Final Report

1. Name and Author

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

Author: Hengxiao Zhu

---

2. Motivation/Rationale for the Project

The quality of a country's higher education system significantly affects its global competitiveness, knowledge innovation, and social development. However, whether higher education excellence—as reflected in prestigious global university rankings—correlates with national education investment remains an open question. This project aims to explore whether countries that invest more in education, particularly as a percentage of their GDP, tend to have more universities ranked in the QS Top 100.

This question is of both academic and policy relevance. For policymakers, understanding this relationship can inform how public resources are allocated to higher education. For students and researchers, it sheds light on structural advantages or limitations embedded in national education funding. Existing literature suggests that while education funding plays a role in shaping university infrastructure and faculty recruitment, global rankings also reflect factors like international collaboration and academic reputation. Therefore, a nuanced data-driven exploration of this relationship is needed.

3. Description of Data Sources

To address the research question, data was collected from multiple reliable sources. The project integrates educational expenditure, GDP, population statistics, and QS World University Rankings across countries for the period 2018–2022.

3a. Exact Data Sources and Extraction Approach

- **QS World University Rankings 2024**
  Source: https://www.topuniversities.com/university-rankings/world-university-rankings/2024
  Extraction: The top 100 universities were scraped using Python (BeautifulSoup), and the number of universities per country was aggregated.
- **World Bank Education Expenditure (% of GDP)**
  Indicator: SE.XPD.TOTL.GD.ZS
  Source: https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS
  Extraction: API-based programmatic download with fallback CSV scraping.
- **World Bank GDP (current US$)**
  Indicator: NY.GDP.MKTP.CD
  Source: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
- **World Bank Population**
  Indicator: SP.POP.TOTL
  Source: https://data.worldbank.org/indicator/SP.POP.TOTL
- **Taiwan (Data Not Available on World Bank)**
  - GDP: Taiwan Ministry of Finance
  - Population: National Statistics Taiwan
  - Education Spending: Estimates based on Taiwan's Ministry of Education Budget Reports
- **Hong Kong SAR**
  - GDP & Population: Census and Statistics Department Hong Kong
  - Education Spending: Not published post-2018 in World Bank; estimated via local government budget reports.
- **Malaysia**
  - GDP & Population: Department of Statistics Malaysia

- Education Spending: Limited data available on World Bank post-2019; estimates interpolated based on local reports.

3b. Modifications and Challenges

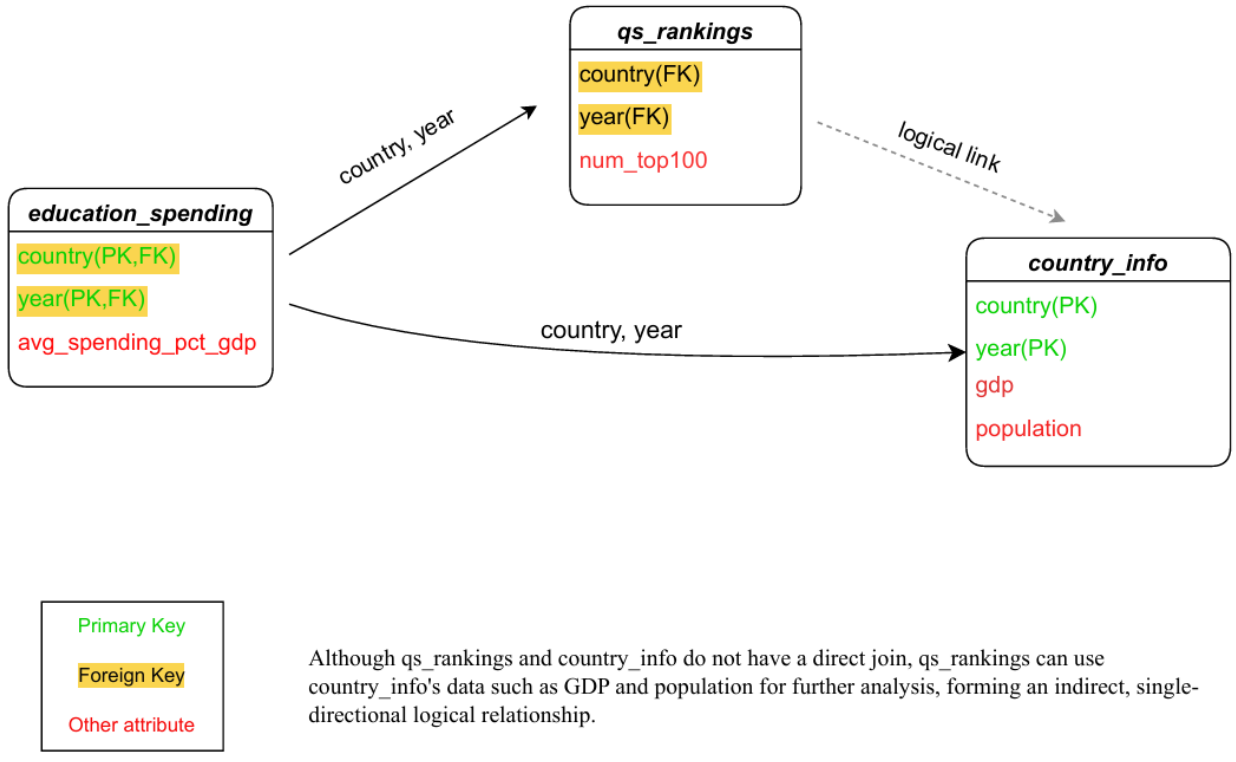
The original plan focused solely on education expenditure and QS Top 100 presence. However, during the data integration phase, it became evident that GDP and population were essential to normalize investment levels and assess per capita influence. Key challenges included:

- **Missing Values**: Some countries lacked complete five-year records for all indicators. Linear interpolation was used to fill internal gaps, while edge missing values (e.g., only 2022 missing) were estimated based on trend slopes.
- **Taiwan/Hong Kong Data**: As non-UN members, their data was missing in World Bank datasets and had to be added manually using local government sources.
- **Country Name Discrepancies**: Ensuring consistency across datasets required mapping country aliases (e.g., "China (Mainland)" vs. "China").

4. Integrated Data Model

To integrate heterogeneous datasets from QS Rankings, World Bank indicators, and manually supplemented sources, I designed a structured data model that aligns key variables across countries and years. The central concept of our model is a country-year pair, which serves as the composite key linking all attributes—namely, education expenditure, GDP, population, and QS performance. The unified data table contains the following fields:

- **Country Name**: The full name of the country or region.
- **Year**: From 2018 to 2022.
- **Education Expenditure (% of GDP)**: Annual spending on education expressed as a percentage of GDP.
- **GDP (current US$)**: Total gross domestic product in nominal US dollars.
- **Population**: Total population count for the year.
- **QS Top 100 Universities**: Number of universities ranked in the QS Top 100 for that year.



[ER Diagram](data models.drawio.pdf)

This model allows for efficient cross-variable analysis by year and by country, supporting visualizations such as time-series trends, scatter plots, and bubble charts. All datasets were merged on a shared structure of country and year to maintain consistency and ensure accuracy across derived metrics.

5. Analyses and Visualizations

5a. Analysis Techniques

I performed data cleaning, interpolation, and aggregation using Pandas and NumPy. Missing values (especially in Malaysia, Hong Kong, and Taiwan) were either fetched via supplementary API calls or filled using linear interpolation where appropriate.I normalized and grouped the data to allow meaningful comparisons across countries, both temporally and relationally.

5b. Visualizations

I created several visualizations to investigate relationships between education expenditure and university rankings:

1. **QS Top 100 University Count by Country**
   A bar chart showcasing the number of QS Top 100 universities per country. I adjusted label spacing and width to accommodate countries with long names.
2. **Education Expenditure Trend (2018–2022)**
   A line plot illustrating year-wise changes in education expenditure across selected countries. Labels were placed using the adjustText library to minimize overlap with lines.
3. **Education Spending vs. QS University Count**
   A scatter plot based on the 5-year average education spending against the number of QS Top 100 universities. Each point is labeled with the country name. This plot highlights that higher spending does not always lead to more top-ranked universities.
4. **Bubble Chart: Education Spending vs. GDP vs. QS Performance**
   A multi-dimensional bubble plot using education expenditure (X-axis), GDP (Y-axis), and number of QS Top 100 universities (bubble size). Country names are shown using adjustText to reduce clutter. This visualization underscores the complex interplay between national wealth, education investment, and global rankings.

6. Conclusions

Through this project, I investigated whether higher national investment in education correlates with better global university rankings. By integrating multi-source data — including GDP, population, and education expenditure — with the QS Top 100 University data, I built a comprehensive dataset that enabled cross-country analysis across five years (2018–2022).

The analysis revealed a nuanced relationship:

- **High investment does not guarantee high rankings**, as seen in countries like Sweden and Denmark, which spend a significant percentage of GDP on education but have relatively fewer universities in the QS Top 100.
- **Economic scale and global visibility** appear to play a stronger role. Countries with large economies (e.g., the United States, United Kingdom, China) are more likely to have more top-ranked universities, even with varying levels of spending.
- **Consistent investment matters.** Countries like Australia and the UK maintain stable education spending and continue to have strong QS representation. This suggests that steady support — rather than short-term spikes — may yield better long-term academic outcomes.
- **Regional disparities persist.** While Europe and North America dominate rankings, rapidly developing regions such as East Asia are catching up, especially with strategic investments (e.g., China’s "Double First-Class" initiative).

Overall, while education expenditure (as % of GDP) provides a partial view, **it alone is not a strong predictor of QS performance**. Instead, a country's global academic ecosystem — including research funding, international collaboration, and language of instruction — also plays a critical role.

7. Future Work

If given more time and resources, this project could be extended in the following directions:

- **Include more variables** such as R&D expenditure, number of academic publications, international faculty ratios, or student mobility rates to better model academic excellence.
- **Switch from % of GDP to absolute education spending**, which may better reflect capacity in large economies like China and the US.
- **Use more advanced statistical models or regression techniques** (e.g., multivariate linear regression, decision trees) to assess causality between investment and ranking outcomes.
- **Improve data completeness** by collaborating with national ministries or academic databases (e.g., UNESCO UIS, OECD Education at a Glance) for more reliable and granular spending data, especially for regions underrepresented in global datasets.
- **Create interactive dashboards** using Plotly, D3.js, or Tableau to allow dynamic filtering of year, country group, or QS rank ranges — making the analysis more exploratory and engaging.
- **Investigate policy-level interventions**, comparing how strategic education investments (like South Korea's SKY university policy or China's "Project 985/211") correlate with ranking improvements over time.

This study serves as a foundational analysis that sheds light on the complex, multi-dimensional relationship between public investment in education and international academic recognition. It opens the door for future research into how countries can optimize their higher education systems for global excellence.