# The Most Dangerous Places in Houston: Crime Analysis on Houston Divisions

Yu-Ting Yeh

## 1 Abstract

This project aims to better understand the public security in Houston, utilizing data from Houston Police Department (HPD). The findings highlight regions with crime by building poisson-Gamma model to derive criminal rate distributions, then visualize crime status between Division using heat map. Besides, to explore the situation between daytime and nighttime, regression model are employed to model the relationship between the cases and time. By offering comprehensive analysis, this project provides insights into public safety across the city of Houston.

## 2 Motivation

The motivation for this project stems from a personal desire to better understand the safety of the city I live in. Last month, I went on my first road trip to Dallas with friends, where we visited the famous Giant Eyeball tourist attraction. Finding a parking was challenging, so we parked farther away. While walking to the spot, we scared. The streets were eerily empty, except for numerous homeless individuals lying along the sidewalks. At that time, we realize the importance of prioritizing personal safety over simply visiting tourist spots. Therefore, this project becomes one of my approaches I get to know the city I live in.

## 3 Background

The Houston Police Department's (HPD) Operations division is structured to deliver comprehensive law enforcement services across the city. The division is organized into four patrol regions(North, East, South, and West). Each is managed by an assistant chief responsible for deploying patrol resources. These regions are further divided into patrol divisions, each led by a captain overseeing specific divisions.[1]

The four patrol regions are

| **Region** | **Divisions** |
|---|---|
| North | Central, North, Northwest |
| West | Midwest, South Gessener, Westside |
| South | Southwest, Clear Lake, South Central, Southeast, Traffic Enforcement |
| East | Eastside, Kingwood, Northeast, Traffic Enforcement |

Table 1: Houston Police Department Patrol Regions and Divisions

HPD supported annual report for public security in October, 2024. [2] Accord to the report, the table shows that aggravated assault dominates the violent category than robbery, sexual assult, kidnapping, murder. As for non-violent category, BMV/TMV (vehicle-related

theft) is the most frequent to Auto Theft, bulgary, theft. In general, HPD summarized that Violent crime totals for year to date October 2024 are trending 1.7% below last year. Non-Violent crime totals for year to date October 2024 are trending 6.6% below last year.

## 4 Data Sourcing & Processing

HPD updates criminal data monthly, capturing details on common criminal cases under the defined divisions A-K. Now, the date in the most updated data is 09/30/2024. After an incident is reported, officers record key details—offense type, location, and time. They then input them into a centralized system for organized storage and retrieval. To date, the data is publicly accessible on HPD website[3].

I retrieved crime data from 01/2023-09/2024. Every record means a crime happened in the past. Each record has details of time, place, beat, longitude, latitude. Detailed description of data list in table 2.

Table 2: HPD dataset description

| **Variable** | **Description** |
|---|---|
| Occurrence Date | Date offense occurred. |
| Occurrence Hour | Hour offense occurred. |
| NIBRS Description | Offense type. |
| Offense Count | Number of offenses committed. |
| Beat | Beat where the offense occurred. |
| Premise | Type of location where the offense occurred. |
| Block Range | Hundred block for address where offense occurred. |
| Street Name | Street name where offense occurred. |
| Street Type | Identifiers of street names. |
| Suffix [Streets] | Directional identifiers for streets. |

To streamline the analysis, I aggregated HPD beats (the place that crime happened) into divisions based on the definitions provided by the Houston Police Department (HPD).Figure 1. For example, beats such as "1A10" and "1A20" were grouped into the divisions "Central".

From figure 1, there are two blank regions surrounded by other regions managed by HPD. They are enclaved cities, Bellaire and West University Place.[5] These two places are independent and maintain its own municipal government and services, including police departments. Consequently, they do not fall under the Houston Police Department's authority. Out of the curiosity of enclaved city, I posed a question on reddit [6] and knew a friend working in Houston police station before. Interesting.

Then, I excluded two types of division in my project.

1. **Enclaved Cities**: Cities such as Bellaire and Memorial Village, maintain their own independent police departments, while geographically within Houston. Consequently, HPD does not provide criminal data for these areas. This is reflected as blank spaces in Figure 1.

2. **IAH and Hobby Airport**: The George Bush Intercontinental and Hobby Airports were excluded due to its minimal residential population (fewer than 100 residents).

The classification of offenses in this project follows the National Incident-Based Reporting System (NIBRS). Offenses are categorized into two groups: Group A and Group B. The details are in Table 3[4].

1. **Group A** (Serious offenses) Group A includes crimes for which law enforcement agencies are required to report all incidents occurring within their jurisdiction, irrespective of whether an arrest was made.

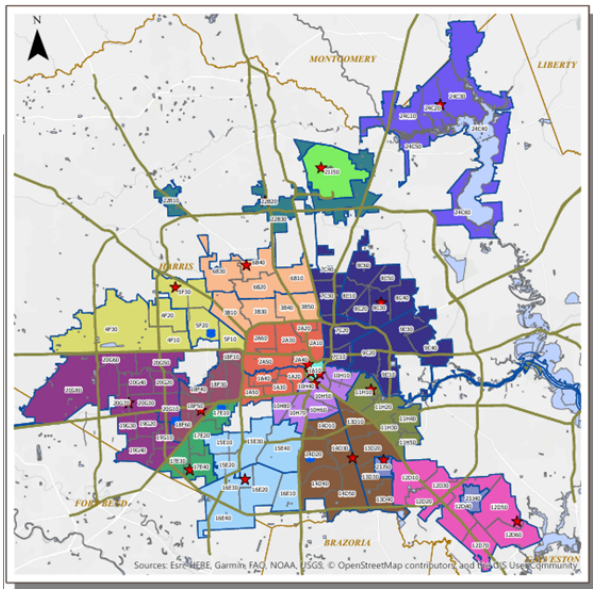



Figure 1: Houston Divisions

2. **Group B** (Less-serious offenses) Group B requires agencies to report only arrest data.

Table 3: NIBRS Offense Code

| Group | Offense Description |
|---|---|
| A | Arson |
| | Aggravated Assault |
| | Simple Assault |
| | Intimidation |
| | Bribery |
| | Burglary/Breaking & Entering |
| | Counterfeiting/Forgery |
| | Destruction/Damage/Vandalism of Property |
| | Drug/Narcotic Violations |
| | Drug Equipment Violations |
| B | Bad Checks |
| | Disorderly Conduct |
| | Driving Under the Influence |
| | Drunkenness |
| | Family Offenses, Nonviolent |

# 5 Method

## 5.1 Poisson-Gamma model

The Poisson-Gamma model is a Bayesian model commonly used for count data. In the context of crime records, each crime instance is considered a count. Then, the Poisson model is used to determine the distribution of average number of cases over a regular basis.

In this project, I decided to summarize the data on a monthly basis, which means 12 records for each month in a year. The model can be described as follows:

$$Y_i \sim \text{Poisson}(\lambda)$$

$$\lambda \sim \text{Gamma}(\alpha, \beta)$$

$$\lambda \mid Y_i \sim \text{Gamma}(\alpha + Y_i, \beta + 1)$$

The posterior distribution can be updated at a time. Summing over $Y_i$, the posterior becomes:

$$\lambda \mid Y_i \sim \text{Gamma}\left(\alpha + \sum_{i=1}^{n} Y_i, \beta + n\right)$$

### 5.2 Criminal Rate

Evaluating safety solely based on the average number of cases can introduce bias, as areas with higher populations naturally tend to report more cases. In criminal analysis, the criminal rate is a commonly used index to describe the safety of a particular area. The criminal rate scales the average number of cases by the estimated population. The formula of criminal rate is

$$\text{Criminal Rate} = \left(\frac{\text{Number of Reported Crimes}}{\text{Total Population}}\right) \times 100,000$$

The scaled average number of cases is defined as:

$$\frac{\lambda_i}{c_i}, \text{ where i denotes each region.}$$

where $\lambda_i$ is the average number of cases, and $c_i$ is the estimated population in the area.

Although the distribution of $\lambda_i/c_i$ is not Gamma distribution, I can sample from the Gamma distribution for $\lambda_i$ and then transform the samples to estimate the mean and credential interval of criminal rate. Alternatively, I can take advantage of the normal approximation, given that $n$ (the sample size) is large.

Using the normal approximation, the scaled criminal rate $\lambda_i/c_i$ follows approximately:

$$\frac{\lambda_i}{c_i} \sim \text{Normal}\left(\mu = \frac{\alpha_i}{c_i \cdot \beta_i}, \sigma^2 = \frac{\alpha_i}{(c_i \cdot \beta_i)^2}\right)$$

### 5.3 Linear Regression model

Using linear regression model with time predictor allows us to investigates the differences in crime situation between daytime and nighttime. The model in bayesian framework is specified as follows:

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

where:

- $Y_i$ is the observed crime rate for the $i$th observation.
- $T_i$ is a binary indicator variable:

$$T_i = \begin{cases} 0, & \text{if } 6:00 \leq \text{time} < 18:00 \text{ (daytime)}, \\ 1, & \text{if } 18:00 \leq \text{time} < 6:00 \text{ (nighttime)}. \end{cases}$$

- $\beta_0 \sim \mathcal{N}(\mu_{\beta_0}, \sigma^2_{\beta_1})$ represents the average crime cases during the daytime ($T_i = 0$).
- $\beta_1 \sim \mathcal{N}(\mu_{\beta_1}, \sigma^2_{\beta_0})$ captures the difference in crime rates between nighttime and daytime.
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2 \sim \text{Inverse-Gamma}(a, b)$.

The posterior distribution for the parameters $\beta_0$, $\beta_1$, and $\sigma^2$ in the Bayesian linear regression model is given by Bayes' theorem:

### 5.4 Markov chain Monte Carlo (MCMC)

In the setting of linear regression models, MCMC facilitates the approximation of posterior distributions for model parameters by sampling from full conditional posterior distributions.

Our goal is to derive the Gibbs sampling updates for the parameters $\beta_0$, $\beta_1$, and $\sigma^2$. The likelihood and prior distributions lead to the following conditional posterior distributions:

The conditional posterior distribution of $\beta_0$ is:

$$\beta_0 \mid \beta_1, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mu_{\beta_0}, \sigma^2_{\beta_0}),$$

where:

$$\mu_{\beta_0} = \frac{\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_1 T_i) + \frac{1}{\sigma^2_{\beta_0}}\mu^{\text{prior}}_{\beta_0}}{\frac{n}{\sigma^2} + \frac{1}{\sigma^2_{\beta_0}}}, \sigma^2_{\beta_0} = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma^2_{\beta_0}}\right)^{-1}.$$

The conditional posterior distribution of $\beta_1$ is:

$$\beta_1 \mid \beta_0, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mu_{\beta_1}, \sigma^2_{\beta_1}),$$

where:

$$\mu_{\beta_1} = \frac{\frac{1}{\sigma^2}\sum_{i=1}^{n} T_i(Y_i - \beta_0) + \frac{1}{\sigma^2_{\beta_1}}\mu^{\text{prior}}_{\beta_1}}{\frac{1}{\sigma^2}\sum_{i=1}^{n} T_i^2 + \frac{1}{\sigma^2_{\beta_1}}}, \sigma^2_{\beta_1} = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} T_i^2 + \frac{1}{\sigma^2_{\beta_1}}\right)^{-1}.$$

The conditional posterior distribution of $\sigma^2$ is:

$$\sigma^2 \mid \beta_0, \beta_1, \mathbf{Y} \sim \text{Inverse-Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 T_i)^2\right),$$

## 6 Result

### 6.1 The type and number of Crime

From 01/2023 to 09/2024, there are 21 months in total. Table 4 and Table 5 present the distribution of cases across all data points. Over the two years, the overall number of cases remains relatively stable. However, February consistently exhibits fewer reported cases in the year. I think it is because there are less days in February (28 days) as the average daily cases in February 2023 is higher than the average in December 2023.

When examining the data by groups, it is evident that in any given month, Group A consistently reports more cases than Group B. This suggests that more serious offenses are reported at a higher frequency compared to less serious ones.

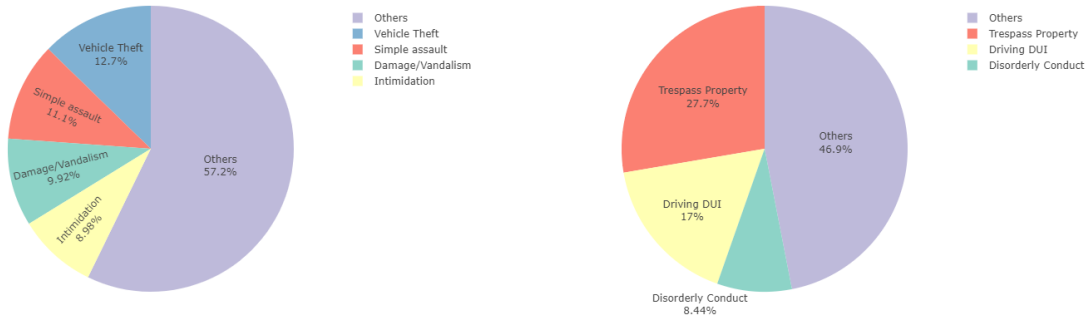
Table 4: 2023 criminal data

| **Group** | **01** | **02** | **03** | **04** | **05** | **06** | **07** | **08** | **09** | **10** | **11** | **12** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group A | 20303 | 18343 | 20514 | 20203 | 21129 | 20139 | 20618 | 20157 | 19578 | 19403 | 18924 | 19147 |
| Group B | 17651 | 15779 | 17727 | 17378 | 18285 | 17345 | 17849 | 17445 | 16929 | 16797 | 16494 | 16640 |
| All | 37954 | 34122 | 38241 | 37581 | 39414 | 37484 | 38467 | 37602 | 36507 | 36200 | 35418 | 35787 |

In Houston, there are 56 types of crimes recorded by HPD. Whenever police recorded a crime in their system, the case would be labeled. From 2023 to 2024, I drew two pie charts of Group A and Group B to display the distribution of crime types. Based on the 2, the most common specified crime in Group A is "Vehicle Theft" (12.7%), followed by "Simple Assault" (11.1%), "Damage/Vandalism" (9.92%), and "Intimidation" (8.8%). In group B, "Trespass Property" (27.7%) constitutes a highest proportion (46.9%) in specified crimes, followed by "Trespass Property" (27.7%), "Driving DUI" (17%) and "Disorderly Conduct" (8.44%).

Table 5: 2024 criminal data

| Group | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 |
|---|---|---|---|---|---|---|---|---|---|
| Group A | 18858 | 18119 | 19354 | 20025 | 20704 | 19788 | 19296 | 19310 | 18009 |
| Group B | 16274 | 15524 | 16541 | 17354 | 17957 | 16954 | 16641 | 16539 | 15502 |
| All | 35132 | 33643 | 35895 | 37379 | 38661 | 36742 | 35937 | 35849 | 33511 |




(a) The types of crime in Group A

(b) The types of crime in Group B

Figure 2: Comparison of the types of crime in Groups A and B

## 6.2 Crime Situation Comparison between Divisions

Before employing poisson-gamma models, I retrieve the parameters of prior distribution from the historic data. I used the monthly averaged number of crime cases from 2021 to 2022 as $\lambda_{prior}$.

After applying poisson-gamma models on overall group (including group A and group B), I got posterior distributions for each divisions $\lambda_i$. I then sampled from posterior distributions and applied transformation to derive the monthly criminal rate distribution. Before scaling by population, the mean number of criminal cases has broader range from 1,000 to 7,000. After scaling, the criminal rate has narrower range from 1000 to 2000, reflecting a more consistent comparison across the divisions.

Table 6: Poisson-Gamma model table

| divisions | Mean | Mean Credible Interval | Criminal Rate | Criminal Rate Credible Interval |
|---|---|---|---|---|
| Central | 4490.94 | (4456.19, 4525.95) | 2205.52 | (2188.45, 2222.71) |
| Clear Lake | 1633.75 | (1611.16, 1656.20) | 1051.80 | (1037.26, 1066.26) |
| Eastside | 1098.38 | (1080.77, 1117.59) | 1203.41 | (1184.12, 1224.46) |
| North | 3538.17 | (3503.84, 3572.55) | 1367.10 | (1353.84, 1380.39) |
| Northeast | 4093.01 | (4056.99, 4129.25) | 2011.80 | (1994.10, 2029.62) |
| Northwest | 2132.63 | (2106.69, 2158.40) | 1359.04 | (1342.51, 1375.46) |
| South Central | 2248.31 | (2220.38, 2275.68) | 2339.36 | (2310.30, 2367.84) |
| South Gessener | 2083.33 | (2058.33, 2109.79) | 1352.43 | (1336.20, 1369.60) |
| Southeast | 2975.93 | (2944.96, 3007.77) | 1681.97 | (1664.47, 1699.97) |
| Southwest | 2554.49 | (2530.05, 2582.77) | 1174.79 | (1163.54, 1187.79) |
| Westside | 7391.35 | (7342.39, 7441.44) | 2168.46 | (2154.10, 2183.15) |

By drawing the posterior distributions of ciminal rate each divisions 3, it is evident that Houston' criminal rates could be classfied into three class. divisions with higher criminal rates ($\geq 2000$) include South Central, Central, Westside, and Northeast. The divisions with an average criminal rate ($1500 - 2000$) is Southeast. divisions with lower criminal rates($\leq$ 1500)are Clear Lake, Eastside, Northwest, South Gessner, and Southwest.

The heatmap 4 reveals that the highest crime rates are concentrated in the central and western parts of Houston, followed by the northern areas. In contrast, the southern regions of Houston appear relatively safer.

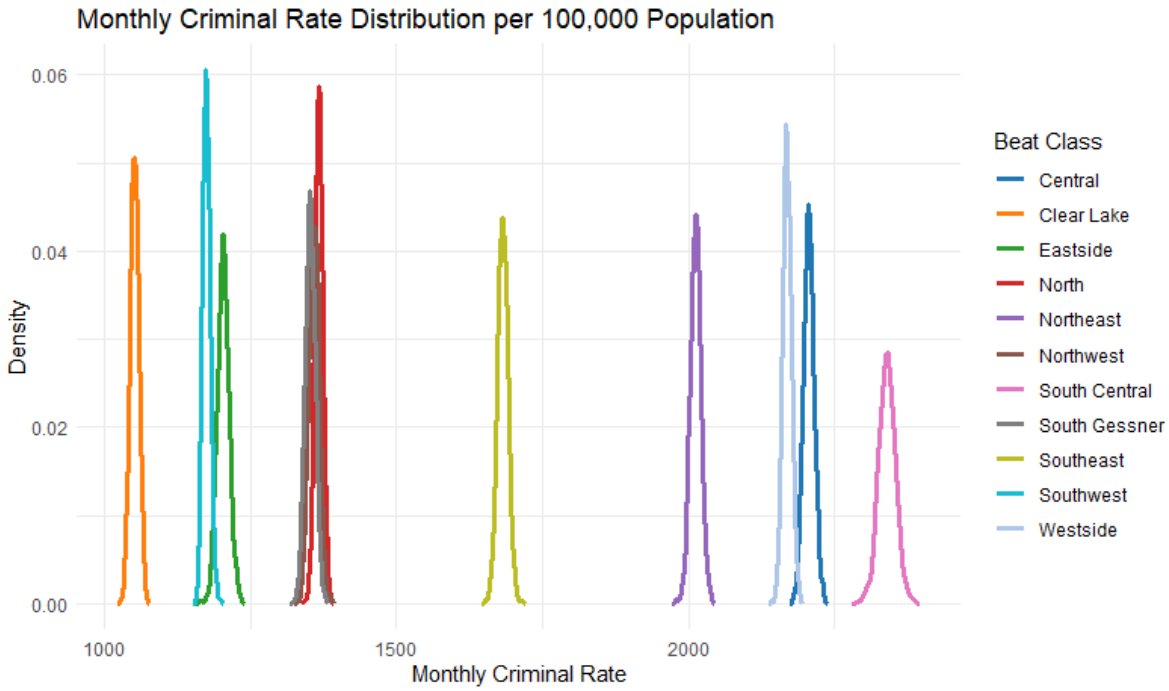



Figure 3: Posterior distributions of criminal rates

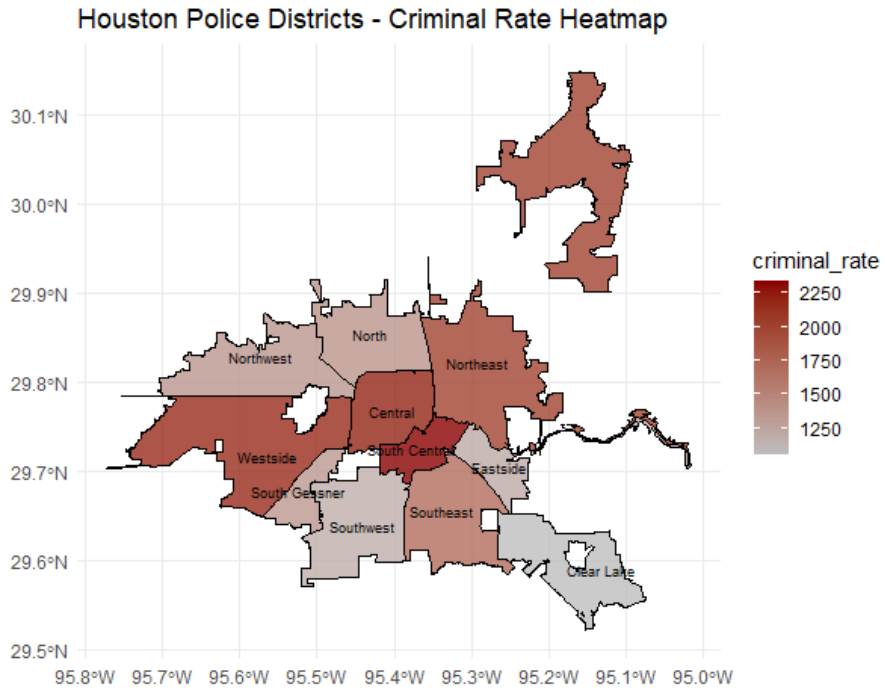



Figure 4: Houston's Heatmap

Then, I applied the pipeline to group A and group B data again, and generated Table 7 and Table 8. The dataset consists of Group A and Group B. The variation in criminal rates between divisions is significantly larger in Group A compared to Group B. The range of criminal rate in Group A is 1000 to 2000. Specifically, the range of criminal rates in Group A spans from 1,000 to 2,000, whereas the range for Group B is below 500. This indicates that the overall variation in divisions' results primarily originates from Group A.

Table 7: Poisson-Gamma model table for Group A

| **Beat Class** | **Mean** | **Mean Credible Interval** | **Criminal Rate** | **Criminal Rate Credible Interval** |
|---|---|---|---|---|
| Central | 3852.33 | (3818.58, 3887.34) | 1891.89 | (1875.32, 1909.09) |
| Clear Lake | 1434.53 | (1413.76, 1454.43) | 923.55 | (910.18, 936.36) |
| Eastside | 898.62 | (881.94, 915.86) | 984.56 | (966.27, 1003.44) |
| North | 3000.06 | (2969.50, 3028.93) | 1159.18 | (1147.38, 1170.34) |
| Northeast | 3466.84 | (3433.73, 3500.90) | 1704.02 | (1687.75, 1720.77) |
| Northwest | 1879.00 | (1854.95, 1905.42) | 1197.41 | (1182.08, 1214.25) |
| South Central | 1916.75 | (1893.06, 1940.76) | 1994.37 | (1969.72, 2019.35) |
| South Gessner | 1826.23 | (1802.70, 1850.83) | 1185.52 | (1170.25, 1201.49) |
| Southeast | 2514.33 | (2486.43, 2544.12) | 1421.08 | (1405.31, 1437.92) |
| Southwest | 2281.93 | (2257.38, 2310.38) | 1049.44 | (1038.15, 1062.52) |
| Westside | 6497.09 | (6452.42, 6541.74) | 1906.10 | (1893.00, 1919.20) |

Table 8: Poisson-Gamma model table for Group B

| **Beat Class** | **Mean** | **Mean Credible Interval** | **Criminal Rate** | **Criminal Rate Credible Interval** |
|---|---|---|---|---|
| Central | 654.91 | (640.57, 669.27) | 321.63 | (314.58, 328.68) |
| Clear Lake | 215.18 | (206.81, 223.17) | 138.54 | (133.14, 143.67) |
| Eastside | 215.69 | (207.57, 223.69) | 236.32 | (227.42, 245.08) |
| North | 555.58 | (542.59, 569.19) | 214.67 | (209.65, 219.93) |
| Northeast | 641.67 | (627.85, 655.23) | 315.39 | (308.60, 322.06) |
| Northwest | 269.71 | (260.75, 279.27) | 171.88 | (166.17, 177.97) |
| South Central | 347.64 | (337.08, 358.37) | 361.72 | (350.73, 372.88) |
| South Gessner | 273.26 | (264.43, 283.44) | 177.39 | (171.66, 184.00) |
| Southeast | 477.09 | (464.46, 489.35) | 269.65 | (262.51, 276.58) |
| Southwest | 289.36 | (279.61, 298.47) | 133.07 | (128.59, 137.26) |
| Westside | 909.10 | (892.25, 924.76) | 266.71 | (261.77, 271.30) |

.

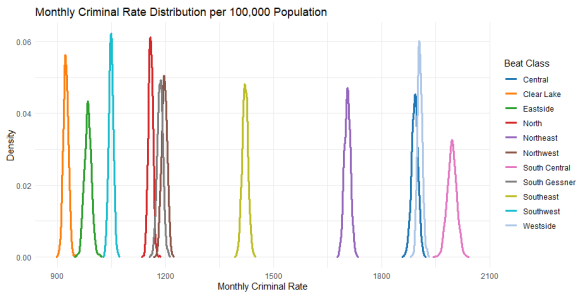
### 6.3 Crime Situation Comparison between Day and Night

This analysis focuses on crime cases in Group A, specifically exploring the differences in crime patterns between day and night. For the choice of priors, the parameters $\mu_{\beta_0}, \sigma^2_{\beta_0}, \mu_{\beta_1}, \sigma^2_{\beta_1}, \alpha_{\sigma^2}, \beta_{\sigma^2}$ were derived from the 2021 2022 crime data as follows:

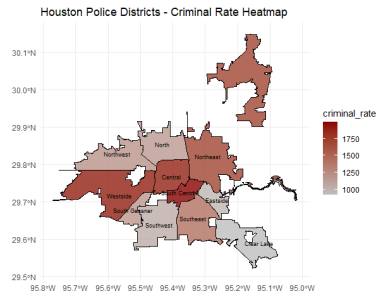
$$
\begin{aligned}
&\mu_{\beta_0} = \text{monthly averaged cases in day},\\
&\mu_{\beta_1} = \text{monthly averaged cases at night} - \text{monthly averaged cases in day},\\
&\sigma_{\beta_1} = \sigma_{\beta_0} = 100,\\
&\alpha_{\sigma^2} = 2, \beta_{\sigma^2} = \frac{\text{Var}(\textbf{monthly averaged cases})}{2}.
\end{aligned}
$$

Using Gibbs sampling, I obtained the posterior distributions of $\beta_1$ for each division. These results are summarized in Table 9 and visualized in Figure 6.

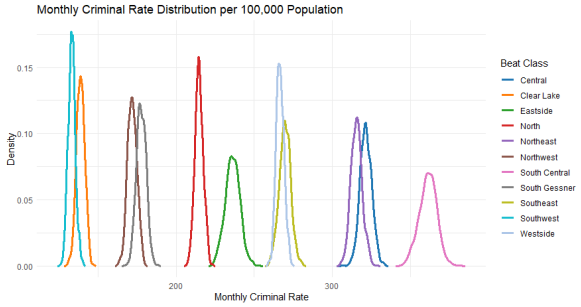
From the posterior distributions of $\beta_1$ shown in Figure 6, it is evident that there is no consistent pattern in crime rates between day and night across all divisions. $\beta_1 = 0$ indicates no significant difference in crime situation between day and night. $\beta_1 > 0$ suggest that crime



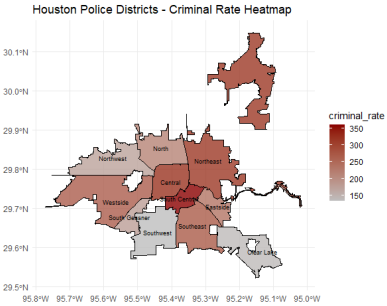

(a) Group A Posterior Distributions




(b) Group A Heatmap




(c) Group B Posterior Distributions




(d) Group B Heatmap

Figure 5: Comparison of Posterior Distributions and Heatmaps for Group A and Group B

situation are worse at night compared to the day. $\beta_1 < 0$ indicates that crime situation are worse during the day compared to the night.

- **Neutral Divisions:** Divisions like Northwest and Southeast have $\beta_1$ distributions centered around zero.
- **Night-Dominant Divisions:** Divisions such as Central and South Central show dangerous nighttime with positive$\beta_1$ values.
- **Day-Dominant Divisions:** Divisions such as Westside and Northeast exhibit dangerous daytime with negative $\beta_1$ values.

# 7 Conclusion

This project provides a overview of crime situation in Houston. In EDA part, there are 30,000 crimes happened monthly in Houston. 20,000 comes from serious offense and 10,000 comes from less serious offense. Among all, Vehicle Theft and assaults are most common cases in Houston. This result is consistent to the report from HPD. [2]

People generally want to avoid the crime type in serious offense (Group A). Based on the insight of my project, the most dangerous places in Houston is from South central, Central to Westside, followed by northern Houston, then southern Houston. To know the exact dangerous places, I should keep collecting the data from the enclaved city, drawing cases location on the maps to check clustering, or looking into the details in group A. Overall, the model has explained the different crime situation in Houston Divisions. In this city, criminal rates can vary significantly based on where you are, potentially by as much as a factor of two.

Regarding safety between day and night, the results are surprising—nighttime cases do not always outnumber daytime cases. Instead, the data reveals that half of divisions are more dangerous during the day, while half of divisions are riskier at night.

The Bayesian framework offers a clear and detailed method for comparing parameters, allowing differences between locations to be assessed based on the degree of overlap in their

Table 9: Table of $\beta_1$ for Different Division

| **Division** | **Mean** | **P**($\beta_1 < 0$) |
|---|---|---|
| Northwest | 9.6911 | 0.2022 |
| South Central | 89.2058 | 0 |
| Southeast | 13.2457 | 0.1583 |
| North | -76.4107 | 1 |
| Clear Lake | -69.8242 | 1 |
| South Gessner | 48.3353 | 0.0004 |
| Westside | -151.6715 | 0.9999 |
| Central | 91.7775 | 0 |
| Northeast | -104.5458 | 1 |
| Southwest | -23.9947 | 0.9185 |
| Eastside | 38.9363 | 0.0001 |

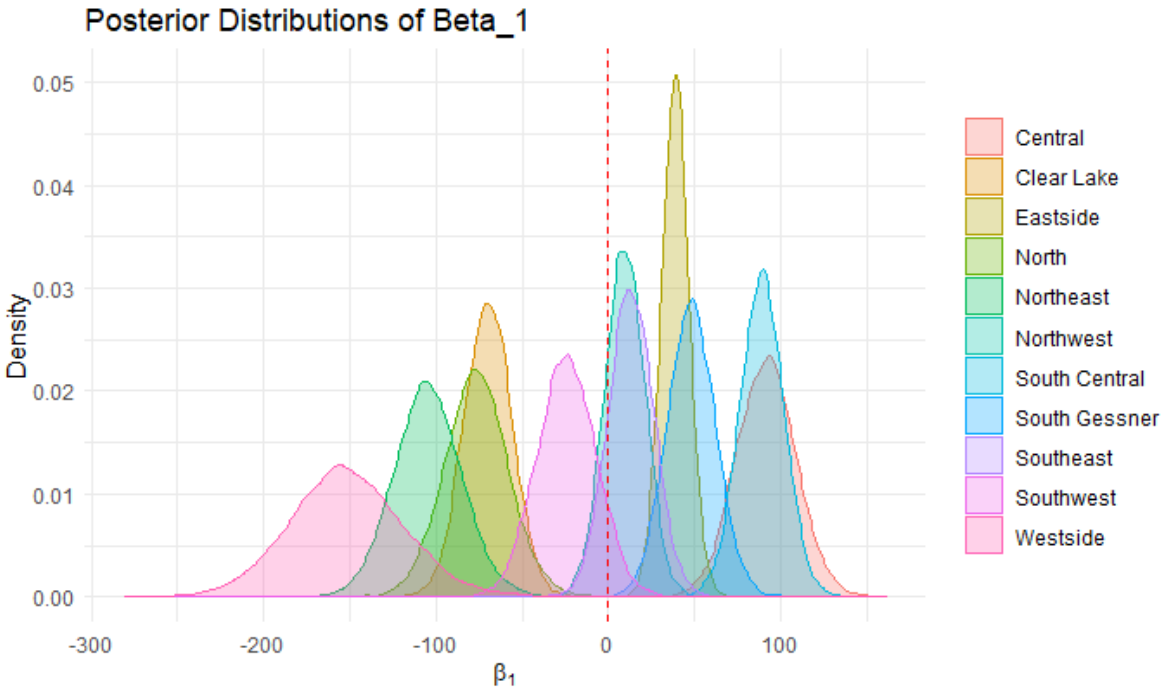



Figure 6: Posterior distributions of criminal rates

distributions. This approach eventually and effectively provides persuasive, data-driven evidence to understand Houston's crime situation.

After I search on the internet, the area near Giant eyeball in Dallas are relatively safe. But, who knows? Ignorance leads to fear. So, It is never a bad thing to get to know things around us.

## References


[1] https://www.houstontx.gov/police/organization/operations/field_operations

[2] https://www.houstontx.gov/police/department_reports/Crimeby_CouncilDistrict/Crimeby_CouncilDistrict.pdf

[3] https://www.houstontx.gov/police/cs/Monthly_Crime_Data_by_Street_and_Police_Beat.htm

[4] https://ucr.fbi.gov/nibrs/2011/resources/nibrs-offense-codes/view

[5] https://www.houstonchronicle.com/neighborhood/bellaire/article/houston-suburbs-bellaire-memorial-18642830.php?utm_source=chatgpt.com

[6] https://www.reddit.com/r/houston/comments/1gt63rb/some_places_are_not_in_police_districts/