D. TENISHA

1st MSC DATA ANALYTICS

DATA ANALYTICS WITH R

Data Collection:

I collected the Tamil Nadu district wise Covid-19 data for March and April month Data (45 days). I selected two districts namely Coimbatore and Erode.

Coding and Output:

```
getwd()
```

```
## [1] "C:/Users/Tenisha/Documents"
```

```
library(xlsx)
```

```
covid_coimbatore<-read.xlsx(file.choose(),1,header = T)
View(covid_coimbatore)
```

```
covid_erode<-read.xlsx(file.choose(),1,header = T)
View(covid_erode)
```

```
head(covid_coimbatore)
```

```
##         Date  Districts Total.Positive.Cases Discharged.Cases
Active.Cases
## 1 2021-03-01 Coimbatore                55799            54737
379
## 2 2021-03-02 Coimbatore                55838            54776
379
## 3 2021-03-03 Coimbatore                55883            54820
380
## 4 2021-03-04 Coimbatore                55931            54871
377
## 5 2021-03-05 Coimbatore                55982            54923
376
## 6 2021-03-06 Coimbatore                56030            54982
364
##   Death.Cases
## 1         683
## 2         683
## 3         683
## 4         683
```

```
## 5         683
## 6         684
```

```
tail(covid_coimbatore)
```

```
##          Date  Districts Total.Positive.Cases Discharged.Cases
Active.Cases
## 40 2021-04-09 Coimbatore                62070            57982
3389
## 41 2021-04-10 Coimbatore                62575            58199
3677
## 42 2021-04-11 Coimbatore                63197            58447
4051
## 43 2021-04-12 Coimbatore                63808            58730
4378
## 44 2021-04-13 Coimbatore                64319            59099
4520
## 45 2021-04-14 Coimbatore                64867            59620
4544
##    Death.Cases
## 40         699
## 41         699
## 42         699
## 43         700
## 44         700
## 45         703
```

```
summary(covid_coimbatore)
```

```
##       Date             Districts         Total.Positive.Cases
Discharged.Cases
##  Min.   :2021-03-01   Length:45          Min.   :55799        Min.
:54737
##  1st Qu.:2021-03-12   Class :character   1st Qu.:56349        1st
Qu.:55297
##  Median :2021-03-23   Mode  :character   Median :57413
Median :55900
##  Mean   :2021-03-23                      Mean   :58409        Mean
:56306
##  3rd Qu.:2021-04-03                      3rd Qu.:60153        3rd
Qu.:57266
##  Max.   :2021-04-14                      Max.   :64867        Max.
:59620
##   Active.Cases   Death.Cases
##  Min.   : 352   Min.   :683.0
##  1st Qu.: 379   1st Qu.:685.0
##  Median : 825   Median :688.0
```

```
##  Mean   :1415   Mean   :689.7
##  3rd Qu.:2193   3rd Qu.:694.0
##  Max.   :4544   Max.   :703.0

head(covid_erode)

##         Date Districts Total.Positive.Cases Discharged.Cases
Active.Cases.
## 1 2021-03-01     Erode                14802            14558
94
## 2 2021-03-02     Erode                14815            14571
94
## 3 2021-03-03     Erode                14826            14588
88
## 4 2021-03-04     Erode                14837            14606
81
## 5 2021-03-05     Erode                14847            14622
75
## 6 2021-03-06     Erode                14860            14635
75
##   Death.Cases
## 1         150
## 2         150
## 3         150
## 4         150
## 5         150
## 6         150

tail(covid_erode)

##          Date Districts Total.Positive.Cases Discharged.Cases
Active.Cases.
## 40 2021-04-09     Erode                15750            15227
373
## 41 2021-04-10     Erode                15831            15236
445
## 42 2021-04-11     Erode                15940            15273
517
## 43 2021-04-12     Erode                16049            15292
607
## 44 2021-04-13     Erode                16136            15361
624
## 45 2021-04-14     Erode                16293            15435
707
##    Death.Cases
## 40         150
## 41         150
```

```
## 42          150
## 43          150
## 44          151
## 45          151
```

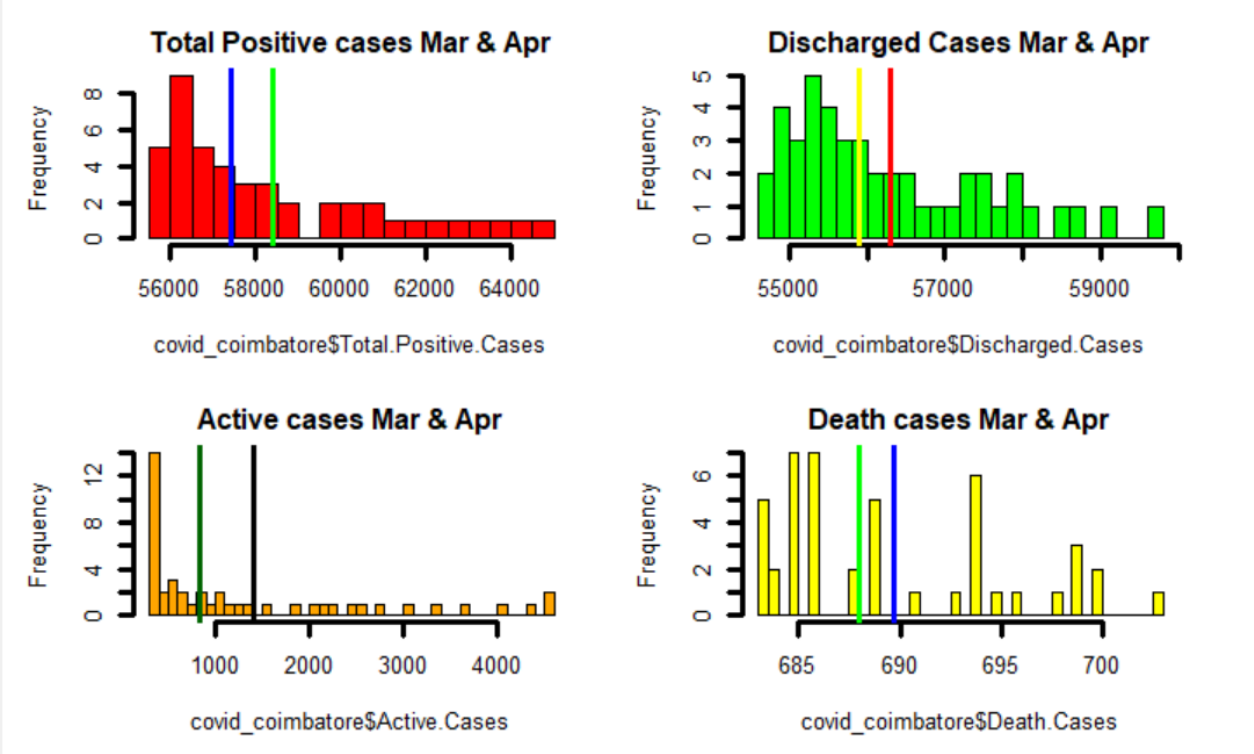
```
summary(covid_erode)
```

```
##       Date              Districts         Total.Positive.Cases
Discharged.Cases
##  Min.   :2021-03-01   Length:45          Min.   :14802        Min.
:14558
##  1st Qu.:2021-03-12   Class :character   1st Qu.:14939        1st
Qu.:14696
##  Median :2021-03-23   Mode  :character   Median :15112
Median :14821
##  Mean   :2021-03-23                      Mean   :15244        Mean
:14893
##  3rd Qu.:2021-04-03                      3rd Qu.:15501        3rd
Qu.:15102
##  Max.   :2021-04-14                      Max.   :16293        Max.
:15435
##  Active.Cases.    Death.Cases
##  Min.   : 75.0   Min.   :150
##  1st Qu.: 94.0   1st Qu.:150
##  Median :141.0   Median :150
##  Mean   :200.4   Mean   :150
##  3rd Qu.:238.0   3rd Qu.:150
##  Max.   :707.0   Max.   :151
```

Histogram for Coimbatore Data:

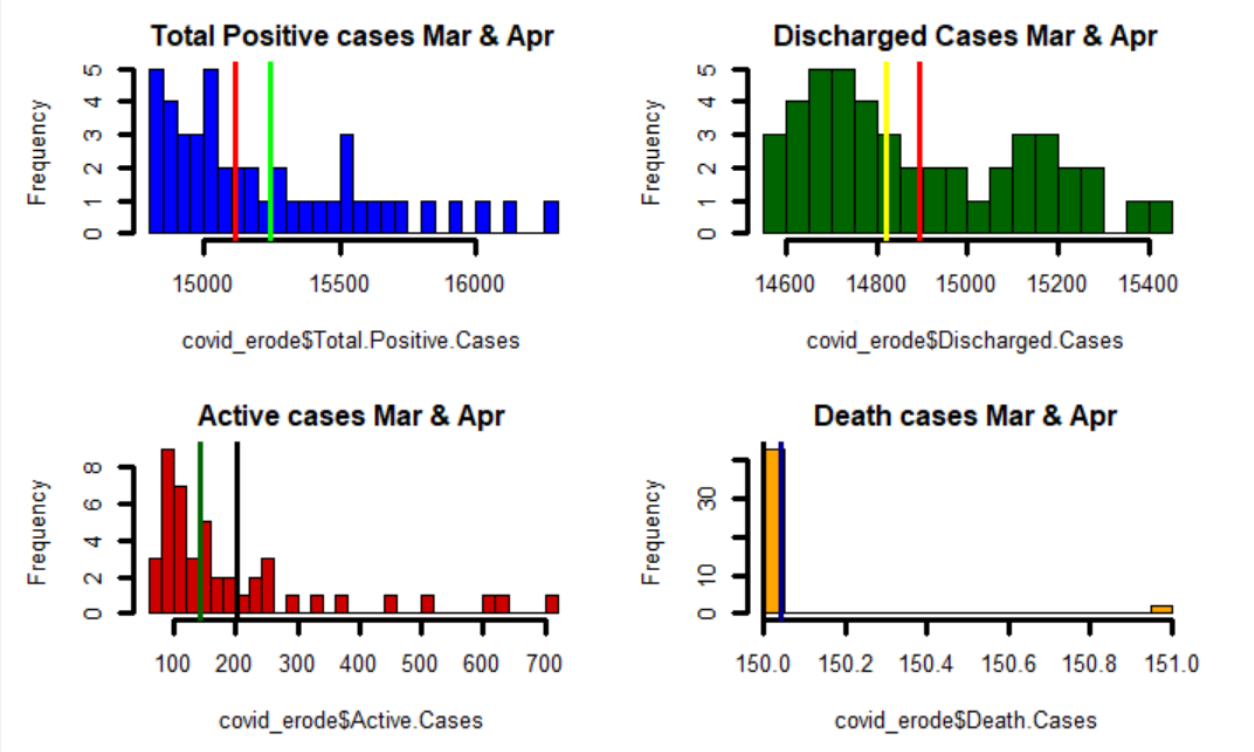
```
attach(covid_coimbatore)
par(mfrow=c(2,2))
hist(covid_coimbatore$Total.Positive.Cases,main = "Total Positive
cases in Coimbatore for March and April",breaks=30,col = "red",lwd=3)
abline(v=mean(covid_coimbatore$Total.Positive.Cases),col = "green",lwd
= 3)
abline(v=median(covid_coimbatore$Total.Positive.Cases),col =
"blue",lwd = 3)
#histogram for total discharged cases
hist(covid_coimbatore$Discharged.Cases,main = "Discharged Cases in
Coimbatore for March and April",breaks=30,col = "green",lwd=3)
abline(v=mean(covid_coimbatore$Discharged.Cases),col = "red",lwd = 3)
abline(v=median(covid_coimbatore$Discharged.Cases),col = "yellow",lwd
= 3)
#histogram for total active cases
hist(covid_coimbatore$Active.Cases,main = "Active cases in Coimbatore
```

```
for March and April",breaks=30,col = "orange",lwd=3)
abline(v=mean(covid_coimbatore$Active.Cases),col = "black",lwd = 3)
abline(v=median(covid_coimbatore$Active.Cases),col = "dark green",lwd
= 3)
#histogram for death cases
hist(covid_coimbatore$Death.Cases,main = "Death cases in Coimbatore
for March and April",breaks=30,col = "yellow",lwd=3)
abline(v=mean(covid_coimbatore$Death.Cases),col = "blue",lwd = 3)
abline(v=median(covid_coimbatore$Death.Cases),col = "green",lwd = 3)
```

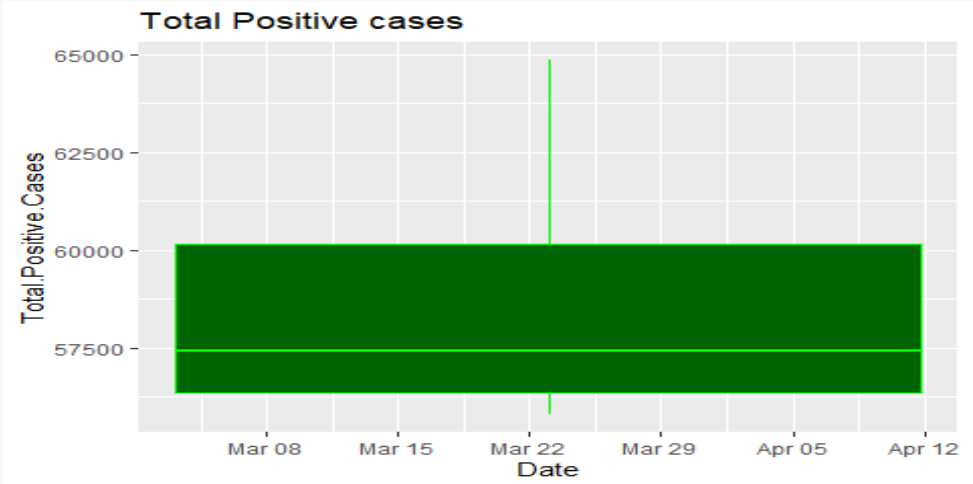



```
 ##histogram for erode data
attach(covid_erode)

## The following objects are masked from covid_coimbatore:
##
##     Date, Death.Cases, Discharged.Cases, Districts,
##     Total.Positive.Cases

par(mfrow=c(2,2))
#histogram for total positive cases
hist(covid_erode$Total.Positive.Cases,main = "Total Positive cases in
Erode for March and April",breaks=30,col = "blue",lwd=3)
abline(v=mean(covid_erode$Total.Positive.Cases),col = "green",lwd = 3)
abline(v=median(covid_erode$Total.Positive.Cases),col = "red",lwd = 3)
#histogram for total discharged cases
hist(covid_erode$Discharged.Cases,main = "Discharged Cases in Erode
```

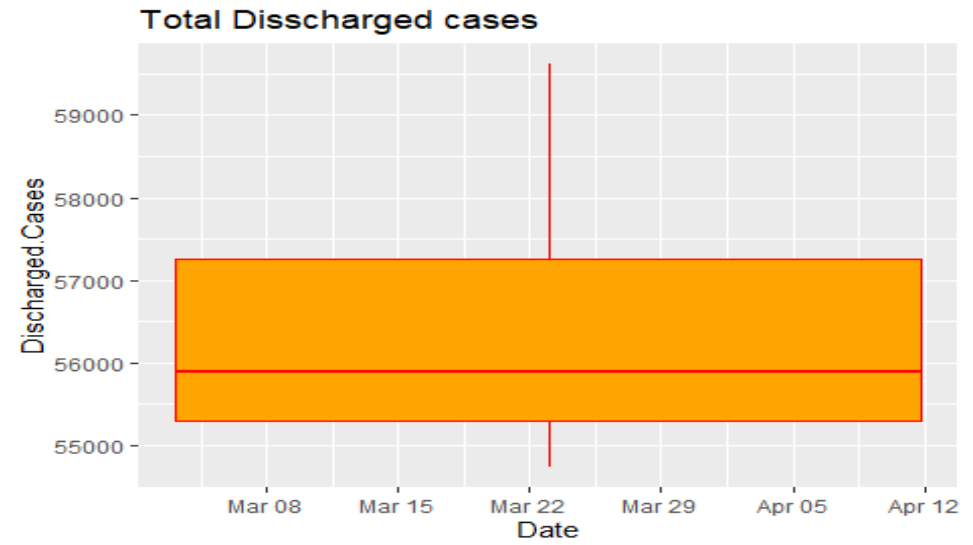
```
for March and April",breaks=30,col = "dark green",lwd=3)
abline(v=mean(covid_erode$Discharged.Cases),col = "red",lwd = 3)
abline(v=median(covid_erode$Discharged.Cases),col = "yellow",lwd = 3)
#histogram for total active cases
hist(covid_erode$Active.Cases,main = "Active cases in erode for March
and April",breaks=30,col = "red3",lwd=3)
abline(v=mean(covid_erode$Active.Cases),col = "black",lwd = 3)
abline(v=median(covid_erode$Active.Cases),col = "dark green",lwd = 3)
#histogram for death cases
hist(covid_erode$Death.Cases,main = "Death cases in erode for March
and April",breaks=30,col = "orange",lwd=3)
abline(v=mean(covid_erode$Death.Cases),col = "blue4",lwd = 3)
abline(v=median(covid_erode$Death.Cases),col = "black",lwd = 3)
```



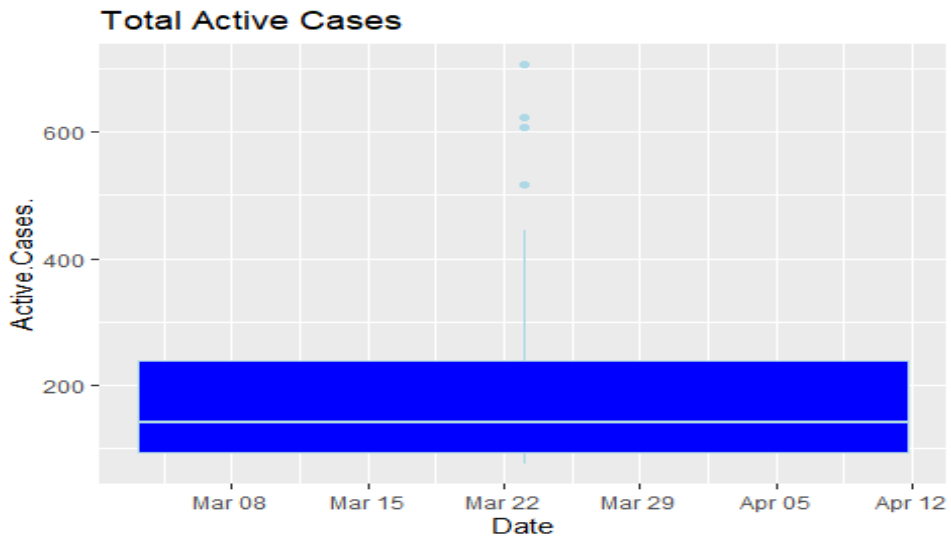

```
##boxplot for Coimbatore data
library(ggplot2)

a<-
ggplot(data=covid_coimbatore,aes(x=Date,y=Total.Positive.Cases))+geom_
boxplot(col="green",fill="dark green")+ggtitle("Total Positive cases")
```



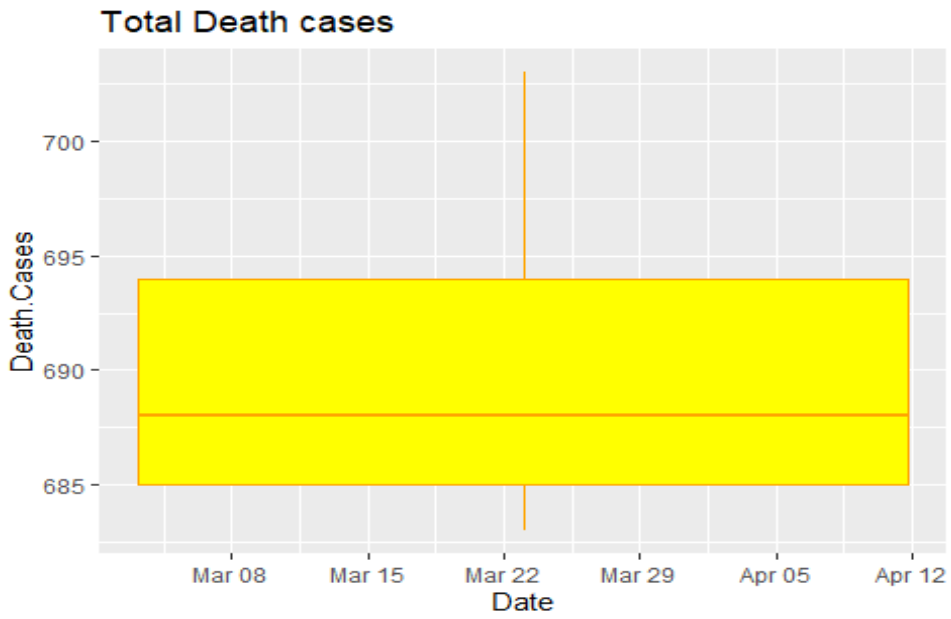

```
b<-
ggplot(data=covid_coimbatore,aes(x=Date,y=Discharged.Cases))+geom_boxp
lot(col="red",fill="orange")+ggtitle("Total Disscharged cases")
b

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```




```
c<-
ggplot(data=covid_coimbatore,aes(x=Date,y=Active.Cases.))+geom_boxplot
(col="light blue",fill="blue")+ggtitle("Total Active Cases")
c

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



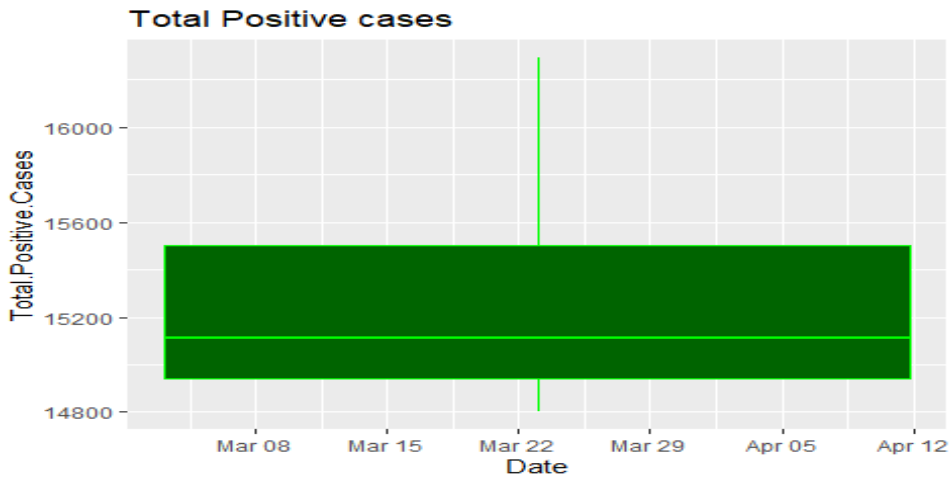

```
d<-
ggplot(data=covid_coimbatore,aes(x=Date,y=Death.Cases))+geom_boxplot(c
ol="orange",fill="yellow")+ggtitle("Total Death cases")
d

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



##boxplot for Erode

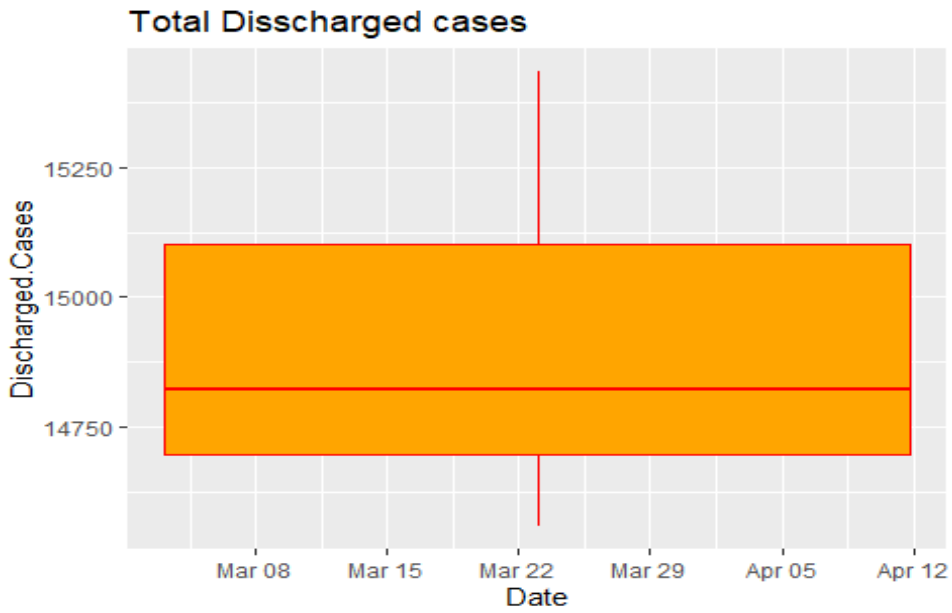
```
e<-
ggplot(data=covid_erode,aes(x=Date,y=Total.Positive.Cases))+geom_boxpl
ot(col="green",fill="dark green")+ggtitle("Total Positive cases")
e
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



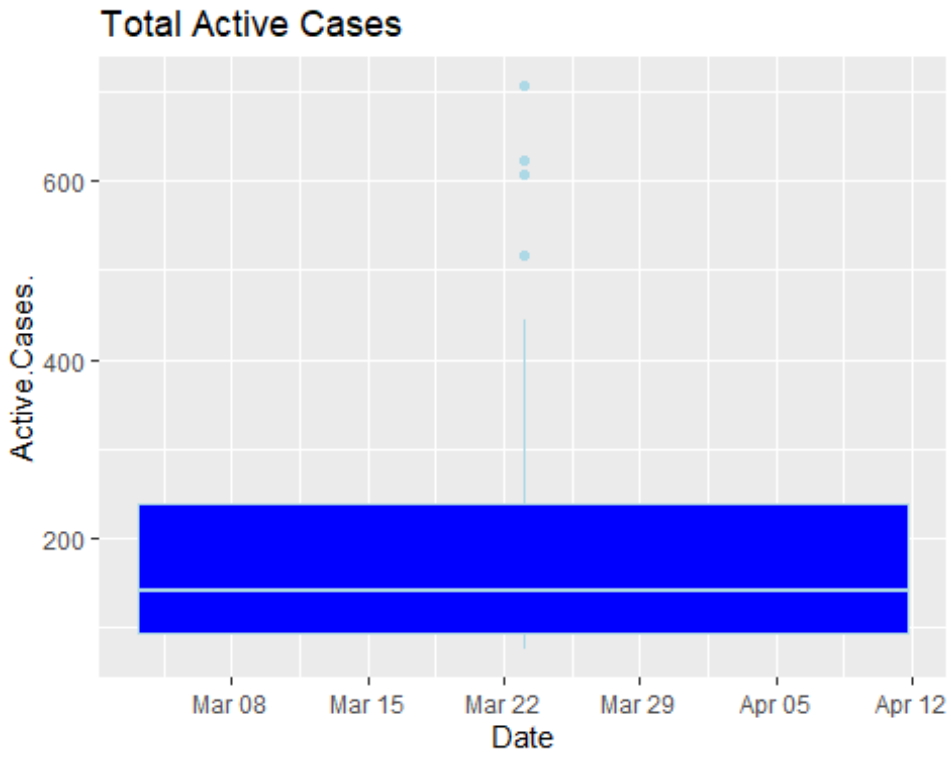

```
f<-
ggplot(data=covid_erode,aes(x=Date,y=Discharged.Cases))+geom_boxplot(c
ol="red",fill="orange")+ggtitle("Total Disscharged cases")
f
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```

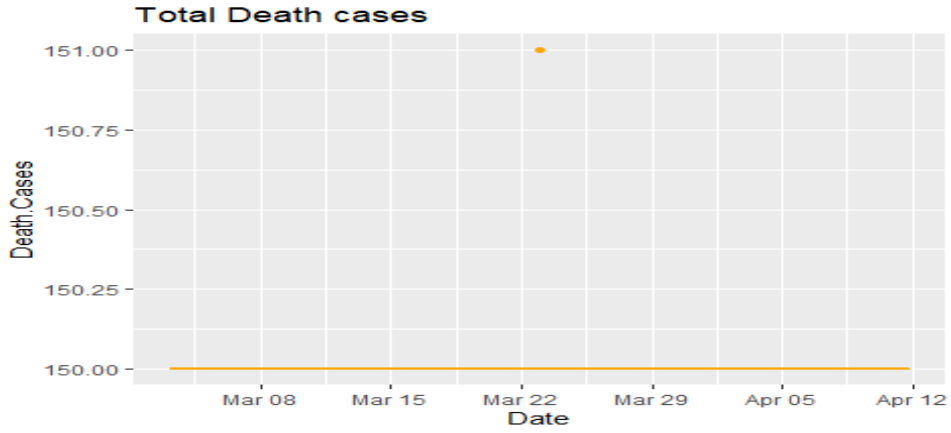



```
g<-
ggplot(data=covid_erode,aes(x=Date,y=Active.Cases.))+geom_boxplot(col=
"light blue",fill="blue")+ggtitle("Total Active Cases")
g
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



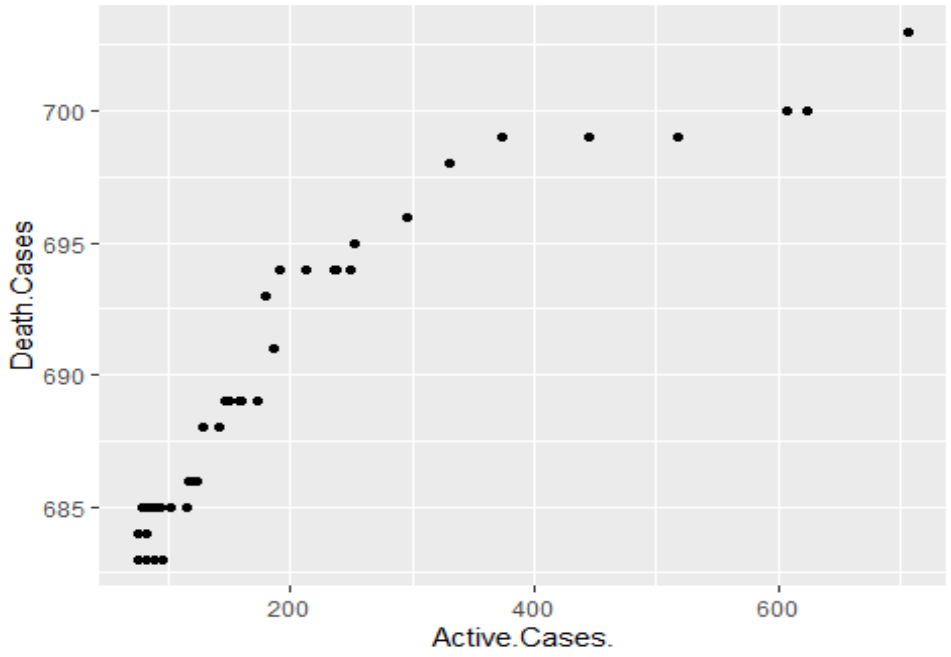

```
h<-
ggplot(data=covid_erode,aes(x=Date,y=Death.Cases))+geom_boxplot(col="o
range",fill="yellow")+ggtitle("Total Death cases")
h

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



##scatter plot for coimbatore

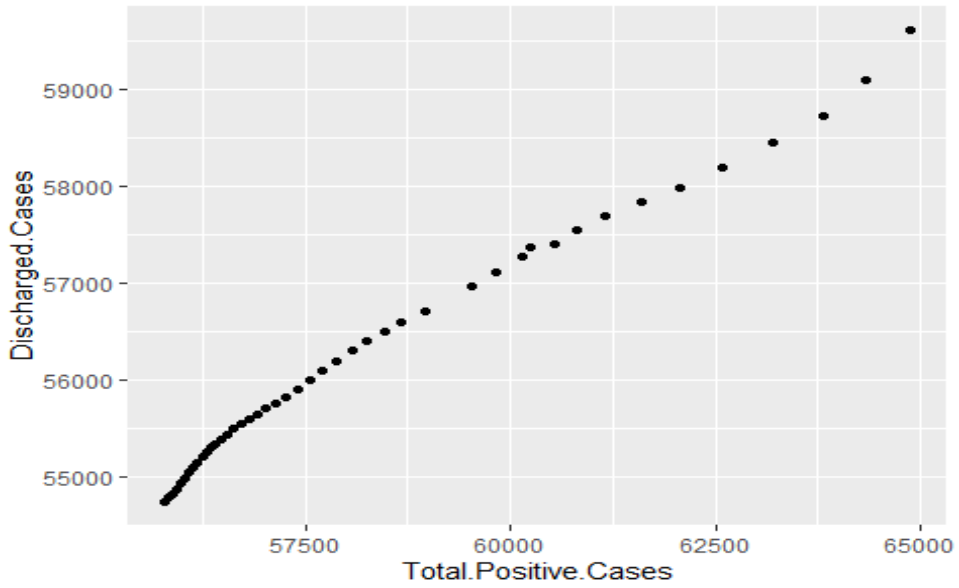
##active cases vs Death cses

```
ggplot(covid_coimbatore,aes(x=Active.Cases.,Death.Cases))+geom_point()
```



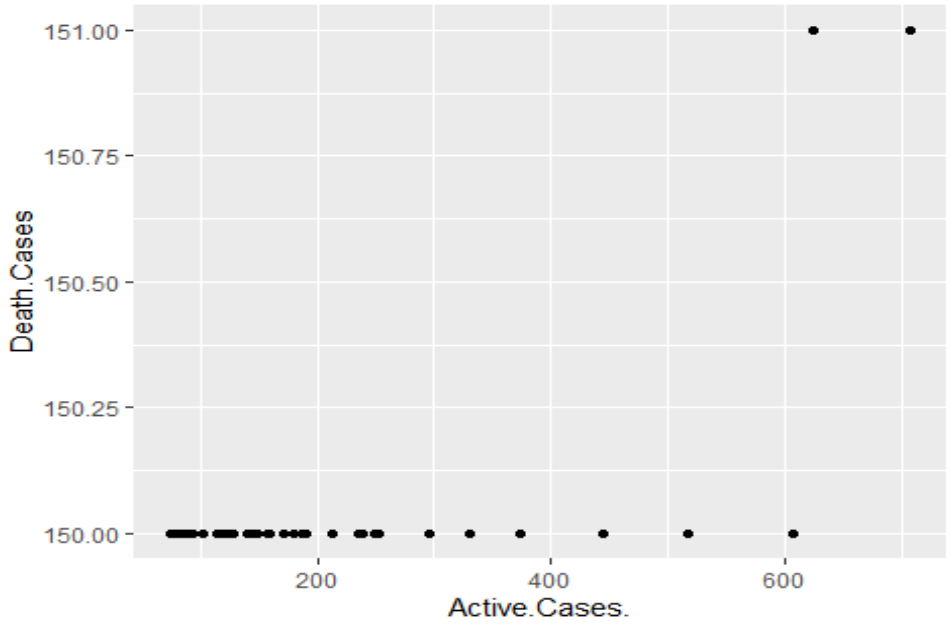
##total positive cases vs Discharged cases

```
ggplot(covid_coimbatore,aes(x=Total.Positive.Cases,Discharged.Cases))+
geom_point()
```



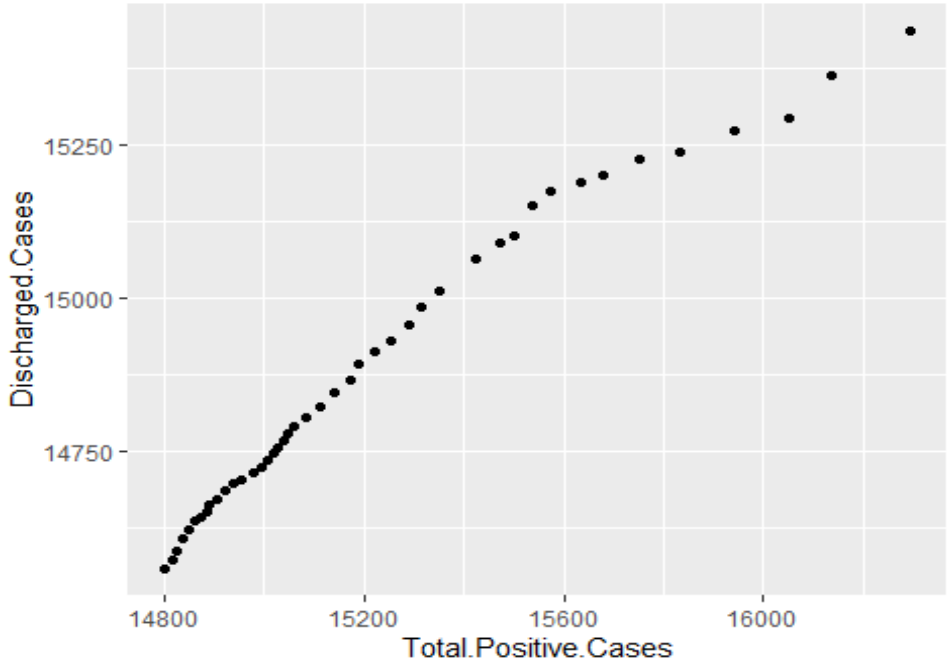
##scatter plot for erode ##active cases vs Death cses

```
ggplot(covid_erode,aes(x=Active.Cases.,Death.Cases))+geom_point()
```



##total positive cases vs Discharged cases

```
ggplot(covid_erode,aes(x=Total.Positive.Cases,Discharged.Cases))+geom_
point()
```

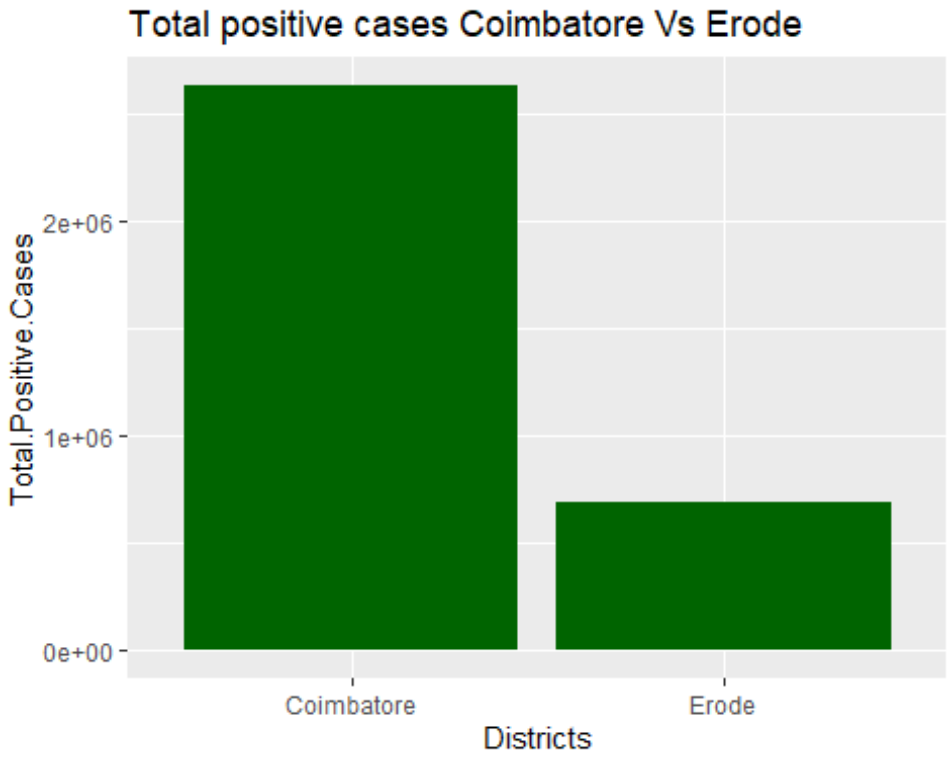


#combining coimbatore and erode datasets

```
colnames(covid_coimbatore)<-colnames(covid_erode)
combined_data<-rbind(covid_coimbatore,covid_erode)
View(combined_data)
```

#coimbatore and erode barplot

```
library(scales)
library(ggplot2)
df3<-data.frame(combined_data)
v<-
ggplot(data=df3,aes(x=Districts,y=Total.Positive.Cases))+geom_bar(stat
="identity",fill="dark green")+ggtitle("Total positive cases
Coimbatore Vs Erode")
v
```



INSIGHTS:

- By comparing the covid data of Coimbatore and Erode, I came to conclusion that Coimbatore has more number of covid cases than the Erode District.
- The histogram of Coimbatore data (i.e) Total Positive cases, Total Discharged cases and total Active cases reveals that they are positively skewed.
- The boxplot shows the quartile, median, min and max values in the dataset.
- The scatter plot shows the relationship between Total positive cases and Discharge cases of both Coimbatore and Erode data. And also relationship between Active cases and death cases.

- The final bar chart shows the vast difference between Coimbatore and Erode covid data.