# STAT 5193

Asitha Mudiyanselage

Graduate Project

## 1 Summary

In this project, I have planned to analyze the most risk areas of getting tornadoes in Oklahoma, USA. Once, I have the final output then risk counties of getting tornadoes can be predicted. Input data were obtained from National Weather Service web site. Tornado data from 2014 to 2018 were extracted form following websites.

- https://www.weather.gov/oun/tornadodata-ok-2014
- https://www.weather.gov/oun/tornadodata-ok-2015
- https://www.weather.gov/oun/tornadodata-ok-2016
- https://www.weather.gov/oun/tornadodata-ok-2017
- https://www.weather.gov/oun/tornadodata-ok-2018

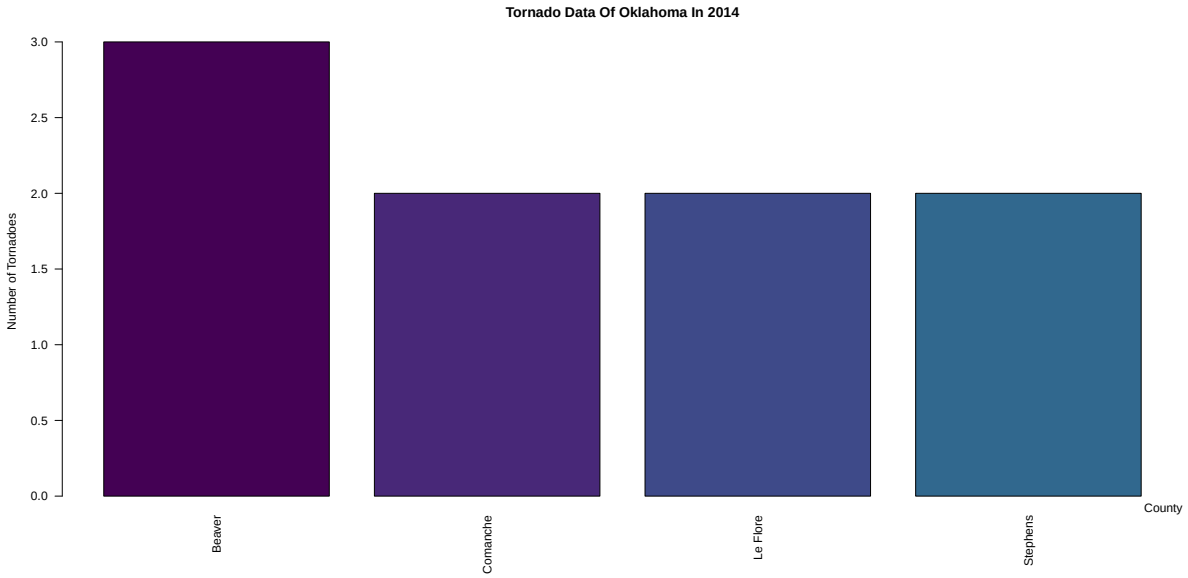
Finally, data were analyzed and visualized using some popular functions in R. The vertical bar charts for number of tornado in each year were plotted. Then as a final outcome, the vertical bar chart was plotted for mean of number of tornadoes for last five years.

In this project I have used following main functions to analyze and visualize the results.

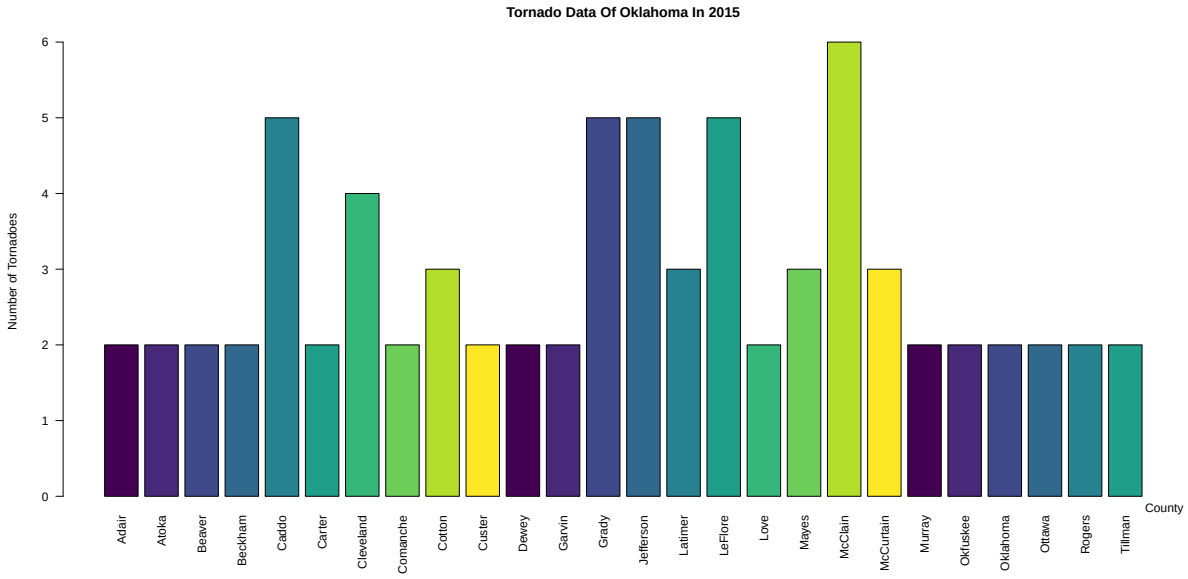
- **aggregate** – used to group by variable within data frame then calculate the summary statistics for each group

- **rbind** – helped to merge two data frames horizontally.
- **table** – used to get frequency counts for categorical variable
- **subset** – used to take small portion from large data set.
- **mtext** – used to write a text in one of the four margins of the current figure.

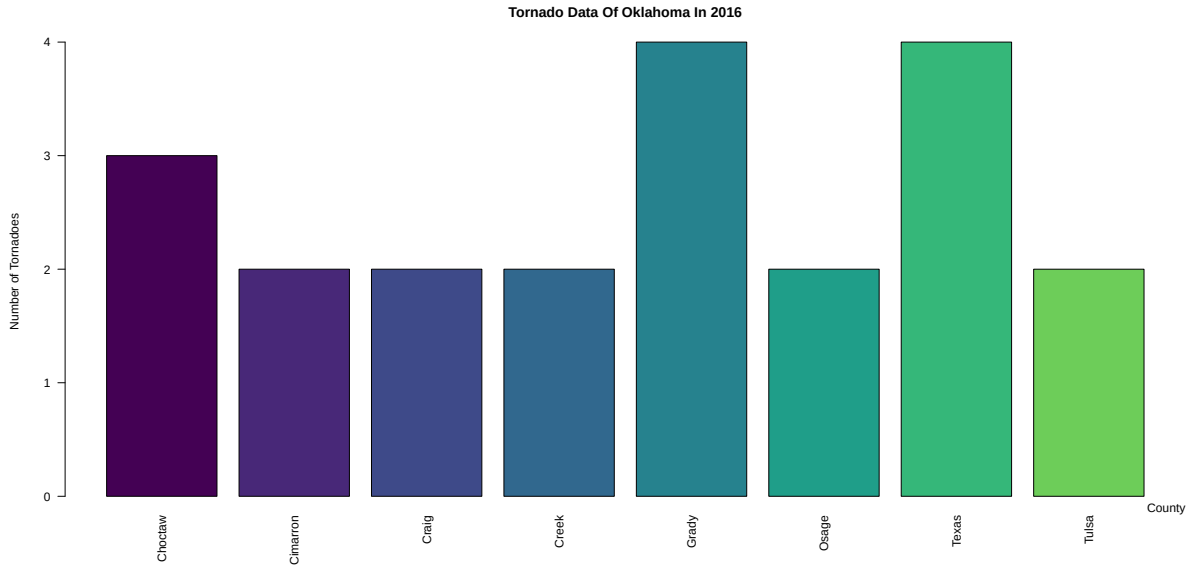
# 2 Results



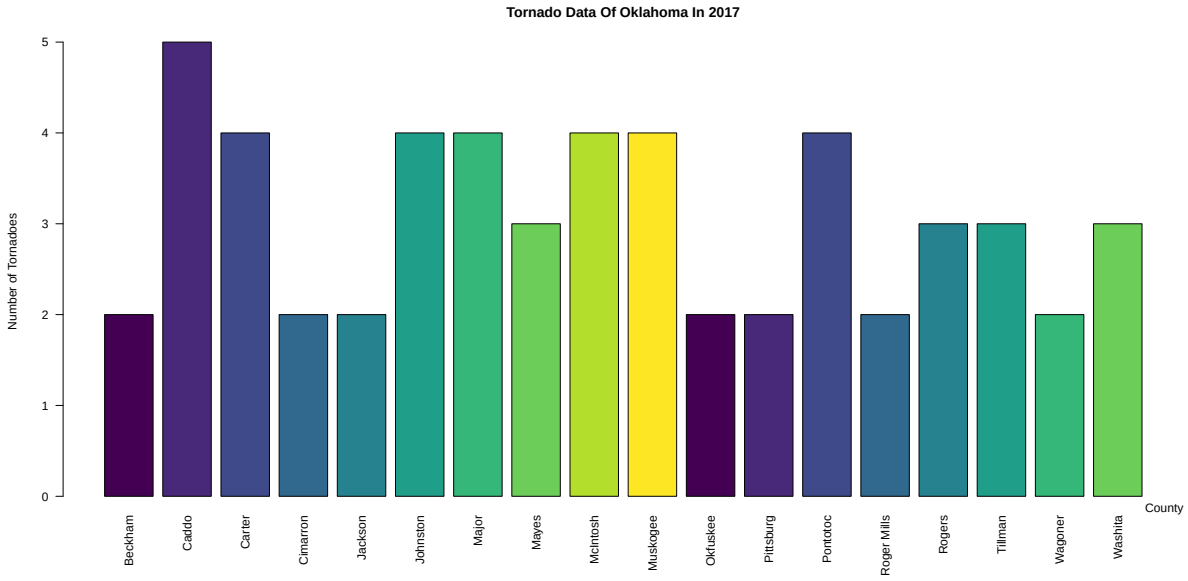

- Beaver county has the highest number of tornadoes in 2014.
- There were only four counties that had more than one tornadoes.



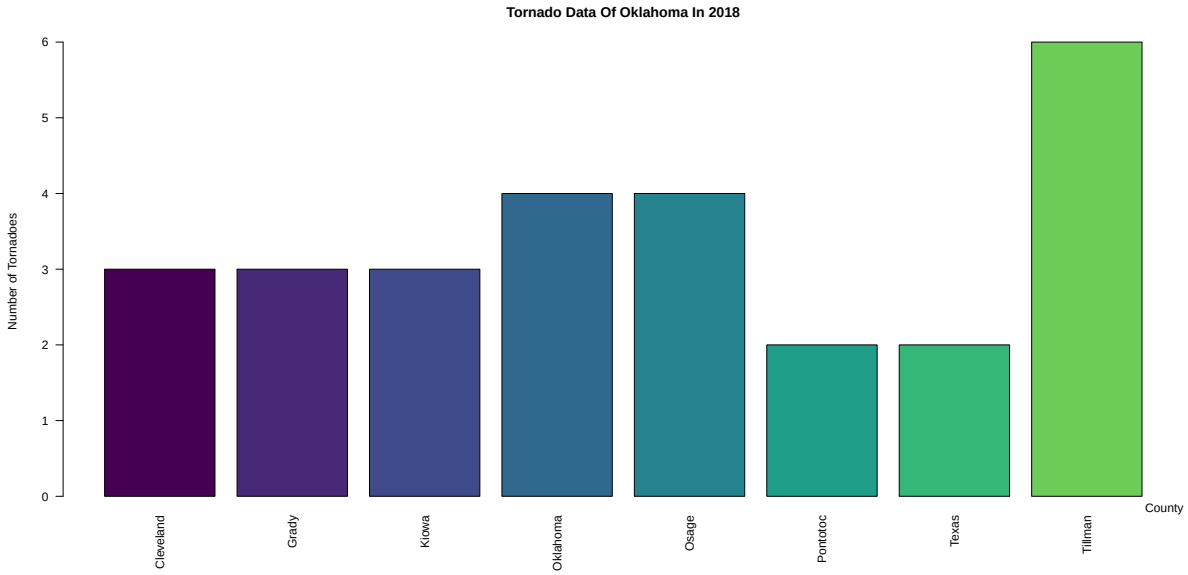

- There were 26 counties which had more than one tornadoes and the Mcclain county has the highest number of tornadoes in 2015.
- Caddo, Cleveland, Grady, Jefferson and Mcclain had more than four tornadoes.



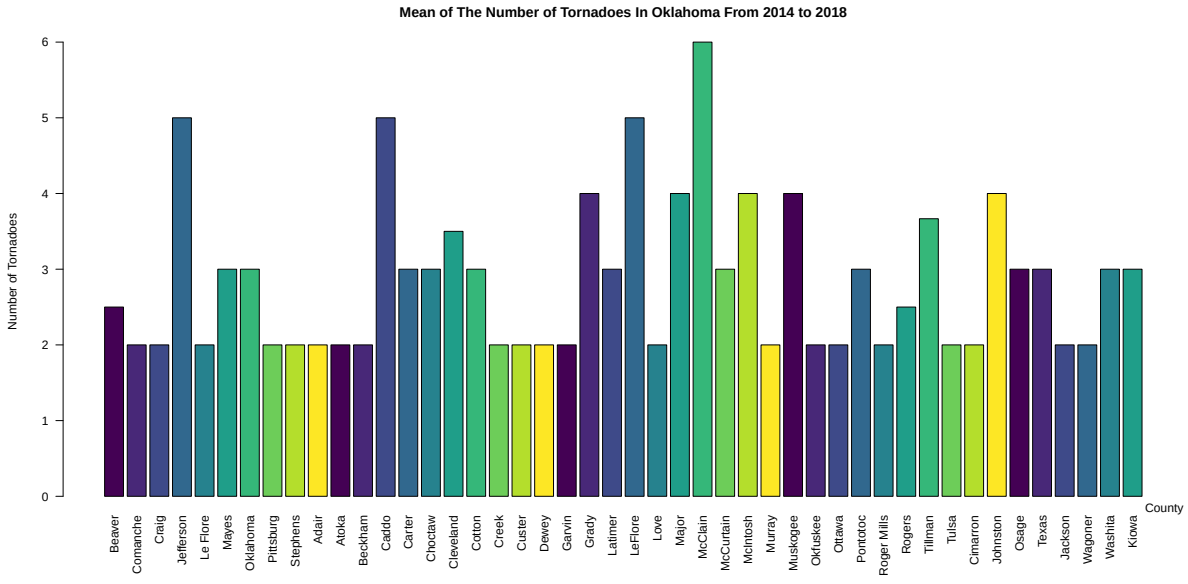

- Grady and Osage had 4 tornadoes in 2016 and only eight counties which had more than one tornadoes.




- Caddo had 5 tornadoes and it was the highest value in 2017.
- In this year there were 18 counties which had tornadoes and seven of them had more than four tornadoes.




- Tillman county had six tornadoes in 2018 and it was the highest.There were only eight counties which had more than one tornadoes.



- Mcclain county has the highest mean of number of tornadoes from 2014-2018.
- According to the vertical bar chat, we can see that there are nine counties which have the mean value more than 4.
- Therefore, We can predict that nine counties have huge risk of getting tornadoes in coming years.

# 3 Appendix

```
> setwd("C:/Users/akaruna/Desktop/R-Assignment/Project_1")
>
> getwd()
[1] "C:/Users/akaruna/Desktop/R-Assignment/Project_1"
>
> # Read the csv files from Project_1 folder.
>
> Data_14 <- read.csv("GRAD_PRO14.csv",header=T)
> Data_15 <- read.csv("GRAD_PRO15.csv",header=T)
> Data_16 <- read.csv("GRAD_PRO16.csv",header=T)
> Data_17 <- read.csv("GRAD_PRO17.csv",header=T)
> Data_18 <- read.csv("GRAD_PRO18.csv",header=T)
>
> # Take the small portion of large data set using subset command.
>
> New_DF14<-subset(Data_14,select=c(Date,County))
> New_DF15<-subset(Data_15,select=c(Date,County))
> New_DF16<-subset(Data_16,select=c(Date,County))
> New_DF17<-subset(Data_17,select=c(Date,County))
> New_DF18<-subset(Data_18,select=c(Date,County))
>
> # Get the frequency count of Counties for each years using table command and
> # use as.data.frame to set the data to data frame.
> # Finally rename the columns and rearrange the index.
>
> New_DF2014<-as.data.frame(table(New_DF14$County))
> colnames(New_DF2014)<- c("County","Tonardo_Count")
> New_DF2014<-New_DF2014[-1,]
>
>
> New_DF2015<-as.data.frame(table(New_DF15$County))
```

```
> colnames(New_DF2015)<- c("County","Tonardo_Count")
> New_DF2015<-New_DF2015[-1,]
>
>
>
> New_DF2016<-as.data.frame(table(New_DF16$County))
> colnames(New_DF2016)<- c("County","Tonardo_Count")
> New_DF2016<-New_DF2016[-1,]
>
> New_DF2017<-as.data.frame(table(New_DF17$County))
> colnames(New_DF2017)<- c("County","Tonardo_Count")
> New_DF2017<-New_DF2017[-1,]
>
> New_DF2018<-as.data.frame(table(New_DF18$County))
> colnames(New_DF2018)<- c("County","Tonardo_Count")
> New_DF2018<-New_DF2018[-1,]
>
> # Then consider the county which has more than one tornado
and visualize the data.
>
> Final_2014<-as.data.frame(New_DF2014[New_DF2014$Tonardo_Count!= 1,])
> rownames(Final_2014) <- NULL
> Final_2014
    County Tonardo_Count
1   Beaver             3
2 Comanche             2
3 Le Flore             2
4 Stephens             2
>
> Final_2015<-as.data.frame(New_DF2015[New_DF2015$Tonardo_Count!= 1,])
> rownames(Final_2015) <- NULL
> Final_2015
       County Tonardo_Count
1       Adair             2
2       Atoka             2
3      Beaver             2
4     Beckham             2
5       Caddo             5
6      Carter             2
7   Cleveland             4
8    Comanche             2
9      Cotton             3
10     Custer             2
11      Dewey             2
12     Garvin             2
13      Grady             5
```

```
14  Jefferson              5
15     Latimer             3
16     LeFlore             5
17        Love             2
18       Mayes             3
19     McClain             6
20  McCurtain              3
21      Murray             2
22    Okfuskee             2
23    Oklahoma             2
24      Ottawa             2
25      Rogers             2
26     Tillman             2
>
> Final_2016<-as.data.frame(New_DF2016[New_DF2016$Tonardo_Count!= 1,])
> rownames(Final_2016) <- NULL
> Final_2016
     County Tonardo_Count
1   Choctaw              3
2  Cimarron              2
3     Craig              2
4     Creek              2
5     Grady              4
6     Osage              2
7     Texas              4
8     Tulsa              2
>
>
> Final_2017<-as.data.frame(New_DF2017[New_DF2017$Tonardo_Count!= 1,])
> rownames(Final_2017) <- NULL
> Final_2017
          County Tonardo_Count
1        Beckham              2
2          Caddo              5
3         Carter              4
4       Cimarron              2
5        Jackson              2
6       Johnston              4
7          Major              4
8          Mayes              3
9       McIntosh              4
10      Muskogee              4
11      Okfuskee              2
12     Pittsburg              2
13      Pontotoc              4
14  Roger Mills               2
```

```
15       Rogers              3
16      Tillman              3
17      Wagoner              2
18      Washita              3
>
> Final_2018<-as.data.frame(New_DF2018[New_DF2018$Tonardo_Count!= 1,])
> rownames(Final_2018) <- NULL
> Final_2018
     County Tonardo_Count
1 Cleveland             3
2     Grady             3
3     Kiowa             3
4  Oklahoma             4
5     Osage             4
6  Pontotoc             2
7     Texas             2
8   Tillman             6
>
> # plot vertical bar chat for each year.
>
> barplot(Final_2014$Tonardo_Count,
+         names.arg = Final_2014$County ,
+         ylab = "Number of Tornadoes",
+         col = viridis(10),
+         main = "Tornado Data Of Oklahoma In 2014",
+         border = "black",
+         las=2)
>
> mtext(side = 1, text = "County", adj = 1)
>
>
> barplot(Final_2015$Tonardo_Count,
+         names.arg = Final_2015$County ,
+         ylab = "Number of Tornadoes",
+         col =viridis(10),
+         main = "Tornado Data Of Oklahoma In 2015",
+         border = "black",
+         las=2)
>
> mtext(side = 1, text = "County", adj = 1)
>
> barplot(Final_2016$Tonardo_Count,
+         names.arg = Final_2016$County ,
+         ylab = "Number of Tornadoes",
+         col = viridis(10),
+         main = "Tornado Data Of Oklahoma In 2016",
```

```
+          border = "black",
+          las=2)
>
> mtext(side = 1, text = "County", adj = 1)
>
> barplot(Final_2017$Tonardo_Count,
+          names.arg = Final_2017$County ,
+          ylab = "Number of Tornadoes",
+          col = viridis(10),
+          main = "Tornado Data Of Oklahoma In 2017",
+          border = "black",
+          las=2)
>
> mtext(side = 1, text = "County", adj = 1)
>
> barplot(Final_2018$Tonardo_Count,
+          names.arg = Final_2018$County ,
+          ylab = "Number of Tornadoes",
+          col = viridis(10),
+          main = "Tornado Data Of Oklahoma In 2018",
+          border = "black",
+          las=2)
>
> mtext(side = 1, text = "County", adj = 1)
>
> # Combine all data frame form 2014 to 2018 into a single data frame.
>
>
> Master_Data<-rbind(Final_2014,Final_2015,Final_2016,Final_2017,Final_2018)
> Master_Data
        County Tonardo_Count
1       Beaver              3
2     Comanche              2
3     Le Flore              2
4     Stephens              2
5        Adair              2
6        Atoka              2
7       Beaver              2
8      Beckham              2
9        Caddo              5
10      Carter              2
11   Cleveland              4
12    Comanche              2
13      Cotton              3
14      Custer              2
15       Dewey              2
```

| | | |
|---|---|---|
| 16 | Garvin | 2 |
| 17 | Grady | 5 |
| 18 | Jefferson | 5 |
| 19 | Latimer | 3 |
| 20 | LeFlore | 5 |
| 21 | Love | 2 |
| 22 | Mayes | 3 |
| 23 | McClain | 6 |
| 24 | McCurtain | 3 |
| 25 | Murray | 2 |
| 26 | Okfuskee | 2 |
| 27 | Oklahoma | 2 |
| 28 | Ottawa | 2 |
| 29 | Rogers | 2 |
| 30 | Tillman | 2 |
| 31 | Choctaw | 3 |
| 32 | Cimarron | 2 |
| 33 | Craig | 2 |
| 34 | Creek | 2 |
| 35 | Grady | 4 |
| 36 | Osage | 2 |
| 37 | Texas | 4 |
| 38 | Tulsa | 2 |
| 39 | Beckham | 2 |
| 40 | Caddo | 5 |
| 41 | Carter | 4 |
| 42 | Cimarron | 2 |
| 43 | Jackson | 2 |
| 44 | Johnston | 4 |
| 45 | Major | 4 |
| 46 | Mayes | 3 |
| 47 | McIntosh | 4 |
| 48 | Muskogee | 4 |
| 49 | Okfuskee | 2 |
| 50 | Pittsburg | 2 |
| 51 | Pontotoc | 4 |
| 52 | Roger Mills | 2 |
| 53 | Rogers | 3 |
| 54 | Tillman | 3 |
| 55 | Wagoner | 2 |
| 56 | Washita | 3 |
| 57 | Cleveland | 3 |
| 58 | Grady | 3 |
| 59 | Kiowa | 3 |
| 60 | Oklahoma | 4 |
| 61 | Osage | 4 |

```
62     Pontotoc              2
63        Texas              2
64      Tillman              6
>
>
> #Calculating the mean of each counties and visualize it.
>
>
> Final_Out<-aggregate(Master_Data$Tonardo_Count,
  by=list(County=Master_Data$County),mean)
>
> barplot(Final_Out$x,
+         names.arg = Final_Out$County ,
+         ylab = "Number of Tornadoes",
+         col =viridis(10),
+         main = "Mean of The Number of Tornadoes
In Oklahoma From 2014 to 2018",
+         border = "black",
+         las=2)
>
> mtext(side = 1, text = "County", adj = 1)
```