# Average Home Purchasing Price Analysis

*Lianghui Li, River He, Zhongshu Wang, Boyue Sun*

*December 6, 2015*

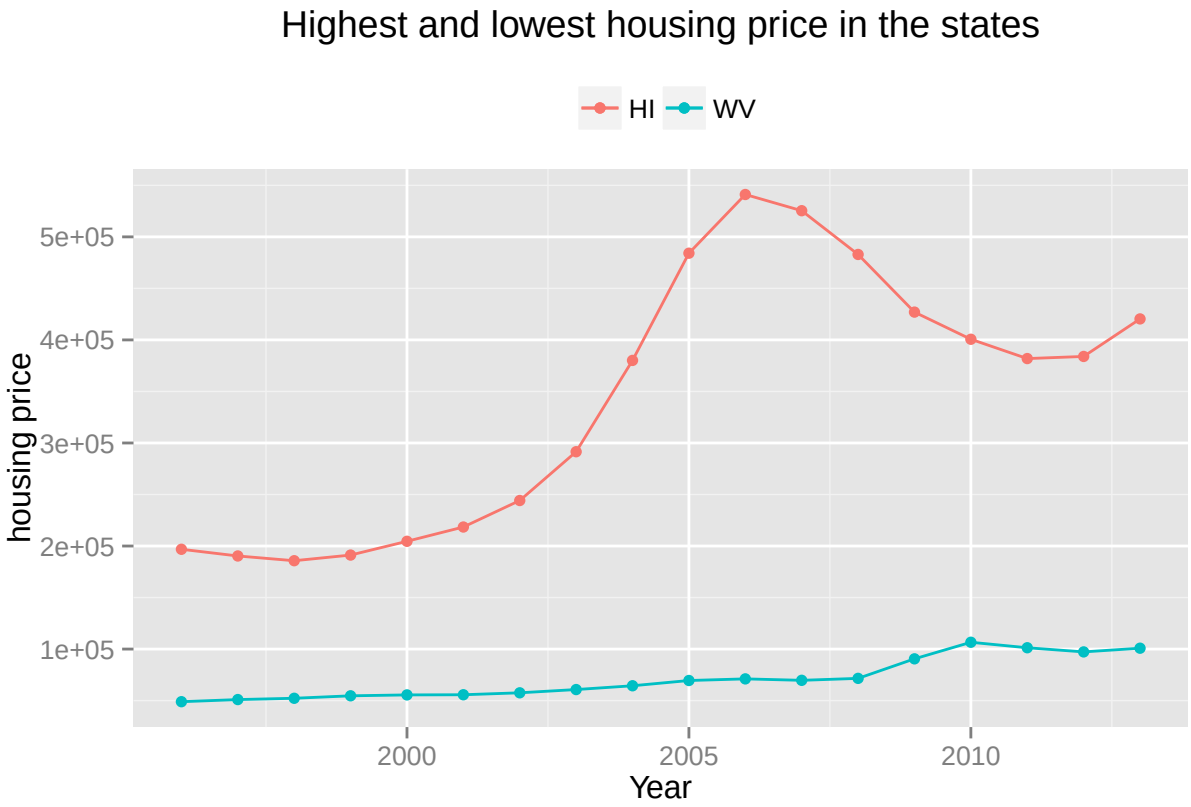
## DATA ANALYSIS

### Housing Affordbility(Lianghui Li)

Colletcting interest rate(per month) from"data360.org"
income distribution(per year) from "census.gov"
and housing price(per month) from "zillow"

#### Time Trend

compare the price trend of minimum average price *WV* and maximum average price *HI*




---

#### General Question

#### How much income to apply for mortgage?

qualify income to apply mortgage in the year 2003, 2013 in state *AK* & *CA*

```
qualify_income('AK', 2003)
```

```
## [1] 40769.97
```

```
qualify_income('CA', 2013)
```

```
## [1] 56190.73
```

**Your income is high enough?**

housing index for the income value 64930 in year 2008, 2010 in state $AL$ & $IL$

```
housing_index(64930, 'AL', 2009)
```

```
## [1] 2.413483
```

```
housing_index(64930, 'DC', 2010)
```

```
## [1] 0.9240816
```

```
#notice that if the year is not between 1996 and 2013, no result will be shown
qualify_income('NJ', 1993)
```

```
## [1] "No sufficient data support"
```

**Is majority of American can afford?**

overall American affordbility for the housing price in $OK$ & $CT$ in 1999 and 2001

```
housing_affordability('OK', 1999)
```

```
## [1] "About 40% American can afford"
```

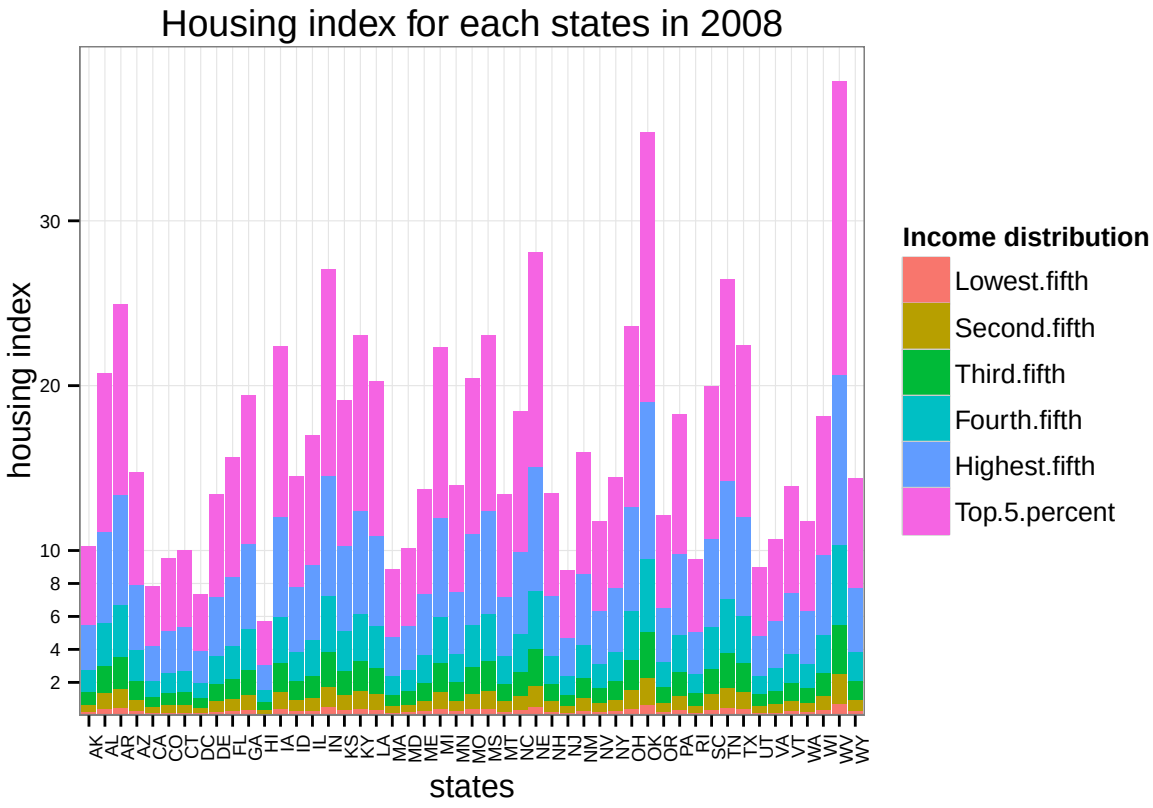
```
housing_affordability('CT', 2008)
```

```
## [1] "Less than 20% American can afford"
```

---

## General Affordbility

Bar plot that show affordbility for each income distribution in the year 2008 for each state




From the graph, we can see that expensive housing states for general American are HI,DC, CA; and cheapest are WV, OK and NE

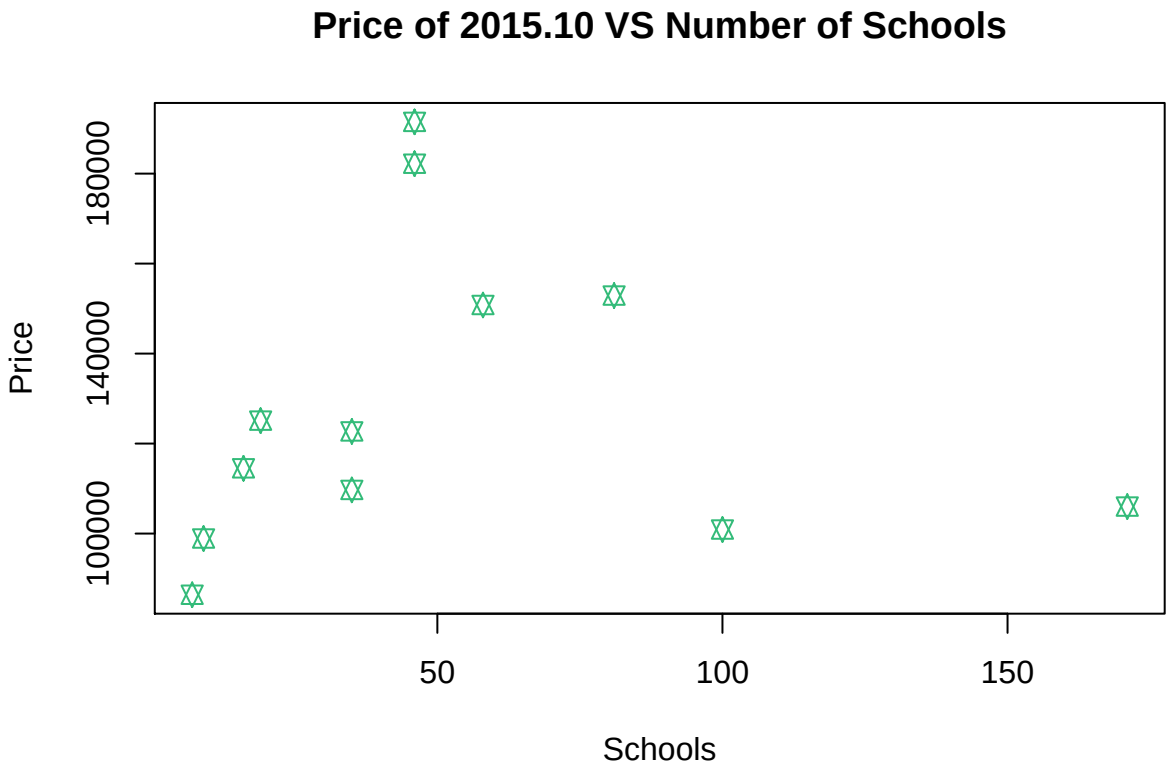
*Limitation* : Since our formula just include two elements(interest rate and income); however, general concept of affordabilty is not limited, but also contains factor such as **number of school**, **transportation cost**, **local tax policy**, etc. These omitted variable could bias our calculating the affordability index.

---

## Realtionship between Housing Price and Education(River He & Boyue)

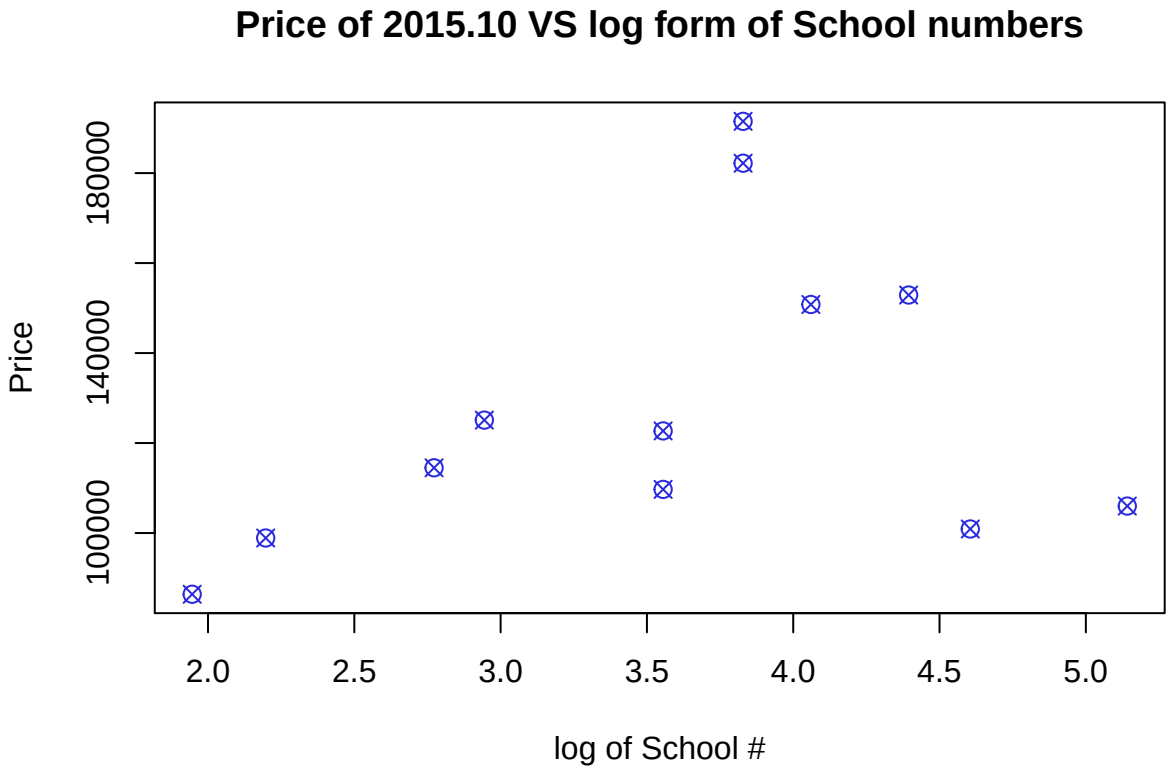
*collecting data from "ftp.cde.ca.gov" and "web.alsde.edu" However, when we first downloaded the data from "web.alsde.edu",we could use it, but somehow the website link has some problems when we use downloadfile command.So we just put AlabamaSchoolListing.csv in our raw data folder*

**relationship between housing price and number of schools in Alabama's counties**




```
## [1] 0.01959473
```

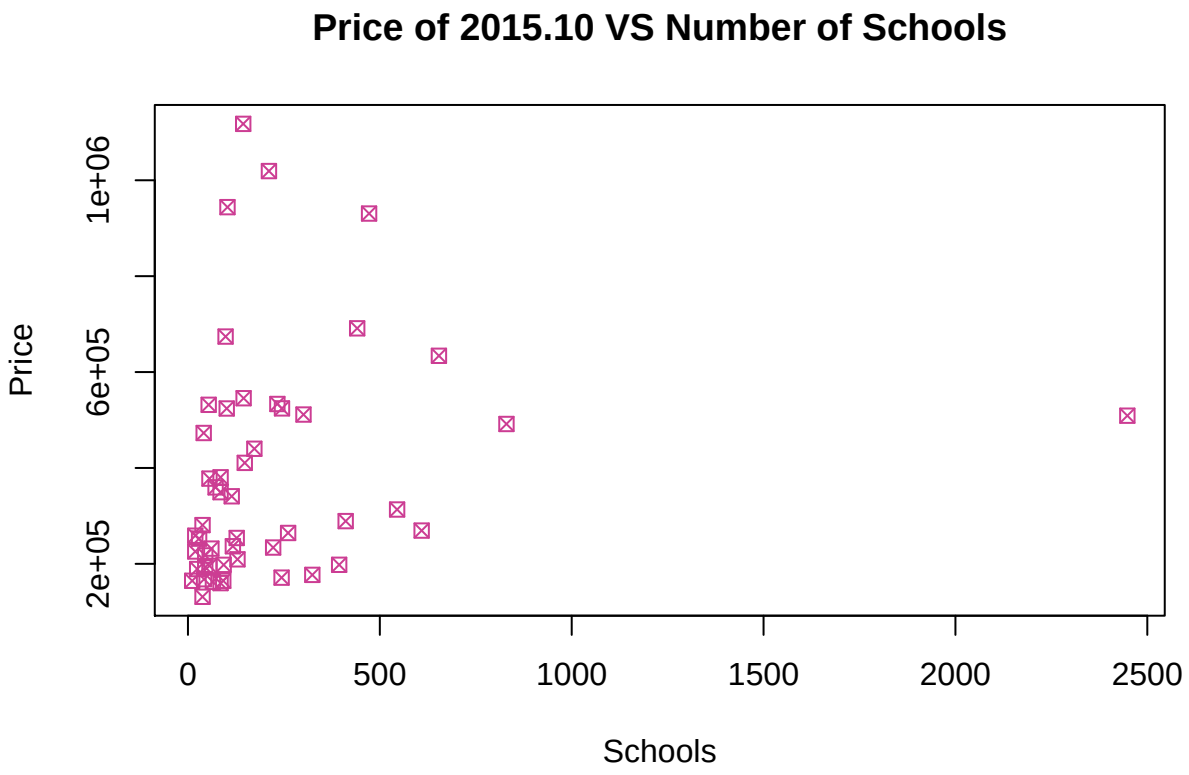
There is a positive relationship between house price and school numbers in Alabama but the correlation is only 0.019 which is very small. We can ignore it. Also from this graph, we found the relationship is not linear,so we decided to use linear-log model.




```
## [1] 0.3374508
```

There is a positive relationship between house price and school numbers in Alabama and the correlation is only 0.3374 which is higher.

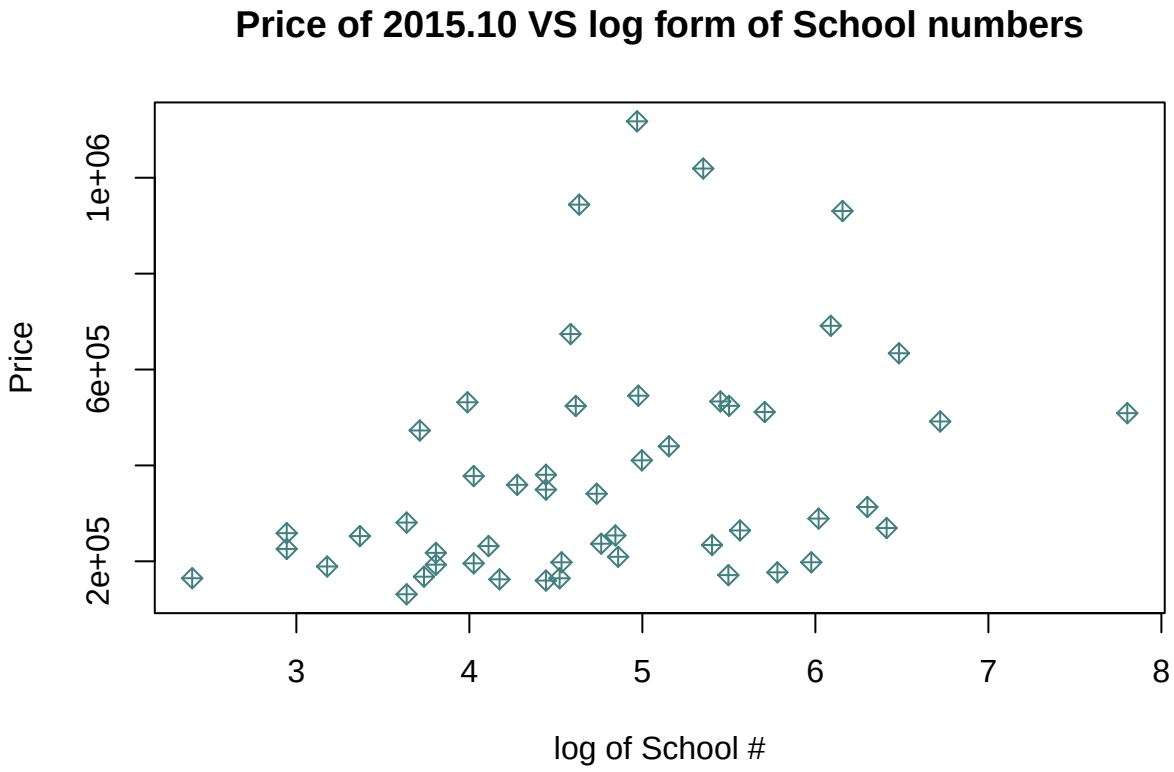
**relationship between housing price and number of schools in California's counties**




```
cor(Ca[,242],Ca[,243])
```

```
## [1] 0.1963823
```

There is a positive relationship between housing price and number of schools in California's counties. However, the correlation is only 0.19, it is really small.




```
cor(Ca[,242],log(Ca[,243]))
```

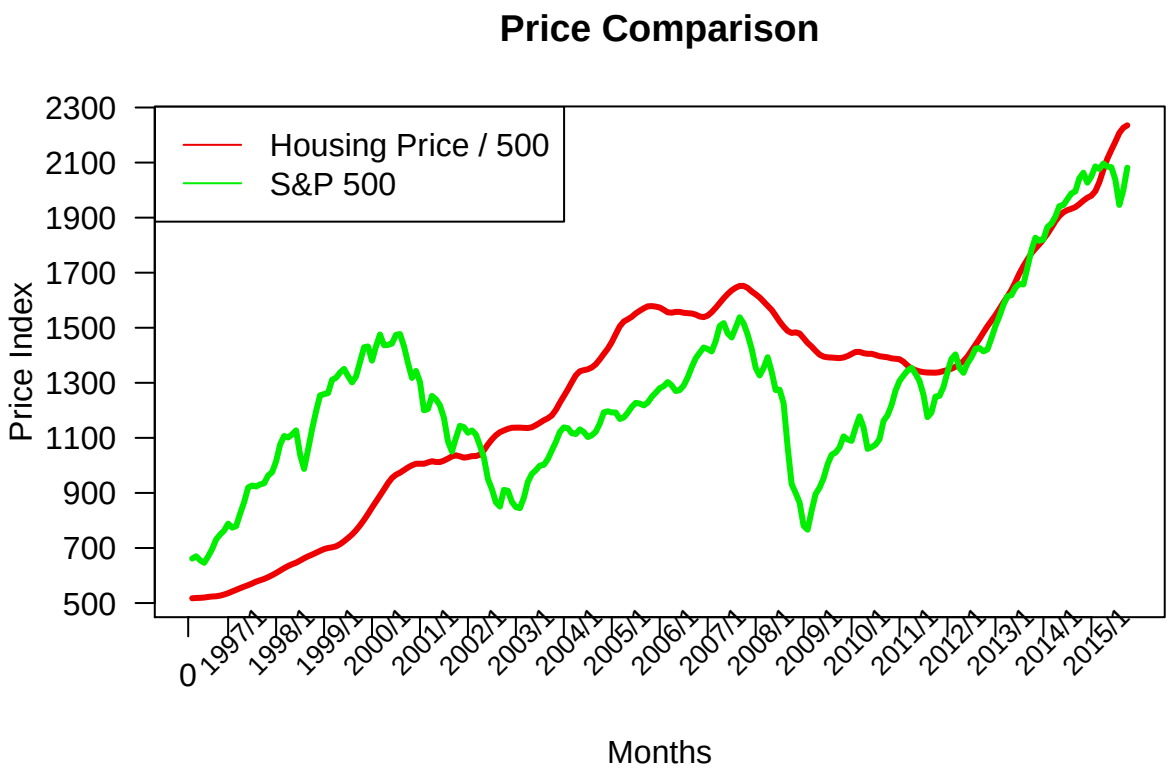
```
## [1] 0.3560753
```

After doing the linear model, we found that the relationship between housing price and number of schools is not linear, so we decided to use linear-log model to construct the graph. There is a positive relationship between housing price and number of schools in California's counties. The correlation in linear-log model is 0.36 which is larger than the correlation we got before.

***Limitation*** **: It is hard to measure the qualities of schools, which may have impact on the housing price. Also we only choose two states in this part, the results may be different from other states.**

### Housing Price and Stock Market(Zhongshu Wang)

collect average housing price from zillow research database s&p500 index from yahoo finance with preselecting month information Take San Francisco as an example, comparing housing price and s&p500. Here, we divided the average housing price by 500 to be more compariable

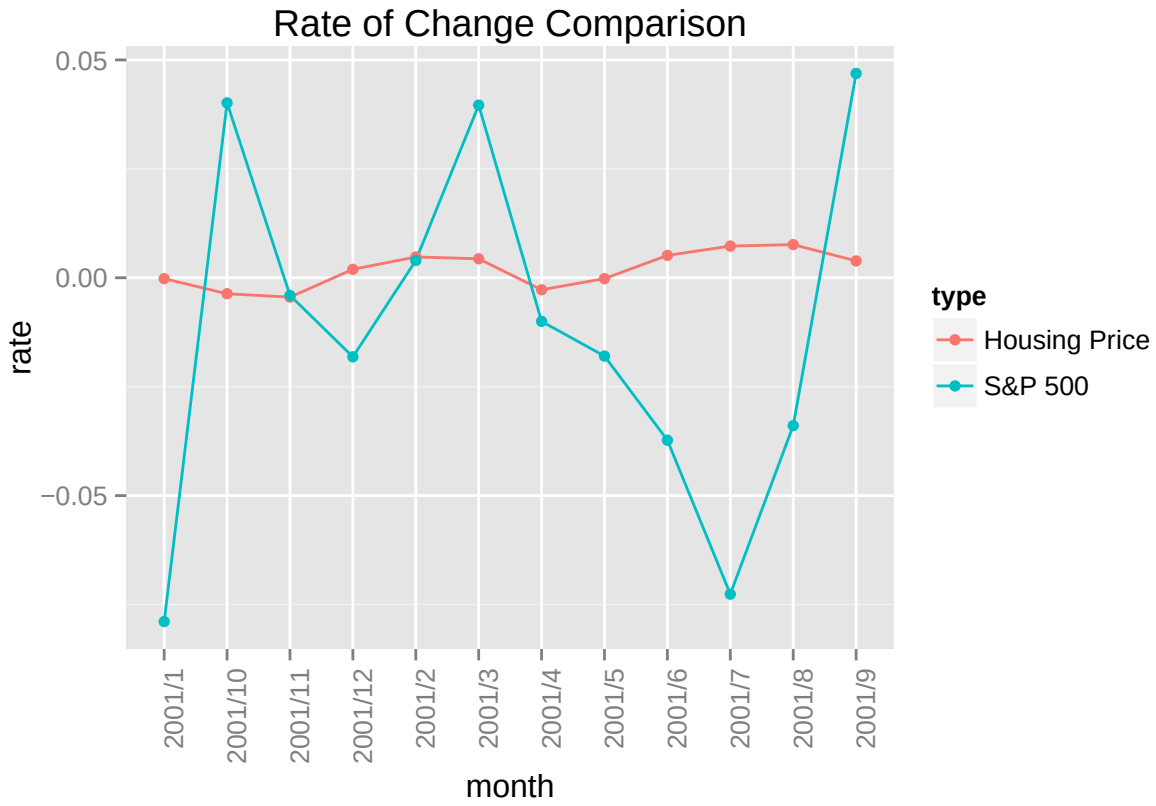



However, this graph can't be develeped into a strong argument

Real Estate and Stock Markets are two investment viechles, which they are subsitutes. Also, Real Estate might affect the stock market as a factor.

Thus, the Dot Com bubble bust in 2001 caused the stock market downtrend, investors lost confidents and looking for other investment viechles. Hence, we conduct an observation of the relationship between housing purchase price and stock price on their rate of change.

Here is the graph of their rate of change during 2001




Graphs shows that rate of change on housing price moved in the opposite direction as the rate of change on s&p 500 Moreover, here is the correlation test

```
cor(sf_rate2001$house_rate,sf_rate2001$sp_rate)
```

```
## [1] -0.2342387
```

There is a week negative correlation between housing price and stock price in 2001

```
cor(sf_rate$house_rate,sf_rate$sp_rate)
```

```
## [1] 0.09625693
```

But, the longrun was more likely to be uncorrelated because other factors which influnced behaviors in both markets

*Limitation* : There are too many confounding variables that influence both market either into the same direction or the different direction. Our research is only based on observation on short term period and a specific region, we don't have sufficient information and skills to do the full economical analysis.