INFSCI 2415 - Information Visualization

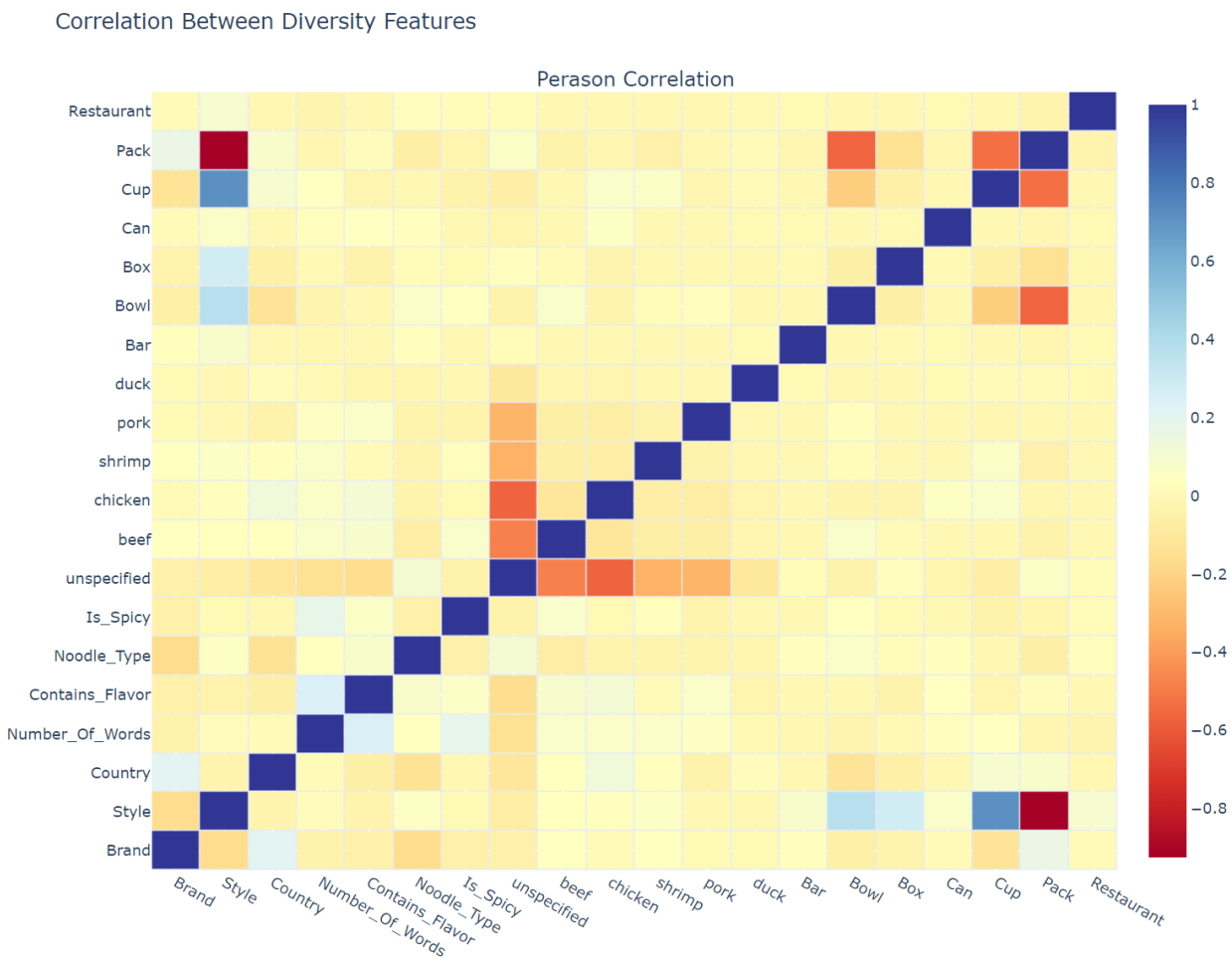
Final Report

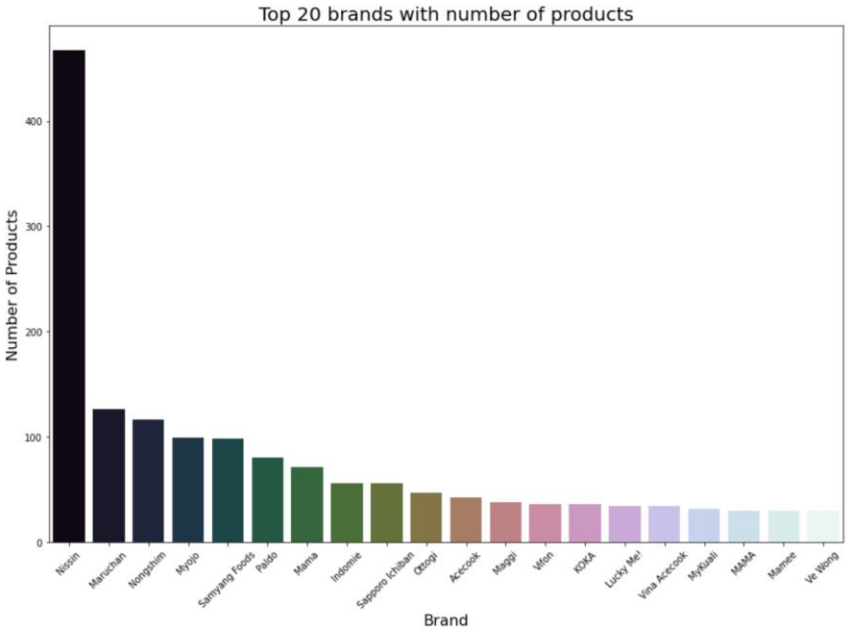
Qisheng Ye

qiy48@pitt.edu

Title: Analysis of Ramen

Figure:





Legend:

In the first above figure, it shows the correlation of different features. The more relevant the features are, the closer the color of the intersecting square is to blue, while the it is not, the closer the color is to orange. The horizontal and vertical coordinates have different features, and the color represents correlation.

In the second above figure, it shows the top 20 brands using the Ramen Ratings (Latest Update Jan 25 2020) data set to count different ramen brands. The top 20 brands are listed horizontally on the x-axis, while different ramen brands are counted on the y-axis and marked with different colors. In the next two word cloud figures, the size and color of the word indicate the frequency of the word.

Findings:

For the first figure, it utilizes pearson correlation: pearson correlation evaluates the linear relationship between two continuous variables. The correlation of different features is shown. If the two variables are increasing or decreasing in parallel then they have a positive correlation between them and if one of the variables is increasing and another one is decreasing then they have a negative correlation with each other. If the change of one variable has no effect on

another variable then they have a zero correlation between them. The correlation is displayed through different colors.

As shown from the second above figure, Nissin company has produced the most products, with more than 400 different products in the data set, while next four companies are quite similar in the number of products. In the following 15 companies, number of products declined at a balanced slope, which means that the Nissan dominates the market. This is in line with the reality because ramen is extremely popular in Japan. From the last two word cloud figures, we can easily find that Nissin is the largest ramen brand and Japan is also the main market of ramen.

Data and Method

The data comes from Ramen Rate in Kaggle. The Ramen Rater is a product review website for the hardcore ramen enthusiast (or "ramenphile"), with over 2500 reviews to date. This dataset is an export of "The Big List" (of reviews), converted to a CSV format.

First of all, because of the large data set, I have preprocessed the brand and country of the data. For example, there are many words that match spicy taste. When there are related words in the description, I will mark the data as ‘Is_Spicy’. This method is used to classify the utensils, meat and noodles. If unknown descriptions are encountered, they will be classified into ‘unspecified’. Through this method, I have processed the data set described by complex ramen noodles. Then I utilized the pearson correlation graph to reflect the correlation between different attributes. The positive and negative correlation differences in the figure are more obvious by selecting a color range with distinct colors.

In addition, by counting the number of products under different brands and countries, I can acquire the top 20 trademarks and countries. The size and proportion of the number of brand products are displayed through the histogram. Through the word cloud function, I successfully used the counts under different attributes to show the diversity and distinguish the frequency.

Statement

Ramen is not only popular in Japan and the Orient, but also a popular "fast food" in many western countries. There is no doubt about this, but which brand is more popular among many brands and which country prefers ramen can only be answered after data analysis. In addition, many questions need to be investigated, including the correlation of ramen attributes, which include not only brand and country, but also ramen varieties, tastes, meat and utensils. If we can find the rule of popular ramen noodles from these data, we will find out the development direction of ramen noodles industry. Since I only try to find out the rule of ramen properties through previous data, there is no risk of data leakage. We can see that in our data, one

company has produced the most products. "Nissin" dominates our product sample. There are more than 400 different products in the data set. We can find that if we look at the next two leading companies, they are quite close to each other in the number of products sold. When we look down, we notice that the number of products is declining at a fairly balanced slope, which means that there is no such huge difference between other companies except leading brands. In the word cloud, we can find the relationship between brand and country. Japan is indeed the most popular place for ramen. From the correlation graph, we learn about the relationship between different attributes, such as the positive or negative correlation between spicy ramen and what meat. In addition, we can also analyze the types of ramen noodles and the relationship between the flavor of ramen noodles and other attributes. Using these relationships, we will better find the development direction of ramen noodles.