Josh Rowe
10/30/22

STAT 331 Week 6 Reflection

I have demonstrated a strong understanding of how to import data from a variety of data types (learning goal WD-1) in the exercise "data_import_practice" lines 12 (read_csv), 21 (read_tsv), 31 (read_delim), and 41 (read_excel). Goal WD-2, selecting columns from a dataset, was demonstrated in my revised lab 4 quarto document (henceforth referred to as lab 4_Revised) line 81, which used the select() function. I demonstrated an understanding of how to filter different rows from a dataframe, goal WD-3, in lab 4_Revised line 64 which used the filter() command. I modified existing variables/created new variables (WD-4) in the same file in lines 76, 100, and 120; lines 76 and 100 updated the variable types of columns, and line 120 created a new column based on the values of other columns. In Week 4 Preview Activity, lines 66 and 69, I used right_join() and left join() (respectively), demonstrating goal WD-5. Lab 4_Revised demonstrated an understanding of filtering joins (WD-6) in lines 47-49, which used the anti_join() function. Learning goals R-1 and R-2 were both demonstrated in lines 27-45 of lab 5, which utilized the here() function to create reproducible code and used new lines for arguments in the ggplot() function to improve readability. Lab 4_Revised line 81 used variable names as arguments in the select() function as opposed to column indices, demonstrating an understanding of learning goal R-3. Lab 4_Revised lines 153 through 162 used ggplot to create a plot with both factors and numeric data types (learning goal DVS-1), and line 118 of the same document used the summarize() function to calculate a numerical summary of a grouped variable (goal DVS-4). Lines 38-45 of lab 5 used the plot modifications jitter (applied with the geom_jitter() function), alpha, and outlier.shape in order to create a clean-looking graphic, demonstrating a firm grasp of learning goal DVS-2. Lines 109 and 110 of lab 5 demonstrate

a mastery of the process used to determine numerical summaries of a variable across multiple groups (using the group_by() function followed by summarize() with a selected summary statistic – in the case of the lines previously referenced, the chosen statistic was the count of rows within each group) – learning goal DVS-5. Mastery of learning goal DVS-3 was shown in challenge 2 lines 27 though 37, which used the viridis package and outlier.shape to create an interesting and original graphic. Learning goals PE-1 and PE-4 were demonstrated in lab 4_Revised. Lines 114 through 117 used multiple inputs to the group_by(), mutate(), and select() functions to avoid unnecessary lines which would have repeated the same command (goal PE-1). Lines 91 through 104 demonstrate an understanding of dplyr's pipe operator (" |> "), the slice_max(), summarize(), and mutate() functions, and grammar of graphics' ggplot2; these are relatively recent tools (dplyr was released 8 years ago, and ggplot2 15 years ago) which are continuously updated.

I have submitted revisions for each assignment marked with anything less than an "S," which represents satisfactory understanding of the concepts at hand. In lab 1 I re-visited the prompt to ensure that I addressed it correctly, and in lab 3 I updated my variable names to improve clarity, discovered the case_when() function to prevent the use of nested if statements, and made use of slice_max() because it was more efficient than filter(). In challenge 3 I revisited the summarize() and across() functions in order to ensure that I used them correctly (in my original submission I had across() create new columns instead of updating existing ones). In lab 4 I once again made use of the slice_max() function, which I had again forgotten but have now made sure to remember.

Due to my demonstration of the aforementioned learning targets and my commitment to revising and improving my work to better understand R, I believe I have earned a course grade of A.