

DATA 602 Proposal for Final Project

DATA 602 Proposal

Mridul Gupta. Prem Patel. Vardaan Bhatia

Purpose: *What is the domain? Within said domain, what are you attempting to investigate? What are practical implications of your analysis and findings? What is/are the population(s)? What variables are of interest? How have data on these variables been observed/recorded?*

The domain of interest of this study is mental health; the focus of this study will be to analyze the various mental health issues prevailing in the society and try to identify patterns for the causes leading to mental health concerns. The problem that we intend to investigate will be related to the personal and professional factors affecting people with mental health disorders based on the dataset from various parts of the world. We will also try to correlate the employment status and income brackets factors and try to build out a relation across multiple factors such as the corporate work, personal life and the location people are based out of.

This study will put emphasis on the changes and trends that can be identified between 1990 and 2016, the range of factors associated with it and how these factors affect an individual's life by being a reason for various mental disorders to develop. Effectively understanding, acknowledging, and going on a path to alleviate these issues is important for an individual's wellbeing and improving their overall quality of life. This project will deal with numerical analysis and understanding the data from a statistical point of view. The focus of this study will be in understanding trends of mental health noticed in our dataset in the attempt to draw logical conclusions through statistical and probabilistic methods to better understand the data we are working with as well as practically introduce concepts learned in class.

Data: *What is the source of your data set? Do you have permission to use the data? Ensure you cite the course of the data and the time-period over which the data was collected. (You do not want to use data that is rather antiquated.) What types of data visualizations to you perceive will be used to visually inspect the data?*

The source of our dataset is from Institute for Health Metrics and Evaluation (IHME), which is an independent global health research center at the University of Washington. Data is collected from year 1990 – 2017.

Our dataset is a collection of varied data points related to Mental Health Disorders across various age groups and countries.

DATA 602 Proposal for Final Project

It consists of data points across multiple mental health disorders by age, level of education, gender and contains prevalence across population along with suicide rates.

The resource available for download on IHME Websites can be used, shared, modified or built upon by non-commercial users in accordance with the **IHME FREE-OF-CHARGE NON-COMMERCIAL USER AGREEMENT**

Data on depression prevalence across several countries, disaggregated by education level and employment status has been gathered from Organisation for Economic Co-operation and Development (OECD)

We will be doing visualizations on RStudio to analyze our sample population which will show us their statistical significance of certain variables in our Data. These visualizations include the following but are not limited to:

- Bar graph
- Scatter plot
- Density Plot

Topic(s) to Investigate: *Outline the focus of your statistical investigation. Provide context to the data to education those who are unfamiliar with the data space you are working within. Provide citations any previous research/studies completed in the domain within which you are playing.*

As per WHO statistics [3], 1 out of every 8 people in the world is suffering from a mental disorder. Mental disorders significantly affect an individual's ability and hamper their thinking, emotional regulation, or behavioral natures. Our dataset [4] is a collection of varied data points related to Mental Health Disorders across various age groups and countries. It consists of data points across multiple mental health disorders by age, level of education, gender and contains prevalence across population along with suicide rates.

The dataset [4] has been sourced from <https://ourworldindata.org/mental-health#> and below are the citations for the author and their work/report.

The resource, as mentioned by the author, is open to use freely with proper citations under the Creative Commons BY License ([link](#)).

DATA 602 Proposal for Final Project

Guiding Questions

- What does the data reveal about a specific group (race, age, gender) suffering from mental illness?
 - Is depression higher in men or women?
 - Which age group has more depression?
- How does each country rank with its population suffering from a specific mental illness?
- Does the education level influence one's likelihood to experience depression symptoms?
- How do the employment status and income brackets affect a person suffering a mental health illness? What does this say about the corporate work environment?

Statistical methods to be used: Provide a list of the types of statistical methods you will use in your analysis. For example, if you are planning on creating a confidence interval for the difference between two population proportions, are you to do so via the “bootstrap technique” or the “conventional method” (or both)?

Some of the Statistical methods we will use in our analysis include but not limited to the following:

- **Bootstrapping and Sampling**
We will be computing the bootstrap samples of our given population based on factors like age, race, gender, education status and provide the distributions of the same. This will allow us to improve our understanding of the behaviour of our data from the known populations.
- **General Probabilistic Analysis**
We will be analyzing the means, medians, and standard deviation of the sample population based on various factors. Additionally, we will be using the data available to draw conclusions based on probabilities regarding the mental ailments that are covered in our dataset.
- **Confidence Interval Evaluation**
We will be performing Confidence Interval Evaluation by looking at different intervals from which we can make informative claims for the entire population with their relationship to the different types of Mental Illness.
- **Hypothesis Testing**
We will be trying to understand and validate the results with the help of Hypothesis testing and determine/predict certain scenarios that might lead to mental health diseases (for ex. people who are employed in IT sector are more prone to mental health diseases).

DATA 602 Proposal for Final Project

References

- [1] GBD results. *Institute for Health Metrics and Evaluation*. Available at: <https://vizhub.healthdata.org/gbd-results/> [Accessed September 29, 2022].
- [2] OECD *Education at a Glance*, *OECD statistics*. Available at: <https://stats.oecd.org/Index.aspx?datasetcode=EAG> (Accessed: October 4, 2022).
- [3] *International Classification of Diseases (ICD) (no date)* World Health Organization. World Health Organization. Available at: <https://www.who.int/classifications/classification-of-diseases> (Accessed: October 4, 2022).
- [4] Dattani, S., Ritchie, H. and Roser, M. (2021) *Mental health, Our World in Data*. Available at: <https://ourworldindata.org/mental-health#data-sources> (Accessed: October 4, 2022).