

McKibben DSC520 Final Project 9.3

Makayla McKibben

2024-08-03

Week 8 Ex. 8.3

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.

Week 9 Ex. 9.3

Code for Importing and Cleaning the Data

```
# Import necessary packages
#install.packages("tidyverse")
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.1
```

```
## Warning: package 'ggplot2' was built under R version 4.4.1
```

```
## Warning: package 'tibble' was built under R version 4.4.1
```

```
## Warning: package 'tidyr' was built under R version 4.4.1
```

```
## Warning: package 'readr' was built under R version 4.4.1
```

```
## Warning: package 'purrr' was built under R version 4.4.1

## Warning: package 'dplyr' was built under R version 4.4.1

## Warning: package 'forcats' was built under R version 4.4.1

## Warning: package 'lubridate' was built under R version 4.4.1

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Import all data files
healthanxiety <-
  read.csv(file = 'healthanxiety_dataset.csv',
           header = TRUE, sep = ",", stringsAsFactors = FALSE)

indicators_anxiety <-
  read.csv(file = 'Indicators_of_Anxiety_or_Depression.csv',
           header = TRUE, sep = ",", stringsAsFactors = FALSE)

global_mental <-
  read.csv(file = 'Mental Health Data Global.csv',
           header = TRUE, sep = ",", stringsAsFactors = FALSE)

mental <-
  read.csv(file = 'Mental Health Dataset.csv',
           header = TRUE, sep = ",", stringsAsFactors = FALSE)

prevalence <-
  read.csv(file = 'prevalence-of-anxiety-disorders-males-vs-females.csv',
           header = TRUE, sep = ",", stringsAsFactors = FALSE)

meds <-
  read.csv(file = 'psychiatric_drug_webmd_reviews.csv',
           header = TRUE, sep = ",", stringsAsFactors = FALSE)

# Check out the data
# First dataset
head(healthanxiety)
```

```
##      subreddit      author      date
## 1 healthanxiety    Nomis176 1/1/2018
## 2 healthanxiety    psychstudent317 1/1/2018
```

```

## 3 healthanxiety      bulk_barn 1/1/2018
## 4 healthanxiety      AutoModerator 1/1/2018
## 5 healthanxiety      parthkhurana7 1/1/2018
## 6 healthanxiety      anxietyislife22 1/1/2018
##
## 1 Final doctor appointment tomorrow, tired of constantly seeking reassurance! Ive decided that Im ha
## 2
## 3
## 4
## 5
## 6
##      automated_readability_index coleman_liau_index flesch_kincaid_grade_level
## 1              9.137909              7.282428              8.616061
## 2              4.399892              7.131386              3.665806
## 3              3.590660              4.930895              4.605802
## 4              4.515682              7.033342              5.059091
## 5              3.034388              5.356252              3.977494
## 6              1.114750              3.142913              2.765000
##      flesch_reading_ease gulpease_index gunning_fog_index      lix smog_index
## 1          72.47409          61.09091          11.345455 39.57576  9.888513
## 2          87.18538          74.16129          5.423656 23.23656  6.427356
## 3          85.64097          72.58491          7.564151 22.68396  8.076483
## 4          78.38364          72.86364          8.945455 26.90909  9.516145
## 5          85.30696          76.72152          7.757637 26.35612  8.648137
## 6          93.05000          81.75000          5.500000 21.25000  6.627428
##      wiener_sachtextformel n_chars n_long_words n_monosyllable_words
## 1          3.8370667      1371          58          250
## 2          0.7243462       136           4           23
## 3          1.0987962       414          10           83
## 4          2.8111318       382          14           65
## 5          2.0245278       644          25          127
## 6          0.3248750       298           9           65
##      n_polysyllable_words n_sents n_syllables n_unique_words n_words sent_neg
## 1              21          15          437          172      330      0.130
## 2              1           3           40           24       31      0.281
## 3              6           8          135           68      106      0.091
## 4             10           8          122           64       88      0.138
## 5             14          15          207           98      158      0.131
## 6              3           8           98           52       80      0.222
##      sent_neu sent_pos sent_compound economic_stress_total isolation_total
## 1      0.749      0.121      -0.3075              0              0
## 2      0.719      0.000      -0.8677              0              0
## 3      0.761      0.148       0.7464              0              0
## 4      0.763      0.098      -0.5824              0              0
## 5      0.781      0.088      -0.7096              0              0
## 6      0.639      0.139      -0.7650              0              0
##      substance_use_total guns_total domestic_stress_total suicidality_total
## 1              3              0              0              0
## 2              0              0              0              0
## 3              2              0              0              0
## 4              0              0              0              0
## 5              0              0              0              0
## 6              0              0              0              0
##      punctuation liwc_1st_pers liwc_2nd_pers liwc_3rd_pers liwc_achievement

```

| | | | | | | |
|------|-----------------------|--------------------------|--------------------------|-------------------|-------------------|-----------|
| ## 1 | 33 | 0 | 1 | 0 | 2 | |
| ## 2 | 4 | 0 | 0 | 0 | 0 | |
| ## 3 | 12 | 2 | 5 | 0 | 2 | |
| ## 4 | 14 | 0 | 3 | 0 | 3 | |
| ## 5 | 25 | 0 | 0 | 0 | 1 | |
| ## 6 | 8 | 0 | 0 | 0 | 1 | |
| ## | liwc_adverbs | liwc_affective_processes | liwc_anger | liwc_anxiety | | |
| ## 1 | 12 | | 16 | 1 | 3 | |
| ## 2 | 0 | | 2 | 0 | 0 | |
| ## 3 | 8 | | 7 | 1 | 0 | |
| ## 4 | 3 | | 7 | 0 | 2 | |
| ## 5 | 5 | | 9 | 2 | 1 | |
| ## 6 | 3 | | 3 | 0 | 0 | |
| ## | liwc_articles_article | liwc_assent | liwc_auxiliary_verbs | liwc_biological | | |
| ## 1 | | 16 | 0 | 31 | 21 | |
| ## 2 | | 0 | 0 | 6 | 3 | |
| ## 3 | | 3 | 0 | 10 | 8 | |
| ## 4 | | 5 | 0 | 6 | 4 | |
| ## 5 | | 11 | 0 | 18 | 17 | |
| ## 6 | | 7 | 2 | 12 | 2 | |
| ## | liwc_body | liwc_causation | liwc_certainty | liwc_cognitive | liwc_common_verbs | |
| ## 1 | 4 | 5 | 2 | 53 | 53 | |
| ## 2 | 1 | 1 | 0 | 4 | 6 | |
| ## 3 | 4 | 2 | 0 | 15 | 23 | |
| ## 4 | 0 | 1 | 1 | 10 | 11 | |
| ## 5 | 7 | 0 | 1 | 18 | 27 | |
| ## 6 | 1 | 0 | 0 | 18 | 19 | |
| ## | liwc_conjunctions | liwc_death | liwc_discrepancy | liwc_exclusive | liwc_family | |
| ## 1 | | 28 | 0 | 4 | 2 | 0 |
| ## 2 | | 2 | 0 | 0 | 1 | 0 |
| ## 3 | | 9 | 0 | 1 | 3 | 0 |
| ## 4 | | 6 | 0 | 1 | 4 | 0 |
| ## 5 | | 14 | 1 | 0 | 7 | 0 |
| ## 6 | | 5 | 0 | 1 | 6 | 0 |
| ## | liwc_feel | liwc_fillers | liwc_friends | liwc_future_tense | liwc_health | liwc_hear |
| ## 1 | 4 | 0 | 0 | 5 | 17 | 1 |
| ## 2 | 0 | 0 | 0 | 0 | 2 | 0 |
| ## 3 | 0 | 0 | 0 | 0 | 5 | 4 |
| ## 4 | 0 | 0 | 0 | 0 | 4 | 0 |
| ## 5 | 0 | 0 | 0 | 0 | 10 | 0 |
| ## 6 | 5 | 1 | 0 | 0 | 1 | 0 |
| ## | liwc_home | liwc_humans | liwc_impersonal_pronouns | liwc_inclusive | liwc_ingestion | |
| ## 1 | 0 | 2 | | 26 | 22 | 0 |
| ## 2 | 0 | 0 | | 4 | 2 | 0 |
| ## 3 | 0 | 0 | | 7 | 3 | 0 |
| ## 4 | 0 | 0 | | 6 | 3 | 0 |
| ## 5 | 0 | 0 | | 10 | 4 | 0 |
| ## 6 | 0 | 0 | | 5 | 4 | 0 |
| ## | liwc_inhibition | liwc_insight | liwc_leisure | liwc_money | liwc_motion | |
| ## 1 | | 0 | 12 | 1 | 0 | 4 |
| ## 2 | | 0 | 1 | 0 | 0 | 0 |
| ## 3 | | 0 | 5 | 0 | 0 | 0 |
| ## 4 | | 0 | 1 | 3 | 0 | 0 |
| ## 5 | | 0 | 1 | 0 | 0 | 2 |

| | | | | | | |
|------|-----------------------|---------------------------|------------------------|------------------|-------------|---------------|
| ## 6 | 0 | 5 | 0 | 0 | 2 | |
| ## | liwc_negations | liwc_negative_emotion | liwc_nonfluencies | liwc_numbers | | |
| ## 1 | 2 | 8 | 0 | 0 | | |
| ## 2 | 0 | 2 | 0 | 0 | | |
| ## 3 | 1 | 3 | 0 | 0 | | |
| ## 4 | 0 | 3 | 0 | 0 | | |
| ## 5 | 3 | 5 | 0 | 1 | | |
| ## 6 | 1 | 0 | 0 | 2 | | |
| ## | liwc_past_tense | liwc_perceptual_processes | liwc_personal_pronouns | | | |
| ## 1 | 8 | 8 | 33 | | | |
| ## 2 | 2 | 0 | 3 | | | |
| ## 3 | 0 | 4 | 12 | | | |
| ## 4 | 1 | 0 | 3 | | | |
| ## 5 | 16 | 1 | 11 | | | |
| ## 6 | 2 | 5 | 4 | | | |
| ## | liwc_positive_emotion | liwc_prepositions | liwc_present_tense | liwc_quantifiers | | |
| ## 1 | 8 | 39 | 39 | 4 | | |
| ## 2 | 0 | 1 | 4 | 0 | | |
| ## 3 | 4 | 11 | 23 | 1 | | |
| ## 4 | 4 | 12 | 8 | 1 | | |
| ## 5 | 4 | 9 | 11 | 4 | | |
| ## 6 | 3 | 11 | 14 | 1 | | |
| ## | liwc_relativity | liwc_religion | liwc_sadness | liwc_see | liwc_sexual | |
| ## 1 | 49 | 0 | 1 | 2 | 0 | |
| ## 2 | 0 | 0 | 0 | 0 | 0 | |
| ## 3 | 10 | 0 | 0 | 0 | 0 | |
| ## 4 | 14 | 0 | 0 | 0 | 0 | |
| ## 5 | 23 | 0 | 1 | 1 | 0 | |
| ## 6 | 9 | 0 | 0 | 0 | 0 | |
| ## | liwc_social_processes | liwc_space | liwc_swear_words | liwc_tentative | liwc_time | |
| ## 1 | 9 | 19 | 0 | 7 | 23 | |
| ## 2 | 1 | 0 | 0 | 1 | 0 | |
| ## 3 | 13 | 3 | 0 | 2 | 6 | |
| ## 4 | 6 | 3 | 0 | 2 | 10 | |
| ## 5 | 2 | 9 | 0 | 5 | 12 | |
| ## 6 | 1 | 2 | 0 | 2 | 6 | |
| ## | liwc_total_functional | liwc_total_pronouns | liwc_work | tfidf_abl | tfidf_abus | |
| ## 1 | 172 | 59 | 5 | 0 | 0 | |
| ## 2 | 15 | 7 | 0 | 0 | 0 | |
| ## 3 | 54 | 19 | 3 | 0 | 0 | |
| ## 4 | 41 | 9 | 3 | 0 | 0 | |
| ## 5 | 83 | 21 | 3 | 0 | 0 | |
| ## 6 | 50 | 9 | 0 | 0 | 0 | |
| ## | tfidf_actual | tfidf_addict | tfidf_adhd | tfidf_advic | tfidf_ago | tfidf_alcohol |
| ## 1 | 0 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 2 | 0 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 3 | 0 | 0 | 0 | 0.1486653 | 0 | 0 |
| ## 4 | 0 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 5 | 0 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 6 | 0 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## | tfidf_almost | tfidf_alon | tfidf_alreadi | tfidf_also | tfidf_alway | tfidf_amp |
| ## 1 | 0.0000000 | 0 | 0 | 0 | 0 | 0 |
| ## 2 | 0.0000000 | 0 | 0 | 0 | 0 | 0 |
| ## 3 | 0.0000000 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | |
|------|-----------------|---------------|-----------------|---------------|---------------|-------------|-------------|
| ## 4 | 0.0000000 | 0 | 0 | 0 | 0 | 0 | |
| ## 5 | 0.0000000 | 0 | 0 | 0 | 0 | 0 | |
| ## 6 | 0.2300638 | 0 | 0 | 0 | 0 | 0 | |
| ## | tfidf_amp.x200b | tfidf_ani | tfidf_anoth | tfidf_anxieti | tfidf_anxious | | |
| ## 1 | | 0 0.0000000 | 0 | 0.1328585 | 0.1961997 | | |
| ## 2 | | 0 0.0000000 | 0 | 0.2188037 | 0.0000000 | | |
| ## 3 | | 0 0.0000000 | 0 | 0.2327669 | 0.0000000 | | |
| ## 4 | | 0 0.1309625 | 0 | 0.2946498 | 0.0000000 | | |
| ## 5 | | 0 0.1109275 | 0 | 0.1247866 | 0.0000000 | | |
| ## 6 | | 0 0.0000000 | 0 | 0.3515198 | 0.0000000 | | |
| ## | tfidf_anymor | tfidf_anyon | tfidf_anyon.els | tfidf_anyth | tfidf_around | tfidf_ask | |
| ## 1 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## 2 | 0 | 0.3999822 | 0.2930602 | 0 | 0 | 0 | |
| ## 3 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## 4 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## 5 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## 6 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## | tfidf_attack | tfidf_away | tfidf_back | tfidf_bad | tfidf_becaus | tfidf_becom | |
| ## 1 | 0 | 0.08158519 | 0.06677145 | 0 | 0.05293436 | 0 | |
| ## 2 | 0 | 0.00000000 | 0.00000000 | 0 | 0.00000000 | 0 | |
| ## 3 | 0 | 0.00000000 | 0.00000000 | 0 | 0.00000000 | 0 | |
| ## 4 | 0 | 0.00000000 | 0.00000000 | 0 | 0.00000000 | 0 | |
| ## 5 | 0 | 0.15325697 | 0.00000000 | 0 | 0.00000000 | 0 | |
| ## 6 | 0 | 0.00000000 | 0.00000000 | 0 | 0.00000000 | 0 | |
| ## | tfidf_befor | tfidf_believ | tfidf_best | tfidf_better | tfidf_bit | tfidf_bodi | |
| ## 1 | 0.07389096 | 0 | 0.0000000 | 0.0767993 | 0 | 0.0000000 | |
| ## 2 | 0.00000000 | 0 | 0.0000000 | 0.0000000 | 0 | 0.0000000 | |
| ## 3 | 0.00000000 | 0 | 0.1596571 | 0.0000000 | 0 | 0.5069656 | |
| ## 4 | 0.00000000 | 0 | 0.0000000 | 0.0000000 | 0 | 0.0000000 | |
| ## 5 | 0.00000000 | 0 | 0.0000000 | 0.0000000 | 0 | 0.0000000 | |
| ## 6 | 0.00000000 | 0 | 0.0000000 | 0.0000000 | 0 | 0.0000000 | |
| ## | tfidf_bpd | tfidf_brain | tfidf_call | tfidf_came | tfidf_care | tfidf_caus | tfidf_chang |
| ## 1 | 0 | 0.09693759 | 0 | 0 | 0 | 0 | 0 |
| ## 2 | 0 | 0.00000000 | 0 | 0 | 0 | 0 | 0 |
| ## 3 | 0 | 0.00000000 | 0 | 0 | 0 | 0 | 0 |
| ## 4 | 0 | 0.00000000 | 0 | 0 | 0 | 0 | 0 |
| ## 5 | 0 | 0.18209629 | 0 | 0 | 0 | 0 | 0 |
| ## 6 | 0 | 0.25647959 | 0 | 0 | 0 | 0 | 0 |
| ## | tfidf_come | tfidf_complet | tfidf_constant | tfidf_control | tfidf_could | tfidf_coupl | |
| ## 1 | 0.1496029 | 0.09104579 | 0.09383877 | 0 | 0 | 0 | |
| ## 2 | 0.0000000 | 0.00000000 | 0.00000000 | 0 | 0 | 0 | |
| ## 3 | 0.0000000 | 0.00000000 | 0.00000000 | 0 | 0 | 0 | |
| ## 4 | 0.0000000 | 0.00000000 | 0.00000000 | 0 | 0 | 0 | |
| ## 5 | 0.0000000 | 0.00000000 | 0.00000000 | 0 | 0 | 0 | |
| ## 6 | 0.1979113 | 0.00000000 | 0.24828065 | 0 | 0 | 0 | |
| ## | tfidf_cri | tfidf_day | tfidf_deal | tfidf_depress | tfidf_diagnos | tfidf_die | |
| ## 1 | 0 | 0.0000000 | 0.08952401 | 0 | 0 | 0 | |
| ## 2 | 0 | 0.0000000 | 0.29487297 | 0 | 0 | 0 | |
| ## 3 | 0 | 0.0000000 | 0.00000000 | 0 | 0 | 0 | |
| ## 4 | 0 | 0.0000000 | 0.00000000 | 0 | 0 | 0 | |
| ## 5 | 0 | 0.1051992 | 0.00000000 | 0 | 0 | 0 | |
| ## 6 | 0 | 0.0000000 | 0.00000000 | 0 | 0 | 0 | |
| ## | tfidf_differ | tfidf_disord | tfidf_doctor | tfidf_doe | tfidf_done | tfidf_dont | |
| ## 1 | 0 | 0.0000000 | 0.3555209 | 0 | 0 | 0 | |

| | | | | | | | |
|------|--------------|--------------|--------------|-------------|-----------------|---------------|--------------|
| ## 2 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## 3 | 0 | 0.0000000 | 0.3114348 | 0 | 0 | 0 | |
| ## 4 | 0 | 0.2207159 | 0.0000000 | 0 | 0 | 0 | |
| ## 5 | 0 | 0.0000000 | 0.3339212 | 0 | 0 | 0 | |
| ## 6 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## | tfidf_drink | tfidf_drug | tfidf_eat | tfidf_els | tfidf_emot | tfidf_end | tfidf_enough |
| ## 1 | 0 | 0 | 0 | 0.0000000 | 0 | 0 | 0 |
| ## 2 | 0 | 0 | 0 | 0.2389321 | 0 | 0 | 0 |
| ## 3 | 0 | 0 | 0 | 0.0000000 | 0 | 0 | 0 |
| ## 4 | 0 | 0 | 0 | 0.0000000 | 0 | 0 | 0 |
| ## 5 | 0 | 0 | 0 | 0.0000000 | 0 | 0 | 0 |
| ## 6 | 0 | 0 | 0 | 0.0000000 | 0 | 0 | 0 |
| ## | tfidf_etc | tfidf_even | tfidf_ever | tfidf_everi | tfidf_everyon | tfidf_everyth | |
| ## 1 | 0 | 0.0573543 | 0 | 0.0000000 | 0 | 0 | |
| ## 2 | 0 | 0.0000000 | 0 | 0.0000000 | 0 | 0 | |
| ## 3 | 0 | 0.0000000 | 0 | 0.1281809 | 0 | 0 | |
| ## 4 | 0 | 0.0000000 | 0 | 0.0000000 | 0 | 0 | |
| ## 5 | 0 | 0.1077395 | 0 | 0.0000000 | 0 | 0 | |
| ## 6 | 0 | 0.0000000 | 0 | 0.0000000 | 0 | 0 | |
| ## | tfidf_experi | tfidf_famili | tfidf_fear | tfidf_feel | tfidf_feel.like | tfidf_felt | |
| ## 1 | 0 | 0 | 0.09771454 | 0.1756274 | 0 | 0 | |
| ## 2 | 0 | 0 | 0.00000000 | 0.0000000 | 0 | 0 | |
| ## 3 | 0 | 0 | 0.00000000 | 0.0000000 | 0 | 0 | |
| ## 4 | 0 | 0 | 0.21670855 | 0.0000000 | 0 | 0 | |
| ## 5 | 0 | 0 | 0.00000000 | 0.0000000 | 0 | 0 | |
| ## 6 | 0 | 0 | 0.00000000 | 0.4646787 | 0 | 0 | |
| ## | tfidf_final | tfidf_find | tfidf_first | tfidf_food | tfidf_found | tfidf_friend | |
| ## 1 | 0.2853735 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 2 | 0.0000000 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 3 | 0.0000000 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 4 | 0.2109644 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 5 | 0.0000000 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 6 | 0.0000000 | 0 | 0.1954272 | 0 | 0 | 0 | |
| ## | tfidf_fuck | tfidf_get | tfidf_give | tfidf_go | tfidf_good | tfidf_got | tfidf_great |
| ## 1 | 0 | 0.04536818 | 0 | 0.09870169 | 0.07283384 | 0.06945061 | 0 |
| ## 2 | 0 | 0.00000000 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0 |
| ## 3 | 0 | 0.00000000 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0 |
| ## 4 | 0 | 0.10061626 | 0 | 0.00000000 | 0.00000000 | 0.15402560 | 0 |
| ## 5 | 0 | 0.08522367 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0 |
| ## 6 | 0 | 0.00000000 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0 |
| ## | tfidf_guess | tfidf_guy | tfidf_happen | tfidf_happi | tfidf_hard | tfidf_hate | |
| ## 1 | 0.09565954 | 0 | 0 | 0 | 0.000000 | 0.0000000 | |
| ## 2 | 0.00000000 | 0 | 0 | 0 | 0.000000 | 0.0000000 | |
| ## 3 | 0.00000000 | 0 | 0 | 0 | 0.000000 | 0.1502026 | |
| ## 4 | 0.00000000 | 0 | 0 | 0 | 0.000000 | 0.0000000 | |
| ## 5 | 0.00000000 | 0 | 0 | 0 | 0.000000 | 0.0000000 | |
| ## 6 | 0.00000000 | 0 | 0 | 0 | 0.205813 | 0.0000000 | |
| ## | tfidf_head | tfidf_health | tfidf_hear | tfidf_heart | tfidf_help | tfidf_high | |
| ## 1 | 0 | 0.1846050 | 0 | 0 | 0 | 0 | |
| ## 2 | 0 | 0.0000000 | 0 | 0 | 0 | 0 | |
| ## 3 | 0 | 0.3234264 | 0 | 0 | 0 | 0 | |
| ## 4 | 0 | 0.2047059 | 0 | 0 | 0 | 0 | |
| ## 5 | 0 | 0.1733893 | 0 | 0 | 0 | 0 | |
| ## 6 | 0 | 0.2442160 | 0 | 0 | 0 | 0 | |

| | | | | | | | |
|------|--------------|-------------|-------------|------------|-------------|-------------|-------------|
| ## | tfidf_home | tfidf_hope | tfidf_hour | tfidf_hous | tfidf_hurt | tfidf_idea | tfidf_im |
| ## 1 | 0 | 0.08624765 | 0.0000000 | 0 | 0 | 0 | 0.281393 |
| ## 2 | 0 | 0.00000000 | 0.0000000 | 0 | 0 | 0 | 0.000000 |
| ## 3 | 0 | 0.00000000 | 0.0000000 | 0 | 0 | 0 | 0.000000 |
| ## 4 | 0 | 0.00000000 | 0.0000000 | 0 | 0 | 0 | 0.000000 |
| ## 5 | 0 | 0.00000000 | 0.1606407 | 0 | 0 | 0 | 0.000000 |
| ## 6 | 0 | 0.00000000 | 0.0000000 | 0 | 0 | 0 | 0.000000 |
| ## | tfidf_issu | tfidf_job | tfidf_keep | tfidf_kill | tfidf_kind | tfidf_know | tfidf_last |
| ## 1 | 0 | 0 | 0 | 0.0000000 | 0 | 0.04772274 | 0.1415657 |
| ## 2 | 0 | 0 | 0 | 0.0000000 | 0 | 0.00000000 | 0.0000000 |
| ## 3 | 0 | 0 | 0 | 0.0000000 | 0 | 0.16721966 | 0.0000000 |
| ## 4 | 0 | 0 | 0 | 0.0000000 | 0 | 0.10583816 | 0.0000000 |
| ## 5 | 0 | 0 | 0 | 0.1804685 | 0 | 0.00000000 | 0.0000000 |
| ## 6 | 0 | 0 | 0 | 0.0000000 | 0 | 0.00000000 | 0.0000000 |
| ## | tfidf_late | tfidf_leav | tfidf_left | tfidf_let | tfidf_life | tfidf_like | tfidf_littl |
| ## 1 | 0.0000000 | 0 | 0 | 0.0000000 | 0.0000000 | 0.0000000 | 0.000000 |
| ## 2 | 0.0000000 | 0 | 0 | 0.0000000 | 0.0000000 | 0.0000000 | 0.000000 |
| ## 3 | 0.0000000 | 0 | 0 | 0.0000000 | 0.0000000 | 0.0000000 | 0.000000 |
| ## 4 | 0.2172411 | 0 | 0 | 0.1969284 | 0.0000000 | 0.0000000 | 0.386036 |
| ## 5 | 0.0000000 | 0 | 0 | 0.0000000 | 0.1152168 | 0.0000000 | 0.000000 |
| ## 6 | 0.0000000 | 0 | 0 | 0.0000000 | 0.0000000 | 0.1122454 | 0.000000 |
| ## | tfidf_live | tfidf_long | tfidf_look | tfidf_lose | tfidf_lost | tfidf_lot | tfidf_love |
| ## 1 | 0 | 0.07684704 | 0 | 0 | 0 | 0.07382478 | 0 |
| ## 2 | 0 | 0.00000000 | 0 | 0 | 0 | 0.00000000 | 0 |
| ## 3 | 0 | 0.00000000 | 0 | 0 | 0 | 0.00000000 | 0 |
| ## 4 | 0 | 0.00000000 | 0 | 0 | 0 | 0.00000000 | 0 |
| ## 5 | 0 | 0.00000000 | 0 | 0 | 0 | 0.13867912 | 0 |
| ## 6 | 0 | 0.00000000 | 0 | 0 | 0 | 0.00000000 | 0 |
| ## | tfidf_made | tfidf_make | tfidf_mani | tfidf_mayb | tfidf_mean | tfidf_med | tfidf_medic |
| ## 1 | 0.08216623 | 0.1159273 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 2 | 0.00000000 | 0.0000000 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 3 | 0.00000000 | 0.0000000 | 0 | 0 | 0.1657204 | 0 | 0 |
| ## 4 | 0.00000000 | 0.0000000 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 5 | 0.00000000 | 0.0000000 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 6 | 0.00000000 | 0.0000000 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## | tfidf_mental | tfidf_might | tfidf_mind | tfidf_mom | tfidf_month | tfidf_move | |
| ## 1 | 0 | 0 | 0.08767951 | 0 | 0.0000000 | 0 | |
| ## 2 | 0 | 0 | 0.00000000 | 0 | 0.0000000 | 0 | |
| ## 3 | 0 | 0 | 0.00000000 | 0 | 0.0000000 | 0 | |
| ## 4 | 0 | 0 | 0.00000000 | 0 | 0.0000000 | 0 | |
| ## 5 | 0 | 0 | 0.00000000 | 0 | 0.2604023 | 0 | |
| ## 6 | 0 | 0 | 0.23198436 | 0 | 0.0000000 | 0 | |
| ## | tfidf_much | tfidf_need | tfidf_never | tfidf_new | tfidf_next | tfidf_night | |
| ## 1 | 0 | 0.06438772 | 0 | 0 | 0.0000000 | 0.0000000 | |
| ## 2 | 0 | 0.00000000 | 0 | 0 | 0.0000000 | 0.0000000 | |
| ## 3 | 0 | 0.00000000 | 0 | 0 | 0.0000000 | 0.0000000 | |
| ## 4 | 0 | 0.00000000 | 0 | 0 | 0.2048134 | 0.0000000 | |
| ## 5 | 0 | 0.00000000 | 0 | 0 | 0.0000000 | 0.4592323 | |
| ## 6 | 0 | 0.00000000 | 0 | 0 | 0.0000000 | 0.0000000 | |
| ## | tfidf_normal | tfidf_noth | tfidf_notic | tfidf_old | tfidf_onc | tfidf_one | tfidf_onli |
| ## 1 | 0 | 0 | 0 | 0 | 0.0000000 | 0 | 0.06604262 |
| ## 2 | 0 | 0 | 0 | 0 | 0.0000000 | 0 | 0.00000000 |
| ## 3 | 0 | 0 | 0 | 0 | 0.0000000 | 0 | 0.00000000 |
| ## 4 | 0 | 0 | 0 | 0 | 0.0000000 | 0 | 0.00000000 |

| | | | | | | | |
|------|--------------------|---------------|---------------|--------------|---------------|----------------|------------|
| ## 5 | 0 | 0 | 0 | 0 | 0.1737179 | 0 | 0.12406040 |
| ## 6 | 0 | 0 | 0 | 0 | 0.0000000 | 0 | 0.00000000 |
| ## | tfidf_pain | tfidf_panic | tfidf_parent | tfidf_part | tfidf_past | tfidf_peopl | |
| ## 1 | 0.0000000 | 0 | 0 | 0 | 0.0828947 | 0.0000000 | |
| ## 2 | 0.6086886 | 0 | 0 | 0 | 0.0000000 | 0.0000000 | |
| ## 3 | 0.0000000 | 0 | 0 | 0 | 0.0000000 | 0.1054718 | |
| ## 4 | 0.0000000 | 0 | 0 | 0 | 0.0000000 | 0.1335123 | |
| ## 5 | 0.3471432 | 0 | 0 | 0 | 0.0000000 | 0.0000000 | |
| ## 6 | 0.0000000 | 0 | 0 | 0 | 0.0000000 | 0.0000000 | |
| ## | tfidf_person | tfidf_place | tfidf_pleas | tfidf_point | tfidf_possibl | tfidf_post | |
| ## 1 | 0.1516435 | 0.09072504 | 0 | 0 | | 0.08316479 | |
| ## 2 | 0.0000000 | 0.00000000 | 0 | 0 | | 0.00000000 | |
| ## 3 | 0.0000000 | 0.00000000 | 0 | 0 | | 0.00000000 | |
| ## 4 | 0.0000000 | 0.00000000 | 0 | 0 | | 0.18444054 | |
| ## 5 | 0.0000000 | 0.00000000 | 0 | 0 | | 0.00000000 | |
| ## 6 | 0.0000000 | 0.00000000 | 0 | 0 | | 0.00000000 | |
| ## | tfidf_pretti | tfidf_probabl | tfidf_problem | tfidf_ptsd | tfidf_put | tfidf_question | |
| ## 1 | 0 | 0 | 0.08309023 | 0 | 0 | 0 | |
| ## 2 | 0 | 0 | 0.00000000 | 0 | 0 | 0 | |
| ## 3 | 0 | 0 | 0.00000000 | 0 | 0 | 0 | |
| ## 4 | 0 | 0 | 0.00000000 | 0 | 0 | 0 | |
| ## 5 | 0 | 0 | 0.00000000 | 0 | 0 | 0 | |
| ## 6 | 0 | 0 | 0.00000000 | 0 | 0 | 0 | |
| ## | tfidf_quit | tfidf_read | tfidf_real | tfidf_realli | tfidf_reason | tfidf_recent | |
| ## 1 | 0 | 0 | 0 | 0.05562930 | 0 | 0 | |
| ## 2 | 0 | 0 | 0 | 0.00000000 | 0 | 0 | |
| ## 3 | 0 | 0 | 0 | 0.09746206 | 0 | 0 | |
| ## 4 | 0 | 0 | 0 | 0.00000000 | 0 | 0 | |
| ## 5 | 0 | 0 | 0 | 0.10449910 | 0 | 0 | |
| ## 6 | 0 | 0 | 0 | 0.14718523 | 0 | 0 | |
| ## | tfidf_relationship | tfidf_rememb | tfidf_right | tfidf_said | tfidf_say | tfidf_scare | |
| ## 1 | 0 | 0 | 0.0000000 | 0.08260083 | 0.0000000 | 0 | |
| ## 2 | 0 | 0 | 0.0000000 | 0.00000000 | 0.0000000 | 0 | |
| ## 3 | 0 | 0 | 0.1299231 | 0.00000000 | 0.1209845 | 0 | |
| ## 4 | 0 | 0 | 0.0000000 | 0.00000000 | 0.0000000 | 0 | |
| ## 5 | 0 | 0 | 0.1393039 | 0.00000000 | 0.0000000 | 0 | |
| ## 6 | 0 | 0 | 0.0000000 | 0.00000000 | 0.0000000 | 0 | |
| ## | tfidf_school | tfidf_see | tfidf_seem | tfidf_self | tfidf_sever | tfidf_shit | |
| ## 1 | 0 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 2 | 0 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 3 | 0 | 0 | 0.1384987 | 0 | 0 | 0 | |
| ## 4 | 0 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 5 | 0 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## 6 | 0 | 0 | 0.0000000 | 0 | 0 | 0 | |
| ## | tfidf_sinc | tfidf_situat | tfidf_sleep | tfidf_social | tfidf_someon | tfidf_someth | |
| ## 1 | 0.0000000 | 0.0976072 | 0 | 0 | 0 | 0.2005516 | |
| ## 2 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | 0.0000000 | |
| ## 3 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | 0.2342431 | |
| ## 4 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | 0.0000000 | |
| ## 5 | 0.1349349 | 0.0000000 | 0 | 0 | 0 | 0.0000000 | |
| ## 6 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | 0.0000000 | |
| ## | tfidf_sometim | tfidf_sorri | tfidf_start | tfidf_stay | tfidf_still | tfidf_stop | |
| ## 1 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## 2 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |

| | | | | | | | |
|------|------------------|---------------|--------------|--------------|---------------|---------------|-------------|
| ## 3 | 0 | 0.0000000 | 0.1076825 | 0 | 0 | 0 | |
| ## 4 | 0 | 0.2155800 | 0.1363107 | 0 | 0 | 0 | |
| ## 5 | 0 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | |
| ## 6 | 0 | 0.2571889 | 0.0000000 | 0 | 0 | 0 | |
| ## | tfidf_stress | tfidf_struggl | tfidf_stuff | tfidf_suicid | tfidf_support | tfidf_sure | |
| ## 1 | 0.1989405 | 0.0000000 | 0 | 0 | 0 | 0.08184181 | |
| ## 2 | 0.3276337 | 0.0000000 | 0 | 0 | 0 | 0.00000000 | |
| ## 3 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | 0.00000000 | |
| ## 4 | 0.0000000 | 0.2013288 | 0 | 0 | 0 | 0.00000000 | |
| ## 5 | 0.1868538 | 0.0000000 | 0 | 0 | 0 | 0.15373903 | |
| ## 6 | 0.0000000 | 0.0000000 | 0 | 0 | 0 | 0.00000000 | |
| ## | tfidf_symptom | tfidf_take | tfidf_talk | tfidf_tell | tfidf_thank | tfidf_therapi | |
| ## 1 | 0.3610230 | 0.06528979 | 0 | 0.0000000 | 0 | 0 | |
| ## 2 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0 | 0 | |
| ## 3 | 0.0000000 | 0.00000000 | 0 | 0.1329733 | 0 | 0 | |
| ## 4 | 0.0000000 | 0.14479787 | 0 | 0.0000000 | 0 | 0 | |
| ## 5 | 0.1695445 | 0.00000000 | 0 | 0.0000000 | 0 | 0 | |
| ## 6 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0 | 0 | |
| ## | tfidf_therapist | tfidf_thing | tfidf_think | tfidf_though | tfidf_thought | tfidf_time | |
| ## 1 | 0 | 0.05768992 | 0.05564094 | 0.08531584 | 0.0000000 | 0.1519041 | |
| ## 2 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0.0000000 | 0.0000000 | |
| ## 3 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0.0000000 | 0.0000000 | |
| ## 4 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0.0000000 | 0.0000000 | |
| ## 5 | 0 | 0.00000000 | 0.00000000 | 0.00000000 | 0.1294994 | 0.0000000 | |
| ## 6 | 0 | 0.15263724 | 0.00000000 | 0.00000000 | 0.0000000 | 0.0000000 | |
| ## | tfidf_tire | tfidf_today | tfidf_told | tfidf_took | tfidf_tri | tfidf_turn | tfidf_two |
| ## 1 | 0.0998051 | 0.08231028 | 0 | 0.0000000 | 0.05895588 | 0 | 0 |
| ## 2 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0.00000000 | 0 | 0 |
| ## 3 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0.10329019 | 0 | 0 |
| ## 4 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0.13075069 | 0 | 0 |
| ## 5 | 0.0000000 | 0.00000000 | 0 | 0.1812327 | 0.00000000 | 0 | 0 |
| ## 6 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0.00000000 | 0 | 0 |
| ## | tfidf_understand | tfidf_us | tfidf_use | tfidf_usual | tfidf_veri | tfidf_want | |
| ## 1 | 0 | 0.0000000 | 0.07499834 | 0 | 0.1432974 | 0.0000000 | |
| ## 2 | 0 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0.0000000 | |
| ## 3 | 0 | 0.1830006 | 0.00000000 | 0 | 0.0000000 | 0.0897336 | |
| ## 4 | 0 | 0.0000000 | 0.00000000 | 0 | 0.3178010 | 0.0000000 | |
| ## 5 | 0 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0.0000000 | |
| ## 6 | 0 | 0.0000000 | 0.00000000 | 0 | 0.0000000 | 0.1355139 | |
| ## | tfidf_way | tfidf_week | tfidf_weight | tfidf_well | tfidf_went | tfidf_whi | tfidf_whole |
| ## 1 | 0.0689263 | 0.06965452 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 2 | 0.0000000 | 0.00000000 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 3 | 0.1207583 | 0.00000000 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 4 | 0.0000000 | 0.30895569 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## 5 | 0.0000000 | 0.00000000 | 0 | 0 | 0.1512142 | 0 | 0 |
| ## 6 | 0.0000000 | 0.18429346 | 0 | 0 | 0.0000000 | 0 | 0 |
| ## | tfidf_wish | tfidf_without | tfidf_wonder | tfidf_work | tfidf_worri | tfidf_wors | |
| ## 1 | 0 | 0.000000 | 0 | 0.0000000 | 0.0000000 | 0 | |
| ## 2 | 0 | 0.000000 | 0 | 0.0000000 | 0.2904661 | 0 | |
| ## 3 | 0 | 0.000000 | 0 | 0.1107566 | 0.0000000 | 0 | |
| ## 4 | 0 | 0.000000 | 0 | 0.0000000 | 0.0000000 | 0 | |
| ## 5 | 0 | 0.000000 | 0 | 0.0000000 | 0.0000000 | 0 | |
| ## 6 | 0 | 0.221473 | 0 | 0.0000000 | 0.0000000 | 0 | |
| ## | tfidf_would | tfidf_wrong | tfidf_x200b | tfidf_year | | | |

```
## 1      0  0.0000000      0 0.05591948
## 2      0  0.0000000      0 0.00000000
## 3      0  0.3326731      0 0.00000000
## 4      0  0.0000000      0 0.00000000
## 5      0  0.0000000      0 0.00000000
## 6      0  0.0000000      0 0.14795297
```

```
# Second dataset
```

```
head(indicators_anxiety)
```

```
##           Indicator      Group      State      Subgroup
## 1 Symptoms of Depressive Disorder National Estimate United States United States
## 2 Symptoms of Depressive Disorder      By Age United States 18 - 29 years
## 3 Symptoms of Depressive Disorder      By Age United States 30 - 39 years
## 4 Symptoms of Depressive Disorder      By Age United States 40 - 49 years
## 5 Symptoms of Depressive Disorder      By Age United States 50 - 59 years
## 6 Symptoms of Depressive Disorder      By Age United States 60 - 69 years
##   Phase Time.Period   Time.Period.Label Time.Period.Start.Date
## 1   1.0          1 Apr 23 - May 5, 2020          04/23/2020
## 2   1.0          1 Apr 23 - May 5, 2020          04/23/2020
## 3   1.0          1 Apr 23 - May 5, 2020          04/23/2020
## 4   1.0          1 Apr 23 - May 5, 2020          04/23/2020
## 5   1.0          1 Apr 23 - May 5, 2020          04/23/2020
## 6   1.0          1 Apr 23 - May 5, 2020          04/23/2020
##   Time.Period.End.Date Value Low.CI High.CI Confidence.Interval Quartile.Range
## 1          05/05/2020  23.5  22.7  24.3          22.7 - 24.3
## 2          05/05/2020  32.7  30.2  35.2          30.2 - 35.2
## 3          05/05/2020  25.7  24.1  27.3          24.1 - 27.3
## 4          05/05/2020  24.8  23.3  26.2          23.3 - 26.2
## 5          05/05/2020  23.2  21.5  25.0          21.5 - 25.0
## 6          05/05/2020  18.4  17.0  19.7          17.0 - 19.7
```

```
# Turn data into a df
```

```
indicators_anxiety <- as.data.frame(indicators_anxiety)
```

```
# Find unique indicators
```

```
unique(indicators_anxiety$Indicator)
```

```
## [1] "Symptoms of Depressive Disorder"
```

```
## [2] "Symptoms of Anxiety Disorder"
```

```
## [3] "Symptoms of Anxiety Disorder or Depressive Disorder"
```

```
# Filter for only anxiety not depression
```

```
indicators_anxiety <- filter(indicators_anxiety,
                             Indicator == 'Symptoms of Anxiety Disorder')
```

```
# Validate that it worked
```

```
head(indicators_anxiety, 8)
```

```
##           Indicator      Group      State
## 1 Symptoms of Anxiety Disorder National Estimate United States
## 2 Symptoms of Anxiety Disorder      By Age United States
## 3 Symptoms of Anxiety Disorder      By Age United States
## 4 Symptoms of Anxiety Disorder      By Age United States
```

```
## 5 Symptoms of Anxiety Disorder      By Age United States
## 6 Symptoms of Anxiety Disorder      By Age United States
## 7 Symptoms of Anxiety Disorder      By Age United States
## 8 Symptoms of Anxiety Disorder      By Age United States
##      Subgroup Phase Time.Period    Time.Period.Label
## 1      United States    1.0          1 Apr 23 - May 5, 2020
## 2      18 - 29 years    1.0          1 Apr 23 - May 5, 2020
## 3      30 - 39 years    1.0          1 Apr 23 - May 5, 2020
## 4      40 - 49 years    1.0          1 Apr 23 - May 5, 2020
## 5      50 - 59 years    1.0          1 Apr 23 - May 5, 2020
## 6      60 - 69 years    1.0          1 Apr 23 - May 5, 2020
## 7      70 - 79 years    1.0          1 Apr 23 - May 5, 2020
## 8 80 years and above    1.0          1 Apr 23 - May 5, 2020
##      Time.Period.Start.Date Time.Period.End.Date Value Low.CI High.CI
## 1              04/23/2020          05/05/2020  30.8   30.0   31.7
## 2              04/23/2020          05/05/2020  40.2   38.1   42.4
## 3              04/23/2020          05/05/2020  34.4   32.6   36.1
## 4              04/23/2020          05/05/2020  34.1   32.1   36.2
## 5              04/23/2020          05/05/2020  31.0   29.0   33.1
## 6              04/23/2020          05/05/2020  24.9   23.6   26.3
## 7              04/23/2020          05/05/2020  16.4   14.8   18.1
## 8              04/23/2020          05/05/2020  14.6   11.5   18.2
##      Confidence.Interval Quartile.Range
## 1              30.0 - 31.7
## 2              38.1 - 42.4
## 3              32.6 - 36.1
## 4              32.1 - 36.2
## 5              29.0 - 33.1
## 6              23.6 - 26.3
## 7              14.8 - 18.1
## 8              11.5 - 18.2
```

```
# Third dataset
head(global_mental)
```

```
##      index      Entity Code Year Schizophrenia.... Bipolar.disorder....
## 1      0 Afghanistan  AFG 1990          0.16056          0.697779
## 2      1 Afghanistan  AFG 1991          0.160312          0.697961
## 3      2 Afghanistan  AFG 1992          0.160135          0.698107
## 4      3 Afghanistan  AFG 1993          0.160037          0.698257
## 5      4 Afghanistan  AFG 1994          0.160022          0.698469
## 6      5 Afghanistan  AFG 1995          0.160076          0.698695
##      Eating.disorders.... Anxiety.disorders.... Drug.use.disorders....
## 1              0.101855          4.828830          1.677082
## 2              0.099313          4.829740          1.684746
## 3              0.096692          4.831108          1.694334
## 4              0.094336          4.830864          1.705320
## 5              0.092439          4.829423          1.716069
## 6              0.09098          4.828337          1.728112
##      Depression.... Alcohol.use.disorders....
## 1      4.071831          0.672404
## 2      4.079531          0.671768
## 3      4.088358          0.670644
## 4      4.096190          0.669738
```

```
## 5      4.099582      0.669260
## 6      4.104207      0.668746
```

```
# Turn data into a df
global_mental <- as.data.frame(global_mental)
# Remove columns that we don't need
colnames(global_mental)
```

```
## [1] "index"      "Entity"
## [3] "Code"       "Year"
## [5] "Schizophrenia...." "Bipolar.disorder...."
## [7] "Eating.disorders...." "Anxiety.disorders...."
## [9] "Drug.use.disorders...." "Depression...."
## [11] "Alcohol.use.disorders...."
```

```
global_mental <- subset(global_mental, select = -c(Schizophrenia....,
                                                    Bipolar.disorder....,
                                                    Eating.disorders....,
                                                    Drug.use.disorders....,
                                                    Depression....,
                                                    Alcohol.use.disorders....))

# Remove rows missing data
global_mental <- global_mental[complete.cases(global_mental),]
# Rename
global_mental <- global_mental %>%
  rename(anxiety = Anxiety.disorders....)
# Check that the data has been trimmed down
head(global_mental, 8)
```

```
##   index      Entity Code Year  anxiety
## 1      0 Afghanistan  AFG 1990 4.828830
## 2      1 Afghanistan  AFG 1991 4.829740
## 3      2 Afghanistan  AFG 1992 4.831108
## 4      3 Afghanistan  AFG 1993 4.830864
## 5      4 Afghanistan  AFG 1994 4.829423
## 6      5 Afghanistan  AFG 1995 4.828337
## 7      6 Afghanistan  AFG 1996 4.828083
## 8      7 Afghanistan  AFG 1997 4.827726
```

```
# Fourth dataset
head(mental)
```

```
##           Timestamp Gender      Country Occupation self_employed
## 1 2014-08-27 11:29:31 Female United States  Corporate
## 2 2014-08-27 11:31:50 Female United States  Corporate
## 3 2014-08-27 11:32:39 Female United States  Corporate
## 4 2014-08-27 11:37:59 Female United States  Corporate      No
## 5 2014-08-27 11:43:36 Female United States  Corporate      No
## 6 2014-08-27 11:49:51 Female      Poland  Corporate      No
##   family_history treatment Days_Indoors Growing_Stress Changes_Habits
## 1              No        Yes    1-14 days              Yes          No
## 2              Yes        Yes    1-14 days              Yes          No
```

```
## 3      Yes      Yes  1-14 days      Yes      No
## 4      Yes      Yes  1-14 days      Yes      No
## 5      Yes      Yes  1-14 days      Yes      No
## 6      No       Yes  1-14 days      Yes      No
##   Mental_Health_History Mood_Swings Coping_Struggles Work_Interest
## 1                      Yes      Medium      No      No
## 2                      Yes      Medium      No      No
## 3                      Yes      Medium      No      No
## 4                      Yes      Medium      No      No
## 5                      Yes      Medium      No      No
## 6                      Yes      Medium      No      No
##   Social_Weakness mental_health_interview care_options
## 1                Yes                      No      Not sure
## 2                Yes                      No      No
## 3                Yes                      No      Yes
## 4                Yes                    Maybe      Yes
## 5                Yes                      No      Yes
## 6                Yes                    Maybe      Not sure
```

```
# Turn data into dataframe
mental <- as.data.frame(mental)
# Remove columns that we don't need
colnames(mental)
```

```
## [1] "Timestamp"      "Gender"
## [3] "Country"        "Occupation"
## [5] "self_employed"  "family_history"
## [7] "treatment"      "Days_Indoors"
## [9] "Growing_Stress" "Changes_Habits"
## [11] "Mental_Health_History" "Mood_Swings"
## [13] "Coping_Struggles"    "Work_Interest"
## [15] "Social_Weakness"     "mental_health_interview"
## [17] "care_options"
```

```
mental <- subset(mental, select = -c(Occupation,
                                     self_employed,
                                     family_history,
                                     Days_Indoors,
                                     Growing_Stress,
                                     Changes_Habits,
                                     Mental_Health_History,
                                     Mood_Swings,
                                     Coping_Struggles,
                                     Work_Interest,
                                     Social_Weakness,
                                     mental_health_interview,
                                     care_options))

# Remove rows missing data
mental <- mental[complete.cases(mental),]
# Check that the data has been trimmed
head(mental, 8)
```

```
##           Timestamp Gender      Country treatment
```



```
## 1 2014-08-27 11:29:31 Female United States Yes
## 2 2014-08-27 11:31:50 Female United States Yes
## 3 2014-08-27 11:32:39 Female United States Yes
## 4 2014-08-27 11:37:59 Female United States Yes
## 5 2014-08-27 11:43:36 Female United States Yes
## 6 2014-08-27 11:49:51 Female Poland Yes
## 7 2014-08-27 11:51:34 Female Australia Yes
## 8 2014-08-27 11:52:41 Female United States No
```

```
# Fifth dataset
head(prevalence)
```

```
## index Entity Code Year
## 1 0 Abkhazia OWID_ABK 2015
## 2 1 Afghanistan AFG 1990
## 3 2 Afghanistan AFG 1991
## 4 3 Afghanistan AFG 1992
## 5 4 Afghanistan AFG 1993
## 6 5 Afghanistan AFG 1994
## Prevalence...Anxiety.disorders...Sex..Male...Age..Age.standardized..Percent.
## 1 NA
## 2 3.556843
## 3 3.548885
## 4 3.542779
## 5 3.538304
## 6 3.535309
## Prevalence...Anxiety.disorders...Sex..Female...Age..Age.standardized..Percent.
## 1 NA
## 2 5.971172
## 3 5.980482
## 4 5.988175
## 5 5.993858
## 6 5.997363
## Population..historical.estimates. Continent
## 1 NA Asia
## 2 12412311
## 3 13299016
## 4 14485543
## 5 15816601
## 6 17075728
```

```
# Turn data into dataframe
prevalence <- as.data.frame(prevalence)
# Remove columns that we don't need
colnames(prevalence)
```

```
## [1] "index"
## [2] "Entity"
## [3] "Code"
## [4] "Year"
## [5] "Prevalence...Anxiety.disorders...Sex..Male...Age..Age.standardized..Percent."
## [6] "Prevalence...Anxiety.disorders...Sex..Female...Age..Age.standardized..Percent."
## [7] "Population..historical.estimates."
## [8] "Continent"
```

```

prevalence <- subset(prevalence, select = -c(Continent,
                                             Population..historical.estimates.))

# Rename columns
prevalence <- prevalence %>%
  rename(index = index, entity = Entity, code = Code, year = Year, male = Prevalence...Anxiety.disorder,
         female =
Prevalence...Anxiety.disorders...Sex..Female...Age..Age.standardized..Percent.)
# Check changes
head(prevalence, 8)

```

```

##   index      entity      code year      male      female
## 1      0  Abkhazia  OWID_ABK 2015         NA         NA
## 2      1  Afghanistan      AFG 1990 3.556843 5.971172
## 3      2  Afghanistan      AFG 1991 3.548885 5.980482
## 4      3  Afghanistan      AFG 1992 3.542779 5.988175
## 5      4  Afghanistan      AFG 1993 3.538304 5.993858
## 6      5  Afghanistan      AFG 1994 3.535309 5.997363
## 7      6  Afghanistan      AFG 1995 3.533797 5.998540
## 8      7  Afghanistan      AFG 1996 3.535415 5.996443

```

```

# Sixth dataset
head(meds)

```

```

##   X      drug_name      date  age gender      time_on_drug reviewer_type
## 1 0 Sertraline Oral 5/12/2024 45-54 Female 1 to less than 2 years Patient
## 2 1 Sertraline Oral 4/21/2024 35-44 Female      less than 1 month Patient
## 3 2 Sertraline Oral 4/16/2024 25-34 Female 2 to less than 5 years Patient
## 4 3 Sertraline Oral 4/11/2024 45-54 Male      less than 1 month Patient
## 5 4 Sertraline Oral 4/8/2024 13-18 Female
## 6 5 Sertraline Oral 3/29/2024 45-54 Female      less than 1 month Patient
##
##           condition rating_overall rating_effectiveness
## 1 Posttraumatic Stress Syndrome      5.0      5
## 2           Depression      1.0      1
## 3 Repeated Episodes of Anxiety      4.3      4
## 4           Panic Disorder      1.7      1
## 5 Major Depressive Disorder      3.0      2
## 6 Posttraumatic Stress Syndrome      2.3      1
##   rating_ease_of_use rating_satisfaction
## 1           5      5
## 2           1      1
## 3           4      5
## 4           3      1
## 5           4      3
## 6           5      1
##
## 1
## 2
## 3
## 4 Of course, take this with a pinch of salt because everyone's chemistry is different, and I am DEFINITELY
## 5
## 6

```

```
# Turn data into dataframe
```

```
meds <- as.data.frame(meds)
```

```
# Remove columns that we don't need
```

```
colnames(meds)
```

```
## [1] "X" "drug_name" "date"  
## [4] "age" "gender" "time_on_drug"  
## [7] "reviewer_type" "condition" "rating_overall"  
## [10] "rating_effectiveness" "rating_ease_of_use" "rating_satisfaction"  
## [13] "text"
```

```
meds <- subset(meds, select = -c(time_on_drug, text))
```

```
# Find conditions that are anxiety related
```

```
unique(meds$condition)
```

```
## [1] "Posttraumatic Stress Syndrome"  
## [2] "Depression"  
## [3] "Repeated Episodes of Anxiety"  
## [4] "Panic Disorder"  
## [5] "Major Depressive Disorder"  
## [6] "Bipolar Depression"  
## [7] "Depressed Mood Disorder Occurring Every Year at the Same Time"  
## [8] "Anxiousness associated with Depression"  
## [9] "Other"  
## [10] "Obsessive Compulsive Disorder"  
## [11] "Extreme Apprehension or Fear of Social Interaction"  
## [12] "Premenstrual Disorder with a State of Unhappiness"  
## [13] "Premature Ejection of Semen"  
## [14] "Problem Behavior"  
## [15] "Schizophrenia"  
## [16] "Osteoporosis"  
## [17] "Aggressive Behavior"  
## [18] "Binge Eating Disorder"  
## [19] "Acute Repetitive Seizures"  
## [20] "Anxious"  
## [21] "Muscle Spasm"  
## [22] "Neuropathic Pain"  
## [23] "Attention Deficit Disorder with Hyperactivity"  
## [24] "Stop Smoking"  
## [25] "Manic"  
## [26] "Convulsive Seizures"  
## [27] "Facial Nerve Pain"  
## [28] "Bipolar I Disorder with Most Recent Episode Mixed"  
## [29] "Nerve Pain"  
## [30] "Bipolar Disorder in Remission"  
## [31] "Epileptic Seizure"  
## [32] "Tonic"  
## [33] "Prevention of Seizures following Head Injury or Surgery"  
## [34] "Inducing of a Relaxed Easy State"  
## [35] "Sleep Disturbance with Extreme Anxiety"  
## [36] "Psychosis caused by Sudden Alcohol Withdrawal"  
## [37] "Seizure with Loss of Normal Tone or Strength"  
## [38] "Chronic Trouble Sleeping"
```

[39] "A Feeling of Restlessness with Inability to Sit Still"
 ## [40] "Tension Headache"
 ## [41] "Calming of Pediatric Patient by Administration of Sedative"
 ## [42] "Symptoms from Stopping Treatment with Opioid Drugs"
 ## [43] "Symptoms from Alcohol Withdrawal"
 ## [44] "Brief Muscle Spasms in an Infant"
 ## [45] "Mania associated with Bipolar Disorder"
 ## [46] "Mental Disorder with Loss of Normal Personality"
 ## [47] "Nausea and Vomiting"
 ## [48] "Hiccups that are Hard to Cure"
 ## [49] "Delirium"
 ## [50] "Tourette"
 ## [51] "Generalized Attack of Muscular Weakness"
 ## [52] "Simple Partial Seizures"
 ## [53] "Rapid Cycle Manic"
 ## [54] "Additional Medications to Treat Depression"
 ## [55] "Multiple Seizure Types"
 ## [56] "Apprehension"
 ## [57] "Combative and Explosive Behavior"
 ## [58] "Schizophrenia With Mood Changes"
 ## [59] "Psychosis caused by a Disease"
 ## [60] "Pervasive Developmental Disorder"
 ## [61] "A Rare Developmental Disorder of Infants"
 ## [62] "Paroxysmal Choreoathetosis"
 ## [63] "Nerve Pain of Tongue and Throat from 9th Cranial Nerve"
 ## [64] "Extreme Discomfort in Calves when Sitting or Lying Down"
 ## [65] ""
 ## [66] "Nicotine Addiction"
 ## [67] "Depression associated with Bipolar Disorder"
 ## [68] "Nausea and Vomiting caused by Cancer Drugs"
 ## [69] "Anxiety associated with an Operation"
 ## [70] "Feeling Restless"
 ## [71] "Additional Medication for Calming"
 ## [72] "Additional Agent to Induce General Anesthesia"
 ## [73] "Sedation with Ability to Respond to Stimulation or Speech"
 ## [74] "Induce Temporary Amnesia"
 ## [75] "Psychosis associated with Alzheimer"
 ## [76] "Prevent Nausea and Vomiting from Cancer Chemotherapy"
 ## [77] "Depression following Delivery of Baby"
 ## [78] "Overweight"
 ## [79] "Bulimia"
 ## [80] "Anorexia Nervosa"
 ## [81] "Muscle Weakness associated with Sleeping Disease"
 ## [82] "Migraine Prevention"
 ## [83] "Disorder characterized by Stiff"
 ## [84] "Agitation associated with Schizophrenia"
 ## [85] "Psychotic Depressive Illness"
 ## [86] "Severe Anxiety"
 ## [87] "Itching"
 ## [88] "Severe Itching"
 ## [89] "Life Threatening Allergic Reaction"
 ## [90] "Additional Medications to Treat Pain"
 ## [91] "Ventricular Premature Beats"
 ## [92] "Heart Ventricle Rhythm Problem"

```
## [93] "High Blood Pressure"
## [94] "Feeling of Apprehension Before an Operation"
## [95] "Involuntary Quivering"
## [96] "Psychosis caused by a Drug"
## [97] "Brief Episode of Schizophrenia with Rapid Onset"
## [98] "Chronic Type of Schizophrenia"
## [99] "Over Excitement"
## [100] "Allergic Conjunctivitis"
## [101] "Inflammation of the Nose due to an Allergy"
## [102] "Inflammation of Skin caused by an Allergy"
## [103] "Sneezing"
## [104] "Hives"
## [105] "Feel Like Throwing Up"
## [106] "Welt from Pressure on Skin"
## [107] "Cluster Headache Prevention"
## [108] "Seizures with Breaks in Consciousness"
## [109] "Petit Mal Seizures"
## [110] "Petit Mal Epilepsy with Multiple Seizure Types"
## [111] "Seizures with Irregular Muscle Contractions"
## [112] "Runny Nose"
## [113] "Additional Medication for Myoclonic Epilepsy"
## [114] "adjunct therapy for obsessive compulsive disorder"
## [115] "Frequent Headaches"
## [116] "Mitral Valve Prolapse Syndrome"
## [117] "Myocardial Reinfarction Prevention"
## [118] "Problems with Bladder Control"
## [119] "Bedwetting"
## [120] "Stuffy Nose"
## [121] "Essential Tremor"
## [122] "Irritable Colon"
## [123] "Nerve Pain after Herpes"
## [124] "Ulcer from Stomach Acid"
## [125] "Behaving with Excessive Cheerfulness and Activity"
## [126] "Recurring Sleep Episodes During the Day"
## [127] "Chronic Muscle or Bone Pain"
## [128] "Diabetic Complication causing Injury to some Body Nerves"
```

```
# The following would qualify as something I don't know how to do
# that could be helpful to know.
# I don't know if there is a way to iterate through the unique items in the
# condition column and find conditions that are anxiety related using R.
# I would imagine it could be done through a loop and doing a partial match
# to condition which contains the letters "anx". I looked through the printed
# conditions myself and then coded the following lines.
```

```
# Remove rows that aren't anxiety related
meds <- subset(meds, condition == "Anxious" |
               condition == "Severe Anxiety" |
               condition == "Repeated Episodes of Anxiety")
# Remove rows that have missing values
meds <- meds[complete.cases(meds),]
# Check changes
head(meds, 8)
```

```
##      X      drug_name      date      age gender reviewer_type
## 3    2 Sertraline Oral 4/16/2024 25-34 Female      Patient
## 8    7 Sertraline Oral 3/26/2024 7-12   Male       Caregiver
## 21   20 Sertraline Oral 9/17/2023 25-34 Female      Patient
## 25   24 Sertraline Oral 9/5/2023 45-54 Female      Patient
## 29   28 Sertraline Oral 8/2/2023 55-64 Female      Patient
## 37   36 Sertraline Oral 3/20/2023 55-64   Male       Patient
## 56   55 Sertraline Oral 9/15/2022 35-44 Female      Patient
## 67   66 Sertraline Oral 7/13/2022 19-24 Female      Patient
##                                     condition rating_overall rating_effectiveness
## 3 Repeated Episodes of Anxiety                4.3                4
## 8 Repeated Episodes of Anxiety                5.0                5
## 21 Repeated Episodes of Anxiety               5.0                5
## 25 Repeated Episodes of Anxiety               1.0                1
## 29 Repeated Episodes of Anxiety               1.3                2
## 37 Repeated Episodes of Anxiety               1.3                1
## 56 Repeated Episodes of Anxiety               4.3                4
## 67 Repeated Episodes of Anxiety               1.7                2
##      rating_ease_of_use rating_satisfaction
## 3                      4                5
## 8                      5                5
## 21                     5                5
## 25                      1                1
## 29                      1                1
## 37                      2                1
## 56                      5                4
## 67                      2                1
```

```
# Summarize all datasets
head(indicators_anxiety, 8)
```

```
##      Indicator      Group      State
## 1 Symptoms of Anxiety Disorder National Estimate United States
## 2 Symptoms of Anxiety Disorder      By Age United States
## 3 Symptoms of Anxiety Disorder      By Age United States
## 4 Symptoms of Anxiety Disorder      By Age United States
## 5 Symptoms of Anxiety Disorder      By Age United States
## 6 Symptoms of Anxiety Disorder      By Age United States
## 7 Symptoms of Anxiety Disorder      By Age United States
## 8 Symptoms of Anxiety Disorder      By Age United States
##      Subgroup Phase Time.Period      Time.Period.Label
## 1 United States 1.0          1 Apr 23 - May 5, 2020
## 2 18 - 29 years 1.0          1 Apr 23 - May 5, 2020
## 3 30 - 39 years 1.0          1 Apr 23 - May 5, 2020
## 4 40 - 49 years 1.0          1 Apr 23 - May 5, 2020
## 5 50 - 59 years 1.0          1 Apr 23 - May 5, 2020
## 6 60 - 69 years 1.0          1 Apr 23 - May 5, 2020
## 7 70 - 79 years 1.0          1 Apr 23 - May 5, 2020
## 8 80 years and above 1.0        1 Apr 23 - May 5, 2020
##      Time.Period.Start.Date Time.Period.End.Date Value Low.CI High.CI
## 1          04/23/2020          05/05/2020 30.8 30.0 31.7
## 2          04/23/2020          05/05/2020 40.2 38.1 42.4
## 3          04/23/2020          05/05/2020 34.4 32.6 36.1
## 4          04/23/2020          05/05/2020 34.1 32.1 36.2
```

```
## 5      04/23/2020      05/05/2020  31.0  29.0  33.1
## 6      04/23/2020      05/05/2020  24.9  23.6  26.3
## 7      04/23/2020      05/05/2020  16.4  14.8  18.1
## 8      04/23/2020      05/05/2020  14.6  11.5  18.2
## Confidence.Interval Quartile.Range
## 1      30.0 - 31.7
## 2      38.1 - 42.4
## 3      32.6 - 36.1
## 4      32.1 - 36.2
## 5      29.0 - 33.1
## 6      23.6 - 26.3
## 7      14.8 - 18.1
## 8      11.5 - 18.2
```

```
head(global_mental, 8)
```

```
## index      Entity Code Year  anxiety
## 1      0 Afghanistan AFG 1990 4.828830
## 2      1 Afghanistan AFG 1991 4.829740
## 3      2 Afghanistan AFG 1992 4.831108
## 4      3 Afghanistan AFG 1993 4.830864
## 5      4 Afghanistan AFG 1994 4.829423
## 6      5 Afghanistan AFG 1995 4.828337
## 7      6 Afghanistan AFG 1996 4.828083
## 8      7 Afghanistan AFG 1997 4.827726
```

```
head(mental, 8)
```

```
##      Timestamp Gender      Country treatment
## 1 2014-08-27 11:29:31 Female United States      Yes
## 2 2014-08-27 11:31:50 Female United States      Yes
## 3 2014-08-27 11:32:39 Female United States      Yes
## 4 2014-08-27 11:37:59 Female United States      Yes
## 5 2014-08-27 11:43:36 Female United States      Yes
## 6 2014-08-27 11:49:51 Female      Poland      Yes
## 7 2014-08-27 11:51:34 Female      Australia    Yes
## 8 2014-08-27 11:52:41 Female United States      No
```

```
head(prevalence, 8)
```

```
## index      entity      code year      male      female
## 1      0      Abkhazia OWID_ABK 2015      NA      NA
## 2      1 Afghanistan      AFG 1990 3.556843 5.971172
## 3      2 Afghanistan      AFG 1991 3.548885 5.980482
## 4      3 Afghanistan      AFG 1992 3.542779 5.988175
## 5      4 Afghanistan      AFG 1993 3.538304 5.993858
## 6      5 Afghanistan      AFG 1994 3.535309 5.997363
## 7      6 Afghanistan      AFG 1995 3.533797 5.998540
## 8      7 Afghanistan      AFG 1996 3.535415 5.996443
```

```
head(meds, 8)
```

```
##      X      drug_name      date    age gender reviewer_type
## 3    2 Sertraline Oral 4/16/2024 25-34 Female      Patient
## 8    7 Sertraline Oral 3/26/2024 7-12  Male      Caregiver
## 21  20 Sertraline Oral 9/17/2023 25-34 Female      Patient
## 25  24 Sertraline Oral 9/5/2023 45-54 Female      Patient
## 29  28 Sertraline Oral 8/2/2023 55-64 Female      Patient
## 37  36 Sertraline Oral 3/20/2023 55-64  Male      Patient
## 56  55 Sertraline Oral 9/15/2022 35-44 Female      Patient
## 67  66 Sertraline Oral 7/13/2022 19-24 Female      Patient
##                                     condition rating_overall rating_effectiveness
## 3  Repeated Episodes of Anxiety                4.3                4
## 8  Repeated Episodes of Anxiety                5.0                5
## 21 Repeated Episodes of Anxiety                5.0                5
## 25 Repeated Episodes of Anxiety                1.0                1
## 29 Repeated Episodes of Anxiety                1.3                2
## 37 Repeated Episodes of Anxiety                1.3                1
## 56 Repeated Episodes of Anxiety                4.3                4
## 67 Repeated Episodes of Anxiety                1.7                2
##      rating_ease_of_use rating_satisfaction
## 3                      4                  5
## 8                      5                  5
## 21                     5                  5
## 25                     1                  1
## 29                     1                  1
## 37                     2                  1
## 56                     5                  4
## 67                     2                  1
```

Data Importing and Cleaning

The first step to take is going to be importing all of the data from the csv files. After that it will be to parse the data in order to condense the datasets down to the appropriate columns. Once the data has been parsed we will remove rows with missing data to create datasets that have all the necessary columns. The code in the above section prints the kept columns and first eight rows of each dataset.

Exceptions

Unfortunately, the dataset that has the symptoms isn't very useable. I don't believe this first dataset is actually useful for me. There's lots of data but not useful for this particular analysis.

Data Discovery

The data from the indicators dataset seems very promising. It breaks down the data by date, age group, ethnicity, gender, education level, and state. We can group the data by these categories throughout the timeframes that all the categories span. We should then make plots over time for all of these values. We can then look to the global_mental dataset and find the prevalence per country. In that prevalence dataset we can average the prevalence between male and female and compare that to the results from the indicators and global_mental datasets. We can then compare the prevalence by male or female from the indicator groups, and compare that to the mental and prevalence datasets sorted by gender. We should average all the results

for female and male from all of the datasets. We can then look at the treatment dataset from WebMD and see how many people are receiving treatment and what medications they are taking. This can also be broken down by gender for more information. To answer our research questions we look at the prevalence of anxiety per location, gender, race, and age. Then we can look at the effectiveness of medication for each group.

Variables and Data Preparation

I think that the datasets have all the appropriate variables aside from the cumulative female and male results. We can compare all of our results to the NIH and WHO findings. And all of the results can be used in comparisons plots for all of the demographic categories.

Machine Learning

I'm not sure how to incorporate machine learning currently. I've removed some categories from the datasets that could potentially be useful for making models that predict anxiety from symptoms, demographics, and comorbidities. If I could make this model I think machine learning would be useful in order to look at more datasets.

For the Future

I think making models and using machine learning to look at new data would be something I'd really like to do. There are some ethical and privacy concerns with doing this though.

Things to Learn

The first dataset has a text field that I'm unsure how to parse for symptoms in R. The way I would approach this would be to parse the whole text field down into individual words. I'd then compare the words individually to a list of symptoms I defined. Alternatively, I'd probably need AI to parse the text field to find all the symptoms. That may be the better way to search through the text field because doing it myself does not account for if someone says these are symptoms they're not experiencing. The string comparison method only looks for the presence of a word not the context. I'd also like to know how to print nice tables of data both in R and to files from R.