Cyber Crimes Trend in India, Fraud: The Major Motive.

Jaswinderpal Singh
M.Sc. Statistics,
Department of Statistics,
Panjab University Chandigarh.

Narinder Kumar
Professor,
Department of Statistics,
Panjab University, Chandigarh.

Abstract

Cyber Crime is any kind of criminal activity that involves the use of computers, networks, or the internet for the exploitation of data and resources. Examples include hacking, identity theft, phishing scams, and distributing malware etc.

Situation is getting worse with every passing year and at the same time lack of awareness in terms of digital literacy is still there. Currently, the active internet users in India are more than 50% of the country's population. The Cyber Crimes started increasing rapidly by the years 2015-2016. Lack of Digital Literacy and Cyber Safety are some of the reasons. Phishing, Identity Theft, Computer Related Offenses, Frauds, Cyber-Bullying, Spamming, data leaks etc. are major Cyber Crimes on the increase in India. India reported 52,974 incidents of cyber crimes in 2021, an increase of nearly six percent from the year before, in which 50035 was the number.

Here, the objective is to analyze the year-wise trend and future prediction of the number of cyber crimes registered in India using the data from 2002 - 2021. Some state wise comparisons are done, keeping the population variation of the states under consideration. Major Cyber Crime Motive is identified and States with similar crime incidence rates are divided into clusters using Statistical Cluster Analysis Techniques.

Keywords: Cyber Crimes, Phishing, Malware, Fraud, Statistics, Cluster Analysis.

2 Introduction

2.1 Cyber Crimes

Cyber crimes are a new class of crimes rapidly increasing due to extensive use of Internet and I.T. enabled services. Any criminal activity with the use of computers, networks, or the internet for the exploitation of data and resources comes under cyber crimes. Examples include hacking, identity theft, phishing scams, and distributing malware etc. Women are commonly targeted for cyber stalking, cyber pornography, impersonation etc.

2.1.1 Major Cyber Crimes in India

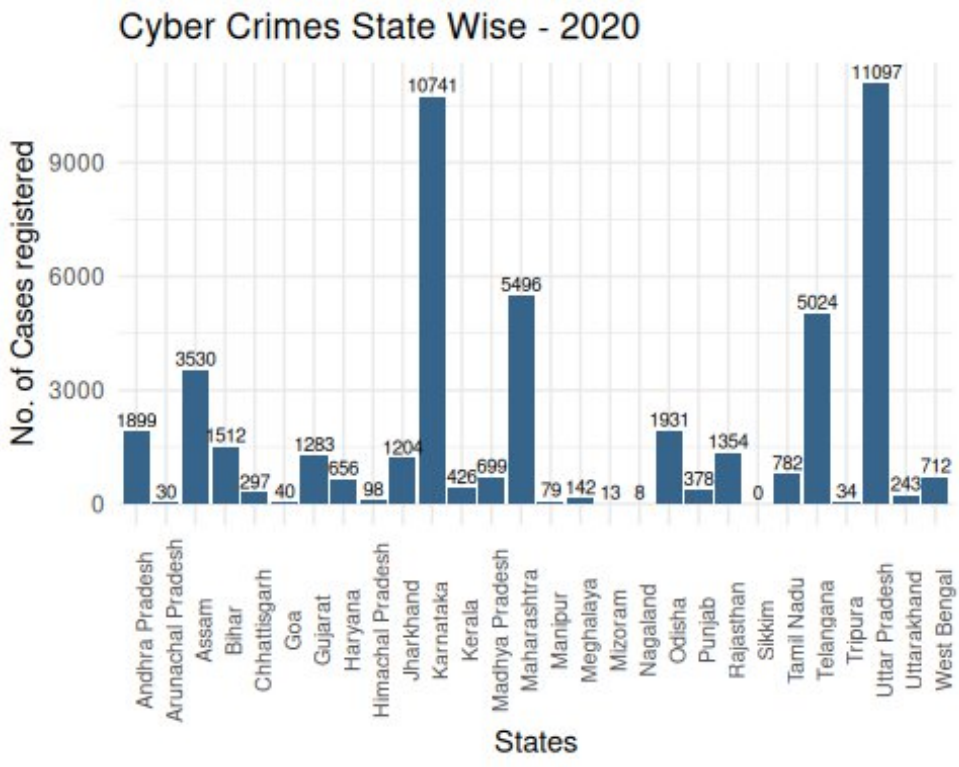
Some major cyber crimes in India include:
1. Hacking: Unauthorized access to computer systems, networks, or websites to steal sensitive information or disrupt operations.
2. Phishing: Attempts to trick individuals into providing personal or financial information through fake emails or websites.
3. Identity Theft: Using someone else's personal information to commit fraud or other crimes.
4. Fraud: Using the internet to scam people out of their money or personal information.
5. Cyberstalking: Harassment or bullying through electronic means.
6. Child pornography: Using the internet to distribute or view child pornography.
7. Ransomware: A type of malware that encrypts a victim's files and demands payment to restore access.
8. Crypto jacking: Unauthorized use of someone's computer or device to mine cryptocurrency.
9. Distributed Denial of Service (DDoS) attacks: Overwhelming a website or network with traffic to make it unavailable.

These are some major cyber crimes reported in India, but there can be various other types of cyber-attacks that are emerging with technology advancements.

3 Cyber Crimes in India: Statistics and Data Visualization

3.1 Cyber Crimes State-Wise:

3.1.1 Year 2020

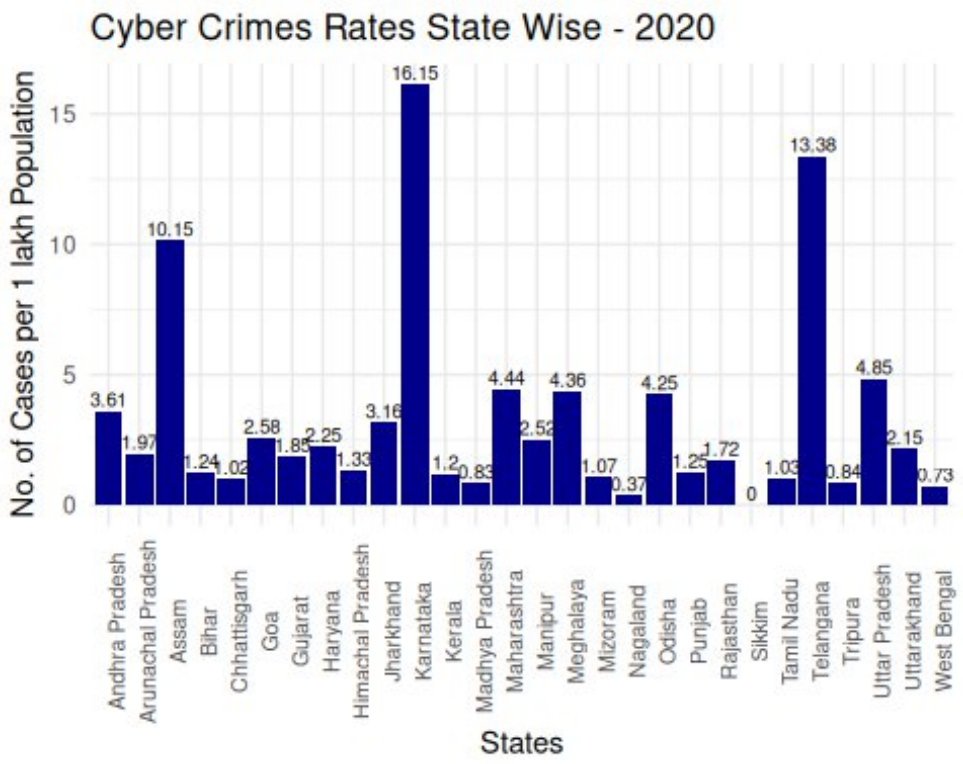


The above figure shows the highest number of cases being registered in Uttar Pradesh, followed by Karnataka, Maharashtra and Telangana. But one must note that the population

in these states may vary. So to get a better comparison, rates can be calculated, which gives the cyber crime cases registered in the state per 1 lakh population. Mid year projected population for each state for the year 2021 is available. Hence the Cyber Crime Rate can be calculated as:

$$Rate = \frac{No.\ of\ cases\ registered}{Mid\ year\ projected\ population\ \ (in\ lakhs)}$$

As far as the rates are concerned, Karnataka and Telangana emerged as the states with highest cyber crime rates followed by Assam, Uttar Pradesh and Maharashtra.

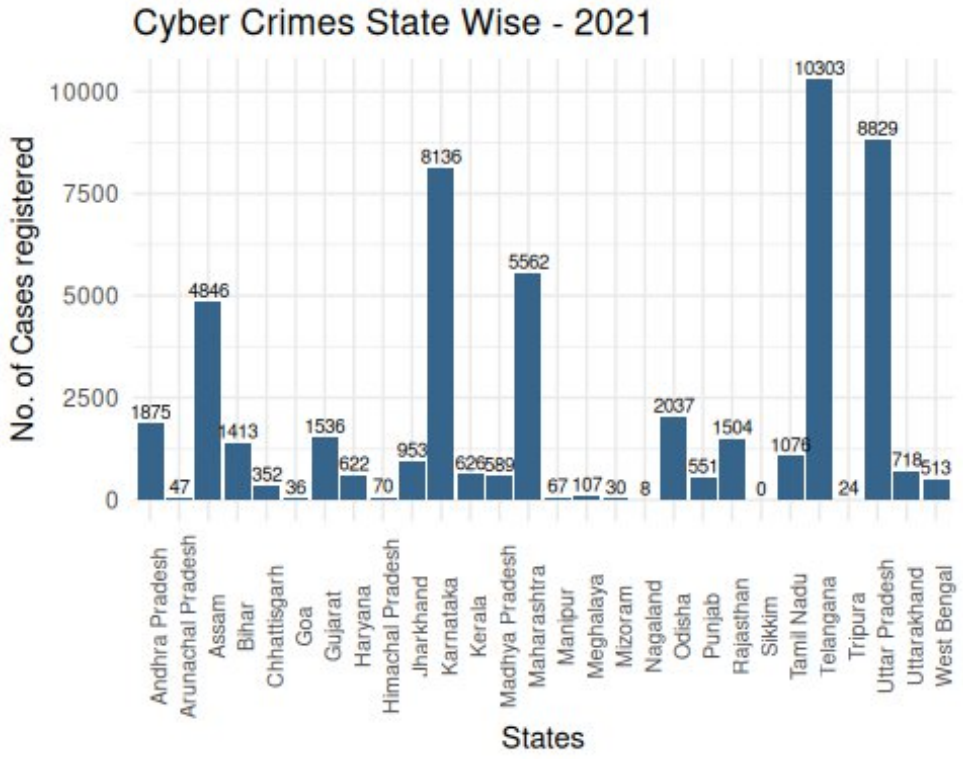


Now, keeping the population variation of the states into account, the better comparative picture here is given the the rates. Now, it can be seen that the state of Karnataka has the maximum cyber crime rate of **16.15** cyber crimes registered per 1 lakh population in the year 2020, followed by Telangana and Assam with cyber crime registeration rate of **13.38** and **10.15** cases per 1 lakh population.

```
## [1] 0.7172579
```

The above value indicates that almost 71.72579 % of the total cyber crimes reported in the country India in 2021 came only from the five states: Telangana, Uttar Pradesh, Karnataka, Maharashtra and Assam.

3.1.2 Year 2021



From the above figure it looks like Telangana tops the Cyber Crimes tally followed by Uttar Pradesh and Karnataka in 2021.

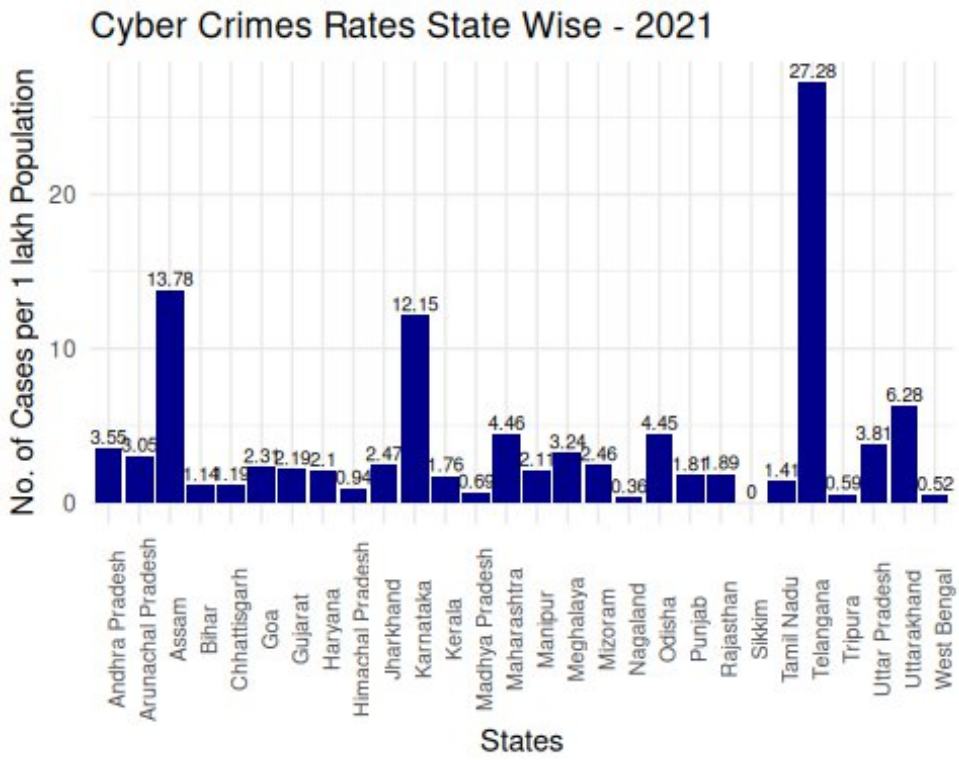
It can be noticed that the cyber crimes in the state of Telangana **almost doubled** in a year, which is a matter of huge concern.

Also, the cyber crimes in Assam are also increasing with every passing year.

However, Karnataka registered slightly lesser number of cyber crimes this year as compared to the previous year.

Despite all, There is no cyber crime case regitered in Sikkim as per the record of National Crime Records Bureau (NCRB) in the past few years.

But for better comparison, let us see the Cyber Crime Rates, i.e. Cyber Crimes in a State per 1 lakh population.



Now, from here it is evident that infact Telangana is the state with highest cyber crime rate of **27.28** cases per 1 lakh population, followed by Assam and Karnataka with rates **13.78** and **12.15** respectively.

```
## [1] 0.7112168
```

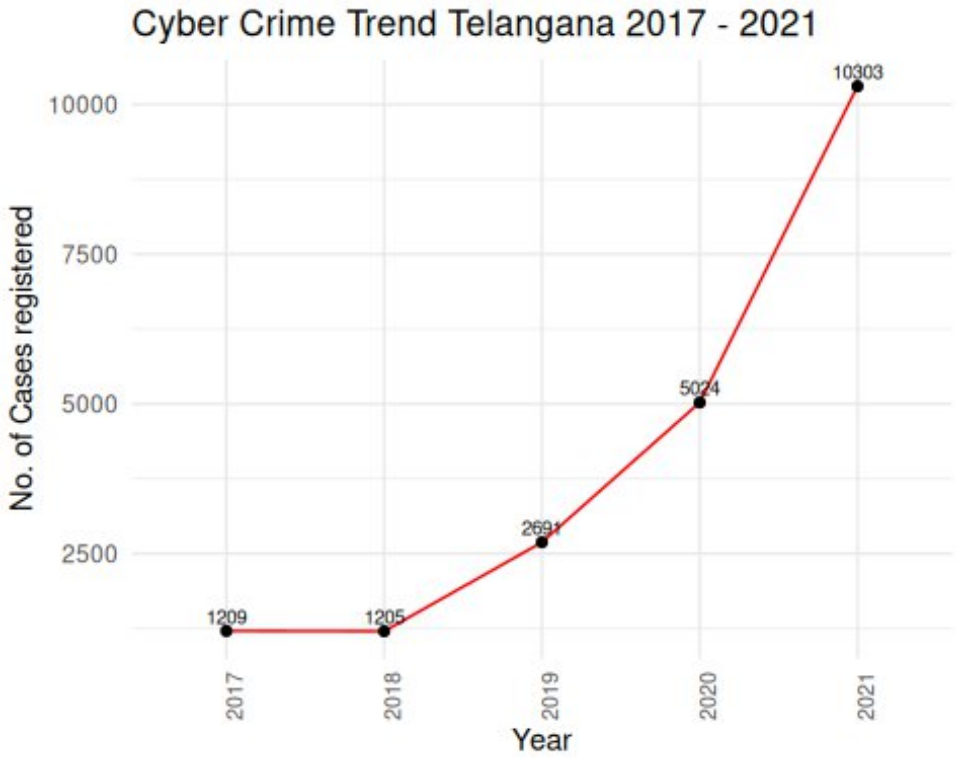
The above value indicates that almost 71.12168 % of the total cyber crimes reported in the country India in 2021 came only from the five states: Telangana, Uttar Pradesh, Karnataka, Maharashtra and Assam. However, they together contributed almost 36% of the Indian Population.

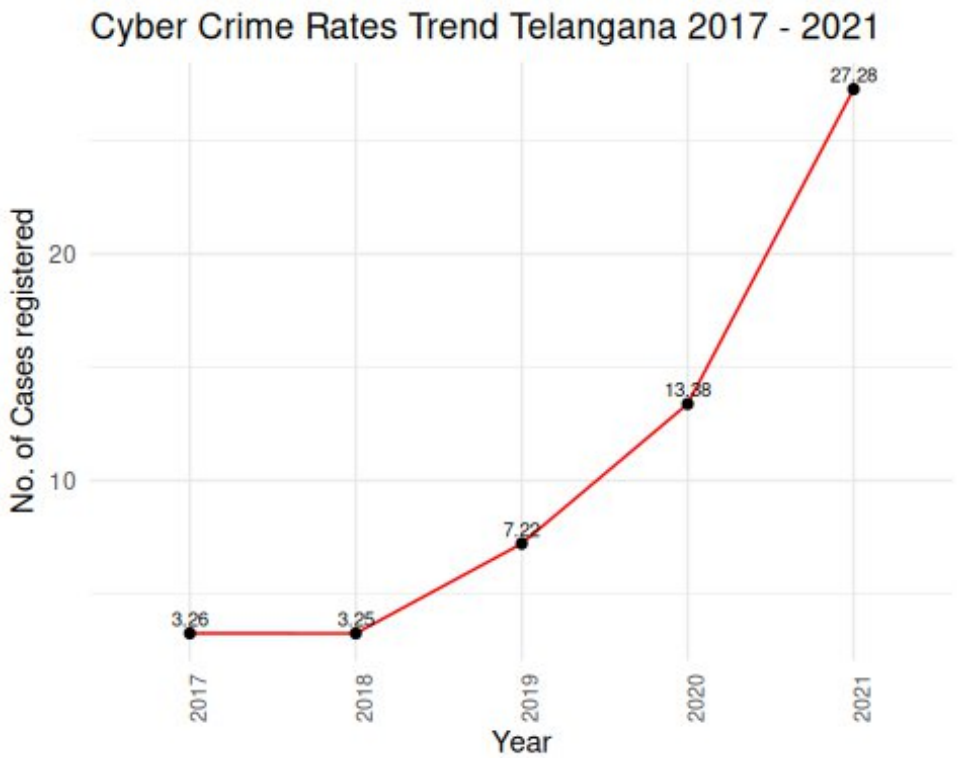
Now, for the past 5 years, the cyber crime trend in these 5 major states is to be viewed. The trend is shown for each state for the years 2017 - 2021. The data can be visualized both in terms of actual crime registered count as well as in terms of rates in these states.

3.1.3 Cyber Crimes Trend: Major States

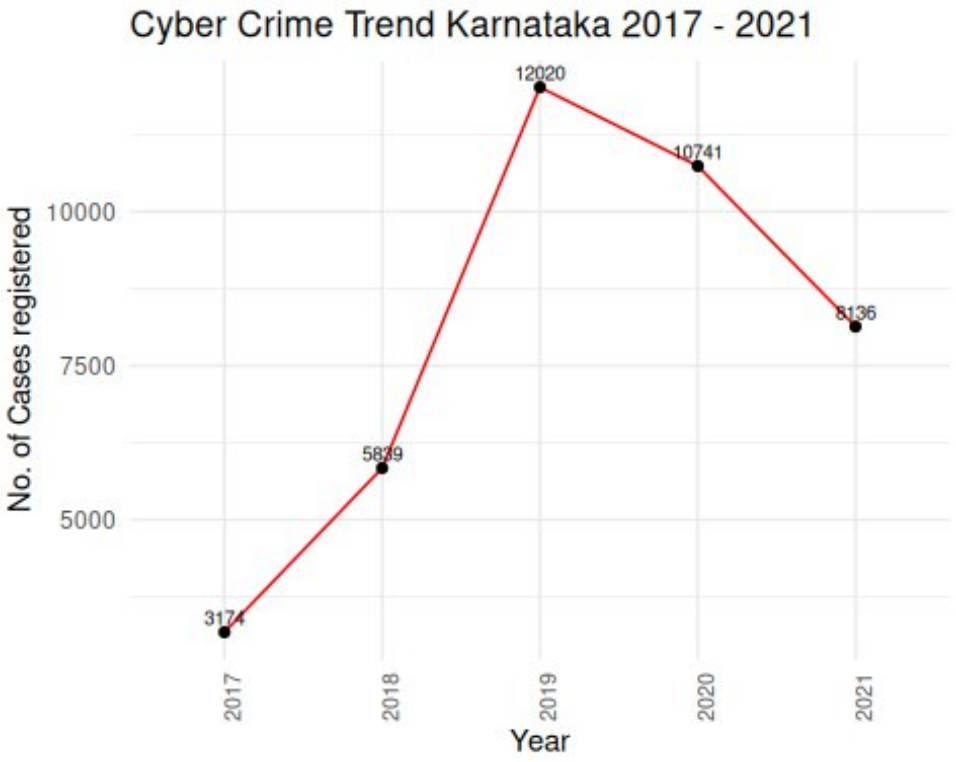
These States can be analyzed further based on last 5 years as follows:

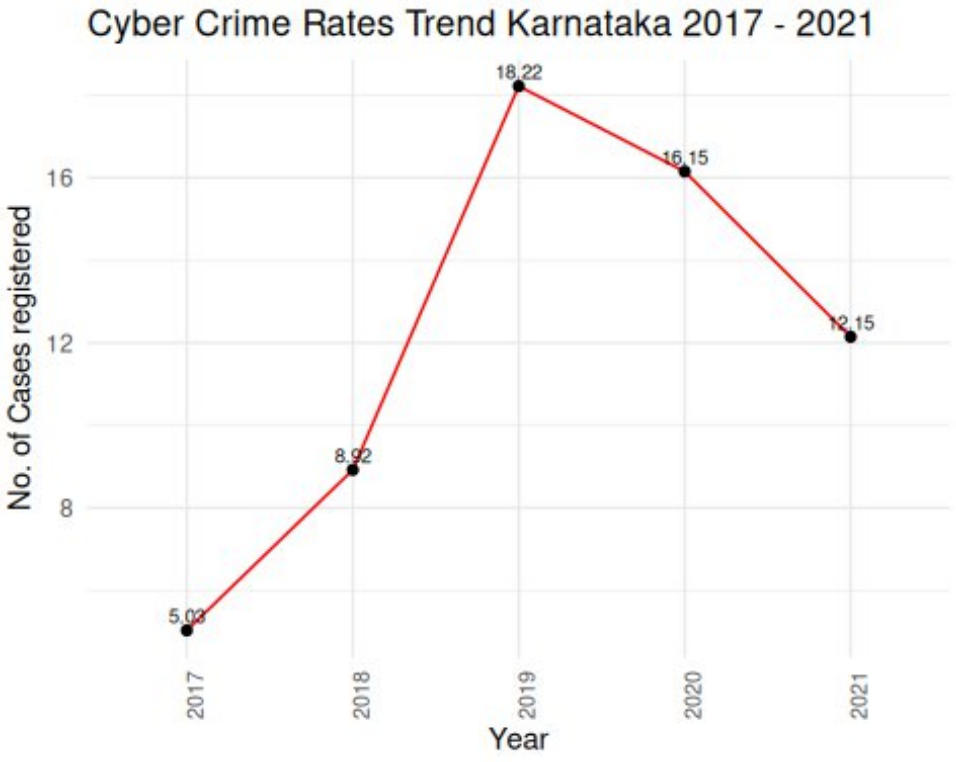
Telangana


Cyber Crime Trend Telangana 2017 - 2021
No. of Cases registered
10000
7500
5000
2500
10303
5024
2691
1209
1205
2017
2018
2019
2020
2021
Year

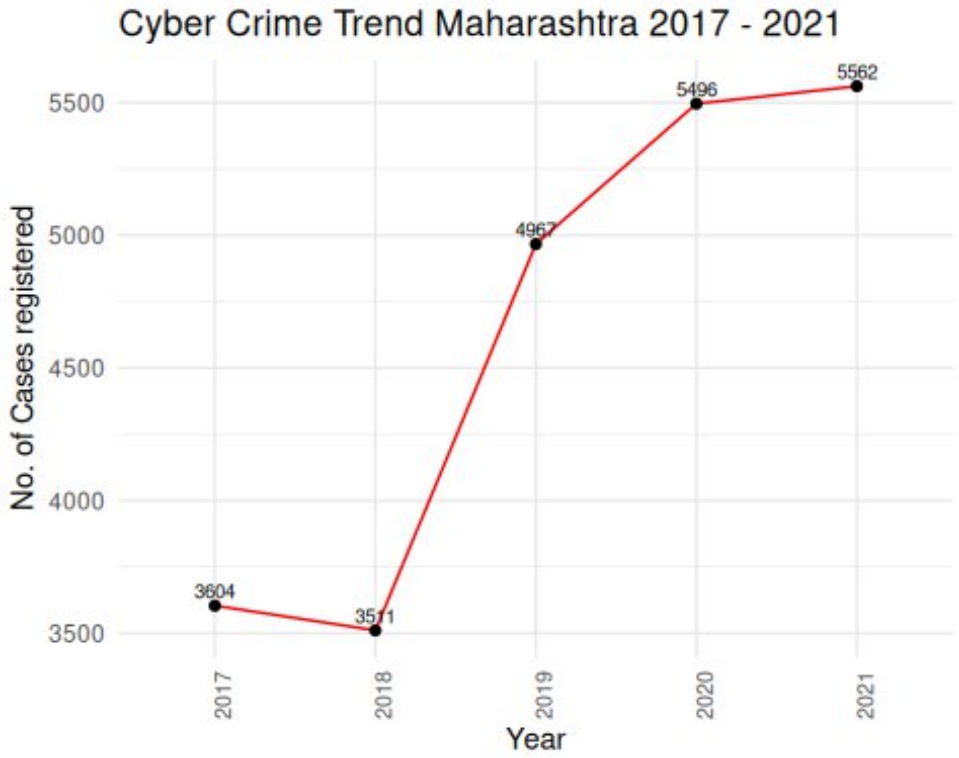

Cyber Crime Rates Trend Telangana 2017 - 2021
No. of Cases registered
20
10
27.28
13.38
7.22
3.26
3.25
2017
2018
2019
2020
2021
Year

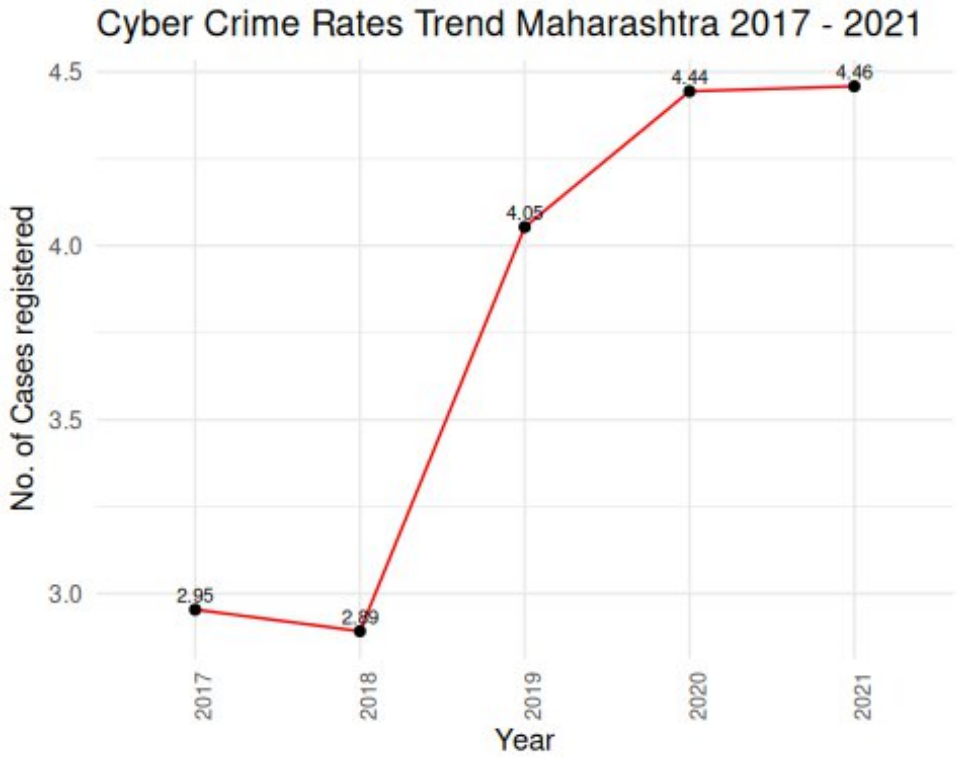
Karnataka


Cyber Crime Trend Karnataka 2017 - 2021
No. of Cases registered
10000
7500
5000
12020
10741
8136
5839
3174
2017
2018
2019
2020
2021
Year

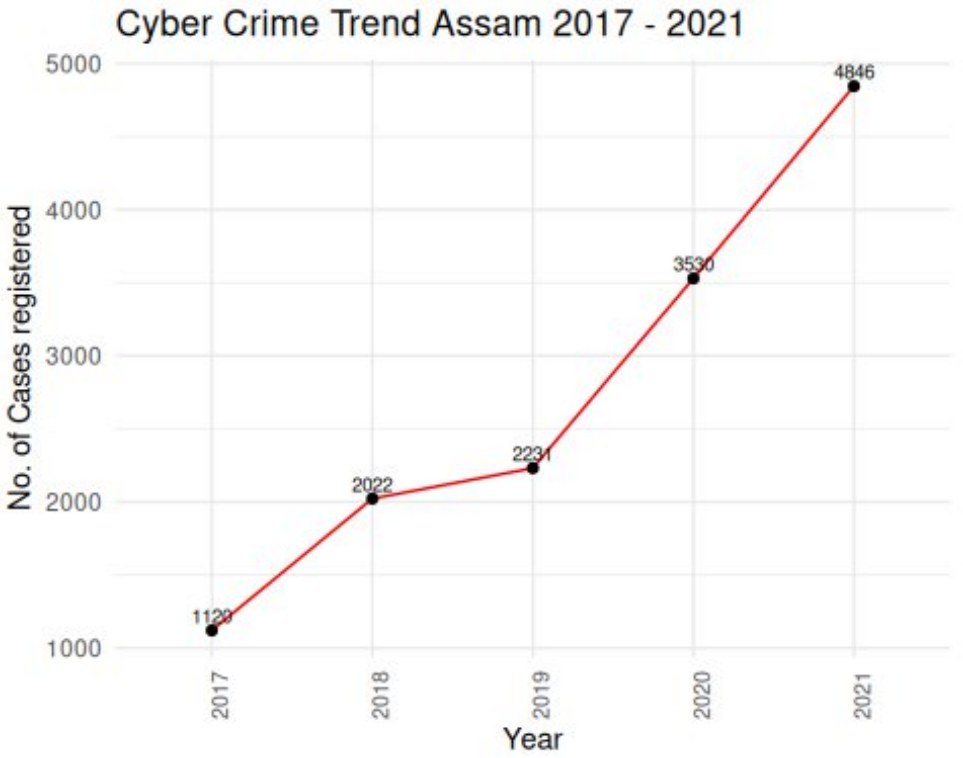

Cyber Crime Rates Trend Karnataka 2017 - 2021
No. of Cases registered
16
12
8
18.22
16.15
12.15
8.92
5.03
2017
2018
2019
2020
2021
Year

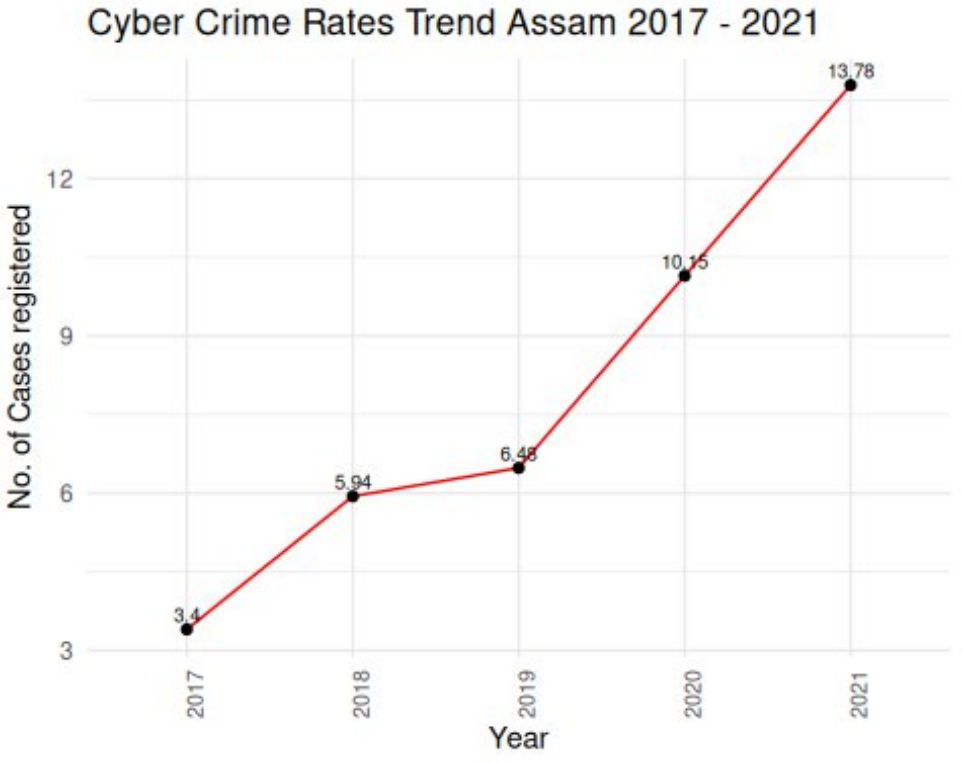
Maharashtra



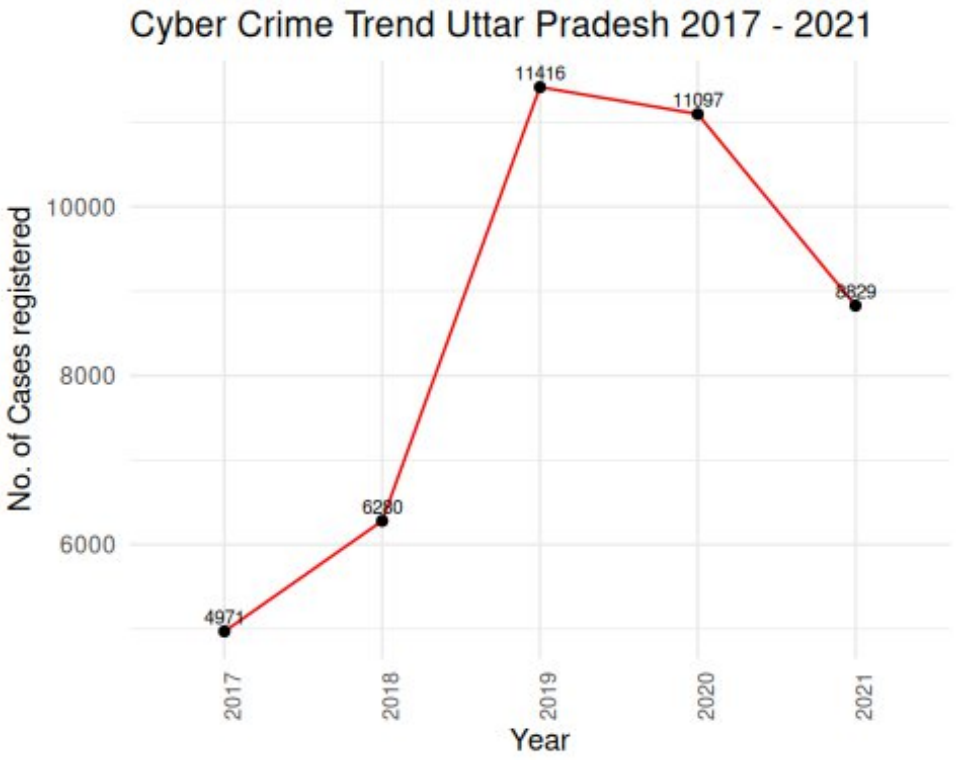


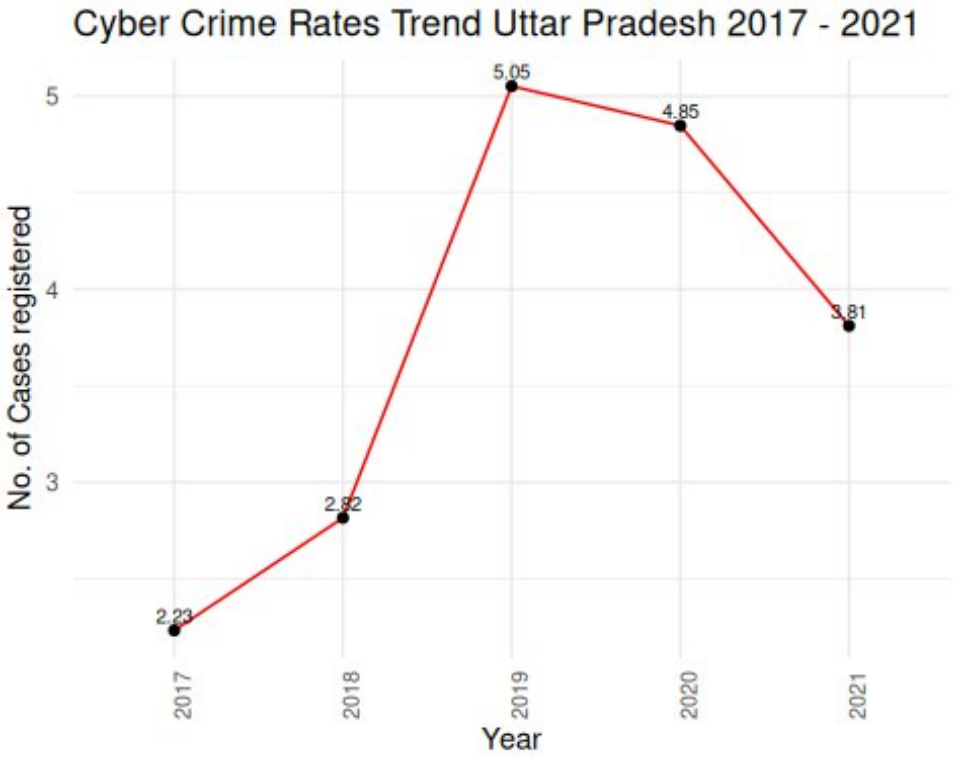
Assam


Cyber Crime Trend Assam 2017 - 2021
No. of Cases registered
5000
4000
3000
2000
1000
4846
3530
2231
2022
1129
2017
2018
2019
2020
2021
Year


Cyber Crime Rates Trend Assam 2017 - 2021
No. of Cases registered
12
9
6
3
13.78
10.15
6.48
5.94
3.4
2017
2018
2019
2020
2021
Year

Uttar Pradesh



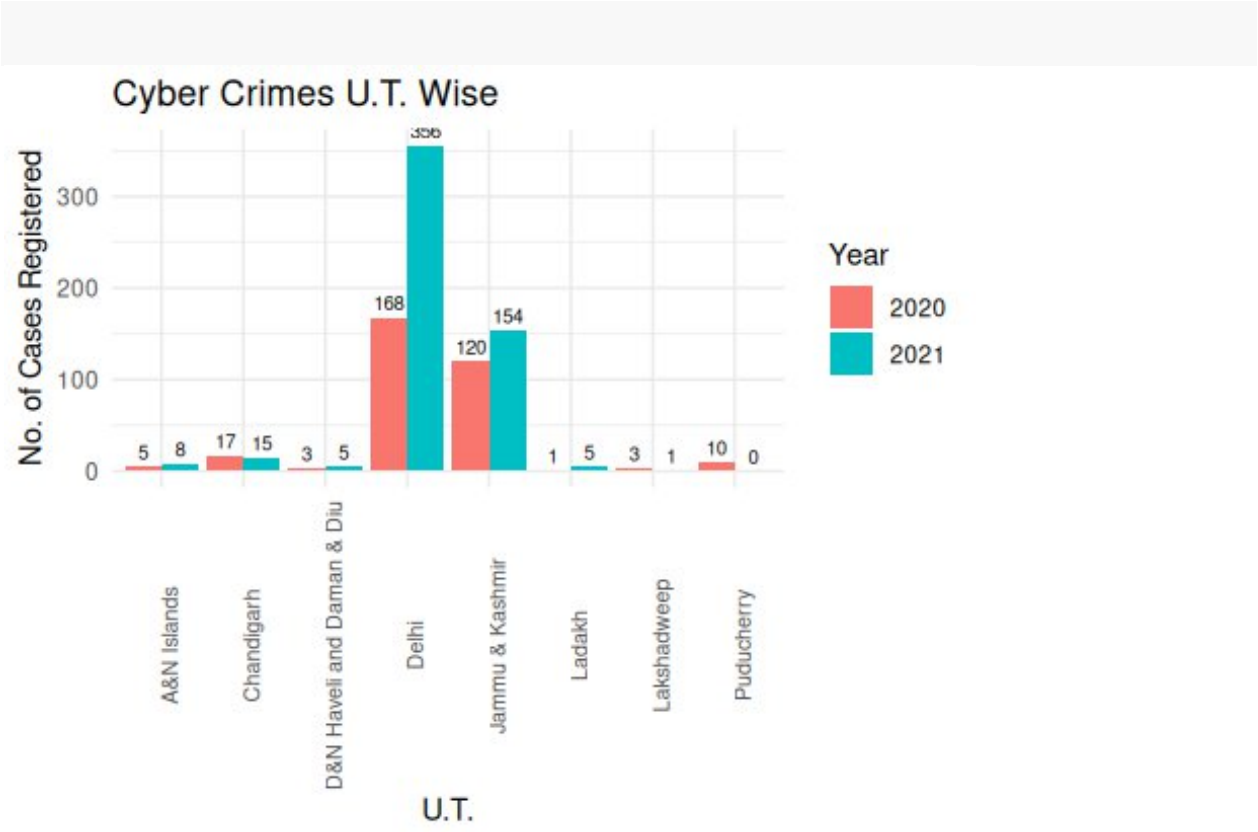


Now, it is evident from that the states of Assam and Telangana are showing rapid increasing trend in Cyber Crimes over the last 5 years. However, Uttar Pradesh and Karnataka are registering some lesser number of cases as compared to previous years. The

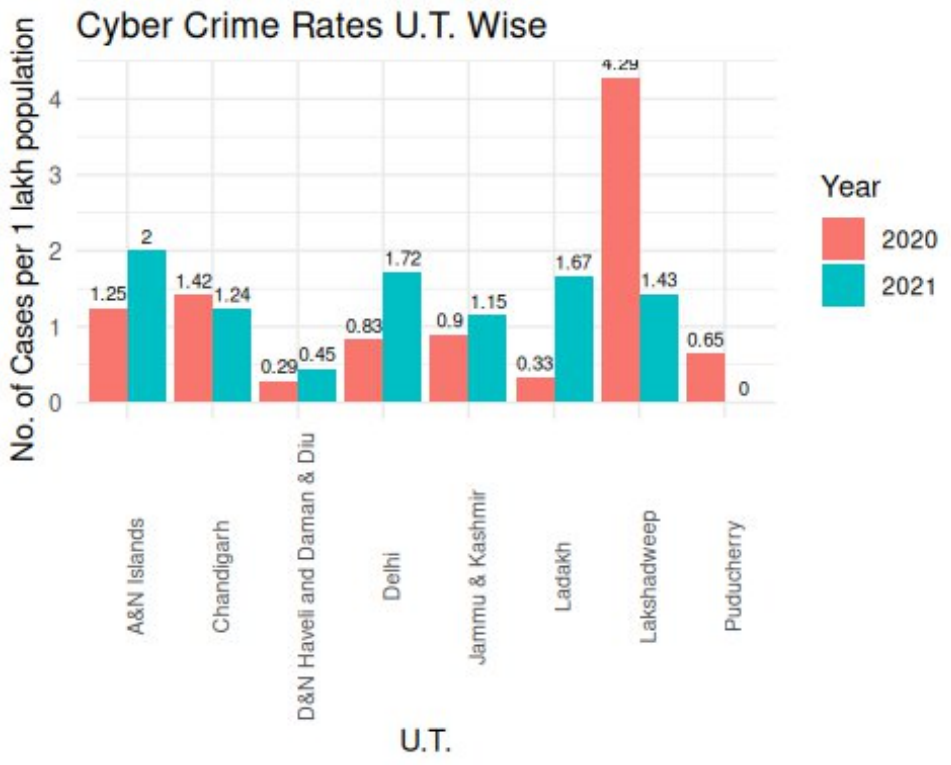
state of Maharashtra is slowly registering more number of cases as compared to previous years.

3.2 Cyber Crimes U.T. - wise

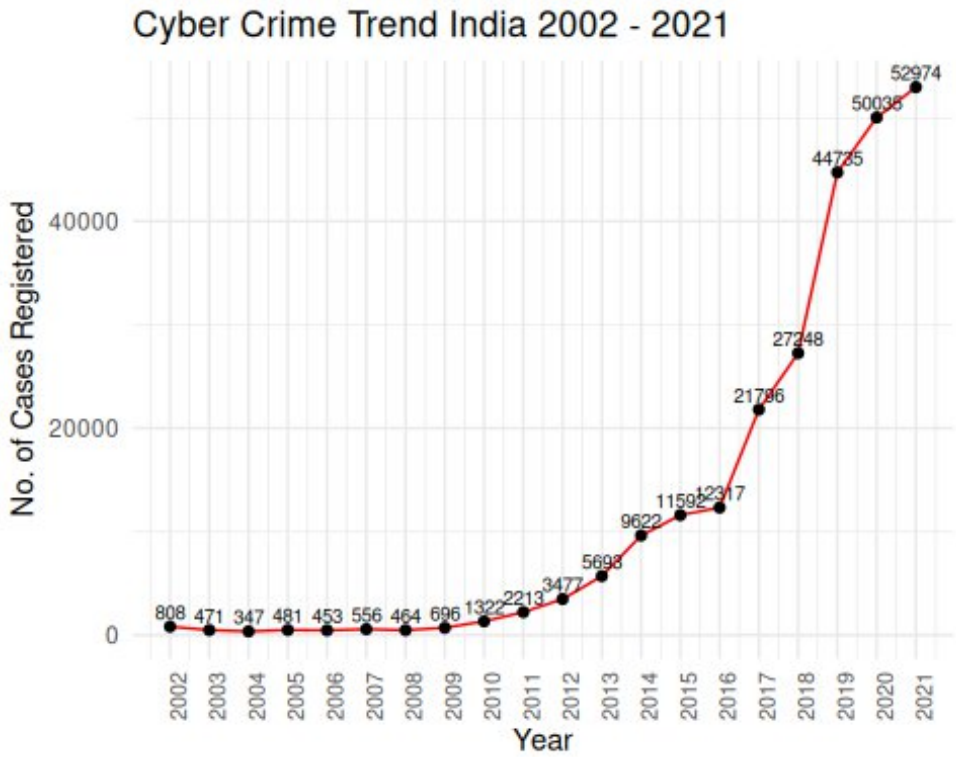
Further, Similar things can be analyzed for the Union Territories also.



In terms of numbers, the Union Territory of Delhi registered the most 356 cyber crime cases in 2021, followed by Jammu&Kashmir and Chandigarh with 154 ans 15 cases respectively. Also, **the cases in Delhi are more than double in 2021 from 2020.** However, in terms of rates, the picture is somewhat like this in all the Union Territories:



3.3 Cyber Crime Trend India



Now, the above plot shows the Trend for the number of Total Cyber Crimes registered in overall India from the year 2002 - 2021.

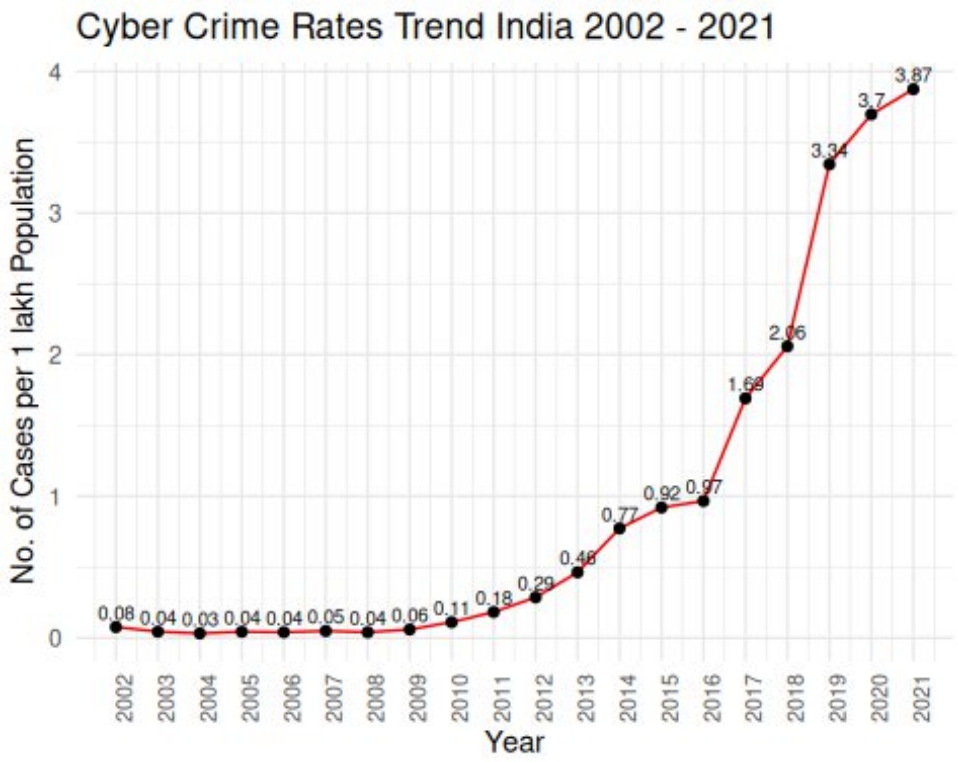
Now, it is evident from the plot, that India registered more than **4 times** cyber crimes in 2021 as compared to 2016.

Also, in a span of an year, India registered **1.5 times** more Cyber Crime cases in 2018 from 2017. And if we see 2018 from 2016, then the registration of the cases is **just above double**.

Also, one may note that initially, from the years 2002 - 2009, the cases registered were much much lower due to the fact that the internet was not that cheaply and easily accessible for the. people of India.
But from the year 2016 to 2017, the cyber crimes were **almost doubled** due to the fact that in this year the telecom sector in India experienced drastic changes with the entry of Jio in the Telecom Market. To tackle with their bussiness strategies, other telecom networks reduced the data charges to a huge extent, leading to a very large increase in the Internet accessibility among the Indian Population.

Also, it is a fact that the population of India also varied in these years. For that matter, we can observe the Cyber Crime Rgistration Rate of India from 2002 - 2021 as follows:

```
cyber_trendr
```



The cyber crime rate prediction for the coming years can be done using the following fitted model:

$$\log(Y_t) = -4.26073 + 0.28345\ t$$

or

$$Y_t = \exp(-4.26073 + 0.28345\ t)$$

where:
$t \geq 1$
Y_t : Predicted Cyber Crime Rate in India for the year $(2001 + t)$

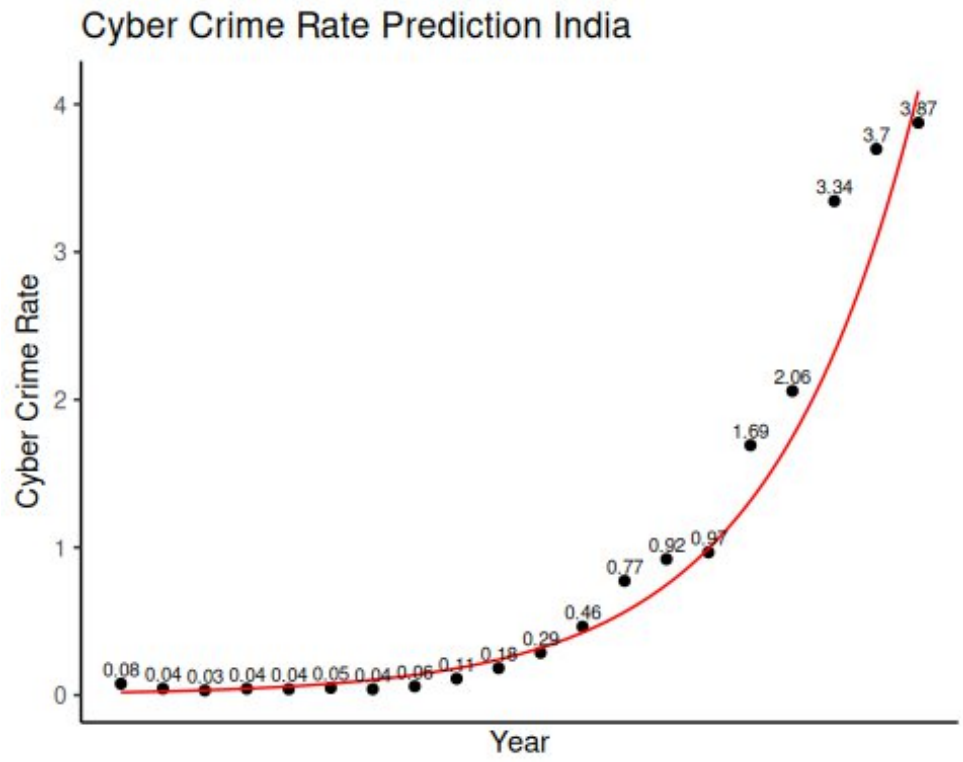
```
summary(model4)
```

```
##
## Call:
## lm(formula = log(rate) ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.93734 -0.29522 -0.01532  0.22008  1.41202
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.26073    0.24395  -17.47 9.85e-13 ***
## t            0.28345    0.02036   13.92 4.47e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5251 on 18 degrees of freedom
## Multiple R-squared:  0.915,  Adjusted R-squared:  0.9103
## F-statistic: 193.7 on 1 and 18 DF,  p-value: 4.475e-11
```

Clearly, from the summary of the model, it is evident that the values of R^2 and Adjusted R^2 do not differ much. Here t denotes the time index. t = 2 corresponds to the year 2002, upto so on t = 20 for the year 2021. The parameter estimates are coming out to be highly significant.

The plot of the fitted curve along with the actual values as dots is shown below:

```
fitplot
```



The Cyber Crime Rates predicted for the coming years using the above model are:

```
rate.prediction
```

```
##    year predicted_rate
## 1  2022          5.429
## 2  2023          7.209
## 3  2024          9.571
## 4  2025         12.707
## 5  2026         16.871
## 6  2027         22.400
## 7  2028         29.741
## 8  2029         39.488
## 9  2030         52.428
## 10 2031         69.609
```

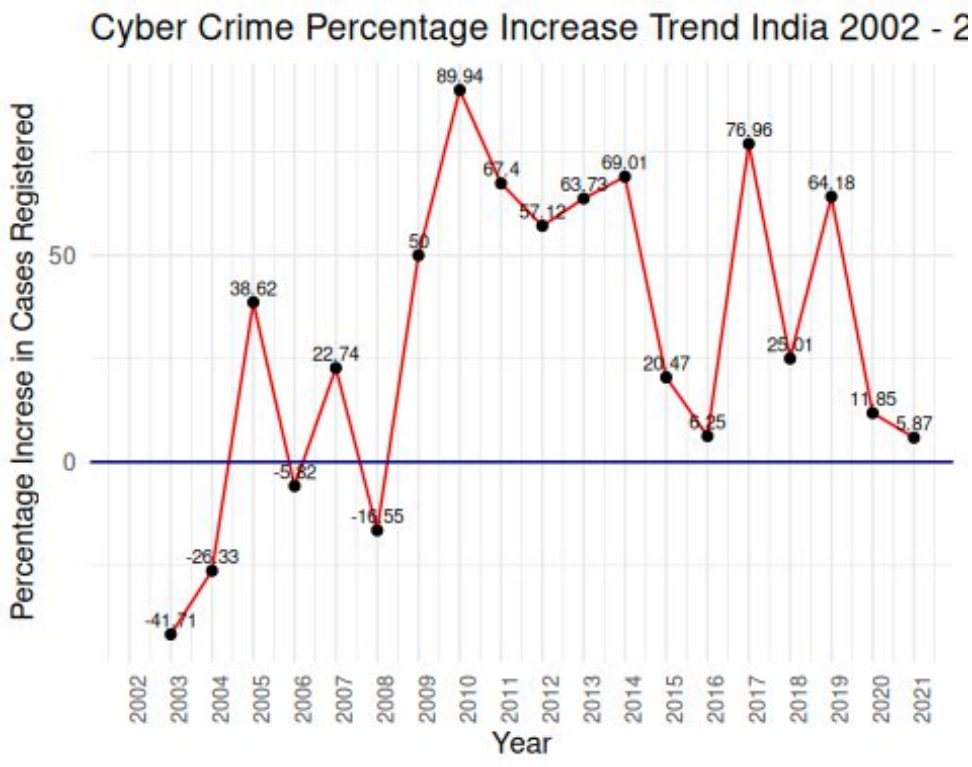
To get a more clear idea about the previous year comparison, percentage increase in the cases can be calculated using the formula:

$$Percentage\ Increase = \frac{Current\ Value - Previous\ Value}{Previous\ value} * 100$$

From the Percentage Increase plot, following facts can be noted:

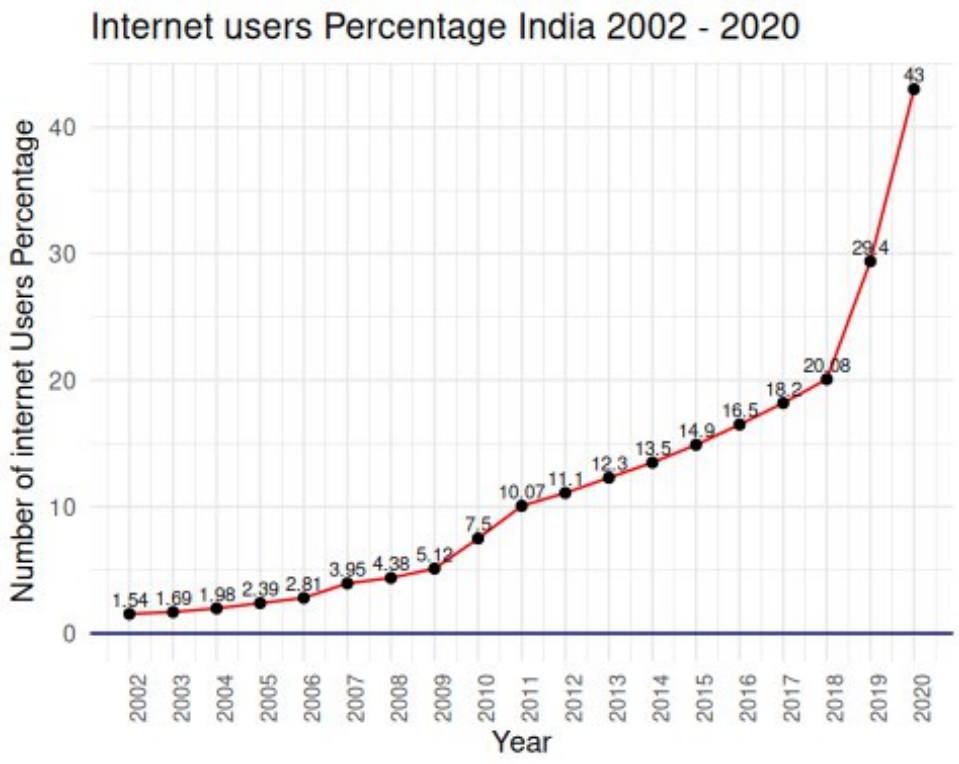
- India registered 5.87% more cyber crimes in 2020 as compared to 2021.
- It must be noted that the any dip in the graph does not mean decrease in the crimes, rather it means that there is less percentage increase in the cases as compared to previous year. However any point below zero, i.e. negative percentage increase indicates that there are lesser number of cases from the previous year.

```
cyber_trendp
```



Internet user-base of India

Besides these cyber crime incidents and trend analysis, one should also have a look on the active internet users in India over the years. Data of active internet users in the country is taken from the world bank and visualized as follows:



4 Cyber Crime Motives

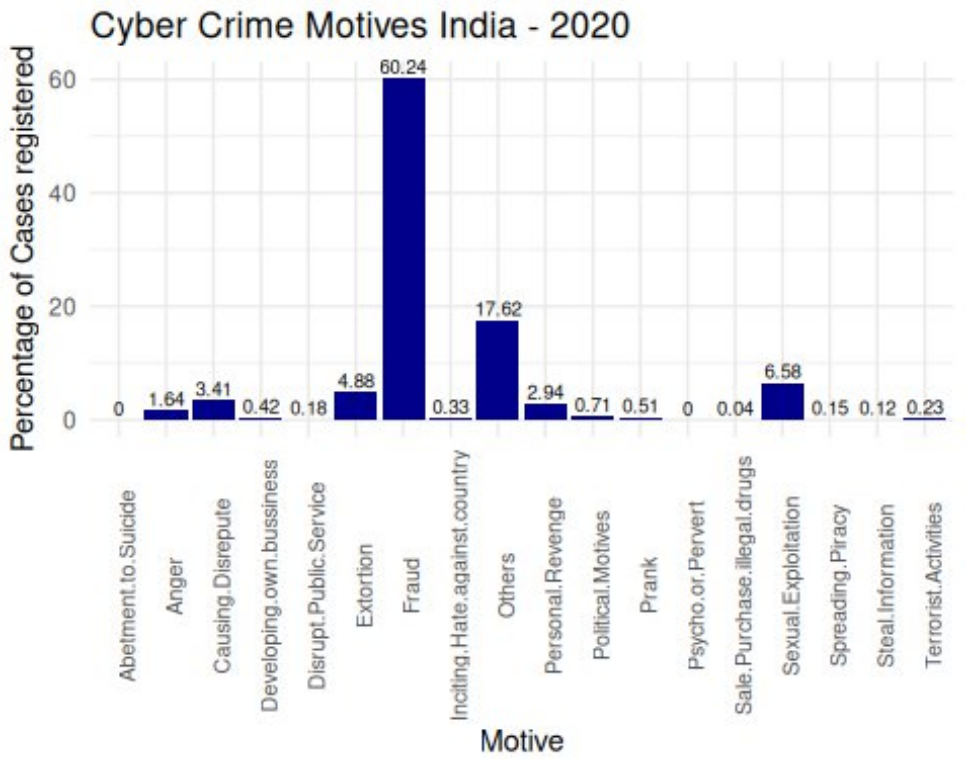
For the various cyber crimes happening in India, the data for the motives behind them is available, which can be observed as follows:

4.1 Cyber Crime Motives Year Wise

4.1.1 Year 2020

From the following diagram, it is evident that, in 2020, the major motive behind the cyber crimes in India was Fraud, contributing to 60% of the total cyber crimes in India.

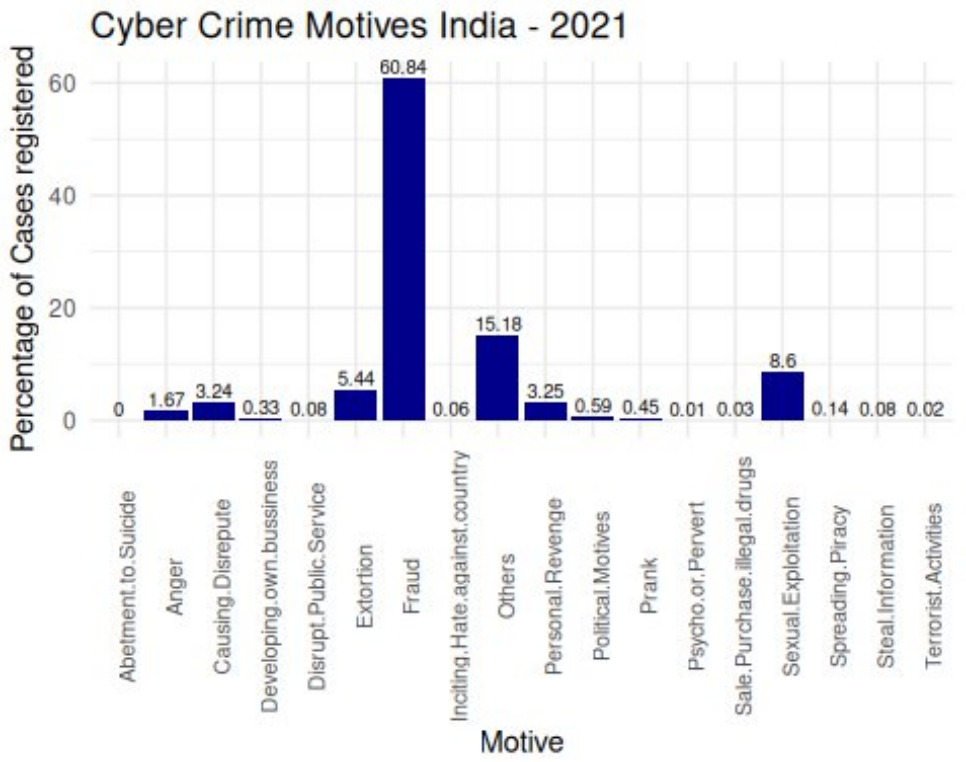
```
source("cybermotives.R")
motives2020
```



4.1.2 Year 2021

Again in the next consecutive year 2021, it can be seen that Fraud emerges as the major cyber crime motive in India, contributing to 60.24% of the total cyber crimes in the country.

```
motives2021
```



4.2 Cluster Analysis

For the fraud cases, different states of India are divided into different clusters based on the fraction of fraud in the total cyber crimes in that states for the years 2017 to 2021.

States are divided into 4 clusters, K-Means Clustering is implemented as Non-Hierarchical cluster analysis and Agglomerative & Divisive Techniques are also implemented under Hierarchical cluster anlaysis.

4.2.1 Distance Matrix Calculations

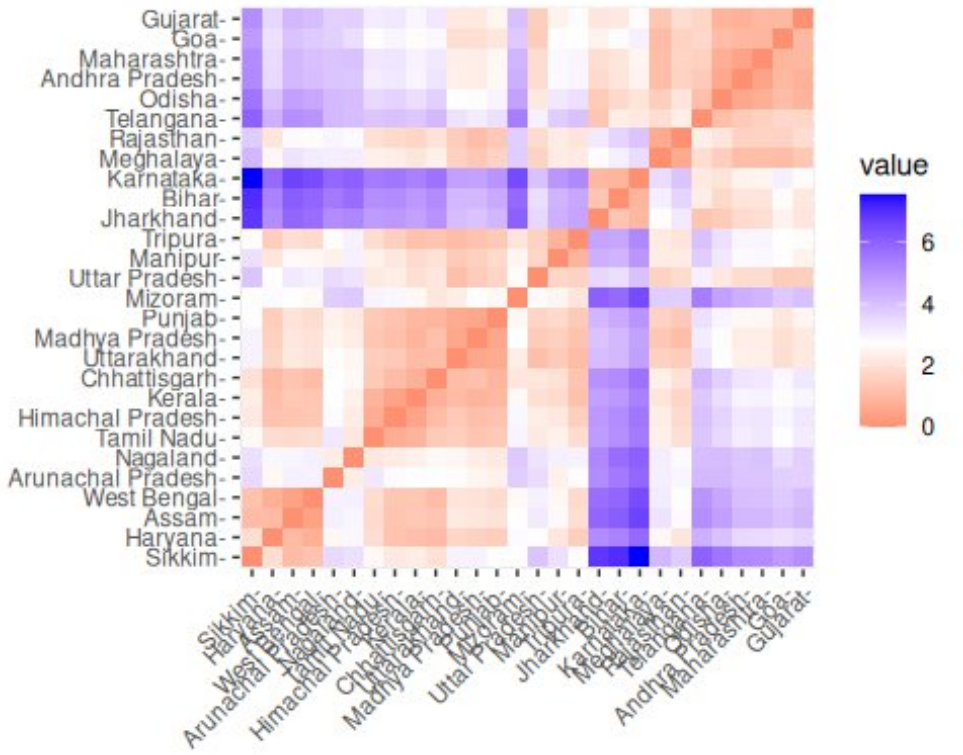
Data has been scaled to centre 0, and unit variability of each year rate (Variables).

```
dist[1:3, 1:3]
```

```
##                   Andhra Pradesh Arunachal Pradesh    Assam
## Andhra Pradesh          0.000000          3.937371 4.223798
## Arunachal Pradesh       3.937371          0.000000 3.029010
## Assam                   4.223798          3.029010 0.000000
```

These are some part of the 28x28 distance matrix of states. This distance matrix can be easily visualized in the following manner:

```
displot
```



The above plot shows the states with most similarity with dark reddish shade to less similarity with light red shades upto dark blue shades with least similar.

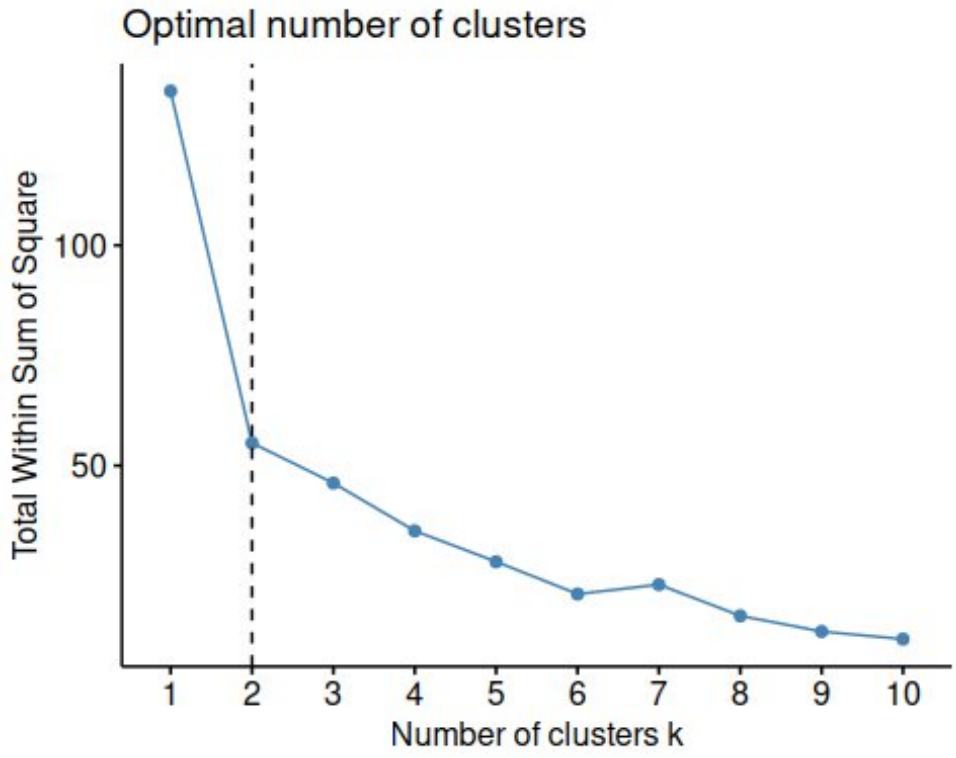
4.2.2 K-Means Clustering (Non-Hierarchical)

To determine the optimum number of clusters, consider the following plot of number of clusters v/s Total within sum of squares.

Within Cluster Sum of Squares must be as minimum as possible. So accordingly the optimum number of clusters are to be chosen.

From the following plot the optimum number of clusters is chosen as that value of k, where the value of Within Cluster Sum of Squares starts flattening down.

```
wssplot
```



According to the above plot, the Total within sum of squares starts flattening after k = 2.

We now proceed with **K-Means clustering** with 2 clusters.

```
km.res$cluster
```

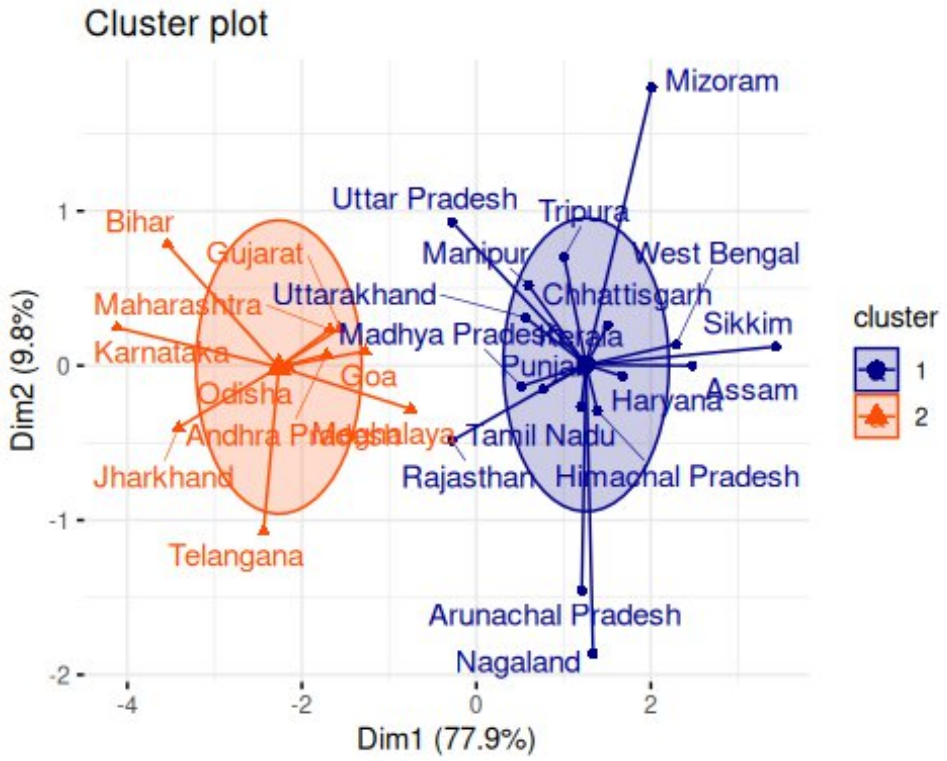
```
##    Andhra Pradesh Arunachal Pradesh             Assam             Bihar
##                 2                 1                 1                 2
##      Chhattisgarh               Goa           Gujarat           Haryana
##                 1                 2                 2                 1
##  Himachal Pradesh         Jharkhand         Karnataka            Kerala
##                 1                 2                 2                 1
##    Madhya Pradesh       Maharashtra           Manipur         Meghalaya
##                 1                 2                 1                 2
##           Mizoram          Nagaland            Odisha            Punjab
##                 1                 1                 2                 1
##         Rajasthan            Sikkim        Tamil Nadu         Telangana
##                 1                 1                 1                 2
##           Tripura     Uttar Pradesh       Uttarakhand       West Bengal
##                 1                 1                 1                 1
```

This is the required clustering of 28 states into 2 clusters.

To have a better picture of the clusters, a look on the cluster plot will be helpful:

4.2.2.1 Cluster Plot

```
clusplot
```



1. As it is evident from the above cluster plot that the states with fewer fraud Cyber Crimes : Arunachal Pradesh and Nagaland are in one cluster.

2. However the states with extensive fraud rates: Karnataka, Bihar,Telangana, Jharkhand,Uttar Pradesh, Maharashtra etc. are in other cluster.

4.2.3 Hierarchical Clustering

4.2.3.1 Agglomerative Clustering

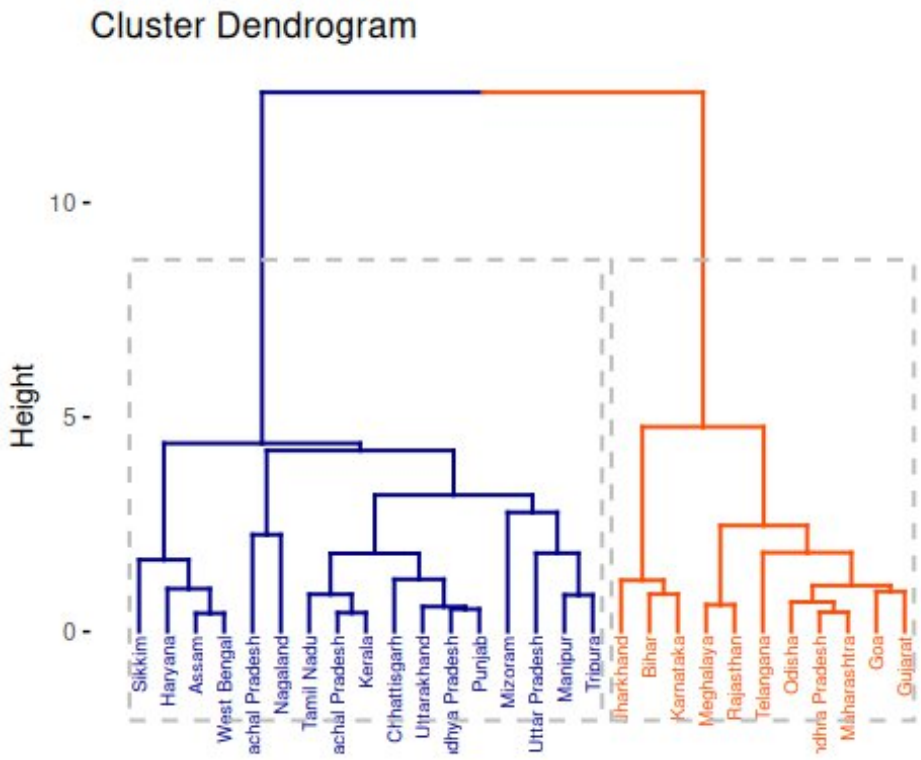
Also, Under the Hierarchical Clustering, using **Agglomerative Custering** with *Ward D2 Linkage*, the states are more likely to be divided into 2 clusters.

The states with major fraction of the Fraud cases are in one Cluster: Telangana, Jharkhand, Bihar etc.

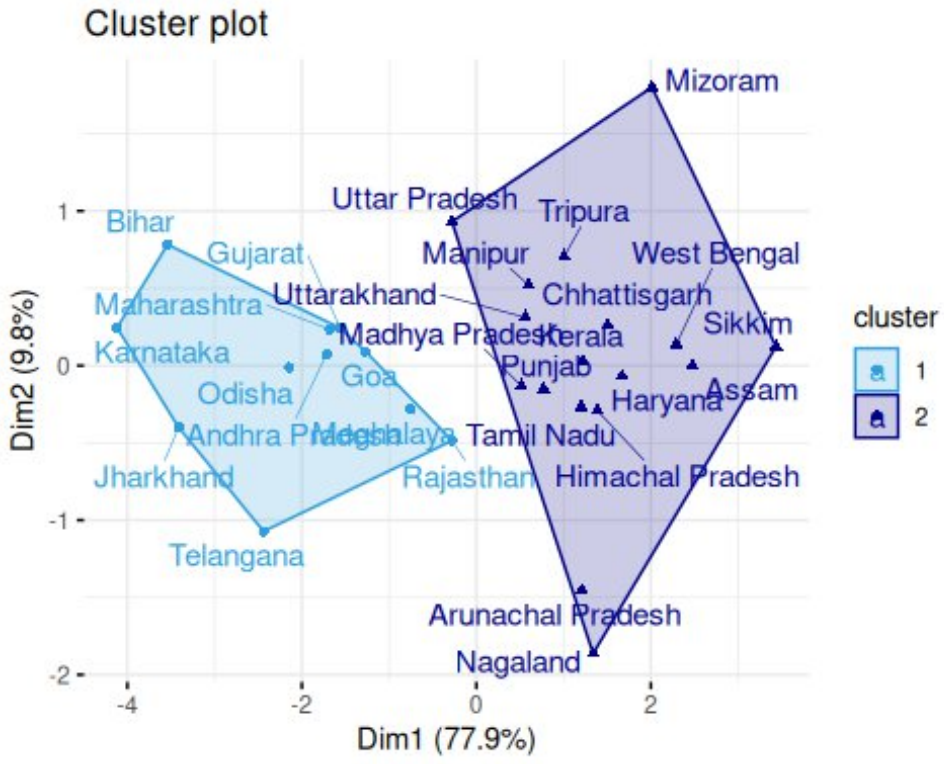
However, the states with comparatively lesser number of frauds are in the other cluster.

The cluster means are given by

```
cameans1 # 1st cluster mean

##  rate2021  rate2020  rate2019  rate2018  rate2017
## 0.5713483 0.6334256 0.5143662 0.4825693 0.4667825

cameans2 # 2nd Cluster mean

##  rate2021  rate2020  rate2019  rate2018  rate2017
## 0.2367204 0.2255833 0.1450668 0.2455529 0.2827396

aglodend
```




```
aglocp
```

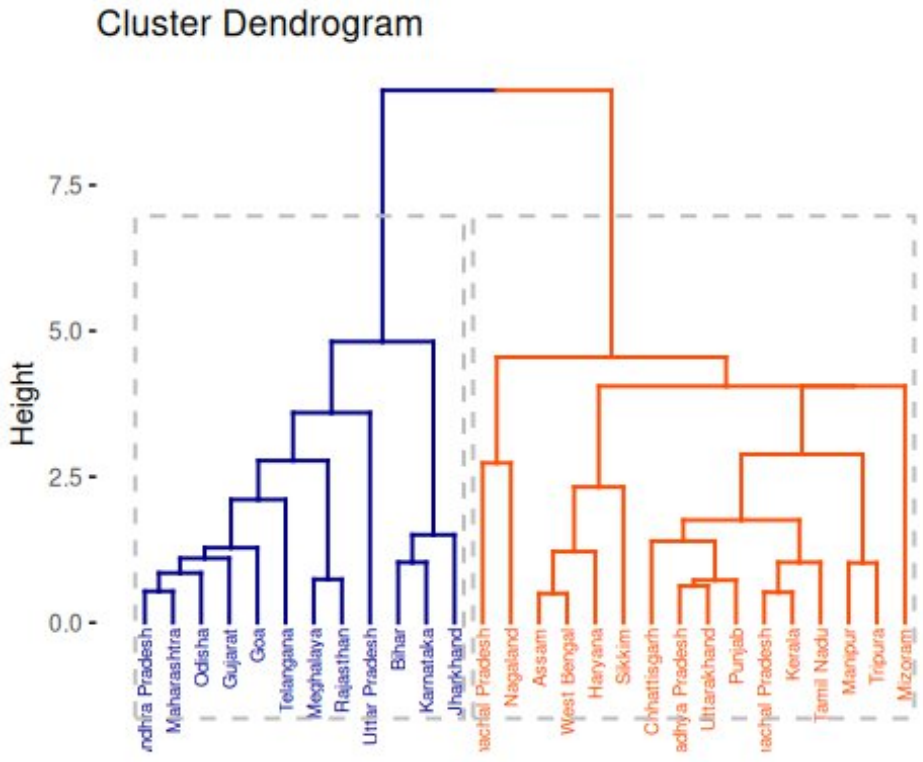


Previously, for better demarkation, we applied K-Means clusteing for 4 clusters. However, under the Hierarchical Clustering, using **Agglomerative Custering** with *Ward D2 Linkage*, the states are more likely to be divided into 2 clusters.

4.2.3.2 Divisive Clustering

On the other hand, if **Divisive Clustering** is applied, the two main clusters so formed are given as follows:

```
dianadend
```



4.3 References:

- Cyber Crimes data: National Crime Records Bureau (NCRB), Ministry of Home Affairs, Govt. of India.
- Individuals using the Internet (% of population) - India: International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database.
- M. Dasgupta, Cyber Crime in India: A Comparative Study.
- Apoorva Bhangla and Jahanvi Tuli, A Study of Cyber Crime and its Legal Framework in India.
- Anupreet Kaur Mokha, A Study on Awareness of Cyber Crimes and Cyber Security.
- Shivani Reddy, Analytical Study on Cyber Crimes in India.