Team members:
Isabel Garcia Pietri
Lucas Barbosa
Jamie Smith

W200 Project 2 Proposal

21st July 2020

Overview

Our research will seek to understand how income inequality (measured by the Gini coefficient) has trended over the last decade for the major regions and countries around the world, and explore its potential relationships with economic growth and globalization.

Goals

The main goal in this study is to analyze income inequality across the world and its relationship with economic growth and globalization.

The final report will answer the following research questions:

1. What is the trend of income inequality across the world in the last decade? Which countries/regions are doing better/worse?
    - Hypothesis: Income inequality is growing over time across the world.
2. What is the relationship between economic growth and income inequality? (E.g. Is economic growth a precondition for income inequality reduction?)
    - Hypothesis: Countries with an increase in economic growth have decreased income inequality.
3. What is the relationship between exposure to globalization (world trade) and income inequality?
    - Hypothesis: Globalization leads to the reduction of income inequality.

Optional research questions:

4. What is the relationship between income inequality and measures of conservatism and liberalism as political constructs?
    - Hypothesis: Liberal governments have less income inequality.
5. What is the relationship between global health and income inequality?

- Hypothesis: Less income inequality leads to a reduction of mortality due to preventable diseases.

Research questions 4 and 5 are going to be developed if time allows. In this case, additional data sets will be included (e.g. World Health Organization data sets).

Workflow

The general workflow that the project will follow is:

1. Data cleaning and merging.
2. Exploratory analysis.
3. Answer research questions.
4. Visualizations and report.

Variables, plots, figures and tables

Most analyses will use the following variables: Country, region, population, Gini coefficient (for 10 years), country's income group, GDP growth (for 10 years), trade indicators (like level of exports and imports as proportion of a country's GDP), among others.

Some of the analyses will rely on countries individually, while others will rely on some grouping, like regions, to get an understanding of the trends around the world.

Some summary metrics will be generated to understand countries with respect to 3 factors:

1. Income inequality performance over time (for instance Gini coefficient difference between years).
2. Economic growth over time (for instance average GDP growth rate).
3. Level of economy's openness to world trade (exports + imports as % of GDP)

Regarding plots, we will rely on a mix of types. Time series lines to see overall trends by region, bar charts to compare a selected number of countries among them, scatter plots to explore the relationship among different variables (e.g. Gini coefficient vs. GDP growth), maps, among others.

The general flow of the final report will sequentially walk through the 3 research topics as outlined above. The primary keys throughout the analysis will be combinations of locations (country, region, global) and year. This will allow us to tie the various data sets together.

GitHub repository

Link: https://github.prod.oc.2u.com/UCB-INFO-PYTHON/Project2_Pietri_Smith_Barbosa.git

Python instructors have access to the team repository.

Data sets

World Income Inequality (published by the academic and research arm of the UN).

Link: https://www.wider.unu.edu/project/wiid-world-income-inequality-database

GDP growth (published by the International Monetary Fund)

Link:
https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/ADVEC/WEOWORLD

World's Trade Data:

Link: https://data.wto.org/