

Galvanize Capstone I Proposal

Chris Sulfrian

Question

Can the amount that a country spends on education per year be correlated with educational outcomes later in life? Some of the things I intend to investigate are: percentage of the labor force with advanced education, pupil to teacher ratios, and then these relations relative to no more than 3 other countries.

H_0 = greater education spending as a percentage of GDP has no result on long term outcomes

H_A = greater education spending does have a positive effect on long term outcomes

The Data

A broad dataset on education statistics from the World Bank, which is quite large (887,000 rows x 70 cols). It came from Kaggle and is similar to the dataset we did our week one case study on. It encompasses 242 countries and regional groupings of countries with numerous metrics that relate to education. Additional columns are labeled by year starting at 1970 and ending in 2100 (spaced every 5 years beyond 2020, and as far as I can tell all empty). Every country/region has the same set of metrics called Indicator Names, though often the data is quite sparse before the mid 1980s. Overall this dataset has hundreds of thousands of NaNs, but I will try to minimize their effect by choosing Indicators that are widely reported and sampling ranges of years that have the highest density of real values.

Minimum Viable Product

- Perform EDA on the dataset to verify the fields chosen are similar in size and scale.
- Perform T-tests on the data to test the H_0 and H_A
- Create a framework that allows easy testing on other countries/groups

MVP+, MVP++

- Repeat for three other countries
- Compare correlations between all four countries

