

Midterm Project

Alexander Hernandez

10/13/2022

Introduction

In the United States, politics have become more divisive than ever. With access to guns at the forefront of public interest with the ever present risk of shootings in public places, the country has become just as split in regard to the policies in place that manage firearms. With this general difference in opinion on guns between Democrats and Republicans, it comes in to question whether the liberal requirements and restrictions on guns has an effect on shootings in more conservative areas. With general trends in political ideologies, a divide may exist between areas that vote for one side or the other. That leaves two clear questions:

- Does the electoral voting result of a state (i.e. Democrat vs. Republican state) result in more instances of mass shootings?
- Does the electoral voting result of a state results in more victims of a mass shooting?
 - Null Hypothesis: There is no difference in victims of mass shootings between red and blue states.
 - Alternative Hypothesis: There is a difference in victims of mass shootings between red and blue states.

With these questions and hypotheses in mind, an investigation into the states (separated by 2020 electoral voting result) mass shootings and their number of deaths can be investigated.

Methods:

1. Import data, filter out District of Columbia, and create 'total' column

```
mass = read.table("C:\\repos\\STAT 50001\\Midterm Project\\Mass shooting data.csv",
                  sep=";", header=TRUE)
mass = mass[mass$State != "District of Columbia",]
mass$total = mass$X..Killed + mass$X..Injured
```

2. Import 'states' loop-up table to 2020 electoral results based on blue or red

```
states = read.table("C:\\repos\\STAT 50001\\Midterm Project\\Red-Blue.csv",
                    sep=";", header=TRUE,
                    fileEncoding="UTF-8-BOM")
```

3. Align incidents of shootings based on state's political alignment in df

```
alignment_match = mass$State
alignment_match[] = states$Political_Alignment[match(mass$State, states$State)]
mass$alignment = alignment_match
```

4. Create pie chart to demonstrate number of mass shooting differences

```
red_amt = nrow(mass[mass$alignment=="Red",])
blue_amt = nrow(mass[mass$alignment=="Blue",])
total_amt = red_amt + blue_amt
red_label = paste("Republican: ", red_amt, " - ",
                  round(100*red_amt/total_amt, 2), "%")
blue_label = paste("Democrat: ", blue_amt, " - ",
                  round(100*blue_amt/total_amt, 2), "%")
pie_labels = c(red_label, blue_label)
pie_data = c(nrow(mass[mass$alignment=="Red",]), nrow(mass[mass$alignment=="Blue",]))
```

5. Run t-test for difference in victims for democrat and republican states

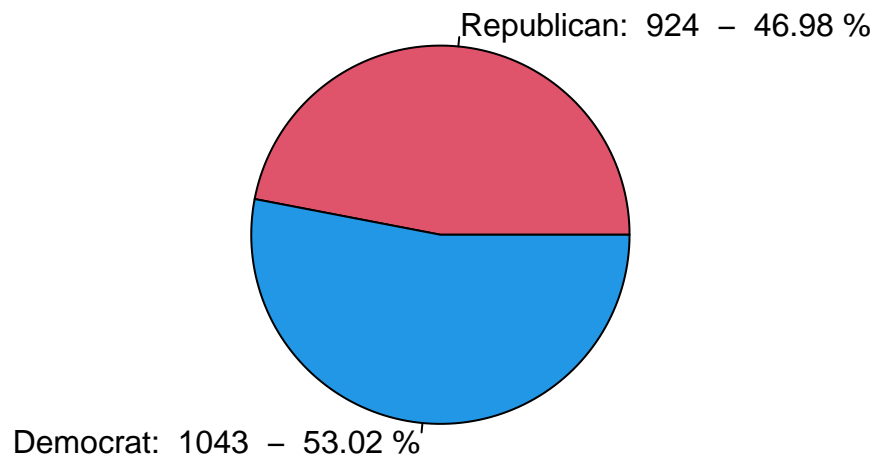
```
new_data = mass[mass$Incident.ID != 946496,]
box_data = log(new_data$total) ~ new_data$alignment
test_data = mass$total ~ mass$alignment
```

Results

Figure 1:

```
pie(pie_data, labels = pie_labels, col = c(2,4),  
    main = "Number of Mass Shootings by Republican and Democrat States")
```

Number of Mass Shootings by Republican and Democrat States

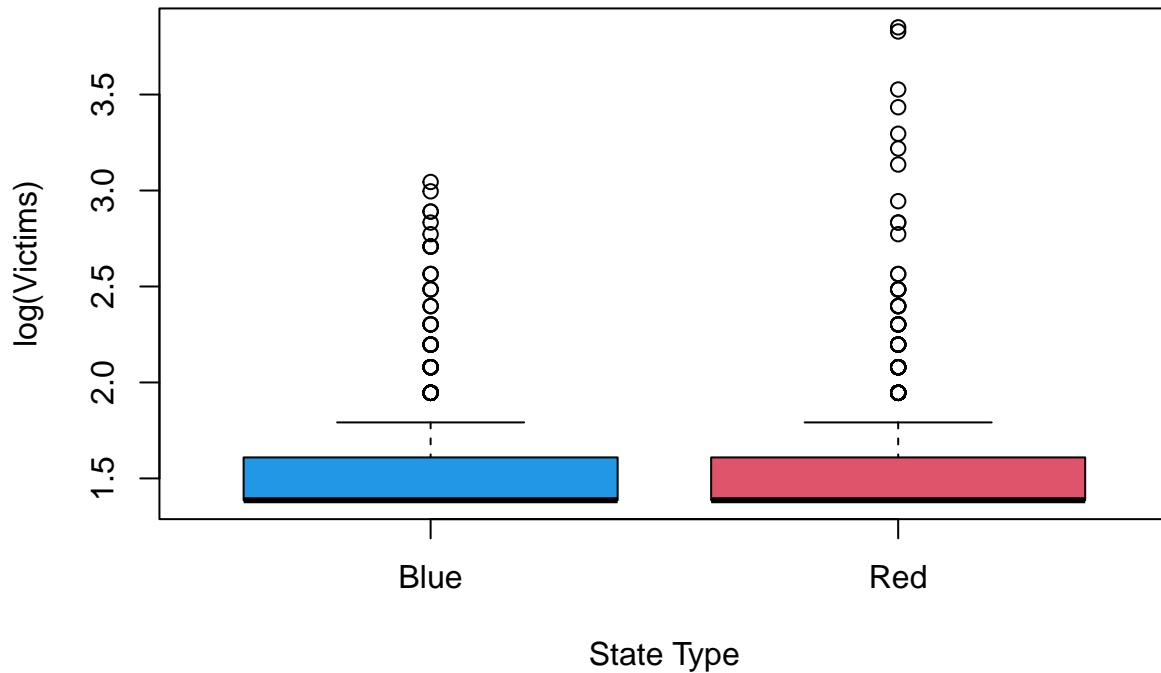


As displayed in this pie chart, the number of mass shootings differs by 6.02% with 924 Republican shootings and 1043 democrat shootings.

Figure 2:

```
boxplot(box_data, col=c(4,2), xlab="State Type", ylab="log(Victims)",  
        main = "Mass Shootings Separated by Blue and Red States")
```

Mass Shootings Separated by Blue and Red States



The box-plot above demonstrates that the vast majority of shootings have between 1-5 victims, or less than $\log(5)$. Notable exceptions are within the outliers, mostly reaching into the 50s but one such example showed itself at 441 victims under Blue (ID: 946496, removed for boxplot visibility).

Figure 3:

```
t.test(test_data)
```

```
##
##  Welch Two Sample t-test
##
## data:  mass$total by mass$alignment
## t = 0.49644, df = 1133.1, p-value = 0.6197
## alternative hypothesis: true difference in means between group Blue and group Red is not equal to 0
## 95 percent confidence interval:
##  -0.7156216  1.2004120
## sample estimates:
## mean in group Blue  mean in group Red
##          5.437200          5.194805
```

From the t-test, a few data points are learned, namely the mean victims of the shootings for Republicans is 5.19 victims while the mean for Democrats is 5.44 victims. With a confidence level of 0.95 (calculated as -0.72 to 1.20), this two-sided t-test reveals a p-value of 0.6197. Given that the p-value must be above 0.05 to reject the null hypothesis, this value is far above what is required for us to claim enough evidence to reject the null that there is no difference between mass shootings in Republican and Democrat states (of which there clearly is).

Discussion

Far from expected when taking a front-facing look at ideologies, the data seems to have been flipped on its head. Whereas Democrat (blue) policies push for gun reform and Republic (red) policies favor liberal gun laws, there is a considerable amount more shootings in blue areas (as shown in the pie chart), at 1043 mass shootings for Democrats and 924 for Republican states. Furthermore, while red states exemplify a greater spread of outliers reaching mass shootings into the mid-tens, blue states showed the highest outlier by far, at 441 victims.

Lastly, the Welch Two-Sample t-test revealed many useful statistics. However, most important are the means and p-value. The calculated means were calculated to be 5.44 for Democrat state mass shootings and 5.19 for Republican mass shootings. This is consistent with the previous pie chart in Figure 1 revealing more shootings and the box-plot in Figure 2 with the excluded 441 victim data-point, removed from the chart but definitely affecting the mean. With the p-value of 0.6197, there is a enough evidence to reject the null hypothesis that the victims between mass shootings in Democrat and Republican states. The alternate hypothesis that there is a difference between the two can be assumed with enough evidence. However, there was a significant outlier within the Democrat mass shooting data, and its removal may change the result somewhat.

Conclusion

Utilizing the p-value of 0.6197 of the t-test for the null hypothesis, “there is no difference in victims of mass shootings between red and blue states,” we can claim that there is enough evidence to reject the null hypothesis. As such, the alternative hypothesis, “there is a difference in victims of mass shootings between red and blue states,” is true, which is valuable to this study of the data.

Self-Reflection

This project was a learning experience. My methods consisting of simple comparison between the two main types categories and performing a t-test on total victims for each of the sides was sound, however it could have been better. Not only did it take quite a while, but identifying the proper steps to take and creating a plan required much forethought on the methods of analysis. On top of my methods, the data required unexpected processing. I had to access another source for the comparison of Red versus Blue states and create the CSV file myself, which perhaps could have been automated or found with more effort. Due to that manual creation, I accidentally caused unexpected errors. Furthermore, certain unexpected regions were included within the data, like the “District of Columbia” which is not eligible for voting. That required data processing.

Lastly, outliers were a big problem that I unfortunately couldn’t ignore completely in this project. As stated in Figure 2, subject ID 946496 resulted in 441 victims, which immensely skewed the data and perhaps greatly changed the p-value, or even resulted in the an incorrect alternative hypothesis if it were to be, “Mass shootings in red states have a greater number of victims than in blue states,” which was disproven through the t-test. This p-value, however, could be skewed even further by other statistics not represented in the dataset, like the difference between population in urban versus rural areas, and crime rates of cities between different states.

In the future, I would like to work with more data that tests disparity between ideologies of areas and how the actual data supports it or not.

Sources

Steven Shepard Senior campaigns and elections editor 12:52 a.m.,. Live election results: The 2020 presidential race. POLITICO. Retrieved October 13, 2022, from <https://www.politico.com/2020-election/results/president/> Usmani, Z.-ul-hassan. (2022, May 25). Mass shooting data.csv. Kaggle. Retrieved October 13, 2022, from <https://www.kaggle.com/datasets/zusmani/us-mass-shootings-last-50-years?select=Mass%2Bshooting%2Bdata.csv>