data_science_assignment

Ridam

03/09/2019

Climate Change

Climate change is very big challenge that we need to address in recent years. There has been trends which depicts depletion of oxygen level, water level and rise in global temperature creating a disbalance in the whole eocsystem. Through this study, some exploratory analysis has been done on the data to depict how human interventions are causing threat to ecology.

For this analysis, the data has been collected from NASA and World Bank Data.[1] [2] The datset contains data collected across various countries over a certain time span. WHO dataset contains 264 countries over a time span of almost 20 years. NASA dataset contained data collected over 200 years.

The first dataset which has been analysed is dataset of population density, different plots has been drawn below to show the the trend of the data. The mean and median of the dataset has been created in the initial case.

```
library(ggplot2)
library(gridExtra)
library(plyr)
library(grid)
library(readxl)
library(dplyr)
```

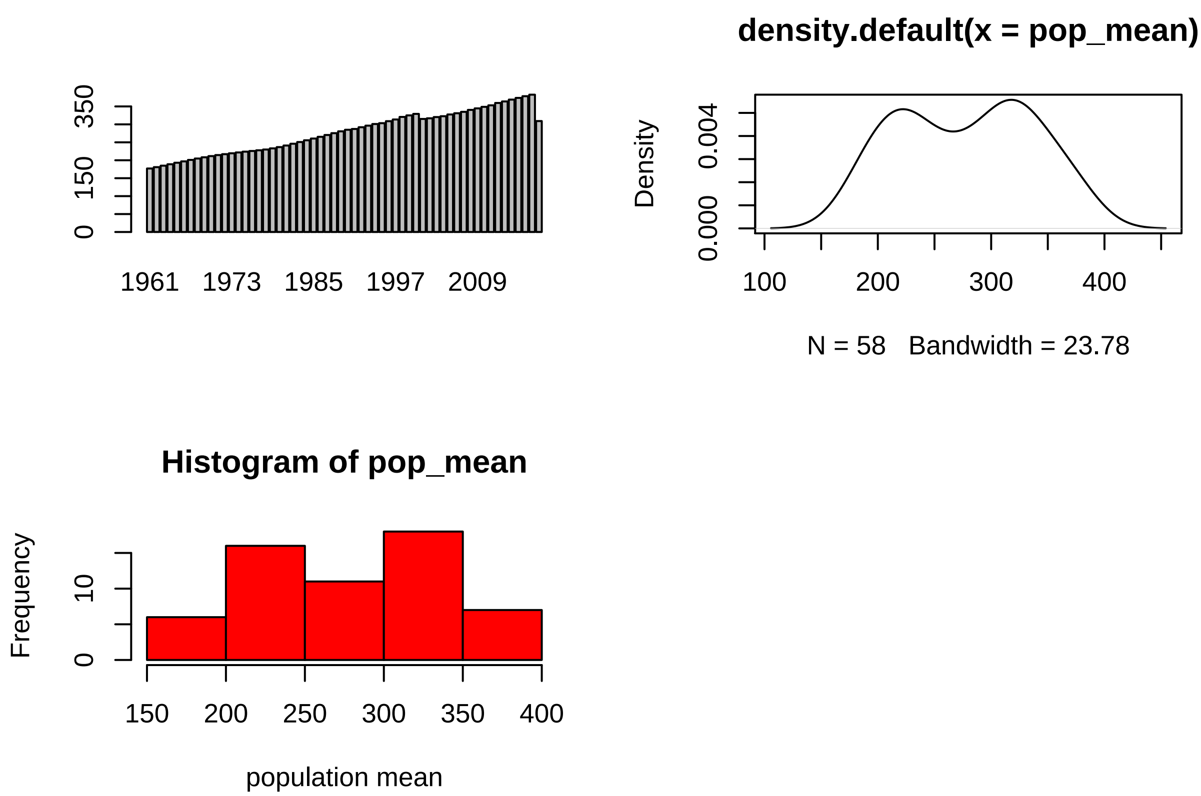
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following object is masked from 'package:gridExtra':
##
##     combine

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(matrixStats)
```

```
##
## Attaching package: 'matrixStats'

## The following object is masked from 'package:dplyr':
##
##     count

## The following object is masked from 'package:plyr':
```

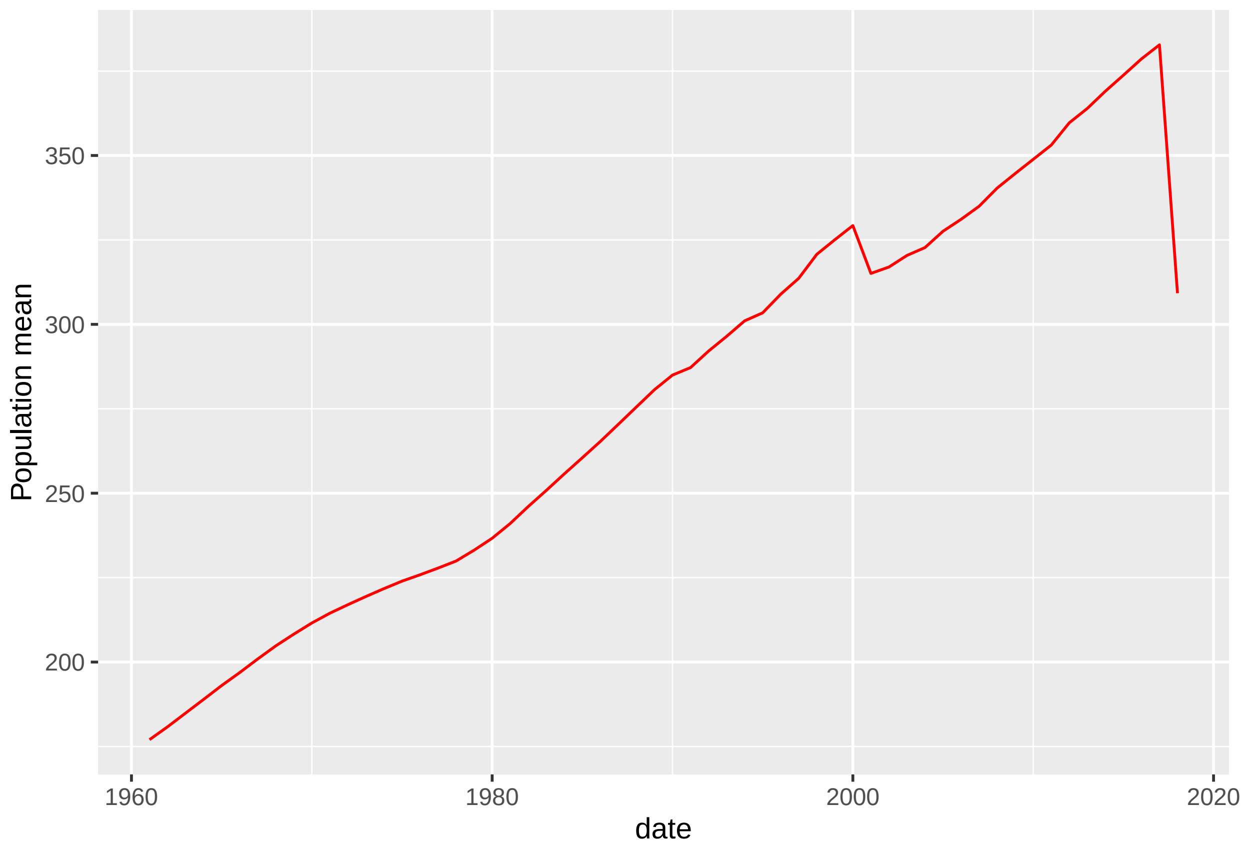
```
##
##     count
```

```r
#reading the excel file
pop_den = read_excel("population_density.xls")

#Since 1960 column doesn't contains any value we will drop this column.
pop_den$'1960' <- NULL

#Finding the mean of all the population and storing it
pop_mean <- colMeans(pop_den[5:62], na.rm=TRUE)

par(mfrow=c(2,2))
#plotting the mean in barplot, density plot and histogram
barplot(pop_mean)
d1 <- density(pop_mean)
plot(d1)
hist(pop_mean, xlab = "population mean", col = "red")
```



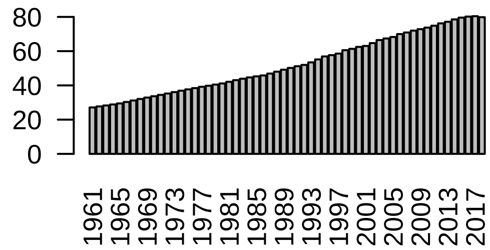
This shows that how population mean of counties range, and the trend of the plot shows that a very high amount of countries has mean more than 300. The distribution shown by the density plot is bimodal distribution.

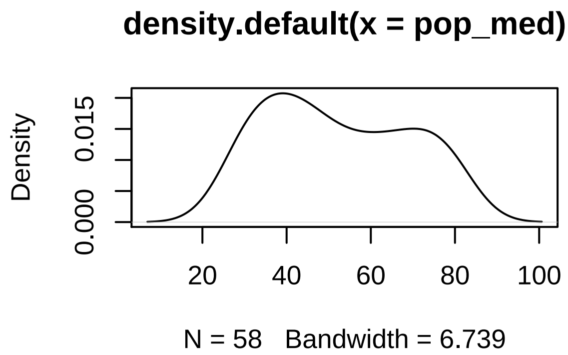
```r
#plotting a line curve using ggplot for mean to understand the curve
date <- 1961:2018
ggplot() + geom_line(aes(x=date,y=pop_mean),color='red') +
  ylab('Population mean')+xlab('date')
```

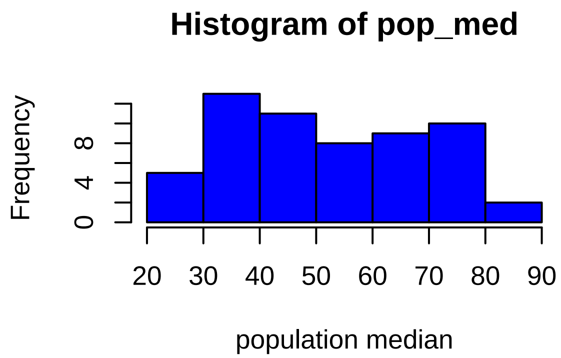


This further state that on increasing time span the population mean increases. But now due to initiative taken by government and other machinary bodies, this rate is getting checked in various countries due to which we can see a fall in population mean in recent years. [3]. The same trend can be seen in the median of population data over the time validating the analysis done on the mean.

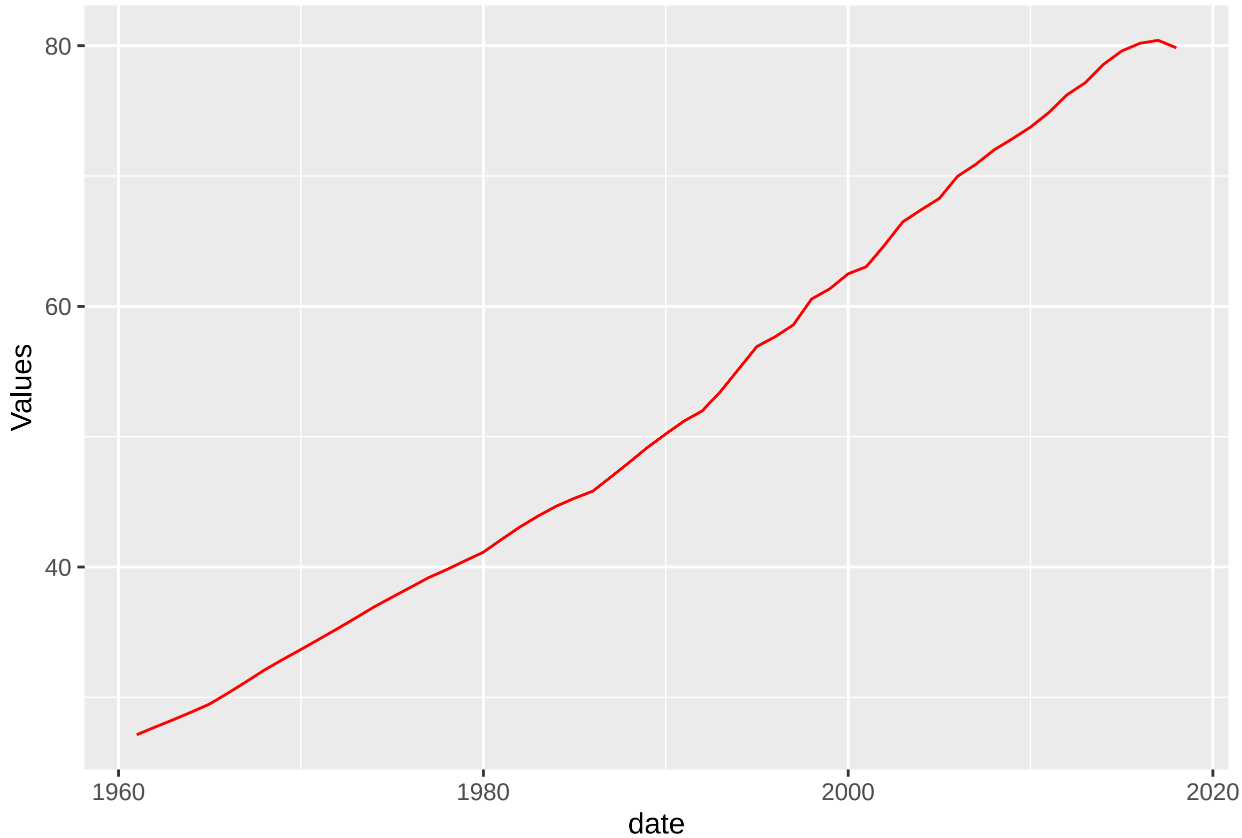
```
#Finding the median of all the population and storing it
pop_med <- colMedians( as.matrix(pop_den[5:62]), na.rm=TRUE)

par(mfrow=c(2,2))
#plotting the median in barplot, density plot and histogram
barplot(pop_med, names = seq(1961, 2018, 1), las = 2)
d2 <- density(pop_med)
plot(d2)
hist(pop_med, xlab = "population median", col = " blue")
```



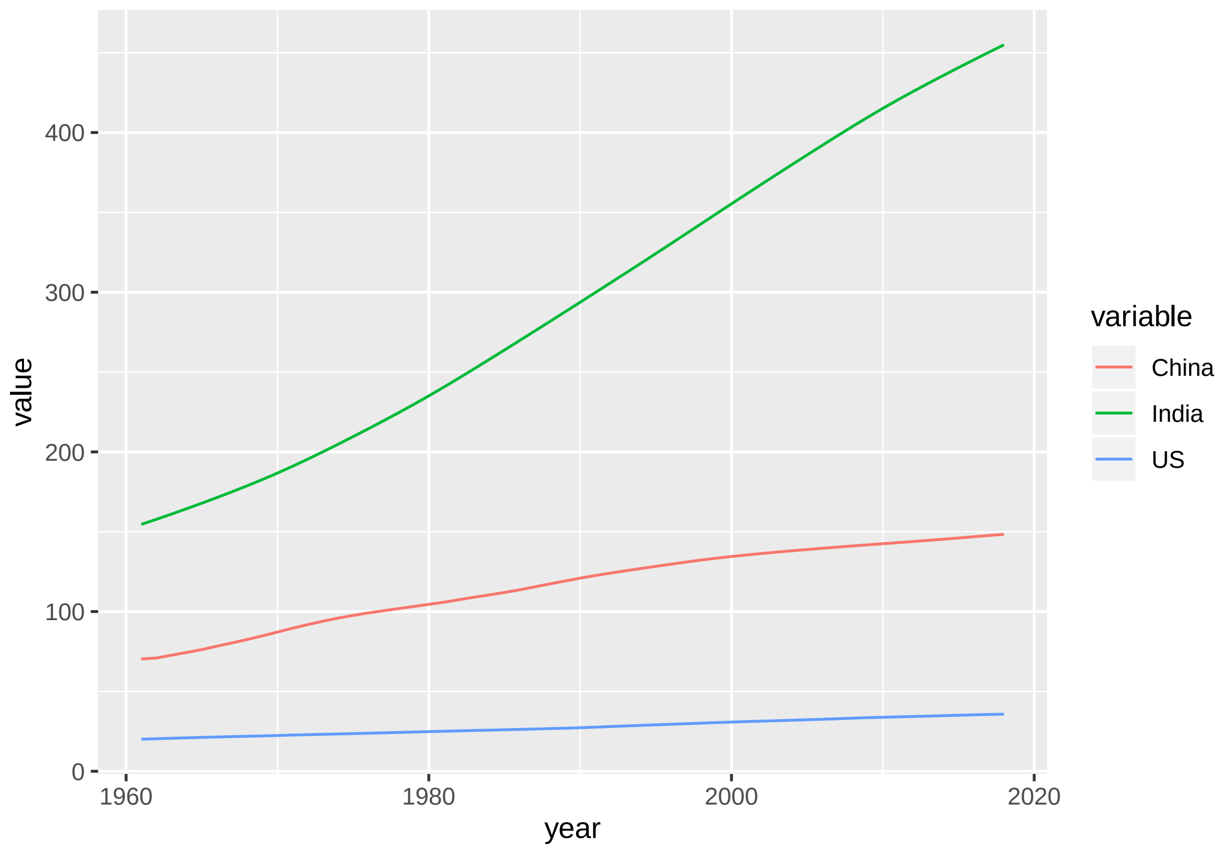





```
#plotting a line curve using ggplot for median to understand the curve
date <- 1961:2018
ggplot() + geom_line(aes(x=date,y=pop_med),color='red') +
  ylab('Values')+xlab('date')
```



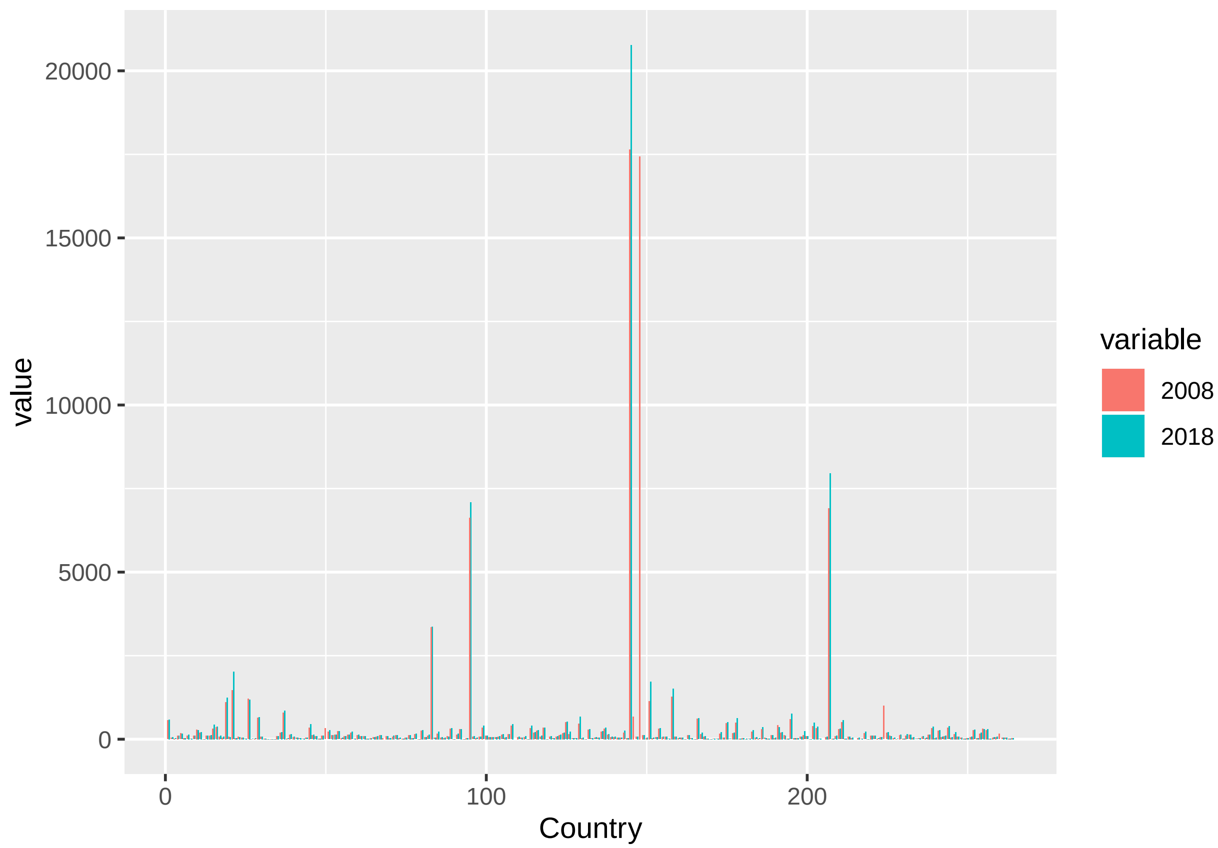

```r
library(reshape2)
#plotting population density of different countries using ggplot
x1 <- 1961:2018
y1 <- pop_den[39,5:62]
y2 <- pop_den[108,5:62]
y3 <- pop_den[250,5:62]
t <- data.frame(cbind(x1, t(y1),t(y2),t(y3)))
colnames(t) <- c("year", "China", "India", "US")
country <- melt(t, id= "year")
ggplot(data = country , aes(x = year, y = value, color = variable)) + geom_line()
```



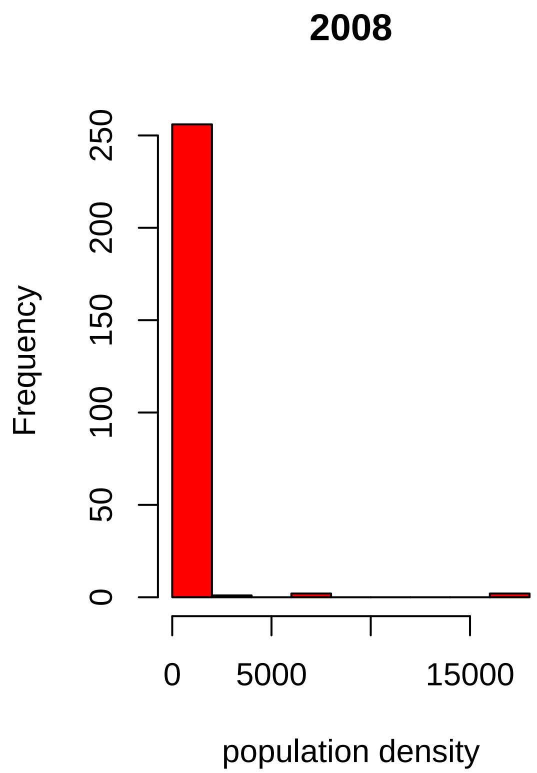
This plot has been shown to depict how population boom is taking place in India. Although China and India had same initial growth rate but they controlled population density whereas our government has not taken any evident step to control this exponential growth. [4]

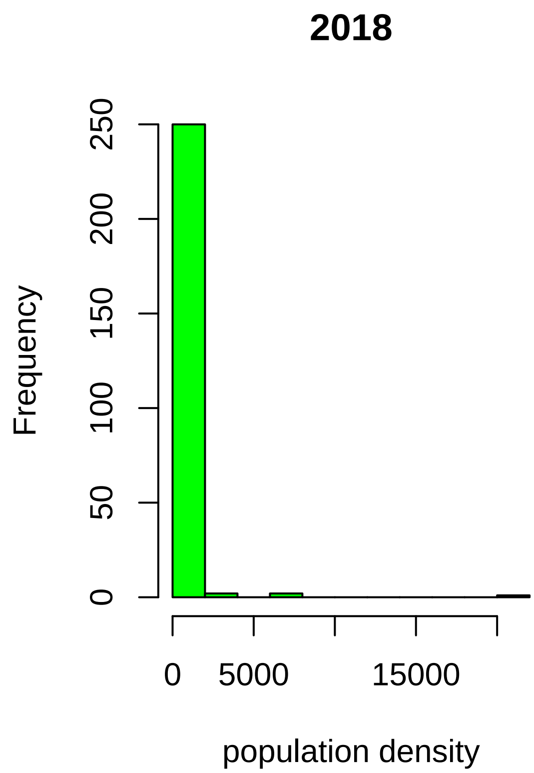
```
#comparison of population density for two years(2018,2008) at a span of 10 years.
x1 <- pop_den$`2008`
x2 <- pop_den$`2018`

y1 <- 1:264 #labelled the countries with numbers
t <- data.frame(cbind(y1, x1, x2))
colnames(t) <- c("Country", "2008", "2018")
dat <- melt(t, id= "Country")
ggplot(data=dat, aes(x=Country,y=value,fill=variable)) +
  geom_bar(position=position_dodge(), stat= 'identity')
```

```
## Warning: Removed 12 rows containing missing values (geom_bar).
```



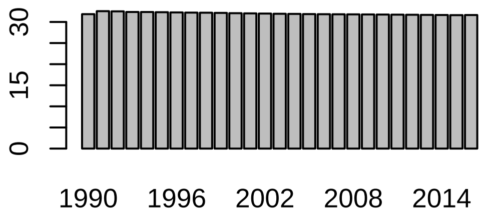

```r
par(mfrow=c(1,2))
hist(x1, main = "2008", xlab = "population density", col = 'red')
hist(x2, main = "2018", xlab = "population density", col = 'green')
```

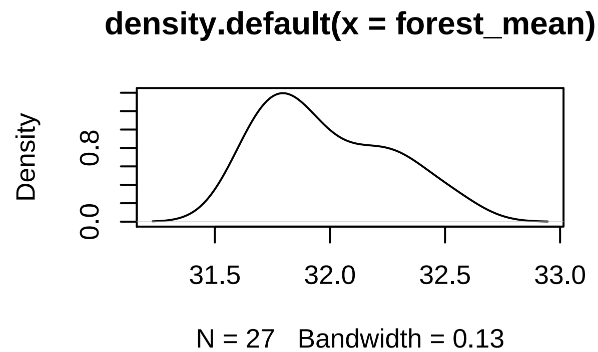


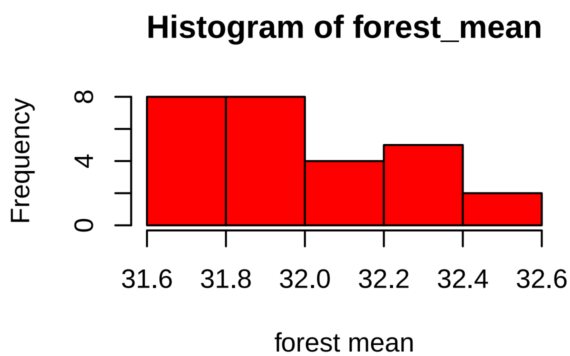


The population density count of countries with more than 20000 has increased over 10 years which is very alarming because in next plots its affect in forest density will be shown. A major portion of the land where population growth is taking place by deforestation. this is leading to fall in ground water level, drought extension, flood and global rise in temperature. Forest land is being used for urbanization.

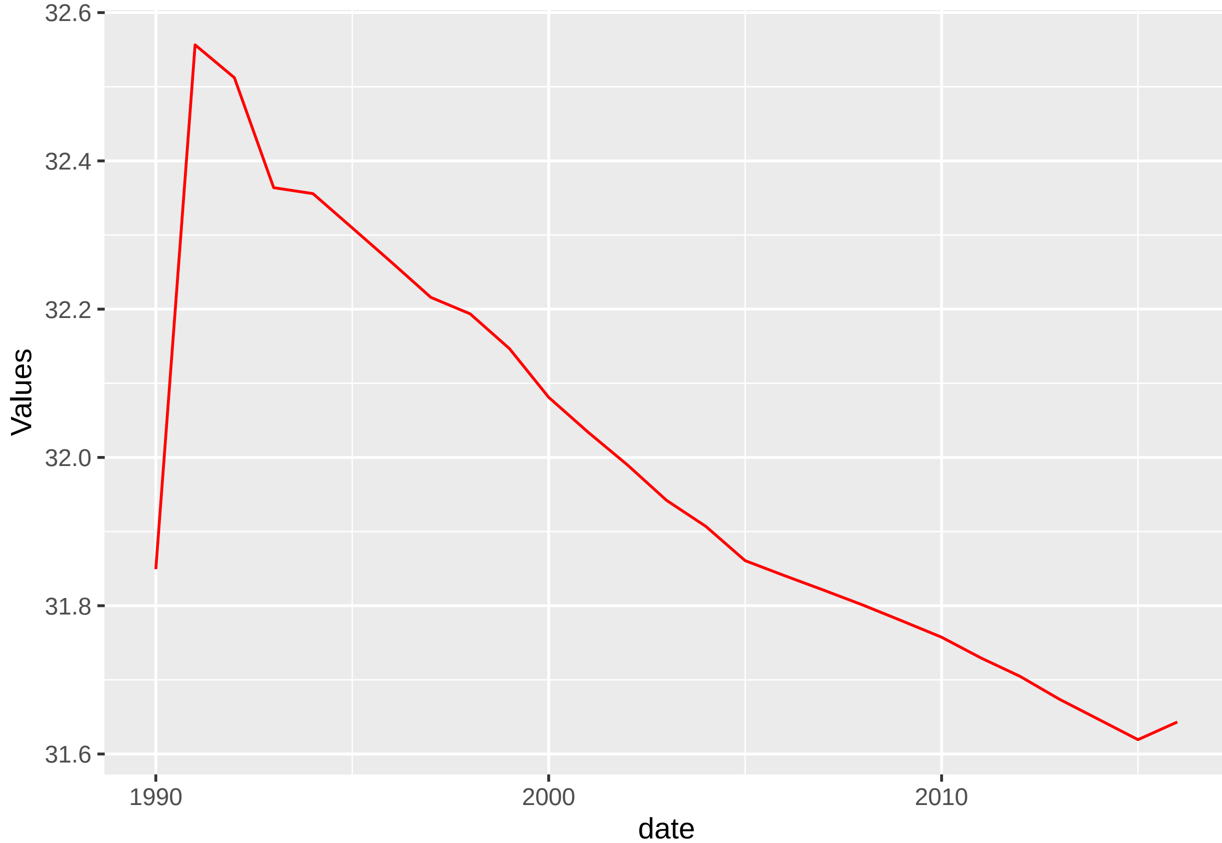
```
#loading the xls file
forest_den = read_excel("forest.xls")

#removing the columns which does not contain any value
forest_den[,5:34] <- NULL
forest_den[,32:33] <- NULL

#Finding the mean of the forest density and storing it
forest_mean <- colMeans(forest_den[5:31], na.rm=TRUE)

par(mfrow=c(2,2))
#plotting the mean in barplot, density plot and histogram
barplot(forest_mean)
d1 <- density(forest_mean)
plot(d1)
hist(forest_mean, xlab = "forest mean", col = "red")
```





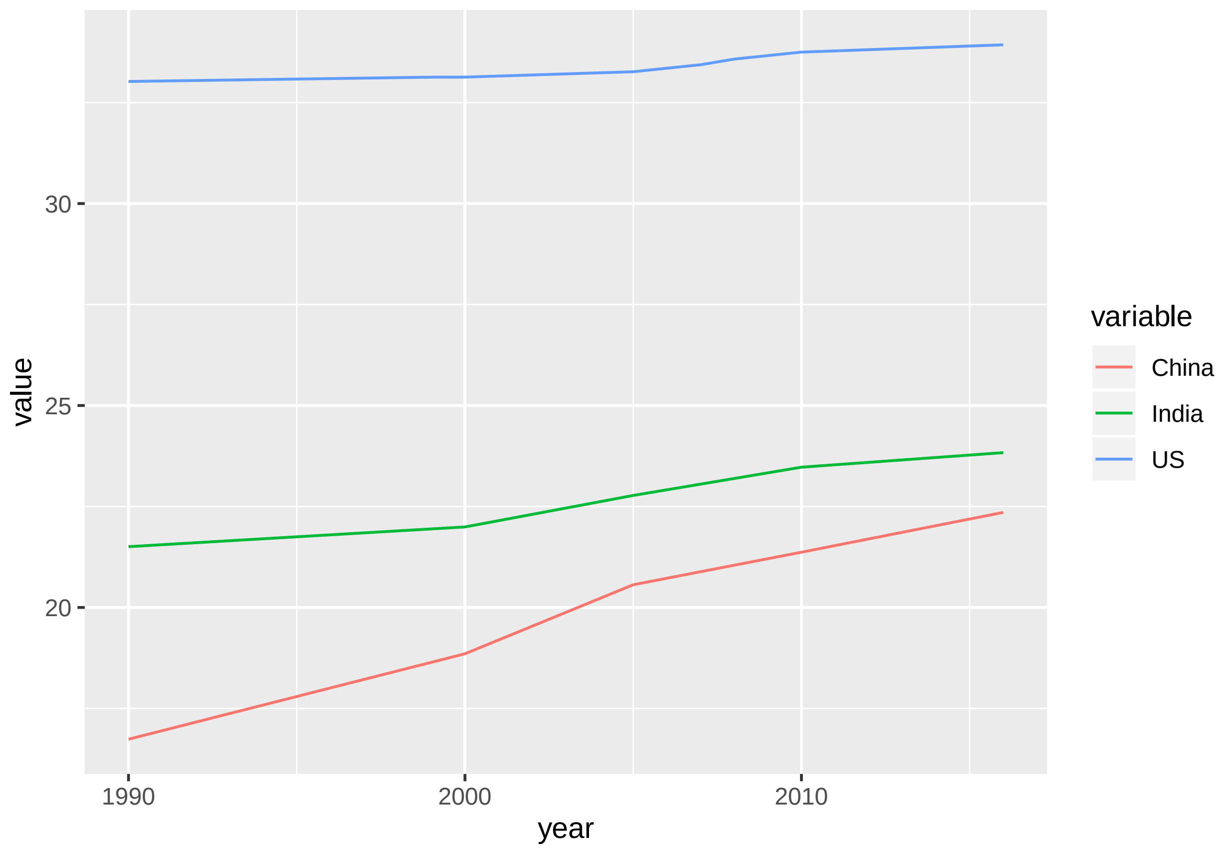



```r
#plotting a line curve using ggplot for mean to understand the curve
date <- 1990:2016
ggplot() + geom_line(aes(x=date,y=forest_mean),color='red') +
  ylab('Values')+xlab('date')
```



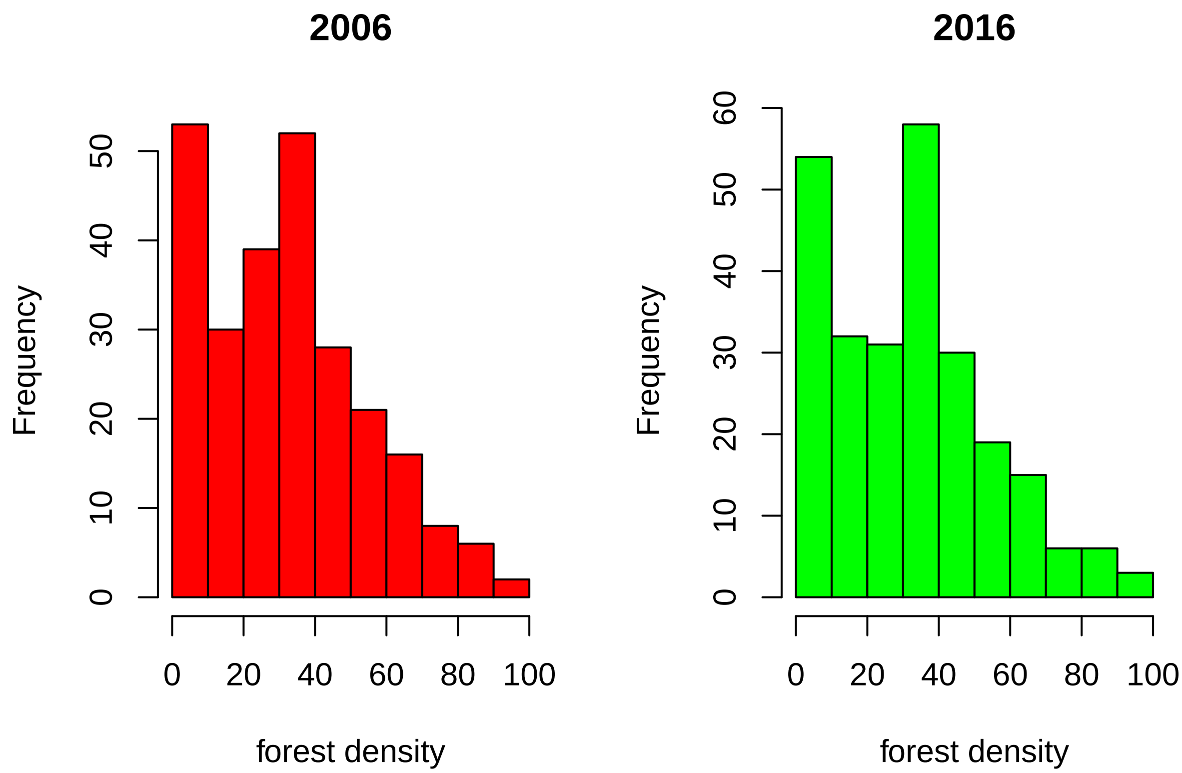
A clear visualization can be seen that the forest land is decreasing with years. Recent debacle that occurred in Brazil is just due to urbanization. Lands are being used for creating building, modernising which is adversely affecting the environment.

```
#plotting forest density of different countries using ggplot
x1 <- 1990:2016
y1 <- forest_den[39,5:31]
y2 <- forest_den[108,5:31]
y3 <- forest_den[250,5:31]
t <- data.frame(cbind(x1, t(y1),t(y2),t(y3)))
colnames(t) <- c("year", "China", "India", "US")
country <- melt(t, id= "year")
ggplot(data = country , aes(x = year, y = value, color = variable)) + geom_line()
```



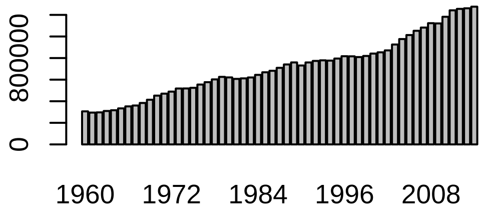
A rapid deforestation has been seen in the developing countries whereas in developed countries and the rate is linear over the years but in case of developed countries it seems to be more or less contant. This practically validates to the fact that urbanisation is leading to this destruction.

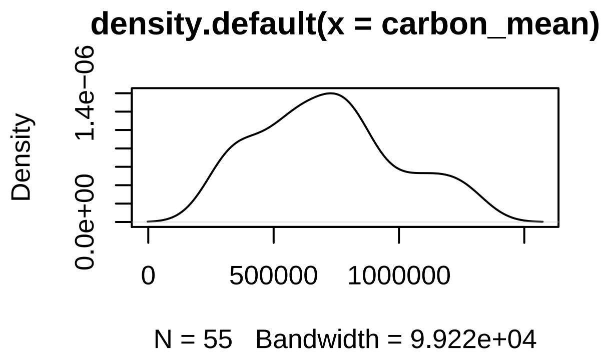
```
#comparison of forest density for two years(2018,2008) at a span of 10 years.
x1 <- forest_den$`2008`
x2 <- forest_den$`2016`
y1 <- 1:264 #labelled the countries with numbers
t <- data.frame(cbind(y1, x1, x2))
colnames(t) <- c("Country", "2006", "2016")
dat <- melt(t, id= "Country")

par(mfrow=c(1,2))
hist(x1, main = "2006", xlab = "forest density", col = 'red')
hist(x2, main = "2016", xlab = "forest density", col = 'green')
```

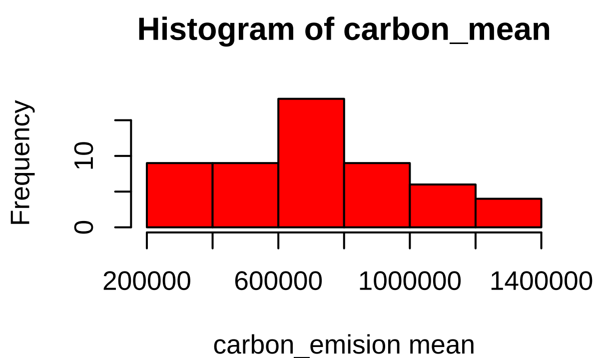


Plant act as lungs for nature. This fast rate of decomposition has increased the rate of carbon dioxide in the nature. This is leading to anamoly change in temperature which has been reported by NASA. Below this we are going to show a case study of how Carbon dioxide concentraion is increasing in the nature.

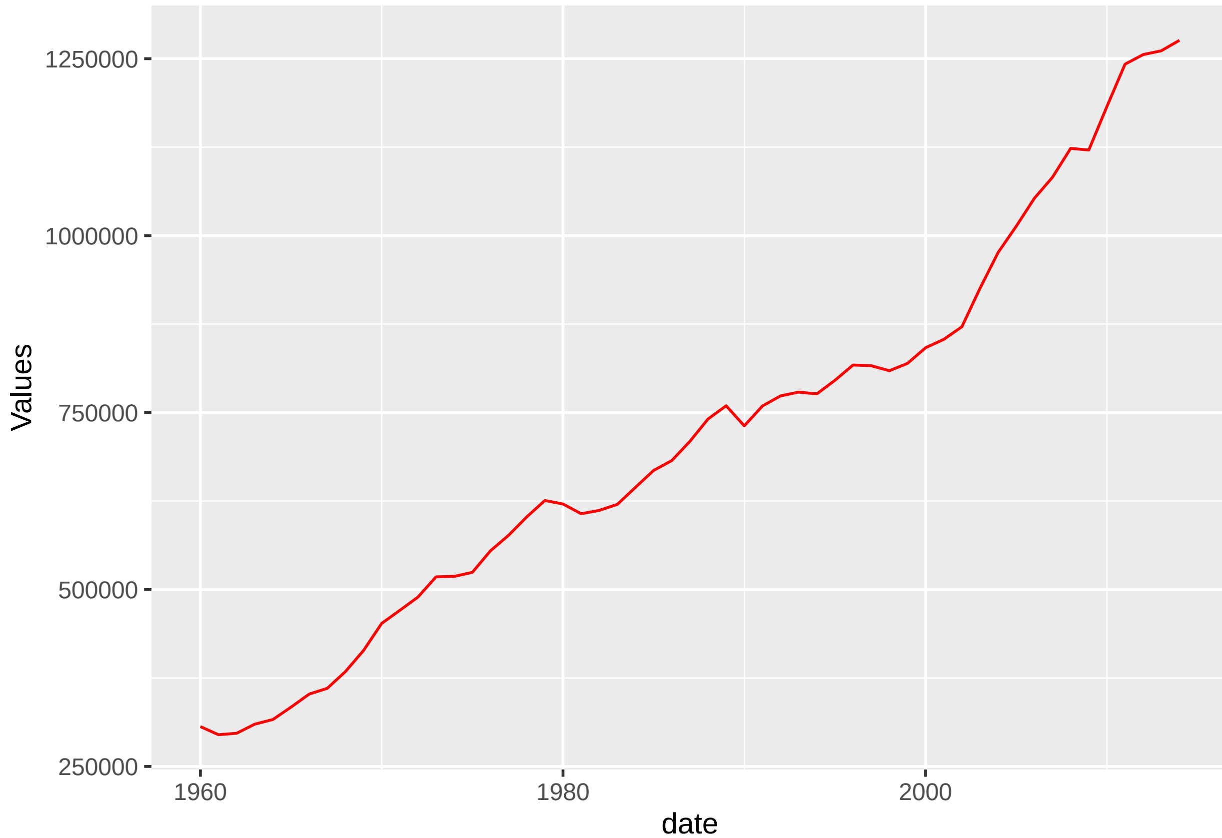
```
carbon_emision = read_excel("carbondioxide.xls")

carbon_emision[,60:63] <- NULL

#Finding the mean of all countries carbon emision and storing it
carbon_mean <- colMeans(carbon_emision[5:59], na.rm=TRUE)

par(mfrow=c(2,2))
#plotting the mean in barplot, density plot and histogram
barplot(carbon_mean)
d1 <- density(carbon_mean)
plot(d1)
hist(carbon_mean, xlab = "carbon_emision mean", col = "red")
```







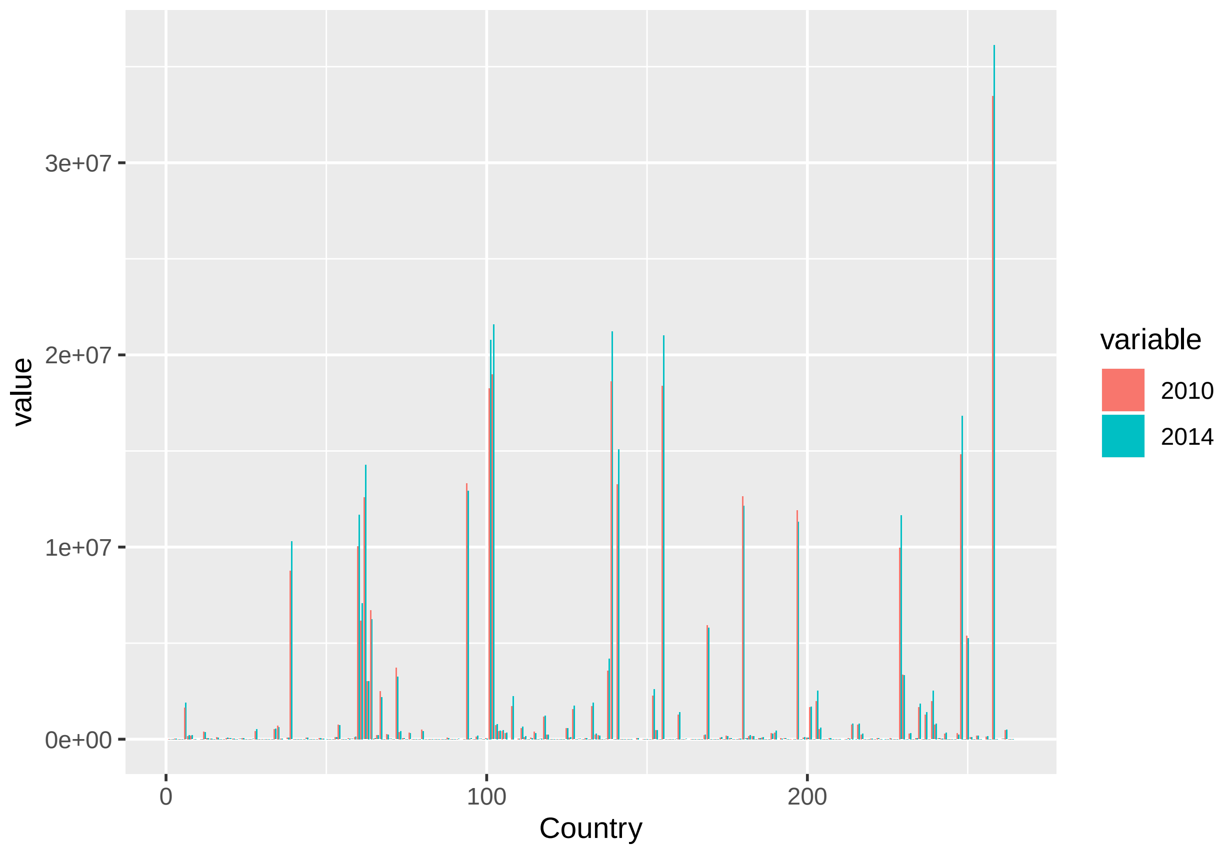

```
#plotting a line curve using ggplot for median to understand the curve
date <- 1960:2014
ggplot() + geom_line(aes(x=date,y=carbon_mean),color='red') +
  ylab('Values')+xlab('date')
```



The bar plot and line plot clearly predicts how the carbon emision is increasing over the time period. Modelling of the line plotting against time clearly depicts that the concentration level of carbon is at exponential rate. The bar plot below shows how within a span of 5 years the level of concentration has increased at such drastic rate creating a debacle in whole ecology.

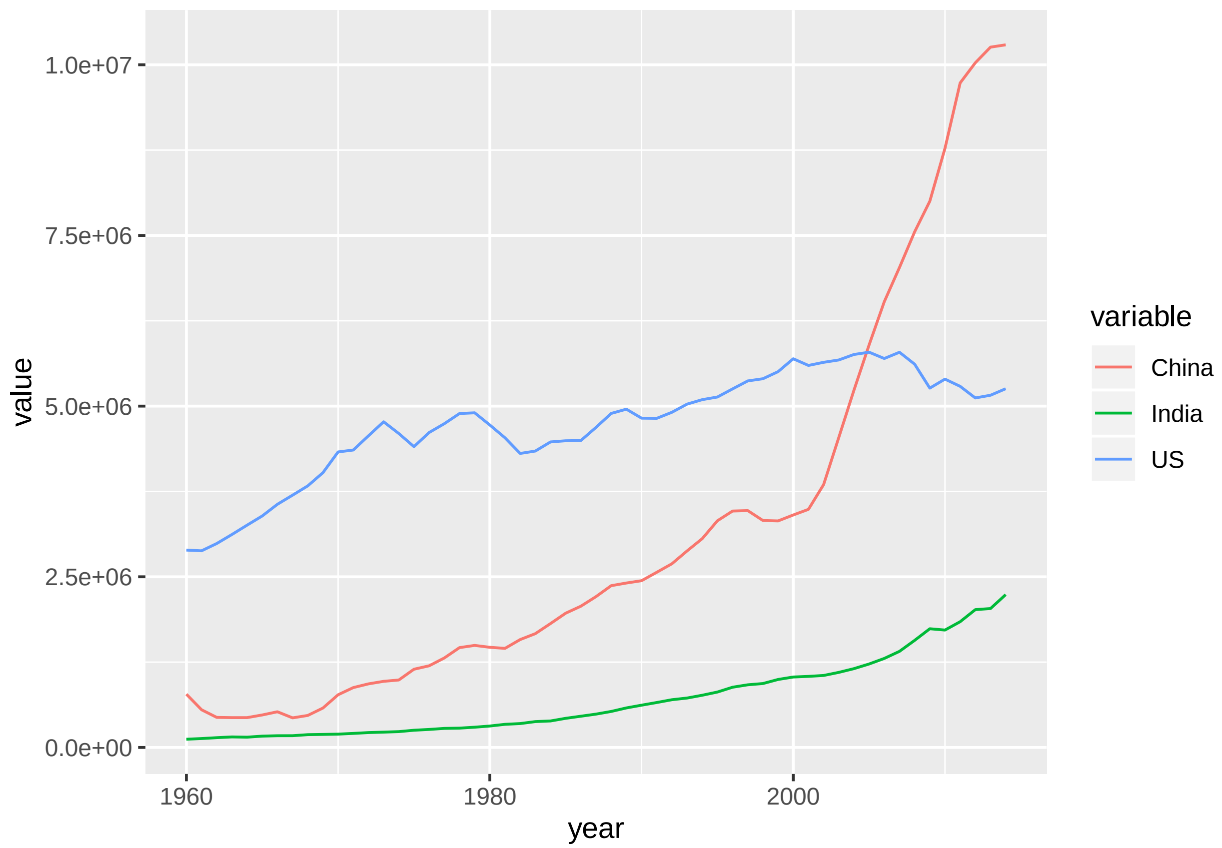
```
#comparison of carbon emision for two years(2010,2014) .
x1 <- carbon_emision$`2010`
x2 <- carbon_emision$`2014`

y1 <- 1:264 #labelled the countries with numbers
t <- data.frame(cbind(y1, x1, x2))
colnames(t) <- c("Country", "2010", "2014")
dat <- melt(t, id= "Country")
ggplot(data=dat, aes(x=Country,y=value,fill=variable)) +
  geom_bar(position=position_dodge(), stat= 'identity')
```

```
## Warning: Removed 28 rows containing missing values (geom_bar).
```



Concentration of carbon dioxide is increasing most in China since urbanization is peak as well as it has huge population which was already depicted in the previous plots. Being a developed country its contributing most to carbon dioxide concentration in this plot. various measure has been taken by chinese government to check this growth rate. [5]

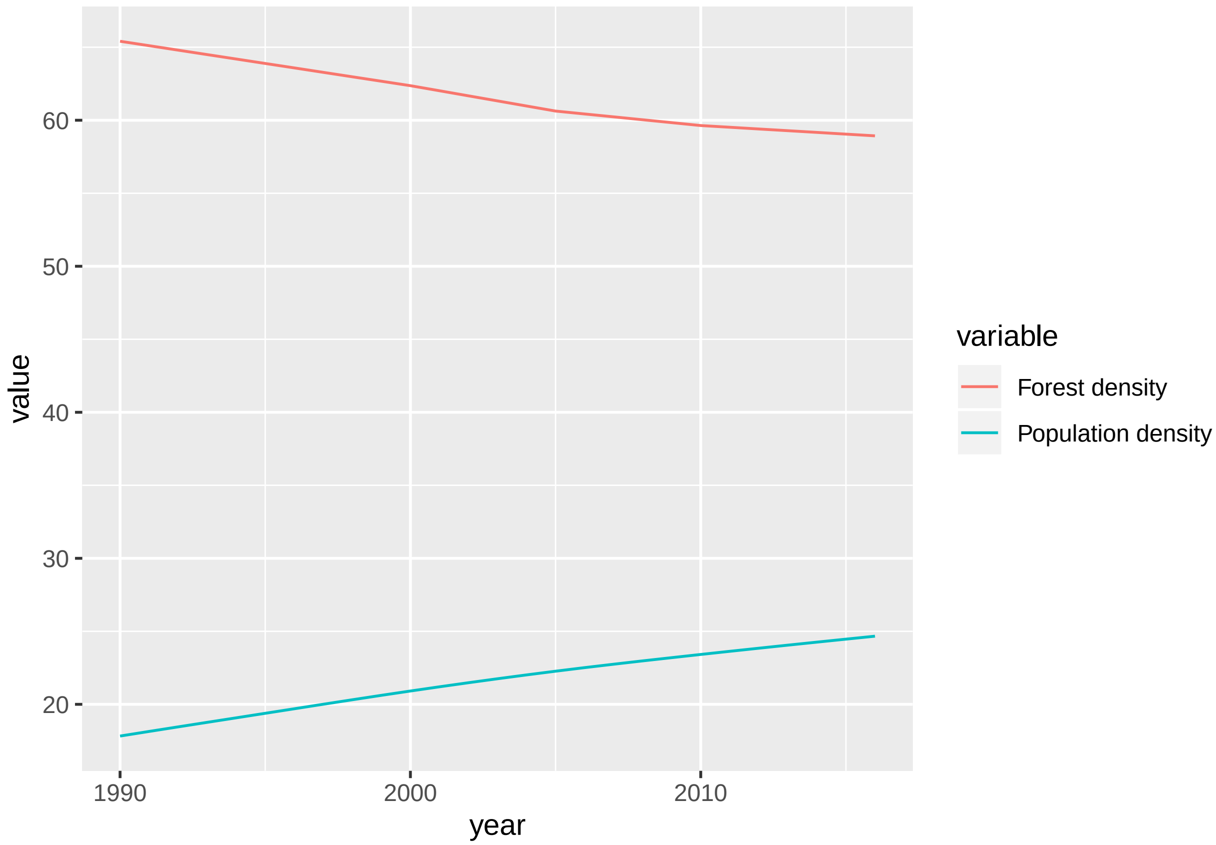
```
#plotting carbon emision of different countries using ggplot
x1 <- 1960:2014
y1 <- carbon_emision[39,5:59]
y2 <- carbon_emision[108,5:59]
y3 <- carbon_emision[250,5:59]
t <- data.frame(cbind(x1, t(y1),t(y2),t(y3)))
colnames(t) <- c("year", "China", "India", "US")
country <- melt(t, id= "year")
ggplot(data = country , aes(x = year, y = value, color = variable)) + geom_line()
```



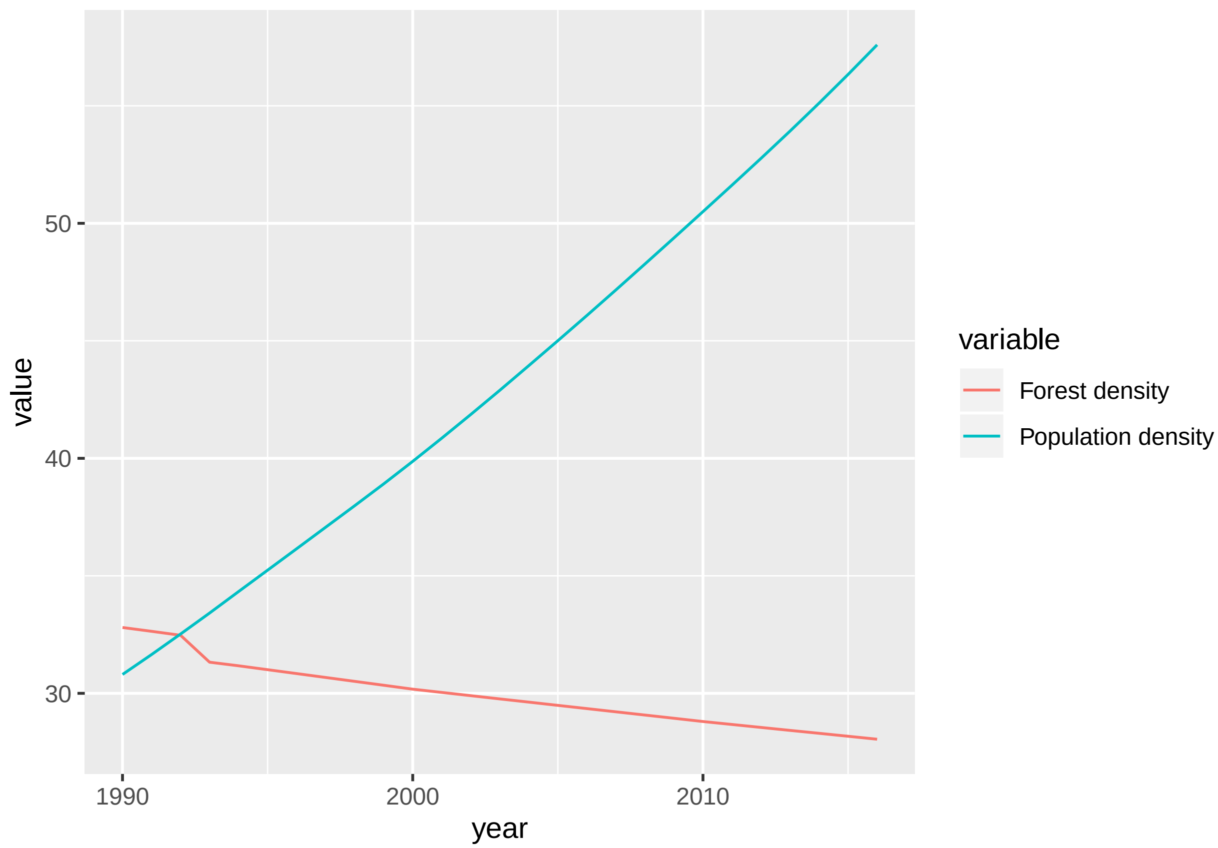
A few case studies with multivariate plots will make this more interesting. The recent debacle that occured in Brazil[6] has lead us to think about environment as primary concerns for every country. A analysis is shown below that how rate of population has directly affected forest denisty. An alarming concern has been seen even in the India and that has been shown in the below plot.

```
#loading the xls file
forest_den = read_excel("forest.xls")
pop_den = read_excel("population_density.xls")
forest_den[,c(5:34, 62:63)] <- NULL
pop_den[,c(5:34, 62:63)] <- NULL

#merging two dataframes
pop_for <- merge( forest_den ,pop_den , by="Country Name")

#Plotting forest density and population density for Brazil
brazil <- pop_for %>% slice(28)
xx = t(brazil[c(5:31)])
yy = t(brazil[c(35:61)])
zz = cbind(xx,yy)
zz = as.data.frame(zz)

x1 <- 1990:2016
y1 <- zz[,1]
y2 <- zz[,2]

t <- data.frame(cbind(x1,y1,y2))
colnames(t) <- c("year", "Forest density", "Population density")
```

```r
country <- melt(t, id= "year")
ggplot(data = country , aes(x = year, y = value, color = variable)) + geom_line()
```




```r
#Plotting forest density and population density for India
india <- pop_for %>% slice(108)
xx = t(india[c(5:31)])
yy = t(india[c(35:61)])
zz = cbind(xx,yy)
zz = as.data.frame(zz)

x1 <- 1990:2016
y1 <- zz[,1]
y2 <- zz[,2]

t <- data.frame(cbind(x1,y1,y2))
colnames(t) <- c("year", "Forest density", "Population density")
country <- melt(t, id= "year")
ggplot(data = country , aes(x = year, y = value, color = variable)) + geom_line()
```



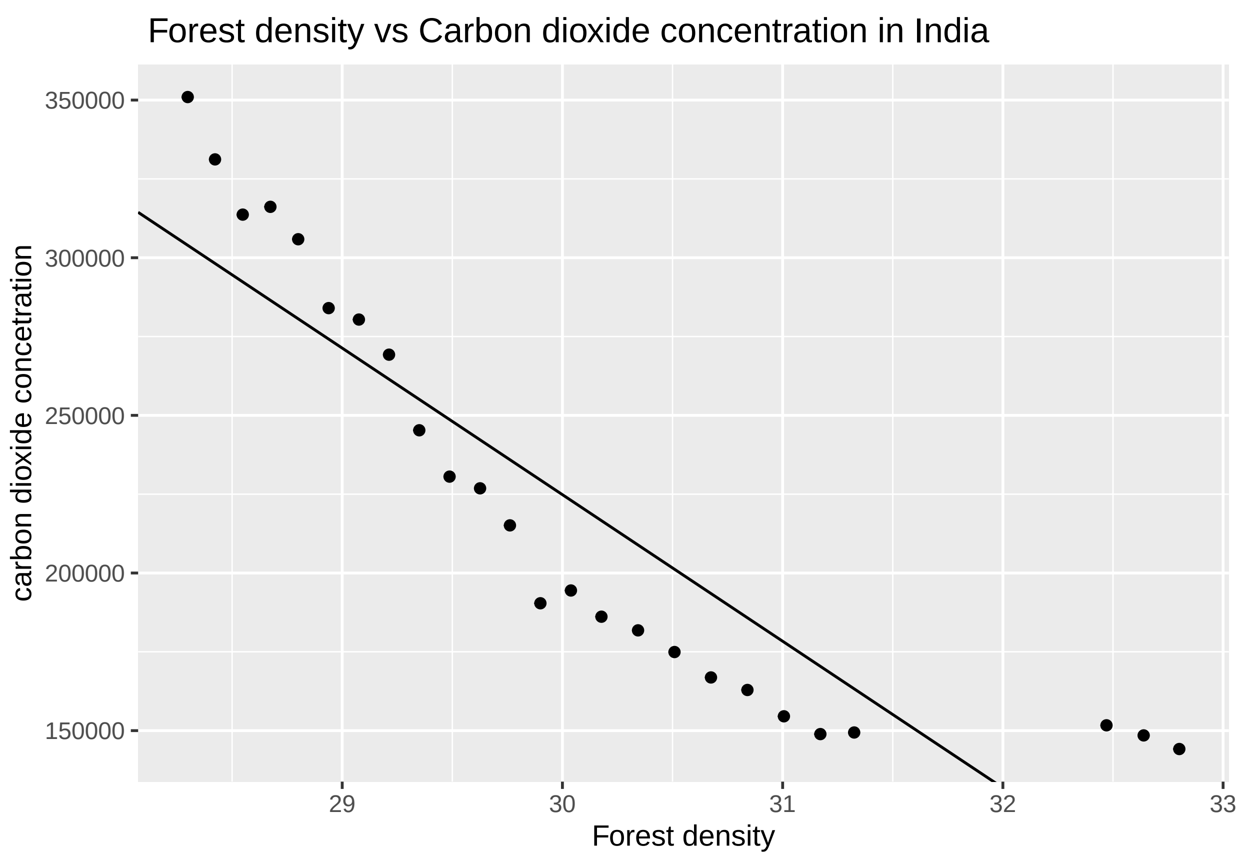
An alarming state of concern has also arisen in India where the country is suffering from acute polution stress, Delhi[7] being a major concern. Various forest acquired lands are being cleared to modernise the country it is actually hampering the country. It adverse affect can be easily visualized from the plot drawn below. It clearly states a negative correlation which depicts that with decrease in forest density, the carbon dioxide concentration in nature is increasing.[8]

```
## carbon_emision and forest density comparison

forest_den = read_excel("forest.xls")
carbon_emision = read_excel("carbondioxide.xls")


forest_den[,c(5:34,62:63)] <- NULL
carbon_emision[,c(5:34,62:63)] <- NULL

for_car <- merge( forest_den , carbon_emision, by="Country Name")

#Plotting forest density vs carbon emision in India in recent past
india <- for_car %>% slice(108)
xx = t(india[c(5:29)])
yy = t(india[c(35:59)])
zz = cbind(xx,yy)
zz = as.data.frame(zz)

x1 <- 1990:2014
y1 <- zz[,1]
y2 <- zz[,2]
```

```r
t <- data.frame(cbind(x1,y1,y2))
colnames(t) <- c("year", "Forest", "carbon")


coef(lm(carbon ~ Forest, data = t))
```

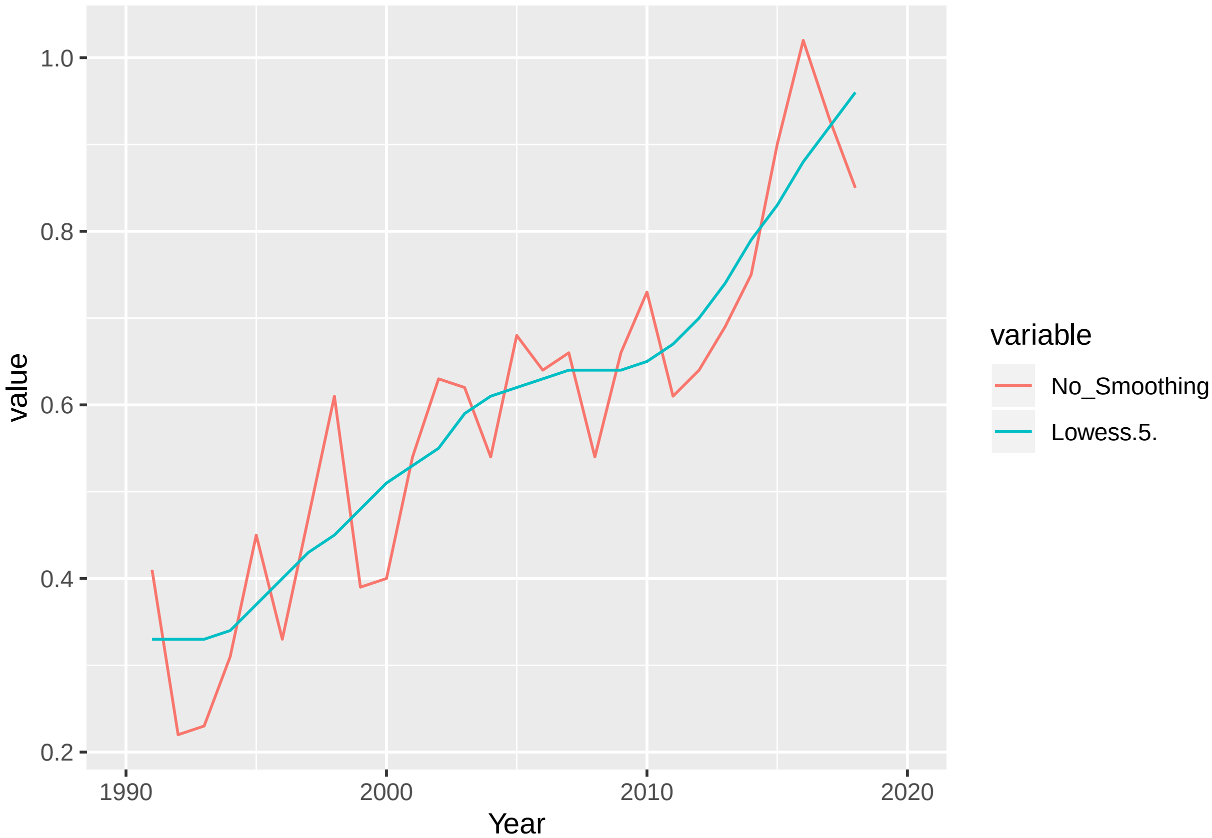
```
## (Intercept)      Forest
##  1619474.05   -46487.65
```

```r
ggplot(data=t, aes(x= Forest, y= carbon))+
  xlab("Forest density")+ ylab("carbon dioxide concetration")+
  geom_point()+
  geom_abline(intercept = 1619474.05, slope = -46487.65   ) +
  ggtitle(" Forest density vs Carbon dioxide concentration in India")
```



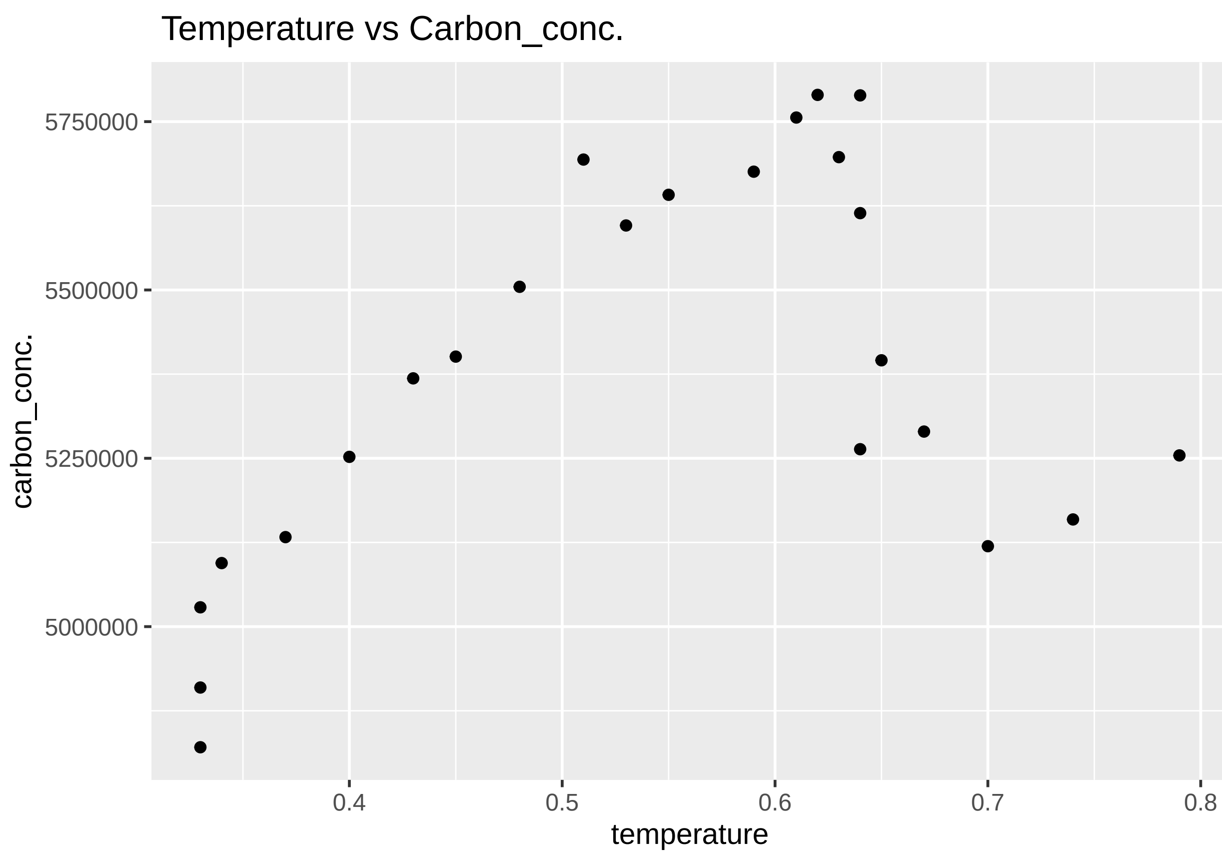
Another Bi-variate plot has been depicted to show the anamoly of temperature in recent years. There has been an adequate increase in global temperature over the time series. This is happening due to global warming , where carbon dioxide concentration holds accountability of this phenomenon.

```r
temperature1 = read_excel("temp.xlsx")
temp2 = temperature1[-c(1:111),]
temp2 = data.frame(temp2)
rise <- melt(temp2, id= "Year")
ggplot(data = rise , aes(x = Year, y = value, color = variable)) + geom_line() + xlim (1990,2020)
```



The scatter below shows a positive correlation which states that the temperature increases with increase in the carbon dioxide concentraion. The carbon dioxide concentration increases due to deforestation. So it can directly stated that deforestation is leading to many adverse effects in environment.

```
year <- 1990:2014
US <- t(carbon_emision[250, 5:29])
rownames(US) = NULL
US_carbon = cbind(year, US)

colnames(US_carbon) <- c("Year","Carbon")

carbon_for <- merge(temp2, US_carbon, by="Year")

ggplot(data=carbon_for, aes(x= Lowess.5. , y= Carbon))+
  xlab("temperature")+ ylab("carbon_conc.")+
  geom_point()+
  ggtitle(" Temperature vs Carbon_conc.")
```



Conclusion

The adverse affect caused due to modernisation is creating a big threat to ecology due to which we must take some precautionary steps to save the environment and ecology. The imbalance and adverse effect of human destruction is leading to climate change and leading to extinction of many creatures. WHO has organised various workshop about ecological concerns. We should also take the responsibilty to for attributing less stress to ecology. The survey clearly shows how fast the deforestation has occured in recent past, and how how it is adversely affecting the ecology by increasing the global temperature, rise in sea level and increase in carbon dioxide concentration. If this continue at this rate then whole earth will extinct just like how in pre historic era. The rate of deforestation is increasing at exponential rate similarly it affects the rise in temperature which is also increasing at linear rate. Many endangered species have extincted due to this anamoly of temperature.

India is facing the major crisis in terms of pollution which is a direct result of deforestation and carbon dioxide (due to automobiles, industries) concentration increase in atmosphere. Several places are facing severe drought, floods, fall in ground water level and rise in temperature. Various reports from WHO and NASA has been publised. Government must take serious measures to stop this degradation elsewhile we will face serious threats in future.