Math 141 Group Project

Carli, Sajid, Tina, Declynn

12/8/2021

Introduction

Since the killing of George Floyd by Minneapolis police officers on May 25, 2020, protests around the country have called for systematic change to prevent such instances of police brutality especially those against black people [Edwards et. al. 2020]. Here, we seek to verify the claims that people of color are disproportionately targeted by police violence. To explore this topic we have two main research questions: (1) Is there racial bias in police shooting deaths? And, (2) does the political leaning of a state impact racial disparity in deaths caused by police shootings?

1. H_0: There is no difference in the proportions of white victims POC victims. Racial status does not affect the use of deadly force by police.

$$p_{white} - p_{poc} = 0$$

H_A: There is a difference in the proportions of white victims POC victims. Racial status does affect the use of deadly force by police

$$p_{white} - p_{poc} \neq 0$$

2. Assuming the population model can be shown by

$$y_{race.ratio} = \beta_0 + \beta_1 x_{political.lean} + e$$

where y is the response, x is the predictor, β_0 is the intercept, β_0 is the slope, and e is the error term.

H_0: There is no linear relationship between states' political alignment (negative values indicates right-leaning while positive values indicate left-lean) and ratio of POC victim

$$\beta_1 = 0$$

H_A:There is a positive linear relationship between states' political alignment (negative values indicate right-leaning while positive values indicate left-lean) and ratio of POC victim. Right leaning states have higher proportion of POC police shooting victims.

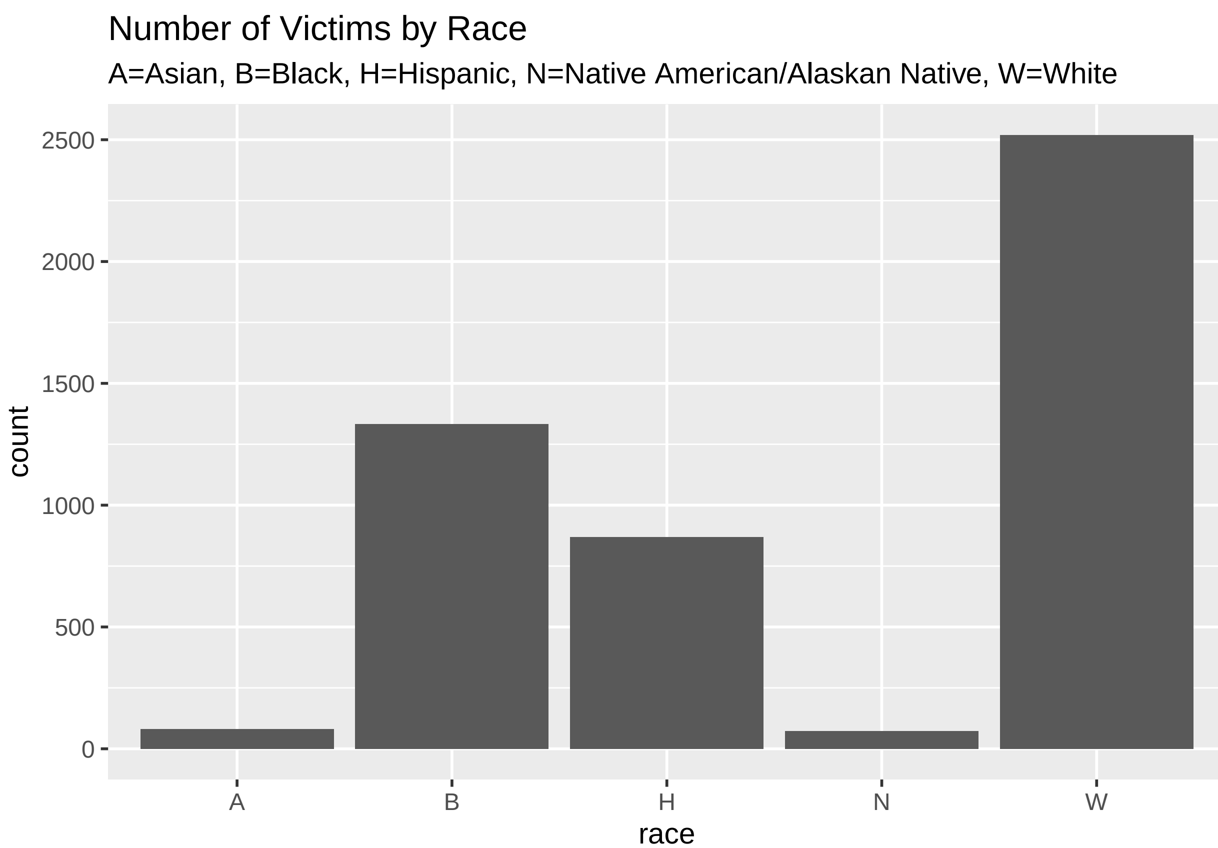
$$\beta_1 \neq 0$$

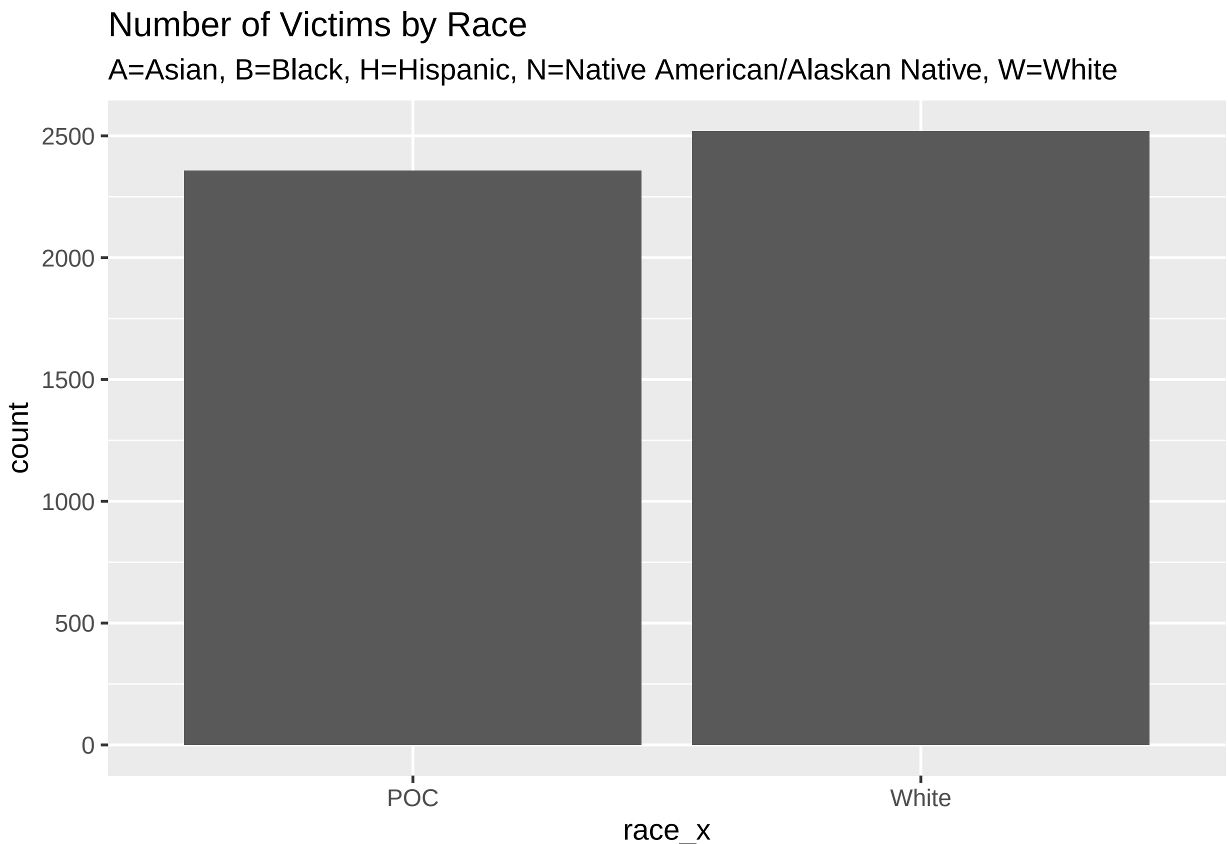
The Washington Post has been compiling a database since Jan. 1, 2015, containing records of every fatal shooting in the United States by a police officer in the line of duty (not off-duty officers) [Tate et. al. 2021]. More than a dozen details about each killing — including the race of the deceased, city and state of the incident, whether the person was armed, and whether the person was experiencing a mental health crisis — have been tracked by culling local news reports, law enforcement websites and social media, and by monitoring independent databases such as Killed by Police and Fatal Encounters.

Political data came from FiveThirtyEight's 2020 partisan lean metric, the average margin difference between how a state or district votes and how the country votes overall [Rakich. 2021]. Positive numbers mean Democratic leans, while negative numbers mean Republican leans. This version of partisan lean, meant to be used for congressional and gubernatorial elections, is calculated as 50 percent of the state or district's lean relative to the nation in the most recent presidential election, 25 percent its relative lean in the second-most-recent presidential election, and 25 percent a custom state-legislative lean based on the statewide popular vote in the four most recent state house elections.

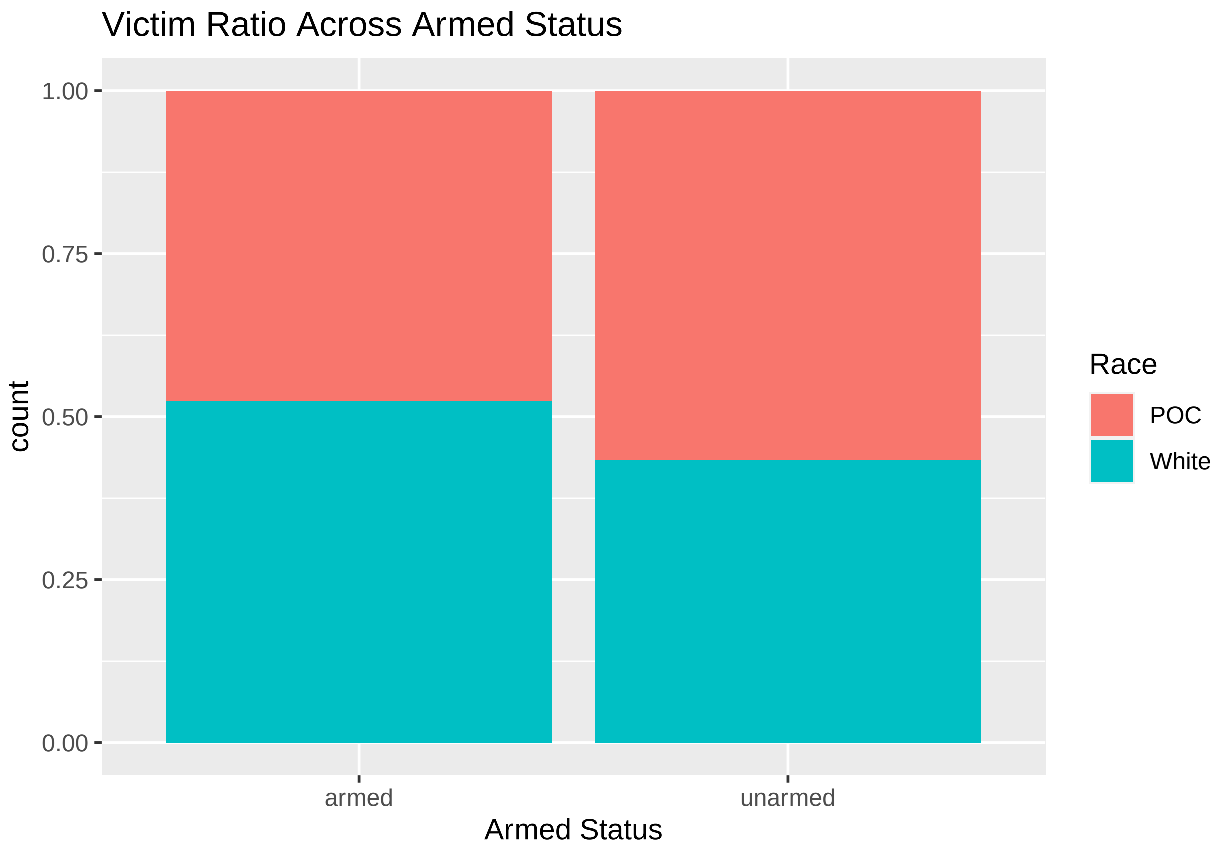
Data Exploration

```
## Rows: 51 Columns: 2

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): state
## dbl (1): 2021

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Rows: 51
## Columns: 4
## $ state  <chr> "Alabama", "Alaska", "Arizona", "Arkansas", "California", "Colo~
## $ `2021` <dbl> -29.594380, -14.620280, -7.587540, -31.836410, 25.454920, 6.367~
## $ po     <chr> "AL", "AK", "AZ", "AR", "CA", "CO", "CT", "DE", "DC", "FL", "GA~
## $ numpoc <dbl> 0.38461538, 0.36111111, 0.51904762, 0.35135135, 0.68208955, 0.4~
```

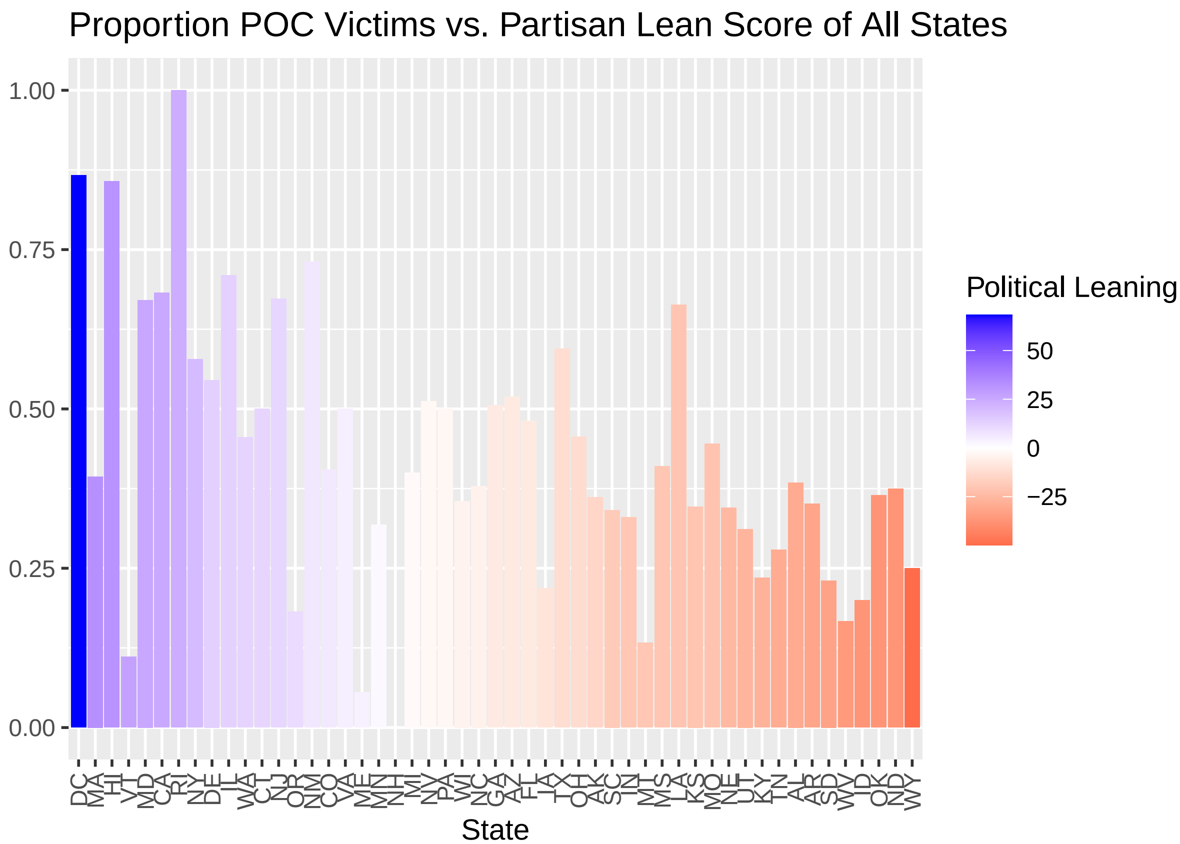


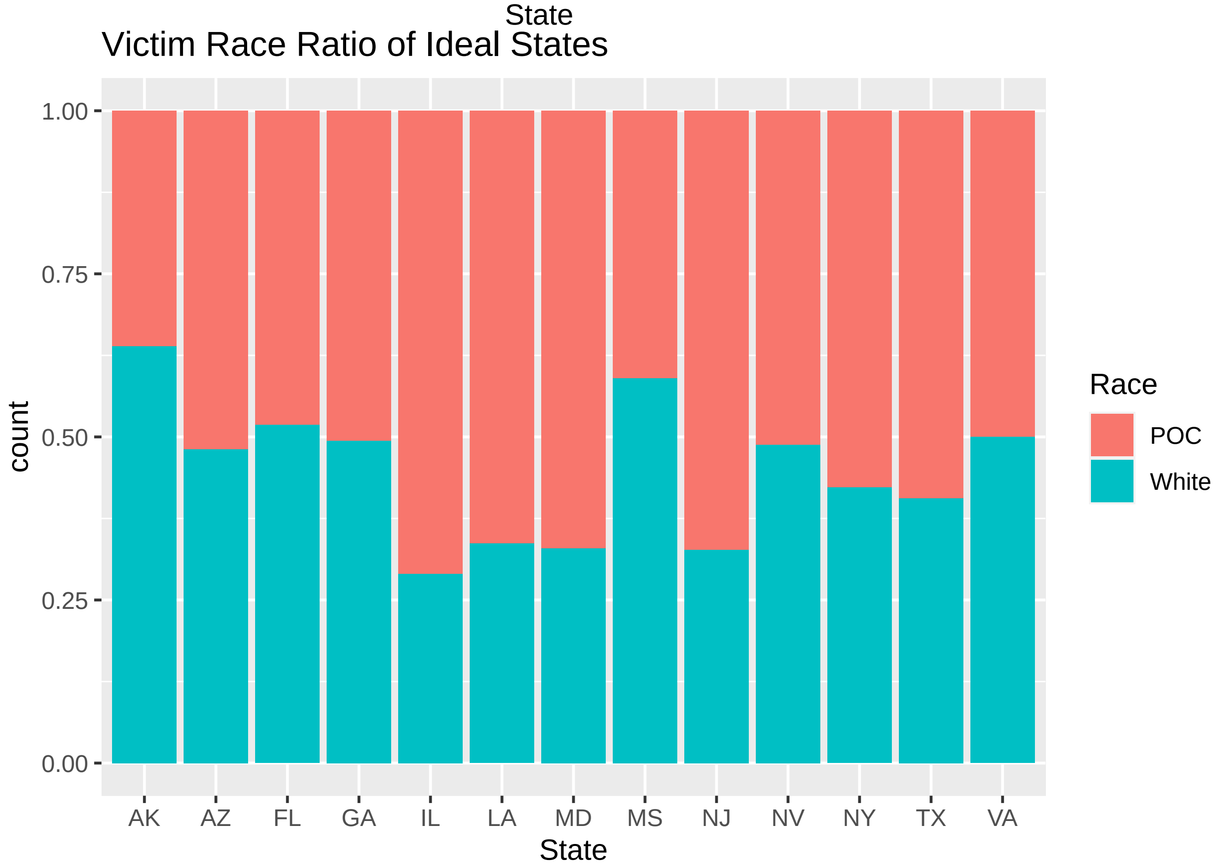


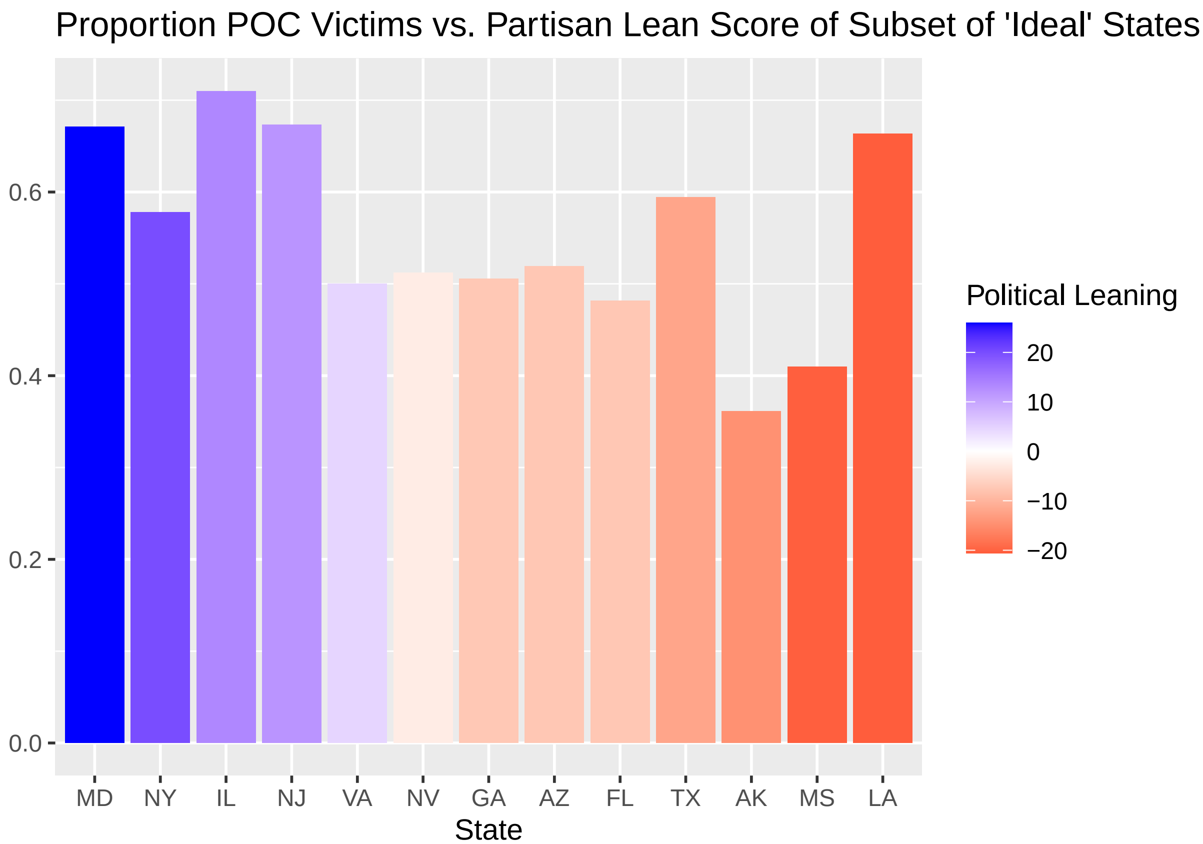
Most victims reported in this dataset are reported as white when broken up by race. When racial minorities are aggregated under the label POC, these numbers are comparible to the number of white victims.



Of unarmed victims, a greater proportion fall under the category of POC while the unarmed victims are closer to a halfway split. This may indicate that POC are perceived as more threatening


Proportion POC Victims vs. Partisan Lean Score of All States
1.00
0.75
0.50
0.25
0.00
DC MA HI VT MD CA RI NY DE WA IL CT NJ OR NM CO VA ME MN NH MI NV PA WI NC GA AZ FL IA TX OH AK SC IN MT MS LA KS MO NE UT KY TN AL AR SD WV ID OK ND WY
State
Political Leaning
50
25
0
−25


Victim Race Ratio of Ideal States
count
1.00
0.75
0.50
0.25
0.00
AK AZ FL GA IL LA MD MS NJ NV NY TX VA
State
Race
POC
White



The ratios of POC and white victims for our subset of state is variable although we would expect a consistent ratio in the absence of racial bias since state populations are approximatly 50:50 white:POC. Our subset of states has a full range of political alignment from left-leaning to right-leaning. There appears to be a slight downward trend in proportion of POC victims as states become more red.

Methods

To adjust for differences in racial populations across states, we chose a subset of states for analysis that have nearly equivalent white and POC populations (40-60% white). These states include Alaska, Arizona, Florida, Georgia, Illinois, Louisiana, Maryland, Mississippi, Nevada, New York, New Jersey, Texas, and Virginia.

The racial identiy of the deceased have been categorized into two divisions: White and POC. The POC includes Asian, Hispanic, Black, Native American. (Other, None, Unknown has been eliminated due to being marginal in number). The armed status of the deceased has been categorized into two divisions: Armed and Unarmed. The Armed category includes Gun, Knife, Toy Weapon, Vehicle.

1. We applied standard bootstrapping procedure for the difference in proportions of race for ideal states with equal population ratio to generate a confidence interval for the average difference in proportions. We found the difference in proportion of POC killed to that of whites for these states only. Then we bootstrapped these differences to estimate a confidence interval for our statistic.

```
set.seed(32)
tab2<-table(dat_select_2$state, dat_select_2$race_x)
prop.table(tab2, margin =1)
```

```
##
##            POC     White
##   AK 0.3611111 0.6388889
##   AZ 0.5190476 0.4809524
##   FL 0.4815951 0.5184049
```

```
##   GA 0.5056818 0.4943182
##   IL 0.7100000 0.2900000
##   LA 0.6633663 0.3366337
##   MD 0.6708861 0.3291139
##   MS 0.4098361 0.5901639
##   NJ 0.6730769 0.3269231
##   NV 0.5121951 0.4878049
##   NY 0.5777778 0.4222222
##   TX 0.5942721 0.4057279
##   VA 0.5000000 0.5000000
```

```
difference <- tibble(
 ideal_states = rep("Ideal", 13),

  POC_minus_WHITE = c(-.2777778, .0120482, .1830986, -.0481928, .038674,
                      .3461538, .1555556, .010989, .1746032, -.2777778,
                      .3493976, .3623954, .3980582)
  )
difference
```

```
## # A tibble: 13 x 2
##    ideal_states POC_minus_WHITE
##    <chr>                  <dbl>
##  1 Ideal                -0.278
##  2 Ideal                 0.0120
##  3 Ideal                 0.183
##  4 Ideal                -0.0482
##  5 Ideal                 0.0387
##  6 Ideal                 0.346
##  7 Ideal                 0.156
##  8 Ideal                 0.0110
##  9 Ideal                 0.175
## 10 Ideal                -0.278
## 11 Ideal                 0.349
## 12 Ideal                 0.362
## 13 Ideal                 0.398
```

```
#observed statistic
mean_diff <- mean(difference$POC_minus_WHITE)
mean_diff
```

```
## [1] 0.1097866
```

```
summary(difference)
```

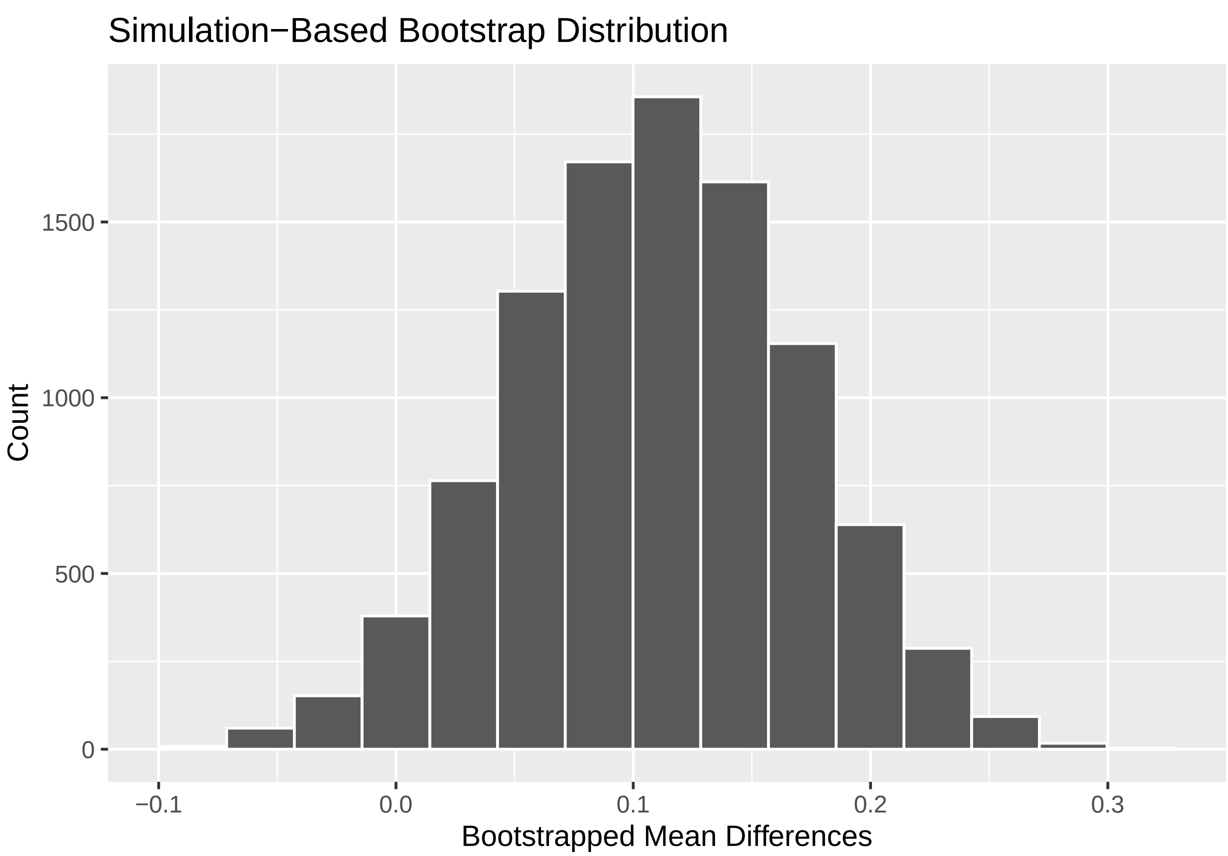
```
##  ideal_states       POC_minus_WHITE
##  Length:13          Min.   :-0.27778
##  Class :character   1st Qu.: 0.01099
##  Mode  :character   Median : 0.15556
##                     Mean   : 0.10979
##                     3rd Qu.: 0.34615
##                     Max.   : 0.39806
```

```
boot_dist <- difference %>%
  specify(response = POC_minus_WHITE) %>%
  generate(reps = 10000, type = "bootstrap") %>% calculate(stat = "mean")
```

```r
boot_dist
```

```
## Response: POC_minus_WHITE (numeric)
## # A tibble: 10,000 x 2
##    replicate   stat
##        <int>  <dbl>
##  1         1 0.142
##  2         2 0.169
##  3         3 0.0552
##  4         4 0.106
##  5         5 0.0841
##  6         6 0.0115
##  7         7 0.0270
##  8         8 0.136
##  9         9 0.0865
## 10        10 0.168
## # ... with 9,990 more rows
```

```r
visualize(boot_dist) +
  labs(x = "Bootstrapped Mean Differences", y = "Count")
```




```r
percentile_ci <- get_ci(boot_dist, level = 0.90)
percentile_ci
```

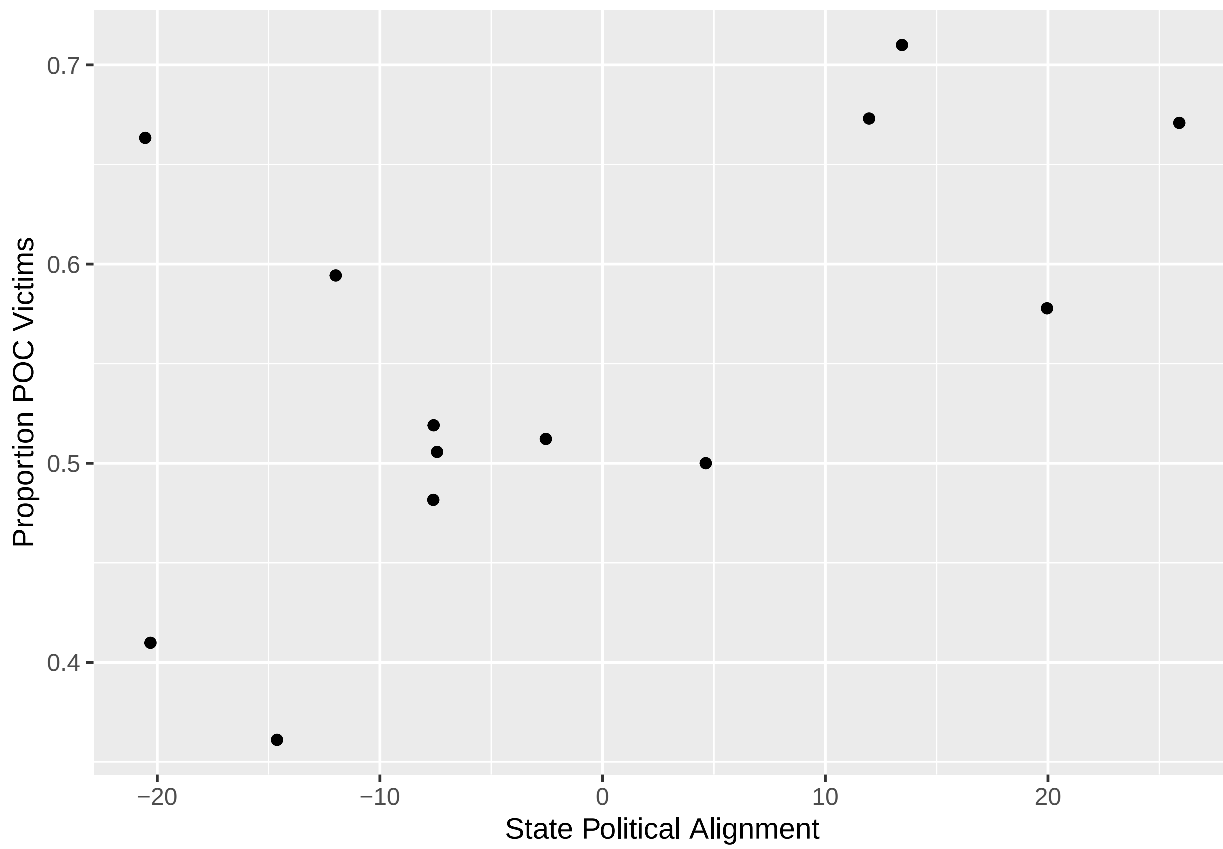
```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1  0.00869    0.207
```

2. Linear regression analysis was performed on state data to understand if there is a relationship between a state's political alignment as given by FiveThirtyEight's partisan lean metric and the proportion of POC victims for that state. The theoretical method for hypothesis testing was used to verify the trend of this relationship.

```
mod<- lm(numpoc~`2021`, dat2_sub)
mod
```

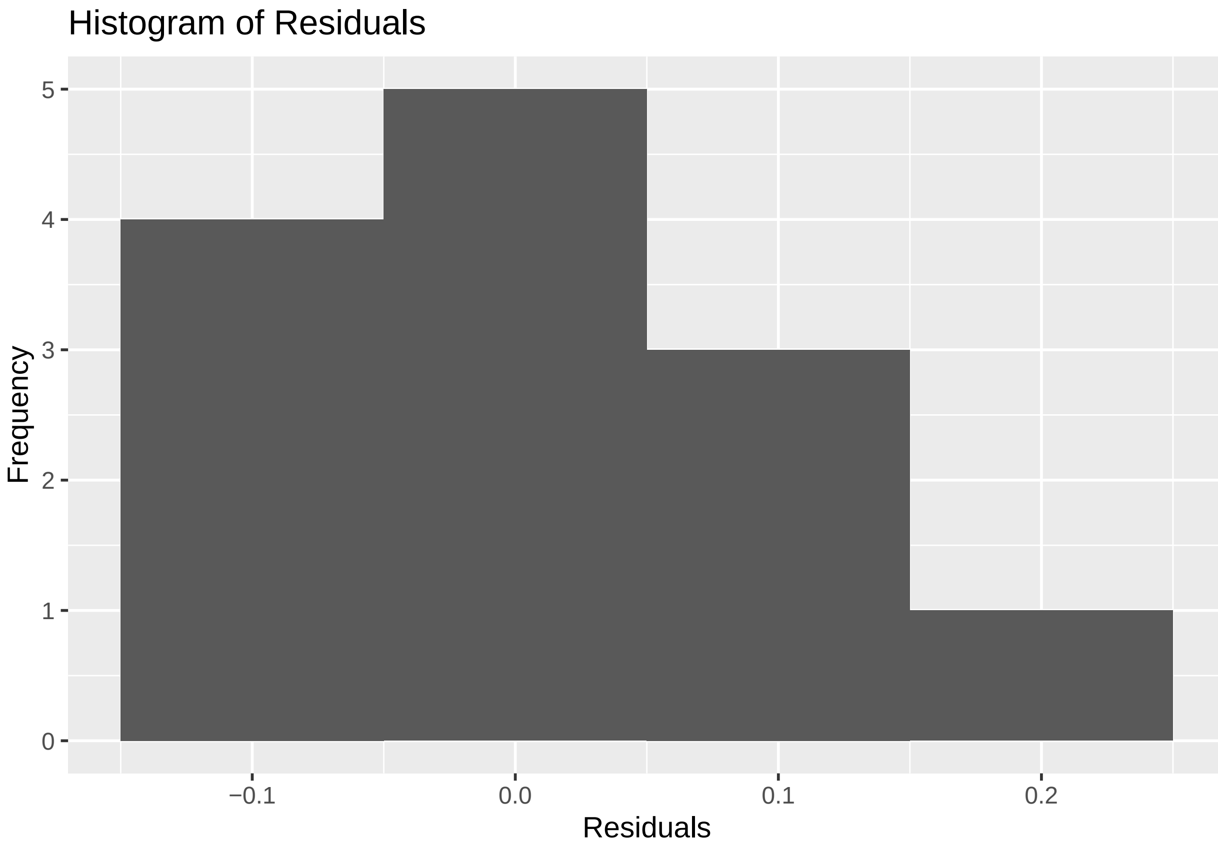
```
##
## Call:
## lm(formula = numpoc ~ `2021`, data = dat2_sub)
##
## Coefficients:
## (Intercept)       `2021`
##    0.557364     0.003999
```

```
#Conditions

#linearity -> scatter plot appears nearly linear
ggplot(dat2_sub, aes(x=`2021`, y=numpoc))+
  geom_point()+
  labs(y="Proportion POC Victims",x="State Political Alignment")
```

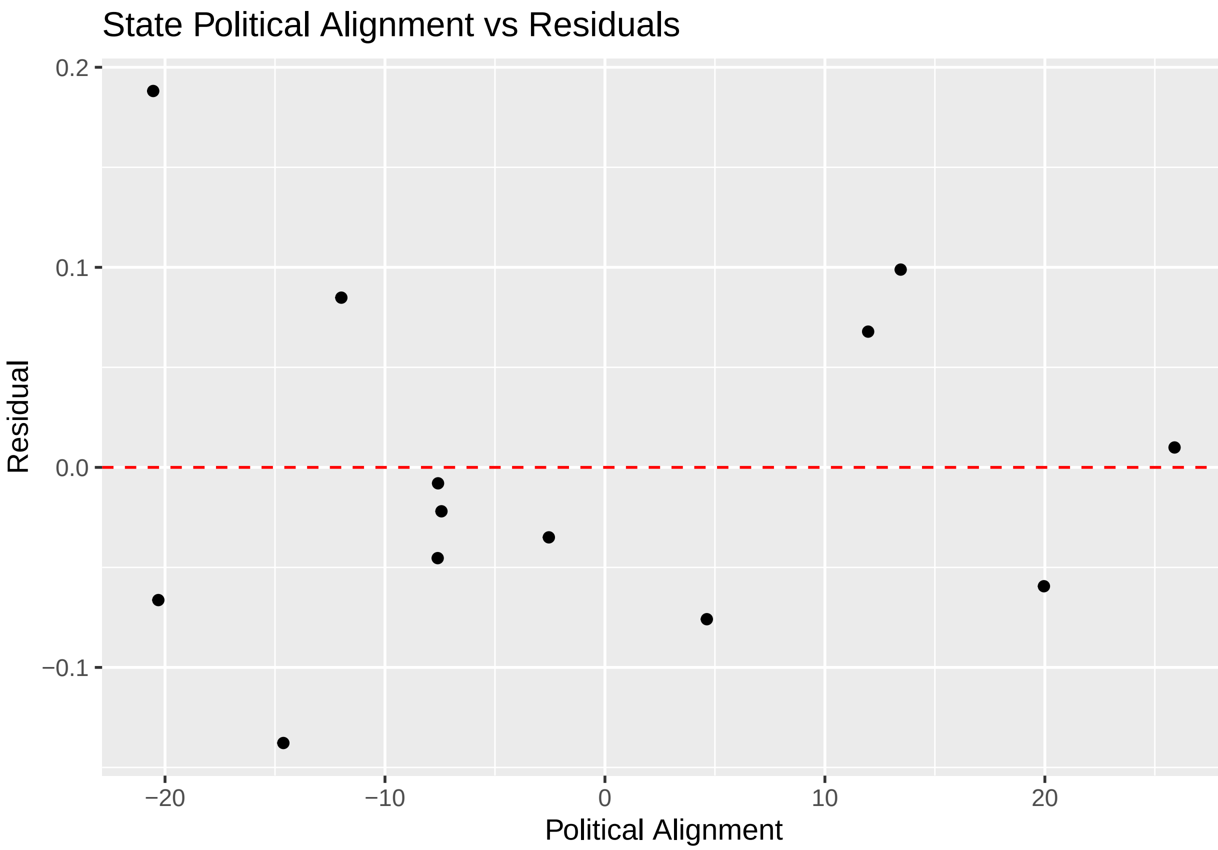



```
#Independence-> data is randomly sampled

#Normally distributed residuals
lm_res <- resid(mod)
lm_res <- data.frame(resid = lm_res)
lm_res <- lm_res %>% mutate(pol = dat2_sub$`2021`)
```

```r
ggplot(data=lm_res, aes(x=resid))+
  geom_histogram(binwidth=0.1) +
  xlab("Residuals") +  ylab("Frequency") + ggtitle("Histogram of Residuals")
```




```r
#Constant or equal variability
ggplot(data=lm_res, aes(x = pol, y =resid))+
  geom_point(size = 1.5) +
  xlab("Political Alignment") +  ylab("Residual") + # x and y label +
  geom_hline(yintercept=0, linetype='dashed', col = 'red') +
  ggtitle("State Political Alignment vs Residuals")
```




```
#p-value
b1=coef(mod)[2]
null=0
SE=0.001755
df=length(dat2_sub$`2021`)-1

T=(b1-null)/SE

pval<- 2*(1-pt(T, df))
paste("p-value:", format(round(pval, 4), nsmall = 2))
```

```
## [1] "p-value: 0.0418"
```

```
#CI
t_star<-qt(0.975, df)
ci<- c(b1-t_star*SE, b1+t_star*SE)

paste("CI:", format(round(ci, 4), nsmall = 2))
```

```
## [1] "CI: 0.0002" "CI: 0.0078"
```

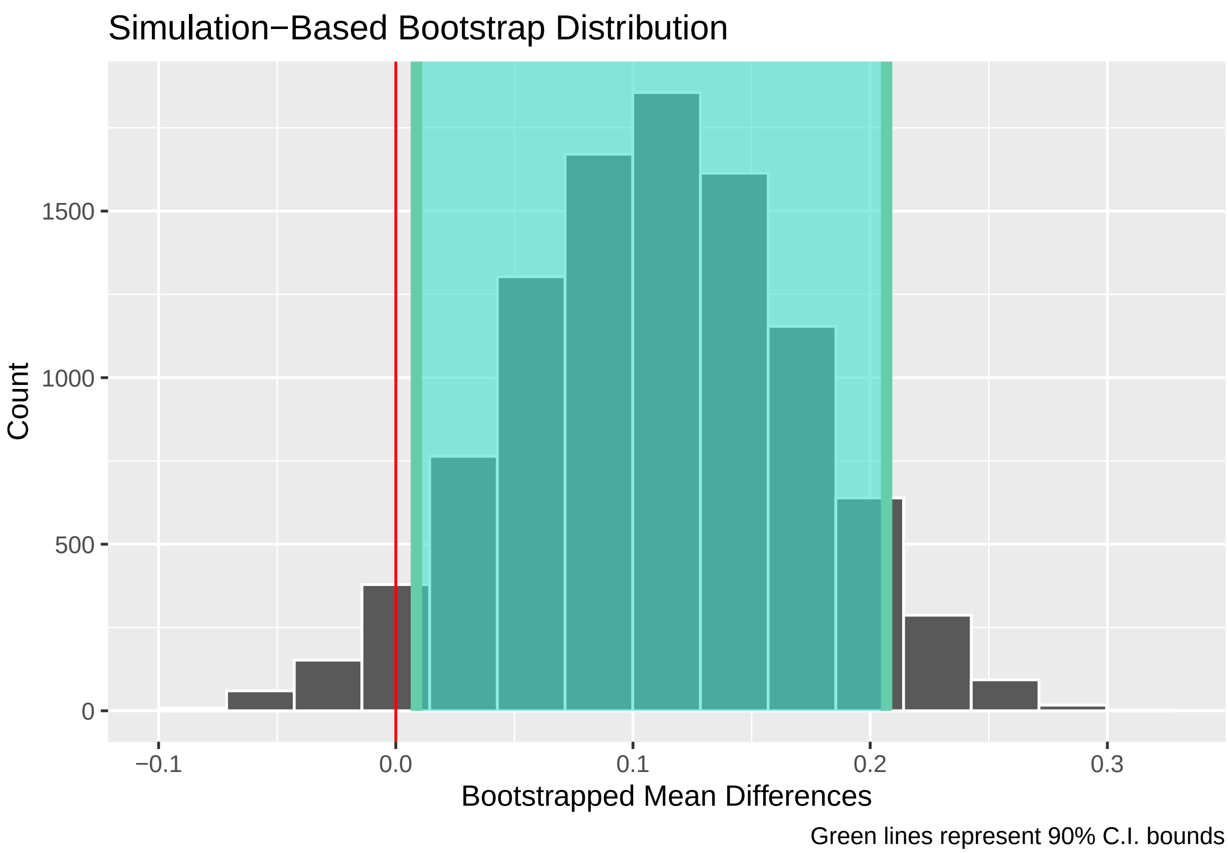
Results and Discussions

We modified our dataset to eliminate the confounding variable of population ratio. It is based on the premise that the number of people killed of a race can roughly be related to the total number of residents of that particular race. Therefore, we adjusted the confounding variable of population by looking only at states with equal population ratio.

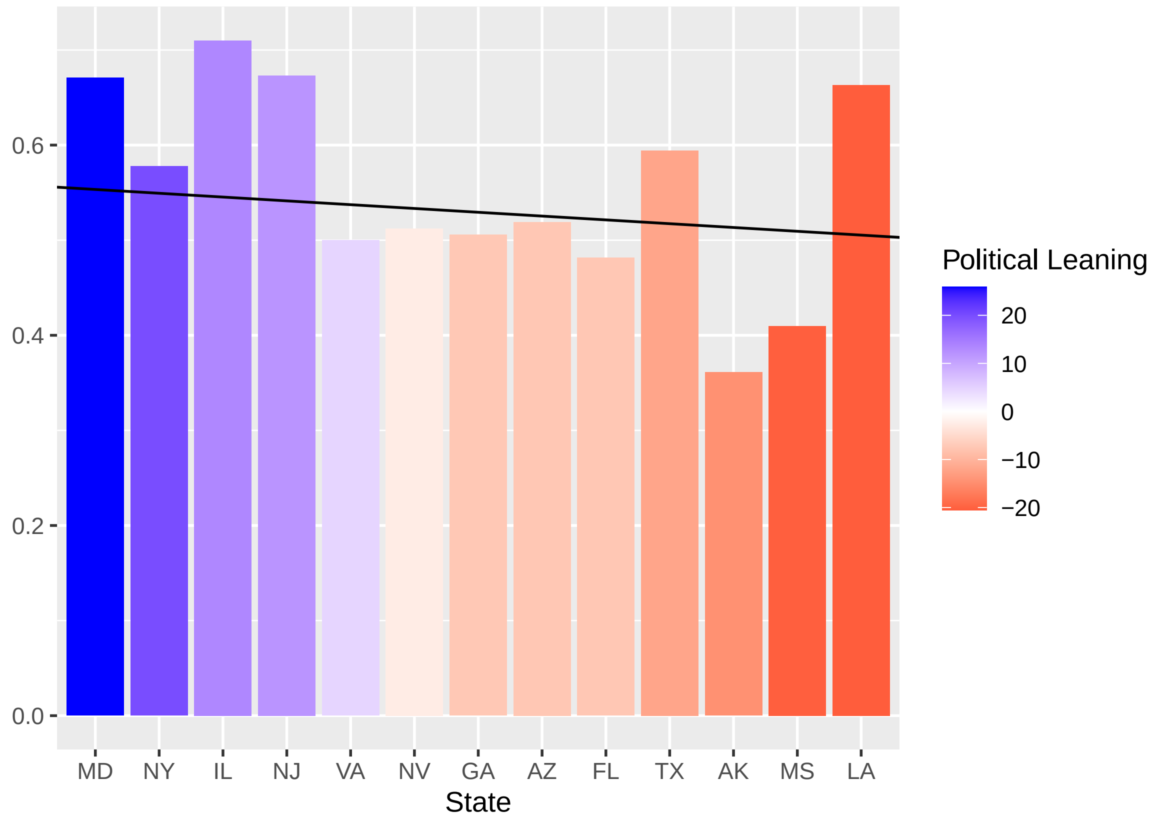
1. Within our subset of states, there is a statistically significant difference in the proportions of white

people killed to the proportion of POC killed by police. We found, with 90% confidence level, that with the absence of population bias, the actual difference in proportion of POC to White people killed lies between 0.0087 and 0.2069. This interval does not contain our null value, 0. Thus, we have enough evidence at 90% confidence to reject the null hypothesis. Since this interval is positive, the data shows that POC are killed more frequently at the hands of police than white people. We can conclude that at 90% confidence there is racial bias in deadly use of lethal force used by the police.

```
visualize(boot_dist) +
  shade_confidence_interval(endpoints = percentile_ci) +
  labs(
    x = "Bootstrapped Mean Differences", y = "Count",
    caption = "Green lines represent 90% C.I. bounds"
  )+ geom_vline(xintercept = 0, color="red")
```



2. For our subset of thirteen "ideal" states (those with approximately equivalent white and nonwhite populations), there does appear to be a positive linear relationship between political alignment as given by FiveThirtyEight's partisan lean metric and proportion of POC victims. Linear analysis of the data give the sample slope as 0.0040 indicating that a 1 unit increase in the partisan lean metric corresponds to an average increase in proportion of POC victim by 0.0040 or 0.4%.



**note: political alignment values decrease along the x-axis to indicate spectrum from left to right politically. Decreasing line of best fit actually has positive slope.

Hypothesis testing of the linear regression using the theoretical method gives a p-value of 0.0418. Since this value is less than 0.05, at 95% confidence, there is enough evidence to reject the null hypothesis and we can assume that there is some linear relationship between political alignment as given by FiveThirtyEight's partisan lean metric and proportion of POC victims. Confidence interval testing indicates that at 95% confidence, the true value of the slope lies between 0.0002 and 0.0078. While these values are small, the entire interval lies above zero indicating that there is some linear trend for these variables.

This data indicates that there is some correlation between political alignment and potential racial bias for police brutality. Red states tend to have a lower proportion of POC victims than blue states.

Conclusions

From our study, it becomes apparent that racial bias exists at the roots of deadly force used by the police. According to the constitution, police are allowed to use deadly force under two circumstances, the first is "to protect their life or the life of another innocent party" — what departments call the "defense-of-life" standard. The second circumstance is to prevent a suspect from escaping, but only if the officer has probable cause to think the suspect poses a dangerous threat to others [Lind, 2014]. The conclusion that more unarmed people of color are killed than white people leads us to an inquiry into why unarmed people of color are deemed as comparatively more "threatening" of the life of an innocent party or the police in duty than white people. However, the scopes of this analysis restrict us from critical examination of the subject. Perhaps a broader study on the nature of police brutality shall be well suited for the investigation. One of the shortcomings of our analysis is that we consider the number of people killed as a representative of the whole population. However, this assumption doesn't take into account that people are not likely to engage in criminal activities indiscriminately. A more rigorous study would take the economic conditions and previous criminal records of the deceased into account.

References

1. Edwards, E., Greytak, E., Ofer, U., Takei, C., & Fernandez, P. (2020, June 30). The other epidemic: Fatal police shootings in the time of covid-19. American Civil Liberties Union. Retrieved December 10, 2021, from https://www.aclu.org/report/other-epidemic-fatal-police-shootings-time-covid-19

2. Lind, D. (2014, August 13). When is it legal for a cop to kill you? Vox. https://www.vox.com/2014/8/13/5994305/legal-police-lethal-force-murder

3. Rakich, N. (2021, May 27). How red or blue is your state? FiveThirtyEight. https://fivethirtyeight.com/features/how-red-or-blue-is-your-state-your-congressional-district/

4. Tate, J., Jenkins, J., & Rich, S. (2021, December 5). Fatal Force: Police shootings database. Washington Post. Retrieved December 10, 2021, from https://www.washingtonpost.com/graphics/investigations/police-shootings-database/