

Project 2

For this project, I am using data from the world bank world development index. In particular, I am using the data on gender issues to find predictive indicators of child marriage. The databank has more than 40 years of data for some countries, but because of the sensitive nature of this topic, many countries have a lot of years missing. To compensate for this, I downloaded the data in a manner which presents each variable as an average of the 15 years from 2000-2014. This will maximize the number of countries I am able to include in the sample without using a range so broad that it is skewed.

After importing the data into stata, I saw that each variable had 3 listings one for percentage of females, one for males, and one for the total. All of these will be collinear in some sense and since I am more focused on women's empowerment I keep the percentage of females version of the indicator. Time permitting I will create a function to rename many of these variables as the World Bank dataset has lengthy and confusing variable labeling practices.

Once I import into python, my first step will be to explore the still missing data. Many of the smaller countries do not even have the child marriage indicator for one of these 15 years. Once I settle on my final list of countries to use in the prediction, I will merge in a region indicator. It makes sense that the region to which the country belongs will greatly impact the allowance of and laws produced regarding child marriage. Next, I will look at missing values in the explanatory variables and define a function to drop or replace the missing values depending on their importance. Depending on these results, I will create a mask using the just the variables without a lot of missing values.

After this, I will use histograms, bar charts and other visualizations to get a general understanding of the data and decide which direction to take the project. I expect that the outcome variable of interest will have some truncation at 0 since many countries have set in place strict laws regarding child marriage in recent years. I have downloaded the worldwide average for each of these indicators as well and will explore variations from the world average and use groupby to look at regional statistics.