McKibben DSC520 Final Project 8.3

Makayla McKibben

2024-07-30

Introduction

Global awareness and tolerance of mental health conditions have been increasing rapidly in the last few years. However, a recent comprehensive analysis of the prevalence and most effective treatments of anxiety is not something I've come across. Mental illness and, specifically, anxiety is a topic that affects a great many people. Whether you are afflicted or have a spouse, relative, or friend who deals with mental illness, the impact of anxiety on a person's quality of life can touch nearly everyone. In order to determine the prevalence of anxiety and the effectiveness of various treatments, we will look at several datasets that provide recent, relevant information. R will be critical in analyzing this much data from these expansive datasets. While the World Health Organization published a study in 2019, and the NIH published a study using data from 2001 - 2004, I'd like to see how a more recent analysis will compare with their findings. This data and our inferences can be checked against the 2019 WHO analysis to see the progression of anxiety's pervasiveness and treatment options in the last five years and against the NIH study to see the change in the last 20 years.

Research questions

What are the most prevalent symptoms of anxiety? How many people experience symptoms of anxiety? Are specific demographics affected more heavily by anxiety? In the last five years, has the number of people affected by anxiety changed from the WHO study's predictions? In the last 20 years, has the number of people affected by anxiety changed from the NIH study? What medications are available to treat anxiety? Which medications appear to be most effective at treating anxiety? How many cases of anxiety are managed by the most effective medications?

Approach

The first part of the analysis will deal with datasets that involve the symptoms that indicate an anxiety disorder. We can establish trends from the number of people experiencing symptoms versus those diagnosed and possibly create theories about its pervasiveness; this will dovetail with the second part of our analysis. The second step will be to work with different global datasets to determine the trends of the prevalence of anxiety disorders and compare the data results between the first two analysis phases to see if our prediction about the number of people affected matches up. The last step will look at a dataset that has ratings of the effectiveness of psychiatric medications, which we will narrow down to anxiety. We can take the trends of prevalence and compare them to the effectiveness of various medications and make a supposition about whether or not certain medications seem more effective at treating anxiety. We can examine all of our datasets to see if specific demographics are particularly susceptible to experiencing anxiety and if certain medications are more effective for different demographic groups as well. We can use the data from the first two parts of our analysis and cross-correlate with the number of people prescribed effective medications to make a conjecture about how many cases of anxiety are well managed.

How your approach addresses (fully or partially) the problem.

The datasets I intend to use are more recent than those from the 2019 WHO study which used data from 2018 and 2019 and the 20-year-old data from the NIH study. I have selected six datasets, two from 2022 and the remaining four updated this year. Once I've analyzed all of this more recent data, I can compare this new information with the WHO and NIH studies and see if there have been unaccounted-for changes.

Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)

Global Mental Health Disorders

Kaggle https://www.kaggle.com/datasets/thedevastator/global-mental-health-disorders "This dataset contains valuable information about the prevalence of mental health disorders including schizophrenia, bipolar disorder, eating disorders, anxiety disorders, drug use disorders, depression, and alcohol use disorders from various countries across the globe." 108554 rows, 11 columns Updated 2022 Missing values left blank

Gender Mental Disorder Prevalence

Kaggle https://www.kaggle.com/datasets/thedevastator/gender-mental-disorder-prevalence-2019 "This dataset provides the gender-based prevalence of mental health disorders around the world in 2019." 56396 rows, 8 columns Updated 2022 Reisha Hermana Missing values left blank

Mental Health Data (Anxiety)

Kaggle https://www.kaggle.com/datasets/michellevp/predicting-anxiety-in-mental-health-data "This dataset appears to contain a variety of features related to text analysis, sentiment analysis, and psychological indicators, likely derived from posts or text data. Additionally, there are features related to psychological aspects such as economic stress, isolation, substance use, and domestic stress. The dataset seems to cover a wide range of linguistic, psychological, and behavioral attributes, potentially suitable for analyzing mental health-related topics in online communities or text data." 1968 rows, 350 columns Updated 2024 Collected 2018-2019 Missing values blank or zero

Indicators of Anxiety or Depression

Kaggle https://www.kaggle.com/datasets/melissamonfared/indicators-of-anxiety-or-depression "This dataset contains information on the indicators of anxiety or depression based on the reported frequency of symptoms during the last 7 days. The data is collected through the Household Pulse Survey, launched by the U.S. Census Bureau in collaboration with five federal agencies." 16093 rows, 14 columns Updated 2024 Collected 2020-2024 No missing values

Mental Health Dataset

Kaggle https://www.kaggle.com/datasets/divaniazzahra/mental-health-dataset "This dataset records a global survey conducted to track trends in mental health. The data covers a range of variables such as levels of stress, depression, anxiety, subjective well-being, and use of mental health services. The survey involved respondents from various demographic backgrounds, including gender, employment status, and geographic region." 292365 rows, 17 columns Updated 2024 Collected 2014-2016 No missing values

WebMD Reviews for Psychiatric Drugs

Kaggle https://www.kaggle.com/datasets/sepidehparhami/psychiatric-drug-webmd-reviews "This dataset consists of unstructured text reviews, categorical ratings, and demographics from patients and caregivers of patients on various psychiatric drugs. The current version of the dataset contains over 61,000 reviews for hundreds of medications used to treat psychiatric disorders." 61321 rows, 13 columns Updated 2024 Collected 2007-2024 Missing data cell is left blank

Anxiety disorders

World Health Organization https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders

Any Anxiety Disorder

NIH https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder

Required Packages

At least: purrr tidyverse Metrics ggplot2

Plots and Table Needs

We will need quite a few plots and tables. At least: Number of people experiencing anxiety symptoms globally % globally by country % by country by demographic % by demographic Number of people diagnosed with anxiety globally % globally by country % by country by demographic % by demographic The most prevalent anxiety symptoms globally % globally by country % by country by demographic % by demographic Most prevalent medications globally % globally by country % by country by demographic % by demographic Ratings of medications Number of people taking specified medication The number of cases we predict will be well-managed

Questions for future steps

Do we anticipate the trends in prevalence we've found to continue over the next five years? The following 10, 20? Do we anticipate the number of people with anxiety that is well-managed to change? How? Are there newer medications that we do not have data on? Do we expect the representation of those diagnosed in specific demographics to change? How?

What do you not know how to do right now that you need to learn to answer your research questions?

I think I have all the resources and information I need to complete this project; it will just take time.