P5 Write Up

Team Members: Joseph Meredith, Kimberly Do, Sabrina Chua, Yujia Gao
Github Link: https://github.com/schua9/4460_Project
Vis Website Link: https://schua9.github.io/4460_Project/
Video Link: https://www.youtube.com/watch?v=gb5P9QkH2No

Data

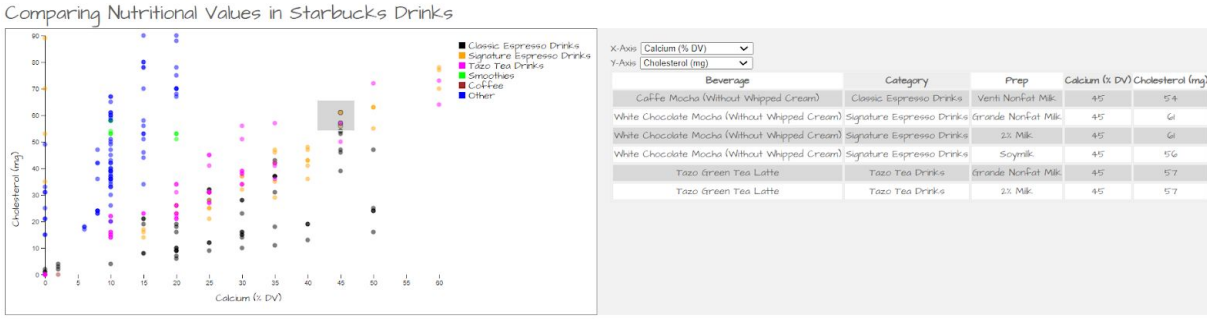
The data that we chose to work with was the compilation of Starbuck drinks and their nutritional values. We decided to work with this data as Starbucks is a popular franchise and health is an increasing concern. This data consisted of Strings for beverages, its categories, and preparation and numerical values for the nutritional facts. Although there were many of the same beverage, there were different forms of preparation that resulted in different nutritional values. With our visualizations, we hoped to gain and give our viewers a better understanding of Starbucks beverages and their nutrition information.

Beverage_	Beverage	Beverage_	Calories	Total Fat (	Trans Fat (	Saturated	Sodium (r	Total Carl	Cholesterc	Dietary Fi	Sugars (g)	Protein (g	Vitamin A	Vitamin C	Calcium ('	Iron (% D\	Caffeine (mg)
Coffee	Brewed Cc	Short	3	0.1	0	0	0	5	0	0	0	0.3	0%	0%	0%	0%	175
Coffee	Brewed Cc	Tall	4	0.1	0	0	0	10	0	0	0	0.5	0%	0%	0%	0%	260
Coffee	Brewed Cc	Grande	5	0.1	0	0	0	10	0	0	0	1	0%	0%	0%	0%	330
Coffee	Brewed Cc	Venti	5	0.1	0	0	0	10	0	0	0	1	0%	0%	2%	0%	410
Classic Esp	Caffe Latte	Short Non	70	0.1	0.1	0	5	75	10	0	9	6	10%	0%	20%	0%	75
Classic Esp	Caffe Latte	2% Milk	100	3.5	2	0.1	15	85	10	0	9	6	10%	0%	20%	0%	75

Small Table of Starbucks Drink Data

Since the data was about Starbucks, we wanted to have a casual and light feel, so we chose lighter colors for our color palette. Our concept was friendly-looking and kept it similar to Starbucks as the audience of our vis would enjoy that style.

Scatter Plot



Screenshot of Scatter Plot

For the scatter plot, we wanted to allow the users to compare the various nutritional facts for the beverages. When making a selection of drinks at a restaurant, users would likely consider a variety of drink options, so we wanted to offer a data visualization that would help a user choose a particular drink (from the many available options) by comparing nutritional content. Rather than small multiples, we decided that allowing the users to select axes of the nutritional values would give a closer look at the specific values they would like to compare. This could give different insights such as if a certain nutritional fact is directly related to another or which types of drinks had a higher nutritional value overall. As well, we concluded that most users may not be interested in seeing all of the nutritional facts and would only look at a select few. Choosing only two selections would also feel less overwhelming. From this, users can see possible correlations in the overall macro and the specific beverages with their corresponding values in the micro.

Kimberly

To let users compare drinks across two selected nutritional values, we plotted all the given Starbucks drinks on a 2D scatter plot. Although we considered using a barplot, we felt

that users would be more interested in contextualizing the nutrition of their beverage across different variables and added sorting and grouping to organize the many beverage types.

First, we implemented two dropdown boxes where users can select what nutritional values they want to appear on the x and y axes. After selecting a variable from the dropdown box, the plots automatically update, the axes resize, and the axes labels change. Initially, the axes did not resize. However, we noticed that because some of the nutritional values (e.g. calories) exceeded 100 while others (e.g. saturated fat) were barely 15, it could be hard to visually compare smaller values while maintaining the same large scale. Thus, we chose to implement automatically resizing axes (that still start at 0 and scale linearly). Together, these two dropdown boxes offer users the freedom to explore the nutritional values across a number of Starbucks beverages for two desired variables.

Second, we color-coded the beverage categories so users could grasp a general idea about the nutritional value across a beverage category. Initially, it was difficult to understand what each point represented when the points were the same color. The categorized coloring system, we saw an improvement of understanding the general nutritional value across beverage categories. As there were many categories, we chose the most significant ones.

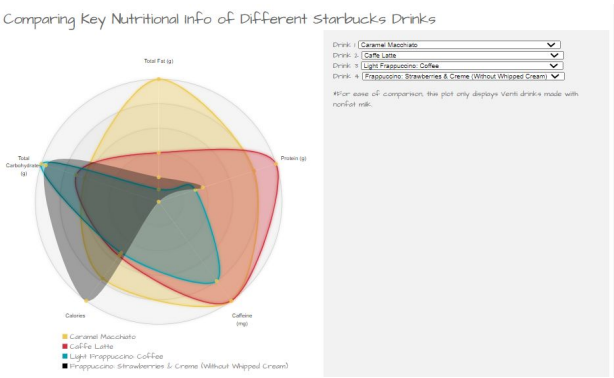
Sabrina

We then added a legend for users to easily see the different beverage categories. At first, we placed the legend inside the graph to save white space, but the points sometimes would overlap the legend, making it difficult to see the dots and the legend. Therefore, we increased the width of the graph so that the legend would be placed to the side.

Furthermore, to gain further insight into the data, we wanted to allow the user to brush over the scatter plot and display a table of the highlighted data. This would display the specific values the viewer would be interested in. Because the scatter plot already had a colorful color palette, we chose to only add a dark outline around the brushed dots and kept the table's overall color scheme monochromatic. For the table, we initially used 1 color but it became harder to read as more data values were highlighted. As a result, we chose a light gray that was dark enough to differentiate the rows and light enough to read the text clearly.

For the table's columns, it was necessary to display the beverage and preparation for context. We debated to keep the category column with the legend, but the blue points labelled "Other" needed more context into their respective categories. For additional columns, adding all the nutritional facts would cause the table to be too long and bulky. To accommodate for that, we altered the table to update 2 additional columns based on the chosen X-axis and Y-axis. As the user had chosen those specific axes, we expect that they would want to see those specific values. Once the table was updated, we realized that there were duplicate columns if the axes were set to the same nutritional fact. We adjusted the table again so if the chosen dropdowns were the same, then it would only display one additional column.

Radar Plot



Screenshot of Radar Plot

In addition to the scatter plot, we created a radar plot for direct comparison of specific Starbucks beverages. The goal of this visualization was to allow users to compare the nutritional values between different drinks and in turn, select the most suitable drink based on personal fitness and nutritional preferences. We selected the radar plot for this task because firstly, it is suitable for comparing a large number of variables across different entities. Secondly, the audience can quickly compare, as the data is represented as shapes and they overlay each other. Lastly, it is able to compile variables that have different scales of measurements and show them in one plot. This is especially important for us because the starbucks drinks dataset contained variables with different units.

Yujia

Considering too many variables on the radar plot might reduce the readability of the visualization and prevent the audience from locating the relevant information, we created the radar plot with 5 selected categories that are people may care most about: total fat (g), total protein (g), total carbohydrates (g), calories (kCal), and caffeine (mg).

To make it easier for the audience to directly compare between different drinks, we put the drinks to be compared in one radar plot. We adjusted the transparency styling of the area representing each drink as well as choosing distinctive colors for the ordinal data to make the visual encoding clearer to the audience. We also limited the maximum number of drinks to be displayed on the radar plot to be 4 to ensure that all the data entries can be read easily without the interference of the messy overlay.

The variables have different units and different scales, so it would be unreadable if we kept a uniform scale across the axes. Thus, we scaled each axis respectively based on the maximum value of each group. We made this decision because the purpose of this vis is for the audience to be able to compare between different drinks, instead of knowing the absolute value of a variable. This way it will be direct for the audience to compare the relevant values for each variable, where the outermost circle value represents 100%, the second outermost circle represents 80%, etc. If the user also wants exact values per category, we added a hover over effect, where the audience can just hover over the data point and read the exact value.

Joseph

While the users cannot choose the categories to be displayed on the radar plot, they can choose the drinks to be displayed on the chart via the dropdown menus to the right of the chart. We felt that having too many dropdown menus would be overwhelming, both visually and functionally, and we thought that it was more important for comparison to choose the drinks over the variables displayed.

The dataset we were working from contained over 200 individual drinks, and we knew that all of these couldn't go in a dropdown menu. Many of the drinks in the dataset were variations of different sizes or prepared with different types of milk. We chose to display only the drinks of the "venti" size, which is Starbucks' largest size, and the drinks prepared with nonfat milk. We chose these because they seemed to encompass the widest variety of different drinks, and by choosing only these variations the dropdown menus contain just under 30 entries. We made special note of this under the dropdown menus so that users would not be deceived.

In addition to this, we also cleaned the data for the radar plot by omitting entries with all zeros for the nutrition information, and by counting "varies" as 0 for the purposes of the chart. This allows our data to display properly across all entries and allows for a smooth experience by the users when comparing different drinks.