Northeastern University

Toronto, Canada

A Report on

Executive Summary of module 1

Under the subject of:

Introduction to Data Analytics

(ALY 6000)

Guided by:

Prof. Mohammad Shafiqul Islam

Submitted By:

Name of Student	NUID	Date of Submission
Parva Patel	002195186	21 January 2022

Index

1. **Scatter plot of the Sales ~ Temperature data.**

Sales data : (7,11,15,20,19,11,18,10,6,22)

Temperature data : (69,81,77,84,80,97,87,70,65,90)

The Sales data and Temperature data was given in the task to present the scatter plot.

Plots play an important role in analysing data as it helps to show relationships between variables or to visualize variation. Scatter plot is most commonly used graph in R language. It is used to determine the relationship between the two numeric variables and is described by positive or negative, linear or nonlinear and strong or weak.

Image below shows the scatter plot of Sales ~ Temperature
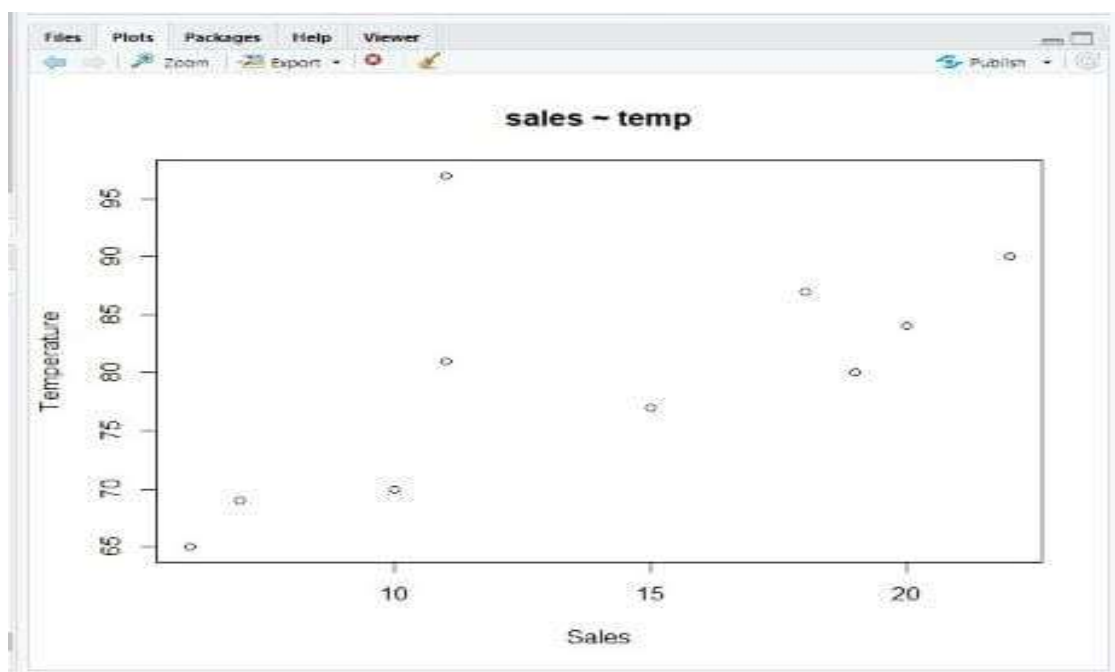
2. The mean Temperature.

Function are modules of code that performs specific task. There are many functions in r language and mean() is one the commonly used function. Mean() function is used to calculate the mean or average of given numeric data.

For example :

Temperature <- c(69,81,77,84,80,97,87,70,65,90)

Mean(Temperature)

Output :

mean(Temperature)

[1] 80

3. Deleting and Inserting a element in a Vector.

In R vector is a basic data structure which shares the elements of same data type in a sequence. It is possible to make changes in the vector such as one can insert and delete the elements from the vector.

To perform delete operation the syntax for it is Sales <- Sales[!Sales %in% c(15)].Here Sales is the vector on which operation is performed and 15 is the element which will get deleted.

To perform insert operation the syntax for it is Sales <- c(Sales[1:2],16,Sales[3:length(Sales)]). Here Sales is the vector on which operation is performed and 16 in the element inserted in 3rd position.

Initial Sales vector :

Sales

[1] 7 11 15 20 19 11 18 10 6 22

After performing delete operation

Sales

[1] 7 11 20 19 11 18 10 6 22

After performing Insert operation

Sales

[1] 7 11 16 20 19 11 18 10 6 22

4. The names vector

Elements of vector can be of the names. For example names <- c("Tom", "Dick", "Harry"). Names is the vector and its elements are Tom, Dick and Harry.

5. Matrix of 5 row by 2 column of 10 integers.

Matrix is a two dimensional data structure which has rows and columns. In R one can create matrix of n number of rows and columns by using the matrix() function.

Example:

Syntax :

Matrix <- matrix(c(1:10),nrow=5,ncol=2)

Output :

	[,1]	[,2]
[1,]	1	6
[2,]	2	7
[3,]	3	8
[4,]	4	9
[5,]	5	10

6. The icsales data frame.

Data frame is a similar data structure to matrix. Moreover, data frame is often used for data analysis as it can store different classes of object in each column. To create a data frame the function used is data.frame().

Syntax :

icSales <- data.frame(Sales,Temperature)

Data frame structure of icSales :

'data.frame':   10 obs. of 2 variables:

 $ Sales   : num 7 11 16 20 19 11 18 10 6 22

 $ Temperature: num 69 81 77 84 80 97 87 70 65 90

7. The summary of the icsales data frame.

When a summary of a large amount of data is accessible, it becomes much easier to work with and relate to that data. Similarly, the r

language has a summary function [summary()] that returns a summary of the data frames.

When the summary() function is used on quantitative data, descriptive statistics such as the minimum, 1st quartile, median, mean, 3rd quartile, and maximum value of the input data are returned.

summary(icSales)

output :

Sales   Temperature

Min.   : 6.00   Min.   :65.00

1st Qu.:10.25   1st Qu.:71.75

Median :13.50   Median :80.50

Mean   :14.00   Mean   :80.00

3rd Qu.:18.75   3rd Qu.:86.25

Max.   :22.00   Max.   :97.00

8. The variables of Student.csv data set.

I learned how to use several functions to import various types of data sets by completing this task. I used the Student.csv dataset in this work, which contains variables like "StudentID", "First", "Last", "Math", "Science" and "Social.Studies".

9. Summary of information about data sets based on the instructions followed.

I was able to learn about the value of various sorts of data and datasets for data analysis by completing this assignment. Each data set has its unique set of traits and functions, resulting in a variety of outcomes.

In this project, I worked on the Sales and Temperature datasets, both of which have the same data type, however the Student.csv file contains a variety of data types, including tables containing qualitative and quantitative data. Apart from that, I learned about a variety of plotting schemes that make data visualisation more easier and aid in the correlation of data sets.

<u>Bibliography</u>

1. **Scatter plot of the Sales ~ Temperature data.**
   https://www.statmethods.net/graphs/scatterplot.html

2. **Deleting and Inserting elements in Vector.**
   https://www.youtube.com/watch?v=rgAvJmvfA2c&ab_channel=Mr.MathExpert

   https://www.youtube.com/watch?v=aRDG2nwiXnY&ab_channel=R-programmingLibrary

3. **Matrix of 5 row by 2 column of 10 integers.**
   https://www.tutorialspoint.com/r/r_matrices.htm

4. **The icsales data frame.**
   https://www.w3schools.com/r/r_data_frames.asp

5. **Importing Student.csv data set.**
   https://www.datacamp.com/community/tutorials/r-data-import-tutorial#csv

   GitHub Username : ParvaPatel10

   GitHub Repository : https://github.com/ParvaPatel10/Module1

<u>Appendix</u>

 # 1. Print your name at the top of the script

```
print("Parva Pareshbhai Patel")
```

 # 2. Install the vcd package

```
r=getOption("repos")
r["CRAN"]="http://cran.us.r-project.org"
options(repos=r)
install.packages("vcd")
```

 # 3. Import the vcd library

```
library(vcd)
```

 # 4. plot a sales ~ temp scatter plot

```
Sales <- c(7,11,15,20,19,11,18,10,6,22)
Temperature <- c(69,81,77,84,80,97,87,70,65,90)
plot(Sales,Temperature,main = "sales ~ temp")
```

5.Mean of Temperature

```r
mean(Temperature)
```

6. Delete the 3 element from the sales vector

```r
Sales <- Sales[-3]
Sales
```

7. Insert 16 as the 3rd element in Sales vector

```r
Sales <- c(Sales[1:2],16,Sales[3:length(Sales)])
Sales
```

8. Vector names with elements Tom, Dick, Harry

```r
names <- c("Tom","Dick","Harry")
names
```

9. 5 rows and 2 column matrix of 10 integers

```r
matrix(c(1:10),nrow=5,ncol=2)
```

10. icSales data frame with Sales and Temperature attributes

```r
icSales <- data.frame(Sales,Temperature)
icSales
```

11. The data frame structure of icSales

```r
str(icSales)
```

12. The summary of the icSales data frame

```r
summary(icSales)
```

13. Import the dataset Student.csv

```r
student<-read.csv("Student.csv",header=TRUE,sep=",")
View(student)
```

Display names of student

ls(student)