Exploring Suicide Rates: Assessing Age, Sex and Race Patterns with Statistical Insights into Suicide Risk*

Sarah Lee

April 21, 2024

This paper investigates the intricate dynamics of suicide rates in the United States, with specific focus on how age, sex, and race intersects. Utilizing linear regression analysis and data from the Centers for Disease Control and Prevention (CDC), we unveil temporal patterns and disparities in suicide rates acrross demographic groups. Our analysis reveals that suicide rates tend to be higher among elderly male individuals within the white racial group; however, we also observe variations across intersections of age, sex and race. These findings offer valuable insights for policymakers, public health officials, and mental health professionals, facillitating the development of more targeted and strategic suicide prevention tailored to the diverse needs of the suicide population.

1 Introduction

The issue of suicide has garnered increasied global attention, affecting individuals, friends, families, and communities worldwide. Suicide rates serve as a critical indicator of a population's mental health and well-being, reflecting underlying social, economic, and cultural factors. Therefore, gaining a comprehensive understanding of suicide distribution across demographics is essential for targeted intervention and prevention strategies by the public health officials and professionals. However, while existing research acknowledges these disparities, a critical gap remains in understanding how specific factors intersect to influence suicide rates. Hence there is a need for specified research that consider each variables and their relationship to further elucidate answers and measures for a better future.

This paper aims to address this gap by examining the complex dynamics of suicide rates in the United States, focusing specifically on the nuanced intersections of age, sex, and race.

*Code and data are available at: https://github.com/sarahhhh02/suicides.git

Through the use of linear regression analysis, we aim to uncover how these demographic variables contribute to suicide risk factors.The estimand here is how age, gender and race are related to the risk factors of suicides. Our goal is to understand how age, sex, and race are related to suicide risk factors. Through this approach, we seek to identify which demographic groups are at heightened risk of suicide prevalence.

Drawing upon suicide datasets soured from the Centers for Disease Control and Prevention (CDC) (Disease Control and Prevention 2022), our study reveals temporal patterns and disparities in suicide rates across various demographic groups. By rigorous analyzing the data, we illuminate shifting dynamics in suicide prevalence among different age cohorts, sexes, and racial groups. Furthermore, by exploring the intersections of these factors, we reveal how they interact to shape the overall landscape of suicide in the United States.

The study focuses on answering the following questions:

- What trend do we see in the total suicide rates in US between 1950-2018?
- How do suicide rates vary across different age groups?
- How do suicide rates vary between sexes?
- How do suicide rates vary across age and sex combined?
- How do suicide rates vary across sex and race combined?
- How do suicide rates vary across age, sex, and race?

Understanding the trends and disparities in suicide rates is crucial for developing effective prevention and intervention strategies tailored to the diverse needs of vulnerable populations. By elucidating the intricate relationships between age, sex, and race in the context of suicide, this paper contributes to the broader discourse on mental health and public health policy. In the subsequent sections, we delve deeper into the methodology employed, present our findings, discuss their implications, and propose avenues for future research, thereby offering valuable insights into addressing this pressing public health issue.

2 Data

2.1 Data Source

The data utilized in this paper focuses on suicide rates in United States from from 1950 to 2018, sourced from the CDC's database on Death rates for Suicide, categorized by sex, race, Hispanic origin, and age (Disease Control and Prevention 2022). This publically acessible data source contains information concerning suicide rates across various demographics, and can be freely downloaded from the CDC's website in multiple export file formats. The database is

compiled by the National Center for Health Statistics (NCHS) and DAE, sampling from a population of 100,000 US residents.

For the analysis of this paper, we utilized the R programming language (R Core Team 2023) for statistical computation and data visualization. The tidyverse package(Wickham et al. 2019) was installed to access to other R packages necessary for this paper's analysis. Additional packages used include dplyr (Wickham et al. 2023), ggplot2 (Wickham 2016), knitr (Xie 2014), rstanarm (Goodrich et al. 2022) and arrow (McKinney et al. 2022).

2.2 Variables of Interest

Our analysis focuses specifically on investigating the demographic factors of age, sex, and race between the years 1950 to 2018. We utilized data from a population of 100,000 US residents, excluding any age-adjusted populations to ensure a more accurate representation of the population. Furthermore, we cleaned the data to separate the sections of the raw data into each selected factors of age, sex, race, and their combination. It is worth noting that Hispanic origin was not included as a specific variable in this research.

3 Model

Through the data analysis in this paper, we identified correlations between suicide rates and demographic factors such as age, sex, and race. To further analyze the implications of these demographic factors with suicide rates, we constructed a linear regression model.

3.1 Model set-up

This model fits a linear regression model where year is the independent variable and suicide rates are the response variable. We run the model in R R Core Team (2023) using the `rstanarm` package of (Goodrich et al. 2022).

$$y_i = \beta_0 + \beta_1 \text{Sex} + \beta_2 \text{Age} + \beta_3 \text{Race} \tag{1}$$

Define

- y_i as the suicide rate estimated by year
- β_0 is the coefficient for intercept
- β_1 is the coefficient for the rate at which suicide rates change with respect to sex
- β_2 is the coefficient for the rate at which suicide rates change with respect to age

- β_3 is the coefficient for the rate at which suicide rates change with respect to race

3.2 Model justification

We anticipate that the linear regression model will aid in understanding the relationship between suicide rates and various demographic variables such as sex, age, and race. Each coefficient in the model represents the change in the response variable associated with a one-unit change in the predictor variable (age, sex, and race), while holding other variables constant. By utilizing this model, we can determine which demographic factors have a significant impact on suicide rates. The significance of each coefficient can be assessed using hypothesis test, such as t-tests.

Moreover, when incorporating multiple predictor variables in the model, we can control for potential confounding variables. To illustrate, by including both age and sex in the model, we can assess the independent effect of each variable on suicide rates while controlling for the other. This model can also capture the interaction effects between predictor variables., demonstrating the influence of age on suicide rates based on the different sexes and racial group.

4 Results

4.1 Total Suicide Rates in United States

The overall trend in Suicide Rates in United States are shown below in Table 1. Indicating us an estimated number of suicide deaths per year.

Table 1: Total Suicide Rates per 100,000 population by Year

Year	Estimate
2018	14.8
2017	14.5
2016	13.9
2015	13.7
2014	13.4
2013	13.0
2012	12.9
2011	12.7
2010	12.4
2009	12.0
2008	11.8
2007	11.5

Year	Estimate
2006	11.2
2005	11.0
2004	11.1
2003	10.9
2002	11.0
2001	10.7
2000	10.4
1999	10.5
1998	11.1
1997	11.2
1996	11.5
1995	11.7
1994	11.8
1993	12.0
1992	11.9
1991	12.2
1990	12.4
1989	12.2
1988	12.4
1987	12.7
1986	12.9
1985	12.4
1984	12.4
1983	12.1
1982	12.2
1981	12.0
1980	11.9
1970	11.6
1960	10.6
1950	11.4

The plot in Figure 1, provides a snapshot of suicide trends by year. We observe a gradual increase in suicides from the year 2000, with the highest numbers recorded by 2018. Table 1 presents the accurate numbers of suicide rate estimates, indicating an estimate of 14.8 suicides per 100,000 individuals in 2018, and the lowest count recorded in 2000 with 10.4 estimations. In the subsequent sections, we will delve into the trends of suicide rates according to each demographic factors.

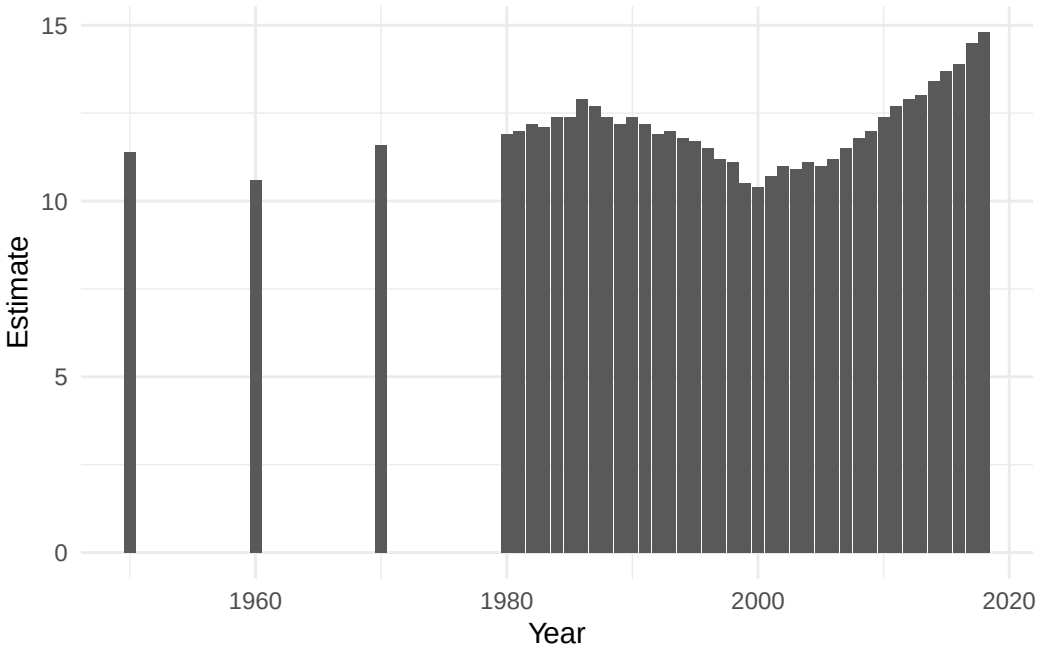

Figure 1: Total Suicide Rates per 100,000 Population by Year

4.2 Trends in Suicide Rates by Sex

There is a notable disparity in suicide rates between the two sexes as illustrated in Figure 2. Males exhibit a significant higher risk of suicide compared to females.

```
Warning in geom_line(data = cleaned_sex, aes(x = YEAR, y = ESTIMATE, color =
SEX, : Ignoring unknown aesthetics: fill
```

The estimated baseline suicide rate, represented by the intercept, is approximately 5.0, indicating the average suicide rates when not considering sex as a factor. However, when factoring in the effects of sex, the estimated suicide rates for males rises to approximately 14.3. That represents 34% increase compared to the average suicide rate. This highlights the significant emphasis on the disparities in suicide rates between sexes.

4.3 Trends in Suicide Rates by Age

Overall, we observe a decline in suicides across all age ranges from 1950 to 1990s, except for the 10-14 age group. However, starting from the 1990s, we notice an increase in suicides extending into the early 2000s, followed by a decrease until around 2010. Subsequently, there is another increase from then until 2018. The 10-14 age group stands out from this pattern, exhibiting consistency with a slight increase starting from the early 2000s.

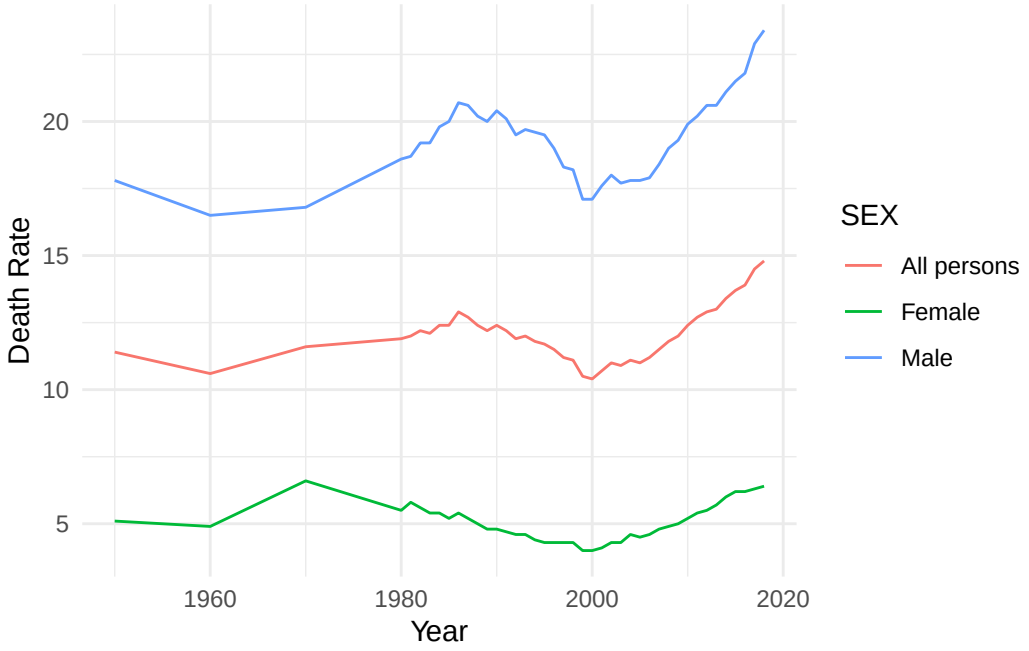

Figure 2: Crude Death Rates per 100,000 Resident Population by Sex

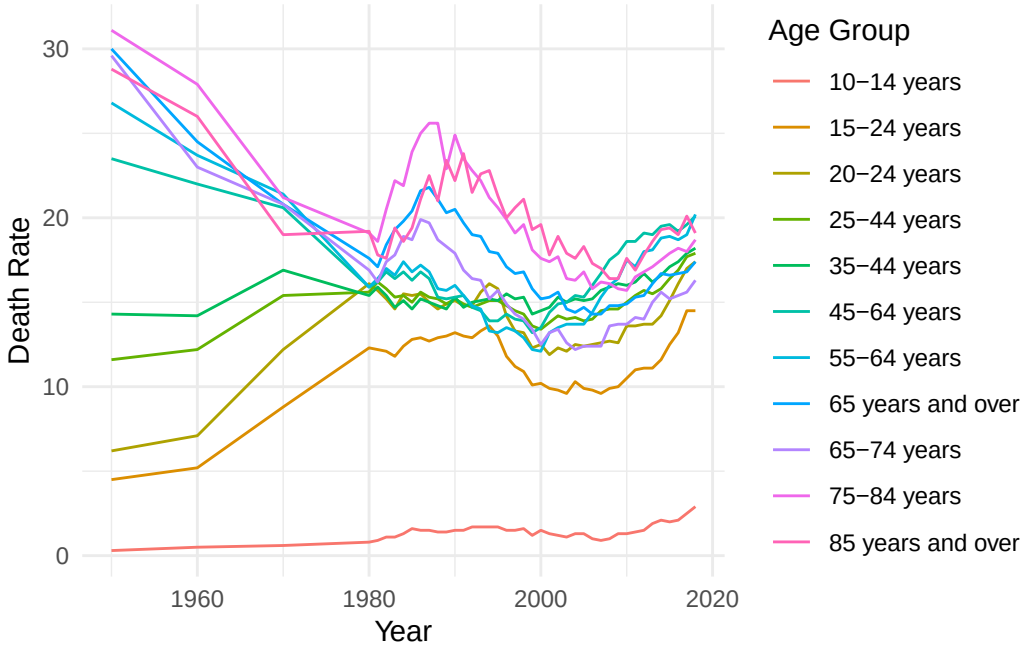

Figure 3: Crude Death Rates per 100,000 Resident Population by Age

Based on the data provided in Figure 3, we observe that, apart from ages 10-14, all other age ranges show a similar pattern in suicide rates. Further analysis reveals that younger age groups, such as those between 15-24 years, tend to have lower suicide rates compared to older age groups. However, there is a notable increase in suicide rates among the elderly population, particularity those aged 75 year and above.

When examining the mean of suicides per age group, we find that the lowest mean among the elderly is 18.6 which is higher than of younger age groups, with the highest mean being 12.4. Moreover, within each age range, there is a variability in suicide rates that may reflect diverse factors that influence suicide risk within specific age groups. these factors include socioeconomic status, mental health issues and access to health support.

4.4 Relation to Suicide Rates by Sex and Age

The estimated baseline suicide rates, when controlling for sex and age predictors, is approximately -9.6. This represents the suicide rate when both sexes and age are at their reference levels. For males, the estimated effects on suicide rates are approximately 21.5, suggesting that males have significantly higher estimated suicide rates compared to females when other variables are held constant.

Furthermore, the plot shown in **?@fig-4** suggests an opposite interaction between the years from 1980 to 2018. While males on average exhibit a concave down pattern, females present a concave up trend.

Visibility we see that suicide rates vary across different age groups. Generally, suicide rates increase with age, as previously observed in Figure 3, with older age groups exhibiting a higher estimated suicides compared to younger age groups.These findings underscore the importance of considering both gender and age when examining suicide rates, as they offer valuable insights into the complex interplay of demographic factors influencing suicide risk. Moreover, they highlight the importance of tailored interventions targeting specific demographic groups to effectively prevent suicide.

4.5 Relation to Suicide Rates by Sex and Race

Displayed below in Figure 6, we can observe the relationship between sex and race through a bar graph. the graph compares females and males among the selected years from 2000, 2010, and 2017 across different racial groups.

The estimated mean response when all predictor variables are zero, is 6.1. Specifically focusing on males, the estimated difference in the response variable between males and females are found to be 10.7, indicating that males have a higher response by this amount compared to females. Additionally, individuals identified as Asian or Pacific Islanders and Black or African American races are estimated to have lower responses by 5.4 and 5.3, respectively. This suggests that

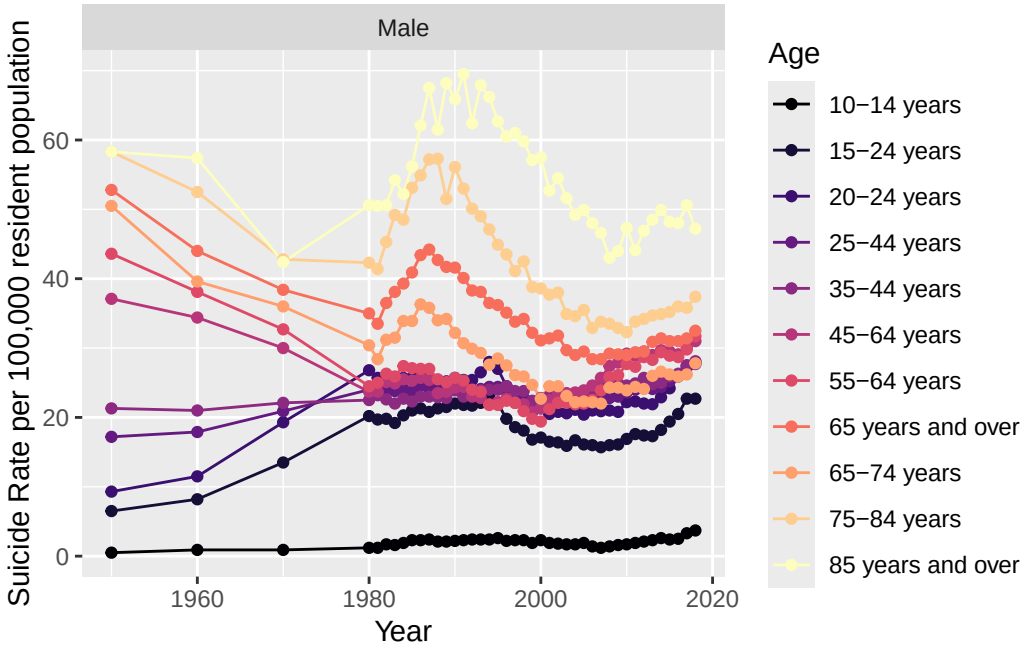

Figure 4: Suicide Rates per 100,000 Population by Sex and Age

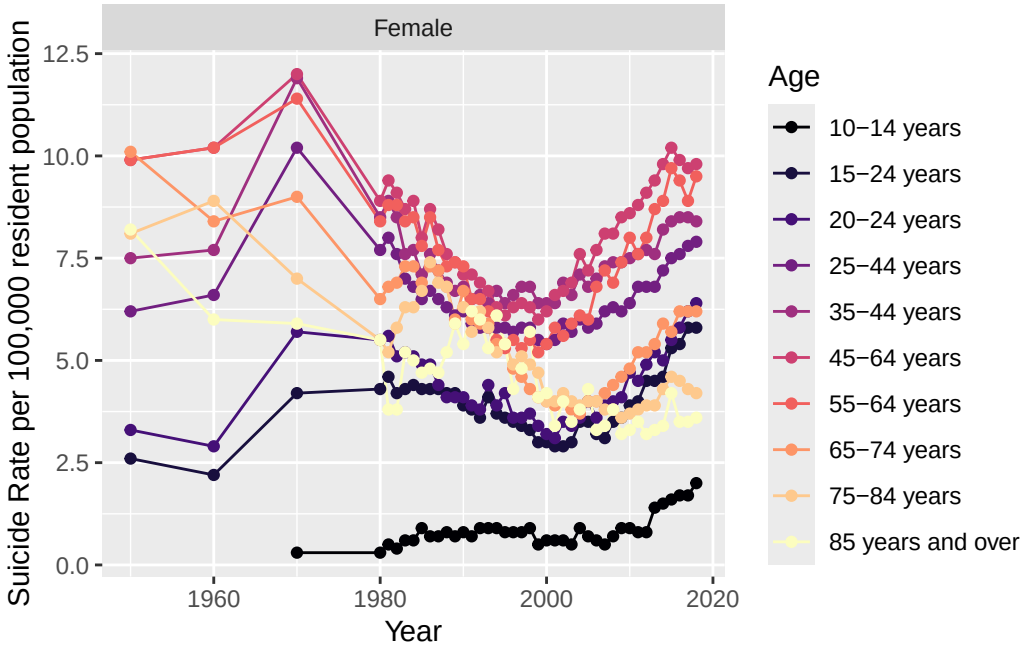

Figure 5: Suicide Rates per 100,000 Population by Sex and Age

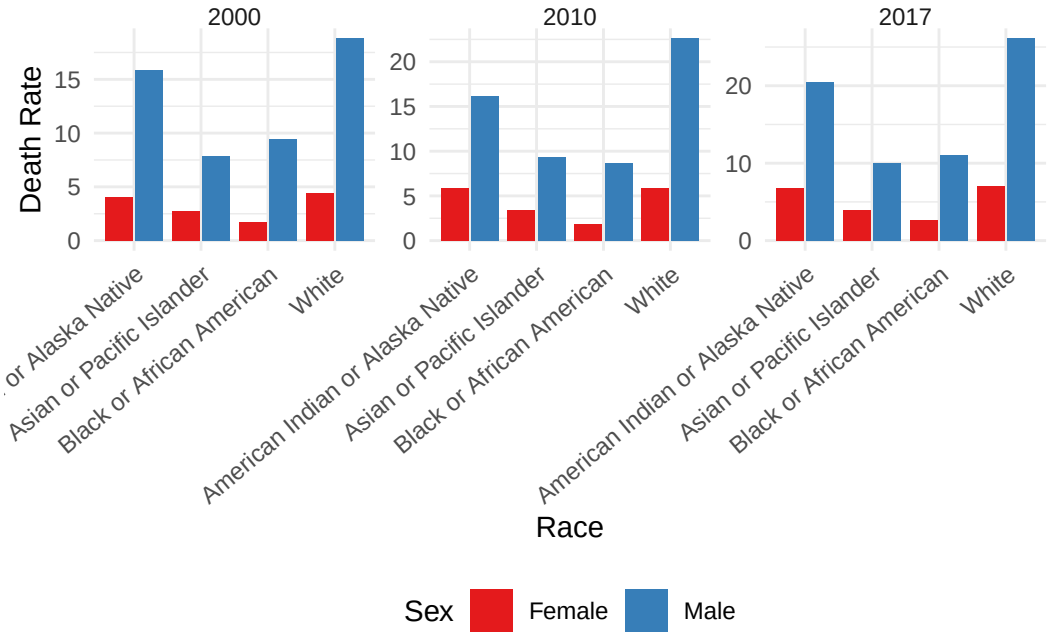

Figure 6: Suicide Rates per 100,000 Resident Population by Sex and Race

individuals of the White race are estimated to have a higher response by 2.1 compared to the other racial groups. Visually, American Indian or Alaska Native individuals are depicted as having the second highest response, as represented in the graph.

4.6 Relation to Suicide Rates in Sex, Age, and Race

Analyzing suicide rates in the United States over the decades, we observe fluctuations based on the combination of race, age and sex. these fluctuations reveal both increases and decreases influenced various factors such as changes in societal attitudes towards mental health, economic conditions, access to healthcare, and public health initiatives. Below, in Figure 7, we visualize age, sex, and race in one graph representing suicide rates.

When examining suicide rates in the United States based on sex, age, and race alone, notable trends emerge. Female suicide rates exhibit a degree of consistency over time, while male rates demonstrate variability across different age brackets and racial backgrounds. It is evident that males consistently show a higher suicide rate compared to females. Age serves as a significant factor influencing suicide rates, with historical data indicating a prevalence of suicide among older adults, particularly those ages 65 and above. Nonetheless, recent years have witnessed a concerning uptick in suicide rates among younger demographics, notably young adults.

Racial disparities also play a role in shaping suicide rates. Historically, White individuals have exhibited higher suicide rates compared to other racial groups in the United States. However,

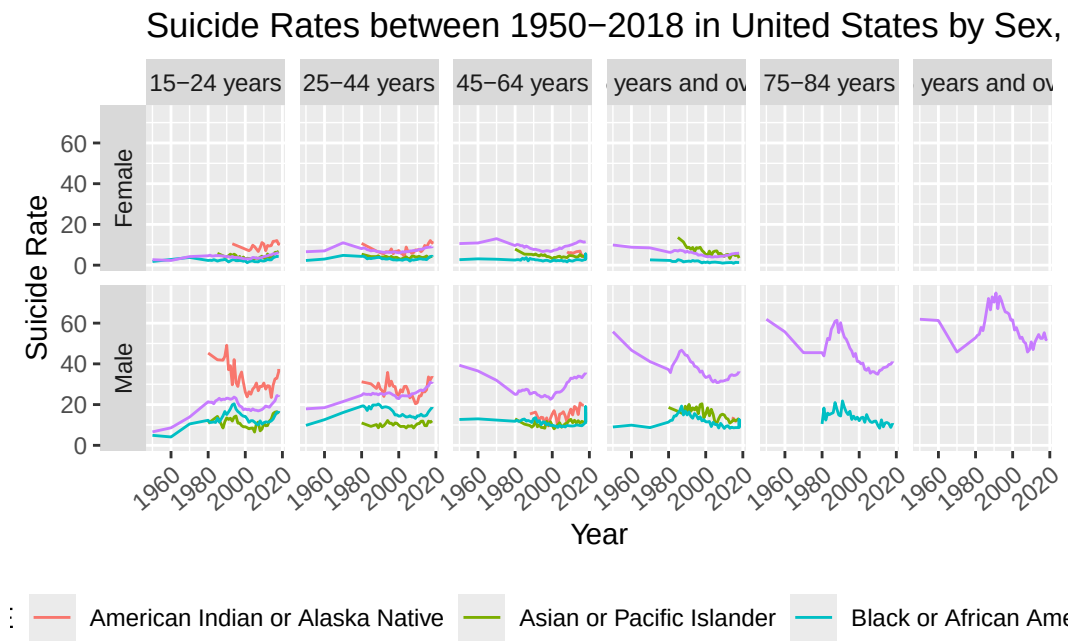

Figure 7: Suicide Rates per 100,000 Resident Population by Sex, Age, and Race

there has been a disproportionate rise in suicide rates among certain racial minority groups, such as Native Americans and Alaska Natives, in recent years.

In our investigation of suicide rates in the United States, we utilized a regression model to discern significant patterns across demographic factors such as age, sex, and race. The summary in Table 2 provides intriguing insights into the relationship between age and the dependent variable.

Notably, individuals aged 65 years and over exhibit a substantial increase in the dependent variable, with coefficients of 2.59. However, this effect varies across different age brackets within this group. While individuals aged 65-74 years display a decrease in the dependent variable, those aged 75-84 years and 85 years and over demonstrate significant increases of 11.23 and 33.91, respectively. These findings underscore the nuanced impact of age on the outcome, with older age groups exhibiting divergent associations. Another key demographic variable examined in the model is sex. The coefficient associated with being male is 14.02, indicating a notable increase in the dependent variable compared to females. This finding suggests a potential gender-based disparity in the outcome, warranting further investigation into the underlying factors driving this divergence. The model also sheds light on the influence of race on the dependent variable. Notably, individuals belonging to Asian or Pacific Islander and Black or African American racial groups exhibit negative coefficients of -9.32 and -10.99, respectively. In contrast, individuals identified as White display a positive coefficient of 0.84. These findings highlight disparities in the outcome across different racial groups, emphasizing the importance of addressing racial equity in various domains.

Table 2: Explanatory models of flight time based on wing width and wing length

	Model
(Intercept)	9.72
	(0.60)
AGE25-44 years	0.98
	(0.50)
AGE45-64 years	−0.16
	(0.51)
AGE65 years and over	2.59
	(0.56)
AGE65-74 years	−1.63
	(1.06)
AGE75-84 years	11.23
	(0.79)
AGE85 years and over	33.91
	(1.09)
SEXMale	14.02
	(0.36)
RACEAsian or Pacific Islander	−9.32
	(0.64)
RACEBlack or African American	−10.99
	(0.61)
RACEWhite	0.84
	(0.60)
Num.Obs.	1266
R2	0.809
R2 Adj.	0.807
Log.Lik.	−4079.480
ELPD	−4091.0
ELPD s.e.	36.1
LOOIC	8181.9
LOOIC s.e.	72.1
WAIC	8181.9
RMSE	6.06

Assessing the overall fit of the model, the R-squared value of 0.809 indicates that a substantial proportion of the variability in the dependent variable is explained by the demographic factors included in the model. Furthermore, measures such as the adjusted R-squared, log likelihood, and root mean square error provide additional insights into the model's predictive performance and goodness of fit.

In conclusion, the findings of the statistical model underscore the significance of demographic factors in shaping outcomes of interest. Age, sex, and race emerge as important predictors, each exhibiting distinct associations with the dependent variable. They provide valuable insights into the complex factors that contribute to suicide risk and resilience in the United States. Various fields, including healthcare, sociology, and public policy, highlighting the importance of considering demographic diversity in research and decision-making processes.

5 Discussion

This paper presents a comphrehensive statistical model of the relationship between demographic factors - age, sex, and race - and their impact on suicide rates. Our objective is to uncover how these variables contribute to variations in the outcome under study. Initially, we conducted separate analysis for each intersection on the demographic factors, we made initial observant for each factor alone. Concluding how demographic factors to discern their individual influence on suicide rates . This provides understanding into how age and sex intersect to shape suicide risks. Subsequently, we explored intersections within the demographic factors to identify any changes in variation.

One significant insight gleaned from our analysis is the varying impact of age on the dependent variable. Older age groups, notably those aged 75-84 years and 85 years and over, exhibit significant increases in the outcome, emphasizing the critical role of age as a predictor in understanding suicide risk. Additionally, our analysis highlights the substantial increase in the dependent variable among males compared to females, indicating potential gender-based disparities that necessitate further exploration and intervention. Overall, our findings suggest that White male individuals among other age groups are at risk of suicide rates. These observations underscore the multifaceted nature of suicide risk and emphasize the importance of tailored interventions that address the unique needs of diverse demographic groups. The intersection of race, age, and sex can significantly influence suicide risk, with older men from certain racial minorities facing distinct challenges and risk factors.

Effective suicide prevention strategies should prioritize targeted support systems, mental health resources, and initiatives aimed at reducing stigma and promoting well-being across all segments of society. Socioeconomic factors, including poverty, unemployment, social isolation, and access to healthcare, may contribute to suicide risk factors across demographic groups, underscoring the need for comprehensive approaches to address these underlying determinants.

5.1 Weaknesses and next steps

Despite these insights, several weaknesses in our study are apparent. Firstly, while our model provides valuable associations between demographic factors and the outcome, it may lack sufficient granularity to capture the full complexity of these relationships. Additionally, the reliance on cross-sectional data limits our ability to establish causality or account for temporal dynamics. Moving forward, future research should strive to address these limitations and expand upon our findings to develop targeted interventions aimed at promoting equity and inclusivity across diverse populations.

Longitudinal studies could provide insights into the temporal trajectories of demographic influences on the outcome, allowing for a more nuanced understanding of these relationships over time. Moreover, incorporating additional variables, such as socioeconomic status or health behaviors, could enhance the model's explanatory power and predictive accuracy. Advanced statistical techniques, such as structural equation modeling or machine learning algorithms, could facilitate the identification of latent patterns and interactions among demographic factors and the outcome.

Qualitative research methods, such as interviews or focus groups, could complement quantitative analyses by providing deeper insights into the lived experiences and perspectives of individuals within different demographic groups. In conclusion, while our paper offers valuable insights into the relationship between demographic factors and suicide rates, there remain opportunities for future research to deepen our understanding of these complex phenomena and inform evidence-based interventions aimed at promoting equity and inclusivity in society.

References

Disease Control, Centers for, and National Center for Health Statistics Prevention. 2022. "Death Rates for Suicide by Sex, Race, and Hispanic Origin: United States, 1999-2019." United States: National Center for Health Statistics. 2022. https://data.cdc.gov/NCHS/Death-rates-for-suicide-by-sex-race-Hispanic-origi/9j2v-jamp/about_data.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

McKinney, Wes et al. 2022. *Arrow: In-Memory Columnar Data*. United States: Apache Software Foundation. https://arrow.apache.org/docs/r/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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 Grolemund, 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.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.