Crime Rate Analysis For New York and Los Angeles

Ariana Netteville

Table of contents

- Introduction
- Primary Data
- Secondary Data
- Attributes
- Data Cleaning
- Homicides in Los Angeles from 1991 to 2003
- Top 5 crimes in each city
- Rates of Crimes
- Most crime
- Conclusion
- Sources

Introduction

This project contains information about crime in two distinct cities, New York City and Los Angeles. Both these cities receive high levels of crime on yearly basis. We will dive further into this topic by exploring and analyzing data obtained from an official government website to see how we can measure these levels of crime to answer our research questions. We will explore the following topics using data wrangling techniques and analysis:

- What are the leading crimes in these cities? (Crime occurs more often in cities with high populations like New York, Los Angeles, etc. Finding out what is the most common crime)

- Which city has the most crime? (This gives us a look into why a certain city has more crime than the other. Does the city with the highest population have more crime due to the fact that there's more people or does it have to do with the police system?)
- What does the progression (rate) of the top crime look like between certain years by city? Is it decreasing, increasing, constant, etc.?

Primary Data

The LAPD crime data from the years 2000 to 2025 was found the data.gov website. This data set was collected by the Los Angeles Police Department's Records Management System. Each case in this data set are crimes reported in the city.

```

Attaching package: 'dplyr'


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

Secondary Data

NYPD Data Set

- The NYPD contains crime information in New York City from 2006 to 2024. This data set was found on data.gov and the data was collected by the New York Police Department's Office of Management Analysis and Planning. Each case in this data set are crimes reported in the city.

Los Angeles Homicides per year

- Homicides in Los Angeles was found on Wikipedia. This data set contains number of homicides in the city from 1991 to 2023.

Attributes

For each data set, our main focus will be on the year and crime description.

Data Cleaning

Both NYPD and LAPD data sets format their dates as follows : mm/dd/yyyy. Los Angeles also includes time stamps along with the dates. To fix this, the functions substr and as.numeric will be used to only show the years these crimes occurred.

New York:

```
  ARREST_KEY ARREST_DATE PD_CD           PD_DESC KY_CD      OFNS_DESC
1  279197226        2023   105 STRANGULATION 1ST   106 FELONY ASSAULT
2  278761840        2023   105 STRANGULATION 1ST   106 FELONY ASSAULT
3  278506761        2023   153            RAPE 3   104           RAPE
    LAW_CODE LAW_CAT_CD ARREST_BORO ARREST_PRECINCT JURISDICTION_CODE AGE_GROUP
1 PL 1211200          F           M              18                 0     25-44
2 PL 1211300          F           K              67                 0     25-44
3 PL 1302503          F           K              77                 0     25-44
  PERP_SEX PERP_RACE X_COORD_CD Y_COORD_CD Latitude Longitude
1        M     WHITE     988210     218129 40.76539 -73.98570
2        M     BLACK     997897     175676 40.64886 -73.95082
3        M     BLACK    1003509     185018 40.67450 -73.93057
                                     Lon_Lat
1                POINT (-73.985702 40.76539)
2                POINT (-73.95082 40.648859)
3 POINT (-73.9305713255961 40.6744956865259)
```

Los Angeles:

```
      DR_NO              Date.Rptd DATE.OCC TIME.OCC AREA   AREA.NAME
1 211507896 04/11/2021 12:00:00 AM     2020      845   15 N Hollywood
2 201516622 10/21/2020 12:00:00 AM     2020     1845   15 N Hollywood
3 240913563 12/10/2024 12:00:00 AM     2020     1240    9    Van Nuys
  Rpt.Dist.No Part.1.2 Crm.Cd                                    Crm.Cd.Desc
1        1502        2    354                              THEFT OF IDENTITY
2        1521        1    230 ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
3         933        2    354                              THEFT OF IDENTITY
                                  Mocodes Vict.Age Vict.Sex Vict.Descent
1                                    0377       31        M            H
2 0416 0334 2004 1822 1414 0305 0319 0400       32        M            H
3                                    0377       30        M            W
  Premis.Cd            Premis.Desc Weapon.Used.Cd
1       501 SINGLE FAMILY DWELLING             NA
```

```
2       102               SIDEWALK            200
3       501 SINGLE FAMILY DWELLING             NA
                       Weapon.Desc Status Status.Desc Crm.Cd.1 Crm.Cd.2
1                                      IC Invest Cont      354       NA
2 KNIFE WITH BLADE 6INCHES OR LESS     IC Invest Cont      230       NA
3                                      IC Invest Cont      354       NA
  Crm.Cd.3 Crm.Cd.4                             LOCATION Cross.Street
1       NA       NA  7800    BEEMAN                        AV
2       NA       NA          ATOLL                         AV     N  GAULT
3       NA       NA 14600    SYLVAN                        ST
      LAT       LON
1 34.2124 -118.4092
2 34.1993 -118.4203
3 34.1847 -118.4509
```

As we can see from the two tables above, both data sets now only contain the year rather than the full date. This will allow for easier data wrangling and analysis.

The Homicide data, obtained from Wikipedia contains footnotes for each of the murder counts. Similarly to the previous data set, substr and as.numeric will be used to only show the numbers without footnotes.

```
# A tibble: 13 x 2
    Year Murders
   <int>   <dbl>
 1  1991    1025
 2  1992    1092
 3  1993    1077
 4  1994     850
 5  1995     838
 6  1996     709
 7  1997     576
 8  1998     426
 9  1999     425
10  2000     550
11  2001     588
12  2002     654
13  2003     515
```

Homicides in Los Angeles from 1991 to 2003

Using the homicide data from Wikipedia, we can determine the rate of murders in Los Angeles through 1991 to 2003 using a simple line plot.

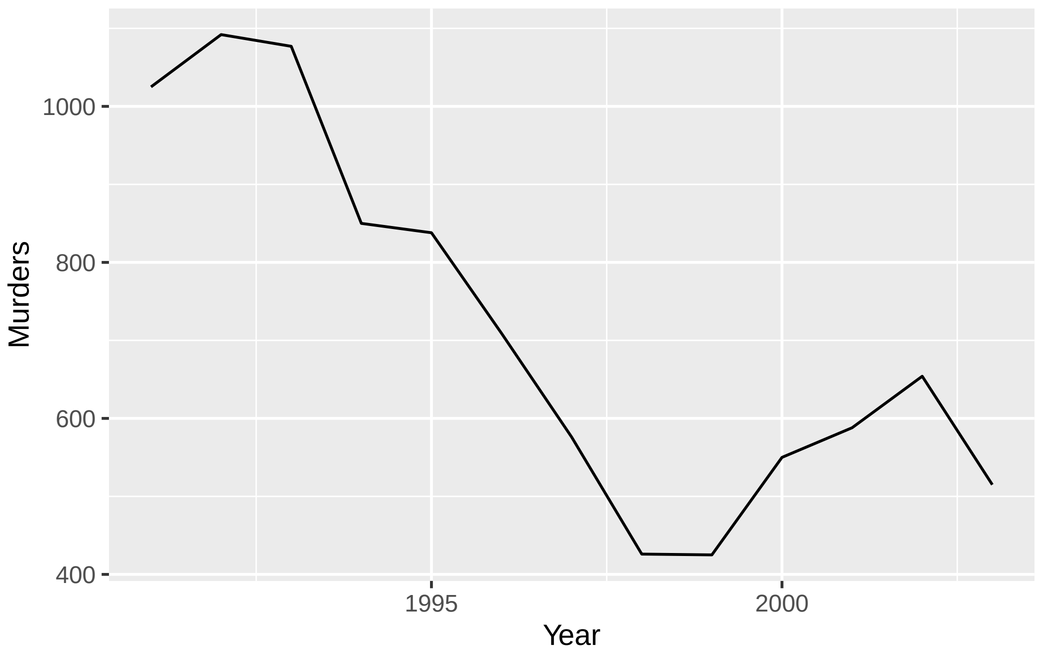

Figure 1: Homicide Rates

Using our homicide data, we can in Figure 1 a steady decline in homicides from 1992 to 1999 with a small increase from 2000 to 2002. Overall, the number of homicides has decreased significantly from the year 1991.

Top 5 crimes in each city

To determine the top crimes in both New York and Los Angeles, we can apply data wrangling techniques to group each crime and count the total number of times that specific crime appears throughout the entire data set. For this project, we will only look at the top 5 crimes.

```
# A tibble: 5 x 2
  Crm.Cd.Desc                                            Total
  <chr>                                                  <int>
1 VEHICLE - STOLEN                                      115190
2 BATTERY - SIMPLE ASSAULT                               74839
3 BURGLARY FROM VEHICLE                                  63517
4 THEFT OF IDENTITY                                      62538
5 VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)  61092
```

According the table, The Los Angeles police department reports that top 5 crimes in the city from the years 2020 to 2023 are Vehicle robberies, Battery, Burglary, Identity theft and vandalism.

```
# A tibble: 5 x 2
  OFNS_DESC                         Total
  <chr>                             <int>
1 DANGEROUS DRUGS                 1144059
2 ASSAULT 3 & RELATED OFFENSES     645227
3 OTHER OFFENSES RELATED TO THEFT  316812
4 PETIT LARCENY                    306203
5 FELONY ASSAULT                   288434
```

According to the table, The New York Police department reports that the top 5 crimes in the city from the years 2021 to 2023 are Drugs, Assault, Other offenses related to theft, Petit Larceny and Felony Assault.

We can see that both cities have some form of assault and theft as some of their top leading crimes.

Rates of Crimes

To determine the rate of crime over the years 2020 to 2024 we will first have to create a new data frame for each state with the same column names and same years. Since New York includes data dating back to 2006, we must filter out the years that are not included in the Los Angeles data set. Furthermore, we can reduce the columns of each data set by only selecting the years and the crime description columns. We will eventually have to combine the two data sets together so it is necessary to include a column with the name of the states so it will be easier to differentiate each state in the plot.

```
`summarise()` has grouped output by 'year'. You can override using the
`.groups` argument.
```

According to the graph in Figure 2 , The rates of crime in Los Angeles starts a steady decrease in 2022 and then rapidly decreases from 2023 to 2024. Meanwhile, New York increases through 2020 to 2024. Additionally, we can notice that Los Angeles has a higher crime count from 2020 to 2023. Does this mean that Los Angeles might have the most crime? We can investigate this further by looking at a side by side bar plot.

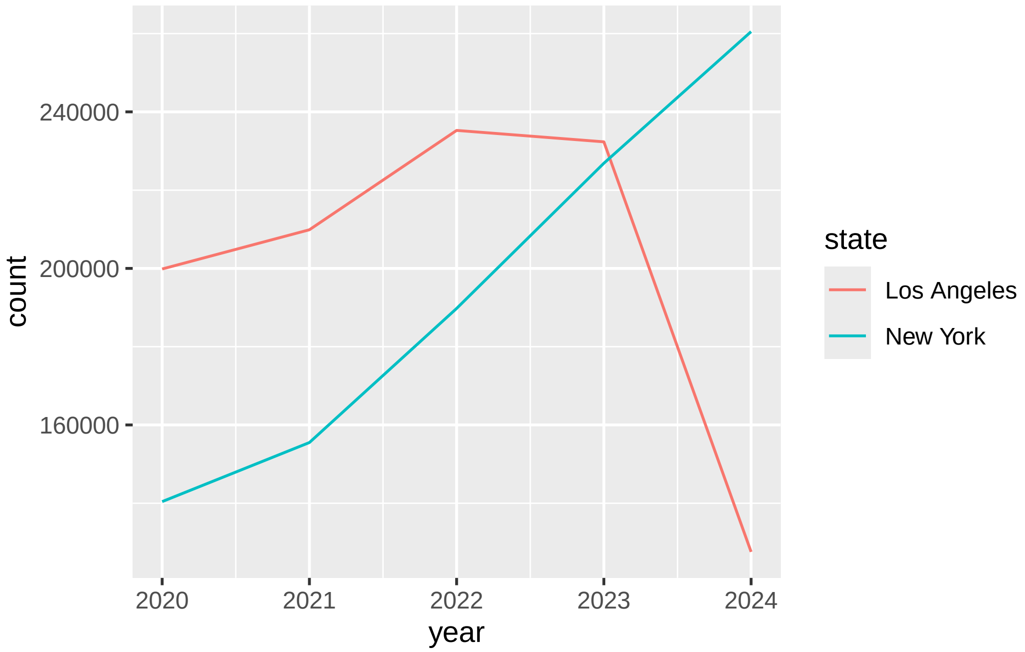

Figure 2: State Crime rates per year

Most crime

We can determine which state has more crime by counting the total number of crimes reported in each state. Since both cities have a large population of people it is expected to have similar crimes numbers in each city.

As we can see from the bar graph in Figure 3, Los Angeles appears to have slightly more crime than New York.

```
# A tibble: 2 x 2
  state         Total
  <chr>         <int>
1 Los Angeles 1004892
2 New York     973069
```

This table confirms that Los Angeles has 31,823 more reported crimes than New York.

Conclusion

After analyzing and comparing crime rates in both New York and Los Angeles, we have discovered several key conclusions. Theft and assault are among the most common crimes in

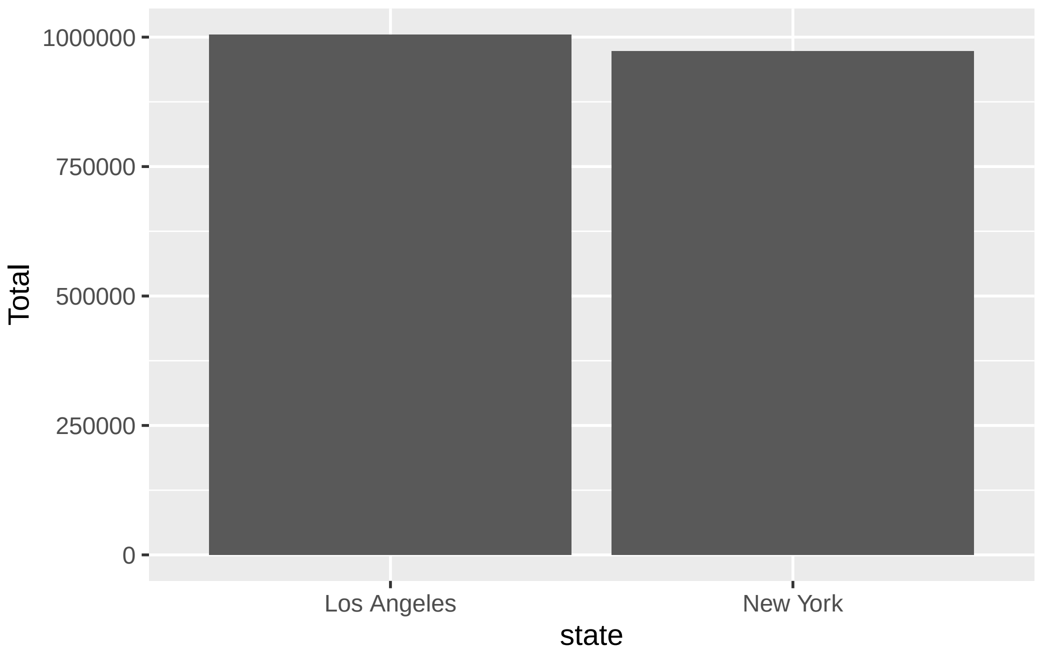

Figure 3: Total crimes by state

both cities, indicating a higher need for prevention against those crimes. This project has also revealed that the city with the most reported crimes is Los Angeles as shown in Figure 3. However, Figure 2 shows that Los Angeles has had a decreasing rate in crime over 2 years while New York crime levels has continue to rise over the past 4 years. We can also see this decreasing pattern in homicide rates in Los Angeles, as seen in Figure 1. In recent years, the homicide rates have almost halved the rates in the early 90's.

Sources

NYPD source -> https://catalog.data.gov/dataset/nypd-arrest-data-year-to-date

LAPD source -> https://catalog.data.gov/dataset/crime-data-from-2020-to-present

LA Homicide Source -> https://en.wikipedia.org/wiki/Crime_in_Los_Angeles

Basic bar plot with ggplot2 source -> https://r-graph-gallery.com/218-basic-barplots-with-ggplot2.html

```
#google's R style
library(tidyr)
library(dplyr)
library(rvest)
```

```r
library(ggplot2)
NYPD_data <- read.csv("~/Downloads/NYPD_Arrests_Data__Historic_.csv")
LAPD_data <- read.csv("~/Downloads/Crime_Data_from_2020_to_Present (1).csv")
URL <- "https://en.wikipedia.org/wiki/Crime_in_Los_Angeles"
ListOfTables <- URL %>%
  read_html() %>%
  html_nodes(css = "table") %>%
  html_table(fill = TRUE)
Homicide_data <- ListOfTables[[4]]

dates <- NYPD_data$ARREST_DATE
NYPD_data$ARREST_DATE <- as.numeric(substr(dates, nchar(dates) - 4 + 1, nchar(dates)))
NYPD_data %>%
  head(3)
dates_la <- LAPD_data$DATE.OCC
LAPD_data$DATE.OCC <- as.numeric(substr(dates_la, 7 , 10))
LAPD_data %>%
  head(3)
murders <- Homicide_data$Murders
Homicide_data$Murders <- as.numeric(substr(murders, 1 , nchar(murders)-4))
Homicide_data
#Homicides per year in Los Angeles
ggplot(data = Homicide_data, aes(x= Year, y = Murders)) +
  geom_line()
LAPD_data_totalCrimes <- LAPD_data %>%
  group_by(Crm.Cd.Desc)%>%
  summarise(Total = n())%>%
  arrange(desc(Total))
LAPD_data_totalCrimes %>%
  head(5)
NYPD_data_totalCrimes <- NYPD_data %>%
  group_by(OFNS_DESC) %>%
  summarise(Total= n())%>%
  arrange(desc(Total))
NYPD_data_totalCrimes %>%
  head(5)
new_NYPD <- NYPD_data %>%
  select(ARREST_DATE, OFNS_DESC)%>%
  filter(ARREST_DATE > 2019) %>%
  mutate(state = "New York")%>%
  rename(year = ARREST_DATE, crime = OFNS_DESC)
new_LAPD <- LAPD_data %>%
  select(DATE.OCC, Crm.Cd.Desc) %>%
  filter(DATE.OCC < 2025) %>%
  mutate(state = "Los Angeles") %>%
  rename(year = DATE.OCC, crime = Crm.Cd.Desc)
state_crimes <- bind_rows(new_LAPD, new_NYPD)
state_crimes_perYear <- state_crimes %>%
  group_by(year, state) %>%
  summarise(count = n())
ggplot(data = state_crimes_perYear, aes(x=year, y =count, color = state)) +
  geom_line()
state_crimes_grouped <- state_crimes %>%
  group_by(state)%>%
  summarise(Total = n())
ggplot(data = state_crimes_grouped, aes(x= state, y=Total)) +
  geom_bar(stat = "identity")
state_crimes_grouped
```