SAS® GLOBAL FORUM 2020 #SASGF

Paper 5200 -2020

What Happens After Police Shootings?

Aravind Dhanabal, Mason Kopasz, Alex Lindsay, Oklahoma State University

ABSTRACT

In the United States, the use of lethal force by police officers has come under extreme scrutiny. Police Departments have instituted new programs such as the wearing of body cameras and recruiting social workers to work together with on-duty officers. Due to limited resources, these cannot be instituted by all departments; therefore, it is important to make data-driven decisions to make the most of those resources. This paper explores the possibility of underlying biases and prejudices that are leading to officers not being held accountable. To accomplish this, police shooting data and the results of the follow-on investigations were examined. The goal was to identify if there is a predictive model to evaluate if a police officer will face Grand Jury Indictment based on the attributes within the dataset. The police disposition (which specifies if an officer faced a Grand Jury Indictment or not) was used as the target variable.

The Decision Tree model that performed best identified the most important variables which were the victim's State, County, Department Involved, Mental Impaired, Age group, Race, and Gender. The result of the Decision Tree model had a 35.86% misclassification rate.

This model can be used to predict the outcomes of police shootings and possibly be used by a government oversight committee, such as the US Justice Department, to investigate those counties and states that have abnormally high perceived unjustified shootings.

INTRODUCTION

The United States is entering a new time in its storied history revolving around social justice. Social Justice can be grouped into 2 major camps: inter-social treatment and unequal government regulation. Inter-social treatment is defined as the treatment of groups of other people based on biases and prejudices. Prejudices can be manifested into categories such as race, gender, nationality, and mental or physical ability. Unequal government regulation involves laws and regulations that purposefully or otherwise create conditions that obstruct, limit or deny groups access to the same opportunities and resources, relative to the rest of society (Pachamama Alliance).

Over the past 30 years, America has seen many high-profile police brutality and murder cases. One of the more famous involves Rodney King and the Los Angeles Police Department. With over 20 police officers present at the scene, Mr. King was beaten with batons. Four Los Angeles Police Department (LAPD) officers were indicted by a Grand Jury. Two months later, all four officers were acquitted. Riots ensued in Los Angeles. More recently, incidents like Michael Brown in Ferguson, Missouri (2014) and Walter Scott in North Charleston, South Carolina (2015) garnered national and international attention. The growing distrust between police officers and the public has culminated in a 22-year low in police confidence in 2019 (Gallup, 2019).

PROBLEM

There is no lack of ideas that aim to reduce the levels of police shooting in America. The approaches can be widely different in their methods and the issues they are attempting to address. For example, the use of body cameras is aimed at Police accountability whereas a program that has social workers accompany on duty officers is targeted at reducing deaths of mentally ill

persons; both have their merit. Unfortunately, restrictions on resources do not allow all projects to be pursued. Although these approaches successful in large, metropolitan cities, the aim is to determine if there are more underlying, societal issues that could be causing the use of lethal force and the follow-on investigation of the officers who used it. The goal of this project is to create a predictive model that can identify the variables that lead to a perceived justified or unjustified police shooting and forecast the shootings across the United States for the next 10 months in an effort to provide key insights to those campaigning for reform.

METHODS

DATA COLLECTION

The dataset used in this project comes from the fatalencounters.org. Beginning January 2000, fatalencounters.org has been tracking civilians that have been the victims of fatal police shooting. They have collected over a dozen data points including gender, age, race, and mental illness. The dataset also contains a brief description of the events surrounding the shooting. Fatalencounters.org compiled these attributes by searching local news reports, law enforcement websites and social media and by monitoring independent databases such as Killed by Police and Fatal Encounters. Additionally, population data was collected from the 2010 census, located at census.gov. The Census data was used to normalize the police shooting data from fatalencounters.org.

DATA PREPARATION & VALIDATION

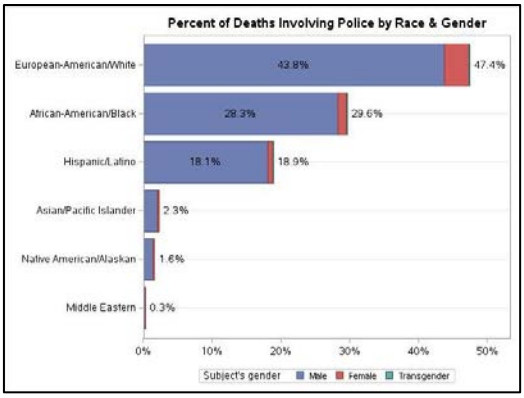
Although fatalencounters.org collected all the data for a specific purpose, there were many variables that were missing or collected/reported in a non-similar fashion. To prepare the data, the ages were binned into nine groups. The first group consisted of persons under 20; ages beyond that were then binned at 10-year intervals. The Disposition variable that recorded the results of the legal proceeding for the officer was not standardized. Therefore, efforts were invested to ensure the results of the legal proceedings fell into one of three categories: Justified, Not Justified, and Unreported. Perceived unjustified police shootings were those which reached a Grand Jury indictment, while perceived justified did not. The Mental Impaired variable required normalization as well. The variable was grouped into one of four categories: yes, no, unknown, drugs/alcohol. The dataset was partitioned into 70% training and 30% validation.

The data from census.gov returned no missing or null values and reported the populations from each state's counties to include males, females, various racial identities, and 5-year age intervals from 0 to 85 and older. There was no need to impute data or removing missing values. The state population was combined using the "first" and "last" to sum up the total, gender, and racial populations of the counties.

ANALYSIS

RACE DEMOGRAPHICS

From the analysis, the largest group of deaths by volume were White Males; they represented 43% of the victims in incidents where Race was specified. However, after standardizing against Population for Whites, Hispanics, and African Americans it is seen that as a proportion of population African Americans & Hispanics are more affected than Whites. African Americans and Hispanic/Latinos were victims of police shootings at 96 and 49 deaths per million, respectively; versus 25 deaths per million for white Americans. This lends merit to the initiatives that have some racial component to drive down deaths involving police.



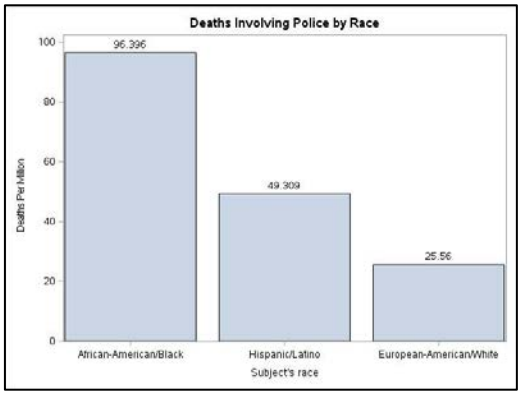
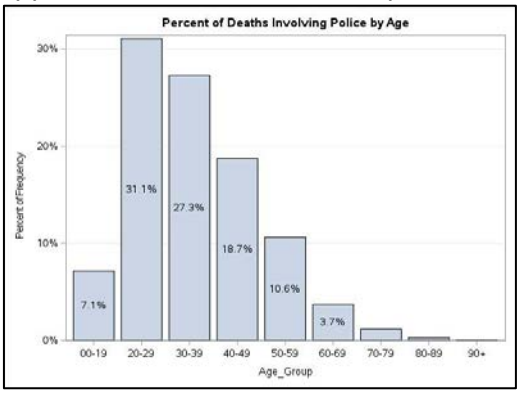


Figure 1 & 2. Deaths by race & Gender, Deaths per million by race.

AGE DEMOGRAPHICS

When looking at age, a large portion of the shooting deaths occurring to persons under 40; they represented around 65% of deaths. When analyzing age by race it can be seen that African Americans and Hispanics are making up larger portions in those prime age bands of 20-39 especially when considering they make up a much lower proportion of the population. This trend proceeds to decrease in their later age bands. This shows that age should be playing a role in any approach or effort to reduce police shooting deaths.



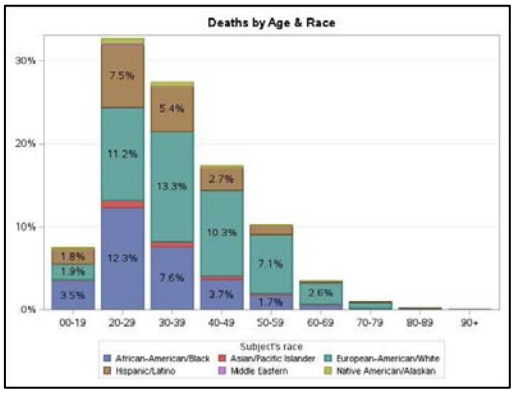
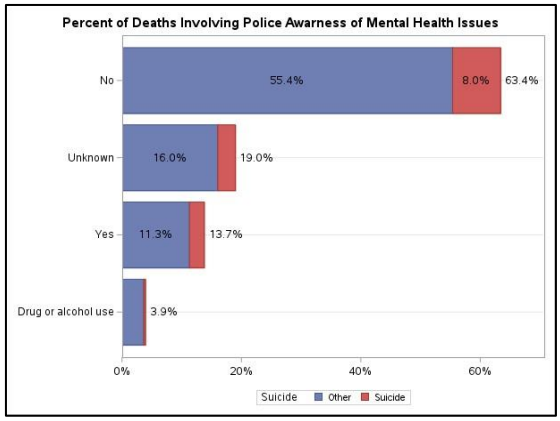


Figure 3 & 4. Deaths by age, Deaths by age & race



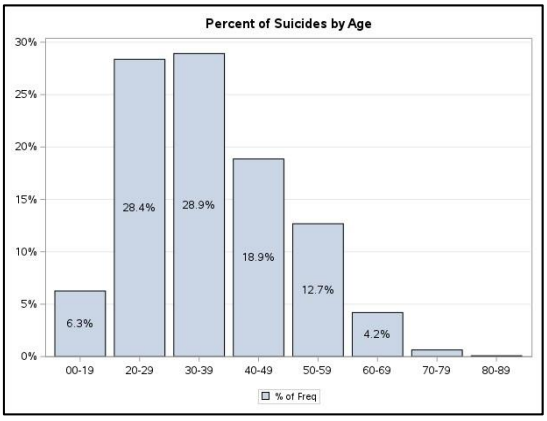


Figure 5 & 6. Deaths involving mental health issues, Deaths for suicide by age.

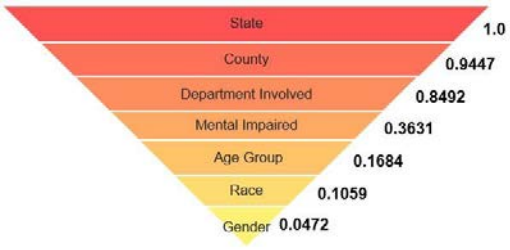
MENTAL HEALTH DEMOGRAPHICS

Looking into the mental health of the victims, police were aware of mental health or drug issues in 17.6 % of the incidents. Within those cases 15% ended with the victims' suicide. Incidents like these reinforce initiatives such as including social workers when mental illness is a factor to drive down preventable deaths. Beyond that 12% of incidents where police identified no mental illness, but suicide was a factor suggests the possibility that these incidents could have been misidentified. When looking strictly at Suicide by age it is seen that once again the young are making up a large proportion of the observations at almost 60%.

DECISION TREE MODEL

After partitioning the data, three decision tree models were created. The models used in analysis were Maximum Tree model, Chi-Squared default, entropy and Gini. The models were built using the Disposition as the target variable with the victim's age group, race, mental impaired, and gender along with the department involved, state, and county the shooting took place used as predictor variables. The splitting criteria of chi-square bested the models built with splitting by entropy, and Gini-index based on the validation assessment. The pruning method was adopted to reduce error and an objective of misclassification rate. The Chi-Squared model was the best compared to other models in terms of ASE and Misclassification rate, which were the two most important metrics used as validation assessment for this analysis.

The model determined there are seven variables which are most significant to determining a perceived justified or unjustified outcome: victim's race, state of shooting, victim's mental illness, county of shooting, gender of victim, age of victim, and city of shooting. The Chi-Squared model is maximized with a validation misclassification of 35.86% and accuracy of 64.14%.



	Predicted: Justified	Predicted: Unjustified	
Actual: Justified	1182 (30.14%)	659 (16.81%)	1841 (46.95%)
Actual: Unjustified	742 (18.92%)	1338 (34.12%)	2080 (53.05%)
	1924 (49.07%)	1997 (50.93%)	3921 (100%)

Figure 7 & 8 Variable Importance, Confusion Matrix **FORECAST**

MODEL

To check for effectiveness of the decision tree model, forecast models were built based on the variable importance from decision tree. Moving Average and Exponential smoothing techniques were used for the forecast models. An exponential smoothing forecast model was chosen for the final selection since it takes account of all past data in the analysis. Police shooting deaths were forecasted for next 10 months and the results were compared with the actual incidents. The actual number of shootings in the United States for the month of November, 97, was near the forecasted value of 94.

GENERALIZATION

The decision tree model created will take an observation and path it towards a decision on whether the disposition will be perceived justified or unjustified. The confusion matrix of the model shows a

correct prediction of the disposition of the police officer approximately 64.14% of the time. The model performed moderately from a statistical point of view. Ideally a model with two possible outcomes would have a better accuracy. However, one important aspect is the variables selected as the most important. The State, County, and Department involved are the three most important to the model and significantly outweigh the variables of Mental impaired, Age group, Race, and Gender. This could be due to a variety of reasons, but it is interesting to note.

The forecast is used to build a projection for shootings in the US from November 2019 to October 2020. The forecast shows the United States is approaching a downward turn in police shootings. This could be due to numerous initiatives at the national, state, and local level to reduce incidents.

FUTURE STUDIES

One major limitation to the model is the restricted scope of the data collected. Since the data was collected from a source that does not have access to police files, the information was limited to what was provided to the public. There were a lot of fields that were either missing completely or were listed as "unreported." The model could be improved if the data was not collected from an open source, but leveraged national police databases, such as those maintained by the Federal Bureau of Investigation or US Justice Department.

Another future study could stem from the forecasted model. Expansion to measure all police responses could prove vital to many police departments. Many departments struggle with having the correct funding from their state's legislature, so having a verified forecast of all reportable incidents could give police departments a leg up when lobbying for expanded staffing and resources.

CONCLUSION

This paper details one approach data science can be applied to Social Justice initiatives in the way police officers carry out lethal force. The study uncovered that African Americans and Hispanic/Latinos are victims of police shootings more often than white Americans per capita. Ages 20-39 are the most prone to becoming victims of police shootings, amassing 65% of all recorded incidents. The predictive model did not result in finding a racial bias for perceived justified versus unjustified police shootings. However, the State, County, Department Involved proved to be the most important variable in deciding if a police shooting will be perceived justified or unjustified. This model could be used in practical application by an oversight committee to ensure there is standardization across police forces for launching investigations into Grand Jury indictments.

REFERENCES

"Social Justice Issues." Examples and Solutions | Pachamama Alliance, https://www.pachamama.org/social-justice/social-justice-issues.

Jones, Jeffrey M. "In U.S., Confidence in Police Lowest in 22 Years." Gallup.com, Gallup, 14 Feb. 2019, https://news.gallup.com/poll/183704/confidence-police-lowest-years.aspx.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Aravind Dhanabal
Oklahoma State University
aravind.dhanabal@okstate.edu

Mason Kopasz
Oklahoma State University
mason.kopasz@okstate.edu

Alexander Lindsay Oklahoma State University
alexander.lindsay@okstate.edu

APENDIX

Variable Importance

Variable Importance				
Variable	Importance	Std Dev Importance	Relative Importance	Count
State	205.04	5.8122	1.0000	36
County	193.71	2.8397	0.9447	55
Department_involved	174.12	2.3686	0.8492	54
Police_awareness	74.4470	18.1311	0.3631	7
agegrp	34.5235	2.4169	0.1684	12
Race	21.7105	0.8163	0.1059	12
Gender	9.6865	1.1434	0.0472	3

Confusion Matrix

Event Classification Table

Data Role=TRAIN Target=Disposition Target Label=Disposition

False Negative	True Negative	False Positive	True Positive
1605	3104	1747	2689

Data Role=VALIDATE Target=Disposition Target Label=Disposition

False Negative	True Negative	False Positive	True Positive
659	1338	742	1182

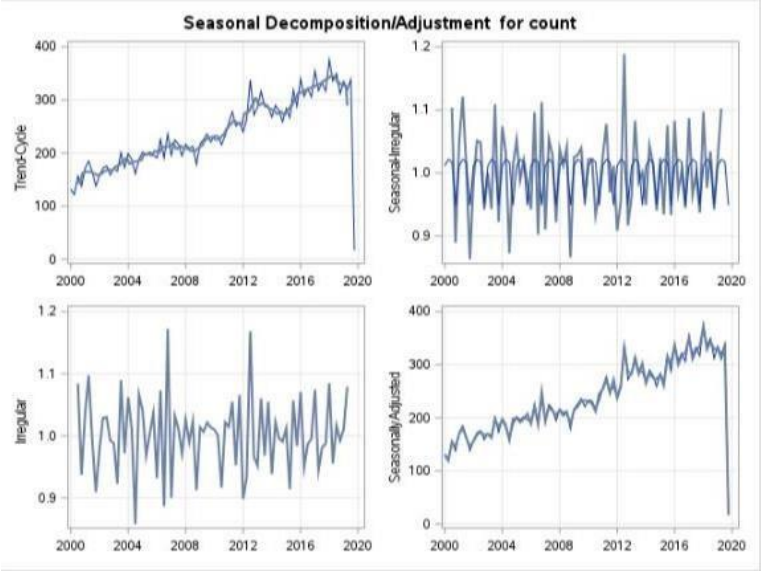
Fit Statistics

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Disposition	Disposition	_NOBS_	Sum of Frequencies	9145	3921	.
Disposition	Disposition	_MISC_	Misclassification Rate	0.364571	0.364193	.
Disposition	Disposition	_MAX_	Maximum Absolute Error	0.828829	0.828829	.
Disposition	Disposition	_SSE_	Sum of Squared Errors	4042.405	1728.685	.
Disposition	Disposition	_ASE_	Average Squared Error	0.221017	0.220439	.
Disposition	Disposition	_RASE_	Root Average Squared Error	0.470125	0.46951	.
Disposition	Disposition	_DIV_	Divisor for ASE	18290	7842	.
Disposition	Disposition	_DFT_	Total Degrees of Freedom	9145	.	.

Forecast model results:

Trend, Seasonality and Cycle charts – overall model



Forecast Chart – Overall model

Date_of_death	Forecast	Standard	95% Confidence Limits	
Nov-19	94	12.15496	70	120
Dec-19	96	12.66427	71	121
Jan-20	104	13.29177	79	130
Feb-20	89	13.90989	64	115
Mar-20	105	14.50841	80	131
Apr-20	91	15.08489	65	117
May-20	95	15.6387	69	121
Jun-20	99	16.16935	72	126
Jul-20	104	16.67537	77	131
Aug-20	105	17.15207	78	132

Forecast by count