

Psychological Health and Stability

Galib Md Azraf Nijhum

Course - ITC510

MS Computer Science

Central Michigan University

Mount Pleasant, MI, USA

Global ID - nijhu1g

Riad Hossain

Course - ITC510

MS Computer Science

Central Michigan University

Mount Pleasant, MI, USA

Global ID - hossai1r

Alexander Richard Cramer

Course - ITC510

MS Computer Science

Central Michigan University

Mount Pleasant, MI, USA

Global ID - cramelar

Date - 2nd Dec, 2023

I. OVERVIEW

This study uses a large dataset covering the years 1990–2017 to investigate trends in mental health. The dataset, which focuses on the prevalence of mental health disorders, offers information on the proportion of the population that suffers from disorders such as eating disorders, drug and alcohol use disorders, bipolar disorders, anxiety disorders, depressive disorders, and schizophrenia. It is a combination of two datasets that each illustrate the prevalence of various mental health disorders in every country across the world for each year. Our goal in this analysis was to uncover temporal trends and relationships between different mental health disorders, as well as further examine the dataset and seek to uncover any statistical anomalies. For those unfamiliar with coding and data modeling, we used Exploratory Data Analysis (EDA) techniques by utilizing the Python coding language and its various libraries to reveal patterns and insights about the dataset. Through visually appealing line graphs, histograms, and scatterplots, we showcase how the prevalence of these disorders changes over the years and explore potential correlations between them. Our analysis sheds light on the evolving landscape of mental health, providing a valuable resource for understanding the distribution and potential interconnections among various mental health conditions.

II. DATASET

We have utilized two different datasets for our project and merged them together into one dataset in order to properly perform Exploratory Data Analysis on the data. The first dataset is from 'Data World' - "Mental Health Depression Disorder Data" [1] containing 6469 rows x 10 attributes and the second one is from 'Kaggle' - "Mental Health" [2] containing 6421 rows x 8 attributes. The two datasets share both the country and year columns and so we merged the two datasets on these fields. For example, each dataset has entries for Afghanistan 1990 with the prevalence of mental health disorders and so on for every country and year. By merging the two datasets, we brought together all the mental health data available for each country yearly that was offered by the two datasets individually. We merged these two datasets using the Python pandas library and named it as "Psychological Health and Stability". Our newly created unique dataset contains 5377

rows x 15 attributes. It includes two different values for the prevalence of each disorder in every country around the world for each year from 1990 – 2017, with the prevalence being measured in (%) and (share of population) - Sex: Both - Age: Age-standardized. Furthermore, the dataset includes numerous key features that allow for an in-depth look at global mental health trends. For one, the dataset includes global coverage with country specific data. The broad reach of the dataset provides an international outlook on mental health. We've obtained wider coverage by combining datasets from two platforms, making it possible to analyze mental health trends more inclusively across various demographics and geographic areas. Another key feature is its longitudinal analysis over nearly three decades. Our dataset, which spans the years 1990 to 2017, makes it possible to do a longitudinal study and identify long-term trends and patterns in the prevalence of mental health issues. This broad chronological coverage is essential to comprehending the historical development of mental health diseases. A third key feature is the multiple dimensions of mental health disorders provided. A wide range of characteristics are included in the dataset, including anxiety, bipolar disorders, eating disorders, schizophrenia, and depressive disorders. With the help of this multi-dimensional approach, many mental health problems may be explored in detail, offering insights into both prevalence and demographic-standardized rates. The dataset's dual measurement units for prevalence offer another unique feature to the dataset. It supports various analytical purposes by including both percentage values and the population share for each mental health issue. This then allows for us to either concentrate on the relative proportions within populations or the percentage prevalence of illnesses. One final unique feature of the dataset is the age and gender standardization. The dataset's analytical robustness is improved by the addition of gender- and age-standardized data (Sex: Both - Age: Age-standardized). This feature makes it possible to compare the prevalence of mental health issues more accurately across various demographic groups, which aids in seeing trends and variances that may be missed by using raw prevalence data.

Entity	Code	Year	Schizophrenia disorders (share of population) - Sex: Both - Age: Age-standardized	Depressive disorders (share of population) - Sex: Both - Age: Age-standardized	Anxiety disorders (share of population) - Sex: Both - Age: Age-standardized	Bipolar disorders (share of population) - Sex: Both - Age: Age-standardized	Eating disorders (share of population) - Sex: Both - Age: Age-standardized	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Depression (%)	Alcohol use disorders (%)
Afghanistan	AFG	1990	0.222006	4.986118	4.713214	0.702023	0.127750	0.160560	0.697779	0.101855	4.826830	1.677982	4.071831	0.672040
Afghanistan	AFG	1991	0.222454	4.988090	4.702100	0.702069	0.123256	0.160312	0.697961	0.099313	4.809740	1.684746	4.079351	0.671768
Afghanistan	AFG	1992	0.221751	4.981346	4.683743	0.700792	0.118844	0.160135	0.698107	0.096690	4.831108	1.694334	4.088358	0.670544
Afghanistan	AFG	1993	0.220987	4.976958	4.673549	0.700087	0.115089	0.160037	0.698257	0.094336	4.800864	1.703520	4.096190	0.669738
Afghanistan	AFG	1994	0.220183	4.977782	4.670810	0.699898	0.111815	0.160022	0.698469	0.092439	4.826423	1.716069	4.099582	0.669260

Fig. 1. Visualization of Unique Dataset

RangeIndex: 5376 entries, 0 to 5375
Data columns (total 15 columns):
Column
0 Entity
1 Code
2 Year
3 Schizophrenia disorders (share of population) - Sex: Both - Age: Age-standardized
4 Depressive disorders (share of population) - Sex: Both - Age: Age-standardized
5 Anxiety disorders (share of population) - Sex: Both - Age: Age-standardized
6 Bipolar disorders (share of population) - Sex: Both - Age: Age-standardized
7 Eating disorders (share of population) - Sex: Both - Age: Age-standardized
8 Schizophrenia (%)
9 Bipolar disorder (%)
10 Eating disorders (%)
11 Anxiety disorders (%)
12 Drug use disorders (%)
13 Depression (%)
14 Alcohol use disorders (%)
dtypes: float64(12), int64(1), object(2)

Non-Null Count Dtype
5376 non-null object
5376 non-null object
5376 non-null int64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64
5376 non-null float64

Fig. 2. Information of Data Type and Size of the Dataset

III. EXPLORATORY DATA ANALYSIS (EDA)

This section is all about Exploratory Data Analysis and Visualization. Several Hypotheses have been incorporated here for both feasible and effective analysis.

Fig 1 comes from the .head() function which displays the dataset with all the column names and the first four rows of the dataset. This helps to illustrate the data we will be performing Exploratory Data Analysis on.

Fig 2 comes from the .info() function which helps to show the count of values for each column as well as the datatype. It displays that each column has 5376 non-null values, as well as datatypes of object, integer, or float.

Fig 3 comes from the .describe() function which displays descriptive statistics about the dataset. It provides mean, standard deviation, minimum, and maximum values, as well as other statistics to better understand the dataset.

Fig 4 is a line graph with a collection of lines representing all the various disorders in our dataset. It has combined the prevalence of each disorder from all the countries in the world for each year from 1990 – 2017 and plotted each point on the graph, connected with a line indicating the disorder it represents. Each line is mostly horizontal across the entire line graph, indicating that the overall presence of each of these disorders globally has stayed relatively consistent over the years. This makes sense when considering the global population as a whole, as it would take significant change to alter the prevalence of any one of these disorders on a global scale.

	Schizophrenia disorders (share of population) - Sex: Both - Age: Age-standardized	Depressive disorders (share of population) - Sex: Both - Age: Age-standardized	Anxiety disorders (share of population) - Sex: Both - Age: Age-standardized	Bipolar disorders (share of population) - Sex: Both - Age: Age-standardized	Eating disorders (share of population) - Sex: Both - Age: Age-standardized	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Depression (%)	Alcohol use disorders (%)
count	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000	5376.000000
mean	0.220000	4.986118	4.713214	0.702023	0.127750	0.160560	0.697779	0.101855	4.826830	1.677982	4.071831	0.672040
std	0.078499	0.099590	0.098434	0.061038	0.023025	0.133282	0.042255	0.164996	0.155028	1.143433	0.444650	0.671746
min	1990.000000	0.188446	1.322333	1.878996	0.181667	0.044780	0.148902	0.314535	0.073908	0.203393	0.368360	2.133903
25%	1996.700000	0.231827	3.057499	3.411289	0.522996	0.094882	0.179452	0.616315	0.121760	3.174200	0.524018	2.960080
50%	2003.000000	0.271644	3.678267	3.932847	0.582961	0.145709	0.198647	0.693954	0.180784	3.548896	0.709000	3.477581
75%	2010.250000	0.284888	4.418555	4.561693	0.867490	0.244037	0.232323	0.834689	0.283283	4.665994	0.908891	3.882559
max	2017.000000	0.462045	7.648889	8.624034	1.507750	0.254046	0.375110	1.206597	0.943991	8.967330	3.452476	6.602754

Fig. 3. Statistical Description of Dataset

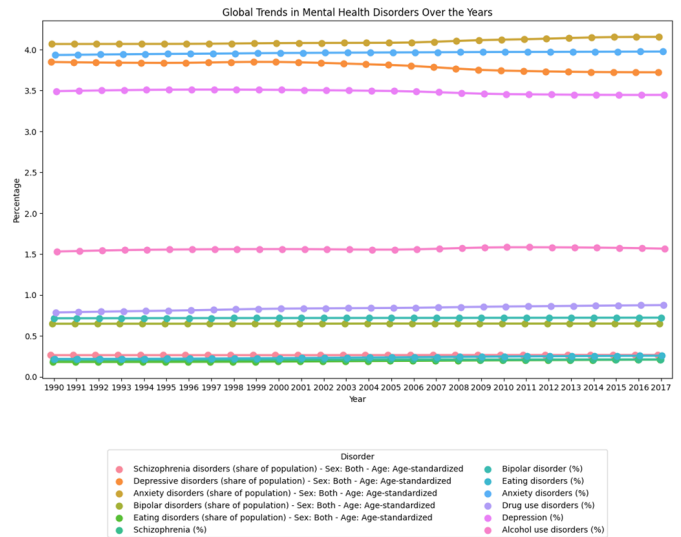


Fig. 4. Global Trends in Mental Health Disorders over the years

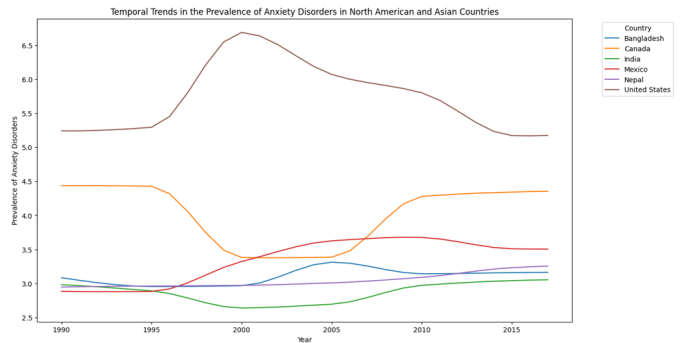


Fig. 5. Temporal Trends in Anxiety Disorders in North America and Asia

Fig 5 shows the temporal trends in the prevalence of anxiety disorders in three North American countries and three Asian countries. The North American countries consistently have much higher rates of anxiety over the years when compared to the Asian countries except for 1990 – 1995 where Mexico dips below the three Asian countries. The United States appears to have the highest rates of anxiety, with India having the lowest rates from 1995 – 2017. There is also much more variation in the prevalence of anxiety in the North American countries over the years, as opposed to the Asian countries which appear to have much more stable rates of anxiety.

Fig 6 plots the correlation coefficients between schizophrenia and bipolar disorder rates for every country across the world. As illustrated by the scatterplot, as well as the correlation coefficient of 0.23, there does not appear to be a clear correlation globally between schizophrenia and bipolar disorder. In individual countries, however, there are both positive and negative correlations. This means that for some countries as the rate of schizophrenia increases so does the rate of bipolar disorder, and vice versa for other countries.

Fig 7 and Fig 8, two graphics show the correlation between anxiety and eating disorders by year for the two columns



Fig. 6. Correlation Between Schizophrenia and Bipolar Disorder in %

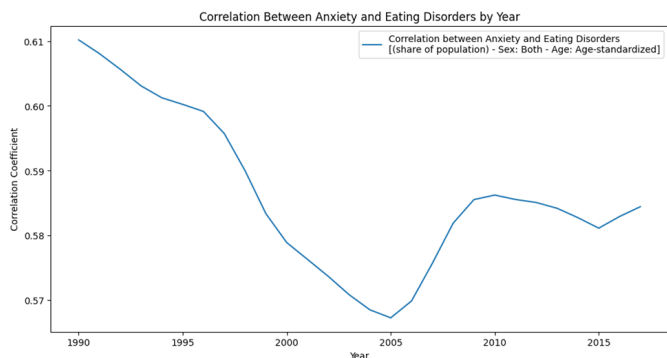


Fig. 7. Correlation Between Anxiety and Eating Disorders Sex: Both - Age: Age-standardized

(share of population) - Sex: Both - Age: Age-standardized and (%), respectively. Both line graphs are slightly different although they show the same general trends over time. Additionally, there is a high positive correlation between anxiety and eating disorders throughout. There is a decrease in correlation between anxiety and eating disorders globally from 1990 – 2005, and then an increase from 2005 – 2010. After this, there is another decrease in the correlation from 2010 – 2015 with the correlation increasing again after 2015. A large difference although is the global minima for the (%)

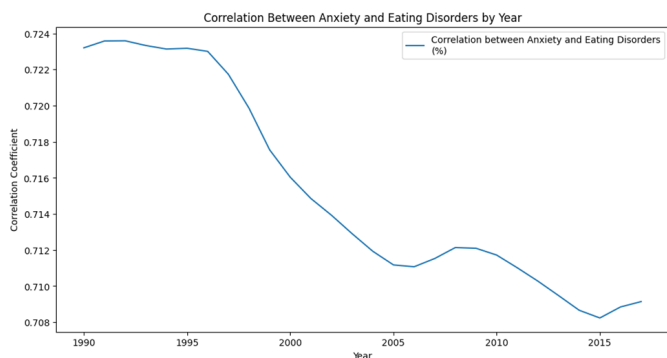


Fig. 8. Correlation Anxiety and Eating Disorders in %

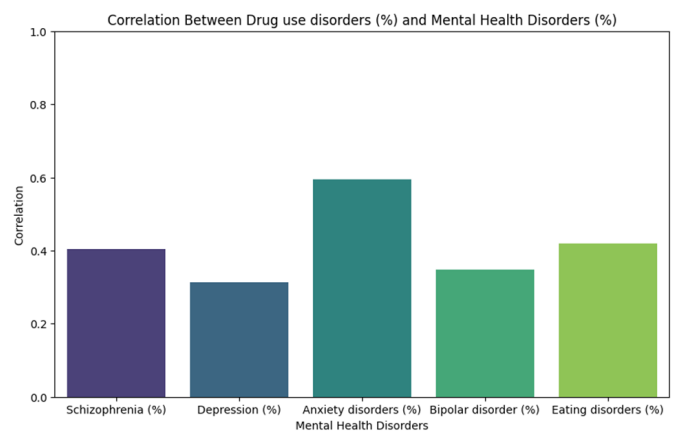


Fig. 9. Correlation Between Drug Use and Mental Health Disorders in %

graph is in 2015 whereas it is in 2005 for the other graph. This shows the extreme similarities between the two different columns and therefore datasets but also how there is some slight variation.

Fig 9 shows the correlations between drug use disorders and various mental health disorders in a histogram. Each bar represents the correlation between that specific mental health disorder and drug use disorder. Anxiety is the mental health disorder with the highest correlation with drug use disorder, with depression having the lowest correlation. However, all of the disorders appear to be within the correlation range of 0.3 – 0.6 so there are no extreme outliers, and all seem to have a moderate positive correlation to drug use disorder.

IV. CONCLUSION

In conclusion, our thorough exploratory data analysis (EDA) of the "Psychological Health and Stability" dataset, which covered the years 1990 to 2017, has provided insightful information about trends in mental health around the world. We generated a unique dataset by combining two separate datasets from "Data World" and "Kaggle," which not only gives an international perspective on mental health but also delivers a thorough and in-depth analysis of various mental health issues. By employing line graphs, histograms, and scatterplots to depict the changing state of mental health, we were able to identify temporal trends and possible relationships between illnesses using Python-based EDA tools. In addition to concentrating on prevalence, our research also examined demographic-standardized rates, highlighting the dual measuring units and age and gender standardization within the dataset. This thorough exploration has contributed to a deeper understanding of the distribution and interconnections among mental health conditions globally. This dataset and EDA may serve as a useful tool for academics, decision-makers, and medical professionals who want to address and lessen the effects of mental health disorders globally, particularly as society continues to struggle with mental health issues.

V. APPENDIX A

Discussed in Jupyter Notebook file.

VI. APPENDIX B

A. *Galib Md Azraf Nijhum*

In this project, I actively contributed to the discovery and acquisition of the 'Mental Health' dataset available on Kaggle. I curated Appendix A, merging two datasets into a cohesive database, and compiled Appendix C, cataloging all project references. My role extended to ensuring the report's coherence and professionalism by using LaTeX for formatting after merging datasets based on shared features. During Exploratory Data Analysis (EDA), I conducted comprehensive statistical analyses, created insightful histogram plots, and led the EDA process. Presenting the EDA findings was part of my responsibilities, having previously prepared the project proposal. Despite challenges in finding suitable datasets with shared features, we successfully merged datasets based on three crucial shared features. Our team's collaborative efforts from dataset discovery, merging, EDA, presentation creation, and report crafting provided invaluable insights, enriching our understanding of data analysis and positioning us favorably for careers in Data Science, equipped with essential knowledge and skills.

B. *Alexander Richard Cramer*

For our group project, I contributed to the Section 1 - Overview, Section 2 - Dataset, Section 3 - EDA, and Section 4 - Conclusion. This project helped me to learn the data science workflow, by finding datasets, merging and cleaning the data, and performing EDA as well as conducting some analysis and write-up regarding the data. Additionally, I learned numerous technical skills throughout such as matplotlib and seaborn formatting for graphics and how to aggregate data in meaningful ways. The most difficult aspect for me was grouping certain fields in order to create the data analysis and illustrations that I wished to. Despite these challenges, I learned valuable lessons from this experience, such as improving my problem-solving skills and understanding the importance of teamwork in data science. I also learned the importance of understanding the entire data science process and the need to carefully examine the dataset. This project was an enjoyable assignment as it challenged me intellectually and allowed me to learn from different perspectives in a collaborative environment.

C. *Riad Hossain*

To contribute to our group project, I identified a relevant dataset titled "Mental Health Depression Disorder Data" from "Data World". After merging two datasets, I conducted some of our exploratory data analysis (EDA) to thoroughly examine our unique dataset. I merged the Jupyter Notebook files with the other analyses done by my group members. During EDA, I presented some correlations, histograms, and assessed the normality of distribution for a specific column value. Employing Python libraries, I generated graphs and figures and documented the process in corresponding markdown. I had

an active contribution to make the presentation files for EDA and Creation of Dataset as well. In the report, I contributed to writing various sections, including the Overview, Dataset, EDA, and Conclusion. Throughout the project, collaboration with my group members fostered strong connections and friendships. Despite encountering some challenges, such as the time-consuming search for similar datasets and common attributes, the overall experience was rewarding. I am pleased with my contribution to the project, and the collaboration with my team members enriched the entire experience. Doing these, I learned a lot of analysis about datasets and creating a new unique dataset from different datasets and so on.

VII. APPENDIX C

REFERENCES

- [1] Data World, "Mental health depression disorder data," <https://data.world/vizzup/mental-health-depression-disorder-data/>, 2023, Accessed on December 4, 2023.
- [2] Mental Health, "Mental health," <https://data.world/vizzup/mental-health-depression-disorder-data/>, 2023, Accessed on December 4, 2023.