THE EVERYTOWN RESEARCH DATABASE: USING SAS® ANALYTIC PROCEDURES TO ANALYZE MASS SHOOTINGS

Jay Iyengar, Data Systems Consultants LLC

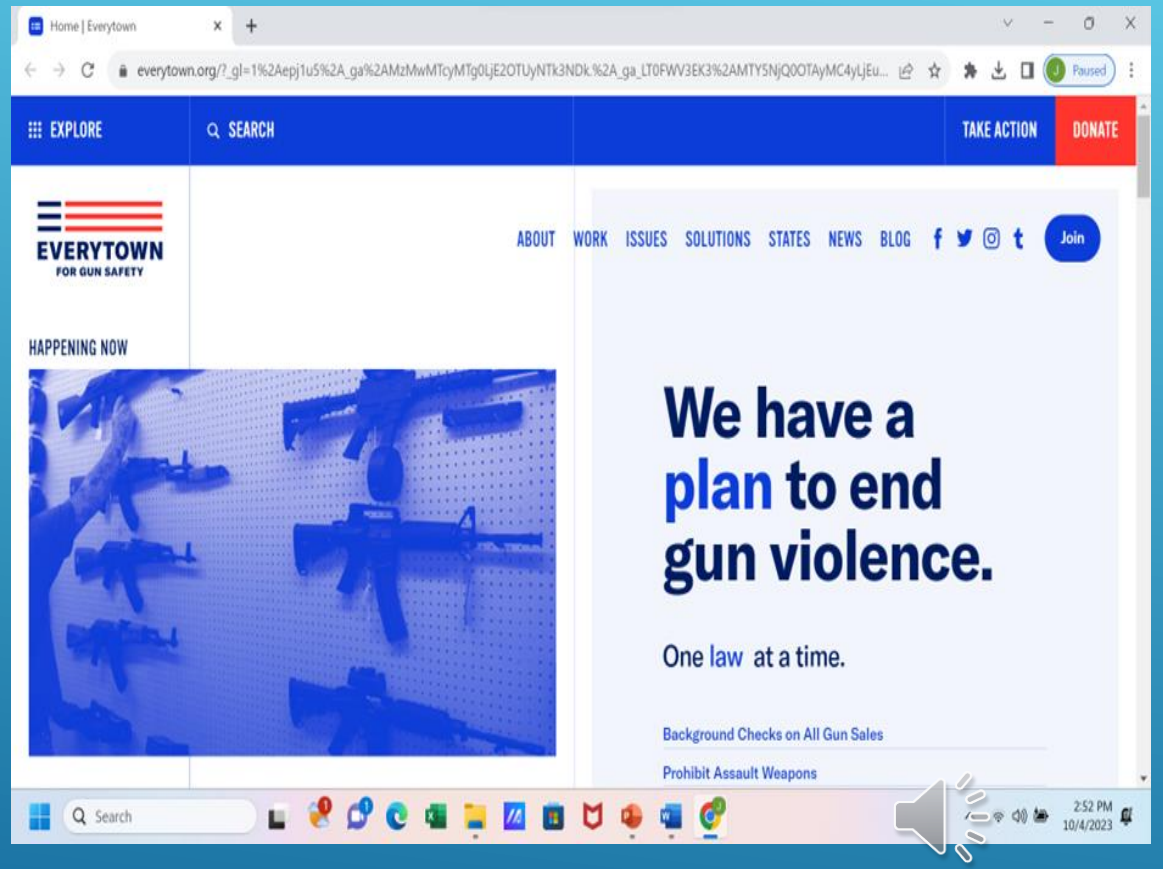
Michigan SAS Users Group – February 22, 2024

INTRODUCTION

Mass Shootings
- Have become a frequent occurrence in US culture.

Mass Shooting data
- Analyze as a data for good or pro-bono project.

Purpose of the Project
- To generate insights which can inform and drive policies and programs directed at reducing levels of gun violence.



INTRODUCTION (CON'T)

Everytown for Gun Safety
- Everytown Research and Policy (conducts empirical research)

Everytown Research and Policy
- conducts quantitative research on gun violence
- releases a database of mass shootings in the United States from 2009-present.

- the database is updated periodically.

Mass shooting
- An shooting incident where 4 or more people are shot and killed or wounded.

IMPORTING THE DATA INTO SAS

Accessing the data

- The database can be downloaded directly from the Everytown website as a CSV file.

File Management

- Minor edits and changes to the file in Excel. Save the file as an XLSX file

SAS Platform and Interface

- SAS On-Demand for Academics and SAS Studio were used as the SAS platform and interface

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


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

CONTENTS OF THE DATA BASE

14 Variables in the database

8 Character variables – State, City, Assault Weapon/HC Magazine, Family Member, Restrictions on Gun Ownership, Warning Signs, Narrative

6 Numeric variables – Date, Number of Fatalities, Number of Injured, Number of People Shot

284 Observations

- 1 Record per Mass shooting

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

PRODUCING AN ANALYSIS DATA SET

Analysis data set
- Contains variables of analytic Interest.

Last Updated
- Administrative variable. Not of analytic Interest

Longitude and Latitude
- Coordinates used for GIS mapping – Outside the scope of the presentation

The analysis data set consists of 11 variables – 7 character, and 4 numeric

Character variables
- Categorical variables which can be used as group or subgroup variables.
- Explanatory variables which provide insights into the factors which drive mass shootings

Narrative
- an unformatted text field which can be manipulated to extract useful information
- Contains location and place where the shooting occurred
  - Place of business, church, school, etc.

VALIDATING THE DATA SET

Examine the SAS data set
-the data portion of the data set.

PROC PRINT
- effective BASE SAS construct to produce data listings and to validate a data set.

```
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;
```

View only a sample of the data set, rather than the entire set of observations.

Limit the number of records using the OBS= option.

SAMPLE LISTING OF THE DATA

Use ODS EXCEL to send output to EXCEL spreadsheet
– Output copied from Excel spreadsheet

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

MANIPULATING THE DATA SET

```
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';
```

Narrative

-Not in a useful format that can be analyzed

Place and location of the shooting

- can be extracted from Narrative using the FIND function

Year and month of the shooting

- can be extracted from the date field.

FREQUENCY TABLES

PROC FREQ
- valid BASE SAS construct for examining data values and validating data sets.

Using PROC FREQ output you can answer exploratory questions about mass shootings.
- How many shootings involved an assault weapon or high-capacity magazine?
- What percentage of shootings were there warning signs present of the shooters' intent or plans?
- Which states had the highest levels of mass shootings?

```
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;
```

PROC FREQ OUTPUT

ORDER=FREQ enables descending order of frequency

Texas
- Had the highest number of shootings.
- Accounted for 11.3% of all mass shootings.

Residential locations
- Accounted for 53 % of all mass shootings

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

GRAPHICS AND VISUALIZATIONS

Graphics and visualizations

- Useful formats and visual aids in understanding the scope of the mass shooting epidemic.

Graphs and Charts

-Trends over time

First Step

- Generate one-way frequencies for year and month of shooting.
- Store the frequency tables in SAS data sets using OUT=option of 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;
```

Second Step

- Produce line graphs using PROC SGLPLOT
  - Part of the ODS GRAPHICS toolset in BASE SAS.
  - Extensive graphics and visualization capabilities.

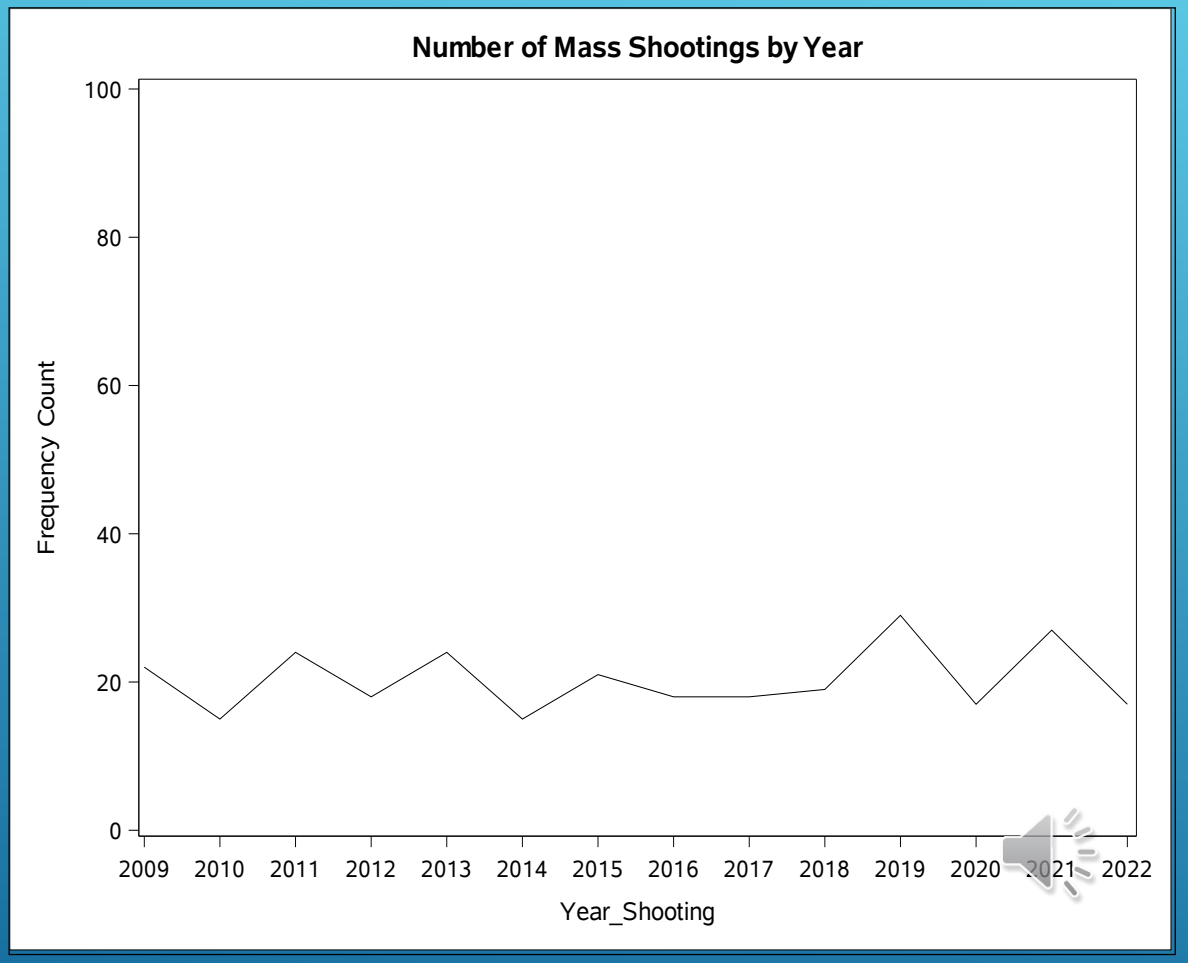
PROC SGPLOT CODE AND OUTPUT

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

The Number of Shootings has fluctuated over time, sharp increases in one year followed by sharp decreases in the next

2020 experienced a sharp decline which is largely due to the Covid-19 pandemic.

2019 and 2021 both had higher levels than other years.



DESCRIPTIVE STATISTICS

Is the number of mass shootings the most useful metric to evaluate the extent of the mass shooting epidemic?

Analyze the number of fatalities (people killed), number injured (people injured), and shooting victims (number of people shot).

- More useful metrics because these variables document the impact of a mass shooting on a population.

BASE SAS Procedures which generate descriptive statistics

- PROC MEANS
- PROC SUMMARY
- PROC UNIVARIATE

PROC MEANS and PROC SUMMARY are very similar.

PROC MEANS OUTPUT

```
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

The Number of Fatalities ranged from 4 to 60

The Number Injured ranged from 0 to 411

The Number of People Shot ranged from 4 to 471.

COMPLEX REPORTS

Computed Location_Group variable based on Location.

- Collapsed specific location values into Location group of ‘Residential’ or ‘Non-Residential’.

```
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;
```

Produce summary report

- Group by State and Location_Group within State.
- Analyze number of fatalities, injured, shooting victims
- ORDER=Freq
- Compute block adds line breaks

PROC REPORT OUTPUT

Limited to states with 5 highest levels of mass shootings

Florida and Texas
- large discrepancies in fatalities between shootings in residential vs. non-residential settings.

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

ADVANCED GRAPHICS

Limit the data to shootings involving Assault Weapons and High-Capacity Magazines

Re-produce linear trends subsetting by year to 2020-2022

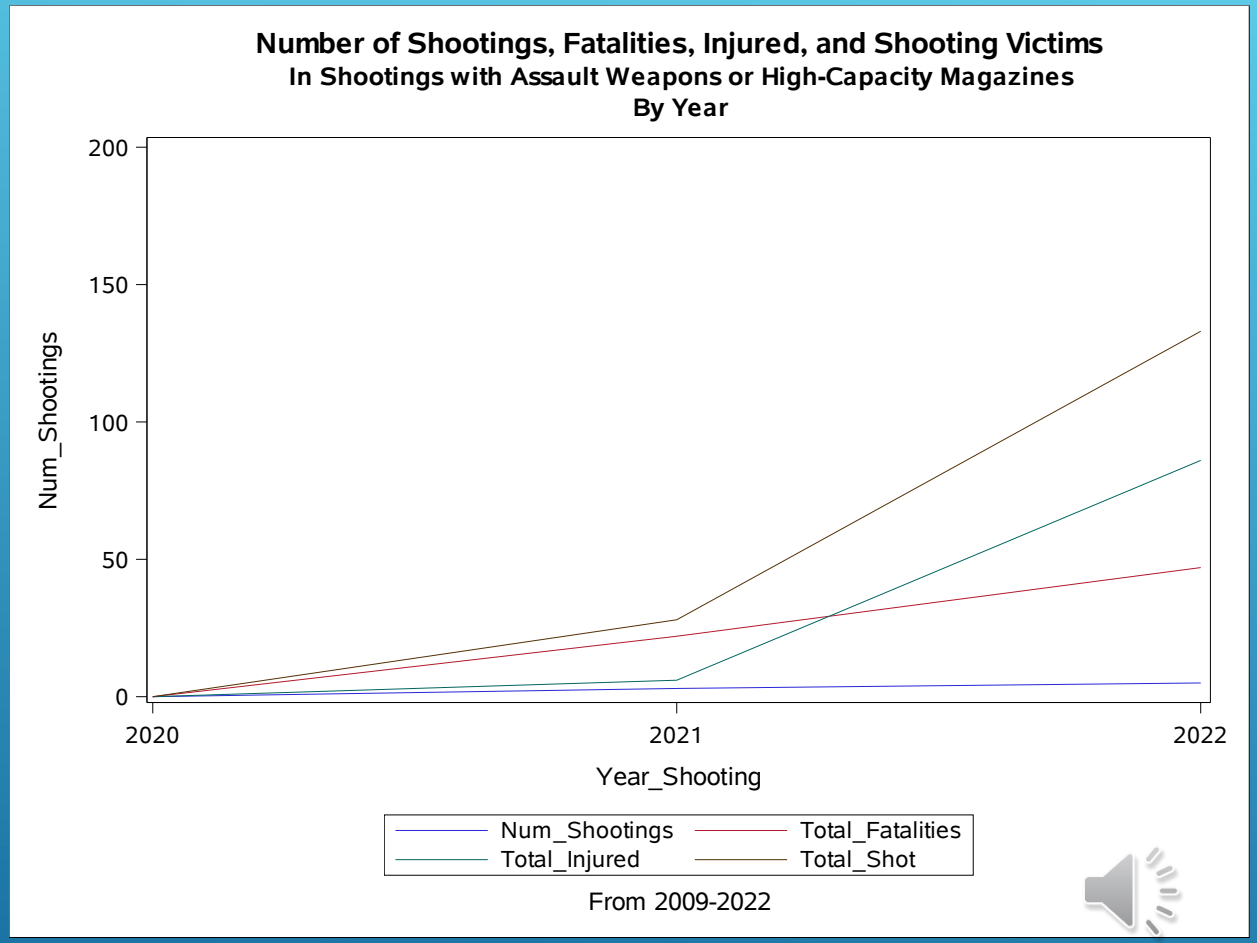
```
Proc SGPlot Data=Shoot_AW_Sum;
   Series X=Year_Shooting Y=Num_Shootings / lineattrs=(pattern=solid);
   Series X=Year_Shooting Y=Total_Fatalities / lineattrs=(pattern=solid);
   Series X=Year_Shooting Y=Total_Injured / lineattrs=(pattern=solid);
   Series X=Year_Shooting Y=Total_Shot / lineattrs=(pattern=solid);

   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;
```

GRAPHICS OUTPUT

Produce separate trend lines using SERIES statement in PROC SGPLOT.

Separate trend lines for
- Number of Fatalities (Red)
- Number Injured (Green)
- Number People Shot (Brown)
- Number of Shootings (Blue)

Increases in all 3 metrics for shootings involving assault weapons or high-capacity magazines.



CONCLUSION

SAS has a multitide of powerful analytic tools and constructs which can be used to display data and show trends, and assess the magnitude of specific events, such as mass shootings.

Using BASE SAS analytic and data validation procedures

- -PROC FREQ
- -PROC MEANS
- -PROC REPORT
- -PROC SGPLOT

you can generate reports and visualizations which can provide insights into the mass shooting epidemic.

CONTACT INFORMATION

Feel free to contact me with any questions you have about my talk.

Jay Iyengar, Director

Data Systems Consultants LLC

Email: datasyscon@gmail.com

https://www.linkedin.com/in/datasysconsult/