

US Mass Shootings- Redesign Project

1. Introduction

This redesigning project consists of a misleading graph related to the Mass Shootings which happened in US which we tried to improve by using R. We have come across the article published on 2nd December 2015 on TruthStream Media which stated that the number of mass shootings in US under the Presidential term of Obama was highest. They released a graph which stated a drastic increase in the shootings under Obama. We could visually sense that the graph published had various issues.

Initially, we identified the issues with the original graph. Later, on further research we could find a relevant data set using which this graph was published. Going forward, we were able to redesign the bad graph and also provide additional possible visualizations for better understanding of the article from the dataset. Hence, we redesigned the graph by suppressing the issues and also made it understandable to the reader.

2. Bad Graph

a) Source

This article has been taken from the website Truthstream Media and was posted on 2nd December 2015. The required visualizations are made from the dataset obtained from Mother Jones database where the data collected is from 1982-2019. The dataset is in the form of CSV file. This dataset includes data related to the number of shootings, location of shootings, weapons obtained, States prone to shooting under each presidential term.

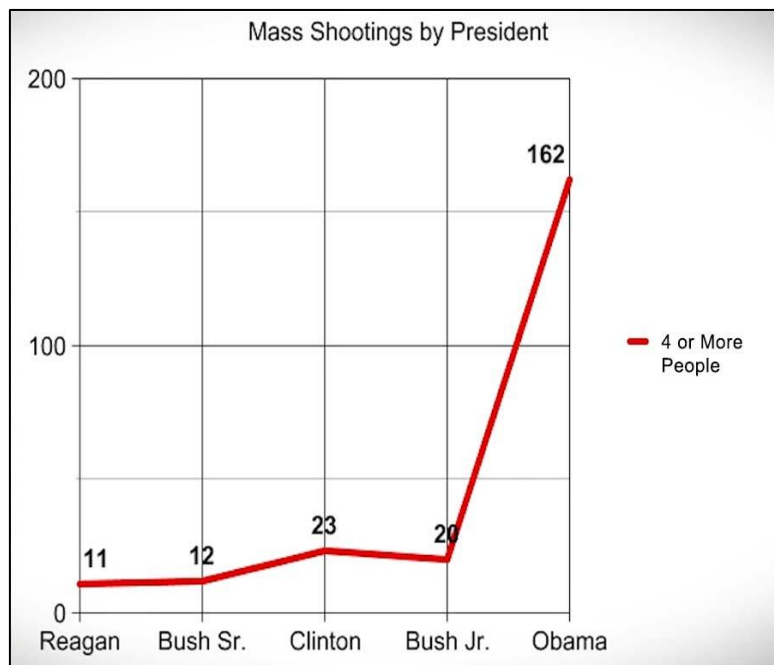


Figure 1: Bad Graph

b) Context

This news report is mainly concentrated on why there were more mass shootings under presidency of Obama than the four previous presidents combined? But there weren't so many mass shootings. Mass shootings have skyrocketed over the past seven years, under the Presidency of Obama. Here the considered the data set which included relevant data from 1982-2019. Hence, the line graph is not an appropriate graph to represent the current context. Also the definition of Mass shooting were a single person is attacking 4 or more people in a single place, was violated while designing the above graph.

c) Weakness

The bad graph we have chosen has the following weakness. Figure 2 clearly represents the issues with the graph. Firstly, the values do not represent time or continuous values; usage of a line graph is a bad choice to represent the data. Secondly, we could observe that the title of the graph is misleading. Also, labeling of the axis is not present; hence it is not possible to interpret what the Y-Axis is trying to explain. For single line plot, usage of legend makes the graph complex and misleading.

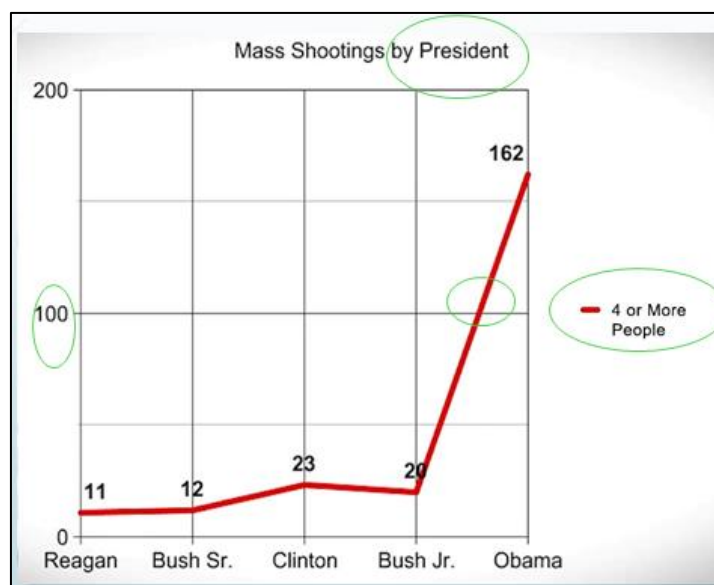


Figure 2: Weakness of the Bad Graph

3. Redesigned Graphs

a) Objective and Improvements

The main objective while redesigning the graph is to make it visually appealing and also understandable by any user who is not aware of the context. Hence, we came to a conclusion that **Bar Graph** is the best way to redesign the line graph as it could represent data without misleading. We have also modified the title and made it less ambiguous for the reader. Graph shown in Figure-3 represents the Frequency of shootings under each presidential term. The Axis labelling clearly shows the presidents along with their term period and also the number of shootings. This graph clearly shows that the value shown for Obama in the original line graph is incorrect whereas the number of shooting under his term is 38 but not 162. We have also used the colour-blind friendly palette “Dark2” palette from the color-brewer website

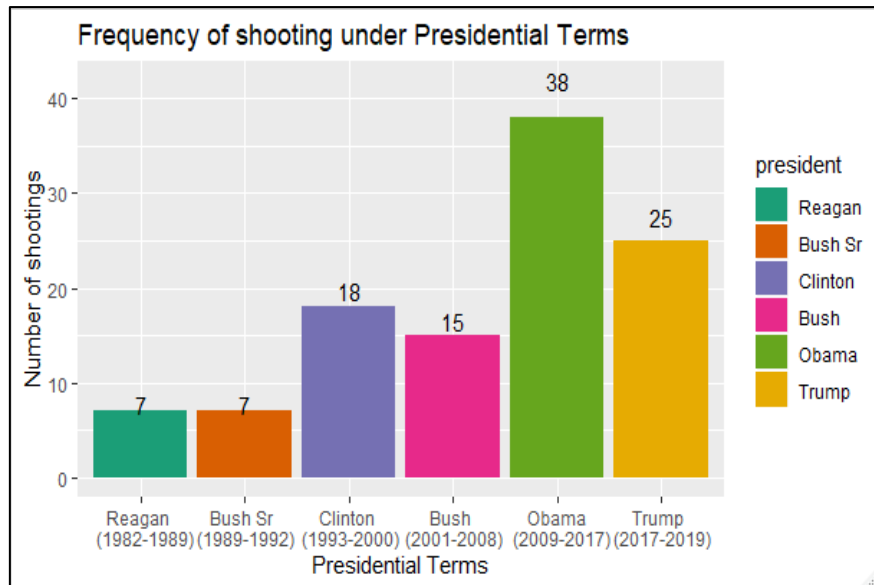


Figure 3: ReDesigned Bad Graph

b) Software Solutions

For redesigning the graph and also to make possible additional visualizations, we have used R studio as the primary software. We installed and loaded packages like ggplot2 and tidyverse to produce visualizations and clean the dataset. Also in-order to show the states prone to more shooting plot, we have installed and loaded the usmap package. Firstly, we loaded the CSV file into data frame and then cleaned the data to avoid Null values. Later, with this data we made necessary visualizations using the ggplot() function.

Code:

```
library(tidyverse)
library(ggplot2)
#installing the usmap package to generate the USmap plot
#install.packages("usmap")
library(usmap)

#Reading the CSV data file into a data frame
shootingData=read.csv(file="F:/chandana/CourseWork/Sem_1/Assignments/Stat515/MidTerm/shootingdata_new_1.csv",header=TRUE)

#There are certain blank rows in the csv file, To exclude the blank rows we use the below
shootingData2<-shootingData[1:110,]
shootingData2

#Checking for null values
is.na(shootingData2)

#Assigning Levels to the Presidents using factor
shootingData2$president<-factor(shootingData2$president,levels=c("Reagan","Bush Sr", "Clinton", "Bush","Obama", "Trump"))
```

```
xlabel<-c('Reagan \n (1982-1989)','Bush Sr \n (1989-1992)','Clinton \n (1993-2000)','Bush \n (2001-2008)','Obama\n (2009-2017)','Trump \n (2017-2019)')
```

```
#Plot between Number of shootings and Presidents
ggplot(shootingData2, aes(x=president,fill=president))+
  geom_bar()+
  labs(x='Presidential Terms',y='Number of shootings',title='Frequency of
shooting under Presidential Terms') +
  scale_fill_brewer(palette = "Dark2")+scale_x_discrete(labels=xlabel)+
  geom_text(stat = 'count', aes(label=..count..), position=position_stack(1.10))
```

```
#Plot between Location of Shooting and number of Shooting with presidency
ggplot(shootingData2, aes(x=president, fill=categories))+
  geom_bar()+
  labs(x='Presidential Terms',y='Number of shootings',title='Frequency of
shooting under Presidential Terms \n with respect to Location') +
  scale_fill_brewer(palette = "Dark2")+scale_x_discrete(labels=xlabel)
```

```
#Plot between Location of Shooting and number of Shooting
ggplot(shootingData2, aes(x=categories,fill=categories))+
  geom_bar()+
  labs(x='Location of the Shootings',y='Number of shootings',title='Frequency of
shooting with respect to Location') +
  scale_fill_brewer(palette = "Dark2")
```

```
#Plot between States and the weapons obtained legally
shootingData2$weapons_obtained_legally[shootingData2$weapons_obtained_1
egally == "Yes "] <- "Yes"
```

```
shootingData2$weapons_obtained_legally<-
factor(shootingData2$weapons_obtained_legally,levels=
c('Yes','No','Unknown','TBD'))
ggplot(shootingData2, aes(x=Location_ab, fill=weapons_obtained_legally))+
  geom_bar()+
  labs(x='States in the US',y='Number of shootings',title='Frequency of
Shootings in each state \n Based on the Weapons Obtained') +
  scale_fill_brewer(palette = "Dark2")
```

```
#Plot between Weapons Obtained and the Presidential terms
shootingData2$weapons_obtained_legally[shootingData2$weapons_obtained_1
egally == "Yes "] <- "Yes"
```

```
shootingData2$weapons_obtained_legally<-
factor(shootingData2$weapons_obtained_legally,levels=
c('Yes','No','Unknown','TBD'))
ggplot(shootingData2, aes(x=president, fill=weapons_obtained_legally))+
```

```
geom_bar()+
labs(x='Presidential Terms',y='Number of shootings',title='Frequency of
shooting under Presidential Terms \n with respect to Weapons Obtained') +
scale_fill_brewer(palette = "Dark2")+scale_x_discrete(labels=xlabel)

#States Prone to Shootings
#using the usmap to show the most prone shooting states in the US
#Considering the Total Victims column to interpret the shootings at that
location

a1<-shootingData2[, c('Location_ab', 'total_victims')]
count2<-aggregate(cbind(count = total_victims) ~ Location_ab,
                  data = a1,
                  FUN = function(x){NROW(x)})

colnames(count2)<-c('state','count')

plot_usmap(data=count2, values='count',lines='black',labels='True')+
labs(title = "Mass shootings in USA from 1982 to 2019")+
scale_fill_continuous(
  low = "orange", high = "red", name = "Mass shooting", label = scales::comma
) + theme(legend.position = "right")
```

c) Challenges

From the data set that we have obtained from the MotherJones database, we could make other visualizations for which certain modifications to the csv file was required. We added column for US State Abbreviations for generating the Map plot. Also we added a column called categories which groups the location of shootings into 5 categories like public venue, workplace, schools, religions and others based on data description. Like, places including restaurants, night clubs, and concerts are categorized into public venue and so on for better visualization. By making these additional changes, we could interpret the following visualizations and get a clear view of the persisting issue.

d) Additional Visualizations

In Figure 4, the bar graph represents the number of shootings with respect to the locations (i.e., public venue, religious, school, workplace and other). The Workplace area is more prone to mass shootings compared to any other location. The second highest area subjected to mass shooting is a public venues and the least shootings occurred at Religious locations.

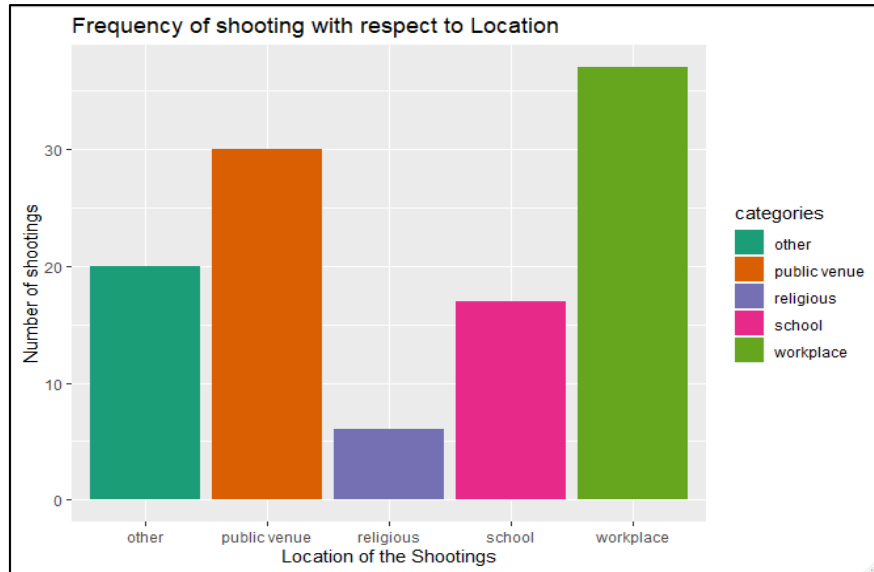


Figure 4: Plot Between Frequency of shooting and Location of shooting

In Figure 5, the graph shows the number of shootings under each presidency term with respect to the locations. As we can see that the number of shootings at public venue is more than any other locations under Obama's presidency. Similarly, the number of shootings at workplace is highest when compared to the other locations under the presidency of Clinton and Trump.

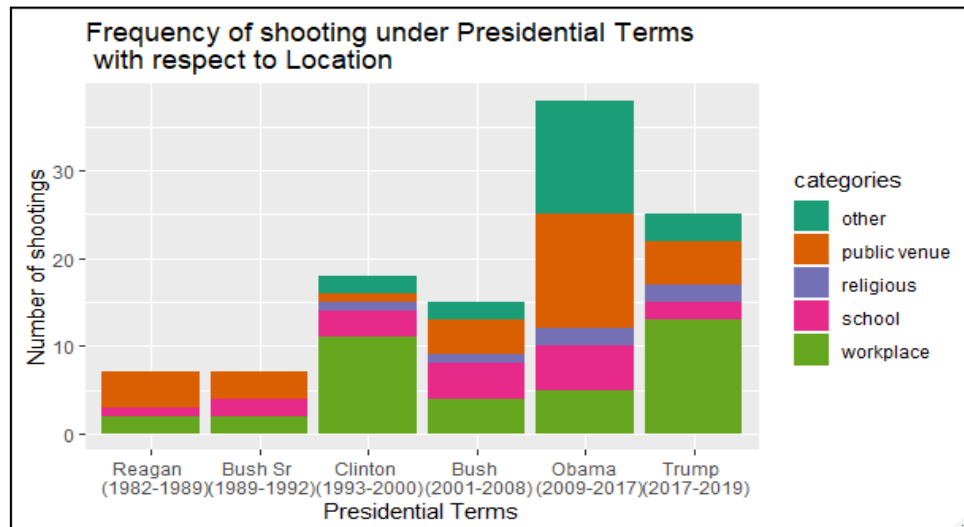


Figure 5: Plot Between Frequency of shooting under Presidential Terms with respect to Location of shooting

In Figure 6, the map plot shows the exact State with their range of mass shootings represented from least to highest using orange and red colors respectively, making California the state with the highest mass shootings. The gray color indicates that there are no mass shootings in that State.

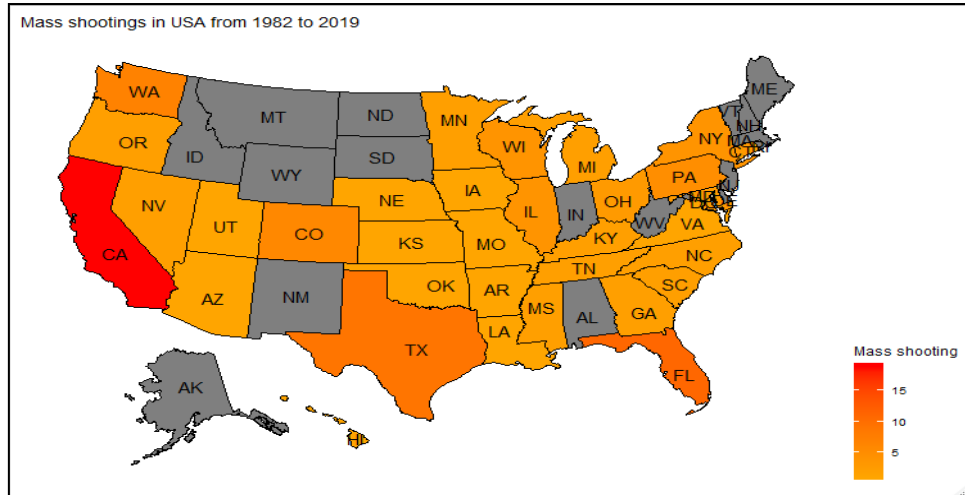


Figure 6: State Prone Map indicating the Mass shootings

Figure 7, the stacked graph represents the number of shooting for each state with respect to whether the weapons obtained legally or not. Hence, shooters obtained weapons legally in every state are more when compared to the weapons obtained illegally.

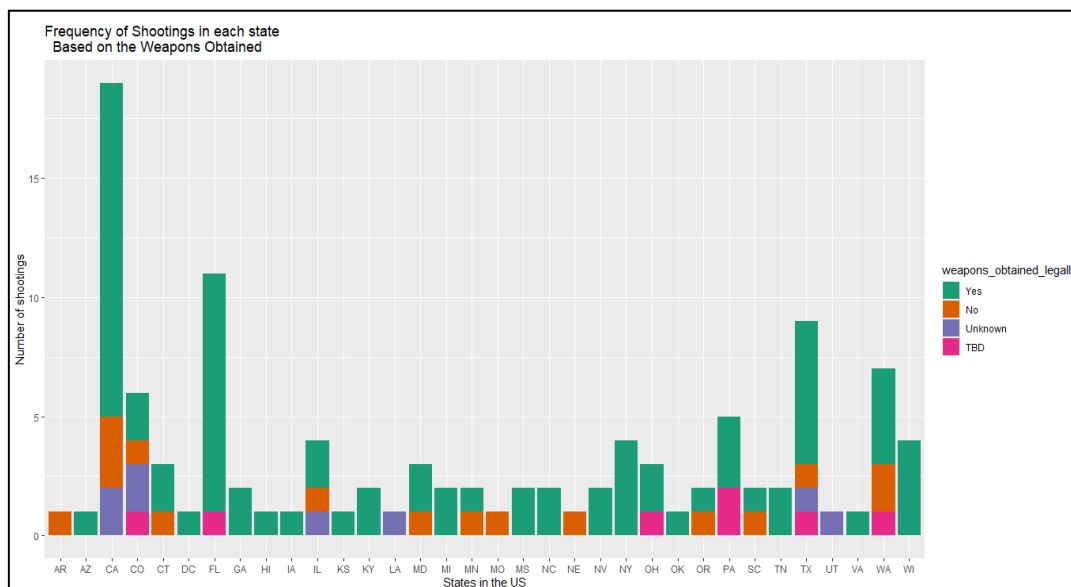


Figure 7: Plot Between Frequency of shooting in each State based on Weapons obtained

Figure 8, the below graph is between the number of shootings under each presidency with respect to whether the weapons obtained legally or not. The weapons obtained legally by shooters under each president is highest than obtained illegally.

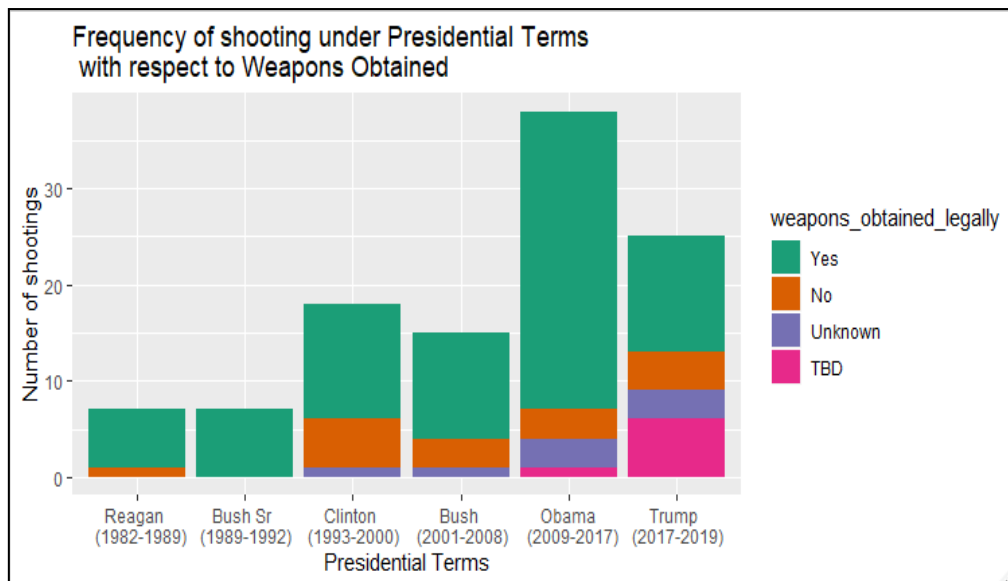


Figure 8: Plot between Frequency of shooting under Presidential Terms with respect to Weapons Obtained

4. Conclusion

After redesigning the graph and also by analyzing the additional visualizations, we could interpret the following conclusions. Firstly, President Obama had the highest number of shootings compared to the other presidents, but not as drastic as shown in the original graph. Also, the original graph did not use the apt definition of mass shootings. However, the overall rate of mass shootings by presidential terms has been increasing. In most of the states, under each presidency, the weapons were obtained legally. Also through our visualization, we could conclude that there is no particular trend for the location of mass shootings.

Contribution:

I have been a part of finding and analyzing the bad graph for re-designing process. Together we have decided on finding the best alternative graph for the original bad graph. Contributed towards data cleaning and CSV file. For additional visualizations, we have tried various plots between data items and presented the appropriate graphs as a part of report. Worked on preparing the presentation for the class and also contributed equally for preparing the documentation. Also cited below references for the report.

References

- DataCamp. (n.d.). *R Documentation*. Retrieved from https://www.rdocumentation.org/packages/usmap/versions/0.5.0/topics/plot_usmap
- DYKES, M. (2015, December 2). *Truthstream Media*. Retrieved from <https://web.archive.org/web/20151204065556/http://truthstreammedia.com/2015/12/02/why->
- Kodali, T. (2015, September 20). *Datascience plus*. Retrieved from <https://datascienceplus.com/using-the-ggplot2-library-in-r/>

Mark Follman, G. A. (2019, August 31). *Mother Jones*. Retrieved from <https://www.motherjones.com/politics/2012/07/mass-shootings-map/>

Troy, T. (2018, December). *National Affairs*. Retrieved from <https://www.nationalaffairs.com/publications/detail/presidents-and-mass-shootings>