Data Science and R – Lab 13

Functions:

`function, apply, lapply, sapply`

1) Write a function `toCelcius()` that converts a temperature given in Fahrenheit (**°F**) to Celsius (**°C**). The formula of conversion is:
   °C = (°F − 32) * 5/9

Determine what input(s) it should take and any output(s) that it should return. Call this function with the values in Fahrenheit (°F) given below. What are the equivalent temperatures in Celsius?

50, 77, 86, 98.6, 32, -40

2) Try to now convert Celsius to Fahrenheit by writing `toFahrenheit()` function and rearranging the formula above. Use these values in Celsius to cross check if your conversion to Fahrenheit is correct.

3) Write a function `div13()` that takes a numerical input and returns TRUE if it is divisible by 13 and FALSE otherwise

4) Save the numbers 1 to 500 into an object named `object`. Apply `div13` to all the numbers in `object` using a loop to iterate over `object`. It should print all the numbers that are divisible by 13 between 1 and 500.

5) Write a function `someFn()` that takes in a single number, x and returns a value given by f(x) below:

$$f(x) = \begin{cases} x^2 + 2x + 3 & \text{if } x<13 \\ x + 3 & \text{if x is divisible by 13 (use } \texttt{div13()}\text{)} \\ x^2 + 4x - 7 & \text{otherwise} \end{cases}$$

Apply `someFn()` with these a vector of numbers containing 0, 10, 20, 26. The result should be 3 123 473 29.

PART B: Working with datasets.

Load the `airquality` dataset from base R.

6) Using `apply()`, find the mean of each column. Do all columns have numeric mean? Apply `summary` to all columns to see why.
Code:_____

7) Find the mean of each columns again, but this time remove the NA values. You can pass `na.rm = TRUE` as an argument to `apply()`

Code: _____

8) Using `apply()`, find the minimum and maximum value of each column. You have to remove NA values. (You can use `range` to get the min and max).

Code: _____

9) First using `na.omit`, remove NA values from `airquality` and save it to `data`. Using `apply()`, find all standardised values of each columns in `data` and save the result in `new_df`.(you can supply function as `function(x) (x-sum(x))/sd(x)`)

Code: _____

10) Find the standard deviation of each column of `new_df` and `data` separately. Are they the same? Why?

Code: _____

Answer: _____

11) Remove NA values of airquality and save it as `data`. You can reuse `data` from 8. Using `data`, find the mean of temperature by month. By observing the results, which month has the highest average temperature?

Code: _____

Answer: _____