

DA5020 – Assignment 1

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

Question 1 — (+5 Bonus Points)

1. What is the difference between an R script and a knitted R document (i.e. the difference between a .R file and a .RMD file). Submit your entire assignment as a knitted R document to receive an **additional 5 bonus points**.
2. When should you use **str()** versus **summary()**?

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

Question 2 — (5 Points)

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

Question 3 — (5 Points)

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

Question 4 — (10 Points)

Display the column names of two categorical variables and two continuous variables.

Question 5 — (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 6 — (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 7 — (10 Points)

Copy the code for the scatterplot in question 6 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**.

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