Name: Yijun Wang

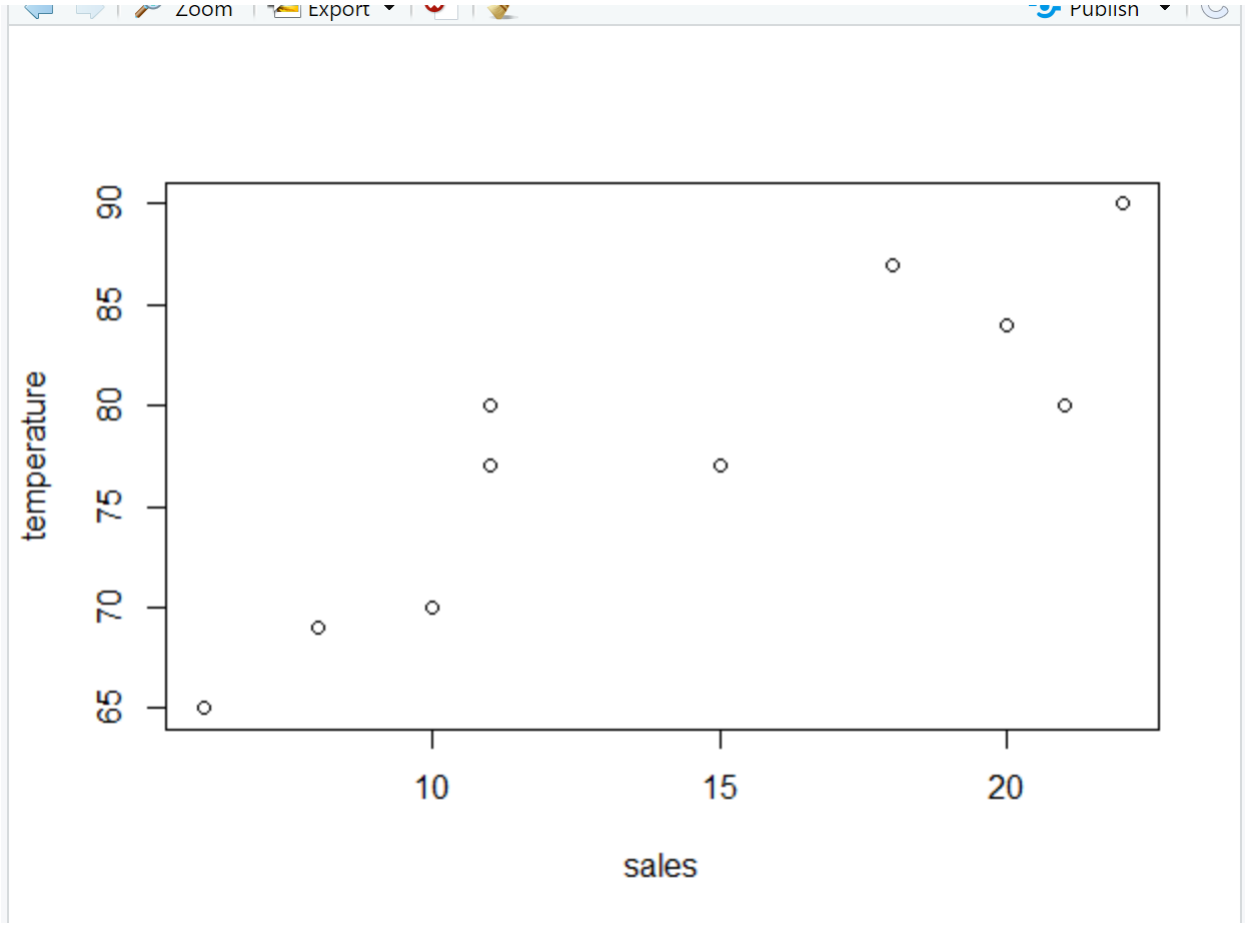
Class: Aly 6000

Date: 09/26/2021

Title: Module 1 Project — Executive Summary Report 1

Key findings about the data based on the Dataset Instruction document

a. A scatter plot of the Sales ~ temp data



b. The mean temperature

```
> mean(temperature)
[1] 77.9
```

c. Display the data after steps 6 and 7

Step 6:

```
> #Delete the 3rd element from the sales vector
> indices <- c(3)
> result <- sales[-indices]
> print(result)
[1]  8 11 20 21 11 18 10  6 22
```

Step 7:

```
[1]  8 11 20 21 11 18 10  6 22
> #Insert 16 as the 3rd element
> sales <- c(result[1:2],16,result[3:9])
> sales
 [1]  8 11 16 20 21 11 18 10  6 22
> 
```

d. Display the names vector

```
> #create a vector names
> names <- c("Tom", "Dick", "Harry")
> names
[1] "Tom"   "Dick"  "Harry"
> 
```

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

```
> #5 row and 2 column matrix of 10 int
> names <- matrix(1:10, nrow=5, ncol=2)
> names
     [,1] [,2]
[1,]    1    6
[2,]    2    7
[3,]    3    8
[4,]    4    9
[5,]    5   10
> 
```

f. Display the icSales data frame

```
> #<icSales> with sales and temp attributes
> icSales <- data.frame(sales, temperature)
> icSales
   sales temperature
1      8          69
2     11          80
3     16          77
4     20          84
5     21          80
6     11          77
7     18          87
8     10          70
9      6          65
10    22          90
> 
```

g. Display the summary of the icSales data frame

```
> #summary of icScales data frame
> summary(icSales)
     sales        temperature
 Min.   : 6.00   Min.   :65.00
 1st Qu.:10.25   1st Qu.:71.75
 Median :13.50   Median :78.50
 Mean   :14.30   Mean   :77.90
 3rd Qu.:19.50   3rd Qu.:83.00
 Max.   :22.00   Max.   :90.00
>
```

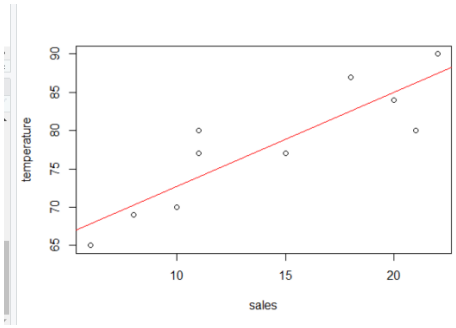
h. Display the variables only from the Student.csv data set.

```
> #Variable names of the Student.csv dataset
> colnames(read.csv("C:\\Users\\junni\\Downloads\\Student.csv"))
[1] "StudentID"      "First"          "Last"           "Math"
[5] "Science"        "Social.Studies"
>
```

i. A summary of the information you learned about the data sets based on the instructions you followed.

From the quartiles, the average temperature is 77.9, the average sale is 14.2, these two numbers are the center position where the numerical values are relatively concentrated. Except that I can use this quartile to not only visually spot outliers in the data, but also to calculate both detail's number. The sales' median is 13.5, minimum number is 6, first quartile is 10.25, mean is 14.30, third quartile is 19.5, and maximum number is 22. The temperature's median is 78.5, minimum number is 65, first quartile is 71.75, mean is 77.9, third quartile is 83, and maximum number is 90. When different sets of data are compared, the data clearly shows the difference distribution of each group, which is highly useful for generating conclusions.

The data distribution in the graphic shows that the relationship between sales and temperature, data is often concentrated near a central value. And use R language to calculate Linear Regression, the result is directly proportional.



This means as the Sales increases, the temperature will increase. Predict the distribution of points appearing at unknown locations, thereby predicting the target value to be evaluated.

Bibliography

Bibliography

datatofish. (2021, July 16). *How to Import a CSV File into R (example included)*. Retrieved from datatofish: https://datatofish.com/import-csv-r/

Kabacoff, R. (2015). *R in Action.* New York: Manning; 2nd edition.

Schork, J. (2021, 01 18). *Read Only Header of File in R (2 Examples)*. Retrieved from statisticsglobe: https://statisticsglobe.com/read-only-header-of-file-in-r

Appendix

1. Print your name at the top of the script

```
1  #Print the name at the tip of the script
2  paste("Yijun Wang")
3
```

2. Install the vcd package(pg 19)

```
4  #Install the vcd package
5  help.start()
6  install.packages("vcd")
7  help(package="vcd")
8
```

3. Import the vcd library(pg 19)

```
9  #Install the vcd library
10 library(vcd)
11 help (Arthritis)
12 Arthritis
13 example(Arthritis)
14 #q()
```

4. Plot a sales ~ temp scatter plot using the data below(pg 9):Sales data: (8,11,15,20,21,11,18,10,6,22)Temperature data: (69,80,77,84,80,77,87,70,65,90)

```
#Plot sales&temperature data
sales <- c(8,11,15,20,21,11,18,10,6,22)
temperature <- c(69,80,77,84,80,77,87,70,65,90)
plot(sales, temperature)
```

5. Find the mean temperature(pg 9)

```
20
21 #The mean of temperature
22 mean(temperature)
```

6. Delete the 3rdelement from the sales vector

```
23 #Delete the 3rd element from the sales vector
24 indices <- c(3)
25 result <- sales[-indices]
26 print(result)
```

7. Insert 16 as the 3rdelement into the sales vector

```
27 #Insert 16 as the 3rd element
28 sales <- c(result[1:2],16,result[3:9])
29 sales
30
```

8. Create a vector <names> with elements Tom, Dick, Harry(pg 22)

```
30
31  #create a vector names
32  names <- c("Tom", "Dick", "Harry")
33  names
```

9. Create a 5 row and 2 column matrix of 10 integers(pg 23) (Kabacoff, 2015)

```
34  #5 row and 2 column matrix of 10 int
35  names <- matrix(1:10, nrow=5, ncol=2)
36  names
37
```

10. Create a data frame <icSales> with sales and temp attributes(pg 26)

```
37
38  #<icSales> with sales and temp attributes
39  icSales <- data.frame(sales, temperature)
40  icSales
41
```

11. Display the data frame structure of icSales(pgs28-31)

```
41
42  #data fram structure of icScales
43  str(icSales)
```

12. Display a summary of the icSales data frame(pgs 28-31)

```
44  #summary of icScales data frame
45  summary(icSales)
46
```

13. Import the dataset Student.csv(pgs 34-37) (datatofish, 2021)

```
46
47  #import the dataset Student.csv
48  read.csv("C:\\Users\\junni\\Downloads\\Student.csv",header=TRUE,sep=",")
49
```

14. Display only the variable names of the Student.csv dataset. (Schork, 2021)

```
49
50  #Variable names of the Student.csv dataset
51  colnames(read.csv("C:\\Users\\junni\\Downloads\\Student.csv"))
52
53
54
55
```