LA Crime Report

Zachary Naumann

7/16/2021

Introduction

The purpose of this report was to determine if there was an effective way to predict the occurrence of violent crime based on various factors. The dataset used was specifically focused on crime in Los Angeles in the year 2020. The dataset, named *la_crime_stats* and containing 276,584 records, was contained in a data frame with 16 variables, with each element representing a reported crime. A preview of the dataset is shown below:

```
##        DR_NO   DATE OCC TIME OCC AREA   AREA NAME Crm Cd       Crm Cd Desc
## 1: 201226015 2020-12-08    17:00   12 77th Street    110 CRIMINAL HOMICIDE
## 2: 200504437 2020-01-10    03:30    5      Harbor    110 CRIMINAL HOMICIDE
## 3: 200704856 2020-01-19    00:15    7    Wilshire    110 CRIMINAL HOMICIDE
## 4: 201221046 2020-09-20    07:05   12 77th Street    110 CRIMINAL HOMICIDE
## 5: 201104267 2020-01-06    07:20   11   Northeast    110 CRIMINAL HOMICIDE
## 6: 201104271 2020-01-06    19:55   11   Northeast    110 CRIMINAL HOMICIDE
##    Vict Age Vict Sex Vict Descent Premis Cd                    Premis Desc
## 1:       31        M            B       101                         STREET
## 2:       44        M            H       109                PARK/PLAYGROUND
## 3:       34        F            B       735 NIGHT CLUB (OPEN EVENINGS ONLY)
## 4:       32        M            B       101                         STREET
## 5:       60        M            W       203                OTHER BUSINESS
## 6:       15        M            H       101                         STREET
##    month day     dow violent
## 1:    12   8 Tuesday       1
## 2:     1  10  Friday       1
## 3:     1  19  Sunday       1
## 4:     9  20  Sunday       1
## 5:     1   6  Monday       1
## 6:     1   6  Monday       1
```

The purpose of each variable is as follows:
1. *DR_NO* - unique identifier for the specific crime
2. *DATE OCC* - date when the crime occurred
3. *TIME OCC* - time when the crime occurred
4. *AREA* - numeric code for the police district the crime occurred in
5. *AREA NAME* - name of the police district the crime occurred in
6. *Crm Cd* - numeric code for the type of crime
7. *Crm Cd Desc* - name of the type of crime
8. *Vict Age* - age of the victim
9. *Vict Sex* - sex of the victim

10. *Vict Descent* - ethnicity of the victim
11. *Premis Cd* - numeric code for the type of location the crime occurred in
12. *Premis Desc* - name of the type of location the crime occurred in
13. *month* - month of the year, as a number (1-12)
14. *day* - day of the month, as a number (1-31)
15. *dow* - day of the week
16. *violent* - indicator used to specify whether a crime was violent or not (non-violent = 0, violent = 1)

There were a number of key steps involved in acquiring the necessary data. Preceding any data analysis, the dataset was partitioned into a training and test set. These were named *train_set* and *test_set* and contained 221,267 and 55,317 entries respectively. After partitioning the dataset, individual variables were tested to see if they had any effect on violent crime. Variables that looked to have an effect were then modeled to determine if they were valid predictors for the occurrence of violent crime.

Methods/Analysis

Dataset Modifications

Several changes were made to the original *la_crime_stats* dataset to facilitate easier data analysis. Sixteen variables were excluded because they were either redundant or referred to information that wasn't relevant to the analysis. The *DATE OCC* field was modified to exclude irrelevant timestamps, and the *TIME OCC* field was modified to show the time in a standard HH:MM format. In addition, the *month*, *day*, *dow*, and *violent* fields were added to streamline analysis.

Definition of 'Violent'

For the purpose of this study, crimes were considered violent if they involved force or threats of force. Crimes were grouped together by *Crm Cd* based on that criterion. Below is the full list of included crimes for reference:

```
##                                        Crm Cd Desc
##  1:                                CRIMINAL HOMICIDE
##  2:                            MANSLAUGHTER, NEGLIGENT
##  3:                                     RAPE, FORCIBLE
##  4:                                   RAPE, ATTEMPTED
##  5:                                           ROBBERY
##  6:                                 ATTEMPTED ROBBERY
##  7:   ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
##  8:     ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER
##  9:      CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT
## 10:            INTIMATE PARTNER - AGGRAVATED ASSAULT
## 11: SHOTS FIRED AT MOVING VEHICLE, TRAIN OR AIRCRAFT
## 12:                SHOTS FIRED AT INHABITED DWELLING
## 13:                               FALSE IMPRISONMENT
## 14:                                        LYNCHING
## 15:                            LYNCHING - ATTEMPTED
## 16:                                RESISTING ARREST
## 17:                           BATTERY ON A FIREFIGHTER
## 18:                            BATTERY POLICE (SIMPLE)
## 19:                           BATTERY - SIMPLE ASSAULT
## 20:                                     OTHER ASSAULT
## 21:                INTIMATE PARTNER - SIMPLE ASSAULT
```

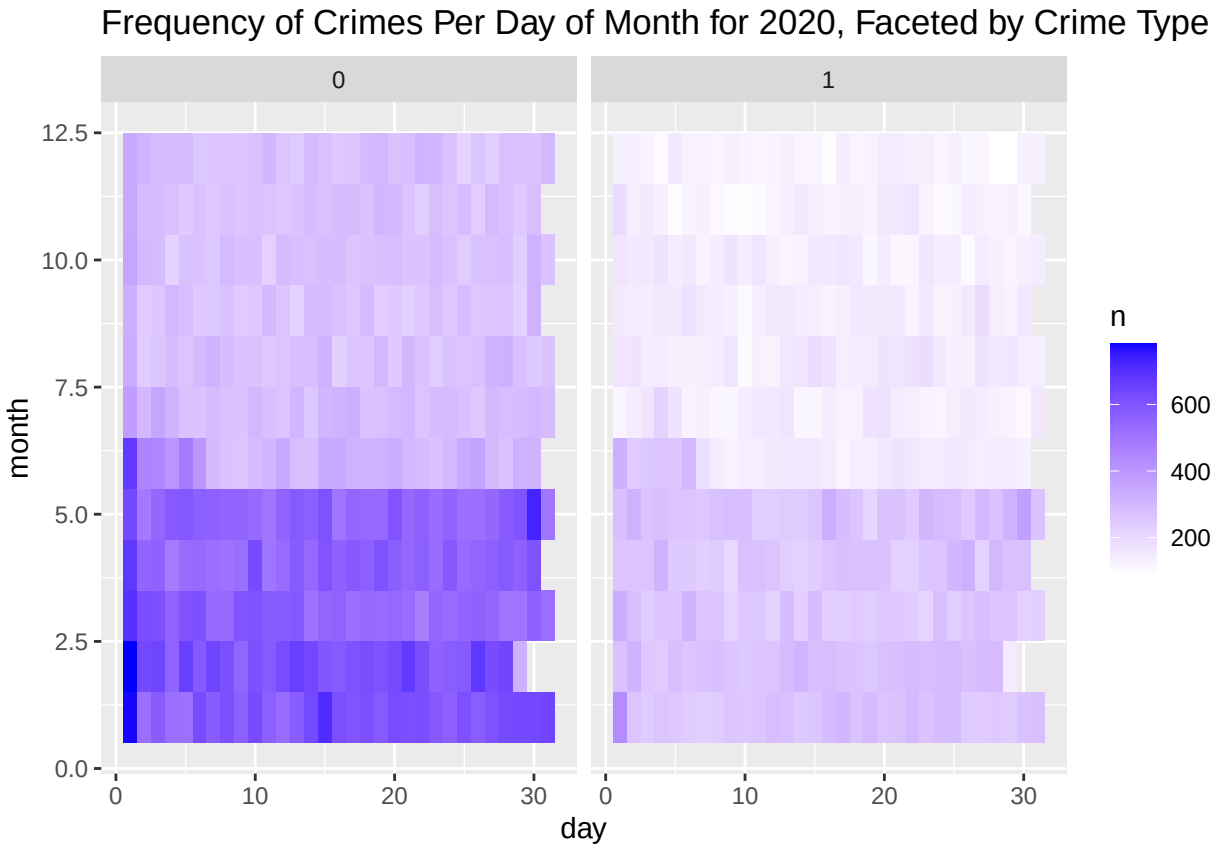
```
## 22:          CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT
## 23:                THROWING OBJECT AT MOVING VEHICLE
## 24:                   DISCHARGE FIREARMS/SHOTS FIRED
## 25:                                       BOMB SCARE
## 26:                     WEAPONS POSSESSION/BOMBING
## 27:                                 BRANDISH WEAPON
## 28:                    BATTERY WITH SEXUAL CONTACT
## 29:                                       KIDNAPPING
## 30:                     KIDNAPPING - GRAND ATTEMPT
## 31:                THREATENING PHONE CALLS/LETTERS
## 32:          CRIMINAL THREATS - NO WEAPON DISPLAYED
## 33:                           CRUELTY TO ANIMALS
##                                          Crm Cd Desc
```

Data Analysis

Overall Crime Frequency

The first thing that was checked in the data analysis process was the frequency of violent and non-violent crimes over the course of the entire year.

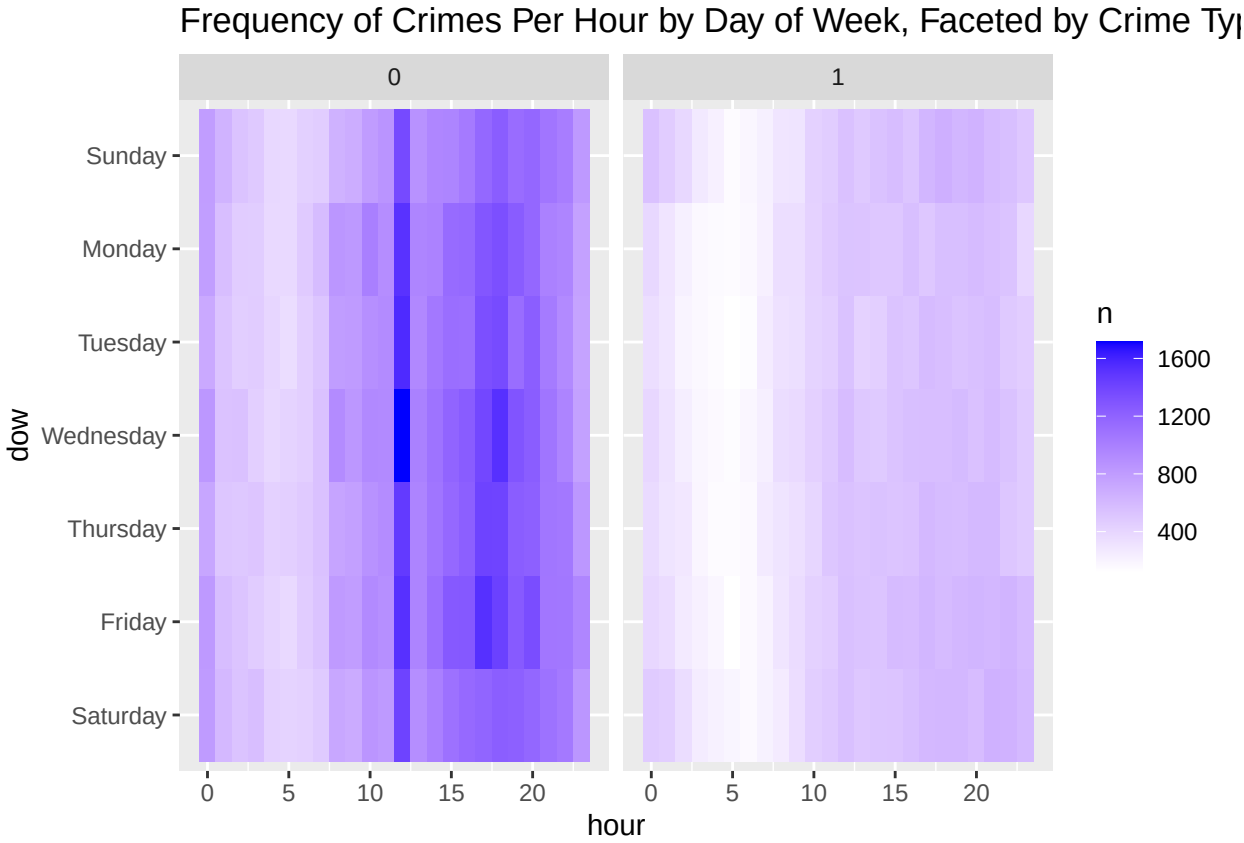
```
## `summarise()` has grouped output by 'violent', 'month'. You can override using the `.groups` argumen
```



For both violent and non-violent crimes, there was a distinct decrease in occurrences starting in June, which coincided with the implementation of lockdowns due to COVID-19.

The frequency of crimes by hour for every day of the week was also examined.

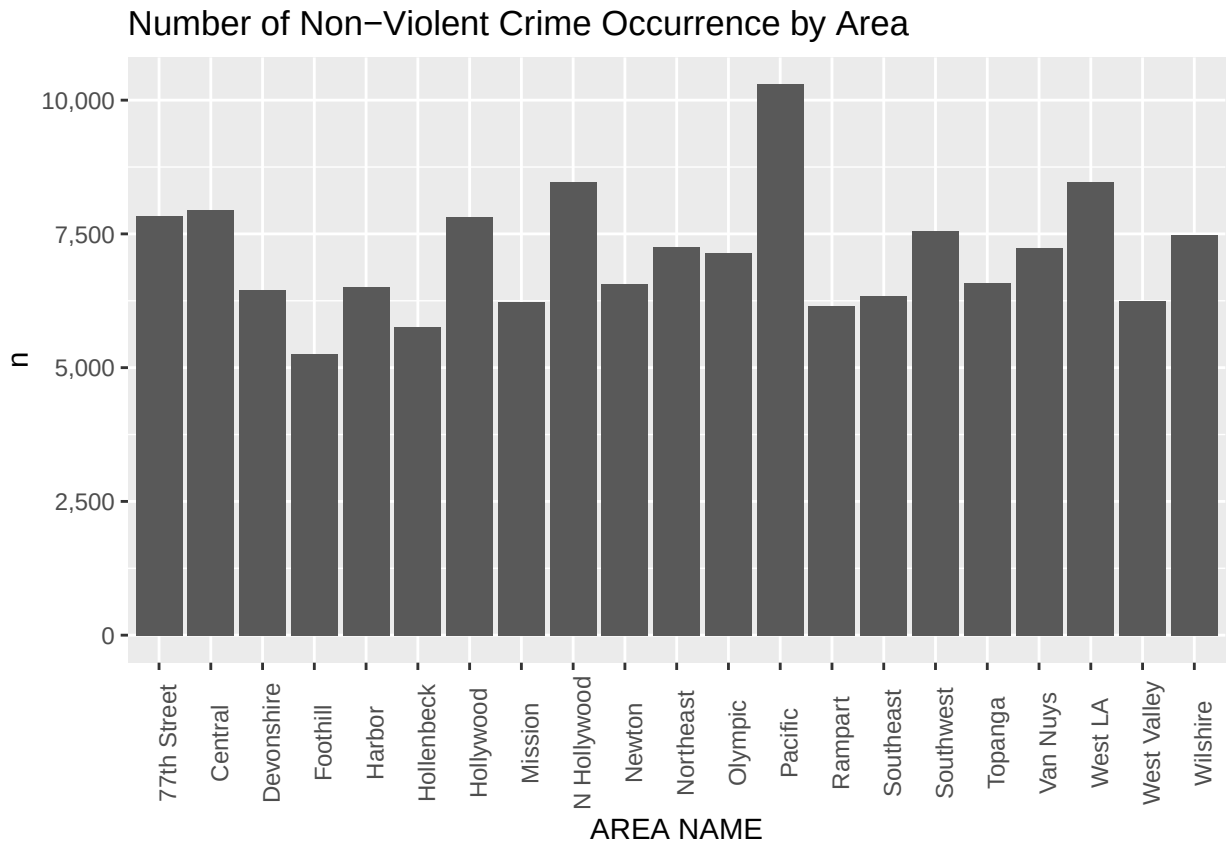
```
## `summarise()` has grouped output by 'violent', 'dow'. You can override using the `.groups` argument.
```

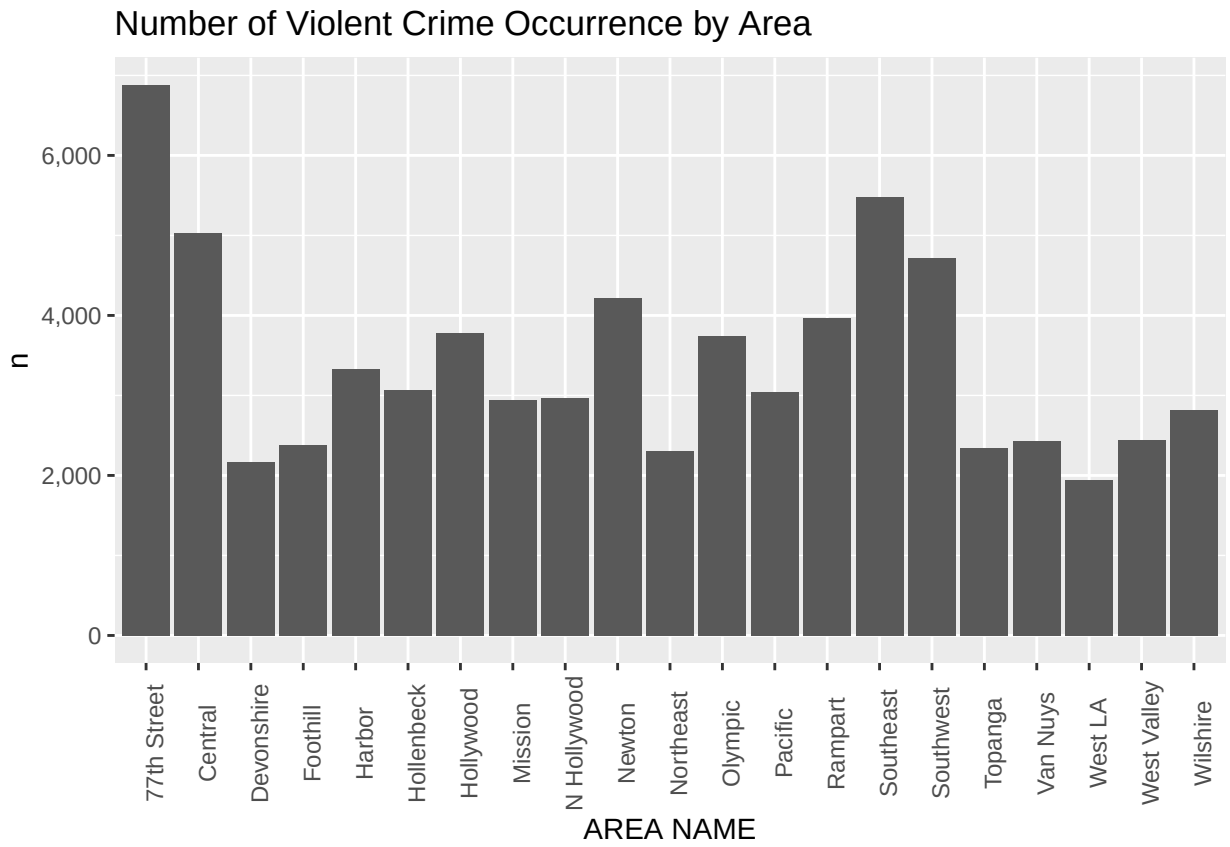


While there was some variation, every day of the week shared a trend toward increased criminal activity in the early evening and decreased criminal activity in the early morning. The type of crime did not have a noticeable effect in either case.

Crimes by Area

The next section of data checked was the relation between crime occurrence and area.

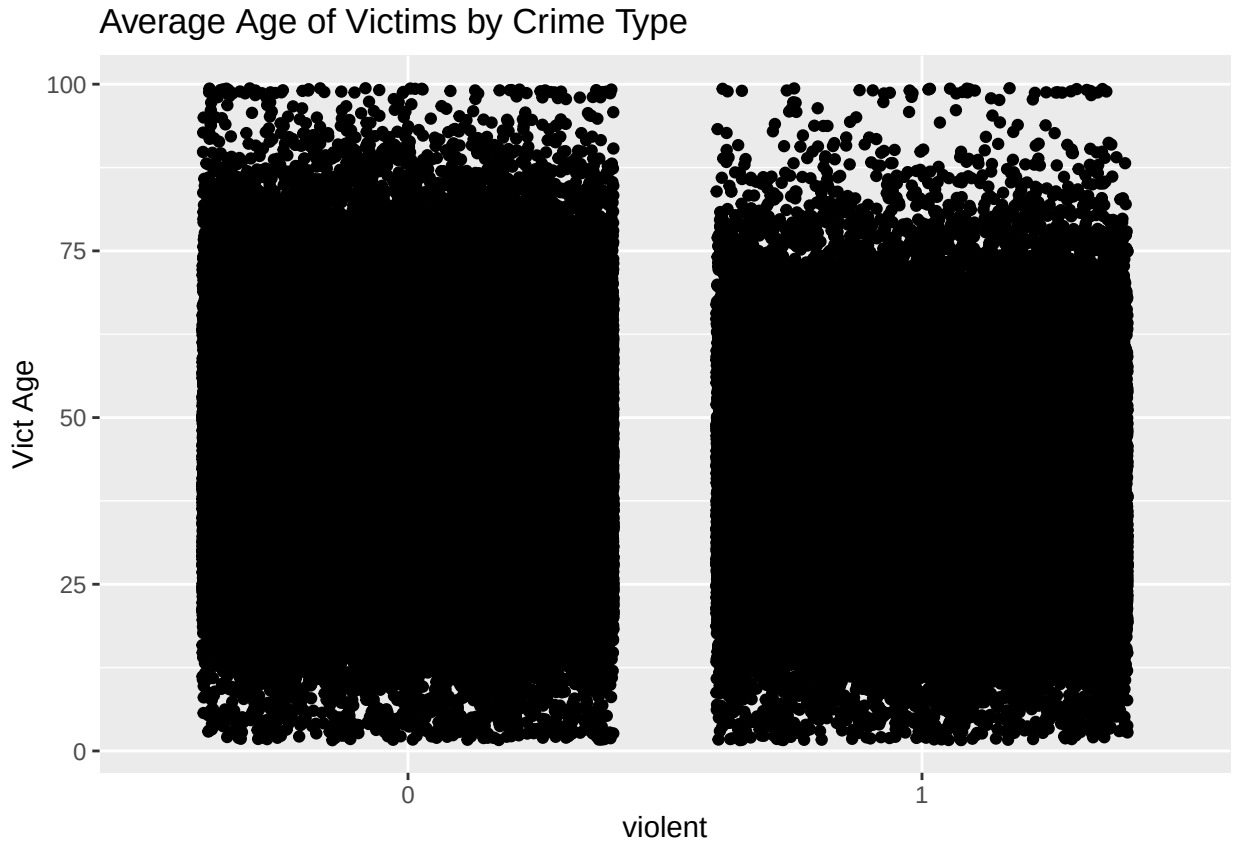


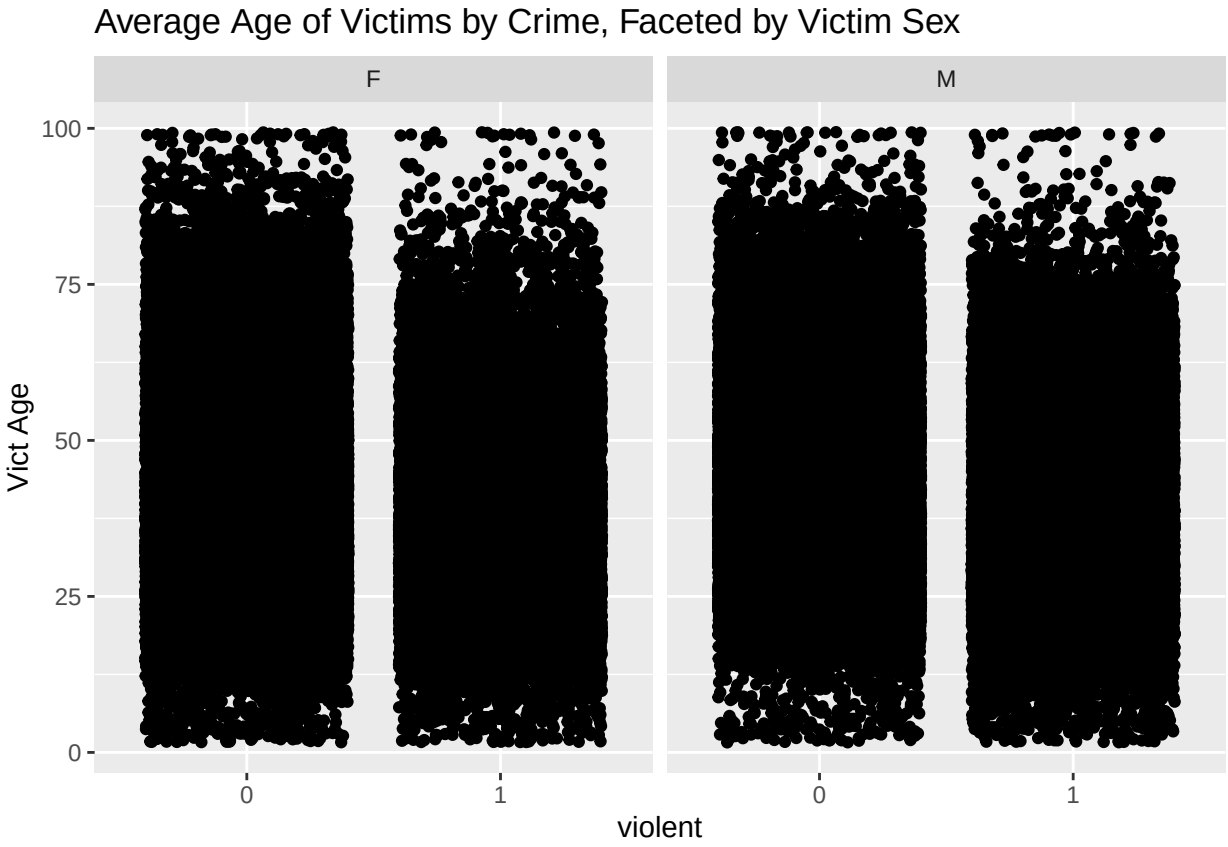


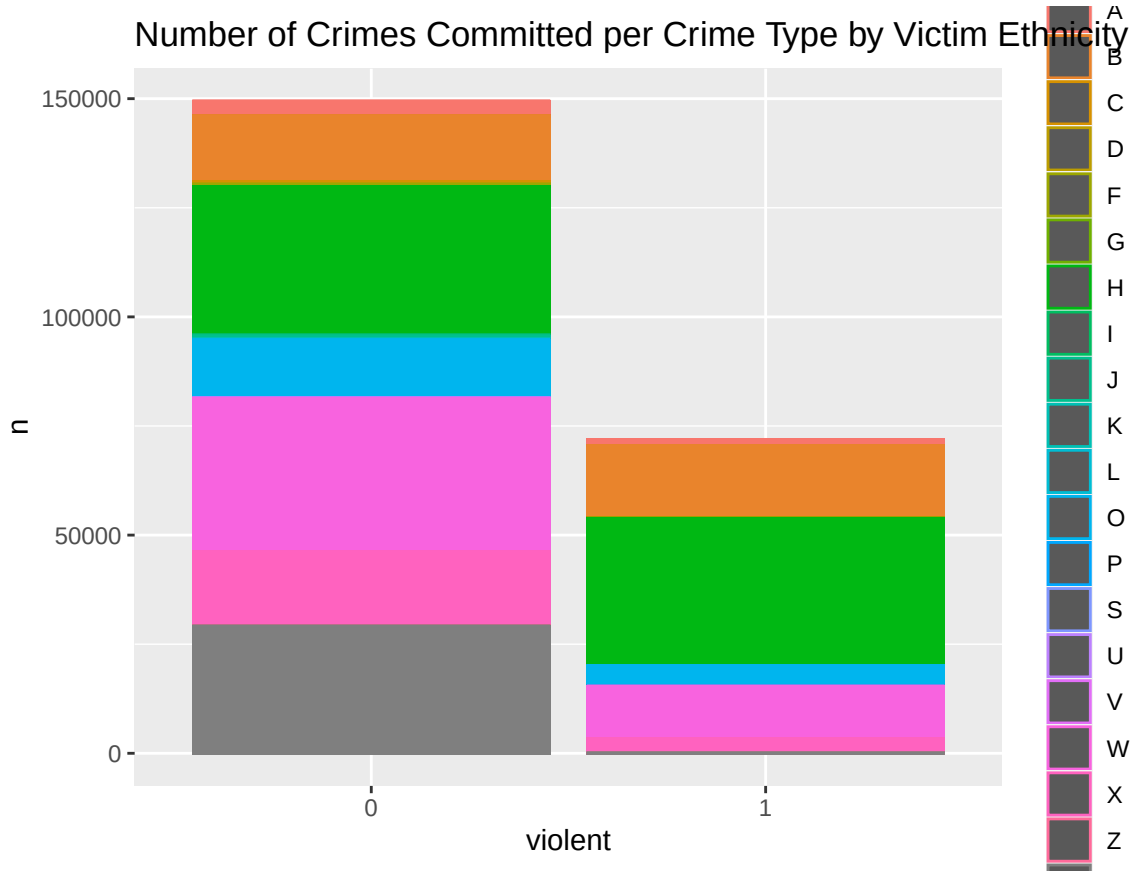
Interestingly, the areas where violent and non-violent crimes were most and least prevalent were completely different. This highlighted area as a potential predictor for the crime prediction model.

Crimes by Victimology

Victimology was also analyzed for potential relationships with crime occurrence.




Average Age of Victims by Crime, Faceted by Victim Sex
F
M
Vict Age
100
75
50
25
0
0
1
0
1
violent

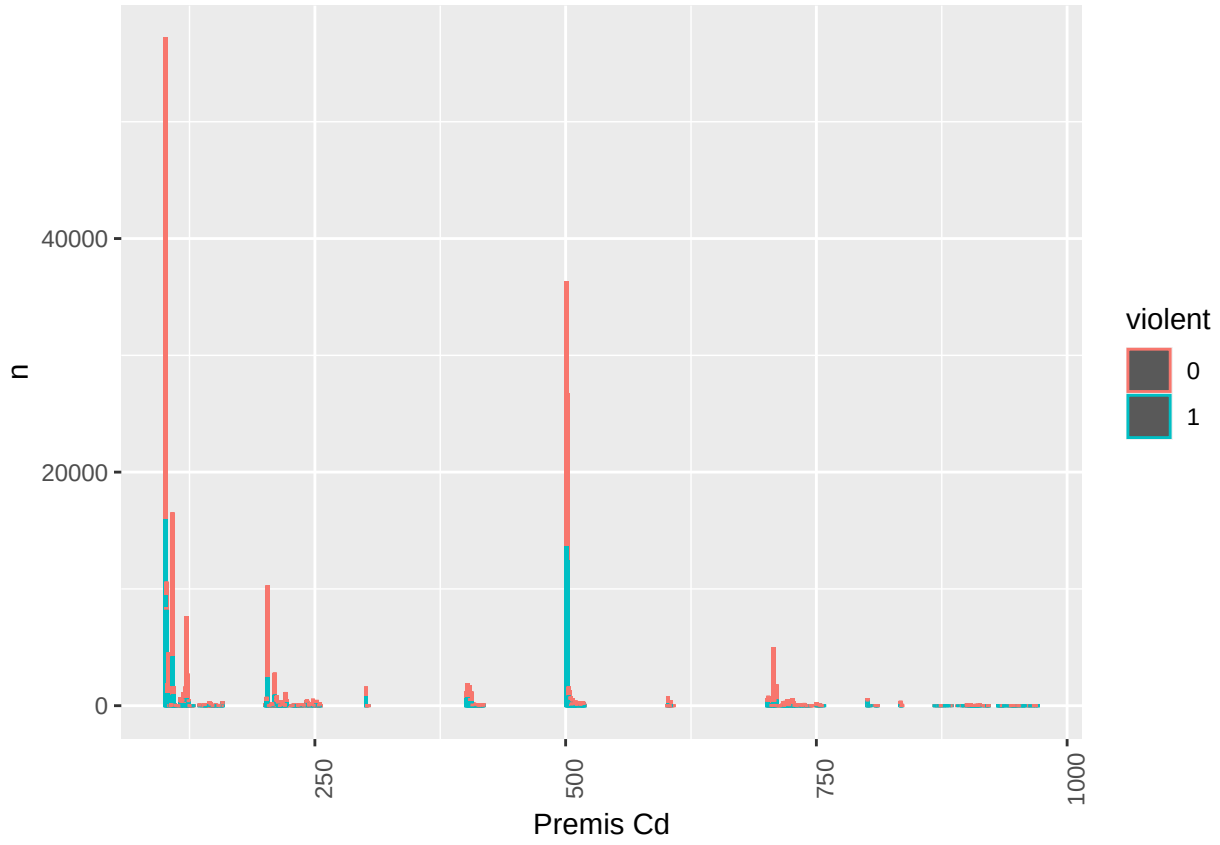


Neither victim age nor sex played a significant role in whether they were targeted for violent crime. Victim ethnicity played a role when the victims were neither Black nor Hispanic/Latino.

Crime by Location

The last potential relationship was between the crime and location type.

```
## Warning: Removed 2 rows containing missing values (position_stack).
```




```
## Selecting by n
```

```
## [1] "MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)"
## [2] "PARKING LOT"
## [3] "SIDEWALK"
## [4] "SINGLE FAMILY DWELLING"
## [5] "STREET"
```

A vast majority of both violent and non-violent crimes occurred in 5 types of locations.

Predictors and Model

Out of all of the variables analyzed, only the *AREA* field had a significant relationship with crime occurrence by crime type. Therefore, the model utilized that field in predicting the occurrence of violent crime. A basic GLM model was implemented that used the *AREA* field to predict whether a crime was violent or not.

Results

```
## Warning in set.seed(1, sample.kind = "Rounding"): non-uniform 'Rounding' sampler
## used
```

The GLM model predicted whether a crime was violent or not with an accuracy of 0.6751451. This is fairly poor accuracy, being correct only about 2/3 of the time. Adding other fields to the model either did nothing or made the accuracy worse. A large part of the problem was that while there were many minor relationships between various fields and the violent crime rate, nothing fit within a normal distribution that could be easily modeled.

Conclusion

The results of this report indicate that the model used was not that effective at predicting whether a crime was violent or not. There was only one field in the dataset that was useful in building the model, which hindered its effectiveness. There were several limitations while working on this model. The first was the effect COVID-19 had on the data. The sudden, sharp decline in crime that occurred midway through the year significantly reduced the number of potential entries that could have led to a more robust analysis. The second limitation was the fact that only a year of data was used. The source of the data actually had data going back up to a decade, but the size of the dataset would have caused problems when trying to run a model. That said, if there was ever any future work done on this topic, the dataset would probably be expanded to include several years' worth of data. Also, the model would be changed to something involving time series analysis, possibly forecasting the rate of violent crime using historical data.

Citatations

P, Sumaia. "Los Angeles Crime Data 2010-2020." Kaggle, 22 June 2021, www.kaggle.com/sumaiaparveenshupti/los-angeles-crime-data-20102020.