Final Project Report

2024-05-01

Group Members

- Nur Arsani
- Isaac Cameron
- Evgenii Ovchinnikov
- Amanda Rosenzweig

Introduction

> Gun violence remains a pressing issue in the United States as it affects communities nationwide. We decided to narrow our focus to the factors that contribute to the lethality in gun violence incidents as it is crucial for informing policy interventions and preventative measures. We will delve into the statistical predictors of lethality in gun violence incidents, examining a factor such as gun characteristics with fatality rates. Additionally, we will explore specific questions regarding the types of guns involved and their relationship with casualty rates, as well as the impact of the status of guns (stolen vs not-stolen) on casualty rates for all states. We will answer these questions by focusing on the type of guns used by categorizing them into long and short guns identified in our dataset and we will employ hypothesis testing and compare two-sample means which will determine if there is a statistically significant difference in the impact of these gun types on casualties. We will also focus on the status of the gun which is stolen or not stolen for all states and employ hypothesis testing to see if there is a significance difference on the impact of the status of guns and rate of casualties. We use t-tests and create graphs to support our claim that specific gun characteristics such as a long barrel and not-stolen increase the fatality rate in gun violence incidents. In other words, guns with long barrels tend to cause statistically higher casualty rates in gun violence incidents compared to short-barrel guns. Additionally, not stolen guns cause statistically higher casualty rates than guns that were reported as stolen.

Background

> This data was collected by the Gun Violence Archives, which is an independent organization focused on the conversion of gun violence in America. Since 2013, they have been collecting data on gun violence through third party sources & crowdsourcing to provide accessible data to anyone looking to educate themselves on the matter. A man named James Ko then took this data and created csv files that

> are better organized and accessible to study for data scientists like ourselves. We combined twelve datasets for each month of 2017. We decided to use the year 2017 dataset because it is the most complete dataset which allows us to analyze it better. Our goal for the rest of the report is to explore the statistical predictors of lethality in gun violence by focusing on factors such as the types of gun (short and long) and the status of the gun (stolen and not-stolen) for all states. We will create visualizations such as bar graphs to answer our questions.

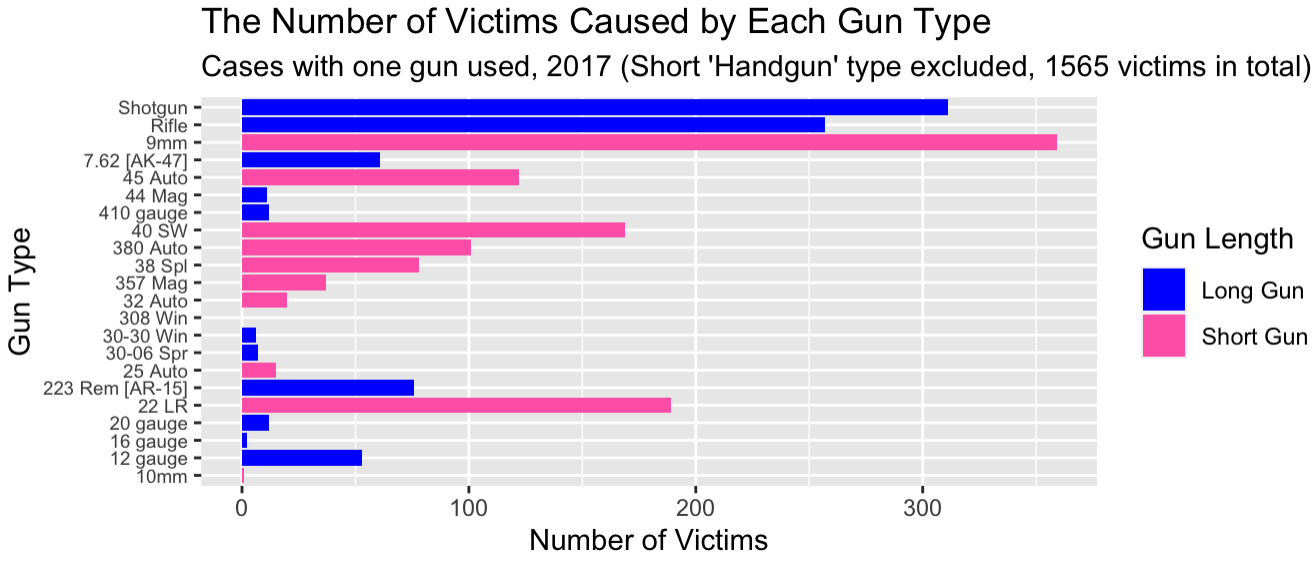
- **The key variables in the final dataset**:
  - `n_killed` *(Number of People Killed)*:
    - Measures: The count of fatalities in each incident.
    - Relation to Key Question: With this variable, we will be able to analyze and conduct statistical inference on the mean number of people killed per incident.
    - Statistics: The number of people killed in all incidents in 2017 in the USA varies from 0 to 27. The average number of people killed per incident is 0.253.
  - `n_injured` *(Number of People Injured)*:
    - Measures: The count of injured people in each incident.
    - Relation to Key Question: With this variable, we will be able to analyze and conduct statistical inference on the mean number of people injured per incident.
    - Statistics: The number of people injured in all incidentcs in 2017 in the US varies from 0 to 25. The average number of people per incident is 0.501.
  - `n_victims` *(Number of People Injured + Number of People Killed)*:
    - Measures: A combined sum of people hit in gun violence cases.
    - Relation to Key Question: We have decided that it would be interesting to explore not only the counts of killed and injured people separately, but the combined number of victims as well. This variable plays exactly this role.
    - Statistics: The total number of injured and killed people in all incidents in 2017 in the US varies from 0 to 47. The average number of victims per incident is 0.754.
  - `gun_stolen` *(Status of Guns)*:
    - Categorizes: Whether the guns involved were reported as stolen.
    - Relation to Key Question: This variable can be used to examine if incidents involving not stolen guns are more likely to be lethal.
    - Note on Variable: For the simplicity of our analysis, we have excluded cases where the status of gun was unknown. Thus, we are working with two statuses: Stolen and Not-Stolen.
  - `gun_type` *(Type of Firearms)*:
    - Measures: The type of guns used in the incident.
    - Relation to Key Question: Different types of firearms might be associated with varying levels of lethality due to their specific capabilities.
    - Note on Variable: There are 23 different types of guns. However, we categorized them into long barrel guns (12) and short barrel guns (11).
    - Thus, further, we introduce new variable `gun_length`, which categorizes guns into “Long Gun” and “Short Gun” types. After categorizing them this way, we can analyze and conduct inference on both length types.
  - `n_guns_involved` *(Number of Guns)*:
    - Measures: The count of guns involved in each incident.
    - In our analysis, we only use cases with 1 gun involved for simplicity purposes. This way, it is easier to identify what specific gun (its length and status) was involved in an incident and caused some specific number of injuries / deaths, which helps us significantly in analysis.

- **Data Source**:
  - "Gun Violence Data by James Ko", retrieved from GitHub:
    - https://github.com/jamesqo/gun-violence-data
  - The original source of the data is Gun Violence Archive's Website:
    - http://www.gunviolencearchive.org
- **Plan for graphing & analysis**:
  - Series of basic summaries of our dataset / primary exploration
  - *Analyzing and answering the 1st question*:
    - We are planning to focus on the two most popular gun types identified in our dataset. To analyze the differences in casualty rates between these types, we plan to employ two-sample means comparison (two types of guns = two independent samples). This approach will allow us to determine if there is a statistically significant difference in the impact of these gun types on casualties.
  - *Analyzing and answering the 2nd question*:
    - To investigate the relationship between the status of a gun (stolen vs. not-stolen) and the rate of casualties, we will compare two-sample means by applying a t-test to evaluate if there is a significant difference in the casualty rates (combining the number of people killed and injured) between incidents involving stolen guns and those that aren't & create graphs that will help visualize any difference in n_victims.

Analysis

- We first tidy up the data by removing columns that we do not need and rearrange them for easier analysis.
- Firstly, we want to address the relationship between the types of guns used in the gun incidents and rate of casualties which are calculated by the average number of people killed per incident. To achieve this, we filter out incidents involving multiple guns, focusing on only one gun involved. This allows us to assess the relationship between the types of guns involved and the rate of casualties more accurately.

Question 1



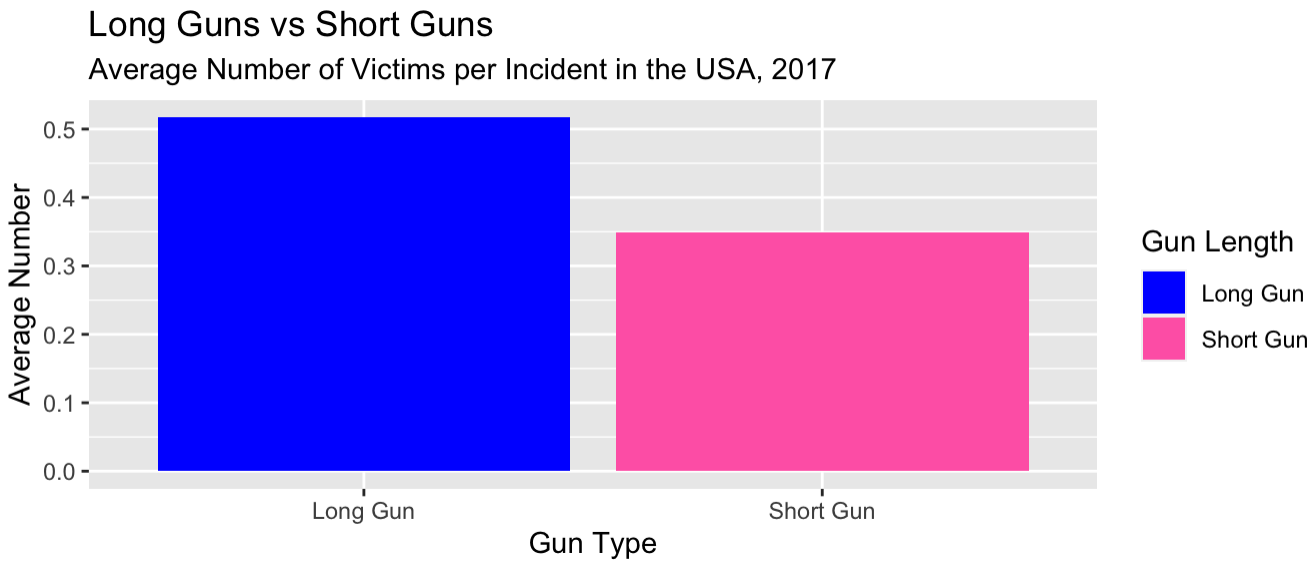
Note: For this bar graph, we have decided to exclude the "Handgun" type of guns. Our data analysis showed that "Handguns" caused the largest number of victims in total (1565 victims in total), so we have decided to exclude it for the graph aesthetics.

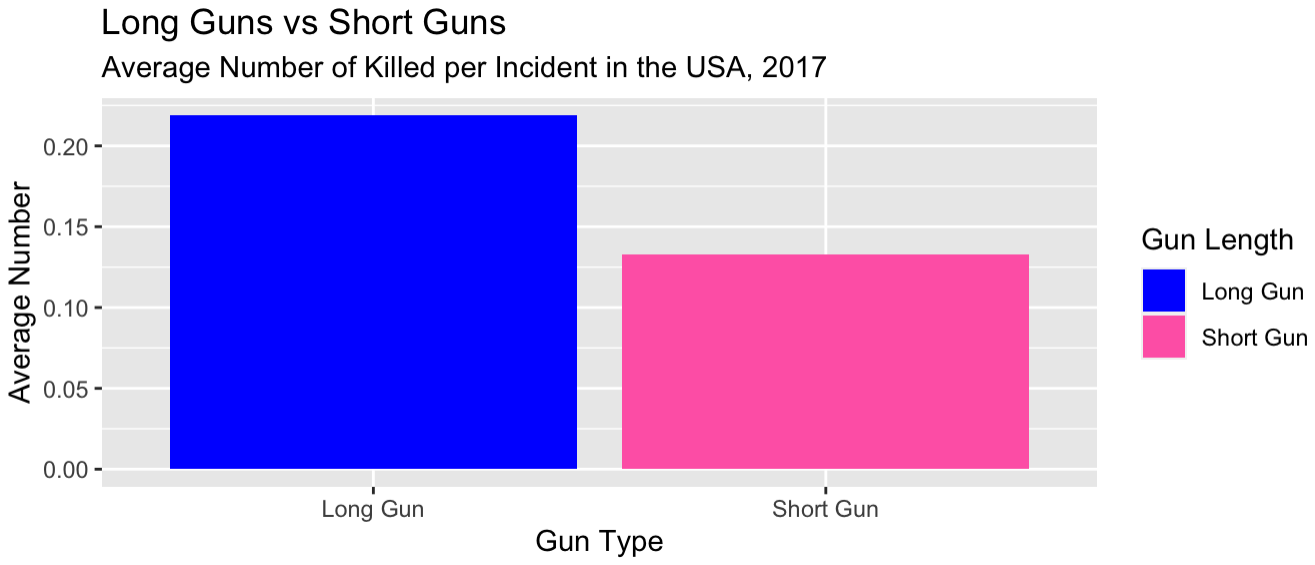
The bar graph shows each individual weapon, colored by its type, and showing the number of total victims hit. We will further categorize these weapon types to better compare long and short barreled guns. We also classified

the types of guns into two categories which are long and short guns to make it easier to employ hypothesis testing to answer our question.

```
## # A tibble: 2 × 5
##   length_type n_case n_injured n_killed n_victims
##   <chr>        <int>     <dbl>    <dbl>     <dbl>
## 1 Long Gun      1563       466      342       808
## 2 Short Gun     7624      1642     1014      2656
```

After filtering out all cases with multiple guns and unknown gun types, we are left with 9,187 cases. Guns with short barrels are clearly far more common, at least in single-gun incidents. In order to determine lethality, We will take the number of victims divided by n to show the amount of people shot per case in order to better understand how the gun type affects lethality in these incidents.





The bar graph illustrates the rate of casualties per incident involving long and short guns. Based on this bar graph the rate of casualties is higher for long guns compared to short guns. This observation suggests that long guns, such as rifles and shotguns, are associated with a greater lethality in gun violence incidents compared to short guns such as handguns, 9mm, etc.

```
##
##  Welch Two Sample t-test
```

```
##
## data:  x and y
## t = -7.7068, df = 1993.2, p-value = 2.02e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2114801 -0.1256820
## sample estimates:
## mean of x mean of y
## 0.3483736 0.5169546
```

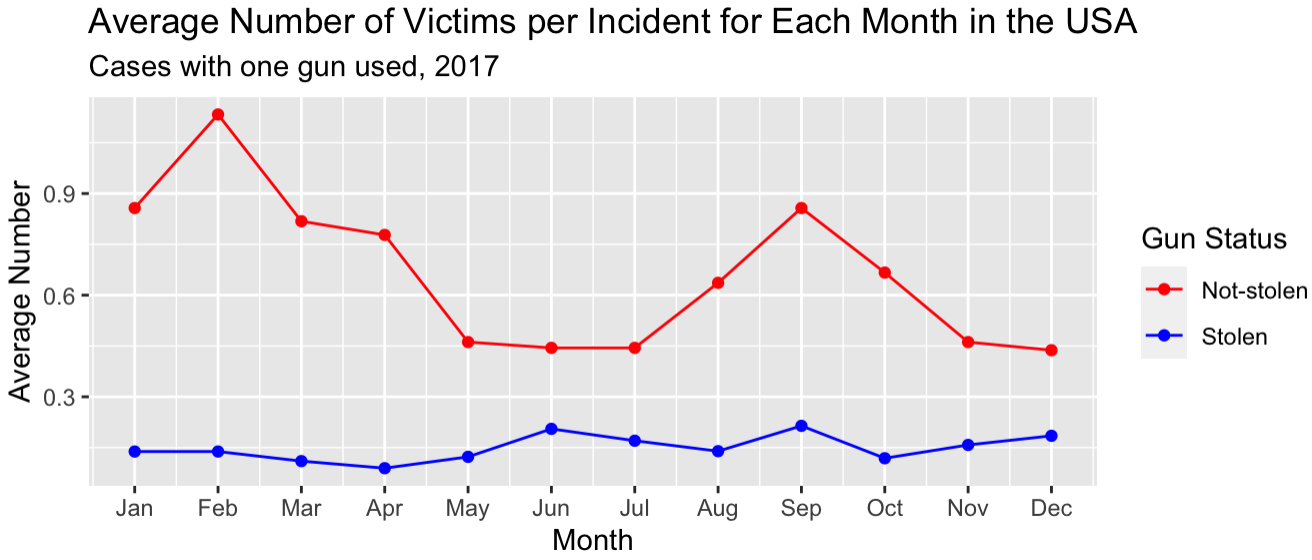
- Hypothesis testing for gun types:
    - This hypothesis test conducted compares the mean number of casualties between incidents involving long and short guns.
    - The null hypothesis: no true difference in means between incidents involving long and short guns.
    - The alternative hypothesis: there is true difference in means between incidents involving long and short guns.
    - The two sample t-test yield a p-value of 2.02e-14. Hence, with such a low p-value, significantly less than 0.05, we reject the null hypothesis in favor of the alternative hypothesis.
    - This result provides statistical evidence that the mean number of casualties is significantly higher in incidents involving long guns compared to incidents involving short guns. This finding aligns with the conclusion drawn from the bar graph, indicating that long guns are more lethal in gun violence compared to short guns.

Question 2

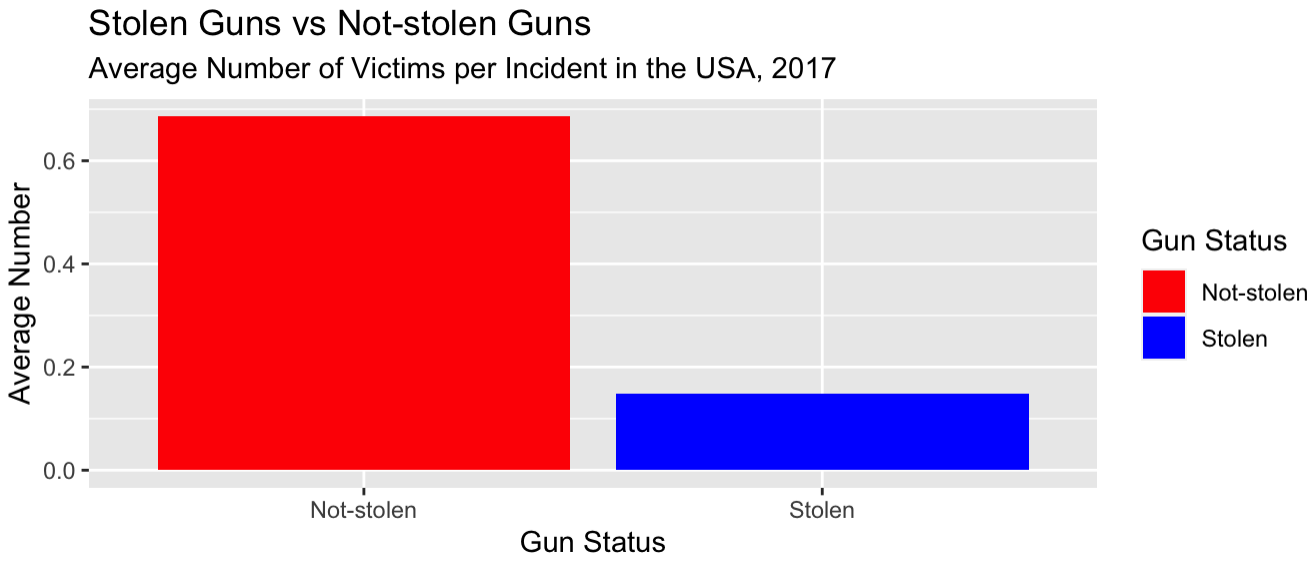
Secondly, we want to explore the status of guns involved in these incidents whether they are stolen or not stolen and see how this impacts the lethality of the case.

```
## # A tibble: 2 × 5
##   gun_stolen n_case n_injured n_killed n_victims
##   <chr>       <int>     <dbl>    <dbl>     <dbl>
## 1 Not-stolen    156        76       31       107
## 2 Stolen       1443       141       74       215
```

Similarly to the gun types table, we filtered out every case that included more than 1 gun as well as all of the unknown statuses. This resulted in 1,599 total cases remaining to be examined. Something that immediately sticks out in this table is the n_victims value of non-stolen guns. This is surprisingly high for the very small proportion of total cases involving the non-stolen weapons.



This scatter plot graph shows the average number of victims per incident categorized by stolen and not-stolen gun status. We can conclude from the graph that not-stolen guns have a higher average number of victims per incidents compared to stolen guns for each month.



The bar graph illustrates the average number of victims per incident for all states categorized whether they are stolen and not-stolen. This bar graph shows that not-stolen guns have a higher average number of victims per incident compared to stolen guns. This suggests that not-stolen guns provide a higher fatality rate in gun incidents compared to stolen guns.

```
##
##  Welch Two Sample t-test
##
## data:  x_2 and y_2
## t = 8.8583, df = 166.85, p-value = 1.166e-15
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4172406 0.6565640
## sample estimates:
## mean of x mean of y
## 0.6858974 0.1489951
```

- Hypothesis testing for gun status:
    - This hypothesis test conducted compares the mean number of victims per incident involving stolen and not-stolen guns.
    - The null hypothesis: no true difference in means between incidents involving stolen and not-stolen guns.
    - The alternative hypothesis: there is true difference in means between incidents involving stolen and not-stolen guns.
    - The two sample t-test yield a p-value of 1.166e-15. Hence, with such a low p-value, significantly less than 0.05, we reject the null hypothesis in favor of the alternative hypothesis.
    - This result provides statistical evidence that indicates a highly significant difference in means between incidents involving stolen guns and those involving not-stolen guns.
    - This finding aligns with the conclusion drawn from the bar graph, indicating that stolen guns result in higher rate of casualties compared to not-stolen guns.

Discussion

- **Interpretation of Question 1**:
    - Based on the results in question one, guns with long barrels such as shotguns and rifles, are significantly less common in gun violence cases yet cause significantly more combined deaths and injuries per case than short barrel guns. The most plausible explanation for the commonality is that handguns are the most common gun to own in the United States by a large margin. By character, they are easy to handle and are often carried along by people for protection. There are countless factors and reasons explaining why long barrel guns cause more injuries and deaths on a per case basis when compared to pistols. The most reasonable interpretation to us was that longer barreled guns usually are higher caliber and intended to be used for hunting or sport. These factors create weapons that will hit humans much harder or even hit multiple victims in some instances because of collateral damage.
- **Interpretation of Question 2**:
    - There was a very high number of unknown status on the guns used in these cases which definitely skewed our data at least slightly. However, stolen guns were far more likely to have a known status probably because it was reported stolen, making it easy to log. But based on the data, despite having more than 9 times as many cases involving stolen guns, the non-stolen gun had exactly half of the victims which caused the huge mean difference seen in the graphs. The best interpretation of these results relate to the idea that someone who owns a gun is probably better at using guns than someone who has stolen a firearm, which would help explain why the non-stolen guns are far more lethal. Another factor we considered is that a stolen weapon is far more likely to involve and attract law enforcement to apprehend the individual or more quickly prevent them from harming others.
- **Shortcomings**:
    - There are two short-comings of in our project. First is the categorization of gun types, particularly instances where the type of firearm is labeled as unknown. This lack of specificity introduces uncertainty and may lead to inaccuracies in our assessment of the relationship between gun types and lethality. The gun type column also provides unclear gun characteristics which lead us to categorizing the gun types into short and long guns for easier data analysis. Future analyses should aim to improve classification of gun types to provide more accurate insights. Other than that there was also missing data for the status of guns. Instead of just having two values for the status of guns which are stolen and not stolen the data provided values of stolen and unknown. Without accurate information on whether the guns involved were stolen, not stolen or unknown, it becomes challenging to draw definitive conclusions about the impact of gun status on casualty rates.
- **Potential future directions**:

    - We should explore additional factors beyond gun types and gun status such as socio-economic background, mental health indicators that could influence the rate of gun violence that is increasing in the United States. However, in order to do that we would need to have an extra dataset since we only chose factors in the datasets that have the strongest pattern for our conclusion.
- **Final summarization**:
    - Gun Types (Long vs Short) - This analysis reveals that incidents involving long guns result in higher casualty rates compared to incidents involving short guns. Long guns, which include rifles, shotguns etc exhibit a greater chance for causing fatalities and injuries in gun violence incidents.
    - Gun Status (Stolen vs not Stolen) - This analysis demonstrates that incidents involving not stolen guns lead to higher casualty rates than incidents involving not-stolen guns. Firearms reported as legal possession are associated with increased lethality in gun violence incidents, indicating a correlation between firearm ownership status and the severity of outcomes.

> In summary, the analysis presents evidence indicating that both gun characteristics which are gun types and gun status play crucial roles in determining the severity of gun violence incidents.