Policing the Police

Our Dataset

Merged from the largest databases of police violence in America:
Mapping Police Violence, Deadspin, The Washington Post

Consists of all recorded police killings from 2013 - 2020

Individual-Specific Variables

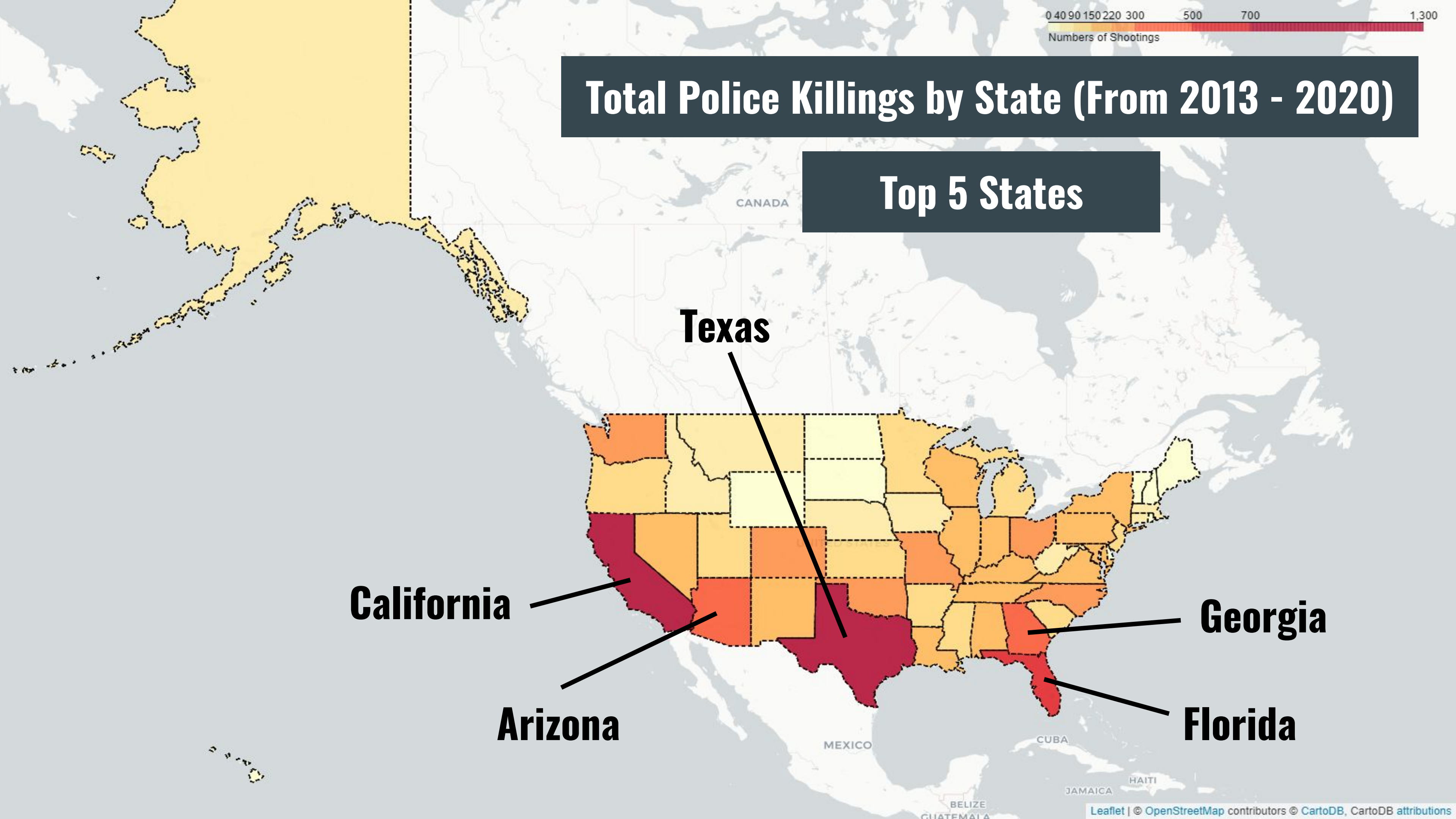
- Age
- Gender
- Race

Location-Specific Variables

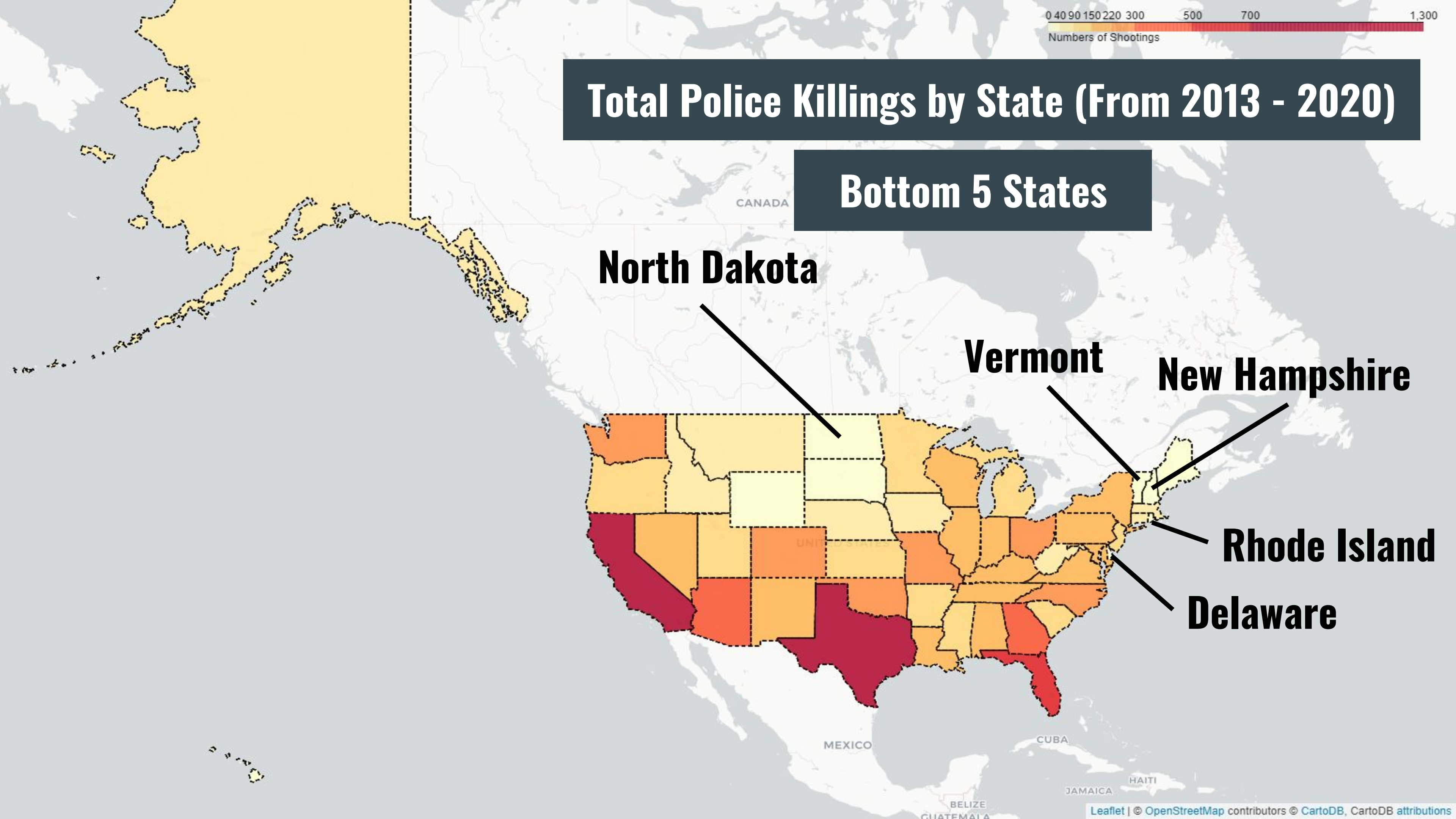
- State
- City
- County
- Geography type: Rural/ Urban/ Suburban

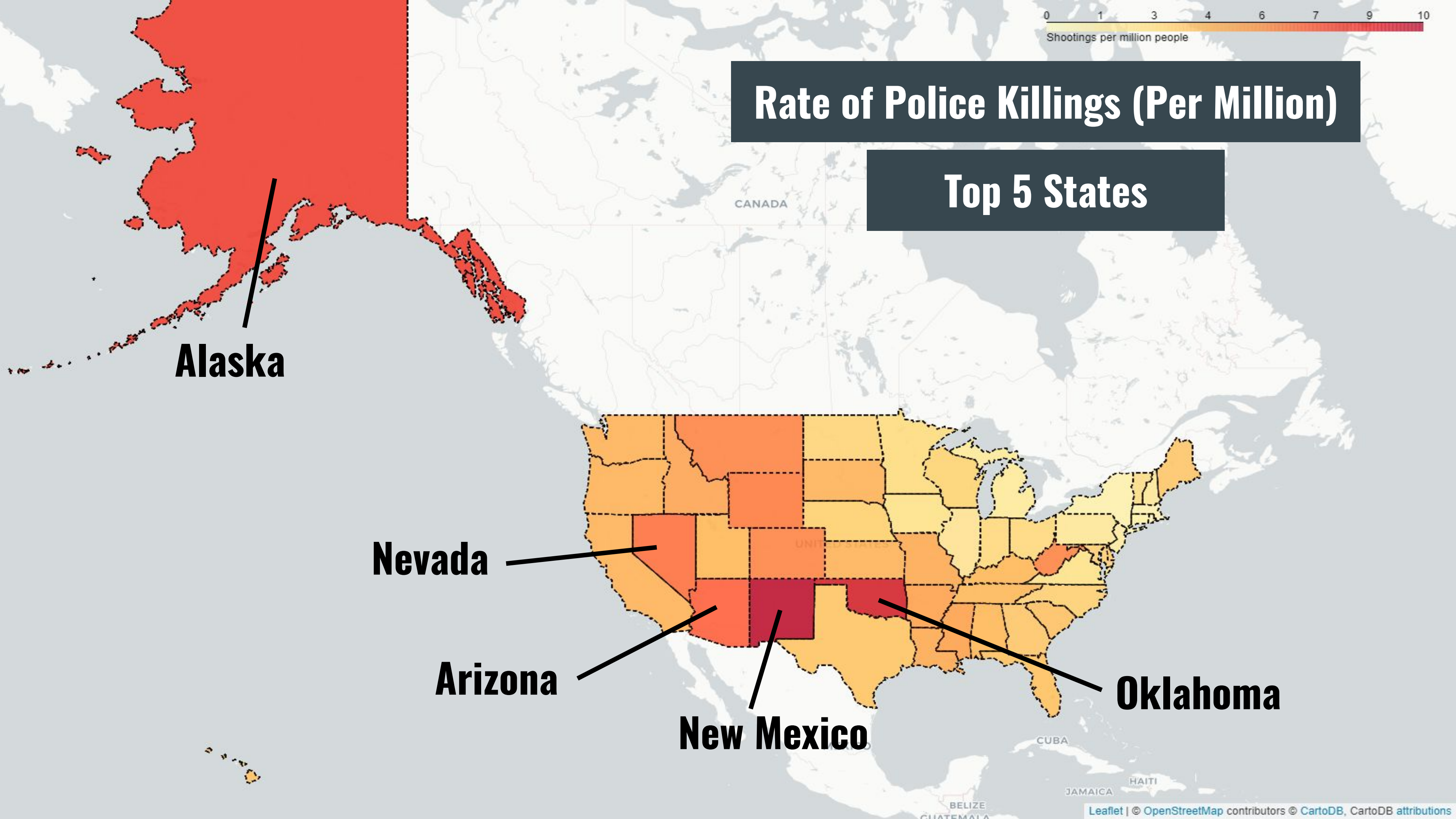
Circumstance-Based Variables

- Weapon on victim
- Whether the victim was perceived to be a threat
- Whether the victim suffered from mental illness
- Cause of death
- Description of death
- Whether the police killing was justified under law
- Police agency responsible

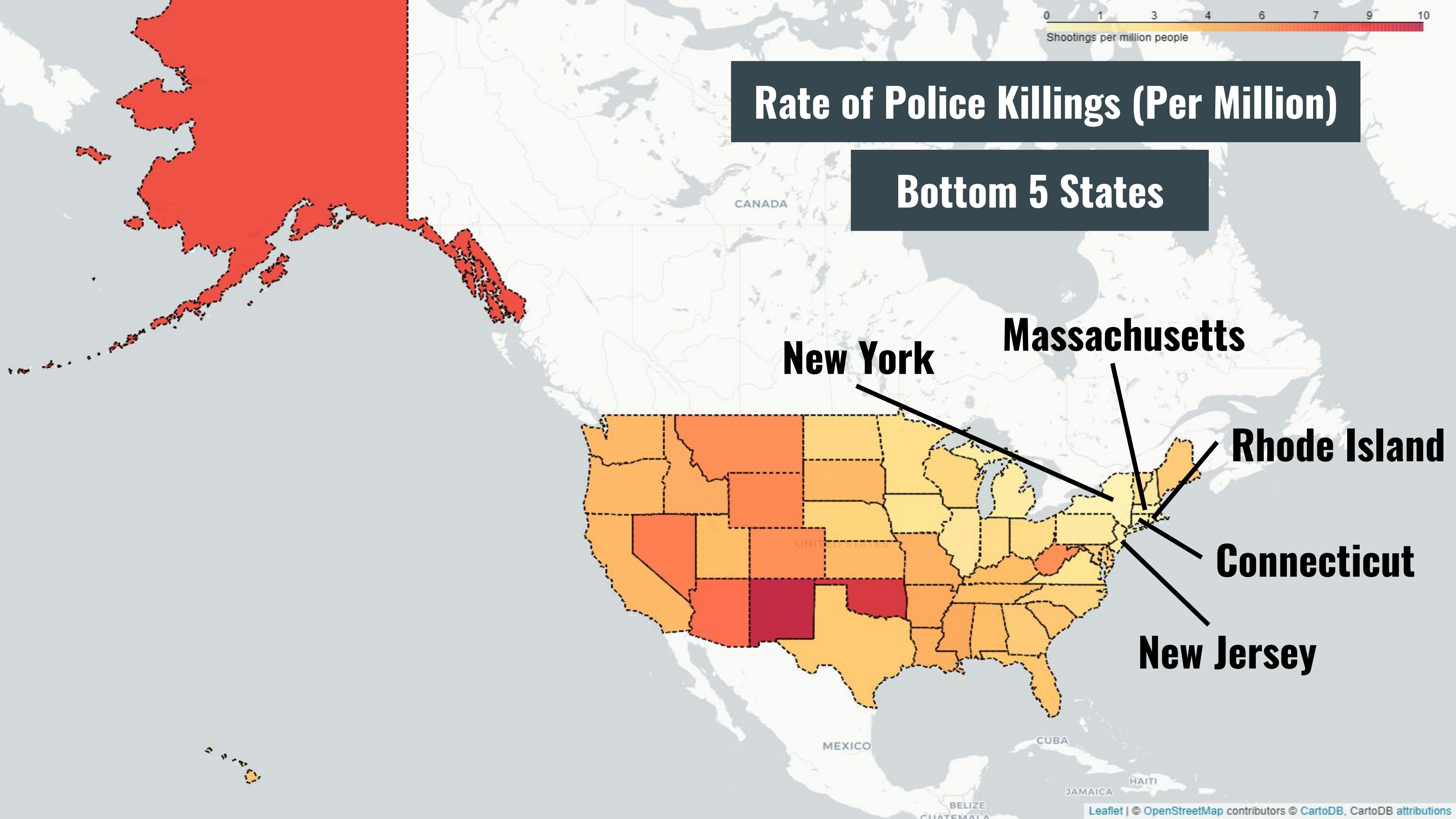


Total Police Killings by State (From 2013 - 2020)
Top 5 States
0 40 90 150 220 300 500 700 1,300
Numbers of Shootings
CANADA
Texas
California
Arizona
Georgia
Florida
MEXICO
CUBA
HAITI
JAMAICA
BELIZE
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions


Total Police Killings by State (From 2013 - 2020)
Bottom 5 States
North Dakota
Vermont
New Hampshire
Rhode Island
Delaware
0 40 90 150 220 300 500 700 1,300
Numbers of Shootings
CANADA
MEXICO
CUBA
JAMAICA
HAITI
BELIZE
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions

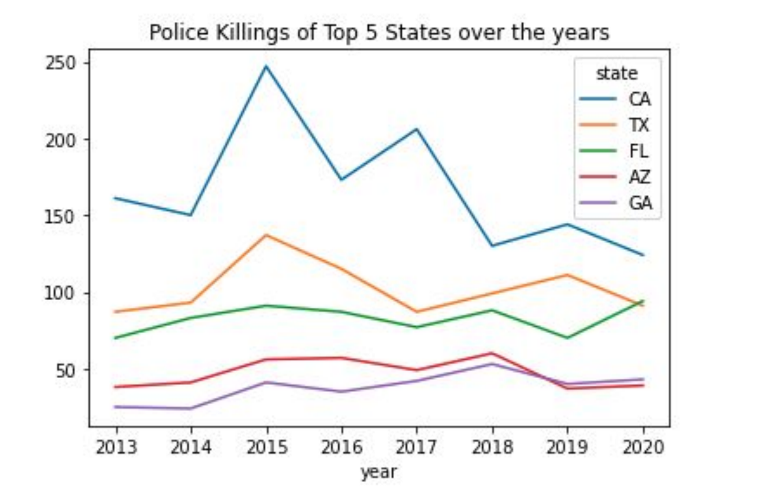


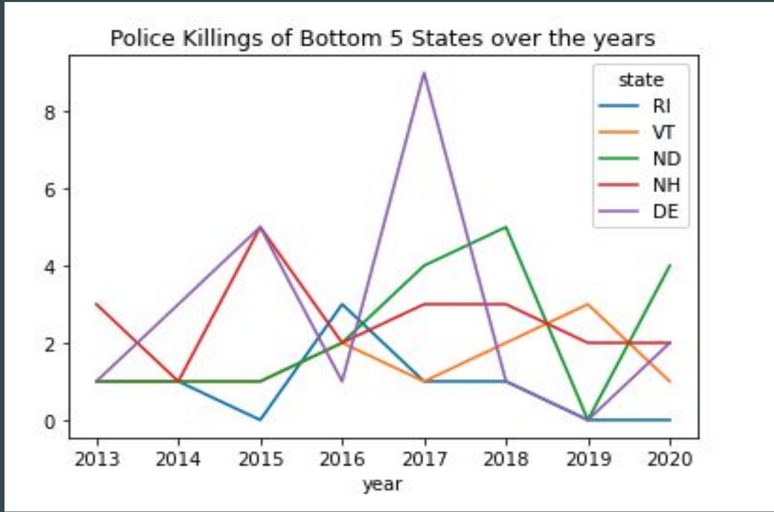
0 1 3 4 6 7 9 10
Shootings per million people
Rate of Police Killings (Per Million)
Top 5 States
CANADA
Alaska
Nevada
Arizona
New Mexico
Oklahoma
CUBA
HAITI
JAMAICA
BELIZE
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions



0
1
3
4
6
7
9
10
Shootings per million people
Rate of Police Killings (Per Million)
Bottom 5 States
CANADA
New York
Massachusetts
Rhode Island
Connecticut
New Jersey
MEXICO
CUBA
JAMAICA
HAITI
BELIZE
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions

How Have Police Killings Changed Over Time?

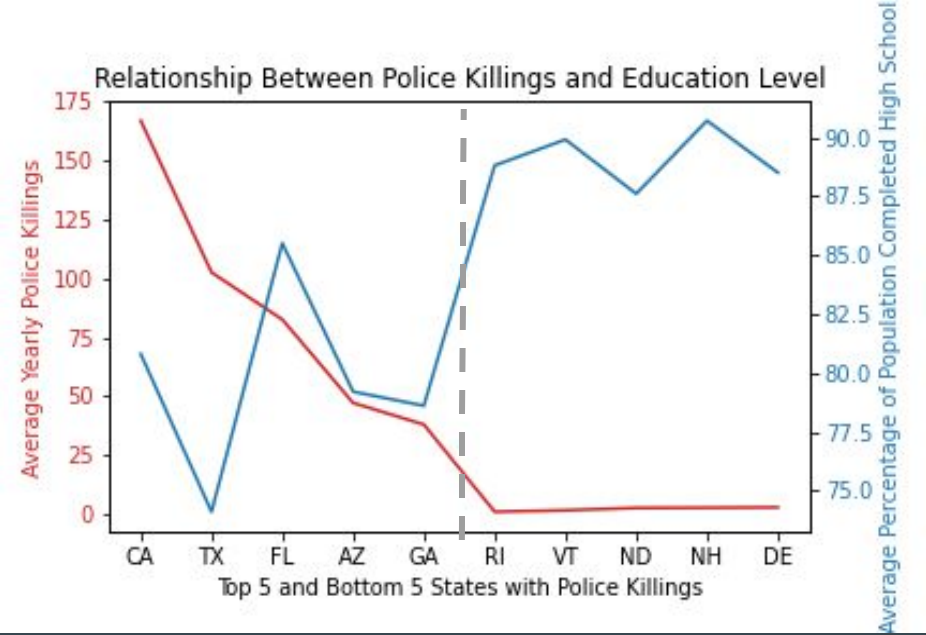




Small fluctuations over the years: Police violence has not improved in these Top 5 states!

Bottom 5 states have stayed relatively low over the years

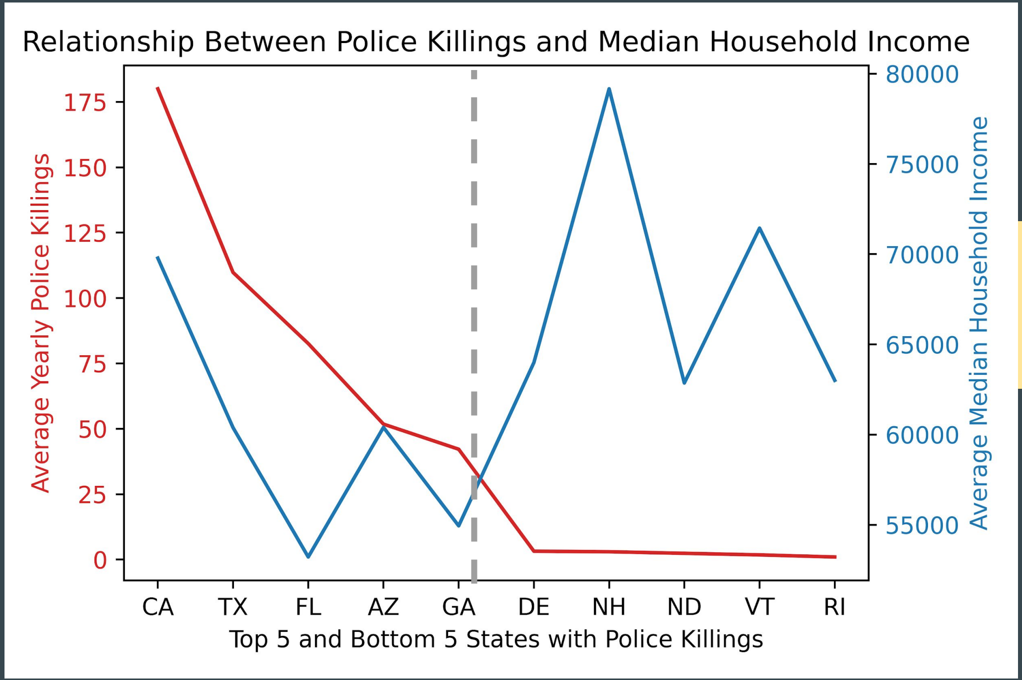
Why Do States Differ In Police Killings?



Negative Correlation:
Higher education level = Lower police killings

Different Education Levels

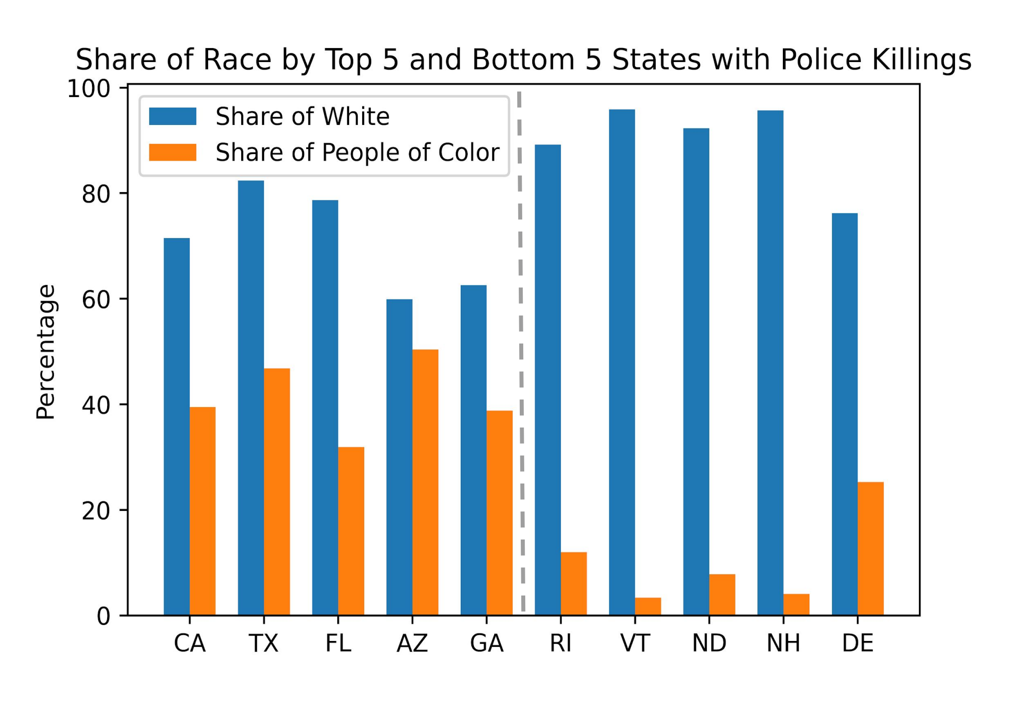
Why Do States Differ In Police Killings?



Negative Correlation:
Higher income level = Lower police killings

Different Income Levels

Why Do States Differ In Police Killings?



Positive Correlation:
Higher share of people of color
= Higher police killings

Different Population Racial Makeup

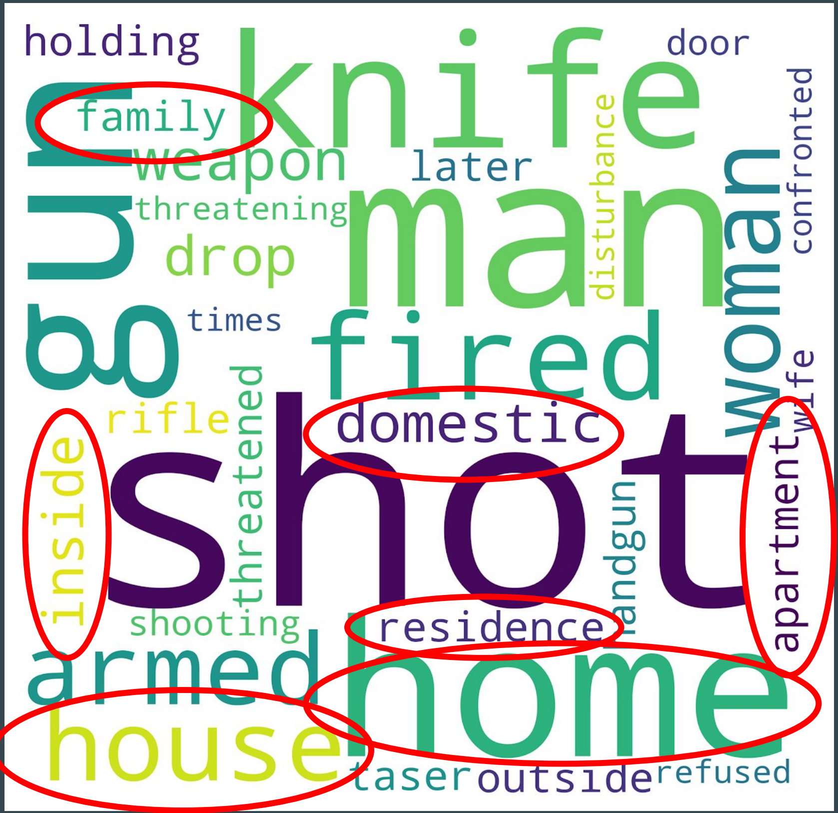
Why Do States Differ In Police Killings?

Create 3 new state-level features:

1) ***state_education_level:*** Proportion of population completed high school in each state
2) ***state_log_income:*** Log of median household income in each state
3) **state_white_share:** Share of Whites in each state

Creating New Features Using Correlations Observed

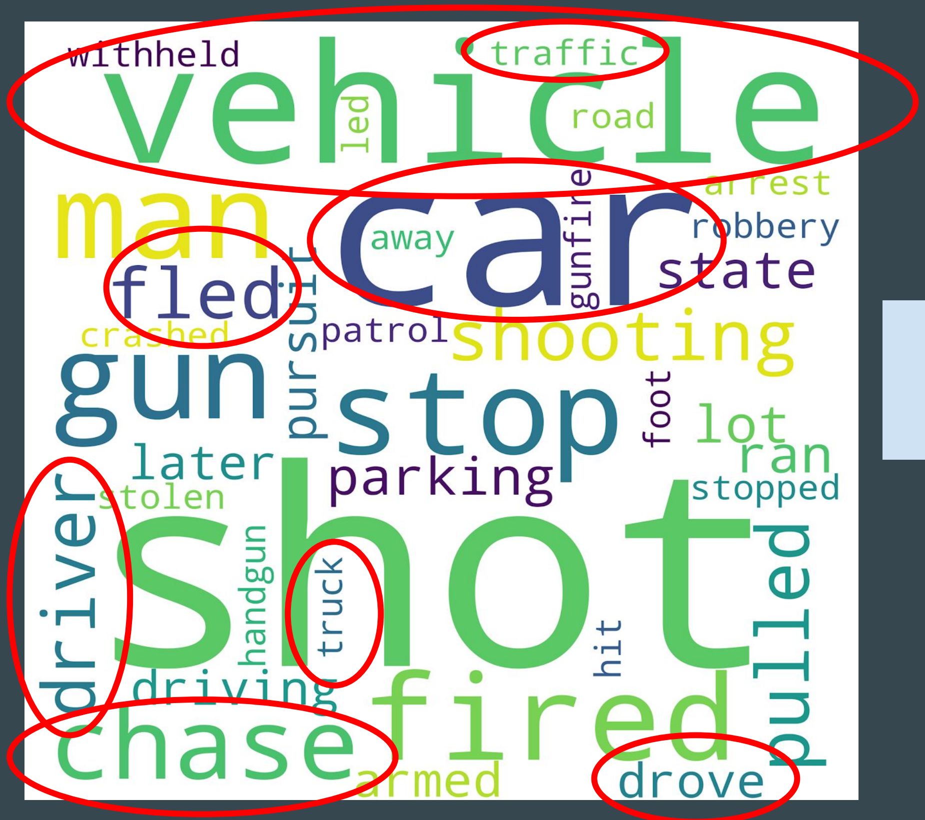
Circumstances Surrounding Police Killings



Circumstance 1: Police killings inside residences

Topic Modelling (LDA) Using *Description of Death* Column

Circumstances Surrounding Police Killing



Circumstance 2: Police killings involving vehicles, where victim is fleeing

Topic Modelling (LDA) Using *Description of Death* Column

Circumstances Surrounding Police Killings

Topic 1: Police killings inside residences

Create new binary feature, ***home_involved***

Topic 2: Police killings involving vehicles, where victim is fleeing

Create new binary feature, ***vehicle_involved***

Create new binary feature, ***was_fleeing***

Creating New Features using Text Mining Results

How Were Victims Killed?

Group into 3 most common causes of death:

- By gunshot
- By taser
- By physical violence (Beaten/ Physical Restraint/ Asphyxiation/ etc)

Create 3 new binary features:

1) ***killed_by_gunshot***
2) ***killed_by_taser***
3) ***killed_by_physical_violence***

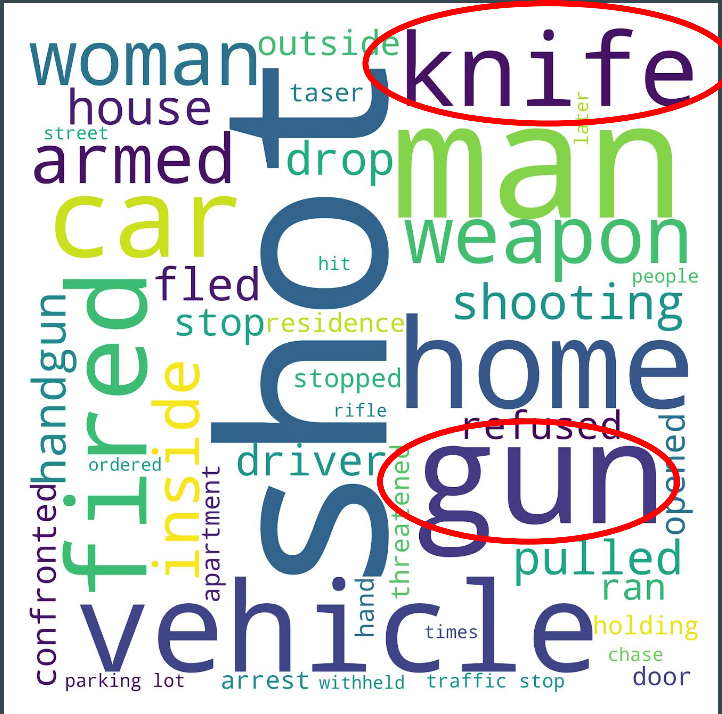
Analysis of *Cause of Death* Column

Gunshot	6982
shot	1437
Gunshot, Taser	237
Taser	221
shot and Tasered	88
Beaten	29
Vehicle	27
Physical Restraint	23
Tasered	13
Physical restraint	9
Asphyxiated	8
Gunshot, Police Dog	5
Other	5
Pepper Spray	4
Taser, Physical Restraint	2
Gunshot, Pepper Spray	2
Taser, Pepper spray, beaten	1
Baton, Pepper Spray, Physical Restraint	1
Bean bag	1
Beaten/Bludgeoned with instrument	1
Bomb	1
Chemical agent/Pepper spray	1
Gunshot, Beanbag Gun	1
Gunshot, Bean Bag Gun	1
Gunshot, Stabbed	1
Taser, Pepper Spray, Beaten	1

Data Transformation to Create Meaningful Features

Were Victims Armed?

Analysis of *Description of Death* Column



Analysis of *Weapon On Victim* Column

Allegedly Armed	5383
Unarmed/Did Not Have an Actual Weapon	1082
gun	886
Unclear	612
Vehicle	505
knife	229
unarmed	103
toy weapon	45
undetermined	40
vehicle	35
unknow	20
sword	9
machet	9
baseba	7
metal	7
gun an	6
hammer	5
hatchet	5
ax	4

We already created a new feature named *vehicle_involved* earlier

Create 2 new binary features:

1) ***armed_with_gun***
2) ***armed_with_knife***

Data Transformation to Create Meaningful Features

Target Encoding: For Categorical Features With Many Levels

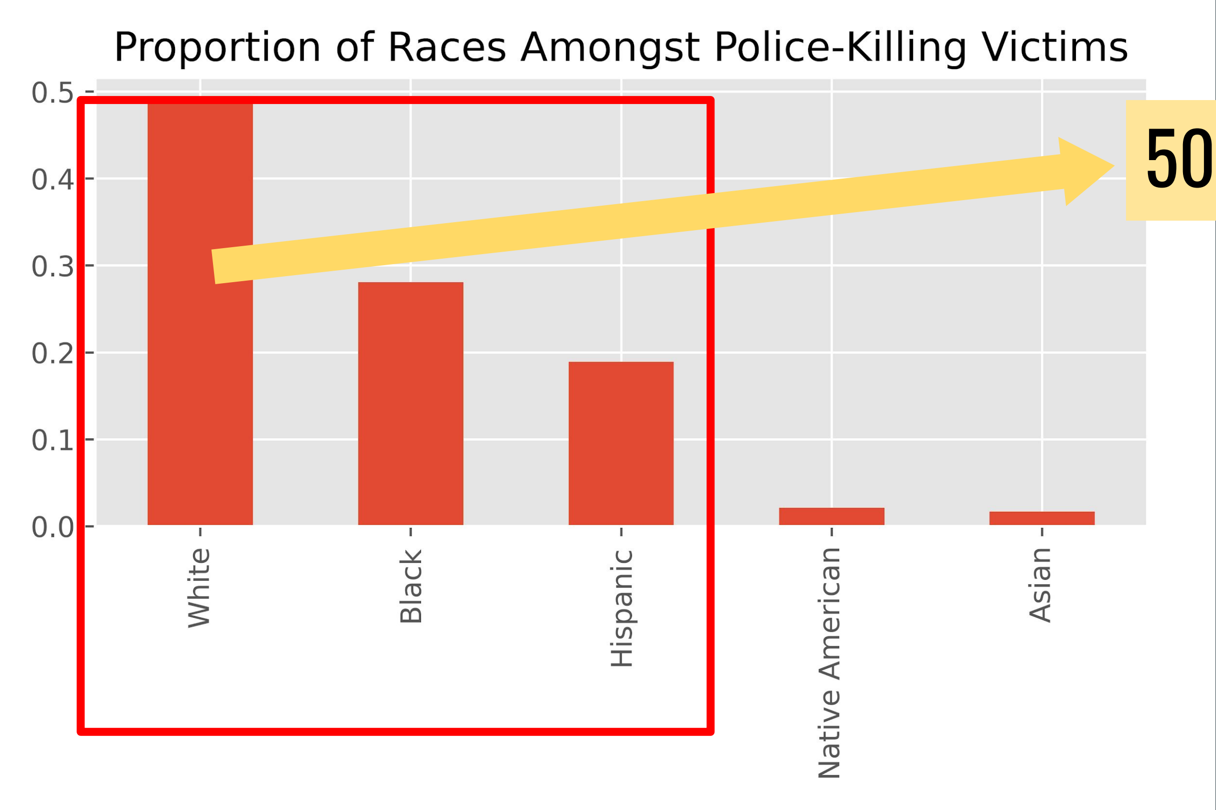
Categorical features encoded:

1) ***state:*** **Consists of 50 unique states**
2) ***police_agency:*** **Consists of 2866 unique police agencies**

What is Target Encoding:
Encode each level with the mean of the target variable for that level

Advantage over One-Hot Encoding:
Does not add to dataset dimensionality

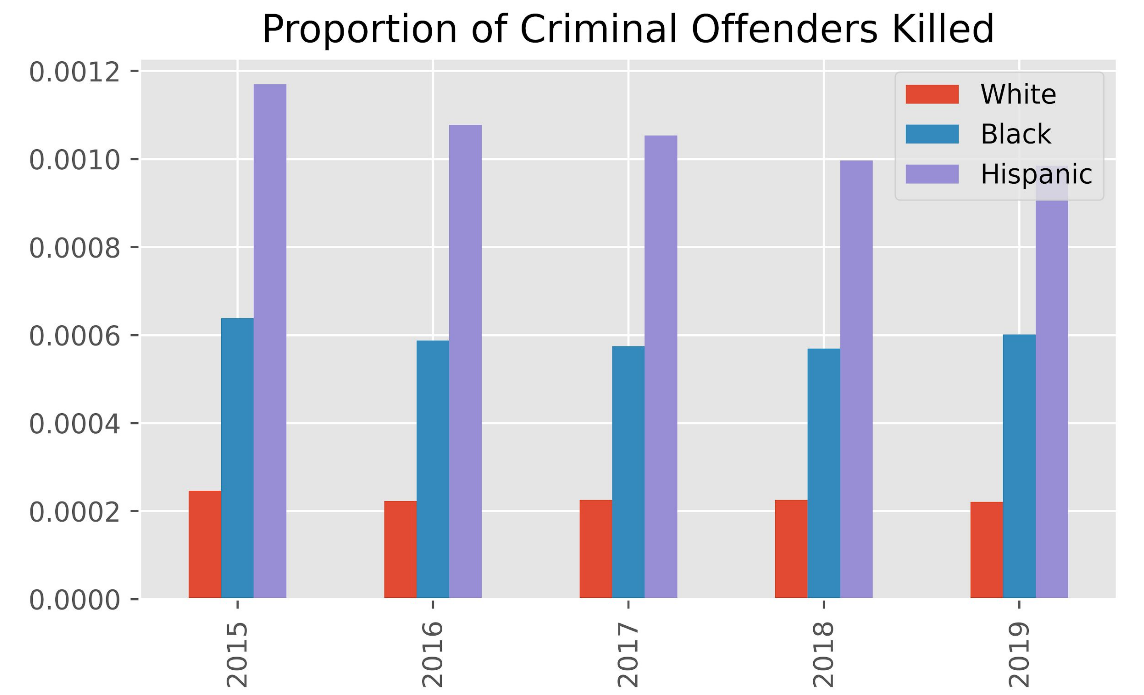
Inspiration for Machine Learning: Does Race Affect How Police Kill?



50% are White Americans!

Does this mean police are **not** racially biased?

Inspiration for Machine Learning: Does Race Affect How Police Kill?



Hispanic people are 4x more likely to be killed than White people.

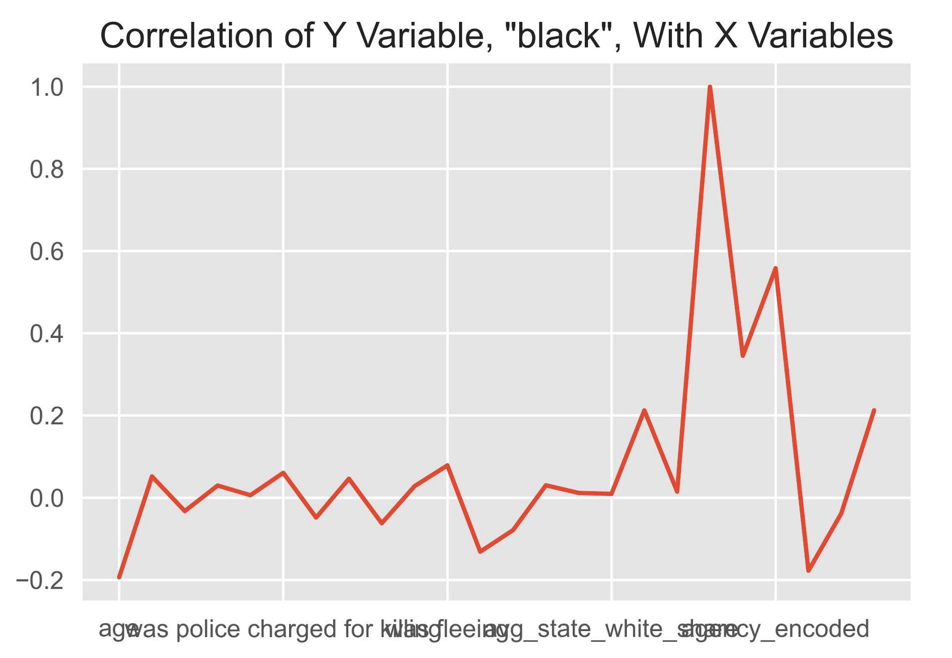
Black people are 2.5x more likely to be killed than White people.

Trend is consistent over time

Predicting the Race of Individuals Killed by Police

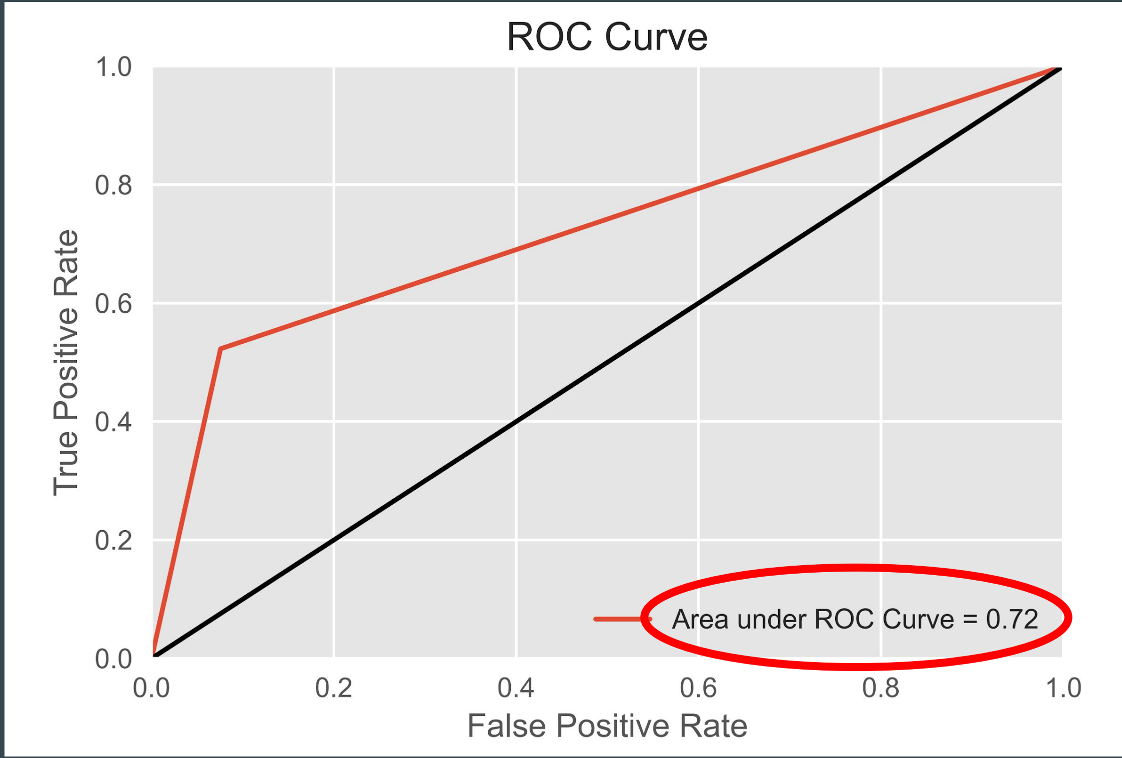
Dependent Variable (Y): ***Black***
Shows whether an individual is Black (0 or 1)

Hypothesis
Given that an individual was killed by the police, there is sufficient difference in the manner they were killed to differentiate whether they are Black



Machine Learning

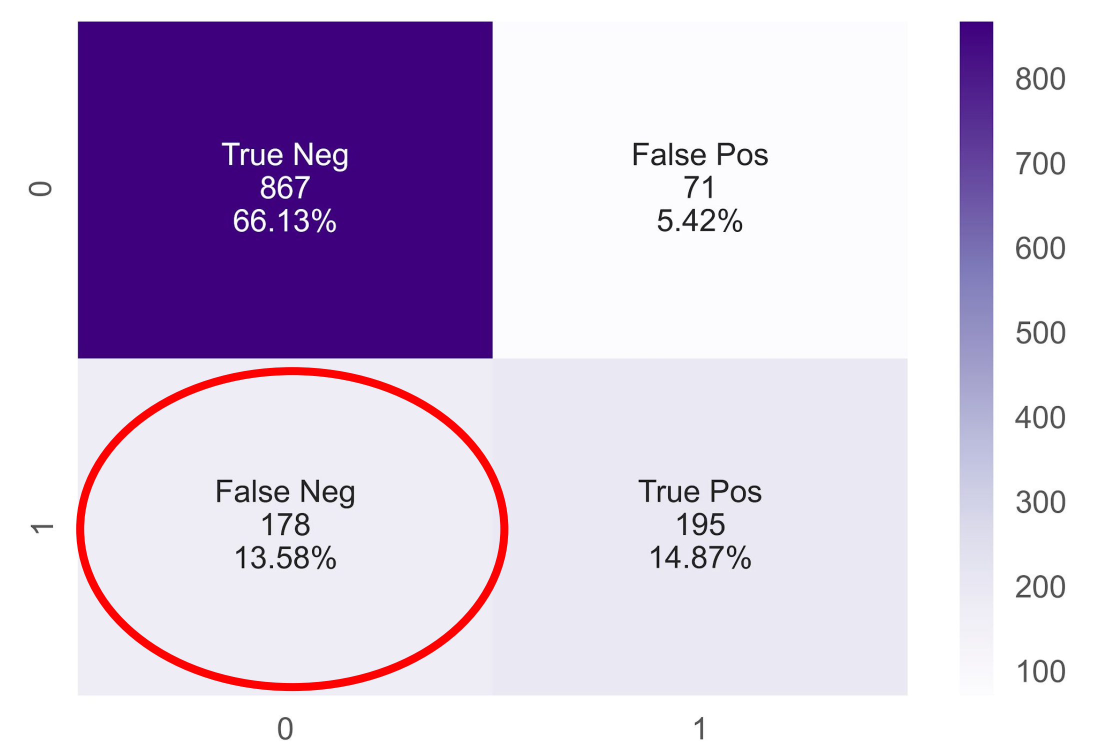
Predicting the Race of Individuals Killed by Police



Accuracy Score: 81%

Model 1: Logistic Regression

Predicting the Race of Individuals Killed by Police



Accuracy Score: 81%

Model 1: Logistic Regression

Predicting the Race of Individuals Killed by Police

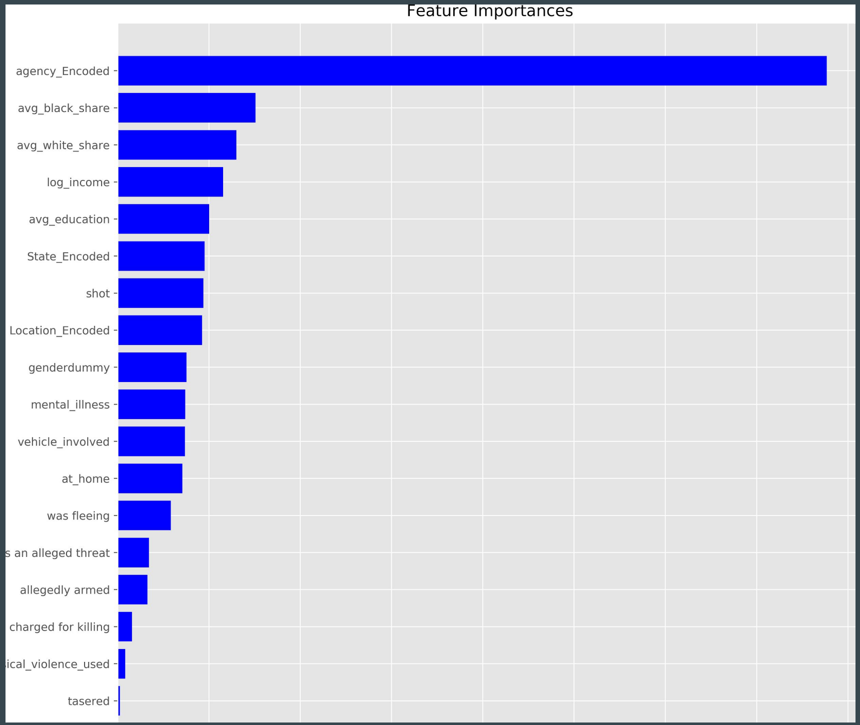
Advantages of Random Forest Model:

1) Does not assume linearity between Y and X variables

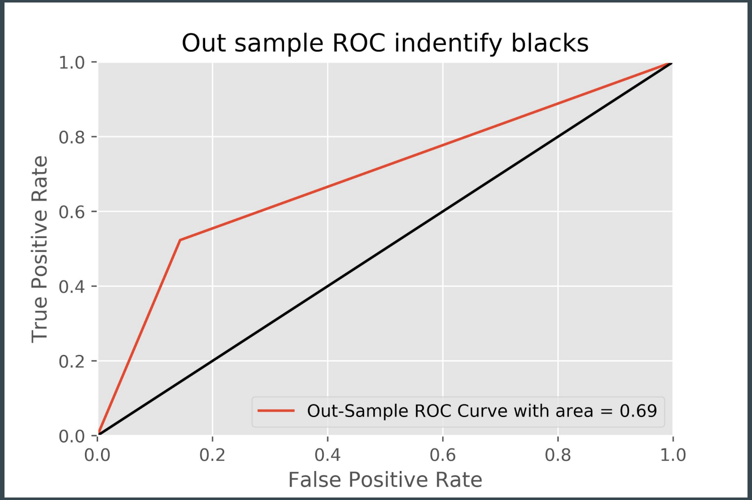
2) Able to handle missing values:
Column ***justified*** has over 4000 rows with missing values
-> We can include it in our Random Forest model but could not in our Logistic Regression model

Model 2: Random Forest

Predicting the Race of Individuals Killed by Police



Accuracy Score: 0.74

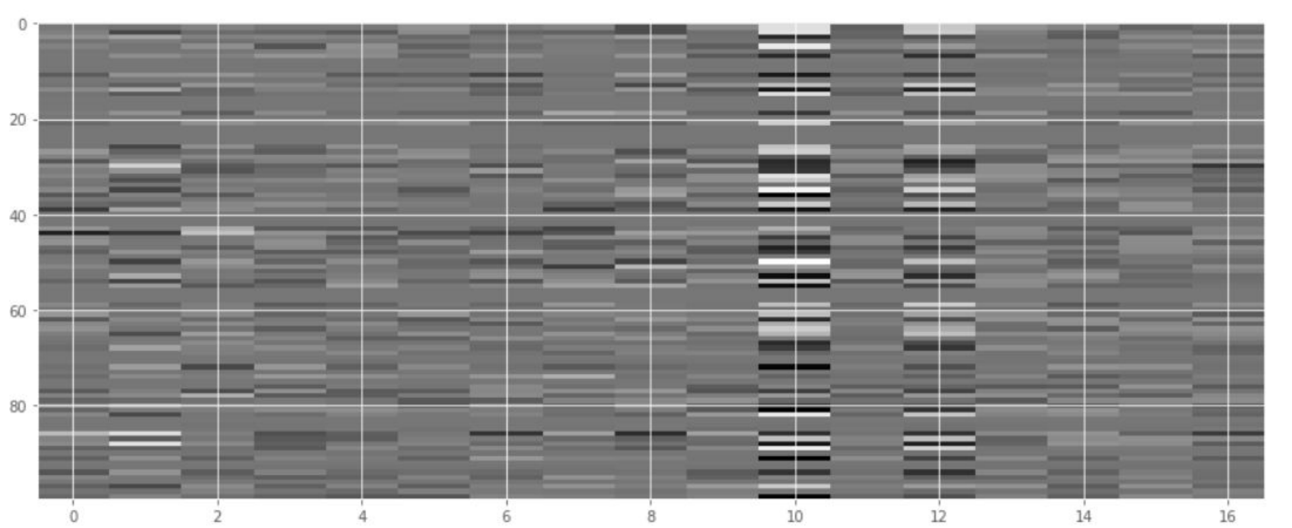


CFM: [825, 126],
[178, 189]

Model 2: Random Forest

Predicting the Race of Individuals Killed by Police

State_Encoded



Location_Encoded

Accuracy Score: 0.81

CFM: [868, 60],
[260, 130]

Model 3: Neural Network