

---

# A Comparative Analysis on Police Related Deaths and Prediction of 2020 Presidential Election

Bon-A Koo<sup>1</sup>, Jana Choe<sup>2</sup>, Yeseo Kim<sup>3</sup>

<sup>1</sup>Northfield Mount Hermon School, Gill, United States

<sup>2</sup>The Governor's Academy, Newburyport, United States

<sup>3</sup>Lakefield College School, Ontario, Canada

## Email address:

bona.koo23@gmail.com (Bon-A K.), jchoespes03@gmail.com (Jana C.), yeseokim4@gmail.com (Yeseo K.)

## To cite this article:

Bon-A Koo, Jana Choe, Yeseo Kim. A Comparative Analysis on Police Related Deaths and Prediction of 2020 Presidential Election.

*International Journal of Data Science and Analysis*. Vol. 6, No. 4, 2020, pp. 105-112. doi: 10.11648/j.ijdsa.20200604.12

**Received:** August 19, 2020; **Accepted:** September 3, 2020; **Published:** September 10, 2020

---

**Abstract:** The recent death of George Floyd once again reminded the Americans of the chronic racial bias when it comes to police using force during an encounter with an alleged criminal or, in some cases, innocent civilians, and promulgated Black Lives Matter (BLM) movements in the United States. In order to verify such police use of excessive force against a particular racial group, we examined datasets regarding cases of police killings, which were collected from 50 states (and Washington, D. C. separately) across the country. To find out the possible factors that might cause frequent police killings against a particular racial group, we analyzed relevant datasets, observing each state's demographics, political ideology, education level, and the frequency of police deaths in respect to each state's frequency of police killings. Although we found numerous factors that might lead such trends in police violence, we discovered a correlation between a state's political ideology and the frequency of police killings of a particular racial group in the corresponding state. In response to such trends, we evaluated the correlation between each state's prevalence of police killings and its presidential election outcome in 2016. Using two machine learning methods, random forest and logistic regression, we further predicted each state's prospective preference toward a particular candidate (Republican or Democrat) and the election outcome of the 2020 presidential election.

**Keywords:** BLM, Police, Violence, Data Analysis, Machine Learning

---

## 1. Introduction

In response to the recent death of George Floyd and the overall perpetuation of police brutality against African-Americans, Black Lives Matter (BLM) movements are extensively propagating in the United States. BLM is seemingly the largest movement in US history, as about 15 million to 26 million people have participated in demonstrations by the week of July 3, 2020 [1].

The term "Black Lives Matter" has been used since July 2013, when a Black community organizer expressed her rage on Facebook about the acquittal of George Zimmerman, a white police officer, in the shooting of Trayvon Martin, a Black teenager. The phrase gained prominence once again in August 2014 when Michael Brown was shot by a white police officer in Ferguson, Missouri. In November 2014, the use of the hashtag accelerated dramatically when the police

officer in question was released, free of charges. Ever since, through social media posts and protests, BLM marks a continuous presence in the States, frequently sparked by racial bias and police killings, underlining the severity of police brutality against the Black population [2].

Undeniably, police violence has long been a major issue in the United States, being one of the leading causes of deaths for young men. Analysts have estimated a mortality rate of 1.8 per 100,000 for young men of ages 25-29; although the number seems insignificant, it is only a few steps behind other causes of deaths such as heart disease (7.0 per 100,000) or cancer (6.3 per 100,000). Researchers have also estimated that the Black men are at a greater risk, being 2.5 times more vulnerable to police violence compared to white men [3]. Considering that the police officers are government employees paid for the protection of the citizens, the exceptionally high rate of police-related fatality, especially among the people of color, seems preposterous.

According to recent research, the Black Lives Matter movement, along with protests against police brutality, had a noticeable correlation with the 2016 election. People who felt "coldly" towards the BLM movement were 66% more likely to vote for Trump than were those who felt "warmly" towards the campaign [4]. The current President of the United States, then a Republican candidate, had expressed support for police and accused the BLM campaign of "dividing America." [5] On the other hand, the Democratic candidate Hillary Clinton publicly announced her support for the movement. Seemingly, in the 2016 election, one's sentiment towards the BLM movement had a clear link with one's vote for a particular candidate, each of whom had a distinct stance towards the issue. In the US today, the progression of statewide BLM movements is similar to that of 2016, only with a greater magnitude. Moreover, prospective candidates of the 2020 election, Joe Biden and Donald Trump, are maintaining identical stances on the issue to those of 2016 candidates, Clinton and Trump respectively. Due to these commonalities, Kevin Drakulich, an associate professor of criminology and criminal justice at Northeastern University, anticipates BLM movements of 2020 and their primary cause, police brutality, will be "important factors in this election - as they were in 2016." [4]

The study aims to identify possible stimuli for police killings and discuss how they might be used for justifications; verify the frequency of police killings against the Black people and the general population in different states; and explore the connection between the prevalence of police killings in a state and the region's unique characteristics such as the education level, demographics, political preference, and frequency of police deaths. The study also strives to anticipate the impact of current BLM movements on the upcoming election of 2020 based on the observation made on each state's characteristics and the results of previous elections.

We acquired relevant data sets from the *US Census Bureau*, *FiveThirtyEight*, and *The Washington Post* for data concerning traits of each state, cases of police death, and cases of police killings respectively [6, 7]. We analyzed the datasets using pandas software library, matplotlib plotting library, and scikit-learn machine learning library.

## 2. Methodology

### 2.1. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) refers to an approach to analyze data sets to find out all main parts of context, often using visual methods. The outcome is usually statistically calculated or visualized to tell the prospect of certain events [8].

In the following sections, we analyze the rate of police killings in respect to race, state, political affiliation, and the number of police deaths.

#### 2.1.1. Justifications of Police Killings

In some or many cases, police killings could be justified by the threat level of the victim, who is also likely an alleged criminal. Thus, we identified three critical factors that might have contributed to the killing: criminal charges of the victim,

the victim's attempts to escape, and the presence of weapons with the victim.

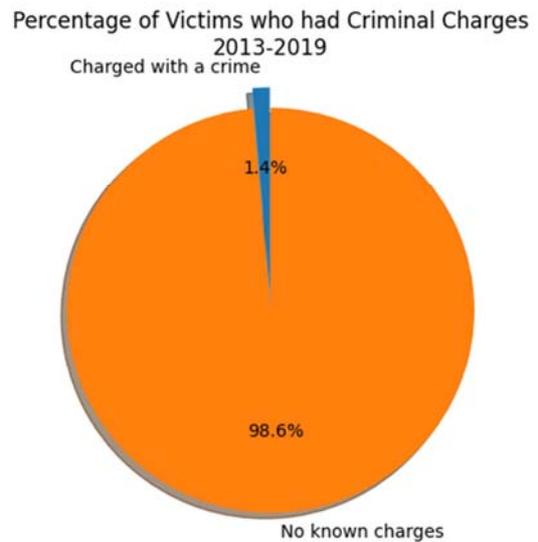


Figure 1. Percentage of victims who had criminal charges.

As shown in Figure 1, only 1.4% of the victims were charged with a crime while the rest was free of charges. However, it is hard to determine whether the police were aware of the victim's criminal charges during the encounter; moreover, a criminal charge cannot account for the immediate threat level of the victim.

Whether the victim was armed during an encounter could define the threat level of one. As shown in Figure 2, 70.9% of the victims were "allegedly armed," possessing a weapon of any type, while 6.7% had a vehicle. 14.2% of the victims were unarmed which questions the necessity of the police to use violence during the encounter, which oftentimes leads to death.

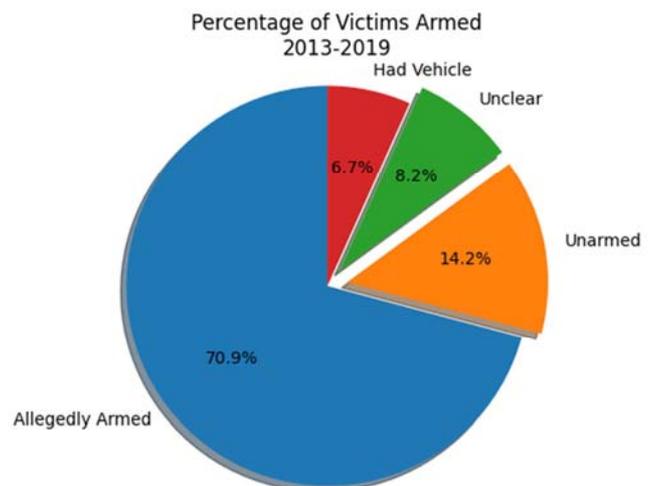


Figure 2. Percentage of victims armed.

Though not a direct indicator of threat level, whether the victim attempted to flee could evince the necessity of Police to use physical force during an encounter. As shown in Figure 3, 20.3% of the victims were fleeing during the encounter either by car, foot, or other means of

transportations. Nonetheless, 42.8% of the victims were not fleeing during an encounter.

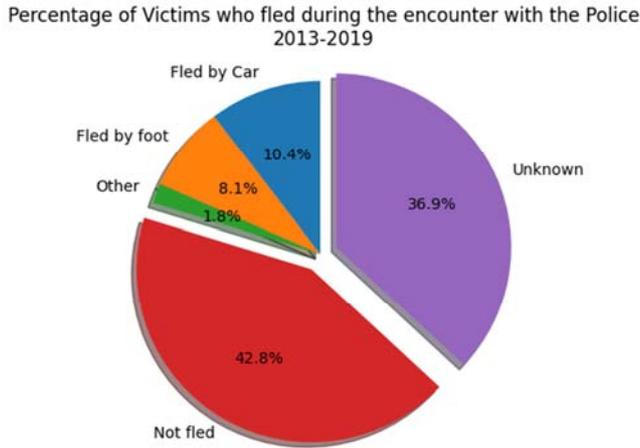


Figure 3. Percentage of victims who fled during the encounter with the police.

2.1.2. Police Killings by Race

We first investigated the total police killings in the United States by race. Figure 4 below shows the percentage of police killings that happened to each race group.

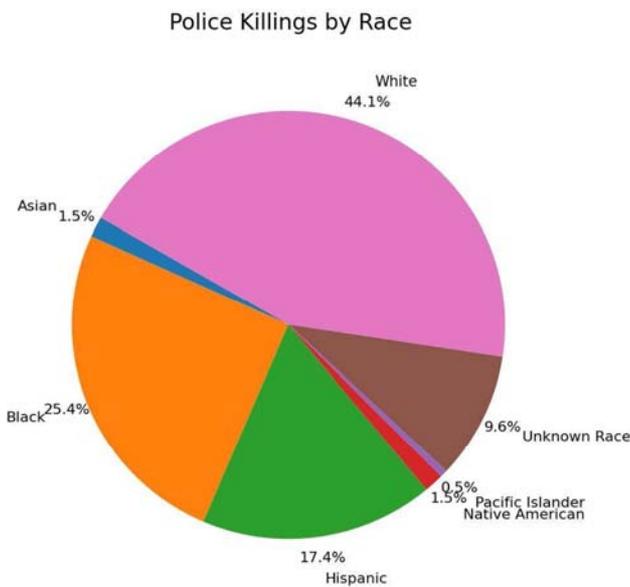


Figure 4. Police killings by race.

According to the figure above, the most number of killings happened to white (44.1%), followed by Black (25.4%) and Hispanic (17.4%). However, as there are significant differences among the population of each race group, we also aimed to investigate the percentage of people killed by the police out of the total number of population by dividing the number of police brutality by population for each racial group.

In the total rate of police killings, the percentage of the Black population outnumbered other racial groups by recording 0.000048%, when other racial groups recorded 0.000017% (white), 0.000023% (Hispanic), and 0.000006% (Asian). The rate of police brutality also depended on the

victim's gender and age. The number of police's extreme usage of force over men was 21.8 times greater than over women. The police brutality seems to have marked the highest in the age group of 18~49 with 5968 cases, followed by 50~69 (1207 cases), 0~17 (154 cases), and 70~109 (128 cases). Thus, the number of cases depended on various kinds of factors, including but not limited to, the victim's race, gender, and age.

We further investigated the trend of victim's races in relation to states' political ideologies. In order to select two representative states from each party (democrat and republican), we examined the approval rating from 2016 voter turnout. Based on the data, California and New York had the highest electoral votes cast of 55 and 29 among the democratic leaning states, while Tennessee and Indiana had the highest electoral votes among the republican leaning states with the cast of 11 [9]. As a result, we selected California and New York for the democratic leaning states, and Tennessee and Indiana for the conservative leaning states.

Figures 5 and 6 represent police killings by race in two liberal states, California and New York.

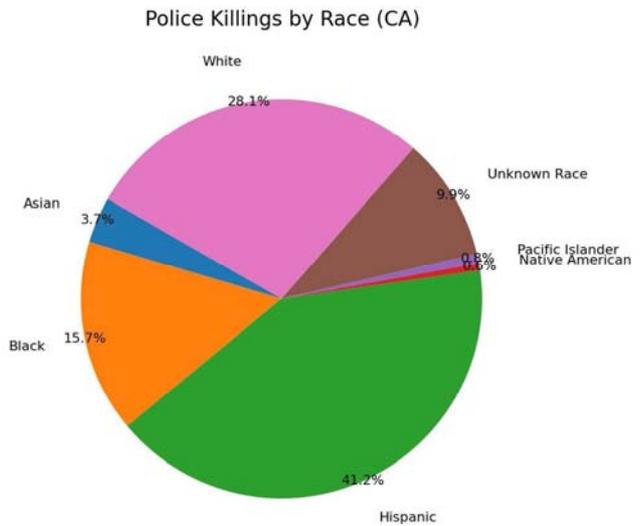


Figure 5. Police killings by race (CA).

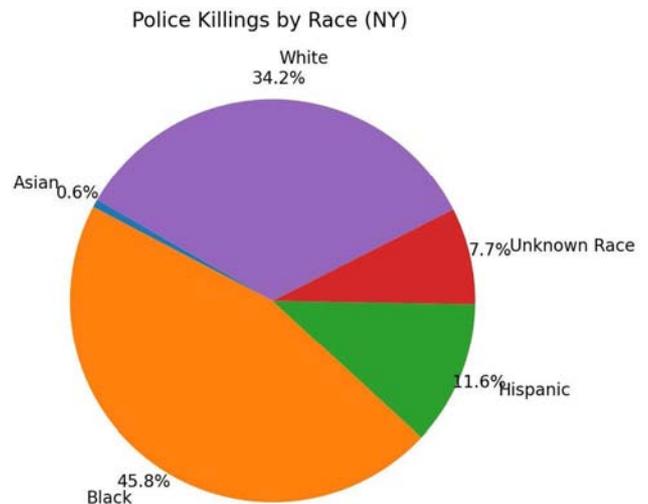


Figure 6. Police killings by race (NY).

The data set of California conveys that the most killings happened to Hispanic group (43.2%), followed by White (28.1%), and Black (15.7%), and the data set of New York conveys that the most killings happened to Black group (45.8%), followed by White (34.2%), and Hispanic (11.6%).

The Figures 7 and 8 represent police killings by race in two conservative states, Tennessee and Indiana.

by races. For example, in the states such as California and Indiana, the police killing rates in black groups were approximately 5 times larger than in white groups, showing that the rates are in fact relatively higher in black groups than in other racial groups.

Police Killings by Race (TN)

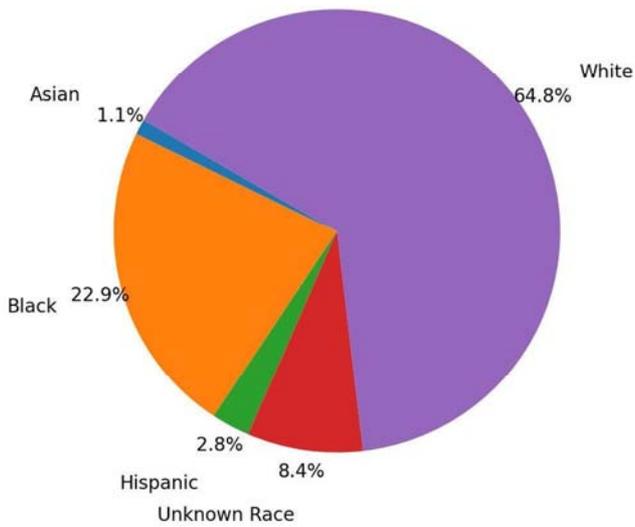


Figure 7. Police killings by race (TN).

Police Killings by Race (IN)

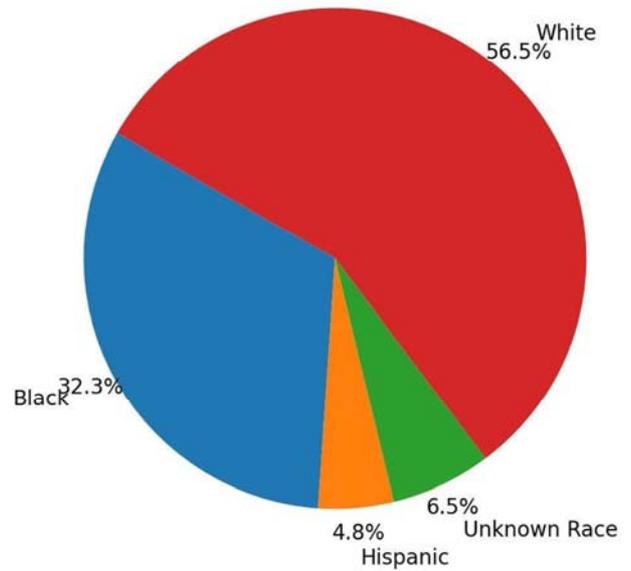


Figure 8. Police killings by race (IN).

The data set of Tennessee conveys that the most killings happened to whites (64.8%), followed by Blacks (22.9%); the data set of Indiana conveys that the most killings happened to whites (56.5%), followed by Blacks (32.3%).

**2.1.3. The Rates of Police Killings by Population**

As shown in the figures in 2.1.2, the number of police killings happened most to the white population in the two conservative states. However, given the fact that white people are the most populated group throughout the four states, the population of four races has to be taken under consideration in order to determine the actual police killings

**2.1.4. Police Killings by Frequency of Police Death**

As we analyzed where and to whom police killings occur most frequently in the states, we have to explore why. As mentioned in 2.1.1, we understand that the threat level of the victim may account for police killing. Expanding this idea further, we inquired whether the frequency of police deaths (number of police officers killed during an encounter) could influence that of police killings.

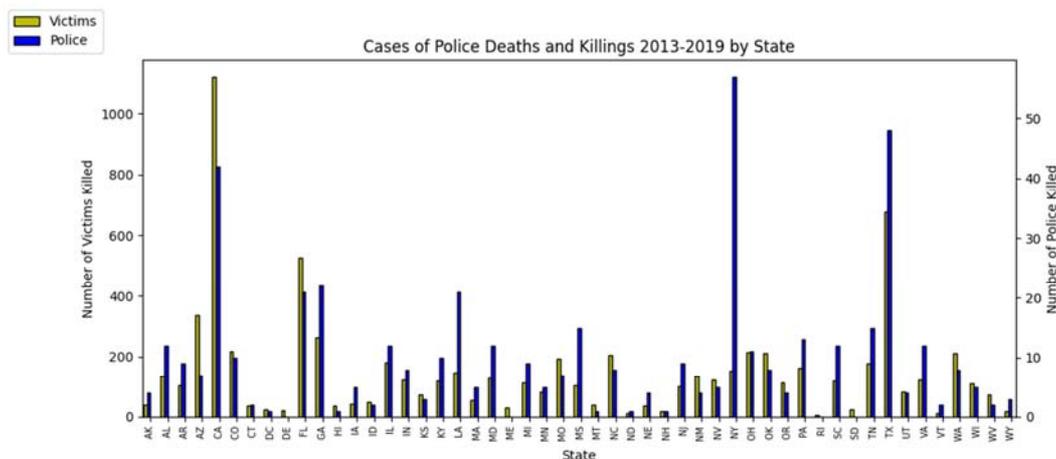


Figure 9. Cases of police deaths and killings 2013~2019.

As shown below in Figure 9, there is a high positive correlation between the frequency of police death and that of killings. There are anomalies like Arizona and California where more victims were killed compared to other states with similar numbers of police deaths. On the other hand, states like Georgia, Louisiana, Mississippi, New York, and Texas had more police killed compared to other states with similar numbers of victims killed by police.

**2.1.5. Police Killings by State and Political Preference**

In order to clarify the connection between public awareness of police killing and election results, we tallied the total number of police killings from 2013 to 2019 by state, which was then labeled by its general preference for Trump or Clinton in the 2016 election (this information was

determined based on the final result of the election, not exclusively on the popular vote) [10].

As shown in Figure 6, California had the most cases of 1123, which is almost twice as many as those of Texas with 677 cases. Florida had the third most cases of 526.

Overall, each state had 147 cases on average. The median value was 113 while the modes were 36, 124, and 136. The range was 1117 as Rhode Island had the least cases of 6.

When taking California out of consideration as an anomaly, the relatively democratic states majority of whom voted for Clinton had 91 cases on average. On the other hand, the apparently conservative states majority of whom voted for Trump had 151 cases on average.

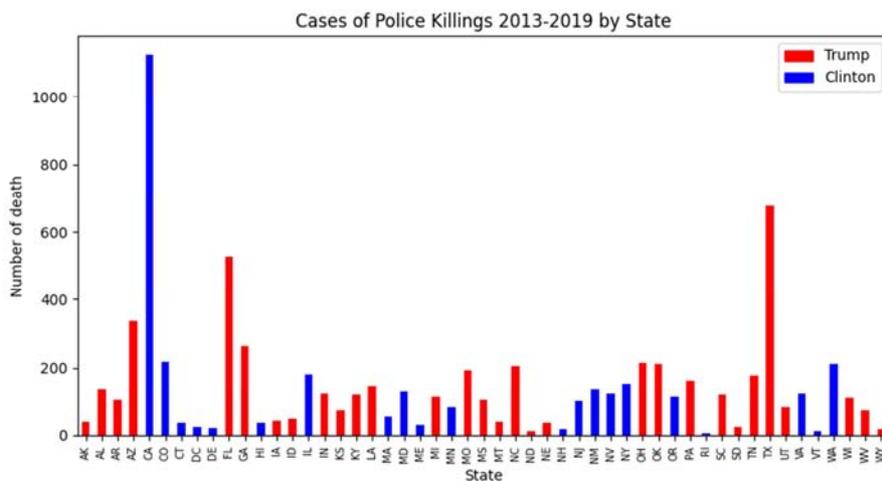


Figure 10. Cases of police killings 2013~2019.

To further inquire about the trend of a state’s political preference in relation to the frequency of police killings against black population, we examined the results of the 2012 and 2016 presidential elections. We identified states that underwent a change in their prevailing political stance over the course of the two elections. Accordingly, there were no states, whose political inclinations were changed from republicans to democrats from 2012 to 2016, whereas there were five states (Florida, Minnesota, Ohio, Pennsylvania,

West Virginia) that changed their political preferences from democrats to republicans [11].

In order to identify the patterns of police killings that target the black population in these five states, we calculated the ratio of police killings of black group to white group for all 50 states. While the average ratio was 2.6403 across all states, that of Florida was 2.8546, Minnesota 3.8503, Ohio 5.4738, Pennsylvania 6.7933, and West Virginia 3.9017, all of which are higher than the average ratio.

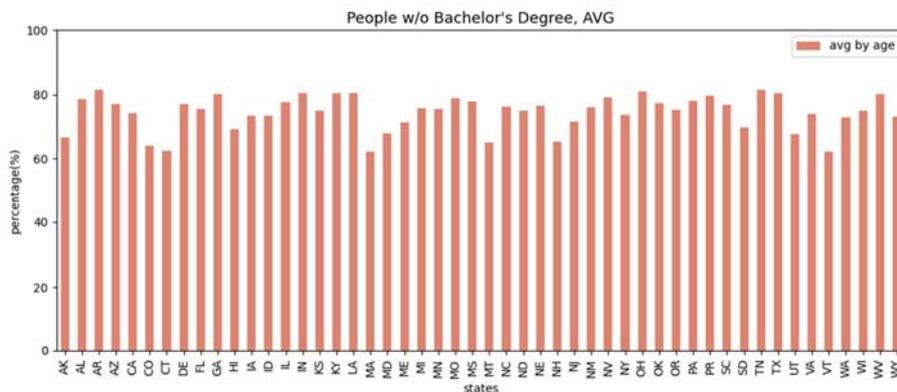


Figure 11. People without Bachelor's degree (Average).

**2.1.6. States Characteristics**

For a further analysis, we examined the level of educational attainment to see the possible correlation of the educational rate and the rate of police killings in the individual states. We determined the education level based on the percentage of people who have not attained a bachelor’s degree from ages 35 to 65 in each state.

The average rate regardless of gender or race ranged from 62.1% (MA) to 81.7% (TN), marking a difference of 19.6% (see Figure 11). In Massachusetts, an estimated 3.14 per 10,000 black men were killed, in comparison to 0.53 per 10,000 white population. In Tennessee, an estimated 3.68 per 10,000 black population were killed, compared to 2.54 per 10,000 white men.

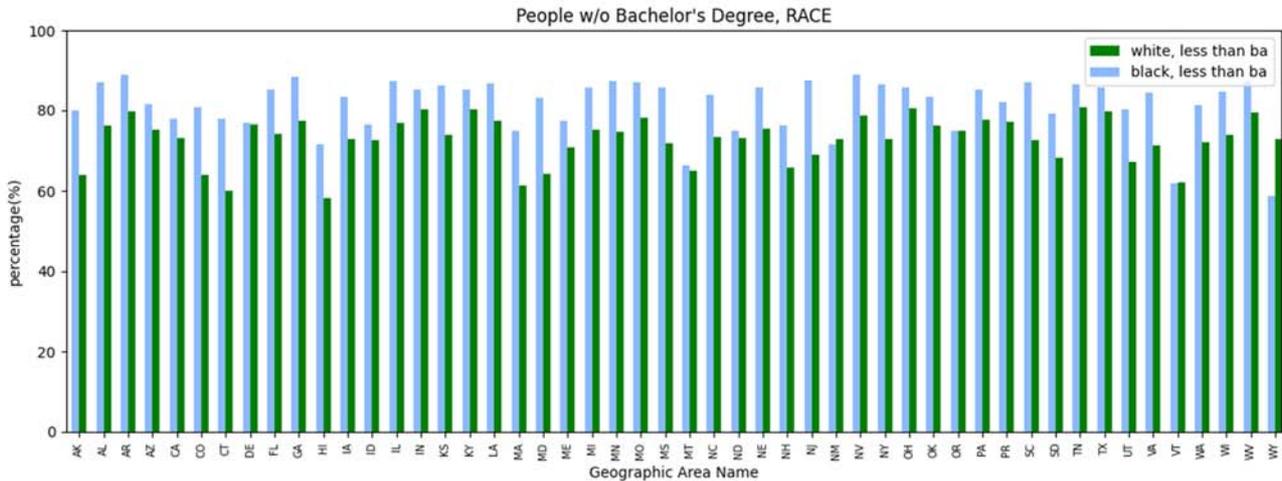


Figure 12. People without Bachelor's degree (Race).

We also obtained an analysis on the educational rate of the same group classified into different races. The average rate ranged from 58.6% (WY) to 89.1% (NV) for the black population, and from 58.1% (HI) to 80.8% (TN) for the white population. An average of 7.91 per 10,000 black men and 3.9 per 10,000 white men were estimated to be killed by police in Nevada.

Minnesota marked the greatest difference in the educational rate between the two races (rate of black population being 25.1% higher). Utah, on the other hand, presented the smallest difference of 0.1% (rate of white

population being higher) between the educational rate.

**2.2. Machine Learning**

Machine learning, a well-known subset of artificial intelligence, refers to the automatic learning of a computer through experience. During this process, a computer studies numerous patterns within a data set and creates its own algorithms to classify or predict the value of a new piece of data [12].

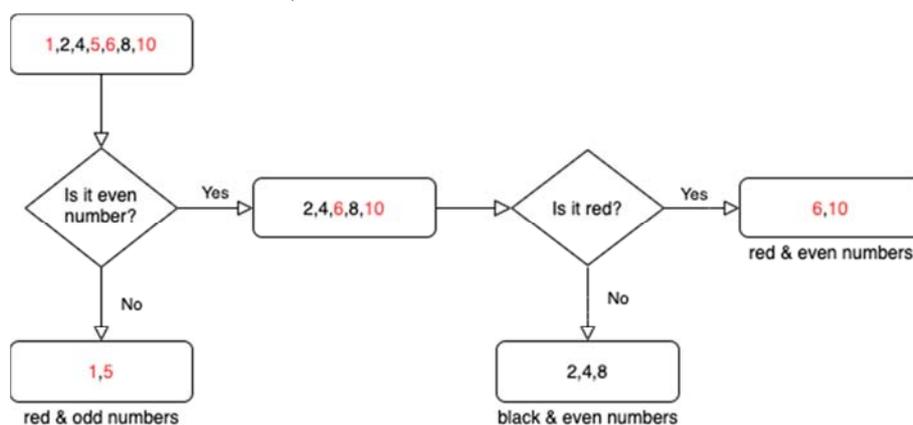


Figure 13. Flowchart of decision tree algorithm.

**2.2.1. Background**

In the second part of our study, we used machine learning as a means to determine the most likely 2020 election result of a state, based on each region’s unique characteristics -

including the frequency of police killing - and previous election results.

In order to obtain the most accurate result, we used a Voting Classifier, a type of machine learning. A Voting Classifier is an example of so-called Ensemble Learning,

using multiple algorithms to derive the final output. We used Soft Voting in particular, in which the probability vectors for each predicted class are summed up and averaged. The class with the highest value wins, becoming the final output [13].

Out of many algorithms used in a Voting Classifier, we specifically used two models: Random forest and Logistic Regression.

The random forest algorithm consists of multiple “decision making trees” which splits the input data into subgroups based on its features. For example, in Figure 13, the tree is trying to classify the 7 different numbers given. The system only uses 2 questions to derive final subgroups: red & odd numbers, black & even numbers, and red & even numbers. In order for the tree to have the most effective design, each subgroup needs to be as distinct as possible, while the components of each have to be as similar as possible [14].

An ensemble of these “trees” create a random forest. Each individual tree comes up with a class prediction and the most voted one becomes the model’s final prediction. [14]

Contrary to the random forest, logistic regression first applies linear regression and then applies logistic function to investigate the probability of an event occurring. The prediction is based on the higher probability of either class [15].

Though more complex, we used the same mechanism to classify each state into prospective preference towards a particular party (Republican or Democrat) in the upcoming 2020 election, by taking education level, demographics, and most importantly, frequency of police killings into account.

### 2.2.2. Results and Predictions

According to the machine learning data that we obtained, democrats will most likely win the presidential election in 2020, gaining support from 27 states, while 24 states for republicans. Compared to the election results in 2016, 7 states— Indiana, Kansas, Michigan, Mississippi, Montana, North Carolina, and North Dakota— changed from Republicans to Democrats, while Minnesota changed from Democrats to Republicans. Taking each state’s number of pledged electors into account, Democrats earned 276 pledged electors while Republicans earned 262 pledged electors. Still, the results leaned to Democrats. In addition, we identified the chances of the states’ political preference for all 51 states (Washington and D. C. separately). As a result, we found out that 10 states— Connecticut, Georgia, Kansas, Montana, New Jersey, North Carolina, North Dakota, Pennsylvania, South Carolina, South Dakota— each won its political preference only by a small difference, indicating higher chances of incorrect predictions in the actual 2020 presidential election.

## 3. Discussion

Analyzing the rate of the police’s use of excessive force regardless of the victim’s race, we delved into possible causes of this occurrence and how it correlates with the environment of each state. Initially, we found that the black citizens were more likely to get killed by a police officer than

were those of other racial groups. Then we recognized a correlation between a state’s political preference and its frequency of police killings. We also identified a high correlation between a state’s frequency of police killings and police deaths. The more were police officers killed, the more they killed civilians. Though the education level of a state had a negligible correlation with the prevalence of police killings in a state, former observations provided us a clue that the public sentiment, whether political preference or the tension between police and civilians, acted as important factors to the outbreak of police violence in a state.

## 4. Conclusion

Using machine learning to comprehensively analyze the datasets, we proposed a prediction on the result of the upcoming 2020 presidential election: the democrats will win by 14 votes of pledged electors ahead of republicans.

The calculated accuracy of this machine learning mechanism was 80%, which could have been improved if we had used more sets of data from the past apart from those of the 2015-2016 which we used so that it forms “training data” along with the result from the 2016 election. Moreover, for a comprehensive yet concise analysis, we could have used data sets concerning individual cities rather than states. By the same token, considering the current public outcry against police violence in the US, we could have included the frequency of protests in a region as an indicator of its holistic sentiment towards police brutality, thereby further developing the analysis of a state’s political preference.

## References

- [1] Buchanan, Larry, et al. “Black Lives Matter May Be the Largest Movement in U. S. History.” *The New York Times*, 8 July 2020, [nyti.ms/2D8PhQY](https://nyti.ms/2D8PhQY).
- [2] Anderson, Monica. “Social Media Conversations About Race.” *Pew Research Center: Internet, Science & Tech*, Pew Research Center, 30 May 2020, [www.pewresearch.org/internet/2016/08/15/social-media-conversations-about-race/](https://www.pewresearch.org/internet/2016/08/15/social-media-conversations-about-race/).
- [3] Williams, Joseph P. “Study: Police Violence a Leading Cause of Death for Young Men.” *U. S. News & World Report*, U. S. News & World Report, 5 Aug. 2019, [www.usnews.com/news/healthiest-communities/articles/2019-08-05/police-violence-a-leading-cause-of-death-for-young-men](https://www.usnews.com/news/healthiest-communities/articles/2019-08-05/police-violence-a-leading-cause-of-death-for-young-men).
- [4] Callahan, Molly. “Many People Who Voted in 2016 Were Motivated by the Black Lives Matter Protests. Will the Same Hold True This Year?” *News Northeastern*, 9 June 2020, [news.northeastern.edu/2020/06/09/the-black-lives-matter-protests-motivated-voters-in-2016-will-they-do-the-same-in-2020/](https://news.northeastern.edu/2020/06/09/the-black-lives-matter-protests-motivated-voters-in-2016-will-they-do-the-same-in-2020/).
- [5] Siddiqui, Sabrina. “Donald Trump Strikes Muddled Note on 'Divisive' Black Lives Matter.” *The Guardian*, Guardian News and Media, 13 July 2016, [www.theguardian.com/us-news/2016/jul/13/donald-trump-strikes-muddled-note-on-divisive-black-lives-matter](https://www.theguardian.com/us-news/2016/jul/13/donald-trump-strikes-muddled-note-on-divisive-black-lives-matter).

- [6] "Census Bureau." *Five Thirty Eight*, 15 July 2019, [fivethirtyeight.com/tag/census-bureau/](https://fivethirtyeight.com/tag/census-bureau/).
- [7] *The Washington Post*, WP Company, [www.washingtonpost.com/wp-srv/metro/data/datapost.html](https://www.washingtonpost.com/wp-srv/metro/data/datapost.html).
- [8] Patil, Prasad. "What Is Exploratory Data Analysis?" *Medium*, Towards Data Science, 23 May 2018, [towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15](https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15).
- [9] Daniel Diorio, Ben Williams. "The Electoral College." *National Conference of State Legislatures*, 6 July 2020, [www.ncsl.org/research/elections-and-campaigns/the-electoral-college.aspx](http://www.ncsl.org/research/elections-and-campaigns/the-electoral-college.aspx).
- [10] "Presidential Election Results: Donald J. Trump Wins." *The New York Times*, The New York Times, 9 Aug. 2017, [www.nytimes.com/elections/2016/results/president](https://www.nytimes.com/elections/2016/results/president).
- [11] "President - Live Election Results." *The New York Times*, The New York Times, 29 Nov. 2012, [www.nytimes.com/elections/2012/results/president.html](https://www.nytimes.com/elections/2012/results/president.html).
- [12] "Machine Learning: What It Is and Why It Matters." *SAS*, [www.sas.com/en\\_us/insights/analytics/machine-learning.html](https://www.sas.com/en_us/insights/analytics/machine-learning.html).
- [13] Mangale, Sanchita. "Voting Classifier." *Medium*, Medium, 18 May 2019, [medium.com/@sanchitamangale12/voting-classifier-1be10db6d7a5](https://medium.com/@sanchitamangale12/voting-classifier-1be10db6d7a5).
- [14] Yiu, Tony. "Understanding Random Forest." *Medium*, Towards Data Science, 14 Aug. 2019, [towardsdatascience.com/understanding-random-forest-58381e0602d2](https://towardsdatascience.com/understanding-random-forest-58381e0602d2).
- [15] Shetty, Badreesh. "An in-Depth Guide to Supervised Machine Learning Classification." *Built In*, 17 July 2019, [builtin.com/data-science/supervised-machine-learning-classification](https://builtin.com/data-science/supervised-machine-learning-classification).