**MSBA SUMMER 2020 R PRACTICE QUESTIONS**

Use R Studio and the starter script “PracticeStarter.R” to perform the following actions in R. You should develop an understanding of the code during the process (e.g. you should be able to recreate the code if asked to do so). Additionally, your answers should conform to standard coding conventions (well commented code, intuitive variable names, etc.). Keep in mind there are many ways to get to the same answer! Also, StackOverflow and Google are your friends. You can email any questions you have about this assignment to [jake.arendsen@emory.edu](mailto:jake.arendsen@emory.edu) and/or [harshit.arora@emory.edu](mailto:harshit.arora@emory.edu).

Finally, we want to acknowledge that this is a slightly difficult assignment, so please do not get discouraged if you run into troubles while performing this. Realistically, this assignment is about at the level we’d expect students to be at the end of the summer, not the very beginning. If you find something just too difficult, just skip it and return to it later. With the exception of question 10, no other question has anything later that depends on its completion.

1. Evaluate each of the following expressions in R
   1. 137 + 614
   2. 39 – 81
   3. 134
   4. Calculate the remainder of
2. Install and load the ggthemes package
3. Using a loop of your choice, perform the following actions
   1. Print “Emory Rocks!” to the console 10 times
   2. Print the names of the countries in rows 610-615 in the gap table (the table that was hardcoded and called gap in the starter script).
4. Set a to be a random number between 1 and 10. Then write an if else statement that prints even if a is even and odd if a is odd.
   1. Now, using the ifelse function, create a string that stores the value to be printed. Then print the string to the console.
5. Write a function that does the following: taking in a single argument x, it should return “positive” if x is a positive number, “negative” if x is a negative number, “0” if x is equal to 0, and “Error: Not a number” if x is not a number. Test your function with at least one number and one non-number.
6. Create a copy of the gap table called “copy”. Confirm that copy and gap are the exact same.
7. Calculate summary statistics (mean, median, min, max, etc.) of the population column of the gap table.
8. Create a simple linear regression model calculating GDP per capita as a function of life expectancy using the gap table. Interpret anything relevant from the output of the model. Additionally, calculate the correlation between these two values.
9. Using ggplot, create a line graph of population over time for Afghanistan. Be sure to properly label your axis and title your graph.
10. Load in the ‘Workers’ & ‘Title’ sheets from the MSBA SQL Tables file from the SQL assignment.
11. Join the Workers and Title tables together
12. Calculate the number of unique Worker Title-Department combinations present in the data.
13. Perform the following manipulations to the joined table
    1. Create a full name column that contains a worker’s full name
    2. Filter the table to only include employees who joined prior to August 2019
    3. Order the table by Salary, least to greatest
    4. Remove all columns except full name, Salary, Title, and department
14. Save the manipulated table as a CSV to your computer

**BONUS:** The excel file BonusFile.csv contains a transformed version of the gap table from this assignment. Perform the necessary data manipulation to transform the gap table into the table as it appears in the excel file. The following tips may prove useful.

* GDP = gdpPercap \* pop
* Finland is not present in the Excel file
* relativePop is a countries’ population divided by the countries’ max population in the dataset

**DOUBLE BONUS**: Write your code to perform the bonus in such a way that it can be performed with a single click of the run button (hint: the pipe operator (%>%) will be essential to performing this)

**HELPFUL RESOURCES**

ggplot:

<https://ggplot2.tidyverse.org/>

<https://rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

<http://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html>

dplyr:

<https://dplyr.tidyverse.org/>

<https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>

<https://4.files.edl.io/b9e2/07/12/19/142839-a23788fb-1d3a-4665-9dc4-33bfd442c296.pdf>

<https://medium.com/@brianward1428/introduction-to-tidyverse-7b3dbf2337d5>