

My Growth as a Data Scientist

Fiona Norton

I have progressed as a student and practitioner of data science by improving my ability to write tidy code, learning to create thoughtful, digestible data visualizations, writing functions, and bettering my understanding of the tidyverse and all the packages it has to offer. This course has instilled in me a habit of always writing tidy code with named arguments, new lines to prevent character wrapping, and white space around commas, arithmetic operations, and the pipe operator. Tidy code is very important to being an effective data scientist because it allows others to read, understand, and replicate your code without errors. My ability to write code in this way can be seen in Lab 5 question 8 and Lab 9 question 3.

Learning how to use the dplyr package to manipulate data has been transformative for me as a data scientist. Since week 3, when we were first introduced to data wrangling and tidy data using dplyr, every activity and lab has improved my ability to filter data, create new variables, and manipulate existing variables. I think some of the problems that best display these skills are questions 2 and 3 in Lab 4. Similarly, since being introduced to ggplot, each week we have learned new features of the function to create more organized, better looking data visualizations. This is shown in Lab 5 question 1, where I revisited and revised an old graph using new skills, and Lab 5 questions 5 and 7, where I use all of the skills I've learned within ggplot.

Writing functions has also been a major step for me. Prior to this course I had only used R for statistical and econometric analysis using the built-in functions, or simply used it like a calculator. Now, I know how to write a function to reduce repetition in my code, I know how to write a vectorized function that can be applied to multiple variables in a data frame, and I understand how to use the map() family for iteration. These skills can be seen in Lab 8, where I was able to write a function to write the "12 Day of Christmas" song using my own functions within the larger function and then iterating through a dataframe with the information for the song.

This course has supplied me with tools that have drastically increased my knowledge of how to use R for data science. I feel much more confident in my ability to organize and analyze data and gather helpful insights.