Homicide Increases in the COVID-19 Era

Bianca Schutz, Kaitlyn Kirt, Kevin Hu, and Marco Stine

Table of contents

1 Introduction

During the COVID-19 pandemic in 2021, homicide rates in Los Angeles increased substantially. A report by the Legal Defense Fund's Thurgood Marshall Institute found that this increase was associated with both pre-pandemic and pandemic-induced economic instability and inequalities.(Moore, Tom, and O'Neil 2022)

Using National Incident-Based Reporting System (NIBRS) data, this report will examine the impact of the pandemic and various socioeconomic factors on homicide rates in LA. We aim to compare the effect of these local characteristics on the incidence of homicide in pre-, mid-, and post-pandemic conditions.

1.1 Data Description

1.1.1 Primary Data: Los Angeles Crime Data

We will utilize NIBRS data from 2018-2024 to examine incident-level data, allowing us to fully capture the circumstances of each event. Our data is obtained from the Los Angeles Police Department (LAPD), who

regularly updates Crime Data on the Data.gov website. The data includes all crime data, which we will subset to focus on homicide, and details such as date, location, and offense type.

The codebook for this dataset can be found here: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data.

To group this data geographically, we will organize the crimes by their latitude and longitude into the appropriate Census Blocks and ZIP Code Tabulation Areas (ZCTAs). Additionally, we will examine the impact of latitude and longitude themselves, as humans do not always behave by administrative unit boundaries and latitude and longitude by themselves might suggest a different pattern.(Krieger et al. 2002)

1.1.2 Secondary Data: Geographic and Socioeconomic Data

The NIBRS data includes two main geographic identifiers: address and coordinates. Some addresses are poorly coded, such as just listing a street and no number, so we instead used latitude and longitude to examine geospatial relationships. We also used these coordinates to group our data into administrative units (Census Blocks and ZCTAs) to perform different levels of geospatial analysis.

We used shapefiles for the state of California, Los Angeles Police Department Divisions, and the Census units to categorize each crime event and to visualize the spatial distribution of the crimes through maps. To categorize the crime coordinate data, we performed a spatial join where we converted the crime coordinates to a spatial object and then joined them based on the geometries of the shape files.

We additionally used publicly available socioeconomic data from the US Census Bureau, the City of Los Angeles Housing Department (LAHD), CalMatters, and Los Angeles County Public Health. We used median household income data from the American Community Survey 5-Year estimates to get an annual measure of income in LA ZCTAs. We used evictions data from both Calmatters and LAHD to gather longer-term data and more recent data, as LAHD only started tracking evictions data in 2023, while Calmatters has tracked LA County data since 2012.

Due to difficulties in obtaining city-level data, the unemployment rate, COVID cases, and number of evictions are for Los Angeles County, while the crime data is for the City of Los Angeles. The city comprises about a third of the county's population and the greater county area largely reflects the trends in the city itself.

To maximize the amount of data available for analysis, we used LA County socioeconomic data from 2019-

1.2 Hypothesis

Based on the 2022 Thurgood Marshall Institute report, we anticipate that the homicide rates have decreased since the end of the pandemic as society has returned to a more "normal" state. We aim to see how the rate of homicide compares to overall crime levels in the city and identify any patterns between socioeconomic factors such as eviction rates and the homicide count.

Our hypothesis is two-fold; first, we predict that since the pandemic, that socioeconomic conditions have improved in Los Angeles over the past few years. Second, we predict that as these socioeconomic measures have bounced back, the rate of homicide has returned to a more "normal" rate.

Additionally, we believe that we will observe a similar relationship between homicide and the overall crime rate; as the number of crimes decreases, so will the number of homicides.

We will also investigate the relationship between homicide and enforcement factors. In particular, we will look at the LAPD budget in comparison to the homicide rate to determine if increased police funding is effective in reducing the number of homicides.

2 Exploratory Data Analysis

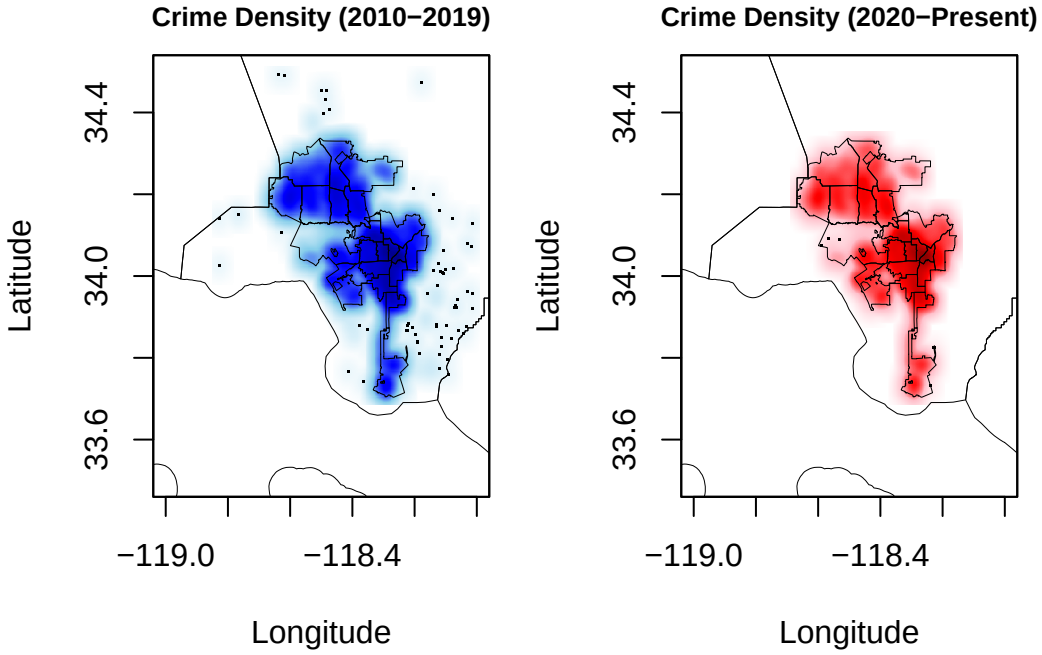

Figure 1: Crime densities in Los Angeles by LAPD Division, 2010-2019 and 2020-Present

Figure 1 shows that both time periods show major clusters of reported crimes in central LA. The 2020–present data appears slightly more dispersed to the north, but overall, the density of crimes has remained the same over time. The West LA and Foothill LAPD divisions seem to have the lowest density of crimes. West LA Division has crimes mostly located in the south of its area. Beverly Hills, which is not one of the LAPD divisions and has its own police force, explains the whiter spot in central LA.

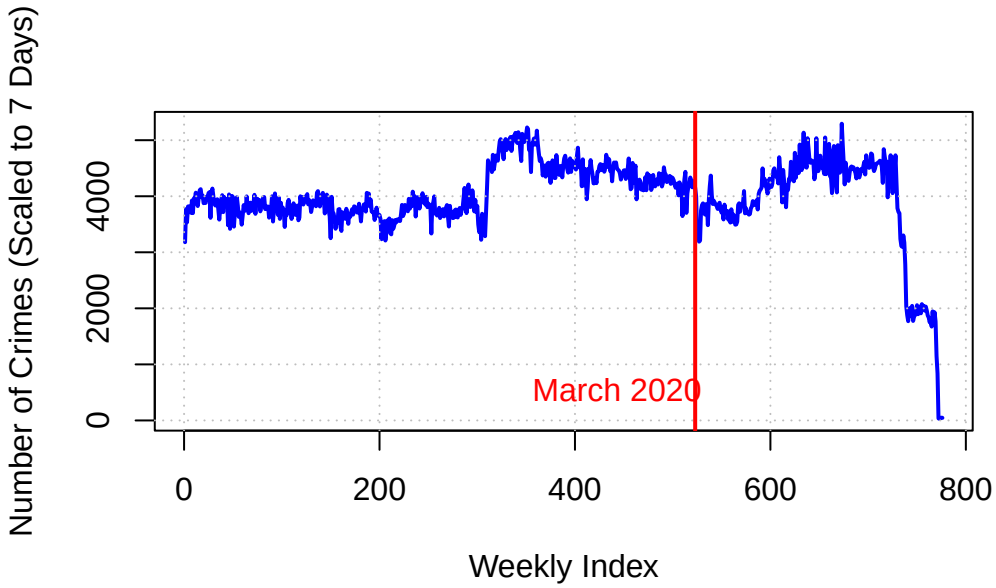

Figure 2: Weekly Crime Counts in Los Angeles from 2010 to present

Figure 2 analyzes the number of crimes over time using time series. The time series shows stable weekly crime counts with occasional large spikes, likely due to shorter weeks. These surges indicate certain weeks where crime reporting jumped significantly, there is a recent interesting surge that seems to mean-revert in 2024. The red line, which denotes March 2020, shows a decline in crimes, likely during the shutdown and quarantine, but it returned to normal levels quickly.

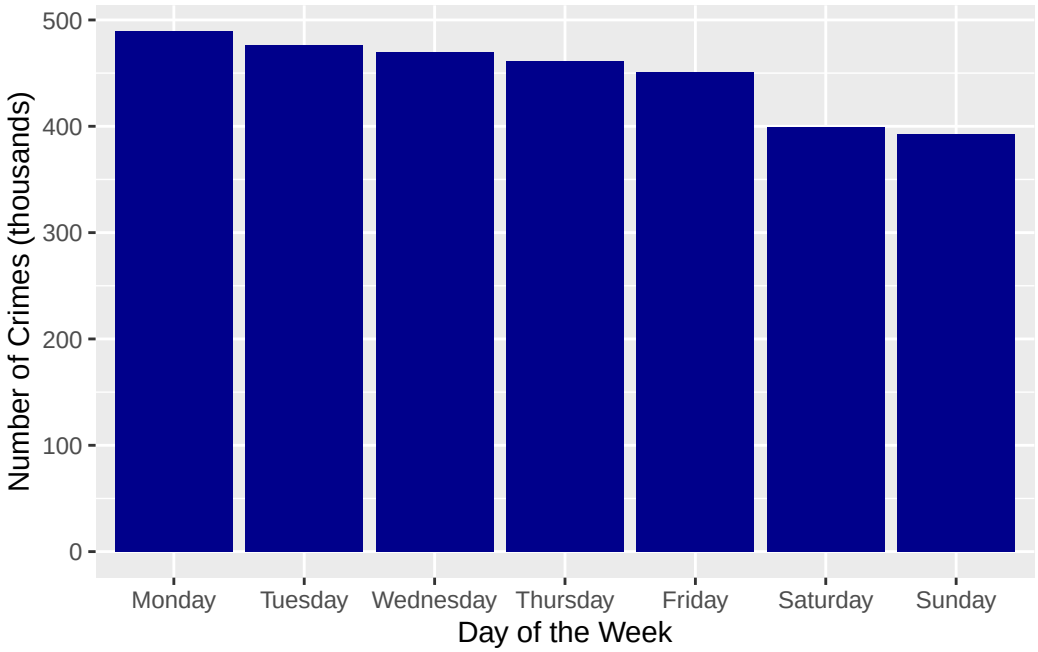

Figure 3: Los Angeles Crimes by Day of the Week

In Figure 3, the number of crimes are rather evenly distributed, with the weekends having slightly lower crime compared to week days, which may be a result from more people around during weekends which indirectly supervises people's behavior. Most crimes occur on Mondays, and the fewest crimes occur on Sunday. The plot analyzes the crime rate frequency frequency by day of the week using a bar graph.

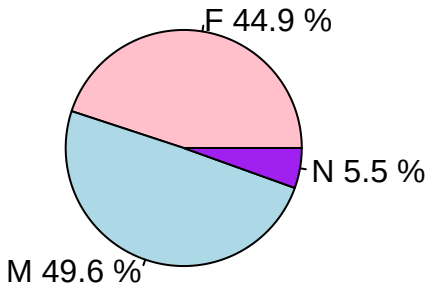

Figure 4: Percent of crime victims by gender, 2010 to present

From the pie chart shown in Figure 4, we can see that around 50% of crime victims are male, 45% are female and 5% are non-binary. This shows that the distribution of crime victims by gender is relatively even, though men do account for a plurality of the victims.

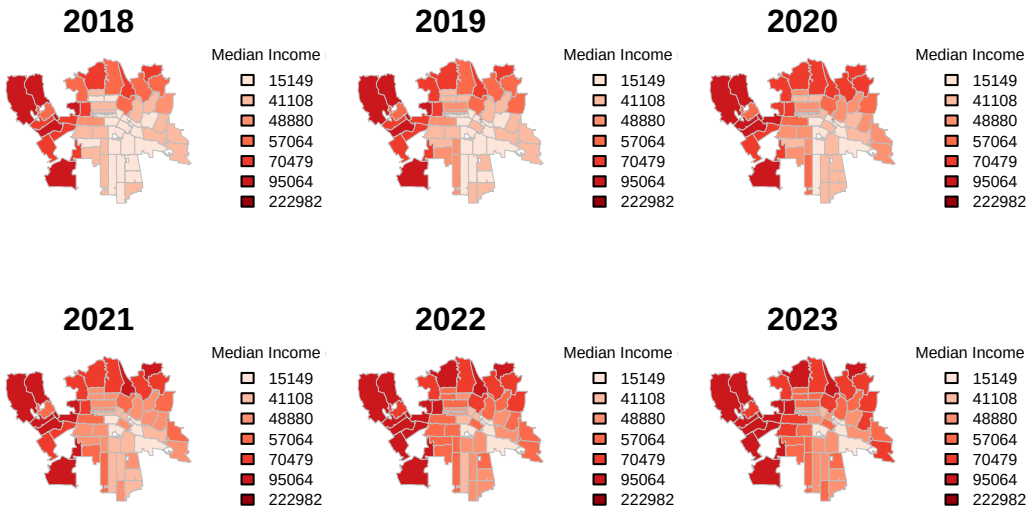

Figure 5: Median household income by year and by ZIP Code Tabulation Area (ZCTA) in Los Angeles

The maps in Figure 5 demonstrate that median income in certain LA zip codes (particularly those to the west) have a much higher median household income than the rest of the city, but over the last six years, the median income in most other LA zip codes has increased. However, there are some that a typical family is at or below the poverty line. We intend to investigate if there is a spatial correlation between these poorer zip codes and the incidence of crimes, particularly homicide, and whether this is affected by pandemic conditions.

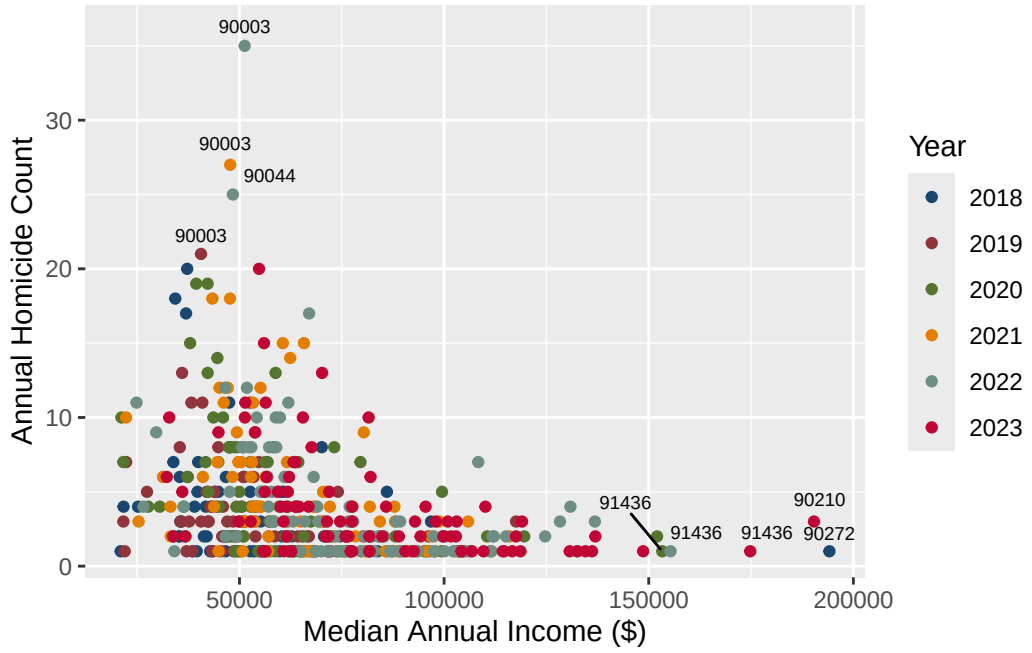

Figure 6: Annual Median Income and Homicide Count by ZCTA in Los Angeles with top 1% income and top 1% homicide rates labelled

Based on Figure 6, we observe a weak inverse relationship between median annual income and homicide count. Specifically, we see that higher homicide rates occur in ZCTA areas with lower median annual income, though most observations are still near or at zero, meaning few homicides typically occur. However,

there are no observations with a median annual income of over $100,000 where the homicide rate is greater than 10, whereas there are several of those observations for ZCTA/year combinations under $100,000. The zip code which stands out most on this figure is 90003, which has 3 of the 4 highest homicide counts in the data. This suggests that there might be a relationship between median income and homicide count, though the relationship might be non-linear or weak, suggesting there are other factors at play.

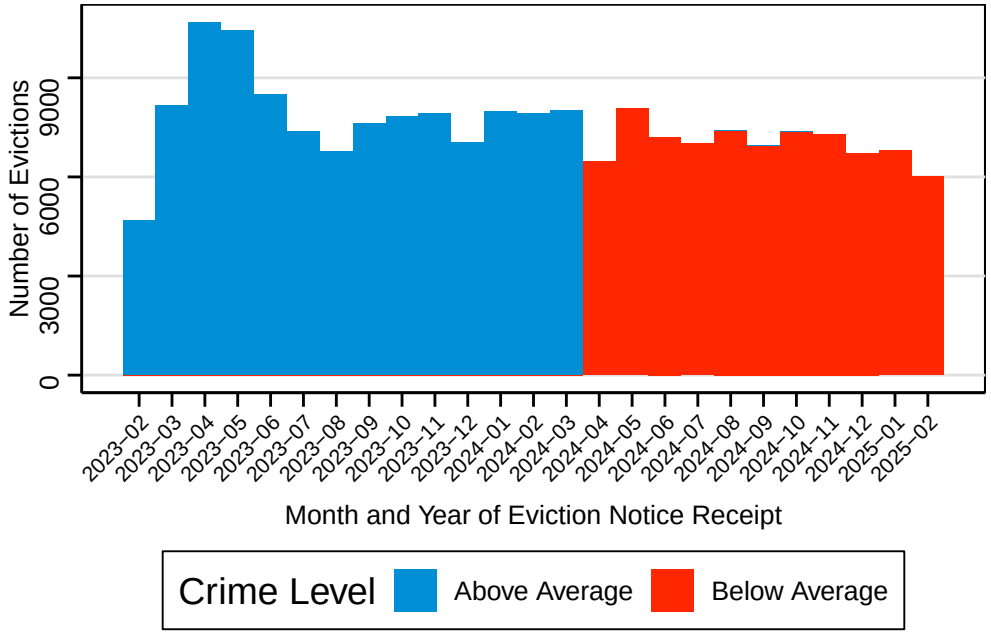

Figure 7: Monthly Frequency of Evictions and Crimes in the City of Los Angeles, February 2023 to February 2025

Figure 7 plots the frequency of evictions in Los Angeles from February 2023 to February 2025, where the City of Los Angeles Housing Department has published case data. By summarizing both the crime and the evictions datasets by month, we see that the number of evictions was highest in early-mid 2023, and since, there have typically been slightly lower frequencies of evictions. Additionally, we split the number of crimes per month based on the average number of crimes of 14,439 to denote above and below average number of crimes per month. There is a clear split between March and April of 2024, as all months before that had above average crime, and all after had below average. Interestingly, the months where there is below average crime, seem to be the ones with slightly lower numbers of evictions.

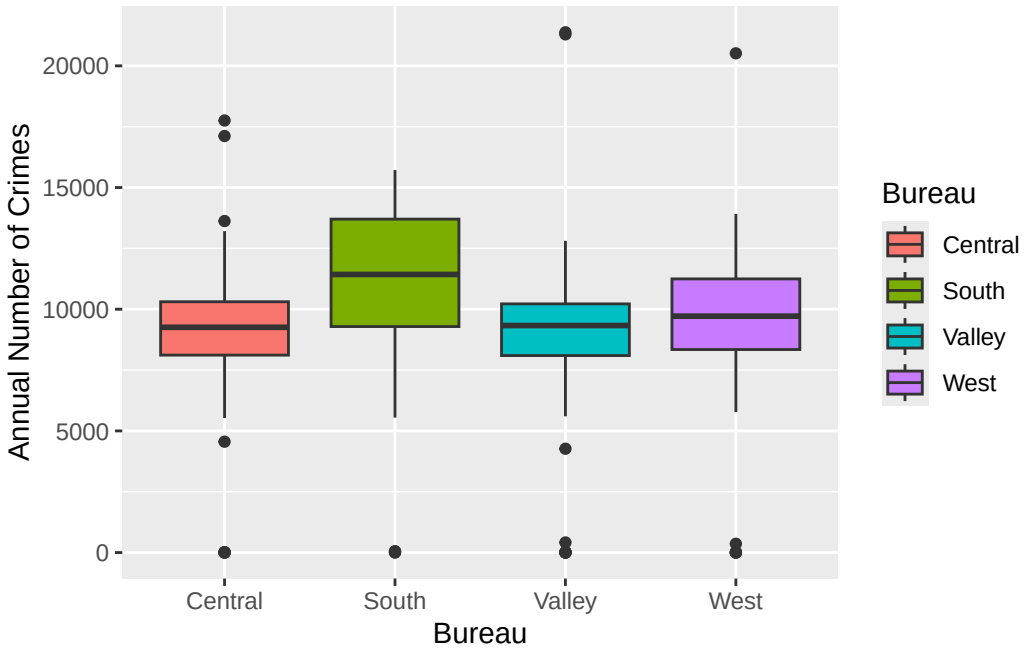

Figure 8: Annual Number of Crimes by Los Angeles Police Department Bureau

Figure 8 shows the number of crimes in the four different bureaus of Los Angeles annually. We categorized the areas into their respective Central, South, Valley, and West Bureaus. The South Bureau has the highest median number of crimes, as well as the largest variability, while the Central, Valley, and West bureaus have relatively similar distributions with lower crime counts.The outliers suggests some years had significantly lower crime counts than others. This visualization helps us understand how crime is distributed across different regions, with the South experiencing the highest fluctuation and overall crime levels.

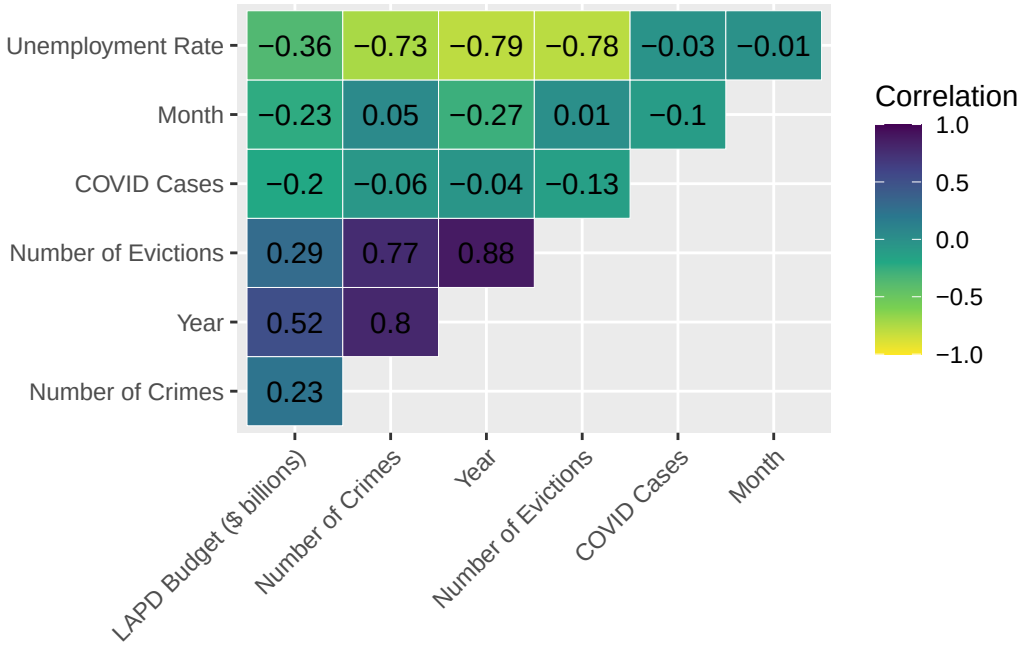

Figure 9: Strength of Correlations Between the Number of Crimes and Different Law Enforcement and Socioeconomic Factors

Figure 9 shows the correlations between different variables that could affect the number of crimes committed. We see that the most significant correlations for the number of crimes are the unemployment rate, year, and number of evictions. Years is positively correlated with number of crimes, meaning that as time passed, the number of crimes increased. Similarly, as the number of evictions increases, so does the number of crimes. Interestingly, as the unemployment rate increases, the number of crimes decreases, which is the opposite effect of what was expected (one would anticipate that higher unemployment would result in more crime). The LAPD budget interestingly had a medium-strength positive correlation with number of crimes, suggesting we should further examine the increase in the LAPD budget compared to the increase in crimes.

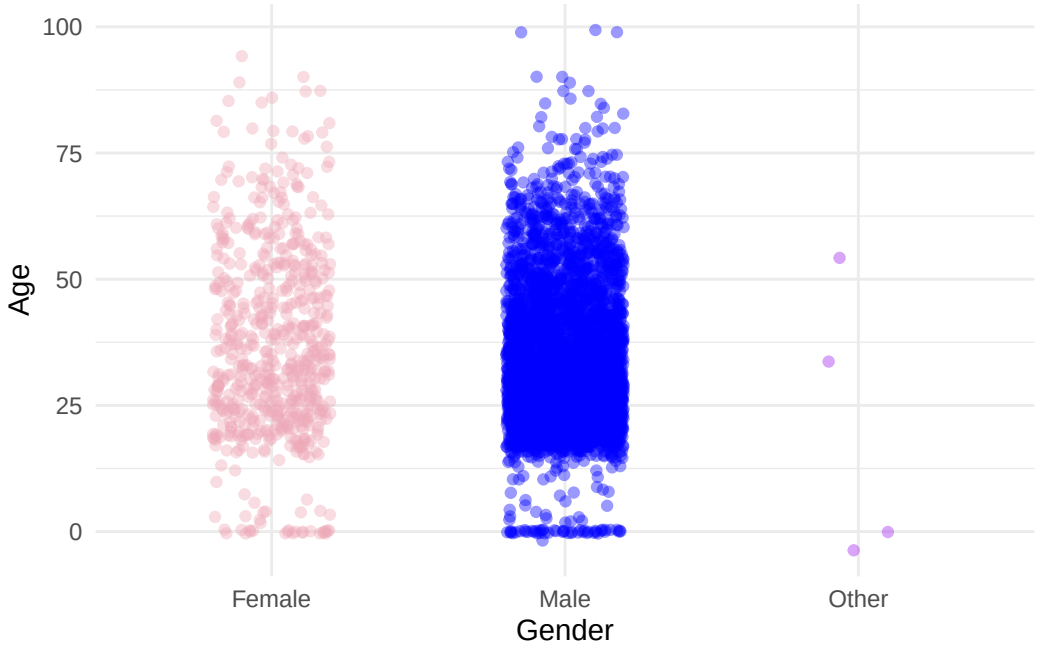

Figure 10: Homicide Victims by Age and Gender in Los Angeles 2010-Present

The jitter plot in Figure 10 shows us both the frequencies and the demographics of homicide victims in Los Angeles from 2010 to present. Based on the density of the points, we see that most homicide victims in the last 15 years have been men between the ages of around 15 to 65. Additionally, we see that there are not many victims who are toddlers and younger children, but there is a good number of babies who have been killed. This is an intricacy in our data that might be interesting to look into further, particularly into the circumstances of those deaths and potential correlations between reports of child abuse and/or domestic violence and infanticide.

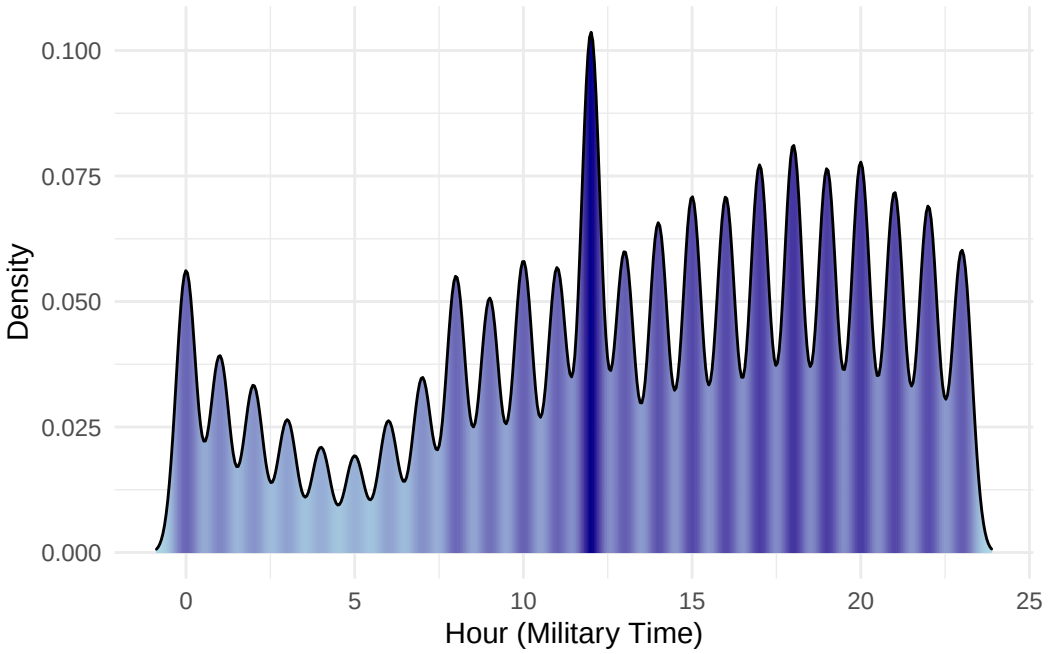

Figure 11: Frequency of Crimes in Los Angeles Throughout the Day

The density plot in Figure 11 shows the frequency of crimes depending on the hour (in military time) that they occurred. Based on the plot, we see that fewer crimes occur early in the morning (with the lowest at 5 AM). The data peaks at noon, meaning the most crimes occur in that hour of the day. Crimes are more frequent after that, until the early morning. The second most common crime occurrence is 6 PM, or 18 hours.

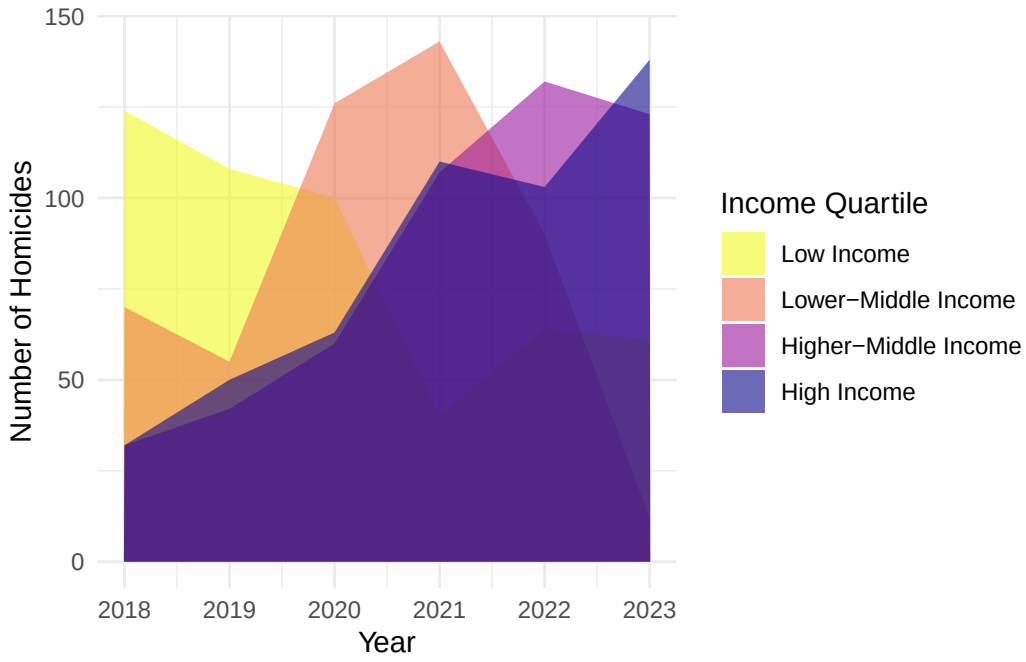

Figure 12: Time series analyzing homicide rate trends over time by median income quartiles

Figure 12 shows that homicide rates were low pre-pandemic and had a sharp increase since the pandemic,

except in areas of low income, where homicide rates were high before the pandemic. In lower income areas, the rates of homicide have decreased significantly since the end of the pandemic. On the other hand, areas of higher-middle and high income have seen large increases in homicide in the six years shown.

3 Statistical Analyses

To examine the relationship between the number of homicides and various socioeconomic and enforcement factors, we will use a variety of statistical analyses to see if certain factors are more influential in causing increases or decreases in homicide rates.

For the purposes of this analysis, homicide rate is quantified as the annual number of homicides in a particular entity, whether that is ZIP Code Tabulation Area (ZCTA) or Census Tract. Additionally, we will examine the influence of general geographic location through the latitude and longitude where the homicide was committed.

3.1 Regression to Examine Weekly Crime Rates

$$\text{Crimes}_i = \beta_0 + \beta_1 \text{Index}_i + \beta_2 \text{CrimesLag}_i + \epsilon_i \tag{1}$$

Table 1: Linear Regression Model Results Estimating Crimes Per Week

Predictors	Estimates	CI	p
(Intercept)	130.85	37.41 – 224.29	**0.006**
Index	-0.05	-0.12 – 0.02	0.149
CrimesLag	0.97	0.95 – 0.99	**<0.001**
Observations	775		
R^2 / R^2 adjusted	0.907 / 0.907		

The lagged crime coefficient (0.81949) suggests that a high crime week is often followed by another high crime week. Since the trend term is not significant, there is no clear drift in weekly crime over time.

3.2 Regression Between Median Income and Crimes per Zip Code

$$\text{HomicideCount}_i = \beta_0 + \beta_1 \text{cali\_median\_income}_i + \epsilon_i \tag{2}$$

Table 2: Linear Regression Model Results Estimating Crimes Based on Median Income

Predictors	Estimates	CI	p
Intercept	8.10	7.12 – 9.08	**<0.001**
Median Income	-0.00	-0.00 – -0.00	**<0.001**
Observations	467		
R^2 / R^2 adjusted	0.129 / 0.128		

Based on the p-value, we observe that median income is not significant at a 0.05 level, meaning it does not significantly decrease the amount of crimes committed. The r^2 value additionally tells us that the income only explains a small portion of the variation in the number of crimes. As stated with the scatterplot, a larger sample and considering other factors might yield different results.

3.3 Negative Binomial for Census Block

$$\text{HomicideCount}_i = \beta_0 + \sum_{j=1}^{k} \beta_j \text{Factor}_j + \epsilon_i \tag{3}$$

Table 3: Annual Homicide Count by Census Block

Predictors	Incidence Rate Ratios	CI	p
(Intercept)	120.27	104.87 – 138.62	**<0.001**
Census Block 2	0.89	0.73 – 1.08	0.230
Census Block 3	0.39	0.32 – 0.48	**<0.001**
Census Block 4	0.11	0.09 – 0.14	**<0.001**
Census Block 5	0.02	0.01 – 0.03	**<0.001**
Census Block 6	0.01	0.00 – 0.04	**<0.001**
Observations	77		
R^2 Nagelkerke	1.000		

3.3.1 Discussion

The results of this model show a significant relationship between Census Block and the homicide rate, except for Census Block 2. Census Block 2 experiences a higher incidence of homicides, but it is not significantly different. The other Census Blocks show a statistically significant decrease in homicide rates. However, the R2 might suggest overfitting.

3.4 Linear Regression Models

Model 1 is defined as:

$$\text{HomicideCount}_i = \beta_0 + \beta_1 \text{LAT\_rounded}_i + \beta_2 \text{LON\_rounded}_i + \epsilon_i \tag{4}$$

Model 2 is defined as:

$$\text{HomicideCount}_i = \beta_0 + \beta_1 \text{year}_i + \epsilon_i \tag{5}$$

Model 3 is defined as:

$$\text{HomicideCount}_i = \beta_0 + \beta_1 \text{LAT\_rounded}_i + \beta_2 \text{LON\_rounded}_i + \beta_3 \text{year}_i + \epsilon_i \tag{6}$$

Model 4 is defined as:

$$\begin{aligned} \text{HomicideCount}_i = {} & \beta_0 + \beta_1 \text{LAT\_rounded}_i + \beta_2 \text{LON\_rounded}_i + \beta_3 \text{year}_i \\ & + \beta_4 (\text{LAT\_rounded}_i \times \text{LON\_rounded}_i) + \beta_5 (\text{LAT\_rounded}_i \times \text{year}_i) \\ & + \beta_6 (\text{LON\_rounded}_i \times \text{year}_i) + \beta_7 (\text{LAT\_rounded}_i \times \text{LON\_rounded}_i \times \text{year}_i) + \epsilon_i \end{aligned} \tag{7}$$

Table 4: Table 4: Geospatial Linear Regression Models Predicting Homicide Count

	Model 1		Model 2		Model 3		Model 4	
Predictors	Estimates	p	Estimates	p	Estimates	p	Estimates	p
(Intercept)	254.07	**<0.001**	-104.25	**0.008**	142.22	**0.010**	-62425.36	0.247
Latitude	-1.05	**0.001**			-1.05	**0.001**	-1147.08	0.060
Longitude	1.83	**<0.001**			1.85	**<0.001**	-596.80	0.240

Year		0.05	**0.007**	0.06	**0.003**	46.28	0.068
Latitude:Longitude						-7.63	0.056
Latitude:Year						0.12	0.533
Longitude:Year						0.43	0.078
Observations	1237	1237		1237		1237	
R^2 / R^2 adjusted	0.057 / 0.055	0.006 / 0.005		0.064 / 0.061		0.069 / 0.064	

Table 4 shows the results of four different linear regression models that were run using homicide counts by year and by latitude and longitude. Longitude and latitude were rounded to two decimal places to summarize the counts but maintain a good level of granularity in the coordinates data. Model 1 looked only at the geospatial aspect of the data and found that both latitude and longitude were significant predictors of homicide rate. As latitude increases, homicide rate decreases, meaning that homicides are more common towards the south and as you move north, homicide counts drop. As longitude increases, the number of homicides increases as well, meaning that homicides are more common to the East.

The second model finds that year is significant as a predictor of homicide rate, and a model looking at both spatial and temporal characteristics also finds year to be significant.

The fourth model, which also considered the interactions between these three features, did not find any significance.

References

Krieger, Nancy, Pamela Waterman, Jarvis T. Chen, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson. 2002. "ZIP Code Caveat: Bias Due to Spatiotemporal Mismatches Between ZIP Codes and US Census–Defined Geographic Areas—the Public Health Disparities Geocoding Project." *American Journal of Public Health* 92 (7): 1100–1102. https://doi.org/10.2105/ajph.92.7.1100.

Moore, Kesha S., Ryan Tom, and Jackie O'Neil. 2022. "The Truth Behind Crime Statistics: Avoiding Distortions and Improving Public Safety." https://www.naacpldf.org/wp-content/uploads/2022-08-03-TMI-Truth-in-Crime-Statistics-Report-FINAL-2.pdf.

Data Sources

Primary Datasets

Crime Data From 2010-2019: https://data.lacity.org/Public-Safety/Crime-Data-from-2010-to-2019/63jg-8b9z/about_data

Crime Data From 2020-Present: https://catalog.data.gov/dataset/crime-data-from-2020-to-present

Secondary Datasets

Los Angeles County Eviction Data: https://calmatters.org/housing/homelessness/2023/11/california-evictions-post-pandemic/

City of Los Angeles Eviction Data: https://housing.lacity.gov/residents/renters/eviction-notices-filed

Median Income Data: https://data.census.gov/table/ACSST1Y2023.S1903?q=income

4 Appendices

```
library(sf) # for map
library(tidyverse) # using only for joining datasets to build map visual
library(RColorBrewer) # for map color scheme
library(tidygeocoder) # to get zip codes of addresses
library(dplyr)
library(RSQLite)

con <- dbConnect(SQLite(), "0_Data/crime_in_la.db")
```

```
query <- dbSendQuery(con, "Select * From crime_la Where year < 2020")
crime_la_2010_2019 <- dbFetch(query,n=-1)
dbClearResult(query)

query <- dbSendQuery(con, "Select * From crime_la Where year > 2019")
crime_la_2020_present <- dbFetch(query,n=-1)
dbClearResult(query)

lapd_shape <- read_sf(dsn = "0_Data/Raw_Data/LAPD_Divisions", layer = "LAPD_Divisions")

lapd_shape <- st_transform(lapd_shape, crs = 4326)

cali <- read_sf(dsn = "0_Data/Raw_Data/ca_counties", layer = "ca_counties")

cali <- st_transform(cali, crs = 4326)

par(mfrow = c(1, 2), mar = c(4, 4, 2, 1))

smoothScatter(
  x    = crime_la_2010_2019$LON,
  y    = crime_la_2010_2019$LAT,
  xlim = c(-119, -118),
  ylim = c(33.5, 34.5),
  nbin = 300,
  xlab = "Longitude",
  ylab = "Latitude",
  main = "Crime Density (2010-2019)",
  colramp = colorRampPalette(c("white", "skyblue", "blue", "darkblue")),
  cex.main=0.8
)

plot(lapd_shape$geometry, lwd=.5, add=TRUE)

plot(cali$geometry, add = TRUE, lwd=.5)

smoothScatter(
  x    = crime_la_2020_present$LON,
  y    = crime_la_2020_present$LAT,
  xlim = c(-119, -118),
  ylim = c(33.5, 34.5),
  nbin = 300,
```

```
  xlab = "Longitude",
  ylab = "Latitude",
  main = "Crime Density (2020-Present)",
  colramp = colorRampPalette(c("white", "pink", "red", "darkred")),
  cex.main=0.8
)

plot(lapd_shape$geometry, lwd=.5, add=TRUE)

plot(cali$geometry, add = TRUE, lwd=.5)

par(mfrow = c(1, 1))
```

```
query <- dbSendQuery(con, "Select Count(*) as Crimes, YearWeek From crime_la Group By YearWeek Order By
weekly_counts <- dbFetch(query,n=-1)
dbClearResult(query)

weekly_counts$Index <- seq_len(nrow(weekly_counts))
# Find the Index for March 2020
march_2020_index <- which(weekly_counts$YearWeek == "2020-09")  # Adjust '202009' to match your data's

# Plot
plot(
  weekly_counts$Index,
  weekly_counts$Crimes,
  type = "l",
  col  = "blue",
  lwd  = 2,
  xlab = "Weekly Index",
  ylab = "Number of Crimes")

abline(h = pretty(weekly_counts$Crimes), v = pretty(weekly_counts$Index),
       col = "gray", lty = "dotted")

# Add vertical line for March 2020
abline(v = weekly_counts$Index[march_2020_index], col = "red", lwd = 2)

# Add a label for March 2020
text(
  x = weekly_counts$Index[march_2020_index] - 80,
  y = 10,
  labels = "March 2020",
  pos = 3,
  col = "red"
)
```

```
# Convert Date.Rptd to Date format (ignore time part)
#crime_la$Date.Rptd
crime_la <- rbind(crime_la_2010_2019, crime_la_2020_present)
# Extract the day of the week
crime_la$Date <- as.Date(substr(crime_la$Date.Rptd, 1, 10), format = "%m/%d/%Y")

crime_la$DayOfWeek <- weekdays(crime_la$Date)
```

```r
# crime_la$DayOfWeek

# Count occurrences of crimes per day
crime_counts <- table(crime_la$DayOfWeek)
# crime_counts

# Order the days correctly
day_order <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")
crime_counts <- data.frame(crime_counts[day_order])
colnames(crime_counts) <- c("Day", "CrimeCount")

crime_counts$thousands <- crime_counts$CrimeCount/1000

# Plot bar chart with horizontal labels
ggplot(crime_counts, aes(x = Day, y = thousands)) +
  geom_col(fill = "blue4") +
  labs(x = "Day of the Week",
       y = "Number of Crimes (thousands)")
```

```r
query <- dbSendQuery(con, "SELECT COUNT(*) As count, Modified_Vict_Sex As `Vict.Sex` FROM (
    SELECT *,
        CASE
            WHEN `Vict.Sex` NOT IN ('M', 'F') THEN 'N'
            ELSE `Vict.Sex`
        END AS `Modified_Vict_Sex`
    FROM crime_la c
    WHERE `Vict.Sex` != '' AND `Vict.Sex` IS NOT NULL) s
    GROUP BY Modified_Vict_Sex
")
gender_counts <- dbFetch(query,n=-1)
dbClearResult(query)

# Create a pie chart using base R
color <- c("F" = "pink", "M" = "lightblue", "N" = "purple")
pie(
  gender_counts$count,
  labels = paste(gender_counts$Vict.Sex,
                 round(gender_counts$count / sum(gender_counts$count) * 100, 1), "%"),
  col = color[gender_counts$Vict.Sex]
)
```

```r
zcta <- read_sf(dsn = "0_Data/Raw_Data/cb_2020_us_zcta520_500k", layer = "cb_2020_us_zcta520_500k")
query <- dbSendQuery(con, "SELECT * FROM zctas_in_la")
LA_zctas <- dbFetch(query,n=-1)
dbClearResult(query)

colnames(LA_zctas) <- c("X", "ZCTA5CE20")
LA_zctas$ZCTA5CE20 <- as.character(LA_zctas$ZCTA5CE20)

# joining the zcta shape file with LA zcta reference file
zcta_filtered <- zcta %>% inner_join(LA_zctas, by = "ZCTA5CE20")

# getting LA median income using joined file
```

```r
query <- dbSendQuery(con, "SELECT ZCTA, S1903_C03_001E, year FROM cali_median_income")
median_household_income <- dbFetch(query,n=-1)
dbClearResult(query)

zcta_with_data <- zcta_filtered %>% inner_join(median_household_income, join_by("ZCTA5CE20" == "ZCTA"))

# remove NAs
zcta_no_NAs <- zcta_with_data %>% filter(!is.na(S1903_C03_001E), S1903_C03_001E != "-")

zcta_no_NAs$S1903_C03_001E <- as.integer(zcta_no_NAs$S1903_C03_001E)
```

```r
# quantiles for the entire dataset (from 2018 to 2023) to divide the median income into groups
qntls <- round(quantile(zcta_no_NAs$S1903_C03_001E, probs = seq(0, 1, length.out = 7)), 0)

# 2x3 plot layout for years 2018-2023, setting margins and outer margins, main title sizes
par(mfrow = c(2, 3), mar = c(1, 1, 1, 5), oma = c(0, 0, 3, 0), cex.main = 1.5)

# color palette for sequential data
palette <- brewer.pal(7, "Reds")

yrs <- 2018:2023

for (yr in yrs) {
  # get just that year's data
  subset_data <- zcta_no_NAs[zcta_no_NAs$year == yr, ]

  # creating bins based on the qtls
  bins <- cut(subset_data$S1903_C03_001E, breaks = qntls, labels = FALSE, include.lowest = TRUE)

  # Plot
  plot(st_geometry(subset_data),
       col = palette[bins],
       border = 'grey',
       axes = FALSE,
       lwd = 0.5)
  title(yr, line = -1)

  # adding legend
  legend("right", inset = c(-.75, 0), legend = qntls, fill = palette, title = "Median Income ($)", bty =
}
```

```r
zcta <- read_sf(dsn = "0_Data/Raw_Data/cb_2020_us_zcta520_500k", layer = "cb_2020_us_zcta520_500k")
query <- dbSendQuery(con, "SELECT i.*, c.*
FROM cali_median_income i
INNER JOIN crimes_with_zips c
ON i.year = c.year AND CAST(i.ZCTA AS int) = c.ZCTA5CE20")
crime_income <- dbFetch(query,n=-1)
dbClearResult(query)

crime_type <- crime_la %>% dplyr::select(year, Crm.Cd.Desc, DR_NO)

crime_income_type <- crime_income[,-c(249, 247)] %>% inner_join(crime_type, join_by("year" == "year", "
```

```r
homicide_income_by_year <- crime_income_type %>%
  filter(Crm.Cd.Desc == "CRIMINAL HOMICIDE") %>%
  group_by(year, ZCTA5CE20) %>%
  summarize(HomicideCount = n(),
            cali_median_income = median(as.numeric(S1903_C03_001E))) %>%
  ungroup()

homicide_income_by_year <- homicide_income_by_year %>% filter(!is.na(HomicideCount) & !is.na(cali_median

homicide_income_by_year$label <- ifelse(
  homicide_income_by_year$cali_median_income > quantile(homicide_income_by_year$cali_median_income, 0.9
  homicide_income_by_year$HomicideCount > quantile(homicide_income_by_year$HomicideCount, 0.99),
  homicide_income_by_year$ZCTA5CE20,
  "")

ggplot(data = homicide_income_by_year, aes(x = cali_median_income,
                                           y = HomicideCount)) +
  geom_point(aes(color = as.factor(year))) +
  ggthemes::scale_color_stata() +
  ggrepel::geom_text_repel(aes(label = label), size = 2.5, nudge_x = 2, nudge_y = 1) +
  labs(x = "Median Annual Income ($)",
       y = "Annual Homicide Count",
       color = "Year")
```

```r
query <- dbSendQuery(con, "SELECT * FROM evictions_city_la")
evictions_la <- dbFetch(query,n=-1)
dbClearResult(query)
# fix mistake
evictions_la$Date.Received <- gsub("2205", "2025", evictions_la$Date.Received)

# convert to date
evictions_la$month_year <- format(as.Date(evictions_la$Date.Received, format="%m-%d-%Y"), "%Y-%m")

# get unique months
unique_months <- data.frame(month_year = sort(unique(evictions_la$month_year)), index = 1:length(unique

evictions_la_indexed <- evictions_la %>% left_join(unique_months, by = "month_year")

# create color scheme
crime_la$month_year <- format(crime_la$Date, "%Y-%m")
crime_la_monthly <- crime_la %>%
  group_by(month_year) %>%
  summarize(num_crimes = n())

evictions_la_indexed_crimes <- evictions_la_indexed %>% left_join(crime_la_monthly, by = "month_year")

avg_crimes <- mean(evictions_la_indexed_crimes$num_crimes)

evictions_la_indexed_crimes$high_low_crime <- as.factor(ifelse(evictions_la_indexed_crimes$num_crimes >


ggplot(data = evictions_la_indexed_crimes, aes(x = index)) +
  geom_histogram(aes(fill = high_low_crime), bins = nrow(unique_months)) +
```

```
  scale_x_continuous(breaks=unique_months$index,
        labels=unique_months$month_year) +
  labs(x = "Month and Year of Eviction Notice Receipt", y = "Number of Evictions", fill = "Crime Level")
  ggthemes::scale_fill_fivethirtyeight() +
  ggthemes::theme_stata(scheme = "s1mono") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.title.x = element_text(vjust = -.75),
        plot.title = element_text(size=12))
```

```
crime_la$year <- lubridate::year(crime_la$DateParsed)
crime_by_area <- crime_la %>% group_by(AREA.NAME, year) %>% summarize(numberOfCrimes = n()) %>% ungroup

query <- dbSendQuery(con, "SELECT * FROM lapd_bureaus")
bureaus <- dbFetch(query,n=-1)
dbClearResult(query)

bureaus$Division <- gsub("77th", "77th Street", bureaus$Division)
bureaus$Division <- gsub("North", "N", bureaus$Division)
bureaus$Division <- gsub("Neast", "Northeast", bureaus$Division)

query <- dbSendQuery(con, "SELECT COUNT(*) FROM crime_la GROUP BY `AREA.NAME`, year")
test <- dbFetch(query,)
crime_by_area <- crime_by_area %>% left_join(bureaus, join_by("AREA.NAME" == "Division"))

ggplot(crime_by_area, aes(y = numberOfCrimes, group = Bureau, x = Bureau)) +
        geom_boxplot(aes(fill = Bureau)) +
  labs(y = "Annual Number of Crimes")
```

```
library(lubridate)
# creating dataset for correlations

# since we do not have pre- and post- pandemic city of LA data, we are instead using LA county data for
query <- dbSendQuery(con, "SELECT * FROM evictions_county_la")
LA_county_evictions <- dbFetch(query,n=-1)
dbClearResult(query)


# covid data
query <- dbSendQuery(con, "
  SELECT
    CAST(strftime('%Y', ep_date) AS INTEGER) AS year,
    CAST(strftime('%m', ep_date) AS INTEGER) AS month,
   CAST(strftime('%d', ep_date) AS INTEGER) AS date,
   cases_14day
  FROM covid_cases
")
covid_14day <- dbFetch(query,n=-1)
dbClearResult(query)
# since there is a 7 day lag in reporting, let's do monthly = cases on the 21st and on the 5th of the n
covid_14day$lagged <- covid_14day$date + 7

covid_la_lagged <- covid_14day %>% filter(lagged %in% c(21, 35)) %>% arrange(month, year)
```

```r
covid_la_monthly <- covid_la_lagged %>% group_by(month, year) %>% summarize(cases = sum(cases_14day)) %

covid_la_monthly$index <- 1:nrow(covid_la_monthly)

# next, unemployment rate
query <- dbSendQuery(con, "
  SELECT
    u.month,
    u.Year,
    u.Value,
    l.police_budget
  FROM unemployment_la AS u
  INNER JOIN lapd_budget AS l ON u.Year = l.year
")
unemployment_lapd_budget <- dbFetch(query,n=-1)
dbClearResult(query)

crime_la_monthly$date <- ymd(paste0(crime_la_monthly$month_year, "-01"))
crime_la_monthly$month <- month(crime_la_monthly$date)
crime_la_monthly$year <- year(crime_la_monthly$date)

correlations_df <- crime_la_monthly %>% inner_join(unemployment_lapd_budget, join_by("year" == "Year", "

correlations_df2 <- correlations_df %>% inner_join(covid_la_monthly, by = c("month", "year"))

correlations_df3 <- correlations_df2 %>% inner_join(LA_county_evictions, join_by("year" == "Year", "mon

correlations <- correlations_df3 %>% dplyr::select(year, month, unemployment_rate, num_crimes, cases, nu

Cor_mat <- round(cor(correlations), 2)

# get only the upper triangle of variables
# (because a correlation heat map shows each correlation value twice)
get_upper_tri <- function(cormat){
  # lower.tri -> Returns a matrix of logicals, so set the values that are true to NA
  cormat[lower.tri(cormat)]<- NA # set the correlation matrix lower triangle to NA
  return(cormat)
}

# reorder them by the strength of the correlation
reorder_cormat <- function(cormat){
  # Use correlation between variables as distance
  dd <- as.dist((1-cormat)/2)
  hc <- hclust(dd) # use hierarchical clustering to reorder
  cormat <-cormat[hc$order, hc$order]
}

# use the functions above to remove the unnecessary parts
Cor_mat<-reorder_cormat(Cor_mat)

Cor_mat_upper<-get_upper_tri(Cor_mat)

# using melt instead of pivot_longer because pivot_longer did not order correctly
```

```r
Cor_mat_upper<-reshape2::melt(Cor_mat_upper, na.rm=TRUE)

# remove the central values
# which show the correlation for each variable with itself
Cor_mat_upper<- Cor_mat_upper %>% filter(value != 1.00)


long_Cor <- Cor_mat_upper %>%
    mutate(Var1 = recode(Var1, year = 'Year', month = 'Month', unemployment_rate =  'Unemployment Rate'
           Var2 = recode(Var2, year = 'Year', month = 'Month', unemployment_rate =  'Unemployment Rate'

ggplot(data = long_Cor, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_viridis_c(direction = -1,limit = c(-1, 1)) +
  theme(axis.title.x =  element_blank(),
        axis.title.y = element_blank()) +
  geom_text(aes(label = value), color = "black", size = 4) +
   theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(size=12)) +
  labs(fill = "Correlation")
```

```r
homicide_by_gender <- crime_la %>% filter(Crm.Cd.Desc == "CRIMINAL HOMICIDE") %>% dplyr::select(Vict.Se

levels(homicide_by_gender$gender) <- c("Female", "Male", "Other")
ggplot(homicide_by_gender, aes(x = gender, y = Vict.Age, color = gender)) +
  geom_jitter(alpha = 0.4, width = 0.2) +
  labs(x = "Gender",
       y = "Age") +
  theme_minimal() +
  scale_color_manual(c("M", "F", "N"), values = c("pink2", "blue", "purple")) +
  theme(legend.position = 'none')
```

```r
crime_la$time <- sprintf("%04d",crime_la$TIME.OCC)
crime_la$Hour <- as.numeric(substr(as.character(crime_la$time), start = 1, stop = 2))

ggplot(crime_la, aes(x = Hour, y = 0, fill = after_stat(density))) +
  ggridges::geom_density_ridges_gradient(scale = 1, alpha = 0.5) +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  labs(x = "Hour (Military Time)",
       y = "Density") +
  theme_minimal() +
  theme(legend.position = 'none')
```

```r
query <- dbSendQuery(con, "
  SELECT i.S1903_C03_001E, x.year, i.ZCTA, x.LAT, x.LON
  FROM (
    SELECT c.LAT, c.LON, CAST(z.ZCTA5CE20 AS INTEGER) AS ZCTA5CE20, CAST(c.year AS INTEGER) AS year
    FROM crime_la AS c
    INNER JOIN crimes_with_zips AS z
      ON c.DR_NO = z.DR_NO
      AND c.LAT = z.LAT
      AND c.LON = z.LON
      AND c.X = z.X
```

```
      AND c.year = z.year
    WHERE c.\"Crm.Cd.Desc\" = 'CRIMINAL HOMICIDE'
  ) AS x INNER JOIN (SELECT m.S1903_C03_001E, CAST(m.year AS INTEGER) AS year, CAST(m.ZCTA AS INTEGER) A
    ON x.ZCTA5CE20 = i.ZCTA AND x.year = i.year
")

homicides <- dbFetch(query,n=-1)
dbClearResult(query)

# Form income quartiles
homicides <- homicides %>% mutate(income_quartile = ntile(S1903_C03_001E, 4))
homicide_trends <- homicides %>% group_by(year, income_quartile) %>% summarize(homicide_count = n(),
          ref = mean(S1903_C03_001E),
          .groups = "drop")

homicide_trends$income_quartile <- as.factor(homicide_trends$income_quartile)
# removing NAs
homicide_trends <- homicide_trends %>% filter(!is.na(homicide_count) & !is.na(income_quartile))

# creating labels
levels(homicide_trends$income_quartile) <- c("Low Income", "Lower-Middle Income", "Higher-Middle Income"
# Plot homicide trends over time by income quartile
ggplot(homicide_trends, aes(x = year, y = homicide_count)) +
  geom_area(aes(fill = as.factor(income_quartile)), alpha = 0.6, position = "identity") +
  labs(
    x = "Year",
    y = "Number of Homicides",
    fill = "Income Quartile"
  ) +
  theme_minimal() +
  scale_fill_viridis_d(option = "plasma", direction = -1)
```

```
library(broom)

weekly_counts$CrimesLag <- c(NA, weekly_counts$Crimes[-nrow(weekly_counts)])

ar1_trend_regression <- lm(Crimes ~ Index + CrimesLag, data = weekly_counts, na.action = na.omit)

sjPlot::tab_model(ar1_trend_regression, dv.labels = "Table 1: Linear Regression Model Results Estimating
```

```
lm_income_crimes <- lm(HomicideCount ~ cali_median_income, data = homicide_income_by_year)

# summary(lm_income_crimes)

sjPlot::tab_model(lm_income_crimes, dv.labels = "Table 2: Linear Regression Model Results Estimating Cri
```

```
library(kableExtra)
query <- dbSendQuery(con, "SELECT * FROM crimes_by_census_block")
crimes_by_cb <- dbFetch(query,n=-1)
dbClearResult(query)

homicides_by_cb <- crimes_by_cb %>% filter(str_detect(Crm.Cd.Desc, "CRIMINAL HOMICIDE")) %>% group_by(BI
```

```r
# shapiro.test(homicides_by_cb$HomicideCount) # cannot use parametric tests. Strong evidence against no

mean_homicides <- mean(homicides_by_cb$HomicideCount)
var_homicides <- var(homicides_by_cb$HomicideCount)
# since variance > mean, use negative binomial
library(MASS)
model <- glm.nb(HomicideCount ~ as.factor(BLKGRPCE), data = homicides_by_cb)

pretty_labels <- c("(Intercept)", paste("Census Block", 2:6))
sjPlot::tab_model(model, pred.labels = pretty_labels, dv.labels = "Table 3: Annual Homicide Count by Ce
```

```r
homicides_latlon_rounded <- homicides %>%
  mutate(LAT_rounded = round(LAT, 2),
         LON_rounded = round(LON, 2))

homicides_lat_long_yr <- homicides_latlon_rounded %>%
  group_by(year, LAT_rounded, LON_rounded) %>%
  summarize(HomicideCount = n()) %>% ungroup()

linear_model1 <- lm(HomicideCount ~ LAT_rounded + LON_rounded, data = homicides_lat_long_yr)
```

```r
linear_model2 <- lm(HomicideCount ~ year, data = homicides_lat_long_yr)

linear_model3 <- lm(HomicideCount ~ LAT_rounded + LON_rounded + year, data = homicides_lat_long_yr)

linear_model4 <- lm(HomicideCount ~ LAT_rounded * LON_rounded * year, data = homicides_lat_long_yr)

sjPlot::tab_model(linear_model1, linear_model2, linear_model3, linear_model4, pred.labels = c("(Interce
                  "Table 4: Geospatial Linear Regression Models Predicting Homicide Count")
```