W200 Fall 2020 | Project 2 Final Report
Title: The Case of Police Shooting Fatality Rates in the US
Team Members: Thomas Gao, Ruby Han, Ruilan Liu

Introduction

In recent years, police brutality has been a hot topic in the United States most notably in the wake of the case of George Floyd in Minneapolis on 25th May 2020.[1] In 2015, The Washington Post began to log records of every fatal shooting by an on-duty police officer in the United States since 1st January 2015. The Post is only documenting shootings in which a police officer in the line of duty shoots and kills a civilian - circumstances that closely parallel the 2014 killing of Michael Brown in Ferguson in which the culmination of Black Lives Matter protest movement began and an increased focus on police accountability nationwide.[2]

The Post is not tracking deaths of people held in police custody, fatality shootings by officers who are off-duty or non-shooting deaths. The database is updated on a regular basis as fatal shootings are reported and as facts emerge about individual cases. In our report, we will be performing analyses on the police shooting fatality rates in the US from The Post in order to determine racial bias correlation strength on shooting rates, COVID-19 impact and whether donning of body cameras affects the fatality shooting rates.

Research Questions

1. **Is there a correlation between the racial composition and the rate of police fatal shootings?**
   We will marry the primary dataset and supplemental demographic dataset to test the first research question. We will present fatality shooting occurrences along with racial composition data.

2. **Are US police shooting fatalities lower in 2020 due to the COVID-19 pandemic?**
   We will compare the fatal shootings from 2015 to 2020 to assess whether shooting fatalities are lower in 2020. As we have not reached the end of 2020, we will only perform data analyses using data from Jan to Oct 2020. In addition, we will perform a complementary cross-sectional analysis to investigate whether states with reduced movement due to lockdown also saw reduced fatal shooting rates.

3. **Will adoption of body cameras help to reduce fatal shootings of police officers?**
   We will compare the proportion of fatal shootings by police officers with and without body cameras. In addition, we will evaluate whether the threat level of suspects will make a difference in police shooting decisions.

Dataset Structure

- **Primary Dataset [3]**

Since 2015, The Post has tracked more than a dozen details about each shooting, including the race of the deceased, circumstances of the shooting, whether the person was armed and whether the person was experiencing a mental-health crisis through culling of local news reports, social media and law

1. https://www.nytimes.com/2020/05/31/us/george-floyd-investigation.html
2. https://www.washingtonpost.com/policeshootings/
3. https://github.com/washingtonpost/data-police-shootings
4. https://en.wikipedia.org/wiki/Firearm_death_rates_in_the_United_States_by_state
5. https://www.bjs.gov/index.cfm?ty=pbdetail&iid=6426
6. https://www.govtech.com/data/Just-How-Common-Are-Body-Cameras-in-Police-Departments.html
7. https://nij.ojp.gov/topics/articles/research-body-worn-cameras-and-law-enforcement#note3
8. Lum, C., et al, Research on body-worn cameras: What we know, what we need to know, Criminology & Public Policy, pp. 93 – 118, March 24, 2019
9. https://www.nap.edu/read/24928/chapter/9#252

enforcement sites. The FBI and Centers for Disease Control and Prevention log fatal shootings by police but officials acknowledge that their data is incomplete. However, The Post seeks to make the database as comprehensive as possible. We will be utilizing this information as our primary dataset resource.

The dataset contains 5826 record entries with 17 columns as of 17th November 2020 from 1st January 2015. The `fatal-police-shootings-data.csv` file contains data about each fatal shooting in csv format. Each row has the following attributes:

Variables	**Description**
`id`	A unique identifier for each victim.
`name`	The name of the victim.
`date`	The date of the fatal shooting in YYYY-MM-DD format.
`manner_of_death`	`shot,shot and Tasered`
`armed`	Indicates that the victim was armed with a weapon that could inflict harm. ● `undetermined`: Unknown whether victim was armed ● `unknown`: The victim was armed with unknown object ● `unarmed`: The victim was not armed
`age`	The age of the victim.
`gender`	The gender of the victim (`M`: Male, `F`: Female, `None`: Unknown).
`race`	`W`: White, non-Hispanic, `B`: Black, non-Hispanic, `A`: Asian, `N`: Native American, `H`: Hispanic, `O`: Other, `None`: Unknown
`city`	The municipality where the fatal shooting took place. Note that this field may contain a county name if a municipality is unavailable or unknown.
`state`	Two-letter postal code abbreviation.
`signs of mental illness`	News reports have indicated the victim had a history of mental health issues, expressed suicidal intentions or was experiencing mental distress at time of shooting.
`threat_level*`	General criteria for `attack` label is most direct and immediate threat to life including incidents in which officers or others were shot at, threatened with a gun, attacked with other weapons or physical force etc. The `attack` category is meant to flag the highest level of threat. The `other` category indicates officers or others facing lower levels of threats. The `other` and `undetermined` categories represent all remaining cases.
`flee*`	News reports have indicated the victim was moving away from officers either by `foot`, `car` or `not fleeing`.
`body_camera`	News reports have indicated an officer was donning a body camera which may have recorded some portion of the incident.
`latitude` and `longitude`	The location of the shooting expressed as WGS84 coordinates, geocoded from addresses. The coordinates are rounded to 3 decimal places meaning that the precision is about 80-100 meters within the contiguous US.

`is_geocoding_exact`	Reflects coordinates accuracy. `true` means that the coordinates are for the location of shooting (within ~ 100 meters), while `false` means that coordinates are for larger region centroid such as the city or country where the shooting occurred.

* The `threat` and `fleeing` columns are not necessarily related. For instance, there was an incident in which the suspect was fleeing and turned to fire a gun at an officer at the same time. In addition, `attack` represents an immediate status prior fatal shots while fleeing could begin slightly earlier and involve a chase.

- **Supplementary Datasets**

    1. 2019 US demographic population data from US Census Bureau site
       https://www.census.gov/quickfacts/fact/table/US/PST045219

    2. 2020 mobility data from Google shows how visits to places, such as grocery stores and parks, are changing in each geographic region by measuring how visits and length of stay at different places change compared to a baseline. The baseline is the median value, for the corresponding day of the week, during the 5-week period Jan 3 – Feb 6, 2020.
       https://www.google.com/covid19/mobility/

Data Exploration and Cleaning

For our main dataset, we started by performing sanity checks and data cleaning in order to build a usable training dataset for our analyses. We collectively decided to remove the name column from the dataset as this is personal information and may violate individual rights. Since we are only analyzing data until Oct. 2020, we have also only extracted the dataset up until the end of October 2020 which resulted in 5743 records. In the sections below, we highlight our observations about missing data, anomalies and how we addressed them for data preparation.

- **Missing Data**

We initially examined the comprehensiveness of the data provided to us. We noticed immediately that there were 619 missing values from the race attribute in the primary dataset. As we are unable to determine the race of the individual due to the missing data, we proceeded to drop those rows in order to evaluate just the first research question on correlation between racial bias versus police shooting rates. However, it is important to note that the missing race data were not dropped when examining both the second and third research questions.

- **Data Types**

We then verified the data types for all datasets in order to plan for how actions or computations will be handled. Once we deemed the data types to be appropriate, we then went on to extract yearly and monthly data in order to parse the primary dataset.

- **Data Transformation**

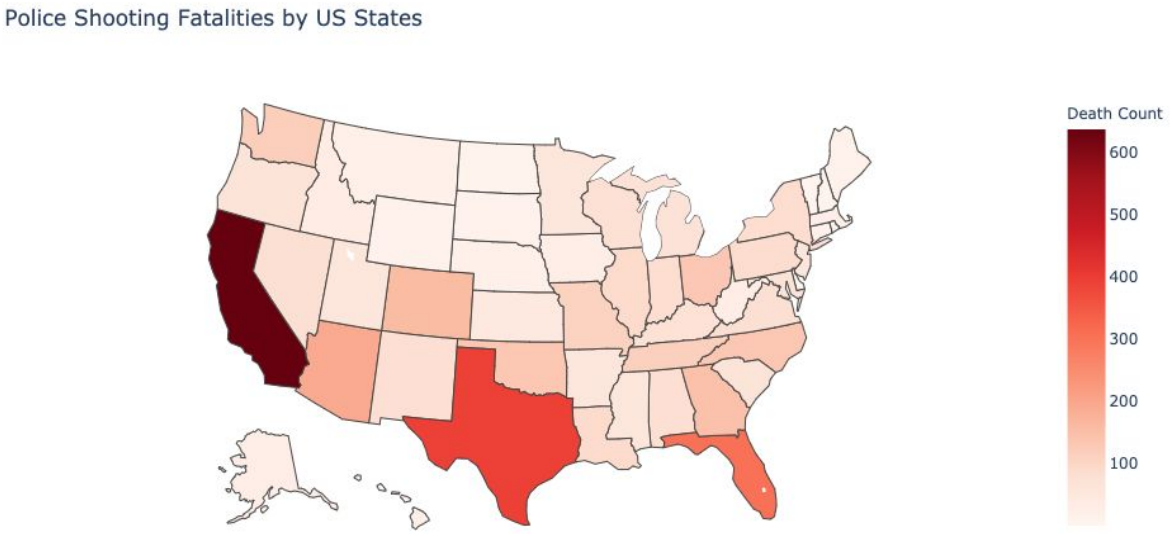
The race elements have been renamed and fully spelt out for clarity and to eliminate ambiguity. The date column has been transformed into a day-of-year variable (day 1 - 366) for alignment across years. In addition, the header columns for the US demographic dataset have been renamed to eliminate whitespaces. A subset of data is produced to generate a total number of fatal shootings at US and state level.

- **Data Anomalies**

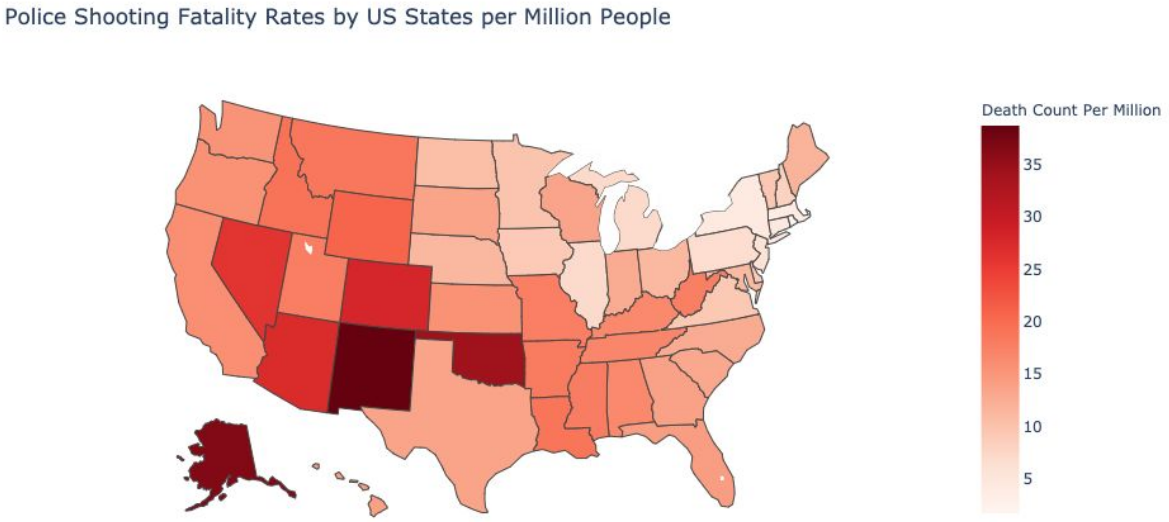
We initially observed that the first row of ID started from 3 instead of 1 in the primary dataset. We then performed sanity checks to verify the total number of missing IDs. A total of 539 missing IDs was discovered and the reason for that is unknown. However, we deemed that this will not affect our subsequent analyses.

- **Initial Data Exploration**

Now that we have performed data cleansing and sanity checks, we will begin exploring our datasets. We started out by attempting to produce an interactive geo map to illustrate the geographical distribution of fatal shooting occurrences using plotly.



Based on the geomap of death count by US states, we observed that California state has by far the highest death count followed by Texas and Florida in terms of the top three states. However, our initial geomap did not take into account the population per state. Thus, we would expect that states with a higher population would have a higher death count via police killings as well. We imported a new supplementary dataset of population per US state and factored that into our next geomap illustration to take into account state population baseline.



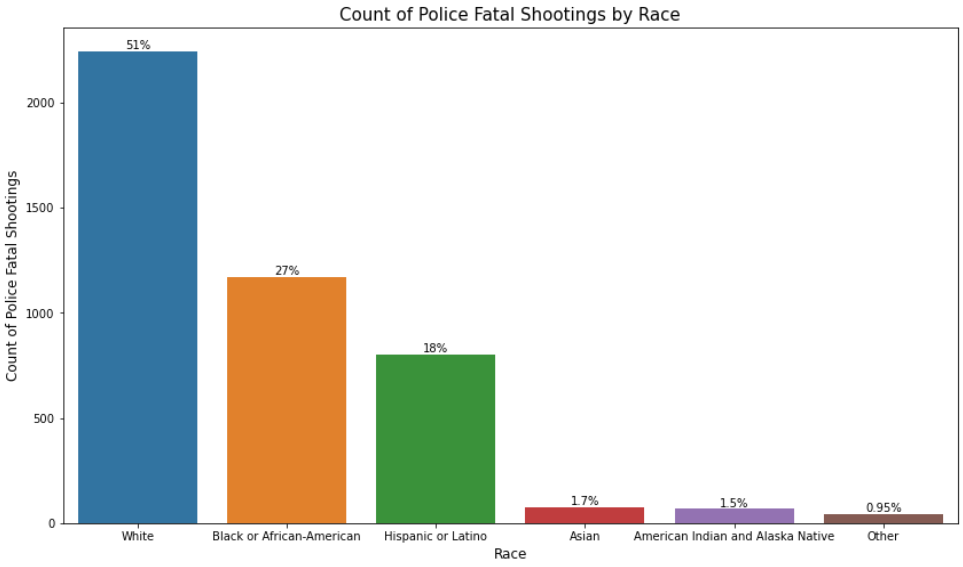
Now that the population is normalized, the color density is a little more evenly distributed across the new geomap. The top three states with the highest death count are New Mexico, Alaska and Oklahoma. The previous top three states (California, Texas and Florida) have a relatively low death rate with data normalization as compared to before.

It is interesting to note that the top three states New Mexico, Alaska and Oklahoma after population normalization have higher overall firearm death rates [4] at 18.1, 23.3 and 19.6 as compared to California, Texas and Florida at 7.9, 12.1 and 12.6. These numbers have been adjusted for population of 100,000 and include suicide, self-defense and accidents. However, we are unable to strongly conclude if there is a direct correlation in terms of officer shooting rates by state. We will now proceed to analyze our research questions.

Analyses: Plots, Charts and Figures

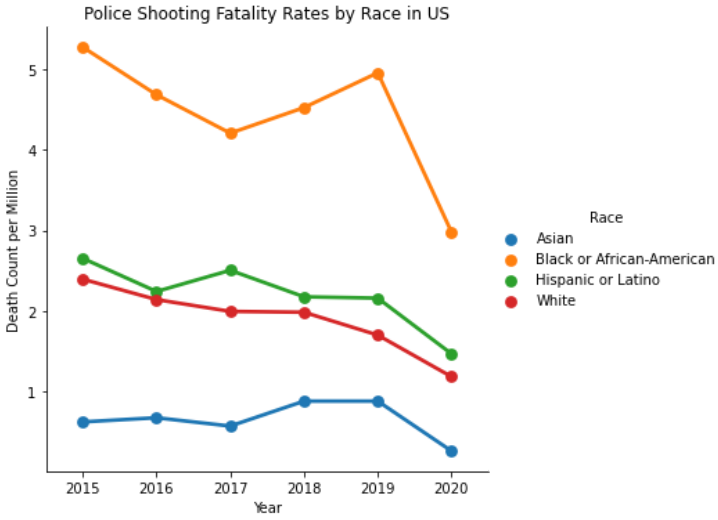
Research Question 1: There is a strong correlation between racial composition and police shooting fatality rates in the US.

An important aspect of this project is to determine whether certain racial groups are disproportionately being fatally shot by police. Rather, is there a strong correlation between racial composition and police shooting fatality rates in the US? We test our first hypothesis on racial disparities using an approach that compares the rate at which racial groups are shot with the US demographic population accounted for as the baseline.



Based on the above bar chart, a little more than half of the individuals shot and killed by police are Whites, followed by Black or African-Americans, Hispanics or Latinos, Asians, American Indians and Alaska Natives, and Others. However, the above result does not reveal a lot of insight as each racial group is not proportionately represented in terms of the total US demographic population. Thus, we incorporated the supplemental demographic dataset from the US Census Bureau (shown in the table below) to calculate death count per million people.

	race	population_in_million	%
0	White	197.128	60.1
1	Black or African-American	43.952	13.4
2	Hispanic or Latino	60.680	18.5
3	Asian	19.352	5.9
4	American Indian and Alaska Native	4.264	1.3
5	Other	2.624	0.8



The ‘American Indian and Alaska Native’ and ‘Other’ groups were both omitted as they were noisy estimates due to small sample size. In addition, we omitted the ‘Other’ group because the classification of this group (consisting of two or more types of race) varies across both the primary and supplementary datasets.

Based on the above plot, we observed that the trend is declining seemingly indicating that shooting rates are decreasing over the years. On the contrary, the trend has been declining due to the missing race data. The below figure displays the frequency of missing race data since 2015.

```
2020    215
2019    141
2018    101
2017     77
2016     56
2015     29
Name: date, dtype: int64
```

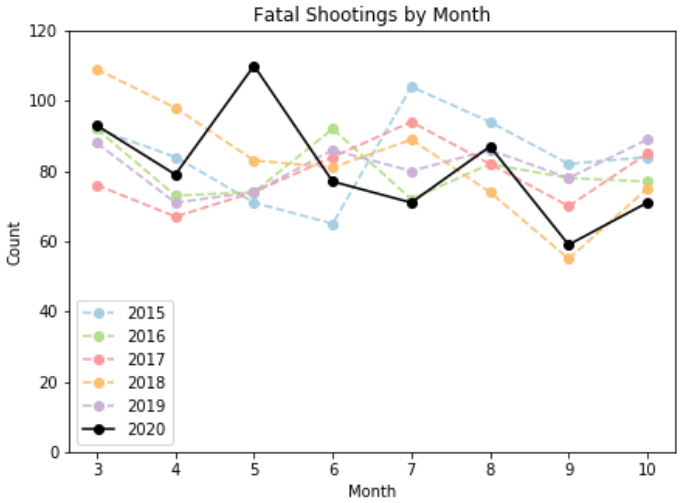
Nonetheless, there is a clear disparity between killings and race. We observed that although almost half of the people shot and killed by police are White Americans, Black Americans are shot at a disproportionate rate. Black Americans account for about 13% of the U.S. population, but are killed by police at more than twice the rate of White Americans based on the line plot. Hispanics are also killed by police at a slightly disproportionate rate in comparison to White Americans albeit at a lower rate compared to Black Americans based on this dataset alone.

What are the implications? Are US police officers targeting a specific community? We do not have data to prove that. We will require a deeper understanding of why crime rates are frequently committed by certain races. Rather, how do we reduce race-specific crime rates? Of course, this is no simple task as crime rates are the result of a complex, large and dynamic set of forces. However, the magnitude of these disparities speak to the importance of this discussion and it suggests that reducing disparities will require identifying and changing socio-historical factors that lead civilians to commit acts of criminal activities.

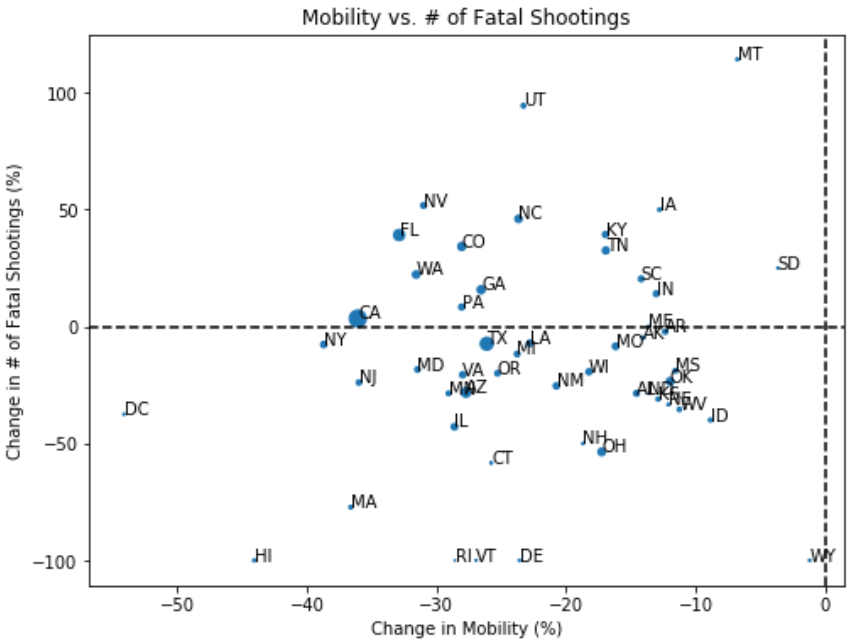
Research Question 2:

Our second hypothesis is that police fatal shootings are lower in 2020 due to the COVID-19 pandemic. The rationale is that reduced outdoor activities have led to fewer policing needs and thus fatal shootings. We test the hypothesis using two methods. The first method is to compare 2020 monthly fatal shootings with those of previous years. The second method is to merge with mobility data and perform cross-sectional comparisons to verify if places with reduced mobility have reduced levels of fatal shootings. However, both methods failed to support the hypothesis.

For the first method, we compared the 2020 line with lines from 2015 to 2019 over the months. We exclude months before COVID and months we don't have complete data yet. In the chart below, we do not observe lower levels of fatal shootings in 2020. In May 2020, there was in fact a spike in fatal shootings, and the increased number of fatal shootings was observed in the weeks following George Floyd's death on May 25.

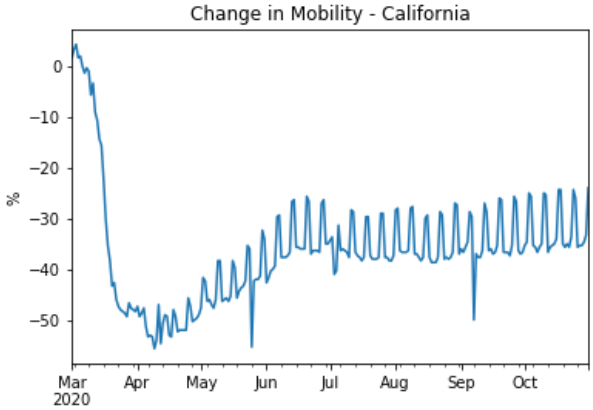


Our second method is to conduct cross-sectional comparisons across states. In the next chart, we plot changes in the number of fatal shootings against changes in mobility. The size of the dots corresponds to the number of fatal shootings for each state from 2015 to 2019. If our hypothesis is true, states with reduced mobility would observe reduced fatal shootings, and thus the scatter plot would show a positive relationship. However, we see only a weak positive linear relationship across the states from visual inspection. We calculated the correlation to be 0.16 with an R-squared value of only 2.6%. Consequently, we deem the relationship not strong enough to support our hypothesis.



We will now describe in detail how we derived the variables in the chart above. To calculate the change in the number of police fatal shootings for each state, we compare the average number of fatal shootings from 2015 to 2019 to that of 2020 and calculate the percent change. Similar to the first method, only data from March to October is used.

To calculate change in mobility, we used the mobility dataset made available by Google. The dataset contains daily information on US counties for six categories of locations and reports mobility relative to Jan. 3 to Feb. 6 mobility before COVID-19 had an impact on the US. Out of the six categories of locations, we filtered for three categories as a proxy for lockdown which include retail and recreation, transit stations and workplaces. Relative mobility is reported in percentages, where 0% indicates the same level of mobility as baseline. We averaged relative mobility across categories, counties and dates and obtained a relative mobility metric for each state.



To sanity check the data, we reviewed the states with largest changes in mobility. DC, Hawaii and New York had the largest reduction in mobility, whereas Wyoming, South Dakota and Montana had the least reduction. This is not surprising based on news reports. To further validate the dataset, we plotted the change in mobility over the months for California in 2020 as shown on the left chart. The steep drop around March and the steady recovery in the coming months align with lockdown strictness in California.

In conclusion, we do not clearly observe the hypothesized effect using either temporal or cross-sectional method. A possible explanation is that both methods are impacted by widespread protests against police brutality this year. While reduction in general mobility might have reduced police fatal shootings, intensified and widespread conflict between police and protestors likely increased the number of police fatal shootings.

Research Question 3:

Our analysis does not show a conclusive result on whether body cameras result in fewer police shootings at a national level. However, we did observe a reduction in fatal shootings in California where adoption of body cameras by the police is prevalent. We believe that the weak result at the national level was likely related to a low level of adoption of body cameras across all states.

The vast majority of fatal shootings across all threat levels are perpetrated by officers who were not donning body cameras at around 87%. However, this does not mean that officers without body cameras are more likely to be involved in fatal shootings as we are missing several critical pieces of information. For example, we need to know the adoption rate of body camera usage, to see how the 13% of shootings perpetrated by officers wearing body cameras is consistent with the percentage of officers who are actually wearing cameras. Another piece of important information would be how many situations the officers were involved in where they might have needed to fire their gun but ultimately did not (although this type of data would be very difficult to collect).

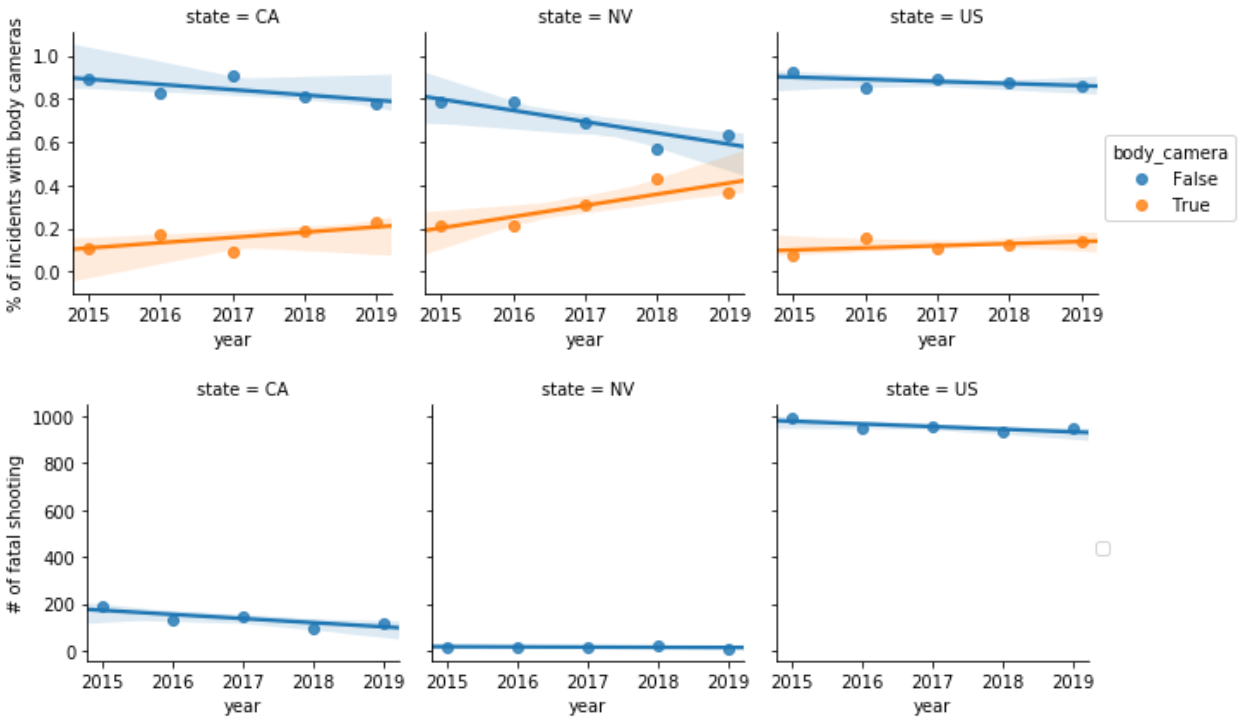
threat_level	attack	other	undetermined
body_camera			
False	88.78%	86.44%	88.72%
True	11.22%	13.56%	11.28%

We attempted to estimate the adoption rate of body cameras via other sources. According to a report of the Bureau of Justice Statistics (BJS) [5]:

1. 47% of general-purpose law enforcement agencies had acquired body-worn cameras in 2016.
2. Overall, in agencies that had acquired body-worn cameras there were 29 body-worn cameras in service per 100 full time officers.

The report suggests that the national average adoption rate per police officer is around 13% (47% x 29%). This estimated adoption rate is nearly identical to the rate of fatal shootings perpetrated by police officers wearing body cameras as noted above. These similar estimates would indicate that an officer wearing a body camera is not less likely to be involved in a fatal shooting.

We can also look at the trends over time, to see if there are other relationships we can detect. A reasonable assumption seems to be that the adoption of body cameras has been increasing over time. As adoption increases, we would naturally expect to see an increase in the percentage of shootings perpetrated by officers wearing body cameras (assuming there are no police officers wearing body cameras, the incidents would all be coded without body cameras; while if all officers are wearing body cameras, all incidents should be coded with body cameras). Therefore, if there was a positive relationship between the adoption of body cameras and a reduction in fatal shootings, we would expect an increase in the percentage of incidents with body cameras and a decline in the total number of fatal shootings over time.



At a national level, we see a very modest increase in the percentage of fatal shootings perpetrated by police officers wearing a body camera, from 7.5% in 2015 to 12.2% in 2018. Based on an estimate from Government Technology [6], 60% of the US law enforcement agencies should be equipped with body cameras by the end of 2018 - very fast growth from 47% adoption rate in 2016. However, the percentage of incidents with body cameras (orange line) does not indicate the steep increase at the national level (state=US). In other words, the percentage of incidents with body cameras doesn't keep up with the body camera adoption rate. Here we want to point out that the number of agencies adopting body cameras does not equate to the number of police using body cameras, which could possibly explain the discrepancy between the high adoption rate of body cameras and the flat trend of incidents with body cameras. Because the large agencies have more budget to adopt body cameras [5], we assume that police officers in big cities will have a higher adoption rate per capita and in turn, the situation is more likely to be improved there. Regardless, we do not see a notable reduction in the number of fatal shootings over years that we would expect if there was a reduction in the likelihood that an officer with a body camera would be involved in a fatal shooting.

We then looked at the data at a more granular level for a few states. According to an evaluation of body camera programs by the National Institute of Justice (NIJ), there was a statistically significant reduction in police use-of-force in California and Nevada [7]. In CA, we observed that the percentage of incidents with body cameras (orange line) increased from 2015 to 2019, and the total number of fatal shootings declined steadily. This would support that the adoption of body cameras could be related to the decline in fatal shootings. In NV however, we do not see the same trend. Once again, results from only 2 states are not representative of the other states.

Hence, given the data at hand we are not able to conclude that deploying body cameras will largely improve the police use-of-force. The average deployment rate at 13% of body cameras is quite small and the body camera policies vary in different agencies – more data and research are needed.

Conclusion

Research Question 1: Is there a correlation between the racial composition and the rate of police fatal shootings?

There is a strong correlation between racial composition and police shooting fatality rates in the US based on this dataset alone. At a high level, minority races such as Black and Hispanic Americans are killed at a higher rate in comparison to White Americans.

Black Americans experience greater poverty levels compared to White Americans per capita in certain communities. In addition, Black Americans are arrested more frequently for violent criminal activities than White Americans. [9] Future improvement opportunities would be to deep dive into certain neighborhoods or communities and normalize crime rates committed by race.

Research Question 2: Are US police shooting fatalities lower in 2020 due to the COVID-19 pandemic?

The second hypothesis is that police fatal shootings have been reduced in 2020 due to COVID-19 lockdown measures. We do not observe strong evidence in support of the hypothesis using both temporal and cross-sectional analysis, and thus cannot reject the null hypothesis that reduction in mobility has no meaningful impact on police fatal shootings

Research Question 3: Will adoption of body cameras help to reduce fatal shootings of police officers?

By analyzing our dataset and referencing all the current studies, we hesitate to conclude that body cameras can effectively reduce police misconduct although in California, it shows positive impact. More research is needed because the average deployment rate of body cameras is quite small and the body camera policies vary in different agencies

Statements of Contribution

- **Ruby Han:** Introduction, Research Question 1 Analysis, Dataset Structure, Data Exploration and Cleaning, Conclusion and Group Discussions
- **Thomas Gao:** Research Question 2 Analysis, Data Exploration and Cleaning, Conclusion and Group Discussions
- **Ruilan Liu:** Research Question 3 Analysis, Data Exploration and Cleaning, Conclusion and Group Discussions