The Power of Efficiency

Fiona Norton

One of my major takeaways from this course has been the importance of efficiency in statistical computing. Efficiency is writing code that does not repeat functions and outputs the information needed in as few steps as possible. In the context of statistical computing and data science, this means not introducing intermittent objects or assigning information to an object unnecessarily. It also means combining steps. For example, when we first learned how to use the five core dplyr functions, we would create separate code chunks for each time we use one of these functions and assign the results back to the original dataframe. Now, we understand that this is a very inefficient way of manipulating data; instead, we can use the pipe operator and combine all the steps into one code chunk. A good example of this is in Lab 9 question 1.

Efficiency allows us to utilize built-in functions or write our own functions to get the information we are looking for in as few steps as possible. This is important because it makes code easier to understand and replicate. If we write inefficient code, there is a lot more room for error. Additionally, it is important to be efficient when writing functions because if you repeat steps within a function, it can cause errors and confusion for anyone else who might try to use your code. Finally, efficiency I most important because it saves time!

I have encountered efficiency in almost every assignment of the quarter, and often received feedback on labs regarding potentially inefficient code. Some of the most eye-opening moments in my understanding of efficiency included Lab 4 question 5 where I realized how I could use optional arguments within functions, like "n = ...", to get the exact information I need without including another step or getting an output that is not exactly what I need. Another one of these "a-ha" moments came in Lab 7 question 3 when, after receiving feedback from Professor Theobold, I realized that it would be more efficient to assign the max and min to objects so that I would not have to repeat those functions and the information would be saved in the previous step.