1 Introduction

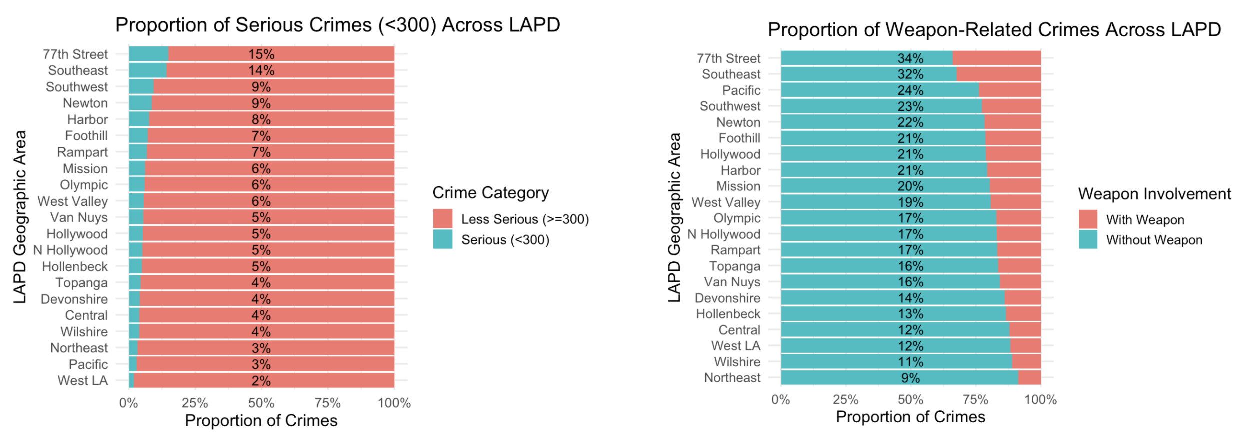
This project is a group-based study[1] conducted under **Option 1**. We apply a Bayesian hierarchical model using Markov Chain Monte Carlo (MCMC) methods to analyze an open-access dataset from the Data.gov platform, maintained by the Los Angeles Police Department (LAPD). The dataset provides detailed records of reported crimes, including the date, location, crime type, and weapon involvement.

2 Data

Our study investigates the spatial variation in crime severity and weapon-involved offenses across different areas of Los Angeles between 2024 and 2025 (up to October). In our analysis, we focus on two key indicators for each district in this time period: Serious crimes rate and the rate that a crime is committed with weapons involved.

Table 1: Definition of Selected Variables from the LAPD Crime Dataset

Variable	**Definition**
Date of Occurrence	Date when crime occurred, recorded in MM/DD/YYYY.
Police Area Code	Identifies one of the 21 LAPD Geographic Areas. Each area represents a distinct administrative region, sequentially numbered from 1 to 21.
Primary Crime Code 1	Indicates the primary and most serious crime committed in an incident.
Weapon Involved Code	Specifies the type of weapon used (guns, knives, bodily force, etc.).



(a) Proportion of Serious Crimes (<300) Across LAPD

(b) Proportion of Weapon-Related Crimes Across LAPD

Figure 1: Comparison of Serious and Weapon-Related Crime Proportions Across LAPD Areas

[1] Group Members: Yiming Gao, Yuran Wang, Molin Yang, Rongsheng Zhang

3 Model

Data and question: Let Police Area be indexed by $i = 1, \ldots, 21$. For rare, independent events occurring at a constant rate, counts with exposure can be modeled as

$$Y_i \mid \lambda_i \sim \text{Poisson}(N_i \lambda_i),$$

where Y_i is number of specific crimes observed for unit i during a fixed period. N_i is Total crimes for unit i. λ_i is the specific crime rate per exposure unit.
Hierarchical prior: To borrow strength across areas and enable shrinkage, we place a shared prior on rates:

$$\lambda_i \mid \gamma \overset{\text{iid}}{\sim} \text{Exponential}(\gamma), \qquad \gamma \sim \text{Gamma}(a, b),$$

so that the data can inform both area specific crime rates and the overall mean rate.
Posterior and computation: By Bayes' rule, the posterior is proportional to likelihood times prior. This hierarchy yields convenient Gibbs full conditionals:

$$\lambda_i \mid \gamma, \mathbf{y} \sim \text{Gamma}(y_i + 1,\ N_i + \gamma) \quad \text{independently for each } i,$$

$$\gamma \mid \boldsymbol{\lambda}, \mathbf{y} \sim \text{Gamma}\left(K + a,\ \sum_{i=1}^{K} \lambda_i + b\right).$$

4 Plan Summary

First, we tailored the original data set to obtain the subset that contains only the variables relevant to our study. Then we calculate the proportion of crimes with *Primary Crime Code 1* < 300 (values below 300 represent serious crimes) and the proportion of crimes with a non-missing *Weapon Involved Code* value (a weapon was involved in the crime).

The next step is to model the full conditionals. Posterior inference proceeds via Gibbs sampling using the conjugate full conditionals. We will implement MCMC sampling in R. Proper weakly-informative hyperpriors (e.g., $a = b = 0.001$) ensure stability and convergence. We report posterior means and credible intervals for each λ_i, estimate the overall average rate $E[1/\gamma]$, and perform posterior predictive checks and prior sensitivity analysis to assess model adequacy.

5 Reference

City of Los Angeles. (2025, November 1). Crime Data from 2020 to Present [Dataset]. Data.gov. https://catalog.data.gov/dataset/crime-data-from-2020-to-present.