Exploring the Relationship Between Proxies of Social Capital Loss and Suicide In the United States (2000-2022)

Research Question: *What is the relationship between social capital loss proxies and suicide rate in the United States?*

Authors: Mario Haya 羅俊杰, Ignatius Harry Cahiadharma 柯昱安, Marcella Alifia Kuswana Putri 潘怡璇, Xóchitl Gutiérrez 顧修齊

Abstract

This study investigates the relationship between the loss of social capital and the increasing suicide rates in the United States from 2000 to 2022. We use Pearson correlation coefficient (PCC), Ordinary Least Squares (OLS), fixed-effect regression model, and both interaction and non-interaction methods to get the causal effect on six different proxies of social capital loss and suicide on a county-level. The results of this study show a moderate positive association between suicide rates and the number of hate groups, as well as a positive correlation between higher median household income and suicide rates. Higher poverty rates are also correlated with increased suicide rates, whereas unemployment rates unexpectedly shows a negative correlation with suicide rates. These findings highlight the complex interplay between socioeconomic factors and suicide rates, suggesting that comprehensive policy responses are needed to address these issues effectively.

The code and data for this paper can be accessed at https://github.com/IgnatiusHarry/Social-Capital-in-the-US

1. Introduction

The United States has become the country with the highest suicide rate among wealthy nations, and has seen a consistent increase in suicide rates in all age groups and demographics during the past two decades. (CDC, 2022). This trend, along with the various contextual level factors, is of great concern to policy makers (Steelesmith et al. 2019). Social capital, defined as the norms, networks, and trust that enable collective action, has been identified as a potential protective factor against suicide. Prior studies have consistently found that higher levels of social capital at the community level are associated with lower suicide rates (Fontanella et al., 2018). This suggests that social cohesion and civic engagement may play a crucial role in supporting individuals and buffering against suicidal behaviors.

However, the loss of social capital can have detrimental effects on individuals and communities, potentially contributing to the rise in suicide. Given its historical institutional racism, hate groups, and racial tensions, the United States is a fit candidate to measure hate group formation. Over the past two decades, the number of hate groups has alarmingly surged by 104.5%, doubling from 599 groups in 2000 to 1,225 in 2022 (CDC, 2022); and we found that the majority of which are self-described as white supremacist and anti-government in nature. The consistent increases in hate groups thus serve as a potential proxy for social capital loss in the United States. This highlights the complex relationship between social capital loss and suicide, which needs further investigation to inform more comprehensive suicide prevention strategies.

To investigate the complex relationship empirically, we employed a robust empirical strategy. First, we use Pearson correlation coefficients (PCC) to examine the bivariate association between social capital loss and county-level suicide rates. Next, we use Ordinary Least Squares (OLS) and fixed-effect regression models to estimate the causal effect of social capital loss on suicide, controlling for relevant social and economic-related factors. Finally, we explore potential interactions between social capital loss and other contextual factors to better understand the complex pathways linking these factors (Tadmon & Bearman, 2023). We utilize a comprehensive county-level dataset of the United States. This allows us to examine the relationship between social capital loss and suicide at the county level.

Our analysis reveals a statistically significant and robust relationship between the loss of social capital and increased suicide rates across US counties. The Pearson correlation coefficients show a moderate positive correlation between the number of hate groups and suicide rates, as well as higher median household income and suicide rates. The regression results then confirm the remaining variables of social capital loss used in the study, have a significant causal effect on suicide.

This research contributes critical insights to the existing literature on the relationship between social capital and suicide. Our analysis identifies key contextual factors, such as population size, number of hate groups, median household income, poverty rate, and unemployment rate. These insights can inform more targeted and effective suicide prevention strategies that address both individual-level vulnerabilities and broader community-level determinants. Our findings have significant implications for policymakers and public health practitioners, highlighting the need for interventions in order to address the growing public health crisis of suicide in the United States.

Finally, our motivation for this study lies in the need establish a more comprehensive understanding of the relationship between social capital loss and suicide rates in the United States. Conclusive research on why suicide rates have dramatically increased in the past two decades is not prevalent, however existing literature suggests that social capital is a proper way to understand this phenomenom. We initially considered analyzing this phenomenon in the European Union, but eventually decided against this due to the United States' unique political climate, characterized by rising racial tensions, divisive rhetoric and the proliferation of hate groups. It is therefore crucial to investigate how such factors may be influencing suicide rates across different communities.

2. Literature Review

2.1 Hate groups and Hate Group Formation

There is scarce sociological literature that has addressed social capital loss in the United States and its potential relationship with increased suicide rates. The most prominent work on this is "Suicide, Social Capital, and Hate Groups in the United States" (2021) by Brendan Szendro. Szendro's work provided a useful framework for our investigation, despite many of its shortcomings. For instance, Szendro (2021) chose to focus on only one proxy of social capital loss: hate groups. We added onto this in two ways: First we increased the number of years considered for hate groups; Szendro used 2010 to 2019, whereas our investigation covered 2000-2020. Finally, it should be noted that Szendro's work found a strong correlation between both variables, where every 1.92 percent increase in suicide rates resulted in one additional hate group. This provided our investigation with a stronger theoretical foundation.

On the other hand, most research on the social determinants of extremism and political radicalization has identified the erosion of community ties, economic insecurity, and feelings of marginalization as key factors that can contribute to the appeal of hate groups and extremist ideologies (Maulana & Wardah, 2023). These are all synonymous with social capital loss. When individuals experience a breakdown in their social support networks and a loss of a sense of belonging, they may be more vulnerable to false promises of extremist groups. Moreover, the decline of social capital and community cohesion can also create an environment where hate groups are able to more easily spread their message, as a lack of strong social bonds and civic engagement can make it harder for communities to collectively resist and counter such threats (Smith, 2018).

Ultimately, the available literature suggests that the loss of social capital in the United States, manifested through declining community involvement, reduced social cohesion, and diminished interpersonal trust, may be a significant contributing factor to the rise of suicide rates as well as the proliferation of hate groups and extremist ideologies. Addressing these complex challenges will require a multifaceted approach that prioritizes the rebuilding of strong, resilient communities and the restoration of social capital at the local and national level.

2.2 Suicide in the United States

Literature on the epidemiology of suicide and suicidal behavior reveals several key trends and risk factors that may be exacerbated by the loss of social capital in the United States (Nock et al., 2008). Existing research has consistently found that higher levels of social capital are associated with better physical and mental health outcomes, including lower suicide rates (Steelesmith et al. 2019). Conversely, the loss of social capital, manifested through declining community involvement and reduced social cohesion can lead to a lack of social support - all of which are well-established risk factors for suicidal behavior.

Other factors unrelated to social issues, such as economic status, household income, and inequality, were also considered in this study. For example, a study by Iemmi et. al. (2016) found a strong negative correlation between income and suicide rates, and for this reason we included household income as one of our variables. Once again, this demonstrates the complexity of the issue and how different factors can contribute to increases in suicide.

Further, other studies have shown that higher levels of unemployment, poverty, and lower median household incomes are associated with elevated suicide rates, particularly in non-metropolitan and economically-distressed areas. (Casant & Helbich, 2022) This relationship can be explained by the significant psychological and social impacts of economic hardship, including increased stress, hopelessness, and social isolation, which also matches the variables chosen for our investigation.

Moreover, research has revealed that population density, or the degree of urbanization, also plays a crucial role in shaping suicide trends (Kegler et al., 2017). Suicide rates tend to be higher in less urban areas, likely due to the compounding effects of limited access to mental health resources, economic opportunities, and social support networks. These disparities in suicide risk factors can create a vicious cycle, as the loss of social capital and community resilience in economically disadvantaged and rural regions can further exacerbate mental health challenges and suicidal behaviors.

Taken together, the existing literature highlights the critical importance of addressing factors such as poverty, unemployment, and community-level dynamics in order to effectively tackle the rising tide of suicide in the United States. Investing in the rebuilding of social capital and community resilience may be a crucial component of these efforts, as strengthening social connections, civic engagement, and a sense of belonging can serve as a vital protective factor against suicidal behaviors.

2.3 Hate Crimes

Our investigation is unique in that it includes other possible determinants of suicide that are not normally considered, such as hate crimes. To the best of our knowledge, it is the first study that includes this variable together with other hate-related variables like hate groups. While there is some literature that has pointed at a possible correlation between hate crimes and suicide, such as the work of Prairie et al. (2023), their study only focuses on adolescents, and thus pertains to an age-related investigation.

Conversely, victims of hate crimes are also known to be at risk of committing suicide, according to several studies (Cramer et al., 2018, Duncan & Hatzenbuehler, 2014). While this correlation is well-established, it is not relevant to our study because we investigate suicides as a whole, in both victims and perpetrators of hate crimes; not just victims.

However, it was ultimately impossible to include this variable as a proxy of social capital loss in the regressions. This is because there were several instances of hate crimes being reported in counties where hate groups did not form, therefore adding thousands of rows with missing values. This ultimately skewed our overall results and thus cannot be used to measure the correlations with other variables.

3. Methodology

3.1 Variables

As mentioned, this research aims to explore the correlation between factors contributing to social capital loss and suicide rates in the United States. The study examines five variables linked to social capital loss, aiming to clarify their influence on the increasing incidence of suicides at the county level from 2000 to 2022:

It should be noted that religious affiliation was originally considered as the sixth proxy for social capital loss, however, this is impossible to measure reliably. The United States Religious Census takes place every 10 years, therefore, only data from 2010 and 2020 was available to us. In other words, adding religious affiliation would make the regression results unreliable and insignificant.

Moreover, the study utilizes cross-sectional panel data regression analysis and computes the Pearson Correlation coefficient (PCC) to substantiate its findings. A cross-sectional study will examine the dynamics of the correlation between risk factors and effects by approaching, observing, or collecting data at one time, namely each research subject will only be observed once. Initially, we analyze the Pearson Correlation, a metric that assesses the magnitude and direction of linear relationships between two variables. Furthermore, to enhance the relevance and depth of our research, we employ Ordinary Least Squares (OLS) and Fixed Effects regression models to mitigate unobserved heterogeneity and offer a nuanced analysis of the temporal dynamics between the variables. These models, specifically the fixed effect regression models, are segmented into specifications with and without lagged variables and interaction terms.

3.2 Research Design

3.2.1 Data collection

Our data was collected from a wide range of secondary sources. For the hate groups variable, data was collected from the Southern Poverty Law Center (SPLC) Hate Map website, which is the only database that actively tracks hate groups in the United States. However, its tracking method is severely inconsistent, with thousands of mistakes, such as mislabeled cities, arbitrary or undefined regions (i.e. "Northeast Texas"), spelling mistakes, and mismatching cities and states. These were all issues addressed during the data cleaning process (see section 3.2.2).

Data for hate crimes (which we were ultimately unable to use in our study) is more consistently published by the Federal Bureau of Investigation (FBI) and can be found on their hate crimes database website (known as "Table 13", which refers to this specific type of crime). However, "Table 13" data only spans 2004 to 2019, but it still provides a reliable measure of hate crime activity. As mentioned, this dataset also represented a significant issue for the investigation, because there were several crimes recorded where hate groups were not recorded, and vice-versa. This created thousands of rows with missing data (*NA* values) that distorted our results. Therefore, we made the decision to ultimately exclude it from the investigation.

Data pertaining to economic variables such as unemployment, median household income, and poverty rates were obtained from the Bureau of Labor Statistics, which is another well-documented database with specific information on exact locations (including cities and counties, which is something that the SPLC database lacked). The data is categorized on a state, city, and county-level, but the latter was chosen for the sake of consistency in our investigation.

Finally, data for suicides in the United States is well-documented and has been consistently by the Centers for Disease Control (CDC), which is available on their website as well. Similarly population data may be collected from various government agencies, including the FBI and the CDC. This process was relatively easier because the previously mentioned sources already included population data within their datasets.

The variables above are used to create US maps. We utilized several packages, including SF, Tmap, and Tigris, to create the map.

3.2.2 Data Cleaning

This study's sample comprises an unbalanced panel dataset of all counties in the United States, excluding its territories and Alaska , from 2000 to 2022.

This research's total number of observations ranges from *43,428* to 74,496, depending on the dataframe analyzed, but the minimum number of observations is the number of United States

counties (around 3,000 depending on the year observed) multiplied by the number of years analyzed. Therefore, a high number of observations represents a significant strength for our analysis. However, this also required substantial time and effort for data cleaning.
The data-cleaning process was significantly time-consuming due to the quality of the sources we used. It can be summarized in the following macro steps:

1. **Deleting Missing Values**: Missing values were removed, including all values corresponding to "Not applicable," "Don't know," and "No answer." However, for the suicide rate, values labeled as "NA" (Not Applicable) were changed to zero, following the CDC website's clarification that zero suicides in a specific area should be recorded as zero.
2. **Adjustment in the Classification Type of Variables**: The downloaded dataset presented most variables as integers or numeric. However, many variables are categorical/ordinal or dates. Adjusting the classification type of these variables was necessary to better manage, visualize, and analyze the data.
3. **Creation of New Variables**: New variables were created for certain variables, such as the suicide rate. For instance, the suicide rate was calculated as (number of deaths/population) *10,000.
4. **Manual correction of mistakes:** The SPLC dataset contained a significant number of mistakes that needed to be manually corrected, which amounted to thousands of additional lines of code in the R-Studio cleaning process. Some of these included manually renaming misspelled cities:

```
# Correct "Housotn" to "Houston"
hategroups_raw$City <- gsub("Housotn", "Houston", hategroups_raw$City)
```

Manually assigning counties to small towns:

```
# Mishiwaka, IN
 hategroups_raw$County[hategroups_raw$City_State_HG == "Mishiwaka, Indiana"]
<- "St. Joseph"
```

Removing non-existent cities:

```
# 1.7.3 "Miami, Massachusetts"
hategroups_raw <- hategroups_raw %>%
        filter(!(str_detect(City, "Miami") & State == "Massachusetts"))
```

Or fixing inconsistent nomenclature:

```
 # Fork McCoy
        hategroups_raw$City[hategroups_raw$City == "Fork McCoy" &
hategroups_raw$State == "Florida"] <- "Fort McCoy"
```

```
    hategroups_raw$City[hategroups_raw$City == "Fort McCroy" &
hategroups_raw$State == "Florida"] <- "Fort McCoy"
    hategroups_raw$City[hategroups_raw$City == "Fort Mccoy" &
hategroups_raw$State == "Florida"] <- "Fort McCoy"
```

5. **Conversion of Excel and CSV files to readable R-Studio data frames:** The conversion of Excel and CSV files in R-Studio can be easily achieved with the *read_excel* and *read.csv* function, but a great deal of manual converting had to take place because many of these tables were disorganized after being transferred into R-Studio.

3.2.3. Variables and Measurement

We established the variables as follows; the dependent variable is suicide, the independent variable is hate groups, and the control variables are population, unemployment rate, poverty rate, and median household income. For more detailed data, we have provided the summary statistics in the Appendix 1. Here are the key variables in our research:

Table 1. Key Variables

Variable Name	Description	Type (Unit)	Measurement
suicide_rate	Suicide Rate; Year: 2000-2022	Numeric (continuous)	Per 100,000 individuals
numberofhategroups	Hate Groups; Year: 2000-2022	Numeric (discrete)	Number of hate groups per county *
unemployment	Unemployment Rate; Year: 2000-2022	Numeric (continuous)	Percentage
poverty	Poverty Rate; Year: 2003-2016	Numeric (continuous)	Percentage
med_household_income	Median Household Income; Year: 2003-2016	Numeric (continuous)	US Dollars
population	Population; Year: 2000-2022	Numeric (discrete)	Person

***Note(s):** Hate groups by type and ideology were also included in the data set but were not specifically used in this analysis. Hate crimes were originally intended to be included as a variable, but this was impossible due to the amount of missing data.

<u>Data Analysis</u>

The data analysis was performed using R version 4.3.1. The specified packages utilized for general data operations encompassed tidyverse (including ggplot2, dplyr, tidyr, etc.). Specifically for map visualization, the tools employed were tmap, purrr, sf, usmap, viridis, and tigris. For data analysis purposes, the packages employed comprised readxl, readr, data.table, corrplot, sandwich, broom, lmtest, and plm.

3.3 Data Visualization Result

We employed three types of visualizations for our findings: Timeline bar graphs showcasing the increase in suicide rates along with the increase of other variables such as hate groups (Figure 1)

A county-level map of the United States showcasing suicide rates and hate groups (Figure 5, Figure 6, Figure 7)

Figure 1:

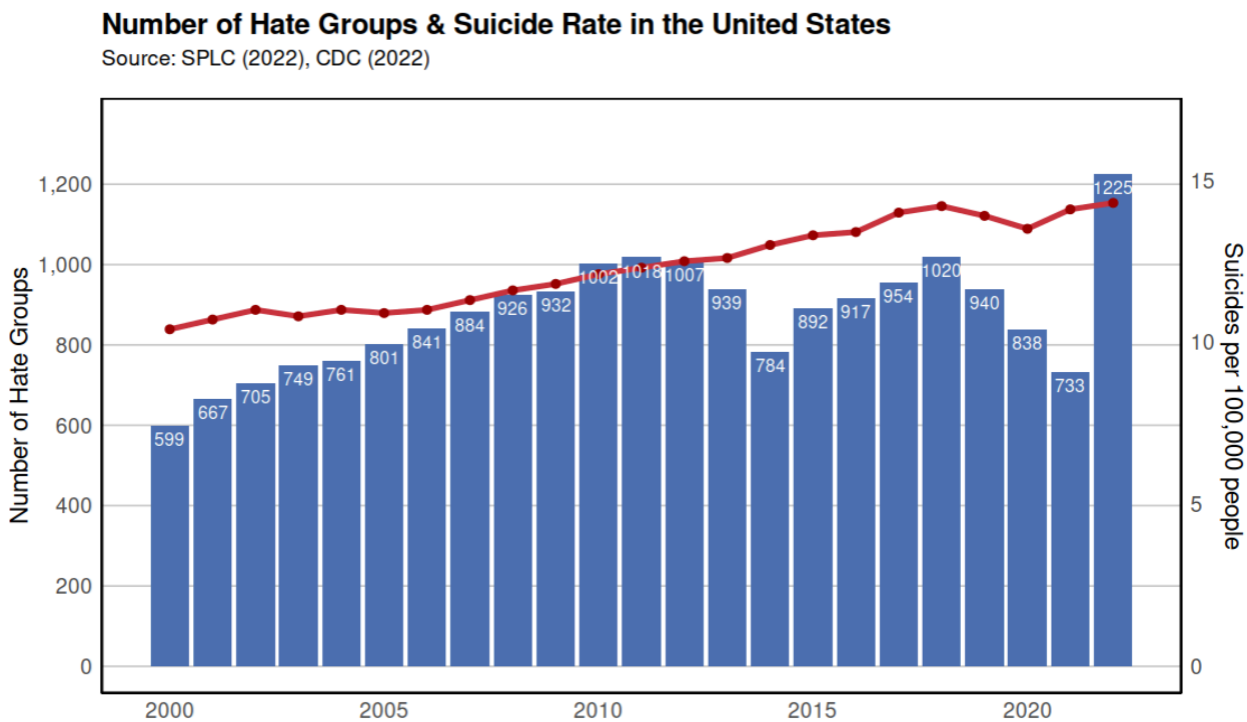
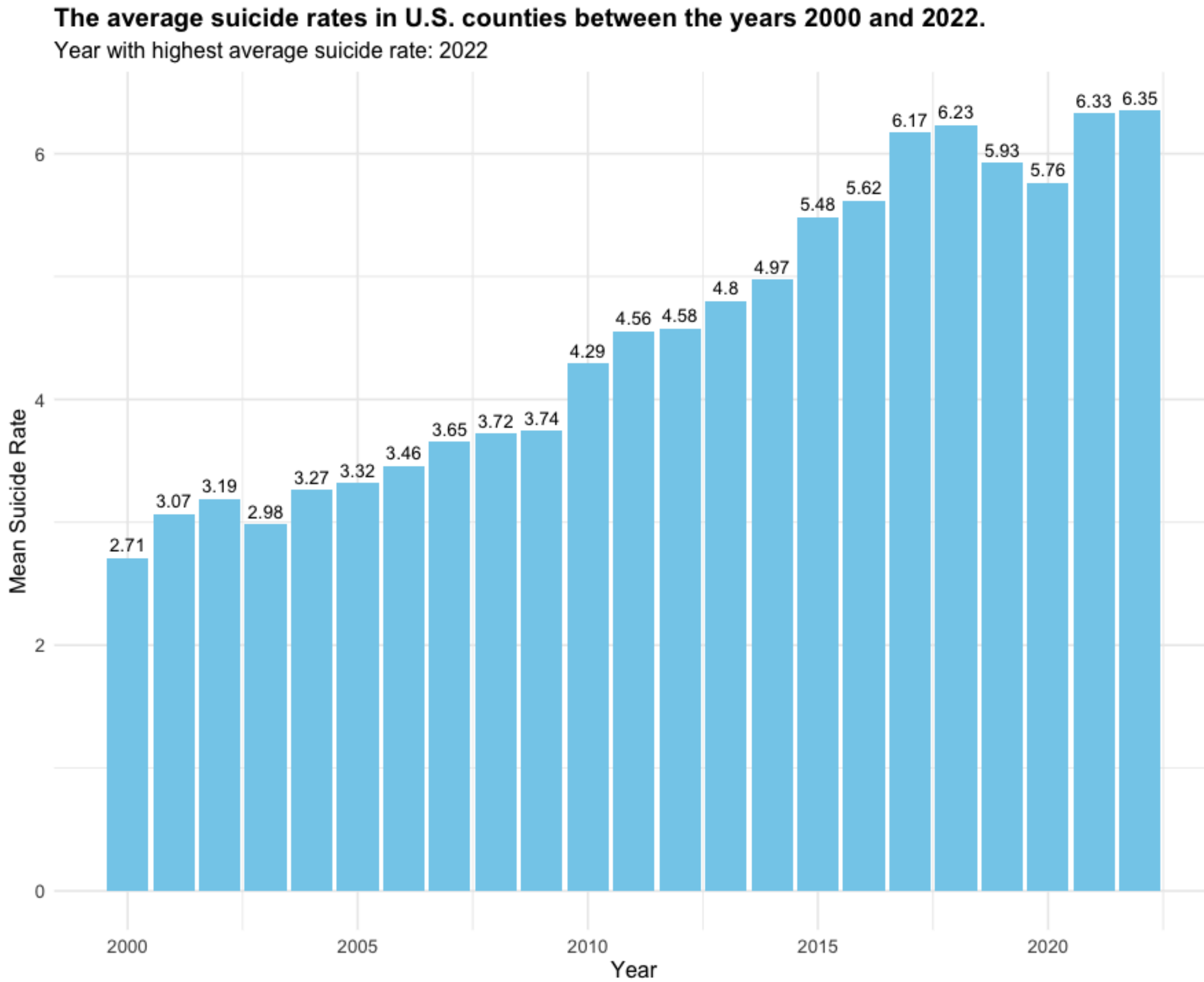


Figure 1 provides an initial demonstration of the potential correlation between suicide rates and hate group formation. Notably, a decrease of both variables in 2018 is apparent, along with their sudden increase in 2022.

Figure 2:

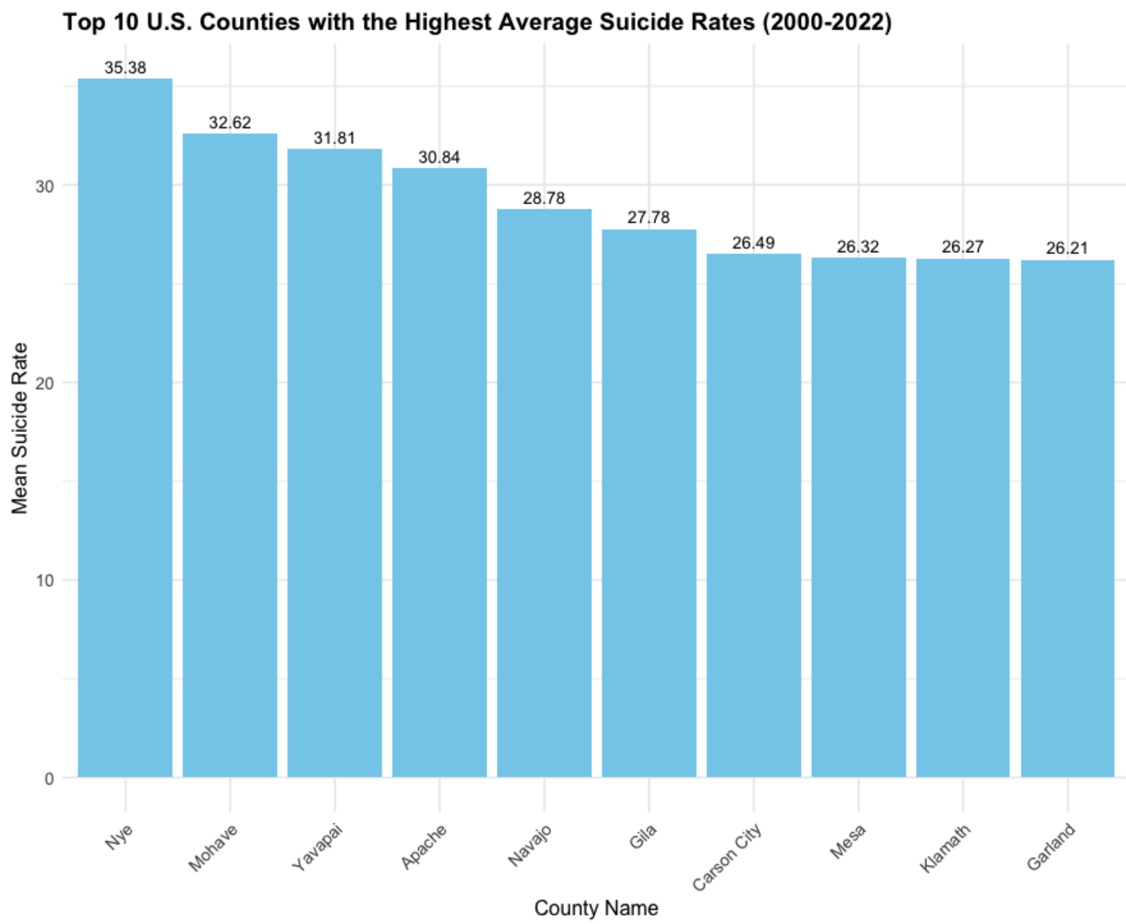


According to the Figure 2, The bar chart illustrates the average suicide rates in US counties from 2000 to 2022, highlighting an upward trend over these years. In 2000, the mean suicide rate was relatively low at 2.71. However, this rate increased steadily, reaching 3.32 in 2005 and 4.29 in 2010. The upward trend became more clear from 2010 onwards, with the rate climbing to 5.48 in 2015. This rise continued through the next several years, peaking at 6.35 in 2022. Additionally, it is known that COVID-19 had a strong negative effect on mental health, particularly in the United States. It is possible that these events also contributed to the surge in suicides between 2020 and 2022.

Economic recessions, particularly the 2008 financial crisis, had not only profound effects on individuals' income, but also their mental health. High unemployment rates and financial instability during such periods have been correlated with increased suicide rates, as financial stress can lead to feelings of hopelessness and despair (Reeves et al., 2012). Additionally, the growing recognition and diagnosis of mental health disorders, such as depression and anxiety, have highlighted the insufficiency in access to mental health services. Despite increased awareness, many individuals still face untreated mental health conditions, heightening their risk of suicide (Nock et al., 2008).

The COVID-19 pandemic has further impacted mental health globally. Stress related to the pandemic, including social distancing, lockdowns, economic uncertainty, and health fears, have contributed to increased anxiety, depression, and suicidal thoughts. The data suggest that the pandemic may have exacerbated the upward trend in suicide rates in the past recent years (Gunnell et al., 2020).

Figure 3:

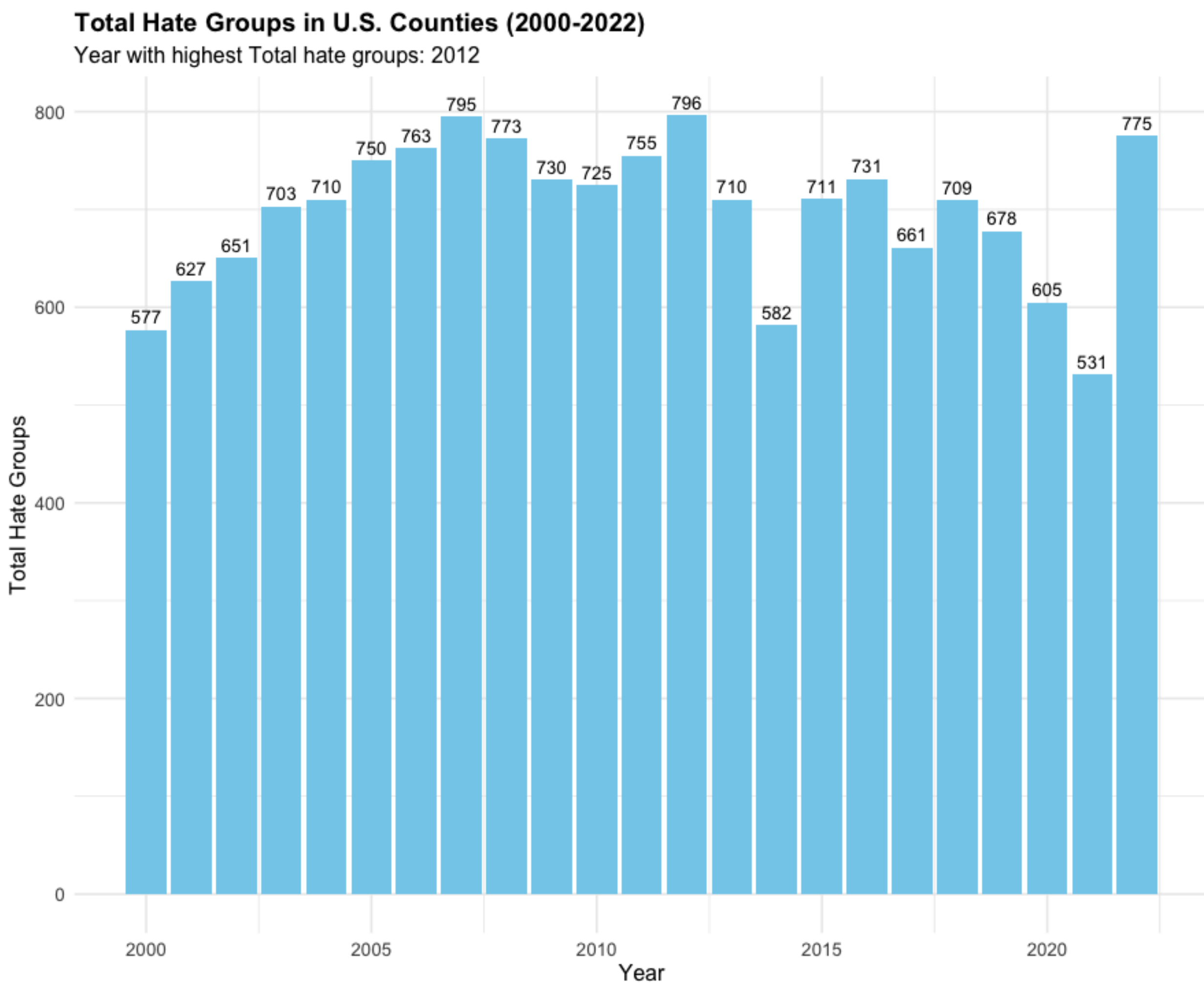


The bar chart, figure 3, provides a stark visualization of the mean suicide rates per 100,000 population across the ten counties in the United States with the highest averages over a 22-year span. Leading this distressing list is Nye County, with an average suicide rate of 35.38, followed by Mohave County at 32.62, and Yavapai County at 31.81. Other counties with alarmingly high rates include Apache County at 30.84, Navajo County at 28.78, and Gila County at 27.78. Carson City, Mesa, Klamath, and Garland counties also feature prominently, with rates ranging from 26.21 to 26.49.

These figures reveal a significant public health crisis, highlighting the urgent need for targeted mental health interventions and resources in these areas. Several factors may contribute to the elevated suicide rates in these regions, including socioeconomic challenges and limited access to mental health care. Additionally, many of these counties encompass rural areas, which often exhibit higher suicide rates compared to urban areas due to increased isolation and fewer healthcare resources (Hirsch, 2006).

Addressing this issue necessitates a multifaceted approach that enhances mental health services, strengthens community support systems, and implements preventive measures tailored to the specific needs of these counties. Comprehensive public health strategies should focus on mitigating risk factors such as financial stress and social isolation, while fostering protective factors like community connectedness and improved access to mental health care (Stone et al., 2018).

Figure 4:



The data above, Figure 4, indicates a fluctuating trend with notable peaks and troughs. The year 2012 stands out as having the highest recorded number of hate groups, reaching a peak of 796. This period marks a significant rise in the presence of hate groups, likely reflecting socio-political events and shifting societal dynamics during that time.

In the early 2000s, the number of hate groups was relatively lower, with 577 groups in 2000. However, there was a steady increase in the following years, reaching 627 in 2001 and continuing to climb, peaking again at 703 in 2002. This upward trend suggests that the early 2000s saw growing polarization and the formation of hate groups, possibly influenced by the aftermath of the September 11 attacks and the subsequent changes in national security and public sentiment.

The mid-2000s to early 2010s show a consistent rise, with the number of hate groups fluctuating around the 700-800 range. For instance, there were 763 groups in 2006, 795 in 2007, and a peak of 796 in 2012. This period could be linked to the economic downturn during the Great Recession, which often exacerbates social tensions and fosters extremist ideologies. According to studies, economic hardship can correlate with increased hate group activity as individuals and communities search for scapegoats (Smith, 2015).

Post 2012, there is a noticeable decline in the number of hate groups, dropping to 582 in 2014. However, this decline was not sustained, as the numbers began to rise again, reaching 711 in 2016. This resurgence coincides with the 2016 U.S. presidential election, suggesting that political rhetoric and campaign dynamics may have played a role in the re-emergence of hate groups. Political analysts often note that election periods can intensify ideological divides, potentially leading to an increase in hate group formation (Jones, 2017).

The data from 2017 to 2022 depicts another wave of fluctuations, with a decrease to 531 in 2020, followed by a sharp increase to 775 in 2022. This recent spike might be attributed to the sociopolitical climate influenced by the COVID-19 pandemic, economic uncertainties, and heightened social movements addressing racial injustices and inequalities. The convergence of these factors likely contributed to the surge in hate group activities as societal stressors and polarized discourses amplified (Dunn, 2023).

Subsequently, several maps are generated to provide clearer visualizations. A specific year is selected based on the preceding diagram to identify the year with the highest suicide rate or incidence of hate groups.

Figure 5:

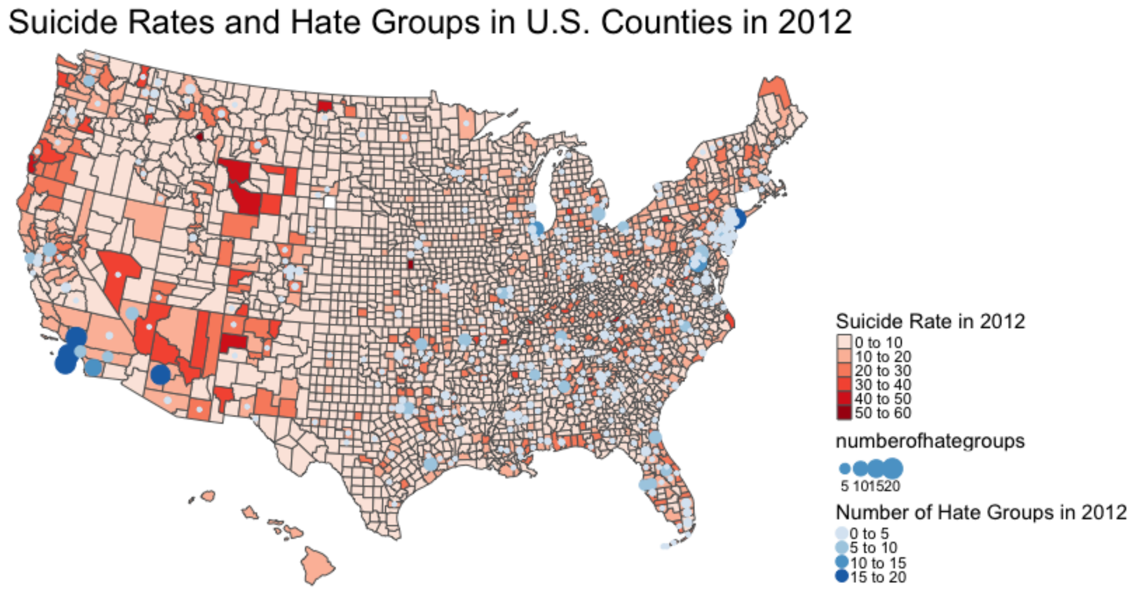


Figure 6:

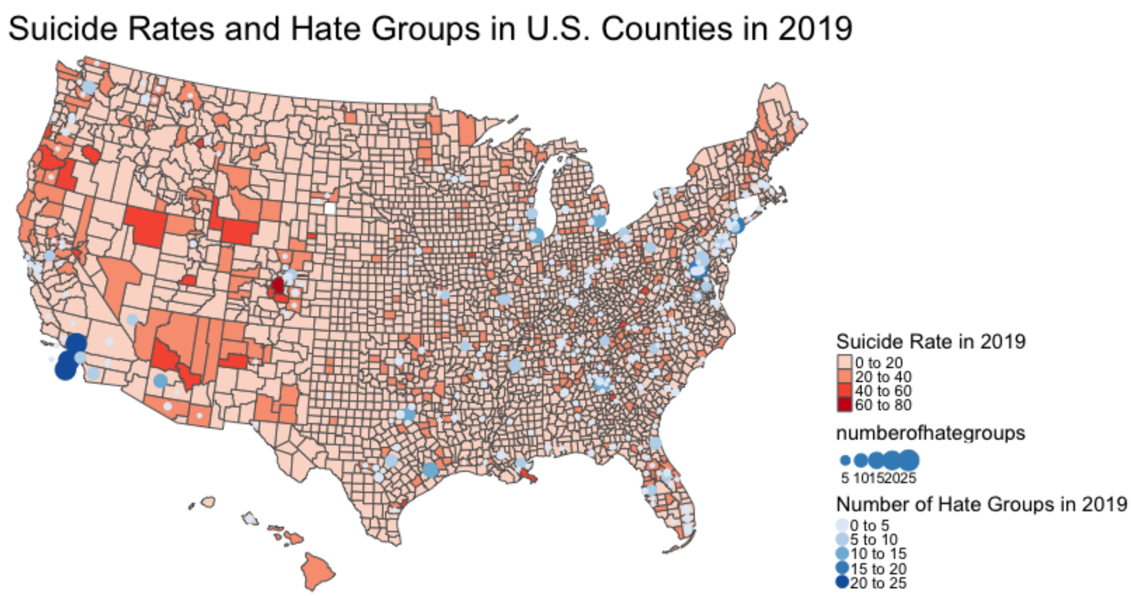
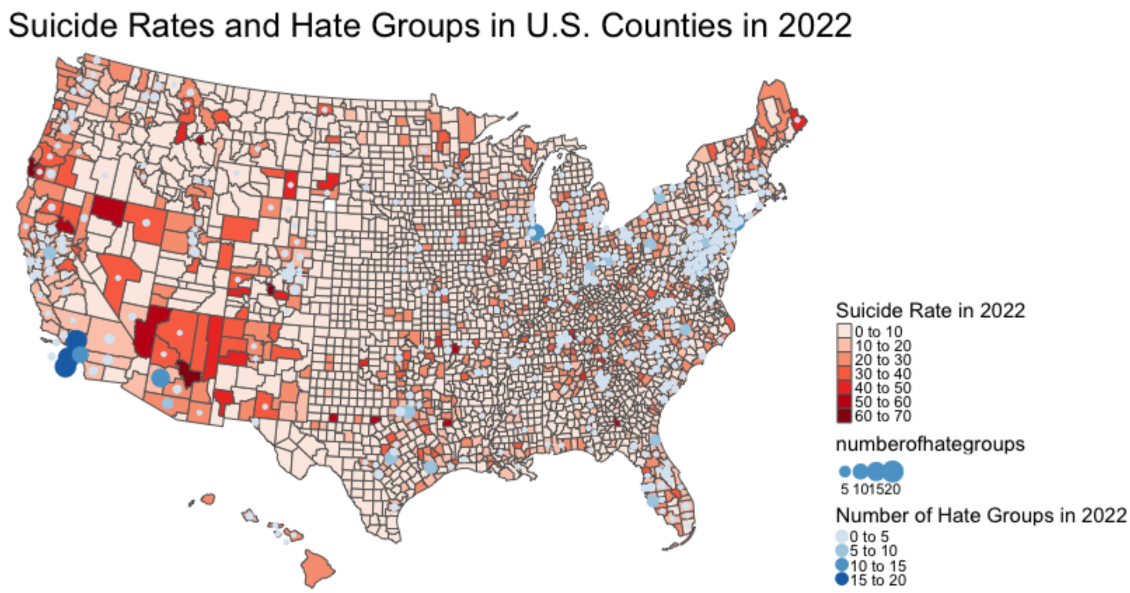


Figure 7:



Then after we get the preliminary result, we will employ several empirical methods for deeper analysis.

4. Empirical Method and Result

4.1 The Pearson Correlation Coefficient

We employ the Pearson Correlation Coefficient (PCC), a statistical measure that quantifies the linear correlation between two data sets to examine the relationship between two variables. According to Rodgers and Nicewander (1988), the PCC measures the strength and direction of the linear relationship between two variables on a continuous scale. The PCC is calculated using the following equation:

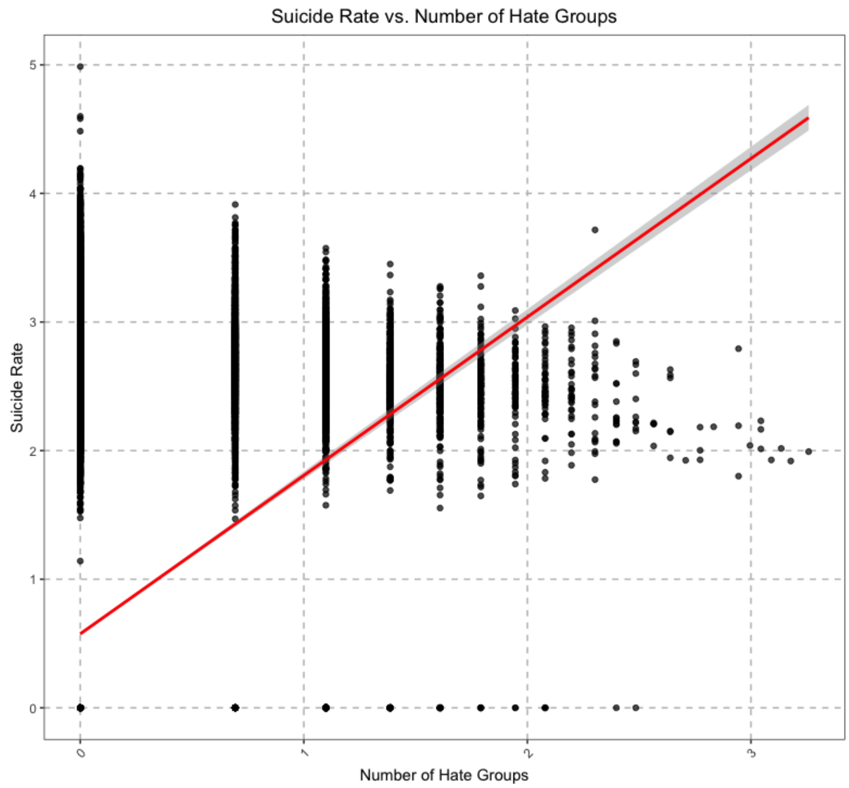
$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Where:

- X_i and Y_i are the individual data points of variables x and y.
- $\bar{x}$ and $\bar{y}$ are the means of variables X and Y, respectively.

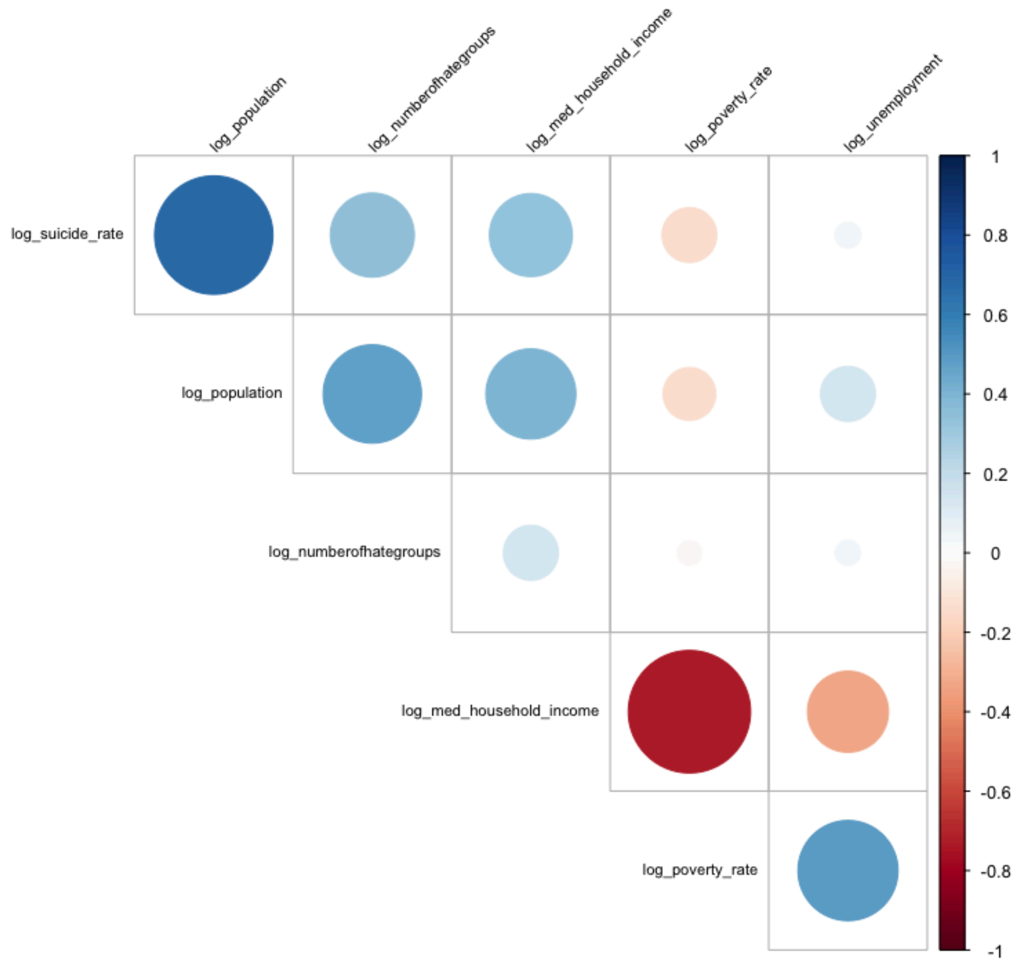
Firstly, we want to see the initial relationship between hate groups and the suicide rate in the U.S. Thus, we employ the Pearson correlation, analyzing 74,496 observations from 2000-2022.

Figure 8. Scatter plot of suicides per 100,000 individuals and hate groups in the United States



According to Figure 8, the scatter plot demonstrates moderate positive associations between suicide rates and the number of hate groups at the US county level. This finding gives preliminary insights into the factors influencing suicide rates across US counties from 2000 to 2022, with additional regression analysis required to thoroughly investigate these correlations. Before proceeding with our investigation, we will also look at the association between this study's variables. Figure 9 shows Pearson's correlation matrix. Correlation coefficients vary from -1 (perfect negative correlation) to 1 (perfect positive correlation), reflecting the strength and direction of the linear relationship between two variables, whereas 0 indicates no linear relationship. This results in a standardized measure of covariance. Moreover, the diagram below, Table 2, illustrates the relationship between the suicide rate and social capital, as proxied by population size, number of hate groups, median household income, poverty rate, and unemployment rate.

Figure 9. Pearson's correlation matrix of key variables



Based on Figure 9, several significant relationships between variables related to social capital loss and suicide rates in the United States were identified. The moderate positive correlation between the logarithm of the suicide rate and the logarithm of the population suggests that regions with larger populations tend to have higher suicide rates. Additionally, there is a positive correlation between the logarithm of the suicide rate and the logarithm of the number of hate groups, indicating that areas with more hate groups may have higher suicide rates. However, contrary to typical expectations, there is a positive correlation with suicide rates, suggesting that higher income areas might have higher suicide rates, potentially due to higher stress or other socioeconomic factors not immediately apparent. Then, the negative correlation suggests that higher poverty rates are associated with lower suicide rates, which is contrary to common assumptions and warrants further investigation. Moreover, the weak positive correlation indicates almost no relationship between unemployment and suicide rates, suggesting that other factors might be more significant in influencing suicide rates. These findings highlight complex and sometimes unexpected relationships between social capital variables and suicide rates. Therefore, we will conduct further research using advanced regression models to delve deeper into these correlations, control for potential confounding factors, and understand the causal pathways driving these relationships. By employing

Ordinary Least Squares (OLS) and Fixed Effects regression analysis, we aim to provide a more nuanced understanding of how population size, number of hate groups, median household income, unemployment, and poverty rates influence suicide rates over time. This approach will help to clarify the complex interplay between social capital loss factors and suicide rates, ultimately informing more effective policy interventions.

Table 2. Pearson's correlation matrix of key variables

Correlation Matrix	*suicide_rate (log)*	*numberofhategroups (log)*	*population (log)*	*med_household_income (log)*	*unemployment (log)*	*poverty_rate (log)*
suicide_rate (log)	*1.00*	*0.35*	*0.69*	*0.34*	*0.03*	*-0.15*
numberofhategroups (log)	*0.35*	*1.00*	*0.47*	*0.15*	*0.03*	*-0.03*
population (log)	*0.69*	*0.47*	*1.00*	*0.40*	*0.15*	*-0.14*
med_household_income (log)	*0.34*	*0.15*	*0.40*	*1.00*	*-0.32*	*-0.73*
Poverty_rate (log)	*0.03*	*0.03*	*0.15*	*-0.32*	*1.00*	*0.49*
Unemployment (log)	*-0.14*	*-0.02*	*-0.14*	*-0.76*	*0.49*	*1.00*

Furthermore, according to Senaviratna and Cooray (2019), substantial collinearity exists when the absolute number of the correlation exceeds 0.8. In this case, according to Table 2, none of the variables exceed the threshold for severe collinearity, suggesting no strong correlation among the variables in the dataset. Therefore, we can proceed with further regression analysis to comprehensively explore these relationships and assess causal effects.

4.2 Ordinary Least Squares (OLS)

Firstly, Ordinary Least Squares (OLS) regression is employed in this study, a robust method widely used in econometrics to analyze linear relationships between variables. According to Wooldridge (2016), OLS is the most commonly used method for estimating linear regression models, minimizing the sum of squared residuals to provide the best linear unbiased estimates under certain conditions. OLS is chosen for its efficiency in estimating parameters and providing unbiased estimates of relationships between social capital proxies and suicide rates. This approach facilitates examining how changes in hate group formation, as a proxy for social capital loss, correlate with variations in suicide rates across U.S. counties from 2000 to 2022. Moreover, Robust standard errors are employed in this research to address potential heteroscedasticity and enhance the reliability of statistical inference. Here is the Ordinary Least Squares (OLS) equation along with the following results:

$$ln(suicide) = \beta_0 + \beta_1 ln(hate\_groups) + \beta_2 ln(income) + \beta_3 ln(poverty) + \beta_4 ln(unemployment) + \beta_5 ln(population) + \varepsilon$$

The intercept term is represented by β0, where the coefficients to be estimated are represented by βn, and the error term is represented by ϵ.

Table 3. Ordinary Least Squares (OLS) Linear Regression

Suicide Rate	*OLS estimates*
	Coef.
Hate Groups (log)	*0.096**** *(0.015)*
Median Household Income (log)	*0.336**** *(0.027)*
Poverty Rate	*0.368**** *(0.099)*
Unemployment (log)	*-0.158**** *(0.012)*
Population (log)	*0.555**** *(0.004)*
R-squared	*0.4786*
F-statistics	*7971****
Number of obs	*43428*
Number of counties	*3103*

Note(s): **** P<0.01, ** P<0.05, * P<0.1*
Robust standard errors are in parentheses
Source(s): *Authors' estimates*

The findings reveal significant associations between various socioeconomic factors and suicide rates. Specifically, the logarithm of hate groups exhibits a positive and statistically significant relationship with the logarithm of the suicide rate (Coefficient = 0.096, $p < 0.01$), suggesting that an increase in hate groups correlates with higher suicide rates. Similarly, higher median household income shows a positive and significant correlation with the logarithm of the suicide rate (Coefficient = 0.336, $p < 0.01$), indicating that regions with greater income levels experience elevated suicide rates.

Moreover, the poverty rate demonstrates a positive and significant association with the logarithm of suicide rate (Coefficient = 0.368, $p < 0.01$), suggesting that higher poverty rates are associated with increased suicide rates. Contrary to expectation, the logarithm of unemployment exhibits a negative and significant relationship with the logarithm of the suicide rate (Coefficient = -0.158, $p < 0.01$), indicating that higher unemployment rates are linked to lower suicide rates.

Additionally, the logarithm of the population shows a robust positive relationship with the logarithm of suicide rate (Coefficient = 0.555, $p < 0.01$), suggesting that larger populations are correlated with higher suicide rates. The model accounts for approximately 47.86% of the variance in suicide rates, as evidenced by the R-squared value (R-squared = 0.4786). The F-statistic (F = 7971, $p < 0.01$) confirms the overall significance of the model.

4.3 Linear Panel Model (PLM)

We also utilize panel-fixed effects to address concerns about potential endogeneity, which could affect both the outcome and the treatment variables. This approach aids in the control of county-specific and time-specific elements that may have an impact on suicide rates and social capital loss variables. Using the panel fixed effects model, we may reduce potential regression analysis biases caused by missing county-specific historical or structural factors that influence these variables. Furthermore, fixed-effects models are more reliable than random-effects models when analyzing policy effects with aggregated data (Wooldridge, 2016). Moreover, Robust standard errors are also employed to address potential heteroscedasticity and enhance the reliability of statistical inference. The plm package makes it easier to estimate linear panel models by including functions for a wide range of model specifications and robust inference methods.

4.3.1 Linear Panel Model (PLM) without Interaction

Here is the Linear Panel Model (PLM) equation along with the following results:

$$ln(suicide)_{it} = \beta_0 + \beta_1 ln(hate\_groups)_{it} + \beta_2 ln(income)_{it} + \beta_3 ln(poverty)_{it} + \beta_4 ln(unemployment)_{it} + \beta_5\, ln(population)_{it} + \varepsilon_{it}$$

The intercept term is represented by β0, where the coefficients to be estimated are represented by βn, and the error term is represented by ϵ. The variables i and t represent counties in the U.S and specified years in this research, respectively.

Table 4. Panel Fixed Effect Regression Model

Suicide Rate (log)	Fixed-effects estimates
	Coef.
Hate Groups (log)	*-0.003* *(0.020)*
Median Household Income (log)	*0.519**** *(0.035)*
Poverty Rate	*1.952**** *(0.231)*
Unemployment (log)	*-0.005* *(0.014)*
Population (log)	*1.273**** *(0.102)*
R-squared	*0.022105*
F-statistics	*182.286****
Number of obs	*43428*
Number of counties	*3103*

Note(s): **** P<0.01, ** P<0.05, * P<0.1*
Robust standard errors are in parentheses
Source(s): *Authors' estimates*

The findings from the PLM analysis reveal that Hate Groups do not exhibit statistical significance in relation to suicide rates. Conversely, Median Household Income and Poverty Rate demonstrate significant associations. Next, this study explores potential interaction effects and utilizes lagged variables through PLM analysis to further elucidate these findings.

Despite the idea that social cohesion might influence mental health outcomes, the result suggests that the presence of hate groups does not have an impact on suicide rates. This result contrasts with some theoretical frameworks that propose capital loss as a potential driver of negative mental health outcomes, including suicide. It is possible that while hate groups signify social fragmentation, their presence does not directly translate into increased suicide rates.

In contrast, the result reveals a significant positive relationship between median household income and suicide rate. This finding suggests that higher income levels, which might be expected to correlate with lower suicide rates, actually associate with higher suicide rates in this dataset. As income disparity widens, individuals may experience higher pressure, contributing to mental health issues and, subsequently, higher suicide rates. This phenomenon aligns with the concept of relative deprivation, where individuals' perceptions of their socio-economic status relative to others can significantly impact their mental health. Lorant et al. (2005) discuss how economic disparities can exacerbate stress and lead to adverse mental health outcomes, supporting the findings of this study.

Similarly, the positive correlation between poverty rates and suicide rates aligns with existing literature. Poverty is a stressor that can lead to various mental health challenges, including depression and anxiety, which are risk factors for suicide. The significant coefficient for the poverty rate underscores the critical impact of economic hardship on mental health. Rehkopf and Buka (2006) emphasize that areas with higher poverty levels often exhibit higher suicide rates, highlighting the detrimental effects of financial stress and limited access to mental health resources. Interestingly, the unemployment rate does not show a significant impact on the suicide rate in this model. This finding might suggest that while unemployment is a critical economic indicator, its direct influence on suicide rates may be more complex and mediated by other factors.

Finally, the model indicates a significant positive relationship between population size and the suicide rate. Larger populations may be associated with greater urbanization and reduced social cohesion, which can adversely affect mental health and increase suicide rates. Urbanization often brings about rapid lifestyle changes and higher stress levels, contributing to mental health challenges. This finding aligns with the notion that as populations grow, the pressures and demands on individuals also increase, potentially leading to higher rates of mental health issues and suicide (Iskander & Crosby, 2021; Casant & Helbich, 2022).

4.3.2 Linear Panel Model (PLM) with Lagged variables and an Interaction Variable

Here is the Linear Panel Model (PLM) equation along with the following results:

$$ln(suicide)_{it-1} = \beta_0 + \beta_1 ln(hate\_groups)_{it-1} + \beta_2 ln(poverty)_{it-1} + \beta_3 ln((hate\ groups)_{it-1} \times ln(income)_{it-1}) + \beta_4 ln(unemployment)_{it-1} + \varepsilon_{it}$$

The equation provided is a panel regression model with lagged variables, linking natural logarithm suicide rates (ln (suicide)) at a given time t - 1 to several predictor variables. i and t denote counties and specific years, respectively.

Table 5. Panel Fixed Effect Regression with Lagged Model and an Interaction Variable

Lagged Suicide Rate (log)	Fixed-effects estimates
	Coef.
Lagged Hate Groups (log)	*-2.500**** *(0.541)*
Lagged Poverty Rate (log)	*2.931**** *(0.230)*
Lagged Hate Groups (log) × Lagged Median Household Income (log)	*0.232**** *(0.050)*
Lagged Unemployment (log)	*-0.001* *(0.014)*
R-squared	*0.0061319*
F-statistics	*182.286****
Number of obs	*43428*
Number of counties	*3103*

Note(s): **** P<0.01, ** P<0.05, * P<0.1*
Robust standard errors are in parentheses
Source(s): *Authors' estimates*

The model above, stated using the within-transformation in R's plm package, investigates the lagged impacts of various social capital proxies on suicide rates while controlling for county-specific fixed effects and potential endogeneity issues.

The estimated model in Table 2 offers numerous noteworthy discoveries. First, there is a substantial negative relationship between lagged hate group numbers and lagged suicide rates. This implies that counties with a larger historical count of hate groups have lower suicide rates. An increase in the lagged poverty rate is positively associated with suicide rates. This suggests that economic suffering contributes to increased suicide risks. Interestingly, the interaction impact between lagged hate groups and lagged median family income reveals a complex relationship. Counties with parallel rises in hate group prevalence and median household income have higher suicide rates. From a theoretical standpoint, the interaction effect aligns with the concept that social capital and socioeconomic factors are intertwined. Higher median household income typically correlates with better access to resources and opportunities, but it can also lead to increased visibility and impact of hate group activities, thus influencing mental health and suicide rates (Putnam, 2000). This finding underscores the importance of a holistic approach in public health strategies that address both economic and social capital dimensions to mitigate suicide risks effectively.

These findings are consistent with earlier research on social capital and mental health, which has demonstrated the complex and sometimes contradictory effects of social networks and community structures on individual well-being. For instance, Kawachi and Berkman (2001) discuss how social cohesion can have both protective and detrimental effects on health outcomes depending on the nature of social ties and the inclusivity of social networks. This is showed in the negative relationship observed between hate groups and suicide rates, suggesting that even negative social capital can have stabilizing effects under certain conditions.

Additionally, the positive association between poverty and suicide rates aligns with Durkheim's (1897) seminal work on the social causes of suicide, which highlighted economic instability as a significant stressor leading to higher suicide rates. The interaction between hate groups and median household income further complicates this relationship, indicating that economic resources alone are insufficient to buffer against the negative impacts of social fragmentation.

Moreover, the robustness checks using robust standard errors, calculated with the HC4 estimator, confirm the significance of our variables. The robust summary shows consistent results, reinforcing the validity of the findings. The significant coefficients for lagged hate group numbers, lagged poverty rate, and the interaction term between lagged hate group numbers and lagged median household income remain significant even after accounting for potential heteroskedasticity.

5. Limitations

Limitations in the analysis include the exclusion of Alaska, as the state's delineation of counties is inconsistent, as well as the exclusion of US colonies and territories. We are only aware of one database on hate groups – Southern Poverty Law Center; thus we were unable to compare across sources. That being said, the dataset on hate groups itself had limitations. We eliminated "statewide" categorized hate groups and all missing values. Mostly we eliminated data on groups with mislabeled cities, unknown locations, or locations that spanned arbitrary regions.

Furthermore, two variables were omitted from this study. The hate crime variable had an issue of an undefined clarification of what constitutes a crime. According to the National Institute of Justice, hate crimes are defined distinctly across jurisdictions. Additionally, law enforcement's inability to categorize some victimizations as hate crimes, and reluctance of victims to engage with law enforcement leads to underreporting and misreporting; with states, such as Alabama, that do not report hate crimes. Therefore providing an inconsistent data set of disproportionate values.

Moreover, the variable for religious affiliation was not used in this study given the limited data. The data found is based on the "US religious congregation census" which only takes place every ten years, limiting our data collection to only two years, in return limiting our measure of religious affiliation. Adding this variable to our research would have then skewed our results negatively. However, the prospect of adding religious affiliation, with sufficient data spanning throughout 2000-2022, would be a valuable asset to our research as a potential proxy of social capital loss and in strengthening the robustness of the models.

Nevertheless, additional potential variables that would strongly contribute to this study include: percentage of county spending devoted to police and public health, non-white population, percentage of population living in rural, urban, and suburban areas, level of education, religious affiliation, and voter turnout. Controlling for these variables can strengthen the analysis by accounting for additional factors that may influence suicide rates. Including these variables would allow for a more holistic understanding of the varied determinants of suicide, thereby improving the accuracy of the model's predictions. Additionally, considering these factors may reveal underlying social, economic, and demographic dynamics that contribute to suicide rates, facilitating the development of more effective public and mental health policies, as well as preventive measures tailored to the unique needs of different communities.

6. Conclusions and Future Research

This paper intends to explore the relationship between five social capital loss proxies and suicide across the United States at a county level. Putnam defines social capital as "features of social organization such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefit" (Putnam, 2000). Social capital is founded on individual actors and their relationships, as well as the social structures in which they are embedded. These associated norms are generally beneficial to those inside the network, but the external effects of social capital are not always positive. Social capital can be used for malevolent purposes such as the formation of hate groups, intolerant to change and differences, becoming less connected to society. Thus, hate groups become a proxy for social capital loss. Furthermore, level of unemployment, poverty, household income, and population size can also serve as proxies for social capital loss, as they can infer potential impacts on social trust, coordination and cooperation.

The analysis across multiple regression models provides varying insights into the factors influencing suicide rates. The Ordinary Least Squares model (OLS) model indicates a positive correlation in the increase of hate groups with higher suicide rates, suggesting counties with a higher prevalence of hate groups experience higher suicide rates. Similarly, counties with greater income levels, higher poverty rates, and larger populations also indicate a positive correlation with elevated suicide rates. In contrast however, the OLS model indicates a negative correlation between unemployment and suicide rates, suggesting higher unemployment rates are linked to lower suicide rates. This model explains about 48 percent (R-squared = 0.4786) of the variation in suicide rates and a highly significant F-statistic ($F = 7971$, $p < 0.01$).

The Panel Fixed Effect Regression Model (PLM) model without interaction terms indicates hate groups do not exhibit statistical significance in relation to suicide rates. On the other hand, both household income and poverty rate show significant correlation with suicide rates, which remains consisting with the previous findings in the OLS model. Moreover, when incorporating lagged variables and interaction terms in the Panel Fixed Effect Regression Model (PLM) model, the results indicate that an increase in the lagged poverty rate is positively associated with suicide rates, suggesting economic suffering contributes to elevated or increased suicide risks. Both household income and population also show a significant correlation with suicide rates.

However, contrary to general belief, this model implies counties with a larger historical count of hate groups and higher unemployment rate have lower suicide rates. Despite these significant findings, the model explains a relatively small portion of the variance in suicide rates, suggesting that other unexamined factors may also play a crucial role. Overall, the findings suggest economic conditions, population size, and presence of hate groups play a critical role in influencing suicide rates. The OLS model provides a more general view with higher explanatory power, whereas the PLM models elucidates nuanced and temporal dimensions of these relationships.

Future research could attempt to explore the relationship between social capital loss and suicide in Europe, given its social and economic similarities to the United States. However, this remains challenging due to the inconsistent and scarce data on proxies such as hate group formation, as most available data focuses on hate speech. The European Statistical Office (Eurostat) provides ample resources on statistics related to suicide, unemployment, crime, immigration and ethnic groups, etc., but none pertaining specifically to hate group formation or hate crimes.

A similar study can also be applied to Canada, given the growing number of hate groups (e.g. Blood and Honour, and Canadian Nationalist Front) and reported hate crimes, alongside lack of proper law enforcement to protect civilians from hate crimes. A lengthy history of racism and white supremacy in Canada is often overlooked, with very little existing contemporary academic literature on right wing organizations. Therefore, this topic would contribute to the much needed research on hate groups and hate crimes in Canada.

Lastly, the upcoming 2024 presidential election in the United States will once again feature Donald Trump and Joe Biden as the primary candidates. Building on our existing research, we propose conducting a study to investigate the correlation between the presence of hate groups and the percentage of voters that support Donald Trump. This analysis can illustrate whether there is a significant relationship between the presence of hate groups and political leanings across counties. Furthermore, it can help us understand to what extent this relationship potentially influences campaign strategies and policy decisions.

Systemic racism permeates all institutions in the United States, contributing to the ongoing issue of social capital loss. Suicide rates in the country are among the highest in the world, accompanied by increasing levels of inflation and limited policies enacted to counteract these trends, making it challenging for citizens to remain afloat. Furthermore, hateful rhetoric from the country’s own politicians perpetuates the continuation of hate groups and discrimination. It is imperative that more research is conducted on the potential causes of suicide and social capital loss in the United States.

Appendix

Appendix 1. Descriptive statistics of key variables

```
> summary(data_clean)
 County_State            Year          County          County.Name           State              STUSPS              ALAND          
 Length:71392       Min.   :2000   Length:71392       Length:71392       Length:71392       Length:71392       Min.   :5.300e+06  
 Class :character   1st Qu.:2005   Class :character   Class :character   Class :character   Class :character   1st Qu.:1.115e+09  
 Mode  :character   Median :2011   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :1.583e+09  
                    Mean   :2011                                                                               Mean   :2.466e+09  
                    3rd Qu.:2017                                                                               3rd Qu.:2.369e+09  
                    Max.   :2022                                                                               Max.   :5.198e+10  

     AWATER          County.Code       STATEFP         COUNTYFP        COUNTYNS         AFFGEOID             Deaths         Population      
 Min.   :0.000e+00   Min.   : 1001   Min.   : 1.00   Min.   :  1.0   Min.   :  23901   Length:71392       Min.   :  0.00   Min.   :      51  
 1st Qu.:7.076e+06   1st Qu.:19048   1st Qu.:19.00   1st Qu.: 35.0   1st Qu.: 481812   Class :character   1st Qu.:  0.00   1st Qu.:   11203  
 Median :1.904e+07   Median :29214   Median :29.00   Median : 79.0   Median : 974108   Mode  :character   Median :  0.00   Median :   25698  
 Mean   :1.453e+08   Mean   :30697   Mean   :30.59   Mean   :103.4   Mean   : 937998                      Mean   : 10.33   Mean   :   98534  
 3rd Qu.:5.645e+07   3rd Qu.:46008   3rd Qu.:46.00   3rd Qu.:133.0   3rd Qu.:1383972                      3rd Qu.: 10.00   3rd Qu.:   66153  
 Max.   :1.405e+10   Max.   :56045   Max.   :56.00   Max.   :840.0   Max.   :2054176                      Max.   :947.00   Max.   :10170292  

  Suicide_Rate     numberofhategroups      UID          unemployment_rate   Unemployed        Employed          Labor Force        poverty_rate   
 Min.   :  0.000   Min.   : 0.0000   Min.   : 10012000   Min.   : 0.600   Min.   :     3   Min.   :     34   Min.   :     38   Min.   :0.000  
 1st Qu.:  0.000   1st Qu.: 0.0000   1st Qu.:190487006   1st Qu.: 3.900   1st Qu.:   276   1st Qu.:   4785   1st Qu.:   5101   1st Qu.:0.114  
 Median :  0.000   Median : 0.0000   Median :292142011   Median : 5.300   Median :   691   Median :  11126   Median :  11852   Median :0.149  
 Mean   :  4.530   Mean   : 0.2248   Mean   :306976624   Mean   : 5.847   Mean   :  2868   Mean   :  46115   Mean   :  48983   Mean   :0.159  
 3rd Qu.:  8.455   3rd Qu.: 0.0000   3rd Qu.:460077016   3rd Qu.: 7.200   3rd Qu.:  1844   3rd Qu.:  29562   3rd Qu.:  31395   3rd Qu.:0.193  
 Max.   :145.364   Max.   :27.0000   Max.   :560452022   Max.   :29.400   Max.   :621950   Max.   :4917685   Max.   :5148584   Max.   :0.620  
                                                         NA's   :37       NA's   :37       NA's   :37        NA's   :37        NA's   :27939  
 median_household_income log_population   log_med_household_income log_unemployment  log_suicide_rate log_numberofhategroups log_poverty_rate
 Min.   :     0          Min.   : 3.932   Min.   : -Inf            Min.   :-0.5108   Min.   :0.0000   Min.   :0.0000         Min.   :0.000   
 1st Qu.: 35424          1st Qu.: 9.324   1st Qu.:10.47            1st Qu.: 1.3610   1st Qu.:0.0000   1st Qu.:0.0000         1st Qu.:0.108   
 Median : 41337          Median :10.154   Median :10.63            Median : 1.6677   Median :0.0000   Median :0.0000         Median :0.139   
 Mean   : 43250          Mean   :10.266   Mean   : -Inf            Mean   : 1.6733   Mean   :0.7587   Mean   :0.1138         Mean   :0.146   
 3rd Qu.: 48790          3rd Qu.:11.100   3rd Qu.:10.79            3rd Qu.: 1.9741   3rd Qu.:2.2465   3rd Qu.:0.0000         3rd Qu.:0.176   
 Max.   :134609          Max.   :16.135   Max.   :11.81            Max.   : 3.3810   Max.   :4.9861   Max.   :3.3322         Max.   :0.482   
 NA's   :27939                            NA's   :27939            NA's   :37                                                NA's   :27939   
```

Appendix 2. Pearson Correlation

```
> # Empirical Method -----------------------------------------------------
> library(tidyverse)
> library(readxl)
> 
> NEW_merged_data <- read_excel("/Users/ignatiusharry/Library/CloudStorage/Dropbox/Data/S2/NCCU/2nd Semester/Big data and social analysis/Final project/
Crime risk analysis/data/Control Variables/NEW_merged_data.xlsx")

> #  1. Pearson Correlation (2003-2016) Balanced Data -------------------------------------------------------------------------
> library(corrplot)
> 
> # Subset the relevant variables from data_clean
> vars <- c("log_suicide_rate", "log_numberofhategroups", "log_population", "log_med_household_income",
+           "log_unemployment",
+           "log_poverty_rate")
> data_subset <- data[vars]
> 
> ggplot(data, aes(x = log_numberofhategroups)) +
+   geom_histogram()
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
> 
> # Calculate correlations
> correlation_matrix <- cor(data_subset)
> # Round the correlation matrix to 2 decimal places
> rounded_correlation_matrix <- round(correlation_matrix, 2)
> 
> # Convert the correlation matrix to data frame for better display
> correlation_df <- as.data.frame(rounded_correlation_matrix)
> 
> # Print the correlation matrix as a table
> print(correlation_df)
                         log_suicide_rate log_numberofhategroups log_population log_med_household_income log_unemployment log_poverty_rate
log_suicide_rate                     1.00                   0.35           0.69                     0.34             0.03            -0.15
log_numberofhategroups               0.35                   1.00           0.47                     0.15             0.03            -0.03
log_population                       0.69                   0.47           1.00                     0.40             0.15            -0.14
log_med_household_income             0.34                   0.15           0.40                     1.00            -0.32            -0.73
log_unemployment                     0.03                   0.03           0.15                    -0.32             1.00             0.49
log_poverty_rate                    -0.15                  -0.03          -0.14                    -0.73             0.49             1.00
```

Appendix 3. Scatter plot of hate groups vs suicide rates

```
> # Sort correlations by absolute value from strongest to weakest for log_suicide_rate row
> sorted_indices <- order(abs(correlation_matrix["log_suicide_rate", ]), decreasing = TRUE)
> sorted_matrix <- correlation_matrix[sorted_indices, sorted_indices]
>
> # Plot sorted correlation matrix using corrplot
> corrplot(sorted_matrix, method = "circle", type = "upper",
+          tl.col = "black", tl.srt = 45, tl.cex = 0.7, diag = FALSE,
+          main = "Pearson's correlation matrix of key variables")
>
> library(ggplot2)
>
> ggplot(data, aes(x = log_numberofhategroups, y = log_suicide_rate)) +
+   # Theme adjustments
+   theme_bw() +  # Use black and white theme for a clean look
+   theme(
+     panel.grid.major = element_line(color = "gray", linetype = "dashed"),  # Gridlines
+     panel.grid.minor = element_blank(),  # Remove minor gridlines
+     plot.title = element_text(hjust = 0.5),  # Center title
+     axis.text.x = element_text(angle = 45, hjust = 1),  # Rotate x-axis labels
+     axis.title.x = element_text(vjust = -0.2),  # Lower x-axis title
+     axis.title.y = element_text(angle = 90),  # Rotate y-axis title
+     legend.position = "bottom"  # Move legend to bottom
+   ) +
+   # Geometries with adjustments
+   geom_point(aes(), alpha = 0.7) +  # Color and size by factor level (optional)
+   geom_smooth(method = "lm", formula = y ~ x, color = "red") +
+   # Enhance labels and title
+   labs(title = "Suicide Rate vs. Number of Hate Groups",
+        x = "Number of Hate Groups",
+        y = "Suicide Rate")
>
> correlation <- cor(data$log_numberofhategroups, data$log_suicide_rate)
> correlation
[1] 0.3463339
```

Appendix 4. OLS Regression Model

```
> # 2. OLS Method (2000-2022) unbalanced data-------------------------------------------------------------
> library(sandwich)
> library(broom)
> library(lmtest)
>
> #change into log
> data_clean <- NEW_merged_data %>%
+   mutate(log_population = log(Population),
+          log_med_household_income = log(median_household_income),
+          log_unemployment = log(unemployment_rate),
+          log_suicide_rate = log(Suicide_Rate + 1),  # Adding 1 to avoid log(0)
+          log_numberofhategroups = log(numberofhategroups + 1), # Adding 1 to avoid log(0)
+          log_poverty_rate = log(poverty_rate + 1))
> # Run the OLS model using lm
> model1 <- lm(log_suicide_rate ~ log_numberofhategroups +
+                #lag_log_numberofhategroups * lag_log_med_household_income +
+                log_med_household_income +
+                log_poverty_rate +
+                log_unemployment +
+                log_population,
+              data = data_clean)
>
> # Summary of the model
> summary(model1)

Call:
lm(formula = log_suicide_rate ~ log_numberofhategroups + log_med_household_income +
    log_poverty_rate + log_unemployment + log_population, data = data_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2856 -0.6312 -0.1971  0.4706  4.9547

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -8.494573   0.298889 -28.420  < 2e-16 ***
log_numberofhategroups    0.095989   0.014063   6.825 8.88e-12 ***
log_med_household_income  0.349761   0.028028  12.479  < 2e-16 ***
log_poverty_rate          0.523661   0.129873   4.032 5.54e-05 ***
log_unemployment         -0.160463   0.012312 -13.033  < 2e-16 ***
log_population            0.554710   0.003808 145.670  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8853 on 43422 degrees of freedom
  (27964 observations deleted due to missingness)
Multiple R-squared:  0.4786,    Adjusted R-squared:  0.4786
F-statistic:  7973 on 5 and 43422 DF,  p-value: < 2.2e-16

>
> # Calculate robust standard errors
> robust_se <- vcovHC(model1, type = "HC4")
>
> # Use coeftest to get robust standard errors
> robust_summary <- coeftest(model1, vcov = robust_se)
>
> # Print the robust summary
> print(robust_summary)

t test of coefficients:

                           Estimate Std. Error  t value  Pr(>|t|)
(Intercept)              -8.4945731  0.2862069 -29.6798 < 2.2e-16 ***
log_numberofhategroups    0.0959891  0.0154290   6.2214 4.974e-10 ***
log_med_household_income  0.3497608  0.0271168  12.8983 < 2.2e-16 ***
log_poverty_rate          0.5236614  0.1205678   4.3433 1.407e-05 ***
log_unemployment         -0.1604627  0.0121605 -13.1954 < 2.2e-16 ***
log_population            0.5547104  0.0039634 139.9582 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 5. Panel Fixed Effect Regression Model

```
plm(formula = log_suicide_rate ~ log_numberofhategroups + log_med_household_income +
    log_poverty_rate + log_unemployment + log_population, data = pnel,
    model = "within")

Unbalanced Panel: n = 3103, T = 12-14, N = 43428

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-3.053436 -0.091803 -0.015773  0.070282  4.343621

Coefficients:
                          Estimate Std. Error t-value Pr(>|t|)
log_numberofhategroups   -0.0038700  0.0174157 -0.2222   0.8242
log_med_household_income  0.5192013  0.0340849 15.2326   <2e-16 ***
log_poverty_rate          1.9523638  0.2161808  9.0312   <2e-16 ***
log_unemployment         -0.0054009  0.0131795 -0.4098   0.6820
log_population            1.2727694  0.0773256 16.4599   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    17243
Residual Sum of Squares: 16861
R-Squared:      0.022105
Adj. R-Squared: -0.05325
F-statistic: 182.286 on 5 and 40320 DF, p-value: < 2.22e-16
>
> # Calculate robust standard errors
> robust_se <- vcovHC(model_no_interaction, type = "HC4")
>
> # Use coeftest to get robust standard errors
> robust_summaryte <- coeftest(model_no_interaction, vcov = robust_se)
>
> # Print the robust summary
> print(robust_summaryte)

t test of coefficients:

                          Estimate Std. Error t value Pr(>|t|)
log_numberofhategroups   -0.0038700  0.0196726 -0.1967   0.8440
log_med_household_income  0.5192013  0.0347542 14.9393   <2e-16 ***
log_poverty_rate          1.9523638  0.2309003  8.4554   <2e-16 ***
log_unemployment         -0.0054009  0.0137223 -0.3936   0.6939
log_population            1.2727694  0.1021281 12.4625   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 6. Panel Fixed Effect Regression with Lagged Model and an Interaction Variable

```
Call:
plm(formula = lag_log_suicide_rate ~ lag_log_numberofhategroups +
    lag_log_poverty_rate + lag_log_numberofhategroups:lag_log_med_household_income +
    lag_log_unemployment, data = pnel, model = "within")

Unbalanced Panel: n = 3103, T = 12-14, N = 43428

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-3.0604215 -0.0671786 -0.0049825  0.0442192  4.3197163

Coefficients:
                                                          Estimate Std. Error t-value  Pr(>|t|)
lag_log_numberofhategroups                              -2.5002178  0.6170350 -4.0520 5.088e-05 ***
lag_log_poverty_rate                                     2.9312047  0.2121372 13.8175 < 2.2e-16 ***
lag_log_unemployment                                    -0.0011699  0.0131084 -0.0892    0.9289
lag_log_numberofhategroups:lag_log_med_household_income  0.2322832  0.0575275  4.0378 5.406e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    17243
Residual Sum of Squares: 17137
R-Squared:      0.0061319
Adj. R-Squared: -0.070428
F-statistic: 62.1927 on 4 and 40321 DF, p-value: < 2.22e-16
>
> # Calculate robust standard errors
> robust_se <- vcovHC(model_with_interaction, type = "HC4")
>
> # Use coeftest to get robust standard errors
> robust_summaryse <- coeftest(model_with_interaction, vcov = robust_se)
>
> # Print the robust summary
> print(robust_summaryse)

t test of coefficients:

                                                          Estimate Std. Error t value  Pr(>|t|)
lag_log_numberofhategroups                              -2.5002178  0.5412631 -4.6192 3.864e-06 ***
lag_log_poverty_rate                                     2.9312047  0.2304893 12.7173 < 2.2e-16 ***
lag_log_unemployment                                    -0.0011699  0.0139502 -0.0839    0.9332
lag_log_numberofhategroups:lag_log_med_household_income  0.2322832  0.0500842  4.6378 3.532e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

References

Casant, J., & Helbich, M. (2022). Inequalities of suicide mortality across urban and rural areas: A literature review. *Multidisciplinary Digital Publishing Institute*, *19*(5), 2669. https://doi.org/10.3390/ijerph19052669

Cramer, R. J., Wright, S., Long, M. M., Kapusta, N. D., Nobles, M. R., Gemberling, T. M., & Wechsler, H. J. (2018). On hate crime victimization: Rates, types, and links with suicide risk among sexual orientation minority special interest group members. *Journal of Trauma & Dissociation*, *19*(4), 476-489.

Duncan, D. T., & Hatzenbuehler, M. L. (2014). Lesbian, gay, bisexual, and transgender hate crimes and suicidality among a population-based sample of sexual-minority adolescents in Boston. *American journal of public health*, *104*(2), 272-278.

Dunn, M. (2023). The impact of COVID-19 on social movements and extremist group formation. *Sociological Review, 51*(2), 245-263.

Durkheim, E. (1897). *Le suicide: Étude de sociologie*. Paris: Félix Alcan.

Ellison, J. M., Semlow, A. R., Jaeger, E. C., & Griffith, D. M. (2021). COVID-19 and mental health: Addressing men's mental health needs in the digital world. *SAGE Publishing, 15*(4). https://doi.org/10.1177/15579883211030021

Fontanella, C. A., Saman, D. M., Campo, J. V., Hiance-Steelesmith, D. L., Bridge, J. A., Sweeney, H. A., & Root, E. D. (2018). Mapping suicide mortality in Ohio: A spatial epidemiological analysis of suicide clusters and area level correlates. *Preventive Medicine, 106*, 177-184. https://doi.org/10.1016/j.ypmed.2017.10.033

Gunnell, D., Appleby, L., Arensman, E., Hawton, K., John, A., Kapur, N., Khan, M., O'Connor, R. C., & Pirkis, J. (2020). Suicide risk and prevention during the COVID-19 pandemic. *The Lancet Psychiatry, 7*(6), 468-471. https://doi.org/10.1016/S2215-0366(20)30171-1

Hirsch, J. K. (2006). A review of the literature on rural suicide. *Crisis, 27*(4), 189-199. https://doi.org/10.1027/0227-5910.27.4.189

Iemmi, V., Bantjes, J., Coast, E., Channer, K., Leone, T., McDaid, D., ... & Lund, C. (2016). Suicide and poverty in low-income and middle-income countries: A systematic review. *The Lancet Psychiatry, 3*(8), 774-783.

Iskander, J. K., & Crosby, A. E. (2021). Implementing the national suicide prevention strategy: Time for action to flatten the curve. *Elsevier BV, 152*, 106734. https://doi.org/10.1016/j.ypmed.2021.106734

Jones, A. (2017). Political rhetoric and hate group activity: A correlational study. *Political Science Quarterly, 132*(4), 639-657.

Kawachi, I., & Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban Health, 78*(3), 458-467.

Kegler, S. R., Stone, D. M., & Holland, K. M. (2017). Trends in suicide by level of urbanization — United States, 1999–2015. *Centers for Disease Control and Prevention, 66*(10), 270-273. https://doi.org/10.15585/mmwr.mm6610a2

Lorant, V., Kunst, A. E., Huisman, M., Bopp, M., & Mackenbach, J. (2005). Socio-economic inequalities in suicide: A European comparative study. *British Journal of Psychiatry, 187*(1), 49-54.

Maulana, I. N. H., & Wardah, T. F. (2023). Fostering community resilience through social capital. *Universitas Merdeka Malang, 1*(1), 1-10. https://doi.org/10.26905/j-tragos.v1i1.9229

Nock, M. K., Hwang, I., Sampson, N. A., & Kessler, R. C. (2008). Mental disorders, comorbidity, and suicidal behavior: Results from the national comorbidity survey replication. *Molecular Psychiatry, 15*, 868-876. https://doi.org/10.1038/mp.2008.29

Nock, M. K., Borges, G., Gilman, S. E., B., C., Kessler, R., & Lee, S. (2008). Suicide and suicidal behavior. *Oxford University Press, 30*(1), 133-154. https://doi.org/10.1093/epirev/mxn002

Prairie, K., Kivisto, A. J., Gray, S. L., Taylor, N., & Anderson, A. M. (2023). The association between hate crime laws that enumerate sexual orientation and adolescent suicide attempts. *Psychology, Public Policy, and Law*, *29*(2), 196.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Reeves, A., Stuckler, D., McKee, M., Gunnell, D., Chang, S. S., & Basu, S. (2012). Increase in state suicide rates in the USA during economic recession. *The Lancet, 380*(9856), 1813-1814. https://doi.org/10.1016/S0140-6736(12)61910-2

Rehkopf, D. H., & Buka, S. L. (2006). The association between suicide and the socio-economic characteristics of geographical areas: A systematic review. *Psychological Medicine, 36*(2), 145-157.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42*(1), 59-66.

Ryan, M. E., & Leeson, P. T. (2011). Hate groups and hate crime. *International Review of Law and Economics, 21*(4). https://doi.org/10.1016/j.irle.2011.08.004

Smith, J. (2015). Economic downturns and the rise of extremist groups. *Journal of Economic Perspectives, 29*(3), 17-35.

Smith, A. (2018). How radicalization to terrorism occurs in the United States: What research sponsored by the National Institute of Justice tells us. Retrieved from https://www.hsdl.org/?abstract&did=811946

Steelesmith, D. L., Fontanella, C. A., Campo, J. V., Bridge, J. A., Warren, K., & Root, E. D. (2019). Contextual factors associated with county-level suicide rates in the United States, 1999 to 2016. *JAMA Network Open, 2*(9), e1910936. https://doi.org/10.1001/jamanetworkopen.2019.10936

Stone, D. M., Simon, T. R., Fowler, K. A., Kegler, S. R., Yuan, K., Holland, K. M., ... & Crosby, A. E. (2018). Vital signs: Trends in state suicide rates—United States, 1999–2016 and circumstances contributing to suicide—27 states, 2015. *Morbidity and Mortality Weekly Report, 67*(22), 617. https://doi.org/10.15585/mmwr.mm6722a1

Southern Poverty Law Center. (n.d.). Groups. Retrieved from https://www.splcenter.org/

Szendro, B. (2021). Suicide, social capital, and hate groups in the United States. *World Affairs, 184*(4), 501-520. https://doi.org/10.1177/00438200211053889

Tadmon, D., & Bearman, P. (2023). Differential spatial-social accessibility to mental health care and suicide. *National Academy of Sciences, 120*(19). https://doi.org/10.1073/pnas.2301304120

U.S. Department of Health & Human Services. (n.d.). Suicide rate by county. Center for Disease Control and Prevention. Retrieved from https://wonder.cdc.gov/controller/datarequest/D158;jsessionid=739799D3CAA72A