

# Creating a Composite Indicator on Gender Issues

Tatiana de Feraudy, Huyen Tran, Matthieu Baillard  
IronHack Paris - Data Analytics - Project 2 - Apr. 24, 2023



# Process

1. Brainstorm of the final indicator
2. Setting up common tools (including github)
3. Paper/pen design of database
4. Choosing indicators
5. Importing/ scraping / downloading data
6. Data cleaning for MySQL (using Excel and Python)
7. Iterating on the final composite indicator
8. Slides



## **Gender Equality: a human right depending on wide-ranging issues**

- Healthcare and reproductive rights
- Education
- Economic empowerment
- Women's Leadership and Government
- LGBT+ Rights
- ...



## Data Collection - Health

- [Life expectancy at birth](#)  
Source: World Bank Gender Data Portal (51 countries missing)
- [Fertility rate](#) (births per woman)  
Source: World Bank Gender Data Portal (7 countries missing)
- [Births attended by skilled health staff](#)  
Source: World Bank Gender Data Portal (34 countries missing)
- [Maternal mortality rate](#)  
Source: World Bank Gender Data Portal (32 countries missing)
- [Obesity](#)  
Source: World Bank Gender Data Portal (32 countries missing) -
- Anemia rate in non-pregnant women



## Data Collection - Socio-economic

- [School enrollment gender parity index](#)  
Source: World Bank Gender Data Portal (32 countries missing)
- [Female labor force with intermediate education](#)  
Source: World Bank Gender Data Portal (41 countries missing)
- [Women, Business and the Law](#)  
Source: World Bank Gender Data Portal (28 countries missing)
- [Literacy rate gender gap \(W/M\)](#)  
Source: World Bank Data Portal



## Data Collection - Socio-economic (2)

- [Saved for old age \(in %, age 15+\)](#)

Source: World Bank Gender Data Portal (62 countries missing)

- [Unemployment & part time employment by gender](#)

Source: World Bank Gender Data Portal (62 countries missing)

- [Gender wage gap](#)

The gender wage gap is defined as the difference between median earnings of men and women relative to median earnings of men. Selected data refers to full-time employees. We reversed the standardized index to reflect the fact that the higher the wage gap, the worse the situation is for gender equality.

Source: OECD (2023), Gender wage gap (indicator). doi: 10.1787/7cee77aa-en (Accessed on 24 April 2023).



# Data Collection - Representation

- [Proportion of seats held by women in national parliaments.](#)  
World Bank Gender Data Portal - 24 countries missing.
- [LGBT+ Rights](#)  
The Equality Index is a crowdsourced measure of the current status of LGBT rights, laws, and freedoms as well as public attitudes towards LGBT people (lesbian, gay, bisexual, transgender), from 0 (Less Equal) to 100 (Most Equal).  
Data retrieved on Apr. 24, 2023 - Scraped via Python with Requests, Pandas, and BeautifulSoup libraries.



# Challenges

Indicators: finding, choosing indicators (based on quality, countries represented, keeping a single measure)

Scraping: finding a scrapable page, doing it !

CSV encoding (ASCII compatibilities, and issues with float formatting).

Normalization of country names (ISO 3166-1 Alpha-3 code) required both joining tables with reference tables, and manual correction of unmatched data. => *Look into country-converter library to make country normalization more efficient and robust.*

MySQL: formatting tables (column names, stdev use, joining several tables) => Excel, Python, but not MySQL





# Demo

Find your country!



# Improvements

Add indicators & manage the “null” issue

Database design for regular update

**Evolution in time**, rather than average of available data for each countries.



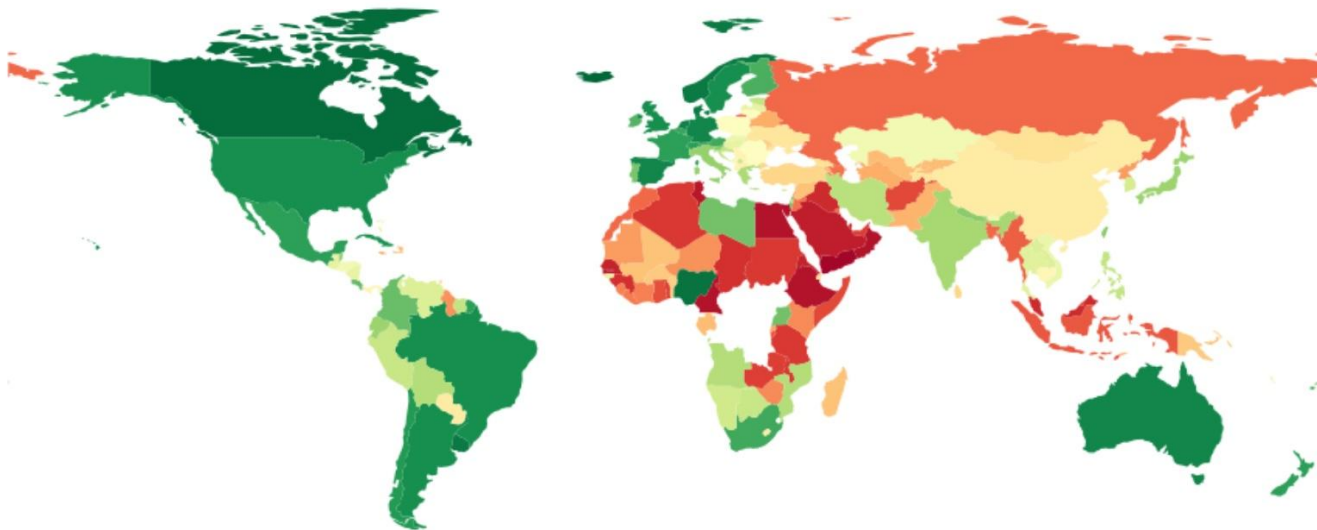
# Improvements

**Basic : Standardization formula**  $z = \frac{x - \mu}{\sigma}$  (where  $\mu$  = avg and  $\sigma$  = stdev) makes indicators with a wider distribution weigh more in the composite indicator. We might prefer the ***normalized score***,  $X_{\text{new}} = (X - X_{\text{min}}) / (X_{\text{max}} - X_{\text{min}})$ .



# Improvements: data viz

LGBT Equality Index



GeoPandas map test for one of our indicators