Motivation for GBV Dashboard

Ability to measure a goal is very important cause that's how we would know where we're coming from and where we're heading. Organizations such as UN, World Bank and Governments refer to this measures as indicators.

The first aspect to fight GBV is to know what we're fighting, what indicators do we want to improve and what indicators should go lower. I would suggest AFD thinks about this, and indentify and measure indicators that would help them achieve their goals against GBV. I came across a few PDFs that might help in making such metrics which are from UN and Rwanda's GMO.

Getting the best indicators would require thorough deliberation and evaluation among experts. But before then, we can take a step by using available public datasets such as World Bank, OECD and UN. UN has a lot of subsidiaries such UNFPA, UNECA among others.

The purpose of this project is to help in making informed decisions, actions, and policy.

My project

Database

The first stage is making a database. The aforementioned organizations are the source of the datasets that were provided. There are more indicators on those sites that might be useful but I find working with those sites quite difficult.

The database has three components:

- countries.csv
- Targets/
- Inputs/

`countries.csv` was derived from github, the purpose is to provide country names (instead of ISO3) in Choropleth map. I didn't use it cause it was not approved, but using it can be done by changing a `country_df_allowed` in line 39 of `dashboard.py`.

`Targets/` contains csv files that I think are good suggestions for target indicators. This is where new target csv with indicators we're trying to measure, increase or reduce. Example of target indicators are `Violence against women Percentage`.

However, a lot of times we can't control target indicators or hit them heads on. We need to direct actions, funds and policies towards controllable indicators that greatly influence the target. These indicators go into `Input/`.

`Targets/` and `Inputs/` are a good start but they would need to be improved. Adding a dataset is easy, it needs to be in csv and first column would be `Country` which contains countries ISO-3 code. ISO-3 is used over name because it is more robust and uniform. You can take a look at the CSVs to get an understanding.

These are the source of each dataset

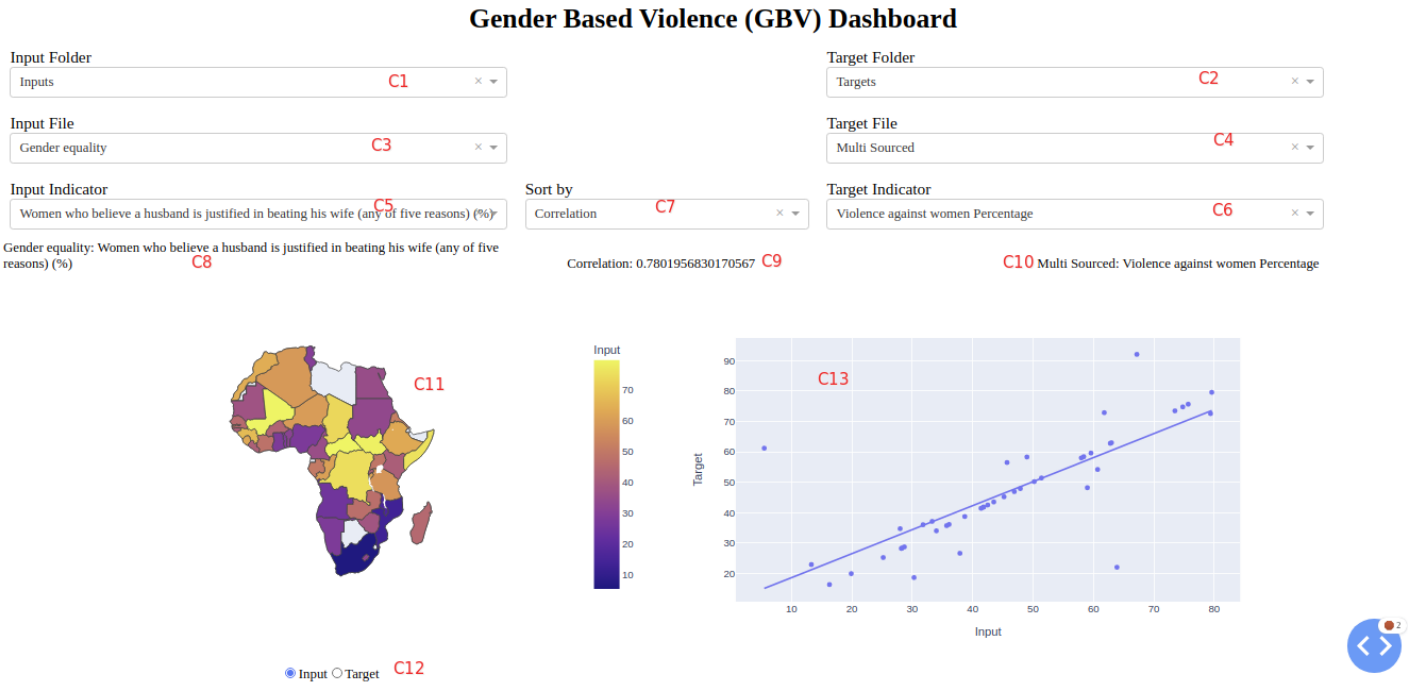
Inputs

- **FGM Attitudes.csv**: `'GBV'/'Bonus_Datasets'/'Copy of FGM-Attitudes-database_Feb-2020.xlsx'`
- **Gender equality.csv**: `'GBV'/'Gender equality Index _World Bank'/`
- **Gender gap index.csv**: `'GBV'/'Global Gender Gap Index'/'detailed-rankings-2013.csv'`
- **Justification of wife beating.csv**: `'GBV'/'Bonus_Datasets'/'Copy of Justification-of-wife-beating-among-adolescents-database_Oct-2019.xlsx'`
- **OECD...**: OECD which is where `GBV/Violence against women Percentage, 2019 or latest available.csv` was sourced
- **Prevention programmes**: `'GBV'/'Bonus_Datasets'/'Prevention programmes Data by country.csv'`
- **UNFPA.csv**: UNFPA which is where `FemaleData/` was sourced.
- **WB...**: From `'GBV'/'World Bank Data'/`, `'World Bank_Gender Based Violence Data'/`

Targets

- **Multi Sourced.csv**: Concatenation of `'GBV'/'Violence against women Percentage, 2019 or latest available.csv'`, `'GBV'/'Global Gender Gap Index'/'detailed-rankings-2013.csv'`, `'GBV'/'Gender Development Index (Female to male ratio of HDI)'/'gdi_female_to_male_ratio_of_hdi_geometric_gdi.csv'`, and `'GBV'/'Level of gender discrimination in social institutions'/'Level of gender discrimination in social institutions_2019.xlsx'`
- **OECD violence against women.csv**: `GBV/Violence against women Percentage, 2019 or latest available.csv` but fetched myself from original source.
- **Proportion of ever-partnered women and girls aged 15-49 years subjected to physical and_or sexual violence by a current or former intimate partner in the previous 12 months.csv**: `'GBV'/'Bonus_Datasets'/'Proportion of ever-partnered women and girls aged 15-49 years subjected to physical andor sexual violence by a current or former intimate partner in the previous 12 months_Country.csv'`

Dashboard



Components:

1. Input Folder (dropdown): This would hardly be changed except for rare cases when targets want to be compared with targets
2. Target Folder (dropdown): This would hardly be changed except for rare cases when inputs want to be compared with inputs
3. Input File (dropdown): This is for selecting a file from selected input folder
4. Target File (dropdown): This is for selecting a file from selected target folder
5. Input Indicator (dropdown): This is for selecting an indicator column from the input file.
6. target Indicator (dropdown): This is for selecting an indicator column from the target file.
7. Sort by (dropdown): How should Input Indicator be sorted. The options are Absolute `Correlation` with the target indicator or `Alphanumeric`
8. input indicator (text): The input indicator selected
9. correlation (text): correlation between input and target indicator. However, this is not always a good measure such as binary data and that's the purpose of the scatter plot.
10. target indicator (text): The target indicator selected
11. Choropleth (figure): This is a Choropleth of indicator chosen. Hover each country to see the input and target value
12. Choropleth (radio): Select between viewing input indicator or target indicator in the Choropleth color.
13. Scatter plot (figure): A scatter plot of input and target indicator with a line to show their correlation.

Component 1 & 2 has a few other applications. Some indicators in input--especially by World bank--are good indicators for targets too. But downside of it is that there are a lot of countries without value. Another application is to investigate where targets or inputs are duplicated.

It is important to note that the dropdown options can be searched. For example, if you want to focus on education indicators, you can simply search for keywords instead of scrolling.

Improvements

These are some of my ideas on improving the project.

- An issue that needs to be worked on are null values. For example indicator X might be available for 2006 and 2018 only for country Nigeria. And a lot of countries don't have a value at all. The first improvement would be better dataset with less null values which would be updated yearly especially for targets. If targets have regularly updated value, a line plot for trajectory can be added. Also based on intuitions, some particular input indicator can be handpicked and updated regularly such as (business related, education etc). It's all about good data.
- For each indicator, I picked the most recent year available. The year (if available) for each indicator are provided as a column named `f'[YEAR] {indicator}'`. If there are more data points this can be used for filtering. Note, the year is included in the hover data.
- Switch to SQL database (e.g sqlite).
- Host the dashboard on a website.
- Add other continents cause GBV is a **global** pandemic.
- Regularly updated data can also help us explore how change in input indicators, policy or funding (e.g `'GBV'/'Bonus_Datasets'/'Humanitarian financing estimates from UNFPA Humanitarian Action Overview reports, 2015-2018'/`) affects the target indicator.
- Getting a hand on intimate partner violence seems interesting. The data is identical to one of the available target indicator but UNFPA has subnational data. Fighting GBV locally and running experiments would help to know what works best.

Footnote

`setup.pdf` shows requirements and how to start the dashboard. Take a look at `report.pdf` to see a practical application or discovery while using the dashboard.

You can contact me at horlasehinde@gmail.com for enquiries or feedback.