The Power of Efficiency

The word “efficient” is typically defined as achieving a maximum desired outcome while expending the least amount of energy or effort possible. While efficiency is a very broad concept, it can particularly be applied to statistical computing and data science. In addition to writing code, data scientists also share code with one another. When a person writes code to be shared with others, there is an expectation that the code is concise and easy for another person to follow. This is where efficiency comes into play.

In statistics, computer science, and data science, there is a well-known principle that appears frequently when writing code, which is Don’t Repeat Yourself (DRY). It refers to writing the least amount of code possible to accomplish a task. This includes creating functions instead of writing the same lines of code down multiple times. Without efficiency, programs can become extremely wordy and difficult to follow. Writing efficient code also saves time on the programmer’s end, as there is less to write and less room for error than if multiple lines of code for the same task are written.

Throughout my time in STAT 331, I have encountered efficiency in various places. Labs 7 and 8 were particularly prevalent for efficiency, as during Weeks 7 and 8 I learned about writing functions and iteration. Question 1 of Lab 7 required us to report the sum of missing values across all variables in the Blackfoot Fish csv file. Rather than using the sum function on each individual column, I used the across function to take the sum of missing values for all columns. This condensed the amount of code I had to write to just a single line, which would still run through any column I specified. I also used efficiency in my code in question 3 of Lab 7, where I created a function that rescaled a function by taking every input, subtracting the minimum value of the vector, and dividing it by the range. By creating this function, I was able to call it multiple times to complete the other tasks of the lab, rather than rewriting the same rescale equation every time I wanted to scale a value. I continued to learn more about efficiency through iteration in Lab 8. In question 7, I used the pmap\_chr function to apply my make\_phrase function to each day number, item, verb, adjective, and location in the Twelve Days of Christmas Song. Learning these concepts and discovering how to create functions and iterate through a list has made my experience as a data scientist a lot more smooth.