NYPD Shooting Project_Week3

J.Kaur

6/20/2021

```
library(magrittr)
library(dplyr)
```

```
##
## 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
```

```
library(readr)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'

## The following object is masked from 'package:magrittr':
##
##     extract
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(chron)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:chron':
##
##     days, hours, minutes, seconds, years

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
library(modelr)
```

The data getting analyzed today is from the NYPD Shooting Incident Rate.We are provided with the perp and victim data for the past 15 years We will looking at the Gender vs Shooting Incident Rates and later predict crime rates for the next few years

```
#Import data
url_link <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"

#Read the data in the csv
complete_data <- read.csv(url_link, header=T, na.strings=c("","NA"))

#Drop irrelevant columns
relevant_data <- subset(complete_data, select=-c(INCIDENT_KEY,STATISTICAL_MURDER_FLAG,LOCATION_DESC,
                                                 JURISDICTION_CODE,X_COORD_CD, Y_COORD_CD,
                                                 Latitude, Longitude, Lon_Lat))

summary(relevant_data)
```

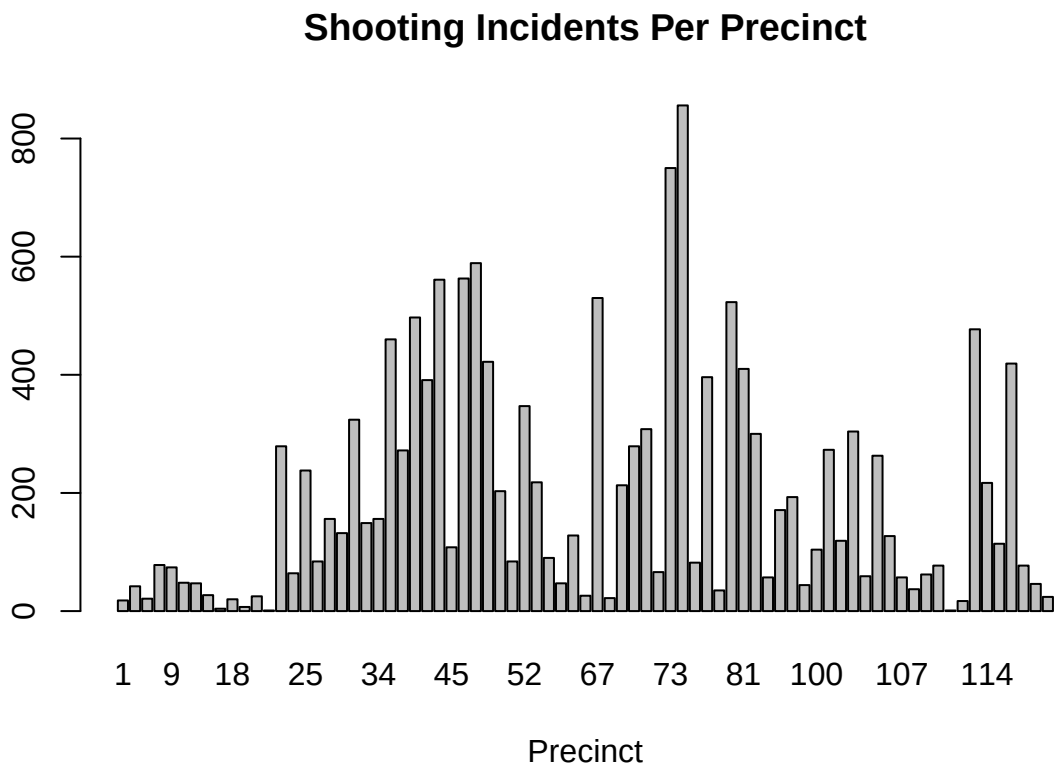
```
##   OCCUR_DATE         OCCUR_TIME            BORO              PRECINCT
##  Length:23568       Length:23568       Length:23568       Min.   :  1.00
##  Class :character   Class :character   Class :character   1st Qu.: 44.00
##  Mode  :character   Mode  :character   Mode  :character   Median : 69.00
##                                                           Mean   : 66.21
##                                                           3rd Qu.: 81.00
##                                                           Max.   :123.00
##  PERP_AGE_GROUP       PERP_SEX           PERP_RACE         VIC_AGE_GROUP
##  Length:23568       Length:23568       Length:23568       Length:23568
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    VIC_SEX            VIC_RACE
##  Length:23568       Length:23568
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
#Data Cleaning
```

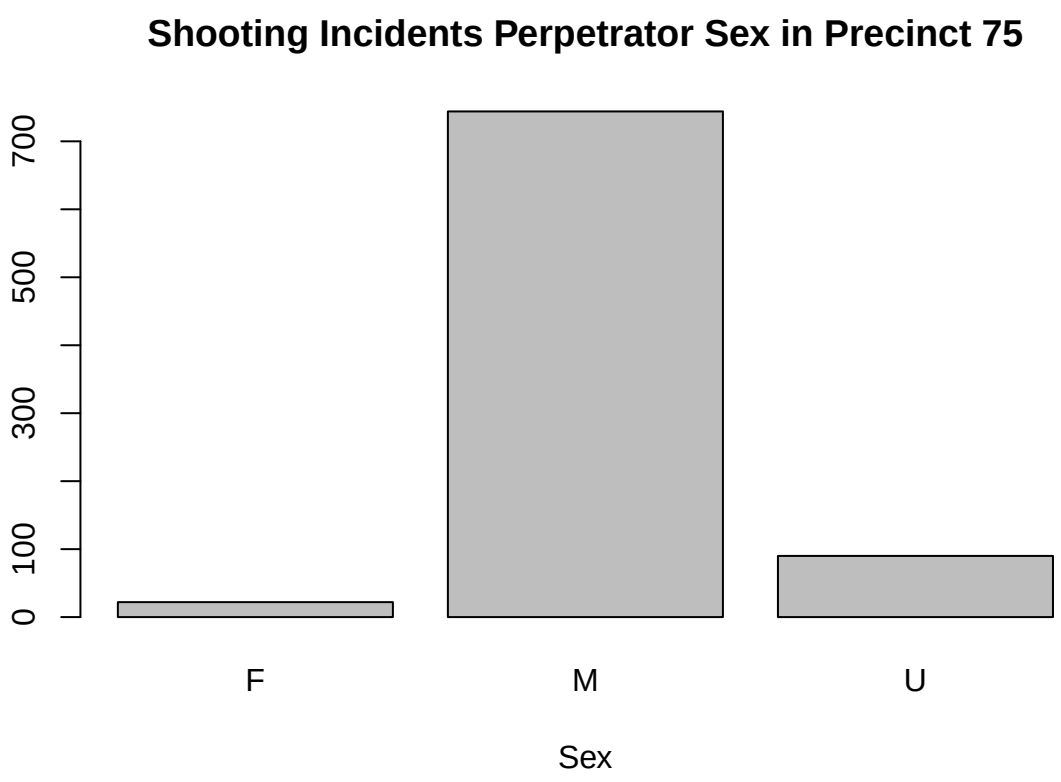
```r
#Date field is character value that needs to be updated
relevant_data[["OCCUR_DATE"]] <- as.Date(relevant_data[["OCCUR_DATE"]], format = "%m/%d/%Y")

#Time field is character value that needs to be updated
relevant_data[["OCCUR_TIME"]] <- chron(times=relevant_data[["OCCUR_TIME"]])

#Remove any rows with NA
relevant_data <- na.omit(relevant_data)

summary(relevant_data)
```

```
##    OCCUR_DATE           OCCUR_TIME           BORO              PRECINCT
##  Min.   :2006-01-01   Min.   :00:00:00   Length:15109       Min.   :  1.00
##  1st Qu.:2008-04-02   1st Qu.:03:39:00   Class :character   1st Qu.: 44.00
##  Median :2010-07-10   Median :15:15:00   Mode  :character   Median : 69.00
##  Mean   :2011-09-26   Mean   :12:47:03                      Mean   : 65.93
##  3rd Qu.:2015-01-04   3rd Qu.:20:35:00                      3rd Qu.: 81.00
##  Max.   :2020-12-29   Max.   :23:59:00                      Max.   :123.00
##  PERP_AGE_GROUP       PERP_SEX          PERP_RACE         VIC_AGE_GROUP
##  Length:15109       Length:15109       Length:15109       Length:15109
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    VIC_SEX            VIC_RACE
##  Length:15109       Length:15109
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

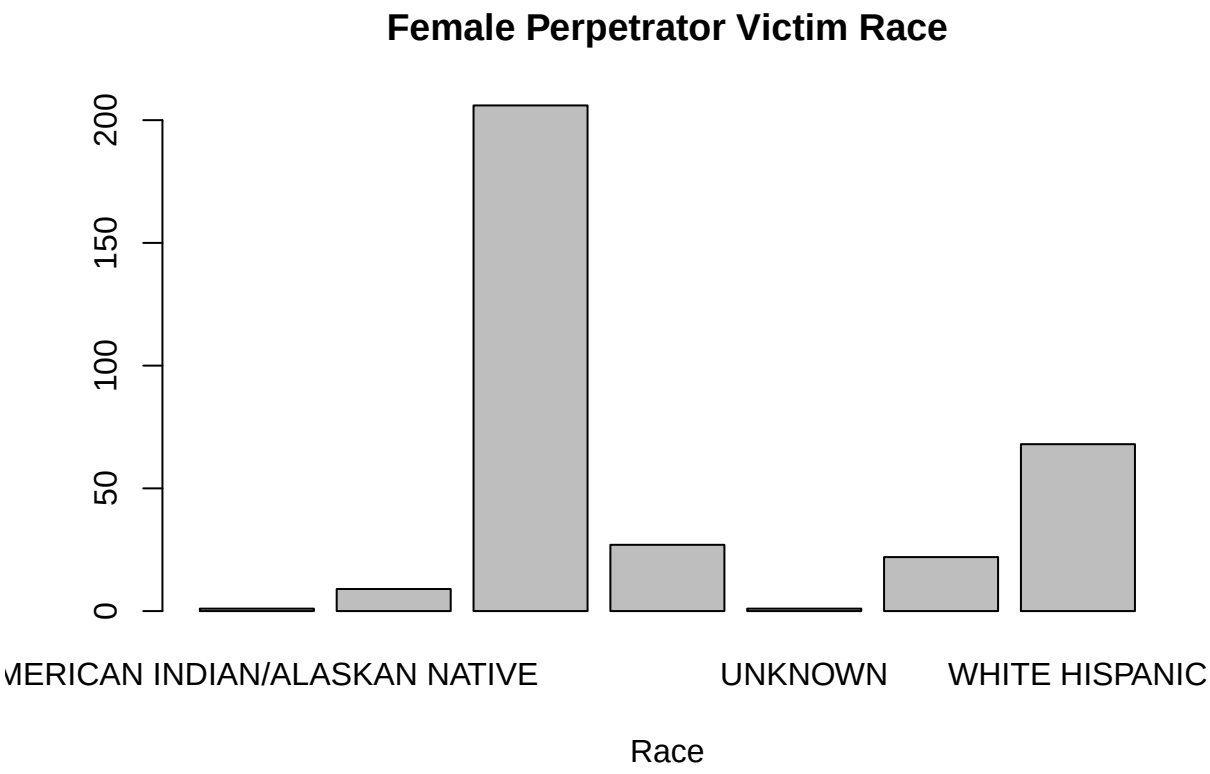
```r
#Data Analyzing
#Add a Year column for modelling purposes
relevant_data[["YEAR"]] <- format(relevant_data[["OCCUR_DATE"]], format="%Y")


#Plotting a bar plot to show precinct with the highest incidence rate (#75)
precinct_data <- table(relevant_data["PRECINCT"])
barplot(precinct_data, main="Shooting Incidents Per Precinct", xlab="Precinct")
```



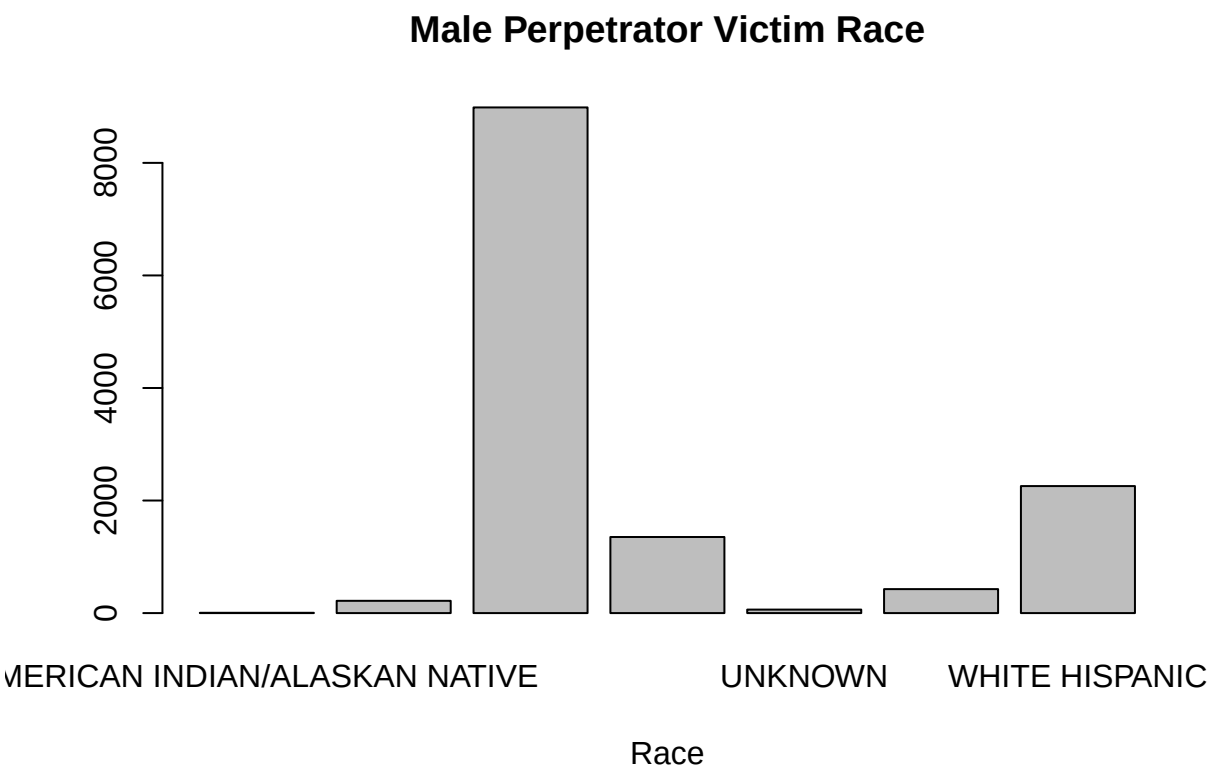

```r
#Looking at precinct 75 data in more detail
precinct_75 <- filter(relevant_data, PRECINCT == 75)

#Plotting a bar plot to show perpetrator sex in precinct 75
perp_sex_75 <- table(precinct_75["PERP_SEX"])
barplot(perp_sex_75, main="Shooting Incidents Perpetrator Sex in Precinct 75", xlab="Sex")
```




```r
#Just look at female perpetrators in all data
female_perp <- filter(relevant_data, PERP_SEX == "F")

#Plotting a bar plot to show female perpetrator victims' race
female_perp_vic_race <- table(female_perp["VIC_RACE"])
barplot(female_perp_vic_race, main="Female Perpetrator Victim Race", xlab="Race")
```



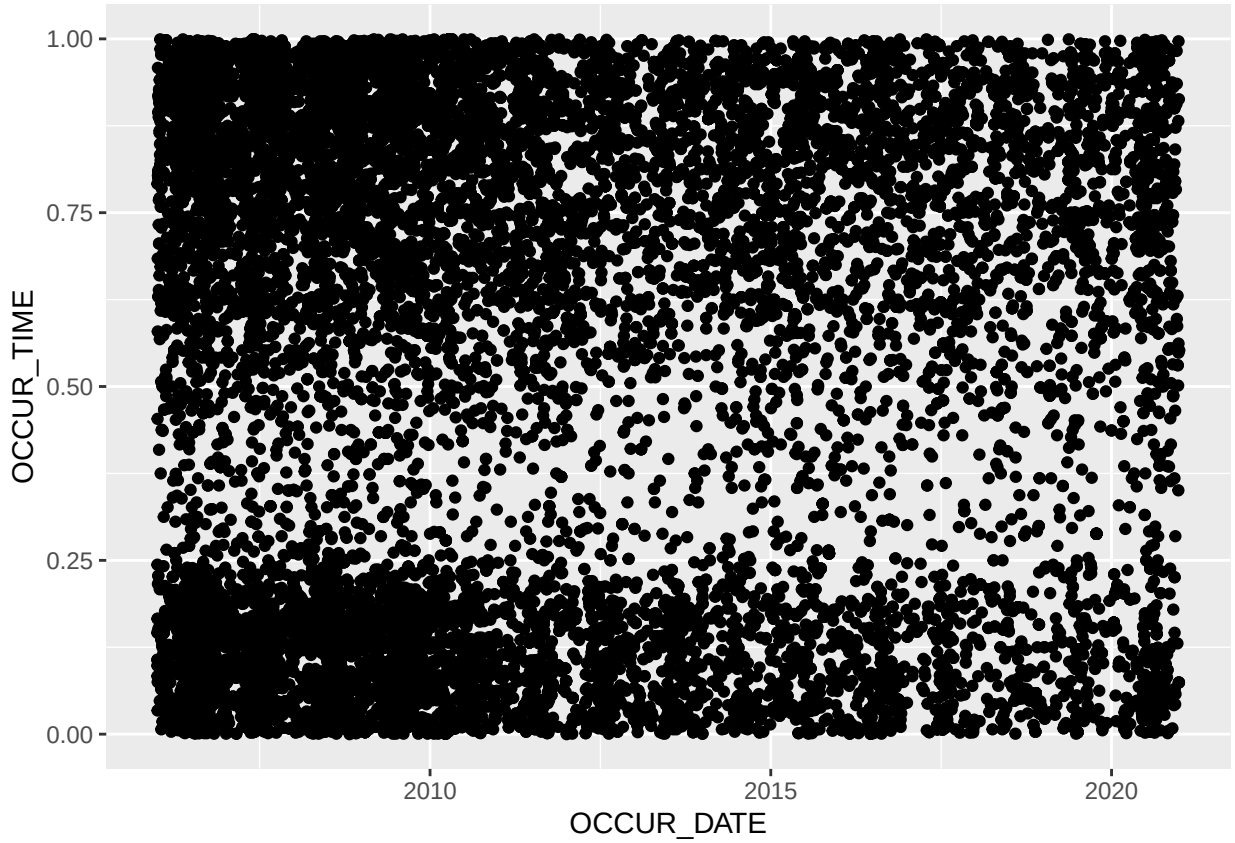

```
#Just look at male perpetrators in all data
male_perp <- filter(relevant_data, PERP_SEX == "M")

#Plotting a bar plot to show male perpetrator victims' race
male_perp_vic_race <- table(male_perp["VIC_RACE"])
barplot(male_perp_vic_race, main="Male Perpetrator Victim Race", xlab="Race")
```



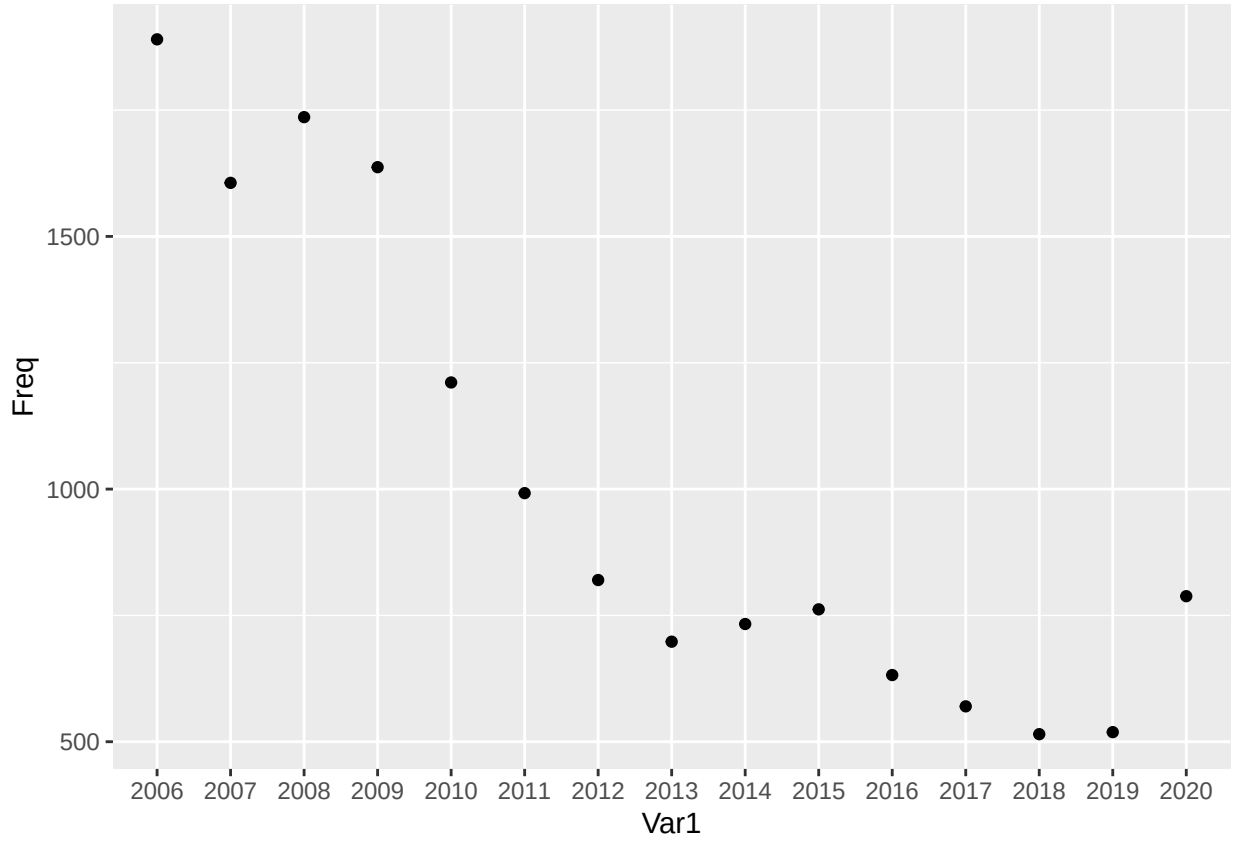
Looking at the above data, there is no significant difference between data with respect to gender. Crime against black community was at an all time high in all cases unfortunately.

```
#Plotting shooting time over date
ggplot(relevant_data, aes(x=OCCUR_DATE, y=OCCUR_TIME)) + geom_point()
```

```
## Don't know how to automatically pick scale for object of type times. Defaulting to continuous.
```




```
#Calculate and plot crime per year
year_crime_data <- data.frame(table(relevant_data[["YEAR"]]))
ggplot(year_crime_data, aes(x=Var1, y=Freq)) + geom_point()
```




```
#Model the change in crime every year
mod_crime <- lm(Freq ~ as.numeric(Var1), data = year_crime_data)

#Predict expected numbers for 2021, 2022, and 2023
new.df <- data.frame(Var1=c(16,17,18))
crime_pred <- predict(mod_crime, new.df)
list(crime_pred)
```

```
## [[1]]
##         1         2         3
## 238.89524 142.84881  46.80238
```

There is a downward trend of crime in the past but it had more or less stabilized but 2020 was an outlier as can be seen the ggplot. This can also be due to the fact that this analysis only captures the resolved incidents. Based on the data, there is still a downward trend for the next few years despite the outlier 2020.

This data used is biased for crimes that were resolved. If the perp was not found and the case is open (missing data), it was removed from the data used above. The other assumption was the gender differences which was shown to be not true.

**Scan to find out the name of the 75th precinct

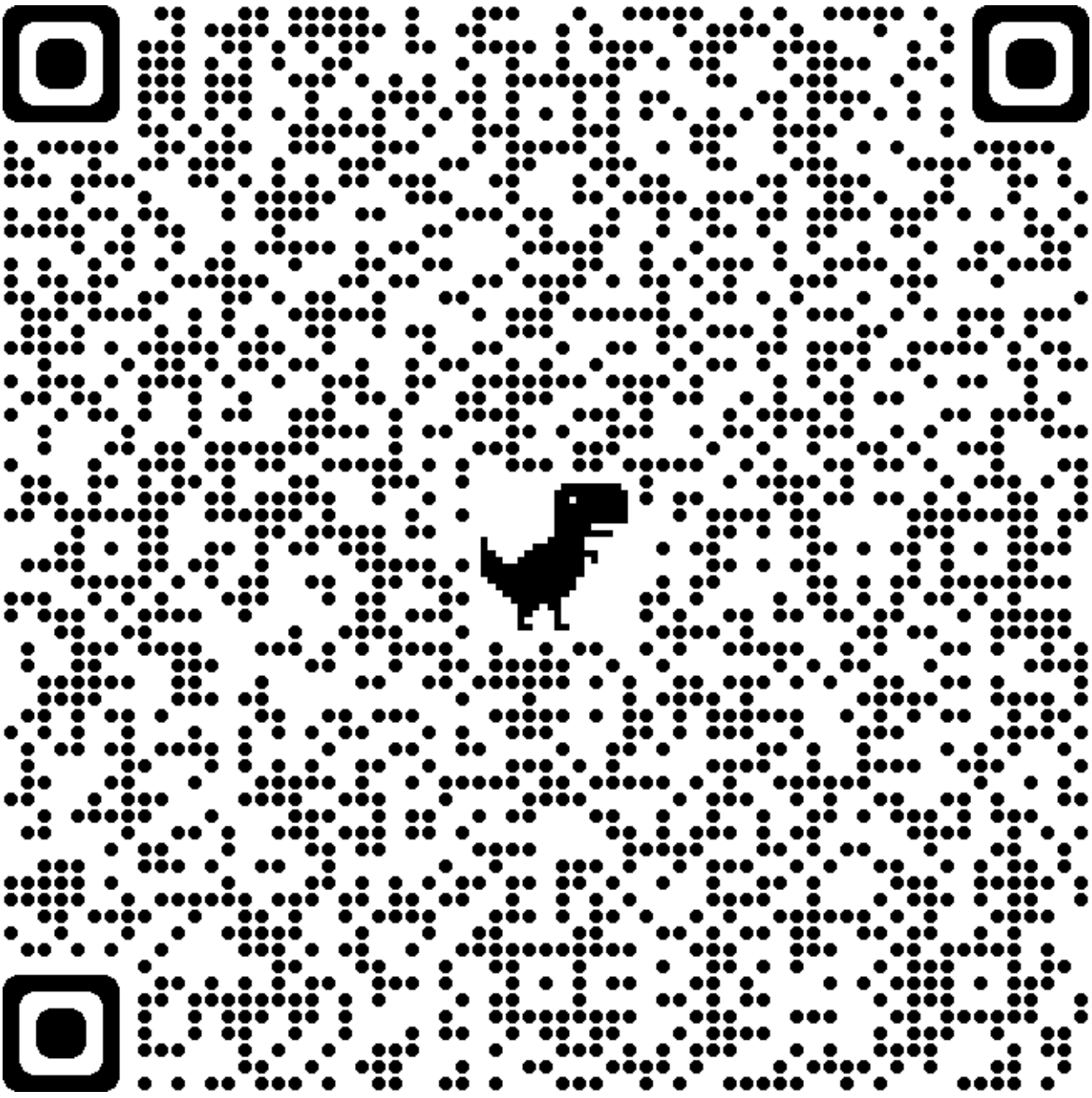


Figure 1: Scan to see the 75th precinct