Housing Price

Di Zhao

1. Description of the data:

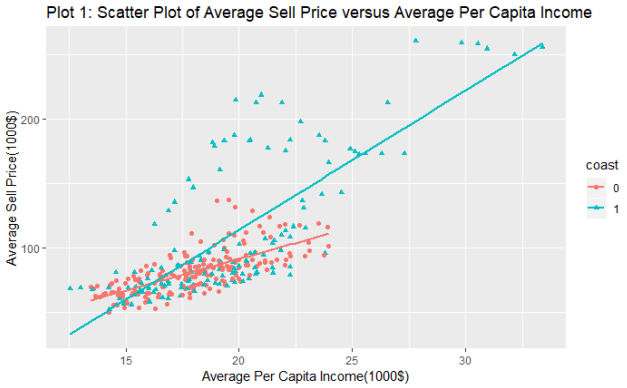
The housing dataset is collected on average sale prices of houses in several metropolitan statistical areas. The description of variables is as follow:

DESCRIPTION	*RANGE/LEVEL*	*DATA TYPE*
AVERAGE PER CAPITA INCOME ($)	From 12,535 to 33,383	Continuous quantitative variable
RENT CONTROL	1: some rent control, 0: no	Qualitative variable
ADJACENT TO A COASTLINE	1: yes, 0: no	Qualitative variable
YEAR	From 1981 to 1989	Discrete quantitative variable
AVERAGE SALE PRICE ($)	From 50,400 to 260,600	Continuous quantitative variable

As far as I am concerned average per capita income would definite affect the average sale price. The richer people are, the more likely they will buy a house. The rent control and year are also major problems that could affect the average sale price. But sometimes, the quality of the real estate environment will affect people's decision about buying a house, and whether adjacent to a coastline or not might be one of those influencing factors. Does the housing economic in adjacent coast area perform better than in not adjacent coast area?

2. Investigation about the study:

Graph:

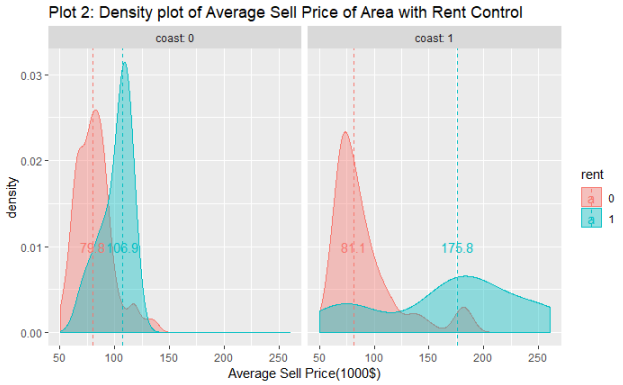


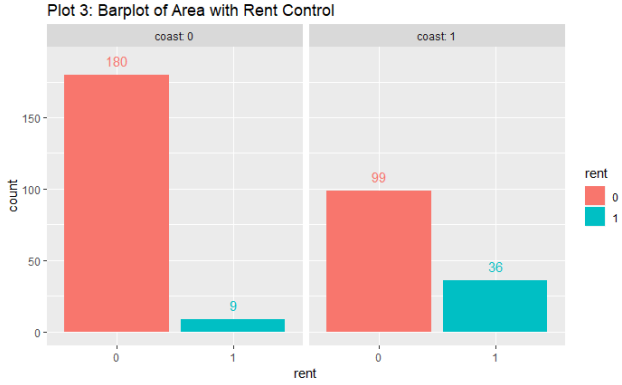
Summary:

$$f(x) = \begin{cases} -6.546 + 4.916x, & \text{when not } adjacent\ to\ coast \\ -102.032 + 10.816x, & \text{when } adjacent\ to\ coast \end{cases},$$

Conclusion: Plot 1 and summary function show us the slopes and the intercepts of the simple linear model for average sell price versus average per capita income by different coast levels are different, which means the effect of per capita income on average sell price depends on different whether adjacent to coast or not. The higher average per capita income area with house in adjacent coast area will have higher sell price, when comparing to not adjacent coast area; the lower average per capita income area with house in adjacent coast area will have lower sell price, when comparing to not adjacent coast area.

Graph:



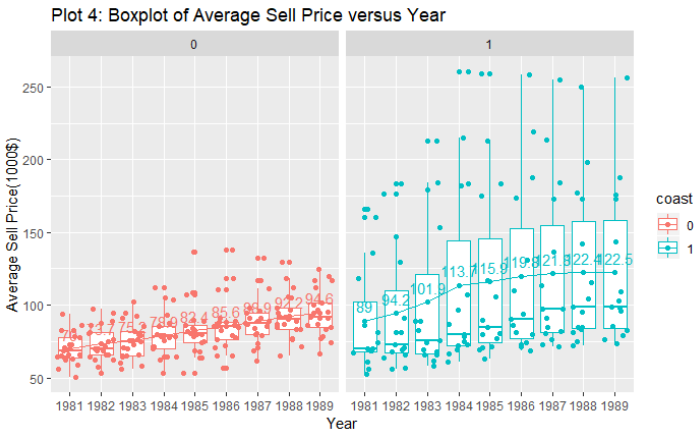


Summary:

COAST	*RENT*	*MEDIAN (1000$)*	*SD (1000$)*	*N*
0	0	79.8	16.50755	180
0	1	106.9	14.07502	9
1	0	81.1	30.49703	99
1	1	175.8	64.41945	36

Conclusion: The area with a sound legal system of rent control will always be rich area, and the price of house will be higher than those without. In plot 3 though the median average sell price for both with or without rent control in adjacent coast area are higher than the median average sell price for both with and without rent control in not adjacent coast area. The people who want to buy the adjacent coast area house are less than not adjacent coast area house.

Graph: Boxplot



Summary:

YEAR	*COAST*	*MEDIAN (1000$)*	*SD (1000$)*
1981	0	68.3	10.58706
1981	1	69.9	37.52756
1982	0	70.2	12.00591
1982	1	72.5	43.14052
1983	0	74.6	12.6356
1983	1	75.7	52.18821
1984	0	78.1	14.92035
1984	1	79.6	63.60979
1985	0	80.7	17.99842
1985	1	84.4	61.42139
1986	0	85.2	18.33309
1986	1	90	60.81602
1987	0	87.6	16.4118
1987	1	97	57.73912
1988	0	90.1	15.49699
1988	1	98.8	53.80031
1989	0	91.4	15.91721
1989	1	98.5	53.28107

Conclusion: Plot 4 shows us the median of the average sell price is increasing with the years, no matter adjacent to coast or not, which means the housing economy is getting better year by year. But in adjacent coast area, the median of average sell price is increasing more than that in not adjacent coast are, which means the housing economy in adjacent coast area booming relatively faster.

3. Summary:

The higher average per capita income area with house in adjacent coast area will have higher sell price, when comparing to not adjacent coast area; the lower average per capita income area with house in adjacent coast area will have lower sell price, when comparing to not adjacent coast area. The median average sell price for both with or without rent control in adjacent coast area are higher than the median average sell price for both with and without rent control in not adjacent coast area. The housing economy in adjacent coast area booming relatively faster. However, housing sell price in adjacent coast area is higher in all aspects than that in not adjacent coast area, it does not mean that housing economy in adjacent coast area is better. For example, the house in adjacent coast area will have higher sell price, but low sell volume. The total sell price in adjacent coast area may not greater than in not adjacent coast area.

Average sell price does not mean we can get conclusion about the housing economic. It's like GDP per capita doesn't show if a country is rich or not. Thus, we need to focus more on the meaning behind the value instead of value itself.

Appendix

```
#input data
setwd("C:/Users/Di/Desktop/566project")
housing <- read.csv("housing.csv", header = T)
names(housing)[1] <- "ind"

housing[2] <- housing[,2]/1000
housing[8] <- 1980 + housing[,8]
#factor
housing[5:8] <- lapply(housing[,5:8], as.factor)
```

#1 Description ##Summary data

```
summary(housing[,c(2:4,9)])
```

```
##       apci          p.growth           rei            price
##  Min.   :12.54   Min.   :-2.054   Min.   :13.00   Min.   : 50.40
##  1st Qu.:16.61   1st Qu.: 3.535   1st Qu.:18.00   1st Qu.: 71.08
##  Median :18.45   Median : 3.964   Median :20.00   Median : 82.45
##  Mean   :18.77   Mean   : 4.268   Mean   :20.42   Mean   : 94.41
##  3rd Qu.:20.32   3rd Qu.: 5.711   3rd Qu.:22.00   3rd Qu.: 97.00
##  Max.   :33.38   Max.   : 8.788   Max.   :29.00   Max.   :260.60
```

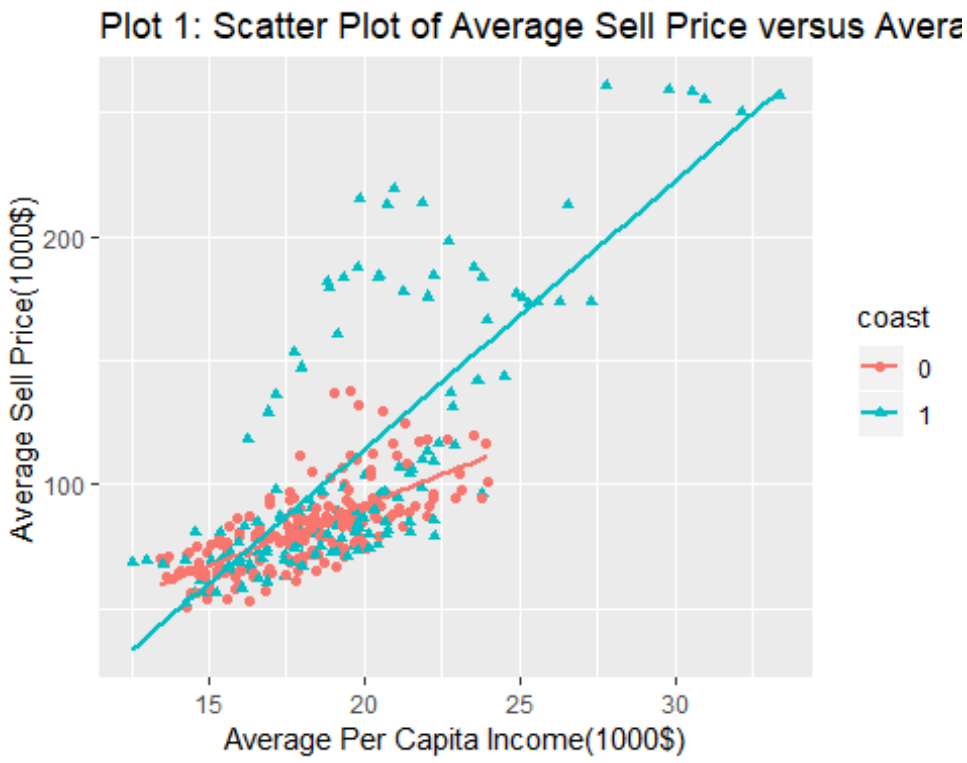
```
str(housing[,-c(2:4,9)])
```

```
## 'data.frame':    324 obs. of  5 variables:
##  $ ind  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ rent : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ coast: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ msa  : Factor w/ 36 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 2 ...
##  $ year : Factor w/ 9 levels "1981","1982",..: 1 2 3 4 5 6 7 8 9 1 ...
```

#rent

#2 investigation ##Plot 1: Relationship between Price and AVERAGE PER CAPITA INCOME depend on Coast

```
#graph
library(ggplot2)
p1 <- ggplot(housing, aes(apci, price)) +
  geom_point(aes(colour = coast, shape = coast)) +
  geom_smooth(aes(colour = coast), se=FALSE, method="lm") +
  ggtitle("Plot 1: Scatter Plot of Average Sell Price versus Average Per Capi
ta Income") +
  xlab("Average Per Capita Income(1000$)") +
  ylab("Average Sell Price(1000$)")
p1
```




```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

#statistics
fit1 = housing %>% group_by(coast) %>% do(model = lm(price ~ apci, data = .))
paste("coast:",levels(housing$coast)[1])

## [1] "coast: 0"

round(fit1$model[[1]]$coefficients,3)

## (Intercept)        apci
##      -6.546       4.916

paste("coast:",levels(housing$coast)[2])

## [1] "coast: 1"

round(fit1$model[[2]]$coefficients,3)
```

```
## (Intercept)        apci
##    -102.032      10.816
```

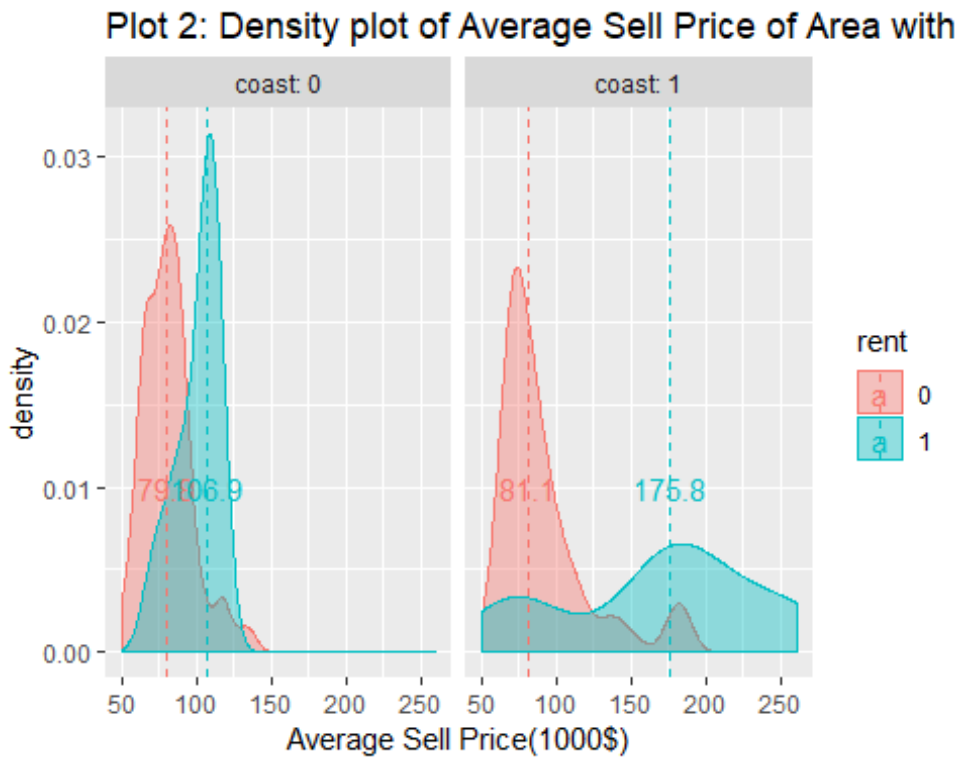
##Plot 2: Boxplot of Average Sell Price of Area with Rent Control

```
library(dplyr)
#statistic
#https://stackoverflow.com/questions/10220510/summary-statistics-by-two-or-more-factor-variables
house <- housing %>%
select(coast, rent, price) %>%
group_by(coast, rent) %>%
summarise(median = median(price), sd = sd(price), n = length(price))
house
```

```
## # A tibble: 4 x 5
## # Groups:   coast [2]
##   coast rent  median    sd     n
##   <fct> <fct>  <dbl> <dbl> <int>
## 1 0     0       79.8  16.5   180
## 2 0     1      107.   14.1     9
## 3 1     0       81.1  30.5    99
## 4 1     1      176.   64.4    36
```

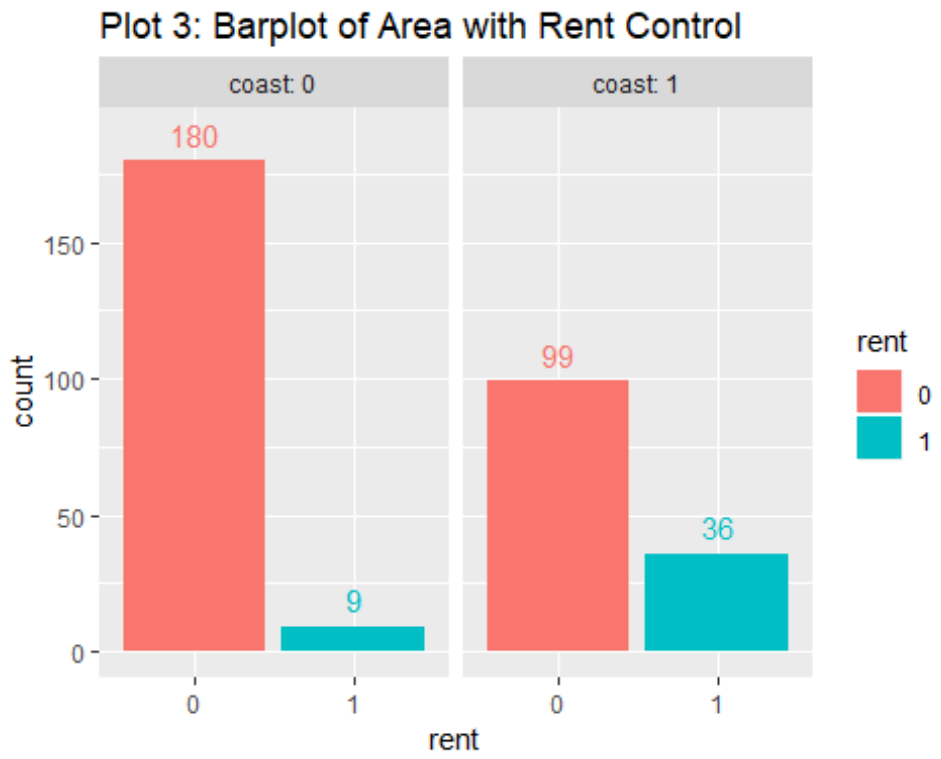
```
write.csv(house, file = "stat2.csv", row.names = T)

#https://stackoverflow.com/questions/23010009/how-to-create-three-way-interaction-graph-in-r-i-have-example-of-graph
#https://ggplot2.tidyverse.org/reference/labellers.html
ggplot(data = housing, aes(x = price)) +
  geom_density(aes(group = rent, colour = rent, fill = rent),
               alpha=0.4) +
  geom_vline(data = house, aes(xintercept=house$median, colour=rent),
             linetype="dashed", size=0.5) +
  geom_text(data = house, aes(label = house$median, y=0.01, x=house$median, colour=rent)) +
  facet_wrap(.~ coast, labeller=label_both) +
  ggtitle("Plot 2: Density plot of Average Sell Price of Area with Rent Control") +
  xlab("Average Sell Price(1000$)")
```



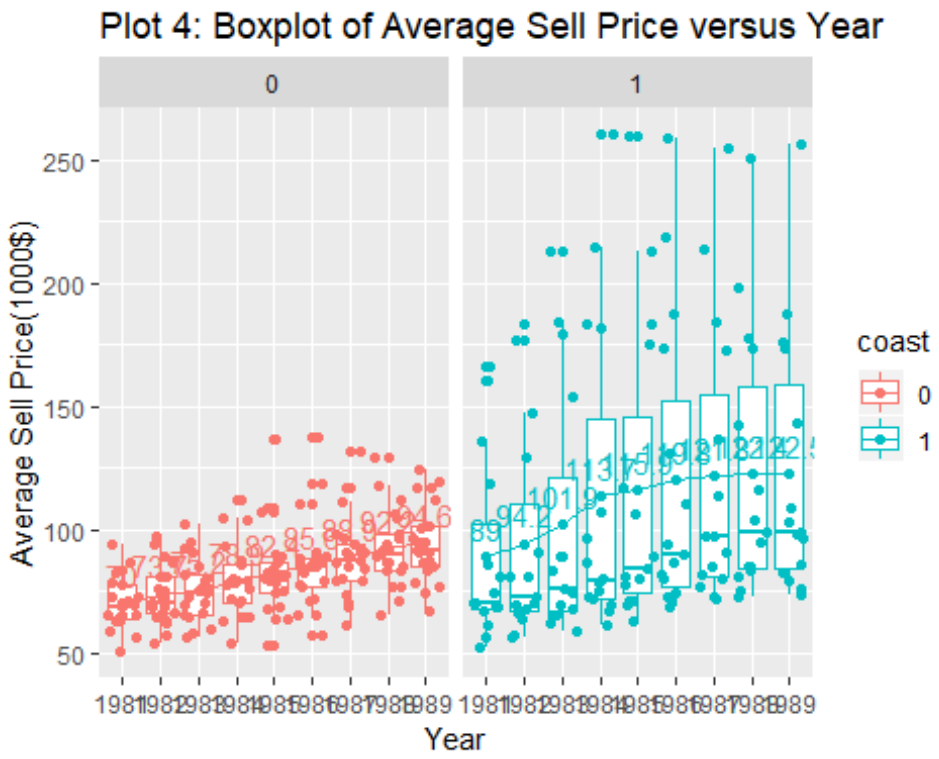
#Plot 3: Barplot of Area with Rent Control

```
#bar plot
ndf <- group_by(housing, rent, coast) %>% summarize( count = n() )
# Stacked barplot with multiple groups
ggplot(data=ndf , aes(x=rent, y = count, fill=rent)) +
  geom_bar(stat="identity", position = "dodge")+
  ggtitle("Plot 3: Barplot of Area with Rent Control") +
  geom_text(data = house, aes(label = n, y=n+10, x=rent, colour=rent)) +
  facet_wrap(.~ coast, labeller=label_both)
```



#Plot 4: Boxplot of Average Sell Price versus Year

```
#graph
#https://sebastiansauer.github.io/vis_interaction_effects/
ggplot(housing) +
  aes(x = year, y = price, color = coast) +
  geom_boxplot() +
  geom_jitter() +
  facet_wrap(~coast) +
  stat_summary(fun.y=mean, geom="line", aes(group=1))  +
  stat_summary(fun.y=mean, geom="point") +
  stat_summary(fun.y=mean, geom="text", show.legend = FALSE,
               vjust=-0.7, aes( label=round(..y.., digits=1))) +
  ggtitle("Plot 4: Boxplot of Average Sell Price versus Year") +
  ylab("Average Sell Price(1000$)") +
  xlab("Year")
```




```
setwd("C:/Users/Di/Desktop/566project")
library(dplyr)
house2 <- housing %>%
select(year, coast, price) %>%
group_by(year, coast) %>%
summarise(median = median(price), sd = sd(price))
house2
```

```
## # A tibble: 18 x 4
## # Groups:   year [9]
##    year  coast median    sd
##    <fct> <fct>  <dbl> <dbl>
##  1 1981  0       68.3  10.6
##  2 1981  1       69.9  37.5
##  3 1982  0       70.2  12.0
##  4 1982  1       72.5  43.1
##  5 1983  0       74.6  12.6
##  6 1983  1       75.7  52.2
##  7 1984  0       78.1  14.9
##  8 1984  1       79.6  63.6
##  9 1985  0       80.7  18.0
## 10 1985  1       84.4  61.4
## 11 1986  0       85.2  18.3
## 12 1986  1       90    60.8
## 13 1987  0       87.6  16.4
## 14 1987  1       97    57.7
## 15 1988  0       90.1  15.5
```

```
## 16 1988  1       98.8  53.8
## 17 1989  0       91.4  15.9
## 18 1989  1       98.5  53.3
```

```
write.csv(house2, file = "stat
```