AIR POLLUTION ANALYSIS

A Mini Project Report Submitted by

Nivedita Suresh Kumar (4NM17CS116)

Pooja Mastappa Naik (4NM17CS125)

Prathika S Shetty (4NM17CS131)

Priya Shetty (4NM17CS139)

Raksha D Shetty (4NM17CS142)

UNDER THE GUIDANCE OF

Mrs. Savitha Shetty

ASST. PROFESSOR GRADE II
Department of Computer Science and Engineering

in partial fulfilment of the requirements for the award of the Degree of

Bachelor of Engineering in
Computer Science & Engineering

From

Visvesvaraya Technological University, Belgaum

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

N.M.A.M. INSTITUTE OF TECHNOLOGY

(An Autonomous Institution under VTU, Belgaum) (AICTE approved, NBA Accredited, ISO 9001:2008 Certified) NITTE -574 110, Udupi District, KARNATAKA.

April 2020

NITTE EDUCATION TRUST | N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India

(ISO 9001:2015 Certified), Accredited with 'A'Grade by NAAC

(: 08258 - 281039 –281263, Fax: 08258 –281265

Department of Computer Science and Engineering

B.E. CSE Program Accredited by NBA, New Delhi from 1-7-2018 to 30-6-2021

CERTIFICATE

"**Air Pollution Analysis**"

is a bonafide work carried out by

Nivedita Suresh Kumar (4NM17CS116)
Pooja Mastappa Naik (4NM17CS125)
Prathika S Shetty (4NM17CS131)
Priya Shetty (4NM17CS139)
Raksha D Shetty (4NM17CS142)

in partial fulfilment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering prescribed by Visvesvaraya Technological University, Belgaum during the year 2019-2020.

It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report.

The Mini project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the Bachelor of Engineering Degree.

Signature of Guide Signature of HOD

ACKNOWLEDGEMENT

We believe that our project will be complete only after we thank the people who have contributed to make this project successful.

First and foremost, our sincere thanks to our beloved principal, **Dr. Niranjan N. Chiplunkar** for giving us an opportunity to carry out our project work at our college and providing us with all the needed facilities.

We express our deep sense of gratitude and indebtedness to our guide **Mrs. Savitha Shetty**, Assistant Professor, Department of Computer Science and Engineering, for her inspiring guidance, constant encouragement, support and suggestions for improvement during the course of our project.

We sincerely thank **Dr. K.R. Udaya Kumar Reddy**, Head of Department of Computer Science and Engineering, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, Nitte.

We also thank all those who have supported us throughout the entire duration of our project.

Finally, we thank the staff members of the Department of Computer Science and Engineering and all our friends for their honest opinions and suggestions throughout the course of our project.

Nivedita Suresh Kumar (4NM17CS116)
Pooja Mastappa Naik (4NM17CS125)
Prathika S Shetty (4NM17CS131)
Priya Shetty (4NM17CS139)
Raksha D Shetty (4NM17CS142)

ABSTRACT

Air pollution has been one of the significant problems to deal with in all the nations today. In South Asia, it is ranked as the sixth most dangerous killer. Keeping this in mind, we‘ve analyzed the air quality data of a few Indian States to find some underlying principles or patterns which might give us an insight into how severe the problem is. The approach is purely data -driven and makes use of Big Data Analytics technologies.

Big data is a new driver of the world‘s economic and societal changes. The world’s data collection is reaching a tipping point for major technological changes that can bring new ways in decision making, managing our health, cities, finance and education. While the data complexities are increasing including data’s volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the ‘value‘ in the data through Big Data Analytics technologies. Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse and complex on a massive scale .

It is interesting to see how data analysis and the day to day instances are coherent and how data analysis can be used to deal with significant problems. Big data has totally changed and revolutionized the way business and organizations work. Leveraging Big Data for environmental protection is still an untapped frontier. In the case of air pollution, once when the key insights on the contributing factors of pollution are identified , the management and prediction becomes far easier.

TABLE OF CONTENTS

INTRODUCTION

In this modern world where we come across new technologies every day, we don't thoroughly examine their effect on the nature around us. Our mini project basically gives a glimpse of how these technologies affect the air. For the analysis, we have picked 5 states, created a dataset based on the factors such as Industry count, Birth-rate, Death-rate, Forest-cover, Population, Vehicle-count and so on and conducted analysis on them.

This Air pollution analysis is a Big Data Analysis project where we have analysed a large dataset not capable of being analysed by a typical database or data analysis software like Excel, Weka etc. After the analysis we have displayed analysed data i.e., graphs in web pages which makes it easy for the user to understand. We hope that this makes some impact on the people about the deteriorating air quality.

Big Data has grasped a lot attention from the market trends, equipment-based performance, and other industry elements. Big data, analytical tools and technologies greatly assist in IT decision making. Even the large organizations find it difficult to deal with the larger datasets in terms of manipulating and managing the Big Data. Big data is particularly a troublesome factor in business analytics since the traditional tools and procedures are not designed to search and analyse massive datasets.

Big Data deals with two classes of data sets, namely structured and unstructured. The records obtained from inventories, orders, and customer information contributes to the structured datasets. The unstructured data set can be obtained from the web, social media, and intelligent devices.

OBJECTIVE

The primary objectives of this project is to collect data pertaining to the determinants of the analysis of the air pollution level of five different states in India during the years 2008-2012 and thereafter calculate and generate reports on the same based on the real time data collected. The determinants for the analysis of the air pollution level are:

- Population density of the states (persons/ sq km)
- Percentage of birth and death rate of the people living in a state
- Count of vehicles owned per family
- Count of industries in the state
- Count of people admitted in the hospital due to air borne diseases
- The total area covered by forests in a state (sq km)
- The concentration of particulate matter in the air (NO_2,SO_2,PM)

TOOLS USED

- **R** - It is a software programming language and software environment for statistical computing and graphics. It is an integrated suite of software facilities for data manipulation, calculation and graphic display. It is an open source application which includes virtually every data manipulation techniques, statistical model and chart. We plan to use R primarily for pre-processing the data, manipulating it into structured form for the analysis and present end results.

- **RStudio** - RStudio is an integrated development environment (IDE) for R language. It is a code editor and development environment. As soon as you create a new script, the windows within your RStudio session adjust automatically so you can see both your script and the results in your console when you run your syntax. R packages are an ideal way to package and distribute R code and data for re-use by others. RStudio includes a variety of tools that makes developing R packages easier and more productive.

- **MongoDB** - MongoDB is a document-oriented NoSQL database used for high volume data storage. Instead of using tables and rows as in the traditional relational databases, MongoDB makes use of collections and documents. Documents consist of key-value pairs which are the basic unit of data in MongoDB. Collections contain sets of documents and function which is the equivalent of relational database tables.

- **CSV** - A CSV is a comma-separated values file, which allows data to be saved in a tabular format. CSVs look like a garden-variety spreadsheet but with a .csv extension. CSV files can be used with any spreadsheet program, such as Microsoft Excel or Google Spreadsheets.

IMPLEMENTATION

The project is implemented in four stages:

- Collection of data
- Import the data into MongoDB
- Design a data analysis program in R using RStudio to create some visual representation of the data
- Present the results of the analysis using web pages

Collection Of Data

The required data of five different Indian states i.e., Delhi, Karnataka, Kerala, Rajasthan and Sikkim over the period of 2008 – 2012 is collected from different sources.

Here is the snapshot of the first few rows of the data collected.

	State	Year	Population.Density..persons.sq.km.	People.with.air.borne.diseases	Industries	Vehicles	Birthrate....	Deathrate....	Forest.cover	SO2	NO2	PM
1	Delhi	2008	10726.0	182256	3026	5899	18.4	4.8	176	7	57	214
2	Karnataka	2008	306.1	1777839	8451	6217	19.8	7.4	36105	13	28	80
3	Kerala	2008	847.7	6599810	5867	4430	14.6	6.6	17113	6	19	47
4	Rajasthan	2008	189.5	698873	6352	5902	27.5	6.8	15626	7	29	130
5	Sikkim	2008	83.0	91256	NA	26	18.4	5.2	3262	20	27	32
6	Delhi	2009	10924.0	196209	2878	6302	18.1	4.4	176	6	50	252
7	Karnataka	2009	310.4	1828754	8541	6953	19.5	7.2	36150	10	27	85
8	Kerala	2009	851.8	6417568	5907	4860	14.7	6.8	17240	5	15	46
9	Rajasthan	2009	193.0	642862	6811	6490	27.2	6.6	15839	6	27	129
10	Sikkim	2009	84.0	92165	46	29	18.1	5.7	3242	20	27	32
11	Delhi	2010	11122.0	249463	3920	6747	17.8	4.2	177	5	55	261

The dataset contains:

- Two categorical variables:
  - **State**
  - **Year**
- Ten numerical variables:
  - **Population.Density..persons.sq.km** – Population per unit area.
  - **People.with.air.borne.diseases** – Count of people admitted in hospital due to air borne diseases.
  - **Industries** – Count of Industries.
  - **Vehicles** – Count of vehicles in thousands.
  - **Birth-rate** - Live births per thousand populations per year.
  - **Death rate** - Number of deaths per thousand population per year
  - **Forest.cover** – Land area that is covered by forests or the forest canopy or open woodland in square kilometres.
  - **SO2** – Concentration of Sulphur dioxide
  - **N02** – Concentration of Nitrogen dioxide
  - **PM** – Particulate Matter Concentrations.

The complete dataset has 25 rows and 12 columns and is stored in a csv file named pollution.csv.

Import The Data Into Mongodb

MongoDB provides the mongoimport utility that can be used to import JSON, CSV or TSV files into a MongoDB database.
The pollution.csv file is imported into the MongoDB database using the command shown in the figure below.

```
C:\Users\Pooja Naik>cd C:\Program Files\MongoDB\Server\4.2\bin

C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db Pollutionanalysis --collection Pollution --type csv --file D:\pollution.csv --headerline
```

--db: Specifies the database to use. We used a database called Pollutionanalysis.
--collection: Specifies the collection name. We used a collection called Pollution.
--type: The input to import.
--headerline: Specifies that the first row in the imported file should be the field names.

Design A Data Analysis Program In R Using RStudio To Create Some Visual Representation Of The Data

To use MongoDB with R, we can use the mongolite package which is a fast and simple MongoDB client for R.

Package '**mongolite**'
Description: High performance MongoDB client based on 'mongo-c-driver' and 'jsonlite'. Includes support for aggregation, indexing, map-reduce, streaming, encryption, enterprise authentication and GridFS.
The code in the following figure establishes a connection to MongoDB collection, retrieves the required data and stores it in an object called x.

```
1 install.packages("mongolite")
2 library(mongolite)
3 con<-mongo(collection = "Pollution",db="Pollutionanalysis",url ="mongodb://localhost",verbose = FALSE )
4 x<-con$find()
5 rm(con)
6 gc()
7
8
```

The function **mongo()** supports the following arguments:

- **Collection**: name of the collection to connect to.
- **db**: name of the database to connect to.
- **url**: address of the MongoDB server in standard URI Format.

- **Verbose**: if TRUE, emits some extra output.

The function returns a mongo connection object that acts as a pointer to a collection on the server.
The **find** method automatically simplifies the collection into a data frame. In the above code, it stores the data frame into an object called x.

rm(con) followed by **gc()** automatically disables the connection to MongoDB server when all the objects using the connection are removed.
The code in the following figure uses arrange function of dplyr package to reorder the rows in an appropriate order.

```
1  library(dplyr)
2  x<-arrange(x,x$Year,x$State)
3
4
```

R Studio provides several functionalities to familiarize with the data set.
class() – Helps in recognizing the class of the data set. **names()** – This function helps in checking out all the variables in the data set.
str() – This helps in understanding the structure of the data set, data type of each attribute and number of rows and columns present in the data.
Summary() – One of the most important functions that help in summarizing each attribute in the dataset.

```
> class(x)
[1] "data.frame"
> names(x)
 [1] "State"                          "Year"                           "Population.Density..persons.sq.km."
 [4] "People.with.air.borne.diseases" "Industries"                     "Vehicles"
 [7] "Birthrate...."                  "Deathrate...."                  "Forest.cover"
[10] "SO2"                            "NO2"                            "PM"
> str(x)
'data.frame':   25 obs. of  12 variables:
 $ State                          : Factor w/ 5 levels "Delhi","Karnataka",..: 1 2 3 4 5 1 2 3 4 5 ...
 $ Year                           : int  2008 2008 2008 2008 2008 2009 2009 2009 2009 2009 ...
 $ Population.Density..persons.sq.km.: num  10726 306 848 190 83 ...
 $ People.with.air.borne.diseases : int  182256 1777839 6599810 698873 91256 196209 1828754 6417568 642862 92165 ...
 $ Industries                     : int  3026 8451 5867 6352 NA 2878 8541 5907 6811 46 ...
 $ Vehicles                       : int  5899 6217 4430 5902 26 6302 6953 4860 6490 29 ...
 $ Birthrate....                  : num  18.4 19.8 14.6 27.5 18.4 18.1 19.5 14.7 27.2 18.1 ...
 $ Deathrate....                  : num  4.8 7.4 6.6 6.8 5.2 4.4 7.2 6.8 6.6 5.7 ...
 $ Forest.cover                   : int  176 36105 17113 15626 3262 176 36150 17240 15839 3242 ...
 $ SO2                            : int  7 13 6 7 20 6 10 5 6 20 ...
 $ NO2                            : int  57 28 19 29 27 50 27 15 27 27 ...
 $ PM                             : int  214 80 47 130 32 252 85 46 129 32
```

```
> summary(x)
       State        Year      Population.Density..persons.sq.km. People.with.air.borne.diseases   Industries      Vehicles     Birthrate....
 Delhi    :5   Min.   :2008   Min.   :   83.0                     Min.   :  91110               Min.   :   46   Min.   :   26   Min.   :14.60
 Karnataka:5   1st Qu.:2009   1st Qu.:  193.0                     1st Qu.: 196209               1st Qu.: 3643   1st Qu.: 4860   1st Qu.:17.30
 Kerala   :5   Median :2010   Median :  314.7                     Median : 698873               Median : 6582   Median : 6302   Median :18.10
 Rajasthan:5   Mean   :2010   Mean   : 2514.8                     Mean   :1725499               Mean   : 5907   Mean   : 5634   Mean   :19.27
 Sikkim   :5   3rd Qu.:2011   3rd Qu.:  860.0                     3rd Qu.:1791660               3rd Qu.: 8446   3rd Qu.: 7228   3rd Qu.:19.50
               Max.   :2012   Max.   :11518.0                     Max.   :6599810               Max.   :12250   Max.   :10910   Max.   :27.50
                                                                                                NA's   :1
  Deathrate....   Forest.cover        SO2              NO2              PM
 Min.   :4.20   Min.   :  176   Min.   : 4.00   Min.   :13.0   Min.   : 32.0
 1st Qu.:5.40   1st Qu.: 3262   1st Qu.: 5.00   1st Qu.:21.0   1st Qu.: 42.0
 Median :6.60   Median :15626   Median : 7.00   Median :27.0   Median : 80.0
 Mean   :6.12   Mean   :14469   Mean   : 9.56   Mean   :30.2   Mean   :109.3
 3rd Qu.:7.00   3rd Qu.:17324   3rd Qu.:11.00   3rd Qu.:31.0   3rd Qu.:168.0
 Max.   :7.40   Max.   :36150   Max.   :20.00   Max.   :59.0   Max.   :261.0
```

For a better analysis, we have defined a function called percentdata that converts all the values of the data set into the same range. The new values are stored in a data frame called percentdataset.

```
#Function that converts the values of a given column in the range of 100
percentdata <- function(b)
{
        maxa<-max(x[,b] ,na.rm=TRUE)
        ab<-((as.numeric(x[,b])/maxa)*100)
        ab
}

#Create a new data frame called percent data that has 1st(state) and 2nd(year) column same as main dataset
percentdataset<-data.frame(x$State,x$Year)

#The values of 3rd cloumn will be converted to the range of 100 by percentdata function
#store it in the 3rd column of percentdataset
percentdataset[,3]<-percentdata(3)

#Similarly for the rest of the columns
percentdataset[,4]<-percentdata(4)


percentdataset[,5]<-percentdata(5)
percentdataset[,6]<-percentdata(6)
percentdataset[,7]<-percentdata(7)
percentdataset[,8]<-percentdata(8)
percentdataset[,9]<-percentdata(9)
percentdataset[,10]<-percentdata(10)
percentdataset[,11]<-percentdata(11)
percentdataset[,12]<-percentdata(12)

#Assign the column names of original dataset to the new dataset
names(percentdataset)<-c(names(x[,1:12]))
```

Here is the snapshot of the first few rows of percentdataset.

	State	Year	Population.Density..persons.sq.km.	People.with.air.borne.diseases	Industries	Vehicles	Birthrate....	Deathrate....	Forest.cover	SO2	NO2	PM
1	Delhi	2008	93.1238062	2.761534	24.7020408	54.0696609	66.90909	64.86486	0.4868603	35	96.61017	81.99234
2	Karnataka	2008	2.6575794	26.937730	68.9877551	56.9844180	72.00000	100.00000	99.8755187	65	47.45763	30.65134
3	Kerala	2008	7.3597847	100.000000	47.6938776	40.6049496	53.09091	89.18919	47.3388658	30	32.20339	18.00766
4	Rajasthan	2008	1.6452509	10.589290	51.8530612	54.0971586	100.00000	91.89189	43.2254495	35	49.15254	49.80643
5	Sikkim	2008	0.7206112	1.382706	NA	0.2383135	66.90909	70.27027	9.0235131	100	45.76271	12.26054
6	Delhi	2009	94.8428547	2.972949	23.4938776	57.7635197	65.81818	59.45946	0.4868603	30	84.74576	96.55172
7	Karnataka	2009	2.6949123	27.709192	69.7224490	63.7305225	70.90909	97.29730	100.0000000	50	45.76271	32.56705
8	Kerala	2009	7.3953811	97.238678	48.2204082	44.5462878	53.45455	91.89189	47.6901798	25	25.42373	17.62452
9	Rajasthan	2009	1.6756381	9.740614	55.6000000	59.4867094	98.90909	89.18919	43.8146611	30	45.76271	49.42529
10	Sikkim	2009	0.7292933	1.396480	0.3755102	0.2658112	65.81818	77.02703	8.9681881	100	45.76271	12.26054
11	Delhi	2010	96.5619031	3.779851	32.0000000	61.8423465	64.72727	56.75676	0.4896266	25	93.22034	100.0000
12	Karnataka	2010	2.7322452	22.321127	87.5265306	82.8964253	69.81818	95.94595	99.9806362	50	37.28814	26.81992

A function named pollution_level is defined to analyze the air pollution level in each state during a particular year using the factors that directly or indirectly play a key role in the pollution level of a place.

```
pollution_level<-function(c)
{

        (sum(percentdataset[c,c(3,4,5,6,10,11,12)],na.rm = TRUE)-percentdataset[c,9])
}

pollution_dataset<-data.frame(percentdataset$State,percentdataset$Year)
i=1
for(i in 1:25)
{
        pollution_dataset[i,3]<-pollution_level(i)
        i=i+1
}
names(pollution_dataset) <- c("State","Year","Pollution_level")
```

A new data frame named pollution_dataset is created to store the generated pollution level value of each state during a particular year.

When conducting data analysis, plotting is critically important. The core plotting and graphics engine in R is encapsulated in the following packages:

- **Graphics**: Contains plotting functions for the “base” graphing systems, including plot, hist, boxplot and many others.

- **grDevices**: Contains all the code implementing the various graphics devices, including X11, PDF, PostScript, PNG etc.

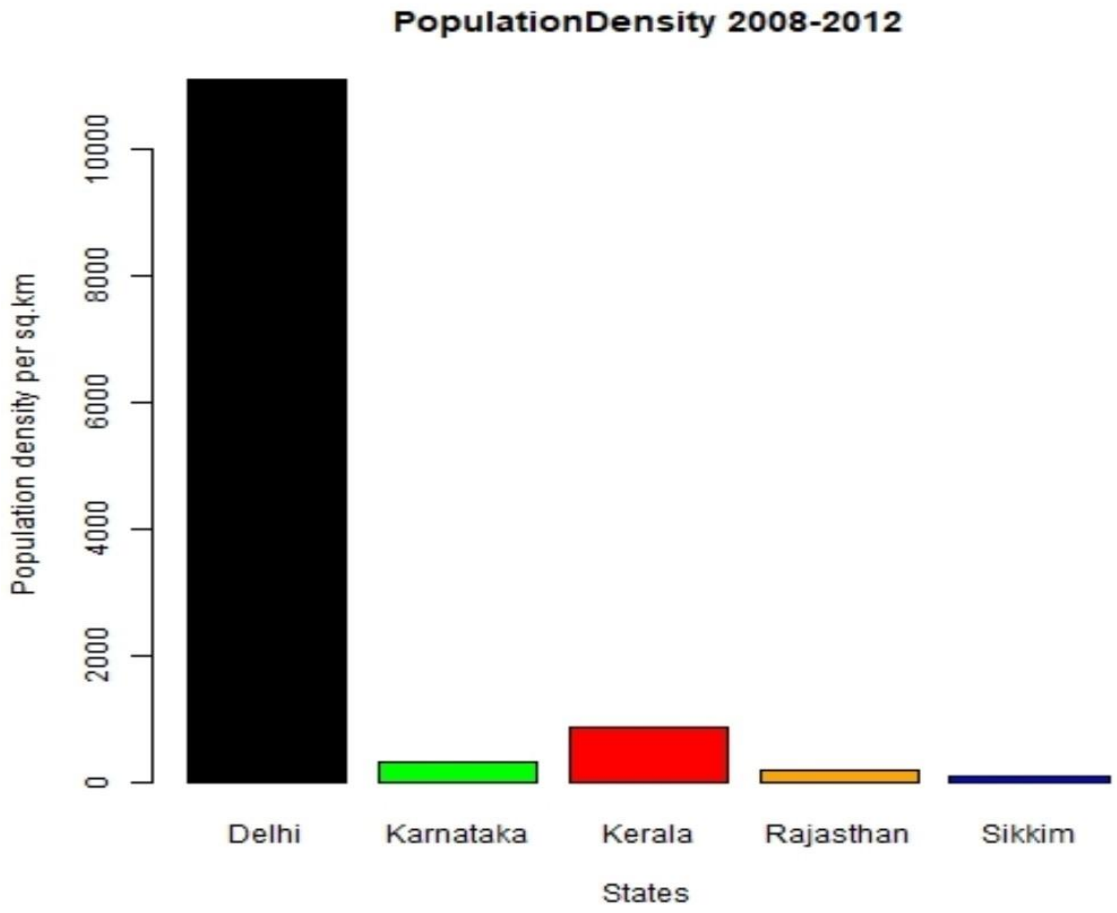
Present The Results Of The Analysis Using Web Pages

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs and maps, data visualization tools provide an accessible way to see and understand trends, outliers and patterns in data.
We've designed web pages to present the results and conclusions of our analysis in a better way. The web pages contain all the visual elements along with a brief conclusion of what they represent which makes the whole process of analysis more feasible.

RESULTS

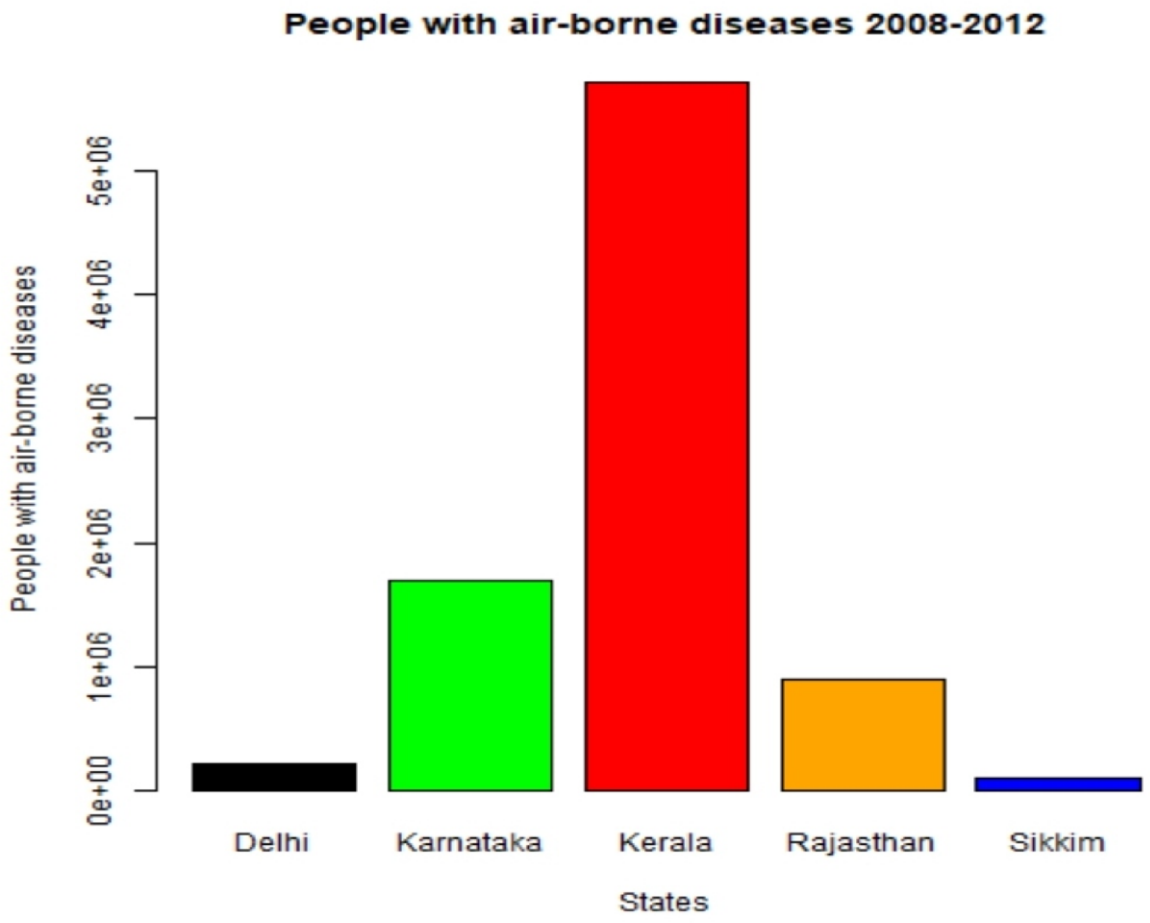
The following bar graphs show the variation of each factor in all five states during the period 2008 to 2012.

FIGURE 1



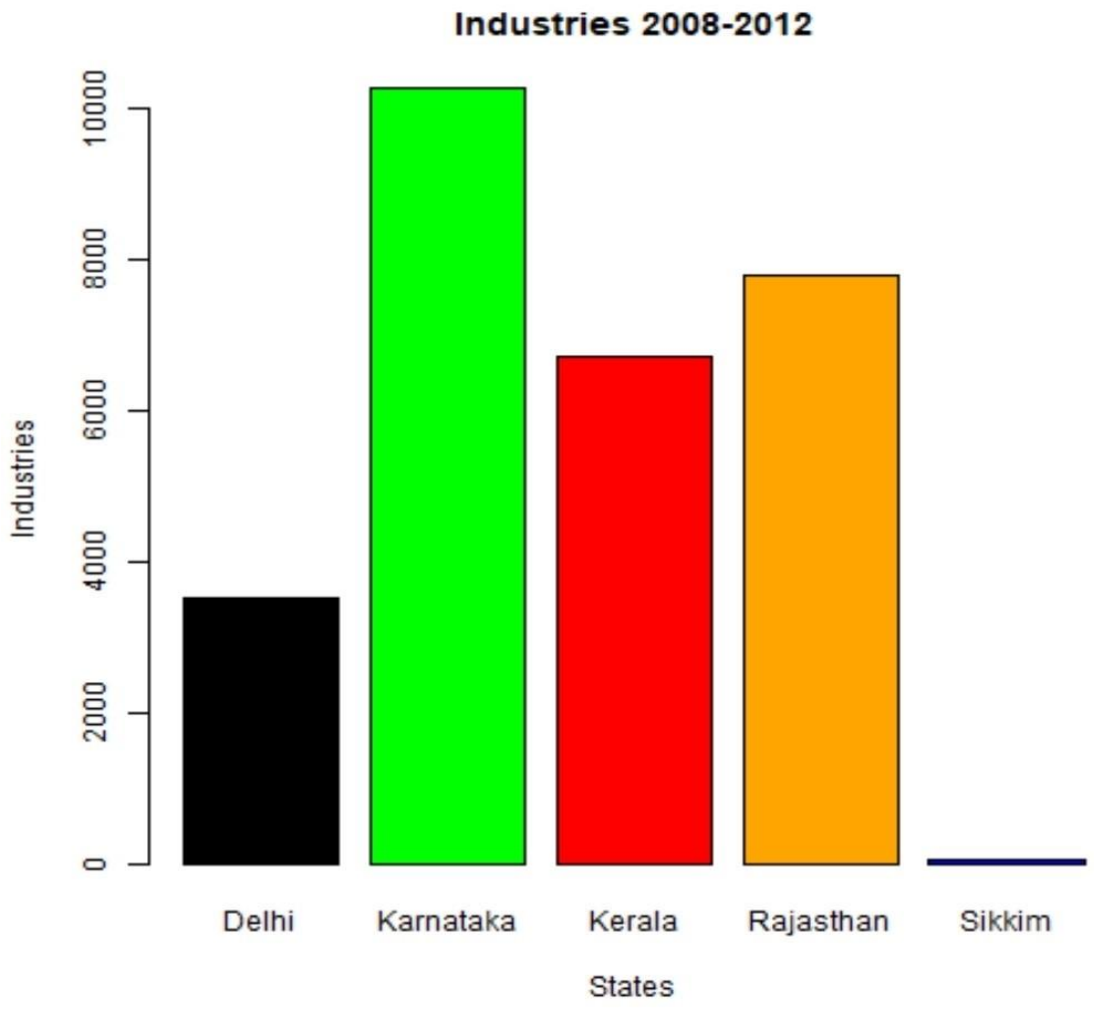

```
#Mean of population of all states from 2008-2012
karnatakapopulation<-c(karnataka$Population.Density..persons.sq.km.)
karnatakapopulationmean<-mean(karnatakapopulation)

keralapopulation<-c(kerala$Population.Density..persons.sq.km.)
keralapopulationmean<-mean(keralapopulation)

delhipopulation<-c(delhi$Population)
delhipopulationmean<-mean(delhipopulation)

rajasthanpopulation<-c(rajasthan$Population)
rajasthanpopulationmean<-mean(rajasthanpopulation)

sikkimpopulation<-c(sikkim$Population)
sikkimpopulationmean<-mean(sikkimpopulation,na.rm = TRUE)

statepopulation<-c(delhipopulationmean,karnatakapopulationmean,keralapopulationmean,rajasthanpopulationmean,sikkimpopulationmean)

#barchart of population
colors=c("black","green","red","orange","blue")
barplot(statepopulation,main="PopulationDensity 2008-2012",ylab="Population density per sq.km ",xlab="States",
names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),col =colors)
```

FIGURE 2



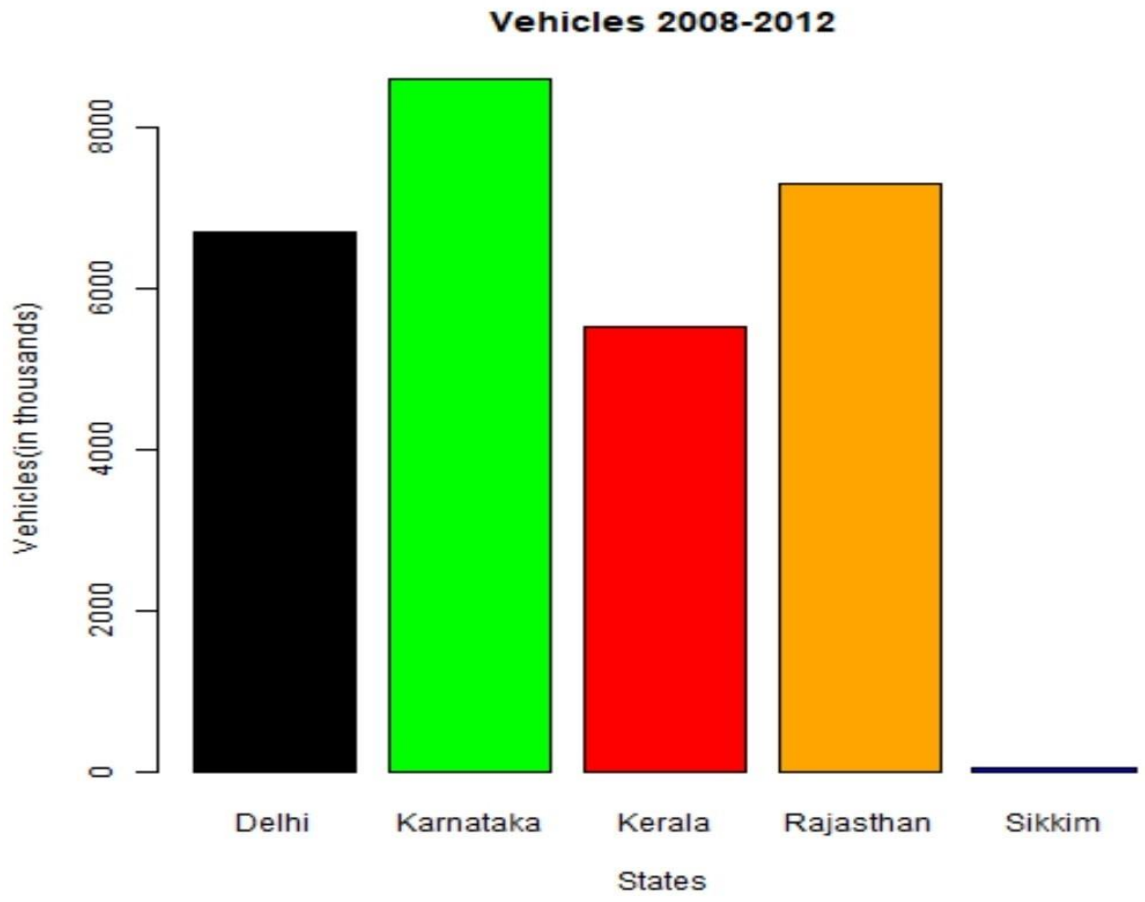

```
#Mean of airborndiseases of all states from 2008-2012
karnatakaairbornediseases<-c(karnataka$People.with.air.borne.diseases)
karnatakaairbornediseasesmean<-mean(karnatakaairbornediseases )

keralaairbornediseases<-c(kerala$People.with.air.borne.diseases)
keralaairbornediseasesmean<-mean(keralaairbornediseases)

delhiairbornediseases<-c(delhi$People.with.air.borne.diseases)
delhiairbornediseasesmean<-mean(delhiairbornediseases)

rajasthanairbornediseases<-c(rajasthan$People.with.air.borne.diseases)
rajasthanairbornediseasesmean<-mean(rajasthanairbornediseases)

sikkimairbornediseases<-c(sikkim$People.with.air.borne.diseases)
sikkimairbornediseasesmean<-mean(sikkimairbornediseases,na.rm = TRUE)

stateairbornediseases<-c(delhiairbornediseasesmean,karnatakaairbornediseasesmean,keralaairbornediseasesmean,rajasthanairbornediseasesmean,sikkimairbornedise

#barchart of airbornediseases
colors=c("black","green","red","orange","blue")
barplot(stateairbornediseases,main="People with air-borne diseases 2008-2012",ylab="People with air-borne diseases",xlab="States",
names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),col =colors)
```

FIGURE 3



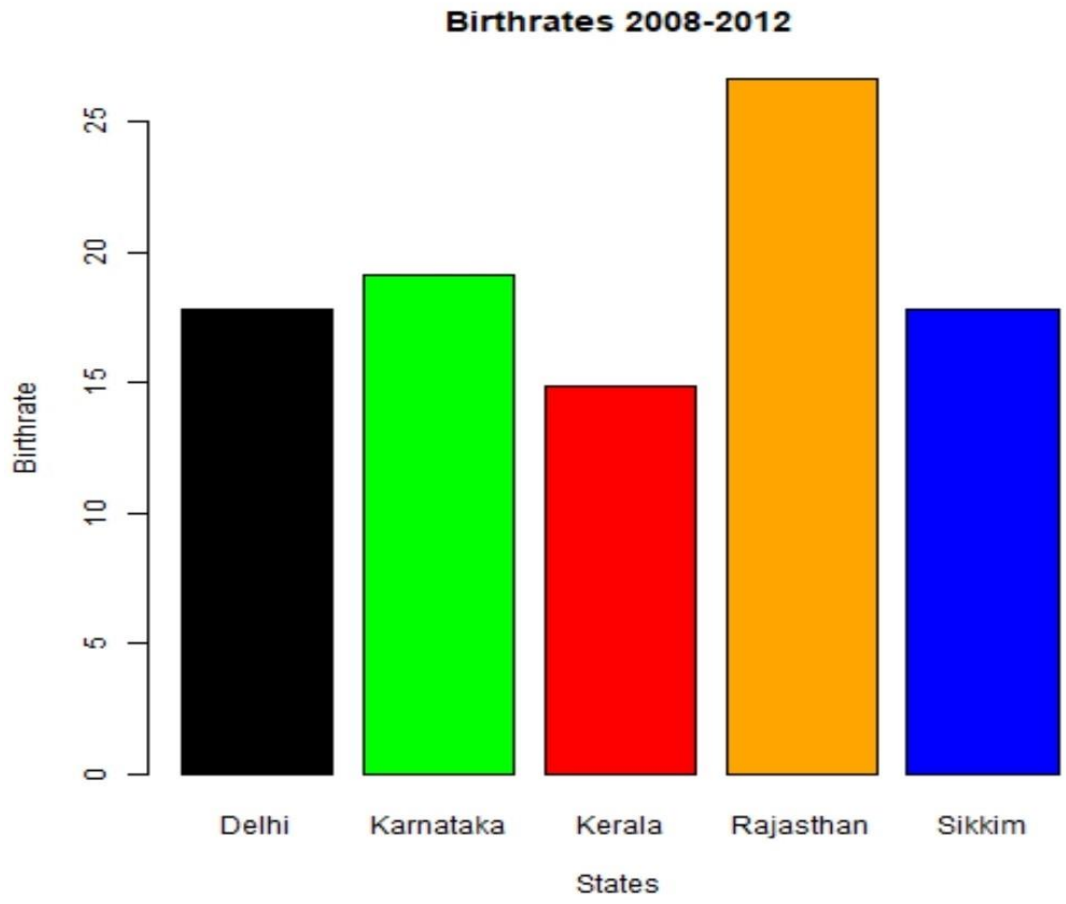

```
#Mean of industries of all states from 2008-2012
karnatakaindustries<-c(karnataka$Industries)
karnatakaindustriesmean<-mean(karnatakaindustries)

keralaindustries<-c(kerala$Industries)
keralaindustriesmean<-mean(keralaindustries)

delhiindustries<-c(delhi$Industries)
delhiindustriesmean<-mean(delhiindustries)

rajasthanindustries<-c(rajasthan$Industries)
rajasthanindustriesmean<-mean(rajasthanindustries)

sikkimindustries<-c(sikkim$Industries)
sikkimindustriesmean<-mean(sikkimindustries,na.rm = TRUE)

stateindustries<-c(delhiindustriesmean,karnatakaindustriesmean,keralaindustriesmean,rajasthanindustriesmean,sikkimindustriesmean)

#barchart of industries
colors=c("black","green","red","orange","blue")
barplot(stateindustries,main="Industries 2008-2012",xlab="States",ylab="Industries",names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),
col =colors)
```

FIGURE 4



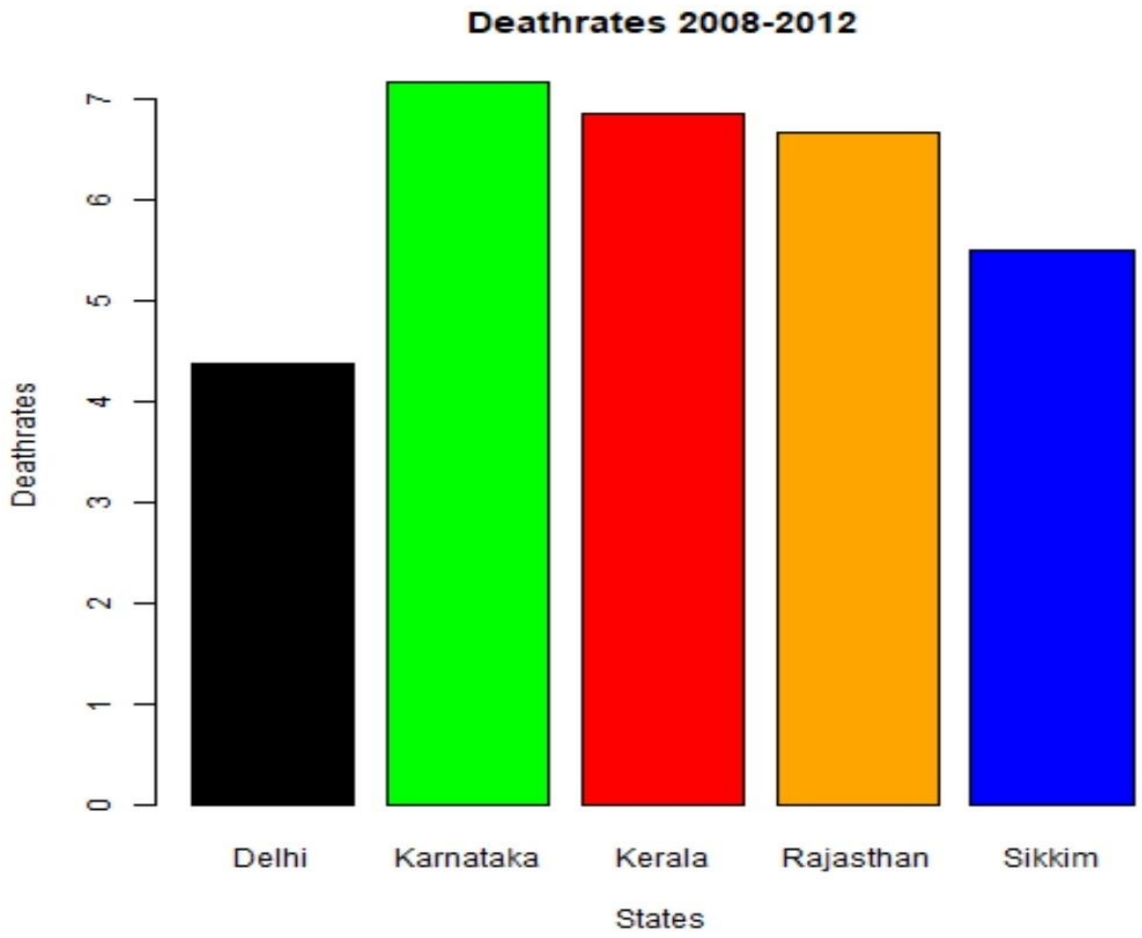

```
#Mean of vehicles of all states from 2008-2012
karnatakavehicles<-c(karnataka$Vehicles)
karnatakavehiclesmean<-mean(karnatakavehicles)

keralavehicles<-c(kerala$Vehicles)
keralavehiclesmean<-mean(keralavehicles)

delhivehicles<-c(delhi$Vehicles)
delhivehiclesmean<-mean(delhivehicles)

rajasthanvehicles<-c(rajasthan$Vehicles)
rajasthanvehiclesmean<-mean(rajasthanvehicles)

sikkimvehicles<-c(sikkim$Vehicles)
sikkimvehiclesmean<-mean(sikkimvehicles,na.rm = TRUE)

statevehicles<-c(delhivehiclesmean,karnatakavehiclesmean,keralavehiclesmean,rajasthanvehiclesmean,sikkimvehiclesmean)

#barchart of vehicles
colors=c("black","green","red","orange","blue")
barplot(statevehicles,main="Vehicles 2008-2012",ylab="Vehicles(in thousands)",xlab="States",names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),
col =colors)
```

FIGURE 5



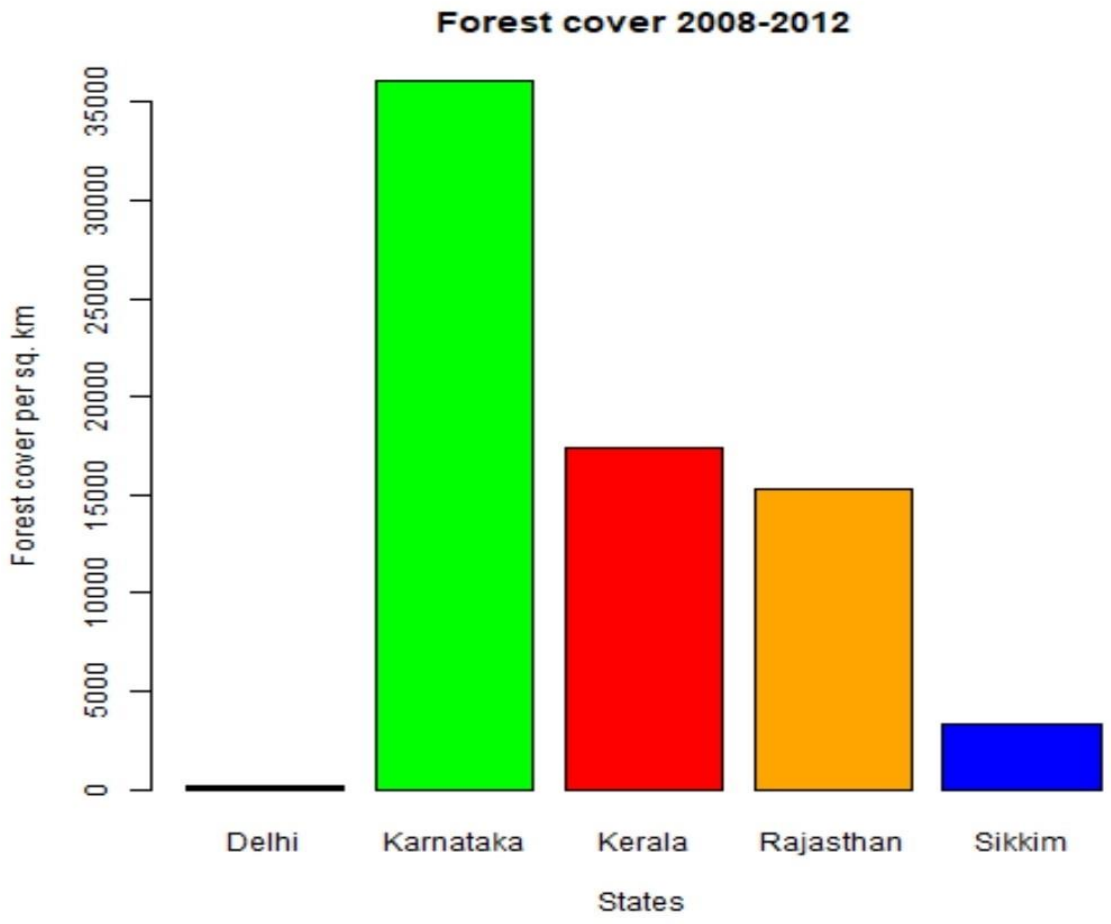

```
#Mean of birthrates of all states from 2008-2012
karnatakabirthrates<-c(karnataka$Birthrate....)
karnatakabirthratesmean<-mean(karnatakabirthrates)

keralabirthrates<-c(kerala$Birthrate....)
keralabirthratesmean<-mean(keralabirthrates)

delhibirthrates<-c(delhi$Birthrate....)
delhibirthratesmean<-mean(delhibirthrates)

rajasthanbirthrates<-c(rajasthan$Birthrate....)
rajasthanbirthratesmean<-mean(rajasthanbirthrates)

sikkimbirthrates<-c(sikkim$Birthrate....)
sikkimbirthratesmean<-mean(sikkimbirthrates,na.rm = TRUE)

statebirthrates<-c(delhibirthratesmean,karnatakabirthratesmean,keralabirthratesmean,rajasthanbirthratesmean,sikkimbirthratesmean)

#barchart of birthrates
colors=c("black","green","red","orange","blue")
barplot(statebirthrates,main="Birthrates 2008-2012",xlab="States",ylab="Birthrate",names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),
col =colors)
```

FIGURE 6



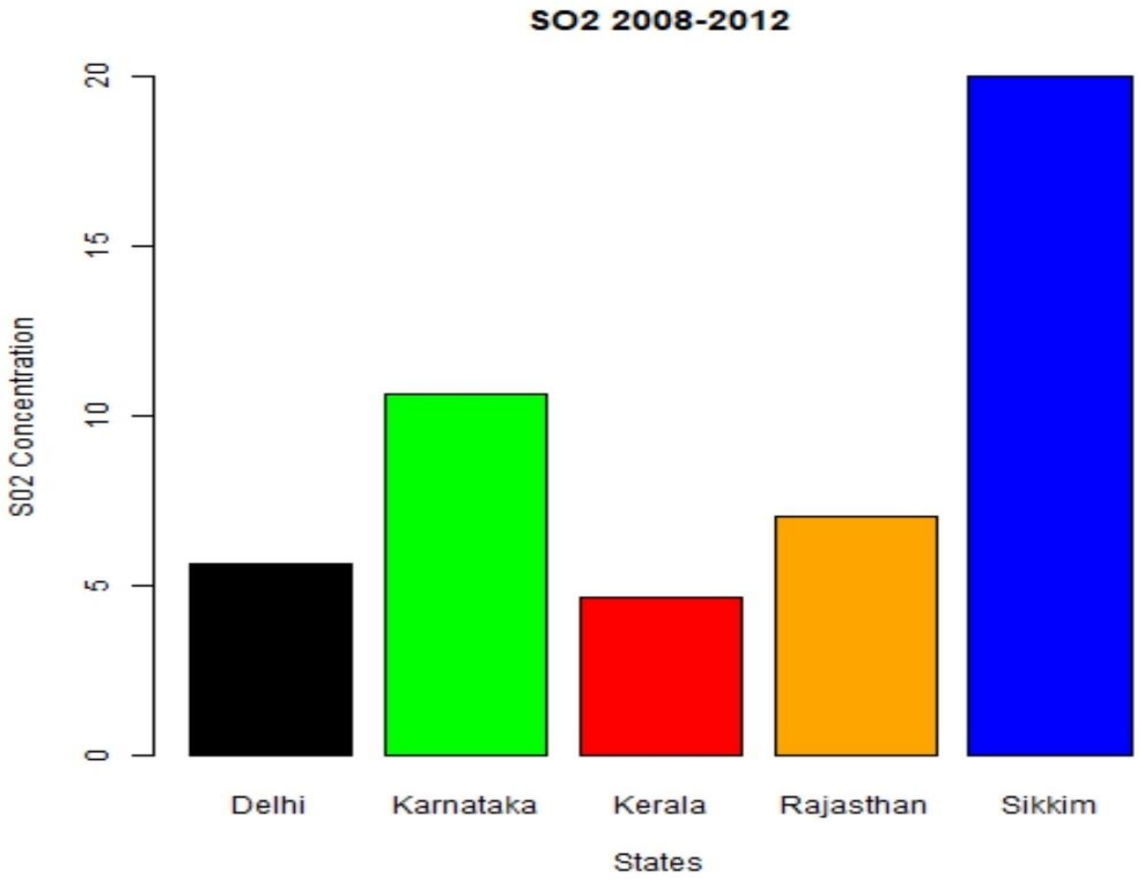

```
#Mean of deathrates of all states from 2008-2012
karnatakadeathrates<-c(karnataka$Deathrate....)
karnatakadeathratesmean<-mean(karnatakadeathrates)

keraladeathrates<-c(kerala$Deathrate....)
keraladeathratesmean<-mean(keraladeathrates)

delhideathrates<-c(delhi$Deathrate....)
delhideathratesmean<-mean(delhideathrates)

rajasthandeathrates<-c(rajasthan$Deathrate....)
rajasthandeathratesmean<-mean(rajasthandeathrates)

sikkimdeathrates<-c(sikkim$Deathrate....)
sikkimdeathratesmean<-mean(sikkimdeathrates,na.rm = TRUE)

statedeathrates<-c(delhideathratesmean,karnatakadeathratesmean,keraladeathratesmean,rajasthandeathratesmean,sikkimdeathratesmean)

#barchart of deathrates
colors=c("black","green","red","orange","blue")
barplot(statedeathrates,main="Deathrates 2008-2012",ylab="Deathrates",xlab="States",names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),
col =colors)
```

FIGURE 7



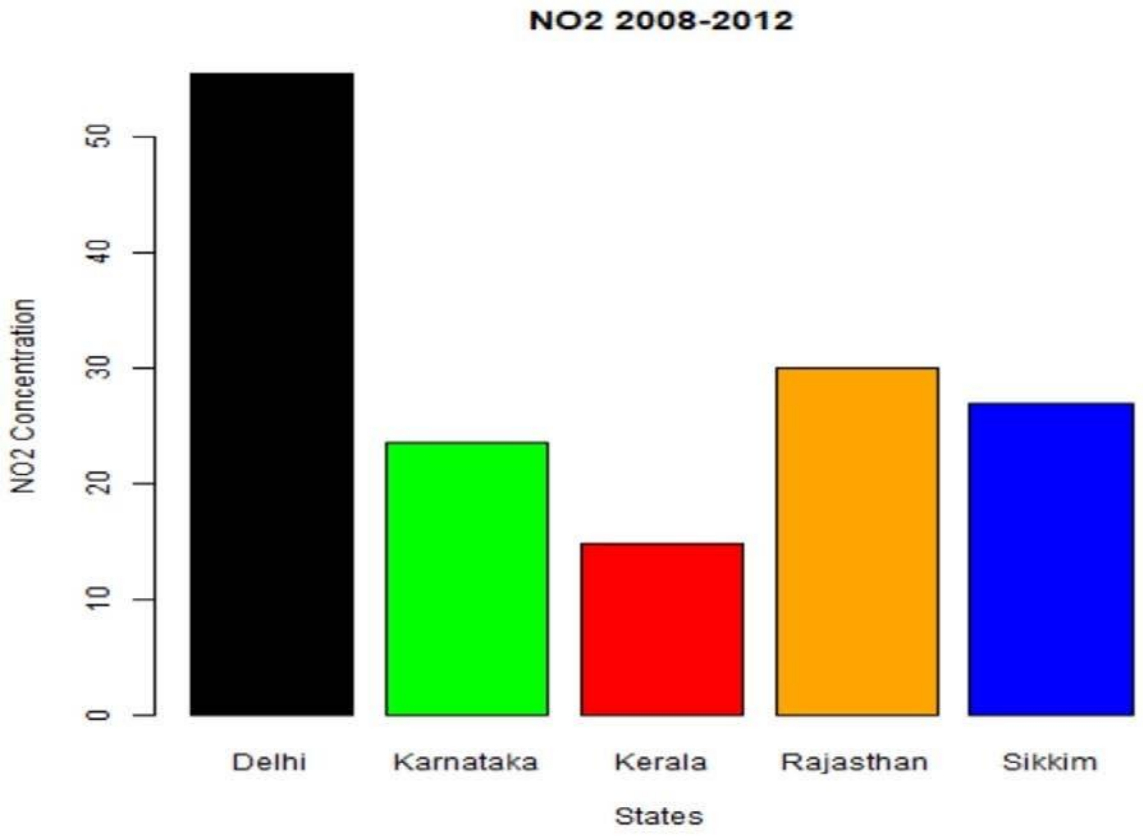

```
#Mean of forestcover of all states from 2008-2012
karnatakaforestcover<-c(karnataka$Forest.cover)
karnatakaforestcovermean<-mean(karnatakaforestcover )

keralaforestcover<-c(kerala$Forest.cover)
keralaforestcovermean<-mean(keralaforestcover)

delhiforestcover<-c(delhi$Forest.cover)
delhiforestcovermean<-mean(delhiforestcover)

rajasthanforestcover<-c(rajasthan$Forest.cover)
rajasthanforestcovermean<-mean(rajasthanforestcover)

sikkimforestcover<-c(sikkim$Forest.cover)
sikkimforestcovermean<-mean(sikkimforestcover,na.rm = TRUE)

stateforestcover<-c(delhiforestcovermean,karnatakaforestcovermean,keralaforestcovermean,rajasthanforestcovermean,sikkimforestcovermean)

#barchart of  forestcover
colors=c("black","green","red","orange","blue")
barplot(stateforestcover,main="Forest cover 2008-2012",xlab="States",ylab="Forest cover per sq. km",
names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),col =colors)
```

FIGURE 8



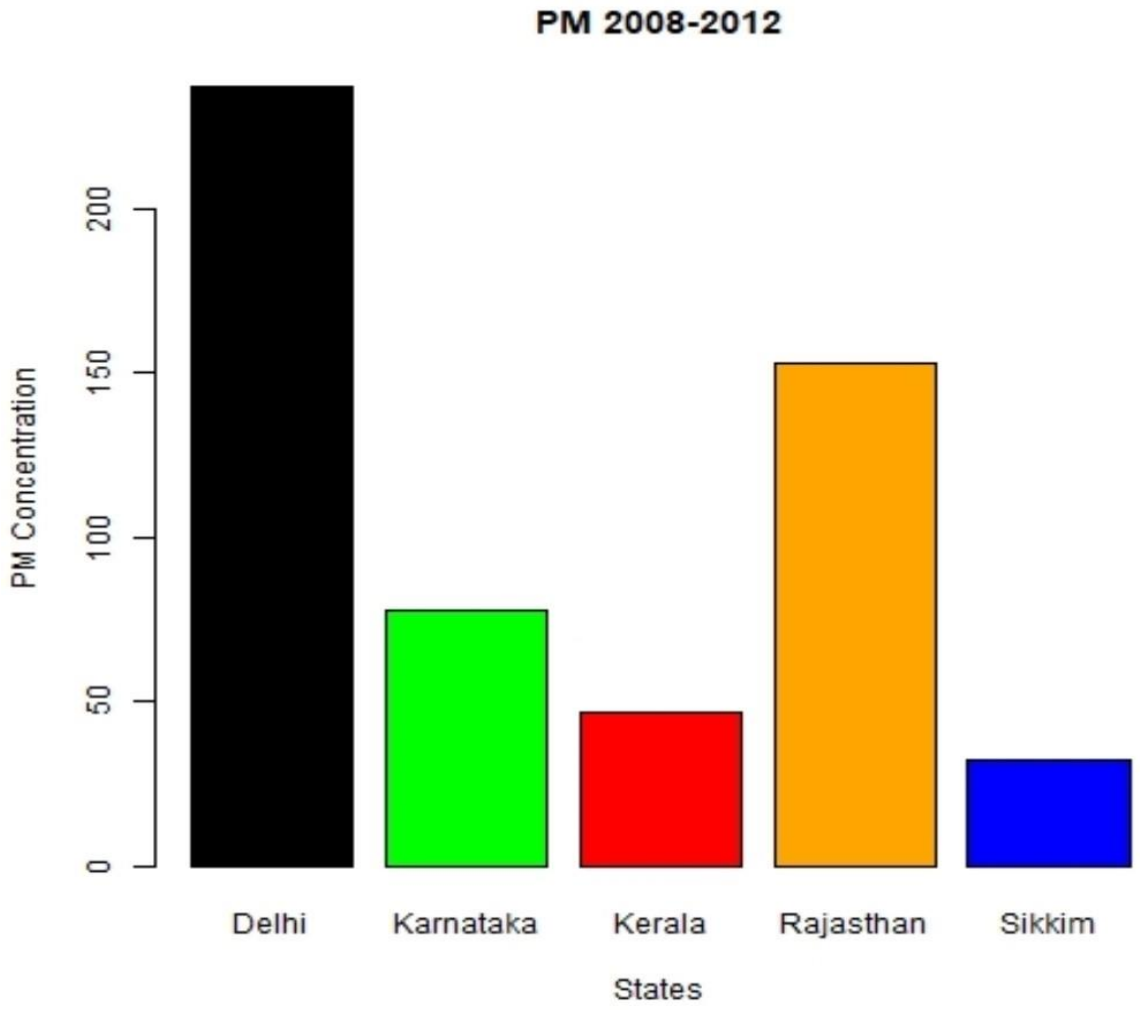

```
#Mean of sulphur of all states from 2008-2012
karnatakasulphur<-c(karnataka$SO2)
karnatakasulphurmean<-mean(karnatakasulphur)

keralasulphur<-c(kerala$SO2)
keralasulphurmean<-mean(keralasulphur)

delhisulphur<-c(delhi$SO2)
delhisulphurmean<-mean(delhisulphur)

rajasthansulphur<-c(rajasthan$SO2)
rajasthansulphurmean<-mean(rajasthansulphur)

sikkimsulphur<-c(sikkim$SO2)
sikkimsulphurmean<-mean(sikkimsulphur,na.rm = TRUE)

statesulphur<-c(delhisulphurmean,karnatakasulphurmean,keralasulphurmean,rajasthansulphurmean,sikkimsulphurmean)

#barchart of so2
colors=c("black","green","red","orange","blue")
barplot(statesulphur,main="SO2 2008-2012",xlab="States",ylab="SO2 Concentration",names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),
col =colors)
```

FIGURE 9




```
#Mean of no2 of all states from 2008-2012
karnatakan<-c(karnataka$NO2)
karnatakanmean<-mean(karnatakan )

keralan<-c(kerala$NO2)
keralanmean<-mean(keralan)

delhin<-c(delhi$NO2)
delhinmean<-mean(delhin)

rajasthann<-c(rajasthan$NO2)
rajasthannmean<-mean(rajasthann)

sikkimn<-c(sikkim$NO2)
sikkimnmean<-mean(sikkimn,na.rm = TRUE)

staten<-c(delhinmean,karnatakanmean,keralanmean,rajasthannmean,sikkimnmean)

#barchart of no2
colors=c("black","green","red","orange","blue")
barplot(staten ,main="NO2 2008-2012",xlab="States",ylab="NO2 Concentration",names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),col =colors)
```

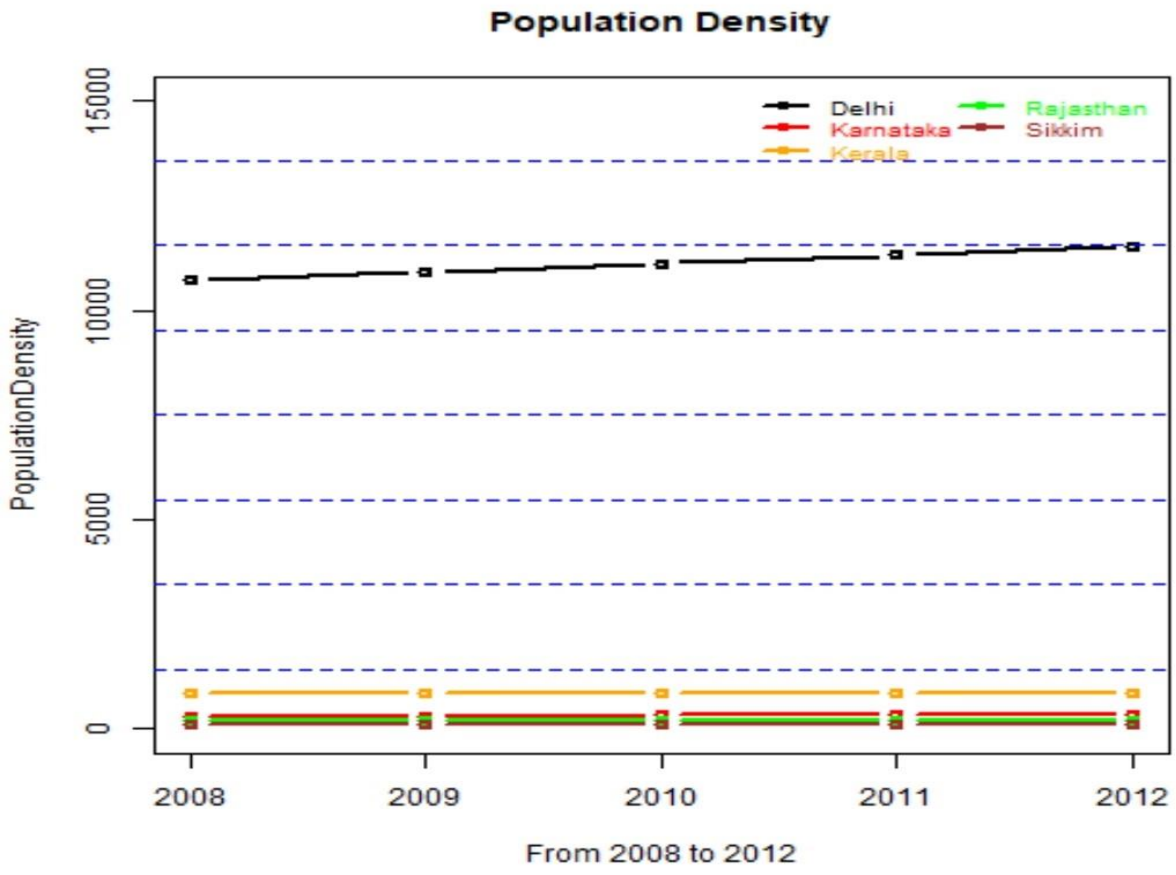
FIGURE 10




```
#Mean of pm of all states from 2008-2012
karnatakapm<-c(karnataka$PM)
karnatakapmmean<-mean(karnatakapm )

keralapm<-c(kerala$PM)
keralapmmean<-mean(keralapm)

delhipm<-c(delhi$PM)
delhipmmean<-mean(delhipm)

rajasthanpm<-c(rajasthan$PM)
rajasthanpmmean<-mean(rajasthanpm)

sikkimpm<-c(sikkim$PM)
sikkimpmmean<-mean(sikkimpm,na.rm = TRUE)

statepm<-c(delhipmmean,karnatakapmmean,keralapmmean,rajasthanpmmean,sikkimpmmean)

#barchart of pm
colors=c("black","green","red","orange","blue")
barplot(statepm,main="PM 2008-2012",xlab="States",ylab="PM Concentration",names.arg = c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),col =colors)
```

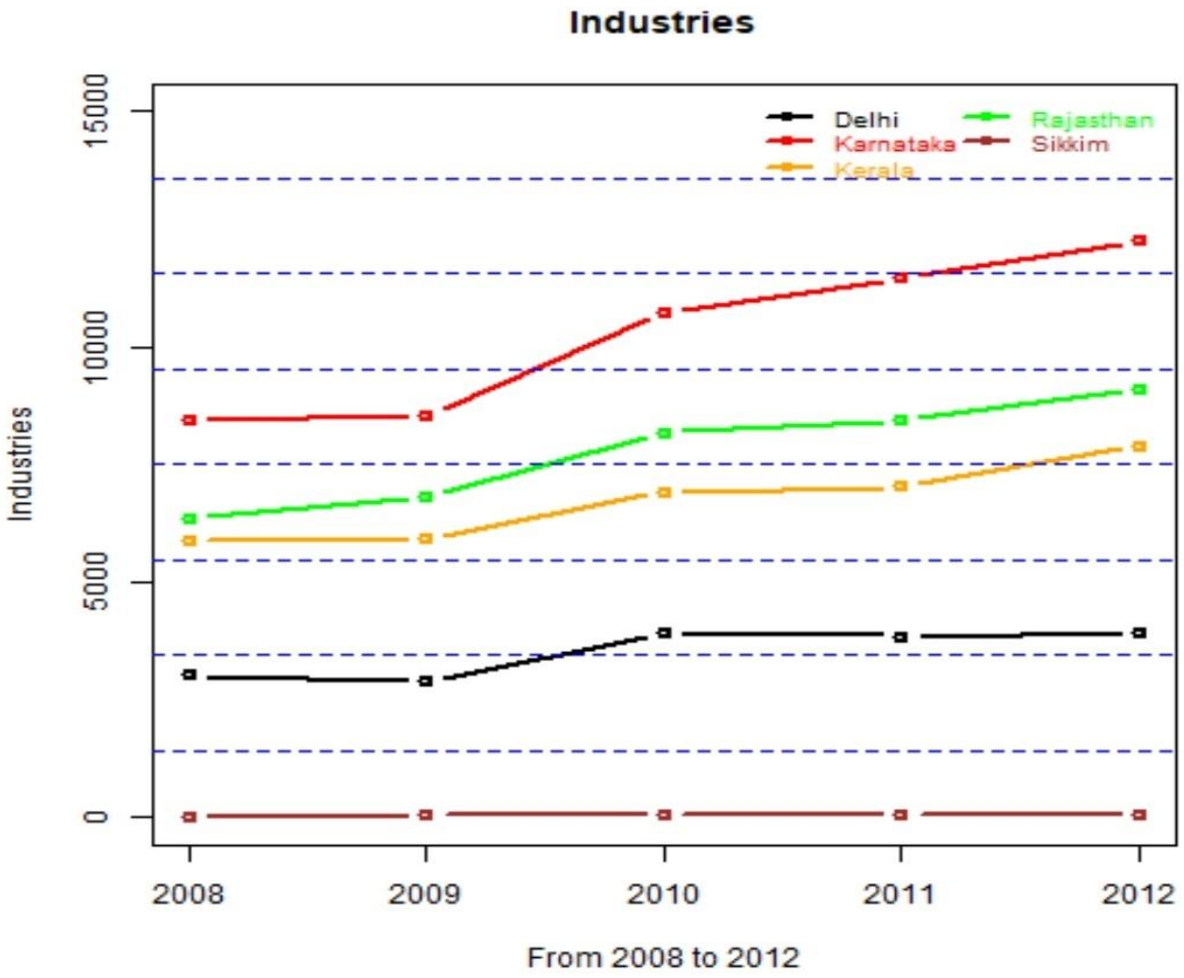
The following line graphs show the variation of each factor in all five states during the period 2008 to 2012.

FIGURE 11



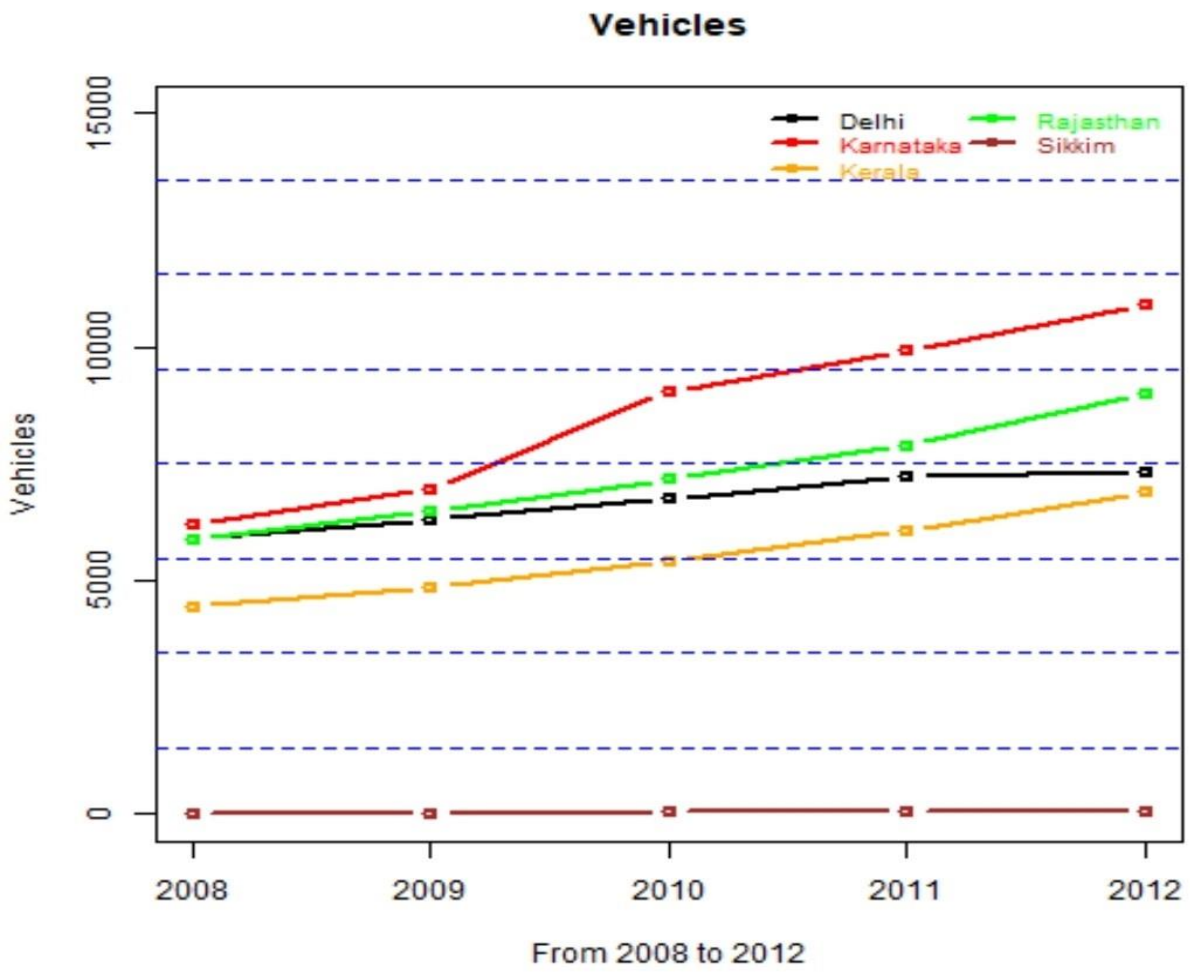

```
PopulationDensity<-x[1:25,c(1,2,3)]

#reshape the PopulationDensity dataset
nn<-reshape(PopulationDensity,timevar = "State",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(PopulationDensity$State))
nn[is.na(nn)]<-0

#plot for PopulationDensity--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,15000),col="black",xlab="From 2008 to 2012",ylab="PopulationDensity",main="Population Density")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

FIGURE 12



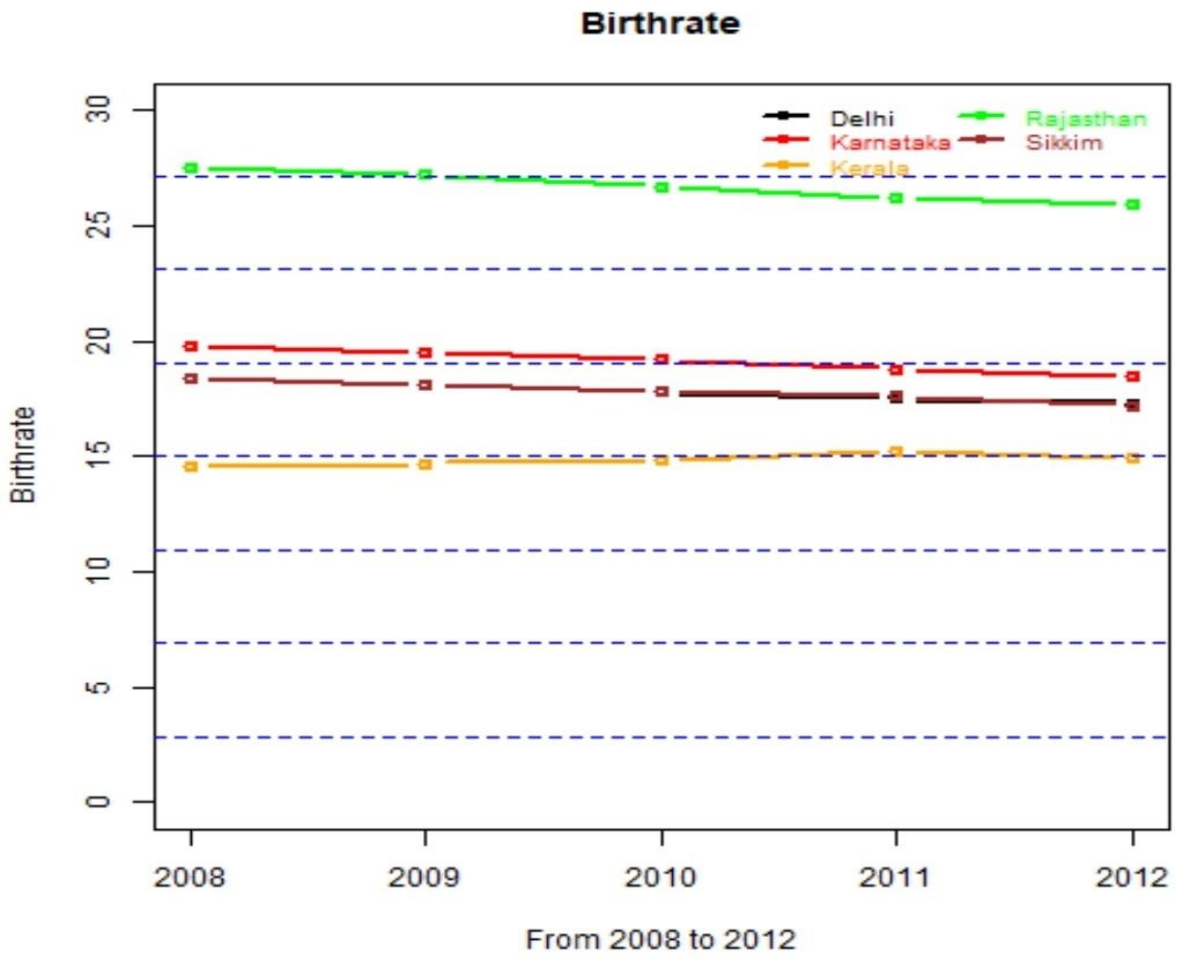

```
#subset of Industries

Industries<-x[1:25,c(1,2,5)]

#reshape the Industries dataset
nn<-reshape(Industries,timevar = "State",idvar="Year",direction = "wide")
names(nn)[c(2:6)]<-as.character(unique(Industries$State))
nn[is.na(nn)]<-0

#plot for Industries--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,15000),col="black",xlab="From 2008 to 2012",ylab="Industries",main="Industries")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

FIGURE 13



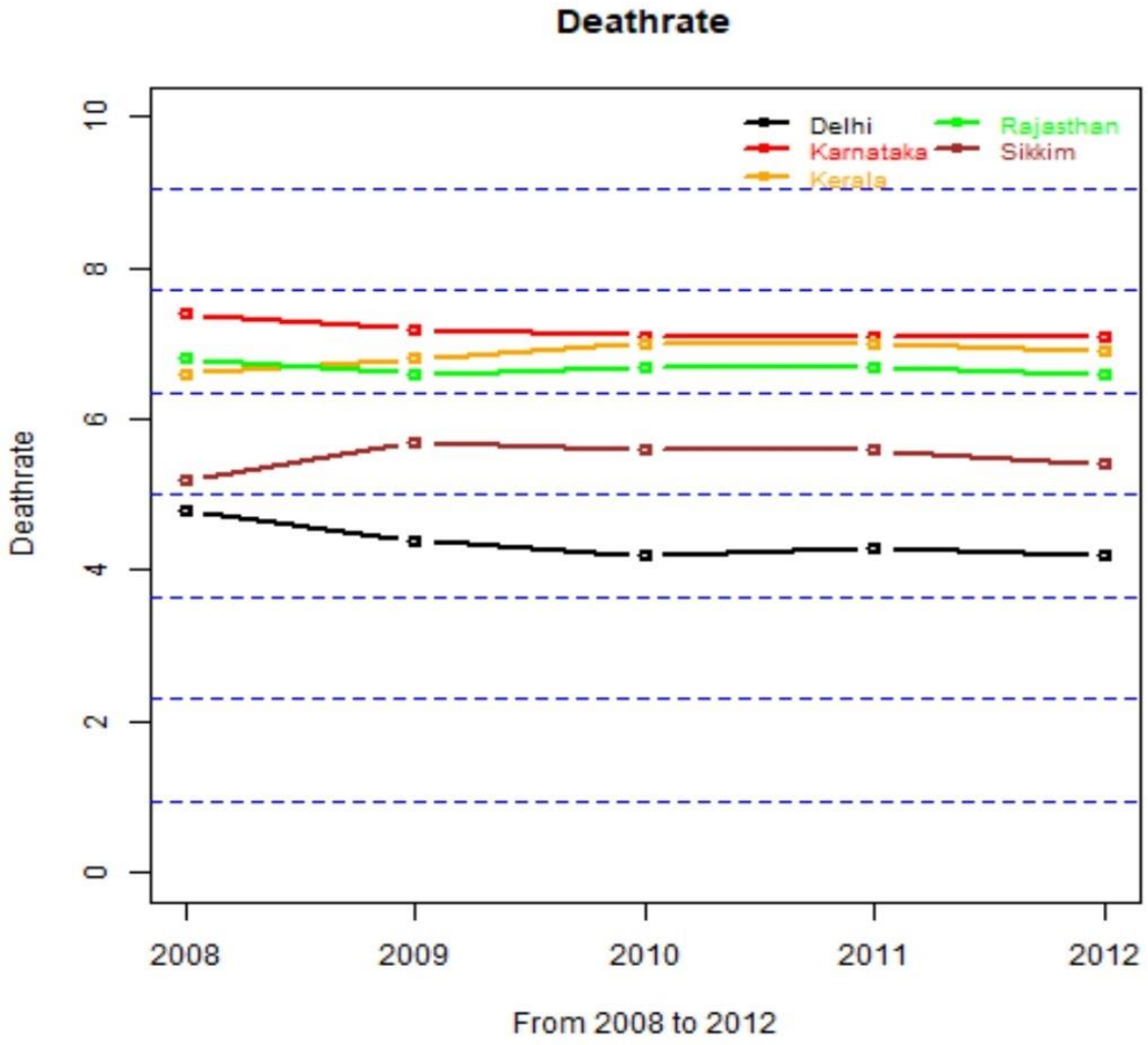

```
#subset of vehicles

Vehicles<-x[1:25,c(1,2,6)]

#reshape the Vehicles dataset
nn<-reshape(Vehicles,timevar = "State",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(Vehicles$State))
nn[is.na(nn)]<-0

#plot for Vehicles--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,15000),col="black",xlab="From 2008 to 2012",ylab="Vehicles",main="Vehicles")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
 ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

FIGURE 14



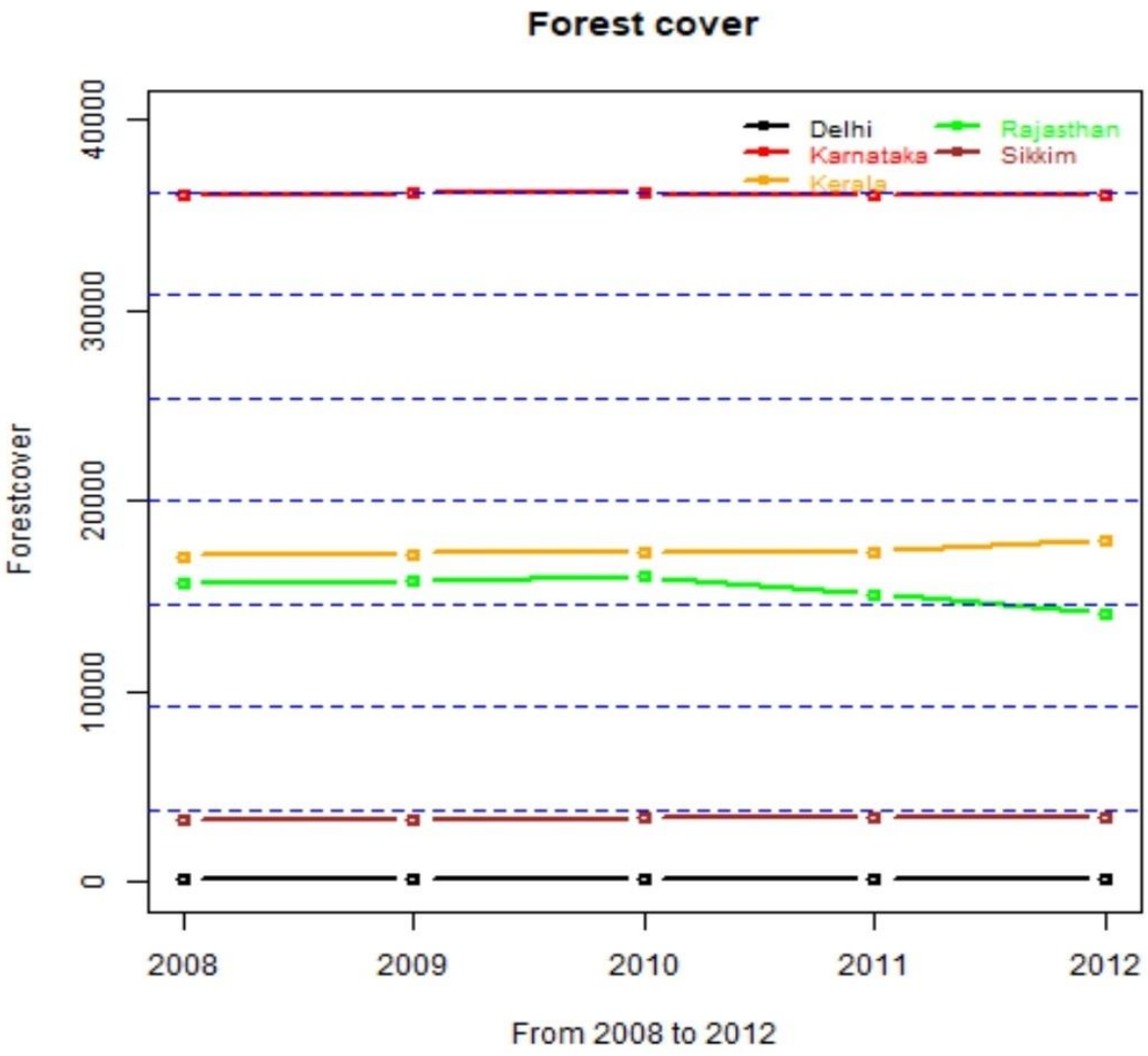

```
#subset of Birthrate

Birthrate<-x[1:25,c(1,2,7)]

#reshape the Birthrate dataset
nn<-reshape(Birthrate,timevar = "State",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(Birthrate$State))
nn[is.na(nn)]<-0

#plot forBirthrate--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,30.0),col="black",xlab="From 2008 to 2012",ylab="Birthrate",main="Birthrate")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

FIGURE 15



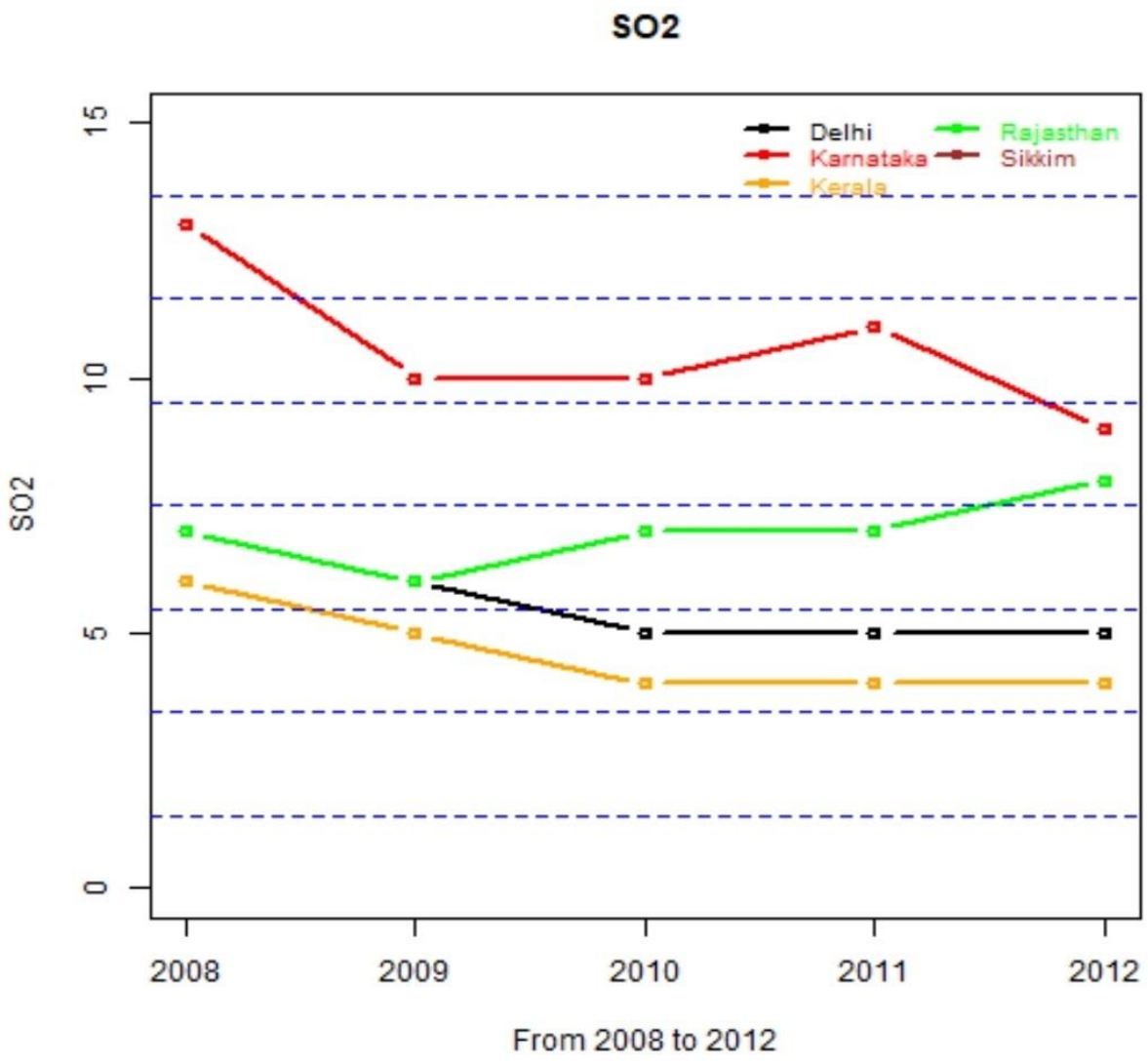

```
#subset of Deathrate

Deathrate<-x[1:25,c(1,2,8)]

#reshape the Deathrate dataset
nn<-reshape(Deathrate,timevar = "state",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(Deathrate$state))
nn[is.na(nn)]<-0

#plot for Deathrate--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,10.0),col="black",xlab="From 2008 to 2012",ylab="Deathrate",main="Deathrate")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

FIGURE 16



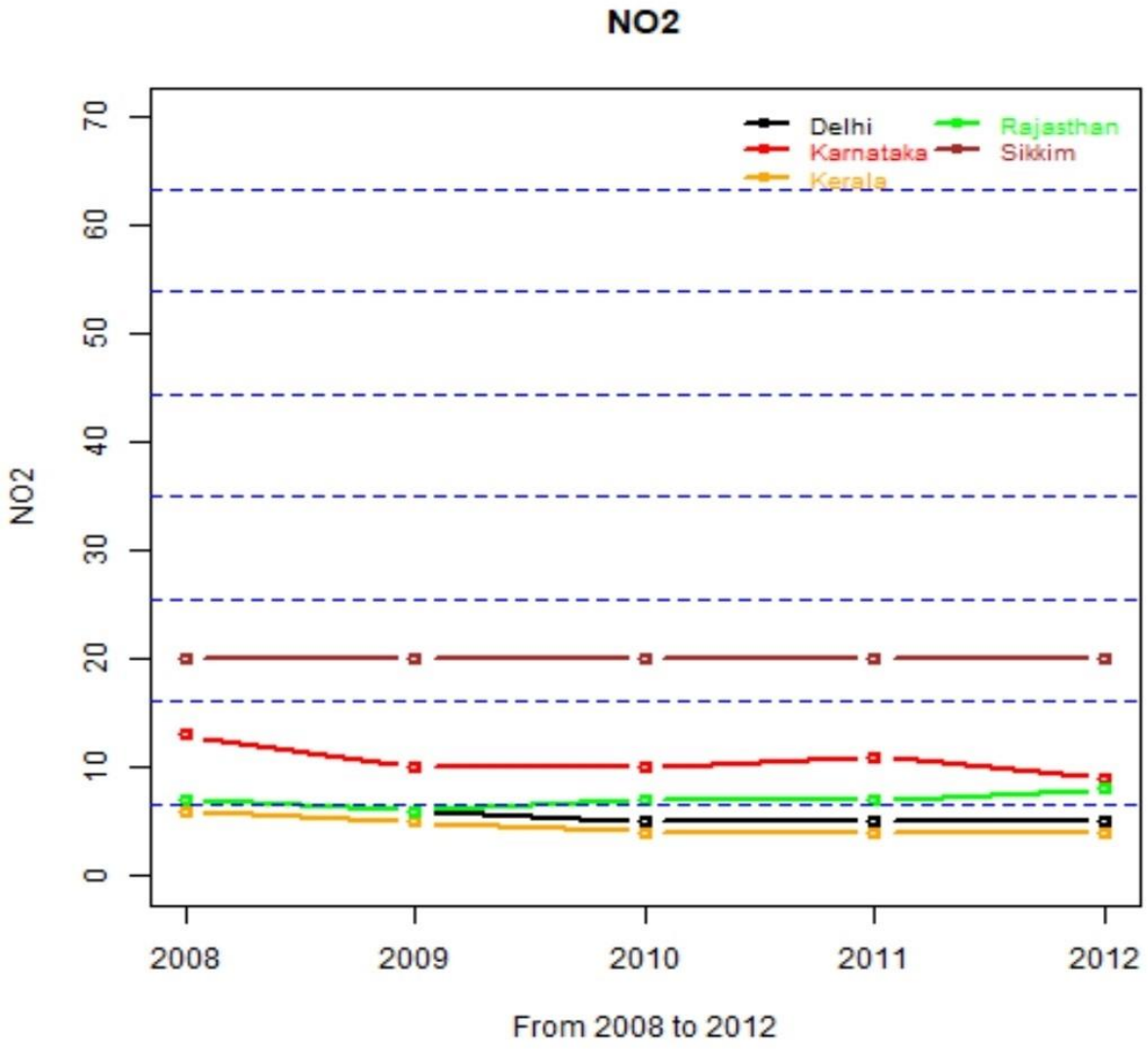

```
#subset of Forest cover

Forestcover<-x[1:25,c(1,2,9)]

#reshape the Forest cover dataset
nn<-reshape(Forestcover,timevar = "State",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(Forestcover$State))
nn[is.na(nn)]<-0

#plot for Forest cover--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,40000),col="black",xlab="From 2008 to 2012",ylab="Forestcover",main="Forest cover")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
 ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

FIGURE 17



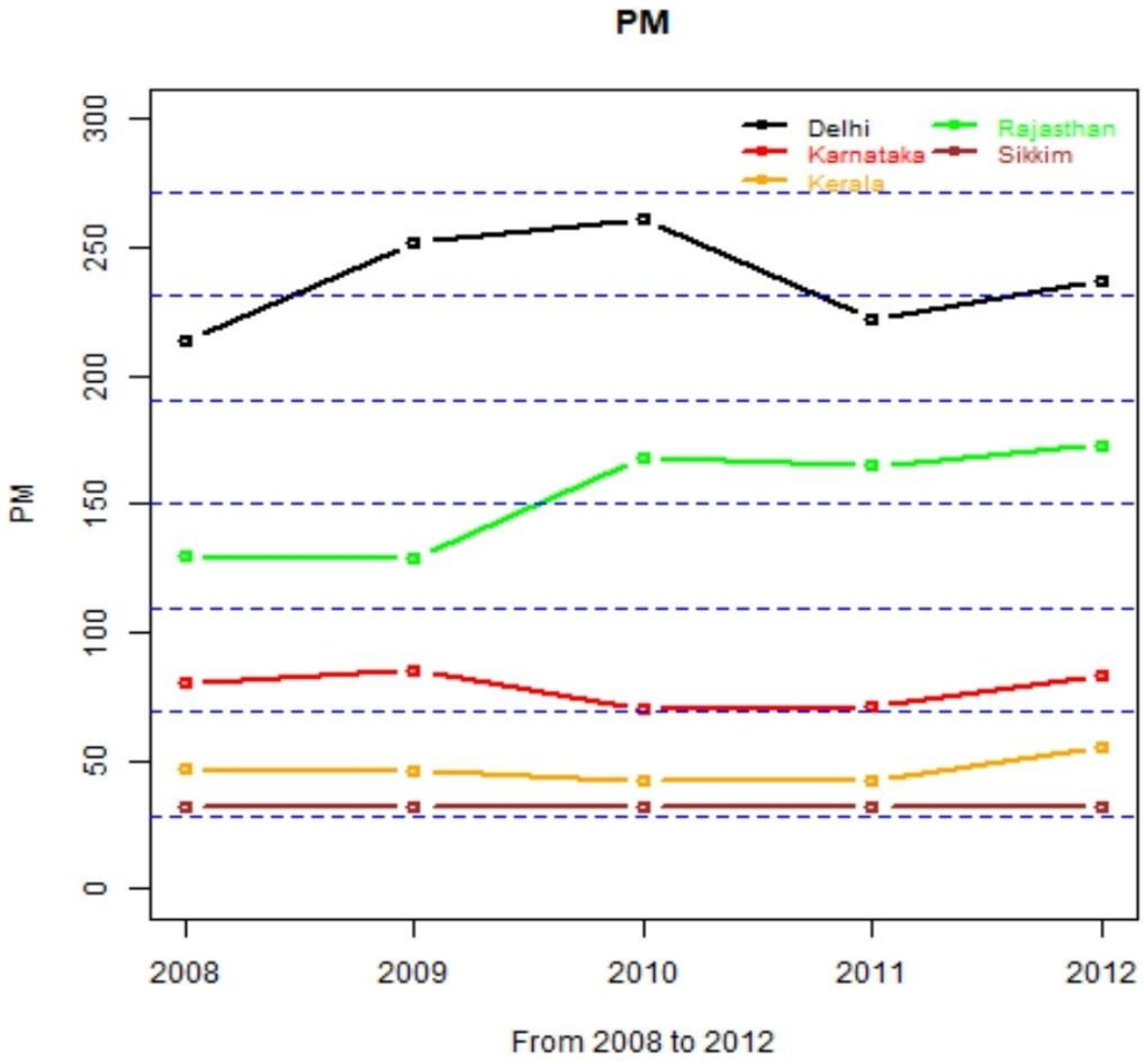

```
#subset of SO2

SO2<-x[1:25,c(1,2,10)]

#reshape the SO2 dataset
nn<-reshape(SO2,timevar = "State",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(SO2 $State))
nn[is.na(nn)]<-0

#plot for SO2--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,15),col="black",xlab="From 2008 to 2012",ylab="SO2",main="SO2")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
  ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

FIGURE 18




```
NO2<-x[1:25,c(1,2,11)]

#reshape the NO2 dataset
nn<-reshape(NO2,timevar = "State",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(NO2$State))
nn[is.na(nn)]<-0

#plot for NO2--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,70),col="black",xlab="From 2008 to 2012",ylab="NO2",main="NO2")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,
 col=c("black","red","orange","green","brown"),ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

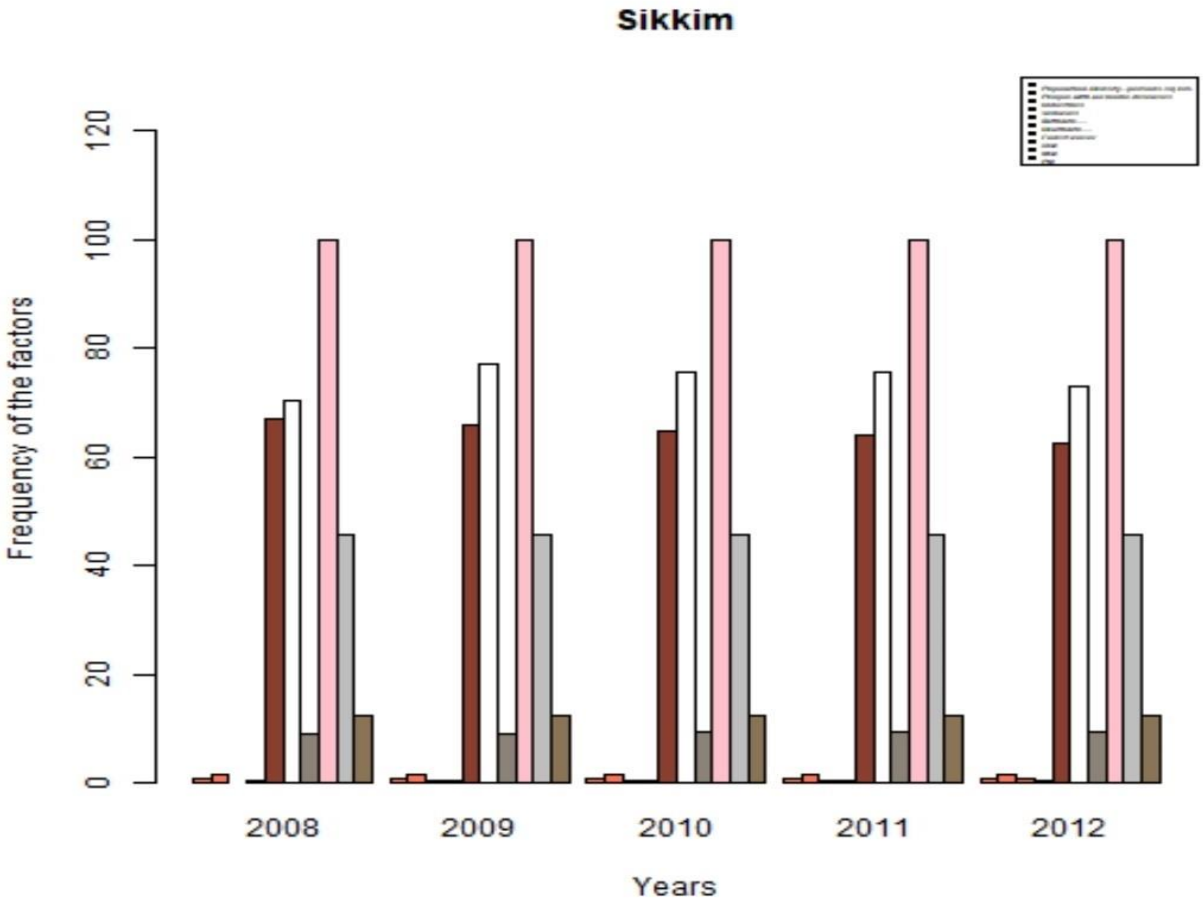
FIGURE 19




```
#subset of PM

PM<-x[1:25,c(1,2,12)]

#reshape the PM dataset
nn<-reshape(PM,timevar = "State",idvar="Year",direction = "wide")
names(nn)[-1]<-as.character(unique(PM$State))
nn[is.na(nn)]<-0

#plot for PM--line graph
plot(nn$Delhi,type="b",lwd=2,xaxt="n",ylim = c(0,300),col="black",xlab="From 2008 to 2012",ylab="PM",main="PM")
axis(1,at=1:length(nn$Year),labels=nn$Year)
lines(nn$Karnataka,col="red",type="b",lwd=2)
lines(nn$Kerala,col="orange",type="b",lwd=2)
lines(nn$Rajasthan,col="green",type="b",lwd=2)
lines(nn$Sikkim,col="brown",type="b",lwd=2)
legend("topright",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","brown"),
 ncol=2,bty="n",cex=0.8,text.col=c("black","red","orange","green","brown"),inset=0.01)
grid(nx=NA,ny=8,lwd=1,lty = 2,col="blue")
```

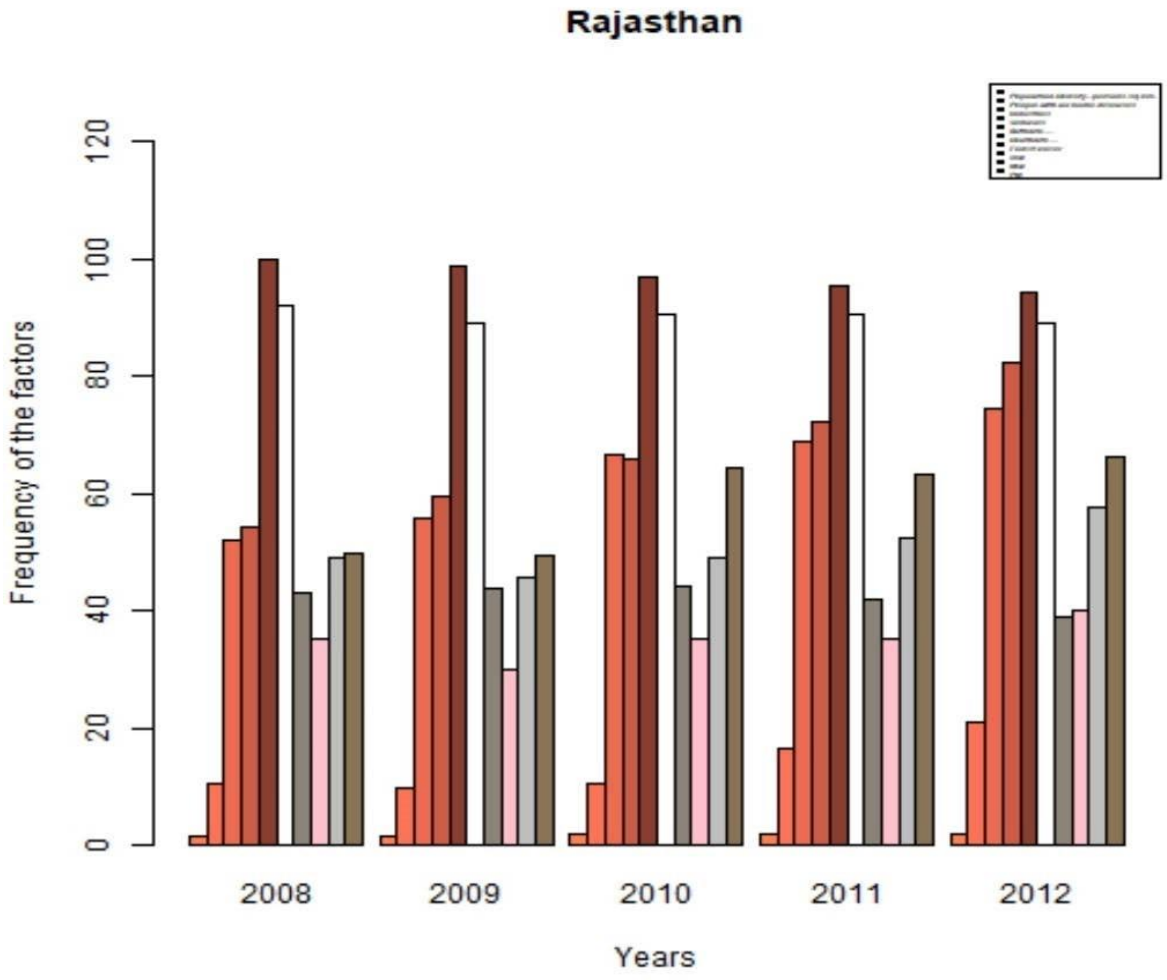
The following grouped bar graphs show the variation in frequency of each factor during the period of 2008-2012 for a particular state.

FIGURE 20



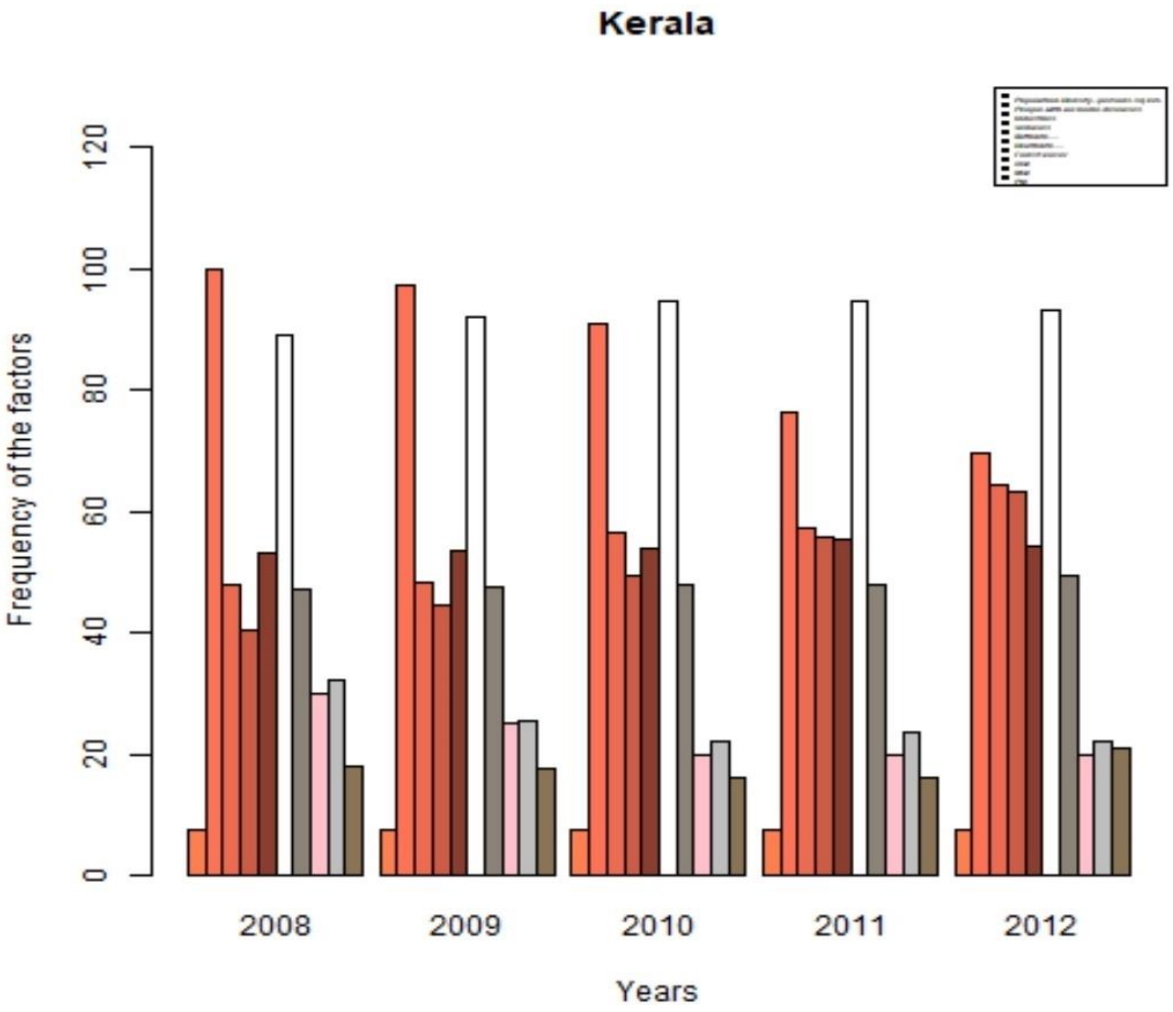

```
#subset the state sikkim from the dataset
percent_sikkim<- subset(percentdataset,percentdataset$State=="Sikkim")

#make an array of colors
color<-c("coral","coral1","coral2","coral3","coral4","white","antiquewhite4","pink","grey","burlywood4")

#choose the required columns(only the factors)
v<-percent_sikkim[,c(3:12)]

#convert the dataset into matrix to plot the chart
v<-as.matrix(v)

#assign the rownames to matrix to-- display it on the chart
rownames(v)<-c("2008","2009","2010","2011","2012")

#Grouped bargraph
barplot(t(v),beside=TRUE,col=color,ylim=c(0,130),main="Sikkim", ylab = "Frequency of the factors",xlab="Years")
legend("topright",rownames(t(v)), fill=color, text.font = 4, cex=0.25 )
```

FIGURE 21



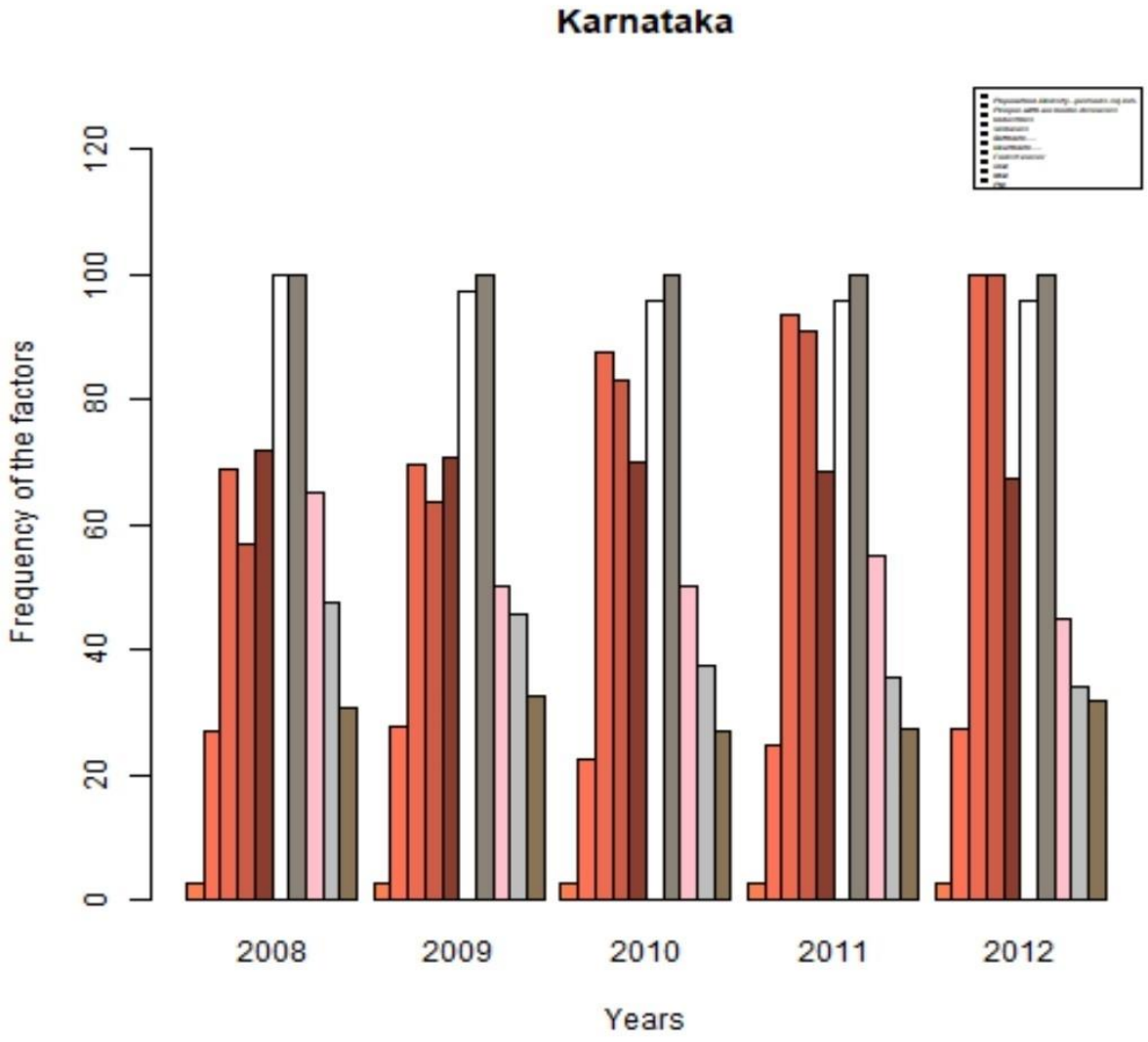

```
#subset the state rajastan from the dataset
percent_rajasthan<- subset(percentdataset,percentdataset$State=="Rajasthan")

#make an array of colors
color<-c("coral","coral1","coral2","coral3","coral4","white","antiquewhite4","pink","grey","burlywood4")

#choose the required columns(only the factors)
w<-percent_rajasthan[,c(3:12)]

#convert the dataset into matrix to plot the chart
w<-as.matrix(w)

#assign the rownames to matrix to-- display it on the chart
rownames(w)<-c("2008","2009","2010","2011","2012")

#Grouped bargraph
barplot(t(w),beside=TRUE,col=color,ylim=c(0,130),main="Rajasthan", ylab = "Frequency of the factors",xlab="Years")
legend("topright",rownames(t(w)), fill=color, text.font = 4, cex=0.25 )
```

FIGURE 22



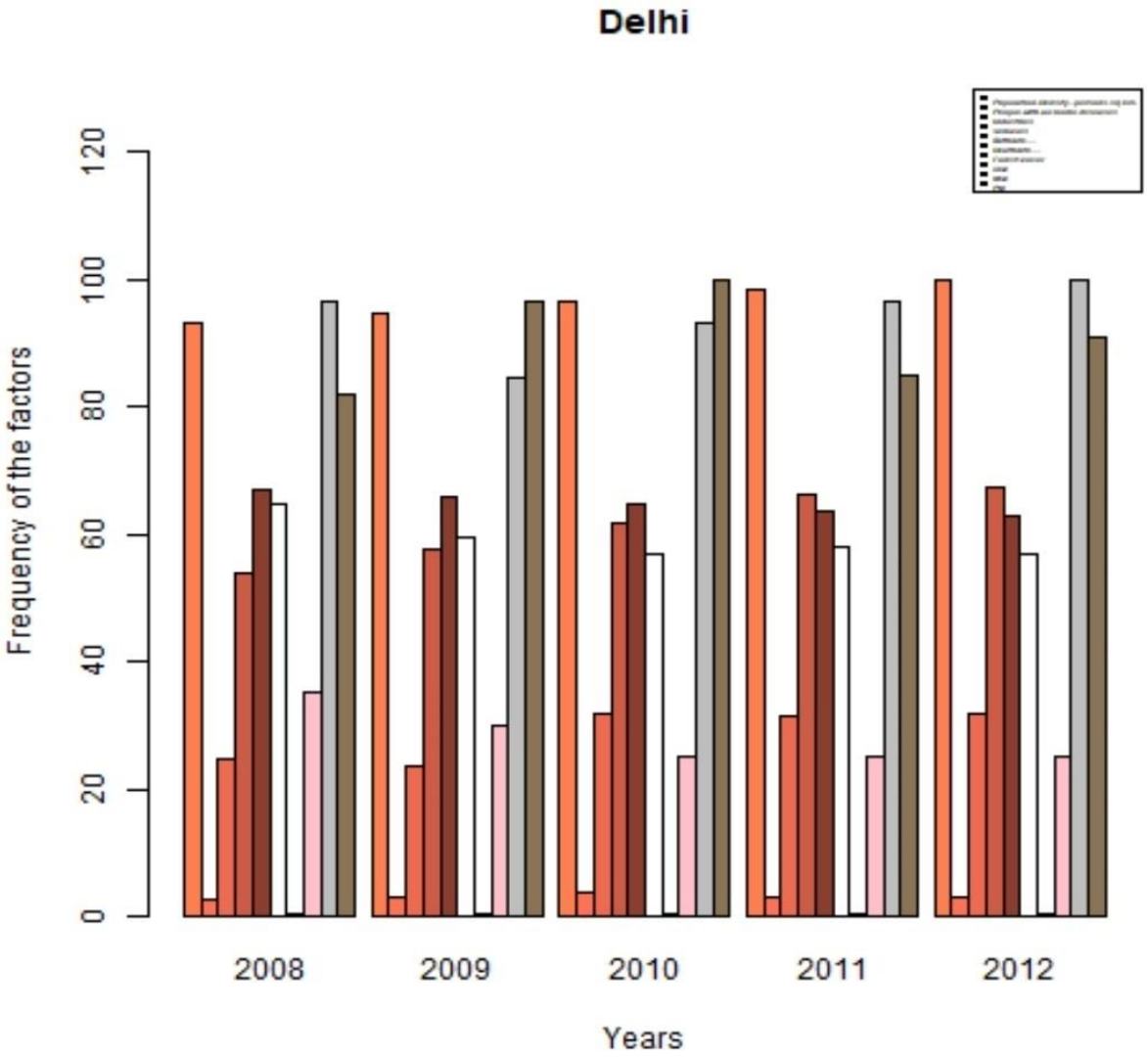

```
#subset the state kerala from the dataset
percent_kerala<- subset(percentdataset,percentdataset$State=="Kerala")

#make an array of colors
color<-c("coral","coral1","coral2","coral3","coral4","white","antiquewhite4","pink","grey","burlywood4")

#choose the required columns(only the factors)
u<-percent_kerala[,c(3:12)]

#convert the dataset into matrix to plot the chart
u<-as.matrix(u)

#assign the rownames to matrix to-- display it on the chart
rownames(u)<-c("2008","2009","2010","2011","2012")

#Grouped bargraph
barplot(t(u),beside=TRUE,col=color,ylim=c(0,130),main="Kerala", ylab = "Frequency of the factors",xlab="Years")
legend("topright",rownames(t(u)), fill=color, text.font = 4, cex=0.25 )
```

FIGURE 23




```
#subset the state karnataka from the dataset
percent_karnataka<- subset(percentdataset,percentdataset$State=="Karnataka")

#make an array of colors
color<-c("coral","coral1","coral2","coral3","coral4","white","antiquewhite4","pink","grey","burlywood4")

#choose the required columns(only the factors)
y<-percent_karnataka[,c(3:12)]

#convert the dataset into matrix to plot the chart
y<-as.matrix(y)

#assign the rownames to matrix to-- display it on the chart
rownames(y)<-c("2008","2009","2010","2011","2012")

#Grouped bargraph
barplot(t(y),beside=TRUE,col=color,ylim=c(0,130),main="Karnataka", ylab = "Frequency of the factors",xlab="Years")
legend("topright",rownames(t(y)), fill=color, text.font = 4, cex=0.25 )
```

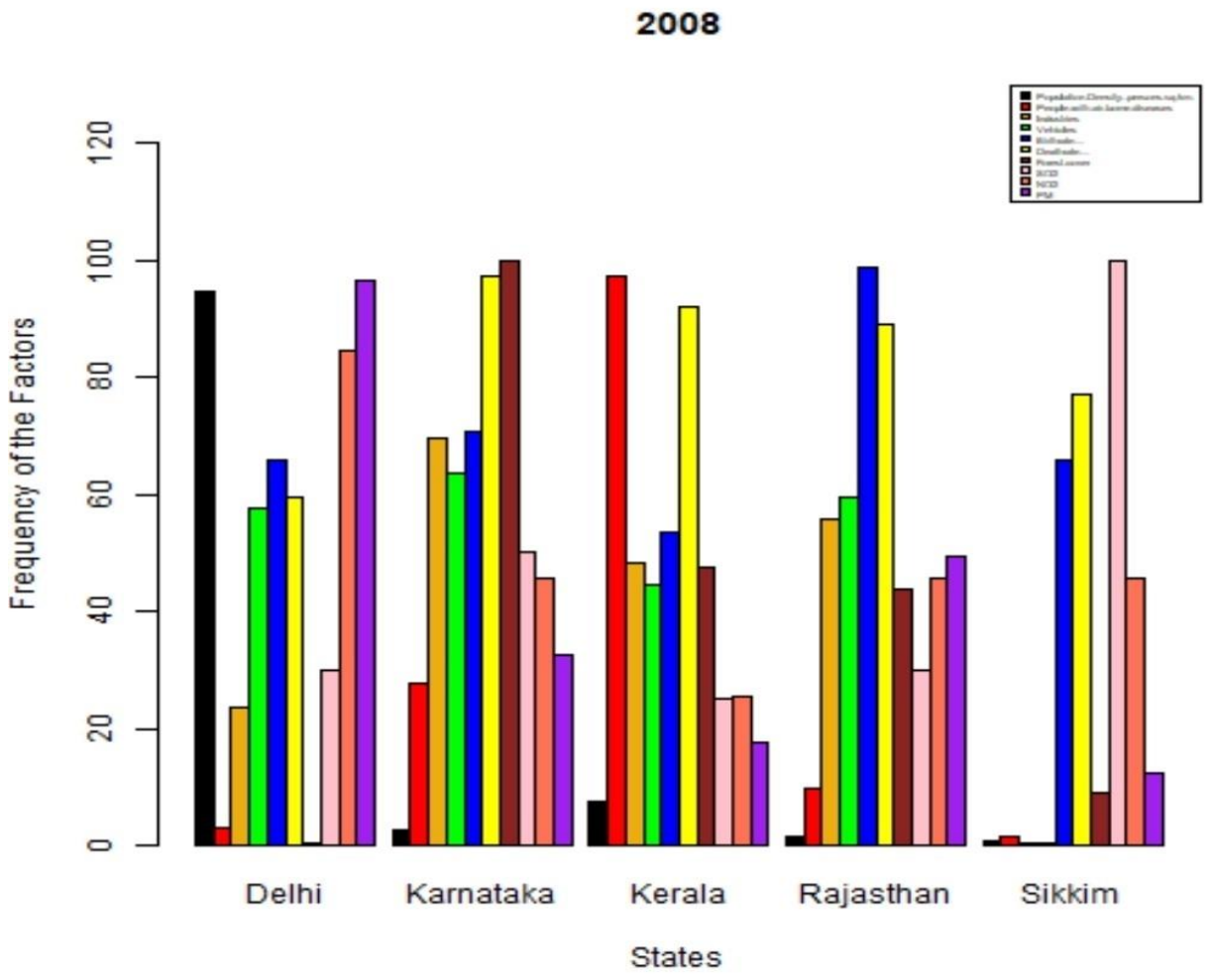
FIGURE 24




```
#subset the state delhi from the dataset
percent_delhi<- subset(percentdataset,percentdataset$State=="Delhi")

#make an array of colors
color<-c("coral","coral1","coral2","coral3","coral4","white","antiquewhite4","pink","grey","burlywood4")

#choose the required columns(only the factors)
z<-percent_delhi[,c(3:12)]

#convert the dataset into matrix to plot the chart
z<-as.matrix(z)

#assign the rownames to matrix to-- display it on the chart
rownames(z)<-c("2008","2009","2010","2011","2012")

#Grouped bargraph
barplot(t(z),beside=TRUE,col=color,ylim=c(0,130),main="Delhi", ylab = "Frequency of the factors",xlab="Years")
legend("topright",rownames(t(z)), fill=color, text.font = 4, cex=0.25 )
```

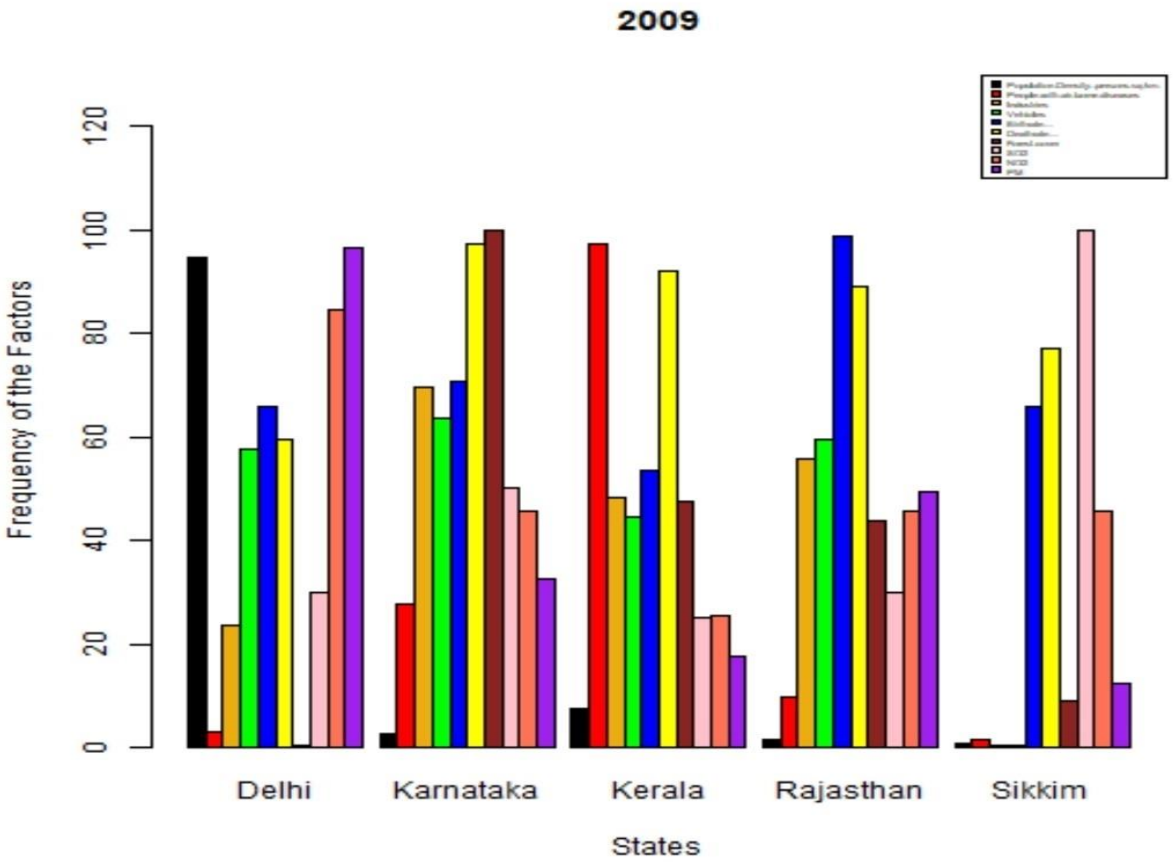
The following grouped bar graphs show the variation in frequency of each factor for all five states during a particular year.

FIGURE 25



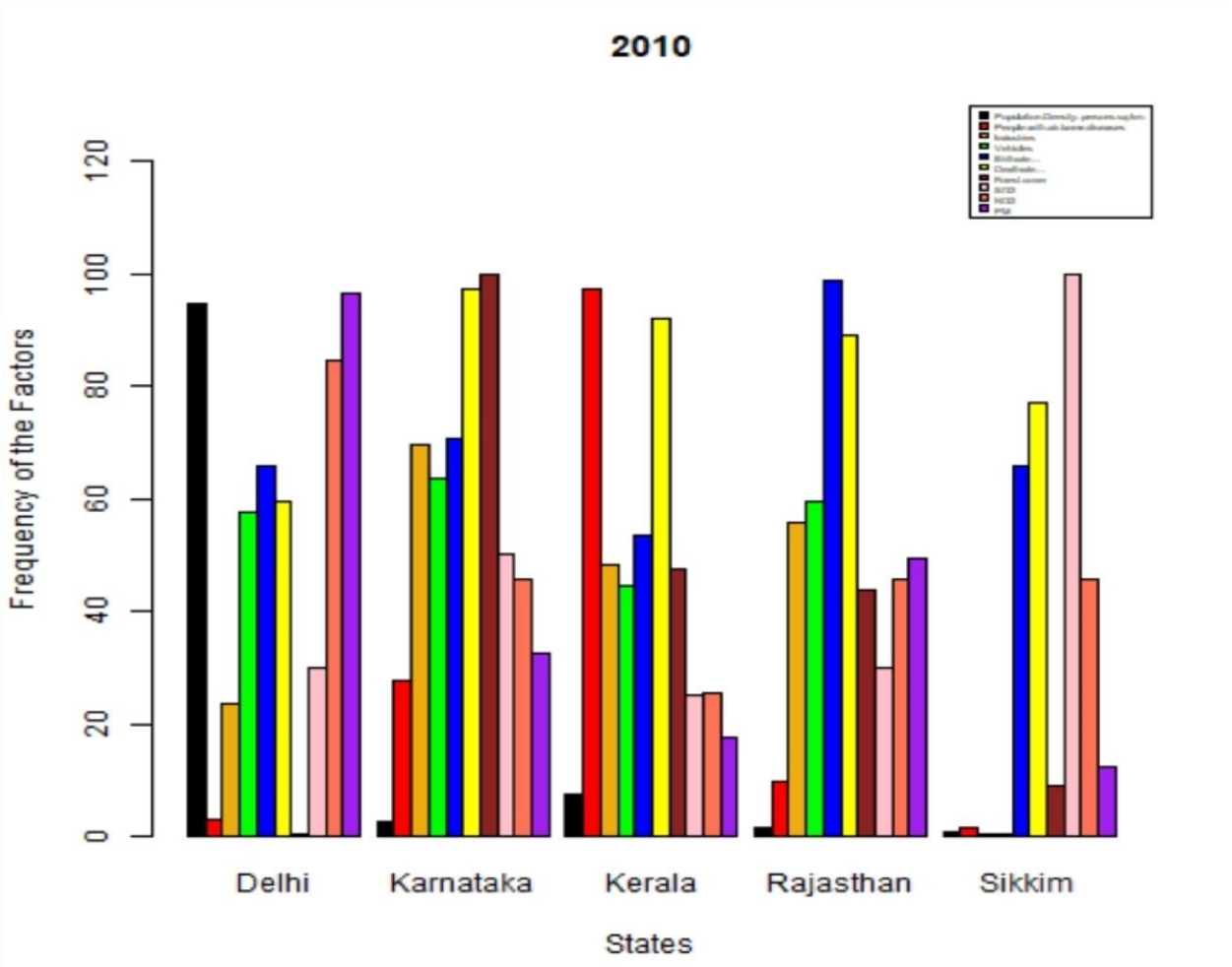

```
#subset the year 2008 from the dataset
percent_eight<- subset(percentdataset,percentdataset$Year =="2008")

#make an array of colors
color<-c("black","red","darkgoldenrod2","green","blue","yellow","brown4","pink","coral1","purple")

#choose the required columns(only the factors)
z<-percent_eight[,c(3:12)]

#convert the dataset into matrix to plot the chart
z<-as.matrix(z)

#assign the rownames to matrix to-- display it on the chart
rownames(z)<-c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim")
png("View_by_year_2008.png")
#Grouped bargraph
barplot(t(z),beside=TRUE,col=color,ylim=c(0,130),main="2008", ylab = "Frequency of the Factors",xlab="States")
legend("topright",rownames(t(z)), fill=color,text.font = 34,cex=0.34)
dev.off()
```

FIGURE 26



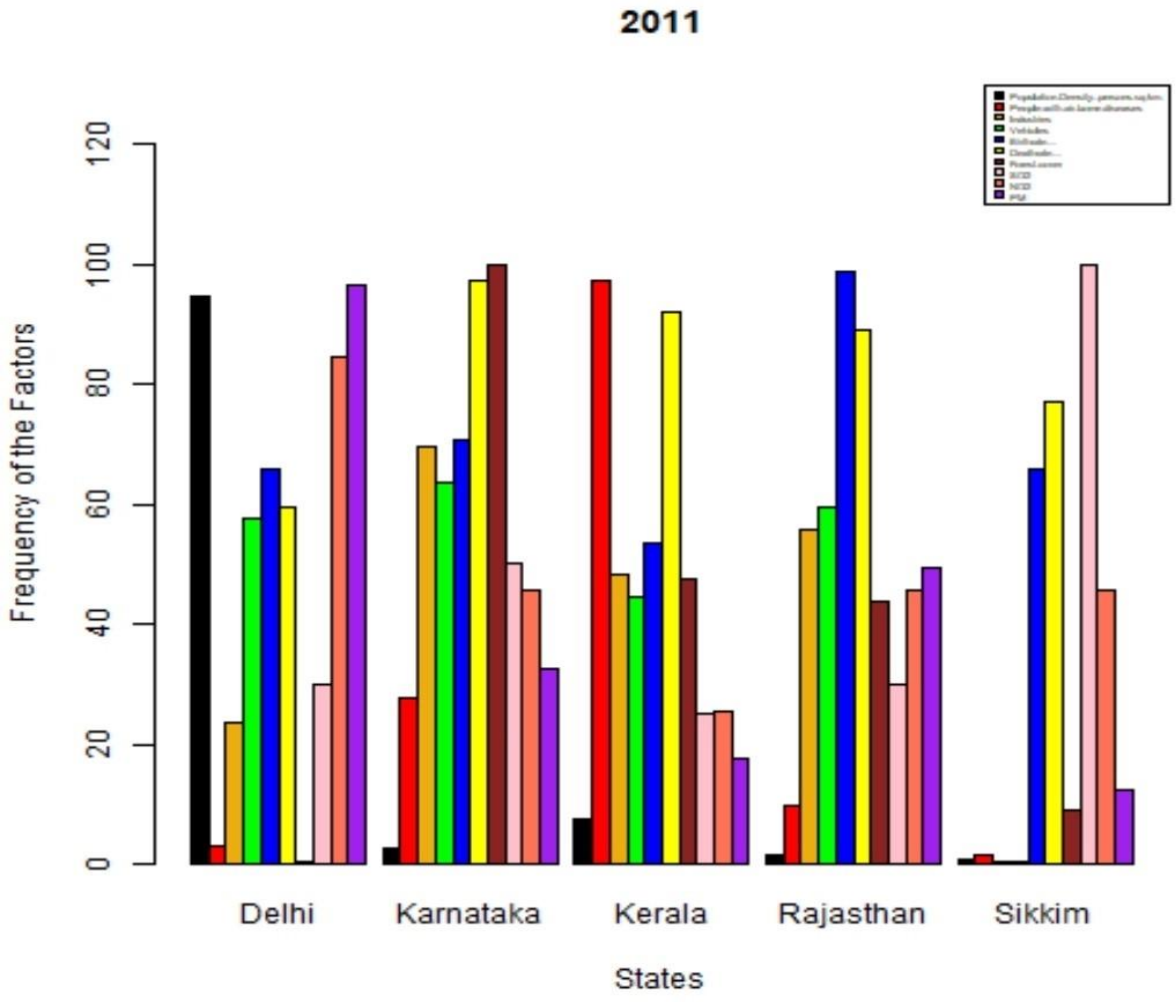

```
percent_nine<- subset(percentdataset,percentdataset$Year =="2009")

#make an array of colors
color<-c("black","red","darkgoldenrod2","green","blue","yellow","brown4","pink","coral1","purple")

#choose the required columns(only the factors)
z<-percent_nine[,c(3:12)]

#convert the dataset into matrix to plot the chart
z<-as.matrix(z)

#assign the rownames to matrix to-- display it on the chart
rownames(z)<-c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim")
png("view_by_year_2009.png")
#Grouped bargraph
barplot(t(z),beside=TRUE,col=color,ylim=c(0,130),main="2009", ylab = "Frequency of the Factors",xlab="States")
legend("topright",rownames(t(z)), fill=color,text.font = 34,cex=0.34)
dev.off()
```

FIGURE 27



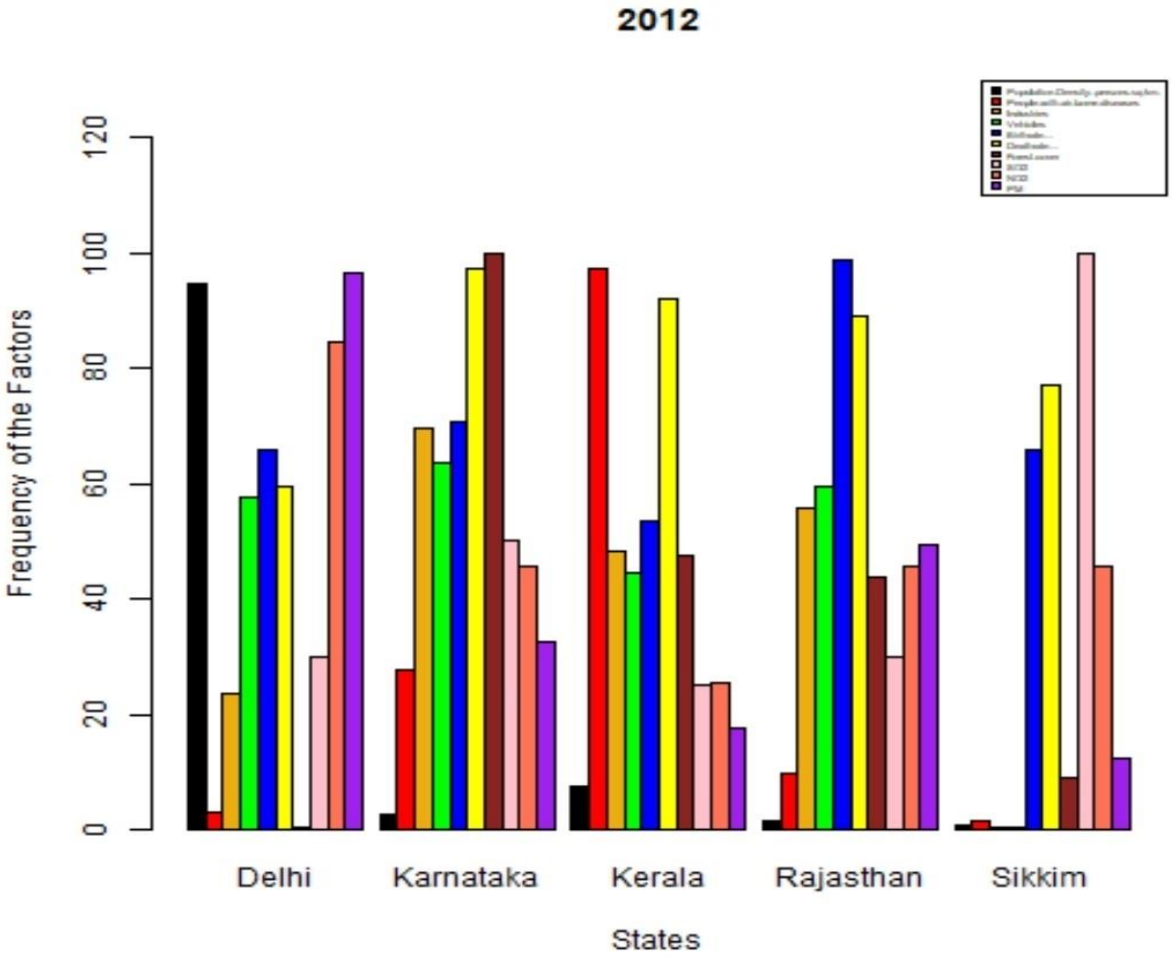

```
percent_ten<- subset(percentdataset,percentdataset$Year =="2010")

#make an array of colors
color<-c("black","red","darkgoldenrod2","green","blue","yellow","brown4","pink","coral1","purple")

#choose the required columns(only the factors)
z<-percent_ten[,c(3:12)]

#convert the dataset into matrix to plot the chart
z<-as.matrix(z)

#assign the rownames to matrix to-- display it on the chart
rownames(z)<-c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim")
png("2010.png")
#Grouped bargraph
barplot(t(z),beside=TRUE,col=color,ylim=c(0,130),main="2010", ylab = "Frequency of the Factors",xlab="States")
legend("topright",rownames(t(z)), fill=color,text.font = 34,cex=0.34)

dev.off()
```

FIGURE 28




```
percent_eleven<- subset(percentdataset,percentdataset$Year =="2011")

#make an array of colors
color<-c("black","red","darkgoldenrod2","green","blue","yellow","brown4","pink","coral1","purple")

#choose the required columns(only the factors)
z<-percent_eleven[,c(3:12)]

#convert the dataset into matrix to plot the chart
z<-as.matrix(z)

#assign the rownames to matrix to-- display it on the chart
rownames(z)<-c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim")
png("2011.png")
#Grouped bargraph
barplot(t(z),beside=TRUE,col=color,ylim=c(0,130),main="2011", ylab = "Frequency of the Factors",xlab="States")
legend("topright",rownames(t(z)), fill=color,text.font = 34,cex=0.34)

dev.off()
```

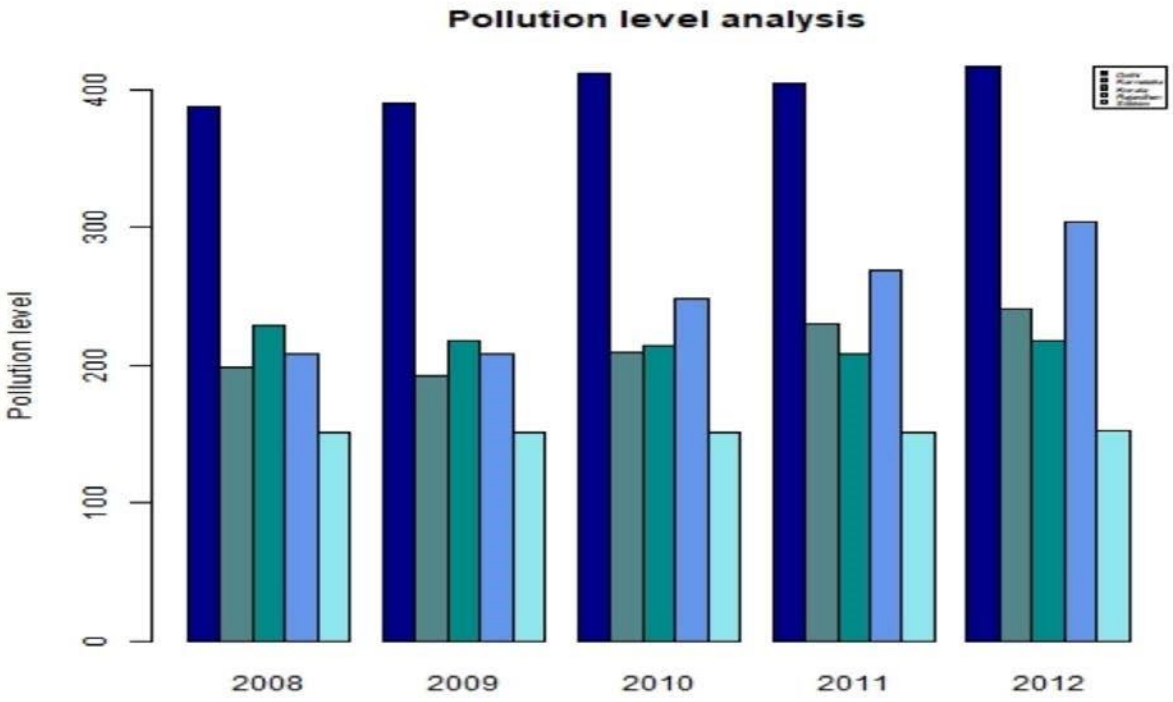
FIGURE 29




```
percent_twelve<- subset(percentdataset,percentdataset$Year =="2012")

#make an array of colors
color<-c("black","red","darkgoldenrod2","green","blue","yellow","brown4","pink","coral1","purple")

#choose the required columns(only the factors)
z<-percent_twelve[,c(3:12)]

#convert the dataset into matrix to plot the chart
z<-as.matrix(z)

#assign the rownames to matrix to-- display it on the chart
rownames(z)<-c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim")
png("2012.png")
#Grouped bargraph
barplot(t(z),beside=TRUE,col=color,ylim=c(0,130),main="2012", ylab = "Frequency of the Factors",xlab="States")
legend("topright",rownames(t(z)), fill=color,text.font = 34,cex=0.34)
dev.off()
```

The below grouped bar graph shows the variation of pollution level in all five states during the period of 2008-2012.

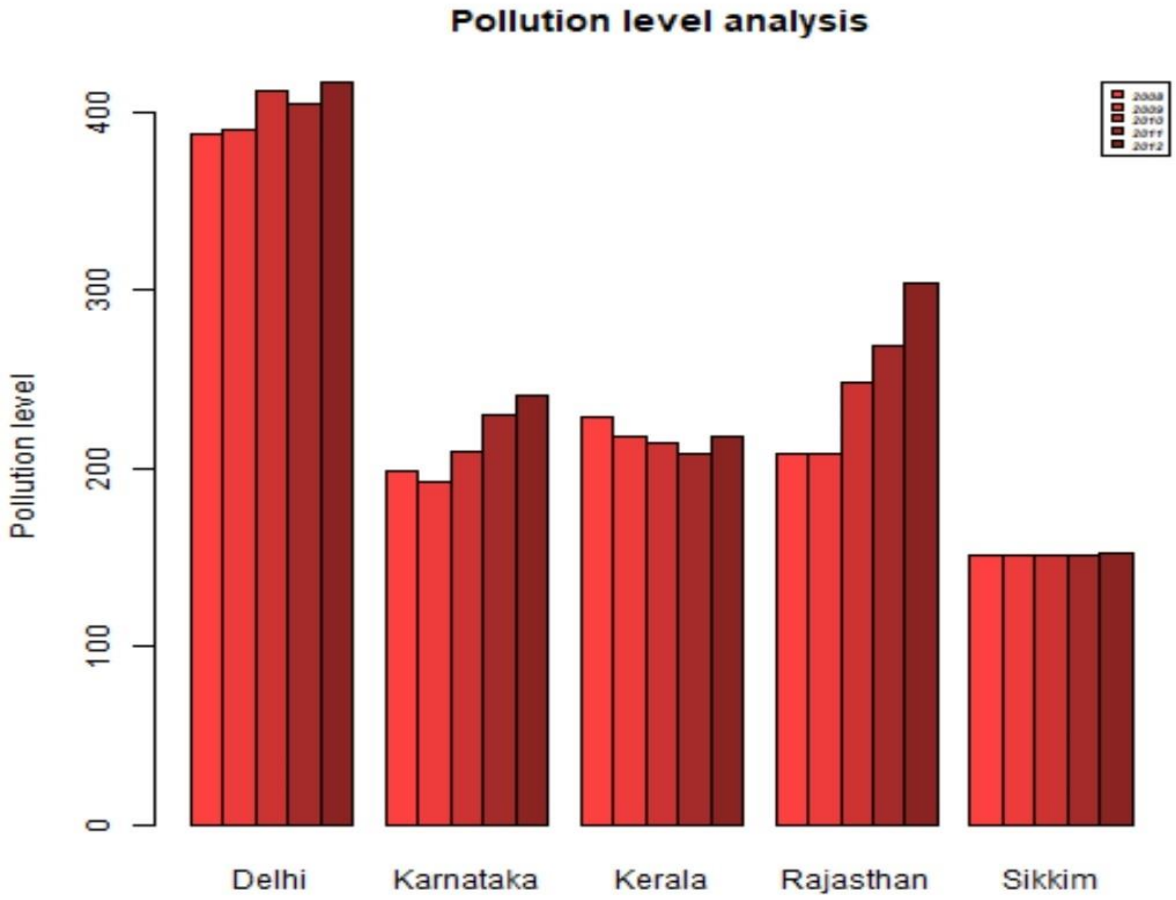
FIGURE 30




```
pp<-reshape(pollution_dataset,timevar = "Year",idvar="State",direction = "wide")
names(pp)[2:6]<-as.character(unique(pollution_dataset$Year))
pp[is.na(pp)]<-0
pp<-pp[,c(2:6)]
pp<-as.matrix(pp)
color1<-c("darkblue","cadetblue4","cyan4","cornflowerblue","cadetblue2")
colnames(pp)<-c("2008","2009","2010","2011","2012")
rownames(pp)<-c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim")
barplot(pp,beside=TRUE,col=color1,main="Pollution level analysis",ylab="Pollution level")
legend("topright",rownames(pp), fill=color1, text.font = 4, cex=0.3 )
```

The below grouped bar graph shows the variation of pollution level during the period of 2008-2012 in all five states.

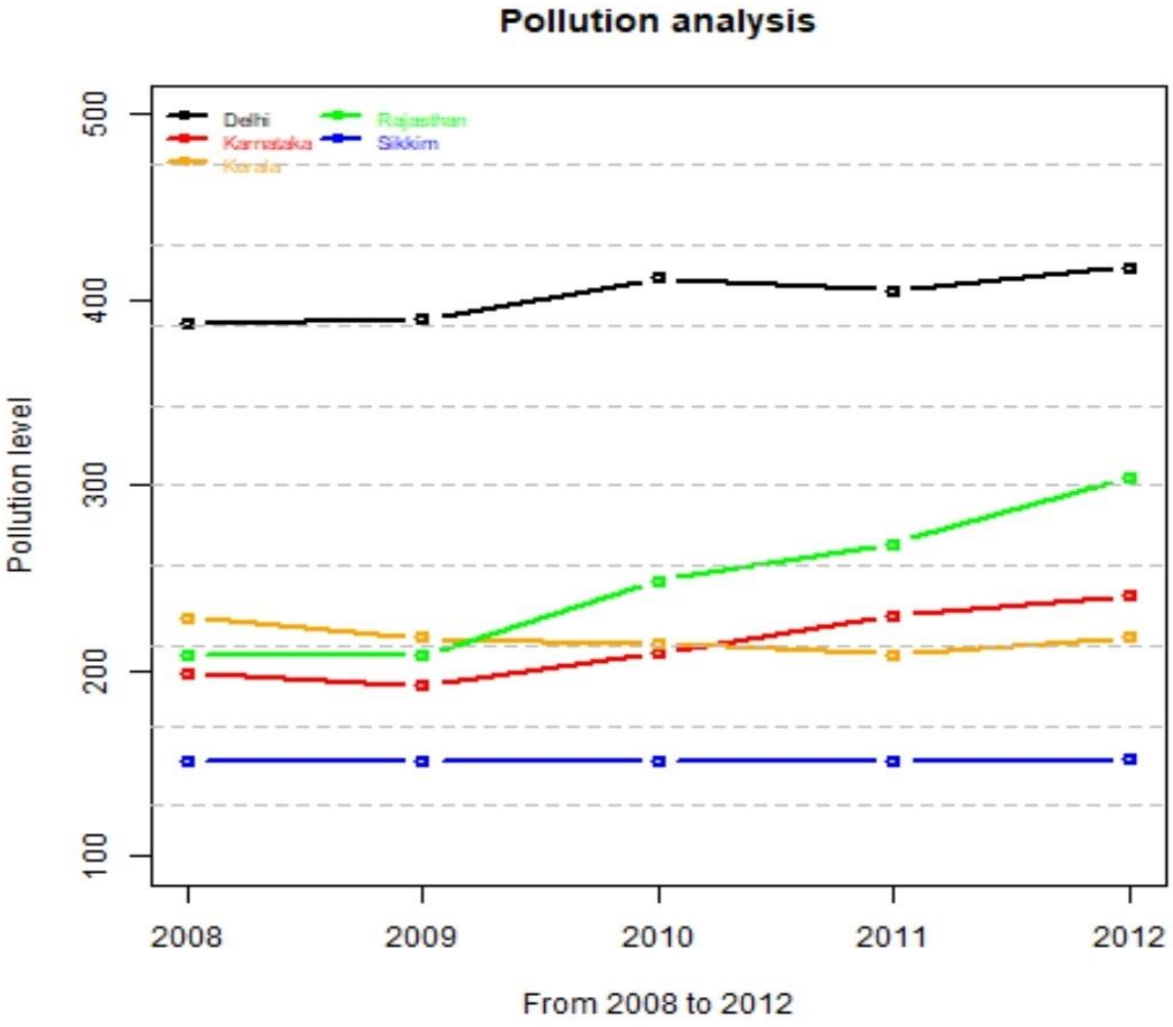
FIGURE 31




```
pp_1<-reshape(pollution_dataset,timevar = "State",idvar="Year",direction = "wide")
names(pp)[2:6]<-as.character(unique(pollution_dataset$State))
pp_1[is.na(pp_1)]<-0
pp_1<-pp_1[,c(2:6)]
pp_1<-as.matrix(pp_1)
color2<-c("brown1","brown2","brown3","brown","brown4")
rownames(pp_1)<-c("2008","2009","2010","2011","2012")
colnames(pp_1)<-c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim")
barplot(pp_1,beside=TRUE,col=color2,main="Pollution level analysis",ylab="Pollution level")
legend("topright",rownames(pp_1), fill=color2, text.font = 4, cex=0.4 )
```

The below line graph shows the variation of pollution level in all five states during the period 2008-2012.

FIGURE 32




```
pp2<-reshape(pollution_dataset,timevar = "State",idvar="Year",direction = "wide")
names(pp2)[2:6]<-as.character(unique(pollution_dataset$State))


pp2[is.na(pp2)]<-0

plot(pp2$Delhi,type="b",lwd=2,xaxt="n",ylim = c(100,500),col="black",xlab="From 2008 to 2012",ylab="Pollution level",main="Pollution analysis")
axis(1,at=1:length(unique(pollution_dataset$Year)),labels=unique(pollution_dataset$Year))
lines(pp2$Karnataka,col="red",type="b",lwd=2)
lines(pp2$Kerala,col="orange",type="b",lwd=2)
lines(pp2$Rajasthan,col="green",type="b",lwd=2)
lines(pp2$Sikkim,col="blue",type="b",lwd=2)
legend("topleft",legend=c("Delhi","Karnataka","Kerala","Rajasthan","Sikkim"),lty=1,lwd=2,pch=21,col=c("black","red","orange","green","blue"),
 ncol=2,bty="n",cex=0.7,text.col=c("black","red","orange","green","blue"),inset=0.01)

grid(nx=NA,ny=10,lwd=1,lty = 2,col="grey")
```

CONCLUSION

From the above analysis, we conclude that:

- The majorly affected state in India during the years 2008-2012 by air pollution is mostly in the northern region i.e. Delhi, which is heavily polluted and requires immediate action.
- Karnataka had initially lower levels of pollutants in the atmosphere (2008-2009) compared to the other states (Delhi, Karnataka and Kerala); however saw a steady increase of the same in the later years (2010-2012). This is largely due to the increase in population density per square km, industries, vehicles, and the concentration of pollutants in the air.
- Kerala was moderately polluted in the early stages (2008) but, later, were taken care of (2009- 2011). The reason for the decrease could be awareness in citizens and government policies.
- Rajasthan saw a massive hike in the air pollution level during 2008-2012 which is directly correlated to the increase in population density per square km, industries, vehicles and particulate matter in the atmosphere.
- The pollution level in Sikkim has been constant throughout the years (2008-2011), however the levels increased by a very small factor in 2012 which can be attributed to the steady increase in population density, industry, vehicles and other determinants.

From the above data analysis approach, we conclude that data analysis is a crucial aspect for a better future. It is interesting to see how data analysis and the day to day instances are coherent and how data analysis can be used to deal with significant problems.

We must find a cure to this significant problem as it is killing our nation slowly.

REFERENCES

- https://towardsdatascience.com/india-air-pollution-data-analysis-bd7dbfe93841
- https://www.researchgate.net/publication/320707293_Air_Quality_Prediction_Big_data_and_Machine_Learning_Approaches
- https://www3.epa.gov/airnow/2018conference/Plenary/karin-tuxen-bettman.pdf
- https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/fighting-airpollution.html
- https://www.sciencedirect.com/science/article/pii/S1877050918307555
- https://www.researchtrend.net/ijet/pdf/39-%20178.pdf
- https://www.analyticsvidhya.com/blog/2016/10/complete-study-of-factors-contributing-toair-pollution/