Homework 3

Overview

Objectives

- Identify a public dataset (formatted as a .csv)
- Practice using R to load and view characteristics of a dataset
- Practice using ggplot2 to create visualizations

Grading

- Uploaded *all* requested files: 5%
- File is properly/clearly formatted: 5%
  - Proper section headers for each part of your homework
  - You clearly indicate which question each of your responses is associated with
  - **Look at your compiled pdf before submitting it. Make sure it looks well-formatted.**
- Part A is worth 45%
- Part B is worth 45%

Deliverables

- .pdf file (generated by knitting your .Rmd file)
- .Rmd file (used to generate your pdf file)
- .csv file (the dataset you chose for this assignment)

Setup

Create a new R Markdown file with the title "Homework 3". For this assignment, you'll need to have the tidyverse collection of R packages installed. If you haven't already, go ahead and install them by running `install.packages("tidyverse")`.

In your R markdown file, create a section heading for each of the following parts of your homework:

- Part A. Visualizing data with ggplot2
- Part B. Loading data in R

Part A. Loading data in R

1. Find a publicly available dataset that is represented as a .csv file. Your chosen dataset should not be trivial (# rows * # columns should be at least 250 cells), and you may not choose the same dataset that I demonstrate in the lecture material thus far. When choosing a dataset, keep in mind that you need to creating at least three different types of plots using your chosen dataset. You will upload the chosen data set as a .csv file along with your homework submission.
2. Tell me about your data set. Where did you get it? Who/what created it?
3. Load your dataset into a data frame (or a tibble, your choice)
4. Print the number of rows and columns (using R code)

5. Print the column names (using R)
6. If your dataset is not tidy, reformat it (using R code) to be tidy.
7. Use the `mutate` function (from the dplyr package) to add a column to your data frame. It's okay if the new column isn't useful.

Part B. Visualizing data with ggplot2

Using ggplot2, create three different *types* of plots (use at least three *different* geoms) using your chosen data set from Part A. Be creative! Feel free to draw inspiration from the R Graph Gallery.

For full credit, you must:

- Create three different plots (use at least three *different* geoms)
- Explicitly label your x- and y-axes on each plot (e.g., using the `labs` function)
- On at least one of your plots use either the color of fill aesthetic