Crime in Los Angeles

Madison Razook
University of Colorado Boulder
mara5374@colorado.edu

David Stone
University of Colorado Boulder
david.stone-1@colorado.edu

1 Problem State/Motivation

Using the City of Los Angeles' "Crime Data from 2010 to the Present", we will look for interesting patterns in the victimhood of violent and property crimes. The Los Angeles Times, which maps crime rates in Los Angeles County, defines violent crime as homicide, rape, aggravated assault, and robbery.[1] Property crimes include burglary, theft, grand theft auto, and theft from vehicle.[2] Los Angeles is the US's second most populous city, with LA County being the most populous county in the country. The area is diverse in terms of both ethnicity and income.[3] Examining what happens in this city will not only benefit its millions of residents, but may also shed some light on crime trends in other areas of the United States. Through our analysis of this data, we seek to find the relationship between victim age, sex, ethnicity and the type of crime they were targeted for. The answers to this research question may help create a short-term solution for crime, in the form of safety advice for those individuals who are most at risk. To truly address crime at its core, however, we must also look at the perpetrator. Understanding who is targeted and for what creates a clearer picture of the offender's motivations. Other possible research areas include the relationship between the time the crime was reported to the police and the likelihood of arrest, as well as any interesting patterns in crime and arrest rates from 2010 to now.

2 Literature Survey

The Los Angeles Police Department (LAPD) devotes a section of its website to research.[4] There are several reports that detail crime rates for 2013-2015, list possible reasons for these numbers, and explain the LAPD's efforts to combat crime. Additionally, the 2016 year-end crime statistics subsection contains charts that document the number and type of crimes committed during the last three months of the year in each division of the city, along with their percent change between months. The last items in the LAPD's research section are year-end use-of-force reviews for 2015 and 2016. The information here examines how the LAPD's treatment of suspects varies according to suspect race, affiliations, and other categories. While the information here is not as extensive as the eight-year data set we will use for our project, it does highlight the most recent crime trends for us. Knowing LAPD's recent tactics may also help us understand any patterns in the arrest rates that we discover.

The Los Angeles Times has also used data maintained by the City of Los Angeles to create public sources of information. In 2009, they launched their Mapping LA project, which provides maps and statistics for all of LA County's neighborhoods.[5] Of interest to our work is the subproject Crime LA.[6] Crime LA is a map showing violent and property crime rates for each neighborhood, highlighting those areas which have recently seen an increase in either type of offense. As of this writing, data from the February 20th-February 26th, 2018 time period provided the map's most recent update. In addition to this service, Crime LA also contains rankings for each neighborhood based on the number of violent or property crimes per 10,000 people, taken from the July 31, 2017- January 28, 2018 time period. Similar to the LAPD's research, this resource quickly updates readers on the most recent criminal activity, but is not able to show long-term crime trends.

A 2010 data analysis done by the Los Angeles County Department of Public Health looked at the effects of homicide on life expectancy by neighborhood and ethnicity in LA County.[7] The study used cause elimination techniques on mortality records and population estimates. Results indicated that the South Service Planning Area (SPA) of LA had higher poverty levels and percentages of black and Latino residents compared to other areas. Homicide was estimated to reduce the life expectancy of black males by 2.1 in LA County and by almost 5 years in low-income urban areas. 82.4% of homicide deaths victimized people between 15-44 years of age, with the majority of these deaths caused by a firearm.

The study concluded with several strategies for homicide reduction based on this information, noting that high homicide rates are often correlated with low levels of social cohesion. With more recent data, we can see if young black males are still most at risk for homicide deaths, and in which areas they are particularly vulnerable. These results could be used to measure the effectiveness of any homicide reduction strategies used since 2010.

A 2013 study published in the University of Pennsylvania Law Review measured the effects of zoning on crime in LA.[8] The authors first looked at crime rates in eight neighborhoods with high levels of crime but different forms of zoned land use. A second method compared two groups of neighborhoods with similar crime trajectories, with one group experiencing zoning changes while the other did not. Zoning changes, mostly in which parcels were converted to land use, led to a significant reduction in crime. The study concluded that mixing residential-only zoning with commercial blocks may serve to reduce crime overall. The results of our project may suggest crime trends that corroborate or disprove this idea.

3 Proposed Work

3.1 Preprocessing

We will first need to clean our data and throw out any data points with missing information. Since this data is transcribed from original paper reports there is guaranteed missing points such as unknown location, victim gender, or victim descent. We will need to transform our data to be more compatible with the tools we our using: Jupyter, Pandas, Matplotlib, SciPy, and scikit-learn which includes converting all numeric values into floating point values so there is never a rounding error or loss of data as well as converting the dates into date objects. Furthermore, we will need to individually label each type of crime as violent or property crime to be able to reduce and split are dataset into these two subsets.

3.2 How our work will differentiate from past work

Our project will differentiate itself from previous work by first focusing on the victims of crimes and their personal attributes and not the suspects of crimes as well as focusing on the long-term trends of the different areas and neighborhoods of Los Angeles over the past 8 years. Furthermore, we will differentiate between violent and property crime to give a more detailed account of the crime rates.

4 Data Set

Our dataset consists of reported and documented crimes with the LAPD dating back to 2010. It includes the date of the crime, the date it was reported, the victim's gender and descent, and area where the crime occurred.
https://catalog.data.gov/dataset/crime-data-from-2010-to-present

5 Evaluation Methods

5.1 Classification

As we see long term crime trends of the different areas of Los Angeles, we can classify the current crime risk as well as predict future risks based on the past. If an area has had a high risk of crime in the past but have been gradually decreasing, we can predict that trend to continue and suggest it has a safe area in the future.

5.2 Clustering

We will cluster and specify the two different types of crime: violent and property. If a location has a general high crime rate but is caused by a majority of property crime and a low rate of violent crime, we can classify that area as safer for a person's well-being than an area with the same amount of general crime that consists of more violent crime.

6 Tools

We will be using Jupyter Notebooks, Pandas, and Matplotlib for cleaning, analyzing, and visualizing our data.

7 Milestones

- Preprocessing, cleaning, and transforming done by week of March 12
- Classifying locations by crime rate by week of March 19

- Classifying victims' gender and descent by likelihood of a crime happening to them by week of March 26
- Repeat classifications after clustering violent and property crime by week of April 2
- Visualize and find trends by week of April 9
- Start putting together final paper by week of April 16

7.1 Milestones Completed

7.1a Preprocessing

- Cleaning the data and throwing out any data points that have missing attributes. Each data point has twenty six attributes and if six out of twenty six are null values, that data point is removed. Because we have a very large amount of data, we decided it would be cumbersome to fill in missing data so we removed these data points completely instead of filling in.

- Transforming the data into consistent readable and graphable values. All integers converted into floating point values so all numerical attributes can be compared and manipulated without attribute type conflicts. All dates were converted into date objects.

- Because the type of crime is a nominal string attribute, we had to find every unique crime code associated with the type of crime and label it as a violent or property crime. This took a long time to do but once it was completed, we could split on the attribute of property or violent crime. This is significant to our motivation as we want to give specific statistics of not only the general crime rate of an area but whether it has a higher chance of violent or property crime.

- Data reduction was completed multiple times to compare specific attributes in more detail. These reductions included separating the attributes of property crime and violent crime which was a complicated process to make lists of the crime codes that were associated with violent and property crime, respectively then sort the data accordingly. Within those subsets we compared the date the crime was committed, the date the crime was reported, and whether or not an arrest was made.

7.1b Classification

We began building a Naive Bayesian classifier to find correlations and find the probability of questions we came up with. These questions used the attributes of location, age, sex, race, and type of crime. Some questions we wanted answered were: What neighborhoods were most likely to have violent or property crimes? Does age and race of the victim correlates with the crime that was committed against them? As the length from when the crime occurred to when the crime was reported increases, does the probability of an arrest decrease?

7.1c Clustering

Our data has real world locations and coordinates where clustering helps immensely to visualize and organize our data. So far, we have used k-means clustering to mine any interesting relationships between location, victim age, weapon used, and time of the incident.

7.2 Milestones Todo

Preprocessing is completed at this point and we are currently in the stages of finding patterns and correlations. We have our cleaned original dataset and two subsets for violent crime and property crime. We need to combine our findings to see if there is a high chance that someone of a specific age, race, and location has a high likeliness of being a victim of either violent or property crime.

Although we have found trends and have ranked areas by the two types of crime: violent and property, we have yet to see a statement that is beyond trivial.

Once we have found correlations and patterns, we will need to also be able to predict future crime patterns. If the violent crime rate has been falling over the past years in specific locations, we can predict those neighborhoods will continue to become safer even if the property crime has stayed the same or has risen.

Finally, we will need to focus on visuals for our final presentation that show readable and understandable

graphs and tables. We will need to experiment with bin size and type of graph to correctly display our findings.

8 Results so far

Although we have yet to have uncovered an insightful and untrivial correlation or pattern, we have started comparing variables to see what significant relationships we can find. Our thought process for figure 1 was looking for a significant difference in the victim's gender and the type of weapon used in the crime. 100s are for guns, 200s for knives, 300s for blunt objects, 400s for fists and feet, and 500s for miscellaneous weapons. The dataset had four options for gender: male, female, and two non-binary categories H and X. Each option was assigned a numeric value: 0 for male, 1 for female, 2 for H, and 3 for X. The results seem to say that the weapon type does not depend on the gender of the victim.

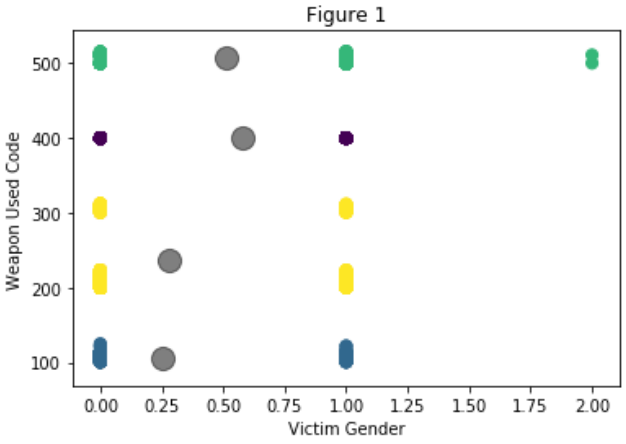


Figure 2 shows the comparison of the victim's age and where the crime occurred. Area ID 1 is for Central L.A. and is the most densely populated and has the highest amount of crime out of around 20 neighborhoods within the city limits. The results do not point towards different neighborhoods having higher crime towards a certain age group.

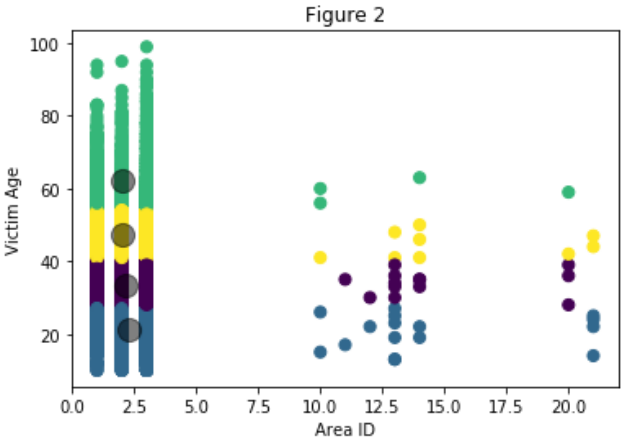
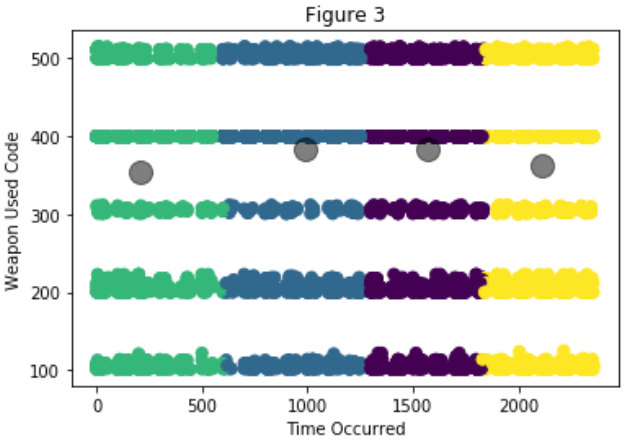
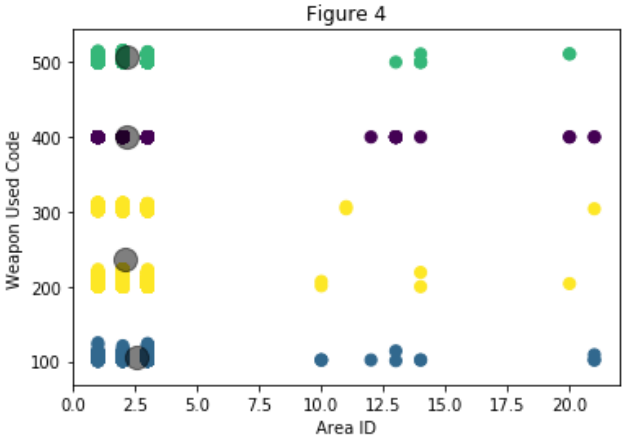


Figure 3 shows the comparison of the type of weapon used and the time the crime occured. This was trying to answer the question of whether or not different weapons were used at different times. The graph shows there is no time of day one weapon is used more than the other.



With Figure 4, we attempted to find an interesting connection between area and weapon used. Areas in the lower range tended to see more weapon use, with 400s being the least used category; beyond that, the clusters here are consistently sized and offer little information.



We will continue to cluster attributes and make comparisons to find interesting and nontrivial observations about our dataset and will also explore multiple types of graphs and clustering algorithms.

REFERENCES

[1] Los Angeles Times. 2018. Violent Crime Ranking - Mapping L.A. Retrieved from http://maps.latimes.com/neighborhoods/violent-crime/neighborhood/list/

[2] Los Angeles Times. 2018. Violent Crime Ranking - Mapping L.A. Retrieved from http://maps.latimes.com/neighborhoods/property-crime/neighborhood/list/

[3] U.S. Census Bureau. 2018 Quick Facts: Los Angeles City, California. Retrieved from https://www.census.gov/quickfacts/fact/table/losangelescitycalifornia/PST045216

[4] Los Angeles Police Department. 2018. Research. Retrieved from http://www.lapdonline.org/research

[5] Los Angeles Times. 2018. Mapping L.A. Retrieved from http://maps.latimes.com/neighborhoods/

[6] Los Angeles Times. 2018. L.A. Crime Maps - Mapping L.A. Retrieved from http://maps.latimes.com/crime/

[7] Redelings, M., Lieb, L. & Sorvillo, F. J Urban Health (2010) 87: 670. DOI: https://doi.org/10.1007/s11524-010-9470-4

[8] Anderson, James M., John M. MacDonald, Ricky Bluthenthal, and J. Scott Ashwood. Reducing crime by shaping the built environment with zoning: an empirical study of Los Angeles. *University of Pennsylvania Law Review* 161, no. 3 (2013): 699-756. JSTOR: http://www.jstor.org/stable/23527820.