

Worksheet

EXPERIMENT – 8

Name- YANA SRINASTAVA

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

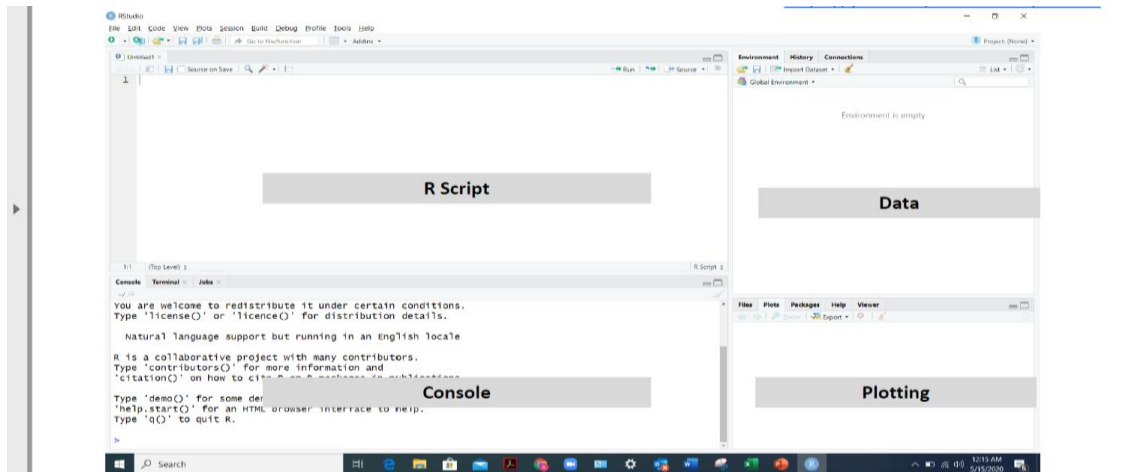
1. Installing R for Windows
 - Go to <https://cran.r-project.org/bin/windows/base/>

Prepared By: Divneet Singh Kapoor & Kiran Jot Singh

- Click on “Download R 4.0.0 for windows”

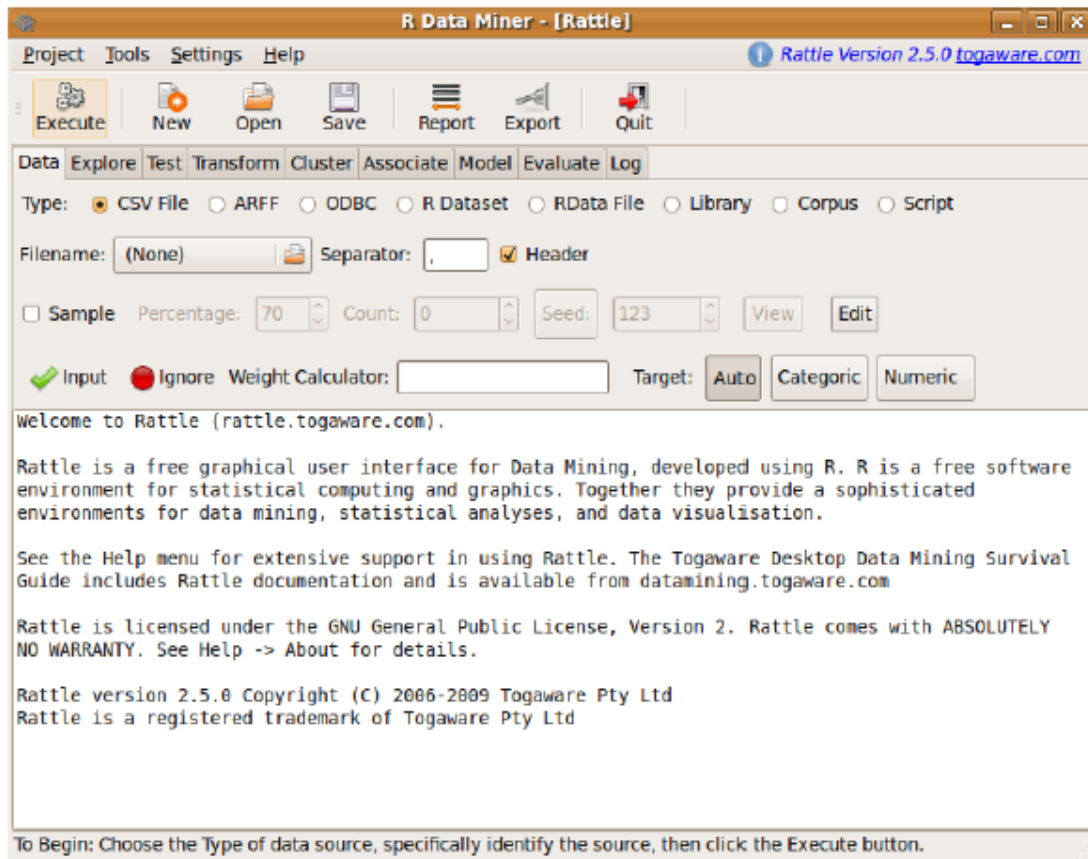
2. Installing RStudio for Windows

- Go to <https://rstudio.com/products/rstudio/download/>
- Click downloaded file to install (select all default options)



3. Install the following packages.

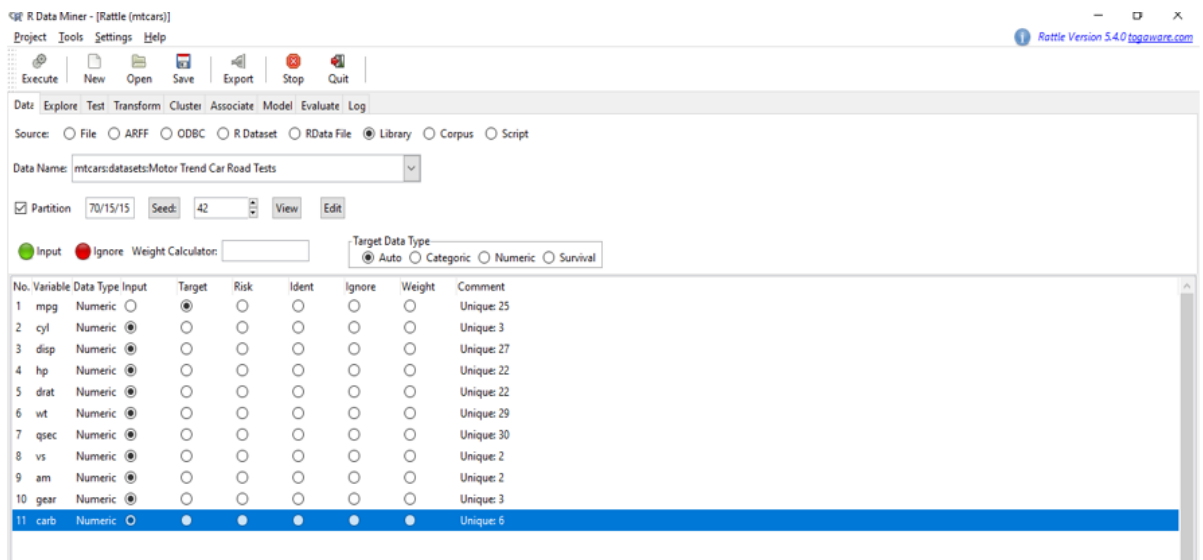
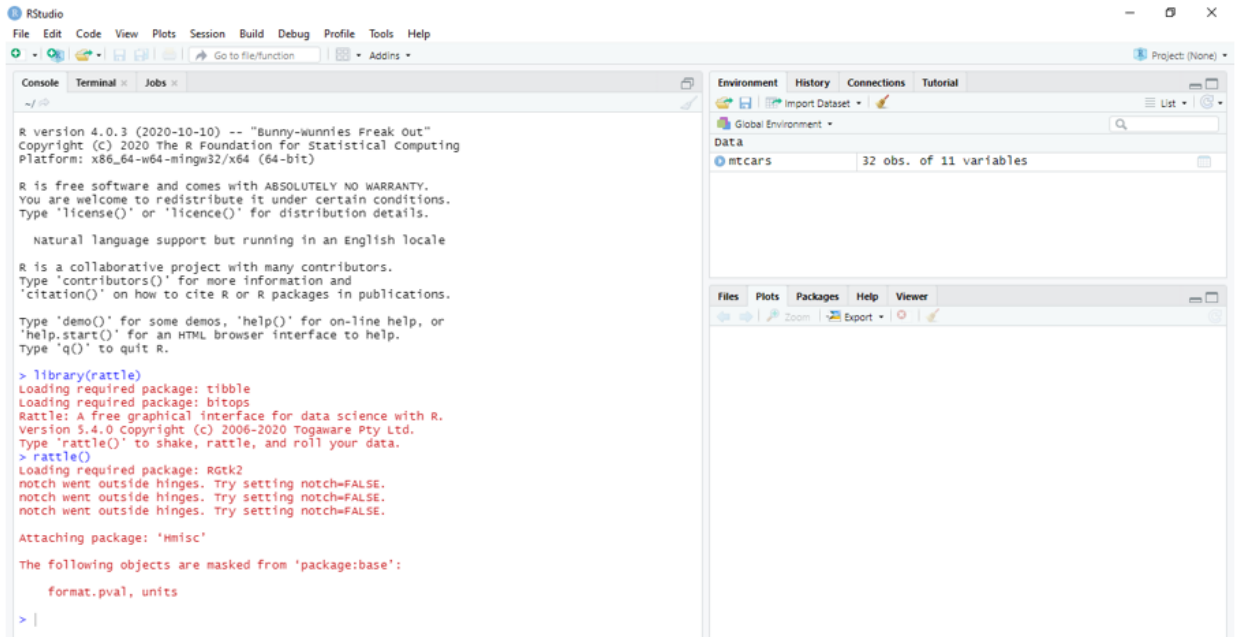
- I. `list_of_packages <- c("tibble", "bitops", "magrittr", "stringi", "XML", "stringr", "Hmisc", "R6", "scales", "lazyeval", "ggplot2", "corrplot", "RGtk2", "cairoDevice", "rattle")`
- II. `install.packages(list_of_packages, repos="https://cloud.r-project.org/", dependencies = TRUE)`
- III. # Additional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



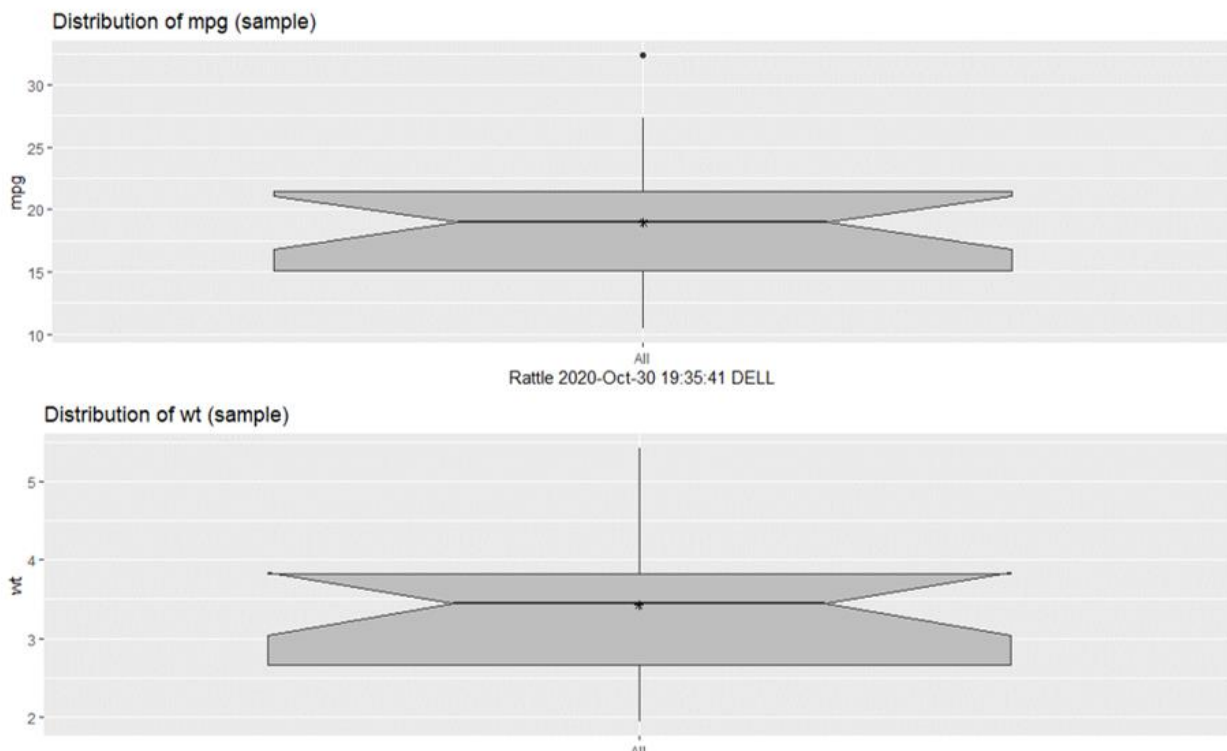
4. Then:

- 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.

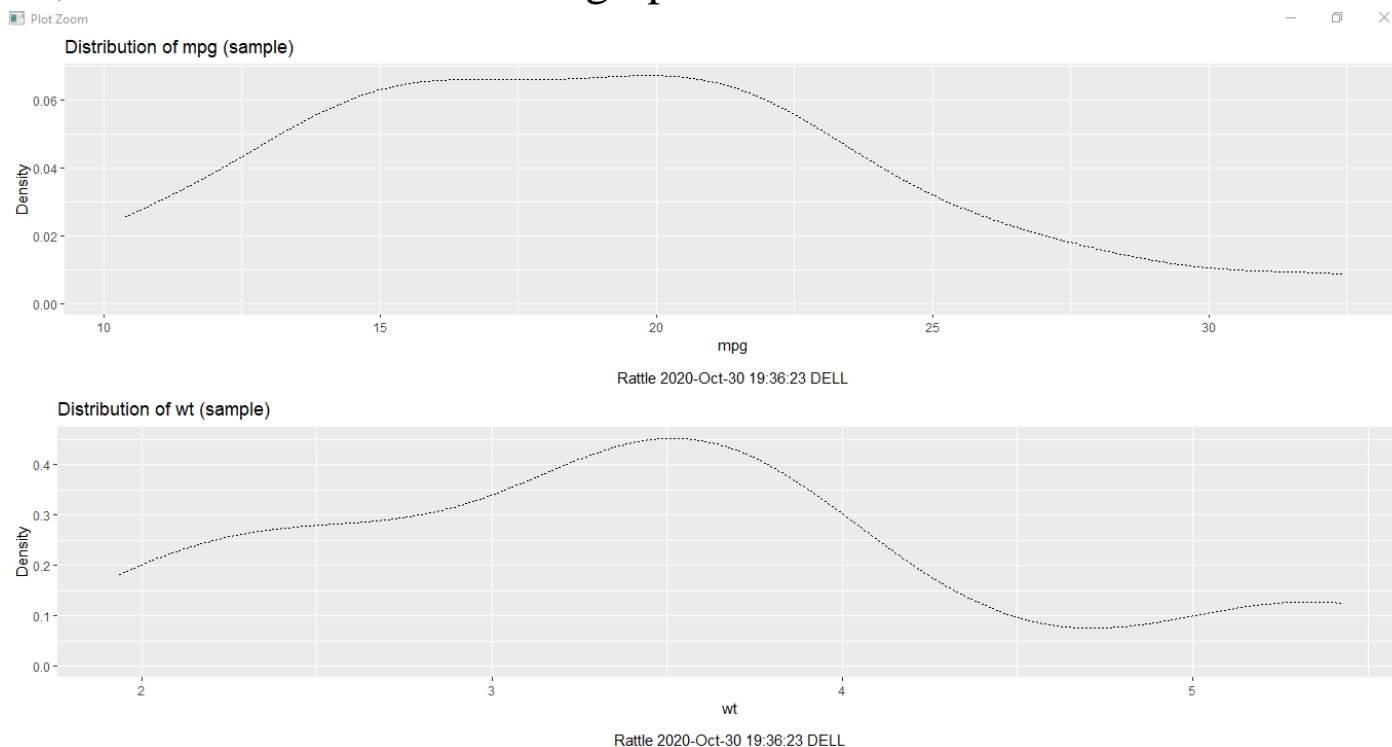
5. Open the RStudio window and execute the command library (rattle),rattle() to get the R Data Miner window.



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.

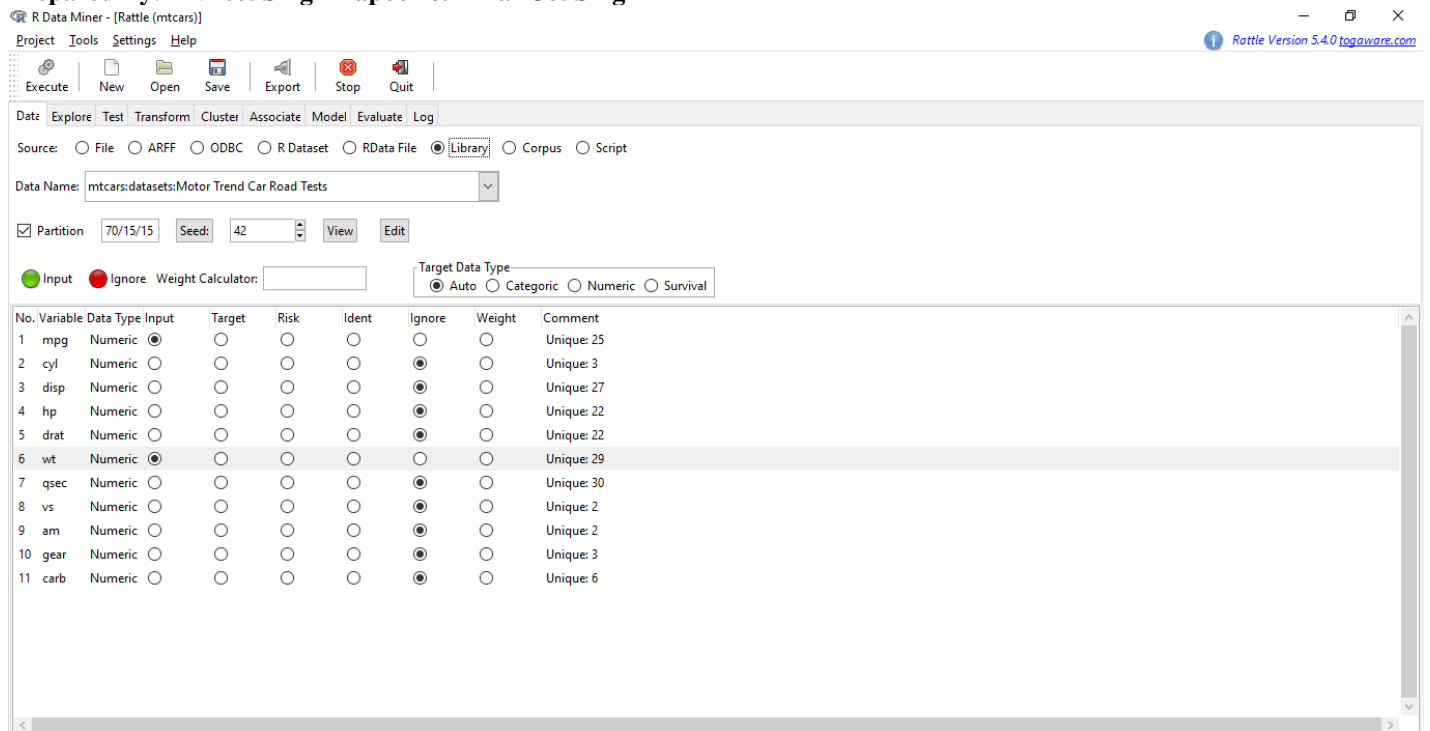


8. Again keeping miles per gallon (mpg) as a target. Under the explore tab, select distribution from type and tick on histogram for mpg and wt, execute the selection. The graphs will be visible on the RStudio.



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.

Prepared By: Divneet Singh Kapoor & Kiran Jot Singh



R Data Miner - [Rattle (mtcars)]
 Project Tools Settings Help

Execute New Open Save Export Stop Quit

Date Explore Test Transform Cluster Associate Model Evaluate Log

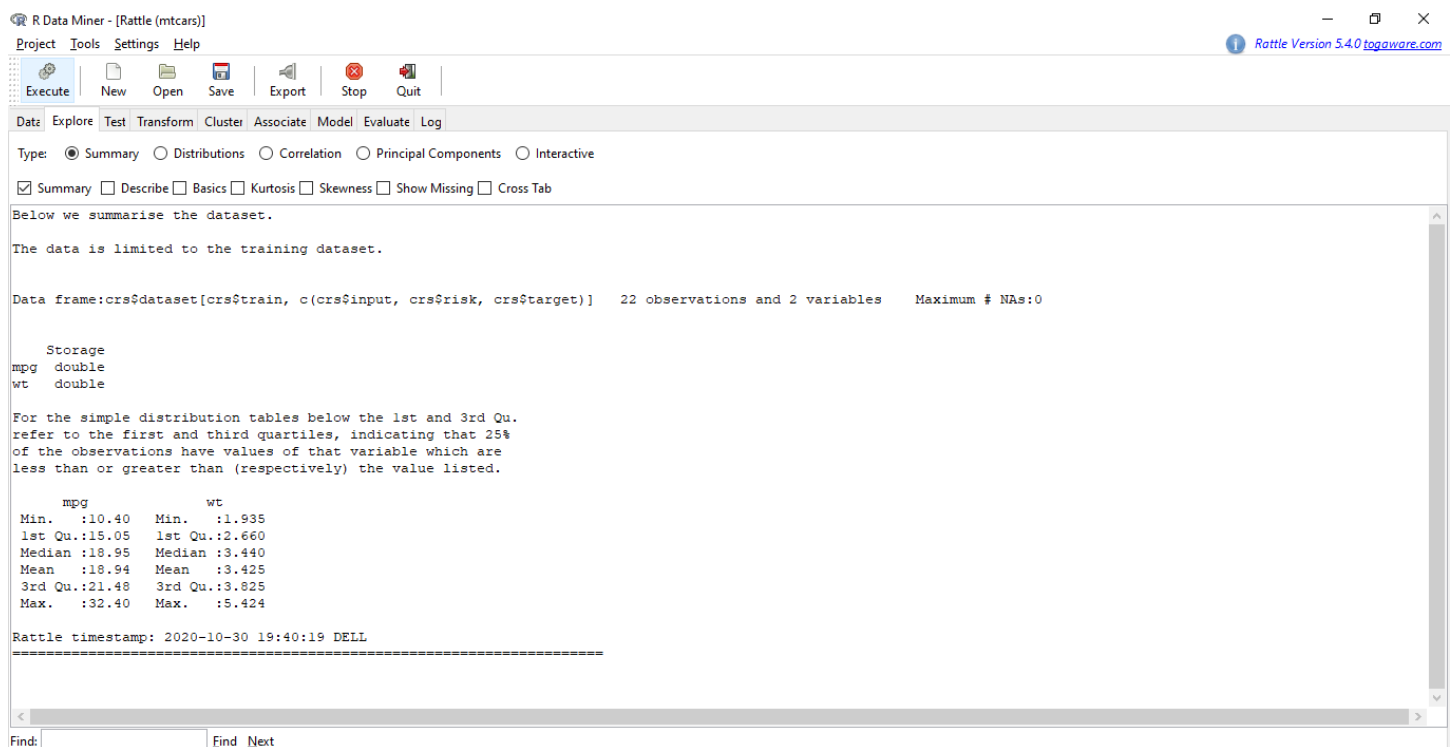
Source: File ARFF ODBC R Dataset RData File Library Corpus Script

Data Name: mtcars:datasets:Motor Trend Car Road Tests

Partition 70/15/15 Seed: 42 View Edit

Input Ignore Weight Calculator: Target Data Type: Auto Categorical Numeric Survival

No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	mpg	Numeric	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 25
2	cyl	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 3
3	disp	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 27
4	hp	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 22
5	drat	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 22
6	wt	Numeric	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 29
7	qsec	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 30
8	vs	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 2
9	am	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 2
10	gear	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 3
11	carb	Numeric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 6



R Data Miner - [Rattle (mtcars)]
 Project Tools Settings Help

Execute New Open Save Export Stop Quit

Date Explore Test Transform Cluster Associate Model Evaluate Log

Type: Summary Distributions Correlation Principal Components Interactive

Summary Describe Basics Kurtosis Skewness Show Missing Cross Tab

Below we summarise the dataset.

The data is limited to the training dataset.

```
Data frame:crs$dataset[crs$train, c(crs$input, crs$risk, crs$target)] 22 observations and 2 variables Maximum # NAs:0
```

```
Storage
mpg double
wt double
```

For the simple distribution tables below the 1st and 3rd Qu. refer to the first and third quartiles, indicating that 25% of the observations have values of that variable which are less than or greater than (respectively) the value listed.

```
mpg          wt
Min.   :10.40  Min.   :1.935
1st Qu.:15.05  1st Qu.:2.660
Median :18.95  Median :3.440
Mean   :18.94  Mean   :3.425
3rd Qu.:21.48  3rd Qu.:3.825
Max.   :32.40  Max.   :5.424
```

Rattle timestamp: 2020-10-30 19:40:19 DELL

Find: Find Next

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