HW1

Yating Liao (7636428840)

2023-01-17

```
d1 = read.csv("VIT.csv")

# Import data and Transfer factors
d1$toilets=as.factor(d1$toilets)
d1$garage=as.factor(d1$garage)
d1$elevator=as.factor(d1$elevator)
d1$storage=as.factor(d1$storage)
d1$zone=as.factor(d1$zone)
d1$category=as.factor(d1$category)
d1$out=as.factor(d1$out)
d1$conservation=as.factor(d1$conservation)
d1$street=as.factor(d1$street)
d1$heating=as.factor(d1$heating)
d1$street=as.factor(d1$street)
d1$good_price=d1$price/1000
str(d1)
```

```
## 'data.frame':    218 obs. of  16 variables:
##  $ price       : int  228000 409000 199000 180000 443600 173000 245000 210000 188982 193000 ...
##  $ good_price  : num  228 409 199 180 444 ...
##  $ area        : num  75.3 100.7 88.9 62.6 146.1 ...
##  $ zone        : Factor w/ 23 levels "Z11","Z21","Z31",..: 14 3 19 23 3 1 16 17 1 1 ...
##  $ category    : Factor w/ 7 levels "2A","2B","3A",..: 6 4 3 5 3 6 3 4 6 4 ...
##  $ age         : int  33 5 14 41 22 35 14 36 37 11 ...
##  $ floor       : int  3 7 8 3 6 4 6 3 4 5 ...
##  $ rooms       : int  5 5 5 4 7 5 4 4 4 4 ...
##  $ out         : Factor w/ 4 levels "E100","E25","E50",..: 1 3 3 3 1 3 3 1 2 3 ...
##  $ conservation: Factor w/ 4 levels "1A","2A","2B",..: 3 1 1 2 1 1 1 4 2 1 ...
##  $ toilets     : Factor w/ 2 levels "1","2": 1 2 2 1 2 1 1 1 1 1 ...
##  $ garage      : Factor w/ 3 levels "0","1","2": 1 2 1 1 1 1 1 1 1 1 ...
##  $ elevator    : Factor w/ 2 levels "0","1": 2 2 2 1 2 1 2 1 1 1 ...
##  $ street      : Factor w/ 4 levels "S2","S3","S4",..: 2 4 1 2 3 3 2 2 3 3 ...
##  $ heating     : Factor w/ 4 levels "1A","3A","3B",..: 2 4 2 1 4 2 4 2 2 2 ...
##  $ storage     : Factor w/ 3 levels "0","1","2": 1 2 1 1 2 2 2 2 2 2 ...
```

```
#1. The number of apartments for each number of garages
levels(d1$garage)
```

```
## [1] "0" "1" "2"
```

```
table(d1$garage)
```

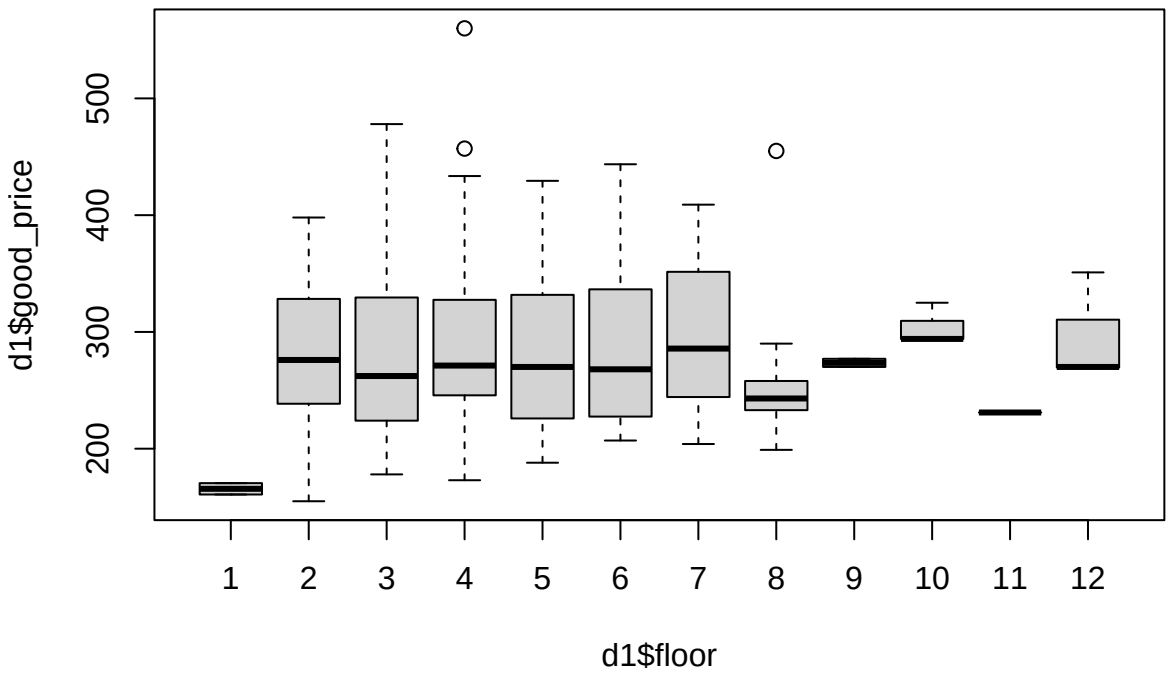
```
##
##   0   1   2
```

```
## 167  49   2
```

```
summary(d1)
```

```
##      price          good_price         area             zone     category
##  Min.   :155000   Min.   :155.0   Min.   : 50.38   Z41    : 18   2A: 4
##  1st Qu.:228500   1st Qu.:228.5   1st Qu.: 75.18   Z53    : 16   2B:14
##  Median :269750   Median :269.8   Median : 86.39   Z52    : 15   3A:61
##  Mean   :280737   Mean   :280.7   Mean   : 88.70   Z61    : 14   3B:77
##  3rd Qu.:328625   3rd Qu.:328.6   3rd Qu.: 99.90   Z62    : 14   4A:36
##  Max.   :560000   Max.   :560.0   Max.   :187.91   Z45    : 13   4B:23
##                                                    (Other):128   5A: 3
##       age            floor           rooms          out      conservation
##  Min.   :  1.0   Min.   : 1.00   Min.   :3.000   E100:122   1A:161
##  1st Qu.: 11.0   1st Qu.: 3.00   1st Qu.:5.000   E25 :  3   2A: 18
##  Median : 16.0   Median : 4.00   Median :5.000   E50 : 87   2B: 36
##  Mean   : 19.9   Mean   : 4.44   Mean   :4.853   E75 :  6   3A:  3
##  3rd Qu.: 24.0   3rd Qu.: 5.00   3rd Qu.:5.000
##  Max.   :118.0   Max.   :12.00   Max.   :7.000
##
##  toilets garage  elevator street   heating  storage
##  1:116   0:167   0: 44    S2: 42   1A:  8   0: 43
##  2:102   1: 49   1:174    S3:107   3A:149   1:174
##          2:  2            S4: 59   3B: 10   2:  1
##                           S5: 10   4A: 51
##
##
##
```

```
# 2.boxplot for the apartment's price for each floor
boxplot(d1$good_price~d1$floor,d1)
```




```
# 3.table showing No.of apartments for each No.of rooms and garages
t1 = table(rooms = d1$rooms,garage = d1$garage)
t1
```

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

```
#4. Average apartment price(mean) for each room and garage
list1 = list(rooms = d1$rooms,garage = d1$garage)
t1=tapply(d1$good_price,list1,mean) # no price
round(t1,2)
```

```
##       garage
## rooms      0     1      2
##     3 230.33    NA     NA
##     4 229.27 279.2     NA
##     5 261.16 344.1 369.25
##     6 358.99 403.5     NA
##     7 443.60 286.0     NA
```

```
#5. The min and max price of apartments with area between 80 and 90 square meters
d2 = d1[d1$area>=80|d1$area<=90,c(2)]
head(d2)
```

```
## [1] 228.0 409.0 199.0 180.0 443.6 173.0
```

```
min(d2)
```

```
## [1] 155
```

```
max(d2)
```

```
## [1] 560
```

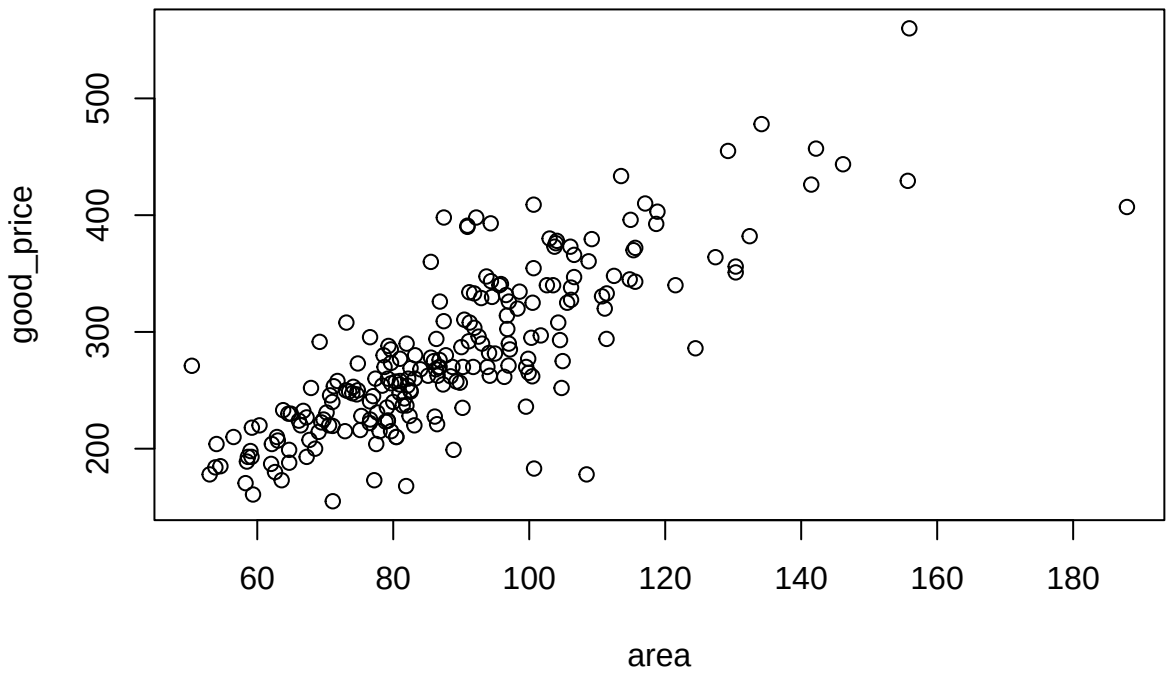
```
#6. The numerical variable that is most correlated with price
d3 = d1[,sapply(d1,is.numeric)]
str(d3)
```

```
## 'data.frame':    218 obs. of  6 variables:
##  $ price     : int  228000 409000 199000 180000 443600 173000 245000 210000 188982 193000 ...
##  $ good_price: num  228 409 199 180 444 ...
##  $ area      : num  75.3 100.7 88.9 62.6 146.1 ...
##  $ age       : int  33 5 14 41 22 35 14 36 37 11 ...
##  $ floor     : int  3 7 8 3 6 4 6 3 4 5 ...
##  $ rooms     : int  5 5 5 4 7 5 4 4 4 4 ...
```

```
d3$price = NULL
cor(d3) #area is most correlated with price
```

```
##             good_price        area         age       floor       rooms
## good_price  1.00000000  0.80914892 -0.27240165  0.02910637  0.52557062
## area        0.80914892  1.00000000 -0.05226235  0.08165600  0.63816604
## age        -0.27240165 -0.05226235  1.00000000 -0.08124348 -0.08274509
## floor       0.02910637  0.08165600 -0.08124348  1.00000000  0.13011339
## rooms       0.52557062  0.63816604 -0.08274509  0.13011339  1.00000000
```

```
plot(good_price~area,d3)
```


good_price
200
300
400
500
60
80
100
120
140
160
180
area