|  |
| --- |
| USF: MUMA College Of Business |
| Factors Affecting Suicide in USA |
| A Statistical Study |

|  |
| --- |
| Subash Chandra Biswal Surendra Varma Sagi Shailendra Singh  11-7-2022 |

Table of Contents

[Executive Summary 2](#_Toc118057632)

[Problem Statement and Significance 3](#_Toc118057633)

[Prior Literature 4](#_Toc118057634)

[Data Source and Preparation 5](#_Toc118057635)

[Data Structure 6](#_Toc118057636)

[Feature Engineering 7](#_Toc118057637)

[Variables 8](#_Toc118057638)

[Exploratory Data Analysis and Visualizations 10](#_Toc118057639)

[Models 16](#_Toc118057640)

[Assumptions Test 17](#_Toc118057641)

[Insights and Recommendations 25](#_Toc118057642)

[References 26](#_Toc118057643)

Executive Summary

Suicide is a major public health concern in the United States (US), affecting individuals across age groups, resulting in severe and prolonged effects on families and communities. In fact, currently it is the 12th leading cause of death in the United States. In 2020, 45,979 Americans died by committing suicide, among those roughly 4 times were men as compared to women. That averages roughly 130 suicides per day. The WHO and the US Surgeon General highlighted the need for more comprehensive data on the occurrence of suicidal thoughts and attempts, for planning national health care policy to reduce suicide-related behaviors.

To understand what contributes the most to suicide ideation or attempts, our project attempts to find the causes and factors that are most responsible for the same. After our analysis, we also attempt to provide some recommendations considering the factors, to find some solution to this problem.

For our analysis, we referred to a lot of prior research as well as considered various data sources. We were able to narrow down some factors that were most common. The features included age, mental health, gun access, economic factors, and so on. We used various data sources such as the CDCs Wonder database, The Bureau of Labor Statistics data for economic factors, as well as a propriety gun friendly index [7] to understand gun laws strictness in a state.

In our analysis we found that the white men were disproportionately affected by suicides as well as the senior population of the United States. We provide a systematic analysis of our process and provide recommendations to Government or Non-Profit organizations to reduce suicide attempts.

# Problem Statement and Significance

Suicide is one of the leading causes of death worldwide. Suicide is an act of self-harm by an individual to end his/her life. The problem with this crisis is that it is not caused by just one factor, but a myriad and complex factors ranging from individual experiences with bullying, mental distress up to societal factors such as unemployment, race & ethnicity factors, etc. Also, suicide affects across age groups, hence the studying the causes becomes far more complex.

Suicide has far reached impacts to not just the individuals but to their loved ones including family and society. Suicide of an individual causes a chain effect of mental distresses among the people that individual connected with during their life. These mental distress factors can trigger suicide ideations in others as well.

The financial toll of suicide on society is also costly, according to the CDC [8]. In 2019, suicide and nonfatal self-harm cost the nation nearly $490 billion in medical costs, work loss costs, value of statistical life, and quality of life costs.

Hence it is a very important crisis to solve in today's modern society. Suicide is not a random phenomenon; it occurs due to several factors in the due course of time in an individual's life. Thus, it is a very preventable crisis, if we provide interventions during the suicide ideation stage and provide better mental health service to those that are affected by it. States and communities can use the Prevention Resource to help make decisions about suicide prevention activities. Strategies range from those designed to support people at increased risk to a focus on the whole population, regardless of risk [8].

# Prior Literature

There is a lot of research in this area. Some of the prior literatures are mentioned below:

Reeves, Stuckler (2012) [1] stated that economic depression or recession is one of the leading causes of suicide rate. Their paper highlights that there was an increase in suicide mortality rate due to the onset of the crisis from 1999 to 2007. As unemployment increases, so do the suicides and this could be observed in the 2007-10 period due to the financial crisis of 2008.

Compton, Druss (2021) [2] highlighted how suicide rates increase more rapidly in rural areas compared to urban areas. They also attributed lack of mental health treatment as one of the leading causes for the increase in suicide rate. In fact, the counties that were classified as mental health shortage areas (HPSAs), experienced a rise in suicide rates compared to non HPSAs.

The research by Kessler, Borges [3] tried to understand suicide ideation. They used a total of 9282 interviews that were conducted in which respondents were asked questions related to suicide ideation/attempts or gesture. The data were then examined for the associations of suicide related behaviors with mental disorders using WHO Composite International Diagnostic Interview (CIDI) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-III and DSM-IV). The associations were analyzed using sociodemographic variables: age, sex, race/ethnicity, education, marital status, employment status, and region of the country.

As mentioned in the paper Understanding Suicide Across the Lifespan [4], Suicidal ideation risk factors broadly falls under two categories: Static risk factors and Dynamic risk factors. Static risk factors are stable/fixed attributes, which imbue a baseline risk; they include sex, race, age, sexual orientation, family history, and personal history of suicide attempts. Dynamic factors fluctuate throughout an individual's life; they include symptoms of mental illness, substance use, firearm possession, and access to healthcare.

Access to guns was also considered a most important factor in suicide deaths. Pritchard C, Hansen [5] described one feature that might be considered cultural’ in the USA compared to Other Western Countries is the easy access to firearms, where in most US States eighteen-year-olds can purchase firearms. This despite evidence that more than 50% of suicides in America are gun-related and children’s suicides are often the result of using the parent’s weapons.

Similarly, Martinez, Hernandez, Khaul [6] stated that firearm suicide is the most common method of suicide in the United States, accounting for one in two suicides among males and one in three among females, suicides would be expected to closely relate to access to firearms. However, even though this suicide method represents a plurality of suicide deaths in the United States, recent data suggest that firearm ownership is decreasing, rather than increasing, in the United States

# Data Source and Preparation

|  |  |  |
| --- | --- | --- |
| ***Link*** | ***Description*** | ***Method*** |
| <https://wonder.cdc.gov/> | WONDER online databases utilize a rich ad-hoc query system for the analysis of public health data. | Filtered the data based on all the causes of deaths which are mapped to suicide, and group by State, year, race, age group from year 2005 to year 2020. |
| <https://www.bls.gov/> | The **Bureau of Labor Statistics** measures labor market activity, working conditions, price changes, and productivity in the U.S. economy to support public and private decision making. | Few Variable data such as CPI is collected from database. Variables such as unemployment, minimum wage, mortgage rate, income tax are collected from multiple files and merged |
| <https://www.az-defenders.com/best-states-for-gun-owners/> | Ranked all states in a proprietary Gun Friendly Index, calculated by assigning scores to multiple parameters and weighing them based on importance to gun owners. At the end, this was then weighted based on the overall culture and sentiment toward firearms in the state. | Collected the score of each state from published report. |

# Data Structure

* Records **17, 898** for years 2005 -2020
* 14 important variables to consider include:
  + State
  + Year
  + Age Group
  + Race
  + Urbanization
  + Population
  + Unemployment Rate
  + Income Tax
  + Mortgage Rate
  + Minimum Wage
  + Consumer Price Index
  + Mental Health Shortage
  + Gun Rule Ranking
  + Divorce rate
  + **Suicide rate**

# Feature Engineering

# Some possible feature engineering we may have to do based on Prior Work research

# Consolidated urbanization areas to 2 types such as Urban/Rural

# Factorized all categorical variables.

# Merged all the data for each variable by state and year.

# Dropped the unwanted variables.

# From different income slabs, only the tax rate for income around USD20,000 or greater than USD 10000 are considered assuming low income leads to suicide.

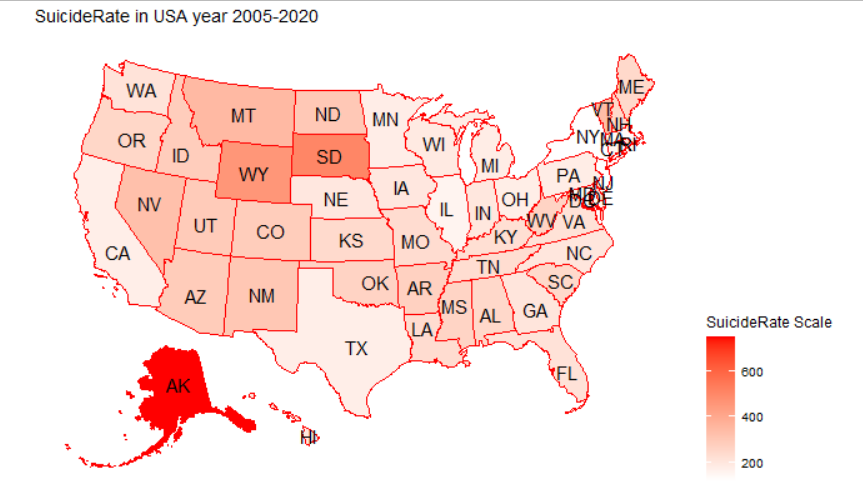
# The mean of unemployment rate and minimum wage of each state for any year is calculated and used.

# Suicide rate is calculated per million people for each state.

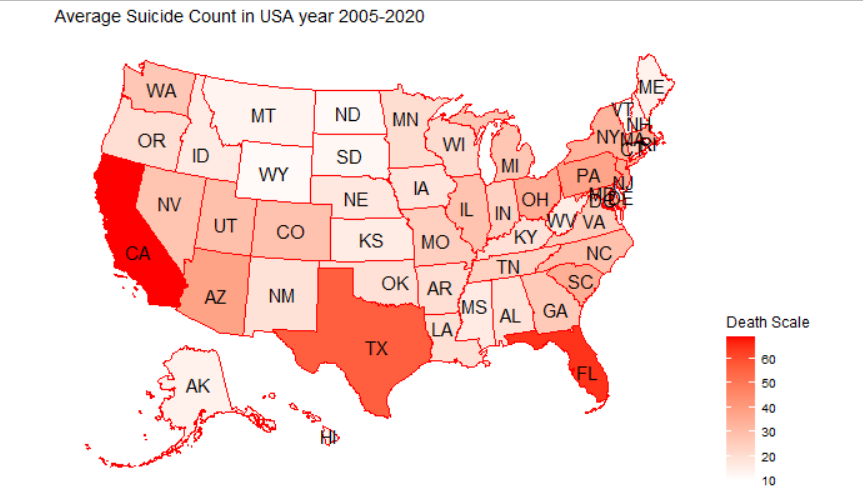
# Variables and Predictor table

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Description** | **Effect** | **Rationale** |
| Year | Years ranging from 2005-2020 | +/- | Suicide rate can vary from year to year. |
| State | All the 51 states of United States | +/- | Suicide rate varies for each state. |
| Urbanization | Indicates whether the region is urban or rural. | - | Studies show that suicide rate is more in rural areas than urban areas, might be because of unemployment and low income. |
| Race | Indicates race of the people. | +/- | Suicide rate varies for each race. |
| Agegroups | Indicates the age group of people | + | Elder people are more inclined towards suicide due to lack of emotional support and better living conditions. |
| Unemprate | Indicates unemployment rate in particular state and year. | + | If the unemployment rate is more, then chances are high for mental health deprivation, and this may lead to suicide. |
| Incometax | Indicates tax rate of each state in each year. | + | More tax rates on the income can affect the lives of people, and this may increase the suicide rate. |
| Minwage | Indicates Minimum wage in each state and year. | + | If the wages are less, it can affect the lives of people, and this may increase the suicide rate. |
| Divorcerate | Indicates divorce rate in each state and year. | + | If the divorce rate is more, it indicates the presence of people with unstable mental condition, and this can lead to increase in suicide rate. |
| Mortgagerate | Indicates the mortgage rate in each state and for each year. | + | More mortgage rates can affect the lives of people, and this may increase the suicide rate. |
| gunscore | Indicates the Gun score of each state, where lower gun score refers to stricter gun laws. | + | More stricter the gun laws, chances are less to use guns and this can reduce the suicide rate. |
| Mental health | Indicates the mental health shortage in each state. | + | More the mental health shortage, chances are more for people committing suicide. |

# Exploratory Data Analysis and Visualizations



The above map shows the suicide rate by state. Alaska has the highest suicide rate among 51 states and New York seems has lowest suicide rate.

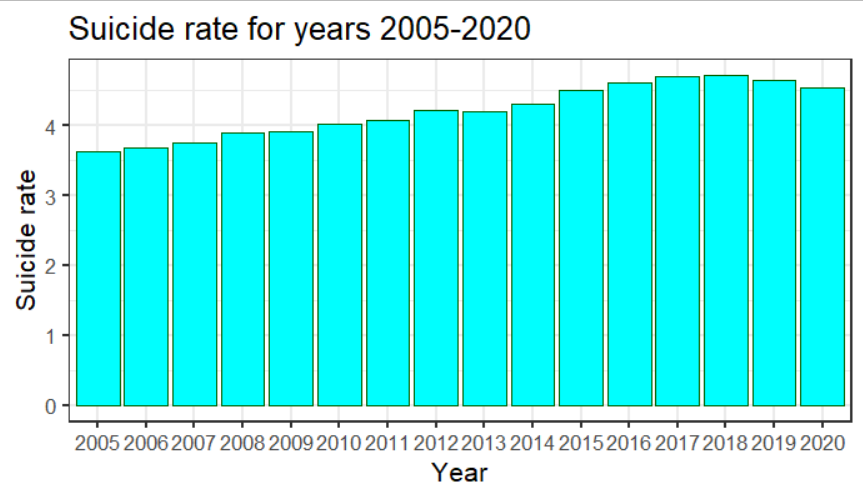


The above map shows the average suicide count by state. It shows that average suicide death count is highest in California, Texas, and Florida (Top 3) , and lowest in North Dakota, South Dakota, and Wyoming (Bottom 3).

Chart, bar chart, histogram

Description automatically generated

The above bar chart shows the suicide death trend from year 2005 to year 2020. While it increased from 2005 to 2018, it decreases after 2018.



The above bar chart shows the suicide rate trend from year 2005 to year 2020. Similar to suicide death count, it increases from 2005 to 2018 but decreases thereafter. But, the rate at which it increases from 2005 to 2018 is more than the rate it decreases after 2018.

Chart

Description automatically generated

The above chart shows the gun score for each state in ascending order. Lower score means stricter gun control policies. It shows that New York has the strictest gun policies and Alaska has the most lenient gun control policies. This exactly reflects the suicide rate trend. The higher the rank in gun score, lower the suicide rate.

Chart, histogram

Description automatically generated

The above chart compares the suicide rate with unemployment rate. It seems that as the unemployment rate increases the impact is reflected in coming years. Higher the unemployment rate higher the suicide rate.

Chart, histogram

Description automatically generated

The above chart compares the suicide rate with mortgage rate. It seems that as the unemployment rate increases the impact is reflected in coming years. Higher the mortgage rate higher the suicide rate.

Chart, bar chart, histogram

Description automatically generated

The above chart compares the suicide rate with divorce rate. Surprisingly as the divorce rate decreases the suicide rate increases. Its possibly because of lower marriage rate. People are remaining single and that is causing stress, less emotional support and so the higher suicide rate.

Chart, bar chart, histogram

Description automatically generated

The above chart compares the suicide rate with income tax rate. It shows similar trend though the suicide rate gets impacted late. Higher the income tax higher the suicide aret in coming years.

Chart, bar chart

Description automatically generated

The above chart compares suicide rate with Consumer Price Index trend. There is no clear trends to relate them.

Chart, bar chart, histogram

Description automatically generated

The above chart compares the suicide rate with minimum wage. It shows that with increase in minimum wage, the suicide rate decreases but it takes time to take effect.

Diagram

Description automatically generated with medium confidence

The above correlation plot shows that there is strong correlation (>0.7) between CPI and divorce rate, mortgage rate and CPI, mortgage rate and divorce rate. To avoid multicollinearity, we can remove the CPI from the model.

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated

The log of suicide rate (DV) is closer to normal distribution. So, we will use exponential model for this data set.

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated

The log transformation of unemployment rate is closer to normal distribution.

The data has multilevel variables such as year and state. So, our model would be multilevel with exponential model.

# Models

Based on the visual analysis of the data, we will use the LMER exponential models from LME4 library for random/mixed effects.

The 3 models are:

**basemodel**<- lmer(log(suiciderate) ~ log(unemprate) + minwage + incometax + mortgagerate+ agegroups + urbanization + divorcerate +race + mentalhealth + (1 | year) + (1 | state), data=dfinal\_new2, REML=FALSE) # **With random effects of year and state**

**random** <- lmer(log(suiciderate) ~ population +log(unemprate)+ minwage + gunscore + incometax + mortgagerate+ agegroups + urbanization + divorcerate + race + mentalhealth + (1 | year) + (1 | state), data=dfinal\_new2, REML=FALSE) # **With population and gunscore**

**interaction** <- lmer(log(suiciderate) ~ population +log(unemprate)+ minwage + incometax + mortgagerate+ agegroups + urbanization + divorcerate + race + mentalhealth + gunscore + divorcerate + population\*log(unemprate) + mentalhealth+ (1 | year) + (1 | state), data=dfinal\_new2, REML=FALSE) # **With interactions**

Below is the stargazer comparison of 3 models.

Table

Description automatically generated

*Random effect with interaction Equation:*

Log(Ysuiciderate)= 4.895 – 0.023Xlog(unemprate) – 0.002Xminwage + 0.004Xgunscore + 0.003Xincometax -0.035Xmortgagerate + 2.265Xagegroups15-24 + 2.530Xagegroups25-34 + 2.625Xagegroups35-44 + 2.716Xagegroups45-54 + 2.585Xagegroups55-64 + 2.422Xagegroups65-74 + 2.607agegroups75-84 + 2.742Xagegroup85+ - 0.137XurbanizationUrban - 0.190Xdivorcerate - 2.231XAsian - 2.191XBlack - 1.519XWhite + 0.295Xmentalhealth

1. If the unemployment rate is increased by 100%, the suicide rate will decrease by 2.3%.
2. For every unit increase in minimum wage the suicide rate will decrease by 0.2%.
3. For every unit increase in gun score, suicide rate will increase by 0.4%.
4. For every unit increase in income tax, suicide rate will increase by 0.3%.
5. For every unit increase in mortgage rate, suicide rate will decrease by 3.5%.
6. For people of age group 15-24, the suicide rate will increase by 22.7% compared to people of age group 0-14.
7. For people of age group 25-34, the suicide rate will increase by 25.3% compared to people of age group 0-14.
8. For people of age group 35-44, the suicide rate will increase by 26.3% compared to people of age group 0-14.
9. For people of age group 45-54, the suicide rate will increase by 27.2% compared to people of age group 0-14.
10. For people of age group 55-64, the suicide rate will increase by 25.9% compared to people of age group 0-14.
11. For people of age group 65-74, the suicide rate will increase by 24.2% compared to people of age group 0-14.
12. For people of age group 75-84, the suicide rate will increase by 26% compared to people of age group 0-14.
13. For people of age group 85+, the suicide rate will increase by 27.4% compared to people of age group 0-14.
14. For urban area the suicide rate decreases by 13.7% as compared to rural area.
15. For unit increase in divorce rate the suicide rate decreases by 19%.
16. For Asian the suicide rate decreases by 22.3% as compared to American Indian or Alaska Natives.
17. For black people the suicide rate decreases by 22% as compared to American Indian or Alaska Natives.
18. For White Americans the suicide rate decreases by 15.2% as compared to American Indian or Alaska Natives.
19. For every unit increase in mental health shortage area, the suicide rate increases by 29.5%.

# Assumptions Test

**Multi-Collinearity: Pass**

**A picture containing text

Description automatically generated**

There is collinearity between population and unemployment rate and so we have used interaction.

**Homoscedasticity: Not Clear**

Chart, scatter chart

Description automatically generated

**Linearity: Pass**

Chart, line chart

Description automatically generated

The plot show quite normality except at both tails.

# Insights and Recommendations

**Insights:**

* American Indian or Alaska Natives have highest risk of suicide. They have 15% more chance of suicide than White Americans and 22% more chance than Black or Asian people.
* People of age 85+ are at highest risk of suicide death.
* People in rural area are at more risk of suicide death than urban area people.

**Recommendation**:

* State gun laws seem to affect access, which in turn is a known factor for most suicides. Hence, stricter gun laws (based on background checks, etc.) could limit access and hence could prevent suicide attempts by non-eligible gun owners.
* Mental health is a contributing factor, hence increasing diagnostic services as well as support services could decrease suicides due to mental health issues such as depression, PTSD, etc.

# References

[1]. 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.

[2]. Ku, B. S., Li, J., Lally, C., Compton, M. T., & Druss, B. G. (2021). Associations between mental health shortage areas and county-level suicide rates among adults aged 25 and older in the USA, 2010 to 2018. General hospital psychiatry, 70, 44-50.

[3]. Kessler RC, Berglund P, Borges G, Nock M, Wang PS. Trends in Suicide Ideation, Plans, Gestures, and Attempts in the United States, 1990-1992 to 2001-2003. JAMA. 2005;293(20):2487–2495.

[4]. Ian H. Steele M.D.,Natasha Thrower M.D.,Paul Noroian M.D.,Fabian M. Saleh M.D. Understanding Suicide Across the Lifespan: A United States Perspective of Suicide Risk Factors, Assessment & Management

[5]. Pritchard C, Hansen L, Dray R, Sharif J. USA Suicides Compared to Other Western Countries in the 21st Century: Is there a Relationship with Gun Ownership? Arch Suicide Res. 2022 Jan 24:1-13. doi: 10.1080/13811118.2021.1974624. Epub ahead of print. PMID: 35068366.

[6]. Martinez-Ales G, Hernandez-Calle D, Khauli N, Keyes KM. Why Are Suicide Rates Increasing in the United States? Towards a Multilevel Reimagination of Suicide Prevention. Curr Top Behav Neurosci. 2020; 46:1-23. doi: 10.1007/7854\_2020\_158. PMID: 32860592; PMCID: PMC8699163.

[7]. “Gun Laws by State: Best States for Gun Owners (2022).” AZ Defenders, 2 Feb. 2022, https://www.az-defenders.com/best-states-for-gun-owners/. “Gun Laws by State: Best States for Gun Owners (2022).” AZ Defenders, 2 Feb. 2022, https://www.az-defenders.com/best-states-for-gun-owners/.

[8]. "Facts About Suicide | Suicide | CDC." https://www.cdc.gov/suicide/facts/index.html, Accessed 29 Oct. 2022.

# Appendix

##Base Data Cleaning

rm(list=ls())

library(stargazer)

library(dplyr)

library(PerformanceAnalytics)

library(tidyr)

library(ggplot2)

library(usmap)

library(viridis)

library(lattice)

library(lme4)

library(readxl)

library(plm)

library(data.table)

library(tidyverse)

library(gapminder)

library(forcats)

theme\_set(theme\_bw(base\_size=16))

setwd("E:/BAIS/3rd Sem/SDM/Project/Data")

##Merging datasets

d <- read\_excel("Suicide Data.xlsx", sheet=1)

d$`Ten-Year Age Groups Code` <- ifelse(d$`Ten-Year Age Groups Code`=="44695","5-14",d$`Ten-Year Age Groups Code`)

d2 <- read\_excel("Suicide Data.xlsx", sheet=2)

d3 <- read\_excel("Suicide Data.xlsx", sheet=3)

d4 <- read\_excel("Suicide Data.xlsx", sheet=4)

d5 <- read\_excel("Suicide Data.xlsx", sheet=5)

d6 <- read\_excel("Suicide Data.xlsx", sheet=6)

d7 <- read\_excel("Suicide Data.xlsx", sheet=7)

d8 <- read\_excel("Suicide Data.xlsx", sheet=8)

d9 <- read\_excel("Suicide Data.xlsx", sheet=9)

colnames(d2) <- c("Year", "MortgageRate")

colnames(d3) <- c("State","Year","UnempRate")

colnames(d4) <- c("State","Year","IncomeTax")

colnames(d5) <- c("Year","State","MinWage")

colnames(d6) <- c("InflationRate","Year","CPI")

colnames(d9) <- c("State","MentalHealth")

table(d6$Year)

table(d5$year)

table(d4$year)

table(d$State)

table(d5$State)

d5 <- d5[d5$State!="U.S. Virgin Islands",]

d5 <- d5[d5$State!="Guam",]

d5 <- d5[d5$State!="Puerto Rico",]

d6 <- d6[d6$Year>2004,]

d6 <- d6[d6$Year!=2021,]

d8 <- d8[d8$Year>2004,]

d8 <- d8[d8$Year!=2021,]

library("dplyr")

df <- merge(d3,d4,by=c("Year","State"))

df2 <- merge(df,d5,by=c("Year","State"))

df3 <- merge(df2,d6,by=c("Year"))

df4 <- merge(df3,d8,by=c("Year"))

df5 <- merge(df4,d2,by=c("Year"))

df6 <- merge(d7,d9,by=c("State"))

df7 <- merge(df5,df6,by=c("State"))

df7 <- df7[order(df7$Year),]

dfinal <- merge(d,df7,by=c("Year","State"))

dfinal$`2013 Urbanization` <- ifelse(dfinal$`2013 Urbanization`=="Large Central Metro" | dfinal$`2013 Urbanization`=="Large Fringe Metro"

| dfinal$`2013 Urbanization`=="Medium Metro" | dfinal$`2013 Urbanization`=="Small Metro", "Urban", "Rural")

dfinal\_new <- dfinal[,-c(4,5,6,8,9,13,17)]

names(dfinal\_new) <- tolower(names(dfinal\_new))

str(dfinal\_new)

names(dfinal\_new)[names(dfinal\_new) == '2013 urbanization'] <- 'urbanization'

names(dfinal\_new)[names(dfinal\_new) == 'ten-year age groups code'] <- 'agegroups'

names(dfinal\_new)[names(dfinal\_new) == 'gfi'] <- 'gunscore'

dfinal\_new$`agegroups` <- ifelse(dfinal\_new$`agegroups`=="44695","5-14",dfinal\_new$`agegroups`)

table(dfinal\_new$agegroups)

library("writexl")

write\_xlsx(dfinal,"E:/BAIS/3rd Sem/SDM/Project/Data/Suicide FinalNew2.xlsx")

dfinal\_new$suiciderate = dfinal\_new$deaths/(dfinal\_new$population/1000000)

##Visualizations

#install.packages("usmap")

library(usmap)

##Grouped for getting total sum between years 2005-2020.

dummy <- aggregate(dfinal\_new$suiciderate, by=list(state=dfinal\_new$state), FUN=mean)

str(dummy)

names(dummy)[names(dummy) == 'x'] <- 'suiciderate'

plot\_usmap(data = dummy, values = "suiciderate", color = "Red", labels = TRUE) +

scale\_fill\_continuous(low = "White", high = "Red", name = "SuicideRate Scale", label = scales::comma) +

labs(title = "SuicideRate in USA year 2005-2020") +

theme(legend.position = "right")

dummy1 <- aggregate(dfinal\_new$death, by=list(state=dfinal\_new$state), FUN=mean)

str(dummy1)

names(dummy1)[names(dummy1) == 'x'] <- 'death'

plot\_usmap(data = dummy1, values = "death", color = "Red", labels = TRUE) +

scale\_fill\_continuous(low = "White", high = "Red", name = "Death Scale", label = scales::comma) +

labs(title = "Average Suicide Count in USA year 2005-2020") +

theme(legend.position = "right")

dummy\_year\_deaths <- aggregate(dfinal\_new$deaths, by=list(year=dfinal\_new$year), FUN=sum)

str(dummy\_year\_deaths)

names(dummy\_year\_deaths)[names(dummy\_year\_deaths) == 'x'] <- 'totaldeaths'

dummy\_year\_deathrate <- aggregate(dfinal\_new$suiciderate, by=list(year=dfinal\_new$year), FUN=mean)

str(dummy\_year\_deathrate)

names(dummy\_year\_deathrate)[names(dummy\_year\_deathrate) == 'x'] <- 'deathrate'

dummy\_year\_deathrate$deathrate <- dummy\_year\_deathrate$deathrate/51

ggplot(dummy\_year\_deaths) +

geom\_bar(aes(x=year, y=totaldeaths),stat="identity", fill="cyan",colour="#006000") +

geom\_line(aes(x=year, y=totaldeaths),stat="identity",color="red") +

labs(title= "Suicide counts for years 2005-2020",

x="Year",y="Suicide Deaths")

ggplot(dummy\_year\_deathrate) +

geom\_bar(aes(x=year, y=deathrate),stat="identity", fill="cyan",colour="#006000") +

geom\_line(aes(x=year, y=deathrate),stat="identity",color="red") +

labs(title= "Suicide rate for years 2005-2020",

x="Year",y="Suicide rate")

##Factorizing Variables

dfinal\_new$state <- as.factor(dfinal\_new$state)

dfinal\_new$year <- as.factor(dfinal\_new$year)

dfinal\_new$urbanization<- as.factor(dfinal\_new$urbanization)

dfinal\_new$race <- as.factor(dfinal\_new$race)

dfinal\_new$agegroups <- as.factor(dfinal\_new$agegroups)

#dfinal\_new$gunscore <- as.factor(dfinal\_new$gunscore)

dfinal\_new$agegroups <- relevel(dfinal\_new$agegroups,"5-14")

str(dfinal\_new)

library(PerformanceAnalytics)

temp <- dfinal\_new[, -c(1,2,3,4,5)] # Numeric variables only

chart.Correlation(temp)

##Created Dummy Variables for Visualization

dummy\_gun\_score <- aggregate(dfinal\_new$gunscore, by=list(state=dfinal\_new$state), FUN=mean)

names(dummy\_gun\_score)[names(dummy\_gun\_score) == 'x'] <- 'gunscore'

##gunscores for different states

ggplot(dummy\_gun\_score)+

geom\_bar(aes(x=reorder(state,+gunscore), y=gunscore),stat="identity", fill="cyan",colour="#006000")+

theme(axis.text.x = element\_text(angle = 90)) + labs(title= "Gun Score for States", x="State", y="Gun Score")

## Unemployment Rate Vs. Death Rate

dummy\_unemp\_year <- aggregate(dfinal\_new$unemprate, by=list(year=dfinal\_new$year), FUN=mean)

names(dummy\_unemp\_year)[names(dummy\_unemp\_year) == 'x'] <- 'unemprate'

dummy\_unemp\_deaths <- merge(dummy\_unemp\_year,dummy\_year\_deathrate,by = c("year"))

str(dummy\_unemp\_deaths)

dummy\_unemp\_deaths$year <- as.factor(dummy\_unemp\_deaths$year)

ggplot(dummy\_unemp\_deaths, aes(x=year))+

geom\_bar(aes(y=unemprate, group=1),stat= "identity", fill="cyan",colour="#006000")+

geom\_line(aes(y=deathrate, group=1),stat="identity",color="red")+scale\_y\_continuous(

"Unemp Rate", sec.axis = sec\_axis(~.\*1,name = "Suicide Rate (In every Million)")) +

labs(title= "Years 2005-2020",

x="Year")

## Mortgage Rate Vs. Death Rate

dummy\_mortgage\_year <- aggregate(dfinal\_new$mortgagerate, by=list(year=dfinal\_new$year), FUN=mean)

names(dummy\_mortgage\_year)[names(dummy\_mortgage\_year) == 'x'] <- 'mortgagerate'

dummy\_mortgage\_deaths <- merge(dummy\_mortgage\_year,dummy\_year\_deathrate,by = c("year"))

str(dummy\_mortgage\_deaths)

ggplot(dummy\_mortgage\_deaths)+ scale\_y\_continuous(

"Mortgage Rate",

sec.axis = sec\_axis(~.\*1,name = "Suicide Rate")

) +

geom\_bar(aes(x=year, y=mortgagerate,group=1),stat="identity", fill="cyan",colour="#006000")+

geom\_line(aes(x=year, y=deathrate,group=1),stat="identity",color="red")+

scale\_y\_continuous(

"Mortgage Rate", sec.axis = sec\_axis(~.\*1,name = "Suicide Rate (In every Million)")) +

labs(title= "Years 2005-2020",

x="Year")

## Divorce Rate Vs. Death Rate

dummy\_divorce\_year <- aggregate(dfinal\_new$divorcerate, by=list(year=dfinal\_new$year), FUN=mean)

names(dummy\_divorce\_year)[names(dummy\_divorce\_year) == 'x'] <- 'divorcerate'

dummy\_divorce\_deathrate <- merge(dummy\_divorce\_year,dummy\_year\_deathrate,by = c("year"))

str(dummy\_mortgage\_deaths)

ggplot(dummy\_divorce\_deathrate)+ scale\_y\_continuous(

"Divorce Rate",

sec.axis = sec\_axis(~.\*1,name = "Suicide Rate")

) +

geom\_bar(aes(x=year, y=divorcerate,group=1),stat="identity", fill="cyan",colour="#006000")+

geom\_line(aes(x=year, y=deathrate,group=1),stat="identity",color="red") +

scale\_y\_continuous(

"Divorce Rate", sec.axis = sec\_axis(~.\*1,name = "Suicide Rate (In every Million)")) +

labs(title= "Years 2005-2020",

x="Year")

## Income Tax Rate Vs. Death Rate

dummy\_tax\_year <- aggregate(dfinal\_new$incometax, by=list(year=dfinal\_new$year), FUN=mean)

names(dummy\_tax\_year)[names(dummy\_tax\_year) == 'x'] <- 'incometax'

dummy\_tax\_deathrate <- merge(dummy\_tax\_year,dummy\_year\_deathrate,by = c("year"))

str(dummy\_tax\_deathrate)

ggplot(dummy\_tax\_deathrate)+ scale\_y\_continuous(

"Income Tax Rate",

sec.axis = sec\_axis(~.\*1,name = "Suicide Rate")

) +

geom\_bar(aes(x=year, y=incometax,group=1),stat="identity", fill="cyan",colour="#006000")+

geom\_line(aes(x=year, y=deathrate,group=1),stat="identity",color="red") +

scale\_y\_continuous(

"Income Tax Rate", sec.axis = sec\_axis(~.\*1,name = "Suicide Rate (In every Million)"))+

labs(title= "Years 2005-2020",

x="Year")

## CPI Vs. Death Rate

dummy\_cpi\_year <- aggregate(dfinal\_new$cpi, by=list(year=dfinal\_new$year), FUN=mean)

names(dummy\_cpi\_year)[names(dummy\_cpi\_year) == 'x'] <- 'cpi'

dummy\_cpi\_deathrate <- merge(dummy\_cpi\_year,dummy\_year\_deathrate,by = c("year"))

str(dummy\_cpi\_deathrate)

ggplot(dummy\_cpi\_deathrate)+ scale\_y\_continuous(

"CPI",

sec.axis = sec\_axis(~.\*1,name = "Suicide Rate")

) +

geom\_bar(aes(x=year, y=cpi,group=1),stat="identity", fill="cyan",colour="#006000")+

geom\_line(aes(x=year, y=deathrate\*100,group=1),stat="identity",color="red")+

scale\_y\_continuous(

" CPI", sec.axis = sec\_axis(~.\*1,name = "Suicide Rate (In every 100 Million)")) +

labs(title= "CPI vs Suicide Rate for years 2005-2020",

x="Year")

## Minimum wage Vs. Death Rate

dummy\_minwage\_year <- aggregate(dfinal\_new$minwage, by=list(year=dfinal\_new$year), FUN=mean)

names(dummy\_minwage\_year)[names(dummy\_minwage\_year) == 'x'] <- 'minwage'

dummy\_minwage\_deathrate <- merge(dummy\_minwage\_year,dummy\_year\_deathrate,by = c("year"))

str(dummy\_minwage\_deathrate)

ggplot(dummy\_minwage\_deathrate)+ scale\_y\_continuous(

"Minimum Wage",

sec.axis = sec\_axis(~.\*1,name = "Suicide Rate")

) +

geom\_bar(aes(x=year, y=minwage,group=1),stat="identity", fill="cyan",colour="#006000")+

geom\_line(aes(x=year, y=deathrate,group=1),stat="identity",color="red")+

scale\_y\_continuous(

" Minimum Wage", sec.axis = sec\_axis(~.\*1,name = "Suicide Rate (In every 1 Million)")) +

labs(title= "Years 2005-2020",

x="Year")

#Models

attach(dfinal\_new)

hist(deaths)

hist(suiciderate)

hist(log(suiciderate))

hist(unemprate)

hist(log(unemprate))

hist(incometax)

hist(log(incometax))

hist(minwage)

hist(log(minwage))

hist(cpi)

hist(log(cpi))

hist(mortgagerate)

hist(log(mortgagerate))

hist(divorcerate)

hist(log(divorcerate))

colSums(is.na(dfinal\_new))

str(dfinal\_new)

dfinal\_new2 <- subset(dfinal\_new,dfinal\_new$year!=2020)

table(dfinal\_new2$year)

dfinal\_2020 <- subset(dfinal\_new,dfinal\_new$year==2020)

dfinal\_2020 <- dfinal\_2020[,-c(6)]

dfinal\_2020 <- dfinal\_2020[complete.cases(dfinal\_2020), ]

dfinal\_2020 <- dfinal\_2020[,-c(15)]

dfinal\_2020 <- dfinal\_2020[,-c(10)]

dfinal\_2020 <- dfinal\_2020[,-c(7)]

attach(dfinal\_new2)

library(lme4)

basemodel<- lmer(log(suiciderate) ~ log(unemprate) + minwage + incometax + mortgagerate+ agegroups + urbanization + divorcerate +

race + mentalhealth + (1 | year) + (1 | state), data=dfinal\_new2, REML=FALSE)

summary(basemodel)

ranef(basemodel)

vcov(basemodel)

**random** <- lmer(log(suiciderate) ~ population +log(unemprate)+ minwage + gunscore + incometax + mortgagerate+ agegroups + urbanization + divorcerate + race + mentalhealth + (1 | year) + (1 | state), data=dfinal\_new2, REML=FALSE)

summary(mixed)

ranef(mixed)

**interaction** <- lmer(log(suiciderate) ~ population +log(unemprate)+ minwage + incometax + mortgagerate+ agegroups + urbanization + divorcerate + race + mentalhealth + gunscore + divorcerate + population\*log(unemprate) + mentalhealth+ (1 | year) + (1 | state),

data=dfinal\_new2, REML=FALSE)

summary(interaction)

ranef(interaction)

library(stargazer)

stargazer(basemodel, mixed, interaction, single.row = TRUE,type = "text")

#Assumptions Test

library(car)

vif(interaction) #Multi-collinearity

plot(interaction) #Homoscedasticity

qqnorm (resid(interaction))

qqline(resid(interaction), col="red")