Analysis of Suicide Risks: Utah State

Asutosh R. Tummuluri, Ramprakash Chilakala

University of South Florida

ABSTRACT

Annually, suicides accounts for more than 800K deaths globally; counteracting is an immediate health issue. Pinpointing the factors of suicides are difficult because of the complexity in reading the data available (which is usually in excel sheets). Even after performing so many statistic and mathematical operations, one can find it challenging to find the true facts because of the outliers, unusual groups, and difficulty in data cleaning.

We used various visualization techniques to analyze the suicide data in the most easy-to-read way and accurately. Our approach of dealing the data improves the way the viewers or the Health Care department understand the underlying risks of suicides just by looking at the visualizations without performing any mathematical operations.

This approach helps in finding the patterns in the suicides, reveal trends that are hidden in the large chunk of data and even helps to predict the suicides and take certain measures to avoid them.

Keywords: Suicides, challenging, risks, visualizations.

1 INTRODUCTION

More than 45,000 people die by suicides annually due to various factors and it is the 10th leading cause of death in the United States. While the risks of other top causes of death decreased, suicides increased by more than 30% from 1999 to 2016. Prevention and prediction of the suicides is an important issue. Identifying the suicide factors is still challenging due to the complexity of a suicide. For example, Suicide attempts, which happen at 15-25 times the rate of suicides is an imperfect factor to consider for prediction because less than 8% of the suicide attempts are turned out to be suicide deaths. Therefore, there is an urgent need to improve our understanding of risks.

We found out that combining public medical conditions (genetic and non-genetic), weapons they possess, gender and age factors comprise some main aspects of a suicide.

To understand and analyze this accurately, we used the suicide data of Utah from 1904 to 2016 taken from the Utah Population Database and the Utah Suicide Genetic Risk Study(USGRS) that exhibit a high incidence of suicide. The USGRS also has deep genealogical data from the past 200 years that lets us to identify suicide risks with extended familial risks. This study gives an opportunity to understand diagnostic, demographic and some genetic characteristics of these suicides and contrasting them with the suicides with no extended familial risk. This database include demographic information such as sex, age, birthdate, and death date. It also has clinical information with a list of more than 10 medical conditions like depression, anxiety, asthma, cardiovascular problems, etc. The data provided is in tables (.xlsx – excel format) which is called raw data, that will give a very little idea by looking at it. The prior solution is to do some mathematical operations and analyzing the data to get an idea, this solution can be understood by only a few people who has knowledge in mathematics and Microsoft Excel. So, our idea is to make the data tell a story by using some interactive and aesthetic visualizations which can be understood by many people, making it easy to solve the problems.

Finally, our data visualizations allowed us to determine demographic and clinical associations with suicide risks.

2 METHODS

A sample of 160 suicides were extracted from the Utah Population Database from 1904 to 2016 with their sex, age, race, medical conditions, and weapons they possess. The data has many null values and outliers. So, we started cleaning to make the most use of the available data. After successful cleaning of the data, we tried to think of the most common questions one can get after looking at the data, like “Which race has the most suicide rate”, “What is the most common medical condition for a suicide”, “Which age group has the most number of suicides” and “What are most common causes of death in a suicide”.

We used HTML to display our visualizations, CSS to add style and D3.js to import the cleaned data, perform operations on them to visualize the answers to the questions with Pie Chart, Bar chart, Stacked Bar chart, area chart, etc.

2.1 Pie Chart

A pie chart is a type of circular graph used to display data. All the sectors in the circle are directly proportional to the fraction of the total circle in every category. The whole pie represents a 100%, while the sectors represent different portions. We used Pie Chart to represent the suicides by race attribute. Each slice of the pie represent each race like (Asian, Black, Latino/Hispanic, White). This gives each race’s share in the total number of suicides. We used the below mathematical formula to get the percentage of suicides in each race:

$$\frac{Number\ of\ suicides\ in\ one\ race}{Total\ number\ of\ suicides} \times 100$$

2.2 Bar Chart and Stacked Bar Chart

A bar chart is used to display the data using bars. Each bar represent a particular category. The length/height of each bar is relative to the aggregation its whole category. We used bar chart to represent the male and female medical conditions with respect to the number of suicides associated with it.

A stacked bar chart is an extension to the standard bar chart. Here, each bar represents more than one categorical value. We used stacked bar chart to represent the number of suicides associated with their sex and race.

We can also use color combinations to distinguish different bars. Here, we used green to represent female and blue to represent male.

We also tried another type of visualization using the same bar chart technique but by using "Names of the category" with "Colors" instead of bars. Here, the size of the text changes according to the number of suicides associated with it. This gave us an opportunity to visualize the cause of death in the suicides in a most aesthetic way.

2.3 Area Chart

An area chart is a combination of line chart and bar chart used to show one or more values of a category over the advancement of another category. Here, we used area chart to represent the number of suicides associated with the person's age at which they committed suicide. The highs and lows of the curve represents the number of suicides associated with the age. This gives us an idea of - which age group prone to the most suicides.

3 Evaluation / Results

A sample of 160 suicides were extracted from the Utah Population Database from 1904 to 2016 with their sex, age, race, medical conditions, and weapons they possess. After cleaning the raw data and applying visualization techniques to the data, we found some interesting facts.

The most suicides are committed by Whites and next comes the Blacks. The least are the Asian. *(figure-1)*

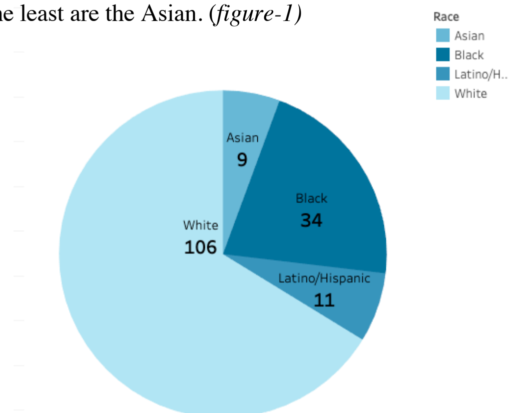

Figure 1: Suicides by Race - Pie chart.

The most common cause of a suicide death is the Gunshot to the head, which accounts 46% of the suicides. Then comes the Hanging and gunshot to the chest. *(figure-2)*

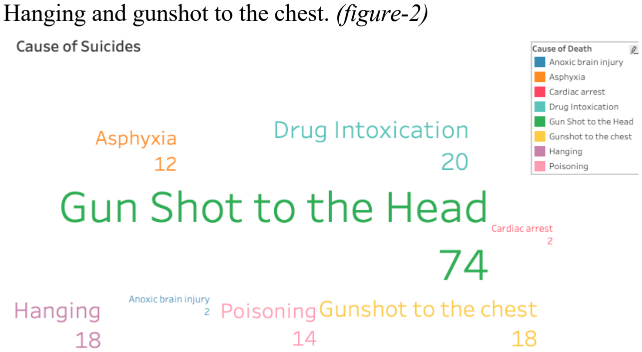

Figure 2: Cause of suicides - Text chart.

Among all the 15 medical conditions, depression is the most common medical condition in both male and female suicides. The least is the asthma, Impulsive control disorder and PD-Cluster C anxiety *(figure-3)*

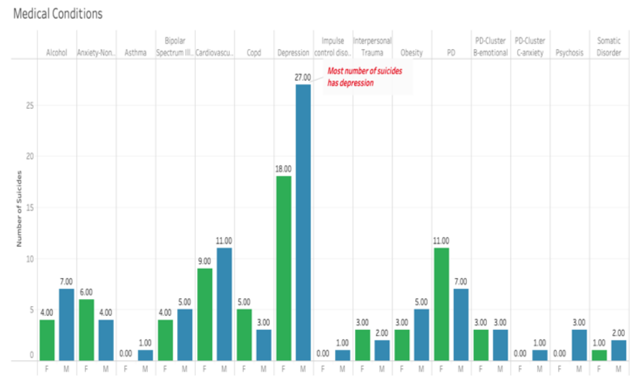

Figure 3: Suicides by Medical Conditions - Bar chart.

The highest number of suicides are committed by the people who are in between ages 17 to 32 which is 45% of the total suicides committed. *(figure-4)*

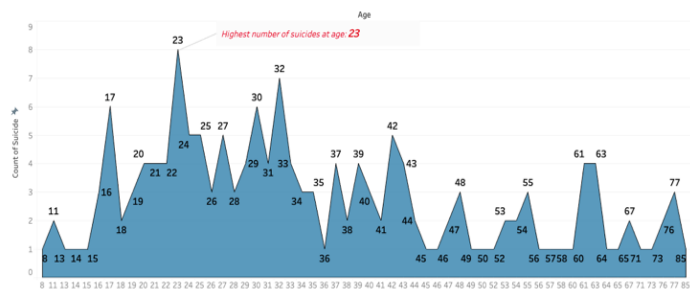

Figure 4: Suicides by Age - Area chart.

White men are prone to the most number of suicides and the least are the Asian women. *(figure-5)*

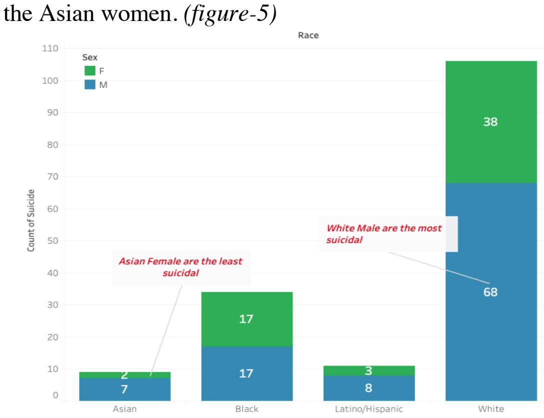

Figure 5: Suicides by Age - Area chart.

4 Conclusion

Understanding of people who are at high risk of suicide death is very important. Our results confirmed that age, sex, race, and weapons they possess can uniquely associate with the suicide deaths in the state of Utah. The visualization may show different conclusions if we take a large dataset. However, in future, one can design and train a model using machine learning techniques which can accurately predict the rate of suicide deaths.

References

[1] Hilary Coon. Extended familial risk of suicide death is associated with younger age at death and elevated polygenic risk of suicide.

[2] Carolina Nobre. Visualizing Multivariate Clinical Data