Lab 1: Data Visualization
BEE 4850/5850

> **Due Date**
>
> Friday, 2/2/26, 9:00pm

> **💡 Tip**
>
> You can find a Jupyter notebook and a Julia 1.11.5 environment in the homework's Github repository. You should feel free to clone the repository and switch the notebook to another language, or to download the relevant data file(s) and solve the problems creating your own notebook or without using a notebook at all. In either of these cases, if you using a different environment, you will be responsible for setting up an appropriate package environment.
>
> Regardless of your solution method, make sure to include your name and NetID on your solution PDF for submission to Gradescope.

Overview

Instructions

The goal of this lab is for you to experiment with ways to visualize data. You may work in groups of up to 2.

You should come to class prepared with a visualization and associated data that you would like to deconstruct and reconstruct.

Load Environment

The following code loads the environment and makes sure all needed packages are installed. This should be at the start of most Julia scripts.

```
import Pkg
Pkg.activate(@__DIR__)
Pkg.instantiate()
```

The following packages are included in the environment (to help you find other similar packages in other languages). The code below loads these packages for use in the subsequent notebook (the desired functionality for each package is commented next to the package).

```
using Random # random number generation and seed-setting
using DataFrames # tabular data structure
using CSVFiles # reads/writes .csv files
using Distributions # interface to work with probability distributions
using Plots # plotting library
using StatsBase # statistical quantities like mean, median, etc
using StatsPlots # some additional statistical plotting tools
```

Problems

Problem 1

Include an image of the original visualization (make sure to provide a reference to the original). Deconstruct the visualization:

- What feature(s) of the data do you think the original author was trying to emphasize?
- What audience were they trying to reach?
- Did they do so successfully?

Identify any "problematic" features of the visualization. These can include questions around data appropriateness or quality or perceptual problems; think in terms of

- Bad taste;
- Bad data;
- Bad perception.

Problem 2

Reconstruct the visual to highlight what you feel are the most important or salient parts of the data and to address any concerns you raised in Problem 1. Why do you think your new visual does a better job?

Note that this problem is likely to involve a fair amount of iteration and experimentation. That's okay! Great visuals are usually the result of a lot of trial and error. What they usually have in common is a clear idea of what message the creator would like to communicate ahead of time, as every "good" choice is downstream of that.

References