

ANA515 Week 5 Activity

Jiaxuan Ren

2023-06-24

```
# Global options that apply to every chunk.  
knitr::opts_chunk$set(echo = TRUE, message=FALSE, tidy=TRUE, tidy.opts=list(width.cutoff=60))
```

Description of the data

The candy-power-ranking dataset measures “How often did a fun-sized candy of a given type win its matchups against the rest of the field?” and I hope to use this dataset to answer research question “What Halloween candy people most prefer?”

Read Data

```
# Load tidyverse library  
library(tidyverse)  
  
# Use read_csv function to read the data 'candy-data.csv'  
mydata <- read_csv("candy-data.csv")
```

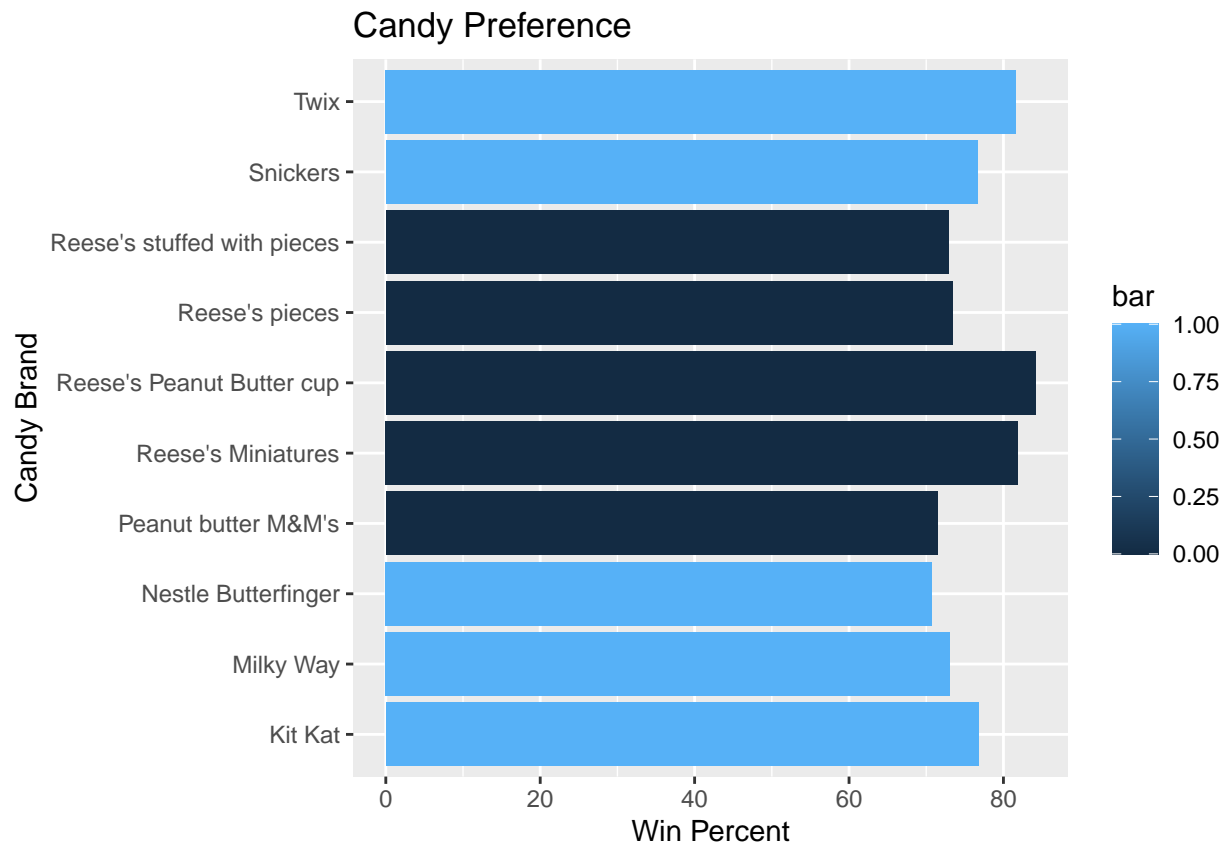
Clean the data

```
# Filter the data to only include top-10 'winpercent' candy  
# and assigned the filtered data into a new object called  
# 'filtered_data'.  
filtered_data <- mydata %>%  
  top_n(10, winpercent)
```

Visualization

1. The figure below shows the top-10 candy brands that win the most matchups “How often did a candy win its matchups against the rest?”. The dark blue denotes whether the candy is a bar or not (shallow blue).

```
ggplot(filtered_data, aes(x = winpercent, y = competitorname,  
  fill = bar)) + geom_col() + labs(x = "Win Percent", y = "Candy Brand",  
  title = "Candy Preference")
```



2. The figure below shows the top-10 candy brands that win the most matchups “How often did a candy win its matchups against the rest?”. The y-axis indicates how many percentages of sugar is contained in the candy

```
ggplot(filtered_data, aes(x = winpercent, y = sugarpercent)) +
  geom_line() + geom_point() + labs(x = "Win Percent", y = "Sugar Percent",
  title = "Candy Preference based on Sugar Percent")
```

