Code book: PSY6422 final module project

The dataset used for the project was obtained via the World Health Organisation (WHO) statistics page. The original dataset provided detailed the crude rate of suicide per 100,000 population in each country globally from the year 2000-2019, including statistics for each gender individually and a combined average for both overall. This was then filtered using the website toggles to only include data for all sexes from specifically European countries. Once filtered it was then exported into raw Excel format ready for import and cleaning.

This was created by the WHO by amalgamating the average reported suicides in each country per year taken from each country’s official death records. It examines not only each sex, both together and individually, but also all age categories. Raw numeric averaged values are given for number of suicides per 100,000 as well as upper and lower bounds for each average.

The link to the website where the original dataset was found can be located here : <https://www.who.int/data/gho/data/themes/mental-health/suicide-rates>

The link to the filtered dataset to only include European countries is here :

blob:null/396b5ef0-7087-4c19-b3db-776007bb2b78

**Variables being observed:**

The original dataset contained many variables unnecessary to the current project – all of which were removed in the data cleaning portion of code. However important variables from the dataset being considered in this analysis that are to be understood include:

* ***Period***– this refers to the year in which the average reported number of suicides per 100,000 was taken. The current dataset reflecting a range of the years 2000-2019.
* ***FactValueNumeric***– this variable describes the average number of suicides per 100,000 population reported that year across all sexes.
* ***Location*** – meaning the country in Europe that the data point originated from (for example Moldova)

**Variables removed**

Multiple useless variables such as dim1 were removed from the raw dataset – however some potentially useful variables were removed by choice. These include

* ***FactValueNumericLower* –** the lower bound of the aforementioned average number of suicides per 100,000 population reported that year across all sexes
* ***FactValueNumericUpper* –** the upper bound of the aforementioned average number of suicides per 100,000 population reported that year across all sexes