**Project Description**

For this project, we will be using the gender statistics hosted by the World Bank on their website. (The world bank data was provided as possible dataset that could be used for the project in the class.) We will be using the data to detect and highlight gender biases that are pervasive in the world in various sectors such as education, healthcare and employment.

Link to the data.

'http://databank.worldbank.org/data/reports.aspx?source=gender-statistics'

Our target audience are the groups that focus on gender-equality and actively seek to influence the political atmosphere to bring attention to

these issues and work to make sure legislations are passed to level the playing field. We will try our best to present the data in a format, that can be used by the government bodies as well.

We are working with data of different countries, according to us thus will help us create a better picture about the topic.

Questions we believe our project will be able to answer: -

* Is gender bias equally prevalent in sectors like education, healthcare and employment. If no, where does it differ?
* How much of a role does education play in combating the gender gap?
* How much of a role does economic factors (GDP of country, per capita income) play in gender bias?
* Does gender bias also depend on age?

**Technical Description**

We will be making a Shiny app for the final project. We will be reading static data that will be downloaded in .csv format from the World bank website. Since the data is huge it would need a lot of wrangling and manipulation. We will initially get rid of the data of the countries of which much of the data is not available. Then we would choose the topics of which the data is available for various countries and which can be used to show gender bias. Some of the data can also need reformatting or removal because it is difficult to use. We will be doing so using the dplyr library. After we are done with the data manipulation we will write the code for the shiny UI and the shiny Server making plots using the ggplot2 and plotly libraries and ensuring all the plots have proper labels and are displaying some sort of gender bias.

We would like to learn about the different plots that can be used to display the data in a more comparative way, so any change can be observed clearly. We will be using statistics to highlight the differences and also try to find scope where we can use simple machine learning (Available in R) for our project. (Although we are not making any predictions so we don’t think it would be helpful for us.)

The major challenges that we anticipate is managing such a large dataset and using appropriate plots to display the data efficiently.