Homework 5: dplyr

Part 1: Look at the hflights dataset and find the percent of flights for each airline that were cancelled by the airline. Put these in order from the highest to the lowest percentage. Your result should be two columns – one containing the name of the airline and the other the percent of flights they cancelled. *(2 points*)

Part two: Download into your Data folder the Complications-Hospital file from the government website <https://data.medicare.gov/data/hospital-compare>. Read the file into R (use read\_csv). How big is the file (rows and columns)? Describe the dataset. (*1/2 point*)

Pick a complication that you would like to study and write all rows for that complication to a .csv file. Only include the following columns: Hospital Name, City, State, Compare to National, Score, Lower Estimate, and Higher Estimate. Explain your results. *( 1 point*)

*(2 points*) Write a function called best<- function(state, complication) that takes as input the two letter state abbreviation and the complication (Measure ID) and outputs the hospital in the state with the lowest score. Check for valid state and complication and give appropriate error messages if invalid input is given. Test your function by making the following function calls:

* best(“DC”,Complication="PSI\_4\_SURG\_COMP")
* best("AP", Complication="PSI\_4\_SURG\_COMP")
* best("DC", "pneumonia")

*(2 points*) Write a function called rankhospitalsinstate <- function(state, complication, n) where n is the number of hospitals to include in the ranking. If n is not passed in, rank all the hospitals in the state. If n is larger than the number of hospitals in the state, rank all the hospitals in the state. Test your function by making the following function calls:

* rankhospitalsinstate("VA","PSI\_15\_ACC\_LAC",10)
* rankhospitalsinstate("VA","PSI\_15\_ACC\_LAC",500)
* rankhospitalsinstate("VA","PSI\_15\_ACC\_LAC")

*( 2 points*) Write a function called rankallhospitals <- function(complication, num=1) which outputs the top num hospitals in the country for the given complication. Test your code by making the following function calls:

* rankall("PSI\_12\_POSTOP\_PULMEMB\_DVT",10)
* rankall("PSI\_14\_POSTOP\_DEHIS")

The R markdown document will produce a .Rmd and .doc or .pdf file. Submit both of those files to me electronically. When you run the R markdown, a by-product should be a .csv file that you write out. Please also submit that file.

Here is a suggested outline for your Markdown document:

Part 1 – answer the Part 1 question.

Part 2: medicare dataset portion:

Introduction - describe the dataset and the question of interest.

Data Acquisition and selection – this is where you import the data and document where the data was obtained. Keep only those columns needed for your analysis. Select a complication and write out a file that contains those rows and columns desired.

Data analysis – include a description of each function. Run each function with selected test cases.

System Environment *(1/2 point*) - use sysInfo() command.