Final Project: A Global Analysis of Gender Inequality

By Lluvia Brambila & Ann Liang

1. Description of the Dataset

This dataset is titled "Gender Inequality Index by Country" from Kaggle. The Gender Inequality Index (GII) is a comprehensive metric designed to assess gender inequality and disparities in a society by accounting for a number of important factors. This indicator sheds light on the disparities and imbalances that people encounter according to their gender. The three main elements of the GII—reproductive health, empowerment, and economic activity—are an extension of the Human Development Index (HDI). One important component of the GII is reproductive health, which includes measures like teenage birth rates and maternal mortality rates. These metrics show the differences in women's health outcomes, particularly with regard to reproductive rights and maternal health.

Ultimately, a high HDI means the country has a high lifespan, the education level is high, and the gross national income per capita is high. However, we have to note that this dataset talks about the HDI in rank; hence, the country with an HDI rank of 1 has the highest HDI. The gender inequality index (GII) sheds light on differences between the sexes in terms of empowerment, employment, and health. In contrast to the human development index (HDI), higher GII scores, however, correlate with less success.

This dataset offers a wealth of historical information on global gender development indices. Crucial columns are as follows:

- **ISO3**: ISO3 for the Country/Territory
- **Country**: Name of the Country/Territory
- **Continent**: Name of the Continent
- **Hemisphere**: Name of the Hemisphere
- **Human Development Groups**: Human Development Groups
- **UNDP Developing Regions**: UNDP Developing Regions
    - **SSA**: Sub-Saharan Africa LAC–Latin America and the Caribbean
    - **EAP**: East Asia and the Pacific
    - **AS**: Arab States
    - **ECA**: Europe and Central Asia
    - **SA**: South Asia
- **HDI Rank (2021)**: Human Development Index Rank for 2021
- **GII Rank (2021)**: Gender Inequality Index Rank for 2021
- **Gender Inequality Index from 1990 to 2021**: Gender Inequality Index from 1990 to 2021

2. Research Questions

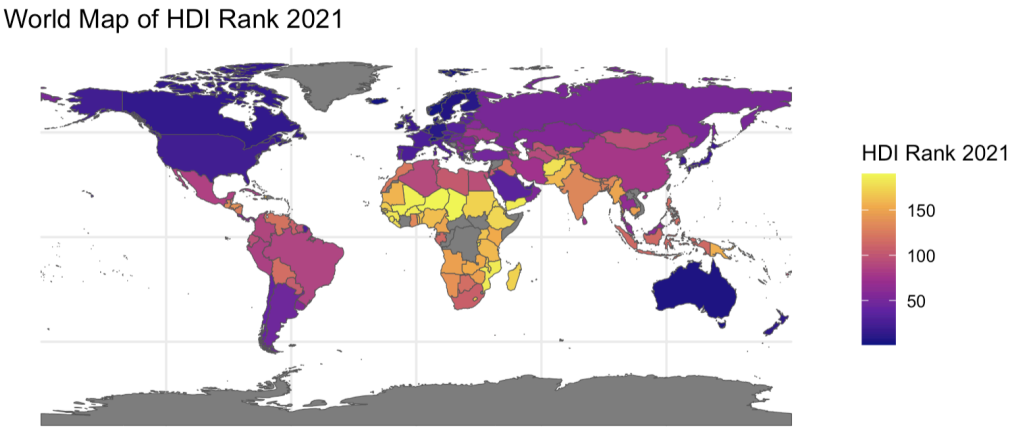
To explore gender inequality across the globe, we aim to address the major inquiry–**How has gender inequality evolved over time and across different regions?** Based on this interest, the analysis spans across the aspects below:

- How does gender inequality vary across continents?
- Which countries have the highest and lowest GII ranks within each continent?
- Does human development impact gender inequality?

- What trends can be observed in GII scores in the U.S. between 1990 and 2021?

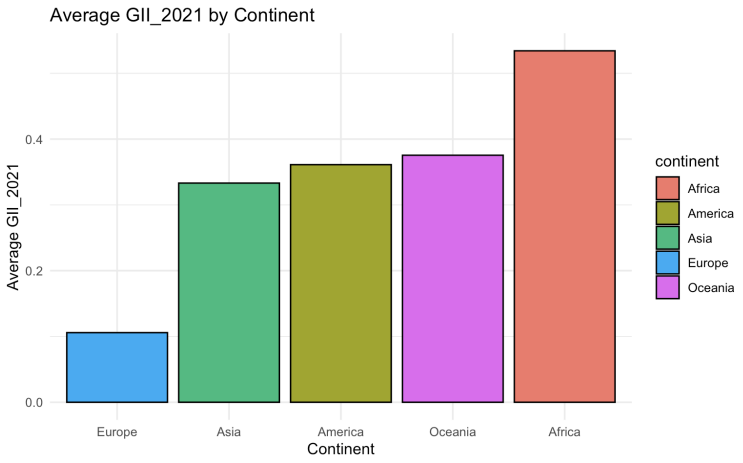
3. Data Analysis

Figure 1: World Map of HDI Rank in 2021



In Figure 1, we used an existing dataset, "world," to merge with our own dataset that we are using in this project to demonstrate the world map diagram so we can clearly see the varying HDI rank in 2021 by country. We had to mutate the name of the country column in the world dataset to match the column name in our own data. We added a key to the diagram, so it was easier to read according to the color-coded countries. There were some missing data in either our own data set or some that didn't match the "world"'s data; for example, there was no data for Antarctica in either dataset even though it is represented visually on the map, but most countries are represented accordingly. We added the minimal theme so the background was clearer, and it gave a better look to the map.

Figure 2: Average GII in 2021 by Continent

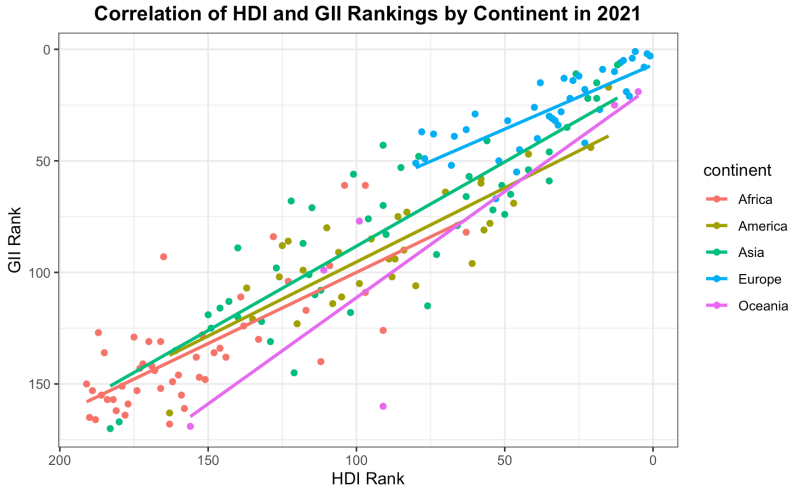


In Figure 2, we created a histogram demonstrating how the average GII in 2021 varied by continent. We used the year 2021 data for this because 2021 was the most recent year that our dataset had data for. We calculated the average GII for 2021 between all the countries within a continent and for all continents and used that data for the diagram. We used the reorder function to reorder the diagrams from smallest to largest so it was easier for the viewer to see the difference between each of the continent's data. We added the minimal theme so the background was clearer, and it gave a better look to the diagram.

Figure 3: Distribution of Countries with Highest and Lowest GII by Continent (1990-2021)

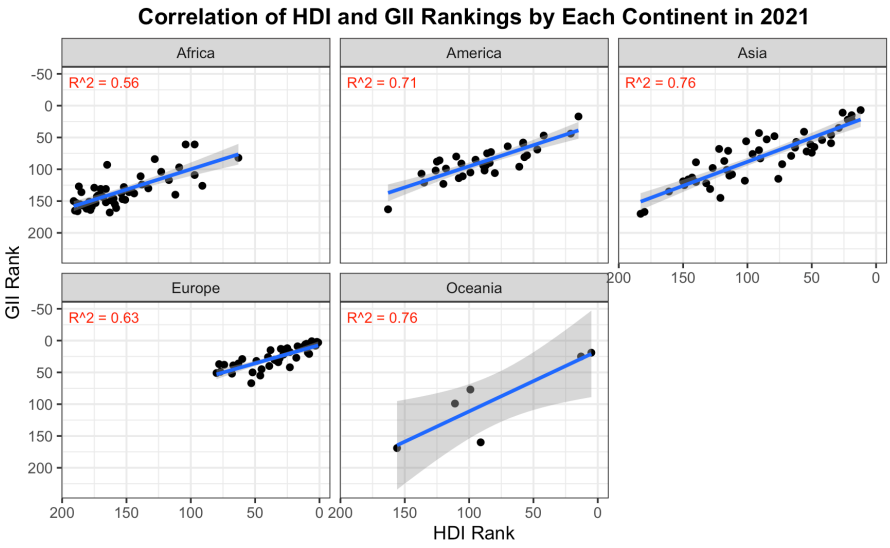
Figure 3 presents a boxplot illustrating the distribution of countries with the highest and lowest GII scores within continents from 1990 to 2021. We filtered each continent to identify countries with the maximum and minimum GII ranks. To process the GII values, we utilized the pivot longer function, as all scores begin with "GII_" followed by the year. We also removed the missing values, reordered the GII values, and adjusted the unit on the y-axis based on the distribution.

Figure 4: Correlation of HDI and GII Rankings by Continent in 2021



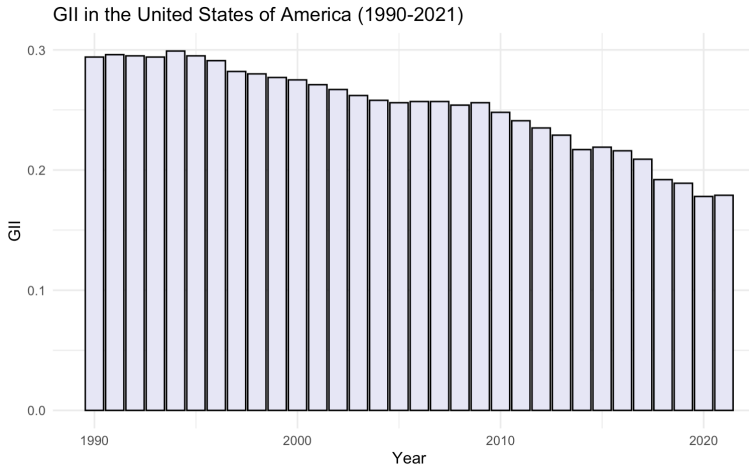
In Figure 4, we created a scatterplot to explore the correlation between HDI and GII rankings across the continents in 2021. Each point on the plot represents one country, and the x and y axes represent the HDI and GII rankings, respectively. To make the interpretation more intuitive, we reversed both axes to align with the perception that lower ranks indicate better performance, with the ranking from the highest as 1 to the lowest as 200. In addition, we added a linear regression line to show the overall trend within each continent. The plot is color-coded by continent, allowing us to visualize any differences in the relationship between HDI and GII rankings across continents.

Figure 5: Correlation of HDI and GII Rankings by Each Continent in 2021



To inspect the strength of the correlations, we calculated the coefficient of determination (R-squared) for each continent's relationship between HDI and GII rankings in 2021 in Figure 5. Using the previous graph, we faceted it to visualize each continent and added annotations on it to display the R-squared values. In the fitted linear regression, we also highlighted the variance between the two rankings with confidence intervals around the lines.

Figure 6: GII in the United States of America (1990-2021)



In Figure 6, we created a visualization demonstrating the United States's GII score change over time (1990-2021). We filtered the data so it would only focus on the United States of America's data, and then we used a pivot longer function when dealing with all the GII values since all of them started with "GII_" and then the year. We used the year values in those variable names to become part of the x-axis.

4. Results

4.1. How does gender inequality vary across continents?

As shown in Figures 1 and 2, continents with more developed countries, such as the USA, Canada, Australia, and most of Europe, exhibit higher HDI rankings and lower GII scores. These regions benefit from better education, advanced medical care, and higher incomes, leading to greater gender equality. In contrast, continents with more developing countries, particularly in Africa, have lower HDI rankings and higher GII scores, reflecting less gender equality. This highlights the significant impact of development on reducing gender inequality.

4.2. Which countries have the highest and lowest GII ranks within each continent?

According to Figure 3, within Africa, Tunisia exhibits the lowest GII score, with a consistently moderate spread across the years, while Liberia has the highest score. In the Americas, Canada displays the lowest score with a moderate distribution over time, while Haiti exhibits the highest score. Additionally, Singapore holds the lowest GII score in Asia, demonstrating consistently high values over the years, whereas Yemen reports the lowest score. In Europe, Denmark shows a persistently low GII score with a moderate spread and an outlier, while Romania presents a higher score with a mild spread. Australia in Oceania records the lowest GII score with a slight distribution, while Papua New Guinea reports the highest score. Countries with the highest GII scores in their respective continents tend to display less variability in the boxplot, indicating a relatively stable circumstance of gender inequality over the years.

4.3. Does human development impact gender inequality?

Based on Figures 4 and 5, HDI significantly impacts GII scores. Figure 4 shows that European countries, with the highest ranks in both HDI and GII, are highly concentrated at the top, indicating strong gender equality. In contrast, other continents like Asia, America, Africa, and Oceania display more varied distributions. Figure 5 reinforces this result, with R-squared values indicating the strength of the relationship between HDI and GII rankings. Africa shows a moderate correlation ($R^2 = 0.56$), Europe a stronger one ($R^2 = 0.63$), the Americas even stronger ($R^2 = 0.71$), and Oceania and Asia the strongest ($R^2 = 0.76$). This highlights a significant link between higher human development and reduced gender inequality, especially in Oceania and Asia.

4.4. What trends can be observed in GII scores in the U.S. between 1990 and 2021?

Figure 6 As we can predict, the United States has become more equal over time. There has been an uprising in women's empowerment in recent years, pushing for equal pay and equal opportunities between men and women, and we can definitely see this reflected in the diagram above; the jump between the more recent years is a bit larger than the jumps between years in the 1900s. In the United States, gender inequality has become more equal over time.

5. Conclusion

In summary, our analysis reveals significant differences in gender equality worldwide, largely tied to a region's level of development. More developed areas like the USA and Europe tend to have less gender inequality, while less developed regions like Africa face greater disparities. Within each continent, specific countries vary in their gender equality levels. The connection between development and gender equality is clear, especially in regions like Oceania and Asia. In the United States, we see some progress toward gender equality over time. Overall, our findings stress the importance of promoting fairness and opportunity for all.