Cyber Crimes in India

Project Report

Jaswinderpal Singh, Hitika, Jasmeen Sharma,
Kajol, Kanishka Jaggi, Kartik Setia
M.Sc. Statistics II Year - 2023

.

Under the Guidance of: Prof. Narinder Kumar
Department of Statistics, Panjab University, Chandigarh

June 30, 2023.

Contents

0.1 Introduction

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. 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, Non-Parametric tests are applied to test the similarity of distributions and randomness of cases, 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.

0.1.1 Cyber Crime

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.

Cyber crimes are a new class of crimes rapidly increasing due to extensive use of Internet and I.T. enabled services.

0.1.2 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.
10. *Spamming:* Sending unsolicited messages through email or other means.

Various studies are done as far as cyber-crimes in India are concerned. M. Dasgupta (2011) has classified the cyber-crimes on different basis and grounds like computer as a target as well as victim. Apoorva Bhangla and Jahanvi Tuli (2021) conducted the study on the legal framework of cyber-crimes in India.

Chapter 1

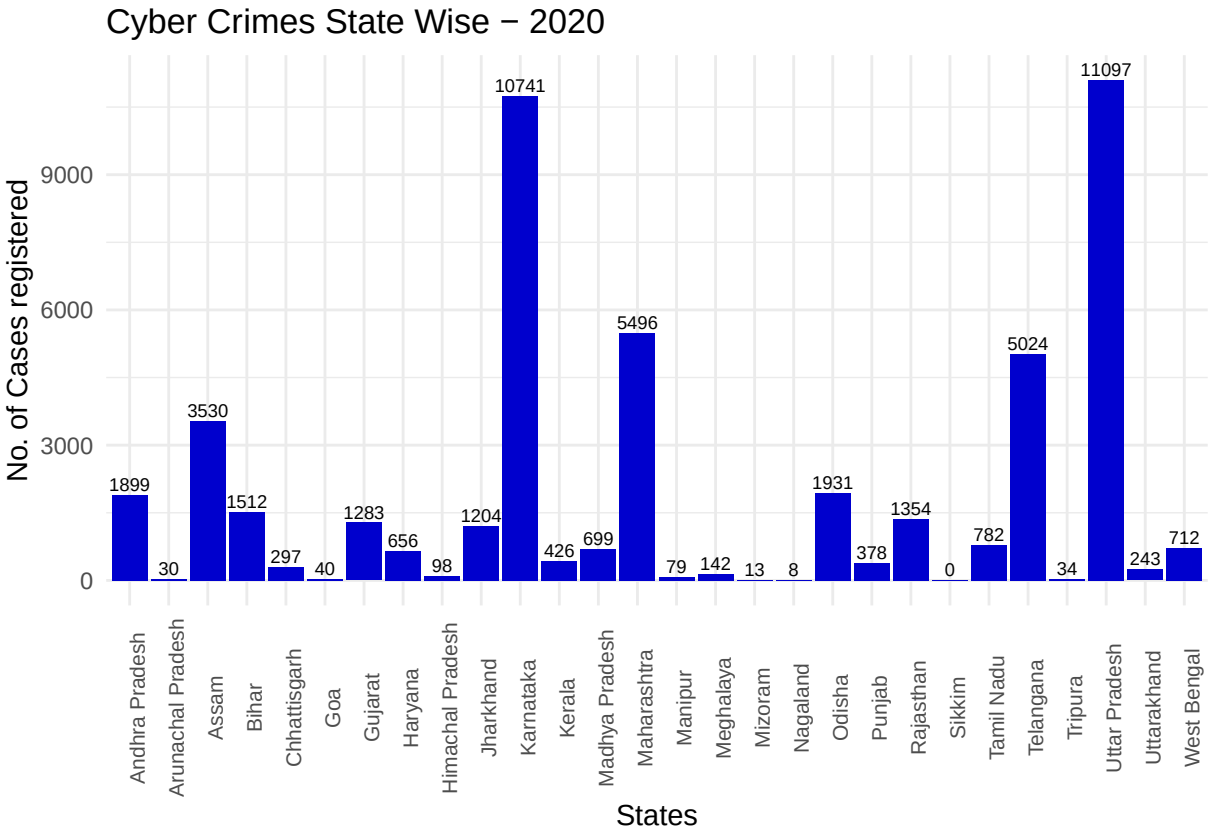
Cyber Crimes in India: Statistics and Data Visualization

1.1 Introduction

In this section, the main objective is to have a visualization of the datasets and to figure out which states and Union Territories have Major role in the Cyber Crimes in India in the past few years. Also, the Cyber Crime registration trend for the country is analyzed over the years and predictions are made for the coming years.

1.2 Cyber Crimes State-Wise:

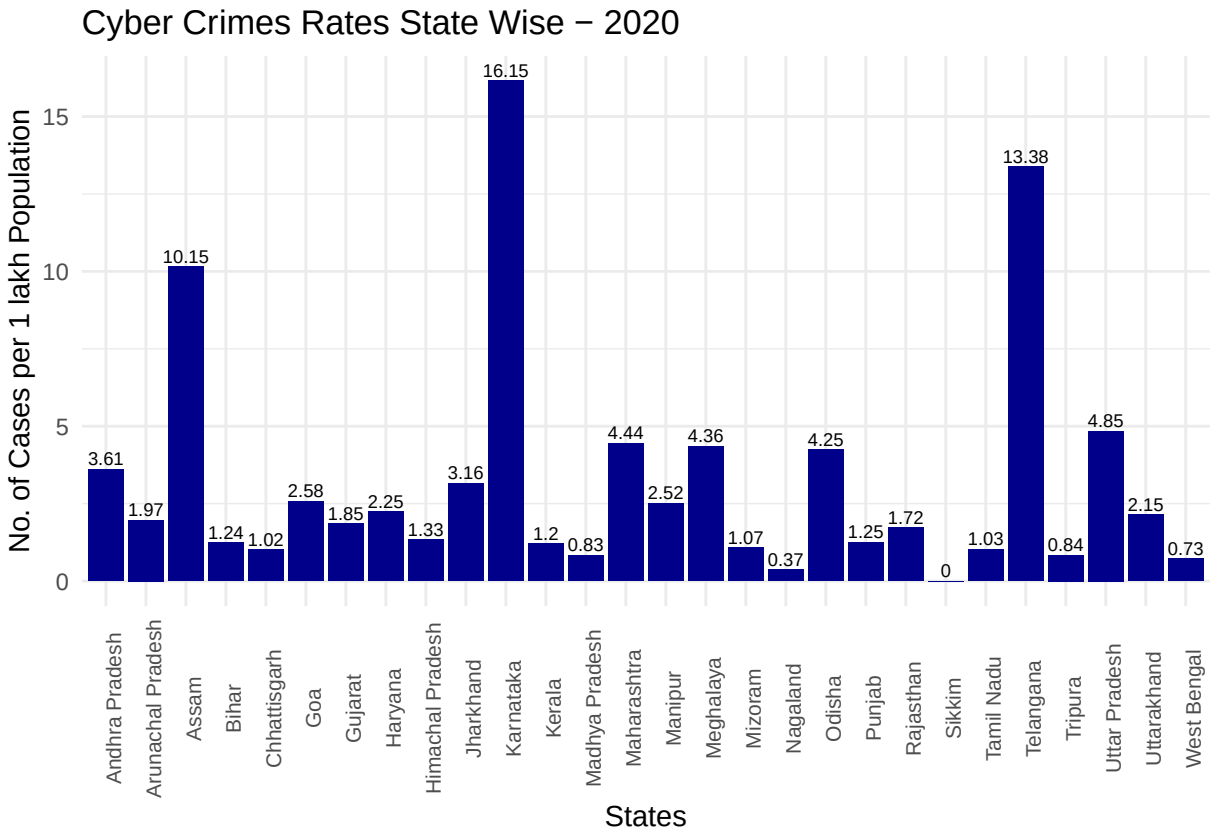
1.2.1 Year 2020



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 2020 is available. Hence the Cyber Crime Rate can be calculated as:

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



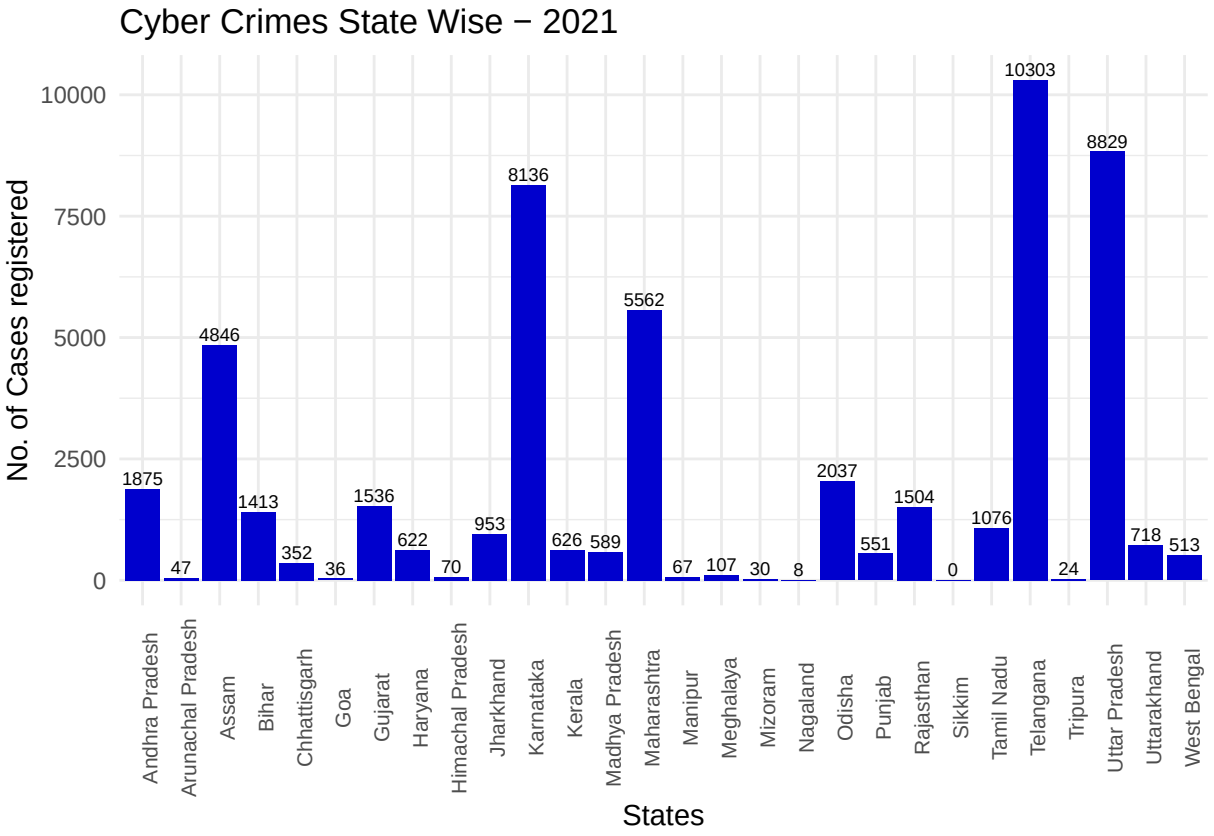
As far as the rates are concerned, Karnataka and Telangana emerged as the states with highest cyber crime rates followed by Assam, Uttar Pradesh and Maharashtra.

Now, keeping the population variation of the states into account, the better comparative picture here is given by 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.

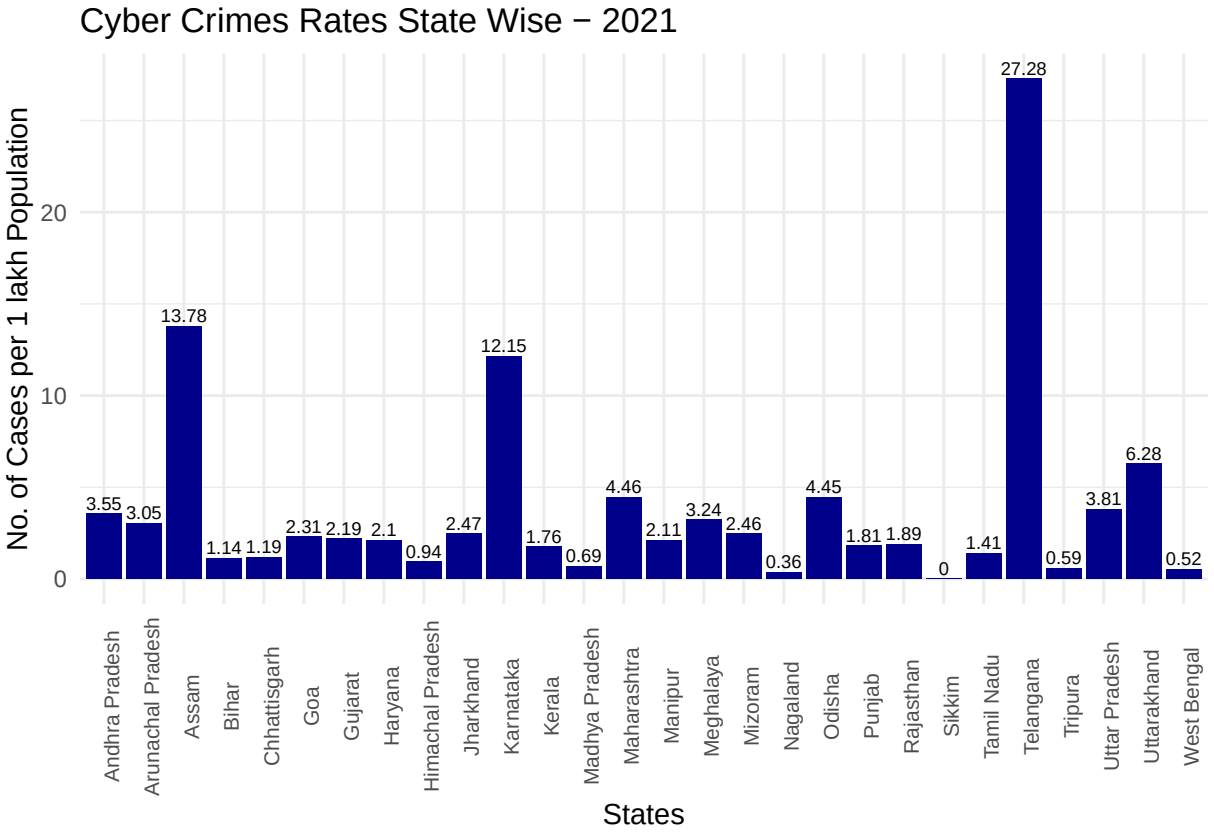
```
## [1] 0.7172579
```

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.

1.2.2 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.



But for better comparison, the Cyber Crime Rates, i.e. Cyber Crimes in a State per 1 lakh population, can be seen. 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.

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.

```
## [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

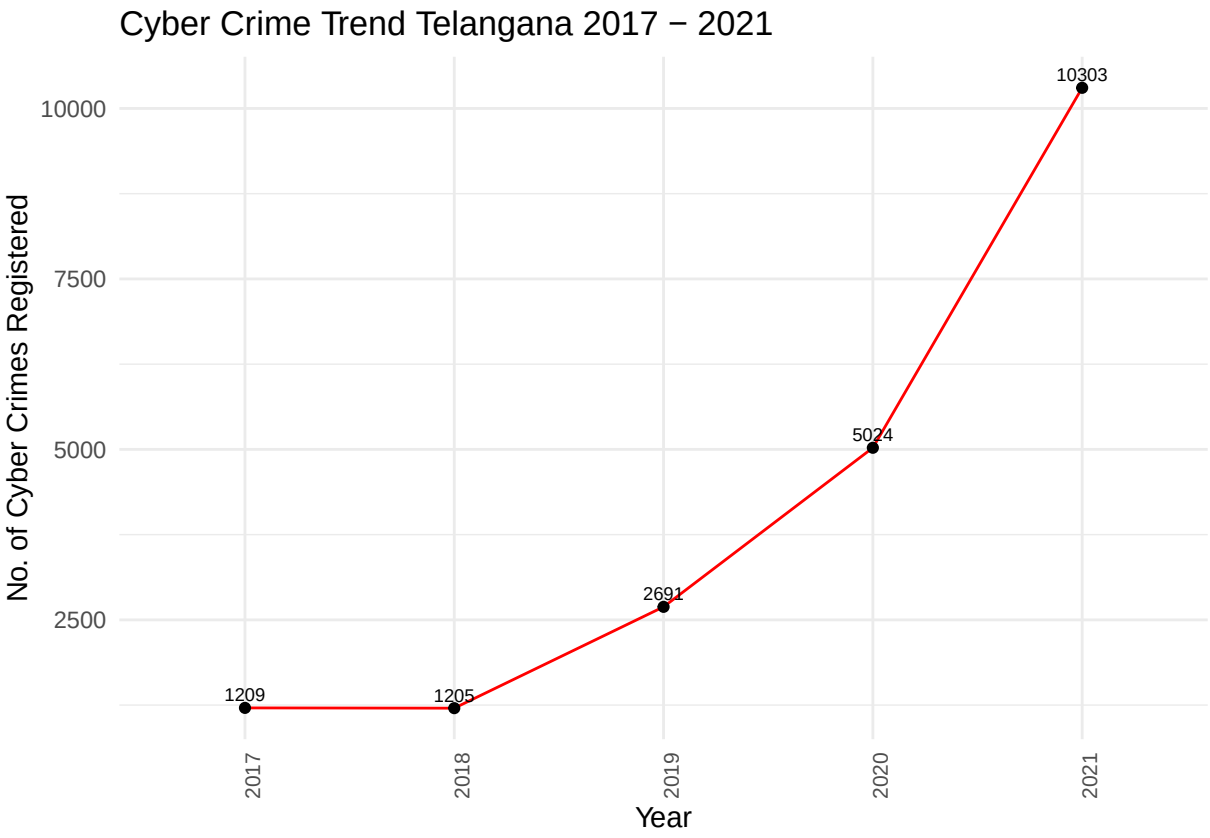
Now, for the past 5 years, the cyber crime trend in these 5 major states is to be viewed. The trend is shown for each state for the years 2017 - 2021. The data can be visualized both in terms of actual crime registered count as well as in terms of rates in these states.

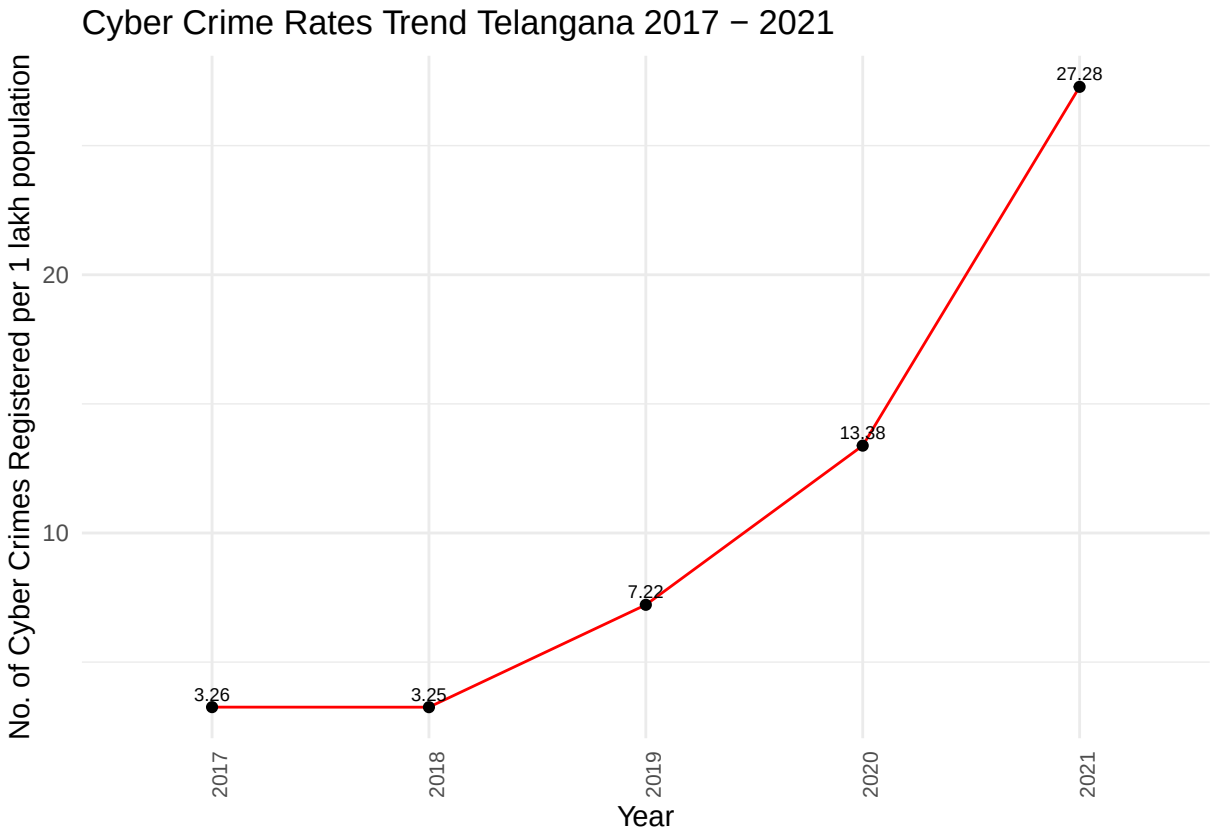
1.2.3 Cyber Crimes Trend: Major States

From the above analysis, it can be seen that the states of Telangana, Uttar Pradesh, Karnataka, Maharashtra and Assam have comparatively higher cyber crime rates as well as larger number of cyber crimes registered in the last two years. Now, it is of interest to analyze these states based on their records of the previous years.

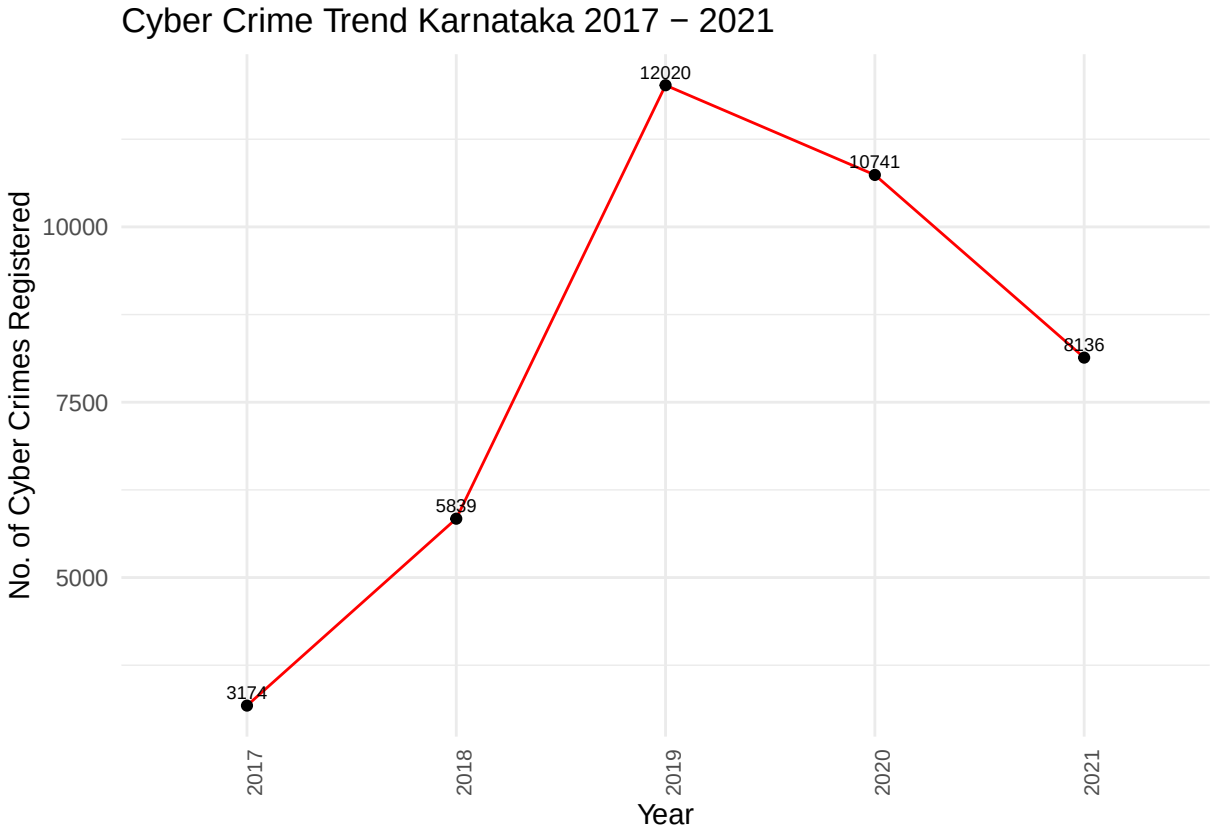
These States can be analyzed further based on last 5 years as follows:

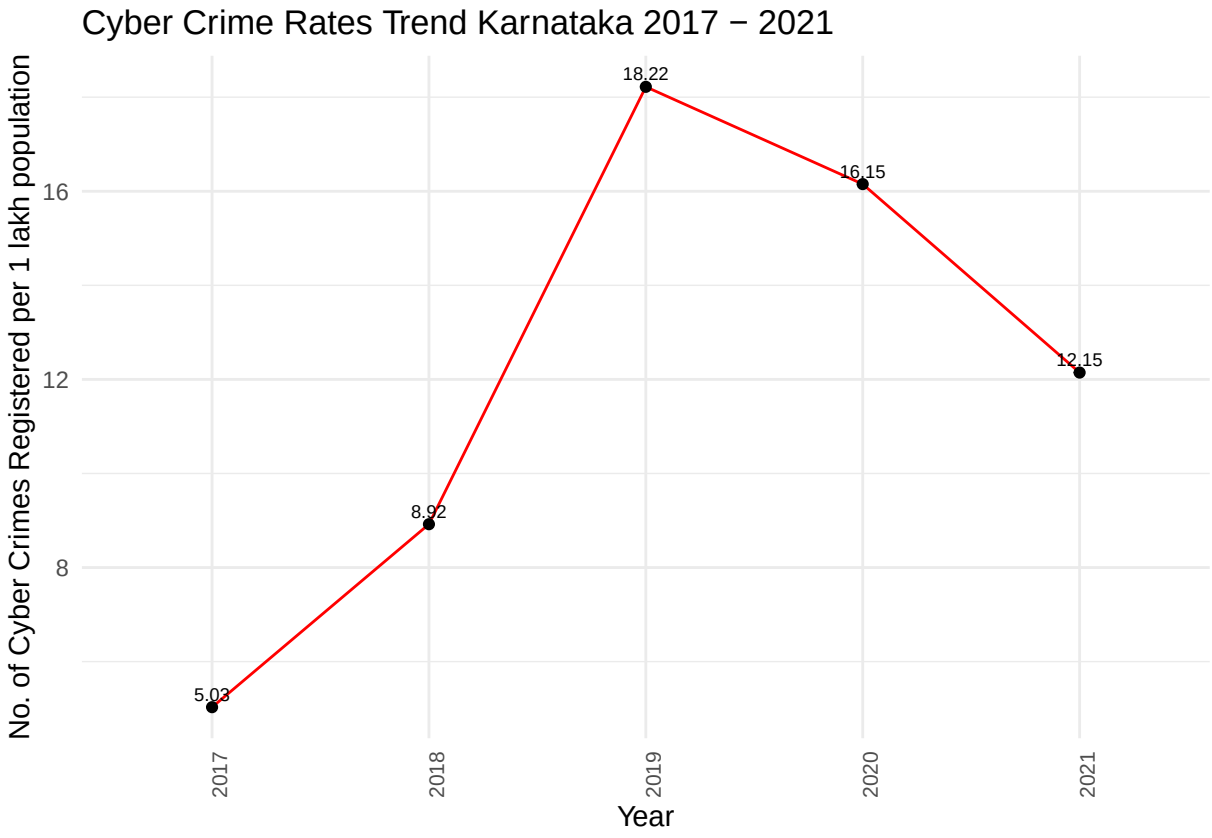
```
##
##  Telangana
```



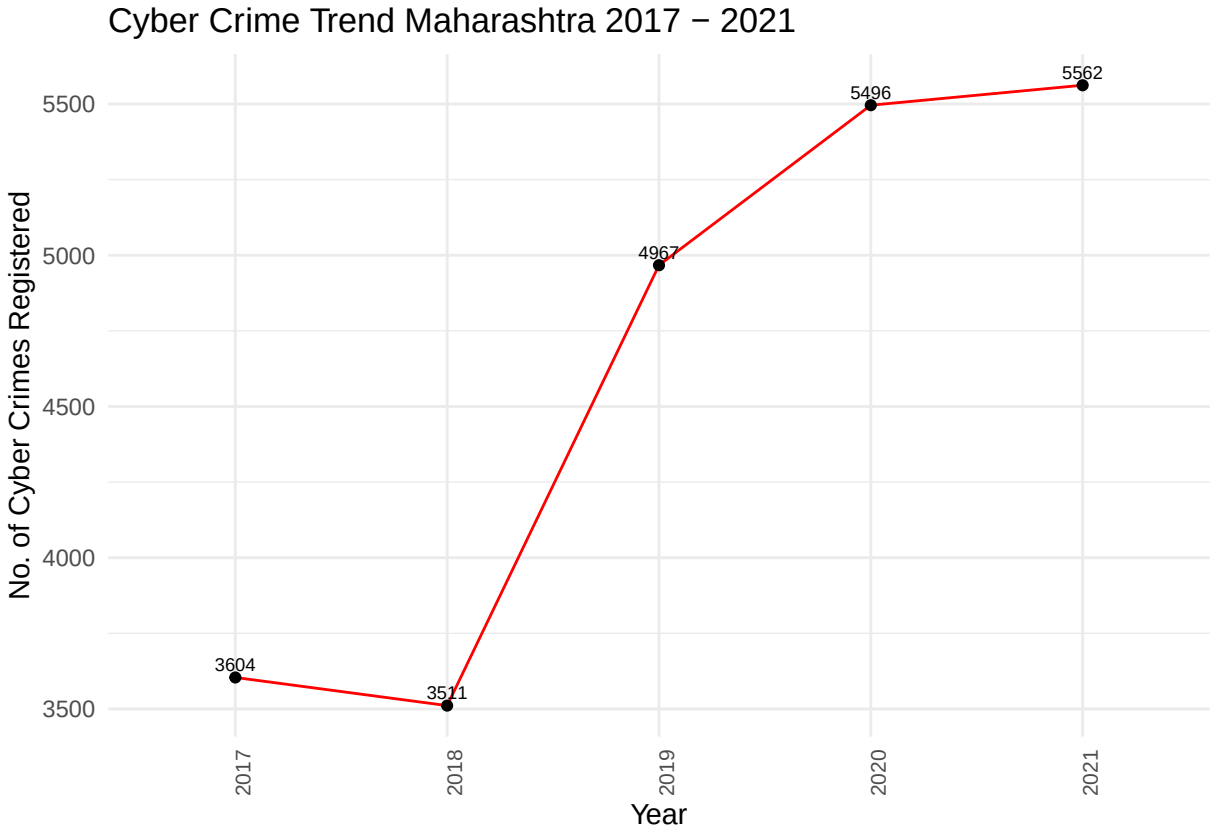

Cyber Crime Rates Trend Telangana 2017 − 2021
No. of Cyber Crimes Registered per 1 lakh population
20
10
27.28
13.38
7.22
3.26
3.25
2017
2018
2019
2020
2021
Year

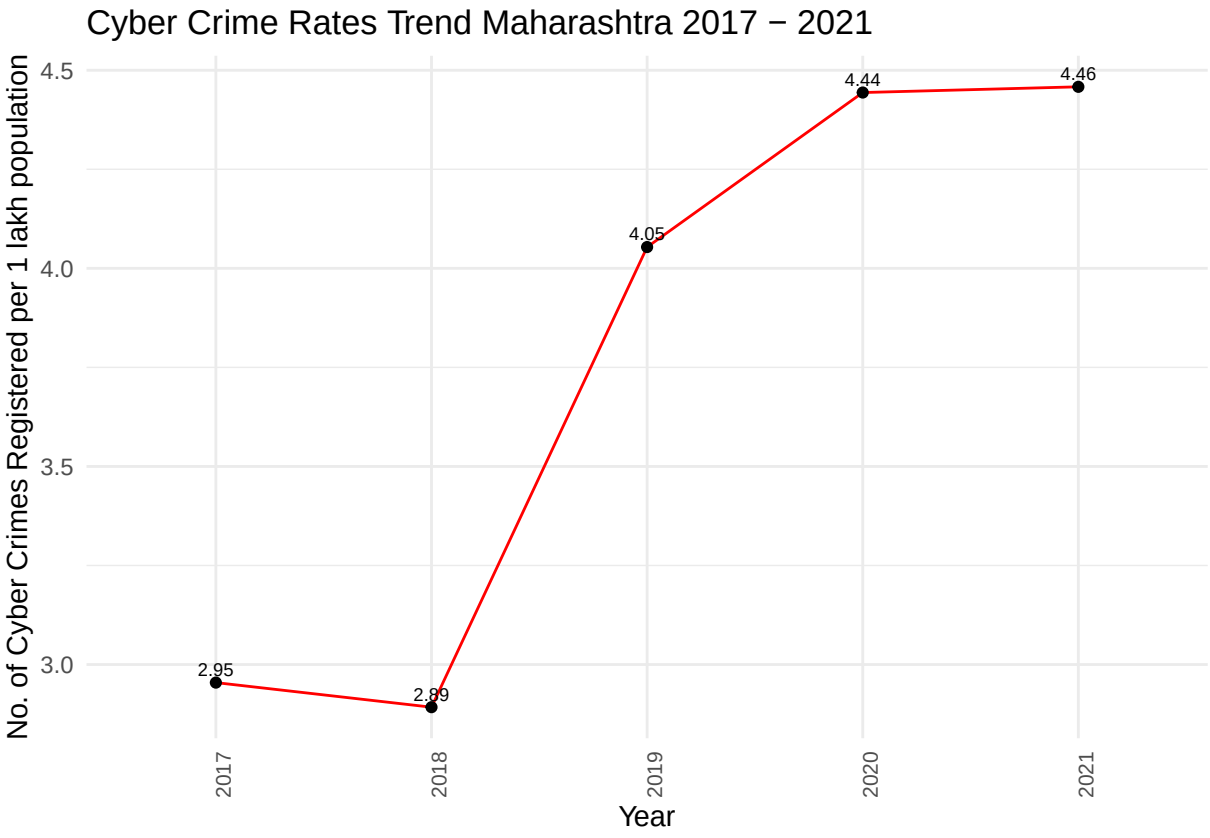
```
##
##
##  Karnataka
```


Cyber Crime Trend Karnataka 2017 − 2021
No. of Cyber Crimes Registered
10000
7500
5000
12020
10741
8136
5839
3174
2017
2018
2019
2020
2021
Year

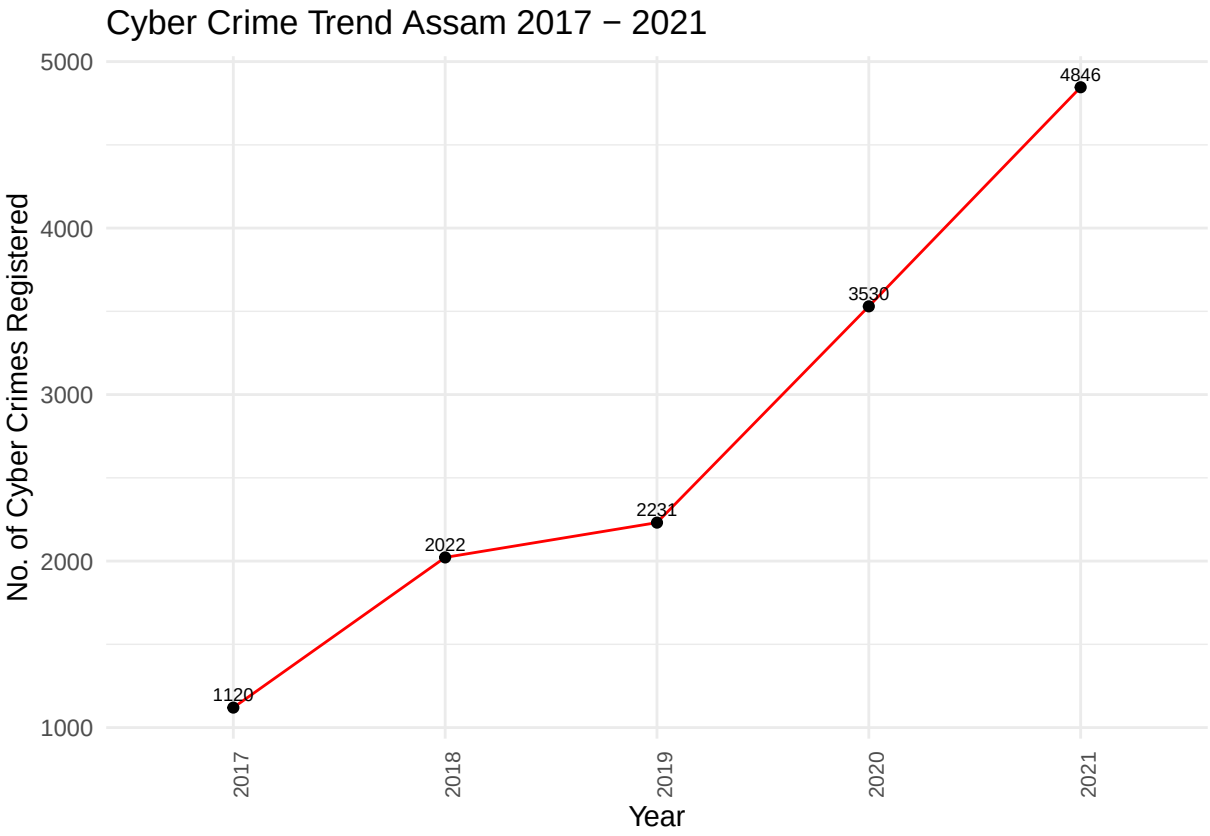


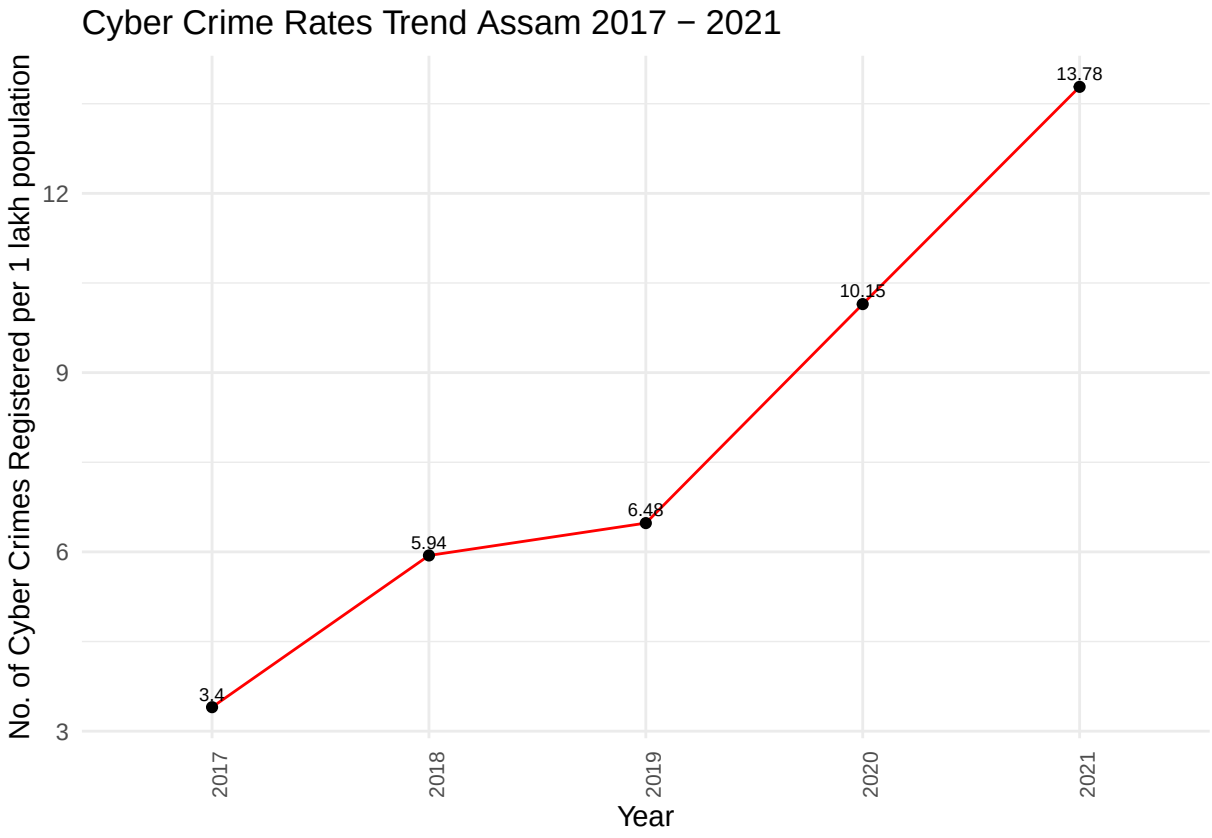

```
##
##
##  Maharashtra
```



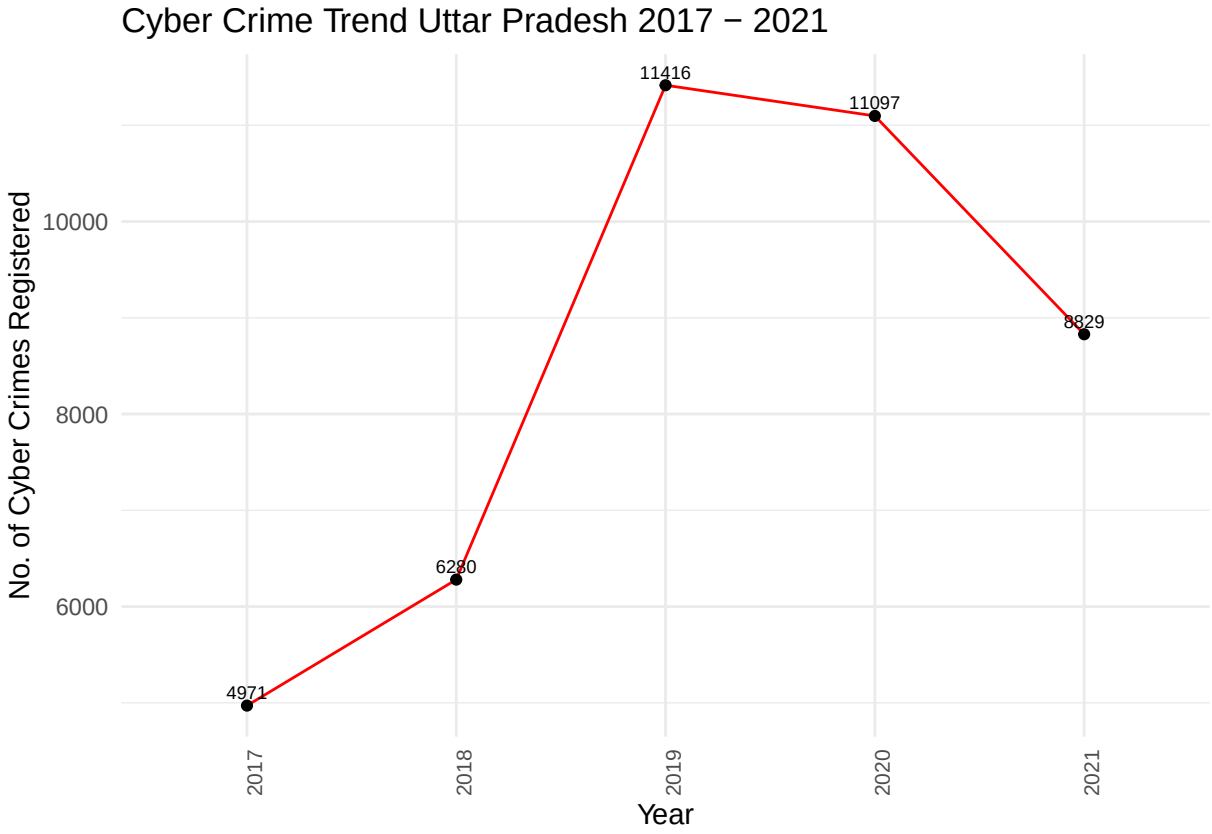

Cyber Crime Rates Trend Maharashtra 2017 − 2021
No. of Cyber Crimes Registered per 1 lakh population
4.5
4.0
3.5
3.0
2.95
2.89
4.05
4.44
4.46
2017
2018
2019
2020
2021
Year

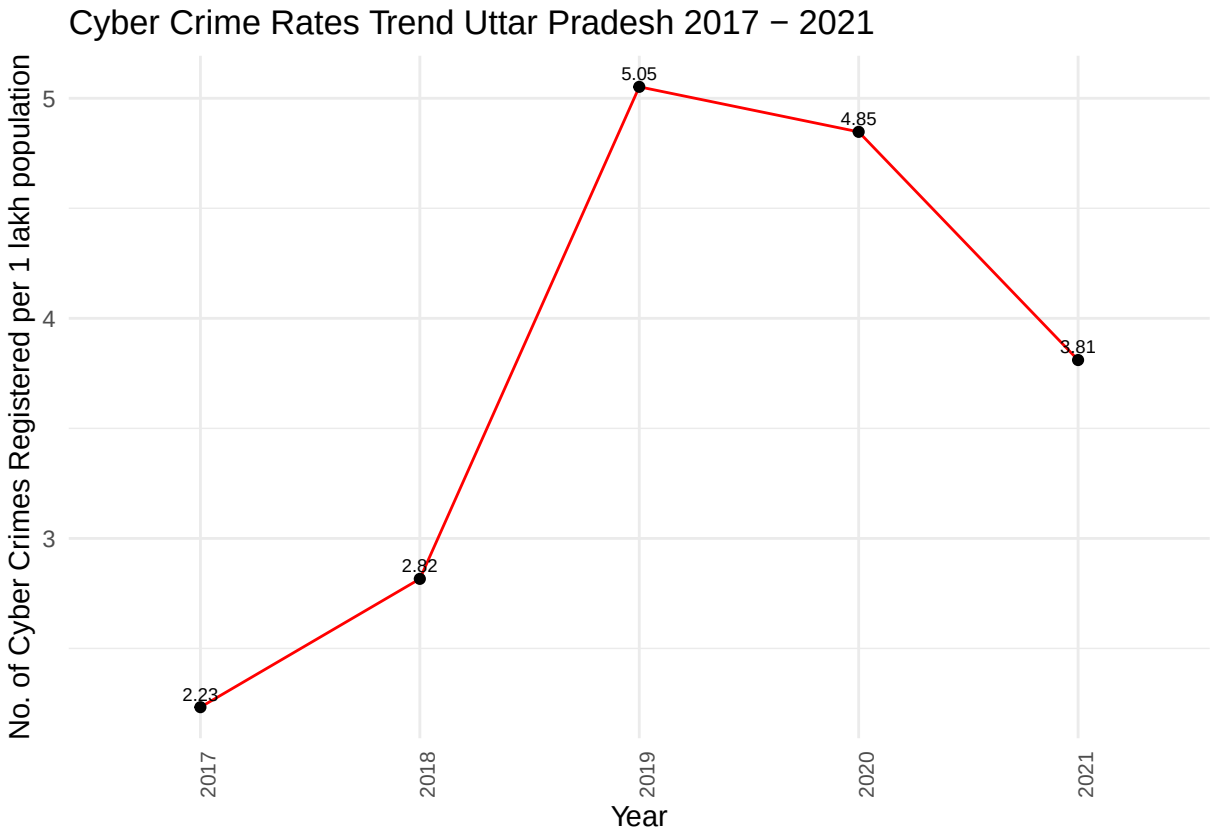
```
##
##
##  Assam
```


Cyber Crime Trend Assam 2017 − 2021
No. of Cyber Crimes Registered
5000
4000
3000
2000
1000
1120
2022
2231
3530
4846
2017
2018
2019
2020
2021
Year



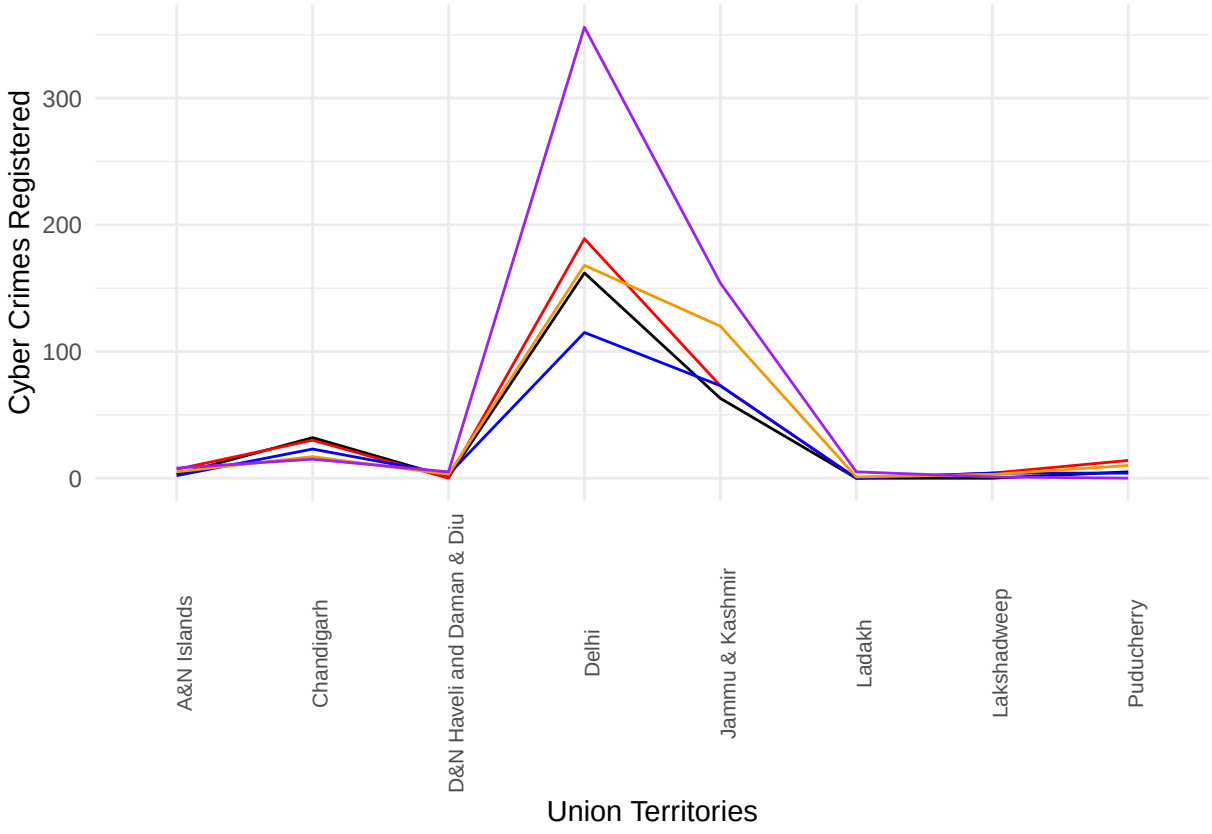

```
##
##
##  Uttar Pradesh
```

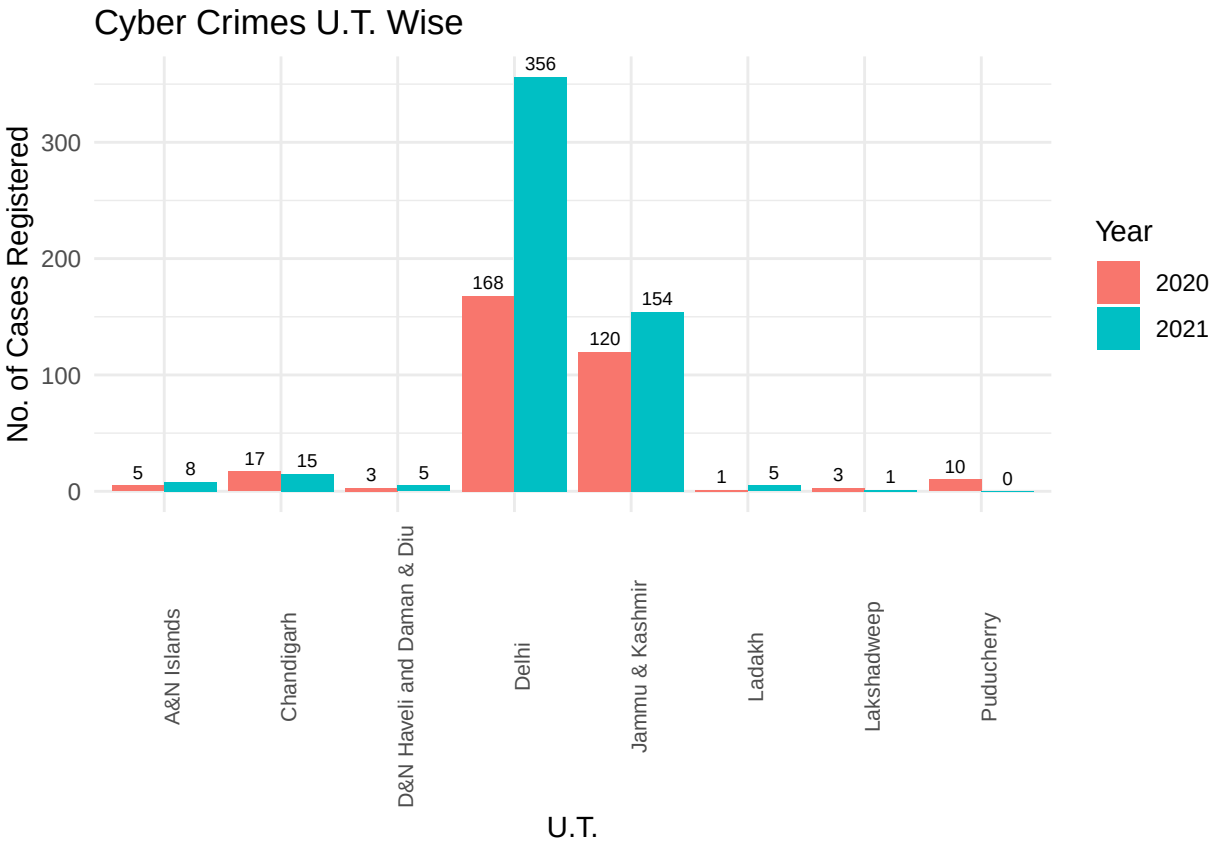




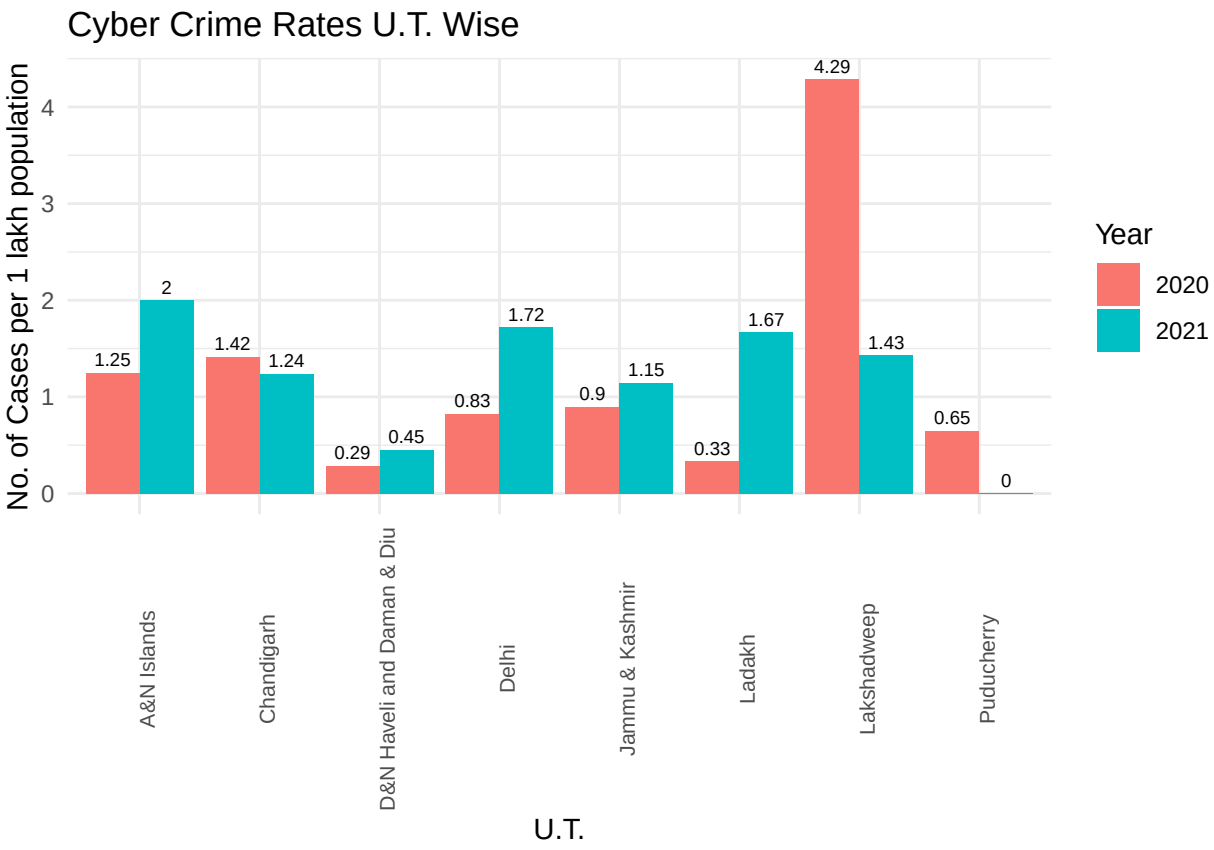
Now, it is evident from that the states of Assam and Telangana are showing rapid increasing trend in Cyber Crimes in last 5 years. However, U.P. and Karnataka are registering some lesser number of cases as compared to previous years.

1.3 Cyber Crimes U.T. - wise

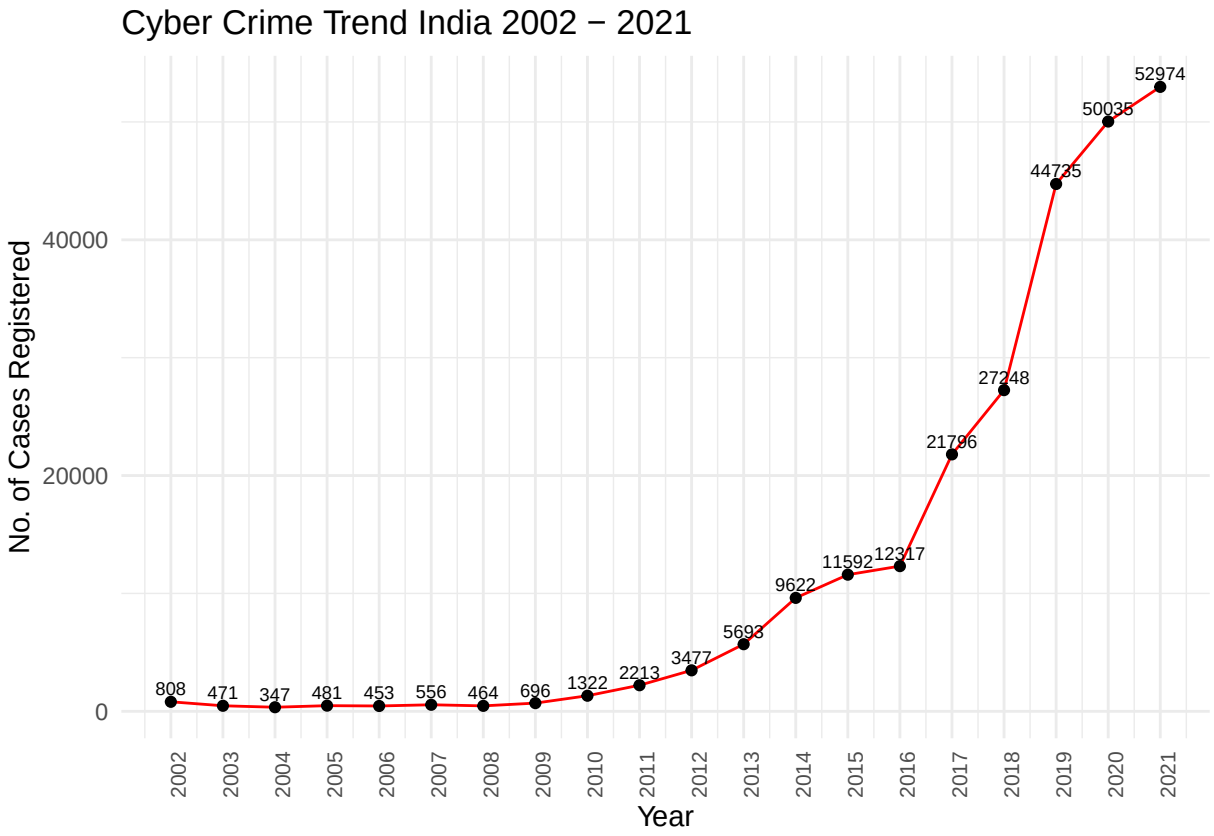




In terms of numbers, the Union Territory of Delhi registered the most 356 cyber crime cases in 2021, followed by Jammu&Kashmir and Chandigarh with 154 ans 15 cases respectively. Also, **the cases in Delhi are more than double in 2021 from 2020.** However, in terms of rates, the picture is somewhat like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



1.4 Cyber Crime Trend India



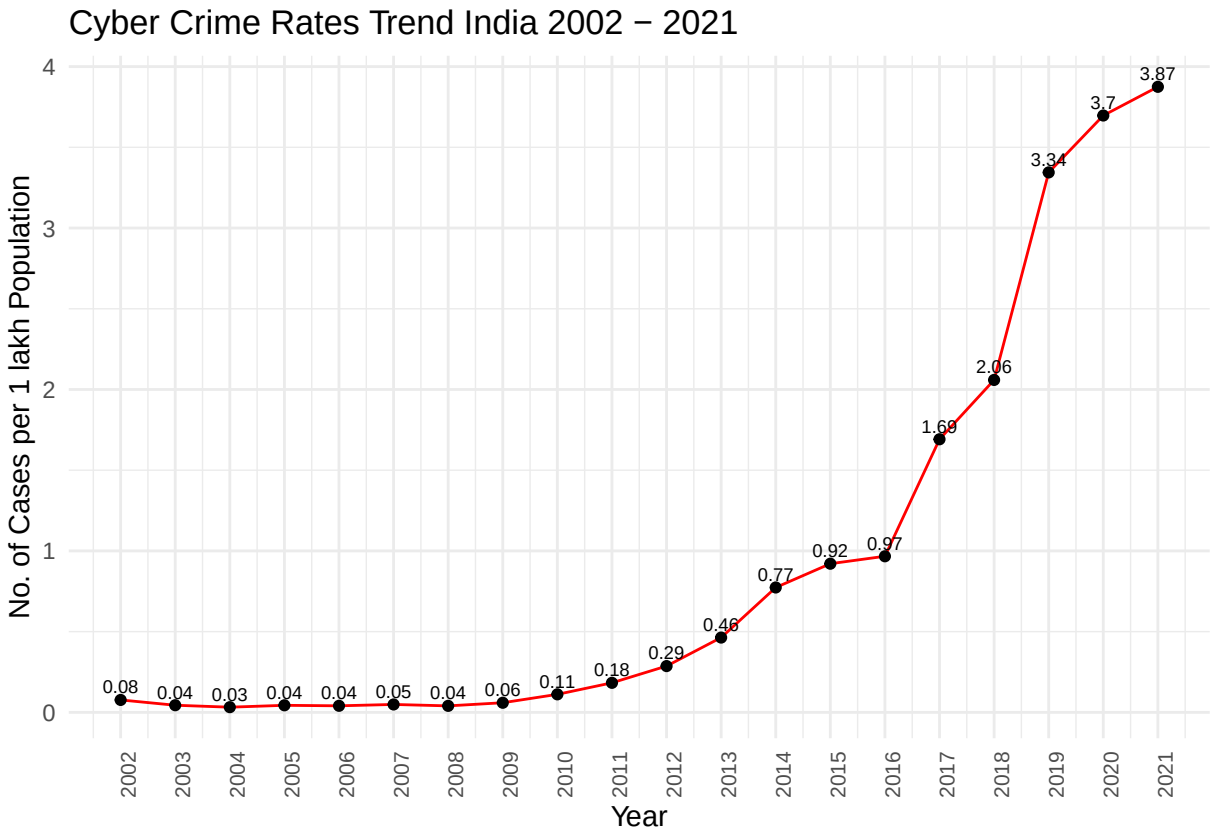
Now, the above plot shows the Trend for the number of Total Cyber Crimes registered in overall India from the year 2002 - 2021.

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:



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

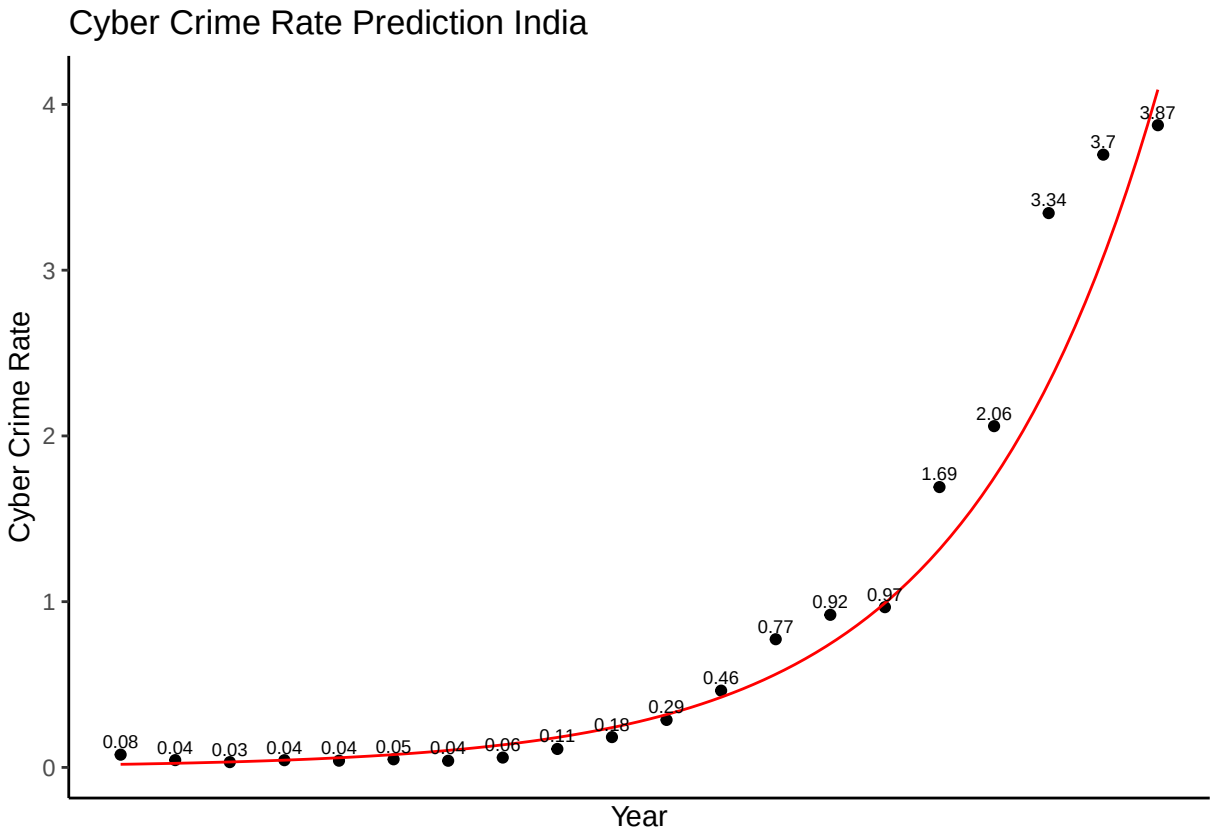
$$Y_t = \exp(-4.26073 + 0.28345\, t)$$

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

```
##
## 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:



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

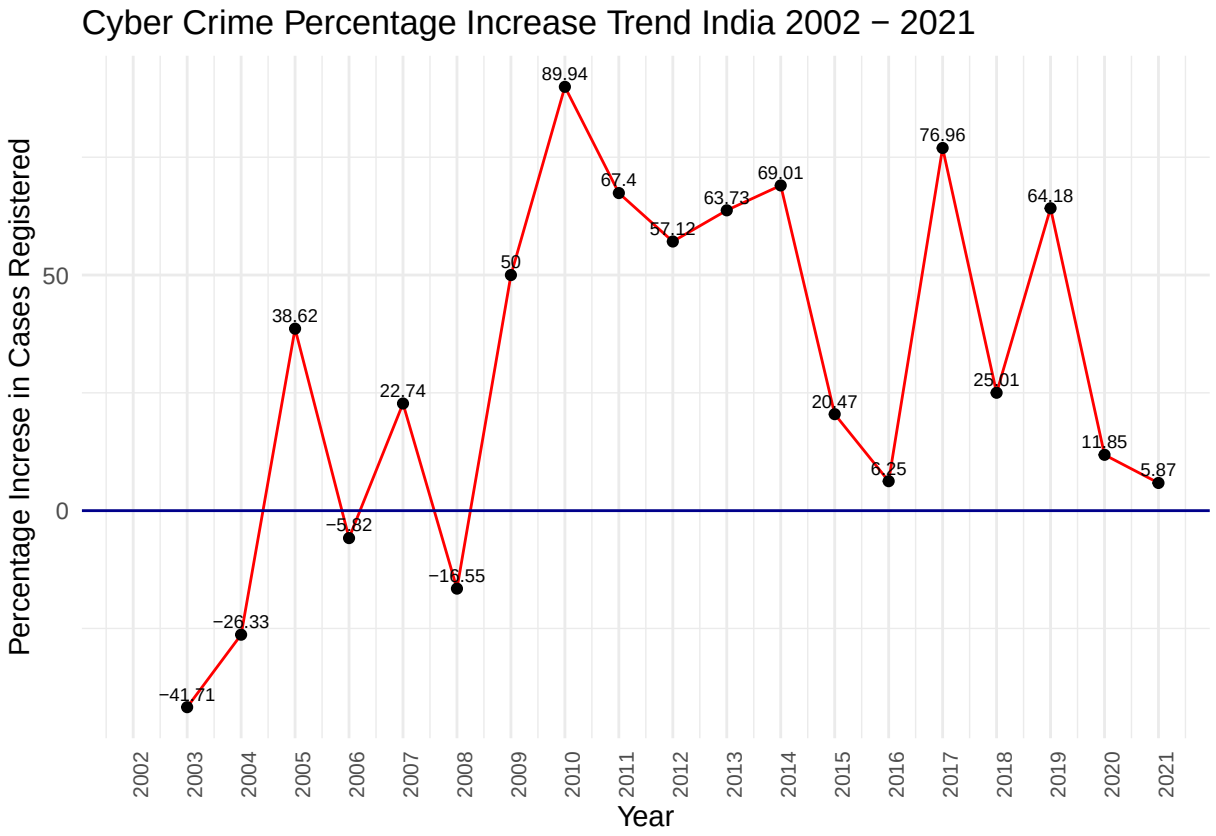
```
##    year predicted_rate
## 1  2022          5.429
## 2  2023          7.209
## 3  2024          9.571
## 4  2025         12.707
## 5  2026         16.871
## 6  2027         22.400
## 7  2028         29.741
## 8  2029         39.488
## 9  2030         52.428
## 10 2031         69.609
```

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

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

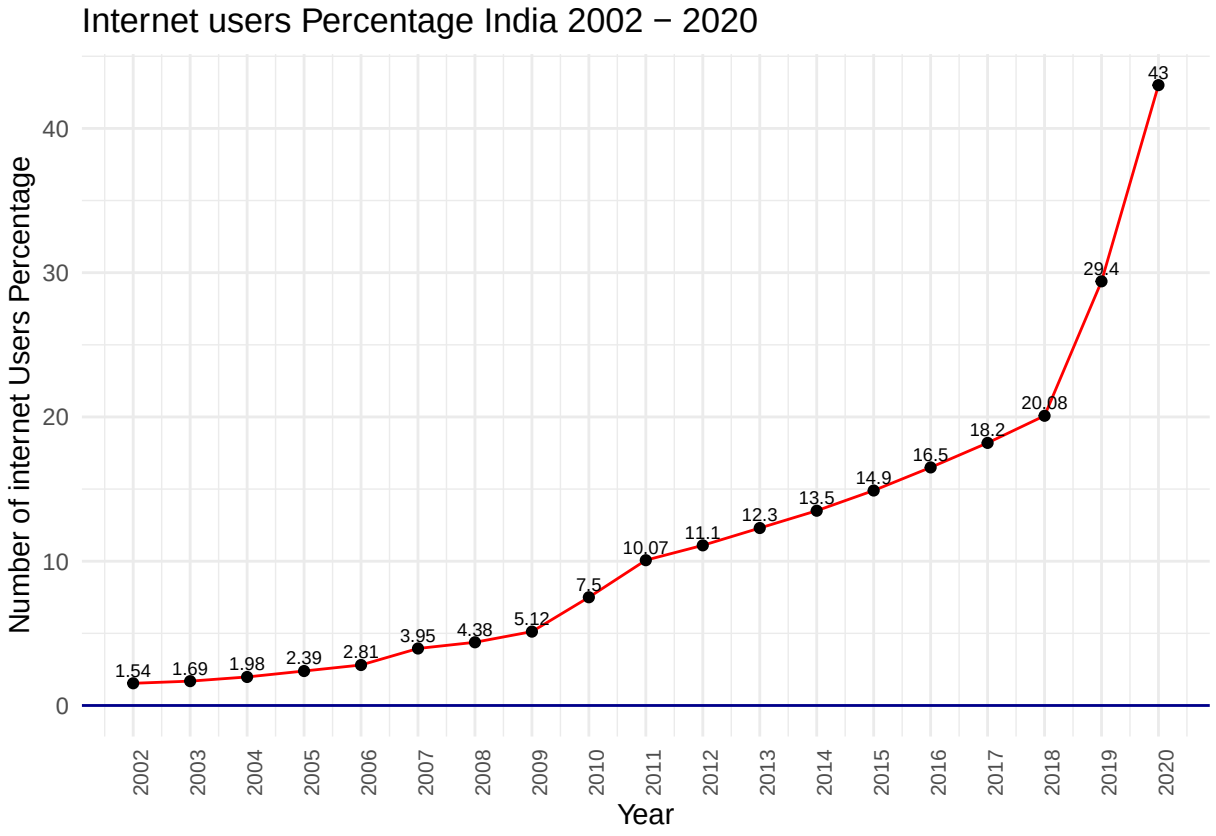
From the Percentage Increase plot, following facts can be noted:

- 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.



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:



It can be seen that the percentage of internet users in the country has increased in the last few years. 43% population of

India had access to the internet by the year 2020. According to a report by Tanushree Basuroy (2022), it was estimated that in 2025, number of active internet users would surpass 900 million in the country. In fact, India was ranked as the second largest online market worldwide in 2019, coming second only to China. The number of internet users was estimated to increase in both urban as well as rural regions, indicating a dynamic growth in access to internet.

Chapter 2

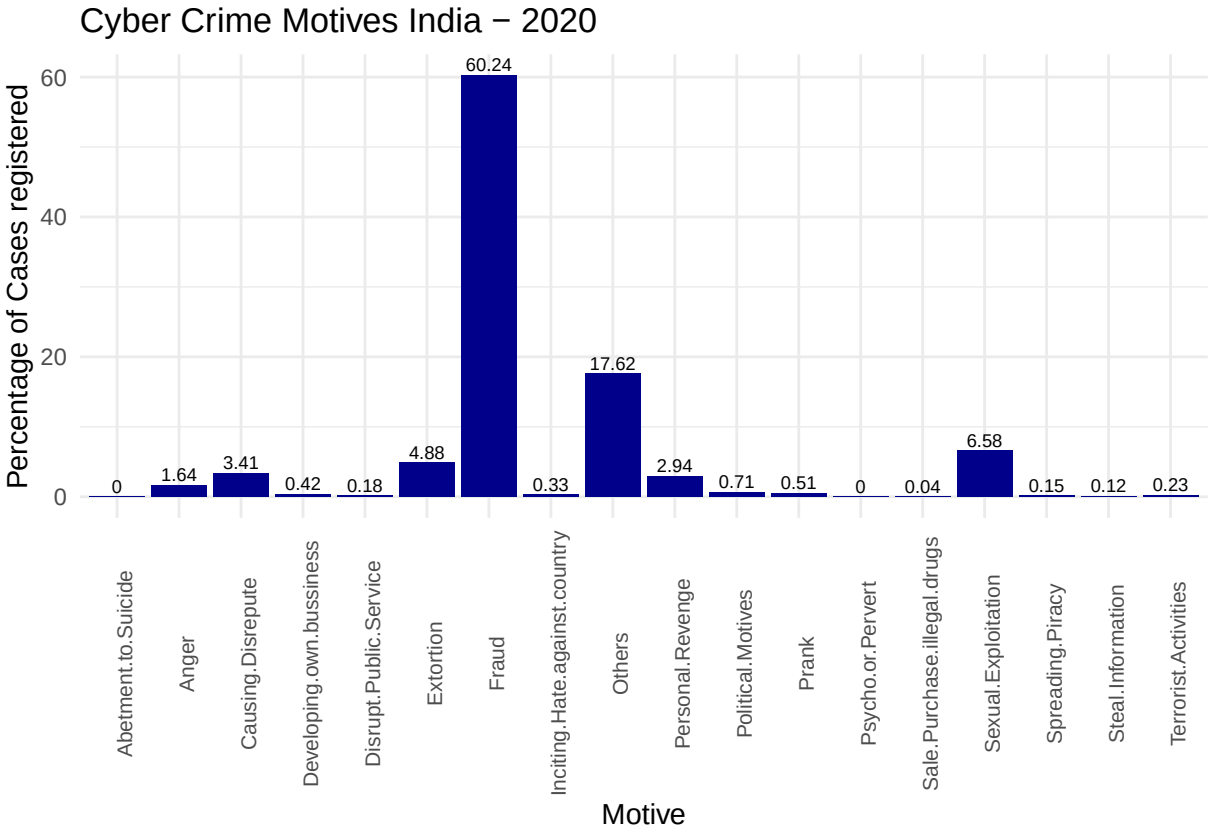
Cyber Crime Motives

2.1 Introduction

In this section, the motives behind the emerging cyber crime incidents in India are analyzed. Major motive is figured out and states are futher analyzed for that for the past few years. Cluster Analysis Technique is used to divide states into two groups (clusters) based on their crime incidence for the major motive keeping mid year projected population under consideration.

2.2 Cyber Crime Motives Year Wise

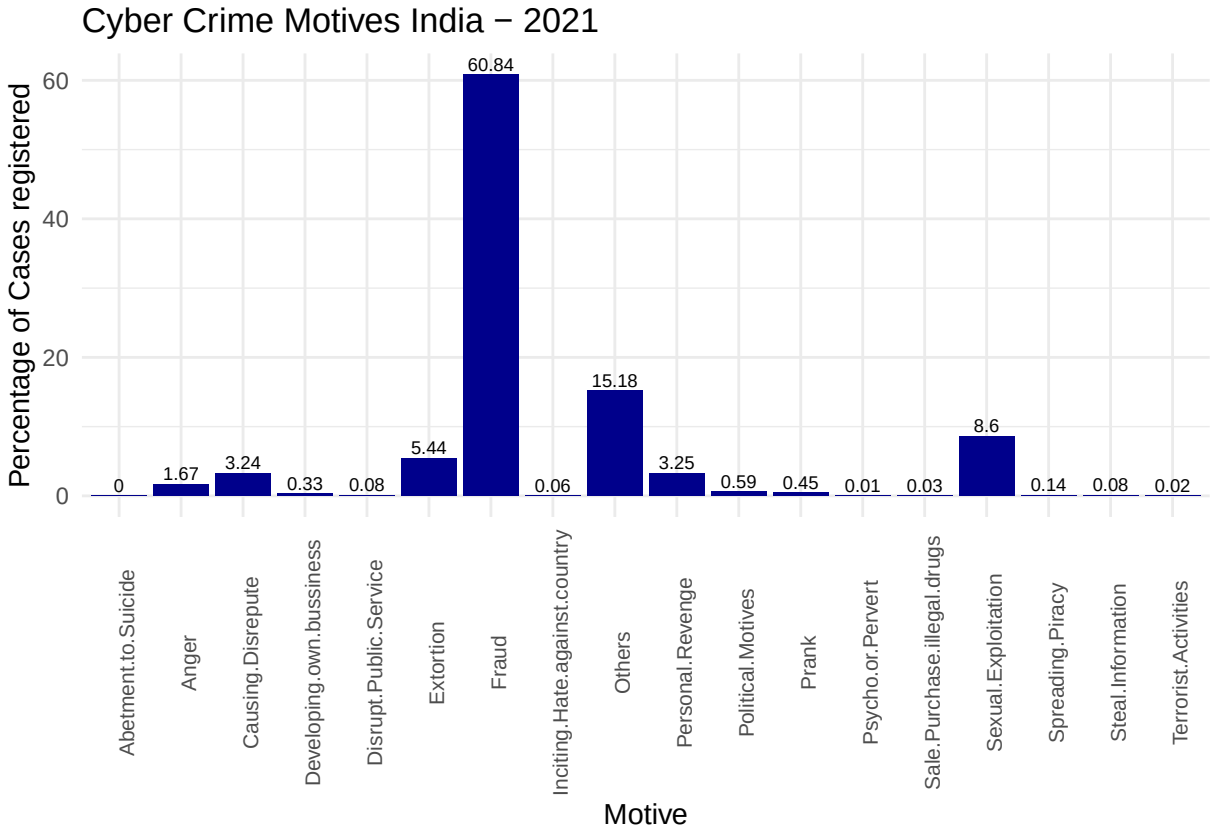
2.2.1 Year 2020



From the above figure, 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.

2.2.2 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.



2.3 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 clusters using:
- K-Means Clustering is implemented (Non-Hierarchical)
- Agglomerative Cluster Analysis (Hierarchical)
- Divisive Techniques (Hierarchical)

2.3.1 Distance Matrix Calculations

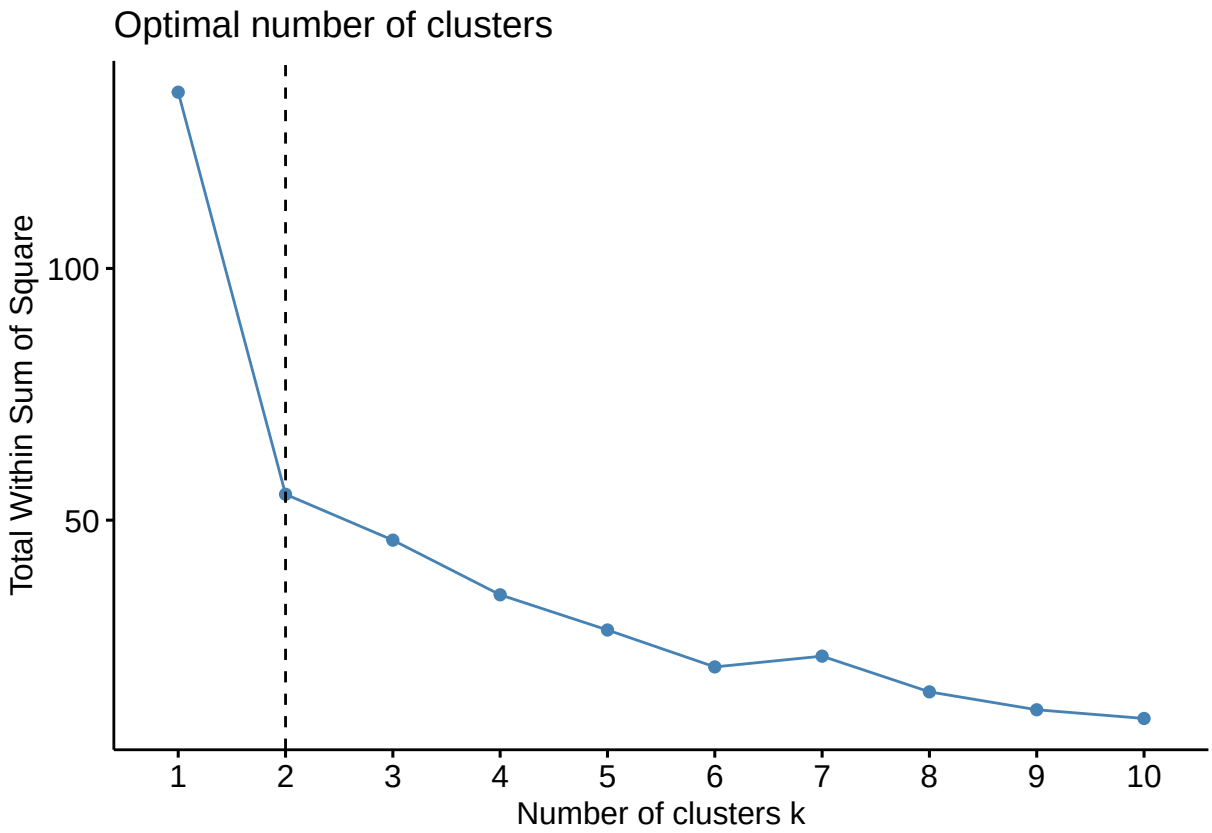
Data has been scaled to centre 0, and unit variability of each year rate (Variables).

```
##                   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
```

This is some part of the 28x28 distance matrix of states.

2.3.2 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.



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.

```
##    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.

Define

$$Z_{ji} = 1, \qquad if\ \mathbf{X_i} \in \textbf{jth cluster}$$

and

$$Z_{ji} = 0, \qquad if\ \mathbf{X_i} \notin \textbf{jth cluster}$$

where $\mathbf{X_i}$ is the ith case vector. i = 1,2,... N.

Let $\mathbf{m_j}$ be the mean vector of jth cluster, and $\mathbf{m}$ be the grand mean.

Let n_j be the number of elements in the jth cluster.

The Within cluster sum of squares vector are given by:

$$WSSV = diag(S_W)$$

,

where

$$S_W = \frac{1}{N}\sum_{j=1}^{g}\sum_{i=1}^{N} Z_{ji}(\mathbf{X_i} - \mathbf{m_j})(\mathbf{X_i} - \mathbf{m_j})'$$

```
## [1] 40.44021 14.70286
```

The Total Within cluster sum of squares are given by:

$$WSS = trace(S_W)$$

, where

$$S_W = \frac{1}{N}\sum_{j=1}^{g}\sum_{i=1}^{N} Z_{ji}(\mathbf{X_i} - \mathbf{m_j})(\mathbf{X_i} - \mathbf{m_j})'$$

```
## [1] 55.14307
```

and Between Cluster Sum of Squares are:

$$BSS = trace(B_W)$$

, where

$$B_W = \frac{1}{N}\sum_{j=1}^{g} n_j(\mathbf{m_j} - \mathbf{m})(\mathbf{m_j} - \mathbf{m})'$$

```
## [1] 79.85693
```

The Total cluster sum of squares are given by:

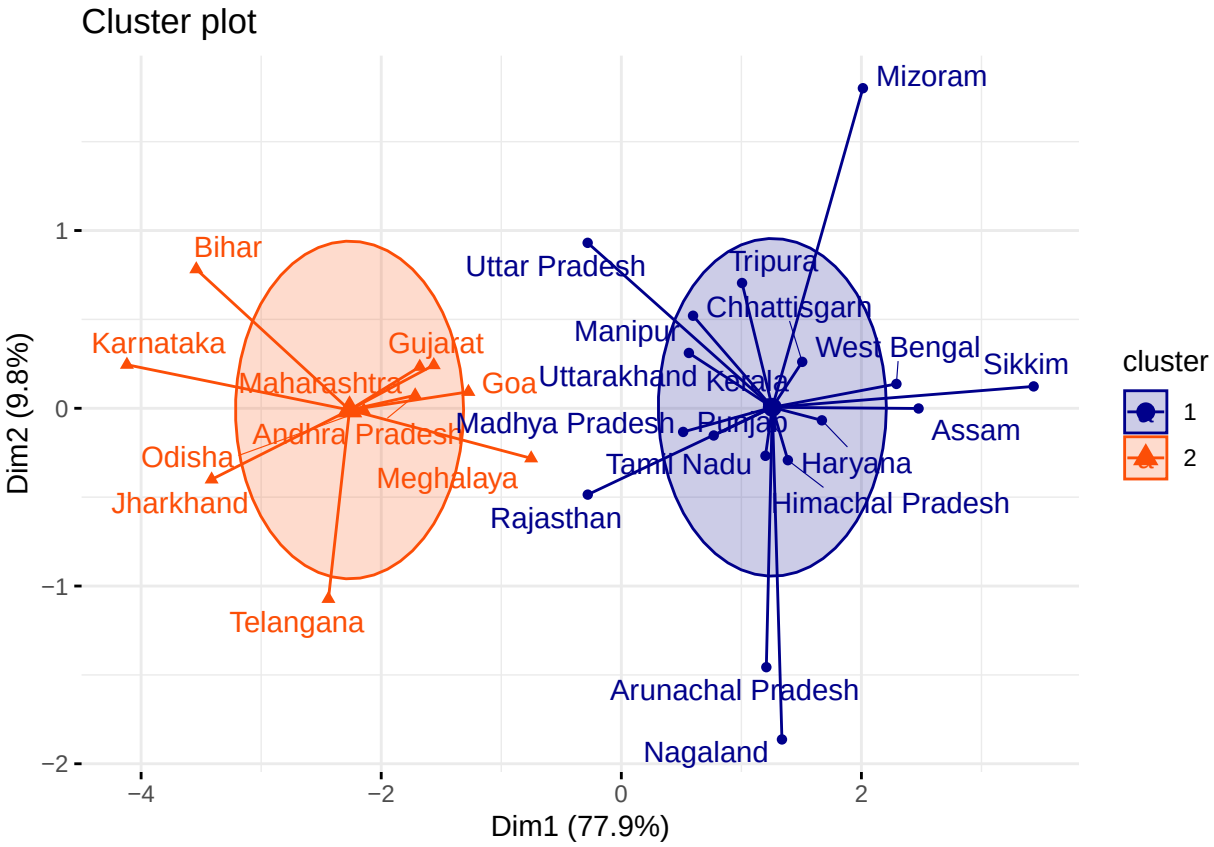
$$TSS = WSS + BSS$$

$\frac{WSS}{TSS}$ must be small.

```
## [1] 0.4084672
```

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

2.3.2.1 Cluster Plot



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.

2.3.3 Hierarchical Clustering

2.3.3.1 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.

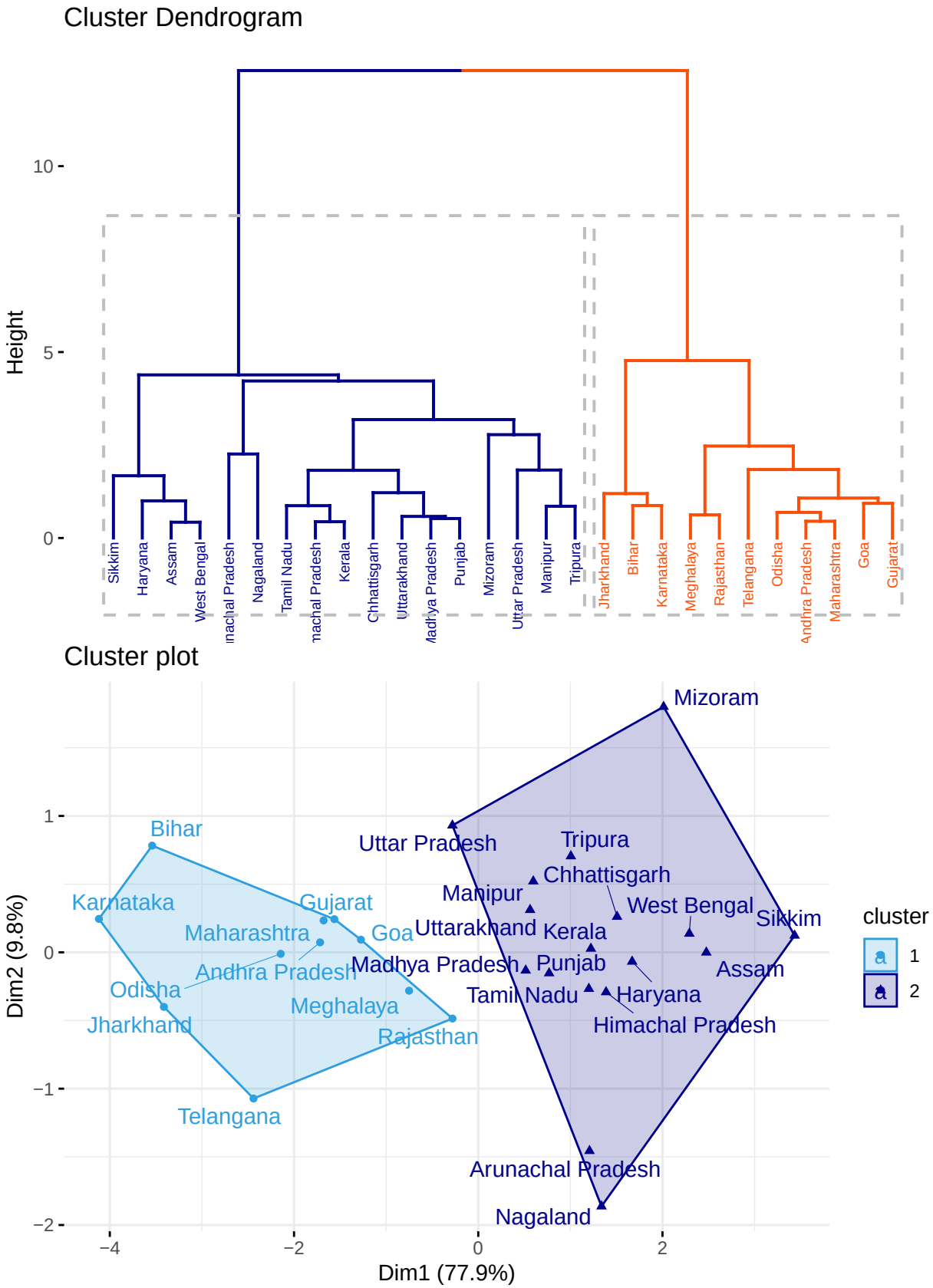
The states with major fraction of the Fraud cases are in one Cluster: Telangana, Jharkhand, Bihar etc.

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

The cluster means are given by

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

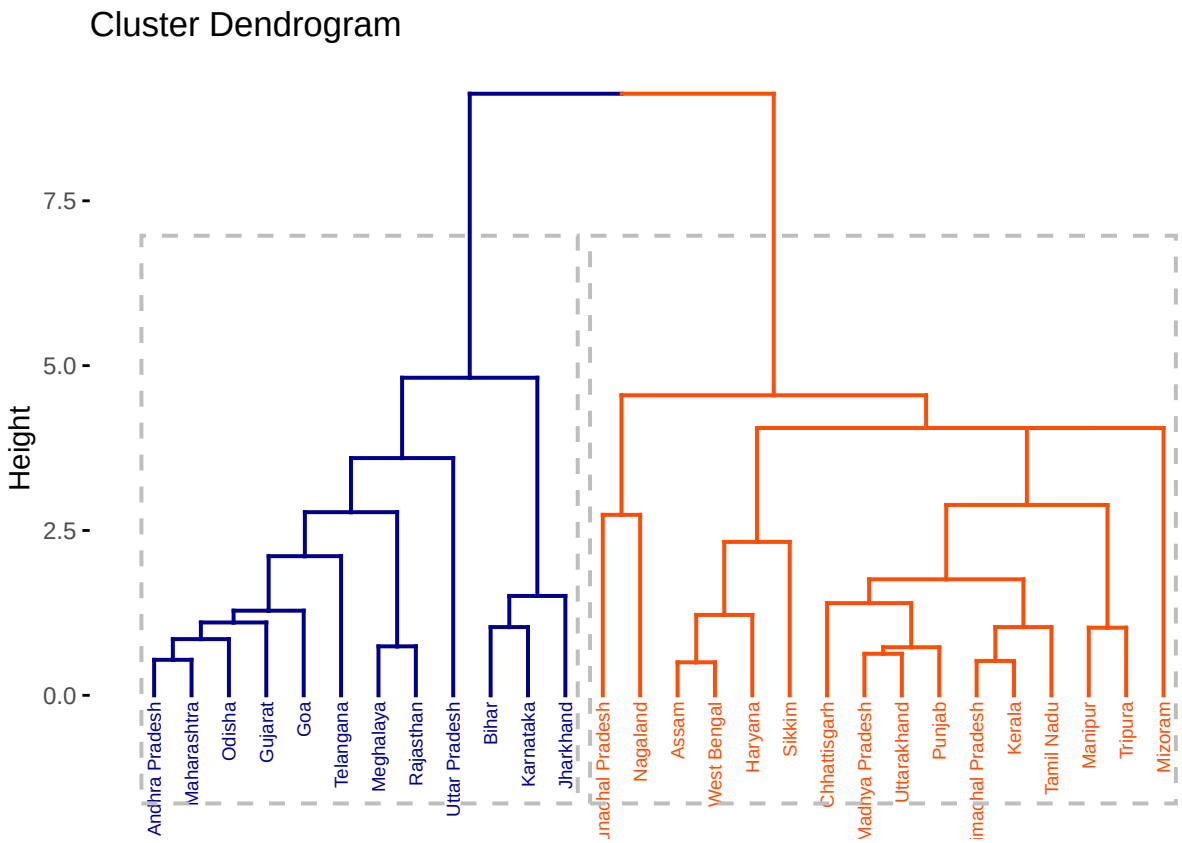
```
##  rate2021  rate2020  rate2019  rate2018  rate2017
## 0.2367204 0.2255833 0.1450668 0.2455529 0.2827396
```



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.

2.3.3.2 Divisive Clustering

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



Chapter 3

Analysis of Cyber crimes under IT Act and IPC Act

3.1 Introduction

3.1.1 Information Technology ACT, 2000

The IT Act was introduced in 17 October 2000. It is the primary law in India dealing with cyber crime and electronic commerce. The original Act contained 94 sections, divided into 13 chapters and 4 schedules. The laws apply to the whole of India. If a crime involves a computer or network located in India, persons of other nationalities can also be indicted under the law. The sections of IT Act included in the analysis are

- Tampering Computer Source Documents (Section 65)
- Computer Related Offenses (Sec 66 & Sec 66 B to E)(TOTAL)
- Cyber Terrorism (Section 66F)
- Publication/ transmission of obscene / sexually explicit act in electronic form

The figure 3.1 shows the rate of cases of Information Technology Act, 2000 being registered in INDIA per 1 lakh population in years 2016-21. One must note that the Computer Related Offenses (Sec 66 & Sec 66 B to E) are reported mostly.

3.1.1.1 Computer Related Offenses (Sec 66 & Sec 66 B to E)

This includes punishment regarding sending offensive messages through communication service,Punishment for dishonestly receiving stolen computer resource or communication, Punishment for identity theft, Punishment for cheating by personation by using computer resource, Punishment for violation of privacy, Penalty and compensation for damage to computer, computer system, etc.

3.1.2 Indian Penal Code , 1862

The Indian Penal Code (IPC) is the official criminal code of India. It is a comprehensive code intended to cover all substantive aspects of criminal law. It came into force in India during the British rule in 1862. The sections of IPC included in the analysis are

- Data theft (Sec.379 to 381)
- Fraud (Sec.420 r/w Sec.465,468-471 IPC)
- Cheating (Sec.420)
- Forgery (Sec.465, 468 & 471)

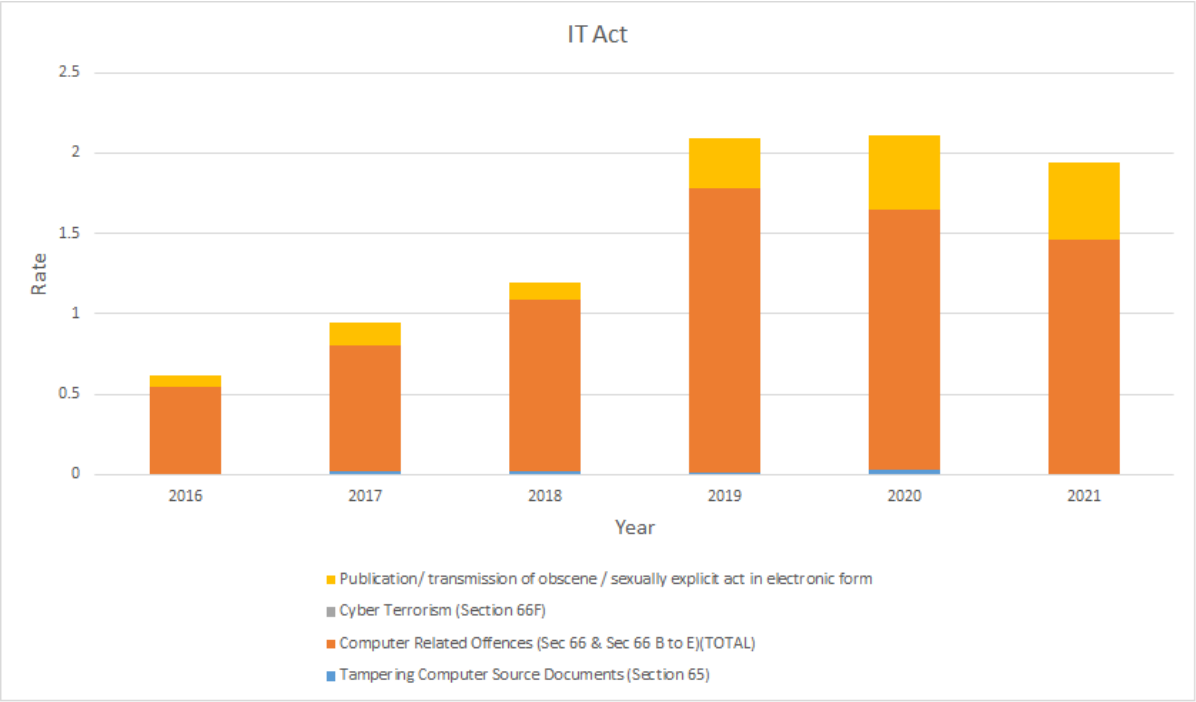


Figure 3.1: rates section-wise yearly it-act cases

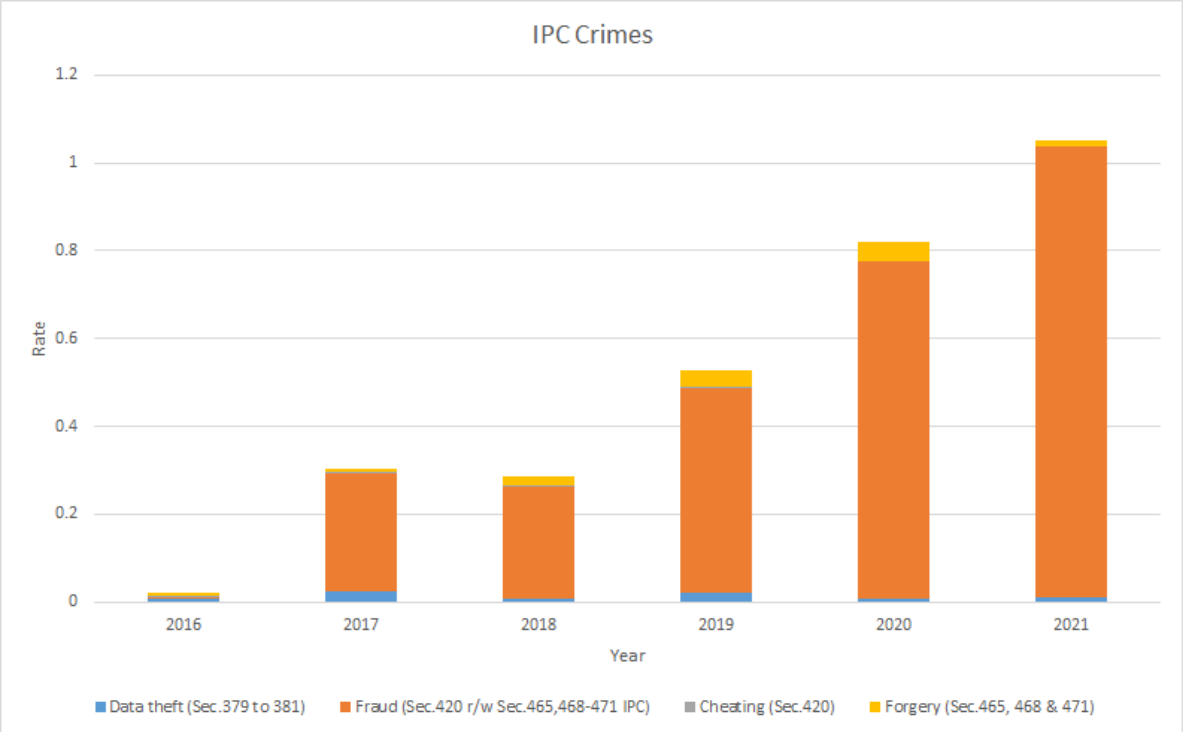


Figure 3.2: rates section-wise yearly ipc-act cases

The figure 3.2 shows the rate of cases of Indian Penal Code (here only those sections are included which include punishments to cyber crimes) being registered in INDIA per 1 lakh population in years 2016-21. One must note that the FRAUD cases are reported mostly.

3.1.2.1 Fraud (Sec.420 r/w Sec.465,468-471 IPC)

This include that whoever cheats and thereby dishonestly induces the person deceived to deliver any property to any person, or to make, alter or destroy the whole or any part of a valuable security, or anything which is signed or sealed, and which is capable of being converted into a valuable security, shall be punished with imprisonment and shall also be liable to fine.

3.2 Percentage share of major crimes under IT Act and IPC

```
##      year       CRO     Total per_share
## [1,] 2016 0.5351648 0.9667975   55.3544
## [2,] 2017 0.7844598 1.6915400   46.3755
## [3,] 2018 1.0685517 2.0589702   51.8974
## [4,] 2019 1.7652248 3.3300190   53.0095
## [5,] 2020 1.6198963 3.6969307   43.8173
## [6,] 2021 1.4566373 3.8746626   37.5939
```

```
##      year       Fraud     Total per_share
## [1,] 2016 0.004395604 0.9667975    0.4547
## [2,] 2017 0.268988693 1.6915400   15.9020
## [3,] 2018 0.253366380 2.0589702   12.3055
## [4,] 2019 0.465976884 3.3300190   13.9932
## [5,] 2020 0.768054263 3.6969307   20.7755
## [6,] 2021 1.024510127 3.8746626   26.4413
```

The above data table shows that Computer Related Offenses (Sec 66 & Sec 66 B to E) ,(here , Computer Related Offenses (Sec 66 & Sec 66 B to E) is named CRO) , has been contributing about 48% and Fraud contribute about 15% of the total cyber crimes each year as per the NCRB data for cyber crimes from year 2016 - 2021. Hence Computer Related Offenses (Sec 66 & Sec 66 B to E) is the major contributor to the cyber crimes in INDIA.

3.3 Randomness and Association

3.3.1 1-sample Run test for randomness

1-sample run test is used to test whether the observations are random

3.3.1.1 Computer related offenses (Sec 66 & Sec 66 B to E)

```
##
##  Runs Test for Randomness
##
## data:  cro_c
## runs = 2, m = 3, n = 3, p-value = 0.2
## alternative hypothesis: true number of runs is not equal the expected number
## sample estimates:
## median(x)
##     17028
```

This shows that year wise incidence of cases of Computer related offenses are random .

3.3.1.2 Fraud

```
##
##  Runs Test for Randomness
##
## data:  fr_c
```

```
## runs = 2, m = 3, n = 3, p-value = 0.2
## alternative hypothesis: true number of runs is not equal the expected number
## sample estimates:
## median(x)
##    4849.5
```

This shows that year wise incidence of cases of Fraud are random .

3.3.2 Kolmogorov Smirnov test

This test is used for testing whether both the samples come from same distribution

Test statistics,

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of the first and the second sample respectively, and sup is the supremum function.

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  y2016 and y2021
## D = 0.14286, p-value = 0.9333
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  y2017 and y2021
## D = 0.14286, p-value = 0.9366
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  y2018 and y2021
## D = 0.10714, p-value = 0.9956
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  y2019 and y2021
## D = 0.14286, p-value = 0.9301
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  y2020 and y2021
## D = 0.14286, p-value = 0.9361
## alternative hypothesis: two-sided
```

This shows that the distribution of computer related offenses throughout the states remains the same in the years 2016 - 2021 i.e. cases are concentrated in specific states and are distrubuted similarily each year.

3.3.3 Kendall Tau Measure of Association

It measures the relationship between two columns of ranked data.

Null hypothesis, The columns are not associated with each other against, The columns are associated with each other

Test statistics,

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{(\text{number of pairs})}$$

Test is to reject null hypothesis for

```
##
##  Kendall's rank correlation tau
##
## data:  cro_c and fr_c
## T = 11, p-value = 0.2722
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.4666667
```

This shows that there is no association between cyber crimes under Computer Related Offenses (Sec 66 & Sec 66 B to E) and Fraud.

3.4 Testing variability and location

Now, checking for variability wrt location and scale for Computer Related Offenses (Sec 66 & Sec 66 B to E) in years 2016-21 using Levene's Test for variability and Kruskal Wallis Test for location

3.4.1 Levene Test

Null hypothesis, variances of all samples are equal against , variances of all samples are not equal

Test statistics,

$$W = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2},$$

where, k: number of different groups to which the sampled cases belong. Ni: Number of elements in different groups. N: total number of cases in all groups

$$Z_{ij} = \left\{ \begin{array}{l} \left|Y_{ij} - \bar{Y}_i\right|, \bar{Y}_i \, is \, the \, mean \, of \, the \, i^{th} \, group \\ \left|Y_{ij} - \tilde{Y}_i\right|, \tilde{Y}_i \, is \, the \, median \, of \, the \, i^{th} \, group \end{array} \right\}$$

where, Y_{ij}: the value of jth case and ith group

Test is to reject null hypothesis for larger values of W

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   5  0.6081 0.6938
##       162
```

Since p-value > 0.05 , for levene's test hence the test is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant variability in the data through 2016-21

Now testing for equality in location using Kruskal Wallis Test

3.4.2 Kruskal Wallis Test

Null hypothesis, location of all populations is same against, location of any two populations differs

Test Statistics,

$$H = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot} - \bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij} - \bar{r})^2}, \text{ where}$$

- N is the total number of observations across all groups
- g is the number of groups
- n_i is the number of observations in group i
- r_{ij} is the rank (among all observations) of observation j from group i
- $\bar{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$ is the average rank of all observations in group i
- $\bar{r} = \frac{1}{2}(N+1)$ is the average of all the r_{ij}.

Test is to reject null hypothesis for large values OF H

```
##
##  Kruskal-Wallis rank sum test
##
## data:  cases by as.factor(year)
## Kruskal-Wallis chi-squared = 1.0211, df = 5, p-value = 0.9608
```

Since p-value > 0.05 , for Kruskal Wallis test, the result is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant change in the locations throughout the years 2016-21

Hence, from both the tests,
Since there is no significant difference in location and variability in different years.We conclude that Computer Related Offenses (Sec 66 & Sec 66 B to E) are statistically same in terms of location and variability for different states.

3.5 State-wise comparison for Cases of Computer Related Offenses (Sec 66 & Sec 66 B to E)

3.5.1 Count of cases in each state

The figure 3 shows the highest number of cases being registered in Karnataka, followed by Uttar Pradesh and Assam. 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 2016 - 2021 is available. Hence the Crime Rate can be calculated.

3.5.2 Crime rate in each state

The figure 4 shows the highest rate of cases being registered in Karnataka wrt 1 Lakh population in respective year, followed by Assam and UttarPradesh. But one must note that the Uttrakhand had observed a peak in year 2021

The crime rate as well as counts of cyber crimes in Karnataka is surprisingly high. The pattern of crime rate wrt 1 Lakh population in respective year in Karnataka can be seen in figure 5.

3.5.3 Year wise increment of Computer Related Offenses (Sec 66 & Sec 66 B to E)

```
##      year       CRO  per_inc
## [1,] 2016 0.5351648       NA
## [2,] 2017 0.7844598  46.5828
## [3,] 2018 1.0685517  36.2150
## [4,] 2019 1.7652248  65.1979
```

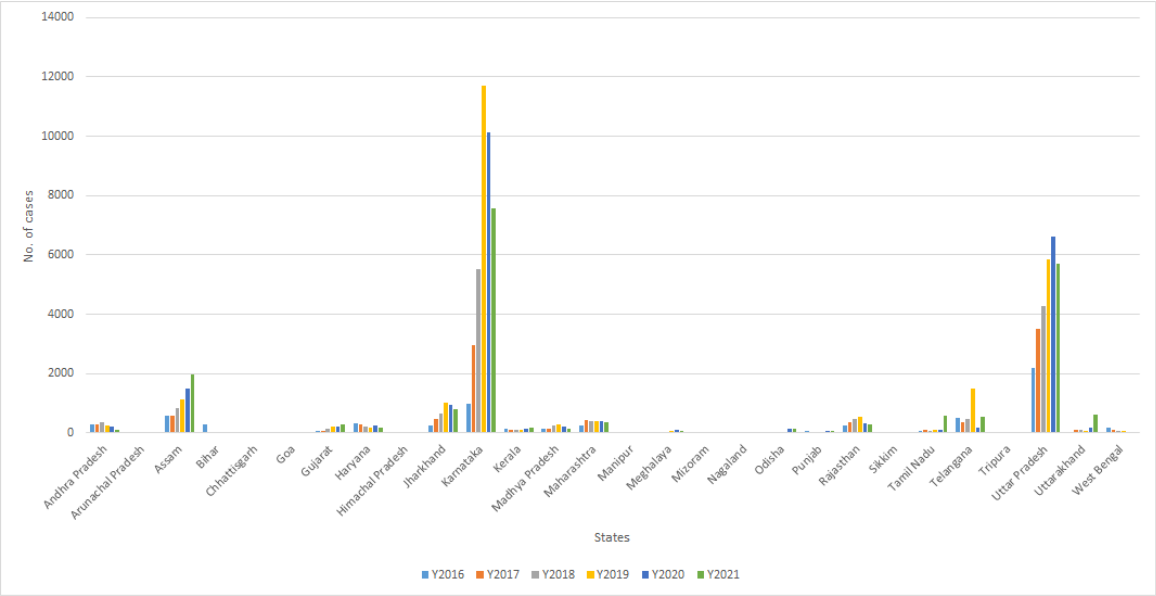


Figure 3.3: counts state-wise yearly CRO cases

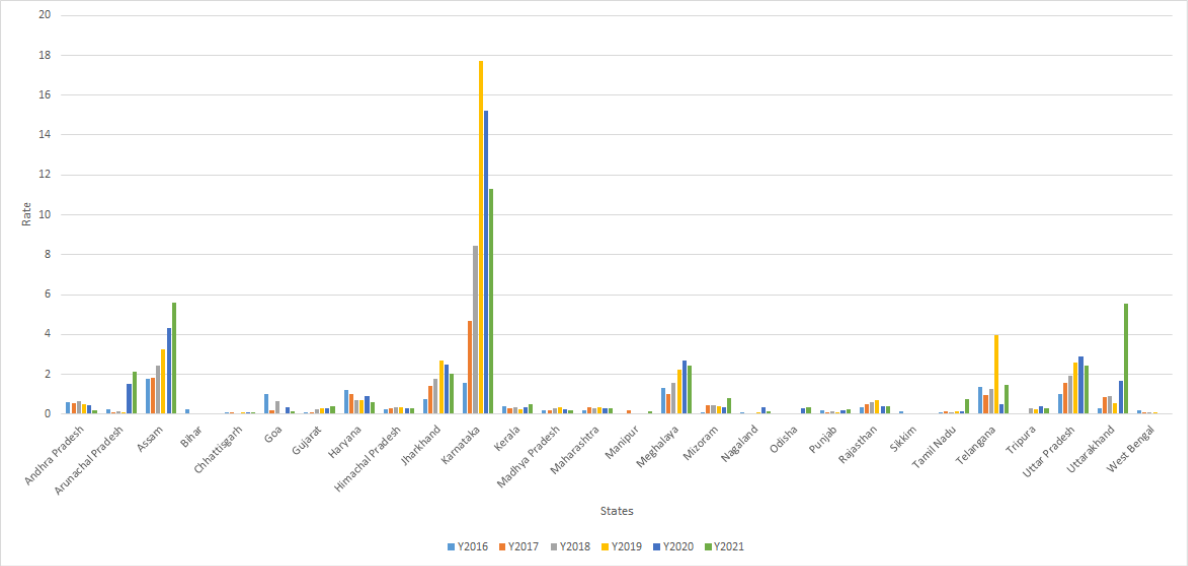


Figure 3.4: rates state-wise yearly CRO cases

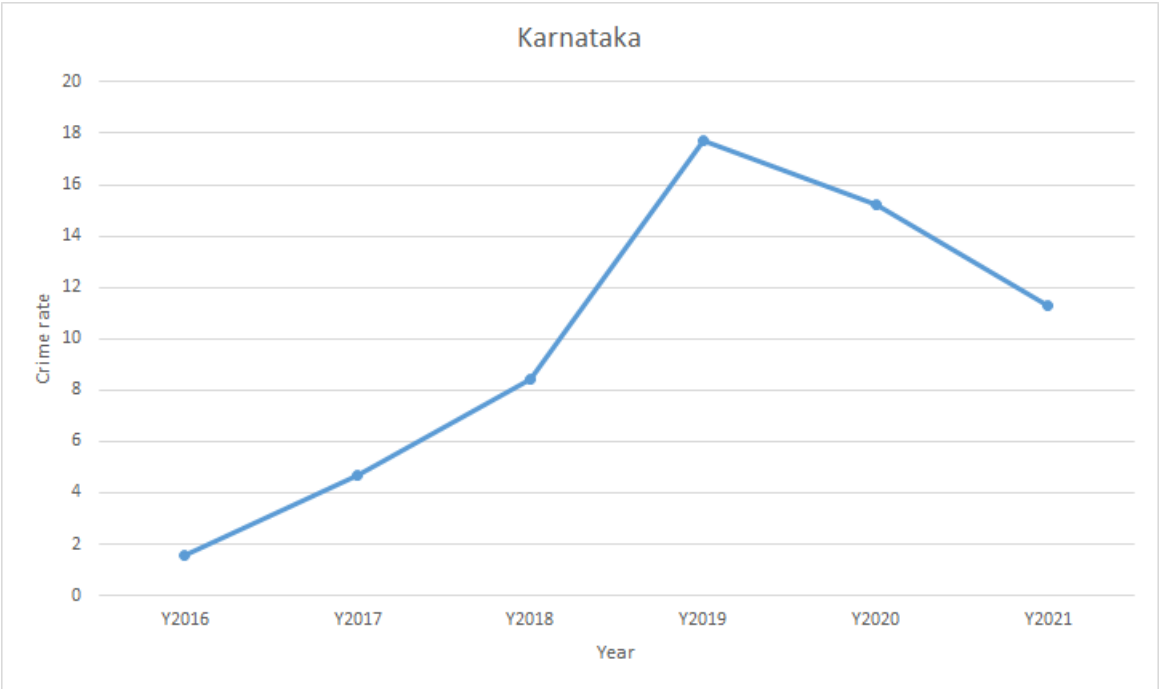


Figure 3.5: karnataka

```
## [5,] 2020 1.6198963  -8.2329
## [6,] 2021 1.4566373 -10.0784
```

The above data table shows the Computer Related Offenses got a peak in year 2019 when they increased by **65 %** as compared to previous year 2018. However a slight % decrease can be seen after that .

Chapter 4

Police Disposal of Cyber Crimes in India

4.1 Introduction

The core duty of the police service is to protect the public by detecting and preventing crime.
Police powers can be grouped into three categories:

Powers to investigate crime. This includes a range of powers to collect evidence needed to identify suspects and support their fair and effective trial.
Powers to prevent crime. This includes a range of powers to maintain public order, prevent anti-social behaviour and manage known offenders/ suspects.
Powers to 'dispose' of criminal cases. These powers allow police officers to dispose of criminal cases outside of court or charge suspects so they can be prosecuted.

Various crimes that are being registered and investigated by different law enforcement agencies are broadly grouped under the following categories for statistical information system.

Broad classification of crimes Crimes under I.T. Act:
1.Tampering computer source documents
2.Computer Related Offences
3.Cyber Terrorism
4.Publication/transmission of obscene / sexually explicit act in electronic form
5.Decryption of Information
6.Un-authorized access/attempt to access to protected computer system
7.Abetment to Commit Offences
8.Attempt to Commit Offences
9.Other Sections of IT Act

Crimes under the Indian Penal Code(IPC)
1.Abetment of Suicide (Online)
2.Cyber Stalking/Bullying of Women/Children
3.Data theft
4.Fraud
5.Credit Card/Debit Card
6.ATMs
7.Online Banking Fraud
8.OTP Frauds
9.Others
10.Cheating
11.Forgery
12.Defamation/Morphing
13.Fake Profile
14.Counterfeiting
15.Currency
16.Stamps

17.Cyber Blackmailing/Threatening
18.Fake News on Social Media
19.Other Offences

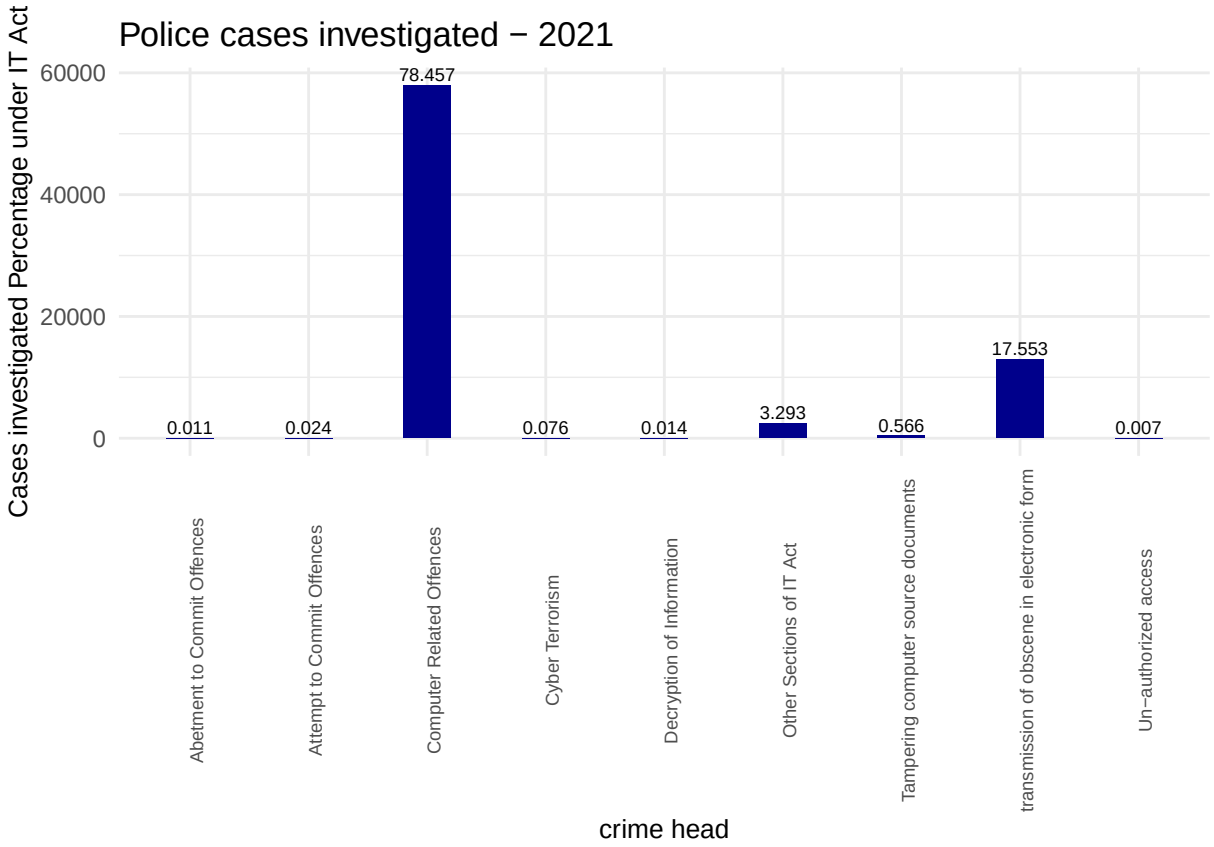
Crimes under the Special and Local Laws (SLL)
1.Gambling Act (Online Gambling)
2.Lotteries Act (Online Lotteries)
3.Copy Right Act
4.Trade Marks Act
5.Other SLL Crimes

4.2 Statistics and Data Visualization

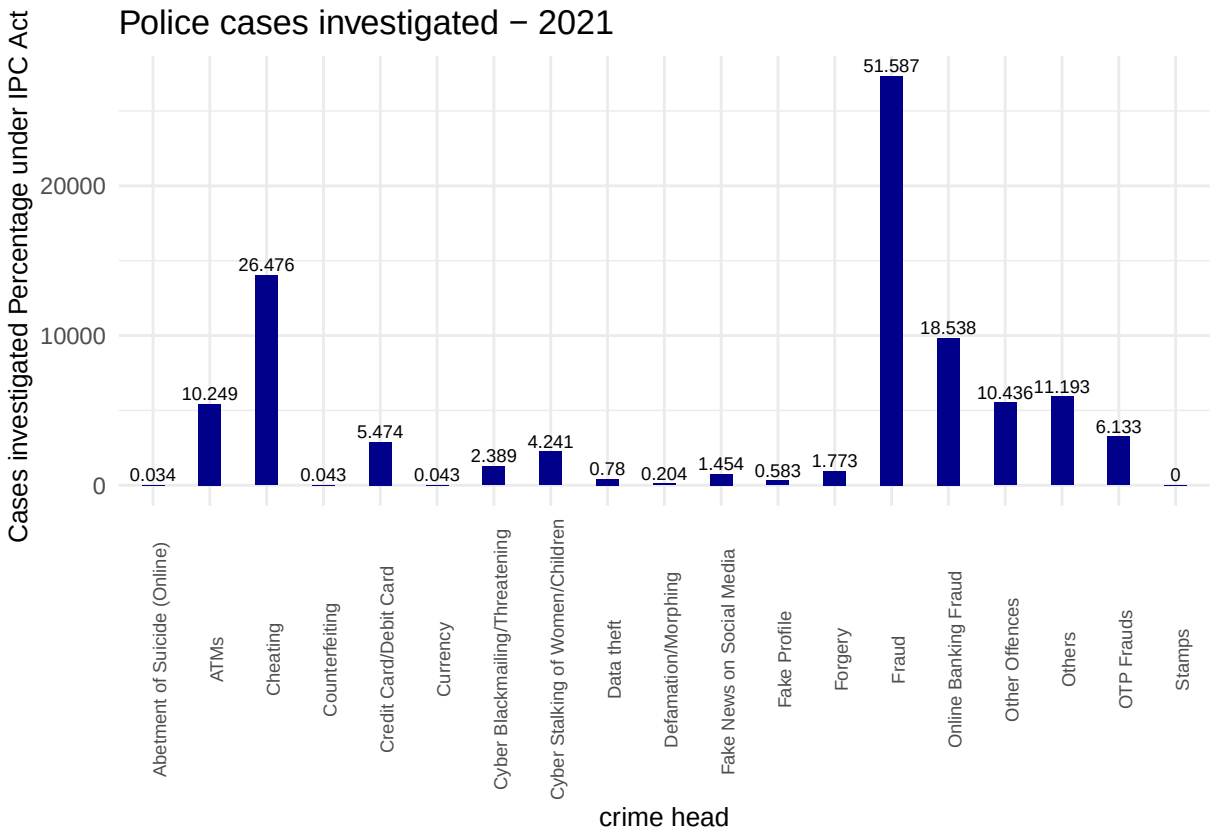
Police Disposal of Cases: Crime Head-Wise

4.3 Total Cases Investigated

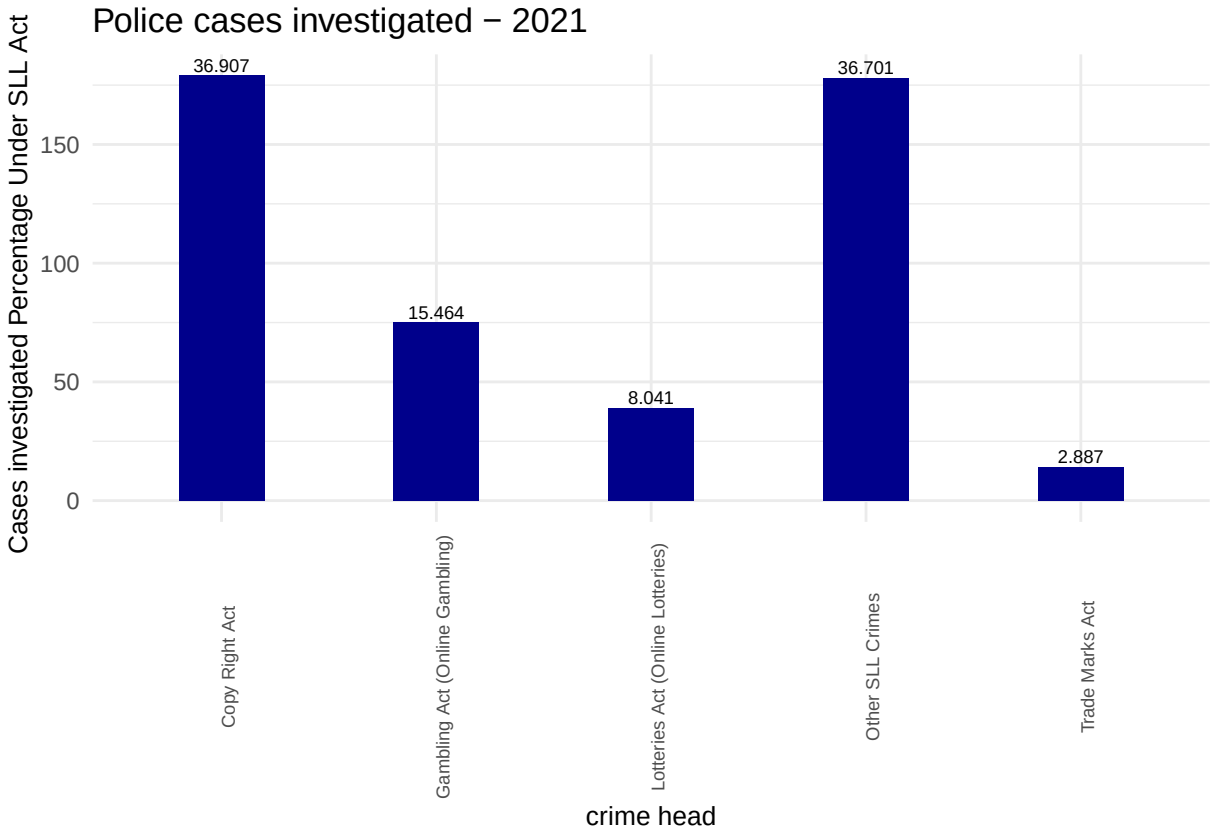
4.3.1 Year 2021



The above figure shows that under IT Act of computer related offences are highest with **78.457 %**.

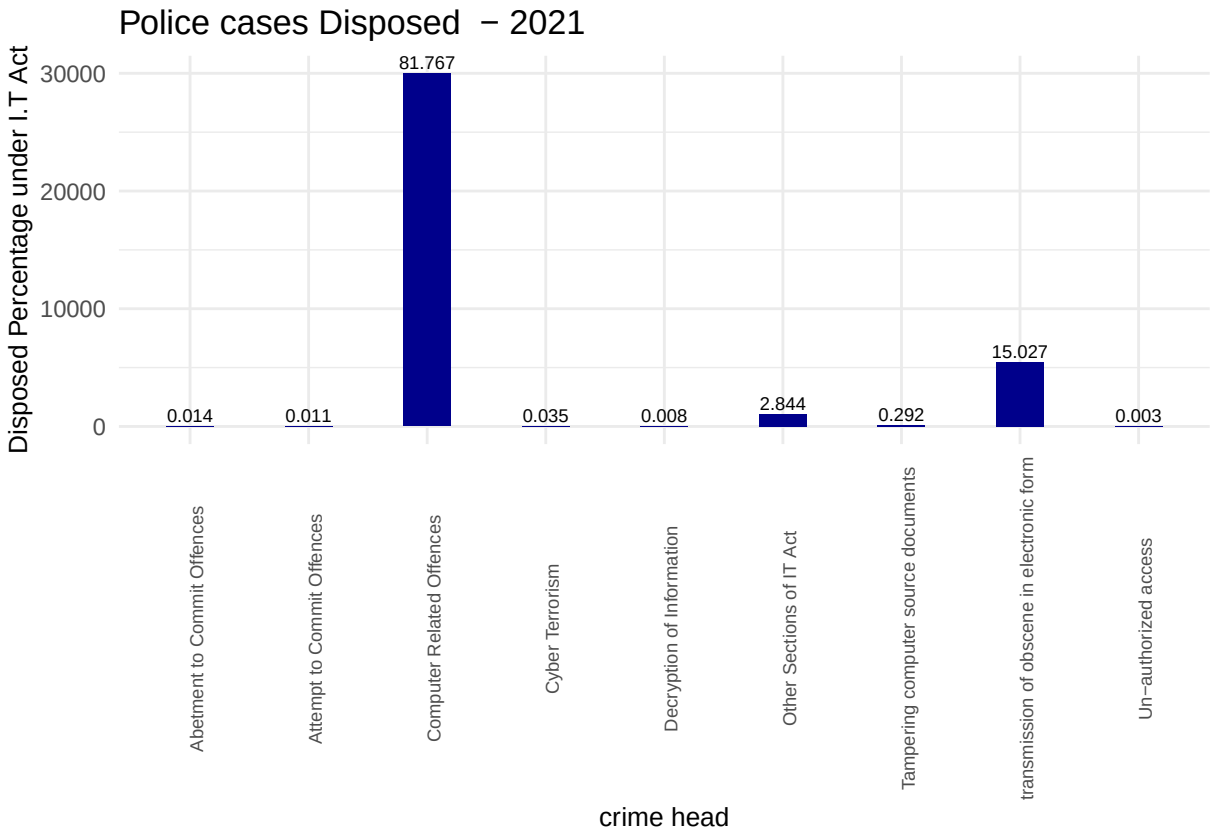


The above figure shows that cases investigated by Police "Under IPC Act" of Fraud are highest with **51.587** %.

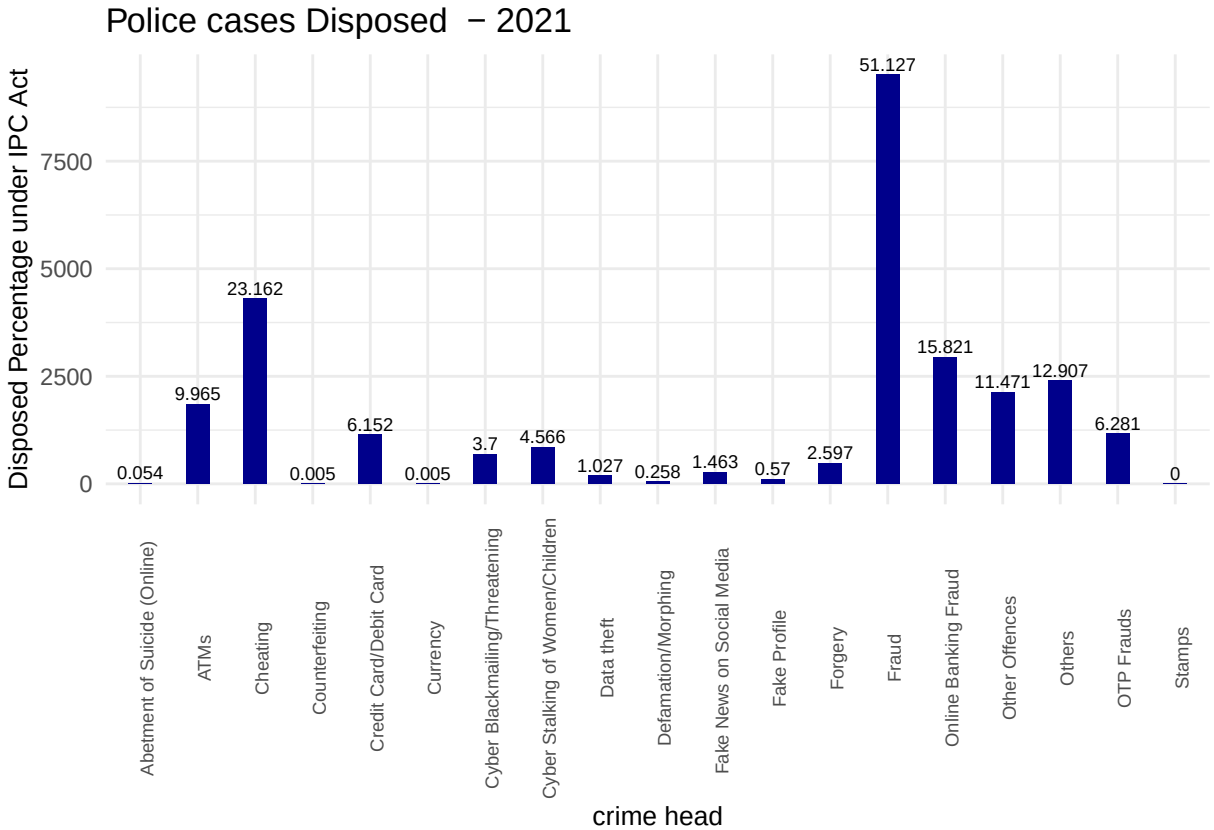


The above figure shows that cases investigated by Police "Under SLL Act" of Copy Right Act are highest with **36.907** %.

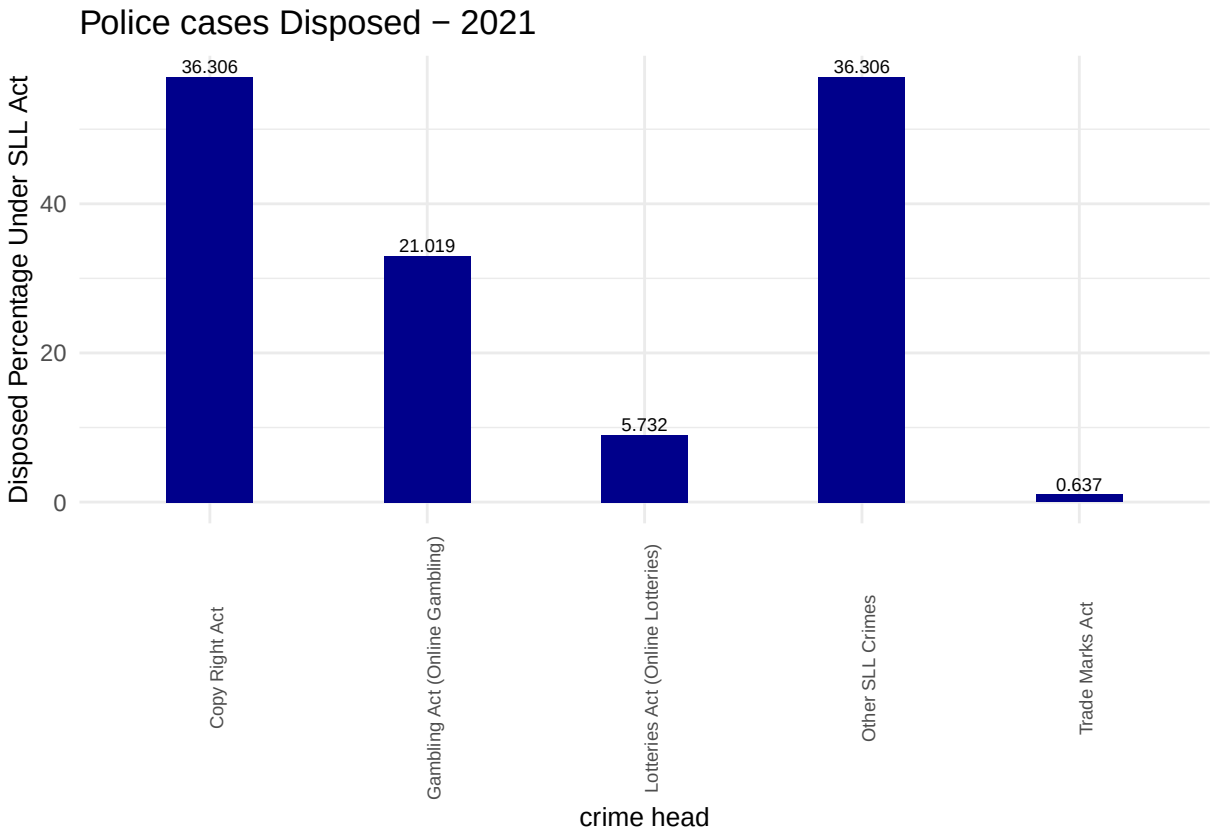
year2021



The above figure shows that cases Disposed under IT Act of computer related offenses are highest with **81.767 %**.

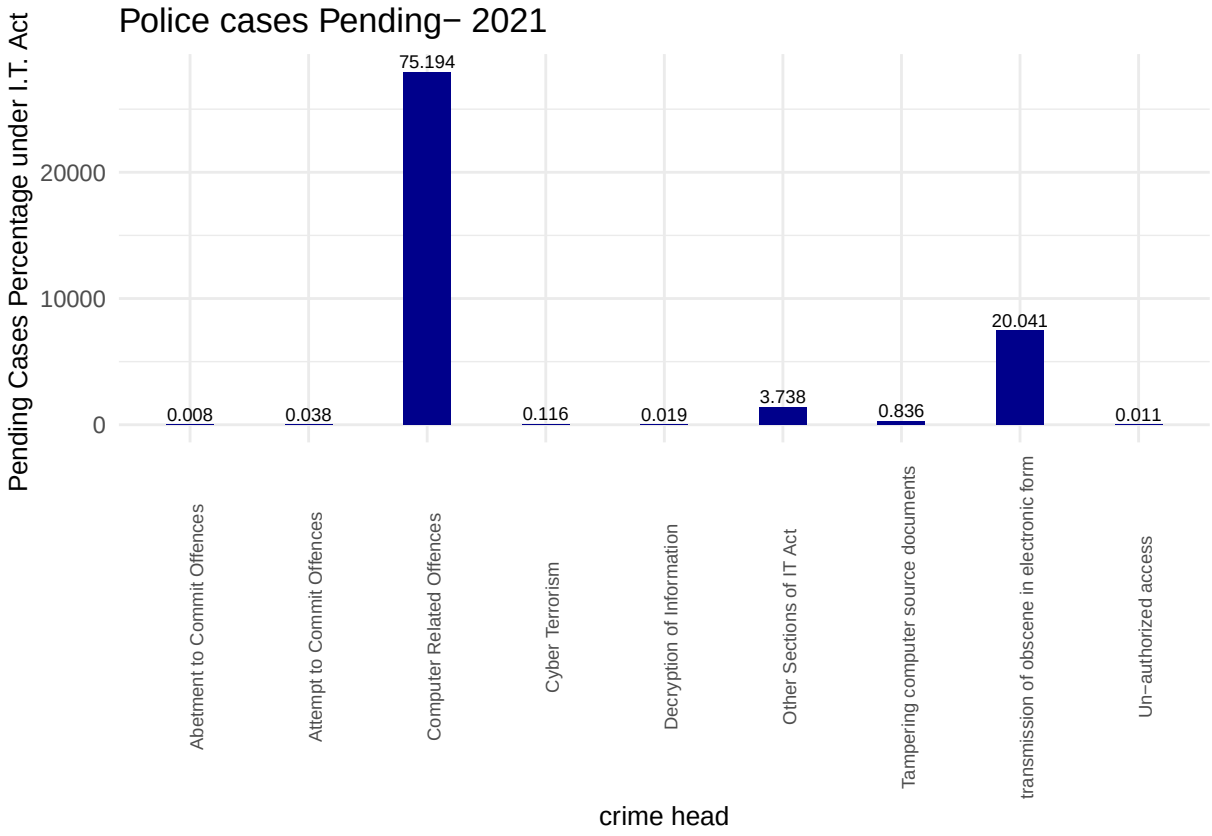


The above figure shows that cases Disposed by Police "Under IPC Act" of Fraud are highest with **51.127 %**.

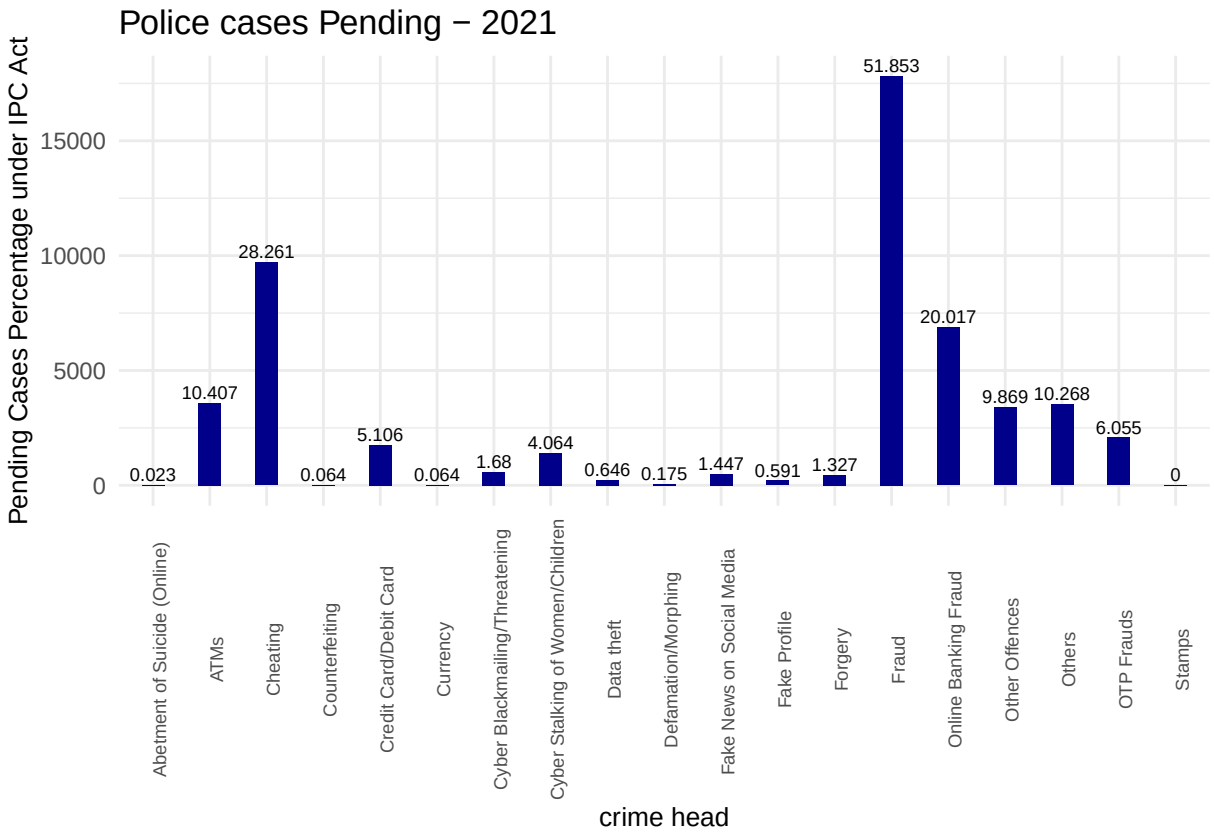


The above figure shows that cases Disposed by Police "Under SLL Act" of Copy Right Act and Other SLL Crimes are highest with **36.306 %**.

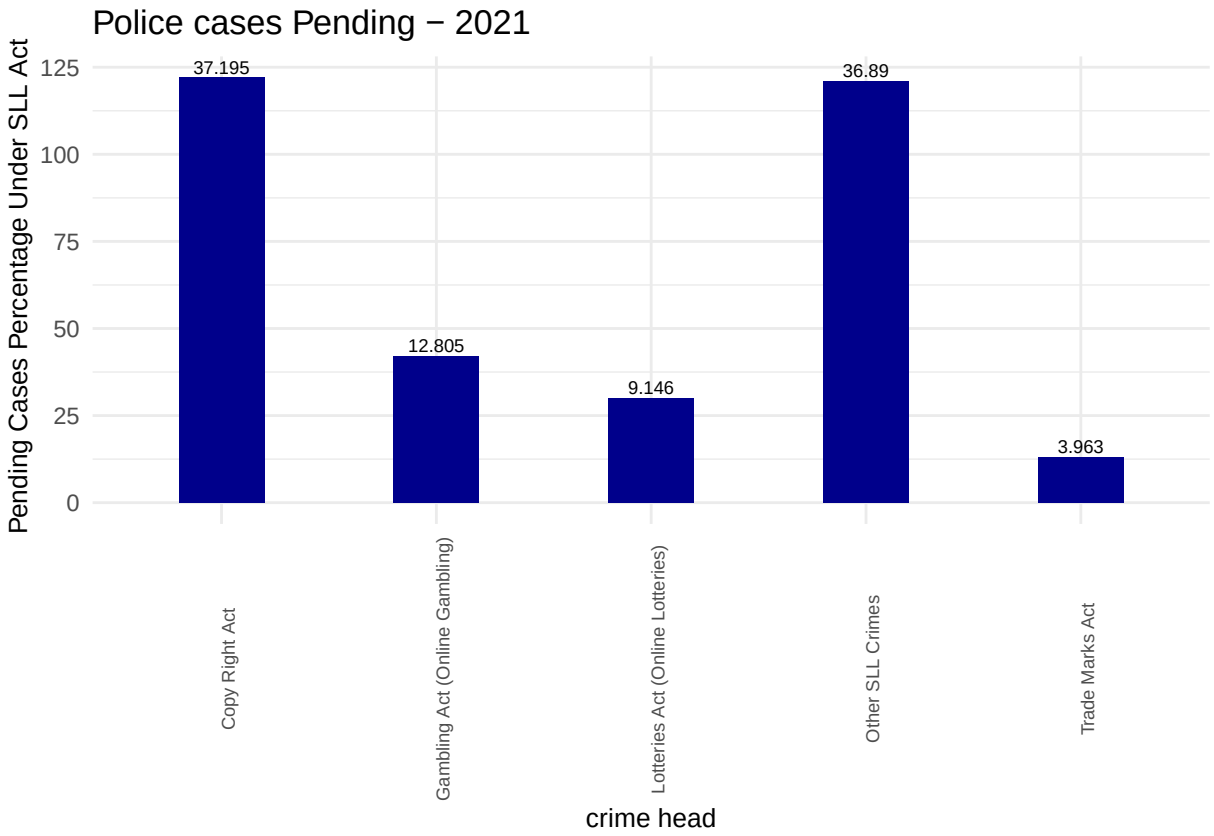
Total Cases Pending year2021



The above figure shows that Cases pending under IT Act of computer related offences are highest with **75.194 %**.

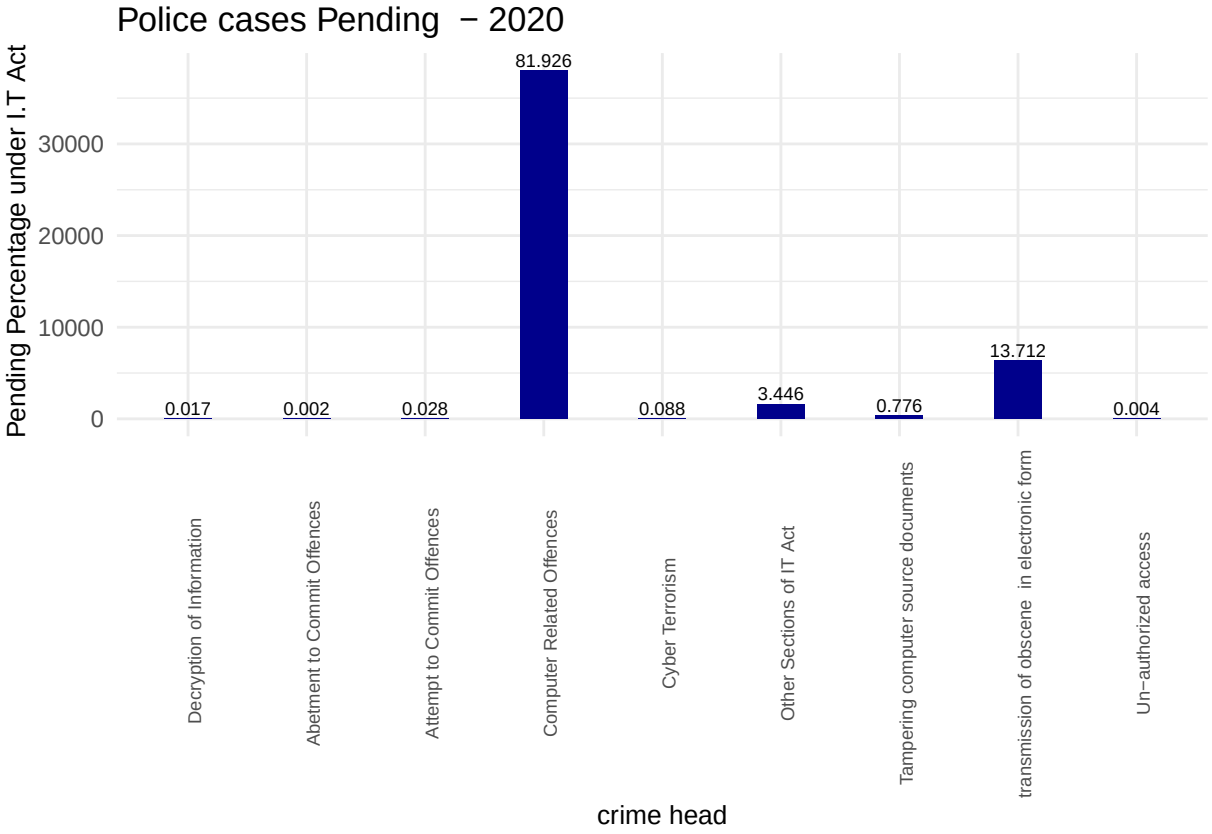


The above figure shows that cases pending by Police "Under IPC Act" of Fraud are highest with **51.853** %.

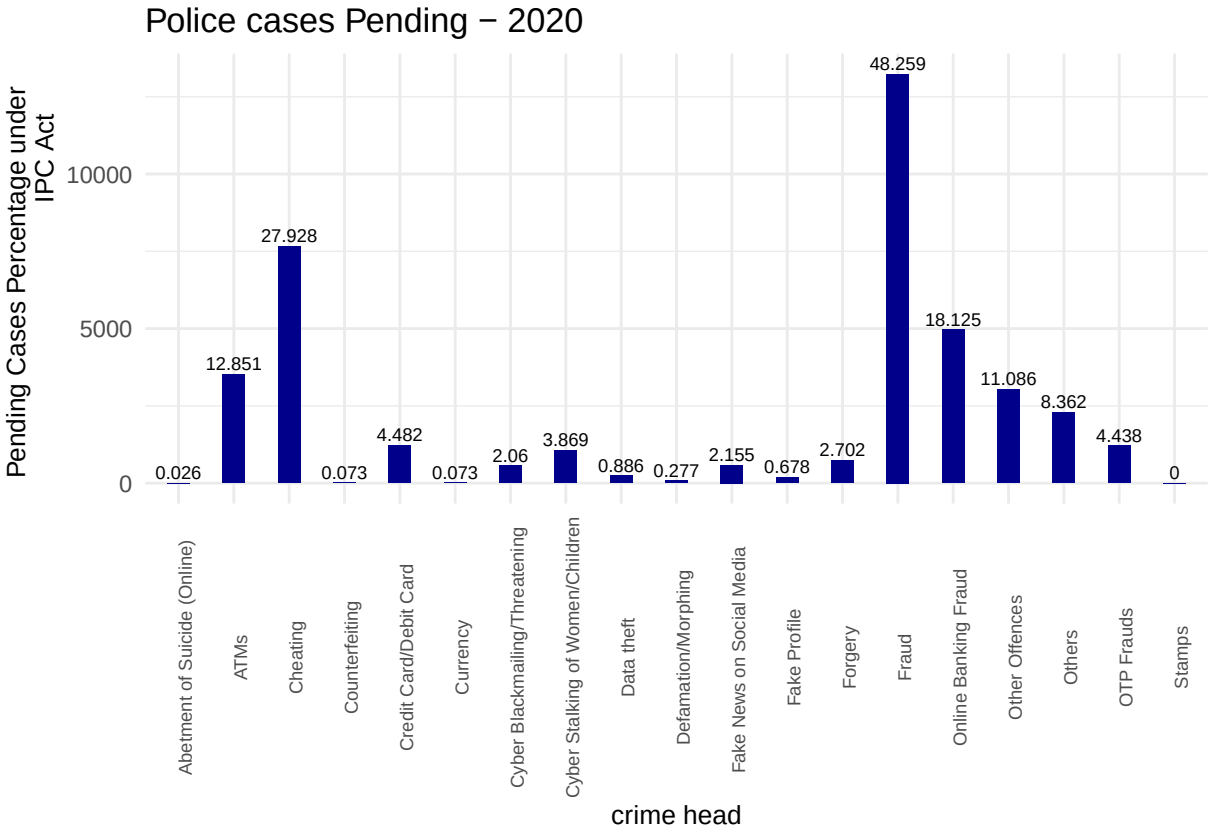


The above figure shows that cases pending by Police "Under SLL Act" of Copy Right Act are highest with **37.195** %.

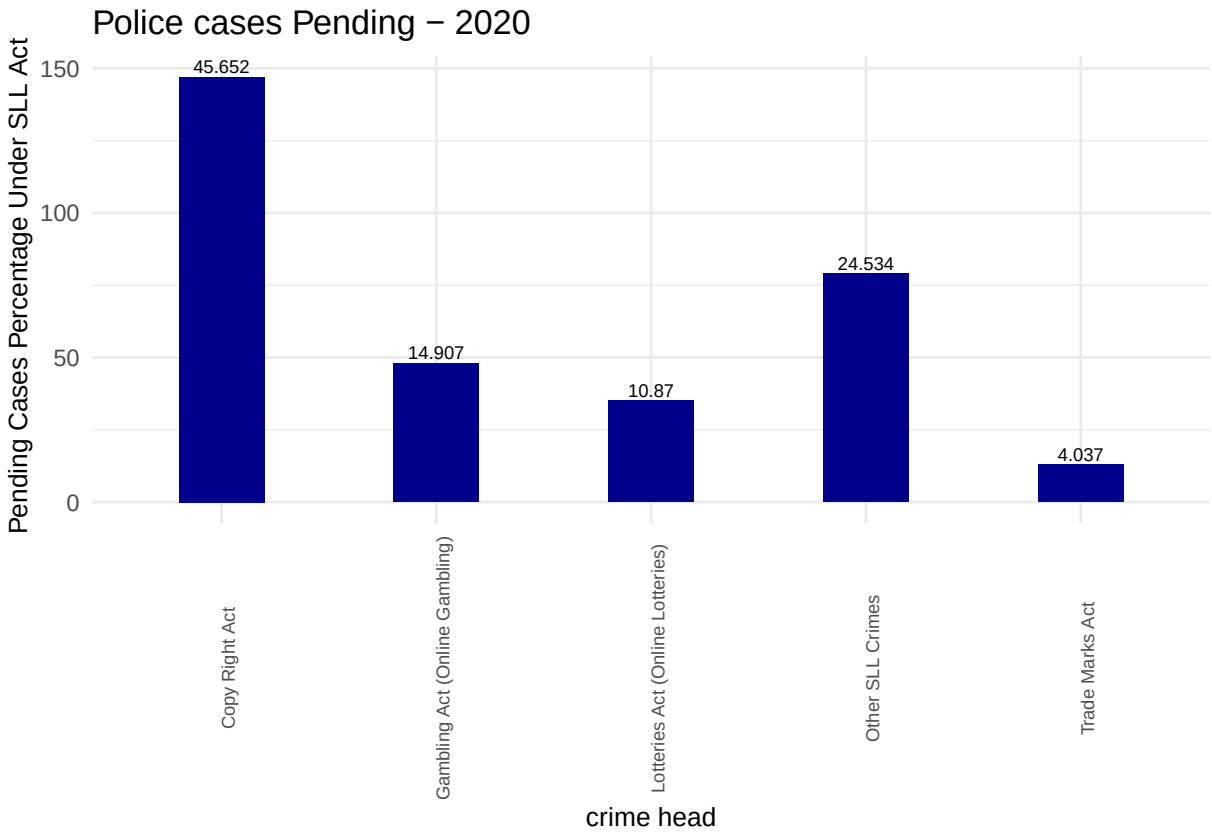
4.3.2 Year2020



The above figure shows that Cases pending under IT Act of computer related offences are highest with **81.926%**.

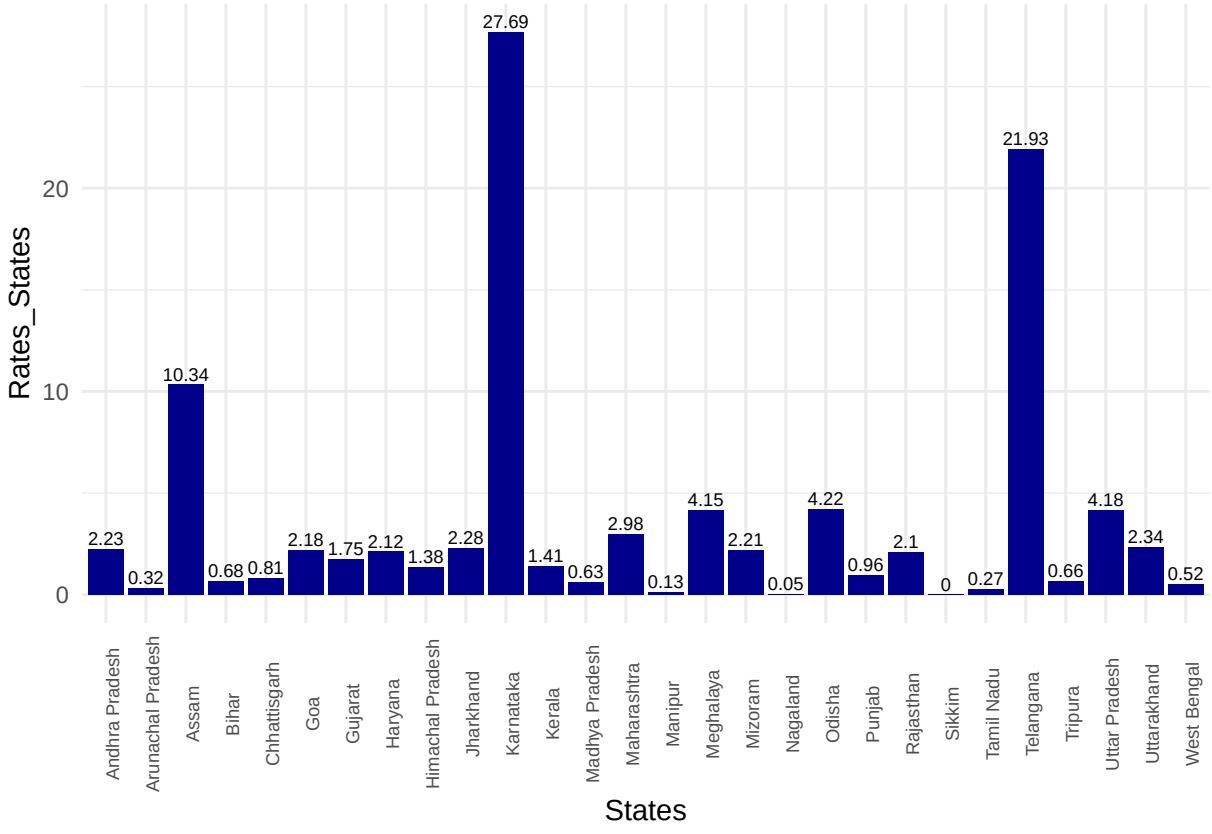


The above figure shows that Cases pending under I.P.C of computer related offences are highest with **48.259 %**.

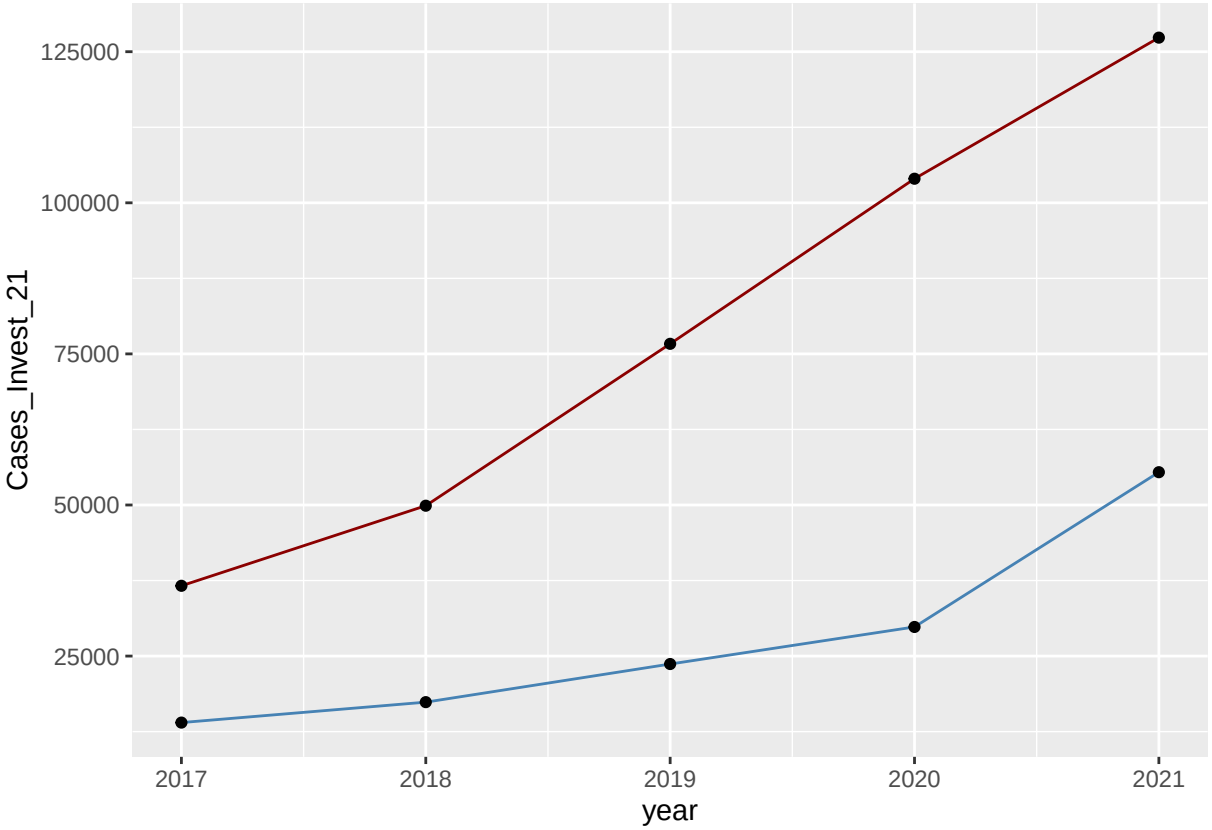


The above figure shows that cases pending by Police "Under SLL Act" of Copy Right Act are highest with **45.652 %**.

4.4 Police Disposal Rate of Cyber Crime Cases (State-wise)



4.5 Total Cyber Crimes for Investigation and their Disposal by Police



Upper line shows Total Cyber Crimes for Investigation and line shows their disposal. This graph shows Total Cyber Crimes for Investigation and their Disposal by Police .

4.5.1 Testing location (crimehead-wise)

```
##
##  Kruskal-Wallis rank sum test
##
## data:  data[, seq(1, 17, 4)]
## Kruskal-Wallis chi-squared = 0.29403, df = 4, p-value = 0.9902
```

Since p-value > 0.05,for Kruskal Wallis test ,the result is not significant .i.e we fail to reject null hypothesis.Therefore there is no significant change in the location throught the years 2017-2021.Hence we conclude that cases investigated crime head wise are statistically same for different years.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  data[, seq(2, 18, 4)]
## Kruskal-Wallis chi-squared = 0.57253, df = 4, p-value = 0.9661
```

Since p-value > 0.05,for Kruskal Wallis test ,the result is not significant .i.e we fail to reject null hypothesis.Therefore there is no significant change in the location throughout the years 2017-2021.Hence we conclude that cases true but insufficient evidence crime head wise are statistically same for different years.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  data[, seq(3, 19, 4)]
## Kruskal-Wallis chi-squared = 0.47127, df = 4, p-value = 0.9762
```

Since p-value > 0.05,for Kruskal Wallis test ,the result is not significant
.i.e we fail to reject null hypothesis.Therefore there is no significant change

in the location throughot the years 2017-2021.Hence we conclude that cases disposed crime head wise are statistically same for different years.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  data[, seq(4, 20, 4)]
## Kruskal-Wallis chi-squared = 0.80944, df = 4, p-value = 0.9372
```

Since p-value > 0.05,for Kruskal Wallis test ,the result is not significant .i.e we fail to reject null hypothesis.Therefore there is no significant change in the location throught the years 2017-2021.Hence we conclude that cases pending crime head wise are statistically same for different years.

4.5.2 Testing variability and location (state-wise)

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   4  1.6938  0.155
##       135
```

Since p-value > 0.05 , for Levene's test hence the test is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant variability in the data through 2017-21
Now testing for equality in location using Kruskal Wallis Test.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  stinvest by as.factor(year)
## Kruskal-Wallis chi-squared = 5.1332, df = 4, p-value = 0.2739
```

Since p-value > 0.05 , for Kruskal Wallis test, the result is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant change in the locations throughout the years 2017-21 Hence, from both the tests,Since there is no significant difference in location and variability in different years. Hence we conclude that cases investigated are statistically same in terms of location and variability for different crimes.

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   4  1.8592 0.1212
##       135
```

Since p-value > 0.05 , for levene's test hence the test is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant variability in the data through 2017-21 Now testing for equality in location using Kruskal Wallis Test

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sttrue by as.factor(year)
## Kruskal-Wallis chi-squared = 2.7508, df = 4, p-value = 0.6004
```

Since p-value > 0.05 , for Kruskal Wallis test, the result is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant change in the locations throughout the years 2017-21 Hence, from both the tests,Since there is no significant difference in location and variability in different years. Hence we conclude that cases true but insufficient evidence are statistically same in terms of location and variability for different crimes.

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   4  1.8592 0.1212
##       135
```

Since p-value > 0.05 , for levene's test hence the test is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant variability in the data through 2017-21

```
##
##  Kruskal-Wallis rank sum test
##
## data:  stdisposed by as.factor(year)
## Kruskal-Wallis chi-squared = 2.7508, df = 4, p-value = 0.6004
```

Since p-value > 0.05 , for Kruskal Wallis test, the result is not significant i.e. we failed to reject null hypothesis. Therefore there is no significant change in the locations throughout the years 2017-21 Hence, from both the tests,Since there is no significant difference in location and variability in different years. Hence we conclude that cases Disposed are statistically same in terms of location and variability for different crimes.

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   4  1.8592 0.1212
##       135
```

Since p-value > 0.05 , for levene's test hence the test is not significant
i.e. we failed to reject null hypothesis. Therefore there is no significant variability in the data through 2017-21

```
##
##  Kruskal-Wallis rank sum test
##
## data:  stpending by as.factor(year)
## Kruskal-Wallis chi-squared = 2.7508, df = 4, p-value = 0.6004
```

Since p-value > 0.05 , for Kruskal Wallis test, the result is not significant
i.e. we failed to reject null hypothesis. Therefore there is no significant
change in the locations throughout the years 2017-21 Hence, from both the tests,Since there is no significant difference in location and variability in different years. Hence we conclude that cases pending are statistically same in terms of location and variability for different crimes.

4.5.3 Paired Wilcoxon signed rank test

Wilcoxon signed rank test
The Wilcoxon test is a non parametric alternative to the t-test for comparing two means
The goal of the test is to determine if two or more sets of pairs are different from one another in a statistically significant manner
Model assume the pairs in the data come from dependent populations

```
##
##  Wilcoxon signed rank exact test
##
## data:  stdata$Cases_Invest_17 and stdata$Cases_Invest_21
## V = 0, p-value = 7.451e-09
## alternative hypothesis: true location shift is not equal to 0
```

The p-value < 0.05. We conclude that the cases investigated in year 2017 are significantly different from cases investigated in year 2021.

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  stdata$Cases_Disposed_17 and stdata$Cases_Disposed_21
## V = 0, p-value = 5.934e-06
## alternative hypothesis: true location shift is not equal to 0
```

The p-value < 0.05. We conclude that the cases Disposed in year 2017 are significantly different from cases Disposed in year 2021.

```
##
##  Wilcoxon signed rank exact test
##
## data:  stdata$Cases_Pending_17 and stdata$Cases_Pending_21
```

```
## V = 0, p-value = 7.451e-09
## alternative hypothesis: true location shift is not equal to 0
```

The p-value < 0.05. We conclude that the cases pending in year 2017 are significantly different from cases pending in year 2021.

```
wilcox.test( stdata$True_cases.but.insufficient.evidence_17, stdata$True_cases.but.insufficient.evid
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  stdata$True_cases.but.insufficient.evidence_17 and stdata$True_cases.but.insufficient.evid
## V = 23.5, p-value = 7.358e-05
## alternative hypothesis: true location shift is not equal to 0
```

The p-value < 0.05. We conclude that the cases true but insufficient evidence in year 2017 are significantly different from cases true but insufficient evidence in year 2021.

4.6 Hierarchical Cluster Analysis

In Hierarchical Cluster Analysis,the observations are grouped together on the basis of their mutual distances.It is visualized through a hierarchical tree,called dendrogram tree.

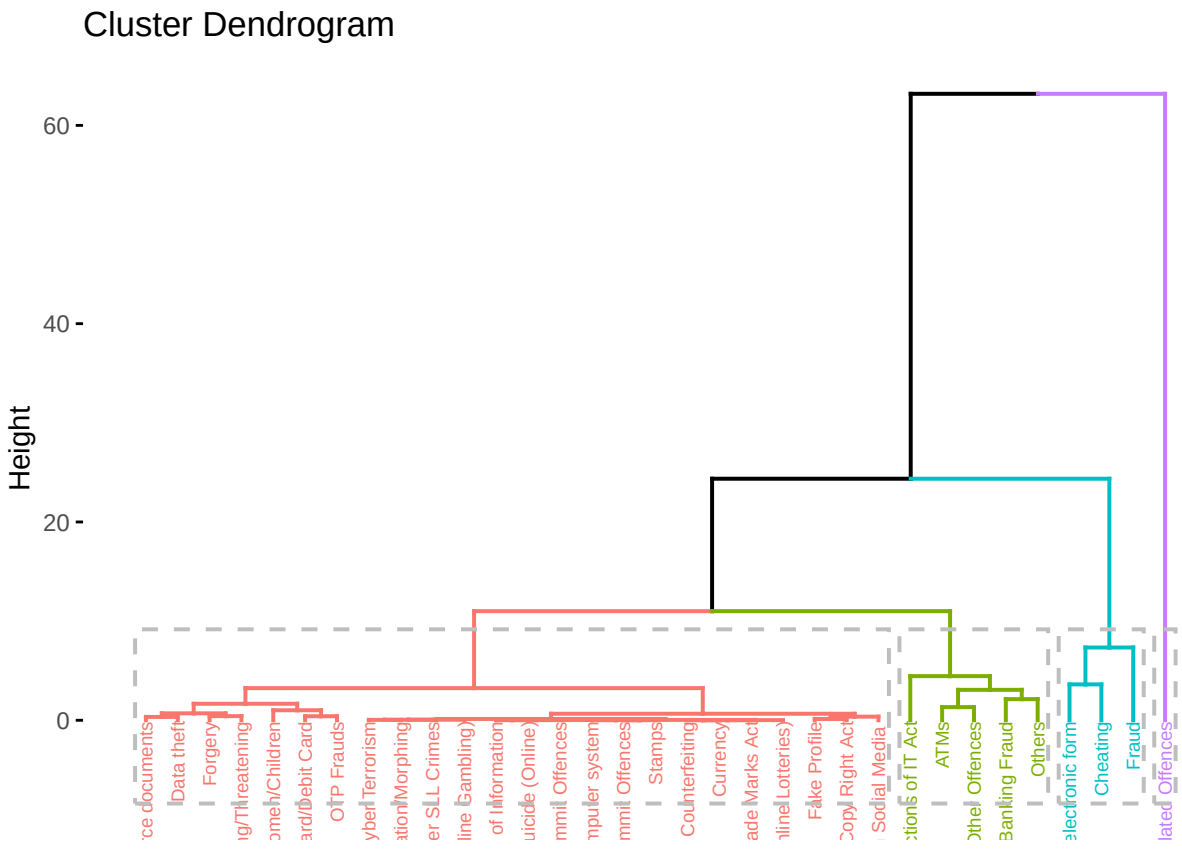
4.6.1 Agglomerative HCA:

Algorithm: Operates by successive merger of cases.

1. Begin with N clusters each containing single cases.

2. At each stage, we merge the two most similar groups to form a new cluster.Thus reducing the no. of clusters to (N-1)

3. Continue till all subgroups are fuse to form a single cluster.

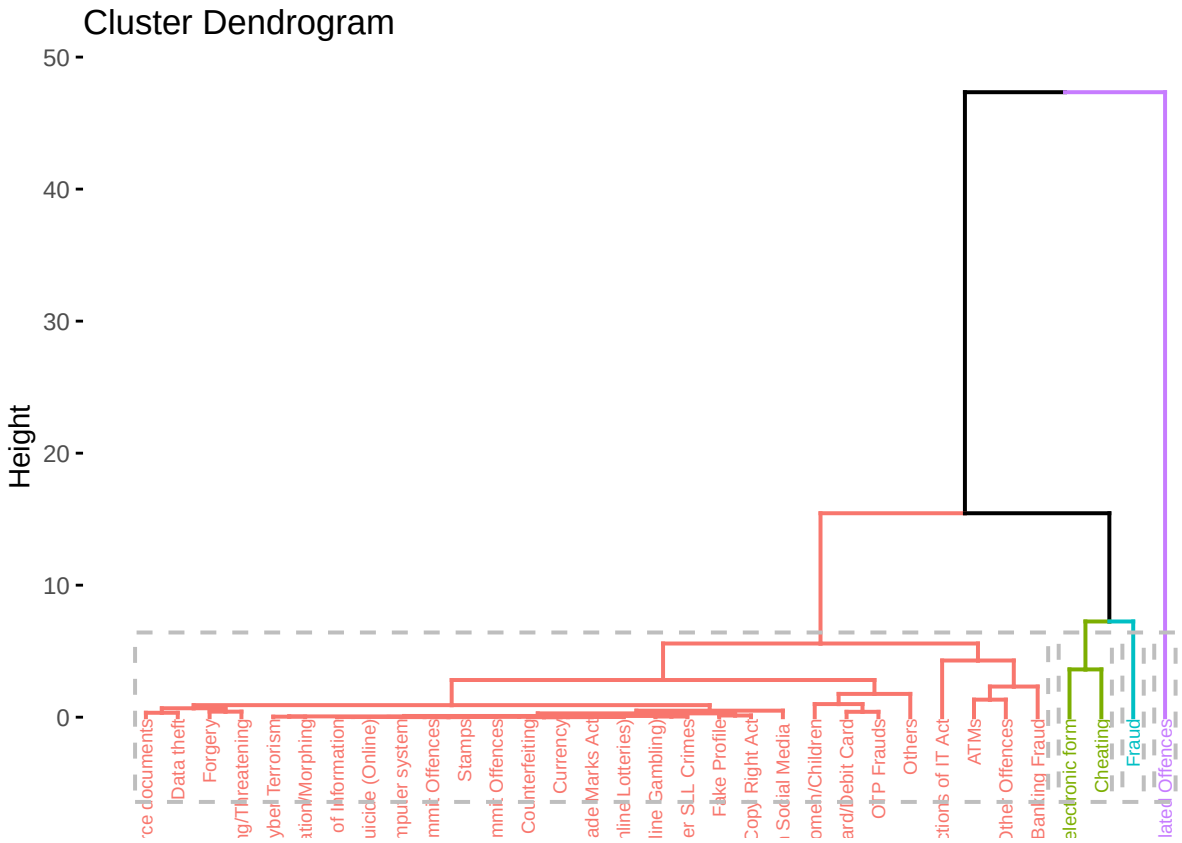


At a particular Distance level of 20 units, we can observe that electronic form, Cheating ,fraud,computer related offence form one cluster, however the crimes with lesser number are in one cluster.

4.6.2 Divisive HCA

1. Divisive method operates by successive splitting of group.
2. Initially starts with a single group(i.e.one single cluster)
3. Group is divided into two groups
4. Continue till there are n groups each with a single object.

On the other hand if Divisive HCA is applied at a particular distance 10 there will be two clusters.
The major four crimes in one cluster and other in second cluster.



Chapter 5

Cyber Crimes against Women and Children

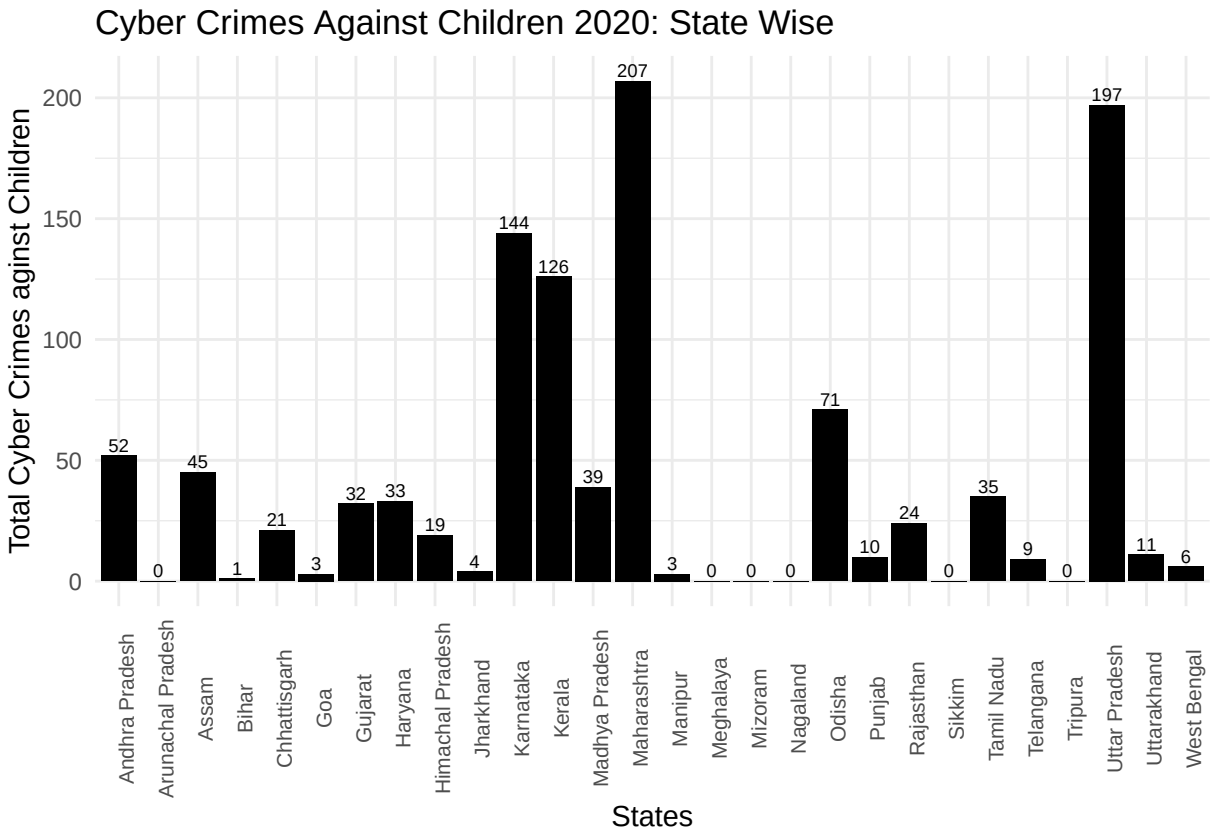
5.1 Introduction

In this section, the main interest is to carry out statistical analysis of the Cyber Crimes happening particularly against women and children.
The major cyber crimes against women for which the data is collected by NCRB include: Cyber Blackmailing/Threatening, Cyber Bullying/Stalking, Cyber Pornography, Fake Profiles, Defamation/Morphing etc.
The major Cyber Crimes against children for which the data is collected by NCRB are: Cyber Blackmailing/Threatening, Cyber Bullying/Stalking, Cyber Child Pornography, Fake Profiles etc.

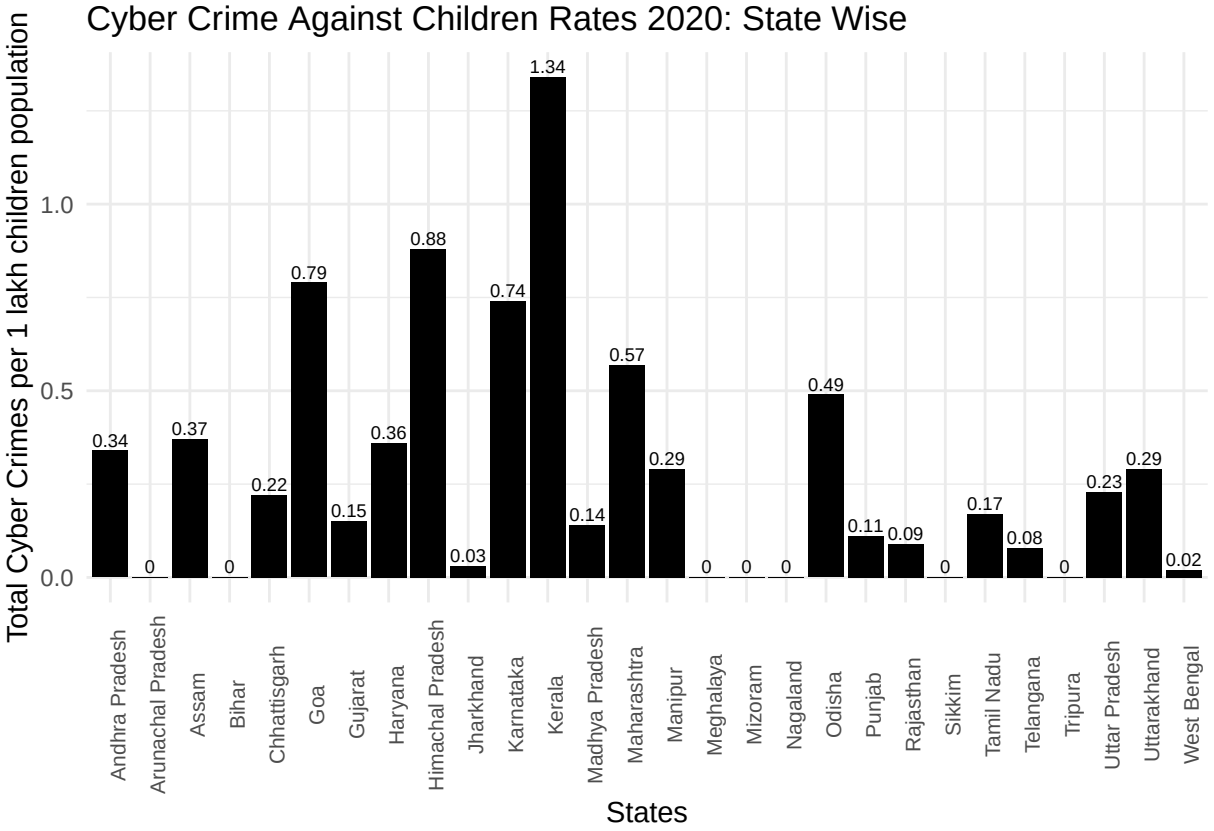
5.2 Cyber Crimes against Children

5.2.1 Year 2020

State-Wise



The states of Maharashtra, Uttar Pradesh, Karnataka, Kerala registered higher cyber-crimes against children in the year 2020.



However, in terms of rates, Kerala has the maximum cyber crime rate against children in the year 2020

5.2.1.1 Runs Test for randomness

```
##
##  Runs Test
##
## data:  child20$rate[1:28]
## statistic = 1.1555, runs = 18, n1 = 14, n2 = 14, n = 28, p-value =
## 0.2479
## alternative hypothesis: nonrandomness
```

Since, p-value > 0.05, the results are coming out be non-significant. Thus, we conclude that the cyber-crime rates against children throughout the states is random.

5.2.1.2 Komlogorov-Smirnov Test

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  child20$rate[1:28] and child21$rate[1:28]
## D = 0.10714, p-value = 0.9927
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  child17$rate[-c(10, 30:39)] and child21$rate[1:28]
## D = 0.60714, p-value = 1.732e-05
## alternative hypothesis: two-sided
```

```
##
```

```
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  child18$rate[-c(10, 30:39)] and child21$rate[1:28]
## D = 0.42857, p-value = 0.009589
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  child19$rate[-c(10, 30:39)] and child21$rate[1:28]
## D = 0.39286, p-value = 0.01865
## alternative hypothesis: two-sided
```

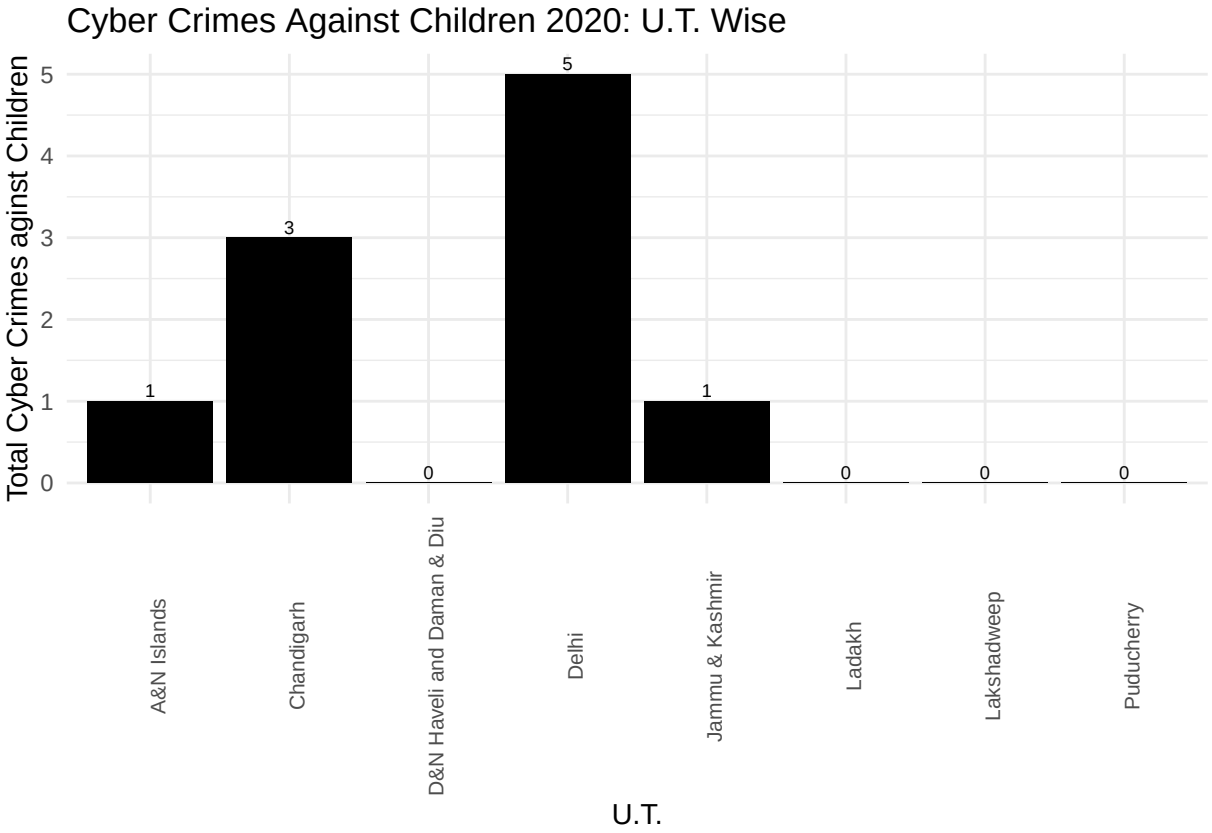
```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  child19$rate[-c(10, 30:39)] and child20$rate[1:28]
## D = 0.39286, p-value = 0.01988
## alternative hypothesis: two-sided
```

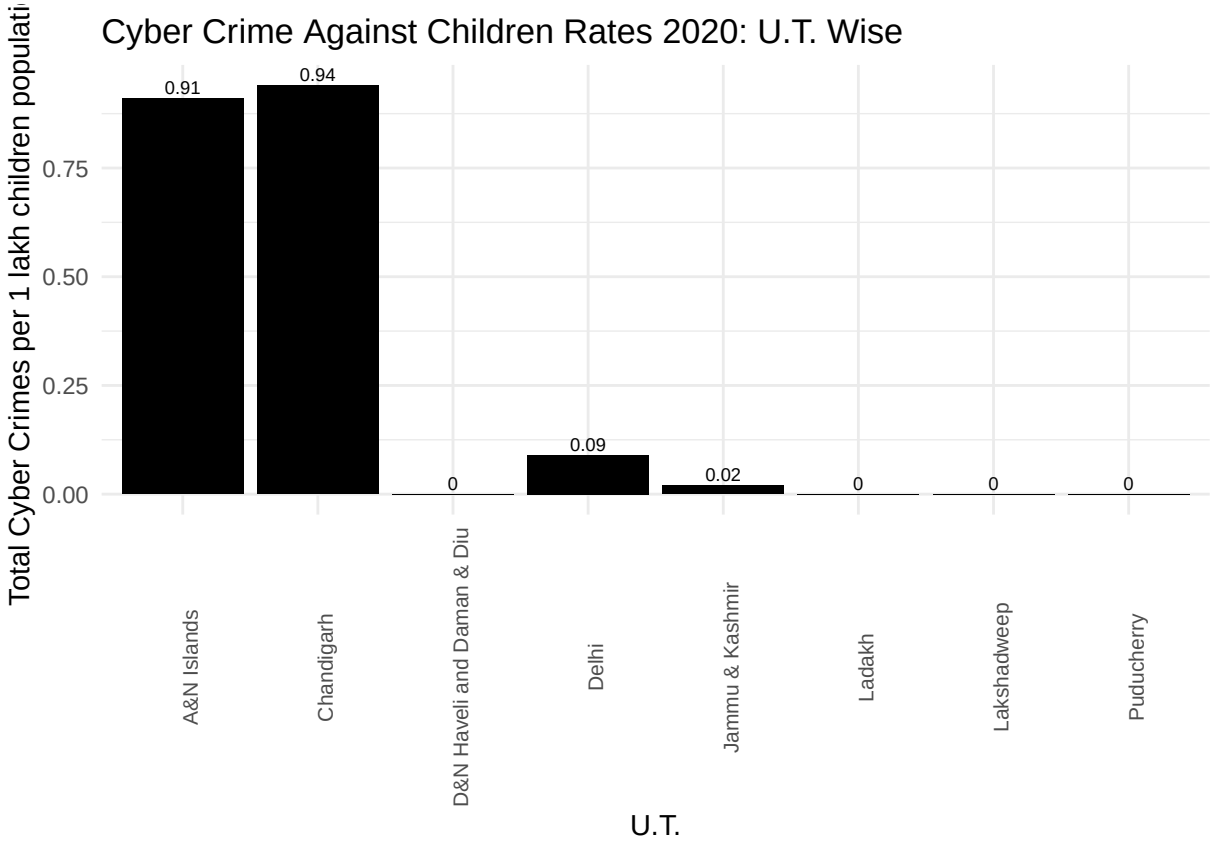
The distribution of cyber-crime rates against children is same for the years 2020 and 2021 throughout the states, however it differs significantly between the years 2021 amd 2017, 2018, 2019.

The distrubution of cases also differs significantly for the consecutive years 2019 and 2020 at 5% level of significance.

U.T. Wise

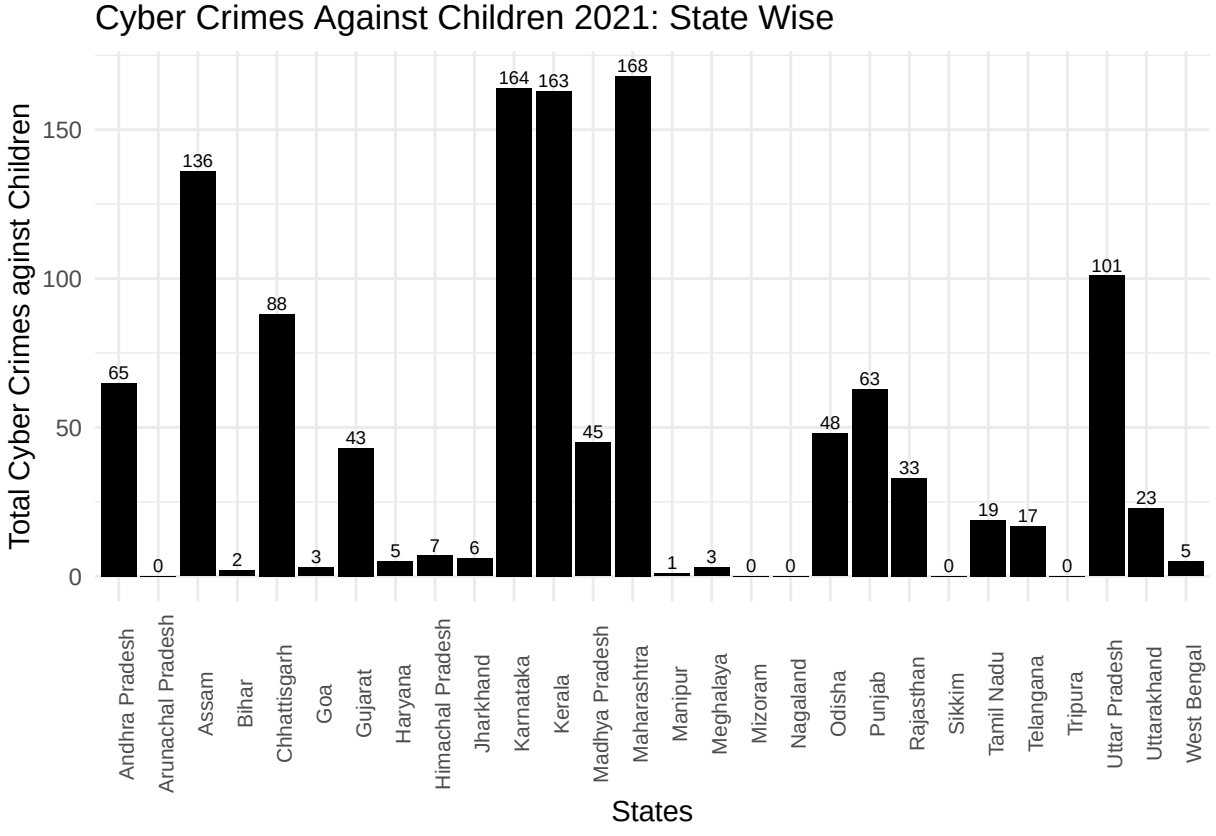
Data Visualization can be done for the Union Territories as follows:





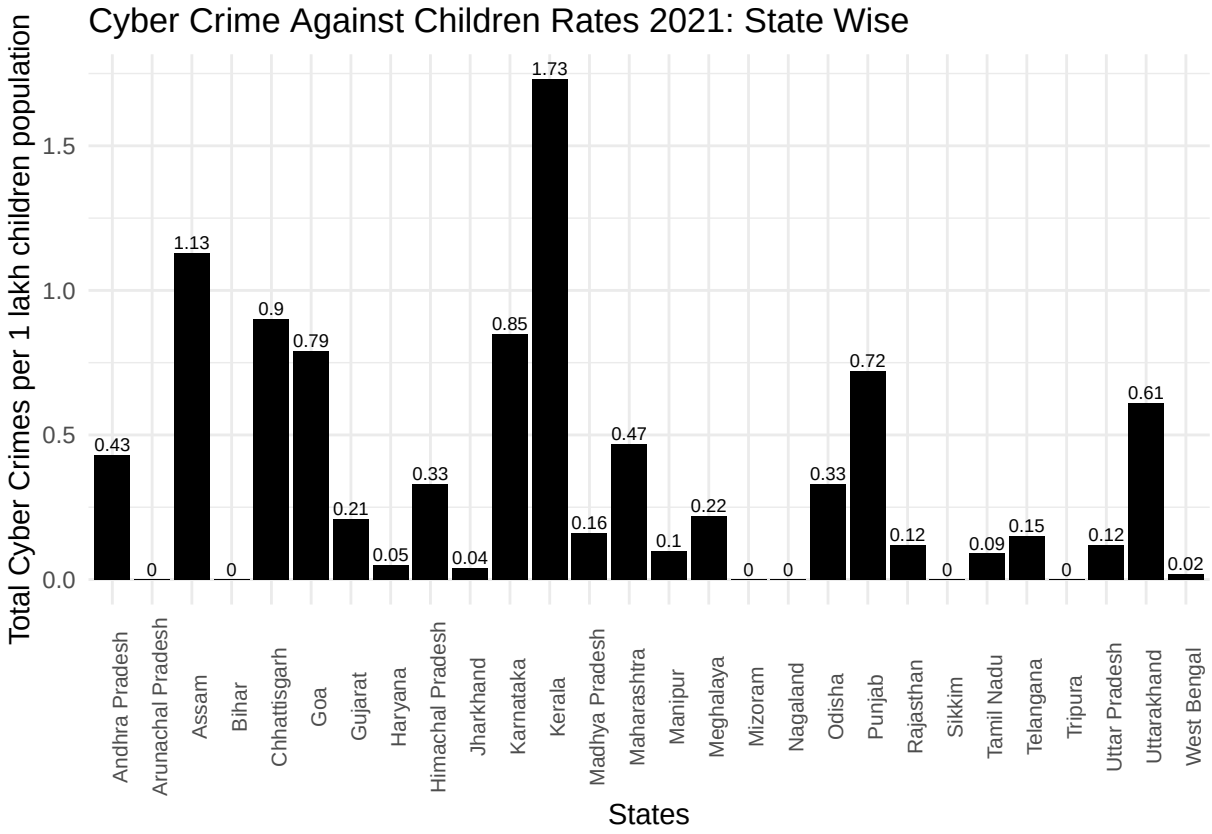
5.2.2 Year 2021

State-Wise



Maharashtra, Karnataka, Kerala, Assam and Uttar Pradesh have comparatively higher cyber crimes against children in

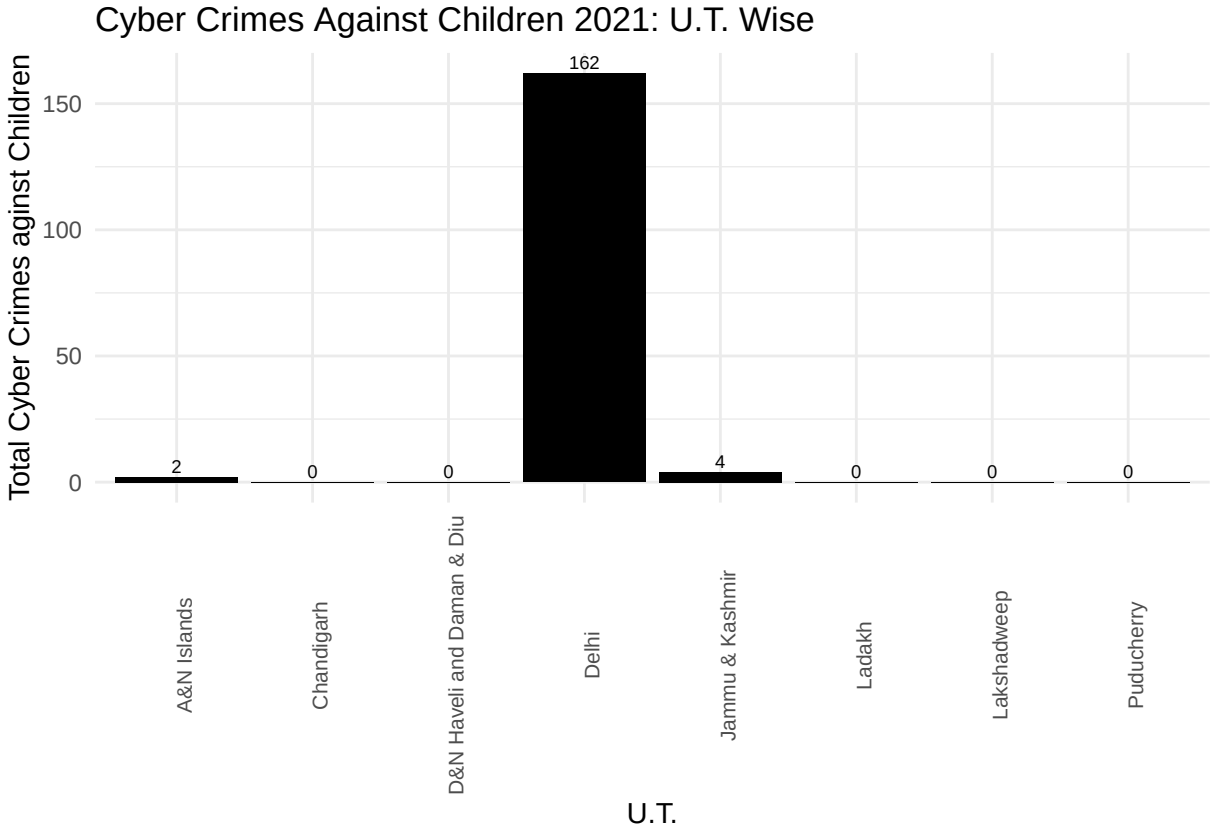
2021.



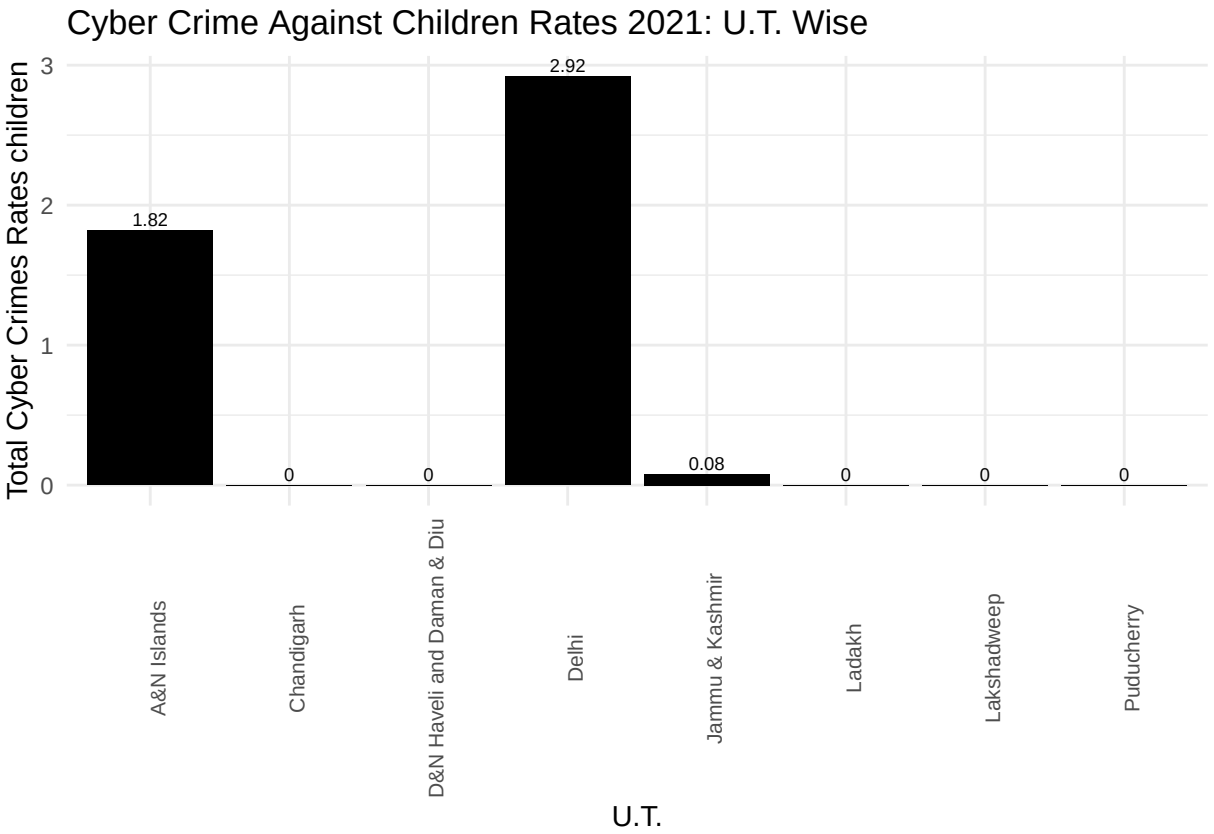
However, in terms of rates, the state of Kerala again has the maximum cyber crime rate against children.

U.T. Wise

Data for the Union Territories can also be viewed as folows:



A huge increase in the cyber crimes against children can be noticed in Delhi as compared to previous year.



5.3 Cyber Crimes against Women

5.3.0.1 Komlogorov-Smirnov Test

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  women20$rate[1:28] and women21$rate[1:28]
## D = 0.21429, p-value = 0.5204
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  women17$rate[-c(10, 30:39)] and women21$rate[1:28]
## D = 0.46429, p-value = 0.003746
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  women18$rate[-c(10, 30:39)] and women21$rate[1:28]
## D = 0.35714, p-value = 0.05419
## alternative hypothesis: two-sided
```

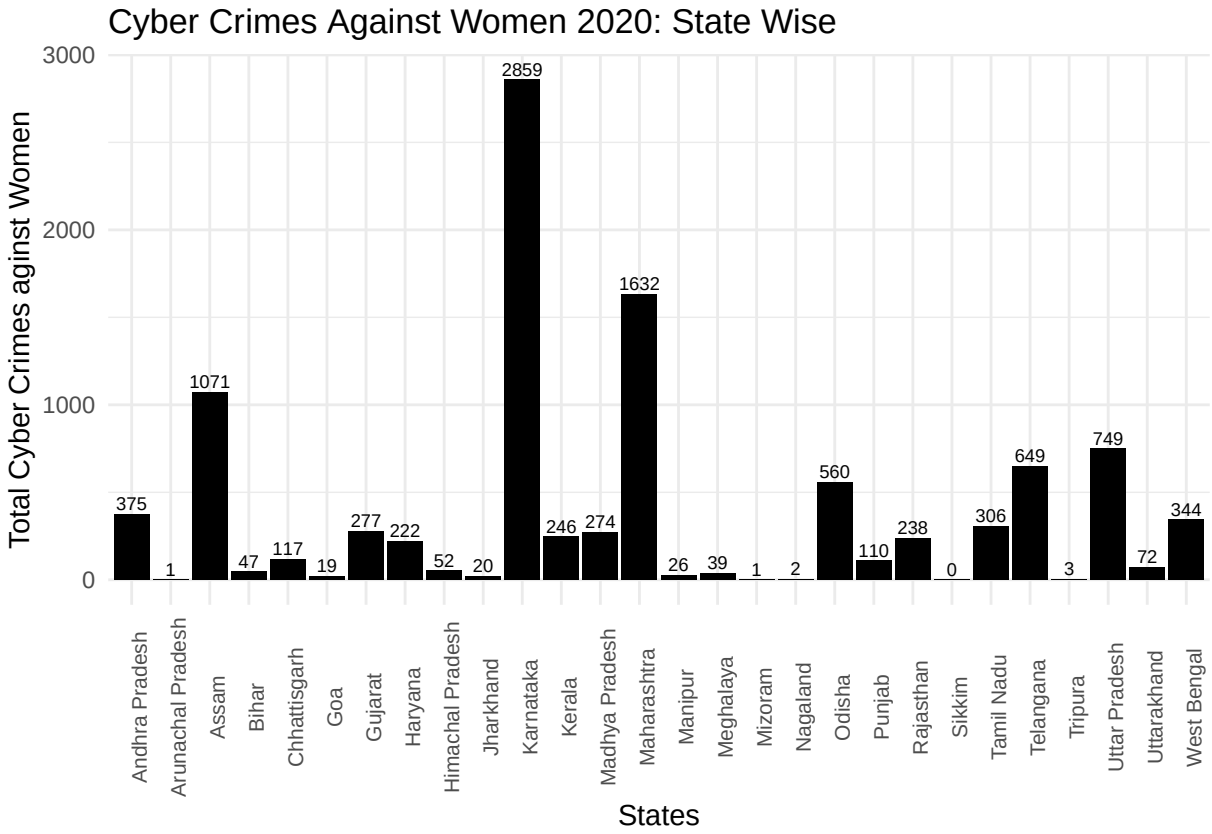
```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  women19$rate[-c(10, 30:39)] and women21$rate[1:28]
## D = 0.39286, p-value = 0.02504
## alternative hypothesis: two-sided
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  women19$rate[-c(10, 30:39)] and women20$rate[1:28]
## D = 0.28571, p-value = 0.1902
## alternative hypothesis: two-sided
```

The distribution of cyber crime rates against women throughout the states is same for the years 2020 and 2021. However, the distribution differs significantly for the years 2017 and 2021.

5.3.1 Year 2020

State-Wise

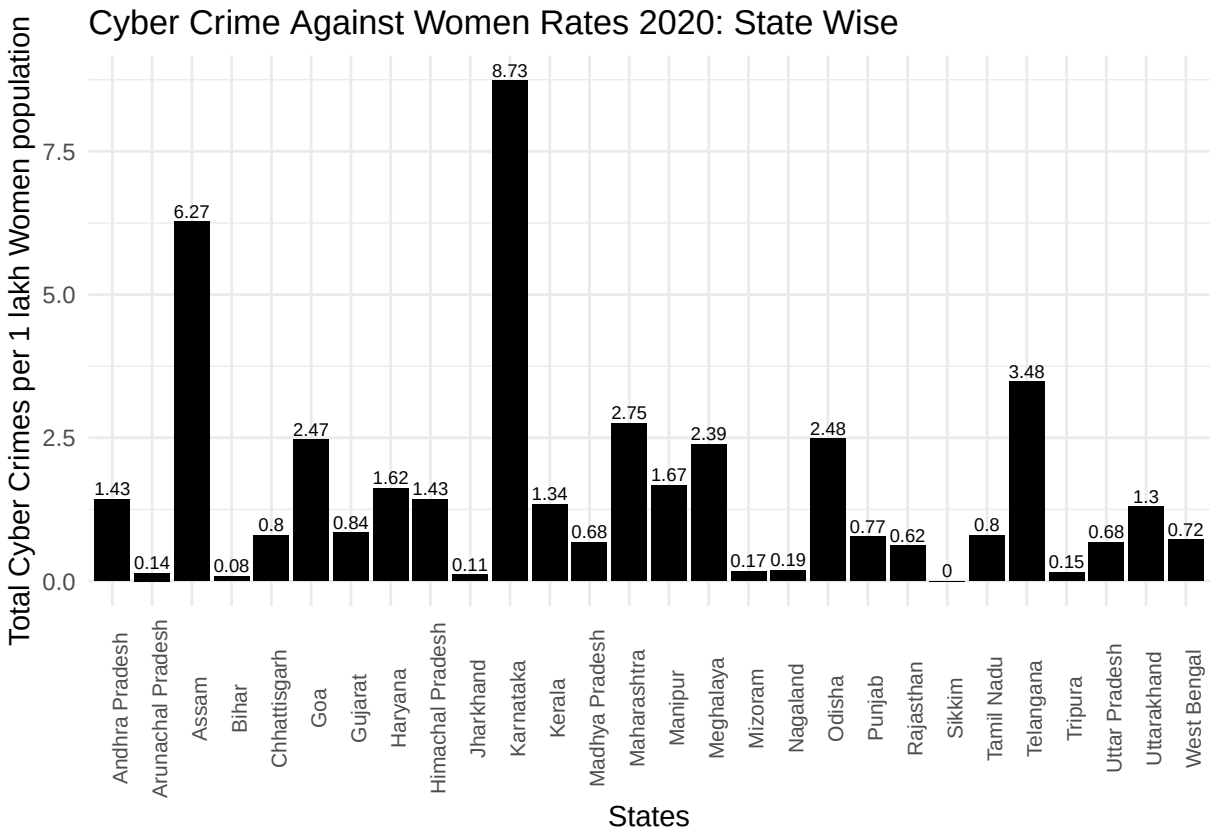


The states of Karnataka, Maharashtra, Assam have comparatively higher cyber crimes against women.

5.3.1.1 Runs Test for randomness

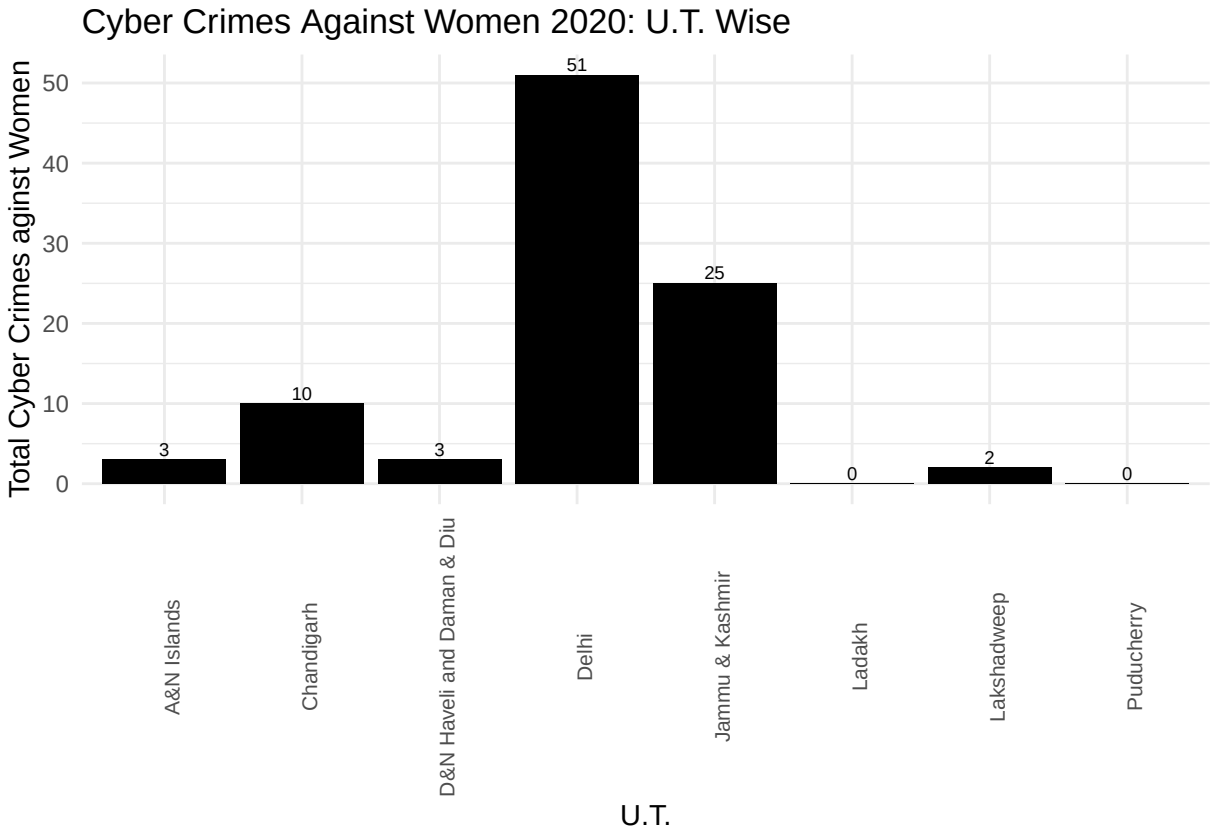
```
##
##  Runs Test
##
## data:  women20$rate[1:28]
## statistic = 0.38516, runs = 16, n1 = 14, n2 = 14, n = 28, p-value =
## 0.7001
## alternative hypothesis: nonrandomness
```

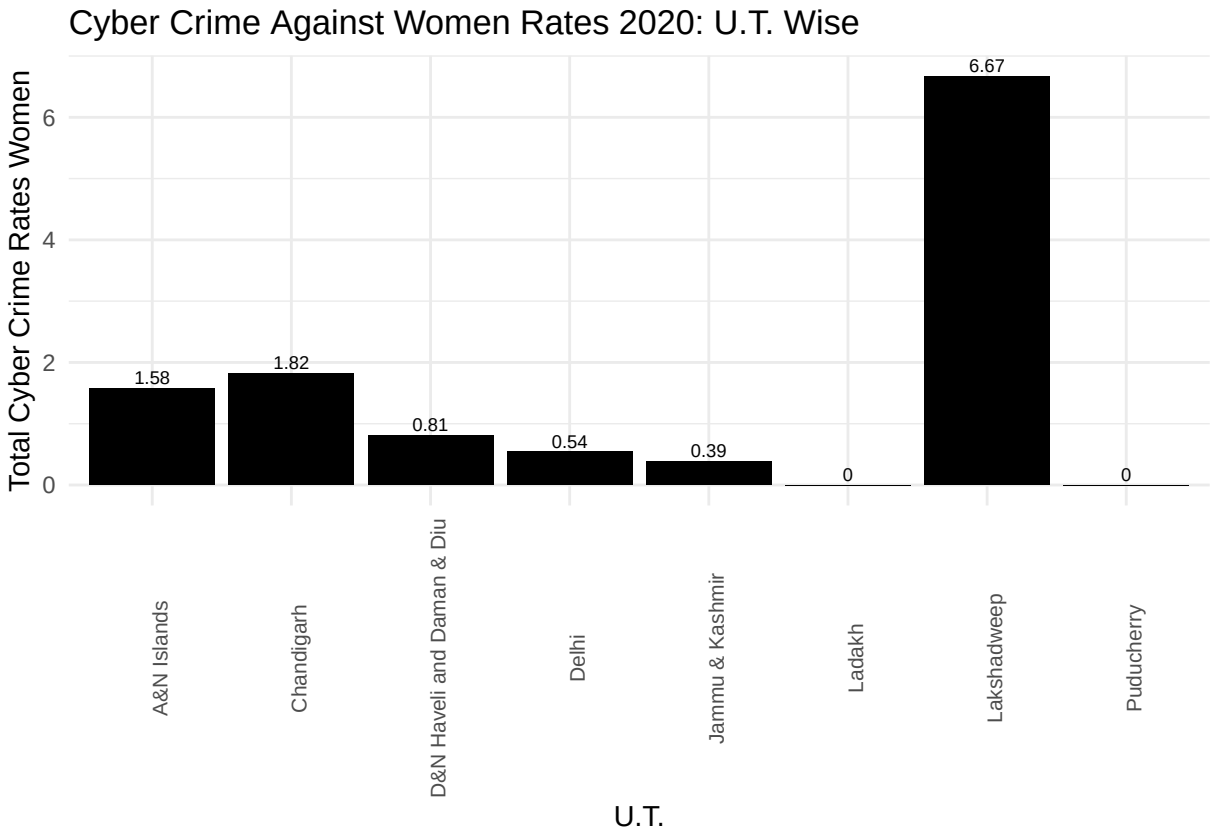
Since, p-value > 0.05, the results are coming out be non-significant. Thus, we conclude that the cyber-crime rates against women throughout the states is random.



In terms of rates, Karnataka has the maximum 8.73 cyber crimes against women followed by Assam and Telangana.

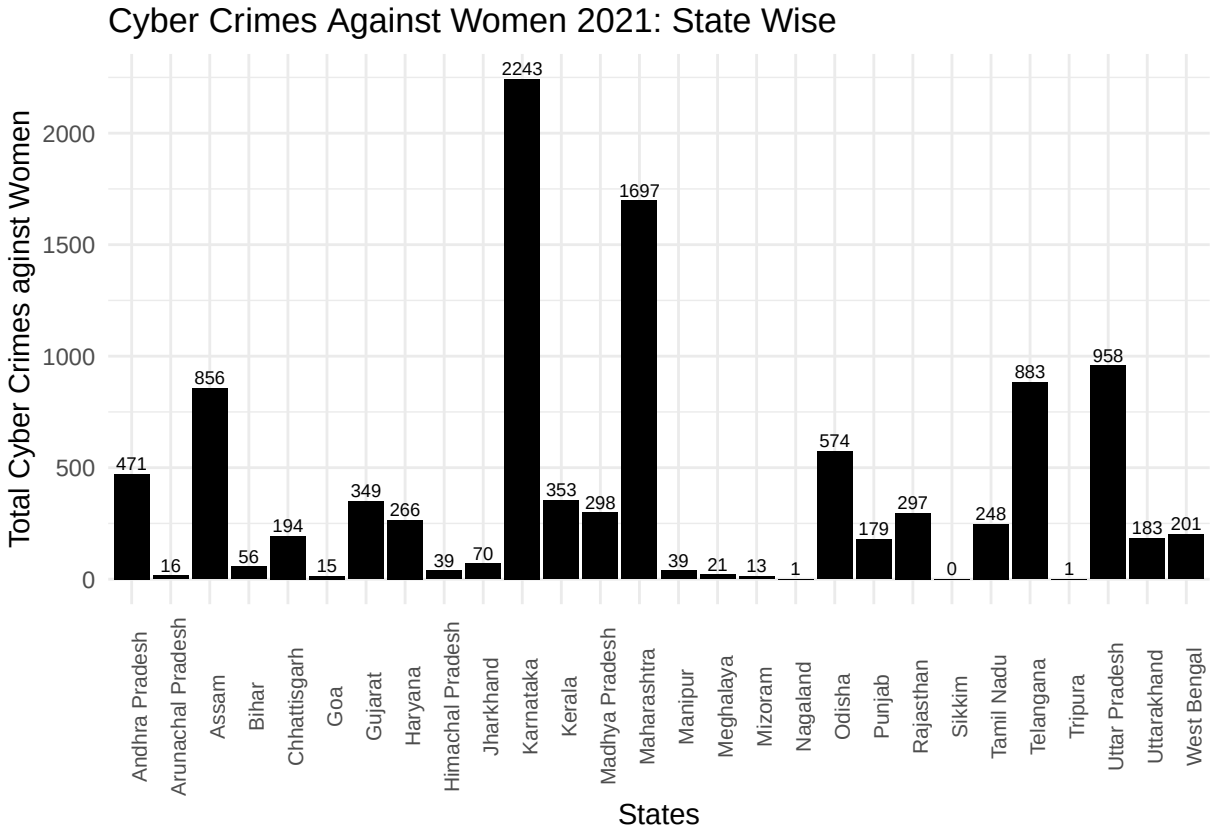
U.T. Wise



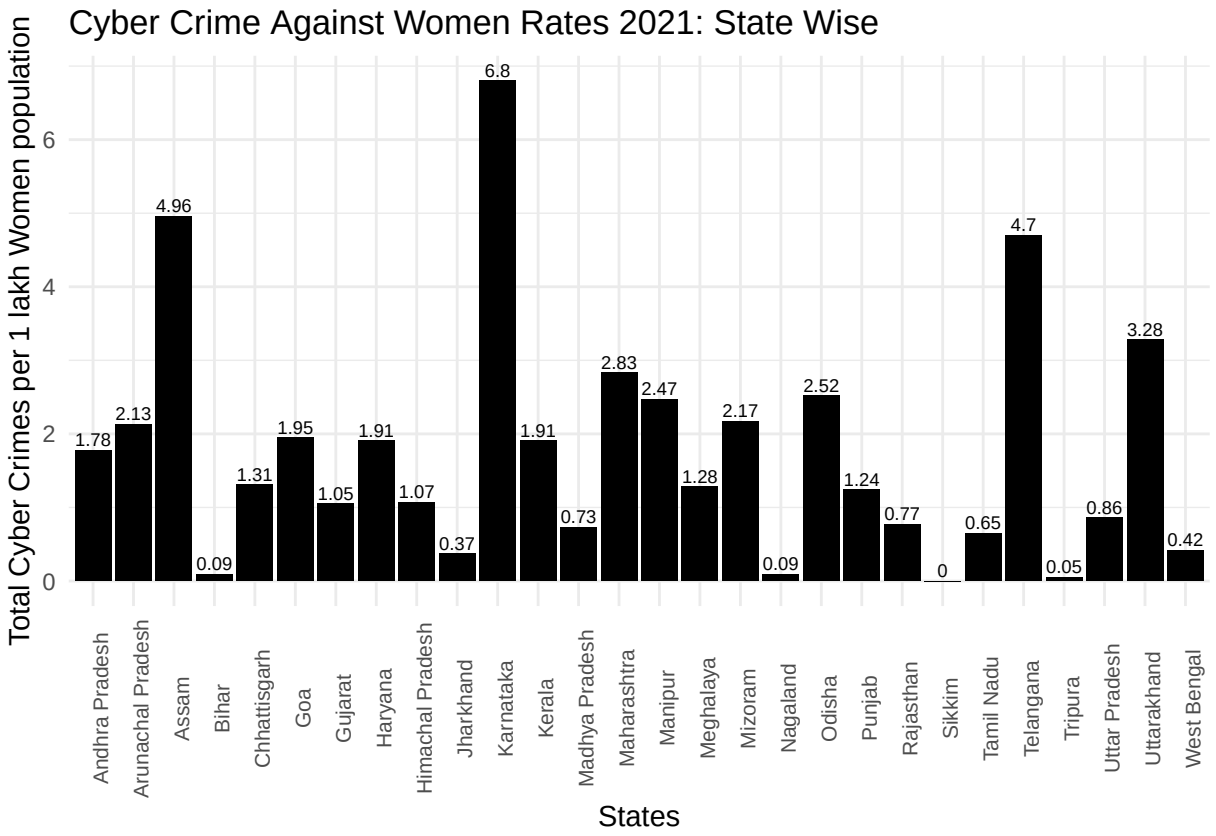


5.3.2 Year 2021

State-Wise

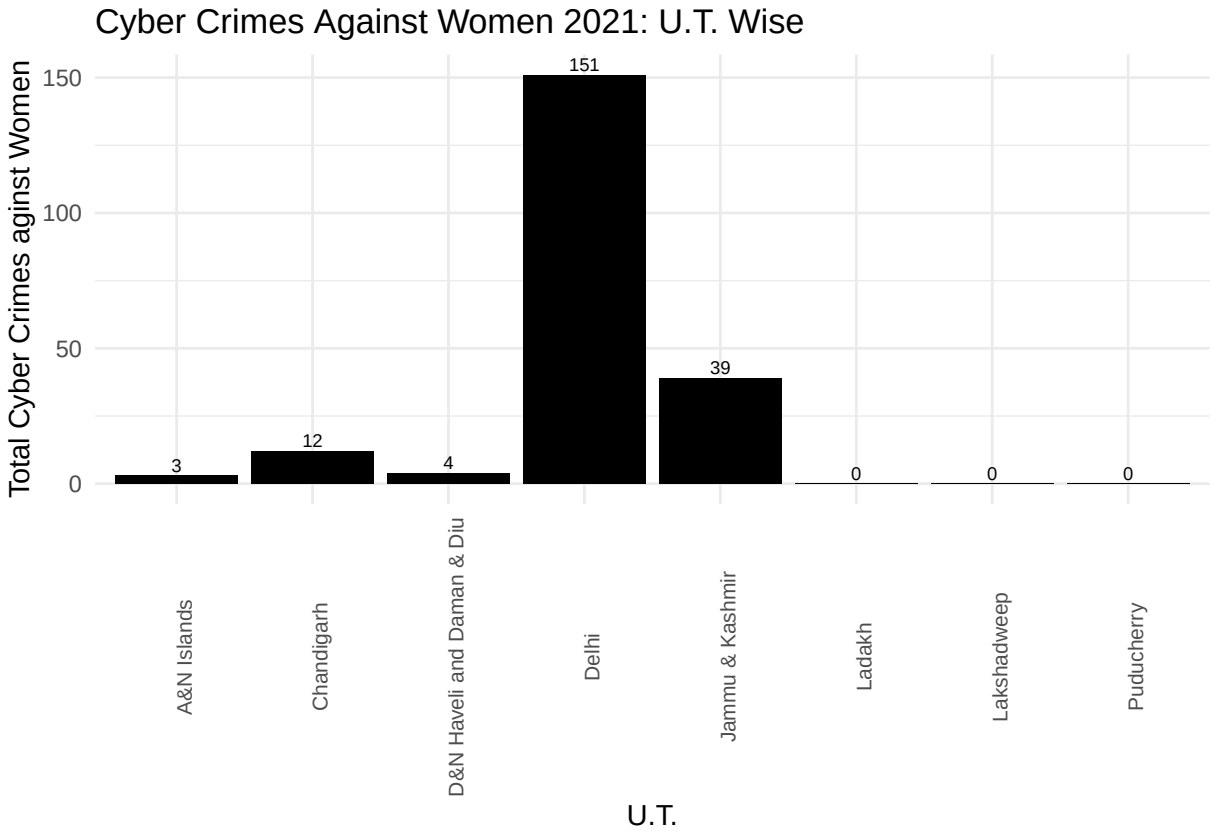


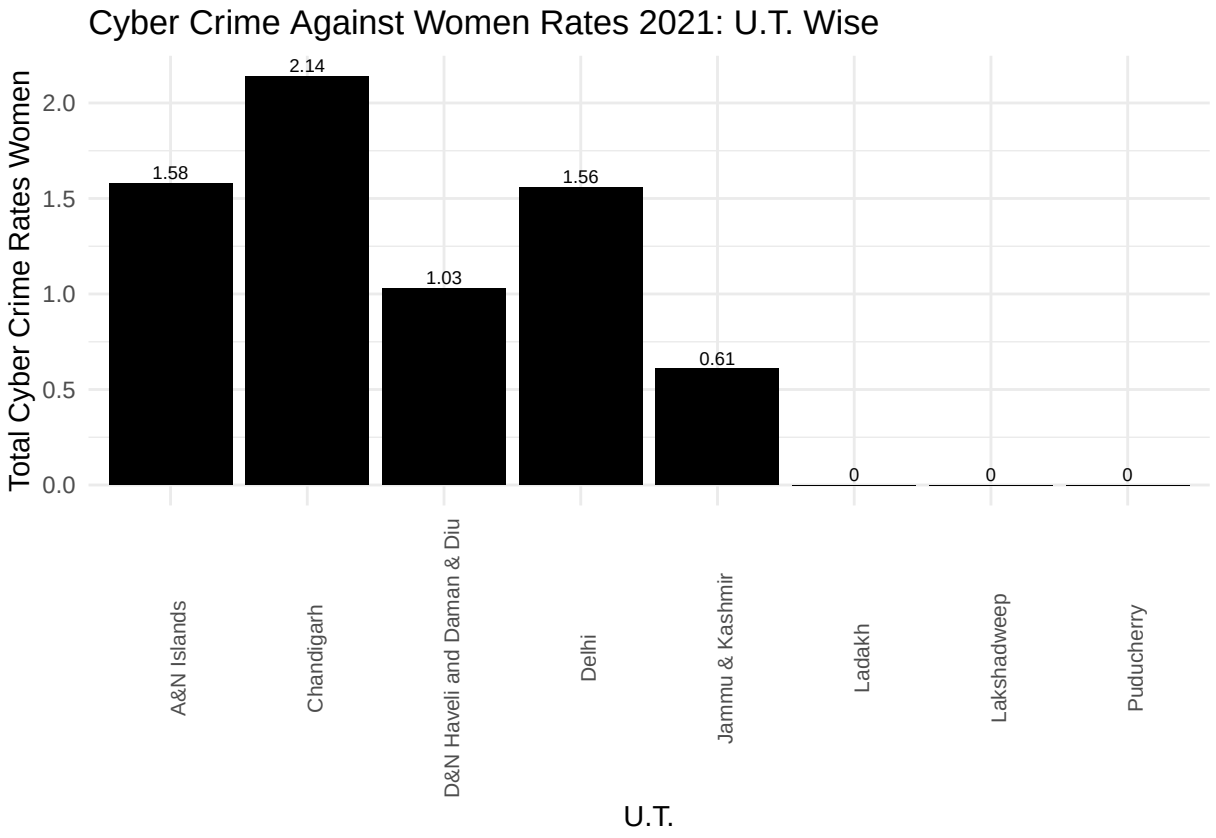
Karnataka, Maharashtra and Assam are again the states with higher cyber crimes against women.



Again, the state of Karnataka has the maximum cyber crime rate against women followed by Assam, Telangana and Uttar Pradesh.

U.T. Wise





5.4 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.
- 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.
- 2022, Tanushree Basuroy, Number of Active Internet users in India 2017-2025.