My Growth as a Data Scientist

Camille Heathcock

I have grown tremendously as a data scientist in this class this quarter. My main takeaway from this class is that I have learned how to think like a data scientist in general. Coding is quite literally another language that one must understand fundamentally in order to successfully write code. I have grown in the sense that it is much easier for me to manipulate data using filtering, selecting, and grouping variables to achieve what I want– specifically using functions from the dplyr package. This area of growth is represented throughout my labs and assignments, but is especially reflected in the last lab we completed, Lab 9. In Lab 9, I was able to use filtering, grouping, and mutating functions within one pipeline operator seamlessly. This can be shown in #1 in part 3, as well as throughout the rest of Lab 9.

Another thing I have taken away from this class is the value of tidy code. I appreciate how much tidiness was emphasized in this class because it has made my code much more organized and readable. I love the format of Quarto documents and how they are separated into code chunks, as well so easily rendered into an HTML document. I also really appreciate the emphasis on efficiency and reproducibility. I will talk about efficiency more in the next essay, but these two concepts have really solidified my understanding of how code should be written. A good artifact for this area of growth would be my revised Lab 8. In Lab 8, I thought I did a good job of separating code into different code chunks, using whitespace around all commas and arithmetic operators, using new lines to eliminate character wrapping, as well as ending the line with every %>% or |> sign. All of these are demonstrated in the 'phrase-function' code chunk.

The last thing I have taken away from this class is how to create an effective visualization and how valuable that can be for understanding data. The "not having to turn your head sideways" idea really clicked for me, and that is something I will try to implement in all visualizations I make in my future. Effective visualizations are so important for data scientists to be able to portray their findings in a digestible format, so I think this is a really important skill that I will continue to use throughout my career. A good artifact for this area of growth would be from Lab 9, #3 in part 3. I think this graph is a great visualization for the question and it is very intuitive for the viewer.