**Course Three Task1 Report**

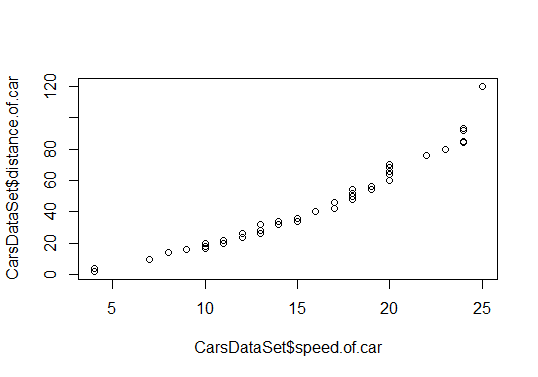
**by**

**Shifeng (Steve) Li**

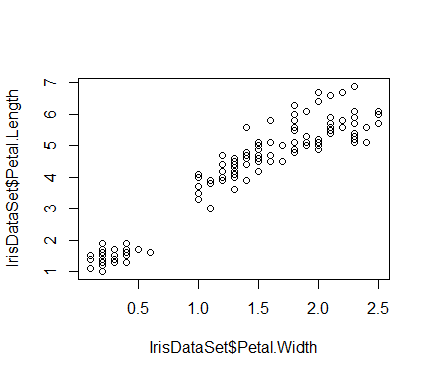
**01/27/2022**

In this project, we used the R programming language and the statistics tools in the R packages to analyze the Cars Data Set and Iris Data Set.

Based on the correlation analysis and plot conducted in the RStudio project, we can see the linear correlation between the speed of car and the distance of the car as the following plot. So basically we can predict the car distance based on the car speed.



Also, for the Iris data set analysis with R, we can see that the Petal width and the Petal Length, have linear correlation as the following plot shows.



Also, shows the quantified correlation coefficient values among the

Iris data features, we see Petal.Width and Petal.Length has correlation coefficient value up to 0.96287. Therefore, Petal.Width can be used to predict Petal.Length

# Sepal.Length Sepal.Width Petal.Length Petal.Width

#Sepal.Length 1.00000 -0.11757 0.87175 0.81794

#Sepal.Width -0.11757 1.00000 -0.42844 -0.36613

#Petal.Length 0.87175 -0.42844 1.00000 0.96287

#Petal.Width 0.81794 -0.36613 0.96287 1.00000

Petal.Width

During running through the data analysis steps in RStudio, encountered several warnings and errors. To resolve the issue, the best practice is to read the warning and error message details. Then study the R Language documentation about the R function or API specification, learn from the samples to the actual usage requirement. Also, search the key words in the warning message and error message to reference how other people resolve the similar issue. After building up enough background knowledge of the related R function/API, then try the solution candidates, often the original issue can be resolved well.

The installation of R and RStudio was straightforward, though have not chosen the customized configuration options yet, the default settings come with the installer provide reasonable functionalities to support going through the initial tutorial project steps.

During running through the data analysis steps in the R script, sometimes the data attribute contents would change as the data set columns are being processed through the functions. So, if need to use the original feature/column values again, may need to use the storage variable to hold the original value, or, need to reload the original data file to the data set.

Also, need to pay attention that the variables in the R language is case sensitive. Therefore, use the right, and self-explained naming convention for the variables will help to reduce the confusion and mistake.

As for doing predictions in R, feel it is similar to what we have done in Jupyter using Python. Eager to explore more to experience power of R and then have more meaningful information to compare the Pros/Cons between the two.

**Uploaded to: https://github.com/UTOct21DaPtSteve/Steve\_Course3**