Revisions

Fiona Norton

Lab 2

Feedback from Dr. Theobold

Question 1:

"Careful! A professional looking report should not display messages about reading in the data or loading in package. These messages are not useful for the reader and make your document look cluttered."

Question 6:

"Technically, both geom_point() and geom_jitter() are plotting points, so you are actually double plotting the points. You only need one of these geoms in your plot!"

Question 11:

"You do not need to use the coord_flip() function to flip the coordinates of your plot. There is a much simpler way. Can you find it?"

My Revision Reflections

One of my takeaways from the feedback I received on this lab was the importance of suppressing unnecessary messages/outputs. When I first turned this in, I did not know that the output that prints in the console when I run my code will be printed in the quarto document automatically unless I specify otherwise. Now I know to use the "messages: false" or "echo: false" functions in the code chunks to prevent this. I also did not understand that using geom_point and geom_jitter at the same time is redundant because they are plotting the same data. I changed my plot in question 6 to only include geom_jitter so that we could see the points more clearly.

Finally, I learned that the coord_flip function is not necessary to use if we want to change the orientation of box plots because we can just switch the x and y in the code and get the same result. I think I previously used coord_flip because I had learned about it in the reading and assumed that it was the best option since I would just have to copy the previous code and add coord_flip to the end.

Lab 3

Feedback from Dr. Theobold

Question 8:

"Consider the nature of the data when plotting. Is every participant on one row? If not, what should you do with the data before plotting it?"

My Revision Reflections

Through the feedback I received from professor Theobold I was able to improve my analysis of the data. I understand the importance of describing the data with both summary statistics and words so that the readers understand the context of the analysis. I also realized that when plotting data where each subject is not in an individual row, it may be helpful to make a transformation before plotting the data, and that it is important to plot what is requested (i.e., demographics).

Lab 4

Feedback from Dr. Theobold

Question 5:

"I want your output to be the five cities with the most avocado sales. You found the top 5 regions and their names are stored in a dataframe. You should use this dataframe to filter rather than typing out the names a second time."

Question 6:

"I want you to pivot your dataframe to obtain the the differences! I also wanted a plot visualizing the differences in avocado prices for these cities."

My Revision Reflections

My revisions for this lab have helped me to fine tune my joining skills and make my code more efficient. I realized that some of my code chunks were printing unnecessary information in my rendered quarto document so I looked through the functions that can be put after a "#|" at the beginning of a code chunk and found the "include" option. This allows me to specify whether or not any of the output should be included! I also figured out some ways to make my code more efficient such as using the "n =" argument in the slice_max/slice_min functions that allows me to display the top n of bottom n rows in a dataframe. Understanding this allowed me to create a "top 5 regions: dataframe without having to type out the names of the regions, and I was ultimately able to delete one of my previous steps in creating the data visualization for these regions. Another way that I was able to make my code more efficient was including the diff() function to display the differences in mean between organic and conventional avocados across California regions, I was then able to create a plot of these differences. Finally I learned how to make adjustments like"n.dodge =" and "labs()" to make my data visualizations easier to understand.

Lab 7

Feedback from Dr. Theobold

Question 3:

"Great function! I'd press you to think about the redundancy in your function. Is the same function being called multiple times? If so, how can you remove this redundancy?"

Question 6:

"Although these are great plots, I was hoping for something very simple, so it is as easy as possible to compare the distribution of lengths with rescaled lengths."

My Revision Reflections

One of the main adjustments that I made to my code after receiving feedback was in my rescale_01() function in order to make it more efficient. Originally I repeated the max() and min() functions but by assigning the max and min to objects, I am able to use them later in the code without actually calling the function. Additionally I realized that sometimes more complex data visualizations are not necessarily better for getting information. To compare the original and rescaled lengths we can just look at their distributions and see that they are virtually the same.