Predicting Race and. Socioeconomic Characteristics from Police Shooting Data

Ryan S. Dunn and Derek Lankeaux

Shiley-Marcos School of Engineering, University of San Diego

# Abstract

Social justice and human rights advocates have taken up efforts to bring awareness and curve the persistence of police violence within marginalized communities. Using available data from multiple sources, our projects aimed to understand the distribution of police shootings, as well as building models that may be able to predict if someone is more likely to be a victim of police shooting. The data uncovered…

Table of Contents

Abstract ……………….…………………………………………………………………………. 2

List of Tables …………………………………………………………………………….……… 4

List of Equations ………………………………………………………………………………... 5

# Predicting Human Characteristics from Historical Police Shootings ………………….…...…… 6

Exploratory Data Analysis and Methodology ……………………..……………….……………. 6

Machine Learning Models …….………………………………………………………………… 8

C5.0 ………………………………………………………………………………………. 8

List of Tables

Table 1. Descriptive Statistics ………………..………………………………………………….. 3

List of Equations

Equation 1 Insert here …………………....………….……………………………………...…… 7

# Predicting Human Characteristics from Historical Police Shootings

# In recent years, police shootings have undergone increased scrutiny, as different socioeconomic groups have begun increasing awareness surrounding the perceived disproportions of police shootings by race. With accompanying protests, calls to action have arisen, suggesting policies that would mirror police that is more in line with European counties such as Iceland and Switzerland, where a staggering zero police killings have taken place. Movements such as “Defund the Police” have also gained traction, to reduce police shootings. There has been significant debate surrounding these protests, as to whether minority race Americans are truly killed by police at a higher rate. These counter-protests have turned this topic into a highly contested public policy issue in need of resolution by data analysis.

# Exploratory Data Analysis and Methodology

The purpose of our analysis was to develop predictive models that could assess the impact of socio-economic and demography attributes of previous police shooting victims. The primary analytics base table was acquired from The Washington Post’s police shooting database, where every fatal shooting by an on-duty police officer has been logged since 2015 (The Washington Post, 2021). To supplement key demographic data, more data was collected from the Kaggle dataset “Fatal Police Shootings in the US”, where the percent over 25 graduation rates, percent below poverty rate, and median household income were attained by city. These supplementary data sets were aggregated to the state level, where a new attribute was joined onto the base table, incorporating these socio-economic variables. These attributes allowed us to append the data with critical socio-economic factors that may drive police altercations. Talk about how we appended the data, assumptions, and basic descriptive statistics. In Figure 1, the distribution of the age of police shooting victims is displayed in bins of 5. The histogram is positively skewed, as intuitively, younger persons(non-adult) are less likely to be a victim of police shootings versus their older counterparts.

**Figure 1**

*Police Shootings Histogram Since 2015 by Age*

Chart, histogram

Description automatically generated

# Figures 2 and 3 display a similar histogram of age, this time with a race overlay (Figure 2) and the corresponding normalized overlay (Figure 3). By providing the data with the race overlay, we can better understand the distribution of race, as relates to the age of police shooting victims. In 2019, whites accounted for 60.1% of the population, Hispanics accounts for 18.5%, and blacks accounted for 12.2% of the U.S total population (Source, 2019). As Figure 3 depicts, for ages 15 – 25 specifically, Black and Hispanic police shootings victims are disproportionate relative to the percent of their total population within the United States. This is most effectively displayed in Figure 3, where the histogram is normalized to display the race distributions within each age bin. Here, there is a clear disparity relative to minority races for age bins 15, 20, and 25. This data supports the assertion that young minorities are disproportionately killed by police.

# Figure 2

# *Histogram of Age with Race Overlay*Chart, histogram Description automatically generatedFigure 3

# *Normalized Histogram of Age with Race Overlay*

Chart, histogram

Description automatically generated

# To further provide evidence of disproportionate police shooting victims by race, Figure 4 displays a boxplot by race, where the distribution of age by race is clearly depicted. Minority races are more heavily concentrated and have lower median ages.

# Figure 4

# *Boxplot of Age by Race*

Chart, box and whisker chart

Description automatically generated

# In addition to age and race, socio-economic attributes such as poverty rate, and the amount of high school graduates in a region are relevant for analysis. Figure 5 displays the correlation between the median poverty rate and percent of population that are high school graduates by state, with a Region overlay. The region overlay helps to provide insight into how different regions of the country have clear differences with respect to education rates and poverty. For example, the Northeast region is primarily contained within a cluster in the top left-hand portion of the scatterplot, indicating a higher education level and lower poverty rate, whereas both the Southeast and Southwest regions exist primarily within the lower right-hand portion of the scatterplot, indicating lower education rates and higher poverty rates.

# Figure 5

# *Scatterplot of Poverty Rate, Percent over 25 HS graduates with Region Overlay*

Chart, scatter chart

Description automatically generated

# Machine Learning Models

**C5.0**

**CART**

**Logistic Regression**

**Neural Network**

**Model Evaluation**

Insert here.

**Implications for Future Research**

Insert here.

# 

# References

The Washington Post. (2021, December). Fatal Force. <https://www.washingtonpost.com/graphics/investigations/police-shootings-database/>

Fatal Police Shootings in the U.S. (Karolina Wullum, 2017). Kaggle. <https://www.kaggle.com/kwullum/fatal-police-shootings-in-the-us>