US Police Killings in 2015

Robert Ludwig, Ayush Srivastava, Gopal Seshadri, Surya Prateek Soni

Introduction

In the United States in 2015 there were 1,146 people killed by the police. Here, we explore the types of individuals killed by police and try to characterize differences between different states and localities to understand the problem of police killing in America.

Motivation for Choosing This Project

America stands out amongst other developed countries as having a particularly high number of people killed by the police. Many of these events make national news in the United States, particularly when the killing seems to be egregious: for example when the victim is a minor, or unarmed or appears to be targeted for their race. Our goal is to characterize victims of police killings to shed light on what typical victims are, and understand typical circumstances in which someone is killed by the police in America with the goal of helping reduce overall violence.

Research Questions

To being with any sort of analysis, we must think of some crucial research questions, which would lead to selection of pertinent data sets, hypothesis and tests. The important questions we want answers for are as following -

1) Distribution -
   a) How do police killings vary with Age, Gender, Cause of death, Armed Info and Race - *Do the univariate analysis for all of them individually*
   b) How do the number of killings vary across Age & Armed Info and Race & Armed Info - *Perform Bivariate analysis*
2) Crime VS Killings -
   a) Find the top 15 cities with maximum killings and check if the killings are related with the general crime in that city
3) Geospatial analysis -
   a) How the distribution of the black people in a particular locality affecting the police killings
4) Forecasting -
   a) Predict the number of police killings that might happen across different number of populations for states and cities and different demographics

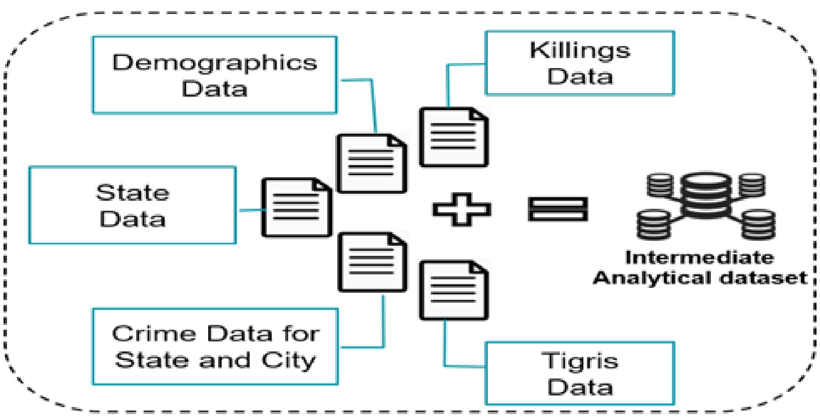
Dataset Used

For our analysis we considered, the police killings data set for the year 2015 from Guardian as our primary data set. Each record in the data set has name, gender, age and race-ethnicity of each victim along with the city, state, latitude, longitude, cause of death and armed info. Even though this data set is informative yet not fully enough to answer our research question.

Attribute	Description
name	Name of deceased
age	Age of deceased
gender	Gender of deceased
raceethnicity	Race/ethnicity of deceased
month	Month of killing
day	Day of incident
year	Year of incident
streetaddress	Address/intersection where incident occurred
city	City where incident occurred
state	State where incident occurred
latitude	Latitude, geocoded from address
longitude	Longitude, geocoded from address
lawenforcementagency	Agency involved in incident
classification	Cause of death
armed	How/whether deceased was armed

To augment our dataset with demographics information that is needed for our analysis, we took the census tract wise demographics information for the year 2015 from PDB_2015_Tract.csv which was extracted from ACS. The data in this data set is at census tract level whereas, the data in our primary data set has only latitude and longitude coordinates. Hence, we have to convert the latitude and the longitude into Geocode of the tract, and this is done using the tigris package. We are using the same to augment the demographics information to the original data set.

We are also analyzing the police killings given crime at the city and the state level, for that purpose we are grouping the original data to the corresponding level of granularity and then augmenting the same with city and state crime data sets.

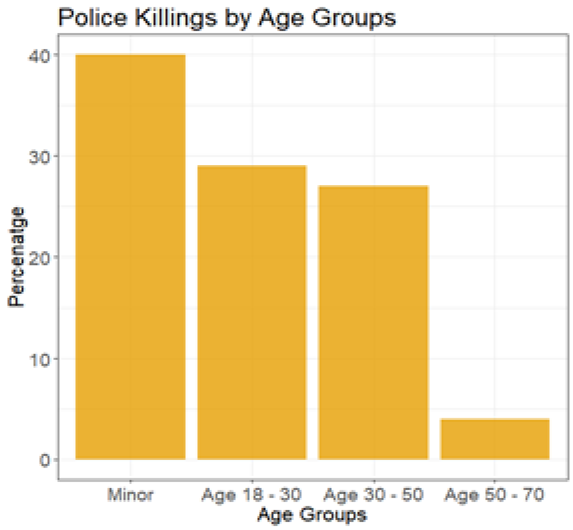


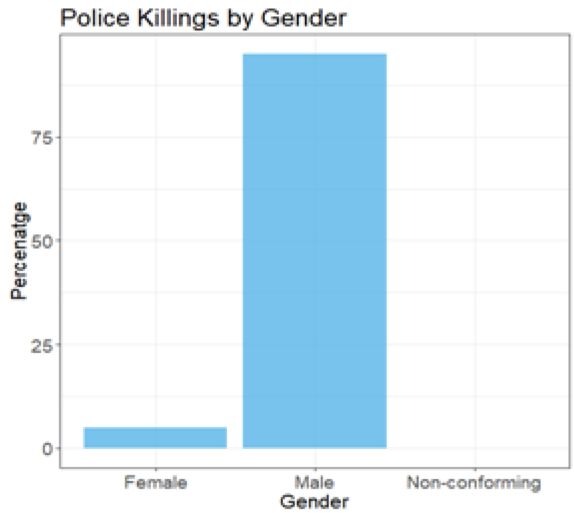
Characterizing Victims

Univariate Analysis

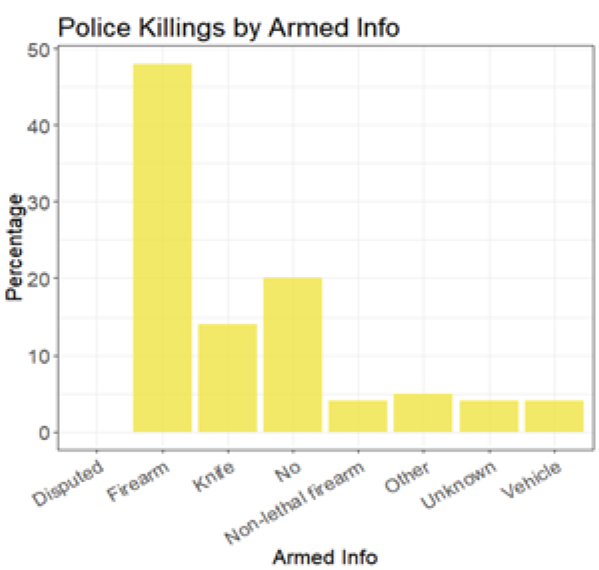
We performed the univariate analysis on police killings with respect to Age, Gender, Cause of Death, Armed Info and Race of the victims. When we plotted police killings by Age groups, we found that 40% of the victims are Minors and around 30% of the victims are in age group of 18-30 years of age. Shockingly high percentage of Minors dictates the need for a different approach in handling the crime situations in the country other than killing them.

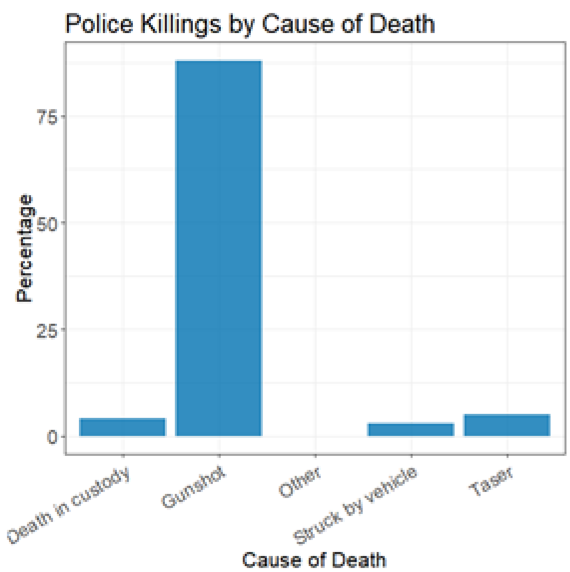
When we plotted the same against Gender, we found that around 95% of the victims are males and around 5% of the victims were females.



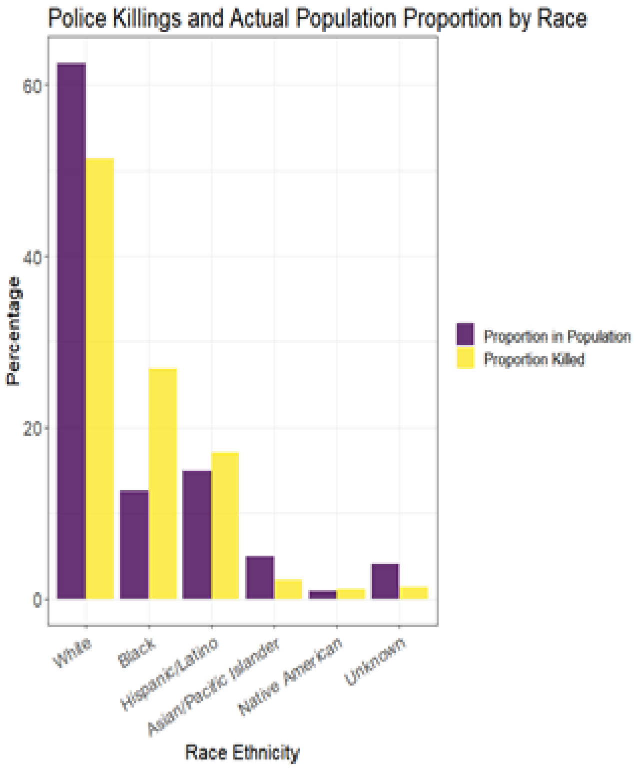


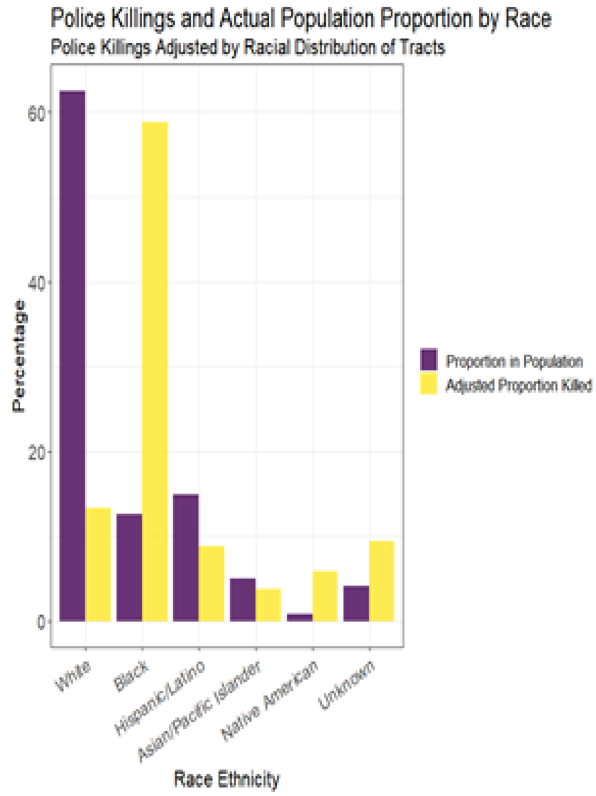
While plotting the Armed Info of the victims, it can be clearly seen that majority of the victims, roughly around 50% were carrying a firearm when they were killed. But shockingly around 20% of the victims were unarmed. After plotting the police killings against the Cause of Death, we found that around 90% of victims were killed from police gunshot whereas around 5% victims were killed by tasers, which was supposed to be used to neutralize the target without killing them.





In order to check the police killings against all the race ethnicities we plotted the proportion killed and proportion in population against the Race ethnicity of the victims, and we found that black people who constitute only 12% of the country's population were killed the most relative to the respective population share. Around 30% of the victims are black. Even though, 50% of the victims are white, it is relatively lesser then their population share. When we adjust the police killings with respect to the racial distribution of the tract, we find that the adjusted proportion killed for black is four times that their population share, and vice-versa when compared to the white population.



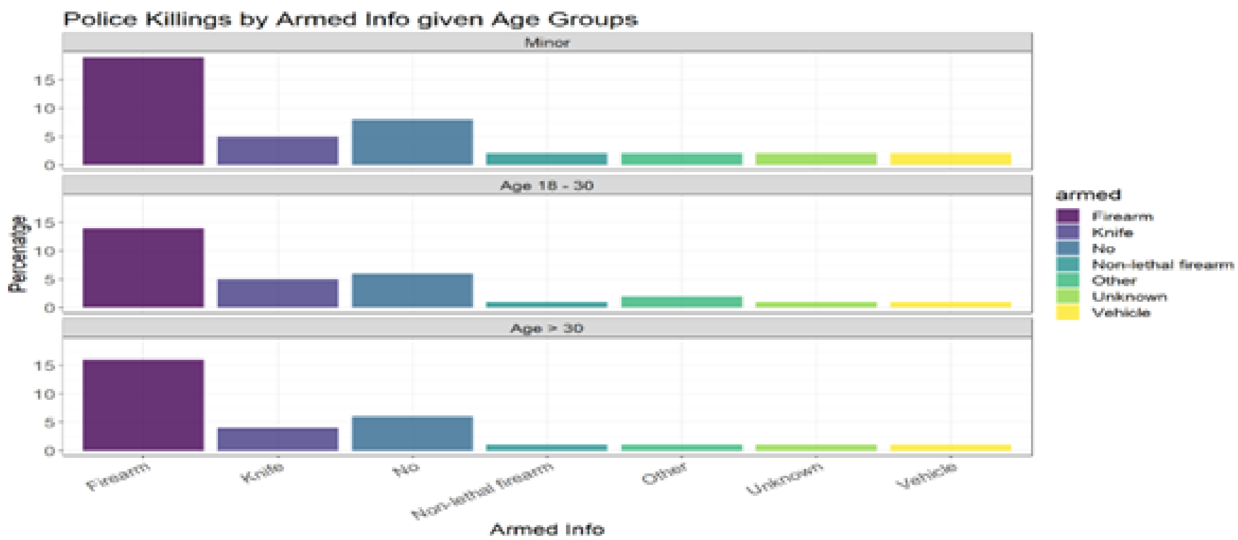


The adjusted count of the police killings is derived by dividing the number of police kills by the population share of the victim's race in that tract.

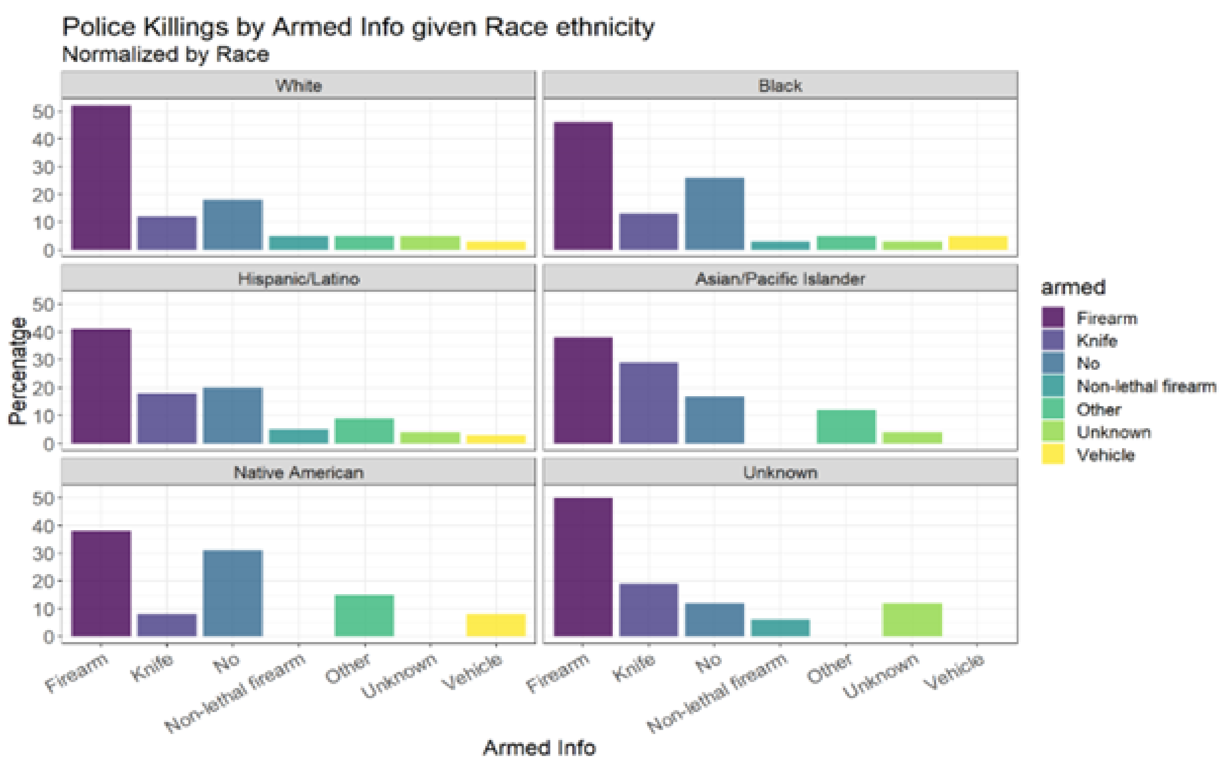
For eg, suppose in a tract we have only 25 people of only 2 racial ethnicities, White and Black. Out of these 25 people we have 20 White people and rest are Black, and the police kills 2 people, one White and one Black out of these 25 people. To get the adjusted have we have 1/(20/25) for whites and 1/(5/25) for blacks. So killing 1 White person out of 20 whites, would effectively result in killing 1.25 White person from that tract but killing 1 Black person out of 5 blacks would result in killing 5 Black person from that tract.

Bivariate Analysis

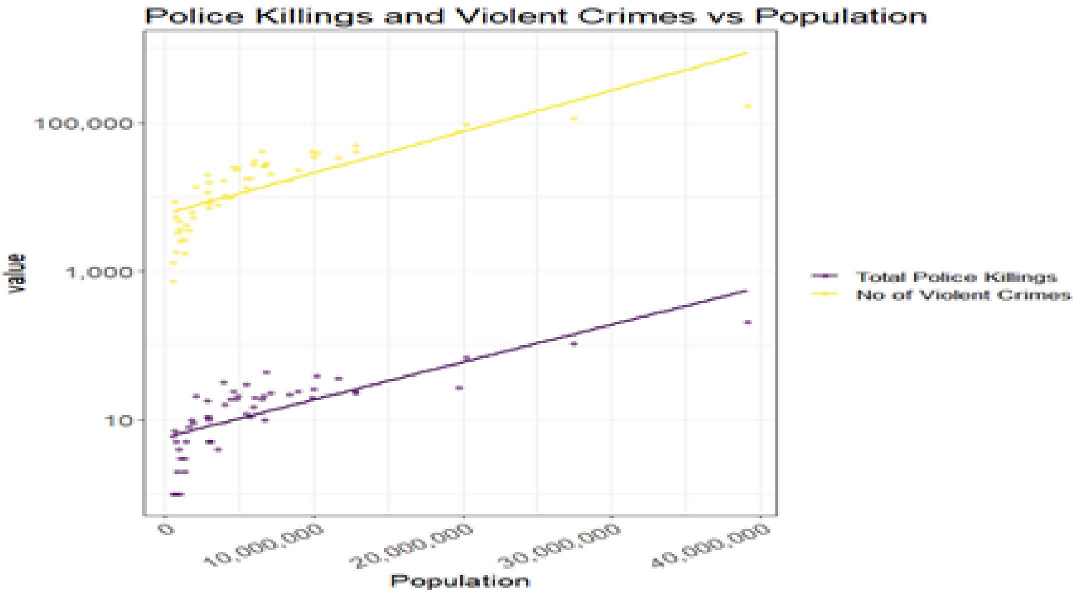
After plotting the Percentage of Police Killings against Armed Info and categorise by Age Group, we can see that Minors constitute the majority of the killing, roughly around 20%, who had a firearm on them. Among Minors, it can be seen that around 10% of the Minors were unarmed. Same pattern can be seen in other Age Groups as well.



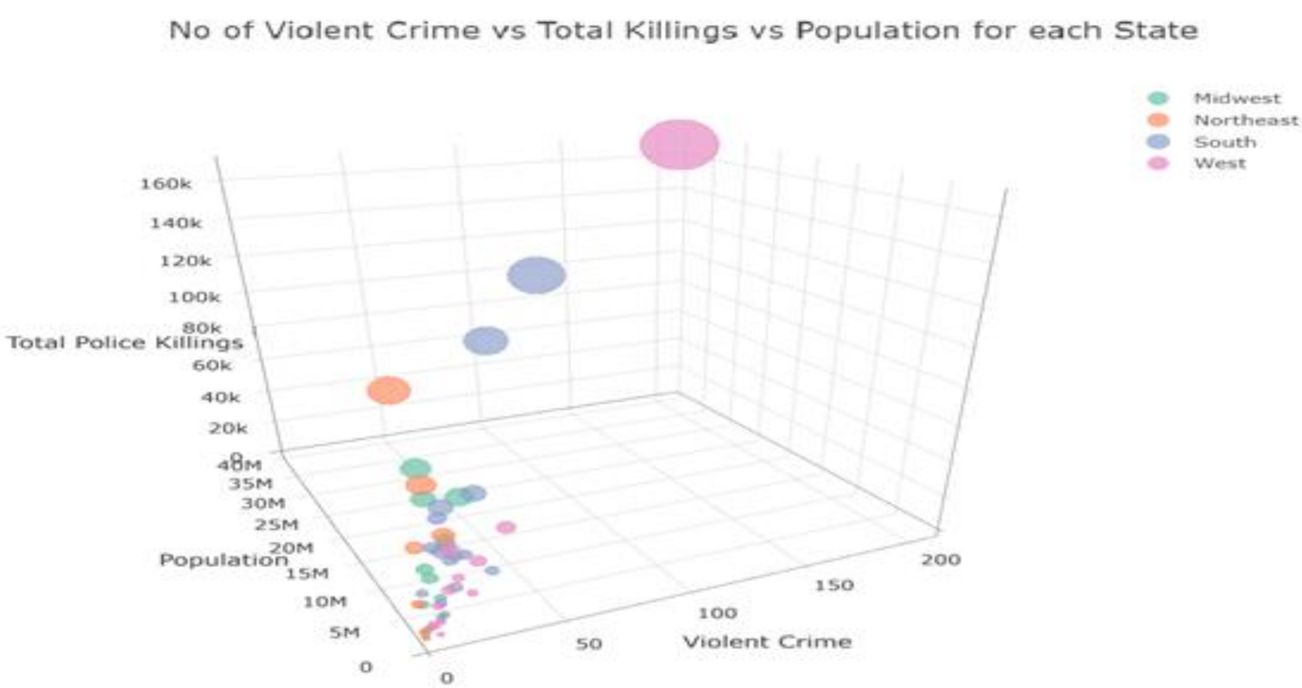
After plotting the Percentage of Police Killings against the Armed Info given Race ethnicity, we found that among all racial groups majority of the victims were having firearms. We also found that unarmed victims were significantly high for Blacks compared to that of White and Hispanic groups. Breaking down by race, we can see that Blacks are more likely to be killed when unarmed when compared to Whites.



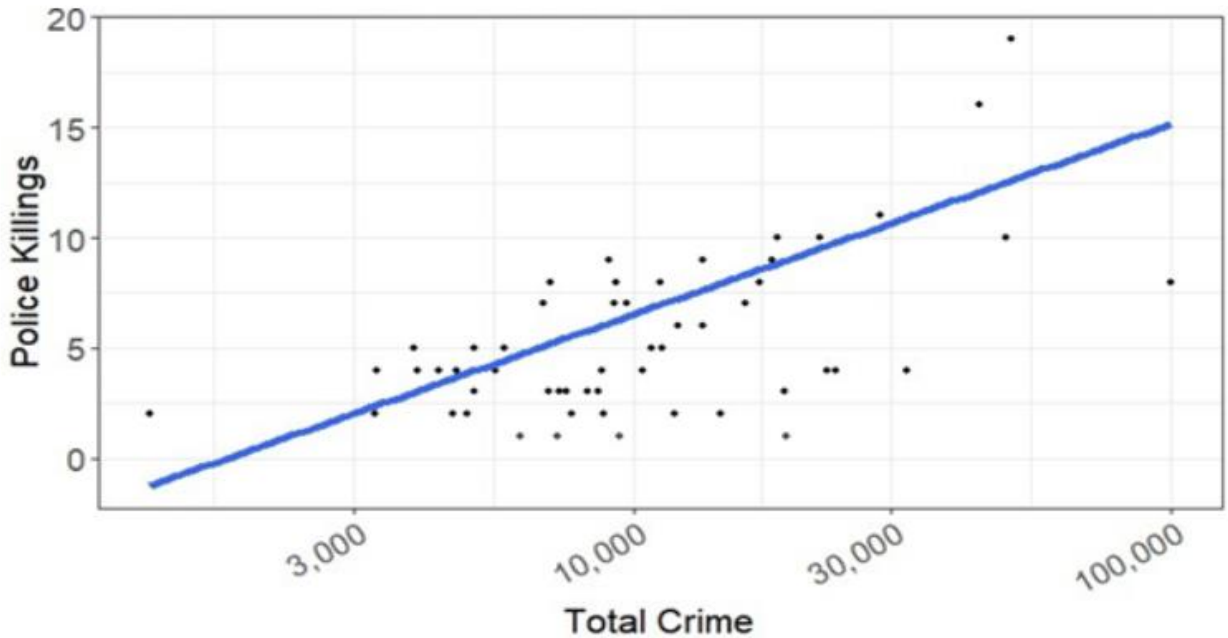
For every States, we have plotted the Total Crimes and Total Police Killings given the Population of each State, after which we found that as the Population of the State increases, the Total Crimes and Total Police Killings also increase, exhibiting a linear relationship between Population and Total Police Killings, and also between Population and Total Crimes.



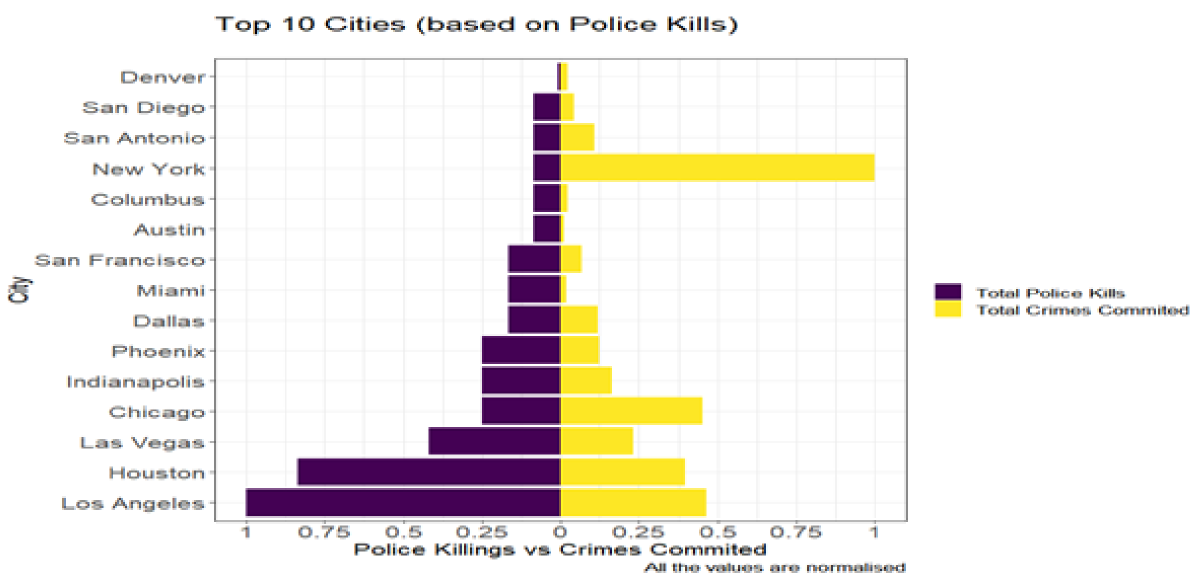
We then plotted the Population, Total Crime and Total Killings in a 3D scatter plot, from where we found a linear relationship between all the three parameters. Interestingly, the States with lower Population (small circles) were in lower heights compared to States with higher Population which clearly shows that for States with lower Population, the Police Killings is also lower.



Then we did an analysis by city, by plotting the Total Police Killings given the Total Crimes for every city in our data set, where again we found a linear relationship between Total Police Killings and Total Crimes. But there are some exceptions where there are relatively more Crimes and less Police Killings.



When we took the top 15 Cities that has more Police Killings, we found that the Cities with relatively high Police Killings may have relatively low Crimes or vice-versa. One such exception is New York, where there are relatively more Crimes and low Police Killings.



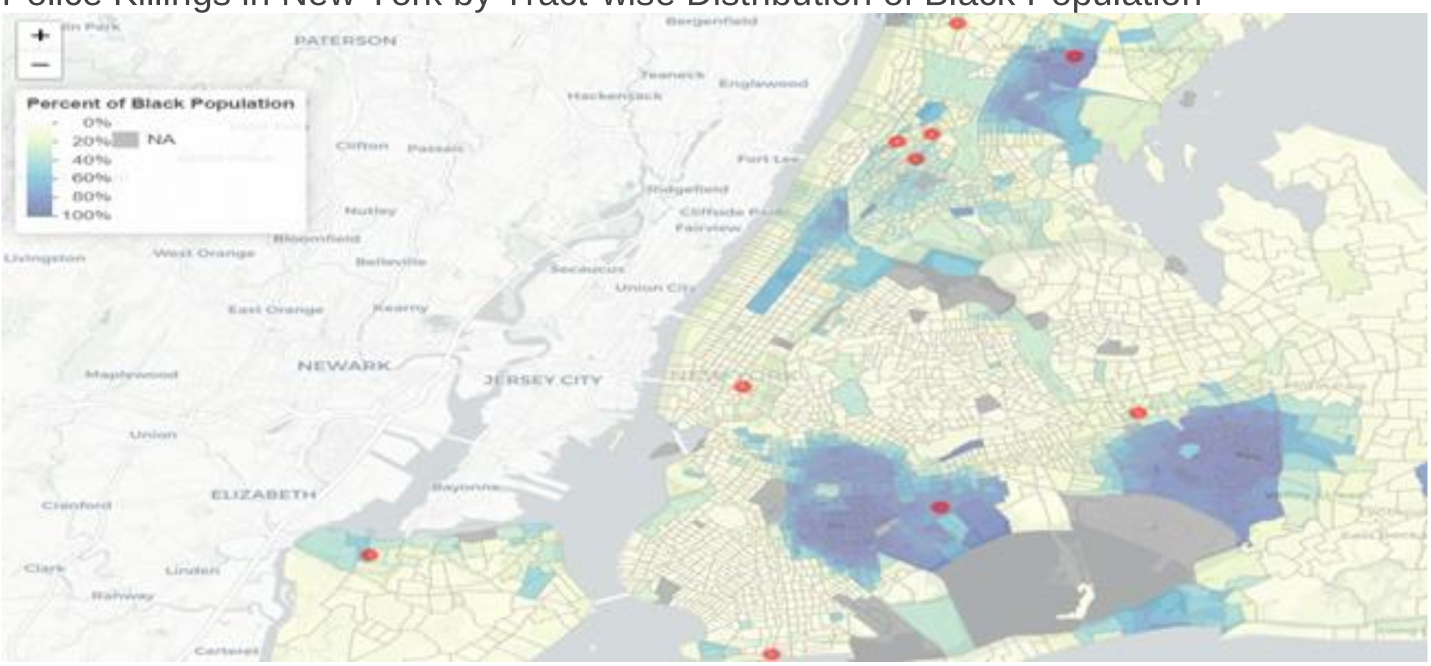
Geospatial Analysis

When we plotted the Police Killings over the map of United States, we found that the greater number of killings are from California. Also, there are relatively more killing in California and States in the West Coast and in the East Coast which are relatively more populated. Less police killings occurred in the States like North Dakota, South Dakota, Montana, Idaho and Wyoming which has significantly less population.

Geographical Distribution of Police Killings

As we noted above, New York is an interesting case, hence we planned to do a deep level analysis on New York. When we did a Geospatial analysis on New York City by distribution of Black population in every tract in the city and by plotting the locations of the Police Killings in the same. We found that most of the killings happened in place with relatively significant Black Population.

Police Killings in New York by Tract-wise Distribution of Black Population



Models Built

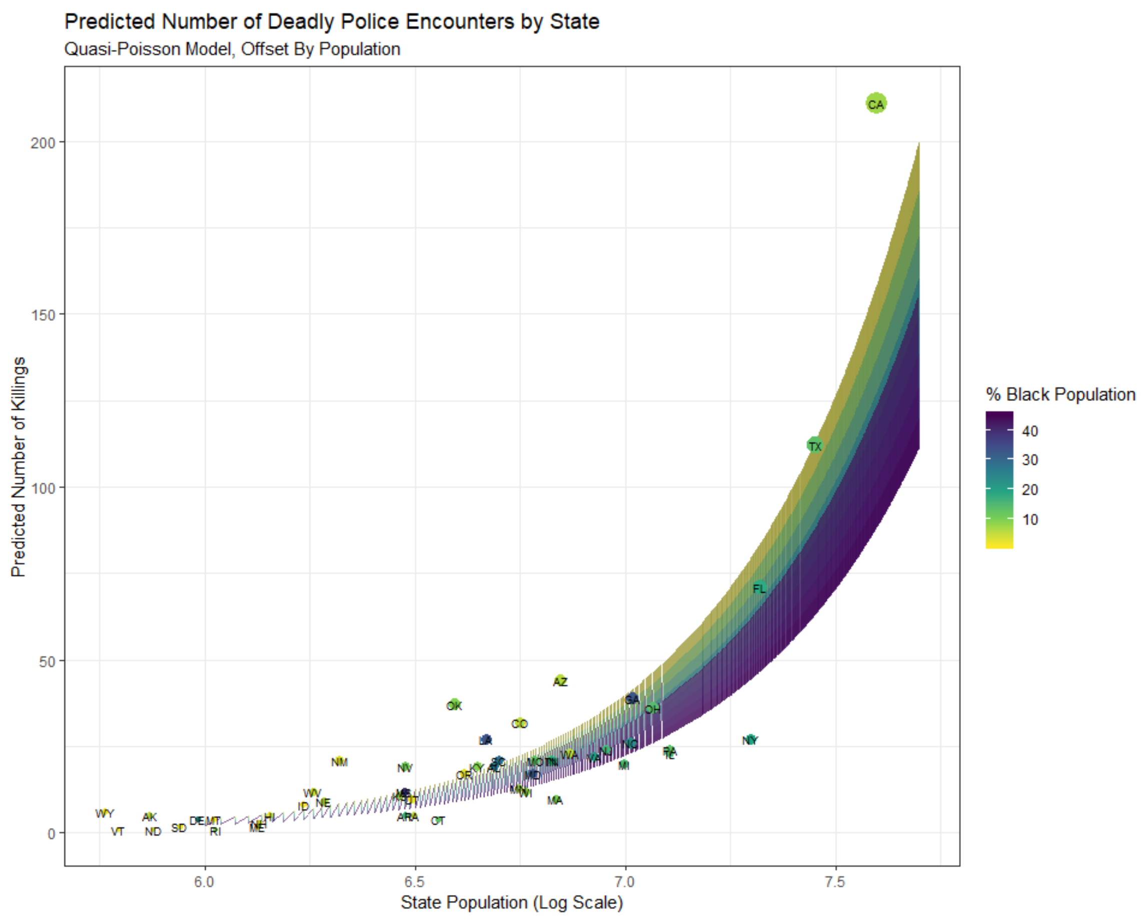
In order to estimate the number of police killings in a given area, we built a Poisson model with the count of police killings in 2015 as the predicted count data. Based on the analysis above, we built two models: one based on state-level aggregations and one based on county-level aggregations.

Our goal is to build a model that explains the effect of demographics on the number of police killings on either a state or a county level. We incorporated the census data as described above, and we used two formulations for predictors: the first was the percentage of black residents taken directly from the census data, and the second was the percentage of non-white residents, calculated from the census data as 100% minus the percentage of white residents. In trying combinations, both of these predictors were almost identical.

For both of the models, we used the population as an offset parameter in the Poisson model because we recognized that the number of violent crimes and the number of police killings are all linearly related. This construction should allow us to incorporate the crimes and population predictors into the model and determine the effect of the demographic variables while accounting for variations in population and crime rate.

State Model

For the state level predictions, we fit a quasi-poisson model as described above, with count of police killings determined by the percentage of black residents in a state, offset by the population.

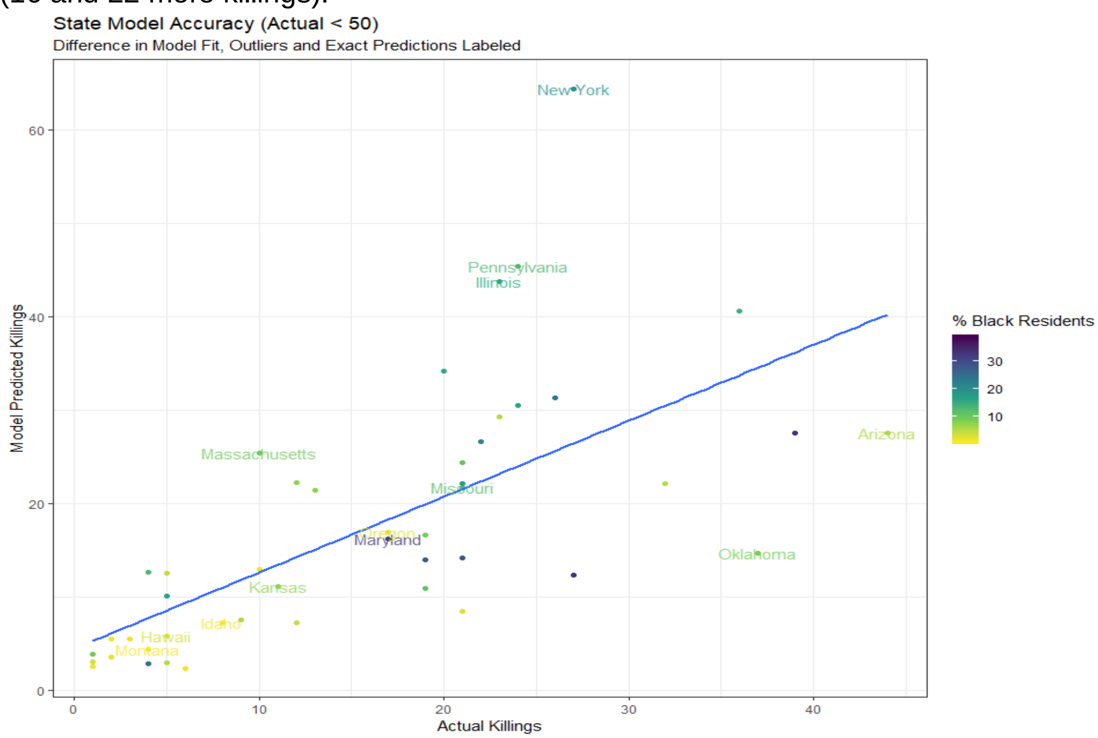


Interestingly, our model predicts that states with smaller black populations have a higher overall level of police violence, and states with a higher black percentage of the population have lower police violence (all else being equal). The coefficient in the regression result is -1.46, indicating that an increase of 1 percentage point of increase in black population corresponds to an increase of 0.25 deaths. The model has a mean absolute difference (mean of absolute value of predicted minus actual) of 8.0 for all 50 states, which drops to 3.8 when looking at the 30 states with fewer than 20 police killings.

This relationship is reflected in our data. There are many states with smaller black populations as a percentage of state population that have outsized numbers of police killings for their population, such as Utah (10 killings, 2% black), Oregon (17, 3% black), and New Mexico (21 killings, 3% black). By contrast, at the high-black-population end of the scale, states like Arkansas, Maryland, and Illinois have fewer killings than peers with similar populations. Looking at the accuracy graph below, there is no clear trend in accuracy with respect to race.

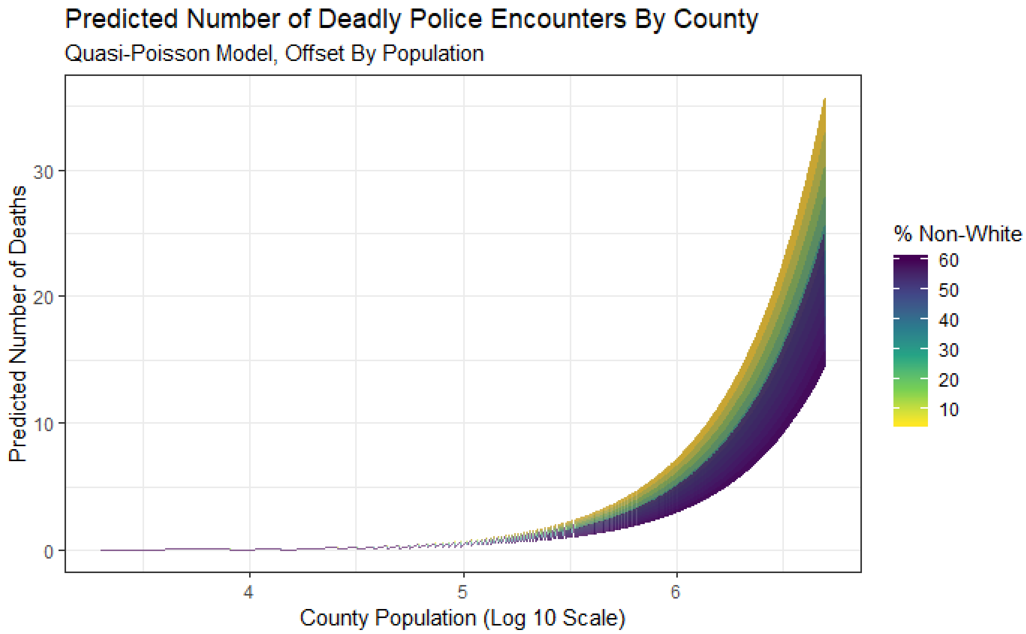
Overall, the model makes reasonable predictions for most states, however the predictions are worst for large states with large numbers of killings such as New York and California. There the model underpredicts killings by 60 in California (predicted 151, actual 211, difference 27%) and overpredicts in New York (predicted 64, actual 27, difference 137%). In the case of New York, this reinforces our earlier finding that New York (in particular New York City where the majority of killings in the state occur) seems to be doing something exceptionally well.

Along those lines, looking at the accuracy plot for the model is perhaps enlightening in this case because it indicates which states stand out as particularly good or particularly violent. New York, Pennsylvania and Illinois are all large population states with fewer than predicted killings (37, 21, and 20 fewer respectively). Arizona and Oklahoma have substantially higher killings than predicted for a state with their population and demographics (16 and 22 more killings).



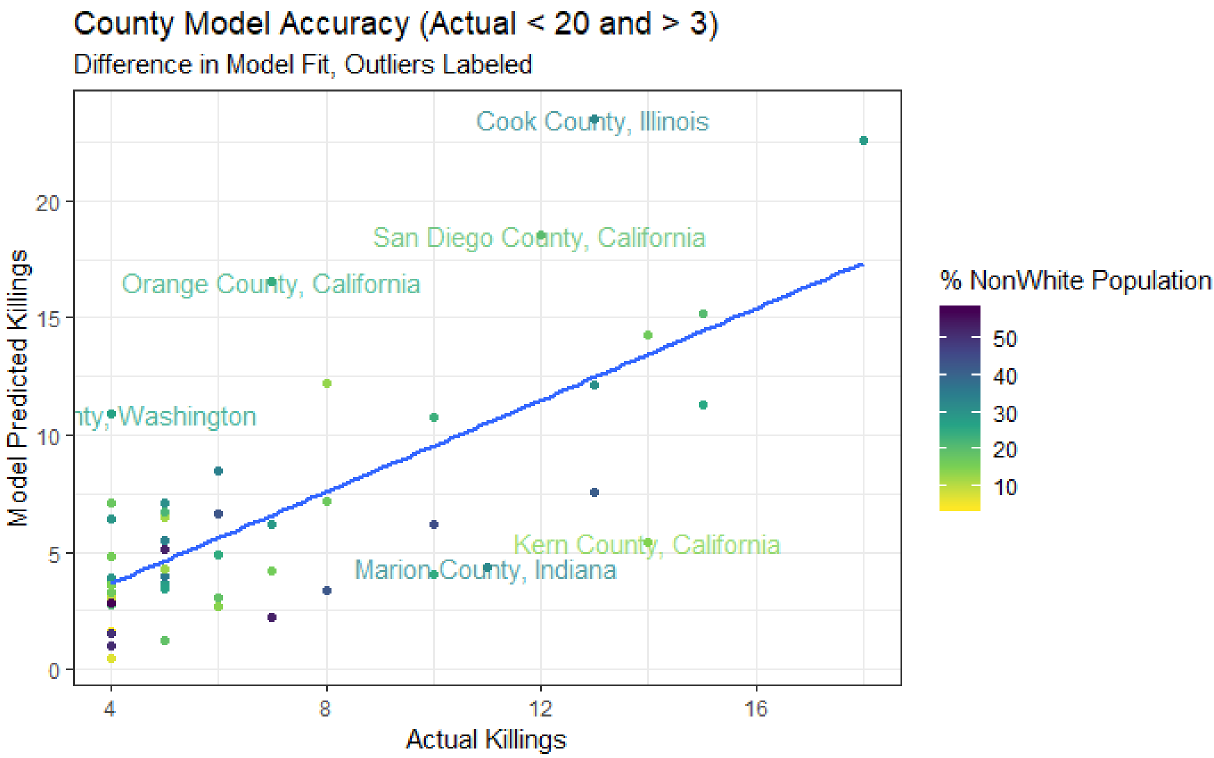
County Model

For the county-level predictions, we fit a quasi-poisson model as well, this time with the non-white percentage of residents in the county as the predictor instead of just looking at the black residents. This model also included the county population as an offset.



We see a similar pattern to the state level, where a higher percentage of non-white residents leads to a lower predicted number of deaths, and a lower percentage of non-white residents predicts a higher number of deaths. Here, however, the coefficient with respect to the non-white population is -1.65, lower than above, which corresponds to an increase of .19 expected killings in a county for every 1 percentage point increase in the non-white residents of a county.

Again, the accuracy graph is illustrative here. We've stricted the range to just those counties with actual killings greater than 3 or less than 20 to remove extreme values, including the many with just a single killing where our model performs typically very well. The main standouts are Cook County, IL (home to Chicago) and Orange County, CA (home to Los Angeles). These have substantially fewer killings than predicted (10 and 9 fewer, respectively). The California county of Kern County (home to the city of Bakersfield) has 9 more killings than predicted, and the home county of Indianapolis, Marion County, also has 7 more than predicted. Once again, there is no clear demographics pattern in the counties that are outliers.



Conclusions

With respect to our original research questions, we can conclude that while the majority of police killings are of white men, our analysis shows that black men are disproportionately killed far outside of their proportion in the areas where they live, and white men are killed disproportionately less. Most victims are unarmed, but many are also unarmed. Finally, the age distribution shows that minors and young adults are most often the victims.

Looking at geography, we noticed that while there is a strong linear relationship between population, violent crime, and police killings, there are some notable breaks from that trend. We investigated New York City as an example of a city that had fewer police killings than expected.

Finally, we built two models to predict what cities and states would have most police violence. After adjusting for population, we found that violence decreases with black population and minority population in both states and counties, and increases with total population. We predict that homogenous states or counties with higher black populations would have fewer killings overall.

Our conclusions are limited by the data we had available. For example, we only have data where killings occurred, but not on police interactions with less violent outcomes. This suggests some possibilities for future work, for example on determining which types of police interactions in which areas are more likely to end in death, and which end with the civilian still alive. Because many localities focus on de-escalation, we might expand our conclusions above by seeing if certain races, ages, or armed-types are more likely to survive. Finally, we might also build another model for individual police forces (county sheriffs vs city police, for example) to see if there are differences in outcomes there.