

Examining Gender Disparity in Education

Goals:

Education is one of the most fundamental aspects of life and plays a crucial role in the economic and social improvement of a society. Historically, there has been an unfortunate disparity in the educational opportunities afforded to men and women in different societies.

The goal of this project is to design and implement an interactive data science application that allows users to visualize the progress that has been made in the field of education in terms of equality of opportunity for both genders. The app examines education data from across the globe over a decade (2000-2010) and helps the user visualize different metrics and configure them as well.

The dataset used for this project is the World Bank Education Statistics Dataset. It can be found at <https://www.kaggle.com/theworldbank/education-statistics>.

Design decisions made:

I designed the application to have two sections: Configuration and Display. In the configuration section, the user chooses the metric that he is interested in and the year for which they want to see the visualization. Additionally, the user can also choose whether to display a legend of the countries in the visualization. In this display section, the created visualization is displayed, and the user can interact with it.

I wanted the users of this application to be able to easily compare statistics for men and women visually. For this purpose, I decided to plot the metrics for different countries as an interactive scatter plot with the axes representing the two genders. Users can zoom and pan the graph to their liking and hover on the different points in the scatter to see relevant information associated with that datapoint. The size of a country's scatter point increases with increase in the gender disparity of the selected metric in that country.

I wanted the users to be able to easily adjust the year in question and therefore, I added a slider to select the year. In this way, a user can quickly change the year to see how the scatter changes over time. I used a drop-down menu to display the supported education statistics that the user can choose from. For toggling the legend, I used a checkbox.

Overview of development:

I worked on this project alone. The technologies used were Python and Streamlit. I divided up the work as follows (and spent a total of 18 hours):

- | | |
|-------------------------------------|-----------|
| 1. Selection of dataset | : 4 hours |
| 2. Exploration of dataset (EDA) | : 2 hours |
| 3. Defining the goals | : 2 hours |
| 4. Making a design | : 2 hours |
| 5. Coding the streamlit application | : 7 hours |
| 6. Creating a write-up | : 1 hour |

I spent a significant portion of time (>60%) on steps 1 and 5 above. The choice of dataset was crucial to the success of the application and the code for the application had to be free from bugs.