

# U.S. Mental Health Dashboard

ISABEL ARVELO\* AND ANNA PLANTINGA

---

## Abstract

In this paper, we present the U.S. Mental Health Dashboard, an R Shiny web application that facilitates exploratory data analysis of U.S. mental health data collected through national surveys. Mental health affects almost every aspect of people’s lives including their social relationships, substance use, academic success, professional productivity, and physical wellness. Even so, mental illnesses are often perceived as less legitimate or serious than physical diseases, and as a result of this stigmatization, many people suffer in silence without access to proper treatment. To address the lack of accessible healthcare information related to mental illness, the U.S. Mental Health Dashboard presents dynamic visualizations, tables, and choropleth maps of the prevalence and geographic distribution of key mental health metrics based on data from the National Survey on Drug Use and Health (NSDUH) and Behavioral Risk Factor Surveillance System (BRFSS). National and state-level estimates are provided for the civilian, non-institutionalized adult population of the United States as well as within relevant demographic subpopulations. By demonstrating the pervasiveness of mental illness and stark health inequities between demographic groups, this application aims to raise mental health awareness and reduce self-blame and stigmatization, especially for individuals that may inherently be at high risk. The U.S. Mental Health Dashboard has a wide variety of potential use cases: to illustrate to individuals suffering from mental illness and those in close proximity to them that they are not alone, identify subpopulations with the biggest need for mental health care, and help epidemiologists planning studies identify the target population for specific mental illness symptoms.

KEYWORDS AND PHRASES: Mental health, Data visualization, Health inequities.

---

## 1. INTRODUCTION

It is fairly commonplace for people to go to the doctor or confide in their friends and family when they are feeling physically unwell. However, mental illnesses are much less likely to be discussed or treated. According to the Mental Health Million Project [12], 45% of individuals with clinical-level mental health challenges in the United States do not seek professional help. Stigma and lack of affordable treatment options are major obstacles that preclude people from discussing and seeking treatment for mental illness [6].

Existing web applications that visualize healthcare data rely on only one survey or are focused on physical illnesses such as cancer and heart disease [2]. Other more comprehensive platforms aimed towards mental health are cumbersome to navigate or answer a small set of specific questions, instead of providing a more high-level view of the mental health landscape [11, 2, 10]. However, there is currently no easy-to-navigate, broad-scale data visualization web application for those who want to learn more about the prevalence and health disparities associated with common symptoms of mental illness. However, such an app would be useful in a wide range of scenarios: for example, it could help people suffering from mental illness and those who care for them know they are not alone, as well as providing an overview of

the scale and distribution of mental illness for public health practitioners and epidemiologists.

To address this gap, we present the U.S. Mental Health Dashboard, an interactive web application for exploratory data analysis that aggregates mental health statistics from two national surveys. We use data from national surveys run by the Department of Health and Human Services (DHHS) [13] and Centers for Disease Control and Prevention (CDC) [3] to visualize various mental health illnesses and key mental health metrics for adults across the United States. The databases are integrated into an interactive web app that allows users to select response variables of interest, produce dynamic visualizations, tables, and choropleths for response variables of interest, and compare results across different subpopulations.

## 2. METHODS

### 2.1 Contributing Datasets

#### 2.1.1 National Survey on Drug Use and Health (2020)

The National Survey on Drug Use and Health (NSDUH) is a national study run by the DHHS that collects information on several health-related issues including tobacco, alcohol, drug use, and mental health in the United States. The population of interest is the civilian, non-institutionalized population aged 12 or older at the time of the survey, i.e.,

---

\*Corresponding author.

excluding active military personnel, people living in institutional group quarters, and homeless individuals. Our analysis only uses data on adults ages 18 and older from the 2020 survey (27,170 observations).

Participants are selected for the survey using an independent, multistage area probability sample within each state and the District of Columbia. Because geographic identifiers such as state are not included in the 2020 public use file, it is only possible to make national estimates. The weights in this file are approximately the inverse probability of selection for each record.

The response variables of interest from this survey include level of psychological distress over the past 30 days, worst psychological distress over the past 30 days, predicted probability of serious mental health illness, and mental illness category as well as indicators for the past month of serious psychological distress, serious suicidal thoughts, suicidal plans, suicide attempt, lifetime major depressive episode, and major depressive episode in last year. Available demographic and social characteristics include gender, educational status, marital status, age, race, employment status, income, and poverty level.

Due to the COVID-19 pandemic, Quarter 4 of 2020 was the first time that the NSDUH utilized web-based interviewing. However, a concerning high number of adults provided usable information on their substance use but did not complete the mental health or later questions (i.e., “break-offs”). The DHHS created additional analysis weights to analyze the unimputed outcomes starting from the mental health and subsequent sections of the questionnaire to account for respondents who broke off the interview before completing these sections. Unfortunately, the public use file does not contain these “break-off analysis weights,” so estimates from 2020 should not be compared longitudinally to other years of survey data.

### 2.1.2 BRFSS: Behavioral Risk Factor Surveillance System (2021)

The Behavioral Risk Factor Surveillance System (BRFSS) is a nationwide system run by the CDC of health-related telephone surveys that provide state-level information about health-related risk behaviors, chronic health conditions, and the use of preventive services among U.S. residents. The population of interest is the noninstitutionalized adult population (18 years or older) residing in private residences or college housing in the United States or participating areas who have a working cellular telephone. Participants are selected through an overlapping, dual-frame landline and cell phone sample. The BRFSS samples high and medium-density strata to obtain a probability sample of all households with telephones. State health departments may directly collect data from their state residents, or they may use a contractor. Person-level analysis weights involve two components: design-based weights and weight adjustment factors. Design weights reflect probabilities of selection at the sample stage. Weight adjustments are performed

using the generalized exponential model (GEM) developed by Folsom and Singh [7], which calibrated the design-based weights to reduce non-response bias, poststratify to known population control totals, and control for extreme weights when necessary.

The response variables of interest from this survey include an indicator for whether individuals were ever told they had a depressive disorder and a self-assessment of general health. Relevant demographic and social variables include health insurance status, age, education, and income.

## 2.2 R Packages Used

### 2.2.1 survey

The main package we used to specify complex survey designs and produce unbiased summary statistics was *survey: Analysis of Complex Survey Samples* [9], written by Thomas Lumley. This package creates *svydesign* objects that ensure that the design information cannot be separated or used with the wrong data.

The fundamental concept that underlies design-based inference is that an individual sampled with sampling probability  $\pi_i$  represents  $\frac{1}{\pi_i}$  individuals in the population.  $\frac{1}{\pi_i}$  is referred to as the sampling weight. All of the analyses run with this package use the Horvitz-Thompson estimator to estimate the population total [8]. For a sample of size  $n$ , the Horvitz-Thompson estimator  $T_X$  for the population total of  $X$  is

$$\hat{T}_X = \sum_{i=1}^n \frac{1}{\pi_i} X_i = \sum_{i=1}^n \hat{X}_i. \quad (2.1)$$

The variance estimate is

$$\widehat{\text{var}}[\hat{T}_X] = \sum_{i,j} \frac{X_i X_j}{\pi_i \pi_j} - \frac{X_i}{\pi_i} \frac{X_j}{\pi_j}, \quad (2.2)$$

which can also be written as

$$\widehat{\text{var}}[\hat{T}_X] = \frac{N-n}{N} x N^2 x \frac{\text{var}[X]}{n}, \quad (2.3)$$

for a population of size  $N$ . The `svymean()` function used to produce mean estimates in the app is estimated by dividing the estimated total by the population size  $N$ . The variance estimate is estimated by dividing the variance estimate for the total by  $N^2$ . The Horvitz-Thompson estimator of the population size is

$$\sum_{i=1}^n \frac{1}{\pi_i}. \quad (2.4)$$

For estimates in a subpopulation, the survey package handles the computational details of domain estimation and sets the sampling weights to 0 for observations outside of the subpopulation.

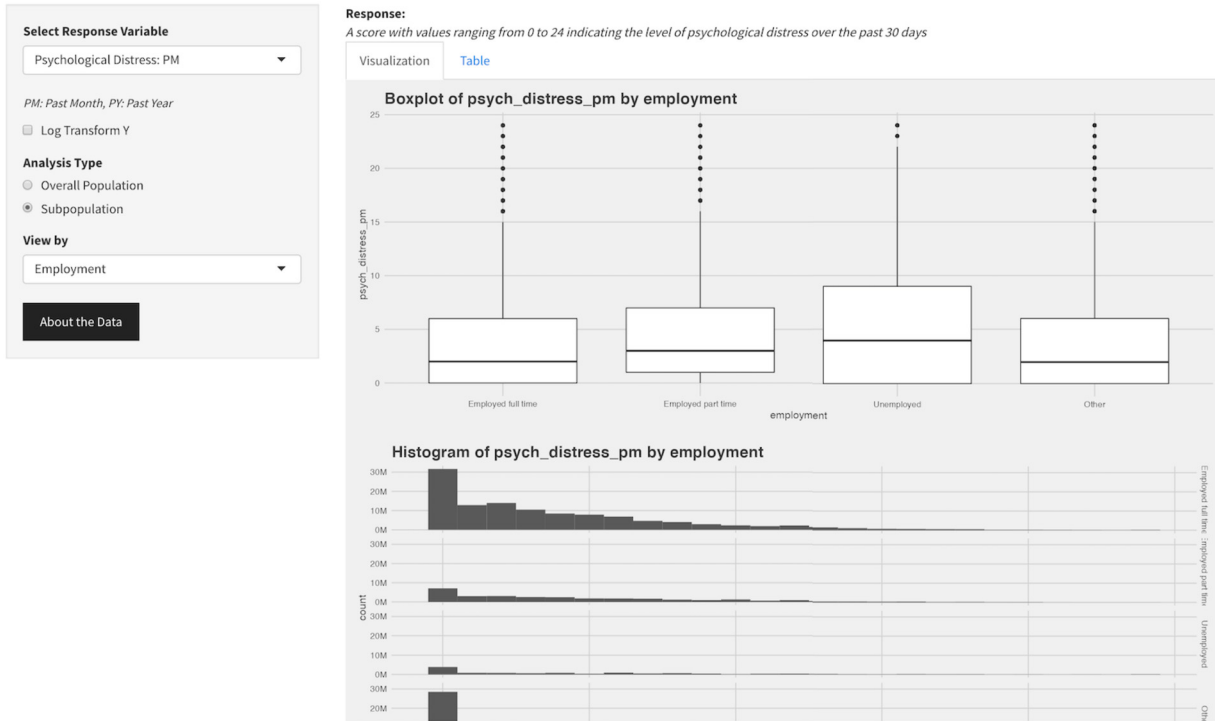


Figure 1: National Dashboard Visualization: Showing distribution of psychological distress within the past month stratified by employment status based on NSDUH survey.

### 2.2.2 ggsurvey

The visualizations on the National tab are implemented with the package, *ggsurvey* [1] that simplifies ggplot2 functions for svydesign objects. The package’s functions call “ggplot2” to make bar charts, histograms, boxplots, and hexplots of survey objects to accurately represent the weighted sample distributions.

### 2.2.3 R shiny

The interactive app is implemented using the *shiny* package [4] and it is deployed on shinyapps.io in its own protected environment with SSL-encrypted access.

## 3. RESULTS

The U.S. Mental Health Dashboard, found at <https://50lulw-isabel-arvelo.shinyapps.io/USMHD/>, has two major tabs: the *National* tab with national-level estimates from the NSDUH survey, and the *State* tab with state-level estimates from the BRFSS survey. Each tab includes a visualization of the response variable of interest as well as a table view to examine specific estimates with more precision.

## 3.1 Major Functions in Shiny App

### 3.1.1 National Level Boxplot, Bar Chart, and Histogram

We begin by visualizing national-level statistics from the NSDUH data set through boxplots and histograms for continuous response variables and bar charts for categorical responses. The visualizations pipe the current input values into the *ggsurvey* functions to dynamically render the histogram and box plots in response to the user’s selected input values. With the log transform Y radio button, the plots remain in the original scale for interpretability, but the breaks in the y-axis for the boxplot and the x-axis for the histogram are spaced according to the log scale. This allows the user to analyze and compare the centers and spreads of the skewed distributions with greater precision.

Figure 1 shows the distribution of psychological distress within the past month without the log transformation.

### 3.1.2 National Level Summary Table

The table of national-level summaries appears on the National tab and includes data on the estimated population mean standard error, and confidence intervals for the population distributions represented in the visualizations. Logistically, we first find the estimated sur-

## U.S. Mental Health Dashboard

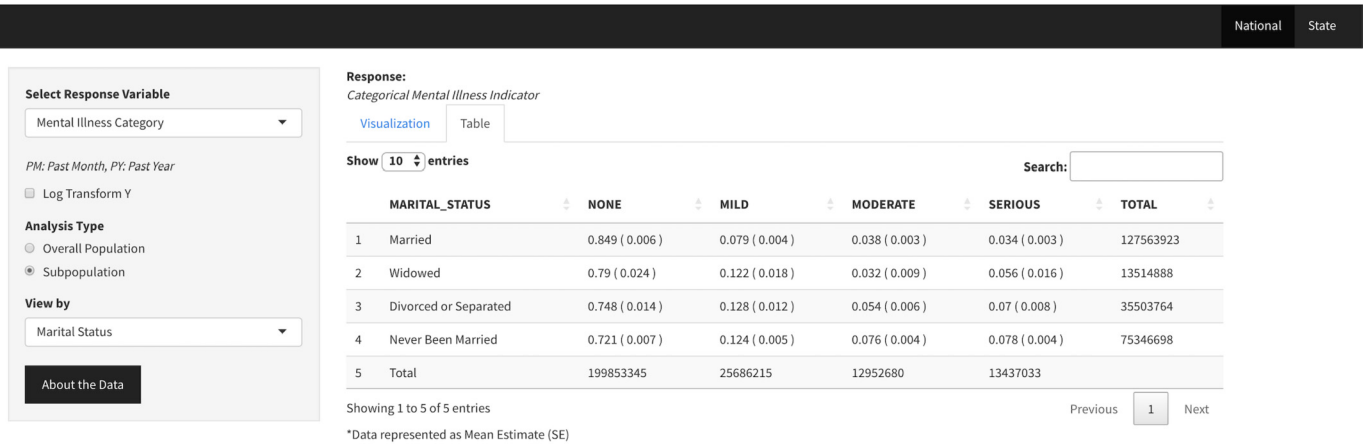


Figure 2: National Dashboard Table: Showing distribution of mental illness category stratified by marital status based on NSDUH survey.

vey mean, standard error, and confidence interval for the response using the Horvitz-Thompson estimators (see Equations (2.1) and (2.2)) as calculated in the *survey* package. The total number of individuals represented in the estimated population for each subgroup is calculated using Equation (2.4), i.e., adding the sampling weights of the individuals in the survey that fit into each category. Formatted data table objects may be sorted by any of the columns, allowing users to order estimates.

Figure 2 shows the table view of the national distribution of the categorical mental illness variable by gender.

### 3.1.3 State-Level Choropleths

The state-level choropleths visually represent the geographic distribution of response variables of interest in the BRFSS across levels of social factors. In order to increase the speed at which the outputs are rendered, we preprocessed the BRFSS survey object to produce lists of tibbles that represent the distribution of the General Health and Depressive Diagnosis variables across health insurance status, education, age, and race (only for General Health). These tibbles are then merged with a simplified shapefile from the *tigris* [14] package that has data on the primary governmental divisions of the 50 states in the United States, as well as the District of Columbia, Puerto Rico, American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, and the U.S. Virgin Islands.

Figure 3 shows the geographic distribution of depressive disorder prevalence by health insurance status.

After the variables are specified, a function is applied over the corresponding tibble to render a choropleth with *leaflet* [5], a JavaScript library used to build web mapping applications. Each state on the leaflet has hover text including the state name, estimated mean of the selected response variable, and estimated standard error.

### 3.1.4 Table of State-Level Summaries

The state table organizes the data visualized in the choropleths into rows that represent the estimated mean of the selected response variable for each state across the different levels of the selected demographic variable (in columns). We combined the reactive list of tibbles produced by the input calls by variable name using a binary function and then spread the rows by the demographic variable to show the state-level mean estimates for the response variables for all states across all levels of the demographic variable.

Figure 4 shows the table view of the state-level estimates of the prevalence of depressive diagnoses as well as the marginal distribution by health insurance status.

## 4. DISCUSSION

In this paper, we have presented an interactive web app for exploratory analysis of public data on the prevalence of mental health disorders and symptoms in the U.S. The mental health crisis adversely affects the quality of life in communities across the United States. The infrastructure and capacity to support individuals struggling with mental illness are lacking, especially in minoritized communities with higher mental health burdens due to systemic unjust policies and practices. Identifying and visualizing which communities and marginalized identities have the highest need for mental health care can help public health professionals allocate resources appropriately and promote solidarity within those groups.

A web app is a useful tool in this scenario because raising general mental health awareness requires reaching a wide audience with varied backgrounds and levels of comfort with quantitative information. Large national surveys such as the NSDUH and the BRFSS include a vast quantity of mental health, social, and demographic data that could be used to

U.S. Mental Health Dashboard

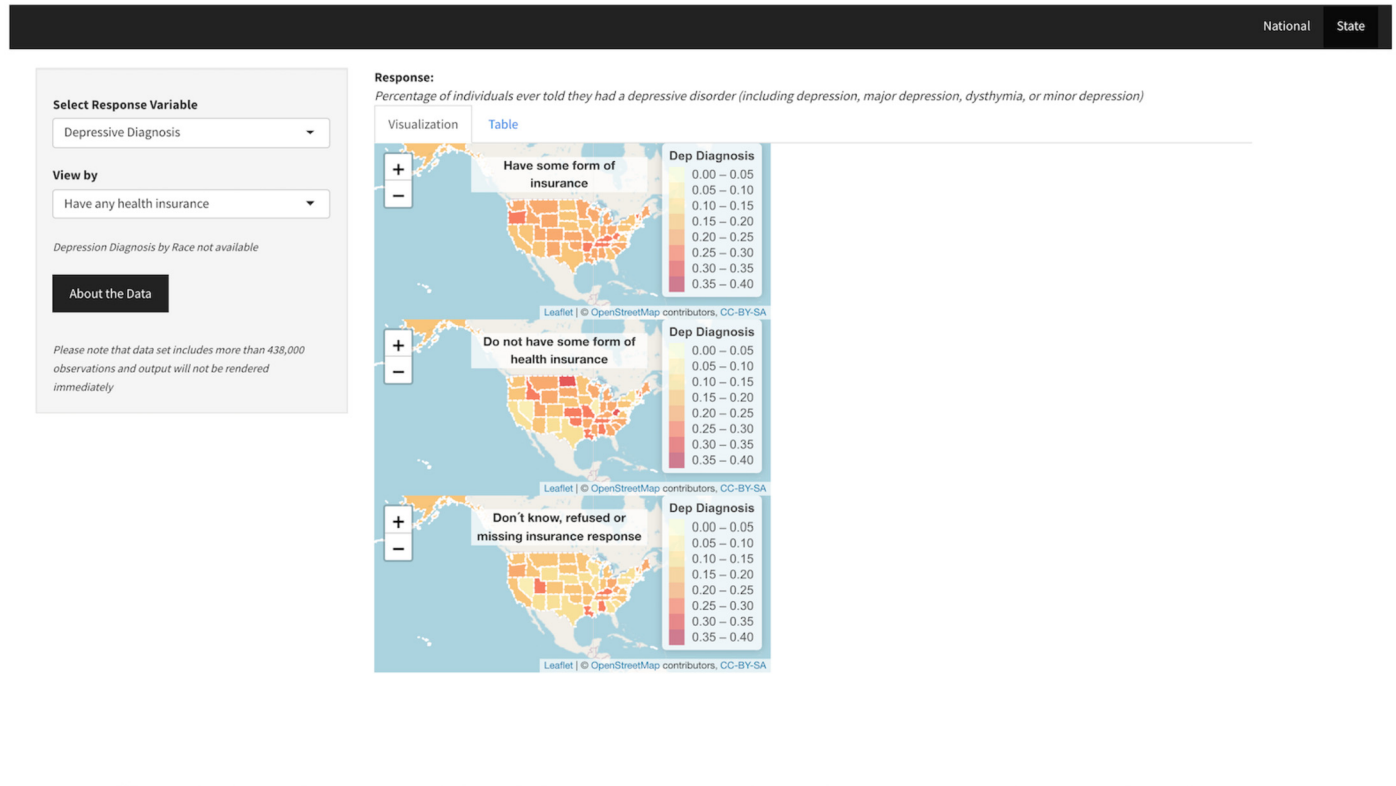


Figure 3: State-level Choropleth: Showing geographic distribution of depressive diagnosis indicator stratified by health insurance status based on BRFSS survey.

answer many different questions, but they require complex survey weighting and analysis approaches that are not accessible to the general public. Instead of trying to anticipate what other stakeholders are interested in learning about, the flexibility of the app allows users to explore specific questions and focus on response variables or demographic factors that are relevant to them. Data visualization gives individuals a clear idea of how to make sense of the information by providing a visual context that makes it easier for audiences to identify and understand trends and patterns.

One of the main aims of this app is to raise awareness and reduce the stigma surrounding mental illness. These figures and visualizations show those suffering that they are not alone and eliminate feelings of self-blame associated with something that often feels uncomfortable or shameful to talk about. By illustrating the pervasiveness of psychological distress, particularly for minoritized populations, we hope to encourage those suffering to seek help and find a support network. Secondly, this app could be among the tools referenced by public health practitioners and epidemiological researchers as they design studies or interventions. High-quality mental health services are unevenly accessible in the United States. The supply of psychiatric residential facility beds, inpatient and outpatient services, and mental health

providers is simply not adequate to meet the demand, especially in communities that need it most. The U.S. Mental Health Dashboard provides quantification and visualization of significant differences in prevalence between different populations to help public health officials prioritize where and how they are investing resources.

In the future, the scope of the U.S. Mental Health Dashboard could be extended by introducing new datasets such as DHHS Mental Health Client-Level Data as well as the CDC National Health Interview Survey for a more comprehensive and robust representation of the prevalence of mental health disorders in the United States. These would allow for the estimation of specific diagnoses of people in mental health treatment facilities and provide more metrics to capture and triangulate the prevalence of anxiety and depression. The app could also be expanded to include treatment options and healthcare coverage available using the National Mental Health Services Survey and Medicare/Medicaid payment and access data to identify the greatest need-to-care gaps so that public health professionals and government officials can prioritize investing resources in these communities. To investigate how structural racism manifests in contemporary health inequities typically assumed to have primarily biological or cultural causes, we could compare the geo-

U.S. Mental Health Dashboard

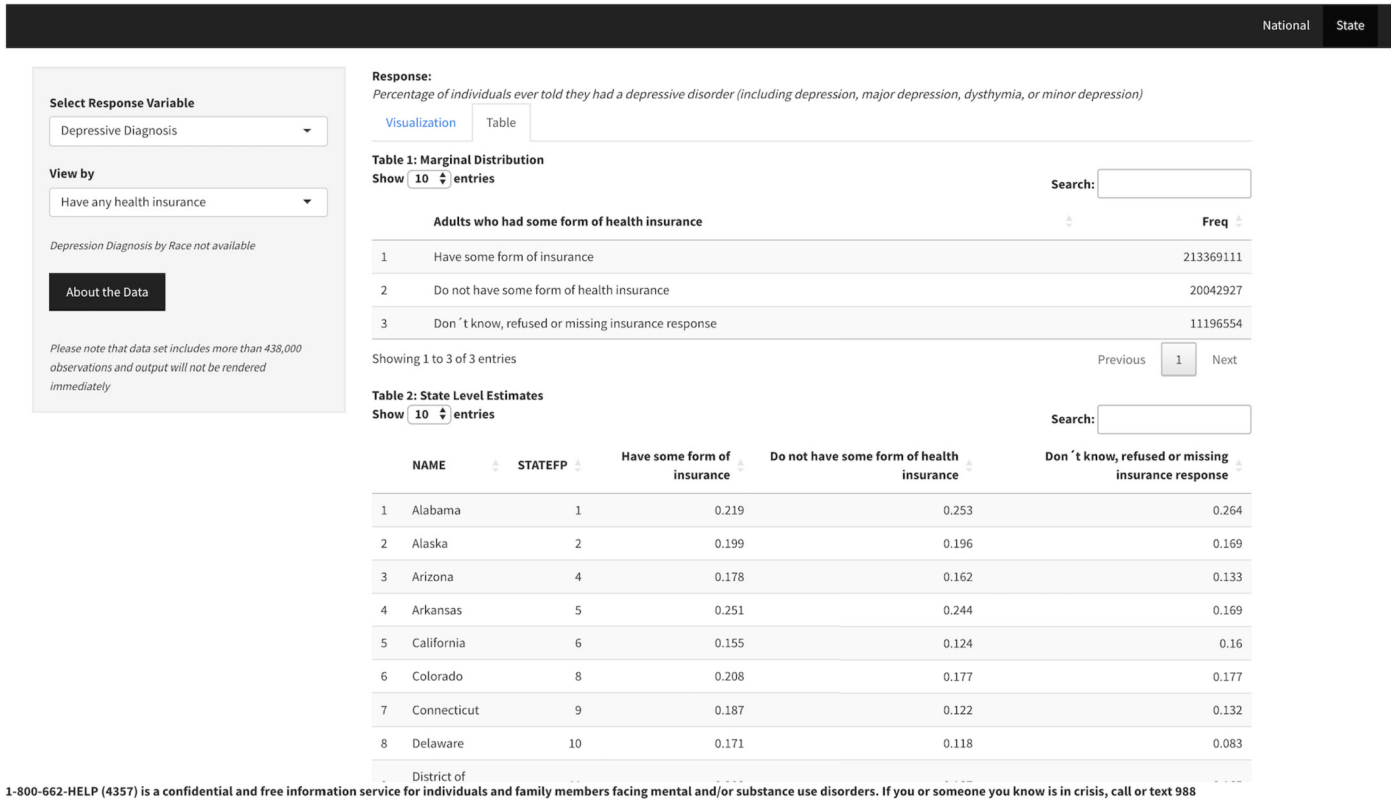


Figure 4: State-level Table: Showing table view of the distribution of depressive diagnosis indicator stratified by health insurance status based on BRFSS survey.

graphic distribution of mental illness prevalence across different subpopulations to historical geospatial data on redlining.

### 5. CONCLUSION

Common misconceptions have caused diseases such as major depressive disorder to be perceived as an indication of personal weakness or fault, and deep-rooted stigmatization and lack of accessible treatments deter individuals from seeking help. However, the ubiquitous feelings of fear and uncertainty that during and since the COVID-19 pandemic left an impression on everyone’s mental well-being. The shared experiences of isolation and trauma associated with a global pandemic have promoted a unique form of solidarity that public health professionals can capitalize on to change how society cares for people suffering from mental health challenges and disorders. Promoting dialogue about these issues is an important start, but awareness must be accompanied by action to turn this opportunity into real change. By providing usable data to policymakers, health-care systems, and healthcare providers, we hope to help guide efforts to improve the mental health and well-being of people and populations.

### ACKNOWLEDGEMENTS

The data and code for the app can be found on <https://github.com/isabelarvelo/U.S.-Mental-Health-Dashboard>.

Accepted 25 October 2023

### REFERENCES

- [1] ALEXANDER, B. (2022). ggssurvey: Simplifying ‘ggplot2’ for Survey Data. R package version 1.0.0. <https://CRAN.R-project.org/package=ggssurvey>.
- [2] CDC-GIS CENTERS FOR DISEASE CONTROL AND PREVENTION (2018). Online Public Health Maps. <https://www.cdc.gov/gis/public-health-maps.htm>.
- [3] CENTER FOR DISEASE CONTROL (CDC) (2021). *Behavioral Risk Factor Surveillance System Survey Data*.
- [4] CHANG, W., CHENG, J., ALLAIRE, J., SIEVERT, C., SCHLOERKE, B., XIE, Y., ALLEN, J., MCPHERSON, J., DIPERT, A. and BORGES, B. (2022). shiny: Web Application Framework for R. R package version 1.7.3. <https://CRAN.R-project.org/package=shiny>.
- [5] CHENG, J., KARAMBELKAR, B. and XIE, Y. (2022). leaflet: Create Interactive Web Maps with the JavaScript ‘Leaflet’ Library. R package version 2.1.1. <https://CRAN.R-project.org/package=leaflet>.
- [6] COOMBS, N. C., MERIWETHER, W. E., CARINGI, J. and NEWCOMER, S. R. (2021). Barriers to healthcare access among US adults with mental health challenges: A population-based study. *SSM-Population Health* **15** 100847.

- [7] FOLSOM, R. E. and SINGH, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the American Statistical Association, Survey Research Methods Section* **598603**.
- [8] LUMLEY, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software* **9** 1–19.
- [9] LUMLEY, T. (2020). *survey: analysis of complex survey samples*. R package version 4.0.
- [10] MENTAL HEALTH AMERICA (2022). *County and State Data Map: Defining Mental Health Across Communities*. Accessed: 2022-09-29.
- [11] MHTTC (2022). Southeast MHTTC Data Visualization Resources Accessed: 2022-09-27.
- [12] NEWSON, J., SUKHOI, O., TAYLOR, J., TOPALO, O. and THIAGARAJAN, T. (March 2021). Mental State of the World 2021. Technical Report, Sapien Labs.
- [13] SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION (SAMHSA) (2020). National Survey on Drug Use and Health 2020.
- [14] WALKER, K. (2022). tigris: Load Census TIGER/Line Shapefiles. R package version 1.6.1. <https://CRAN.R-project.org/package=tigris>.
- Isabel Arvelo. Data Science Institute, Vanderbilt University, United States. E-mail address: [ica1@williams.edu](mailto:ica1@williams.edu)
- Anna Plantinga. Department of Mathematics and Statistics, Williams College, United States. E-mail address: [amp9@williams.edu](mailto:amp9@williams.edu)