

Final Grade Reflection

Sebastian Boksem

After looking over the last 10 weeks of this course discussion of my midterm portfolio, I would consider giving myself a grade of an A. I think this decision stems from my ability to successfully learn the targets that we were given throughout the quarter. I would say this class has took me on a journey to learn some of the most difficult, yet at the same time rewarding material than any other course I have taken. I think as a result, I have grown in aspects such as revising my thinking and being more collaborative.

I would like to think that I have accomplished most if not all of the learning targets throughout this quarter. I have accomplished WD-1 in Labs 3, 4, and 7 through the use of a CSV file and referencing it through my files in the first step of every lab (Importing Data section). I have also achieved importing xlsx files through PracticeActivityWeek4 with the readxl package. I have accomplished WD-2 through Practice Activity 3 in step 1 in which I narrowed the data set down to certain variables. I have accomplished WD-3 in the “Finding Bieber” section of Lab 3. Here, I filtered rows from both numeric and character data types. I accomplished WD-4 in Lab 3 in question 4 by changing multiple variables to a factor type. Furthermore in question 6, I created a new variable to determine whether the participant was white or non-white. I completed WD-5 through the use of a left join on question 6 on Lab 4 to connect the different types of regions/types of avocado combinations with the original volume from the cleaned avocado data. I have achieved WD-6 through question 2 (cleaning data) by incorporating both anti-joins to remove total_us rows and also semi-joins to only list California cities. I have completed WD-7 by pivoting in Lab9 “Pennsylvania Values” section with a pivot wider. In Lab4 question 7 I used a Pivot Longer. I have accomplished R-1 through the use of my Lab8.html in which the final lab report is looking professional looking with different headers, and also being reproducible with the indication of the read.cs being through a dropbox link. I have successfully done R-2 in PracticeActivity6, code chunk 1. For the result, I added # before every step to signal what I did for documentation. It is also tidy is respect to using proper returns and spacing. I have achieved R-3 in the “Familiar Words” section by slicing_max to avoid any changes in inputs. I have also created achieved this in Lab7Complete in the “Adding Stops” section by creating a function that will stop if the variables are not the correct type, or the length is greater than a certain amount. I achieved DVS-1 on question 7 of Lab5 question 1 and 2 in which I made not only a time series plot, which is based upon a numeric and a factor variable. I also in Lab7 completed a histogram in which a character variable

was used. I completed DVS-2 in Lab 5, Question 2.3 in which I reordered the legend so that the first one would have the highest mean, connecting it better to the line graph. I have accomplished DVS-3 in Lab7 in the “Visualizing Missing Data” section in which I make a time graph upon which entries were marked as NA. I completed DVS-4 to find numerical summaries in Lab 3 in the “Familiar Words” section in which I found a medians of different ratings. I have completed DVS-5 in Lab7 in which I summarized (a count) across every single variable the number of NA values in each. I have accomplished DVS-6 on Lab9 in the “Pennsylvania Values” section in which I created a table for the counts between the two states. I achieved DVS-7 in Lab9Challenge “Importing Data Set” section in which I added a filter at the bottom for the reader to select in addition to making the table red. I have achieved PE-1 in Lab 3 question 4 by using a across function to mutate variables into a factor data type. I completed PE-2 in Lab7 “Writing a Function” section by creating a resealing function to re scale the length in the fish data set. I have achieved PE-3 in Lab 8 “Full Song” section by creating an an iteration to sing the full 12 days of Christmas song. I checked off PE-4

After taking into account the first six weeks of my STATS 331, I would consider giving myself a grade of an A-. I think this decision stems from my ability to try and fully learn the targets attached,. As a result, I think my learning has followed in other aspects such as revising my thinking and being collaborative in my group, peer reviews, and Discord. Over these last 6 weeks, I feel like I have ended up being able to successfully checkoff most learning targets. For example, with working with data, I have successfully been able to filter rows from a data base for a variety of data types. I think a clear example of this lies in my Lab 3 submission, Hip-Hop. Under the Finding Bieber question, I was able to find a single participant out of 10752 rows by different uses of grouping and filtering both factor and numbers. Not only this, in the process of filtering these factor variables, I needed to turn them into factors originally, which was accomplished and done in a program efficiency way. Instead of creating a mutate with changing the variable to a factor for all the variables I needed to do, I instead followed to create conciser code by including an across function (and did that in the labs that followed). I feel like I also accomplished parts of “Working with Data” by the use of different joins, as well as filtering joins inside more specifically my Lab 4: Avocado Prices. In this lab, A key factor was the use of filtering joins to remove certain aspects of regions that were not needed. My ability to filter out certain regions has clearly been shown in the `clean_avocado` data set, which I was able to remove rows that were not connected to any metro area (and following create new data sets surrounding US total in addition to California regions) As a result of this, I was further able to analyze and make visualizations about their respective Avocados sales. I believe this also demonstrates my ability to be able to cycle between the use of filtering joins and a `filter()` statement. These use of working with data through the use of joins also connects to the learning targets associated with “Data Visualization and Summarization”. In Lab4 surrounding the mean average price for different regions within California, I calculated numerical summaries of variables, then following that, I further used a left join to connect back to the initial data after grouping to create different visualizations such as a box plot. I feel like this highlights my abilities to creates these summaries, while also highlighting my ability to differentiate between what kind of joins exists and use each one in their own unique way.

I have completed all the challenge assignments to date in order to highlight my creative ability in creating these data visualizations. I feel like a good example in which outside learning was used could stem from the Lab 4 challenge problem in which we were attempting to connect each California Region's housing prices to the average price of an avocado. I ended up creating a scatter plot showcasing the differences and found a slight association that as housing prices increases so did the price of avocados. This helped me overall extend my learning outside this class by being able to create a data set of my own from outside resources which previously I did not really know how to do.

Historically, I have completed every reflection attempt given the opportunity and have received check marks on all labs and challenge assignments, just waiting on a formatting reflection on Lab4. This class has overall been great in terms of teamwork for me. My team have done a really influential job attempting to try and complete the practice assignments and if one of us ever has any questions about labs we tend to answer and try to help one another after class hours. For peer reviews, I have reviewed and given helpful syntax suggestions to every person I have been assigned to, although I will say at certain times it can be hard due to us having the same coding knowledge as one another. I have started to be more active in Discord to answer questions. One example of this was during Lab4: Avocado Prices in which a person had a question concerning how to reference a non-syntactically correct column. I previously spent the last 30 minutes looking into how to solve issues like that, and seeing there question arose I told them that back ticks were needed for certain columns like that. In addition, I also just used the code that they had (even though my code for the problem was different) and adjusting it with the back ticks to actually see if it contained the correct answer for them. I think this was important as rather than just copying and pasting my own code, it allowed them to use that small knowledge gained from me to get closer to the answer to the problem on there own.

One thing I believe that would negatively affect my grade and would note for improvement over these next three weeks is my ability to answer the practice assignments on time. To my knowledge, I believe I have turned in two practice assignments later than when they were suppose to be due. To be honest, I think a small part of it has to be wanting to give myself a small break throughout the week whereas the the weekend and start of the week I would be working on the labs, challenge assignments, and the preview activities immensely. As a result, I found my self a little burnt out and waited until the weekend to complete these activities before the lab work. This is something I hope to improve on in the last part of the quarter and have started to designate times on Wednesday to complete the assignments. This class has been one of the most challenging, but at the same time rewarding classes I have taken. A goal I had from the start of the quarter was to understand R in a greater ability than I did in previous classes. I think as of now I have accomplished that immensely and hope to continue to do so until the end of the quarter!