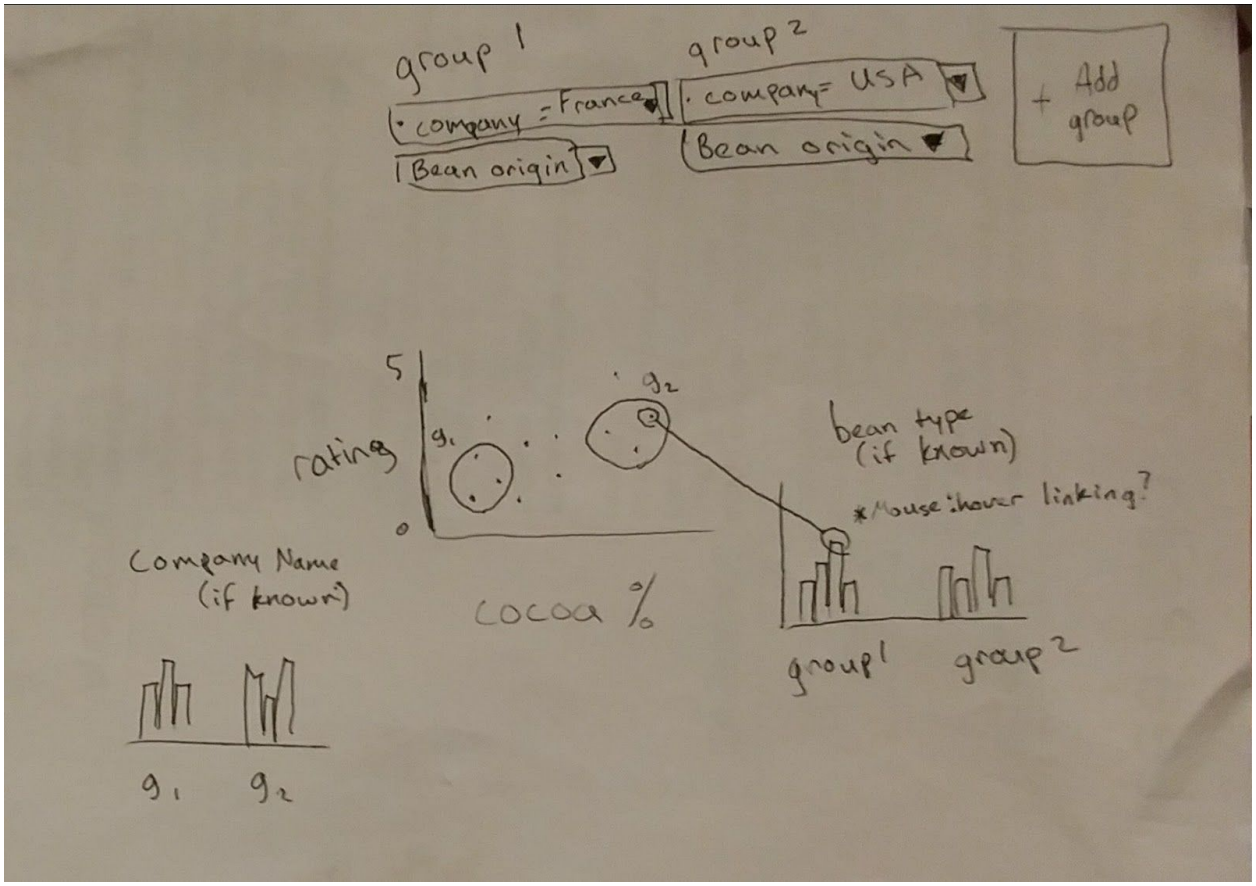
Lee Polla
2-17-18
Assignment 3: D3



The dataset that I found was a chocolate data set from kaggle. The data was about chocolate ratings of of different dark chocolates with varying cocoa percentages. Information included company names and locations, bean origins and bean types, and the cocoa percentage and rating. The bean types was often unknown in the data set, so I worked with company location and bean origin to create groupings in visualizations of the cocoa percentage to rating. Company location and bean origin country were the broadest categories and could be broken down into specific companies or to a specific bean origin within a country.

If a dark chocolate connoiseur or dark chocolate producer wanted to know where to find the beans or to learn which companies make the best chocolates then they can make groups by location or bean origin to see their ratings and cocoa percentages. Then users can get a more detailed breakdown of which specific companies or which specific bean origins are being shown on the scatter plot by either hovering over a point for a tooltip or viewing a side plot with information about where that chocolate came from compared to the counts for other places in the same group. If users hover over a point then it will be emphasized on the other plot. The user will also be able to compare between different groups on the same plot to compare the performance of different bean origin and company combinations.

After working with the data, I learned that there were about a thousand specific chocolate origins and over 400 specific companies who manufacture chocolate. This made grouped bar charts infeasible for my sideplot visualizations because there would possibly be too many bars to fit horizontally and possibly too small or too large frequency counts to be displayed on a single y axis. So I opted for a pie chart to display this information because it is always the same size and conveys relative information while also providing the exact count numbers through a tooltip. Since it was possible that too many different groups could show up in the pie chart and become very crowded, I opted not to show any text labels for these groups but did allow for the pie chart tooltip to linger after moving the cursor. To save space and allow for the user to focus on only two linked plots at once, I decided to make only one pie chart at a time to visualize a single group that can be chosen with additional dropdown selectors.

For the groupings to be shown together on the scatter plot, I decided to settle for only two groups because it could become too noisy or hard to keep track of variables later but some work was done with the intent of being reusable for any number of groupings. There were a lot of combinations that had no chocolates in them but all of the controls for making any groups are still available even if they show nothing. Different groups will appear in either blue or orange on the scatterplot, I recommend looking at combinations of Venezuala, USA and France with the group selectors to see the largest subsets of data rather than the many subsets with only one chocolate.

I spent roughly 15 hours on this assignment and the part that took the longest was figuring out how to use pie charts.