HW1

Shijie Gao, USC ID:6037-6293-25

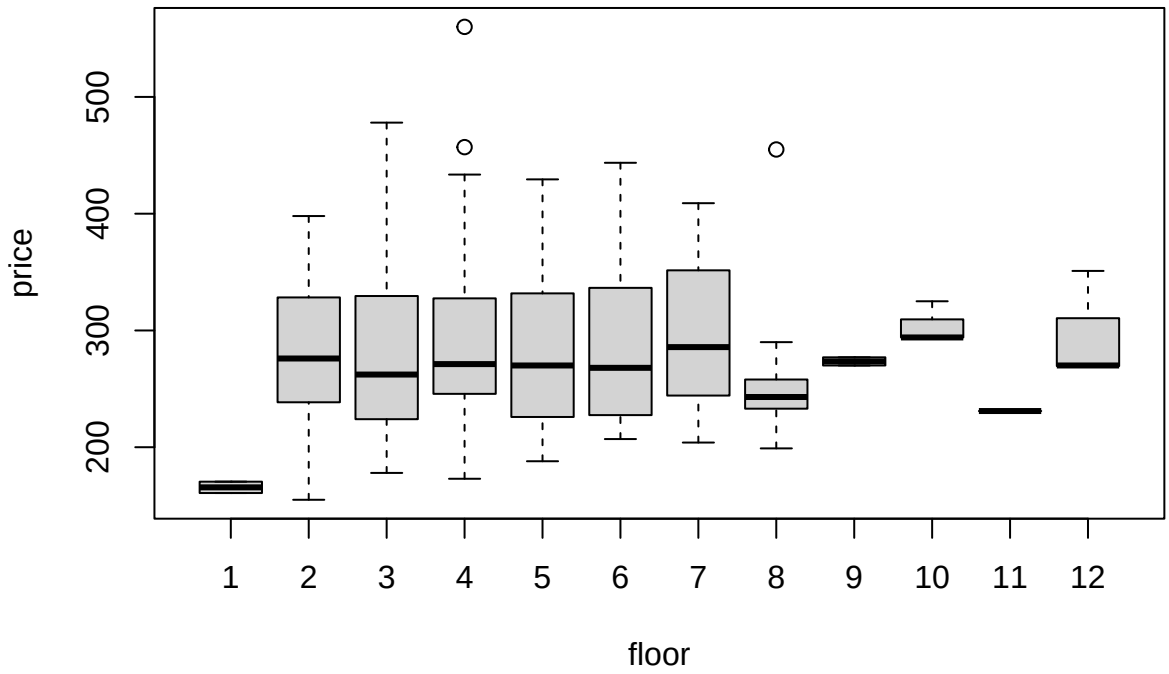
2023-01-15

```
setwd("C:/Users/GAOSHIJIE/Desktop")
data = read.csv("VIT.csv")
data["price"] = data["price"] / 1000
data[c("toilets", "garage", "elevator", "storage")] <- lapply(data[c("toilets",
                                    "garage", "elevator", "storage")], factor)

#Question 1:
table(data$garage)
```

```
##
##   0   1   2
## 167  49   2
```

```
#Question 2:
boxplot(price~floor, data)
```



```r
#Question 3:
table(data$rooms, data$garage)
```

```
##
##       0   1   2
##   3   3   0   0
##   4  46   5   0
##   5 104  35   2
##   6  13   8   0
##   7   1   1   0
```

```r
#Question 4: mean price per type of room & garage
aux = list(data$rooms, data$garage)
signif(tapply(data$price, aux, mean), 2) #up to 2 sig. digits
```

```
##     0   1   2
## 3 230  NA  NA
## 4 230 280  NA
## 5 260 340 370
## 6 360 400  NA
## 7 440 290  NA
```

```r
#Question 5: find max and min price s.t. 80<=area<=90
d = data[data$area <= 90,]
d = d[d$area >= 80,]
d_max = max(d$price)
d_min = min(d$price)
d_max
```

```
## [1] 398
```

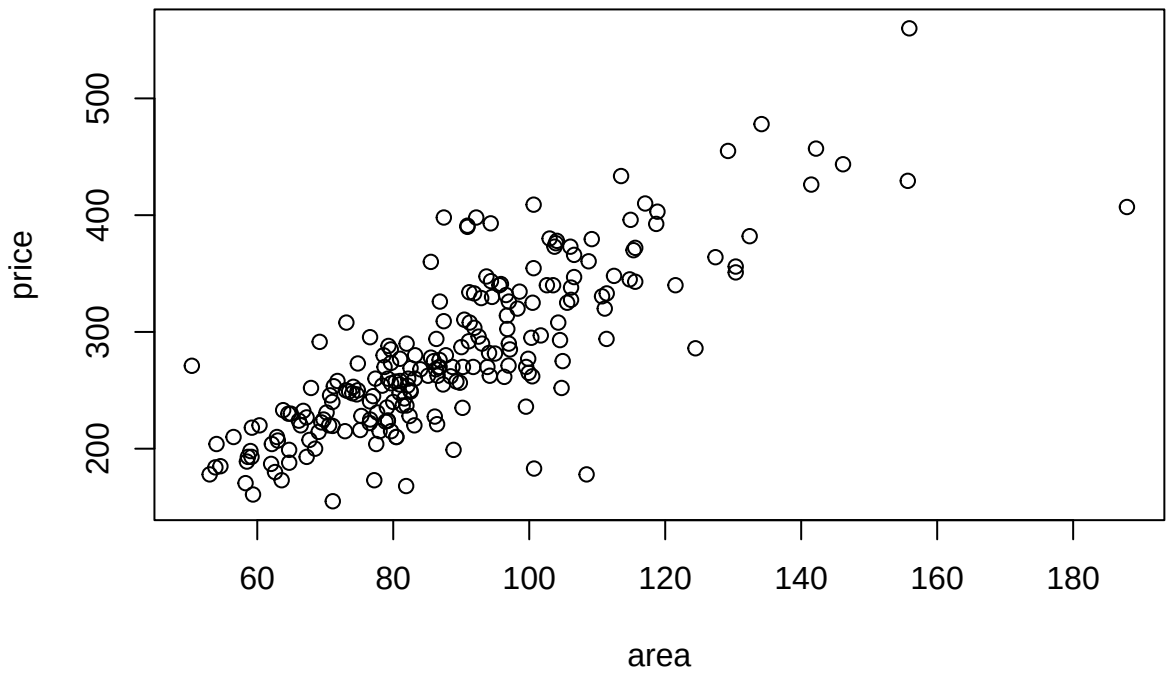
```r
d_min
```

```
## [1] 168
```

```r
#Question 6: Draw a scatterplot for price and most cor variable with price
data_num = data[,sapply(data, is.numeric)] #choose numeric var
#calculate cormatrx
cor_matrix = as.data.frame(cor(data_num, method = "spearman",  use = "everything"))
#choose max cor except price
cor_most = order(cor_matrix$price, decreasing = TRUE)[2]
var_most_name = names(cor_matrix)[cor_most] #var_name
var_most_name
```

```
## [1] "area"
```

```r
#So area is the most correlated variable with price
plot(price~area, data_num)
```




```
#Question 7: Calculate avg diff price between S4 and S2
#get avg price of S4
S4_mean = apply(data[data$street == "S4",]["price"], 2, mean)
#get avg price of S2
S2_mean = apply(data[data$street == "S2",]["price"], 2, mean)
diff = S4_mean - S2_mean
diff
```

```
##    price
## 10.67302
```

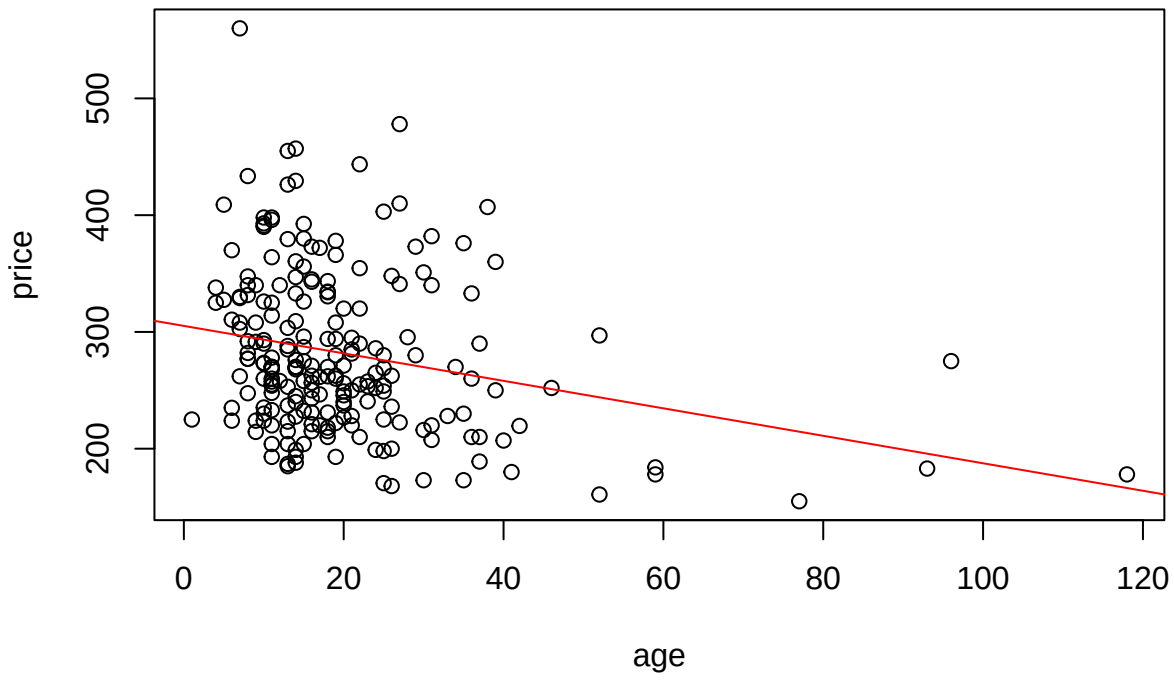
```
#S4 is 10.67302 thousands expensive than S2 on average

#Question 8:
plot(price~age, data) #plot all points

diffage_mean = aggregate(price ~ age, data, mean) #calculate mean for different age
m1 = lm(price~age, diffage_mean) #linear regression
coefficients(m1) #get coef
```

```
## (Intercept)         age
##   305.15944    -1.17657
```

```
abline(m1, col = "red") #draw the regression line
```

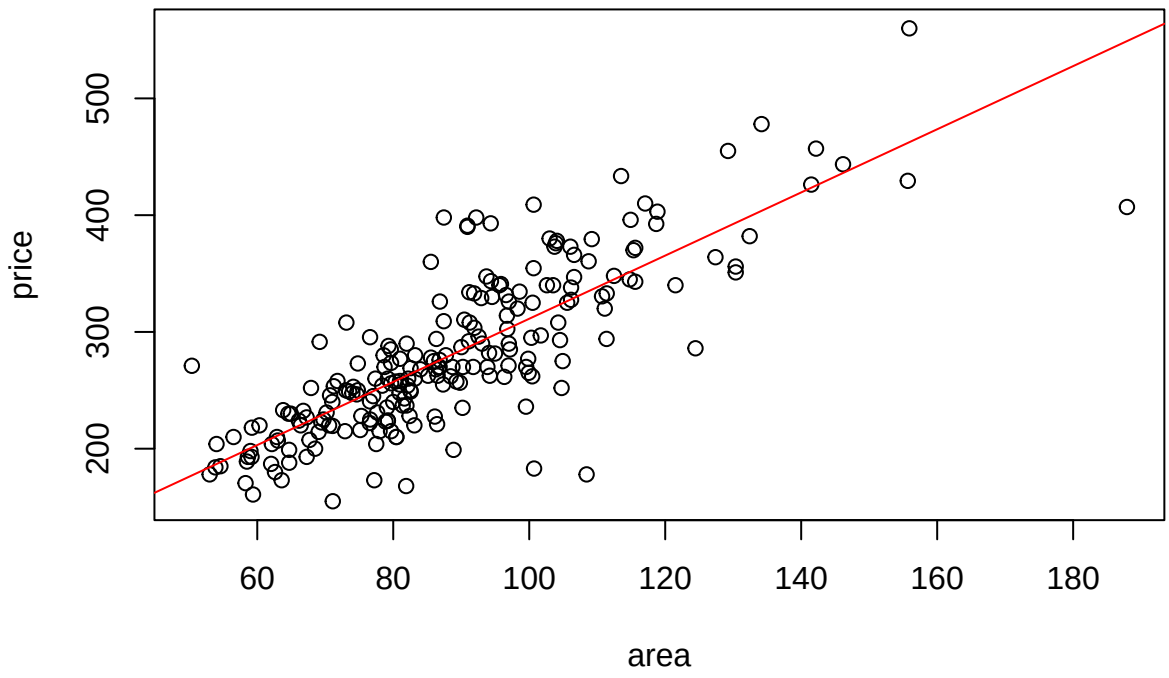



```
# 1.17657 thousand decrease for each additional year

#Question 9:
plot(price~area, data) #plot all points
m2 = lm(price~area, data)
coefficients(m2) #get coef
```

```
## (Intercept)        area
##   40.817980    2.704749
```

```
abline(m2 ,col = "red") #draw the regression line
```




```
# 2.704749 thousand increase for each additional square meter
```