Trends in U.S. County-Level Crime Rates (2002-2014)

Will Clary , Canace Finley & Naa Adoley Acquaye

2024-11-11

Project Overview

This report analyzes crime patterns in the United States from 2002 to 2014, examining both violent and property crimes. The analysis includes visualization of trends, patterns, and distributions of different crime types over the years.

Dataset Description

- Source: National Neighborhood Data Archive (NaNDA)
- Years: 2002-2014
- Coverage: All U.S. counties (50 states including Alaska and Hawaii)
- Type: FBI Part I criminal offenses
- Access: Inter-university Consortium for Political and Social Research (ICPSR)

Key Variables

- Geographic Identifiers:
  - STCOFIPS (Five-digit county code)
  - State and County FIPS codes
- Crime Categories:
  - Violent Crimes (Murder, Rape, Robbery, Assault)
  - Property Crimes (Burglary, Larceny, Vehicle Theft, Arson)
- Population Data:
  - County population figures
  - Agency reporting statistics

Target Audience

- Law Enforcement:
  - Resource allocation
  - Strategic planning
  - Prevention initiatives
- Policymakers:
  - Evidence-based decisions
  - Trend analysis
  - Policy effectiveness
- Community Members:
  - Local crime awareness
  - Informed decision-making
  - Safety planning

Learning Objectives

- Analyze crime rate evolution (2002-2014)
- Compare violent vs property crimes
- Identify high/low crime areas
- Explore population-crime correlations
- Examine temporal patterns

Data Preparation

First, the necessary libraries will be loaded.

```
# Load required libraries
library(tidyverse)
library(readr)
library(ggplot2)
library(dplyr)
library(viridis)
library(scales)
library(sf)
library(tigris)
library(leaflet)
library(dplyr)
library(viridis)
library(shiny)
library(plotly)
library(maps)
library(gridExtra)
library(egg)
library(RColorBrewer)
library(shinyWidgets)# For pickerInput
library(magrittr)
library(DT)
#install.packages("shinyWidgets")
```

```
load("C:/Users/masho/OneDrive - University of Denver/DATA VISUALIZATION/38649-0001-Data.rda")
Crime_data <- da38649.0001

# Checking data structure
glimpse(Crime_data)
```

```
## Rows: 41,118
## Columns: 25
## $ STCOFIPS <fct> 01001, 01001, 01001, 01001, 01001, 01001, 01001, 01001, 01001…
## $ FIPS_ST  <fct> 01, 01, 01, 01, 01, 01, 01, 01, 01, 01, 01, 01, 01, 01, 01, 0…
## $ FIPS_CTY <fct> 001, 001, 001, 001, 001, 001, 001, 001, 001, 001, 001, 001, 0…
## $ YEAR     <dbl> 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2…
## $ STUDYNO  <dbl> 4009, 4360, 4466, 4717, 23780, 25114, 27644, 30763, 33523, 34…
## $ EDITION  <dbl> 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2…
## $ PART     <fct> (4) Crimes-County, (4) Crimes-County, (4) Crimes-County, (4) …
## $ IDNO     <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2…
## $ CPOPARST <dbl> 44959, 46808, 47929, 48917, 50316, 52092, 52417, 52947, 56219…
## $ CPOPCRIM <dbl> 24424, 45749, 45898, 47669, 49052, 50582, 50884, 51369, 54571…
## $ AG_ARRST <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 13, 13, 13, 13, 13, 13…
## $ AG_OFF   <dbl> 2, 4, 2, 3, 4, 3, 4, 4, 4, 4, 3, 3, 3, 10, 13, 10, 9, 7, 13, …
## $ COVIND   <dbl> 100.0000, 100.0000, 100.0000, 100.0000, 100.0000, 100.0000, 9…
## $ INDEX    <dbl> 1344, 1829, 1758, 1727, 1814, 1804, 1786, NA, NA, NA, NA, NA,…
## $ MODINDX  <dbl> 1357, 1832, 1761, 1728, 1817, 1809, 1799, NA, NA, NA, NA, NA,…
## $ VIOL     <dbl> NA, NA, NA, NA, NA, NA, NA, 210, 143, 134, 139, 153, 144, NA,…
## $ PROPERTY <dbl> NA, NA, NA, NA, NA, NA, NA, 1279, 1521, 1698, 1765, 1667, 162…
## $ MURDER   <dbl> 0, 3, 0, 0, 1, 1, 3, 0, 0, 2, 3, 2, 2, 6, 1, 1, 5, 1, 10, 5, …
## $ RAPE     <dbl> 18, 29, 11, 15, 16, 14, 20, 17, 14, 13, 15, 38, 18, 56, 80, 4…
## $ ROBBERY  <dbl> 35, 34, 38, 28, 49, 35, 24, 32, 26, 36, 34, 31, 37, 39, 44, 3…
## $ AGASSLT  <dbl> 48, 95, 103, 67, 85, 66, 71, 161, 103, 83, 87, 82, 87, 193, 2…
## $ BURGLRY  <dbl> 172, 326, 358, 356, 343, 393, 379, 190, 211, 361, 447, 397, 3…
## $ LARCENY  <dbl> 1023, 1235, 1162, 1149, 1226, 1194, 1175, 989, 1232, 1223, 12…
## $ MVTHEFT  <dbl> 48, 107, 86, 112, 94, 101, 114, 100, 78, 114, 85, 121, 129, 1…
## $ ARSON    <dbl> 13, 3, 3, 1, 3, 5, 13, 0, 1, 7, 108, 4, 14, 0, 2, 0, 0, 0, 3,…
```

Missing Value Analysis

```
# Check for missing values
#is.na(Crime_data)
missing_values <- colSums(is.na(Crime_data))
print(missing_values[missing_values > 0])
```

```
##    INDEX  MODINDX     VIOL PROPERTY
##    19064    19064    22054    22054
```

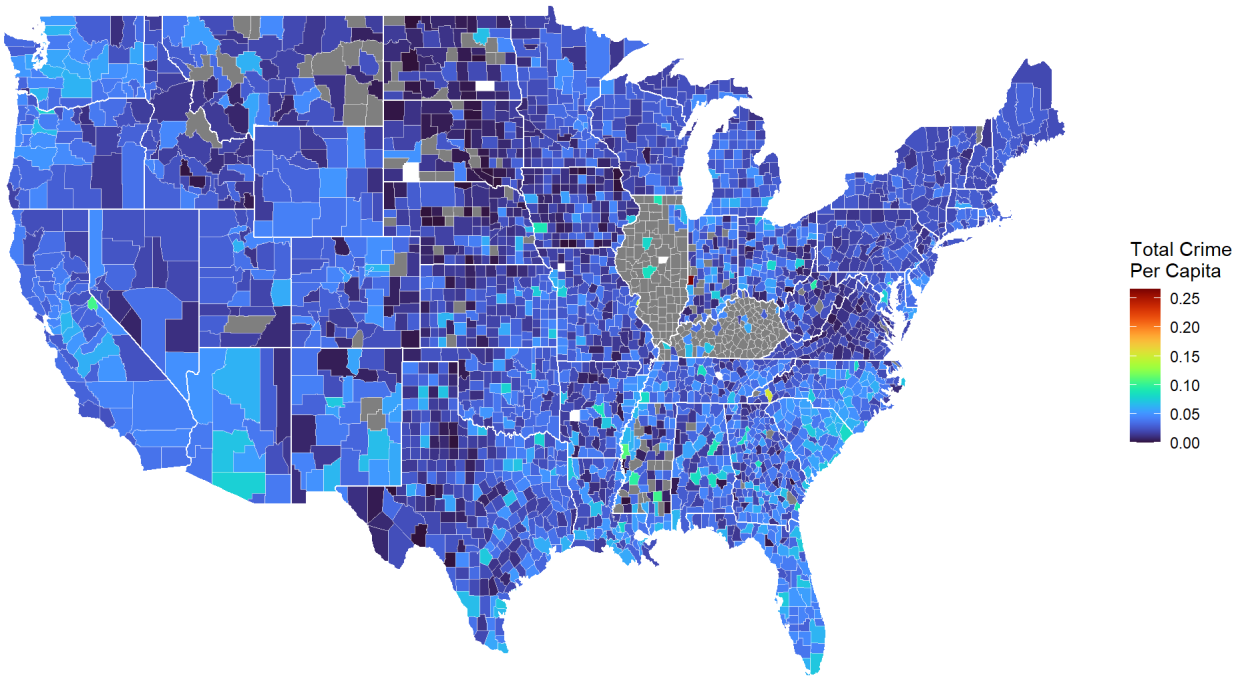
Geographic Crime Analysis

Crime Rate Maps

2002 Crime Rate Map | 2008 Crime Rate Map | 2014 Crime Rate Map

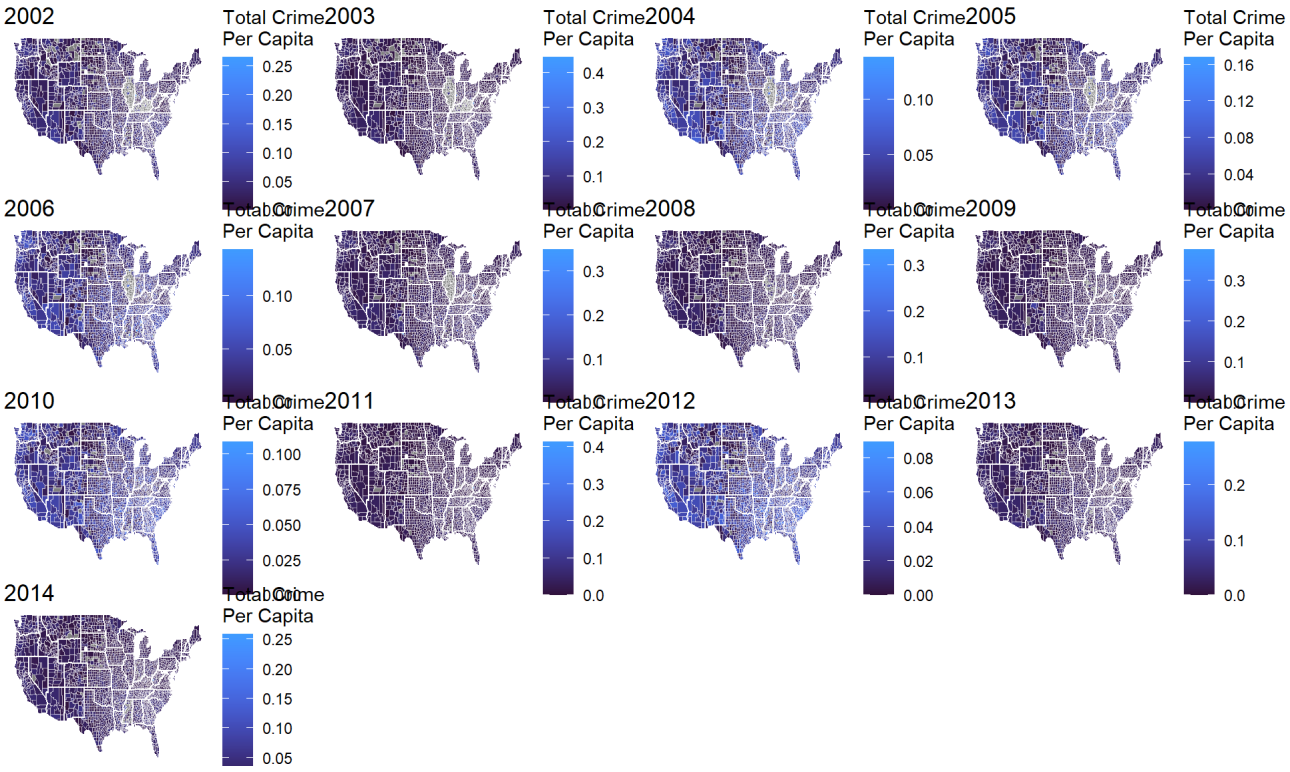
```
Main_States <- map_data("county")
FIPSData <- read.csv("C:/Users/masho/Downloads/US_FIPS_Codes_final (1).csv", colClasses = c(STCOFIPS = "character"))
MergedCounties <- inner_join(Main_States, FIPSData, by = c("subregion", "region"))
CountyCrime <- inner_join(MergedCounties, Crime_data, by = "STCOFIPS")
StateOutline <- map_data("state")

Crime02 <- CountyCrime %>%
  filter(YEAR == 2002) %>%
  group_by(STCOFIPS) %>%
  mutate(TOTALCRIME = MURDER + RAPE + ROBBERY + AGASSLT + BURGLRY + LARCENY +
           MVTHEFT + ARSON) %>%
  mutate(PerCapita02 = TOTALCRIME/CPOPCRIM)

ggplot(data = Crime02, aes(long, lat, group = group)) +
  geom_polygon(aes(fill = PerCapita02),
               colour = alpha("white", 1/2), linewidth = 0.2) +
  geom_polygon(data = StateOutline, aes(x = long, y = lat, group = group),
               color = "white", fill = NA) +
  scale_fill_viridis(labels = scales::comma,
                     name = "Total Crime\nPer Capita",
                     option = "H", begin = 0, end = 1) +
  theme_void() +
  ggtitle("2002 Crime Rate per Capita")
```

2002 Crime Rate per Capita



Animated Crime Rate Maps

```
# Create dynamic map for all years
years <- unique(CountyCrime$YEAR)
years_plot <- list()

crime_year <- CountyCrime %>%
  group_by(STCOFIPS) %>%
  mutate(TOTALCRIME = MURDER + RAPE + ROBBERY + AGASSLT + BURGLRY + LARCENY +
           MVTHEFT + ARSON) %>%
  mutate(PerCapita = TOTALCRIME/CPOPCRIM)

for(YEAR_ in years) {
  years_plot[[as.character(YEAR_)]] <- ggplot(
    data = crime_year %>% filter(YEAR == YEAR_),
    aes(long, lat, group = group)
  ) +
    geom_polygon(aes(fill = PerCapita),
                 colour = alpha("white", 1/2), linewidth = 0.2) +
    geom_polygon(data = StateOutline, aes(x = long, y = lat, group = group),
                 color = "white", fill = NA) +
    scale_fill_viridis(
      labels = scales::comma,
      name = "Total Crime\nPer Capita",
      option = "H",
      begin = 0,
      end = 0.2
    ) +
    ggtitle(YEAR_) +
    theme_void()
}

# Display maps in a grid
do.call(grid.arrange, c(years_plot, ncol = 4))
```



Crime Pattern Analysis

Heatmap of Crime Types Over Time

This visualization shows the standardized patterns of different crime types across years.

```
# Preparing data for heatmap
crime_long <- Crime_data %>%
  select(YEAR, MURDER, RAPE, ROBBERY, AGASSLT, BURGLRY, LARCENY, MVTHEFT, ARSON) %>%
  # Calculating yearly totals for each crime type
  group_by(YEAR) %>%
  summarise(across(everything(), sum, na.rm = TRUE)) %>%
  # Reshaping data from wide to long format
  pivot_longer(cols = -YEAR,
               names_to = "Crime_Type",
               values_to = "Count") %>%
  # Calculate z-scores for better comparison across crime types
  group_by(Crime_Type) %>%
  mutate(z_score = scale(Count)) %>%
  ungroup()
# Adding  interactive controls in sidebar or above plot
selectInput("color_scheme", "Select Color Scheme:",
            choices = c("Plasma" = "plasma",
                        "Magma" = "magma",
                        "Viridis" = "viridis",
                        "Inferno" = "inferno"),
            selected = "plasma")
```

Select Color Scheme:

Plasma

```
sliderInput("year_range", "Select Year Range:",
            min = min(crime_long$YEAR),
            max = max(crime_long$YEAR),
            value = c(2002, 2014),
            step = 1,
            sep = "")
```

Select Year Range:

```
checkboxGroupInput("crime_types", "Select Crime Types:",
                   choices = unique(crime_long$Crime_Type),
                   selected = unique(crime_long$Crime_Type))
```

Select Crime Types:

- [x] MURDER
- [x] RAPE
- [x] ROBBERY
- [x] AGASSLT
- [x] BURGLRY
- [x] LARCENY
- [x] MVTHEFT
- [x] ARSON

```
# Creating interactive heatmap
renderPlotly({
  # data will be filtered based on inputs
  filtered_data <- crime_long %>%
    filter(
      YEAR >= input$year_range[1],
      YEAR <= input$year_range[2],
      Crime_Type %in% input$crime_types
    )

  # Creating the plot
  p <- ggplot(filtered_data,
              aes(x = YEAR, y = Crime_Type, fill = z_score,
                  text = paste("Year:", YEAR,
                               "\nCrime Type:", Crime_Type,
                               "\nZ-Score:", round(z_score, 2)))) +
    geom_tile() +
    scale_fill_viridis(
      option = input$color_scheme,
      name = "Z-Score\n(Standardized Count)",
      guide = guide_colorbar(title.position = "top")
    ) +
    scale_x_continuous(breaks = unique(filtered_data$YEAR)) +
    labs(
      title = paste("Crime Patterns:", input$year_range[1], "-", input$year_range[2]),
      subtitle = "Standardized counts shown as z-scores for better comparison",
      x = "Year",
      y = "Crime Type"
    ) +
    theme_minimal() +
    theme(
      axis.text.x = element_text(angle = 45, hjust = 1),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      plot.title = element_text(size = 14, face = "bold"),
      plot.subtitle = element_text(size = 10, color = "gray50"),
      legend.position = "right",
      legend.title = element_text(angle = 0)
    )

  # Convert to plotly for interactivity
  ggplotly(p, tooltip = "text") %>%
    layout(hoverlabel = list(bgcolor = "white"))
})
```

Interactive Data Table

```
selectInput("state", "Select State:",
            choices = unique(crosswalk$state_name),
            selected = NULL,
            multiple = FALSE)
```

Select State:

ALABAMA

```
uiOutput("countyUI")
```

```
pickerInput("crime_type", "Select Crime Type(s):",
            choices = crime_choices,
            selected = "MURDER",
            options = list(`actions-box` = TRUE),
            multiple = TRUE)
```

Select Crime Type(s):

MURDER

```
sliderInput("year", "Select Year Range:",
            min = 2002, max = 2014,
            value = c(2002, 2014),
            step = 1,
            sep = "")
```

Select Year Range:

Summary

```
renderPrint({
  req(filtered_data())
  selected_data <- filtered_data()

  state_counties <- paste(unique(selected_data$county_name), collapse = ", ")

  cat("Data Summary:\n",
      "State: ", input$state, "\n",
      "County(s): ", state_counties, "\n",
      "Crime Type(s): ", paste(input$crime_type, collapse = ", "), "\n",
      "Year Range: ", input$year[1], " to ", input$year[2], "\n")
})
```

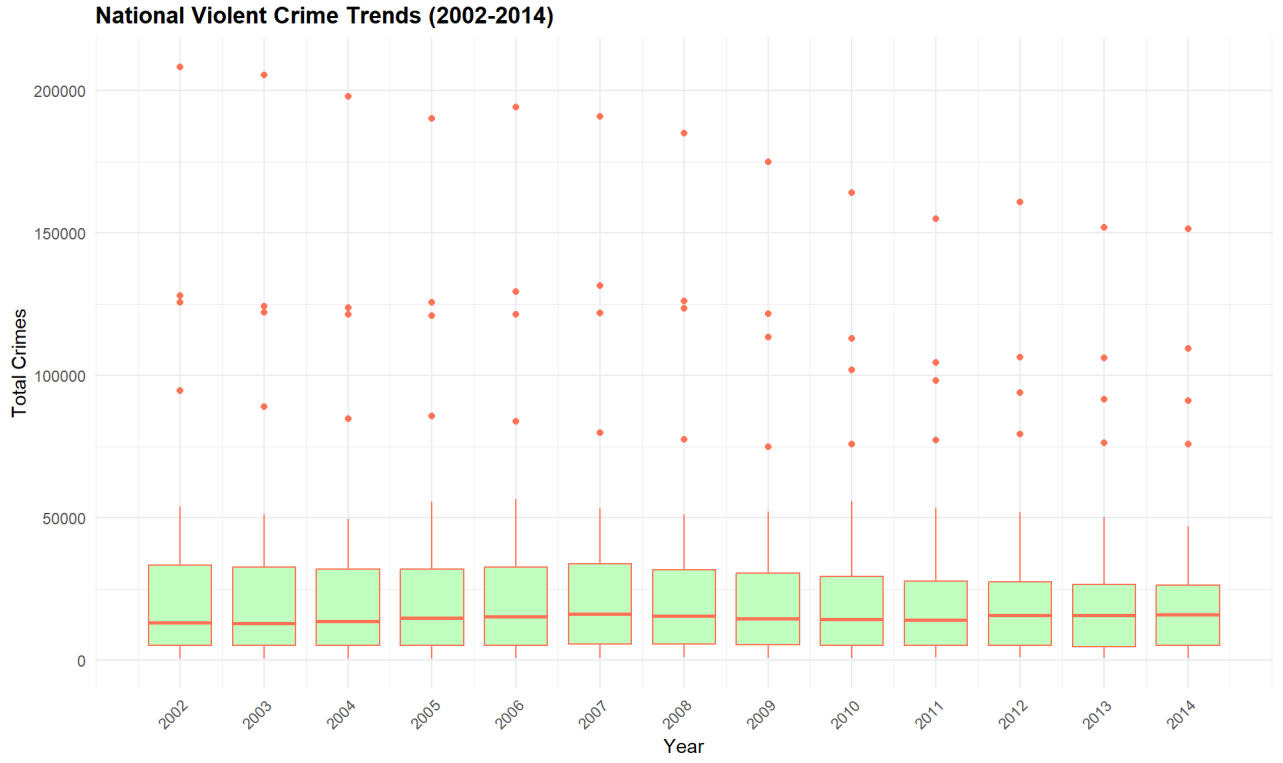
Crime Data Table

```
# County selection UI
observeEvent(input$state, {
  req(input$state)
  print(paste("State selected:", input$state))

  counties <- unique(crosswalk$county_name[crosswalk$state_name == input$state])
  print(paste("Counties found:", length(counties)))

  output$countyUI <- renderUI({
    pickerInput("county", "Select County:",
                choices = counties,
                options = list(`actions-box` = TRUE),
                multiple = TRUE)
  })
})

# Data filtering
filtered_data <- reactive({
  req(input$state)
  req(input$county)
  req(input$crime_type)

  print("All inputs received")

  selected_counties <- unique(crosswalk$county_fipcode[crosswalk$county_name %in% input$county])
  selected_counties <- sprintf("%05d", as.numeric(selected_counties))

  filtered <- crime_data_clean %>%
    filter(state_name == input$state,
           YEAR >= input$year[1],
           YEAR <= input$year[2]) %>%
    filter(Combined_FIPS %in% selected_counties) %>%
    group_by(YEAR, state_name, county_name) %>%
    summarise(across(all_of(input$crime_type), sum, na.rm = TRUE),
              .groups = 'drop')

  print(paste("Filtered rows:", nrow(filtered)))

  if (nrow(filtered) == 0) {
    return(NULL)
  }

  return(filtered)
})

# Data table output
renderDT({
  req(filtered_data())
  print("Rendering table")

  display_data <- filtered_data() %>%
    rename(State = state_name, County = county_name)

  selected_crime_columns <- intersect(input$crime_type, names(display_data))

  display_data <- display_data %>%
    select(YEAR, State, County, all_of(selected_crime_columns))

  print(paste("Final table rows:", nrow(display_data)))

  if (nrow(display_data) == 0) {
    return(NULL)
  }

  datatable(display_data,
            options = list(pageLength = 10,
                           scrollX = TRUE,
                           scrollY = "400px",
                           dom = 'Bfrtip',
                           buttons = c('copy', 'csv', 'excel', 'pdf', 'print')),
            extensions = 'Buttons')
})
```

Box Plot

Box plot of national violent crimes (MURDER, RAPE, ROBBERY, AGASSLT).

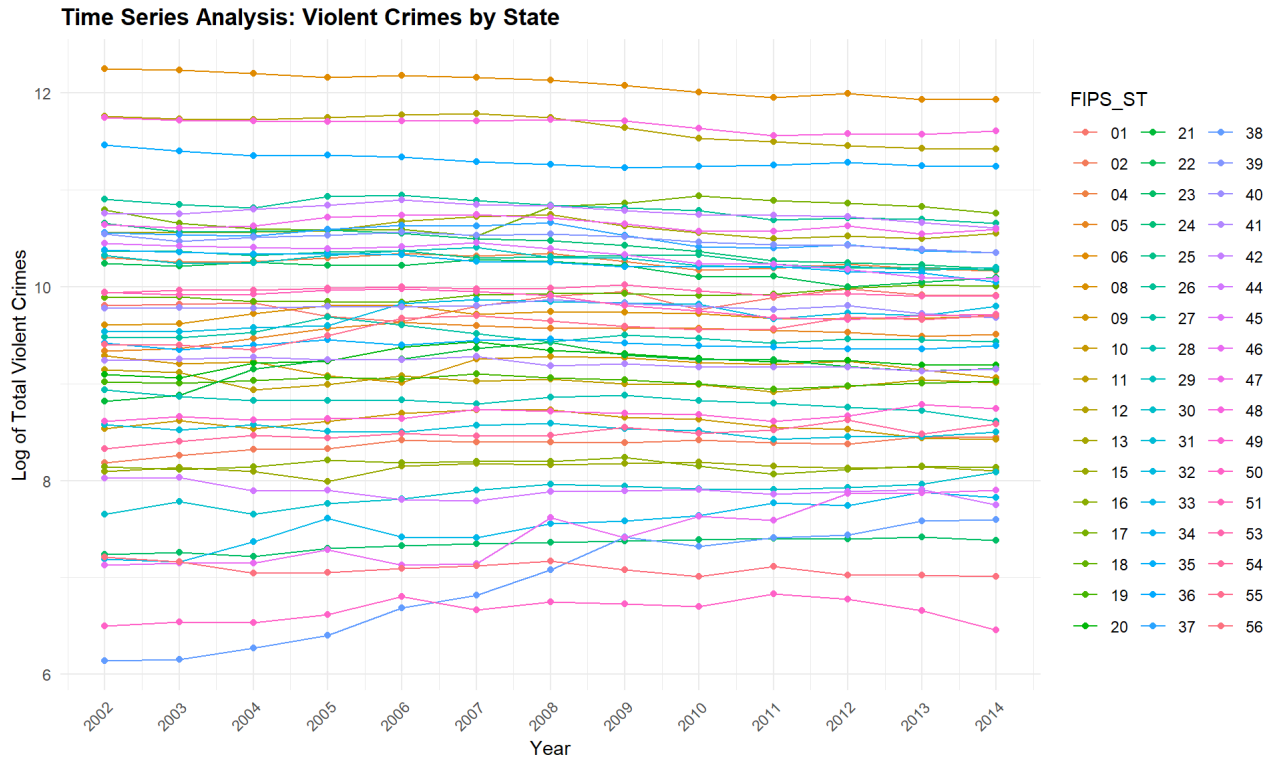
```
Data2 <- Crime_data %>%
  mutate(TotalViol = MURDER + RAPE + ROBBERY + AGASSLT) %>%
  group_by(YEAR, FIPS_ST) %>%
  summarize(TotalViol = sum(TotalViol))

ggplot(data = Data2, aes(group = YEAR, x = YEAR, y = TotalViol)) +
  geom_boxplot(color = "coral1", fill = "darkseagreen1") +
  labs(x = "Year", y = "Total Crimes",
       title = "National Violent Crime Trends (2002-2014)") +
  scale_x_continuous(breaks = seq(2002, 2014, 1)) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold")
  )
```



Line Graph

#Line graph with year by year trend examination (Violent Crime by State)

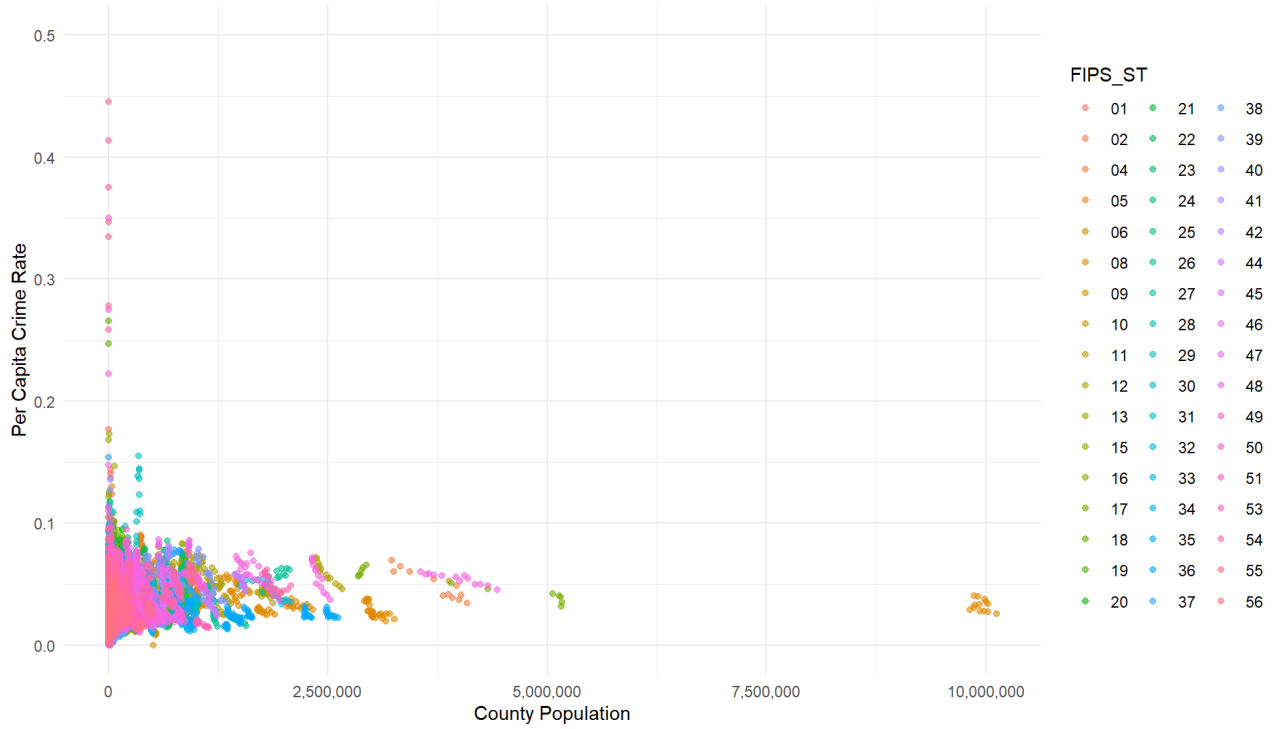
```
Data3 <- Crime_data %>%
  mutate(TotalViol = MURDER + RAPE + ROBBERY + AGASSLT) %>%
  group_by(YEAR, FIPS_ST) %>%
  summarize(TotalViol = sum(TotalViol))

ggplot(data = Data3, aes(group = FIPS_ST, color = FIPS_ST, x = YEAR, y = log(TotalViol))) +
  geom_line() +
  geom_point() +
  labs(x = "Year", y = "Log of Total Violent Crimes",
       title = "Time Series Analysis: Violent Crimes by State") +
  scale_x_continuous(breaks = seq(2002, 2014, 1)) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold"),
    legend.position = "right"
  )
```



Population and Crime Analysis

County Population vs Crime Rates

```
Data4 <- Crime_data %>%
  mutate(TotalCrime = MURDER + RAPE + ROBBERY + AGASSLT + BURGLRY + LARCENY +
           MVTHEFT + ARSON) %>%
  mutate(PerCapita = TotalCrime / CPOPCRIM) %>%
  group_by(STCOFIPS) %>%
  filter_all(all_vars(is.finite(PerCapita)))

ggplot(data = Data4, aes(x = CPOPCRIM, y = PerCapita, color = FIPS_ST)) +
  geom_point(alpha = 0.6) +
  labs(x = "County Population", y = "Per Capita Crime Rate",
       title = "Crime Rates per County") +
  scale_y_continuous(limits = c(0, 0.5)) +
  scale_x_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold"),
    legend.position = "right"
  )
```



Crime Rates Analysis

Bubble Charts

Analysis of violent and property crime rates in relation to population.

```
# Calculate crime rates

# First, let's calculate crime rates per 100,000 population and summary data will be created
crime_summary <- Crime_data %>%
  group_by(YEAR) %>%
  summarize(
    total_population = sum(CPOPCRIM, na.rm = TRUE),
    # Violent crimes (sum of murder, rape, robbery, aggravated assault)
    violent_crimes = sum(MURDER + RAPE + ROBBERY + AGASSLT, na.rm = TRUE),
    # Property crimes (sum of burglary, larceny, motor vehicle theft)
    property_crimes = sum(BURGLRY + LARCENY + MVTHEFT, na.rm = TRUE)
  ) %>%
  mutate(
    violent_rate = (violent_crimes / total_population) * 100000,
    property_rate = (property_crimes / total_population) * 100000
  )
# Adding interactive controls
selectInput("crime_view", "Select Crime Type:",
            choices = c("Violent Crimes" = "violent",
                        "Property Crimes" = "property"),
            selected = "violent")
```

Select Crime Type:

Violent Crimes

```
# Adding the color selection with predefined colors
selectInput("bubble_color", "Select Color:",
            choices = c("Magenta" = "#FF69B4",
                        "Blue" = "#0000FF",
                        "Green" = "#00FF00",
                        "Purple" = "#800080",
                        "Red" = "#FF0000"),
            selected = "#FF69B4")
```

Select Color:

Magenta

```
# Creating the interactive bubble chart
renderPlotly({
  # rate will be chosen based on selection
  rate_col <- if(input$crime_view == "violent") {
    crime_summary$violent_rate
  } else {
    crime_summary$property_rate
  }

  # Creating  plot
  p <- ggplot(crime_summary,
              aes(x = YEAR,
                  y = rate_col,
                  size = total_population/1000000,
                  text = paste("Year:", YEAR,
                               "\nRate:", round(rate_col, 2),
                               "\nPopulation:", format(total_population, big.mark=",")))) +
    geom_point(alpha = 0.6, color = input$bubble_color) +
    scale_size_continuous(name = "Population\n(Millions)",
                          range = c(10, 30)) +
    scale_x_continuous(breaks = unique(crime_summary$YEAR)) +
    labs(title = paste(input$crime_view, "Rate vs. Year (2002-2014)"),
         subtitle = "Bubble size represents total population",
         x = "Year",
         y = paste(input$crime_view, "per 100,000 Population")) +
    theme_minimal() +
    theme(
      plot.title = element_text(size = 14, face = "bold"),
      plot.subtitle = element_text(size = 10, color = "gray50"),
      axis.text.x = element_text(angle = 45, hjust = 1)
    )

  # Convert to plotly for interactivity
  ggplotly(p, tooltip = "text") %>%
    layout(hoverlabel = list(bgcolor = "white"))
})
```

```
# Calculating and displaying percentage change
renderText({
  rate_col <- if(input$crime_view == "violent") {
    "violent_rate"
  } else {
    "property_rate"
  }

  start_rate <- crime_summary %>%
    filter(YEAR == 2002) %>%
    pull(!!sym(rate_col))

  end_rate <- crime_summary %>%
    filter(YEAR == 2014) %>%
    pull(!!sym(rate_col))

  percent_change <- ((end_rate - start_rate) / start_rate) * 100

  paste("Percentage Change (2002-2014):", round(percent_change, 2), "%")
})
```

```
# Creating the bubble chart for property crimes
# Add interactive controls
selectInput("property_color", "Select Bubble Color:",
            choices = c("Green" = "#00FF00",
                        "Blue" = "#0000FF",
                        "Purple" = "#800080",
                        "Orange" = "#FFA500",
                        "Teal" = "#008080"),
            selected = "#00FF00")
```

Select Bubble Color:

Green

```
# Add year range selector
sliderInput("year_range", "Select Year Range:",
            min = 2002, max = 2014,
            value = c(2002, 2014),
            step = 1)
```

Select Year Range:

```
renderPlotly({
  # Filter data based on year range
  filtered_data <- crime_summary %>%
    filter(YEAR >= input$year_range[1] & YEAR <= input$year_range[2])

  # Creating the plot
  p <- ggplot(filtered_data,
              aes(x = YEAR,
                  y = property_rate,
                  size = total_population/1000000,
                  text = paste("Year:", YEAR,
                               "\nProperty Crime Rate:", round(property_rate, 1),
                               "\nPopulation:", format(total_population, big.mark=","),
                               "million"))) +
    geom_point(alpha = 0.6, color = input$property_color) +
    scale_size_continuous(name = "Population\n(Millions)",
                          range = c(10, 30)) +
    scale_x_continuous(breaks = unique(filtered_data$YEAR)) +
    labs(title = "Property Crime Rate vs. Year (2002-2014)",
         subtitle = "Bubble size represents total population",
         x = "Year",
         y = "Property Crimes per 100,000 Population") +
    theme_minimal() +
    theme(
      plot.title = element_text(size = 14, face = "bold"),
      plot.subtitle = element_text(size = 10, color = "gray50"),
      axis.text.x = element_text(angle = 45, hjust = 1)
    )

  # Convert to plotly
  ggplotly(p, tooltip = "text")
})
```

```
# Calculating and displaying percentage change dynamically
renderText({
  filtered_data <- crime_summary %>%
    filter(YEAR >= input$year_range[1] & YEAR <= input$year_range[2])

  start_rate <- filtered_data %>%
    filter(YEAR == min(YEAR)) %>%
    pull(property_rate)

  end_rate <- filtered_data %>%
    filter(YEAR == max(YEAR)) %>%
    pull(property_rate)

  percent_change <- ((end_rate - start_rate) / start_rate) * 100

  paste("Percentage Change between",
        min(filtered_data$YEAR), "and",
        max(filtered_data$YEAR), ":",
        round(percent_change, 2), "%")
})
```

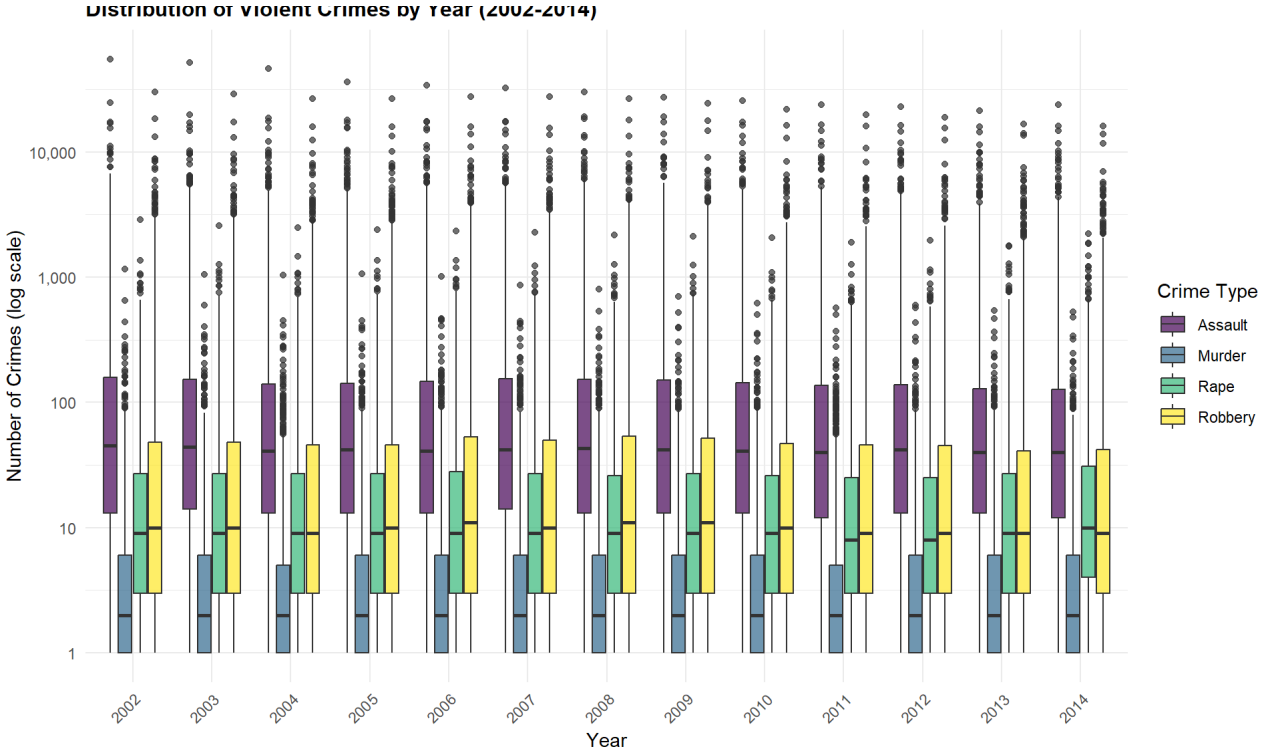
Distribution of Crimes

Side by Side Box plots

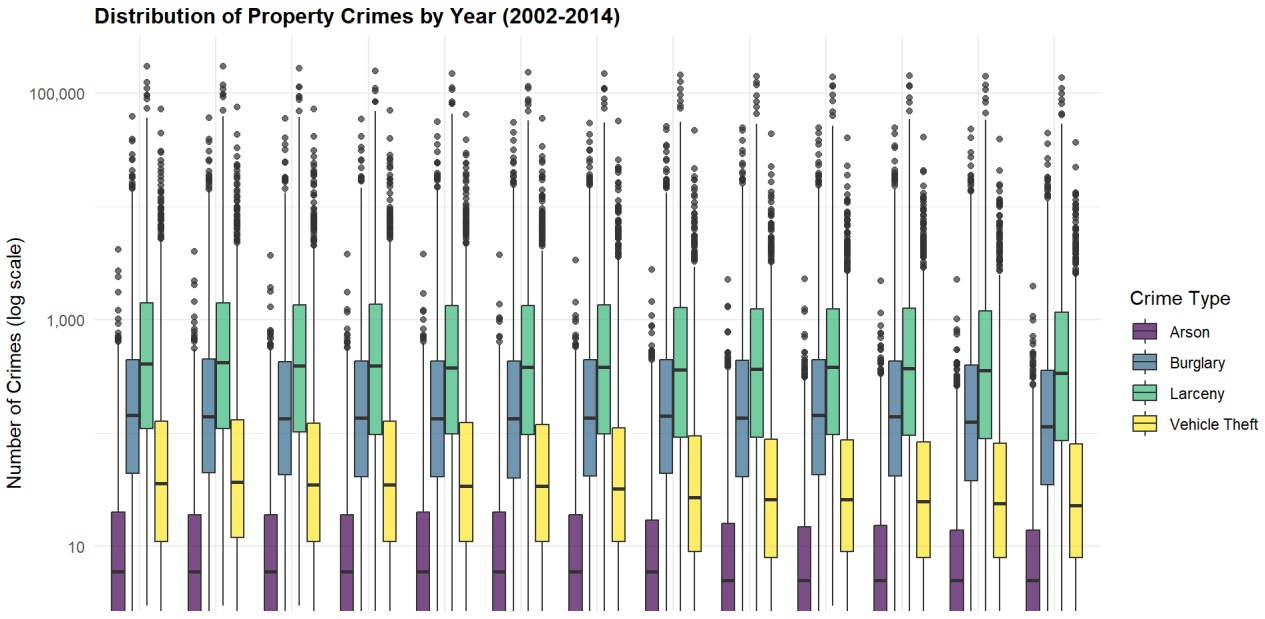
Analysis of the distribution of different crime types using side by side box plots to compare distributions across years

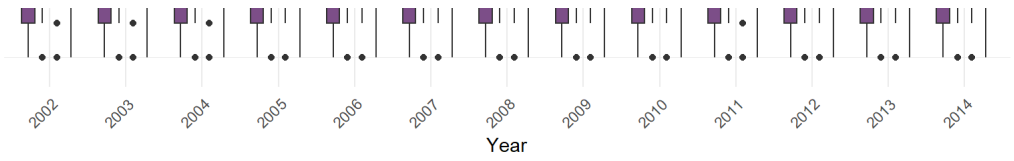
```
crime_box_data<-Crime_data %>%
  # This is to select relevant columns and convert to long format
  select(YEAR, MURDER, RAPE, ROBBERY, AGASSLT, BURGLRY, LARCENY, MVTHEFT, ARSON) %>%
  pivot_longer(
    cols = -YEAR,
    names_to = "Crime_Type",
    values_to = "Count"
  ) %>%
  # Removing any NA or infinite values
  filter(!is.infinite(Count), !is.na(Count)) %>%
  # Creating more readable crime type labels
  mutate(Crime_Type = case_when(
    Crime_Type == "MURDER" ~ "Murder",
    Crime_Type == "RAPE" ~ "Rape",
    Crime_Type == "ROBBERY" ~ "Robbery",
    Crime_Type == "AGASSLT" ~ "Assault",
    Crime_Type == "BURGLRY" ~ "Burglary",
    Crime_Type == "LARCENY" ~ "Larceny",
    Crime_Type == "MVTHEFT" ~ "Vehicle Theft",
    Crime_Type == "ARSON" ~ "Arson"
  ))

# Creating two box plots: one for violent crimes and one for property crimes
# Violent Crimes Box Plot
crime_box_data %>%
  filter(Crime_Type %in% c("Murder", "Rape", "Robbery", "Assault")) %>%
  ggplot(aes(x = as.factor(YEAR), y = Count, fill = Crime_Type)) +
  geom_boxplot(alpha = 0.7) +
  scale_fill_viridis_d() +
  scale_y_log10(labels = scales::comma_format()) +
  labs(
    title = "Distribution of Violent Crimes by Year (2002-2014)",
    x = "Year",
    y = "Number of Crimes (log scale)",
    fill = "Crime Type"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 12, face = "bold"),
    legend.position = "right"
  )
```

Distribution of Violent Crimes by Year (2002-2014)




```
# Property Crimes Box Plot
 crime_box_data %>%
  filter(Crime_Type %in% c("Burglary", "Larceny", "Vehicle Theft", "Arson")) %>%
  ggplot(aes(x = as.factor(YEAR), y = Count, fill = Crime_Type)) +
  geom_boxplot(alpha = 0.7) +
  scale_fill_viridis_d() +
  scale_y_log10(labels = scales::comma_format()) +
  labs(
    title = "Distribution of Property Crimes by Year (2002-2014)",
    x = "Year",
    y = "Number of Crimes (log scale)",
    fill = "Crime Type"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 12, face = "bold"),
    legend.position = "right"
  )
```






```
# Calculating summary statistics
summary_stats <- crime_box_data %>%
  group_by(Crime_Type, YEAR) %>%
  summarise(
    Median = median(Count, na.rm = TRUE),
    Q1 = quantile(Count, 0.25, na.rm = TRUE),
    Q3 = quantile(Count, 0.75, na.rm = TRUE),
    Mean = mean(Count, na.rm = TRUE),
    SD = sd(Count, na.rm = TRUE),
  .groups = "drop"  # This removes the grouping after summarise
  ) %>%
  ungroup()
print(head(summary_stats))
```

```
## # A tibble: 6 × 7
##   Crime_Type  YEAR Median    Q1    Q3  Mean    SD
##   <chr>      <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Arson       2002      2     0    11  24.1 125.
## 2 Arson       2003      2     0    10  22.8 115.
## 3 Arson       2004      2     0    10  21.7 105.
## 4 Arson       2005      2     0    11  21.5 102.
## 5 Arson       2006      2     0    11  22.2 103.
## 6 Arson       2007      2     0    11  20.8  95.7
```

National Crime Trends

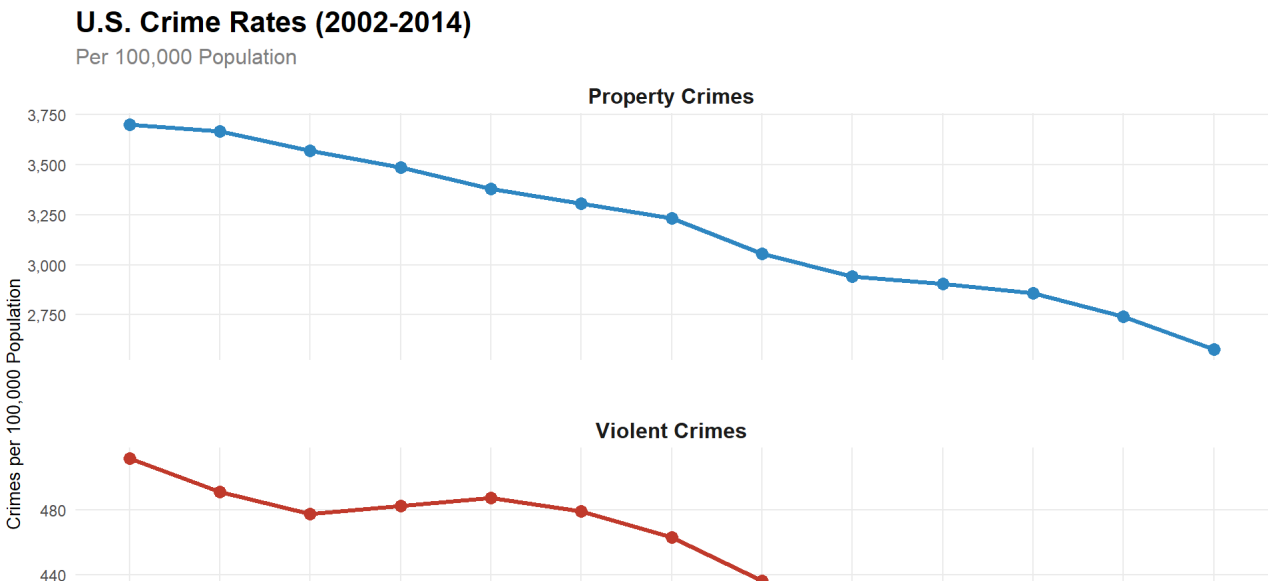
Time series plot

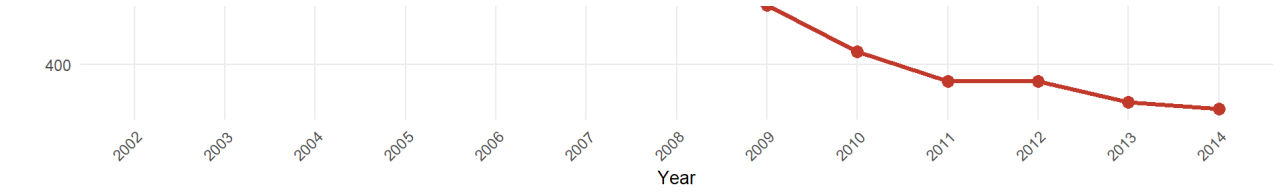
This is the Analysis of overall national crime trends over the study period.

```
# Calculating national trends
# First, let's calculate national totals for each year
national_trends <- Crime_data %>%
  group_by(YEAR) %>%
  summarize(
    # Calculating violent crimes total
    violent_crimes = sum(MURDER + RAPE + ROBBERY + AGASSLT, na.rm = TRUE),
    # Calculating property crimes total
    property_crimes = sum(BURGLRY + LARCENY + MVTHEFT, na.rm = TRUE),
    # Calculating total population for rates
    total_population = sum(CPOPCRIM, na.rm = TRUE)
  ) %>%
  # Calculating rates per 100,000 population
  mutate(
    violent_rate = (violent_crimes / total_population) * 100000,
    property_rate = (property_crimes / total_population) * 100000
  ) %>%
  # Converting to long format for plotting
  pivot_longer(
    cols = c(violent_rate, property_rate),
    names_to = "crime_type",
    values_to = "rate"
  ) %>%
  # This is to make crime type labels more readable
  mutate(
    crime_type = case_when(
      crime_type == "violent_rate" ~ "Violent Crimes",
      crime_type == "property_rate" ~ "Property Crimes"
    )
  )
national_trends
```

```
## # A tibble: 26 × 6
##     YEAR violent_crimes property_crimes total_population crime_type       rate
##    <dbl>          <dbl>           <dbl>            <dbl> <chr>           <dbl>
##  1  2002        1369845         9916958        267856616 Violent Crimes   511.
##  2  2002        1369845         9916958        267856616 Property Crimes 3702.
##  3  2003        1330239         9930556        270863727 Violent Crimes   491.
##  4  2003        1330239         9930556        270863727 Property Crimes 3666.
##  5  2004        1322312         9887114        277036402 Violent Crimes   477.
##  6  2004        1322312         9887114        277036402 Property Crimes 3569.
##  7  2005        1344216         9711433        278616385 Violent Crimes   482.
##  8  2005        1344216         9711433        278616385 Property Crimes 3486.
##  9  2006        1373094         9527971        281849827 Violent Crimes   487.
## 10  2006        1373094         9527971        281849827 Property Crimes 3381.
## # ℹ 16 more rows
```

```r
# Creating the time series plot
ggplot(national_trends, aes(x = YEAR, y = rate, color = crime_type)) +
  # Add lines
  geom_line(size = 1.2) +
  # Add points
  geom_point(size = 3) +
  # Use different scales for violent and property crimes
  facet_wrap(~crime_type, scales = "free_y", nrow = 2) +
  # Customize colors
  scale_color_manual(values = c("Property Crimes" = "#2E86C1",
                                "Violent Crimes" = "#C0392B")) +
  # Format axis labels
  scale_x_continuous(breaks = 2002:2014) +
  scale_y_continuous(labels = scales::comma) +
  # Add labels
  labs(
    title = "U.S. Crime Rates (2002-2014)",
    subtitle = "Per 100,000 Population",
    x = "Year",
    y = "Crimes per 100,000 Population",
    color = "Crime Type"
  ) +
  # Customize theme
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12, color = "gray50"),
    axis.title = element_text(size = 10),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "none",
    panel.grid.minor = element_blank(),
    strip.text = element_text(size = 12, face = "bold"),
    panel.spacing = unit(2, "lines")
  )
```





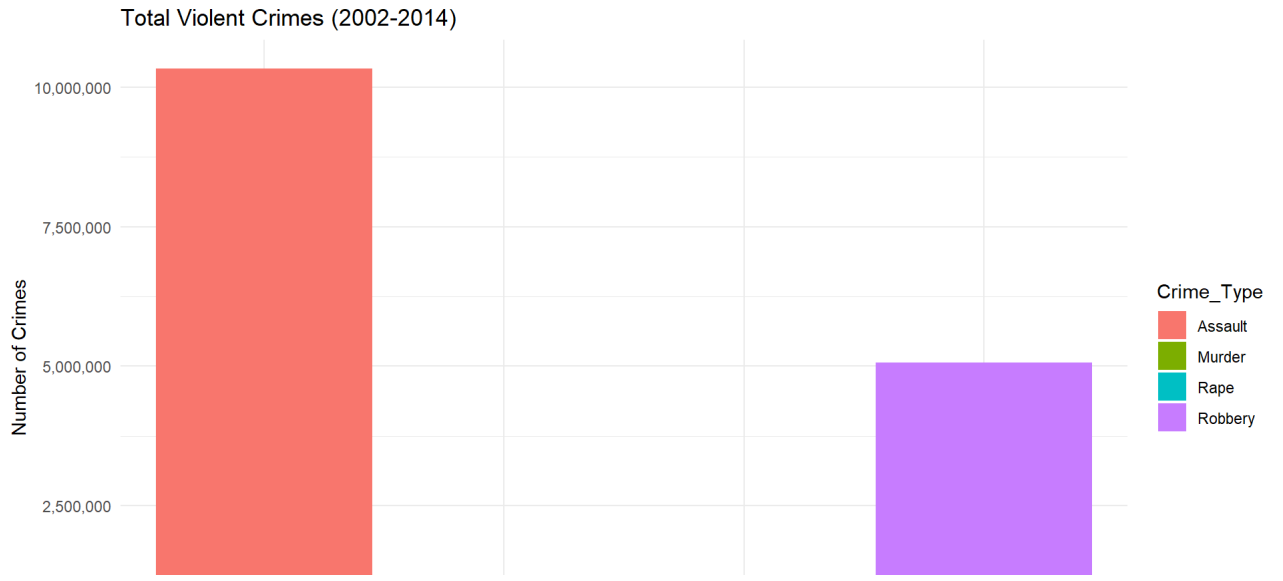

```
# Print summary statistics
summary_stats <- national_trends %>%
  group_by(crime_type) %>%
  summarise(
    Mean_Rate = mean(rate),
    Percent_Change = ((last(rate) - first(rate)) / first(rate) * 100)
  )
print(summary_stats)
```

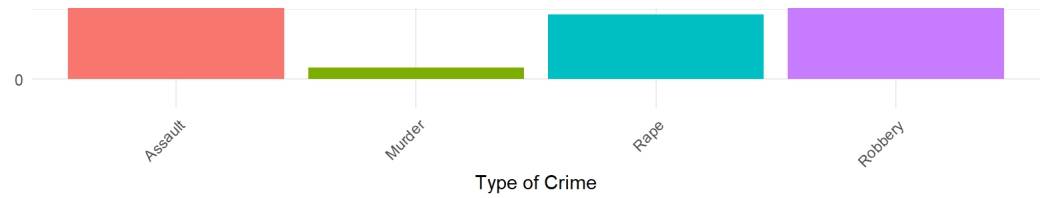
```
## # A tibble: 2 × 3
##   crime_type      Mean_Rate Percent_Change
##   <chr>               <dbl>          <dbl>
## 1 Property Crimes     3186.          -30.4
## 2 Violent Crimes       443.          -27.1
```

Bar Plot

##Simple Bar Plot of Major Crime Types

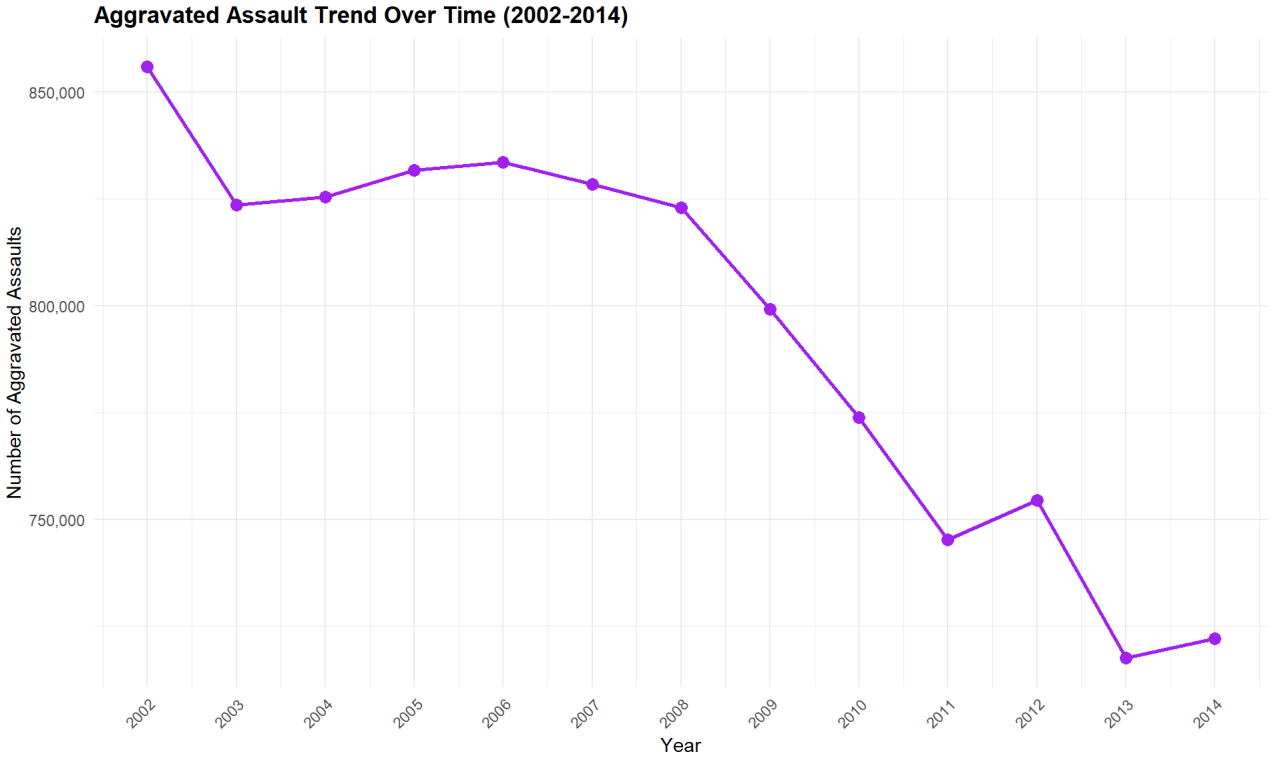
```
# Create simple bar plot of crime totals
Crime_data %>%
  summarise(
    Murder = sum(MURDER, na.rm = TRUE),
    Rape = sum(RAPE, na.rm = TRUE),
    Robbery = sum(ROBBERY, na.rm = TRUE),
    Assault = sum(AGASSLT, na.rm = TRUE)
  ) %>%
  pivot_longer(everything(),
               names_to = "Crime_Type",
               values_to = "Total") %>%
  ggplot(aes(x = Crime_Type, y = Total, fill = Crime_Type)) +
  geom_col() +
  scale_y_continuous(labels = comma) +
  theme_minimal() +
  labs(
    title = "Total Violent Crimes (2002-2014)",
    x = "Type of Crime",
    y = "Number of Crimes"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```





Trend for Aggravated Assault

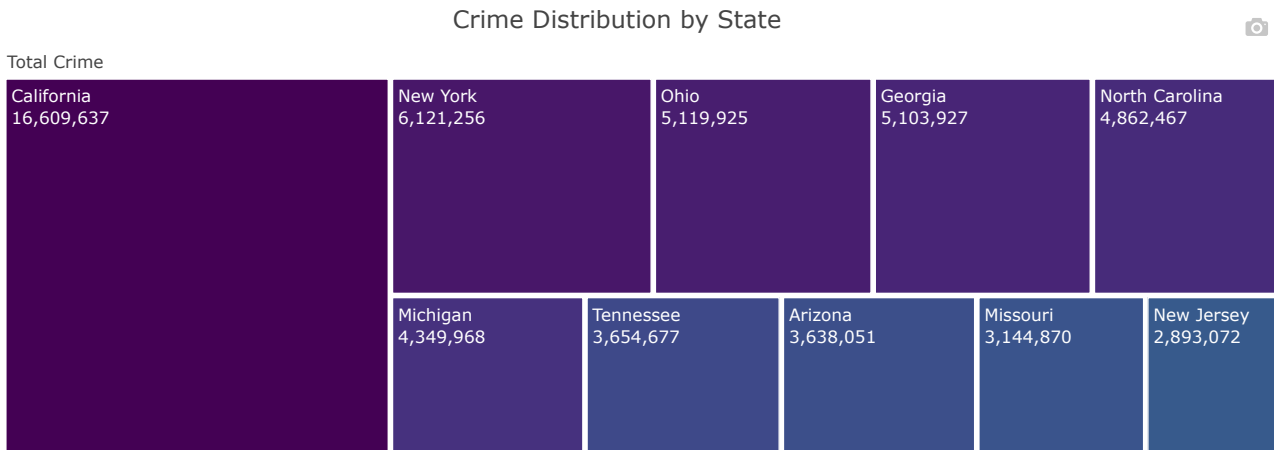
```
#Showing trend for Aggravated Assault
Crime_data %>%
  group_by(YEAR) %>%
  summarise(
    Total_Assaults = sum(AGASSLT, na.rm = TRUE)
  ) %>%
  ggplot(aes(x = YEAR, y = Total_Assaults)) +
  geom_line(color = "purple", size = 1) +
  geom_point(color = "purple", size = 3) +
  theme_minimal() +
  scale_y_continuous(labels = comma) +
  # Fix x-axis to show all years
  scale_x_continuous(breaks = 2002:2014) +  # This will show every year
  labs(
    title = "Aggravated Assault Trend Over Time (2002-2014)",
    x = "Year",
    y = "Number of Aggravated Assaults"
  ) +
  theme(
    plot.title = element_text(face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1)  # Angle the year labels for better readability
  )
```

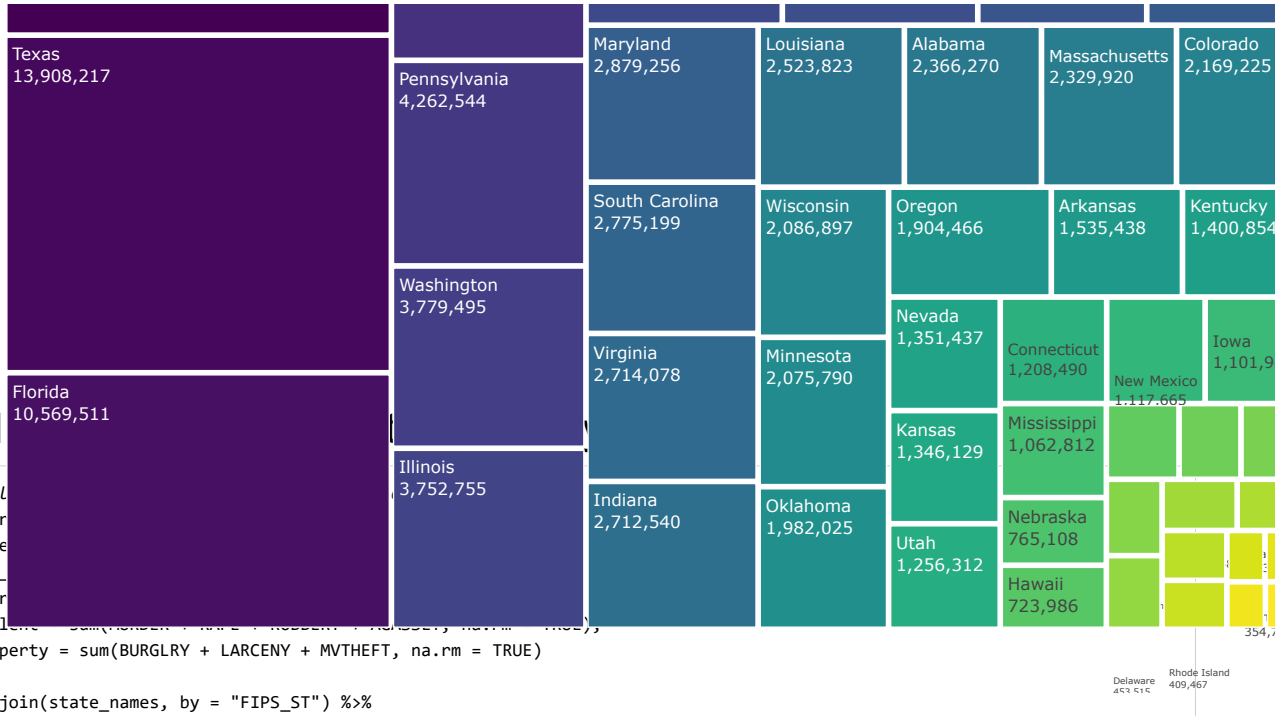


Treemaps

##Treemap for exploring total crimes by state

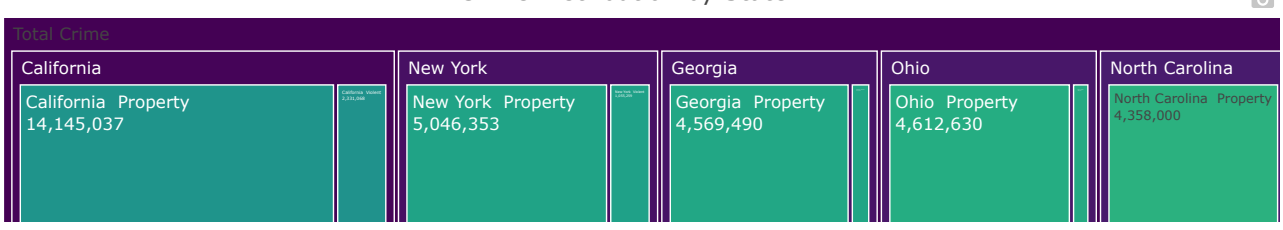
```r
library(plotly)
library(dplyr)

# Creating state names look up with all states from code book
state_names <- data.frame(
  FIPS_ST = c("01", "02", "04", "05", "06", "08", "09", "10", "11", "12",
              "13", "15", "16", "17", "18", "19", "20", "21", "22", "23",
              "24", "25", "26", "27", "28", "29", "30", "31", "32", "33",
              "34", "35", "36", "37", "38", "39", "40", "41", "42", "44",
              "45", "46", "47", "48", "49", "50", "51", "53", "54", "55", "56"),
  State = c("Alabama", "Alaska", "Arizona", "Arkansas", "California",
            "Colorado", "Connecticut", "Delaware", "DC", "Florida",
            "Georgia", "Hawaii", "Idaho", "Illinois", "Indiana",
            "Iowa", "Kansas", "Kentucky", "Louisiana", "Maine",
            "Maryland", "Massachusetts", "Michigan", "Minnesota", "Mississippi",
            "Missouri", "Montana", "Nebraska", "Nevada", "New Hampshire",
            "New Jersey", "New Mexico", "New York", "North Carolina", "North Dakota",
            "Ohio", "Oklahoma", "Oregon", "Pennsylvania", "Rhode Island",
            "South Carolina", "South Dakota", "Tennessee", "Texas", "Utah",
            "Vermont", "Virginia", "Washington", "West Virginia", "Wisconsin", "Wyoming")
)

# Calculating total crimes by state with complete state coverage
crime_tree <- Crime_data %>%
  mutate(FIPS_ST = as.character(FIPS_ST)) %>%
  group_by(FIPS_ST) %>%
  summarise(
    Total = sum(MURDER + RAPE + ROBBERY + AGASSLT +
                BURGLRY + LARCENY + MVTHEFT + ARSON, na.rm = TRUE)
  ) %>%
  # Joining with state names to ensure all states are included
  right_join(state_names, by = "FIPS_ST") %>%
  # Replace any NAs with 0
  mutate(Total = coalesce(Total, 0)) %>%
  arrange(desc(Total))

# Creating treemap
plot_ly(
  type = "treemap",
  ids = paste0("id_", seq_along(crime_tree$State)),  # Unique IDs for each state
  labels = crime_tree$State,
  parents = rep("Total Crime", nrow(crime_tree)),
  values = crime_tree$Total,
  textinfo = "label+value",
  hoverinfo = "label+value+percent parent",
  marker = list(
    colors = viridis::viridis(nrow(crime_tree))
  ),
  branchvalues = "total"
) %>%
  layout(
    title = "Crime Distribution by State",
    width = 1000,
    height = 800
  )
```

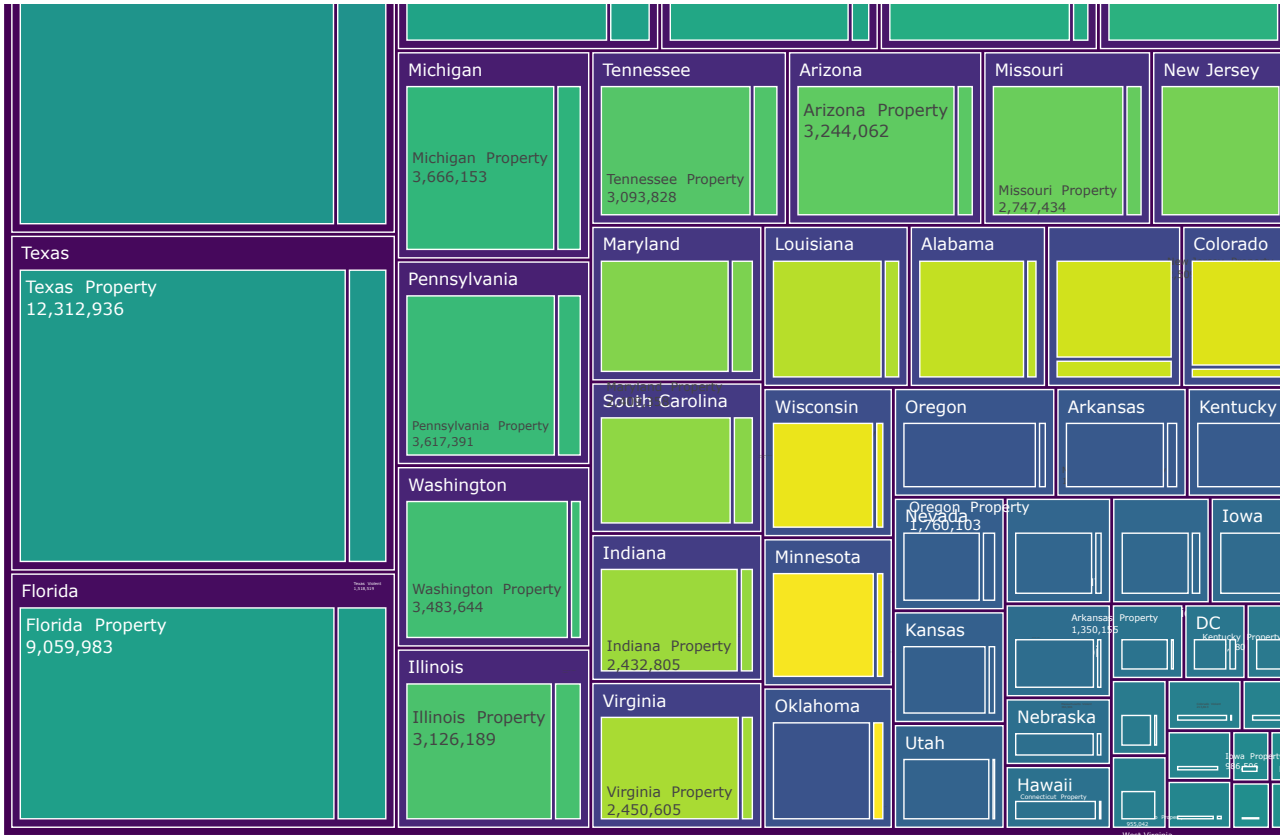




Total

```
# Calcul
crime_tr
  mutate
  group_
  summar
    Violent = sum(MURDER + RAPE + ROBBERY + AGASSLT, na.rm = TRUE),
    Property = sum(BURGLRY + LARCENY + MVTHEFT, na.rm = TRUE)
  ) %>%
  left_join(state_names, by = "FIPS_ST") %>%
  mutate(
    Total = Violent + Property
  ) %>%
  arrange(desc(Total))

# Creating the treemap data
plot_ly(
  type = "treemap",
  labels = c(
    "Total Crime",
    crime_tree$State,
    paste(rep(crime_tree$State, each = 2),
          c(" Violent", " Property"))
  ),
  parents = c(
    "",
    rep("Total Crime", nrow(crime_tree)),
    rep(crime_tree$State, each = 2)
  ),
  values = c(
    sum(crime_tree$Total),
    crime_tree$Total,
    c(rbind(crime_tree$Violent, crime_tree$Property))
  ),
  textinfo = "label+value",
  hoverinfo = "label+value+percent parent",
  marker = list(
    colors = viridis::viridis(nrow(crime_tree) * 2 + 1)
  ),
  branchvalues = "total"
) %>%
  layout(
    title = "Crime Distribution by State",
    width = 1000,
    height = 800
  )
```




Michigan
Michigan Property
3,666,153
Tennessee
Tennessee Property
3,093,828
Arizona
Arizona Property
3,244,062
Missouri
Missouri Property
2,747,434
New Jersey
Texas
Texas Property
12,312,936
Pennsylvania
Pennsylvania Property
3,617,391
Maryland
Louisiana
Alabama
Colorado
South Carolina
Wisconsin
Oregon
Arkansas
Kentucky
Washington
Washington Property
3,483,644
Oregon Property
1,760,103
Nevada
Iowa
Indiana
Indiana Property
2,432,805
Minnesota
Kansas
Arkansas Property
1,350,155
DC
Florida
Florida Property
9,059,983
Illinois
Illinois Property
3,126,189
Virginia
Virginia Property
2,450,605
Oklahoma
Nebraska
Utah
Hawaii