

# 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. Students may want to complete the Data Camp "Introduction to R" before attempting this assignment and will earn bonus points by completing it.

## Question 1 — (+5 Bonus Points, All instructions must be followed) (Both questions must be completed to obtain the bonus points)

1. Complete the data camp "Introduction to R". A link to Data Camp will be shared within the course, so that you can access the resource for free using your Northeastern email. Please submit the statement of accomplishment from data camp to earn credit (must also have your name on it).
2. Additionally, explain what a data frame is and provide your own examples using R code. Include this example in your Markdown file.

## Question 2 — (2.5 Points)

Create a dataframe with **3 variables**. The 3 variables should be cars, mpg, cost.

The variable cars should contain 3 inputs. (Truck, Car, SUV)

The variable mpg should contain 3 inputs (11, 30, 24)

The variable cost should contain 3 inputs (45000,25000,35000)

Print the dataframe

## Question 2a — (2.5 Points provide all output)

Select row 1 in column 3, what was selected?

Select rows 1 through 3, what was selected?

Select the last column, what was selected?

**Use the built-in data set mtcars, to answer questions 3-5 below**

## 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. Your submission must contain two files: the .Rmd file and a knitted PDF or HTML (from the .rmd). Name your .Rmd file, DA5020.A1.FirstName.LastName.Rmd and your PDF/HTML DA5020.A1.FirstName.LastName.{pdf,html}, where FirstName.LastName is your first and last name.
- The .Rmd file must be fully commented and properly "chunked" R code and detailed explanations. Make sure that it is easy to recognize which question you answer and that your code runs from beginning to end (because that is how we will test it.) Code that doesn't execute, stops, throws errors will receive no points. If the TAs have to "debug" your code or spend any effort getting it to run, substantial points will be deducted.
- Not submitting a knitted PDF or HTML will result in reduction of 30 points.
- Not submitting the .Rmd file (or both) will result in a score of 0.
- Describe answers for full marks