DS 5500 Homework 2 - Due Oct. 28

Instructions

Create a new public Github repository for this homework assignment. The repository should include all of the code necessary to reproduce your submitted solutions. Do not include the raw data in the repository.

Use the README.md of the Github repository to present your solutions. Answer all questions completely for full credit, including figures and tables where appropriate. For each problem, either provide relevant inline code snippets, or cite the source file where the relevant code lives (with line numbers if appropriate).

Describe any data processing steps (transformation, filtering, etc.) you perform when solving each problem, providing reasoning where appropriate. You may need to be creative when deciding how to approach each problem, as there may not be a single "correct" solution.

Your solutions should be posted as a ***public*** *note* on Piazza in the *hw1* folder with the title "[hw2] - your name name". Include in the body of the note a link to the Github repository with your solutions.

Overview

This homework asks you to visualize data from the Gapminder Foundation (https://www.gapminder.org).

Download the full Gapminder dataset from the Open Numbers Github repository:

- https://github.com/open-numbers/ddf–gapminder–systema_globalis

Problem 1

Confirm your group for the second project component. Check in with me if the group is different from the first project component.

Post a ***public*** *note* on Piazza in the *project* folder with the title "[group] all group members' last names". Include in the body of the note all of the group members' full names.

Problem 2

Choose and critique one of the visualization by one of your fellow classmates for HW 1 Problem 2 (distribution of income across countries and continents over time). Include a link to the original.

Describe the visualization and how it is similar and/or different from yours. Is it easy to interpret? Does it effectively visualize what is being asked? Why or why not?

Problem 3

Choose and critique one of the visualization by one of your fellow classmates for HW 1 Problem 2 (relationship between income, life expectancy, and child mortality over time). Include a link to the original.

Describe the visualization and how it is similar and/or different from yours. Is it easy to interpret? Does it effectively visualize what is being asked? Why or why not?

Problem 4

Choose and fit one or more models to quantify the relationship betweem income (GDP per capita) and life expectancy over time. Justify your choice of model and comment on its appropriateness. (You are not required to handle the autocorrelation of time series, but should comment on how this impacts your analysis.)

Visualize the model(s) and comment on what they tell you about the relationship between income and life expectancy over time.

Problem 5

Choose and fit one or more models to quantify the relationship betweem income (GDP per capita) and child mortality over time. Justify your choice of model and comment on its appropriateness. (You are not required to handle the autocorrelation of time series, but should comment on how this impacts your analysis.)

Visualize the model(s) and comment on what they tell you about the relationship between income and child mortality over time.