NYPD

Ben Lewis

2025-01-10

Step 1: Start an Rmd Document

Start an Rmd document that describes and imports the shooting project dataset in a reproducible manner.

```
library(tidyverse)
```

```
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
df <- read_csv(url)
```

Step 2: Tidy and Transform Your Data

Add to your Rmd document a summary of the data and clean up your dataset by changing appropriate variables to factor and date types and getting rid of any columns not needed. Show the summary of your data to be sure there is no missing data. If there is missing data, describe how you plan to handle it.

```
summary(df)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:28562       Length:28562       Length:28562
##  1st Qu.: 65439914   Class :character   Class1:hms         Class :character
##  Median : 92711254   Mode  :character   Class2:difftime    Mode  :character
##  Mean   :127405824                      Mode  :numeric
##  3rd Qu.:203131993
##  Max.   :279758069
##
##  LOC_OF_OCCUR_DESC     PRECINCT     JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:28562       Min.   :  1.0   Min.   :0.0000    Length:28562
##  Class :character   1st Qu.: 44.0   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 67.0   Median :0.0000    Mode  :character
##                     Mean   : 65.5   Mean   :0.3219
##                     3rd Qu.: 81.0   3rd Qu.:0.0000
##                     Max.   :123.0   Max.   :2.0000
##                                     NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:28562       Mode :logical           Length:28562
##  Class :character   FALSE:23036             Class :character
##  Mode  :character   TRUE :5526              Mode  :character
##
##
```

```
##
##
##    PERP_SEX           PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:28562       Length:28562       Length:28562       Length:28562
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_RACE            X_COORD_CD        Y_COORD_CD        Latitude
##  Length:28562       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000068   1st Qu.:182912   1st Qu.:40.67
##  Mode  :character   Median :1007772   Median :194901   Median :40.70
##                     Mean   :1009424   Mean   :208380   Mean   :40.74
##                     3rd Qu.:1016807   3rd Qu.:239814   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :59
##    Longitude         Lon_Lat
##  Min.   :-74.25   Length:28562
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :59
```

```
df <- df[, c("INCIDENT_KEY", "OCCUR_DATE")]
df$OCCUR_DATE = mdy(df$OCCUR_DATE)
```

```
df$OCCUR_YEAR = year(df$OCCUR_DATE)
df$OCCUR_MONTH = month(df$OCCUR_DATE)
summary(df)
```

```
##   INCIDENT_KEY         OCCUR_DATE           OCCUR_YEAR    OCCUR_MONTH
##  Min.   :  9953245   Min.   :2006-01-01   Min.   :2006   Min.   : 1.000
##  1st Qu.: 65439914   1st Qu.:2009-09-04   1st Qu.:2009   1st Qu.: 5.000
##  Median : 92711254   Median :2013-09-20   Median :2013   Median : 7.000
##  Mean   :127405824   Mean   :2014-06-07   Mean   :2014   Mean   : 6.805
##  3rd Qu.:203131993   3rd Qu.:2019-09-29   3rd Qu.:2019   3rd Qu.: 9.000
##  Max.   :279758069   Max.   :2023-12-29   Max.   :2023   Max.   :12.000
```

```
sum(is.na(df))
```

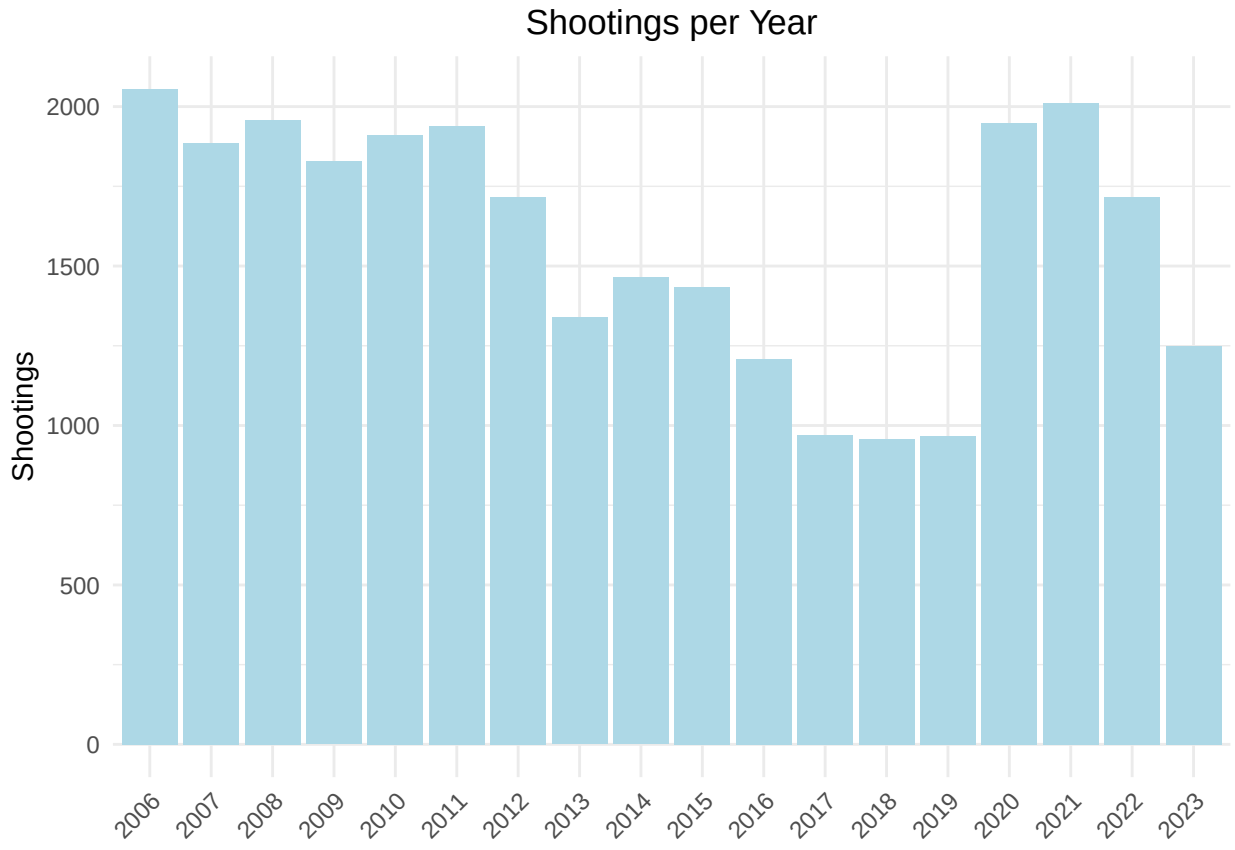
```
## [1] 0
```

Step 3: Add Visualizations and Analysis

Add at least two different visualizations & some analysis to your Rmd. Does this raise additional questions that you should investigate?

```r
year_counts = as.data.frame(table(df$OCCUR_YEAR))
colnames(year_counts) <- c("Year", "Shootings")


ggplot(year_counts, aes(x = Year, y = Shootings)) +
  geom_bar(stat = "identity", fill = "lightblue") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(hjust = 0.5)
  ) +
  labs(
    title = "Shootings per Year",
    x = NULL
  )
```



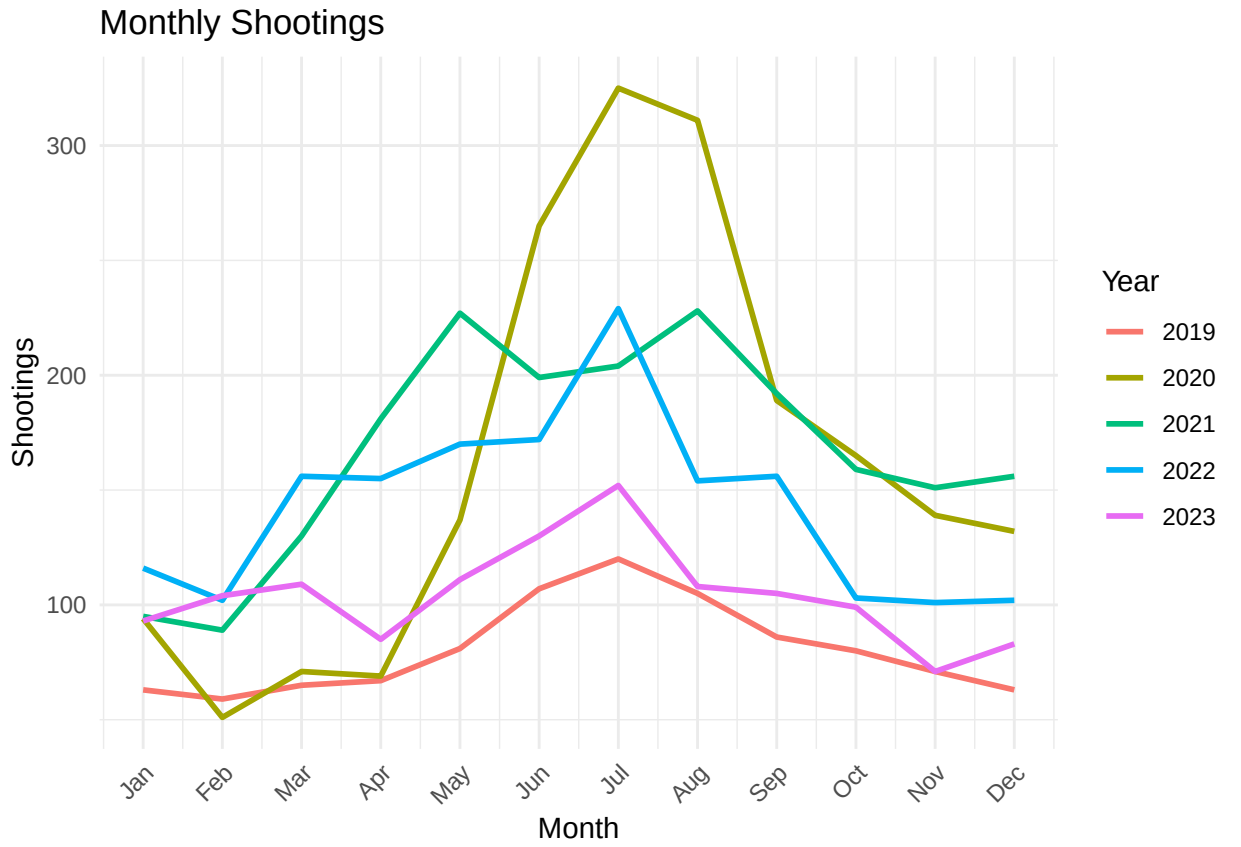

```r
df_2019_to_2023 <- df %>%
  filter(OCCUR_YEAR >= 2019 & OCCUR_YEAR <= 2023)

monthly_shootings <- df_2019_to_2023 %>%
  group_by(OCCUR_YEAR, OCCUR_MONTH) %>%
  summarise(incident_count = n(), .groups = "drop")

ggplot(monthly_shootings, aes(x = OCCUR_MONTH, y = incident_count, color = as.factor(OCCUR_YEAR), group
  geom_line(size = 1) +
```

```r
  scale_x_continuous(breaks = 1:12, labels = month.abb) +  # Show months as abbreviations
  theme_minimal() +
  labs(
    title = "Monthly Shootings",
    x = "Month",
    y = "Shootings",
    color = "Year"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



Step 4: Add Bias Identification

Write the conclusion to your project report and include any possible sources of bias. Be sure to identify what your personal bias might be and how you have mitigated that.