Stat 272 Final Report

JonJeng Thao and Otto Weigel

1 Introduction

Burglaries are a common occurrence in the majority of U.S. cities. In 2023, the burglary rate was about 250.7 cases per 100,000 people nationally. Several factors have been identified in relation to local fluctuations in burglary rates, particularly the type of residence and residential density (Vandeviver and Bernasco, 2019). Places that had been recently burgled were also more likely to be burgled again over places that had not experienced a burglary in the recent past, potentially leading to a positive feedback loop.

Aside from burglary rates, another aspect of interest is the time it takes people to report a crime to the police. In recent years, many people have become less trusting of police, which may affect how quickly someone reports a burglary to the police or if they report it at all. For example, someone who has previously had a poor experience with police or knows someone who has may be more apprehensive towards reporting a crime. Similarly, overall distrust of police has been found to decrease the rate at which people report crimes to police, though this depends on the type of crime (Davis and Henderson, 2003).

By identifying factors that affect the rate of reporting and time to report, we can begin to determine potential strategies to alleviate wariness to report. Based on previous research, we expect that factors such as ethnicity/descent and age could have an effect on the time it takes for someone to report a burglary.

2 Materials and Methods

To carry out this study, we used crime data from the Los Angeles Police Department. The LAPD logs each crime that is reported in Los Angeles, and this dataset is a collection of all the crimes committed in LA between 2021 and 2023. Demographic, time, and location data were collected, with the most important being the victim age, victim sex, victim descent/race, the date and time of the crime, the date the crime was reported, and the type of weapon used in the crime.

These variables were then adjusted to make them more workable than previously. The time of the crime was originally collected as numeric military time, but we changed it to a binary variable "isDay", where a one indicates that the crime occurred during the hours of 6:00 am and 6:00 pm and a zero indicates 6:00 pm to 6:00 am. We implemented a similar change with the type of weapon, creating a binary of whether a weapon was used or not called "weaponUsed". To find the number of days it took people to report the crime, we subtracted the date occurred from the date reported. Due to the significant skewing of this data and the large amount of zeros, we decided to make

another binary variable ("sameDay") with a positive result indicating that the victim reported the crime the same day and a negative result indicating it took at least a day for the victim to report the crime.

From these additional variables, we then filtered the dataset, looking only at crimes reported as burglaries. We also filtered for cases where the victim's age was known, as we wanted to use age as a potential predictor variable and having unknown values would complicate this. We also filtered the victim sex variable to just include men and women, as it allowed us to format it as a binary variable and the two groups made up the vast majority of the reported crimes. Similarly, we filtered victim descent to include people recorded as asian, black, hispanic, other, and white (A, B, H, O, and W, respectively), because, like the victim sex variable, these victim descents made up the majority of the burglaries and it helped simplify the categorical variable. We also filtered the burglaries for premise code, looking at only residential burglaries that happened at either single family residences or multi-unit dwellings. Lastly, we looked only at burglaries where the burglary was reported within a month (30 days) of its occurrence. Like the rest of the variables we filtered, by far most of the burglaries were reported within a month and it was unlikely that someone would report a burglary over a month after it happened. In certain cases, it also appeared burglaries that were reported a long time after they happened might have been due to a clerical policy, and this filtering helped to eliminate any potential recording errors.

Initially, we planned to model the count data of the "days.to.report" variable using a linear regression, but the data were very skewed and thus did not meet the criteria for linear regression. However, because logistic regressions do not have the same criteria, we decided to model the probability of someone reporting a burglary on the same day the burglary happened. We used "sameDay" as the dependent variable and the time of day (day or night), time occurred, weapon use, victim sex, victim descent, victim age, and premise code as the explanatory variables. To select the best model, we used the stepwise selection method. To assess individual models against one another, drop in deviance tests were utilized.

3 Results

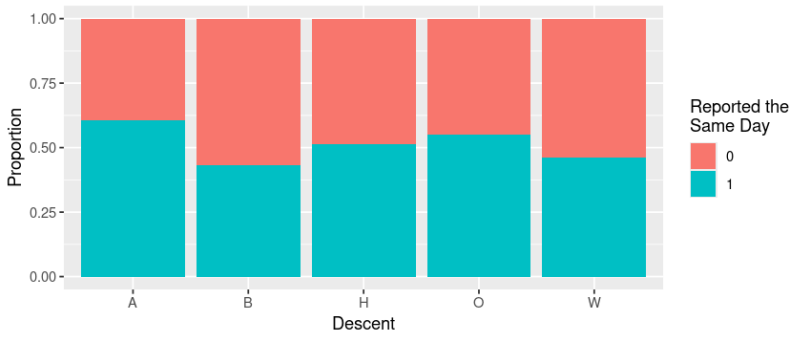

Figure 1: Bar chart of the proportion of people reporting a burglary within the same day, by descent

Using stepwise selection, we determined that the best model from our predictor variables was:

$$log(\frac{\pi}{1-\pi}) = 0.113 + 0.375 \times SingleFamily - 0.604 \times Black - 0.331 \times Hispanic - 0.223 \times Other - 0.577 \times White + 0.264 \times Weapon + 0.0752 \times Daytime - 0.11 \times Man \quad (1)$$

Table 2 has a more complete view of the coefficients, their standard errors, test statistics, and subsequent p-values.

This model takes the baseline demographic as an Asian woman living in a multi-unit building who has been burgled at night. Our results from the logistic regression model suggest that 'singleFamily' is the most significant variable in determining whether a burglary victim reports their house on the first day. The exponentiation of 'singleFamily' was 1.45. Meaning that burglary victims living in single family homes are 45% more likely to report their house burglary on the same day than victims living in multi-unit housings.

Asian victims were the base case for the variable 'Victim Descent'. This coupled with the fact that all the descent factor coefficients were negative means that Asian victims were more likely to report burglaries on the first day compared to victims of other descents. The next variable that affects same day burglary reports are the descent binary variables: Other, Hispanic, White, and Black. **Table 1** displays the change in odds of the different descents as a percentage of the baseline odds, holding all other variables constant. Interestingly, the decrease in odds of a same day report between white and black people compared to Asian people was very similar. Generally, distrust of police is more often seen among people of color, so there may be another factor influencing people's decisions to report burglaries immediately or at a later date.

Descent	Odds Per 100 Asians Same Day Reports
Asian	100 Same Day Reports
Black	54.7 Same Day Reports
White	56.1 Same Day Reports
Hispanic	73.3 Same Day Reports
Other	80.0 Same Day Reports

Table 1: Table of the odds of different descents reporting a burglary as a percentage of the odds of an Asian person reporting a burglary.

The use of a weapon is a significant variable in determining whether a burglarized home reported on the same day. Victims who were burglarized by a perpetrator with a weapon are 30% more likely to report on the same day. This could be due to safety reasons. When weapons are involved, it most likely puts the victims at more risk. This danger could affect the victim to feel the need to report sooner rather than later.

The daytime variable appeared to have a somewhat smaller affect on the odds of reporting a burglary the same day when compared to other variables, and was the least significant of the variables in the model. Nonetheless, this variable was significant beyond the 99% level. Overall, the odds of someone reporting a burglary that happened during the day the same day is 7.8% higher than for a burglary that happened at night, assuming all other factors remain the same. This could potentially be due to the fact that people would be asleep at the time of the burglary if the burglary was at night and thus would not notice the crime until the next day when they woke up.

The last factor in the model was the victim's sex. Interestingly, we found that the odds of a woman reporting a burglary the same day were about 11.6% higher than the odds of a man doing the same, provided other factors remained constant. This could be due to men potentially feeling embarrassed to report that someone was able to break into their house and steal things, or trying to solve the burglary themselves before deciding to report it to law enforcement.

Though all of the variables are statistically significant beyond the 99% level, our model is not a particularly good predictor of whether someone will report a burglary the same day. We found a C-statistic of .578, indicating a poor classifier as it is close to .5. Figure 2 shows the ROC curve of the model, with a fairly linear slope, further proof that this model is not a good predictor of the data.

To be able to extrapolate this model to a broader scope, we need to meet the conditions for logistic regression. Because all of the predictors in the model were binary predictors, the linearity condition is met for logistic regression. Randomness and independence, on the other hand, are not met. Burglaries are often based on location factors and the same person can commit multiple burglaries, potentially at the same house, which makes these observations possibly linked and thus not independent. This similarly makes the data potentially not random. From these conditions not being met, we must conclude that we cannot extrapolate this model beyond our dataset.

4 Discussion

Looking a bit deeper into the racial aspect of house burglaries, although Asian victims were the most likely to report burglaries on the same day, they were the group with the smallest sample size, consisting of only 8. 34% of the entire sample (**Table 3**). This is not to mention that California has a very unique and diverse ethnic demographic compared to the rest of the United States.

In our results, we discussed from **Table 2**, that single family residences were the most significant variable in determining whether a victim reports a house burglary on the same day. If we make an assumption that most people living in houses make more money than people living in multi-unit properties, then we may be able to infer that in one way, income could be a factor of whether someone reports their house burglarized on the same day.

Interestingly enough, looking at **Table 4**, the true positivity rate for a single family house to be reported on the same day is significantly greater than a multi-unit building to be reported on the same day. This means that more single family homes are being burglarized, and the percent of them reporting on the same day is also significantly higher. The difference in the same-day reporting for the building types is nearly 11%.

Not all neighborhoods are treated equally in the city of Los Angeles. The neighborhood with the most reported burglaries in Los Angeles was West LA. Moreover, according to Niche.com, the median household income in that neighborhood was roughly $127,000! Furthermore, the neighborhood with the second highest reported burglaries is Pacific Palisades with a median income of around $208,000! Although the data set does not include data on the financial status of the victims, we think this might be something worth looking into.

This data set included LAT and LON variables corresponding to the approximate latitude, and longitude of the reported burglary. In **Figure 3**, we include a map of Los Angeles, California from ZeeMaps.com. In **Figure 4**, we plotted the Latitude and Longitude of the data set, then colored each point depending on whether it was

reported on the same day. Comparing the plots, there seems to be a similarity in the outline of the shapes. The data set's plot seems to be slightly distorted due to the ranges of Lon and Lat. Overall, the plot shows that the distribution of whether a burglary reported same day does not seem to be randomly distributed. We can see that geography has an impact on whether someone reports burglaries on the same day. On the top left, it looks like most people reported same day than people on the right.

5 Appendix

Variable	Estimate	Standard Error	Statistic	P Value
(Intercept)	0.113	0.0556	2.02	0.0429
Single Family Residence (Y/N)	0.375	0.0297	12.6	1.84e-36
Black (Y/N)	-0.604	0.058	-10.4	2.14e-25
Hispanic (Y/N)	-0.311	0.0543	-5.73	9.78e-09
Other (Y/N)	-0.223	0.055	-4.05	5.06e-05
White (Y/N)	-0.577	0.0504	-11.5	2e-30
Use of a weapon (Y/N)	0.264	0.039	6.77	1.33e-11
Daytime (Y/N)	0.0752	0.0266	2.83	0.00471
Male Victim (Y/N)	-0.11	0.0264	-4.15	3.36e-05

Table 2: Table of variables selected for logistic regression using stepwise selection

Descent	Count	Percent total
White	9745	39.06%
Hispanic	4991	20.13%
Black	3555	13.98%
Asian	2072	8.34%
Other	4587	18.46%
Total	24950	100%

Table 3: Table of Ethnic Demographics

Building Type/Report Same Day	Yes	No	Total	True Positivity Rate
Single Family House	9058	8689	17747	51.04%
Multi-Unit Building	2918	4285	7203	40.51%

Table 4: Table of Same Day Reports for House Type

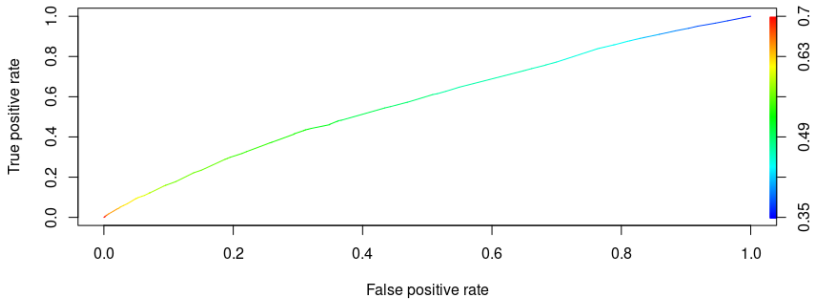

Figure 2: ROC curve of model

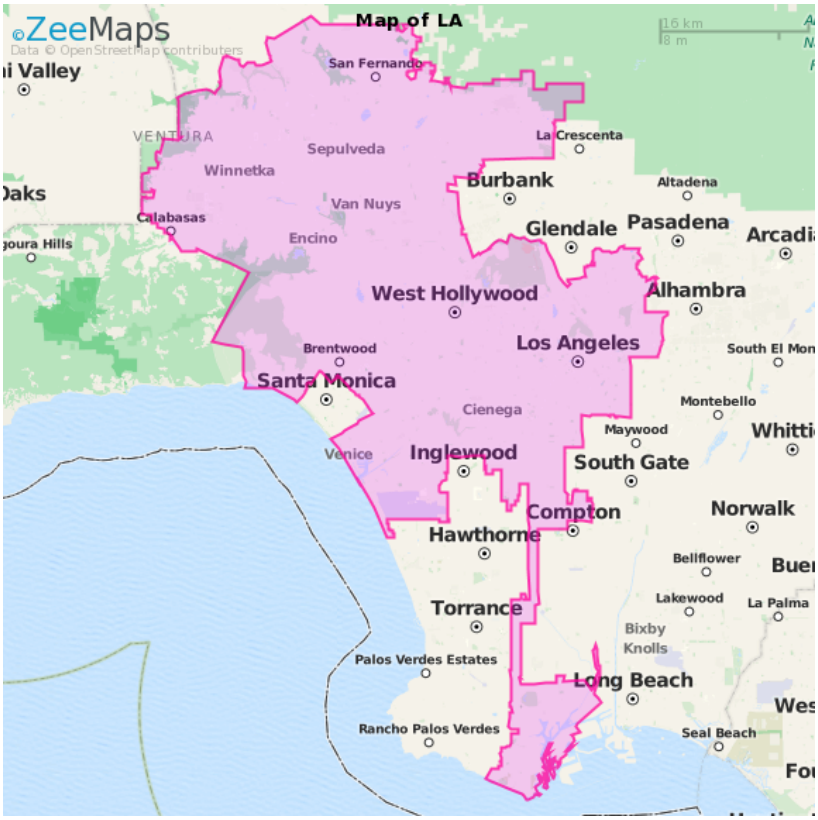

Figure 3: Map of Los Angeles, California

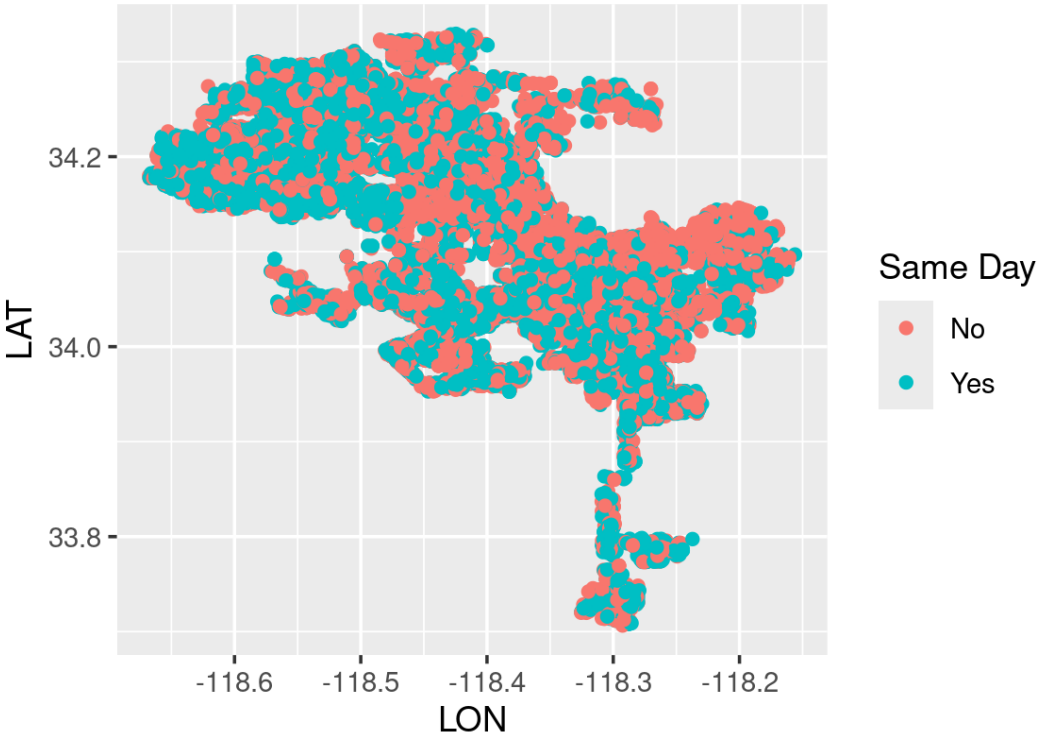

Figure 4: Map of Same Day Reports in Los Angeles, California

Figure 5: Count of property type in dataset

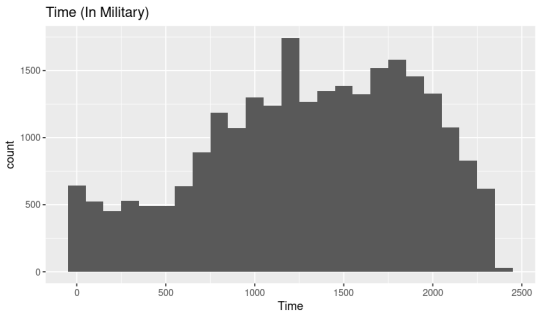

Figure 6: Histogram of time when burglaries occur

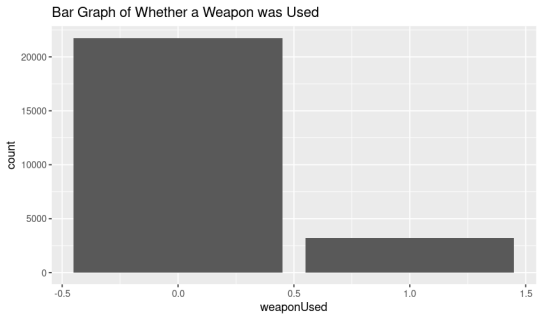

Figure 7: Bar chart of burglaries where weapons are used or not

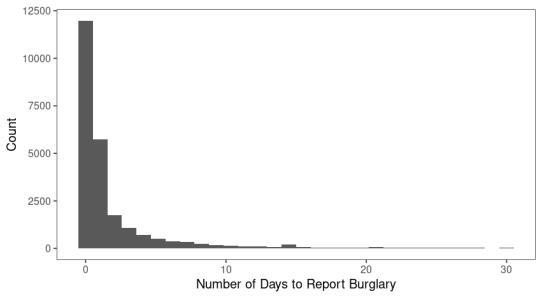

Figure 8: Histogram of days to report a burglary

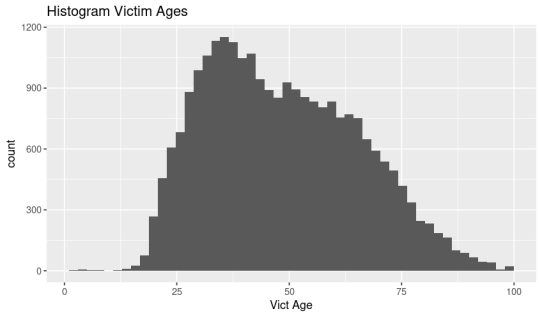

Figure 9: Histogram of victim ages

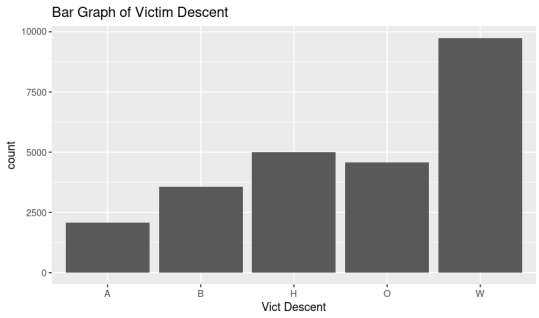

Figure 10: Bar chart of victim descents

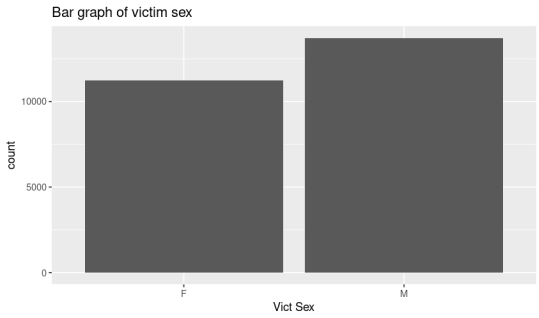

Figure 11: Bar chart of victim sex