

final_181

Sarah Mirza

2024-10-24

Violent and Nonviolent Crime Graphs

First, read in the CSV file. Reading in the CSV file is in its own cell as it is a relatively large file and reducing the amount of times it must be read in can significantly reduce time, storage, and cost.

```
# read in CSV as data frame
crime_data <- read.csv("/Users/sarahmirza/Documents/GitHub/LA-Crime-Data/Crime_Data_from_2020_to_Present.csv")
```

Preview the data in its own cell to avoid excessively reading in the CSV. Going forward we will use the data frame to do our further analyses.

```
# preview the crime data frame
head(crime_data)
```

```
##      DR_NO      Date.Rptd      DATE.OCC TIME.OCC AREA
## 1 190326475 03/01/2020 12:00:00 AM 03/01/2020 12:00:00 AM      2130      7
## 2 200106753 02/09/2020 12:00:00 AM 02/08/2020 12:00:00 AM      1800      1
## 3 200320258 11/11/2020 12:00:00 AM 11/04/2020 12:00:00 AM      1700      3
## 4 200907217 05/10/2023 12:00:00 AM 03/10/2020 12:00:00 AM      2037      9
## 5 220614831 08/18/2022 12:00:00 AM 08/17/2020 12:00:00 AM      1200      6
## 6 231808869 04/04/2023 12:00:00 AM 12/01/2020 12:00:00 AM      2300     18
## AREA.NAME Rpt.Dist.No Part.1.2 Crm.Cd
## 1 Wilshire      784      1      510
## 2 Central      182      1      330
## 3 Southwest    356      1      480
## 4 Van Nuys     964      1      343
## 5 Hollywood   666      2      354
## 6 Southeast   1826     2      354
## Crm.Cd.Desc      Mocodes Vict.Age
## 1 VEHICLE - STOLEN      1822 1402 0344      47
## 2 BURGLARY FROM VEHICLE      0344 1251      19
## 3 BIKE - STOLEN      0325 1501      19
## 4 SHOPLIFTING-GRAND THEFT ($950.01 & OVER)      1822 1501 0930 2004      28
## 5 THEFT OF IDENTITY      1822 0100 0930 0929      41
## Vict.Sex Vict.Descent Premis.Cd      Premis.Desc
## 1 M      0      101      STREET
## 2 M      0      128      BUS STOP/LAYOVER (ALSO QUERY 124)
## 3 X      X      502 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)
## 4 M      0      405      CLOTHING STORE
## 5 M      H      102      SIDEWALK
## 6 M      H      501      SINGLE FAMILY DWELLING
## Weapon.Used.Cd Weapon.Desc Status.Desc Crm.Cd.1 Crm.Cd.2 Crm.Cd.3
## 1 NA      AA Adult Arrest      510      998      NA
```

```
## 2      NA      IC Invest Cont      330      998      NA
## 3      NA      IC Invest Cont      480      NA      NA
## 4      NA      IC Invest Cont      343      NA      NA
## 5      NA      IC Invest Cont      354      NA      NA
## 6      NA      IC Invest Cont      354      NA      NA
## Crm.Cd.4      LOCATION Cross.Street      LAT
## 1      NA 1900 S LONGWOOD      AV      34.0375
## 2      NA 1000 S FLOWER      ST      34.0444
## 3      NA 1400 W 37TH      ST      34.0210
## 4      NA 14000 RIVERSIDE      DR      34.1576
## 5      NA      1900 TRANSIENT      34.0944
## 6      NA 9900 COMPTON      AV      33.9467
## LON
## 1 -118.3506
## 2 -118.2628
## 3 -118.3002
## 4 -118.4387
## 5 -118.3277
## 6 -118.2463
```

Load in the required plotting library, ggplot2.

```
library(ggplot2)
```

Create a frequency table to see which areas have the highest rates of crime. Using base R, xtabs will count the frequency of how many times each area has had a specific primary crime code. To do this, select the columns you want using a tilde followed by a + to cross reference the columns, then after the comma, provide the data frame you would like to use. For ease of use, typecast the xtabs table to a data frame for viewing and further analyses.

```
#hist(crime_data$TIME.OCC)

#hist(crime_data$Crm.Cd)

# data frame containing the frequency of each crime committed per area
crime_by_region <- data.frame(xtabs(~AREA.NAME+Crm.Cd, crime_data))

head(crime_by_region) # preview data frame
```

```
## AREA.NAME Crm.Cd Freq
## 1 77th Street 110 237
## 2 Central 110 102
## 3 Devonshire 110 29
## 4 Foothill 110 41
## 5 Harbor 110 80
## 6 Hollenbeck 110 144
```

Classify categories of non-violent crime and analyze their geographic distribution to identify regions in Los Angeles with the highest rates of these crimes. Create a list of violent crime codes as determined by the Los Angeles Police Department.

```
violent_crime_codes <- c(110,113,121,122,815,820,821,210,220,230,231,235,236,250,251,761,926,435,436,437)
```

Go through the crime_data data frame and determine whether the primary crime code is found in the violent crimes list, or not, and add the crime data row to new, respective data frames-one for violent crimes and one for nonviolent crimes.

```
# if the main crime code matches a value in the violent crimes list, add it to violent crimes data frame
violent_crimes <- crime_data[crime_data$Crm.Cd %in% violent_crime_codes,]
head(violent_crimes)
```

```
##          DR_NO          Date.Rptd          DATE.OCC TIME.OCC AREA
## 10 211904005 12/31/2020 12:00:00 AM 12/31/2020 12:00:00 AM      1220  19
## 12 221908151 04/12/2022 12:00:00 AM 10/01/2020 12:00:00 AM         1  19
## 18 210705560 11/27/2020 12:00:00 AM 11/27/2020 12:00:00 AM      1800   7
## 26 221105176 02/02/2022 12:00:00 AM 02/09/2020 12:00:00 AM      1200  11
## 30 220808837 04/28/2022 12:00:00 AM 03/30/2020 12:00:00 AM      1630   8
## 36 211220472 09/27/2021 12:00:00 AM 11/26/2020 12:00:00 AM      2000  12
##          AREA.NAME Rpt.Dist.No Part.1.2 Crm.Cd
## 10      Mission          1974         2    624
## 12      Mission          1988         1    821
## 18      Wilshire          776         1    230
## 26      Northeast        1132         2    930
## 30      West LA           842         2    624
## 36 77th Street          1259         2    930
##
##                                Crm.Cd.Desc
## 10                                BATTERY - SIMPLE ASSAULT
## 12 SODOMY/SEXUAL CONTACT B/W PENIS OF ONE PERS TO ANUS OTH
## 18                                ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
## 26                                CRIMINAL THREATS - NO WEAPON DISPLAYED
## 30                                BATTERY - SIMPLE ASSAULT
## 36                                CRIMINAL THREATS - NO WEAPON DISPLAYED
##
##          Mocodes Vict.Age Vict.Sex Vict.Descent Premis.Cd
## 10          0416         26         M             H        502
## 12 0913 2024 1817 0360 1258 0507         8         F             H        501
## 18          1309 0400        31         F             O        101
## 26          1912 0913        30         F             W        501
## 30          2000 1813 0913 0416        24         F             O        501
## 36 0913 0400 0443 1814 2000        29         F             H        502
##
##          Premis.Desc Weapon.Used.Cd
## 10 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)        400
## 12          SINGLE FAMILY DWELLING        400
## 18          STREET        307
## 26          SINGLE FAMILY DWELLING        500
## 30          SINGLE FAMILY DWELLING        400
## 36 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)        511
##
##          Weapon.Desc Status Status.Desc Crm.Cd.1
## 10 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)      IC Invest Cont        624
## 12 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)      IC Invest Cont        812
## 18          VEHICLE      AA Adult Arrest        230
## 26          UNKNOWN WEAPON/OTHER WEAPON      AO Adult Other        930
## 30 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)      IC Invest Cont        624
## 36          VERBAL THREAT      AO Adult Other        930
##
##          Crm.Cd.2 Crm.Cd.3 Crm.Cd.4          LOCATION
## 10          NA          NA          NA 9000      CEDROS          AV
## 12          821          NA          NA 13400     RANGOON          ST
## 18          NA          NA          NA 4500      LOMITA          ST
## 26          NA          NA          NA 2800      WAVERLY          DR
## 30          NA          NA          NA 1200 S     WESTGATE          AV
## 36          NA          NA          NA 7800 S     SAN PEDRO          ST
##          Cross.Street      LAT      LON
```

```
## 10          34.2336 -118.4535
## 12          34.2285 -118.4258
## 18          34.0452 -118.3351
## 26          34.1084 -118.2639
## 30          34.0474 -118.4635
## 36          33.9679 -118.2695
```

```
# if the main crime code does not match a value in the violent crimes list, add it to nonviolent crimes
non_violent_crimes <- crime_data[!crime_data$Crm.Cd %in% violent_crime_codes,]
head(non_violent_crimes)
```

```
##          DR_NO          Date.Rptd          DATE.OCC TIME.OCC AREA
## 1 190326475 03/01/2020 12:00:00 AM 03/01/2020 12:00:00 AM    2130    7
## 2 200106753 02/09/2020 12:00:00 AM 02/08/2020 12:00:00 AM    1800    1
## 3 200320258 11/11/2020 12:00:00 AM 11/04/2020 12:00:00 AM    1700    3
## 4 200907217 05/10/2023 12:00:00 AM 03/10/2020 12:00:00 AM    2037    9
## 5 220614831 08/18/2022 12:00:00 AM 08/17/2020 12:00:00 AM    1200    6
## 6 231808869 04/04/2023 12:00:00 AM 12/01/2020 12:00:00 AM    2300   18
## AREA.NAME Rpt.Dist.No Part.1.2 Crm.Cd
## 1  Wilshire          784          1    510
## 2   Central          182          1    330
## 3 Southwest          356          1    480
## 4  Van Nuys          964          1    343
## 5 Hollywood          666          2    354
## 6 Southeast         1826          2    354
##          Crm.Cd.Desc          Mocodes Vict.Age
## 1          VEHICLE - STOLEN                      0
## 2      BURGLARY FROM VEHICLE      1822 1402 0344      47
## 3          BIKE - STOLEN          0344 1251      19
## 4 SHOPLIFTING-GRAND THEFT ($950.01 & OVER)      0325 1501      19
## 5          THEFT OF IDENTITY 1822 1501 0930 2004      28
## 6          THEFT OF IDENTITY 1822 0100 0930 0929      41
## Vict.Sex Vict.Descent Premis.Cd          Premis.Desc
## 1      M          0      101          STREET
## 2      M          0      128      BUS STOP/LAYOVER (ALSO QUERY 124)
## 3      X          X      502 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)
## 4      M          0      405          CLOTHING STORE
## 5      M          H      102          SIDEWALK
## 6      M          H      501      SINGLE FAMILY DWELLING
## Weapon.Used.Cd Weapon.Desc Status Status.Desc Crm.Cd.1 Crm.Cd.2 Crm.Cd.3
## 1      NA          AA Adult Arrest      510      998      NA
## 2      NA          IC Invest Cont      330      998      NA
## 3      NA          IC Invest Cont      480      NA      NA
## 4      NA          IC Invest Cont      343      NA      NA
## 5      NA          IC Invest Cont      354      NA      NA
## 6      NA          IC Invest Cont      354      NA      NA
## Crm.Cd.4          LOCATION Cross.Street          LAT
## 1      NA 1900 S LONGWOOD          AV      34.0375
## 2      NA 1000 S FLOWER          ST      34.0444
## 3      NA 1400 W 37TH          ST      34.0210
## 4      NA 14000 RIVERSIDE          DR      34.1576
## 5      NA          1900 TRANSIENT      34.0944
## 6      NA 9900 COMPTON          AV      33.9467
##          LON
## 1 -118.3506
```

```
## 2 -118.2628
## 3 -118.3002
## 4 -118.4387
## 5 -118.3277
## 6 -118.2463
```

Pie Charts

For both the nonviolent and violent data frames, create tables to track the number of occurrences in their respective data frames. Sort them, then plot them in a pie chart to show how much each area contributes to the types of crimes that occur in Los Angeles.

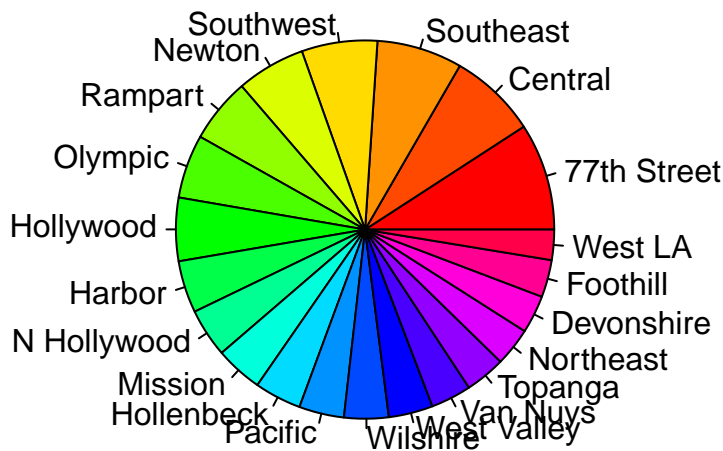
```
# pie chart to show violent crime distribution in LA
violent_crimes_table <- table(violent_crimes$AREA.NAME) # add all the area names from the violent crimes

violent_crimes_sorted <- sort(violent_crimes_table,decreasing=TRUE) # sort table by most crimes first

#View(violent_crimes_top10)

# display pie chart based on data
violent_pie <- pie(violent_crimes_sorted,
  main = "Crimes by Area",
  col = rainbow(length(violent_crimes_sorted)),
  labels = names(violent_crimes_sorted))
```

Crimes by Area



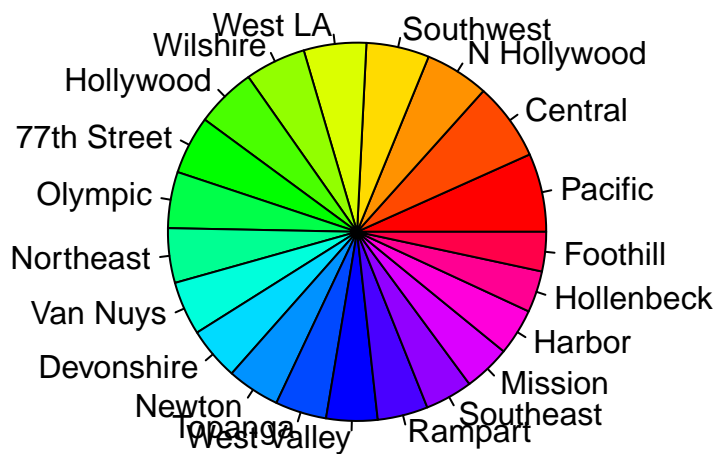
```
#head(violent_crimes_sorted)

# pie chart of non violent crimes
nonviolent_crimes_table <- table(non_violent_crimes$AREA.NAME) # add all area names from nonviolent crimes

nonviolent_crimes_sorted <- sort(nonviolent_crimes_table,decreasing=TRUE) # sort in decremting order

# create pie chart based on nonviolent crimes
nonviolent_pie <- pie(nonviolent_crimes_sorted,
  main = "Crimes by Area",
  col = rainbow(length(nonviolent_crimes_sorted)),
  labels = names(nonviolent_crimes_sorted))
```

Crimes by Area



```
#nonviolent_crimes_sorted
```

Frequency Distribution Graph

Use the dplyr library to filter the data by counting the frequency and ordering the data to have highest number of occurrences first. The data frames should now show how often each type of crime is committed per precinct. Add a new column to each data frame indicating the type of crime and negate the frequencies on the nonviolent data frame. Bind the data frames together. Sort the data frames so they appear in the order that they occurred in their respective data frames-sorted in decreasing frequency.

```
library(dplyr) # for filtering
```

```
##
## 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
# count the number of times each area occurs in the violent crime data frame
violent_by_area <- violent_crimes %>%
  count(AREA.NAME) %>%
  arrange(desc(n)) # sort by most crime

#violent_by_area

# count the number of times each area occurs in the nonviolent crime data frame
nonviolent_by_area <- non_violent_crimes %>%
  count(AREA.NAME) %>%
  arrange(desc(n)) # sort by highest crime rate

#nonviolent_by_area

# add new column to classify crime type for merging
```

```

violent_by_area$Type <- "Violent Crime"
nonviolent_by_area$Type <- "Nonviolent Crime"

# assign column names
colnames(violent_by_area) <- c("Area","Crimes","Type")
colnames(nonviolent_by_area) <- c("Area","Crimes","Type")

# negate all the nonviolent crime totals for flipped graph
nonviolent_by_area$Crimes <- -nonviolent_by_area$Crimes

# combine data frames
crimes_by_area <- rbind(violent_by_area,nonviolent_by_area)

#head(crimes_by_area)

#sort by the order they appear in their data frames (sorted by decreasing frequency)
crimes_by_area$Area <- factor(crimes_by_area$Area, levels = unique(crimes_by_area$Area))

head(crimes_by_area)

```

```

##           Area Crimes           Type
## 1 77th Street  25988 Violent Crime
## 2   Central  21260 Violent Crime
## 3 Southeast  20731 Violent Crime
## 4 Southwest  18352 Violent Crime
## 5   Newton  16714 Violent Crime
## 6   Rampart  15703 Violent Crime

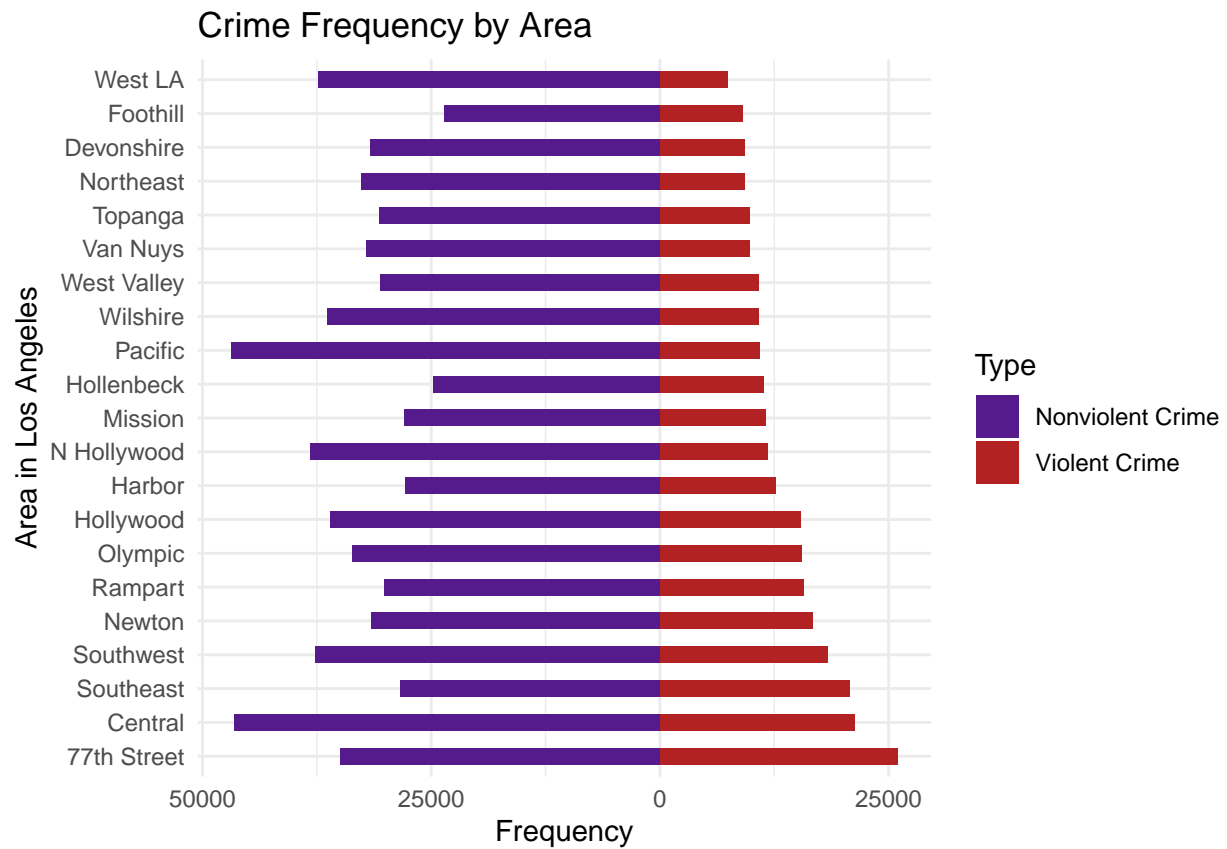
```

Create a bar chart flipped on its access to show the distribution of crimes. Nonviolent crime crime will be represented with negative values and the violent crimes will be represented with positive values. The crimes will be ordered by least violent crimes to most violent crimes from top to bottom. The y-axis will represent every area in LA, once the coordinates are flipped. Red will indicate violent crime and nonviolent will be in purple. Save the plot.

```

# crime frequency plot
crime_freq_plot <- ggplot(crimes_by_area, aes(x = Area, y = Crimes, fill = Type)) +
  geom_bar(stat = "identity",width=0.5) + # bar chart
  coord_flip() + # flip x and y axis
  labs(title = "Crime Frequency by Area",
       x = "Area in Los Angeles",
       y = "Frequency") +
  theme_minimal() +
  scale_y_continuous(labels = abs) +
  scale_fill_manual(values = c("Violent Crime" = "firebrick", "Nonviolent Crime" = "purple4"))
crime_freq_plot

```



```
ggsave("CrimeFrequency.png",plot=crime_freq_plot) # save plot
```

```
## Saving 6.5 x 4.5 in image
```