

# Final Term Project

Informatics in Engineering Management

EM 624

Fall 2021

**Pratik Shinde**  
**[pshinde1@stevens.edu](mailto:pshinde1@stevens.edu)**

## Table of Contents

Abstract.....	4
Introduction .....	5
Research Question .....	5
Literature .....	6
About the Dataset.....	7
Data Description, Processing & Usage .....	7
Methodology .....	9
Results .....	10
Analysing the Intents of Deaths.....	10
Suicidal Trends.....	11
Suicides as per the Age Groups.....	11
Susceptibility of One Gender to Suicide Count .....	12
Educational Qualification & Suicidal Counts .....	13
Race & Suicides .....	14
Gender Specific Analysis for Years 2012, 2013, 2014 .....	15
Race Specific Analysis for Years 2012, 2013, 2014.....	16
Education Qualifications Specific Analysis for Years 2012, 2013, 2014.....	18
Critique .....	19
Suggestions.....	19
Conclusion.....	20

References .....	21
------------------	----

## Abstract

Suicide is a leading cause of mortality and a serious public health problem in the United States. The consequences of a suicide are felt by family, friends, and communities. Schools are vital in our communities for assisting children and families, as we learned across the country during the COVID-19 epidemic. While schools are expected to deliver instruction, they also give chances for adolescents to participate in physical exercise as well as academic, social, mental, and physical health services, all of which can assist reduce stress and guard against bad consequences.

Many variables can either enhance or decrease the risk of suicide. Suicide has been linked to several types of injury and violence. People who have been exposed to violence, such as child abuse, bullying, or sexual assault, are more likely to commit suicide. Suicidal thoughts and actions can be reduced by being linked to family and community support and having simple access to health care.

The analysis of the given dataset would not only give insights on the factors which need to be taken care of to keep the number of suicides as low as possible, but also can be used to enlighten people of the society, or that particular group which is more susceptible to this horrible life taking activity.

## Introduction

Suicide is defined as self-inflicted death with the purpose to die. A suicide attempt occurs when someone intentionally damages oneself with the purpose of ending their life but does not die as a result of their actions (Facts About Suicide, 2021).

Every year, 703,000 individuals commit suicide, with many more attempting to commit suicide. Every suicide is a tragedy that impacts whole families, towns, and nations, as well as the people who are left behind. Suicide affects people of all ages and was the world's fourth biggest cause of mortality among 15- 29-year-olds in 2019. Suicide is a worldwide problem that occurs in all parts of the world, not only in high-income countries. In fact, in 2019, low- and middle-income nations accounted for more than 77 percent of global suicides. Suicide is a severe public health issue, but it may be avoided with early, evidence-based, and frequently low-cost treatments. A comprehensive multisectoral suicide prevention plan is required for effective national responses (Organization, 2021).

## Research Question

This research will look at the data supplied to see if there are any trends or connections between characteristics like ethnicity, gender, education, or age and suicide events using guns in the United States between 2012 and 2014. The investigation should take into account the elements that influence the occurrences and strive to answer the following questions:

- Does age play a role in occurrence of suicides?
- Is one gender more prone to suicide than the other one?
- Is any combination of factors indicative of more suicides than other?

Suicide may be avoided, and everyone can help save lives and build healthy and strong people, families, and communities. Suicide prevention necessitates a multifaceted public health strategy. Suicidal thoughts and actions can be reduced by being linked to family and community support and having simple access to health care. The research questions are aimed at finding out the major factors responsible for the suicides that have taken place over the years from 2012 to 2014. The data from the csv file will help in analysing the factors & hence formulating some strategies to help prevent such horrible act from happening at the same time.

## **Literature**

Suicide is defined as the deliberate act of causing one's own death. Risk factors include mental disorders (such as depression, bipolar disorder, autism spectrum disorders, schizophrenia, personality disorders, and anxiety disorders), nihilistic beliefs, physical disorders (such as chronic fatigue syndrome), and substance use disorders (such as alcohol use disorder and benzodiazepine use and withdrawal). Some suicides are spontaneous acts resulting from stress (e.g., financial or academic challenges), relationship issues (e.g., breakups), or harassment/bullying. Those who have tried suicide before are at a higher risk of doing so again. Limiting access to suicide techniques such as weapons, pills, and poisons; treating mental problems and substance abuse; attentive media reporting on suicide; and improving economic conditions are all effective suicide prevention strategies. Despite the fact that crisis hotlines are ubiquitous, they have not been well researched (WHO, 2021).

Some people are impacted more than others. Lesbian, homosexual, and bisexual students, as well as female students, were shown to have higher levels of these sentiments. Roughly half of lesbian, gay, and bisexual students, as well as nearly a third of students unsure of their sexual identity, said they had seriously considered suicide, significantly higher than heterosexual students. In 2019, the number of black students who attempted suicide increased by nearly half (Prevention, 2021).

## About the Dataset

URL: [https://github.com/fivethirtyeight/guns-data/blob/master/full\\_data.csv](https://github.com/fivethirtyeight/guns-data/blob/master/full_data.csv)

The dataset can be found at the above given URL. It is a csv file which has details about deaths with intents from the year 2012 to 2014. The intent shows a number of factors such as homicide, suicide & accidental. The focus of this research is on the suicidal data and analyse it further with the given parameters of the csv.

Each row in the csv file gives info about someone's demise with answers to all the possible factors which could be resultant for this horrible thing happening to any person. The data has all kinds of info so first part of the process was to sort the required data for the research.

## Data Description, Processing & Usage

The vast dataset with 100,799 entries needed to be sorted and filtered out to get what was required for the analysis & apart from that the csv file has a total of 10 columns amongst which some were selected, this was done step by step as follows:

- The first thing to do was analysing what percentage of the total deaths were caused by suicide and what were the other factors responsible. The
- To analyse the suicidal data, we can just omit the data regarding demises with any other intent of death by applying a filter to the "intent" column, turning on the non-required values & then deleting the rows containing the same. Within less than a minute a dataset for just suicidal case would be ready to analyse for all the 3 years.
- To analyze if any specific gender is inclined towards the occurrence of suicides, the "sex" column was analyzed so as to form a pie chart with the distinctive bifurcation of total cases in the form of percentage of whole.
- The factor of age is one to be thought of & with such vast and precise data in the file it is the factor to look out for as it could give some much-needed inputs. To analyze the same, I created some age group range so as to quantify the data in a usable format and then plot the graphs. The age groups are as follows:
  - 0-24 years
  - 25-34 years
  - 35-44 years

- 45-54 years
  - 55-64 years
  - 65-74 years
  - 75-84 years
  - 85+ years
- To analyze correlation with any other factor I considered race at first and created a graph for suicides according to the race specified in the csv file. The races given in the file are:
  - Asian/ Pacific Islander
  - Native American/Native Alaskan
  - White
  - Black
  - Hispanic
- Educational qualification defines the quality of life and the pressure on an individual as it defines the type of career one is going to pursue and each has a different pressure on mind as per the requirements. The column “education” gives information about the educational qualification and the analysis could be performed from there. The qualifications from the file are:
  - BA+
  - Some college
  - HS/ GED
  - Less than HS
  - Unknown (some of the cases have unknown or undisclosed qualification)
- The next question is about the chronology of the suicides and for the same, I filtered out the data from the csv files into 3 different files for the years 2012, 2013 & 2014.



## Methodology

The whole process was carried out with the help of various libraries such as pandas, matplotlib & numpy. The steps are as follows:

1. Importing the libraries to be used in the analysis.
2. Then for each variable that was to be analyzed, I imported the variables through pandas data frame.

For example:

```
df = pd.read_csv('All1.csv')
df['sex']
```

This command line would read in the column of 'sex' for the analysis related to gender. Similarly, I read all the column when needed.

3. Defining functions for index and counters. I defined a total of 7 functions to identify the index of any item from the list defined as counter for that particular variable. The other function was aimed at counting values into that list of counters using a for loop.

One example for education counter is as follows:

```
def education_level(parts):
    """
    - The function gets the index value for each educational type so as to append to
    the list of education_counter

    - The file had some variables named NA which means no data, NA cant be used at
    gives a null value so the variable
    has been changed to "Unknown".
    """
    if parts == 'BA+':
        return 0
    if parts == 'Some college':
        return 1
    if parts == 'HS/GED':
        return 2
    if parts == 'Unknown':
        return 3
    if parts == 'Less than HS':
        return 4

# function for appending to education counter
def education_counter(dff, counter):
    """
    - The function iterates through the 'education' column from the specified csv
    file in the form of 'df'
    - The counter is updated on detecting the educational level denoted in the loop
    """
    for row in dff['education']:
        if row == 'BA+':
            index = education_level(row)
            counter[index] += 1
        elif row == 'Some college':
            index = education_level(row)
            counter[index] += 1
        elif row == 'HS/GED':
```

```

        index = education_level(row)
        counter[index] += 1
    elif row == 'Unknown':
        index = education_level(row)
        counter[index] += 1
    elif row == 'Less than HS':
        index = education_level(row)
        counter[index] += 1

```

The above type of function was used in education counters, intent counters, race counters & gender counters. Other functions defined are for plotting pie charts, bar charts.

4. The bar graphs with multiple input are the part of the code where I faced problems. As there were more than 1 bar graph to be plotted in one single graph with multiple inputs from the csv it was a very confusing part of the analysis.

## Results

### Analysing the Intents of Deaths

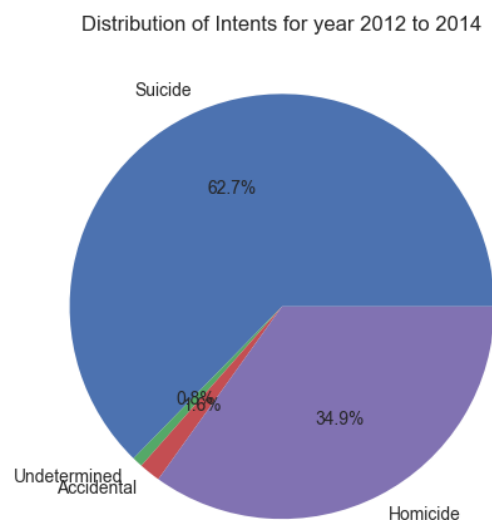


Figure 0.1 Intent of Demises

The graph analyses the cause of deaths and this is where the research begins into analysing the cause of suicides and further analysing its factors as the data depicts 62.7% of the demises in the period of 2012-2014 are caused due to suicides. That is a high percentage and reflects the amount of mental tension & strain in the society, hence it is necessary to identify the factors and fix the issue with possible measures.

## Suicidal Trends

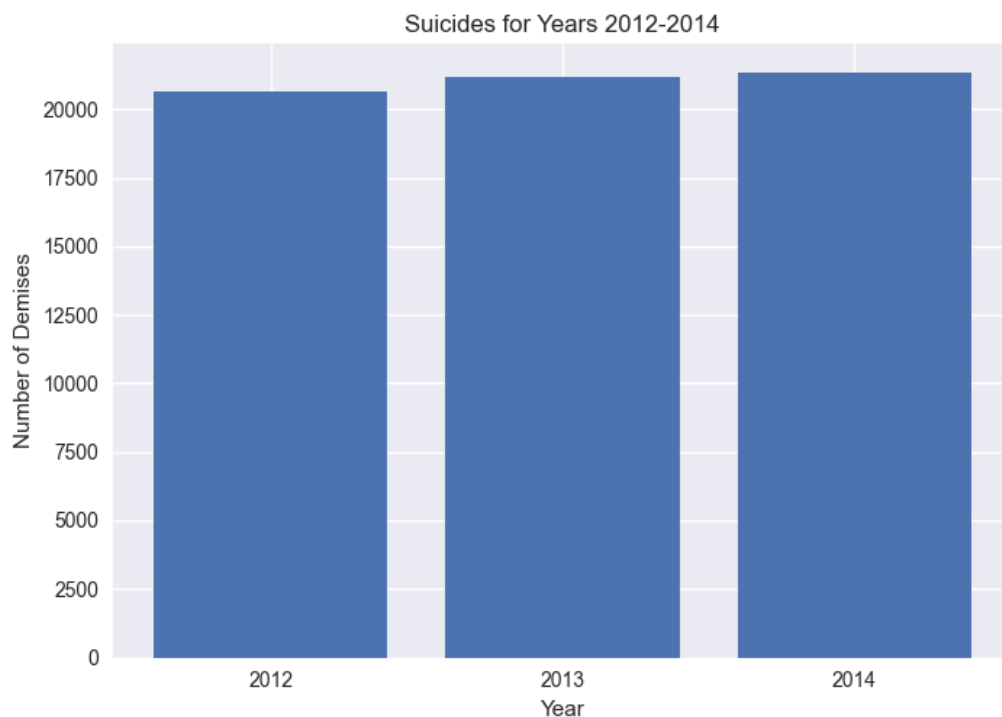


Figure 0.2 Suicides from 2012 to 2014

The number of suicides is increasing since 2012 and there is no drop in the cases. This is not a good sign for the society.

The total number of suicides that took place from the year 2012 to 2014 are: 63175

The count of suicides for year 2012 is

Suicide 20666

The count of suicides for year 2013 is

Suicide 21175

The count of suicides for year 2014 is

Suicide 21334

## Suicides as per the Age Groups

The age bracket '45-54' & '55-64' are the ones with the highest number of suicides. The age group of 25-65 are the ones with bars touching the heights. The corporate

environment, as well as the strain to earn a living, is the primary cause of suicide among this age bracket, which is America's working-class.

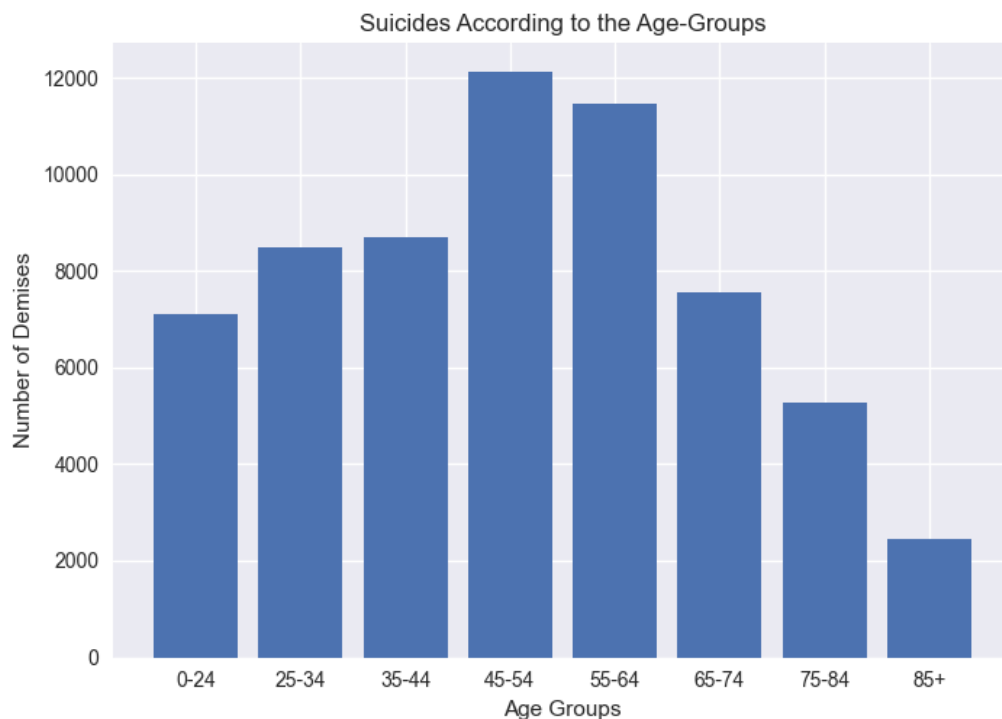


Figure 0.3 Suicides as per the Age Groups

There are 7 suicide cases with unknown ages.

### Susceptibility of One Gender to Suicide Count

MALES. Yes, the males are most prone to suicide as & when compared to the females in this particular dataset and time frame.

86.25 % males committed suicide from the year 2012 to 2014

13.75 % females committed suicide from the year 2012 to 2014

When medical practitioners and public health professionals try to prevent and treat suicidal ideation in their patients, gender differences must be acknowledged, and methods for dealing with risk factors for suicidal ideation in men and females must be developed. In other words, guys who have had a terrible life event or who have a lower level of life satisfaction should be given greater attention.

Gender Distribution of Suicides

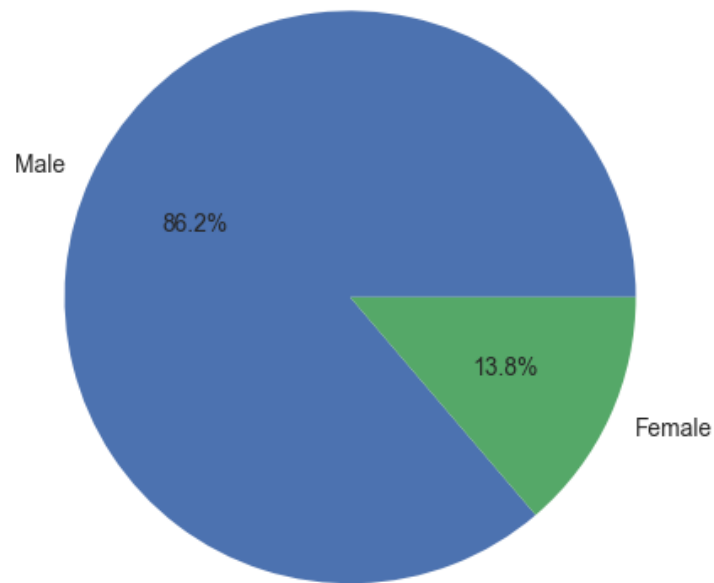


Figure 0.5 Gender Susceptibility to Suicide

### Educational Qualification & Suicidal Counts

Suicides as per the Educational Backgrounds

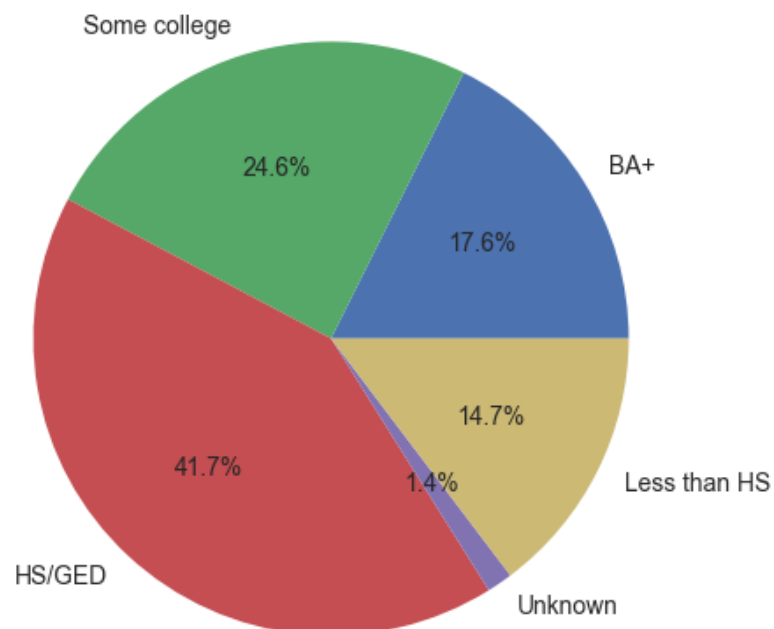


Figure 0.4 Suicides & Educational Qualifications

The count of suicides per education class is

HS/GED	26321
Some college	15533
BA+	11147
Less than HS	9292
Unknown	882

Adolescence is a period in which young individuals can have a good start in life. Adolescents' reports of poor mental health are on the rise. Connecting with adolescents and forming deep ties can help to safeguard their mental health. Schools and parents can help kids develop these protective bonds and grow into healthy adults. The suicide rate is highest amongst the high school students as per the dataset with a count of 26,321 (41.7%).

College diploma students & BA+ category is the highest after high school students with 24.6% & 17.6% of the total suicides. When we compare this with the age group dataset the 0-24 age group which includes both of these categories is one of the highest in terms of number of suicides.

### Race & Suicides

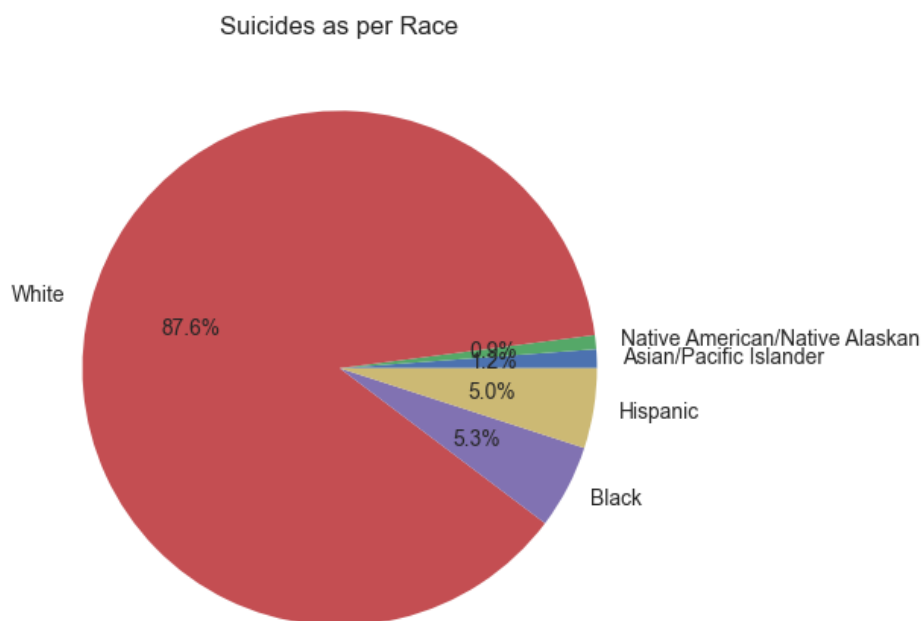


Figure 0.6 Suicides as per Racial Classes

According to recent research, racial prejudice is so painful that it is connected to the ability to commit suicide, which is thought to be a pre-requisite for taking one's own life. The capacity to emotionally and mentally reinterpret a violation, on the other hand, can help to lessen its negative consequences (Fickman, 2020).

The data gives a clear indication of how 87.6% of the Whites are susceptible to suicides and the Hispanic & Black account for 10.3% of the total count.

The count of suicides per racial class is

White	55372
Black	3332
Hispanic	3171
Asian/Pacific Islander	745
Native American/Native Alaskan	555

### Gender Specific Analysis for Years 2012, 2013, 2014

86.66 % males committed suicide from the year 2012 to 2014

13.34 % females committed suicide from the year 2012 to 2014

86.14 % males committed suicide from the year 2012 to 2014

13.86 % females committed suicide from the year 2012 to 2014

85.94 % males committed suicide from the year 2012 to 2014

14.06 % females committed suicide from the year 2012 to 2014

The suicide rate amongst males has been constant around 86%. The percentage is coming slightly down year by year from 2012 to 2014 in the males. At the same point females see an increase in the cases.

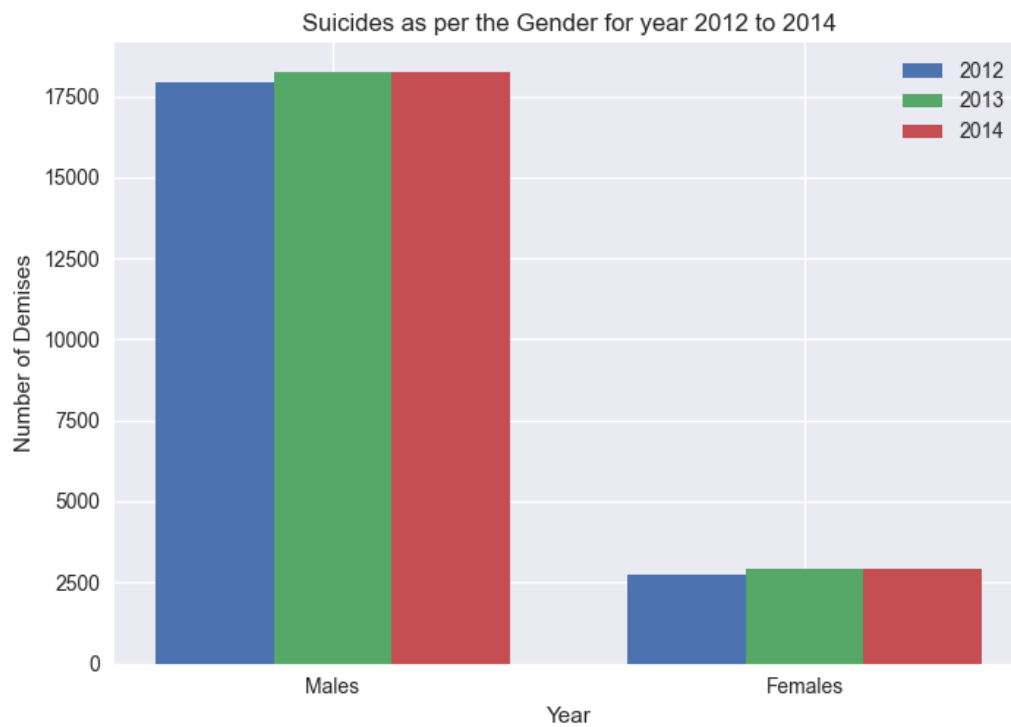


Figure 0.8 Suicides as per the Gender (2012-2014)

### Race Specific Analysis for Years 2012, 2013, 2014

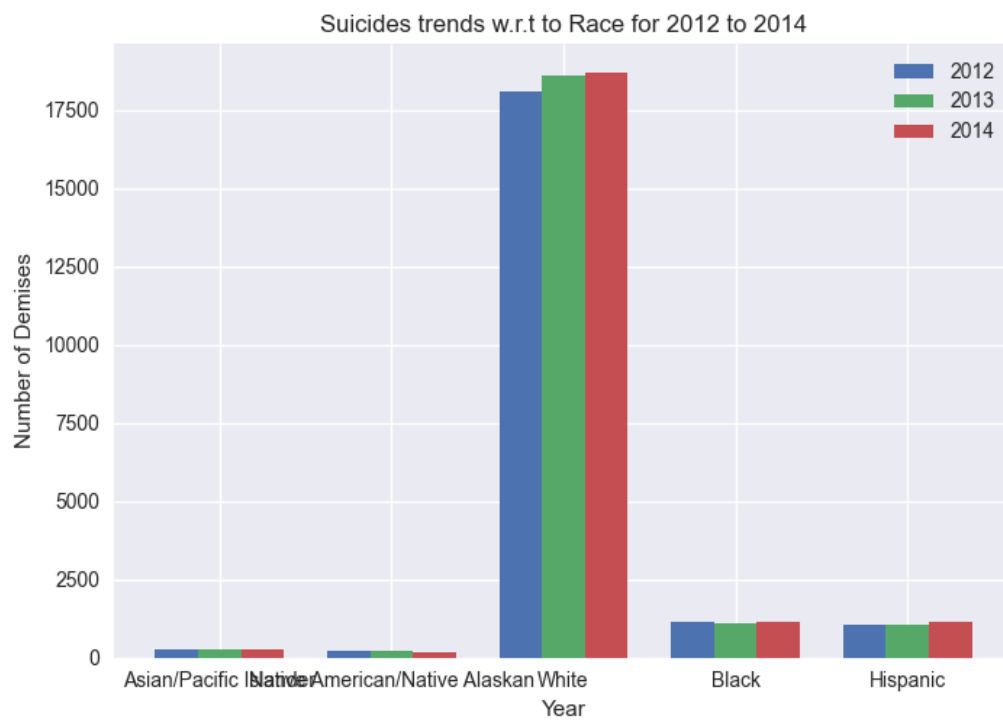


Figure 0.7 Race & Suicides (2012, 2013, 2014)



The count of suicides per racial class in year 2012 is

White	18085
Black	1128
Hispanic	1023
Asian/Pacific Islander	242
Native American/Native Alaskan	188

-----

The count of suicides per racial class in year 2013 is

White	18610
Black	1075
Hispanic	1034
Asian/Pacific Islander	267
Native American/Native Alaskan	189

-----

The count of suicides per racial class in year 2014 is

White	18677
Black	1129
Hispanic	1114
Asian/Pacific Islander	236
Native American/Native Alaskan	178

The study data indicates a higher percentage of deaths amongst the Whites in the given time period. Whereas the Native Americans are the one least susceptible to suicides.

Looking at the racism scenario around the world these results are quite tragic and a little bit unbelievable. As everyone is aware of the wide spread protests of “black lives matte” and the reason behind it, the data doesn’t really relate to the real-life situations. This creates a question on the credibility of the data.

## Education Qualifications Specific Analysis for Years 2012, 2013, 2014

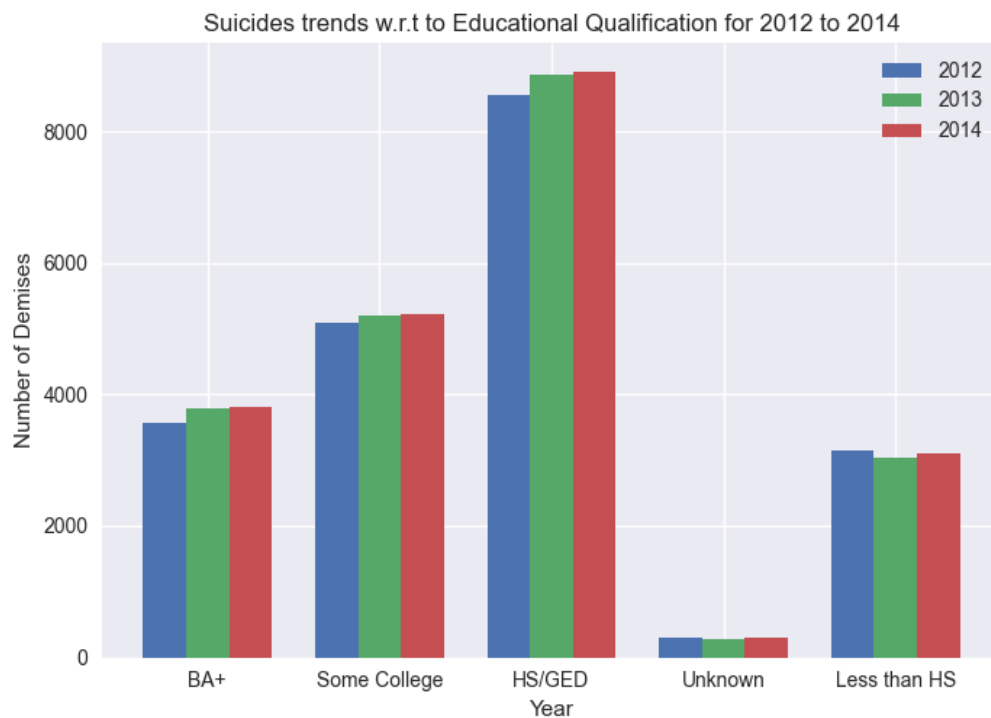


Figure 0.9 Educational Qualifications & Suicides (2012, 2013, 2014)

Supporting the overall educational qualification data, the trends when looked at yearly show the same major place of high schools as a problem for the society. The weak willpower of the kids in the society needs to be changed by counselling and other such events which helps in growth of a child.

## Critique

The csv file has a vast data regarding the demises in various circumstances. But at some point, the results don't match other standard reports and research papers which show a different trend.

The race factor is the one which gives shocking results. According to Frickman of University of Houston, suicide is the third highest cause of death among Black young adults aged 15 to 24, with over 3,000 Black Americans dying by suicide each year. But the facts in the research of this dataset are stated otherwise, showcasing Whites as the highest susceptible class to suicides.

In predicting suicide capacity in Black individuals, felt discrimination accounted for statistically significant variation above and beyond both emotions of sadness and non-discriminatory stresses. Perceived prejudice was not particularly connected with the ability to commit suicide in white people (Fickman, 2020).

Walker is a major researcher in the United States focusing in culture, race, mental health, and suicide, and is the author of the newly released "The Unapologetic Guide to Black Mental Health". When the research statistics of a top researcher give the indications of racial discrimination being a major cause for higher suicides in the black community, how can the dataset hence retrieved from the provided link be found credible.

## Suggestions

In order to prevent suicide, everyone has a responsibility to play. For example, religion groups may seek to prevent suicide by cultivating life-preserving cultures and norms, offering perspective and social support to community members, and assisting individuals in navigating life's challenges to discover a lasting sense of hope, meaning, and purpose.

Mental Healthcare centres should be setup in every neighbourhood or the already present hospitals should carry out campaigns and programs to create awareness about mental health because even if a person is struggling, sometimes he doesn't know.

A platform on the technological base can be another solution to tackle the problem. Such platforms could be used for surveying and analysing real time data not only for kids but also working professionals.

## Conclusion

The analysis clearly suggests a relation between **age, gender, education, race & suicides**. One gender, a specific age group & one particular race being more susceptible to the cause of this horrible thing like suicide to a person. So, some policies targeting these particular groups could be used to tackle the growing problem of suicides in the society.

Despite the fact that past suicide attempts are one of the most powerful risk factors for suicide, the great majority of persons who try suicide—nine out of ten—do not die by suicide. A rising number of people who have had suicide thoughts are writing and speaking about it, connecting with one another, and sharing their paths to healing and recovery. These kinds of activities help reduce Suicides to a greater extent and should be promoted.

Data Analysis as a tool helps identify the major factors, people affecting and different groups who need help with such serious conditions. This practice should take place on a bigger and a funded platform. Because what good is all this technology if we don't preserve the lives of those who make the world go round?

## References

- Facts About Suicide*. (2021, 11 28). Retrieved from Central for Disease Control & Prevention: <https://www.cdc.gov/suicide/facts/index.html>
- Fickman, L. (2020, August 3). *Racial Discrimination Linked to Suicide*. Retrieved from University of Houston: <https://uh.edu/news-events/stories/august-2020/08032020-rheeda-walker-racism-and-suicide-african-americans.php>
- Organization, W. H. (2021, 11 28). *Suicide*. Retrieved from World Health Organization: <https://www.who.int/news-room/fact-sheets/detail/suicide>
- Prevention, C. f. (2021, 11 28). *Mental Health*. Retrieved from Centers for Disease Control and Prevention: <https://www.cdc.gov/healthyyouth/mental-health/index.htm>
- WHO. (2021, 11 28). *Suicide*. Retrieved from World Health Organization : <https://www.who.int/en/news-room/fact-sheets/detail/suicide>