**Overview**

**Theoretical Framework**

**Data Selection**

**Imputation of Missing Data**

Every variable I selected, with the exception of compulsary schooling had at least one missing value. I addressed them as follows:

* For all variables gathered from the UNESCO Institute of Statistics, the World Bank and the Kaggle dataset, if a country had data for a particular variable for any year from 2011 to 2020, but not for 2021, I took an average of the years available instead, as while it would likely not be completely accurate, it would serve as good approximation.
* For the variables related to free schooling, I assumed that no data meant that the respective country had none. I felt this was a reasonable solution as less than 25 countries were missing this information and mostly occured with countries with no formal schooling.
* I decided to drop the variables related to teacher attrition, due to a lack of data. Less than 1/3 of courtries had this data available for any of the previos 10 years.
* For the ratio between qualified teachers to pupils and the percentage of current teachers who are currently qualified, I replaced missing values with the global average.

**Multivariate Analysis**

**Normalisation**

**Weighting and Aggregation**

**Links to other Indicators**

**Visualisation of Data**