Northeastern University

Course: DA5020

Assignment: Module 2 - Programming in R

Total Points: 100

Date Due: Posted on Blackboard

Learning Objectives

In this assignment, you will learn how to:
- read and write simple text files
- load data selectively
- control conditional and iterative execution
- create data frames and vectors
- make R code readable

Tasks

Before diving into the programming problems, study the data files that are provided for the assignment:

1. (10 Points) Load the data file into an appropriate data object of your choice. The file is compressed, so you can either uncompress the file after downloading or use one of the various R functions to load a zipped (compressed) file. Look at the other questions to determine what object is most appropriate: data frame versus vector? Figure out how to deal with missing values. Add your strategy as a comment to the functions that you write in the next few tasks.

2. (5 points) Assume your R script may be run multiple times. You do not want R to read the data file into your environment every time you run the script since this is expensive. If the data object already exists in your environment, then your script should recognize this and not reload the data file. Optimize your code and make sure the data file is not reloaded repeatedly.

3. (30 Points) Write a function called `TotalNumDelays(Carrier)` that finds and returns the total number of delays of a carrier. You must calculate the number of delays, not the total number of minutes all flights were late in aggregate.

4. (30 Points) Write a function called `TotalDelaysByOrigin(Origin)` that finds and returns the total number of delays for a particular airport.

5. (25 Points) Write a function called `AvgDelay(Carrier,Dest)` that calculates and returns the average arrival delay for a carrier flying into the specified destination airport.

Clarifications

ADDITIONAL DATA DESCRIPTION: Each record within the AirlinesDelays.txt is for a specific flight. There are seven columns that represent a delay for a flight: DEP_DEL, ARR_DELAY, CARRIER_DELAY, WEATHER_DELAY, NAS_DELAY, SECURITY_DELAY and LATE_AIRCRAFT_DELAY. You can tally the number of delays for a flight by counting the number of these variables that are greater than 0.

CODING STYLE: Part of the assignment is to practice good coding standards: consistent naming, formatting with indentation, testing input values, dealing with errors, commenting on the code, etc.. You need to submit an R script, which is a text file having an .R extension, i.e., To help you and us read the code you must follow the coding standards mentioned above. This applies to ALL assignments going forward. If your code is not readable and/or does not run 20% will be automatically deducted from your assignment grade.

CODE FUNCTIONALITY: The functions must return the result as a return value which is then printed from the calling function or main script. This makes the function truly a "function" versus a "procedure". Functions return values, procedures do work and do not return values. Returning a value also makes it easier to deal with error conditions. For example, you could return -1 if the data could not be found. The calling function can then test for that condition.

The file can either be uncompressed after it's downloaded or you can use various R functions to load a zipped (compressed) file.

Deliverables

You need to submit an .R extension file. Be sure to state all the assumptions and give explanations as comments in the .R file wherever needed to help us assess your submission. **Please name the submission file LAST_FirstInitial_2.R** for example for John Smith's 2nd assignment, the file should be named Smith_J_2.R.