



Toronto, Canada

Executive summary of Module 1

Introduction to Data Analysis
(ALY 6000)

Guided by:
Prof. Mohammad Shafiqul Islam

Submitted by:

Name:	NUID:	Submission:
Dhruvang Patel	002195090	20 th January, 2022

INDEX

- i. A scatter plot of the Sales ~ temperature data
- ii. The mean temperature
- iii. Display the data after steps 6 and 7
- iv. Display the names vector
- v. Display the 5 row by 2 column of 10 integers
- vi. Display the icSales data frame
- vii. Display the summary of the icSales data frame
- viii. Display the variables only from the Student.csv data set.
- ix. A summary of the information you learned about the data sets based on the
- x. instructions you followed.

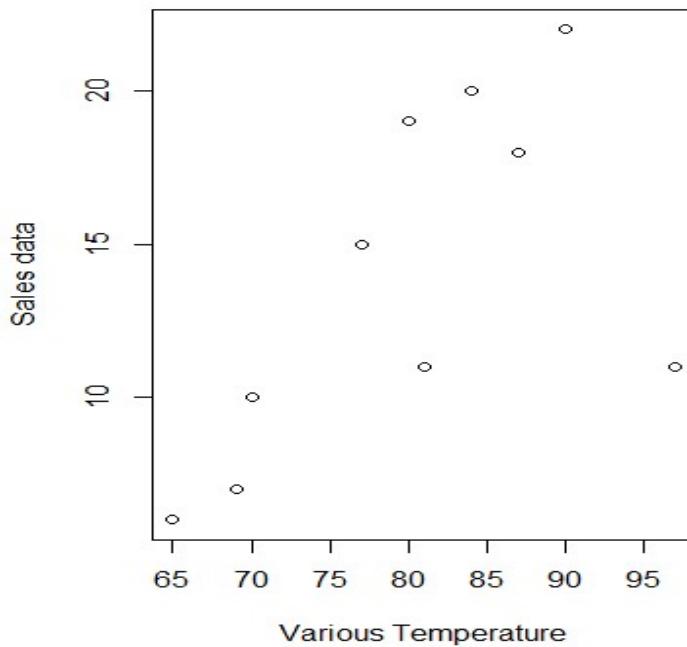
i. A scatter plot of the Sales ~ temperature data

The data for sales and temperature is given as follows:

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

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

From this I came to know that scatterplots provide a visual depiction of the relationship between two variables data, which can help reveal linear and non-linear patterns.



ii. Mean Temperature

To find mean we have to add all the scores and divide it by total number of scores.

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

Here I found the mean of the above data and the mean is 80.

iii. Display the data after steps 6 and 7

Step 6: Deleting an element.

So, in this command I deleted the 3rd element from sales data.

The syntax for deleting the element is:

mylist[-number]

The program I performed to remove third element

Sales <- Sales[-3]

Step 7: Inserting an element

Syntax for adding an element is:

A <- c(A[value], number, A[value])

For adding 16 value at 3rd position on sales data

Sales<-c(Sales[1:2],16,Sales[3:9])

iv. Display the names vector

Vectors are one-dimensional arrays that can hold numeric data, character data, or logical data. The combined function c() is used to form the vector.

Syntax:

A <- c("name1", "name2", "name3")

The program I performed to display the names is

name <-c("Tom", "Dick", "Harry")

Output:

"Tom" "Dick" "Harry"

v. Display the 5 row by 2 column of 10 integers

I used the matrix() function to create a matrix and nrow and ncol can be used for getting appropriate value for rows and column respectively

The task I performed to make matrix of 5 rows and 2 columns is:

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

Output:

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

vi. Display the icSales data frame

A data frame is a table array-like structure in which is made from different datasets.

The program I did for icSales dataframe is:

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

Output:

	Sales	Temperature
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

vii. **Display the summary of the icSales data frame**

Summary is a generic function used to provide summaries of data related to the individual object that was fed into it.

The input for the following code is:

```
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

viii. **Display the variables only from the Students.csv dataset**

I used ls() to list all variables that are created in the environment.

The output I got after using the ls(students) function is:

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

ix. **A summary of the information I learned.**

From this module I walked away with a plethora of information regarding programming language R. I learned different types of statistics, datasets-dataframes, and vectors. I also plotted different types of graphs to have a better understanding of data. I also did different types of mathematical expressions which helped with easy analysis of data. I also did modification of data like adding and deleting the values in a dataset and also imported the dataset from external.

BIBLIOGRAPHY

- <https://youtu.be/VmOIVFXBsY>
- <https://www.datamentor.io/r-programming/matrix/>
- https://www.tutorialspoint.com/r/r_data_frames.htm
- <https://www.r-bloggers.com/2009/11/r-tutorial-series-summary-and-descriptive-statistics/>
- <https://www.edureka.co/community/50251/>

APPENDIX

```
#name
print("Dhruvang Patel")

#Install vcd package
r=getOption("repos")
r["CRAN"]="https://cran.r-project.org/"
options(repos=r)
install.packages("vcd")

#Import library
library(vcd)

#Load Sales data
Sales <- c(7, 11, 15, 20, 19, 11, 18, 10, 6, 22)

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

#Plot Data
plot(Sales ~ Temperature,
      xlab = "Various Temperature",
      ylab = "Sales data")
#Mean
mean(Temperature)

#Remove 3rd element
Sales <- Sales[-3]
```

Sales

```
#Insert element
```

```
Sales <- c(Sales[1:2], 16, Sales[3:9])
```

```
Sales
```

```
#Create name
```

```
name <- c("Tom", "Dick", "Harry")
```

```
name
```

```
#Creating matrix
```

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

```
#Create Dataframes
```

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

```
icSales
```

```
#Dataframe structure
```

```
structure(icSales)
```

```
#summary of Dataframe
```

```
summary(icSales)
```

```
#Import students data
```

```
library(readxl)
```

```
Student <- read_excel("~/R/ALY 6000/Module 1/Student.xlsx")
```

```
View(Student)
```

```
#display names of students
```

```
ls(Student)
```

GitHub username: dhruvang186

GitHub repository: <https://github.com/dhruvang186/Module1>