“Understanding Gun Violence in the United States”

Alex Bendeck, Thea Dowrich, Ashley Murray and Matty Pahren

Introduction

In recent years, we have seen a rapid increase in the rate of gun violence incidents in the United States. The United States is one of a few countries in which the right to bear arms is constitutionally protected, which makes gun control policy intended to protect our constituents difficult to enact. In recent years, there has been a push to implement more progressive gun control policies at the federal and state levels, but there is still a long way to go before we can ensure the safety and security of our communities.

Since the horrific Sandy Hook shooting of 2012, there have been over 2,000 mass shootings in the U.S. Mass shootings are incidents in which “four or more people were shot”[1]. Many gun violence incidents go unreported, however, so this number could potentially be an underestimation.

An important characteristic of gun violence incidents is the number of individuals injured and killed. Events are largely characterized by their fatalities; thus, it is important to understand what factors influence an event’s deaths and injuries. In media reports, many journalists discuss the perpetrator’s weapon of choice, as military-grade guns have the capacity to cause substantially more harm than handguns. As mentioned before, states have tried to implement their own gun safety policies, and as such, there are varying policies across the United States. Therefore, the location of the gun violence incident can also play a role in how many individuals were hurt in the midst of the attack.

In short, the goal of our project is to understand the causes of gun violence incidents and how they differ between our 50 states. Our primary goal of our model is to predict the number of injuries and deaths in a shooting event and its relationship with a states’ policies. Moreover, we hope to predict the number of shootings in a given year. We hope to get a greater picture of the problem that is plaguing our society today and gain insights that can inform policy decisions going forward.

Our data is all recorded gun violence incidents in the United States between January 2013 and December 2018. We will use Poisson regression to predict the number of shootings, as well as the number of injuries and fatalities in state in a given year. The range of injuries in our dataset is between 0 and 53, with a mean value of 0.494. The range of deaths in our dataset is between 0 and 50, with a mean of 0.2523. Since we have such small occurrences of injuries and deaths, Poisson regression makes sense as it is intended for incidents with few counts. Additionally, in understanding the importance of different incident-level characteristics on the severity of a shooting, we will be using a logistic regression to determine the probability that people are injured in a given instance of gun violence. Interpretation is of utmost importance in our modeling strategies because we want our findings to be understood by policymakers. Therefore, we will not be using random forests and other less interpretable, though highly predictable models of that variety.

Data

The main dataset utilized in this analysis was hosted on Kaggle, and was originally downloaded from the Gun Violence Archive (GVA). The Gun Violence Archive is a “not for profit corporation formed in 2013 to provide free online public access to accurate information about gun-related violence in the United States. GVA will collect and check for accuracy, comprehensive information and gun-related violence in the U.S. and then post and disseminate it online” [1]. Within the dataset, there are more than 260k gun violence incidents, alongside specific information about each of the incidents.

Alongside the dataset containing the gun violence incidents, we will be utilizing a dataset from `State Firearm Laws` that includes information concerning state-specific gun laws that will be merged by the predictor `state` with the original dataset. Additionally, in order to control for the varying population sizes across different states, we will also be utilizing a dataset from the `United States Census Bureau` that hosts information concerning state population totals for the specified years that the gun violence incidents took place.

We examined an extensive list of variables in our models, and a full table of them can be found in the appendix.

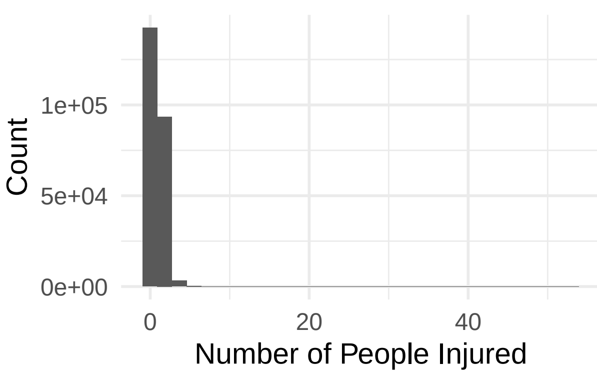
Data Cleaning

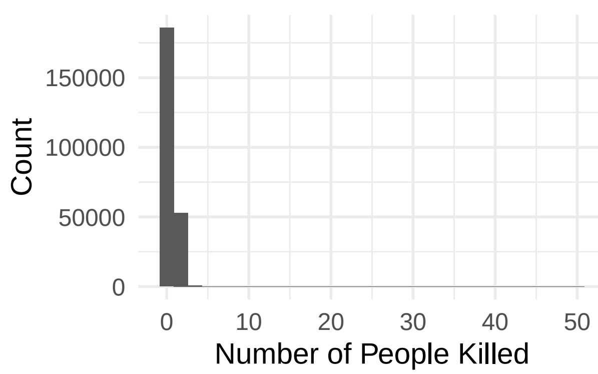
Some of the columns in the data pulled from Kaggle was unusable at first, so the data had to be cleaned. Initially, the incident characteristics column was a large chunk of text containing various characteristics of the gun violence incident, such as whether the incident was a mass shooting, was a result of domestic violence, or occurred when someone was under the influence of alcohol or drugs. To make this information usable in our model, we made indicator variables which equal one if a certain phrase is in the larger incident characteristic text and 0 if not. Ultimately, we ended up with 21 new variables which are: whether or not the gun incident was a mass shooting, if it involved an unlawfully purchased firearm, if it happened at a bar or a club, if it involved an assault weapon, if it involved the TSA, if an officer was involved, if the weapon was used in defense, if drugs were involved, if a murder or suicide happened, if there was a kidnapping, if people involved in the incident were under the influence, if the incident was a result of domestic violence, if the incident was a result of an armed robbery, if people involved had an open carry permit, if it was a drive-by shooting, if the incident was a result of negligent discharge, if the incident was part of a raid, if there was gang involvement, if people involved had a concealed carry permit, if an animal was harmed, and if the suspect used an illegally owned gun.

Additionally, some of the other data columns such as participant age group, participant gender, and participant status were recorded as a dictionary of information which made it hard to use. To fix this problem, we ran through each observation and recorded the total occurrences of a category. For example, we looked through all participants in a shooting, including victims and suspects, and recorded the number of males involved, females involved, and the number of participant's whose gender is unknown in three separate columns. We did the same for age, except the columns were child, teenager or adult.

Exploratory Data Analysis

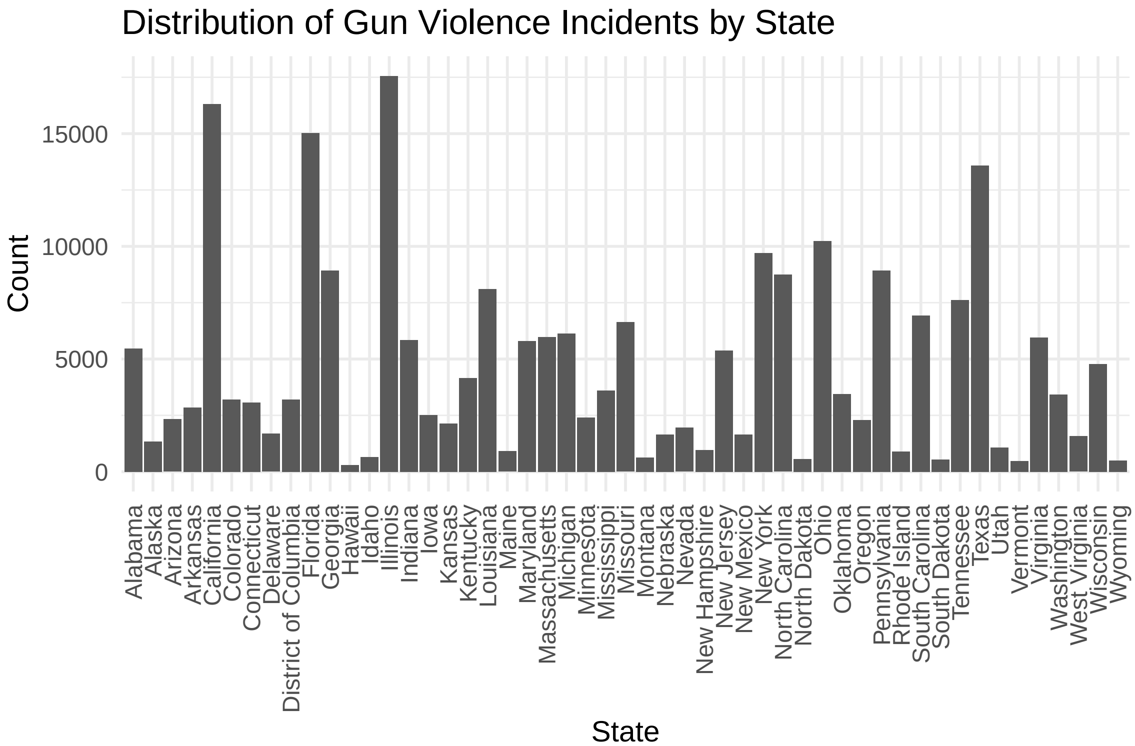
To start, we examine the distribution of people injured and killed in each shooting. The plots below show that both of these variables are extremely right-skewed and have a large number of observations at or close to zero.



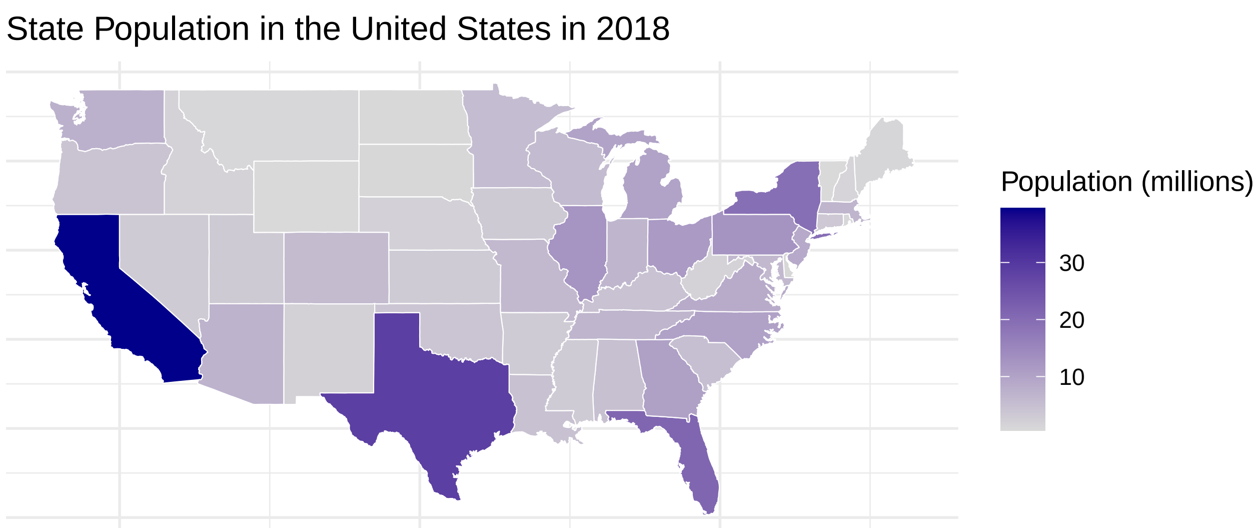


Next, we examine the distribution of gun violence incidents by state. From this plot, we can see that states like California, Florida, Illinois, and Texas have the most number of gun violence incidents, while states like

Hawaii, Wyoming, North dakota, and Vermont have very few incidents recorded.



Based on the previous plot, a state's size seems like it could impact the number of gun violence incidences that have happened there. Therefore, we next examined population by state. From the plot below, it is clear that California, Texas, and Florida are some of the most populous states. Thus, it makes sense that there might be more instance of gun violence there just because they have more people.

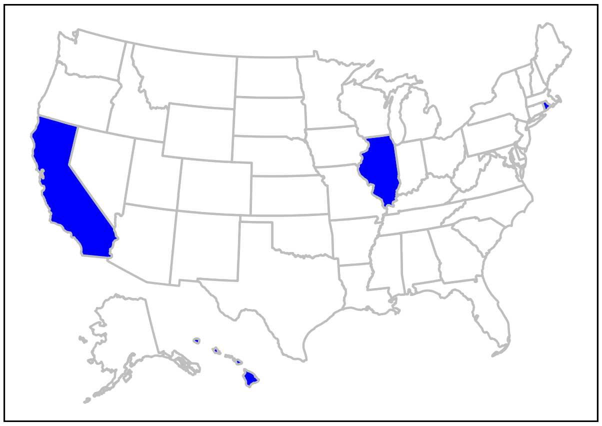


Additionally, we are interested in whether or not the passage of certain gun laws has an impact on the number of gun violence incidents in a given state.

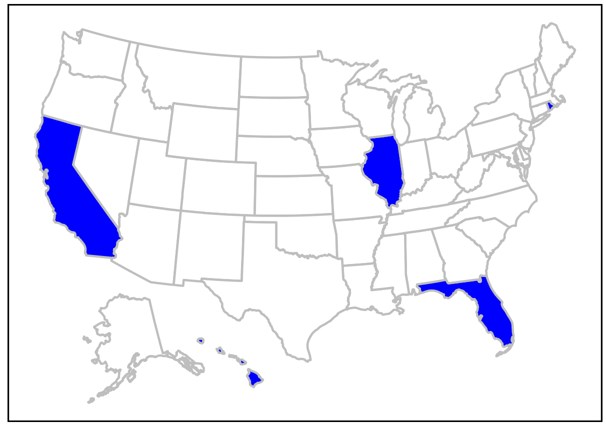
From the graph below, it is clear that not many states require a waiting period when purchasing a gun. However, Florida notably adopted this policy between 2013 and 2018. In Florida, a gun buyer now has to wait 3 business days between the time they purchased the gun and when they can receive said gun. Florida actually passed several new gun laws in 2018, following the shooting at Marjory Stoneman Douglas High

School in Parkland where 17 people were killed [2]. Florida now also requires people to be 21 instead of 18 in order to purchase a gun, and they banned bump stocks, which allow semi-automatic weapons to work more like automatic weapons.

States With a Waiting Period in 2013

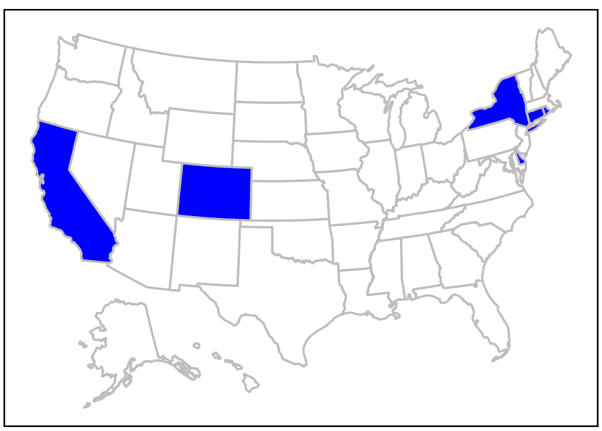


States With a Waiting Period in 2018

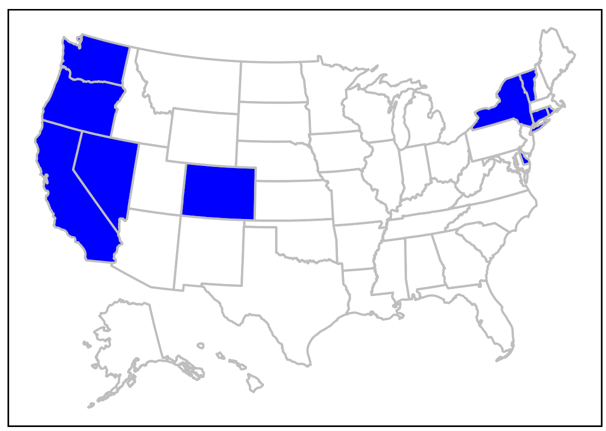


Gun violence is a large point of discussion in American politics, and there have been a large movement to make gun laws more strict, since the US appears to have more issues with gun violence than other countries. One law that is gaining more attention by states is the requirement of universal background checks. The maps below show that a couple more states have required background checks in 2018 compared to 2013, and since 2018, other states including New Jersey, New Mexico, Virginia have also adopted these policies [3].

Background Checks Required in 2013



Background Checks Required in 2018



Modeling Number of Shootings Per Year

One of our goals is to determine which gun laws at the state level are significant predictors of high numbers of shootings. In order to do this, we consider the total number of shootings in each of the 50 U.S. states in the years 2014, 2015, 2016, and 2017 from the Kaggle gun violence dataset and augment it with the active gun laws in each state during those years from the gun laws dataset. From a modeling approach, it seems clear that a Poisson regression model is appropriate. Poisson regression is often used when the response variable is some type of count per unit time or space, and in this case the response variable is the number of shootings in a given state in one year. Intuitively, one important issue that must be dealt with is the difference in population between states. For example, we would expect heavily populated California to have many more shootings than sparsely populated Wyoming simply because there are more people in California.

Our first idea of a model to fit for this question was a Poisson regression model that included the state population (for each year) as a predictor. However, it quickly became clear that this model was not an ideal fit. Consider the issue of interpreting this coefficient. While it makes sense that more densely populated states will have more shootings, this does not really provide meaningful or actionable interpretations. For example, state legislators can't simply work to reduce the population in their states in order to limit shootings. We next considered that a model which included an intercept for each state (i.e. a categorical predictor for state) might fit better than the continuous population predictor for this reason. However, this would not account for one last wrinkle in this dataset: repeated data observations from the same states over time. Since data points from the same states are highly correlated, running a regular Poisson model (even with an intercept for each state) will not account for these repeated measures, meaning that our your standard error estimates and p-values would not necessarily be reliable. Finally, in order to both account for differences between states (such as population) and repeated data points from states over time, our final model is a mixed-effects Poisson regression model with a random intercept for each state.

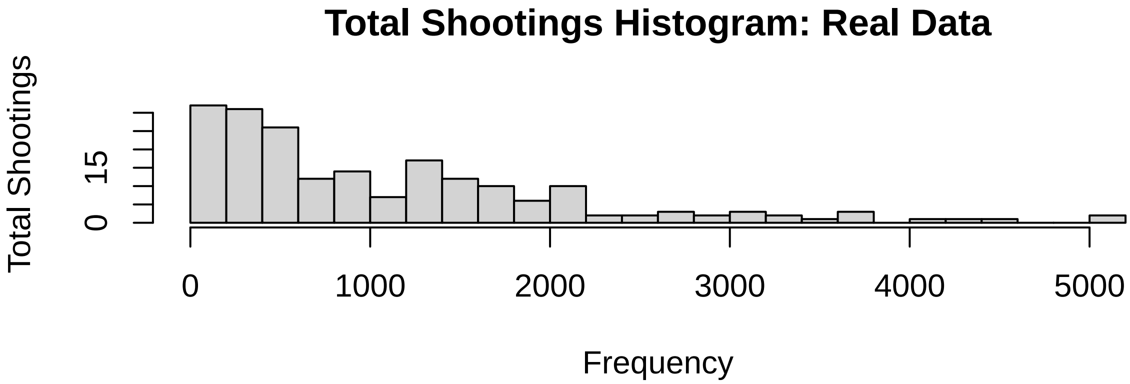
Table 1: Yearly Shootings Model Coefficients

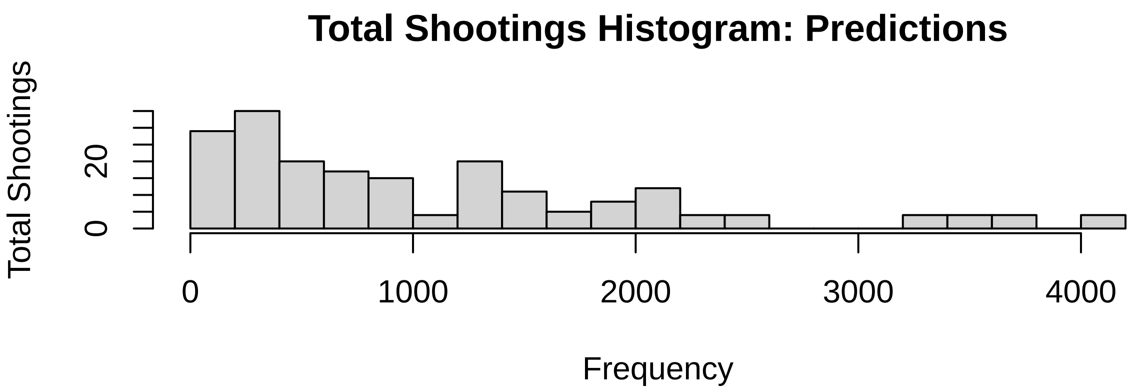
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.6136	0.1503	43.9890	0.0000
registration	-2.2816	0.9929	-2.2980	0.0216
permitconcealed	-0.1217	0.0178	-6.8246	0.0000
assault	0.9068	0.4274	2.1216	0.0339
universal	0.3745	0.0553	6.7673	0.0000
gunshow	-0.4372	0.0773	-5.6547	0.0000

We initially fit this mixed-effects Poisson regression model with more than a dozen categorical predictors denoting the presence of commonly-discussed gun control legislation, such as background checks and gun show sale regulations. We then conducted backwards selection using BIC to select a subset of predictors to produce a more parsimonious model. Now let's interpret the coefficients that were selected to be in the final model. Three predictors suggest actions that can be taken by states to lessen the number of expected occurrences of shootings. Holding all else constant, a state requiring gun owners to register their firearms with the state will have $e^{-2.28} \approx 0.10$ times the number of shootings as a state that does not. Additionally, a state that requires a permit to carry concealed weapons is expected to have $e^{-0.12} \approx 0.89$ times the number of shootings as a state that does not. Also, a state that requires universal background checks upon purchase of any firearm at a gun show is expected to have $e^{-0.44} \approx 0.64$ times the number of shootings as a state that does not. However, two other laws seem to have the opposite of the intended effect of reducing gun violence. Holding all else constant, a state that bans the sale of assault weapons is expected to have $e^{0.91} \approx 2.48$ times the number of shootings as a state that does not. Furthermore, a state that requires universal background checks upon purchase of firearms at gun dealers is expected to have $e^{0.37} \approx 1.45$ times the number of shootings as a state that does not.

The main takeaways from this part of our analysis are as follows. We started with more than a dozen different gun control laws in our model that are meant to curb gun violence incidents. We found that most of the laws had no effect on occurrences of gun violence, three had a significant effect of reducing the number of incidents, and two actually had a significant effect of increasing the number of incidents. Based only on this data and model, we would recommend that legislators require gun registration with the state at the time of purchase, require universal background checks at gun shows, and require permits to carry concealed weapons in order to reduce gun violence. However, the results of this analysis still leave some questions unanswered. For example, why would it be that requiring universal background checks at gun shows is correlated with fewer shootings, whereas universal background checks at more conventional retailers has the opposite effect? Outside the scope of this analysis, one could imagine investigating more closely the source of firearms that are used in shootings. For example, are these guns often stolen, bought at gun shows, or bought at gun retailers? Answers to these questions may provide more insight into why we see the trends that are present in the data and how legislators might interpret them. Similar contextualization is needed to understand why banning assault weapons in a state might actually result in more gun violence. Opponents of assault weapons bans often argue that criminals will find ways to acquire such weapons regardless of laws that limit their sale, though we would again need additional data to make a detemination on this. Regardless, this is certainly a complex issue, but the results of this analysis can provide insight into what kind of next steps to take or investigate in order to curb gun violence.

In order to assess how well this model fits, we should first compare the distribution of the original data observations to that of the predictions from this model.





From these plots, we see that the model is a reasonably good fit for the data. One concern we had was that perhaps we would need to use a zero-inflated Poisson model rather than a standard Poisson model. However, it seems that the current model is doing a decent job. For both the actual data and predicted values, there are a similar number of data observations in the smallest histogram bin, indicating that the model is not underestimating the prevalence of small data observations.

We also check a plot of the residuals against the predicted values to further check the model fit. This plot is visible in the Appendix. Overall, the residual plot indicates no serious issues with the model. One potential area of concern is overdispersion, where the mean and variance of the data observations are not the same, breaking the assumption that the data follow a Poisson distribution. However, we see only marginal evidence of overdispersion for very high predicted values, and even then the overdispersion is very slight. Overall, we conclude that this model is a reasonably good fit for the data.

Modeling Occurrence of Injuries

Another one of our goals is to predict which characteristics of a shooting make it more or less likely to result in injuries. It was somewhat surprising to us that more than half of the shooting incidents in our dataset resulted in no injuries. To accomplish this, we fit a logistic regression model where each data row was a shooting and the response variable was a binary indicator of whether or not any injuries were sustained. The model includes twenty different predictors involving various characteristics of the circumstances surrounding the shooting.

Table 2: Injuries Model Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.1049	0.0057	-18.3115	0.0000
unlawful_purchase	-2.0397	0.2129	-9.5821	0.0000
bar_club	0.7691	0.0334	23.0148	0.0000
assault_weapon	-0.6822	0.0725	-9.4033	0.0000
tsa	-8.4112	0.9978	-8.4299	0.0000
officer_involved	-0.3620	0.0186	-19.4208	0.0000
defensive_use	0.2744	0.0263	10.4157	0.0000
drug_involvement	-0.7438	0.0245	-30.3497	0.0000
murder_suicide	-1.1828	0.0498	-23.7621	0.0000
kidnapping	-0.6973	0.0533	-13.0763	0.0000
under_influence	-0.9434	0.0492	-19.1675	0.0000
domestic_violence	-0.0712	0.0237	-3.0054	0.0027
armed_robbery	-0.1083	0.0163	-6.6559	0.0000
open_carry	-1.5106	0.0227	-66.5994	0.0000
drive_by	0.5127	0.0191	26.8486	0.0000
negligent_discharge	1.0746	0.0291	36.9292	0.0000
raid	-1.9983	0.0357	-56.0135	0.0000
gang_involvement	0.4302	0.0298	14.4252	0.0000
concealed_carry	-0.2960	0.0709	-4.1764	0.0000
animal	-1.2156	0.0977	-12.4401	0.0000
illegal_possession	-1.2348	0.0449	-27.4940	0.0000

We found that essentially all of the predictors we included in the model were significant. This seemed strange at first, and we checked to see if anything weird was happening with collinearity between predictors, but the predictors were not really correlated at all. We believe that the significance of the predictors is actually a result of the large number of data observations with which this model was fit (> 220,000). As a result, we will interpret only the predictors with the largest z-values. Holding all else constant, for the predictors with the largest effects of reducing the odds of injuries, shootings where at least one of the involved weapons was previously issued an open carry permit multiplies the odds of having at least one injury by $e^{-1.51} \approx 0.22$. If a shooting occurred as part of a police raid, this multiplies the odds of having at least one injury by $e^{-2.00} \approx 0.14$. If the shooting occurred as part of a drug sale or other drug-related transaction, this multiplies the odds of having at least one injury by $e^{-0.74} \approx 0.48$. Holding all else constant, for the predictors with the largest effects of increasing the odds of injuries, the incident occurring as the result of an accidental or negligent firing of a gun multiplies the odds of having at least one injury by $e^{1.07} \approx 2.91$. The incident being a drive-by shooting multiplies the odds of having at least one injury by $e^{0.51} \approx 1.67$. The shooting occuring at a bar or club multiplies the odds of having at least one injury by $e^{0.77} \approx 2.16$.

There are several insights that can potentially be gleaned from this model. Recommendations for policymakers from this model would be to try and prevent the more deadly circumstances from arising. Accidental or negligent firings of a gun are very highly correlated with injuries during shooting incidents. This suggests that requiring some kind of training or permit when purchasing a gun could help reduce gun casualties by ensuring that gun buyers have the expertise to properly handle their firearm. This will reduce the likelihood of negligent gun firing in the first place. Similarly, it seems as though shootings which occur at bars or clubs are particularly susceptible to resulting in injuries, perhaps because many people are in a relatively small space. This would suggest that perhaps more stringent security checks and presence of security guards at clubs may be needed to help reduce the occurrence of club shootings, since when these shootings occur they often lead to casualties. It is not extremely surprising that police raids are less likely to lead to injuries, since police receive training on how to use weapons and are more likely to know how to avoid injury when being shot at.

To see if the model is an adequate fit for the data, we can look at the binned Pearson residuals plotted against

Table 3: Confusion Matrices with 0.5 (left) and 0.47 (right) Prediction Cutoffs

	Actual 0	Actual 1		Actual 0	Actual 1
Predicted 0	31971	20328	Predicted 0	15809	4849
Predicted 1	3107	4853	Predicted 1	19269	20332

the predicted probabilities. This plot is included in the Appendix. The residual plot does not show anything of major concern. Besides one outlier data point with a very small predicted probability, the residuals are nicely centered around 0 with no worrisome patterns.

Although our primary goal in fitting this model was inference, it is also worthwhile to see how this model performs in terms of prediction. To do this, we re-fit the model using only data from the years 2014-2016, leaving the 2017 data as a hold-out test set. First, we marked all data points from the test set with a predicted probability at or above 0.5 as predicted to be a 1 (indicating at least one injury) and the rest to be 0. As shown below, this resulted in around 61% predictive accuracy. Looking at the confusion matrix below, we see that this model predicted many false negatives; in other words, it predicted many incidents to not have any injuries when in fact injuries were sustained. We subsequently tried adjusting the predicted probability threshold, previously 0.5, for marking predictions as either a 1 or 0. We found that once the threshold reaches 0.47 or lower, there is a large shift where instead of seeing many false negatives, there are instead many false positives, as shown in the second confusion matrix below. The predictive accuracy is again around 60%. It seems that many shootings have a predicted probability of having injuries just under 0.5, which accounts for this shift when we change the prediction threshold. Overall, this model does not do a particularly good job at prediction, and it seems that for many shootings there are around even odds of at least one person sustaining an injury. However, as mentioned earlier, the main purpose of this model was inference rather than prediction. We are satisfied with the conclusions we were able to draw from this model, so we are willing to make the tradeoff of slightly subpar predictive accuracy.

Modeling the Total Number of Injuries and Deaths per Year

Gun violence incidents ravage our communities due to the damage they can inflict on others. Many persons can be injured and even killed at the hand of the gun, and we have seen these effects in our national media streams. Therefore, another one of our modeling goals was to determine how many individuals are injured or killed in a gun violence incident in each state, every year. We sought to numerate the violence caused by these events using gun control law data by state. In that, we wanted to determine if variable gun legislation influenced the amount of harm that could come from a gun violence incident.

To do so, we used a dataset from Kaggle for shootings in all 50 states between the years of 2013 and 2018. This dataset was then joined to another that contained data on the gun laws by each state and year. The modeling strategy was built upon a Poisson regression framework. Poisson models are often used when the response variable is a count per unit time, and in our case, the numerical values for `total_injured` and `total_killed` are counts per year.

The model building strategy was equivalent for each response variable, as they are essentially the same. The initial step was to create a Poisson model that included all variables, including state. However, this output returned over 10 NA coefficients, indicating that the variables are correlated with one another. It can be readily understood that this is due to the state variable. The legislation in our dataset is at a state-level and as such, the predictors will be highly related to one another. Compare this to federal legislation, such as the Constitution. The Constitution protects the right to bear arms and is independent of what state one is in. However, there are certain regulations that come with this right that are entirely dependent upon your location. Therefore, the next step was to turn state into a random effect to account for our troublesome circumstance. Additionally, our data is per year, so this variable was removed at the outset. Further, it is unlikely that any state legislature would be in a situation in which they had the total individuals injured without the total killed and vice versa and so these effects were removed initially.

The final models to predict `total_injured` and `total_killed` were mixed effects Poisson, as shown below.

Table 4: Total Injured Model Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.4532	0.1950	27.9641	0.0000
mass_shootings	0.0194	0.0007	27.1670	0.0000
invcommitment	-0.4908	0.1028	-4.7743	0.0000
invoutpatient	0.4542	0.1271	3.5747	0.0004
danger	0.5159	0.1081	4.7743	0.0000
loststolen	0.1884	0.0267	7.0680	0.0000
universal	-0.1476	0.0790	-1.8680	0.0618
ccrenewbackground	-0.1583	0.0317	-4.9933	0.0000
mcdv	-0.2248	0.0650	-3.4589	0.0005
dvro	0.2755	0.0346	7.9724	0.0000
stalking	-0.1651	0.1588	-1.0394	0.2986

The final predictors in the `total_injured` model are `mass_shootings`, `invcomittment`, `invoutpatient`, `danger`, `loststolen`, `universal`, `ccrenewbackground`, `mcdv`, `dvro` and `stalking`. It is important to understand what these predictors mean in context, so interpretations of these coefficients will be provided. Holding all else constant, for each additional mass shooting a state has each year, the total number of persons injured changes by a multiplicative factor $e^{0.0194} \approx 1.02$. For states that do not prohibit firearm possession to those who have been involuntarily committed to an outpatient facility, the total number of persons injured will be $e^{-0.4908} \approx 0.612$ times higher than states that do. On the same wavelength, states that do not prohibit firearm possession to those who have been involuntarily committed to an inpatient facility, the total number of persons injured will be $e^{0.4542} \approx 1.57$ times higher than states that do. In states that do not prohibit firearm possession for those that have been deemed by the court to be a danger to oneself or others, the total number of persons injured will be $e^{0.5159} \approx 1.68$ times higher than states that. For states that do not require reporting of lost/stolen guns by firearm owners, the total number of persons injured will be $e^{0.1884} \approx 1.21$ times higher than states that do. States that do not mandate universal background checks see the total number of persons injured to be $e^{-0.1476} \approx 0.863$ times higher than states with them. In states that do not require new background checks for concealed permit renewal, the total number of persons injured will be $e^{-0.1583} \approx 0.853$ times higher than states that do. In states that do not prohibit persons who have been convicted of misdemeanor domestic violence, the total number of persons injured will be $e^{-0.2248} \approx 0.799$ times higher than states that do. For states that do not have a law automatically prohibiting domestic violence-related restraining order subjects from obtaining a firearm have, the total number of persons injured will be $e^{0.2755} \approx 1.32$ times higher than states that do have this prohibition. Finally, for states that do not have a law prohibiting firearm possession to persons with a stalking conviction, the total number of persons injured will be $e^{-0.1651} \approx 0.848$ times higher than states that do have this prohibition.

Table 5: Total Killed Model Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.8255	0.1742	27.7000	0.0000
mass_shootings	0.0106	0.0010	10.4798	0.0000
danger	0.3239	0.0853	3.7979	0.0001
permit	0.0412	0.0322	1.2793	0.2008
waitingh	-0.2735	0.0858	-3.1886	0.0014
microstamp	-0.2884	0.0651	-4.4330	0.0000
mcdv	-0.2208	0.0813	-2.7161	0.0066
dvro	0.1950	0.0519	3.7583	0.0002

The final predictors in the `total_killed` model are `mass_shootings`, `danger`, `permit`, `waitingh`, `microstamp`, `mcdv`, and `dvro`. It is important to understand what these predictors mean in context, so interpretations of these coefficients will be provided. Holding all else constant, for each additional mass shooting a state has each year, the total number of persons killed changes by a multiplicative factor $e^{0.0106} \approx 1.01$. In states that do not prohibit firearm possession for those that have been deemed by the court to be a danger to oneself or others, the total number of persons killed will be $e^{0.3239} \approx 1.38$ times higher than states that. For states that do not require a permit to purchase all firearms, the total number of persons killed will be $e^{0.0412} \approx 1.51$ times higher than states that do. States that do not have a waiting period on all handgun purchases from dealers see the total number of persons killed to be $e^{-0.2735} \approx 0.761$ times higher than states with them. In states that do not require all sold handguns have ballistic fingerprinting or microstamping for identification, the total number of persons killed will be $e^{-0.2884} \approx 0.749$ times higher than states that do. In states that do not prohibit persons who have been convicted of misdemeanor domestic violence, the total number of persons killed will be $e^{-0.2208} \approx 0.802$ times higher than states that do. For states that do not have a law automatically prohibiting domestic violence-related restraining order subjects from obtaining a firearm have, the total number of persons killed will be $e^{0.195} \approx 1.22$ times higher than states that do have this prohibition.

These results present great opportunity for legislative action. States that do not require a permit to purchase all firearms are significantly correlated with high fatalities. This potentially signals to lawmakers on the fence about gun control policies that a firearm permit is a great first step to protect our communities. Further, protecting constituents from those with a history of violence via prohibiting firearm purchase for persons who have been convicted of misdemeanor domestic violence from significantly reduces the amount of damage done in a gun violence incident.

At the outset of model construction, collinearity was a prevalent issue in both situations. However, all predictors in both models have variance inflation factor values of less than 10. In fact, all predictors have values less than seven.

To further verify this model's performance, we can look to the distribution of the residuals. It can be seen that the residuals are relatively normally distributed and randomly scattered indicating that our model is a good fit. These plots are included in the Appendix.

Conclusion

From the four models that were fit, a lot of information can be derived concerning the impact of gun violence. With the modeling of the number of shootings per year, we fit a Poisson regression with a random effect for each state. And with our model selection approach, it was found that most of the gun law predictors were not significant in determining the number of shootings per year. However, we had found that three had a significant effect of reducing the number of incidents, and two had a significant effect of increasing the number of incidents. With the modeling of the occurrence of injuries, we fit a logistic regression model where each data row was a shooting, with a response variable of binary indicator of whether or not any injuries were sustained. From this model, it was found that all of the predictors were significant.

There was a limitation in the fact that the predictors were seen as significant given the fact that there are a lot of gun violence occurrences, and the fact that a lot of the predictors were highly correlated with each other. For future work, it would be interesting to see whether other gun laws outside of the ones included in this analysis influence gun-related violence.

Citations

[1] https://www.gunviolencearchive.org

[2] Melendez, Steven. “Here’s a List of Gun Control Laws Passed since the Parkland Shooting.” Fast Company, Fast Company, 13 Feb. 2019, www.fastcompany.com/90306582/heres-a-list-of-gun-control-laws-passed-since-the-parkland-shooting.

[3] “Universal Background Checks.” Giffords, 28 Oct. 2020, giffords.org/lawcenter/gun-laws/policy-areas/background-checks/universal-background-checks/.

[3] Introduction to Creating Maps with ggplot2, remiller1450.github.io/s230s19/Intro_maps.html.

Appendix

Variable Definitions

Table 6: Incident Variables

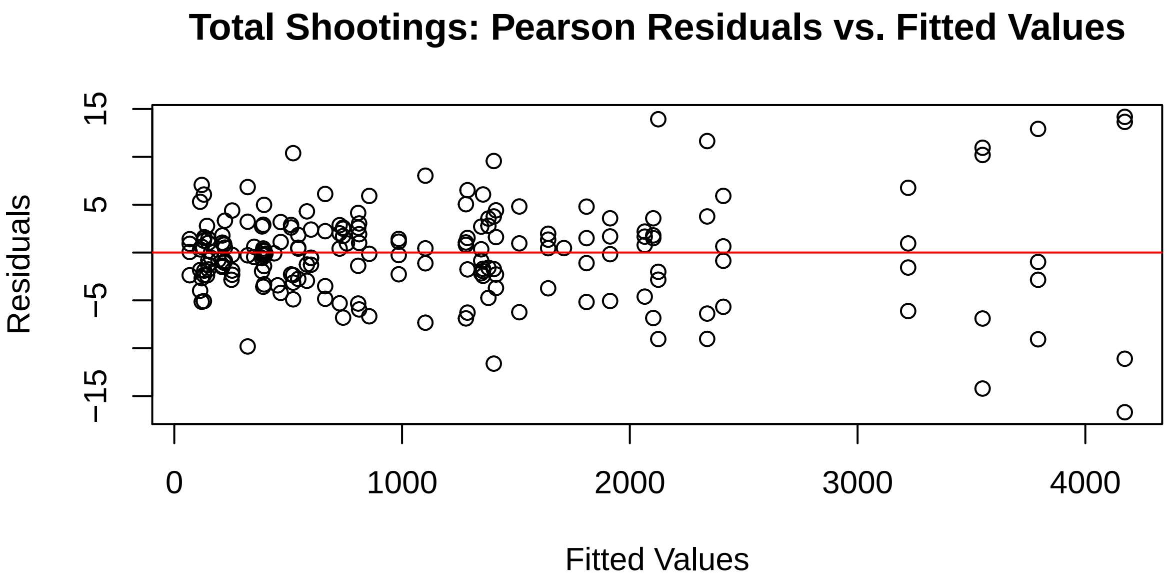
Variable	Description
State	State in which the incident took place
Year	Year of the incident
Killed	Number of people killed in an incident
Injured	Number of people injured in an incident
Mass Shooting	Whether or not the incident was a mass shooting
Unlawful Purchase	Whether or not the incident involved an unlawfully purchased firearm
Bar/Club	Whether or not the incident took place at a bar or a club
Assault Weapon	Whether or not the incident involved an assault weapon
TSA	Whether or not the incident involved the TSA
Officer Involved	Whether or not the incident involved a law enforcement officer
Defensive Use	Whether or not a gun was used defensively in the incident
Drug Involvement	Whether or not a drug transation was involved in the incident
Murder/Suicide	Whether or not the incident was a murder or suicide
Kidnapping	Whether or not the incident involved a kidnapping
Under Influence	Whether or not persons involved in the incident were under the influence of drugs or alcohol
Domestic Violence	Whether or not the incident was a result of domestic violence
Armed Robbery	Whether or not the incident was part of an armed robbery
Open Carry	Whether or not persons involved in the incident had an open carry permit
Drive By	Whether or not the incident was a drive-by shooting
Negligent Discharge	Whether or not the incident was a result of negligent discharge of a firearm
Raid	Whether or not the incident was part of a raid
Gang Involvement	Whether or not the incident involved gangs
Concealed Carry	Whether or not persons involved in the incident had a concealed carry permit
Animal	Whether or not an animal was involved in the incident
Illegal Possession	Whether or not persons involved in the incident involved an illegally possessed firearm
Participant Gender - Male	The number of males involved in the incident (suspencts and victims)
Particpant Gender - Female	The number of females involved in the incident (suspencts and victims)
Participant Gender - Unknown	The number of people of unknown gender involved in the incident (suspencts and victims)
Age Group - Child	The number of children involved in the incident (suspects and victims)
Age Group - Teen	The number of teenagers involved in the incident (suspects and victims)
Age Group - Adult	The number of adults involved in the incident (suspects and victims)
Age Group - Unknown	The number of people of unknown age group involved in the incident (suspects and victims)

Table 7: State-Level Variables

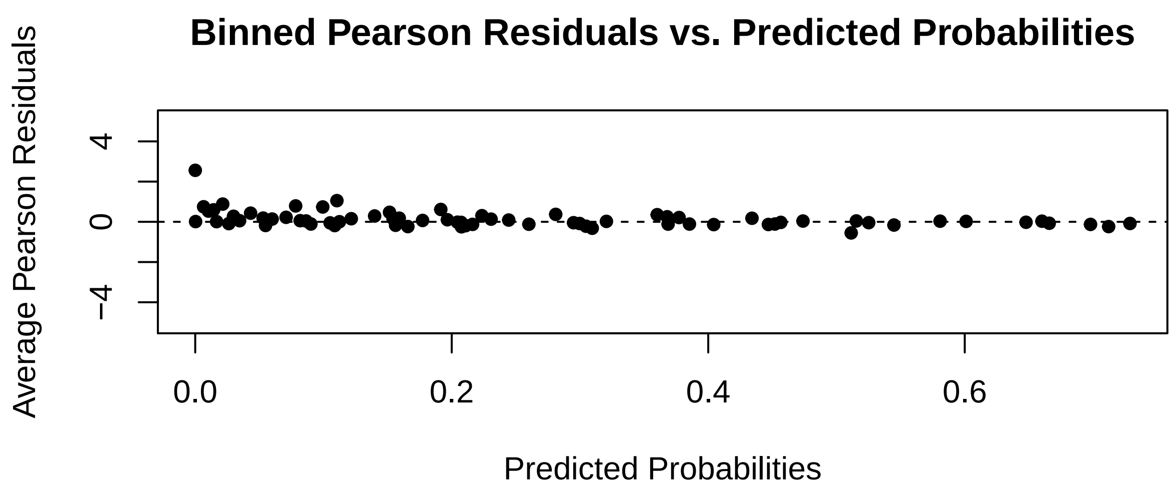
Variable	Description
State	State where a given law is in effect
Year	Year in which a given law is in effect
Population	State's population

Variable	Description
Dealer	State dealer license required for sale of all firearms
Liability	Dealers are liable for damages resulting from illegal gun sales
Waiting - Handgun	Waiting period is required on all handgun purchases from dealers
Permit	A license or permit is required to purchase all firearms
Fingerprint	Buyers must be fingerprinted at point of purchase
Training	Safety training or testing required prior to issuing a firearm license or permit
Registration	Gun owners must register their firearms with the state
Handgun Sale - Age 21	Purchase of handguns from licensed dealers and private sellers restricted to age 21 and older
Open Carry Handgun	No open carry of handguns is allowed in public places
Permit Concealed	Permit required to carry concealed weapons
Assault	Ban on sale of assault weapons beyond just assault pistols
Felony	Firearm possession is prohibited for all people with a felony conviction
Violent	Firearm possession is prohibited for people who have committed a violent misdemeanor punishable by less than one year of imprisonment
Universal	Universal background checks required at point of purchase for all firearms
Gun Show	Background checks required for all gun show firearm sales at point of purchase
State Checks	State conducts separate background checks, beyond the federal NICS check, for all firearms
Inpatient Commitment	Firearm possession is prohibited for people who have been involuntarily committed to an inpatient facility
Outpation Commitment	Firearm possession is prohibited for people who have been involuntarily committed to an outpatient facility
Danger	Firearm possession is prohibited if person is deemed by court to be a danger to oneself or others
Lost/Stolen	Mandatory reporting of lost and stolen guns by firearm owners
Concealed Carry Background Check	Concealed carry permit renewal requires a new background check
Domestic Violence Misdemeanor	People convicted of a misdemeanor crime of domestic violence against a spouse, ex-spouse, or cohabitating partner are prohibited from possessing firearms
Domestic Violence Restraining Order	State law automatically prohibits domestic violence-related restraining order (DVRO) subjects from possessing firearms
Stalking	A stalking conviction is prohibitive for firearm possession

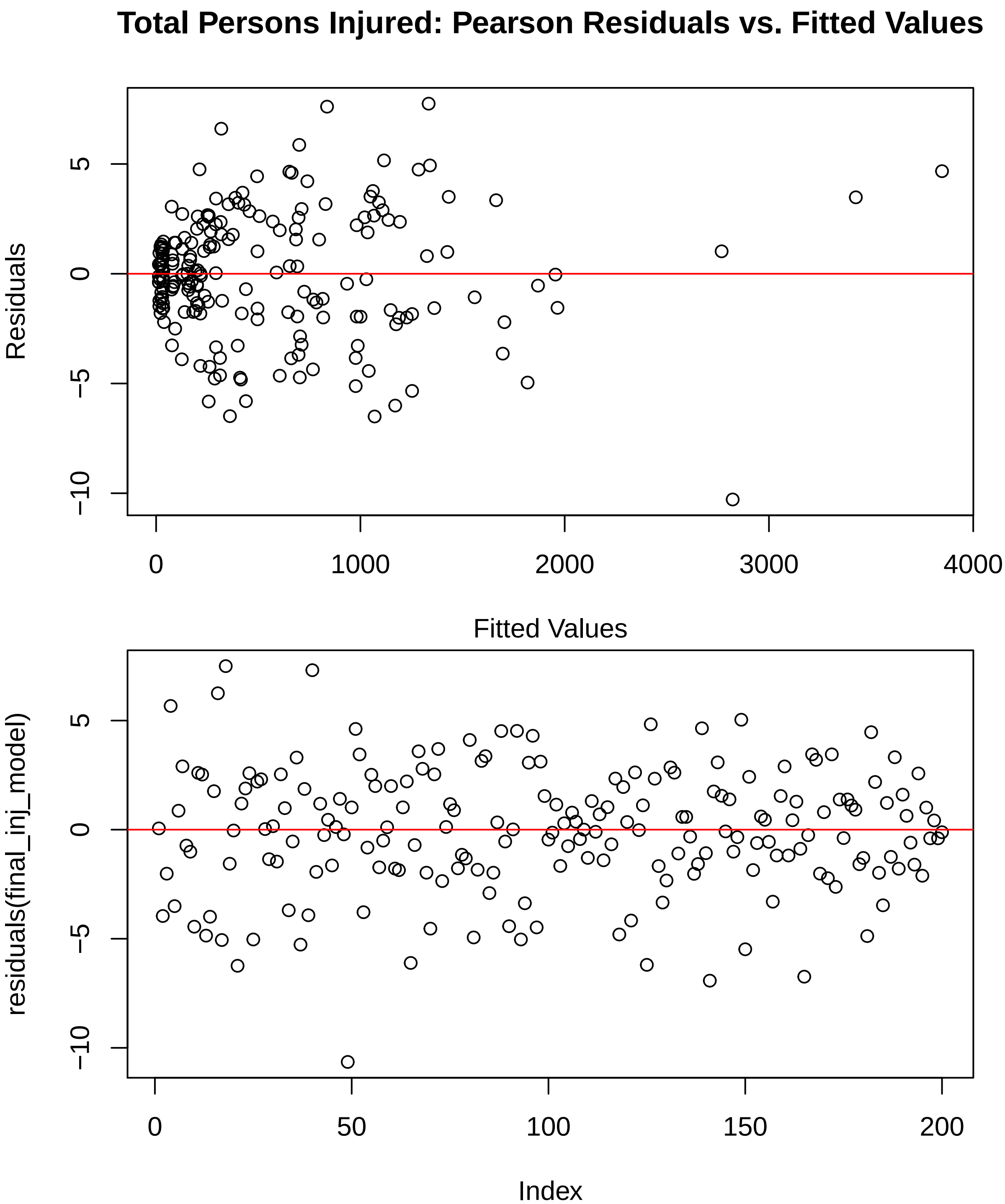
Total Shootings Residuals

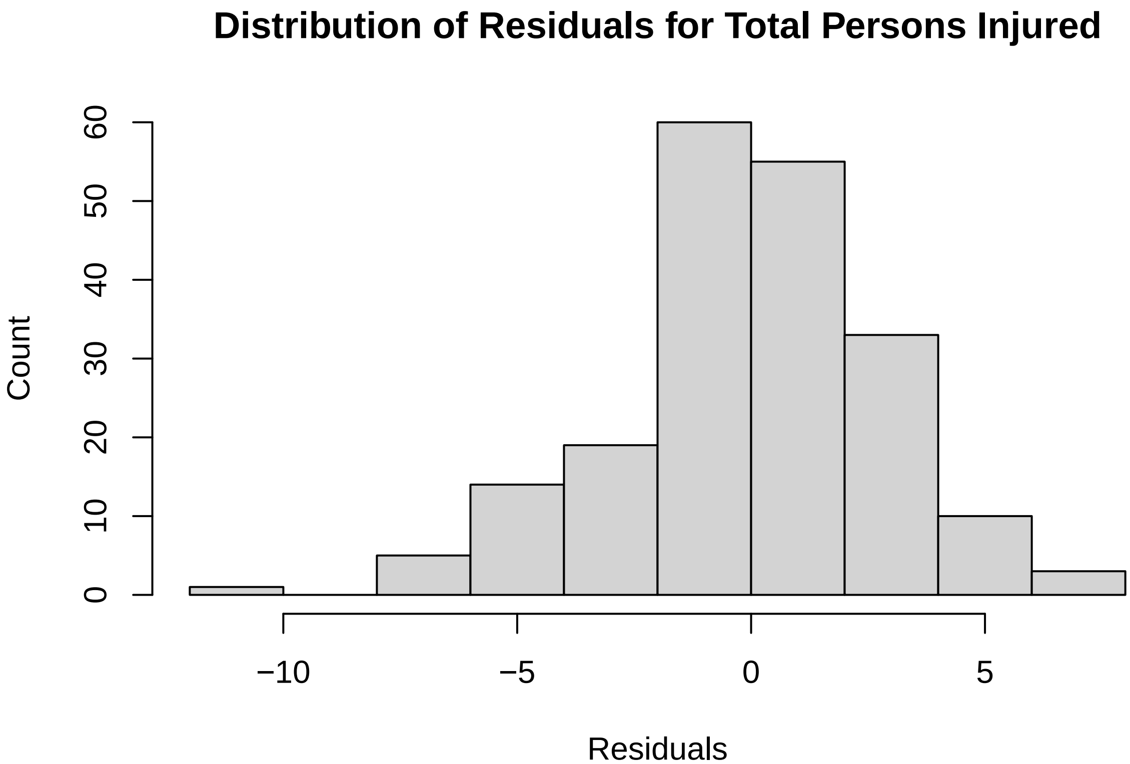


Injuries Residuals



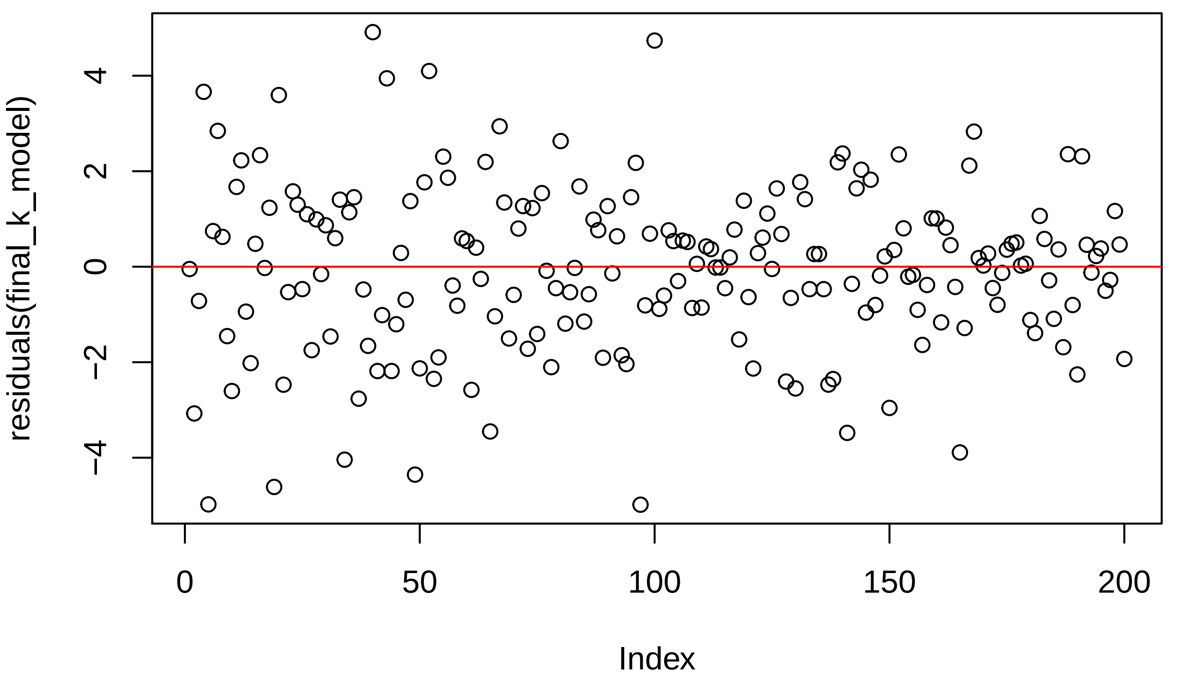
Total Injured Residuals

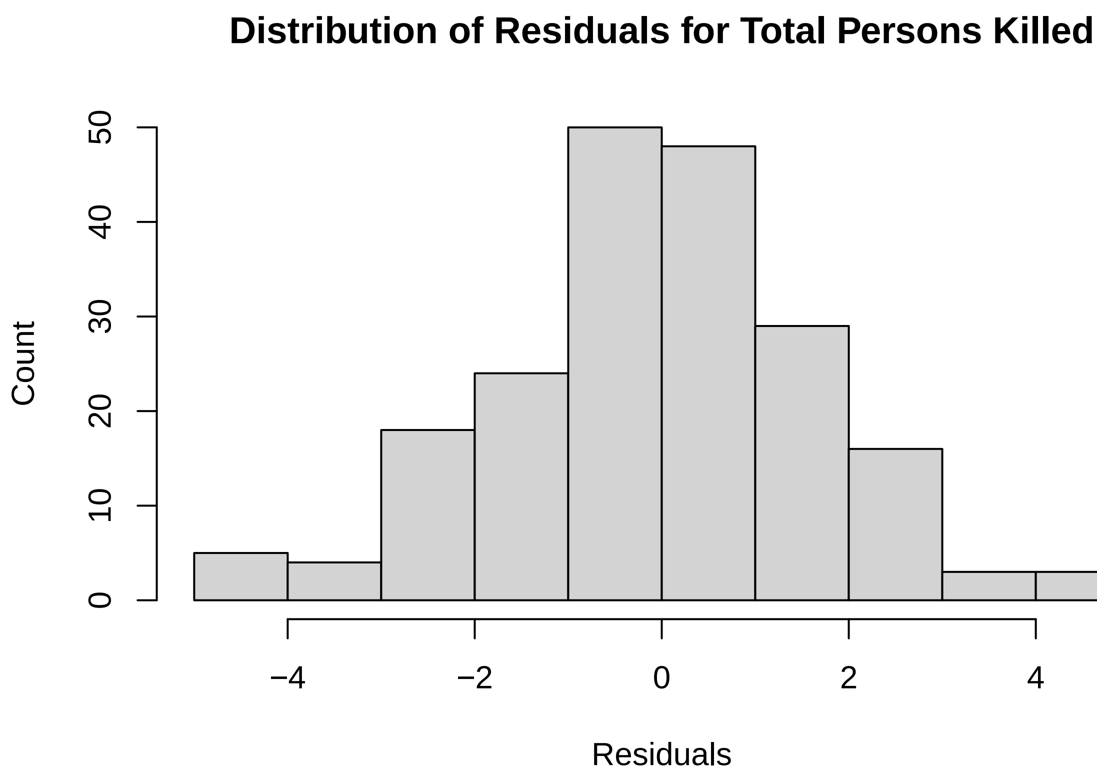




Distribution of Residuals for Total Persons Injured
Count
0
10
20
30
40
50
60
−10
−5
0
5
Residuals

Total Killed Residuals




Distribution of Residuals for Total Persons Killed
Count
0
10
20
30
40
50
−4
−2
0
2
4
Residuals