Interactive statistical graphing and analysis of poverty, population, unemployment and education data of US states

Dinesh Reddy Poddaturi, Yue Zhang

Outline

- Download the data

- Clean the data using techniques we learned

- Visualize the data using Shiny interactivity

Data

- https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/

- 4 files (poverty, population, unemployment rate and education)

- Each file has about 3300 rows and 50 columns with a file size of 1.5 Mb

Techniques used to clean the data

- *"tidyverse"* package is used to clean the data and extract the relevant data for analysis

- Select the variables of interest from each file

- 'split-apply-combine' strategy to clean the data and make it tidy

- 'gsub' function to replace any special characters in the names of the variables of interest

- These steps makes our data into nice long format. This data is used to do analysis and do visualization in the project

Shiny Application

Discussion

- Came to a conclusion that education and unemployment rate are related to poverty

- Population doesn't seem to be very significant

- Cannot get to proper conclusion without any numerical analysis

- Could fit a linear model with unemployment rate, education, birth rate, and death rate as explanatory variables