

Exercise Set 03: Spurious Correlations

BEE 4850/5850, Fall 2024

Due Date

Friday, 2/9/24, 9:00pm

Tip

You can find a Jupyter notebook, data, and a Julia 1.9.x environment in [the exercise'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 without using a notebook. 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 exercise is for you to find datasets and reason about the relationships (or lack thereof!) between variables.

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 DataFrames # tabular data structure
using CSVFiles # reads/writes .csv files
using Plots # plotting library
using StatsBase # statistical quantities like mean, median, etc
using StatsPlots # some additional statistical plotting tools
```

Problem

Find a single or multiple datasets (don't just pull from [Spurious Correlations!!](#)) where two or more variables appear to be correlated, but this correlation is likely spurious. Plot the relevant variable(s) and show they are correlated through any needed quantitative and/or qualitative means. Explain why you think the correlation is spurious.

References