Lesson 13 Data Visualization for Stats!

Goals for Today

- More plots: plotting lines, making box plots, and puttings things in the order that you like.

Readings for Today

None!

Required

None!

Optional

None!

Important Links and Files

- The file we worked on in class is available here

Written Assignment #3

Details

- **Due Date**:

    - 2021-10-29 (**this Friday!**) prior to the start of class (i.e., needs to be submitted before 1PM EST).
- **Format**:
    - You'll submit this via Google Classroom. Again, it can be in whatever format you like - so long as it uploads! Take care if you're trying to use a Google Doc, however, as the copy/paste function of code seems to be a little funky.
- **Working Style**
    - You can do this individually or as a group - it is entirely up to you! If you do work in groups, please make note in the document that you did so and list everyone's name.

Data

In the spirit of the holiday - we're going to examine some *halloween candy* data!

The data is available here. You can gather more information about the data set here. You can use the code below to download the data.

```
library(tidyverse)
candy <- read_csv(url("https://raw.githubusercontent.com/kylescotshank/applied_data
```

Note: for full credit, you *must* provide the functions you used to obtain your answers!

Questions

Answer these as thoroughly as you can and please provide the code that you've used to generate your answer.

1. Generate a facet grid of bar plots showing the distribution of each variable. Your output should show a grid of bar plots where you get a bar for TRUE and FALSE counts of each variable. **Note: this will be much, much easier if you remember to use the `pivot_longer` command)**. Provide a written interpretation of your visualization.

2. Generate a plot where you show `competitorname` of each candy bar and that candy bars value for `pricepercent`. Make sure that this plot has `competitorname` on the **y-axis** and the `pricepercent` on the **x-axis**. Order the chart so that the values go from largest to smallest (top to bottom). **Note: you'll need to use coord_flip() to do this! Make sure you**

look up that function with `?coord_flip` to see how it works works. You'll also need to reorder the bars - to do this, use the `reorder()` function. Use `?reorder` to see how it works.

3. Run a t-test that compares the average of `sugarpercent` between candy bars that are chocolate versus those that are not. Report the results and interpret your findings. Visualize the differences between these groups with a boxplot.

4. Create a scatterplot showing the variable `sugarpercent` on the x-axis and `winpercent` on the y-axis. Color the points based on whether or not the candybar is fruity or chocolate. Add linear regressions that show the different relationships for chocolate and fruity candybar types. Provide a written interpretation of your visualization.

5. Run a regression with `winpercent` as the dependent variable and everything else as independent variables. Print the `summary()` of the output of this model. Plot all of the coefficients of this model as a bar chart from largest to smallest. Color the bars based on whether or not the coefficients are positive or negative. **Hint: you'll want to use the `tidy()` function to easily get the coefficients out of your summary to plot them!**. Provide a written interpretation of your visualization.