Final Project

2023-04-25

First we will import our datasets

```
library(ggplot2)
library(readr)
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
```

```
#Recorded Police Deaths
police_deaths <- read_csv("/Users/jeremy/Desktop/Final Project/police_deaths_USA_v7.csv")
```

```
## Rows: 25638 Columns: 13
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (11): Rank, Name, Day_Of_Week, Cause, Department, State, Tour, Badge, W...
## dbl   (1): Age
## date  (1): End_Of_Watch
##
## 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.
```

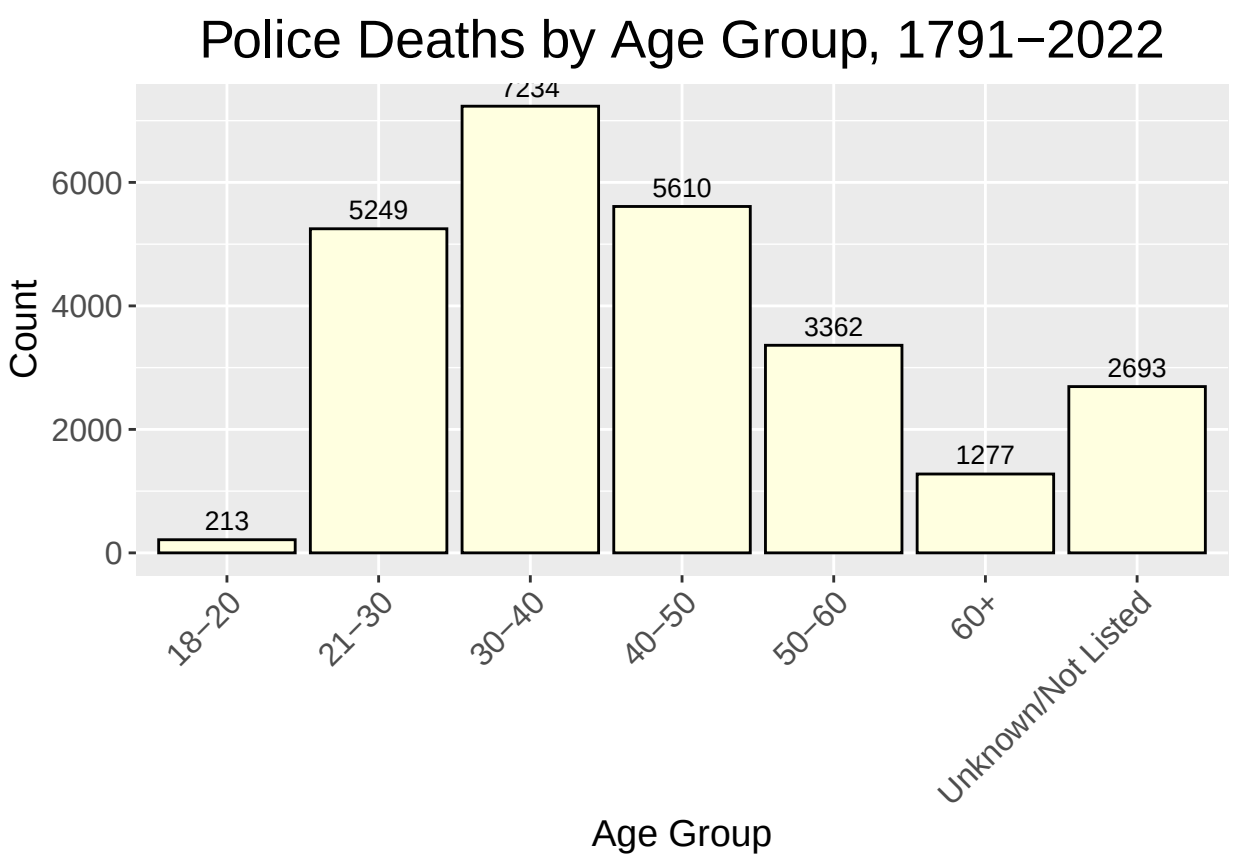
```
#Recorded K9 Deaths
k9_deaths <- read_csv("/Users/jeremy/Desktop/Final Project/k9_deaths_USA_v7.csv")
```

```
## Rows: 506 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): Rank, Name, Breed, Gender, Day_Of_Week, Cause, Department, State,...
## dbl   (1): Age
## date  (1): End_Of_Watch
##
## 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.
```

Identifying total police deaths by age group, 1791-2022

```
library(ggplot2)
```

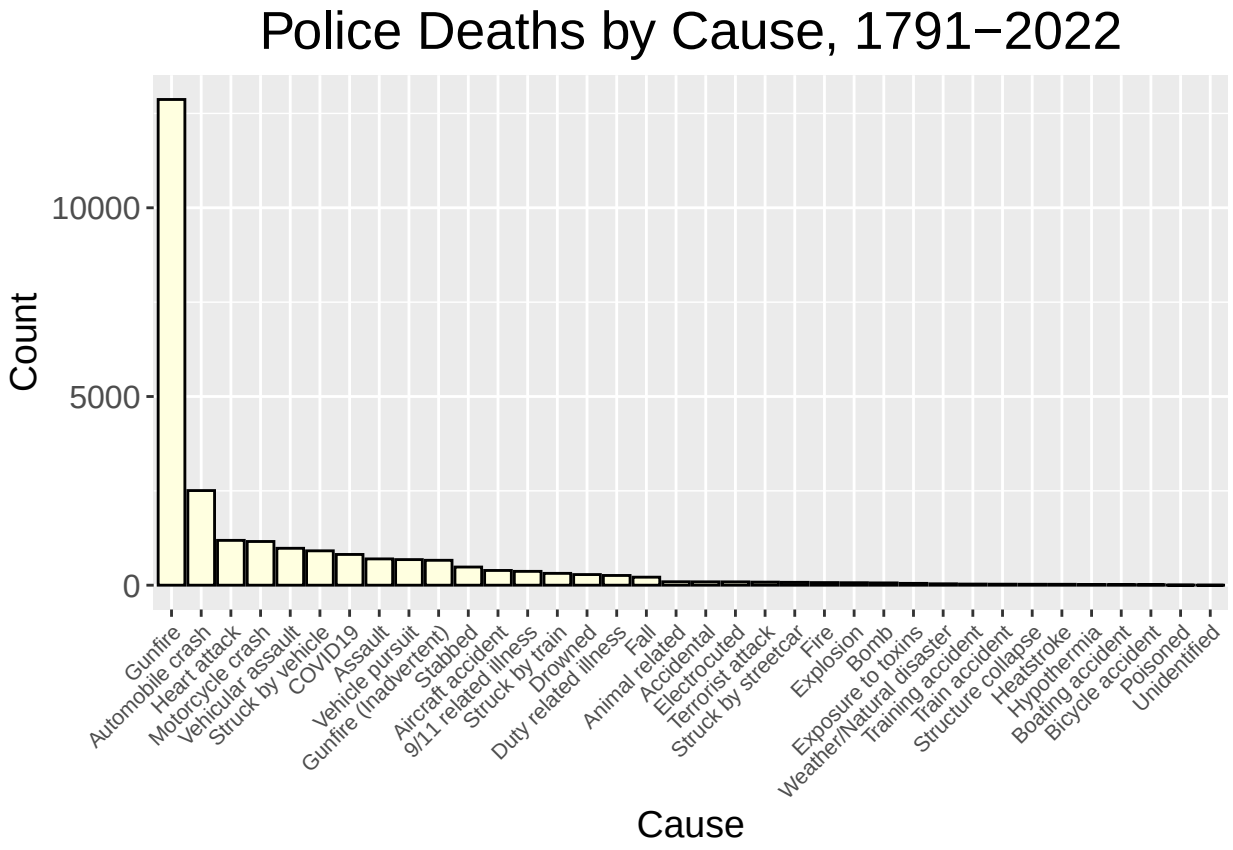
```
age_ranges <- cut(police_deaths$Age, breaks = c(18, 21, 30, 40, 50, 60, Inf),
                  labels = c("18-20", "21-30", "30-40", "40-50", "50-60", "60+"))
barcolor <- "lightyellow"
ggplot(data = police_deaths, aes(x = age_ranges)) +
  geom_bar(fill = barcolor, color = "black") +
  labs(title = "Police Deaths by Age Group, 1791-2022", x = "Age Group", y = "Count") +
  scale_x_discrete(labels = c("18-20","21-30", "30-40", "40-50", "50-60", "60+", "Unknown/Not Listed")) +
  theme(plot.title = element_text(hjust = 0.5, size = 20),
        axis.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 12),
        axis.text.y = element_text(size = 12),
        plot.background = element_rect(fill = "white")) +
  labs(width = 5, height = 9) +
  geom_text(stat = "count", aes(label = after_stat(count)), vjust = -0.5, size = 3.5)
```



Identifying cause of police deaths, 1791-2022

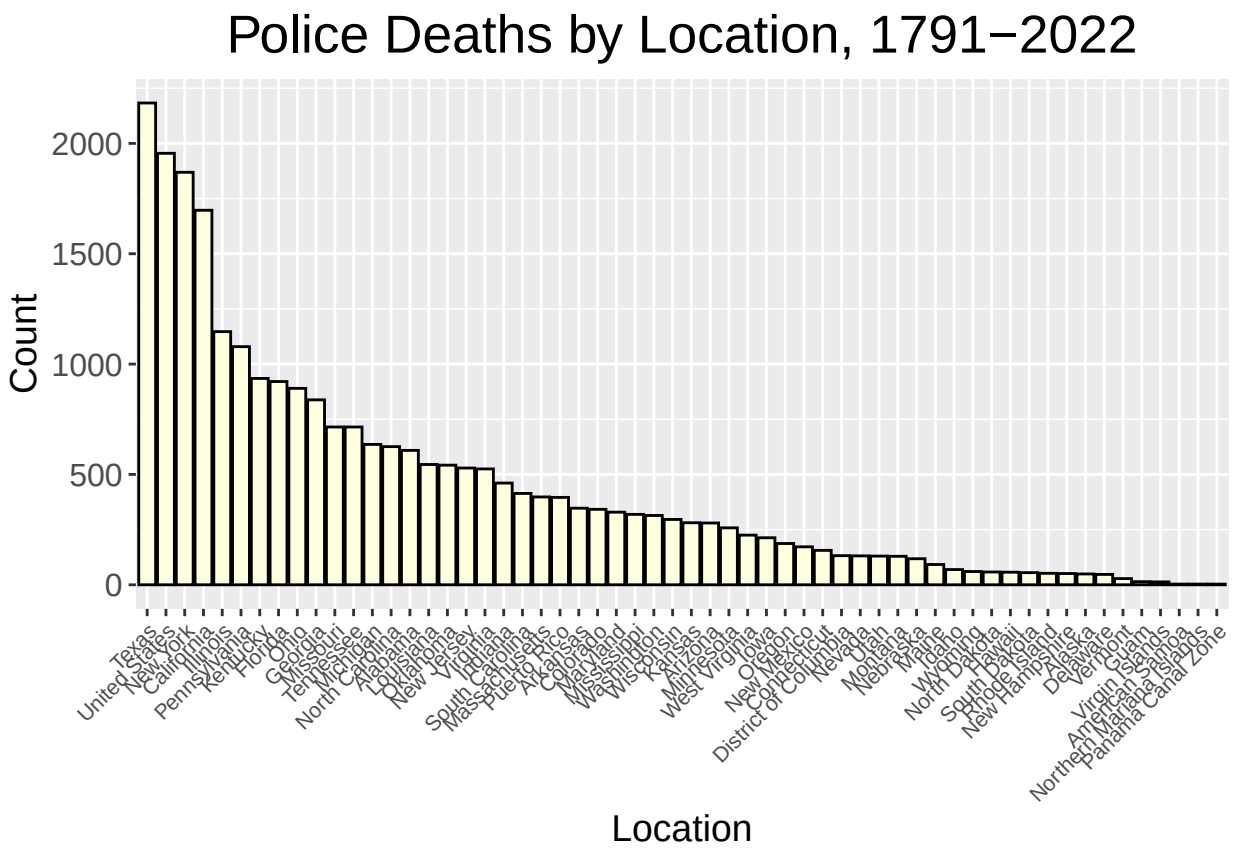
```
cause_freq <- table(police_deaths$Cause)
cause_freq_df <- as.data.frame(cause_freq)

ggplot(cause_freq_df, aes(x=reorder(Var1, -Freq), y=Freq)) +
  geom_bar(fill = "lightyellow", color = "black", stat="identity", position = position_stack(vjust = 0.5
  labs(title = "Police Deaths by Cause, 1791-2022", x = "Cause", y = "Count") +
  theme(plot.title = element_text(hjust = 0.5, size = 20),
        axis.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
```

```
        axis.text.y = element_text(size = 12),
        plot.background = element_rect(fill = "white")) +
  labs(width = 5, height = 9)
```



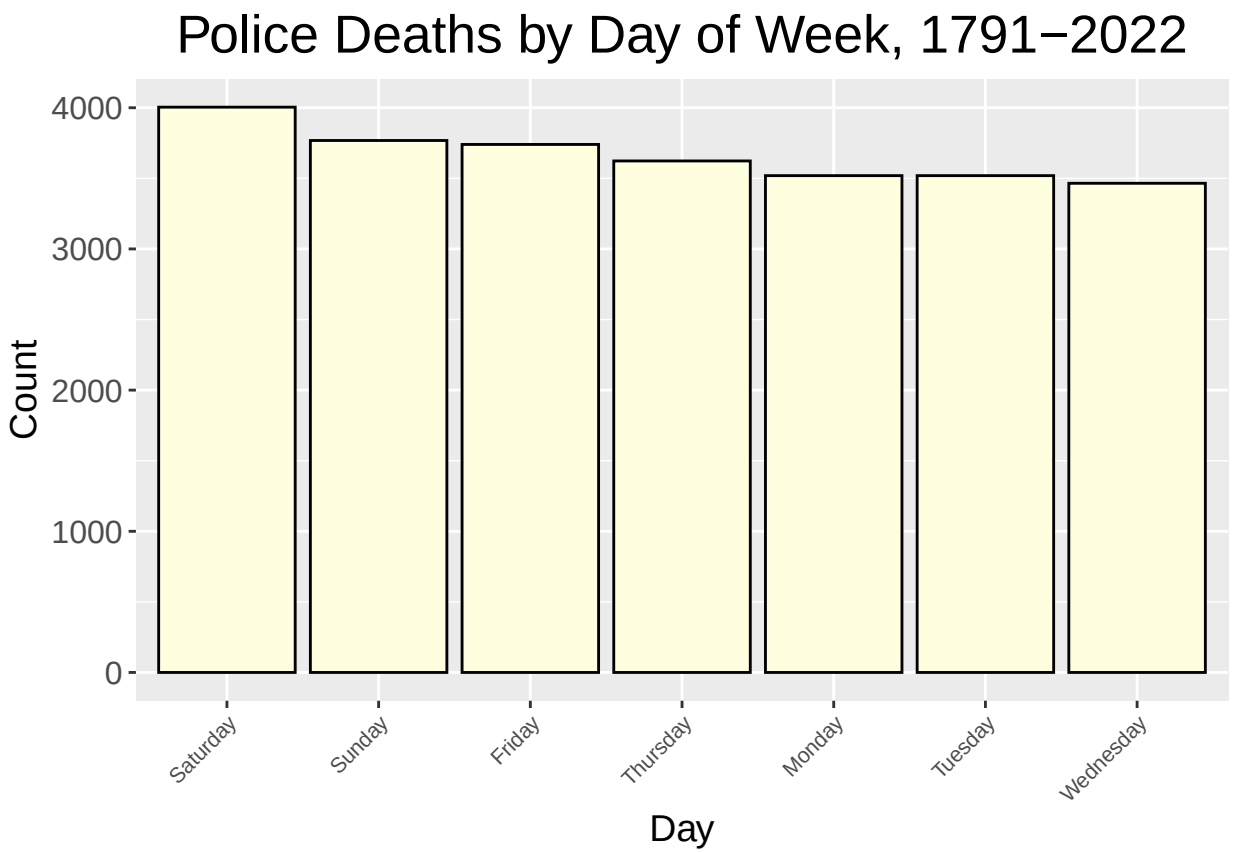
Identifying location of police deaths, 1791-2022

```
state_freq <- table(police_deaths$State)
state_freq_df <- as.data.frame(state_freq)

ggplot(state_freq_df, aes(x=reorder(Var1, -Freq), y=Freq)) +
  geom_bar(fill = "lightyellow", color = "black", stat="identity", position = position_stack(vjust = 0.5
  labs(title = "Police Deaths by Location, 1791-2022", x = "Location", y = "Count") +
  theme(plot.title = element_text(hjust = 0.5, size = 20),
        axis.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(size = 12),
        plot.background = element_rect(fill = "white")) +
  labs(width = 5, height = 9)
```



Identifying day of week of police deaths,l 1791-2022

```
day_freq <- table(police_deaths$Day_Of_Week)
day_freq_df <- as.data.frame(day_freq)

ggplot(day_freq_df, aes(x=reorder(Var1, -Freq), y=Freq)) +
  geom_bar(fill = "lightyellow", color = "black", stat="identity", position = position_stack(vjust = 0.5
  labs(title = "Police Deaths by Day of Week, 1791-2022", x = "Day", y = "Count") +
  theme(plot.title = element_text(hjust = 0.5, size = 20),
        axis.title = element_text(size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(size = 12),
        plot.background = element_rect(fill = "white")) +
  labs(width = 5, height = 9)
```



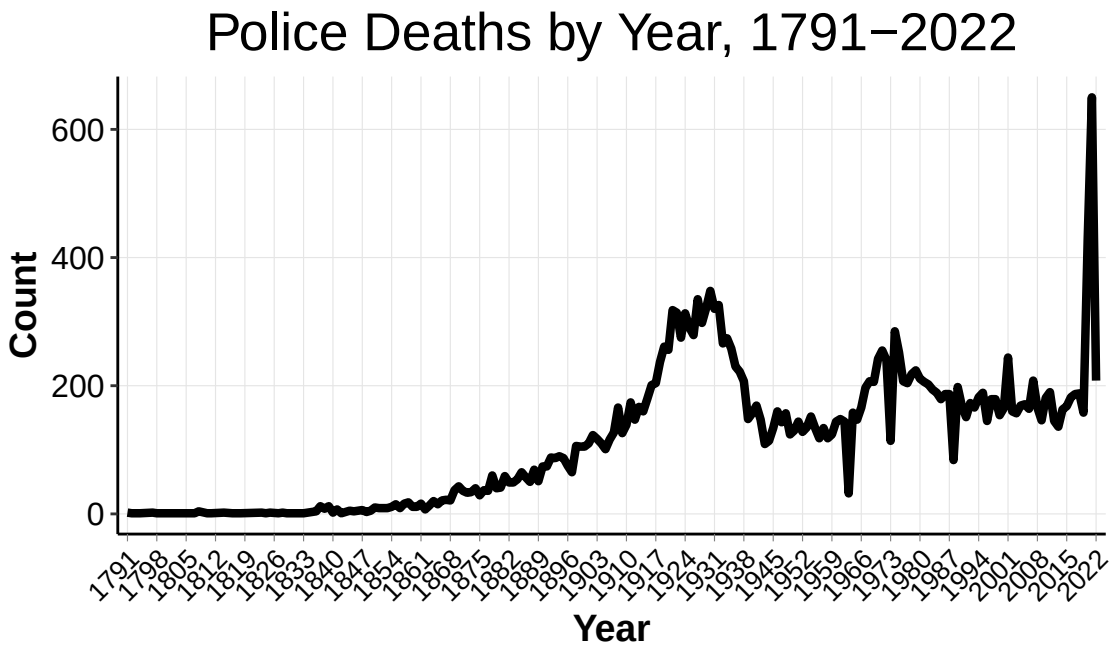
Identifying police deaths over time:

```
library(dplyr)

police_deaths$End_Of_Watch <- as.Date(police_deaths$End_Of_Watch)

death_count_by_year_covid <- police_deaths %>%
  group_by(year = lubridate::year(End_Of_Watch)) %>%
  summarize(count = n())

ggplot(death_count_by_year_covid, aes(x = year, y = count)) +
  geom_line(color = "black", size = 1.5) +
  labs(x = "Year", y = "Count", title = "Police Deaths by Year, 1791-2022") +
  theme_classic() +
  theme(plot.title = element_text(hjust = 0.5, size = 20),
        axis.title = element_text(size = 14, face = "bold"),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10, color = "black"),
        axis.text.y = element_text(size = 12, color = "black"),
        plot.background = element_rect(fill = "white"),
        panel.grid.major.x = element_line(color = "gray90", size = 0.2),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(color = "gray90", size = 0.2),
        panel.grid.minor.y = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank(),
        legend.position = "none",
        plot.margin = unit(c(1,1,2,1), "cm")) +
  scale_x_continuous(breaks = seq(1791, 2022, by = 7), expand = c(0.01, 0)) +
```

```
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10, color = "black"),
        axis.ticks.x = element_line(color = "gray50", size = 0.2))
```

```
## 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.
```

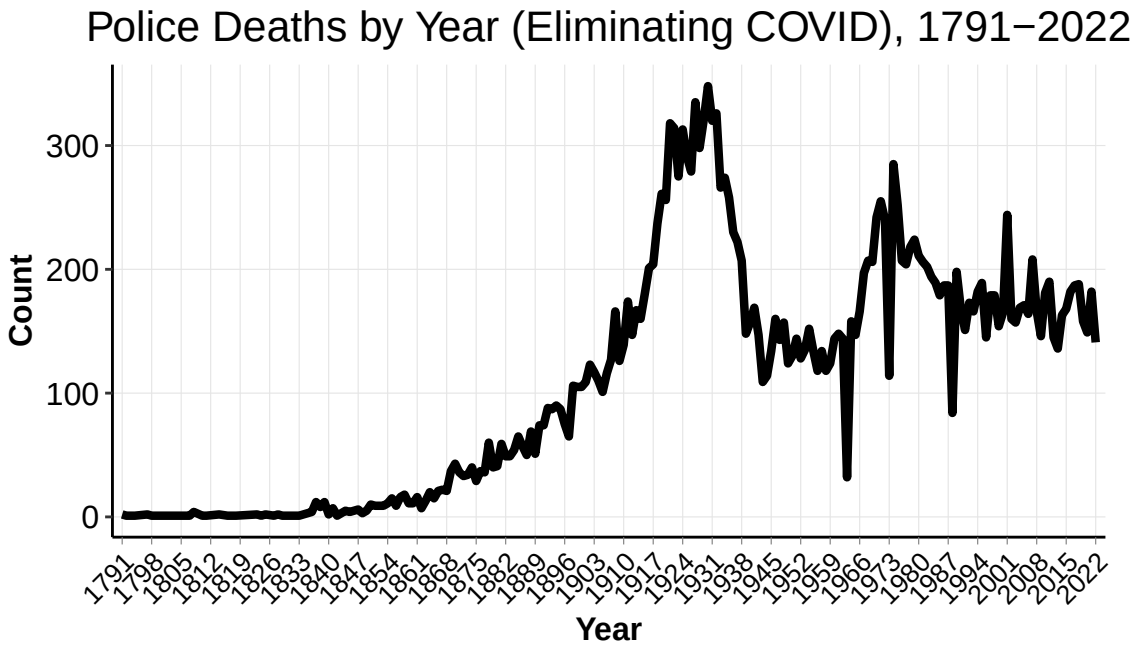
```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Identifying police deaths by year eliminating COVID-19 from cause.

```
police_deaths_filtered <- police_deaths %>%
  filter(Cause != "COVID19")

police_deaths_filtered$End_Of_Watch <- as.Date(police_deaths_filtered$End_Of_Watch)

death_count_by_year <- police_deaths_filtered %>%
  group_by(year = lubridate::year(End_Of_Watch)) %>%
  summarize(count = n())

ggplot(death_count_by_year, aes(x = year, y = count)) +
  geom_line(color = "black", size = 1.5) +
  labs(x = "Year", y = "Count", title = "Police Deaths by Year (Eliminating COVID), 1791-2022") +
  theme_classic() +
  theme(plot.title = element_text(hjust = 0.5, size = 16),
        axis.title = element_text(size = 12, face = "bold"),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10, color = "black"),
```

```
        axis.text.y = element_text(size = 12, color = "black"),
        plot.background = element_rect(fill = "white"),
        panel.grid.major.x = element_line(color = "gray90", size = 0.2),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(color = "gray90", size = 0.2),
        panel.grid.minor.y = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank(),
        legend.position = "none",
        plot.margin = unit(c(1,1,2,1), "cm")) +
  scale_x_continuous(breaks = seq(1791, 2022, by = 7), expand = c(0.01, 0)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10, color = "black"),
        axis.ticks.x = element_line(color = "gray50", size = 0.2))
```



Plotting the two against one another (covid vs. no covid)

```
ggplot() +
  geom_line(data = death_count_by_year, aes(x = year, y = count), color = "green", size=1) +
  geom_line(data = death_count_by_year_covid, aes(x = year, y = count), color = "red", size=1) +
  labs(x = "Year", y = "Count", title = "Police Deaths by Year, 2018-2022") +
  theme_classic() +
  xlim(2018, 2022) +
  theme(plot.title = element_text(hjust = 0.5, size = 16),
        axis.title = element_text(size = 12, face = "bold"),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10, color = "black"),
        axis.text.y = element_text(size = 12, color = "black"),
        plot.background = element_rect(fill = "white"),
        panel.grid.major.x = element_line(color = "gray90", size = 0.2),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(color = "gray90", size = 0.2),
        panel.grid.minor.y = element_blank(),
        panel.border = element_blank(),
```

```
        panel.background = element_blank(),
        legend.position = "none",
        plot.margin = unit(c(1,1,2,1), "cm"))
```

```
## Warning: Removed 205 rows containing missing values (`geom_line()`).
## Removed 205 rows containing missing values (`geom_line()`).
```