

The Final Project CareerFoundry

Marta Majer

1. Data Source

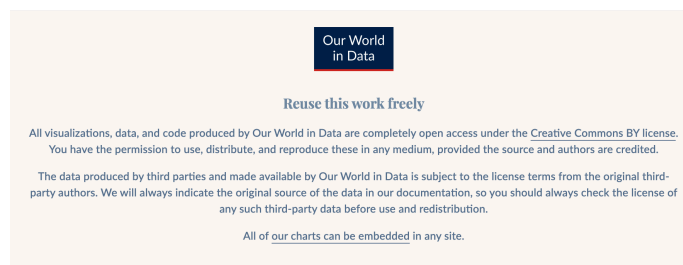
Database names - used in the analysis. A few more are available, but won't be relevant.

01_mental_illnesses_prevalence
02_burden_disease_from_each_mental_illness
04_adult_population_covered_in_primary_data_on_the_prevalence_of_mental_illnesses
05_anxiety_disorders_treatment_gap
06_depressive_symptoms_across_us_population

Data Source

Summarize the data source. Include whether it's internal or external data, who owns the data, and how trustworthy it is.

The data was sourced from [Kaggle](#), where it has a usability score of 10.00. Upon researching the scientist behind the dataset, I found their work in the [Google Scholar](#) database and on [Our World in Data](#), where they are currently employed. Given this background, I believe the external data is trustworthy.



Data Collection

Summarize the data collection method. Is it administrative data, usage data, or survey data? Is it collected manually or automatically? Is there a time lag?

The mental health data is based on survey data, collected through screening questionnaires. These surveys ask individuals about symptoms that align with those found in diagnostic manuals. While formal diagnoses require a healthcare professional, these surveys can be administered more easily and broadly without direct involvement from medical personnel, allowing for more widespread data collection. The data is likely collected manually through participants' self-reports. Depending on the survey schedule, there may be some time lag in capturing real-time data, as it reflects the period during which the survey was administered.

2. Data Profile

Data Overview

Write an overview of the data contents. What variables are included?

This dataset provides insights into the prevalence, impact, treatment coverage, and symptoms of mental health disorders. Key variables include:

1. General Information:

- **Entity:** Geographical region or country.
- **Code:** Unique identifier for the country or region.
- **Year:** Year of data collection (two datasets that base their relevance on time are from 1990-2019, the treatment gap is 2002-2017 and symptom frequency was just collected in 2014)

2. Prevalence of Mental Health Disorders (as a percentage of the population):

Schizophrenia disorders, Depressive disorders, Anxiety disorders, Bipolar disorders, Eating disorders.

3. Disability-Adjusted Life Years (DALYs):

- DALYs represent the overall disease burden, calculated as the total years lost due to premature death and years lived with disability, typically expressed per 100,000 citizens.
- DALYs effectively capture the impact of mental health disorders on both life expectancy and quality of life.

4. Treatment Coverage:

- **Potentially adequate treatment:** Proportion receiving effective treatment.
- **Other treatments:** Proportion receiving less effective treatments.
- **Untreated:** Proportion not receiving any treatment.

5. Symptom List (for depressive disorders):

- Appetite change
- Depressed mood
- Difficulty concentrating
- Loss of interest
- Low energy
- Low self-esteem
- Psychomotor agitation
- Sleep problems
- Suicidal ideation

Limitations

Be sure to note any limitations of the data set. Could the data be biased? Is it collected infrequently? Could it contain manual errors?

Self-Disclosure: Respondents may not feel comfortable sharing their mental health symptoms, leading to underreporting. This discomfort can vary across cultures.

Memory Recall: Individuals might struggle to accurately remember past symptoms, particularly older adults recalling experiences from years ago.

Potential Bias: Social desirability and stigma can influence responses, creating bias in the data.

Manual Errors: Data entry mistakes can occur, impacting the reliability of the findings.

Not Full Data: not all countries are fully covered by the survey - only North America has 100%

Relevancy

Use the project objective and your hypothesis to determine the relevancy of the data set to your project.

I chose this dataset because I'm currently pursuing a master's degree in psychology, and mental health is a topic I'm passionate about. Combining my new analytical skills with the subject of mental disorders makes this project particularly engaging for me. I aim to explore the differences in how various countries approach mental health, as well as relationships between factors such as years, symptoms, and treatment coverage. This will help me gain valuable insights and deepen my understanding of global mental health trends.

3. **Questions to explore**

- How have illness ratios changed over the years? What are the trends in mental health disorder prevalence over time?
- What is the distribution of mental illnesses across countries? How do different countries compare in terms of prevalence rates?
- Which mental health disorder is the most common, and where is it most prevalent? Identify the most widespread mental illness and the regions or countries where it's most prominent.
- How are people experiencing symptoms of depression in the US? (only North America has surveys that covered 100% of people experiencing the disease officially, yet we have symptoms only for the United States available).
- Which countries have the highest DALYs (Disability-Adjusted Life Years) due to mental health disorders?
- What is the relationship between DALYs and prevalence ratios? Are countries with higher prevalence also experiencing a higher burden of disease? **Ideal for correlation analysis.**
- Which countries show the fastest improvement in mental health indicators? Where are we seeing the largest reductions in prevalence or DALYs over time?)