Shelby	15	60,600	60,500	60,700	60,800	61,000	61,300	61,400	61,500	62,600

PART C. DATA VISUALIZATIONS

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

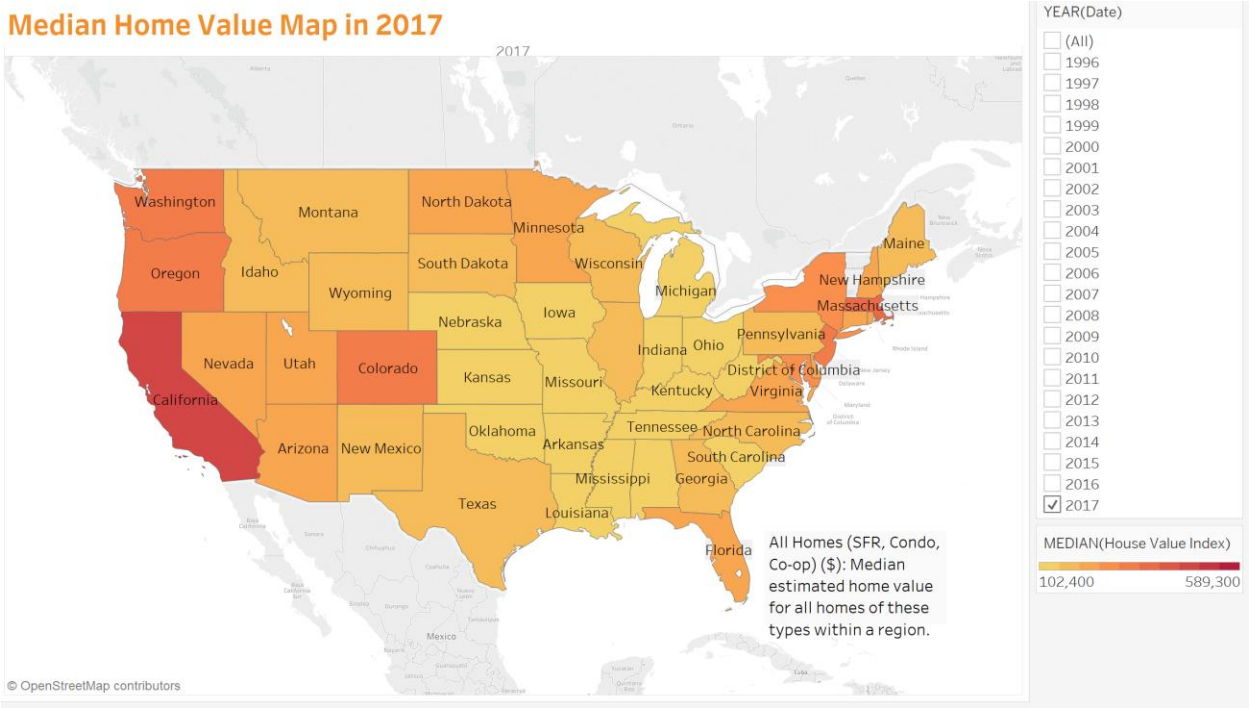


As depicted in the line chart above the median house value in the United States s of January 2017 is $195,300. This that the house values have almost doubled since 1996. At its peak level the median house prices have reached $196,500 in May 2007. However, as evident from the chart, the US house value 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*

2. Which States have the highest median home value in 2017?



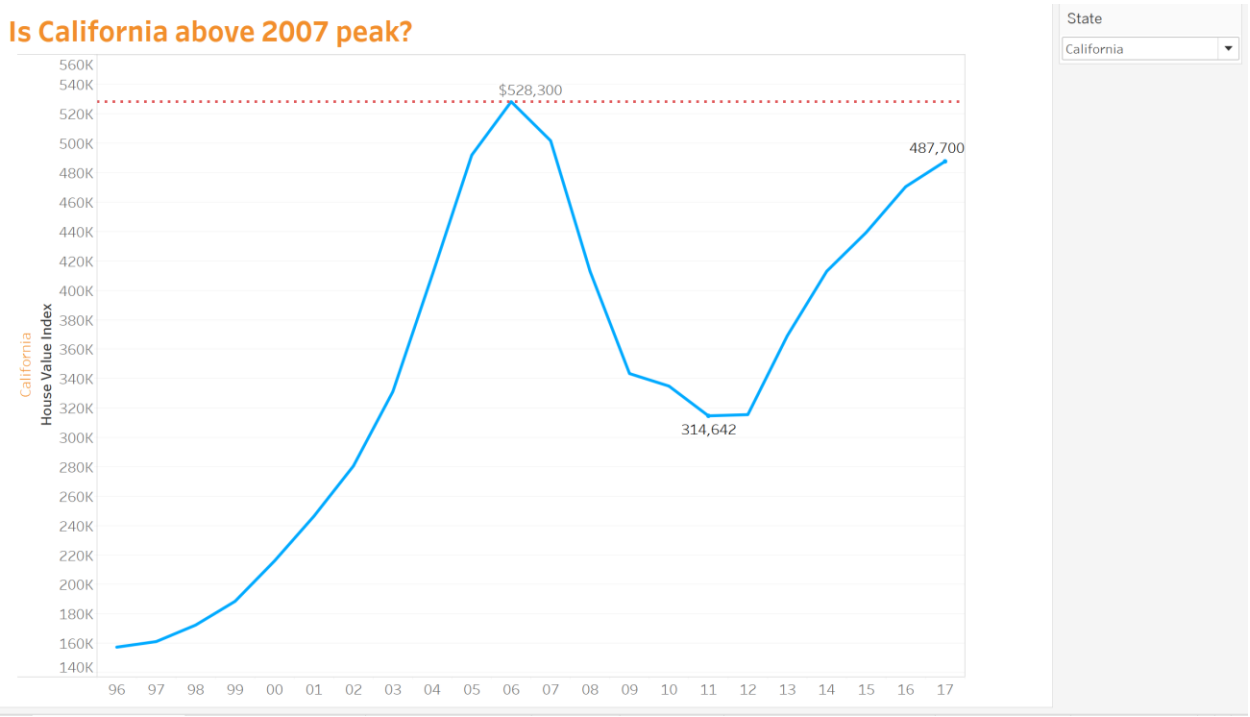
As we can easily see on the above map the majority of states with the highest median house values in 2017 are located in the west of the United States. (Please note that although Hawaii and Alaska, are not immediately visible on this map, they could be easily seen by zooming out). The lowest house value 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*

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

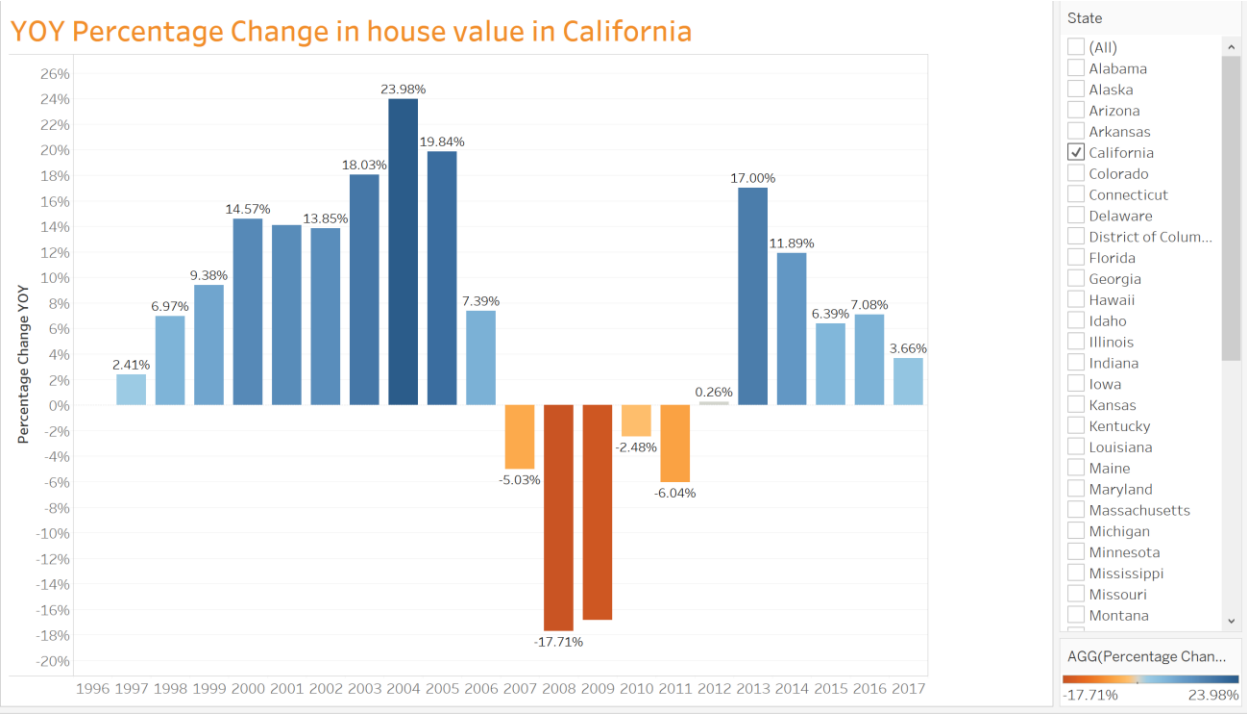


As of January 2017 median house value in California is $487,700. Therefore, interestingly enough California house prices still have not reached their 2006 peak of $528,300. Although they have regained most of the value lost since the 2011 bottom of $314,642. We can also do the same analysis for any of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

Categories used

- *Reference line*
- *Dates*

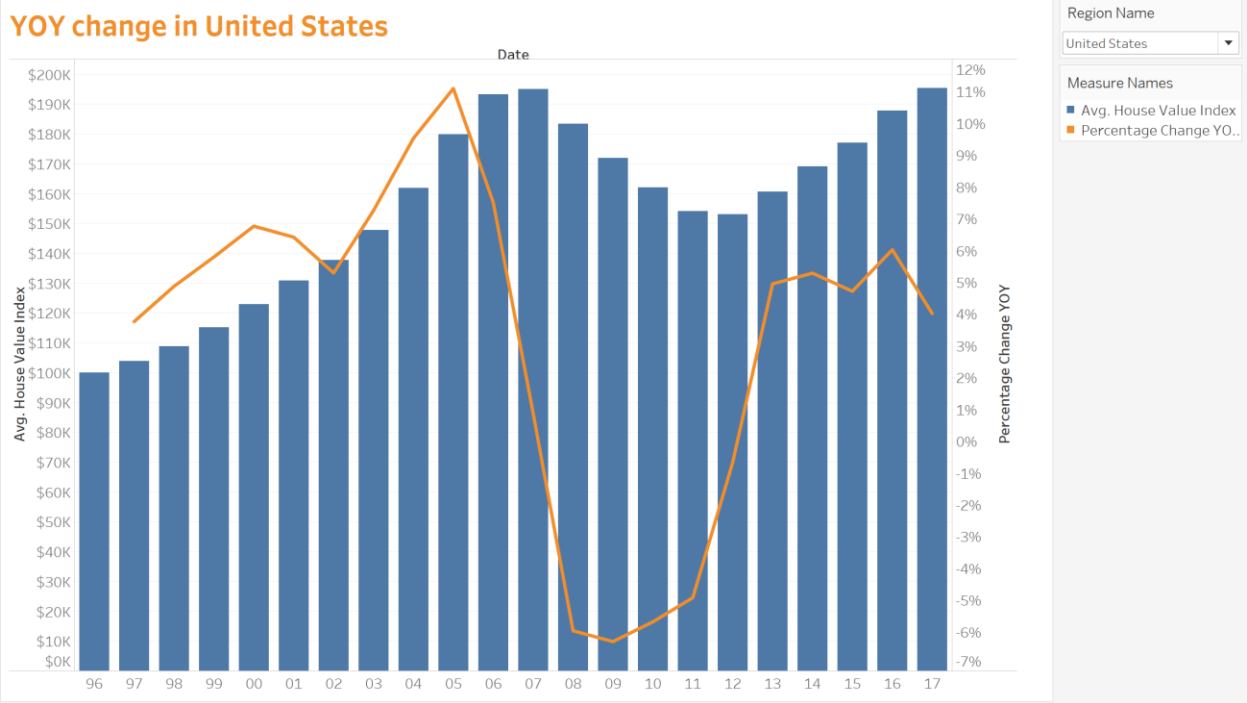
4. **What is the annual percentage change in home value appreciation in each state?**



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*

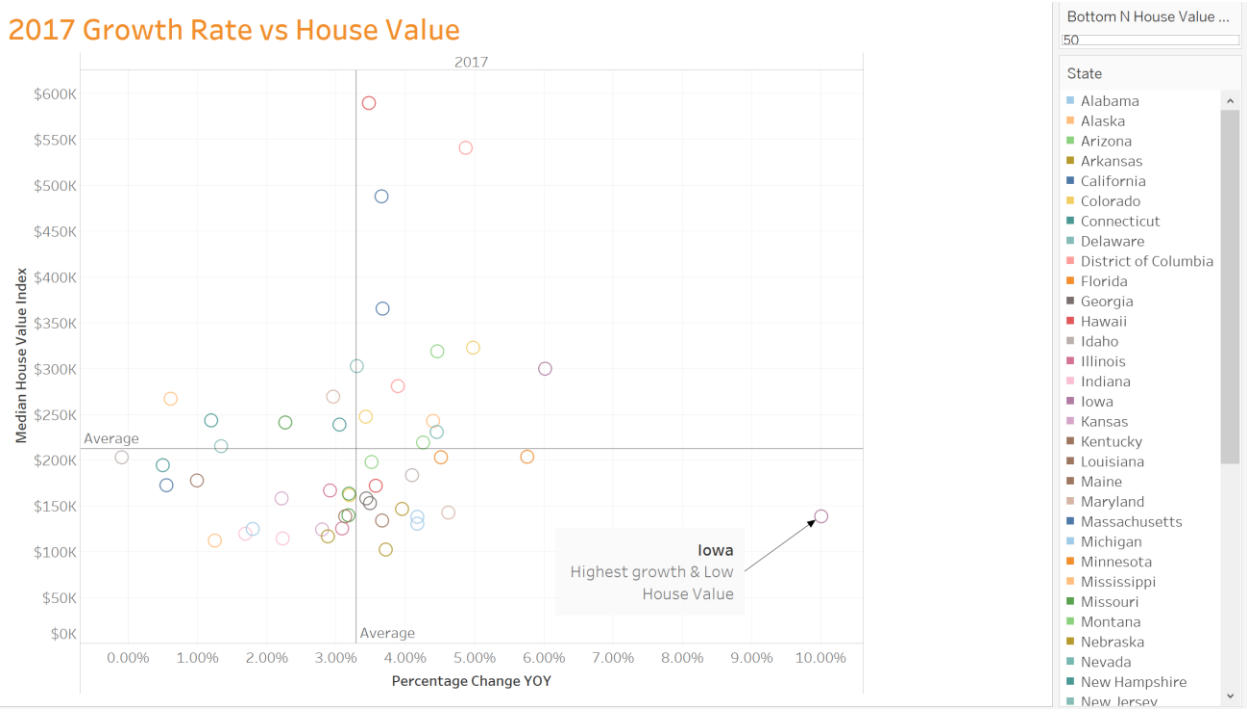


This chart is similar to the one above but due to the dual axis we can see both the average house value in any given year as depicted by the blue bar charts and the corresponding year over year percent change which is shown by orange line chart. This view can be shown for the United States in general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

Categories used:

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

5. **Which areas exhibit the fastest growth and have the lowest house prices?**



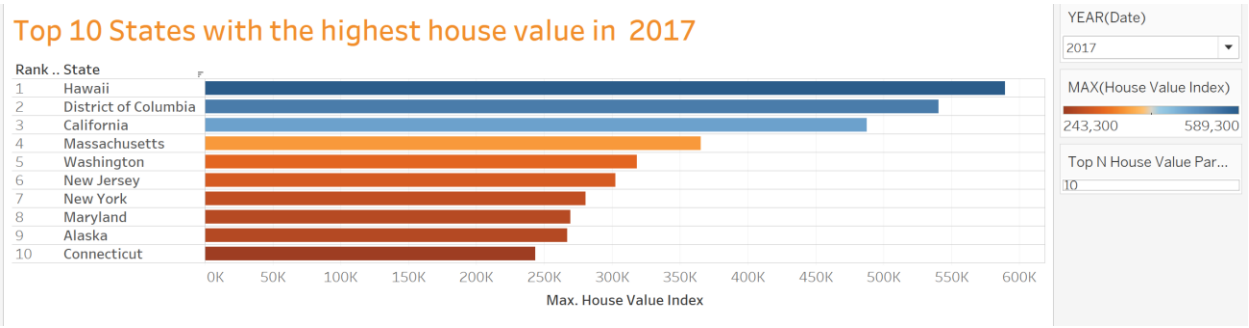
The scatter plot above makes it easy to see which states have the highest percentage of growth in 2017 as compared to 2016. For example, we can tell that the states represented by the circles in the bottom right quadrant have higher than average percentage of growth and lower than average house values. 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)*

6. **What are the top 10 States with the highest house value index?**

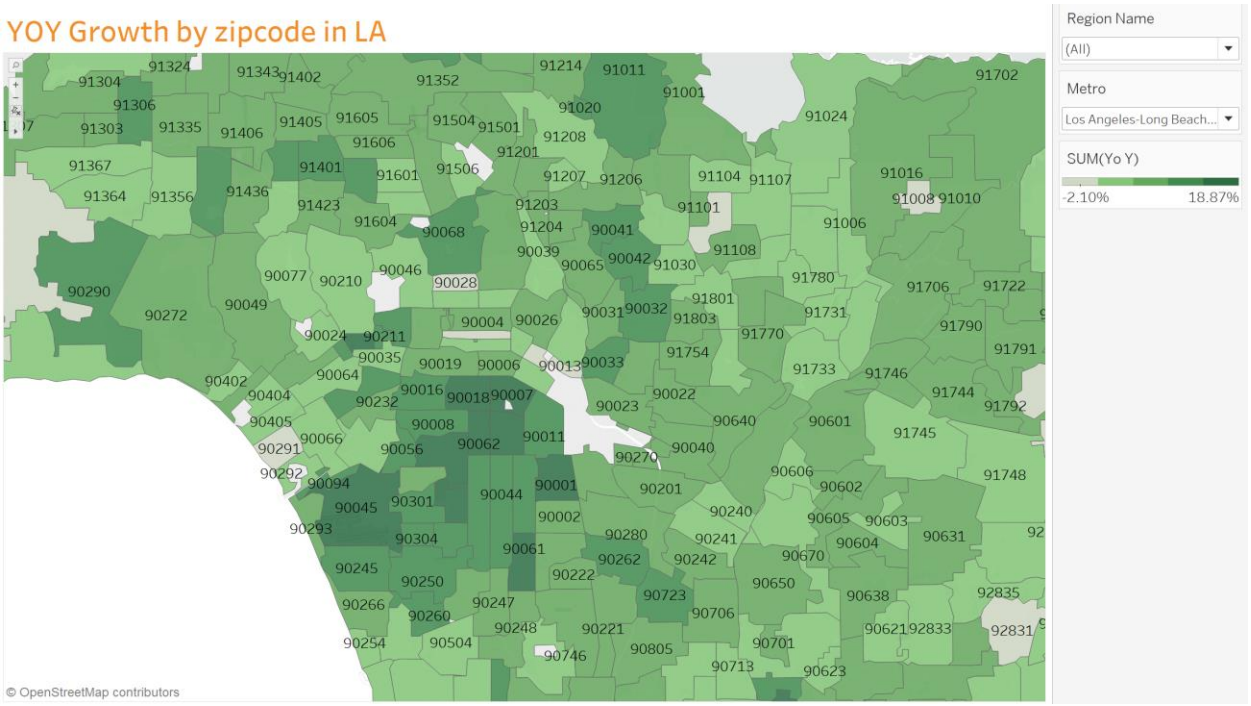


The bar chart above ranks the top 10 states with the highest house value index. As we can see Hawaii has the highest house value index of $589,300, followed by District of Columbia in the second place and California in third.

Categories used:

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

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

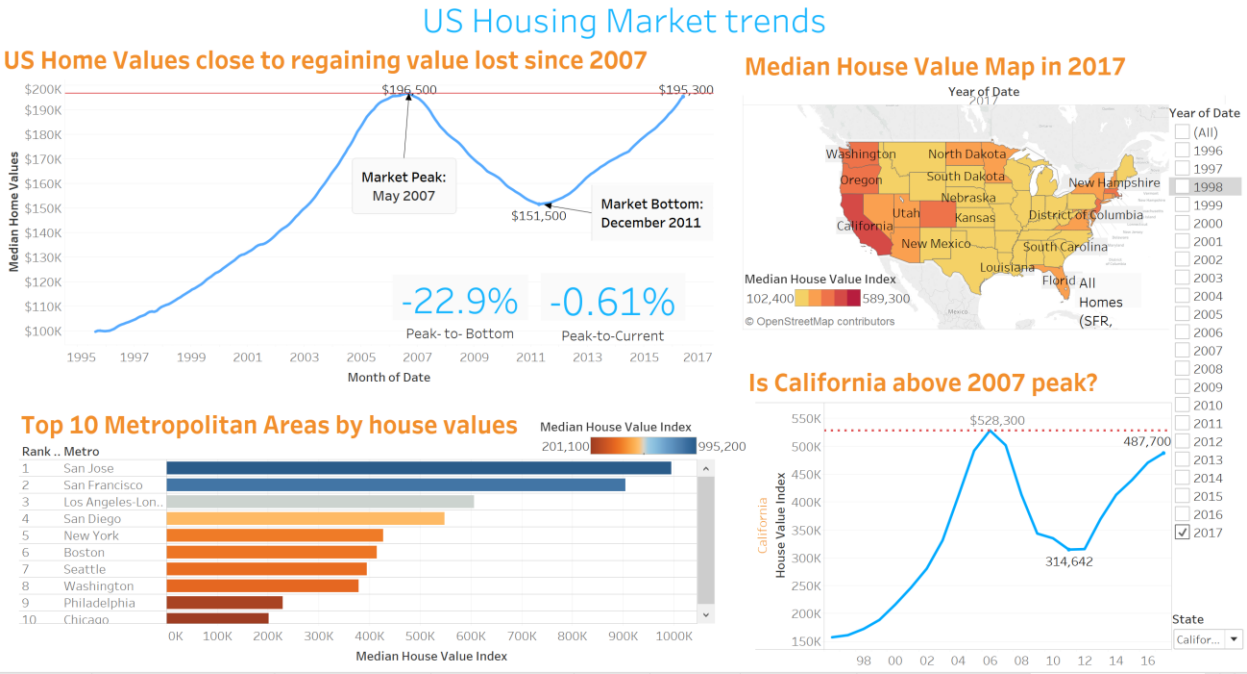


Looking at the map of the Greater Los Angeles area we can see that some neighborhoods have gained as much as 18.87% as compared to last year. Most of the zip codes in the area exhibit steady growth with very few exceptions as can be seen by mostly darker green areas on the map.

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.

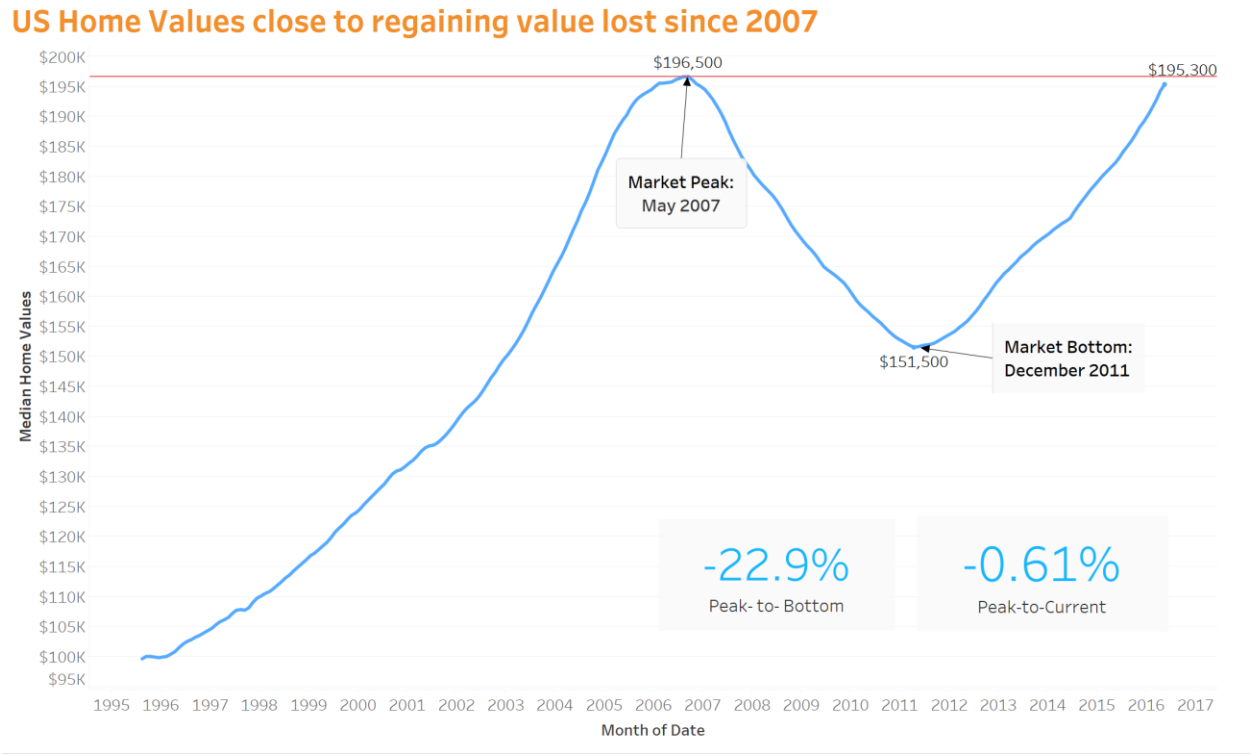


Figure 1

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 gains[2].

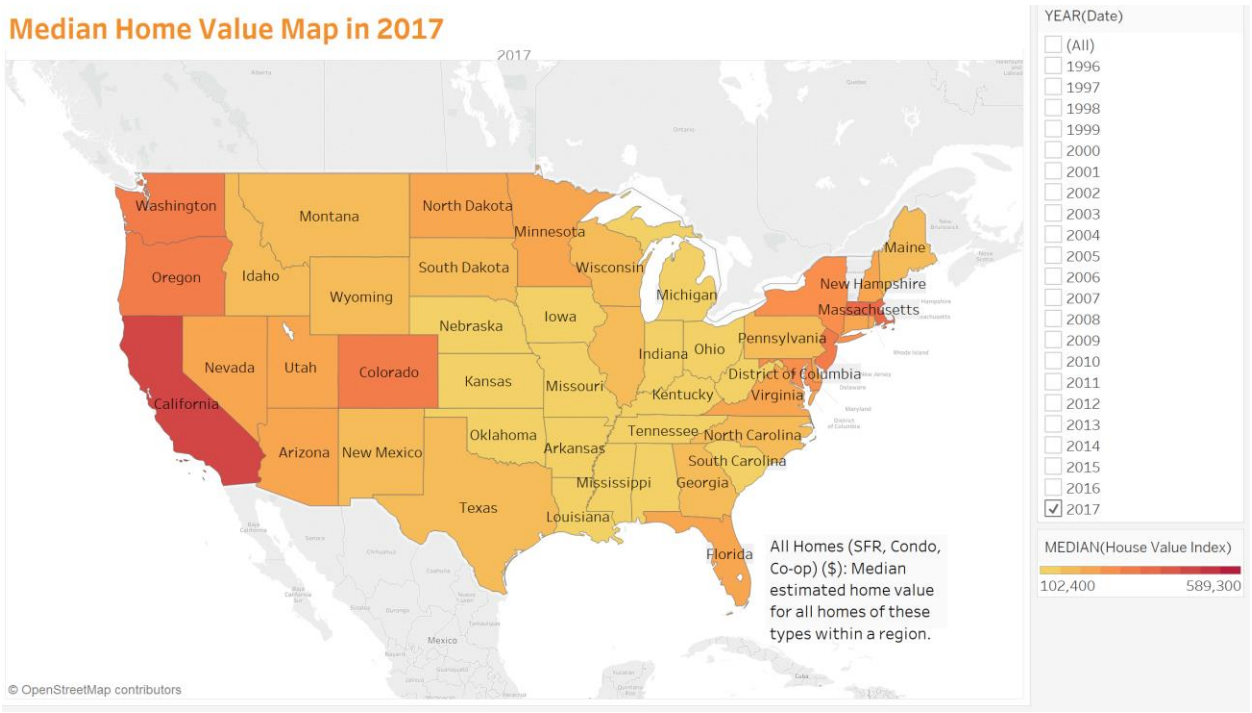


Figure 2

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.

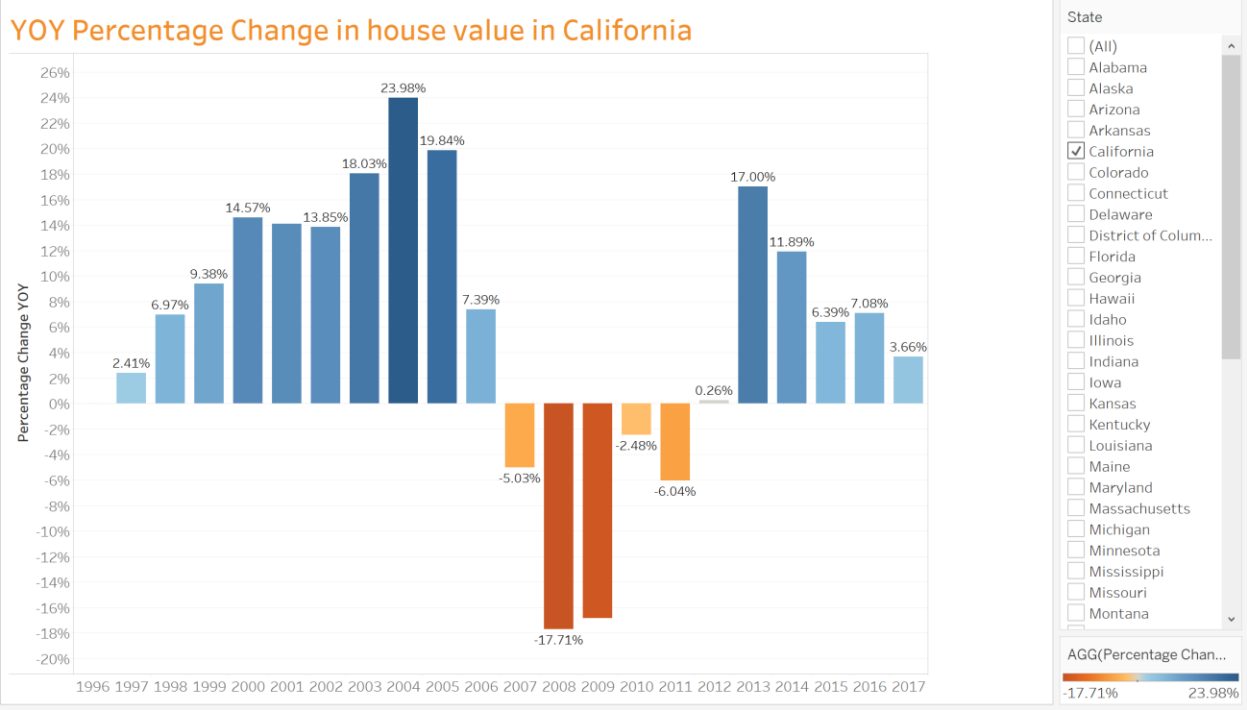


Figure 3

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.

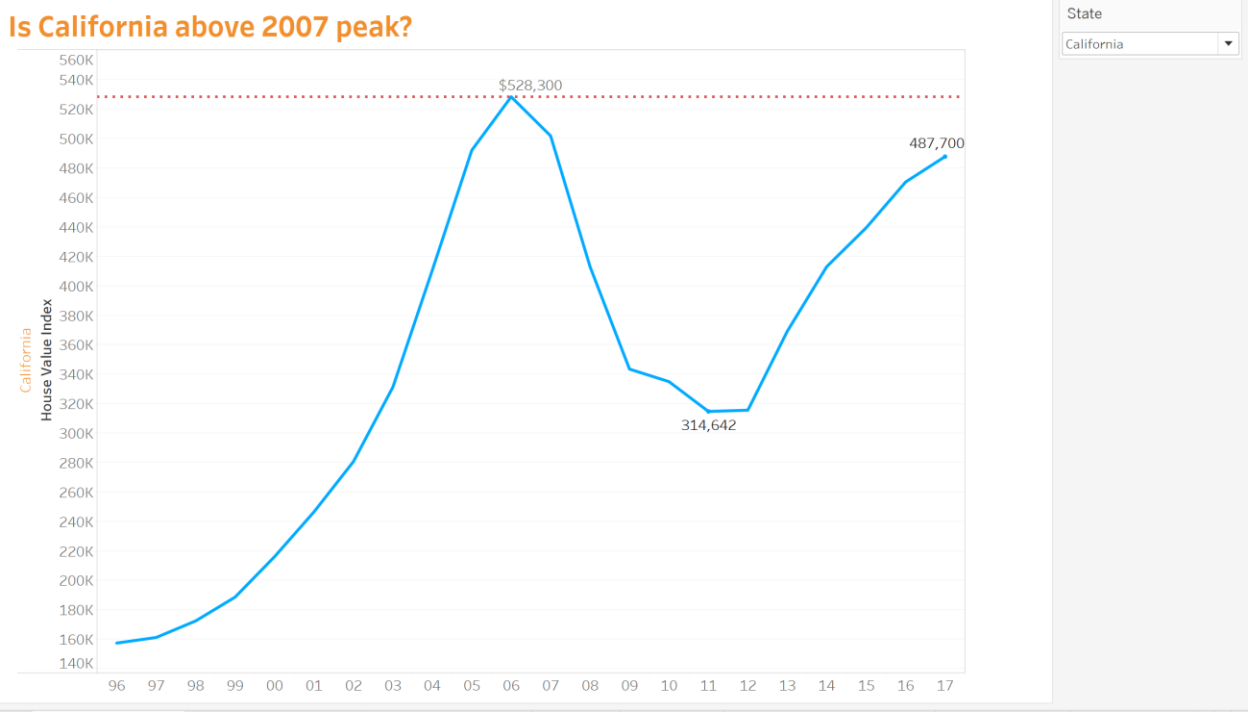


Figure 4

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.

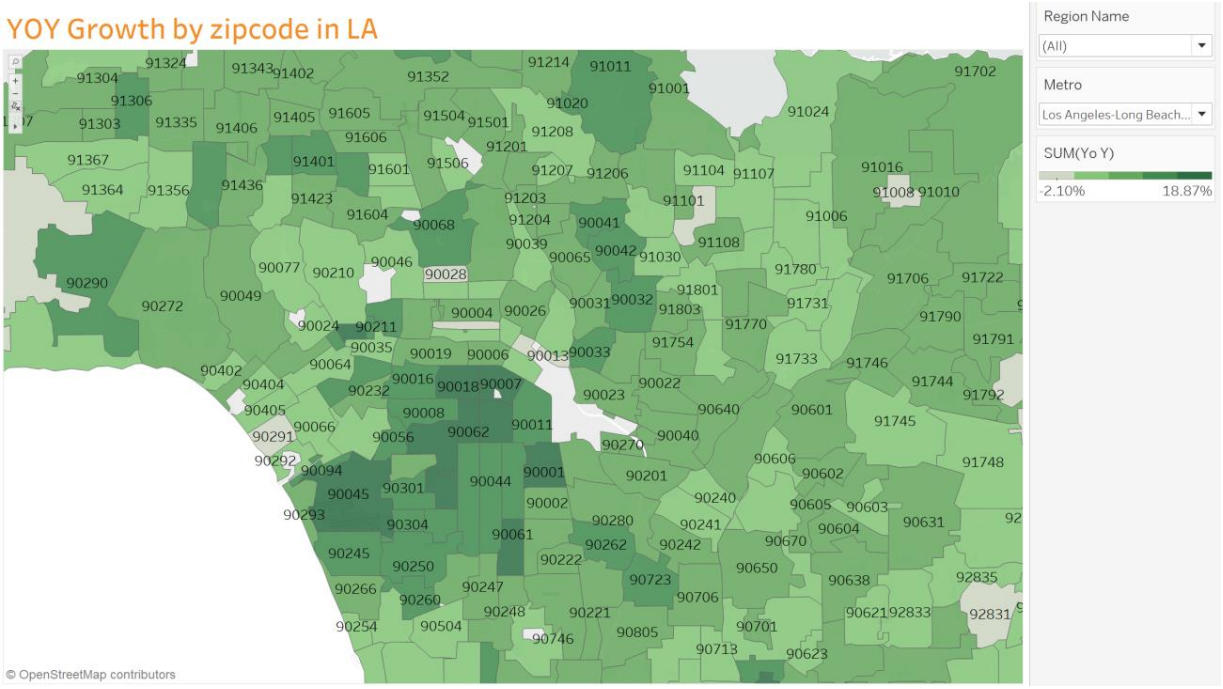


Figure 5

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.