US Mass Shootings
(1966-2017)

Paul Szkilnyk
Jeremy Potts
Mass Shooting Dataset

Kaggle dataset summary: Mass Shootings in the United States of America (1966-2017) The US has witnessed 398 mass shootings in last 50 years that resulted in 1,996 deaths and 2,488 injured. The latest and the worst mass shooting of October 2, 2017 killed 58 and injured 515 so far. The number of people injured in this attack is more than the number of people injured in all mass shootings of 2015 and 2016 combined. The average number of mass shootings per year is 7 for the last 50 years that would claim 39 lives and 48 injured per year.
Mass Shooting Dataset

```
mydata = read.csv("shoot.csv", header=TRUE, sep="", na.strings=c("", "NA"))
summary(mydata)
library(plyr)
library(ggplot2)
library(ggmap)
library(maps)
library(mapdata)
```
<table>
<thead>
<tr>
<th>State</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>37</td>
<td>25</td>
<td>21</td>
<td>16</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>FL</td>
<td>25</td>
<td>1</td>
<td>21</td>
<td>16</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>TX</td>
<td>21</td>
<td>2.50</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
<tr>
<td>GA</td>
<td>16</td>
<td>1</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
<tr>
<td>WA</td>
<td>16</td>
<td>1</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
<tr>
<td>AZ</td>
<td>12</td>
<td>1</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>37</td>
<td>25</td>
<td>21</td>
<td>16</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>FL</td>
<td>25</td>
<td>1</td>
<td>21</td>
<td>16</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>TX</td>
<td>21</td>
<td>2.50</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
<tr>
<td>GA</td>
<td>16</td>
<td>1</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
<tr>
<td>WA</td>
<td>16</td>
<td>1</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
<tr>
<td>AZ</td>
<td>12</td>
<td>1</td>
<td>5.00</td>
<td>5.762</td>
<td>9.000</td>
<td>12.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Causes</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>psycho</td>
<td>68</td>
<td>65</td>
<td>44</td>
<td>18</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>terrorism</td>
<td>65</td>
<td>44</td>
<td>18</td>
<td>13</td>
<td>38</td>
<td>77</td>
</tr>
<tr>
<td>anger</td>
<td>44</td>
<td>18</td>
<td>13</td>
<td>38</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>frustration</td>
<td>18</td>
<td>13</td>
<td>38</td>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>domestic dispute</td>
<td>13</td>
<td>38</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Other)</td>
<td>38</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA's</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fatalities</td>
<td>Injured</td>
<td>Total Victims</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>0.000</td>
<td>0.000</td>
<td>3.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Qu.</td>
<td>1.000</td>
<td>1.000</td>
<td>4.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3.000</td>
<td>3.000</td>
<td>5.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.437</td>
<td>6.176</td>
<td>10.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>5.500</td>
<td>5.000</td>
<td>9.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>59.000</td>
<td>527.000</td>
<td>585.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Frequency of Shootings by Year

```r
year = mydata$Year
year.freq = table(year)
pie(year.freq)
barplot(year.freq, las=2)
```
Frequency of Shootings by Year
Frequency of Shootings by Year
Frequency of Shootings by Year

\[ yr = \text{count}(\text{mydata}, \ 'year') \]
\[ yr1 = yr[\text{order}(-yr$\text{year}),] \]
\[ \text{mysum1} = \text{cumsum}(yr1$freq) \]
\[ \text{mysum1}[6] \]
\[ \text{mysum1}[42] \]
\[ \text{mysum1}[6] \div \text{mysum1}[42] \]

\[ > \ yr1 \]
\[ \begin{array}{ccc}
1 & 1966 & 2 \\
2 & 1971 & 1 \\
3 & 1972 & 1 \\
4 & 1974 & 2 \\
5 & 1976 & 2 \\
6 & 1979 & 2 \\
\end{array} \]

\[ > \text{mysum1}[6] \]
\[ [1] \ 191 \]

\[ > \text{mysum1}[42] \]
\[ [1] \ 323 \]

\[ > \text{mysum1}[6] \div \text{mysum1}[42] \]
\[ [1] \ 0.5913313 \]
Frequency of Shootings by Year

```r
> table(year)
year
  2   1   1   2   2   2   2   2   3   2   3   1   6

> yr2 = year[order(-year$year),]
Error in year$year : $ operator is invalid for atomic vectors

yr = count(mydata, 'year')

> yr

<table>
<thead>
<tr>
<th>year</th>
<th>freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1966</td>
<td>16</td>
</tr>
<tr>
<td>1971</td>
<td>14</td>
</tr>
<tr>
<td>1972</td>
<td>67</td>
</tr>
<tr>
<td>1974</td>
<td>69</td>
</tr>
<tr>
<td>1976</td>
<td>10</td>
</tr>
</tbody>
</table>

> class(year)
[1] "integer"

> class(yr)
[1] "data.frame"
```
Frequency of Shootings by Month

```r
month = mydata$Month
month.freq = table(month)
pie(month.freq)
barplot(month.freq)
```
Frequency of Shootings by Month
Frequency of Shootings by Month
Monthly Temperature relative to Incident

```r
month1 = count(mydata, 'month')
avgt = read.csv("temp.csv", header=TRUE, sep="","")
avgt1 = merge(avgt, month1)
plot(avgt1$month, avgt1$AvgTemp, type = "o")
plot(avgt1$AvgTemp, avgt1$freq)
abline(lm(avgt1$freq ~ avgt1$AvgTemp))
summary(lm(avgt1$freq ~ avgt1$AvgTemp))
```
Monthly Temperature
Monthly Temperature relative to Incident
Monthly Temperature relative to Incident

```r
> summary(lm(avg1$freq ~ avg1$AvgTemp))

Call:
  lm(formula = avg1$freq ~ avg1$AvgTemp)

Residuals:
          Min     1Q Median     3Q    Max
-9.710 -5.289  -1.379  5.344  17.097

Coefficients:
                          Estimate  Std. Error   t value  Pr(>|t|)
(Intercept)              53.9532     9.3386    5.777 0.000178 ***
avg1$AvgTemp            -0.5113      0.1703   -3.003 0.013281 *

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.583 on 10 degrees of freedom
Multiple R-squared: 0.4741,  Adjusted R-squared: 0.4216
F-statistic: 9.016 on 1 and 10 DF,  p-value: 0.01328
```
Frequency of Shootings by State

```r
state = mydata$State
state.freq = table(state)
pie(state.freq)
barplot(state.freq, las=2)
```
Frequency of Shootings by State

<table>
<thead>
<tr>
<th>State</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>37</td>
</tr>
<tr>
<td>FL</td>
<td>25</td>
</tr>
<tr>
<td>TX</td>
<td>21</td>
</tr>
<tr>
<td>GA</td>
<td>16</td>
</tr>
<tr>
<td>WA</td>
<td>16</td>
</tr>
<tr>
<td>AZ</td>
<td>12</td>
</tr>
</tbody>
</table>
Frequency of Shootings by State

- CA: 37
- FL: 25
- TX: 21
- GA: 16
- WA: 16
- AZ: 12
### Frequency of Shootings by State

```r
state = mydata$State
state1 = count(mydata, 'state')
stt = read.csv("states.csv", header=TRUE, sep="","
stt1 = stt$state
stt2 = count(stt,'state')
stt3 = merge(state1,stt2, by="state", all=TRUE)

> sttt3

<table>
<thead>
<tr>
<th>state</th>
<th>freq.x</th>
<th>freq.y</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AL</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>AR</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>AZ</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>CA</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>WV</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WY</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ND</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>NH</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>RI</td>
<td>NA</td>
<td>1</td>
</tr>
</tbody>
</table>

48 Levels:
```
State Population relative to Incident

```r
stt = read.csv("states.csv", header=TRUE, sep="",""
state1 = count(mydata, 'state')
pops = merge(state1, sttt, all=TRUE)
pops1 = pops[order(pops$pop),]
plot(pops$pop, pops$freq)
abline(lm(pops$freq ~ pops$pop))
summary(lm(pops$freq ~ pops$pop))
> pops1

<table>
<thead>
<tr>
<th>state</th>
<th>freq</th>
<th>pop</th>
<th>SD</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WY</td>
<td>1</td>
<td>585501</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>VT</td>
<td>1</td>
<td>624594</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>DC</td>
<td>1</td>
<td>681170</td>
<td>51</td>
<td>2</td>
</tr>
<tr>
<td>AK</td>
<td>1</td>
<td>741894</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>ND</td>
<td>NA</td>
<td>757952</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>DE</td>
<td>1</td>
<td>952065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>2</td>
<td>1042520</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI</td>
<td>NA</td>
<td>1056426</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>1</td>
<td>1331479</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>NA</td>
<td>1334795</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
State Population relative to Incident
State Population relative to Incident

> summary(lm(pops$freq ~ pops$pop))

Call:
  lm(formula = pops$freq ~ pops$pop)

Residuals:
     Min      1Q  Median      3Q     Max
-7.0753 -1.4384 -0.5172  1.2484  8.7314

Coefficients:
                           Estimate Std. Error t value   Pr(>|t|)    
(Intercept)              9.462e-01  5.948e-01   1.591   0.119     
pops$pop                 8.675e-07  6.032e-08  14.382  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.037 on 46 degrees of freedom
(3 observations deleted due to missingness)
Multiple R-squared:  0.8181, Adjusted R-squared:  0.8141
F-statistic: 206.8 on 1 and 46 DF,  p-value: < 2.2e-16
Frequency of Shootings by Cause

cause = mydata$Cause
cause.freq = table(cause)
pie(cause.freq)
barplot(cause.freq)
Frequency of Shootings by Cause

- Psycho: 68
- Terrorism: 65
- Anger: 44
- Frustration: 18
- Domestic dispute: 13
- Unknown: 38
- Unemployment: 77
- Racism: 5
- Religious radicalism: 2
- Revenge: 0
- Suspension: 0
- Terrorism: 0

(fractional data not shown)
Frequency of Shootings by Cause

- psycho: 68
- terrorism: 65
- anger: 44
- frustration: 18
- domestic dispute: 13
- (Other): 38
- NA's: 77
plot(mydata$Year, mydata$TotalVictims)
abline(lm(mydata$TotalVictims ~ mydata$Year))

Victims per Year
Fatalities per Year

```r
plot(mydata$Year, mydata$Fatalities)
abline(lm(mydata$Fatalities ~ mydata$Year))
```
Plotting Incidents on a Map

```r
states <- map_data("state")
ggplot() +
geom_polygon(data=states,
aes(x = long, y = lat,
fill = region, group = group,
color = "white") +
coord_fixed(1.3) +
guides(fill=FALSE) +
geom_point(data=mydata,
aes(x=Longitude, y=Latitude),
color="red") +
coord_cartesian(xlim = c(-125, -65),
ylim = c(25, 50))
```
Plotting Incidents on a Map
Plotting Incidents on a Map