2012-2014
Gun Deaths in the United States

Erin Tsai

MSBA 305

Section C1

Table of Contents

I. Introduction

As seen in the past year, mass shootings continue to be a problem in the United States. After this past weekend, as of August 6, 2019, the United States has seen about 112 people killed in mass shootings in the past 216 days. That comes down to approximately one death every other day. (Padilla, 2019.)

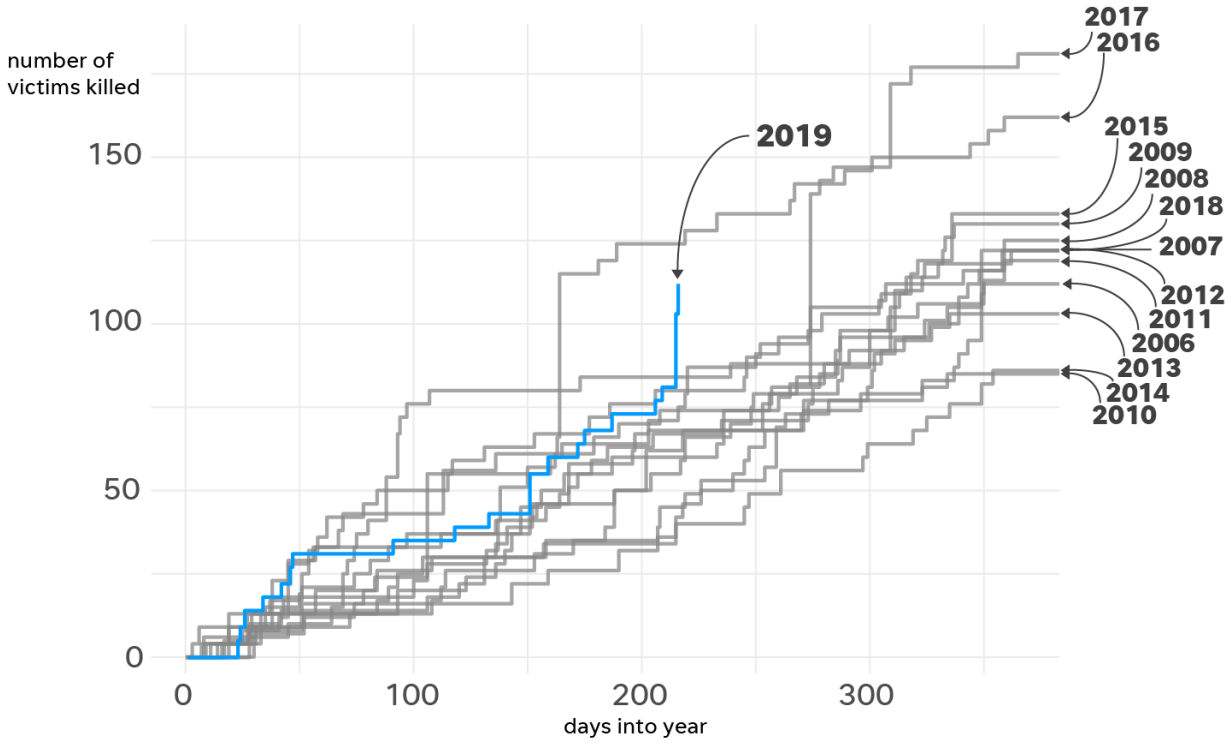


Table 1.1 - Chart on Mass Shooting Victims as of August 2019. (Padilla, 2019).

The number of mass killings in the United States has become a growing problem. However, mass killings are not the only problem in the United States. In general, gun deaths has become a severe problem in the United States. People in the United States live in constant fear that they will become the next victim prone to gun deaths.

Articles have made generalizations on race, ethnic, religious background and shootings, usually about the shooter. For instance, as Chang states, “Whenever a mass shooting receives

national attention, public discourse immediately turns to terrorism and mental health. Yet the overwhelming focus on these two explanations appears to ignore several significant facts. The vast majority of mass shootings in the United States are committed by men, regardless of national, ethnic, or religious background." (Chang, 2018).

It is interesting to note that although most articles discuss information about the shooter, we may also be interested to look into victim data. As Howard notes, "The findings presented by Feldman et al. show that, overall, individuals are at highest risk for being killed by a police officer in neighborhoods with the highest concentrations of racial/ethnic minorities. This finding is consistent with the broad literature on racial/ethnic health disparities, which typically shows similar health disparity gradients along racial/ethnic and socioeconomic lines." (Howard, 2019). However in recent years, there has been a growing focus on black men being targeted. McGroarty states: "According to the study, black men were the demographic most at risk of being killed during interactions with law enforcement, but people with mental illness across all racial groups faced heightened dangers when interacting with law enforcement. In fact, their study found that 23 percent of all people killed by police officers displayed evidence of mental illness. Overall, the risk of death as a result of police intervention was seven times greater for people with mental illness than for those without." (McGroarty, 2019).

Based on the growing concern, we are analyzing a data set on gun deaths data in the United States for 2012-2014 gun deaths. We are considering various characteristics to determine if they could potentially be contributing factors. As necessary, we will compare with U.S. Census data to determine if they are possibly contributing factors to gun deaths.

II. Data Set

The data set analyzed in this paper is from Kaggle. It is titled Gun Deaths in the US: 2012-2014. It originated from the Centers for Disease Control and Prevention (CDC). The data set has 101,000 rows and 11 columns. The data provides information about the victim's age, sex, race, education, the aggressor's intent, time (month and year), place of death, and whether or not the police were present at place of death.

This data set was chosen because of the ease of use. It was previously cleaned and has very minimal number of null values. Additionally, the data has been formatted for ease of use for analytics purposes. Moreover, because of the small number of columns, it allows for a more focused study.

III. Descriptive Analytics

We review the data based on the different categories to determine if there are possible contributing factors to victim deaths. Are there certain races that are more prone to being victims of gun death? Are there places that are more common than others? We consider education in the first instance.

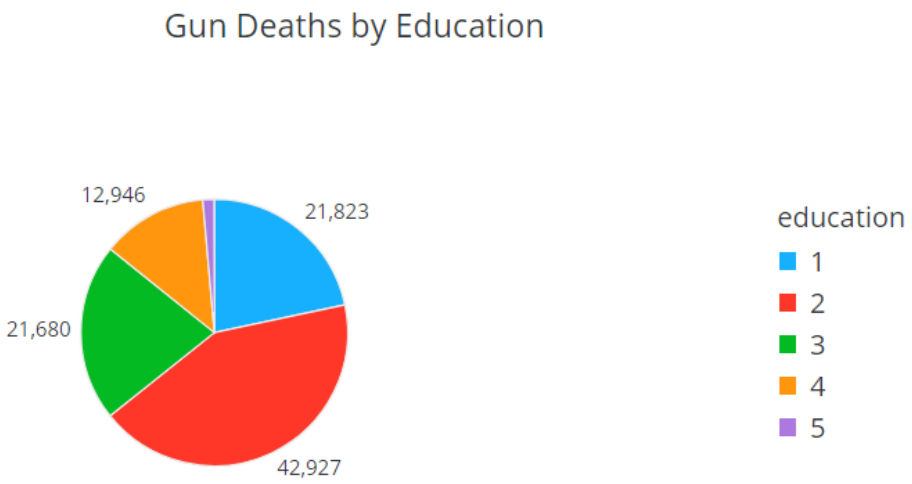


Table 3.1 2012-2014 Gun Deaths by Education

Education in this data set is labeled using the values 1-5 to represent as follows:

1: Less than High School

2: Graduated from High School or equivalent

3: Some College

4: At least graduated from College

5: Not available

Based on the pie chart above, the largest percentage of gun deaths are in level 2, which would indicate that the victims had graduated high school or equivalent. What we want consider is how this compares to the population. According to the U.S. Census Bureau for 2013, approximately 29.5% of the population above 18 years of age was a high school graduate. [U.S. Census Bureau, 2013] We compare to 2013, as 2012-2014 covers 2013. So 2013 would provide us with a good approximation of the data without needing to go into the specifics. This tells us that the breakdown of the gun deaths by education does not seem to directly correlate to the population. What this means is that we should consider whether or not education level plays a part into determining risk of gun deaths.

Secondly, we consider the data by race and sex.

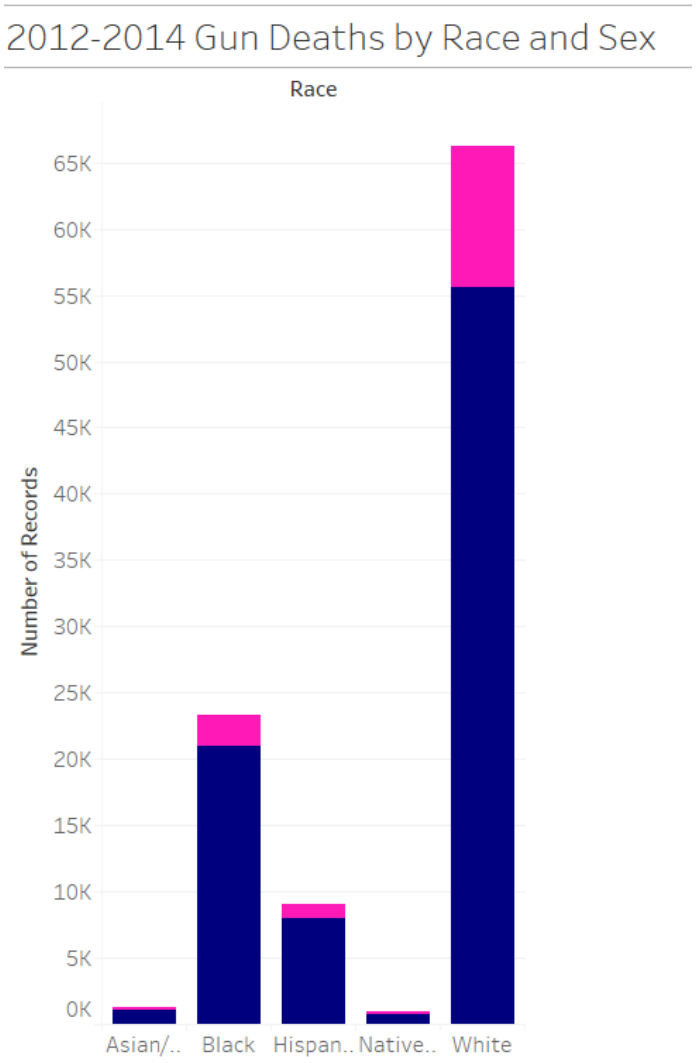


Table 3.2 2012-2014 Gun Deaths by Race and Sex [Blue for Male, Pink for Female]

We note distinctly, that the majority of gun deaths are against White victims. We can also see the distinctly large difference between number of men killed versus females killed. The difference is approximately the same throughout all races. There is no one race, where the discrepancy between the number of males killed and the number of females killed greatly differ.

Although it appears that the victims are mostly White, we must also consider what that means as a population as a whole. As a basic approximation, we consider the U.S. Census data

for 2013. The breakdown of races is: White, Black/African American, American Indian/Native, Asian, Pacific Islander, or Two or more races. Naturally, this will be hard to compare to our data, because the race breakdown is different. Because Hispanics are not separately categorized under the generalized fact we are considering at this time, the comparison is quite difficult. However, we can compare the number of Black/African American to the whole population as an initial comparison. In 2013, there were 41,623,897 Black/African Americans in the Resident Population, which makes up approximately 13% of the whole population in the United States. (U.S. Census Fact Finder, June 2014.) We distinctly note that in our data, the Black victims who died make up approximately a 23% of the data points. This is much higher than the actual population of Black residents. This seems to indicate that people who are racially Black are more often victims of gun deaths.

The U.S. Census determined that White people make up about 77.7% of the population in the United States. This includes the Hispanic population. This seems to be approximately in line with the percentage of gun deaths. Based on the data set, the white victim deaths make up approximately 74.5% of the deaths in the data set, when we combine the White race with Hispanic, as Hispanics are included in the U.S. Census Bureau values. To us, this does not seem to indicate that white persons are more prone to gun deaths, unlike the black persons who experience a higher percentage of gun deaths.

Next, we take a look at Age and Intent of the Gun Deaths.

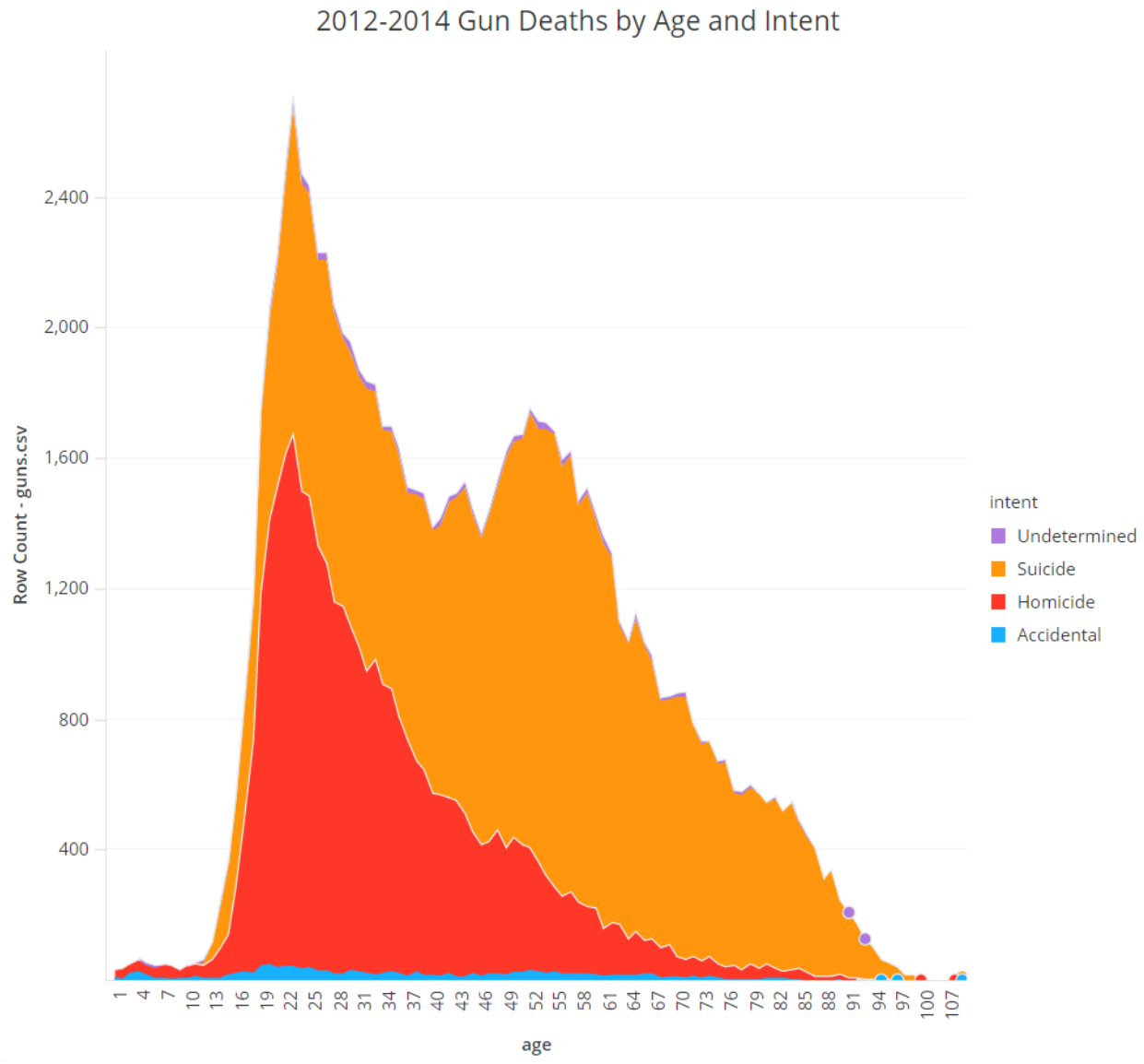


Table 3.3: 2012-2014 Gun Deaths by Age and Intent

As noted by the areas of the types of intent of gun deaths, Suicide seems to be the primary intent of the Gun Deaths. We note that Suicide numbers seem higher between the ages of 40 and 95 as opposed to the ages of 10-40. Between the ages of 10-40, there is a lot more homicides than suicides. Because of this, we want to consider the possibilities of why that is. Why are there less homicides as the victims are aged older? The older we get, are we less prone go gun deaths? Or, in the alternative, is it possible that older people are less likely to be killed by

homicide as opposed to suicide because the older we get, the more careful we are around guns and are more aware of the dangers of guns. These are questions we will consider in the future.

We note that as the age value is higher, the overall number of gun deaths is also far smaller. But, I believe that as a population, the greatest concern should be amongst the sharp peak in gun deaths in the ages of 20s and 30s. Why is this? Is there anything we can do to prevent this or minimize this? Unfortunately, this data is limited and cannot provide us much more information past the actual past information provided. However, this is something we should keep in mind in future analysis.

We further consider whether or not there is a trend over time based on the data provided between 2012-2014. Because of the minimum number of years in the data, we are cognizant of the value we can obtain from considering time trends. However, we will consider it, initially, to see if any information can be obtained.

A quick chart based on the 2012-2014 data is provided below in **Table 3.4**.

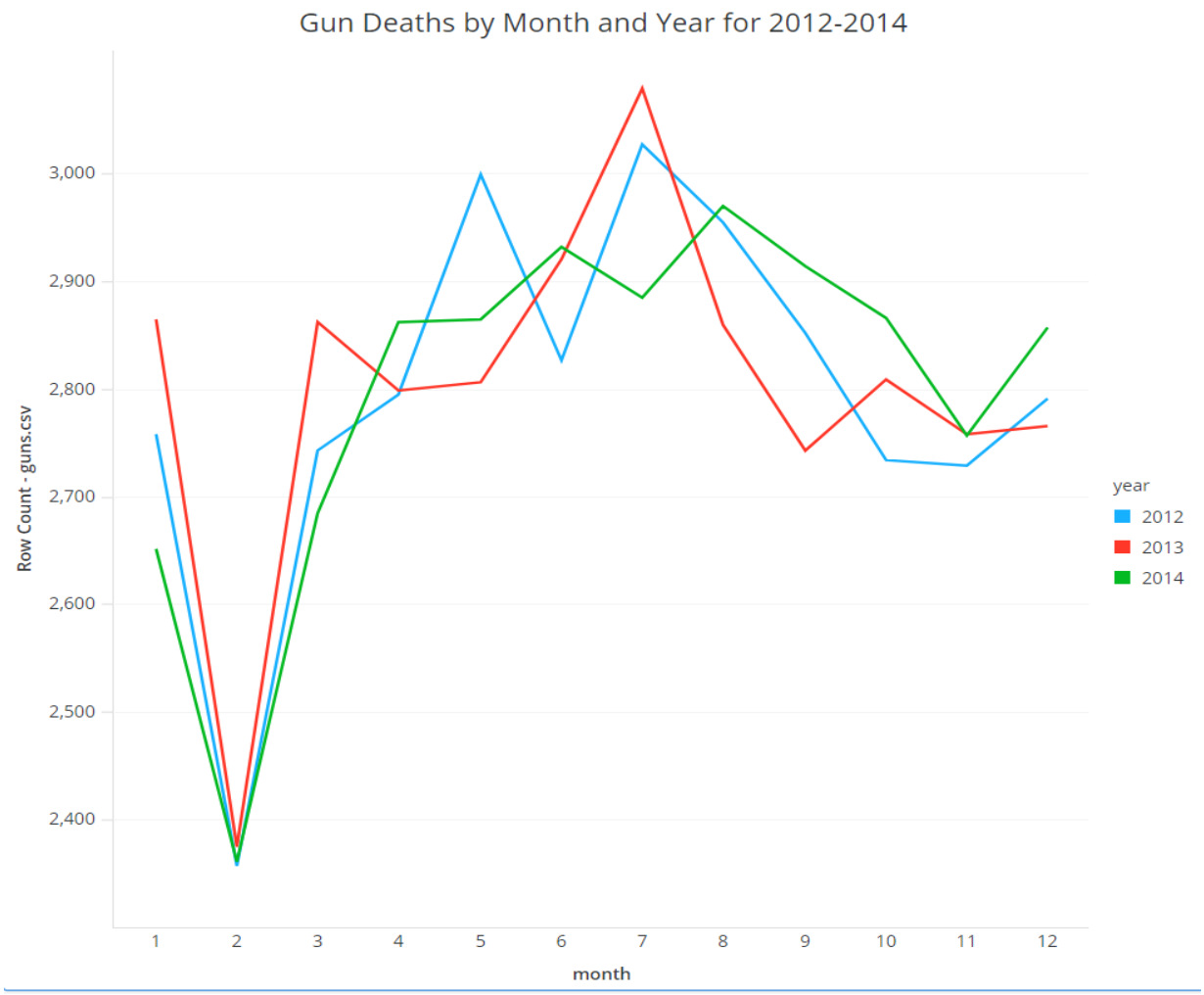


Table 3.4 – 2012-2014 Gun Deaths by Month and Year

In this chart, we analyze the trends of gun deaths in 2012-2014 by month. We note that the overall trend over the course of the year seems to follow the same pattern. We note a distinct dip of gun deaths every February. Although this dip is obvious, we also note that the y-axis has limited range, meaning that all these values seem fairly close to one another.

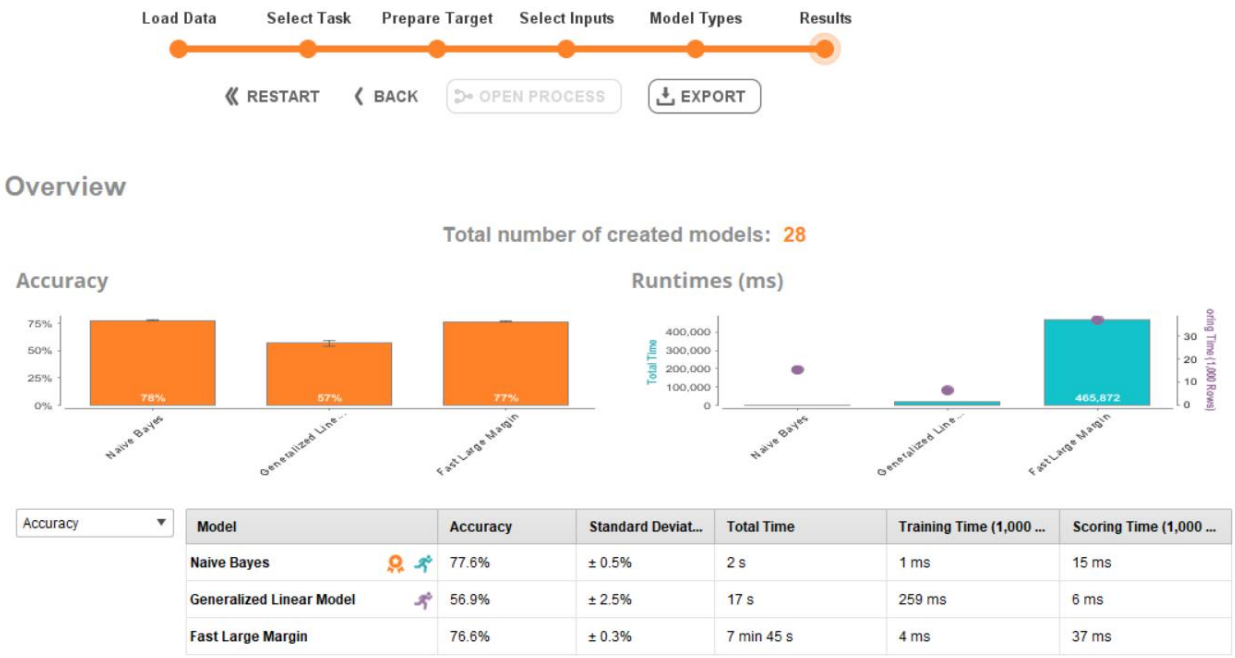
Moreover, we note that there is no clear trend over the years from 2012-2014. There does not seem to be a distinct increase or decrease over the years.

The descriptive analysis we have considered show interesting trends, that we will analyze to see if they continue to hold through predictive analytics and prescriptive analytics.

IV. Predictive Analytics

Next, we tackle predictive analytics. First, we consider predicting race of the victim. We noted in the earlier review of the Descriptive Analytics that although White persons are the majority race of victim deaths, compared to population data, Black persons seem to have an unusually high percentage of victim deaths. We want to consider if this can be predicted based on other characteristics of the gun death. We consider police presence, gender of victim, age of victim, place, education, and intent as characteristics affecting the data.

A quick run of the Naive Bayes Model, Generalized Linear Model, and Fast Large Margin model yield the following overview results:

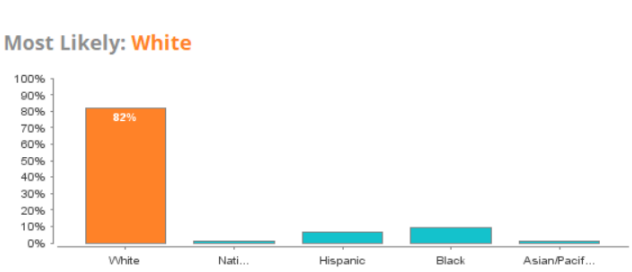


Model	Accuracy	Standard Deviat...	Total Time	Training Time (1,000 ...	Scoring Time (1,000 ...
Naive Bayes	77.6%	± 0.5%	2 s	1 ms	15 ms
Generalized Linear Model	56.9%	± 2.5%	17 s	259 ms	6 ms
Fast Large Margin	76.6%	± 0.3%	7 min 45 s	4 ms	37 ms

Table 4.1 – Overview of Auto Model from RapidMiner

We note that the Naive Bayes and Fast Large Margin models have the highest accuracies. For ease of use, we will focus on the Naive Bayes Model.

Utilizing the Naive Bayes Simulator, if we enter the following data: Age: 22; Education: 2.299; Intent: Suicide; Place: Home; Police: 0.014; Sex: M; we get the following prediction:



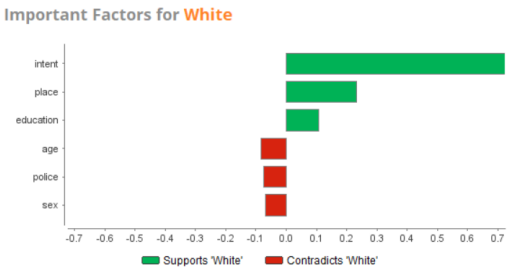


Table 4.2 Prediction based on Chosen Data

This agrees with our data above under Descriptive Analytics. We note that through this model, we get the following performance data:

accuracy: 77.56% +/- 0.51% (micro average: 77.56%)

	true Asian/Pacific ...	true White	true Native Ameri...	true Black	true Hispanic	class precision
pred. Asian/Pacifi...	0	0	0	0	1	0.00%
pred. White	258	17281	158	1574	1068	84.96%
pred. Native Amer...	0	0	0	0	0	0.00%
pred. Black	120	1514	73	4962	1410	61.42%
pred. Hispanic	6	180	8	93	93	24.47%
class recall	0.00%	91.07%	0.00%	74.85%	3.62%	

Table 4.3 Naive Bayes Model Performance for Race

Here, we see a distinct note that with respect to precision for predicting White persons, the model performs extremely well. However, with respect to Black, the precision falls drastically. We also note that with respect to Asians and Native Americans, we have no precision. We consider whether that is due to the fact that there are less data points, or if it is due to other factors, such as outside factors affecting the data.

We note, interestingly enough, that when we consider Weights by Correlation, we see a high Weight for Intent.

Weights by Correlation

Attribute	Weight
intent = Homicide	0.275
intent = Suicide	0.268
education	0.130
sex = M	0.038
police	0.023
intent = Undetermined	0.009
intent = Accidental	0.008

Table 4.4 Weights by Correlation for Race

In detail, we see this again in the Correlation Table below.

Correlations

Attribut...	educati...	intent = ...	intent = ...	intent = ...	intent = ...	police	race = ...	race = ...	race = ...	race = ...	sex = M
education	1	-0.018	-0.255	0.257	-0.008	-0.029	0.055	-0.195	-0.139	0.250	-0.076
intent = A...	-0.018	1	-0.095	-0.168	-0.012	-0.015	-0.011	-0.012	-0.001	0.012	0.004
intent = ...	-0.255	-0.095	1	-0.948	-0.066	0.163	0.017	0.563	0.178	-0.612	-0.020
intent = ...	0.257	-0.168	-0.948	1	-0.117	-0.154	-0.014	-0.549	-0.175	0.598	0.023
intent = ...	-0.008	-0.012	-0.066	-0.117	1	-0.011	0.001	-0.015	-0.000	0.011	-0.020
police	-0.029	-0.015	0.163	-0.154	-0.011	1	0.004	0.008	0.048	-0.039	0.032
race = A...	0.055	-0.011	0.017	-0.014	0.001	0.004	1	-0.064	-0.037	-0.164	-0.015
race = Bl...	-0.195	-0.012	0.563	-0.549	-0.015	0.008	-0.064	1	-0.171	-0.756	0.066
race = Hi...	-0.139	-0.001	0.178	-0.175	-0.000	0.048	-0.037	-0.171	1	-0.437	0.026
race = W...	0.250	0.012	-0.612	0.598	0.011	-0.039	-0.164	-0.756	-0.437	1	-0.070
sex = M	-0.076	0.004	-0.020	0.023	-0.020	0.032	-0.015	0.066	0.026	-0.070	1

Table 4.5 – Correlation for Race

The ones with large correlation values are for Intent. This is interesting to us, as we did not previously consider the relationship between intent and race, but rather, we had considered the relationship between intent and age.

As noted above, the Generalized Linear Model did not perform as well.

Generalized Linear Model - Performance

Criterion
accuracy
classification error

Table View Plot View

accuracy: 56.90% +/- 2.54% (micro average: 56.90%)

	true Asian/Pacific ...	true White	true Native Ameri...	true Black	true Hispanic	class precision
pred. Asian/Pacifi...	15	663	7	266	104	1.42%
pred. White	240	13594	149	3042	1339	74.03%
pred. Native Amer...	3	370	8	153	62	1.34%
pred. Black	85	2635	68	2476	785	40.93%
pred. Hispanic	38	1661	24	717	295	10.79%
class recall	3.94%	71.84%	3.12%	37.21%	11.41%	

Table 4.6 – Performance of Generalized Linear Model

However, we note that there is an actual value in precisions for all races, instead of only a few like we did in the Naive Bayes Model. This might be of relevance to us in applying predictions. As discussed before, this is not the focus of our study.

With respect to the Fast Large Margin Model, we can see that with respect to Black victim deaths versus all others, we note the following:

Fast Large Margin - Model

Binary Models

- Asian/Pacific Islander vs. all other
- White vs. all other
- Native American/Native Alaskan vs. all other
- Black vs. all other
- Hispanic vs. all other

```
  0.040 * education
+ 0.241 * intent = Accidental
- 0.514 * intent = Homicide
+ 0.517 * intent = Suicide
+ 0.256 * intent = Undetermined
+ 0.623 * police
- 0.253 * sex = M
+ 0.512
```

Table 4.7 – Fast Large Margin Model

It appears that intent is a controlling factor in the data, in addition to police. And although the accuracy is better than the Generalized Linear Model, it has no precision for other races besides White and Black.

Fast Large Margin - Performance

Criterion: accuracy / classification error

Table View / Plot View

accuracy: 76.55% +/- 0.35% (micro average: 76.55%)

	true Asian/Pacific ...	true White	true Native Ameri...	true Black	true Hispanic	class precision
pred. Asian/Pacifi...	0	0	0	0	0	0.00%
pred. White	238	16670	155	1252	1036	86.15%
pred. Native Amer...	0	0	0	0	0	0.00%
pred. Black	146	2305	84	5377	1536	56.91%
pred. Hispanic	0	0	0	0	0	0.00%
class recall	0.00%	87.85%	0.00%	81.11%	0.00%	

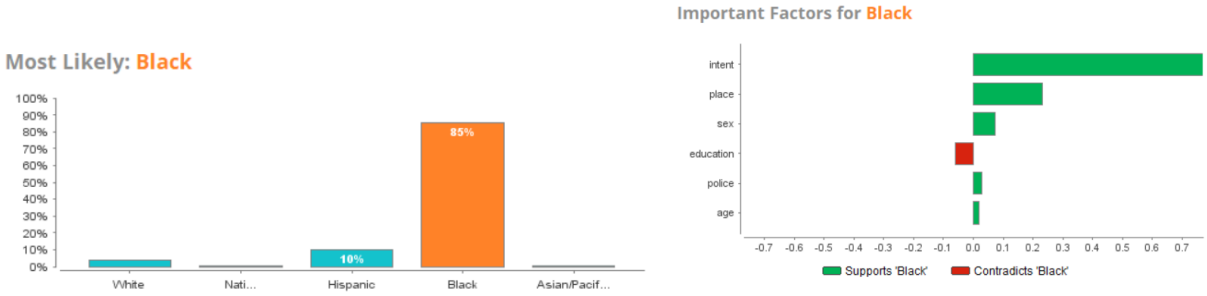
Table 4.8 – Fast Large Margin – Performance

Therefore, we must be cognizant of the Performance in utilizing these models. For instance, we would not use these models to make predictions about the race of the victim being Asian, as that would give us little or no information based on the Performance.

V. Prescriptive Analytics

Through prescriptive analytics, we are attempting to seek the best course of action. In our case, we are considering how we can optimize parameters to make the best predictions.

Using RapidMinder, we attempted to Optimize the Naive Bayes Model for Black race. The run time is fairly quick. When running the Naive Bayes Simulator for the following data: Age: 37; Education: 1.786; Intent: Homicide; Place: Street; Police: 0.000; Sex: M, the Simulator predicted the following:



This also agrees with the data in our Descriptive Analytics. Based on the accuracy of this Model, we would be fairly confident in predicting the race of the victim. Additionally, we note how relevant the factors are based on the data.

It appears based on our Predictive Analytics that Intent is an important factors for determining Race of victim.

Because we previously did not consider the relationship of race along with intent together, this is something to consider for future analysis. It would be interesting to note whether there are any particular connections we can make as far as making predictions to minimize harm in the future based on these connections.

Further review into the Optimized Simulation, we note that if simulate where police presence is detected at time of death, the predicted race becomes Hispanic a majority of the time. This is of particular interest to us, as we noted in the Introduction that media has been portraying black men to be those who are subject to death most often in police presence. This is interesting to consider, because we wonder if there is further analysis that we can conduct to determine a better understanding of why this happens in the Simulation.

VI. Conclusion

Based on this study, we have learned a few interesting things about our data from 2012-2014. We have learned that our data seems to indicate that a lot of the gun deaths experienced in the United States are due to Suicide. However, with different age groups, the trend is different. For younger people, the majority of gun deaths experienced is homicide instead of suicide.

We have also learned that when comparing the race of victims gun deaths to race of the population in the United States, there seems to be a rather large number of black victims. And in analyzing the race of the victims of gun deaths, we note that there is a large correlation with intent. This leads us question why intent is so closely linked with the race of the victims. With more data over time, and more characteristics considered, we can identify a lot of relevant factors.

Based on this study alone, we are able to gain the knowledge of knowing how these various factors analyzed here, including education, age, intent, and race, have an affect on predicting the race of the victim.

Because of the current situation in the United States, with the large numbers of mass shootings happening, it would be interesting to consider the victim data versus the shooter data in

the next study. Because victim data is not every often analyzed in the current media, this study has provided us with information that we can use in future analysis and applications.

VII. References

Chang, S. (2018). Our National Psychosis: Guns, Terror, and Hegemonic Masculinity. *Harvard Civil Rights-Civil Liberties Law Review*, *53*(2), 495–532. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=a9h&AN=133547552&site=ehost-live&scope=site

Howard, J. T. (2019). To Protect and Serve? New Data on Police-Related Deaths Reveal a Persistent American Dilemma. *American Journal of Public Health*, *109*(3), 349–350. https://doi.org/10.2105/AJPH.2018.304936

McGroarty, M. (2019). Mental health crises significant factor in police shootings.*Chicago Policy Review (Online)*, Retrieved from https://search.proquest.com/docview/2220771713?accountid=25283

Padilla, R., Petras, G., Loehrke, J., Thorson, M., Gelles, K., & Sullivan, S. (2019, Aug. 6). *After El Paso and Dayton*, 112 killed in mass shootings over 216 days. Retrieved from: https://www.usatoday.com/in-depth/news/2019/08/06/el-paso-texas-dayton-ohio-shooting-victims-mass-killings-by-numbers/1922454001/

U.S. Census Bureau (2013), *Table 1 of Education Attainment in the United States: 2013*. Retrieved from: https://www.census.gov/data/tables/2013/demo/educational-attainment/cps-detailed-tables.html

U.S. Census Bureau Fact Finder (June 2014), *Annual Estimates of Resident Population by Sex, Race, and Hispanic Origin for the United States, States, and Counties, April 1, 2010 to July 1, 2013*. Retrieved from: https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2013/PEPSR6H?slice=Hisp~tothisp!Year~est72013