Data Science and R – Lab 8

Use the **Editor** panel to type your R code and work out the code/output of the following questions.

```
# Loops & Conditionals
```

1) Load any ONE of these datasets from the preloaded datasets in R

`iris` `attenu` `infert` `OrchardSprays` `Theoph`

Code: _____

2) Choose a column with **factored data**. How many different levels are there and how are they distributed (how many observations for each level)?

Code: _____

3) Create a new data frame `df` which is a subset of your data frame with no missing values, i.e. remove all 'NA' values.

Code: _____

4) Using `df`, find the column with the **highest variance**. Iterate over all numeric columns, and keeping track of the largest variance so far and its index (Hint: use `for` and `if-statement-else` constructs)

Code: _____

5) Taking the column with the highest variance, divide it into three class levels. Determine the suitable cutoff values for "low", "average", and "high" or "small", "medium" and "large" by looking at its `summary()`. Create a **new column** in `df` with a suitable name and assign the values using the `ifelse` construct

Code: _____

6) Convert this new column into an **ordered factor** with the three levels that you have created in 5)

Code: _____

7) Find the first numeric column of `df` (use `is.numeric()`). Consider all the values that are less than 10 in that first column of `df` as an error. Filter them out, compute and print the `average` of the remaining values in that column. If all the remaining values are less than 10, then print out "All numbers are less than 10" instead of the average. (Use `for, if, break, next`)

Code: _____

8) This time using `df`, starting from the 1st column, find the first column name with a **mean of less than 10.** Print the **mean and the name of the column**. (Hint: use `break` or `next`)

Code: _____