R PROGRAMMING

DAY 3 LAB MANUAL

INDHUMATHI V

192124215

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

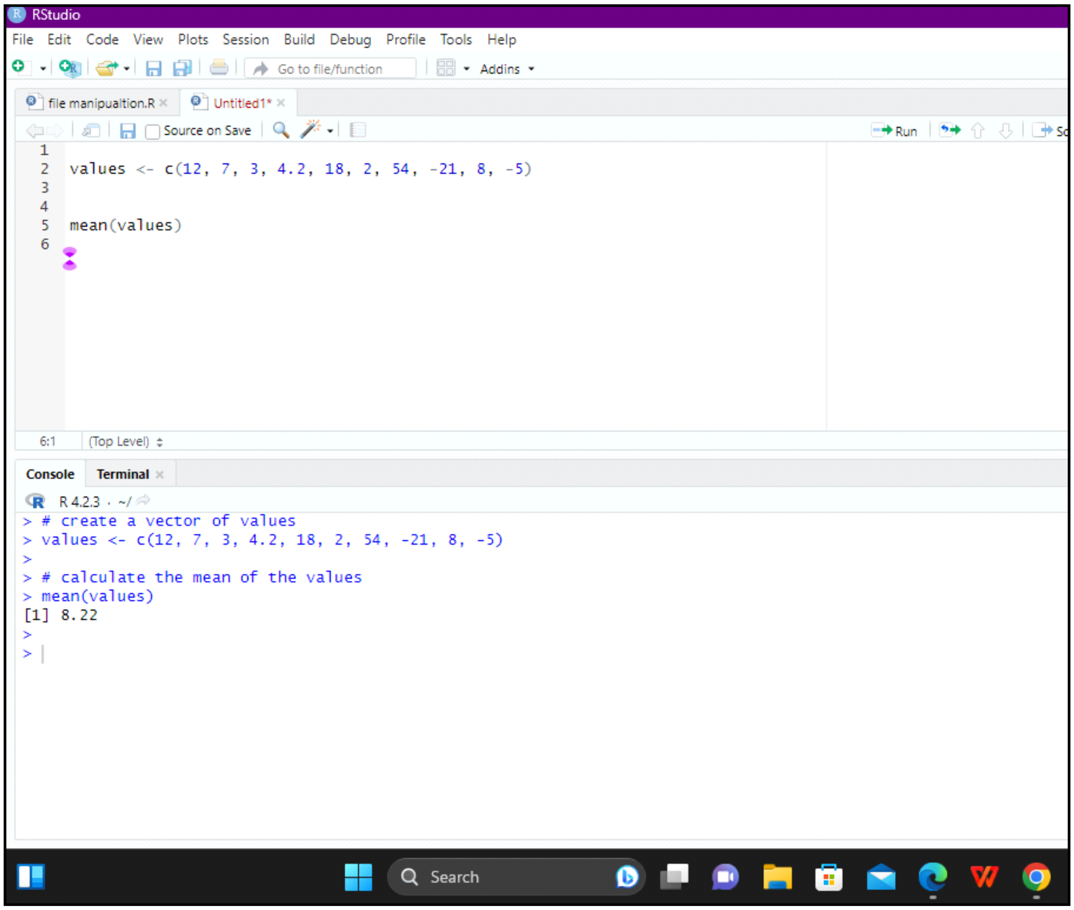
program:

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

mean(values)
```

output:

```
> # create a vector of values

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

>

> # calculate the mean of the values

> mean(values)

[1] 8.22

>
```

SCREENSHOT:



```

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


mean(values)

```

```
> # create a vector of values
> values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
>
> # calculate the mean of the values
> mean(values)
[1] 8.22
>
> |
```

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

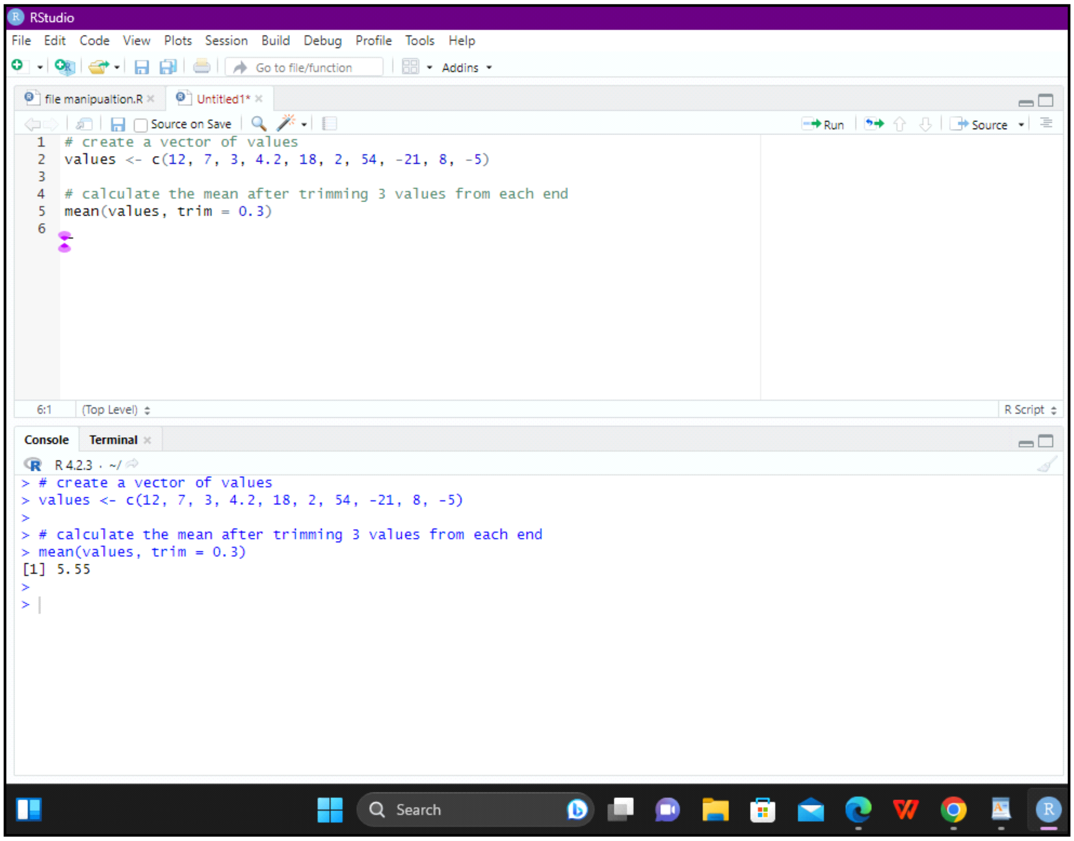
PROGRAM:

```
# create a vector of values
values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

# calculate the mean after trimming 3 values from each end
mean(values, trim = 0.3)
```

OUTPUT:

```
> # create a vector of values
> values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
>
> # calculate the mean after trimming 3 values from each end
> mean(values, trim = 0.3)
[1] 5.55
>
>
```

SCREENSHOT;



```
# create a vector of values
values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

# calculate the mean after trimming 3 values from each end
mean(values, trim = 0.3)
```

```
> # create a vector of values
> values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
>
> # calculate the mean after trimming 3 values from each end
> mean(values, 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)

PROGRAM:

# create a vector of values

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

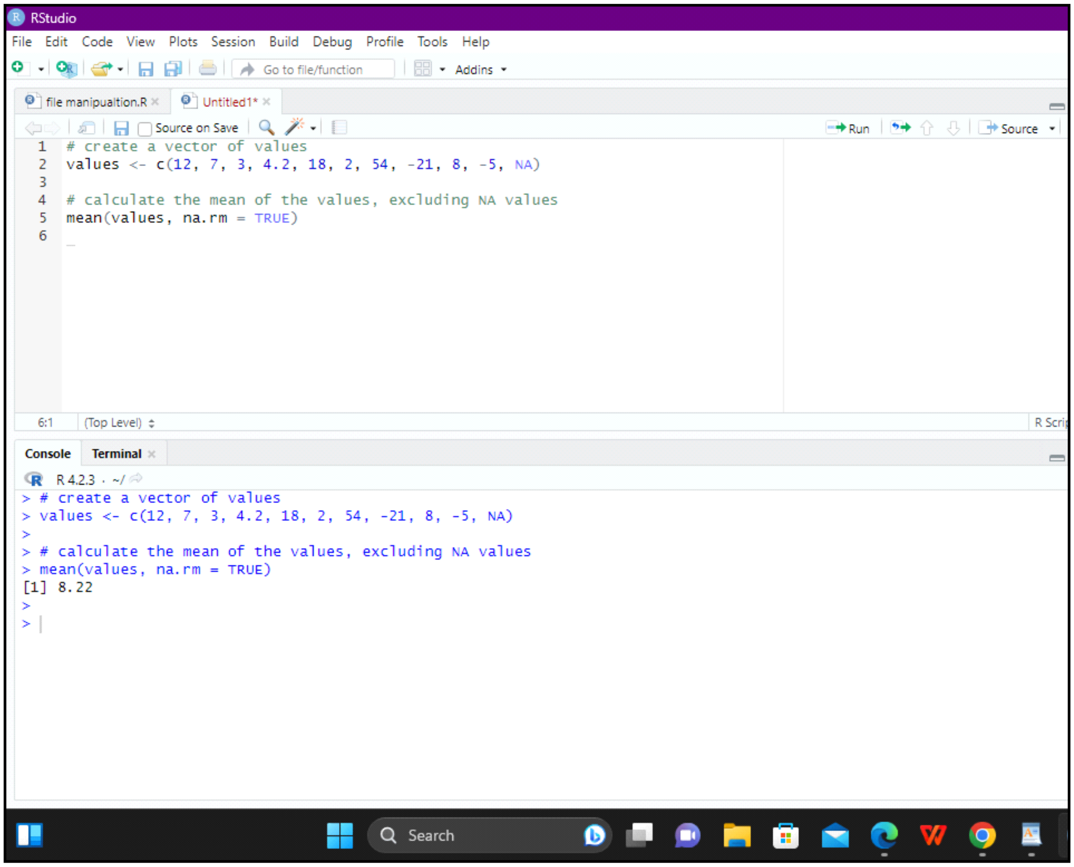
# calculate the mean of the values, excluding NA values

mean(values, na.rm = TRUE)

OUTPUT:

```
> # create a vector of values

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

>

> # calculate the mean of the values, excluding NA values

> mean(values, na.rm = TRUE)

[1] 8.22

>
```

SCREENSHOT:



```
# create a vector of values
values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5, NA)

# calculate the mean of the values, excluding NA values
mean(values, na.rm = TRUE)
```

```
> # create a vector of values
> values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5, NA)
>
> # calculate the mean of the values, excluding NA values
> mean(values, na.rm = TRUE)
[1] 8.22
>
> 
```

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

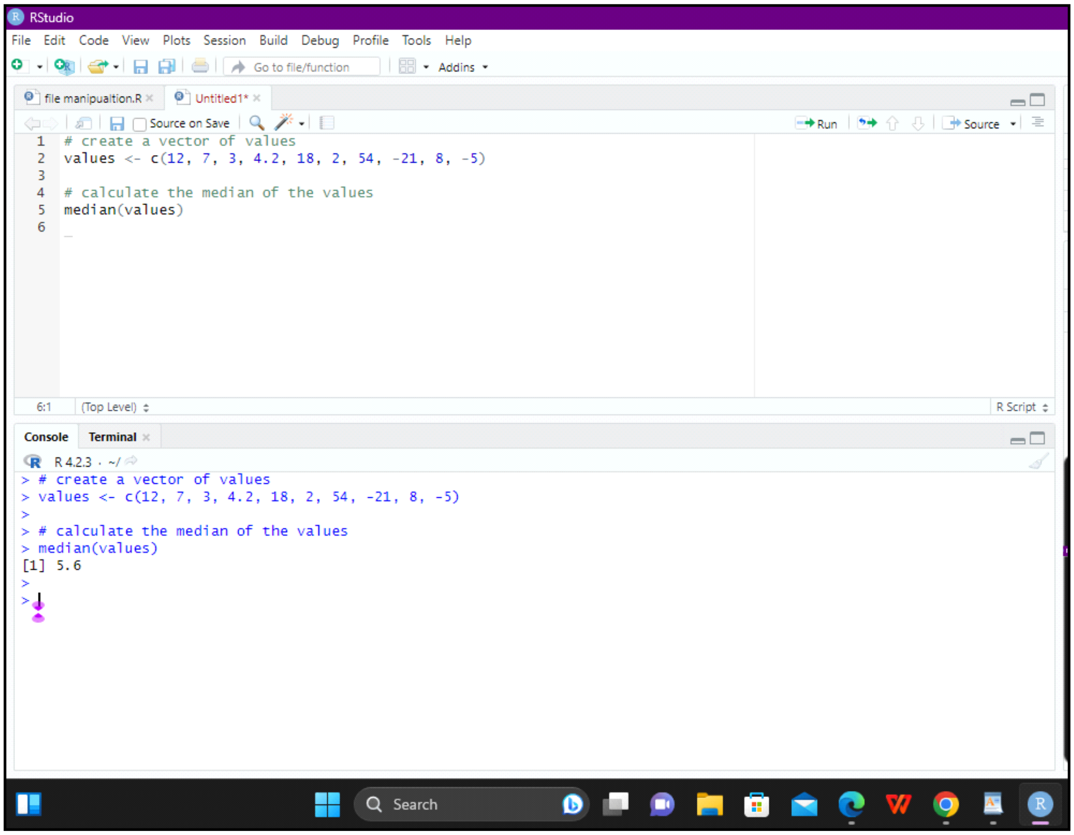
PROGRAM:

```
# create a vector of values
values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

# calculate the median of the values
median(values)
```

OUTPUT:

```
> # create a vector of values
> values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
>
> # calculate the median of the values
> median(values)
[1] 5.6
```

SCREENSHOT:



```
# create a vector of values
values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)

# calculate the median of the values
median(values)
```

```
> # create a vector of values
> values <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
>
> # calculate the median of the values
> 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) , (&quot;o&quot;,&quot;it&quot;,&quot;the&quot;,&quot;it&quot;,&quot;it&quot;)

PROGRAM:

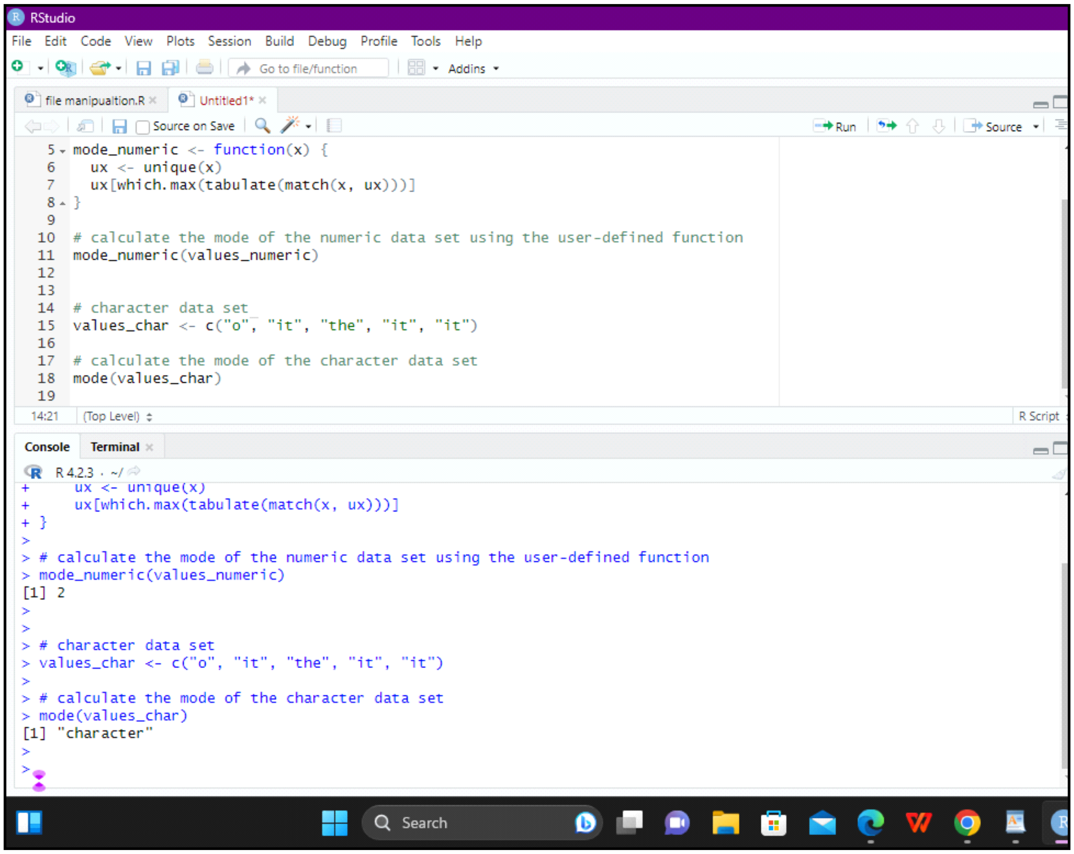
```
# numeric data set

values_numeric <- c(2, 1, 2, 3, 1, 2, 3, 4, 1, 5, 5, 3, 2, 3)


# user-defined function to calculate mode for numeric data set

mode_numeric <- function(x) {

  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

# calculate the mode of the numeric data set using the user-defined function
mode_numeric(values_numeric)



# character data set
values_char <- c("o", "it", "the", "it", "it")

# calculate the mode of the character data set
mode(values_char)
```

OUTPUT:

```
> # numeric data set
> values_numeric <- c(2, 1, 2, 3, 1, 2, 3, 4, 1, 5, 5, 3, 2, 3)
>
> # user-defined function to calculate mode for numeric data set
> mode_numeric <- function(x) {
+     ux <- unique(x)
+     ux[which.max(tabulate(match(x, ux)))]
+ }
>
> # calculate the mode of the numeric data set using the user-defined function
> mode_numeric(values_numeric)
[1] 2
>
```

```
>
> # character data set
> values_char <- c("o", "it", "the", "it", "it")
>
> # calculate the mode of the character data set
> mode(values_char)
[1] "character"
>
>
```

SCREENSHOT:



```
mode_numeric <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

# calculate the mode of the numeric data set using the user-defined function
mode_numeric(values_numeric)


# character data set
values_char <- c("o", "it", "the", "it", "it")

# calculate the mode of the character data set
mode(values_char)
```

```
+     ux <- unique(x)
+     ux[which.max(tabulate(match(x, ux)))]
+ }
>
> # calculate the mode of the numeric data set using the user-defined function
> mode_numeric(values_numeric)
[1] 2
>
>
> # character data set
> values_char <- c("o", "it", "the", "it", "it")
>
> # calculate the mode of the character data set
> mode(values_char)
[1] "character"
>
>
```

UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION

Exercise: 4

i) Find the car which gives maximum city miles per gallon

ii) Find the cars which gives minimum disp in compact and subcompact class

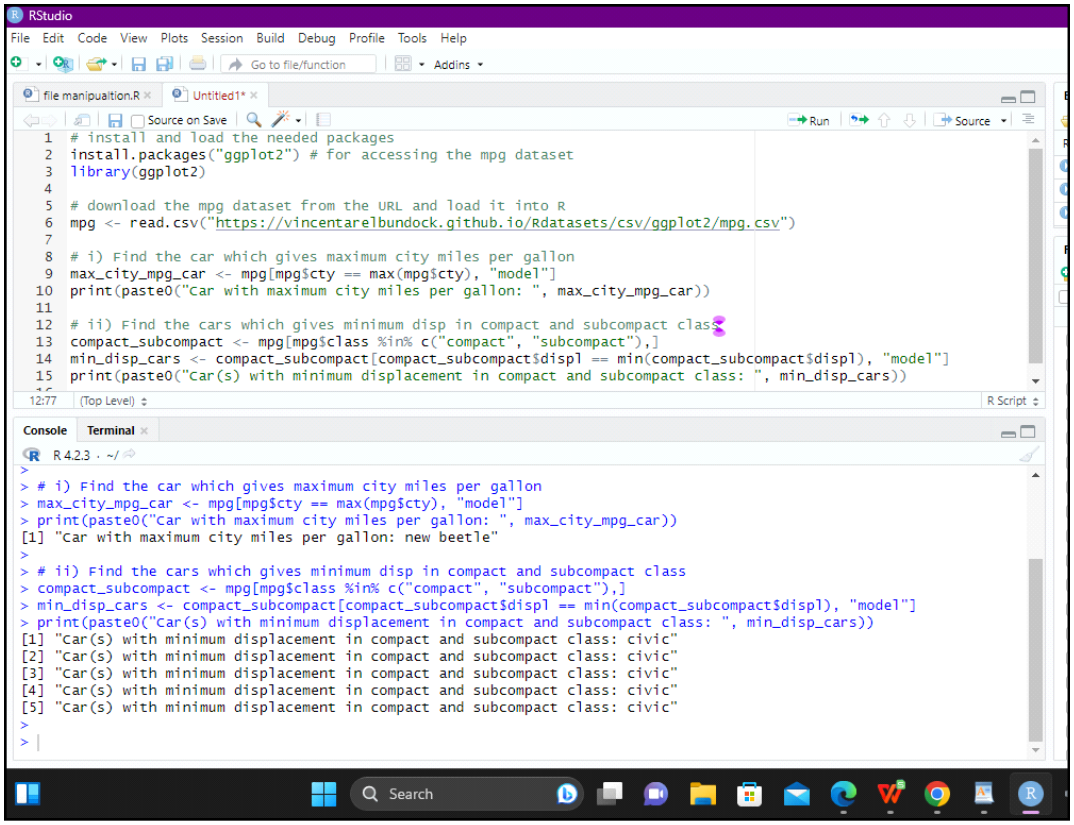
PROGRAM:

```
# install and load the needed packages
install.packages("ggplot2") # for accessing the mpg dataset
library(ggplot2)

# download the mpg dataset from the URL and load it into R
mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")

# i) Find the car which gives maximum city miles per gallon
max_city_mpg_car <- mpg[mpg$cty == max(mpg$cty), "model"]
print(paste0("Car with maximum city miles per gallon: ", max_city_mpg_car))

# ii) Find the cars which gives minimum disp in compact and subcompact class
compact_subcompact <- mpg[mpg$class %in% c("compact", "subcompact"),]
min_disp_cars <- compact_subcompact[compact_subcompact$displ == min(compact_subcompact$displ), "model"]
print(paste0("Car(s) with minimum displacement in compact and subcompact class: ", min_disp_cars))
```

OUTPUT:

```
> # download the mpg dataset from the URL and load it into R
> mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")
>
> # i) Find the car which gives maximum city miles per gallon
> max_city_mpg_car <- mpg[mpg$cty == max(mpg$cty), "model"]
> print(paste0("Car with maximum city miles per gallon: ", max_city_mpg_car))
[1] "Car with maximum city miles per gallon: new beetle"
>
> # ii) Find the cars which gives minimum disp in compact and subcompact class
> compact_subcompact <- mpg[mpg$class %in% c("compact", "subcompact"),]
> min_disp_cars <- compact_subcompact[compact_subcompact$displ ==
min(compact_subcompact$displ), "model"]
> print(paste0("Car(s) with minimum displacement in compact and subcompact class: ",
min_disp_cars))
[1] "Car(s) with minimum displacement in compact and subcompact class: civic"
[2] "Car(s) with minimum displacement in compact and subcompact class: civic"
[3] "Car(s) with minimum displacement in compact and subcompact class: civic"
[4] "Car(s) with minimum displacement in compact and subcompact class: civic"
[5] "Car(s) with minimum displacement in compact and subcompact class: civic"
>
```

SCREENSHOT:



```
# install and load the needed packages
install.packages("ggplot2") # for accessing the mpg dataset
library(ggplot2)

# download the mpg dataset from the URL and load it into R
mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")

# i) Find the car which gives maximum city miles per gallon
max_city_mpg_car <- mpg[mpg$cty == max(mpg$cty), "model"]
print(paste0("Car with maximum city miles per gallon: ", max_city_mpg_car))

# ii) Find the cars which gives minimum disp in compact and subcompact class
compact_subcompact <- mpg[mpg$class %in% c("compact", "subcompact"),]
min_disp_cars <- compact_subcompact[compact_subcompact$displ == min(compact_subcompact$displ), "model"]
print(paste0("Car(s) with minimum displacement in compact and subcompact class: ", min_disp_cars))
```

```
>
> # i) Find the car which gives maximum city miles per gallon
> max_city_mpg_car <- mpg[mpg$cty == max(mpg$cty), "model"]
> print(paste0("Car with maximum city miles per gallon: ", max_city_mpg_car))
[1] "Car with maximum city miles per gallon: new beetle"
>
> # ii) Find the cars which gives minimum disp in compact and subcompact class
> compact_subcompact <- mpg[mpg$class %in% c("compact", "subcompact"),]
> min_disp_cars <- compact_subcompact[compact_subcompact$displ == min(compact_subcompact$displ), "model"]
> print(paste0("Car(s) with minimum displacement in compact and subcompact class: ", min_disp_cars))
[1] "Car(s) with minimum displacement in compact and subcompact class: civic"
[2] "Car(s) with minimum displacement in compact and subcompact class: civic"
[3] "Car(s) with minimum displacement in compact and subcompact class: civic"
[4] "Car(s) with minimum displacement in compact and subcompact class: civic"
[5] "Car(s) with minimum displacement in compact and subcompact class: civic"
>
> |
```

Exercise: 5

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

i) Find the standard deviation of city milles per gallon

ii) Find the variance of highway milles per gallon

PROGRAM:

# load the ggplot2 package

library(ggplot2)

```
# load the mpg dataset

data(mpg)


# i) Find the standard deviation of city miles per gallon

sd_city_mpg <- sd(mpg$cty)

print(paste0("Standard deviation of city miles per gallon: ", sd_city_mpg))


# ii) Find the variance of highway miles per gallon

var_highway_mpg <- var(mpg$hwy)

print(paste0("Variance of highway miles per gallon: ", var_highway_mpg))
```

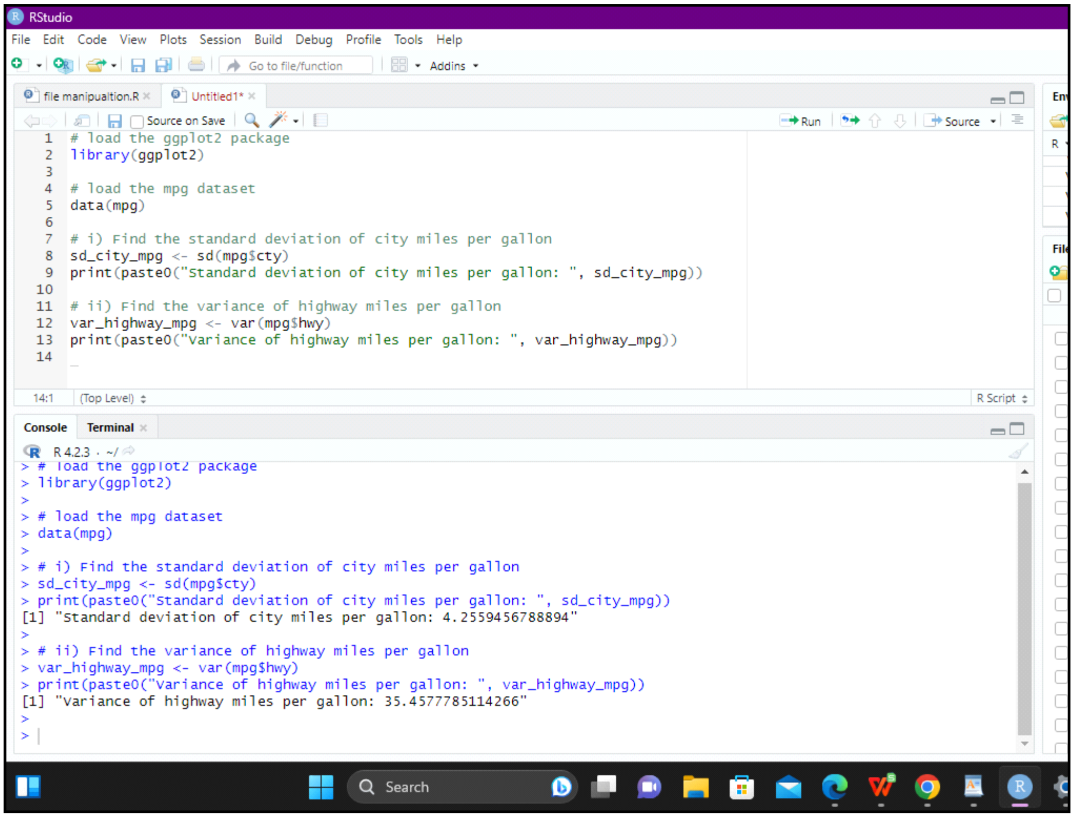
OUTPUT:

```
 # load the ggplot2 package

> library(ggplot2)

>

> # load the mpg dataset

> data(mpg)

>

> # i) Find the standard deviation of city miles per gallon

> sd_city_mpg <- sd(mpg$cty)

> print(paste0("Standard deviation of city miles per gallon: ", sd_city_mpg))

[1] "Standard deviation of city miles per gallon: 4.2559456788894"

>

> # ii) Find the variance of highway miles per gallon

> var_highway_mpg <- var(mpg$hwy)

> print(paste0("Variance of highway miles per gallon: ", var_highway_mpg))

[1] "Variance of highway miles per gallon: 35.4577785114266"
```

>

>

SCREENSHOT:



```
# load the ggplot2 package
library(ggplot2)

# load the mpg dataset
data(mpg)

# i) Find the standard deviation of city miles per gallon
sd_city_mpg <- sd(mpg$cty)
print(paste0("Standard deviation of city miles per gallon: ", sd_city_mpg))

# ii) Find the variance of highway miles per gallon
var_highway_mpg <- var(mpg$hwy)
print(paste0("Variance of highway miles per gallon: ", var_highway_mpg))

```

```
> # load the ggplot2 package
> library(ggplot2)
>
> # load the mpg dataset
> data(mpg)
>
> # i) Find the standard deviation of city miles per gallon
> sd_city_mpg <- sd(mpg$cty)
> print(paste0("Standard deviation of city miles per gallon: ", sd_city_mpg))
[1] "Standard deviation of city miles per gallon: 4.2559456788894"
>
> # ii) Find the variance of highway miles per gallon
> var_highway_mpg <- var(mpg$hwy)
> print(paste0("Variance of highway miles per gallon: ", var_highway_mpg))
[1] "Variance of highway miles per gallon: 35.4577785114266"
>
> |
```

Exercise 6

Use the same dataset and perform the following queries

i) Find the range of the disp in the data set mpg

ii) Find the Quartile of the disp in the data set mpg

iii) Find the IQR of the disp column in the data set mpg

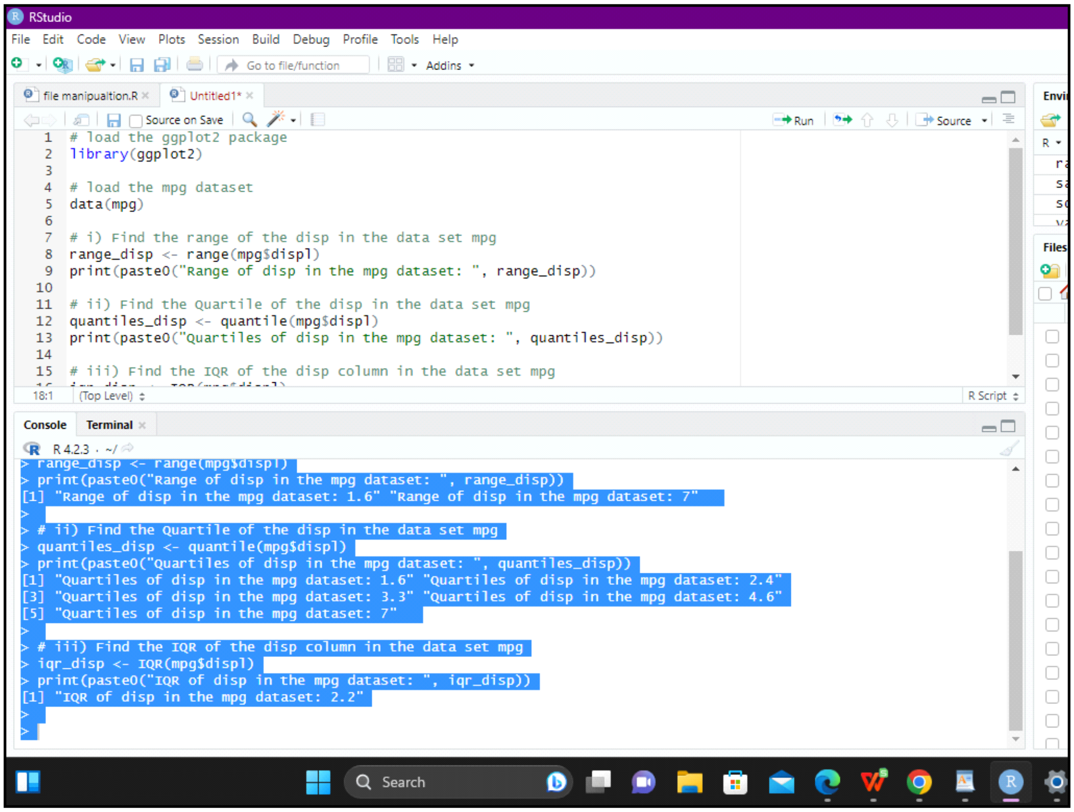
PROGRAM:

```
# load the ggplot2 package
library(ggplot2)

# load the mpg dataset
data(mpg)

# i) Find the range of the disp in the data set mpg
range_disp <- range(mpg$displ)
print(paste0("Range of disp in the mpg dataset: ", range_disp))

# ii) Find the Quartile of the disp in the data set mpg
quantiles_disp <- quantile(mpg$displ)
print(paste0("Quartiles of disp in the mpg dataset: ", quantiles_disp))

# iii) Find the IQR of the disp column in the data set mpg
iqr_disp <- IQR(mpg$displ)
print(paste0("IQR of disp in the mpg dataset: ", iqr_disp))
```

OUTPUT:

```
> # load the ggplot2 package
> library(ggplot2)
>
> # load the mpg dataset
> data(mpg)
>
> # i) Find the range of the disp in the data set mpg
> range_disp <- range(mpg$displ)

> print(paste0("Range of disp in the mpg dataset: ", range_disp))

[1] "Range of disp in the mpg dataset: 1.6" "Range of disp in the mpg dataset: 7"

>

> # ii) Find the Quartile of the disp in the data set mpg

> quantiles_disp <- quantile(mpg$displ)

> print(paste0("Quartiles of disp in the mpg dataset: ", quantiles_disp))

[1] "Quartiles of disp in the mpg dataset: 1.6" "Quartiles of disp in the mpg dataset: 2.4"

[3] "Quartiles of disp in the mpg dataset: 3.3" "Quartiles of disp in the mpg dataset: 4.6"

[5] "Quartiles of disp in the mpg dataset: 7"

>

> # iii) Find the IQR of the disp column in the data set mpg

> iqr_disp <- IQR(mpg$displ)

> print(paste0("IQR of disp in the mpg dataset: ", iqr_disp))

[1] "IQR of disp in the mpg dataset: 2.2"

>

>
```

SCREENSHOT:



```
# load the ggplot2 package
library(ggplot2)

# load the mpg dataset
data(mpg)

# i) Find the range of the disp in the data set mpg
range_disp <- range(mpg$displ)
print(paste0("Range of disp in the mpg dataset: ", range_disp))

# ii) Find the Quartile of the disp in the data set mpg
quantiles_disp <- quantile(mpg$displ)
print(paste0("Quartiles of disp in the mpg dataset: ", quantiles_disp))

# iii) Find the IQR of the disp column in the data set mpg
iqr_disp <- IQR(mpg$displ)
```

```
> range_disp <- range(mpg$displ)
> print(paste0("Range of disp in the mpg dataset: ", range_disp))
[1] "Range of disp in the mpg dataset: 1.6" "Range of disp in the mpg dataset: 7"
>
> # ii) Find the Quartile of the disp in the data set mpg
> quantiles_disp <- quantile(mpg$displ)
> print(paste0("Quartiles of disp in the mpg dataset: ", quantiles_disp))
[1] "Quartiles of disp in the mpg dataset: 1.6" "Quartiles of disp in the mpg dataset: 2.4"
[3] "Quartiles of disp in the mpg dataset: 3.3" "Quartiles of disp in the mpg dataset: 4.6"
[5] "Quartiles of disp in the mpg dataset: 7"
>
> # iii) Find the IQR of the disp column in the data set mpg
> iqr_disp <- IQR(mpg$displ)
> print(paste0("IQR of disp in the mpg dataset: ", iqr_disp))
[1] "IQR of disp in the mpg dataset: 2.2"
>
>
```

Exercise 7

#Install Library

library(e1071)

a. 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

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

PROGRAM:

# load the ggplot2 package

library(ggplot2)

# load the mpg dataset

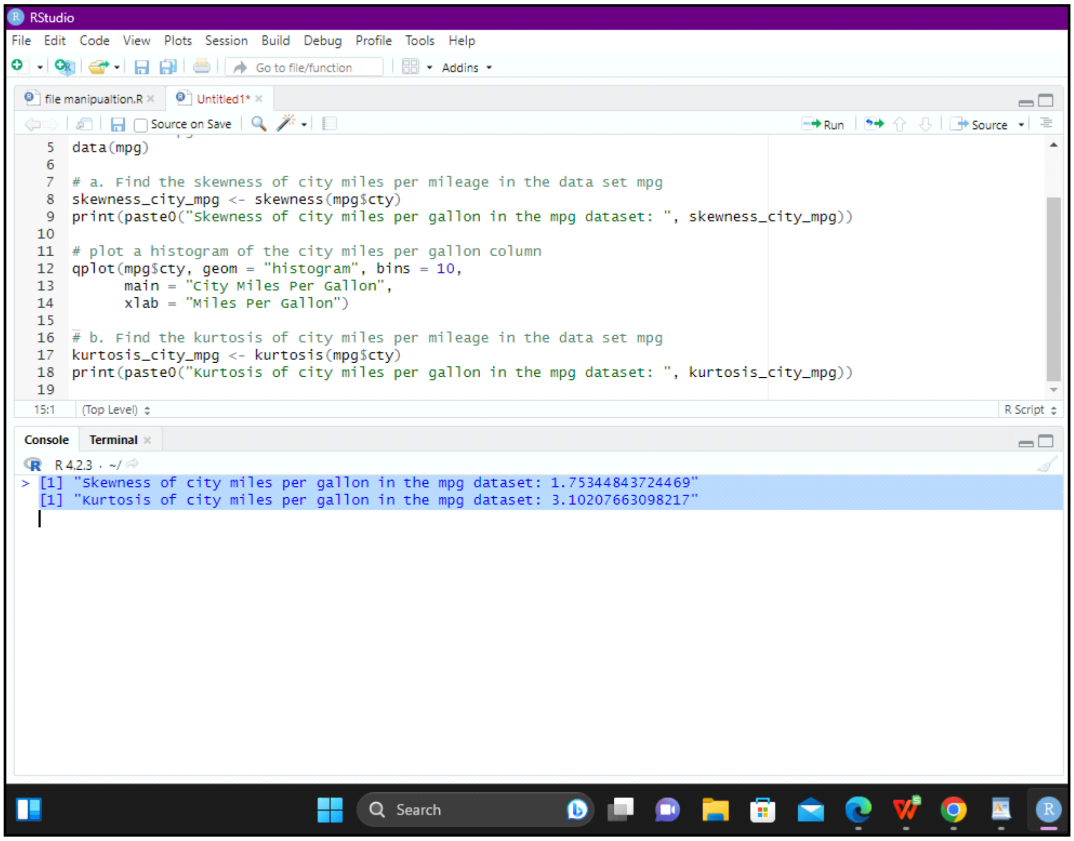
data(mpg)

```
# a. Find the skewness of city miles per mileage in the data set mpg

skewness_city_mpg <- skewness(mpg$cty)

print(paste0("Skewness of city miles per gallon in the mpg dataset: ",
skewness_city_mpg))


# plot a histogram of the city miles per gallon column

qplot(mpg$cty, geom = "histogram", bins = 10,

      main = "City Miles Per Gallon",

      xlab = "Miles Per Gallon")


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

kurtosis_city_mpg <- kurtosis(mpg$cty)

print(paste0("Kurtosis of city miles per gallon in the mpg dataset: ",
kurtosis_city_mpg))
```

OUTPUT:

```
[1] "Skewness of city miles per gallon in the mpg dataset:
1.75344843724469"

[1] "Kurtosis of city miles per gallon in the mpg dataset:
3.10207663098217"
```

SCREENSHOT



```
data(mpg)

# a. Find the skewness of city miles per mileage in the data set mpg
skewness_city_mpg <- skewness(mpg$cty)
print(paste0("Skewness of city miles per gallon in the mpg dataset: ", skewness_city_mpg))

# plot a histogram of the city miles per gallon column
qplot(mpg$cty, geom = "histogram", bins = 10,
      main = "City Miles Per Gallon",
      xlab = "Miles Per Gallon")

# b. Find the kurtosis of city miles per mileage in the data set mpg
kurtosis_city_mpg <- kurtosis(mpg$cty)
print(paste0("Kurtosis of city miles per gallon in the mpg dataset: ", kurtosis_city_mpg))

```

```
> [1] "Skewness of city miles per gallon in the mpg dataset: 1.75344843724469"
  [1] "Kurtosis of city miles per gallon in the mpg dataset: 3.10207663098217"
```