

Task 6.1: Sourcing Open Data

Parvin Rahimi

Data Source

Gender equality affects the society as a whole and can reliably indicate how equal and democratic a society is towards all groups. Its absence on the other hand is a significant indicator of lack of tolerance and understanding in that society. A study of the predominance of gender-based violence in a society can tell us a lot about the situation of marginalized gender groups and whether they are violently discriminated against. Since I couldn't find a wholesome open data on gender-based violence in all gender groups including non-binary people, I decided to use the data on violence on women ages 15 to 49.

The dataset is an open and external data and is published by Gender Statistics of The World Bank and is collected annually¹. It could also be considered a benchmarking data for countries when developing anti violence measures for specific gender groups in order to specify what to focus on in order to bring change and make the society safer for all genders.

“The main sources of data on violence against women are: (i) specialized surveys on violence against women, for example surveys using the WHO multi-country study instrument and methodology or the European Union (EU)-wide survey on violence and (ii) modules on violence against women within larger national health surveys, such as the DHS and RHS. A small number of data points came from other surveys such as national crime victimization surveys or MICS. (Further details [here](#))

Source: The United Nations Inter-Agency Working Group on Violence Against Women Estimation and Data (VAW-IAWGED) (WHO, UN Women, UNICEF, UNSD, UNFPA, UNODC)²

Data Profile

- Data is clean, consistent and has no missing values or duplicates. (see Python scripts)
- Irrelevant entries before 2000 were deleted.
- The type of data of Year and Value columns were changed:
Year from int64 to string(object)
Value from float to int64

Variable	Description	Time Variant/ Invariant	Structured/Unstructured	Qualitative/Quantitative	Binary/Nominal/Ordinal	Discrete/Continuous
Indicator Name	survey Questions	invariant	unstructured	qual	nominal	n/a
Indicator Code	specific code for the survey question	invariant	structured	qual	nominal	n/a
Country Name	name of the country/region	variant	structured	qual	nominal	n/a
Country Code	code of the country or region	variant	structured	qual	nominal	n/a
Year	year the data was collected and registered	variant	structured	quan	nominal	discrete
Value	answer to the survey question in form of quantitative value	variant	structured	quan	binary/ordinal	discrete

¹ <https://datacatalog.worldbank.org/search/dataset/0037654/Gender-Statistics>

² <https://genderdata.worldbank.org/topics/violence>

Limitations and Ethics

Data entries are anonymized, and thus most probably do not violate data privacy laws at analysis and sharing level. It is unclear if this was also the case when collecting the data.

Because of the source of this administrative data, it is trustworthy and reliable. This, however, doesn't mean that the data is not biased. Collecting data on (domestic) violence and finding the right survey questions to gather the data is very challenging. Some of the biases that may occur in this case are:

- Domestic violence is a cultural issue that may be understood differently across cultures.
- Depending on the atmosphere that the data was collected, and the survey conducted, the interviewees may not be completely honest, especially if they fear legal or cultural repercussions.
- Some small local institutions that have carried out the survey locally may be corrupted or have cultural biases.
- The gender and the position of the interviewees may affect the responses the women give.
- If the data were collected manually, it can also be prone to human error.

Questions to explore

Objective: Finding in which regions in Europe there has been an increase in violence on women since 2000?

Hypothesis: If a woman lives in Eastern Europe, then it is more likely that that person will experience gender-based violence compared to a woman living in other subregions of Europe.

- How is the overall historical situation in Europe in relation to violence on women since 2000?
- Is there a regional geographic element to violence on women in Europe?
- How is the situation regionally?
- Are there any regional differences in Europe?
- Since when do we see these differences?
- Do some countries within a region behave differently than the rest of the region?

Note: U.N. Geoscheme of Europe: 4 Subregions

Eastern Europe³	Western Europe	Northern Europe	Southern Europe
Russia	Germany	Iceland	Spain
Belarus	Netherlands	Norway	Portugal
Ukraine	Belgium	Sweden	Italy
Poland	Luxemburg	Finland	Greece
Czechia	Austria	Estonia	Andorra
Slovakia	Switzerland	Latvia	Malta
Hungary	Monaco	Lithuania	Montenegro
Romania	Liechtenstein	Denmark	Bosnia and Herzegovina
Moldova	France	United Kingdom	Albania
Bulgaria		Ireland	North Macedonia
			Serbia
			San Marino
			Vatican
			Croatia
			Slovenia

³ Eastern Europe is the largest and most populous subregion of Europe.