SESUG 2023 - Paper 147

# The Everytown Research database: Using SAS® analytic procedures to analyze mass shootings

Jay Iyengar, Data Systems Consultants LLC

## ABSTRACT

With mass shootings occurring every week, It can accurately be stated that mass shootings in the U.S. have reached the level of an epidemic. Everytown Research and Policy conducts independent methodically rigorous research and supports evidence-based policies to reduce the incidence of gun violence. In 2009, Everytown Research started assembling a Mass Shooting database which records key data on every mass shooting in the U.S. In this paper, I examine and explore the database using SAS® procedures to produce a series of tables, reports, graphics and visualizations. The goal of this project is to generate insights from the SAS analytics that guide the building of effective programs and policies to reduce the epidemic of mass shootings.

## INTRODUCTION

Everytown for Gun Safety is the largest Gun Violence Prevention organization in the United States. Everytown for Research & Policy is the research arm of Everytown for Gun Safety. Everytown for Research and Policy conducts independent research and rigorous analyses and supports evidence-based solutions to the Gun violence epidemic. The definition of a mass shooting can vary widely, with some definitions counting dozens of mass shootings a year, while others count multiple shootings per day. Everytown Research defines a mass shooting as an incident where 4 or more people are shot and killed or wounded. Based on this definition, Everytown alleges that 19,000 people have been shot and wounded or killed since 2015. The Everytown Research database on mass shootings consists of records of mass shootings from 2009 to the present using their definition of a mass shooting. The database gets updated periodically as new mass shooting incidents occur.

## THE DATABASE

The mass shooting database can be downloaded directly from the Everytown Research website. It's downloaded in CSV file format, which can be opened in Microsoft Excel. Before we analyze the data, it's necessary to extract the data, import it into SAS, and create SAS data sets. For this project, the SAS Studio interface was used, along with SAS On-demand for academics, a free non-commercial version of SAS.

In SAS Studio, I used PROC IMPORT to extract the data into SAS and create SAS datasets. The PROC IMPORT code is provided below in Figure 1. Running the PROC IMPORT code produces a SAS data set called EVERYTOWN.

```
FILENAME REFFILE 'MassShootingsData/everytownresearch_massshootings_data.xlsx';

PROC IMPORT DATAFILE=REFFILE
       DBMS=XLSX
       OUT=WORK.EveryTown;
       GETNAMES=YES;
RUN;
```

**Figure 1. PROC IMPORT code used to read CSV file into SAS data set**

## VARIABLES IN THE DATABASE

Once the data has been read from the CSV file and converted into a SAS data set, the next step is to run a PROC CONTENTS. PROC CONTENTS provides the number of variables and observations in your data set, along with a list of variables and variable attributes. The list of variables from the PROC CONTENTS output is displayed in Figure 2 below.

| | | | |
|---|---|---|---|
| **Data Set Name** | WORK.EVERYTOWN | **Observations** | 284 |
| **Member Type** | DATA | **Variables** | 14 |
| **Engine** | V9 | **Indexes** | 0 |
| **Created** | 08/16/2023 17:08:06 | **Observation Length** | 912 |

| Variable | Type | Len | Format | Informat | Label |
|---|---|---|---|---|---|
| AssWpn_HC_Mag | Char | 12 | $12. | $12. | AssWpn_HC_Mag |
| City | Char | 24 | $24. | $24. | City |
| Date | Num | 8 | MMDDYY10. | | Date |
| Family_Member_Fatality | Char | 12 | $12. | $12. | Family_Member_Fatality |
| Last updated | Char | 24 | $24. | $24. | Last updated |
| Latitude | Num | 8 | BEST. | | Latitude |
| Longitude | Num | 8 | BEST. | | Longitude |
| Narrative | Char | 764 | $764. | $764. | Narrative |
| Number_Fatalities | Num | 8 | BEST. | | Number_Fatalities |
| Number_Injured | Num | 8 | BEST. | | Number_Injured |
| Number_PeopleShot | Num | 8 | BEST. | | Number_PeopleShot |
| Restriction_GunOwnership | Char | 12 | $12. | $12. | Restriction_GunOwnership |
| State | Char | 2 | $2. | $2. | State |
| Warning_Signs | Char | 12 | $12. | $12. | Warning_Signs |

**Figure 2. PROC CONTENTS - listing of variables in the Everytown database**

From the PROC CONTENTS output, you'll notice the data set contains 14 variables, including the date the shooting occurred, the city and state where the shooting occurred, and quantitative variables on the number of people shot, the number injured, and the number of fatalities there were. Also included is Narrative, a free-format text variable, which lists details of the shooting incident.

Of the 14 variables in the dataset, 8 variables are character variables and 6 variables are numeric. Not all the variables in the data set are of interest from an analytic standpoint. Last updated is an administrative variable which defines the date the data set was last updated. Latitude and Longitude are of analytic interest to the extent we can plot and graph the coordinates and produce a GIS map of the shootings. Producing GIS Maps is outside the scope of this paper.

Removing the 3 extraneous variables leaves us with 11 variables in our analysis data set, 7 character variables and 4 numeric variables. The character variables are categorical variables with a discrete set of values. These variables will provide us with groups or subgroups for our analysis, with the exception of the Narrative variable. In providing a text description of the shooting incident, Narrative includes other important details of the shooting, such as the location and place where it occurred.

Location and place where the shooting occurred might be important factors which assist our understanding of the shootings. As an unformatted text variable, Narrative can be manipulated to extract information on the location and the type of facility, and store them in variables.

## EXAMINING AND VALIDATING THE DATA SET

Besides examining the descriptor portion of the data set using PROC CONTENTS, it's important to view the data portion of the SAS data set. PROC PRINT and PROC FREQ are both good constructs to examine the data values of the data set. PROC PRINT generates a record-level listing of the data set. Using the OBS= data set option, I selected the first 20 records of the data set and produced a detail report using PROC PRINT. The SAS Code and Output are provided in Figure 3 below.

```
Proc Print Data=EveryTown(Obs=20);
    Var Date State City Number_PeopleShot Number_Injured
        Number_Fatalities Family_Member_Fatality
        Restriction_GunOwnership Warning_Signs AssWpn_HC_Mag;
    Title1 'Mass Shootings - United States';
    Title2 '2009 to Present';
Run;
```

| Date | State | City | Shot | Injured | Fatalities | Family_Fatal | Restrictions | Signs | Assault_HC |
|---|---|---|---|---|---|---|---|---|---|
| 01/27/2009 | CA | Wilmington | 6 | 0 | 6 | Yes | No | No | No |
| 02/14/2009 | NY | Brockport | 5 | 1 | 4 | No | No | Yes | No |
| 03/05/2009 | OH | Cleveland | 6 | 1 | 5 | Yes | Yes | No | No |
| 03/10/2009 | AL | Coffee and Geneva County | 14 | 4 | 10 | Yes | No | No | Yes |
| 03/15/2009 | FL | Miami | 4 | 0 | 4 | Yes | No | No | Yes |
| 03/21/2009 | CA | East Oakland | 5 | 1 | 4 | No | Yes | Yes | Yes |
| 03/29/2009 | CA | Santa Clara | 6 | 1 | 5 | Yes | No | No | No |
| 03/29/2009 | NC | Carthage | 11 | 3 | 8 | No | No | Yes | No |
| 04/03/2009 | NY | Binghamton | 17 | 4 | 13 | No | No | Yes | Yes |
| 04/04/2009 | WA | Graham | 5 | 0 | 5 | Yes | No | Yes | No |
| 04/06/2009 | AL | Green Hill | 4 | 0 | 4 | Yes | No | Yes | No |
| 04/19/2009 | MD | Middletown | 4 | 0 | 4 | Yes | No | No | No |
| 06/22/2009 | KS | Kansas City | 4 | 0 | 4 | Yes | Yes | Yes | No |
| 08/27/2009 | GA | Lawrenceville | 5 | 1 | 4 | Yes | No | No | No |
| 11/01/2009 | NC | Mount Airy | 4 | 0 | 4 | No | Yes | Yes | Yes |
| 11/05/2009 | TX | Fort Hood | 45 | 32 | 13 | No | No | No | Yes |
| 11/09/2009 | OK | Oklahoma City | 4 | 0 | 4 | No | Yes | No | No |
| 11/12/2009 | AR | Pearcy | 5 | 0 | 5 | No | No | Yes | No |
| 11/26/2009 | FL | Jupiter | 6 | 2 | 4 | Yes | Yes | No | No |
| 11/28/2009 | KS | Osage | 4 | 0 | 4 | Yes | No | Yes | Yes |

**Figure 3. PROC PRINT code and output, EveryTown data set**

The PROC PRINT output was generated using the ODS EXCEL destination. A copy of the output was copied and pasted from the produced excel workbook. The variable names have been shortened from the SAS data set variable names to fit the output on the page. The output doesn't include the 3 variables dropped from the dataset (Latitude, Longitude, and Last_Updated) and doesn't include Narrative.

The data set includes four categorical variables which contain binary (YES/NO) values. Family_Fatal indicates whether a family member of the shooter was killed in the shooting. Restrictions defines if the shooter was legally barred from owning firearms. Signs documents if there were warning signs of the shooting which were present. Assault_HC indicates if there was an Assault Weapon and/or High-Capacity Magazine used in the shooting.

These additional variables provide useful subgroup variables in the analysis data set which might reveal significant insights in the data set.

## DATA MANIPULATION

Before we generate frequency tables on the variables, it's prudent to derive a few new variables which may provide us with analytic insights into mass shootings. Earlier, we discussed the variable Narrative, and how the shooting location can be extracted from this variable. In addition, in order to look at trends in mass shootings over time, its worthwhile to have variables such as Month and Year that the shooting took place. We have the date when the shooting occurred. We just need to extract the month using the MONTH() function, and the year using the YEAR() function.

We found a number of words present in the Narrative variable which describe the location or place of a mass shooting. However, in order to search a long string for a substring or a specific word, we need to use SAS character string functions, such as the FIND function or the INDEX function. Both the FIND() and the INDEX() function return a numeric value indicating the position of the word in the string, if the word is found in the substring. If the word is not found the functions return a 0.

In figure 4 below, is an excerpt from a SAS program which manipulates the EveryTown data set and uses the FIND function to search Narrative for specific words in an IF-THEN-ELSE series. The function takes a third argument, which is the 'i' modifier. The 'i' modifier is used to ignore case when searching for words.

```
Data ET.EveryTown;
      Length Location $25;
 Set EveryTown;

 Year_Shooting = Year(Date);
 Month_Shooting = Month(Date);

 If (FIND(Narrative, 'home', 'i')>0 and FIND(Narrative, 'nursing home', 'i')=0) or
     FIND(Narrative, 'house', 'i')>0 or FIND(Narrative, 'homes', 'i')>0 or
     FIND(Narrative, 'apartment', 'i')>0 or FIND(Narrative, 'residence', 'i')>0 or
     FIND(Narrative, 'residences', 'i')>0 or FIND(Narrative, 'trailer', 'i')>0 or
     FIND(Narrative, 'townhouse', 'i')>0 or FIND(Narrative, 'garage','i')>0 or
     FIND(Narrative, 'party','i')>0 or FIND(Narrative, 'Social Gathering','i')>0 or
     FIND(Narrative, 'backyard', 'i')>0 Then Location='Residence';
```

**Figure 4. DATA STEP excerpt using FIND and other functions**

We search for records where the FIND() function returns a value greater than 0 (>0), because we know this represents the position of the substring within the string. I could've also used the FINDW() function which searches for the presence of a word (not a substring) in a substring.

In the DATA STEP, I also use the YEAR() and MONTH() functions to create new variables for the year and month the shooting occurred, respectively. The program in Figure 4 is just an excerpt, as the entire program is provided in the appendix of this paper.

## FREQUENCY TABLES

Now that we have derived new variables Location and Month and Year of the shooting, we need to generate frequency tables on these and other variables in the data set for validation and to produce analytic insights. Usually we only produce frequency tables for character variables. To produce frequency tables we use PROC FREQ.

With the ORDER= option, there is the capability to display the frequencies in specific sorted order. In our code, we use the ORDER=FREQ option to display the results by descending order of frequency. In Figure 5, which is provided below, we include the PROC FREQ code we used to produce the frequency tables, including the ORDER=FREQ option.

```
Proc Freq Data = ET.EveryTown Order=Freq;
     Tables AssWpn_HC_Mag Warning_Signs
            Family_member_fatality Restriction_GunOwnership
            State Year_Shooting Month_Shooting
             Location / List Missing;
   Title1'One-Way Frequencies';
   Title2'Everytown Research database';
Run;
```

| State | Frequency | Percent |
|---|---|---|
| TX | 32 | 11.27 |
| CA | 30 | 10.56 |
| FL | 17 | 5.99 |
| IL | 15 | 5.28 |
| OH | 14 | 4.93 |
| WA | 10 | 3.52 |
| AZ | 9 | 3.17 |
| NC | 9 | 3.17 |
| GA | 8 | 2.82 |
| IN | 8 | 2.82 |

| Location | Frequency | Percent |
|---|---|---|
| Residence | 151 | 53.17 |
| Other Place | 56 | 19.72 |
| Professional Office | 27 | 9.51 |
| Unknown | 16 | 5.63 |
| School | 7 | 2.46 |
| Restaurant | 5 | 1.76 |
| Retail Store | 4 | 1.41 |
| Bar/Nightclub | 3 | 1.06 |
| Place of Worship | 3 | 1.06 |
| Beauty Salon | 2 | 0.70 |
| Government Facility | 2 | 0.70 |

**Figure 5. PROC FREQ code and output for State and Location variables**

In Figure 5, two Frequency Tables are listed below the PROC FREQ SAS code. Frequency tables for only 2 variables, State and Location, are provided. The frequencies for the remaining variables are provided in the appendix at the end of the paper. Further, only the first ten rows of the SAS output have been included in the tables in Figure 5 for the sake of conserving space.

In reviewing the output in Figure 5, there some interesting findings to point out. In the last 14 years, the states with the highest number of mass shootings were Texas, California, and Florida. Texas had the highest number respectively at 32. This finding makes sense, considering Texas has one of the highest levels of gun ownership amongst its population.

For the other variable, Location, the findings perhaps weren't quite as interesting. Locations with the highest number of mass shootings were Private Residences (151) and Other Places (56). It's worthwhile to note, however, that 9.5% of the mass shootings took place in professional offices, and 2.5% of mass shootings took place in schools. This percentage may be relatively small by itself. However, the numbers may be significant by looking at cross-tabulations with other variables.

## TRENDS IN THE DATA

In analyzing trends in mass shootings, it's prudent to use graphical visualizations to represent trends over time. In the data manipulation phase of this project, we created two new variables, Year and Month from the date the mass shooting occurred. Visualizations give us the ability to use different types of graphs and charts to depict trends, such as line graphs, bar charts, pie charts, etc.

In the previous section, we used PROC FREQ to generate frequency counts on the number of mass shootings by a number of variables in the EveryTown database. Many SAS analytic procedures have the capability to store results in an output SAS data set which contains summary data. With PROC FREQ, you can produce an output data set pertaining to each frequency table being produced. To perform this, you need to specify a data set using the OUT= option on each TABLES statement.

Figure 6 below shows the SAS code using the OUT=option on the TABLES statement for PROC FREQ.

```
Proc Freq Data = ET.EveryTown Noprint;
   Tables Year_Shooting / Out=Shootings_Year List Missing;
   Tables Month_Shooting / Out=Shootings_Month List Missing;
Run;
```

**Figure 6. PROC FREQ code for creating output data sets**

The code in Figure 6 creates 2 separate SAS data sets for year and month, respectively. Now the data is in a summary format where it can be plugged into other SAS procedures to produce visualizations.

In BASE SAS, ODS Graphics give us the capability of producing high-quality visualizations which depict analytic trends in events such as mass shootings. Using PROC SGPLOT, you can produce a variety of types of graphs and charts, such as line graphs and bar charts. Taking the output data sets from Figure 6, we then plugged them into PROC SGPLOT to produce a vertical line plot graphing the number of mass shootings by Year. In Figure 7 below is the SAS code and output.

```
Proc SGPlot Data=Shootings_Year;
   VLine Year_Shooting / Response=Count;
   YAxis Min=0 Max=100;
   Title1 'Number of Mass Shootings by Year';
Run;
```

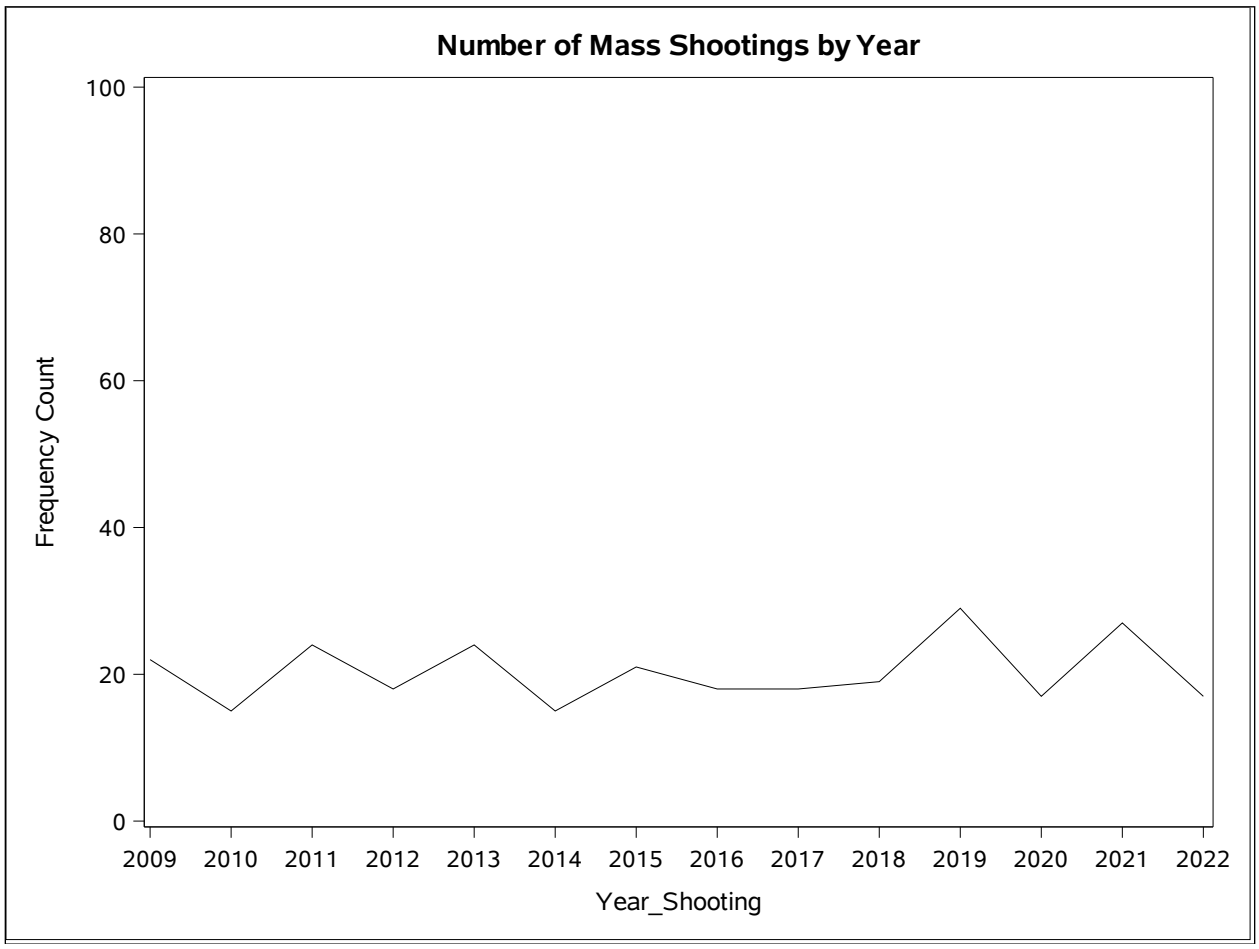



**Figure 7. PROC SGPLOT Output - Line graph of Mass Shootings by Year**

The vertical line graph in Figure 7 plots the number of mass shootings on the Y axis, and year the shooting occurred on the X axis.

From 2019-2022, the number of mass shootings in the United States stayed relatively constant overall. The graph does show a lot of fluctuation in the number of graphs from one year to the next. In some years, sharp increases in shootings, were followed by sharp decreases in shootings or decreases in shootings over the next 1-2 years.

The number of shooting incidents stayed within the range of between 18-30 events per year over this time frame. However, you'll notice that the 2 years with the highest number of shootings were in 2019, and 2021, respectively. The years with the 2 highest levels of shootings both were in the last 5 years.

## DESCRIPTIVE STATISTICS – NUMERIC VARIABLES

Besides a frequency count of the number of mass shootings, statistics on the number of fatalities, the number of people injured, and the number of people shot can provide useful metrics in measuring the extent of the mass shooting epidemic. Number of fatalities is specifically the number of people killed in the shooting incident.

BASE SAS includes many procedures to analyze numeric variables. These procedures include PROC MEANS, PROC SUMMARY and PROC UNIVARIATE as well as others. I selected PROC MEANS to run basic descriptive statistics on the number of fatalities, number injured, and number of people shot using the Everytown database. The SAS code and output are presented below in Figure 8.

```
Proc Means Data=ET.EveryTown N Mean Median Std Min Max Maxdec=1;
    Var Number_Fatalities
           Number_Injured
           Number_PeopleShot;
    Title1'Number of Deaths, Injuries, and Total Individuals Shot';
    Title2'Everytown Research database';
Run;
```

| Variable | Label | N | Mean | Median | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|---|
| Number_Fatalities | Number_Fatalities | 284 | 5.6 | 4.0 | 5.1 | 4.0 | 60.0 |
| Number_Injured | Number_Injured | 284 | 3.6 | 0 | 25.2 | 0 | 411.0 |
| Number_PeopleShot | Number_PeopleShot | 284 | 9.3 | 5.0 | 29.2 | 4.0 | 471.0 |

**Figure 8. Proc Means SAS code/output - Statistics on Fatalities, Injured and People Shot**

In the PROC MEANS statement, you list the specific statistics to be computed and displayed in the SAS Output. I selected N (Frequency count), Mean (Average), Median, Minimum and Maximum values, as well as the Standard Deviation.

The SAS Output in Figure 8 shows the average number of fatalities in a mass shooting was 5.6, while the range in the number of fatalities went from the minimum of 4.0 to a maximum of 60. The average number of people injured was 3.6, which ranged from a minimum of 0 to a maximum of 411 people. The average number of people shot in a mass shooting was 9.3, with a range going from a minimum of 4 to a maximum of 471 people shot.

Apart from the raw number of mass shooting incidents, the number of fatalities, people injured and people shot provides a more accurate depiction of the magnitude and impact of a mass shooting. For instance, the number of mass shooting may stay relatively constant from one year to the next. However, if the number of fatalities is increasing, then you can make an argument that mass shootings are actually getting worse.

## COMPLEX REPORTS AND VISUALIZATIONS

In Figure 5, earlier in the paper, we looked at a frequency distribution for location of the mass shooting. From the output, we could tell that shootings occurred in various locations in public, as well as in private residences. The public locations consist of Schools, Retail stores such as Walmart, Hospitals, Gas Stations, Grocery Stores, Churches and other places of worship, Bars and Nightclubs, Restaurants, Professional Offices, Beauty Salons, and Government agencies.

Taking a look at each of the specific locations individually doesn't draw out differences in mass shootings. However, if we we're to lump the specific categories into a general group of non-residential settings, this may reveal some patterns when compared to residential settings.

I decided to manipulate the data further, and compute a new variable, Location_group. Using The DATA STEP and IF-THEN-ELSE conditional logic, I created a new variable, Location_group, based on the variable Location. I coded all of the specific public locations and settings as 'Non-Residential'. Settings of 'Residence' were coded to 'Residential'.

In Figure 8 in the previous section, we computed descriptive statistics for the number of fatalities, people injured, and people shot. From an analytic standpoint, it might be worthwhile to look at these analytic variables within categories of Location_Group. Further, analyzing these variables within geographic levels, such as State, might yield state-to-state or even regional differences.

PROC REPORT is a powerful BASE SAS reporting construct for producing complex reports. It can construct reports of multiple dimensions, columnar reports, and detail as well as summary reports. I used PROC REPORT to generate a summary report of the analytic variables by State and Location_Group. The PROC REPORT code is listed in Figure 9 below.

```
Proc Report Data=EveryTown2 Headline Headskip;
    Column State Location_Group Number_Fatalities Number_Injured Number_PeopleShot;

    Define State / Group;
    Define Location_Group / Group;
    Define Number_Fatalities / Analysis Sum Order=Freq;
    Define Number_Injured / Analysis Sum Order=Freq;
    Define Number_PeopleShot / Analysis Sum Order=Freq;

    Compute After State;
        Line ' ';
    EndComp;

    Where State in('TX', 'CA', 'FL', 'IL', 'OH');

    Title1'Number of Fatalities, Injured, and People Shot';
    Title2'By State and Location Group';

    Footnote1'From 2009-2023';
Run;
```

**Figure 9. PROC REPORT code - Report by State and Location_Group**

In PROC REPORT, the variables that will be displayed in the report need to be included in the COLUMN statement. Thus, In the PROC REPORT code in Figure 9, there are 5 variables in the COLUMN statement. State and Location have been defined as Group variables in DEFINE statements.

The remaining variables; Number_Fatalities, Number_Injured, and Number_PeopleShot are defined as analysis variables. The SUM and ORDER=FREQ options have been specified in the DEFINE statements for each of these variables to indicate that the values for the variables should be summed within categories of State and Location. Also, the data in the output will be ordered according to the frequency count for State and Location.

I inserted a line break after each distinct value of State, using a COMPUTE BLOCK. You can also create line breaks using the BREAK AFTER statement, but not in SAS Studio which outputs in HTML. The output in the report is restricted to the 5 states with the highest numbers of shootings, Texas, California, Florida, Illinois, and Ohio. The report produced by the code is presented below in Figure 10.

| State | Location_Group | Number of Fatalities | Number of People Injured | Number of Shooting Victims |
|---|---|---|---|---|
| CA | Non-Residential | 81 | 13 | 94 |
| | Private Residence | 76 | 63 | 139 |
| | | | | |
| FL | Non-Residential | 98 | 83 | 181 |
| | Private Residence | 30 | 1 | 31 |
| | | | | |
| IL | Non-Residential | 33 | 60 | 93 |
| | Private Residence | 32 | 6 | 38 |
| | | | | |
| OH | Non-Residential | 33 | 19 | 52 |
| | Private Residence | 36 | 2 | 38 |
| | | | | |
| TX | Non-Residential | 126 | 149 | 275 |
| | Private Residence | 82 | 16 | 98 |
| | | | | |

**Figure 10. Number of fatalities, injured and People Shot by State and Location_Group**

By reviewing the output in Figure 10, some interesting patterns are revealed. In CA, IL, and OH, there was a relatively equal split in fatalities between Private Residential, and Non-Residential locations. The numbers injured and shot were substantially higher in non-residential settings in both Ohio and Illinois. However, in CA the opposite was true, with numbers substantially higher in private residential settings.

In Florida and Texas, the level of fatalities, injured, and people shot were dramatically higher in non-residential settings versus private residences. These discrepancies indicate that in these two states increased attention should be given to public safety and security in commercial and public buildings and facilities. There are other variables in the database which should be explored which can reveal insights into the trends of mass shootings in the US. Specifically, the use of an assault weapon or high-capacity magazine involved in the mass shooting.

From a rational perspective, the use of assault weapons carries with it the potential to inflict exponentially more fatalities, injuries, and shooting victims, then a regular handgun, because its an automatic or semi-automatic weapon and can carry many more rounds in one magazine. To document the impact that assault weapons really have, however, its necessary to examine empirical evidence through data.

The database contains the variable AssWpn_HC_Mag, which records if the mass shooting was perpetrated using an assault weapon or high-capacity magazine. Subsetting our data set where records have the value AssWpn_HC_Mag='Yes' and then summarizing the variables for fatalities, injured, and shooting victims would yield some interesting trends. To produce this analysis, I chose PROC SQL. In Figure 11 below is the PROC SQL code.

```
Proc Sql;
   Create Table Shoot_AW_Sum As
   Select Year_Shooting,
          Count(*) as Num_Shootings,
          Sum(Number_Fatalities) as Total_Fatalities,
          Sum(Number_Injured) as Total_Injured,
           Sum(Number_PeopleShot) as Total_Shot
   From ET.Everytown
   Where AssWpn_HC_Mag='Yes'
   Group By Year_Shooting;
Quit;
```

**Figure 11. PROC SQL code - analysis of shootings with assault weapons by year.**

For the analysis, I grouped the results by year of the shooting. Also, I included a count of the number of shootings as one of the variables, and created a new SAS data set, Shoot_AW_Sum. To depict trends in shootings with assault weapons, I used this summary data set to create line graphs and charts with PROC SGPLOT. The PROC SGPLOT code is presented in Figure 12 below.

```
Proc SGPlot Data=Shoot_AW_Sum;
   Series X=Year_Shooting Y=Num_Shootings / lineattrs=(pattern=solid);
   Series X=Year_Shooting Y=Total_Fatalities / lineattrs=(pattern=solid);
   Series X=Year_Shooting Y=Total_Injured / lineattrs=(pattern=solid);
   Series X=Year_Shooting Y=Total_Shot / lineattrs=(pattern=solid);

   YAxis Min=0 Max=200;
   Where 2020<=Year_Shooting<=2022;
   Title1 'Number of Shootings, Fatalities, Injured, and Shooting Victims';
   Title2 'In Shootings with Assault Weapons or High-Capacity Magazines';
   Title3 'By Year';
   Footnote1 'From 2020-2022';
Run;
```

**Figure 12. PROC SGPLOT Code to produce line graphs**

One of the capabilities of PROC SGPLOT is to graph multiple series and linear trends on the same axes. Using the SERIES statement, I graphed separate linear trends for number of shootings, number of fatalities, number of people injured and number of shooting victims. The graph produced by the PROC SGPLOT code is displayed in Figure 13 below.

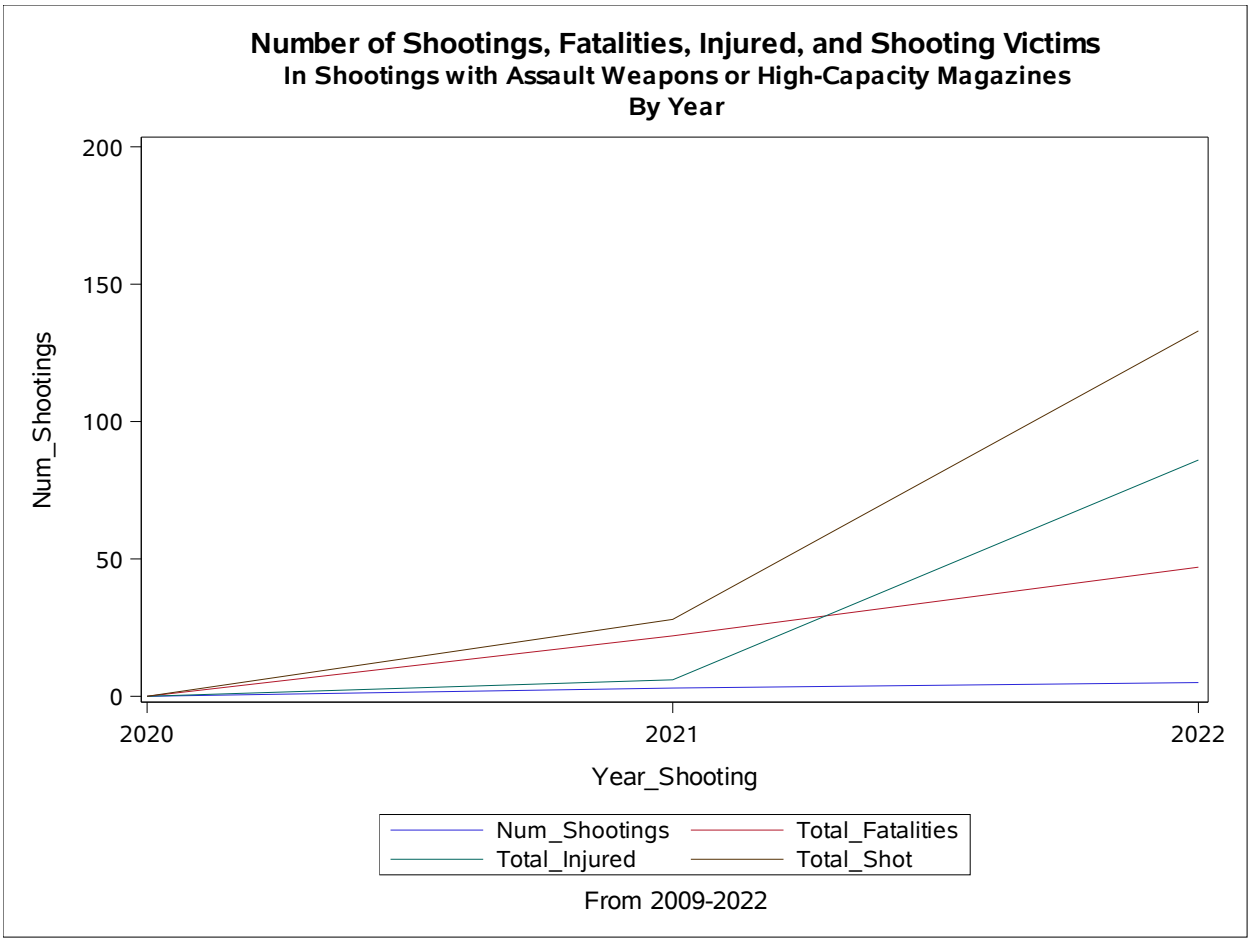



**Figure 13. Graph of Shootings/Fatalities/Injured/People Shot by Year for 2020-2022**

In order to make the trends in the data more visible, I limited the data in the graph to the last three years, 2020 to 2022. The data points mapped on the graph are limited to shooting incidents involving assault weapons or high-capacity magazines. I also included a trend line for the number of shootings.

The graphed lines clearly show increases in shooting victims (brown), people injured (green), and fatalities (red) over the last three years from 2020 to 2022. Shootings with assault weapons were at a low point in 2020 due to the pandemic, when there were restrictions in place regarding access to public and commercial building and facilities, such as retail stores, grocery stores, airports, churches, etc.

Both the number of injured in shootings, and the number of people shot show dramatic increases from 2021, where they were at a level of less than 25 for shooting victims, and under 10 for injured persons to 2022, where they rose to a level close to 150 for shooting victims, and a least 75 for persons injured in a shooting. In addition, an examination of the raw number of mass shootings (blue line) with an assault weapon shows an increase.

## CONCLUSION

EveryTown Research and Policy has compiled and released a database of mass shootings full of valuable information. Through BASE SAS and the DATA STEP, this database can be manipulated to produce an analysis data set. Using BASE SAS analytic and reporting procedures this data can be validated and analyzed to generate insights that address the problem of mass shootings and gun violence. With ODS graphics and PROC REPORT, visualizations and complex reports can be produced which depict the state of mass shootings. These reports highlight subgroups to focus on for policy and program efforts directed at reducing the level of mass shootings.

## REFERENCES

EveryTown for Gun Safety Support Fund, EveryTown Research and Policy. “Mass Shootings in the United States, an EveryTown for Gun Safety Support Fund Analysis. Mass Shootings with Four or more people Killed, Download the Data”, Updated March 2023. https://everytownresearch.org/mass-shootings-in-america/

## ACKNOWLEDGMENTS

The author would like to thank Kelly Smith, SESUG 2023 Operations Chair, Mel Alexander, SESUG 2023 Academic Chair, Abbas Tavakoli and Mostafa Zahed, Statistics, Analytics, and Reporting Section Co-Chairs, and the SESUG Executive Committee and Conference Team for accepting my abstract and paper and for organizing this great conference.

## CONTACT INFORMATION

Jay Iyengar, Director
Data Systems Consultants LLC
datasyscon@gmail.com
Linkedin: https://www.linkedin.com/in/datasysconsult/

Jay Iyengar is Director of Data Systems Consultants LLC. He is a SAS consultant, trainer, and SAS Certified Advanced Programmer. He’s been an invited speaker at several SAS user group conferences (WIILSU, WCSUG, SESUG) and has presented papers and training seminars at SAS Global Forum, Pharmaceutical SAS Users Group (PharmaSUG), and other regional and local SAS User Group conferences (MWSUG, NESUG, WUSS, MISUG). He was co-leader and organizer of the Chicago SAS Users Group (WCSUG) from 2015-19. His industry experience includes Health care, Public Health, International Trade, Marketing and Education.

## TRADEMARK CITATION

## Appendix I – SAS Code/Log

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
 NOTE: ODS statements in the SAS Studio environment may disable some output features.
 73
 74         FILENAME REFFILE '/home/iyenj/MassShootingsData/Everytown_091623.xlsx';
 75         LIBNAME ET '/home/iyenj/MassShootingsData';
 NOTE: Libref ET was successfully assigned as follows:
       Engine:        V9
       Physical Name: /home/iyenj/MassShootingsData
 76
 77         OPTIONS ORIENTATION=LANDSCAPE;
 78
 79         Proc Import Datafile=REFFILE
 80         DBMS=XLSX
 81         OUT=EveryTown
 82         Replace;
 83         GETNAMES=YES;
 84         Run;

 NOTE: One or more variables were converted because the data type is not supported by
       the V9 engine. For more details, run with options MSGLEVEL=I.

 NOTE: The import data set has 299 observations and 14 variables.
 NOTE: WORK.EVERYTOWN data set was successfully created.
 NOTE: PROCEDURE IMPORT used (Total process time):
       real time           0.05 seconds
       user cpu time       0.05 seconds
       system cpu time     0.00 seconds
       memory              3913.78k
       OS Memory           28156.00k
       Timestamp           09/20/2023 07:04:49 PM
       Step Count                        24  Switch Count  2
       Page Faults                       0
       Page Reclaims                     1181
       Page Swaps                        0
       Voluntary Context Switches        16
       Involuntary Context Switches      1
       Block Input Operations            120
       Block Output Operations           896


 85
 86         /* Contents of Mass Shooting Data set */
 87         Proc Contents Data=EveryTown;
 88         Run;

 NOTE: PROCEDURE CONTENTS used (Total process time):
       real time           0.06 seconds
       user cpu time       0.06 seconds
       system cpu time     0.00 seconds
       memory              3487.53k
       OS Memory           28204.00k
       Timestamp           09/20/2023 07:04:50 PM
       Step Count                        25  Switch Count  0
```

```
      Page Faults                       2
      Page Reclaims                     1216
      Page Swaps                        0
      Voluntary Context Switches        4
      Involuntary Context Switches      0
      Block Input Operations            240
      Block Output Operations           24


90         ODS Excel File = "/home/iyenj/MassShootingsData/EveryTown_Sample.xlsx";
91
92         /* Print out a Sample of the Shooting Records */
93         Proc Print Data = EveryTown (Obs=30);
94           Var Date State City Number_PeopleShot Number_Injured Number_Fatalities
                 Family_Member_Fatality Restriction_GunOwnership Warning_Signs
95               AssWpn_HC_Mag;
96           Title1 'Mass Shootings - United States';
97           Title2 '2009 to Present';
98         Run;

NOTE: There were 30 observations read from the data set WORK.EVERYTOWN.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.15 seconds
      user cpu time       0.16 seconds
      system cpu time     0.00 seconds
      memory              5247.71k
      OS Memory           40012.00k
      Timestamp           09/20/2023 07:04:50 PM
      Step Count                        26  Switch Count  2
      Page Faults                       0
      Page Reclaims                     1304
      Page Swaps                        0
      Voluntary Context Switches        21
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           40


99
100        ODS Excel Close;
NOTE: Writing EXCEL file: /home/iyenj/MassShootingsData/EveryTown_Sample.xlsx
101
102        /* Manipulate the Data set */
103    Data ET.EveryTown;
104     Length Location $25;
105        Set EveryTown;
106
107        Year_Shooting = Year(Date);
108        Month_Shooting = Month(Date);
109
110        If (FIND(Narrative, 'home', 'i')>0 and
111            FIND(Narrative, 'nursing home', 'i')=0) or
112            FIND(Narrative, 'house', 'i')>0 or
113            FIND(Narrative, 'homes', 'i')>0 or
114                FIND(Narrative, 'apartment', 'i')>0 or
```

```
115              FIND(Narrative, 'residence', 'i')>0 or
116              FIND(Narrative, 'residences', 'i')>0 or
117              FIND(Narrative, 'trailer', 'i')>0 or
118              FIND(Narrative, 'townhouse', 'i')>0 or
119              FIND(Narrative, 'garage','i')>0 or
120              FIND(Narrative, 'party','i')>0 or
121              FIND(Narrative, 'Social Gathering','i')>0 or
122              FIND(Narrative, 'backyard', 'i')>0 Then Location='Residence';
123
124    Else If FIND(Narrative, 'School', 'i')>0 or
125            FIND(Narrative, 'high school', 'i')>0 or
126            FIND(Narrative, 'College', 'i')>0 Then Location='School';
127
128  Else If FIND(Narrative, 'Association', 'i')>0 Or
129          FIND(Narrative, 'business', 'i')>0 Or
130          FIND(Narrative, 'newspaper', 'i')>0 Or
131          FIND(Narrative, 'office', 'i')>0 Or
132          FIND(Narrative, 'place of employment', 'i')>0 Or
133          FIND(Narrative, 'Co.', 'i')>0 Or
134          FIND(Narrative, 'Company', 'i')>0
                      Then Location = 'Professional Office';
135
136  Else If FIND(Narrative, 'Church', 'i')>0 Or
137          FIND(Narrative, 'synagogue', 'i')>0 Or
138          FIND(Narrative, 'temple', 'i')>0 Then Location='Place of Worship';
139
140  Else If FIND(Narrative, 'Restaurant', 'i')>0 Or
141          FIND(Narrative, 'coffee shop', 'i')>0 Then Location='Restaurant';
142
143  Else If FIND(Narrative, 'Municipal Center', 'i')>0 Or
144          FIND(Narrative, 'army Base', 'i')>0 Or
145          FIND(Narrative, 'Navy Yard', 'i')>0 Or
146          FIND(Narrative, 'Naval Reserve', 'i')>0
                               Then Location='Government Facility';
147
148  Else If FIND(Narrative, 'Hospital', 'i')>0 Or
149    FIND(Narrative, 'Nursing Home', 'i')>0 Then Location='Healthcare Facility';
150
151           Else If FIND(Narrative, 'Walgreens', 'i')>0 Or
152              FIND(Narrative, 'Walmart', 'i')>0 Or
153              FIND(Narrative, 'store', 'i')>0 Or
154              FIND(Narrative, 'Pharmacy', 'i')>0 Then Location='Retail Store';
155
156           Else If FIND(Narrative, 'Health Spa', 'i')>0 Or
157              FIND(Narrative, 'Hair Salon', 'i')>0 Or
158              FIND(Narrative, 'Barbershop', 'i')>0 Then Location='Beauty Salon';
159
160           Else If FIND(Narrative, 'bar', 'i')>0 Or
161              FIND(Narrative, 'nightclub', 'i')>0 Then Location='Bar/Nightclub';
162
163           Else If FIND(Narrative, 'Supermarket', 'i')>0 Or
164              FIND(Narrative, 'Market', 'i')>0 Then Location='Grocery Store';
165
166    Else If FIND(Narrative, 'gas station', 'i')>0 Then Location='Gas station';
167
```

```
168   Else If FIND(Narrative, 'movie theatre', 'i')>0 Then Location='Cinema';
169
170   Else If FIND(Narrative, 'Music Festival', 'i')>0 Then Location='Concert';
171
172   Else If FIND(Narrative, 'Airport', 'i')>0 Then Location='Airport';
173
174   Else If FIND(Narrative, 'Indian Reservation', 'i')>0 Then Location='Indian
      Reservation';
175
176   Else If FIND(Narrative, 'Bank', 'i')>0 Then Location='Bank';
177
178   Else If FIND(Narrative, 'Shopping Mall', 'i')>0 Then Location='Shopping Mall';
179
180   Else If FIND(Narrative, 'X', 'i')>0 Then Location='Unknown';
181
182   Else Location='Other Place';
183
184   Label AssWpn_HC_Mag = 'Assualt Weapon or High-Capacity Magazine'
185         City = 'City where shooting occurred'
186         Location='Place where shooting incident occurred'
187         Date = 'Date of Shooting'
188         Family_Member_Fatality = 'Fatalities are Family Member of the Shooter'
189         Last_updated = 'Shooting Record Last Updated'
190         Latitude = 'Geographic Latitude of Shooting'
191         Longitude = 'Geographic Longitude of Shooting'
192         Narrative = 'Descriptive details of Shooting'
193         Number_Fatalities = 'Number of People Killed (Fatalities)'
194         Number_Injured = 'Number of People Injured'
195         Number_PeopleShot = 'Number of Shooting Victims'
196         Restriction_GunOwnership = 'Shooter Restricted from Gun Ownership'
197         State = 'State where Shooting Occurred'
198         Warning_Signs = 'Warning Signs of Shooting\Shooter Present';
199   Run;

NOTE: There were 299 observations read from the data set WORK.EVERYTOWN.
NOTE: The data set ET.EVERYTOWN has 299 observations and 17 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              1677.25k
      OS Memory           32812.00k
      Timestamp           09/20/2023 07:04:50 PM
      Step Count                        27  Switch Count  1
      Page Faults                       0
      Page Reclaims                     293
      Page Swaps                        0
      Voluntary Context Switches        33
      Involuntary Context Switches      0
      Block Input Operations            32
      Block Output Operations           784
```

```
201         /* One-Way Frequencies - Character Variables */
202         Proc Freq Data = ET.EveryTown Order=Freq;
203            Tables Location / List Missing;
204            Title1'One-Way Frequencies';
205            Title2'Everytown Research database';
206         Run;

NOTE: There were 299 observations read from the data set ET.EVERYTOWN.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.03 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1172.28k
      OS Memory           32812.00k
      Timestamp           09/20/2023 07:04:50 PM
      Step Count                        28  Switch Count  4
      Page Faults                       0
      Page Reclaims                     302
      Page Swaps                        0
      Voluntary Context Switches        44
      Involuntary Context Switches      0
      Block Input Operations            800
      Block Output Operations           264

208         /* One-Way Frequencies - Character Variables */
209         Proc Freq Data = ET.EveryTown Order=Freq;
210            Tables AssWpn_HC_Mag
211                   Warning_Signs
212                   Family_member_fatality
213                   Restriction_GunOwnership
214                   State
215                   Year_Shooting
216                   Month_Shooting / List Missing;
217
218            Title1'One-Way Frequencies';
219            Title2'Everytown Research database';
220         Run;

NOTE: There were 299 observations read from the data set ET.EVERYTOWN.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.10 seconds
      user cpu time       0.10 seconds
      system cpu time     0.00 seconds
      memory              1308.28k
      OS Memory           32812.00k
      Timestamp           09/20/2023 07:04:50 PM
      Step Count                        29  Switch Count  4
      Page Faults                       0
      Page Reclaims                     184
      Page Swaps                        0
      Voluntary Context Switches        35
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           352
```

```
223        /** Number of Mass Shootings by Year and Month **/
224        /** Frequency Tables and Graphs/Trend Lines   **/
225        Proc Freq Data = ET.EveryTown Noprint;
226           Tables Year_Shooting / Out=Shootings_Year List Missing;
227           Tables Month_Shooting / Out=Shootings_Month List Missing;
228        Run;

NOTE: There were 299 observations read from the data set ET.EVERYTOWN.
NOTE: The data set WORK.SHOOTINGS_YEAR has 15 observations and 3 variables.
NOTE: The data set WORK.SHOOTINGS_MONTH has 12 observations and 3 variables.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              1733.81k
      OS Memory           33332.00k
      Timestamp           09/20/2023 07:04:50 PM
      Step Count                        30  Switch Count  6
      Page Faults                       0
      Page Reclaims                     291
      Page Swaps                        0
      Voluntary Context Switches        45
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           792

230        Proc SGPlot Data=Shootings_Year;
231           VLine Year_Shooting / Response=Count;
232           YAxis Min=0 Max=100;
233           Title1 'Number of Mass Shootings by Year';
234        Run;

NOTE: Since no format is assigned, the numeric category variable will use the default of
BEST6.
NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           2.43 seconds
      user cpu time       0.09 seconds
      system cpu time     0.02 seconds
      memory              19635.90k
      OS Memory           48808.00k
      Timestamp           09/20/2023 07:04:52 PM
      Step Count                        31  Switch Count  3
      Page Faults                       0
      Page Reclaims                     6205
      Page Swaps                        0
      Voluntary Context Switches        432
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           824

NOTE: There were 15 observations read from the data set WORK.SHOOTINGS_YEAR.
```

```
236         Proc SGPlot Data=Shootings_Month;
237            VLine Month_Shooting / Response=Count;
238            YAxis Min=0 Max=100;
239            Title1 'Number of Mass Shootings by Month';
240         Run;

NOTE: Since no format is assigned, the numeric category variable will use the default of
BEST6.

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.24 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              2831.37k
      OS Memory           50860.00k
      Timestamp           09/20/2023 07:04:53 PM
      Step Count                        32  Switch Count  3
      Page Faults                       0
      Page Reclaims                     933
      Page Swaps                        0
      Voluntary Context Switches        256
      Involuntary Context Switches      5
      Block Input Operations            0
      Block Output Operations           512

NOTE: There were 12 observations read from the data set WORK.SHOOTINGS_MONTH.

241
242          /* Descriptive Statistics - Numeric Variables */
243         Proc Means Data=ET.EveryTown MAXDEC=2 N Mean Median Std Min Max;
244            Var Number_Fatalities
245                Number_Injured
246                Number_PeopleShot;
247
248            Title1'Number of Deaths, Injuries, and Total Individuals Shot';
249            Title2'Everytown Research database';
250         Run;

NOTE: There were 299 observations read from the data set ET.EVERYTOWN.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              7222.06k
      OS Memory           55612.00k
      Timestamp           09/20/2023 07:04:53 PM
      Step Count                        33  Switch Count  3
      Page Faults                       0
      Page Reclaims                     1592
      Page Swaps                        0
      Voluntary Context Switches        53
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8
```

```
253         Title2;
254         Title3;
255
256
257         /** Total Number of Fatalities, Injuries and Shootings by Year **/
258         /** Descriptive Stats and Graphs/Trend Lines      **/
259         Proc Sql;
260           Create Table Total_Fatal_Inj_Shot_Yr As
261            Select Year_Shooting,
262                   Sum(Number_Fatalities) as Total_Fatal,
263                   Sum(Number_Injured) as Total_Injured,
264                   Sum(Number_PeopleShot) as Total_Shot
265           From ET.EveryTown
266           Group By Year_Shooting;
NOTE: Table WORK.TOTAL_FATAL_INJ_SHOT_YR created, with 15 rows and 4 columns.

267         Quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              5909.53k
      OS Memory           55084.00k
      Timestamp           09/20/2023 07:04:53 PM
      Step Count                        34  Switch Count  2
      Page Faults                       0
      Page Reclaims                     293
      Page Swaps                        0
      Voluntary Context Switches        22
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272

269  Proc SGPlot Data=Total_Fatal_Inj_Shot_Yr;
270       Series X=Year_Shooting Y=Total_Fatal;
271       Series X=Year_Shooting Y=Total_Injured;
272       Series X=Year_Shooting Y=Total_Shot;
273
274       YAxis Min=0 Max=1000;
275
276  Title1 'Total Number of Fatalities, Injured and People Shot in Mass Shootings';
277  Title2 'By Year';
278  Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.29 seconds
      user cpu time       0.06 seconds
      system cpu time     0.01 seconds
      memory              3304.53k
      OS Memory           51884.00k
      Timestamp           09/20/2023 07:04:53 PM
      Step Count                        35  Switch Count  3
      Page Faults                       1
      Page Reclaims                     753
```

```
       Page Swaps                         0
       Voluntary Context Switches         285
       Involuntary Context Switches       1
       Block Input Operations             160
       Block Output Operations            680

NOTE: There were 15 observations read from the data set WORK.TOTAL_FATAL_INJ_SHOT_YR.

 279
 280         Proc SGPlot Data=Total_Fatal_Inj_Shot_Yr;
 281            VLine Year_Shooting / Response=Total_Fatal;
 282            YAxis Min=0 Max=200;
 283            Title1 'Total Number of Fatalities in Mass Shootings';
 284            Title2 'By Year';
 285         Run;

NOTE: Since no format is assigned, the numeric category variable will use the default of
BEST6.
NOTE: PROCEDURE SGPLOT used (Total process time):
       real time           0.19 seconds
       user cpu time       0.05 seconds
       system cpu time     0.01 seconds
       memory              3359.56k
       OS Memory           52140.00k
       Timestamp           09/20/2023 07:04:53 PM
       Step Count                         36  Switch Count  4
       Page Faults                        0
       Page Reclaims                      838
       Page Swaps                         0
       Voluntary Context Switches         258
       Involuntary Context Switches       1
       Block Input Operations             0
       Block Output Operations            552

NOTE: There were 15 observations read from the data set WORK.TOTAL_FATAL_INJ_SHOT_YR.

 286
 287         Proc SGPlot Data=Total_Fatal_Inj_Shot_Yr;
 288            VLine Year_Shooting / Response=Total_Injured;
 289            YAxis Min=0 Max=500;
 290            Title1 'Total Number of Injuries in Mass Shootings';
 291            Title2 'By Year';
 292         Run;

NOTE: Since no format is assigned, the numeric category variable will use the default of
BEST6.
NOTE: PROCEDURE SGPLOT used (Total process time):
       real time           0.17 seconds
       user cpu time       0.06 seconds
       system cpu time     0.01 seconds
       memory              3197.09k
       OS Memory           52140.00k
       Timestamp           09/20/2023 07:04:53 PM
       Step Count                         37  Switch Count  4
       Page Faults                        0
```

```
      Page Reclaims                     696
      Page Swaps                        0
      Voluntary Context Switches        260
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           568

NOTE: There were 15 observations read from the data set WORK.TOTAL_FATAL_INJ_SHOT_YR.

 293
 294        Proc SGPlot Data=Total_Fatal_Inj_Shot_Yr;
 295           VLine Year_Shooting / Response=Total_Shot;
 296           YAxis Min=0 Max=650;
 297           Title1 'Total Number Shot in Mass Shootings';
 298           Title2 'By Year';
 299        Run;

NOTE: Since no format is assigned, the numeric category variable will use the default of
BEST6.
NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.19 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              3193.46k
      OS Memory           52140.00k
      Timestamp           09/20/2023 07:04:54 PM
      Step Count                        38  Switch Count  4
      Page Faults                       0
      Page Reclaims                     687
      Page Swaps                        0
      Voluntary Context Switches        267
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           544

NOTE: There were 15 observations read from the data set WORK.TOTAL_FATAL_INJ_SHOT_YR.

 300
 301        /** Average Number of Fatalities, Injuries and Shootings by Year **/
 302        /** Descriptive Stats and Graphs/Trend Lines    **/
 303        Proc Means Data=ET.EveryTown Mean Maxdec=2 Noprint;
 304            Var Number_Fatalities Number_Injured Number_PeopleShot;
 305        Class Year_Shooting;
 306        Output Out=Fatal_Inj_Shot_Year
 307           Mean(Number_Fatalities) = Avg_Fatal
 308           Mean(Number_Injured) = Avg_Injured
 309           Mean(Number_PeopleShot) = Avg_Shot;
 310        Run;

 NOTE: There were 299 observations read from the data set ET.EVERYTOWN.
 NOTE: The data set WORK.FATAL_INJ_SHOT_YEAR has 16 observations and 6 variables.
 NOTE: PROCEDURE MEANS used (Total process time):
       real time           0.00 seconds
       user cpu time       0.00 seconds
       system cpu time     0.01 seconds
```

```
      memory              8231.57k
      OS Memory           57408.00k
      Timestamp           09/20/2023 07:04:54 PM
      Step Count                        39  Switch Count  3
      Page Faults                       0
      Page Reclaims                     1632
      Page Swaps                        0
      Voluntary Context Switches        27
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


311
312        Proc SGPlot Data=Fatal_Inj_Shot_Year;
313           VLine Year_Shooting / Response=Avg_Fatal;
314           YAxis Min=0 Max=30;
315           Title1 'Average Number of Fatalities per Mass Shooting';
316           Title2 'By Year';
317        Run;

 NOTE: Since no format is assigned, the numeric category variable will use the default
of BEST6.
 NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.22 seconds
      user cpu time       0.08 seconds
      system cpu time     0.01 seconds
      memory              3196.75k
      OS Memory           51884.00k
      Timestamp           09/20/2023 07:04:54 PM
      Step Count                        40  Switch Count  4
      Page Faults                       0
      Page Reclaims                     724
      Page Swaps                        0
      Voluntary Context Switches        267
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           544

 NOTE: There were 16 observations read from the data set WORK.FATAL_INJ_SHOT_YEAR.

318
319        Proc SGPlot Data=Fatal_Inj_Shot_Year;
320           VLine Year_Shooting / Response=Avg_Injured;
321           YAxis Min=0 Max=30;
322           Title1 'Average Number of Injuries per Mass Shooting';
323           Title2 'By Year';
324        Run;

 NOTE: Since no format is assigned, the numeric category variable will use the default
of BEST6.
 NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.19 seconds
      user cpu time       0.08 seconds
      system cpu time     0.01 seconds
```

```
      memory              3187.78k
      OS Memory           51884.00k
      Timestamp           09/20/2023 07:04:54 PM
      Step Count                        41  Switch Count  4
      Page Faults                       0
      Page Reclaims                     686
      Page Swaps                        0
      Voluntary Context Switches        264
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           632

 NOTE: There were 16 observations read from the data set WORK.FATAL_INJ_SHOT_YEAR.

 325
 326        Proc SGPlot Data=Fatal_Inj_Shot_Year;
 327           VLine Year_Shooting / Response=Avg_Shot;
 328           YAxis Min=0 Max=50;
 329           Title1 'Average Number Shot per Mass Shooting';
 330           Title2 'By Year';
 331        Run;

 NOTE: Since no format is assigned, the numeric category variable will use the default
of BEST6.
 NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.21 seconds
      user cpu time       0.08 seconds
      system cpu time     0.00 seconds
      memory              3215.46k
      OS Memory           51884.00k
      Timestamp           09/20/2023 07:04:54 PM
      Step Count                        42  Switch Count  4
      Page Faults                       0
      Page Reclaims                     686
      Page Swaps                        0
      Voluntary Context Switches        267
      Involuntary Context Switches      17
      Block Input Operations            0
      Block Output Operations           552

 NOTE: There were 16 observations read from the data set WORK.FATAL_INJ_SHOT_YEAR.
```

## Appendix II – SAS Output

| Assualt Weapon or High-Capacity Magazine | | | | |
|---|---|---|---|---|
| AssWpn_HC_Mag | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| No | 190 | 66.90 | 190 | 66.90 |
| Under review | 44 | 15.49 | 234 | 82.39 |
| Yes | 50 | 17.61 | 284 | 100.00 |

| Warning Signs of Shooting\Shooter Present | | | | |
|---|---|---|---|---|
| Warning_Signs | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| No | 81 | 28.52 | 81 | 28.52 |
| Under review | 100 | 35.21 | 181 | 63.73 |
| Yes | 103 | 36.27 | 284 | 100.00 |

| Fatalities are Family Member of the Shooter | | | | |
|---|---|---|---|---|
| Family_Member_Fatality | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| No | 114 | 40.14 | 114 | 40.14 |
| Under review | 44 | 15.49 | 158 | 55.63 |
| Yes | 126 | 44.37 | 284 | 100.00 |

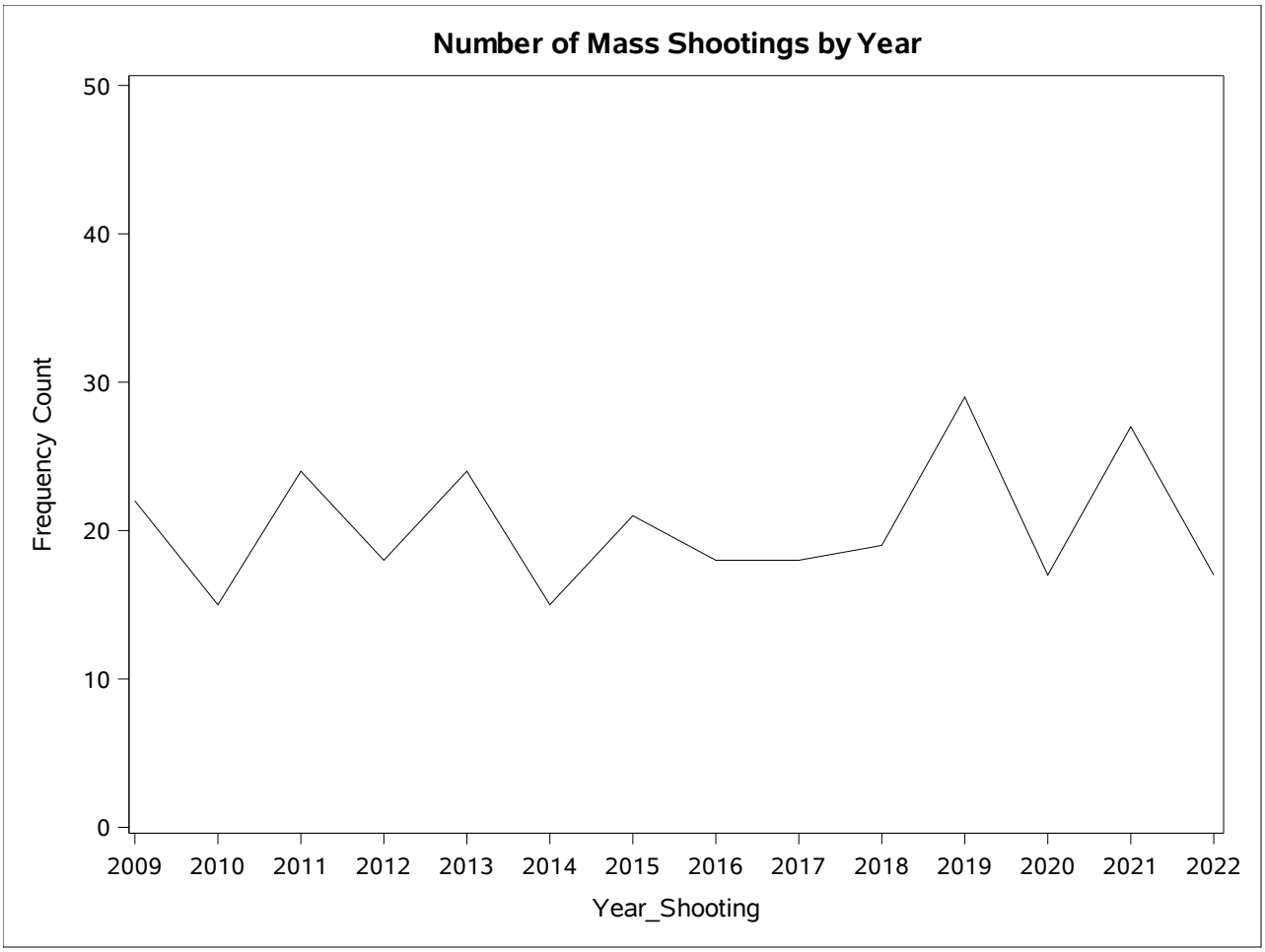
| Shooter Restricted from Gun Ownership | | | | |
|---|---|---|---|---|
| Restriction_GunOwnership | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| No | 166 | 58.45 | 166 | 58.45 |
| Under review | 44 | 15.49 | 210 | 73.94 |
| Yes | 74 | 26.06 | 284 | 100.00 |

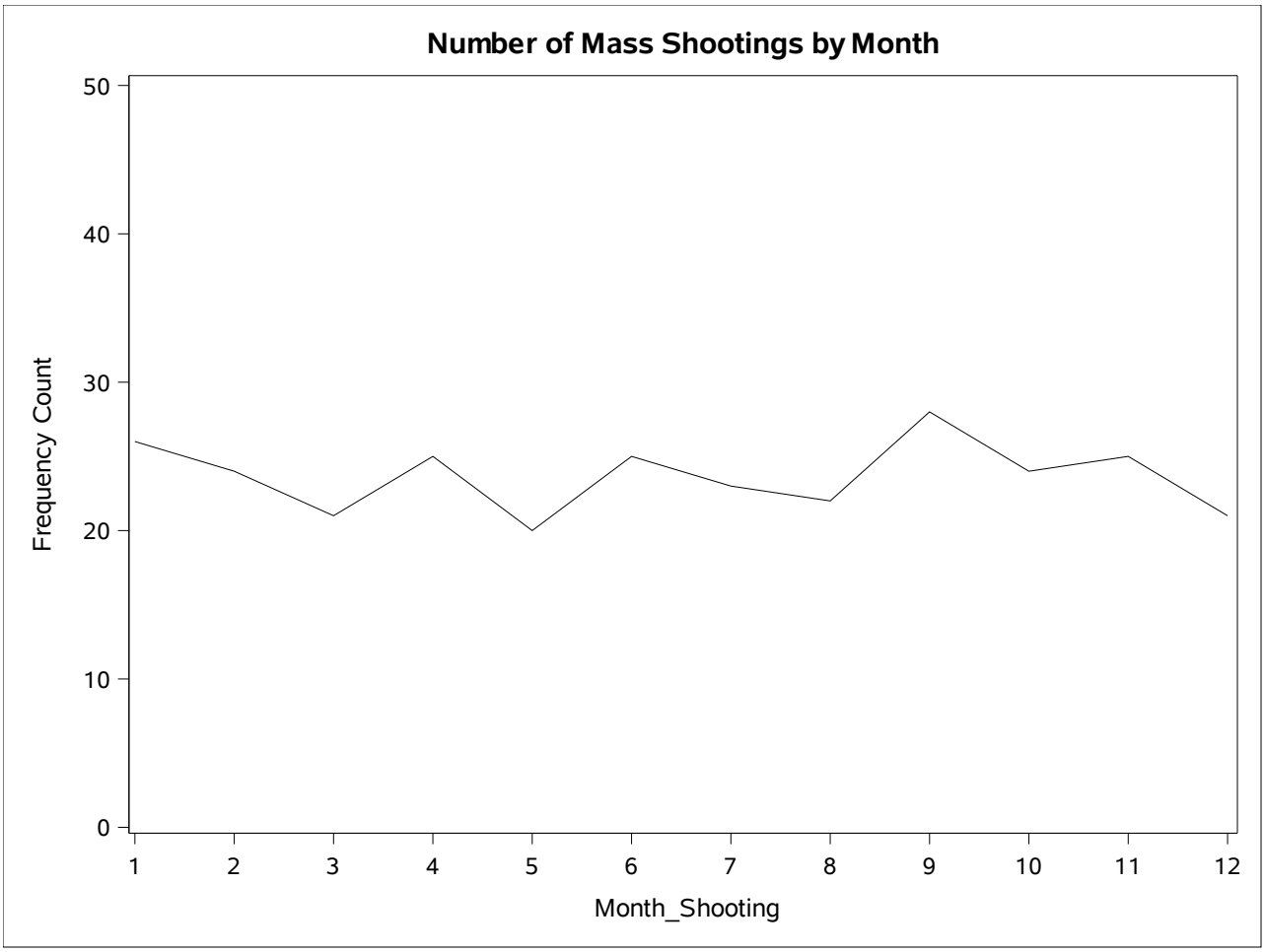
| State where Shooting Occurred | | | | |
|---|---|---|---|---|
| State | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| AK | 1 | 0.35 | 1 | 0.35 |
| AL | 5 | 1.76 | 6 | 2.11 |
| AR | 2 | 0.70 | 8 | 2.82 |
| AZ | 9 | 3.17 | 17 | 5.99 |
| CA | 30 | 10.56 | 47 | 16.55 |
| CO | 4 | 1.41 | 51 | 17.96 |
| CT | 2 | 0.70 | 53 | 18.66 |
| DC | 2 | 0.70 | 55 | 19.37 |
| DE | 1 | 0.35 | 56 | 19.72 |
| FL | 17 | 5.99 | 73 | 25.70 |
| GA | 8 | 2.82 | 81 | 28.52 |
| ID | 1 | 0.35 | 82 | 28.87 |
| IL | 15 | 5.28 | 97 | 34.15 |
| IN | 8 | 2.82 | 105 | 36.97 |
| KS | 6 | 2.11 | 111 | 39.08 |
| KY | 2 | 0.70 | 113 | 39.79 |
| LA | 3 | 1.06 | 116 | 40.85 |
| MA | 2 | 0.70 | 118 | 41.55 |
| MD | 5 | 1.76 | 123 | 43.31 |
| ME | 1 | 0.35 | 124 | 43.66 |
| MI | 8 | 2.82 | 132 | 46.48 |
| MN | 4 | 1.41 | 136 | 47.89 |
| MO | 7 | 2.46 | 143 | 50.35 |
| MS | 4 | 1.41 | 147 | 51.76 |
| MT | 1 | 0.35 | 148 | 52.11 |
| NC | 9 | 3.17 | 157 | 55.28 |
| ND | 2 | 0.70 | 159 | 55.99 |
| NJ | 3 | 1.06 | 162 | 57.04 |
| NM | 4 | 1.41 | 166 | 58.45 |
| NV | 4 | 1.41 | 170 | 59.86 |
| NY | 6 | 2.11 | 176 | 61.97 |

| State where Shooting Occurred | | | | |
|---|---|---|---|---|
| State | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| OH | 14 | 4.93 | 190 | 66.90 |
| OK | 6 | 2.11 | 196 | 69.01 |
| OR | 1 | 0.35 | 197 | 69.37 |
| PA | 8 | 2.82 | 205 | 72.18 |
| SC | 8 | 2.82 | 213 | 75.00 |
| SD | 1 | 0.35 | 214 | 75.35 |
| TN | 5 | 1.76 | 219 | 77.11 |
| TX | 32 | 11.27 | 251 | 88.38 |
| UT | 2 | 0.70 | 253 | 89.08 |
| VA | 7 | 2.46 | 260 | 91.55 |
| VT | 1 | 0.35 | 261 | 91.90 |
| WA | 10 | 3.52 | 271 | 95.42 |
| WI | 7 | 2.46 | 278 | 97.89 |
| WV | 5 | 1.76 | 283 | 99.65 |
| WY | 1 | 0.35 | 284 | 100.00 |

| Year_Shooting | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| **2009** | 22 | 7.75 | 22 | 7.75 |
| **2010** | 15 | 5.28 | 37 | 13.03 |
| **2011** | 24 | 8.45 | 61 | 21.48 |
| **2012** | 18 | 6.34 | 79 | 27.82 |
| **2013** | 24 | 8.45 | 103 | 36.27 |
| **2014** | 15 | 5.28 | 118 | 41.55 |
| **2015** | 21 | 7.39 | 139 | 48.94 |
| **2016** | 18 | 6.34 | 157 | 55.28 |
| **2017** | 18 | 6.34 | 175 | 61.62 |
| **2018** | 19 | 6.69 | 194 | 68.31 |
| **2019** | 29 | 10.21 | 223 | 78.52 |
| **2020** | 17 | 5.99 | 240 | 84.51 |
| **2021** | 27 | 9.51 | 267 | 94.01 |
| **2022** | 17 | 5.99 | 284 | 100.00 |

| Month_Shooting | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| **1** | 26 | 9.15 | 26 | 9.15 |
| **2** | 24 | 8.45 | 50 | 17.61 |
| **3** | 21 | 7.39 | 71 | 25.00 |
| **4** | 25 | 8.80 | 96 | 33.80 |
| **5** | 20 | 7.04 | 116 | 40.85 |
| **6** | 25 | 8.80 | 141 | 49.65 |
| **7** | 23 | 8.10 | 164 | 57.75 |
| **8** | 22 | 7.75 | 186 | 65.49 |
| **9** | 28 | 9.86 | 214 | 75.35 |
| **10** | 24 | 8.45 | 238 | 83.80 |
| **11** | 25 | 8.80 | 263 | 92.61 |
| **12** | 21 | 7.39 | 284 | 100.00 |


Number of Mass Shootings by Year
Frequency Count
50
40
30
20
10
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Year_Shooting


Number of Mass Shootings by Month
Frequency Count
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
Month_Shooting