State/UT-wise Details of Crime against Women

Statistical Analysis Report

Name:-

Muhibullah Noorzad (69)

Asim Hussain (14)

Deepak Kumar Dev (20)

Course Name:

Master of Computer Application (MCA)

Subject Name:

Mathematical Techniques for computer Applications

Subbmited to:-

Dr. Vasudha Bhatnagar

Department of computer science

University of Delhi

Delhi-110007, India

State/UT-wise Details of Crime against Women

Crime against women is a critical issue that affects many societies worldwide. According to statisticcal Analysis, women are more likely to be victims of different forms of violence, including domestic violence, rape, sexual assault, and harassment. It is essential to collect and analyse data on these crimes to better understand the scope of the problem, identify patterns and trends, and develop effective interventions to prevent them. As a student, I chose the crime against women data set for several reasons, which I will outline in this report.

Reasons for Choosing Crime against Women Data Set:

The Importance of the Issue: Crime against women is a significant issue that affects the lives of many women worldwide. By analysing the data set, I hope to contribute to the understanding of the extent and nature of the problem, which can inform policy and program development.

Broad Range of Crimes: The crime against women data set includes a wide range of crimes, including domestic violence, sexual assault, rape, and harassment. By analysing these different types of offenses, I can gain a more comprehensive understanding of the issue.

Availability of Data: The data set is readily available, making it accessible for analysis. This allows for a wider range of research and analysis.

Opportunity for Data Analysis: As a student, I am interested in data analysis, and the crime against women data set provides an excellent opportunity to analyse data and identify trends and patterns.

Social Relevance: Crime against women is a socially relevant topic, and the analysis of the data set can provide insights that can inform public discourse and policy development.

There will be several steps to perform the analysis

1. Acquire the data: Collect and import the data into software for analysis.
2. Clean and prepare the data: Check for missing values, errors, and outliers and remove or replace them.

3. Exploratory Data Analysis (EDA): Study the data, calculate basic statistics, and create visualizations to gain insights and identify patterns.
4. Data Transformation: Transform the data if needed to meet the requirements of the statistical models.

Interpretation and Conclusions: Interpret the results, draw meaningful conclusions, and present them in a clear and concise manner.

1) Acquire the Data:

https://www.kaggle.com/datasets/greeshmagirish/crime-against-women-20012014-india/code

```
#install.packages('dplyr')
#install.packages('ggplot2')

df<-read.csv("crimes_against_women_2001-2014.csv",sep=",",header=TRUE,)
```

Output :

STATE/UT	DISTRICT	Year	Rape	Kidnappir	Dowry De	Assault or	Insult to n	Cruelty by	Importation of Girls
ANDHRA F	ADILABAD	2001	50	30	16	149	34	175	0
ANDHRA F	ANANTAP	2001	23	30	7	118	24	154	0
ANDHRA F	CHITTOOR	2001	27	34	14	112	83	186	0
ANDHRA F	CUDDAPA	2001	20	20	17	126	38	57	0
ANDHRA F	EAST GOD	2001	23	26	12	109	58	247	0
ANDHRA F	GUNTAKA	2001	0	0	0	1	0	0	0
ANDHRA F	GUNTUR	2001	54	51	7	139	129	378	0
ANDHRA F	HYDERABA	2001	37	39	24	118	27	746	0
ANDHRA F	KARIMNA	2001	56	49	62	414	81	224	0
ANDHRA F	KHAMMAI	2001	47	30	17	180	336	172	0
ANDHRA F	KRISHNA	2001	37	21	10	208	72	265	0
ANDHRA F	KURNOOL	2001	29	47	13	141	107	92	0
ANDHRA F	MAHABOO	2001	59	27	14	176	41	69	0
ANDHRA F	MEDAK	2001	35	20	26	100	25	192	0
ANDHRA F	NALGOND	2001	35	19	31	188	59	214	0
ANDHRA F	NELLORE	2001	46	80	10	207	228	287	0
ANDHRA F	NIZAMAB	2001	21	21	19	55	15	228	0
ANDHRA F	PRAKASH	2001	19	12	5	140	100	119	0
ANDHRA F	RANGA RE	2001	72	83	37	113	55	421	7
ANDHRA F	SECUNDEF	2001	0	0	1	0	1	0	0
ANDHRA F	SRIKAKUL	2001	8	12	6	38	47	108	0
ANDHRA F	VIJAYAWA	2001	25	48	2	84	122	520	0
ANDHRA F	VIJAYAWA	2001	1	0	0	1	1	0	0
ANDHRA F	VISAKHA F	2001	12	12	3	67	48	99	0

Head of our data

```
> head(df,3)
  X0       STATE.UT  DISTRICT Year  Rape Kidnapping.and.Abduction Dowry.Deaths
1  1 ANDHRA PRADESH  ADILABAD 2001    50                        30           16
2  2 ANDHRA PRADESH ANANTAPUR 2001    23                        30            7
3  3 ANDHRA PRADESH  CHITTOOR 2001    27                        34           14
  Assault.on.women.with.intent.to.outrage.her.modesty Insult.to.modesty.of.Women
1                                                 149                         34
2                                                 118                         24
3                                                 112                         83
  Cruelty.by.Husband.or.his.Relatives Importation.of.Girls
1                                 175                    0
2                                 154                    0
3                                 186                    0
>
```

This information appears to be of woman who faced different kind of crime from 2001 to 2014

STATE/UT: The name of the state or union territory (Discrete Data Type)

Year: The year the data was collected (Ordinal data type)

2) Clean and prepare the data:

Dealing with missing values

When dealing with a missing value, we first look at the dataset's columns. If we discover any gaps in the data there, the NA values are produced as an output, which may not be ideal for all models.

```
#removing the null value

data<-na.omit(data)
```

Operation for Cleaning Data

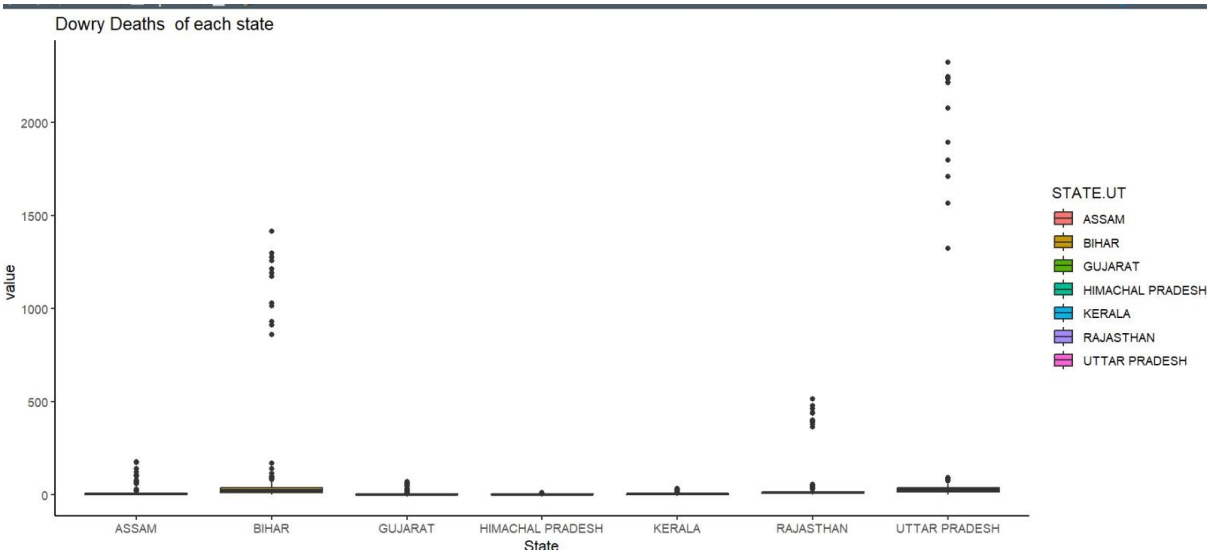
Finding the minimum, maximum, median, and other values of the columns in the given dataset by looking at the summary of the dataset and looking for any N/A values.

```
> #for print the summary  of the data
> summary(df)
   STATE.UT            DISTRICT              Year           Rape
 Length:10677       Length:10677       Min.   :2001   Min.   :   0.00
 Class :character   Class :character   1st Qu.:2004   1st Qu.:   8.00
 Mode  :character   Mode  :character   Median :2008   Median :  22.00
                                       Mean   :2008   Mean   :  57.99
                                       3rd Qu.:2011   3rd Qu.:  44.00
                                       Max.   :2014   Max.   :5076.00
 Kidnapping.and.Abduction  Dowry.Deaths
 Min.   :    0.00         Min.   :   0.00
 1st Qu.:    6.00         1st Qu.:   1.00
 Median :   20.00         Median :   5.00
 Mean   :   69.89         Mean   :  20.18
 3rd Qu.:   49.00         3rd Qu.:  16.00
 Max.   :10626.00         Max.   :2469.00
 Assault.on.women.with.intent.to.outrage.her.modesty Insult.to.modesty.of.Women
 Min.   :    0.0                                     Min.   :   0.00
 1st Qu.:   10.0                                     1st Qu.:   0.00
 Median :   34.0                                     Median :   2.00
 Mean   :  113.5                                     Mean   :  27.42
 3rd Qu.:   85.0                                     3rd Qu.:  12.00
 Max.   :10001.0                                     Max.   :4970.00
 Cruelty.by.Husband.or.his.Relatives Importation.of.Girls
 Min.   :    0.0                     Min.   : 0.0000
 1st Qu.:   11.0                     1st Qu.: 0.0000
 Median :   50.0                     Median : 0.0000
 Mean   :  209.2                     Mean   : 0.1753
 3rd Qu.:  144.0                     3rd Qu.: 0.0000
 Max.   :23278.0                     Max.   :83.0000
```

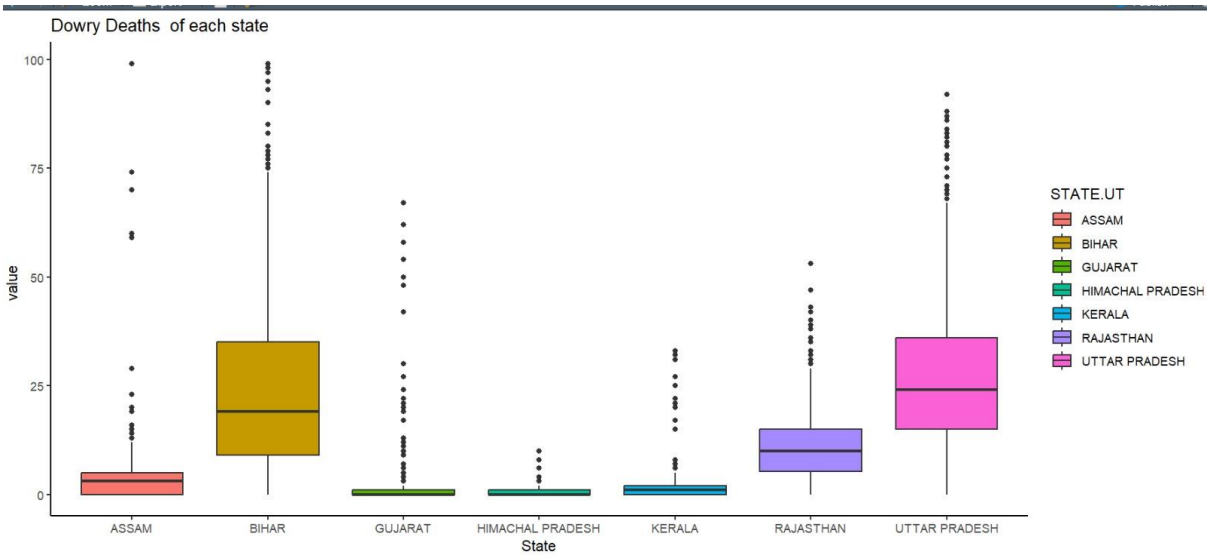
```
> cat("Number of columns : ",ncol(df))
Number of columns :  10
> cat("Number of rows : ",nrow(df))
Number of rows :  10677
> # for print the minimum value of the data kindnapping columns
> cat("Minimum type crime Reported : ",min(df$Kidnapping.and.Abduction))
Minimum type crime Reported :  0
> # for print the maximum value of the data kindnapping columns
> cat("Maximum type crime Reported ",max(df$Kidnapping.and.Abduction))
Maximum type crime Reported  10626
> #for print the the number mean kidnapping columns
> cat("Mean : ",mean(df$Kidnapping.and.Abduction))
Mean :  69.88836
> #for print the the number median  kidnapping columns
> cat("Median : ",median(df$Kidnapping.and.Abduction))
Median :  20
> #for print the  quantile 0.25
> quantile(df$Kidnapping.and.Abduction, 0.25)
25%
  6
> #for print the  quantile 0.50
> quantile(df$Kidnapping.and.Abduction, 0.50)
50%
 20
> #for print the  quantile 0.75
> quantile(df$Kidnapping.and.Abduction, 0.75)
75%
 49
> #for print the fivenum
> fivenum(df$Dowry.Deaths)
[1]    0    1    5   16 2469
```

Query 1.

State wise comparison of Dowry Deaths from 2001 to 2014?



After removing outliers

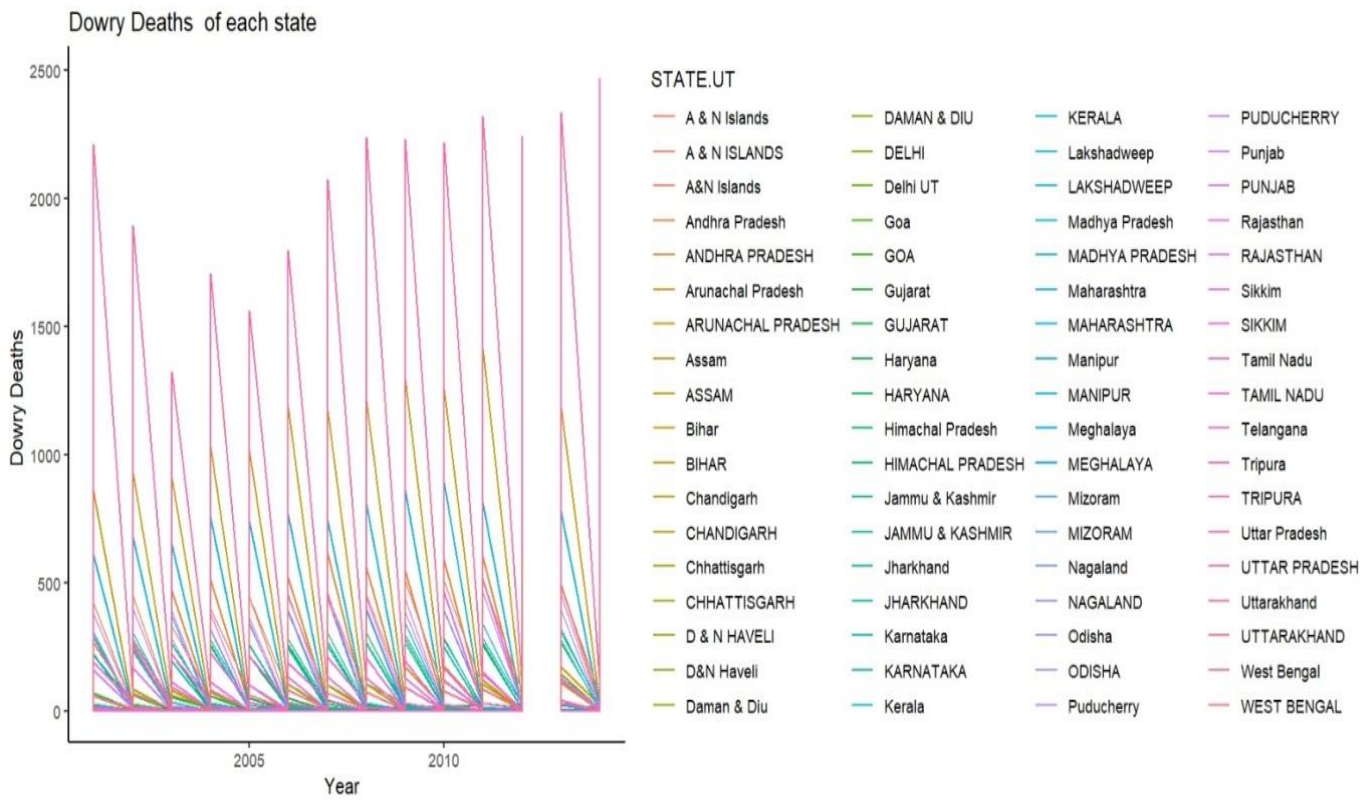


Conclusion:-

Deaths due to the dowry system occurred in Assam, Bihar, Gujarat, Himachal Pradesh, Kerala, Rajasthan, and Uttar Pradesh. We found Bihar and Uttar Pradesh to be the highest among them and Kerala to be the least among them. It may be due to cultural ignorance or illiteracy among the people of Bihar and Uttar Pradesh. And least in Kerala due to its high literacy rate.

Query 2.

Total number of Dowry Death over the years?

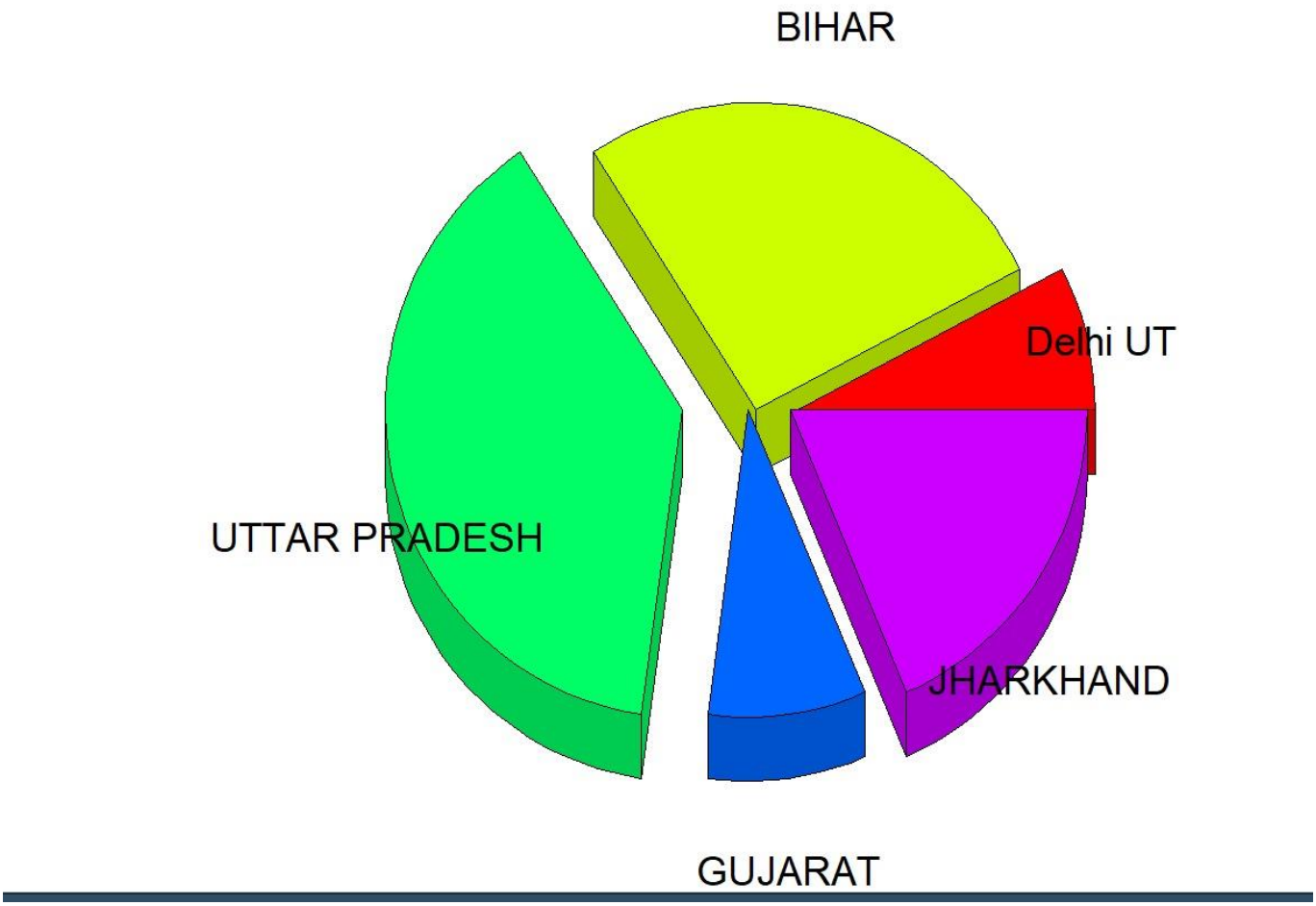


Conclusion:-

We have compared the number of deaths due to the dowry system in each state of India. 2001–2005, the cases of dowry deaths were declining. After 2005, the number of cases increased until 2009, and after that cases each year are constant, and after that cases each year are constant.

Query 3.

State wise comparison of total Rape Cases from 2001 to 2014?

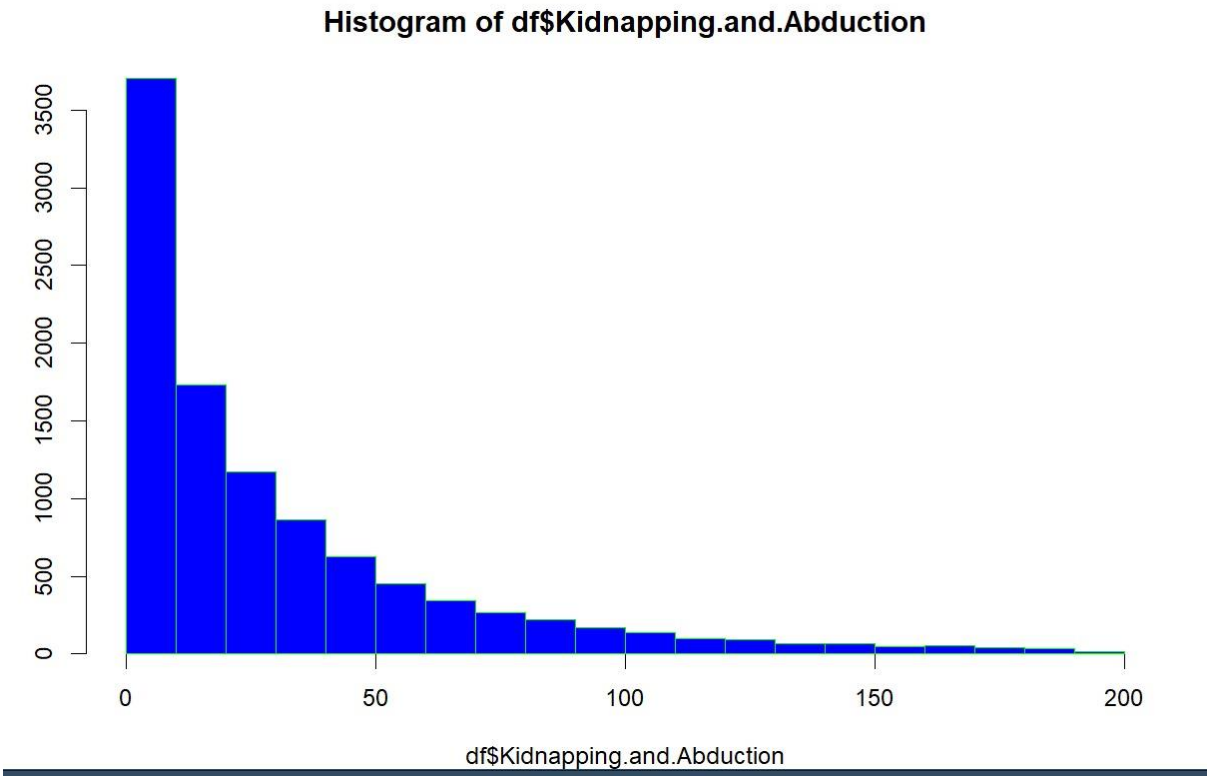


Conclusion:-

When we analysed the rape cases among Bihar, Uttar Pradesh, Delhi, UT, Jharkhand, and Gujarat, we discovered that Uttar Pradesh has the most cases. So that signifies that law and order in Uttar Pradesh is not good, and in terms of women's safety, Uttar Pradesh is not performing well

Query 4.

Frequency of Kidnapping and Abduction from all states.

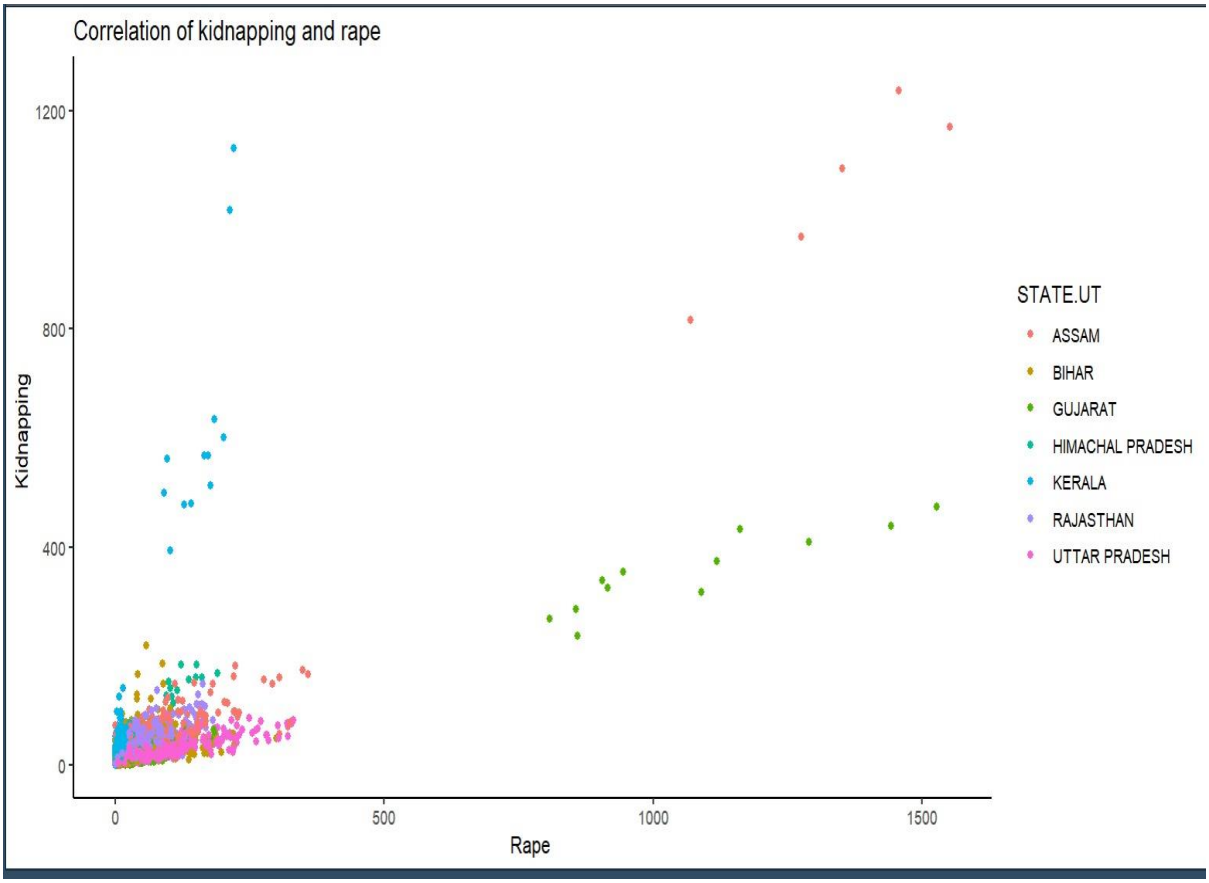


Conclusion:-

Here we have analysed the kidnapping and abduction cases using histograms. Which shows that most of the cases are under 50, and that 50% of the cases are under 10.

Query 5.

Correlation in Rape and Kidnapping in different states.

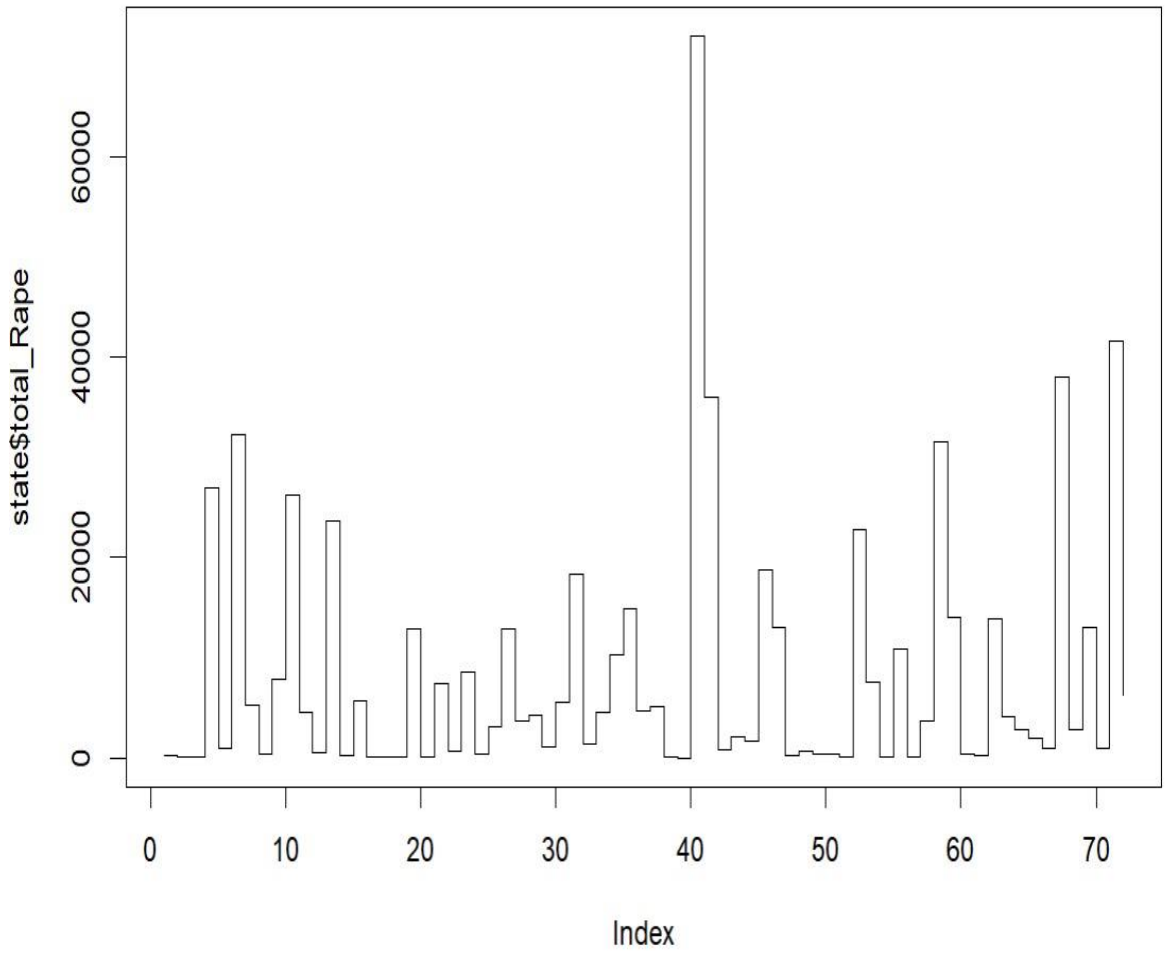


Conclusion:-

This is the correlation between rape and kidnapping in Assam, Bihar, Gujarat, Himachal Pradesh, Kerala, Rajastan, and Uttar Pradesh. So this scatter plot shows that Himachal Pradesh has a high number of kidnapping cases, and Uttar Pradesh has a high number of rape cases, but Assam has both high numbers of kidnapping and rape cases.

Query 6.

Total Rape cases from 2001 to 2014 recoded.

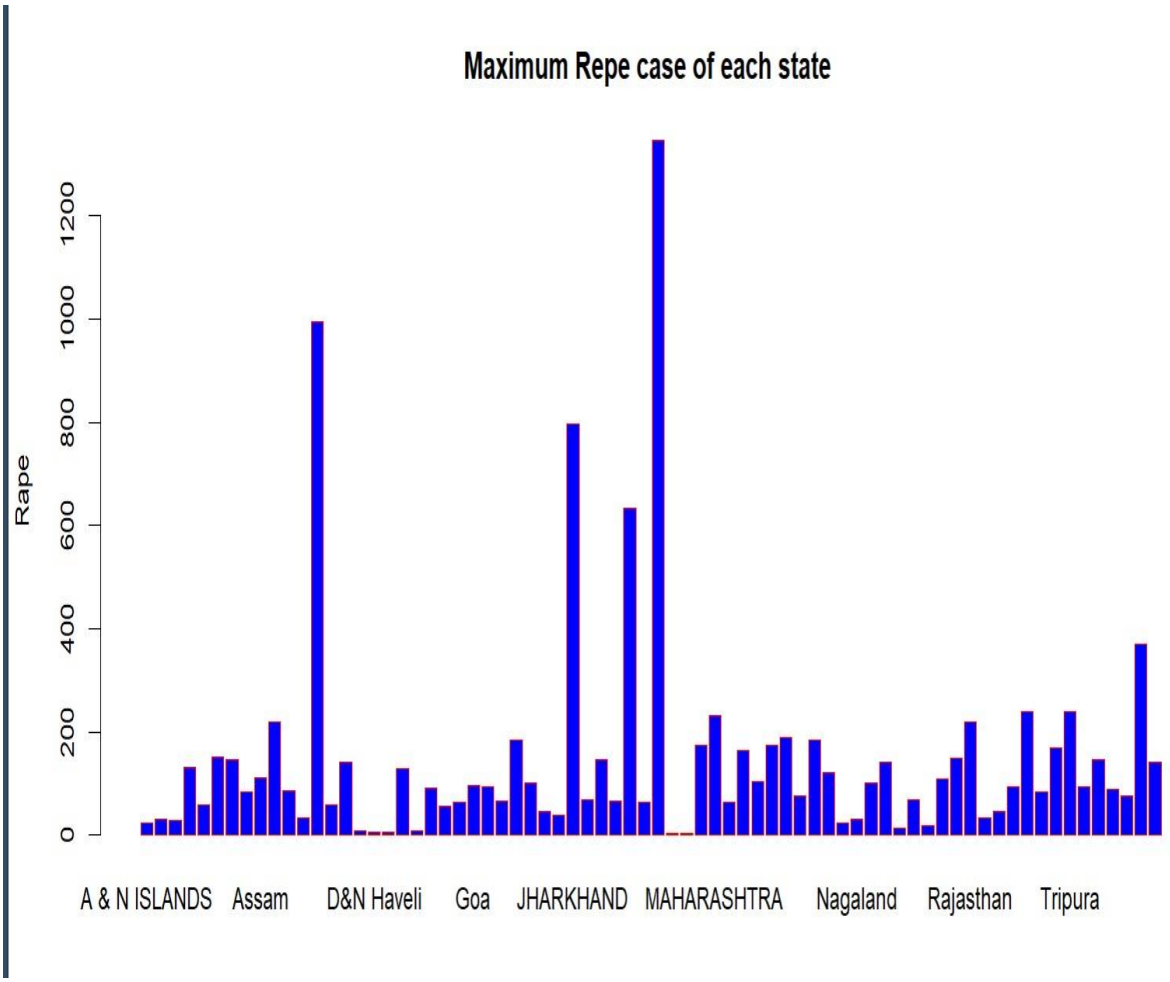


Conclusion:-

Here we have shown the total number of rape cases that happened during 2001–2014 among all the Indian states and UT. This data shows that Uttar Pradesh is the worst state among all states, and Kerala is the best among all states and UT.

Query 7.

Maximum Rape Case of the states.

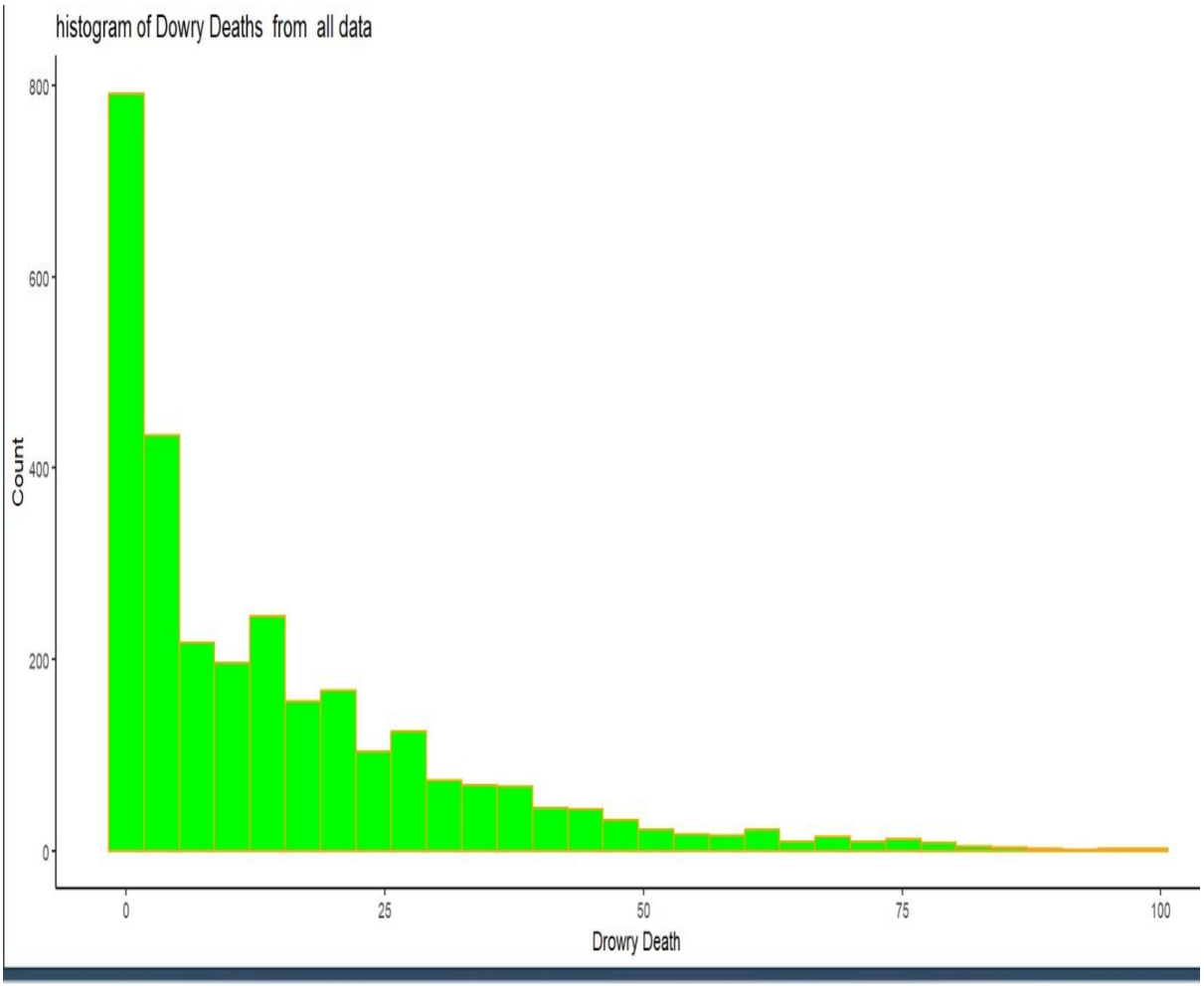


Conclusion:-

We have compared the rape cases among the Indian states, which are: A&N Islands, Assam, D&N Haveli, Goa, Jharkhand, Maharashtra, Nagaland, Rajasthan, and Tripura.

Query 8.

Deaths due to Dowry from all the data.

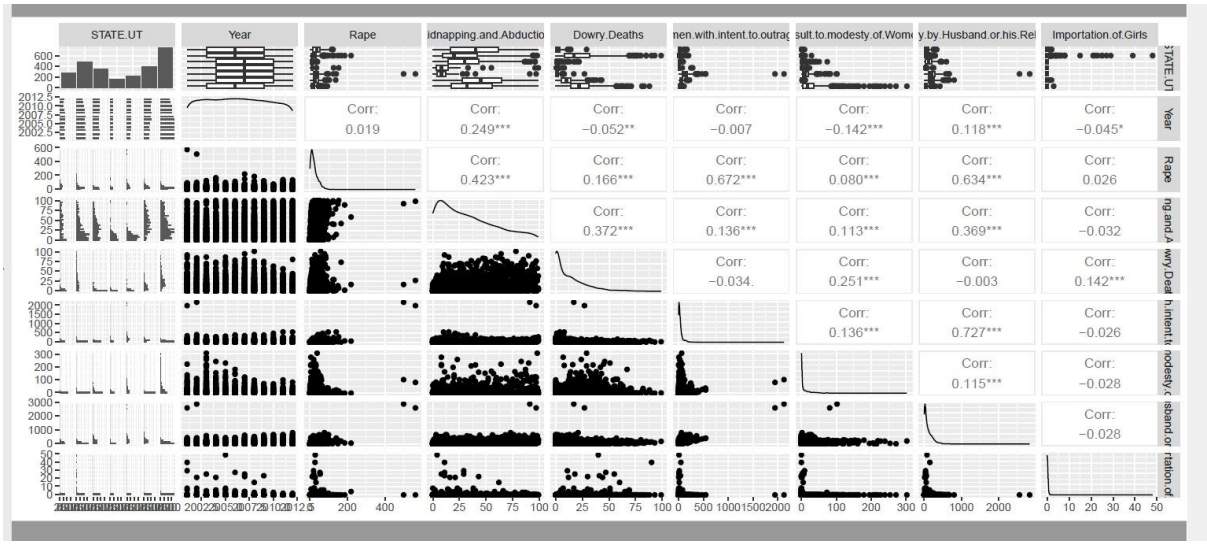


Conclusion:-

Here, we have analysed the dowry death cases in Himachal Pradesh, Kerala, Rajasthan, Assam, Gujarat, Uttar Pradesh, and Bihar using histograms. Which shows that most of the cases are under 25, and that 50% of the cases are under 10.

Query9.

Correlation of all coloumns from the data.



Conclusion:-

This shows the correlation of the entire column in the data sheet, in which we observe the differing correlation in cases and with the states also in numeric and plotted form.