JF_assignment2

Jeremy Fields

2026-01-29
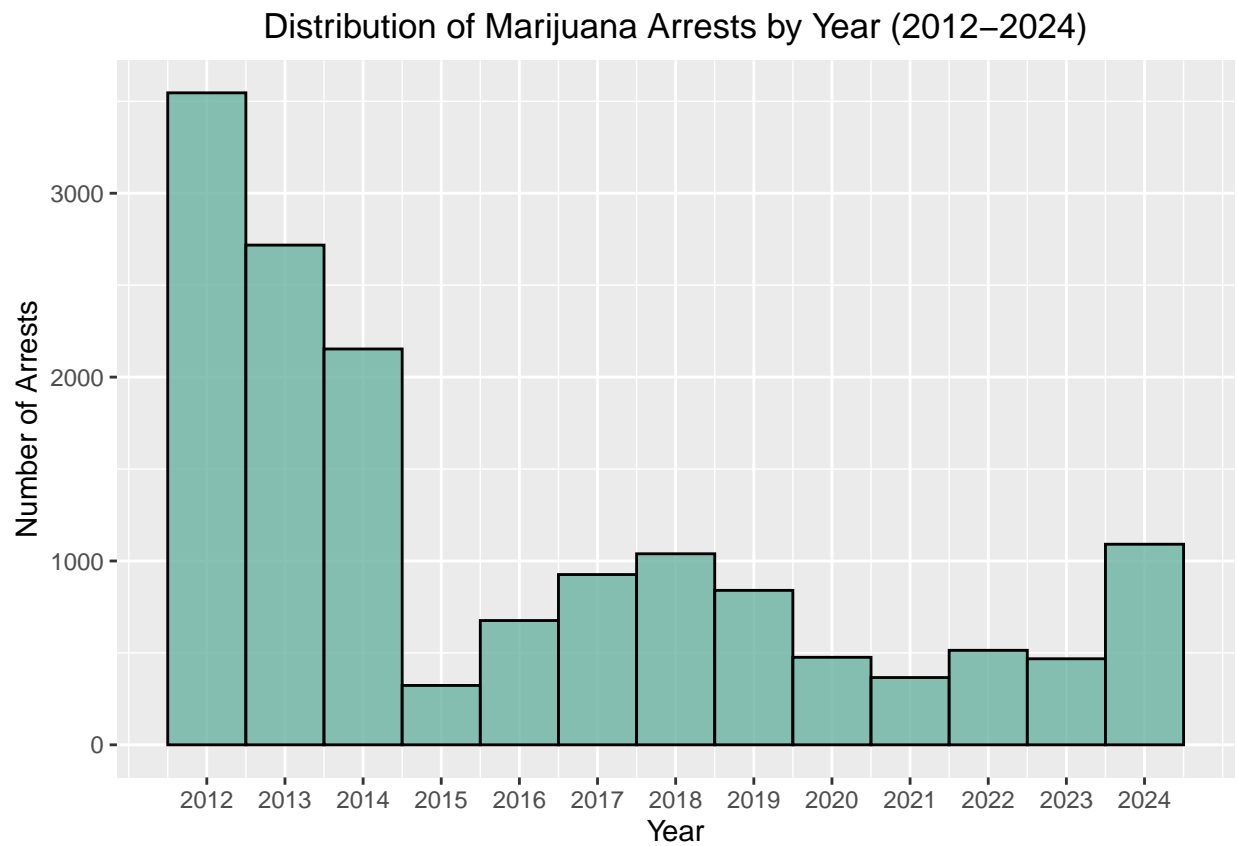
Distribution of Marijuana Arrests by Year (2012–2024)

Figure 1 - Histogram

- Provides the clearest view of arrest volume
- Distribution shows largest concentration of arrests between 2012-2014
- Distribution shows a large drop in 2015 (coinciding with legalization under Initiative 71)
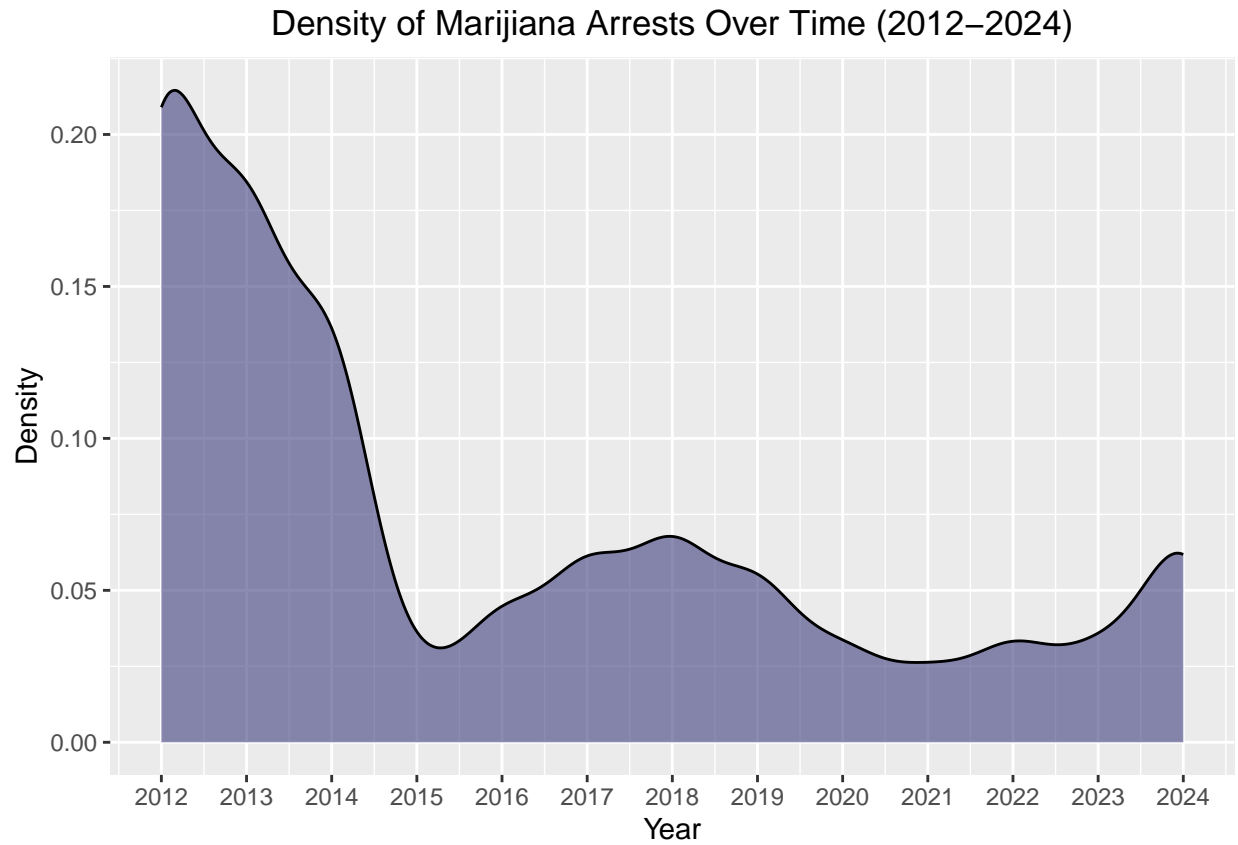
Figure 2 - Density plot

- This density plot shows the shape of the distribution
- Emphasizes the high probability of an arrest occuring in the 2012-2014 range
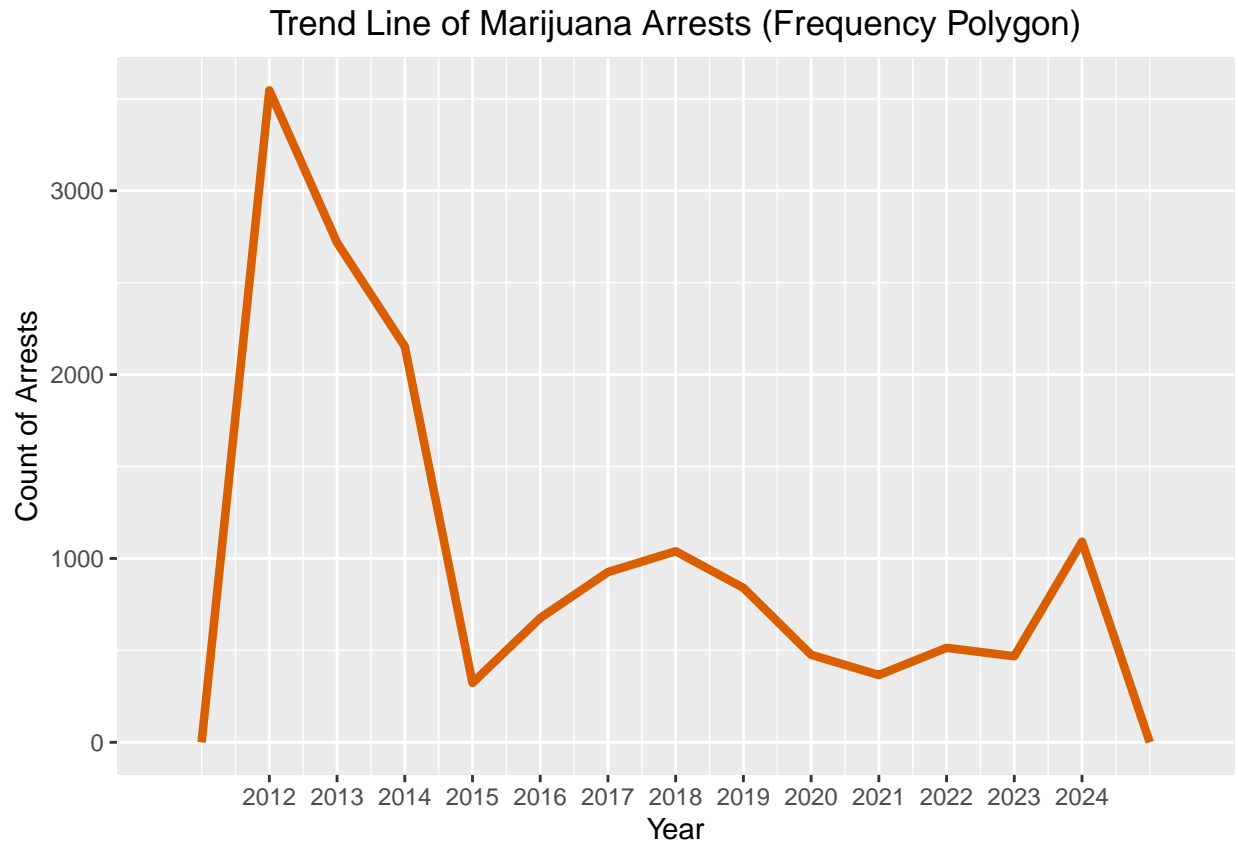
Figure 3 - Frequency polygon

- The frequency polygon plot provides a clean look at the slope of decline
- This provides useful in order to identify and visualize the rate of change, specifically the sharp downward slope occurring between 2014 and 2015

Appendix: Code

```r
# Load data
m_arrests_dc <- read.csv("~/INFO526 - Data Viz/Marijuana_Arrests.csv",
                         comment.char="#")

# Import libraries
library(dplyr)
library(ggplot2)

# Cleaning logic
m_cleaned <- m_arrests_dc %>%
  mutate(
    # Clean RACE: Trim spaces and group unknowns
    Race_Clean = trimws(RACE),
    Race_Status = case_when(
      is.na(Race_Clean) | Race_Clean == "" | Race_Clean == "U" ~ "Unknown",
      TRUE ~ "Valid"
    ),

    # Clean ETHNICITY: Trim spaces and group unknowns
    Eth_Clean = trimws(ETHNICITY),
    Eth_Status = case_when(
      is.na(Eth_Clean) | Eth_Clean == "" | Eth_Clean == "U" ~ "Unknown",
      TRUE ~ "Valid"
    ),

    # Clean SEX: Group unknowns
    Sex_Status = case_when(
      is.na(SEX) | SEX == "" | SEX == "U" ~ "Unknown",
      TRUE ~ "Valid"
    )
  ) %>%
  # filter for the relevant years
  filter(!is.na(YEAR), YEAR >= 2012, YEAR <= 2024)
# First plot (histogram)
ggplot(m_cleaned, aes(x = YEAR)) +
  geom_histogram(binwidth = 1, fill = "#69b3a2", color = "black", alpha = 0.8) +
  scale_x_continuous(breaks = seq(2012, 2024, 1)) +
  labs(title = "Distribution of Marijuana Arrests by Year (2012-2024)",
       x = "Year",
       y = "Number of Arrests") +
  theme(plot.title = element_text(hjust = 0.5))
# second plot (density plot)
ggplot(m_cleaned, aes(x = YEAR)) +
  geom_density(fill = "#404080", alpha = 0.6) +
  scale_x_continuous(breaks = seq(2012, 2024, 1)) +
  labs(title = "Density of Marijiana Arrests Over Time (2012-2024)",
       x = "Year",
       y = "Density") +
  theme(plot.title = element_text(hjust = 0.5))
# third plot (frequency polygon)
ggplot(m_cleaned, aes(x = YEAR)) +
```

```r
geom_freqpoly(binwidth = 1, color = "#d95f02", size = 1.5) +
scale_x_continuous(breaks = seq(2012, 2024, 1)) +
labs(title = "Trend Line of Marijuana Arrests (Frequency Polygon)",
     x = "Year",
     y = "Count of Arrests") +
theme(plot.title = element_text(hjust = 0.5))
```