Name: Yijun Wang

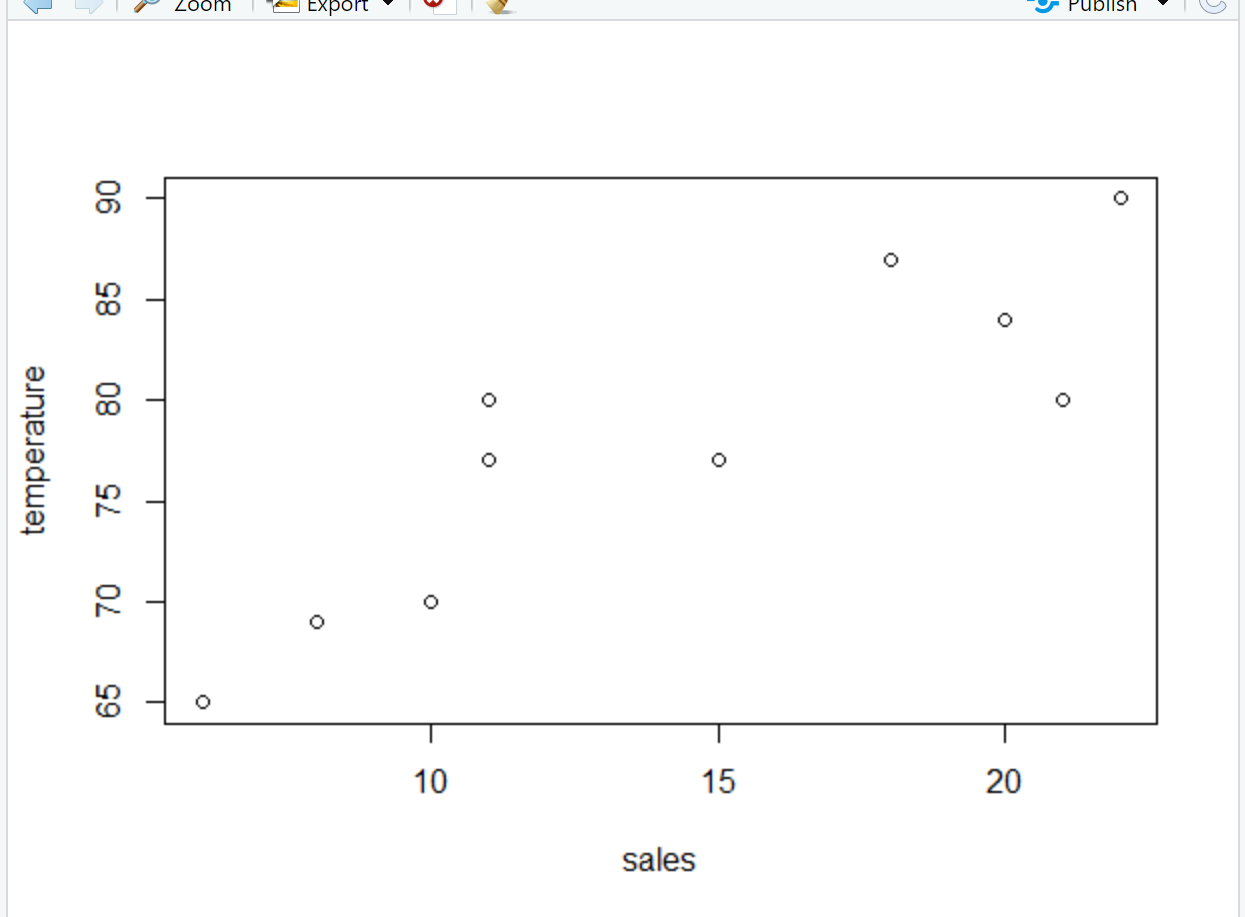
Class: Aly 6000

Date: 09/26/2021

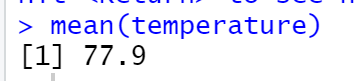
Title: Module 1 Project — Executive Summary Report 1

Key findings about the data based on the Dataset Instruction document

1. A scatter plot of the Sales ~ temp data

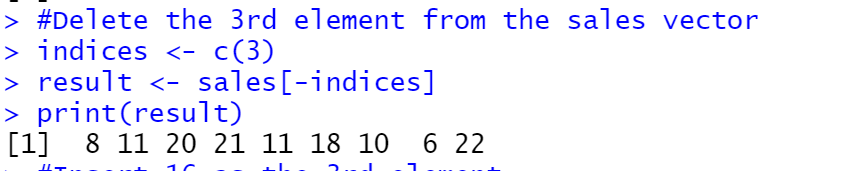


1. The mean temperature

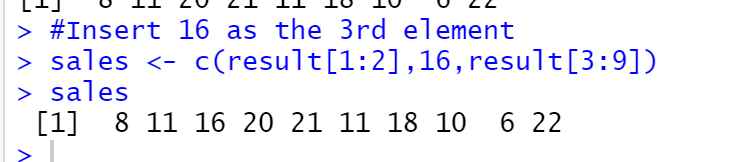


1. Display the data after steps 6 and 7

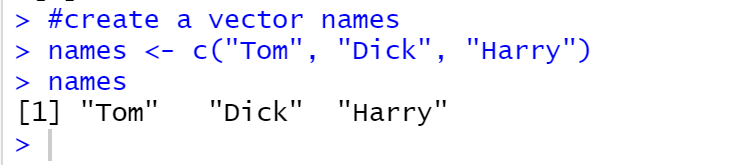
Step 6:



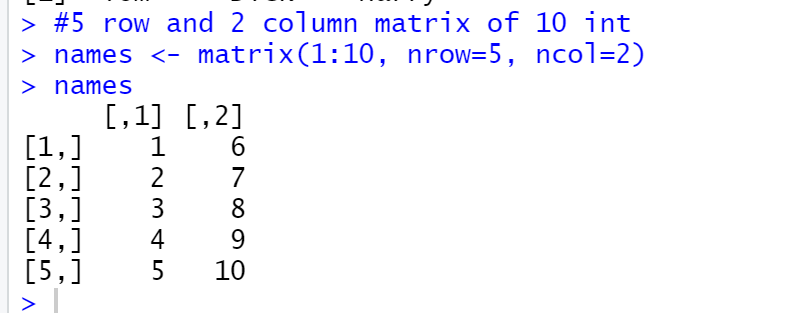
Step 7:



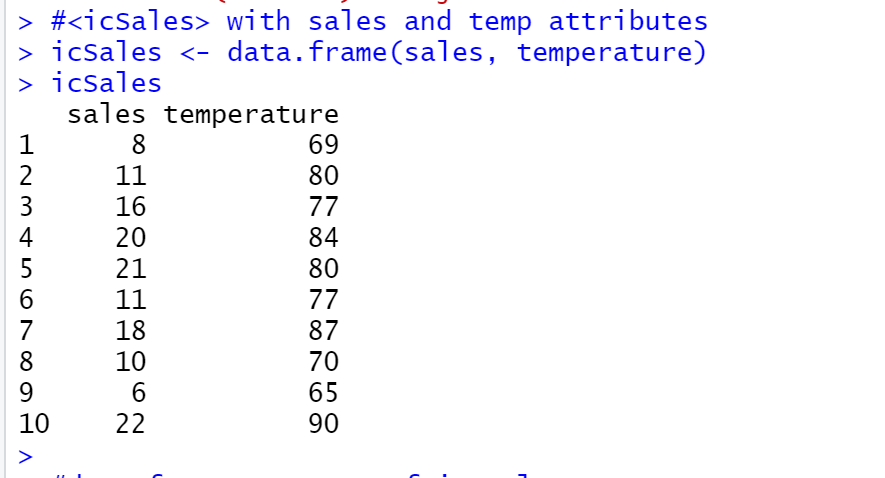
1. Display the names vector



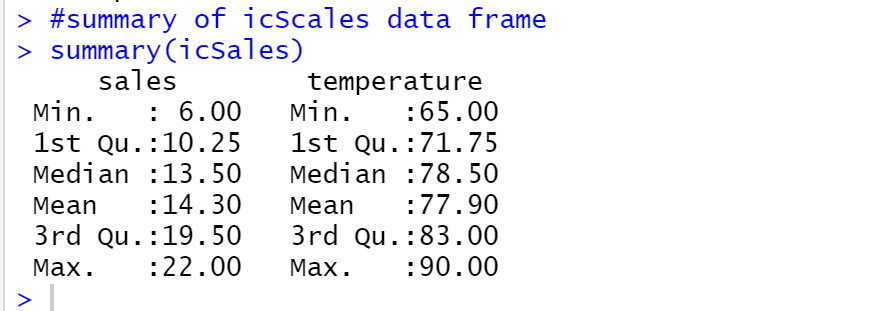
1. Display the 5 row by 2 column of 10 integers



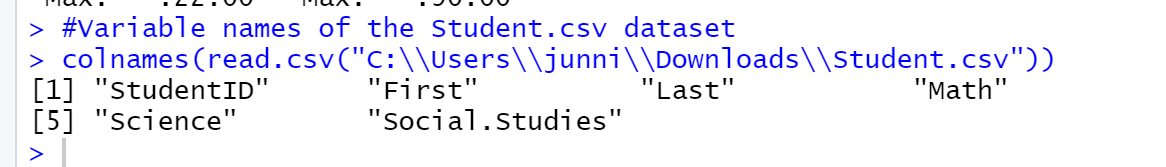
1. Display the icSales data frame



1. Display the summary of the icSales data frame



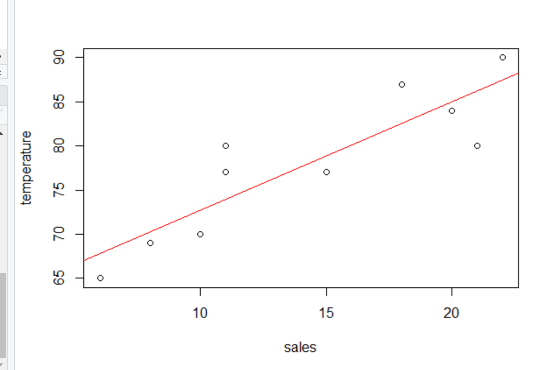
1. Display the variables only from the Student.csv data set.



1. 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

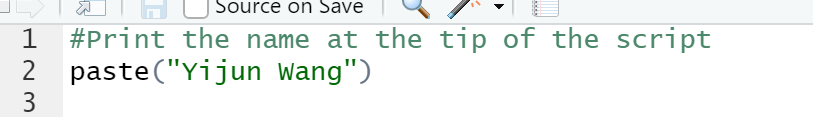
datatofish. (2021, July 16). *How to Import a CSV File into R (example included)*. Retrieved from datatofish: https://datatofish.com/import-csv-r/

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

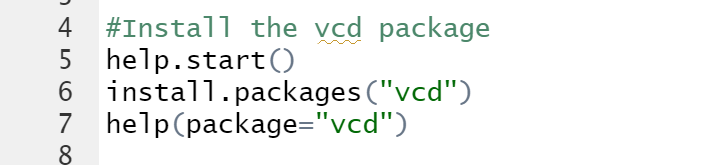
Schork, J. (2021, 01 18). *Read Only Header of File in R (2 Examples)*. Retrieved from statisticsglobe: https://statisticsglobe.com/read-only-header-of-file-in-r

Appendix

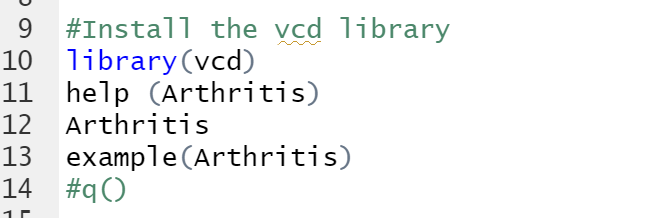
1. Print your name at the top of the script



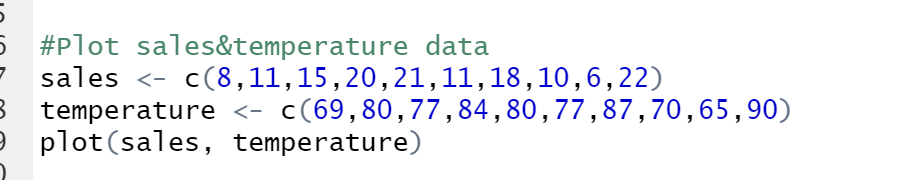
1. Install the vcd package(pg 19)



1. Import the vcd library(pg 19)



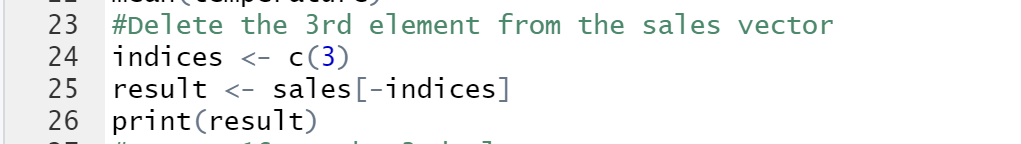
1. Plot a sales ~ temp scatter plot using the data below(pg 9):Sales data: (8,11,15,20,21,11,18,10,6,22)Temperature data: (69,80,77,84,80,77,87,70,65,90)



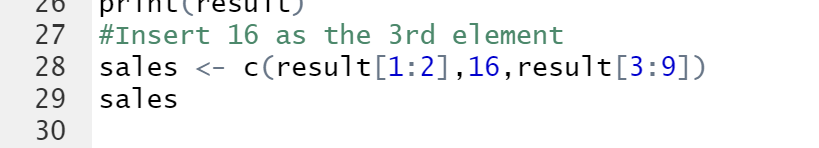
1. Find the mean temperature(pg 9)



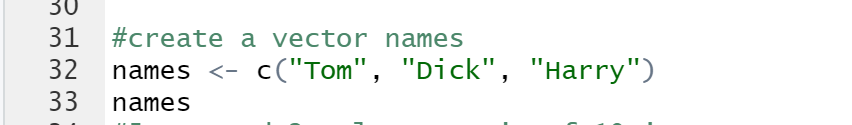
1. Delete the 3rdelement from the sales vector



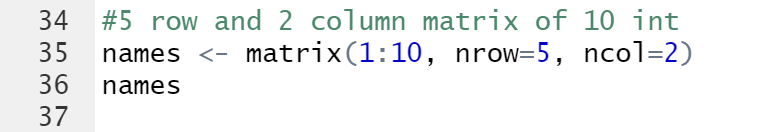
1. Insert 16 as the 3rdelement into the sales vector



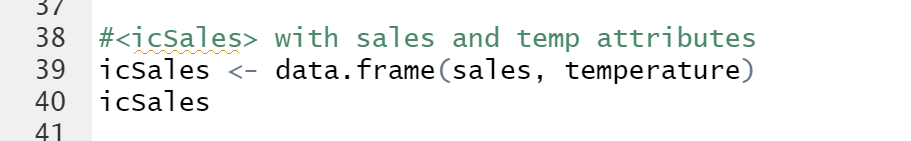
1. Create a vector <names> with elements Tom, Dick, Harry(pg 22)



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



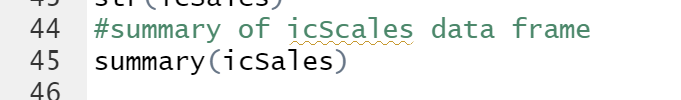
1. Create a data frame <icSales> with sales and temp attributes(pg 26)



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



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



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



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

