The Victims of Los Angeles

by Stephen Li and Amil Ahsan

Abstract:

Los Angeles is the second largest city in the United States, and home to over 3,976,324 people. Given Los Angeles's prominence in California, it is a popular spot for tourists to visit and for people to settle down and live, so it is important to learn the potential risks to visitors in Los Angeles. We seek to find if there is a correlation between some of the victim's demographics, and what type of crime is committed. We used histograms, box plots, linear models, and logistic regression to determine whether age and gender of an individual were good predictors to determine if they were more likely to being a victim of a crime. From our results, we determined that gender was not a significant variable; however, age proved to have a notable effect. Older people seem to be more likely to experience harmless crimes and younger people were more likely to experience a crime with an actual threat.

Introduction:

Our study revolves around the city of Los Angeles, the second largest city in the United States and is home to over 3,976,324 people (as of the 2016 ACS census). Due to the city's rich culture and proximity to prominent tourist attractions such as Disneyland, Universal Studios, and Hollywood, Los Angeles can be considered a tourist hotspot. Given Los Angeles's high population, it is more likely that criminal activity such as theft, assault, vandalism, rape, etc. will occur at a higher rate than other cities. Unfortunately, the city's crime rate has also increased since 2015, and as a result Los Angeles has 83.7% more crime than other cities in the United States (http://www.city-data.com/).

While the chance that you will be the victim of a crime is fairly low, it is still important to recognize the potential risks if you are planning to live in or visit Los Angeles. We will perform studies, analyzing whether certain demographics of people are more susceptible to threats than others; that is whether they are more susceptible to being physically attacked or threatened.

We classified crimes into two categories, the first of which consists of all crimes that don't pose any direct threat onto the victim (class B). These would include crimes such as vandalism and theft. The second category is composed of crimes where the victim can be physically harmed (class A). Assault, murder, and rape are a few examples that fit into this one. We hypothesized that younger people will experience the majority of class A crimes, whereas older people will be the primary targets of class B crimes. Our justification for this line of thinking lies in the fact that younger people tend to go out more and that older people are more vulnerable to easier crimes like theft. Keeping in mind that there is a social standard of woman being more valued and respected than men, we believe that men will be victimized more than women.

Methods and Materials:

Our data covers all crimes reported by the Los Angeles Police Department from 2010 to the present day. Since the data is incredibly large, we randomized the data and selected a sample of about 10,000 people. One problem we ran into was defining the predictors, since much of the data was categorical and hard to define such as location, time, weapons, etc. We defined a threat as any action that directly impacts the victim's feelings of safety and involve the use of some kind of weapon, physical threats, or intimidation (class A). Threats were operationally defined as '1' and crimes that did not involve threats (class B) were defined as '0'.

Similarly, we defined males as ‘1’ and females as ‘0’. Since age was not a categorical variable, we were able to leave it as is.

In order to analyze the data, we first wanted to identify whether age had an effect on being targeted by any crime in general (both class A and B). Using age as the x-axis and number of victims as the y-axis, we created a simple histogram to portray the extent to which each age group is affected by crime.We then used a box plot to study the distribution of class A and class B crimes with age on the y-axis. This will allows us to contextualize the effect of age on whether a victim would be involved in a crime with threats or without threats.

Next, we fitted a linear model and a logistic regression with a binomial distribution using age as our predictor and threat as our response variable, plotting the predicted results against age. This determined whether age was a significant variable, and to see if we could find a correlation between age and whether there was a threat or not. In other words, we wanted to investigate whether age would have an influence on the likelihood of being the victim of a class A crime. We also used this analysis to see if there was going to be a significant difference between using logistic regression and linear models.

Gender was the next factor to consider. To get an initial idea of how gender had a role, we first conducted a similar analysis to the age histogram. Taking gender as a categorical independent variable and the number of victims as the dependent variable, we created a histogram to show the distribution. We then performed a logistic regression with a binomial distribution, using gender as our predictor and threat as our response variable. Similarly, this was also done to see if gender was a significant variable and if there was a correlation between age and threat.

Finally, we used a linear regression using gender and age as our predictors, which was done to compare the age and gender variables.

Results:

To get an idea for what the rest of the experiment would look like, the age and gender histograms were found first. We found that age does have a role in the crime rate, but gender does not.

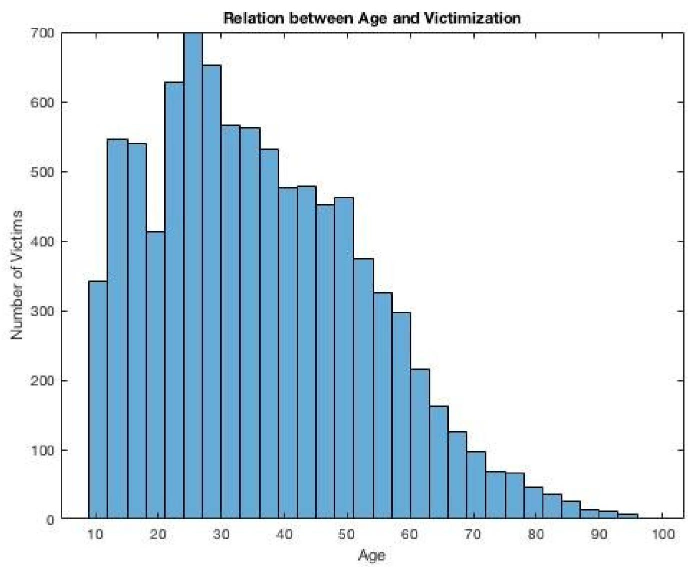


Figure 1: Relation between Age and Victimization. A correlation between age and victimization can be observed in this spread of data. The rate peaks at around the 20-40 range and slowly descends to a value of 0.

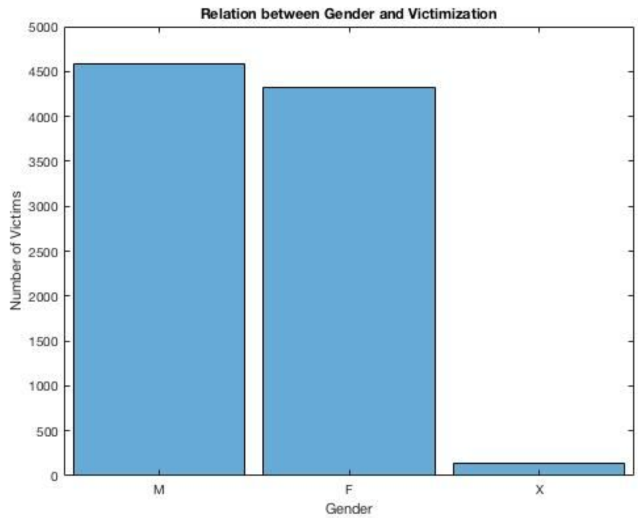


Figure 2: Relation between Gender and Victimization. This shows the amount of crime that each gender experienced from our dataset. The distribution is roughly half-half so it can be assumed that gender does not affect general crime in any way. The third bar can be neglected because it represents blanks or unknowns in the dataset, while we are testing strictly for males and females.

By looking at the box plot we observe that the distribution of ages between crimes that involve threats and crimes that do not involve threats is fairly similar, but significant enough to consider studying. Both that data tended to skew towards younger ages, but moreso for class A crimes. A notable difference is that the upper adjacent for class B crimes is 88 and 78 for class A crimes. Median is 35 for class B crimes and 32 for class A crimes.

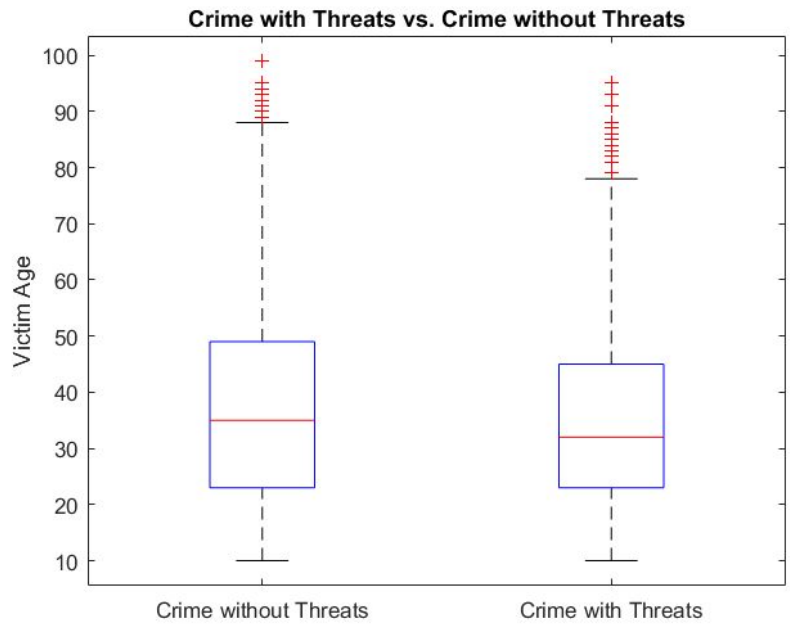
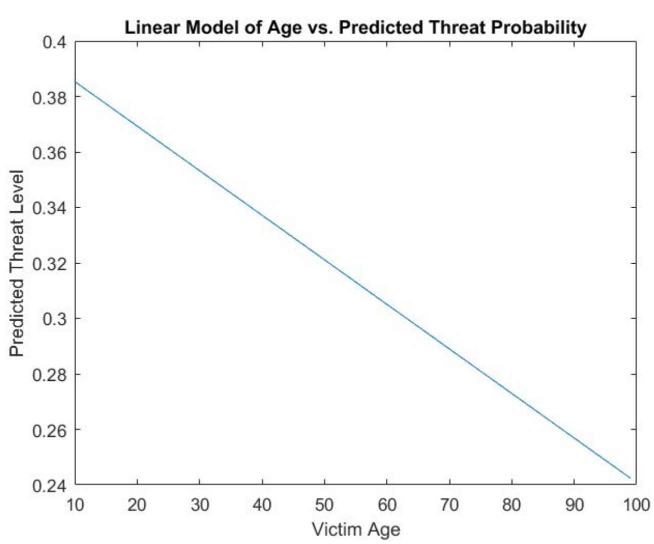


Figure 3: *Boxplot of Crimes with Threats vs. Crime without Threats. The y-axis represents the age of the victim. The boxplots show the upper and lower ends of the data, the median and the 25th and 75th percentile. The lines represent the outliers.*

For the linear model and logistic regression of age, both models show that victim age is a significant variable. For the linear model our p-value was 3.8292e-08 and for the logistic regression our p-value was 4.1085e-08. Furthermore, if we look at the graphs they appear to be very similar, both visually and value wise. Both graphs shows a line with a downward slope.



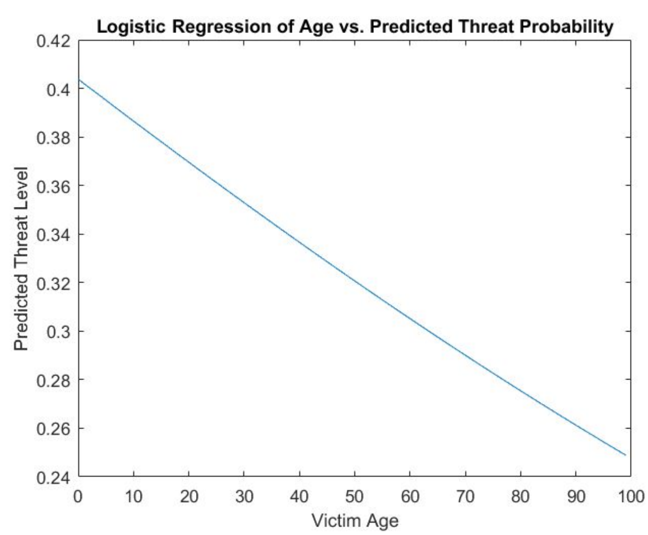


Figure 4: Linear Regression of Age vs. Predicted Threat Probability. The line shows that the likeliness of being the victim of a class A crime goes down as you get older.

Figure 5: Logistic Regression of Age vs. Predicted Threat Probability. The curve is very slight, so the graph resembles a line, but the results are similar to Figure 2.

For the logistic regression of gender, we obtained a p-value of 0.30109 for gender, with '0' for females and '1' for males. This shows that gender was not a significant variable for our analysis. By plotting the data we observe a line with an upward slope and an extremely small curve.

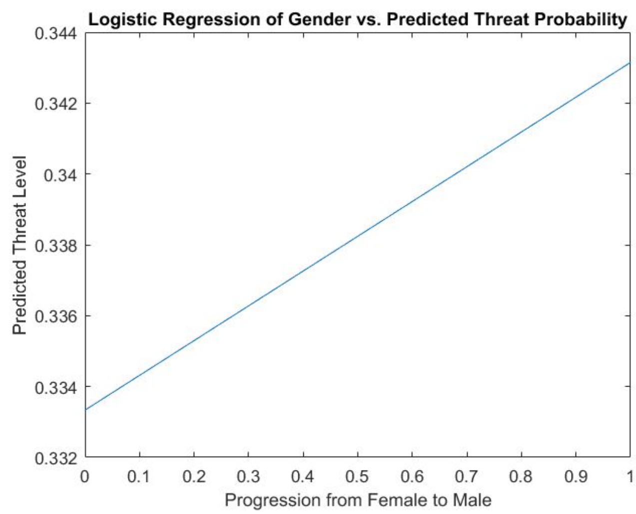


Figure 6: Logistic regression of Gender vs. Predicted Threat Probability. It appears to look like a line, but the small difference between the minimum and maximum of the graph suggests that gender has a very small effect on the possibility of a threat.

For the logistic regression of age and gender, age continues to be a significant variable since its p-value is 1.0868e-08 while our p-value for gender is 0.054445.

Discussion and Conclusions:

From our analysis, we observed that gender most likely does not have a significant impact on whether a person will be victimized and how a person will be victimized. After using logistic and linear regressions and comparing its effect with age, it has consistently shown to be insignificant enough judging by our p-values. However, age did have an effect, from our results from the histogram, box plot, and linear and logistic regressions. Overall, class A crimes tended to skew a bit toward the younger population. In comparison to the 2016 census data of Los Angeles, we see that the distribution for class A crimes matches Los Angeles's population a bit more. This shows that older people may be victims of theft more than younger people, since that is what most of the class B crimes were categorized as.

Looking at our linear and logistic regression plots, we can see that it is much more likely overall that you will experience a crime without a threat since the maximum of the lines were closer to '0' than to '1'. We also observe that age has decent but not dramatic effect on whether a crime involves a threat or not. The difference between the maximum and the minimum is 0.143 for the linear regression and 0.1543 for the logistic regression.

While age did meet our hypothesis, we were a bit surprised to see that gender did not have a significant effect. This was due to our preconception that men can be viewed as potential targets more easily than women. In the future, we think there a number of improvements that could be made. One must remember that this is only a small portion of the whole dataset. Although the dataset is reasonably large, our data analysis will most likely differ than that of the whole dataset. Our methods of analysis also are vulnerable to outliers, even if we analyzed the entire dataset. To account for this, one might want to choose a more outlier resistant method. A more skilled data analyst could have explored the data using methods such as support vector machines, best subset regression, or just refine our existing analysis to perform a more complex analysis.

References

"Crime Data From 2010 To Present | Los Angeles - Open Data Portal." Data.lacity.org. N. p., 2017. Web. 30 Nov. 2017.

"Crime In Los Angeles, California (CA): Murders, Rapes, Robberies, Assaults, Burglaries, Thefts, Auto Thefts, Arson, Law Enforcement Employees, Police Officers, Crime Map." *City-data.com*. N. p., 2017. Web. 30 Nov. 2017.

"Census Profile: Los Angeles, CA." *Census Reporter*. N. p., 2017. Web. 30 Nov. 2017.