This analysis aims to use the goals of Exploratory Data Analysis (EDA) to answer the question: Does policing differ by community demographics, namely race and income level?

The analysis uses data on police accountability, arrests, and demographics for California law enforcement agencies (namely police and sheriff departments) for 2016 to 2018. The data were retrieved from the INFO 290 course website, and originally produced by Campaign Zero and their project 8 Can't Wait (Campaign Zero, n.d.).

Prior to beginning the analysis, the data were reviewed to determine whether any data cleaning was needed. In reviewing the data, it was determined that the police arrest csv file provided incorrect labels for three of the population by race columns, based on a comparison against U.S. Census Bureau data (U.S. Census Bureau, 2020). Specifically, the column containing the Hispanic population count was labeled as 'White Population', the column containing the white population count was labeled as 'Black Population', and the column containing the Black population count was labeled as 'Hispanic Population'. These labels were corrected within Jupyter Notebook. No other data cleaning was performed.

To explore how policing might differ by race, for the three California regions—Central, Northern, and Southern—the analysis compared the average proportion of Black people within the population against the average proportion of drug arrests where the person arrested was Black. As shown in Figure 1, in all regions, police arrest Black people for drug arrests at higher rates than would be expected based on the proportion of Black people within the population. The difference between the two proportions is greatest in Northern California, where proportion of drug arrests of Black people is nearly three times that of the proportion of Black people within the population.

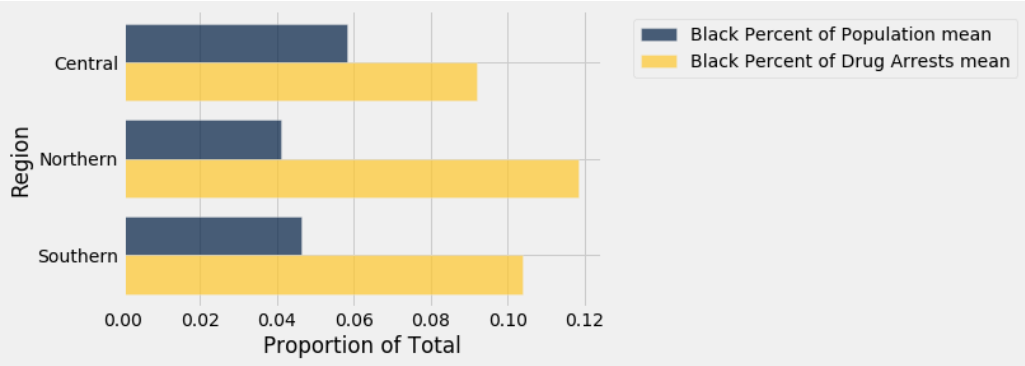

Figure 1. Bar chart of average proportion of Black people within population and average proportion of drug arrests where police arrest a Black person by three regions in California.

Digging deeper into these findings, the analysis compared the distribution of the proportion of Black people that make up the population within each law enforcement jurisdiction against the distribution of the proportion of drug arrests where law enforcement arrested a Black person for each jurisdiction. Figure 2 overlays histograms for these two metrics, indicating that police arrest Black people at higher rates than would be expected based solely on population racial makeup, as the drug arrest histogram is shifted to the right relative to the population histogram. The drug arrest distribution is more right-tailed than the population distribution, indicating that arrests are particularly high relative to population figures in some areas. Oakland represents one of the more extreme examples, where police arrested a Black person in 72% of drug arrests, despite the fact that Black people made up only 24% of the population.

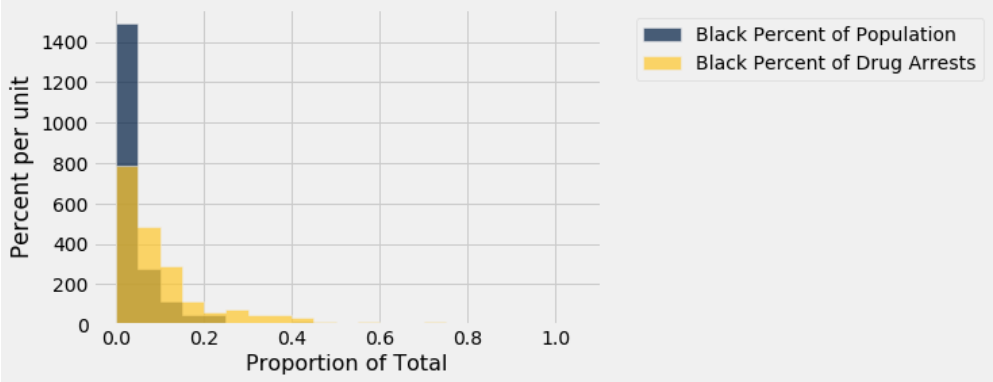

Figure 2.Histograms of the proportion of the population in each law enforcement agency jurisdiction made up by Black people and the proportion of drug arrests in each law enforcement agency jurisdiction made up by Black people in California.

In comparison, the distribution of the proportion of white people that make up the population of a given jurisdiction more closely mirrors the distribution of the proportion of drug arrests where police arrest a white person, as shown in Figure 3, below.

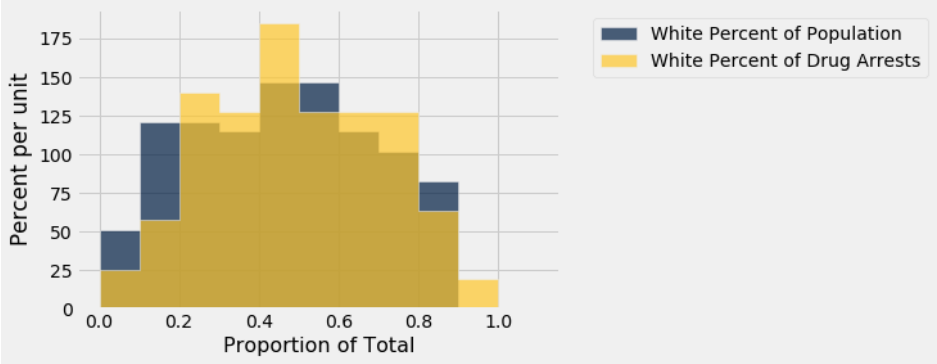

Figure 3.Histograms of the proportion of the population in each law enforcement agency jurisdiction made up by white people and the proportion of drug arrests in each law enforcement agency jurisdiction made up by white people in California.

The data also indicate that policing varies by the wealth of the community being policed. Specifically, police appear to be most violent in poorer areas. As shown in Figure 4, the agencies with the five highest police violence scores police jurisdictions where median income is less than $55,000; the mean of the median income of these jurisdictions is about 25% lower than the mean of all law enforcement jurisdictions. Similarly, the agencies with the ten highest violence scores police communities where the median income is less than $80,000, and the mean of the median income of these jurisdictions is about 13% lower than the mean of all jurisdictions.

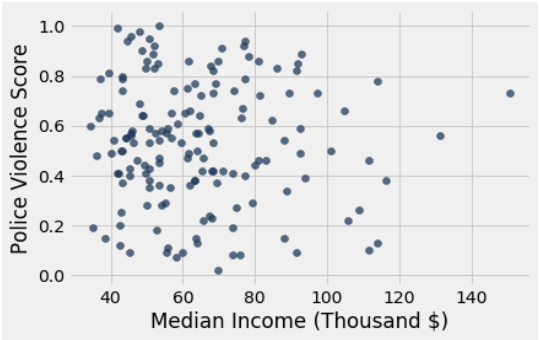

Figure 4. Scatterplot of police violence score for law enforcement agencies in California vs. median income of the jurisdiction that they police.

This EDA indicates that policing outcomes appear to differ based on the demographics—specifically race and income levels—of the individuals and communities being policed. Future analyses might explore the relationship between additional metrics of policing (e.g., accountability, violence, etc.) and community sociodemographic variables (e.g., income, race, estimated population of undocumented immigrants). Analyses may also explore the relationship between accountability measures (e.g., whether a jurisdiction ensures police oversight and discipline for misconduct, erases misconduct records, etc.) and policing outcomes (e.g., police-caused deaths, serious injuries, shootings, etc.).

References

U.S. Census Bureau. (2020). QuickFacts. https://www.census.gov/quickfacts/

Campaign Zero. n.d. Join Campaign Zero. https://www.joincampaignzero.org/#vision