Homicide Increases in the COVID-19 Era

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

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)

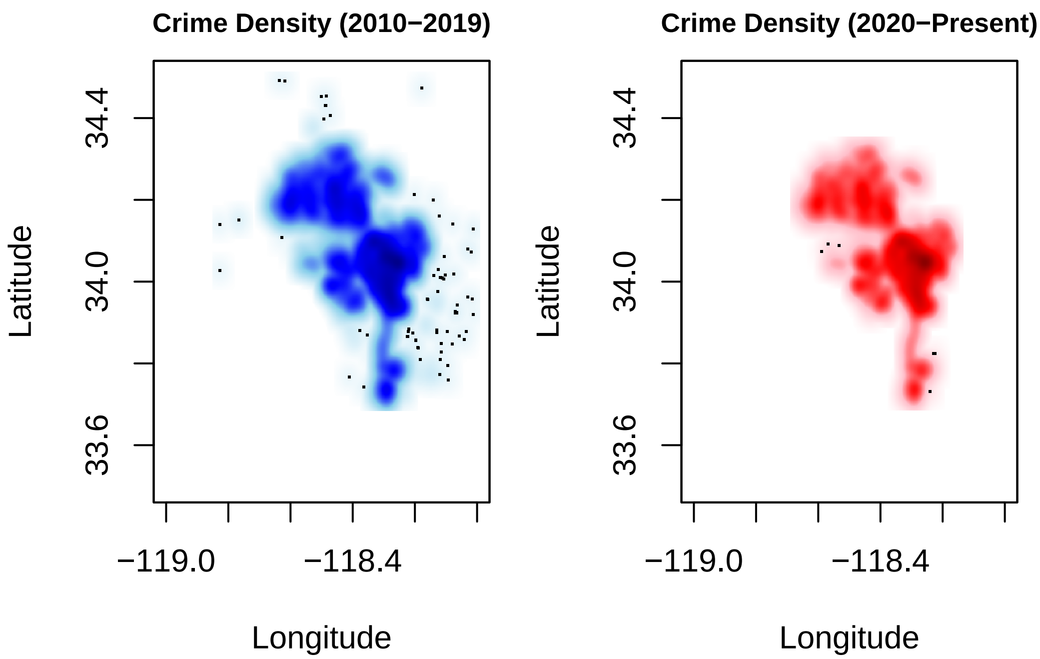
Using National Incident-Based Reporting System (NIBRS) data, this report will examine the impact of the pandemic and various socioeconomic factors on homicide rates in LA. We aim to compare the effect of these local characteristics on the incidence of homicide in pre-, mid-, and post-pandemic conditions.

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.

2 Exploratory Data Analysis

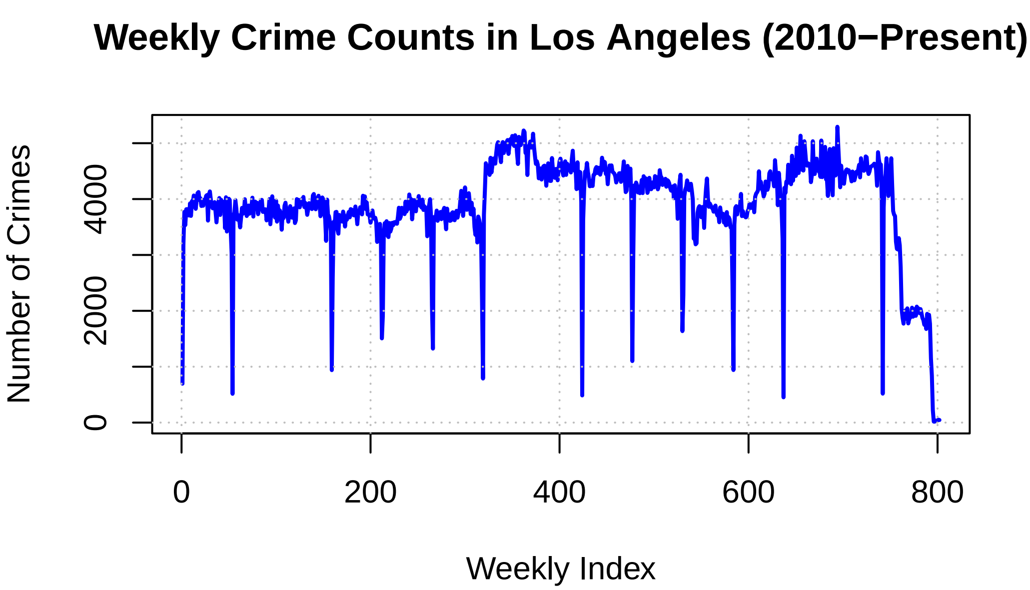
2.1 Examining Crime Density from 2010–2019 and 2020-Present



2.1.1 Interpretation

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 patterns remain similar looking.

2.2 Examining Number Of Crimes Over Time



2.2.1 Interpretation

The time series shows stable weekly crime counts with occasional large spikes. These surges indicate certain weeks where crime reporting jumped significantly, there is a recent interesting surge that seems to mean-revert in 2024.

2.3 Crime rate frequency by day of the week

```
[1] "C"
```

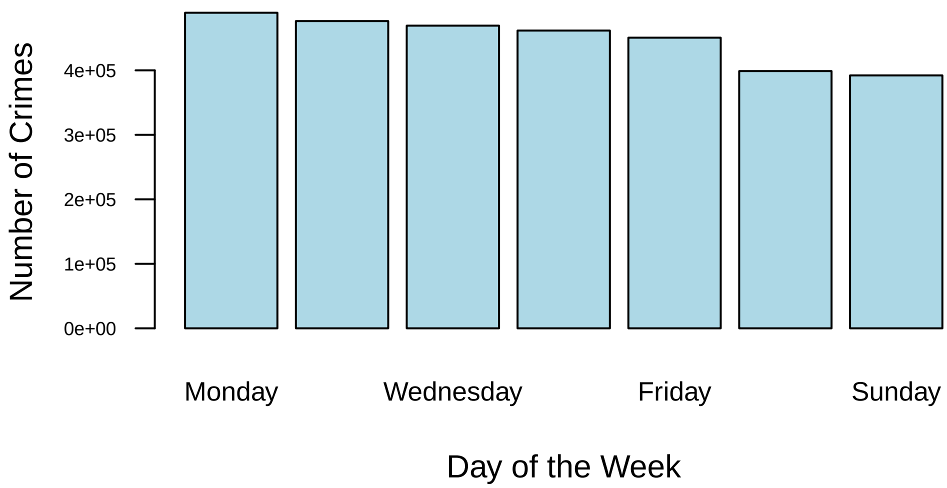
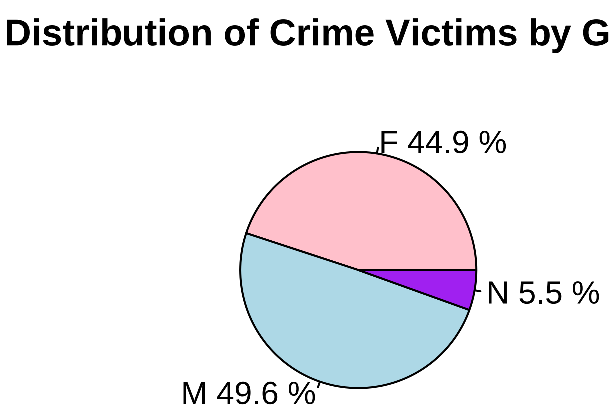


Figure 1: Crime Occurrences by Day of the Week

2.3.1 Interpretation

The number of crimes are rather evenly distributed, with the weekends having slightly lower crime compared to week days.

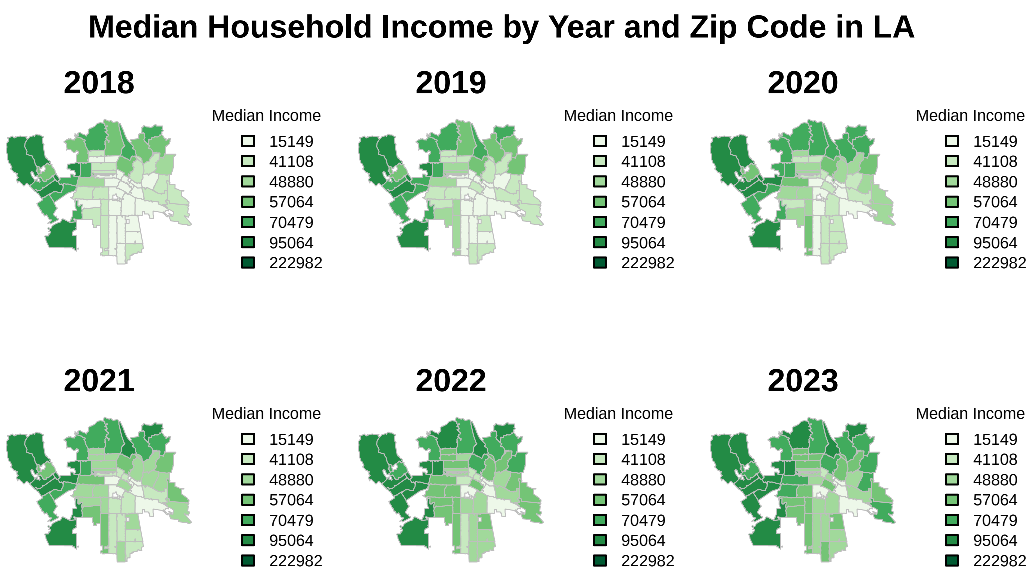
2.4 Gender Distribution for crimes over time



2.4.1 Interpretation

From the pie chart, we can see that around 50% of the crime is commited by male, 45% by female and 5% by non-binary.

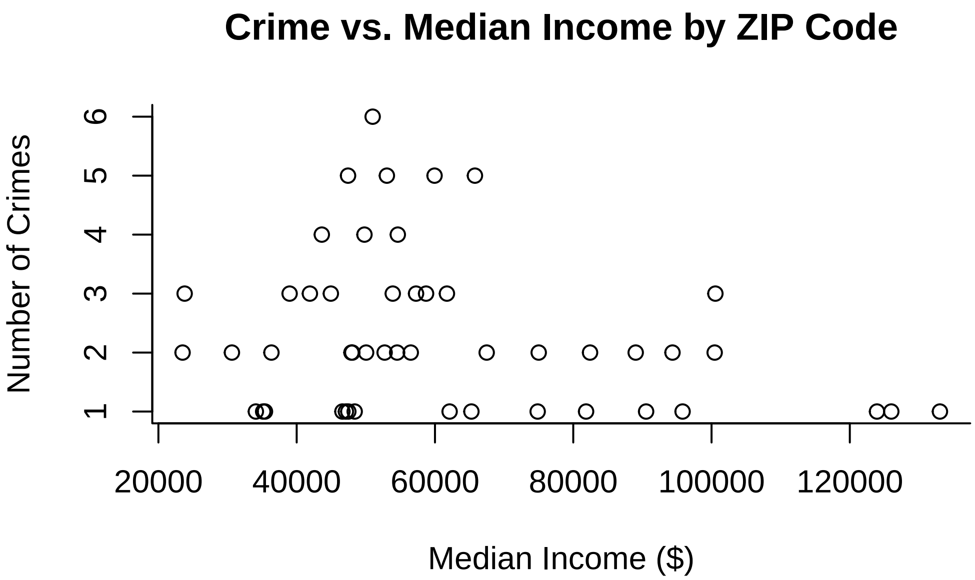
2.5 Median Household Income by Zip Code in Los Angeles by Year



2.5.1 Interpretation

The maps 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.

2.6 Crime Rates Compared to Median Household Income

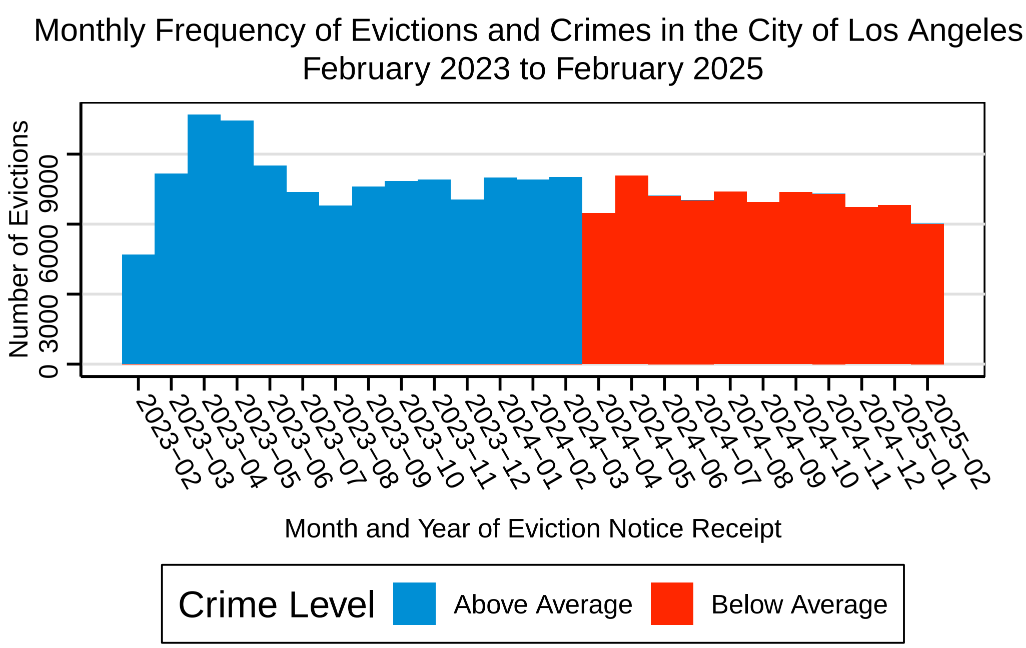


2.6.1 Interpretation

Based on this plot, we do not observe a clear trend between the number of crimes committed and the median income in that zip code. However, this scatterplot only accounts for a small portion of the dataset and averages across pre- and post-pandemic conditions. Future visualizations and analyses could consider pandemic vs non-pandemic years as variables.

3 Exploratory Data Analysis With ggplot2

3.1 Number of Evictions and Crime Levels Per Month



3.1.1 Interpretation

This histogram 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.

3.2 Boxplot - Kaitlyn

```
crime_by_area <- crime_la %>% group_by(AREA.NAME) %>% summarize(numberOfCrimes = n()) %>% ungroup()

bureaus <- read.csv("Raw_Data/Division_Bureau_Ref.csv")

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

crime_by_area <- crime_by_area %>% left_join(bureaus, join_by("AREA.NAME" == "Division"))
```

3.3 geom_tile for correlations - Bianca

3.4 geom_jitter

3.5 Plot 5

3.6 Plot 6

4 Regression Analyses

4.1 Regression to Examine Weekly Crime Rates

Table 1: Linear Regression Model Results Estimating Crimes Per Week

Predictor	B	SE	t	p
(Intercept)	711.40	86.455	8.23	$<$0.001
Index	-0.09	0.072	-1.25	0.210
CrimesLag	0.83	0.020	41.19	$<$0.001

4.1.1 Interpretation

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.

4.2 Regression Between Median Income and Crimes per Zip Code

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

Predictor	B	SE	t	p
(Intercept)	3.12	0.493	6.33	<0.001
cali_median_income	0.00	0.000	-1.78	0.082

4.2.1 Interpretation

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.

References

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

5 Appendices

Data Wrangling

6 Reading in Raw Datasets

```
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

# Primary datasets: Crime Data 2020-Present and 2010-2019
crime_la_2010_2019 <- read.csv("Raw_Data/Crime_Data_from_2010_to_2019.csv")

crime_la_2020_present <- read.csv("Raw_Data/Crime_Data_from_2020_to_Present.csv")

# Combine the two datasets
crime_la <- rbind(crime_la_2010_2019, crime_la_2020_present)
```

```
# Secondary datasets
california_median_income_2018_2023 <- list()
yrs <- 2018:2023
for (i in 1:length(yrs)) {
  obj <- read.csv(paste0("Raw_Data/ACSST5Y",yrs[[i]],".S1903-Data.csv"), skip = 0)
  obj <- obj[-1,]
  obj$year <- yrs[[i]]
  california_median_income_2018_2023[[i]] <- obj
}

cali_median_income <- do.call(rbind, california_median_income_2018_2023)

# adding a variable that's just the ZCTA zip code
cali_median_income$ZCTA <- stringr::str_sub(cali_median_income$NAME, -5)

write.csv(cali_median_income, "cali_median_income.csv")
```

Zip Codes for Crimes

```
library(sf)
library(dplyr)

crime_la_2010_2019 <- read.csv("Raw_Data/Crime_Data_from_2010_to_2019.csv")

crime_la_2020_present <- read.csv("Raw_Data/Crime_Data_from_2020_to_Present.csv")

# Combine the two datasets
crime_la <- rbind(crime_la_2010_2019, crime_la_2020_present)

crime_la$DateParsed <- as.POSIXct(crime_la$Date.Rptd,
```

```r
                                  format = "%m/%d/%Y %I:%M:%S %p",
                                  tz = "America/Los_Angeles")

crime_la$YearWeek <- format(crime_la$DateParsed, "%Y-%U")

crime_la$year <- as.integer(stringr::str_sub(crime_la$YearWeek, 1, 4))

coords <- crime_la %>% select(LAT, LON, DR_NO, year)

zcta <- read_sf(dsn = "Raw_Data/cb_2020_us_zcta520_500k", layer = "cb_2020_us_zcta520_500k")

  points <- st_as_sf(coords, coords = c("LON", "LAT"), crs = 4326)  # WGS84

  zcta <- st_transform(zcta, 4326)  # Convert ZCTA to WGS84 if needed
  cat("Starting Zip Code Retrieval \n")

  result <- st_join(points, zcta)

  cat("Finished Zip Code Retrieval \n")

  write.csv(result, "crime_zip_files/all_zips.csv")

crime_zctas <- read.csv("Raw_Data/all_zips.csv")
colnames(crime_zctas) <- c(colnames(crime_zctas)[-c(1, length(colnames(crime_zctas)))],"LON", "LAT")

crime_zctas$LON <- as.numeric(gsub('c(', "", crime_zctas$LON, fixed = TRUE))
crime_zctas$LAT <- as.numeric(gsub(')', "", crime_zctas$LAT, fixed = TRUE))
```

```r
duplicates <- as.integer(as.character(data.frame(table(crime_zctas$universal_idx)[table(crime_zctas$uni

# there are approximately 200 rows that are duplicated (100 incidents) as the area lies on the intersec
duplicate_data <- crime_zctas %>% filter(universal_idx %in% duplicates)

rev_geo_duplicates <- reverse_geocode(duplicates2, lat = LAT, long = LON)
# the above query identifies the address of these coordinates as KFC, 8644, Balboa Boulevard, Northridg
# therefore, we will only keep the 91325 zip code rows.

crime_zctas_no_duplicates <- crime_zctas %>% mutate(duplicate = ifelse(universal_idx %in% duplicate_dat

# it's empty now
# duplicates <- table(crime_zctas_no_duplicates$universal_idx)[table(crime_zctas_no_duplicates$universa

# next, check for NAs
sum(is.na(crime_zctas_no_duplicates$ZCTA5CE20)) # there are 3724 NAs. Let's find them a zip code

# some also have lat long = 0. We'll find those separately
no_zipcode_not0 <- crime_zctas_no_duplicates %>% filter(is.na(ZCTA5CE20) & LON != 0)

no_zipcode_not0_zips <- reverse_geocode(no_zipcode_not0, lat = LAT, long = LON)

no_zipcode_not0_zips$ZCTA5CE20 <- stringr::str_extract(no_zipcode_not0_zips$address, "\\d{5}")

no_zipcode_not0_zips <- no_zipcode_not0_zips %>% select(-address)
```

```r
crime_zctas2 <- crime_zctas_no_duplicates %>% filter(!is.na(ZCTA5CE20))

crime_zctas3 <- rbind(crime_zctas2, no_zipcode_not0_zips)

crime_zctas4 <- crime_zctas3 %>% filter(!is.na(ZCTA5CE20))

write.csv(crime_zctas4, "crimes_with_zips.csv")

no_zipcode_0s <- crime_zctas_no_duplicates %>% filter(is.na(ZCTA5CE20) & LON == 0) %>% pull(DR_NO)

main_dataset_0s <- crime_la %>% filter(DR_NO %in% no_zipcode_0s)

summary(main_dataset_0s)

summary(crime_la)
```

Figure 1 - Examining Crime Density from 2010–2019 and 2020-Present

```r
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
)

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
)

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

Figure 2 - Examining Number Of Crimes Over Time

```
crime_la$DateParsed <- as.POSIXct(crime_la$Date.Rptd,
                                  format = "%m/%d/%Y %I:%M:%S %p",
                                  tz = "America/Los_Angeles")

crime_la$YearWeek <- format(crime_la$DateParsed, "%Y-%U")

weekly_counts <- aggregate(DR_NO ~ YearWeek, data = crime_la, FUN = length)
names(weekly_counts)[2] <- "Crimes"

weekly_counts <- weekly_counts[order(weekly_counts$YearWeek), ]

weekly_counts$Index <- seq_len(nrow(weekly_counts))

plot(
  weekly_counts$Index,
  weekly_counts$Crimes,
  type = "l",
  col  = "blue",
  lwd  = 2,
  xlab = "Weekly Index",
  ylab = "Number of Crimes",
  main = "Weekly Crime Counts in Los Angeles (2010-Present)"
)

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

Figure 3 - Crime rate frequency by day of the week

```
# Convert Date.Rptd to Date format (ignore time part)
#crime_la$Date.Rptd
crime_la$Date <- as.Date(substr(crime_la$Date.Rptd, 1, 10), format = "%m/%d/%Y")

# Set locale to English (for weekday names)
Sys.setlocale("LC_TIME", "C")

# Extract the day of the week
crime_la$DayOfWeek <- weekdays(crime_la$Date)
# 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 <- crime_counts[day_order]
```

```
# Plot bar chart with horizontal labels
barplot(crime_counts, col = "lightblue",
        xlab = "Day of the Week", ylab = "Number of Crimes",
        names.arg = names(crime_counts), las = 1, cex.names = 0.8, cex.axis = 0.6)
```

Figure 4 - Gender Distribution for crimes over time

```
# Remove missing values and empty strings in Vict.Sex column
crime_la <- subset(crime_la, Vict.Sex != "" & !is.na(Vict.Sex))

# Replace all non-"F" and non-"M" values with "N"
crime_la$Vict.Sex <- ifelse(crime_la$Vict.Sex %in% c("F", "M"),
                                  crime_la$Vict.Sex, "N")

# Count occurrences of each gender
gender_counts <- table(crime_la$Vict.Sex)

# Create a pie chart using base R
color <- c("F" = "pink", "M" = "lightblue", "N" = "purple")
pie(
  gender_counts,
  labels = paste(names(gender_counts),
                 round(gender_counts / sum(gender_counts) * 100, 1), "%"),
  col = color[names(gender_counts)],
  main = "Distribution of Crime Victims by Gender"
)
```

Figure 5 - Median Household Income by Zip Code in Los Angeles by Year

```
zcta <- read_sf(dsn = "Raw_Data/cb_2020_us_zcta520_500k", layer = "cb_2020_us_zcta520_500k")
LA_zctas <- read.csv("Raw_Data/zctas.csv") #obtained using zipcodeR::search_city("Los Angeles", "CA") a

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
median_household_income <- cali_median_income %>% select(ZCTA, S1903_C03_001E, year)

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)
```

```
# 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, "Greens")

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 =
}

# main title for all plots
mtext("Median Household Income by Year and Zip Code in LA", outer = TRUE, cex = 1, font = 2)
```

Figure 6 - Crime Rates Compared to Median Household Income

```
# using latitude and longitude to obtain the zip codes of the crime locations
# crime_la$year <- as.integer(stringr::str_sub(crime_la$YearWeek, 1, 4))
# id_lat_long_crime_la <- subset(crime_la, select = c("LAT", "LON", "DR_NO", "year"))

# taking a random sample to reduce load on model
#set.seed(123)
#sampled_ids <- id_lat_long_crime_la[sample(1:nrow(id_lat_long_crime_la), 500),]
#rownames(sampled_ids) <- NULL
#zip_codes <- reverse_geocode(sampled_ids, lat = LAT, long = LON, verbose = TRUE)
#write.csv(zip_codes, "sampled_crimes.csv")
zip_codes <- read.csv("sampled_crimes.csv")

zip_codes$ZCTA5CE20 <- stringr::str_extract(zip_codes$address, "\\d{5}")
```

```r
crime_income_sample <- zip_codes %>% inner_join(unique(subset(zcta_no_NAs, select = c("ZCTA5CE20", "S19

# this takes a while, so loading in this csv instead

crime_income_sample <- zip_codes %>% inner_join(unique(subset(zcta_no_NAs, select = c("ZCTA5CE20", "S19

crime_income_per_zip <- crime_income_sample %>%
  group_by(ZCTA5CE20) %>%
  summarize(number_of_crimes = n(),
            cali_median_income = mean(S1903_C03_001E)) %>%
  ungroup()

plot(x = crime_income_per_zip$cali_median_income,
     y = crime_income_per_zip$number_of_crimes,
     type = "p",
     ylab = "Number of Crimes",
     xlab = "Median Income ($)",
     main = "Crime vs. Median Income by ZIP Code",
     bty = "l")
```

Crime ~ Week Regression Code

```r
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)

ar1_trend_regression %>%
tidy() %>%
  mutate(
    p.value = scales::pvalue(p.value)
  ) %>%
  knitr::kable(
    caption = "Linear Regression Model Results Estimating Crimes Per Week",
    col.names = c("Predictor", "B", "SE", "t", "p"),
    digits = c(0, 2, 3, 2, 3),
    align = c("l", "r", "r", "r", "r")
  )
```

Crime ~ Income Regression Code

```r
lm_income_crimes <- lm(number_of_crimes ~ cali_median_income, data = crime_income_per_zip)

summary(lm_income_crimes)

lm_income_crimes %>%
tidy() %>%
  mutate(
```

```
    p.value = scales::pvalue(p.value)
  ) %>%
  knitr::kable(
    caption = "Linear Regression Model Results Estimating Crimes Based on Median Income",
    col.names = c("Predictor", "B", "SE", "t", "p"),
    digits = c(0, 2, 3, 2, 3),
    align = c("l", "r", "r", "r", "r")
  )
```