R Dataframes Assignment – text for exercise 5 part 3 – Sam Duffield and Stacey Stevens

In order to ensure that each line of the code operated as expected after we had fixed the file, we ran them one at a time and inspected the results displayed or changes to the data frame. First, we changed the code for importing the original frame so that it could be easily adjusted to fit with various working directory locations as long as a data folder with the data in question was present. We then adjusted the first summarize function code to calculate the desired mean instead of the maximum value for the groups in the data set. We also added another pipe command to the second code to ensure that the data of interest would be grouped and summarized properly, even though there were other ways to fix this problem.

This ensured that the two codes had parallel structure and could be compared to each other. The only differences were the desired grouping variable, making it easier to compare the code. It would take a long time to calculate the average volumes for the data frame without code, we would have taken a subset of data with the identical structure and ensure that the code functioned properly. In order to compensate for any null values in the data frame, we also added mean(volume, na.rm=TRUE) to the code to remove any null values so that the averages would only represent data that had been collected successfully.

Using multiple methods with different code could also provide an increase in confidence that the numbers that were determined are correct. Using SQL that we learned this semester could help to verify the results found using R in class. While we were compiling the code, we were able to run each section separately so that we could observe the desired result at each step. Although this does not help verify the final result, it gave us confidence that each line is having the desired result on the data frame we were manipulating. By comparing these stages whenever there is an opportunity between different data manipulation methods, we would be able to watch where problems arise during our analysis. The high degree of precision in the calculated values also gave us confidence that we had found the correct answer, and they would help even more if we had done an analysis with multiple methods.