

DA5020 – Assignment 1

This assignment provides you with practice in R and data exploration. Use R Studio to answer the questions below and clearly label each question in your script.

Use the built-in data set `mtcars`, to answer questions 2-5 below

Question 1 — (5 Points)

Display the number of rows and columns in the dataset. Hint: helpful functions are `nrow()`, `ncol()` and `dim()`.

Question 2 — (5 Points)

Display the first 3 rows of the dataset and display the last 5 rows. Hint: `head()` and `tail()` are helpful functions.

Question 3 — (10 Points)

Inspect the data and identify two categorical variables and two continuous variables. What are their names and what type of data do they contain?

Question 4 — (10 Points)

Type and execute the code below in R studio. Does the code display a chart? If so, explain the chart that was displayed; if the chart is not displayed, indicate what is missing and how would you update the code to show a scatterplot showing the relationship between `disp` and `mpg`.

```
ggplot(mtcars, aes(x=disp, y=mpg))
```

Use the built-in data set `mpg`, to answer questions 6 & 7 below

Question 5 — (10 Points)

Create a scatterplot showing the relationship between `displ` and `hwy`. Color each point based on the number of cylinders e.g. `cyl`. Explain the scatterplot.

Question 6 — (5 Points)

Copy the code for the scatterplot in question 5 and update it to apply `facet_wrap` using the `drv` column. What does the resulting charts indicate about rear wheel drive vehicles when compared to 4wd and front-wheel drive? Hint: use the help function to get more information on `facet_wrap`, e.g. `?facet_wrap`.

Question 7 — (5 points +5 bonus points)

1. What is the difference between an R script and RMD script?
2. Optional/Bonus question: Convert your R script to a Rmd file. Knit the Rmd file and submit it along with the knitted PDF/HTML to receive an **additional 5 bonus points**.

Submission Details

- Name your .R script using the following format: DA5020.A1.FirstName.LastName.R, where *FirstName.LastName* is **your first and last name**. If you submit an Rmd file, you should use DA5020.A1.FirstName.LastName.Rmd
- Ensure that you number each question correctly.
- We must be able to run your code. You will not receive credit for any code that does not run.

Note: For the questions that require an explanation, indicate your response as comments (using #) if you are submitting .R script.