Prepared By: Divneet Singh Kapoor & Kiran Jot Singh

Worksheet

EXPERIMENT – 8

Name- YANA SRIVASTAVA

UID-20BCS2279 SECTION – 23 B

Aim:

Explore, visualize, transform and summarize input datasets for building classification models.

Requirements:

(i) Rattle
(ii) R Data Miner.

Expected Outcome:

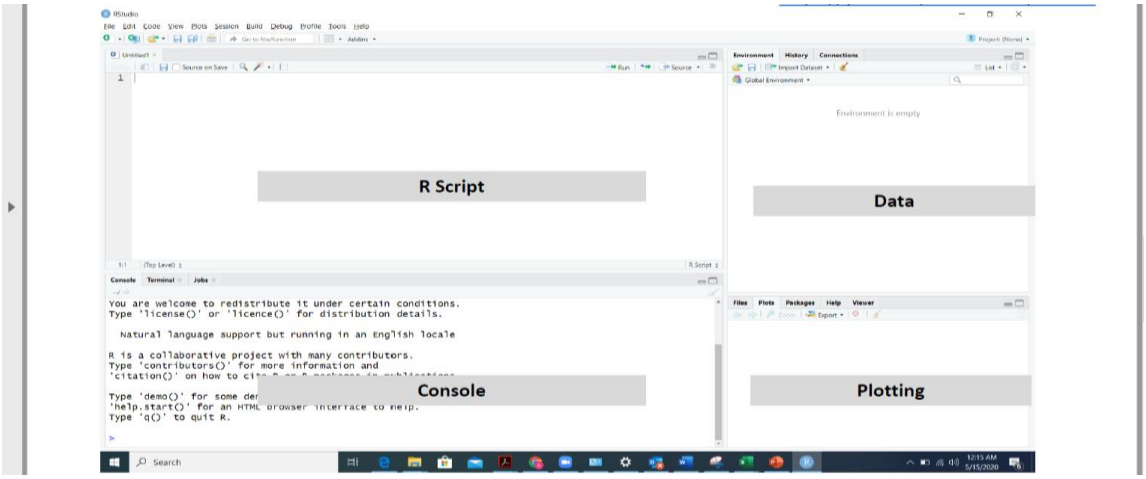
- Data mining combines concepts, tools, and algorithms from machine learning and statistics for the analysis of very large datasets, so as to gain insights, understanding, and actionable knowledge.
- Rattle uses the Gnome graphical user interface as provided through the **RGtk2** package. It runs under various operating systems, including GNU/Linux, Macintosh OS/X, and MS/Windows.

Installation Steps

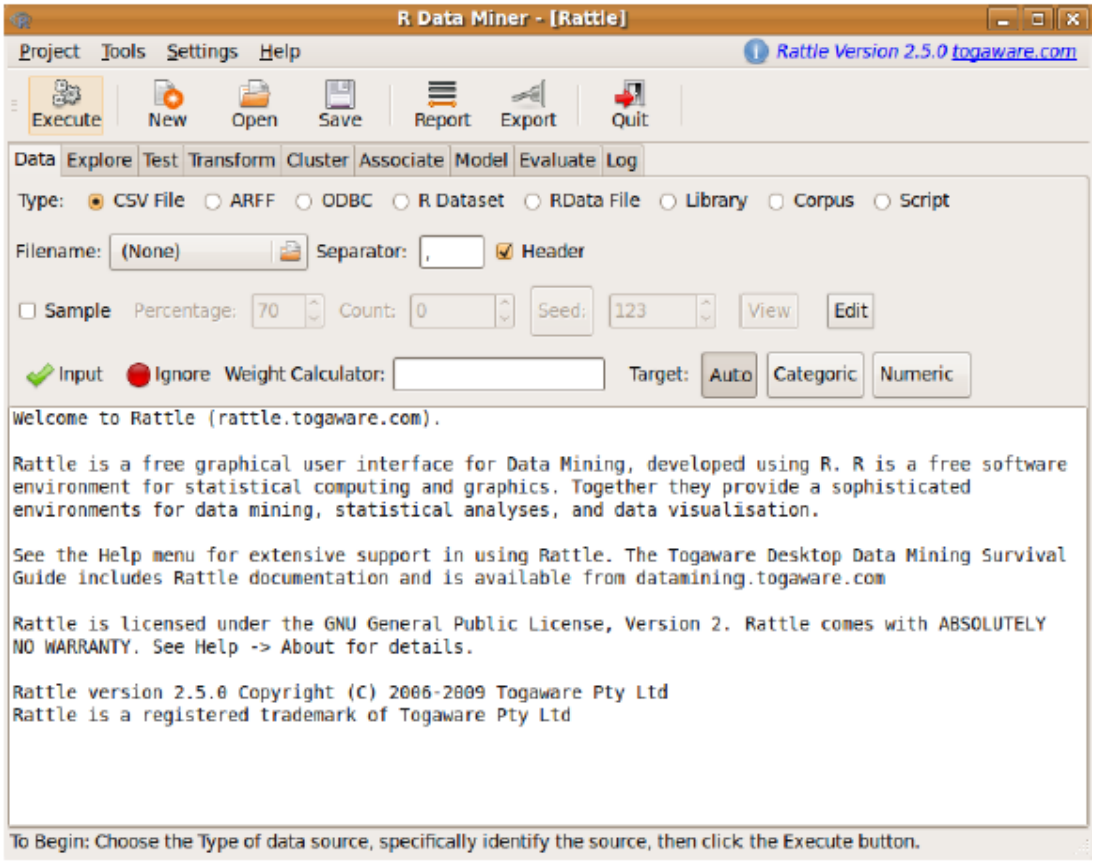
1. Installing R for Windows
    - Go to https://cran.r-project.org/bin/windows/base/

- Click on “Download R 4.0.0 for windows”

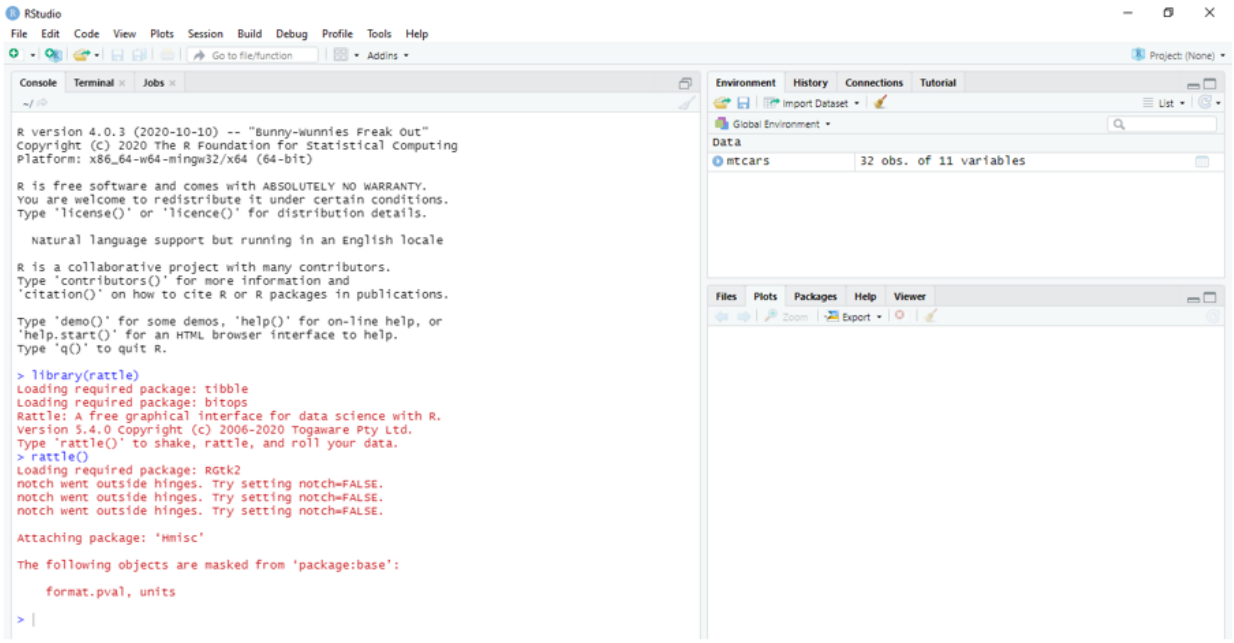
2. Installing RStudio for Windows
   - Go to https://rstudio.com/products/rstudio/download/
   - Click downloaded file to install (select all default options)

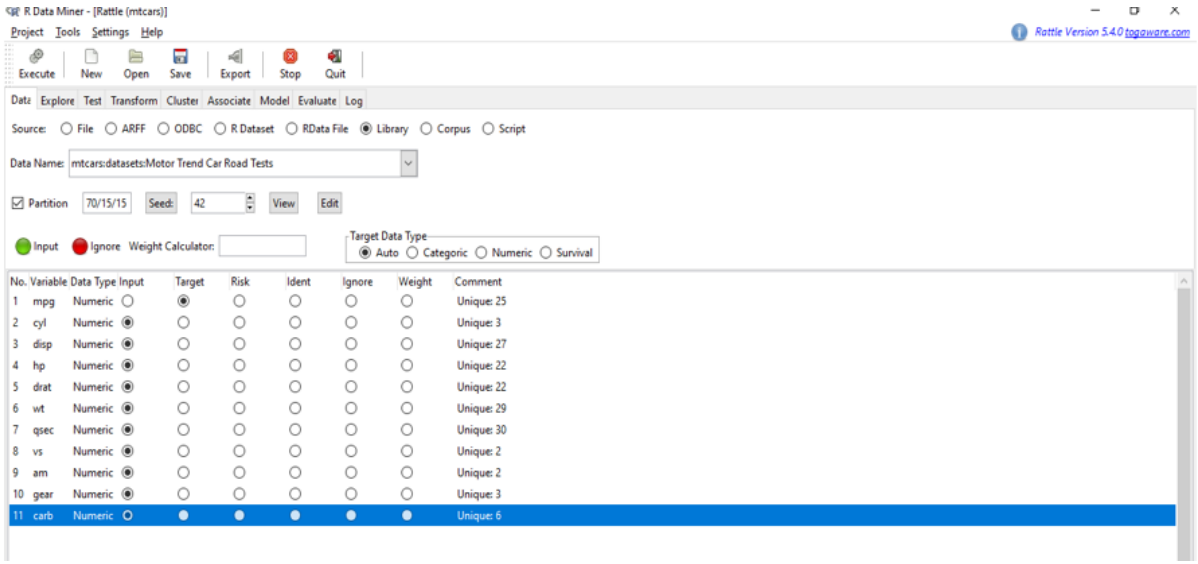


3. Install the following packages.
   I. list_of_packages <- c("tibble", "bitops", "magrittr", "stringi", "XML", "stringr", "Hmisc","R6", "scales", "lazyeval", "ggplot2", "corrplot", "RGtk2","cairoDevice","rattle")
   II. install.packages(list_of_packages, repos="https://cloud.r-project.org/", dependencies = TRUE)
   III. # Additional packages
   install.packages(c("tidyselect", "doBy", "ellipse", "mlbench", "amap", "ggdendro", "fpc", "randomForest", "DAAG", "arules"), dependencies = T, type = "binary")
   IV. Install Rattle
   library(rattle)
   rattle()



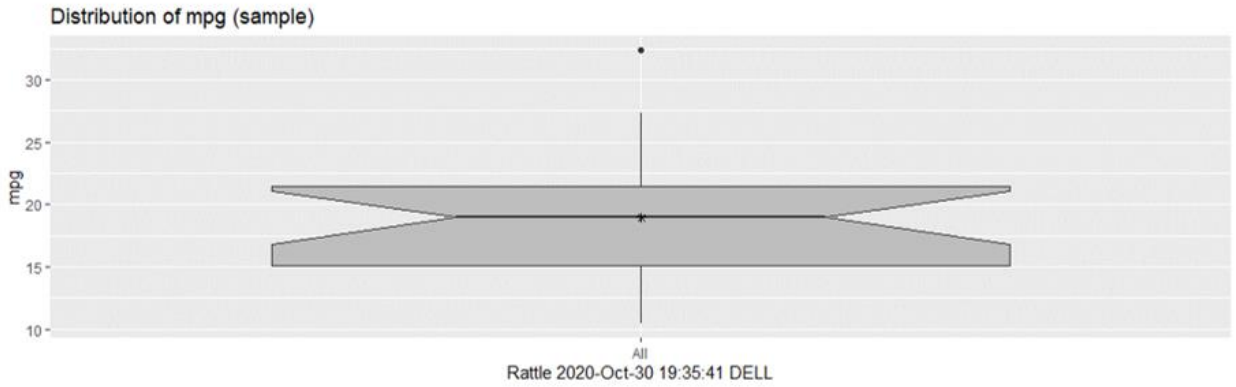
4. Then:
   I. Click on the **Execute** button;
   II. Click on **Yes** within the resulting popup;
   III. Click on the **Model** tab;
   IV. Click on the **Execute** button.
5. Open the RStudio window and execute the command library (rattle),rattle() to get the R Data Miner window.

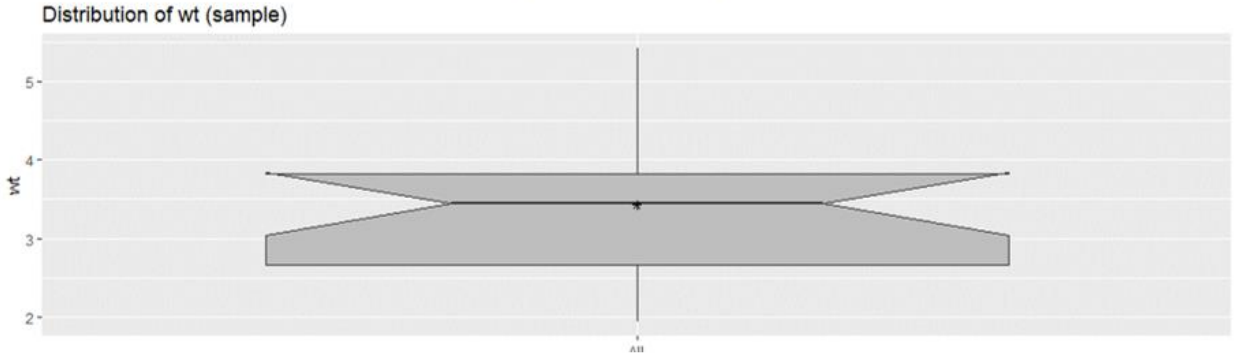




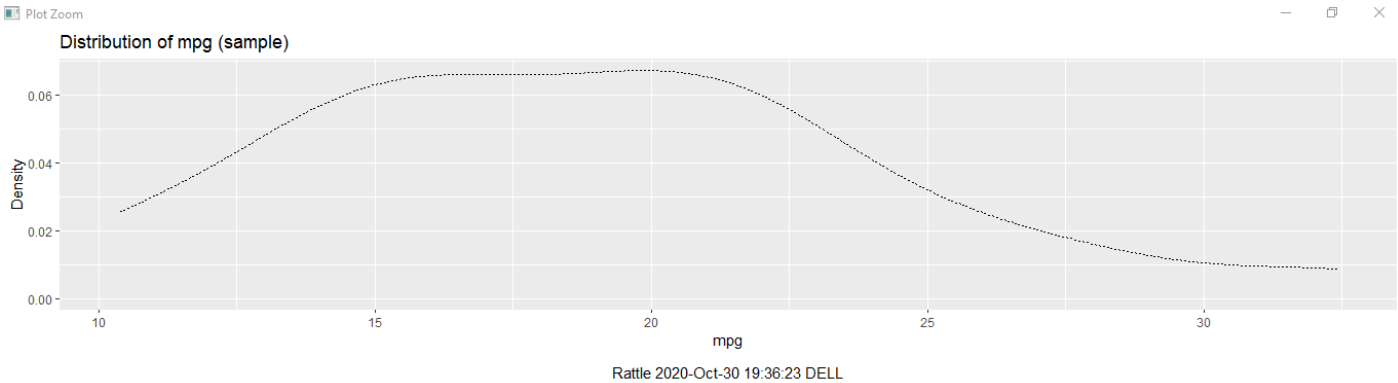
6. After the R Data Miner window opens, from the library of data sets available, select the mtcars:datasets: Motor Trend Car Road Test and execute the data.

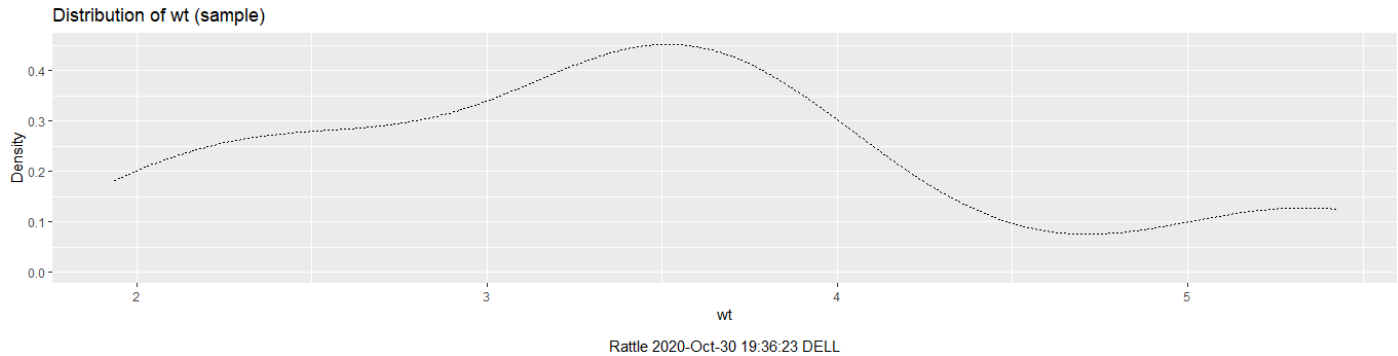
7. Now select miles per gallon (mpg) as a target. Under the explore tab, select distribution from type and tick on boxplot for mpg and wt, execute the selection. The graphs will be visible on the RStudio.



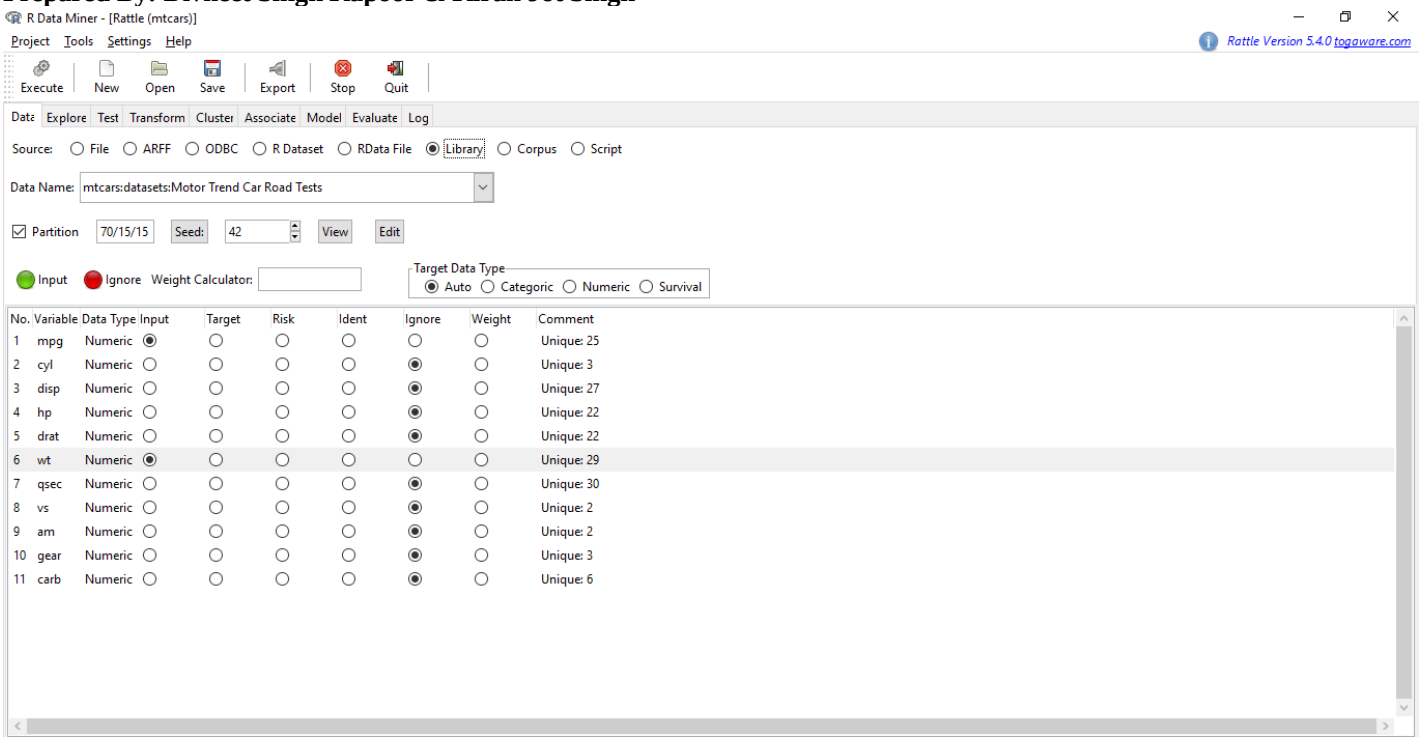


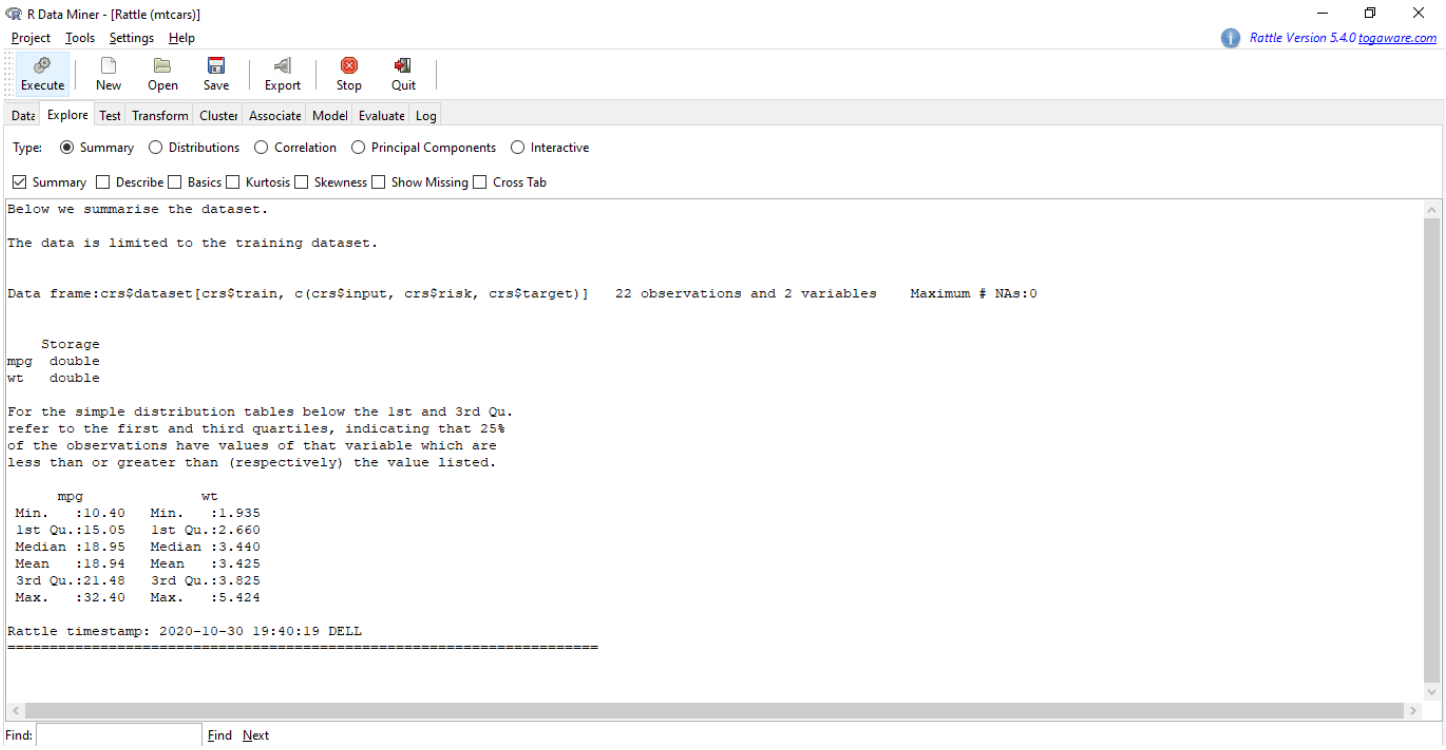
8. Again keeping miles per gallon (mpg) as a target. Under the explore tab, select distribution from type and tick on histogram for mpg and wt, execute the selection. The graphs will be visible on the RStudio.





9. For the summary statistics for mpg and wt, under the data tab select mpg and wt as inputs and ignore the rest of the variables. Execute the data. Under the explore tab, select summary from type and execute the selection.





CONCLUSION :

Rattle continues to undergo development, extending in directions dictated by its actual use in data mining and from suggestions and code offered by its user population.

Prepared By: Divneet Singh Kapoor & Kiran Jot Singh

Thank You Mam