Technological Institute of Tijuana

Academic Subdirectorate

Systems and Computing Department

SEMESTER: August - December 2021

CAREER: Computer Systems Engineer

MATTER: Data Mining

JOB NAME: Unit 2 - Practice 1

STUDENT NAME AND CONTROL NUMBER:

Castro Cebreros Alejandro - 16211341

Márquez Millán Seashell Vanessa - 17212153

TEACHER NAME: Jose Christian Romero Hernandez

DATE OF DELIVERY: October 29, 2020

1-.Find the interesting dataframe

```
df <- read.csv(file.choose())
```

2-.Read the csv and analyze the data with R

```
df
tail(df)
nrow(df)
head(df)
str(df)
ncol(df)
```

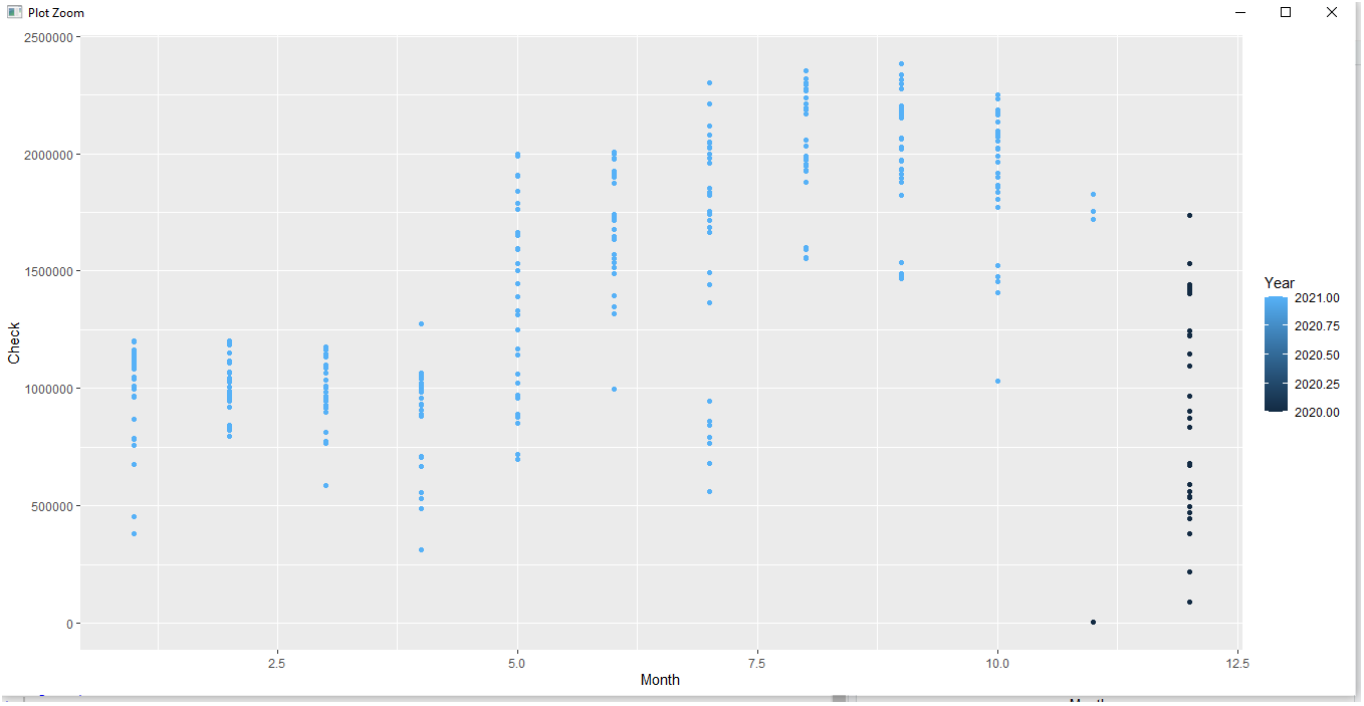
Now we change then columns names because was so large

```
colnames(df) <-
c("Year","Month","Check","OrgAct","UnicMob","CheckCum","OrgAct+","UnicMob+"
,"Delec+","Country")
```

3-.GRAPHICS

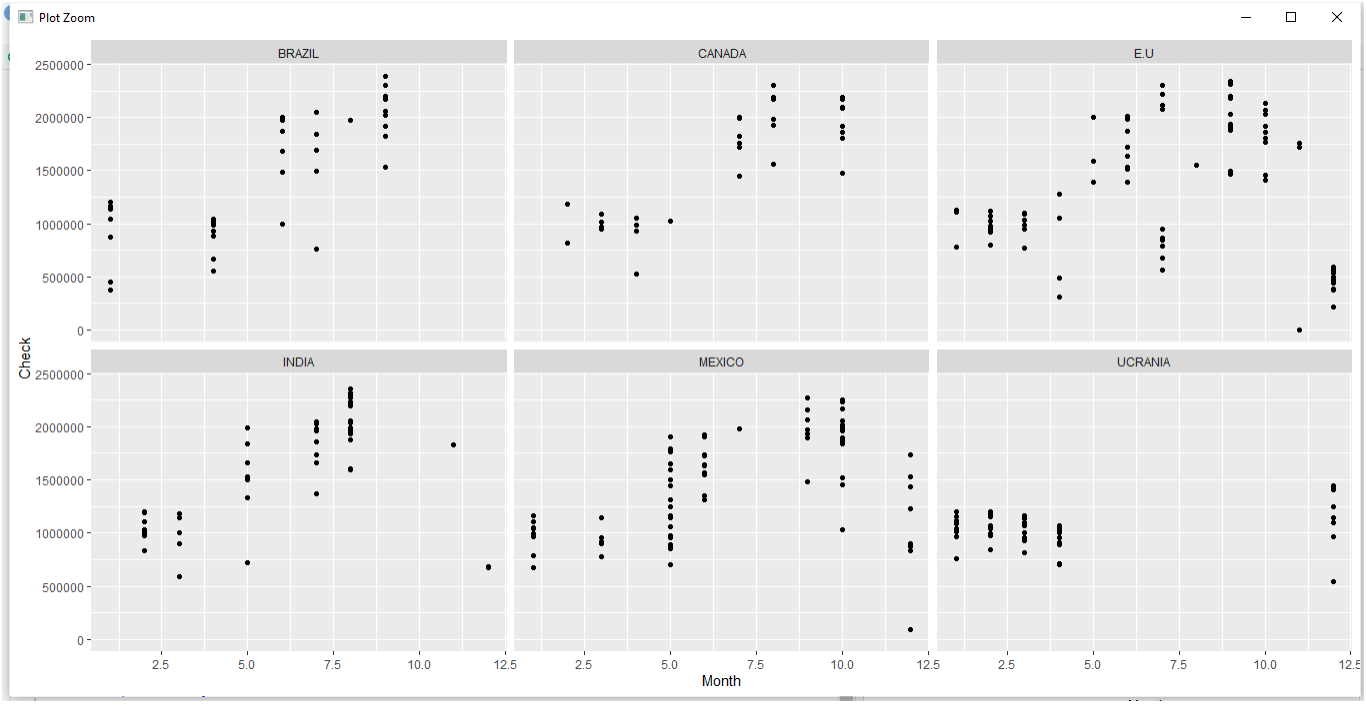
Dispersion grapgic

```
ggplot(df, aes(x = Month, y = Check,
                   color=Year)) +
  geom_point()
```



Facet graphiac

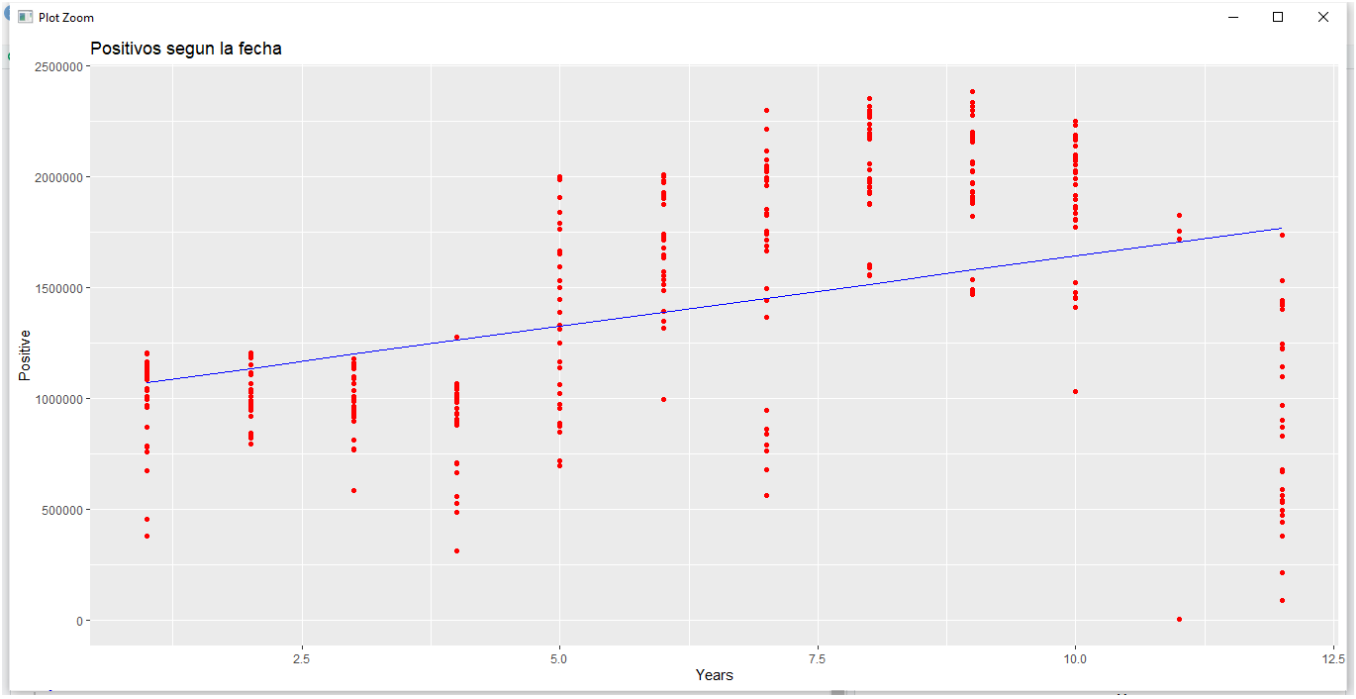
```
ggplot(data = df) +
  geom_point(mapping = aes(x = Month, y = Check)) +
  facet_wrap(~ Country, nrow = 2)
```



Simple regression graph

```
regressor = lm(formula = Check ~ Month,
               data = df)
summary(regressor)

ggplot() +
  geom_point(aes(x=df$Month, y=df$Check),
             color = 'red') +
  geom_line(aes(x = df$Month, y = predict(regressor, newdata = df)),
            color = 'blue') +
  ggtitle('Positivos segun la fecha') +
  xlab('Years') +
  ylab('Positive')
```


Plot Zoom
Positivos segun la fecha
2500000
2000000
1500000
1000000
500000
0
Positive
2.5
5.0
7.5
10.0
12.5
Years