515 Project 2

Sydney Reuter

17 April 2024

Research Question

This project aims to examine how police shootings in the United States have changed since January 2015 in three main ways. These ways are event counts (number of shootings), victim characteristics (race and presence of a body camera), and place (by state). The project then combines these findings to further explore how police shootings have changes statistically.

Literature Review

This article examined the impact of a new policy in the Dallas Police Department (DPD) that required police to document when they point their gun directly at citizens. Narrative reports of officer-involved shootings from 2003 to 2018 were collected from DPD and qualitatively coded. Chi-squared tests analyzed the coded data to identify specific characteristics of shootings over time. The authors found that the policy was associated with a permanently lower proportion of shootings, specifically shootings that occur because an officer mistook an item for a gun. This decrease was gradual. This article is relevant to the current project as body cameras, analyzed in this project, may be considered a form of reporting and therefore may have the same effect as the DPD policy.

Shjarback, J. A., White, M. D., & Bishopp, S. A. (2021). Can police shootings be reduced by requiring officers to document when they point firearms at citizens? Injury Prevention, 27(6), 508–513. https://doi.org/10.1136/injuryprev-2020-043932 (https://doi.org/10.1136/injuryprev-2020-043932)

Data

Data from this project is obtained from the Washington Post and the US Census Bureau. The Washington Post Fatal Force Database provides data on police-involved killings since 2015. Variables include the date, location, victim characteristics (race, gender, age, name), event characteristics (if the victim was armed, fleeing, or threatening), if mental illness was related, and if a body camera was worn. Data from the 2022 US Census, specifically variables pertaining to race, population, and location, was also utilized.

Methods

This project was created using R and R Markdown.

Data was obtained in two different methods. Washington Post data was obtained by CSV file. After loading data in R the file was then cleaned and additional variables were added (date, month, year, and yearmonth). Race values were streamlined to include Black, Asian, Hispanic, White, Native American, Unknown, and Other. Race codes that included multiple races were grouped into “Other”.

Five summary tables were created using the Washington Post data. These tables included a testing variable (yearmonth, state, race, body camera, and race/body-camera), the count of police shootings per variable, and the percentage of police shootings per variable.

US Census Data was obtained by using the “tidycensus” package and a unique API key. Race variable codes were then converted to plain English and summary tables for race and state were made. These summary tables were then combined with the Washington Post “race” and “state” summary tables, which included the variables race or state, count of shootings, count of population, percentage of shootings, and percentage of population.

Finalized summary tables included: Washington Post yearmonth, Washington Post body camera, Washington Post race/body-camera, Washington Post and Census state, and Washington Post and Census race.

Analytical methods were divided into three groups. The first group, event count, utilized the yearmonth summary table. The second group, victims, utilized the race, body camera, and race/body-camera, summary tables. The third group, places, utilized the state summary table.

Event count was analyzed in two ways. First, a line chart was created to graph event counts (y) over time (x). Second, a linear regression was performed to examine if the slope of this model was significant (with event counts as y and date as x).

Victims was analyzed in four ways. First, a bar graph was created to examine the percent of police shootings (y) per race (x), accounting for body camera presence. Second, an outlier test was done on the race/body-camera summary table to examine if any races differed from the group in regard to the presence or absence of a body camera. Third, a proportions test was done on the body camera summary

table to see if there was a significant difference between the proportion of shootings with and without a body camera. Lastly, an outlier test was performed with the race summary table to see if any police shooting percentages (within races) significantly differed from the group.

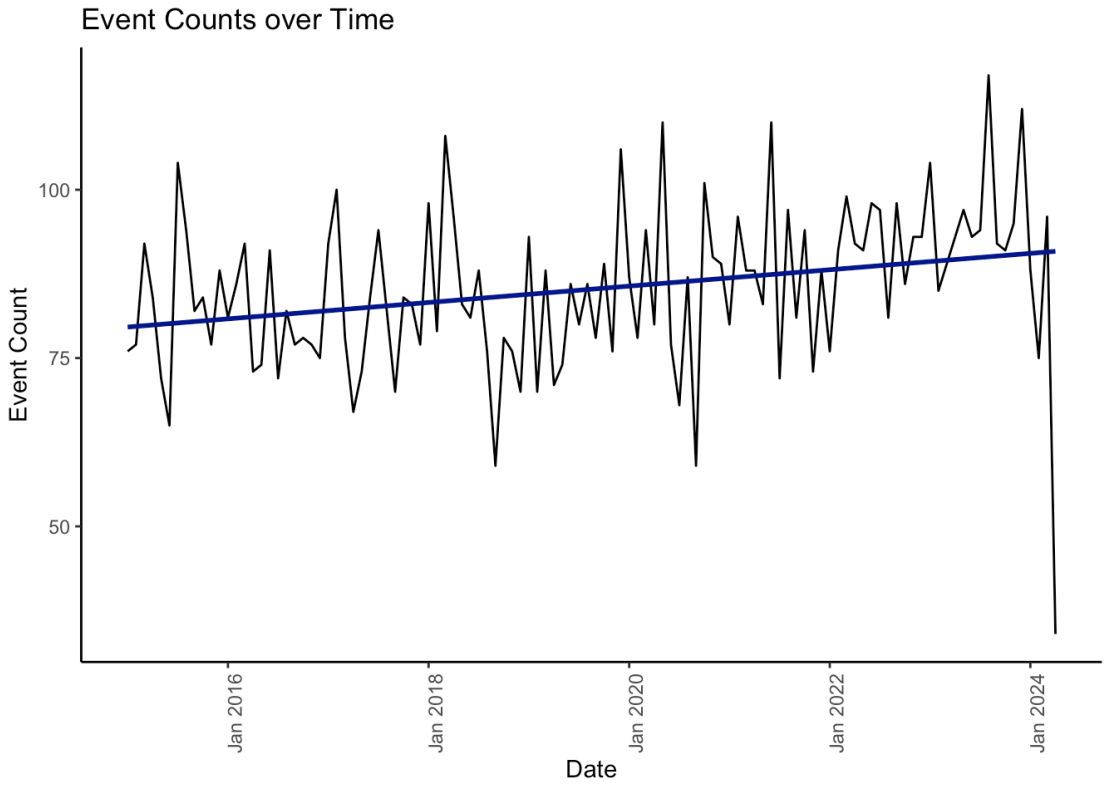
Place was analyzed in three ways. A contour map was created to visualize police shooting hotspots across the US. Additionally, an outlier test was performed on the state summary table to examine if any states significantly differed in terms of percentage of police shootings. A boxplot was also created to visualize the results of the outlier test.

The integration of these three areas of analysis consisted of producing a line chart of event counts (y) per date (x) and three linear regressions (with event counts as y and date as x). These were grouped by states. To focus the analysis on problem areas, only states that were found to have a statistically significant, positive Difference (discussed below in the analysis section) were utilized. These states were located in hot spots identified from the contour map.

Analysis

Event Count

Line Chart



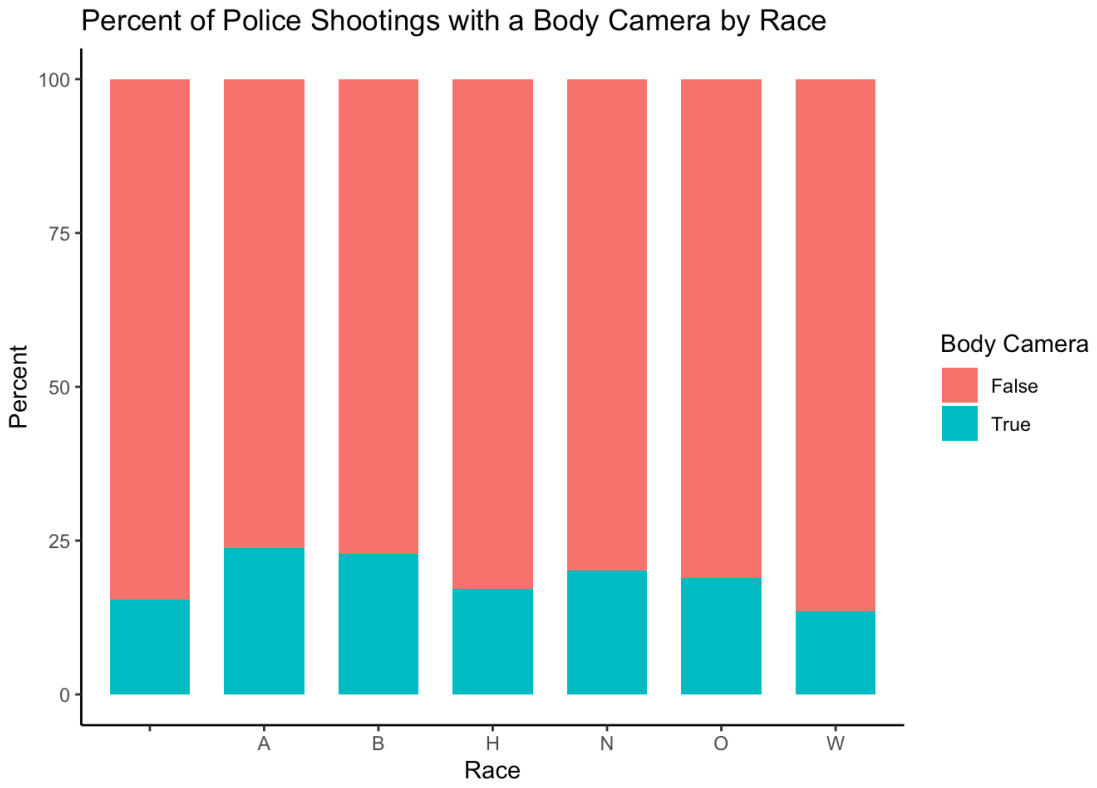
Linear Regression

```
##
## Call:
## lm(formula = count ~ yearmonth, data = wapo.yearmonth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -56.857  -6.913   0.772   6.452  26.954
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2370.8615   835.3126  -2.838   0.0054 **
## yearmonth       1.2161     0.4136   2.940   0.0040 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.79 on 110 degrees of freedom
## Multiple R-squared:  0.07287,    Adjusted R-squared:  0.06444
## F-statistic: 8.646 on 1 and 110 DF,  p-value: 0.003997
```

As seen in the line chart above, the number of police-involved shootings per yearmonth has increased since January 2015. According to the linear regression performed, this increase (slope) is statistically significant at the 0.05 significance level (p = 0.0221). For each month that passes, there is approximately 1 more police-involved shooting.

Victims

Bar Chart



As seen in the bar chart, each race has an approximately equal proportion of shootings that occur with and without a body camera present. Races with a higher proportion of no-body-camera shootings are Native American, White, and Unknown.

Outlier Test (Race and Body Camera)

```
## # A tibble: 7 × 3
##   race  difference IsOutlier
##   <chr>      <dbl> <chr>
## 1 ""         -69.1 n
## 2 "A"        -52.4 n
## 3 "B"        -54.3 n
## 4 "H"        -65.7 n
## 5 "N"        -59.7 n
## 6 "O"        -62.2 n
## 7 "W"        -73.0 n
```

According to the outlier test for race/body camera, there are no outliers. This indicates that the variable “Difference” (which is the proportion of shootings without a body camera subtracted from those with a body camera) is not statistically different for any of the races; therefore, the proportion of shootings with or without a body camera is not significantly different across races. This is reflected in the bar graph above.

Proportions Test

```
##
##  1-sample proportions test with continuity correction
##
## data:  7905 out of 9497, null probability 0.5
## X-squared = 4195.2, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.8246679 0.8397977
## sample estimates:
##         p
## 0.8323681
```

According to the proportions test, the sample proportion of 0.8324 (no body camera present) is significantly different than 0.5 ($p < 0.001$). This indicates that there is a significant difference in the proportion of shootings with and without a body camera. There is a disproportionate amount of shootings that occur with no body camera present.

Outlier Test (Race)

```
## # A tibble: 5 × 5
##   Race            Shooting.PCT Race.PCT    DIFF IsOutlier
##   <chr>                  <dbl>    <dbl>   <dbl> <chr>
## 1 Asian                   1.76     6.98  -5.22  n
## 2 Black                  23.6     14.2    9.35  y
## 3 Hispanic               16.3     18.5   -2.15  n
## 4 Native American         1.35     2.02  -0.668 n
## 5 White                  45.0     73.3  -28.2   y
```

According to the outlier test for race, both Black and White are outliers. This indicates that the variable “Difference” (which is percentage of total shootings minus percentage of total population) for both Black and White is significantly different from the rest of the races Given that these two groups are significantly different from the others, analysis on the Difference variable can be done.

A difference of 0 indicates that the proportion of shootings that involves a specific race is equal to the proportion of the U.S. population that is that race. A positive difference indicates there are proportionally more shootings, and a negative difference indicates that there are proportionally less shootings.

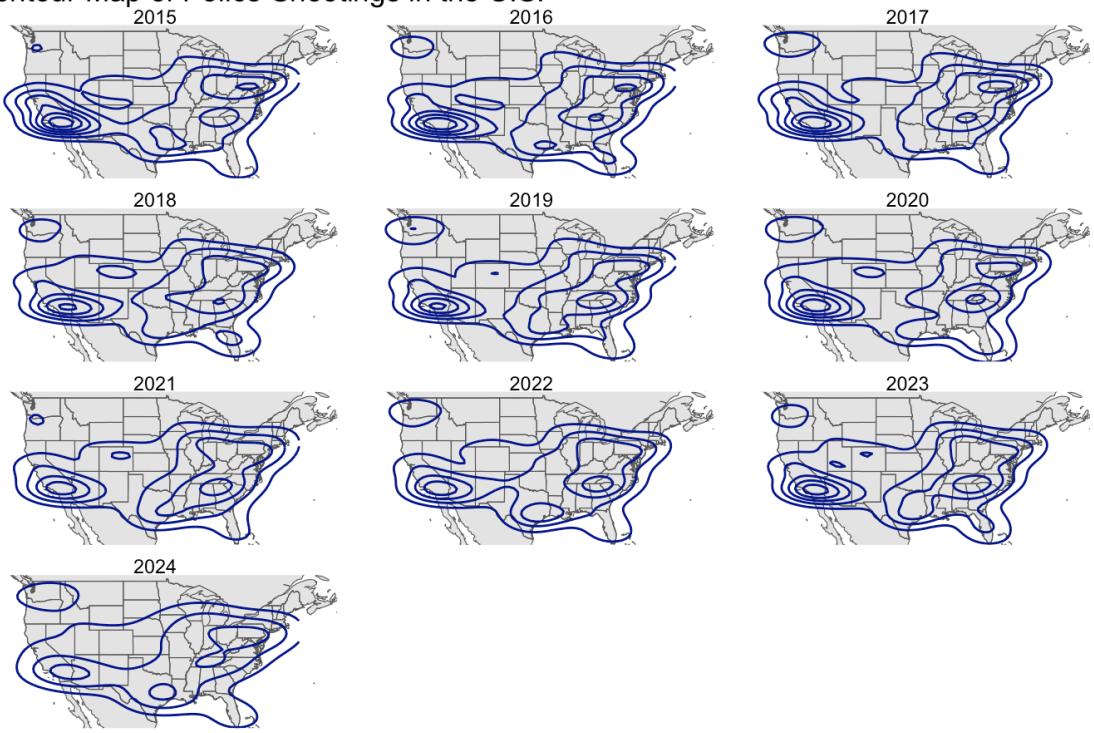
The race Black has a difference of 9.31, which indicates that Black individuals account for more police-involved shooting victims (23.5%) than is proportional to their presence in the US population (14.2%). Additionally, the race White has a difference of -28.2, which indicates that they account for less police-involved shooting victims (45.0%) than is proportional to their presence in the US (73.3%). While other races do not have a difference of 0, they are not specifically discussed in this analysis as they are not statistically different from the group;

however, it may be worthwhile to note that the only race with a positive difference was Black, and all other races had a negative difference. This indicates that the only race that is involved in disproportionately more police-involved shootings (compared to their proportion of the US population) is Black.

Places

Contour Map

Contour Map of Police Shootings in the U.S.

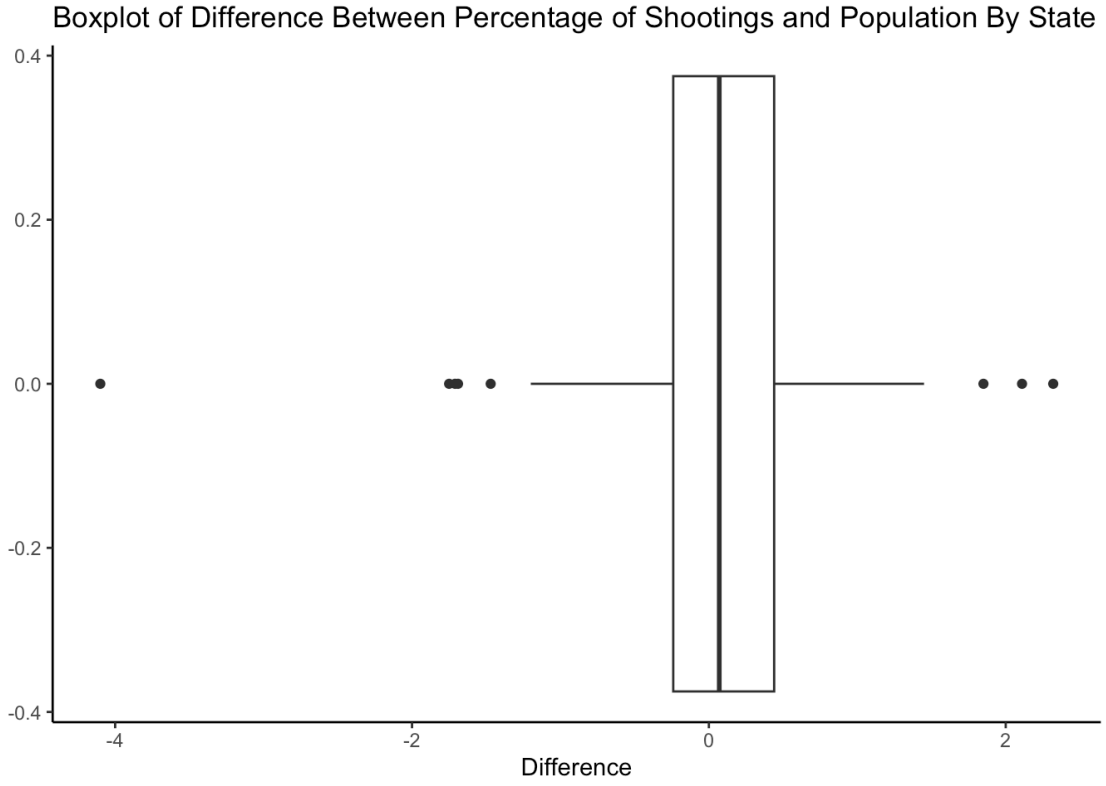


As seen in the contour map, hot spots for police-involved shootings have remained relatively consistent since 2015. These spots are Southern California (including the Phoenix region), the DMV region (DC, Maryland, and Northern Virginia), and the intersection of Southeastern states (specifically Tennessee, North Carolina, South Carolina, Georgia, and Alabama). Additionally, Southeast Texas and the Northwest are sometimes labeled as hot spots. Recently, Denver, Colorado has emerged as a hot spot.

The largest hot spots include Southern California, the DMV area, and the intersection of the Southeastern states. These hot spots also consistently appear regardless of year.

Cold spots include the most Northern states (specifically Montana and the Dakotas). Additionally, the general Midwest (Michigan, Wisconsin, Illinois, and Indiana) appears to be more of a cold spot than a hot spot; however, some activity appears to still be present.

Boxplot



The boxplot above displays the variable Difference. Difference is calculated by subtracting the state population (which is the percentage it makes up of the total US population) from the percentage of shootings that occur in that state. The mean Difference is very close to 0, but is slightly positive. A positive Difference indicates a high proportion of shootings (when compared to state population). There are approximately 8 outliers, 5 with negative Differences and 3 with positive Differences. These are discussed in greater detail below.

Outlier Test (States)

```
## # A tibble: 51 × 5
##    Shooting.PCT State                State.PCT Difference IsOutlier
##           <dbl> <chr>                    <dbl>      <dbl> <chr>
##  1         0.63 Alaska                    0.22     0.41   n
##  2         1.9  Alabama                   1.5      0.4    n
##  3         1.4  Arkansas                  0.9      0.5    n
##  4         4.47 Arizona                   2.15     2.32   y
##  5        13.9  California               11.8      2.11   y
##  6         3.58 Colorado                  1.73     1.85   y
##  7         0.31 Connecticut               1.08    -0.77   n
##  8         0.3  District of Columbia      0.2      0.1    n
##  9         0.21 Delaware                  0.3     -0.09   n
## 10         6.38 Florida                   6.47    -0.0900 n
## 11         3.75 Georgia                   3.21     0.54   n
## 12         0.47 Hawaii                    0.43     0.0400 n
## 13         0.61 Iowa                      0.95    -0.34   n
## 14         0.87 Idaho                     0.55     0.32   n
## 15         2.07 Illinois                  3.82    -1.75   y
## 16         1.95 Indiana                   2.03    -0.0800 n
## 17         0.96 Kansas                    0.88     0.0800 n
## 18         1.75 Kentucky                  1.35     0.4    n
## 19         1.94 Louisiana                 1.39     0.55   n
## 20         0.62 Massachusetts             2.09    -1.47   y
## 21         1.28 Maryland                  1.84    -0.56   n
## 22         0.44 Maine                     0.41     0.0300 n
## 23         1.81 Michigan                  3.01    -1.2    n
## 24         1.03 Minnesota                 1.7     -0.67   n
## 25         2.55 Missouri                  1.84     0.71   n
## 26         1.34 Mississippi               0.88     0.46   n
## 27         0.63 Montana                   0.33     0.3    n
## 28         2.9  North Carolina            3.13    -0.23   n
## 29         0.21 North Dakota              0.23    -0.0200 n
## 30         0.49 Nebraska                  0.59    -0.1    n
## 31         0.3  New Hampshire             0.41    -0.11   n
## 32         1.08 New Jersey                2.77    -1.69   y
## 33         2.08 New Mexico                0.63     1.45   n
## 34         1.66 Nevada                    0.93     0.73   n
## 35         1.88 New York                  5.98    -4.1    y
## 36         2.79 Ohio                      3.52    -0.73   n
## 37         2.63 Oklahoma                  1.19     1.44   n
## 38         1.59 Oregon                    1.26     0.33   n
## 39         2.17 Pennsylvania              3.88    -1.71   y
## 40         0.08 Rhode Island              0.33    -0.25   n
## 41         1.61 South Carolina            1.54     0.0700 n
## 42         0.32 South Dakota              0.27     0.05   n
## 43         2.8  Tennessee                 2.07     0.73   n
## 44         9.48 Texas                     8.75     0.730  n
## 45         1.26 Utah                      0.98     0.28   n
## 46         1.81 Virginia                  2.58    -0.77   n
## 47         0.15 Vermont                   0.19    -0.04   n
## 48         2.69 Washington                2.3      0.39   n
## 49         1.59 Wisconsin                 1.76    -0.17   n
## 50         0.96 West Virginia             0.54     0.42   n
## 51         0.32 Wyoming                   0.17     0.15   n
```

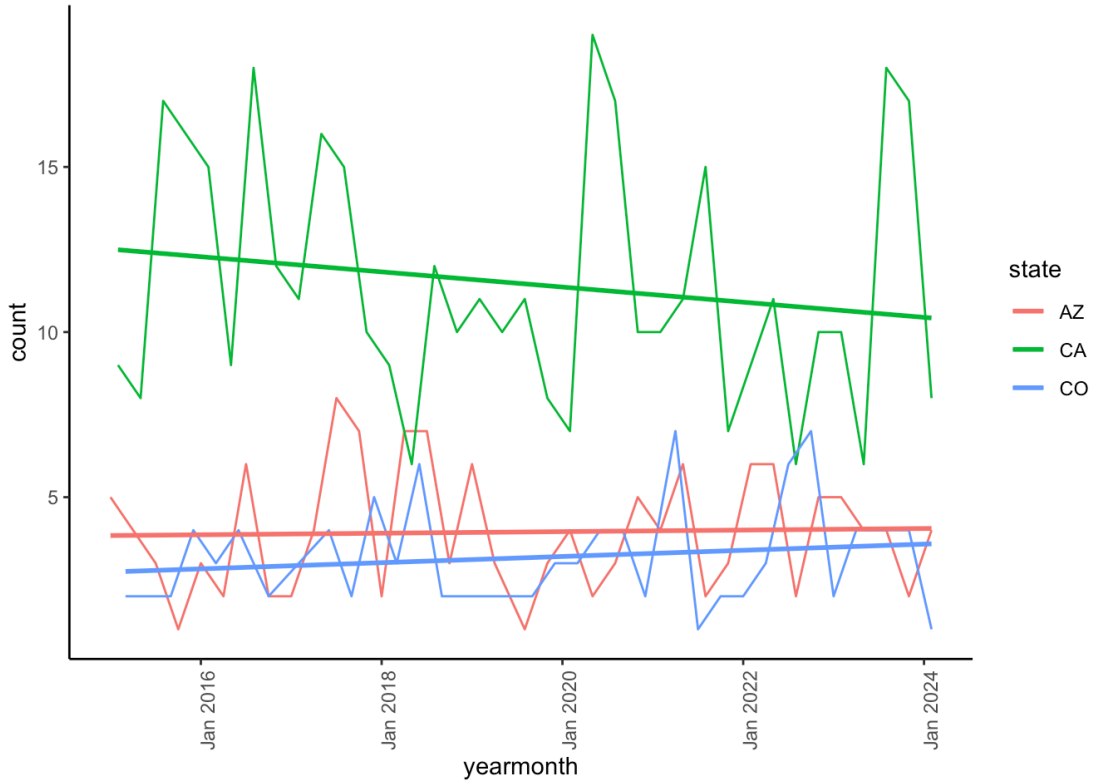
The outlier test examines if the Difference between state population and police shooting percentage is significantly different within states. A Difference of 0 indicates that the population proportion and shooting proportion are equal. A positive Difference indicates a higher proportion of shootings (compared to state population), and a lower Difference indicates a lower proportion of shootings. While no state has a Difference of 0, only states that are considered outliers are analyzed. States that are outliers are considered significantly different from the population (all US states).

The outliers with a positive Difference were Arizona, California, and Colorado. Unsurprisingly, both Arizona and California are considered to be located within hot spots identified on the contour map displayed previously. Additionally, Colorado is an emerging hot spot as shown on the contour map.

The outliers with a negative Difference were Illinois, Massachusetts, New Jersey, New York, and Pennsylvania. According to the contour map Illinois does not appear to be located in a hot spot or a cold spot. Massachusetts, New Jersey, New York, and Pennsylvania are located closely to the DMV hot spot; however, the contour map does not specifically include these states in this hot spot.

Integration

Line Chart



As seen in the line chart above, police-involved shootings for Arizona, California, and Colorado have not increased greatly over time. Conversely, police-involved shootings appear to have decreased for California and remained constant for Arizona. Only Colorado shows an increase in police shootings, which is supported by the contour map previously shown, as Colorado is an emerging hot spot.

Linear Regression

```
##
## Call:
## lm(formula = count ~ yearmonth, data = wapo.state.AZ)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.955 -1.754  0.058  1.164  8.116
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  55.65901  140.99271   0.395    0.694
## yearmonth    -0.02565    0.06981  -0.367    0.714
##
## Residual standard error: 1.99 on 109 degrees of freedom
## Multiple R-squared:  0.001237,   Adjusted R-squared:  -0.007926
## F-statistic: 0.135 on 1 and 109 DF,  p-value: 0.714
```

```
##
## Call:
## lm(formula = count ~ yearmonth, data = wapo.state.CA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2034 -2.5669 -0.5384  1.9560  9.4834
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 797.9915   270.1950   2.953  0.00384 **
## yearmonth    -0.3893     0.1338  -2.910  0.00438 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.815 on 110 degrees of freedom
## Multiple R-squared:  0.07146,    Adjusted R-squared:  0.06302
## F-statistic: 8.466 on 1 and 110 DF,  p-value: 0.00438
```

```
##
## Call:
## lm(formula = count ~ yearmonth, data = wapo.state.CO)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6009 -1.1652 -0.1234  1.0039  4.7128
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -165.71649  110.43130  -1.501    0.136
## yearmonth      0.08365    0.05468   1.530    0.129
##
## Residual standard error: 1.498 on 104 degrees of freedom
## Multiple R-squared:  0.02201,    Adjusted R-squared:  0.01261
## F-statistic:  2.34 on 1 and 104 DF,  p-value: 0.1291
```

The linear regressions above support the conclusions drawn from the line graph. Both Arizona and California have a negative slope, indicating a decrease in police-involved shootings. However, Colorado has a positive slope, which indicates an increase in police-involved shootings. Only California has a statistically significant slope of -0.4063 (p = 0.0034).

In terms of victim characteristics and place, areas with hot spots may have a higher proportion of Black individuals. Additionally, these jurisdictions may not utilize body cameras as frequently as jurisdictions located outside of hot spots. These characteristics may drive up the number of police-involved shootings in these areas. While worth consideration, this project is unable to examine these hypotheses as the data is insufficient.

So What

This project has confirmed that police-involved shootings have increased since January 2015, and that most occur without a body camera. Additionally, Black individuals are disproportionately victims of police-involved shootings and White individuals are disproportionately not victims of police-involved shootings. This confirms previous knowledge regarding police-involved shootings. Additionally, multiple hot spots for police-involved shootings were found. These were Southern California, the DMV area, the intersection of the Southeastern states, Southeast Texas, and Denver, Colorado.

Given this information, practice and policy should aim to decrease the number of police-involved shootings, with special regard to body camera use and race of the subject. Two main practices should be implemented to reduce police-involved shootings.

First, body cameras should be implemented more frequently. There are multiple reasonings for this. Police shootings without cameras far outnumber those with cameras. Additionally, cameras can act as a form of reporting, which is proven to reduce use of firearms by police (Shjarback et al., 2021). Based on this information, increasing body camera use may decrease the number of police shootings. Jurisdictions within known hot spots should prioritize implementing this policy, as these areas are at greatest risk of police-involved shootings.

Second, mandatory reporting of when an officer draws their service weapon and/or discharges their service weapon should be widely adopted. However, this reporting should also require the officer to record the race of the subject. This practice may have two effects. First, an increase in reporting will drive down police-involved shooting rates, as seen in Shjarback (2021). Second, using the logical framework by Shjarback (2021), requirements to report subject race may hold police “more accountable” and reduce shootings involving Black individuals. Like the policy regarding body cameras, jurisdictions located in hot spots should prioritize this implementation for the same reasonings as listed above.

A successful integration of these practices would be a consistent use of body cameras across the majority of jurisdictions, especially those located in hot spots. Body cameras should be worn and recording during every police-citizen encounter. Additionally, thorough reporting that includes race of the subject would be completed each time a service weapon is drawn and/or discharged (excluding training purposes). This reporting would then serve as an extra piece of evidence during investigations of these incidents. Ideally, these practices would have an “accountability” effect on officers, resulting in a decrease of police-involved shootings overall, especially those involving Black individuals.