

Suicide Death Rates by Race, Sex, and Age in the United States*

Disproportionate Suicide Rates by Men Ages 85 and Older and White Individuals

Julian Marrero

April 18, 2024

This report accesses the death rate of suicides in the United States. The study employed linear regression to evaluate the relationship between age, sex, and race to the death rate by suicide. The regression analysis found that men had higher death rates by suicide compared to women, and particularly men at the ages of 85 years and older had the highest suicide death rate. It was also found that individuals identifying as White had the highest suicide death rate compared to other ethnicities (i.e., Asian/Pacific Islander, African American, American Indian/Alaska Native). The findings from this paper allow us to identify groups at high risk of suicide in order to implement specialized measures to combat this.

1 Introduction

In the United States, suicide represents a significant public health issue, shaped deeply by demographic variables such as age, sex, and race. The eleventh leading cause of death is by suicide (National Institute of Mental Health 2024). More specifically, for individuals aged 15-24, suicide is the third leading cause of death in the United States (National Institute of Mental Health 2024). This paper explores how these factors intersect to influence suicide rates, providing insights into the complexities of vulnerability and risk within diverse populations. A linear regression model was utilized to conduct statistical analysis focusing on how the demographic factors play a role with the death rate by suicide. With the aim to uncover patterns and disparities that can inform targeted prevention strategies, understanding these demographic influences is crucial for developing effective interventions and policies tailored to the specific needs of different groups, ultimately striving to reduce the incidence of suicide across the nation.

*Code for R scripts and data for this analysis are available at: <https://github.com/julianmarrero/Death-Rates-For-Suicide-In-The-United-States>

The parameter of interest is the expected difference in suicide death rates per 100,000 people across different races, sexes, and age groups in the United States. Specifically, the difference in the expected suicide death rates between males and females, adjusting for race and age group. The difference in the expected suicide death rates among different racial groups (e.g., White vs. Black or African American vs. Asian or Pacific Islander vs. American Indian or Alaska Native), adjusting for sex and age group. The difference in the expected suicide death rates among different age groups (e.g., 10-14 years, 15-24 years, 25-44 years, 45-64 years, 65 years and over), adjusting for sex and race.

From the analysis conducted, a general trend was that the suicide rate increased as the age group increased. Within these age groups, men had a higher death rate by suicide compared to women. When looking at the suicide rates by race, White Americans had significantly the highest suicide rate, followed by Black or African Americans, American Indians or Alaska Natives, and Asian or Pacific Islanders.

The linear regression analysis was conducted using the statistical programming language R (R Core Team 2020). The following packages were used in the analysis: tidyverse (Wickham et al. 2019), dplyr (Wickham et al. 2023), ggplot2 (Wickham 2016), readr (Wickham, Hester, and Bryan 2024), and stats (R Core Team 2023).

The paper begins with a discussion of the data used in this report, shown in Section 2. In this section, there is a discussion on the variables in the dataset, the context of the data, whether other datasets could have been used, the data construction and cleaning process, and what was being measured in the study. This is followed by a detailed account of the linear model used in the analysis, see Section 3. The results of the linear regression are then presented and explained in Section 4. The paper then concludes with a discussion of the findings, the limitations of the analysis, and the future research to be conducted, see Section 5.

2 Data

2.1 Overview of Dataset

The dataset used for the analysis had the following variables:

- **YEAR:** The year the data was recorded, ranging from 1950 to the most recent year available. This allows for trend analysis over time.
- **AGE:** Categorized into groups (e.g., 10-14 years, 15-24 years, etc.), facilitating age-specific analysis which is crucial given the varying suicide risks across different life stages.
- **SEX:** Male or Female, important for understanding gender differences in suicide rates.

- **RACE:** Including categories like White, Black or African American, Asian or Pacific Islander, and American Indian or Alaska Native. This division helps in assessing racial disparities in mental health outcomes.
- **ESTIMATE:** The suicide rate per 100,000 population, adjusted for age. This standardization allows for comparison across different population groups.

2.2 Broader Context

The dataset is critical for understanding public health trends related to mental health and societal impacts across different demographic groups in the United States. It supports a range of analyses, from public health surveillance to informing targeted suicide prevention strategies. Given the stigma and complexity surrounding mental health and suicide, having detailed and reliable data is crucial for effective policy-making and community support initiatives.

2.3 Comparative Analysis with Similar Datasets

While other datasets related to health statistics and mortality due to various causes exist (e.g., WHO's Global Health Estimates, CDC's Wide-ranging Online Data for Epidemiologic Research (WONDER)), the selected dataset is unique in its specific focus on suicide, combined with detailed demographic breakdowns. These features make it particularly valuable for focused studies on suicide prevention, unlike more generalized health datasets that might not provide detailed suicide-specific data or the necessary demographic granularity.

2.4 Data Construction and Cleaning

Certain variables, like ESTIMATE, were constructed to provide age-adjusted rates, ensuring comparability across different demographic segments irrespective of their age distributions—a crucial factor in epidemiological analyses concerning health outcomes.

The data cleaning included filtering to focus on relevant demographic breakdowns, handling missing data, and ensuring consistent categorization across variables like sex and race, which might otherwise vary due to changes in societal definitions and recording practices over time. Missing data was removed from the analysis as it may skew the data in a particular way, thus it was removed.

2.5 Measurement of Suicide Death Rates

The suicide death rate is typically calculated as the number of suicide deaths in a given year per 100,000 population, adjusted for age. Age adjustment is crucial because it allows for comparisons across different populations and over time, controlling for the effects of age distribution that may vary significantly between different groups or change over time. The dataset measures suicide rates in deaths per 100,000 population. This standardization helps in making the data comparable across different sizes of populations.

The number of suicide deaths, which comes from the National Vital Statistics System (NVSS) mortality files. These files compile death certificate data from all U.S. states and territories, which are filled out by medical examiners, coroners, and attending physicians. Population estimates come from the U.S. Census Bureau. These estimates are critical as they provide the base population required to calculate death rates accurately.

2.6 Measurement Reliability and Validity

The accuracy of suicide classification on death certificates can be influenced by the ability and willingness of officials to recognize and record suicide as the cause of death. There can be stigma or legal implications associated with recording a death as a suicide, potentially leading to underreporting. The training and practices of certifying officials might vary widely, impacting the consistency of data on suicide across different regions and over time.

Changes in data collection methods, coding practices (e.g., ICD code revisions), or definitions of suicide across different time periods can affect the comparability of suicide rates over time. Regional differences in reporting practices or in the prevalence of factors associated with suicide risk can affect the measurement's reliability.

The extent to which the NVSS mortality data captures all suicide deaths depends on the completeness and accuracy of the cause of death recorded. There's an ongoing debate about whether all suicides are accurately identified or if some are misclassified as accidents or undetermined deaths. Suicide death rates are used as a proxy to measure the underlying construct of suicide risk within the population. The construct validity hinges on whether the rates accurately reflect true differences in suicide risk across different demographic groups or are biased by reporting practices or data collection methods.

3 Model

3.1 Model Overview

The linear regression model in this analysis was employed to estimate the relationship between suicide death rates (as the dependent variable) and several key demographic factors (independ-

dent variables) such as sex, race, and age groups. Here is how the model was constructed:

- Dependent Variable: ESTIMATE, representing the suicide rate per 100,000 population, age-adjusted.
- Independent Variables:
 - Sex: Categorical variable with two levels (Male, Female).
 - Race: Categorical variable including groups such as White, Black or African American, Asian or Pacific Islander, and American Indian or Alaska Native.
 - Age_Group: Categorical variable with several age bands such as 10-14 years, 15-24 years, etc.

The model used for estimation is a linear regression model specified as:

$\text{ESTIMATE} \sim \text{factor}(\text{Sex}) + \text{factor}(\text{Race}) + \text{factor}(\text{Age_Group})$ where:

- ESTIMATE represents the suicide rate per 100,000 population.
- $\text{factor}(\text{Sex})$, $\text{factor}(\text{Race})$, and $\text{factor}(\text{Age_Group})$ are treated as fixed effects to account for systematic differences in suicide rates based on these demographic characteristics.

This setup allows the model to quantify the impact of each categorical variable on the suicide rate, controlling for the influences of the others. The coefficients resulting from this regression give the estimated differences corresponding to each level of sex, race, and age group compared to their respective reference categories, thereby addressing the defined estimands.

3.2 Model Fitting

The model was fit using the Ordinary Least Squares (OLS) method, which is the most common approach in linear regression. OLS estimates the parameters by minimizing the sum of the squared differences between the observed values and the values predicted by the model.

3.3 Interpretation of Coefficients

Each coefficient in the regression model represents the expected change in the suicide rate associated with each category of the independent variables, holding all other variables constant. For example, the coefficient for Male in the Sex variable indicates how much higher or lower the suicide rate is for males compared to females (the reference category), controlling for race and age. Similarly, coefficients for different races show differences in suicide rates relative to a reference race (typically the first alphabetically unless otherwise specified).

3.4 Assumptions

Several key assumptions were made when using this linear regression model: (1) The relationship between the independent and dependent variables is linear. (2) Observations are independent of each other. (3) Constant variance of error terms irrespective of the value of the independent variable. (4) The residuals of the model are normally distributed.

These assumptions were checked using diagnostic plots like Residual vs. Fitted, Normal Q-Q plots, Scale-Location plots, and Residual vs. Leverage plots. Such checks help validate the appropriateness of the model and indicate whether modifications or transformations are needed.

4 Results

4.1 Summary Statistics

Figure 1 displays the summary statistics generated from the applied linear regression model. Once the data was loaded, it was then filtered by the rows where the STUB_NAME column matches “Sex, age and race,” focusing on data segmented by these demographics. I then removed any rows with missing values in the data frame to ensure clean data for analysis. I then extracted the sex, race, and age information from each row. I then performed a linear regression analysis based on my discussed model to represent the relationship between the suicide rate (ESTIMATE) and the factors of sex, race, and age group.

From Figure 1, the intercept row indicates the baseline suicide rate (without considering sex, race, or age group) is 9.5924 per 100,000 population. The factor(Sex)Male row suggest being male is associated with an increase in the suicide rate of 13.9575 per 100,000 population compared to females. The following three rows depicting different ethnic groups suggests that Asian or Pacific Islander groups have a significantly lower suicide rate by 9.1691 per 100,000 population. Black or African American groups have a significantly lower suicide rate by 11.2777 per 100,000 population. White groups have a higher suicide rate by 1.2909 per 100,000 population. The following rows display the effect of age on the suicide death rate. The 25-44 years group has a higher suicide rate, but it’s not statistically significant as $p > 0.05$. The 45-64 years group has a lower rate, but this is not statistically significant for the same reason. The “65 years and over” group has a significantly higher rate by 2.6592 per 100,000 population. The “65-74 years” group has a significantly higher rate by 2.4185 per 100,000 population. The “75-84 years” group has a significantly higher rate by 11.4157 per 100,000 population. The “85 years and over” group has a significantly higher rate by 33.7877 per 100,000.

The Residual Standard Error has 6.141 on 1271 degrees of freedom, indicating the average size of the residuals. Multiple R-squared is 0.8092, suggesting that approximately 80.92% of the variability in suicide rates can be explained by the model. The adjusted R-squared is 0.8077,

Residuals:

Min	1Q	Median	3Q	Max
-18.241	-4.361	-0.112	3.814	28.300

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.5924	0.6262	15.319	< 2e-16
factor(Sex)Male	13.9575	0.3769	37.030	< 2e-16
factor(Race)Asian or Pacific Islander	-9.1691	0.6391	-14.348	< 2e-16
factor(Race)Black or African American	-11.2777	0.6030	-18.702	< 2e-16
factor(Race)White	1.2909	0.6029	2.141	0.03244
factor(Age_Group)25-44 years	0.9712	0.5118	1.898	0.05798
factor(Age_Group)45-64 years	-0.1961	0.5242	-0.374	0.70837
factor(Age_Group)65 years and over	2.6592	0.5552	4.790	1.87e-06
factor(Age_Group)65-74 years	2.4185	0.8017	3.017	0.00261
factor(Age_Group)75-84 years	11.4157	0.8121	14.057	< 2e-16
factor(Age_Group)85 years and over	33.7877	1.0666	31.677	< 2e-16

Residual standard error: 6.141 on 1271 degrees of freedom

Multiple R-squared: 0.8092, Adjusted R-squared: 0.8077

F-statistic: 539.1 on 10 and 1271 DF, p-value: < 2.2e-16

Figure 1: Summary Statistics of Regression Model

which is a modified version of R-squared that has been adjusted for the number of predictors in the model. It's very close to the R-squared, which suggests that the model does not have unnecessary additional variables. The F-statistic is 539.1 on 10 and 1271 degrees of freedom with a p-value $< 2.2e-16$, indicating that the overall model is statistically significant. The distribution of residuals, with a median close to zero, suggests that the model may be well-calibrated, but the range indicates some large errors in prediction for certain observations.

4.2 Suicide Rates by Sex and Age Group

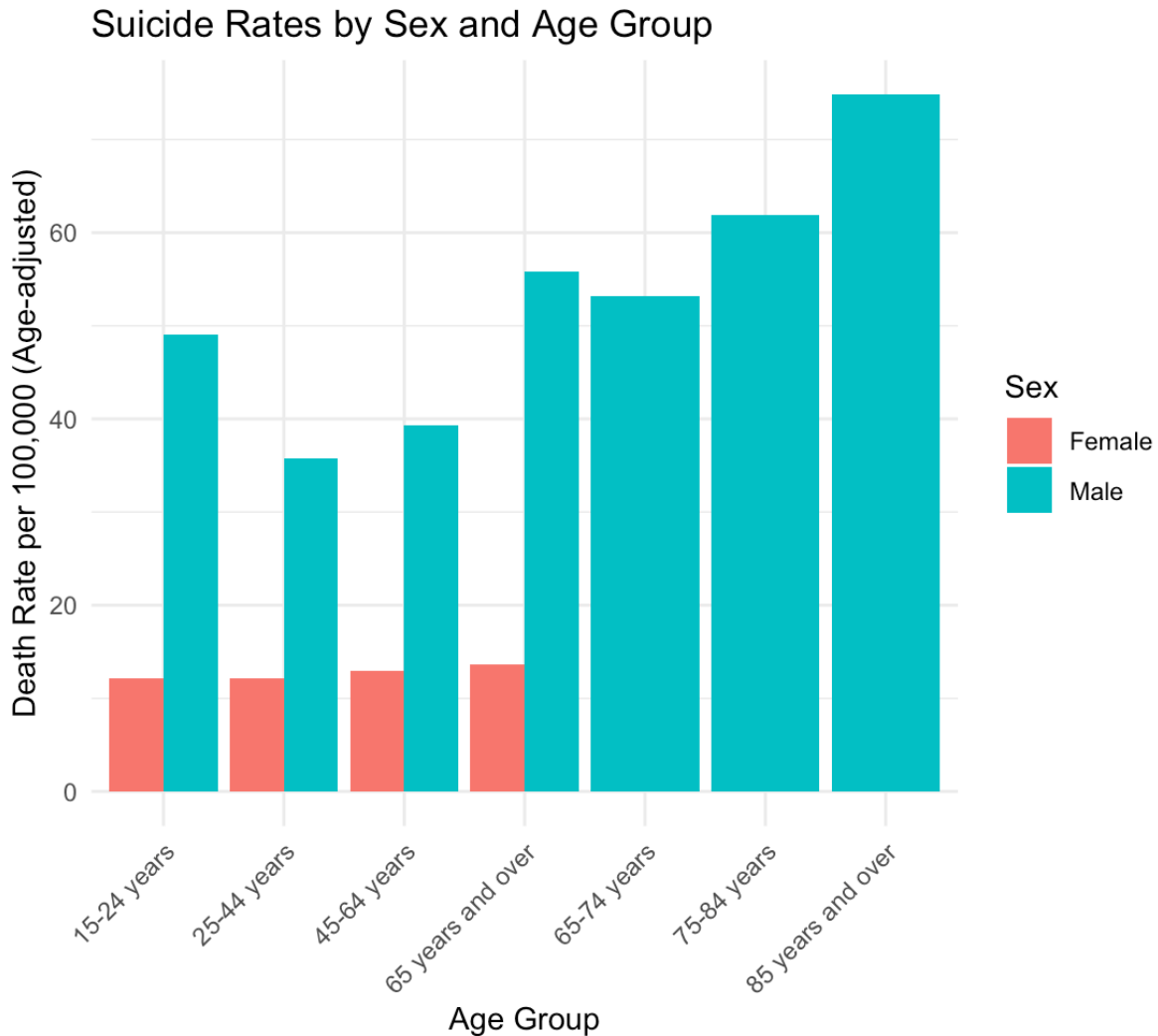


Figure 2: Suicide Rates by Sex and Age Group

Figure 2 displays the suicide rates by sex and age group. For the age groups 65-74 years and

the groups above, there was not sufficient data collected for women, thus only data for men in these age groups were displayed. The filtered data used to generate the summary statistics of the regression model shown in Figure 1 was used to generate the plot shown in Figure 2.

As we can see from Figure 2, there is a clear increasing trend in suicide rates with age for both males and females, especially for males. Each successive age group generally has a higher suicide rate than the preceding one. For every age group, males have a higher suicide rate than females. The disparity between males and females appears to be consistent across all age groups. The highest suicide rates for both sexes are observed in the “85 years and over” age group. The suicide rates for the youngest age group presented (“15-24 years”) are the highest among the younger age groups (up to 45-64 years old). The gap between male and female suicide rates appears to increase with age. Particularly in the oldest age group, “85 years and over,” the suicide rate for males is substantially higher than for females.

4.3 Suicide Rates by Race

Figure 3 displays the suicide rates by race. The races are divided into American Indian or Alaska Native, Asian or Pacific Islander, Black or African American, and White. The filtered data used in the Figure 1 was used to generate this plot.

From Figure 3, we see that the American Indian or Alaska Native group has the lowest suicide rate among the categories presented, as indicated by the shortest bar. The suicide rate for the Asian or Pacific Islander group is slightly lower than that of American Indian or Alaska Native but still among the lower rates compared to other groups. The Black or African American group shows a higher suicide rate than the previous two groups, as depicted by the taller bar. However, the White racial category exhibits the highest suicide rate by a significant margin, as indicated by the tallest bar in the chart.

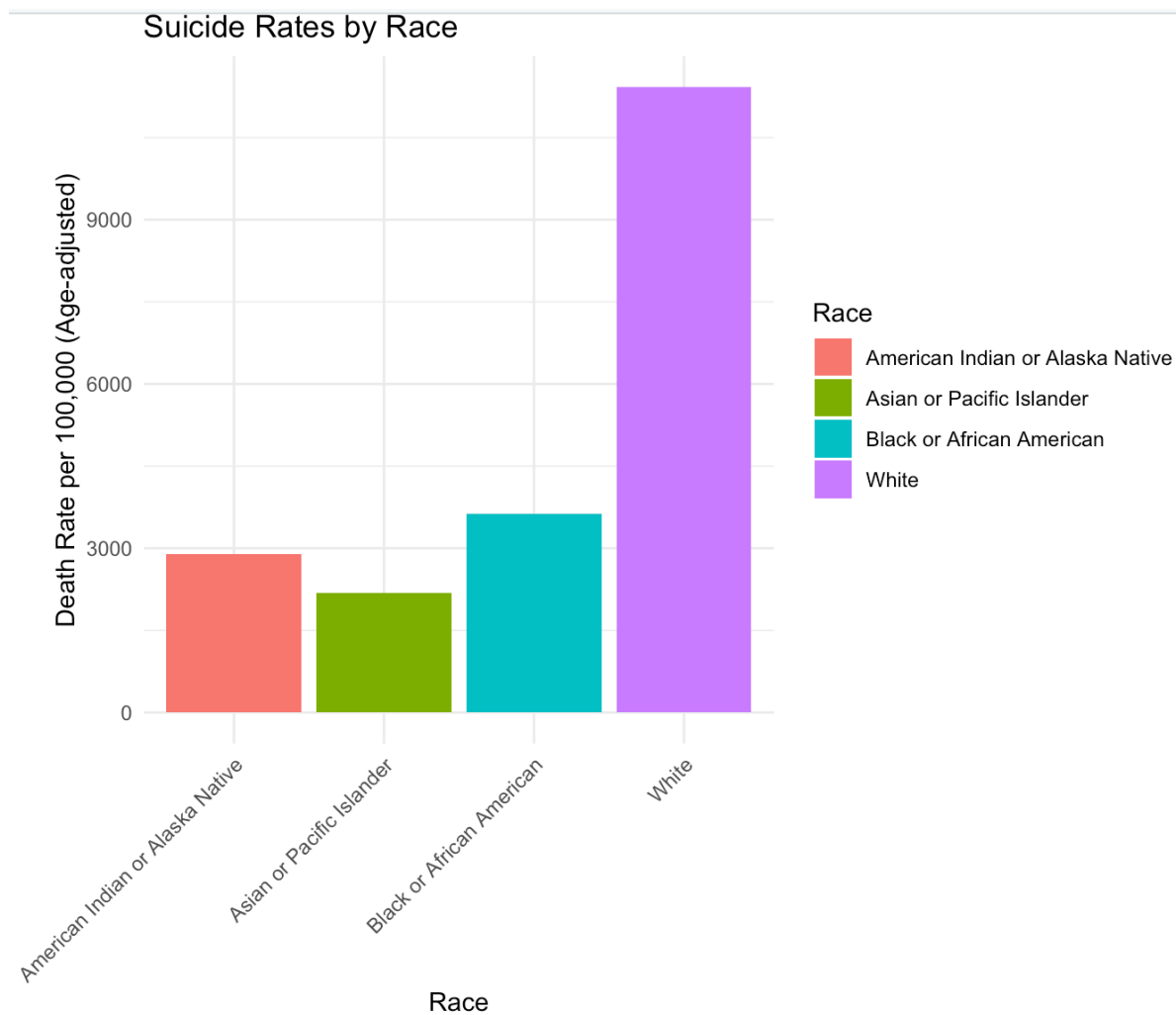


Figure 3: Suicide Rates by Race

5 Discussion

5.1 Findings

The data on suicide rates by sex, age, and race provide a reflection of broader sociocultural dynamics and health disparities. From Figure 2, males exhibit higher rates than females across all age groups, and suicide rates increase with age, with a marked peak in the oldest age group. The significantly higher rates of suicide among males may speak to societal norms regarding masculinity and mental health, where males may be less likely to seek help or express emotional distress. The rising rates with age, particularly the spike in the oldest age group, could indicate societal neglect of the elderly (social isolation), a lack of support systems in place for this demographic, or potentially the age-related psychological changes. These trends might also be interpreted as a reflection of the effectiveness or reach of mental health resources and public health interventions across different segments of the population.

From Figure 2, a particularly interesting observation is the unexpected suicide rate for males in the 15-24 years age group. This breaks the trend of a general increase in suicide rate as the age group increases. This may indicate certain pressures and struggles that individuals in this age group face compared to the others. For most people, these years play a significant role in developing us as individuals. In the early stages of this age group, this is a pivotal moment in a large proportion of people's lives as this is the time they start to decide on what they want to do with their lives. People in this age range are transitioning from childhood to adulthood, considering potential university opportunities and deciding on what career path they want to pursue. Although, this may be an exciting time for many, it can also be daunting time as this is typically the greatest change a person experiences at the time. In the Western world, people typically move out of their home to pursue their higher education or career, which comes with a lot of stress. Thus, this transitioning period may indicate the unforeseen pressure that young people face during this period, and the importance of employing measures to guide them through this process.

Additionally, as shown in Figure 3, there are notable differences in suicide rates among racial groups, with the highest rates observed in the White population. Differences in suicide rates across racial groups can be indicative of economic inequality and the varying degrees of access to healthcare services. For example, systemic barriers to healthcare, including mental health services, could disproportionately affect certain races, contributing to the observed disparities. High suicide rates in specific racial groups might also reflect stresses related to socioeconomic status, including employment instability, housing insecurity, and educational opportunities.

The data underscore the need for targeted public health strategies. High suicide rates in particular demographic groups suggest a need for tailored prevention programs that address the unique challenges faced by those groups. It highlights the importance of public health policies that are not only inclusive but also specifically designed to reach and support individuals who may be at greater risk due to demographic factors.

Comparatively, understanding suicide rates within the U.S. context can provide a benchmark for global health studies. Variations in suicide rates across countries may be influenced by factors such as cultural attitudes toward suicide, the availability and societal acceptance of mental health resources, and national healthcare policies. Thus, this analysis can contribute to a global dialogue on mental health and suicide prevention.

5.2 Limitations

The first limitation is the presence of reporting and selection bias within the dataset used. Cultural and societal factors can lead to systematic under reporting or misclassification of suicides, which varies by demographic factors such as race, sex, and age. There may be biases in who gets identified in official statistics as having died by suicide, influenced by factors such as race and socioeconomic status. Data from death certificates might not uniformly cover all segments of the population, particularly if there are sub-populations less likely to access healthcare or have their deaths formally registered.

As seen in Figure 3, the White group had a significantly higher rate of suicide compared to any other group. This may be due to the fact that the White population in the United States makes up approximately 75.5% of the total population compared to 13.6% for Black or African Americans, 1.3% for American Indian or Alaska Natives, 6.6% for Asian or Pacific Islanders. (U.S. Census Bureau 2023). As the White population is significantly the largest, the pool of reporting is in their favor, which may skew the findings.

The analysis shows associations between demographic factors and suicide rates but does not establish causality. Without understanding the direct causal mechanisms, interventions may not effectively target the root causes of the observed disparities.

The use of aggregated data can mask individual variability and sub-group differences within larger racial, age, or sex categories. For example, cultural diversity within the Asian or Pacific Islander group or generational differences within the “65 years and over” category may be overlooked.

5.3 Future Research

Future research that can be completed is the analysis of certain signs or indicators that an individual may commit suicide. This report has been able to identify which specific groups (seperated by age, race, and sex) are the most vulnerable to suicide. However, researching specific signs that an individual may be suicidal would have a great benefit in identifying high-risk individuals.

By having a greater understanding about the causality of suicide, we can then develop and implement prevention strategies and mental health interventions that are tailored to the unique

needs of each demographic group. This requires a granular approach that considers the cultural appropriateness and accessibility of mental health services.

Engaging with policymakers and community organizations to address broader social determinants is crucial. Policies that address economic inequality, improve access to mental health services, and foster social inclusion can help mitigate suicide risks.

References

- National Institute of Mental Health. 2024. “Suicide.” <https://www.nimh.nih.gov/health/statistics/suicide>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- . 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- U.S. Census Bureau. 2023. “QuickFacts.” <https://www.census.gov/quickfacts/fact/table/US/PST045223>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. <https://readr.tidyverse.org>.