

Crime

Corey Michaud

2023-09-24

The data in this notebook was taken from <https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8> and consists of crime data from Los Angeles, CA (2020-2023).

All visualizations, statistics, and insights discussed are based on this dataset and time frame mentioned above only. The objective is to gain a comprehensive understanding of crime in Los Angeles through data exploration and visualization.

Loading Libraries

```
# Libraries
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lubridate)
```

Exploratory Data Analysis

```
# Importing dataset
crime <- read.csv("crime.csv")
head(crime)
```

```
##      DR_NO      Date.Rptd      DATE.OCC TIME.OCC AREA
## 1  10304468 01/08/2020 12:00:00 AM 01/08/2020 12:00:00 AM    2230    3
## 2  190101086 01/02/2020 12:00:00 AM 01/01/2020 12:00:00 AM     330    1
## 3  200110444 04/14/2020 12:00:00 AM 02/13/2020 12:00:00 AM    1200    1
## 4  191501505 01/01/2020 12:00:00 AM 01/01/2020 12:00:00 AM    1730   15
## 5  191921269 01/01/2020 12:00:00 AM 01/01/2020 12:00:00 AM     415   19
## 6  200100501 01/02/2020 12:00:00 AM 01/01/2020 12:00:00 AM      30    1
##      AREA.NAME Rpt.Dist.No Part.1.2 Crm.Cd
## 1   Southwest      377      2      624
## 2    Central      163      2      624
## 3    Central      155      2      845
```

```

## 4 N Hollywood          1543          2      745
## 5      Mission          1998          2      740
## 6      Central          163           1      121
##
##                               Crm.Cd.Desc                               Mocodes
## 1                               BATTERY - SIMPLE ASSAULT                0444 0913
## 2                               BATTERY - SIMPLE ASSAULT                0416 1822 1414
## 3                               SEX OFFENDER REGISTRANT OUT OF COMPLIANCE 1501
## 4                               VANDALISM - MISDEAMEANOR ($399 OR UNDER) 0329 1402
## 5 VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)              0329
## 6                               RAPE, FORCIBLE 0413 1822 1262 1415
## Vict.Age Vict.Sex Vict.Descent Premis.Cd
## 1          36      F           B      501
## 2          25      M           H      102
## 3           0      X           X      726
## 4          76      F           W      502
## 5          31      X           X      409
## 6          25      F           H      735
##
##                               Premis.Desc Weapon.Used.Cd
## 1                               SINGLE FAMILY DWELLING                  400
## 2                               SIDEWALK                               500
## 3                               POLICE FACILITY                        NA
## 4 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)                        NA
## 5                               BEAUTY SUPPLY STORE                    NA
## 6                               NIGHT CLUB (OPEN EVENINGS ONLY)        500
##
##                               Weapon.Desc Status Status.Desc Crm.Cd.1
## 1 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE) AO Adult Other      624
## 2                               UNKNOWN WEAPON/OTHER WEAPON IC Invest Cont 624
## 3                               AA Adult Arrest      845
## 4                               IC Invest Cont      745
## 5                               IC Invest Cont      740
## 6                               UNKNOWN WEAPON/OTHER WEAPON IC Invest Cont 121
## Crm.Cd.2 Crm.Cd.3 Crm.Cd.4 LOCATION
## 1      NA      NA      NA 1100 W 39TH PL
## 2      NA      NA      NA 700 S HILL ST
## 3      NA      NA      NA 200 E 6TH ST
## 4      998      NA      NA 5400 CORTEEN PL
## 5      NA      NA      NA 14400 TITUS ST
## 6      998      NA      NA 700 S BROADWAY
## Cross.Street LAT LON
## 1              34.0141 -118.2978
## 2              34.0459 -118.2545
## 3              34.0448 -118.2474
## 4              34.1685 -118.4019
## 5              34.2198 -118.4468
## 6              34.0452 -118.2534

```

```

# Changing Date.Rptd and DATE.OCC to date format
crime$Date.Rptd <- as.Date(crime$Date.Rptd, format = "%m/%d/%Y")
crime$DATE.OCC <- as.Date(crime$DATE.OCC, format = "%m/%d/%Y")

times_char <- sprintf("%04d", crime$TIME.OCC)
datetime <- strptime(times_char, format = "%H%M")
crime$TIME.OCC <- format(datetime, format = "%H:%M")

```

```
head(crime)
```

```
##      DR_NO  Date.Rptd  DATE.OCC TIME.OCC AREA  AREA.NAME Rpt.Dist.No
## 1  10304468 2020-01-08 2020-01-08  22:30   3  Southwest      377
## 2  190101086 2020-01-02 2020-01-01  03:30   1   Central      163
## 3  200110444 2020-04-14 2020-02-13  12:00   1   Central      155
## 4  191501505 2020-01-01 2020-01-01  17:30  15 N Hollywood    1543
## 5  191921269 2020-01-01 2020-01-01  04:15  19   Mission     1998
## 6  200100501 2020-01-02 2020-01-01  00:30   1   Central      163
##  Part.1.2 Crm.Cd                                Crm.Cd.Desc
## 1         2      624                                BATTERY - SIMPLE ASSAULT
## 2         2      624                                BATTERY - SIMPLE ASSAULT
## 3         2      845                                SEX OFFENDER REGISTRANT OUT OF COMPLIANCE
## 4         2      745                                VANDALISM - MISDEAMEANOR ($399 OR UNDER)
## 5         2      740 VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)
## 6         1      121                                RAPE, FORCIBLE
##      Mocodes Vict.Age Vict.Sex Vict.Descent Premis.Cd
## 1         0444 0913      36         F         B         501
## 2         0416 1822 1414      25         M         H         102
## 3             1501         0         X         X         726
## 4             0329 1402      76         F         W         502
## 5             0329      31         X         X         409
## 6 0413 1822 1262 1415      25         F         H         735
##      Premis.Desc Weapon.Used.Cd
## 1             SINGLE FAMILY DWELLING         400
## 2             SIDEWALK         500
## 3             POLICE FACILITY         NA
## 4 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)         NA
## 5             BEAUTY SUPPLY STORE         NA
## 6             NIGHT CLUB (OPEN EVENINGS ONLY)         500
##      Weapon.Desc Status Status.Desc Crm.Cd.1
## 1 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)    AO  Adult Other      624
## 2             UNKNOWN WEAPON/OTHER WEAPON    IC  Invest Cont      624
## 3             UNKNOWN WEAPON/OTHER WEAPON    AA  Adult Arrest      845
## 4             UNKNOWN WEAPON/OTHER WEAPON    IC  Invest Cont      745
## 5             UNKNOWN WEAPON/OTHER WEAPON    IC  Invest Cont      740
## 6             UNKNOWN WEAPON/OTHER WEAPON    IC  Invest Cont      121
##  Crm.Cd.2 Crm.Cd.3 Crm.Cd.4                                LOCATION
## 1         NA         NA         NA 1100 W 39TH                                PL
## 2         NA         NA         NA  700 S HILL                                ST
## 3         NA         NA         NA  200 E 6TH                                ST
## 4         998         NA         NA 5400 CORTEEN                                PL
## 5         NA         NA         NA 14400 TITUS                                ST
## 6         998         NA         NA             700 S BROADWAY
##  Cross.Street      LAT      LON
## 1             34.0141 -118.2978
## 2             34.0459 -118.2545
## 3             34.0448 -118.2474
## 4             34.1685 -118.4019
## 5             34.2198 -118.4468
## 6             34.0452 -118.2534
```

```
# Finding dimension of crime dataset
dim(crime)
```

```
## [1] 788767      28
```

This dataset has 788,767 rows and 28 columns.

```
# Applying NA function to each column
crime_is_na <- sapply(crime, function(x) sum(is.na(x)))
crime_is_na[which(crime_is_na > 0)]
```

```
##      Premis.Cd Weapon.Used.Cd      Crm.Cd.1      Crm.Cd.2      Crm.Cd.3
##           9       514250          10       730596       786831
##      Crm.Cd.4
##       788710
```

Premis.Cd has some missing values. Weapon.Used.Cd has so many missing values because not all recorded crimes involve a weapon, and Crm.Cd.2-4 have many missing values because they will only have a value if more than 1 crime was committed in the same instance. I will not be removing any rows that contain missing values because if I do, there will not be nearly as much data. As a result, I will account for missing values during statistical tests.

What are the different types of crimes recorded?

```
# Finding length of crime list
length(table(crime$Crm.Cd.Desc))
```

```
## [1] 138
```

There are 138 different types of crimes that have been committed.

```
# Sorting the number of occurrences for each crime
sorted_crimes <- table(crime$Crm.Cd.Desc) %>% sort(decreasing = TRUE)
```

There are too many crimes to print out the name of each one, but if you'd like you can download this rmd and print the sorted_crimes variable.

```
head(sorted_crimes)
```

```
##
##                                VEHICLE - STOLEN
##                                84390
##                                BATTERY - SIMPLE ASSAULT
##                                62631
##                                THEFT OF IDENTITY
##                                50404
##                                BURGLARY FROM VEHICLE
##                                48479
## VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)
##                                48199
##                                BURGLARY
##                                47994
```

Here are the 5 highest occurring crimes.

```
tail(sorted_crimes)
```

```
##
## FIREARMS RESTRAINING ORDER (FIREARMS RO)
##                                4
##                                FAILURE TO DISPERSE
##                                3
```

```
##          GRAND THEFT / AUTO REPAIR
##                               3
##          PICKPOCKET, ATTEMPT
##                               3
##    DISHONEST EMPLOYEE ATTEMPTED THEFT
##                               2
##          INCITING A RIOT
##                               1
```

Here are the 5 lowest occurring crimes.

How many of the crimes were only attempted?

```
# Finding number of crimes with "ATTEMPT" in title
sum(grepl("ATTEMPT", crime$Crm.Cd.Desc, ignore.case = TRUE))
```

```
## [1] 12723
```

Out of all the crimes committed, 12,723 of them were attempts.

How many crimes involved stolen property?

```
# Finding number of crimes with "STOLEN" in title
sum(grepl("STOLEN", crime$Crm.Cd.Desc, ignore.case = TRUE))
```

```
## [1] 98087
```

Out of all the crimes committed, 98,087 of them involved stolen property.

Which days had the most amount of crime? The least?

```
# Sorting dates of crimes based on how many occurred that day
date_occ_counts <- crime %>% count(DATE.OCC) %>% arrange(desc(n))
```

```
head(date_occ_counts)
```

```
##    DATE.OCC    n
## 1 2022-12-02 1128
## 2 2020-01-01 1101
## 3 2022-10-01 1054
## 4 2023-02-01 1039
## 5 2023-02-02 1030
## 6 2022-12-01 1027
```

The date with the most amount of crime is 12/02/2022 with 1128 instances.

```
tail(date_occ_counts)
```

```
##    DATE.OCC    n
## 1331 2020-03-21 427
## 1332 2023-08-27 422
## 1333 2020-03-29 421
## 1334 2020-03-22 417
## 1335 2020-04-09 415
## 1336 2023-08-28 150
```

The date with the least amount of crime is 03/21/2020 with 427 instances. If you noticed, half of the lowest crime occurrence dates is in March 2020, which happens to be when the United States had lock downs for

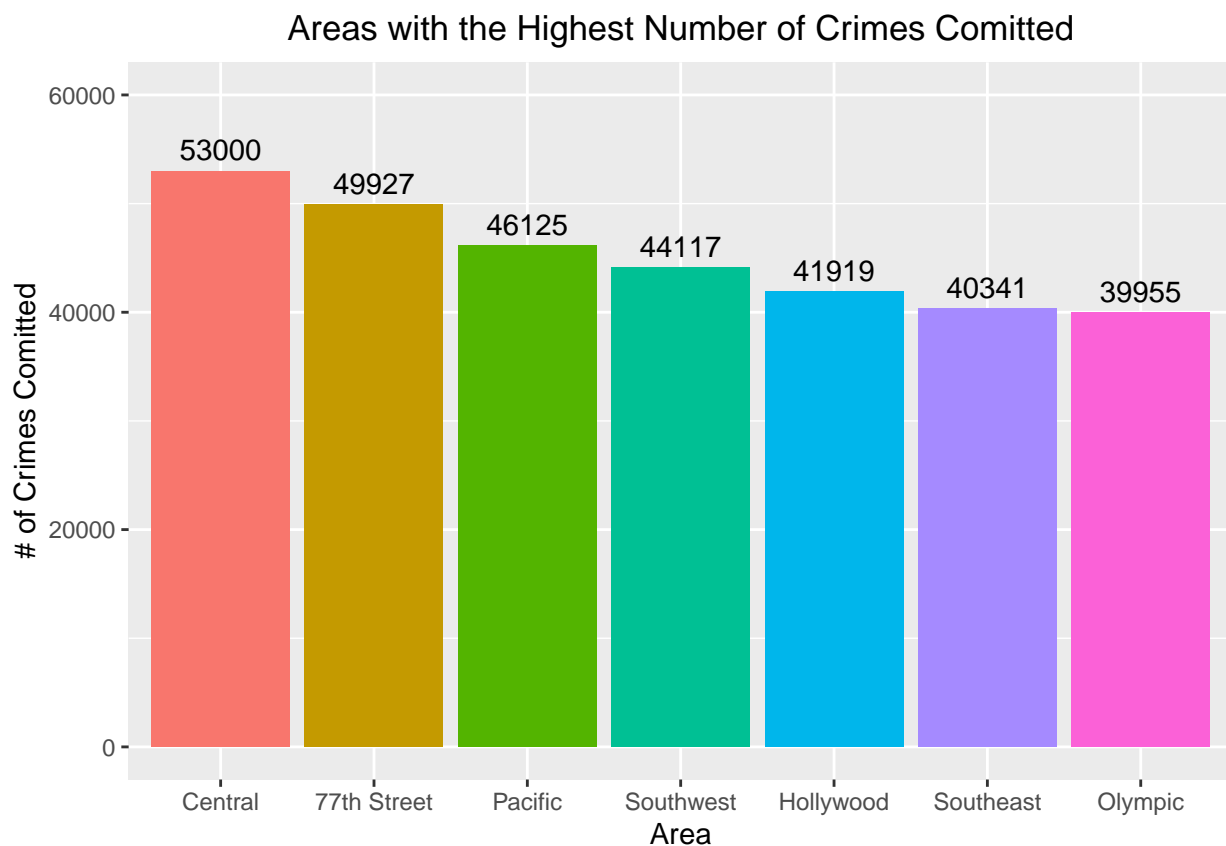
COVID-19.

Graphing the Data and Conducting Statistical Analysis

Which areas had the most amount of crime? The least?

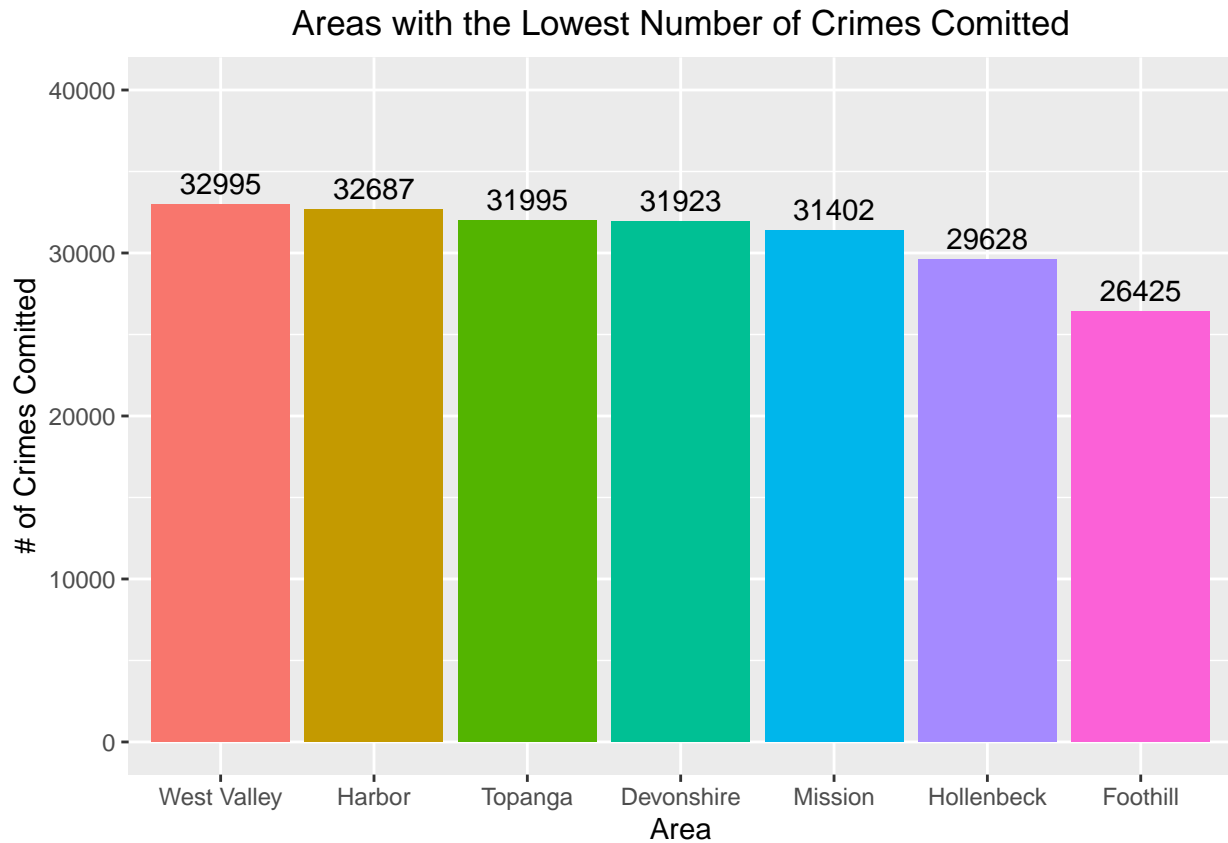
```
area_name_counts <- crime %>% count(AREA.NAME) %>% arrange(desc(n))
```

```
ggplot(head(area_name_counts, 7), aes(x = fct_inorder(AREA.NAME), y = n, fill = fct_inorder(AREA.NAME))) +  
  geom_bar(stat = "identity") +  
  geom_text(aes(label = n), vjust = -0.5) +  
  labs(x = "Area", y = "# of Crimes Comitted", title = "Areas with the Highest Number of Crimes Comitted") +  
  theme(plot.title = element_text(hjust = 0.5), legend.position = "none") +  
  scale_y_continuous(limits = c(0, 60000))
```



The area with the highest amount of crime is Central with 53,000 reported cases.

```
ggplot(tail(area_name_counts, 7), aes(x = fct_inorder(AREA.NAME), y = n, fill = fct_inorder(AREA.NAME))) +  
  geom_bar(stat = "identity") +  
  geom_text(aes(label = n), vjust = -0.5) +  
  labs(x = "Area", y = "# of Crimes Comitted", title = "Areas with the Lowest Number of Crimes Comitted") +  
  theme(plot.title = element_text(hjust = 0.5), legend.position = "none") +  
  scale_y_continuous(limits = c(0, 40000))
```



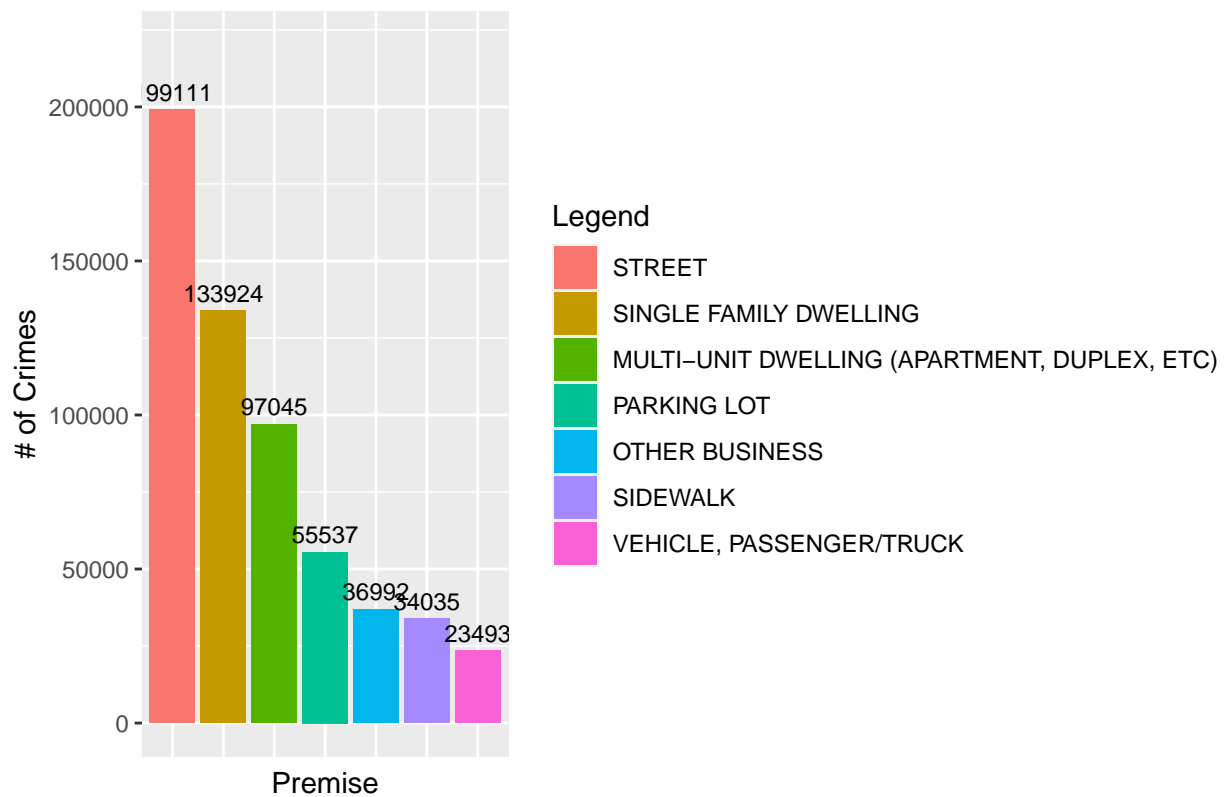
The area with the lowest amount of crime is Foothill with 26,425 reported cases.

What premise had a crime occur the most? The least?

```
premise_counts <- crime %>% count(Premis.Desc) %>% arrange(desc(n))
```

```
ggplot(head(premise_counts, 7), aes(x = fct_inorder(Premis.Desc), y = n, fill = fct_inorder(Premis.Desc))) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = n), vjust = -0.5, size = 3) +
  labs(x = "Premise", y = "# of Crimes", title = "Premise Where Crime Occured the Most") +
  theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  scale_y_continuous(limits = c(0, 220000)) +
  guides(fill = guide_legend(title = "Legend"))
```

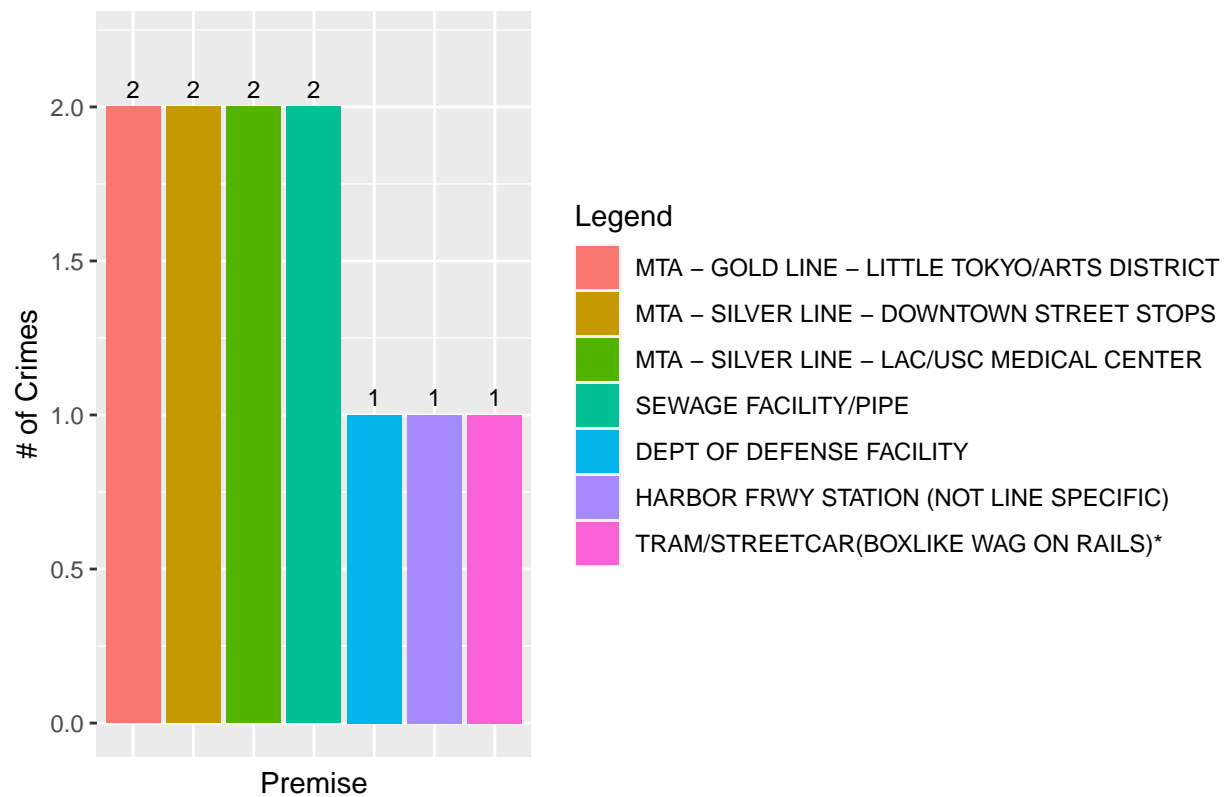
Premise Where Crime Occured the Most



The premise where the most crime occurs are streets.

```
ggplot(tail(premise_counts, 7), aes(x = fct_inorder(Premis.Desc), y = n, fill = fct_inorder(Premis.Desc))) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = n), vjust = -0.5, size = 3) +
  labs(x = "Premise", y = "# of Crimes", title = "Premise Where Crime Occured the Least") +
  theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  scale_y_continuous(limits = c(0, 2.2)) +
  guides(fill = guide_legend(title = "Legend"))
```


Premise Where Crime Occured the Least

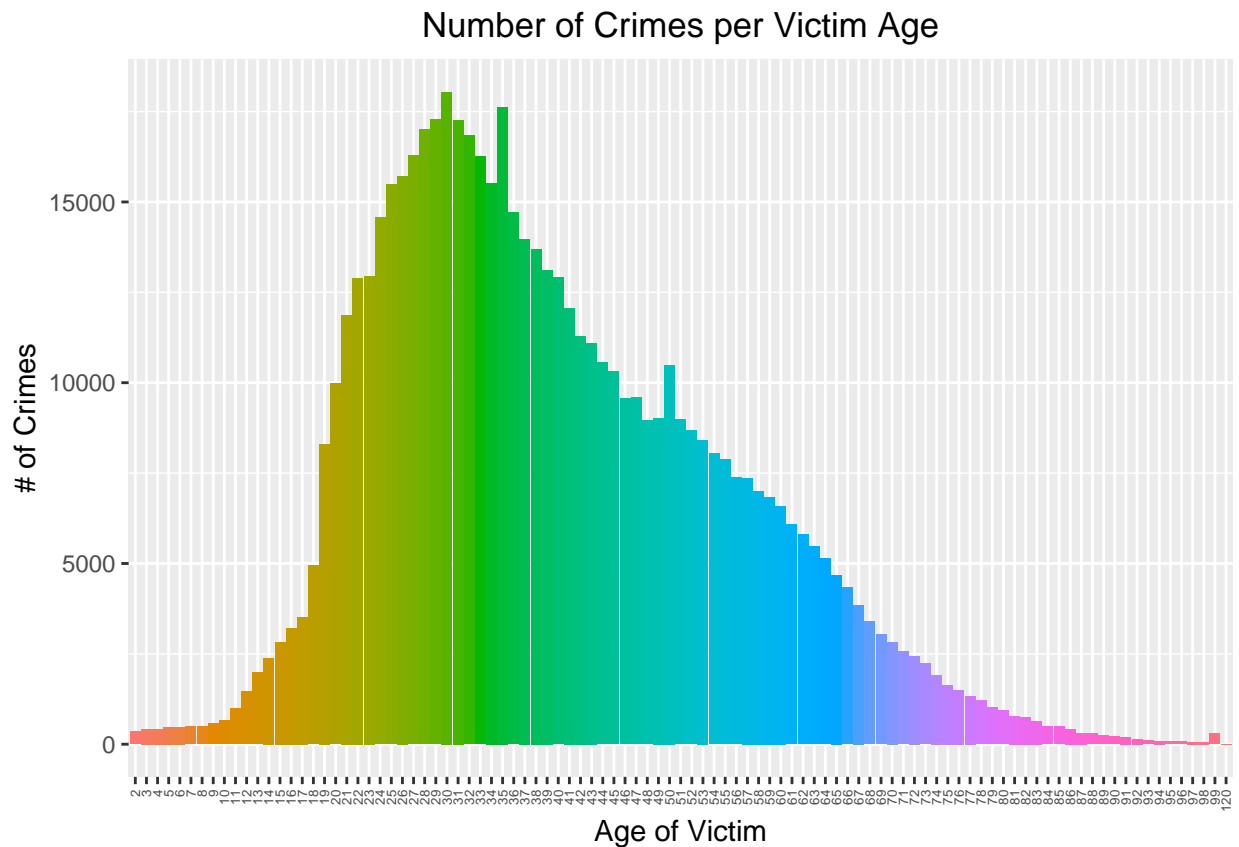


The premise where the least crime occurs are trams/streetcars.

Show the number of crimes committed per age of victim

```
crime_per_age <- table(crime$Vict.Age)[5:103] %>% as.data.frame()
```

```
ggplot(crime_per_age, aes(x = Var1, y = Freq, fill = Var1)) +
  geom_bar(stat = "identity") +
  labs(x = "Age of Victim", y = "# of Crimes", title = "Number of Crimes per Victim Age") +
  theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_text(angle = 90, vjust = 0.5, hju
```

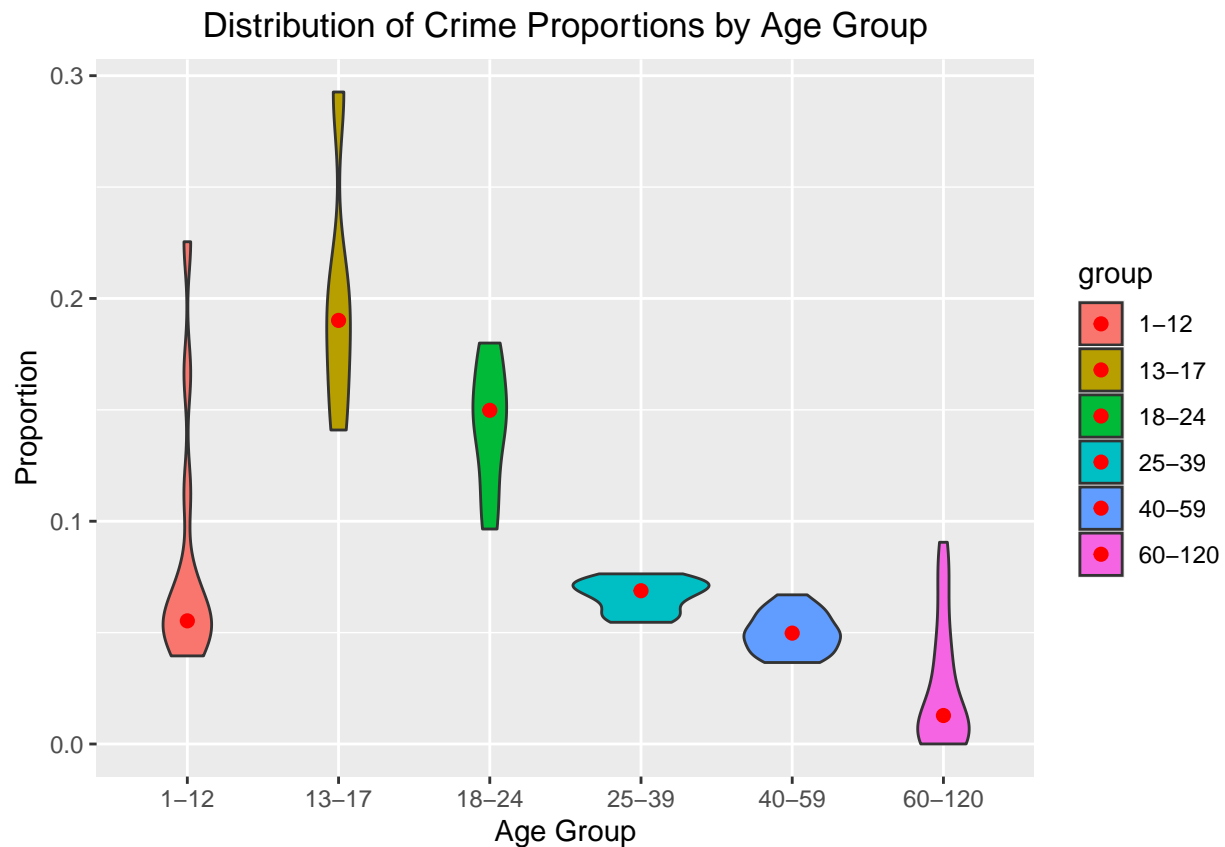


We can see most victims are aged 19-50, where more children and elderly people are less victimized.

```
# Creating age groups
crime_per_age$group <- cut(as.numeric(crime_per_age$Var1), breaks = c(0, 12, 17, 24, 39, 59, 120), labels = c("0-12", "13-17", "18-24", "25-39", "40-59", "60-120"))

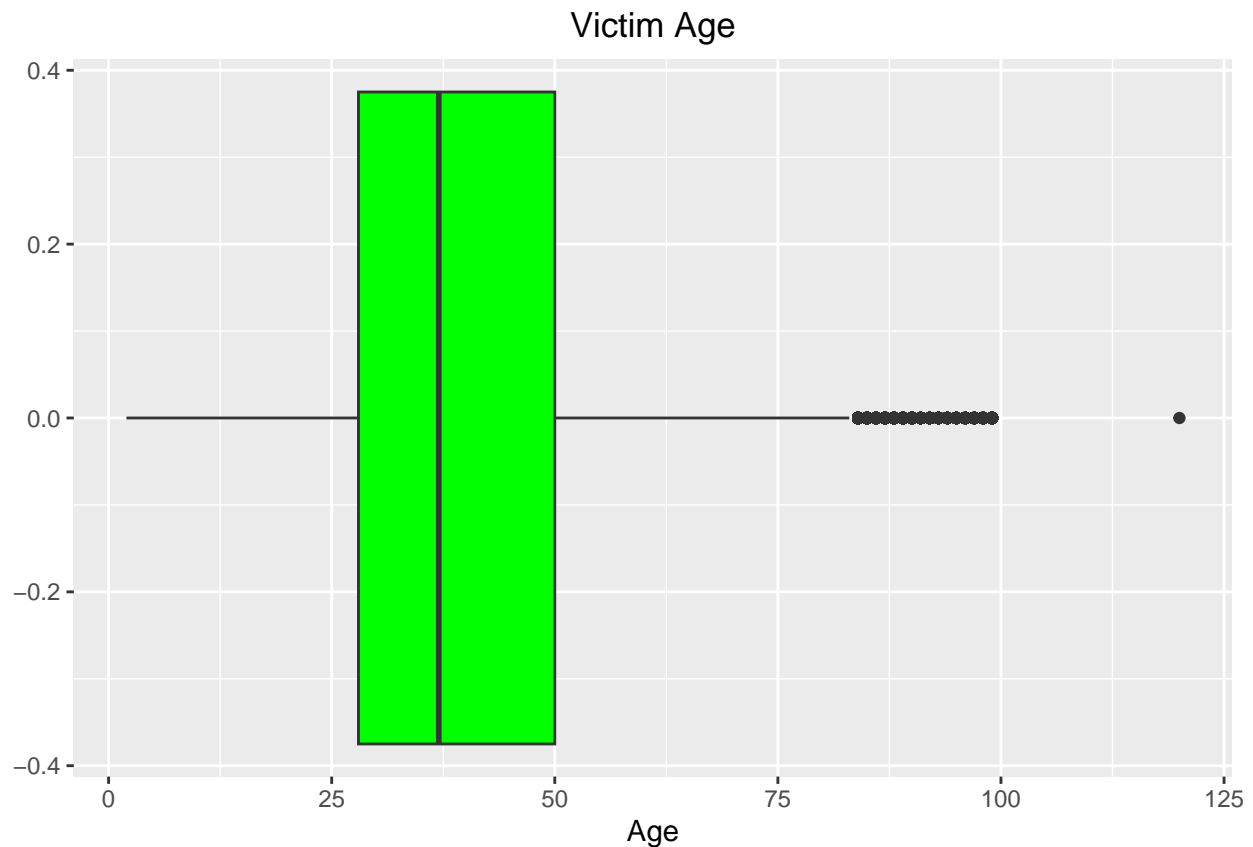
# Finding age group proportions
crime_per_age <- crime_per_age %>% group_by(group) %>% mutate(Proportion = Freq / sum(Freq))

ggplot(crime_per_age, aes(x = group, y = Proportion, fill = group)) +
  geom_violin() +
  stat_summary(fun = median, geom="point", size=2, color="red") +
  labs(x = "Age Group", y = "Proportion", title = "Distribution of Crime Proportions by Age Group") +
  theme(plot.title = element_text(hjust = 0.5))
```



From the violin chart, we can see the age group with the highest number of victims is 13-17. The red dots, representing the median, show the 13-17 age group takes up about 20% of the data. The lowest proportion group is the 60-120 year old group, likely due to the older individuals having less chance to become victim of a crime and it is skewed.

```
ggplot(crime[crime$Vict.Age > 0,], aes(x = Vict.Age)) +
  geom_boxplot(fill = "green") +
  labs(x = "Age", title = "Victim Age") +
  theme(plot.title = element_text(hjust = 0.5))
```



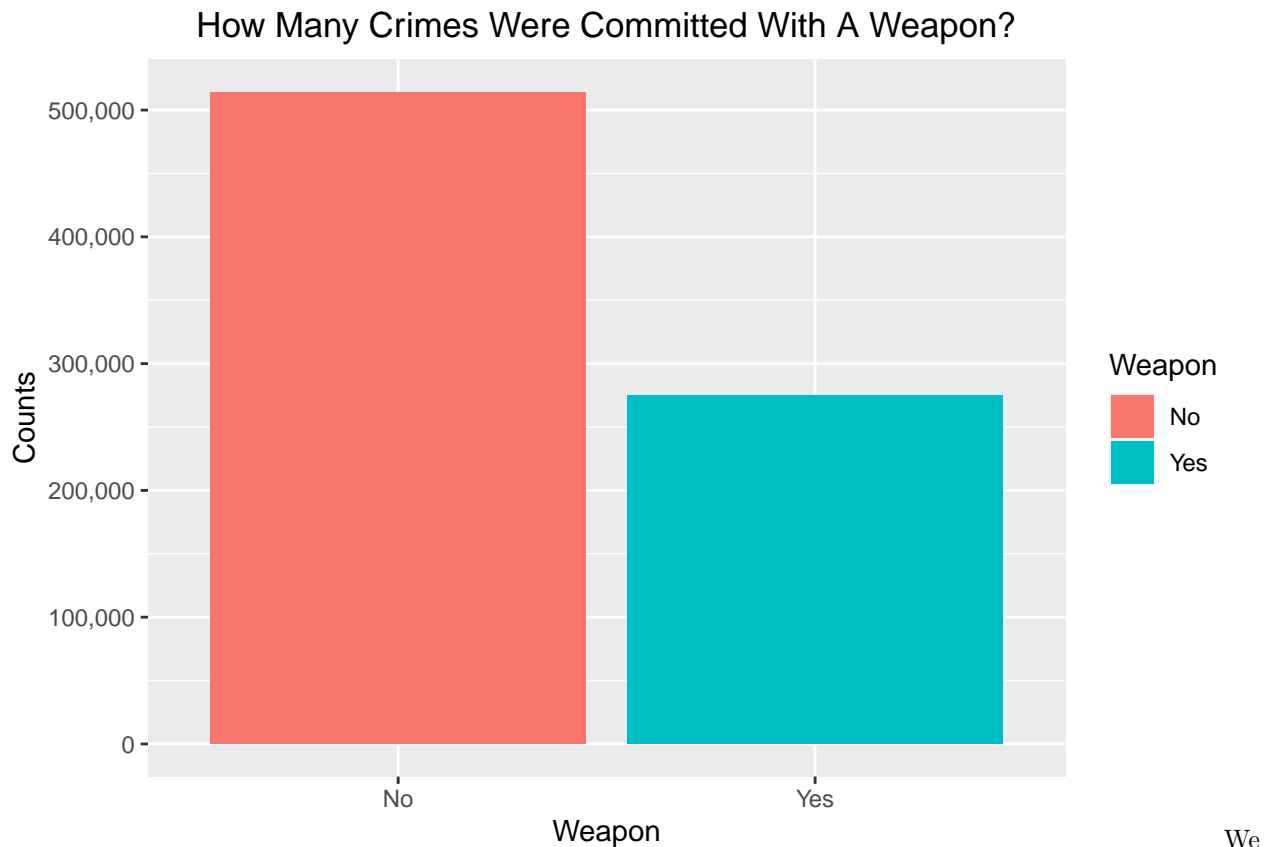
Median age of victim seems to be around 37 years old with an IQR of approximately 28-50 years old.

How many crimes occurred with a weapon or force?

```
# Finding number of crimes committed with a weapon
crime_weapon_na_count <- sum(is.na(crime$Weapon.Used.Cd))
crime_weapon_count <- 788767 - crime_weapon_na_count

weapons <- data.frame(
  Weapon = c("Yes", "No"),
  cc = c(crime_weapon_count, crime_weapon_na_count)
)

ggplot(weapons, aes(x = Weapon, y = cc, fill = Weapon)) +
  geom_bar(stat = "identity") +
  labs(x = "Weapon", y = "Counts", title = "How Many Crimes Were Committed With A Weapon?") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(labels = scales::comma)
```



can see there's about half as many crimes committed with a weapon than without one.

Is there a difference in the proportion of women who are victim of a crime to men?

```
male_victims <- subset(crime, Vict.Sex == "M")
female_victims <- subset(crime, Vict.Sex == "F")

prop.test(
  x = c(nrow(female_victims), nrow(male_victims)),
  n = c(788767, 788767)
)

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(nrow(female_victims), nrow(male_victims)) out of c(788767, 788767)
## X-squared = 3338, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.04642327 -0.04337765
## sample estimates:
##  prop 1    prop 2
## 0.3688390 0.4137394
```

With a p-value close to 0 and an alpha level of 0.05, we can reject the null hypothesis and state that the true proportion of victims of a crime who are woman or who are men are different.

Are woman victims of certain crimes more often than men?

```
gender_crime <- subset(crime, Vict.Sex == "M" | Vict.Sex == "F")

# Using Chi-Square test of Independence
chisq.test(table(gender_crime$Vict.Sex, gender_crime$Crm.Cd.Desc))

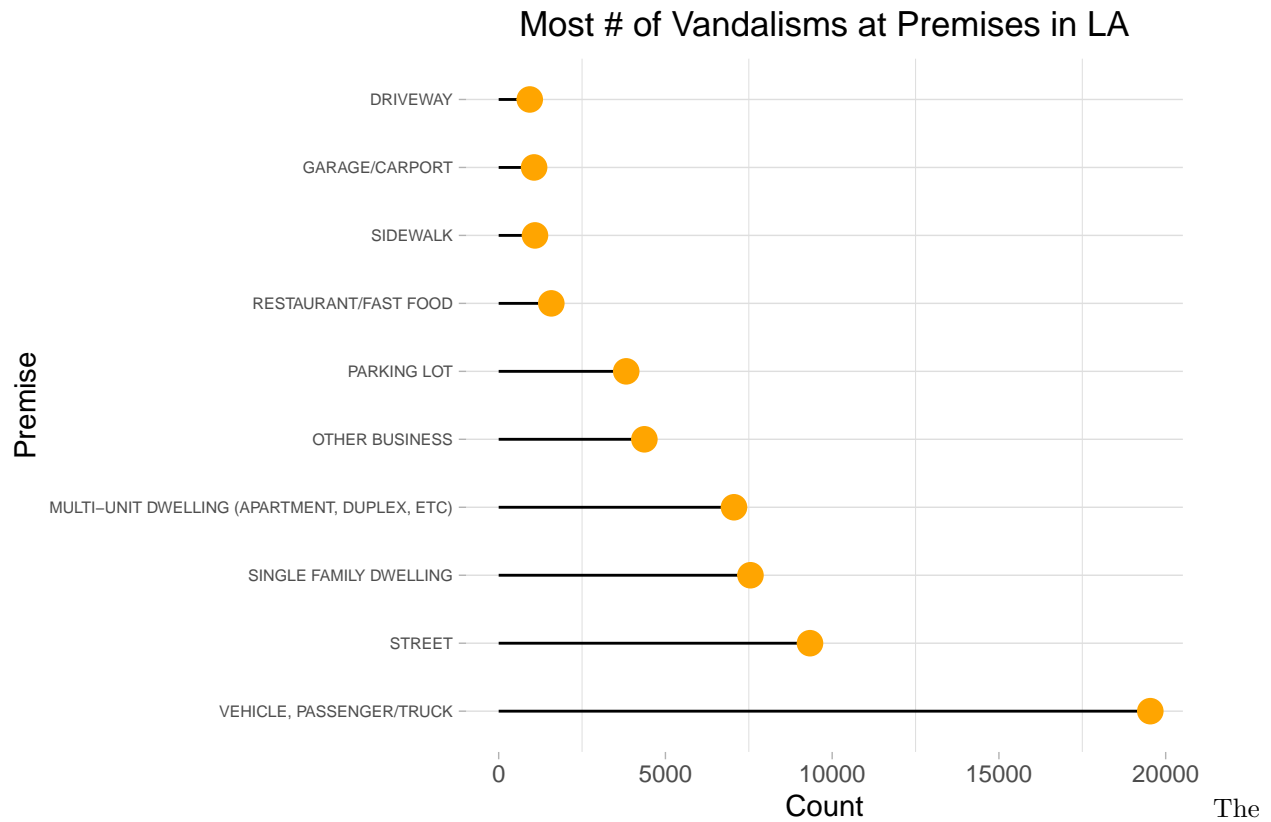
##
## Pearson's Chi-squared test
##
## data:  table(gender_crime$Vict.Sex, gender_crime$Crm.Cd.Desc)
## X-squared = 63413, df = 134, p-value < 2.2e-16
```

With a p-value close to 0 and an alpha level of 0.05, we can reject the null hypothesis, indicating a statistically significant association between gender and crime type on an individual.

What premise had the most vandalisms? The least?

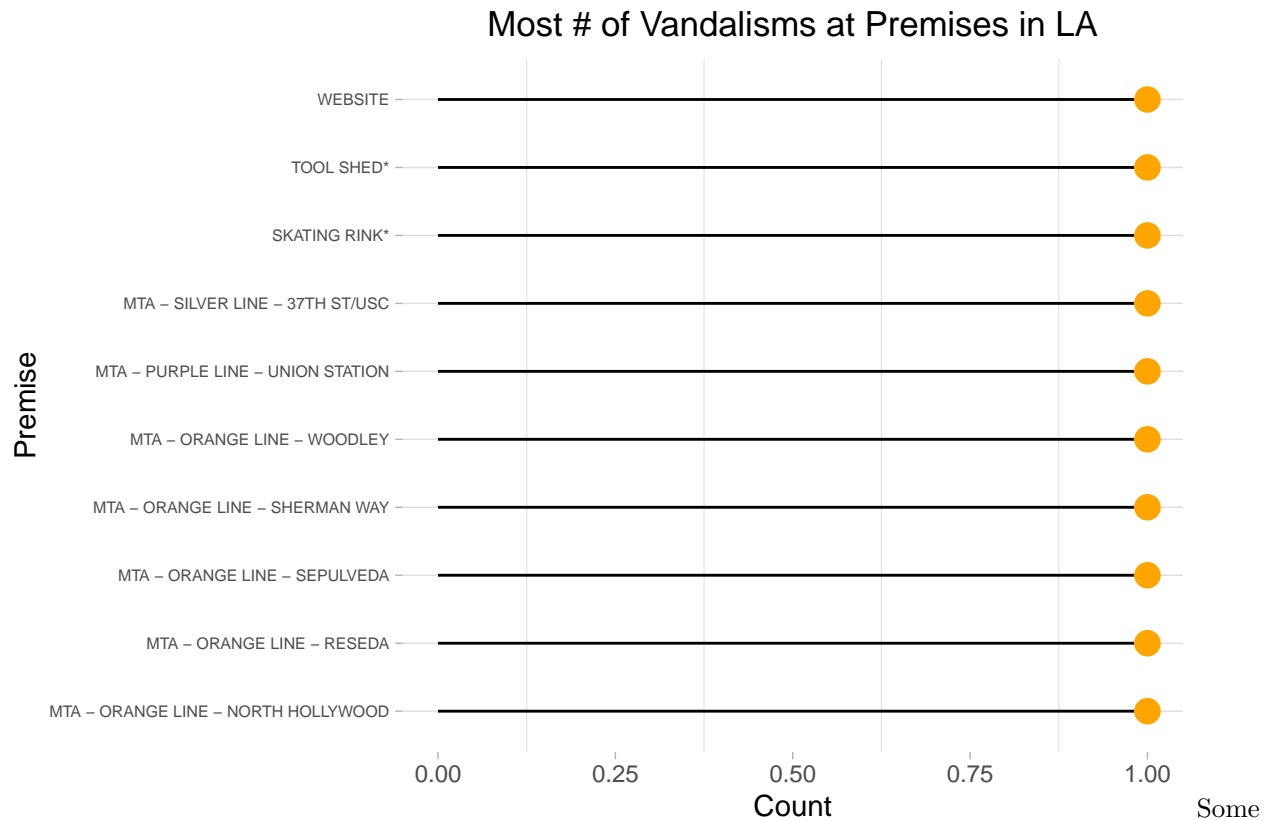
```
# Counting number of vandalism reports at each premise
premise_vandalisms <- crime %>% filter(grepl("VANDALISM", Crm.Cd.Desc, ignore.case = TRUE)) %>% group_by(Premis.Desc)

ggplot(head(premise_vandalisms, 10), aes(x = fct_inorder(Premis.Desc), y = Count)) +
  geom_segment(aes(x = fct_inorder(Premis.Desc), xend = Premis.Desc, y = 0, yend = Count)) +
  geom_point(color = "orange", size = 4) +
  coord_flip() +
  labs(x = "Premise", y = "Count", title = "Most # of Vandalisms at Premises in LA") +
  theme_light() +
  theme(axis.text.y = element_text(size = 6), plot.title = element_text(hjust = 0.5),
        panel.grid.major.x = element_blank(), panel.border = element_blank(),
        plot.margin = margin(c(10,20,10,10), unit = "pt")
  )
```



premise with the highest number of reported vandalism are vehicles (passenger/trucks).

```
ggplot(tail(premise_vandalisms, 10), aes(x = fct_inorder(Premis.Desc), y = Count)) +
  geom_segment(aes(x = fct_inorder(Premis.Desc), xend = Premis.Desc, y = 0, yend = Count)) +
  geom_point(color = "orange", size = 4) +
  coord_flip() +
  labs(x = "Premise", y = "Count", title = "Most # of Vandalisms at Premises in LA") +
  theme_light() +
  theme(axis.text.y = element_text(size = 6), plot.title = element_text(hjust = 0.5),
        panel.grid.major.x = element_blank(), panel.border = element_blank(),
        plot.margin = margin(c(10,20,10,10), unit = "pt"))
)
```



MTA lines have the lowest number of vandalisms, with only 1 recorded, as well as a skating rink and tool shed having only 1 instance. In fact, there's many more which only have 1 reported vandalism.

What percent of crimes are still being investigated?

```
sum(grepl("Cont", crime$Status.Desc, ignore.case = TRUE)) / nrow(crime)
```

```
## [1] 0.8005457
```

80.05% of crimes are still being investigated.

Are certain ethnic groups victims of crimes more often than others?

```
descent_table <- table(crime$Vict.Descent)[3:21]
prob = rep(1/length(descent_table), length(descent_table))

chisq.test(descent_table, p = prob)
```

```
##
## Chi-squared test for given probabilities
##
## data: descent_table
## X-squared = 2288176, df = 18, p-value < 2.2e-16
```

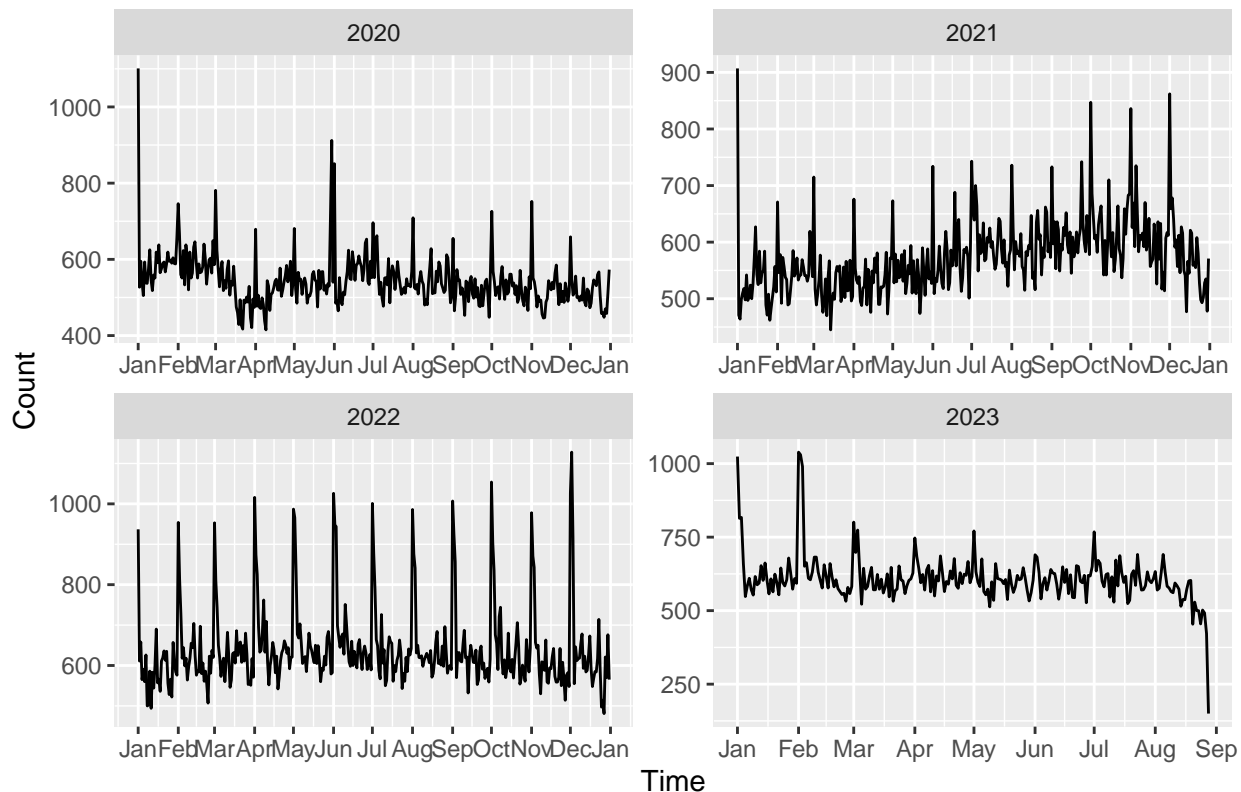
With a p-value is close to 0, and alpha = 0.5, we can reject the null and state that ethnic groups are not equally as likely to be targeted for a crime.

Are there any seasons where crime is more frequent? Less frequent?

```
# Grouping the date and number of crimes
daily_crime <- crime %>% group_by(DATE.OCC) %>% summarize(Count = n())
```

```
ggplot(daily_crime, aes(x = DATE.OCC, y = Count)) +
  geom_line() +
  facet_wrap(~ year(DATE.OCC), ncol = 2, nrow = 2, scales = "free") +
  labs(x = "Time", title = "Crime over Time") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_x_date(date_breaks = "1 month", date_labels = "%b")
```

Crime over Time



It seems like the beginning of each month there's a spike in crime where the biggest spike is in January.

Is there a certain month that has more crime yearly than others?

If we take a look at crime from the graph above, we can see that there's not really a specific month that generally has more or less crime than others.

Has crime increased over time in any area?

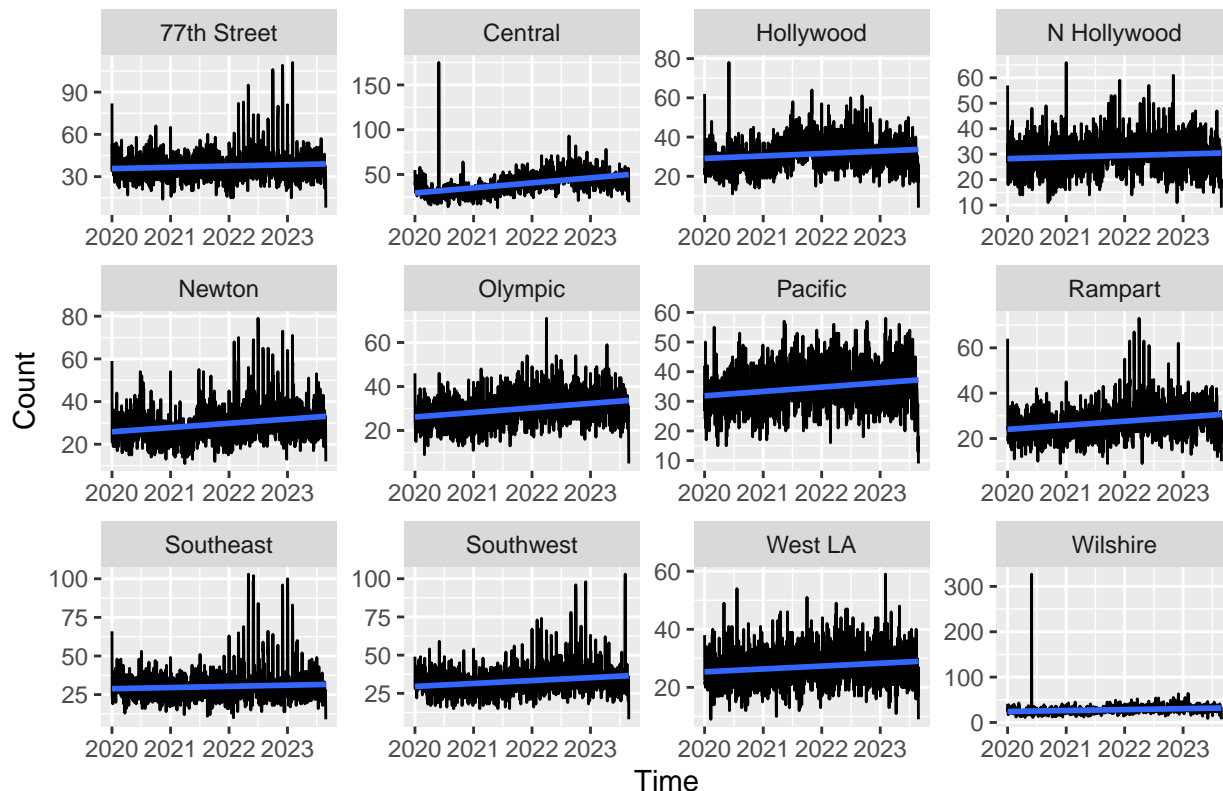
```
# Grouping the top 12 areas by date and number of crimes
selected <- area_name_counts[1:12,1]
filtered_crime <- crime %>% filter(AREA.NAME %in% selected)
crime_time_area <- filtered_crime %>% group_by(DATE.OCC, AREA.NAME) %>% summarize(Count = n())

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

```
ggplot(crime_time_area, aes(x = DATE.OCC, y = Count)) +
  geom_line() +
  facet_wrap(~ AREA.NAME, scales = "free") +
  labs(x = "Time", title = "Crime over Time") +
  theme(plot.title = element_text(hjust = 0.5)) +
  stat_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Crime over Time



No area seems to have a significant increase in crime over time, but some areas like Central, Newton, Olympic, Pacific, and West LA do have some increase in crime. The areas all seem to have some major spikes in crime on certain days, but nothing that overlaps with each other.

Are there certain times of the day where crime occurs more? What about day vs night?

```
# Convert military time to hours and minutes
hours <- as.numeric(substr(crime$TIME.OCC, 1, 2))
minutes <- as.numeric(substr(crime$TIME.OCC, 4, 5))

# Calculate the intervals
interval <- sprintf("%02d:%02d", floor((hours * 60 + minutes) %/% 30 / 2), ((hours * 60 + minutes) %/% 30 %/ 2))
interval_df <- as.data.frame(table(interval))

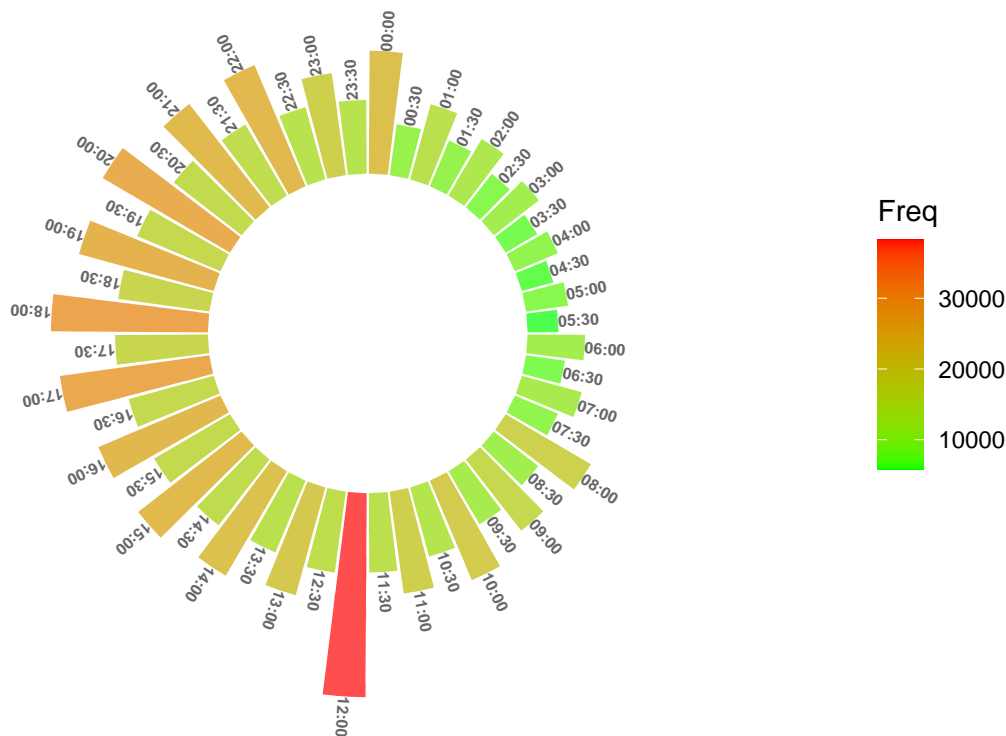
number_of_bar <- nrow(interval_df)
angle <- 90 - 360 * (as.numeric(interval_df$interval) - 0.5) / number_of_bar

interval_df$angle <- ifelse(angle < -90, angle + 180, angle)
```

```

# Plot of crime frequency per 30 minutes
ggplot(interval_df, aes(x = as.factor(interval), y = Freq, fill = Freq)) +
  geom_bar(stat = "identity", alpha = 0.7) +
  geom_text(aes(x = interval, y = Freq, label = interval, hjust = 0),
            color = "black", fontface = "bold", alpha = 0.6, size = 2, angle = angle,
            inherit.aes = FALSE) +
  scale_fill_gradient(low = "green", high = "red") +
  theme_minimal() +
  theme(
    axis.text = element_blank(),
    panel.grid = element_blank(),
    axis.title = element_blank(),
    plot.margin = unit(rep(-0.5, 4), "cm")
  ) +
  coord_polar(start = 0) +
  scale_y_continuous(limits = c(-30000, 40000))

```



We can see that as the time of day increases, the frequency of crime increases, but there is a major spike at noon (12:00).