**DS 710**

**Homework 10**

**R assignment**

1. In this problem, you will work with the “cleaned” version of the US News and World Report data on colleges and universities, which you created in Homework 9.
2. Read the data into R and attach it.  Use *length* and *which* to determine how many schools have a per-student instructional expenditure higher than their out-of-state tuition.  Then use control flow to answer the same question.  (Check that the two methods give the same answer!)
3. Use *system.time* to compare the running times of the two methods you wrote in part a.  Iterate each method enough times that you can see a difference in the running times.  Report the user time + system time for each method.  Which is more efficient?
4. Consider three different methods of finding the mean of each numeric column of the data:
5. Using apply() and the built-in function mean()
6. Using apply() and a function you write, called mymean(), which takes the sum of all of the non-missing values and divides by the number of non-missing values, without using the built-in function mean().
7. Using a for() loop to iterate over the numeric columns, and a for() loop inside it to iterate over the values within that column, without using the built-in functions mean() or sum().

Which do you expect to be most efficient?  Explain your answer in 1-3 sentences.

d. Write functions for each of the 3 methods in part c. Apply them to the US News and World Report data and check that all 3 methods give the same answers.

e. Use *microbenchmark* to compare the median running time of the methods in part c.  Write 1-3 sentences describing which method is most efficient.

* It’s OK if the result doesn’t match your prediction in part c, as long as your prediction was well-justified. Different people may get different results, depending on details of your code.
* If you get the warning message, “Could not measure a positive execution time,” double-check that you included parentheses at the end of each function call.

Submit a .docx, .pdf, or .r file to GitHub, containing your R code, R output, and written interpretations and explanations.