STAT 436 hw4

Yichen Lin

May 2023

:Link to Rmd
Link to dataset

1 Discussion

1.1 What is the essential question that your visualization is supposed to inform?

I chose the McDonald's menu dataset on kaggle. The main goal of this shiny app is to analyze the inner relationship between the nutritional composition inherent in different categories of McDonald's products through pca downscaling, and observe whether there are nutritional similarities and differences between categories.

1.2 How do aspects of your design support exploration of the essential question? Were there trade-offs you had to make so that certain features were more clearly visible?

Users can select the main components they want to explore as well as the type of product. By looking at the nutritional elements in each principal ingredient and the scatter plot distribution of different categories of products on the principal ingredients, users can understand which categories of products have similar nutritional values and which categories have completely different nutritional values.

While designing the app, the trade-offs I encountered were mainly about how users choose which two principal components are the horizontal and vertical coordinates of the scatter plot and how to choose the size of the two plots. After trying different layouts, I chose to use the first two principal components entered by the user as the axes, while setting the facet statement to a maximum of two columns per row to prevent the problem of too narrow images by default.

1.3 What are your key findings? How do they relate to your prior understanding?

My finding is that the proportion of nutrition of some McDonald's products, especially coffee&tea, desserts, shakes and snacks, are not consistent with what people think in daily lives.

The first principal component has negative values for most nutrients, and the second principal component has high values for sugar content, fat content and carbohydrate ratio, and low values for the other components. According to the scatter diagram, snacks and desserts had essentially greater values on the first principal component than coffee and shakes on the first principal component, and less values on the second principal component than coffee and shakes on the second principal component. So generally speaking, drinks and shakes are much more likely to make people put on weight.

1.4 How did you create the visualizations? Were there any data preparation steps?

Before the shiny statement, there are several preprocessing steps. I removed the columns that are not needed in the shiny app and performed a pca downscaling on the dataset to get the result. This way the app can use the pca results combined with the original dataset to hit the visualization purpose.