

Candy Data Cleaning

11/8/2017

Background

This data comes from the *538 Ultimate Halloween Candy Power Rankings* (<http://fivethirtyeight.com/features/the-ultimate-halloween-candy-power-ranking/>); however, we need to clean this data before we can use it.

Variable definitions:

The variables (V1 - V13) included in the data (in order) are as follows (Note: 1 indicates “Yes” and 0 indicates “No”):

1. What’s the name of the candy?
2. Does it contain chocolate?
3. Is it fruit flavored?
4. Is there caramel in the candy?
5. Does it contain peanuts, peanut butter or almonds?
6. Does it contain nougat?
7. Does it contain crisped rice, wafers, or a cookie component?
8. Is it a hard candy?
9. Is it a candy bar?
10. Is it one of many candies in a bag or box?
11. The percentile of sugar it falls under within the data set.
12. The unit price percentile compared to the rest of the set.
13. The overall win percentage according to 269,000 matchups.

Cleaning task

Upload the `candy.csv` file into your **RStudio** session and *Import* the data just like you would for your class’ campaign data. Then:

1. Give each variable a short, descriptive name.
2. Change the names of the candy to a **character** class variable.
3. Change at least one of the variables with 1/0 to “Yes”/“No”.
4. For the variable(s) you chose in step 3, how many of the candies in the data were “Yes”?
5. Create a plot, using the variable(s) you chose in step 3, and one of the last three variables (Variables 11 - 13).
6. Interpret your plot.