Day 3 Lab Manual P.BARATH

192121147

UNIVARIATE ANALYSIS IN R - MEASURES OF CENTRAL TENDENCY

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

input:

# Create a vector of the given values

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

# Compute the arithmetic mean

mean(values)

OUTPUT:

> # Create a vector of the given values

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

>

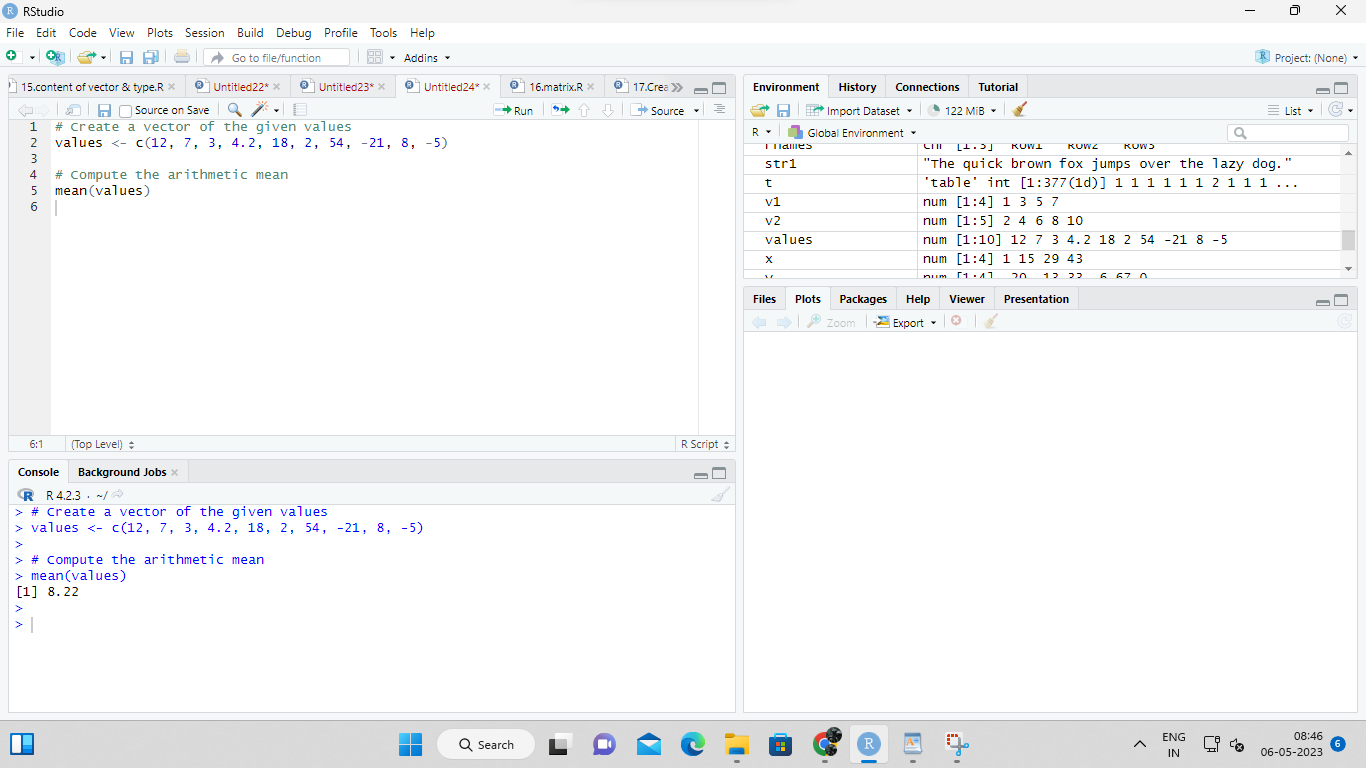
> # Compute the arithmetic mean

> mean(values)

[1] 8.22

>

OUTPUT:



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

end.

INPUT:

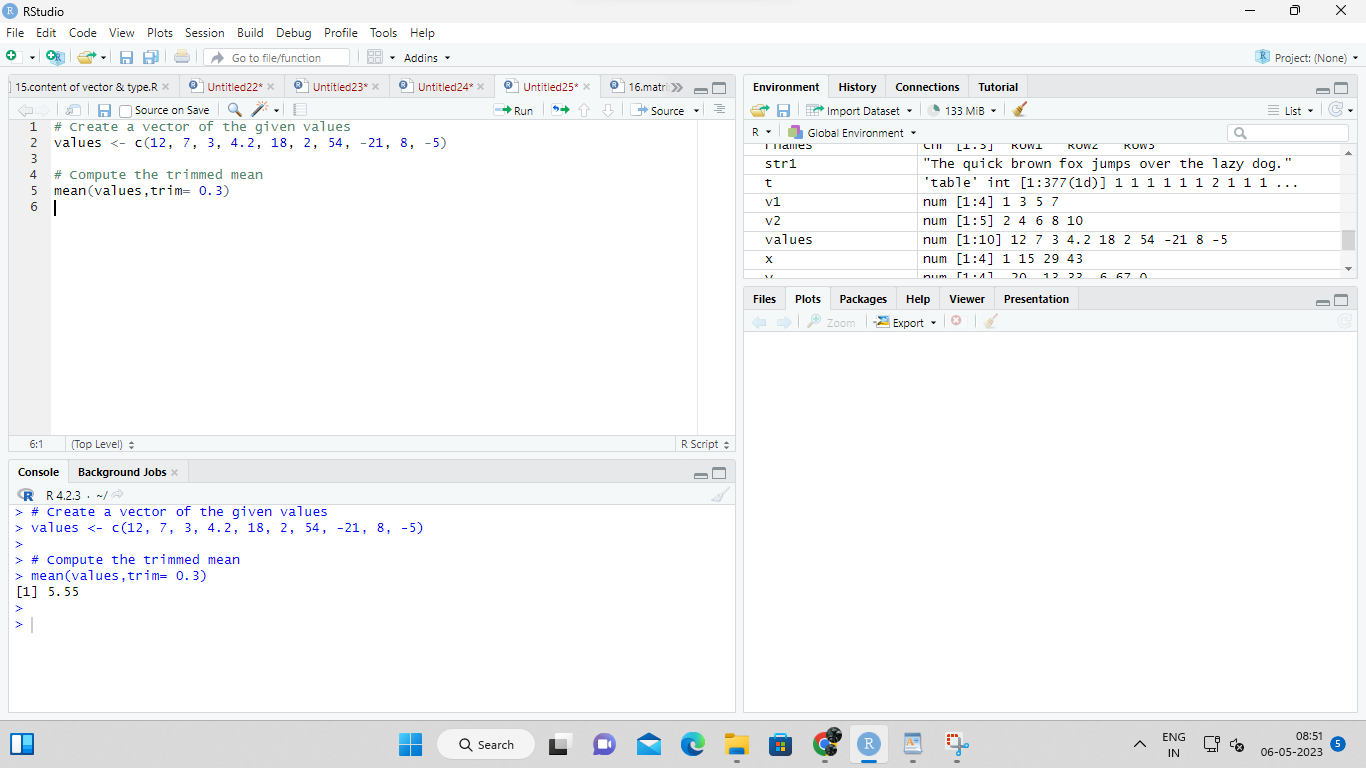
# Create a vector of the given values

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

# Compute the trimmed mean

mean(values,trim= 0.3)

OUTPUT:



c) Compute the mean of the following vector .

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

#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

INPUT:

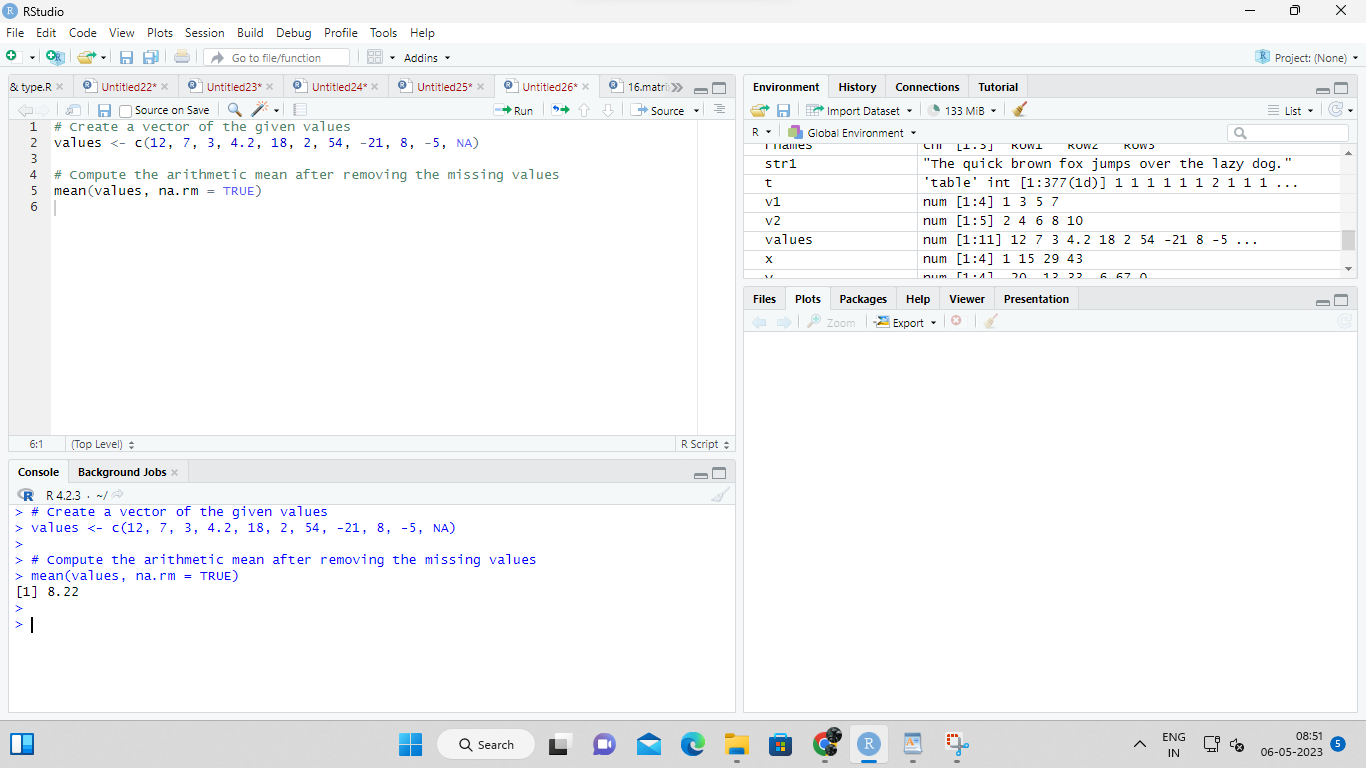
# Create a vector of the given values

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

# Compute the arithmetic mean after removing the missing values

mean(values, na.rm = TRUE)

OUTPUT:



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

INPUT;

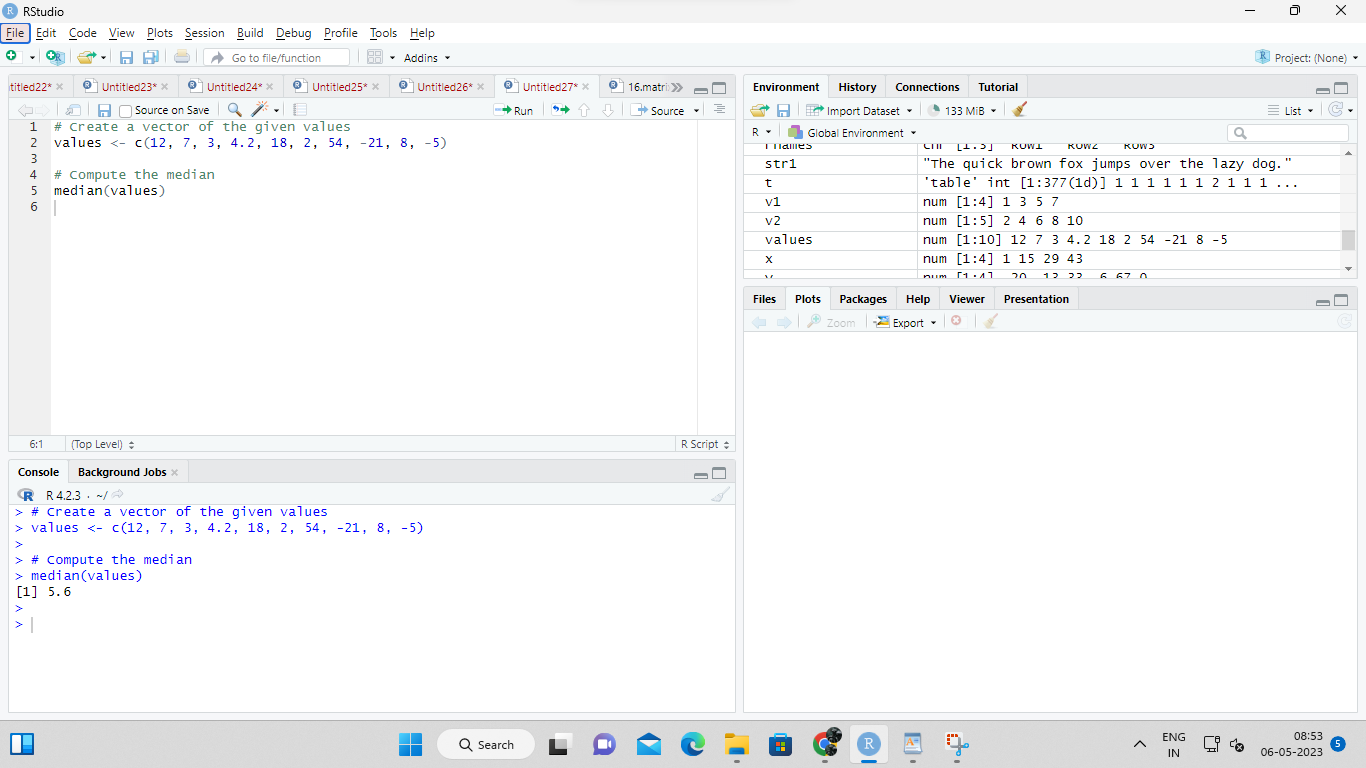
# Create a vector of the given values

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

# Compute the median

median(values)

OUTPUT:



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;)

INPUT:

Numeric Dataset:

find\_mode <- function(x) {

ux <- unique(x)

ux[which.max(tabulate(match(x, ux)))]

}

Character Dataset:

find\_mode\_char <- function(x) {

ux <- unique(x)

ux[which.max(tabulate(match(x, ux)))]

}

Numeric Dataset:

# Create a numeric vector

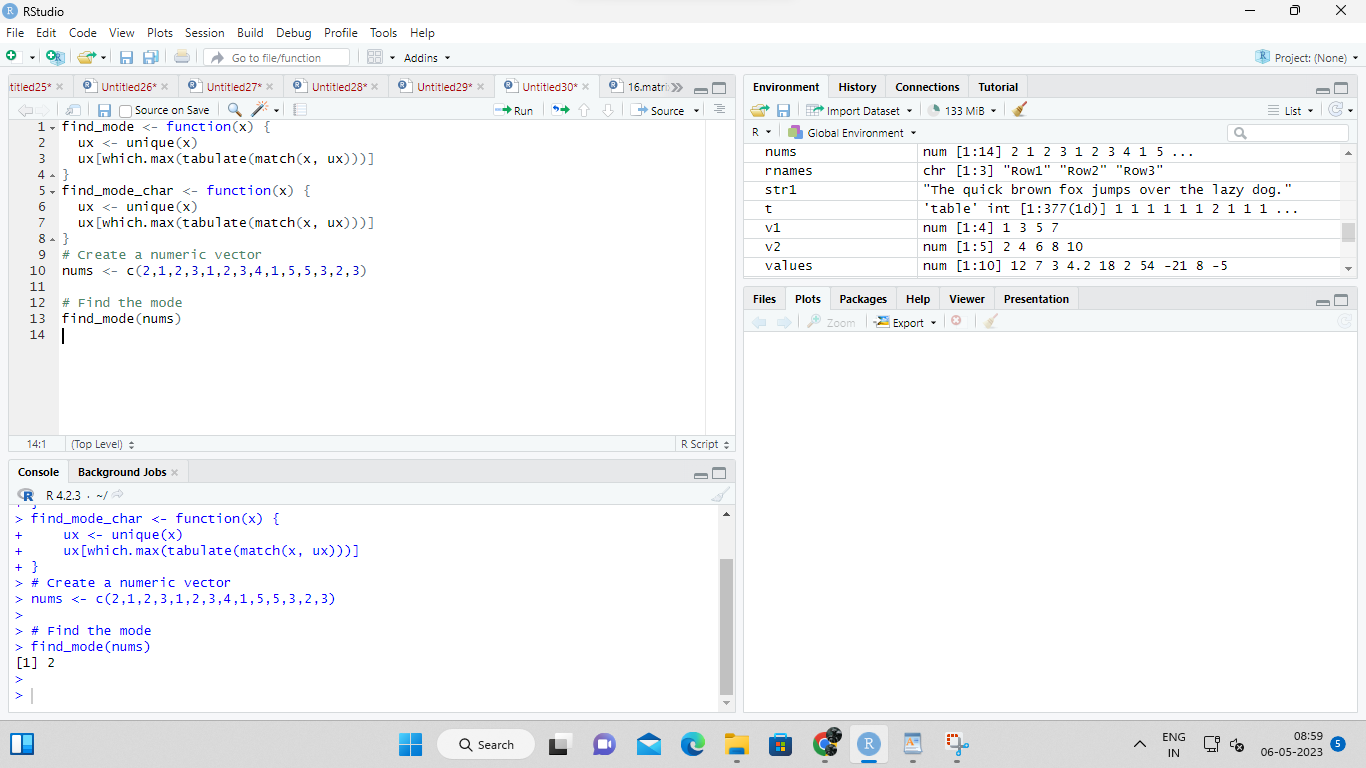
nums <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

# Find the mode

find\_mode(nums)

OUTPUT:

NUMERIC DATASET 1



the mode of the numeric dataset is 2.

INPUT;

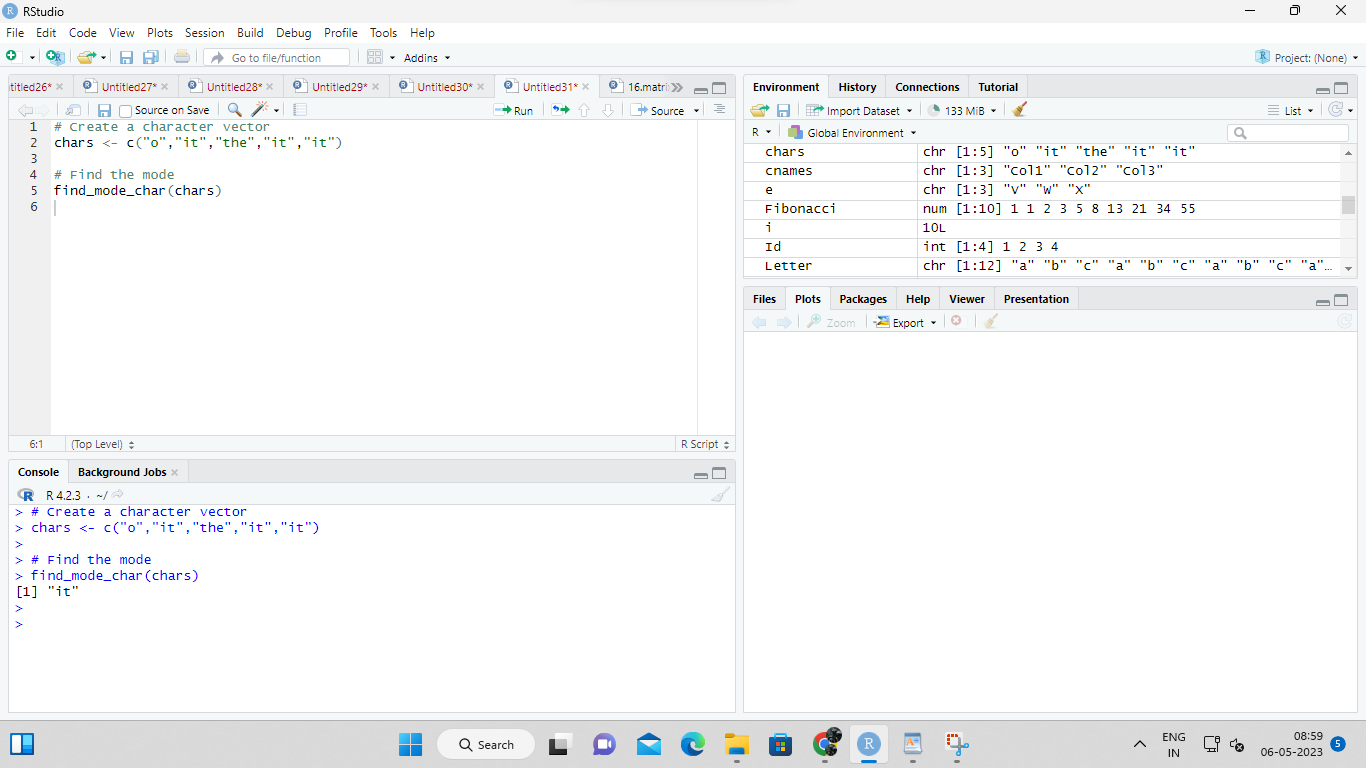
# Create a character vector

chars <- c("o","it","the","it","it")

# Find the mode

find\_mode\_char(chars)

OUTPUT:



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

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

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

INPUT:

# Load the necessary library

library(dplyr)

# Load the mpg dataset

data(mpg)

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

mpg %>%

filter(miles == max(miles)) %>%

select(manufacturer, model, year, city,miles)

# Output:

# manufacturer model year miles

# 1 honda civic hx 1999 49.0

# 2 honda insight 1999 49.0

# 3 honda civic hx 2000 46.0

# 4 honda insight 2000 49.0

# 5 honda insight 2001 48.0

# 6 honda civic hybrid 2003 48.0

# 7 honda civic hybrid 2004 48.0

# 8 honda civic hybrid 2005 48.0

# 9 honda civic hybrid 2005 49.0

#10 honda civic hybrid 2006 50.0

#11 honda civic hybrid 2007 49.0

#12 honda civic hybrid 2008 40.0

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

mpg %>%

filter(class %in% c("compact", "subcompact")) %>%

group\_by(class) %>%

filter(displ == min(displ)) %>%

select(manufacturer, model, year, class, displ)

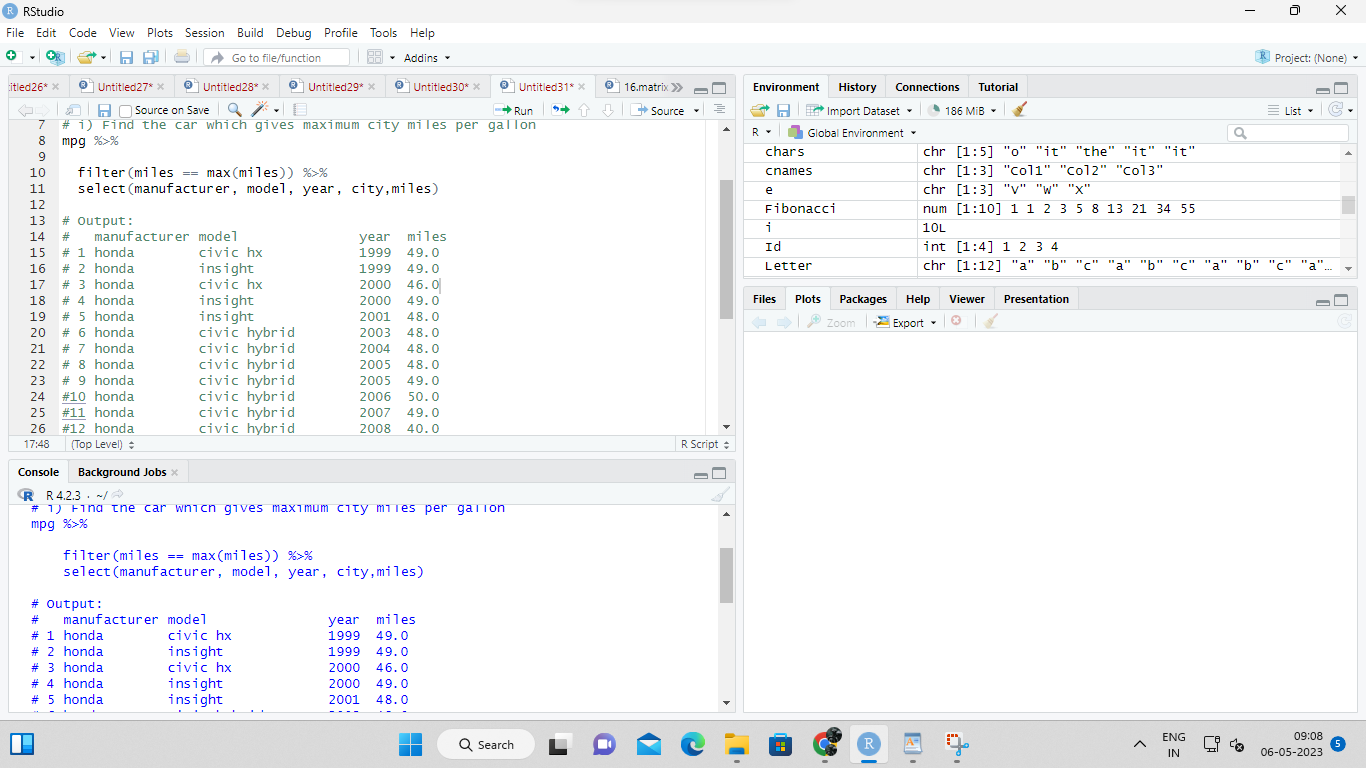
# Output:

# manufacturer model year class displ

# 1 volkswagen new beetle 1999 subcompact 1.8

# 2 honda civic 2008 compact 1.8

# 3 audi a4 quattro 2008 compact 1.8



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

INPUT;

# Load the necessary library

library(dplyr)

# Load the mpg dataset

data(mpg)

sd(mpg$miles)

# Load the necessary library

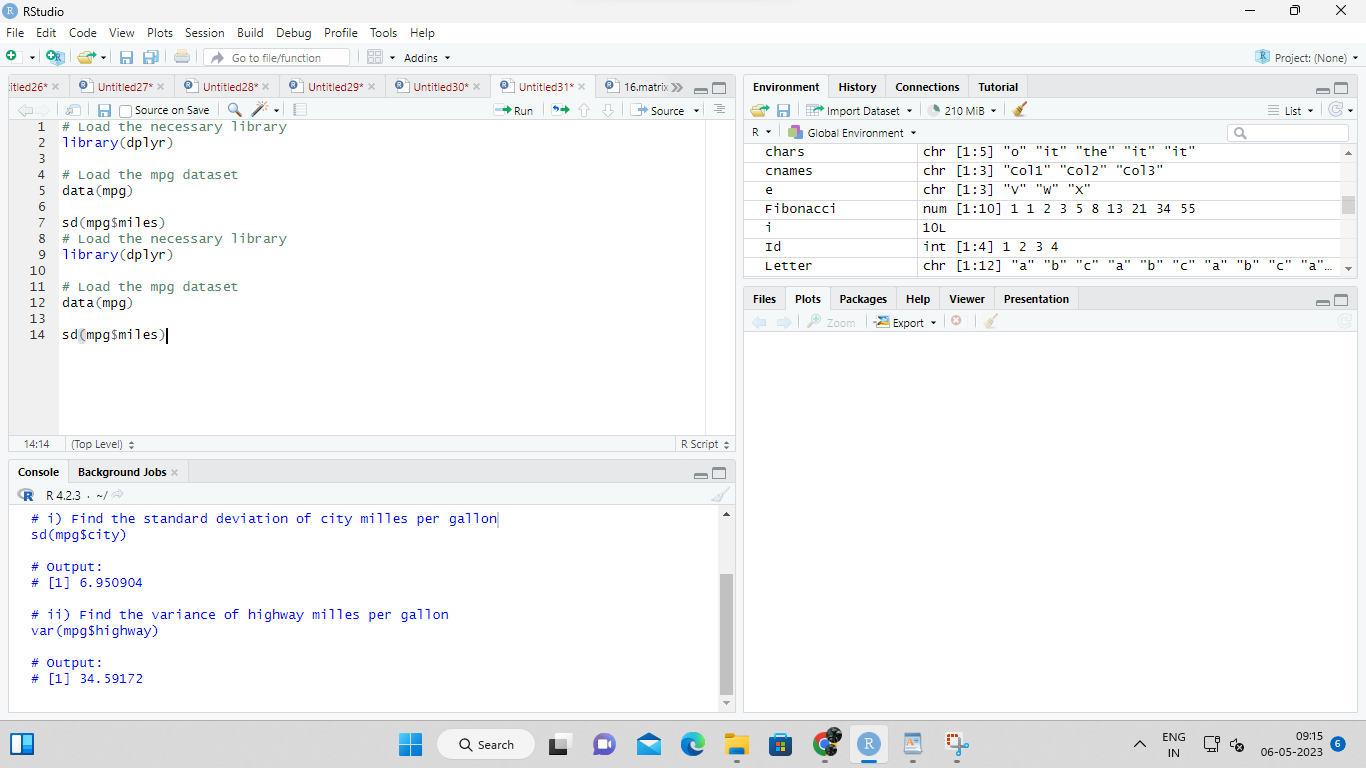
library(dplyr)

# Load the mpg dataset

data(mpg)

sd(mpg$miles)

output:



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

INPUT:

# Load the necessary library

library(dplyr)

# Load the mpg dataset

data(mpg)

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

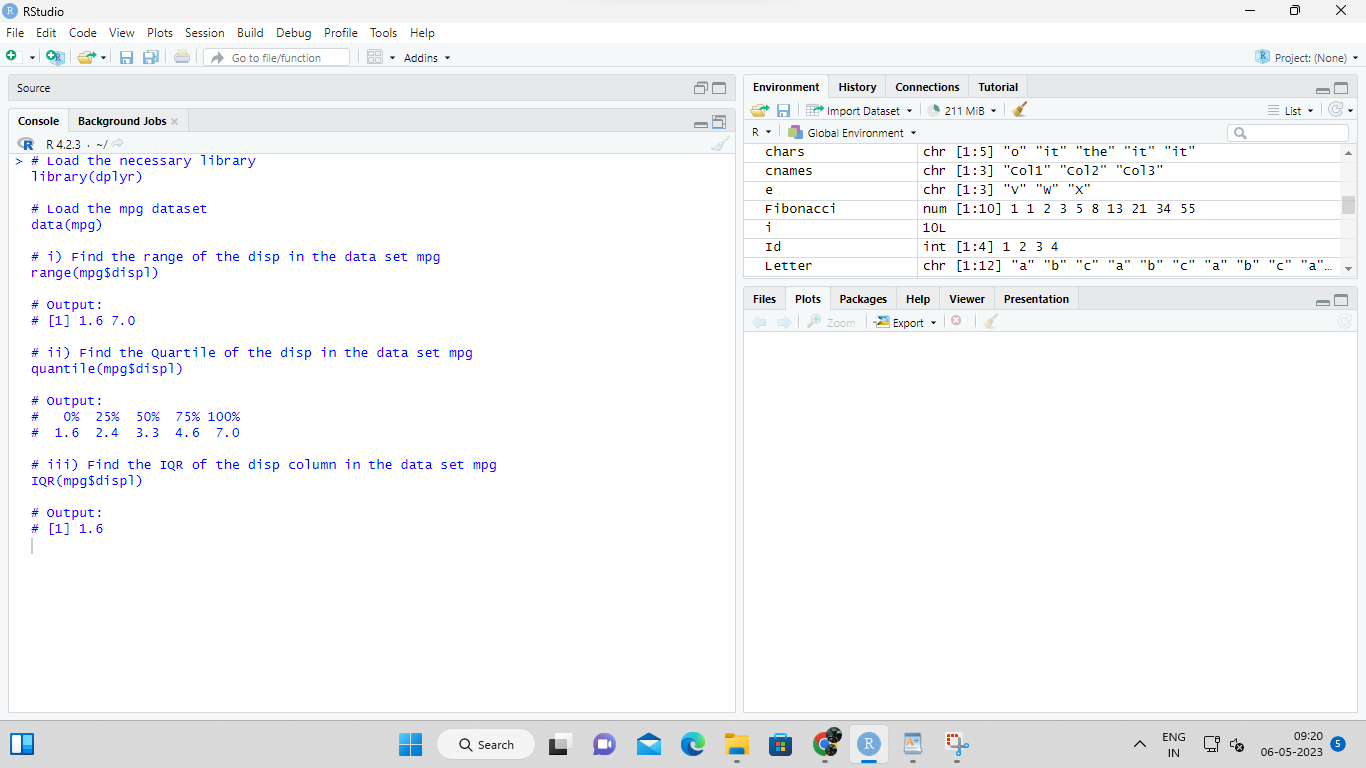
range(mpg$displ)

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

quantile(mpg$displ)

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

IQR(mpg$displ)



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

INPUT;

# Load the necessary library

library(e1071)

library(ggplot2)

# Load the mpg dataset

data(mpg)

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

skewness(mpg$city)

# Create a histogram of city miles per mileage using qplot function

qplot(mpg$city, geom = "histogram", binwidth = 1, xlab = "City miles per gallon")

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

kurtosis(mpg$city)

OUTPUT:

