

Introduction

Hello, my name is Jaeyoung and I'd like to introduce my teammates Jaehun and Yunseo. We don't have time so let's dive right in.

I'm an **Asian, male, 23**. I'm trying to travel to LA city.

I'm really excited but I heard that LA is different from the old days. it is getting more dangerous with lots of gangs in the city and drugs being more abundant. And you know that "gun" is legal in the states. So I was worried about what would happen to me when I traveled.

I decided to prepare for my journey by finding the **pattern** and **behavior** of crime happening in LA.

let me briefly Introduce our dataset and show how we pre-processed it.

we used various amounts of data to show what we wanted.

Our core data is crime data in La City in 2023 provided by the LAPD.

it has Two hundred thirty-two thousand rows and 31 columns of features including the district, crime type, occurrence date, and so on. We extracted features and also made new features to better understand the data.

we also used LAPD division from laxity geohub. Since the LAPD has its own unique district classification. we tried to incorporate zipcode data to easily access external data to be apply it on each district. So it was possible to use data like an American community survey based on the zipcode to access the population, median income, unemployment, etc in each district.

The last data we used was sheriff and police station data also lacity geohub to better understand the occurrence of crime on the map.

Now we are ready to engage in exploratory data analysis on crime data in La City.

The research question that we want to impose is what are the key patterns and trends in crime across the city in 2023? we tried to approach this question in a multilateral way visualizing it using r shiny.

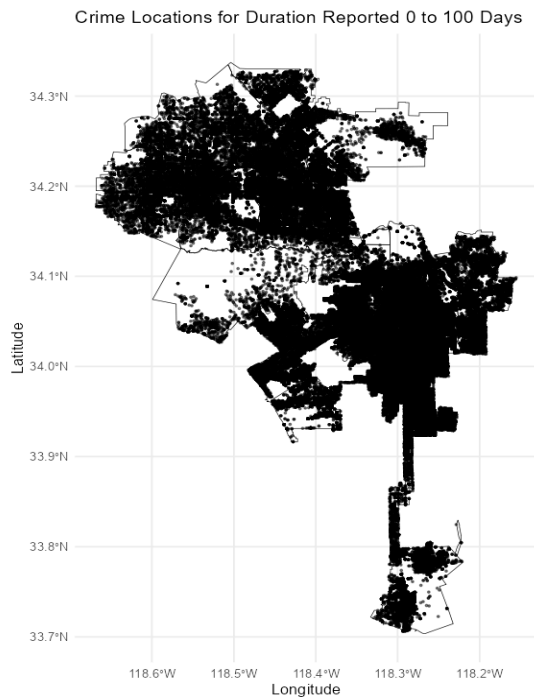
So, just before hearing our insight in data let's demonstrate how to use our r shiny. Let me introduce jaehun~~~

Key Findings

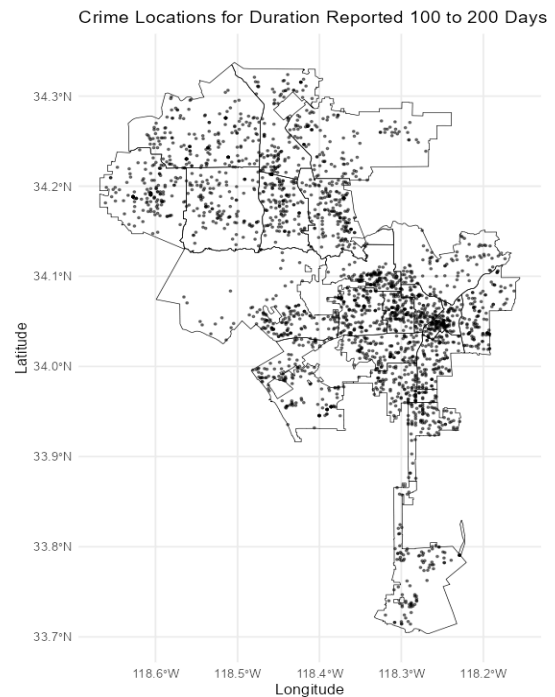
In the first part of our analysis—referring to page 1 and the second, third, and fourth options on page 2—we explored the relationship between the duration of crime reporting and the factors that influence it.

As a starting point, we formed the hypothesis: "Are there any noticeable trends when there is a significant gap between the time a crime occurs and when it is reported?" However, when we examined the plots for 0-100 days and 100-200 days, we found no distinct differences between the areas. (Refer to the map on page 1 for different durations).

Previous Plot



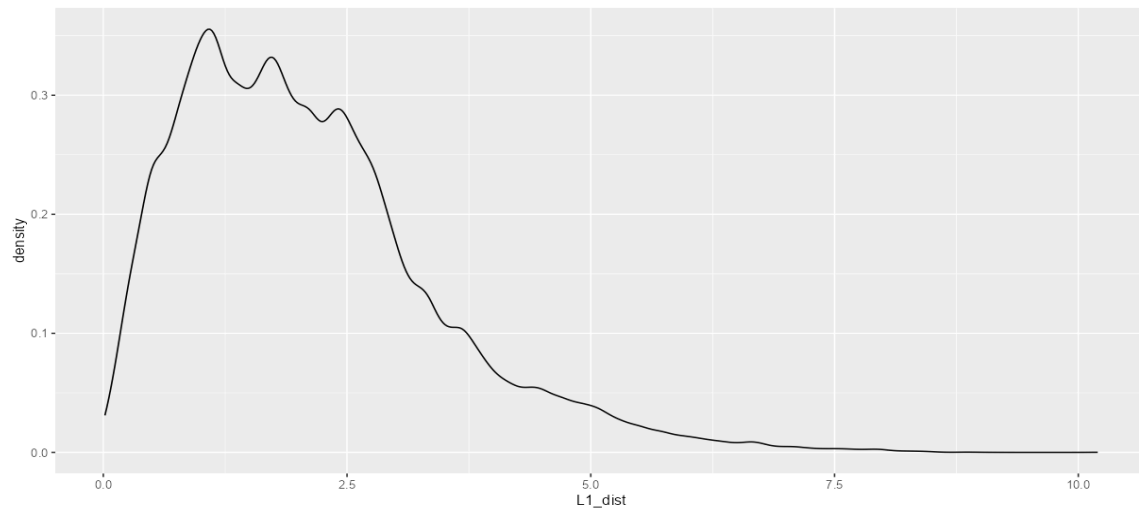
Current Plot



Next, we investigated whether the distance to the nearest police station affects reporting time. The data revealed that 75% of crimes occur within 3 kilometers of a police station. This suggests two things:

1. Police station locations are well-planned since most crimes happen within 3 kilometers.
2. The first quartile (Q1) distance exceeds 1 kilometer, indicating that police coverage is effective. Otherwise, the Q1 would be lower, implying criminals committing crimes directly near police stations. (Refer to the graph and table on page 1 for police station distances).

Crime Count in Selected Interval: 232209
Percentage of Total Crimes: 100%



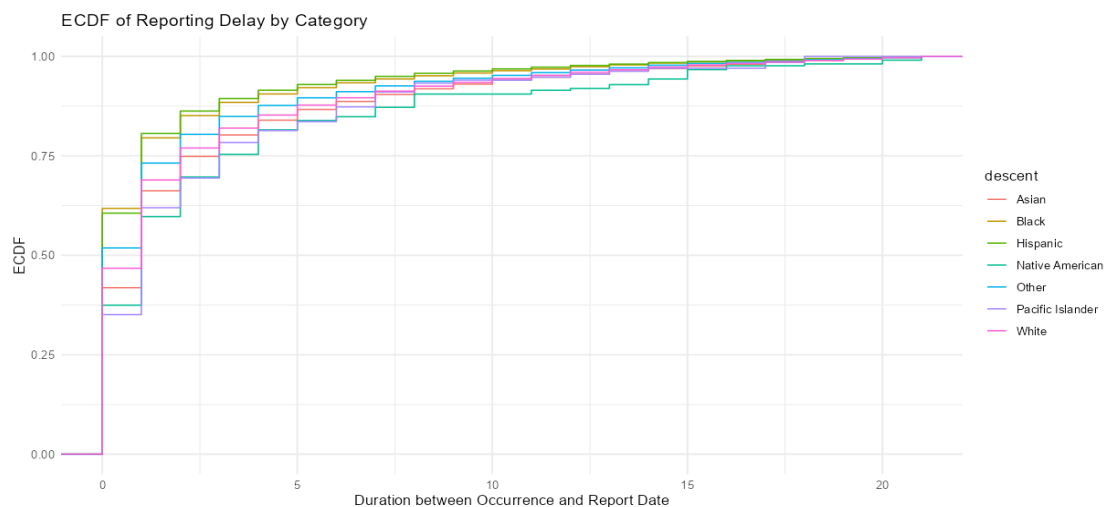
	Min	1st Quantile	Median	Mean	3rd Quantile	Max	Mode
0~6 days	0.01	1.09	1.85	2.07	2.75	10.19	1.08
7~49 days	0.01	1.18	2.01	2.20	2.91	8.84	2.41
50~100 days	0.02	1.21	2.07	2.25	2.96	8.24	3.06
101~365 days	0.02	1.22	2.04	2.26	2.96	8.39	2.46
365+ days	0.03	1.26	2.13	2.31	2.92	7.71	1.79

We then explored other factors influencing reporting duration. Examining the EDCF of victim ethnicity, we found that crimes involving Asian, Pacific Islander, and Native American victims are reported less frequently, whereas crimes against Whites, Blacks, and Hispanics are reported more quickly. (Refer to the table below option 2 on page 2).

We also checked whether crime type impacts reporting duration. The duration appears similar across different crime types, suggesting crime type is not a significant factor. Although the mean values are high, this is largely due to outliers with very high maximum durations.

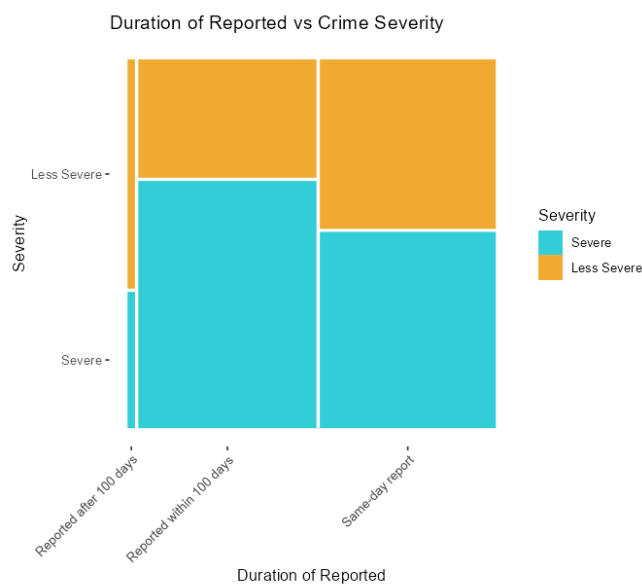
Explore Temporal and Socioeconomic Crime Patterns

Dive deeper into how crime relates to time, demographic factors, and socio-economic conditions.



category	0 day	1 day	2 day	3 day	4 day	5 day	6 day	7 day	8~49 day	50~100 day	101~365 day	365+ day
Asian	0.38	0.60	0.68	0.73	0.76	0.79	0.80	0.82	0.95	0.97	1.00	1.00
Black	0.58	0.74	0.80	0.83	0.85	0.86	0.87	0.88	0.96	0.98	1.00	1.00
Hispanic	0.57	0.76	0.81	0.84	0.86	0.88	0.89	0.90	0.97	0.98	1.00	1.00
Native American	0.33	0.52	0.61	0.66	0.71	0.73	0.74	0.76	0.94	0.97	1.00	1.00
Other	0.47	0.67	0.74	0.78	0.80	0.82	0.83	0.85	0.95	0.97	1.00	1.00
Pacific Islander	0.32	0.56	0.62	0.70	0.73	0.75	0.79	0.82	0.96	0.97	1.00	0.00
White	0.42	0.62	0.70	0.74	0.77	0.79	0.81	0.83	0.95	0.97	1.00	1.00

Another hypothesis was that severe crimes might take longer to be reported because criminals involved in these crimes are more careful to avoid getting caught. However, the mosaic plot shows that crimes reported within 100 days tend to be more severe than those reported on the same day. Interestingly, crimes reported after 100 days are less severe compared to both groups. We interpreted this to mean that crimes reported after 100 days are likely less significant and therefore delayed.

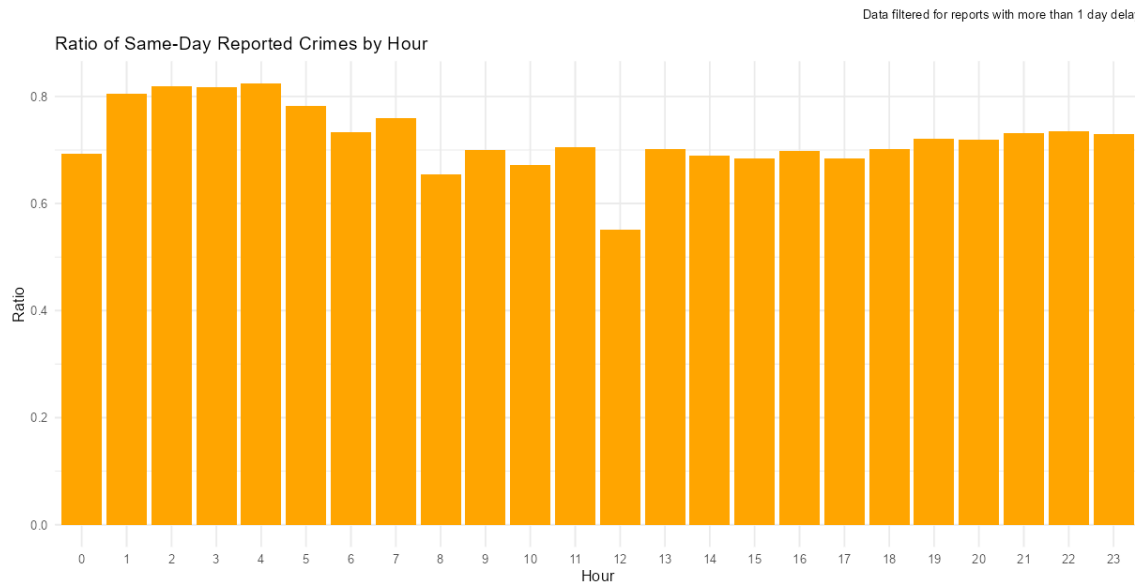
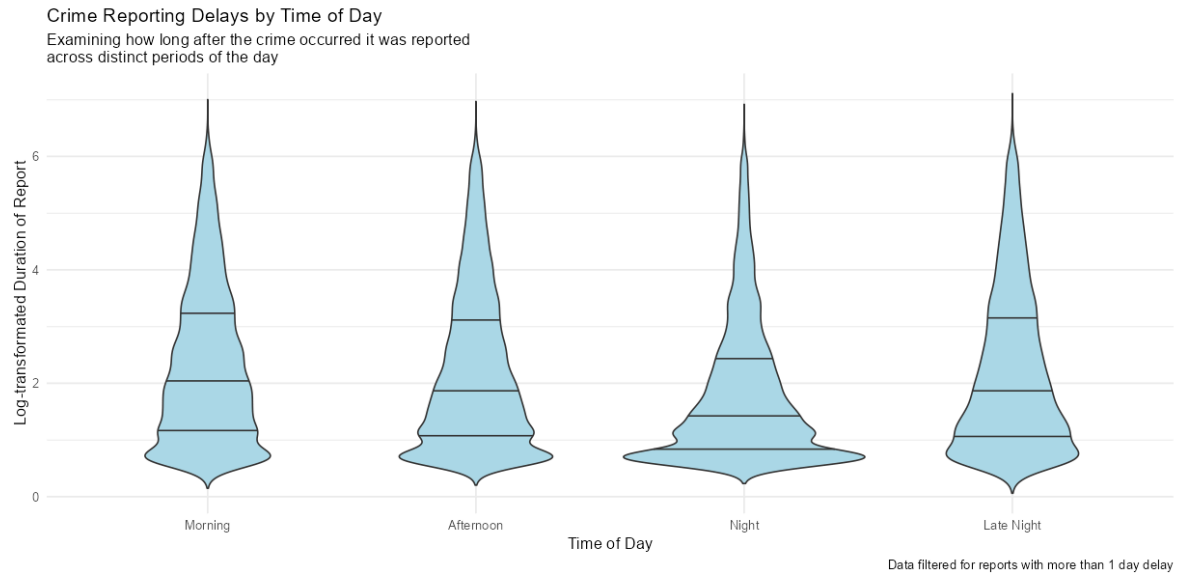


Contingency Table:

	Reported after 100 days	Reported within 100 days	Same-day report	Total
Severe	1709	77447	60744	139900
Less Severe	2875	37023	52411	92309
Total	4584	114470	113155	232209

[시간대별 보고시간 → pg2 옵션3]

Our final hypothesis about reporting duration was that crimes occurring late at night would be reported later. We filtered the data for crimes reported within one day and expected to see longer delays for late-night crimes. However, the data showed no difference in reporting times across different time periods, disproving this hypothesis.



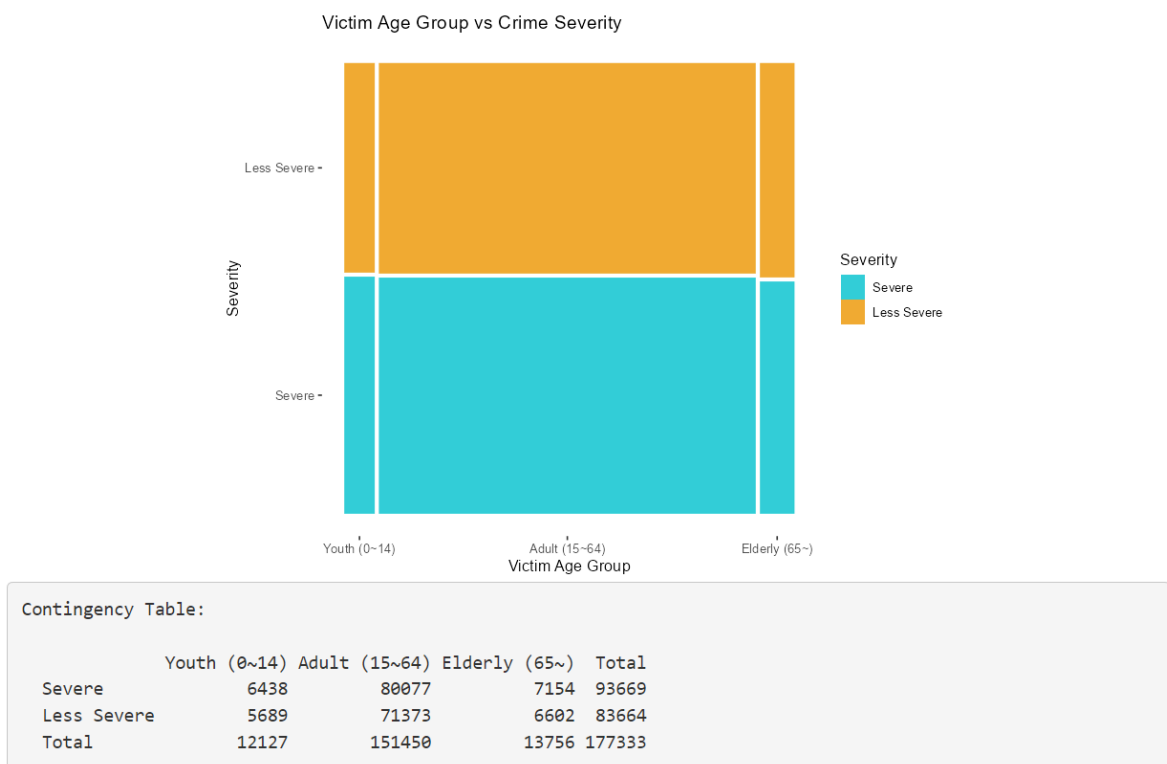
After these analyses, we shifted our perspective to explore the dataset more closely. Please refer to the remaining options on page 2 to follow this exploration.

Since crimes are categorized as severe or less severe, we checked whether the severity of a crime is linked to victim characteristics. Initially, we hypothesized that crimes involving female victims would be more severe. However, the mosaic plot revealed that males are more vulnerable to severe crimes, contradicting our assumption.

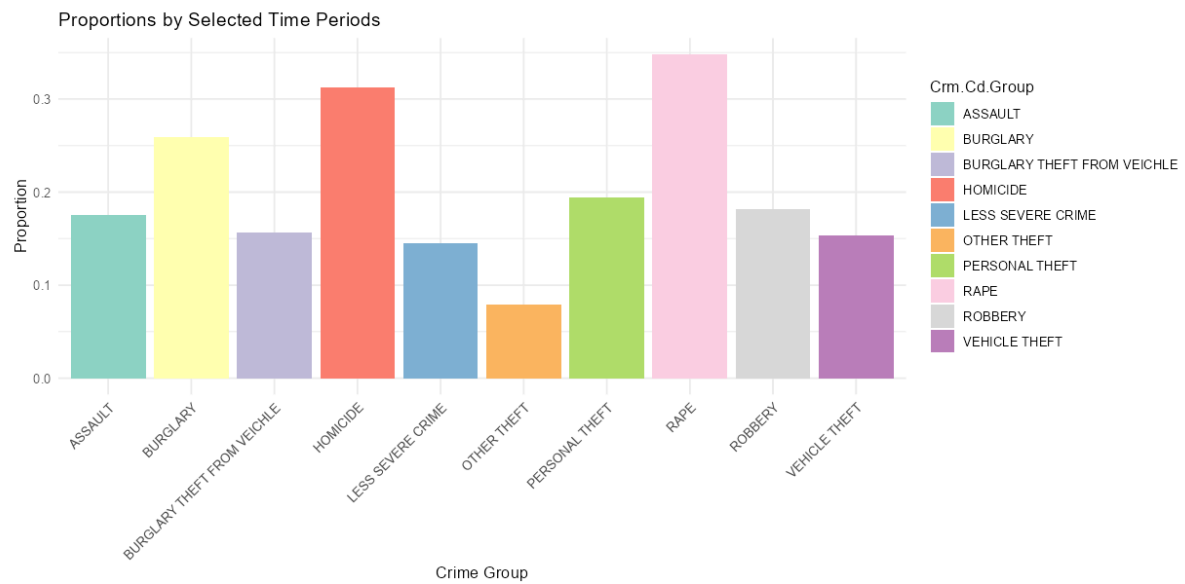


We also considered victim age as a potential factor, but the data showed no significant variation in severity across different age groups

.

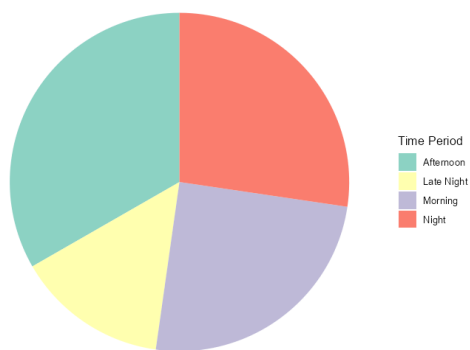


The bar graph of crime group proportions during late-night hours shows unequal distributions of different crimes.

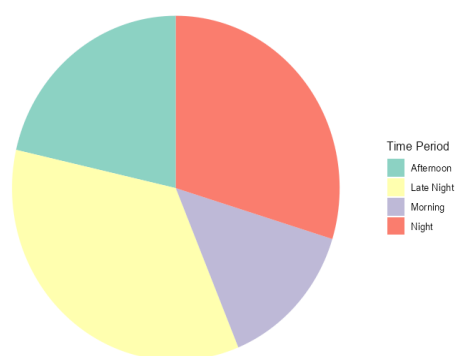


Pie charts reveal that severe crimes like rape are more common at night, while less severe crimes are more frequent in the afternoon.

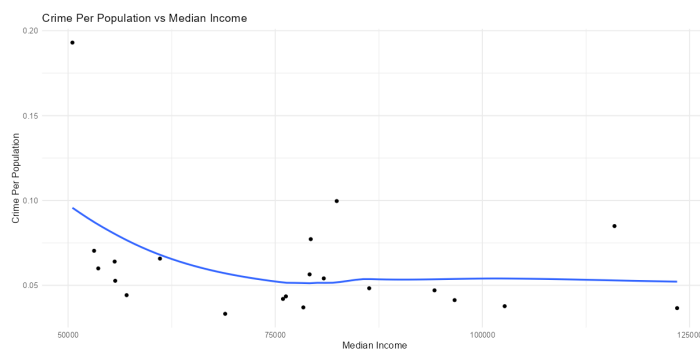
Crime Composition by Time: LESS SEVERE CRIME

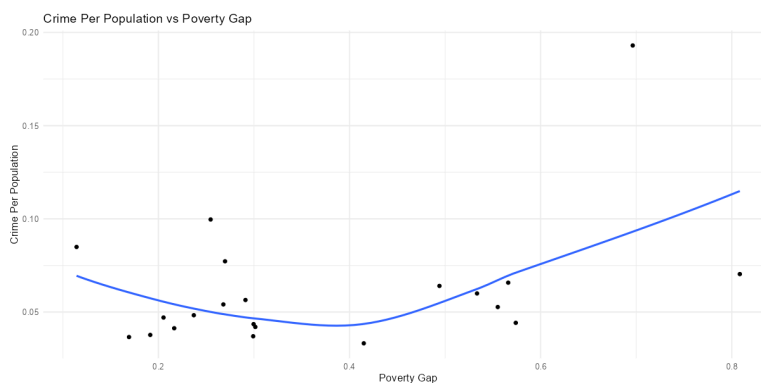
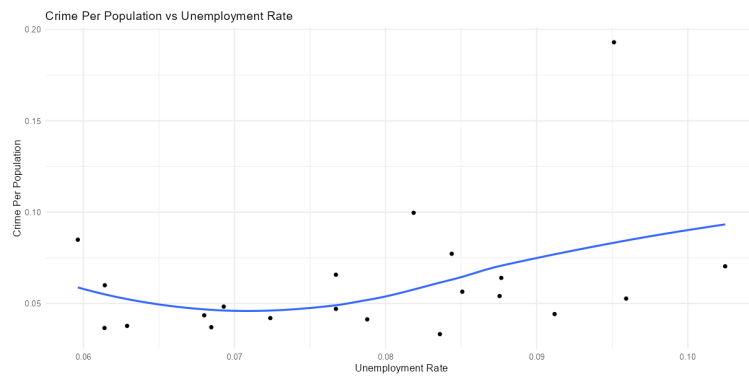


Crime Composition by Time: RAPE



Lastly, we examined whether socioeconomic factors influence crime. However, we couldn't draw a clear conclusion due to the limited number of areas analyzed and the presence of a few outlier points.





ENDING

From the first two pages of r shiny, we could find a few key variables to understand the pattern and trend of crime in la city. We thought that using these variables to predict the likelihood of crime happening might be an interesting way to interact with data. so we used a few models such as multinomial logistic regression, randomforest, and xgboost. we will not dive deep into the model but focus on the result of the model and answer the question that we started from.

Again I'm an asian male 23 who wants to travel la city. Lets take a specific arguments to predict the probability of crime happening. everyone now knows that im adult. male. asian. and i want to travel at hollywood at night. if we click the area in the map. we can find out the probability of crime happening. also we can know what type of crime we are likely to experience.

From data preparation to eda and trying a simple model it was possible for us to gain a better understanding of the crime in LA city and to answer the research question to find the pattern and trend of crime across LA city.