

# Crimes in India

---

This project involved various tasks, which has been broadly grouped into 4 phases, as discussed below:

## **Phase 1: Data Collection/preparation part**

This phase requires data collection from various sources, so as to work on the project. The data collected were:

- 1.1 about reported crimes in India (2001 – 2012): This data was available online and downloaded from the following GitHub link:

[https://github.com/ravikesh2426/ML-File/blob/9236b5858591e296fecc0d302922082b76b540f1/India\\_Crime%20-Zip.zip](https://github.com/ravikesh2426/ML-File/blob/9236b5858591e296fecc0d302922082b76b540f1/India_Crime%20-Zip.zip)

- 1.2 about the population of each state: This data was web scrapped using SELENIUM from the following wikipedia link:

[https://en.wikipedia.org/wiki/List\\_of\\_states\\_in\\_India\\_by\\_past\\_population](https://en.wikipedia.org/wiki/List_of_states_in_India_by_past_population)

- 1.3 about literacy rate in each state: This data was also web scrapped from the following Wikipedia link:

[https://en.wikipedia.org/wiki/List\\_of\\_Indian\\_states\\_and\\_union\\_territories\\_by\\_literacy\\_rate](https://en.wikipedia.org/wiki/List_of_Indian_states_and_union_territories_by_literacy_rate)

- 1.4 about area of each state: This data was available at the following Wikipedia link:

[https://en.wikipedia.org/wiki/List\\_of\\_states\\_and\\_union\\_territories\\_of\\_India\\_by\\_area](https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_area)

## **Phase 2: State/UT wise analysis**

This phase involved analysing crimes in India on following parameters:

- 2.1 To analyse Literacy Rate vs Total Crimes.

Total crimes do not take into consideration the effect of higher population, so instead of comparing literacy rate with total crimes, a better metric to compare the effect of literacy on crime would be to compare literacy rate and crime rate. Crime rate is the total crime committed per 100,000 of population. This way it considers the effect of more crimes in larger population. Therefore, to see the effect of just literacy on crime across all the states in India, crimes rate was calculated using the datasets of crime from sections 1.1 and 1.2 of Phase 1.

Crime rate was calculated using following formulae:

$$\text{Crime rate} = (\text{Total reported crimes} / \text{Total population}) \times 100,000$$

Fig. 2(a) and Fig. 2(b) show for the year 2001 and 2011 respectively, the following two plots:

- (i) States of India and Literacy rate (%)
- (ii) States of India and Crime rate (crimes per 100,000 population)

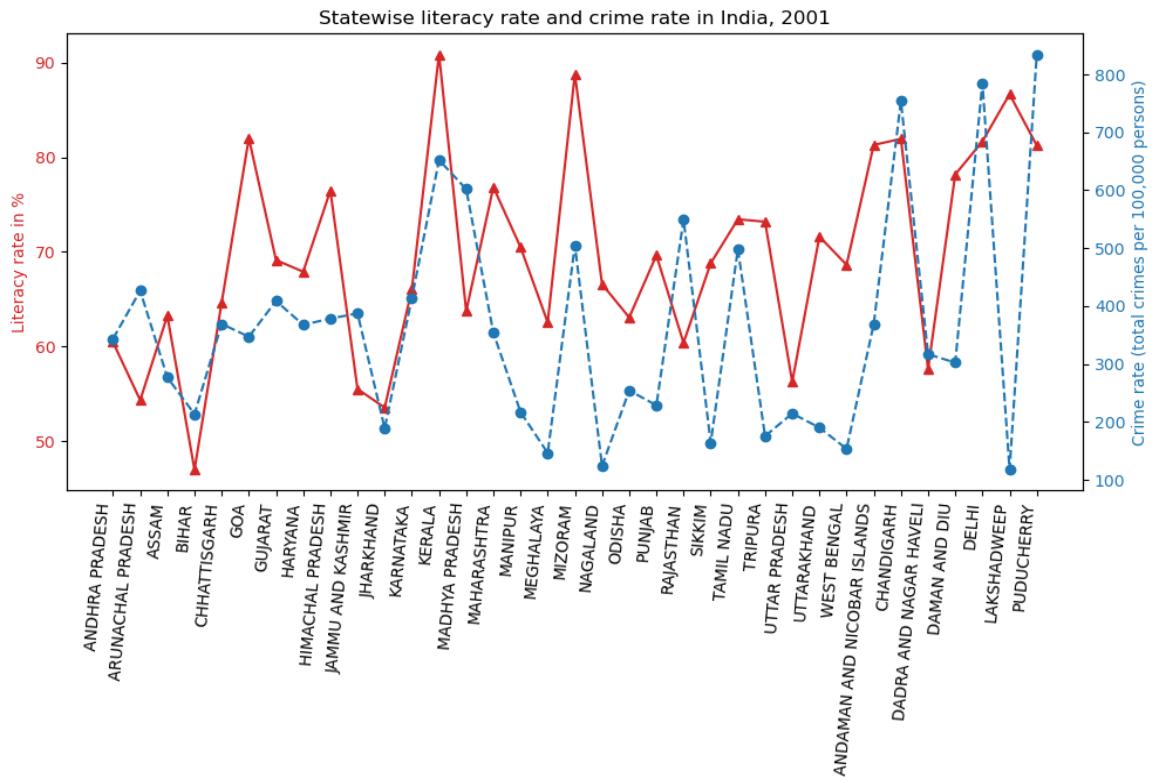


Fig. 2(a): States of India, Literacy rate and Crime rate (2001)

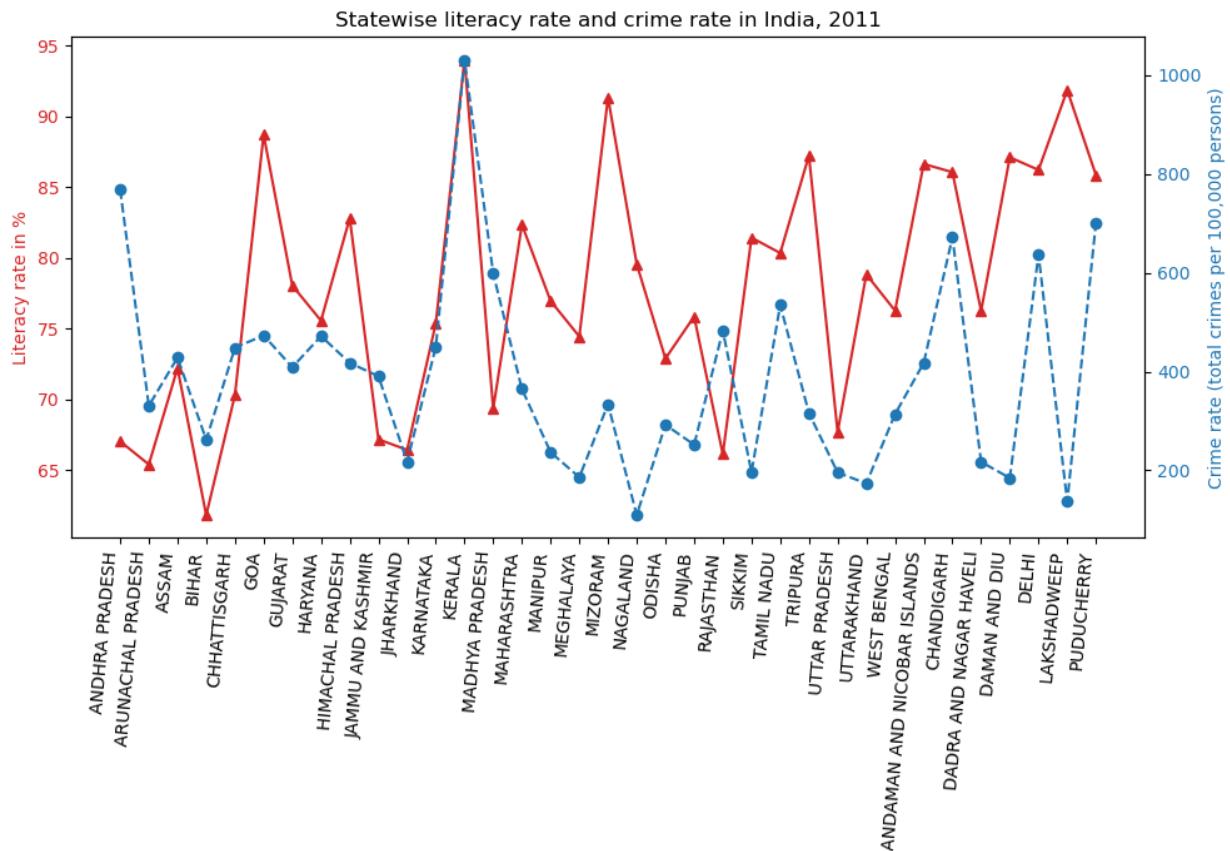


Fig. 2(b): States of India, Literacy rate and Crime rate (2011)

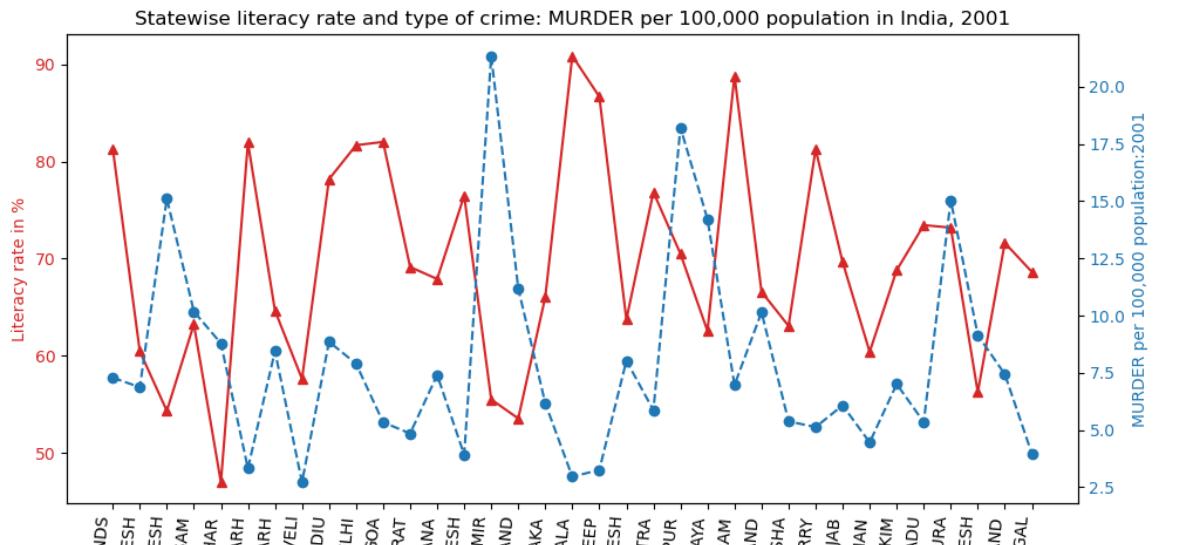
Table 1: Observations from Fig. 2(a) and Fig. 2(b)

	Lowest crime rates	Highest crime rates	Lowest literacy rates (%)	Highest literacy rates (%)
2001	LAKSHADWEEP (117.46), NAGALAND (124.00), MEGHALAYA (146.28), WEST BENGAL (153.48), SIKKIM (163.87), TRIPURA (175.51), JHARKHAND (188.87) and UTTARAKHAND (190.20).	CHANDIGARH (753.98), DELHI (785.24) and PUDUCHERRY (834.62).	BIHAR (47.00 %), JHARKHAND (53.56 %), ARUNACHAL PRADESH (54.34 %), JAMMU AND KASHMIR (55.52 %), UTTAR PRADESH (56.27 %) and DADRA AND NAGAR HAVELI (57.63 %).	PUDUCHERRY (81.24 %), ANDAMAN AND NICOBAR ISLANDS (81.30 %), DELHI (81.67 %), CHANDIGARH (81.94 %), GOA (82.01 %), LAKSHADWEEP (86.66 %), MIZORAM (88.80 %) and KERALA (90.86 %).
2011	NAGALAND (109.36), LAKSHADWEEP (136.58), UTTARAKHAND (173.45), DAMAN AND DIU (184.43), MEGHALAYA (185.90), UTTAR PRADESH (195.54) and SIKKIM (196.15).	PUDUCHERRY (701.02), ANDHRA PRADESH (768.55) and KERALA (1031.14)	BIHAR (61.80 %), ARUNACHAL PRADESH (65.38 %), RAJASTHAN (66.11 %), JHARKHAND (66.41 %), ANDHRA PRADESH (67.02 %), JAMMU AND KASHMIR (67.16 %), UTTAR PRADESH (67.68 %) and MADHYA PRADESH (69.32 %).	MIZORAM (91.33 %), LAKSHADWEEP (91.85 %) and KERALA (94.00 %).

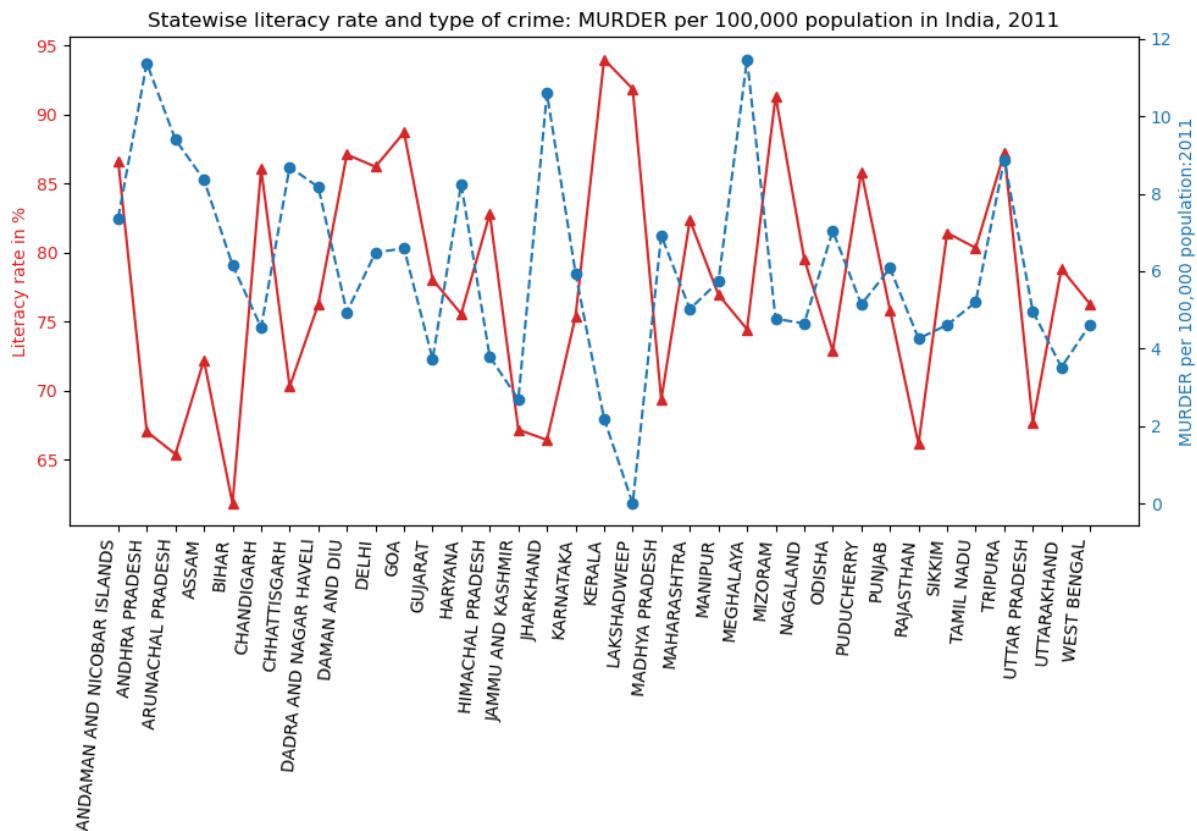
## 2.2. To analyse the type of crime vs each state vs Literacy rate.

The different type of crimes which have been analysed are:

Murder, Attempt to Murder, Culpable Homicide Not Amounting to Murder, Rape, Custodial Rape, Other Rape, Kidnapping & Abduction, Kidnapping And Abduction Of Women And Girls, Kidnapping And Abduction Of Others, Dacoity, Preparation And Assembly For Dacoity, Robbery, Burglary, Theft, Auto Theft, Other Theft, Riots, Criminal Breach Of Trust, Cheating, Counterfeiting, Arson, Hurt/Grievous Hurt, Dowry Deaths, Assault On Women With Intent To Outrage Her Modesty, Insult To Modesty Of Women, Cruelty By Husband Or His Relatives, Importation Of Girls From Foreign Countries, Causing Death By Negligence, Other IPC Crimes

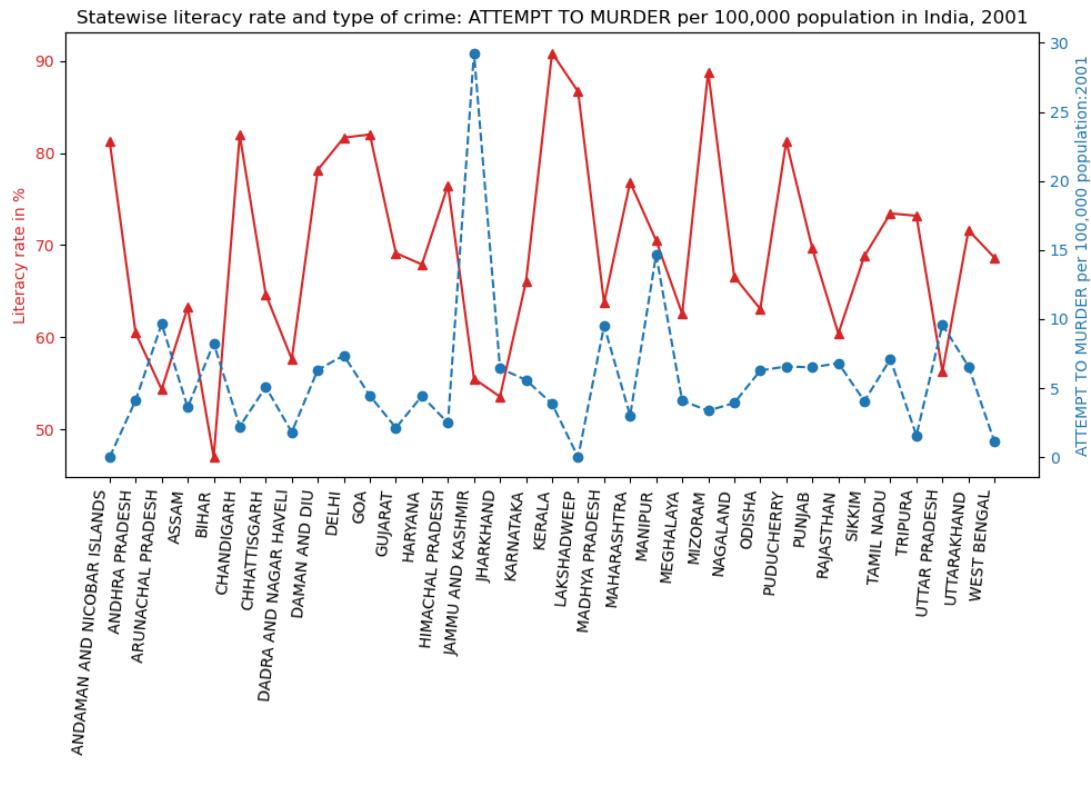


(a)

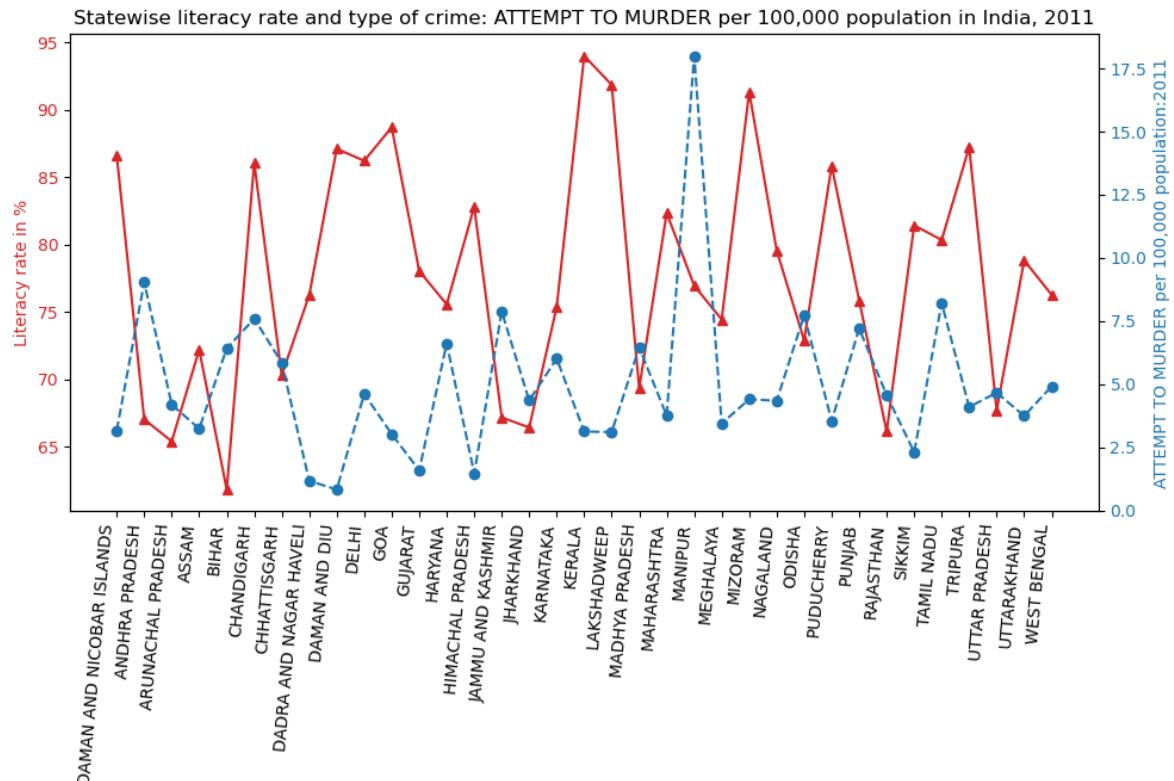


(b)

Fig. 3: State wise literacy and murder (a) 2001, (b) 2011



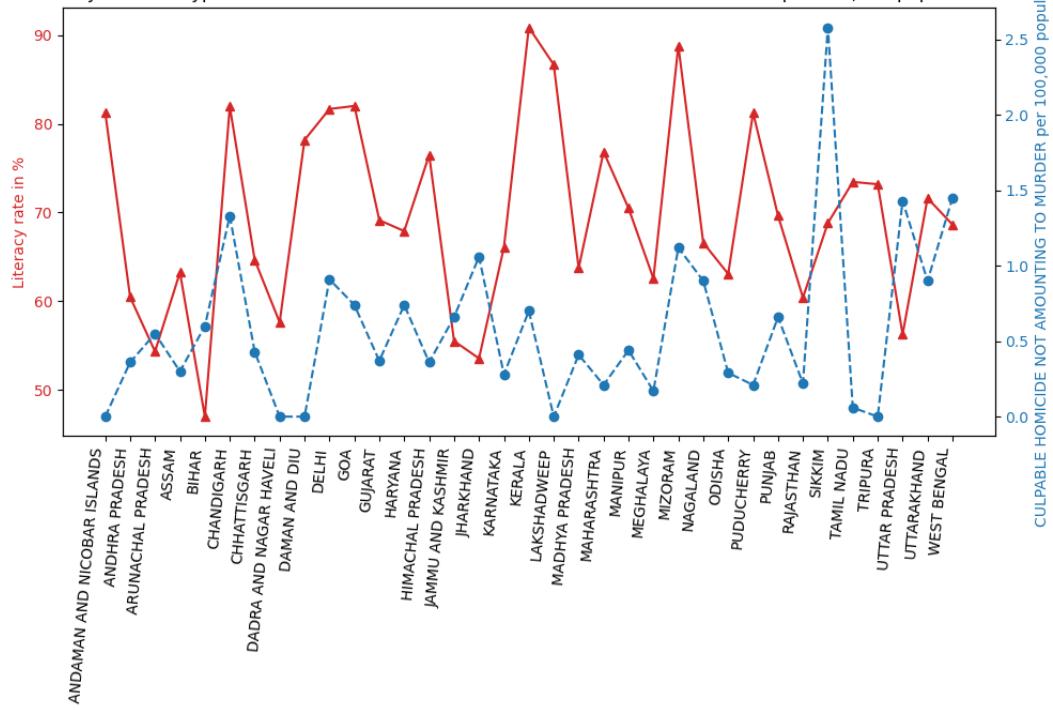
(a)



(b)

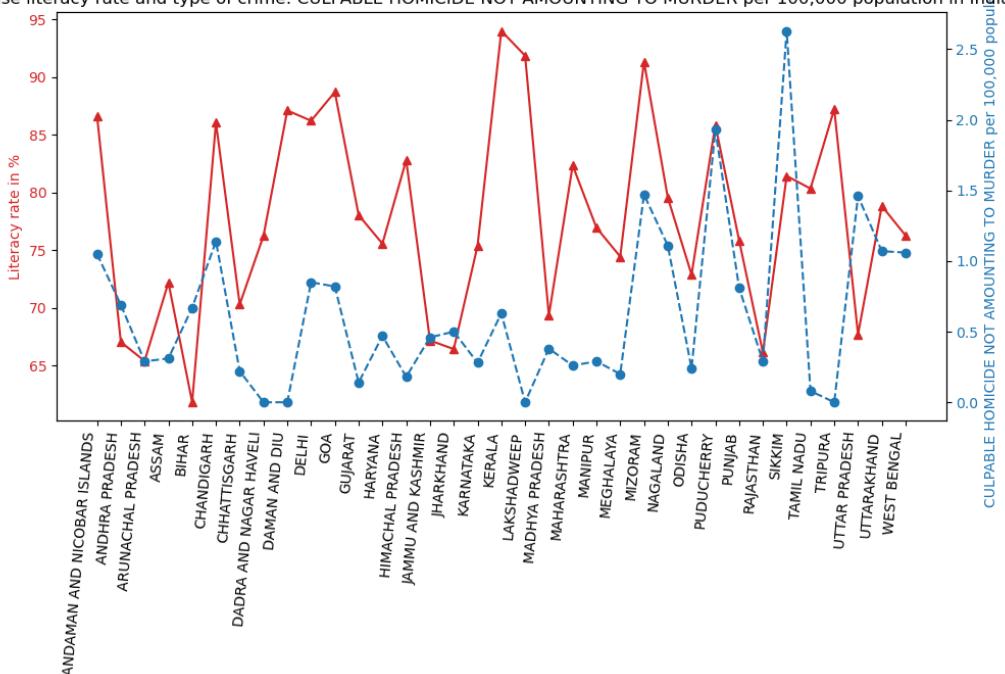
Fig. 4: State wise literacy and attempt to murder (a) 2001 (b) 2011

Statewise literacy rate and type of crime: CULPABLE HOMICIDE NOT AMOUNTING TO MURDER per 100,000 population in India, 2001



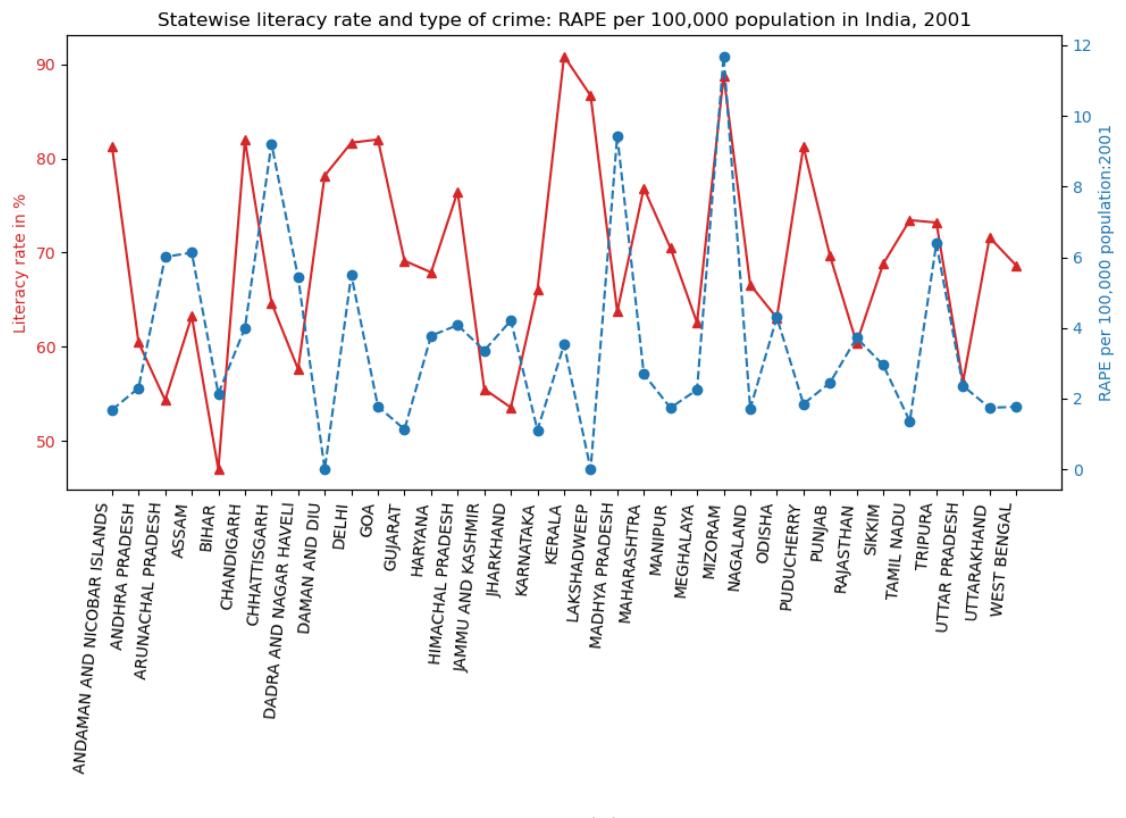
(a)

Statewise literacy rate and type of crime: CULPABLE HOMICIDE NOT AMOUNTING TO MURDER per 100,000 population in India, 2011

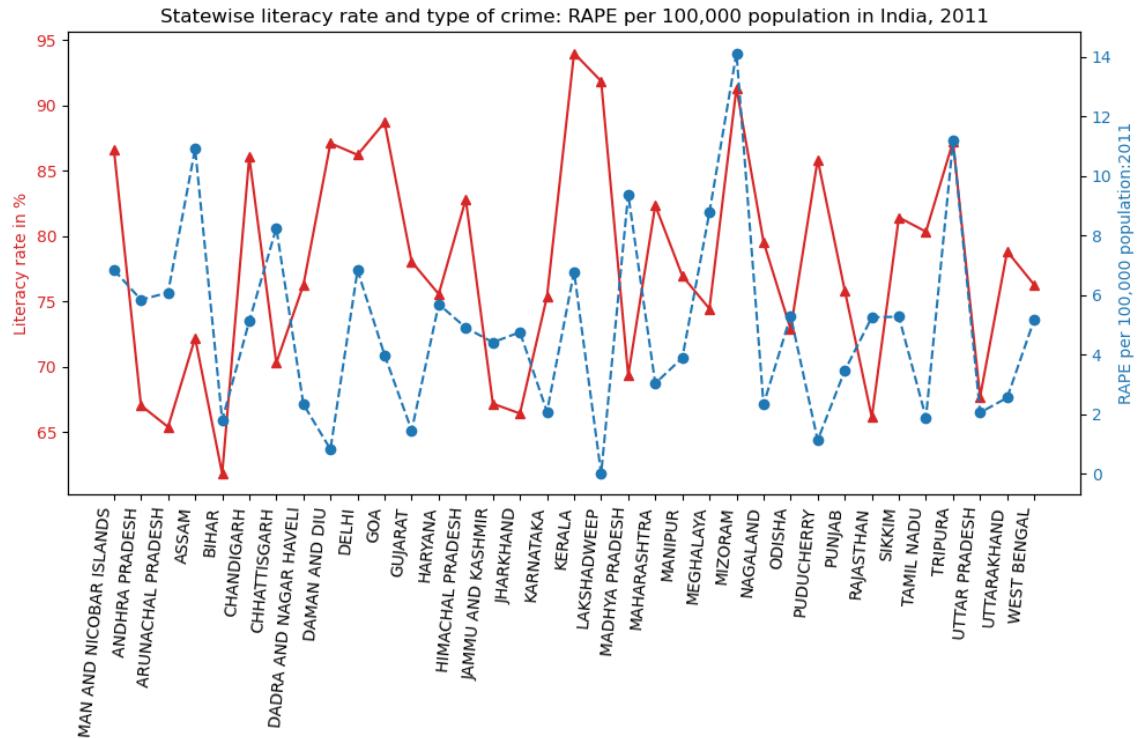


(b)

Fig. 5: State wise literacy and culpable homicide not amounting t to murder (a) 2001 (b) 2011

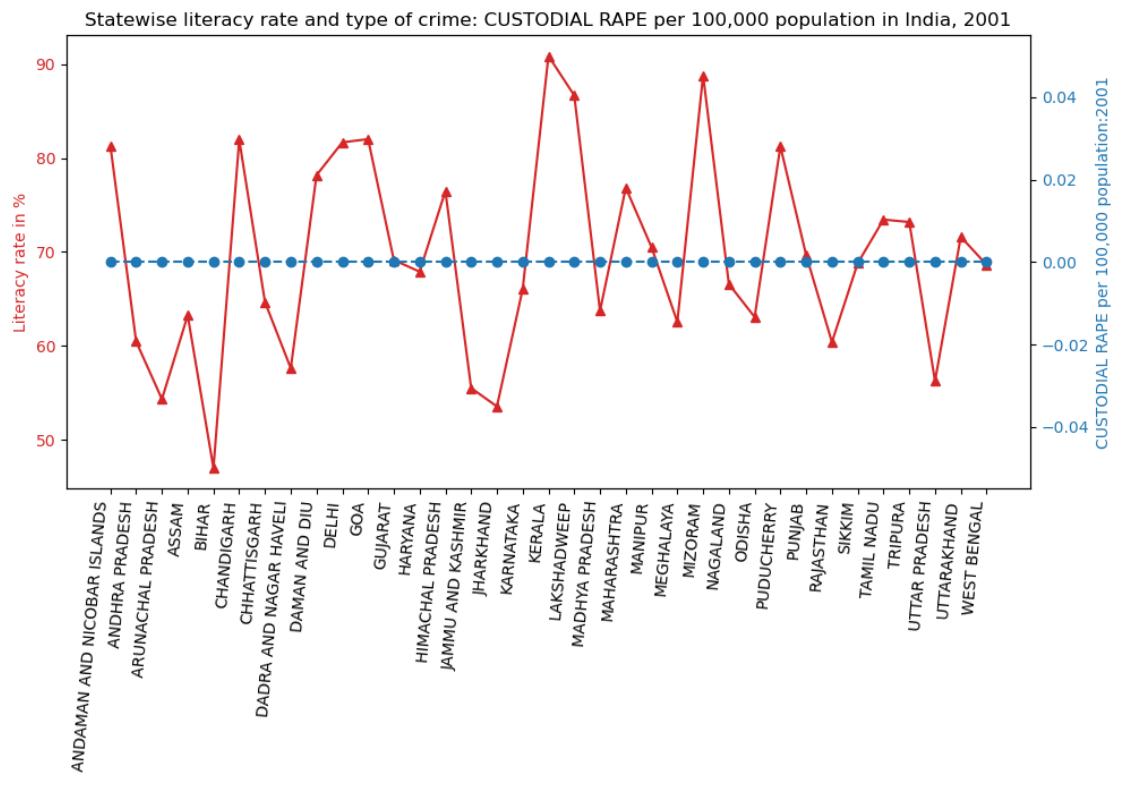


(a)

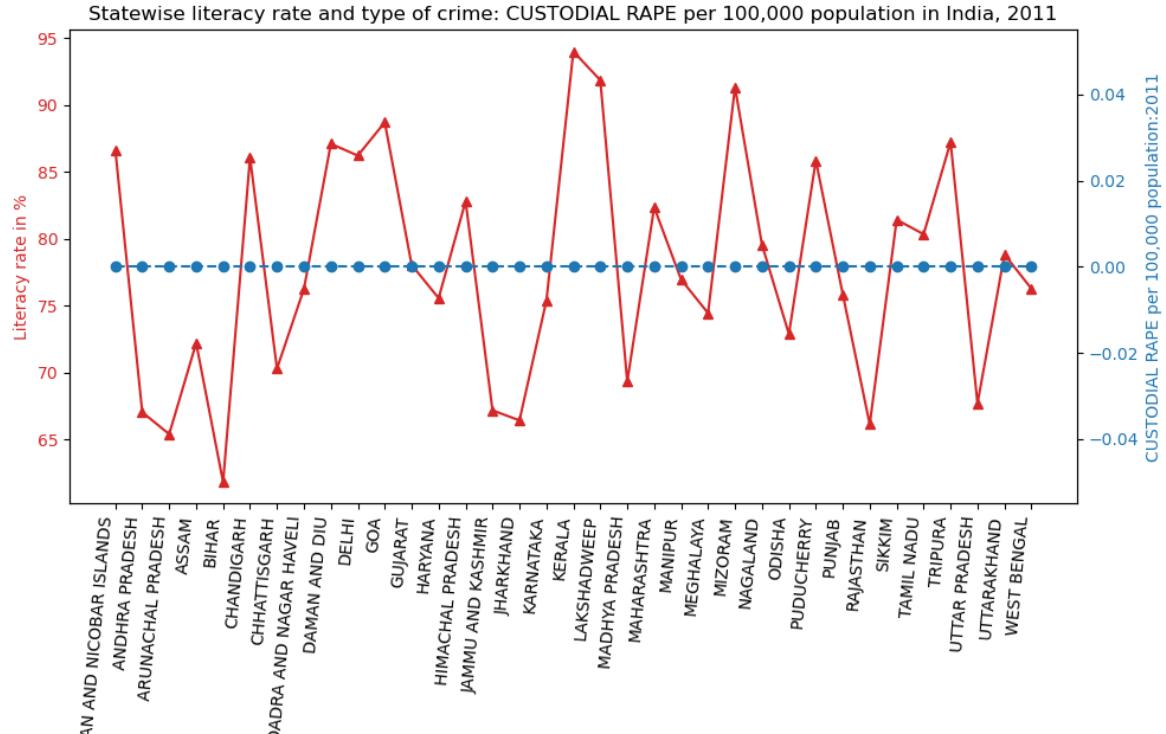


(b)

Fig. 6: State wise literacy and rape (a) 2001 (b) 2011

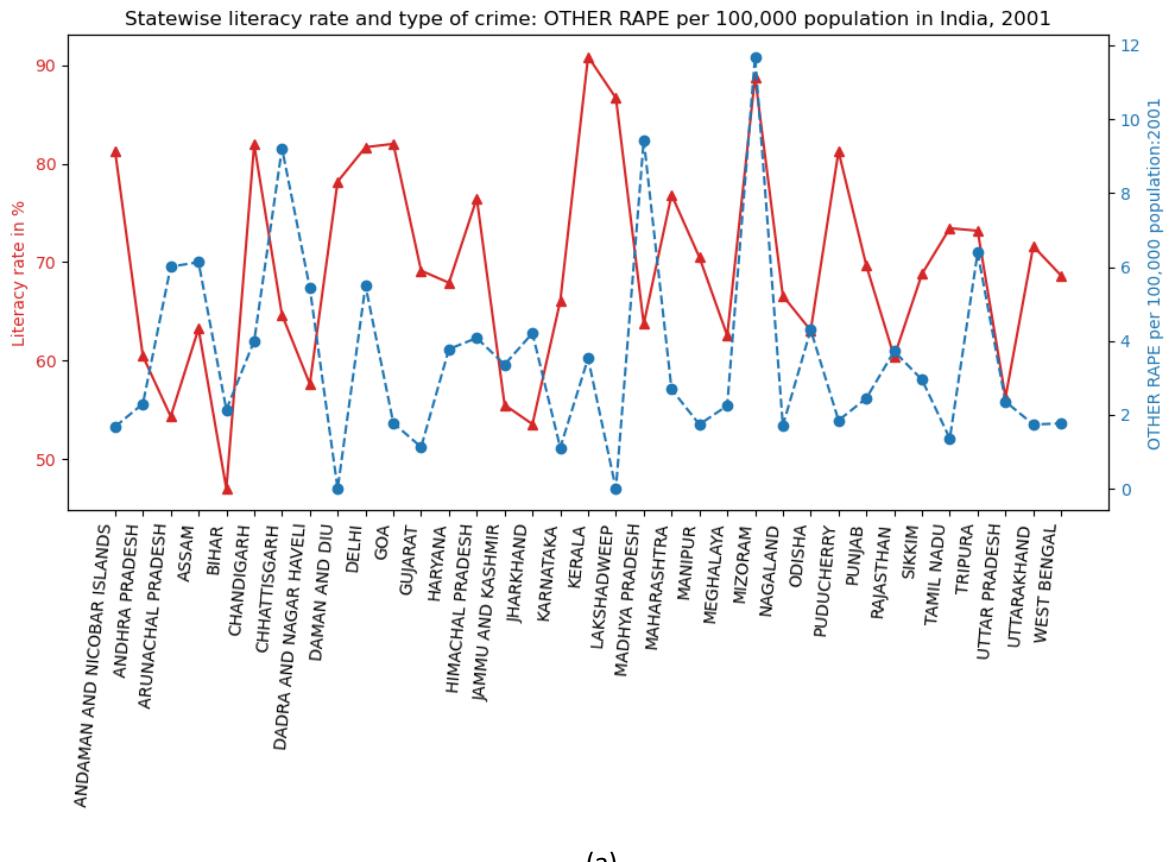


(a)

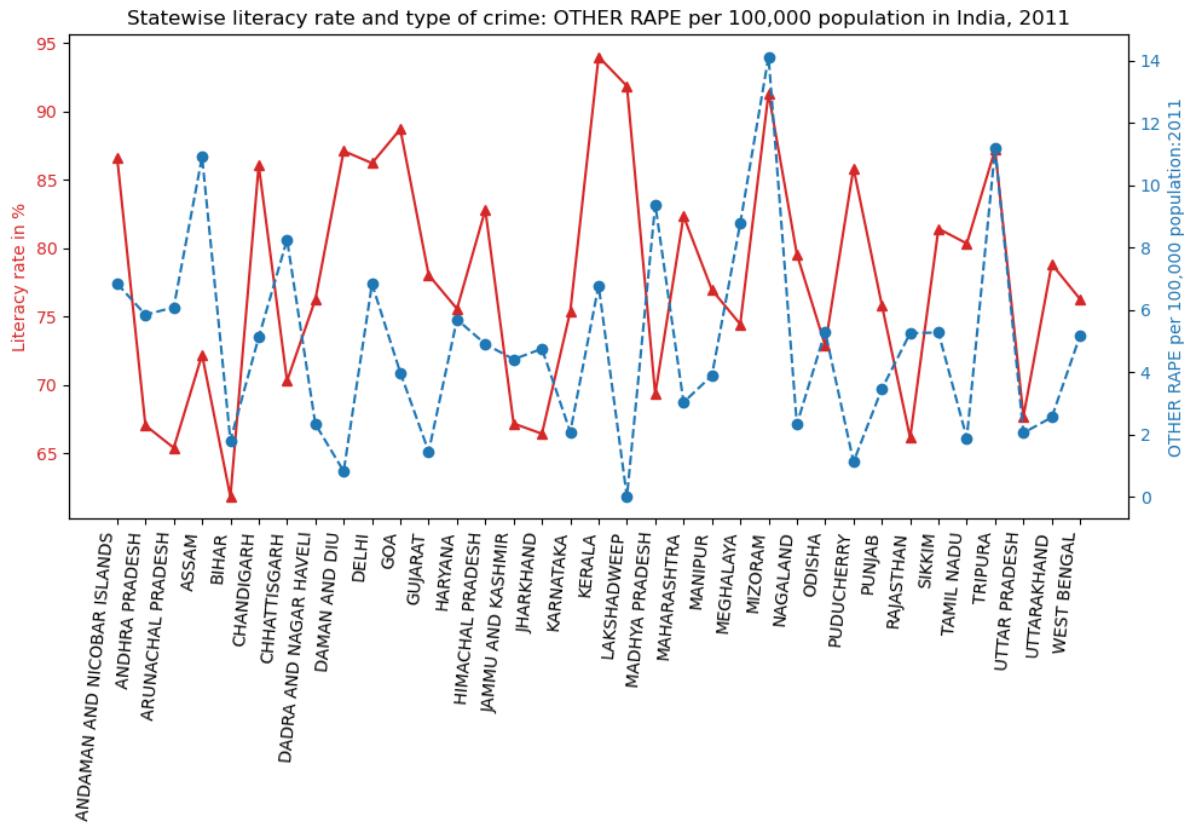


(b)

Fig. 7: State wise literacy and custodial rape (a) 2001 (b) 2011

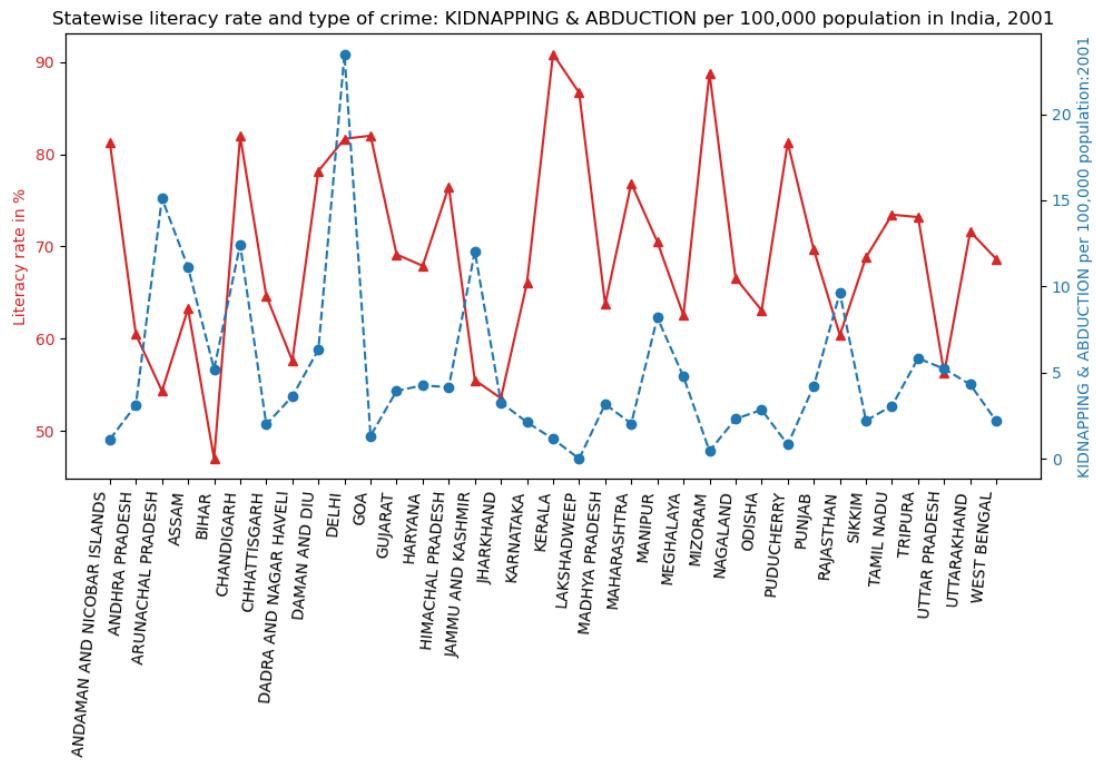


(a)

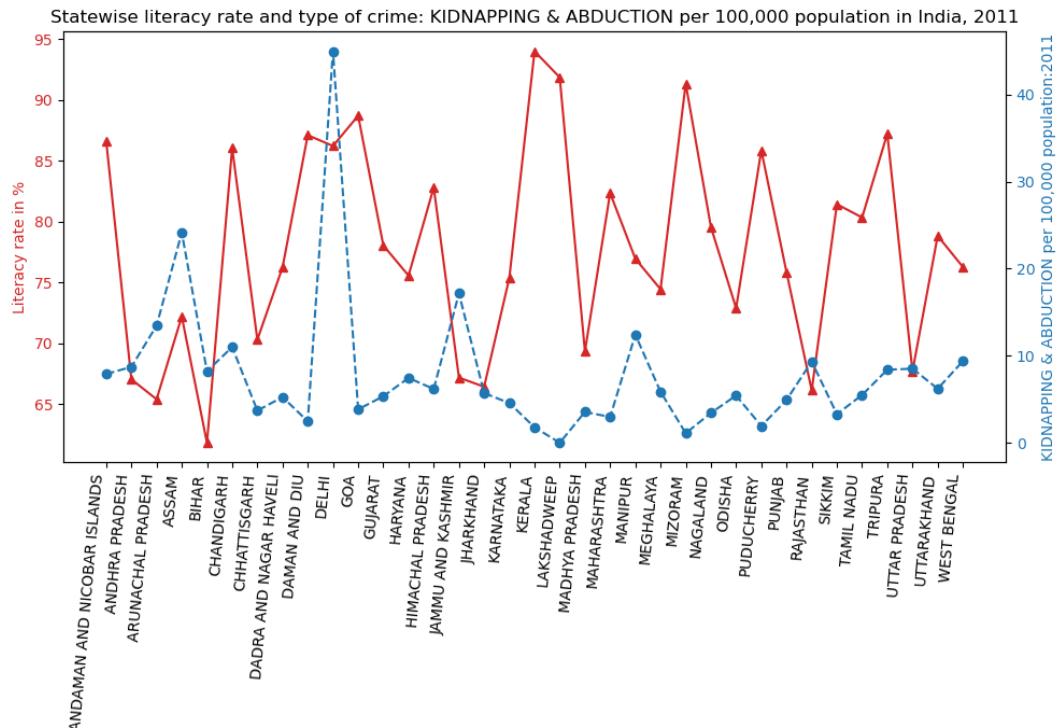


(b)

Fig. 8: State wise literacy and other rape (a) 2001 (b) 2011



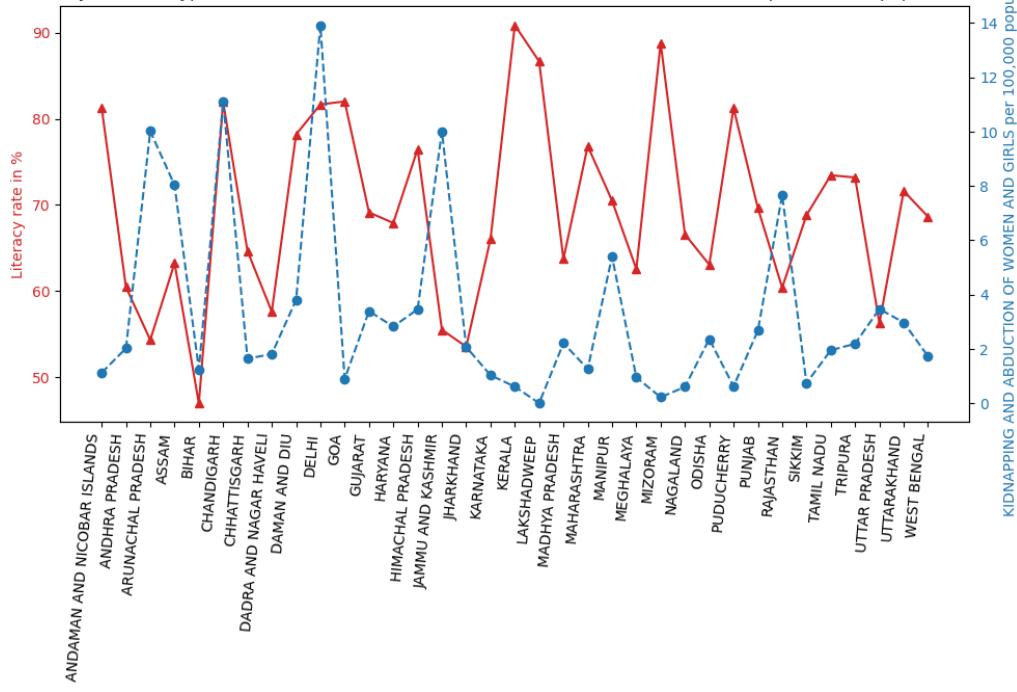
(a)



(b)

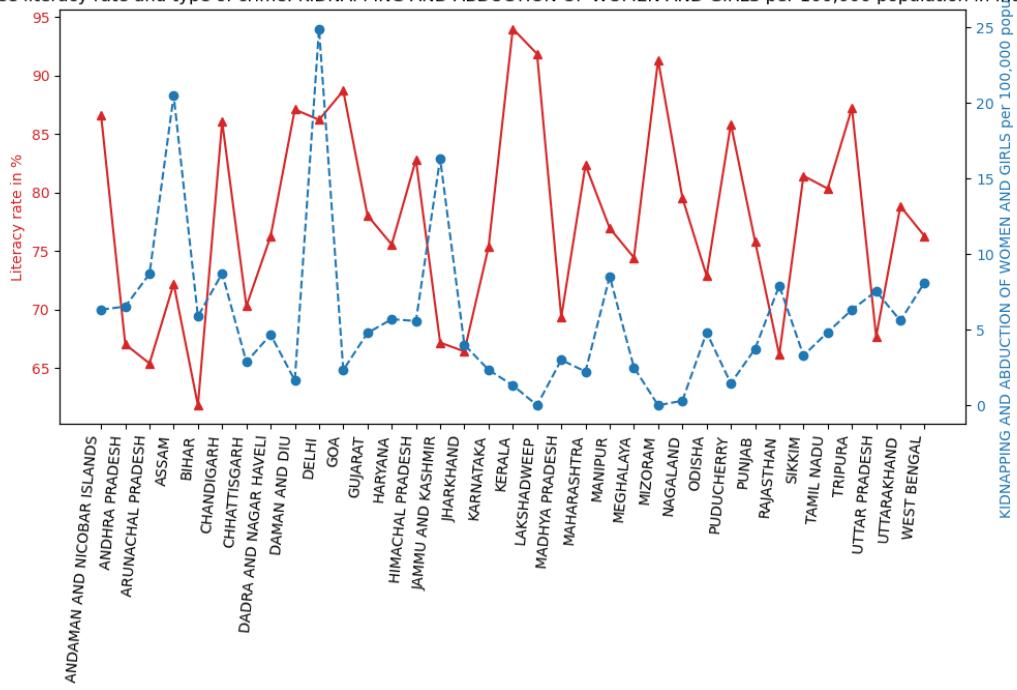
Fig. 9: State wise literacy and kidnapping and abduction (a) 2001 (b) 2011

Statewise literacy rate and type of crime: KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS per 100,000 population in India, 2001



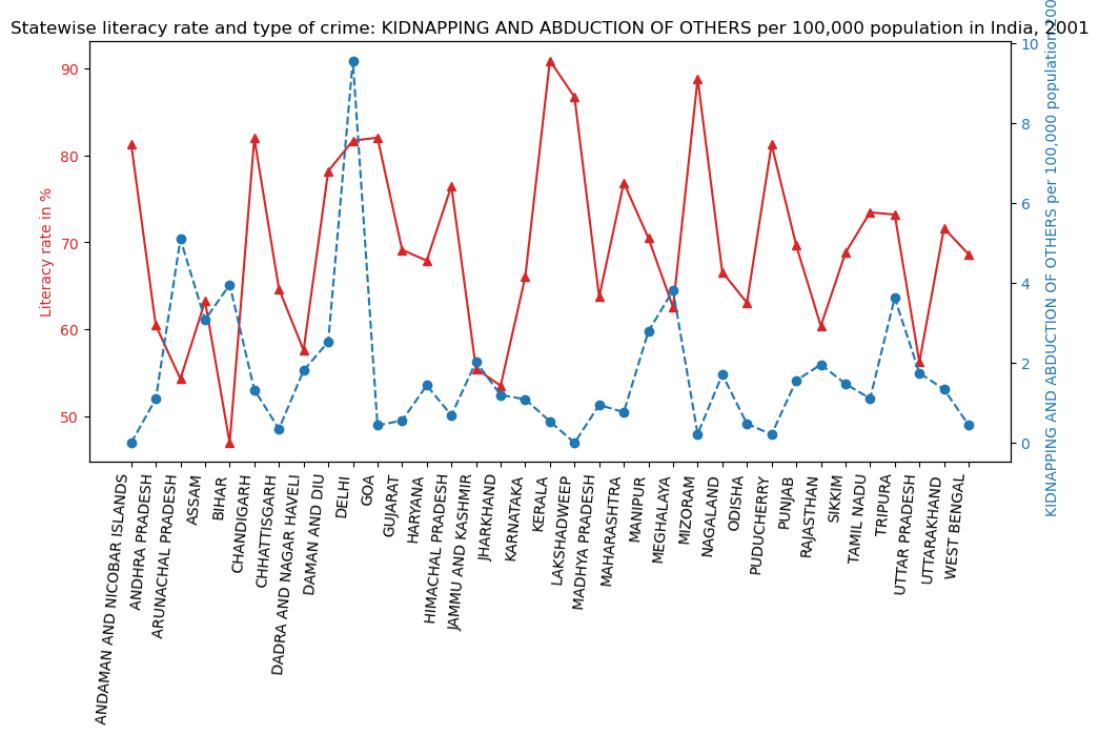
(a)

Statewise literacy rate and type of crime: KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS per 100,000 population in India, 2011

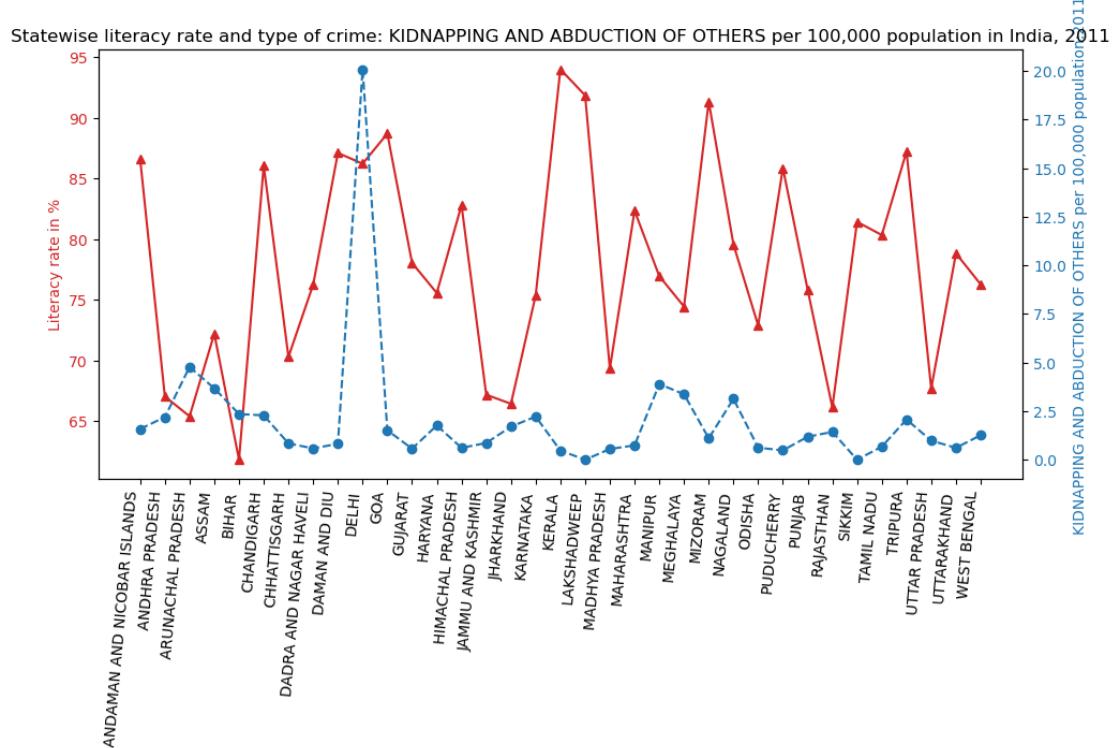


(b)

Fig. 10: State wise literacy and kidnapping and abduction of women and girls (a) 2001 (b) 2011

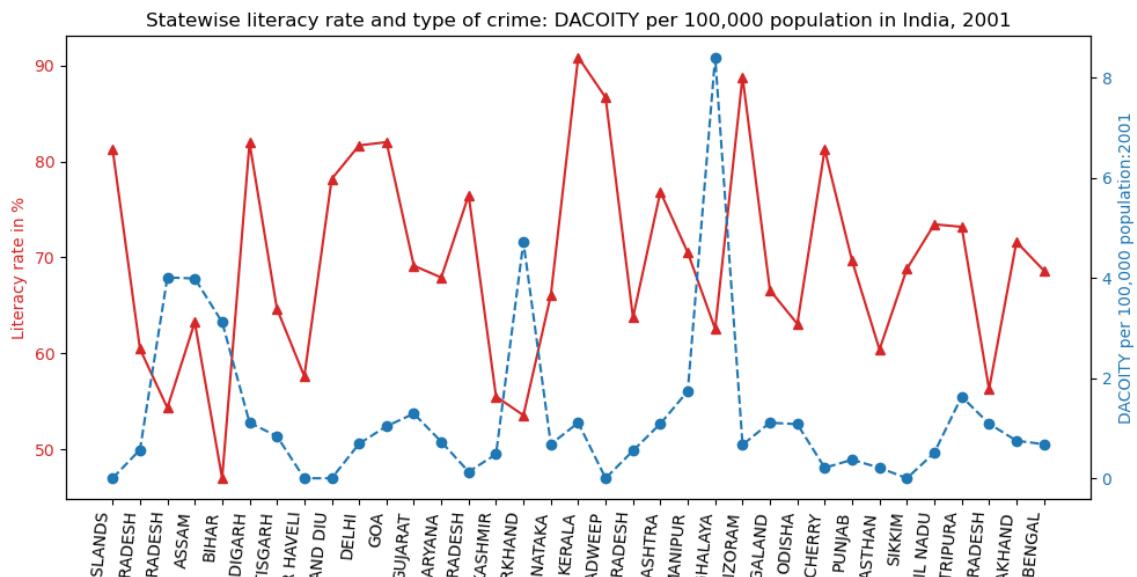


(a)

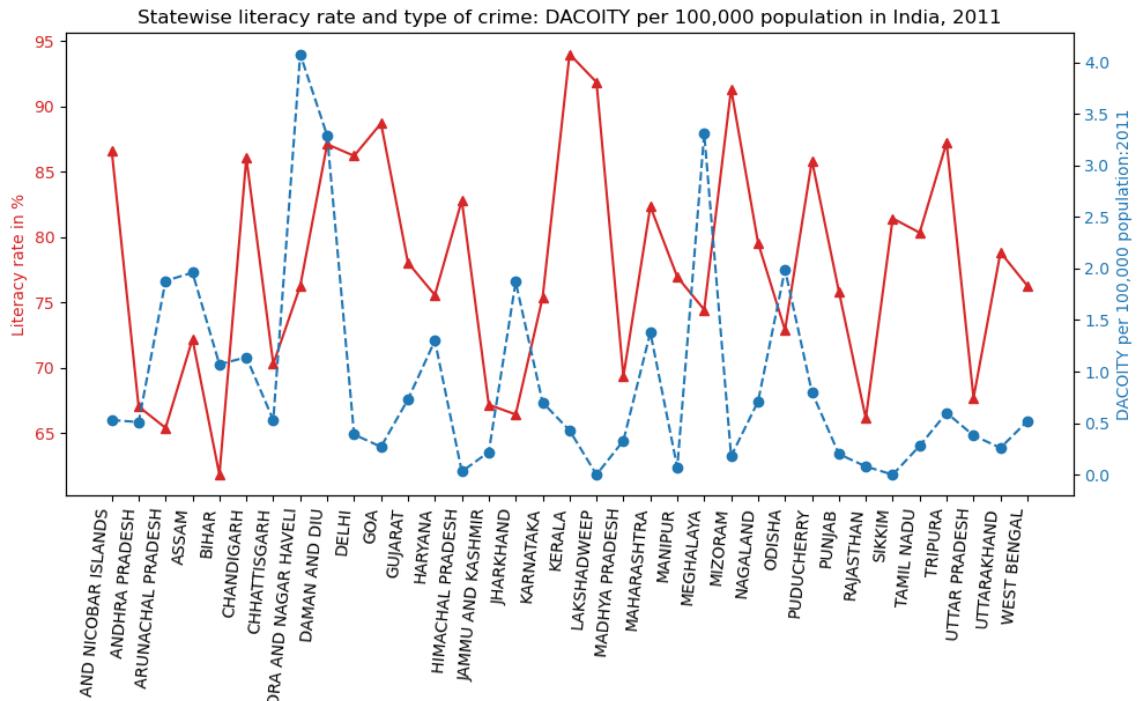


(b)

Fig. 11: State wise literacy and kidnapping and abduction of others (a) 2001 (b) 2011



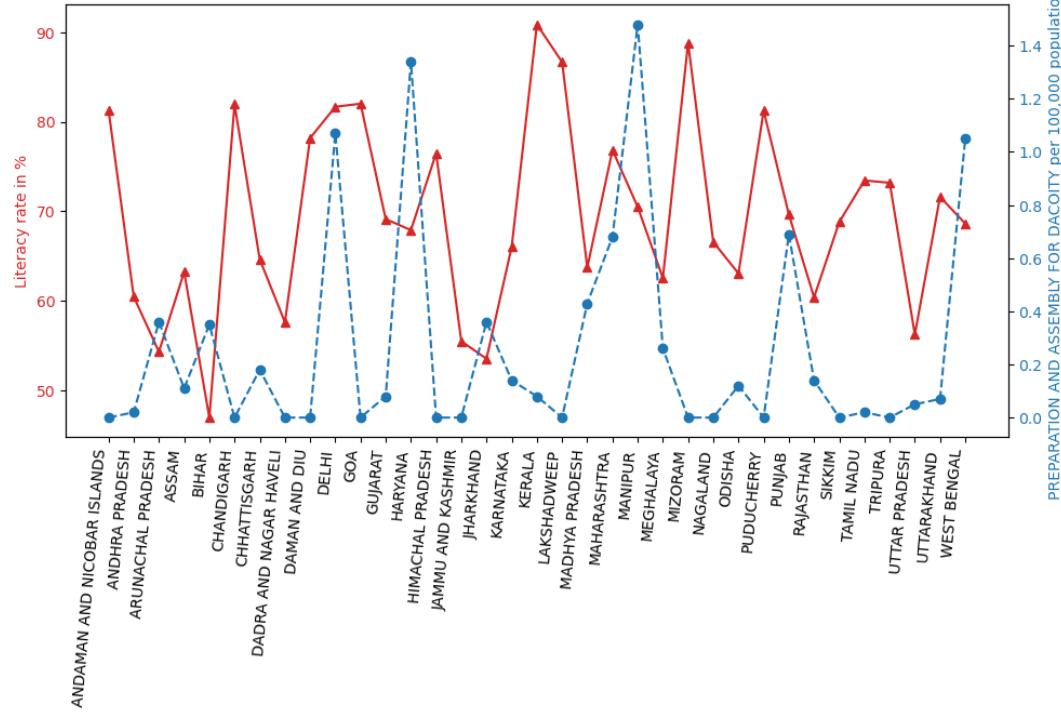
(a)



(b)

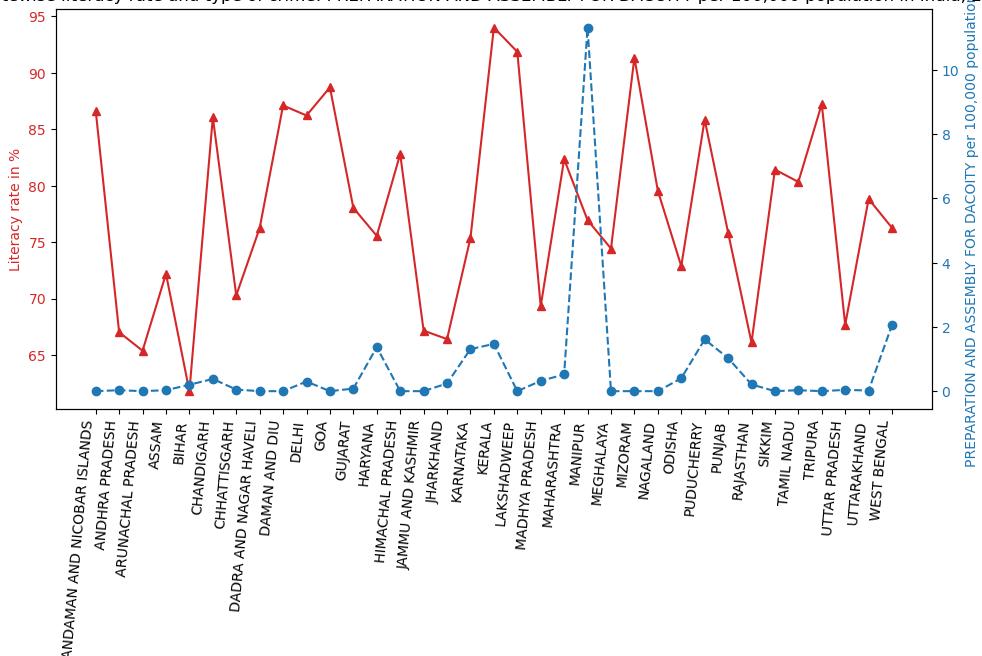
Fig. 12: State wise literacy and dacoity (a) 2001 (b) 2011

Statewise literacy rate and type of crime: PREPARATION AND ASSEMBLY FOR Dacoity per 100,000 population in India, 2001



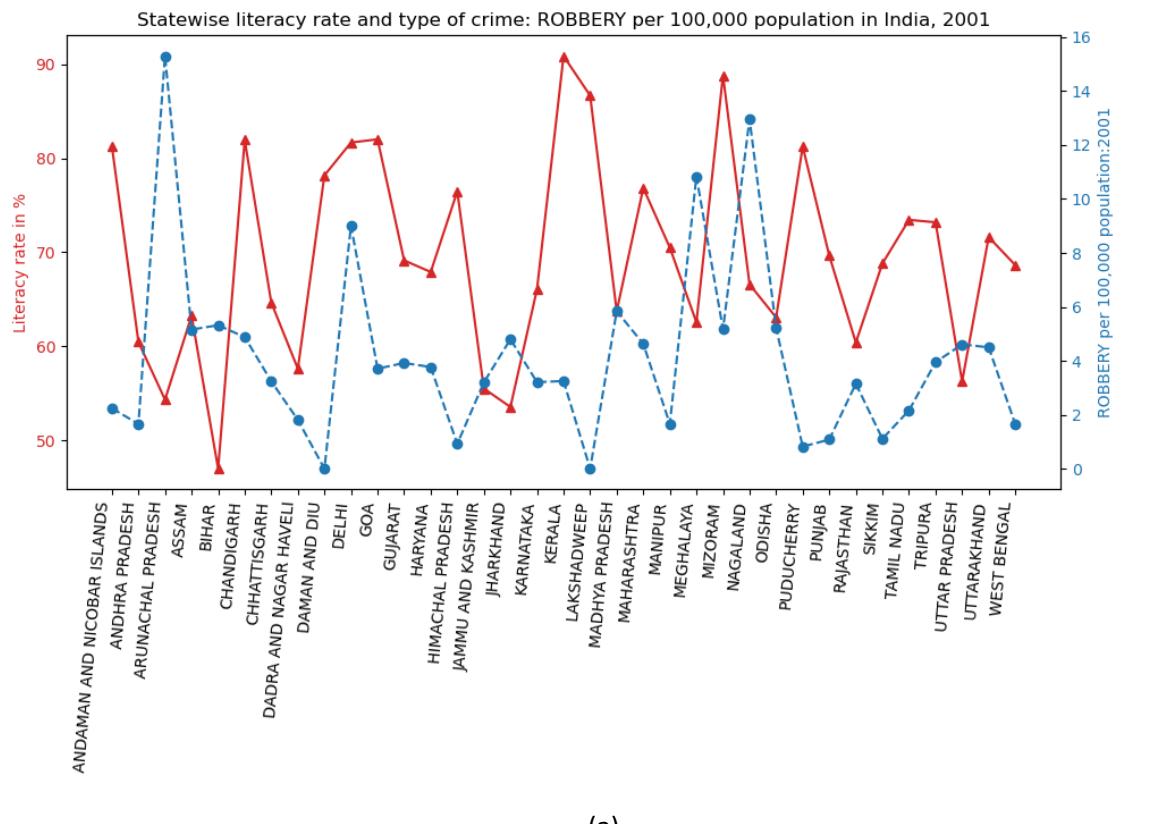
(a)

Statewise literacy rate and type of crime: PREPARATION AND ASSEMBLY FOR Dacoity per 100,000 population in India, 2011

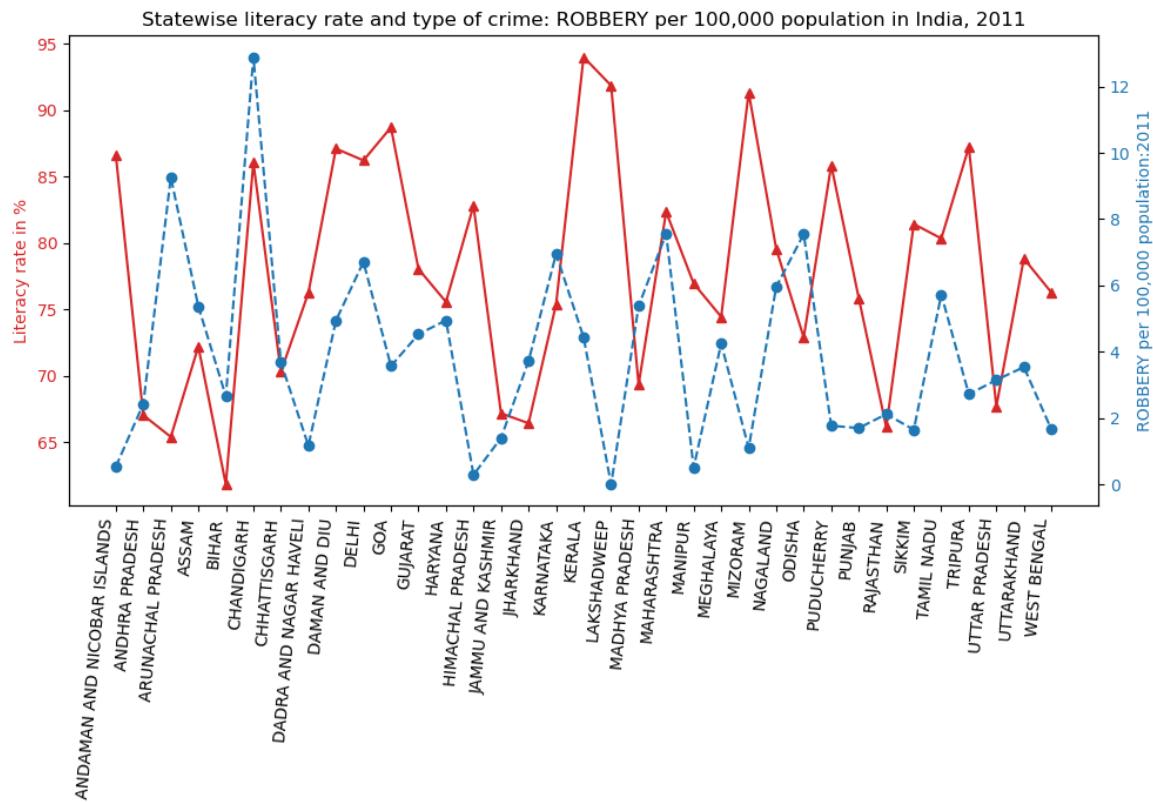


(b)

Fig. 13: State wise literacy and preparation and assembly for dacoity (a) 2001 (b) 2011



(a)



(b)

Fig. 14: State wise literacy and robbery (a) 2001 (b) 2011

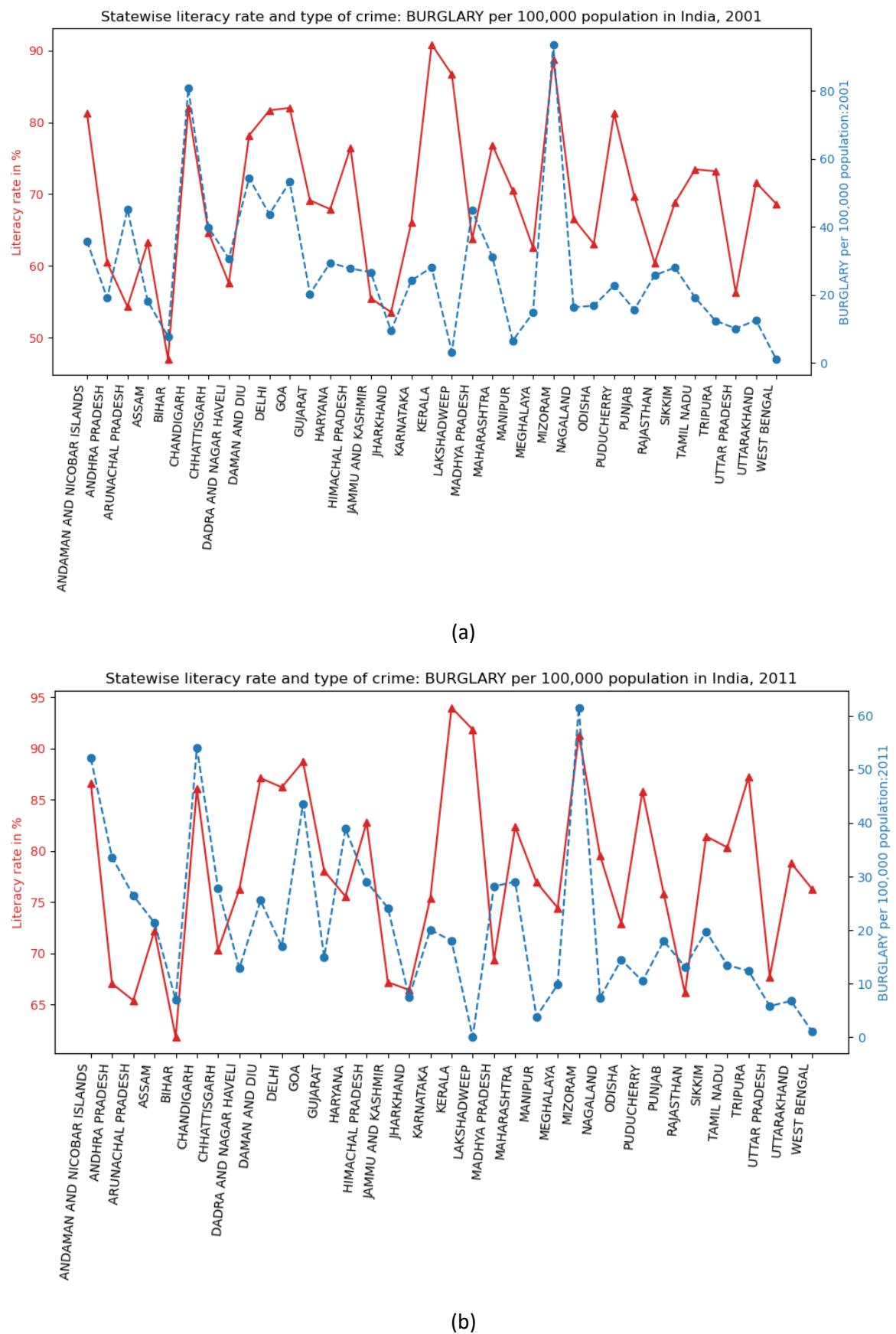
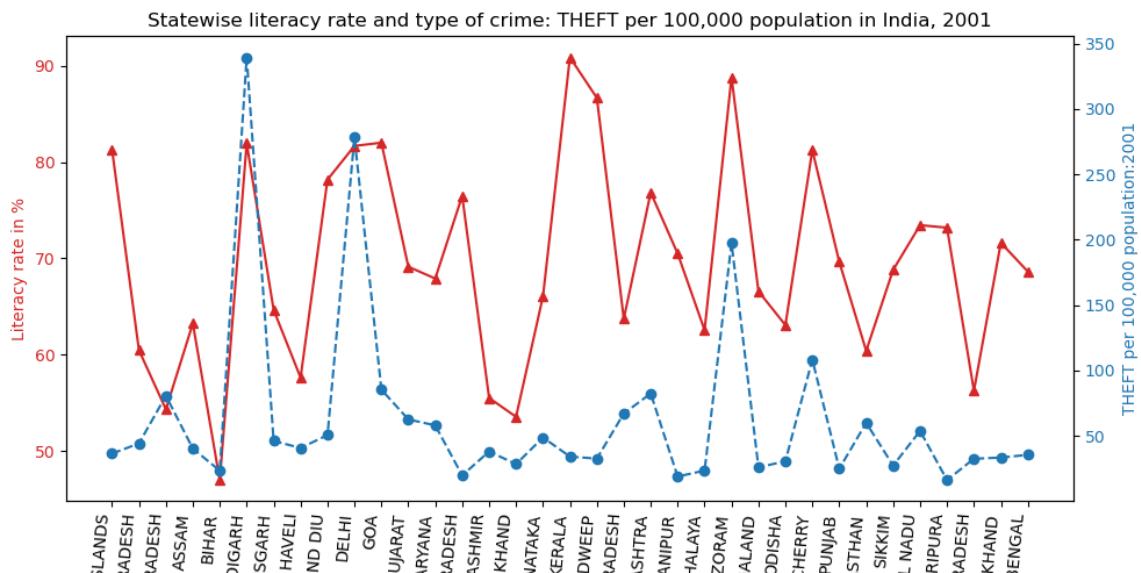
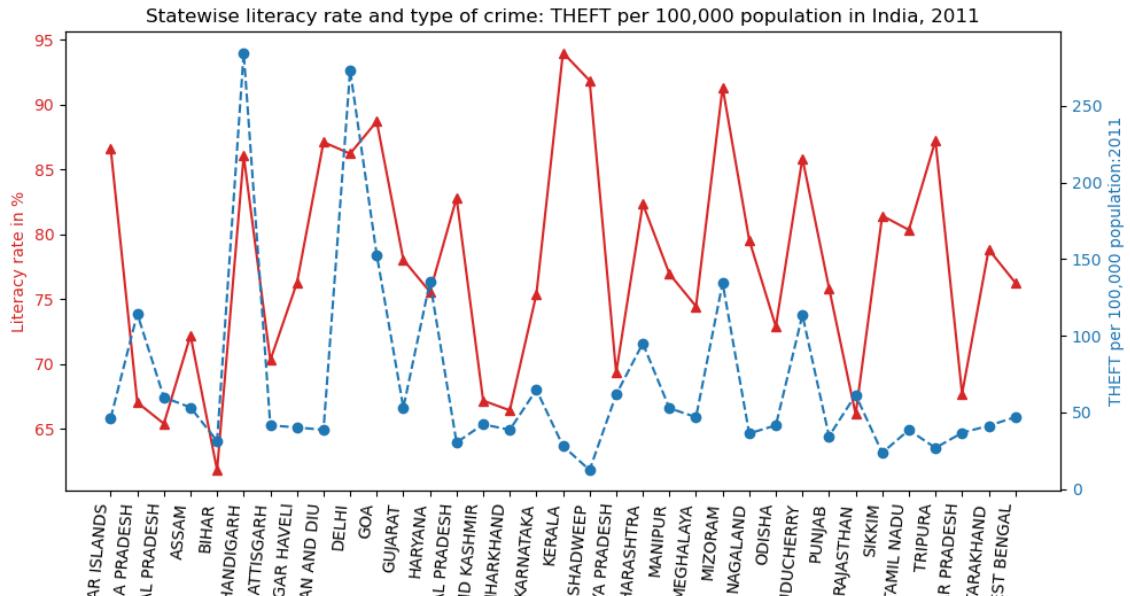


Fig. 15: State wise literacy and burglary (a) 2001 (b) 2011

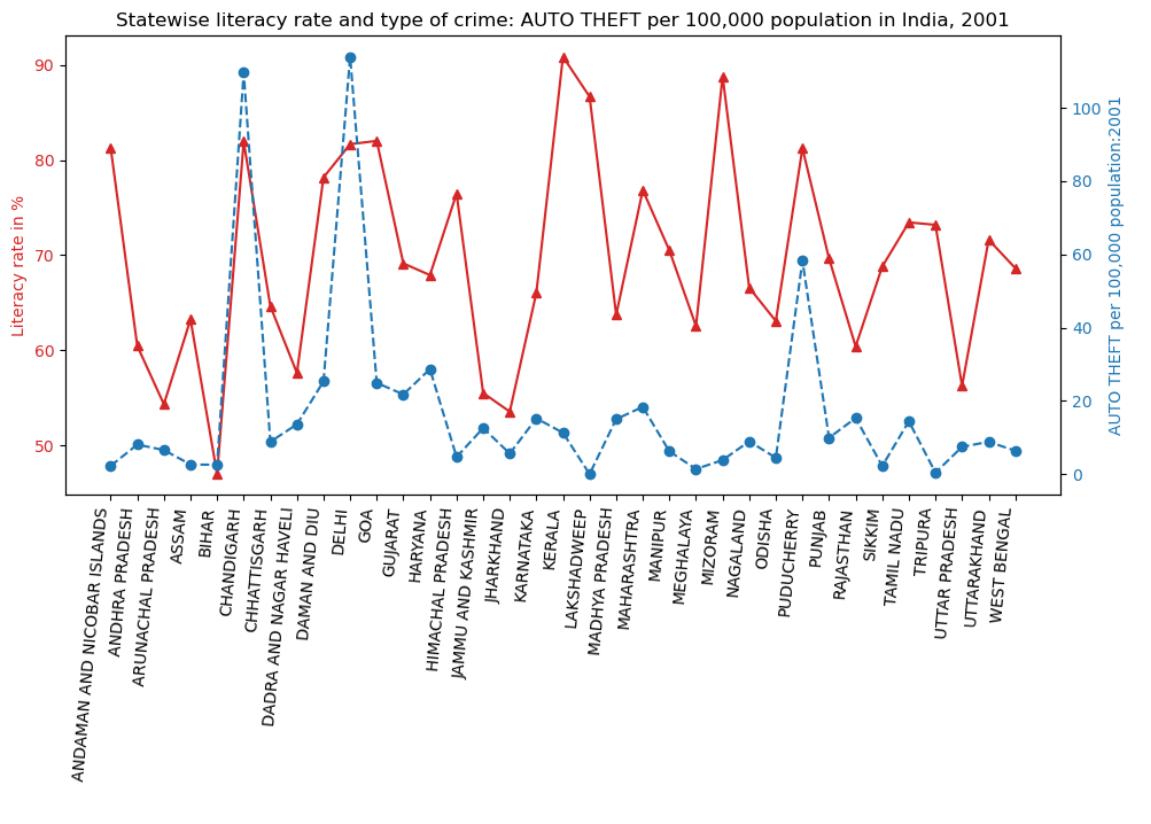


(a)

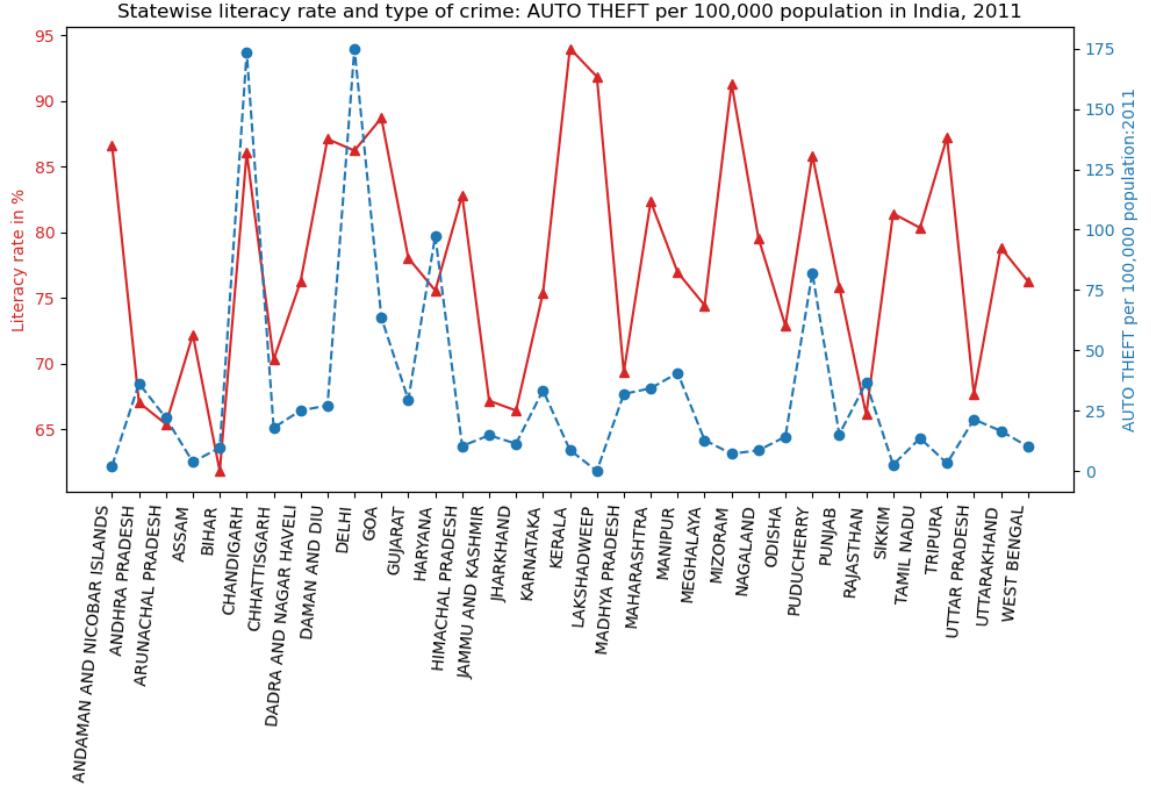


(b)

Fig. 16: State wise literacy and theft (a) 2001 (b) 2011

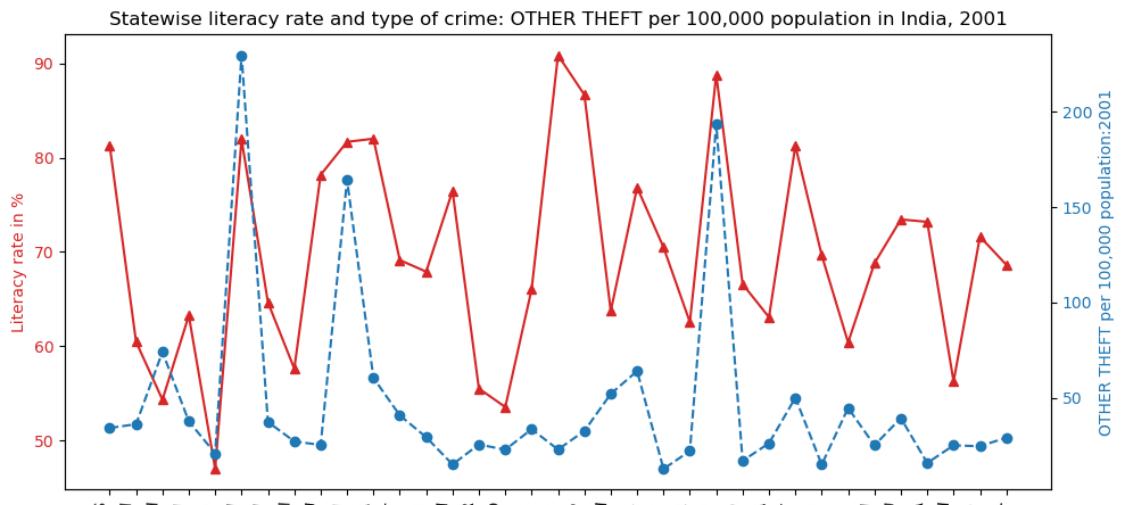


(a)

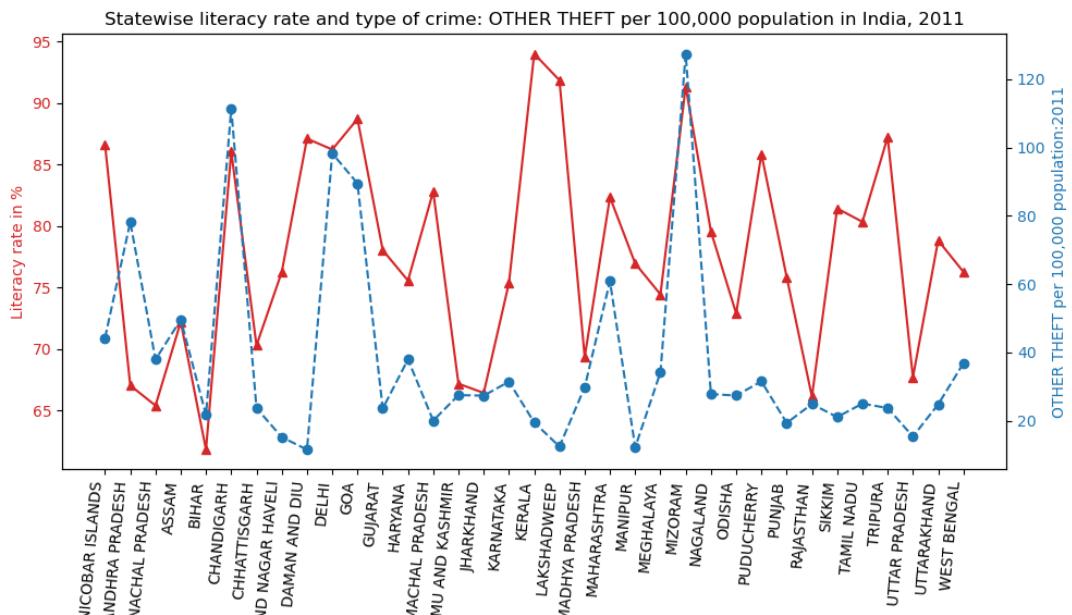


(b)

Fig. 17: State wise literacy and auto theft (a) 2001 (b) 2011

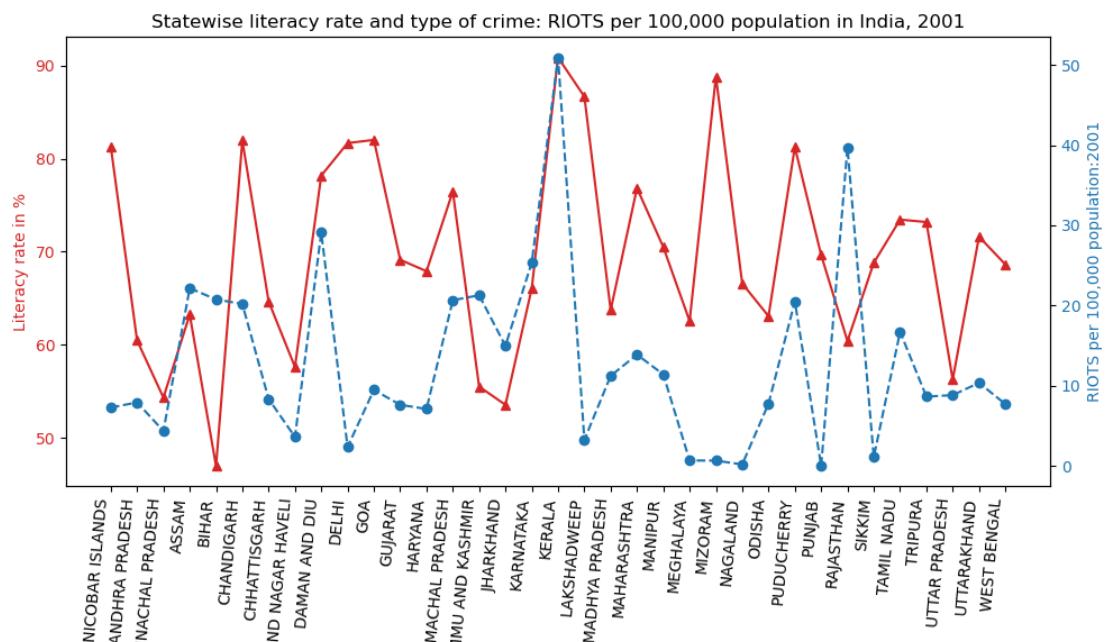


(a)

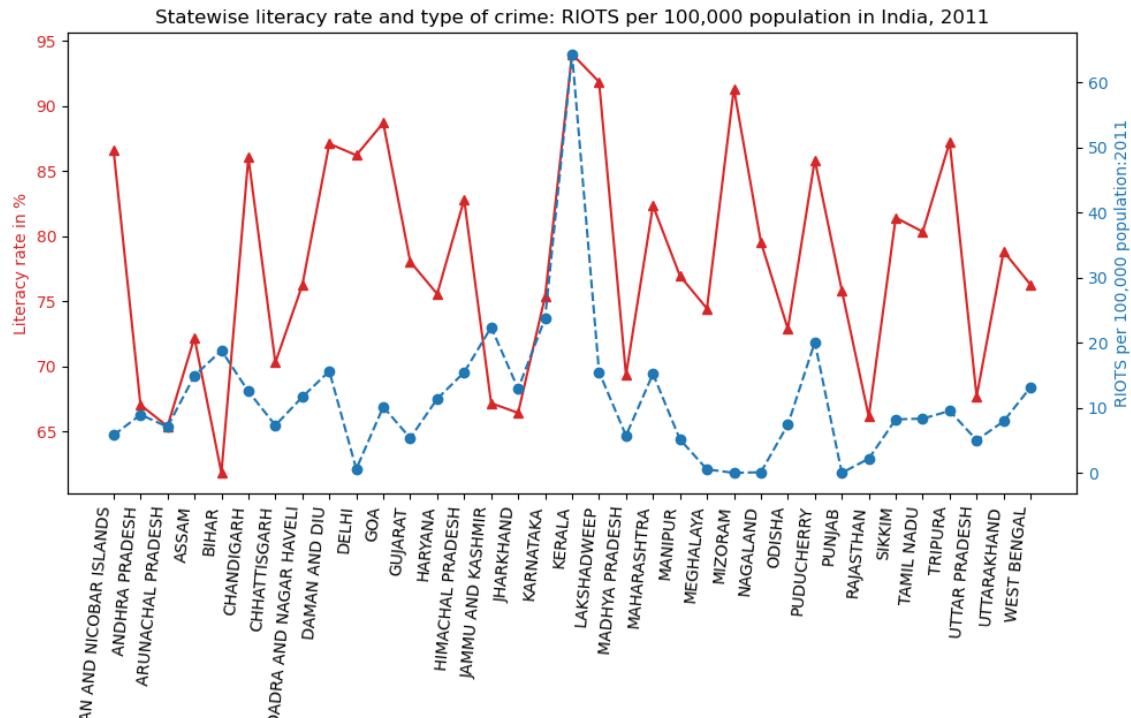


(b)

Fig. 18: State wise literacy and other theft (a) 2001 (b) 2011

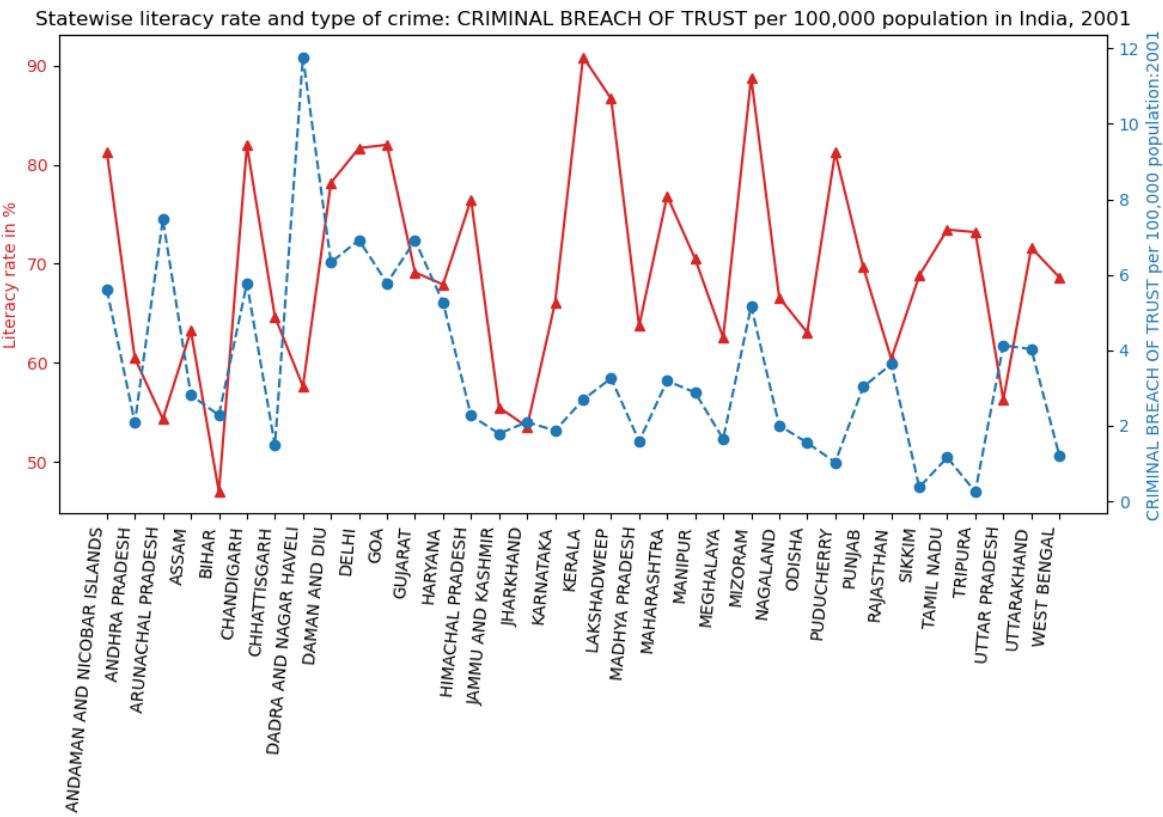


(a)

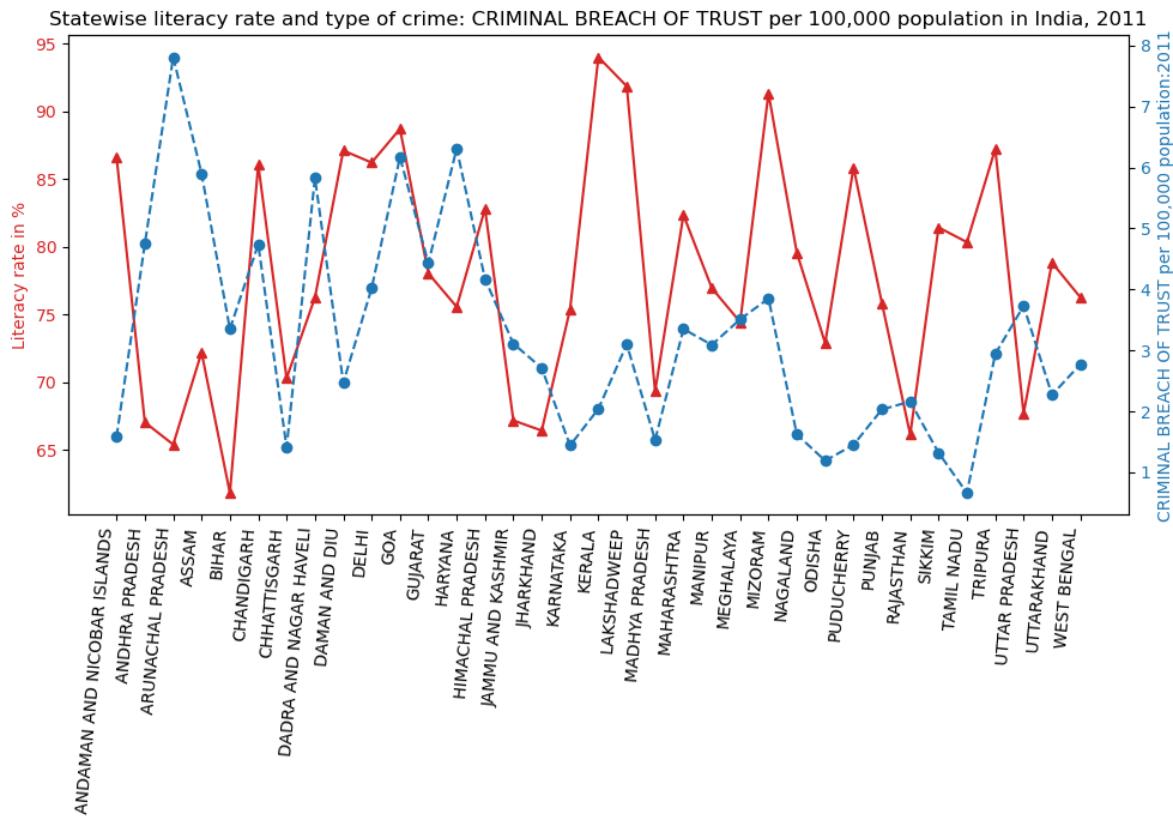


(b)

Fig. 19: State wise literacy and riots (a) 2001 (b) 2011

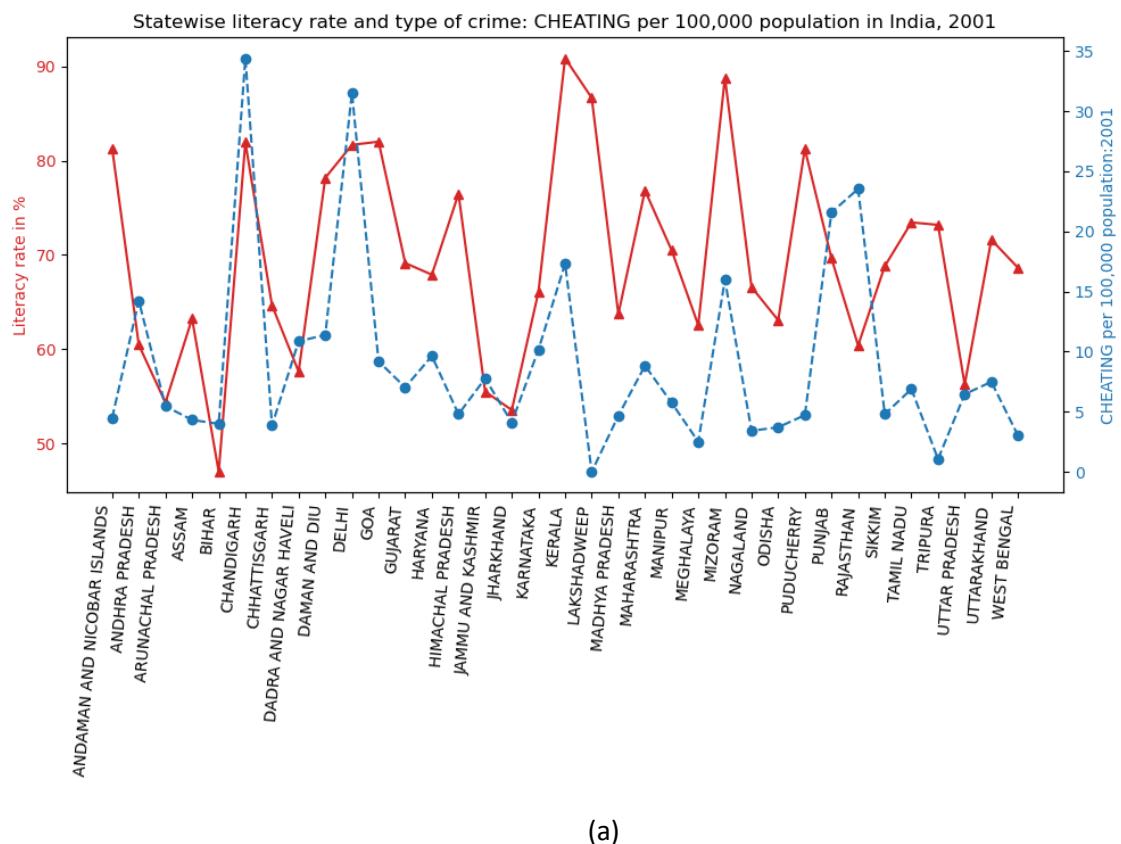


(a)

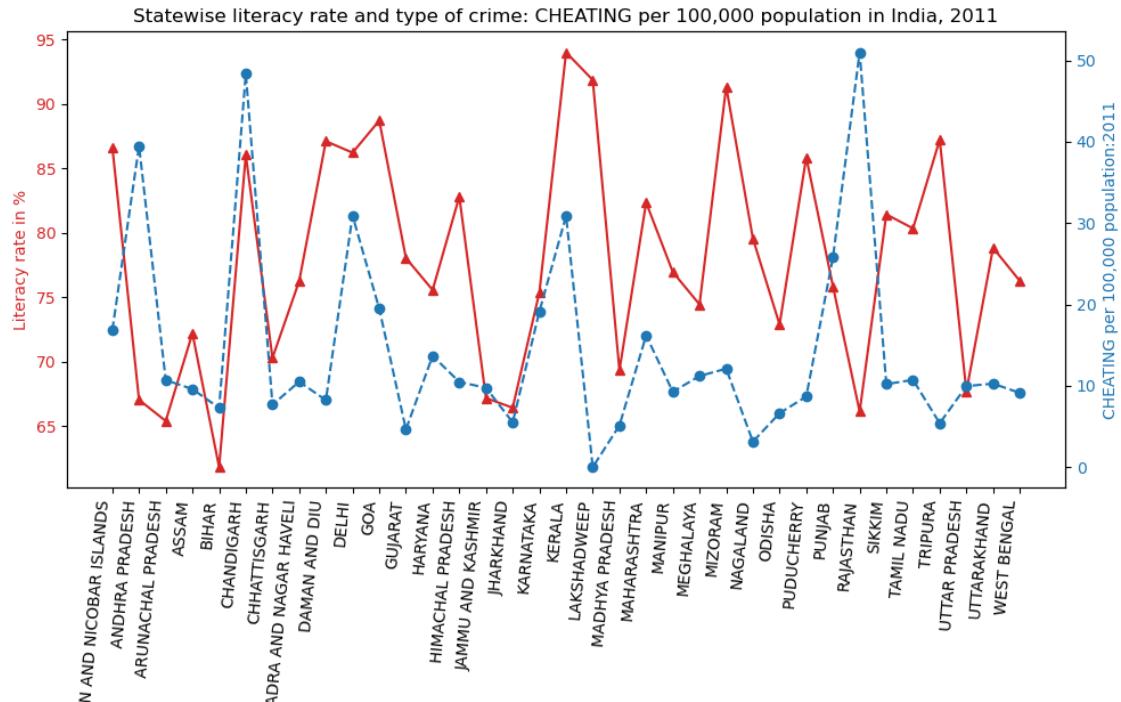


(b)

Fig. 20: State wise literacy and criminal breach of trust (a) 2001 (b) 2011

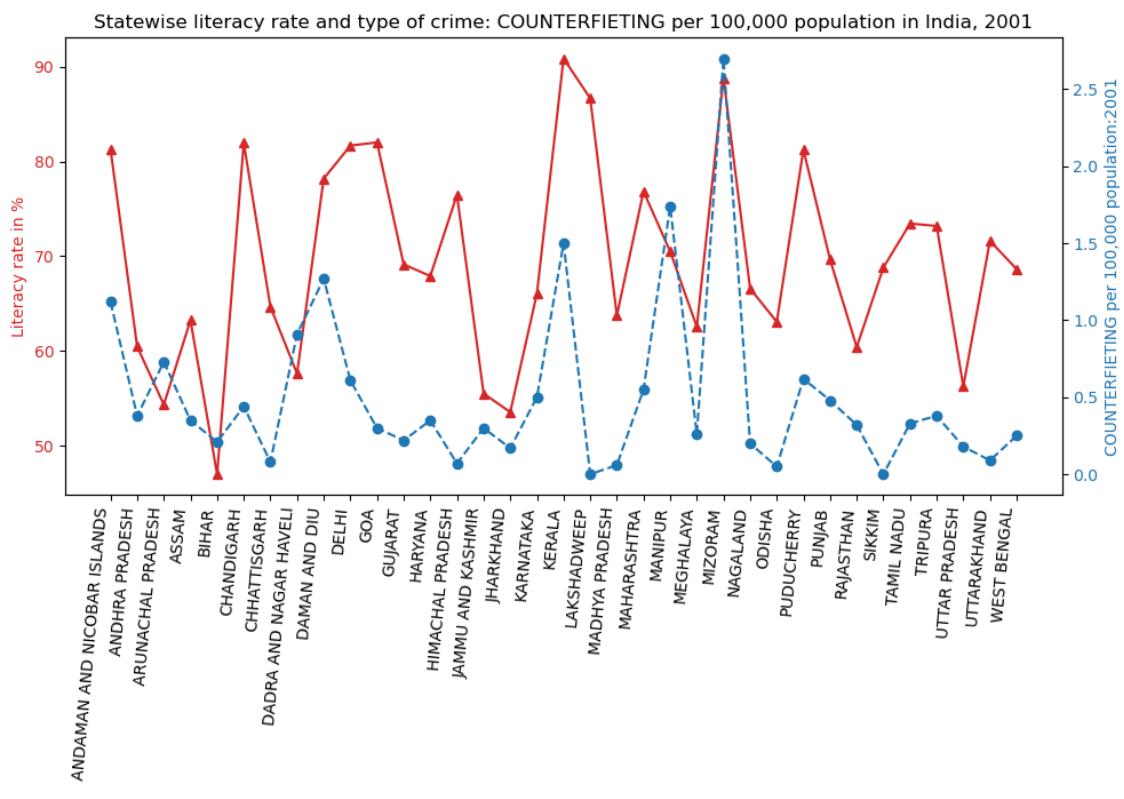


(a)

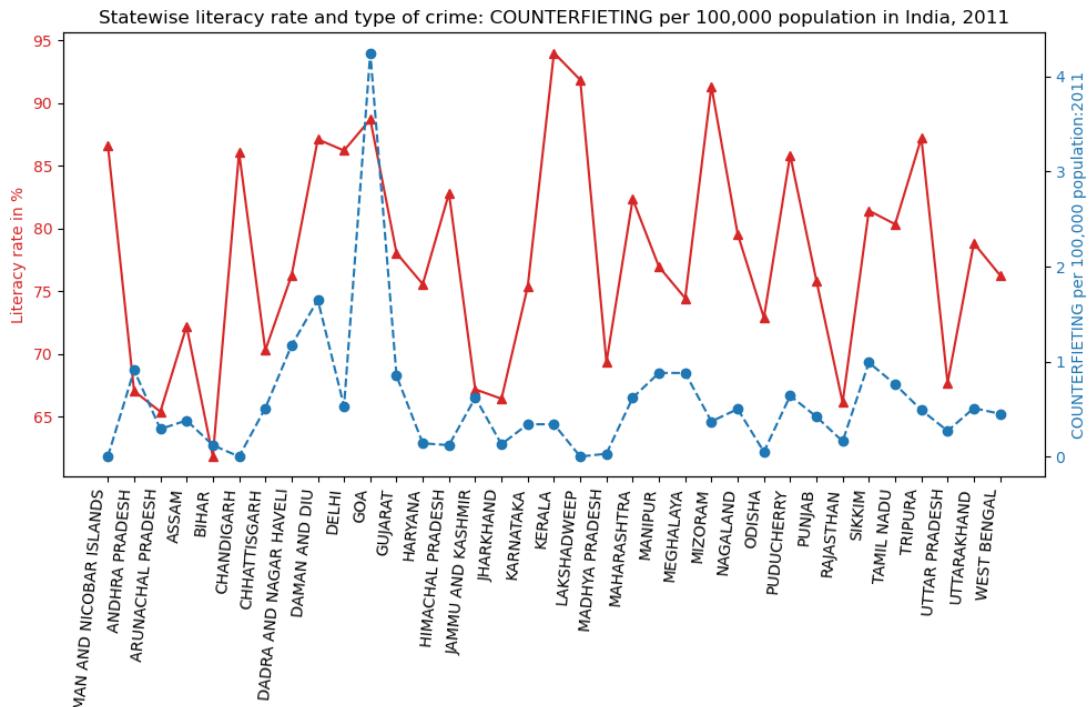


(b)

Fig. 21: State wise literacy and cheating (a) 2001 (b) 2011

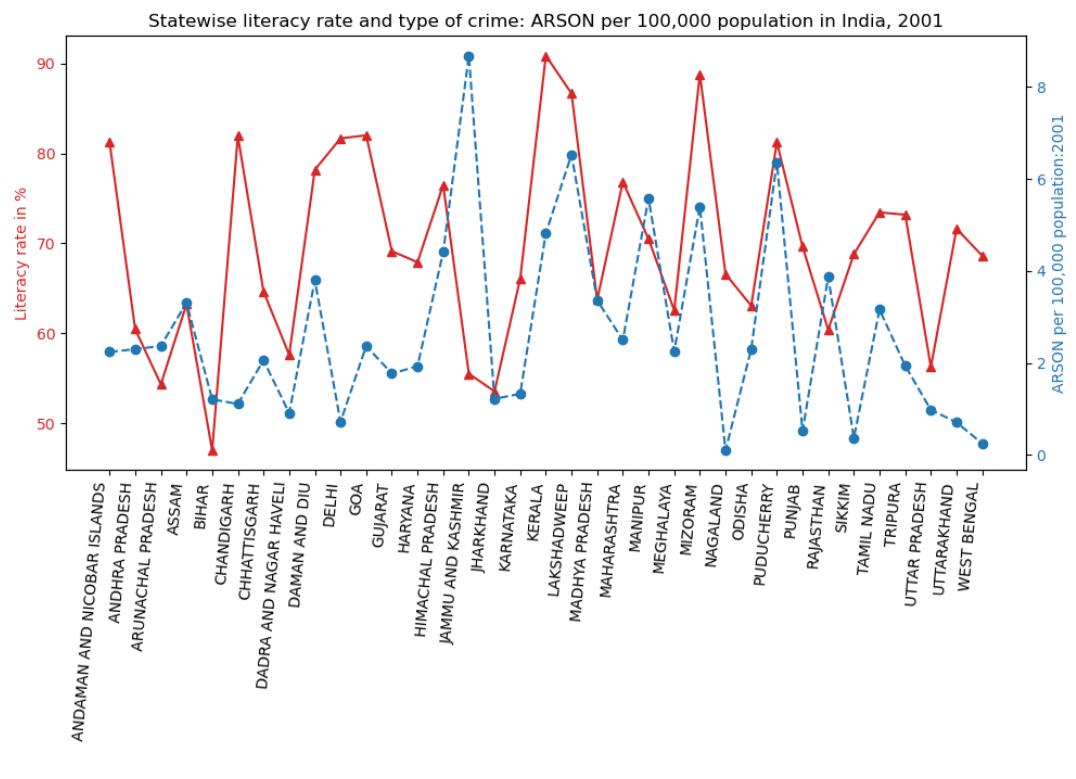


(a)

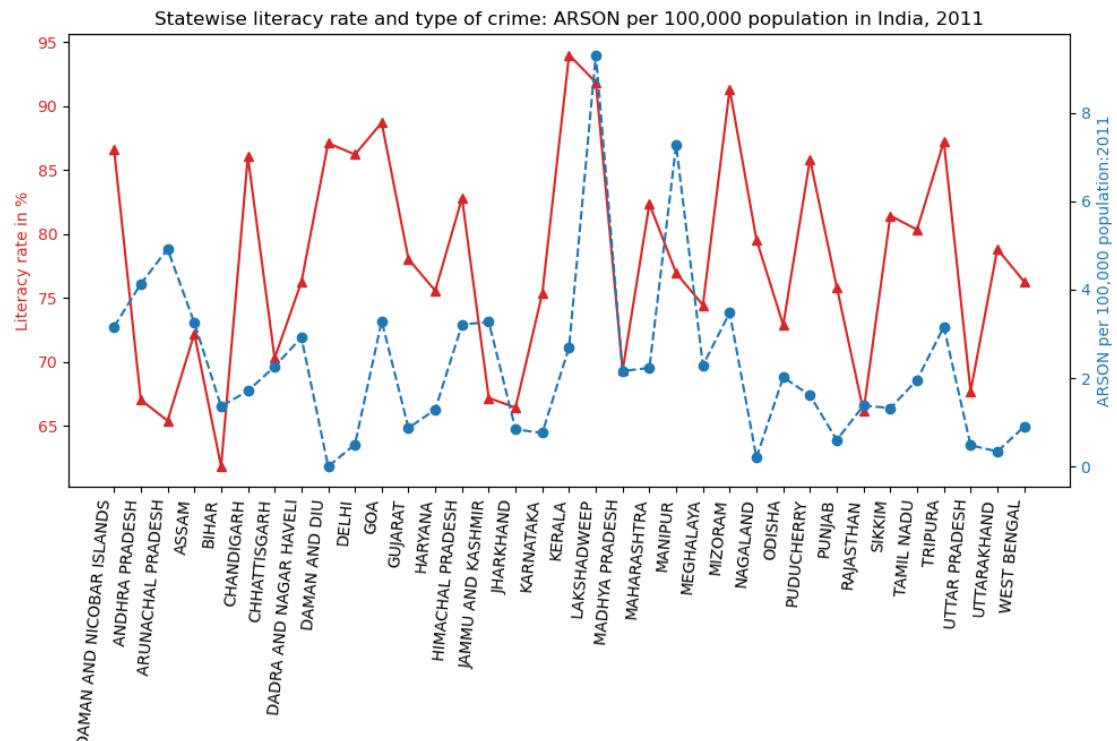


(b)

Fig. 22: State wise literacy and counterfeiting (a) 2001 (b) 2011

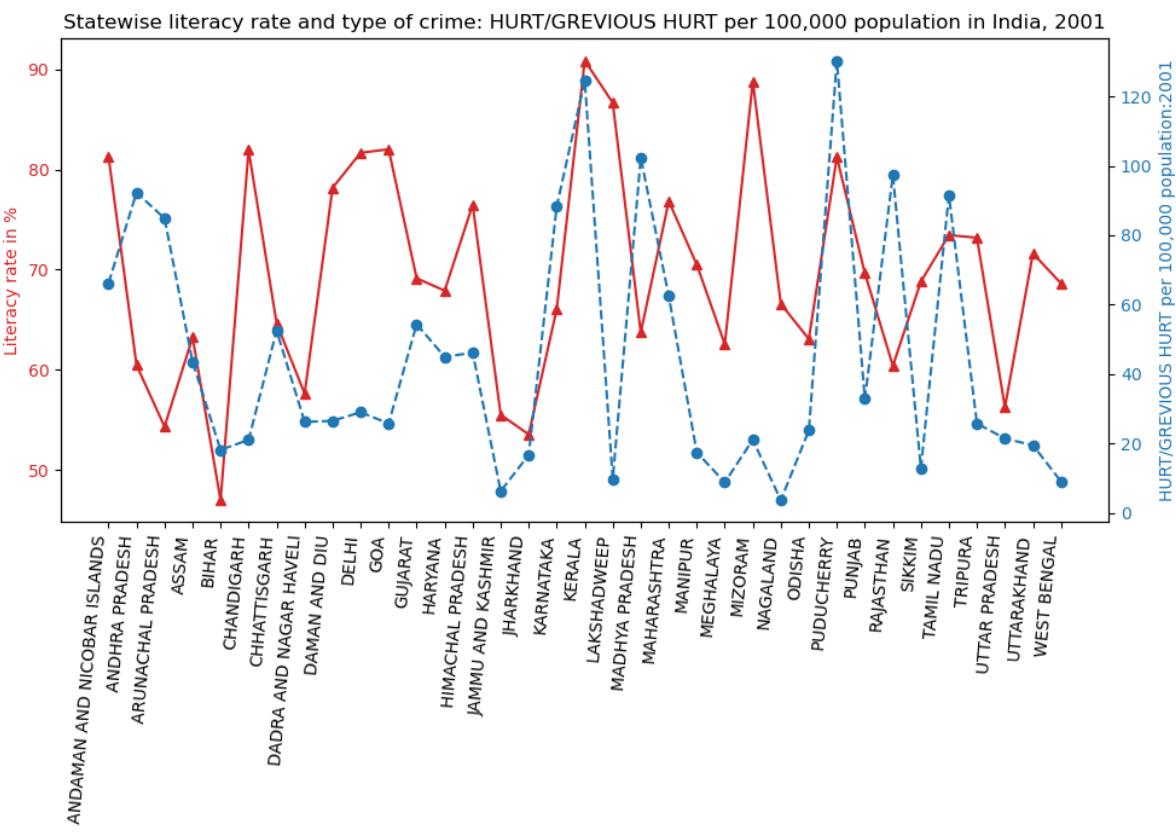


(a)

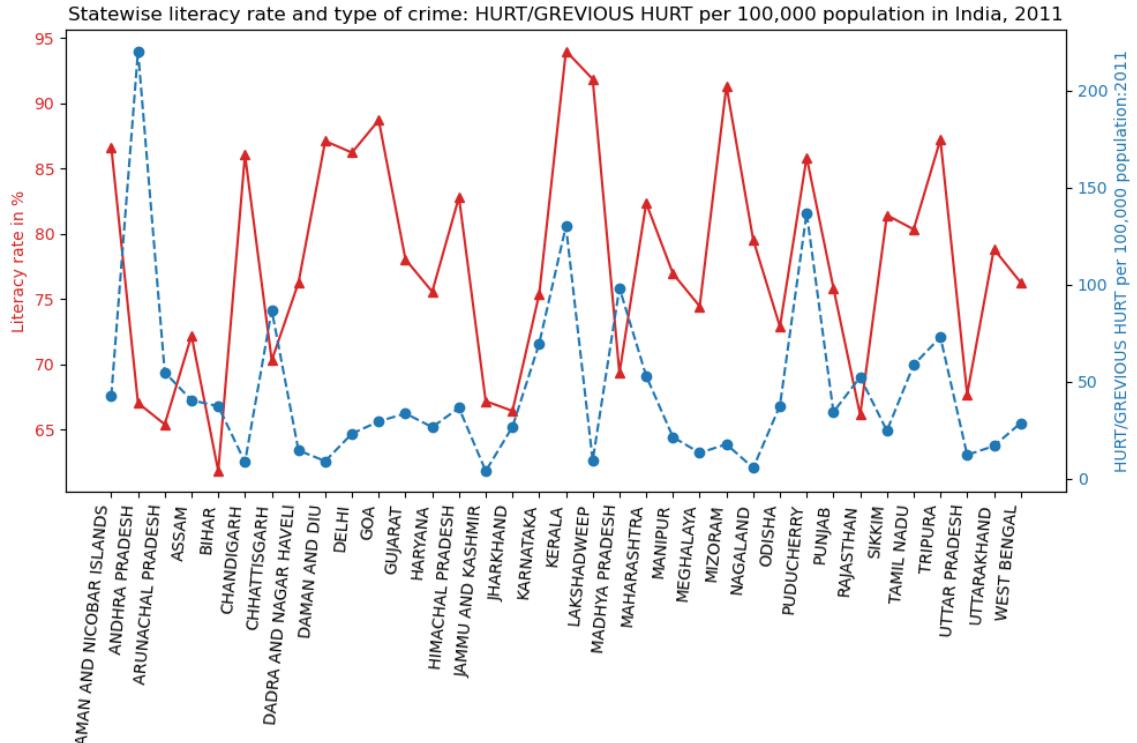


(b)

Fig. 23: State wise literacy and arson (a) 2001 (b) 2011

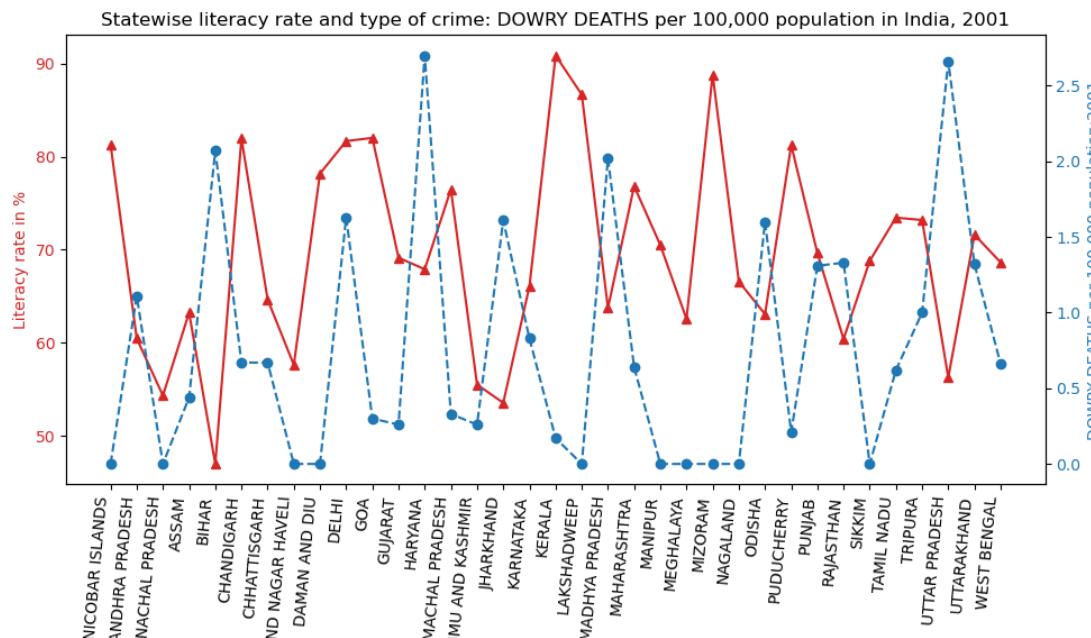


(a)

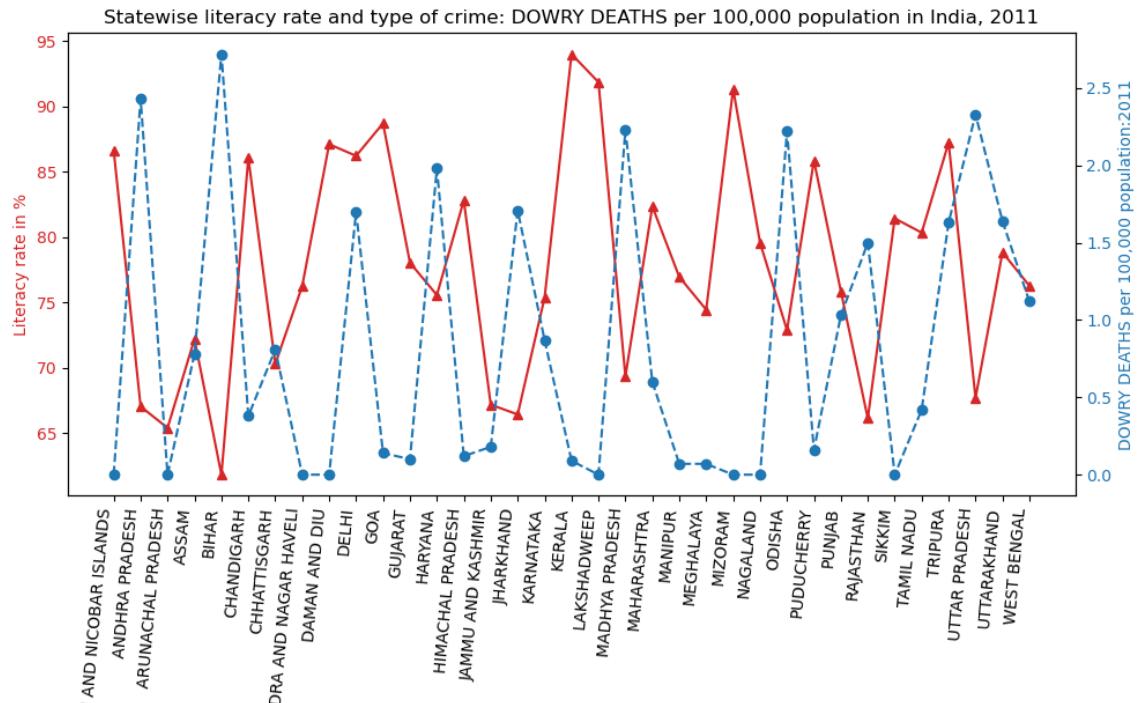


(b)

Fig. 24: State wise literacy and hurt/grievous (a) 2001 (b) 2011



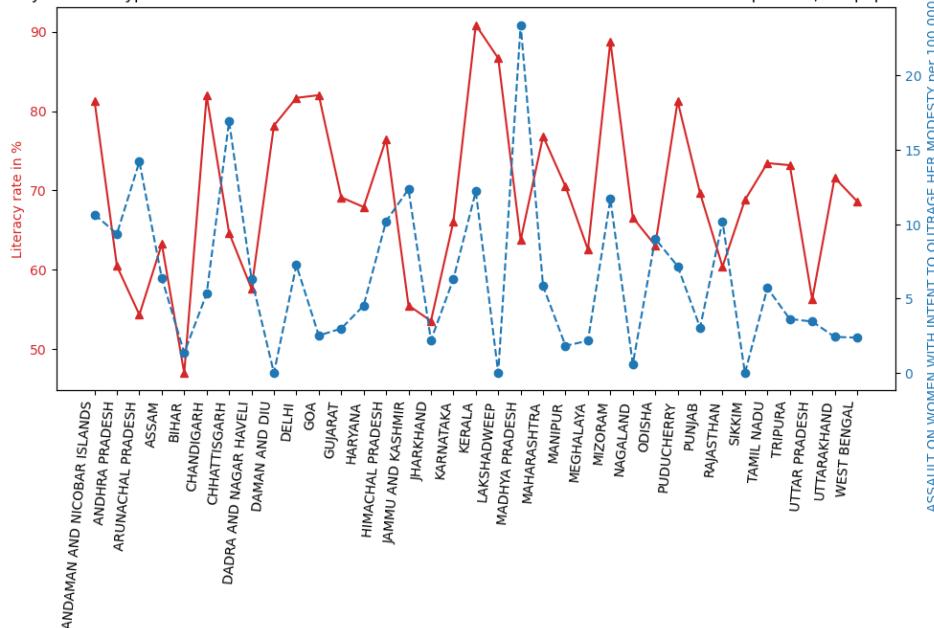
(a)



(b)

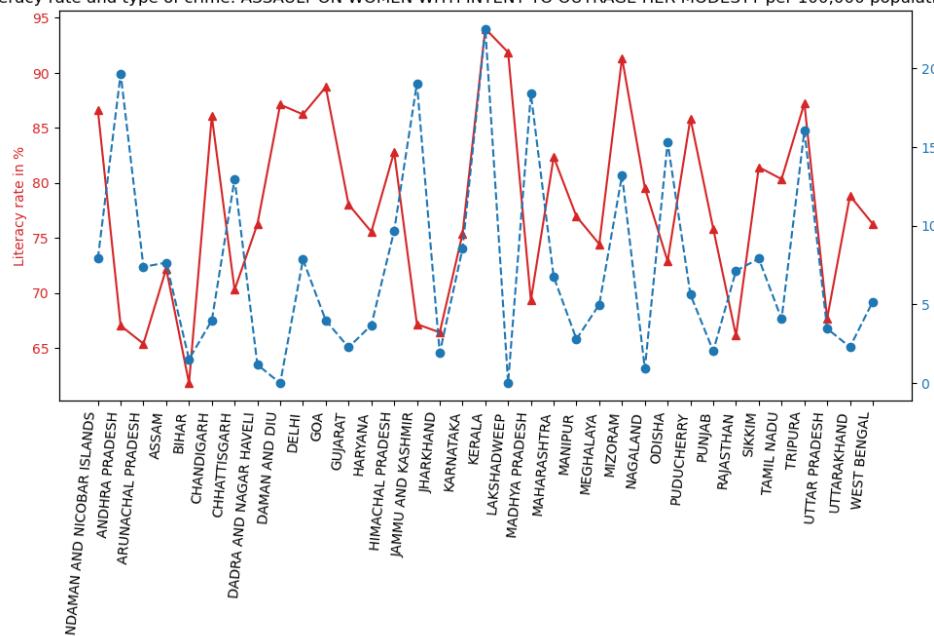
Fig. 25: State wise literacy and dowry deaths (a) 2001 (b) 2011

Statewise literacy rate and type of crime: ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY per 100,000 population in India, 2001



(a)

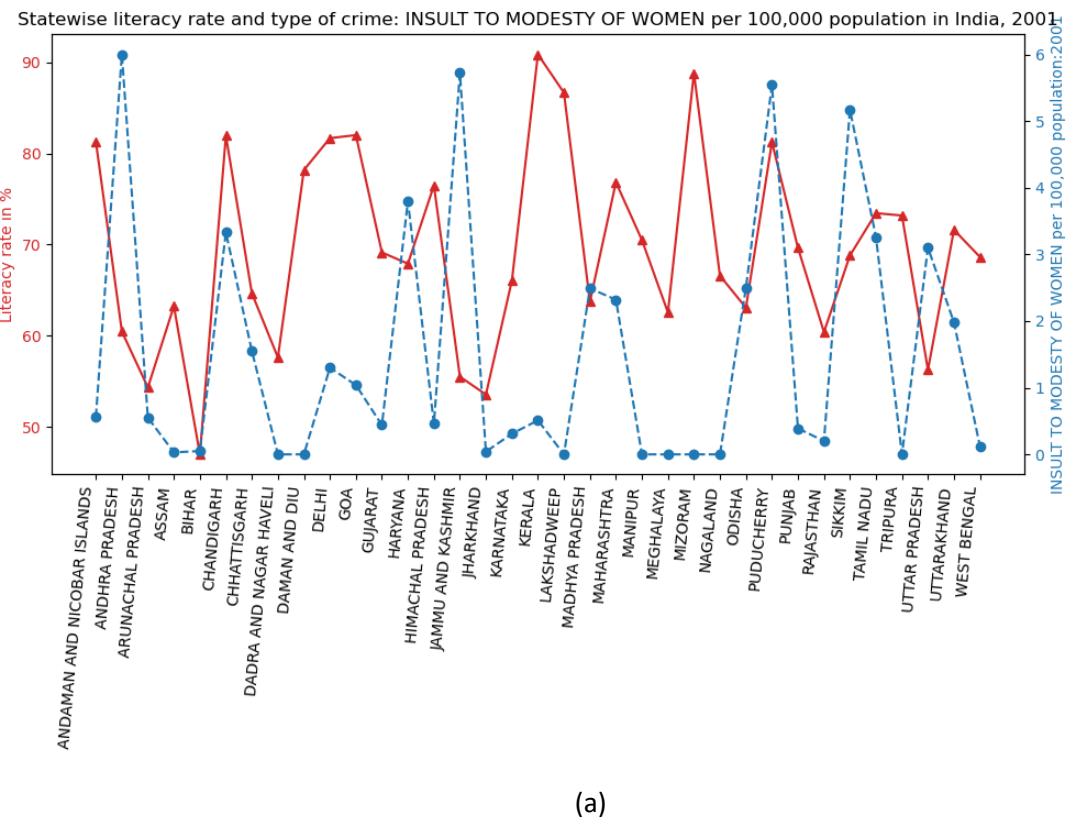
Statewise literacy rate and type of crime: ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY per 100,000 population in India, 2011



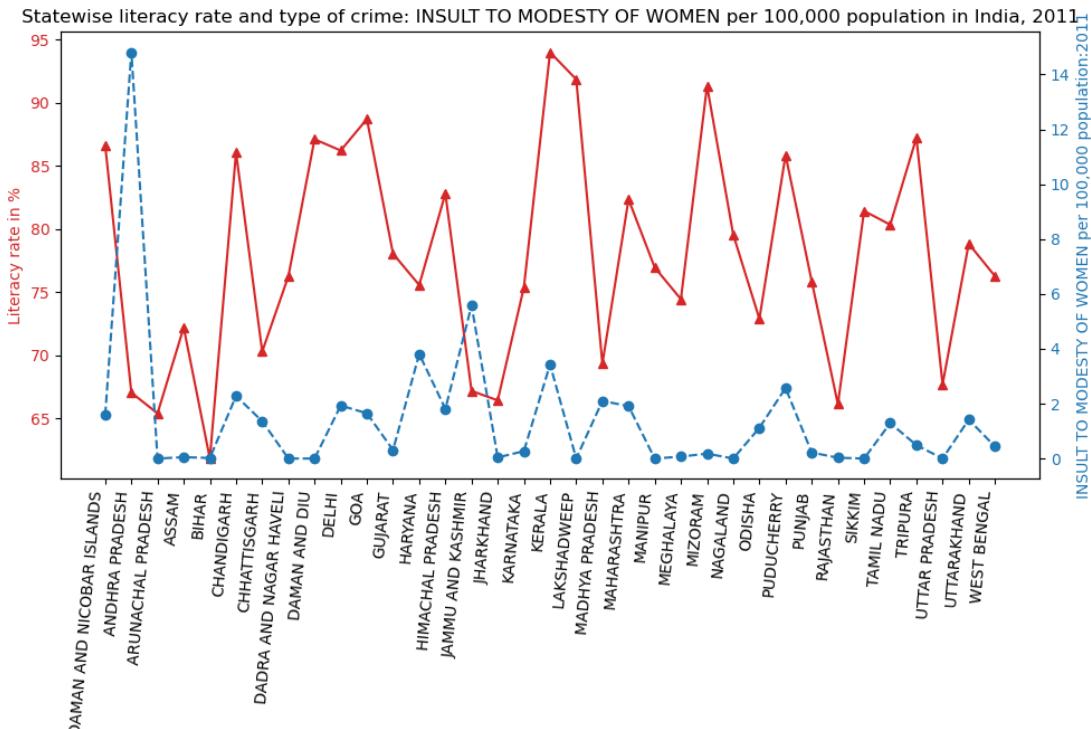
(b)

Fig. 26: State wise literacy and assault on women with intent to outrage her modesty

(a) 2001 (b) 2011



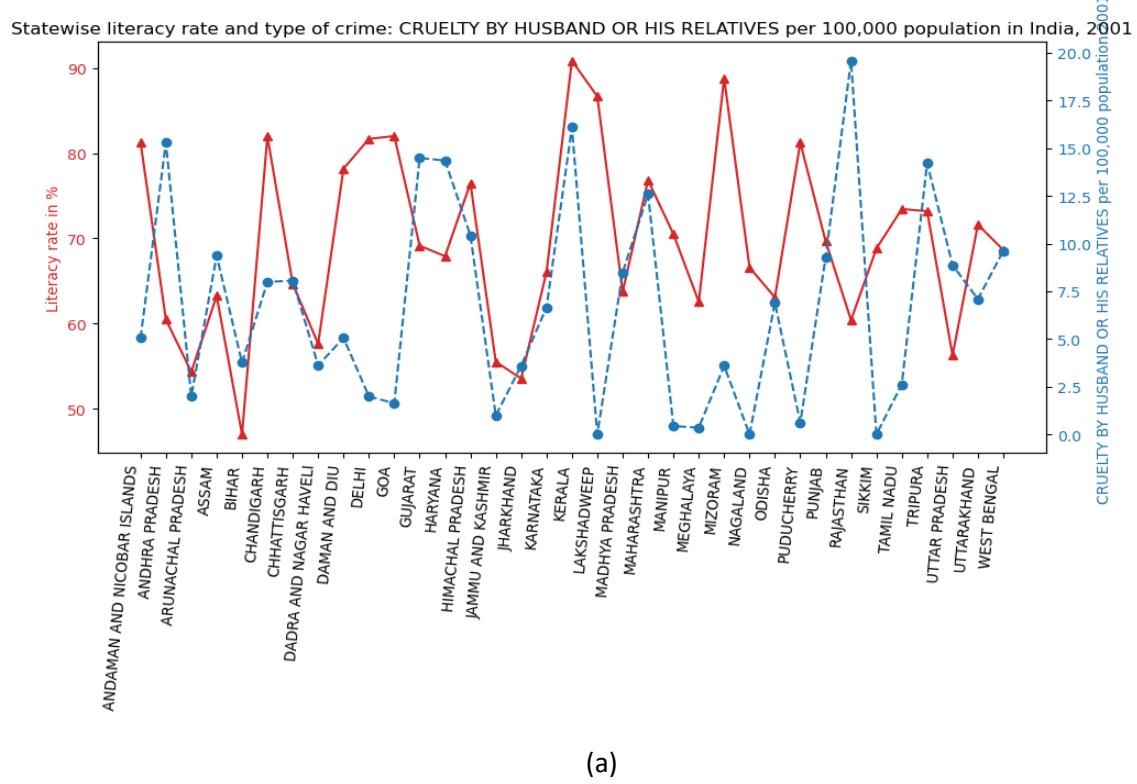
(a)



(b)

Fig. 27: State wise literacy and insult to modesty of women

(a) 2001 (b) 2011



(a)

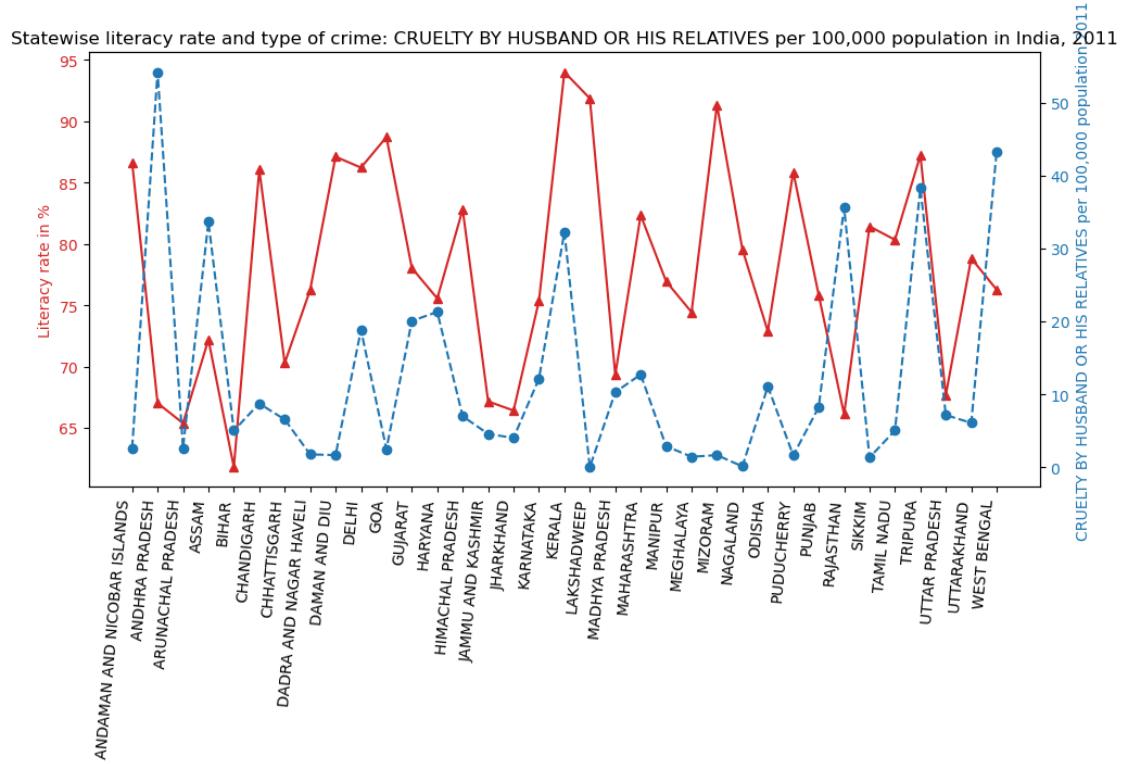
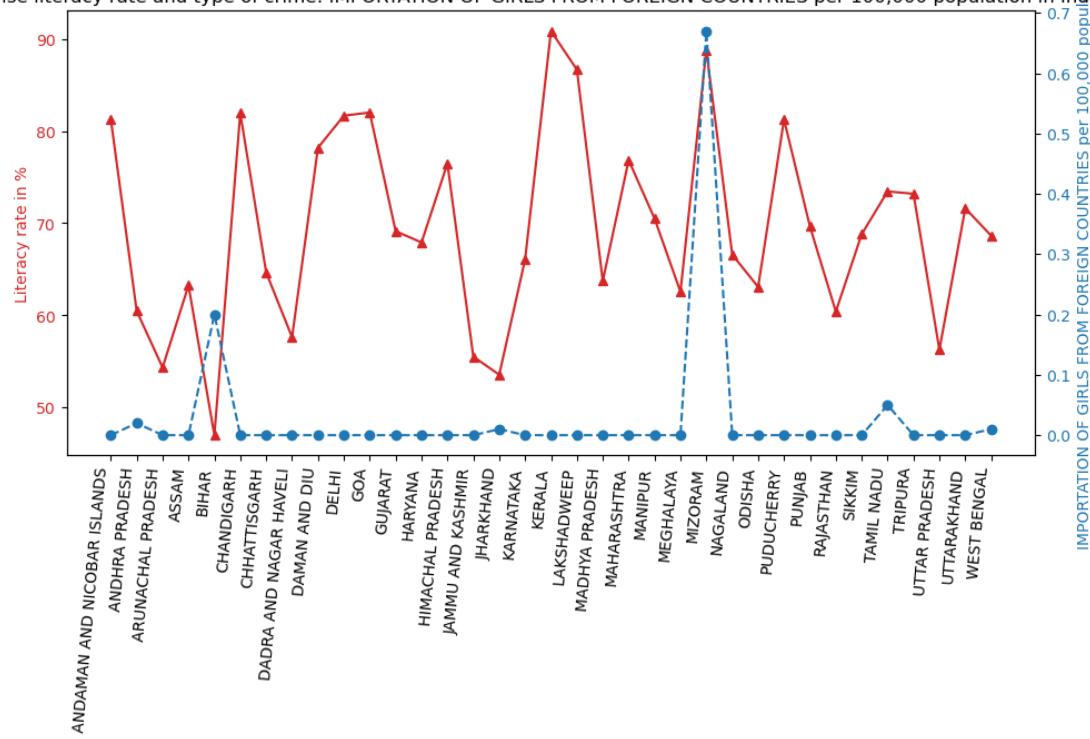


Fig. 28: State wise literacy and cruelty by husband or his relatives

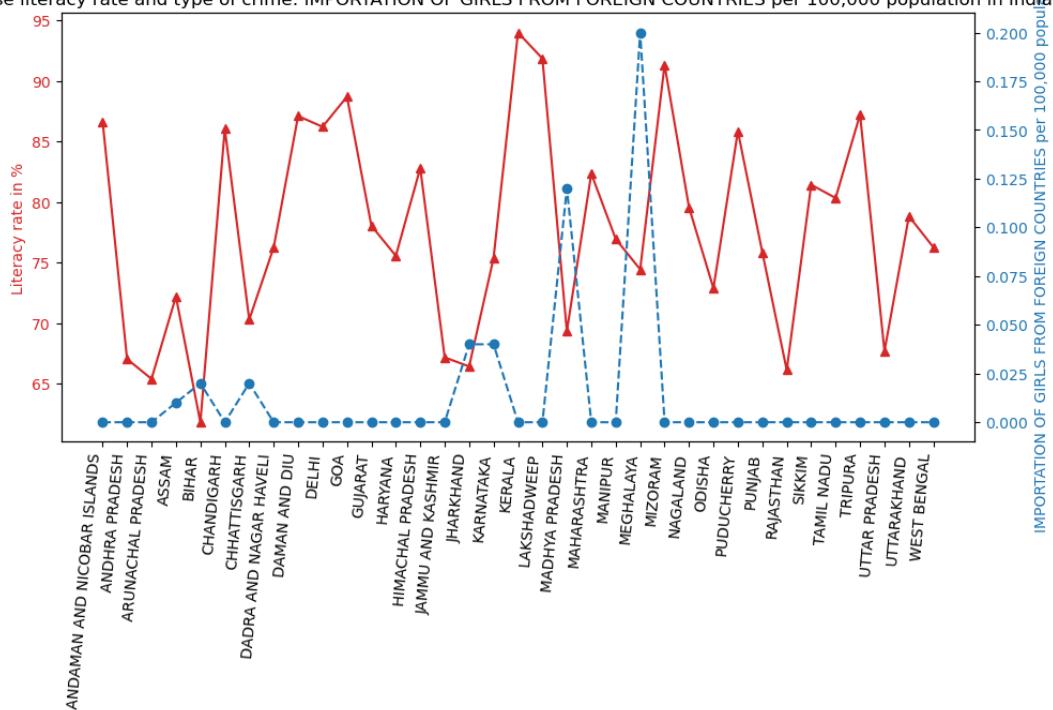
(a) 2001 (b) 2011

Statewise literacy rate and type of crime: IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES per 100,000 population in India, 2001



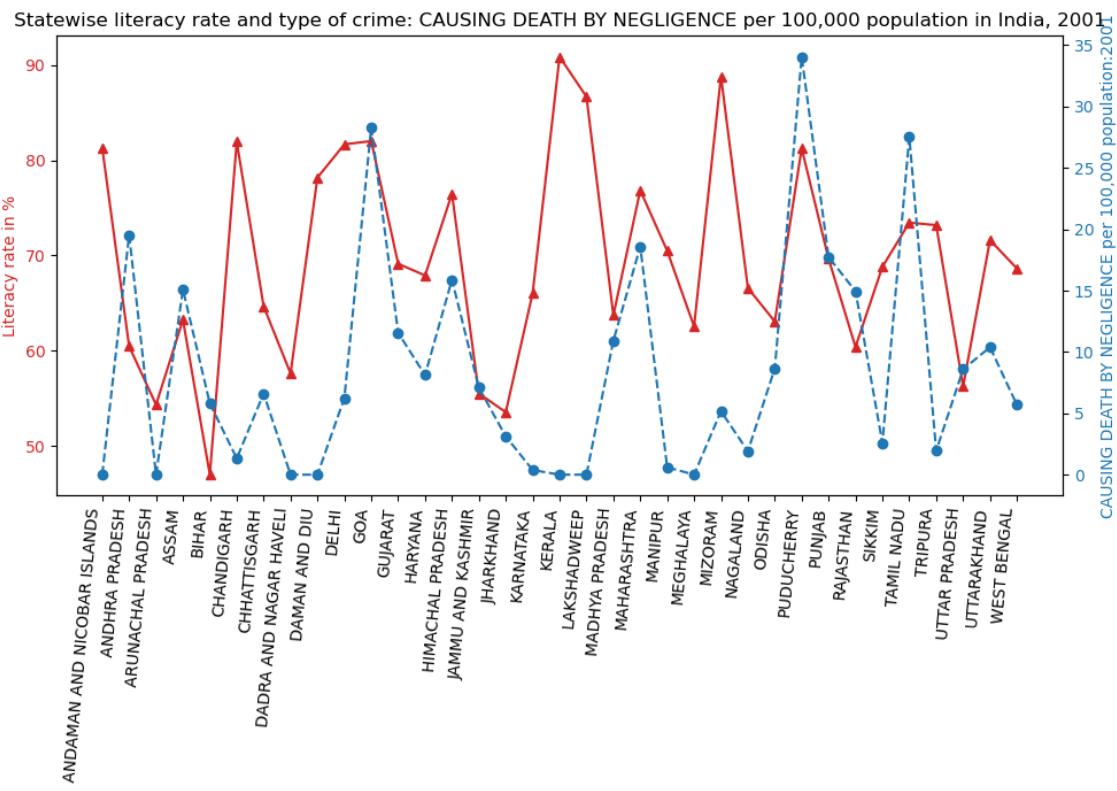
(a)

Statewise literacy rate and type of crime: IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES per 100,000 population in India, 2011

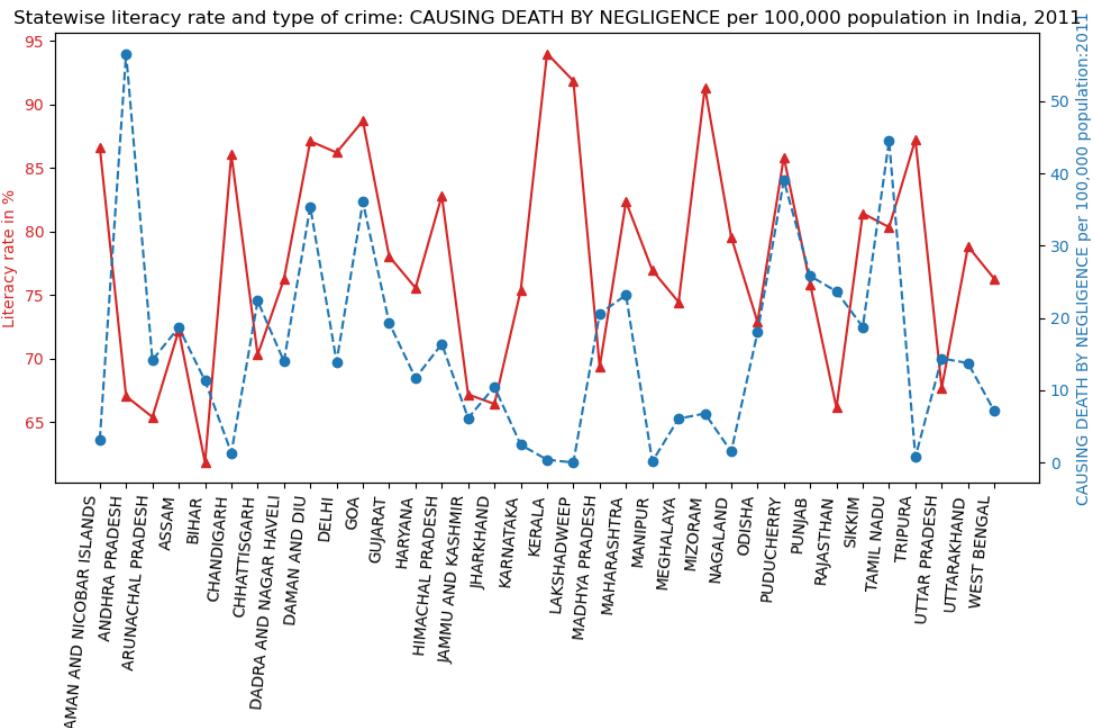


(b)

Fig. 29: State wise literacy and importation of girls from foreign countries (a) 2001 (b) 2011

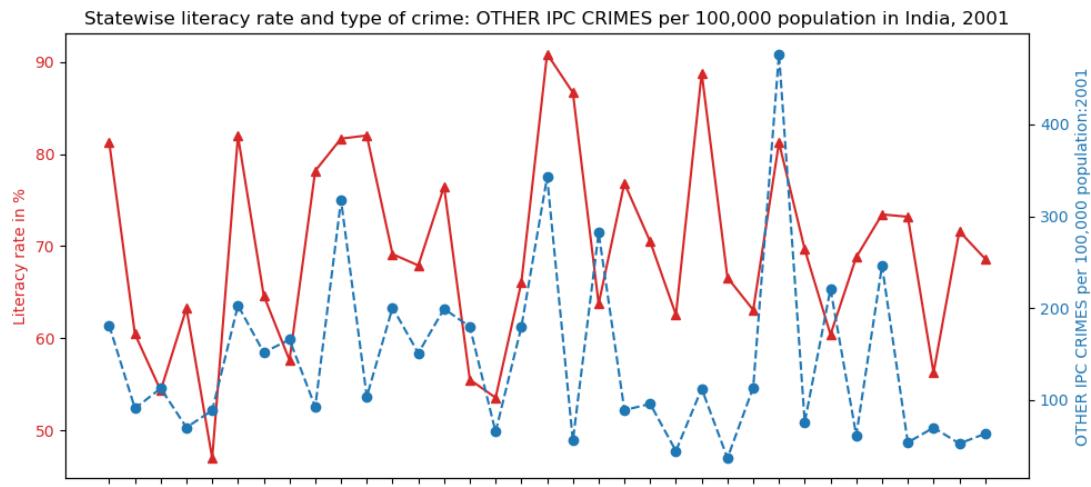


(a)

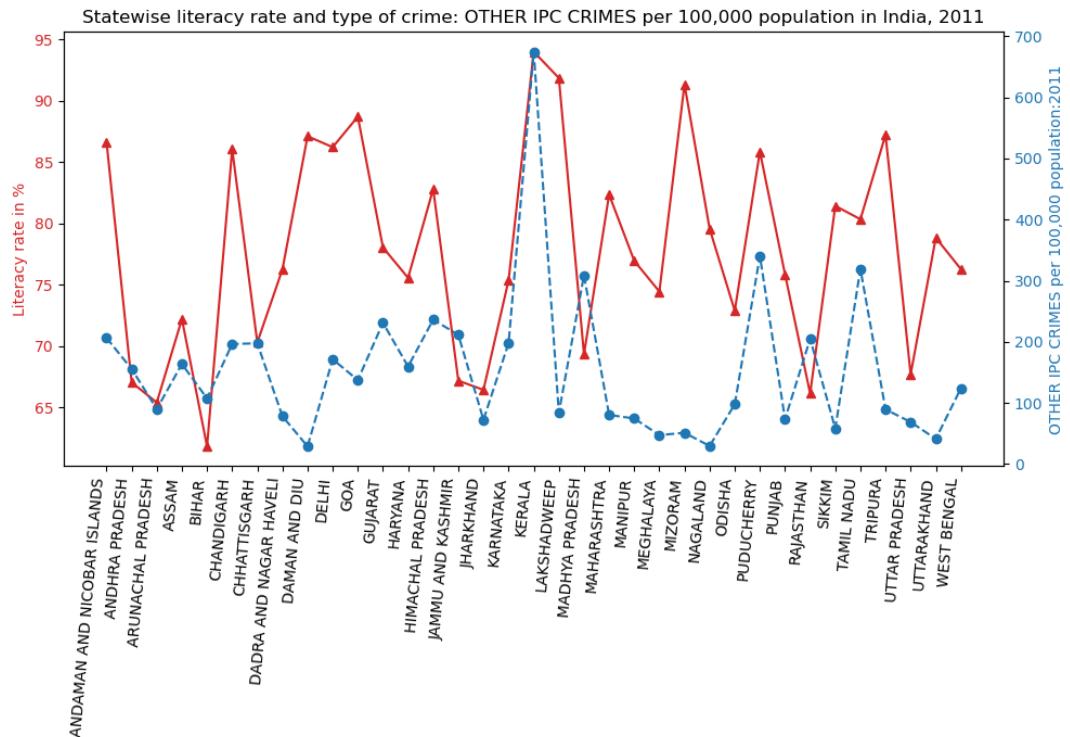


(b)

Fig. 30: State wise literacy and causing death by negligence (a) 2001 (b) 2011



(a)



(b)

Fig. 31: State wise literacy and other IPC crimes (a) 2001 (b) 2011

Table 2: Observations made in Fig.3 to Fig. 31

	2001		2011	
	Lowest	Highest	Lowest	Highest
Murder	DADRA AND NAGAR HAVELI (2.72), KERALA (2.96), LAKSHADWEEP (3.26), CHANDIGARH (3.33), HIMACHAL PRADESH (3.92), and WEST BENGAL(3.97).	TRIPURA (15.04), ARUNACHAL PRADESH (15.11), MANIPUR (18.22) and JAMMU AND KASHMIR (21.35).	LAKSHADWEEP (0.00), KERALA (2.19), JAMMU AND KASHMIR (2.69), UTTARAKHAND (3.52), GUJARAT (3.73) and HIMACHAL PRADESH (3.79).	JHARKHAND (10.59), ANDHRA PRADESH (11.37) and MEGHALAYA (11.47).
Attempt to Murder	ANDAMAN AND NICOBAR (0.00) ISLANDS and LAKSHADWEEP (0.00).	MANIPUR (14.64) and JAMMU AND KASHMIR (29.27).	DAMAN AND DIU (0.82)	MANIPUR (18.00)
Rape	LAKSHADWEEP (0.00) and DAMAN AND DIU (0.00)	CHHATTISGARH (9.21), MADHYA PRADESH (9.44) and MIZORAM (11.69).	LAKSHADWEEP (0.00) and DAMAN AND DIU (0.82)	ASSAM (10.91), TRIPURA (11.17) and MIZORAM (14.12).
Kidnapping & Abduction	LAKSHADWEEP (0.00), MIZORAM (0.45) and PUDUCHERRY (0.82).	DELHI (23.49) and ARUNACHAL PRADESH (15.11).	LAKSHADWEEP (0.00)	DELHI (44.97), ASSAM (24.15) and JAMMU AND KASHMIR (17.16).
Dacoity	ANDAMAN AND NICOBAR ISLANDS, SIKKIM, DADRA AND NAGAR HAVELI, DAMAN AND DIU, LAKSHADWEEP	MEGHALAYA (8.41), JHARKHAND (4.72) and ARUNACHAL PRADESH (4.01)	SIKKIM (0.00) and LAKSHADWEEP (0.00)	DADRA AND NAGAR HAVELI (4.08), DAMAN AND DIU (3.29) and MEGHALAYA (3.31)
Dowry Deaths	ANDAMAN AND NICOBAR ISLANDS, SIKKIM, ARUNACHAL PRADESH, NAGALAND, MIZORAM, MEGHALAYA, MANIPUR, DADRA AND NAGAR HAVELI, DAMAN AND DIU and LAKSHADWEEP	MADHYA PRADESH (2.02), BIHAR (2.07), UTTAR PRADESH (2.66) and HARYANA (2.70)	ANDAMAN AND NICOBAR ISLANDS, SIKKIM, ARUNACHAL PRADESH, NAGALAND, MIZORAM, DADRA AND NAGAR HAVELI, DAMAN AND DIU and LAKSHADWEEP (no cases)	ODISHA (2.22), MADHYA PRADESH (2.23), UTTAR PRADESH (2.33), ANDHRA PRADESH (2.43) and BIHAR (2.72)

Assault on Women with Intent to Outrage Her Modesty	SIKKIM, LAKSHADWEEP and DAMAN AND DIU (no cases)	CHHATTISGARH (16.92) and MADHYA PRADESH (23.39)	LAKSHADWEEP and DAMAN AND DIU (NIL cases)	ODISHA (15.29), TRIPURA (16.02), MADHYA PRADESH (18.36), JAMMU AND KASHMIR (19.03), ANDHRA PRADESH (19.64) and KERALA (22.50)
Cruelty by Husband or His Relatives	SIKKIM, LAKSHADWEEP and NAGALAND (no cases)	ANDHRA PRADESH (15.29), KERALA (16.09) and RAJASTHAN (19.59)	LAKSHADWEEP (Nil cases)	DELHI (18.80), GUJARAT (20.05), HARYANA (21.28), KERALA (32.21), ASSAM (33.66), RAJASTHAN (35.61), TRIPURA (38.25), WEST BENGAL (43.29) and ANDHRA PRADESH (54.17).

### 2.3. To analyse year-on-year total crime rate.

The year on year total crime rate for all the states and UT together has been shown in Fig. 32. A steady increase in crime rate can be observed from Fig. 32, except for the year 2003.

In the year 2003, there was some decrease in crime rate observed in the country compared to 2002. Also, a very little decrease in crime rate was observed between 2004 and 2005. The exact reason for the same needs to be explored.

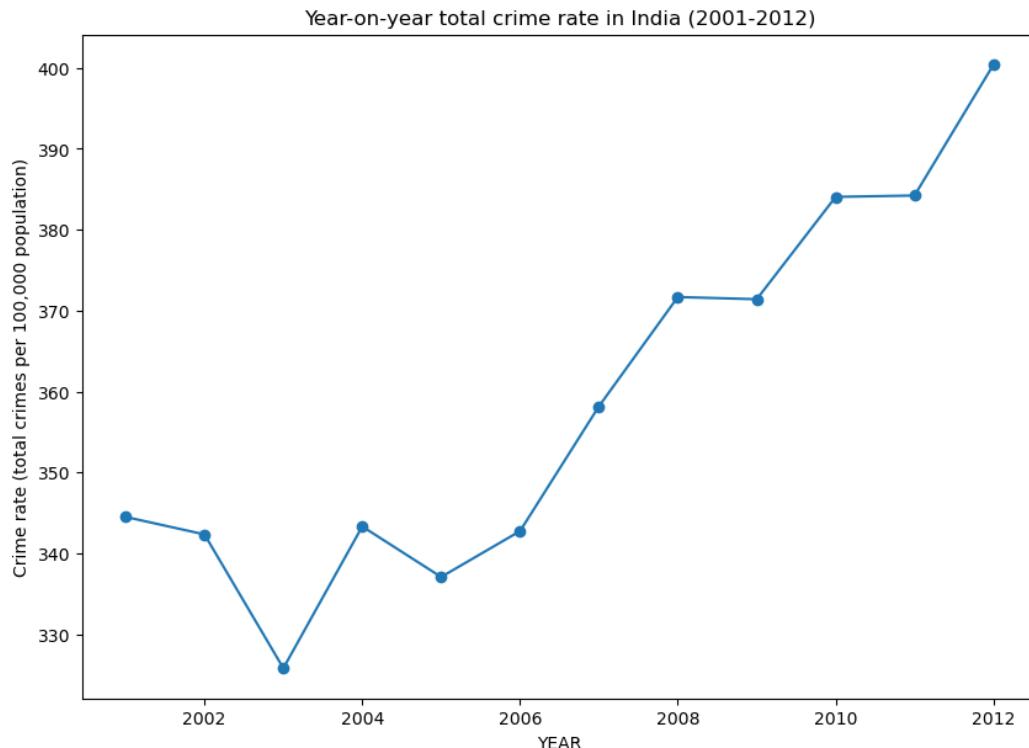


Fig. 32.: Year on year crime rate (2001- 2012)

## 2.4. To analyse area vs overall crime.

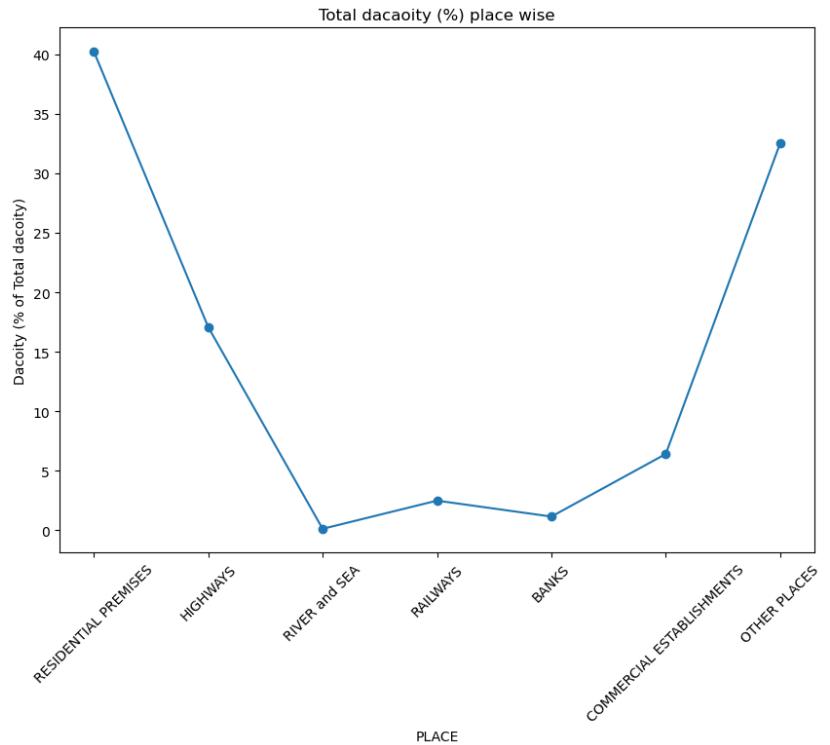


Fig. 33: Dacoity (%) in various areas or localities

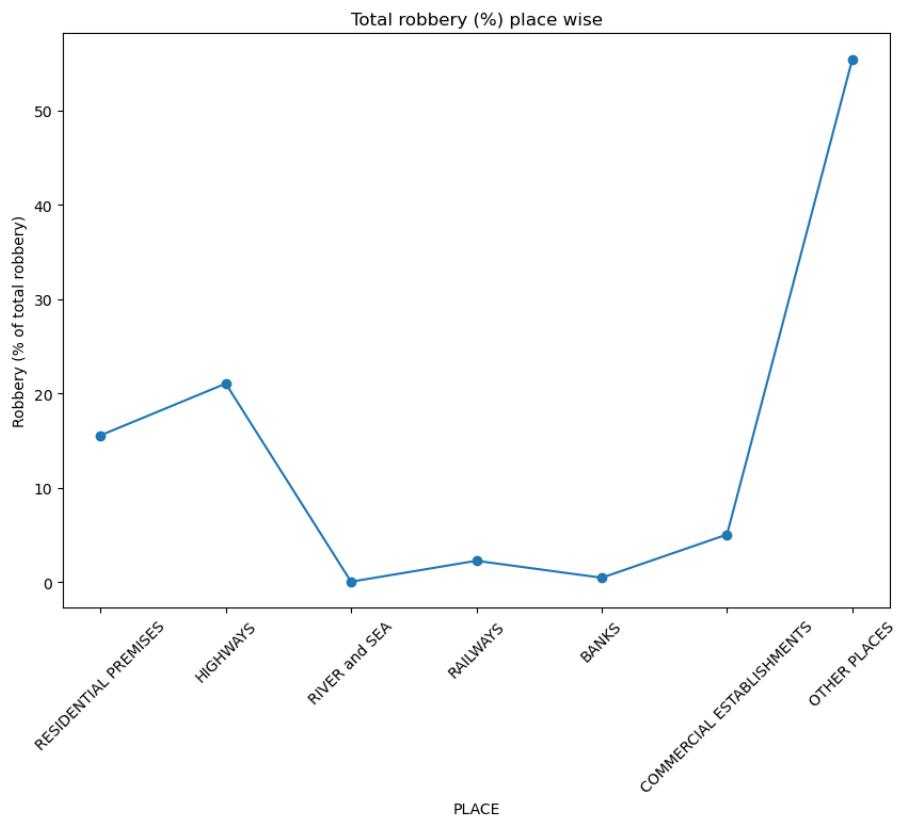


Fig. 34: Robbery (%) in various areas or localities

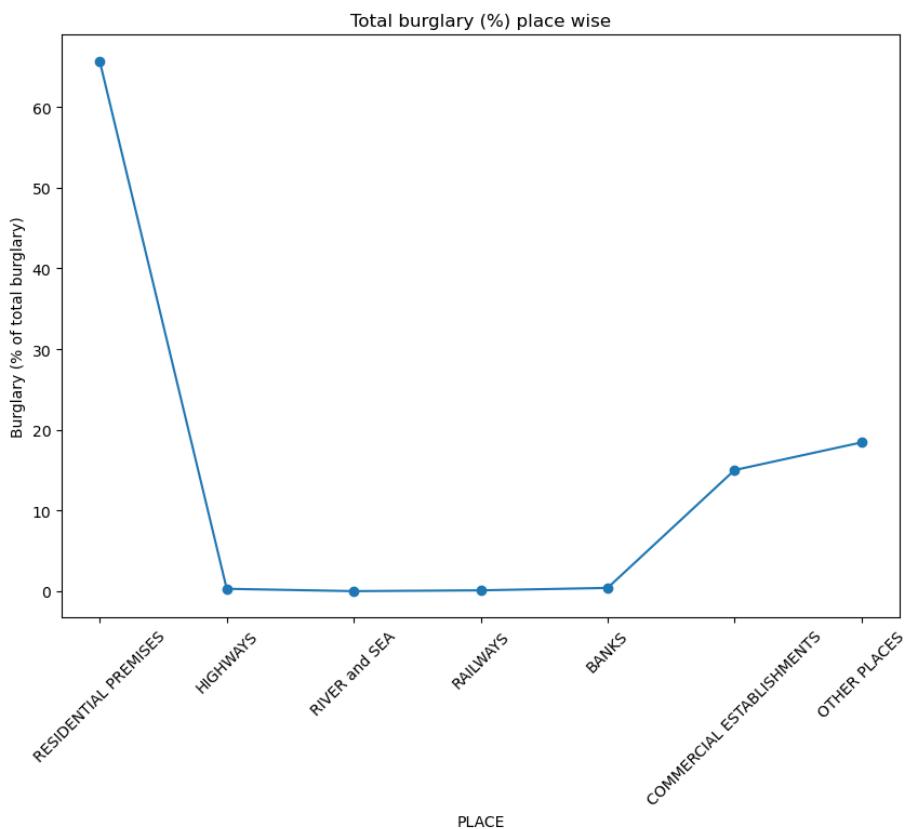


Fig. 35: Burglary (%) in various areas or localities

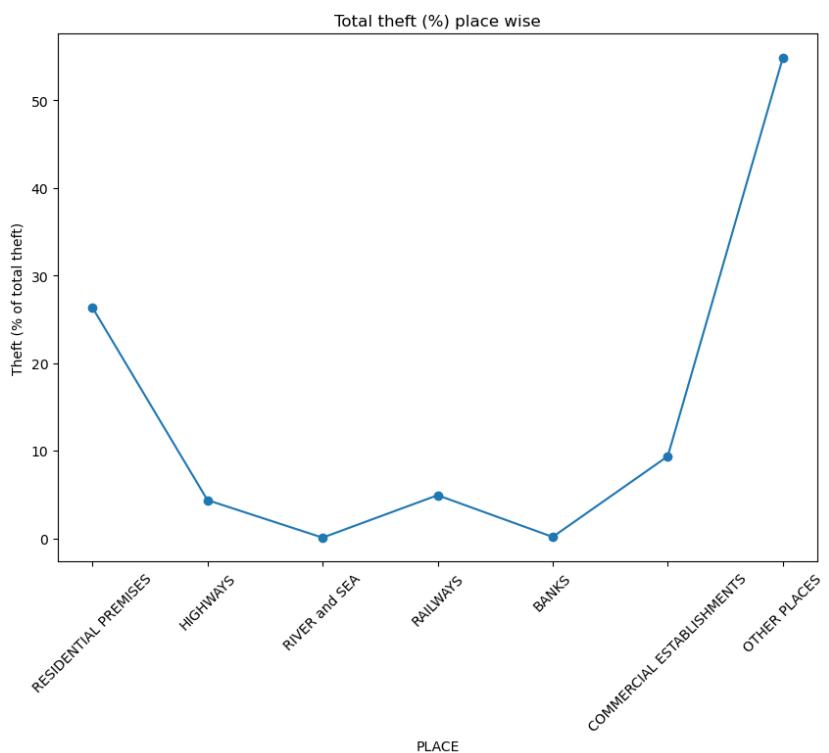


Fig. 36: Theft (%) in various areas or localities

Various areas and locations were analysed for Dacoity, Robbery, Burglary and Theft. The areas included in the analysis have been categorised as:

RESIDENTIAL PREMISES, HIGHWAY', RIVER and SEA, RAILWAYS, BANKS, COMMERCIAL ESTABLISHMENTS, and OTHER PLACES

These have been shown by various plots in Fig 33 to Fig.36. It can be observed from these plots that the highest percentage of dacoity and burglary has been registered in residential premises. However, robbery and theft have been registered in area labelled as "other places".

## 2.5. To analyse Population vs overall Crime.

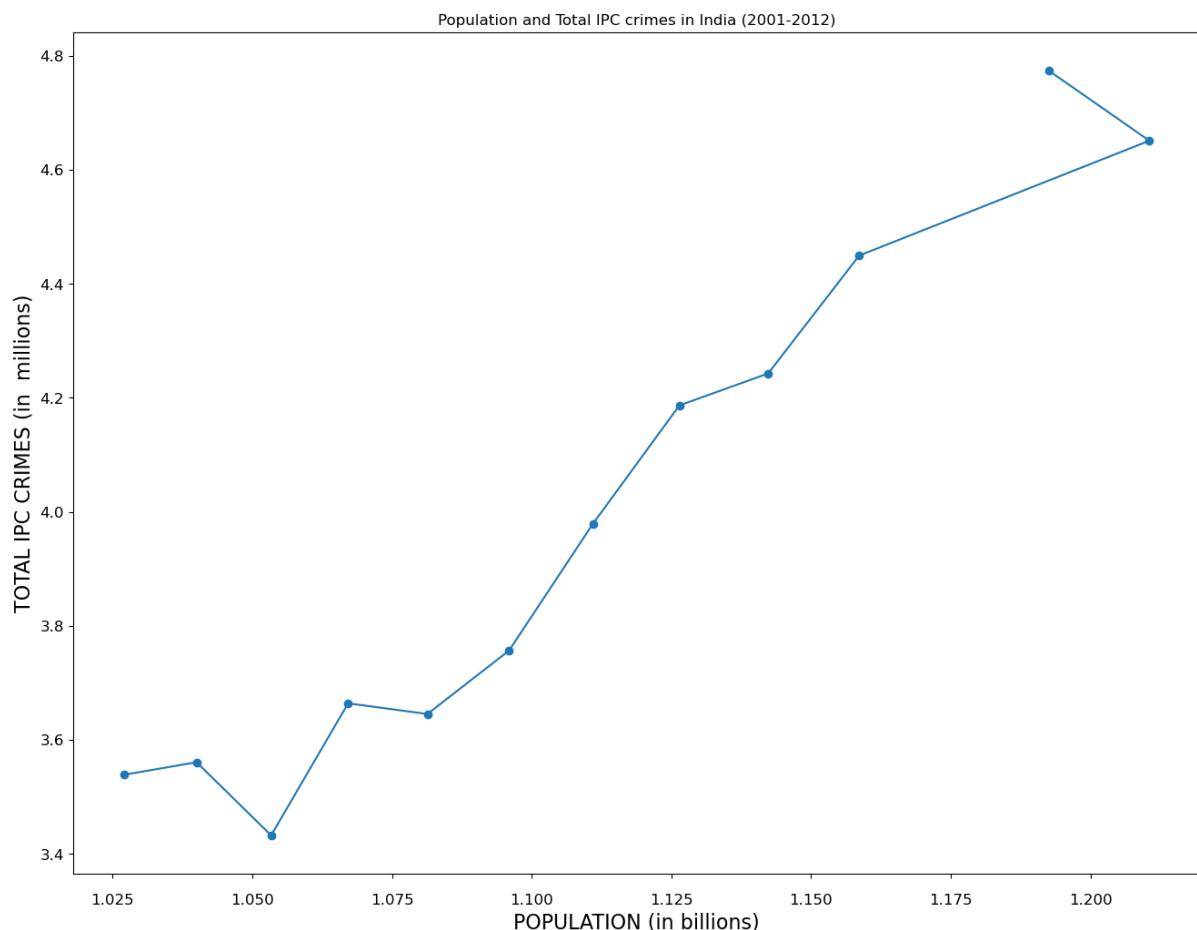


Fig. 37: Population (2001-2012) and total IPC crimes

It can be observed from Fig. 37, that there has been almost a steady increase in crime as the population increased between 2001 and 2012.

### Sensitive, moderate and peaceful districts:

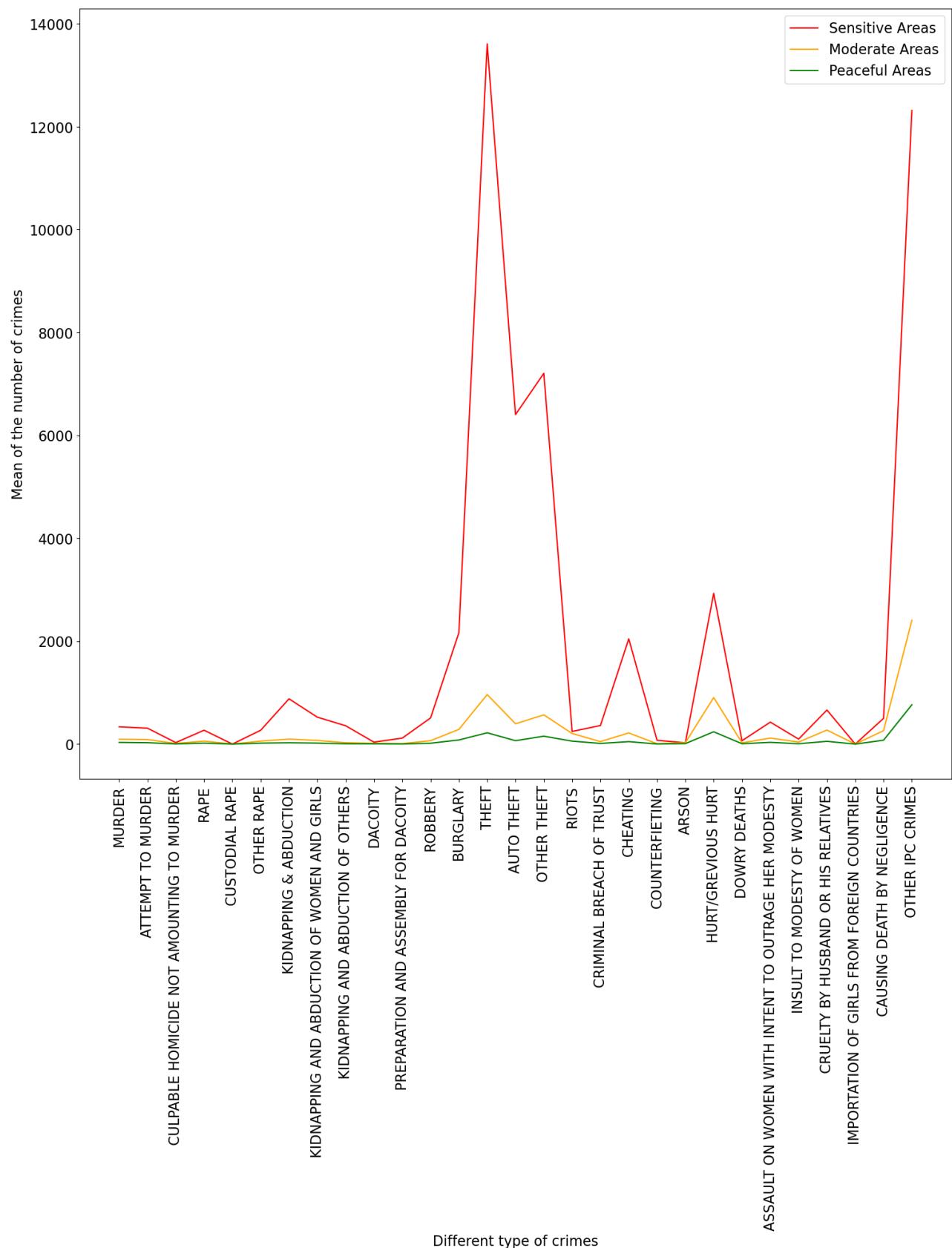


Fig. 38: Sensitive, moderate and peaceful districts clustered based upon the different crimes

Based upon there are 4 districts in the country which fall under sensitive areas. These are: 'MUMBAI', 'DELHI UT TOTAL', 'BANGALORE COMM.R.', 'MUMBAI COMM.R.'

It can be observed from Fig. 38, that:

- (i) The mean value of the crime "THEFT" is observed to be maximum in sensitive areas compared to peaceful and moderate areas.
- (ii) Crimes like "CHEATING" and "HURT/GREVIOUS HURT" are also relatively higher in sensitive areas.

Conclusion:

- (i) It is expected that crime rate should be low in states with high literacy. This is observed in case of LAKSHADWEEP in both the years (2001 and 2011). It is a UT with one of the highest literacies and lowest crime rate.  
However, in 2011, KERALA which has the highest literacy has also the highest crime rate in the country. ANDHRA PRADESH, which is one of the lowest literacy states in the country is among one of the highest crime rate states.
- (ii) Despite increase in literacy, there has been a drastic increase in the crime rate in ANDHRA PRADESH (more than doubled) and KERALA between 2001 and 2011.
- (iii) MIZORAM and DELHI have seen a significant decline in crime rates between 2001-2011.
- (iv) Even though there has been a considerable decline in crime rate in PUDUCHERRY between 2001-2011, still it is one of the highest crime rate states in the country as on 2011.
- (v) DADRA AND NAGAR HAVELI have reported an increase of 18.61 % points in literacy, which is the highest increase in the decade 2001-11. This is followed by BIHAR, which has reported an increase in 14.8 % point. TRIPURA has also reported an increase in 14.03 % points in literacy in this duration.
- (vi) JAMMU AND KASHMIR reported a drastic reduction in murder as a crime between 2001 and 2011. It became a state with the highest murder to one of the lowest murder states.
- (vii) In general, this decade (2001-2011) observed a drastic reduction in murder as the crime. Where there were 4 states in the year 2001 with more than 15 murders per 100,000 population, there were no states with more than 15 murder per 100,000 population and there were only 3 states with more than 10 murder per 100,000 population.
- (viii) In 2001, JAMMU AND KASHMIR had the highest attempt to murder cases with more than 29 cases per 100,000 population. This number drastically reduced to 7.87 by 2011.

- (ix) One of the highest literate north-eastern state of MIZORAM reported the highest rape cases in the country in both 2001 and 2011.
- (x) DELHI reported the highest cases of kidnapping and abduction in the country in both 2001 and 2011.
- (xi) DADRA AND NAGAR HAVELI reported the highest increase in dacoity per 100,000 population from 2001 to 2011. It became a UT with zero cases of dacoity to the highest cases by 2011
- (xii) SIKKIM and LAKSHDWEEP were the states with NO cases of dacoity consistently in 2001 and 2011.
- (xiii) HARYANA and BIHAR were the states with the highest dowry deaths in the country in 2001 and 2011 respectively, whereas ANDHRA PRADESH reported the highest increase in dowry deaths per 100,00 population between 2001 and 2011.
- (xiv) MADHYA PRADESH and KERALA were the states with highest cases of "ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY per 100,000 population" in 2001 and 2011 respectively. In fact, KERALA reported the highest increase in cases of assault on women with intent to outrage her modesty per 100,000 population between 2001 and 2011.
- (xv) RAJASTHAN and ANDHRA PRADESH topped the list of states in "CRUELTY BY HUSBAND OR HIS RELATIVES per 100,000 population" in 2001 and 2011 respectively. In fact, ANDHRA PRADESH reported a very drastic fourfold increase in this crime from 2001 to 2011. On the flip side LAKSHDWEEP was consistently a UT with no cases of "CRUELTY BY HUSBAND OR HIS RELATIVES per 100,000 population" in 2001 and 2011.
- (xvi) 'MUMBAI', 'DELHI UT', 'BANGALORE COMM.R.', 'MUMBAI COMM.R. are the districts in sensitive areas.
- (xvii) Among these DELHI UT (total IPC crimes = 633174) has the highest crimes. This data represents the sum total of all IPC crimes from 2001 to 2012. so, it is the most unsafe district.
- (xviii) Districts LONGLENG(Nagaland) and ANJAW (Arunachal Pradesh) are among the least crime districts. In the same 12 years duration there have been total IPC crimes 84 and 85 respectively. Therefore, these districts can be referred to as the safest districts in the country.

#### **Limitations:**

The districts which fall under sensitive areas are also among the most populated districts of the country. So, clustering these districts considering total crime will not give a fair idea of the crime in these districts. A better metric to cluster the districts based on crimes would be to consider crime rate, which basically normalizes the crimes based on the population. Crime rate considers the number of crimes per 100,000 of the population.