Information Visualization Final Report

Crime Against Women in India

Ruchita Rathi, Janine Heiser, Ankita Bhosle

Problem Statement and Motivation

Crime has always been one of the larger problems in India, especially since it is the second most populous country in the world, and a bane of India's development efforts. Certain cultural peculiarities, such as sex being a taboo in India, the skewed sex ratio, age-old customs like Sati and Dowry and the lower status women hold in society leave women at higher risk to become victims of sexualized violence. There is also lack of awareness, inadequate infrastructure and insufficient programs in place for their safety and protection. Sexualized violence is most severe in rural areas, but some urban areas like Tier 1 cities, and surprisingly, even the capital, are recognized as high risk zones.

Most recently, a series of very brutal crimes against young girls (e.g. the Delhi Gang Rape Case, Scarlet Keeling Rape and Murder) drew national and global attention to a longstanding problem. In response to these horrific acts, the safety of women has suddenly become a national priority; the issue has gained visibility through different journalistic media such as articles by influential columnists, blog posts, short films, celebrities campaigning about the issue, and social media propaganda.

These motivations compel us to analyze crime data from India with the goal of gaining an insight into the nature and depth of the problem. We use data from the Open Government Data (OGD) Platform India (http://data.gov.in). Our hope is that this project becomes a drop in the ocean of existing material to raise some awareness about the issue.

Project Goals

We aim to create a series of visualizations to allow exploration and analysis of data about Crime Against Women in India. Using several categories of data, we attempt to illuminate the following:

- Show which types of crimes are the most rampant and whether that aligns with preconceived notions about most common crimes.
- Provide historical data about crimes against women to give the viewer an overall sense of the historical trend in various crimes.
- Provide general profiling information about the perpetrators of crimes against women; specifically, we look at gender and age. Specially, we wanted to show that women, particularly in domestic abuse cases can harm other women.
- Show the geographical concentrations of crimes (especially since our audience here is not familiar with the states and general geography of the country)
- Show the role of external variables like literacy rate and sex ratio on crime.

We had also originally included information about the disparity between cases reported and actual convictions (and effectively, acquittals) to give a brief insight into the legal proceedings. However, in the user testing phase of our project, users found the visualization for this ineffective, so we removed it from our main project (but documented it in our Process section.

Related Work

One of us used this data for our Open Data project at the ISchool and had done some preliminary research using IPython notebook and Python libraries like PANDAS, NumPy, SciPy, Mapplotlib, etc. A link to the notebook is available at: http://nbviewer.ipython.org/github/chalenge/crime-against-women/blob/master/_notebooks/WomenCrimeAnalysis.ipynb

We also explored existing visualizations created by various other teams using the same dataset:
- Crime against Women in India - 8 questions using RCharts, GoogleVis
- Crime against Women - IndiData
- Crime against Women - Tableau
- Crime against Women - Tableau - Crime Statistics across India

Additionally, the following few visualizations were exclusively based on mapping to show unsafe areas in different cities:
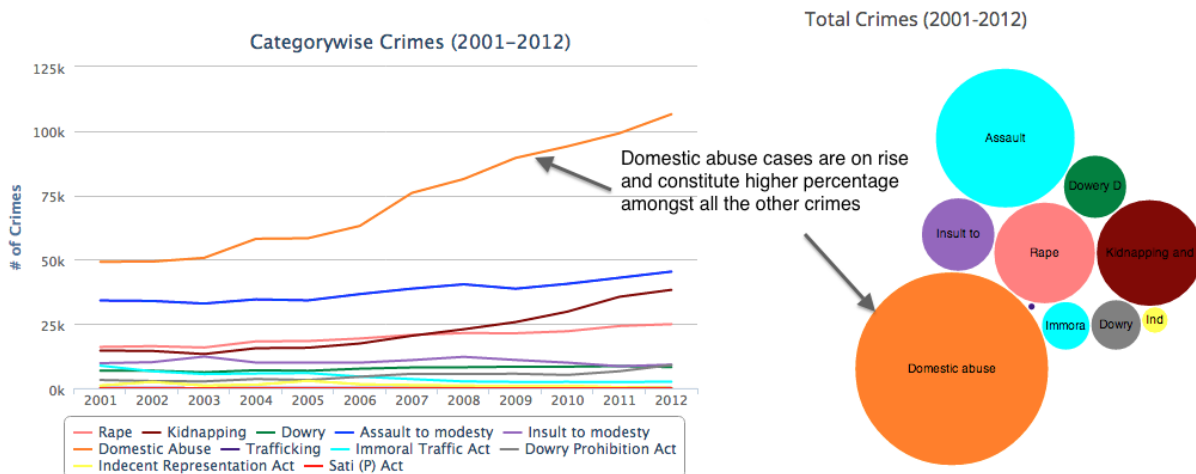- Akshara crowdmap
- Safecity project

-

Description

With this visualization, we seek to understand the statistics for crimes against women in India.

Some of our specific findings:

- The most reported and convicted crimes were under the domestic abuse category (Cruelty by husband or his relatives). It also appears that the same category has had the most dramatic rise in number of reported crimes over the years. This could be because either 1) Number of cases of domestic abuse has actually risen dramatically 2) Women are beginning to speak up more about domestic abuse than before. Since the institution of family is strong in India, the compassion factor makes this a particularly complex category of crime. The concept of 'rape by husband' is very new in India and probably many women, especially in rural areas, are not even aware that it is a violation. Apparently, this is changing.
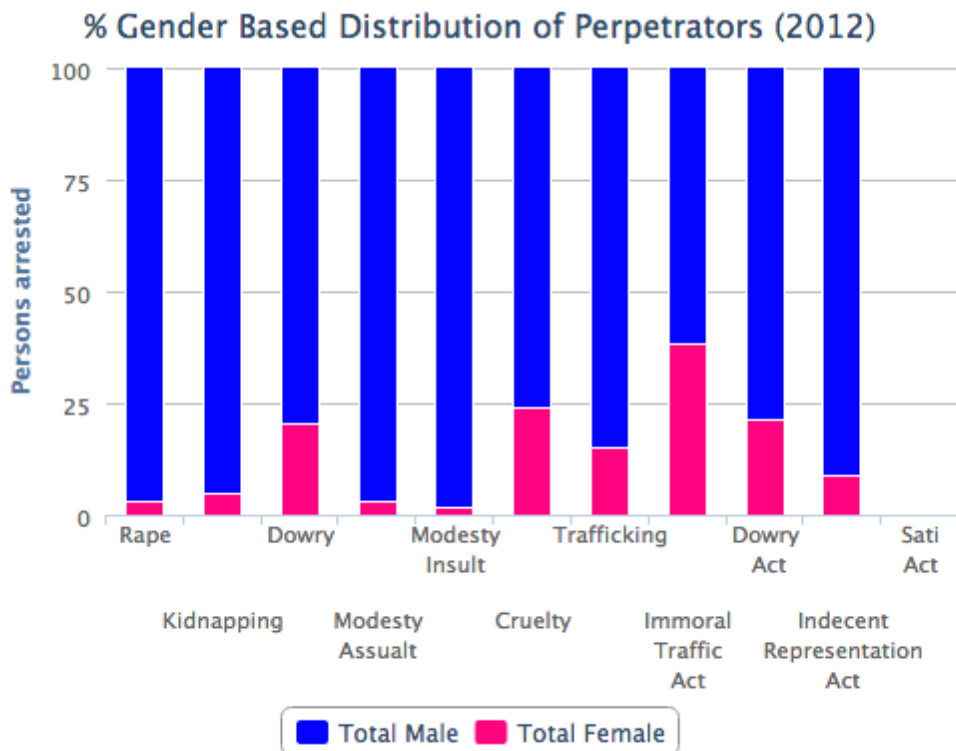

- Since the popular Indian media is rife with news of rapes, it appears to be the most rampant and high profile of all the crimes. No other crime has been given as much attention in the media, with a significant increase in the last few years owing to cases like the Delhi Gang Rape, Scarlett Keeling Case in Goa, etc. which made international headlines. The hype in the media compels us to believe that rapes are on a dramatic rise. However, the data reveals that rapes have more or less followed a flat trend.
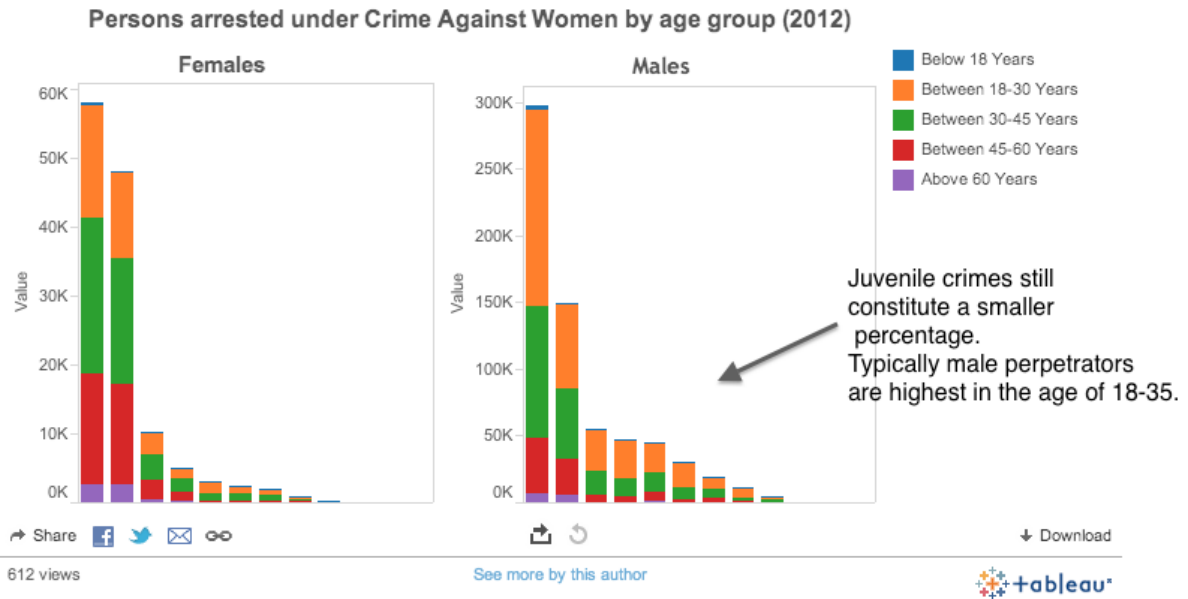
- This raises a lot of questions - if the number of rapes was always the same (or this high), why all the sudden media attention? Where was the media in earlier years? Could the data hiding the truth for the last few years?. There could be any one or more of multiple possible explanations.
- One explanation to this trend is probably that Indian society puts a lot of social stigma on rape, with ignorant politicians passing statements questioning the integrity of the female victims. These kinds of things tend to sensationalize the crime. Also, if foreign victims are involved, it draws even more attention. This situation seems to have improved slightly over last couple of years.

We also tried to complete the perpetrator's profile using age and gender data and tried to answer the following questions:

- What is the ratio of male to female offenders for different crimes?
  This ratio seems smallest for Immoral trafficking: the flesh trade is known to be dominated by women, with females being both lead criminals as well as accomplices to the same. The highest ratio is for insult to the modesty of women (molestation), which would not typically involve many women.
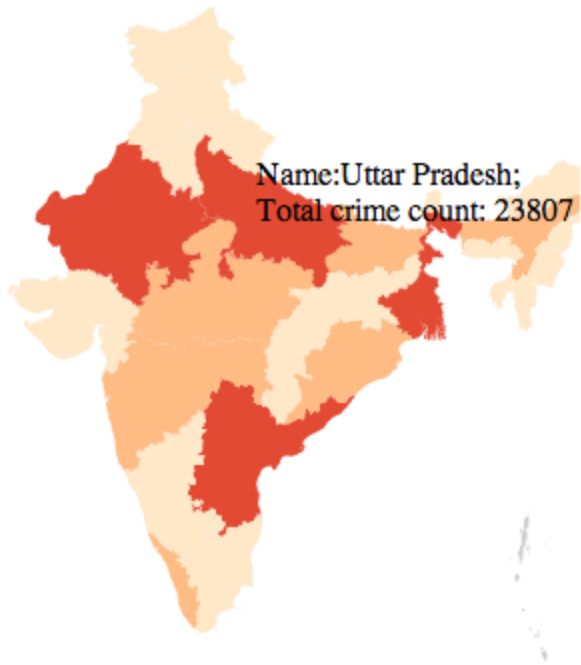
- Based on the Delhi gang rape verdict for the juvenile offender, we wanted to evaluate the ratio of juvenile perpetrators. It seems like males below 18 years are the least likely to be involved in crimes against women. The typical age group of male offenders seems to be 18-30 years which is when they are known to be most aggressive, while for women it is 30-45 years which is the typical age bracket of relatives and in laws committing crimes against women, women ringleaders of flesh trade, etc.
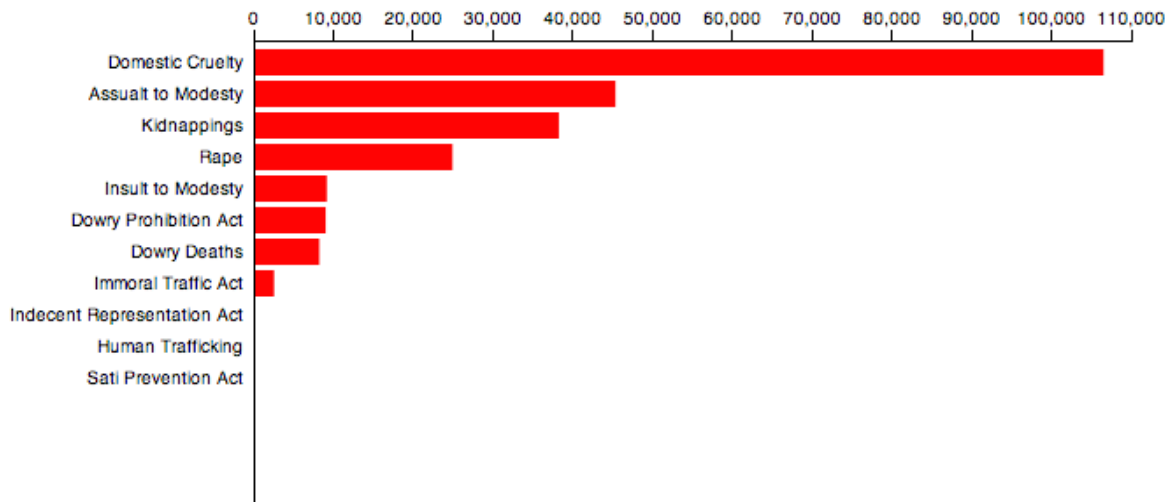
Persons arrested under Crime Against Women by age group (2012)

We also tried to map the crime hotspots in India.
- A chloropleth map of India allowed us to instantly visualize the states having the highest cases registered under Crime against Women. West Bengal, Andhra Pradesh, Uttar Pradesh and Rajasthan (in that order) seem to be the crime 'hotspots' going by cases registered. The far north, north east and west of the country seem relatively safer as compared to the center and north. One hypothesis for the capital New Delhi being termed as one of the most unsafe cities in India is probably because of the exodus of immigrant population from neighboring states like Uttar Pradesh and Rajasthan.

Name:Uttar Pradesh;
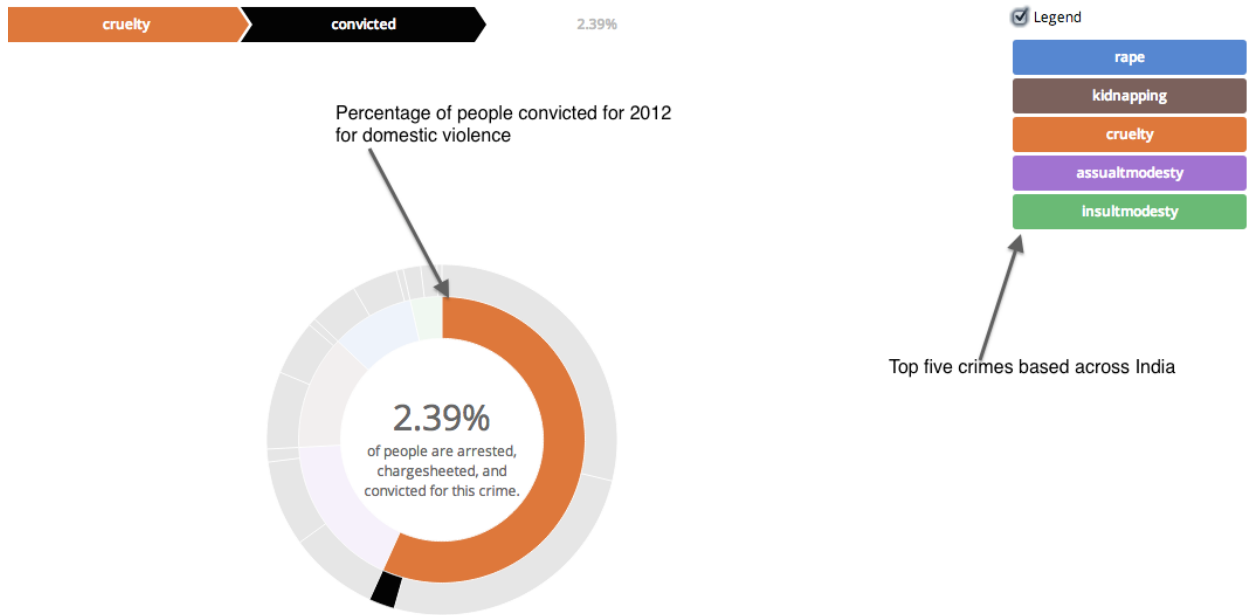Total crime count: 23807

- We also used a hierarchical bar chart to understand the breakdown of crime statistics by state and the long tail of distribution of these crimes.

A bar chart showing values by state, with a horizontal axis scaled from 0 to 20,000 (increments of 2,000).

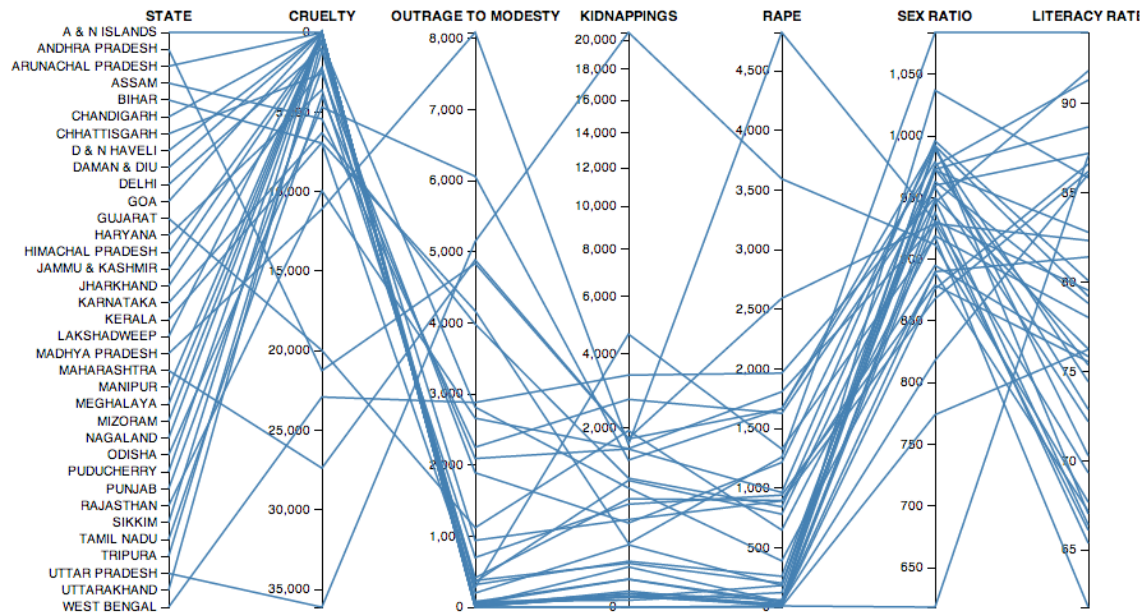| State	Approximate Value
WEST BENGAL	~19,500
ANDHRA PRADESH	~13,300
RAJASTHAN	~13,200
UTTAR PRADESH	~7,600
MAHARASHTRA	~7,400
GUJARAT	~6,600
ASSAM	~6,300
KERALA	~5,200
MADHYA PRADESH	~4,000
BIHAR	~3,700
KARNATAKA	~3,700
HARYANA	~3,000
ODISHA	~2,500
DELHI	~1,900
TAMIL NADU	~1,900
PUNJAB	~1,300
JHARKHAND	~1,300
CHHATTISGARH	~900
TRIPURA	~800
UTTARAKHAND	~300
JAMMU & KASHMIR	~300
HIMACHAL PRADESH	~200
CHANDIGARH	~0

- Based on the top categories of crimes generated from the above bar chart, we also wanted to understand India's judicial process of arresting, charge-sheeting, convicting the perpetrators. We also observe that the number of total perpetrators keeps decreasing through the stages from from arrest through chargesheeted to conviction.
- A small percentage of people arrested for various different categories of crimes actually end up getting convicted. This reveals a gaping hole in the judicial system of India, which can be attributed to either loopholes in the legal systems or lack of talented lawyers. Alternatively, it could be the case that acquittals happen due to insufficient evidence, victims turning hostile (institution of family is strong in India, women are very likely to develop compassion or fear towards their perpetrators if they belong to the same family and consequently withdraw their statements.) or maybe even false positive cases (though this seems contrary to all existing evidence).

cruelty → convicted 2.39%

Percentage of people convicted for 2012 for domestic violence

2.39%
of people are arrested, chargesheeted, and convicted for this crime.

Legend

rape
kidnapping
cruelty
assualtmodesty
insultmodesty

Top five crimes based across India

Based on the media reports on low literacy rates and dwindling sex ratio's impacts in the states of Haryana, we wanted to examine the impact of these external factors on the crime trends. An interesting question here is whether you would expect higher values of literacy rate (more awareness) and sex ratio (more women in society) to be associated with higher (more reporting of crimes) or lower (reduced number of crimes) number of cases registered.

Data

We used Crime against Women data from the Open Government Data Platform of India (data.gov.in), which had various catalogs, downloadable in common formats like xls, csv, json, xml, ods and also accessible via APIs. The most current data available is for the year 2012. We also used data from Wikipedia for Literacy Rate and Sex Ratio of Indian states for the same year. Most of our data was in fairly usable form, except for some minor wrangling that we had to do to suit the formats our visualizations required.

Tools used

We extensively used Tableau for our exploratory data analysis and some of the final visualizations. We used Highcharts for some of the easier visualizations where we just had to plug in the data and D3.js for more complex visualizations. We also used Python and Javascript to wrangle some of our data and a parallax scrolling responsive framework for our UI.

Process

We were originally Team TubesTrends, where we planned to analyze trends across various social media platforms. All our data was scraped off the web and in intermittent samples, which posed many complications. Also, the data just did not have enough dimensions as to afford rich analyses. Processing and wrangling the data was also very time consuming and not proportionately rewarding. After talking to the instructor - Prof. Marti Hearst - we decided to shelve our original idea.

We then decided to work with this concept and dataset one of us was using for the Open Data class, namely, Crime against Women in India. We were constrained for time so prioritization of tasks and smart division of work were on top of our list, which had to be balanced with collaboration to the degree that a project of this nature required. We decided to modularize the work by the catalogs of data and individual comfort levels with different tools. Our process subsequently spanned over the following phases:

- Spending time with the data, getting to know it.
- Brainstorming about potentially interesting insights we could get from it, EDA and bringing out our narrative
- Assignment of work modules
- Brainstorming over design choices, consensus
- Developing the visualizations
- Integration, hosting, and usability assessment, incorporating the feedback from peer reviews
- Usability Tests
- Final Report

We created some redundancy with our designs, so as to have enough choices to get feedback from users about the type of visualizations that work for the users. We kept all our work for the Final Presentation, and based on people's feedback, and on results of usability tests, and removed some visualizations out from our final deliverable. One of these was our sunburst diagram to show the legal process of crimes - from registration through laying the chargesheet and finally conviction. This particular visualization did not seem to work for most users, and had to be demonstrated by us explicitly. Even after that, they did not find it very useful, and it also seemed to violate some visualization principles. We also left out a bar chart showing number of men and women perpetrators arrested for various crimes against women, because it seemed redundant in the light of another chart that served almost the same purpose and communicated all the same information.

Apart from these two, we had success with all the other visualizations, and they seemed to be clear enough to users.

Results

Usability tests

For the usability tests of visualizations, we recruited four users from the following demographic profiles:
- Age: 25 - 50
- Nationality: Indian, American
- Sex: Males and Females

We asked the users to evaluate the visualizations based on the following tasks:
- Asked the users to complete usability assessment of the visualizations and think out aloud while evaluating the visualization.
- Asked them if they understand the visualization and what purpose they think it fulfils and if they think it fulfils the purpose well.
- Also asked the users if they think one visualization works better than the other in complementary visualizations
- We also asked the users to provide feedback on what they could have done differently

The findings for the usability study are as listed below:

User 1:

Demographics: 30-35 years of age, Male, Indian origin, Product development at an IT company, High technical competency
- The user was taken under the assumption that cases for Rape and Outrage to modesty would have a higher rate of cases registered based on recent media reports. However, he was surprised to see that domestic violence was on rise since last decade.
- He was also interested in seeing the profile of perpetrator broken down based on age and gender roles. He was especially interested in the ratio of male to female perpetrators across all the crimes.
- He also liked the crime hotspot provided in one map and was able to quickly relate the visualization of number of cases registered for 2012 with the crime hotspots in 2012. He also exclaimed the fact that because many cases of domestic crimes are reported in West Bengal could be probably because the sex ratio of male to female is higher and also probably because females

in West Bengal are economically more stable than their other counterparts in other states.
- The user was shocked to see the sunburst visualization on the number of persons arrested, chargesheeted, and convicted for top four crimes. However, the user did experience small amount of confusion in interpreting the sunburst visualization citing issues like understanding the total percentage that is shown in the message.
- The user seemed to understand the parallel coordinate visualization to correlate external factors like sex ratio and literacy rates to the top crimes against women. He was however initially overwhelmed with the number of factors that impact the crime statistics.

User 2:
Demographics: 50-55 years of age, Male, Indian origin, Small business owner in India familiar with the issue of violence against women, Medium technical competency
- The user was taken under the assumption that cases for Rape and Outrage to modesty would have a higher rate of cases registered based on recent media reports. However, he was surprised to see that domestic violence was on rise since last decade.
- The user liked the crime hotspots provided in one map but was not able to relate the map with the hierarchical bar chart. At one point, when the user clicked on the category of bar chart, he felt confusion about the value he had selected.
- The user liked the sunburst visualization because it affirmed his analysis about violence against women for the year 2012. However, he stated that the visualization would have been more powerful to interpret as a simple bar chart.
- The user seemed interested in exploring the parallel coordinates visualization to understand the impact of external factors on crime statistics. He was surprised to see how increasing literacy rate shows decreasing trend for cruelty and other crimes. He also cited the government of India's "Sarva Siksha Abhiyan's" contribution to uplift the number of females taking primary education in India.

User 3:
Demographics: 35 year-old, Female, American, familiar with the issue of sexualized violence in a US context, but unfamiliar with sexualized violence in India

- The user was surprised to see that domestic violence was so much higher than other crimes against women.
- For a few types of crimes, such as Sati and Dowry Death, the user was unsure of what those crimes entailed and suggested that there be some kind of legend on the visualization that explained what the crimes involved for American audiences.
- The user spent over 5 minutes attempting to figure out the meaning/purpose of the sunburst diagram. In the end, the user could not figure out what the chart was trying to show, and gave up and moved on to the next visualization in frustration.
- The user commented how colorful and inviting the website was; specifically, the user said the colors made them feel hopeful, even though the topic was extremely depressing.

User 4:
Demographics: Early 30s male, American, student.
- User noted that he liked the bubble chart; he felt it gave an overall sense of the magnitude of the different types of crime.
- The user commented that he was surprised to see that women committed acts of violence against other women. Nonetheless, he noted that women perpetrators tended to be older women engaging domestic abuse; this result made sense to him, as it seemed likely that a mother in-law might be horrible to their daughter in law, younger female family member, etc.
- When looking at the map of the India, the user commented that they would have like seen city names/points to give more context to the map.

Demo
- Working link to the visualization:
  http://people.ischool.berkeley.edu/~ruchitarathi/infoviz-final-caw/

Source Code
- Github repository: https://github.com/j9recurses/infovizfinal

Contributions

Task	Steps	Team Member
Background research	1. Evaluate existing visualizations 2. Needs assessment (what is involved, affordances of the interface) 3. Summarization of findings	Ankita
Exploratory data analysis	Exploratory data analysis using Tableau	Ankita
Getting to know our data	1. Exploratory data analysis 2. Storyboarding	Ankita, Ruchita, Janine
Prototyping	1. Creating multiple alternative designs 2. Going over design principles covered in class for this step	Ankita, Ruchita, Janine
Usability tests for prototypes	1. Recruit members 2. Schedule slots 3. Carry out tests 4. Document results, feed back into design process, finalize designs	Ankita, Janine, Ruchita
Development using Highcharts	1. Skeletons of highchart visualizations	Ankita
Development/ Implementation using D3.js	1. Skeletons of D3 visualizations 2. Development in D3.js	Ruchita, Janine
UX/UI/Backend	UI and backend development	Janine, Ruchita
UI testing	1. crossbrowser testing 2. usability assessment for the web application	Ankita

Credits

- Parallax responsive UI framework:
  https://github.com/jalxob/cool-kitten/pulls
- Site logo and background PSD files:
  http://www.danajunebaker.com/