U.S. Mass Shootings From 1966 - 2017

Deekshitha Balaji

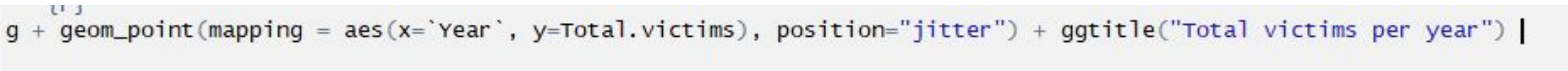
Introduction

- Since 1966, 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. 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 claiming 39 fatalities and 48 injured per year.

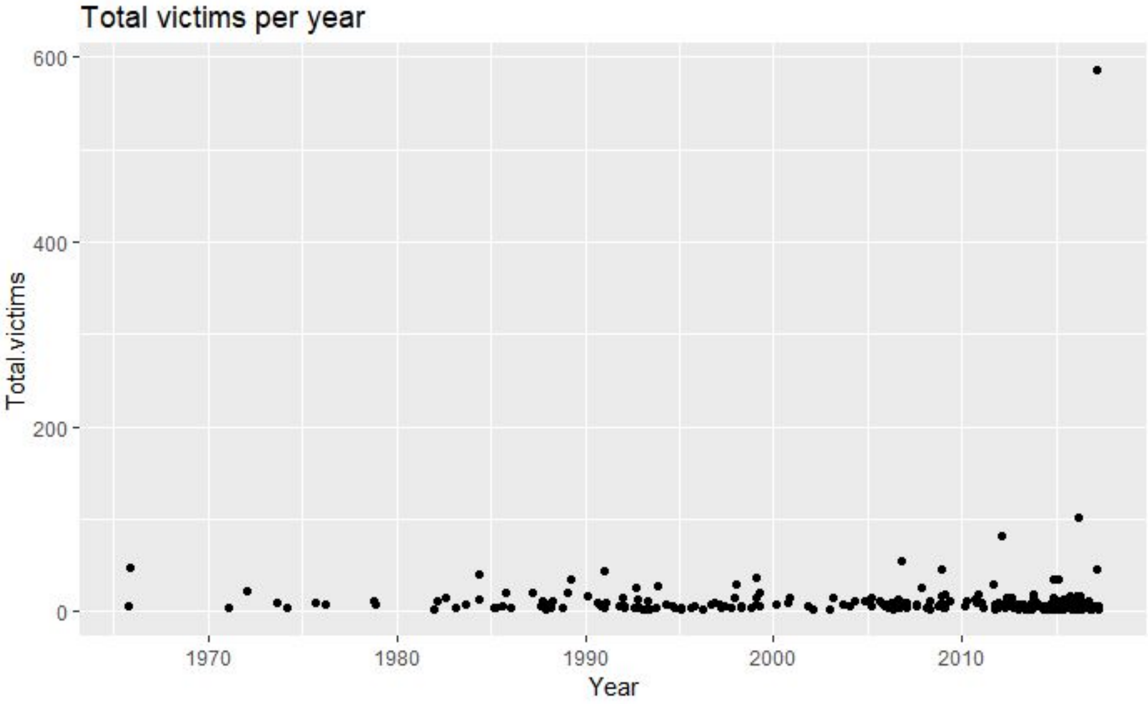
- The data set[1] has 22 variables in total, including information about the total number of victims, mental health issues, age of the shooter, target, the cause, and age of the shooter. It is important to note that since our data comes from such a wide range of years, there is a high chance of incorrect data and/or many outliers.

- We will examine these data and answer the following questions.
  - How have the occurrence of mass shootings increased over these fifty years?
  - Have the total victims (fatalities and injured) increased as well?
  - Which states saw the most mass shootings?
  - What were the reasons for all of these mass shootings in the first place?
  - How much did mental health play a role in these mass shootings?

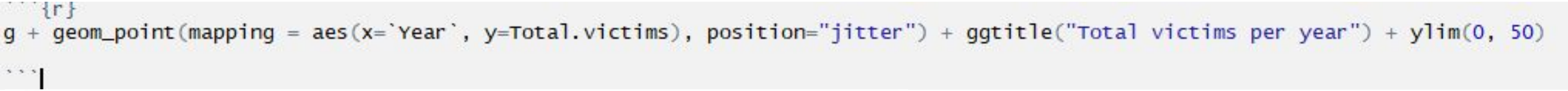
How have the total number of mass shooting victims increased?



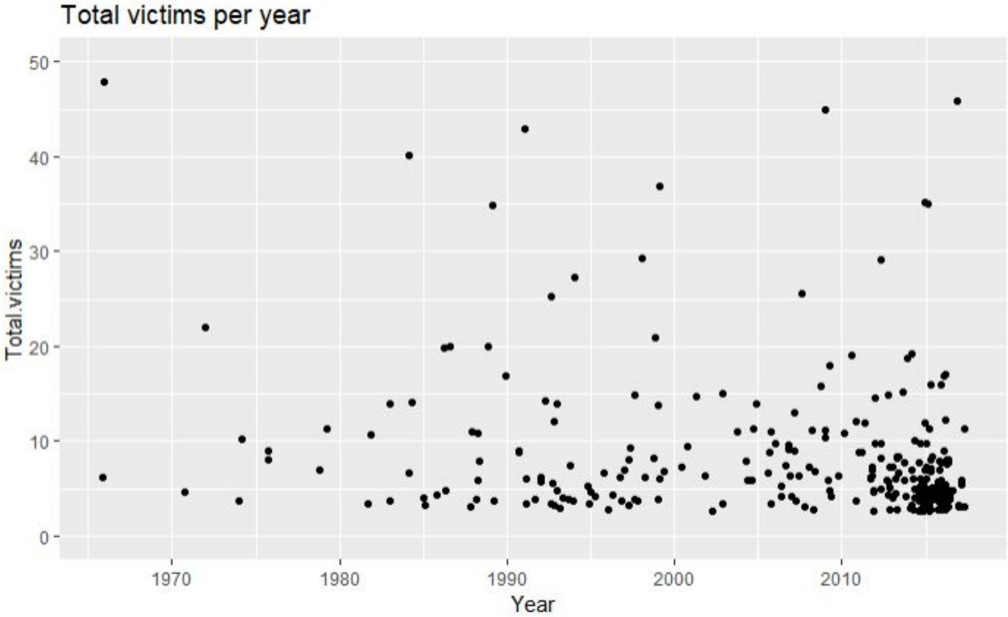
```
g + geom_point(mapping = aes(x=`Year`, y=Total.victims), position="jitter") + ggtitle("Total victims per year")
```



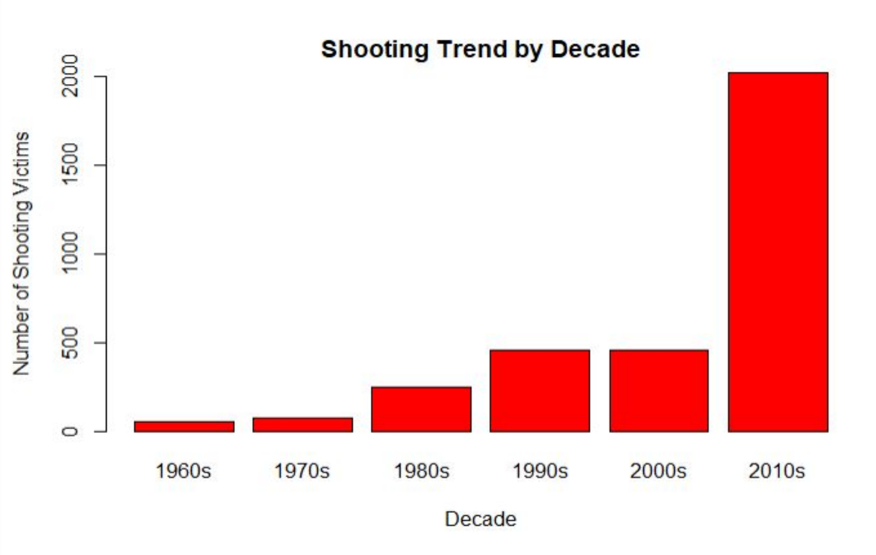
- We answer this question using ggplot to plot the total number of victims (injured + fatalities) versus year.
- The plot shows a slight increase in the number of victims, with the one outlier point corresponding to the Las Vegas shooting in 2017.



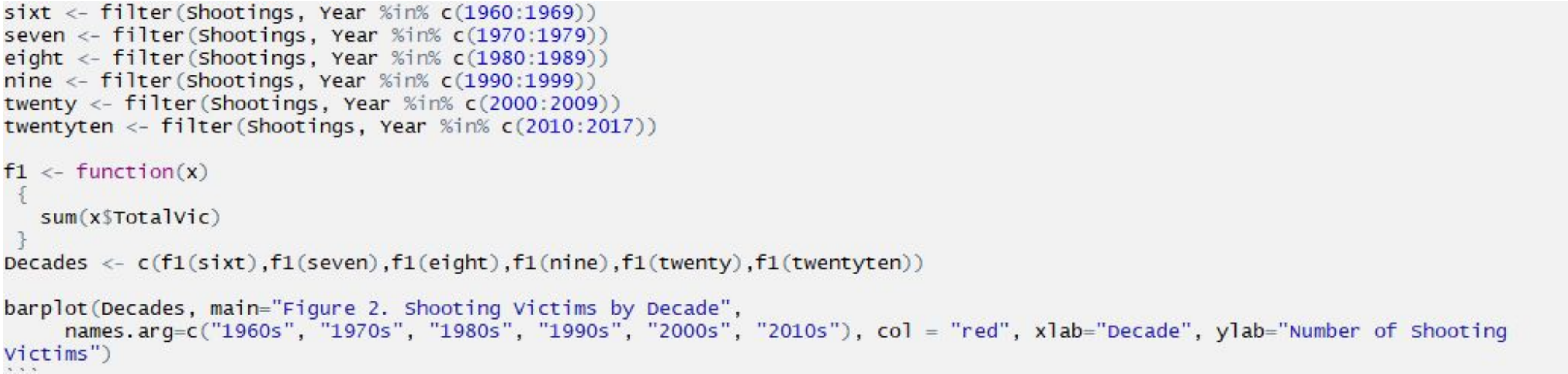
~~~
```{r}
g + geom_point(mapping = aes(x=`Year`, y=Total.victims), position="jitter") + ggtitle("Total victims per year") + ylim(0, 50)

```
~~~



- This plot shows the number of victims per year without the number of victims involved in the Las Vegas shooting in 2017.

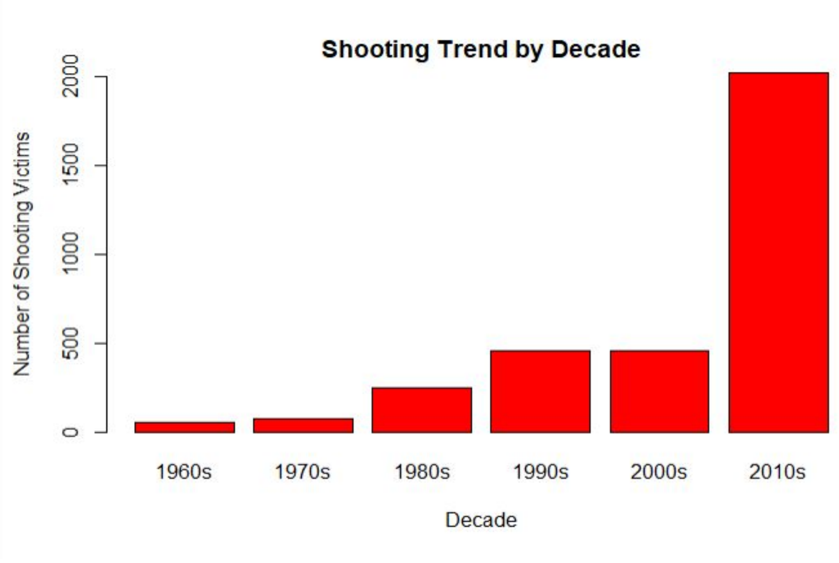


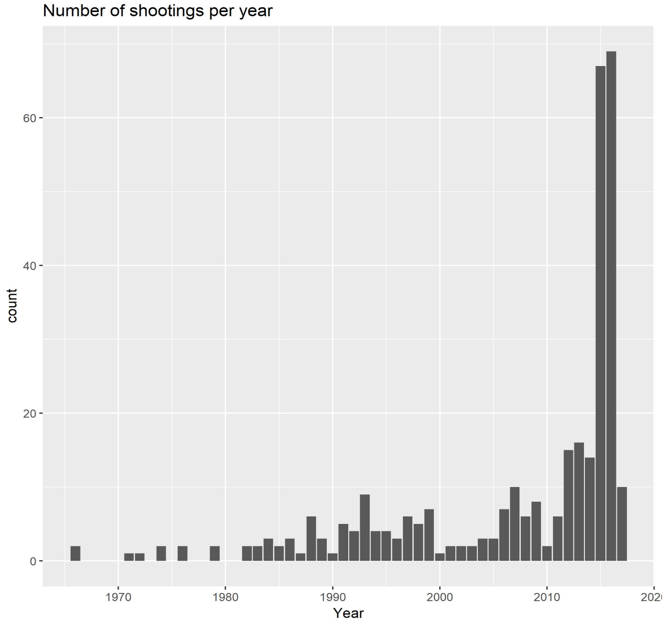
- This plot shows that total number of victims per decade has greatly increased in the 2010s compared to previous decades.
- The function shown below sums the total victim count for every decade.



```
sixt <- filter(Shootings, Year %in% c(1960:1969))
seven <- filter(Shootings, Year %in% c(1970:1979))
eight <- filter(Shootings, Year %in% c(1980:1989))
nine <- filter(Shootings, Year %in% c(1990:1999))
twenty <- filter(Shootings, Year %in% c(2000:2009))
twentyten <- filter(Shootings, Year %in% c(2010:2017))

f1 <- function(x)
 {
   sum(x$Totalvic)
 }
Decades <- c(f1(sixt),f1(seven),f1(eight),f1(nine),f1(twenty),f1(twentyten))

barplot(Decades, main="Figure 2. Shooting Victims by Decade",
     names.arg=c("1960s", "1970s", "1980s", "1990s", "2000s", "2010s"), col = "red", xlab="Decade", ylab="Number of Shooting
Victims")
```

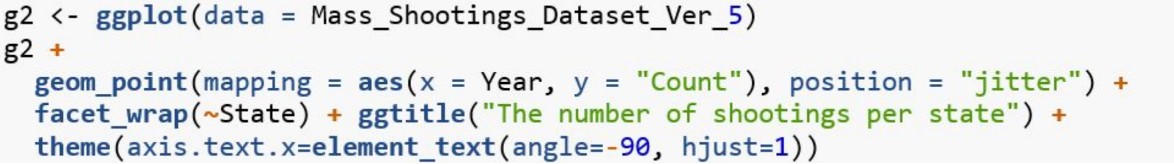
The Rise of Mass Shootings From 1966 - 2017



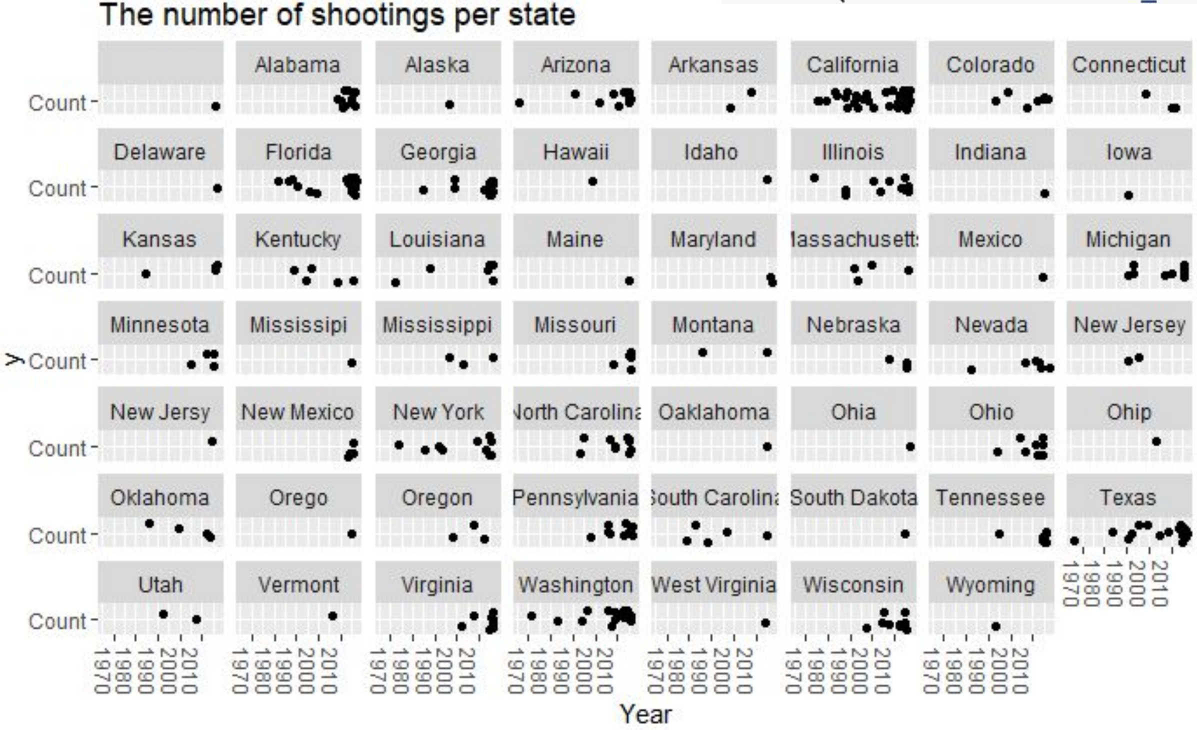


- These two bar plots indicate that both the number of shootings and number of victims has increased over the last 50 years, with the 2010s seeing a dramatic increase.

How many shootings in each state?

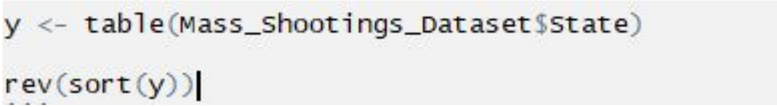


```
g2 <- ggplot(data = Mass_Shootings_Dataset_Ver_5)
g2 +
  geom_point(mapping = aes(x = Year, y = "Count"), position = "jitter") +
  facet_wrap(~State) + ggtitle("The number of shootings per state") +
  theme(axis.text.x=element_text(angle=-90, hjust=1))
```

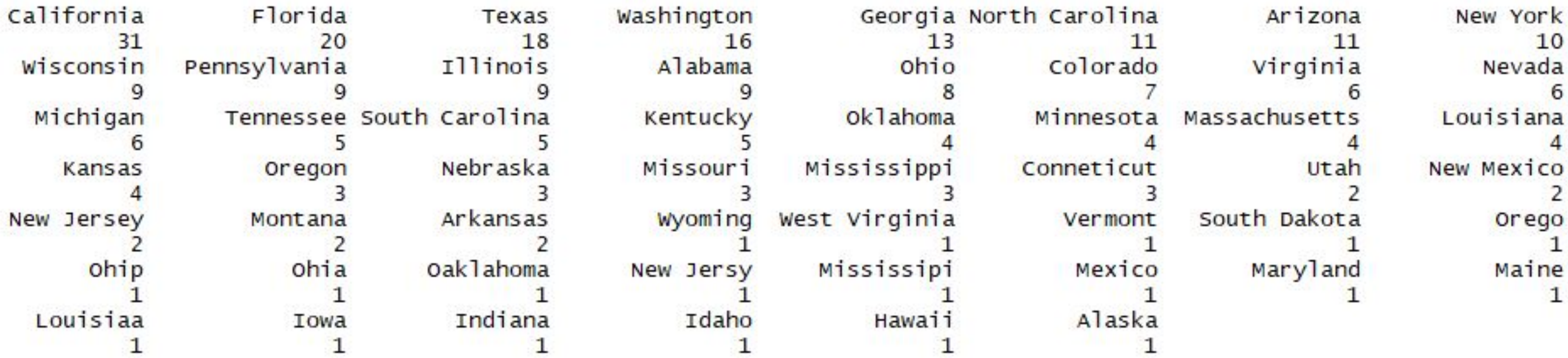


The number of shootings per state

- This figure shows that the states with the greatest amount of mass shootings are California, Florida, and Texas.



```r
y <- table(Mass_Shootings_Dataset$State)

rev(sort(y))
```



```
California        Florida          Texas     Washington        Georgia North Carolina        Arizona       New York 
        31             20             18             16             13             11             11             10 
 Wisconsin   Pennsylvania       Illinois        Alabama           Ohio       Colorado       Virginia         Nevada 
         9              9              9              9              8              7              6              6 
  Michigan      Tennessee South Carolina       Kentucky       Oklahoma      Minnesota  Massachusetts      Louisiana 
         6              5              5              5              4              4              4              4 
    Kansas         Oregon       Nebraska       Missouri    Mississippi     Conneticut           Utah     New Mexico 
         4              3              3              3              3              3              2              2 
New Jersey        Montana       Arkansas        Wyoming  West Virginia        Vermont   South Dakota          Orego 
         2              2              2              1              1              1              1              1 
      Ohip           Ohia      Oaklahoma      New Jersy      Mississipi         Mexico       Maryland          Maine 
         1              1              1              1              1              1              1              1 
  Louisiaa           Iowa        Indiana          Idaho         Hawaii         Alaska 
         1              1              1              1              1              1 
```

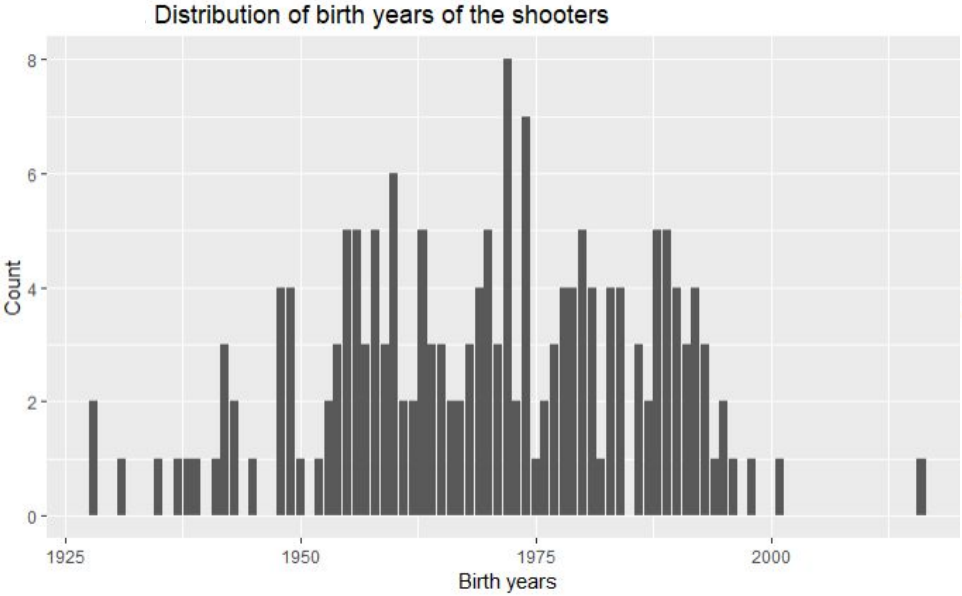
- This table shows, numerically, the number of mass shootings in each state with California being the highest.

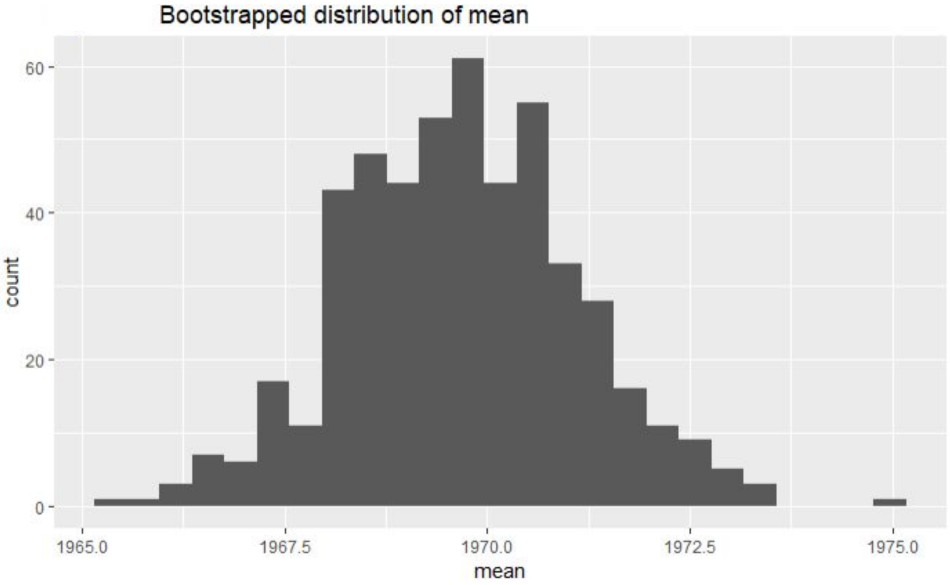
Something interesting to investigate might be to compare when the shooter was born to the shooting that was done.

- For example, if a 30 year old man performed a shooting in 2010, it would mean he was born and raised during the 80s.
- Evaluating shooters based on their age in relation to the shooting can reveal whether something has occurred in our society to create more "shooters" after a certain date, or if the number of people willing to commit such a crime has been constant and have simply had greater access or incentive to commit their crimes in recent years.
- The following graphs indicate that most of the mass shooters were born in the mid to latter-half of the 20th century, particularly in the mid-70s

Was there a particular period of time when a shooter was more likely to have been born?

- Shooters' birth-years are centered around 1970.

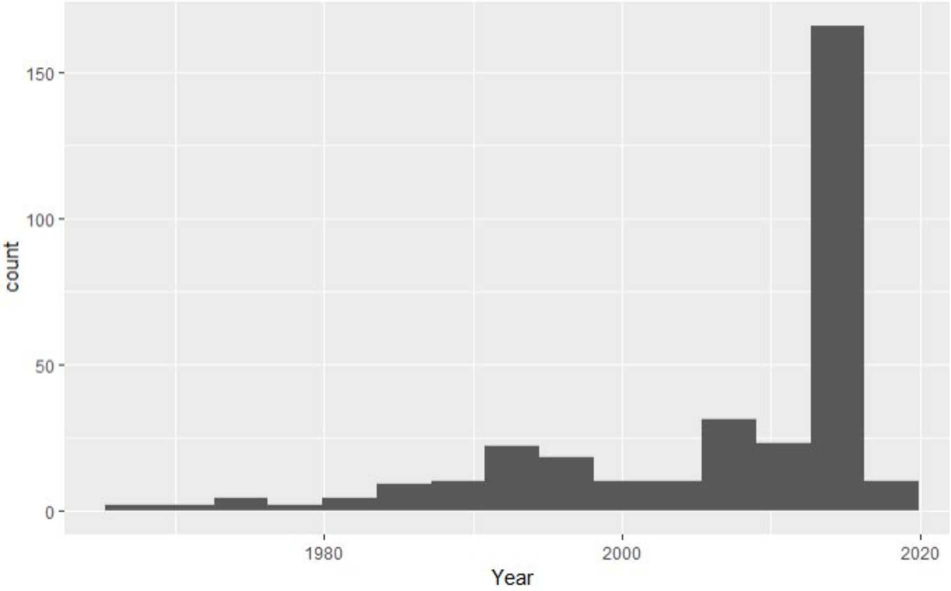


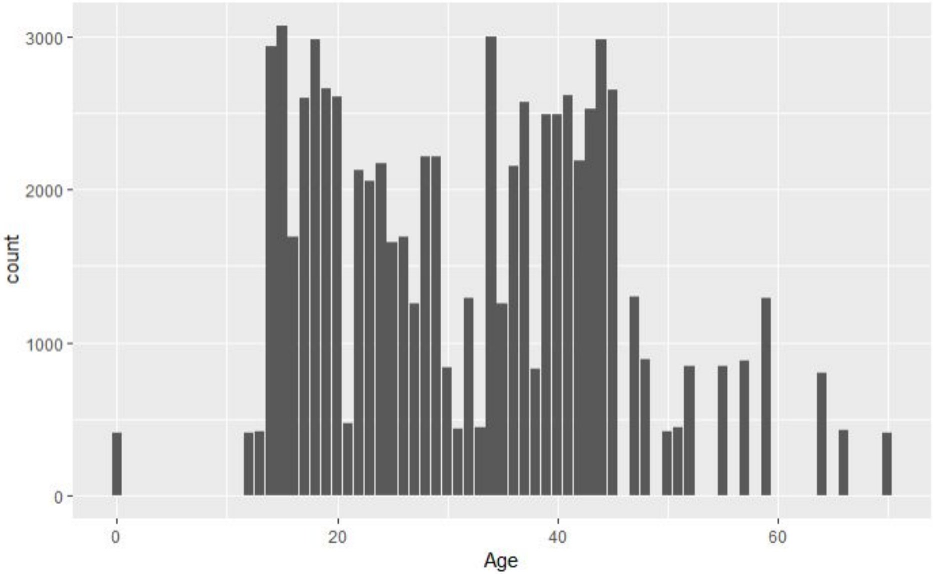


Code

```
SampBorn<-Shootings %>%filter(!is.na(Age), !is.na(Year)) %>% mutate(YearBorn = Year - Age)
Samp2<-SampBorn$YearBorn
z<-do(500)*mean(~YearBorn, data= sample_n(SampBorn, size = 150, replace = T))
ggplot(z, aes(mean))+geom_histogram(bins = 25)+ggtitle("Figure 2. Bootstrapped distribution of mean")
z<-do(500)*(sample_n(SampBorn, size = 150, replace = T))
ggplot(z, aes(YearBorn))+ geom_bar()
ggplot(z, aes(Age)) + geom_bar()
```

But does the total number of shootings have a similar distribution? What about the age distribution?

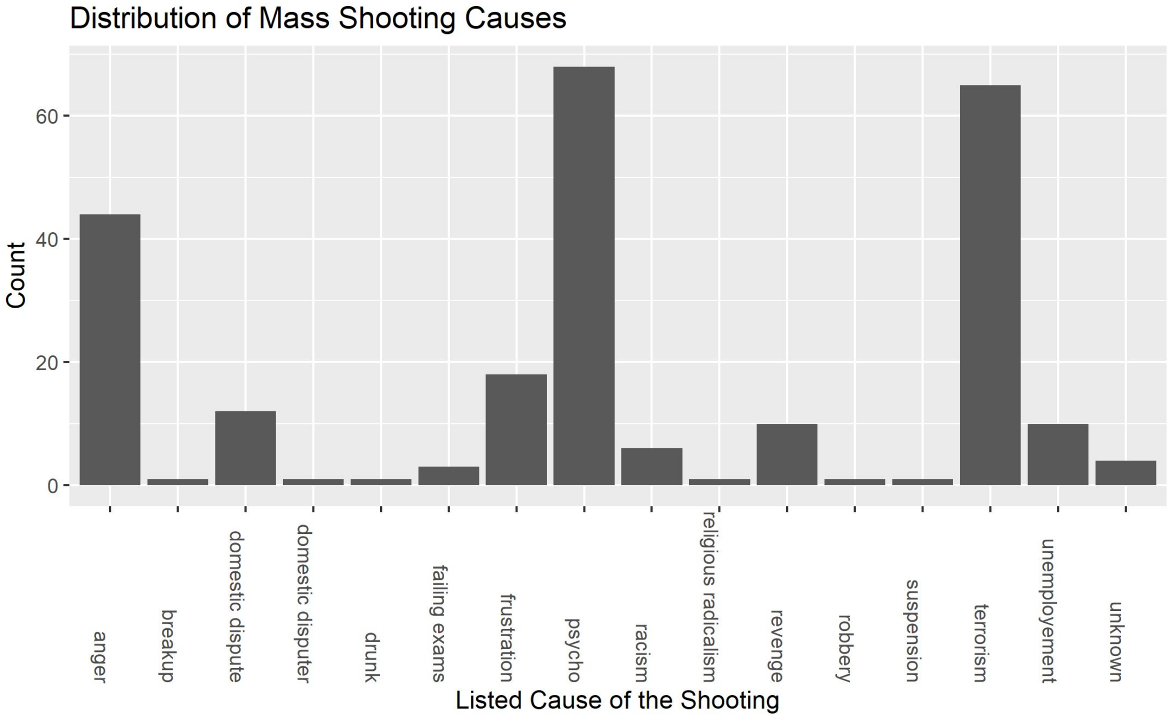




- This figure, given that the mean birth year for mass shooters is around 1970, might suggest that most mass shooters are middle-aged.

- This figure, on the other hand, does not indicate that most mass shooters are middle aged

An Examination Into The Cause of Mass



- This bar graph shows the distribution of the various causes of mass shootings from 1966 - 2017.

- According to the graph, the top causes for mass shootings are mental health problems (psycho), terrorism, and anger; where mental health problems are the leading cause.

Examining mental health

- According to the National Institute of Mental Health, 1 in 5 Americans struggle with mental health issues. [2]
- We used bootstrapping techniques to check the confidence interval for the mean proportion of shooters with mental health issues and found that with 95% confidence between 42.95% and 67.05% of shooters have mental health issues.
- To do this, we had to filter out the "unclear" from our mental health variable, and then set it as a factor and relevel it from 3 levels to 2, so that we could convert it into logical values to take the bootstrapped mean. After bootstrapping we had to lapply the mean on the list to be able to use qdata, since qdata would not take a list.

Code for obtaining the mental health variable

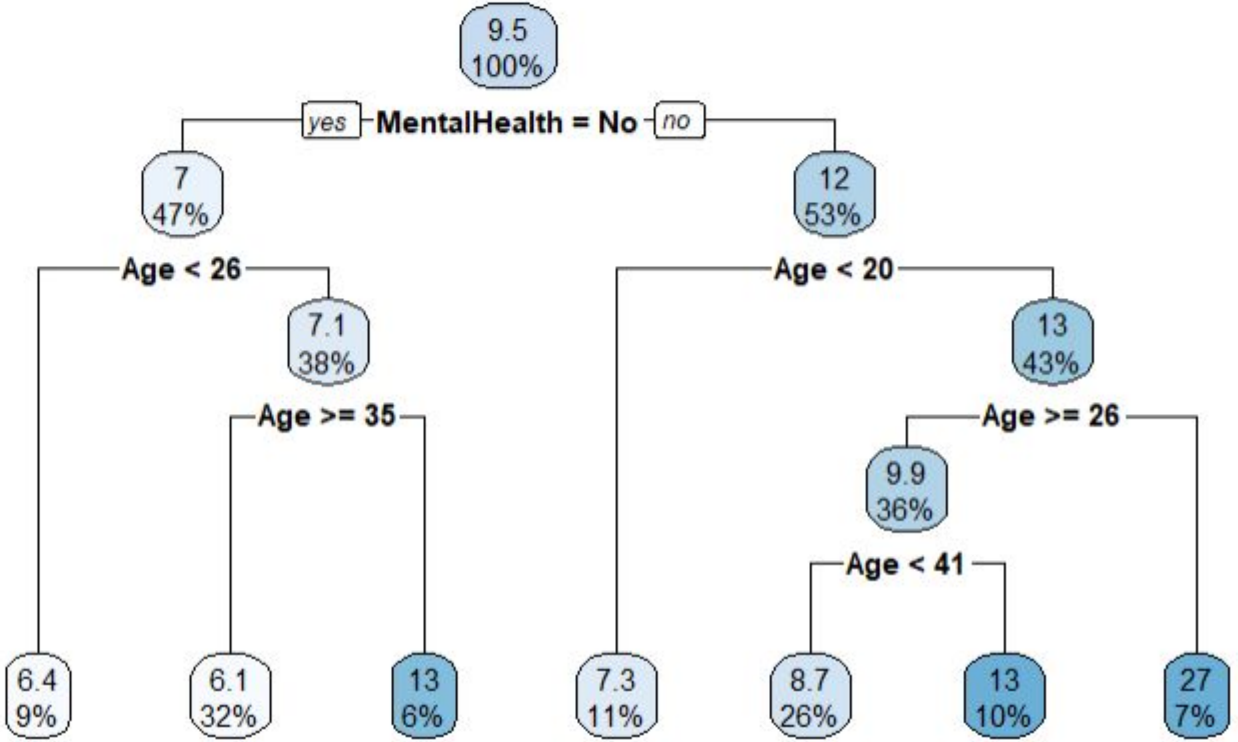
```
a<-Shootings%>%filter(MentalHealth %in% c("No", "Yes"))
a$MentalHealth<-as.factor(a$MentalHealth)
a$MentalHealth<-factor(a$MentalHealth)
levels(a$MentalHealth)<-c(FALSE, TRUE)
a$MentalHealth<-as.logical(a$MentalHealth)
AvgMental<-do(100)*mean(~MentalHealth, data = sample_n(a, size = 50, replace = T))
lapply(AvgMental, mean)
qdata(~mean, AvgMental, p = c(.025, .975))
```

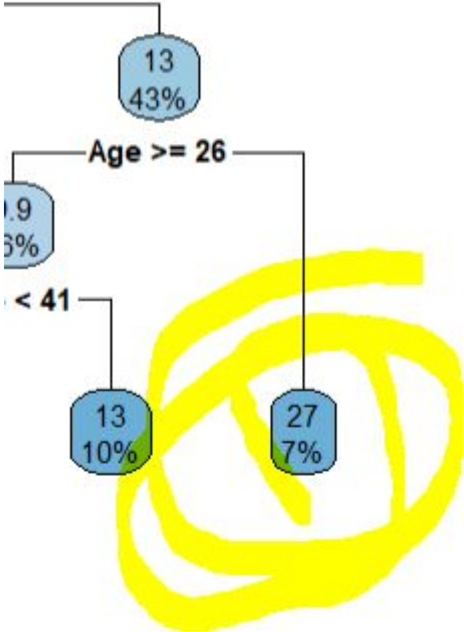
Mental health, continued

Furthermore, we filtered all mental health options other than “Yes” and “No” and put it into a model against Total Victims and found that if a person did have mental health, that on average there would be 4.73 more victims, with a high F value.

```
          Analysis of Variance Table
Response: TotalVic
                    Df  Sum Sq Mean Sq F value   Pr(>F)
MentalHealth   1  1108.5 1108.48  11.037 0.001065 **
Residuals     197 19785.2  100.43

          lm(formula = TotalVic ~ MentalHealth, data = a)
Coefficients:
(Intercept)  MentalHealthYes

     6.968          4.730
```

Total victims using rpart as a predictor based on mental health and age





What we noticed:

People between the ages of 20 and 26 with no mental health issues tend to kill significantly more people (almost triple the mean) but just how accurate is our model?

```
partmod<-rpart(TotalVic~ MentalHealth + Age, a)
rpart.plot(partmod)
```

Examining the model

Our first thought was that perhaps there was an outlier skewing our data, so we immediately checked the Las Vegas Shooting, however that shooting had been listed as “unclear” in regards to mental health and had already been filtered out. Just to be sure, we used the same variables and data frame to create a linear model and check the residuals to see if any had gone beyond Cook’s distance- none had.

However, since the tree didn’t have a categorical variable as the response we were unsure of how to test its accuracy. We were able to use confidence interval testing to determine that with 95% confidence the mean number of victims is between 6 and 21.

Code

```
AvgVic<-do(100)*mean(~TotalVic, data = sample_n(Shootings, size = 50, replace = T))

qdata(~mean, AvgVic, p = c(.025, .975))

treemod<-lm(TotalVic~ MentalHealth + Age, TrainA)
plot(treemod)
```

A better tree!

We used randomForest to create a forest of trees to find the best predictor to try to determine the cause of the shooting based on the mental health (Y/N) of the shooter and the year the shooting took place.

Using a training/testing set and the predict function, we tested the accuracy of the tree, and despite cause having 13 different levels, the randomForest tree was accurate approximately 40% of the time.

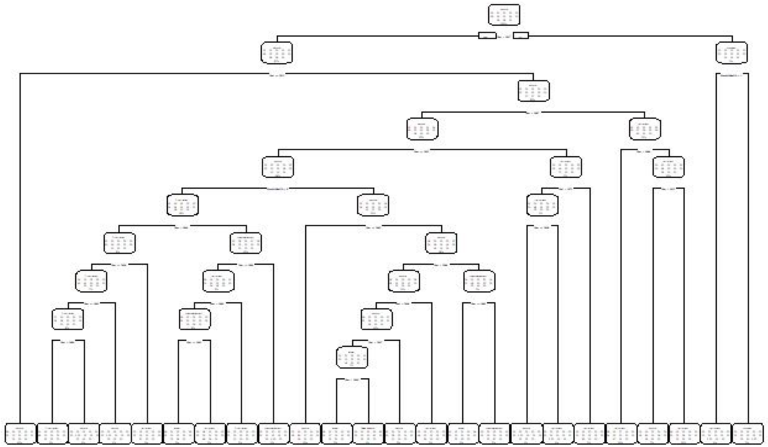
Code

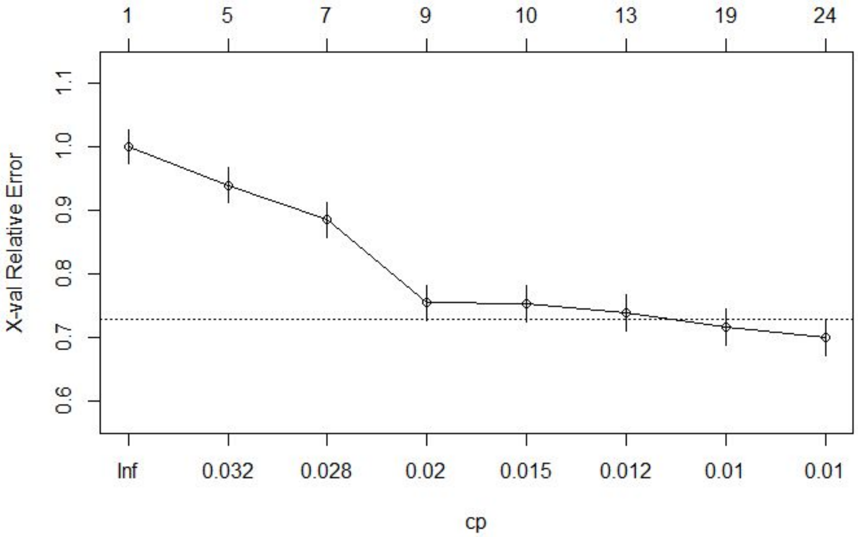
```
a<-a%>%replace_with_na(replace = list(Cause = ""))
a<-filter(a, !is.na(Cause))
a2<-sample(a, size = 1000, replace = TRUE)
index<-sample(1000, 1000*.75)
TrainA<-a2[index,]
TestA<-a2[-index,]
TrainA$MentalHealth<-as.integer(TrainA$MentalHealth)
TestA$MentalHealth<-as.integer(TestA$MentalHealth)
TrainA$Cause<-droplevels(TrainA$Cause)
TestA$Cause<-droplevels(TestA$Cause)
partmod<-randomForest(Cause~  Year + MentalHealth, data = TrainA)
TestA$pred<-predict(partmod, TestA, type = "class")
mean(TestA$pred == TestA$Cause)
```

An example tree from the forest

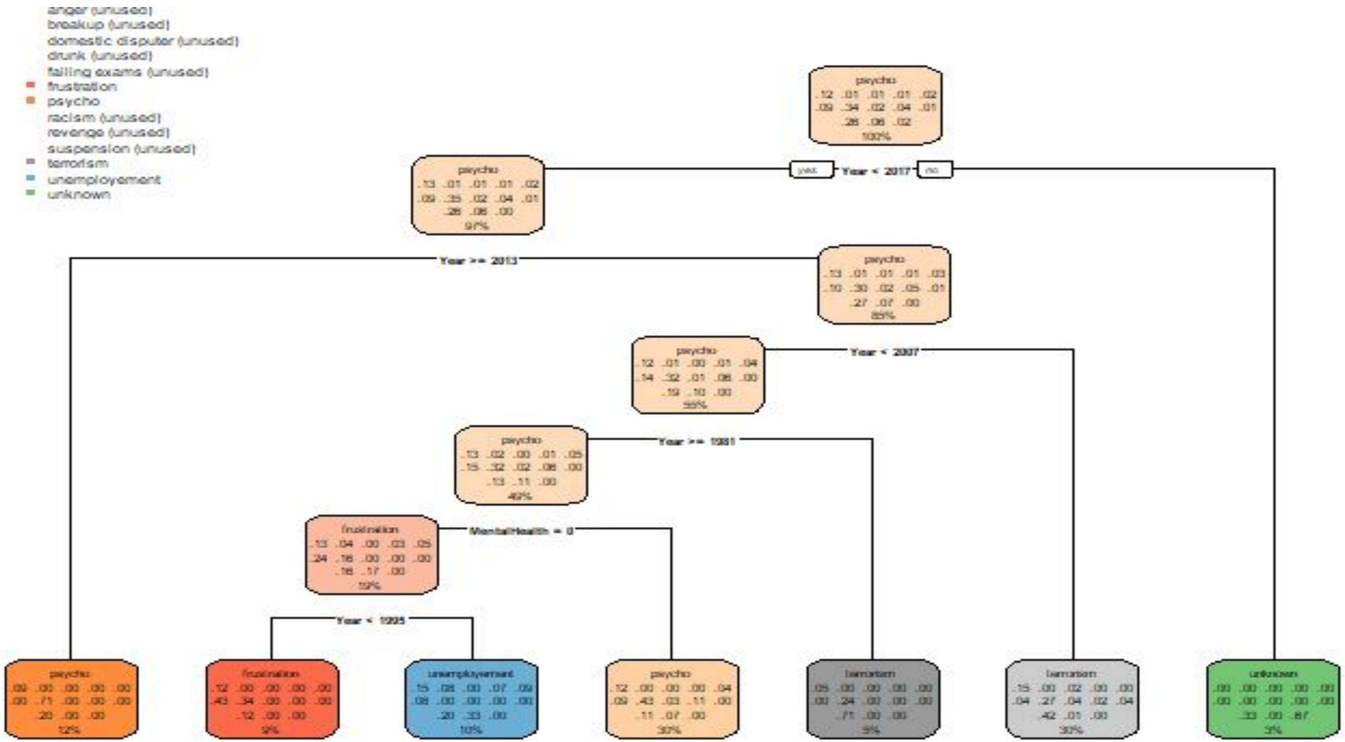
So the first tree is unhelpful- too much growth! So we looked at the CP (complexity parameter) plot:

```
examplepart<-rpart(Cause~ Year + MentalHealth, data = TrainA)
rpart.plot(examplepart)
```





After pruning: Still difficult to read, but a better model!



Code

```
examplepart<-rpart(Cause~ Year + MentalHealth, data = TrainA)
rpart.plot(examplepart)
plotcp(examplepart)
examplepart<-rpart(Cause~ Year + MentalHealth, data = TrainA, cp = .03)
rpart.plot(examplepart)
```

Conclusions

- Our investigation shows that mass shootings in the United States have seen an unprecedented increase over the last fifty years, and that one of the major leading causes is mental health issues.
- The National Alliance of Mental Illness reports that 1 in 5 people experience mental health issues every year.
- This statistic and our findings suggest that possible preventative measures for future mass shootings include increased awareness of mental health, and funding to programs and facilities which help those in need; and imposing greater restrictions, or ban individuals who have mental health issues from obtaining firearms.

References

https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years/home

https://www.nami.org/learn-more/mental-health-by-the-numbers