Visualising Crimes Against Women in India

Cheng Leng Lee
School of Information Systems
Singapore Management University
cllee.2017@smu.edu.sg

Zi Yuan Lim
School of Information Systems
Singapore Management University
ziyuan.lim.2017@smu.edu.sg

Sean Chai Shong Hee
School of Information Systems
Singapore Management University
sean.chai.2017@smu.edu.sg

Abstract - In this study, we analyse the phenomenon of crimes against women in India. In recent years, the incidence of violence against women has risen even as India experiences rising affluence and development. We conducted time-series and state-level analysis on the incidence of crimes against women, and explored possible relationships between various socioeconomic factors and the different crime types. Our final product is an interactive visualisation tool which users can navigate to derive insights on the issue of crimes against women in India.

Index Terms - Crime, Women, India, Inequality

I. INTRODUCTION

According to the Thomson Reuters Foundation Annual Poll, India is ranked as the world's most dangerous country for women. [1] This is not surprising, as India has had a long-standing history of violence against women, which is deeply rooted in certain cultural practices such as female infanticide and acid attacks. Even with the increasing public outcry regarding such discrimination and the enactment of laws protecting women, the number of crimes committed against women is increasing steadily over the years. Hence, our project aims to look into the trends of crimes against women in India at various geographic levels and analyse the effects of various socioeconomic factors on the rate of crime against women to gain a better understanding of this issue.

II. MOTIVATION AND OBJECTIVES

India is one of the world's fastest growing economies, and is projected to be the third largest economy in the world. Despite its rapid growth and development, women in India still suffer from long-standing gender inequality and are the victims of brutal and inhumane crimes even in 2019. [2] Hence, there is a need to analyse various socioeconomic factors to garner insights on the root causes for crimes against women to understand why this phenomenon is so.

Our objectives of this project are as follows:

1. Provide an overview of the issue of crimes against women in India
2. Draw comparisons to study the differences in crime rates between different states
3. Study the effect of various socioeconomic factors on the number of crimes committed against women
4. Assess the effectiveness of the law enforcement in case handling

We hope to achieve these objectives through the development of interactive visualisation dashboards, which can help us to understand the increasing trend of crimes against women, and what factors may contribute to such crime rates.

III. RELATED WORKS

We researched on existing data visualizations for crimes against women in India. Majority of such visualisations lacked depth and were simplistic in nature. However, we identified a few visualizations that were insightful and aesthetically pleasing.
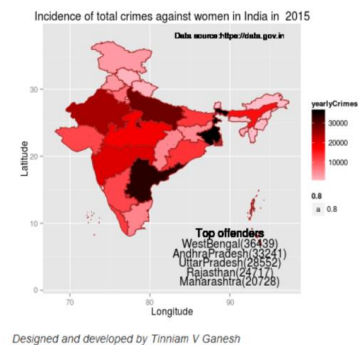

Fig. 1. Incidence of crimes by state, India

Figure 1 allows for visualisation and comparison for the total crimes committed at state-level. The usage of a choropleth map allows us to compare the magnitude of crimes against women in different states. However, it does not account for the difference in population size within each state, and merely takes the absolute number of crimes in each state as analysis.
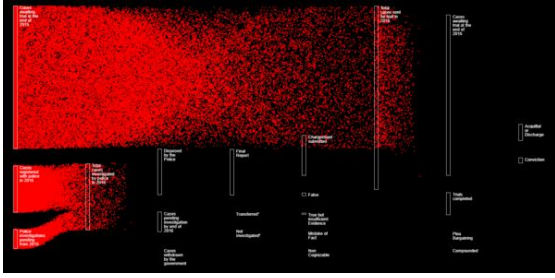
Fig. 2. Rape Trial process, India

Figure 2 depicts a modified Sankey plot that shows the efficacy of the justice system in India in handling rape cases. Each dot represents a single rape case in India, and the dots will travel to show the final outcome of the case - whether it ends in conviction or acquittal, or is dropped in the middle of the process.

These visualisations tend to be more descriptive in nature, providing insight as to what happened instead of identifying patterns that allow us to understand why the crime rate is as such. Our project research therefore aims to provide various forms of analysis to identify factors leading to crime against women.

IV. APPROACH

For our visualization approach, we adopted a 5-phase process. This consists of data definition, aggregation, preparation, analysis and finally interpretation.

Data definition is the process of identifying and understanding what sort of data we would require. This is done via research on the subject matter. We performed data aggregation by collecting relevant data from credible and reliable sites. Afterwhich, we sift through and filter out data we deem unnecessary for our analysis. We then perform data analysis using suitable visualisation methods before interpreting the results. Section 5 will elaborate on the stage of collating, exploring and preparing data while Section 6 will explain the design considerations of planning and creating the visualisation. Finally, Section 7 explores the final visualisation dashboards developed for our tool, elaborating on the purpose of the visualisation, as well as the design rationale behind it.

V. DATA COLLECTION, EXPLORATION & PREPARATION

Embarking on our project, we first sourced for suitable datasets that could allow us to study the problem at a granular level. From the Indian Open Government Data Platform, we were able to obtain district-level offences of violence against women. Besides that, we procured the 2011 India Census Data from Kaggle, which consists of a detailed reporting of various demographic and socioeconomic data of the country. Synthesising these two datasets would allow us to gain a more comprehensive understanding of the issue at hand, while analysing its causes and effects.

We first used Tableau to create preliminary visualisations to explore the data in greater detail. This helped us to discover existing trends and envision possible plots to visualise and explain the existence of such trends.

During which, we transformed our data to a suitable format for our preliminary visualisations to be plotted. This involved combining the district-wise datasets for crimes against women across different years, and reconciling the differences in data format. We also aggregated the district-level data to state-level, as some of the data we obtained was only available at state-level. Hence, we made the decision to aggregate our data to state level in order to overcome this decision, even though we might be compromising the granularity of data in the process.

Furthermore, the district-level data for crimes against women was presented in different data formats on the Open Government Data platform, possibly due to the inclusion of more acts protecting women being introduced in 2014. Hence, we had to standardise the data formats prior to plotting our visualisations.

When generating our visualisations in R, significant data cleaning was required such that our data was in the correct format for the different R packages used in preparing our visualisations. To do so, we utilised R packages such as sf, tidyverse and tidygraph.

VI. DESIGN CONSIDERATIONS

Following data preparation, we sourced for different visualisations that could present our intended analysis as clearly as possible.
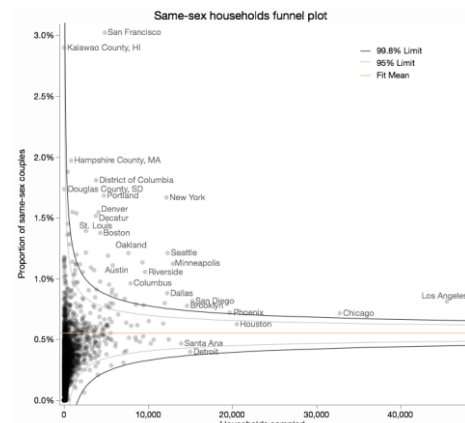
1. Funnel Plot

Fig. 3: Funnel Plot Comparing Same-Sex Couples

We studied the usage of this funnel plot, which shows the number of same-sex households as a proportion of the total number of households sampled in a US state. This was similar to the intention we had in comparing across different Indian states according to their crime rates. This visualisation was able to simultaneously display sample statistics and the

corresponding sample size for multiple cases, which we felt was suitable for our analysis.
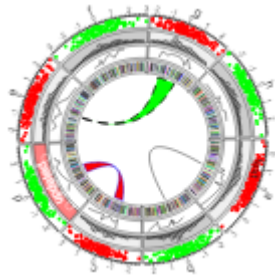
2. Circos Plot

Fig. 4: Circos Plot

The next plot we considered for use in our dashboard design was the circos plot. The circos plot allows for different types of analysis to be shown in one plot across the different tracks, hence we considered using this plot as an overview or summary for our visualisation tool.

However, after much consideration, we decided against it as we felt that the plot was confusing to interpret for the analyst, and might not be useful as our analysis was across different datasets, which may be difficult to synthesise into one plot.
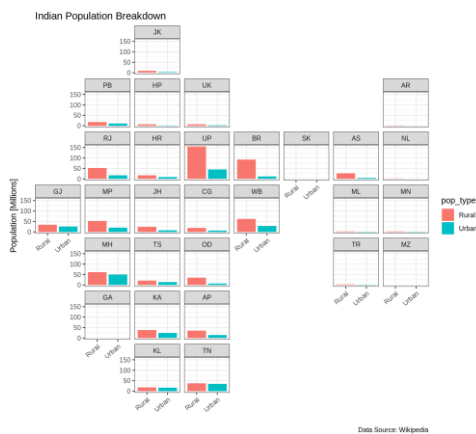
3. Geofacet Plot

Fig. 5: Geofacet Plot for Indian Population

The geofacet plot above is used to study the gender breakdown across different Indian states. We felt that the advantages of using this plot was that it is easy to interpret since the plots are arranged geographically, which can allow for comparison at a glance. This plot however, can prove cumbersome should the values of one of the graphs be significantly higher than the other graphs. This will result in graphs with poor scaling and difficulty in analyzing the results proper.

VII. VISUALISATION WALKTHROUGH

A. Homeage

The user is greeted by the overview page, which states the objectives of our visualisation tool and provides background information to the issue.
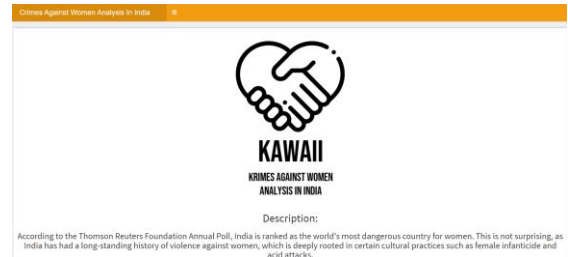

Fig. 6: Homepage of Visualisation Tool

B. Time-Series Analysis of Crimes Against Women
The main objective of this page is to allow analysts to explore the yearly changes in crime rates, as well as the distribution of crime rates across the different crime types. This visualization presents users with a high level overview of Crime breakdown in India.

1. Line Graph

Fig. 7: Yearly view of crime rates from 2011-2015

The yearly trend of the number of cases of crimes against women is displayed in a line graph, which can be filtered by crime type. This can be easily done by clicking on the labels below the line graph, to toggle the crime types which are reflected on the graph. The scale of the line graph changes accordingly as well.

2. Bar Graph Interfacing with Treemap

Fig. 8: Crime Type Breakdown per Year

The page also shows a bar graph which interacts with a treemap, showing the crime type breakdown for each year on hover. This shows the part-whole relationship for each year of analysis. This allows for easy identification of crime types which constitute the largest portion of the total crimes.

3. Design Rationale

The rationale for using both line graph and bar graph is because the line graph clearly shows the trend for each crime type over the years, while the bar graph shows the individual components of the crime rate for each year when it interfaces with the treemap.

C. State-Level Comparison of Crimes Against Women

The main objective of this dashboard is to provide a comparison of crime rates across different states in India. To do so, a choropleth map is presented beside a funnel plot.
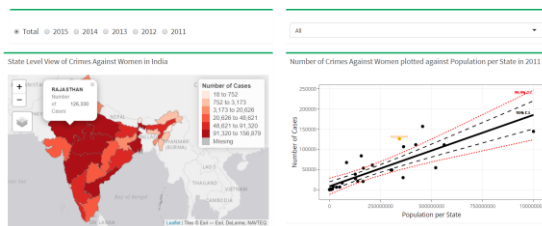
1. Choropleth Map

Fig. 9 : State-Level View of Crime Rates

The choropleth map shows each Indian state, shaded in proportion to the absolute number of crimes against women reported in that state. When the user clicks on a particular state, the relative position of that state is highlighted on the funnel plot.

2. Funnel Plot

Fig. 10: Funnel Plot showing Statistics of State

The funnel plot serves as a means to identify the states which have high incidence of crimes against women, while considering the population size within each state. The addition of statistical measures, such as the 95% and 99.9% confidence interval of the number of crimes for the differing population sizes allows the user to quickly identify any outliers which have higher crime occurrences than other states of similar population size.

Hovering over a particular point on the funnel plot will display the state, as well as the corresponding population size and number of crime cases in the tooltip. This provides the analyst with more information without compromising the simplicity of the plot.

3. Interactive Elements

The above visualisations can be filtered by year and crime type according to the filters placed above the plots, allowing for a more detailed analysis of the individual crime rates. This would be useful for analysts who wish to study a particular year or crime type of interest.
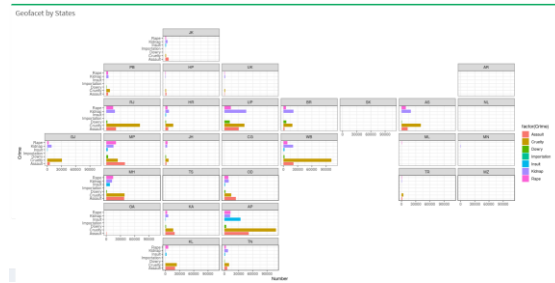
4. Geofacet Plot

Fig. 11: Geofacet Plot for Crimes in India's States

The second tab of this dashboard features a geofacet plot, which shows the crime type breakdown per state in India at a glance. This allows for easy comparison and identification of states where a certain type of crime is more prevalent.

D. Analysis of Socioeconomic Factors on the Incidence of Crimes Against Women

This page serves to analyse the possible relationships between various socioeconomic factors and the crime rate in a particular state using a parallel coordinate map, with the potential to analyse at district-level using the correlation heatmap and scatterplot.

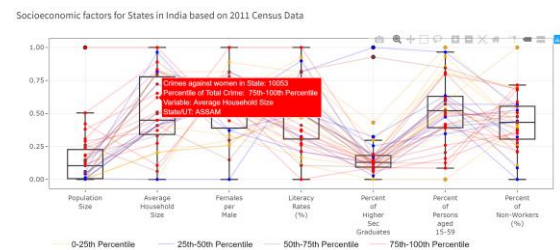*1. Parallel Coordinate Map with Boxplot*

Fig. 12: Parallel Coordinate Map for 7 socioeconomic factors against the crime rates of different states

Analysts will be able to see data points for the socioeconomic factors pertaining to the various states in India. They will be able to filter the plot by state to isolate data points for the selected state. Users will also be able to filter the data by the percentile of total crime. Hovering over each point displays the tooltip.
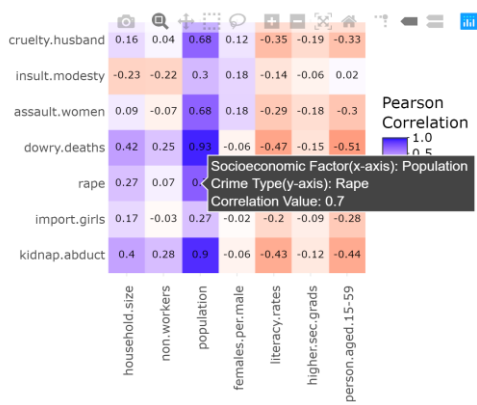
2. Correlation Heatmap

Fig. 13: Correlation Heatmap for Crime Types against Socioeconomic Factors

The correlation heatmap shows how the socioeconomic factors correlate with the different crime types. Hovering over each point will displays a tooltip. Clicking on each point will populate the adjacent scatter plot with the selected variables. Correlation heatmap will display information pertaining to the state selected.
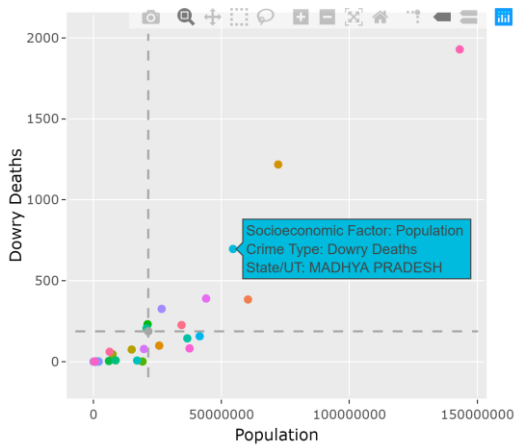

Fig. 14: Scatter plot showing statistics for the selected state

The scatter plot displays data points for the variables selected in the correlation heatmap. Hovering on the data points displays the tooltip. Each data point represents an individual state (when the filter is set to select all) or an individual district (when filter is set to a distinct state).

3. Interactivity Elements

Added interactivities between the plots allowed for easy filtering of data, which aided our visual exploration of these correlations. Users will be able to draw in depth analysis between socioeconomic factors within the states and their respective crime types through effective filtering.

E. Analysis of Police Crime Disposal for Reported Cases

1. Sankey Diagram

A sankey diagram is used to show the flow of reported cases of crime against women in a police investigation. This allows the user to evaluate the effectiveness of the law enforcement in handling such cases. We decided to develop this plot as we believe this is an important factor which can serve as a deterrent to further incidence of crimes if the justice system is efficient and effective.
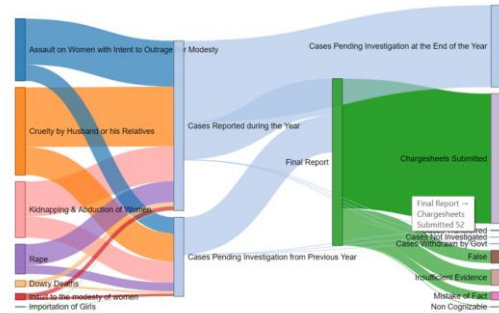

Fig. 15: Sankey Diagram showing Percentage of Crimes at Each Stage

The user can mouseover the flows to view the percentage of crimes which reach a particular stage in the investigation process. This complements the visual indication of the number of cases at each stage, according to the size of the flow.

VIII. KEY FINDINGS & INSIGHTS

1. We generally see increasing crime rates from 2011 to 2015 (224,335 to 306,078) with the largest increase in Assault and Kidnapping, showing an increment by almost two fold. One would expect that as countries become more developed, crime rates tend to fall. However, the inverse holds true for India. We feel this could be attributed to two possibilities. Firstly, whilst India is becoming more developed, it is still largely a patriarchal society thus women still suffer from violence.
Secondly, the development of India could postulate for greater women's rights and thus women are more likely to report crimes as compared to the past. This would suggest that crimes against women in previous years were actually underreported due to women having lesser rights and being fearful to speak out.

2. From the state-level analysis, one can observe that the states of West Bengal, Rajasthan, Delhi, Assam, Madhya Pradesh and Andhra Pradesh experience higher incidences of crimes against women than the 99.9% confidence interval, as seen from the funnel plot. Some of these states are of the least developed states in India based on the "MDI" [3], which can explain why their crime rates are abnormally high. However, that being said, richer states like Delhi also experience higher than average crime rates, which postulates that economic development is unlikely to be the only driver of crime against women.

3. We observed that there is a high correlation between population size and most of the different crime types. Namely, there was an almost perfect direct linear relationship between population size and dowry deaths (0.93), as well as kidnapping and abduction (0.9) across the whole of India in 2011. Other crime types also showed strong correlations with population size (>0.5). This result was aligned with our expectations, as we believe higher population sizes would generally result in higher crimes against women.

5. There were two interesting observations we made based on the correlation heatmap. Firstly, females per male ratio did not play a significant factor in the happenings of the different crimes. Secondly, we observed that there was a negative correlation for persons aged between 15-59 with the different crime types. This meant that crimes against women rose in states that had a population that was older than 60 years of age.

6. Taking a look at West Bengal through our parallel coordinate plot, we see that West Bengal is near the mean for multiple socioeconomic factors, with only its population being above the 3rd quartile mark. Yet West Bengal remains as the state with the highest crimes against women. In such instances, we can conclude that there are other inherent factors that are affecting the crime rates in the state.

7. Only around 50% of reported cases involve the submitting of chargesheets, which means the accused is charged and is submitted to the court for trial. The other half are either transferred, dismissed or have not been investigated even at the end of the year. This shows how the law enforcement may not be very efficient in India. The speed of which crimes are prosecuted could potentially act as an enabler for even more crimes. However, we do not have domain understanding on the other crimes that are prosecuted yearly by the Indian Police, as such it is difficult for us to ascertain if the Police are acting in an efficient manner.

VIIII. CONCLUSION AND FUTURE WORK

The increasing prevalence of violence against women in India is a worrying phenomenon. Given the increasing awareness of this issue on a global scale, this research aims to provide relevant information and identify possible causes for this issue. This could potentially complement government efforts in curbing this phenomenon.

In the future, we could further extend our research by conducting all our analysis by district-level, which could provide more detailed and relevant insights for the authorities of each state due to the greater data granularity. We could also study other socioeconomic factors such as the caste system and poverty level, which could allow us to derive more meaningful and useful relationships between these factors and the crime rates in India.

REFERENCES

[1]https://poll2018.trust.org/

[2]https://www.theguardian.com/world/2019/aug/01/india-ruling-party-bjp-expels-politician-rape-kuldeep-singh-sengar

[3]https://timesofindia.indiatimes.com/india/Odisha-Bihar-least-developed-states-Goa-and-Kerala-top-chart-Rajan-panel-report/articleshow/23094131.cms