Significance Testing and Propensity-Score Analysis to the Effects of Gun Regulation on Rates of Gun-Related Injury

Abstract

Background

In our highly politicized society there are arguments about gun violence: about if it is responsible or not to increase regulations. On both sides of the argument, people claim that their ideal system is the one that results in less harm. To a layperson, either argument may sound appealing: that it is good to take guns out of the hands of criminals or that it serves as an inherent deterrent to have guns more easily available. In order to test if there is validity on either side of the debate, an investigation was held on known data about how regulation interacts with the frequency and injury rates of gun incident.

Methods

Using a database of recorded gun incidents from the years 2013 to 2018, two separate questions were tested. First, on a statewide basis, it was tested if there is a significant difference in the effect of regulation on the likelihood of an injury, given a gun-related incident had occurred. Next, it was tested whether a state's level of regulation would alter the frequency of incidents occurring within the state.

Results

Using 218,748 incidents in 11,148 cities as a baseline and a significance level of 0.05, it was analyzed that there is not a significant difference in the chance of injuries once an incident has occurred. Also, increasing regulation did not display a clear association with increasing or decreasing gun violence.

Conclusions

Based on the results of this study, it may be concluded that a state's level of gun regulation does not directly have a significant impact on the harm that firearms cause in the state borders. It was also concluded that the regulation level was not a significant factor in predicting the rate of incidents occurring. The most relevant predictor of the rate of gun violence is the population level.

Introduction

Description of Data Set

The data set was acquired via Kaggle.com, using a database of all gun incidents that were recorded from January 2013 to March 2018. It was acquired by the data set's creator using a program to get a comprehensive record based on reports from the Gun Violence Archive. Some incidents ended up not being recorded in the data set due to issues they made in the code or to being unable to fit in the scraping methods (such as being recorded in a pdf rather than text). It still had created a sample of approximately 240 thousand incidents to work with. After cleaning the data, the total number of incidents that are usable in the set are 218,748, representing 11,148 cities across the nation. Removing some areas of the data may bias results, particularly in how areas with extremely low population density are completely disregarded at some threshold. Also the data set cannot account for cities that have no gun incidents as there is nothing to report, but there are enough observations to perform analytics and for the purposes of the study, areas that do not have reported incidents are not necessarily of interest.

Of the data set, we aim to study a few potential predictors of gun violence. In the data set, the study examined if an injury occurred, if a gun was stolen, and if a mass shooting occurred (defined as four or more injuries or deaths). Additionally, some extra variables were added: the city's population (used from a separate data set obtained through SimpleMaps.com, aggregating population from census data), a Frequency variable (by city) that counts the total number of incidents that appear in the data set, and an arbitrary factor of the state's Gun Regulation level. The regulation factor was determined by studying the number and severity of statewide gun control laws and assigning the state a level from 1 to 5. The research was done via estimates of a state's gun regulation (Gifford's Law Center, "Gun Laws by State."). Additionally, in some analyses, regulation was split between low and high regulation, where high regulation states were defined as any state with a Regulation level of 3 or higher.

The database used took reports from the Gun Violence Archive, which collects data about incidents from over 2,500 law enforcement, government, corporate and media sources. Two assumptions are needed in order for the analysis of this data to be meaningful. As there are a portion of incidents that do not get reported, this study assumes that the rates of violence in non-reported incidents closely resemble the rates within reported incidents. Thus, missing data does not bias the obtained results.

A second assumption is that all gun incidents are viewed without considering whether or not defensive gun use (DGU) was involved. This is because the rates of DGU are widely debated and many studies result in dramatically different conclusions on the exact rate.

The National Crime Victimization Survey argues that less than one percent of all gun incidents involve DGU, regardless of what the victimized population is (David Hemenway, “The epidemiology of self-defense gun use: Evidence from the National Crime Victimization Surveys 2007–2011“, 2013). On the opposition, Gary Kleck, the most famous advocate in favor of DGU, has a 2018 paper wherein he states that research from the Center of Disease Control argue in favor of his surveys, where he concludes that there is a significantly higher number of annually unreported DGU cases than reported gun incidents (Gary Kleck, “What Do CDC's Surveys Say About the Frequency of Defensive Gun Uses?”, 2018).

Both of these are different estimates from what the data set reports, as the Gun Violence Archive’s data shows that annually there is an approximate 3% of reported incidents that involve DGU of some type. Due to the difficulty in estimating the exact proportion of DGU, especially in regards to non-reported incidents, this study simply aims to view all gun incidents as occurring under the same circumstances. This study will not treat that aspect of gun violence and will instead focus on the question of reducing gun violence by reducing the total number of incidents.

Statistical Analysis

In order to determine firstly if gun regulations had some sort of effect on the harm caused by crime (to analyze regulations against certain types of firearms on the basis that it will reduce the ability of violent criminals to harm innocents), mosaic plots were made comparing potential predictors of injury and an analysis of the Pearson residuals was conducted. A small Pearson residual (having a magnitude lower than 2) is a strong indicator that the two analyzed factors are independent. A plot was made that shows the interaction of regulation and at least one injury (be it a fatality or simply anything classified as an injury).

The aim of this work was to predict if gun regulation provides significant support for keeping more dangerous weapons out of the hands of criminals and preventing harm once an incident has occurred. This is why analyzing the harm rates related to stolen firearms is important, as an illegally acquired firearm suggests that regulation may not be having the intended effect on the presence of weapons in an area. While mass shootings are common and often garner a large amount of attention, the overwhelming majority of incidents only have 0 or 1 victims harmed. As such, a clearer picture is drawn from analyzing whether or not any injury has occurred.

Next, the aim was to analyze the total number of reported incidents and the effect regulation had on them. In this specific area, it was important to additionally consider population. There is a slight correlation between states with high regulation and states having a high population (a coefficient of approximately 0.33, indicating a weak relationship). This agrees well with common sense rationale that states with higher population have more reported incidents, so the aim was to determine if regulation has an effect once population is controlled for.

For the initial analysis, a t-test with an alpha level of 0.05 was performed to check if there was any difference in the frequency of gun violence between the low regulation states vs the high regulation states.

T-tests

2-tailed

```
data:  lowreg2$freq and hireg2$freq
t = -0.4426, df = 1118.4, p-value = 0.6581
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.751140   8.057295
sample estimates:
mean of x mean of y
 19.68195  22.02887
```

Low > High

```
data:  lowreg2$freq and hireg2$freq
t = -0.4426, df = 1118.4, p-value = 0.6709
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -11.0762      Inf
sample estimates:
mean of x mean of y
 19.68195  22.02887
```

High > Low

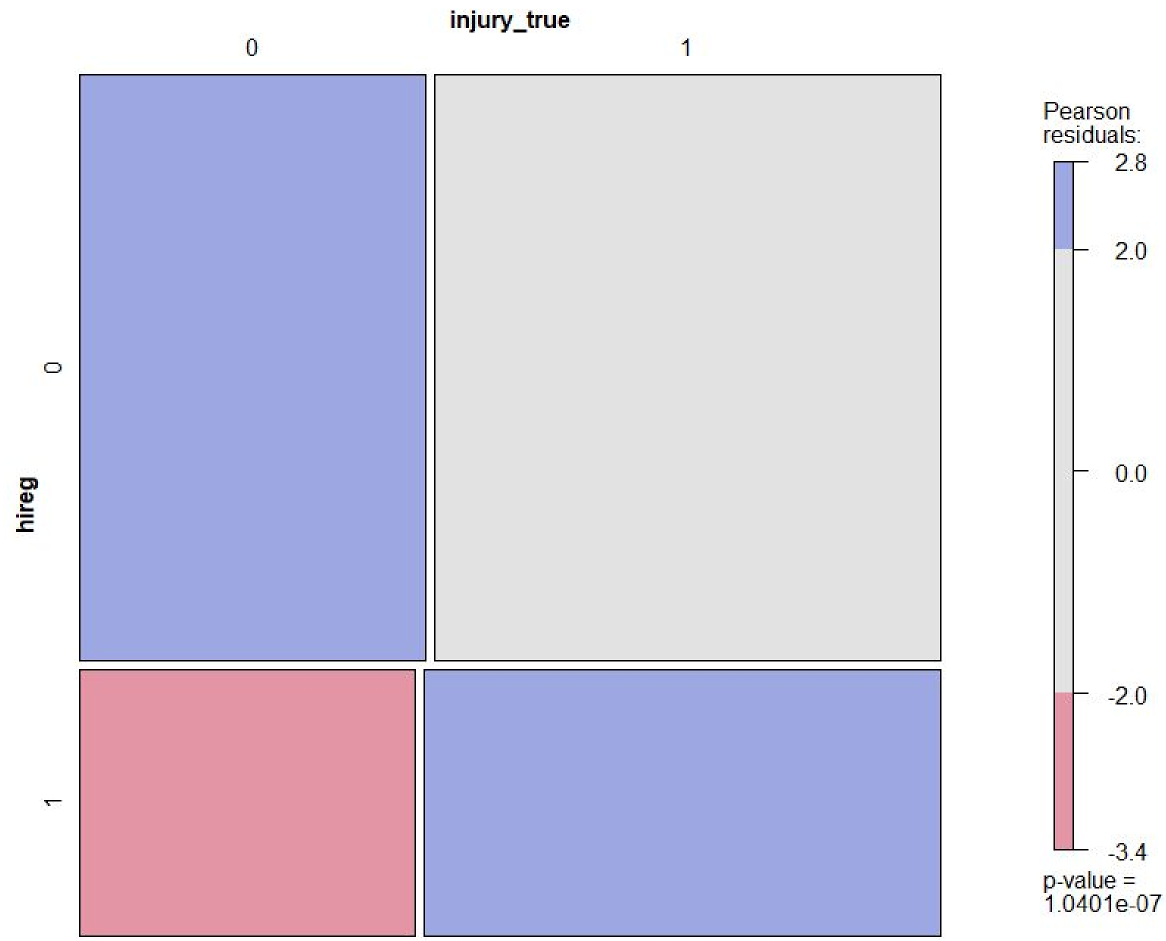
```
data:  lowreg2$freq and hireg2$freq
t = -0.4426, df = 1118.4, p-value = 0.3291
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 6.382354
sample estimates:
mean of x mean of y
 19.68195  22.02887
```

Next, a linear model was made, regressing number of gun incidents against regulation level and population, both with and without interactions between regulation and population. The intent with the linear model was to test if regulation has a low enough p-value to be considered an important predictor or if population was the only relevant one, especially once their interactions were accounted for.

Results

Results of Harm Rate Analysis

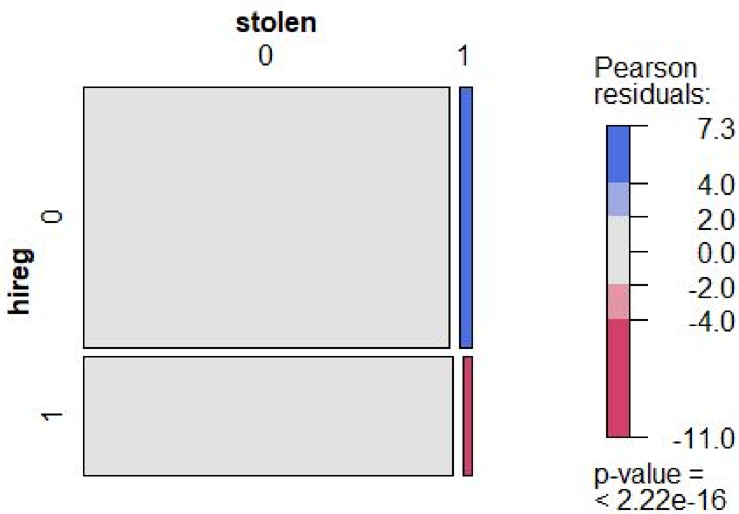
The mosaic plots argue that for the most part there is not a significant link between regulation level and the incidence of a gun crime leading to injury. The plots make it appear that there is a higher total amount of gun crime in areas with low regulation, but an important piece of information to keep in mind is that the majority of the cities are not marked as high regulation. In brief, the reason the frequency of gun crimes in low regulation areas is approximately twice as high as in high regulation has little to do with the total rates and more to do with simply how many cities are in low regulation areas. This introduces one of the consistent issues within the analysis: that there are numerous confounding factors that make it difficult to draw a concrete conclusion.

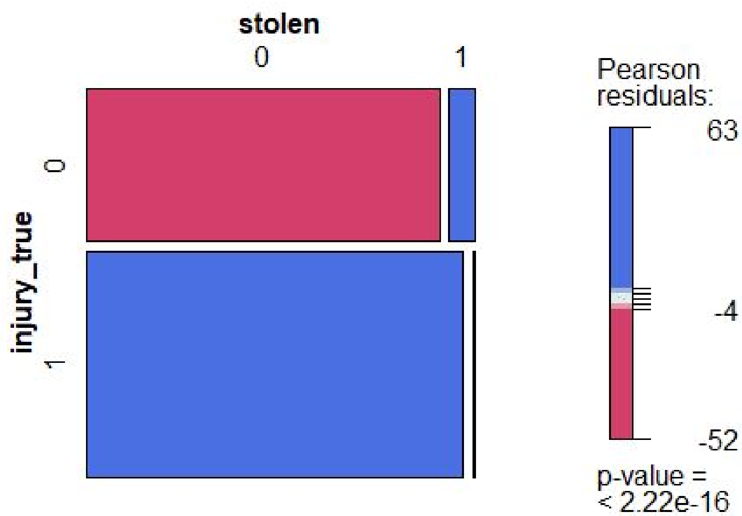


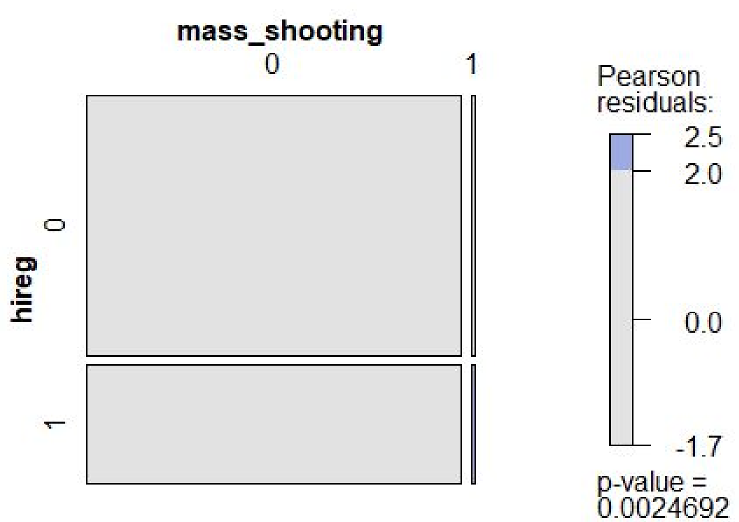
The Pearson residuals suggest that once a crime has occurred, there is a slightly higher chance that injury may occur in a city with high regulation. However, the rate of injuries themselves are not significantly different. There are a variety of possible explanations behind this, since the mosaic plots do not account for things such as the population.

As for the other predictors, it is not observable that the regulation level has an effect on important predictors. There is a higher probability of a firearm being stolen in a low regulated state, as well as a lower probability of an incident involving a stolen firearm resulting in an injury occurring. Whether it may be related to an assailant having their weapon taken from them, or a self-injury using a weapon owned by a relative or friend, the illegal procurement of weapons is a very broad term such that it is difficult to draw a conclusion on an incident simply if it involves a stolen weapon. There is potential confusion in this manner, as the definition of a stolen firearm is not clear. Additionally, much of the data does not have knowledge on if a weapon was stolen or not, so how one interprets the unknown firearms makes it difficult to draw a strong link between stolen weapons and harm rates.

Mass shootings are very uncommon as a whole, so there may not be adequate data in the set to draw off of, but given the available data, the mosaic analysis shows there may be a slightly higher correlation of mass shootings with high regulation. Again, possibly because there are confounding factors (the population, mainly), it gives a very weak Pearson residual, and even if there is a correlation it appears that it is strictly because there are slightly fewer incidents in high regulation states - that is to say, the percentages of shootings that are classed as mass shootings is not noticeably different between regulation levels.



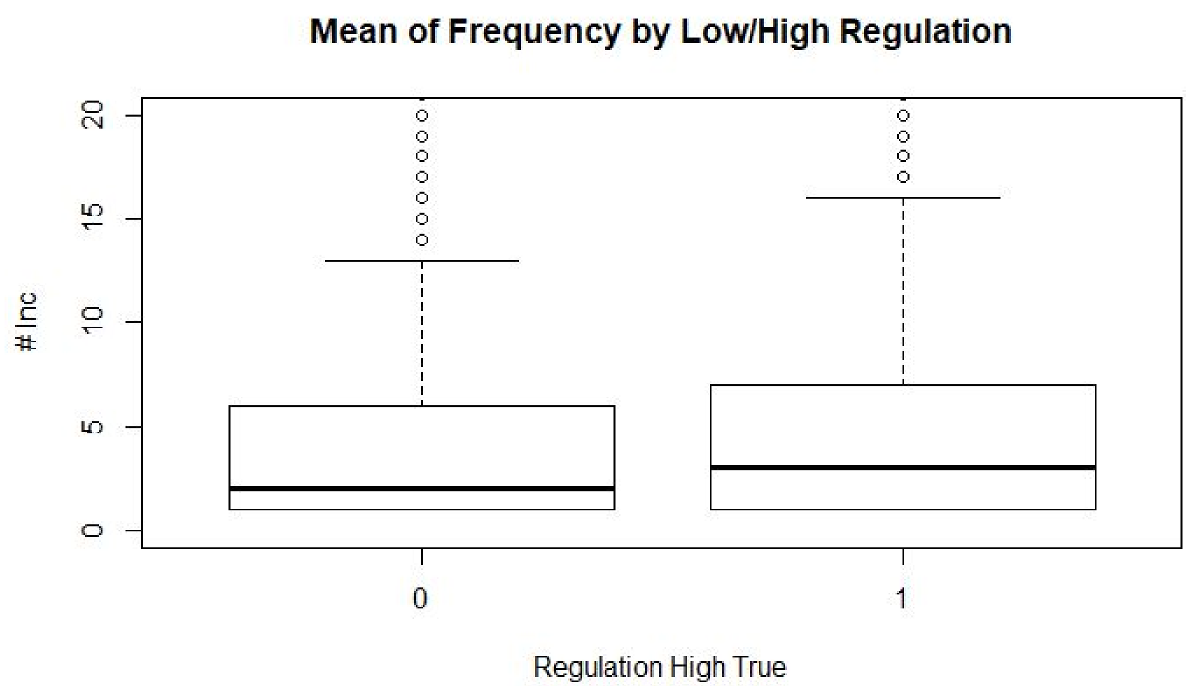


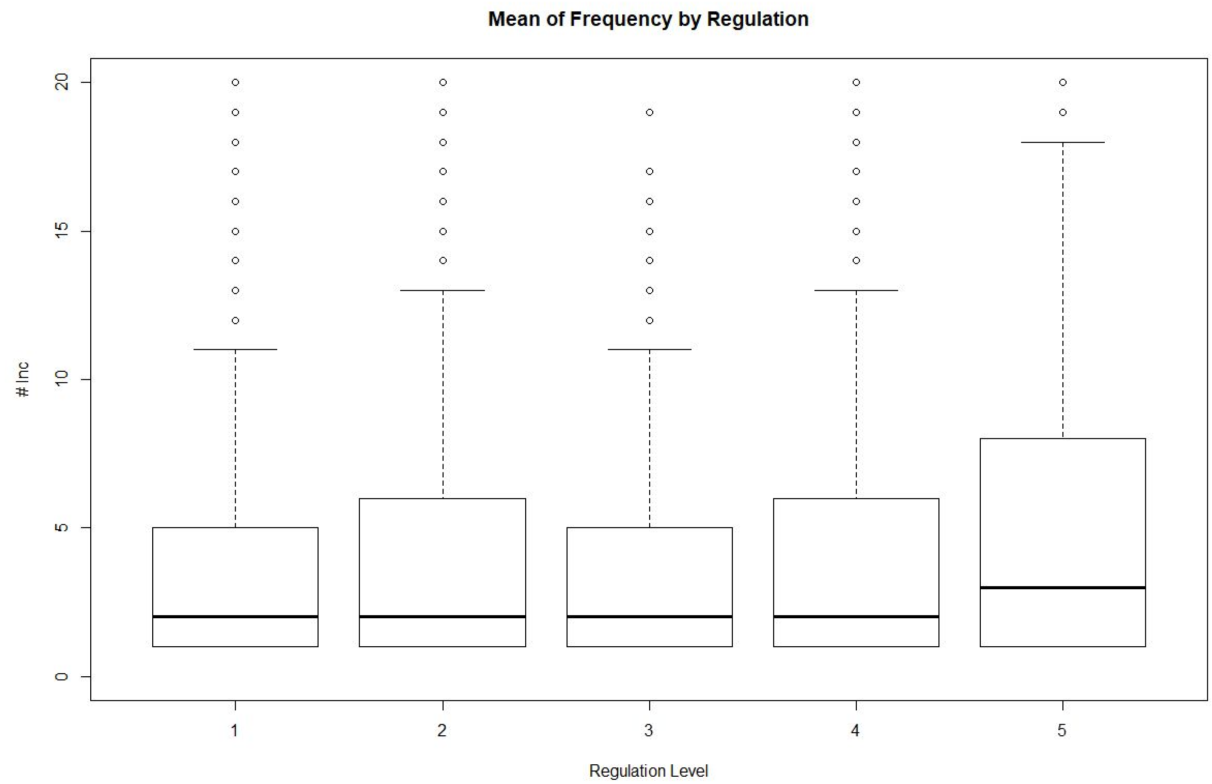


Regardless of these individual comparisons, it is ultimately concluded that there may be evidence suggesting a slightly higher chance of an injury occurring within a state with high regulation. Be that as it may, the difference is very small and the overall rate of injury does not end up significantly different based on regulation levels. While this is something to keep in mind, what is more significant to analyze will be the effect of regulation on the rate of incidents occurring.

Results of Frequency Analysis

The result of the t-test returned a two-sided p-value of 0.6581. This does not support that there is a difference in the frequency of gun incidents based solely on the level of regulation. To gain easier visualization, box plots of the means also were created to measure frequency as it related to regulation. Of the following plots, a large number of outlier incidents were present on all regulation levels. This is because many areas have very low numbers of gun incidents such that their means are very low (under 5 incidents per city on each regulation level), but there are so many cities with a high number of incidents that outliers are common.





As the box plots shown suggest, it does not appear that there is any significant effect on the rates of gun crime based on regulation, although there is a seemingly slight increase in some of the higher regulation areas. This may be due to there being a strong correlation between the regulation level and the population density, which the t-test and box plots do not account for, so further testing was performed.

While a variety of linear models were tested, they all returned somewhat inconclusive results. Taken marginally, the regulation level (1,2,3,4, or 5) appears to be irrelevant to number of incidents, but when interacted with population it displayed an effect. If testing for the regulation level as a numerical variable, the indication was that there was a strictly negative linear relationship based on the interaction. By the glms shown, the regulation does not seem to have any direct effect, but there is a slight indication that some regulation levels may slow the effect of gun crime occurrence at higher population levels.

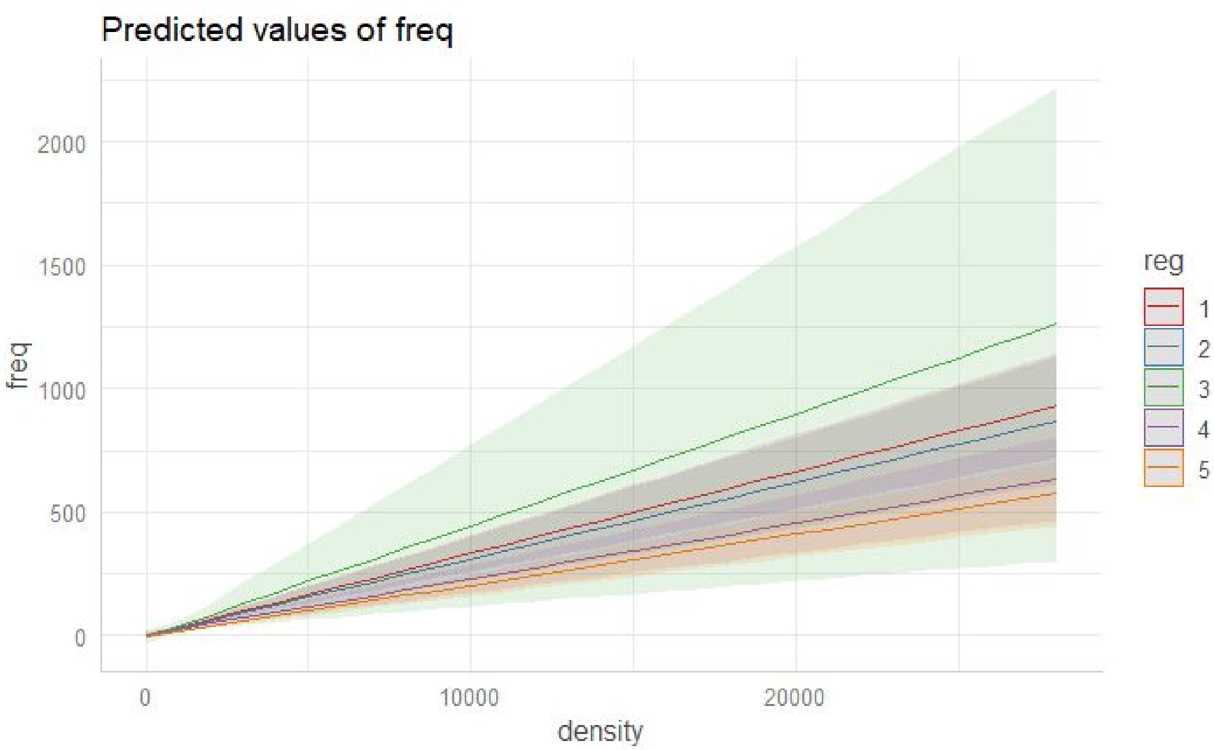
Regulation as numeric

```
Call:
glm(formula = freq ~ as.numeric(reg) * density, data = data2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
 -492.4    -17.4     -8.5     -2.4  10705.7

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.411365   3.510856   0.402  0.68769
as.numeric(reg)         -0.480143   1.340606  -0.358  0.72024
density                  0.037302   0.004210   8.859  < 2e-16 ***
as.numeric(reg):density -0.003426   0.001056  -3.245  0.00118 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 23726.93)

    Null deviance: 270158395  on 11147  degrees of freedom
Residual deviance: 264412941  on 11144  degrees of freedom
AIC: 143952

Number of Fisher Scoring iterations: 2
```

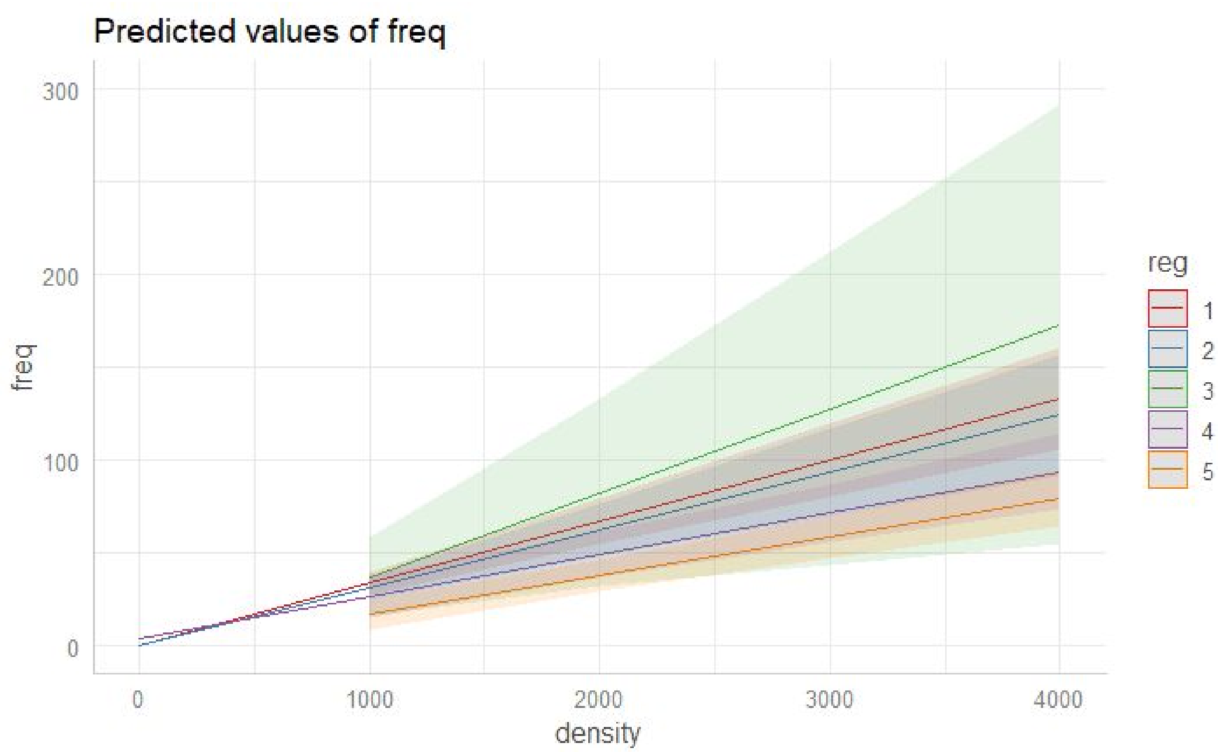
Regulation as categorical

```
Call:
glm(formula = freq ~ reg * density, data = data2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
 -474.0    -17.7     -8.6     -2.2  10706.0

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.851543   2.876131   0.296  0.76718
reg2         -0.294749   4.795707  -0.061  0.95099
reg3         -8.536408  14.804807  -0.577  0.56422
reg4          3.356831   7.546666   0.445  0.65647
reg5         -4.333677   5.910653  -0.733  0.46345
density       0.033133   0.003940   8.410  < 2e-16 ***
reg2:density -0.002126   0.006213  -0.342  0.73216
reg3:density  0.012154   0.018284   0.665  0.50624
reg4:density -0.010634   0.005078  -2.094  0.03626 *
reg5:density -0.012443   0.004675  -2.662  0.00779 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 23735.26)

    Null deviance: 270158395  on 11147  degrees of freedom
Residual deviance: 264363328  on 11138  degrees of freedom
AIC: 143962

Number of Fisher Scoring iterations: 2
```

Below is a plot of predicted frequency of gun crimes based on the population, at each of the five regulation levels (as categories) with 95% confidence intervals displayed as color-coded shading on the graph.

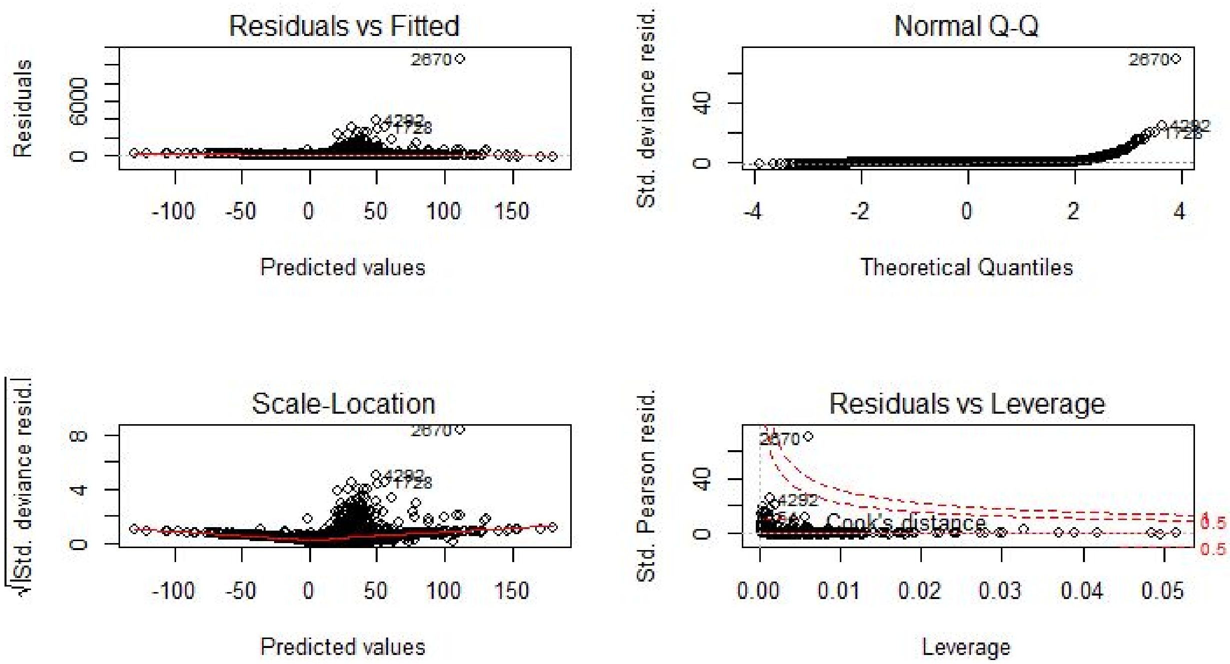


This first prediction model makes it appear that there is a linear connection between regulation level and the frequency of incidents (with regulation 3 being uncertain since there are fewer observations within it). However, the model is misleading because it covers all population densities but most are in the bottom most third of the graph. Note the density approaches nearly 30,000 at the end of the graph. Within the data set, the median population density is 437 and any population higher than 1343.45 is an outlier population. In fact, of the low regulation cities, the only one that has a higher than 10,000 population density is Arlington, Virginia at 11892.6. The following graph is the same as the one above but zoomed in such that it ends just after the cutoff for population density outliers.

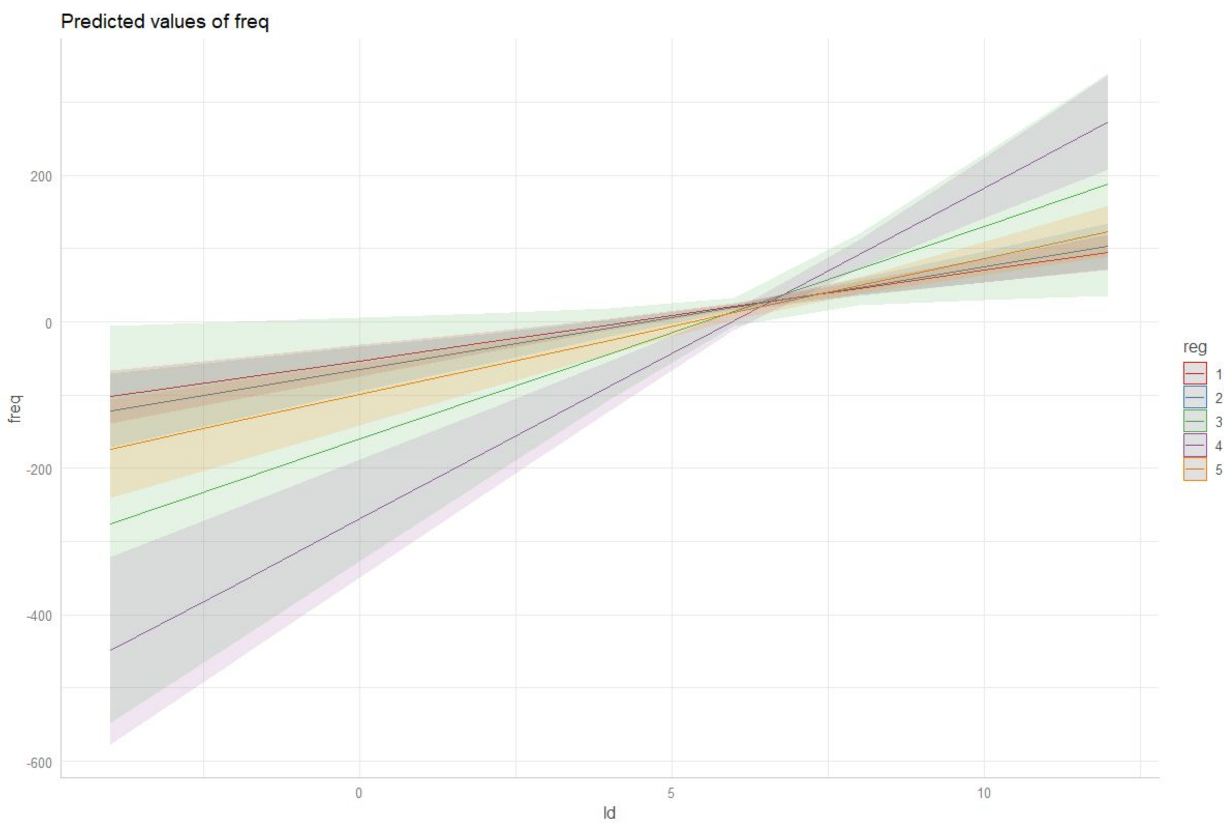


What this model shows is that as we examine the majority of the population, the rates of gun violence are very consistent regardless of regulation. There may be slight evidence in rates of crime being influenced by regulation as the population increases past the national outlier point. That said, there is not sufficient evidence to argue there is a direct relationship between regulation and frequency.

To help control for these complications that the population density has on the model, the data was manipulated once more, taking the natural log of the population to create a new model where the logged density was plotted against the frequency. This model showed that only the regulation levels 1 and 4 were significant, but that was including the severe outlier of Chicago, which itself is the only point within the residuals that is shown as being extreme enough to unbalance the data.



Residuals vs Fitted
Residuals
Predicted values
Normal Q-Q
Std. deviance resid.
Theoretical Quantiles
Scale-Location
Predicted values
Residuals vs Leverage
Std. Pearson resid.
Cook's distance
Leverage



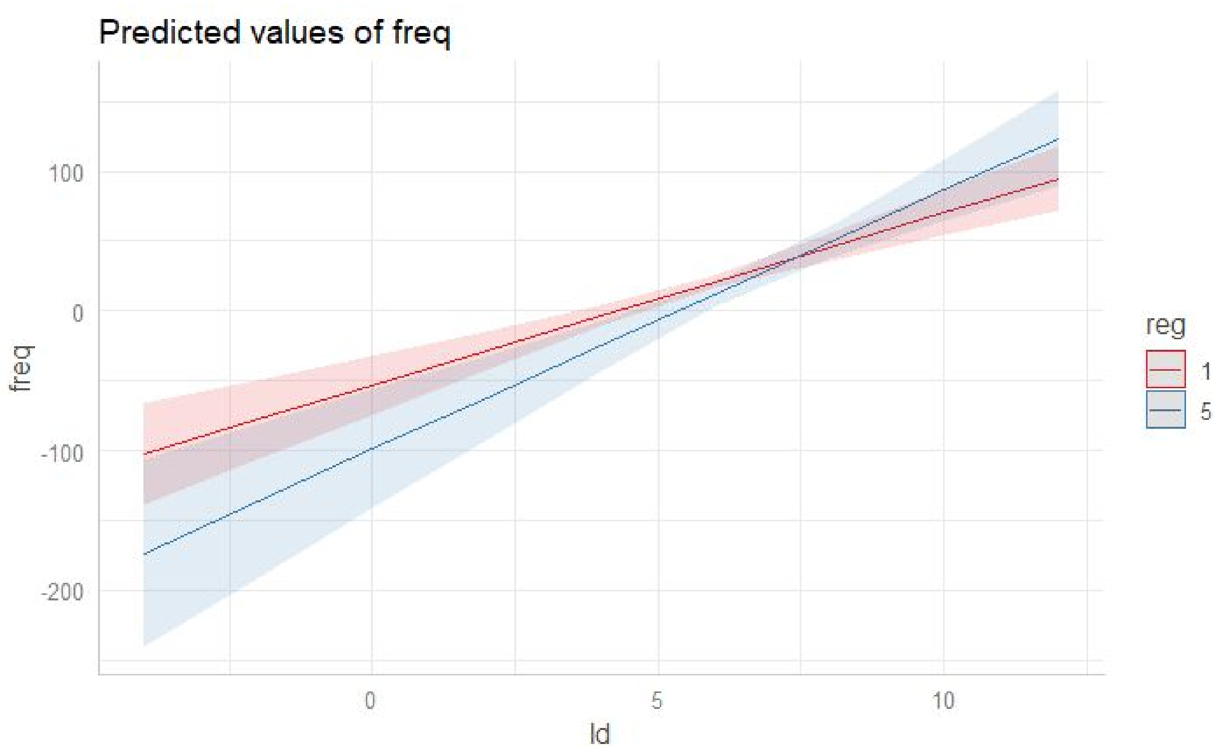
Predicted values of freq
freq
ld
reg
1
2
3
4
5

Due to it being a singular point that skews the analysis, a final model was made that removed Chicago from the data set, and this one resulted in a glm that takes regulation levels 1 and 5 as having significant effects at the 0.05 level. As such, to get a better look at the difference between those effects, the two, the last plot only examines regulation levels 1 and 5.

```
Call:
glm(formula = freq ~ reg * ld, data = dataln)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
 -176.3    -23.2    -14.2     -2.5  10702.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -53.075     10.863  -4.886 1.04e-06 ***
reg2         -11.894     18.912  -0.629   0.5294
reg3        -107.402     85.482  -1.256   0.2090
reg4        -215.587     42.558  -5.066 4.13e-07 ***
reg5         -46.159     24.306  -1.899   0.0576 .
ld            12.340      1.874   6.586 4.73e-11 ***
reg2:ld        1.708      3.180   0.537   0.5911
reg3:ld       16.678     13.580   1.228   0.2194
reg4:ld       32.761      6.382   5.134 2.89e-07 ***
reg5:ld        6.226      3.719   1.674   0.0941 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 23898.51)

    Null deviance: 270156015  on 11139  degrees of freedom
Residual deviance: 265990391  on 11130  degrees of freedom
AIC: 143935

Number of Fisher Scoring iterations: 2
```

Remove Chicago

```
Call:
glm(formula = freq ~ reg * ld, data = dataln2)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
 -74.0   -22.5   -13.8    -2.6  3894.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -53.075      8.178  -6.490 8.95e-11 ***
reg2         -11.894     14.238  -0.835   0.4035
reg3        -107.402     64.355  -1.669   0.0952 .
reg4         -33.475     32.100  -1.043   0.2970
reg5         -46.159     18.299  -2.522   0.0117 *
ld            12.340      1.411   8.748  < 2e-16 ***
reg2:ld        1.708      2.394   0.714   0.4754
reg3:ld       16.678     10.224   1.631   0.1028
reg4:ld        3.302      4.815   0.686   0.4929
reg5:ld        6.226      2.800   2.224   0.0262 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 13545.07)

    Null deviance: 153627219  on 11138  degrees of freedom
Residual deviance: 150743106  on 11129  degrees of freedom
AIC: 137597

Number of Fisher Scoring iterations: 2
```



This plot's results are notable, as they do not seem to show a difference at any population level between the rates of gun crime and the regulation, be it minimal or maximized. Between log density rates of 2.5 and 5, there is a minute disconnect that shows at the 95% confidence level, those two rates have a significant difference.

Conclusions

The data ultimately indicates that the regulation level alone is not a significant predictor of a city's level of gun violence. In terms of both the frequency of incidents and the harm those incidents cause, the study presents inadequate reason to believe that regulation levels are a key predictor. The more important factors are probably ones that would be more difficult to predict. For instance, since regulation alone did not adequately predict the rates of harm caused by firearms, it would be more important to look at public opinion on guns or the prevalence of firearm smuggling. Most importantly, it examines the regulation levels as broad categories for simplicity when simply "more regulation" may not be as significant since it also does not answer which possible regulations have an effect on crime rates. These could have stronger effects on the violent crime rates independent of regulation and population.

To examine a means to reduce gun violence, a follow-up to this would be to more closely examine population-dense cities more deeply. To check for instance which regulations each city imposes, whether gun crimes break those laws or not, as well as possibly to check if nearby areas where those regulations are not present exist (i.e. if someone traveled out of state to acquire a type of firearm that is illegal in their state). Those sort of questions would be a way to narrow down which regulations are effective or ineffective. Also if they are ineffective, if it is due to one's ability to avoid the regulation.

Returning to the question: is it possible to reduce the rate of harm caused by gun crime simply by increasing the regulations on firearms? At most, the data analysis supports a hypothesis that regulation has no significant effect on the rates of gun violence.

Sources

Gifford's Law Center. "Gun Laws by State." Gifford's Law Center.
https://lawcenter.giffords.org/browse-gun-laws/ (accessed February 9, 2019)

Simplemaps.com. "US Cities Database." SimpleMaps.com.
https://simplemaps.com/data/us-cities (accessed July 22, 2019)

Hemenway, David. "The epidemiology of self-defense gun use: Evidence from the National Crime Victimization Surveys 2007–2011" ScienceDirect.com.

https://www.sciencedirect.com/science/article/pii/S0091743515001188 (accessed May 11, 2019)

Kleck, Gary. "What Do CDC's Surveys Say About the Frequency of Defensive Gun Uses?" SSRN.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3194685 (accessed May 11, 2019)