Project Report - Mental Health on the Map

Shefali Verma, Tessa Joseph, Anvi Rakshe

Abstract

This project seeks to examine the underlying factors contributing to mental health distress within the American population, focusing on the data from 2022 covering all 50 states. Through the analysis of state-wide health metrics provided by America's Health Rankings, we aim to find patterns shared among states that report high levels of mental health issues. This project also aims to find a way to identify at-risk communities that might be underreporting mental health concerns due to cultural or stigma-related barriers. By grouping communities based on mental health outcomes and correlating factors, such as location and environmental conditions, we aspire to understand the broader influences on mental health.

Introduction

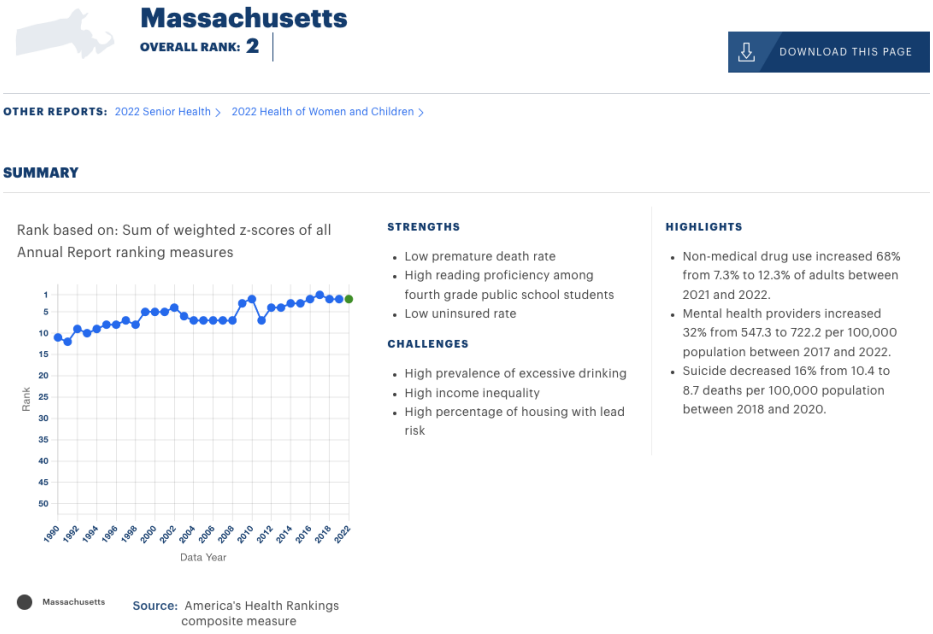
Mental well-being often takes a back seat in health discussions, primarily due to the discomfort and stigma associated with it. Nevertheless, it is necessary to prioritize mental health issues as a critical public health concern, given their status as leading causes of morbidity and mortality globally. While existing interventions such as medications and therapy offer relief to those already grappling with mental health problems, this reactive approach only addresses individual cases.

As medical professionals and researchers pursue cures, there is a notable lack of emphasis on the proactive prevention of mental health challenges at a broader level. Adding to this issue, certain communities may harbor stigmas surrounding mental health, leading to potential underreporting or even completely disregarding cases of mental distress within their population.

Through a comprehensive examination of health and socioeconomic characteristics across American states, coupled with Frequent Mental Distress data, there may be an opportunity to predict the likelihood of mental health distress in communities where underreporting of mental health issues may be a concern.

We ask ourselves, what are the indicators of mental distress within a population? And how can we identify other communities with low positive mental-health outcomes?

We scraped data from [America's Health Rankings.org](). This site offers in-depth health-related data of each state in the United States. It not only describes the States' mental health data, like their average number of poor mental health days per month, or percentage of adults who experience frequent mental distress, but also contains comprehensive data about the State's health behaviors, access to clinical care, social and economic data, and physical environment/infrastructural data.



Massachusetts

OVERALL RANK: 2

DOWNLOAD THIS PAGE

OTHER REPORTS: 2022 Senior Health › 2022 Health of Women and Children ›

SUMMARY

Rank based on: Sum of weighted z-scores of all Annual Report ranking measures

Source: America's Health Rankings composite measure

STRENGTHS

- Low premature death rate
- High reading proficiency among fourth grade public school students
- Low uninsured rate

CHALLENGES

- High prevalence of excessive drinking
- High income inequality
- High percentage of housing with lead risk

HIGHLIGHTS

- Non-medical drug use increased 68% from 7.3% to 12.3% of adults between 2021 and 2022.
- Mental health providers increased 32% from 547.3 to 722.2 per 100,000 population between 2017 and 2022.
- Suicide decreased 16% from 10.4 to 8.7 deaths per 100,000 population between 2018 and 2020.

WEBSITE: mass.gov/orgs/department-of-public-health

Measure	Rating	Value	Rank
Social and Economic Factors *	+++++	0.542	4
Community and Family Safety - Annual *	+++++	0.651	8
Firearm Deaths †	+++++	3.8	2
Occupational Fatalities	+++++	3.3	6
Public Health Funding	+++++	$185	7
Violent Crime	++++	309	18
Economic Resources - Annual *	+++	0.335	23
Economic Hardship Index	+++++	25	4
Crowded Housing †	+++	2.2%	24
Dependency (Ages <18 or >64) †	+++++	36.9%	2
Less Than High School Education †	+++	8.9%	26
Per Capita Income †	+++++	$49,746	1
Poverty †	++++	11.5%	17
Unemployment †	++	6.6%	38
Food Insecurity	+++++	8.4%	9
Homeownership †	+	63.2%	44
Homeownership Racial Disparity †	++	39.2	33

Data Description

Our analysis revolves around a comprehensive dataset that looks at various factors potentially affecting mental health across all 50 U.S. states. The data, sourced from America's Health Rankings Website for the year 2022, includes a wide spectrum of indicators that may have implications on the prevalence of mental health issues at the state level.

For scraping the America's Health Rankings Website:
We use the function: scrape_state_data(state_code), which builds a dataframe for each state based on the state code put through the function. This is the only function we have for scraping the website. The dataframes for each state will then be concatenated and cleaned:

We first use a dictionary with all the states and respective state codes to generate 50 dataframes using the scrape_state_data function. Then we combined all 50 dataframes together to make one comprehensive one. Then we clean the dataframe, getting rid of rows with missing values, making all values numerical, etc.
This is what the a part of the final data frame looks like:

Category	AL	AK	AZ	AR	CA	CO	CT	DE	FL	GA	HI	ID	IL	IN	IA	KS	KY	
Social and Economic Factors	-0.377	-0.286	-0.177	-0.742	-0.126	0.299	0.438	0.265	0.146	-0.051	0.397	-0.035	0.059	-0.067	0.491	-0.006	-0.294	
Community and Family Safety - Annual	-0.398	-0.909	-0.426	-1.070	0.077	0.027	0.729	-0.008	-0.205	-0.299	0.924	0.631	-0.028	-0.534	0.196	-0.538	0.063	
Firearm Deaths	23.400	24.000	17.100	22.500	8.700	15.700	6.100	14.200	13.900	17.700	3.500	18.100	13.900	17.300	11.200	17.100	20.300	
Occupational Fatalities	5.700	8.600	4.000	5.900	3.200	3.600	3.600	4.300	4.500	5.500	4.100	4.300	3.400	6.900	5.600	6.000	5.900	
Public Health Funding ($)	129.000	449.000	79.000	128.000	138.000	127.000	126.000	152.000	79.000	107.000	241.000	172.000	109.000	76.000	161.000	87.000	110.000	1
Violent Crime	454.000	838.000	485.000	672.000	442.000	423.000	182.000	432.000	384.000	400.000	254.000	243.000	426.000	358.000	304.000	425.000	259.000	6

The components of our dataset are:

- Frequent Mental Distress Score: This score is represented as the percentage of adults in the state who reported their mental health was not good for more than 14 days in the past month. We will use this as the primary indicator of mental health in a population.
- Health Distress Indicators: We look at data that reflects the frequency and intensity of mental distress experienced by the population.
- Healthcare Access: Parameters evaluating the availability and quality of mental healthcare services, including the number of mental health providers per capita and the percentage of residents with health insurance coverage.
- Socioeconomic Factors: Data points considering the impact of employment rates, education levels, and income inequality on mental health.
- Lifestyle and Community Metrics: Variables assessing the role of physical activity, diet, substance abuse, and community support in mental health.
- Environmental Influences: Information on factors such as air quality and urbanization that correlate with mental health outcomes.
-

Pipeline Overview

Data retrieval from the America's Health Rankings Website:

- Our `scrape_state_data(state_code)` function builds a dataframe for each state based on the two-letter state code put through the function.

The Big DataFrame: Generation and Concatenation

- Then, we use a dictionary with all the states and respective state codes to generate 50 dataframes using the `scrape_state_data` function. Then we combine the data frames together to make one comprehensive one.
- The third code cleans the dataframe, getting rid of rows with missing values, making all values numerical, etc.

Cleaning the Dataframe:

- We changed missing values to NaN, get rid of rows with too many missing values, made all values numerical by removing “$” and “%” symbols, and removed unnecessary symbols like "†" and "*" from feature names.

Method

We are left with an extremely large dataframe with over 100 health factors which may or may not correlate with frequent mental distress. Our goal is to determine which health factors actually do have a correlation with mental distress and isolate them. Then, run those factors through a machine learning model to be able to predict a frequent mental distress value from the other factors. Finally, we wanted to be able to run example values throughout our model, and receive a fairly accurate predicted Frequent mental distress Score.

Our first step was visualizing the relationship between each factor and frequent mental distress. This would give us a general idea of whether there are factors which are correlated.

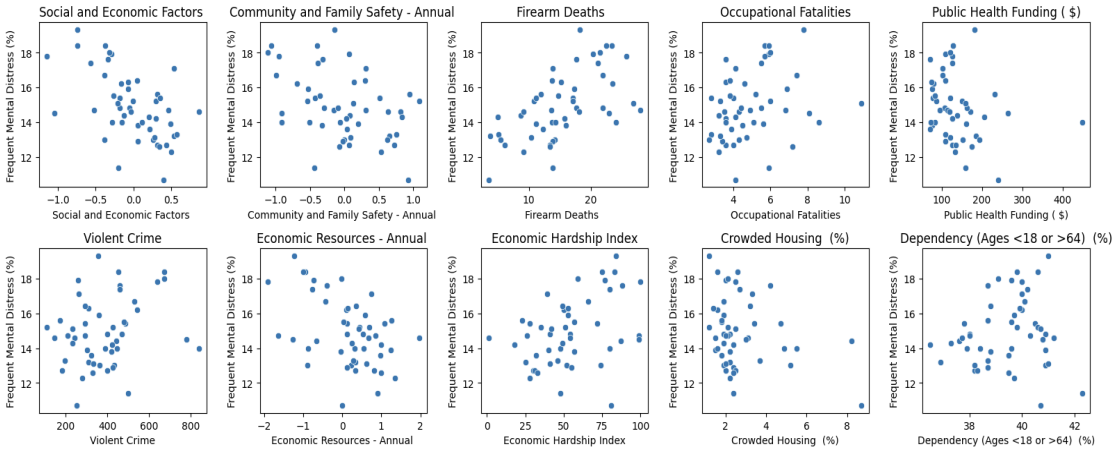


Fig. 1 Sample of Scatter-plots Displaying the Relationship Between all Given Health Metrics and Frequent Mental Distress. See Appendix A for Complete Scatterplot Matrix.

We then constructed a correlation matrix using *scaled* values to determine which health metrics actually have any relationship with Frequent Mental Distress. This prevents any single variable from disproportionately affecting outcomes. We only considered those with a correlation of >|0.50|.

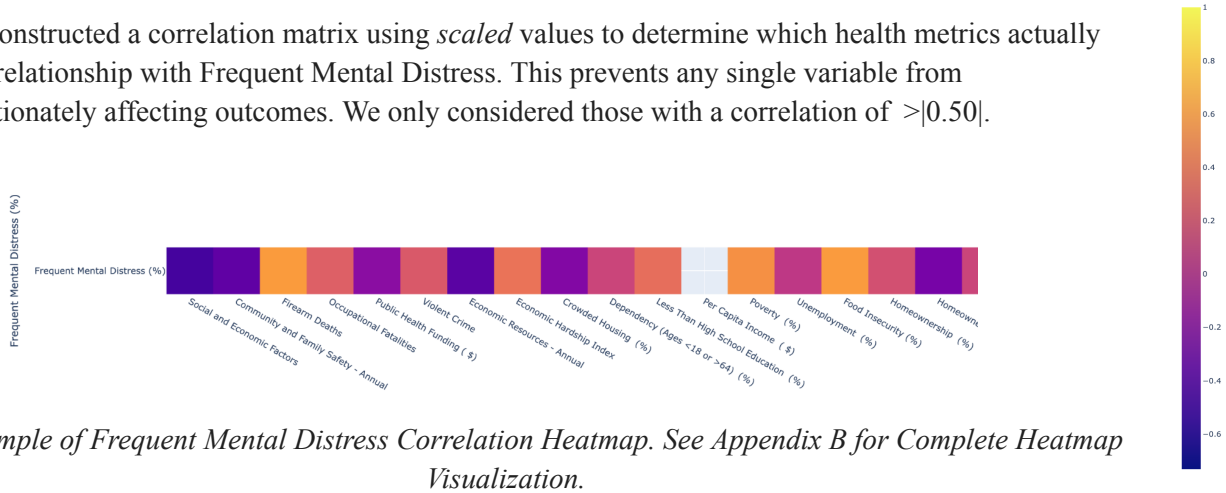


Fig 2. Sample of Frequent Mental Distress Correlation Heatmap. See Appendix B for Complete Heatmap Visualization.

Finally, we used the K-nearest neighbors method to model the relationship between our chosen features and predict mental distress outcomes from their values. Our data has already been scaled, reducing error and ensuring that each metric will contribute proportionally to the prediction model. The K-nearest neighbors method is suitable for our data because K-nearest neighbors takes into account the proximity of data points and by identifying and analyzing patterns among various health metrics, the method will effectively capture the factors that may play a role in mental distress. Our final function predict_mental_distress(selected_feature_values) will take any values for our chosen health metrics, and return a predicted frequent mental distress score. This function explores various scenarios by inputting different combinations of health metric values, enabling an adaptable analysis.

Table 1. True values of State Frequent Mental Distress Scores vs. our Model's Predicted Values

Index	True Values	Predicted Values
IA	13.9	13.80
WA	15.4	13.96
WI	13.6	14.38
VA	14.7	13.76
ME	15.2	15.06
NH	14.6	13.64
NV	17.6	15.72
NE	13.1	13.06
NC	13.8	14.80
MA	13.2	13.12

Discussion

Individual Categories

We found that the health and social metrics which had the highest correlations with Frequent mental distress were:

Social and Economic Factors,Firearm Deaths, percentage of people in poverty, Food insecurity, Social support, Adverse Childhood experiences, Access to clinical care and dental visits, flu vaccinations, at-risk behaviors, teen births, Smoking and Tobacco Usage, e-cigarette usage, general health outcomes, drug usage, physical health and distress, chronic conditions, arthritis, cardiovascular, kidney, and pulmonary diseases, depression, diabetes, and high blood pressure.

While this list may seem fairly obvious, it is necessary to note that the features were included as long they had a correlation with Frequent Mental distress of over 0.5. And while they may have a correlation with mental distress, there is no conclusion for causation.

Machine Learning Model

Our model had an R-squared value of 0.71, and a Mean Squared Error of 0.8445. The mean squared is fairly low, indicating a good fit of the model. The R^2 value being 0.71 indicates a fair fit, however considering the number of variables factored in, the R^2 score may be acceptable.

We are also able to see the scores model predicted for known values of mental distress, which are fairly similar to the real values.

Testing the model

We input arbitrary feature values and received an appropriate predicted Frequent mental distress score, which answers our original question: how can we identify other communities with low positive mental-health outcomes?

By comparing the model's predicted output to the measures of central tendency and standard deviations of the states' Frequent mental distress scores, we can gain an understanding of whether the tested population has relatively better or worse mental health outcomes than the American average and by how much.

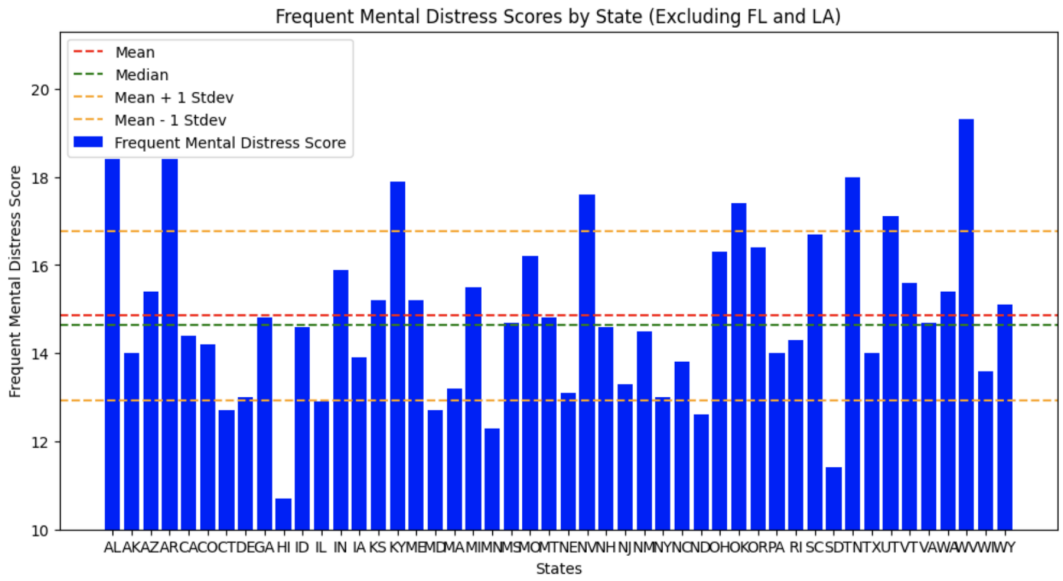


Fig. 3. Chart displaying Frequent Mental Distress scores of all states excluding Florida and Louisiana (no reported Frequent Mental Distress scores). The mean, median and standard deviations of the states Frequent Mental Distress Scores are shown with red, green and yellow dashed lines, respectively.

It is important to keep in mind however, that this data is based on American populations. It may be true that people in other countries and cultures may experience mental health issues due to completely different factors (ongoing wars, better weather, cultural behavior, etc). This predictive model may only be fit for those communities who experience mental distress in a similar way to Americans.

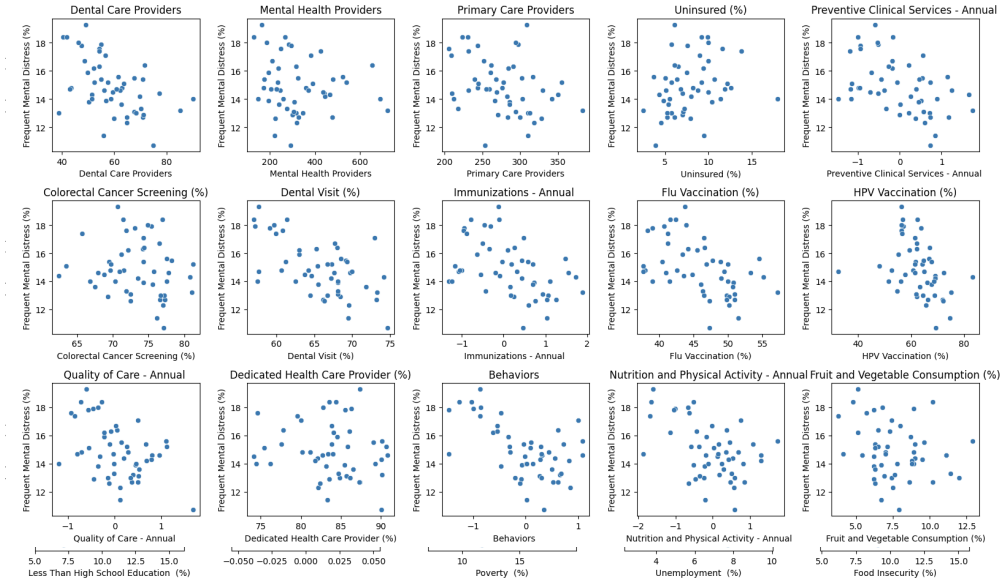
Takeaway

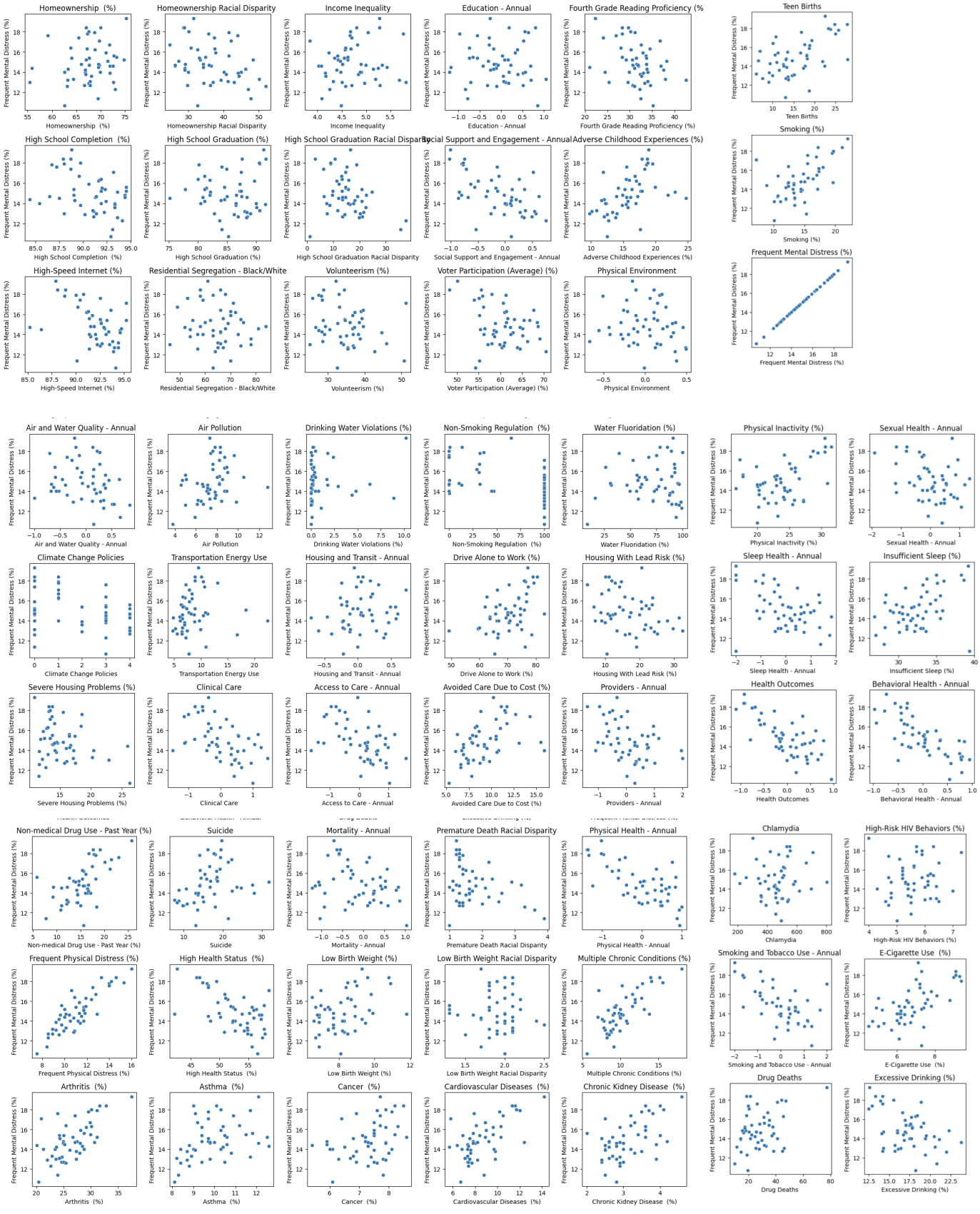
Our team remains optimistic about the potential insights that could be gained from our analysis. The visualizations, including the correlation heatmap and the scatter plots, suggest that while some relationships between mental health distress and various socioeconomic and health-related factors are evident, the complexity of mental health issues demands an approach that considers many elements. Since we have many factors that are in play, more advanced techniques, such as Random Forests, could provide a deeper understanding of the most influential factors so that we may further facilitate the identification of commonalities among affected groups and foster the development of targeted public health strategies.

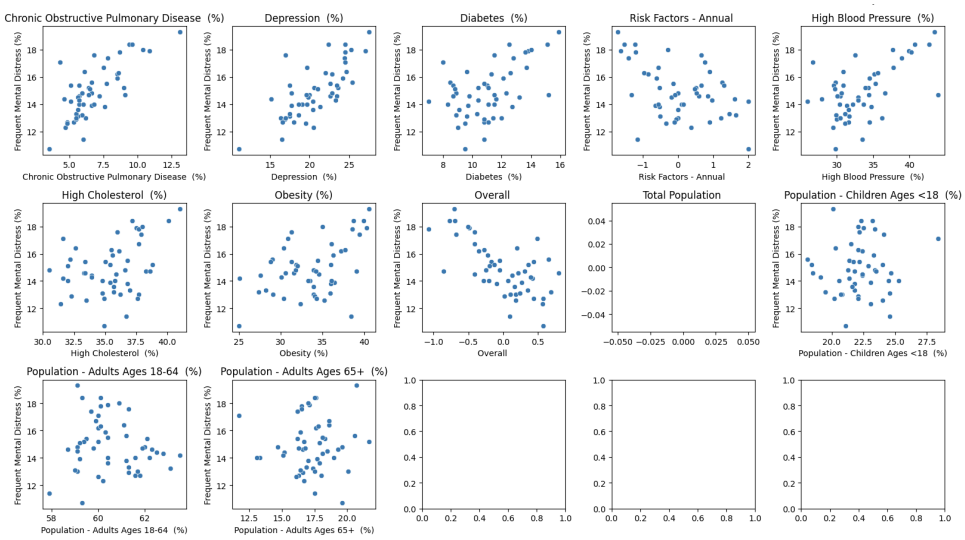
References

U.S. Census Bureau. (n.d.). 2022 Annual Report. America's Health Rankings. https://www.americashealthrankings.org/

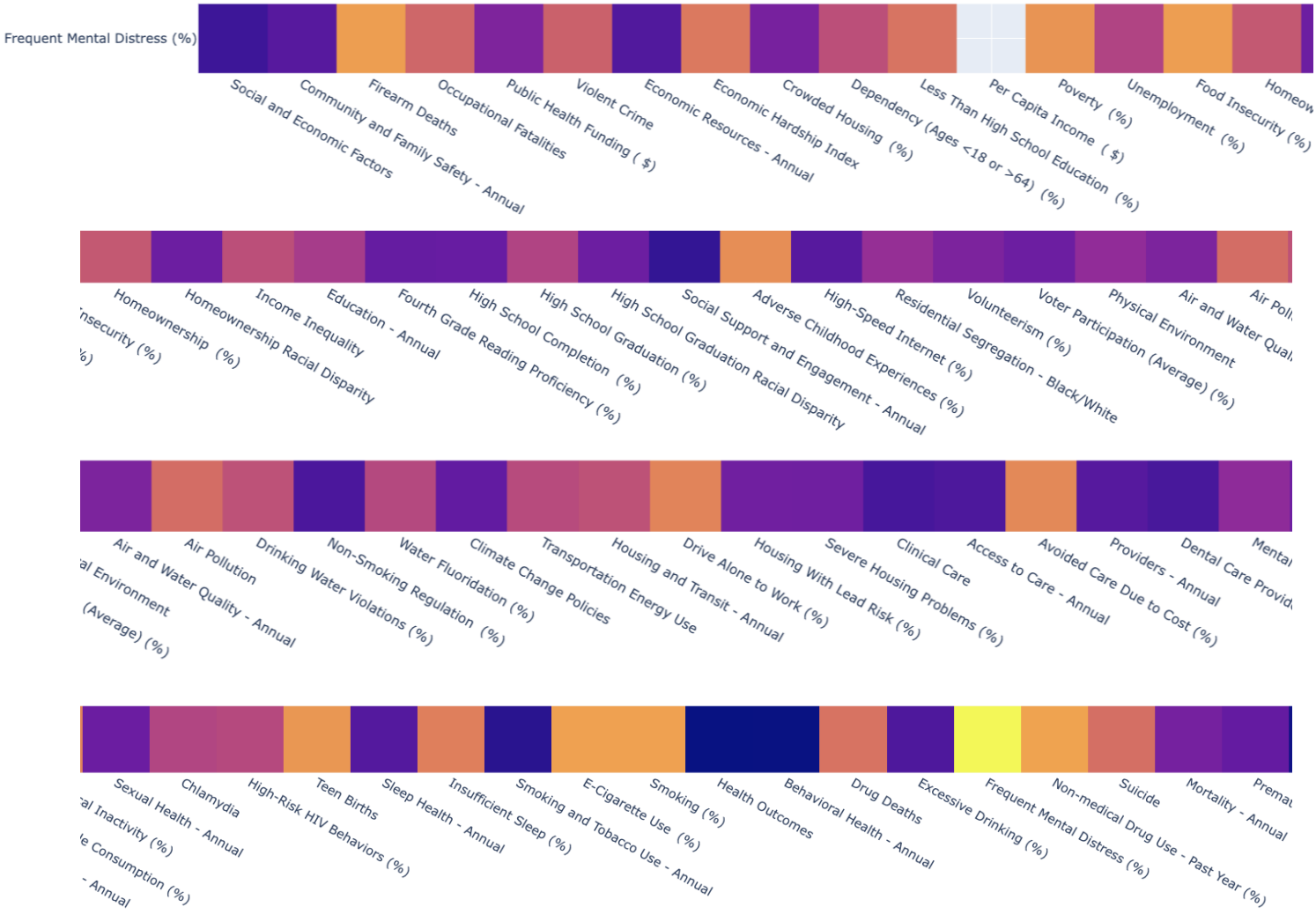
Appendix A.




Homeownership (%)
Homeownership Racial Disparity
Income Inequality
Education - Annual
Fourth Grade Reading Proficiency (%
Teen Births
High School Completion (%)
High School Graduation (%)
High School Graduation Racial Disparity
Social Support and Engagement - Annual
Adverse Childhood Experiences (%)
Smoking (%)
High-Speed Internet (%)
Residential Segregation - Black/White
Volunteerism (%)
Voter Participation (Average) (%)
Physical Environment
Frequent Mental Distress (%)
Air and Water Quality - Annual
Air Pollution
Drinking Water Violations (%)
Non-Smoking Regulation (%)
Water Fluoridation (%)
Physical Inactivity (%)
Sexual Health - Annual
Climate Change Policies
Transportation Energy Use
Housing and Transit - Annual
Drive Alone to Work (%)
Housing With Lead Risk (%)
Sleep Health - Annual
Insufficient Sleep (%)
Severe Housing Problems (%)
Clinical Care
Access to Care - Annual
Avoided Care Due to Cost (%)
Providers - Annual
Health Outcomes
Behavioral Health - Annual
Non-medical Drug Use - Past Year (%)
Suicide
Mortality - Annual
Premature Death Racial Disparity
Physical Health - Annual
Chlamydia
High-Risk HIV Behaviors (%)
Frequent Physical Distress (%)
High Health Status (%)
Low Birth Weight (%)
Low Birth Weight Racial Disparity
Multiple Chronic Conditions (%)
Smoking and Tobacco Use - Annual
E-Cigarette Use (%)
Arthritis (%)
Asthma (%)
Cancer (%)
Cardiovascular Diseases (%)
Chronic Kidney Disease (%)
Drug Deaths
Excessive Drinking (%)


Chronic Obstructive Pulmonary Disease (%)
Depression (%)
Diabetes (%)
Risk Factors - Annual
High Blood Pressure (%)
High Cholesterol (%)
Obesity (%)
Overall
Total Population
Population - Children Ages <18 (%)
Population - Adults Ages 18-64 (%)
Population - Adults Ages 65+ (%)
Frequent Mental Distress (%)

Appendix B.


Frequent Mental Distress (%)
Social and Economic Factors
Community and Family Safety - Annual
Firearm Deaths
Occupational Fatalities
Public Health Funding ($)
Violent Crime
Economic Resources - Annual
Economic Hardship Index
Crowded Housing (%)
Dependency (Ages <18 or >64) (%)
Less Than High School Education (%)
Per Capita Income ($)
Poverty (%)
Unemployment (%)
Food Insecurity (%)
Homeownership (%)
Homeownership Racial Disparity
Income Inequality
Education - Annual
Fourth Grade Reading Proficiency (%)
High School Completion (%)
High School Graduation (%)
High School Graduation Racial Disparity
Social Support and Engagement - Annual
Adverse Childhood Experiences (%)
High-Speed Internet (%)
Residential Segregation - Black/White
Volunteerism (%)
Voter Participation (Average) (%)
Physical Environment
Air and Water Quality - Annual
Air Pollution
Drinking Water Violations (%)
Non-Smoking Regulation (%)
Water Fluoridation (%)
Climate Change Policies
Transportation Energy Use
Housing and Transit - Annual
Drive Alone to Work (%)
Housing With Lead Risk (%)
Severe Housing Problems (%)
Clinical Care
Access to Care - Annual
Avoided Care Due to Cost (%)
Providers - Annual
Dental Care Provid
Sexual Health - Annual
Chlamydia
High-Risk HIV Behaviors (%)
Teen Births
Sleep Health - Annual
Insufficient Sleep (%)
Smoking and Tobacco Use - Annual
E-Cigarette Use (%)
Smoking (%)
Health Outcomes
Behavioral Health - Annual
Drug Deaths
Excessive Drinking (%)
Frequent Mental Distress (%)
Non-medical Drug Use - Past Year (%)
Suicide
Mortality - Annual

APPENDIX B. (cont'd)