University of California Berkeley
Stat 133 Final Project Report

Bay Area Housing Data Analysis

August 9,2016

Group 5: Xinhao Li, Kentaro Ino, Yunshan Guo, Kevin Khuu

Table of Content

1. Introduction

Housing markets are becoming increasingly significant in shaping the economic and social well-being of many Americans. In this paper, we are restricting our analysis of the housing market to the Bay Area as we feel current economic trends such as the Tech Boom have been very receptive to the housing market. Substantial variation exists across neighborhoods in the type of housing available, the quality of public services, the level of tax burdens, and the quality of life. Consequently, households confront important tradeoffs between different types of housing, neighborhood characteristics, and accessibility to place of work. Since housing expenditures are a large component of every household's budget, the availability of housing and its price assume considerable importance to quality of life. Moreover, housing markets play a central role in the process of metropolitan development, both affecting and reflecting other forces at work in this sphere. Urban development patterns, in turn, #are crucial to our future welfare in many ways.

Our research is intended to present housing conditions in the San Francisco Bay Area in the context of historic trends and expectations for future. Specifically, this research aims to identify changes in the housing market in terms of significant characteristics of housing and their corresponding effect on home values. After collecting a large set of relevant data from sources, we focused mainly on three different parts of the housing market: the major growth tendency of housing prices prominent in each county, the difference in housing prices between different counties, and the relationship between population and income average and housing price.

With this analysis, we can clearly map how the local housing environment. And the housing prices could be a good metric to extrapolate the general health of the economy in the Bay Area. According to the results from analysis, we are able to interpret the extent of how large historical events, such as the economic recession in 2008, affected the housing market. Besides, by analyzing both demographic trends and population intensity, trends in housing prices could be revealed.

There are several websites which contain straightforward housing price information on specific locations, however, these sites only recorded a limited, non-representative amount of data. Through our research, we manage to glean a substantial amount of housing data from each county to precisely demonstrate the average housing price for each county, which presents a overview of residential market on Bay Area.

2. Data Collection and Wrangling process

2.1 Raw Data Collection

Xinhao suggested Quandl, a search engine primarily for numerical data, which offered access to several million financial, economic and social datasets We used the API to glean housing data from Zillow and Economic data from the Federal Reserve Economic Data. By inputting the code indicating the area category and area code number for the relevant US counties, Quandl compiled a dataset containing all of the included area categories and code numbers. In order to obtain a large representative sample size, instead of manually inputting codes, we constructed a lookup code script only selecting the information about the Bay Area to do a loop for searching and avoid repeating the input process. Similarly, we used the same technique to obtain GDP and population growth data on Bay Area. We also included Sacramento County as a foil to Bay Area counties to see if certain economic trends were also felt in counties not centered around the tech boom. One glaring omission that readers might notice from this paper is the absence of Santa Clara data from the Price-to-Rent Ratio graphs and the Home foreclosure rate graphs. Unfortunately, due to complications from the Quandl API regarding access requests, we had to exclude Santa Clara county data.

Below is a sample of raw data obtained from Quandl

Value stands for the value of the variable Type.

Type A stands for average price for all homes)

	Date	Value	City	County	Metro	Type
c1	5/31/16	272200	Sacramento	Sacramento	Sacramento	A
2	4/30/16	269000	Sacramento	Sacramento	Sacramento	A
3	3/31/16	266600	Sacramento	Sacramento	Sacramento	A
4	2/29/16	262300	Sacramento	Sacramento	Sacramento	A
5	1/31/16	259600	Sacramento	Sacramento	Sacramento	A
6	12/31/15	256600	Sacramento	Sacramento	Sacramento	A
7	11/30/15	253100	Sacramento	Sacramento	Sacramento	A
8	10/31/15	252000	Sacramento	Sacramento	Sacramento	A
9	9/30/15	278600	Sacramento	Sacramento	Sacramento	A

2.2 Aggregation Process

Since we wanted to present the growth tendency of the Bay Area housing market during the past two decades, we inner joined all the separate datasets by year and county so that we have a master data file containing all of the relevant housing data needed to conduct our analysis.

Below is a sample cleaned data (twoB stands for the average price of two bedroom properties for that specific county, while threeB stands for the average price of three bedroom properties)

Population is in units of 1000s

region	subregion	Year	Pop	Income	twoB	threeB
california	alameda	1996	1359.099	28535	152197.7778	226967.3611
california	alameda	1997	1380.383	29971	171414.4444	239140.1042
california	alameda	1998	1405.903	32234	189932.7778	267451.5625
california	alameda	1999	1427.114	34513	211057.7778	298254.6875
california	alameda	2000	1450.086	39093	277123.3333	380213.5417
california	alameda	2001	1468.652	38991	318203.8889	426436.9792
california	alameda	2002	1460.438	39619	338772.7778	449368.75
california	alameda	2003	1451.418	41226	374987.2222	493705.2083
california	alameda	2004	1441.496	43140	438918.3333	570041.1458
california	alameda	2005	1435.87	44745	521633.8889	678390.1042
california	alameda	2006	1437.154	47382	537928.3333	690273.4375
california	alameda	2007	1447.863	48133	512087.2222	667007.2917

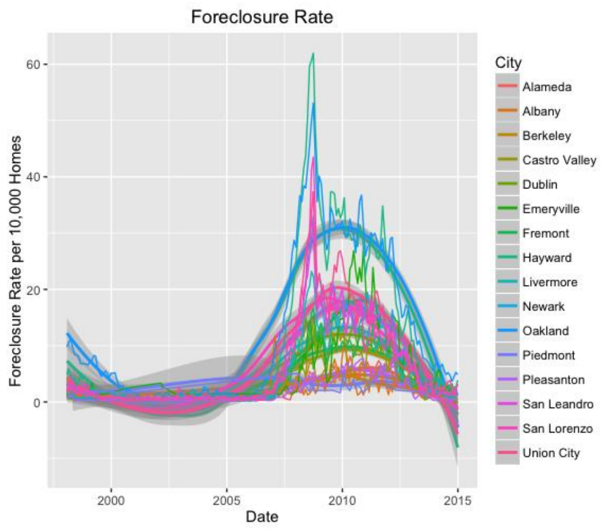
3. Analysis and Data Visualization

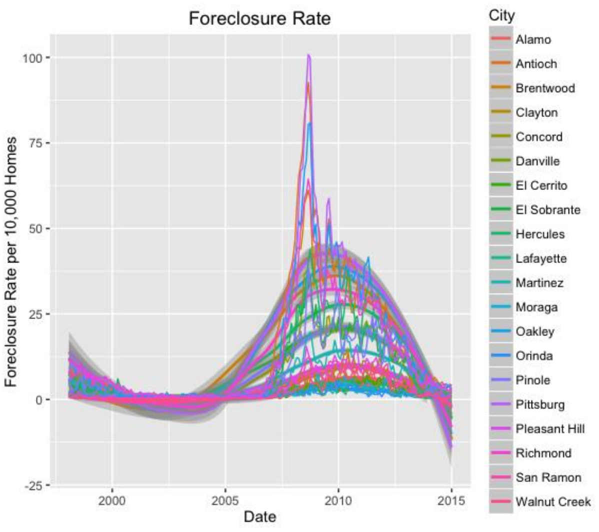
3.1 Identifying Major Factors Affecting Housing Price

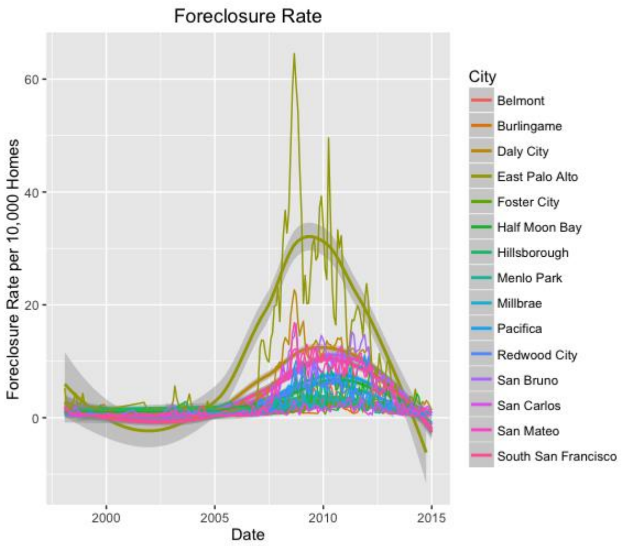
I.Home Foreclosure Rates

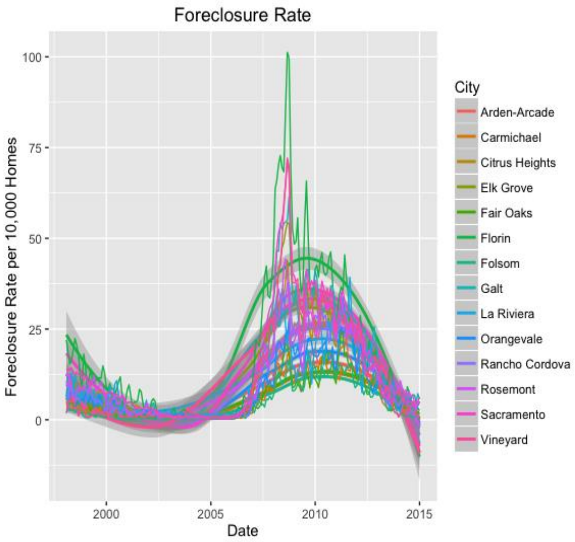
Measuring Home Foreclosure Rates We can see that home foreclosure rates remained relatively low and stable up until the 2008 Recession. This spike happens to coincide with the end of teaser rates for Adjustable Rate Mortgages. Home foreclosure rates have since lowered

Another factor is Price-to-Rent Ratio. The Bay area has a very high average Price-to-Rent Ratio. Markets with very high average Price-to-Rent ratios. However, Price-Rent-Ratio says nothing about the absolute cost of living in the Bay Area, which is one of the highest in the country. This data is of Alameda , Contra Costa, San Mateo and Sacramento county.



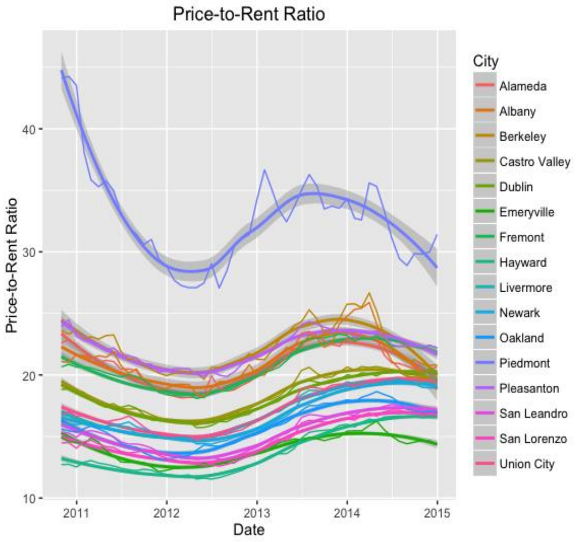


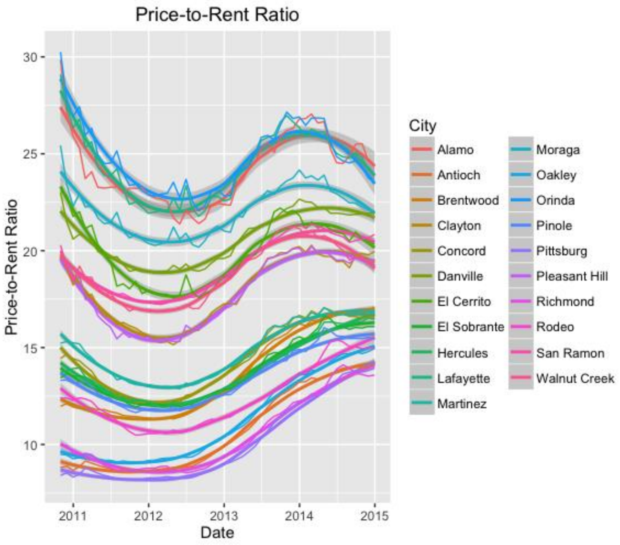


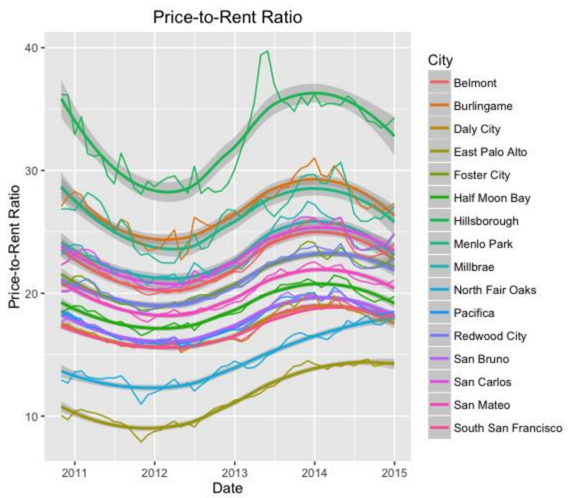


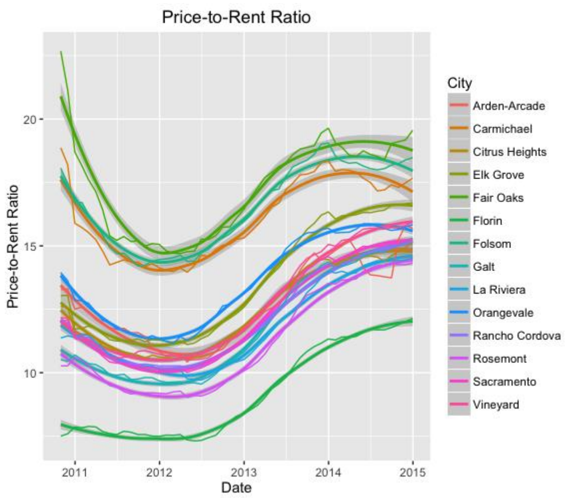
II. Price to Rent Ratio

Price-to-Rent Ratio is another important factor. We can see that cities with higher income are less filled with rental properties. The Bay Area has always had a higher price-to-rent ratio than the average city.









What does Price-to-Rent Ratio say about. Price-to-Rent Ratio is a great measure of living expenses. A price-to-Rent Ratio of a property is just the value of the property divided by the annual rent that could be gained by that property. For example, the Price-to-Rent ratio states that

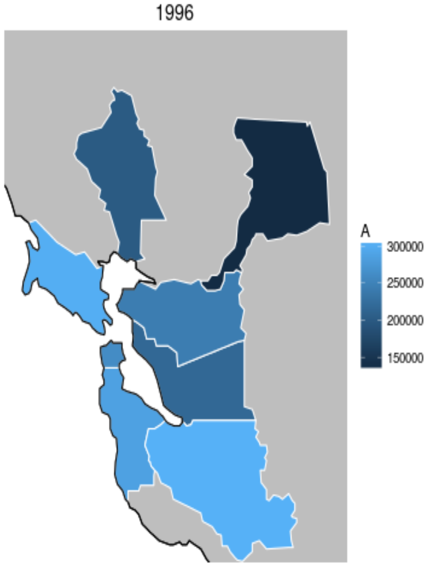
3.2 Compare and analysis the trend of data changes during past two decades

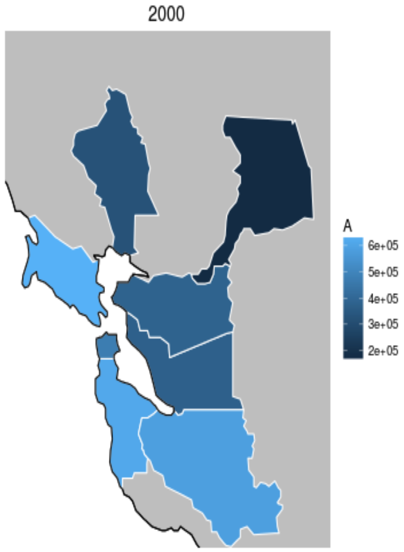
Below are plots depicting the average price of ALL homes in each different county, selected by specific years to show the historical economic events and their effect on the housing market From early 1990s from 1990 to 1995, housing prices did not change a lot. They increased excitingly during the dotcom boom before regressing immensely during the Great Recession However, we can see that for the latest Year in our dataset,2014, we can see that housing prices have recovered and surpassed pre-Great Recession levels.

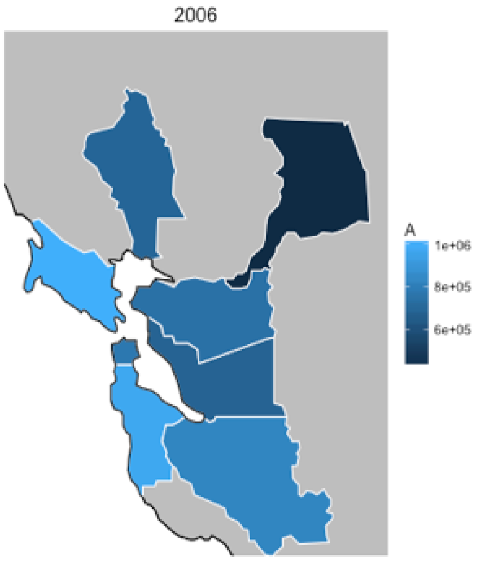
When analyzing the present Bay Area housing market, one of the more intuitive questions people might come up is: Which counties are the most expensive to live in? Before using real data to illustrate, an individual might suspect that San Francisco and Santa Clara would have higher average incomes and property values than counties in the North and East Bay, given Tech’s boom and concentration around Silicon Valley

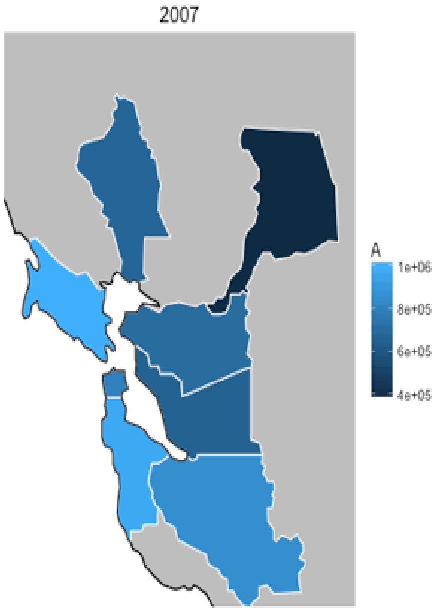
Housing data of each county on specific year was plotted as a map to visualize the price distribution of all homes between different counties. The result mainly confirmed the previous assumption and the fact so far is: Santa Clara is the first place followed by San Francisco, San Mateo and Marin. Like any other economic problem, we can boil this one down to supply and demand. Silicon Valley has always been the Mecca of tech, but

With all of this recent discontent from at tech giants moving to these cities, most of the attention has been placed on the demand side of the problem: people are upset that the presence of too many software engineers has been inflating their rents. But there's also a supply side to the issue that needs to be considered: Maybe there isn't enough housing in the city to go around. If people want to live in this area, and they don't want price to go up, then there needs to be more housing for constructed. However, the area might be limited; San Francisco is located on a peninsula, there's pretty much only one way for the city to add new housing units: by growing vertically but supply is still not meeting demand, which leads to higher housing prices.









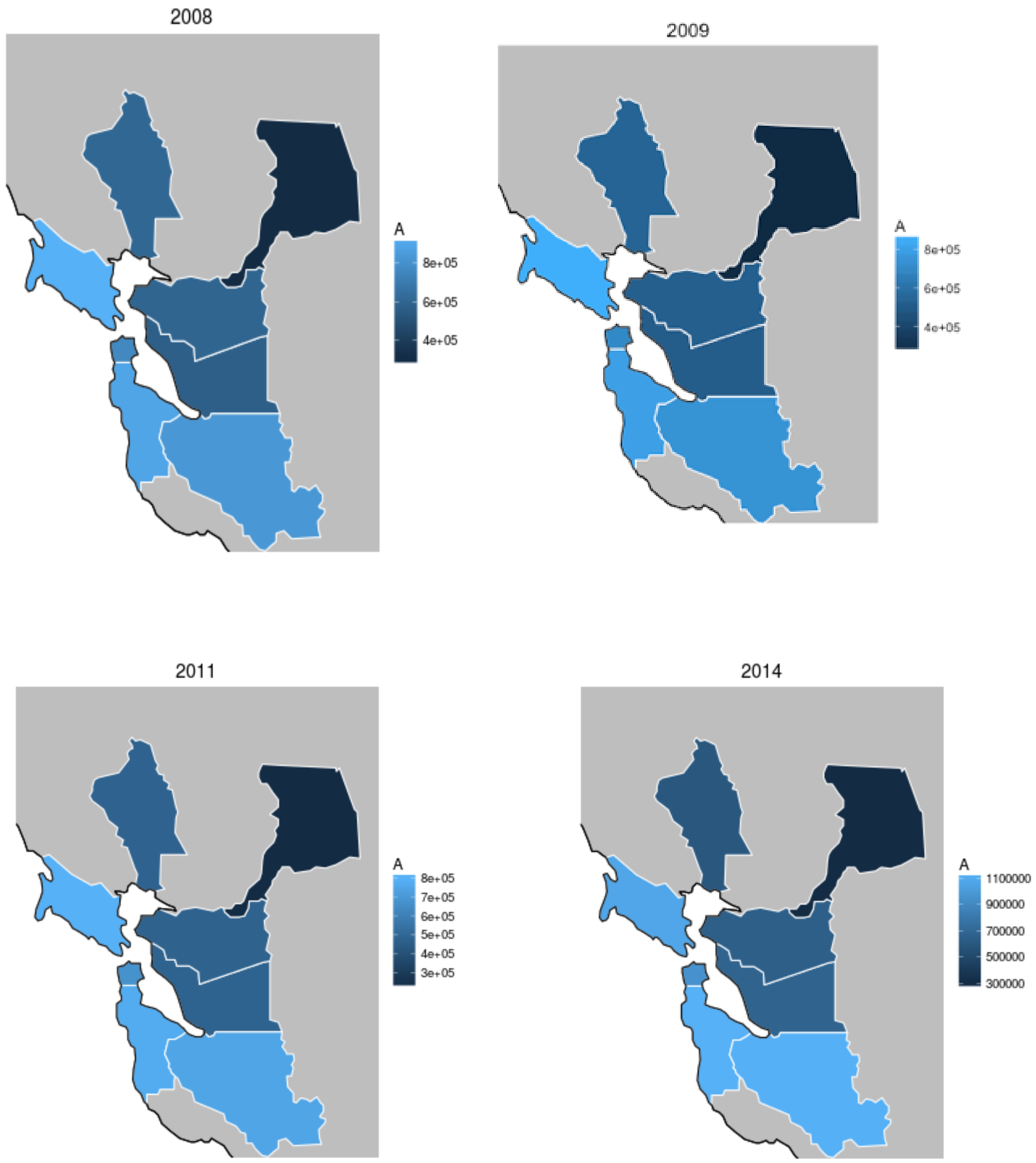
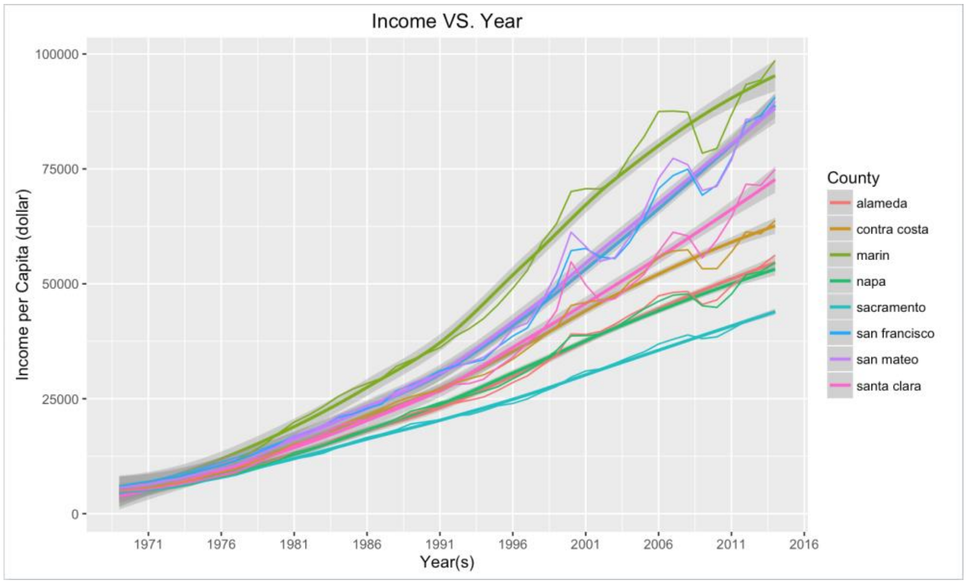


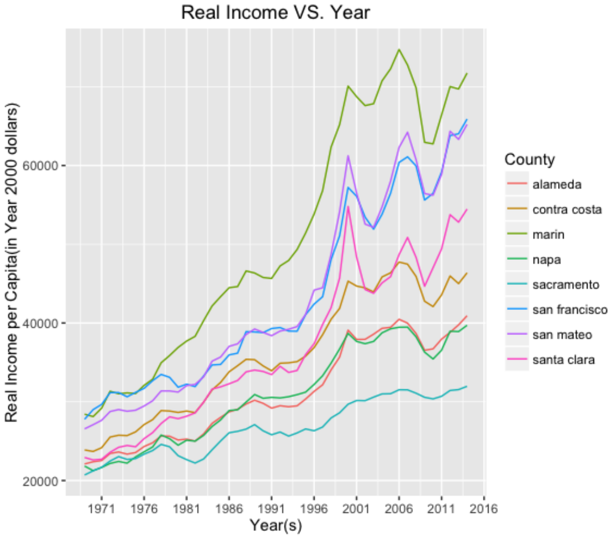
Figure-1: illustrating the housing price distribution among Bay Area through brightness difference. The highest price region started at Santa Clara and switch to Marin later due to the financial crisis, and then went back to Santa Clara after more technical industry development in the whoel Bat Area.

As shown in the figure 1, Santa Clara and Marin counties occupied the major part of the housing market in 1990's. As time goes by, housing prices in Marin County have maintained their stand at high position in housing market and kept stable, and there is little significant change, since Marin is a traditional rich suburbs. However, Santa Clara experienced a great change among the past two decades. It dropped from the top position during the financial crisis and then came back because of the tech boom.

3.2 Special Case Analysis



As we can see there has been a sizeable upper trend across all counties in the Bay area since the 1970s, but this picture does not take into account other Quality of Life factors, such as Inflation and Cost of Living expenses.

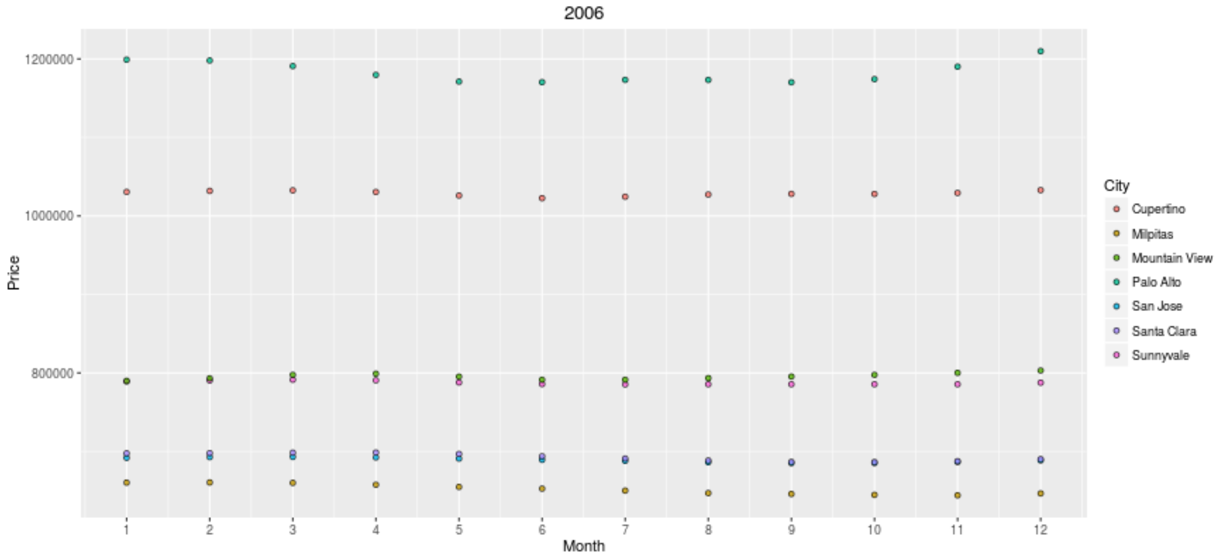


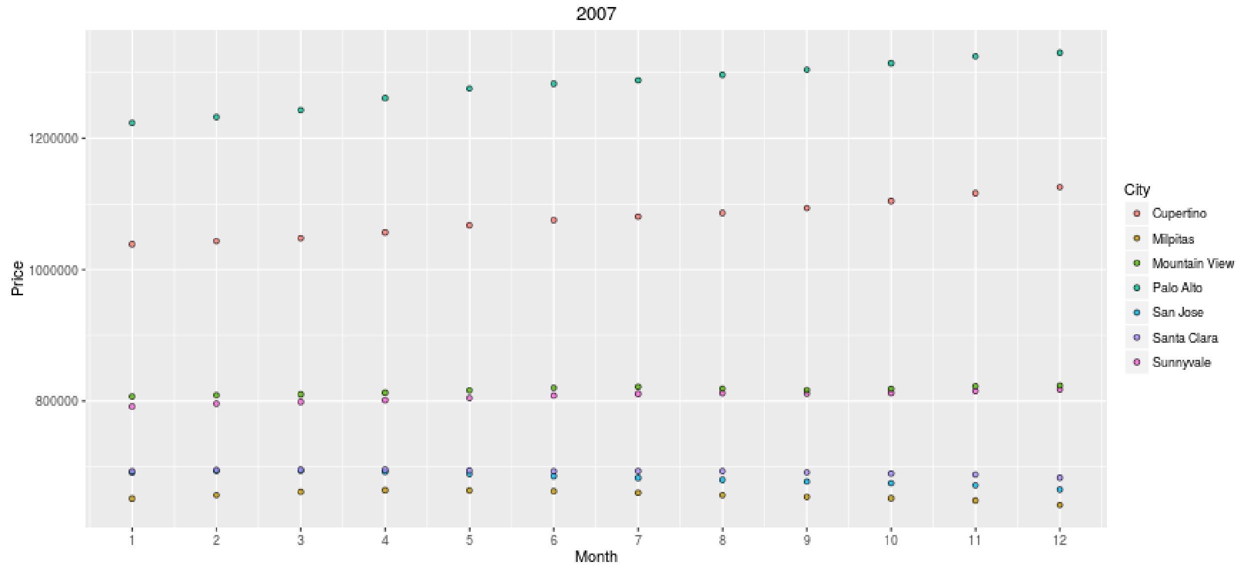
However, Real Income/Buying power in 2014 hasn't recovered from 2000 values even though nominal income grew ~20% on from 2014 values.

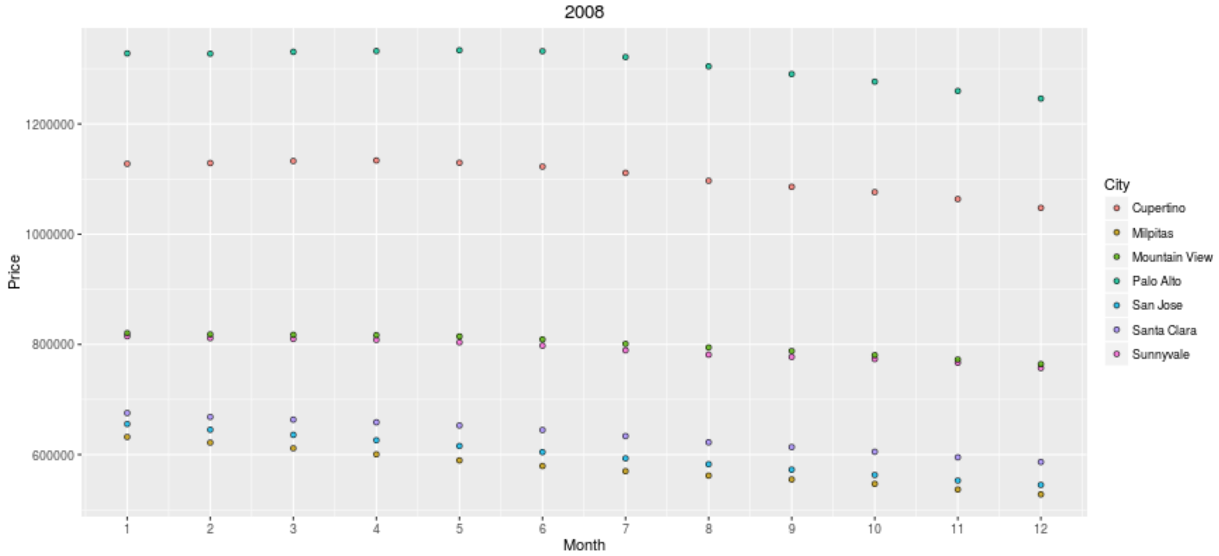
We can link this back to housing. One of the reasons that buying power in the Bay Area hasn’t grown in 14 years may be because the increase in wages may have been used to pay increases in Rent. One of the major groups that comprise part of the CPI index is Housing costs(rent of primary residence, owners' equivalent rent).

Theoretically, housing price should go up along with the time line. However, from approximately 2008 through the early 2009, the U.S. went into fairly severe recession that has taken a toll on many aspects of American Life. From fallings stocks to rising unemployment, most all American families have experienced some ill effects from the economic downturn over the last couple of years. In addition to the many different economic metrics that indicate a slowdown in the economy, this particular recession has been characterized by falling real estate values as well. In fact, the National Association of Realtors reports that home values nationwide have slipped almost 20% nationwide in only 3 years.

Another interesting thing to note is that Average Buying Power in the Bay Area for most counties has only recently recovered to Dot Com boom levels.




2007
Price
1200000
1000000
800000
Month
1 2 3 4 5 6 7 8 9 10 11 12
City
Cupertino
Milpitas
Mountain View
Palo Alto
San Jose
Santa Clara
Sunnyvale


2008
Price
1200000
1000000
800000
600000
Month
1 2 3 4 5 6 7 8 9 10 11 12
City
Cupertino
Milpitas
Mountain View
Palo Alto
San Jose
Santa Clara
Sunnyvale

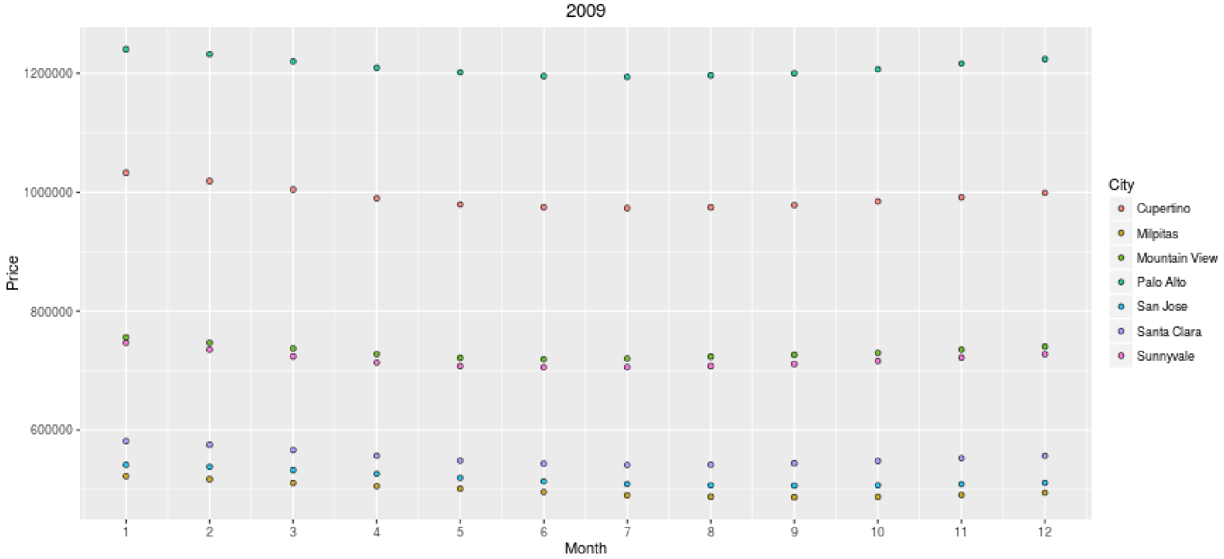

2009
Price
1200000
1000000
800000
600000
Month
1 2 3 4 5 6 7 8 9 10 11 12
City
Cupertino
Milpitas
Mountain View
Palo Alto
San Jose
Santa Clara
Sunnyvale

Figure-2. Price change trend of Santa Clara County by year, a typical case to show historical event, such as financial crisis, could have had a significant effect on housing market.

As discussed above, population and income are two of the major factors that would affect the housing price. However, in 2008, affected by financial crisis, housing prices plummeted even though population grew. Although economic recessions are particularly painful to those employed or a mortgage they cannot pay for, they do afford us an opportunity to research changes in behavior in the midst of an unhealthy economy or real estate market. For example, in a good economy when house prices are consistently rising, most consumers are comfortable paying what is perceived to be full market value for a given property because the inherent assumption is that values will continue to rise. In fact, in the early to mid-2000's many consumers took advantage of unconventional mortgage-backed securities because the underlying assumption was that real estate was secure and past trends showed their value rising higher. As we know now, our national real estate market was indeed more susceptible to decline than anybody would have guessed. However, now that we find ourselves in a declining real estate market, there is an opportunity to understand what inherent characteristics of real estate continue to elevate certain properties to the top of the real estate market and what characteristics serve to depreciate others.

4. Conclusion

Through the use of statistical analysis and computing data, we were able to isolate major factors that mirrored trends in the housing market. In addition, we were able to see the differences in these trends among different counties. Our map_app helped us to identify which characteristics of housing were the most significant at each point in time as well as showing the trend among the two decades.

Limitations and Further Questions:

Professor Do suggested that we separate the income data into quintiles, in order to see trends amongst all income brackets. Unfortunately, we could not separate the income dataset accurately into quintiles as the dataset consisted of the average income of cities, not individual income within this city. A further search into appropriate individual income data using this data could've confirm the Professor's hypothesis.

Our datasets are only 2014 recent, many rates may have grown this threshold during 2014 to 2016.

Besides, the results of life expectancy and income analysis revealed that there were some similarities and some difference in regards to which characteristics of housing were significant in determining the house value. Although it is clear that housing market is complicated, which are affected by various factors and we are also very interested in exploring the accurate inner relation to see how housing market could change, we were unable to reach that due to lack of relevant data. This could potentially lead to further research and data analysis. It remains to be seen whether the income growth was localized within Silicon Valley. A further study could be given comparing other tech hubs like Seattle, Portland, New York City, and Austin to see if the growth was matched in these other cities.