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 behavior and manage known offenders/ suspects.
- **Powers to 'dispose' off 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 Cyber Crimes under I.T. Act:**

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

- **Crimes under the Indian Penal Code (IPC)**

Abetment of Suicide (Online)	Cyber Stalking/Bullying of Women/Children	Cyber Blackmailing/Threatening		Credit Card/Debit Card
ATMs	Online Banking Fraud	Frauds	OTP Frauds	Cheating
Forgery	Defamation/Morphing	Fake Profile	Counterfeiting	Currency
Stamps	Data Theft	Fake News on social media		Other Offenses

- **Crimes under the Special and Local Laws (SLL)**

Gambling Act (Online Gambling)	Lotteries Act (Online Lotteries)
Copy Right Act	Trade Marks Act
Other Special and Local Laws (SLL) Crimes	

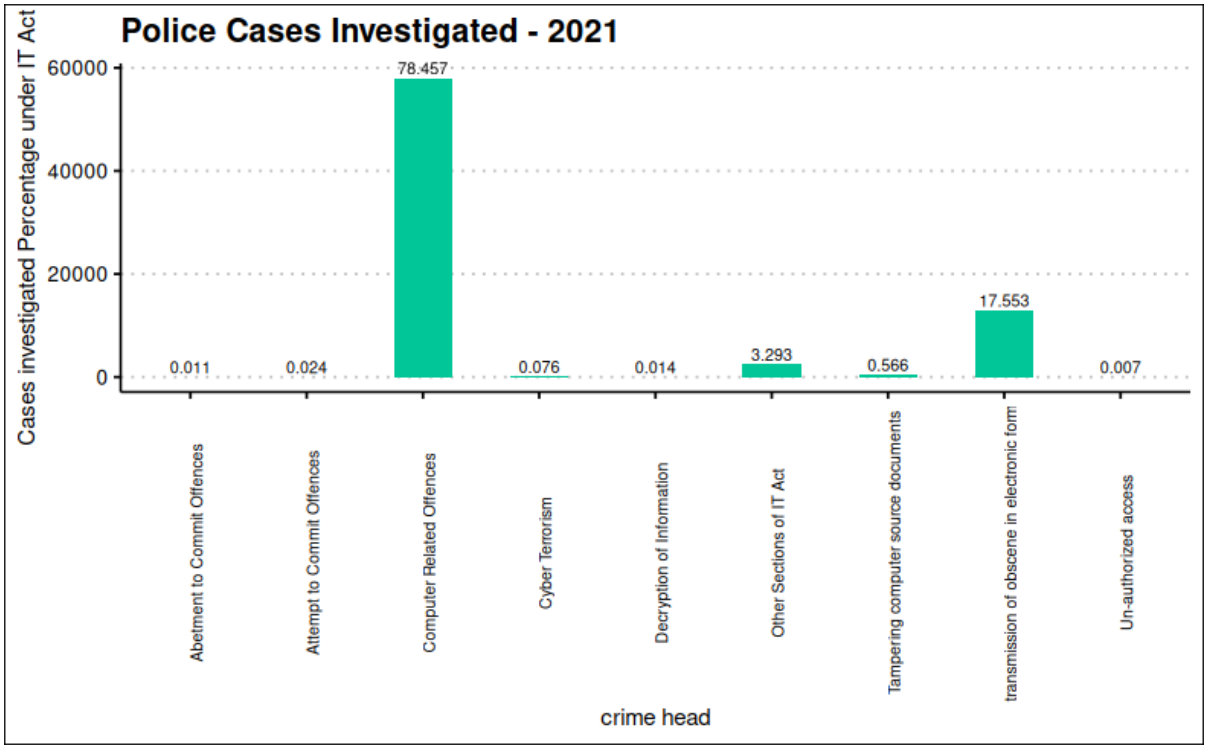
4.2 Statistics and Data Visualization

Police Disposal of Cases: Crime Head-Wise

Plotting the graphs for cases investigated by police, case disposed by police cases, pending by police for the year 2021 on crime-head and graph of cases pending by police for the year 2020 under IT Act, Under IPC Act, Under SLL Act.

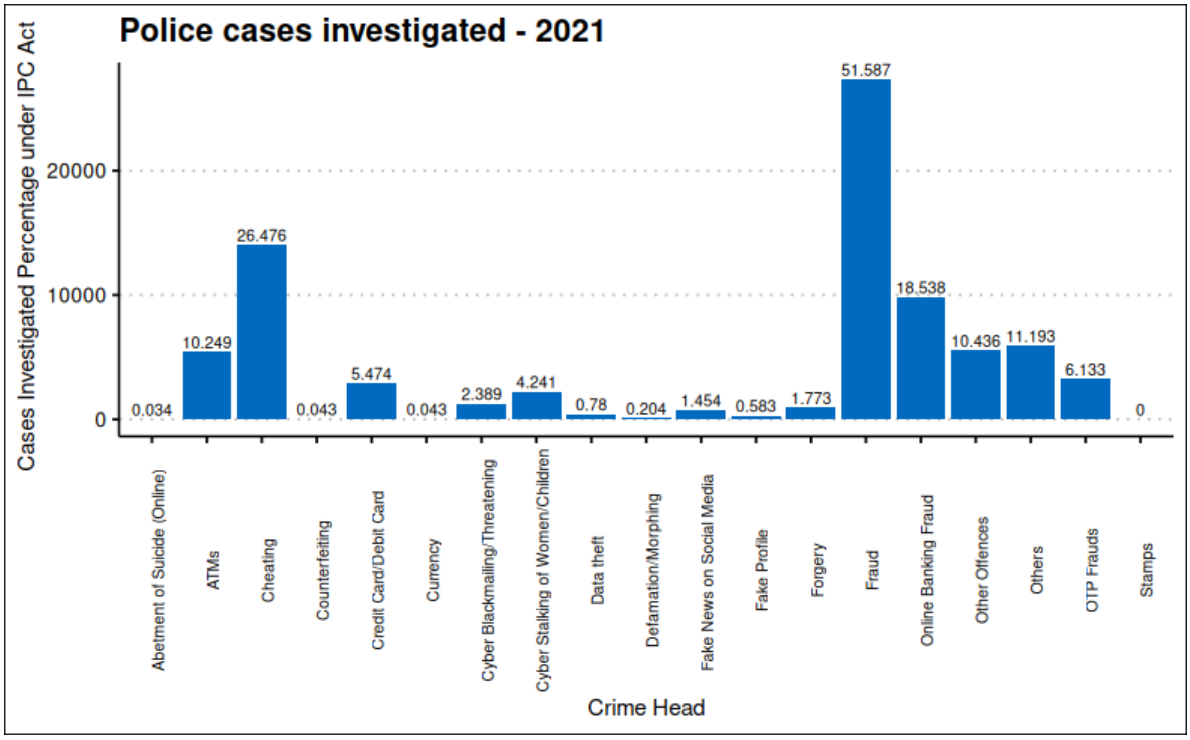
4.2.1 Total Cases Investigated

Year 2021



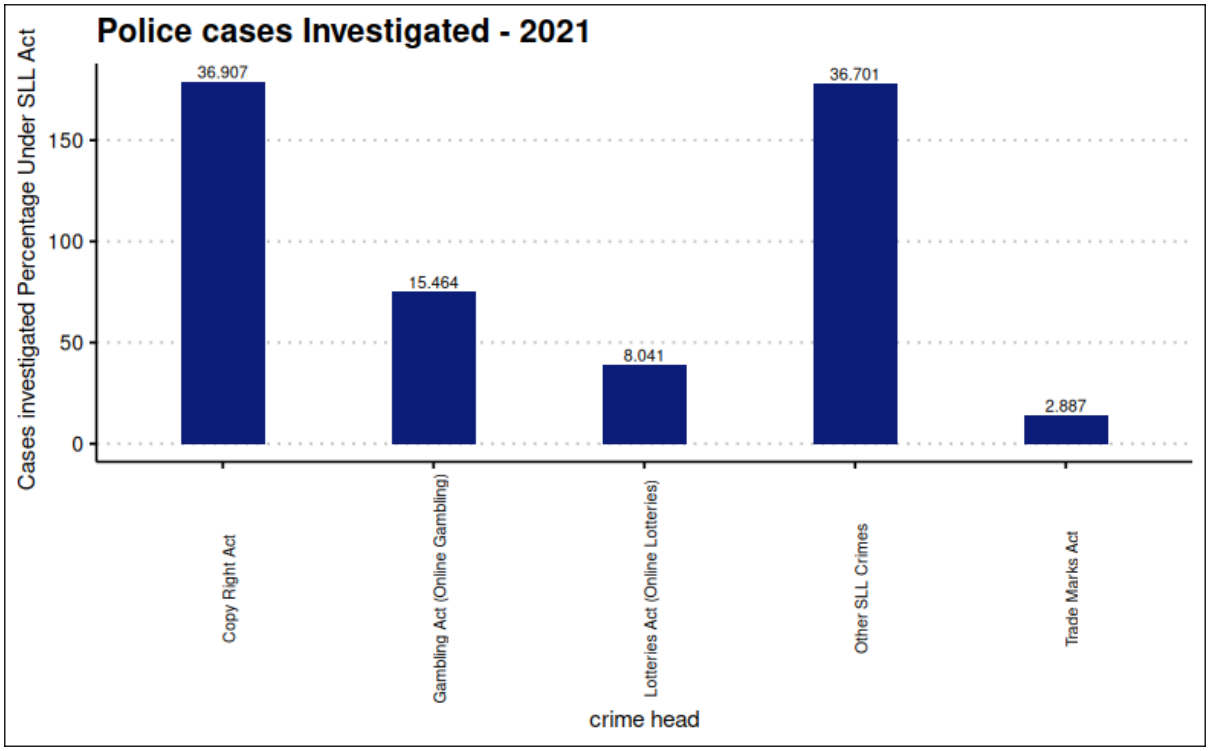
The police cases investigated in the year 2021 are shown in the above plot in the sense of Crime Head-Wise.

It is evident from the above figure shows that under IT Act of computer related offenses 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 cases investigated percentage of cheating is the second highest with **26.476 %**

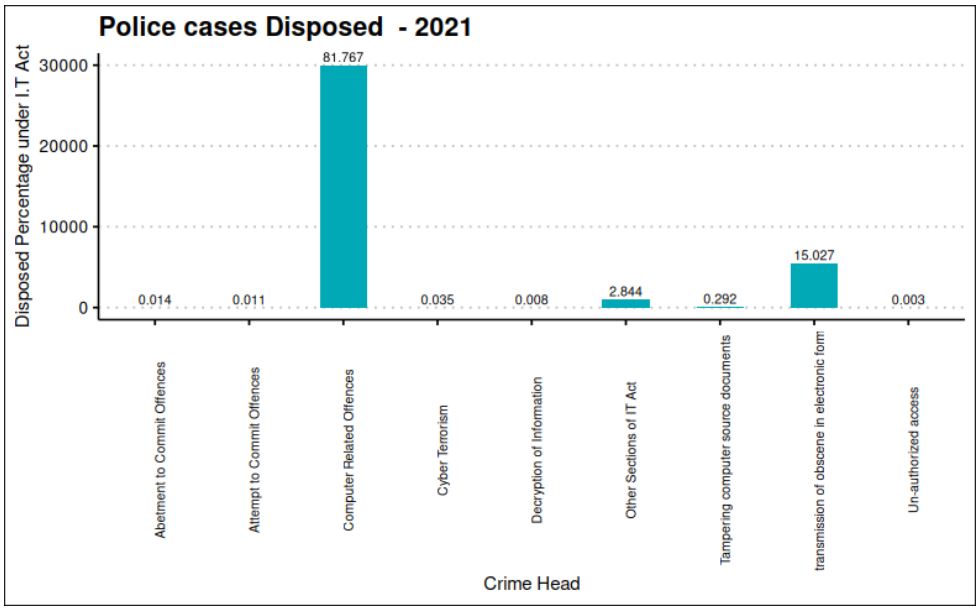


The above figure shows that cases investigated by Police "Under SLL Act" of Copy Right Act are highest with **36.907 %**.

Just with a slight difference, the second highest in this category is Other SLL Crimes with **36.701 %**.

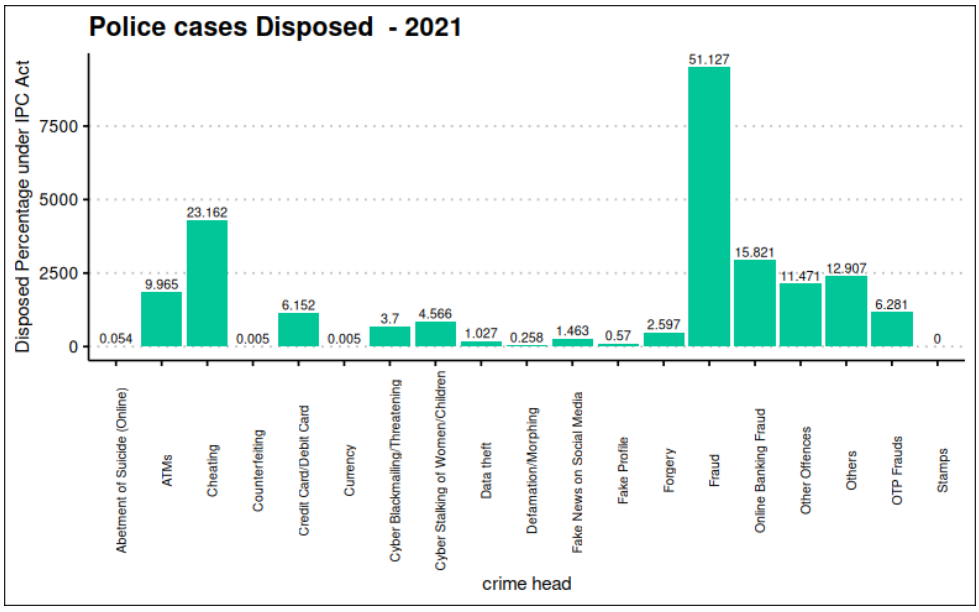
4.2.2 Total Cases Disposed

Year 2021



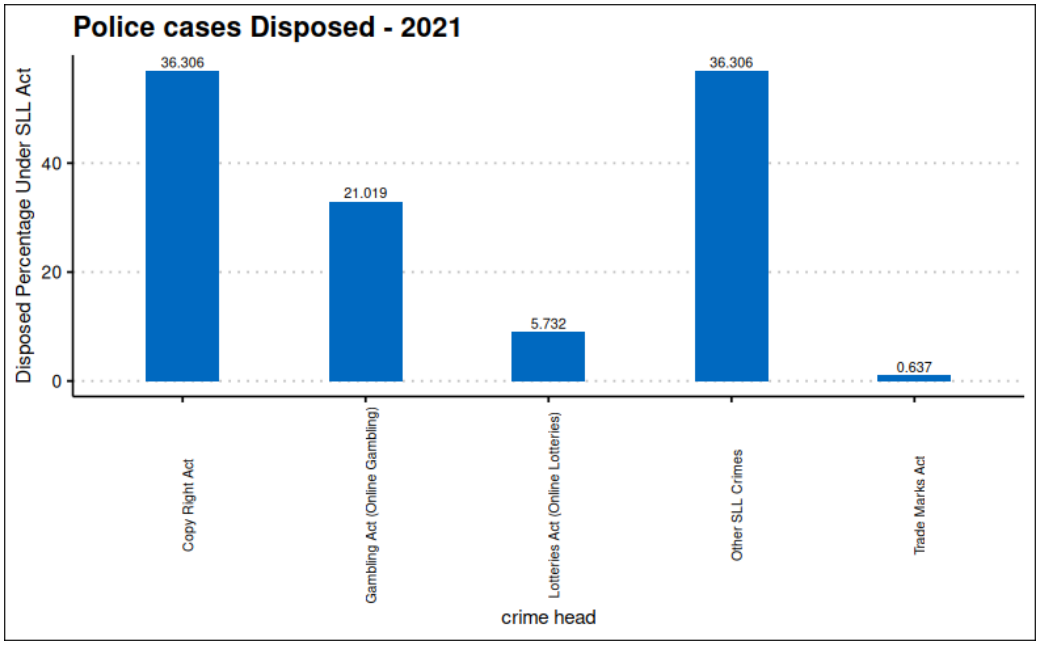
The above figure shows that cases Disposed under IT Act of computer related offenses are highest with **81.767%**.

The second highest in this category is 'Transmission of Obscene material in electronic form' with **15.027 %.**



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

Also, cases Disposed by Police "Under IPC Act" of Cheating are the second highest with **23.162%**.

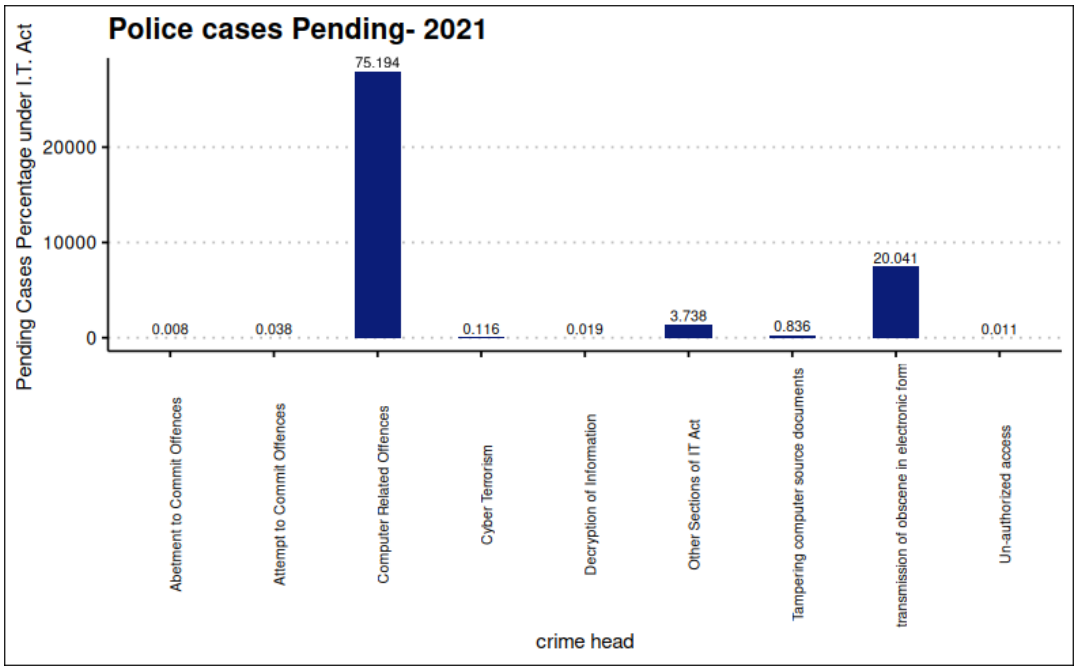


The above figure shows that cases Disposed by Police "Under SLL Act" of Copy Right Act and Other SLL Crimes are highest both with **36.306%**.

Also, the second highest in this category is the cases Disposed by Police "Under SLL Act" of Gambling Act (Online Gaming) with **21.019 %.**

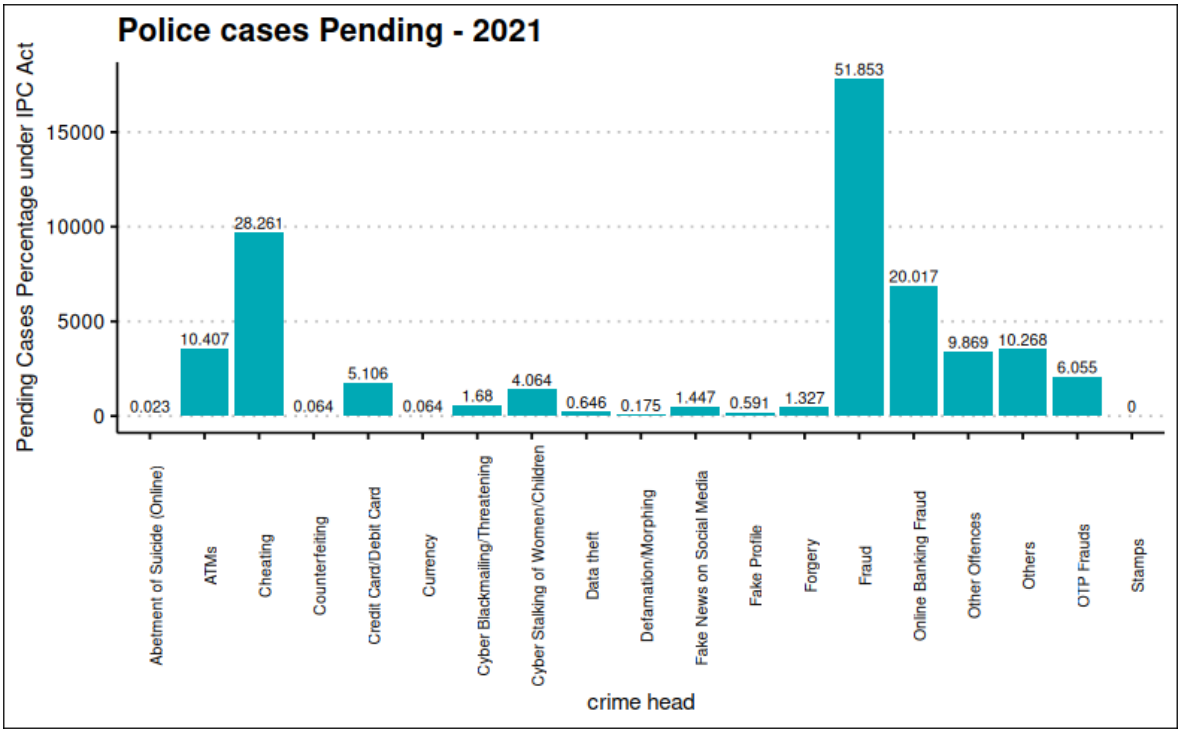
4.2.3 Total Cases Pending

Year 2021



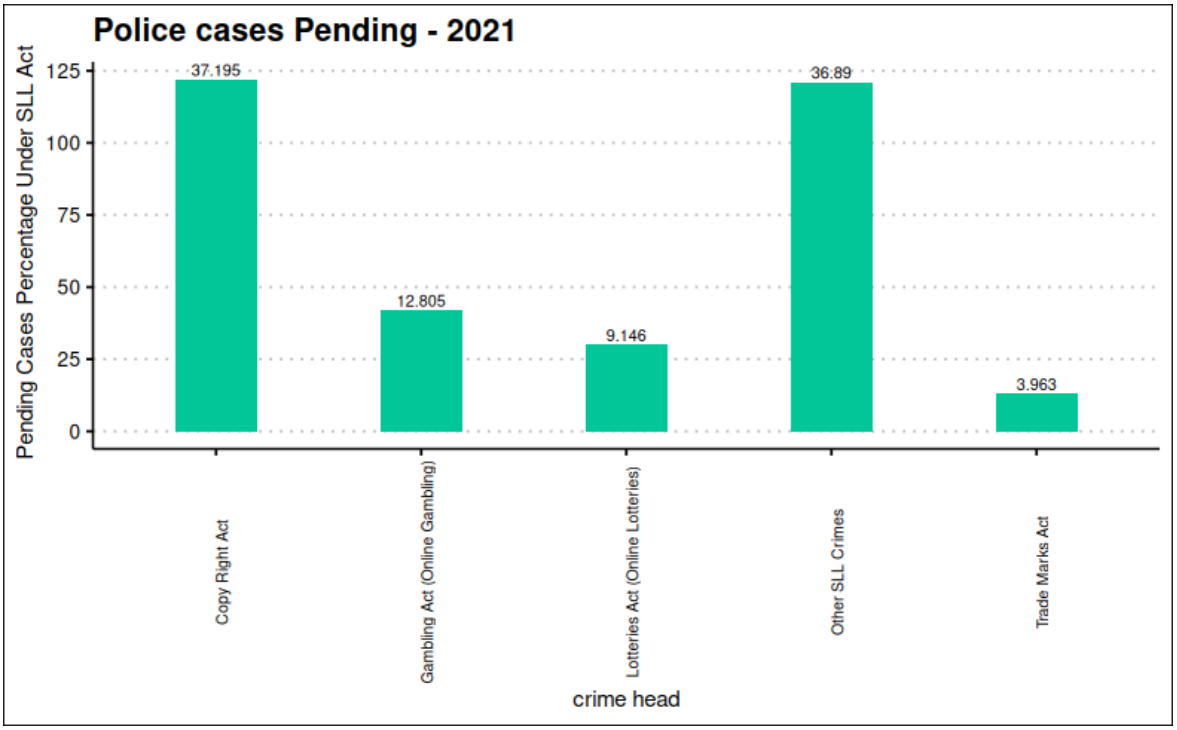
The above figure shows that Cases pending under IT Act of computer related offenses are highest with **75.194%**.

Also, it can be noted that the Computer Related Offenses are more prevalent in the recent years.



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

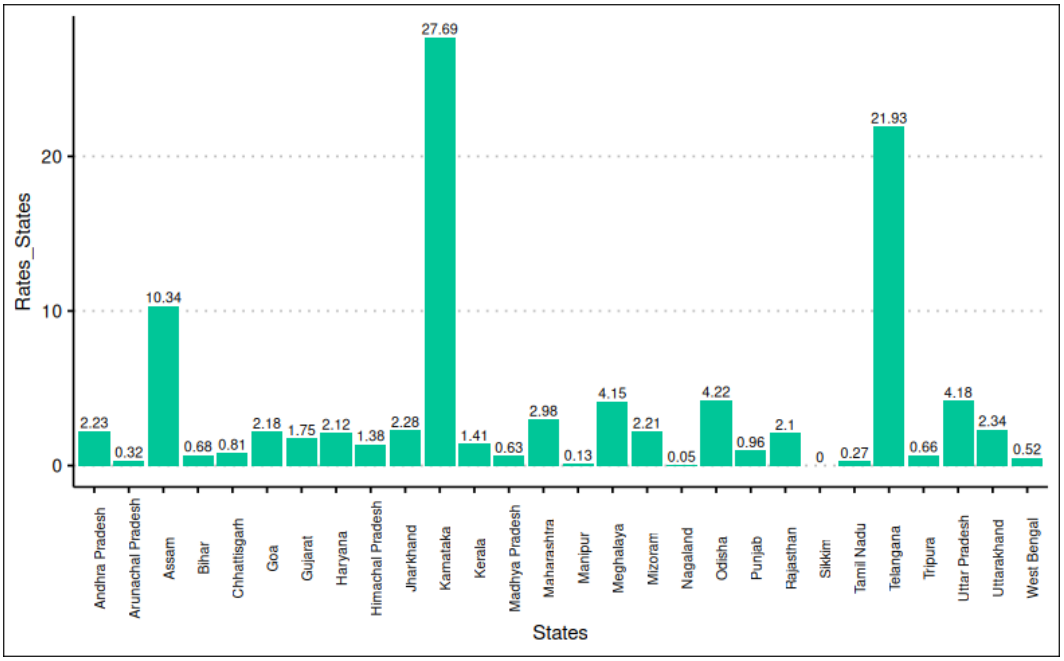
Majority of Cyber Crimes in India are due to Fraud as discussed in Chapter 2.



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

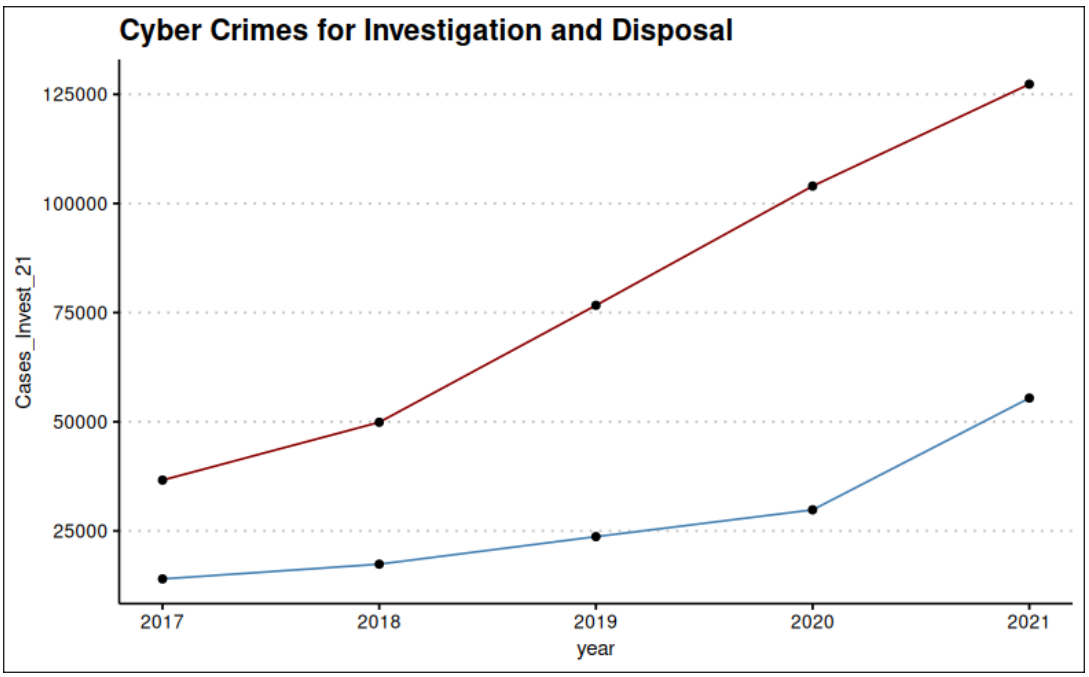
Along with that the cases pending by Police "Under SLL Act" of Other SLL Crimes are the second highest, on just a slight difference with **36.89%.**

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



The Police Disposal Rate of Cyber Crimes is highest in the state of **Karnataka** with **27.69** cases disposed per 1 lakh population, followed by the state of Telangana and Assam with the rates of 21.93 and 10.34 respectively.

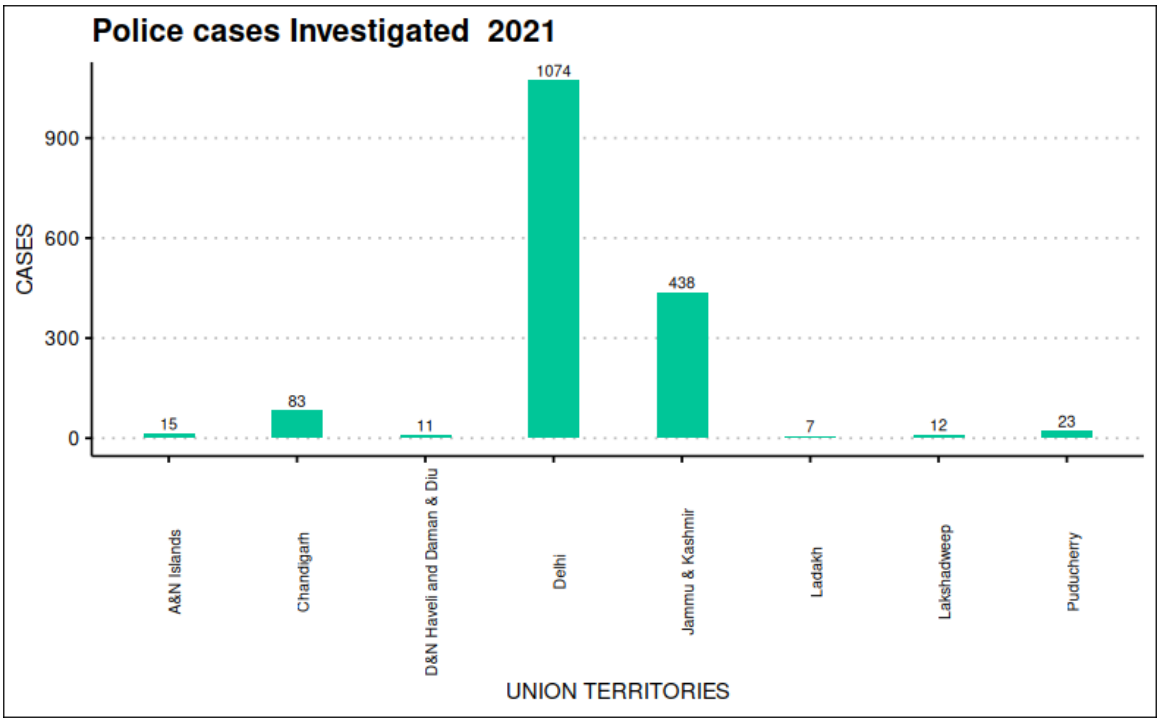
Total Cyber Crimes for Investigation and their Disposal by Police



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

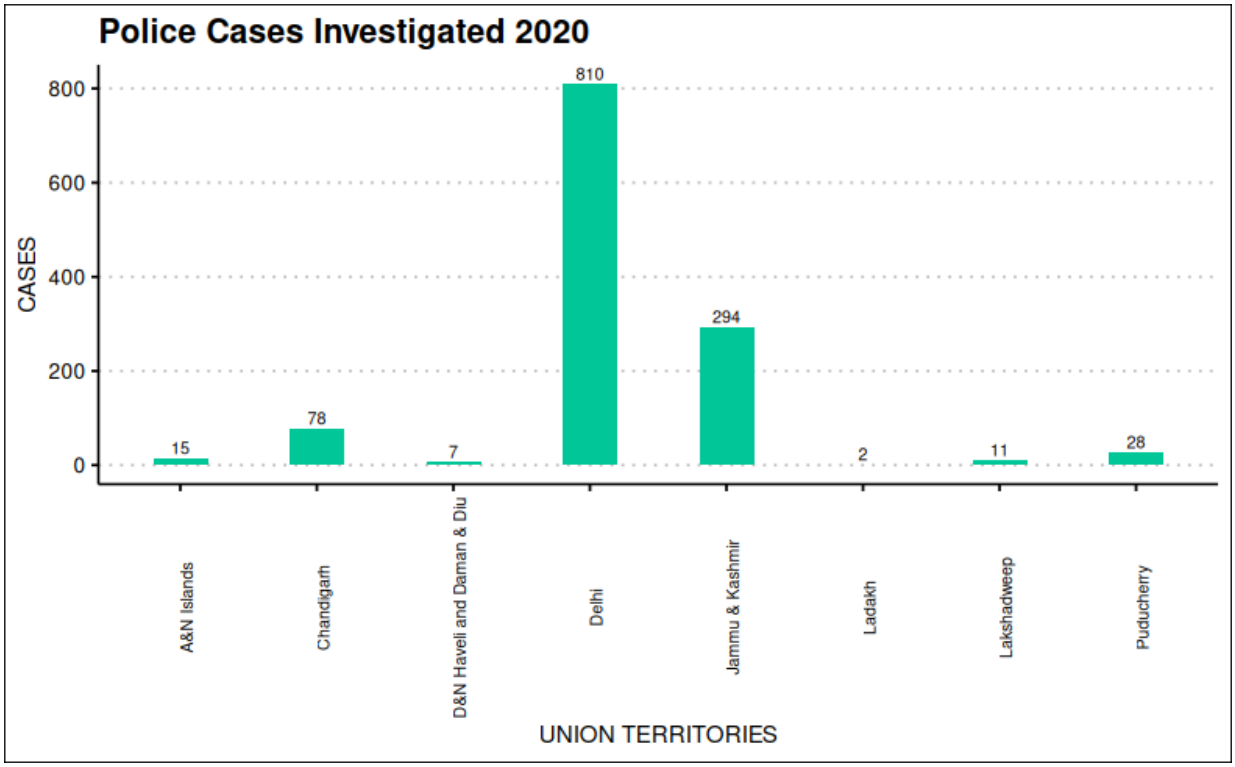
Union Territories

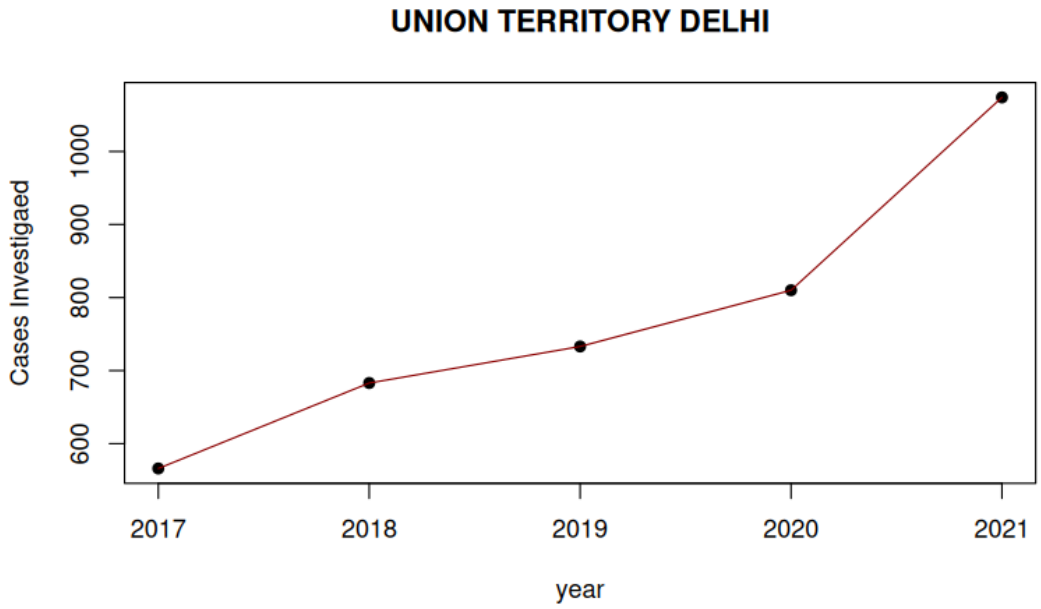
Year 2021



The above figure shows that cases investigated in 2021 are more in Delhi than the other union territories. Police Investigated **1074** cyber-crime incidents in the Union Territory of Delhi in the year 2021, followed by **438** investigations of the same in Jammu and Kashmir.

Year 2020

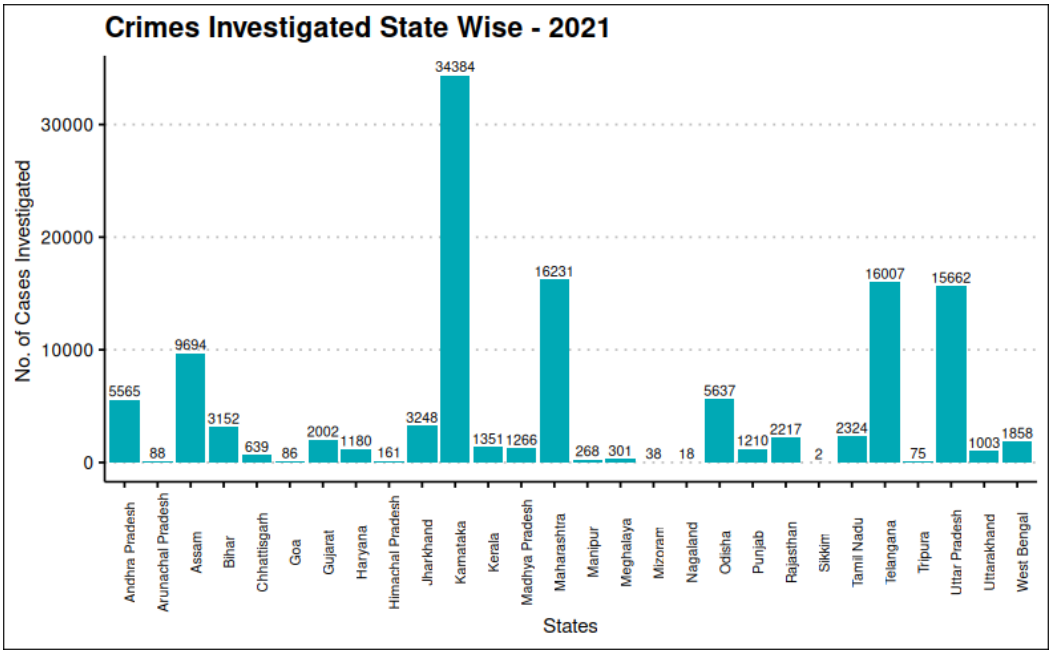




Now, it is of interest to analyze the Union Territory of Delhi based on the performance of past 5 years. Delhi continuously shows an increasing trend as far as the Police Investigation of Cyber Crimes are concerned.

States

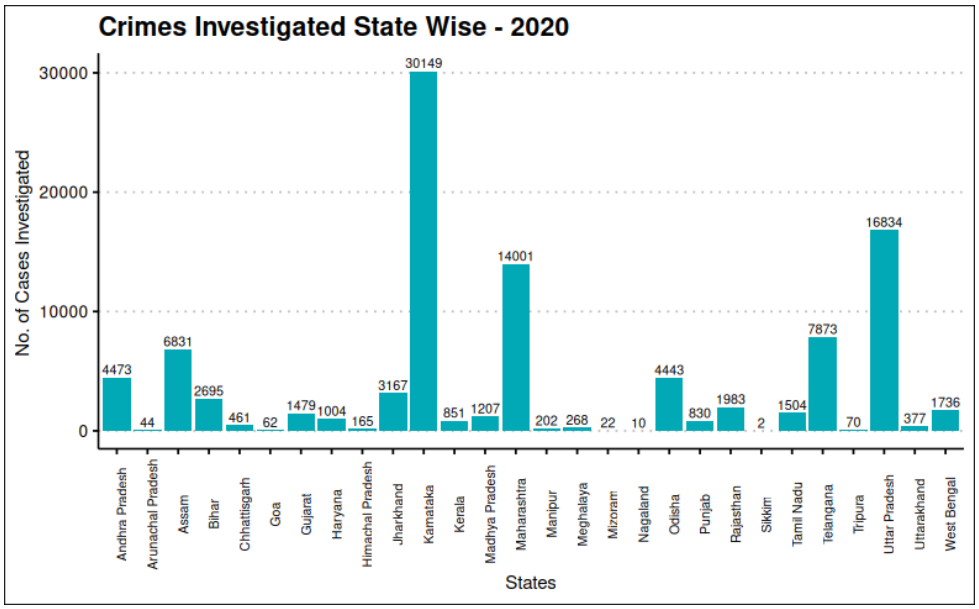
Year 2021



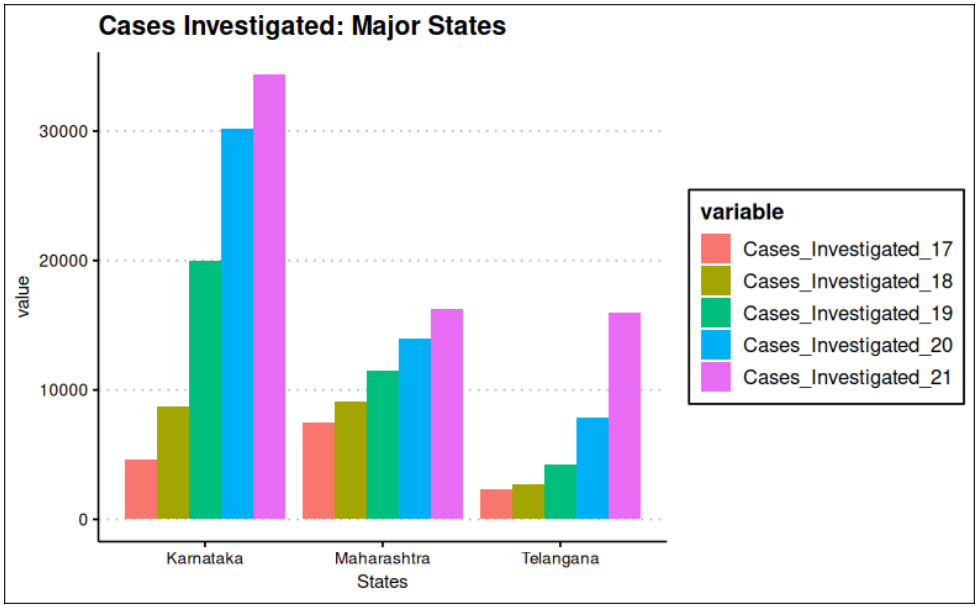
The most cyber crime incidents were investigated in the state of Karnataka, where police investigated **34,384** incidents during 2021.

After that the number is quite high in the states of Maharashtra, Telangana, Uttar Pradesh and Assam.

Year 2020



From the graphs of year 2021 and 2020 it has been observed that cases investigated in the **three major states** Karnataka, Maharashtra, Telangana have the higher number of cases investigated by Police than the other states.



This graph shows that cases investigated in Karnataka, Maharashtra, Telangana are increasing from year 2017 to 2021.

4.3 Testing location (Crime Head-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. 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 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. We fail to reject null hypothesis. Therefore, there is no significant change in the location throughout the years 2017-2021. 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. 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 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. 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 pending crime head wise are statistically same for different years.

4.3.1 Testing variability and location (state-wise)

Levene's Test and Kruskal Wallis Test

```
## 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. 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. 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. 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. We failed to reject null hypothesis. Therefore, there is no significant change in the locations throughout the years **2017-21**. 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. 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. We failed to reject null hypothesis. Therefore, there is no significant difference in 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. 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, we conclude that cases pending are statistically same in terms of location and variability for different crimes.

4.3.2 Paired Wilcoxon Signed-Rank test

The Paired Wilcoxon Signed-Rank test is a non-parametric alternative to the paired t-test for comparing the average difference in case of dependent samples.
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.
Test assume that 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.

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  stdata$True_cases.but.insufficient.evidence_17 and
stdata$True_cases.but.insufficient.evidence_21
## 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.

4.4 Hierarchical Cluster Analysis

There are two major techniques under Hierarchical Cluster Analysis

1. Agglomerative Hierarchical Cluster Analysis.
2. Divisive Hierarchical Cluster Analysis.

In Hierarchical Cluster Analysis (HCA), the observations are grouped together based on their mutual distances. It is visualized through a hierarchical tree, called *dendrogram tree*. Objects in the dendrogram are linked together based on their similarity.

Variables under consideration:

• Cases Investigated
• Cases Disposed
• Cases Pending
• Cases True but Insufficient evidence.

Data Structure:

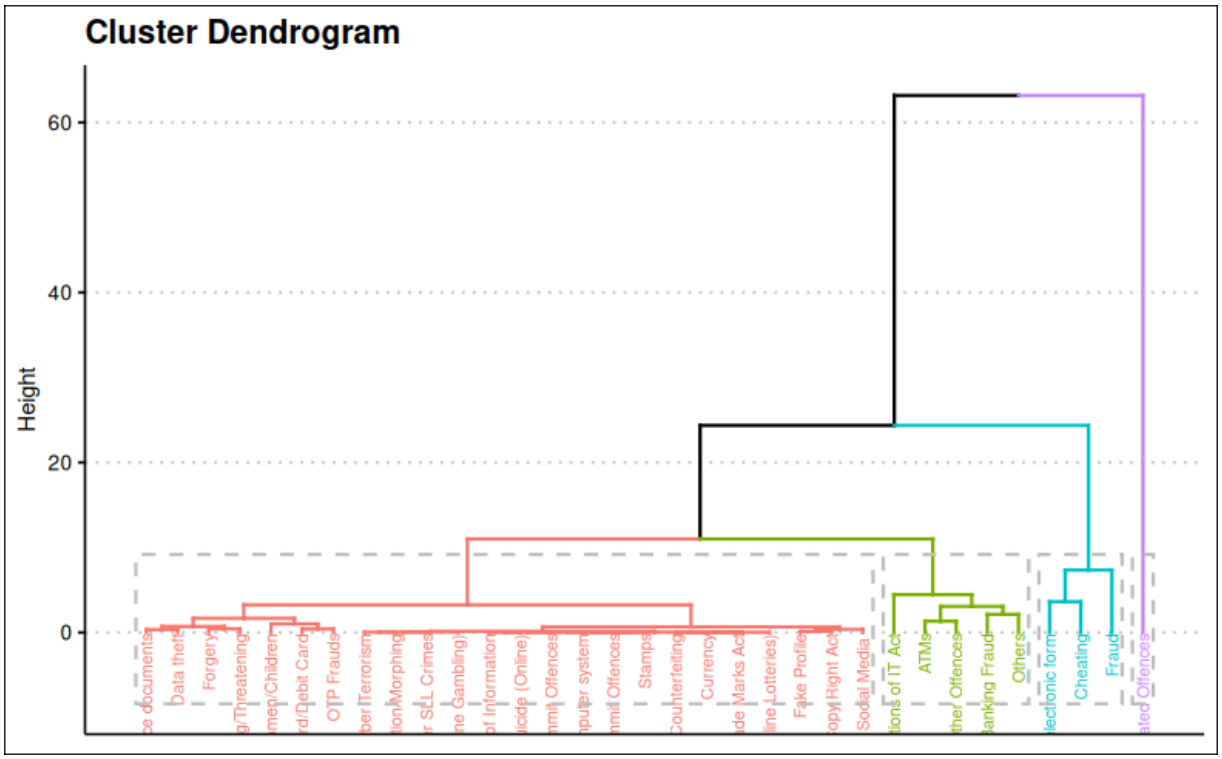
• Each Crime Head corresponds to a vector of order 20×1 using the observations of the above variables for the years 2017 - 2021.
• The vectors are made for the Crime Heads, then the Agglomerative HCA and Divisive HCA is applied.
• There are 33 Crime Heads used in this study.

Agglomerative HCA:

Agglomerative approach operates by successive merger of cases.

- First step involves the distance matrix calculations, in which the distance of each object from every other object is calculated and a distance matrix is formed. Euclidean distance after standardizing or Mahalanobis distance are popular for this matter.
- This Algorithm begins with N clusters each containing single cases at the initial stage.
- At each stage, two most similar groups are merged to form a new cluster, thus reducing the no. of clusters by one.
- Now the fusion process is repeated till we are left with a single cluster with all the objects in it.

- The various cluster distance measures that can be used are: Single Linkage, Complete Linkage, Median Linkage, Centroid Linkage, Average Linkage and Wald D2 Linkage etc.



In the dendrogram displayed above, each leaf corresponds to one object. As we move up the tree, objects that are like each other are combined into branches, which are themselves fused at a higher height.

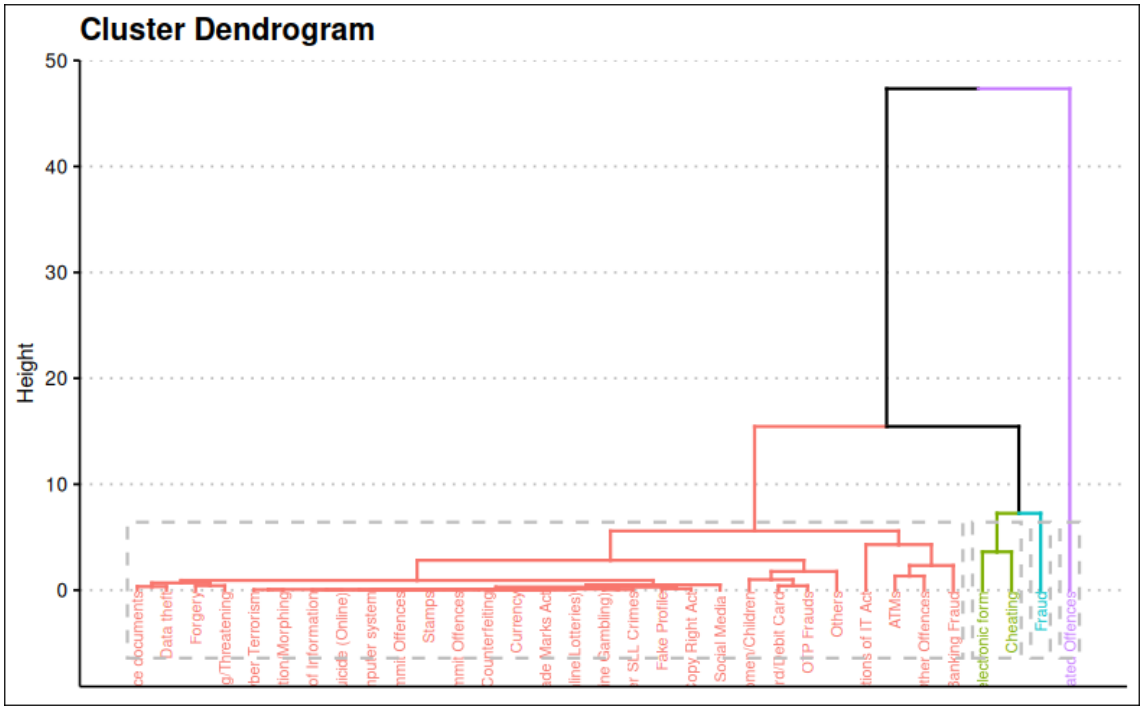
The height of the fusion, provided on the vertical axis, indicates the similarity/distance between two objects/clusters. It has been observed that computer related offenses, electronic form, Cheating, fraud are higher forming one cluster.

Divisive HCA

Divisive approach operates just as opposite as that of Agglomerative.

- First step involves the distance matrix calculations, in which the distance of each object from every other object is calculated and a distance matrix is formed. Euclidean distance after standardizing or Mahalanobis distance are popular for this matter.
- This Algorithm begins with a single cluster containing all the cases at the initial stage.
- At each stage, two most distinct groups are formed in the sense of maximum distance.
- Now the division process is repeated till we are left with N clusters each with a single object in it.

If Divisive HCA is applied the major four crimes computer related offenses, electronic form, Cheating, fraud in one cluster and others in another cluster.



Interpretation

The dendrogram tree just flattens down after a distance level of 10 units, which in turn divides the Crime Heads into 3 clusters.

Cluster 1

```
##  [1] "Tampering computer source documents"
##  [2] "Data theft"
##  [3] "Forgery"
##  [4] "Cyber Blackmailing/Threatening"
##  [5] "Cyber Terrorism"
##  [6] "Defamation/Morphing"
##  [7] " Decryption of Information"
##  [8] "Abetment of Suicide (Online)"
##  [9] "Un-authorized access/attempt to  access to protected computer
system"
## [10] "Abetment to Commit Offences"
## [11] "Stamps"
## [12] "Attempt to Commit Offences"
## [13] "Counterfeiting"
## [14] "Currency"
## [15] "Trade Marks Act"
## [16] "Lotteries Act (Online Lotteries)"
## [17] "Gambling Act (Online Gambling)"
## [18] "Other SLL Crimes"
## [19] "Fake Profile"
## [20] "Copy Right Act"
## [21] "Fake News on Social Media "
## [22] "Cyber Stalking/Bullying of Women/Children"
## [23] "Credit Card/Debit Card"
```

```
## [24] "OTP Frauds"
## [25] "Others"
## [26] "Other Sections of IT Act"
## [27] "ATMs"
## [28] "Other Offences"
## [29] "Online Banking Fraud"
```

In this cluster, the Crime Heads with comparatively lower incidence rate are included.

Cluster 2

```
## [1] "Publication/transmission of obscene / sexually explicit act in
electronic form"
## [2] "Cheating"
## [3] "Fraud"
```

Cluster 3

```
## [1] "Computer Related Offences"
```

The crime incidence of Computer related offenses is very high as compared to the other crime heads. Hence its distance comes out to be very high from the other crime heads. Consequently, it is alone in the third cluster.