GROUP 13
CRIME PATTERN ANALYSIS OF INDIA (2001-2010)

GROUP MEMBERS

Kusum Bunkar
kusum@iitk.ac.in | 19111047

Nitesh Trivedi
nitesht@iitk.ac.in | 19211266

Renuka Ahirwar
renuka@iitk.ac.in | 19111071

Virendra Nishad
viren@iitk.ac.in | 19111099

Yash Kumar
yashk@iitk.ac.in | 19211274

December 6, 2020

1 Abstract

With the increased availability of digital data on crimes committed across India crime pattern analysis has evolved as finding crime hot-spots and cold-spots. This project is centered with the district wise analysis of crimes committed across India. Focus has been narrowed down to four major types of crimes namely crimes on SC, crimes on ST, crimes on Women, and crimes on Children. Analysis reports top five hot-spots and cold-spots districts/states for each type of crime. Our analysis is backed by the inferences drawn out of census data.

2 Problem Statement

We aim to analyse hot-spot and cold-spot districts/states in crimes across India for years 2001 to 2010. We have considered four major types of crimes(against SC,ST,Women and Children). We also aim to back our analysis by finding relation with the census data.

3 Introduction and Motivation

Crime pattern analysis plays an important role in construction of policies to overcome crimes committed in a region. The crime pattern analysis focuses on detection of hot-spot areas; the areas with high density of crime. Based on certain requirements policies are framed for hot-spot regions. The policies help to deploy necessary resources to the hot-spot areas. Studies state that 10% people commit about 50% crimes, the statement indeed reflects the importance of crime pattern analysis. Crimes get diversified based on geographical location, in specifically urban crime pattern has been analyzed. Some studies have also considered the spatial element of crime and have linked geographical information with crime data records enabling GIS(Geographic information system) to become an additional tool in the war against crime .

India ranks 65th in crimes with crime index of 44.2 https://www.numbeo.com/crime/rankings_by_country.jsp. Crime pattern analysis can benefit India in deciding policies for districts/states, which in-turn leads to deployment of resources in efficient manner. The efficient utilization of resources can overcome crimes in India in a cost effective manner. Motivated with the effectiveness and importance of crime pattern analysis we present our work on district/state wise crime analysis in India.

4 Dataset

- District wise crime data record of India for the year 2001-2014.

- Census Data for the period 2001 to 2010.
- Data has been shared on Open Govt Data Platform India portal under Govt. Open Data License - India
- Census Data released (and owned by) the Registrar General and Census Commissioner of India under the Ministry of Home Affairs, Government of India.
- Unemployment Rate data by Ministry of Statistics and Programme Implementation.
- Poverty Rate data by Reserve Bank of India.

5 Data pre-processing

Following steps were followed to pre-process the data:

- The missing state/districts crime count were made zero by adding a new entry for that particular year.
- Several district's names were changed with common name if they had any other alias.
- Districts of union territories were combined together as one single district named by the corresponding union territory's name.
- Spelling mistakes in district's name were handled with correct name.
- The number of districts in the crime files were made equal in number for every year(2001-2010) using above mentioned steps.
- Census Data state/district names were made similar to the pre-processed crime files.

Note: All the pre-processing data is commented in the attached pre-processing code.

6 Methodology

We first find the extremum to extract hot-spot/cold-spot districts and states for each type of file. Then we find growth rate of extracted extremum districts/states for years 2001 to 2010. Going ahead we use co-relation statistics to find insights.

6.1 Finding extremum

- We will use mean as reference for average crime count.
- Then we will calculate deviation of crime count from mean for each district.
- For zscore method we will normalise b the standard deviation of the Total crimes.
- For Population case we will normalize the deviation by population size.
- We will pick district/state showing maximal positive deviation as leading spot for the respective crime.
  By using above method we will find the following results -
  - Top 5 states and districts (in India) per crime year wise.
  - Top 5 districts (in every state) per crime year wise.
  - Top 5 crimes in every state per year.
  - Top 5 crimes in India per year.

6.2 Change in crime per year to identify prevalent crimes

- We will observe increase/decrease in total crime and overall crime across categories per year in that particular district.

- This change will help us to determine the effective growth rate of crime and also in which category. The formula for growth rate of crime used is mentioned below :

$$\text{Growth rate of crime for two consecutive years} = \frac{\text{Total crimes in current year - Total crimes in previous year}}{\text{Total crimes in previous year}} \times 100$$

- We would consider crime data for each subcategory separately. This will give us a clear picture about which subcategory of crime has contributed to overall crimes in that particular category. For example, let's say crime against women is 1000 for a year with dowry, rape, kidnapping cases as 500, 300, 200 respectively. So we can conclude that dowry has a major contribution in overall women crime.

$$\text{Contribution of a subcategory} = \frac{\text{Total cases in the subcategory of crime}}{\text{Total cases of the crime}} \times 100$$

6.3 Correlation

- Based on correlation of attributes like literacy rate, population density, strength of security personnel, other crimes etc with the respective crime, we will conclude possible causes for crime.

6.4 Inferential analysis

- To develop insights, it is crucial to study the relation of districts/states with high crime rates with the districts and states with lower crime rates and the factors associated(literacy rate, non workers, unemployment and population density).

- For each subcategory of crimes, we extracted particular number of columns(4-5) based on the correlation values.

- The union of census columns obtained for each subcategory above were used to form a set which were used to make heatmap with total crimes and subcategories of crimes.

- Finally, the selection of the census attributes was done using the value obtained from the correlation matrix obtained in the previous step. In order to fetch 5 census attributes, we used different values in the condition(example for SCs it was 0.45, etc.)

7 RESULTS AND ANALYSIS

Our analysis is centered towards four types of crimes:

1. Crime against Women
2. Crime against Children
3. Crime against SCs
4. Crime against STs

We apply our methodology on each of the crime files in order mentioned above.

7.1 Women

7.1.1 Extremum

We have extracted top five hot-spots/cold-spots districts and states under two dimensions: Density(districts/states total crimes are normalized by population) and zscore Normalization. The following tables show the top hot-spots/cold-spots for crimes against women.

Population			
Districts		**States**	
Top	Low	Top	Low
DELHI	DEHRADUN	DELHI	HARYANA
HYDERABAD CITY	MADHUBANI	ANDHRA PRADESH	CHHATTISGARH
CHITTORGARH	SIHORE	MADHYA PRADESH	JHARKHAND
KARIMNAGAR	ETAH	RAJASTHAN	PUNJAB
KOTA	HISSAR	WEST BENGAL	JAMMU & KASHMIR

Table 7.1: Density based top 5 hot-spots/cold-spots districts and states.

Zscore			
Districts		**States**	
Top	Low	Top	Low
DELHI	TAMENGLONG	UTTAR PRADESH	LAKSHADWEEP
24 PARGANAS SOUTH	LONGLENG	ANDHRA PRADESH	DAMAN & DIU
MUMBAI	ANJAW	MADHYA PRADESH	NAGALAND
24 PARGANAS NORTH	KIPHIRE	WEST BENGAL	D & N HAVELI
HYDERABAD CITY	CHANDEL	RAJASTHAN	PUDUCHERRY

Table 7.2: zscore based top 5 hot-spots/cold-spots districts and states.

By observing the tables 7.1 and 7.2 we can spot the difference in the extracted hot-spots/cold-spots districts and states. As key observations are:

- Only two hot-spot districts fall in the intersection of both the methods.
- Uttar Pradesh does not appear in top 5 hot-spots when calculated using density method while zscore method covers it as a hot-spot district.
- Haryana appears as a cold-spot district when density method is used.
- Most cold-spot districts belong to Union Territories calculated using zscore method.

7.1.2 Growth Rate

For the top hot-spots/cold-spots states and districts, the following plots show the growth rate in consecutive years for the period 2001-2010.

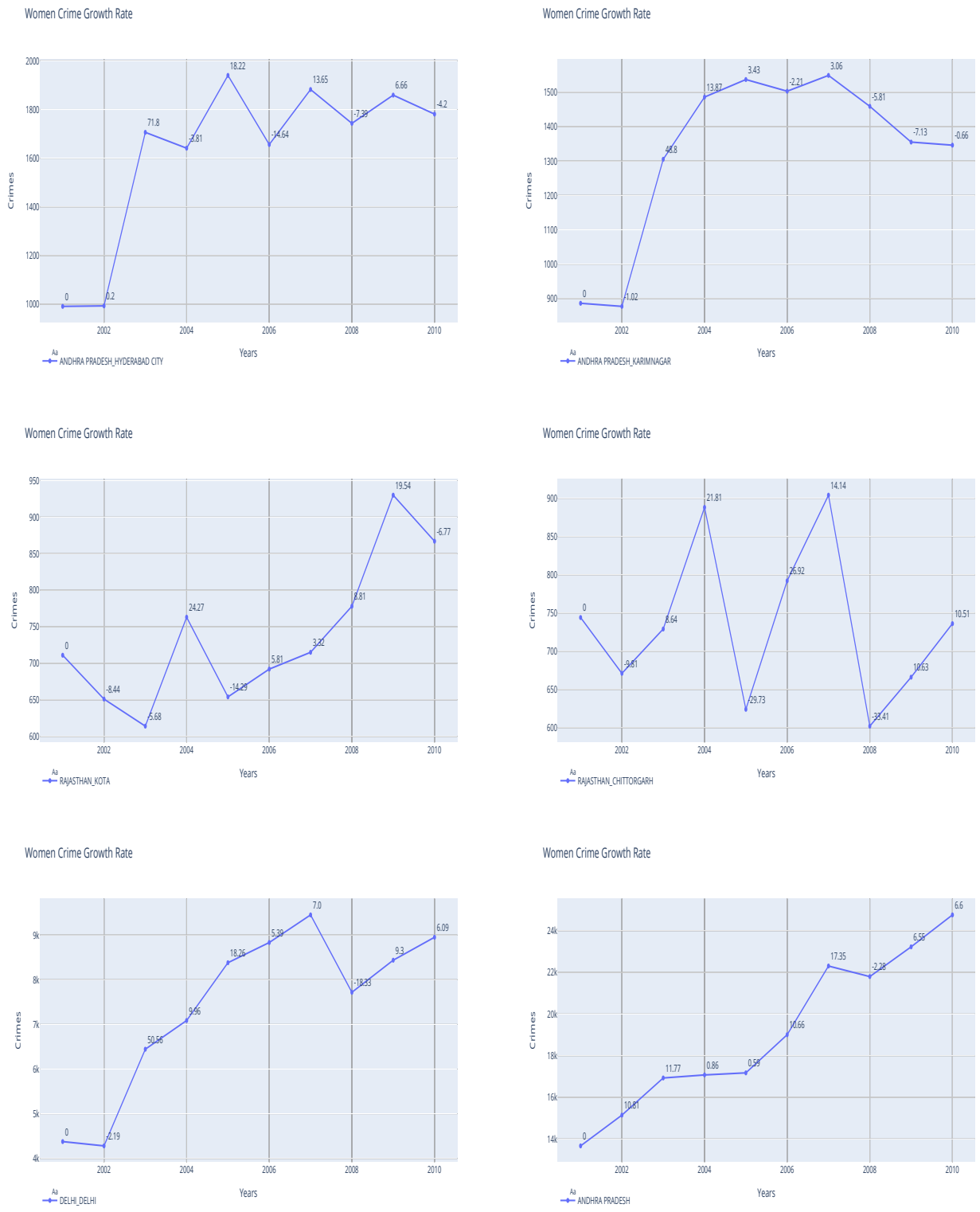

Figure 7.1: (Trend of Crimes against Women in hot-spot districts(1-5 fig number) for the period 2001-2010 (Note: Numbers on the line plots represent the Growth Rate between two consecutive years)

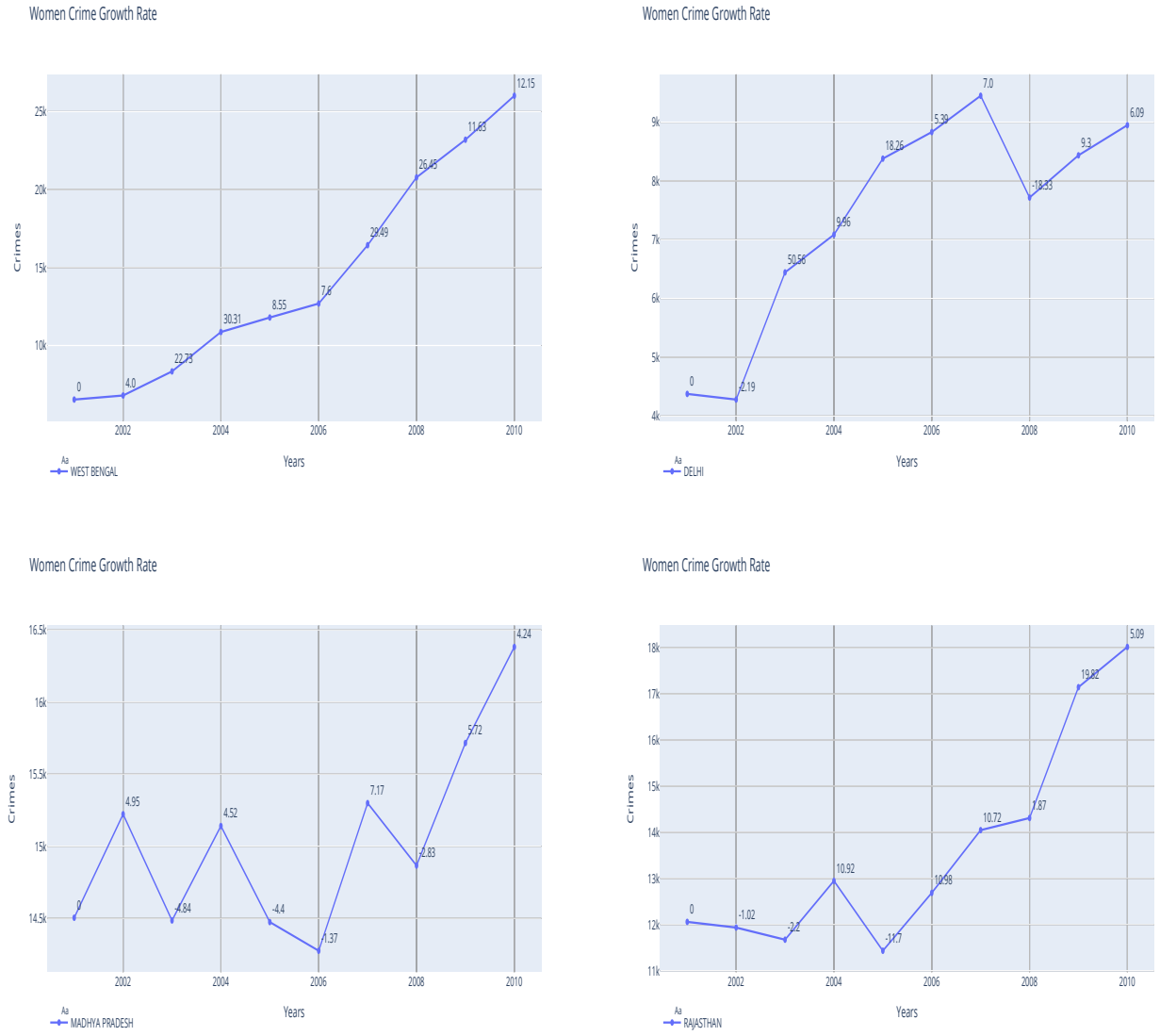

Figure 7.2: (Trend of Crimes against Women in hot-spot states(6-10 fig number) for the period 2001-2010 (Note: Numbers on the line plots represent the Growth Rate between two consecutive years)

As observed from the plots above, crimes in the states have mostly steeply increased over the period of 10 years. The following are the observations:

- For districts, we can see there is a district namely, Rajasthan's Chittoorgarh over the period of 10 years the crime has decreased. In the year 2000, having the total mark as 750 and in 2010 the count came below 750.
- For other districts, with gradual ups and downs the total crime count increased over the period of 10 years.
- Hyderabad city experiences highest growth rate in year 2003 from previous year.
- In case of states, West Bengal is showing a steep increase in consecutive years as well and starting with total crime count below 10000 ending with count of above 25000.
- Looking at growth rates of hot-spots districts/states we can state that crimes against women have significantly increased.

Note: In growth rate calculation, if a previous year had zero number of crimes we assumed growth rate to be zero instead of infinite.

7.1.3 Contribution of sub category

The plot below shows the contribution of sub-category in the total crimes against women.

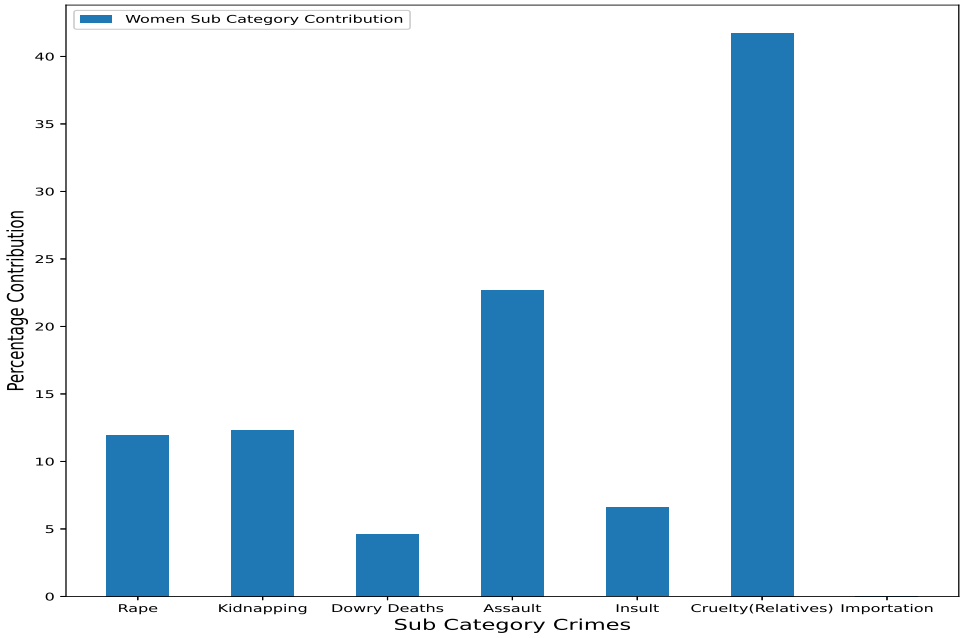

Figure 7.3

7.1.4 Inferences

To develop insights we found correlation between census data and total crimes on women. From correlation matrix, we extracted top 5 census attributes and plotted these attributes for hot-spots and cold-spots districts and states as shown here. Below is a heatmap between census data and different crimes against women. This was used to extract relevant attributes from census data.

The attributes extracted from the heatmap are :

- Literates
- Households with 5 members.
- Households with latrine facility within the premises.
- Power of Parity Rs 45000 to 90000.
- Higher Education.

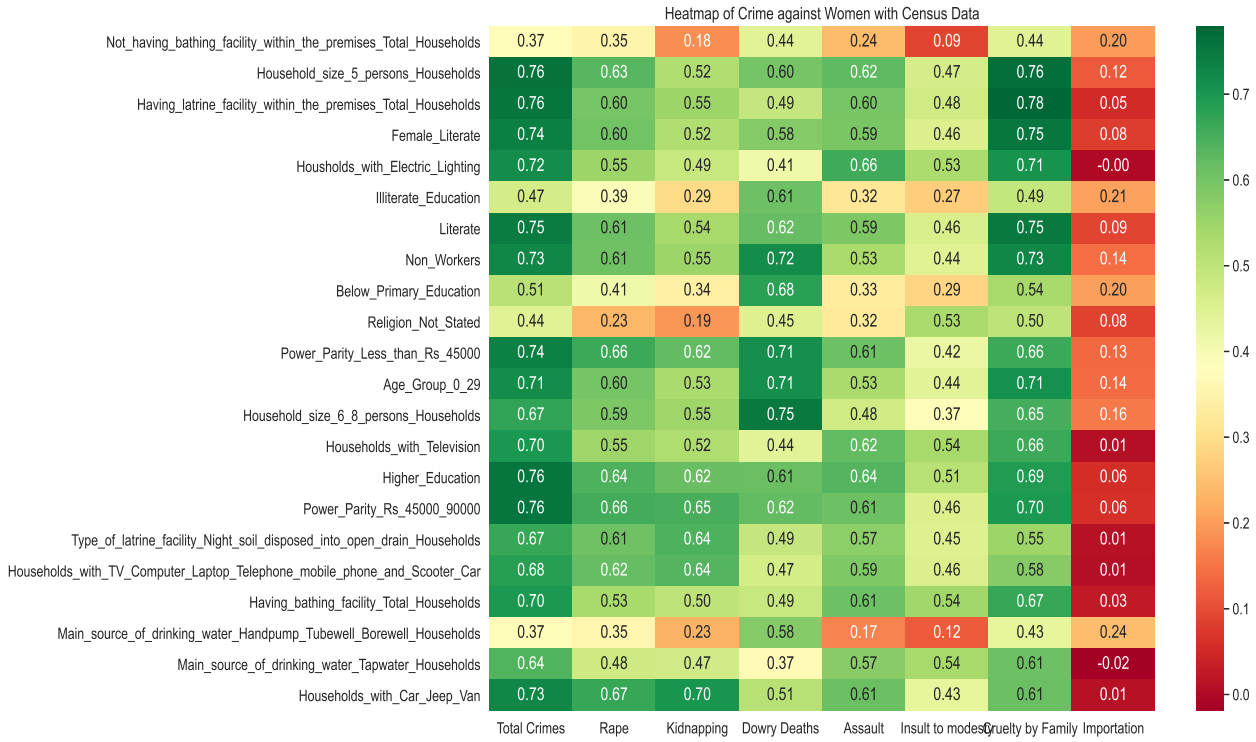

Figure 7.4: Correlation between crimes on women and census attribute

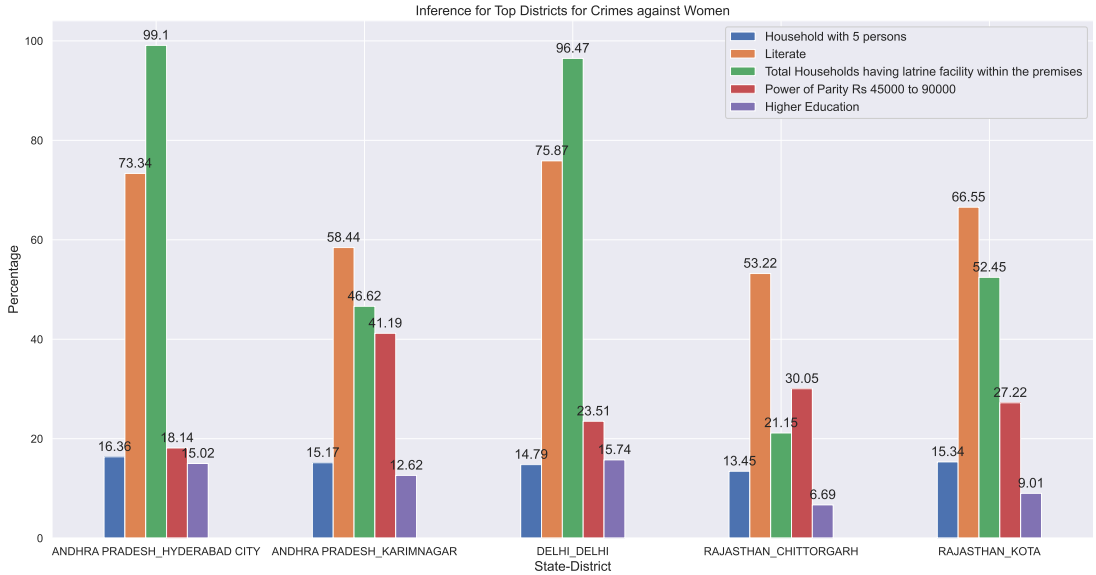

Figure 7.5: Top 5 hot-spot districts with percentage value of census attributes

Here are the plots presented for the Top states/districts with the census attributes extracted from the heatmap above.

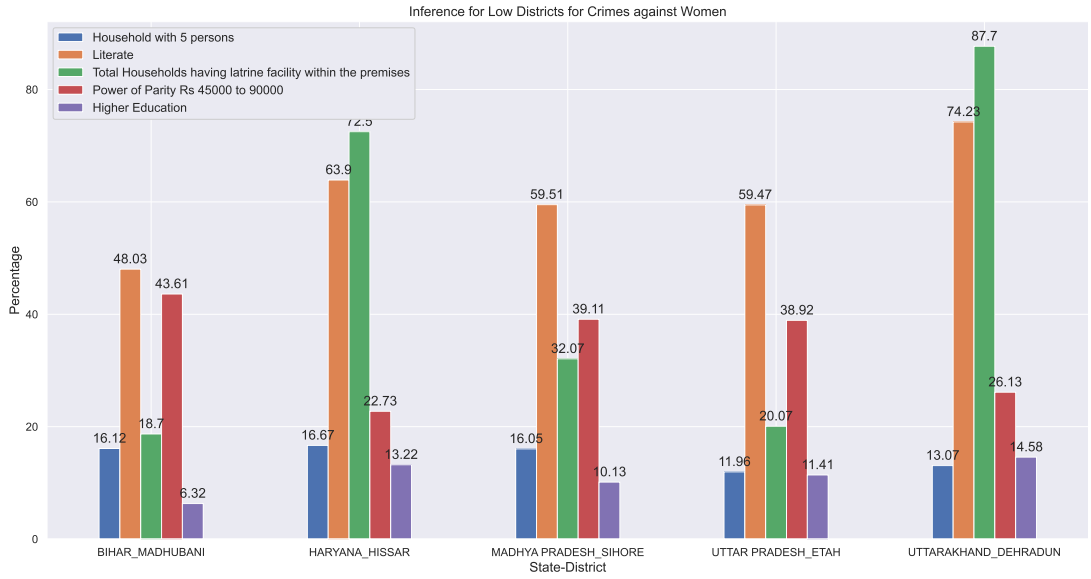

Figure 7.6: Low 5 cold-spot districts with percentage value of census data

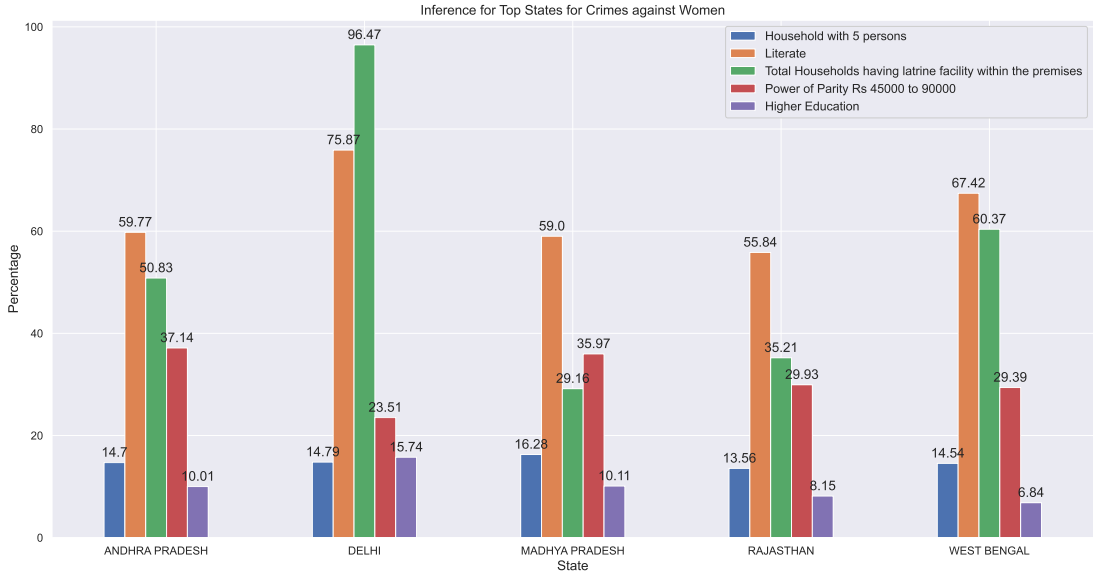

Figure 7.7: Top 5 hot-spot states with percentage value of census data

Inference drawn from the plots are:

- In major cases, districts appearing in top five hot-spot regions are cities.
- Power of Parity is on a higher side in low districts/states as compared to that in top districts/states. This can be considered as one the probable reason for crime against women.

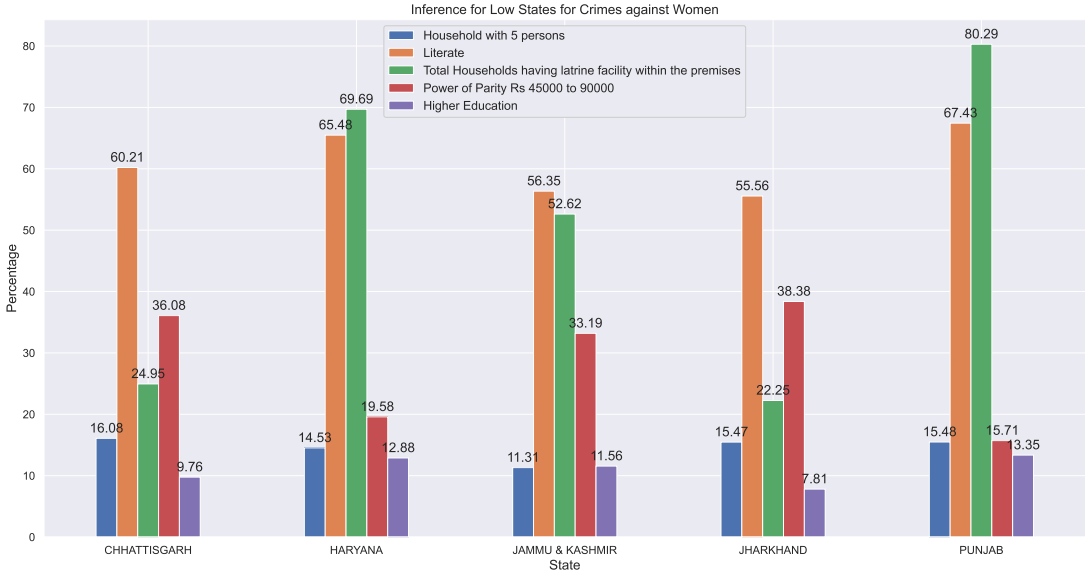

Figure 7.8: Top 5 cold-spot states with percentage value of census attributes

- Higher Education is slightly high in major cases in low districts/states as compared to that in the top districts/states, clearly showing it's impact in total crimes against women.

- Other than the stated factors all other highly correlated factors stand indiscriminating between hot-spots and cold-spot regions.

7.2 Children

7.2.1 Extremum

Table 7.3 and Table 7.4 represents density based and zscore(volume) based hot-spots/cold-spots districts/states for crimes on Children.

Population			
Districts		**States**	
Top	Low	Top	Low
DELHI	RAIGARH	MADHYA PRADESH	BIHAR
INDORE	CHITTOOR	DELHI	WEST BENGAL
KHANDWA	AMRELI	CHHATISGARH	TAMIL NADU
BETUL	CHENNAI	UTTAR PRADESH	KERLA
RAJGARH	SHIVPURI	MAHARASHTRA	KARNATAKA

Table 7.3: Density based top 5 hot-spots/cold-spots districts and states.

Zscore			
Districts		**States**	
Top	Low	Top	Low
DELHI	KAMENG WEST	MADHYA PRADESH	LAKSHADWEEP
PUNE	TAWANG	MAHARASHTRA	DAMAN & DIU
MUMBAI	KAITHAL	UTTAR PRADESH	NAGALAND
INDORE	JAMMU	DELHI	D & N HAVELI
NAGPUR	KARGIL	ANDHRA PRADESH	PUDUCHERRY

Table 7.4: Zscore based top 5 hot-spots/cold-spots districts and states.

We made the following key observations:

- Delhi and Indore are the hot-spot districts in case of crimes against children in both zscore and population normalized method.
- Madhya Pradesh, Delhi, UP and Madhya Pradesh are hot-spot states in case of both zscore and population normalized method
- Betul and Khandwa are in the top 5 hot-spot districts in case of population normalized as the population of these districts is very less.

7.2.2 Growth Rate

As observed from the plots below, crimes in the states has mostly steeply increased over the period of 10 years. The following are the observations:

- In mostly all the districts there is a significant change in crime from year 2003 to 2004, the maximum is in the Indore district of Madhya Pradesh.
- There are three districts of Madhya Pradesh - Indore, Betul and Khandwa which are present in the top 5 hot-spot districts in crime against children.
- In case of states, Madhya Pradesh and Maharashtra are the two states in which the number of crimes against children has always increased from the year 2001 to 2010.
- In Chhattisgarh, the number of crimes against children has continuously increased from the year 2007 to 2010.
- Overall, Madhya Pradesh has the maximum increase in the number of crimes against children.

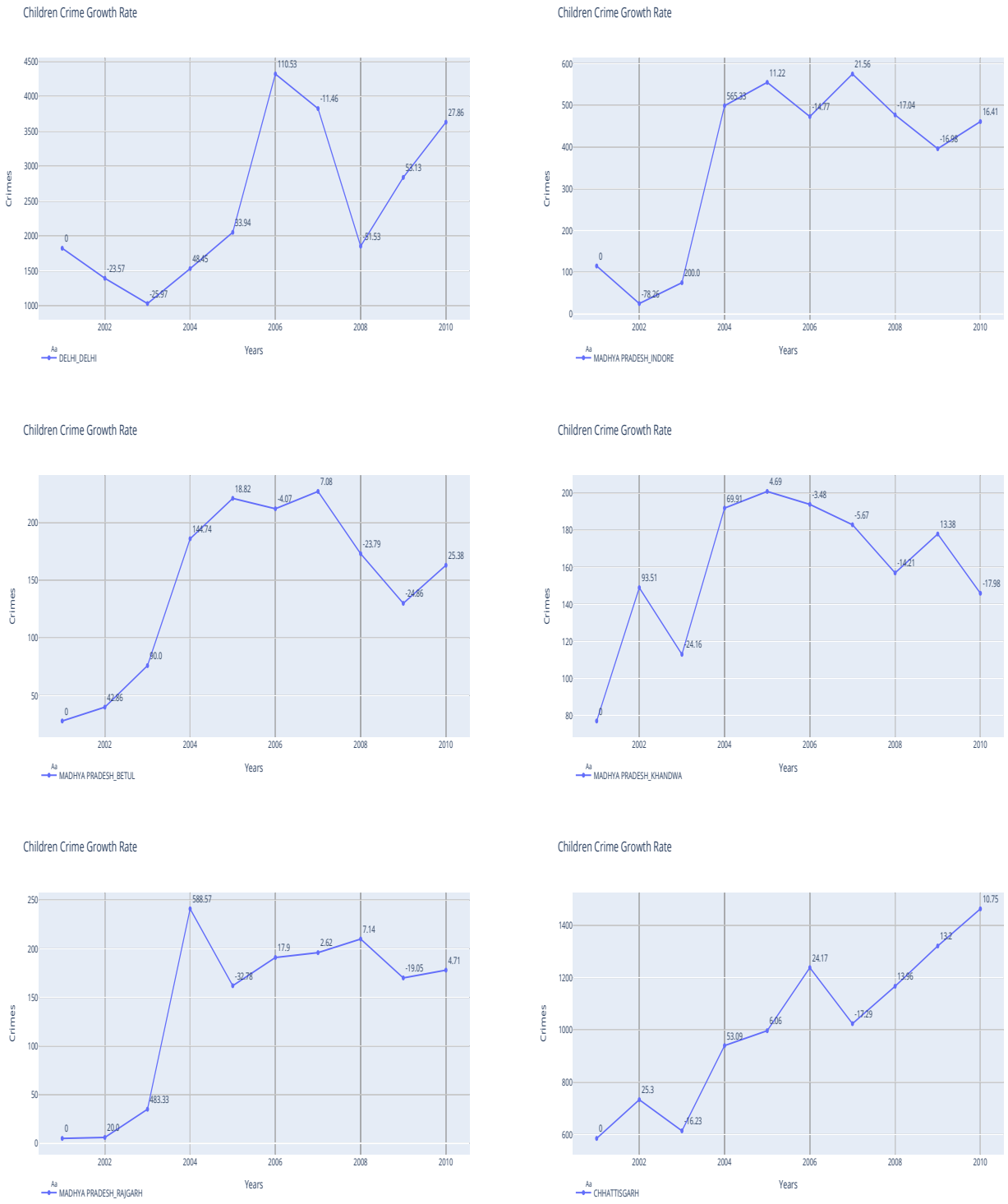

Figure 7.9: (Trend of Crimes against Women in hot-spot districts(1-5 fig number) for the period 2001-2010 (Note: Numbers on the line plots represent the Growth Rate between two consecutive years)

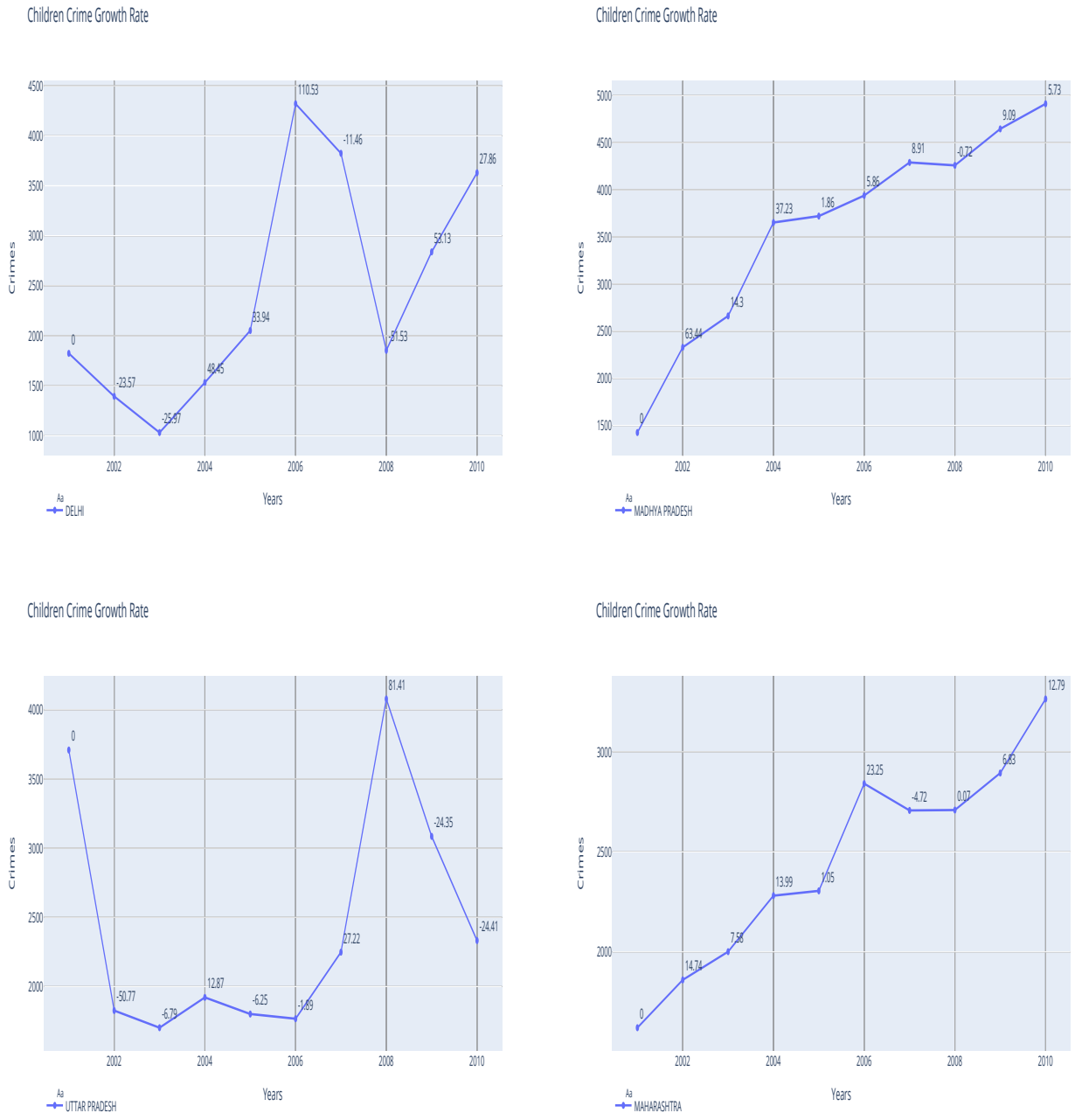

Figure 7.10: (Trend of Crimes against Women in hot-spot states(6-10 fig number) for the period 2001-2010 (Note: Numbers on the line plots represent the Growth Rate between two consecutive years)

7.2.3 Contribution of sub crimes

The plot below shows the contribution of the subcategory in the total crimes against children. Below bar plot represents the percentage contribution of sub crimes in overall crimes against children.

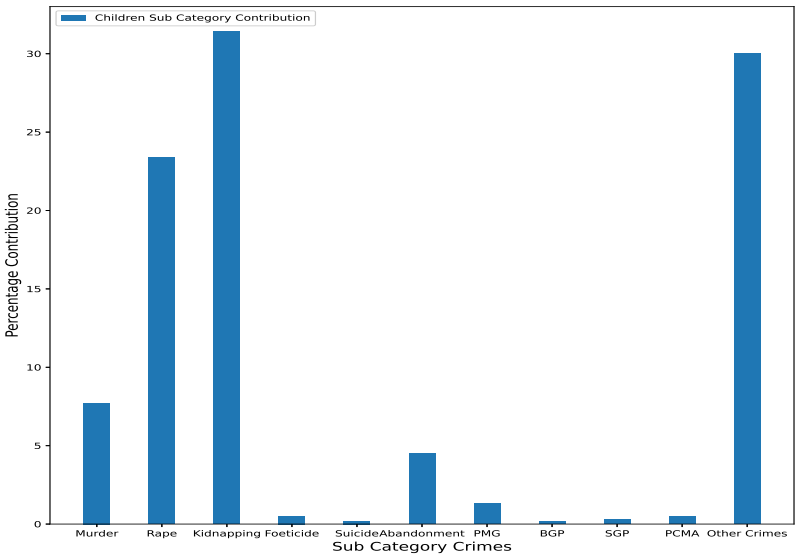

Figure 7.11

7.2.4 Inference

Below is a heatmap between census data and different crimes on children. This was used to extract relevant attributes from the census data.

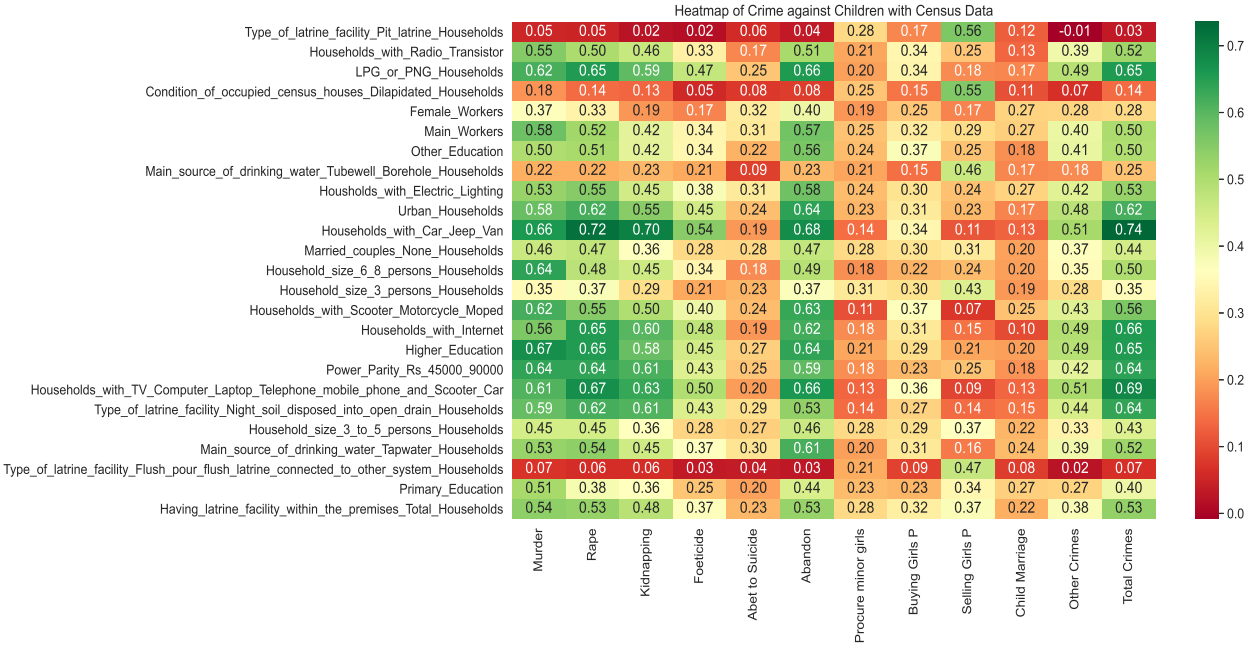

Figure 7.12

The attributes extracted from the heatmap are :

- Households with Internet
- Households with LPG or PNG
- Households with TV, Computer, Laptop, Telephone, mobile, Scooter and Car
- Households with Car, Jeep and Van
- Higher Education.

Below are the plots presented for the Top states/districts with the census attributes extracted from the heatmap above.

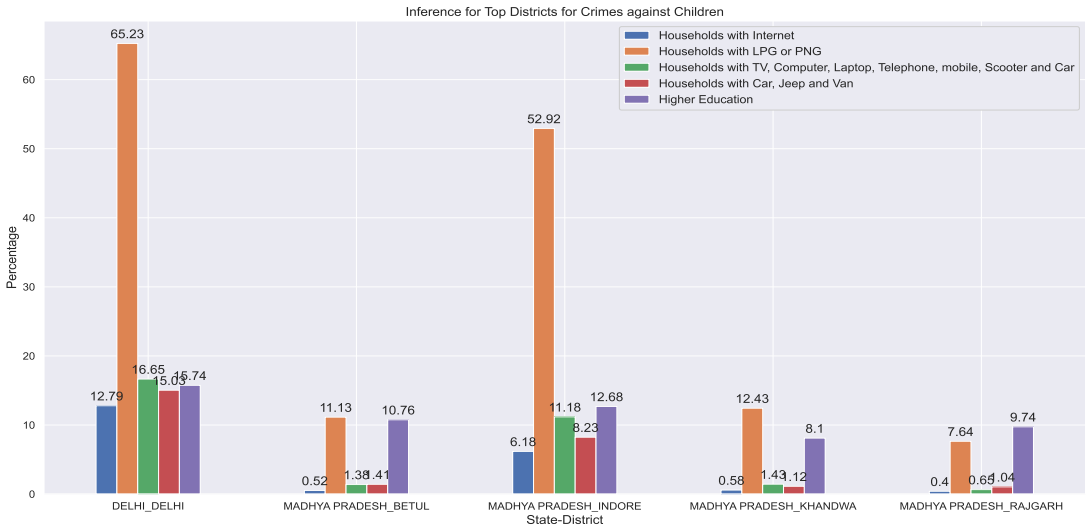

Figure 7.13

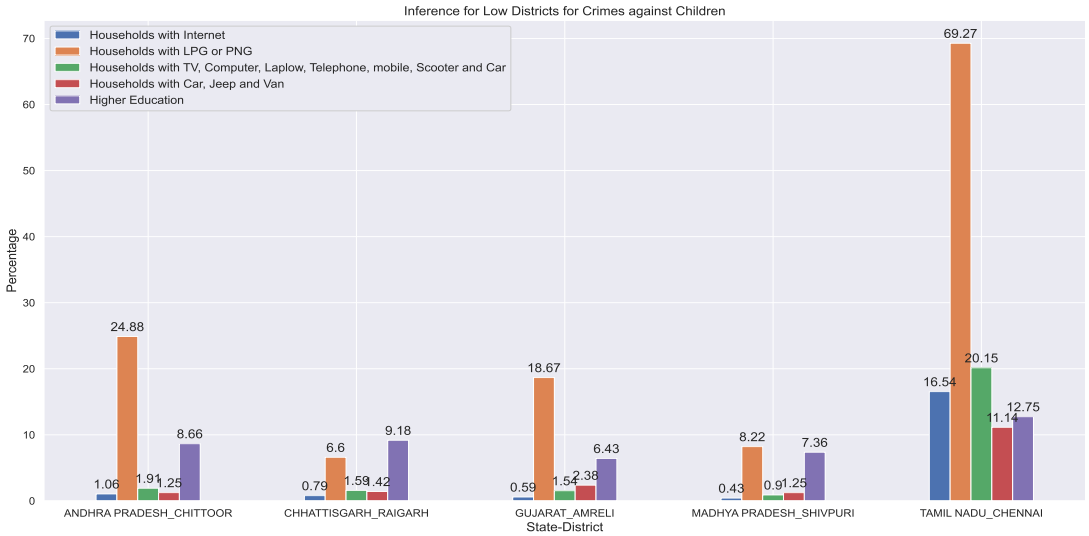

Figure 7.14

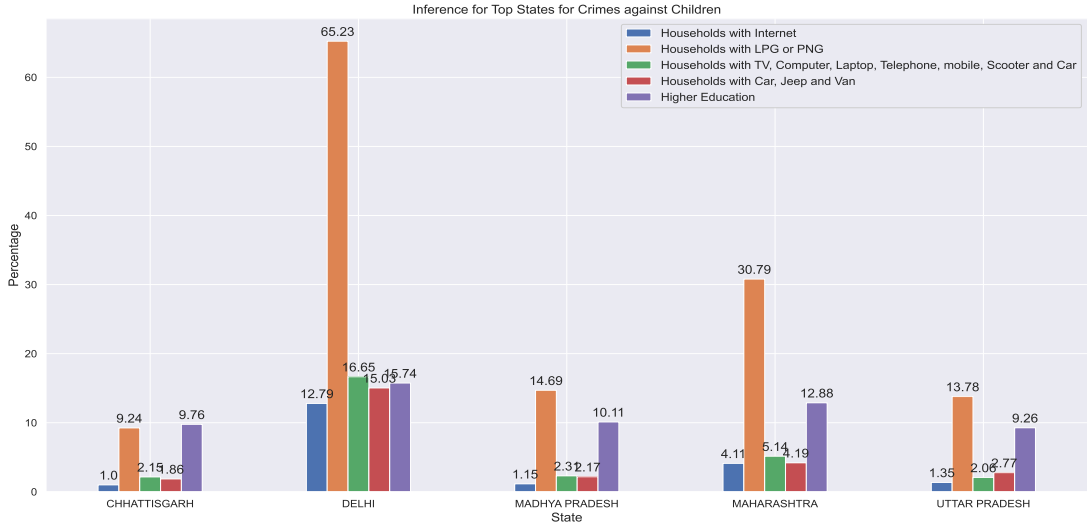

Figure 7.15

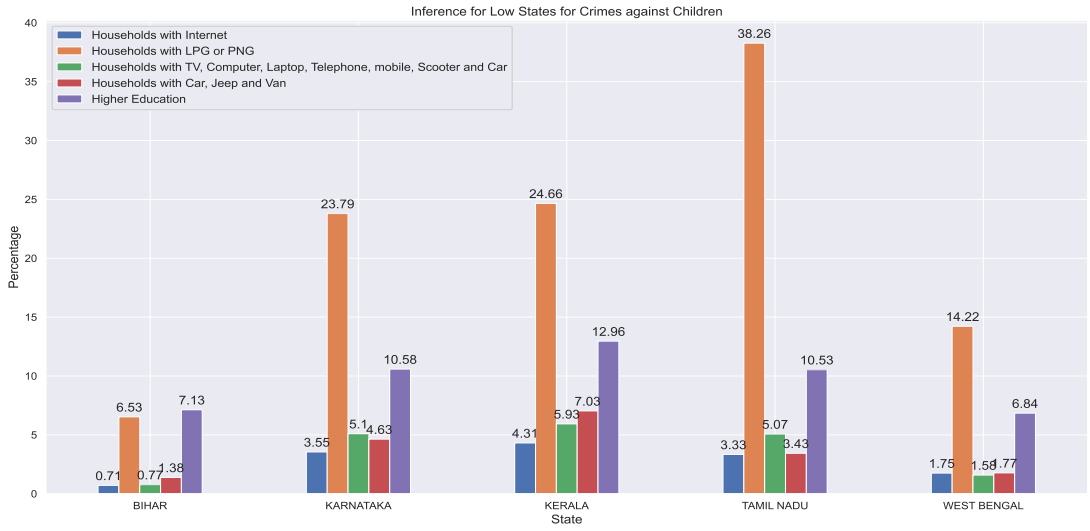

Figure 7.16

Inference drawn from the plots are:

- Delhi and Indore are the top hot-spot districts in case of crimes against children and the census data attributes like households with LPG or PNG are also high.
- However, there is exception in case of Betul, Khandwa and Rajgarh which are also in the hot-spot district but the attributes of census data like households with LPG or PNG, households with Internet are quite less.
- There is only one attribute that is Higher Education which can be accounted for majority of crimes as it is somewhat constant in all districts.

7.3 SC

7.3.1 Extremum

Table 7.5 and Table 7.6 represents density based and zscore(volume) based hot-spots/cold-spots districts/states for crimes on Children.

Population			
Districts		**States**	
Top	Low	Top	Low
BHARATPUR	KAUSHAMBI	MADHYA PRADESH	MAHARASHTRA
UJJAIN	ADILABAD	RAJASTHAN	WEST BENGAL
GUNA	PUNE	ANDHRA PRADESH	PUNJAB
DEWAS	DAVANAGERE	UTTAR PRADESH	HARYANA
DHOLPUR	AMBEDKAR NAGAR	BIHAR	JHARKHAND

Table 7.5: Density based top 5 hot-spots/cold-spots districts and states.

Zscore			
Districts		**States**	
Top	Low	Top	Low
BHARATPUR	NALBARI	UTTAR PRADESH	MIZORAM
UJJAIN	NARAYANPUR	RAJASTHAN	DAMAN & DIU
KANPUR	JAMMU	MADHYA PRADESH	D & N HAVELI
SITAPUR	RAJOURI	ANDHRA PRADESH	CHANDIGARH
GANGANAGAR	RAMBAN	BIHAR	GOA

Table 7.6: zscore based top 5 hot-spots/cold-spots districts and states.

We make following key observations:

- Although mostly the results for both methods fall in the intersection for hot-spot districts while on the other hand for cold-spots districts results vary.
- Haryana comes in cold-spots for population normalized value on crimes whereas the zscore method excludes it.
- Punjab comes in cold-spots for population method which is not included in the zscore results.
- Following the zscore method, Pune is not a cold-spot but it is covered as a cold-spot by the density based method.

7.3.2 Growth Rate

We have analysed the growth rate of hot-spots/cold-spots districts/states for crimes on SC over the years 2001 to 2010. Below plots represent the growth rate of crimes on SC over the period 2001 to 2010

As observed from the plots above, crimes in the states has mostly steeply increased over the period of 10 years. The following are the observations:

- Districts of Madhya Pradesh(Ujjain, Guna, and Dewas) experience decrease in crimes on SC over the ten years.
- Crimes on SC have increased in districts of Rajasthan(Dholpur and Bharatpur) over the period of ten years.
- Hyderabad city experiences highest growth rate in the year 2003 from previous year.
- In case of hot-spot states, crimes on SC in Uttar Pradesh have decreased while for Bihar there is a gradual increase in crimes against SCs.

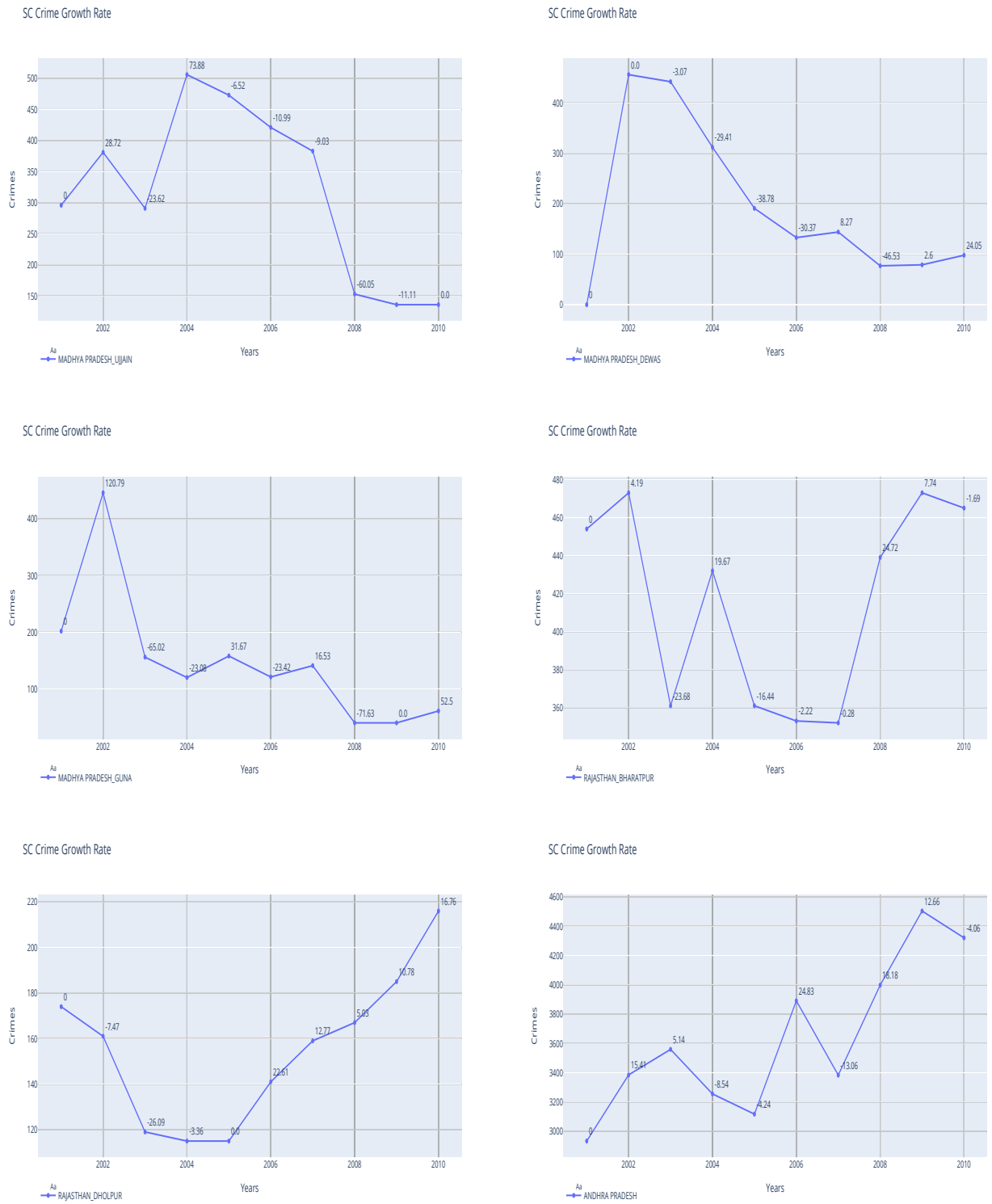

Figure 7.17: Trend of Crimes against SC in top districts(1-5 fig) for the period 2001-2010 (Note: Numbers on the line plots represent Growth Rate for two consecutive years)

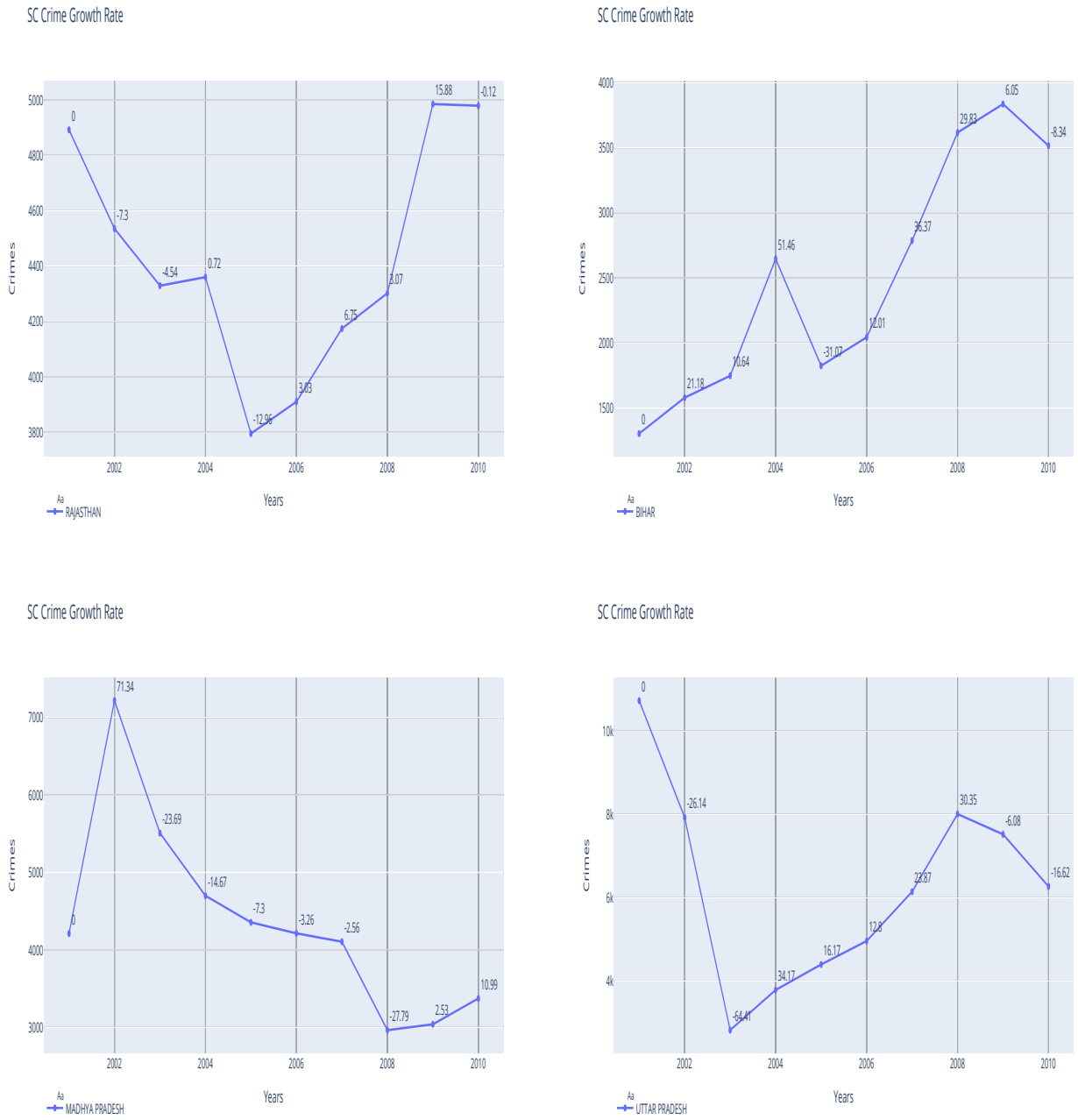

Figure 7.18: Trend of Crimes against SC in top states(6-10 fig) for the period 2001-2010 (Note: Numbers on the line plots represent Growth Rate for two consecutive years)

7.3.3 Contribution of sub category

The plot here shows the contribution of sub-category in the total crimes against SCs.

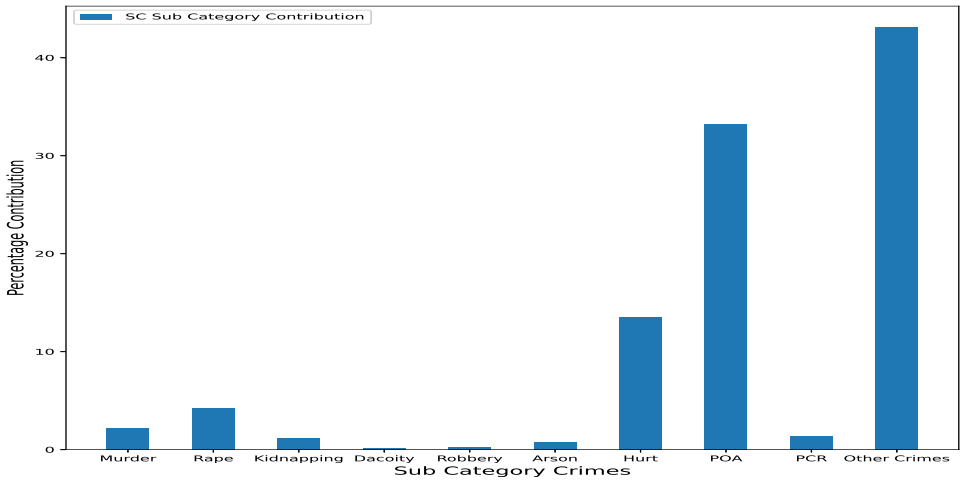

Figure 7.19

7.3.4 Inference

Below is a heatmap between census data and different crimes against SCs. This was used to extract relevant attributes from census data.

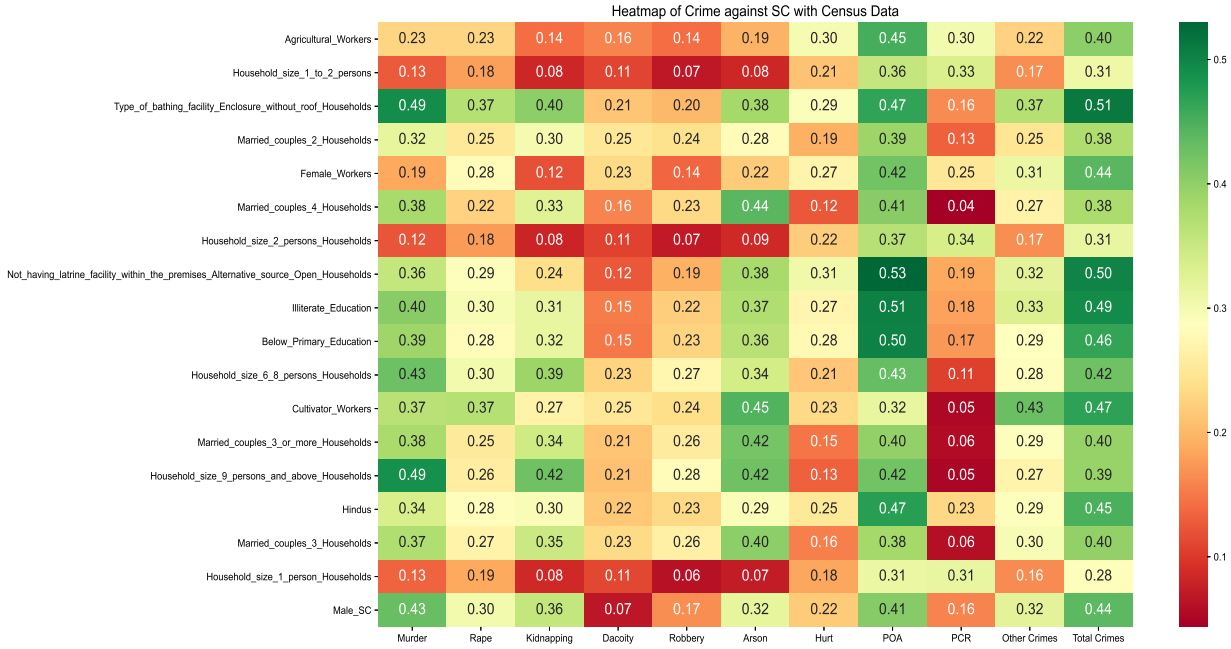

Figure 7.20

The attributes extracted from the heatmap are :

- No latrine facility within the premises.

- Cultivator Workers.
- Below Primary Education roof
- Bathing facility enclosure without roof.
- Illiterate.

Below are the plots presented for the Top states/districts with the census attributes extracted from the heatmap above.

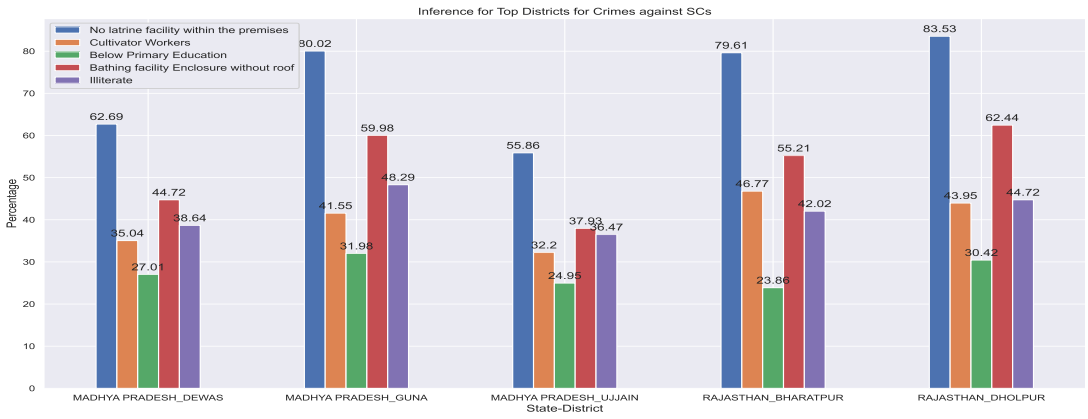

Figure 7.21

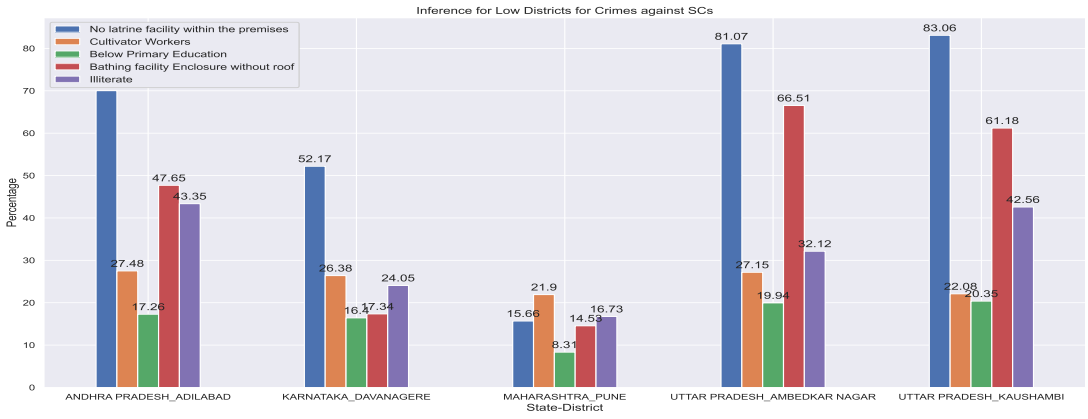

Figure 7.22

Inference drawn from the plots are:

- Considering Illiteracy as a factor, wherever it is low in percentage that state/district is having lo crime count against SCs. Example, State Punjab(28 percent Illiteracy) is a cold-spot but Bihar(41 percent Illiteracy) is a hot-spot. In districts, Pune(16 percent Illiterates) is a cold-spot and Bharatpur(42 percent Illiterates) is a hot-spot for crimes against SCs. Therefore, decreasing illiteracy will surely decrease the crime count against SCs.
- For the factor, Bathing Facility without roof wherever it's percentage is more like in Ujjain District(38 percent) is a hot-spot but in Pune(14 percent) is a cold-spot. For states, Madhya Pradesh with 45 percent such cases is a hot-spot but Maharashtra with 24 percent is a cold-spot for crimes against SCs. Thus, making bathing facilities with rooftops will decrease crime counts against SCs

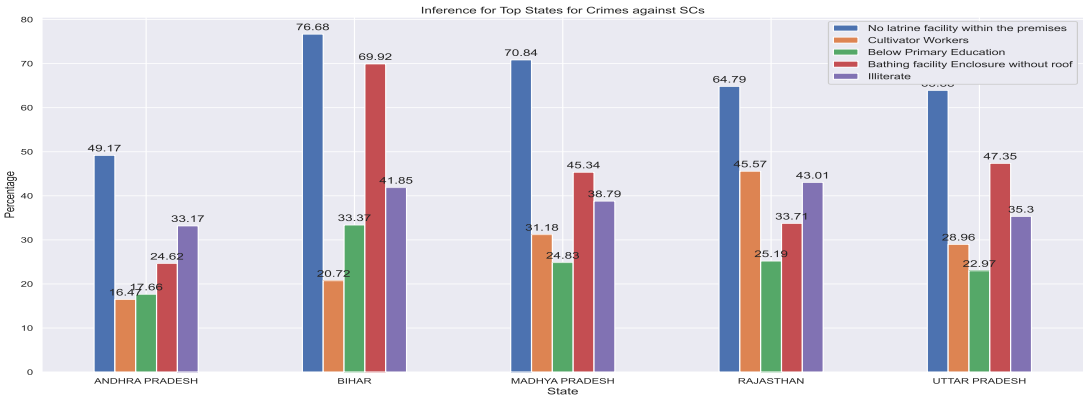

Figure 7.23

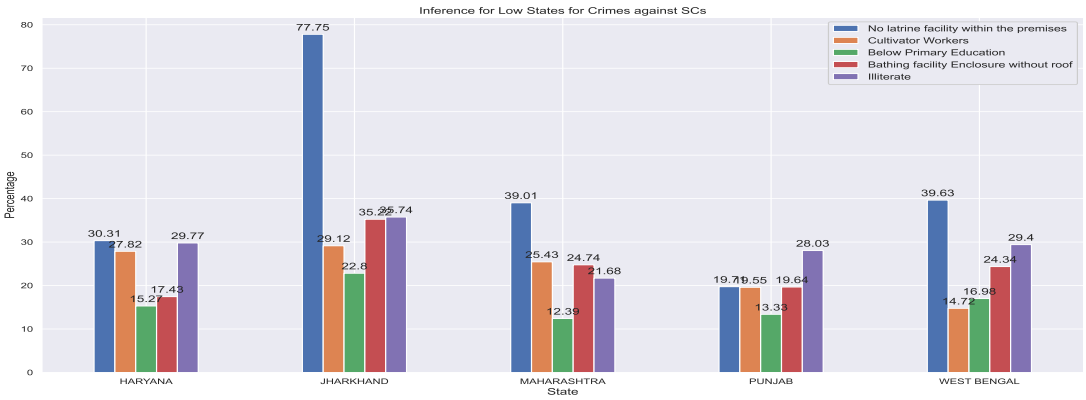

Figure 7.24

- Below Primary Education is also a very significant factor to consider for crime against SCs. Example, Uttar Pradesh(23 percent people having below primary education) is a hot-spot and West Bengal(16 percent people with below primary education) is a cold-spot for crime against SCs. In districts, Dewas district of Madhya Pradesh(27 percent) is hot-spot and district Adilabad of Andhra Pradesh(17 percent people with below primary education) is a cold-spot for crime against SCs. Therefore, providing people with Primary Education will surely decrease crime cases against SCs.
- Cultivator Workers is again a factor clearly showing the difference between hot/cold spots. Example, District Dholpur of Rajasthan(43 percent cultivator workers) is a hot-spot whereas Kaushambi of Uttar Pradesh(22 percent cultivator workers) is a cold-spot. Additionally, state Rajasthan(45 percent cultivator workers) is a hot-spot and Jharkhand(29 percent cultivator workers) is a coldpot. This indicates that, majority of the cultivator worker population comprise of SCs and are vulnerable of crimes.
- All the factors very well show the difference between hot-spots and cold-spot regions.

7.4 ST

7.4.1 Extremum

Population			
Districts		**States**	
Top	Low	Top	Low
GWALIOR	JHABUA	RAJASTHAN	WEST BENGAL
BHARATPUR	AMRAVATI	MADHYA PRADESH	ASSAM
TONK	SRIKAKULAM	ANDHRA PRADESH	NAGALAND
KARIMNAGAR	VALSAD	CHHATTISGARH	MEGHALAYA
SIKAR	SUNDARGARH	ODISHA	BIHAR

Zscore			
Districts		**States**	
Top	Low	Top	Low
JAIPUR	TAWANG	MADHYA PRADESH	LAKSHADWEEP
NALGONDA	GOPALGANJ	RAJASTHAN	MEGHALAYA
BETUL	NORTH GOA	ANDHRA PRADESH	GOA
MAYURBHANJ	SOUTH GOA	ODISHA	DAMAN & DIU
SEONI	MANDI	CHHATTISGARH	JAMMU & KASHMIR

We make following key observations:

- hot-spot districts extracted from both methods differ completely.
- West Bengal is a top hot-spot state when extracted with normalized population method.
- Bihar is out of top hot-spot states using density based method.

7.4.2 Growth Rate

As observed from the plots below, crimes in the states has mostly steeply increased over the period of 10 years. The following are the observations:

- District Karimnagar of Andhra Pradesh, the crime count has decreasde over the period of ten years.
- Crimes on ST have increased in districts of Rajasthan(Sikar and Bharatpur) over the period of ten years.
- State Odisha, Madhya Pradesh and Chhattisgarh experienced loss in growth rate over the period 10 years.
- Crimes on ST in Rajasthan have increased over the period of ten years while of states Madhya Pradesh and Chattisgarh have decreased significantly.

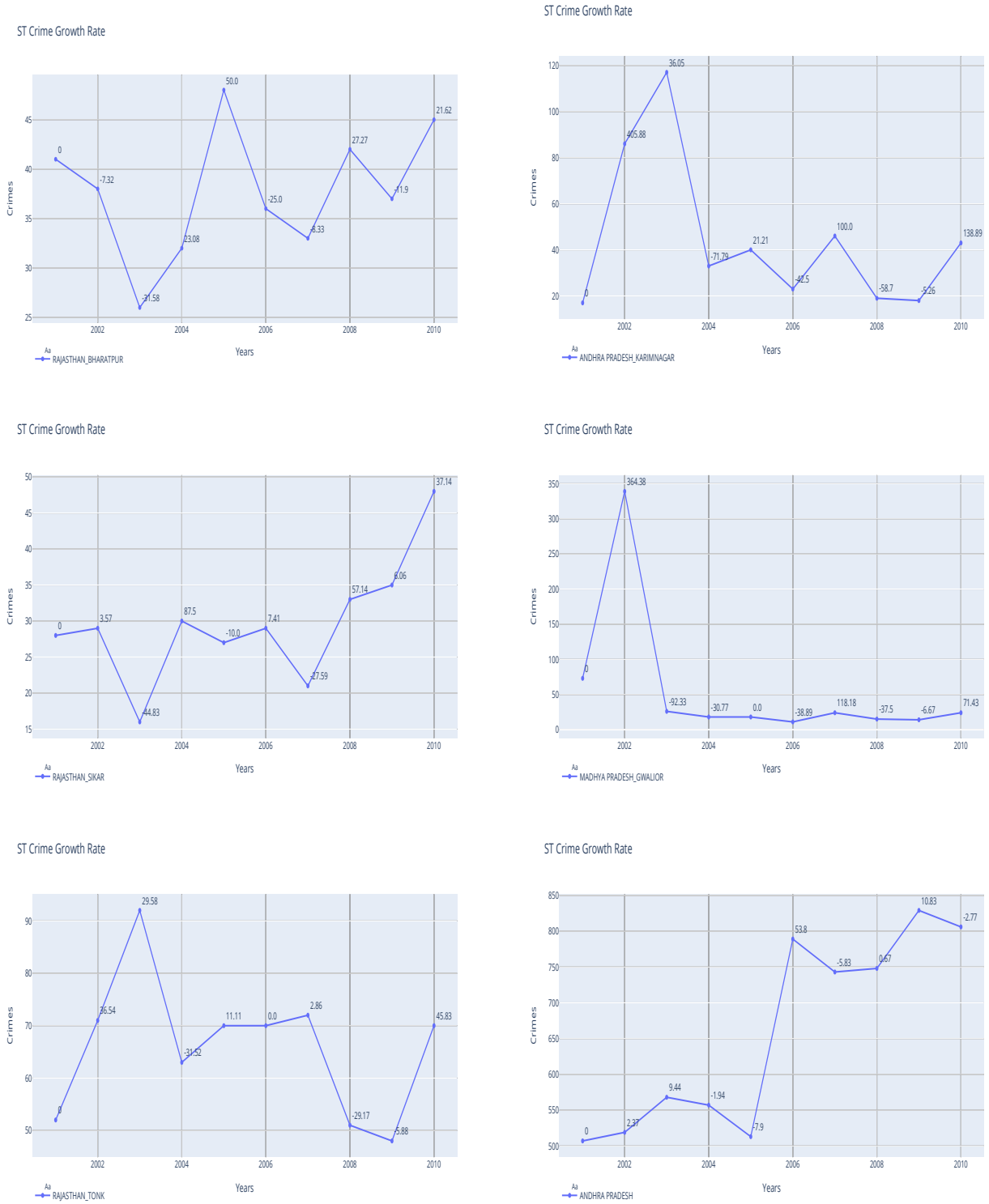

Figure 7.25: Trend of Crimes against ST in top districts(1-5 fig) for the period 2001-2010 (Note: Numbers on the line plots represent Growth Rate for two consecutive years)

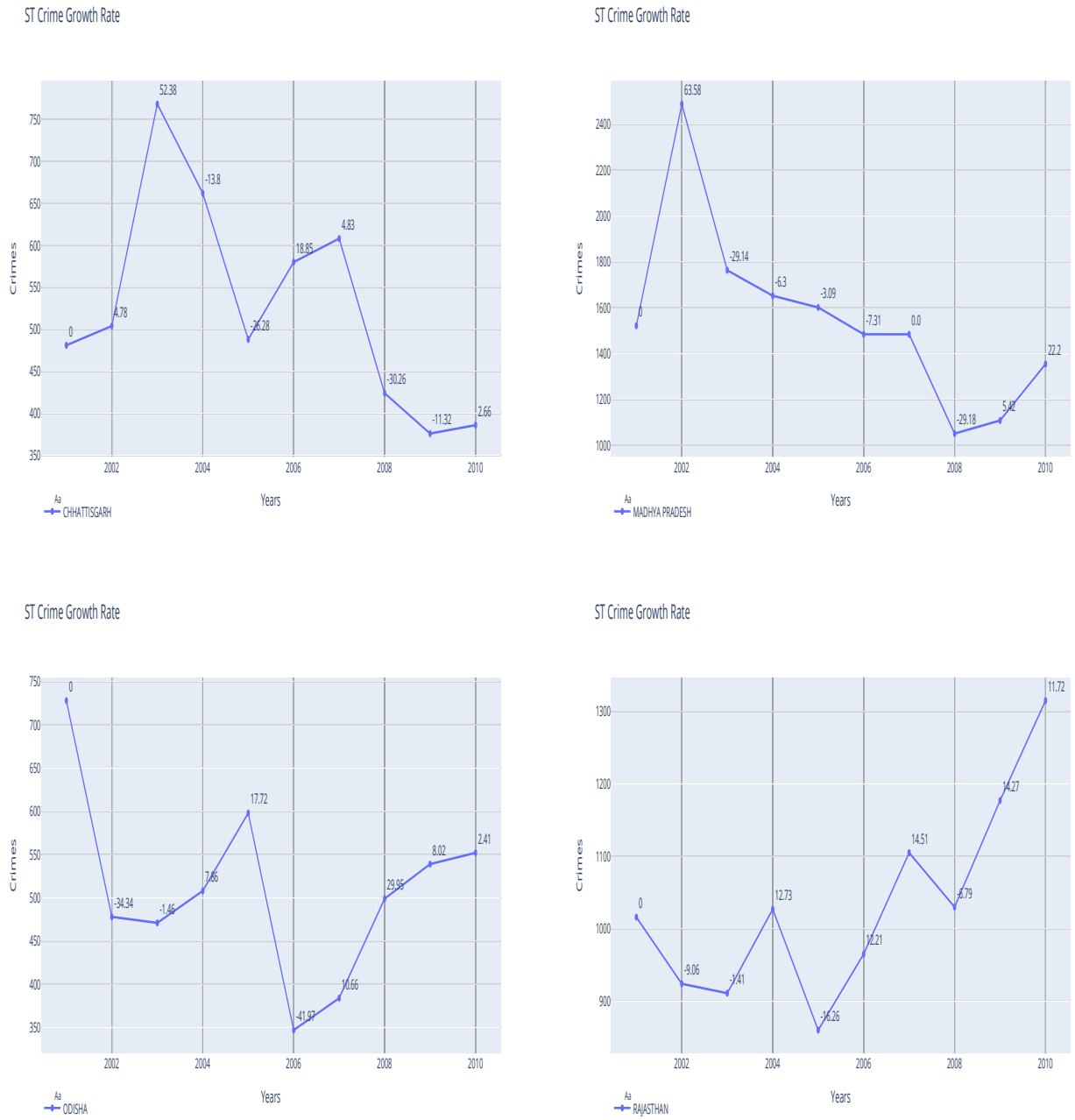

Figure 7.26: Trend of Crimes against ST in top states(6-10 fig) for the period 2001-2010 (Note: Numbers on the line plots represent Growth Rate for two consecutive years)

7.4.3 Contribution of sub category

Below plot represents the percentage contribution of a sub category in total crimes committed against ST.

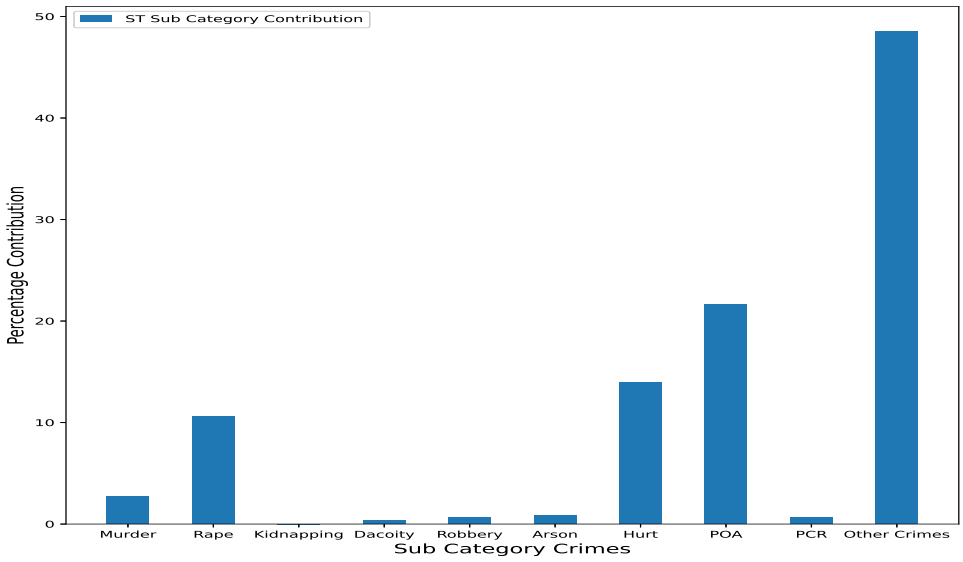

Figure 7.27

7.4.4 Inference

Below is a heatmap between census data and different crimes against STs. This was used to extract relevant attributes from census data.

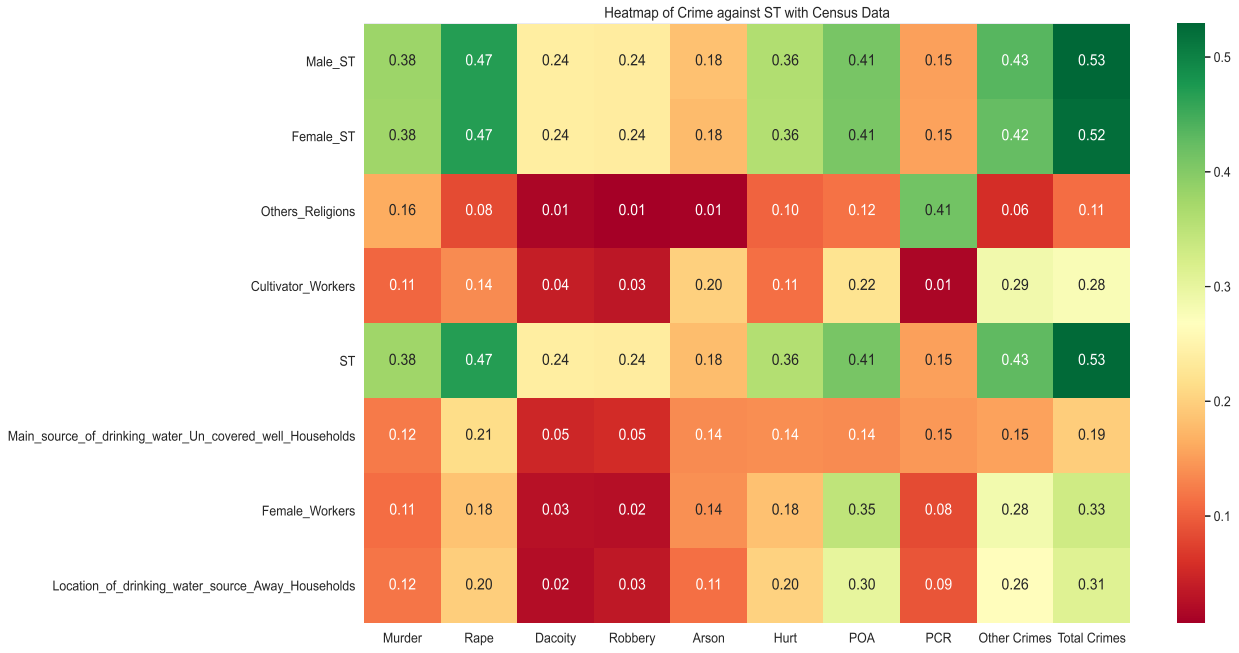

Figure 7.28

The attributes extracted from the heatmap are :

- No latrine facility within the premises.
- Cultivator Workers.
- Below Primary Education roof
- Bathing facility enclosure without roof.
- Illiterate.

Below are the plots presented for the Top states/districts with the census attributes extracted from the heatmap above.

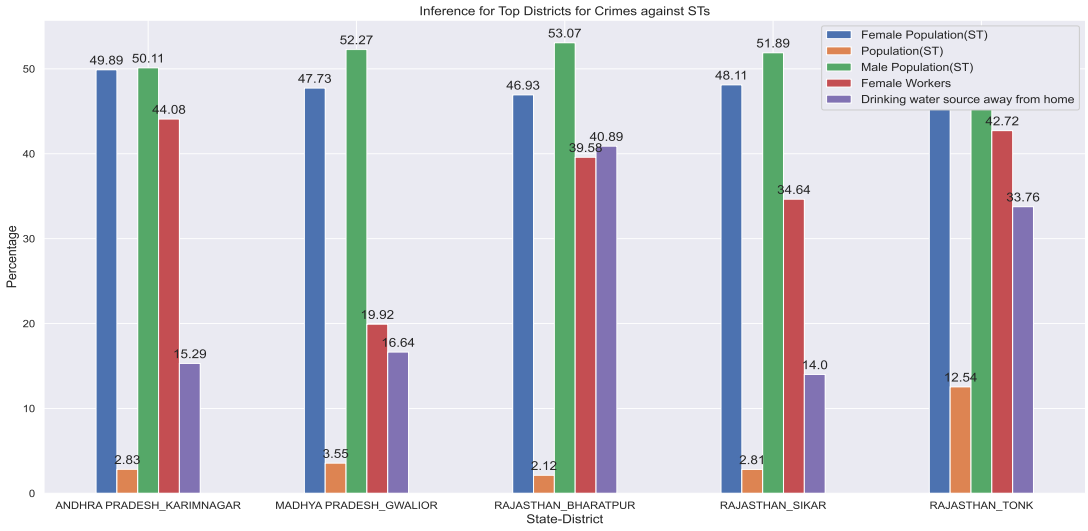

Figure 7.29

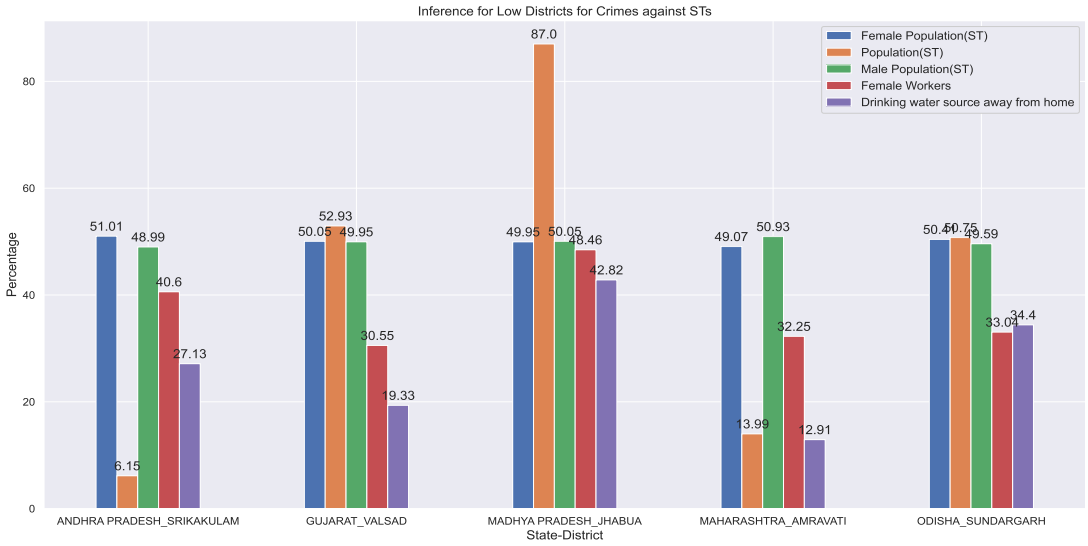

Figure 7.30

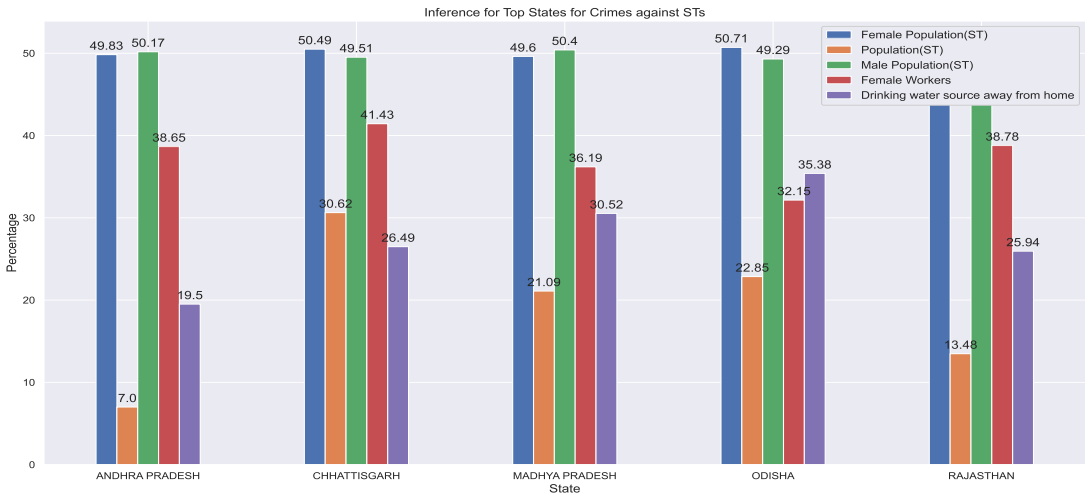

Figure 7.31

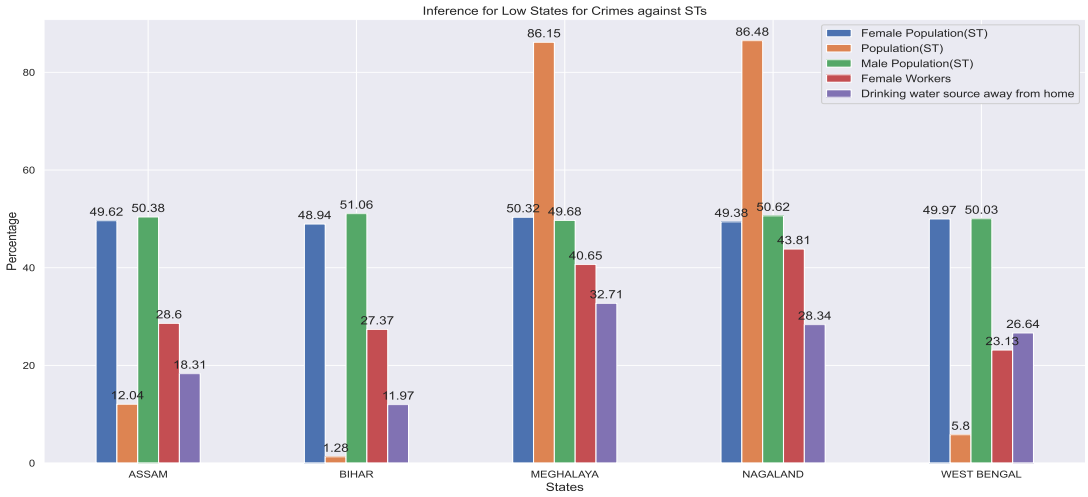

Figure 7.32

Inference drawn from the plots are:

- Comparative look at plots of cold-spot and hot-spot district shows that districts with more ST population have less crimes than districts with less ST population.
- hot-spot districts/states have more male ST population than cold-spot districts.
- Majorly hot-spot districts have high female workers than the cold-spot districts.

8 DISCUSSION

Our analysis can be summarized on four types of crimes we have analyzed:

- **SC:** Majorly crimes on SC are committed in regions with high illiteracy rate. Also in regions with more population having below primary education. We found ”bathing facility without roof” in high correlation

with crimes on SC, which can be considered as an factor affecting crimes on SC in rural areas. Crimes related to "Prevention of atrocities" has major contribution in crimes on SC.

- **ST:** Regions with less ST population experience more number of crimes on ST. Also hotspot regions have more female ST working as compared to coldspot regions, which can be a probable reason for crime against ST womens. Rape on ST women is the third most contributed subcategory in ST crimes.
- **Children:** Crimes on children are more frequent in urban areas compared to rural areas. Regions with more higher educated population experience less crimes on children. One key observation is crimes on children are highly correlated with number of households with car and jeep.
- **Women:** Crimes on women are more frequent in urban areas as compared to rural areas. Power of parity is highly correlated with coldspot regions i.e., regions with high power parity have less number of crimes on women. Cruelty by husband or relatives is most contributed sub category in crime against women.

9 Future Directions

- We have only considered positively correlated census attributes while some negatively correlated attributes might help to explain analysis better.
- Comparative analysis can be improved by considering urban areas and rural areas separately.
- Prediction model could be trained to predict crimes in future.
- Other crimes might be considered which will expand breadth of the analysis.
- Data could be made more consistent using interpolation or regression model to predict missing values.
- Availability of more recent data may improve our results.