CRIME IN INDIA

Data Analysis over the various crimes in India

By Group 15,

Hanmant Lokare(hl447@rit.edu)

Rajkumar Pillai(rlp4867@rit.edu)

Dipesh Virkar(dv1462@rit.edu)

Table of Contents

INTRODUCTION

A Report published by National Crime Record Bureau (NCRB) set-up in 1956 responsible for collecting and analyzing crime data as defined by Indian Penal Code (IPC). NCRB compared the crime rates of 1953 to 2013 and concluded that there is a significant increase in rate of rapes, dowry, molestation, cruelty by husband over the period. Official sources show that rape cases in India has doubled between 1990 and 2008. Cruelty by husband and his family is a major contributor towards crime against women. Official statistics show that there has been dramatic increase in the number of reported crimes against women.

Rape in India has been described by Radha Kumar as one of the India's most common crimes against women. Official sources show that rape cases in India has been doubled between 1990 and 2008. In most of the rape cases, the culprit is known to the victim.

AIM

Our aim is to help the constitution to make strict laws against the crime which is very frequent and put more efforts on making justice with women's suffering from crime in specific areas.

SUMMARY

In this project we have analyzed the data of crimes in India over the period of 10 years.

We have followed steps as follows:

Data Collection - To analyze the data of crimes we found a dataset of different types of crimes in different states in India over the period of 10 years.

Data Cleaning and Preparation - This dataset had some NULL values and redundant values in the dataset which we deleted, and the NULL values were considered as 0.

Data Analysis - For data analysis we visualized distribution of various of crimes against states, years.

KEY FINDINGS

1. Crime rate in India from 2001 to 2013.

   We observed an **increase** in the crime rate from

2.Distribution of Various Crimes in India

   Cruelty by Husband / Family is the most frequent crime in India

   Importation of Girls is the least frequent crime in India

3.State and year wise distribution of specific crimes

   Andhra Pradesh is the state with most number of crimes committed

   As the crime rate is increasing every year, **2013** has the most number of crimes

4.Most safe and unsafe state in India

   Andhra Pradesh is the state with most number of crimes committed

   Lakshadweep is the state with least number of crimes

5.Total crimes committed on women

   About **50%** of the total crimes committed every year account for the crimes committed against women

DATA CLEANING AND PREPARATION

In this section of report, we will discuss about pre-processing of data.

Firstly, the dataset was collected from https://www.kaggle.com/rajanand/crime-in-india.This dataset was published by National Crime Records Bureau(NCRB), Government of India on their website and it has been shared on open Government Data Platform India Portal Under. The whole dataset was in comma separated file (*.csv), we converted it into database by using **SQL server Import and Export Wizard**. This wizard helps us to convert all the required file into tables of our database.

There were few datasets corrupted so we dropped that dataset on the bases of how important that data was. Also, the dataset called Arrest under crime against women contains an contains an anomalous data which dropped from the table.

	Area_Name	Year	Group_Name	Sub_Group_Name	Persons_Acquitted	Persons_against_whom_cases_Compounded_or_Withd...
27...	Kerala	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	106
27...	Lakshadweep	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	0
27...	Madhya Pradesh	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	30
27...	Maharashtra	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	251
27...	Manipur	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	0
27...	Meghalaya	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	1
28...	Mizoram	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	0
28...	Nagaland	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	0
28...	Odisha	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	31
28...	Puducherry	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	9
28...	Punjab	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	45
28...	Rajasthan	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	36
28...	Sikkim	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	0
28...	Tamil Nadu	2002	Immoral Traffic (Preventio...	08. Immoral Traffic Prevention ...	1956	45

Fig.1

As shown in the Fig.1 in column person_acquitted contains same figure and that is not possible. So, we dropped all the rows which has that number.

Similarly, all the dataset contains NULL values in some rows, which we replaced by using transform tab in rattle by mean value. This technique helped us to overcome the missing values in the dataset.

There were many attributes in tables which were not necessary for our analysis, we dropped that attribute again by using rattle transform tab's cleanup options.

CRIME RATE IN INDIA OVER THE YEARS

It's very important to analyze the overall crime rate in India. It gives a broad perspective of overall distribution of crime in past years. Our goal is to find the rate of crime over the year. To get that data we will use 42_District_wise_crimes_committed_against_women_2001_2013 and add all the crimes committed over each year using following query in SSMS.

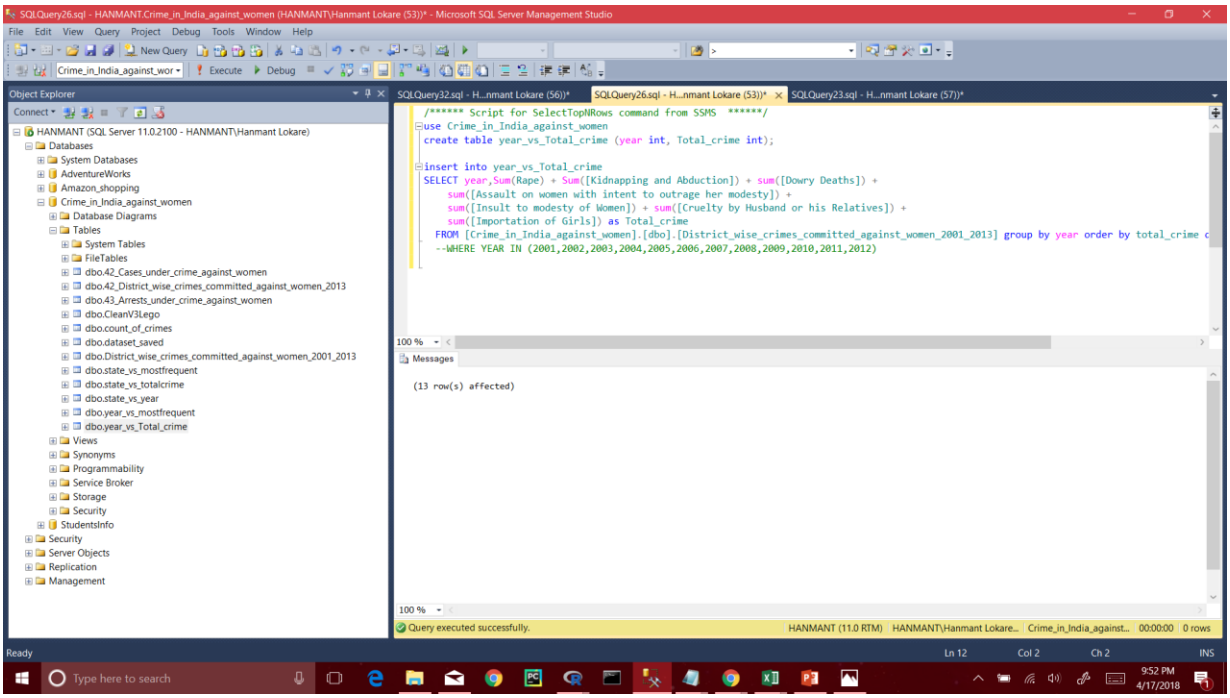

Fig. 2

Next, we must plot a graph of year vs Total crime. For this we will use R. The following code is used to plot the graph in Fig. 3.

```
con<-odbcConnect("Crime_in_India_against_women")

> res <- sqlQuery(con, 'select * from year_vs_Total_Crime')

> scale<-res[,1]

> serial<-res[,2]

> plot(scale,serial,
xlab="YEAR",ylab="TOTAL_CRIMES",type="p",col='red',main="Distribution of Total
Crime vs Year")

> lines(scale[order(scale)], serial[order(scale)])
```

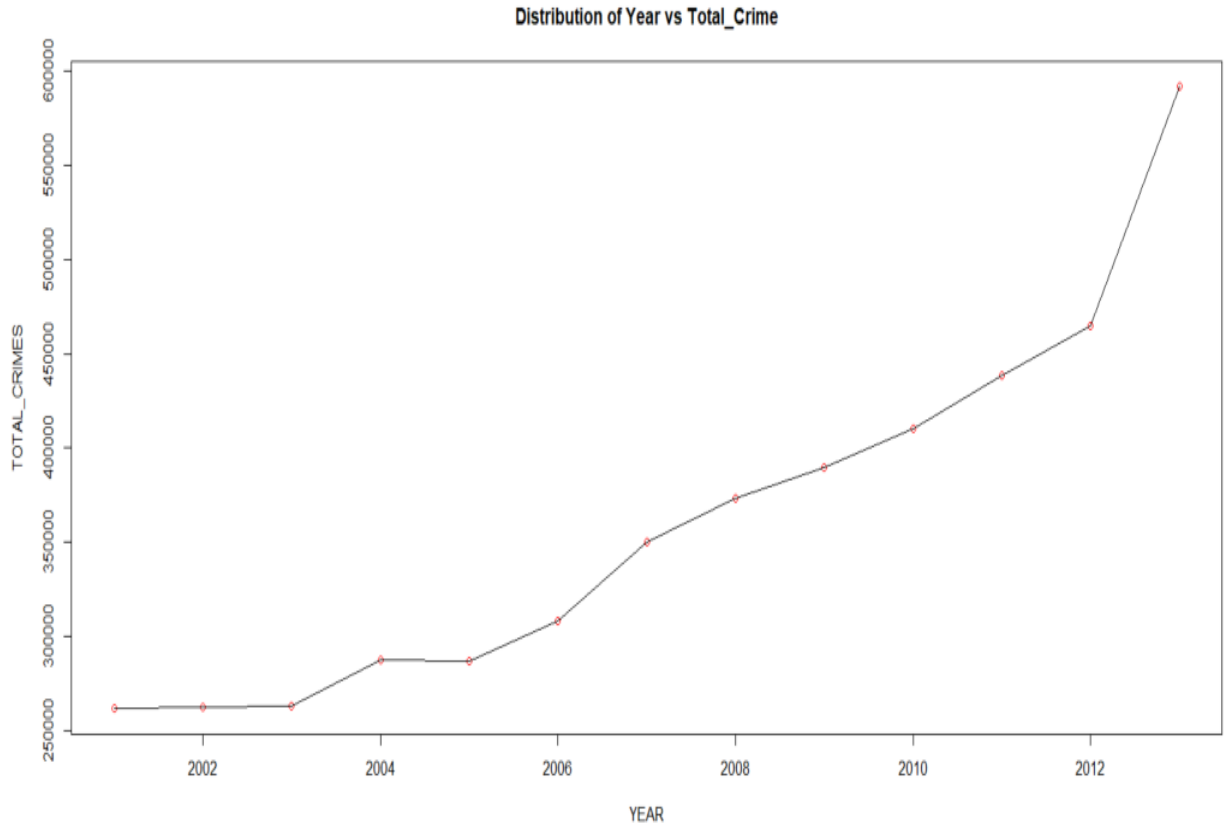

Fig.3

- From the above line graph, we can observe the exponential growth of crimes committed in India from the period of 2001 to 2012.
- We can also observe the sudden growth of crime from approximately from 45,000 and reached around 60,000 2012 to 2013.

DISTRIBUTION OF NUMBER OF CRIMES IN EVERY STATE

After analyzing the data distribution over the years now we will distribution of crime in all the states over the period of 10 years

To obtain a count of all crimes in each state we used sum function in SQL to count the crimes in every state and every year.

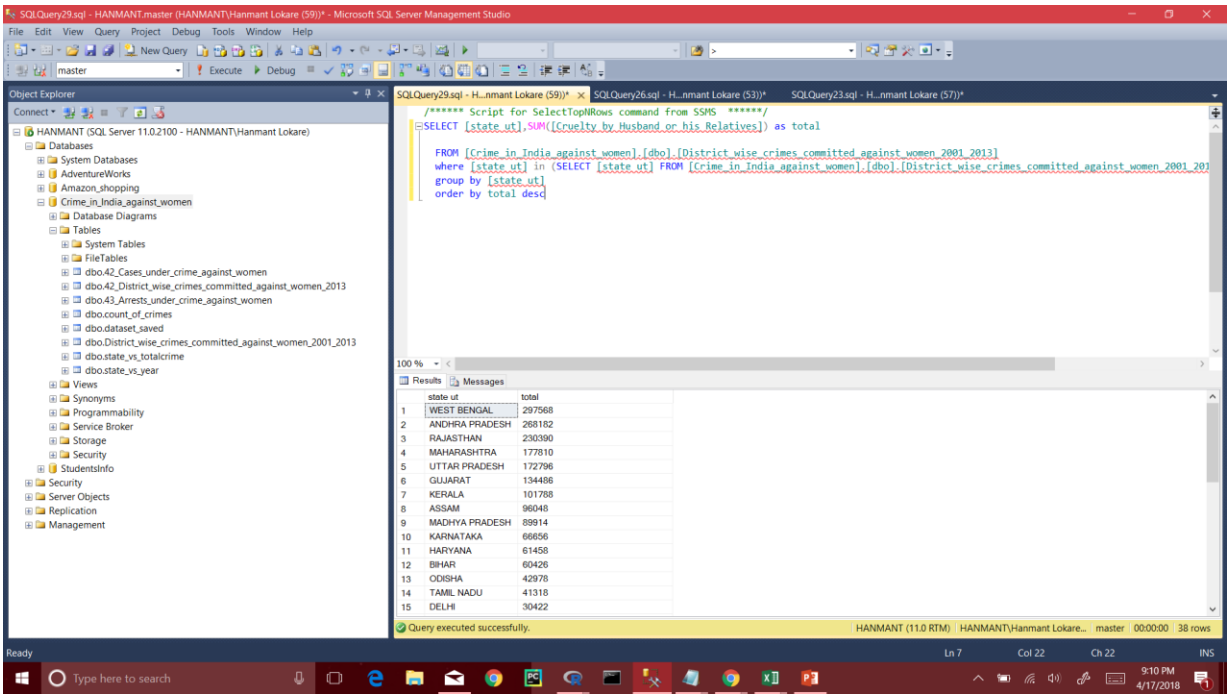

Fig. 4

To visualize the above the data we used a simple bar graph to show state vs count of crime.

The following R code was used to create a bar graph using barplot() function.

R-Code:

```
> con<-odbcConnect("Crime_in_India_against_women")
> res <- sqlQuery(con, 'select * from state_vs_totalcrime')
> scale<-res[,1]
> serial<-res[,2]
> par(mar=c(15, 4, 3, 1))
> barplot(serial,horiz=FALSE,names.arg=scale,las=2,ylim=c(0,250000),main="Bar plot of crimes in every state over the years")
```

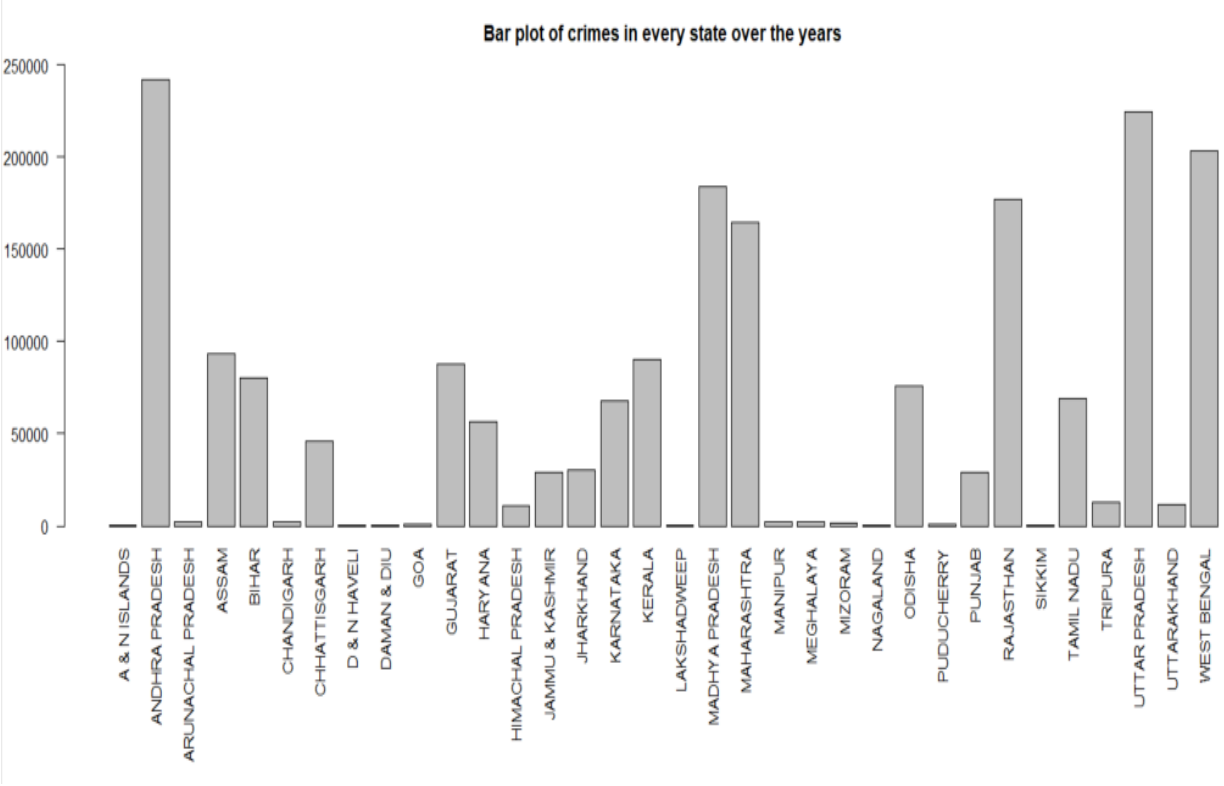

Fig. 5

- The above graph shows us the distribution of number of crime committed in different states over the years.
- Andhra Pradesh and Uttar Pradesh have the most number of Crimes Committed.
- The union territories and small states have a less number of Crimes Committed

DISTRIBUTION OF VARIOUS CRIMES COMMITTED (on Women) OVER THE COUNTRY

Now that we have learned about overall crime in India we will focus on crimes related to women. Firstly, we will study about the distribution of different types of crimes on women in India. So, to obtain the total number of crime against women we first calculated the sum of all types of crimes over the period of 10 years by using SSMS.

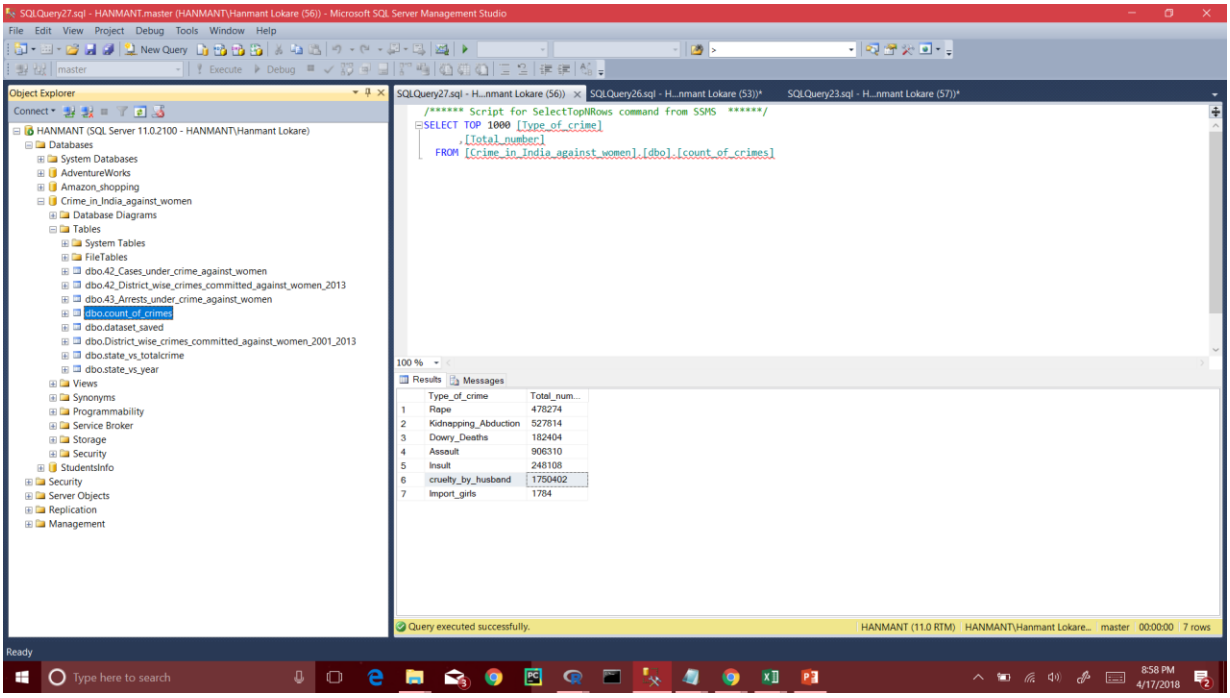

Fig. 6

So to represent this result in a graph we used a Pie Graph by using the R command pie()

The R code for the graph is as following:

```
> con<-odbcConnect("Crime_in_India_against_women")
> res <- sqlQuery(con, 'select * from count_of_crimes')
> scale<-res[,1]
> serial<-res[,2]
> pct<-round(serial/sum(serial)*100)
> scale<-paste(scale,pct)
> scale<-paste(scale,"%",sep="")
> pie(serial, scale, main="Pie Chart of All crimes")
```

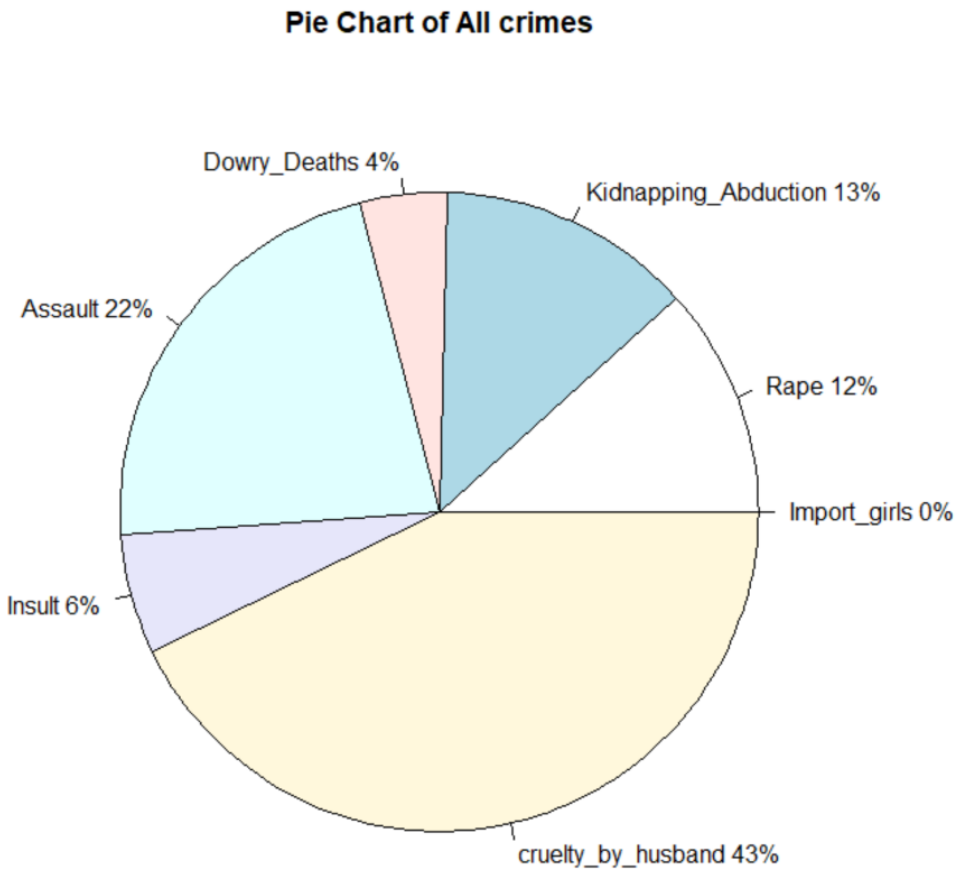

Fig. 7

- As we can see in the above Pie Chart, **Cruelty by Husband** is the most frequent crime committed in India on women over the years 2001-2013 accounts for about **43%** of the total crime committed on women.
- In this Pie chart we can see that **Importation of Girls** for various purposes holds **0%** over the total count as it has a negligible value as compared to the maximum count.
- Other major crimes on women include Rapes, Assaults, Deaths by Dowry and Kidnapping

Most frequent Crime – Cruelty by Husband / Relatives

After analyzing the different types of crimes over the country we found that Cruelty by husband or family is the most frequent crime in India, Lets analyze the crime count over the years.

To do that we can count the number of Cruelty by husband cases over the entire country every year.

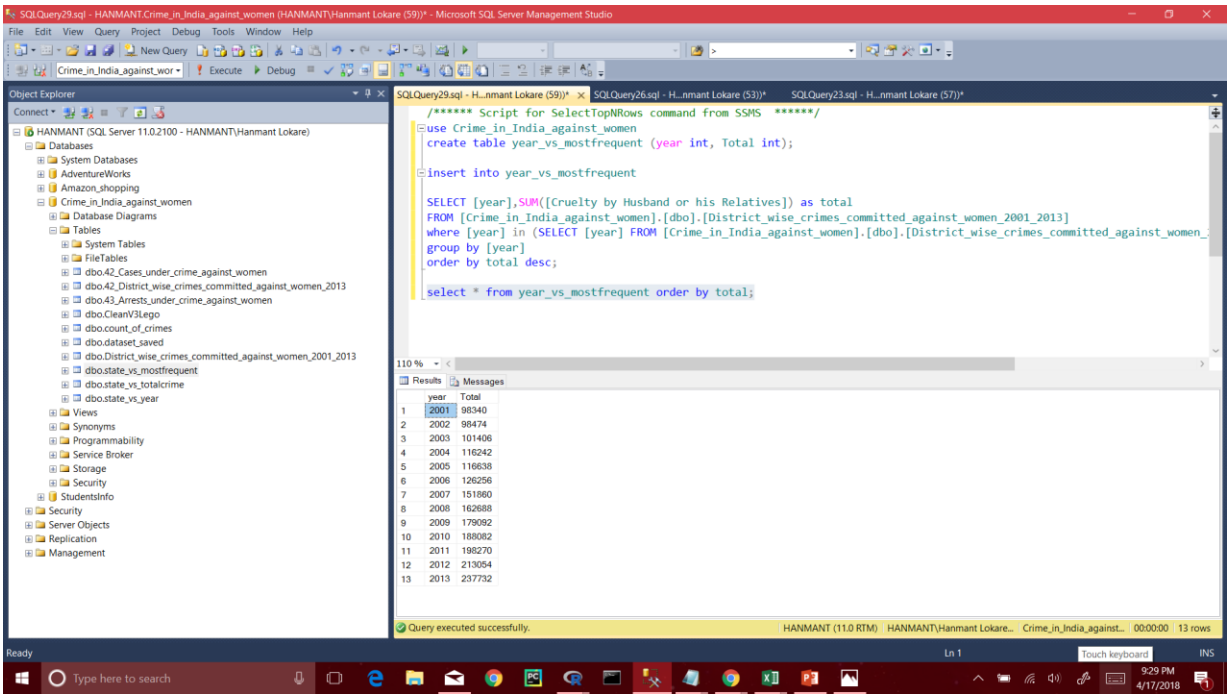

Fig.8

To visualize the above data, we will use a horizontal bar graph.This can be done using the following R code.

```
> con<-odbcConnect("Crime_in_India_against_women")
> res <- sqlQuery(con, 'select * from year_vs_mostfrequent')
> scale<-res[,1]

> p<-ggplot(data=df, aes(x=year, y=Total)) +
+   geom_bar(stat="identity",fill="steelblue")+
+ theme_minimal()+
+ scale_x_continuous(breaks=c(2001:2013),labels=factor(2001:2013))+
+ scale_y_continuous(limits=c(0,250000))
> p
> p + coord_flip()
```

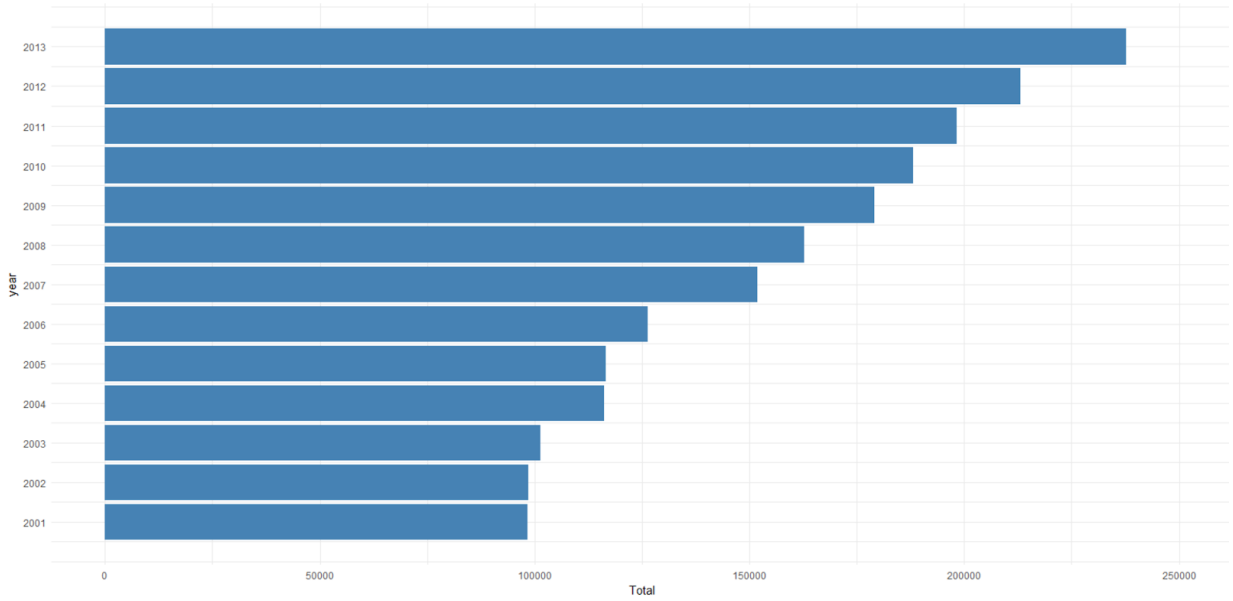

Fig.9

In this above graph we can see that the maximum number of crimes have occurred in the year 2013. The count has been increasing since the year 2001.

Most frequent Crime – Cruelty By Husband / Relatives versus every state.

We will now count the number of cruelty by husband crimes for every state over the period of 10 years. To do this we can use the sum function and count the number of cruelty by husband crimes in every state.

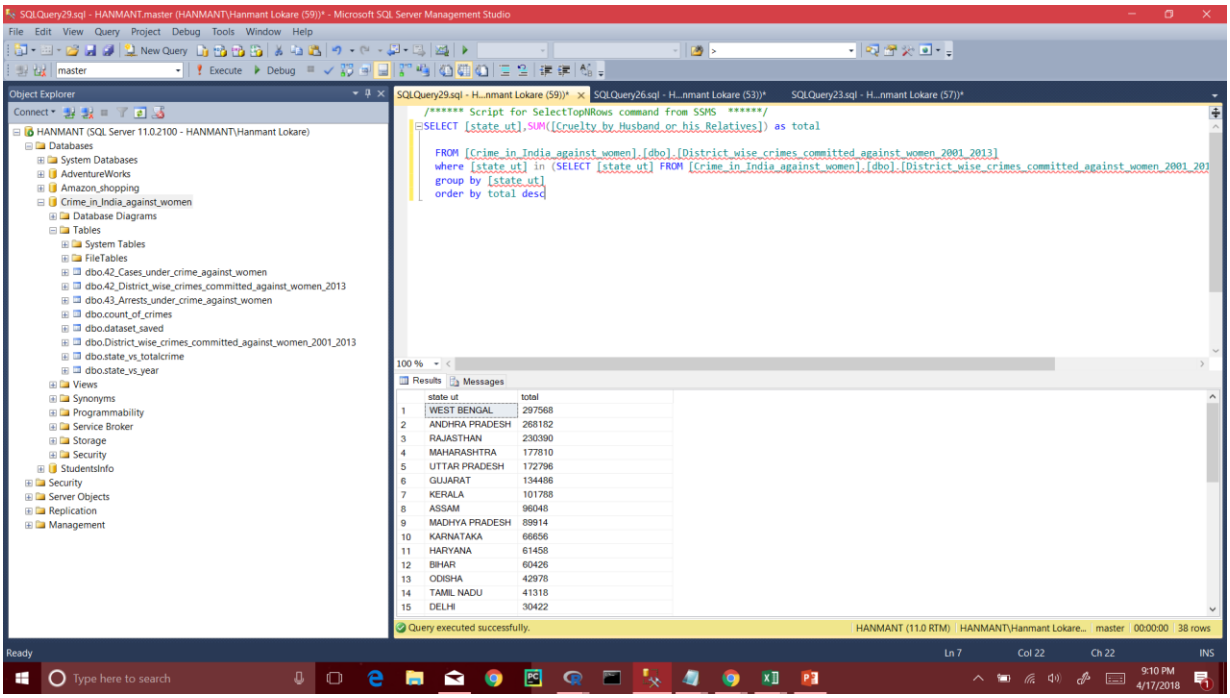

Fig.10

To visualize this we have used a lollipop chart which is similar to a bar graph in R.
The following code was used to generate the above graph.

R-code:

```
> con<-odbcConnect("Crime_in_India_against_women")
> res <- sqlQuery(con, 'select * from state_vs_mostfrequent')
> library(ggplot2)
> theme_set(theme_bw())
> cty_mpg<-res
> ggplot(cty_mpg, aes(x=state, y=Total)) +
+   geom_point(size=3) +
+    geom_segment(aes(x=state,
+                     xend=state,
+                      y=0,
+                      yend=Total)) +
+    labs(title="Lollipop Chart",
```

```
+  subtitle="Total_vs_state")+
+  theme(axis.text.x = element_text(angle=45, vjust=0.45))
```

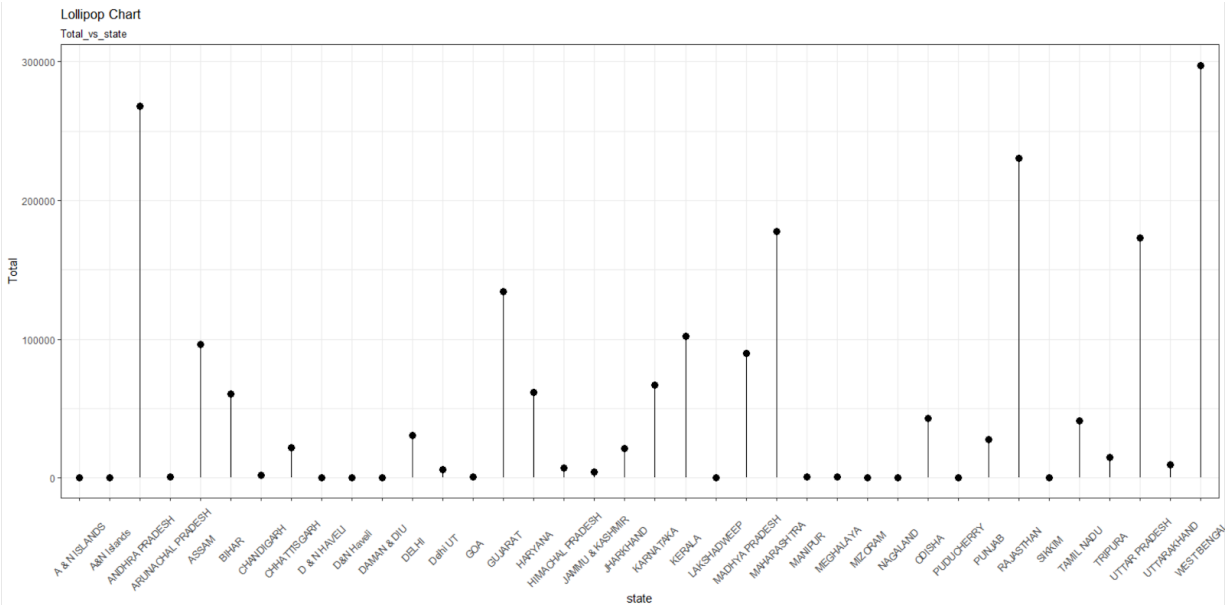

Fig.11

- This graph shows us that the most number of Cruelty by husband crimes were committed in West Bengal.
- The lease number of Cruelty by husband crimes were committed in Lakshadweep.

DISTRIBUTION OF RAPE (on Women) YEARWISE

We would now turn our focus to the most recent crime cases in India on women which is Rape. So, we will try to find the distribution of rape counts over the years to see if the number of rape counts is increasing or decreasing. To find it we used all the entries of rape cases from our data and calculated the sum of rape cases every year.

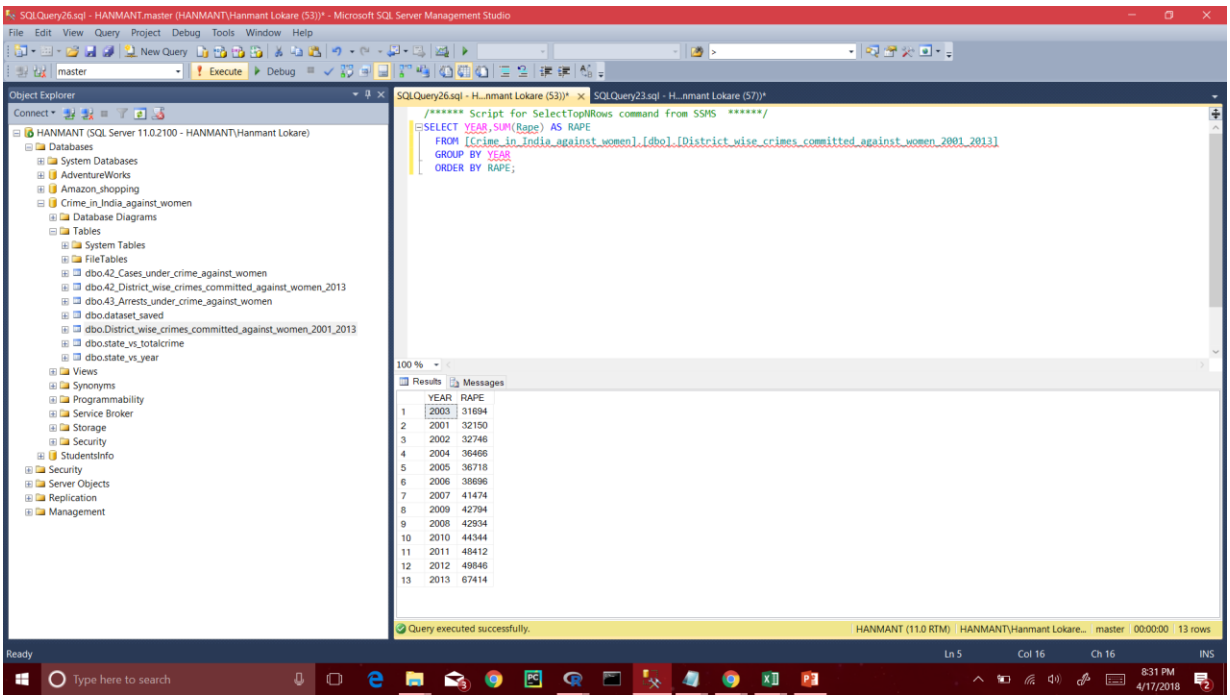

Fig. 12

To represent the data, we used a R function plot () to draw the above graph.

R-code:

*> res <- sqlQuery(con, 'select * from year_vs_rape')*

> scale<-res[,1]

> serial<-res[,2]

> plot(scale,serial, xlab="Year",ylab="Rape",type="p",col='blue',main="Distribution of Rape vs Year",ylim=c(0,35000),pch=24,bg='red')

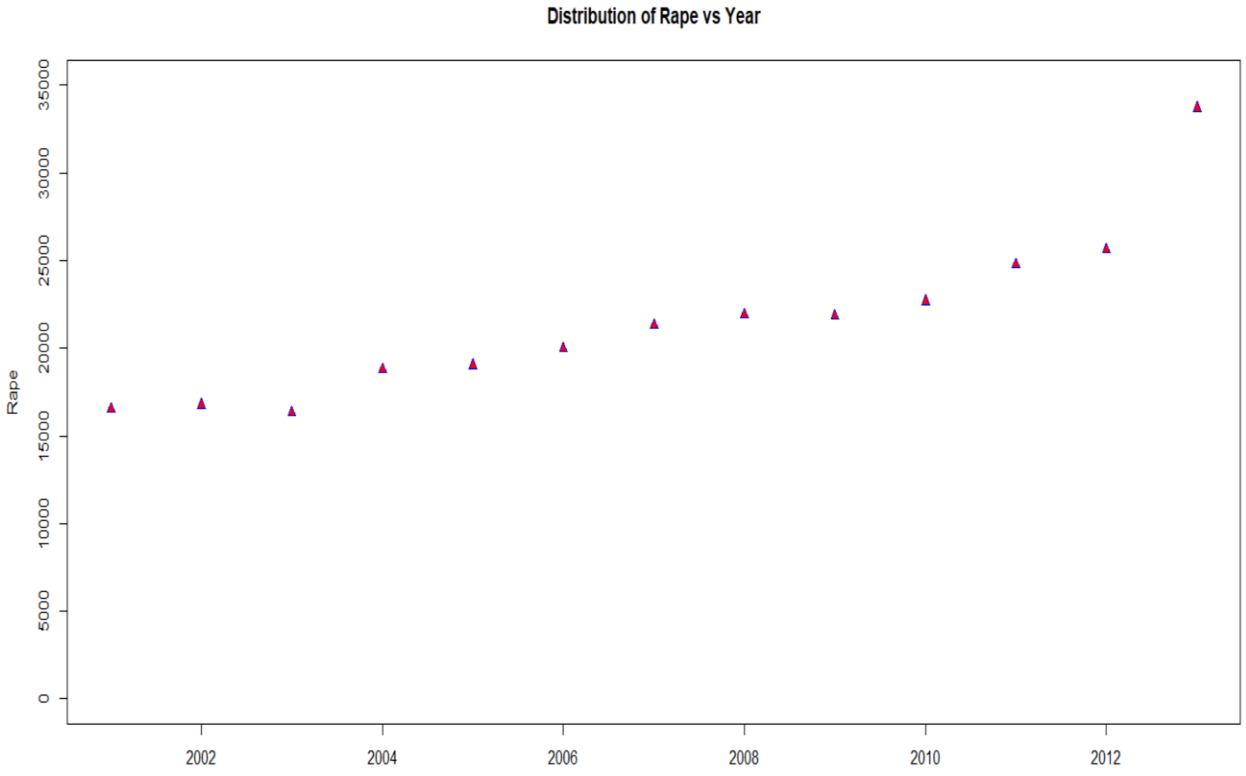

Fig.13

- In the above graph we see an increasing number of rapes every year.
- The highest number of rapes in 2013 are about 30000.
- This increasing rate in the number of crimes should be considered seriously and strict laws should be made for the crime of rape.

DISTRIBUTION OF RAPE VICTIMS AGEWISE

After analyzing the data for rape count over year we will now see age wise distribution of rape victims in India. We will use the following query in SSMS to extract the data.

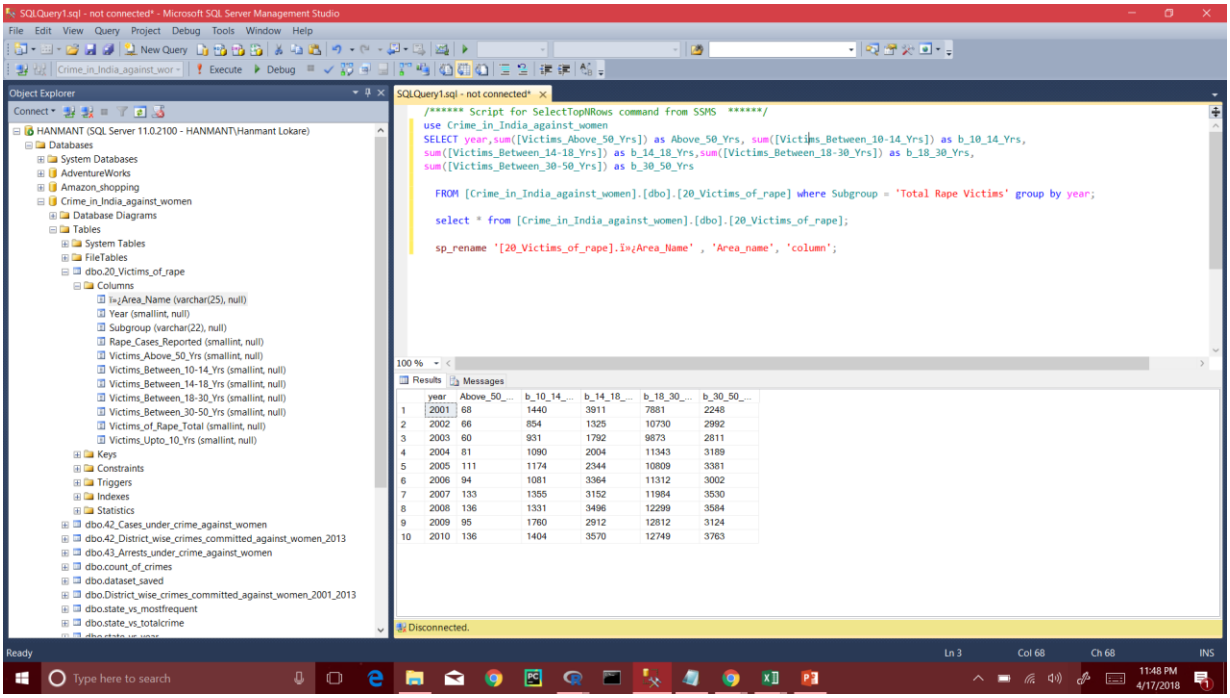

Fig.14

To visualize the above data, we again used a Pie chart using the Pie function in R.

Rcode:

```
> res <- sqlQuery(con, 'select * from age_vs_rape')
> scale<-res[,1]
> slices<-c(scale[,1],scale[,2],scale[,3],scale[,4],scale[,5],scale[,6])
> lbls <- c("(Above_50)","(Between_10_14)", "(Between_14_18)", "(Between_18_30)",
"(Between_30_50)", "(Upto_10)")
 > lbls <- paste(lbls, pct)
> lbls <- paste(lbls,"%",sep="")
> pie(slices,labels = lbls, col=rainbow(length(lbls)),main="Rape_victims_age_wise")
```

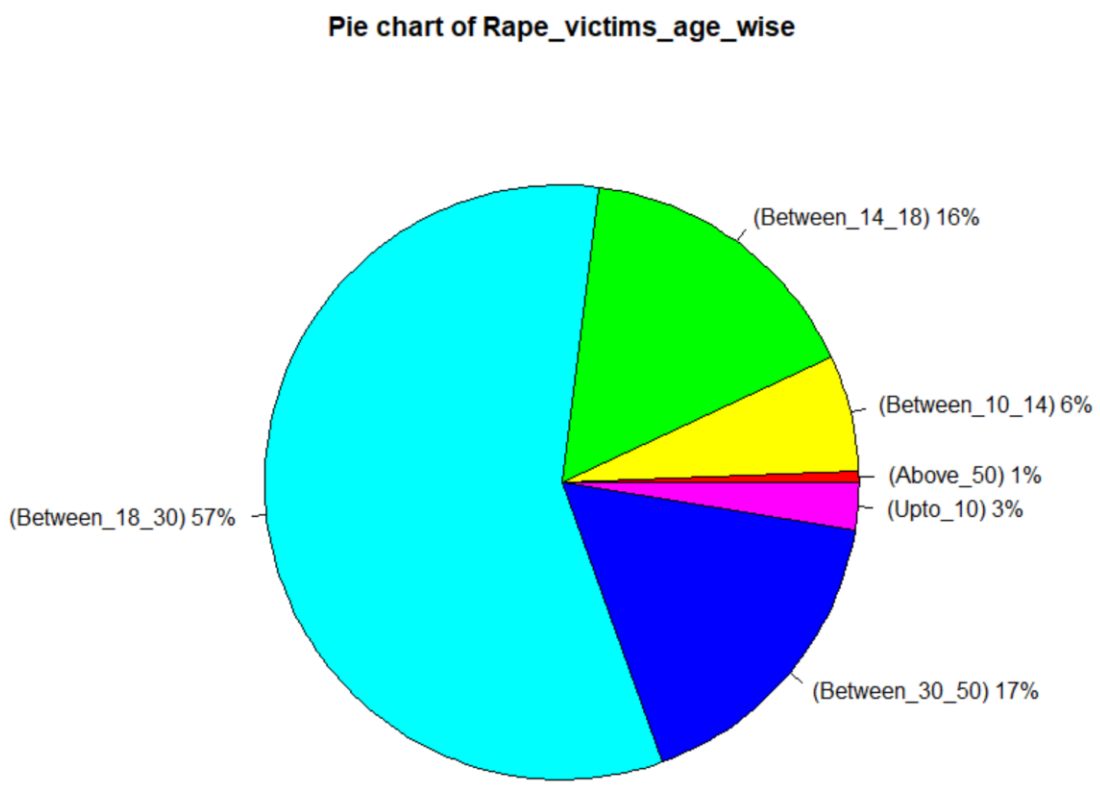

Fig. 15

- The above pie chart shows that majority rape victims were aged in between 18-30 years which accounts for about 57% of the total rape victims
- The woman aged above 50 shows only 1% of the total rape victims
- We can also see that rape against young girls up to 18 years accounts for about 23% of the total count

DISTRIBUTION OF RAPE STATEWISE

We will now analyze the distribution of rape crimes in different states in India. To do this we will count the sum of rape cases in each state of India over the period of 10 years

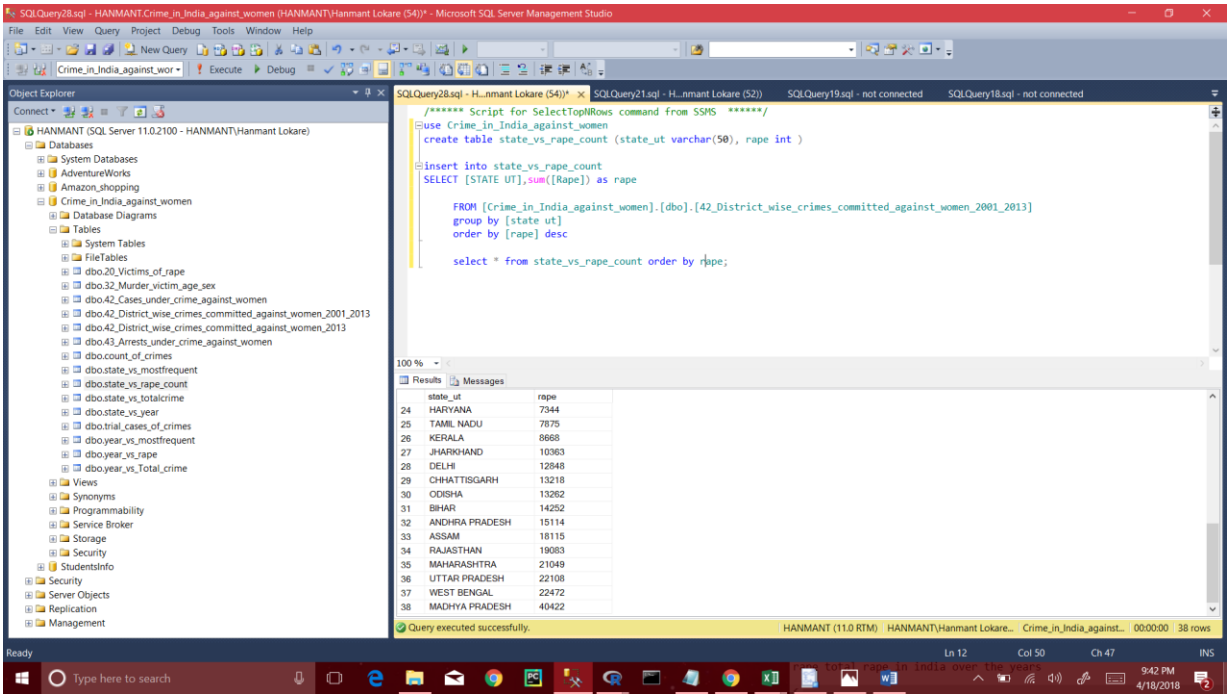

Fig.16

To visualize this data we have used a Doughnut plot in R using the following code.

R-Code:

```
> res <- sqlQuery(con, 'select * from state_vs_rape_count')
> scale<-res[,1]
> serial<-res[,2]
> dat = data.frame(count=serial, category=scale)
> dat$fraction = dat$count / sum(dat$count)
> dat = dat[order(dat$fraction), ]
> dat$ymax = cumsum(dat$fraction)
> dat$ymin = c(0, head(dat$ymax, n=-1))
> p1 = ggplot(dat, aes(fill=category, ymax=ymax, ymin=ymin, xmax=4, xmin=3)) +
+ geom_rect() +
+ coord_polar(theta="y") +
+ xlim(c(0, 4)) +
+ annotate("text", x = 0, y = 0, label = "") +
+  labs(title="Doughnut Plot of Rape_Count in every state")
```

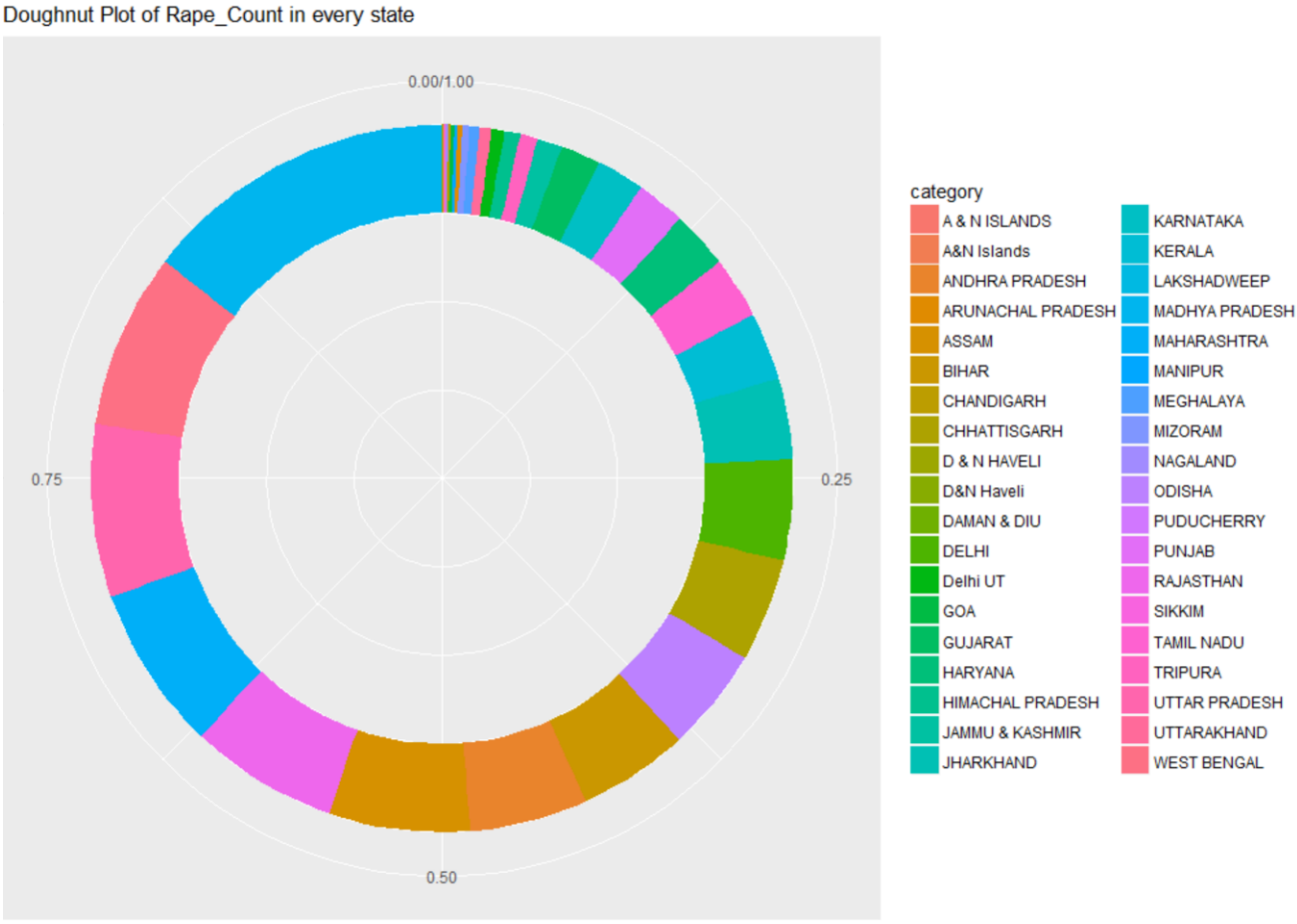

Fig.17

- In this doughnut plot, we can see the distribution of rape counts in every state throughout the course of 10 years.
- Different colors represent different states in the graph.
- The maximum number of rape counts is in the state of Madhya Pradesh.

DISTRIBUTION OF TOTAL CRIME AND CRIME AGAINST WOMEN

We will now compare the total number of crimes and crime against women. We can do this by counting the sum of crimes in every year and sum of crimes in against women every year.

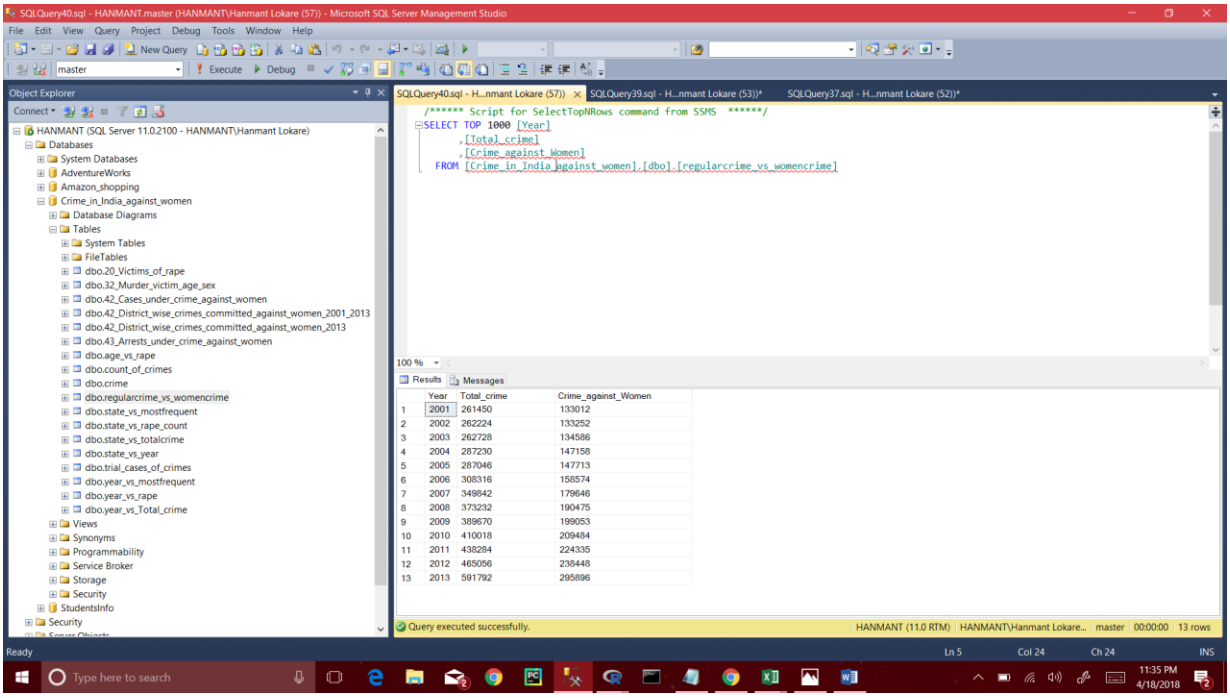

Fig.18

To visualize the above data we have used stacked bar graph. This can be done using the following R code.

R-code

```
> data
      2001   2002   2003   2004   2005   2006   2007   2008   2009   2010
women   133012 133252 134586 147158 147713 158574 179646 190475 199053 209484
regular 128438 128972 128142 140072 139333 149742 170196 182757 190617 200534
      2011   2012   2013
women   224335 238448 295896
regular 213949 226608 295896
> barplot(data, col=colors()[c(23,89,12)] , border="white", space=0.04, font.axis=2,
ylab="Total_crimes",xlab="Year",legend=rownames(data),args.legend=list(x="top", bty='n',
inset=c(-0.10,0)),ylim=c(0,600000),main="Stacked Bar Plot of Crime_against_women and
regular_crimes")
```

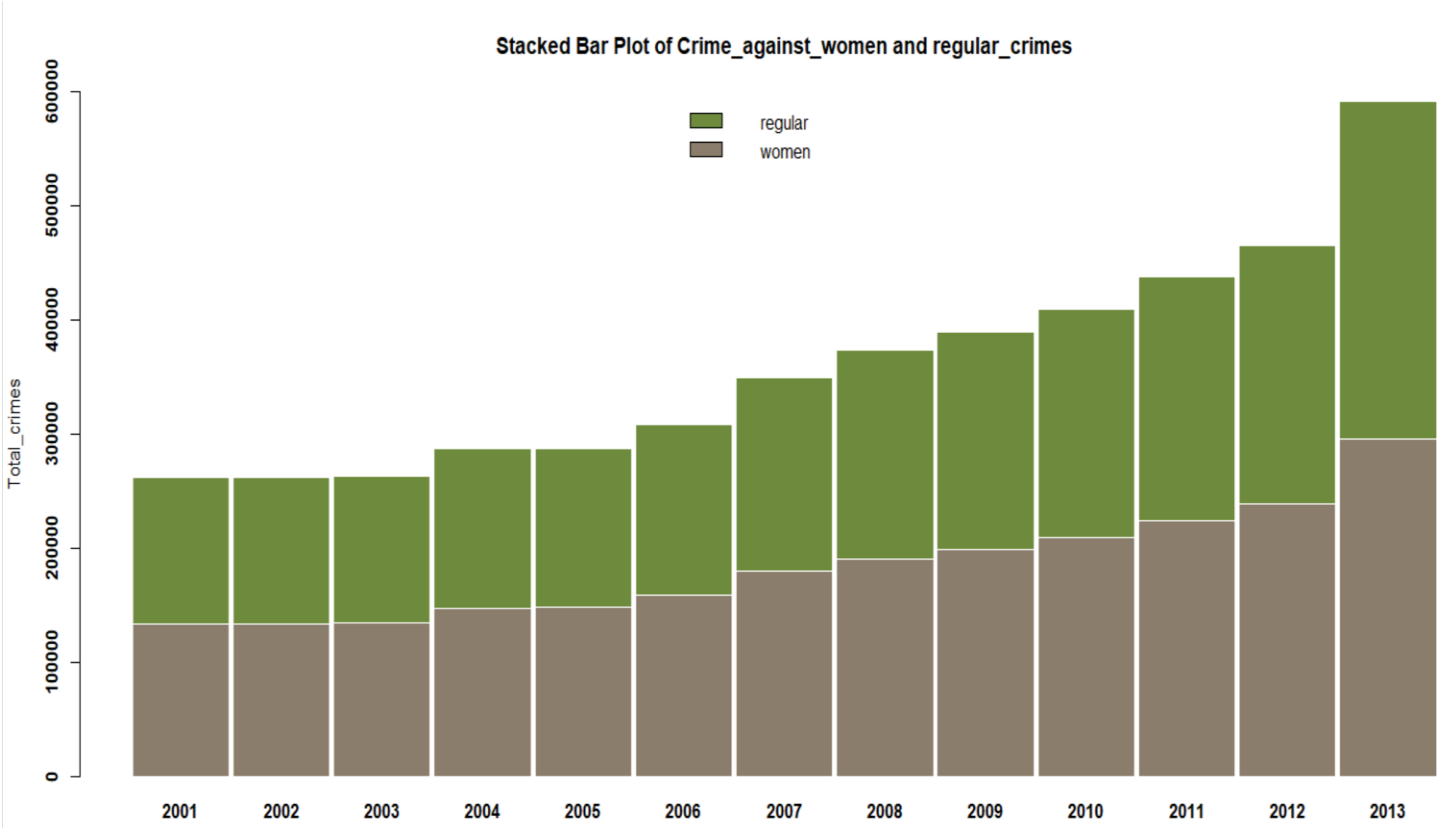

Fig.19

- In this above stacked bar graph we can clearly see that crime against women is about 50% of the total crime every year.
- It also shows us that crime rate against women increases proportionally with the increasing number of total crimes.

SCHEMA

Murder_Victims_age_sex(Area_name varchar, year int,group_name varchar,sub_group_name varchar,Victims_Total,Victims_Above_50_Yrs int, Victims_Between_10_Yrs int,Victims_Between_15-18_Yrs int,Victims_Between_18-30_Yrs int,Victims_Between_30-50_Yrs int)

42_Cases_under_crime_against_women(Area_Name,Year int,Group_Name varchar,Sub_group_name varchar)

42_District_wise_crimes_committed_against_women_2001_2013(STATE UT varchar,DISTRICT varchar,Year int,Rape int,Kidnapping and Abduction int,Dowry Deaths int,Assault on women with intent to outrage her modesty int,Insult to modesty of Women int,Cruelty by Husband or his Relatives int,Importation of Girls int)

42_District_wise_crimes_committed_against_women_2013(STATE UT varchar,DISTRICT varchar,Year int,Rape int,Kidnapping and Abduction int,Dowry Deaths int,Assault on women with intent to outrage her modesty int,Insult to modesty of Women int,Cruelty by Husband or his Relatives int,Importation of Girls int)

43_Arrests_under_crime_against_women(Area_Name varchar,year int,Group_name varchar,Sub_group_name varchar)

age_vs_rape(Above_50 int,Between_10_14 int,Between_14_18 int,Between_18_30 int,Between_30_50 int,Upto_10 int)

count_of_crimes(Type_of_crime varchar, Total_number int)

crime(year int,crime_against_women int)

regularcrime_vs_womencrime(year int,Total_crime int,Crime_against_Women int)

state_vs_mostfrequent(state varchar, Total int)

state_vs_rape_count(state_ut varchar,rape int)

state_vs_totalcrime(state varhcar, total_crime int)

state_vs_year(state varchar, year int,TOTAL_CRIMES int)

year_vs_mostfrequent(year int,Total int)

year_vs_rape(year int, Rape_count int)

year_vs_Total_crime(year int,Total_crime int)

CONCLUSION

After analyzing the crime data of India over the period of ten years we can conclude and answer the following questions:

1. Is the crime rate in India decreasing or increasing?

The crime rate in India is increasing since the past 10 years or so. The government has been unable to control crimes in the country. Even after putting several measures to stop the crimes, there are loopholes which have allowed these crimes to flourished, and hence the law system should be rechecked.

2. Which state has the most/least number of Crimes?

Andhra Pradesh and **Uttar Pradesh** are the most criminal states in India**.**

Lakshadweep has the least number of crimes.

3. Most frequent crime in India?

Cruelty by husband / Family is the most frequent crime in India and that is in the state of Andhra Pradesh.

4.How does the most frequent crime trend in India?

As expected this crime shows an upward increasing trend over the years

5.Are women safe in India?

Crime against women accounts for about 50% of the total crimes committed every year.

These crimes are Cruelty, Rape, Abductions, Importation

Lakshadweep has accounted as the least crime against women and so is the safest state for women

Whereas **Andhra Pradesh** as the most unsafe state for women.

6. Other serious crimes in India?

Over the past couple of years, the number of rape cases have increased. This must be taken seriously. Even after putting efforts to put new laws and acts against rape, there has been no effect on the count of rapes.

REFERENCES

1. https://www.thenewsminute.com/article/bengaluru-chennai-record-highest-crimes-among-south-metros-india-ncrb-data-72460
2. https://en.wikipedia.org/wiki/Crime_in_India
3. https://www.kaggle.com/rajanand/crime-in-india
4. https://www.statmethods.net/graphs/index.html
5. https://en.wikipedia.org/wiki/National_Crime_Records_Bureau
6. https://www.w3schools.com/sql/
7. https://stackoverflow.com/questions/16327971/subtracting-one-data-frame-from-the-other
8. https://stackoverflow.com/questions/8005154/conditionally-remove-dataframe-rows-with-r
9. https://support.discountasp.net/kb/a1179/how-to-import-a-csv-file-into-a-database-using-sql-server-management-studio.aspx
10. https://www.statmethods.net/graphs/index.html
11. https://en.wikipedia.org/wiki/National_Crime_Records_Bureau
12. https://www.w3schools.com/sql/
13. https://stackoverflow.com/questions/16327971/subtracting-one-data-frame-from-the-other
14. https://stackoverflow.com/questions/8005154/conditionally-remove-dataframe-rows-with-r
15. http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html#Lollipop%20Chart
16. https://www.statmethods.net/graphs/density.html
17. http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization