- 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)

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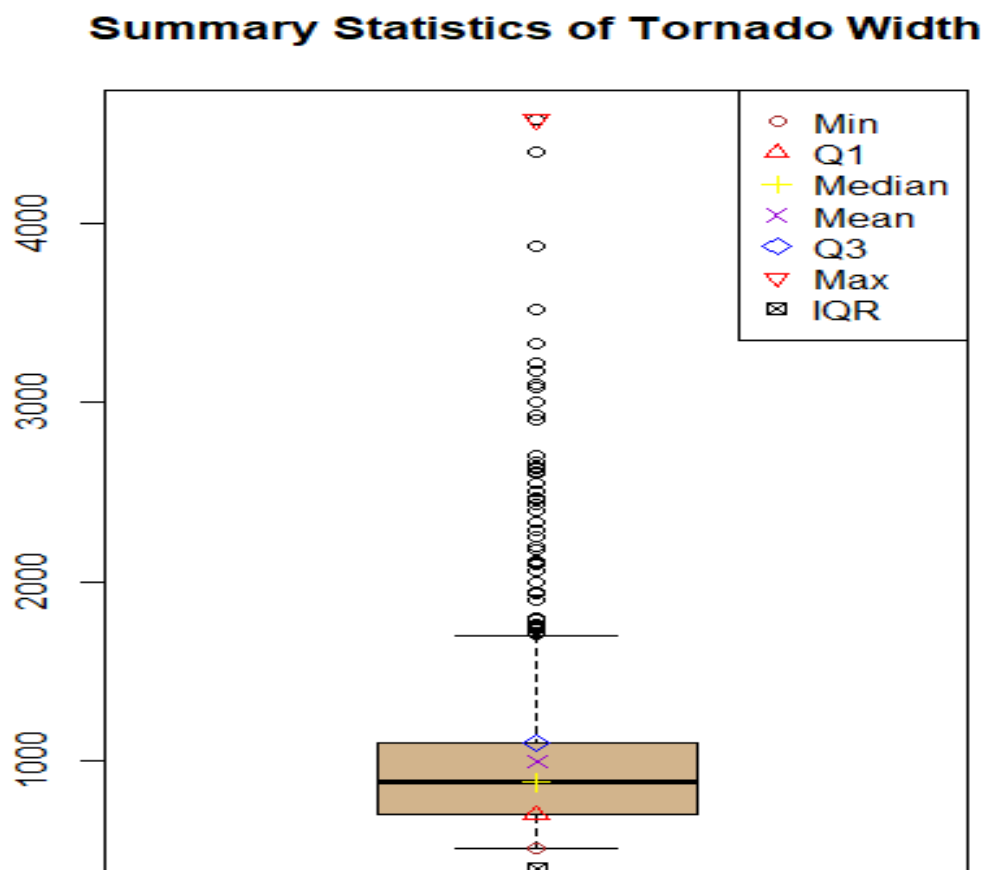
df1len <- subset(df1,subset=len>30)

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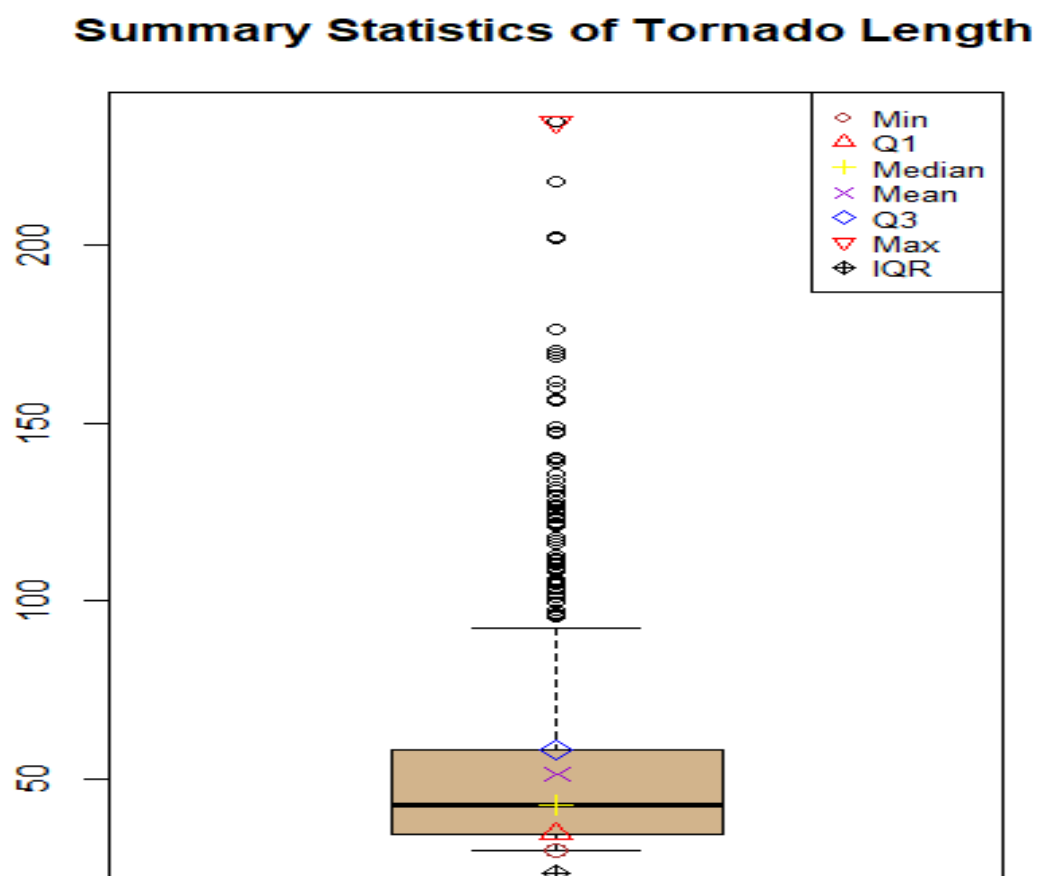
BoxPlot for Tornado Width(entire width to show the range):

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')

- o Create a table (crosstab) showing the mean number of injuries and fatalities for each tornado magnitude

cro_mean(df1$inj,df1$mag)

cro_mean(df1$fat,df1$mag)

```
> cro_mean(df1$inj,df1$mag)

|           | df1$mag |   |     |     |     |    |       |
|           |      -9 | 0 |   1 |   2 |   3 |  4 |     5 |
| --------- | ------- | --| --- | --- | --- | --| ----- |
| df1$inj   |       0 | 0 | 0.3 | 1.7 | 9.5 | 61 | 219.6 |
> cro_mean(df1$fat,df1$mag)

|           | df1$mag |   |     |     |     |     |      |
|           |      -9 | 0 |   1 |   2 |   3 |   4 |    5 |
| --------- | ------- | --| -- | --- | --- | --- | ---- |
| df1$fat   |       0 | 0 |  0 | 0.1 | 0.5 | 4.2 | 22.8 |
> |
```
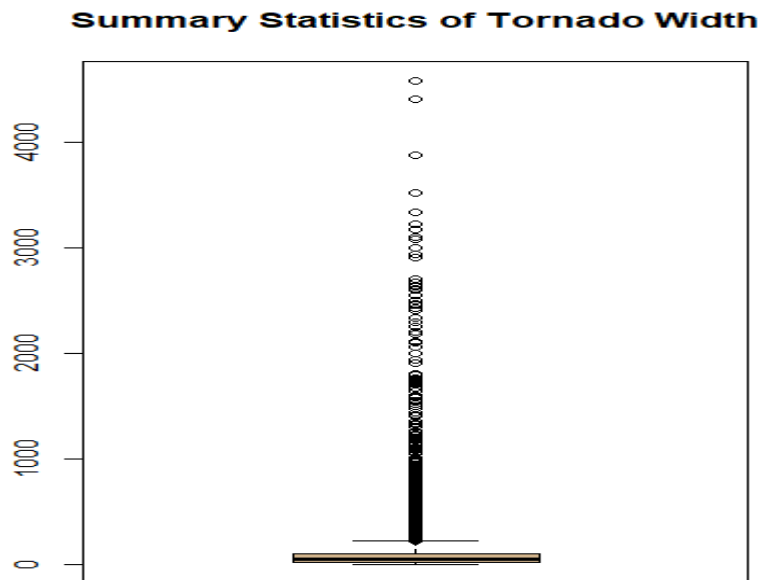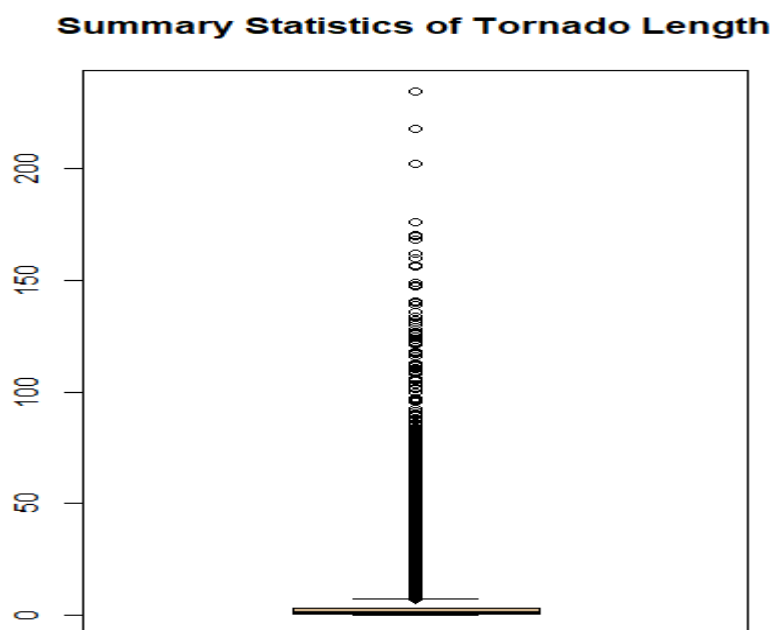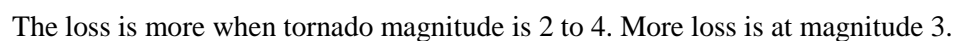
- o Plot the loss for each tornado magnitude

Magnitude <- df1$mag

Loss <- df1$loss

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

The loss is more when tornado magnitude is 2 to 4. More loss is at magnitude 3.

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

Loss for each Tornado Magnitude
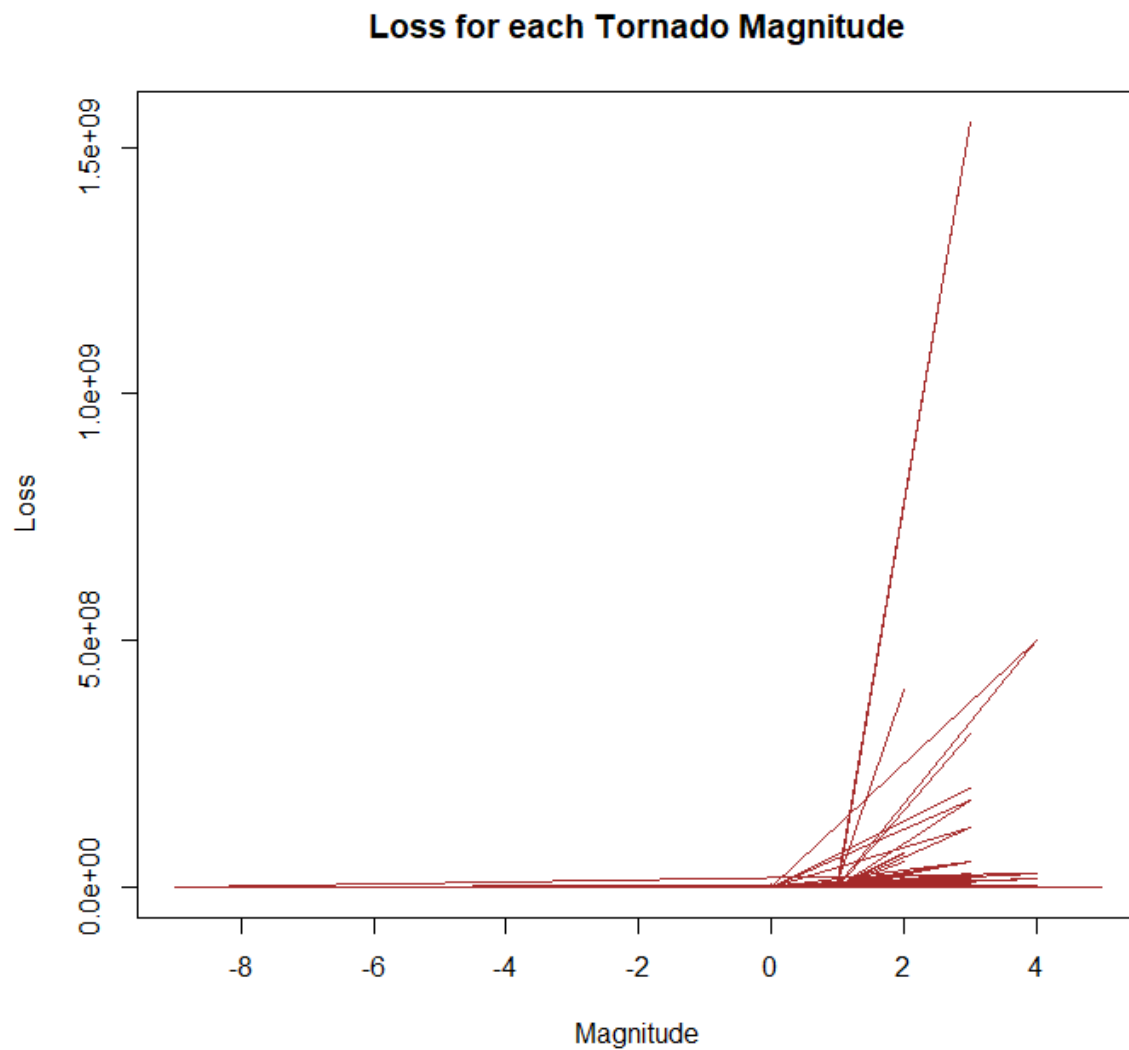
     o   Which 3 states have the highest injuries and fatalities?

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"
>
```
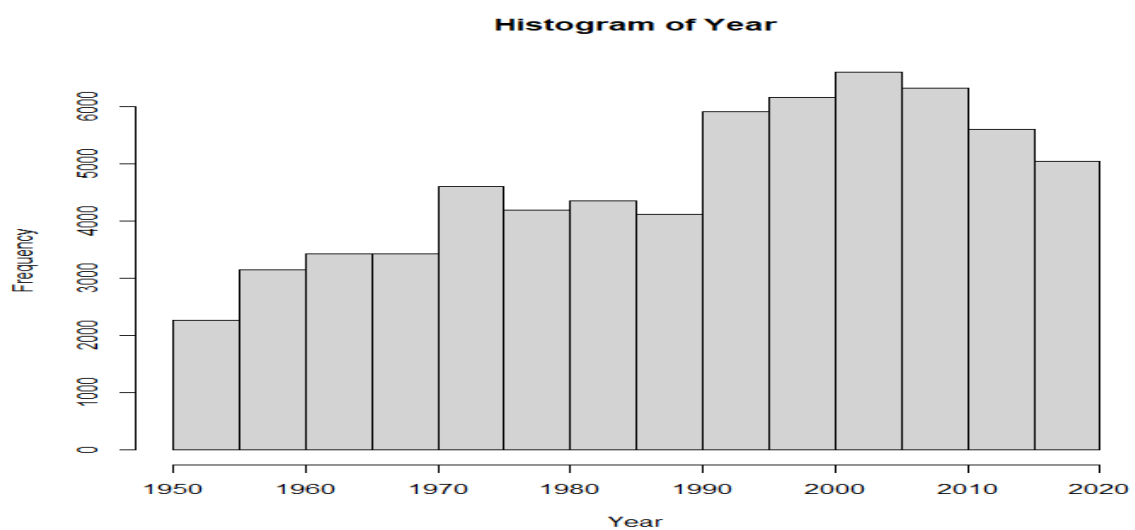
Top three states with highest fatalities: Alabama, Texas, Mississippi

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

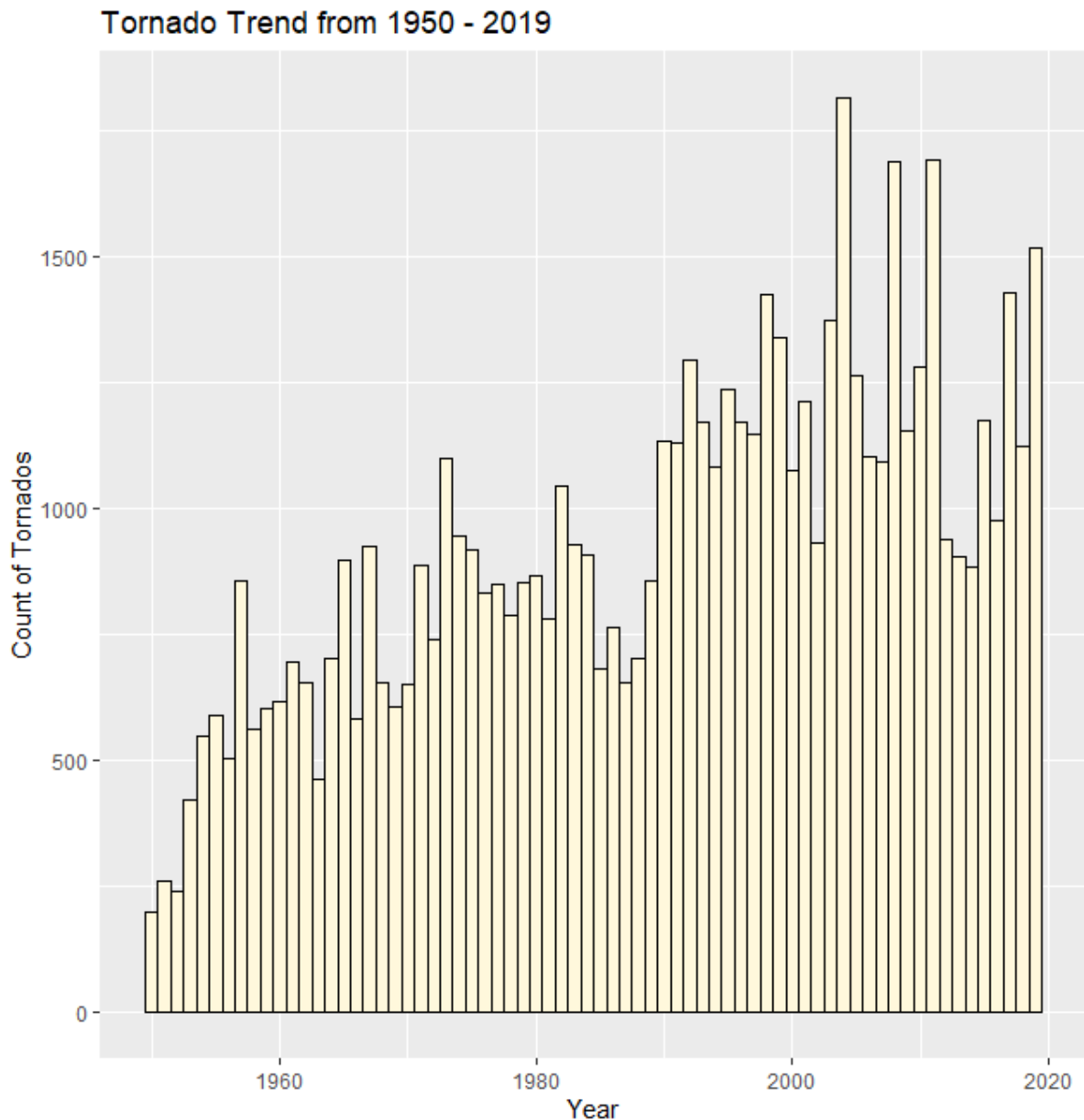o Is there a trend in the number of tornados from 1950 through 2019?

```
Year <- df1$yr
hist(Year)
```


Histogram of Year

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

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

p
```


Tornado Trend from 1950 - 2019

According to the histogram above, in an overall view, there is a positive correlation between the
number of tornados 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 tornados 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.