assignment\_6\_r.R

vnath

2023-05-04

rm(list=ls())  
cat("/014")

## /014

setwd("C:/Users/vnath/Documents/Uni/STTN 317/assignment\_6")  
  
library(ggplot2)  
library(corrplot)

## corrplot 0.92 loaded

library(olsrr)

## Warning: package 'olsrr' was built under R version 4.2.3

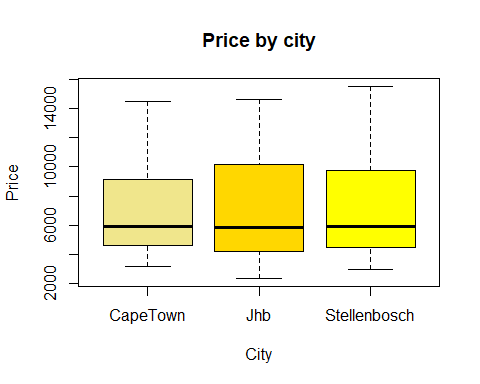
##   
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':  
##   
## rivers

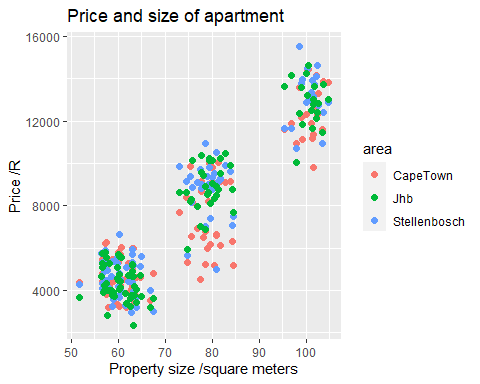
df <- data.frame(read.csv("rent\_data.csv"))  
attach(df)  
  
# ===== Question 1 ===== #  
head(df)

## area bedr bathr kitchen prop\_sqr\_m dist\_to\_varsity dist\_to\_bookshop price  
## 1 CapeTown 1 1 old 62.5 8.3 12.5 4617  
## 2 CapeTown 1 1 old 59.2 12.2 16.4 5380  
## 3 CapeTown 1 1 old 58.9 7.6 12.1 4441  
## 4 CapeTown 1 1 old 60.3 11.2 15.2 3235  
## 5 CapeTown 1 1 old 66.8 11.7 15.5 3537  
## 6 CapeTown 1 1 old 62.5 12.0 15.0 3970

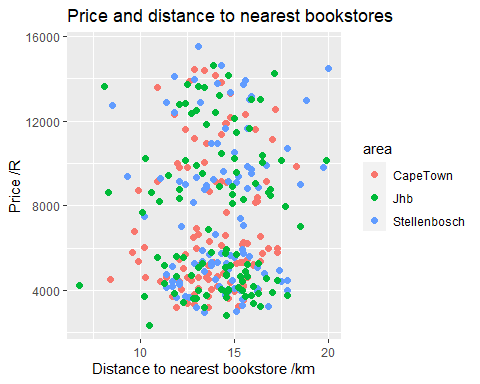
# ===== Question 2 ===== #  
  
# price and city  
boxplot(price~area, main="Price by city", xlab="City", ylab="Price", col=c("khaki","gold","yellow"))



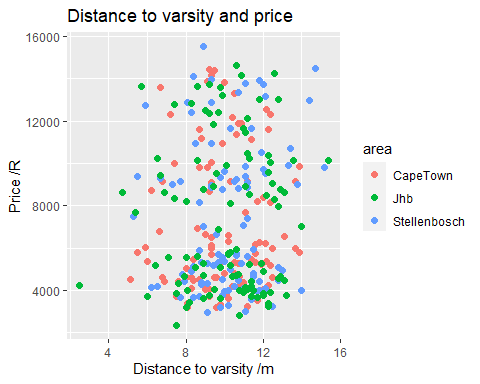
# price and size of apartment  
ggplot(df, aes(x=prop\_sqr\_m, y=price, color=area)) + geom\_point(size=2, ) + ggtitle("Price and size of apartment") +   
 labs(x="Property size /square meters", y="Price /R")



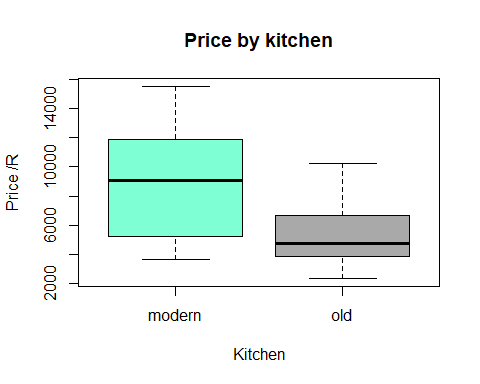
# distance to nearest academic bookshop and price  
ggplot(df, aes(x=dist\_to\_bookshop, y=price, color=area)) + geom\_point(size=2) + ggtitle("Price and distance to nearest bookstores") +   
 labs(x="Distance to nearest bookstore /km", y="Price /R")



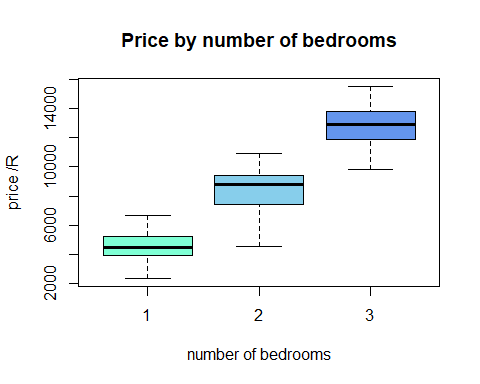
# distance to varsity and price  
ggplot(df, aes(x=dist\_to\_varsity, y=price, color=area)) + geom\_point(size=2) + labs(x="Distance to varsity /m", y="Price /R") + ggtitle("Distance to varsity and price")



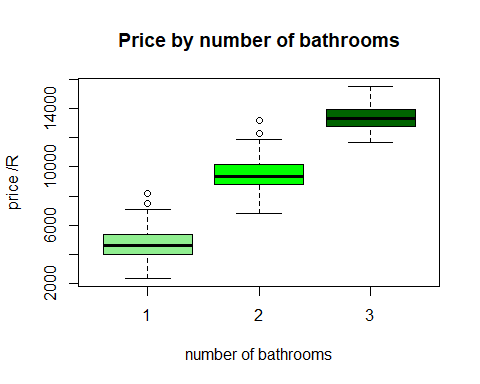
# kitchen and price  
boxplot(price~kitchen, main="Price by kitchen", xlab="Kitchen", ylab="Price /R", col=c("aquamarine","darkgrey"))



# bedroom, bathroom, and price  
boxplot(price~bedr, main="Price by number of bedrooms", xlab="number of bedrooms", ylab="price /R", col=c("aquamarine","skyblue","cornflowerblue"))



boxplot(price~bathr, main="Price by number of bathrooms", xlab="number of bathrooms", ylab="price /R", col=c("lightgreen","green","darkgreen"))



# ===== Question 3 ===== #  
df2 <- subset(df, select = c(-kitchen,-area))  
round(cor(df2),3)

## bedr bathr prop\_sqr\_m dist\_to\_varsity dist\_to\_bookshop price  
## bedr 1.000 0.903 0.982 0.048 0.025 0.940  
## bathr 0.903 1.000 0.892 0.036 0.009 0.948  
## prop\_sqr\_m 0.982 0.892 1.000 0.031 0.009 0.922  
## dist\_to\_varsity 0.048 0.036 0.031 1.000 0.959 0.055  
## dist\_to\_bookshop 0.025 0.009 0.009 0.959 1.000 0.031  
## price 0.940 0.948 0.922 0.055 0.031 1.000

head(df2)

## bedr bathr prop\_sqr\_m dist\_to\_varsity dist\_to\_bookshop price  
## 1 1 1 62.5 8.3 12.5 4617  
## 2 1 1 59.2 12.2 16.4 5380  
## 3 1 1 58.9 7.6 12.1 4441  
## 4 1 1 60.3 11.2 15.2 3235  
## 5 1 1 66.8 11.7 15.5 3537  
## 6 1 1 62.5 12.0 15.0 3970

# ===== Question 4 ===== #  
# first we test for normality of the errors  
Model1 <- lm(price ~ bedr + bathr + prop\_sqr\_m + dist\_to\_varsity + dist\_to\_bookshop, data=df)  
summary(Model1)

##   
## Call:  
## lm(formula = price ~ bedr + bathr + prop\_sqr\_m + dist\_to\_varsity +   
## dist\_to\_bookshop, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2183.7 -603.7 8.6 642.0 2175.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 668.061 821.080 0.814 0.417   
## bedr 2455.221 361.456 6.793 6.11e-11 \*\*\*  
## bathr 2518.138 159.410 15.797 < 2e-16 \*\*\*  
## prop\_sqr\_m -23.935 16.880 -1.418 0.157   
## dist\_to\_varsity -7.601 85.496 -0.089 0.929   
## dist\_to\_bookshop 27.726 82.886 0.335 0.738   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 862.1 on 294 degrees of freedom  
## Multiple R-squared: 0.9373, Adjusted R-squared: 0.9362   
## F-statistic: 878.5 on 5 and 294 DF, p-value: < 2.2e-16

anova(Model1)

## Analysis of Variance Table  
##   
## Response: price  
## Df Sum Sq Mean Sq F value Pr(>F)   
## bedr 1 3078364830 3078364830 4141.8539 <2e-16 \*\*\*  
## bathr 1 183920754 183920754 247.4602 <2e-16 \*\*\*  
## prop\_sqr\_m 1 1650061 1650061 2.2201 0.1373   
## dist\_to\_varsity 1 494121 494121 0.6648 0.4155   
## dist\_to\_bookshop 1 83161 83161 0.1119 0.7382   
## Residuals 294 218510668 743234   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# ===== Question 5 ===== #  
library(car)

## Warning: package 'car' was built under R version 4.2.3

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.2.3

df\_scaled <- as.data.frame(scale(subset(df, select=c(-area, -kitchen))))  
mod <- lm(price ~ bedr + bathr + prop\_sqr\_m + dist\_to\_varsity + dist\_to\_bookshop, data=df\_scaled)  
ols\_coll\_diag(mod)

## Tolerance and Variance Inflation Factor  
## ---------------------------------------  
## Variables Tolerance VIF  
## 1 bedr 0.03108588 32.168946  
## 2 bathr 0.18333770 5.454416  
## 3 prop\_sqr\_m 0.03450297 28.983010  
## 4 dist\_to\_varsity 0.07998524 12.502307  
## 5 dist\_to\_bookshop 0.08017862 12.472153  
##   
##   
## Eigenvalue and Condition Index  
## ------------------------------  
## Eigenvalue Condition Index intercept bedr bathr prop\_sqr\_m  
## 1 2.85708330 1.000000 0 3.700902e-03 0.0203903322 4.068256e-03  
## 2 1.95426251 1.209122 0 1.911098e-05 0.0001717637 3.803404e-05  
## 3 1.00000000 1.690291 1 0.000000e+00 0.0000000000 0.000000e+00  
## 4 0.13070054 4.675443 0 3.049152e-02 0.9495496696 5.114740e-02  
## 5 0.04089285 8.358681 0 5.187234e-04 0.0053952322 9.127751e-05  
## 6 0.01706080 12.940818 0 9.652697e-01 0.0244930022 9.446550e-01  
## dist\_to\_varsity dist\_to\_bookshop  
## 1 9.374577e-05 0.0000523952  
## 2 2.031351e-02 0.0204521486  
## 3 0.000000e+00 0.0000000000  
## 4 3.912041e-04 0.0002936754  
## 5 9.760967e-01 0.9783340508  
## 6 3.104841e-03 0.0008677300

Model2 <- lm(price ~ bathr + prop\_sqr\_m + dist\_to\_varsity + dist\_to\_bookshop, data=df\_scaled)  
ols\_coll\_diag(Model2)

## Tolerance and Variance Inflation Factor  
## ---------------------------------------  
## Variables Tolerance VIF  
## 1 bathr 0.20379716 4.906840  
## 2 prop\_sqr\_m 0.20445487 4.891055  
## 3 dist\_to\_varsity 0.08008749 12.486345  
## 4 dist\_to\_bookshop 0.08018544 12.471092  
##   
##   
## Eigenvalue and Condition Index  
## ------------------------------  
## Eigenvalue Condition Index intercept bathr prop\_sqr\_m dist\_to\_varsity  
## 1 1.9793640 1.000000 0 0.009916505 0.009826280 0.0165565450  
## 2 1.8715737 1.028394 0 0.043946751 0.044221497 0.0038769880  
## 3 1.0000000 1.406899 1 0.000000000 0.000000000 0.0000000000  
## 4 0.1082024 4.277051 0 0.939379270 0.944274148 0.0005197068  
## 5 0.0408599 6.960079 0 0.006757474 0.001678076 0.9790467602  
## dist\_to\_bookshop  
## 1 0.0161784135  
## 2 0.0043276777  
## 3 0.0000000000  
## 4 0.0006069677  
## 5 0.9788869412

# ===== Question 6 =====#  
Model3 <- lm(price ~ dist\_to\_varsity + bedr + area + kitchen, data=df)