

Title: "NYPD_Shooting_by_Vic_Age"

Author: "JG"

Date: "2023-08-21"

Output: pdf_document

To begin install packages

tidyverse, knitr, lubridate

Load 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
```

Read csv from url

```
## Rows: 27312 Columns: 21
## -- Column specification -----
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## dbl  (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl  (1): STATISTICAL_MURDER_FLAG
## time (1): OCCUR_TIME
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Clean data

```
# Clean up data
```

```
## Drop unnecessary columns
```

```
nypd_data <- drop_na(nypd_raw_data) %>% select(-c(INCIDENT_KEY, LOCATION_DESC, X_COORD_CD, Y_COORD_CD, I
```

```

## Change the date type
nypd_data <- nypd_data %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))

## Filter out spurious age groups
nypd_data<-nypd_data %>%
  filter(!VIC_AGE_GROUP=="1022")

## Unique values in Victim Age group
unique(nypd_data$VIC_AGE_GROUP)

# Verify no missing values
sum(is.na(nypd_data))

## Change logical boolean into int and add DEATHS and INJURIES columns

nypd_data <- nypd_data %>% mutate(STATISTICAL_MURDER_FLAG = case_when(STATISTICAL_MURDER_FLAG == FALSE ~ 0, STATISTICAL_MURDER_FLAG == TRUE ~ 1))
nypd_data <- nypd_data %>% mutate(DEATHS = case_when(STATISTICAL_MURDER_FLAG == FALSE ~ 0, STATISTICAL_MURDER_FLAG == TRUE ~ 1))
nypd_data <- nypd_data %>% mutate(INJURIES = case_when(STATISTICAL_MURDER_FLAG == FALSE ~ 0, STATISTICAL_MURDER_FLAG == TRUE ~ 1))

nypd_murder_age <- nypd_data %>%
  group_by(VIC_AGE_GROUP, OCCUR_DATE) %>%
  summarize(STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG)) %>%
  select(VIC_AGE_GROUP, OCCUR_DATE, STATISTICAL_MURDER_FLAG) %>%
  ungroup()

```

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

```

nypd_injury_VIC_AGE_GROUP <- nypd_data %>%
  group_by(VIC_AGE_GROUP, OCCUR_DATE) %>%
  summarize(INJURIES = sum(INJURIES)) %>%
  select(VIC_AGE_GROUP, OCCUR_DATE, INJURIES) %>%
  ungroup()

```

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

```

## Add new columns
# Cumulative murders by VIC_AGE_GROUP
nypd_murder_age <- nypd_murder_age %>%
  mutate(CUMMURDER = ave(nypd_murder_age$STATISTICAL_MURDER_FLAG, nypd_murder_age$VIC_AGE_GROUP, FUN = sum))

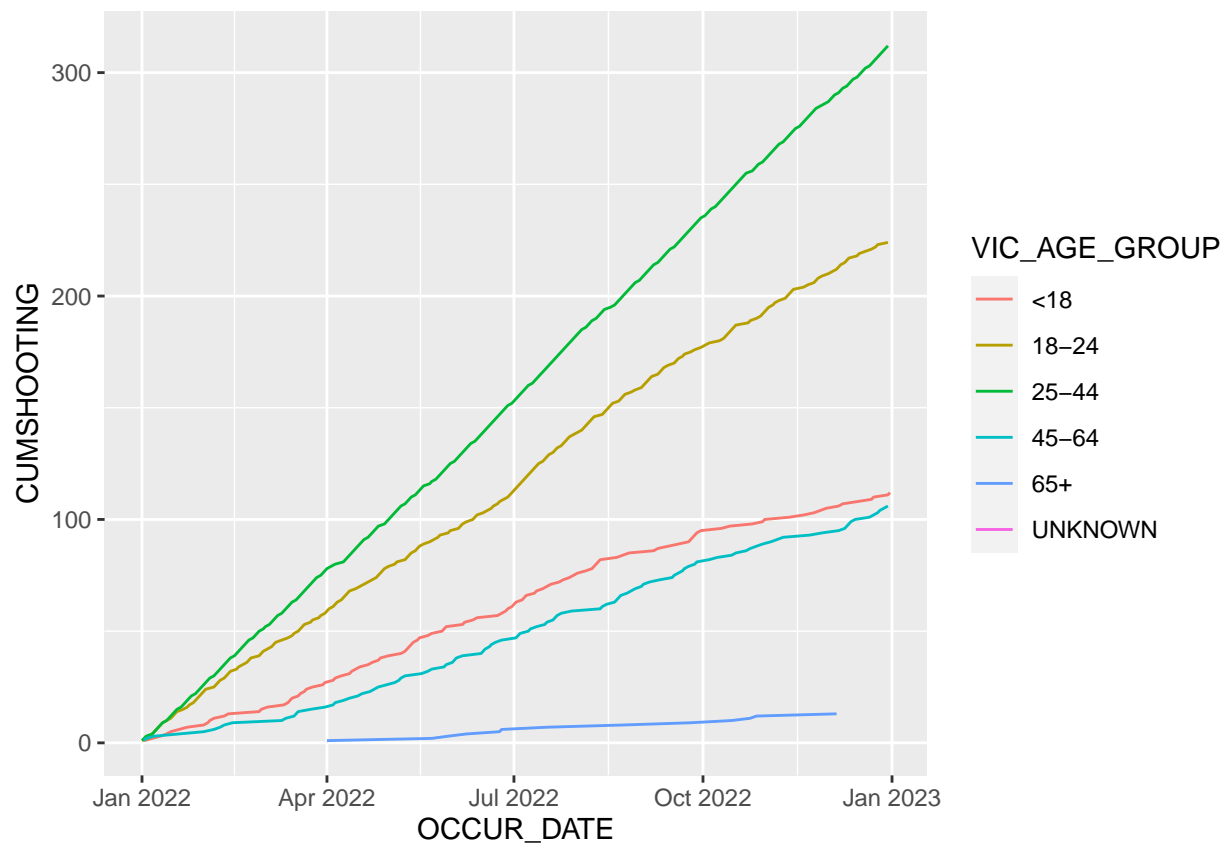
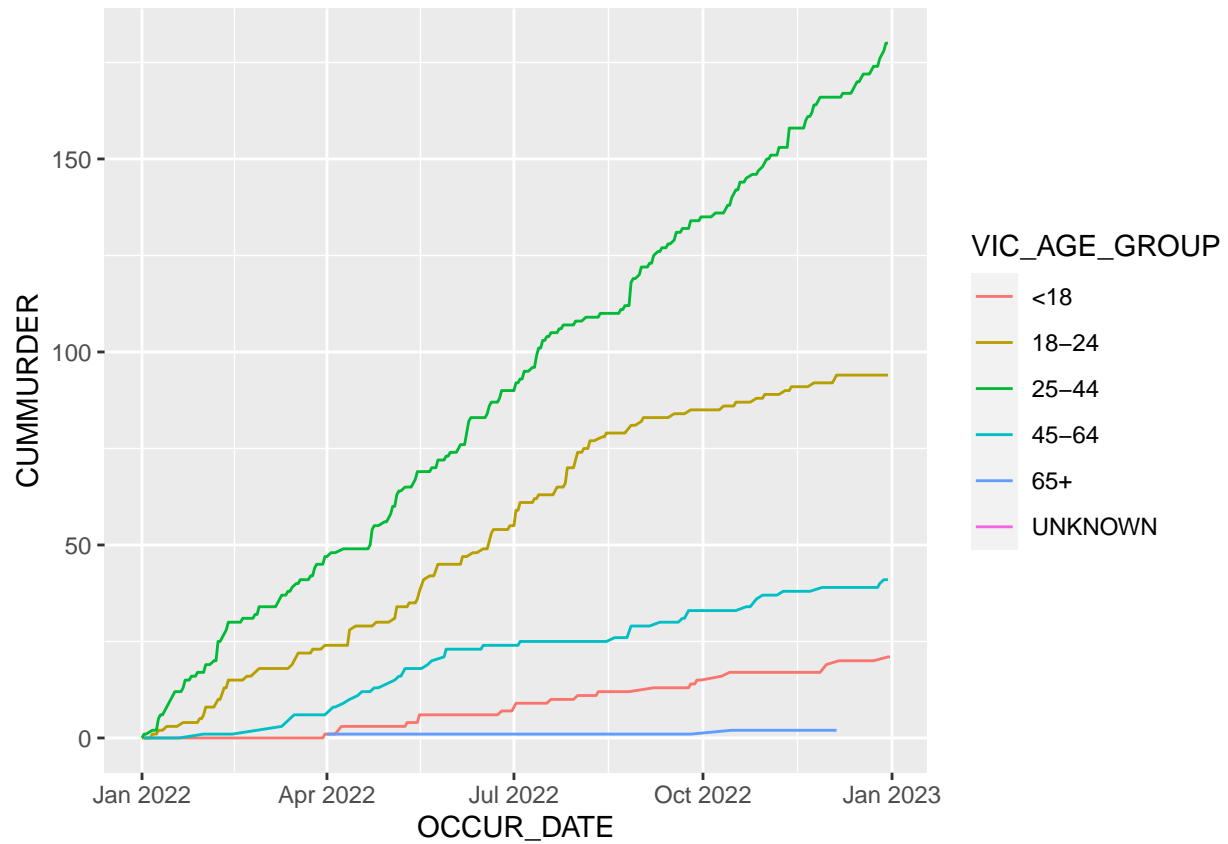
nypd_murder_age$SHOOTING=1

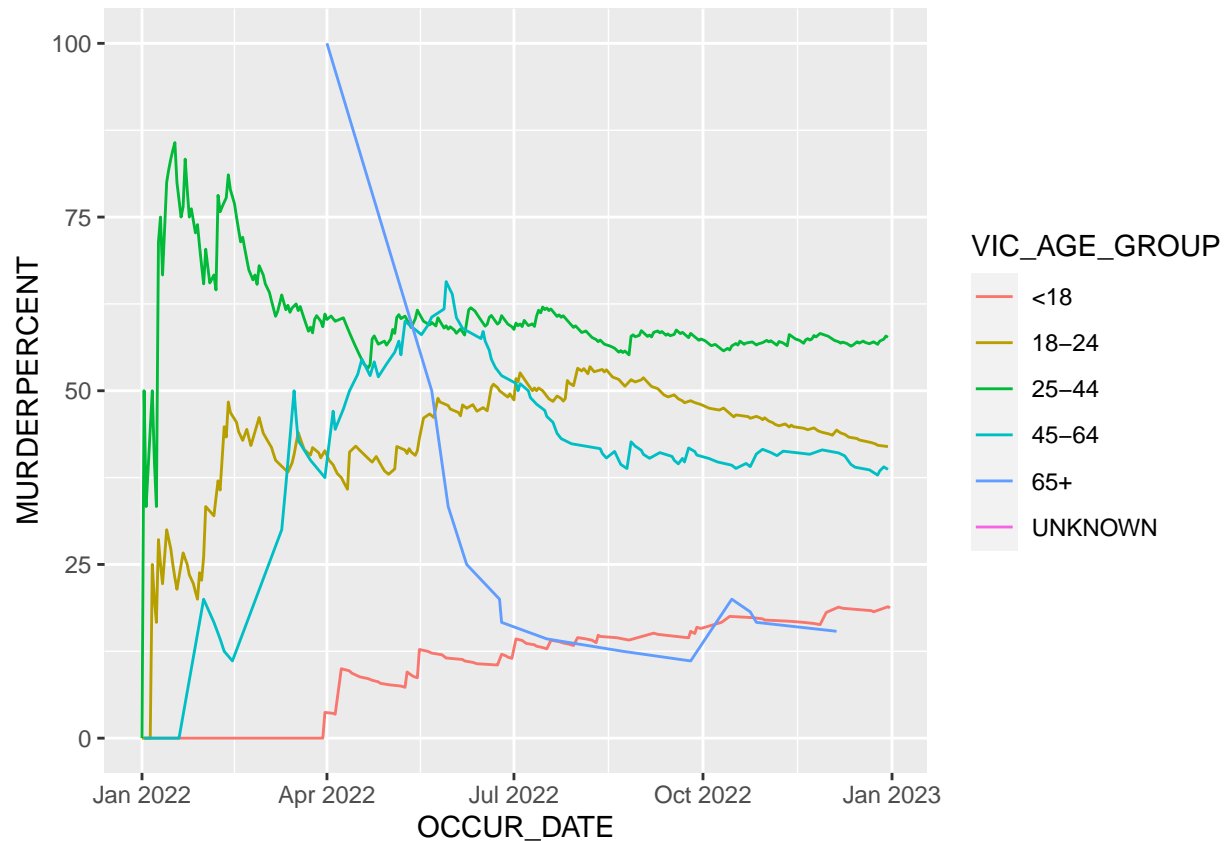
# Cumulative Shootings by VIC_AGE_GROUP
nypd_murder_age <- nypd_murder_age %>%
  mutate(CUMSHOOTING = ave(nypd_murder_age$SHOOTING, nypd_murder_age$VIC_AGE_GROUP, FUN = cumsum))

# Percentage of shootings are murders by VIC_AGE_GROUP
nypd_murder_age$MURDERPERCENT <- with(nypd_murder_age, CUMMURDER/CUMSHOOTING *100)

```

Plot data





Model

Modeling Data

```
mod <- lm(CUMSHOOTING ~ CUMMURDER, data = nypd_murder_age)
summary(mod)
```

```
##
## Call:
## lm(formula = CUMSHOOTING ~ CUMMURDER, data = nypd_murder_age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.621 -15.362  -1.934   6.783  58.018
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19.82590    1.06736   18.57  <2e-16 ***
## CUMMURDER    1.62650    0.01423  114.26  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.33 on 766 degrees of freedom
## Multiple R-squared:  0.9446, Adjusted R-squared:  0.9445
## F-statistic: 1.306e+04 on 1 and 766 DF, p-value: < 2.2e-16
```

```
nypd_murder_age %>% slice_min(CUMSHOOTING)
```

```
## # A tibble: 6 x 7
##   VIC_AGE_GROUP OCCUR_DATE STATISTICAL_MURDER_F~1 CUMMURDER SHOOTING CUMSHOOTING
##   <chr>         <date>          <dbl>      <dbl>    <dbl>      <dbl>
## 1 18-24         2022-01-01            0          0          1          1
## 2 25-44         2022-01-01            0          0          1          1
## 3 45-64         2022-01-02            0          0          1          1
## 4 65+          2022-04-01            1          1          1          1
## 5 <18          2022-01-02            0          0          1          1
## 6 UNKNOWN      2022-12-21            0          0          1          1
## # i abbreviated name: 1: STATISTICAL_MURDER_FLAG
## # i 1 more variable: MURDERPERCENT <dbl>
```

```
nypd_murder_age %>% slice_max(CUMSHOOTING)
```

```
## # A tibble: 1 x 7
##   VIC_AGE_GROUP OCCUR_DATE STATISTICAL_MURDER_F~1 CUMMURDER SHOOTING CUMSHOOTING
##   <chr>         <date>          <dbl>      <dbl>    <dbl>      <dbl>
## 1 25-44         2022-12-30            0        180          1        312
## # i abbreviated name: 1: STATISTICAL_MURDER_FLAG
## # i 1 more variable: MURDERPERCENT <dbl>
```

```
x_grid <- seq(0, 3000)
new_df <- tibble(CUMSHOOTING = x_grid)
nypd_pred <- nypd_murder_age %>% mutate(pred = predict(mod))
```

```
# nypd_pred
nypd_pred %>% ggplot() +
  geom_point(aes(x = OCCUR_DATE, y=CUMSHOOTING), color= "blue")+
  geom_point(aes(x = OCCUR_DATE, y = pred), color = "red")+
  theme(legend.position = "bottom",
        axis.text.x=element_text(angle=90)) +
  labs(title="Cumulative Shootings by age", y=NULL)
```

Cummulative Shootings by age

