Data Challenge 3

Data Science I

For our third data challenge, we will be using the coffee ratings data from Tidy Tuesday. Please follow the link and read about the variables and familiarize yourself with the dataset. Remeber to use functions and iteration if you are performing the same operation 2 or more times!

1) Write a function that returns a table with the mean and standard deviation of the aroma through moisture columns of the coffee ratings data for different categories of coffee. Produce the table for (a) species, (b) country of origin, and (c) binned mean altitude in meters.

- Write a function that returns the mean and sd of a vector
- Write a function that uses the function defined in the first step and returns the mean and sd of the aroma through moisture columns over a specific grouping as a table (Hint: To make a table in Rmarkdown checkout the function knitr::kable)
- Wrangle the coffee ratings data frame
- Using `map` to calculate tables for the three groups (species, country of origin, binned mean altitude in meters) in one step

2) Write a function that makes a scatterplot of any two variables from the columns aroma:moisture and colors by a particular grouping. Make the plot for (1) aroma versus flavor colored by species and (2) balance versus body colored by binned mean altitude. (Hint: Remove outliers to make a more compelling plot!)

- Write a function that plots two variables from a data frame and colors by group
- Use `pmap` to iterate over a list of function arguments to produce two plots:
  - Aroma versus flavor colored by species
  - Balance versus body colored by binned mean altitude

Rubric:

Code Style (10 points) Is code organized well and commented?

Submission (10 points) Was the data challenge submitted as an html document on Canvas? Does the html document look aesthetically pleasing? Did the homework contain a link for a GitHub repository? Did the repository contain the code for the assignment?

Question 1 (40 points) Does the code produce the three tables? Are the tables well organized and aesthetically pleasing? (20 points) How concise is the code? Does the code use iteration and functions when appropriate? (20 points)

Question 2 (40 points) Does the code produce the two scatterplots? Are the scatterplots well made (all scatterplots should have a title, good axis and legend titles without underscores, ...)? (20 points) How concise is the code? Does the code use iteration and functions when appropriate? (20 points)