Data Visualization

Qiyu Wang

10/12/2020

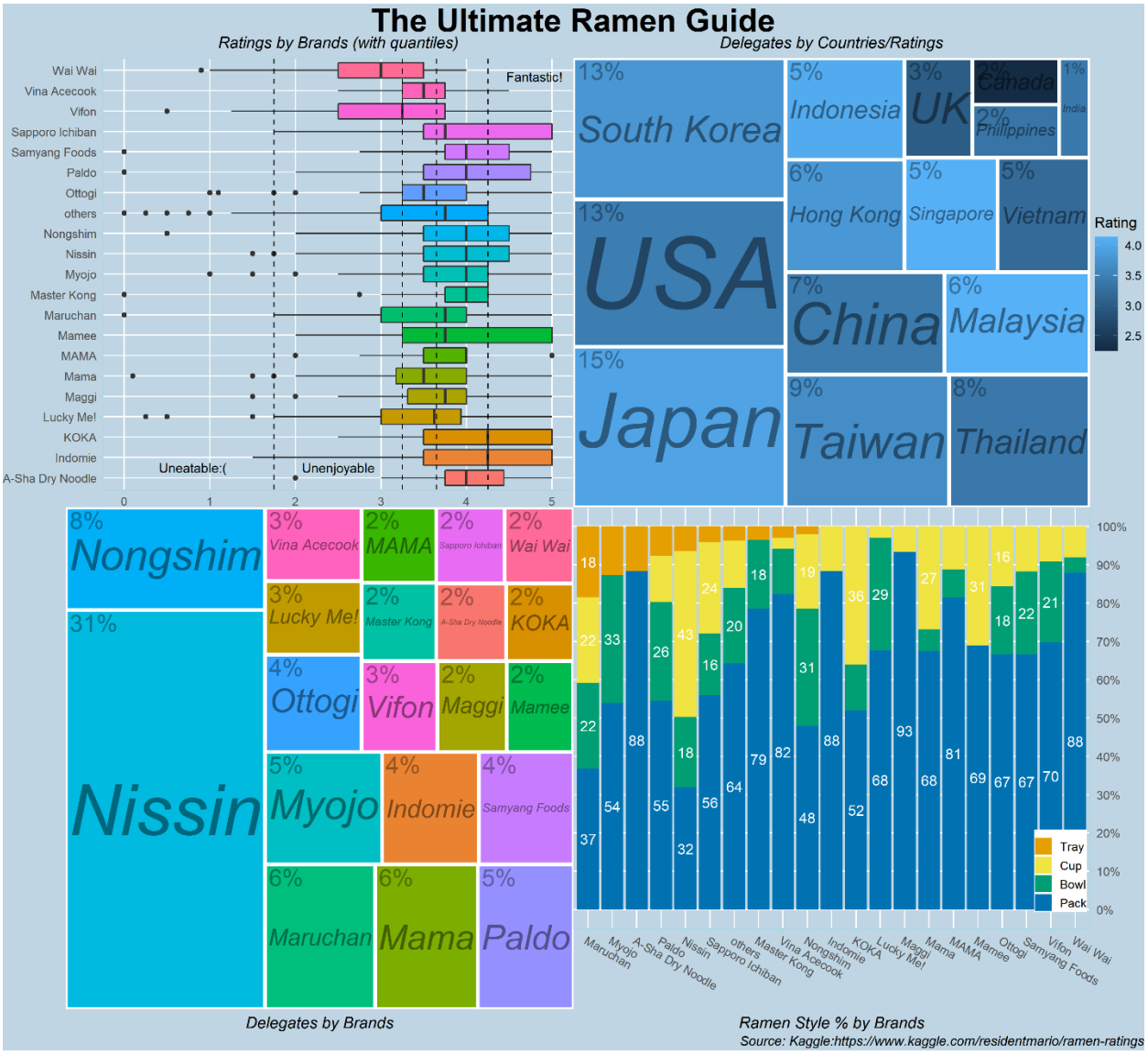
The Ultimate Ramen Guide

According to a research article by Lígia posted on ACM digital library and multiple online surveys/researches, Information system, computer science and engineering are some of those majors that easily hold students overly stressed or even depressed, especially for those students who have 4 homework, 1 project, and two exams overlapping within a week. For those overly stressed students, a bowl of delicious Ramen with eggs and Chashu could be their momentary escape, and this visualization should be the ultimate guide for those to pick their perfect fit of ramen.

For the project, I used a Kaggle dataset called "Ramen Ratings", which records 2580 different Ramen and their corresponding Ratings(stars), Country of origin, Brand, and package style. The purpose of this visualization is to help our students find their best available suit of ramen instantly based on Brand, Country of origin, and package style without any struggling.

The three key considerations here are Availability, Deliciousness, and Preference of Packaging, which is the aims to be addressed in this plot. I decided boxplot should be great to show the distribution of ratings, and treemap/stacked barplot should be nice to show the percentage information. To visualize availability, I used two Treemaps, to show the percentage of different ramens available out of total (2580) grouped by Country/Brand. Based on those two Treemaps, it should become a lot easier to identify the most approachable options by brand/country. The color scale in the country treemap is scaled upon average ratings, so we could quickly see which country's ramens are overall tastier. For the fills in brand treemap, the colors are assigned based on brands, which match the color fills in boxplot, so the brand treemap could be used as legend for boxplot. For example, I would definitely say Japanese ramens is a perfect balance of availability and deliciousness, while Malaysian ramens taste better but not as approachable as Japanese ramens.

Since brand is a more common metrics when it come to picking ramen, I introduced a boxplot to further explore which brand has better Deliciousness in detail. The boxplot shows the five-point summary of the measurement of Deliciousness by ratings distribution. The four dashed lines added are (from right to left) overall 75% quantile, average, 25% quantile and (25% quantile – 1.5*IQR) which shows the outliers (uneatable ramens). The fill colors are corresponding to the ones in brand treemap, so we can easily switch over there to see the Availability information. Finally, I introduced a stacked barplot to show the availability of packaging styles by brands. So, if anyone has a specific preference of packaging style, he could look into the brand accordingly. For example, if someone love Tray ramen, it would be definitely easier for him/her to find it with the brand "Maruchan".



Reference:

1. Data Source: Kaggle - Ramen-Ratings: https://www.kaggle.com/residentmario/ramen-ratings
2. Treemap Reference: https://cran.r-project.org/web/packages/treemapify/vignettes/introduction-to-treemapify.html
3. Theme Reference: https://www.datanovia.com/en/blog/ggplot-themes-gallery/
4. The Prevalence of Anxiety and Depression Symptoms among Computer Science Students – by: Lígia Maria Soares Passos: https://dl.acm.org/doi/abs/10.1145/3328778.3366836