

# *Bay Area Housing Data Analysis*

---

Housing Hacker (Group 5):  
Xinhao Li, Kentaro Ino, Tina Guo, Kevin Khuu

# Agenda

- *Introduction*
- *Data Collection and Wrangling Process*
- *Analysis and Data Visualization*
- *Conclusion*

# Introduction

*Housing markets are significant in shaping the economic and social well-being of many Americans.*

- ❑ Housing expenditures comprise a major large component of every household's budget, so the availability of housing and its price assume considerable importance.
- ❑ Housing markets play a central role in the process of metropolitan development, both affecting and reflecting other forces at work in this sphere.



# Introduction

## *Our Research:*

- ❑ Tracking conditions in the specific San Francisco Bay and Sacramento housing markets in the context of historic trends and expectations for future.
- ❑ Identify changes in the housing market in terms of significant characteristics of housing and their corresponding effect on home values.
- ❑ Interpret the extent of how pivotal economic events, can affect the housing market.

# Data Collection and Wrangling Process

## *Raw Data Collection*

- ❑ We acquired our data using Quandl, a search engine for numerical data offering access to several million financial, economic and social datasets to scrap raw housing data.
- ❑ Constructed a lookup code script only selecting the information about Bay Area, Using a for loop searching and avoid redundant code.



# Data Collection and Wrangling Process

## *Raw Data Collection*

*Below is a sample of raw data obtained from Quandl (Type A stands for average price for all types and value stands for the numeric value of that type)*

	Date	Value	City	County	Metro	Type
1	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



## Data Collection and Wrangling Process

### *Aggregation Process*

- ❑ Spread housing types to present more details.
- ❑ Gather data by type and year.
- ❑ Inner joined by year and county to obtain a general dataset containing all the housing data of Bay Area.



# Data Collection and Wrangling Process

## *Aggregation Process*

*Below is a sample cleaned data (twoB stands for two-bedroom, threeB stands for three-bedroom)*

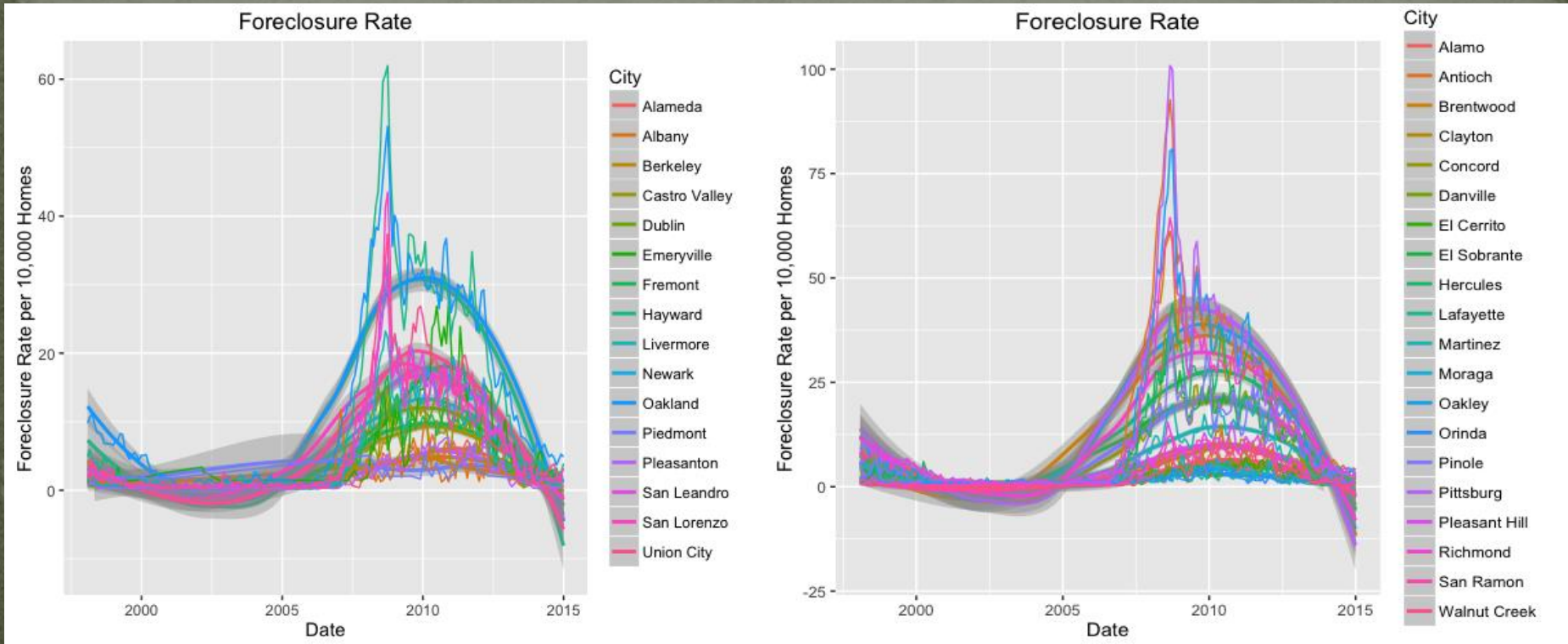
region	subregion	year	Pop	Income	twoB	threeB
california	alameda	1996	1359.099	28535	152197.777	226967.361
california	alameda	1997	1380.383	29971	171414.444	239140.104
california	alameda	1998	1405.903	32234	189932.777	267451.562
california	alameda	1999	1427.114	34513	211057.777	298254.687
california	alameda	2000	1450.086	39093	277123.333	380213.541
california	alameda	2001	1468.652	38991	318203.888	426436.979
california	alameda	2002	1460.438	39619	338772.777	449368.75
california	alameda	2003	1451.418	41226	374987.222	493705.208
california	alameda	2004	1441.496	43140	438918.333	570041.145



# ■ Analysis and Visualization

*Identify Major factors affecting housing price*

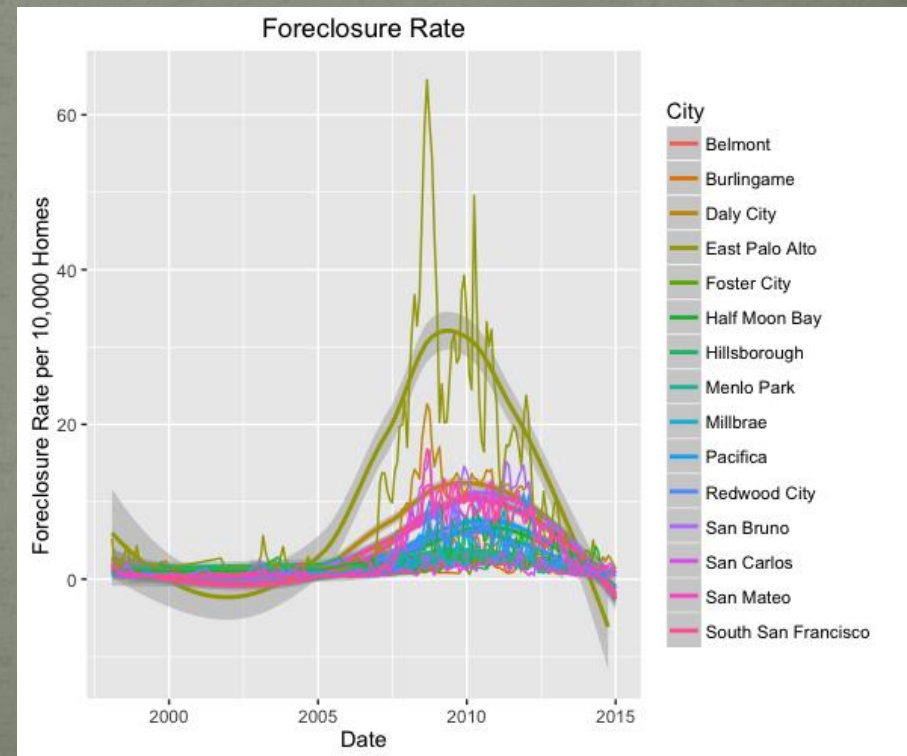
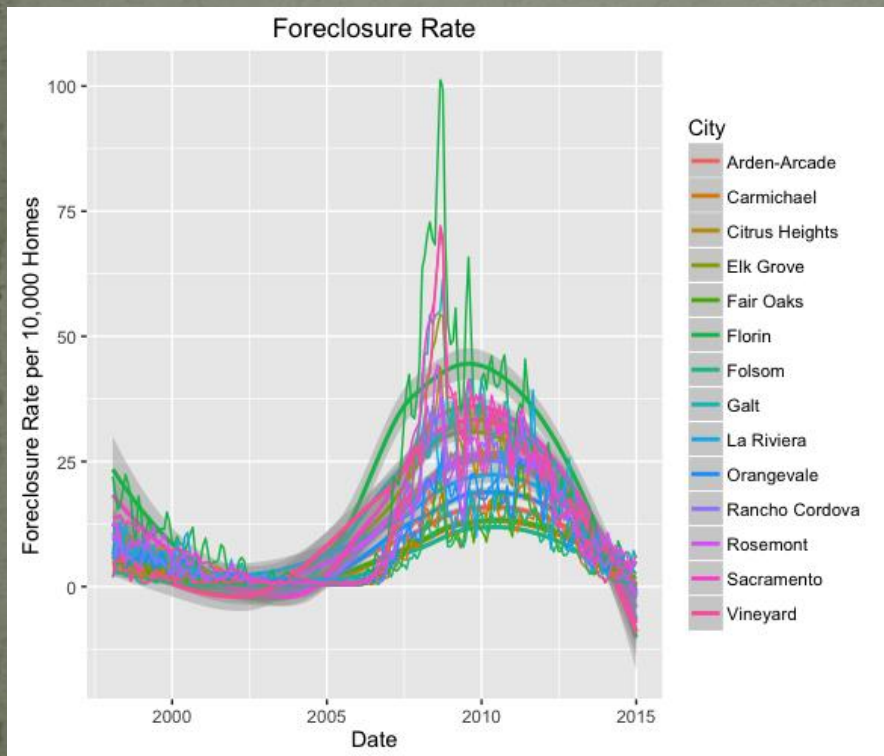
## ■ Home Foreclosure Rates



# Analysis and Visualization

*Identify Major factors affecting housing price*

## Home Foreclosure Rates



# Analysis and Visualization

*Identify Major factors affecting housing price*

## Home Foreclosure Rates

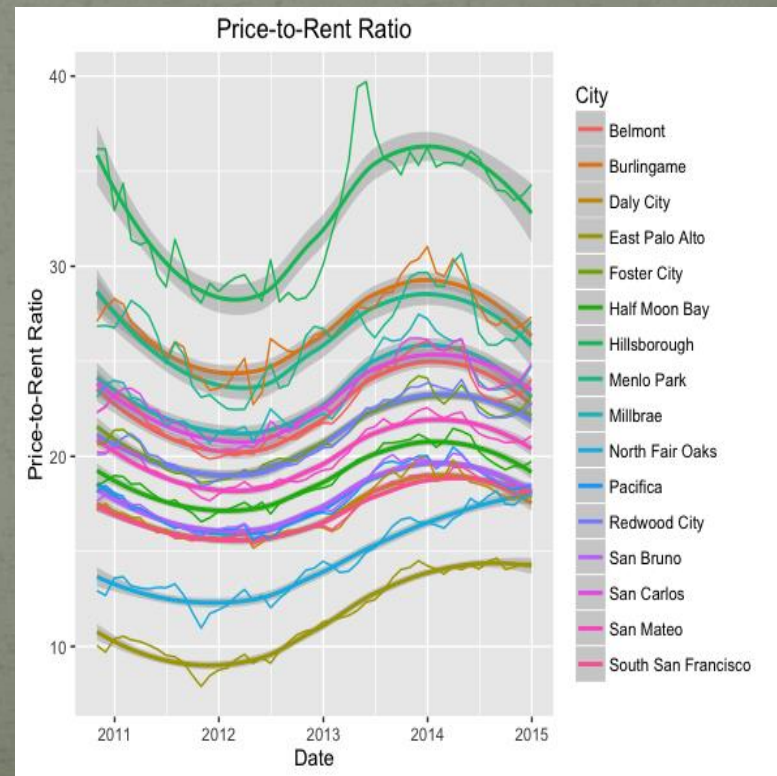
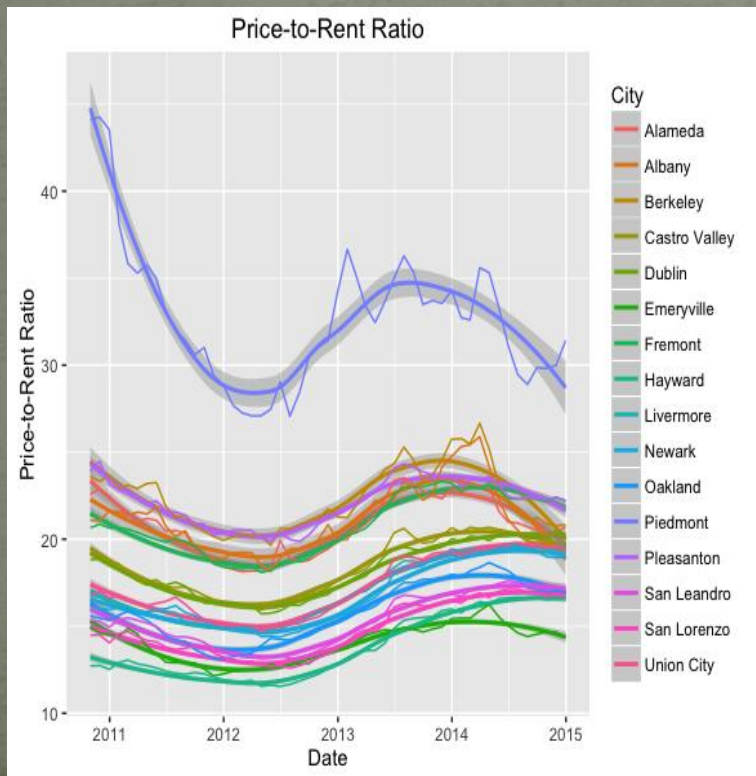
- Home foreclosure rates spiked immensely during the 2008 Recession.
- Same general trend for all of the counties. However, this trend was less pronounced in Silicon Valley compared to the rest of the Bay Area.



# ■ Analysis and Visualization

*Identify Major factors affecting housing price*

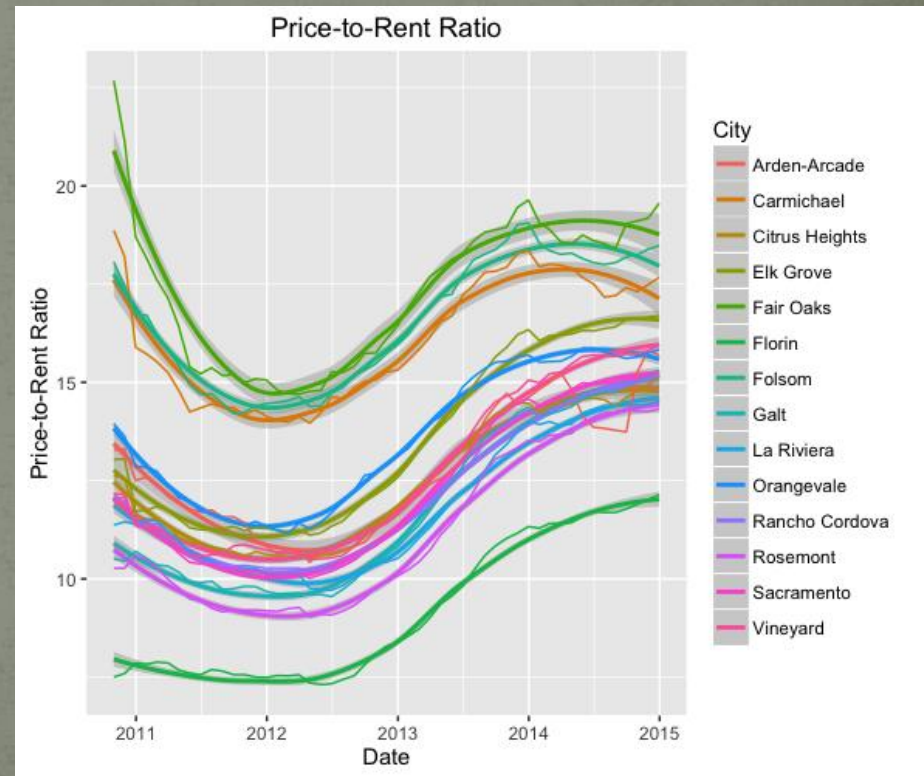
## □ Price to Rent Ratio





# Analysis and Visualization

*Identify Major factors affecting housing price*



# Significance of Price-to-Rent Ratio

- It is an indicator of whether it is better to buy or rent in a specific area. According to the New York Times, 15 is the threshold.
- Many of the Bay Area counties have significantly higher than average PRR ratios. However, we can see that these higher than average PRR ratios tend to be focused in higher income counties. These PRR ratios have also reached extreme highs compared to the 15 year average.



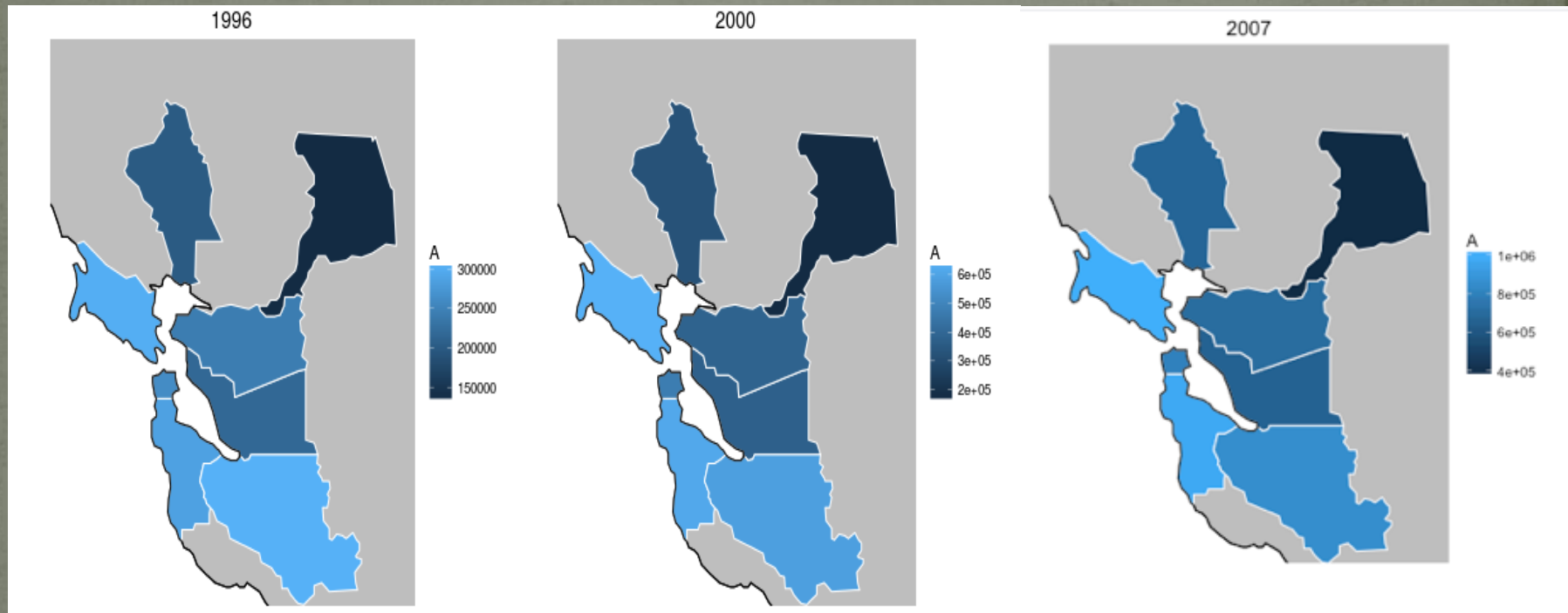
# Analysis and Visualization

*Compare and Analysis trend of data changes*

- ❑ During 1980s to 1990s, housing price grew at a stable rate. After that, it experienced a dramatically growth. Overall, the housing price of Bay Area increases.
- ❑ San Francisco and Santa Clara would have higher price than other counties. Average housing prices are higher in the South and North Bay compared to the East Bay.

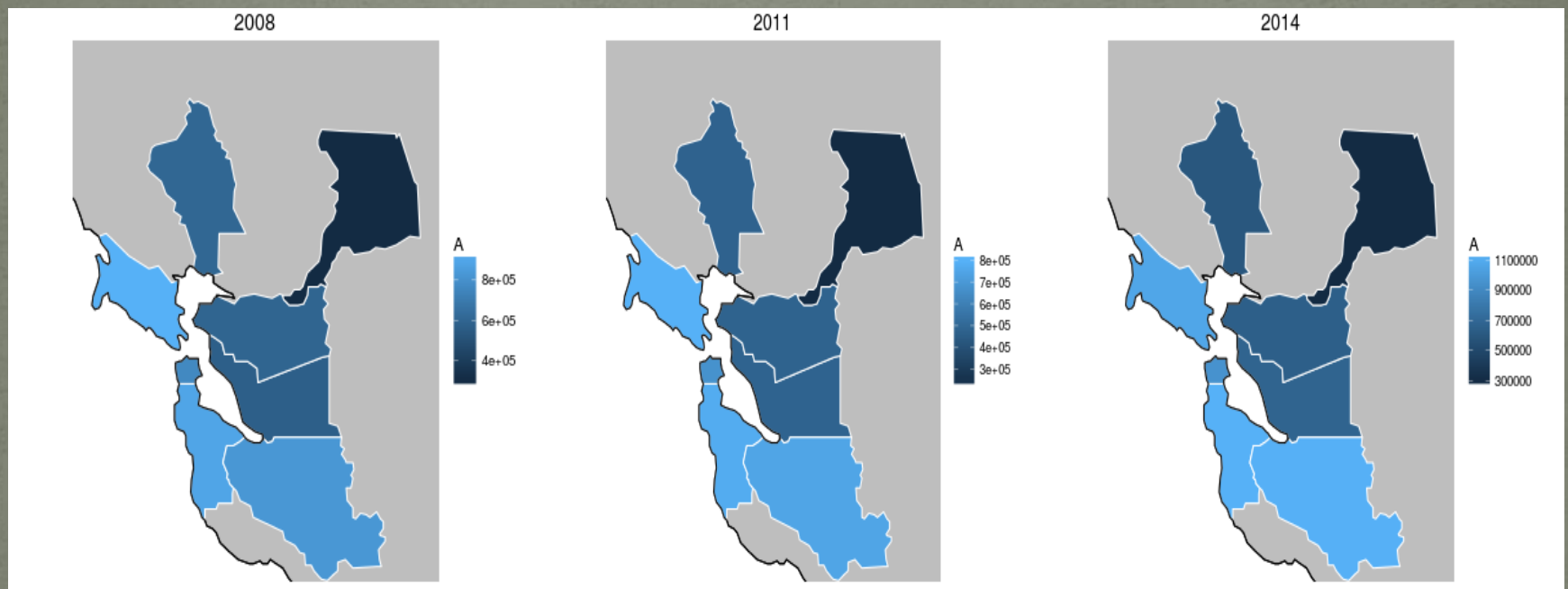
# ■ Analysis and Visualization

*Compare and Analysis trend of data changes*



# ■ Analysis and Visualization

*Compare and Analysis trend of data changes*





# Analysis and Visualization

## *Compare and Analysis trend of data changes*

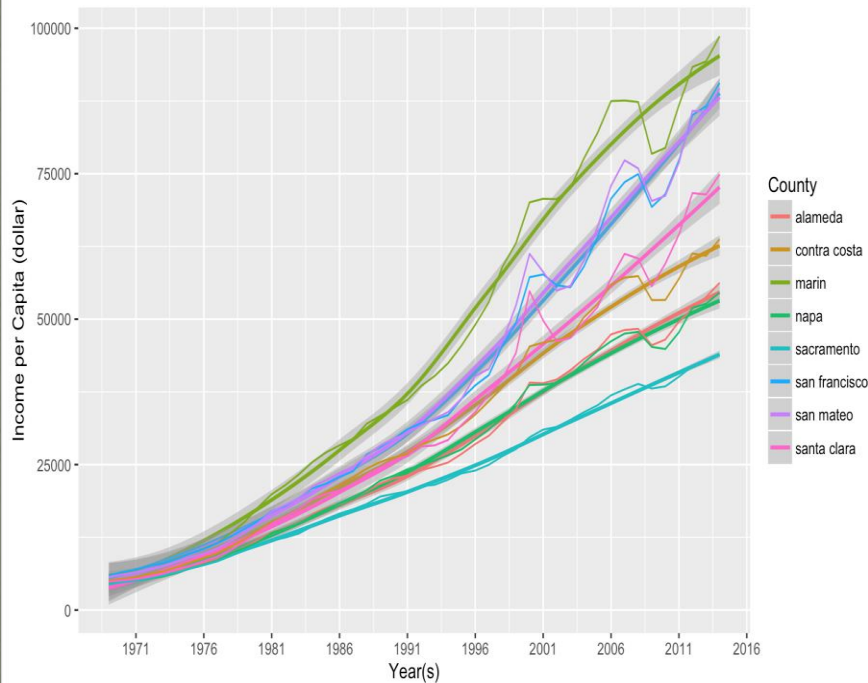
- ❑ Santa Clara and Marin counties occupied the major part of the housing market in 1990's. As time goes by, housing price in Marin County is still stand at high position in housing market and kept stable, and there is little significant change, since Marin is a traditional rich suburbs.
- ❑ Santa Clara experienced a great variation among the past two decades. It dropped from highest average after the dot com boom and remained behind Marin county until 2014 when it retained its former throne.



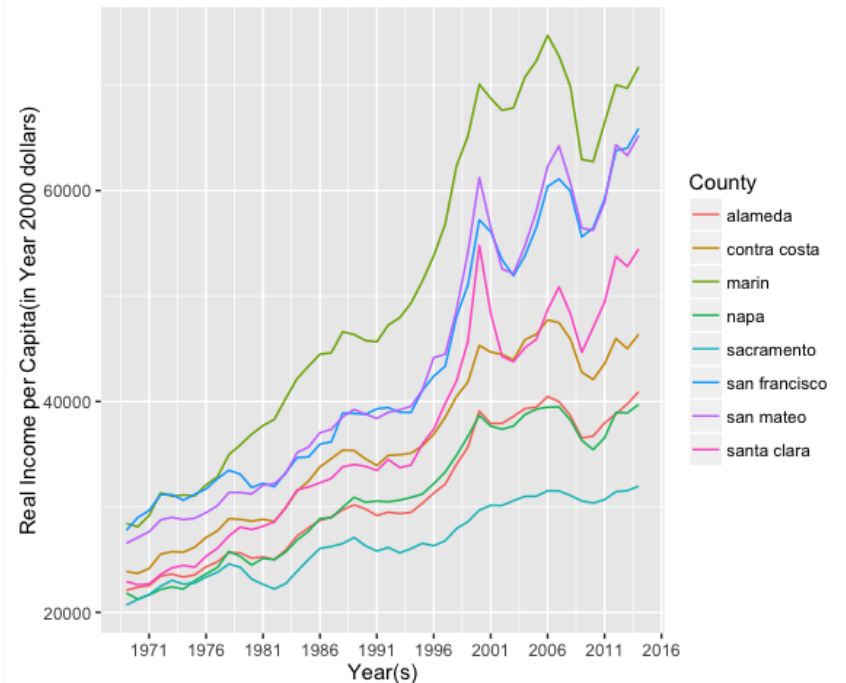
# Analysis and Visualization

## *Special Case Analysis*

Income VS. Year



Real Income VS. Year



# Analysis and Visualization

## *Special Case Analysis*

- ❑ Upward trend in income across all countries even after burst in DotCom bubble and 2008 Great Recession
- ❑ However, after adjusting for inflation against Year 2000 dollars(using Consumer Price Index) we can see that Real Income for most countries have only recovered to Year 2000 levels.



# Analysis and Visualization

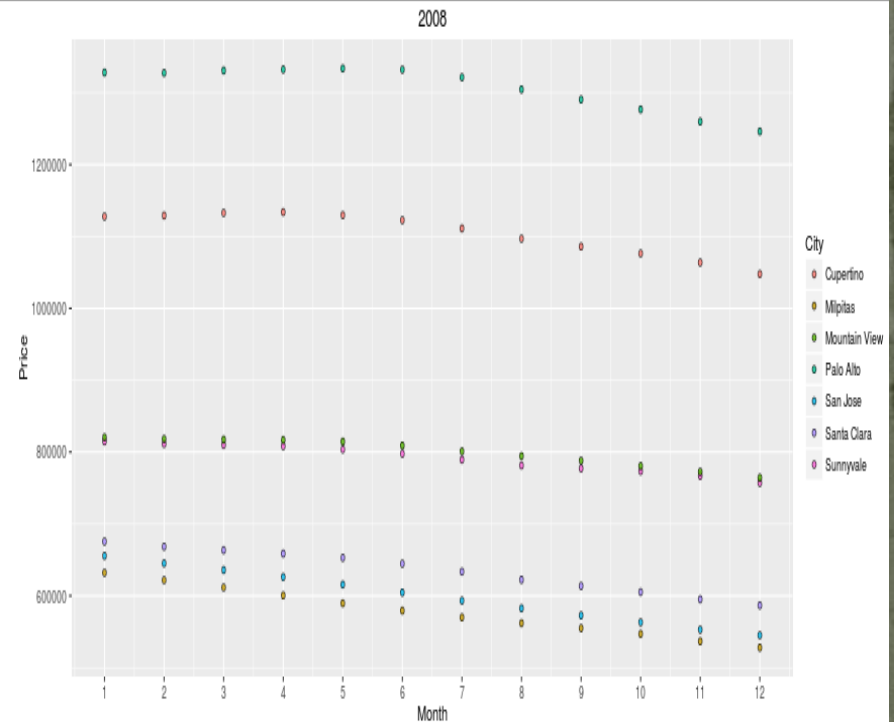
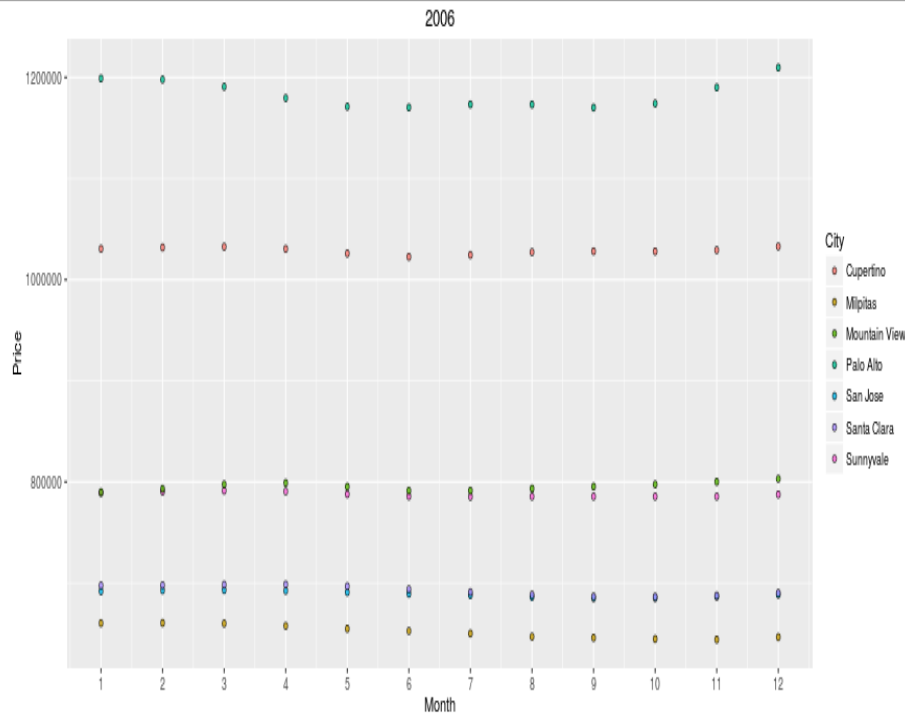
## *Special Case Analysis*

- ❑ What could be some of the reasons for this?
- ❑ The increase in income may have been met with an increase in housing and Cost of Living, which explains the stagnancy of buying power.



# Analysis and Visualization

## *Special Case Analysis*



# Analysis and Visualization

## *Special Case Analysis*

We can see that average prices for some of these properties have recovered and surpassed pre-Great Recession levels.



## Conclusion

- ❑ We can see that the housing market has recovered to post recession levels.
- ❑ Housing market is complicated 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.

## Limitations and Further questions.

- ❑ There were some intriguing questions that we couldn't answer with our data. For example, our FRED data was based on averages seeing trends between different income quintiles.
- ❑ Our analysis is only restricted to the Bay Area. Some further analysis could be focused on other emerging tech hubs like Seattle, Portland, New York City and Austin and a comparative analysis could be performed to see if some of these trends of the Bay area were replicated on any scale.

# Questions?\*\*\*



\*\*No guarantee of answer~