DSCI 510 Project Proposal

Title: Does Spending More on Education Lead to Better Global University Rankings?

Author: Hengxiao Zhu (zhuhengx@usc.edu)

Objective

The project will investigate whether countries with a higher share of education in GDP will have more top global universities. Specifically, the project aims to determine whether national investment in higher education, measured as a percentage of GDP, correlates with a country's representation in the QS World University Rankings Top 100. The project will explore this relationship using a visual descriptive analysis approach that avoids overly complex models. These findings may shed light on the relationship between national investment in education and global academic standing.

Data Sources

1. Education Expenditure by Country (as % of GDP)

Source: World Bank Open Data

URL: https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS

Method: Retrieved via API or CSV

Description: Annual public spending on education as a percentage of GDP by country. The most recent 5 years will be averaged to ensure stability.

2. QS World University Rankings (Top 100)

Source: QS Top Universities

URL: https://www.topuniversities.com/university-rankings/world-university-rankings/2024

Method: Extracted using pandas.read_html() or BeautifulSoup for web scraping

Description: University names, rankings, and country of origin from the Top 100 global universities.

3. GDP and Population

Source: World Bank

GDP URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

Population URL: https://data.worldbank.org/indicator/SP.POP.TOTL

Method: CSV download or World Bank API

Description: Provides economic and demographic context; may be used to normalize results (e.g., universities per capita).

Research Questions and Data Integration

1. **Data Collection and Cleaning**
   Retrieve and clean each dataset. Align country names and select consistent year ranges for comparison (e.g., latest QS rankings vs. average education spending from the past 5 years).
2. **Data Integration**
   Merge datasets by country. Count the number of universities each country has in the **QS Top 100** and align that count with corresponding education expenditure data.
3. **Exploratory Analysis**
   Compute correlation between education expenditure (% of GDP) and number of Top 100 universities. Identify outlier countries (e.g., high spending but low representation, or vice versa).
4. **Visualization**
   a) **Scatter Plot**: % of GDP spent on education (x-axis) vs. number of Top 100 universities (y-axis).
   b) **Bar Chart**: Top 10 countries by university count, overlaid with their spending levels.
5. **Interpretation:**
   Finally, interpret the results to see if a trend emerges. For example, the project will discuss whether higher spenders generally have more top universities, and note any outliers or regional patterns. The emphasis is on insight from data rather than advanced modeling.