

Annabel Consilvio and Hannah Fowler
Paul Ruvolo
Software Design- Section 2
13 March 2015

Project Overview

For this project, we analyzed data taken from a survey given to participants who attended the World Cup in 2014. The survey required background information including age, level of education, gender, and residential location. We specifically filtered preferences for various ethnic foods from 48 different countries using the given background information for each participant. With this, we determined which country created the most favorable food altogether and for various demographics.

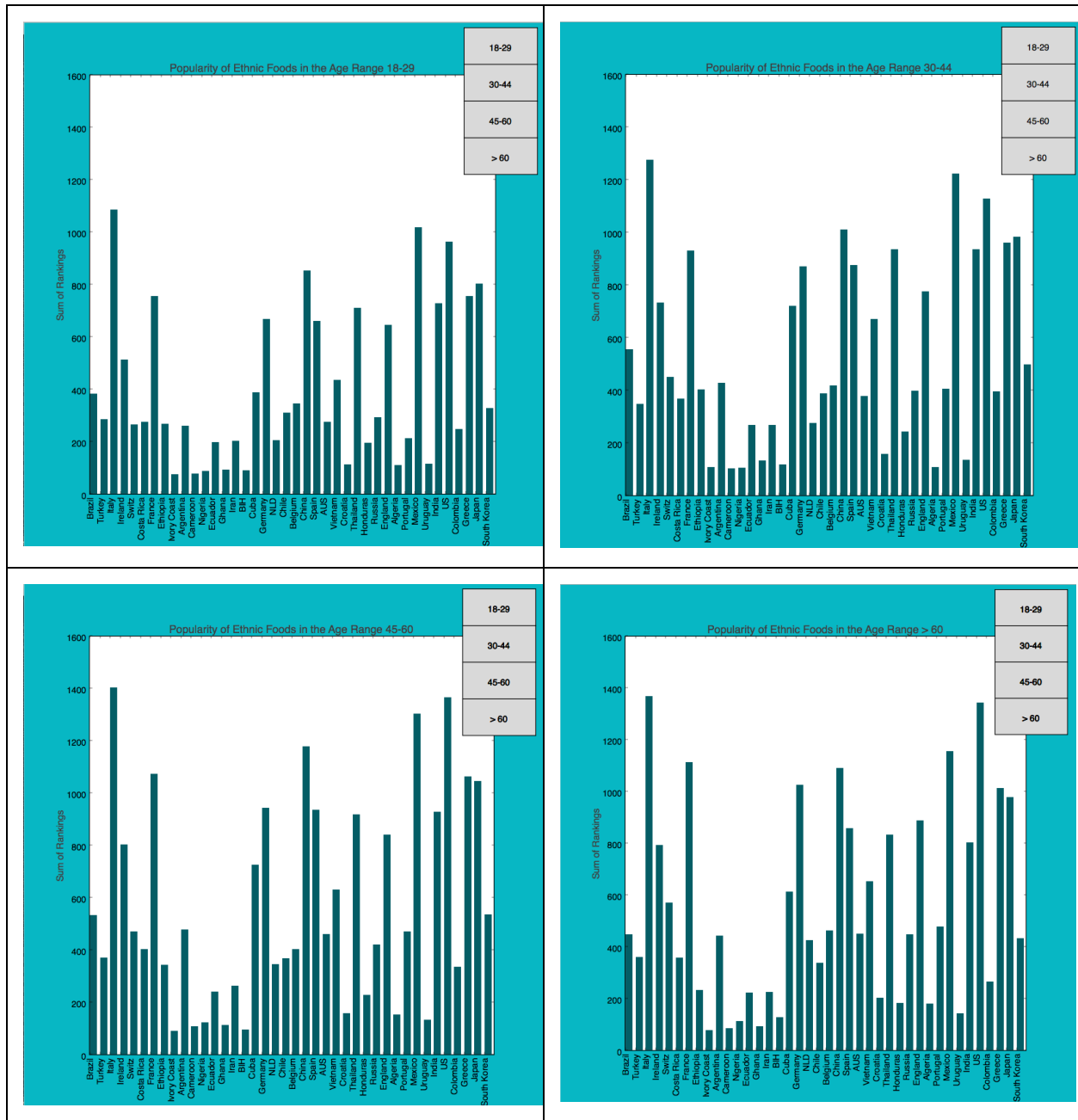
Results

From our analysis, we found that Italian food was the most loved among all ages. Additionally, we found that the younger generation (under 44) preferred Mexican food second to Italian food while the older generation (over 44) chose US food as its second favorite. In general, almost all ages did not rank the food from the Ivory Coast, Cameroon, Nigeria, Ghana, Croatia, Algeria, and Uruguay very high. However, as these are not very commonly eaten foods in America, it is likely that the surveyed population may never have eaten these foods, which is why they were ranked so much lower. Some of the other bigger winners were China, Germany, France, and Japan.

These results are plotted below in the form of bar graphs. Each of the four shown below represent a different age demographic polled (~1300 people were polled in total). We were able to make the data interactive by clicking the buttons in the top right hand corner which each correspond to a different age range. When they are clicked, the bar graphs update according to this age range. In the future, we would like to implement the gender and location data also accumulated in the data set. We would also like to make an interactive world map to indicate where in the world food is the most loved.

There were some slight problems with the data set we used for this project. The most important problem we had with it was that we had to assume that N/A data points could be called 0. When the data was taken, if a person was not asked a specific question about a country, the data would return N/A, but the data would also return N/A if the person said they didn't know about the food. Because of this ambiguity, we had to assume that any N/A data point was equal to 0. If the data had been taken a different way initially, our results might more accurately represent people's preferences in ethnic foods.

If you are interested in seeing larger images of the ones shown below, please see our github repository.



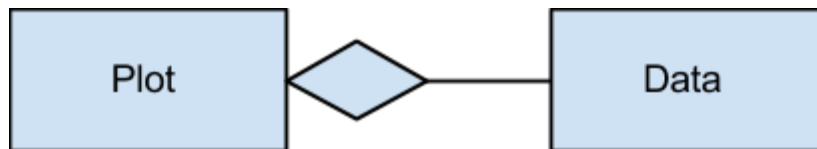
Implementation

We split our code into two main classes: the Data class and Plot class. The former filters the data and the latter visualizes it. When we initialized our Plot class, we referenced the Data class so Plot.

In the Data class, we filtered the data from a csv file such that any N/A elements were removed and replaced with a 0. From there, we summed each column by adding an element ranking if the ranking was given by a participant in our specified age range. Finally, the class combines the list of countries and list of sums into a dictionary so we could easily determine the total ranking for each country.

In the Plot class, we took our data filtered through the Data class and created a function that would generate a bar plot according to an age range. Furthermore, we created different functions that correspond to a button for each age range so that when the function is called, i.e., when the button is clicked, the bar plot is created.

Plot is composed without belonging to Data. The UML diagram of this is shown below.



Reflection

Throughout this project, understanding how to create bar plots with matplotlib went smoothly, since the documentation for it was fairly straightforward. However, actually making the data interactive by incorporating buttons was much more difficult. We considered various libraries for visualizing our data, but Bokeh did not have enough explanation to use it. Even though matplotlib helped create plots, there were very few examples to help show how to implement buttons into our final product. From this process, we have learned a new library which can be more useful over plotting figures in Matlab which isn't as customizable. Beforehand, it would have been nice to know Bokeh's library so that we could have created a nicer looking graph, but consequently, we had to understand Matplotlib's appearance methods. Overall, we can now say that we have gained a new skill of reading documentation and understanding it so we may use an unknown library.

Because we understand each other's work ethic and work well together, we spent the majority of the project pair programming and working at least in the same room. The rest of the time, because conflicts came up, one person would work on the project, but we would then come together, catch up, and continue on the same path. Throughout our given time, we didn't have any issues working together, and simply enjoyed chai tea while coding.