

The Power of Efficiency

Jake McDonald

Code efficiency is important to all data scientists and can mean a myriad of different things. Generally, efficiency in code is when the coding methodology uses very little computing power to execute. In R, there is the base code that comes with the language and there are also various packages that one can add on to get access to functions and lines of code that people have made within R. The one our class focused on, Tidyverse, allows us to do various syntaxes and functions that make our code more efficient not only for the computer, but the human reading the code as well. These things include the pipe operator, which allows us to commit various functions to the same object. Additionally, they provide functions that aid with data wrangling, factor type variables, string variables, visualization, data-sets, and date variables. These functions are more efficient than base code R because they do the things the data-scientist want by using more condensed code. It makes it easier for the human to read the code because the syntax of the language is easier to digest.

Another way a data scientist can make their coding more efficient than other packages is creating functions that utilize iteration. Iteration is when the computer performs the same task until it is complete. Without a function, a data scientist would have to repeat the same lines of code until the desired task is finished. A function, however, allows the data scientist to iteration in the function so that the function itself runs that same line of code until the job is done. An 'a-ha' moment that I had while learning iteration was Lab 8, which can be seen in the supporting artifacts section. This was a very difficult lab for me to understand, but when I finally was able to get it to iterate the lyrics of the song, it was a great learning moment. Additionally, learning how to create a function, which is used in another created function.