

Module 1 Project

Course name: ALY6000 Introduction to Analytics

Student name: Sanjana Mohile

Date: 2022-01-22

Title: Executive Summary Report 1

Professor: Mohammad Shafiqul Islam

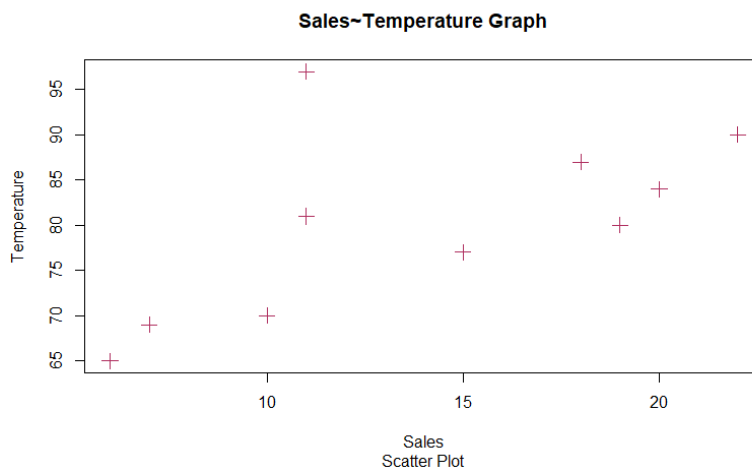
NUID: 002123793

A. A Scatter plot of Sales ~Temperature Data ^[6]

Code –

```
Sales_data <- c(7,11,15,20,19,11,18,10,6,22)
Temperature_data <- c(69,81,77,84,80,97,87,70,65,90)
x <- Sales_data
y <- Temperature_data
?plot
plot(x,y, main = "Sales~Temperature Graph", sub = "Scatter Plot",
      xlab = "Sales", ylab = "Temperature", type = "p",pch=3 , cex = 1.5,col = "maroon")
```

Output –



B. The mean temperature

Code –

```
mean_temp = mean(Temperature_data)
> print(paste("The mean temperature is : ",mean_temp))
```

Output –

```
[1] "The mean temperature is : 80"
```

C. Display the data after steps 6 and 7

a. Delete the third element in the sales data. ^[6]

Code –

```
Sales_data[-3]
```

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

b. Insert 16 as the third element into sales vector

Output –

```
Sales_data[3] <- 16
```

```
> print(Sales_data)
```

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

D. Display the names vector

Code –

```
names <- c('Tom','Dick','Harry')
```

```
> print(names)
```

Output –

```
[1] "Tom" "Dick" "Harry"
```

E . Display the 5 row by 2 column of 10 integer. ^[2]

Code –

```
matrix_a <- matrix(data, nrow = 5, ncol = 2, byrow = TRUE)
```

```
> matrix_a
```

Output –

```
      [,1] [,2]
[1,] 10  20
[2,] 30  40
[3,] 50  60
[4,] 70  80
[5,] 90 100
```

F. Display the icSales data frame

Code –

```
icScales <- data.frame(Sales_data, Temperature_data)
```

```
> print(icScales)
```

Output –

Sales_data	Temperature_data
1	7
2	11
3	16
4	20
5	19
6	11
7	18
8	10
10	6
10	22

G. Display the summary of icSales data frame

Code –

```
summary(icScales)
```

Output –

Sales_data	Temperature_data
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

H. Display the variables only from the Student.csv data set.^[5]

Code –

```
colnames(Student)
```

Output –

```
[1] "StudentID"  "First"      "Last"      "Math"      "Science"    "Social.Studies"
```

I. Summary –

The Sales data and Temperature data variables are both of the double data type. The `typeof()` function was used to verify this. They each had ten observations, with the lowest being 6 and the highest being 22, for sales data, and a maximum of 90 and a minimum of 65 for temperature data. The mean temperature was recorded to be 80. In the student data set Mary O'Leary received the top marks in all three subjects: math, science, and social studies. Dan Thornton received the lowest marks in Math and Science, while Bob Smith received the lowest marks in Social Studies.

Bibliography –

1. <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/plot>
2. <https://www.guru99.com/r-matrix-tutorial.html>
3. https://www.tutorialspoint.com/r/r_data_frames.htm
4. <https://www.geeksforgeeks.org/how-to-import-data-from-a-file-in-r-programming/>
5. <https://stackoverflow.com/questions/44152087/r-data-table-accessing-column-with-variable-name>
6. <https://www.educba.com/scatterplots-in-r/>
7. <https://stackoverflow.com/questions/652136/how-can-i-remove-an-element-from-a-list>

Appendix –

Q1. Print your name on the top of the script

```
> print("Sanjana Sandeep Mohile")
[1] "Sanjana Sandeep Mohile"
```

Q2. Install the vcd package

```
> install.packages("vcd")
```

There is a binary version available but the source version is later:

```
  binary source needs_compilation
vcd 1.4-8 1.4-9          FALSE
```

installing the source package ‘vcd’

.....

The downloaded source packages are in

```
‘C:\Users\mohil\AppData\Local\Temp\RtmpwpgyVO\downloaded_packages’
```

I found another way to install packages.

On the packages tab, in the R window we have a tab written as install. Click on the “install” tab and write the name of the package you wish to install.

You will see the package getting downloaded in the Console window.

Q3. Import the vcd package

```
library(vcd)
Loading required package: grid
```

Another way to import any library is to search the library in the packages and tick the square box that is given prior to it's name.

Q4. Plot a sales~temperature scatter plot using the data

```
# Scatter plot of sales~temperature
Sales_data <- c(7,11,15,20,19,11,18,10,6,22)
Temperature_data <- c(69,81,77,84,80,97,87,70,65,90)

x <- Sales_data
y <- Temperature_data

?plot
plot(x,y, main = "Sales~Temperature Graph", sub = "Scatter Plot",
      xlab = "Sales", ylab = "Temperature", type = "p",pch=3 , cex = 1.5,col = "maroon")
```

Q5.Find the mean temperature

```
> print(paste("The mean tempature is : ",mean_temp))
[1] "The mean tempature is : 80"
```

Q6.Delete the 3rd element from the sales vector.

```
Sales_data[-3]
[1] 7 11 20 19 11 18 10 6 22
```

Q7.Insert 16 as 3rd element in the sales vector.

```
> Sales_data[3] <- 16
> print(Sales_data)
[1] 7 11 16 20 19 11 18 10 6 22
```

Q8.Create a vector <names> with elements Tom, Dick, Harry.

```
> names <- c('Tom','Dick','Harry')
> print(names)
[1] "Tom" "Dick" "Harry"
```

Q9.Create a 5 row and 2 column matrix of 10 integers.

```
data <- c(10,20,30,40,50,60,70,80,90,100)
> matrix_a <-matrix(data, nrow = 5, ncol = 2, byrow = TRUE)
```



```
> matrix_a
     [,1] [,2]
[1,]  10  20
[2,]  30  40
[3,]  50  60
[4,]  70  80
[5,]  90 100
```

Q10. Create a dataframe <icSales> with sales and temp attributes
 icSales <- data.frame(Sales_data, Temperature_data)

Q11. Display the dataframe structure of icSales

```
> print(icSales)
  Sales_data Temperature_data
1         7             69
2        11             81
3        16             77
4        20             84
5        19             80
6        11             97
7        18             87
8        10             70
9         6             65
10       22             90
```

Q12. Display a summary of icSales dataframe

```
summary(icSales)
  Sales_data  Temperature_data
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
```

Q13. Import the dataset Student.csv

```
Student <- read.csv("C:/Users/mohil/OneDrive/Desktop/Studies/ALY-6000 Introduction to Analytics/Module 1/Student.csv", header = TRUE, sep = ",")
```

Warning message:

```
In read.table(file = file, header = header, sep = sep, quote = quote, :
```

```
incomplete final line found by readTableHeader on 'C:/Users/mohil/OneDrive/Desktop/Studies/ALY-6000 Introduction to Analytics/Module 1/Student.csv'
```

```
print(Student)
```

	StudentID	First	Last	Math	Science	Social.Studies
1	11	Bob	Smith	90	80	67
2	12	Jane	Weary	75	NA	80
3	10	Dan	Thornton, III	65	75	70
4	40	Mary	O'Leary	90	95	92

Q14. Display only the variable names of Student.csv

```
colnames(Student)
```

```
[1] "StudentID" "First" "Last" "Math" "Science" "Social.Studies"
```

My Github username : SanjanaMohile

Github Repository : <https://github.com/SanjanaMohile/ALY6000-Module1.git>