Exploring Child-Involved Crimes: Insights from the LAPD Crime Dataset

Jocelyn Yang, Yiming Xue, Xinyang Wang, Janel Wu, Yichen Wang

UCLA Department of Statistics

Background

- Research has shown that violence can significantly undermine children's health and their ability to grow into adults who can create sound families and communities.
- Children who have been exposed to criminal violence are more likely to display antisocial behaviors as they grow up.
- Using LAPD crime data, we focused on features ranging from time, location, and victim identities, aiming to better understand the possible factors behind child-involved crimes and predict potential solutions to reduce child-involved crime rates.

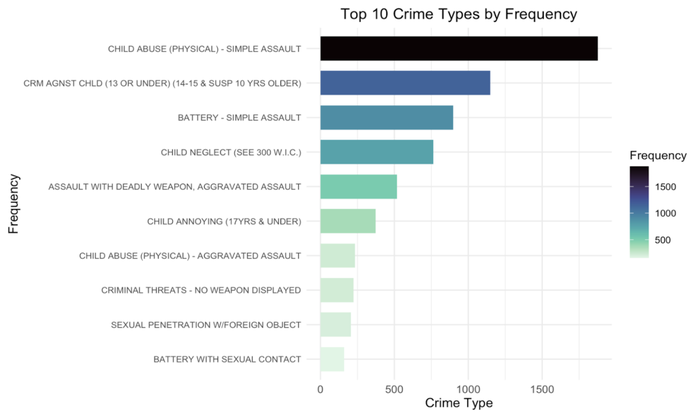
Data Overview

Scope and Size: This dataset documents crime incidents in the City of Los Angeles from January 2020 to November 2024, with 990K rows, 28 columns, and each row as a crime incident. Crime types containing the term "CHILD" result in a subset of 8,258 cases.

Data Filetering: We set the victim age range between 1 and 12. The upper limit of 12 aligns with the definition used by the National Honor Society. The lower limit of 1 was chosen based on our inference that missing age values in the dataset have been auto-imputed as 0.

Data Missingness: The missingness of the dataset contains 1 missing value in Premis.Cd (i.e., location type) and 1793 in Weapon.Desc (i.e., weapon type if used). We applied listwise deletion, excluding the respective cases with location missing values. For analyses involving the weapons, we retain these cases in the dataset. We want to investigate the pattern of missingness to determine if the absence of weapon information is systematically related to other variables in the dataset.

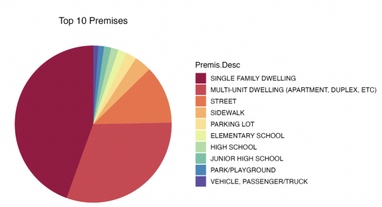
Crime Type Frequency:

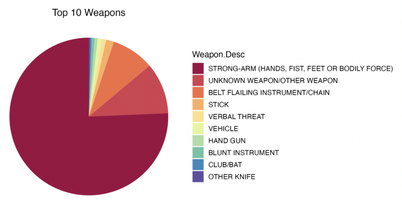


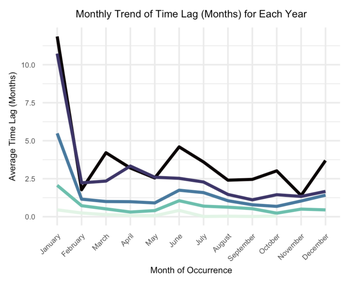
Annual Frequencies of Child-Related Crimes:

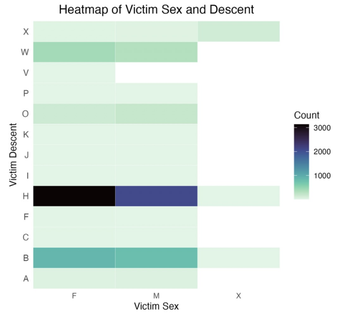
Year	2020	2021	2022	2023	2024
Cases Reported	1930	1797	2051	1890	589
Cases Occurred	1583	1727	2088	2098	761

Main Findings : EDA









Main Findings: Model Testing

1. Relationship Between Time Delay and Crime Status:

- **Aim:** To examine if there is a significant difference in arrest status (Arrested vs. Not Arrested) based on the delay between crime occurrence and report (short vs. long delays).
- **Method:** Logistic regression

```
Call:
glm(formula = Arrested ~ delay_cat, family = binomial, data = df_hyp1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.49348    0.04364 -34.222  < 2e-16 ***
delay_catshort  0.39375    0.05502   7.156  8.3e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **Conclusion:** With a p-value of 8.3e-13 and positive estimate, short delays are significantly associated with a higher likelihood of arrest. Moreover, the odds of arrest increase by approximately **48%** when the reporting delay is short.

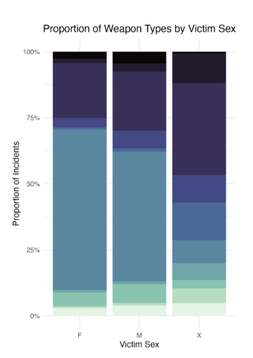
```
                Odds_Ratio      2.5 %     97.5 %
(Intercept)      0.2245896 0.2060157 0.2444596
delay_catshort   1.4825234 1.3314151 1.6519474
```

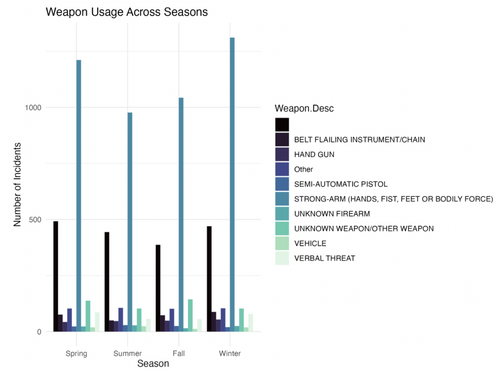
- **Implication:** Timely reporting of crimes, especially those involving children, enhances the likelihood of arrests and improves justice outcomes.

More Findings

2. Weapon Involvement by Time of Crime:

- **Aim:** Investigate whether the season (Winter, Spring, Summer, Fall) in which a crime occurs is associated with different types of weapons used in crimes involving children.
- **Method:** Chi-Square Test





- **Conclusion:**
  - A significant association was found between the sex and the type of weapon used in child-involved crimes (p-value $< 2.2e\text{-}16$).
  - A significant association was found between the season and the type of weapon used in child-involved crimes (p-value $= 0.0005792$).
  - Winter showed a higher prevalence of physical assaults, potentially due to more indoor interactions during colder months.

3. Predicting Crime Type:

- **Aim:** Predict the type of crime (Assault vs. Neglect/Abuse) based on factors like location type, time of occurrence, and victim demographics.
- **Method:**
  Random Forest Model with 300 trees and 2 variables per split.
- **Data Processing:**
  - Location: The type of location where the crime occurred, such as residential, commercial, school, vehicle, or public spaces.
  - Time: Categorized into Night, Morning, Afternoon, and Evening .
  - Crime Type: Classified into categories such as Assault and Neglect/Abuse.
- **Conclusion:** The Random Forest model achieved an accuracy of **66.75%**, given the information of location, time, and victim characteristics (gender, age, ethnicity).

Reference

• Sharkey, P. T., Tirado-Strayer, N., Papachristos, A. V., Raver, C. C. (2012). The effect of local violence on children's attention and impulse control. *American Journal of Public Health, 102*(12), 2287–2293.
• National Health Service (NHS). (n.d.). Inclusive content: Age. Retrieved December 4, 2024, from https://service-manual.nhs.uk/content/inclusive-content/age