Growth as a Data Scientist

Based on my (albeit limited) experiences, a great deal of data science work seems to involve transferring and altering datasets. Prior to this course, I had only done this in MATLAB and did not have much experience with post-processing my work once a dataset was edited into a usable form. I also tend to become disorganized, and this course helped me not only learn the importance of keeping good organizational habits, but develop specific skills to maintain easy-to-follow files and projects.

Importing data into R's workspace and filtering necessary information was such a common part of this course that it became commonplace. Prior to taking this course, however, these were not skills that I was completely confident with (in any programming language). While I have previously worked with several different types of datasets, processing data to make it work for me and my team always seemed like an arduous process, and it was one that I often dreaded. After these last ten weeks, where I completed that process almost twice per week, I feel much more confident both importing different types of data and altering them to be more useful. In "Lab 7_Revised" I demonstrated my ability to process imported data by using functions such as arrange, mutate, and slice_tail to take a cumbersome dataset and filter out only the data I needed. Prior to this course I would often leave unnecessary information in my dataset and would have trouble displaying the information which I did use. Labs such as lab 7 helped me grow these skills, which I believe will make me a better data scientist moving forward.

This course also taught me the importance of maintaining clear organization. I often end up at the end of the quarter with a folder on my computer full of jumbled file names and no well-organized subfolders. That was not a possibility with this course, and looking

back at my work this quarter I can see why; I keeping consistent file path and name structures makes it incredibly easy to share work and navigate to important files. If I want to review a specific assignment from this course, I never have to spend time wondering where it might be. This also makes it easy to share my work with other people, because the organizational structure is apparent enough to understand without spending much time to figure it out. Lab 8 exemplifies this ideal type of organization and carries it into the file itself; each code chunk is given a header, and completes one task, making it easy for the code to be read even by people who were not present when it was written.

This course taught me that data science is much more than just working with large datasets and writing chunks of code to make graphs. I learned the importance of reducing data, processing it to make it easier for your code to handle, and of maintaining clear organizational habits to make your work both reproducible and easy to share.