Northeastern University

Course: DA5020

Assignment: Module 3 - Data Shaping - B

Total Points: 100

Date Due: Posted on Blackboard

Learning Objectives

In this assignment, you will learn how to:

- read and write CSV files
- load and convert dates
- calculate with dates
- summarize date data

Tasks

Before diving into the programming problems, study the data files that are provided for the assignment.The files contains dates and related data on birds strikes to aircraft.

1. (20 Points) Load the data file "BirdStrikes" into an appropriate data object of your choice. How many bird strikes did not have a “Reported: Date” assigned, *i.e.*, for where there is no value for "*Reported: Date*".

2. (20 Points) Which year had the most bird strikes? Write a function to calculate.

3. (20 Points) How many bird strikes were there for each year? Place the result into a data frame.

4. (20 Points) Write a function that calculates the number of birds strikes per airline and then put those results into a dataframe called AirlineStrikes. Write another function that accepts the dataframe AirlineStrikes as an argument, and returns the airline that has the most bird strikes. (Note: UNKNOWN occurs most frequently but is a marker for missing data. Your function needs to return a valid airline, i.e. the second most occurring value.)

5. (10 Points) Comment on the time and space complexity of your functions. What would happen if the data set were 2x, 10x, 100x, 1000x bigger than it is now? How would that affect memory use and run time of your code? Express your complexity estimate using big-O notation. Write the comment in the Comment field of the submission.

6. (10 Points) Choose one of your functions and use `system.time()` to measure the execution time for the original sized data, 2 times the original size and 4 times the original size by duplicating the data set. Do the measured results match your expected results reported in answer 5? Can you explain why the answers may vary?

Note

Create a function for problem 1 to 4, only then you will be able to do problem 6.

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_3B.R** for example for John Smith's 3rd assignment, the file should be named Smith_J_3B.R.