

- Review IDMA Chapter 4 and author slide presentation

Reviewed IDMA Chapter 4 and author slide presentation

- Review IDMA Chapter 8 (R) and author slide presentation

Reviewed IDMA Chapter 8 (R) and author slide presentation

- Load the 1950-2019_actual_tornadoes.csv dataset into an R dataframe; display a few records

Load Dataset:

```
df1 <- read.csv('1950-2019_actual_tornadoes.csv',header=TRUE,sep=",")
```

Display of few records:

```
head(df1)
```

```
tail(df1)
```

```

D:/Spring 2021/AIT 580_Prof.Harry Foxwell/S. week5/
# ... WITH 43 MORE ROWS
> df1 <- read.csv('1950-2019_actual_tornadoes.csv',header=TRUE,sep=",")
> head(df1)
  om  yr mo dy      date      time tz st stf stn mag inj fat loss closs slat slon elat
1  1 1950  1  3 1950-01-03 11:00:00  3 MO 29  1  3  3  0  6  0 38.77 -90.22 38.83
2  2 1950  1  3 1950-01-03 11:55:00  3 IL 17  2  3  3  0  5  0 39.10 -89.30 39.12
3  3 1950  1  3 1950-01-03 16:00:00  3 OH 39  1  1  1  0  4  0 40.88 -84.58  0.00
4  4 1950  1 13 1950-01-13 05:25:00  3 AR  5  1  3  1  1  3  0 34.40 -94.37  0.00
5  5 1950  1 25 1950-01-25 19:30:00  3 MO 29  2  2  5  0  5  0 37.60 -90.68 37.63
6  6 1950  1 25 1950-01-25 21:00:00  3 IL 17  3  2  0  0  5  0 41.17 -87.33  0.00
  elon len wid ns sn sg fl f2 f3 f4 fc
1 -90.03 9.5 150 2 0 1 0 0 0 0 0
2 -89.23 3.6 130 1 1 1 135 0 0 0 0
3  0.00 0.1 10 1 1 1 161 0 0 0 0
4  0.00 0.6 17 1 1 1 113 0 0 0 0
5 -90.65 2.3 300 1 1 1 93 0 0 0 0
6  0.00 0.1 100 1 1 1 91 0 0 0 0
> tail(df1)
  om  yr mo dy      date      time tz st stf stn mag inj fat loss closs slat
65157 618536 2019 12 29 2019-12-29 15:53:00  3 MS 28  0  1  0  0 10000  0 33.0978
65158 618537 2019 12 29 2019-12-29 16:03:00  3 MS 28  0  1  0  0 75000  0 33.1628
65159 618538 2019 12 29 2019-12-29 16:13:00  3 MS 28  0  1  0  0 10000  0 33.2598
65160 618539 2019 12 29 2019-12-29 16:32:00  3 MS 28  0  0  0  0 5000  0 33.4720
65161 618540 2019 12 29 2019-12-29 17:13:00  3 MS 28  0  1  0  0 150000  0 32.5268
65162 618541 2019 12 29 2019-12-29 18:50:00  3 AL  1  0  0  0  0 0  0 34.7541
  slon elat elon len wid ns sn sg fl f2 f3 f4 fc
65157 -89.5429 33.1410 -89.4411 6.60 500 1 1 1 7 0 0 0 0
65158 -89.4323 33.2339 -89.3298 7.70 900 1 1 1 7 0 0 0 0
65159 -89.2778 33.2879 -89.2208 3.82 200 1 1 1 19 0 0 0 0
65160 -89.0315 33.4888 -88.9910 2.61 200 1 1 1 105 0 0 0 0
65161 -89.1628 32.5581 -89.1215 3.23 125 1 1 1 101 0 0 0 0
65162 -87.0777 34.7946 -87.0041 5.07 50 1 1 1 83 0 0 0 0
>

```

- Use R to answer *and interpret* the following:
 - Display the range and related summary statistics for the length and width of tornados (Boxplots)

```
df1wid <- subset(df1,subset=wid>500)
```

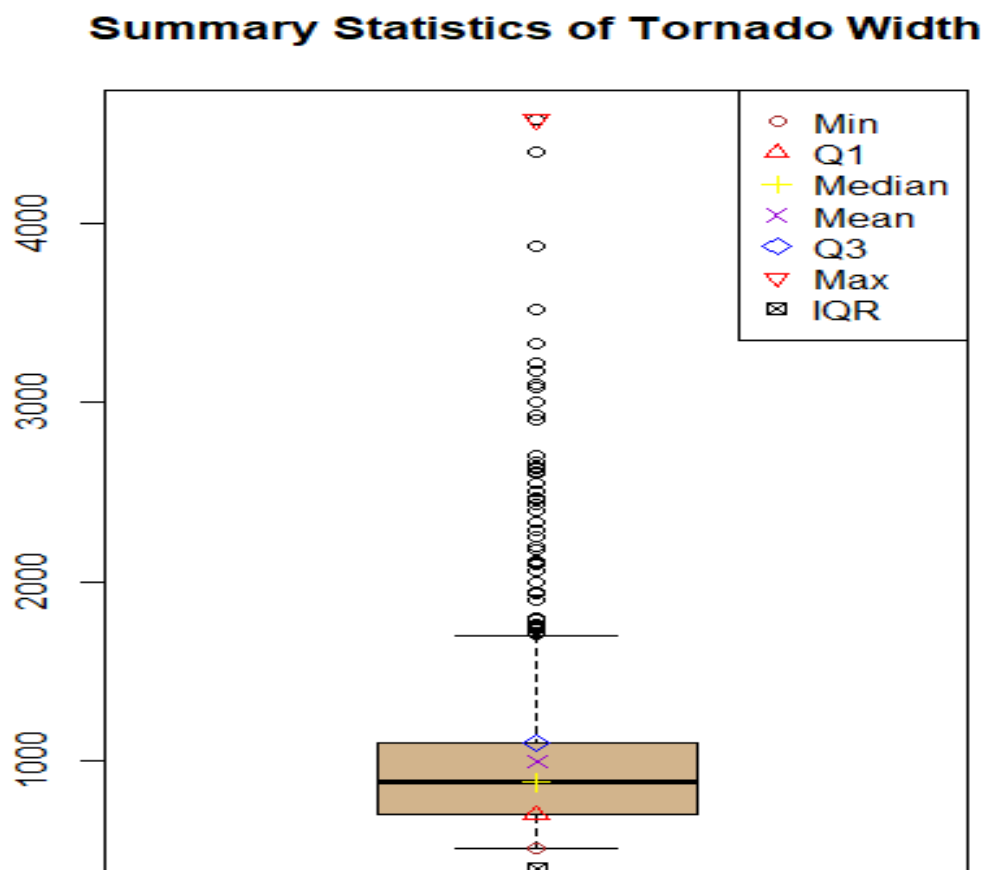
```
df1wid
```

```

boxplot(df1wid$wid,col='tan',main="Summary Statistics of Tornado Width")
range(df1wid$wid)
summary(df1wid$wid)
points(min(df1wid$wid),cex=1.0,pch=1,col='brown')
points(quantile(df1wid$wid,0.25),cex=1.0,pch=2,col="red")
points(median(df1wid$wid),cex=1.0,pch=3,col='yellow')
points(mean(df1wid$wid),cex=1.0,pch=4,col='darkviolet')
points(quantile(df1wid$wid,0.75),cex=1.0,pch=5,col='blue')
points(max(df1wid$wid),cex=1.0,pch=6,col='red')
points(quantile(df1wid$wid,0.75)-quantile(df1wid$wid,0.25),cex=1.0,pch=7,col='black')
legend("topright",c("Min","Q1","Median","Mean","Q3","Max","IQR"),cex=1.0,pch=c(1,2,3,4,5,6,7),
col=c('brown',"red","yellow",'darkviolet','blue','red','black'), text.font=1)

```

BoxPlot for Tornado Width greater than 500:



```

df1len <- subset(df1,subset=len>30)

df1len

boxplot(df1len$len,col='tan',main="Summary Statistics of Tornado Length")

range(df1len$len)

summary(df1len$len)

points(min(df1len$len),cex=1.2,pch=1,col='brown')

points(quantile(df1len$len,0.25),cex=1.2,pch=2,col="red")

points(median(df1len$len),cex=1.2,pch=3,col='yellow')

points(mean(df1len$len),cex=1.2,pch=4,col='darkviolet')

points(quantile(df1len$len,0.75),cex=1.2,pch=5,col='blue')

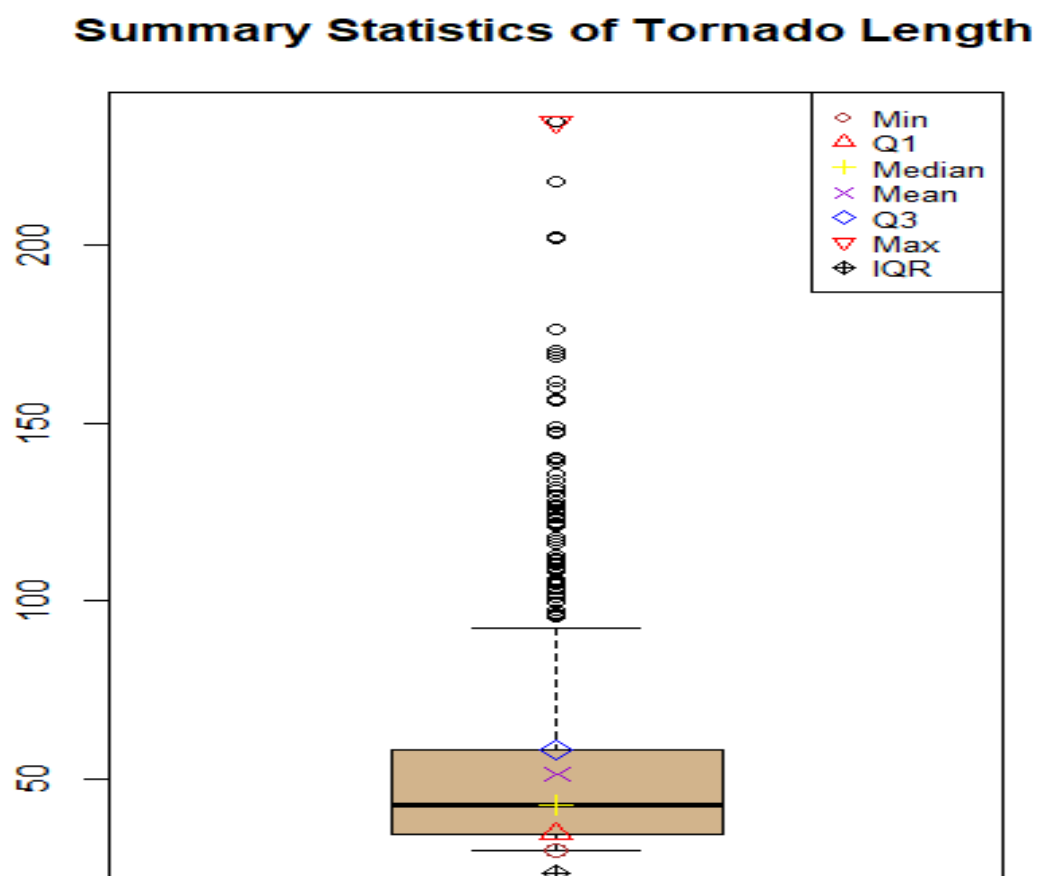
points(max(df1len$len),cex=1.2,pch=6,col='red')

points(quantile(df1len$len,0.75)-quantile(df1len$len,0.25),cex=1.0,pch=9,col='black')

legend("topright",c("Min","Q1","Median","Mean","Q3","Max","IQR"),cex=0.8,pch=c(1,2,3,4,5,6,9),
col=c("brown","red","yellow","darkviolet","blue","red","black"), text.font=1)

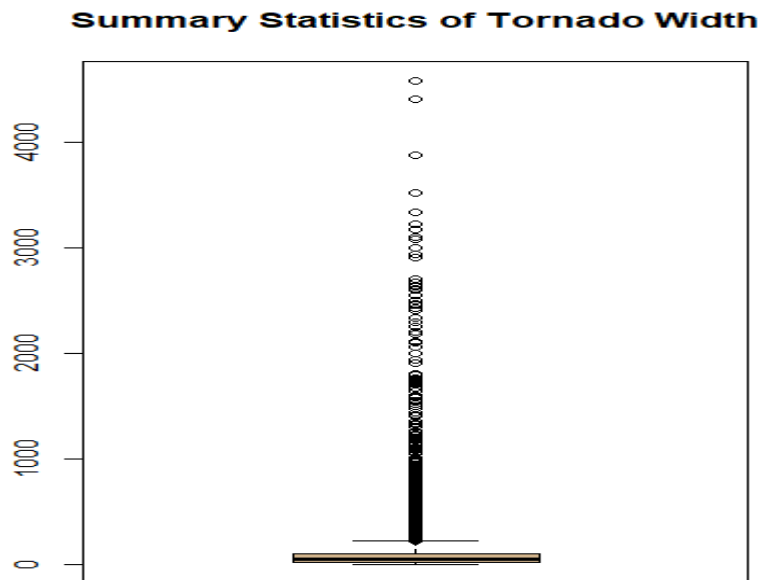
```

BoxPlot for Tornado Length greater than 30:



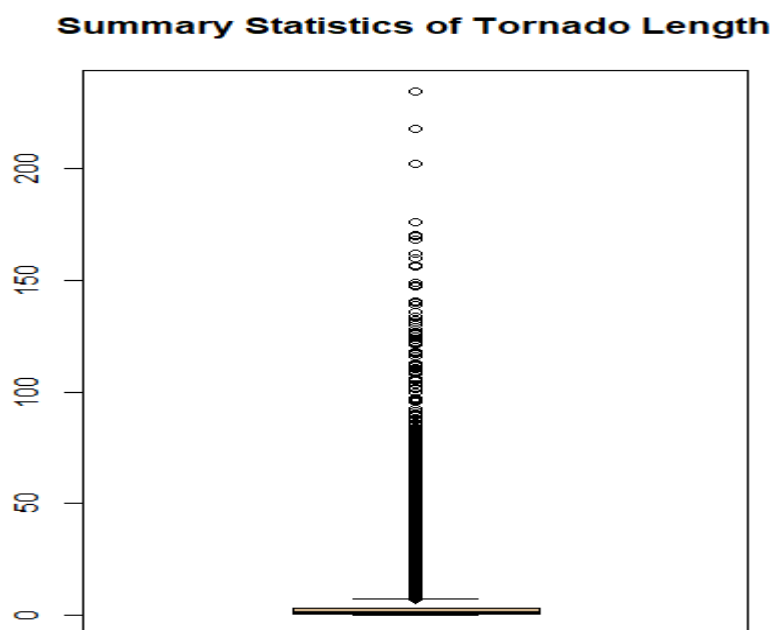
BoxPlot for Tornado Width(entire width to show the range):

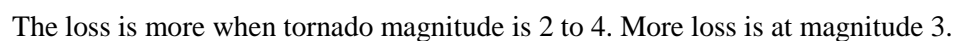
```
boxplot(df1$wid,main="Summary Statistics of Tornado Width",col='tan')
```



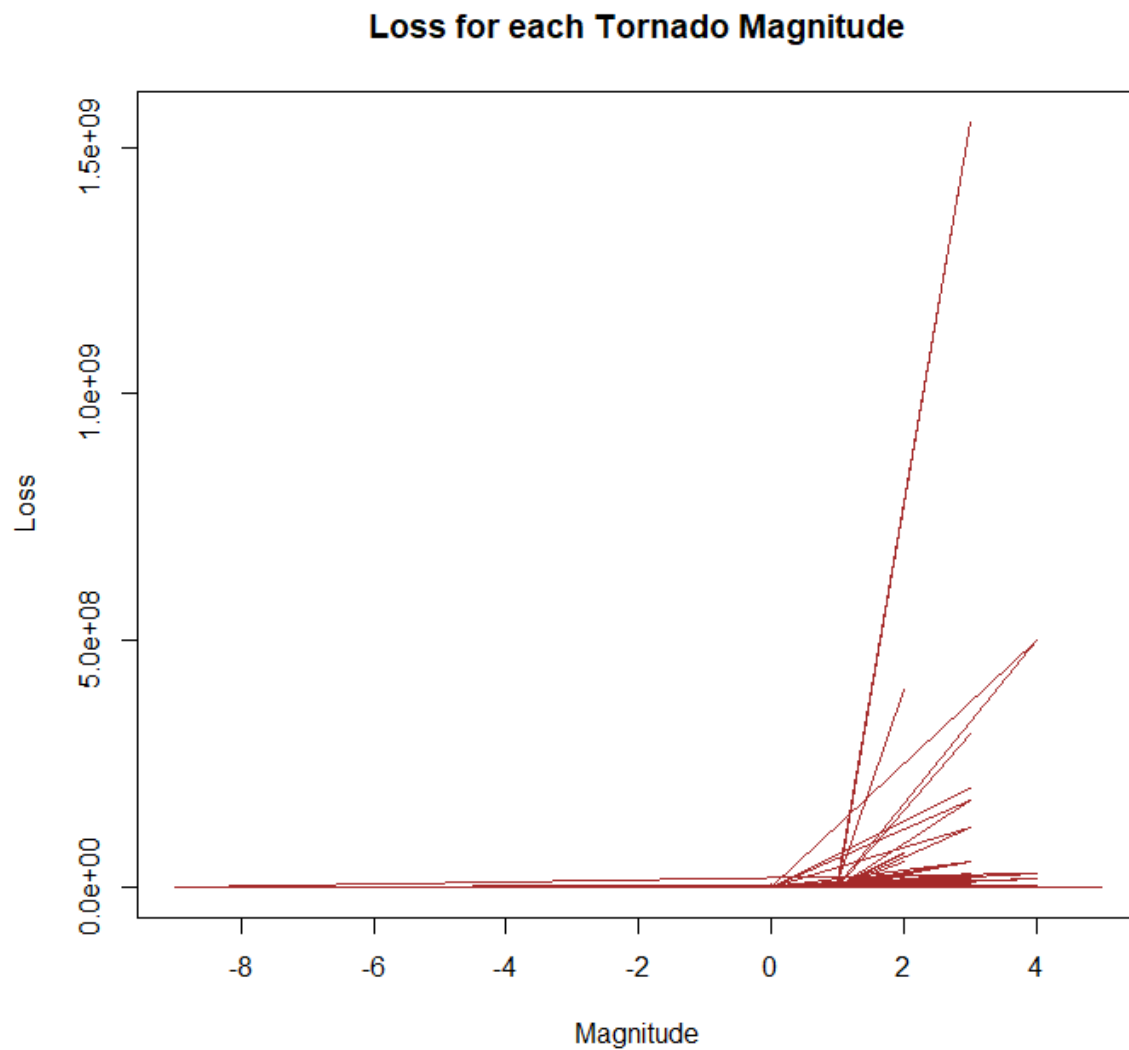
BoxPlot for Tornado Length(entire length to show the range):

```
boxplot(df1$len,main="Summary Statistics of Tornado Length",col='tan')
```





```
plot(Magnitude, Loss, col='brown', main='Loss for each Tornado Magnitude', type='line')
```



- Which 3 states have the highest injuries and fatalities?

```
Injuries <- df1$inj
```

```
States <- df1$st
```

```
Fatalities <- df1$fat
```

```
s <- aggregate(Injuries~States,df1,sum)
```

```
s
```

```
ordering <- s[order(s[, 'Injuries'], decreasing=TRUE),]
```

```
ordering
```

```
top3withhighinj <- ordering[1:3, 'States']
```

```
top3withhighinj
```

```
fat <- aggregate(Fatalities~States,df1,sum)
```

```
fat
```

```
ordering <- fat[order(fat[, 'Fatalities'], decreasing = TRUE),]
```

```
ordering
```

```
top3withhighfat <- ordering[1:3, 'States']
```

```
top3withhighfat
```

```
> top3withhighfat
[1] "AL" "TX" "MS"
> top3withhighinj
[1] "TX" "AL" "MS"
```

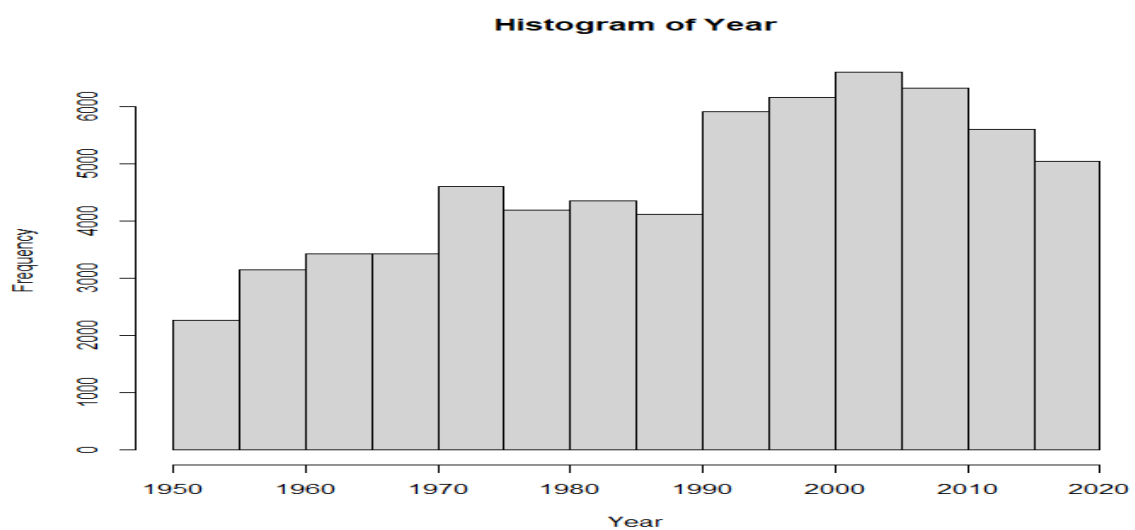
Top three states with highest fatalities: Alabama, Texas, Mississippi

Top three states with highest injuries: Texas, Alabama, Mississippi

- Is there a trend in the number of tornados from 1950 through 2019?

```
Year <- df1$Yr
```

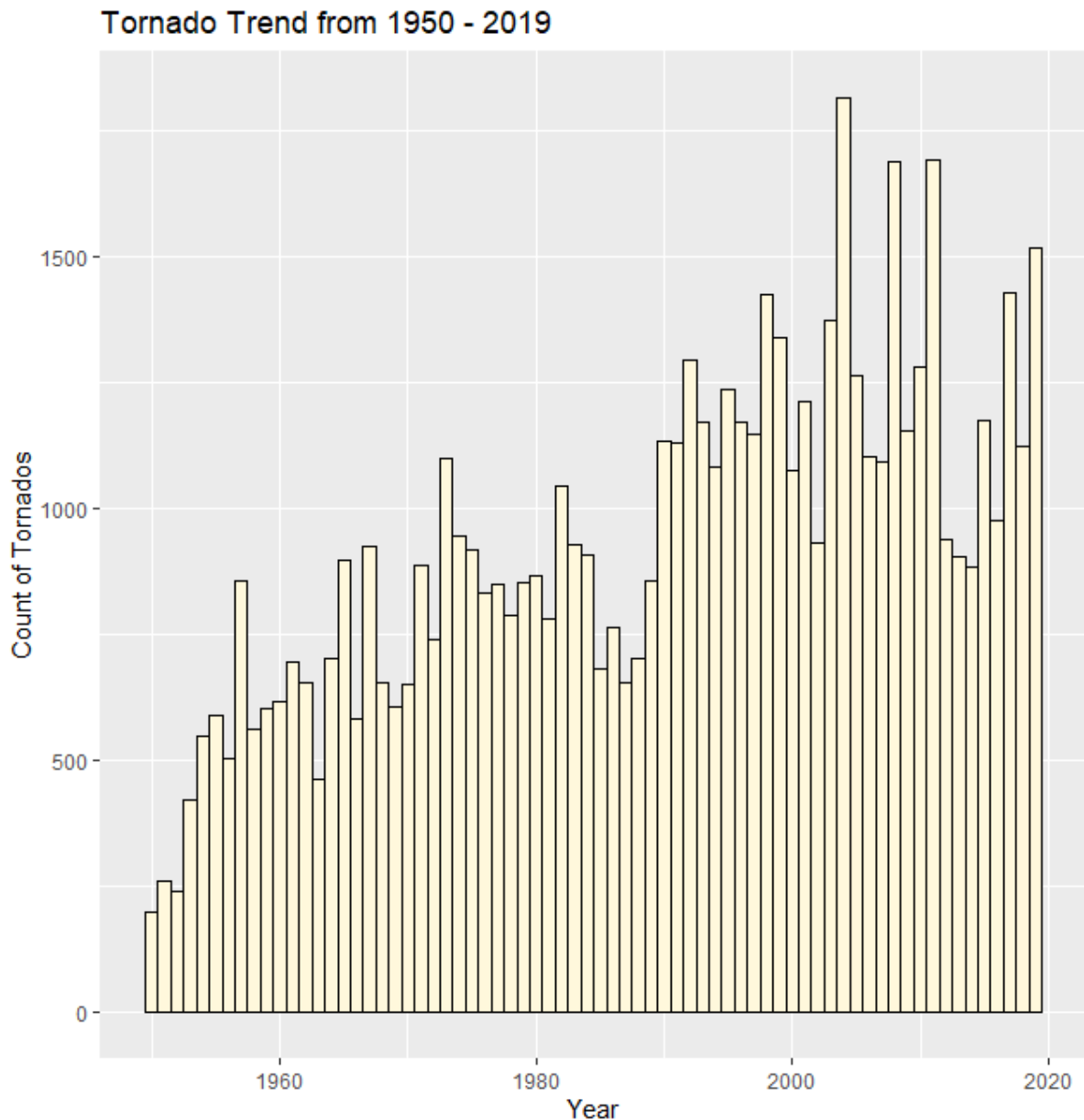
```
hist(Year)
```



```
p <- ggplot(data=df1) +geom_histogram(mapping=aes(x = yr),
binwidth=1,fill="cornsilk",color="black") +

labs(x = "Year", y = "Count of Tornadoes",title = "Tornado Trend from 1950 - 2019")
```

p



According to the histogram above, in an overall view, there is a positive correlation between the number of tornadoes occurring and the year in which they occurred. But if we get into detail, from 1950 to 1973, there was an overall rise in number of tornadoes occurring (peak 1), after that there is a decrease from 1974 to 1989. Then again, there was a rise till 2004 (peak 2) and then declined. If we observe, this is a bimodal distribution, it has two peaks (modes) in the years 1973 and 2004. The trend is not constant, it is changing now and then from positive correlation to negative correlation.