Exploratory Data Analysis For Gender Inequality

Step 1:

We chose to study the gender_development.csv and gender_inequality.csv files.

Our initial questions that we seek to answer with our data analysis are:

1. What factors significantly affect the gender inequality (especially in regards to the female population) in countries around the world?
2. Is gender development in different countries tied to education, income and/or life expectancy?

Step 2:

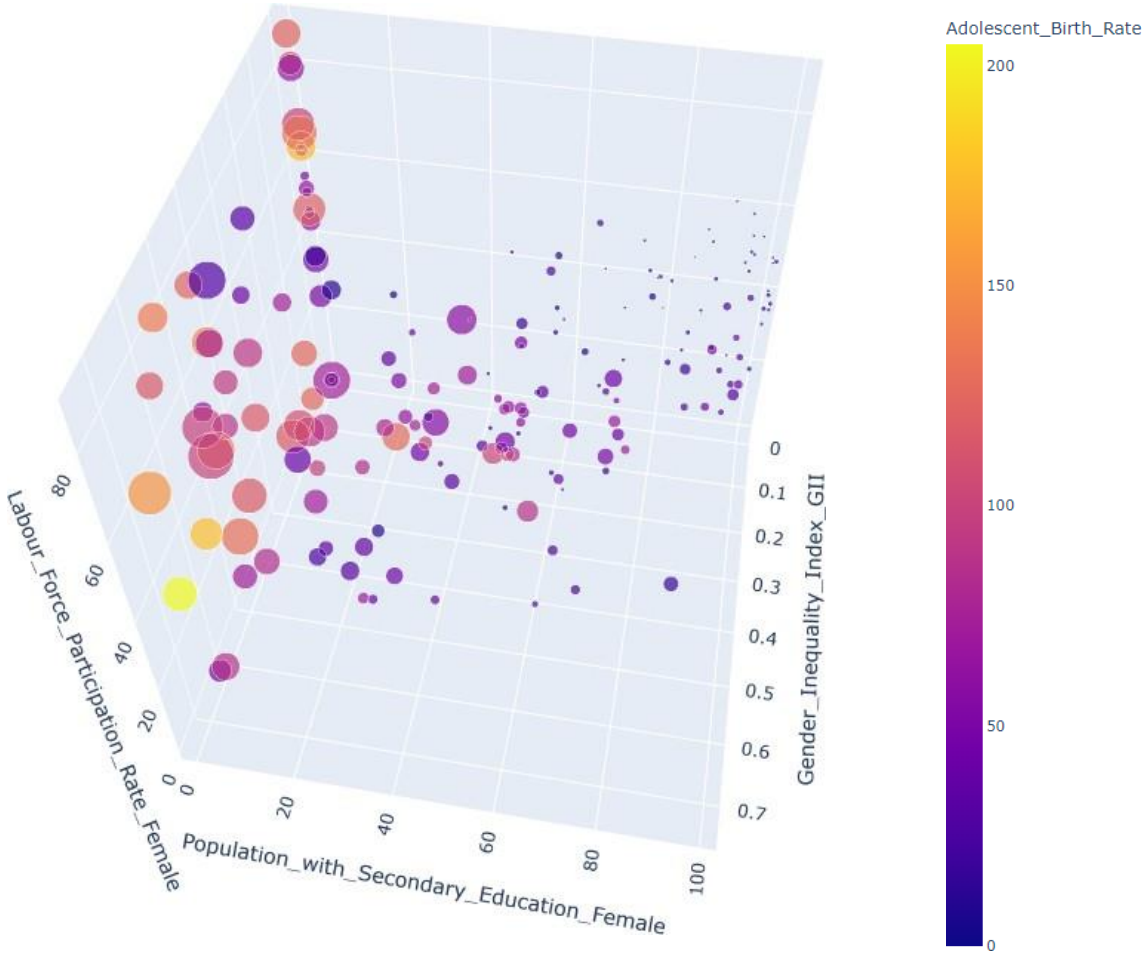
When attempting to answer our questions, there are some parts of the data that need to be adjusted. Before creating any of the visualizations, we created a female_education.csv comprising of the top 10 and bottom 10 countries based on the female secondary education population. We also added the ISO 3166-1 alpha-3 country code for all countries in gender_development.csv in order to create the map based visualization.

Exploratory Data Analysis For Gender Inequality

Step 3:

Visualization 1:

Comparison of female labour force rate, female population with secondary education, gender inequality index (GII), adolescent birth rate and maternal mortality ratio

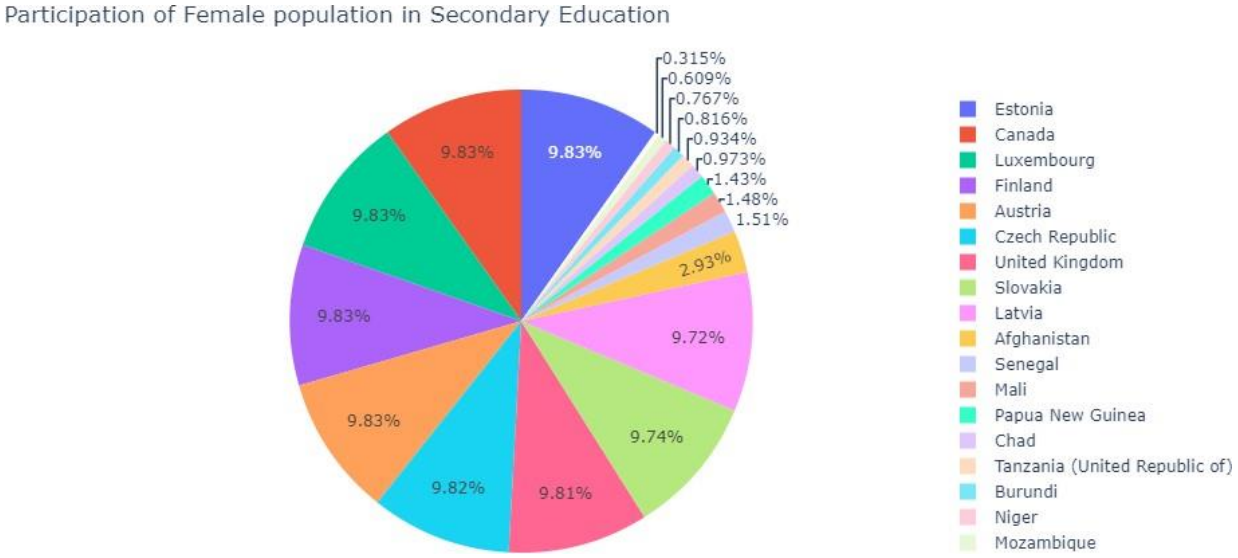


For the first visualization based on gender inequality, we decided to make an interactive 5D scatter plot. This plot shows the relation between female labour force rate, female population with secondary education, gender inequality index (GII), adolescent birth rate and maternal mortality ratio. The interactivity and the 5D nature of the plot lend well to analysing these variables, as the user can freely move the plot around and hover over any point to see its exact values. The color gradient from purple to yellow makes the data visually distinct and easy to parse. From the plot, we can perceive that with increasing female secondary education population, the gender inequality index trends towards 0. As a result, we can presume that a

Exploratory Data Analysis For Gender Inequality

higher educated female population contributes to more gender equality. We can also observe that as the GII decreases and female education rate increases, the maternal mortality ratio and adolescent birth rate decrease. This is to be expected, as educating the female population and creating an environment of gender equality increases awareness of maternal complications/contraception and better equips the female population, causing less adolescent pregnancies and less general maternal mortalities. There is conclusion we can determine from female labour force participation as the data is quite spread out. We can also observe a vertical line of outlier data in one of the vertices of the plot which muddies the data. This plot allows us to visualize all our variables in relation to the female population and determine that education is a great contributor to a lowered GII, despite the outlier data.

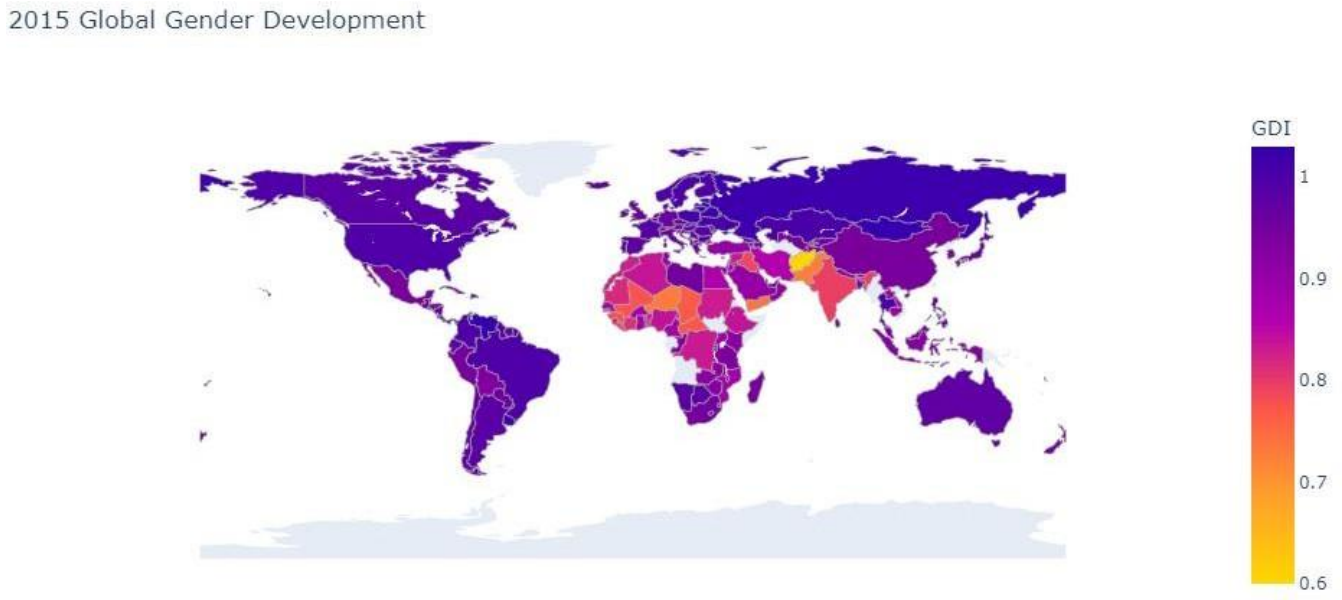
Visualization 2:



For the second visualization based on gender inequality, we decided to make an interactive percentage pie chart. As mentioned above in step 2, we created a separate data set with the top 10 and bottom 10 countries based on their female population with secondary education and GII in order to better determine the effect of education on GII. From the chart, we clearly see that the top 10 countries with the lowest amount of inequality have significantly more female population with secondary education than the bottom 10 countries with the highest GII. This reinforces our conclusion that access to education is the biggest contributor to more gender equality. This chart, while helping us answer our question, could have been designed better. The numbers are a little bit confusing if the user doesn't hover over a pie slice and the colors are chosen at random with no meaning.

Exploratory Data Analysis For Gender Inequality

Visualization 3:

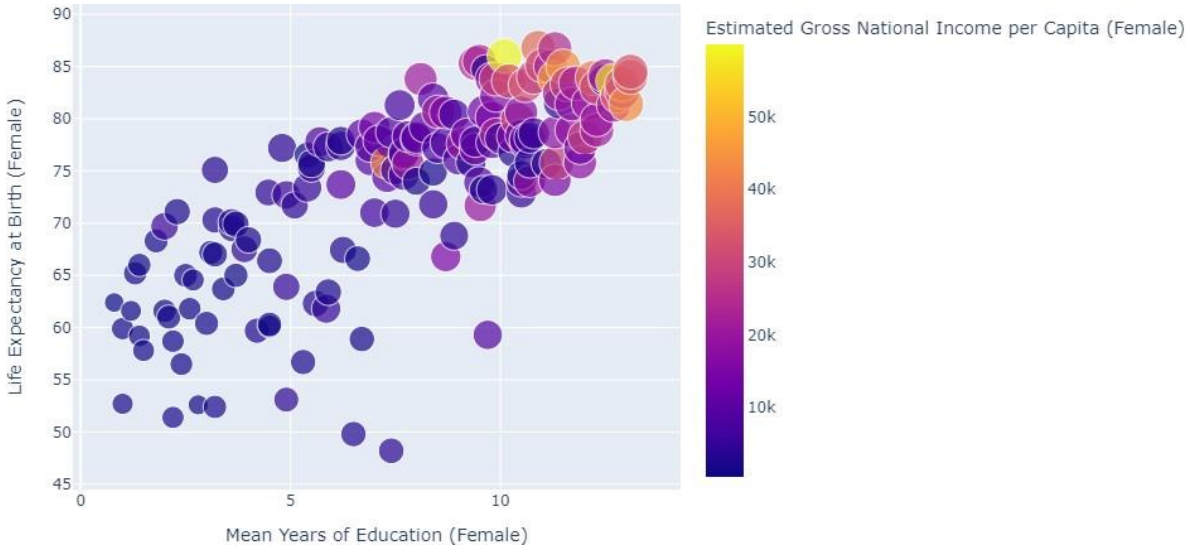


For the first visualization based on gender development, we decided to visualize a map of the world in terms of each country's GDI (Gender Development Index). The heat map of the world provides an immediately clear picture of the data, with the same visually distinct colors as the first visualization, but inverted as it allows for easier viewing in this context. The heat map is also interactable and display more information when the user hovers over a specific country. This map is the reason we needed to add ISO 3166-1 alpha-3 country codes to gender_development.csv. We can observe that most countries have GDIs between 0.8 and 1 which indicates a good amount on gender development. However, some areas around Africa and the Middle-East have lower GDIs. While this style of heatmap is useful for visualizing the data globally, there is no effective way to gauge the trends in the additional variables of the dataset.

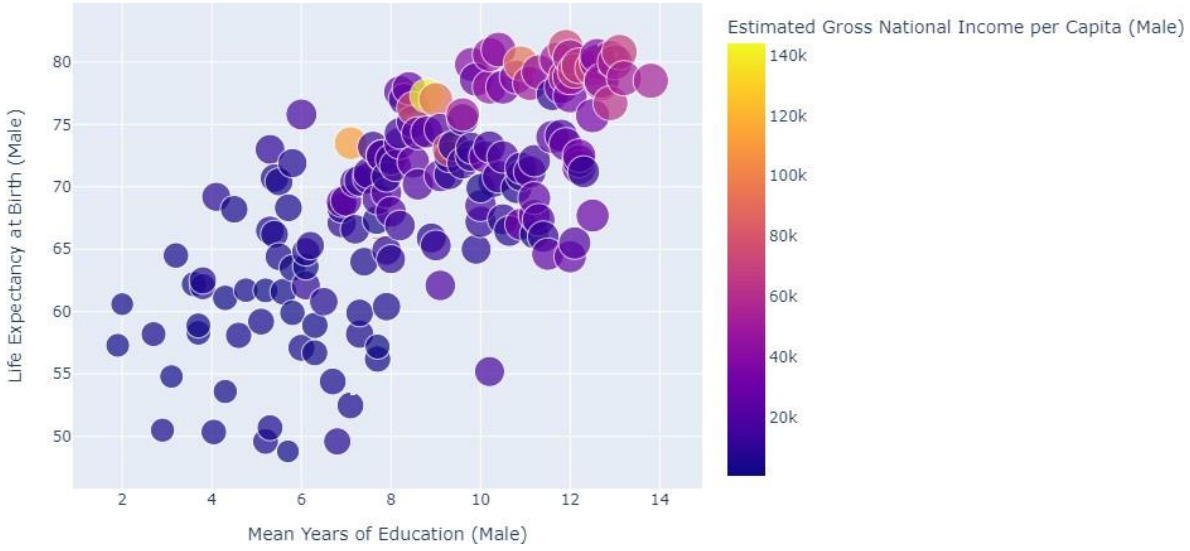
Exploratory Data Analysis For Gender Inequality

Visualization 4:

Comparison of Life Expectancy at Birth, Mean Years of Education, Estimated Gross National Income per Capita and Human development index for female populations of the world



Comparison of Life Expectancy at Birth, Mean Years of Education, Estimated Gross National Income per Capita and Human development index for male populations of the world



Exploratory Data Analysis For Gender Inequality

For the second visualization based on gender development, we decided to make an interactive 4D scatter plot. This plot shows the relation between life expectancy at birth, mean years of education, estimated gross national income per capita and human development index for both male and female populations. The plot provides a clear picture of the data and the trends between variables, with the same visually distinct colors as the first visualization. As with the first visualization, the interactivity and the 4D nature of the plot lend well to analysing these variables, as the user can freely hover over any point to see its exact values as well as zoom in to better understand specific clusters. Throughout both male and female plots, we can observe that higher life expectancy and years of education increase the GDI of countries as well as the income. This is to be expected, as an increase in life span and education facilitates the development of a nation and the generation of income. We can also observe that female populations tend to live longer and generate slightly larger income per capita than male populations, as well as having more consistent amounts of years of education. It must be noted that, before creating the plots, we removed any data with null values as the plots would have difficulty ordering the data otherwise. Thus, there may be some error in our results and due to this and the limited amount of data with few clear trends, we can only say, when answering our question, that life expectancy and education increase the gender development of a country.