Group Name: Global Citizens

Project Title: Gender Equality and Standard of Living

* **Concept: Anthony Brown**

This project compares indicators for gender equality in multiple countries with the standard of living in each country. The goal is to ascertain whether there is a positive correlation between gender equality and standard of living for countries around the world. To measure gender equality, we collected data for educational attainment, labor participation, and monthly income by gender.

Our source for educational attainment and labor participation data was the World Bank’s Gender Data Portal: [*https://genderdata.worldbank.org/*](https://genderdata.worldbank.org/). Monthly income by gender was sourced from the International Labor Organization (ILO) an agency of the United Nations: [*https://ilostat.ilo.org/topics/women/*](https://ilostat.ilo.org/topics/women/). For standard of living, we used GDP per capita data collected from the World Bank: [*https://data.worldbank.org/indicator/NY.GDP.PCAP.CD*](https://data.worldbank.org/indicator/NY.GDP.PCAP.CD)*.*

* **Data collection and limitations - Justin Caldwell**
  + Once we figured out the concept of our project we discussed exactly what data we needed to collect.
  + We decided on collecting GDP data, education data, salary, and wage data. Originally, we sought to collect data for the last five years focusing on North, Central, and South America.
  + In our first attempt at collecting this data, we realized that we needed to parse data from different sources, and would not be able to get education and wage data for the last five years for all countries in North, Central, and South America.
  + We pivoted and decided to use a sample of the countries that had solid data. We also decided to reduce the time frame of five years to the year 2020 because that year had the most available data.
* **Data Preparation - Ajeeta Sen**
  + Our project reads three different CSV files stored as three different data frames containing 1) Income data 2) GDP per Capita and 3) Gender Statistics respectively.
  + GDP and Gender Stats data frames were merged to create a single dataframe using a **l**eft join on the Country name.
  + The data frame was further refined to only include columns required for our analysis and to drop empty cells and rows.
  + Income Data was refined to exclude rows and columns that were not relevant and remove nulls.
  + Male and female income data was switched from row-based to column-based.
  + Data types were changed when necessary - US dollar to numerics, series to lists.
  + Male and Female Income Ratios were computed and GDP data were merged to create a new data frame with no null values.
  + Male and Female at least Secondary education percentage ratios were computed and merged with the data frame containing income ratios to be used for statistical analysis.

**Data Analysis and Interpretation: Owusu Bekoe**

We analyzed the data by asking three main questions. The first question explored the correlation between various education indicators of the sampled countries by gender and their GDP per capita. We hoped to understand whether the incremental attainment of education based on gender matched the countries’ economic performance. In other words, does the attainment of higher education for women as compared to men tell us anything about a country’s economic performance? We compared each of the following to GDP per capita: bachelor’s degree (male, female), lower and upper secondary (male, female), and primary education attainment (male, female). Using st.pearsonr, we found a positive correlation (coefficient over 0.5) between all the education indicators, except primary school education, and GDP per capita. We visualized our calculations using scatter plots. We consider these findings with skepticism as the correlation is negligible because the R-values were close to 0 and the data set comes with several limitations that prevent us from drawing strong conclusions. Nevertheless, our data analysis suggests that typically higher educational attainment by males and females reflects in the economic performance of our sampled countries.

As part of our first question, we also investigated whether countries’ economic performance for the year 2020 reflected the ratio of female-to-male labor force participation. Using the same statistical analysis as the above, we found that there is a positive (although not strong enough) relationship between female-to-male labor force participation and GDP per capita. The correlation coefficient for this comparison is 0.49. We suspect that a much bigger data set would give us a better idea of the trends. We showed our calculations using scatter plots.

Our second question focused on the relationship between the gender pay gap and GDP per capita. Essentially, we investigated whether countries with higher female-to-male income ratios performed better economically. We found a negative correlation between the female-to-male income ratios and GDP per capita. The correlation coefficient for this comparison was -0.24. Once again, we suspect that the data limitations obstruct us from seeing the global trend more clearly. We further compared male and female income by country through a double bar chart. The bar chart showed a higher pay gap in wealthy countries like Germany, the Netherlands, and the United States. A closer inspection of this data might reveal that, perhaps, the difference between male vs female income appears significantly greater for Germany, The Netherlands, and the US because they generally have higher income rates (as measured in USD) as compared to the other countries.

Our third question explored the relationship between the ratio of female-to-male educational attainment and the ratio of female-to-male income. We found a weak but positive correlation: a coefficient of 0.31. The scatter plot used to visualize our findings showed a slight upward trend, reiterating the impediment of our small sample size/data set.

**Conclusions: Anthony Brown**

Through collecting and analyzing data for multiple countries, we did find positive correlations between gender equality indicators for education, labor participation, income and GDP per capita, which indicate that gender equality may play a role in elevating a country’s standard of living. However, the correlation in the data we analyzed was not strong enough to be conclusive. It is possible that expanding the data to a broader set of countries would reveal more consistent trends. We also were unable to analyze any potential trends by region as the countries for which we were able to identify usable data were scattered randomly throughout the globe. Another potential approach would be to view data for multiple years to determine if gender equality influences standard of living over time.