



STEVENS
INSTITUTE *of* TECHNOLOGY
THE INNOVATION UNIVERSITY®

Informatics for Engineering Management

EM 624 Fall 2021 – Final Project

ANALYSIS ON SUICIDE IN USA

Author: HARSH SHAH

Guidance: Dr. CARLO LIPPIZZI

Date: 28th November

Overview of Suicide Cases in U.S.

We hear often that someone committed suicide, also sometimes we get to know the reason behind it. What is Suicide? - Suicide is death caused by injuring oneself with intent to die. As many factors leads to increase in number of suicides like depression, money problems, child abuse, bullying, sexual violence etc. Major suicide cases involves guns to injure themselves with intent to die. According to Google suicide rates increased 33% between 1999 and 2019. Suicide is the 10th leading cause of death in the United States which is not a good number. It was responsible for about 48000 deaths in 2019 which is about one death every 11 minutes. And there are many who get suicidal thoughts on regular basis and are undergoing a treatment or therapy. Being connected to community or your loved ones or family can lead to decrease in this thoughts and ultimately leads to decrease in suicides.

Suicide is a spot on the socio-economical life of the citizen of any country. To end a life due to failure is never a good solution. We can always have good times awaiting our struggle. However, this is a data analysis done on suicides using guns for year 2012, 2013, 2014. Here we will see any trends in suicides and try to understand are there any relations between factors.

After all this information some questions may arise in our mind, Does gender play any role? ,Is there any particular community or race?, What are the major reasons?. In this project I have done a detailed data analysis on suicides involving guns and data is of between year 2012 and 2014. And on the basis of this I will answer the following questions and present the final conclusions and analyze whether there are any linkages or trends between factors like race, gender, education or age.

Following are the questions which are going to answer:

- To analyze the number of cases for suicides, homicides, accidental etc to have a clear idea of the data for further analysis only on suicides.
- To analyze which year has highest number of suicide cases.
- Is one gender more prone to suicide than the other one?
- Which age has highest number of suicides?
- Is one Race more prone to suicide than any other?
- Does education have any role in suicides ?
- What was the place where suicide took place? (Just to have idea, maybe used later)

By answering this question we will have a basic idea about the data then further we will try to find linkage between factors and try to answer the following questions

- Does age play any role in suicides . Does age play any role in suicides over Race?
- Does one gender is more prone to suicides than other for a particular race?
- Is combination of age and race is more prone to suicide than other?
- To know increases in cases of race over 3 years.
- To analyze the relation between gender, age and number of suicides with this we can have clear idea that, Does age make difference leading to suicide?, And if yes, whether it is for male or female.
- From WHO statistics file we try to find about number of cases increased from 2005 to 2011 and distributed it with respect to gender.
- From same data we compared total suicides and suicides with the use of guns and tried to find a answer whether implementing gun laws can make a difference in reduction of suicides.
- Also extracted information from article of Washington post to support our claim to implement strict gun laws.

Research Motivation:

Suicide is always the hot topic in the society. Research helps us have an idea about the correlation between some factors and maybe helping our ability to identify people at risk for suicide. People should have idea of information related to suicide to prevent it. I am self-motivated towards this topic in order to help prevent it in any possible way. This analysis report will thus give idea about various factors related to suicides and try to understand is there a pattern.

Data Description

URL: Data from github

[mailto: https://github.com/fivethirtyeight/guns-data/blob/master/full_data.csv](https://github.com/fivethirtyeight/guns-data/blob/master/full_data.csv)

URL: Data from Kaggle

Article: [mailto:https://www.washingtonpost.com/graphics/business/wonkblog/suicide-rates/](https://www.washingtonpost.com/graphics/business/wonkblog/suicide-rates/)

Source of data: Github Guns Death In America Project, WHO suicides report over various countries.

Dataset:

The Guns Death dataset contains 100,798 samples and it contains 10 different variables, out of which we used only 6 of them as month, police, place, Hispanic are dropped as they are of no use for our analysis. Raw file was converted to CSV file from Github, cleaning was not required as the data in file was already clear. This data is from year 2012 to 2014. This dataset contains data of suicides, homicides, accidental and underdetermined but our research is only on suicide so we will work on data related to only suicide. In this the 6 main variables are year, intent, sex, age, race and education.

The WHO suicide statistics data contains data of all countries and variables like gender and age group is stated. We are working only with the gender and calculate total number of cases from this in order to give conclusion related to our study. There was no need to clean the data as it was in form of CSV file.

Paragraphs were extracted from website of Washington post article and the data was clean using the Stopwords text file and all digits and punctuations were removed and unrelated words for analysis were also removed and further analysis is done on this clean data.

- In Year it is indicated in which suicide occurred like 2012, 2013 and 2014.
- Intent if filled with data whether it was a suicide, homicide, accidental or undetermined.
- Sex includes the gender of the person.
- Race includes the race of person like White, black, Native American, Asian/Pacific Islander and Hispanic.
- Age includes the age of the person and range of age is being used while plotting the data to have understanding.
- Education gives us the understanding related to the education level of the person.

Data Preparation:

- Firstly the raw data is converted in to CSV file and saved to respective folder.
- Then with the help of PANDAS library it is converted into Data Frame and data with intent suicide is only printed.
- Data was cleared and no need of cleaning it.
- Some of the data in file is missing and indicated with 'NA'.
- Variables like: place, police, Hispanic, month are not used because it is not useful to us in analyzing.
- Empty list for every factor is created to store total values at every point.

- For WHO suicides file was open using CSV reader and converted to lists to read data and initializing variables to store data for various conditions.
- For data from article we used “BeautifulSoup” and extracted all paragraphs and then data was clean and “stopwords” and unnecessary words were removed. Even punctuation digits were removed which were of no use in analysis.
- Also here words like ‘guns’, ‘Americans’ etc, are removed as they are of no use.

Methodology

- We used Python for our data analysis and IDE Pycharm is used.
- First we read the csv file using the pandas and converting it into Data Frame.
- Next we imported the necessary libraries like “PANDAS”, “CSV”, “MATPLOTLIB.PYPLOT”, “BeautifulSoup”, “Wordcloud”
- Then we converted intent column values to list for the requirement of our first output.
- Next we work with the data of suicides cases only.
- Using pandas we narrow down our columns to suicides only and for every factor we calculated related values.
- Then using ‘if’ functions we conditioned the various cases and find the statistics. For some factors function was also defined
- Count list for every factor is created to count the values.
- Hence for every factor we got count list and with this we plot the Pie-chart with percentage indicated on it for our required output.
- And then initializing this output list into graph plot conditions.
- For WHO data we initialize variable to store total cases values and same using “if” conditions we compare only total suicides with data from guns used. And also we compare gender cases increase over the years and also compare which gender has more tendency for suicide.
- For our recommendations section we took a article from Washington post and try to analyze it with the help of WordCloud. For this using “Beautifulsoup” we extracted all paragraph and cleaned the whole data and converted into list first and then to string in order to plot the Wordcloud.

Result

- Out of our data following are the cases for suicides, homicides, accidental respectively.

```
The number of suicides are 63175  
The number of homicides are 35176  
The number of accidental are 1639
```

Fig. 1

Hence we will work only on cases related to suicide. (Fig. 1- Output from python)
This indicates guns are majorly used in committing suicides. About 63.18% gun used cases are Suicides, which is a big number to taken into consideration

- **Year wise distribution of Suicide cases:**

```
There were total 20666 suicide cases in 2012.  
There were total 21175 suicide cases in 2013.  
There were total 21334 suicide cases in 2014
```

Fig .2 (Image from python output)

Pie chart- Suicide cases distribution over 3 years

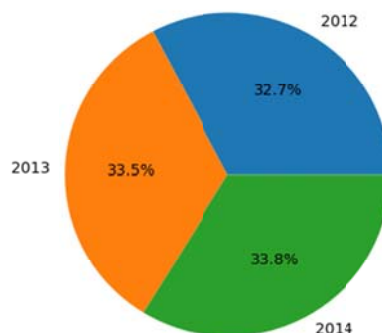


Fig. 3

Here Fig. 3 indicates cases distribution over the years and according to this there are almost same cases over the years. And there is no vast increase or decrease in number of suicides.

- **Gender Wise Distribution**
Is one gender more prone to suicide than the other one?

```
There were total 54486 males who committed suicides.  
There were total 8689 females who committed suicides.
```

Fig.4 (Image from python output)

Here below is the pie chart which indicates Distribution over gender for 3 years and it indicates that 86.2 % of Suicides are committed by males. And only 13.8% of suicides were of females, which is vast difference. Hence we can conclude that males are more prone to suicides than females. For further analysis we also extracted data from 2005 to 2011 (Fig.6) and found that number of males committing suicides are almost 5 times that of women. This cases are increasing over years.

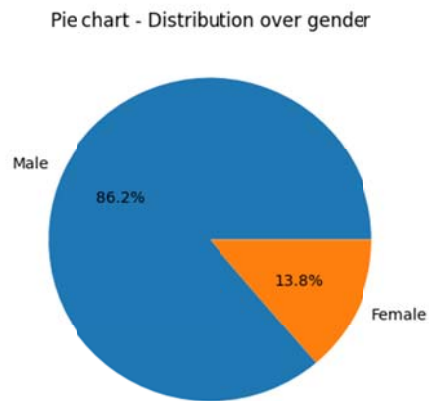


Fig.5

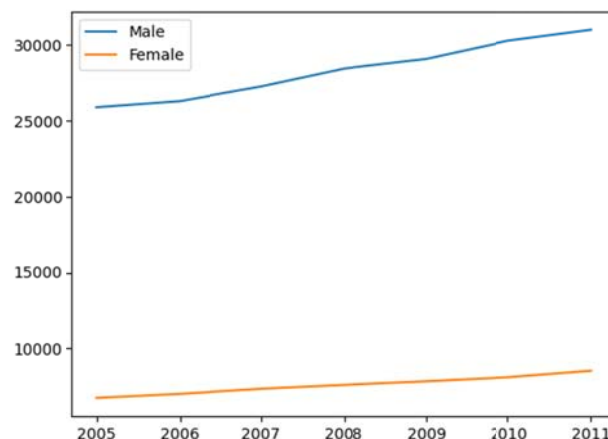


Fig.6

- **Race Wise Distribution:**
Is one race more prone to suicide than any other?

```
There were total 745 suicide cases for Asian/Pacific Islander race.
There were total 55372 suicide cases for White race.
There were total 555 suicide cases for Native American/Native Alaskan race.
There were total 3332 suicide cases for Black race.
There were total 3171 suicide cases for Hispanic race.
```

Fig. 7 (Output from python)

Pie chart- Suicide case distribution over race

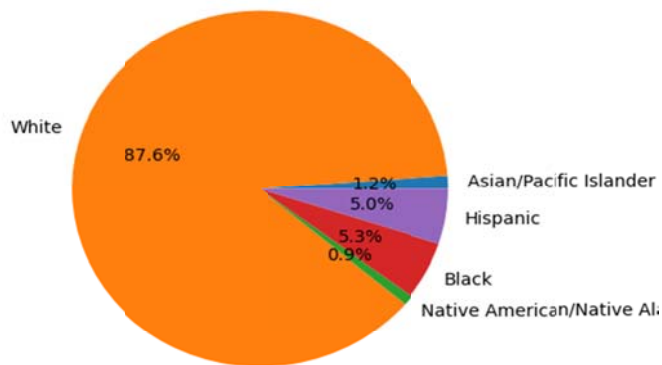


Fig. 8

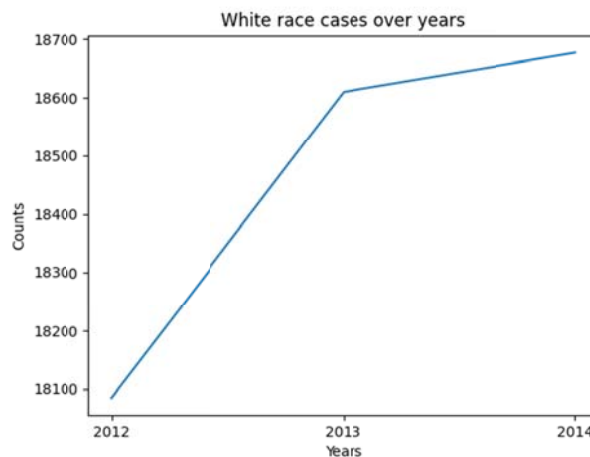


Fig. 8A

Here in above Fig.7 it clearly indicates that 87.6% of cases are of White race which is huge number hence white people are more prone to suicides than other races. And black are only 5.3% which is not more in compare to total number of cases and in compare to white race, and 5% were Hispanic race. And 1.2% were Asian/Pacific Islander and 0.9% were Native American/Native Alaska.

Also plotted if there is a increase in cases of white race people over 3 years. There was only a increase of about 550 cases in 2013 and 100 in 2014.

➤ Education wise distribution:

Does education have any role in suicides ?

Here from figure 9 it seems that maximum people were HS/GED educated which is about 41.7% and second highest people went to some college and 17.6% people are BA + in education hence majority of them are educated. And only 14.7% are less than HS. 1.4% cases are not stated in data. Here people with HS/GED as there education are prone to suicide than others.

Pie chart- Suicide case distribution over education

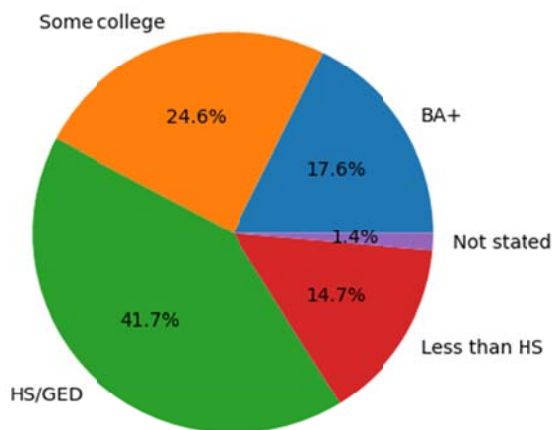


Fig. 9

➤ **Age wise analysis:**

Does age play a role in occurrence of suicides?

It seems maximum cases were 13049 and between 50-60 age group. Second most number cases were 10095 and between 40-50 age group. Third most number of cases were 9185 and between 60-70 age group. Fourth most number of cases were 8831 and between 20-30 age group. Fifth most number of cases were 8147 and between 30-40 age group.

Here from figure it is clearly visible that maximum number of cases are between 50-60 age group. And the second most number of cases are between 40-50 age group. Third most number of cases are in 60-70 age group. And fourth number of cases are in 20-30 age group. And fifth most number of cases are between 30-40 age group. Hence here people with age group of 50-60 are more prone of committing suicide than any other age group.

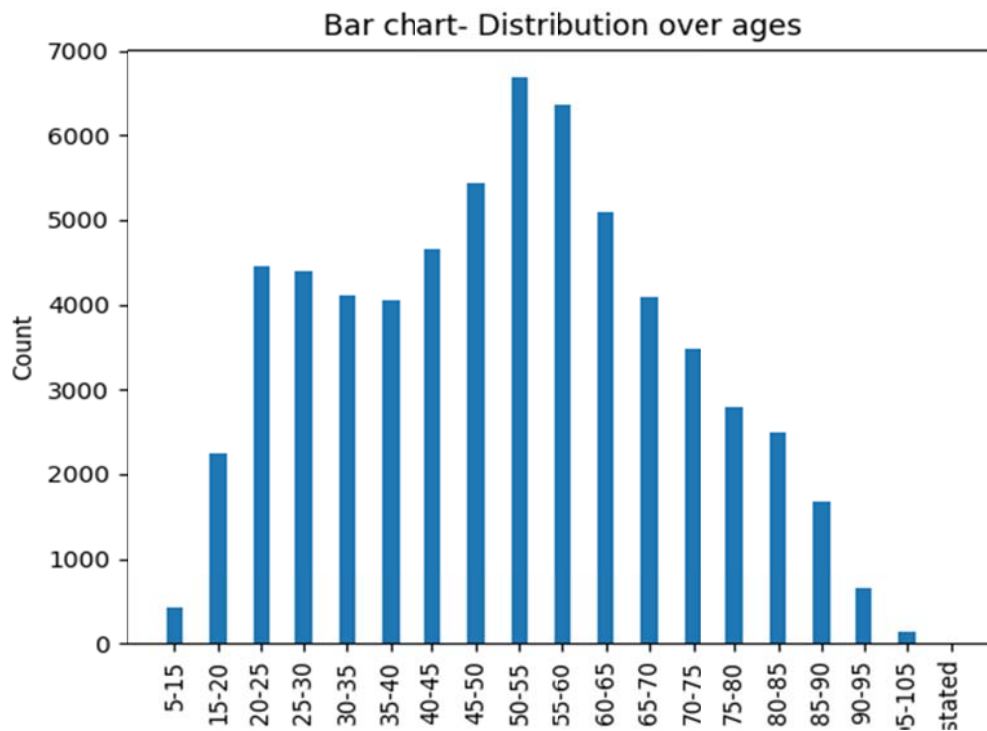


Fig. 10

- **Comparison with factor : Age vs Race comparison.**
Is combination of age and race is more prone to suicide than other?

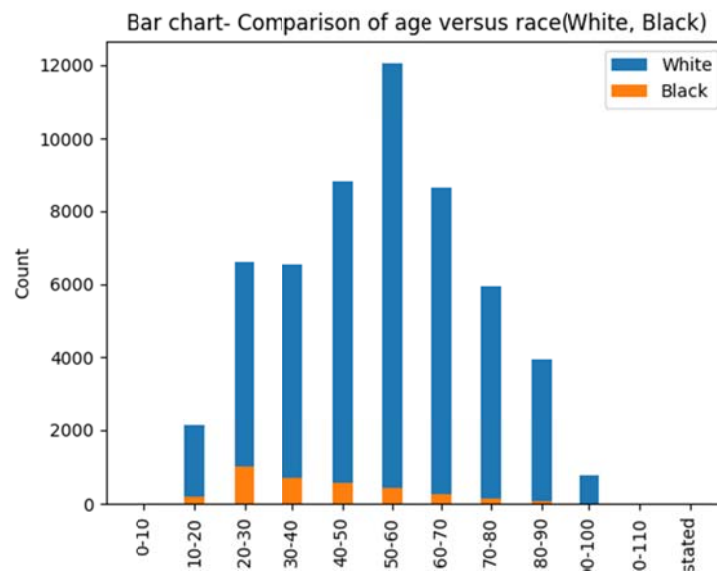


Fig. 11

Hence from fig. 11 it seems that for white race also maximum suicides were between 50-60. And then to 40-50 and 60-70 are almost same. And age group 20-30 and 30-40 are also same. In compare to white race there very negligible suicides among black community and all other.

➤ **White race and Gender**

Does one gender is more prone to suicides than other for a particular race?

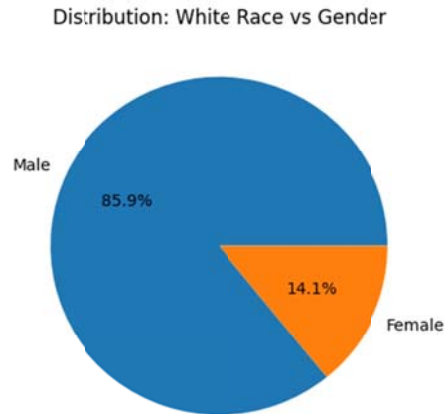


Fig. 12

Here from figure 12 it seems that out of total guns used for suicides there are 87.6% of white race out of which there are 85.9% cases that are white male and only 14.1% cases that are white female. Hence white males are more prone to suicide than white females or any other race.

➤ **White race and Education**

Does any race and education is more prone to suicides than any other?

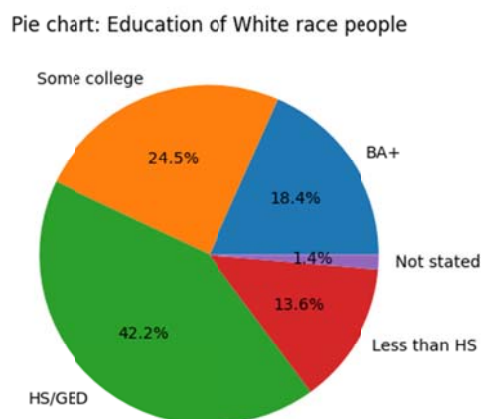


Fig. 13

Here from the above Fig. 8 there are 87.6% of white people and in that almost 42.2 % of them have HS/GED as their education hence they are prone to suicide and 24.5% of them went to some college, 18.4% have BA+ as their education and 13.6% have education less than 13.6%. Hence if a person is white and have education as HS/GED he has more chances of committing suicide than any other race or other education people.

➤ **Total cases compared with cases with guns used:**

This analysis was done in order to have an idea that out of total cases in , how many cases guns were used as a mode for suicide.

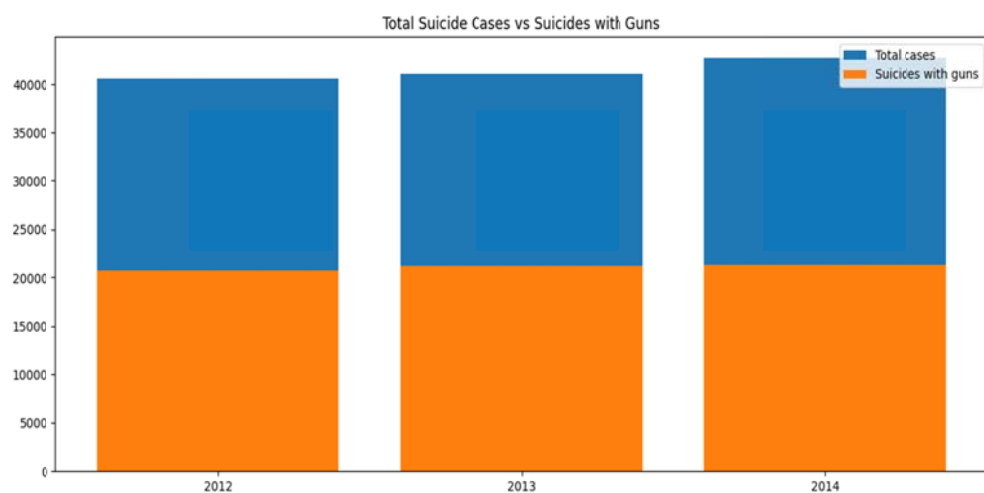


Fig. 14

By seeing Fig.14 it seems that in almost 50% cases out of total guns were used as a mode for suicide. Hence and as per upper data maximum cases are of males hence males use more firearm as a mode for suicide. By implementing strict gun laws we can greatly reduce the suicide rate. Maybe they can use another method but atleast it gets delayed. According to this people having guns have more tendency to perform suicide as majority cases are of suicides from above data.

For this also picked a article from Washington post and try to analyze it by plotting WordCloud.

According to figure. 15 method word appears the most hence the article tries to point that method of suicide is important and for that guns are used. And it is the most lethal method to commit suicide. Majority of men use guns as there weapons for suicide. According to various experts by changing policy can have a decrease in rate of suicide. Hence according to them if you change the method there is still a chance a life can be saved. But with access of guns it's the most lethal way for suicide. And here important factor are men as they use firearms more and in major cases weapon is firearm and gender is male as stated in our report.

Message for Suicide support people

“Are you thinking about killing yourself?”- I know this is not an easy question to ask but for prevention this question is must.

First thing the person with suicidal thoughts should is tell the reason or your internal feeling to your friends or family member. This will help in reduction of suicidal thoughts and may get relieve from thoughts. Keep yourself away from lethal weapons like guns.

You should immediately contact National Suicide Prevention Lifeline number **(1-800-273-TALK)** and the Crisis Text Line **(741741)**

Message for people around them: Listen to their problems, try to restrict their access to lethal weapons, inform the respective mental health authorities immediately. If you are nearby try to contact 911.

Message 1: **“Virtual health care services, like telehealth, are available for those looking to access behavioral health care”**

Message 2: **“Social connectedness is a key, and all Americans can play a role in supporting others”**

Message 3: **“Help is available for those who maybe experiencing a mental health or suicidal crisis.”**

Message 4: **“Acting now can help to prevent negative mental health impacts of pandemic”**

References

1. [mailto: https://github.com/fivethirtyeight/guns-data/blob/master/full_data.csv](mailto:https://github.com/fivethirtyeight/guns-data/blob/master/full_data.csv)
2. <https://www.washingtonpost.com/graphics/business/wonkblog/suicide-rates/>
3. <https://www.nimh.nih.gov/health/publications/5-action-steps-for-helping-someone-in-emotional-pain>
4. <https://www.kaggle.com/szamil/who-suicide-statistics>
5. https://theactionalliance.org/covid-19/messaging/key_messages