Challenge 4 Reflection

Professor Comments:

I’d like to see a reference to where you obtained these data!

We use boxplots to plot the relationship between one numerical variable and one categorical variable. What type of plot is more suitable for two numerical variables?

Careful! Look at your plot title! Is that what you are plotting? It seems like you might want some jittering in your points to alleviate the overplotting!

I added a brief description to how I obtained the data. I looked for the average home price of average home size (2000-2500). At first, I didn’t think about how the avocados dataset was gathered from 2015-2018. I had originally compared average avocado price to 2022 home prices. I also reworked my ggplot step to be a scatterplot instead of a boxplot. I was also considering facet wrapping by year, but I thought it made it more difficult to see the trends in prices.

I also reworked the title as well as the jitter opacity to make the data eaiser to see.

My argument ended up staying the same. There was no direct linear relationship between avocado and home prices. Using phrasing from checking the linearity assumption in simple linear regression, I came to a conclusion similar to that of a Lack of Fit F-test.