

Data Exploration

In this assignment, you will design a visualization for a small data set and provide rigorous rationale for your design choices. You are free to use any graphics or charting tool.

A dataset (provided in csv format) includes attributes for each toffee along with its ranking. For binary variables, 1 means yes, 0 means no. The data contains the following fields. We have aggregated and wrangled the data to produce a data table with the following information.

Number of records: 12

Data fields:

Toffee: Name of toffee.

chocolate: Does it contain chocolate?

fruity: Is it fruit flavored?

caramel: Is there caramel in ity?

peanutyalmondy: Does it contain peanuts, peanut butter or almonds?

nougat: Does it contain nougat?

crispedricewafer: Does it contain crisped rice, wafers or a cookie component?

hard: Is it hard?

bar: Is it a bar?

pluribus: Is it one of many candies in a bag or box?

sugarpercent: The percentile of sugar it fall under within a larger data set of 86 toffees.

pricepercent: The unit price percentile compared with a larger data set of 86 toffees.

classwinpercent: The win percentage based on all the pairwise ranking matchups in our class.

Your task is to download this data and design a static visualization that you believe effectively communicates following aspects of the data:

- Using a pair plot, illustrate the relationships of sugar percentage, price percentage, and win percentage. Use chocolate variable to differentiate between plots based on its presence.
- comparison between sugar percentage and price percentage.
- Which type of toffee has higher sugar content?
- What are the top 10 toffees. who won the toffee power ranking?
- correlations between the characteristics and the popularity (probability of winning)
- correlations of toffees with the target variable “winpercent”.
- Plot the chart to show which is the most expensive and which is the least expensive.

While you must use the data set given, you are free to filter, transform and augment the data as you see fit. Such transformations may include (but are not limited to) computing percentages or averages, grouping elements into new categories, and/or removing data that are not relevant to your driving question. You are also free to incorporate external data as you see fit. Your chart should be interpretable.