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

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

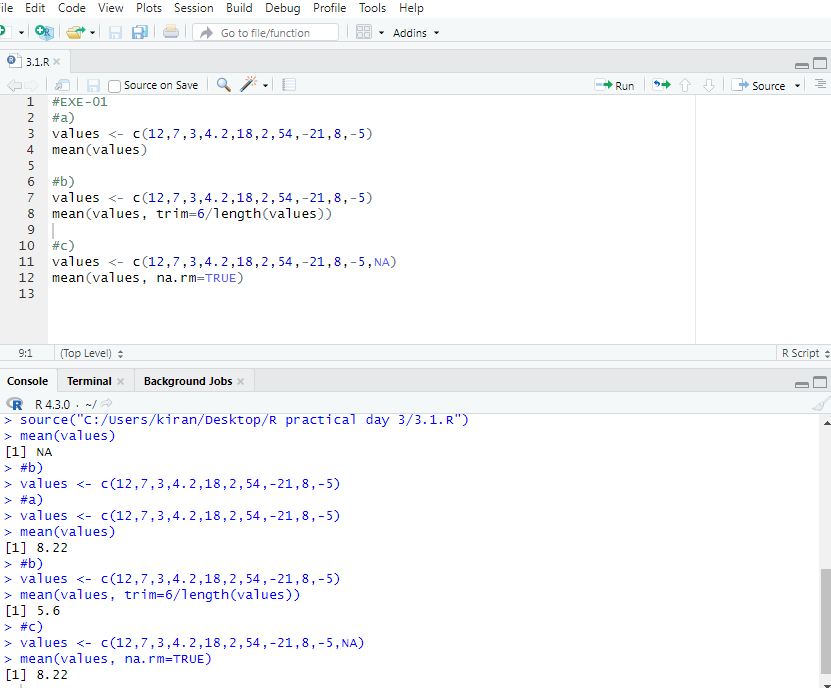
end.

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.



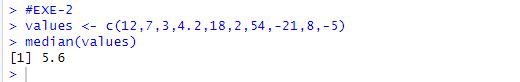
#To drop the missing values from the calculation use na.rm = TRUE

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

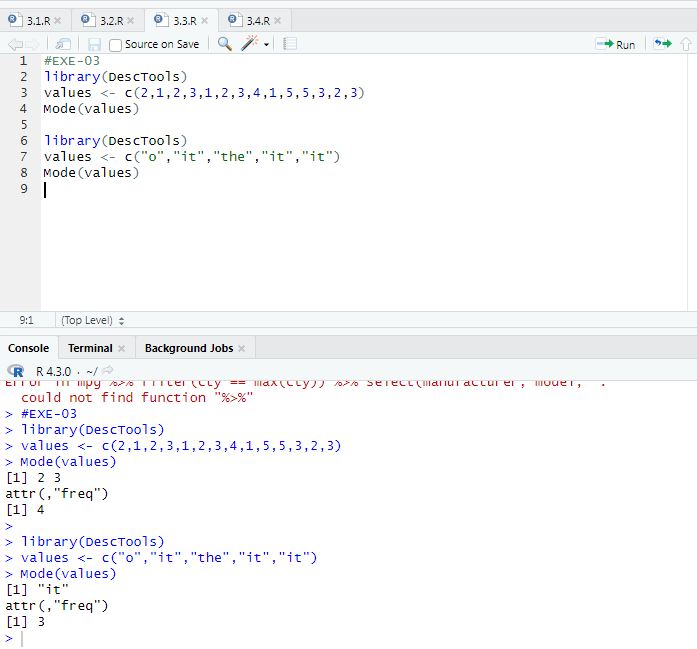




III. MODE

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")



UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION

Exercise: 4

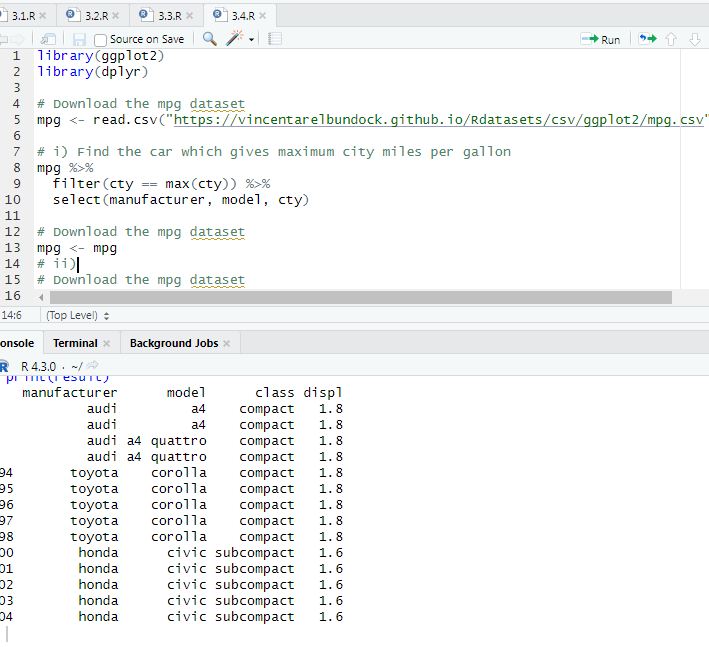
Download mpg dataset which contains Fuel economy data from 1999 and 2008 for 38 popular models of car from the URL given below.

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

Answer the following queries

Find the car which gives maximum city miles per gallon

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

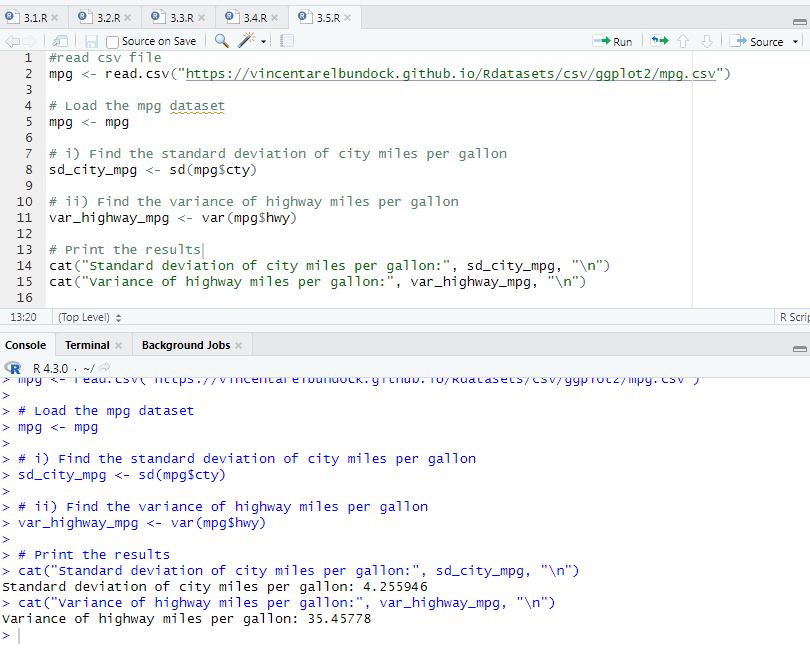


Exercise: 5

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

Find the standard deviation of city milles per gallon

Find the variance of highway milles per gallon



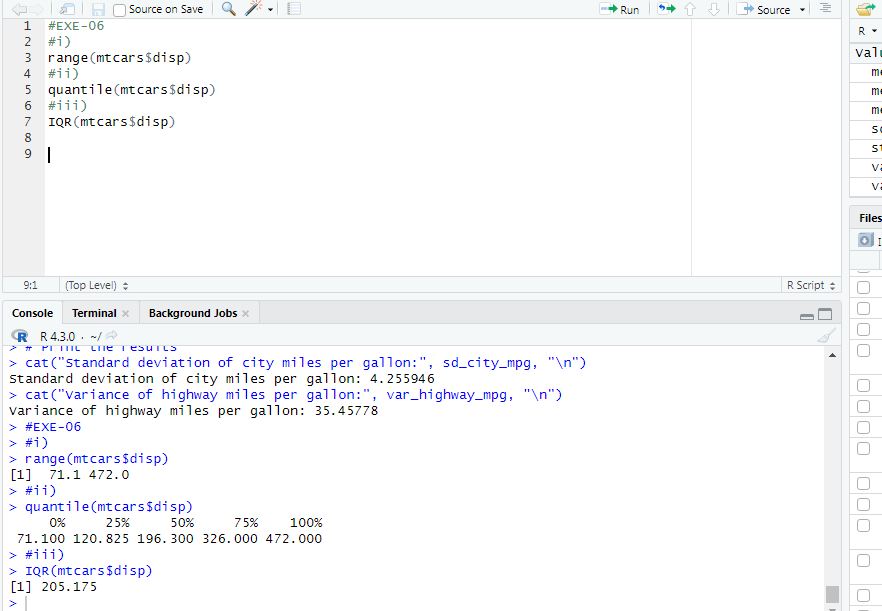
Exercise 6

Use the same dataset and perform the following queries

Find the range of the disp in the data set mpg

Find the Quartile of the disp in the data set mpg

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



Exercise 7

#Install Library

library(e1071)

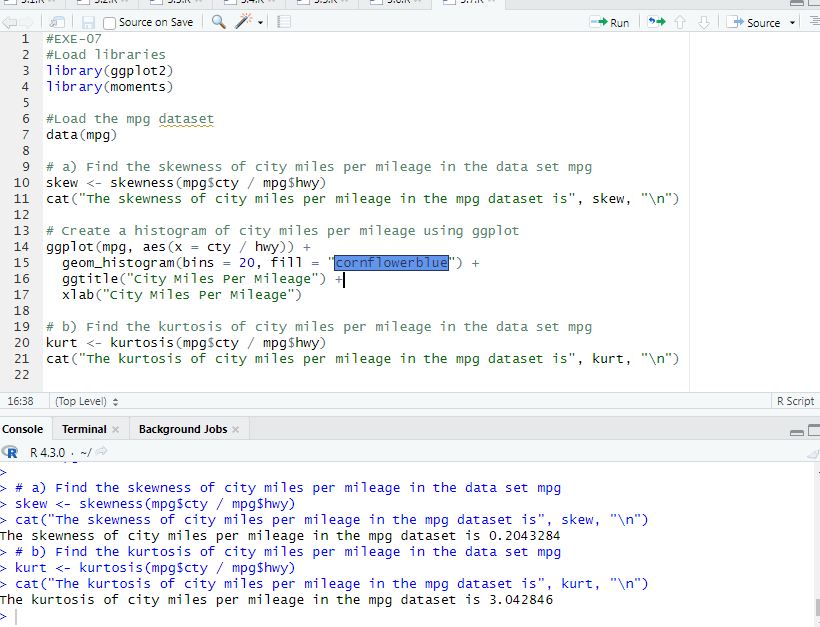
Find the skewness of city miles per mileage in the data set mpg ?

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

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

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

BIVARIATEANALYSIS IN R -COVARIANCE,CORRELATION,CROSSTAB



Exercise: 8

Reference Status Gender TestNewOrFollowUp

1 KRXH Accepted Female Test1 New

2 KRPT Accepted Male Test1 New

3 FHRA Rejected Male Test2 New

4 CZKK Accepted Female Test3 New

5 CQTN Rejected Female Test1 New

6 PZXW Accepted Female Test4 Follow-up

7 SZRZ Rejected Male Test4 New

8 RMZE Rejected Female Test2 New

9 STNX Accepted Female Test3 New

10 TMDW Accepted Female Test1 New

Load the dataset and Create a data frame and name it as dataframe1

Load the function for crosstab

xtabs(~colname , data=Data frame name )

data <- matrix(c("KRXH", "Accepted", "Female", "Test1", "New",

+ "KRPT", "Accepted", "Male", "Test1", "New",

+ "FHRA", "Rejected", "Male", "Test2", "New",

+ "CZKK", "Accepted", "Female", "Test3", "New",

+ "CQTN", "Rejected", "Female", "Test1", "New",

+ "PZXW", "Accepted", "Female", "Test4", "Follow-up",

+ "SZRZ", "Rejected", "Male", "Test4", "New",

+ "RMZE", "Rejected", "Female", "Test2", "New",

+ "STNX", "Accepted", "Female", "Test3", "New",

+ "TMDW", "Accepted", "Female", "Test1", "New"), ncol=5, byrow=TRUE)

>

>

> dataframe1 <- data.frame(Reference=data[,1], Status=data[,2], Gender=data[,3], TestNewOrFollowUp=data[,5])

>

> print(dataframe1)

Reference Status Gender TestNewOrFollowUp

1 KRXH Accepted Female New

2 KRPT Accepted Male New

3 FHRA Rejected Male New

4 CZKK Accepted Female New

5 CQTN Rejected Female New

6 PZXW Accepted Female Follow-up

7 SZRZ Rejected Male New

8 RMZE Rejected Female New

9 STNX Accepted Female New

10 TMDW Accepted Female New

> library(stats)

> status\_gender\_table <- xtabs(~ status + gender, data = dataframe1)

Error in eval(predvars, data, env) : object 'status' not found

> status\_gender\_table <- xtabs(~ Status + Gender, data = dataframe1)

> print(status\_gender\_table)

Gender

Status Female Male

Accepted 5 1

Rejected 2 2

> dataframe2 <- xtabs(~Reference+Status, data=dataframe1)

> print(dataframe2)

Status

Reference Accepted Rejected

CQTN 0 1

CZKK 1 0

FHRA 0 1

KRPT 1 0

KRXH 1 0

PZXW 1 0

RMZE 0 1

STNX 1 0

SZRZ 0 1

TMDW 1 0

> table <- xtabs(~Reference+Status, data=dataframe1)

> print(table)

Status

Reference Accepted Rejected

CQTN 0 1

CZKK 1 0

FHRA 0 1

KRPT 1 0

KRXH 1 0

PZXW 1 0

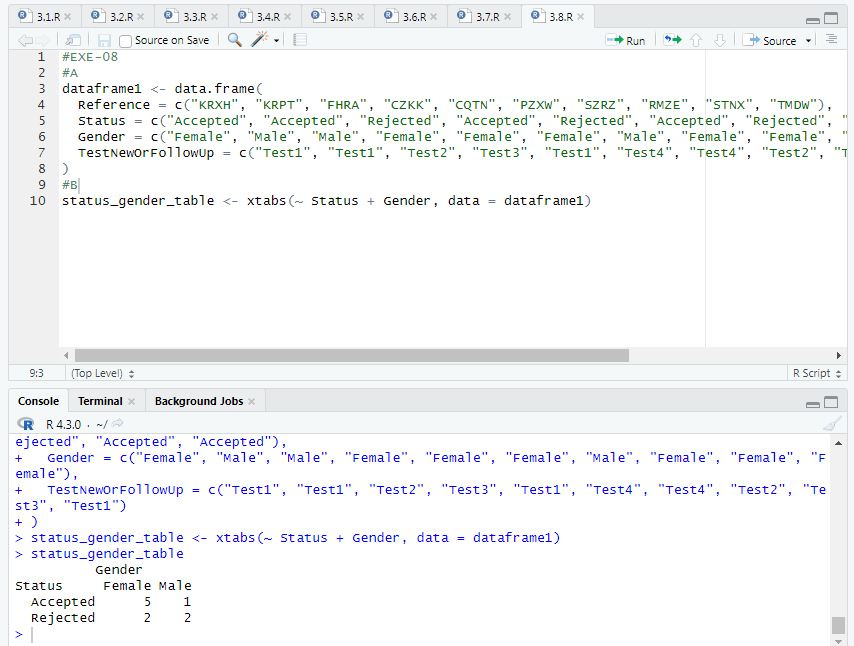
RMZE 0 1

STNX 1 0

SZRZ 0 1

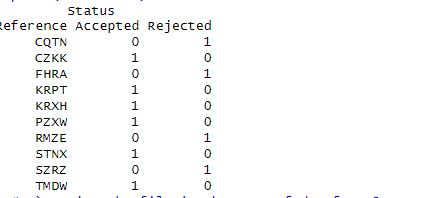
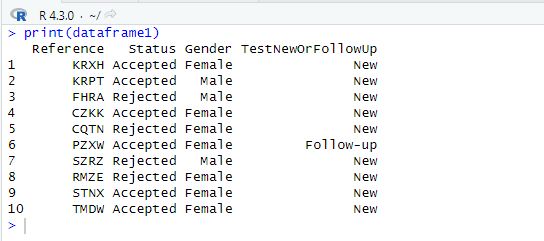
TMDW 1 0

> save(dataframe2, file="dataframe2.RData")



Exercise: 9

Use Two Categorical Variables and Discover the relationships within a dataset

Next, using the xtabs() function, apply two variables from “dataframe1 “, to create a table delineating the relationship between the “Reference” category, and the “Status” category.Save the file in the name of dataframe2

Exercise: 10

Use the same data frame using three Categorical Variables create a Multi-Dimensional Table

Apply three variables from “dataframe1” to create a Multi-Dimensional Cross-Tabulation of “Status“, “Gender“, and “Test“.

Exercise: 11

Row Percentages

The R package “tigerstats” is required for the next two exercises.

1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.

2) Enclose the xtabs() formula in the tigerstats function, “rowPerc()” to display row percentages for “Status” by “Test“.

Exercise 12

Column Percentages

1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.

2) Enclose the xtabs() formula in the tigerstats function, “colPerc()” to display row percentages for “Status” by “Test“.

VISUALIZATION IN R

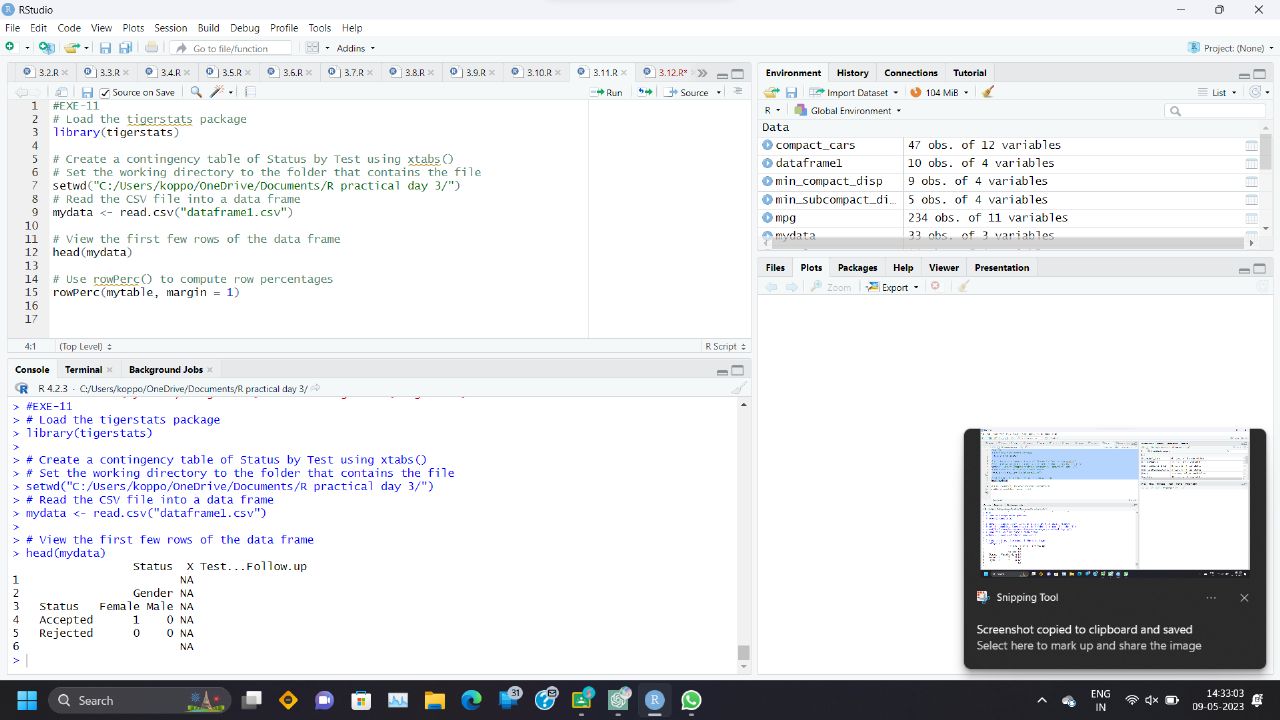
Write a program for creating a pie-chart in R using the input vector(21,62,10,53). Provide labels for the chart as ‘London’, ‘New York’, ‘Singapore’, ‘Mumbai’. Add a title to the chart as ‘city pie-chart’ and add a legend at the top right corner of the chart.

Create a 3D Pie Chart for the dataset “political Knowledge” with suitable labels,colours and a legend at the top right corner of the chart.

Write a program for creating a bar chart using the vectors H=c(7,12,28,3,41) and M=c(“mar”, “apr”, “may”, “jun”, “jul”). Add a title to the chart as “Revenue chart”.

Make a histogram for the “AirPassengers“dataset, start at 100 on the x-axis, and from values 200 to 700, make the bins 200 wide

Create a Boxplot graph for the relation between "mpg"(miles per galloon) and "cyl"(number of Cylinders) for the dataset "mtcars" available in R Environment.



VISUALIZATION IN R