

Worksheet-4c in R

Plotting the Graph using ggplot2

Worksheet for R Programming

Instructions:

- Use RStudio or the RStudio Cloud accomplish this worksheet.
- Save the R script as *RWorksheet_lastname#4c.Rmd*.
- On your own *GitHub repository*, push the Rmd file, as well as this pdf worksheet to the repo you have created before. The repository will be the same as the previous one. There is no need for you to create another one.
- Do not forget to comment your Git repo on our VLE
- Accomplish this worksheet by answering the questions being asked and writing the code manually.

1. Use the dataset **mpg**

A data frame with 234 rows and 11 variables:

```
#' \describe{
#'   \item{manufacturer}{manufacturer name}
#'   \item{model}{model name}
#'   \item{displ}{engine displacement, in litres}
#'   \item{year}{year of manufacture}
#'   \item{cyl}{number of cylinders}
#'   \item{trans}{type of transmission}
#'   \item{drv}{the type of drive train, where f = front-wheel drive, r = rear wheel drive, 4 = 4wd}
#'   \item{cty}{city miles per gallon}
#'   \item{hwy}{highway miles per gallon}
#'   \item{fl}{fuel type}
#'   \item{class}{"type" of car}
#' }
```

"mpg"

- Download and open the mpg file. Upload it to your OWN environment
- a. Show your solutions on how to import a csv file into the environment.
 - b. Which variables from **mpg** dataset are categorical?
 - c. Which are continuous variables?

2. Which manufacturer has the most models in this data set? Which model has the most variations? Show your answer.
 - a. Group the manufacturers and find the unique models. Show your codes and result.
 - b. Graph the result by using `plot()` and `ggplot()`. Write the codes and its result.

2. Same dataset will be used. You are going to show the relationship of the `model` and the `manufacturer`.
 - a. What does `ggplot(mpg, aes(model, manufacturer)) + geom_point()` show?
 - b. For you, is it useful? If not, how could you modify the data to make it more informative?

3. Plot the `model` and the `year` using `ggplot()`. Use only the top 20 observations. Write the codes and its results.

4. Using the pipe (`%>%`), group the `model` and get the number of cars per model. Show codes and its result
 - a. Plot using `geom_bar()` using the top 20 observations only. The graphs should have a title, labels and colors. Show code and results.
 - b. Plot using the `geom_bar() + coord_flip()` just like what is shown below. Show codes and its result.

5. Plot the relationship between `cyl` - *number of cylinders* and `displ` - *engine displacement* using `geom_point` with aesthetic color = *engine displacement*. Title should be “**Relationship between No. of Cylinders and Engine Displacement**”.
 - a. How would you describe its relationship? Show the codes and its result.

6. Plot the relationship between `displ` (engine displacement) and `hwy` (highway miles per gallon). Mapped it with a continuous variable you have identified in #1-c. What is its result? Why it produced such output?

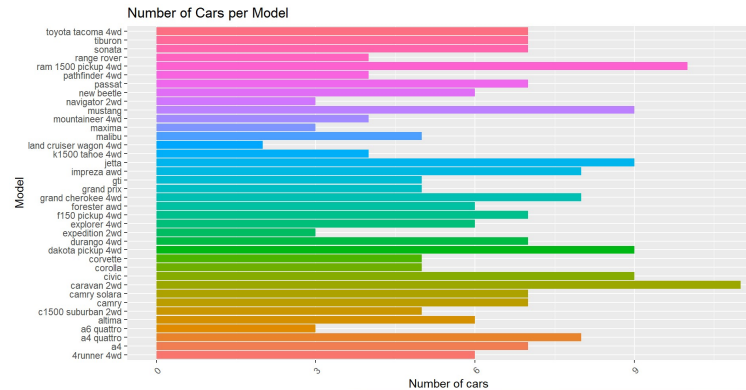


Figure 1: Car Model

6. Import the traffic.csv onto your R environment.

a. How many numbers of observation does it have? What are the variables of the traffic dataset the Show your answer.

b. subset the traffic dataset into junctions. What is the R codes and its output?

c. Plot each junction in a using `geom_line()`. Show your solution and output.

7. From alexa_file.xlsx, import it to your environment

a. How many observations does alexa_file has? What about the number of columns? Show your solution and answer.

b. group the **variations** and get the total of each variations. Use `dplyr` package. Show solution and answer.

c. Plot the variations using the `ggplot()` function. What did you observe? Complete the details of the graph. Show solution and answer.

d. Plot a `geom_line()` with the date and the number of verified reviews. Complete the details of the graphs. Show your answer and solution.

e. Get the relationship of variations and ratings. Which variations got the most highest in rating? Plot a graph to show its relationship. Show your solution and answer.

===== END =====