Late Night Dangers: Exploring the Link Between Time of Day and Weapon-Related Crime Victimization

Aaron Peterseil, Fengruo Zhang, Ken Deng, Shreya Arvind, Thomas Blakely

Contents

1 Introduction

While overall crime in California has been decreasing since 2022, there has been an uptick in the use of firearms in homicides and aggravated assaults, according to data collected by the California Department of Justice and reported by the LA Times (Department of Justice Report, LA Times). Naturally, public safety remains a high-priority issue in any government. Solutions to public safety issues, however, are quite controversial, as seen in the discourse around gun laws and law enforcement.

The purpose of this paper, however, is not to focus on specific policy measures to combat this increase in violent crime. Instead, this paper highlights the use of firearms in crime in the Los Angeles area to help policymakers draft more effective policies based on these results. Our analysis is broken down into three parts:

(1) An analysis of crime rates.

(2) An examination of firearm usage frequency in crime.

(3) A model to predict what factors influence the occurrence of gun-related crime.

2 Methods

2.1 Dataset

The results and conclusion of this paper are based on crime data collected between February 10, 2020, and March 6, 2024, by the Los Angeles Police Department (LAPD) (Crime Data from 2020 to Present). The data was downloaded on March 6, 2024. We note that, as of Mach 7, 2024, the LAPD adopted a new Record Management System to comply with FBI mandates. As a consequence, future versions of this dataset may not be representative of what we used.

The LAPD crime data contains records of 910,707 incidents spread across 28 variables. Information covered includes the date and time of the incident, the geographic area where the crime occurred in relation to the 21 Los Angeles Community Police Stations as well as its exact location, details on the crime including if a weapon was used and potential motivation for the crime, information on the victim, and the current status of the case.

2.2 Data Cleaning

Some modifications were needed to align the dataset with our research goals. First, we created a binary isFirearm variable based on the 'Weapon. Desc' column. Of the 80 unique weapon descriptions in the dataset, 22 were firearms, and 3 were imitation firearms (Air Pistol, Simulated Gun, etc.). The value of the isFirearm variable was determined based on whether the weapon was one of the firearms or an imitation firearm. The rationale for including imitation firearms with firearms in our analysis is that:

(1) Due to their close appearance to real firearms, they often result in dangerous situations where they are mistaken for firearms (Hoops & Teret, 2017).

(2) Crime descriptions of these imitation guns included accounts of them being used to commit "ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT."

As crime does not necessarily scale linearly with increases in population, we included a population variable in our dataset as well (Oliveira, 2021). Our numbers for this variable were based on the estimates provided by the LAPD of the population under the jurisdiction of 19 Los Angeles Community Police Stations (LAPD Online). Data on the population Olympic and the Topanga police stations, however, did not contain population data. This is likely due to them being established in recent years. Olympic data was found from the Rampart Division Wikipedia, but no data can be found on Topanga. Thus for Topanga police station, we used the population of the town. Following this, the variables were then converted to factors for the purpose of analysis.

Other than these two variables, we utilized five other variables from the original dataset for our analysis: TIME.OCC, AREA, Vict.Age, Vict.Sex, Vict.Descent. We limited our variable choice to these five for several reasons. Our first reason is to prevent overlap, as some variables in the original dataset contain the same information in the realm of modeling–AREA and AREA.NAME, for example. Second, some variables can be viewed as post-hoc in relation to a crime incident. For example, the Status.Desc variable explains the current state of the criminal investigation, which has limited value as we are more interested in what factors influence a person's likelihood of being a victim of a firearm crime.

Of those five variables, Vict.Sex and Vict.Descent required further cleaning. The Vict.Sex variable included records of "H", "-", and NA, which weren't in the original dataset's codebook. These incidents were converted to "X" to properly denoted sex status as unknown. Similarly, the Vict.Descent variable included records of "-" and NA, which we converted to "X" to denote as unknown. Following these changes, no further cleaning was required.

2.3 Exploratory Data Analysis

To begin the analysis of crime rates, we counted the total number of crimes committed in each of the 21 areas of Los Angeles from 2020 to the present. Our goal was to explore the range and distribution of crime rates and gain an understanding of which areas are most crime-ridden.

Statistic	Value
Mean Number of Crimes	43367
Median Number of Crimes	42981
Lowest Number of Crimes	30123
Highest Number of Crimes	62071

Table 1: Table for Number of Crimes

Looking at some summary statistics, the average number of crimes is 43,367. The lowest number of crimes is about 30,000, and the highest is about 62,000.

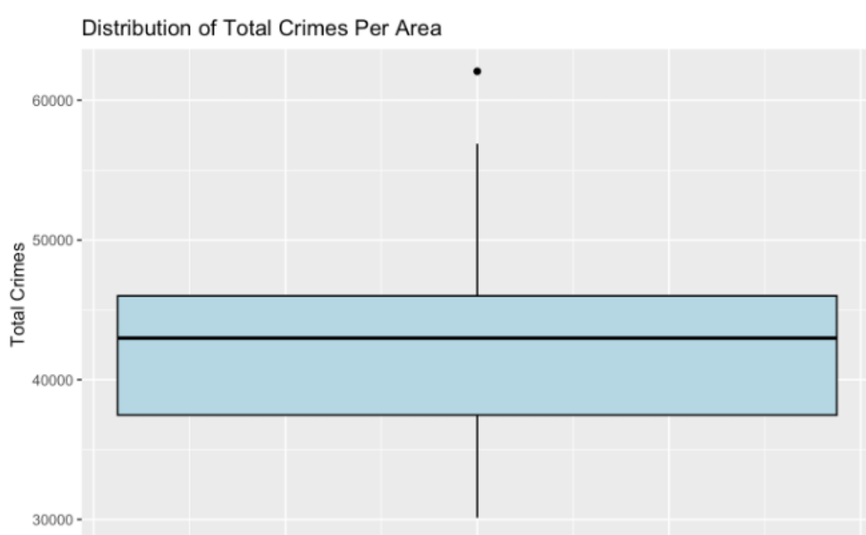

Figure 1: Number of Crimes vs Population Area

The histogram of total crimes per area shows the most common crime rates in the data. We observe that the most frequent number of crimes per area ranges from 35,000 to 40,000. The shape of the plot suggests that the data is right-skewed and not normally distributed.

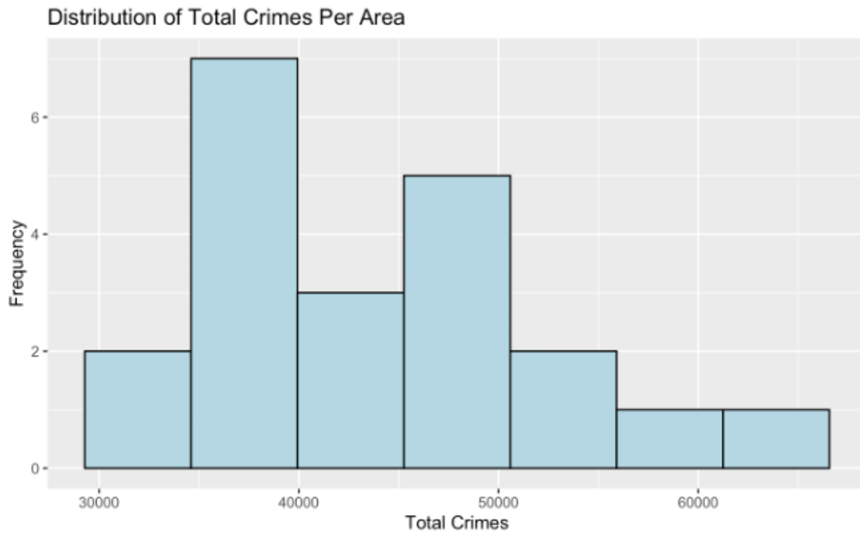

Figure 2: Number of Crimes vs Population Area

The box plot of total crimes per area illustrates the distribution of crime rates. We observe that the median number of crimes is about 43,000. The plot shows that approximately 50% of the data lies between 37,000 and 46,000. There is one outlier depicted by the dot at the top of the plot. This represents the highest number of total crimes, which is about 62,000.

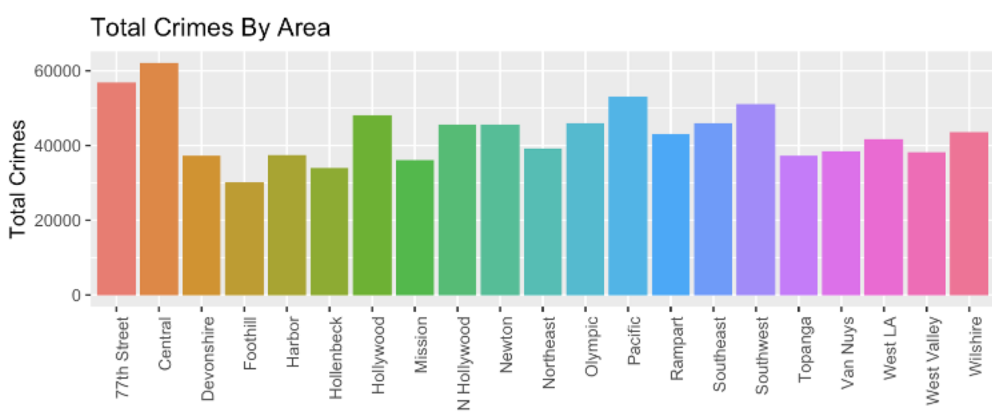

Figure 3: Number of Crimes vs Population Area

The bar plot of total crime by area illustrates the crime rate in each area of Los Angeles. From this visualization, we note that Central has the highest number of total crimes, followed by 77th Street. On the other hand, Foothill has the lowest number of total crimes, followed by Hollenbeck.

3 Results

3.1 Crime Rate Analysis

One of the research questions we wanted to answer with the L.A. crime data was which areas in L.A. are more dangerous than others. To answer this question, we broke it down into two parts:

(1) Are there proportionally more crimes in certain areas?

(2) Of the crimes committed, is the proportion of crimes that use a firearm higher in certain areas?

To answer the first question, we wanted to determine if the number of crimes committed in a given area was proportional to the area's population. If the area did not affect how many crimes were committed, then the number of crimes should scale exactly with the number of people living there. To visualize this, we plotted the number of crimes versus the population of each area.

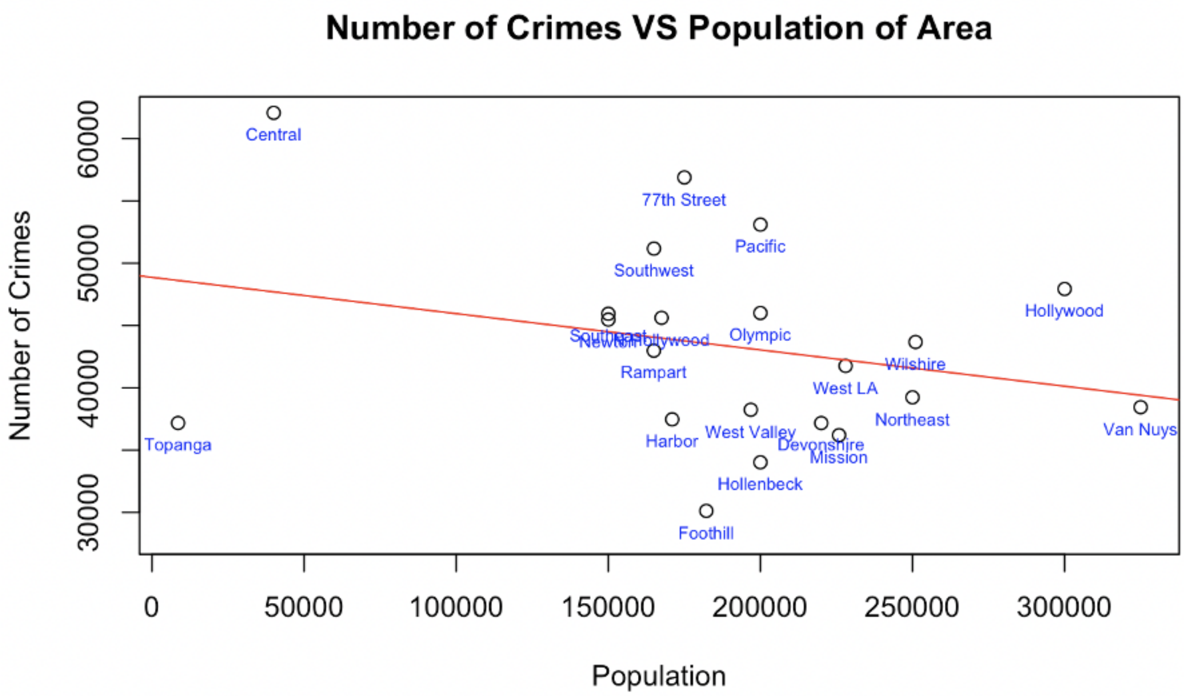

Figure 4: Number of Crimes vs Population Area

As shown by the above plot, there is little direct correlation between the number of crimes and the population of each area–with what there is being slightly negative. The correlation coefficient of this relationship is -0.2666, indicating a weak negative correlation between the variables. This corresponds with an R^2 value of 0.07108, which suggests that only about 7% of the variation in the number of crimes in an area can be explained by the population of the area.

To answer the second question, we performed a chi-squared test of significance to determine if the area had an effect on the number of crimes in the area, which we normalized to be per 1000 people and per year. The table used for the chi-squared test is as follows:

The Chi-square test had a resulting p-value of approximately 0 ($p < 0.001$), suggesting that there is statistically significant evidence of a relationship between the area and the proportion of crimes that use a firearm. We then ran pairwise comparison tests on the different areas to determine which were statistically significant. Topanga and Central were by far the most significant and overrepresented.

77th Street	Central	Devonshire	Foothill	Harbor	Hollenbeck	Hollywood	Mission	N Hollywood	Newton
32.50686	155.17750	16.89636	16.53166	21.91228	17.01050	15.97800	16.02885	27.23463	30.31667
Northeast	Olympic	Pacific	Rampart	Southeast	Southwest	Topanga	Van Nuys	West LA	West Valley
15.69440	23.00250	26.54900	26.05525	30.62933	31.01455	434.29907	11.82923	18.31535	19.42745
Wilshire									
17.40080									

Table 2: Table for Crime Per Area

3.2 Use of Firearms in Crime

Following our analysis of the general crime patterns in the Los Angeles area, we then proceeded with a similar method of testing. This time, however, we looked at the proportion of firearm crime to total crime in order to identify if any area had a disproportionately high rate of firearm usage. The data table we used is:

	77th Street	Central	Devonshire	Foothill	Harbor	Hollenbeck	Hollywood	Mission	N Hollywood	Newton	Northeast	Olympic
NoFirearm	51019	59947	36231	28948	35514	31889	46373	34864	44404	42023	37994	44369
YesFirearm	5868	2124	941	1175	1956	2132	1561	1337	1214	3452	1242	1636

	Pacific	Rampart	Southeast	Southwest	Topanga	Van Nuys	West LA	West Valley	Wilshire
NoFirearm	52101	41211	41409	48460	36346	37548	41247	37142	42403
YesFirearm	997	1770	4535	2714	830	897	512	1099	1273

Table 3: Count of Firearm-Related and Non-Firearm-Related Crimes by Area

The resulting Chi-squared test similarly had a p-value of approximately 0 ($p < 0.001$), suggesting evidence of a statistical difference between the areas and their proportion of firearm use in crimes. We then ran pairwise comparison tests on the different areas to determine which were statistically significant. 77th Street and Southeast were both significant and overrepresented. These findings can be seen through the following visualization.

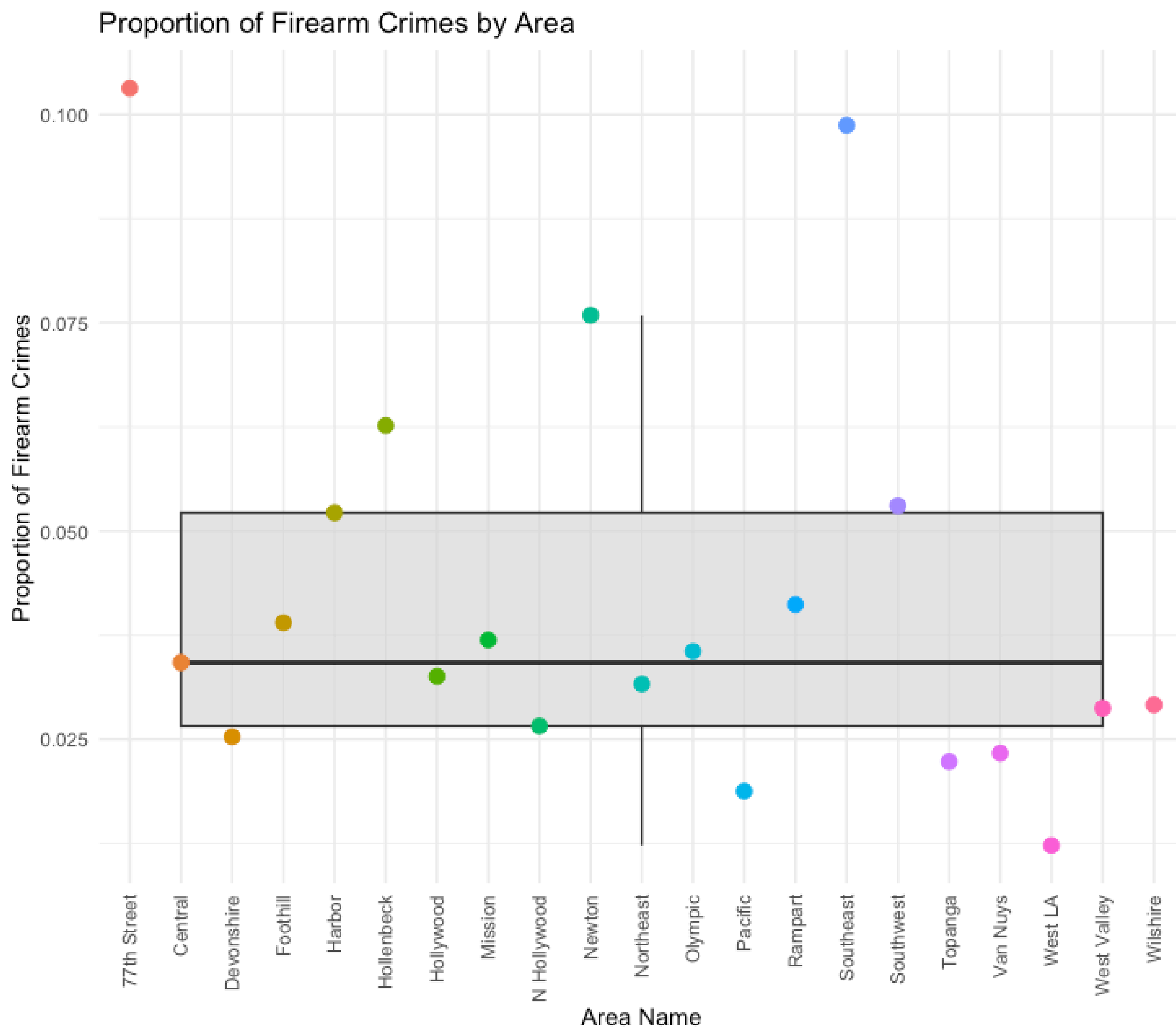

Figure 5: Proportion of Firearm Crimes by Area

3.3 XGBoost Model to Predict Firearm Use

To conclude our analysis of the LA crime data, we sought to create a predictive model to determine the likelihood that a crime involves a firearm. As mentioned previously, the features selected for this model were: TIME.OCC, AREA, Vict.Age, Vict.Sex, Vict.Descent, and population, and the outcome variable is isFirearm. We then partitioned our dataset using an 80-20 train-test split and created our XGBoost model. The resulting model had a high accuracy with roughly 95.6% predictive power. From this model, it was revealed that Vict.Sex_M and population were the most influential predictors–indicating that these features highly predict whether a crime involved a firearm or not. Below is the "importance plot" of our model.

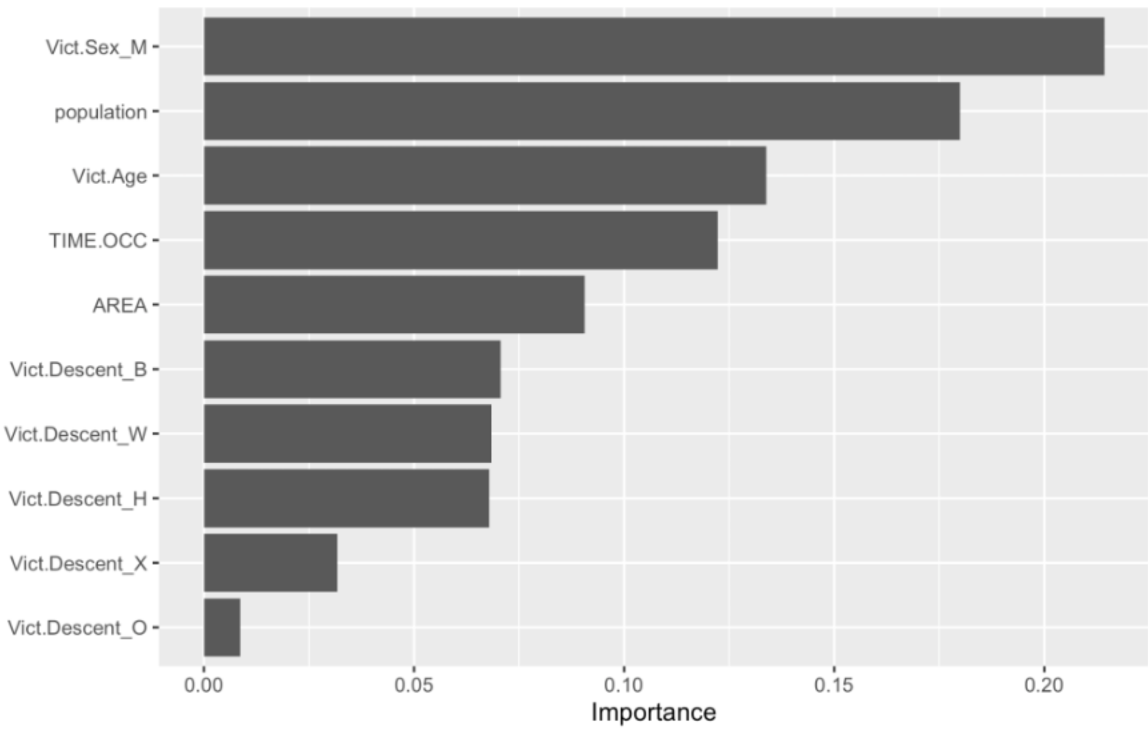

Figure 6: XGBoost Importance Plot

4 Discussion

4.1 Analysis and Limitations

We begin our discussion of our findings by first looking at the results of our Crime Rate Analysis. Our analysis identified that Topanga and Central had a disproportionately high crime rate after populations were standardized. While working through the numbers, however, we're skeptical of our Topanga findings as its crime rate is over 10 times the third highest crime rate in District 77th Street. As such, we believe our population value for Topanga does not properly capture the area that the Topanga police district covers. However, as there are no specific population estimates for this area, we believe it is of high importance that population data for the Topanga district be tabulated and accessible.

In our analysis of the use of firearms in crime, we found that the areas surrounding the 77th Street and Southeast police station had a disproportionately high rate of firearm use in crime. It is alarming, considering that these areas use firearms at roughly 3 times the median rate of our data. However, we do note that these police stations are right next to each other, roughly 2.5 miles apart. We hypothesize that, as these areas are close to each other and between Compton and Skid Row, we might be seeing an offshoot of illegal activities from those areas into these neighborhoods. Regardless, from our data, we cannot come to any direct conclusions, but the alarmingly high rate of firearm use in these areas indicates that further data on these areas must be gathered.

Finally, with our XGBoost model, we find that the results are fairly consistent with current research. For instance, our importance plot identified that being a male victim was a significant predictor of whether a firearm was used in a crime. A report by Allwood et al. in 2023 found that men were victims of gun violence at roughly three times the rate of women (6% vs. 18%). In addition, that same report found that black people were victims of gun violence at the highest rate, followed by Latinx and then whites. In our model, we found that black was the race with the highest importance in determining if a crime involved a firearm or not, followed by whites, then Latinx. However, it is important to note that in the report by Allwood et al., the difference in gun violence percentage was 9% to 7.9% between Latinx and whites. In our importance plot, we see that white and Latinx have almost the same amount of importance, so we believe our model still represents the current published literature.

But what is most interesting about our importance plot is that victim race had the least predictive value. Of our six variables, the victims' race ranked last in predictive power. This could imply that maybe the other five variables had enough predictive power that the victim's race did not need to be accounted for, or it could mean that in the general populace, people of different races will be victims of gun violence at roughly the same percentage.

4.2 Recommendations for Further Research

To build off our study, we would recommend that further research be done on police districts–namely in the development of quantifying population parameters like population density and socioeconomic status. Research by Hipp and Rousell (2013) indicates that population density and its relation to crime are complex and non-linear, and thus, devising a one-size-fits-all policy based on an area's population will likely not lead to meaningful intervention. In addition, socioeconomic factors like unemployment and inequality have long been shown to be significant determinants of crime (Buonanno, 2003). As socioeconomic data is typically not framed in a way relating to police districts, research into this area can be fruitful.

5 References

[1] Allwood, M. A., Robinson, J. N., & Kim, H. (2023). Youth Exposure to Gun, Knife, and Physical Assaults: Assessing PTSD Symptoms Across Types of Assaults, Race, Ethnicity, Sex, and Context. Journal of Interpersonal Violence, 38(21-22), 11545-11568. https://doi.org/10.1177/08862605231185300

[2] Bonta, R. (2023, June). Crime in California 2022. Open Justice Data. https://data-openjustice.doj.ca.gov/sites/default/files/2023-06/Crime%20In%20CA%202022f.pdf

[3] Buonanno, P. (2003). The socioeconomic determinants of crime. A review of the literature. Working Paper Dipartimento di Economia Politica, Università di Milano Bicocca; 63. https://boa.unimib.it/handle/10281/22981

[4] Department, L. A. P. (2024, March 6). Crime data from 2020 to present: Los Angeles - open data portal. Crime Data from 2020 to Present | Los Angeles - Open Data Portal. https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8

[5] Hipp, J. R., & Roussell, A. (2013). Micro- and macro-environment population and the consequences for crime rates. Social Forces, 92(2), 563–595. https://doi.org/10.1093/sf/sot091

[6] Home of the Los Angeles Police Department. LAPD Online. (2024, March 6). https://www.lapdonline.org/

[7] Katherine E. M. Hoops, M. (2017, November 1). BB guns that look like real guns. JAMA Pediatrics. https://jamanetwork.com/journals/jamapediatrics/fullarticle/2653918?utm_campaign=scite&utm_source=scite&utm_medium=referral

[8] Los Angeles Times. (2023, October 6). A troubling California trend: More violent crimes with guns even as restrictions tighten. Los Angeles Times. https://www.latimes.com/california/story/2023-10-06/a-troubling-trend-in-california-more-violent-crimes-are-being-committed-with-guns-even-as-restrictions-tighten

[9] Oliveira, M. (2021, December 1). More crime in cities? on the scaling laws of crime and the inadequacy of per capita rankings-a cross-country study - crime science. BioMed Central. https://crimesciencejournal.biomedcentral.com/articles/10.1186/s40163-021-00155-8

[10] Wikimedia Foundation. (2024, February 2). LAPD Rampart Division. Wikipedia. https://en.wikipedia.org/wiki/LAPD_Rampart_Division