Northeastern University

Course: CS6020

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. Load the data file "BirdStrikes" into an appropriate data object of your choice. The files contains dates and related data on birds strikes to aircraft.

1. (20 Points) How many bird strikes were **not** reported, *i.e.*, for which 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) Which airline had the most bird strikes? Write a function to calculate the number of birds strikes per airlines and then put those results into a data frame. Determine the answer from the data frame by writing another function to which you pass the frame. (Note that the most bird strikes could not be associated with an airline, so UNKNOWN is actually the most common, but that's not an actual airline but a marker for a missing value; in that case report the second most as it's an actual airline.)

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 now? How would that

affect memory use and run time of your code? Can you express your complexity estimate using big-O notation? Write the comment in the Comment field of the submission.

6. (10 Points) Use `system.time()` to measure each function and then experiment if you were to double and then quadruple the input file size. Can you capture this data in a data object and plot it? From the plot can you surmise its growth function? How about constructing a regression model?