**Day 3 Lab Manual**

**UNIVARIATE ANALYSIS IN R - MEASURES OF CENTRAL TENDENCY**

**Exercise:**

**I. ARITHMETIC MEAN**

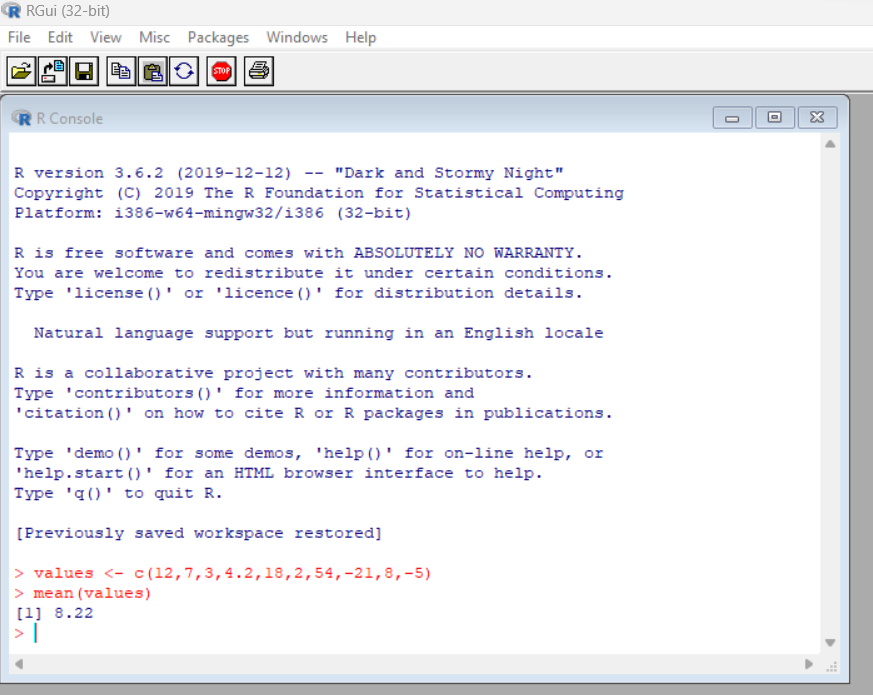
a) Write suitable R code to compute the average of the following values.

12,7,3,4.2,18,2,54,-21,8,-5

values <- c(12,7,3,4.2,18,2,54,-21,8,-5)

mean(values)

[1] 8.22



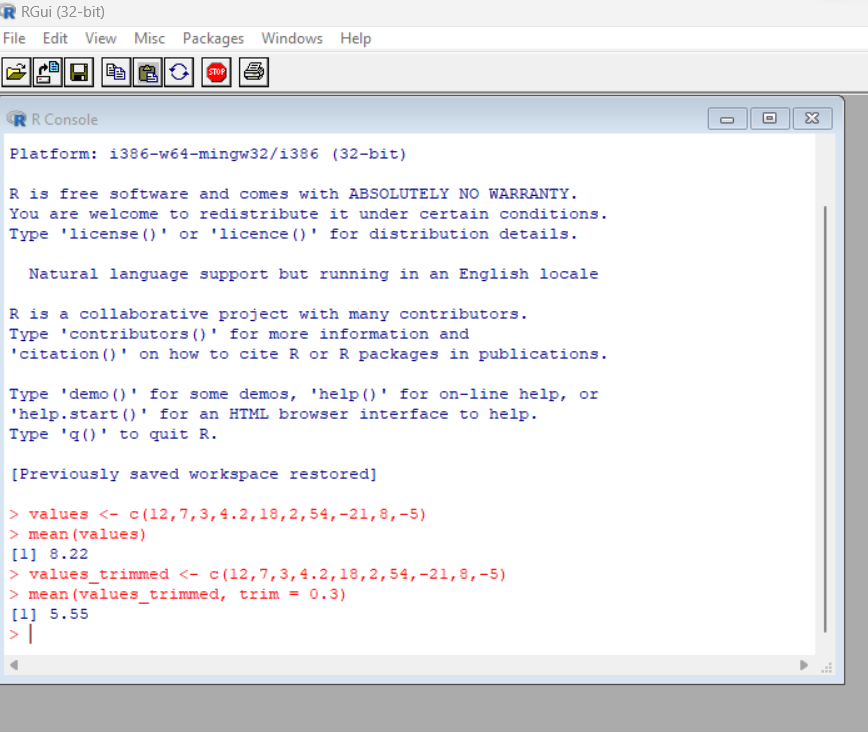
b) Compute the mean after applying the trim option and removing 3 values from each

end.

values\_trimmed <- c(12,7,3,4.2,18,2,54,-21,8,-5)

> mean(values\_trimmed, trim = 0.3)

[1] 5.55



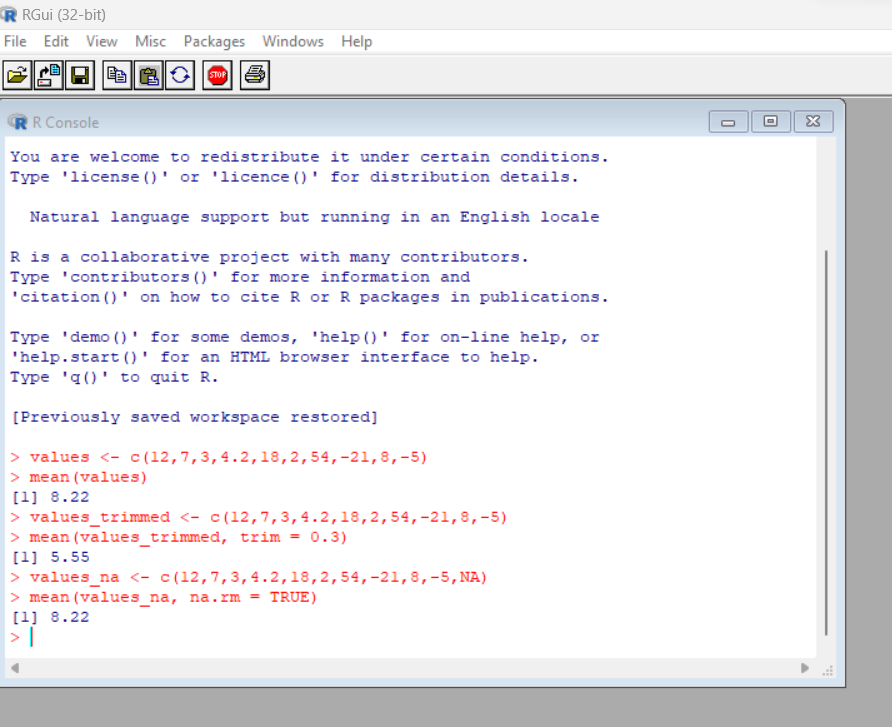
c) Compute the mean of the following vector .

(12,7,3,4.2,18,2,54,-21,8,-5,NA)

> values\_na <- c(12,7,3,4.2,18,2,54,-21,8,-5,NA)

> mean(values\_na, na.rm = TRUE)

[1] 8.22



#If there are missing values, then the mean function returns NA.

# Find mean dropping NA values.

#To drop the missing values from the calculation use na.rm = TRUE

**II.MEDIAN**

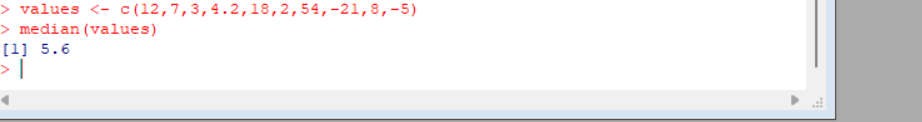
Write suitable R code to compute the median of the following values.

12,7,3,4.2,18,2,54,-21,8,-5

> values <- c(12,7,3,4.2,18,2,54,-21,8,-5)

> median(values)

[1] 5.6



**III. MODE**

Calculate the mode for the following numeric as well as character data set in R.

(2,1,2,3,1,2,3,4,1,5,5,3,2,3) , ("o","it","the","it","it")

> # Create the function.

> getmode <- function(v) {

+ uniqv <- unique(v)

+ uniqv[which.max(tabulate(match(v, uniqv)))]

+ }

>

> # Create the vector with numbers.

> v <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

>

> # Calculate the mode using the user function.

> result <- getmode(v)

> print(result)

[1] 2

>

> # Create the vector with characters.

> charv <- c("o","it","the","it","it")

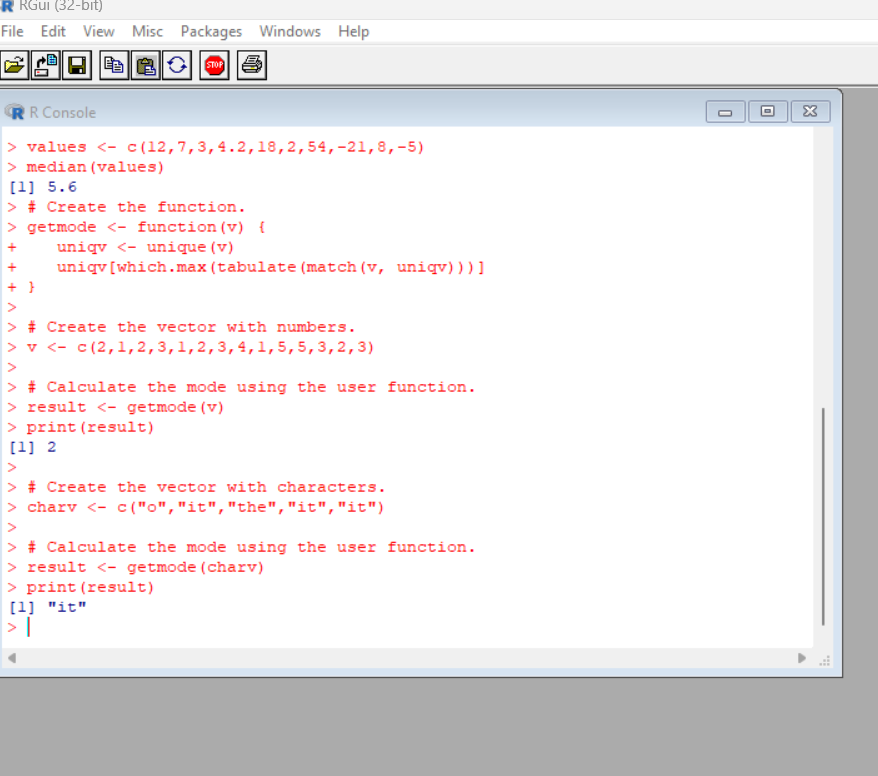
>

> # Calculate the mode using the user function.

> result <- getmode(charv)

> print(result)

[1] "it"



**UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION**

**Exercise: 4**

Download mpg dataset which contains Fuel economy data from 1999 and 2008 for 38 popular models of car from the URL given below.

<https://vincentarelbundock.github.io/Rdatasets/datasets.html>

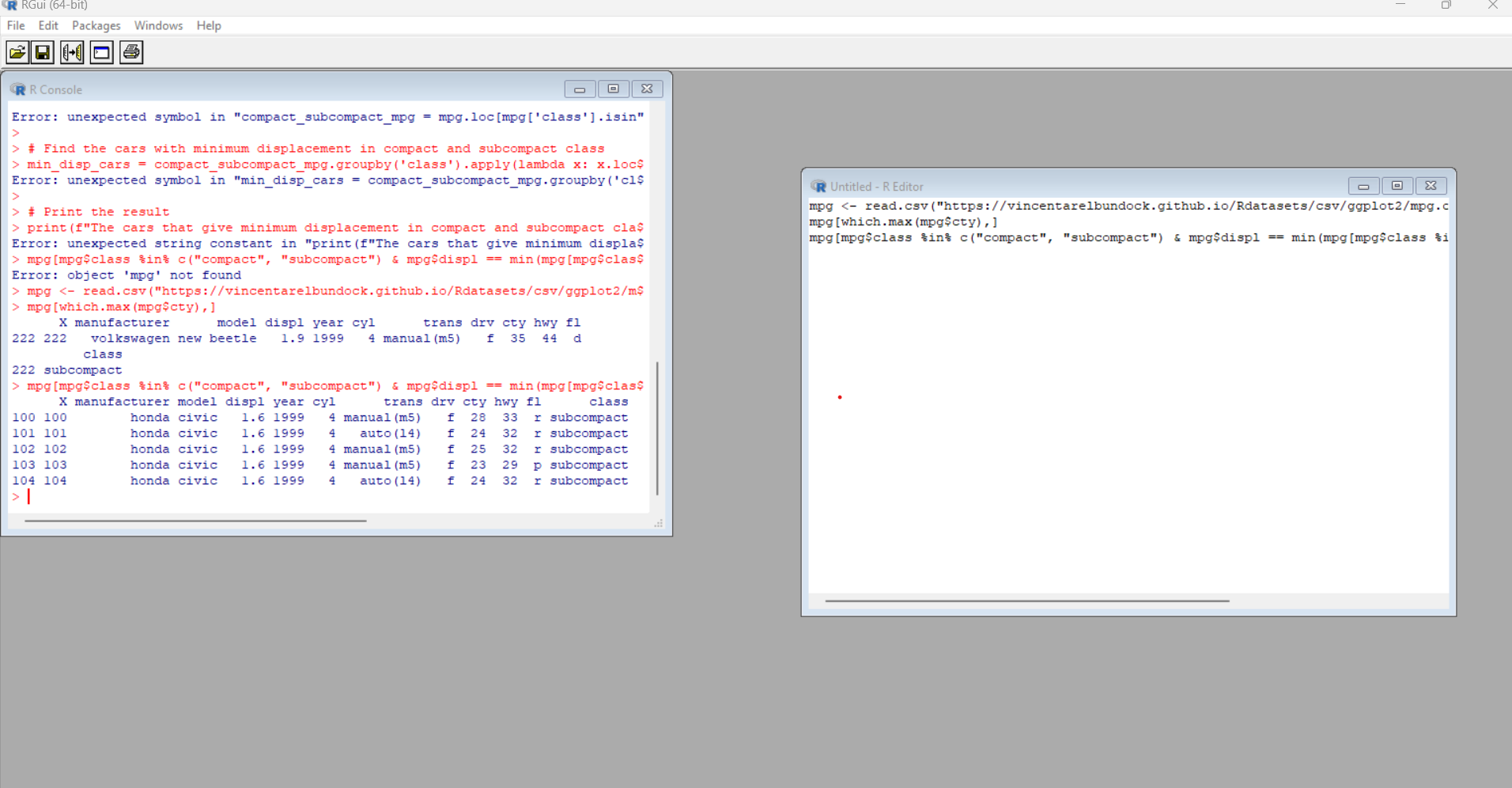
Answer the following queries

1. Find the car which gives maximum city miles per gallon
2. Find the cars which gives minimum disp in compact and subcompact class

mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")

mpg[which.max(mpg$cty),]

mpg[mpg$class %in% c("compact", "subcompact") & mpg$displ == min(mpg[mpg$class %in% c("compact", "subcompact"),]$displ),]



**Exercise: 5**

Use the same dataset as used in Exercise 4 and perform the following queries

1. Find the standard deviation of city milles per gallon
2. Find the variance of highway milles per gallon

library(dplyr)

data(mpg)

# find the standard deviation of city miles per gallon

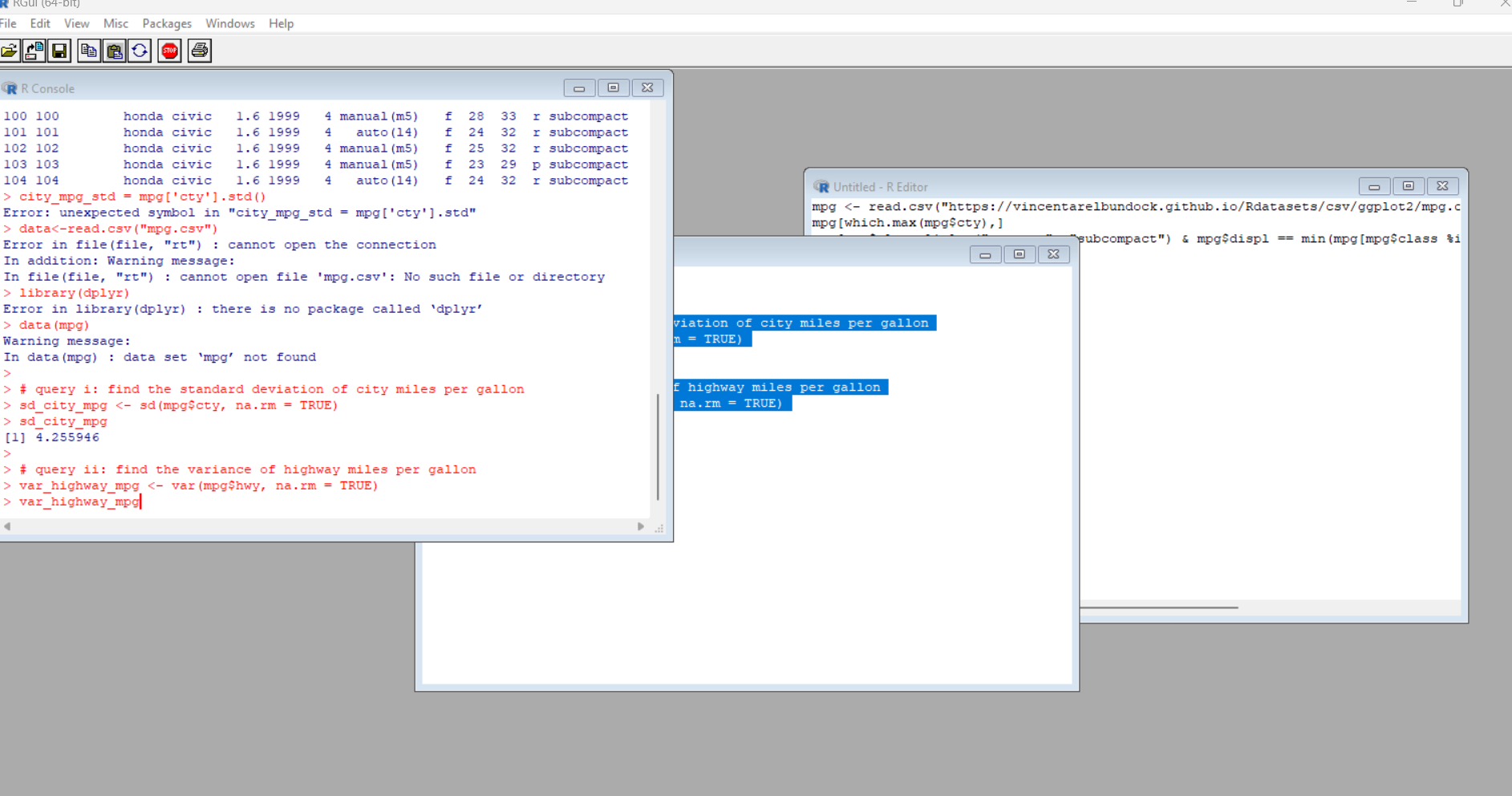
sd\_city\_mpg <- sd(mpg$cty, na.rm = TRUE)

sd\_city\_mpg

# find the variance of highway miles per gallon

var\_highway\_mpg <- var(mpg$hwy, na.rm = TRUE)

var\_highway\_mpg



**Exercise 6**

Use the same dataset and perform the following queries

1. Find the range of the disp in the data set mpg
2. Find the Quartile of the disp in the data set mpg
3. Find the IQR of the disp column in the data set mpg

library(dplyr)

data(mpg)

# query i: find the range of the disp column

disp\_range <- range(mpg$displ, na.rm = TRUE)

disp\_range

# query ii: find the quartiles of the disp column

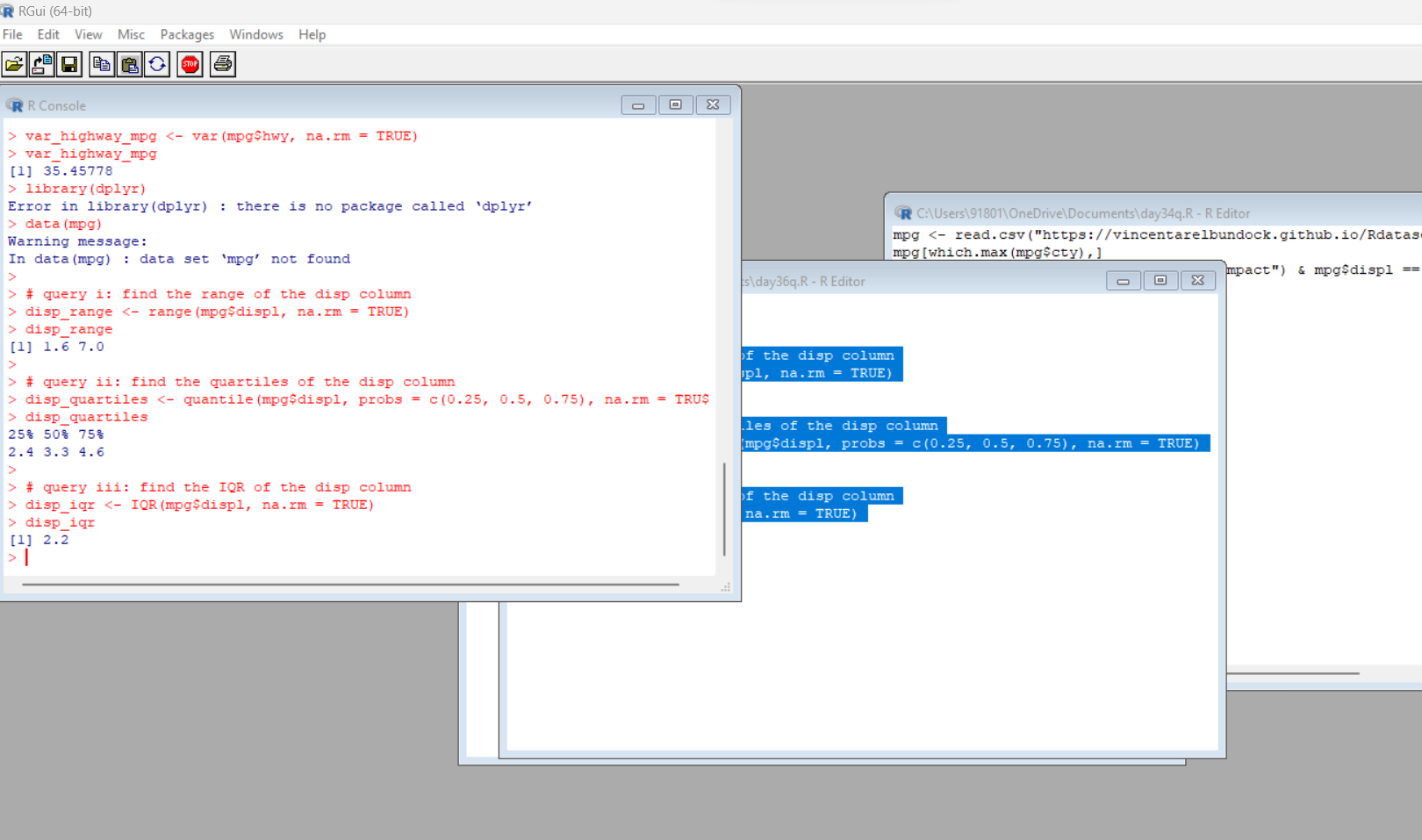
disp\_quartiles <- quantile(mpg$displ, probs = c(0.25, 0.5, 0.75), na.rm = TRUE)

disp\_quartiles

# query iii: find the IQR of the disp column

disp\_iqr <- IQR(mpg$displ, na.rm = TRUE)

disp\_iqr

****

**Exercise 7**

#Install Library

library(e1071)

1. Find the skewness of city miles per mileage in the data set mpg ?

Use qplot function and display the graph for the city miles per mileage column

1. Find the kurtosis of city miles per mileage in the data set mpg

# load the mpg dataset

library(dplyr)

data(mpg)

# query a: find the skewness of city miles per gallon

skew\_city\_mpg <- skewness(mpg$cty, na.rm = TRUE)

skew\_city\_mpg

# plot a histogram of city miles per gallon

library(ggplot2)

qplot(mpg$cty, geom = "histogram", binwidth = 2) +

labs(x = "City miles per gallon", y = "Frequency")

# query b: find the kurtosis of city miles per gallon

kurt\_city\_mpg <- kurtosis(mpg$cty, na.rm = TRUE)

kurt\_city\_mpg

