Untangling the Threads of Health, Social Context, and Economic Factors in Suicide Rate

Lisa Wang

1007652746

April 24, 2024

Instructor: Prof. Khazra

University of Toronto

1 Introduction

According to the World Health Organization (WHO), more than 700,000 people take their own lives every year, with countless others attempting suicide. Each act of suicide profoundly impacts families, communities, and countries, leaving enduring scars on those left behind. Suicide spans all ages and was the fourth leading cause of death among 15-29-year-olds globally in 2019 [9]. Although mental health conditions like depression and anxiety are commonly linked to suicide, it is a complex issue with biological, psychological, social, and environmental roots.

Reflecting this complexity, an ecological study was performed to analyze suicide rates from 2001 with healthcare access in the US [8]. Their multivariate models found positive bivariate associations with age-adjusted suicide rates and Native American ethnicity and a higher proportion of uninsured residents. Negative bivariate associations were observed between the suicide rate and factors such as per capita income, the higher population density of physicians, higher federal aid for mental health, and a higher proportion of African Americans. Their findings support the view that clinical intervention is a crucial element in the prevention of suicide.

Beyond demographics and healthcare access, a 2015 study by Machado DB, Rasella D, and dos Santos DN [5] revealed a positive correlation between the suicide rate in Brazil and income inequality, as measured by the Gini Index. This connection underscores income disparity as a community-level risk factor influencing suicide rates. In the United States, an analysis of data spanning from 1999 to 2018 noted a concerning 35 percent rise in the age-adjusted suicide rate, with marked discrepancies across different genders, ages, and levels of urbanization [3]. Such trends highlight the critical need for interventions tailored to specific at-risk demographics.

Previous studies have shown that the suicide rate is associated with a wide range of factors spanning demographics, healthcare access, and socioeconomic status. This study delves into US suicide rates, using more recent data from the 2020 Community Health Rankings to probe deeper into the relationship between suicide and a combination of demographics, healthcare, and socioeconomic factors. This includes examining mental health service accessibility, sleep patterns, the percentage of the population uninsured, per capita incomes, and racial compositions. Additionally, it incorporates factors such as food resource access and unemployment, allowing for a more accurate estimation of the association between these factors and the US suicide rate. Based on this background, the initial hypothesis is that a combination of lower socioeconomic status, poor access to mental health services, and higher unemployment rates is associated with higher suicide rates.

Preliminary exploratory data analysis has demonstrated a surprisingly minimal association between suicide rates and the availability of mental health providers. Moreover, it highlights a negative relationship between suicide rates and factors such as per capita income and access to healthy food. The investigation further reveals stark regional variations in these parameters across the United States, pointing to a broader narrative of income inequality and resource access disparity. By building multivariate linear regression models, estimates of each factor's association with the suicide rate were obtained. The model also incorporates interactions of unemployment and demographic factors, aiming to understand the predictors' collective impact on the US suicide rate.

By employing machine learning techniques such as building regression trees and Random Forest models, the research has further quantified the importance of each variable, revealing the significant impact of racial composition and socioeconomic status on suicide rates. This study aims to leverage insights from robust analytical methods to pinpoint the determinants of suicide rates and catalyze improvements in mental health resources. In doing so, it seeks to contribute to the broader effort to reduce suicide rates and steer the formulation of evidence-based preventative policies to tackle this public health challenge.

2 Context and Data

To explore the complex relationship between demographics, socioeconomic factors, healthcare access, and their association with suicide rates in the United States, this study utilized the 2020 Community Health Ranking panel data, which records measurements at the county level. Additionally, daily weather data from the National Oceanic and Atmospheric Administration (NOAA) Global Summary of the Day (GSOD) were merged with the health ranking data. To prepare the dataset for regression analysis, a data-cleaning process was performed that involved removing unnecessary variables, addressing missing values, and renaming variables.

2.1 Summary of Important Variables

`suicide_rate`: The dependent variable is the age-adjusted suicide rate in the US measured in 2020.

`mental_health_provider_rate` (independent variable): Records the number of mental health providers per 100,000 population.

The availability and accessibility of mental health providers play a crucial role in influencing suicide rates. A shortage of mental health professionals can result in inadequate support for individuals grappling with mental health challenges, leaving them without the necessary interventions and

`percent_insufficient_sleep` (independent variable): Measures the percentage of adults who report that they sleep less than 7 hours per night on average.

Sleep plays a fundamental role in maintaining mental well-being, and insufficient sleep is likely associated with an increased risk of mental health disorders, including depression and anxiety. Consistent lack of sleep can contribute to elevated stress levels, impaired cognitive function, and emotional instability. The cognitive and emotional consequences of insufficient sleep may compromise an individual's ability to cope with stressors and navigate daily life, potentially increasing susceptibility to suicidal thoughts and behaviours. In analyzing factors associated with suicide rates and seeking interventions to reduce them, including a measure of the percentage of adults experiencing insufficient sleep allows for exploration of the impact of inadequate sleep and the potential mental health problems it may induce on suicide rates.

`pct_limited_healthfood_access` (independent variable): Measures the percentage of the population that is low-income and does not live close to a grocery store.

Low-income individuals residing in areas without convenient access to grocery stores face obstacles in maintaining a healthy diet, as they may rely on less nutritious and more readily available but often less healthy food options. The stressors associated with economic hardship and limited access to nutritious food can compound, potentially increasing the vulnerability of individuals facing these living conditions to mental health challenges, thus influencing suicide risk. Additionally, the lack of accessible grocery stores may contribute to social and economic isolation. These individuals are likely to encounter more challenges and struggle with their basic needs, which can deplete their mental resources, potentially adding more stress to daily life compared to others, further impacting mental health. Including this measurement will explore the relationship between this variable and the suicide rate. It may further support the analysis of how this measurement differs across states and counties, helping me incorporate regional differences into my analysis.

`per_capita_income` (independent variable): Measures the 2020 average income earned per person in US dollars.

People's income is another key factor potentially linked to suicide rates through various channels, including financial strain, stress, access to mental and physical health services as well as educational and employment opportunities.

Low income often correlates with financial strain, job insecurity, and economic hardship. Persistent financial stressors can contribute to elevated levels of psychological distress, anxiety, and depression, all of which are risk factors for suicide.

Higher-income individuals generally have more opportunities for greater access to mental health services, such as counselling and psychotherapy. Conversely, those with lower incomes may face barriers to accessing these critical resources, leading to undiagnosed or untreated mental health conditions.

Moreover, income influences access to education and employment opportunities. Limited access to quality education and employment prospects can contribute to feelings of hopelessness and despair, increasing the risk of suicidal ideation.

The aforementioned suggests that the lower-income group faces more difficulties and is more susceptible to mental health issues, potentially leading to a higher likelihood of suicide. However, it's also plausible that individuals with lower incomes may find ways to cope and reduce mental depletion due to these factors. On the other hand, wealthier individuals may face challenges in managing their wealth and encounter other struggles, increasing their susceptibility to stress, anxiety, or other mental problems that may lead to suicide. The uncertainty regarding the possible association between income and suicide rate underscores the importance of including this variable to explore and better understand the relationship.

`percent_unemployed` (independent variable): This variable measures the percentage of the population that is without employment. Economic despair, often a consequence of unemployment, is a significant stressor that can escalate the risk of mental health issues and consequently, may lead to an elevated suicide rate. Unemployment reflects not just personal economic difficulty but also broader economic conditions, which necessitates its consideration in analyzing regional disparities in suicide rates.

`percent_uninsured` (independent variable): Access to healthcare is crucial for the treatment of mental health conditions. The lack of health insurance may impede individuals from seeking help for mental health issues, possibly exacerbating risks associated with suicidal behaviour.

`num_primary_care_physicians` (independent variable): The availability of primary care physicians is an indicator of access to initial health consultation and ongoing health management. A higher number of primary care physicians may correlate with better mental health outcomes and potentially lower suicide rates.

`demographic_composition_racial_groups` (independent variable): Demographic factors, including racial and ethnic backgrounds, can influence the prevalence of mental health conditions. Different groups may have varying levels of susceptibility to suicide due to cultural, socioeconomic, and biological factors, as well as differing access to mental health resources.

2.2 Dealing with Missing Values

There are 584,204 observations in the filtered dataset, and approximately 1.1% of the total observations (6,493 observations) are missing. Handling missing data can be approached in various ways, such as deletion or imputation. Considering the diverse nature of counties in the US, which vary in population size, health, education metrics, geographic diversities, and many unobserved factors, imputing the missing suicide rate using mean values or values from other counties may introduce bias if these imputed values do not accurately represent the true suicide rate at the county level. To ensure accurate future analysis, the study drops the missing observations assuming that they are missing completely at random, resulting in a cleaned dataset with 577,711 observations in total.

2.3 Summary Statistics Table

Summary Statistics for Key Variables

	mean	std	min	25%	50%	75%	max
suicide_rate	18.106670	7.099866	5.115770	13.632098	17.042262	20.943218	118.974198
mental_health_provider_rate	173.111809	161.876042	4.121160	63.102100	128.817270	230.031820	2123.029710
percent_insufficient_sleep	33.555299	3.736848	23.028348	30.982083	33.499922	36.230907	46.707783
pct_limited_healthfood_access	7.121382	5.431051	0.000000	3.555368	6.098241	9.380374	71.844209
per_capita_income	25326.094053	6031.272948	9286.000000	21298.000000	24350.000000	28089.000000	64746.000000
percent_unemployed_CDC	7.376378	2.722742	1.100000	5.500000	7.000000	8.800000	29.900000
percent_uninsured	10.843639	4.791373	2.262724	7.003586	9.954634	13.641932	31.207924
num_primary_care_physicians	113.515557	363.406736	0.000000	9.000000	21.000000	69.000000	7397.000000
percent_black	9.011645	12.888742	0.127479	0.957137	3.184433	11.236182	73.989419
percent_american_indian_alaska_native	2.043551	6.630184	0.093411	0.372695	0.605010	1.204601	92.515200
percent_asian	1.827316	3.129565	0.049116	0.544709	0.873171	1.734573	42.952310
percent_native_hawaiian_other_pacific_islander	0.139783	0.492440	0.000000	0.038956	0.067540	0.124953	12.974729
percent_hispanic	9.519534	12.954145	0.610451	2.518589	4.701351	10.294709	96.359551
percent_non_hispanic_white	76.105154	18.835014	2.691288	64.981958	82.098666	91.288416	97.887219
population_density_per_sqmi	267.507415	1047.570924	0.711831	34.416533	69.124519	172.970039	28069.675983

Table 1: Summary Statistics Table

The summary statistic table above contains the summary statistics for the original data and the newly merged variables. The key observations are:

- The mean suicide rate within the US is around 18.11. The standard deviation is approximately 7.10, showing a 39.21% variation around the mean rate. The lowest suicide rate recorded is 5.12, while the highest reaches 118.97, signifying a considerable dispersion in suicide rate within the country.

- The descriptive statistics for the mental health provider rate variable reveal an average rate of approximately 173.11, with a standard deviation of 161.88, nearly as large as the mean value, indicating a high variation of 93.5% around the mean. The minimum rate is 4.12, while the maximum rate reaches 2,123.03. Both the standard deviation and rate range suggest significant variation in mental health provider rates across the US. Given the likely association between access to mental health programs and the suicide rate, this implies that the dispersion in mental health provider rates might explain the observed large variation in the suicide rate.

- Approximately 33.6% of individuals in the US experience less than 7 hours of sleep per night on average, highlighting a significant portion of the US population facing potential sleep challenges. The standard deviation is around 3.74%, suggesting an 11.1% deviation from the mean. This emphasizes the heterogeneity in sleep habits. This percentage ranges from 23.03% to 46.71%. Exploring the factors contributing to this range and the association between this and the suicide rate can reveal regional, demographic, or lifestyle influences on sleep habits.

- The mean value for the percentage of the population with limited access to healthy food is around 7.1%, indicating that an average of 7.1% of the population is in the low-income group and does not live close to a grocery store. The standard deviation is around 5.4%, showing a relatively high deviation of 79.1 percent from the mean. This percentage measurement also has a wide range, with the minimum percentage being 0% and it goes up to 71.84%. The large dispersion observed suggests significant income inequality across the US. Areas with limited access to healthy food may be disproportionately populated by individuals with lower incomes.

- In 2020, the mean per capita income for the US population was approximately \$25,326.09, with a standard deviation of \$6,031.27. This deviation suggests an approximate 23.9% variation around the mean. The recorded range of monthly wages spans from a minimum of \$9,286 to a maximum of \$64,746, indicating a considerable dispersion in individuals' yearly income. Since one's income level is linked to their ability to access physical and mental healthcare, quality education, and other opportunities. The observed differences in individuals' income may underscore social disparities and economic inequalities within the population.

- The mean unemployment rate in the data is approximately 7.38%, with a relatively moderate standard deviation of about 2.72, indicating that while there is some variation in unemployment rates across the dataset, it is not extremely wide. The range, however, is quite broad, with the lowest observed unemployment rate near 1.10% and the highest near 29.90%, suggesting that there are areas with very low unemployment as well as areas with extremely high unemployment. The median unemployment rate is 7%, which is slightly lower than the mean, indicating a slight skew in the distribution towards higher values.

- On average, 10.84% of the population in the dataset is uninsured, with a standard deviation of about 4.79, which is quite high relative to the mean. This indicates a considerable disparity in insurance coverage among different areas. The minimum percentage of uninsured individuals is around 2.26%, while the maximum is notably high at 31.21%. The data is somewhat right-skewed, as the median (9.95%) is below the mean, and the 75th percentile is substantially higher than the median, indicating a significant proportion of the data has a higher percentage of uninsured individuals.

- There is a large mean number of primary care physicians (approximately 113.52), but the standard deviation is very high (around 363.41), pointing to a wide variation in the availability of primary care physicians across different areas. The distribution is likely highly right-skewed, considering the median is only 21 compared to a mean that is over five times higher and with a maximum value of 7397. This suggests that while some areas may have an abundance of primary care physicians, many areas have far fewer than the average.

2.4 Data Visdualization

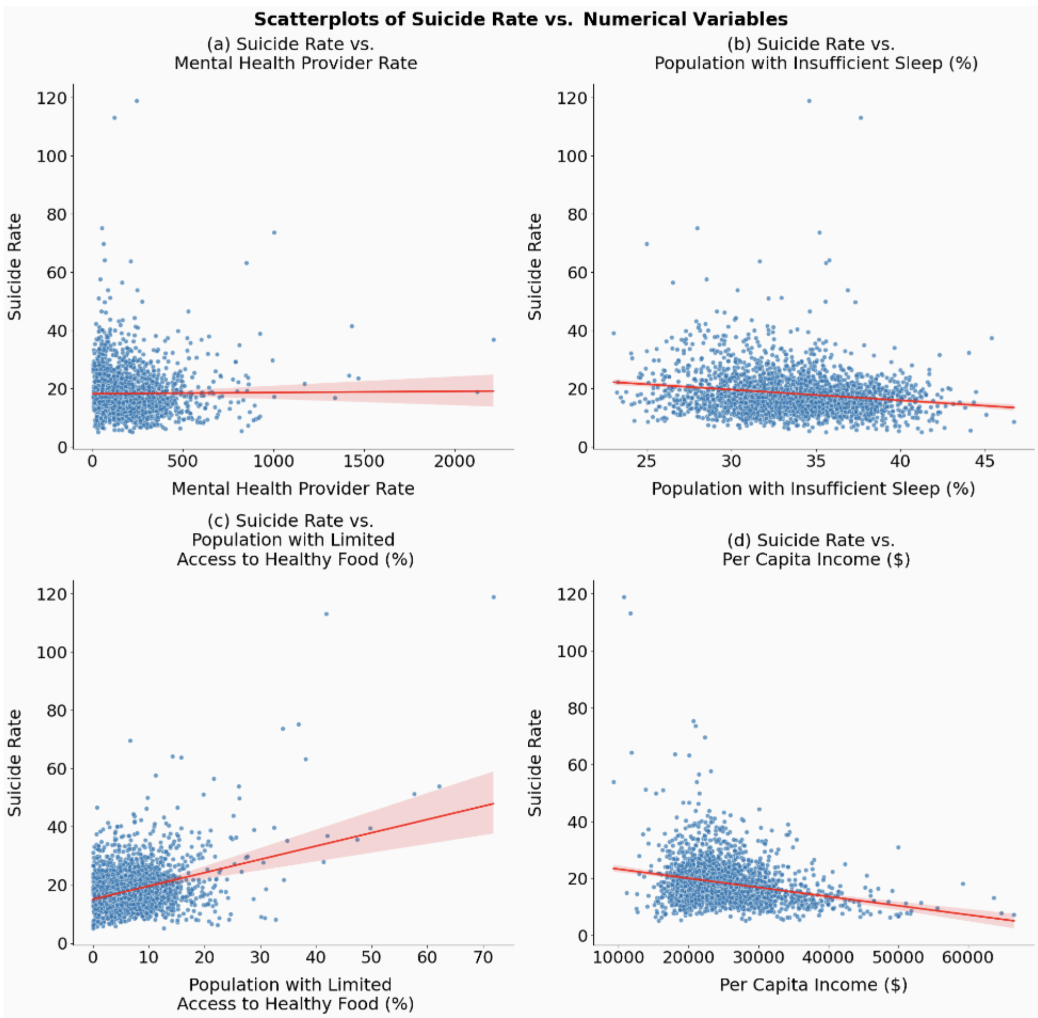

Figure 1: Scatterplots of Suicide Rate vs. Numerical Variables

`Scatterplot Analysis:` The scatterplot illustrates the potential relationship between suicide rates and key variables mental health provider rate, population with insufficient sleep, population with limited access to healthy food and per capita income.

Panel (a) displays the relationship between mental health providers' rates and the suicide rate. Most of the data points are clustered where the mental health provider rate is between the minimum rate and 1000, due to the collected observations having low mental health provider rates. There are a few observations with a low mental health rate that are associated with an extremely high suicide rate. Additionally, there are a few observations with high mental health provider rates that correspond to a relatively lower suicide rate. These observations are likely problematic, such as leverage points, outliers, or influential points. By fixing the mental health provider rate, we can observe variations in the suicide rate, suggesting that varying the mental health provider rate does not affect the suicide rate. Thus, we cannot conclude that there is a strong association between the mental health provider rate and the suicide rate.

Many factors, such as affordability, quality of care, and the complexity of mental health issues, may lead to the observation that there is no obvious association between mental health provider rates and the suicide rate. Mental health services can be expensive; therapy sessions, medication, and other treatments can quickly accumulate costs, making them inaccessible to individuals with limited financial resources. While mental health providers play a crucial role in suicide prevention and intervention, the mere presence of a mental health provider doesn't guarantee effective treatment. The quality of care, as well as the competence of providers, can significantly affect outcomes.

In Panel (b), the relationship between the suicide rate and the percentage of the population with insufficient sleep is explored. Observations are randomly scattered with no discernible association between the variables. Varying the percentage of the population with insufficient sleep shows a random range of suicide rates. Similar to Panel (a), a few observations stand out, suggesting potential problems. Consequently, no association can be concluded, indicating that the lack of sleep might not directly influence people's intent to commit suicide.

Sleep plays a crucial role in regulating mood, and sleep deprivation can impair cognitive function. Individuals experiencing sleep deprivation may have difficulty coping with stressors and may be more prone to impulsive or risky behaviours, including suicidal thoughts and actions. However, it is important to recognize that this association might not always be universally applicable. Some individuals may have a higher tolerance for sleep deprivation and may not experience significant negative effects on their mental health or well-being. Individuals with strong social support networks, effective coping mechanisms, and positive life circumstances may be better equipped to manage sleep deprivation without experiencing a significant increase in suicide risk. Thus, it's possible to observe a lack of association between hours of sleep and the suicide rate.

Panel (c) plots the association between the suicide rate and the percentage of the population with limited access to healthy food. The data points exhibit a positive trend: higher percentages of the population with limited access to healthy food correspond to higher suicide rates, indicating a positive correlation. Since this population is characterized by low income and limited access to grocery stores, it implies that lower-income individuals with less access to basic needs have a higher suicide rate compared to wealthier individuals with easier access to food resources.

Factors such as nutritional deficiencies and socioeconomic disparities may explain this observed association. Limited access to healthy food can result in inadequate intake of essential nutrients, which are important for brain health and mood regulation. Thus, nutritional deficiencies can be linked to an increased risk of depression and other mental health disorders associated with suicide risk. Moreover, limited access to healthy food is more prevalent in socioeconomically disadvantaged communities characterized by poverty, food insecurity, and a lack of variety of resources. These communities may also face higher rates of unemployment, housing instability, and other social stressors that contribute to mental health disparities and suicide risk.

Panel (d) depicts the relationship between the US's per capita income and suicide rate. Low per capita income is associated with high suicide rates, while observations with per capita income exceeding \$500,00 are linked to significantly lower suicide rates. This aligns with the findings in Panel (c), suggesting that higher-income groups exhibit lower suicide rates.

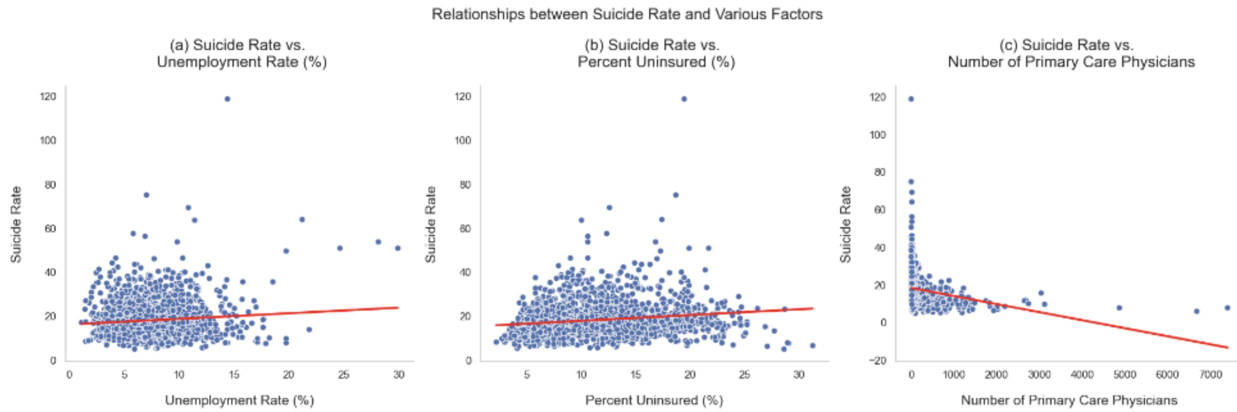

Figure 2: Scatterplots of Suicide Rate vs. Numerical Variables

The scatterplot above illustrates the relationships between the suicide rate and three different factors: unemployment rate, percent uninsured, and the number of primary care physicians.

`Suicide Rate vs. Unemployment Rate (Panel a):` The upward trend of the red line indicates a positive correlation. As the unemployment rate increases, there appears to be a slight increase in the suicide rate. This suggests that higher unemployment rates may be associated with higher suicide rates.

`Suicide Rate vs. Percent Uninsured (Panel b):` There is a positive correlation between the percent uninsured and suicide rates, as indicated by the best-fit line. While the scatter of points is widespread, the trend suggests that as the percentage of uninsured individuals in a population increases, so does the suicide rate. This could imply that lack of insurance, which often limits access to mental health services, may contribute to higher suicide rates.

`Suicide Rate vs. Number of Primary Care Physicians (Panel c):` The red trend line indicates a negative correlation; as the number of primary care physicians increases, the suicide rate appears to decrease. The data points become less dense as the number of primary care physicians increases, suggesting limited access to healthcare for many individuals in the US.

The findings support strategies that improve economic stability and healthcare access as potential ways to mitigate suicide risk. There may be a need for targeted mental health services in areas with high unemployment and uninsured rates.

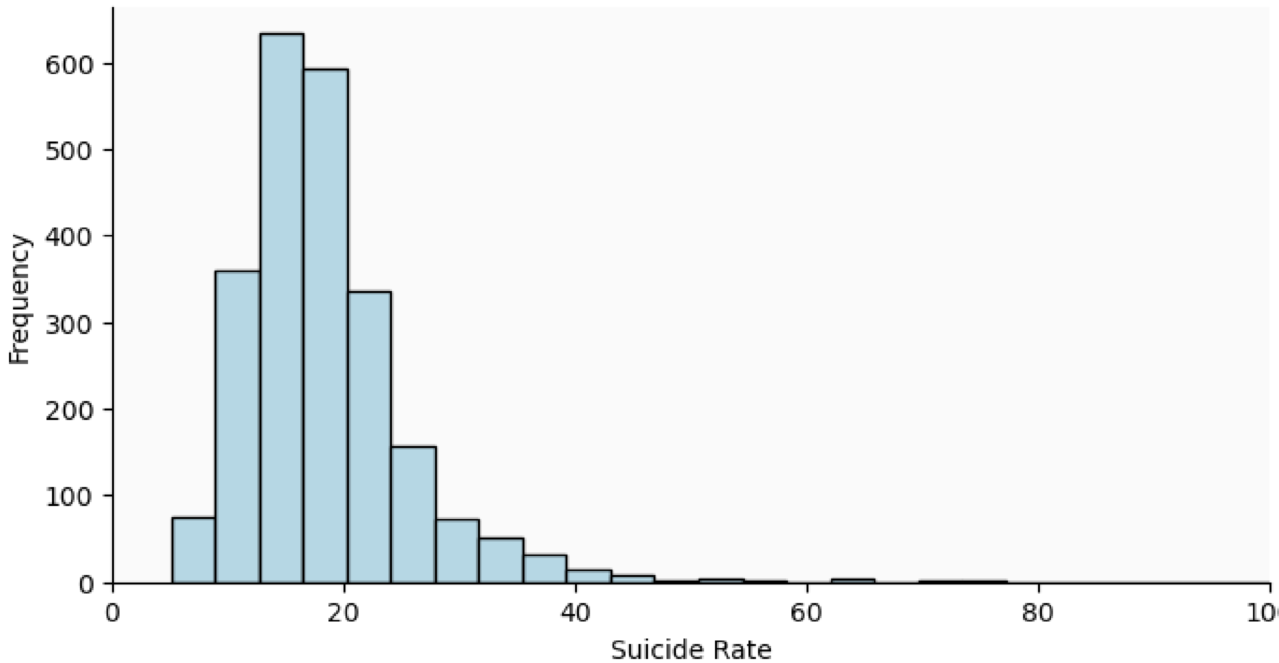

Figure 3: The Distribution of Suicide Rate Among US Counties

Analysis of Suicide Rate Distribution: The analysis of suicide rate data for the US reveals a notable right-skewed distribution, as depicted in the histogram above. This skewness indicates that the majority of regions in the country exhibit relatively lower suicide rates, forming a cluster on the left side of the distribution. However, the presence of a rightward tail shows the existence of specific areas with significantly higher suicide rates, contributing to the overall skewness of the distribution.

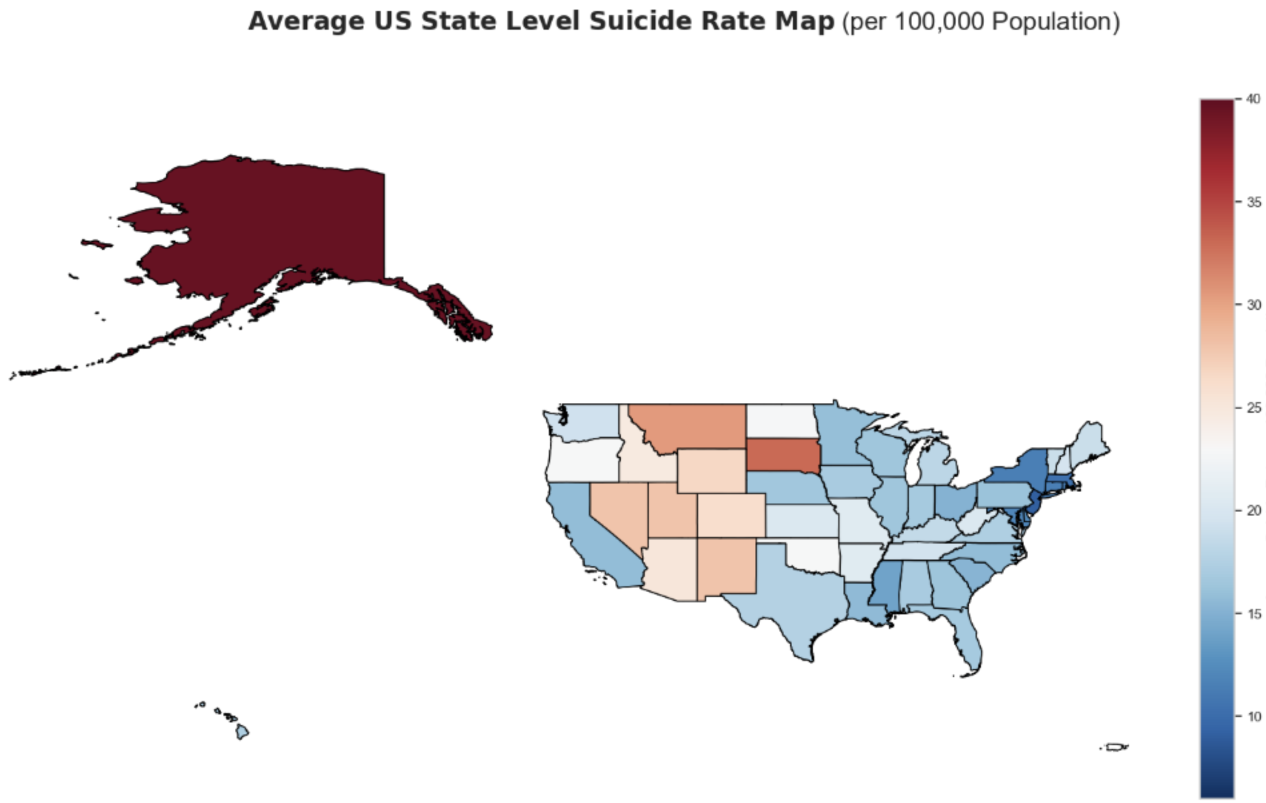

Figure 4: Suicide Rate Map

This choropleth map above describes the US state-level suicide rate. The map uses a colour spectrum ranging from dark blue to dark red. Regions in dark blue represent areas with the lowest suicide rates, while dark red regions represent areas with the highest suicide rates.

Most of the regions are coloured in a relatively light blue, indicating that these areas have suicide rates of around 15 to 20. There are a few states in dark blue located in the northeast of the US, representing extremely low suicide rates. These states are New York, Massachusetts, and New Jersey. They have higher population densities and relatively higher economic activities compared to other states. New York, for example, is a hub for finance, media, culture, fashion, and technology. It is also a top tourist destination, attracting numerous visitors each year. Additionally, it is home to prestigious universities and colleges, making it a global center for higher education and research. Similarly, New Jersey offers a variety of cultural attractions and has a strong economy driven by industries such as pharmaceuticals, telecommunications, and finance. It is home to major corporations and research institutions, contributing to its economic vitality [7]. Given the numerous economic opportunities that these urban and suburban areas offer, these factors likely contribute to greater financial stability and lower levels of stress, reducing the risk of suicide.

There are around nine states located in the west of the US that are coloured orange and red, indicating that these regions have relatively high suicide rates above 25. These states are Montana, South Dakota, Wyoming, Nevada, Utah, Colorado, Arizona and New Mexico. Mining and agriculture have been significant industries in these states, shaping their economies and landscapes. Many of these states have large rural areas where individuals may experience social isolation and limited access to mental health support resources, possibly due to geographic barriers and shortages of mental health professionals. Moreover, indigenous communities in states like South Dakota, Montana, Arizona, and Utah may face unique challenges related to historical trauma, displacement, and ongoing social and economic disparities [4]. These factors can contribute to higher rates of mental health issues and suicide within these communities.

Particularly noteworthy is Alaska, shaded in the darkest red, signifying the highest suicide rate among all states. Alaska's extreme suicide rates can be partially attributed to its unique environmental and social conditions. It is geographically isolated with a harsh climate, which can exacerbate feelings of loneliness and lead to seasonal affective disorder during long winters with very short days. Additionally, many communities in Alaska are remote and only accessible by plane or boat, limiting access to healthcare and social services, including mental health support. Economic factors also play a role; despite a wealth of natural resources, many areas face high living costs with limited job opportunities. Cultural factors amongst indigenous populations, similar to those in states like South Dakota and Montana, also contribute to the complexity of mental health challenges in Alaska. After examining the individual socio-economic and geographical influences on suicide rates in various regions, including the starkly high rates in Alaska, it becomes clear that suicide is a complex issue with multifaceted causes. The map shows that suicide rates vary widely across the states in the US. This variation can be attributed to many factors, including differences in policy regarding resource allocation, accessibility to healthcare services, and the demographic and geographic makeup of each state. The involvement of such complex and interrelated factors suggests that any analysis of suicide rates should take regional characteristics into account. Therefore, when considering policy changes or interventions, it is crucial to address these regional disparities to ensure that measures are effective and context-specific. This tailored approach is essential for mitigating risk factors and supporting at-risk populations in a manner that corresponds with their specific local realities.

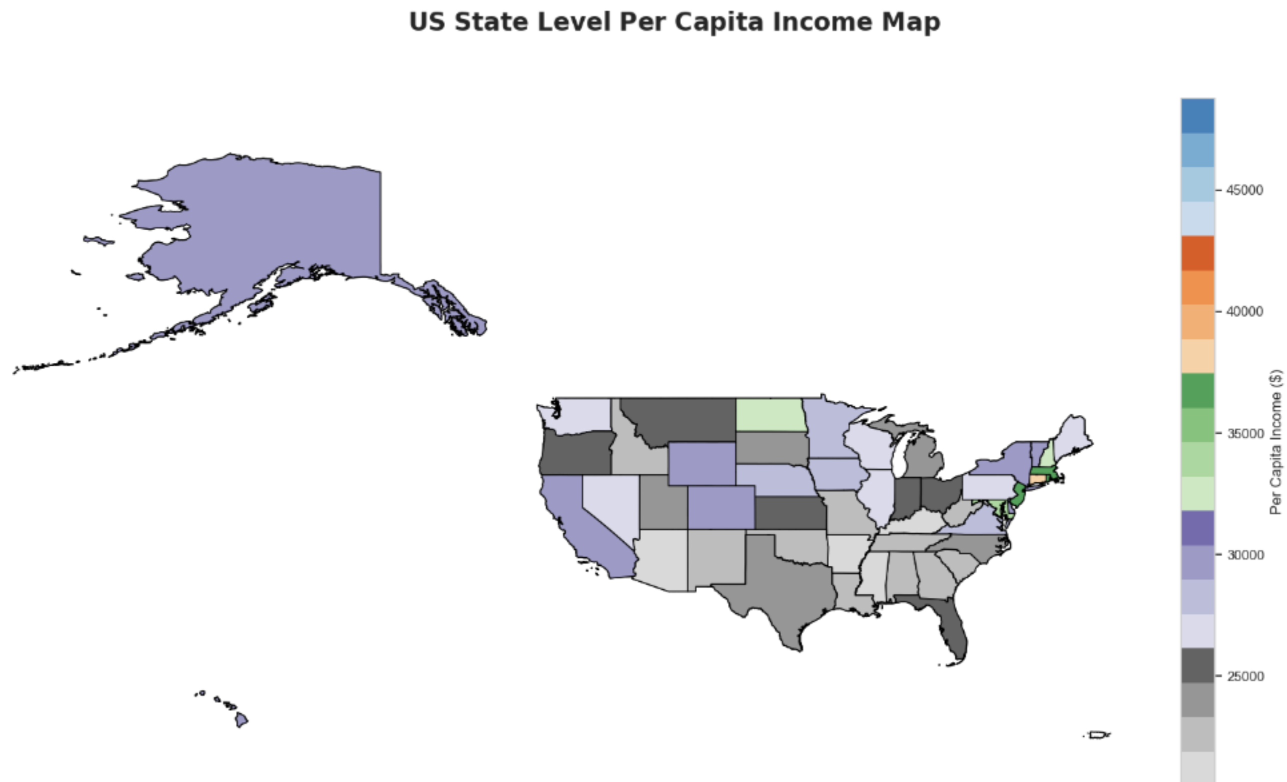

Figure 5: Per Capita Income Map

This choropleth map describes the US state-level per capita income. The map colours the US state-level per capita income based on five different income ranges. Most states are coloured in grey, representing that these states have a per capita income of or below $25,000.

Some states are coloured in purple, such as Alaska, California, Nevada, Wyoming, Colorado, Virginia, New York, Vermont, and Delaware, with state-level per capita income in the $30,000 range. A few states are coloured in green: North Dakota, Maryland, New Jersey, Massachusetts, Rhode Island, and New Hampshire. These states have a per capita income of around $35,000. Several factors can contribute to this, such as education and innovation, industry diversity, as well as access to capital and resources. All these states boast strong educational institutions, research facilities, and innovation ecosystems, including universities, research centers, and technology clusters, which are likely to drive economic growth. While each state's economic strength varies, they all benefit from diverse industries that contribute to income growth due to the mentioned factors. These various industries are likely to provide a range of job opportunities, suggesting that these areas are associated with higher wages and thus have higher per capita income values.

According to the map and data, Connecticut has the highest per capita income, possibly because it is home to a significant financial services and insurance sector, with major companies such as insurance giants Aetna, The Hartford, and Travelers headquartered in the state. These industries offer high-paying jobs and contribute to income growth in Connecticut [1].

The visual representation depicts the variation in per capita income across the US. States with higher per capita incomes, often characterized by robust economic drivers like advanced industries and strong financial sectors, contrast with those with lower incomes that may lack such economic advantages. The disparities highlighted in this map reflect the complexity of economic structures and the diverse factors contributing to the financial health of a region.

Per capita income serves as a crucial indicator for examining socioeconomic health and well-being across the states. The map therefore not only provides an overview of economic standings but also emphasizes the importance of considering income as a significant factor in the analysis of suicide rates.

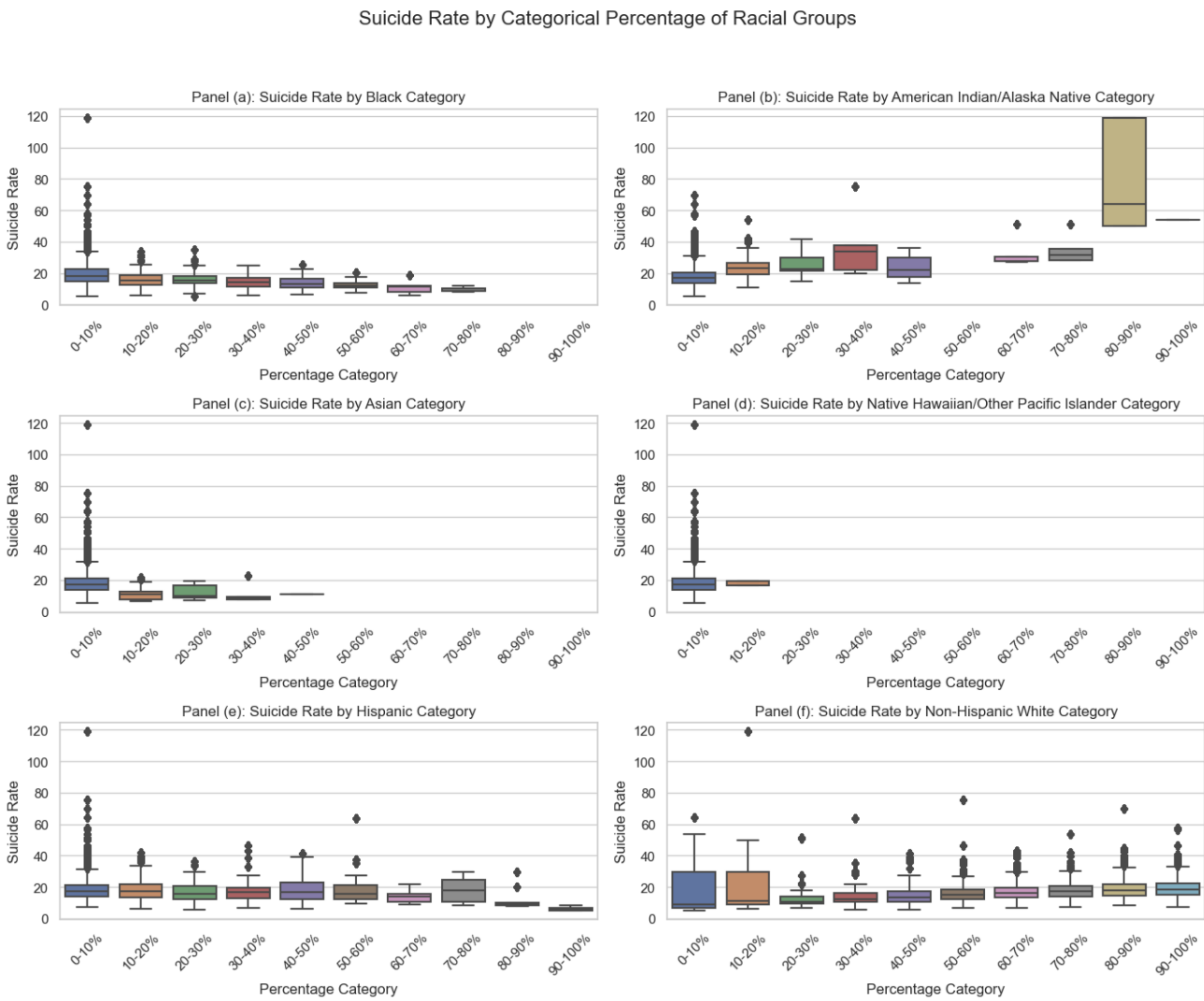

Figure 6: Suicide Rate by Racial Groups

The boxplot depicted the distribution of suicide rates among different racial groups. The plots show a wide range of suicide rates within most racial categories, as indicated by the spread of the data points and the length of the boxes. This suggests a high level of variability within each racial group that may be influenced by factors not directly captured by race alone.

The plot for the American Indian and Alaska Native categories (Panel b) shows a significant increase in suicide rates in areas with higher percentages of this population. The box for the 90% category is notably higher than the other categories. This may imply that areas with a high concentration of these populations are particularly at risk and could benefit from targeted mental health and suicide prevention resources.

The suicide rates in areas with higher percentages of Asian (Panel c) and Hispanic populations (Panel e) do not show a noticeable increase with higher racial concentration. This may suggest that for these groups, factors other than racial concentration are more significant in influencing suicide rates.

For the Black (Panel a) and Non-Hispanic White (Panel f) populations, the suicide rates do not exhibit a clear trend as the percentage increases. However, the variability seems to decrease in higher percentage categories for the Black population (Panel a). These findings suggest that interventions may need to be tailored to the specific needs of racial groups, with a particular focus on American Indian and Alaska Native populations (Panel b) in high-concentration areas. Moreover, the lack of a clear trend in suicide rates for other groups suggests that socioeconomic, cultural, or environmental factors might play a larger role than racial concentration alone.

3 Regression Analysis

In this section, the study explores the relationship between the US suicide rate and various economic and health-related variables through a series of multivariate linear regression models. Given the complexity and multifactorial nature of suicide rates, understanding the influence of different predictors is crucial. Ordinary Least Squares (OLS) regression provides a robust statistical method to quantify these relationships. By fitting an OLS model, factors that significantly impact the suicide rate can be assessed while adjusting for other variables in the model. This approach provides clarity on the individual contribution of each predictor while controlling for confounders.

The regression coefficients derived from OLS models express the expected change in the suicide rate for change in the predictor, holding other variables constant. By employing OLS regression, this study aims to uncover significant economic and health-related factors that correlate with changes in the suicide rate, providing insights that could inform targeted interventions and policies.

In the first regression result table (Table 2), six models have been fitted:

Specification 1:

$$\begin{aligned} suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\ &+ \beta_3 percent\_uninsured_i + u_i \end{aligned} \tag{1}$$

Specification 2:

$$\begin{aligned} suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\ &+ \beta_3 percent\_uninsured_i + \beta_4 pct\_limited\_healthfood\_access_i + u_i \end{aligned} \tag{2}$$

Specification 3:

$$\begin{aligned} suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\ &+ \beta_3 percent\_uninsured_i + \beta_4 pct\_limited\_healthfood\_access_i + \\ &\beta_5 per\_capita\_income_i + u_i \end{aligned} \tag{3}$$

Specification 4:

$$\begin{aligned} suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\ &+ \beta_3 percent\_uninsured_i + \beta_4 pct\_limited\_healthfood\_access_i \\ &+ \beta_5 per\_capita\_income_i + \beta_6 percent\_insufficient\_sleep_i + u_i \end{aligned} \tag{4}$$

Specification 5:

$$
\begin{aligned}
suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\
&+ \beta_3 percent\_uninsured_i + \beta_4 pct\_limited\_healthfood\_access_i \\
&+ \beta_5 per\_capita\_income_i + \beta_6 percent\_insufficient\_sleep_i \\
&+ \beta_7 percent\_unemployed_i + u_i \qquad (5)
\end{aligned}
$$

Specification 6:

$$
\begin{aligned}
suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\
&+ \beta_3 percent\_uninsured_i + \beta_4 pct\_limited\_healthfood\_access_i \\
&+ \beta_5 per\_capita\_income_i + \beta_6 percent\_insufficient\_sleep_i \\
&+ \beta_7 percent\_unemployed_i + \beta_8 percent\_uninsured_i + u_i \qquad (6)
\end{aligned}
$$

To comprehensively understand the determinants of the US suicide rate, the initial regression model (Table 2 Specification 1) was fitted with a key variable of interest—the mental health provider rate. This foundational model established a baseline from which the impact of additional variables could be evaluated.

Moving to Table 2 Specification 2, the number of primary care physicians was incorporated. This expanded model aimed to include broader aspects of access to care. With the inclusion of this variable, changes were observed in both the adjusted R-squared value and the coefficients. The adjusted R-squared, which reflects the proportion of variance in the dependent variable explained by the model, improved slightly.

Moving to Specifications 3 and 4, financial-related variables, including the percentage of unemployed and per capita income, were added, reflecting economic influences on the suicide rate. Including these variables increased the adjusted R-squared value, suggesting an improvement in the model's fit. The statistical significance of all these variables indicates their association with the suicide rate.

Specifications 5 and 6 introduced health-related variables: percentages of the population with insufficient sleep and limited access to healthy food into the analysis. The inclusion of these variables further improved the adjusted R-squared value, suggesting that health-related factors significantly contribute to the variability in suicide rates.

This iterative model-fitting process paid particular attention to the mental health provider rate. As variables were added across specifications, the coefficient for the mental health provider rate remained statistically significant at the 1% significance level, with slight variations in its estimated value. This underscores the persistent association of mental health resources with suicide rates. Among all six specifications, Specification 6 has the highest adjusted R-squared value with the inclusion of the variables of interest. This suggests that these predictors are strongly associated with the suicide rate, and they should be retained in the model to understand the relationship between these factors and the suicide rate.

	Dependent variable: suicide_rate					
	(1)	(2)	(3)	(4)	(5)	(6)
const	15.124***	15.378***	14.654***	22.427***	44.167***	41.927***
	(0.027)	(0.026)	(0.033)	(0.074)	(0.117)	(0.117)
mental_health_provider_rate	0.001***	0.003***	0.003***	0.004***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
num_primary_care_physicians		-0.004***	-0.004***	-0.003***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
pct_limited_healthfood_access						0.222***
						(0.002)
per_capita_income				-0.000***	-0.000***	-0.000***
				(0.000)	(0.000)	(0.000)
percent_insufficient_sleep					-0.612***	-0.582***
					(0.003)	(0.003)
percent_unemployed_CDC			0.117***	-0.102***	0.217***	0.131***
			(0.003)	(0.004)	(0.004)	(0.004)
percent_uninsured	0.267***	0.241***	0.230***	0.157***	0.090***	0.044***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	584204	577711	577711	577711	577711	577711
R^2	0.032	0.077	0.079	0.101	0.178	0.202
Adjusted R^2	0.032	0.077	0.079	0.101	0.178	0.202
Residual Std. Error	6.987 (df=584201)	6.660 (df=577707)	6.653 (df=577706)	6.575 (df=577705)	6.287 (df=577704)	6.193 (df=577703)
F Statistic	9521.668*** (df=2; 584201)	16163.346*** (df=3; 577707)	12454.712*** (df=4; 577706)	12953.097*** (df=5; 577705)	20851.946*** (df=6; 577704)	20924.990*** (df=7; 577703)
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression Result (1)

The exploratory data analysis previously conducted suggests the potential association between race and the suicide rate, as evidenced by the variation in suicide rates among different racial groups in the US. In subsequent specifications, race was incorporated into the analysis. Models of varying sizes were fitted and compared; interactions were added to refine these models, leading to the final model formulation.

Specification 1:

$$
\begin{aligned}
suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\
&+ \beta_3 pct\_limited\_healthfood\_access_i + \beta_4 per\_capita\_income_i \\
&+ \beta_5 percent\_american\_indian\_alaska\_native_i + \beta_6 percent\_asian_i \\
&+ \beta_7 percent\_black_i + \beta_8 percent\_hispanic_i + \beta_9 percent\_insufficient\_sleep_i \\
&+ \beta_{10} percent\_native\_hawaiian\_other\_pacific\_islander_i \\
&+ \beta_{11} percent\_non\_hispanic\_white_i + \beta_{12} percent\_unemployed_i \\
&+ \beta_{13} percent\_uninsured_i + u_i
\end{aligned} \tag{7}
$$

Specification 2:

$$
\begin{aligned}
suicide\_rate_i = \beta_0 &+ \beta_1 num\_primary\_care\_physicians_i + \beta_2 pct\_limited\_healthfood\_access_i \\
&+ \beta_3 per\_capita\_income_i \\
&+ \beta_4 percent\_american\_indian\_alaska\_native_i + \beta_5 percent\_asian_i \\
&+ \beta_6 percent\_black_i + \beta_7 percent\_hispanic_i + \beta_8 percent\_insufficient\_sleep_i \\
&+ \beta_9 percent\_native\_hawaiian\_other\_pacific\_islander_i \\
&+ \beta_{10} percent\_non\_hispanic\_white_i + \beta_{11} percent\_unemployed_i + u_i
\end{aligned} \tag{8}
$$

Specification 3:

$$
\begin{aligned}
suicide\_rate_i = \beta_0 &+ \beta_1 mental\_health\_provider\_rate_i + \beta_2 num\_primary\_care\_physicians_i \\
&+ \beta_3 pct\_limited\_healthfood\_access_i + \beta_4 per\_capita\_income_i \\
&+ \beta_5 percent\_american\_indian\_alaska\_native_i + \beta_6 percent\_asian_i \\
&+ \beta_7 percent\_black_i + \beta_8 percent\_hispanic_i + \beta_9 percent\_insufficient\_sleep_i \\
&+ \beta_{10} percent\_native\_hawaiian\_other\_pacific\_islander_i \\
&+ \beta_{11} percent\_non\_hispanic\_white_i + \beta_{12} percent\_unemployed_i \\
&+ \beta_{13} percent\_uninsured_i + \beta_{14} unemployed \times american\_indian\_i \\
&+ \beta_{15} unemployed \times asian\_i + \beta_{16} unemployed \times black\_i \\
&+ \beta_{17} unemployed \times hispanic\_i + \beta_{18} unemployed \times native\_hawaiian\_i \\
&+ \beta_{19} unemployed \times non\_hispanic\_white\_i + u_i
\end{aligned}
\tag{9}
$$

Building upon the model from Table 2, Specification 6 (Equation (3)), Specification 1 in Table 3 was established by including demographic factors such as the percentages of various racial groups among the array of variables. This specification represented the full model, integrating all relevant variables and thereby providing a detailed examination of the potential influences on the US suicide rate.

To examine the necessity of such a complex model, Specification 2 was derived as a pared-down version. This reduced model retained only the main variables of interest, such as the mental health provider rate, the percentage with limited access to healthy foods, per capita income, the percentage of the unemployment population, and race-related variables, aiming to identify the most impactful predictors. Specifications 1 and 2 yielded similar adjusted R-squared values. The choice between the full and reduced models was guided by the Hausman specification test, which compares the coefficients of the complex model to those of the simpler one. A significant test result would indicate that the more complex model's specific estimators are inconsistent under the reduced model, suggesting that the full model is the preferable specification. Indeed, the Hausman test resulted in a small p-value, proving that the full model (Specification 1) was more appropriate. This indicates that the additional variables provided information about the suicide rate not captured by the main variables alone.

In Specification 3 of Table 3, interaction terms between the unemployment rate and the percentage of each racial group were added to the already established Specification 1. This addition aims to investigate the potential variations in the relationship between unemployment and suicide rates across different racial demographics. These interaction terms are crucial for examining if the stress associated with unemployment, a well-known risk factor for suicide, has varying effects on different racial groups. The regression results for Specification 3 indicate that all the coefficients for the interaction terms are statistically significant at the 1% significance level. This suggests that the impact of unemployment on suicide rates interacts significantly with racial demographics. The analysis from this specification points to possible disparities in social and economic support systems among racial groups, underscoring the necessity for targeted policy interventions.

This process revealed that the coefficient for the mental health provider rate varied across different specifications, illustrating how its relationship with the suicide rate changes when accounting for various demographic contexts. The adjusted R-squared values showed incremental improvements from Specification 1 to Specification 3, indicating a progressively better fit of the model to the data. Specification 3, which has the highest adjusted R-squared value, is therefore considered the best fit and explains the most variation in the US suicide rate. Consequently, Specification 3 is the preferred model as per this analysis.

	Dependent variable: suicide_rate		
	(1)	(2)	(3)
const	17.702***	48.193***	75.655***
	(0.848)	(0.842)	(2.626)
mental_health_provider_rate	0.001***	-0.001***	0.001***
	(0.000)	(0.000)	(0.000)
num_primary_care_physicians	-0.000***		-0.000***
	(0.000)		(0.000)
pct_limited_healthfood_access	0.209***	0.210***	0.220***
	(0.002)	(0.002)	(0.002)
per_capita_income	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
percent_american_indian_alaska_native	0.310***	0.105***	-0.248***
	(0.008)	(0.009)	(0.026)
percent_asian	-0.183***	-0.470***	-1.000***
	(0.009)	(0.009)	(0.030)
percent_black	-0.076***	-0.287***	-0.552***
	(0.008)	(0.008)	(0.026)
percent_hispanic	-0.023***	-0.208***	-0.460***
	(0.008)	(0.008)	(0.025)
percent_insufficient_sleep	-0.325***	-0.388***	-0.375***
	(0.003)	(0.003)	(0.003)
percent_native_hawaiian_other_pacific_islander	1.307***	0.777***	0.413***
	(0.026)	(0.027)	(0.102)
percent_non_hispanic_white	0.107***	-0.134***	-0.480***
	(0.008)	(0.008)	(0.026)
percent_unemployed_CDC	0.324***	0.282***	-7.285***
	(0.004)	(0.004)	(0.328)
percent_uninsured	0.254***		0.214***
	(0.002)		(0.002)
unemployed_american_indian			0.076***
			(0.003)
unemployed_asian			0.113***
			(0.004)
unemployed_black			0.068***
			(0.003)
unemployed_hispanic			0.061***
			(0.003)
unemployed_native_hawaiian			0.124***
			(0.015)
unemployed_non_hispanic_white			0.080***
			(0.003)
Observations	577711	584204	577711
R^2	0.371	0.358	0.381
Adjusted R^2	0.371	0.358	0.380
Residual Std. Error	5.499 (df=577697)	5.690 (df=584192)	5.458 (df=577691)
F Statistic	26214.828*** (df=13; 577697)	29584.620*** (df=11; 584192)	18675.706*** (df=19; 577691)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: Regression Result (2)

3.1 Preferred Specification Analysis

The final model's estimates offer insights into how various factors are associated with the suicide rate. Limited access to healthy foods emerges as one of the key factors, with the estimated model coefficient suggesting that inadequate nutritional options, where the population with healthy food access increases by 1 percentage point are associated with an increase in suicide rates of approximately 0.22 per 100,000 people. This underscores the broader implications of access to resources on mental health and well-being. This result aligned with the previous hypothesis that the stressor associated with economic hardship and limited access to nutritious food can compound and likely increase the vulnerability of individuals facing these living conditions to mental health challenges, thus influencing suicide risk.

The result reveals another significant association between suicide and the percentage of the population uninsured. The statistical significance of the estimated coefficient suggests a reliable and nonrandom association between these variables, controlling for other factors in the model. One percentage point increase in the uninsured population is associated with a suicide rate increase of approximately 0.214 per 100,000 people. The observed association between the uninsured population percentage and the suicide rate underscores several critical implications for public health policy. Primarily, it suggests that expanding health insurance coverage could serve as an effective suicide prevention strategy by reducing a major stressor linked to higher suicide rates. For healthcare providers and public health officials, this association may guide more targeted interventions and screenings, particularly in demographics with higher uninsured rates. Furthermore, ongoing research and monitoring are essential to explore the underlying mechanisms of this association and to evaluate the impact of health policy changes. Lastly, enhancing public awareness about the critical role of health insurance in mental health could motivate increased coverage efforts and inform individuals about accessible mental health resources, illustrating the broader societal and economic benefits of comprehensive health insurance policies in mitigating suicide risks.

Another interesting finding comes from the mental health provider rate. The initial hypothesis is that clinical intervention is a crucial element in the prevention of suicide. However, a counter-intuitive observation came from the positive estimated coefficient of 0.001 from the mental health provider rate, indicating that increasing the rate of mental health providers is positively linked to suicide rates. The observation that an increase in the rate of mental health providers correlates with higher suicide rates highlights several important considerations. It suggests that the presence of more mental health professionals might be a reactive measure in response to increased demand from higher suicide rates, rather than a preventative strategy. This raises questions about the direction of causality, indicating the need for further studies to determine whether higher suicide rates attract more providers, or vice versa. Moreover, the magnitude of this estimated coefficient also suggests that clinical intervention might not be the strongest indicator when considering factors like resource access and insurance coverage. The finding highlights that merely increasing the number of providers does not guarantee effective mental health care; barriers such as cost, insurance coverage, and social stigma might hinder access to these services.

This model also provides interesting associations between unemployment, different racial groups, and their interaction effects. The coefficient for the percentage of unemployment is -7.285, indicating that for every one percentage point increase in unemployment, the suicide rate decreases by 7.285 units, assuming the racial composition and other factors remain constant. Looking at the coefficient for the interaction between unemployment and the American Indian Alaska Native group, the estimated value of 0.076 suggests that the negative effect diminishes as the percentage of this racial group in the population increases. Similar interaction effects between other racial groups and unemployment were also observed. Although these results may seem counter-intuitive, several factors could lead to this outcome, such as stronger social bonds and community support associated with unemployment, changes in work-related stress, and government intervention and support. During times of unemployment, strong family ties and community support networks provide resilience against the mental health impacts of unemployment. These social support systems may offer emotional support, shared resources, or coping mechanisms that mitigate the stress typically associated with job loss. Moreover, unemployment might reduce the pressures and anxiety associated with employment, such as long working hours, job dissatisfaction, or toxic work environments, which are significant risk factors for mental health issues and suicide. Unemployment might also change an individual's daily routine in ways that could decrease the risk of suicide. For instance, individuals might have more time to engage in physical activity, pursue hobbies, or spend time with loved ones, all of which can improve mental health. Additionally, in response to high unemployment rates, governments may implement more robust social safety nets, including unemployment benefits, healthcare subsidies, and other forms of financial assistance, which can reduce financial stress and provide a buffer against the despair that might lead to suicide.

Overall, the preferred model, with an adjusted R-squared of 0.38, accounts for a significant portion of the variability in suicide rates, underscoring the complex relationship among demographic, economic, and health-related factors. Based on the relative magnitude of the estimated coefficients, this model suggests that the percentage of uninsured individuals and access to resources are relatively stronger predictors of the suicide rate. The interaction terms further reveal that the association between unemployment and suicide rates varies among different racial groups. Additional analysis should focus on the data by community to explore whether local policies and economic conditions play a significant role.

4 Machine Learning

Building upon the regression analysis, the study further utilizes the capabilities of machine learning to provide additional insights into the factors affecting the US suicide rate. To this end, a regression tree was fitted as a non-linear, non-parametric model that offers a visual representation of decision rules and their hierarchical relationships. This model highlights the complex interaction effects more intuitively, potentially revealing non-linear patterns that the linear regression models may overlook.

Subsequently, a Random Forest model was built to estimate the relative importance of each predictor variable. This method not only improves prediction accuracy but also offers robustness against overfitting by averaging the results of numerous trees, providing a quantifiable measure of each factor's impact on the suicide rate.

4.1 Regression Tree

Objective Function:

$$\min_{j,s}\left[\sum_{i:x_{i,j}\leq s,x_i\in R1}(y_i-\hat{y}_{R1})^2+\sum_{i:x_{i,j}>s,x_i\in R2}(y_i-\hat{y}_{R2})^2\right]$$

In the context of a regression tree, the objective function guides the algorithm to divide the data into groups that make the outcomes within each group similar. This is done by finding the best feature to split the data on, in a way that minimizes the differences within these groups, measured by the Mean Squared Error (MSE).

Minimizing MSE means the algorithm tries to find splits that make the predicted suicide rates as close as possible to the actual rates within each group formed after the split. The process continues, breaking the data down into smaller groups until adding more splits doesn't improve the predictions or until they reach the set limit to prevent the tree from becoming overly complex.

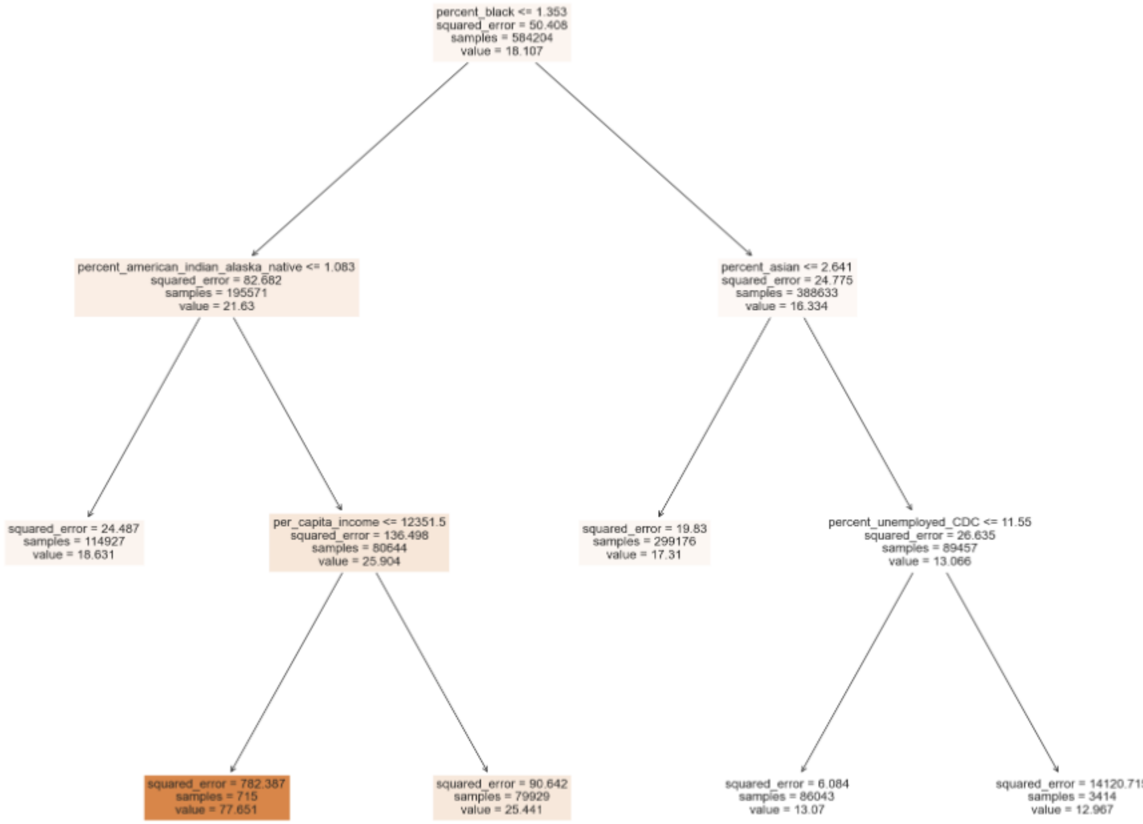

Figure 7: Regression Tree

The initial split of the regression tree in Figure 7 is based on the percentage of the black population at a value of $\leq 1.353\%$, signifying this variable as the most significant predictor within the dataset. The tree indicates that different levels of the percentage of the Black population are associated with varying suicide rates.

After the initial split, the percentage of American Indian and Alaska Natives and the percentage of the Asian population are identified as the next level of predictors with thresholds at $\leq 1.083\%$ and $\leq 2.641\%$, respectively. Considering the left child node of the percentage of American Indian and Alaska Natives, it is further divided into one terminal node and another node, which itself is split based on per capita income. When the percentage of this population is $\leq 1.083\%$, the terminal node is reached with the predicted suicide rate of 18.631 per 100,000 people. On the other hand, when the population of American Indian and Alaska Natives is more than 1.083%, per capita income becomes another key factor determining the suicide rate. For regions with a high percentage of American Indian and Alaska Natives and a per capita income threshold at $\leq 12,351.5$, the terminal node with the highest predicted suicide rate of 77.651 is reached. If the per capita income exceeds $12,351.5$, the suicide rate decreases to 25.441, which is the second-highest predicted suicide rate among all terminal nodes.

The split based on per capita income highlights the importance of income levels in regions with a high percentage of American Indian and Alaska Natives. It further depicts the income and financial constraints faced by these communities, indicating that financial constraints are a significant factor that leads to the highest predicted suicide rates. This also suggests that suicide rates have a non-linear relationship with per capita income, where certain income levels may correlate with higher or lower rates of suicide, reinforcing the complexity of socio-economic factors in public health outcomes.

For the right child node that splits based on the percentage of the Asian population, regions where this demographic exceeds 2.641% encounter unemployment as a significant determinant of suicide rates. This indicates that in areas with a high percentage of the Asian population, unemployment is a critical issue. This node's emphasis on unemployment suggests that joblessness may have a profound impact on mental health within the Asian community.

Unlike a linear regression model, which assigns a global coefficient to each predictor, a regression tree identifies distinct subgroups within the population that may have varying risk levels for suicide. The Mean Squared Error (MSE) in regression trees reflects the prediction error within the subsets of data at each node. The relatively large MSE values observed may be primarily due to the omission of relevant predictors since suicide rates are influenced by a complex set of factors, and the model may not have incorporated enough variables to fully capture this complexity and variance. Although a high MSE value limits the regression tree's predictive ability, fitting this tree can still reveal insights not observed with linear regression models. This regression tree has helped uncover the varied challenges faced by different communities. The identified issue of low income among American Indian and Alaska Natives suggests that policies aimed at assisting these low-income groups could be crucial in reducing suicide rates and improving mental health conditions. Meanwhile, in regions with high percentages of Asian and Black populations, unemployment has emerged as a significant issue. This highlights the potential benefits of targeted unemployment assistance and policies addressing job loss challenges within these communities.

4.2 Random Forest Model

The importance matrix from the Random Forest model provides a clear visualization of the variables most influential in predicting the suicide rate, as shown in Figure 8 below. The percentage of American Indian Alaska Natives and the percentage of the Black population emerge as the most significant predictors of the suicide rate, which aligns with the findings from fitting the regression tree. Their higher importance values suggest that racial demographics play a crucial role in modelling suicide risk within this dataset.

The number of primary care physicians and the percentage of the uninsured population are also influential, highlighting the importance of healthcare access and insurance status. This could reflect the role of healthcare systems in suicide prevention. On the other hand, predictors like the percentage of Native Hawaiian and Other Pacific Islander populations are found to be less predictive in this model. This finding points to the need for further research to understand why certain variables, such as mental health provider rates, are less influential—whether because of data limitations or because other factors overshadow their impact.

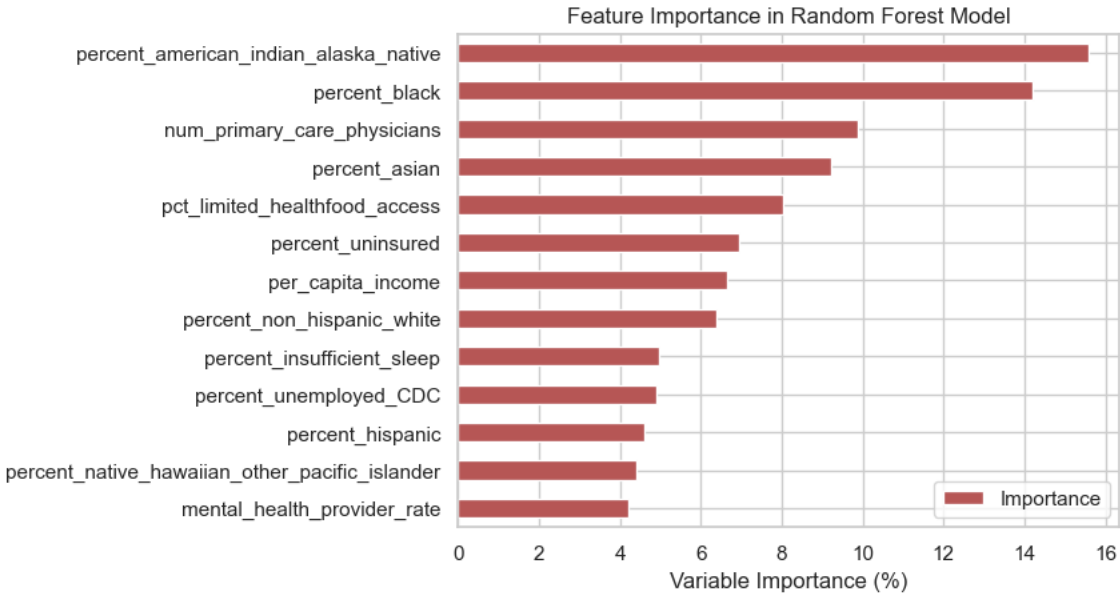

Figure 8: Importance Matrix

5 Limitation

5.1 Existence of Potential Confounders

While our regression models and machine learning analyses provide valuable insights into the factors influencing suicide rates, we acknowledge several limitations inherent to our study design and dataset. Primarily, the potential for unaddressed confounders exists, which could impact the relationships observed between suicide rates and the evaluated predictors such as mental health provider rates, access to healthy foods, and socioeconomic variables.

Given the complexity of the factors contributing to suicide, which include unmeasured psychological, genetic, and situational variables, our models may not fully account for all underlying influences. For instance, personal life events and undisclosed mental health issues could act as confounders. These elements might skew the observed associations or mask other significant relationships.

While our findings contribute to the understanding of suicide dynamics and offer a basis for targeted interventions, they should be interpreted with caution. The potential presence of these confounders suggests a need for further research incorporating more comprehensive data.

5.2 Limitation in Model's Prediction Accuracy

The regression tree under Section 4.1 indicated relatively large MSE values for certain branches, suggesting that the model's ability to accurately predict suicide rates across different subsets of data might be limited.

This could be due to the complexity of suicide as a public health issue, influenced by unmeasured factors such as personal crises or psychological conditions, which are not represented in the available predictors.

Further studies incorporating a broader range of variables and more granular data are necessary to enhance the predictive accuracy and provide more reliable insights for policy-making and suicide prevention strategies.

5.3 Limitations in Generalizability to Non-US Populations

This study utilizes data exclusively from the United States as provided by the 2020 Community Health Rankings. While this dataset allows for a detailed examination of the relationship between suicide rates and various predictors within the US, it inherently limits the generalizability of this study's findings to populations in other countries. Different countries have distinct public policies, cultural contexts, healthcare systems, and socio-economic structures, all of which can significantly influence the association between the predictors studied and suicide rates.

For instance, the impact of healthcare access on suicide rates may differ markedly in countries with universal healthcare systems compared to the US system. Similarly, socioeconomic factors such as unemployment may have different implications for mental health and suicide in societies with varying levels of social safety nets. Furthermore, cultural attitudes towards mental health and suicide can vary widely, affecting both the prevalence of suicide and the effectiveness of preventive measures.

Given these variations, the associations and predictions reported in this study should be applied to non-US populations with caution. Future research should be conducted across different populations to verify whether the observed relationships hold in diverse environments.

6 Conclusion

This study has provided a detailed examination of how demographics, healthcare access, and socioeconomic factors influence suicide rates across the United States. Utilizing data from the 2020 Community Health Rankings, this study brings fresh insights into this complex issue. The use of descriptive statistics, bar plots, and scatterplots has revealed substantial regional disparities in suicide rates, income levels, and access to healthy food resources. State-level data analyses have highlighted variations in suicide rates and identified socioeconomic status as a crucial determinant. For instance, the negligible relationship between suicide rates and the availability of mental health providers, along with the non-significant link between sleep insufficiency and suicide rates, presents counter-intuitive findings that indicate the need for further investigation. In contrast, the positive correlation between limited access to healthy foods and increased suicide rates suggests that environmental and socioeconomic factors play a crucial role in suicide risk.

Key findings from the regression model indicate that limited access to nutritious food and lack of health insurance are relatively stronger predictors of increased suicide rates. This aligns with the results of previous studies and suggests that economic hardships and inadequate health coverage exacerbate mental health challenges. Surprisingly, higher numbers of mental health providers correlate with increased suicide rates, possibly reflecting a reactive surge in mental health services in response to rising suicide incidences rather than a preventative measure. Additionally, the interaction between unemployment rates and racial demographics emphasizes the complex nature of these relationships, where community support potentially mitigates the negative impacts typically associated with unemployment. Furthermore, this study utilizes machine learning techniques, including regression trees and Random Forest models, to reveal non-linear patterns and variable importance that traditional models might overlook. These insights underline the necessity for targeted public health interventions and policies that not only expand access to critical resources but also address the specific needs of diverse populations to effectively reduce suicide risks.

Looking forward, there are several areas where further research could prove beneficial. First, expanding the scope of this study to include international data could help validate these findings across different cultural and economic contexts. Additionally, a deeper dive into the specific types of mental health services, their effectiveness, and their direct impact on suicide rates could help discover clearer pathways for intervention. Further study could integrate additional variables such as the Gini index and other racial-related variables, and explore associations that might improve the regression model. These additional variables could offer deeper insights into how socioeconomic factors intersect with suicide risk.

7 References

[1] **Connecticut Government**. (n.d.). *Official website of the state of Connecticut*. Retrieved March 1, 2024, from `https://portal.ct.gov/`

[2] **County Health Rankings & Roadmaps** (n.d.). Rankings data documentation. Retrieved from `https://www.countyhealthrankings.org/health-data/\protect\penalty\z@methodology-and-sources/rankings-data-documentation`

[3] **Hedegaard, H., Curtin, S. C., & Warner, M.** (2020). *Increase in suicide mortality in the United States, 1999–2018*. National Center for Health Statistics Data Brief, No. 362.

[4] **International Work Group for Indigenous Affairs (IWGIA)**. (n.d.). Retrieved from `https://www.iwgia.org/en/`

[5] **Machado, D. B., Rasella, D., & dos Santos, D. N.** (2015). Impact of Income Inequality and Other Social Determinants on Suicide Rate in Brazil. *PLoS ONE*, 10(4), e0124934. DOI: `https://doi.org/10.1371/journal.pone.0124934`

[6] **Morley, S. A.** (1981). The Effect of Changes in the Population on Several Measures of Income Distribution. *The American Economic Review*, 71(3), 285–294. Retrieved from `http://www.jstor.org/stable/1802779`

[7] **New Jersey Government**. (n.d.). *Official website of the state of New Jersey*. Retrieved March 1, 2024, from `https://www.nj.gov/`

[8] **Tondo, L., Albert, M. J., & Baldessarini, R. J.** (2006). Suicide Rates in Relation to Health Care Access in the United States: An Ecological Study. *Journal of Clinical Psychiatry*, 67(4), 517-523.

[9] **World Health Organization**. (2022, March 10). *Suicide*. Retrieved from `https://www.who.int/news-room/fact-sheets/detail/suicide`