

Study for Suicide Phenomenon and Prevention in the US with Data Analysis

Tianchong Chen *

Business Data Analytics, Arizona State University, 1151 S Forest Ave Tempe, AZ 85281, United States

*Corresponding author: Tchen119@asu.edu

Abstract. Suicide is a global problem and the US has a very high suicide rate, it is a very serious social problem and the leading cause of death in the United States. To prevent further deterioration of the suicide problem, many scholars focus on suicide data for research, which is essential for current situation in US. Based on this limitation in the other research, in this project, an open suicide data set was used to analyse the characteristics of suicide groups and the methods of suicide. This study also used multiple linear regression modelling to analyse the reasons for their suicides with good accuracy. By analysing the data, the proposed methods can identify the groups that are more likely to commit suicide and the methods they prefer to use, as well as the reasons that lead them to commit suicide, and the certain recommendations were made to reduce the suicide rate. It can help people to understand suicide better and can help the government to give people better guidance and take appropriate preventive measures.

Keywords: Suicide prevention, data analysis, linear regression.

1. Introduction

According to surveys nearly 800,000 people worldwide die by suicide each year. This figure may also be underestimated because suicide is illegal or stigmatised in some countries. Suicide is one of the leading causes of death worldwide. This is why suicide has become a serious public health issue [1]. In the United States, suicide is also a serious problem, with 45,979 people dying by suicide in 2020, about one death every 11 minutes [2]. Suicidal behaviour is a social problem. It not only affect the health of friends, relatives, colleagues and the community, but also has a huge economic cost to society.[2]Despite efforts to reduce suicide in the United States, the suicide rate is still on the rise and the annual cost to society is more than \$70 billion [3]. So the analysis of suicide is very important.

Based on the negative impact of suicide on American society, many social organisations and academics conducted researches on this issue. Their researches generally focuses on a particular aspect. For example, Yu and Chen used suicide data in the US from 1999-2017 to develop an APC model to analyse the age patterns of whites in the US who committed suicide by different methods, and finally to target different methods of suicide in specific age groups to reduce the additional suicide mortality among whites [4]. Rosalia's Bayesian analysis of suicide data identified points of change in the suicide rate series and concluded that suicide rates tend to level off as the level of economic and social development increases [5]. Kerry Shaw studied on the relationship between guns and suicide [6]. The Centers for Disease Control and Prevention (CDC) gave an analysis of suicide data and recommendations for suicide prevention [7].

From the above study, it can be found that scholars tend to focus on a particular aspect of suicide e.g. the causes of suicide or the methods of suicide. This study will build on their work to provide a comprehensive analysis. Using global suicide data to do the first step of analysis and then extract the data for the US, the proportion of people who commit suicide will be analysed by age, gender and ethnicity. Methods of suicide will be investigated and factors influencing suicide will be examined using regression models. Finally, recommendations will be given based on the results of the analysis.

The rest of this paper is organized as follows:section 2 provides our approach and data set. Experimental details and discussion will be given in section 3. The conclusions of this work and future work are discussed in section 4.

2. Datasets and Methods

2.1 Data sources

In this project, a public data set from Kaggle was used, which includes information about suicide worldwide from 1985 to 2016 [6]. The original data includes some basic information, including country, year, sex, age group, count of suicides, population, suicide rate, country-year composite key, HDI for year, gdpforyear, gdppercapita and generation (based on age grouping average).

2.2 Proposed model

Multiple Linear Regression Model: a regression model that estimates the relationship between a quantitative dependent variable and two or more independent variables using a straight line [7, 8]. The formula can be found in equation (1):

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \epsilon \quad (1)$$

X is the independent variable. Y is the dependent variable. This model is used for the identification of causes of suicide.

2.3 Evaluation indicator

R square that can determine the fit of a linear regression was chose in this study [9, 10]. A measure of model fit, R-squared can take any values between 0 to 1, the closer it is to 1, the better the model fit is. Th Formula is can be found in equation (2):

$$R \text{ square} = \frac{SS_{\text{regression}}}{SS_{\text{total}}} \quad (2)$$

$SS_{\text{regression}}$ is the sum of squares due to regression (explained sum of squares). SS_{total} is the total sum of squares. It is use to evaluate the multiple Linear Regression Model.

3. Results and Discussion

3.1 Data analysis

From the original data set, the ten countries with the highest number of suicides in the world from 1985 to 2016 were shown in Figure 1, they are Russia, USA, Japan, France, Ukraine, Germany, Republic of Korea, Brazil, Poland and United Kingdom, with the United States coming in second place.

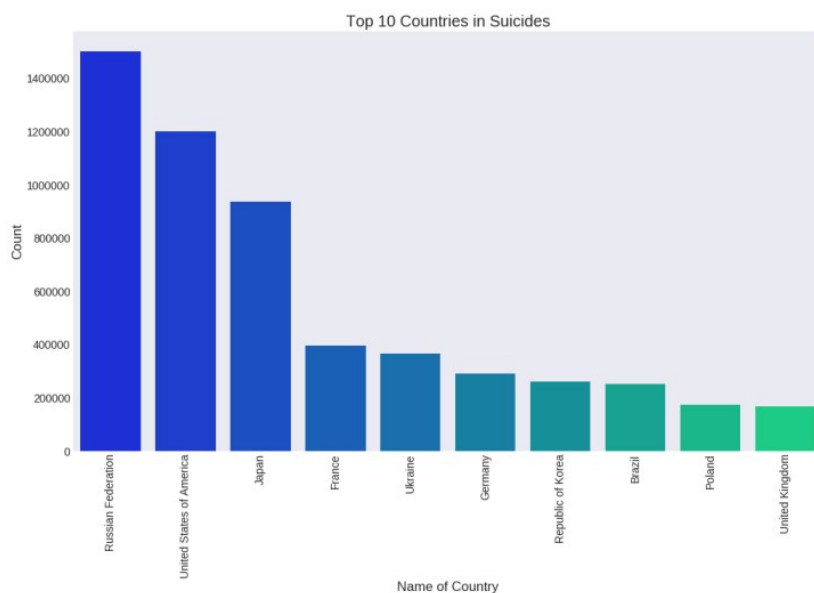


Figure 1. Top countries with the highest number of suicides

3.1.1 Demographic characters of suicide in the US

This study then filtered out the US data for analysis. From 1990 to 2020, about 1.1 million Americans died by suicide shown in Figure 2. The annual number of suicides in the United States showed an overall upward trend, increasing by 1.5 times from 30,000 to 45,000. Based on these facts, suicide is a serious problem in the United States which deserves the attention of the government.

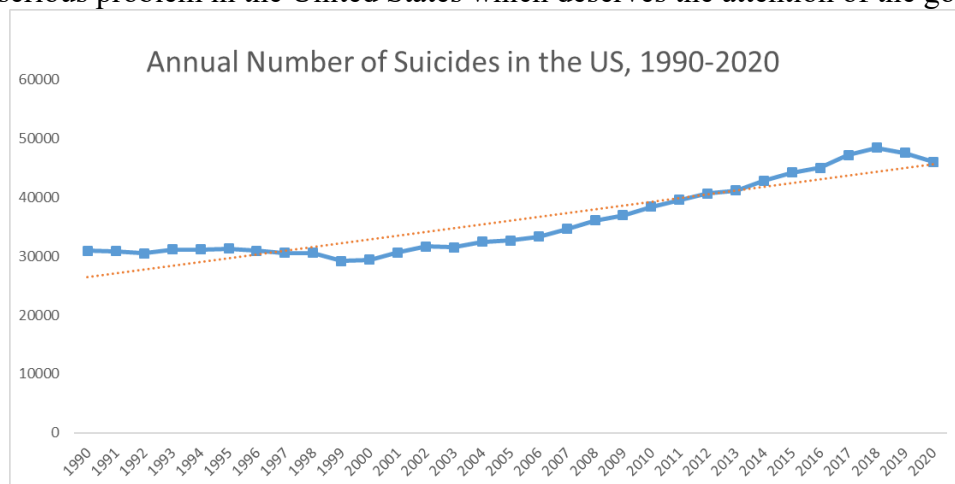


Figure 2. Annual number of suicides in the US from 1990-2020

Some groups have higher suicide rates than others. Suicide rates vary by sex, race, age, and many other factors. The Figure 3-5 show suicide rates by gender, race and age in the United States in 2020. Men are more likely to commit suicide than women. The proportion of males is 22.5% and the proportion of females is 5.6%. The suicide rate for men is about four times higher than for women.

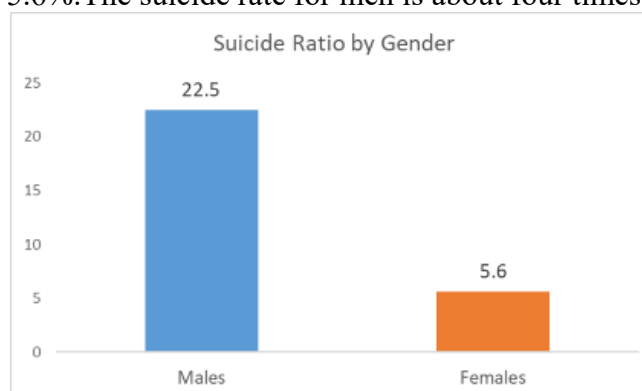


Figure 3. Suicide Ratio by Gender.

By race, the groups with the highest rates were Indian/Alaska Native and non-Hispanic White populations. Their respective percentages are 15.7% and 14.6%. The suicide rate for other races including Hispanic/Latino, Blacks/African, Asian/Pacific Islanders is basically around 7%.

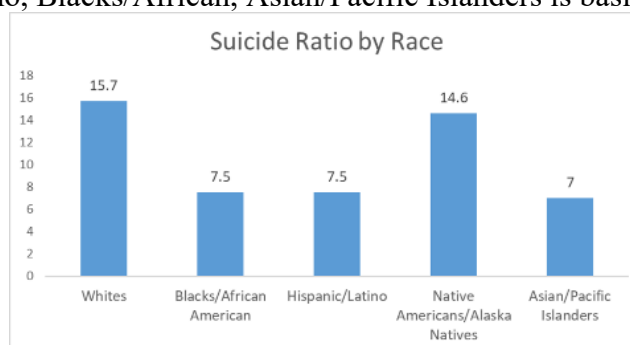


Figure 4. Suicide Ratio by Race

Suicide affects all ages. Adults aged 75 and older account for fewer than 10 % of all suicides but have the highest suicide rate. Adults aged 55–74 have lower rates of suicide than most of other adult

groups. The suicide rate for youth and young adults ages 15–24 years was lower than other age groups. However, suicide is the third leading cause of death for young people.

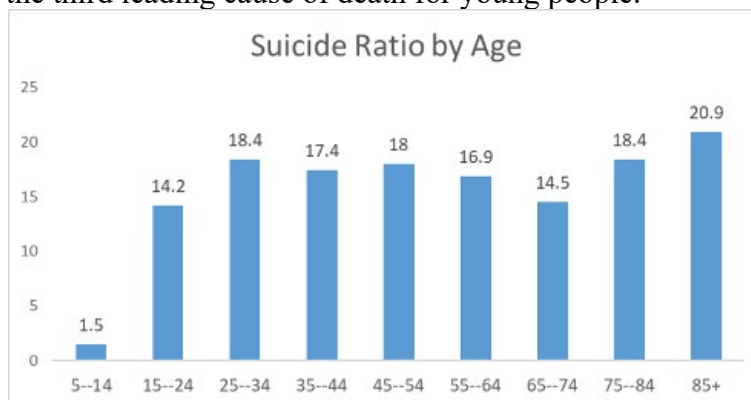


Figure 5. Suicide Ration By age

3.1.2 Method of suicide

In terms of method of suicide, the main forms of suicide in the United States are Firearm, Suffocation/Hanging, Poisoning, Fall, Cut/Pierce and Other methods. Their proportional distribution is shown in Figure 6, respectively 53%, 27%, 12%, 2%, 2%, 4%. Firearm is the most used method of suicide in the United States. More than 50 percent of Americans commit suicide with guns. This is a very important piece of information then this study move on to examine what type of gun they use to commit suicide.

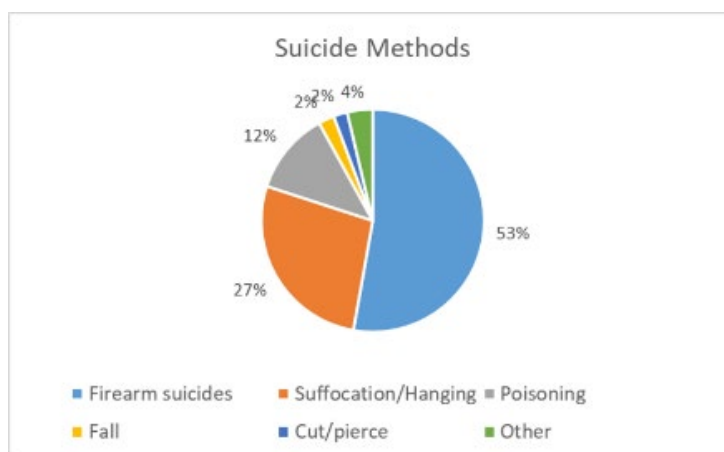


Figure 6. Suicide Methods.

Figure 7 shows states with higher gun ownership rates have significantly higher suicide rates than states with lower gun ownership rates. States with high gun ownership rates have a firearm suicide rate of 10.3, while states with low gun ownership rates only have a firearm suicide rate of 2.1, indicating a positive correlation between gun ownership and suicide rates.

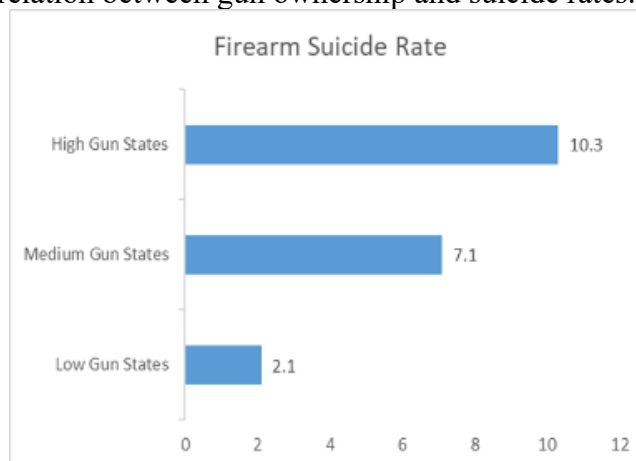


Figure 7. Firearm Sucide Rate.

The top five and last five states for suicide rates in the US are shown in the Figure 8, it can be surmised that states which is represented in blue are higher gun ownership states, which is Wyoming, Alaska, Montana, New Mexico and Idaho. In contrast, states which is represented in green are lower gun ownership states, which is Massachusetts, Rhode Island, New York, New Jersey and District of Columbia.

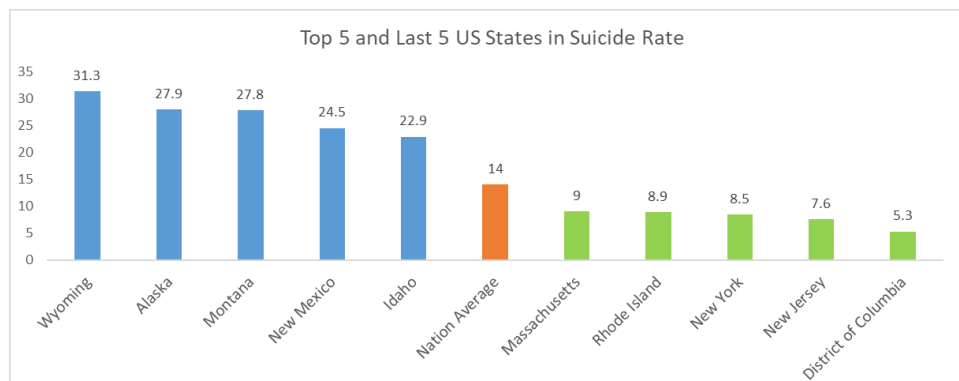


Figure 8. Top 5 and last 5 States in Suicide Rate.

Besides, handguns are used more in suicide than long gun. As shown in Figure 9, it has a ratio of 73% while the other ratio is 27%.

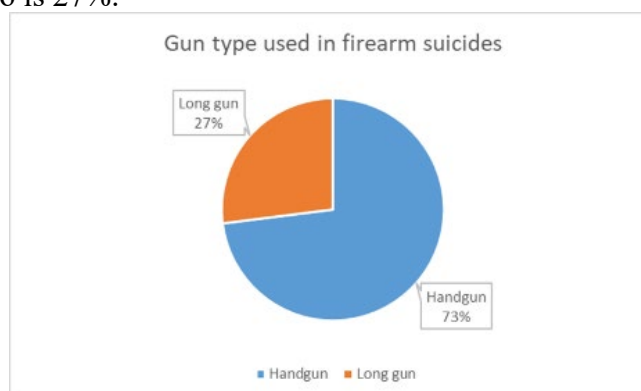


Figure 9. Gun type used in firearm suicides.

3.1.3 Factors affecting suicide rate

From the data set, it can be observed that there are two factors that affect suicide, one is internal cause of family relationships, and the other is the external cause of economic status.

Table.1. Factors affecting suicide rate

State	Divorce rate	Capita.personal income	Suicide rate
Alabama	3.7	46479	16
Alaska	3.2	63502	27.5
Arizona	2.9	49648	17.6
Arkansas	3.6	47235	19
California	2.5	70192	10

Based on this, a Multiple Linear Regression Model is made. The dependent variable is the suicide rate of each state in the United States in 2020 shown in Table 1, and the independent variables are the divorce rate and per capita.personal income of each state in 2020, representing family relationships and economic status respectively. This is achieved by the following equation (3).

$$Y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + u \quad (3)$$

Where, Y, X1, X2 represent Suicide Rate by state in 2020, Divorce Rate by state in 2020, Per Capita Personal Income by state in 2020, respectively.

The results are shown in Figure 10.

OLS Regression Results						
Dep. Variable:	Suicide Rate	R-squared:	0.431			
Model:	OLS	Adj. R-squared:	0.407			
Method:	Least Squares	F-statistic:	18.15			
Date:	Fri, 04 Mar 2022	Prob (F-statistic):	1.35e-06			
Time:	22:45:32	Log-Likelihood:	-141.80			
No. Observations:	51	AIC:	289.6			
Df Residuals:	48	BIC:	295.4			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	14.3932	5.714	2.519	0.015	2.904	25.883
Divorce Rate	3.9606	1.015	3.902	0.000	1.920	6.002
Per Capita Personal Income	-0.0001	6.93e-05	-2.099	0.041	-0.000	-6.12e-06

Figure 10. QLS regression results.

Then, Substitute the results into the formula:

$$Y = 14.3932 + 3.9606 * X1 - 0.0001 * X2 + u \quad (4)$$

From the Figure 10, the adjusted R square of the model is 43.1, showing that the model fits well. The coefficients of both variables are significant at the 5 level. The coefficient of the divorce rate is negative, indicating that the increase in the divorce rate will also increase the suicide rate. The coefficient of per capita income is positive, indicating that the increase in per capita income will reduce the suicide rate.

3.2 Discussion

From the results of the above experiment, the suicide rate in the United States continues to rise, with the group more likely to commit suicide being white males and those over the age of 75, a group that should then be widely focused. A large proportion of their suicides are by gun and there is a positive ratio between gun ownership and suicide rates. Then gun use remains a problem and states with high gun ownership rates should do something to prevent people from committing suicide. Finally, divorce rates and personal income are two important factors that influence suicide, so states can regulate this issue to reduce suicide rates. Based on this analysis, the following recommendations were made:

1) First, identify and support people at risk, then strengthen access and delivery of suicide care. It is necessary to find out those who have a high tendency to commit suicide in time, such as elderly people without financial sources and teenagers who encounter problems in school, and then they should be given care and help, psychological counseling and subsidies. Schools, companies and communities must do their part to reduce suicide rates.

2) The second point is to implement stricter gun control measures. The statistics prove that high gun ownership rate leads to high suicide rate. It is impossible to ban guns in The United States, but guns can be strictly controlled. Before buying guns, strict background checks can be carried out to ensure that those who buy guns cannot have criminal records or suicidal tendencies.

3) The third point is to strengthen economic supports. Financial support is also important in reducing suicide rates. Economic pressure is a main reason why many people choose to commit suicide. Therefore, the government should provide financial subsidies to those who are under economic pressure, provide relief and jobs to the unemployed. At the same time, raising the minimum wage is also quite necessary. When people's economic problems are solved, the suicide rate will surely decrease.

Our study is more comprehensive than other studies on suicide and gives more comprehensive recommendations. It is a further extension of previous studies in that it covers various aspects of the characteristics of the suicide population, suicide patterns, states with high suicide rates, and causes of suicide.

4. Conclusions

The experiment completes the analysis of the suicide data. Using data analysis tools and multiple linear regression models this study got the results of the study. It can be concluded that men, whites and Native Americans who are prone to suicide and over 75 years of age have the highest suicide rates. Gun suicide tends to be their most common method of suicide and handguns are used more often than long guns in suicides, and gun suicide rates are positively correlated with gun ownership rates. Increases in divorce rates increase suicide rates but increases in per capita income decrease them.

In the future, this study will continue to crawl the latest suicide data in the United States for research. Creating a webpage to visualize the data and make the information more visible to the government and individuals. And this study will be working on more models to find more causes of suicide and ways to reduce the suicide rate. Further study will also try to work with a suicide prevention organisation to provide advice to help those who want to commit suicide to give up the idea.

References

- [1] Ritchie, Hannah, Max Roser, and Esteban Ortiz-Ospina. Suicide [J]. Our World in Data. N.p., 2022. Web. 30 Apr. 2022.
- [2] Centers for disease control and prevention. Facts About Suicide [R]. <https://www.cdc.gov/suicide/facts/index.html>, 2022
- [3] Moutier, Christine, Patrick J. Kennedy, The Rate Of Suicides In The United States Is Growing – What Can We Do? [R]. World Economic Forum, 2019.
- [4] Yu, B., and X. Chen. Age Patterns of Suicide with Different Methods for US Whites: APC Modelling Analysis of the 1999–2017 National Data [J]. *Epidemiology and Psychiatric Sciences*, vol. 29, 2020, p. e180., doi:10.1017/S204579602000092X.
- [5] Condorelli, Rosalia. A Bayesian Analysis Of Suicide Data [J]. *Quality & Quantity* 47.2 (2011): 1143-1161., doi:10.1007/s11135-011-9608-9.
- [6] Kaggle. Suicide Rates Overview 1985 To 2016. Kaggle.com.
- [7] Bevans, R. An introduction to multiple linear regression [M]. Statistics (2020).
- [8] Weisberg, Sanford. Applied linear regression [J]. Vol. 528. John Wiley & Sons, 2005.
- [9] Teoh, S. H., Yang, Y. G., & Zhang, Y. R-square and market efficiency [J]. Available at SSRN 926948, 2009
- [10] Akossou, A. Y. J., and Rodolphe Palm. Impact of data structure on the estimators R-square and adjusted R-square in linear regression [J]. *Int. J. Math. Comput* 20.3 (2013): 84-93.