Mapping Police Violence

Farrukh Masood · Follow

Published in CS334-Project-Group 2 · 24 min read · Mar 20, 2021

CS 334 — Principles and Techniques of Data Science

Written by Nabeel Aamer, Farrukh Masood, Sayyed Muhammad Ahsan Imran, and Maria Asif.

Introduction

Law enforcement agencies across the US have failed to provide the general population with basic information about the number of lives they have taken. While there is a mandate to have this data reported officially by the police, it will, in reality, take several years before the data is brought to the general public.

In a country where 3 people are killed by police every day, this dataset aims to provide some insights into police violence patterns across the country. It includes information on over 8,000 killings by the police nationwide since 2013. This data only includes active uniformed police officers acting in the law enforcement capacity and does not include killings by vigilantes or security guards.

About the Data

The information meticulously sources from the three largest, most comprehensive, and impartial crowdsourced databases on police killings in the country: FatalEncounters.org, the U.S. Police Shootings Database, and KilledbyPolice.net. While there are undoubtedly police killings that are not included in the dataset (namely, those that go unreported by the media), these estimates suggest that this dataset captures 92% of the total number of police killings that have occurred since 2013.

This dataset will be used to understand key insights about police killings in the hopes of providing greater transparency and accountability for police departments in the United States. We will focus on the following:

> Are there any inherent underlying relationships and biases that influence the Police killings in the United States?

In addition to this, to be able to answer our question, we will explore some topics which are listed below.

Sub-topic 1: How do police determine the threat level, and are there any biases involved?

Sub-topic 2: Is there an inherent difference between how police react in different states, especially when they are divided along party lines such as Republican and Democrat.

Sub-topic 3: Are the police biased towards any particular racial group?

Sub-topic 4: Do significant events such as Black Lives Matter and Election Month have a difference in Police Killings?

To begin with, we need to first clarify what counts as a police killing.

Police Killing: A case where a person dies as a result of being shot, beaten, restrained, intentionally hit by a police vehicle, pepper-sprayed, tasered, or otherwise harmed by police officers, whether on-duty or off-duty.

To start our analysis with the data, we begin by importing the relevant libraries and the excel file dataset into a pandas data frame called Police_killing.

The original data set consisted of 8941 rows and 36 columns. It also contained 3 other sheets, each of which have been discussed later in this blog.

To understand the shape and size of the dataset, we use the following command method.

```
print("The size is: " , Police_killing.size)
print("Number of rows and columns are: ", Police_killing.shape, '\n')
```

```
The size is:  321876
Number of rows and columns are:  (8941, 36)
```

Now, let's look at the different columns in the dataset to understand the nature of the dataset.

```
In [162]: Police_killing.columns

Out[162]: Index(['Victim's name', 'Victim's age', 'Victim's gender', 'Victim's race',
         'URL of image of victim', 'Date of Incident (month/day/year)',
         'Street Address of Incident', 'City', 'State', 'Zipcode', 'County',
         'Agency responsible for death', 'ORI Agency Identifier (if available)',
         'Cause of death',
         'A brief description of the circumstances surrounding the death',
         'Official disposition of death (justified or other)',
         'Criminal Charges?',
         'Link to news article or photo of official document',
         'Symptoms of mental illness?', 'Armed/Unarmed Status',
         'Alleged Weapon (Source: WaPo and Review of Cases Not Included in WaPo Database)',
         'Alleged Threat Level (Source: WaPo)', 'Fleeing (Source: WaPo)',
         'Body Camera (Source: WaPo)', 'WaPo ID (If included in WaPo database)',
         'Off-Duty Killing?',
         'Geography (via Trulia methodology based on zipcode population density: http://jedkolko.com/wp-content/uploads/2015/0
5/full-ZCTA-urban-suburban-rural-classification.xlsx )',
         'MPV ID', 'Fatal Encounters ID', 'order', 'Encounter Type (DRAFT)',
         'Initial Reported Reason for Encounter (DRAFT)',
         'Names of Officers Involved (DRAFT, currently have data for 2020 and 2017 only)',
         'Race of Officers Involved (DRAFT)',
         'Known Past Shootings of Officer(s) (DRAFT)',
         'Call for Service? (DRAFT)'],
        dtype='object')
```

And here is a snapshot of 5 random rows from the data frame to understand how data has been recorded.

```
In [163]: random_sample = Police_killing.sample(n = 5)
          random_sample
Out[163]:
```

	Victim's name	Victim's age	Victim's gender	Victim's race	URL of image of victim	Date of Incident (month/day/year)	Street Address of Incident	City	State	Zipcode	...
5056	Bradley Carey	54	Male	White	http://www.fatalencounters.org/wp-content/uplo...	2016-08-03	17729 Telegraph Rd	Detroit	MI	48219.0	...
8894	Frankie Pitt	45	Male	Black	http://wric.images.worldnow.com/images/2057402...	2013-01-13	Chippenham Parkway ramp and Hull Street Road	Midlothian	VA	23112.0	...
2820	Joseph Robbins	25	Male	White	NaN	2018-08-08	158 Hospital Dr	Carthage	TN	37030.0	...
3422	Name withheld by police	Unknown	Male	Black	NaN	2018-02-02	1018 Cleveland Ave SW	Atlanta	GA	30344.0	...
3340	Timothy Smothers Jr.	31	Male	White	NaN	2018-03-02	13301 GA-133	Quitman	GA	31643.0	...

5 rows × 36 columns

To explore the data, we need to look at the number of missing values in each column. For this, we use the following method:

```
In [168]: Police_killing.isnull().sum()
          # Total values: 8837
TA-urban-suburban-rural-classification.xlsx )       18
MPV ID
241
Fatal Encounters ID
128
order
7814
Encounter Type (DRAFT)
7176
Initial Reported Reason for Encounter (DRAFT)
7644
Names of Officers Involved (DRAFT, currently have data for 2020 and 2017 only)
8275
Race of Officers Involved (DRAFT)
8793
Known Past Shootings of Officer(s) (DRAFT)
8908
Call for Service? (DRAFT)
7642
```

We can see that the last few columns were almost empty, and there are several missing values in other columns. Before moving on to the analysis, it is essential to clean our data to extract meaningful inferences from it.

Data Cleaning

Data cleaning is a pivotal step in the data science cycle. Before we can derive meaningful insights and inferences from our data, we need to validate its correctness and ensure that it is in a standardized and useable format. Thus, we will divide the cleaning process into various stages before moving to the analysis.

Dropping columns

One crucial step is to get rid of empty columns. The following 2 columns are almost empty; hence we will remove them from the data set.

- Off-Duty Killing?
- order

Next, we can notice that the columns at the end are merely rough draft columns (as indicated by their names). Hence we need to drop them too.

- Encounter Type (DRAFT)
- Initial Reported Reason for Encounter (DRAFT)
- Names of Officers Involved (DRAFT, currently have data for 2020 and 2017 only)

- Race of Officers Involved (DRAFT)
- Known Past Shootings of Officer(s) (DRAFT)
- Call for Service? (DRAFT)

Lastly, after careful analysis of the dataset, we can observe that 5 columns are not needed for our analysis, so we need to drop the following columns too:

- URL of image of victim
- Link to news article or photo of official document
- Street Address of Incident
- ORI Agency Identifier (if available)
- Alleged Weapon (Source: WaPo and Review of Cases Not Included in WaPo Database)

We can drop all of these columns using the following method:

```
In [174]:
dfclean = Police_killing.drop(columns = ['Call for Service? (DRAFT)',
                              'Known Past Shootings of Officer(s) (DRAFT)',
                              'Race of Officers Involved (DRAFT)',
                              'Names of Officers Involved (DRAFT, currently have data for 2020 and 2017 only)',
                              'Initial Reported Reason for Encounter (DRAFT)',
                              'Encounter Type (DRAFT)',
                              'order',
                              'Off-Duty Killing?',
                              'URL of image of victim',
                              'Street Address of Incident',
                              'ORI Agency Identifier (if available)',
                              'Alleged Weapon (Source: WaPo and Review of Cases Not Included in WaPo Database)',
                              'Link to news article or photo of official document'])

print("New number of columns: ",len(dfclean.columns))

New number of columns:  23
```

We can observe that we are left with 23 columns in the dataset after dropping the columns mentioned above. Furthermore, the cleaned dataset is now named "dfclean".

Checking values for each column

Now, one by one, we check each column to see its unique values. Since 'Name' and 'Age' need not be categorized, we leave these two columns as they are. For the "Victim's gender" column, we notice that there exist both "Male" and "Male " (with space at the end). Hence, we standardize the two different types into one category — "Male".

```
In [178]: dfclean["Victim's gender"].unique() #Consolidating the Male anomily
          dfclean["Victim's gender"] = dfclean["Victim's gender"].replace('Male ', 'Male')
          dfclean["Victim's gender"].value_counts()

Out[178]: Male           8471
          Female          445
          Transgender       9
          Unknown           6
          Name: Victim's gender, dtype: int64
```

Next, when we observe the "Official disposition of death (justified or other)" column, we see that it has 154 different values!

However, we notice some similarities in them — most of them tell the verdict (justified/not justified) followed by some details. Hence, we apply regex to reduce the number of values. We do so by:

We can now notice that the number of different values has been reduced from 154 to only 10!

This helps us understand the data much better, and it will help us with the visualizations (we need to wait a bit for that part, though).

Similarly, we face a similar issue in the “Criminal Charges?” column. Since we want to know only whether the individual was charged or not, we again apply regex to reduce the number of values from 39 to just 4.

Similarly, we observe the unique values for all the columns one by one and clean data in various ways which are listed.

- Symptoms of mental illness? : There were different spellings for unknown, so they were grouped under “Unknown”.

- Armed/Unarmed Status: This column also had various values, which were grouped under four broad categories — “Unarmed,” “Allegedly Armed,” “Vehicle,” and “Unclear.”
- Three other columns — “Alleged Threat Level (Source: WaPo)”,” Fleeing (Source: WaPo)”, and “Body Camera (Source: WaPo)” were also edited to consolidate data into clearly defined categories.
- The “Cause of death” column had numerous values combined into 6 categories using regex and replace function, as shown below.

Here is how the values were categorized:

Other: [‘Pepper Spray’, ‘Bean bag’, ‘Bomb’,’ Chemical agent/Pepper spray’],

Beaten: Beaten/Bludgeoned with instrument

Asphyxiated: ‘Physical Restraint’, ‘Physical restraint’, ‘Baton, Pepper Spray, Physical Restraint’]

Taser: Tasered’, ‘Taser, Physical Restraint’, ‘Taser, Beaten’,’ Taser, Pepper spray, beaten’,’ Taser, Baton’,’ Taser, Pepper Spray, Beaten’]

Gunshot: ‘Gunshot, Taser’,’ Gunshot, Police Dog’’Gunshot, Pepper Spray’,’ Gunshot, Taser, Baton’,’ Gunshot, Vehicle’,’ Gunshot, Stabbed’,’ Gunshot, Taser, Beanbag Shotgun’,’ Gunshot, Taser, Pepper spray’,’ Gunshot, Unspecified Less Lethal Weapon’,’ Gunshot, Beanbag Gun’,’ Gunshot, Bean Bag Gun’.

Lastly, for this sheet, we need to categorize the victim’s race column since there are a number of values.

Now that we’ve thoroughly cleaned our data for **sheet one**, let’s dive deeper into it.

To begin with, we want to see the distribution of gender when it comes to Police killings. Surprisingly, 94.8% of the victims were males, whereas 5%

of the victims were females. The remaining 0.2% was divided amongst the transgender community and unknown.

By Race:

This graph divides the overall victims by race. In the years 2013 to 2021, roughly 3800 whites and 2250 blacks were killed by the police. Around 1500 Hispanics were also killed, and 1000 people were not identified along racial lines.

When we take population into consideration, the perspective changes. According to the United States Census Bureau, whites account for nearly 75% of the US population but only 43% of police killings victims.

Comparing this to the African Americans, who account for about 13% of the population but 25% of police killings victims.

Cause of death:

When talking about the cause of death for the victims, more than 95% of the people were killed due to being shot at by the police department. The others were killed by being tasered multiple times, killed by a vehicle, choked or beaten to death, and other methods such as poison or bombs.

Official Status of the cases:

It is the police department's duty to follow up on an incident involving a suspect's death, 64% of the cases have yet to be investigated, according to the police departments. In contrast, only 11% of them are justified, and even fewer are cleared. Most of the 8890 cases are still open to date.

Was the suspect armed?

A person was coded as **Unarmed/Did Not Have a Weapon** in the database if they were one or more of the following:

- not holding any objects or weapons when killed
- holding household/personal items that were not used to attack others (cellphone, video game controller, cane, etc.)
- holding a toy weapon (BB gun, pellet gun, air rifle, toy sword)
- an innocent bystander or hostage killed by a police shooting or other police use of force

- a person or motorist killed after being intentionally hit by a police car or as a result of hitting police stop sticks during a pursuit

A person was coded as having a **Vehicle** as a weapon if they were one or more of the following:

- a driver who was killed while hitting, dragging or driving towards officers or civilians
- a driver who was driving and/or being pursued by police at high speeds, presenting a danger to the public

A person was coded as **Allegedly Armed** in the database if they were alleged to have possessed objects or weapons in the circumstances other than those stated above. As seen above, in most cases, the suspects were allegedly armed.

Killed by the police each year:

This graph shows that the number of people killed by the police has been a constant of about 1100 or so killed each year in all of United States, the 2021 year has the lowest number because data is only until the second month of the year (February).

The second sheet contains data of members of specific races being killed by the police. Let's first have a look at the columns.

There are 51 columns and 102 rows in this dataset. Let's see what type of data is stored in the columns to decide our next step. We will do that using the **info** function.

Since most columns contain numbers, first, we need to check for null values in the data.

As we can see, there are numerous null values in the data set, and this is problematic since it will create a hindrance in our analysis. Let's again have a look at the data to understand the issue. Furthermore, since the last two rows contain cumulative statistics for all the police departments, they had to be removed as they did not offer any substantial information.

What we can see is that wherever the count is 0, there is a null value. Hence, we replace the null values with 0s.

After replacing the missing values with 0 within the relevant columns, we note that there are still 3 columns with a huge amount of unreported data. The violent crimes columns for the years 2013–2019 have some unreported values. To account for those cases, we will take the value from the average violent crimes reported column as a substitute. We cannot fill in these values with 0s.

Similarly, values for the total arrests are missing for some years, and these will be accounted for by taking the average values in the Estimated Average Arrests per Year column. However, some police departments(Jacksonville and Raleigh) do not have information on the average arrests per year and the individual arrests for each year. Since there is no appropriate number to replace those NaNs, we will have to drop those two departments from our analysis.

Now, we drop the two rows.

Almost done! We've accounted for all NaN values in this dataset!

Just one more issue. For the Black-white disparity and Hispanic white disparity, a few values tend to infinity. We need to change them to NaNs. This is because as white people were not killed, the denominator is zero, giving invalid values.

Moving on,

The figures above represent the number of people killed in a particular state between 2013 and 2020. The NaNs in this case highlight that over the course of seven years, there were no killings of a particular race of people. We believe that this is a very strong assumption and highly unlikely as it is difficult to imagine 0 incidents over the course of seven years, and therefore we will keep these values as NaNs.

It's time to analyze the police killings. Let's have a look at the data for the different police departments.

Top 10 cities for Number of Killings by Major Police Departments.

We wanted to see the general trend of the number of killings varying in different cities. The bar plot above shows the number of killings by Major Police departments in a descending fashion. The highest number of police killings were reported to be in the vicinity of Los Angeles Police Department.

Note: There is only one major police department in each city.

Top 10 cities for Number of Violent Crimes Reported by Major Police Departments.

We wanted to see the general trend of the number of violent crimes reported in different cities of the USA. The bar plot above shows the number of violent crimes reported to Major Police departments in descending order. New York seems to be home to the highest number of violent crimes reported.

Top 10 cities for the Average Number of Arrests by Major Police Departments.

We wanted to see the general trend of Average Arrests in different cities of the USA. The bar plot above shows the number of Average Arrests reported to Major Police departments in descending order. Again, New York seems to have the highest number of Arrests, and other cities are pretty far behind.

Top 10 cities for Killings per 10,000 arrests by Major Police Departments

This visualization shows a combination of Arrests and Killings by Major Police Departments in different cities of the USA, arranged in descending order. It shows that Irving has the highest number of Killings per 10,000 arrests.

Following this,

The column name for the Black Male Population was renamed to something much simpler.

Moreover, there were no major issues in this dataset, and this did not require cleaning despite the fact that there were two NaN in the row, which contained information about the national average homicide rate.

This visualization shows the average police homicide rate for black men (per 100,000) of 15 police departments in different cities. The dotted black line shows the national average homicide rate for black men in all the police departments in the United States, according to the FBI statistics. This visualization allows us to compare different police departments in terms of their average homicide rate for black men and highlight certain departments that may be discriminating against blacks.

After understanding the entire dataset, we are now ready to begin our analysis for the main problem statement as highlighted above the various sub-topics.

Exploring our sub-questions in detail:

Sub-topic 1: How do police determine the threat level, and are there any biases involved?

When determining the threat level of an individual, the police categorized them as ‘attack’, ‘non-attack’, and ‘other’. The majority were labeled as **‘attack’ (around 65%)**, and the others were either non-attack or unknown. Since they were not labeled as an attack, we assume that it was not an aggressive stance and, therefore it was not an attack. Thus the variables ‘non-attack’ and ‘other’ are combined to **‘non-attack’.** After this categorization of data under the **“Alleged Threat Level (Source: WaPo)”** column, we are left with just 2 categories — **attack** and **non-attack**. Since the target variable is to see who is labeled as ‘attack’ and the data is split in a ratio of approximately **2:1 (attack - 4241, non-attack-2327)**, a good model can be fit on to the data.

In order to analyze what factors come into play when the police determine the threat level of a person, the following factors were chosen:

“Victim’s age”, “Victim’s gender”, “Victim’s race”, “Criminal Charges?”, “Symptoms of mental illness?”, “Armed/Unarmed Status”, “Location_Type”

The reason we chose these factors was that these factors are usually most evident when police killings are analyzed. Along with this, these factors are relevant and will help us determine whether the police are biased in the way it determines the threat level of an individual.

To carry out our machine learning model on this data, we first had to make use of ordinal encoding to ensure we had a classification model. All values for age were also converted to integers.

Next, we split the data into our feature set and label. The label, as mentioned before, was the alleged threat level. To confirm that a correlation existed between the features chosen, we made a correlation matrix.

To split the data set into train data set and test data, we used the unbiased function. The shape of both data sets following an **80–20 split** is as follows:

To fit and test our model, we used **KNN**. The main advantages of using KNN are as follows:

- Easy to interpret
- Quick calculation time
- Simple algorithm
- High accuracy — do not need to compare with better-supervised learning models

- No assumptions about data — no need to make additional assumptions, tune several parameters, or build a model. This makes it crucial in nonlinear data case

The original K value for our model was set at **4.** To test our model with this KNN value, the accuracy, F1 value, and precision scores were calculated as shown below.

In order to make sure our model chose the **optimal value of K**, it was run 40 times to observe the value of K at which the loss is minimized. This value came out to be **8.** We made a correlation matrix to visualize the test statistics.

The test results were as follows:

Here the accuracy predicts how accurate our model is in determining the correct threat level for the individuals. Precision basically is a measure of the values our model correctly predicts as an attacker versus the total predicted values for the attackers. Lastly, as seen **91%** of actual attackers were identified by the model (recall score).

Given this feature set and parameters, our machine learning model predicted the **attackers** more accurately than the non-attackers. We believe this is because of the fact that the attacker had more concrete actions when threatening the police whereas the non-attackers were classified under a very wide set of conditions and therefore our model could not interpret the underlying factors.

This model however is useful as it can quickly help the police determine if an individual will be threatening towards the police or not. Based on our model and the underlying goal, **it is preferable to have false positives** (a non-attacker being predicted as an attacker) rather than false negatives(an attacker predicted as a non-attacker). Hence, false negatives should be kept as low as possible. The intuition behind this stems from the reasoning that it is better than the police to remain vigilant for a non-attacker rather than put their guard down for someone who can actually be a potential attacker.

However, on the other hand, it is troublesome to label a non-attacker as an attacker as it means more serious charges for someone who is less guilty and does not pose a threat. Going by that, it is unfair to jail an innocent.

Hence, considering that our model is **67% accurate**, we can say that while there are other underlying factors which have not been accounted for in this model, the chosen features in fact do determine how the police view the threat level of individuals.

Sub-topic 2: Is there an inherent difference between how police react in different states especially when they are divided along party lines such as Republican and Democrat?

For this subtopic, we must refer to a permutation test to find out whether the null hypothesis holds, or we have to consider the alternative hypothesis. The null hypothesis states that there is no difference between the red and blue states as per the **results of the 2016 US election** between Donald J Trump and Hilary Clinton.

States won by the Republican presidential nominee have been labeled as red whereas the states won by the Democrats are the blue ones. There are two distinct groups available for hypothesis testing, **group A consists of the blue states** whereas **group B consists of red states.** Our test statistic for this comparison would be the difference in the average killing rate of a certain ethnicity (for example black, white & Hispanic). Large values of this statistic will tend to favor the alternative hypothesis and vice versa.

Simulation under the null hypothesis: We know that if the null hypothesis is true, all rearrangements of the killing rates among the two groups should be equal i.e the difference in their mean should be zero or close to it. So, the elements in both groups are shuffled at random to create two more groups which have now been allotted values at random (from the previous groups A and B). In the end, we find the difference in the average value of both groups and label it as the difference in means of the newly created random groups. Now, this process is repeated 100 times to create a whole simulation of values. These values are plotted to form a histogram and the observed value (test statistic) is also plotted on the same histogram. Since the observed value is not far away from zero, we can accept the null hypothesis. This result is valid for all ethnicities tested: Black, white and Hispanic.

Plot A (White People) shows a histogram for the simulated values of the average difference between the two groups (average difference in rate of killing of white people in US). The red line on the plot depicts the actual observed value of the average difference between the red group of states and the blue group of states. Y-axis shows the frequency of simulated values whereas the X-axis shows the difference in means of the two groups. (test statistic)

The histogram visually shows that the observed value is in alignment with the simulated values. **This means that the permutation test had led us to an acceptance of the null hypothesis.**

Plot B (Black People) shows a histogram for the simulated values of the average difference between the two groups (average difference in rate of killing of black people in the US). The red line on the plot depicts the actual observed value of the average difference between the red group of states and the blue group of states. Y-axis shows the frequency of simulated values whereas the X-axis shows the difference in means of the two groups.

The histogram visually shows that the observed value is in alignment with the simulated values. **This means that the permutation test had led us to an acceptance of the null hypothesis.**

Plot C (Hispanic People) shows a histogram for the simulated values of the average difference between the two groups (average difference in the rate of killing of Hispanic people in the US). The red line on the plot depicts the actual observed value of the average difference between the red group of states and the blue group of states. Y-axis shows the frequency of simulated values whereas the X-axis shows the difference in means of the two groups.

The histogram visually shows that the observed value is in alignment with the simulated values. **This means that the permutation test had led us to an acceptance of the null hypothesis.**

Clustering was done to segment the various police departments in major US cities in order to understand the nature of killings carried out by these police departments. Clustering allows us to group certain data points that are similar to each one another in a single group and can help us to identify which points are similar to one another and their characteristics. Euclidean distance was chosen as the distance metric to evaluate the closeness between two points. We chose the **Euclidean distance** because it is relatively easier to compute and provides us the straight line difference between two points.

For this model, we wanted to see the behavior of these police departments with respect to the following two features: **Black-White Disparity and Hispanic-White Disparity.** These features show the disparity in terms of the number of Black and Hispanics killed relative to the number of white people adjusted for the population of each of these groups. Additional features included in the model were the **number of people killed for each of these three races**. For our model, we tried various values for the number of clusters to be made, and ultimately we settled on making **three** clusters. The following graph show our results:

Based on the results we can say the following:

1. Police departments in Republican states such as Texas, Florida, Ohio, Indiana, Florida, Louisiana and Georgia had a high amount of black-white disparity.
2. Democratic states such as Illnios, Washington, New Jersey, Maryland, California and District of Columbia had a high amount of black-white disparity.
3. Some of the aforementioned departments also had significant Hispanic-white disparity.

4. The two police departments with significantly high Hispanic-white disparity were in Long Beach and Philadelphia.

Police Departments in Cluster 0

Police Departments in Cluster 1

Sub-topic 3: Are the police biased towards any particular racial group?

To determine whether the police are biased towards any racial group, we will plot a graph of their respective population percentages and the percentage of killings of that race.

The graph shows the proportion of various races in the United States in terms of the US population and people killed over a span of 8 years. From a population perspective, 61.07% of the US population is White, 12.3% is Black, 17.81% is Hispanic and the remaining 8.82% belongs to other communities such as Asian and Native American. In terms of killings over this span of time, 43.34% of killings were of white people, 24.88% killings were of black people, 17.1% of killings involved Hispanics, and other races accounted for 14.7% killings. It can be inferred from this graph that despite accounting for a smaller proportion of the populations, both Black and other races do suffer disproportionately in terms of the number of people killed by police, whereas whites who account for a whopping **61%** of the population only represent **43%** of the police killings.

The above graph shows the black-white disparity of the **top 50 police departments** in the United States. According to this graph, we see that several big cities in America have a substantial black-white disparity such as Chicago, Minneapolis, Boston, Washington, San Francisco, and New York. All of these major cities have a black-white disparity value of more than **8**, with Chicago and Minneapolis having a value greater than **20.** Overall, this graph indicates that there seems to be a general trend across police departments in various American cities of treating the black population differently.

The map above represents the percentage of black people killed by the police in US states from 2013–2021. There are several key observations that can be made from this map. Firstly, there are several Republican states in the East coast, which have a significant proportion of black people killed. Black killings account for over **53%** of killings in Louisiana and states such as Florida, North Carolina, South Carolina, Mississippi, Alabama and Georgia all had more than **30%** of killings of the African-American population. Certain Democratic states also had a high proportion of black people killed with Illinois, New York, New Jersey and Virginia having a figure greater than **40%**. Moving on, we notice that several states on the West Coast such as California, Oregon and Washington have relatively lower figures of proportions of African American killed. The same also applies to several central Republican states such as Utah, Nebraska and Kansas.

Interestingly, several states have a significantly high percentage of African-American killings relative to their percentage population as shown below:

1. Texas: 22% of killings,11.8% of population
2. Florida: 31% of killings, 17% of population
3. New York: 46% of killings, 26% of population
4. Pennsylvania: 33% of killings, 11.2% of population
5. Illinois: 54% of killings, 14.2% of population
6. Ohio: 39% of killings, 14.3% of population
7. North Carolina: 35% of killings, 22% of population
8. Michigan: 35% of killings, 14.1% of population

Overall it can be said that throughout the US, there has been a substantial proportion of police killings that involve the African-American community and in some of these cases the proportion of Black population in that state is not as large, which may indicate that the police are biased against the black community.

Sub-Question 4: Do significant events such as Black Lives Matter and Election Month have a difference in Police Killings?

To determine whether significant effects have an effect on police killings, we first map out the number of killings by month over the 8 year period under observation to observe any peaks and dips.

As can be seen from the graph above, there are a few significant peaks and dips in the number of killings over the span of **8** years. Specifically, we will explore the following anomalies:

- **Jan — Feb 2013**
- **July 2015**
- **June/July 2020**
- **October 2020**

Looking at the first peak, it turns out that the **lowest number of police killings** during the 8 years under observation were in late January 2013 which corresponds perfectly with the start of the second term presidency of President Barack Obama. Right after this, we note that the graph rises steeply till July 2013 — the time when **Black Lives Matter** movement began.

Moving on to June-July 2015, we can observe a spike in the number of killings. It is interesting to note that this is exactly the same time as when **same-sex marriages were legalized** in the United States and the **LGBT movement** gained momentum.

In mid-2020, in the months of June and July, we observe another rise in killings — this time the number being the **highest number of police killings** from 2013 to 2020. Unfortunately, one of the people killed during this time was George Floyd, whereas 19 other people were shot during the **BLM** protests in June 2020.

The number of killings again decreased sharply following this event, and in fact, this time corresponded to the **second-lowest number of killings** during the 8 years. As can be seen on the graph, this corresponds to October-November 2020, the dates of the US elections 2020. However, it is interesting to note that there was **no major difference** during the presidential elections of 2016.

To conclude, our analysis of the sub-questions provides us with a strong base to answer our main question. While statistically there is no evidence of a bias against a certain community in Republican versus the Democratic States, we do come to realize that there is a disproportion in the number of killings of certain ethnicities **(Black and Others)** when compared to the percentage they make up of the population. Furthermore, while race was one of the factors that determine how the police labels an individual as threatening or not, there are other factors that play a role as well so we can not conclusively determine how much race has a role to play. An in-depth analysis of the findings of our model and hypothesis could be helpful in tracing patterns and behaviors of police killings to lower the biases involved.

Written by Farrukh Masood

Follow

1 Follower · 3 Following

No responses yet

What are your thoughts?

Respond

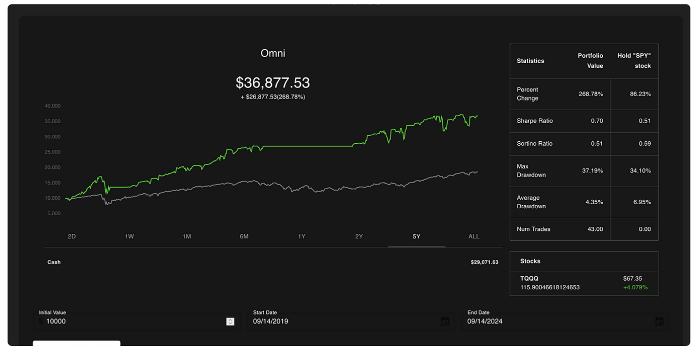
Recommended from Medium

 Jessica Stillman

Jeff Bezos Says the 1-Hour Rule Makes Him Smarter. New…

Jeff Bezos's morning routine has long included the one-hour rule. New…

Oct 30, 2024 · 20K · 538



In DataDrivenInvestor by Austin Starks

I used OpenAI's o1 model to develop a trading strategy. It is…

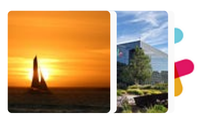
It literally took one try. I was shocked.

Sep 15, 2024 · 8.5K · 216

Lists

Staff picks

803 stories · 1581 saves

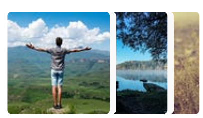


Stories to Help You Level-Up at Work

19 stories · 924 saves

Self-Improvement 101

20 stories · 3238 saves



Productivity 101

20 stories · 2739 saves

In Publishous by Desiree Peralta

OnlyFans is Finally Dead

And I'm happy about it.

Oct 8, 2024 20K 408

In Level Up Coding by Jacob Bennett

The 5 paid subscriptions I actually use in 2025 as a Staff Software...

Tools I use that are cheaper than Netflix

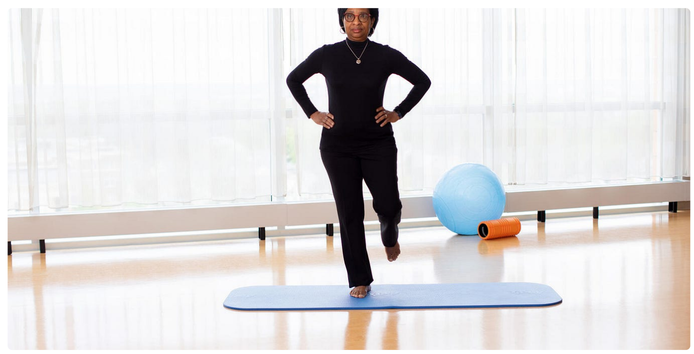
Jan 7 6.3K 130

Harendra

How I Am Using a Lifetime 100% Free Server

Get a server with 24 GB RAM + 4 CPU + 200 GB Storage + Always Free

Oct 26, 2024 8.7K 132



F. Perry Wilson, MD MSCE

How Old Is Your Body? Stand On One Leg and Find Out

According to new research, the time you can stand on one leg is the best marker of...

Oct 23, 2024 21K 547

See more recommendations