

FinalProject_STAT650

Muhammad Sualeh Alam (tz3502) Humza Shah

Introduction:

Basic Background about NICS System

NICS (National Instant Criminal Background Check System) is the system through which people undergo checks before acquiring a firearm. NICS is a [national database](#) maintained by the **FBI**. NICS checks are point-of-sale background checks that happen each time a firearm is purchased. NICS searches criminal and mental health history and other records to determine if the person is legally prohibited from owning a firearm.

Motivation behind choosing this dataset:

We have been interested in real-world government data to solve real problems and shape policy changes for the betterment of the society and public health related outcomes. The **NICS** is a national database maintained by the **FBI**, we very intrigued by this dataset since last year and wanted to work on it so we could be challenged with real-world problems and suggest genuine real-world solutions. Since we have seen increased gun shooting incidents (especially recent *high-profile shooting incident*) due to various socio-political reasons occurring over the recent months. We wanted to discover and investigate the key factors linked to increased weapon sales to the public and their usage in the overall

Data Description:

The dataset covers the time period **from January 2017 to December 2021**, providing monthly background check data for each U.S. state and firearm category.

Each column represents the type of transaction submitted to the *National Instant Criminal Background Check System (NICS)*. In addition, each type of transaction is broken down by the type of firearm—handgun, long gun, and other. The types of firearms are defined by the Bureau of Alcohol, Tobacco, Firearms and Explosives as follows:

Primary Variables of Interest:

1. Permit

A background check before issuing a permit to purchase or carry a firearm. Done by law enforcement or other agencies. May not indicate a firearm was purchased.

2. Permit Recheck

A repeated background check on someone who already has a permit. These are ongoing rechecks required in some states.

3. Handgun

A background check related specifically to a **handgun purchase**.

4. Long Gun

A background check for a **long gun** (rifle or shotgun).

5. Other

Firearms that are not handguns or long guns. Includes silencers, receivers, pistol-grip shotguns, grenades launchers, etc.

6. Multiple

A background check where **more than one firearm type** (e.g., both handgun and rifle) is involved in a single transaction.

7. Admin

Administrative checks that don't relate to a sale or transfer. It could include system testing, data corrections, etc.

8. Private Sale Handgun

Private Sale refers to Checks related to person-to-person (private) sales, often facilitated by a licensed dealer to meet legal requirements. Background check for a private sale of a handgun.

9. Private Sale Long Gun

Background check for a long gun private sale.

10. Private Sale Other

Background check for Other-type private sale.

Questions for Analysis:

1. ***What is the overall trend of different firearm type permits (Declining or Rising in US) across 2017-2020 ?***
2. ***Are handguns replacing long guns in popularity over time (2017-2021) ?***
3. ***In which year did the United States experience the peak in total firearm-related permit background checks and what might this indicate about public sentiments?***
4. ***What are the top 10 states with the most and least gun permits across US (2017-2021) ?***
5. ***Which state had the highest total background checks for handguns, long guns, other guns and multiple firearm types from 2017 to 2020?***

6. ***Which states had the highest number of firearm permit background checks from 2017 to 2020?***
7. ***Is the “Other” firearms category growing, and what might this suggest about trends in niche/tactical firearms?***
8. ***Do permit rechecks (renewals or ongoing monitoring) vary significantly across states, and what does this indicate about state-level firearm monitoring systems?***
9. ***Do states with high handgun background checks also have high firearm-related incidents (crime/suicide)?***

Data Preparation Steps:

1. Converting **PDF** format data into **Excel** format using online PDF to Excel converter
2. Combining and cleaning all different worksheets into a single dataframe for analysis
3. Creating column of Year and Month_Year for each worksheet for ease

```
# Takes 1 minute to run this cell (to extract data till January 2017)

# Extracted data from Jan 2017 - Dec 2020 (2017-2021)

# Get all worksheet names from Excel
sheet_names <-
excel_sheets("C:/Users/ihate/Downloads/f24/data/NICS_Firearms.xlsx")

sheet_names_to_read <- sheet_names[-1] #remove first sheet as its blank sheet

# To extract numbers from sheets name
table_numbers <- as.numeric(gsub("Table ", "", sheet_names_to_read))

# logic to select only sheets with increment of 3
keep_indices <- (table_numbers %% 3 == 1)
filtered_sheet_names <- sheet_names_to_read[keep_indices]

# print(head(filtered_sheet_names))

# Reducing length of sheets for analyzing only 5 years (40 works)
filtered_sheet_names= filtered_sheet_names[1:48] # WORKING till year 2017

data_list <- list()

for(sheet in filtered_sheet_names) {
  # Reading sheets one by one to a temp df
  df_temp <-
read_excel("C:/Users/ihate/Downloads/f24/data/NICS_Firearms.xlsx",
           sheet = sheet, skip=0,
           col_names=c("State", "Permit", "Permit_Recheck", "HandGun",
                        "LongGun", "Other","Multiple", "Admin",
                        "Prepawn_HG", "Prepawn_LG", "Prepawn_Other",
```

```

        "Redemption_HG", "Redemption_LG",
        "Redemption_Other", "Return/Disp_HG",
        "Return/Disp_LG",
        "Return/Disp_Other", "Rentals_HG",
        "Rentals_LG", "PrivateSale_HG",
"PrivateSale_LG",
        "PrivateSale_Other", "RetToSeller_HG",
        "RetToSeller_LG", "RetToSeller_Other",
"Totals"))

# Adding Year and Months to dataframe based on each Excel worksheet
df_temp['Month_Year'] = df_temp$'State'[2]
df_temp['Year'] = as.numeric(gsub("\\D", "", df_temp$'State'[2]))

# Removing first 5 rows (empty) and last row (TOTALS row) from the
dataframe
df_temp <- df_temp[-c(1:4, nrow(df_temp)), ]

# Removing first 4 rows (empty rows)
# df_temp = df_temp[-c(1:4), ]

# cat("SheetName ",sheet, "\n") # commenting out rn to display sheet
names

  data_list[[sheet]] <- df_temp
}

combined_df <- bind_rows(data_list) # combine multiple df into single df
View(combined_df)

# To view number of rows and columns of df
print(dim(combined_df))

[1] 2640    28

head(combined_df,2)

# A tibble: 2 × 28
  State   Permit Permit_Recheck HandGun LongGun Other Multiple Admin
Prepawn_HG
  <chr>   <chr>   <chr>           <chr>   <chr>   <chr> <chr>   <chr> <chr>
1 Alabama 33421   505             31103   28933   1855  1388     0     43
2 Alaska  441     3             3658   3822    430   233     0     0
# i 19 more variables: Prepawn_LG <chr>, Prepawn_Other <chr>,
#   Redemption_HG <chr>, Redemption_LG <chr>, Redemption_Other <chr>,
#   `Return/Disp_HG` <chr>, `Return/Disp_LG` <chr>, `Return/Disp_Other`
<chr>,
#   Rentals_HG <chr>, Rentals_LG <chr>, PrivateSale_HG <chr>,
#   PrivateSale_LG <chr>, PrivateSale_Other <chr>, RetToSeller_HG <chr>,

```

```
# RetToSeller_LG <chr>, RetToSeller_Other <chr>, Totals <chr>,  
# Month_Year <chr>, Year <dbl>
```

The NICS Firearms Background Checks dataset has:

Columns = **28**

Rows = **2640**

The extracted dataset contains NICS Background Check data from **2017 January till 2020 December**

Data Cleaning

```
sum(is.na(combined_df))
```

```
[1] 0
```

There are no missing values in this dataset.

```
sum(duplicated(combined_df))
```

```
[1] 0
```

There are no duplicate values as well

```
str(combined_df)
```

```
tibble [2,640 × 28] (S3: tbl_df/tbl/data.frame)  
  $ State           : chr [1:2640] "Alabama" "Alaska" "Arizona" "Arkansas"  
  ...  
  $ Permit          : chr [1:2640] "33421" "441" "9043" "3878" ...  
  $ Permit_Recheck  : chr [1:2640] "505" "3" "846" "900" ...  
  $ HandGun         : chr [1:2640] "31103" "3658" "25093" "10987" ...  
  $ LongGun         : chr [1:2640] "28933" "3822" "14562" "12719" ...  
  $ Other           : chr [1:2640] "1855" "430" "1942" "587" ...  
  $ Multiple        : chr [1:2640] "1388" "233" "1395" "510" ...  
  $ Admin           : chr [1:2640] "0" "0" "0" "9" ...  
  $ Prepawn_HG      : chr [1:2640] "43" "0" "12" "12" ...  
  $ Prepawn_LG      : chr [1:2640] "13" "0" "4" "15" ...  
  $ Prepawn_Other   : chr [1:2640] "1" "0" "3" "2" ...  
  $ Redemption_HG   : chr [1:2640] "2447" "123" "1082" "1043" ...  
  $ Redemption_LG   : chr [1:2640] "1121" "84" "349" "935" ...  
  $ Redemption_Other: chr [1:2640] "9" "1" "4" "2" ...  
  $ Return/Disp_HG  : chr [1:2640] "25" "49" "170" "0" ...  
  $ Return/Disp_LG  : chr [1:2640] "0" "31" "9" "0" ...  
  $ Return/Disp_Other: chr [1:2640] "0" "0" "0" "0" ...  
  $ Rentals_HG      : chr [1:2640] "0" "0" "0" "0" ...  
  $ Rentals_LG      : chr [1:2640] "0" "0" "0" "0" ...  
  $ PrivateSale_HG  : chr [1:2640] "30" "9" "21" "2" ...  
  $ PrivateSale_LG  : chr [1:2640] "23" "10" "8" "14" ...  
  $ PrivateSale_Other: chr [1:2640] "5" "1" "3" "1" ...
```

```

$ RetToSeller_HG : chr [1:2640] "2" "0" "0" "0" ...
$ RetToSeller_LG : chr [1:2640] "3" "1" "1" "2" ...
$ RetToSeller_Other: chr [1:2640] "0" "0" "0" "0" ...
$ Totals : chr [1:2640] "100927" "8896" "54547" "31618" ...
$ Month_Year : chr [1:2640] "December - 2020" "December - 2020"
"December - 2020" "December - 2020" ...
$ Year : num [1:2640] 2020 2020 2020 2020 2020 2020 2020 2020 2020
2020 2020 ...

```

We can see that the columns are not in proper format so we will address them in the next step and fix them for our analysis.

Data Transformation

In this step, we will be transforming *character* columns into *numerical* format for further analysis.

```

# Convert chr columns to numeric
combined_df <- combined_df %>%
  mutate(across(c("Permit", "Permit_Recheck", "HandGun", "LongGun",
                  "Other", "Multiple", "Admin", "Prepawn_HG",
                  "Prepawn_LG", "Prepawn_Other", "Redemption_HG",
                  "Redemption_LG", "Redemption_Other", "Return/Disp_HG",
                  "Return/Disp_LG", "Return/Disp_Other", "Rentals_HG",
                  "Rentals_LG", "PrivateSale_HG", "PrivateSale_LG",
                  "PrivateSale_Other", "RetToSeller_HG",
                  "RetToSeller_LG", "RetToSeller_Other", "Totals"),
             as.numeric))

str(combined_df)

tibble [2,640 × 28] (S3: tbl_df/tbl/data.frame)
 $ State      : chr [1:2640] "Alabama" "Alaska" "Arizona" "Arkansas"
...
 $ Permit      : num [1:2640] 33421 441 9043 3878 26034 ...
 $ Permit_Recheck : num [1:2640] 505 3 846 900 1788 ...
 $ HandGun     : num [1:2640] 31103 3658 25093 10987 59909 ...
 $ LongGun     : num [1:2640] 28933 3822 14562 12719 39389 ...
 $ Other       : num [1:2640] 1855 430 1942 587 6932 ...
 $ Multiple    : num [1:2640] 1388 233 1395 510 0 ...
 $ Admin       : num [1:2640] 0 0 0 9 0 0 0 0 0 1 ...
 $ Prepawn_HG  : num [1:2640] 43 0 12 12 0 0 0 0 0 16 ...
 $ Prepawn_LG  : num [1:2640] 13 0 4 15 0 0 0 0 0 5 ...
 $ Prepawn_Other : num [1:2640] 1 0 3 2 0 0 0 0 0 0 ...
 $ Redemption_HG : num [1:2640] 2447 123 1082 1043 450 ...
 $ Redemption_LG : num [1:2640] 1121 84 349 935 273 ...
 $ Redemption_Other : num [1:2640] 9 1 4 2 20 0 0 0 0 10 ...
 $ Return/Disp_HG : num [1:2640] 25 49 170 0 1 268 0 43 2 987 ...
 $ Return/Disp_LG : num [1:2640] 0 31 9 0 2 38 0 0 0 95 ...
 $ Return/Disp_Other: num [1:2640] 0 0 0 0 14 2 0 0 6 21 ...

```

```

$ Rentals_HG      : num [1:2640] 0 0 0 0 0 0 0 0 0 0 ...
$ Rentals_LG      : num [1:2640] 0 0 0 0 0 0 0 0 0 0 ...
$ PrivateSale_HG  : num [1:2640] 30 9 21 2 523 0 0 81 0 400 ...
$ PrivateSale_LG  : num [1:2640] 23 10 8 14 27 0 0 42 0 333 ...
$ PrivateSale_Other: num [1:2640] 5 1 3 1 0 0 0 6 0 94 ...
$ RetToSeller_HG  : num [1:2640] 2 0 0 0 59 0 0 1 0 38 ...
$ RetToSeller_LG  : num [1:2640] 3 1 1 2 21 0 0 1 0 42 ...
$ RetToSeller_Other: num [1:2640] 0 0 0 0 0 0 0 0 0 1 ...
$ Totals          : num [1:2640] 100927 8896 54547 31618 135442 ...
$ Month_Year      : chr [1:2640] "December - 2020" "December - 2020"
"December - 2020" "December - 2020" ...
$ Year            : num [1:2640] 2020 2020 2020 2020 2020 2020 2020 2020
2020 2020 ...

```

We have successfully fixed the columns and transformed them into their correct format. Now we will move towards the exploratory data analysis.

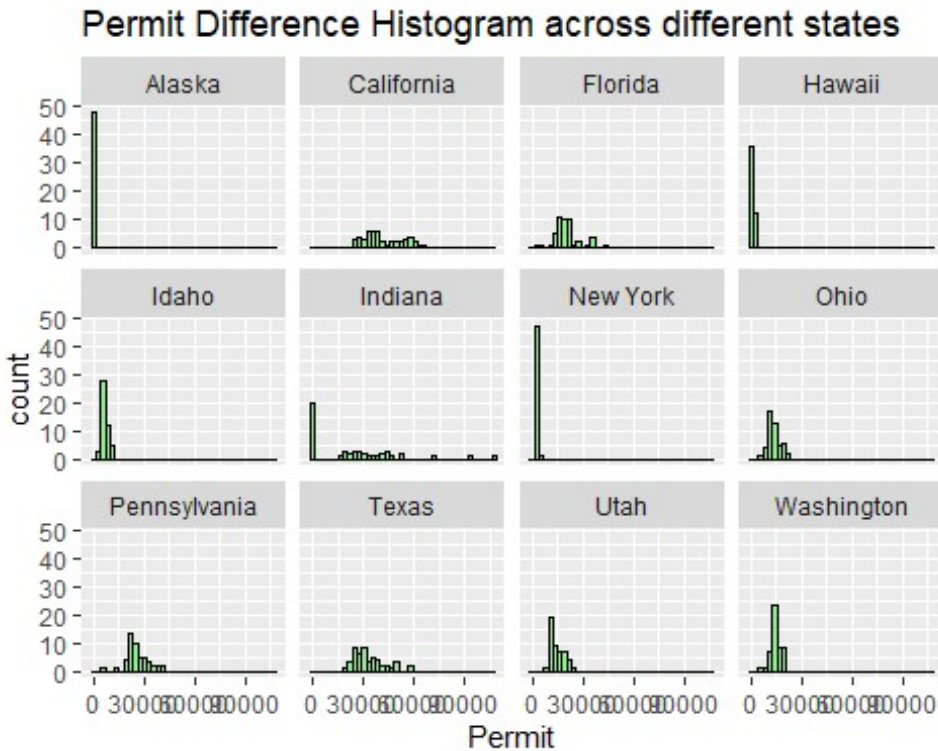
Exploratory Data Analysis (EDA)

EDA is a very crucial early step in the data analysis workflow that helps understand the overall data. Therefore, we will be performing *preliminary analysis* to understand the datasets characteristics, formats, and potential issues before moving on to answering critical research questions.

```

ggplot(combined_df %>% filter(State %in% c('Texas', 'California', 'Florida',
                                           'Alaska', 'Washington', 'Idaho',
                                           'Ohio', 'New York', 'Hawaii',
                                           'Pennsylvania', 'Indiana',
                                           'Utah'))),
  aes(x = Permit)) +
  geom_histogram(bins = 40, col = "black", fill = "lightgreen") +
  labs(title = ("Permit Difference Histogram across different states")) +
  facet_wrap(vars(State))

```

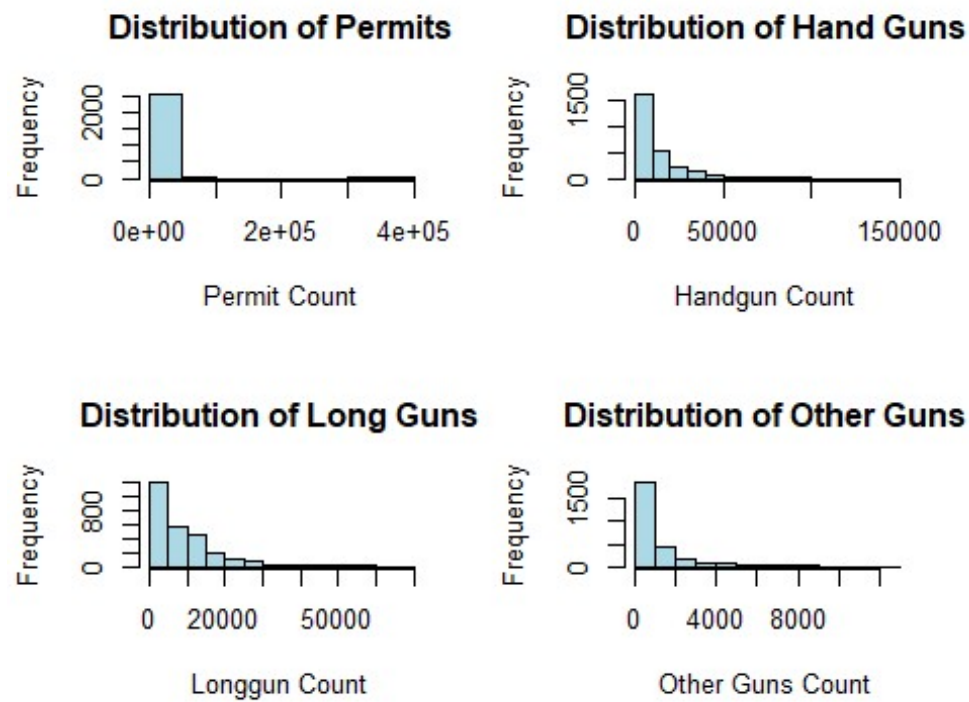


For the sake of simplicity, we handpicked few states (**12**) out of the total 50 states, to see how the permits distribution of the states vary across state to state. The histogram shows that there is a mix and match of *approximate normal* and mostly right-skewed distribution for Permits across different states.

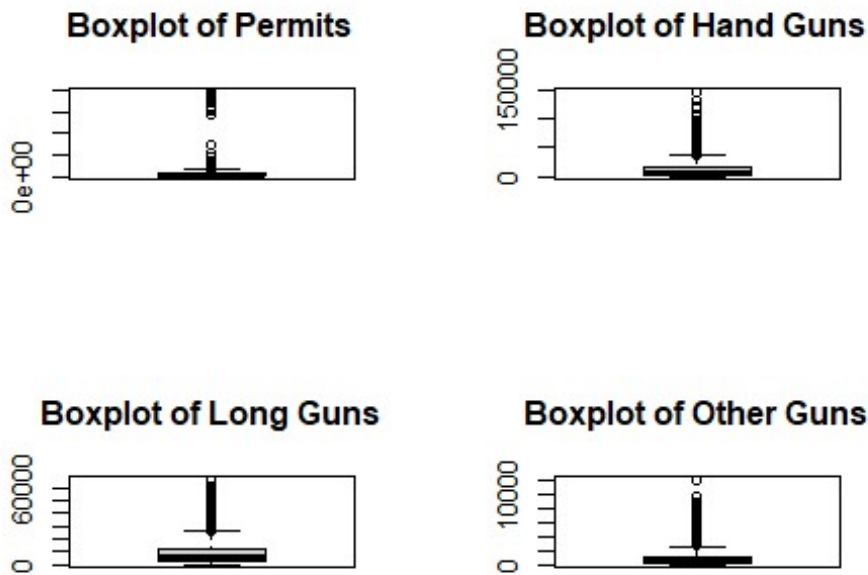
Now we will analyze to understand the distribution of other variables of interest like “Permit Count”, “HandGun”, “Longgun”, and “Other Guns” which will be used for answering the research questions.

```
par(mfrow = c(2, 2)) # To show graphs together

hist(combined_df$Permit, main = "Distribution of Permits", xlab = "Permit
Count", col = "lightblue")
hist(combined_df$HandGun, main = "Distribution of Hand Guns", xlab = "Handgun
Count", col = "lightblue")
hist(combined_df$LongGun, main = "Distribution of Long Guns", xlab = "Longgun
Count", col = "lightblue")
hist(combined_df$Other, main = "Distribution of Other Guns", xlab = "Other
Guns Count", col = "lightblue")
```

```
# Boxplots to check for outliers
boxplot(combined_df$Permit, main = "Boxplot of Permits")
boxplot(combined_df$HandGun, main = "Boxplot of Hand Guns")
boxplot(combined_df$LongGun, main = "Boxplot of Long Guns")
boxplot(combined_df$Other, main = "Boxplot of Other Guns")
```



We can see almost all of the histograms of our variables of interest being right skewed.

Notably, in the boxplots, we can observe *significant outliers* and *extreme values*. We will need to investigate these points and handle them.

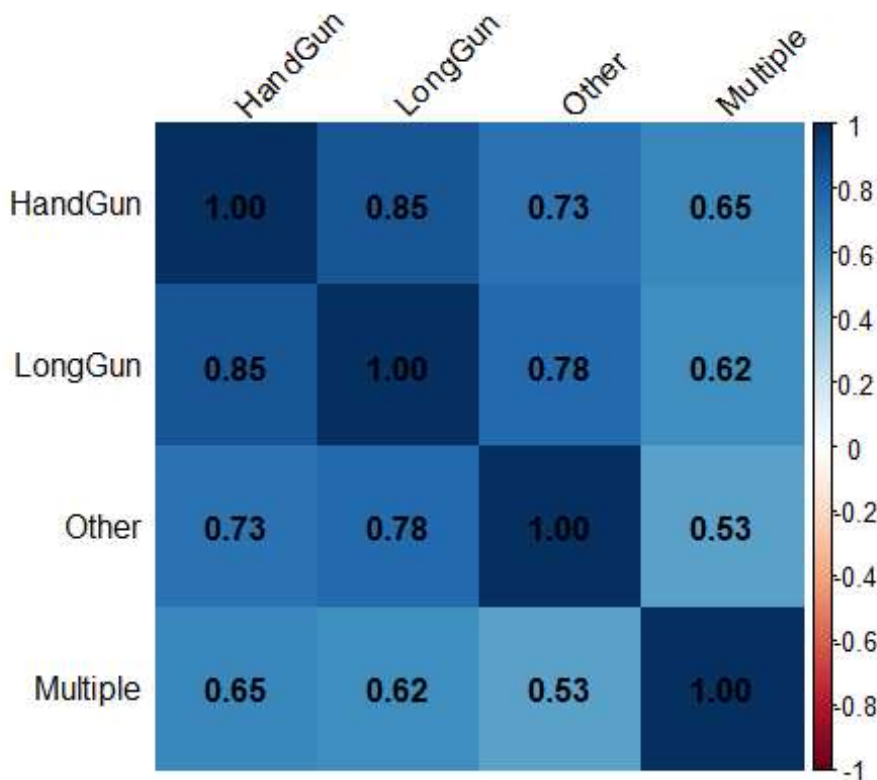
Dealing with Outliers:

At first I thought of removing the outliers but then after giving much thought, I believe that since this data involves real-world data and comes from a reputable source (**FBI NICS Firearms**), it's very likely that most of the extreme values in this dataset is actual real data and real-world data is supposed to be messy. Therefore for our analysis we will keep the outliers and will not blindly remove them as they may be representing useful data and removing it blindly can impact our analysis and lead to inaccurate results.

Correlation Matrix for different gun types

```
cor_matrix <- cor(combined_df[c("HandGun", "LongGun", "Other", "Multiple")],
                  use = "complete.obs")

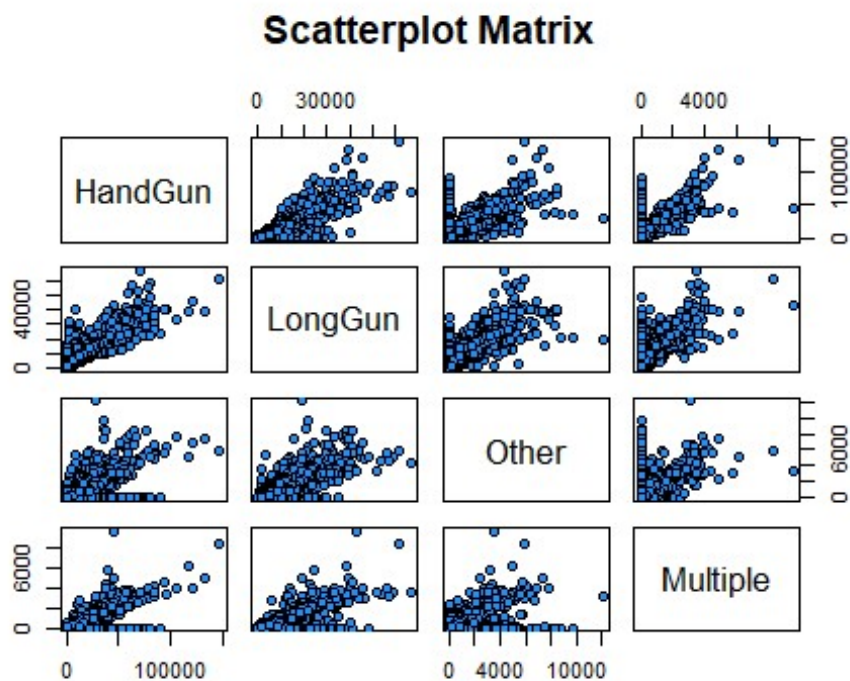
# Plot heatmap of correlation matrix
corrplot(cor_matrix, method = "color", addCoef.col = "black",
         tl.col = "black", tl.srt = 45)
```



As we can see from the correlation matrix, that there is a high correlation between the gun types (Handgun, Long gun, Other, Multiple) which suggests shared seasonal fluctuations. For example, a higher correlation between different gun type permits suggests that people might be more likely to get gun permits based on socio-political instability, or maybe due to certain hunting seasons.

Scatterplot Matrix

```
pairs(combined_df[c("HandGun", "LongGun", "Other", "Multiple")],  
      main = "Scatterplot Matrix", pch = 21, bg = "dodgerblue")
```



To further confirm and strengthen our claim we are also visualizing scatterplots of different gun types. This further strengthens that all gun permits shows **strong to moderate** correlation between each other.

Explanatory Data Analysis

Now we will be communicating our findings, revealing interesting patterns, and answering the previously stated research questions.

Research Question 1:

What is the overall trend of different firearm type permits (Declining or Rising in US) across 2017-2020?

```
yearly_summary <- combined_df %>%
  group_by(Year) %>%
  summarize(Total_Permits = sum(Permit),
            Total_Handguns = sum(HandGun),
            Total_LongGuns = sum(LongGun),
            Total_Other = sum(Other))

print(yearly_summary)

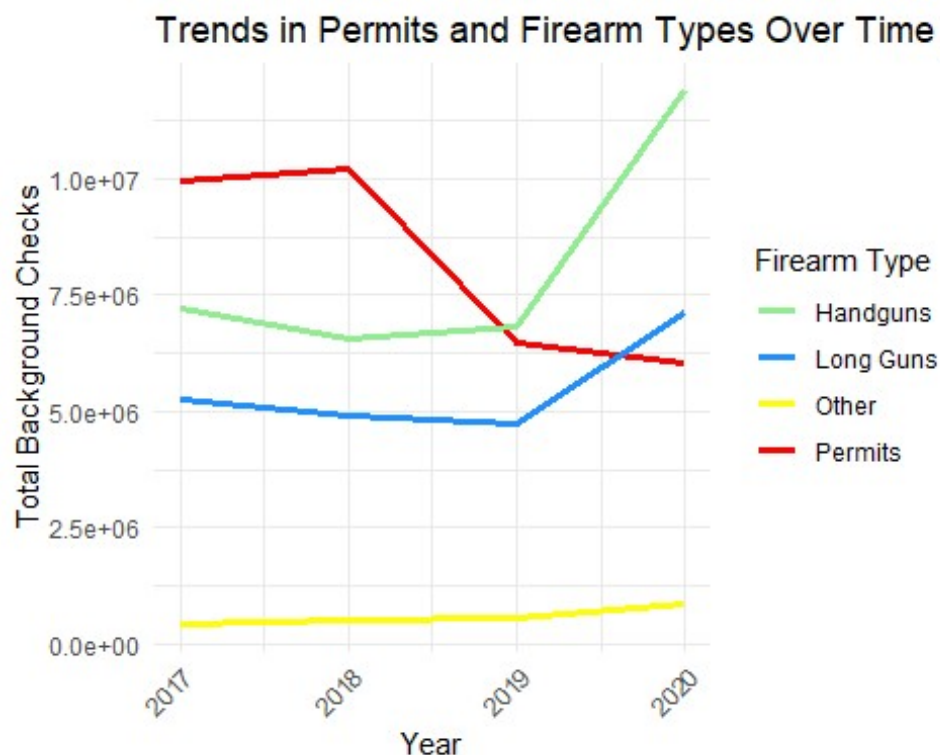
# A tibble: 4 × 5
  Year Total_Permits Total_Handguns Total_LongGuns Total_Other
```

	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2017	9941684	7226979	5234757	399993
2	2018	10229162	6576111	4916533	494299
3	2019	6482780	6802167	4744298	543538
4	2020	6027166	11897521	7132864	854836

```

ggplot(yearly_summary, aes(x = Year)) +
  geom_line(aes(y = Total_Permits, color = "Permits"), size = 1.2) +
  geom_line(aes(y = Total_Handguns, color = "Handguns"), size = 1.2) +
  geom_line(aes(y = Total_LongGuns, color = "Long Guns"), size = 1.2) +
  geom_line(aes(y = Total_Other, color = "Other"), size = 1.2) +
  scale_color_manual(
    name = "Firearm Type",
    values = c(
      "Permits" = "red", "Handguns" = "lightgreen",
      "Long Guns" = "dodgerblue", "Other" = "yellow"
    )
  ) +
  labs(
    title = "Trends in Permits and Firearm Types Over Time (2017-2020)",
    x = "Year",
    y = "Total Background Checks"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



The above line graph suggests that gun interest, particularly in **handguns** permits rose tremendously in 2020, which might be due to COVID'19 Pandemic or election year uncertainty or economic insecurity, but at the moment we are not sure and this is only a

speculation. We can clearly see that there is a huge and growing demand for personal protection among american people.

However, the apparent decline in **Permits** background check may be due to the reporting differences or perhaps state level permit checks which are not reported to **NICS** rather than a true decline in gun permit background checks. We will be investigating and explaining this in more detail in the next analysis.

Research Question 2:

Are handguns replacing long guns in popularity over time (2017-2021) ?

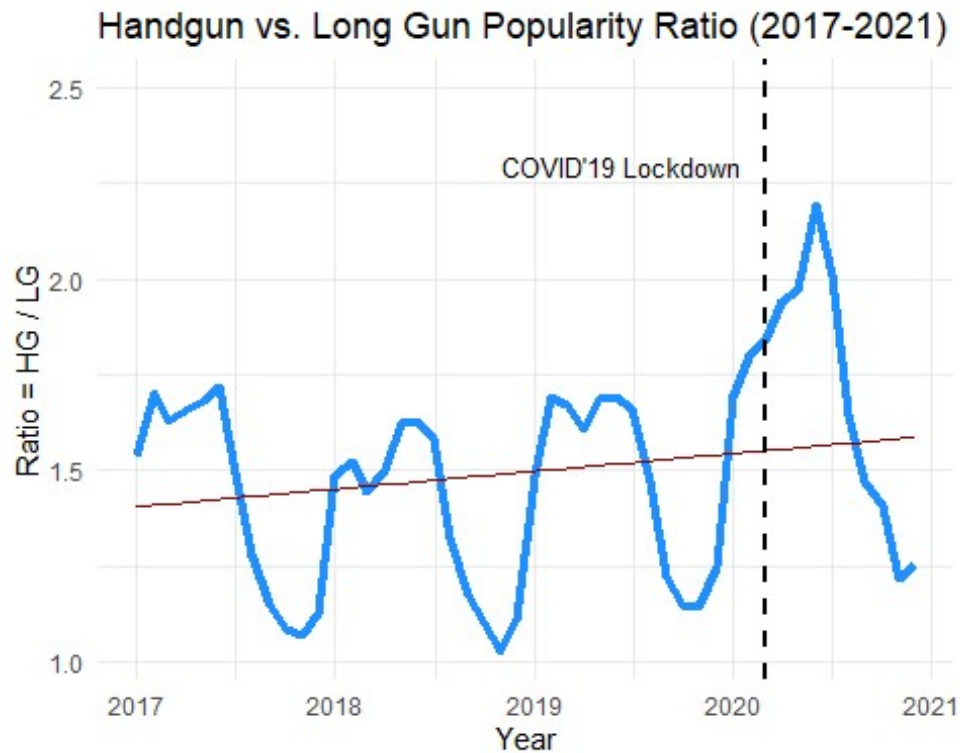
$$\text{Hand Gun Ratio} = \frac{\text{Total HandGun}}{\text{Total LongGun}}$$

```
ratio_df <- combined_df %>%
  group_by(Date = parse_date_time(Month_Year, orders = "B Y")) %>%
  summarise(
    Total_Handguns = sum(HandGun, na.rm = TRUE),
    Total_LongGuns = sum(LongGun, na.rm = TRUE)
  )

# Creating ratio = HG / LG
ratio_df = ratio_df %>% mutate(HG_LG_Ratio = Total_Handguns / Total_LongGuns)

ggplot(ratio_df, aes(x = Date, y = HG_LG_Ratio)) +
  geom_line(color = "dodgerblue", size=1.5) +
  geom_smooth(method = "lm", se = FALSE, color = "darkred", size=0.6) +
  geom_vline(xintercept = as.POSIXct("2020-03-01"), linetype =
"dashed",size=1) +
  annotate("text", x = as.POSIXct("2020-03-01"), y = 2.5,
    label = "COVID'19 Lockdown", vjust = 4, size = 3.5, hjust = 1.1) +
  labs(title = "Handgun vs. Long Gun Popularity Ratio (2017-2021)",
    x = "Year", y = "Ratio = HG / LG") + theme_minimal()

`geom_smooth()` using formula = 'y ~ x'
```



From the line graph, we can clearly see that the graph has an *upward trend* which means handguns are becoming more and more popular which may suggest that more American people are purchasing handguns for **self-defense**. We also see a spike in ratio starting in year 2020, likely to be triggered by political uncertainty, and the **COVID-19 pandemic** as claimed previously. However, it's important to point out that this remains speculative and cannot be confirmed solely from this data alone. This trend could reflect changing attitude towards self defense and concerns around personal safety. The linear regression line also shows a *gradual increase* in the HG/LG ratio which confirms a shift in preference toward handguns. In conclusion, our previous claim is further strengthened through this analysis that handguns are getting more popular among American population than long guns.

Research Question 3:

In which year did the United States experience the peak in total firearm-related permit background checks and what might this indicate about public sentiments?

To further verify our claim that COVID 2020 was the peak moment when all majority of firearm permits surged, we will perform the analysis with the combined permits count of all different permits to get a true picture.

Total Permits

$$= \text{Permits} + \text{HG Permit} + \text{LG Permit} + \text{Multiple Permit} + \text{AdminPermit} + \text{OtherPermit}$$

```

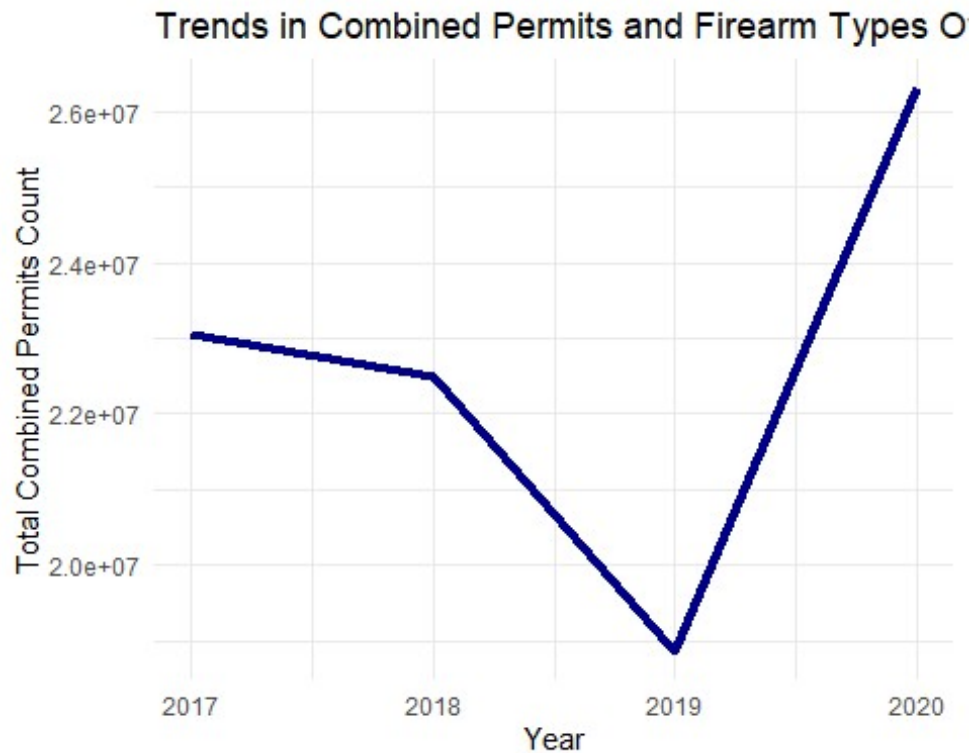
yearly_summary <- combined_df %>%
  group_by(Year) %>%
  dplyr::summarize(
    Total_Permits = sum(Permit, na.rm = TRUE),
    Total_Handguns = sum(HandGun, na.rm = TRUE),
    Total_LongGuns = sum(LongGun, na.rm = TRUE),
    Total_Other = sum(Other, na.rm = TRUE),
    Total_Multiple = sum(Multiple, na.rm = TRUE),
    Total_Admin = sum(Admin, na.rm=TRUE))

print(yearly_summary)

# A tibble: 4 × 7
  Year Total_Permits Total_Handguns Total_LongGuns Total_Other
Total_Multiple
  <dbl>          <dbl>          <dbl>          <dbl>          <dbl>
<dbl>
1  2017          9941684          7226979          5234757          399993
236167
2  2018          10229162          6576111          4916533          494299
257088
3  2019          6482780          6802167          4744298          543538
256524
4  2020          6027166          11897521          7132864          854836
393814
# i 1 more variable: Total_Admin <dbl>

ggplot(yearly_summary, aes(x = Year)) +
  # geom_line(aes(y =
(Total_Permits+Total_Handguns+Total_LongGuns+Total_Other+Total_Multiple) )
  geom_line(aes(y =
(Total_Permits+Total_Handguns+Total_LongGuns+Total_Other+Total_Multiple+Total
_Admin) ), color = "darkblue", size = 1.5) +
  labs(title = "Trends in Combined Permits and Firearm Types Over Time",
        x = "Year", y = "Total Combined Permits Count") +
  theme_minimal()

```

The combined permits analysis confirms that the year **2020** actually experienced the **highest** number of total firearm-related permit background checks, which includes both standard and administrative types. This trend likely reflects increased public concern for safety, as well as an increased desire for security and firearm access during a year marked by the **COVID-19 pandemic, protests, and political uncertainty**. The continued high levels in 2021 (near the end of 2020) suggest that this was not because of only one factor but due to a combination of various factors which contributed to this extent of growth and firearm demand from the public. This pattern further supports our previous hypothesis that **external crises** trigger increased gun interest

Research Question 4:

What are the top 10 states with the most and least gun permits across US (2017-2021) ?

```
state_permits <- combined_df %>%
  # filter(State != 'Kentucky') %>%
  group_by(State) %>%
  dplyr::summarize(Total_Permits = sum(Permit, na.rm = TRUE)) %>%
  arrange(desc(Total_Permits)) # Sort the data by desc total permits

# Filtering the top 10 states
top_10_states <- state_permits[1:10, ]

top_10_states <- top_10_states %>%
```

```

mutate(Percentage = Total_Permits / sum(Total_Permits) * 100)

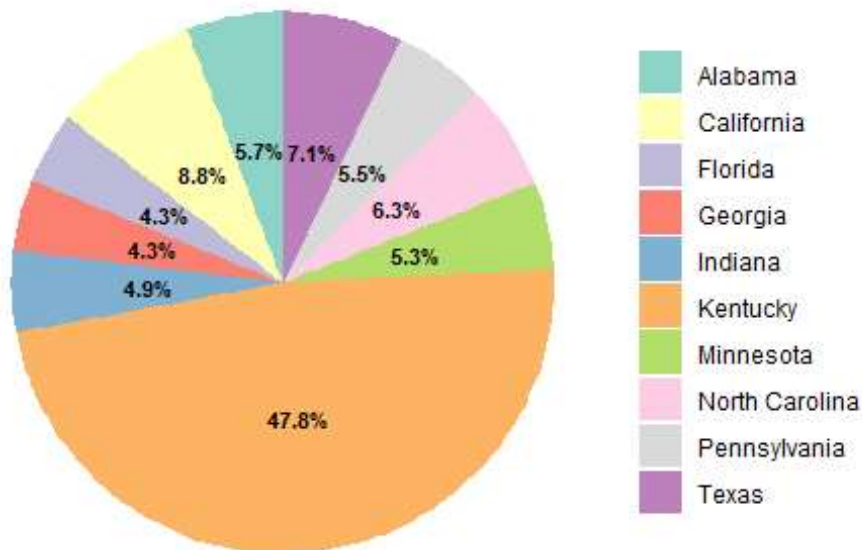
print(top_10_states)

# A tibble: 10 × 3
  State      Total_Permits Percentage
  <chr>          <dbl>         <dbl>
1 Kentucky      10820756         47.8
2 California     1986061          8.78
3 Texas          1610447          7.12
4 North Carolina 1430574          6.32
5 Alabama        1298749          5.74
6 Pennsylvania    1252996          5.54
7 Minnesota       1189867          5.26
8 Indiana         1102058          4.87
9 Florida         972540          4.30
10 Georgia        964232          4.26

# Pie chart
ggplot(top_10_states, aes(x = "States", y = Total_Permits, fill = State)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(
    title = "Total Permits by State (2017-2021)",
    caption = "Data source: FBI Firearms Checks (NICS)"
  ) +
  theme_void() +
  theme(
    legend.title = element_blank(),
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12),
    plot.caption = element_text(size = 10)
  ) +
  geom_text(aes(label = paste0(round(Percentage, 1), "%")),
    position = position_stack(vjust = 0.5),
    color = "black", fontface = "bold", size = 3) +
  scale_fill_manual(values = brewer.pal(10, "Set3")) # Using different
colors

```

Total Permits by State (2017-2021)



Data source: FBI Firearms Checks (NICS)

State of **Kentucky** has the most permits by state, followed by **California** and **Texas**.

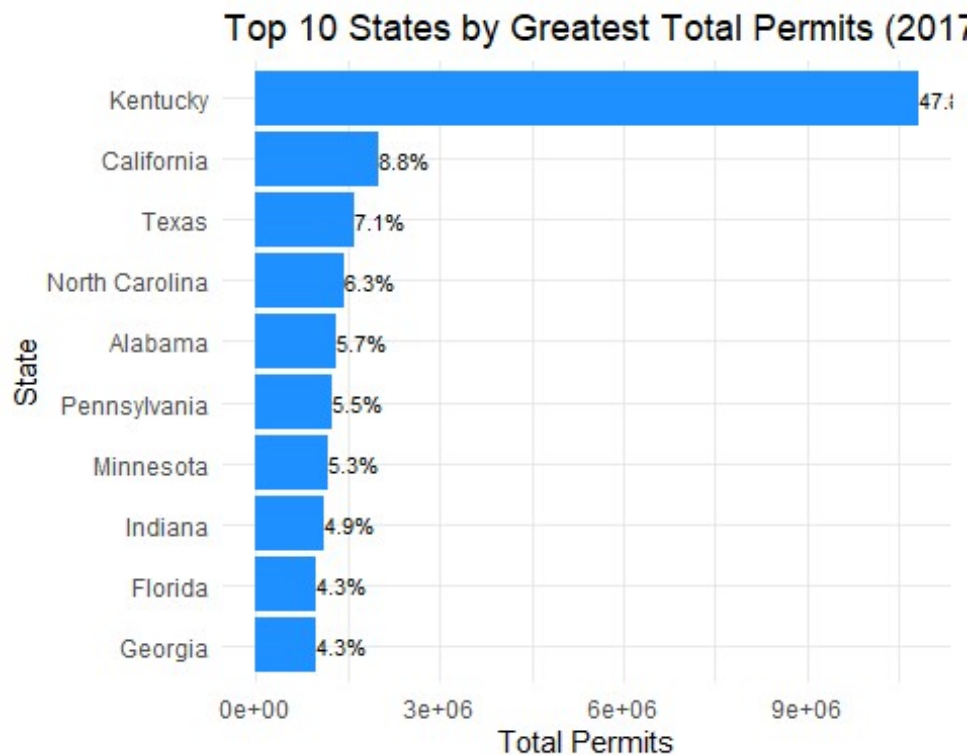
Kentucky recorded the highest number of total permits rechecks in 2017–2018 because it is the only state that requires **monthly automatic rechecks** of its concealed carry permit holders. This is why the state of Kentucky permit counts are so *highly inflated*, which does not represent a true picture.

```
# Summarize by State
state_summary <- combined_df %>%
  group_by(State) %>%
  dplyr::summarize(Total_Permits = sum(Permit),
                   Total_Handguns = sum(HandGun),
                   Total_LongGuns = sum(LongGun)) %>%
  arrange(desc(Total_Permits))

# Filtering top 10 states by Total Permits
state_summary_top10 <- state_summary %>%
  arrange(desc(Total_Permits)) %>%
  head(10) %>%
  mutate(Percent = Total_Permits / sum(Total_Permits) * 100)

ggplot(state_summary_top10, aes(x = reorder(State, Total_Permits),
                                y = Total_Permits)) +
  geom_bar(stat = "identity", fill = "dodgerblue") +
  geom_text(aes(label = paste0(round(Percent,1), "%")),
            hjust = 0, size = 3) + # text of percentage
```

```
coord_flip() +
labs(title = "Top 10 States by Greatest Total Permits (2017-2021)",
     x = "State",
     y = "Total Permits") +
theme_minimal()
```



The bar graph shows similar result that the State of **Kentucky** has the most total permits by state, followed by **California** and **Texas**.

Now we will move on to analyzing the states with the *least total gun permits* across US.

To better analyze we take **log(Total Permits)** because of huge gap in permit counts for states like *Vermont* and *Delaware*, otherwise the graph shows an unreliable picture of states.

```
# Filtering bottom 10 states by Total Permits
state_summary_top10 <- state_summary %>%
  # filter(Total_Permits !=0) %>%
  arrange((Total_Permits)) %>%      # using ascending
  head(10)

print(state_summary_top10)
```

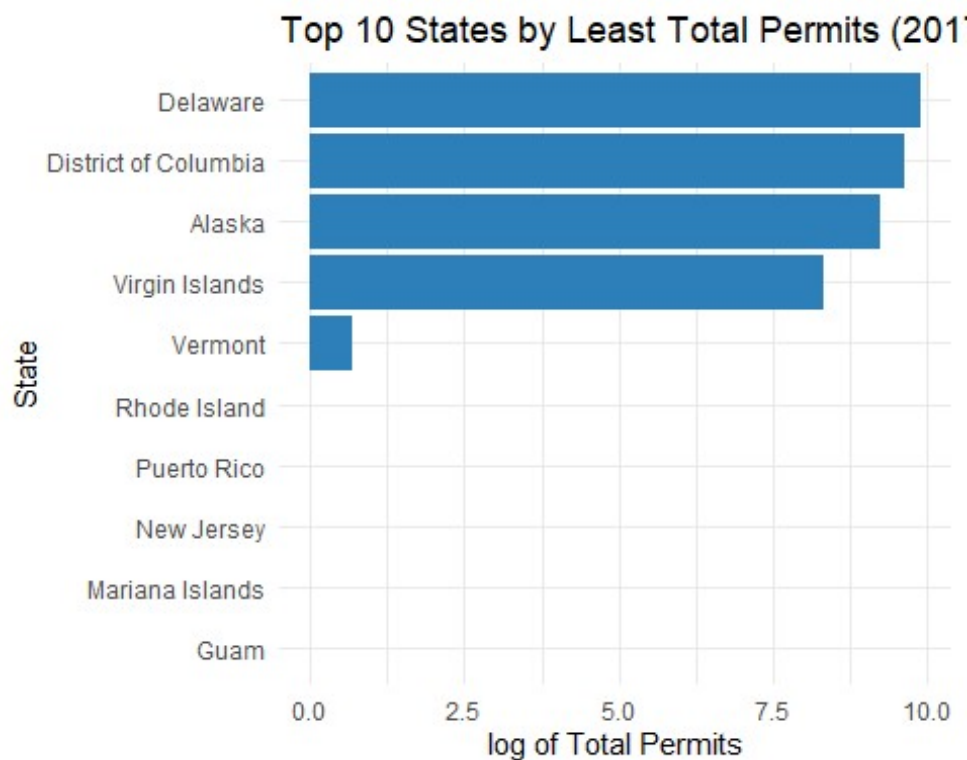
A tibble: 10 × 4

State	Total_Permits	Total_Handguns	Total_LongGuns
<chr>	<dbl>	<dbl>	<dbl>
1 Guam	0	5836	2676

2	Mariana Islands	0	411	266
3	New Jersey	0	259157	190097
4	Puerto Rico	0	90286	9232
5	Rhode Island	0	59785	39508
6	Vermont	2	81164	79390
7	Virgin Islands	4063	1454	158
8	Alaska	10370	143909	134839
9	District of Columbia	15084	6322	205
10	Delaware	19788	104430	77145

Transforming Total_Permits on log scale for clarity otherwise not showing counts

```
ggplot(state_summary_top10, aes(x = reorder(State, Total_Permits),
                                y = log(Total_Permits))) +
  geom_bar(stat = "identity", fill = "#2c7fb8") +
  coord_flip() +
  labs(title = "Top 10 States by Least Total Permits (2017-2021)",
       x = "State", y = "log of Total Permits") + theme_minimal()
```



We have discovered some interesting insights from the above bargraph. We can see that states *Guam*, *Mariana Islands*, *New Jersey*, *Puerto Rico* and *Rhode Islands* have a total of 0 Permit Background checks from the year 2017-2021. This shows that these 5 states have the most strictest gun laws in all across US.

Rhode Island:

In Rhode Island, a specific permit is not required to own a rifle, shotgun, or handgun in Rhode Island, but a handgun safety certificate (*Blue Card*) is necessary to purchase a handgun. This may explain why Rhode Island showed 0 total permits in the above analysis.

Puerto Rico:

Several factors, including legislative changes and social issues, affected gun ownership in Puerto Rico in 2019. In 2018 because of restrictive licensing system, obtaining a gun license in Puerto Rico was a highly difficult and expensive process. It was a “may-issue” system, meaning permits were granted on a discretionary basis and were rarely issued to ordinary citizens. Furthermore, the permit application involved high fees, fingerprinting, multiple required affidavits, membership in a gun club, and a gun safety course.

Guam:

After researching online I found out that something interesting which is consistent with our results that Guam has some of the **strictest gun laws** in the United States, requiring residents to obtain a *Firearms Identification Card (FID)* and go through an FBI background check to possess firearms.

New Jersey:

New Jersey, and Rhode Island also have a high percentage of their populations with less than **1%** holding permits, according to USCCA (U.S. Concealed Carry Association). This agrees with our analysis and supports it.

Mariana Island:

In 2018, the U.S. territory of American Samoa had one of the lowest civilian firearms per capita counts in the world, with an estimated 0.7 civilian firearms per 100 people. The Commonwealth of the Northern Mariana Islands (CNMI) also had strict gun laws in 2018, requiring a *Firearm Owner's Identification Card (FOID)* with training and background checks. This study also agrees with our analysis and shows that Mariana Island is very strict regarding gun laws.

For this analysis, we will be filtering out states which had 0 total permits and will be comparing the rest of the states.

```
# Filtering top 10 states by Total Permits
state_summary_top10 <- state_summary %>%
  filter(Total_Permits !=0) %>%
  arrange((Total_Permits)) %>%
  head(10)
```

```
print(state_summary_top10)
```

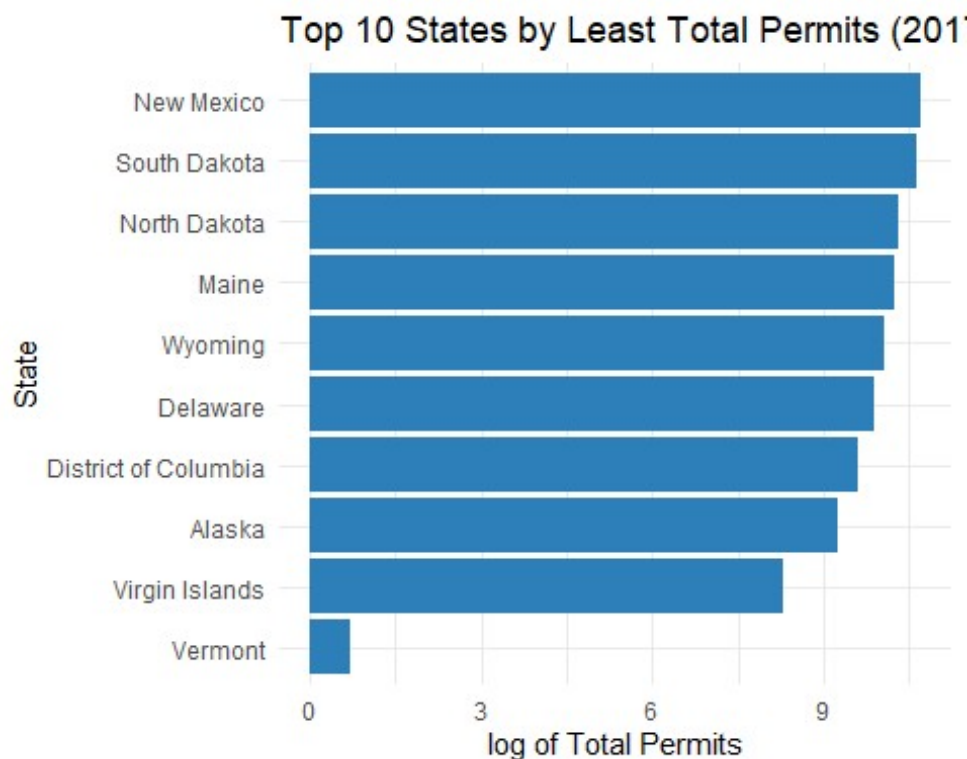
```
# A tibble: 10 × 4
```

State	Total_Permits	Total_Handguns	Total_LongGuns
<chr>	<dbl>	<dbl>	<dbl>

1	Vermont	2	81164	79390
2	Virgin Islands	4063	1454	158
3	Alaska	10370	143909	134839
4	District of Columbia	15084	6322	205
5	Delaware	19788	104430	77145
6	Wyoming	23241	103668	109579
7	Maine	28420	176588	175516
8	North Dakota	29993	89924	133656
9	South Dakota	41534	135461	176563
10	New Mexico	44951	320985	217932

Transforming Total_Permits on log scale for clarity otherwise not showing Total Permits

```
ggplot(state_summary_top10, aes(x = reorder(State, Total_Permits),
                                y = log(Total_Permits))) +
  geom_bar(stat = "identity", fill = "#2c7fb8") +
  coord_flip() +
  labs(title = "Top 10 States by Least Total Permits (2017-2021)",
       x = "State", y = "log of Total Permits") + theme_minimal()
```



The state with the lowest total gun permits is **Vermont**. After searching online about Vermont I found out something interesting which makes sense:

Vermont has close to zero total permits because the state does not require gun permits because it has “*constitutional carry (permitless carry)*,” allowing residents to carry

handguns without a permit, and it has historically had permissive gun laws, therefore the results from our analysis aligns with the real-world.

The state of **Vermont** has historically had very minimal restrictions on firearms. For example, there is no permit required to purchase a firearm.

NOTE:

Some states have “Constitutional Carry” laws that don’t require a permit to carry a concealed weapon, which means *a permit holder may not appear on statistics* for permit rates.

Research Question 5:

Which state had the highest total background checks for handguns, long guns, other guns and multiple firearm types from 2017 to 2020?

```
firearm_totalsummary <- combined_df %>%
  group_by(State) %>%
  summarise(
    Total_Handgun = sum(HandGun, na.rm = TRUE),
    Total_LongGun = sum(LongGun, na.rm = TRUE),
    Total_Other = sum(Other, na.rm = TRUE),
    Total_Multiple = sum(Multiple, na.rm = TRUE)
  )

state_hg=
firearm_totalsummary$State[which.max(firearm_totalsummary$Total_Handgun)]
state_lg=
firearm_totalsummary$State[which.max(firearm_totalsummary$Total_LongGun)]
state_o =
firearm_totalsummary$State[which.max(firearm_totalsummary$Total_Other)]
state_m=
firearm_totalsummary$State[which.max(firearm_totalsummary$Total_Multiple)]

max_hg = max(firearm_totalsummary$Total_Handgun)
max_lg = max(firearm_totalsummary$Total_LongGun)
max_other = max(firearm_totalsummary$Total_Other)
max_multiple = max(firearm_totalsummary$Total_Multiple)

# creating a dataframe with max States and max Permits
top_states <- data.frame(
  Firearm_Type = c("Handgun", "LongGun", "Other", "Multiple"),
  State = c(state_hg, state_lg, state_o, state_m),
  Total = c(max_hg, max_lg, max_other, max_multiple)
)

print(top_states)
```

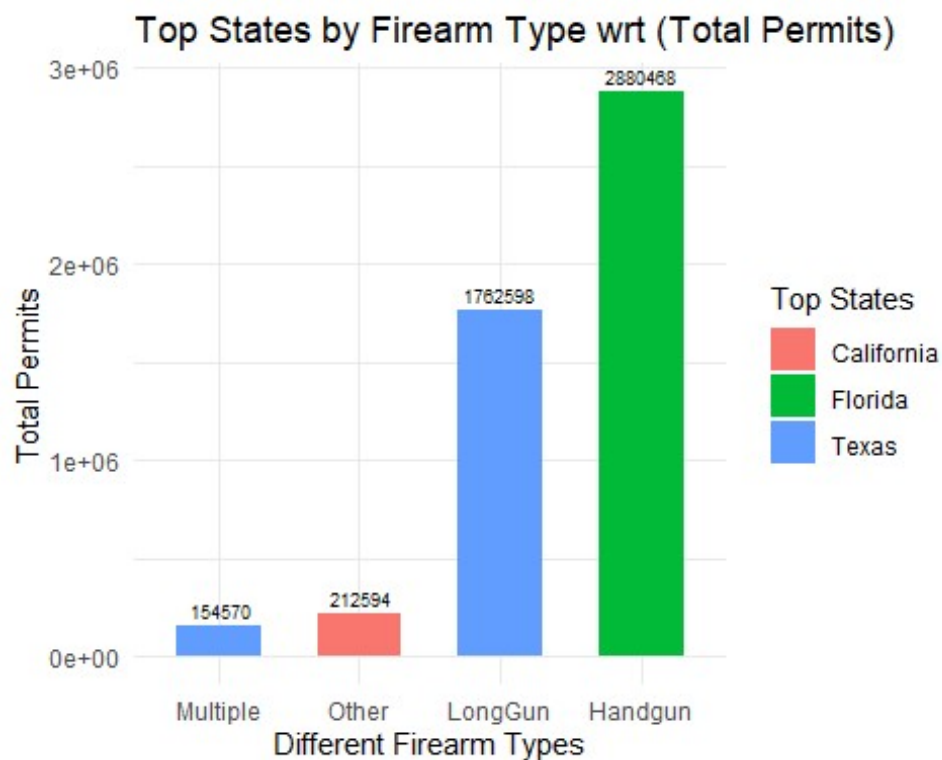


```

Firearm_Type    State    Total
1      Handgun    Florida 2880468
2      LongGun    Texas  1762598
3        Other California  212594
4      Multiple    Texas  154570

ggplot(top_states, aes(reorder(x = Firearm_Type, Total), y = Total, fill =
State)) +
  geom_col(width = 0.6) +
  geom_text(aes(label = paste0(format(Total))),
            vjust = -0.5, size = 2.5) +
  labs(
    title = "Top States by Firearm Type wrt (Total Permits)",
    x = "Different Firearm Types",
    y = "Total Permits",
    fill = "Top States"
  ) +
  theme_minimal()

```



Florida:

From the above analysis we can conclude that Florida leads in Handgun background check permits with approximately **2.9 million** permits, which tells us that Florida is a gun-friendly state and the state policies of Florida encourages handgun purchases for people for self-defense. Moreover, a high rate of civilian handgun ownership indicates that people in Florida are well equipped for self defense.

Texas:

As expected Texas is famous for its gun culture and leads in both categories of “Long Guns” and “Multiple-Gun” purchases suggesting a supportive culture of firearm ownership which includes rifles, and shotguns for hunting or sporting or gun collection. Furthermore, the “Multiple” category points to higher number of purchases of multiple guns by Texans and they are surely not to be messed with.

California:

California leads in the ‘Other’ category which may include (e.g silencers, grenades, etc). This aligns with California’s strict gun laws and regulatory system which requires permits for even unusual firearms which are not even guns.

Research Question 6:

Which states had the highest number of firearm permit background checks from 2017 to 2020?

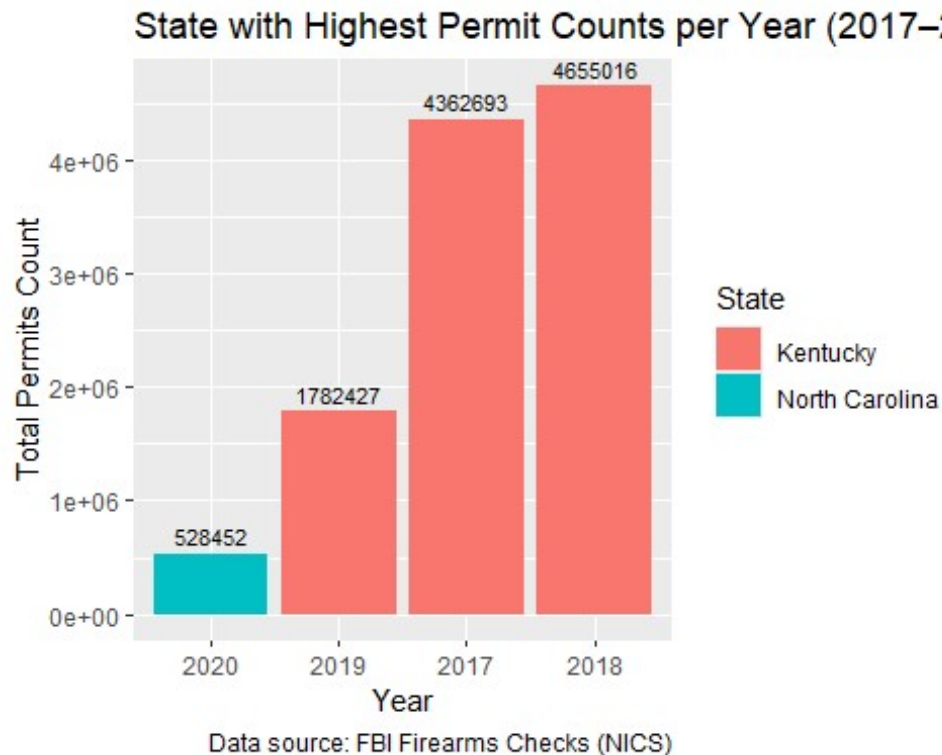
```
permit_statesummary_year <- combined_df %>%
  group_by(Year, State) %>%
  # filter(State != 'Kentucky') %>%
  summarise(Total_Permits = sum(Permit, na.rm = TRUE))

top_states_by_year <- permit_statesummary_year %>%
  group_by(Year) %>%
  filter(Total_Permits == max(Total_Permits)) # To find state with max
permits

print(top_states_by_year)

# A tibble: 4 × 3
# Groups:   Year [4]
   Year State      Total_Permits
  <dbl> <chr>      <dbl>
1  2017 Kentucky    4362693
2  2018 Kentucky    4655016
3  2019 Kentucky    1782427
4  2020 North Carolina  528452

ggplot(top_states_by_year, aes(x = fct_reorder(factor(Year), Total_Permits),
                               y=Total_Permits, fill = State)) +
  geom_col()+geom_text(aes(label = paste0(Total_Permits)), vjust = -0.5, size
= 3) +
  labs(title = "State with Highest Permit Counts per Year (2017-2020)", x =
"Year",
       y = "Total Permits Count", fill = "State",
       caption = "Data source: FBI Firearms Checks (NICS)"
)
```



State of Kentucky has consistently had the highest permits count in 2017-2019 with approximately **4.36 million** permits in 2017, **4.65 million** permits in 2018 and **1.78 million** permits in 2019. However, in the year 2020 North Carolina state has showed the greatest number of permit count. One notable observation is that, the total permit count of Kentucky is extremely higher than other states therefore we will have to investigate it further to find out why.

Why permit counts are so high in Kentucky?

After thoroughly researching and reading articles and news online, I learned that the state of Kentucky runs *monthly permit rechecks* through the NICS (as per FBI documentation). That means every month, existing permit holders are rechecked which explains the significantly higher counts of permits for state of Kentucky as the total permits are artificially inflated and now the high numbers starts making much more sense. Therefore, the high numbers in Kentucky do not reflect demands of new permits issued, but rather existing monthly administrative background checks. On the other hand, North Carolina in 2020 gives a more realistic number for background checks for new permits as compared to Kentucky.

Now we will be removing the state of *Kentucky* from our analysis because of artificially inflated permit counts (due to automated monthly rechecks) and redoing the analysis without the state of Kentucky.

```
permit_statesummary_year <- combined_df %>%
  group_by(Year, State) %>%
```

```
filter(State != 'Kentucky') %>%
summarise(Total_Permits = sum(Permit, na.rm = TRUE))
```

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

```
top_states_by_year <- permit_statessummary_year %>%
  group_by(Year) %>%
  filter(Total_Permits == max(Total_Permits)) # To find state with max
permits
```

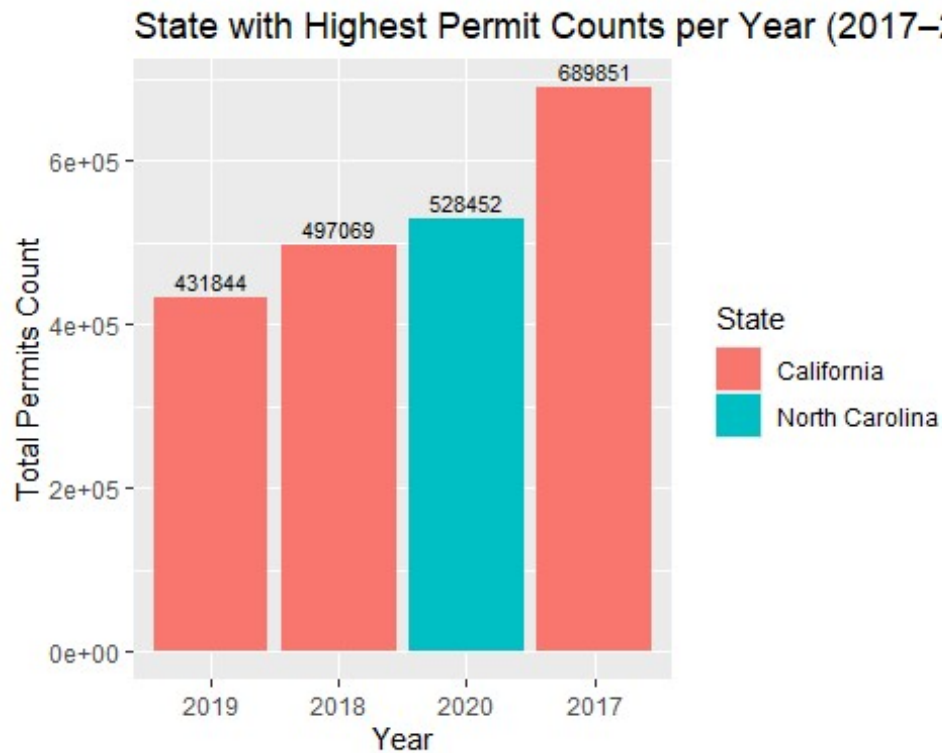
```
print(top_states_by_year)
```

```
# A tibble: 4 × 3
```

```
# Groups:   Year [4]
```

	Year	State	Total_Permits
	<dbl>	<chr>	<dbl>
1	2017	California	689851
2	2018	California	497069
3	2019	California	431844
4	2020	North Carolina	528452

```
ggplot(top_states_by_year, aes(x = fct_reorder(factor(Year), Total_Permits),
                               y=Total_Permits, fill = State)) +
  geom_col()+geom_text(aes(label = paste0(Total_Permits)), vjust = -0.5,size
= 3) +
  labs(title = "State with Highest Permit Counts per Year (2017-2020)", x =
"Year",
       y = "Total Permits Count", fill = "State")
```



From the bargraph, we can see that the state of **California** dominated the total permits count in the years 2017-2019, while in the year 2020 the state of **North Carolina** remains the state with the highest number of permit counts.

Logical reasoning behind California dominating permit counts (2017-2019):

As we already know that California has among the strongest gun permit background check laws in the U.S, requiring them for all firearm purchases and transfers, including private sales. The state was the first to require background checks for ammunition purchases. Moreover, state's large population is also a big factor as its the most populous state in the U.S. When combined with the state's comprehensive regulatory requirements, strong gun laws, this naturally leads to a higher overall number of background checks in California.

Research Question 7:

Is the “Other” firearms category growing, and what might this suggest about trends in niche/tactical firearms?

$$\text{Other Permit Ratio} = \frac{\text{Total Other Permits}}{\text{Total Permit Checks}}$$

```
# Calculate total and proportion of 'Other' over time
other_df <- combined_df %>%
  group_by(Date = parse_date_time(Month_Year, orders = "B Y")) %>%
  summarise(
    Total_Other = sum(Other, na.rm = TRUE),
```

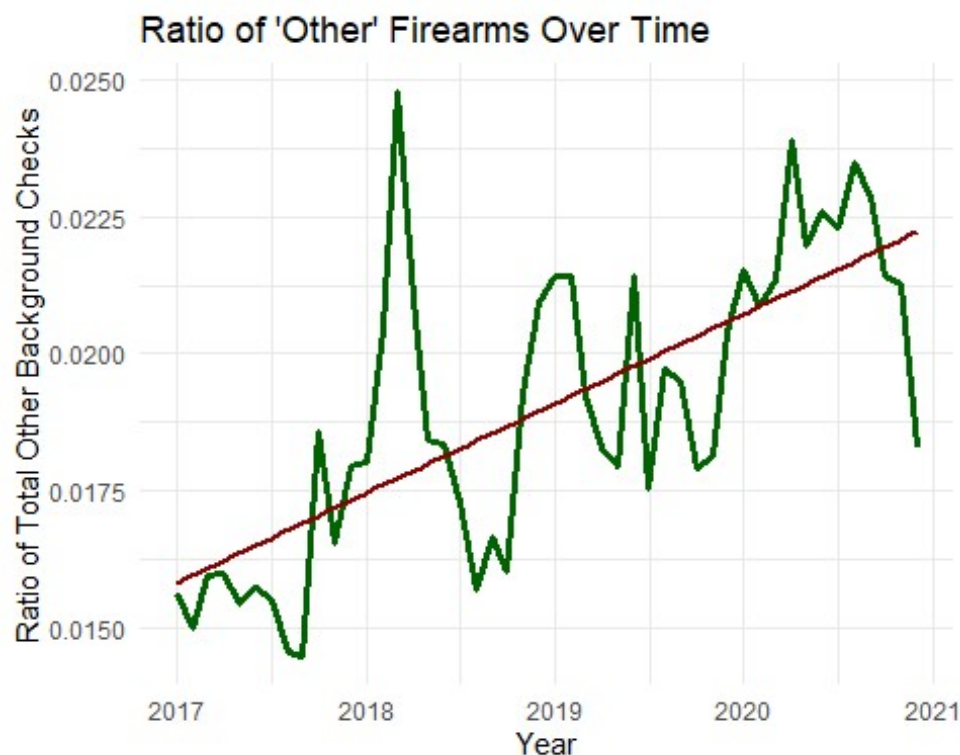
```

    Total_Checks = sum(Totals, na.rm = TRUE)
  ) %>%
  mutate(Other_Proportion = Total_Other / Total_Checks)

ggplot(other_df, aes(x = Date, y = Other_Proportion)) +
  geom_line(color = "darkgreen", size = 1.2) +
  geom_smooth(method = "lm", se = FALSE, color = "darkred") +
  labs(title = "Ratio of 'Other' Firearms Over Time",
       x = "Year", y = "Ratio of Total Other Background Checks") +
  theme_minimal()

`geom_smooth()` using formula = 'y ~ x'

```



From the plot, we can clearly see a **very sharp rise** in Other Permit Category ratio which suggests increasing consumer interest in non-traditional firearms, possibly driven by trends in regulatory gaps. The “Other” category includes firearms that don’t fit legally into traditional handgun or long gun classifications, which can create **loopholes** in firearm regulation and oversight. These types of other weapons like *AR-styled pistols*, *pen guns*, *ghost guns*, and *silencers* are extremely dangerous and can often **bypass stricter regulation** due to technical definitions. People may acquire these weapons without the same **scrutiny** applied to traditional handguns/rifles and can misuse them in criminal activities, as parts can be **disposed off easily**.

Recommendations:

If this growth continues unaddressed, the “Other” category could become a major source for illegal firearm possession and criminal misuse because regulatory laws are outdated with loopholes and there is limited public knowledge about these weapons. This is a **very alarming situation** and this category may evolve into a *regulatory blindspot* in future, essentially undermining public safety efforts and law enforcement investigations. Therefore, addressing this now with targeted policies and strict gun-control policies is not only realistic but extremely essential for national security and preventing future crimes.

Research Question 8:

Do permit rechecks (renewals or ongoing monitoring) vary significantly across states, and what does this indicate about state-level firearm monitoring systems?

This is a very critical question because states which conduct frequent gun permit rechecks may indicate better monitoring system and stronger public safety systems, while other states may only do one-time background checks.

We will be calculating the permit recheck ratio as followed for our analysis:

$$\text{Recheck Ratio} = \frac{\text{Total Rechecks}}{\text{Total Permits}}$$

Permit Recheck Ratio reflects how actively a particular state monitors firearm permit holders over time.

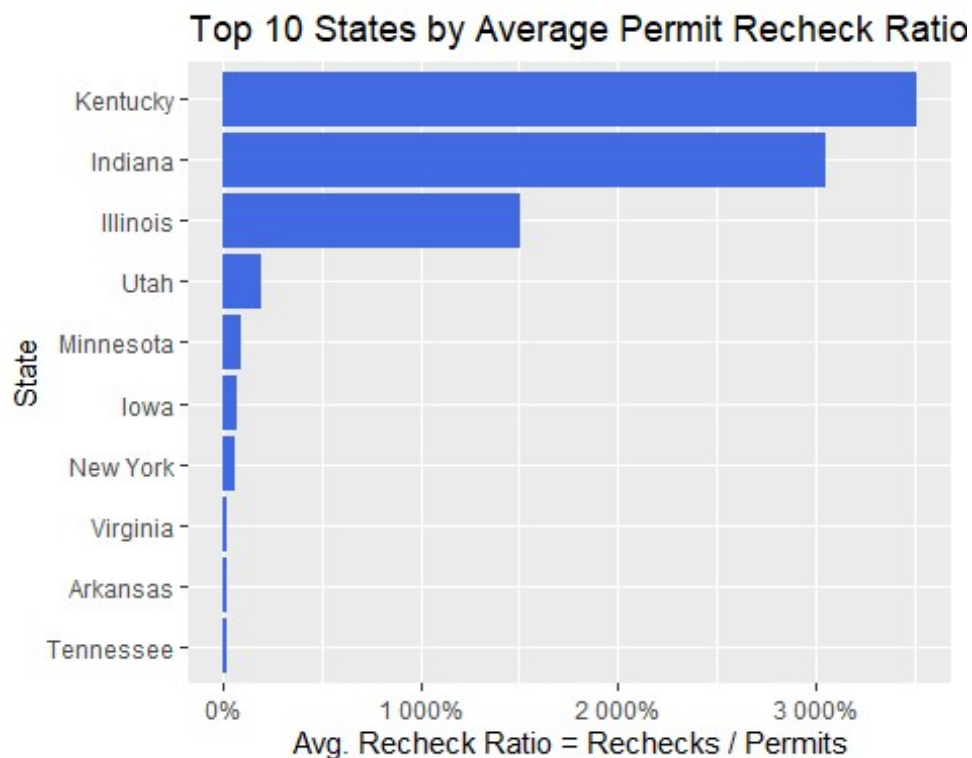
```
permitrecheck_summary <- combined_df %>%
  group_by(State, Year) %>%
  summarise(
    Total_Permits = sum(Permit, na.rm = TRUE),
    Total_Rechecks = sum(Permit_Recheck, na.rm = TRUE),
  )

# Finding Permit Recheck Ratio
permitrecheck_summary = permitrecheck_summary %>%
  mutate(Recheck_Ratio = Total_Rechecks / Total_Permits)

avg_recheck_ratio_by_state <- permitrecheck_summary %>%
  group_by(State) %>%
  summarise(
    Avg_Recheck_Ratio = mean(Recheck_Ratio, na.rm = TRUE),
    Total_Permits = sum(Total_Permits),
    Total_Rechecks = sum(Total_Rechecks),
  ) %>%
  arrange(desc(Avg_Recheck_Ratio))
```

```
# Extracting top 10 states with the highest ratio
topstate_avg_recheck_ratio = avg_recheck_ratio_by_state %>%
  slice_max(Avg_Recheck_Ratio, n = 10)

ggplot(topstate_avg_recheck_ratio, aes(x = reorder(State, Avg_Recheck_Ratio),
                                       y = Avg_Recheck_Ratio)) +
  geom_col(fill = "royalblue") +
  coord_flip() + scale_y_continuous(labels = percent_format(accuracy = 1)) +
  labs(title = "Top 10 States by Average Permit Recheck Ratio (2017-2020)",
       x = "State", y = "Avg. Recheck Ratio = Rechecks / Permits")
```



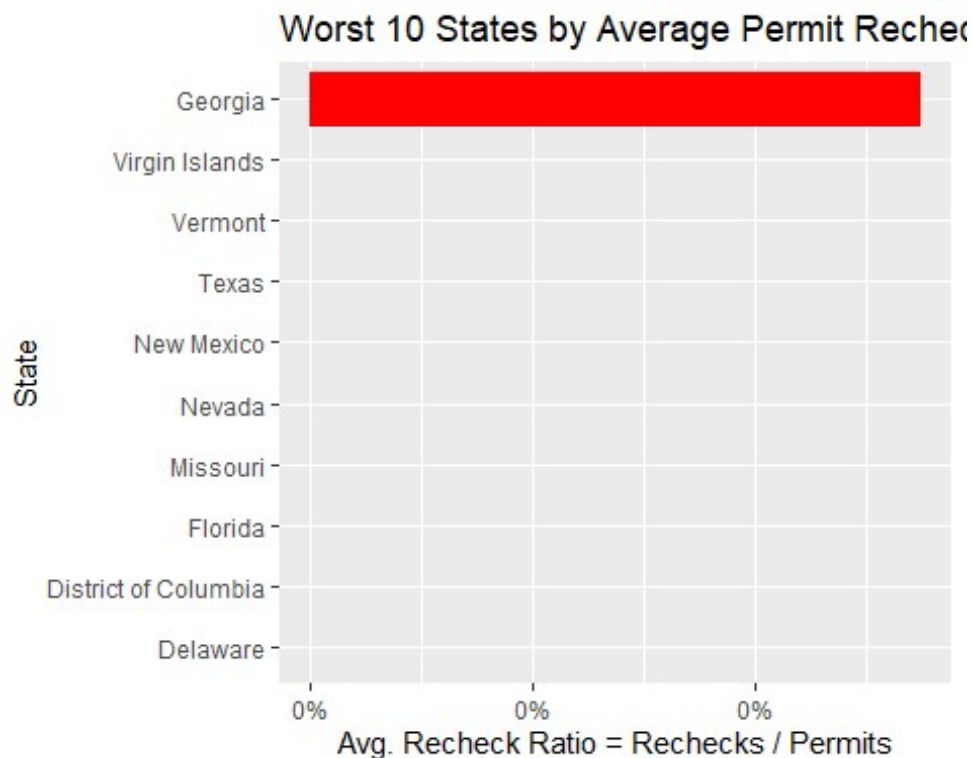
We can also see there is a huge variation in how different US states conduct permit rechecks. As previously seen, **Kentucky** again **dominates** in average permit recheck ratio as compared to all other states with a huge margin. This is likely due to the automated monthly rechecks of all permit holders which shows a very strict and effective monitoring system. An interesting observation is that **Indiana** comes 2nd in terms of highest average recheck ratio and is also quite close to Kentucky. In contrast, many other states show low permit rechecking ratio, potentially pointing to weaker monitoring systems, and should be strengthened and better controlled for public safety. Lastly, states with average recheck ratio close to 1 indicates these states likely perform only one-time checks at issuance.

Now let's analyze the worst performing states in terms of average recheck ratio.

```
# Extracting worst 10 states with lowest ratio
worststate_avg_recheck_ratio = avg_recheck_ratio_by_state %>%
  slice_min(Avg_Recheck_Ratio, n = 10)
```



```
ggplot(worststate_avg_recheck_ratio, aes(x = reorder(State,
Avg_Recheck_Ratio),
                                         y = Avg_Recheck_Ratio)) +
  geom_col(fill = "red") +
  coord_flip() + scale_y_continuous(labels = percent_format(accuracy = 1)) +
  labs(title = "Worst 10 States by Average Permit Recheck Ratio (2017-2020)",
       x = "State", y = "Avg. Recheck Ratio = Rechecks / Permits")
```



From the graph we can observe that the worst performing states with respect to average recheck ratios are *Delaware, District of Columbia, Florida, Missouri, Nevada, New Mexico, Texas, Vermont, Virgin Islands* and *Georgia*. These worst performing states indicates that they lack proactive monitoring systems for permit rechecks, which raises serious public safety concerns, policy gaps, therefore suggesting major policy and regulatory changes to control firearm permits.

Safety Comparison of Kentucky compared to the worst performing states:

If we were to compare the states safety of Kentucky with the worst performing states (in terms of average recheck ratio) then we can see that Kentucky turns out to be a much safer state in most aspects with majority of the worst performing states like *New Mexico, Georgia, Nevada, Delaware, Florida*, etc. Kentucky's permit recheck system is likely contributing positively to its relatively better safety metrics, which may suggests some association. However, causation cannot be established at this point because many other factors influence crimes and safety such as poverty, criminal justice, demographics, and

many more. In conclusion, states with weak monitoring systems should consider more active permit rechecks because even partial increases in permit recheck frequency could help in reducing crimes, gun deaths and ultimately contribute to enhanced public safety.

Research Question 9:

Do states with high handgun background checks also have high firearm-related incidents (crime/suicide)?

This is a very important question which needs to be addressed, by examining whether states with more handgun background checks also experience higher firearm death rates because I personally believe handgun access is a critical factor in both firearm crimes and suicides. Therefore, let's investigate it.

To answer this research question, we will be using an [external dataset](#) of **CDC Firearm Mortality** of all states in the US. In this dataset we will be using **CDC Firearm Mortality Rate** as an *indicator* for death rates across all states. The reason for using this mortality rate indicator is because it controls for population differences (across all states) so it's much better than using *total_deaths* for comparing risk across states accurately and gives a fair comparison across states. The total death counts is influenced by **population size** and big states like (CA, TX, etc) will naturally have more deaths and more checks.

```
# Reading an external dataset of firearm deaths by CDC
cdc_mortality <-
read_csv("C:/Users/ihate/Downloads/f24/data/cdc_firearm_mortality.csv",
show_col_types = FALSE)

# For finding abbreviation of states using state name with abbrev
state_lookup <- data.frame(
  STATE = state.abb,
  State = state.name,
  stringsAsFactors = FALSE
)

# Adding state names in cdc df to join on a common column
cdc_mortality = cdc_mortality %>%
  left_join(state_lookup, by = "STATE")

# Since our NICS Background Check data is only from year=2017 to year=2020
cdc_mortality = cdc_mortality %>%
  filter(YEAR >= 2017 & YEAR <= 2020)

cdc_summary <- cdc_mortality %>%
  group_by(State) %>%
  summarise(
    Avg_Firearm_Death_Rate = mean(RATE, na.rm = TRUE),
    Total_Firearm_Deaths = sum(DEATHS, na.rm = TRUE)
  )
```

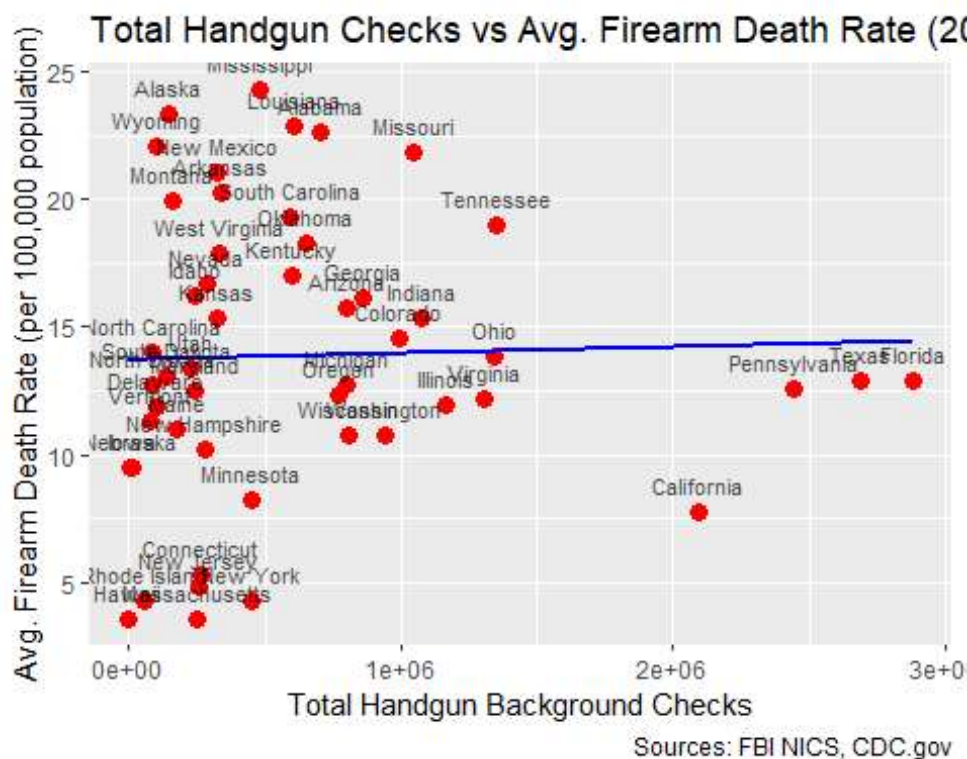
```

handgun_summary <- combined_df %>%
  group_by(State) %>%
  summarise(Total_Handgun_Checks = sum(HandGun, na.rm = TRUE))

# Merging both datasets using State as common variable
merged_data <- handgun_summary %>%
  inner_join(cdc_summary, by = "State")

ggplot(merged_data, aes(x = Total_Handgun_Checks, y = Avg_Firearm_Death_Rate,
                        label = State)) +
  geom_point(color = "red", size = 3) +
  geom_smooth(method = "lm", color = "blue", se = FALSE) +
  geom_text(size = 3, vjust = -1, alpha=0.7) +
  labs(
    title = "Total Handgun Checks vs Avg. Firearm Death Rate (2017-2020)",
    x = "Total Handgun Background Checks",
    y = "Avg. Firearm Death Rate (per 100,000 population)",
    caption = "Sources: FBI NICS, CDC.gov"
  )

```



```

correlation <- cor(merged_data$Total_Handgun_Checks,
                  merged_data$Avg_Firearm_Death_Rate, use = "complete.obs")

print(paste("Correlation b/w Avg Firearm DeathRate indicator and HandGun

```

```

checks:",
      round(correlation, 2)))

[1] "Correlation b/w Avg Firearm DeathRate indicator and HandGun checks:
0.03"

correlation <- cor(merged_data$Total_Handgun_Checks,
                  merged_data$Total_Firearm_Deaths, use = "complete.obs")

print(paste("Correlation b/w Avg Firearm Deaths and HandGun checks:",
            round(correlation, 2)))

[1] "Correlation b/w Avg Firearm Deaths and HandGun checks: 0.89"

```

From the analysis of 2017-2020 FBI NICS Total Handgun background checks and CDC firearm mortality death rates, we can see a **slight positive** relationship ($r \sim 0.03$) which is **almost negligible** and we find **no meaningful correlation** between Total handgun checks vs firearm mortality death rate. This could be explained because the high handgun background check does not necessarily means that people are owning greater number of handguns, it is merely a background check and does not confirms whether a gun was sold or not. Therefore, higher handgun background checks alone does not directly cause higher firearm mortality deaths. Additional factors such as state regulations, permit rechecks, illegal sales, mental health accessibility, and economic conditions likely play a more bigger role. Lastly, few outlier states like *California, Florida, Texas, Pennsylvania, Alabama, and Mississippi* could also be observed.

Comment on Correlation Coefficients:

The correlation between handgun background checks and firearm death rates is very **weak** ($\rho \sim 0.03$), suggesting no clear relationship.

The correlation with firearm deaths ($\rho \sim 0.89$) is likely driven by population size, **not by a true relationship** between handgun access and mortality risk.

Comments about outlier states:

California:

California state shows an interesting pattern of **very high** handgun background checks but **lower mortality death rate**. This could be explained due to California's stronger state regulations with respect to guns and better enforcement of law and order. This might also be due to better mental health infrastructure and better knowledge of gun usage among the public.

Mississippi and Alabama

Both states show **low to moderate** handgun background checks but **extremely high** mortality death rate. These might suggest illegal or private sales of guns and weaker gun regulations in these states which needs to be addressed.

Pennsylvania, Texas and Florida

These states show the **highest** handgun background checks but moderate firearm death rate, which might suggest somewhat strict regulations and gun laws due to which crimes are controlled. Furthermore, it might also have to do with better mental health resources and ease of accessibility for suicide or mental health illnesses.

Limitations:

1. The FBI's National Instant Criminal Background Check System (NICS) counts initiated background checks, not the number of firearms sold. Therefore, a one-to-one correlation cannot be assumed.

References

```
# # Handling and Accessing Columns with Spaces in R
# # https://www.statology.org/r-read-csv-column-names-with-spaces/
#
# # Extracting Numbers from String in Base R
# # https://www.r-bloggers.com/2024/06/extracting-numbers-from-strings-in-r/

# Looping through multiple worksheets in excel in R
# https://stackoverflow.com/questions/45215910/how-to-loop-through-multiple-excel-sheets-and-append-sheetname-in-new-column

# Converting list of lists into a dataframe
# https://stackoverflow.com/questions/29674661/r-list-of-lists-to-data-frame

# Merging two dataframes
# https://stackoverflow.com/questions/8169323/r-concatenate-two-dataframes

# Removing first 5 rows in R Dataframe
# https://stackoverflow.com/questions/37770938/how-to-remove-first-n-rows-in-a-data-set-in-r

# Gun Deaths and Mortality Rate State
# https://www.cdc.gov/nchs/state-stats/deaths/firearms.html

# Gun Violence Deaths
# https://publichealth.jhu.edu/center-for-gun-violence-solutions/annual-gun-violence-data

# Gun Violence Deaths Dashboard
# https://giffords.org/lawcenter/resources/gun-violence-statistics/gun-violence-data-dashboards/

# Gun Violence By State
# https://worldpopulationreview.com/state-rankings/gun-violence-by-state
```

```
# Inspired by Data Analysis Formatting from this Medium article
# https://medium.com/@maikl.botros/chapter-iii-investigating-flight-delays-
and-cancellations-dataset-09e730bf6c55

# Research about Kentucky monthly background checks
# https://www.wsaz.com/content/news/Kentucky-performs-monthly-background-
checks-on-concealed-carry-permit-holders-361738231.html

# Explanation of why Kentucky is ranked 1st in FBI NICS Background Checks
# https://www.lpm.org/news/2015-12-07/kentuckys-background-checks-for-gun-
owners-stand-out

# Kentucky permit rechecks skewed by high volume of frequent rechecks
# https://whyy.org/articles/gun-background-checks-are-on-pace-to-break-
record-in-2019/

# Working with Dates in R lubridate
# https://stackoverflow.com/questions/67786097/ordering-dates-in-r-with-
lubridate

# US States abbreviation in R
# https://stackoverflow.com/questions/5411979/state-name-to-abbreviation

# Research Paper Concealed Carry Permit Holders Across the US: 2017
#
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3004915#:~:text=November%
202016%20election.-,%E2%96%A0,%E2%96%A0

# Adding annotation to your graph in R
# https://ggplot2.tidyverse.org/reference/annotate.html
# https://r-graph-gallery.com/233-add-annotations-on-ggplot2-chart.html
```