The Power of Efficiency

To me, I think efficiency is essentially the process of being able to produce something faster in comparison to ones self. I think the ones self is an important note to keep in the definition especially in a data science role as there is a huge discrepancy in skill levels especially in something like R. Examples of being efficiency could lie at every one of these levels. Maybe for example, a beginner, wanting to be able to select certain features from a data set and instead of manually selecting those observations, they might use the filter() function. A intermediate skilled statistical user might create there own function to be able to repeat some sort of calculation they would like to perform on a data set. An advanced statistical user could try and optimize their own code to make it run milliseconds faster by condensing it down. Every person has this sort of idea to make ones code more efficient, however I think it is much more important of an idea that no one is completed their programming in terms of efficiency. At the end of a project, everyone could strive to improve their own work and having that motivation I think is a key to learning data science and statistical computing in general. Efficiency is important especially in the data science realm as if given unlimited time, I feel like anyone could accomplish any task given to them. Anyone could spend days trying to manually find summary statistics about certain data. However, I feel like nobody would really want to do that and so instead they try to find ways to speed up the process. I think this also connects to general knowledge about learning R and its fundamentals. if a slow task is consistently being used in someones everyday lives, it will come to a certain point where someone will go out of there way to learn it in a faster and more meaningful process.

I think I have most certainly faced a couple of "a-ha" moments in this quarter. I think a fundamental example could be shown in using the across functionin my Cleaning Data section of Lab 3. I went from copying and pasting every single categorical variable with a mutating as.factor() to using one mutate with an across. I think it is also important to note that I was not really hesitant in learning this new way of being efficient. I think this overall stems from the fact that with learning something new, I did not really understand nor memorize it the first try. However, with doing it multiple times (through reflections as well), it finally clicked with my and now towards the ending labs such as Lab 7, an across function is being used. A second example of this "a-ha" moment would be the classic use of creating a function. This is highlighted clearly in Lab7 with using it to create new re-scaling measurements. Of course, initially I could just mutate this certain calculation on multiple different vectors, however,

that is not beneficial to ones self. I think also with the creation of functions, it opens a new idea of being efficient within a function which creates a new strive to improve ones self on efficiency-wise. The last example that I feel like I gained with efficiency was something that I did not really realized until the end of the quarter until I started looking at the polished HTML reports I was submitting with my labs. I think looking at the HTML from someone who has no clue what the lab is about is incredibly inefficient. In multiple sections, I include warnings that should not be there, or have multiple tibbles that make no sense on its own. I think having this perspective of your own final work being visible efficient is important to, especially for the future in which everyone will be looking at the work that you will do. Towards the end of the quarter, Lab 8 for example (attached HTML), my lab reports look a lot more polished, allowing the reader to clearly and quickly look at what I am addressing with my code. Overall, efficiency is something I have dealt with a lot throughout this quarter and it is something I hopefully will try to improve on in the future.