Visualisation Project

31187366
PRASHANT HARIPRASAD JAJORIA

CRIMES IN INDIA

Tutor: Farah Kabir 04-P1

Table of Contents

Introduction

'Data is the new oil', you probably would have heard this quote if you work in the field of Data Science. This Quote was originally coined by Mathematician Clive Humby, in 2006, but was in spotlight in the recent years when The Economist published a report in 2017, titled 'The world's most valuable resource is no longer oil, but data.' This is so true in today's era of Technology where many Computer applications are driven by data. Data plays a fundamental role in the field of Artificial Intelligence and Machine Learning. Companies across the world are now investing more capital and resource into deriving value from the organizational data. Data exploration comes in handy wherein one can gets to know more about the data by doing cleaning the data, treating missing values, detecting outliers and performing Statistical tests. It is important that the insights and results obtained from the Data Exploration process is communicated in a form which the target audience can understand. This where the field of Data Visualization comes to the rescue. So, what is Data Visualization and why is it so important. Data Visualization is visually representing your information using static or interactive graphs, charts and maps that helps to see patterns and trends which otherwise would have been difficult to see.

Problem Definition

In India it is not uncommon these days to find news of a Theft or Riots on TV, social media, or newspaper. Everyday a different type of crime is reported in the news. These news cases reported are from different state and district of India.

As the news of criminal activities have become a prominent feature in the daily, it demands a deeper investigation. Due to advancement in technology, it is possible today to have data of crimes for past years.

In this report I have considered data of crimes committed in different districts of India from year 2001 to 2011. It is interesting to know what are the types of crimes that take place in every state of a country and which of the state is the most dangerous. Does higher Literacy rate ensure fewer criminal activities? Is poverty one of the factor behind crimes in India?

Intended Audience

- Government officials

- Politicians

- Students / Researchers

- News reporters

- Non-Governmental Organizations

Results

• • •

HIGHEST NO OF THEFT CASES WERE REPORTED IN INDIA FROM 2001 TO 2012

TOTAL OF 3732k CASES OF THEFT REPORTED FROM 2001 TO 2012

THEFT AND HURT/GREVIOUS HURT ARE THE TOP 2 CRIMES IN INDIA

CASES OF CUSTODIAL RAPE ARE THE LEAST IN THE GIVEN YEARS WITH ONLY 26 CASES REPORTED FROM 2001 TO 2012

DELHI, THE CAPITAL OF INDIA HAS THE HIGHEST CRIME RATE OF 7543 CASES / 100,000 PEOPLE

STATE OF NAGALAND HAS THE LOWEST CRIME RATE WITH VALUE OF 663 CASES / 100,000 PEOPLE

THE DISTRICT OF UTTAR KANNADA IN THE STATE OF KARNATAKA HAS THE HIGHEST REPORTED NO OF CASES WITH TOTAL OF 350,347

THE DISTRICT OF KIPHIRE IN THE STATE OF NAGALAND HAS THE LEAST REPORTED CASE IN INDIAWITH ONLY 195 CASES REPORTED FROM 2001 TO 2012

Questions

1. What are the most common crimes committed by people in India?

2. Which State recorded the highest Crime rate? And which district had the most Criminal cases within the state?

3. How does Literacy rate affect the Crimes in India?

4. Is there a relation between Poverty and Crimes?

Additional Questions

1. Which is the state with lowest Crime rate?

2. Which is the state with least Crime cases reported?

Five design sheet methodology is used to systematically analyze all the design layouts. It has been used by students, companies and website designers to have a final layout that encompasses the best aspect of each design sheet developed.

Sheet 1

The process of Five design sheet starts with Brainstorming, a way to engage ideas and thoughts to achieve the final goal of the project. In this sheet, all the ideas regarding the graph/plots to be used to answer a question are discussed. Each question can be answered using various plots/charts. All the ideas discussed in Brainstorming part are then filtered based on time required to complete the project, implementation hurdles and removes any redundant ideas that are repeated. The ideas after going through filtering process, are categorized based on the type of question it answers. All the categorized ideas are then used in the Design sheet 2,3 and 4 to show detailed implementation of the layout.

Using the Five design approach, I started the design process by first brainstorming the ways in which I can show the data regarding Crimes in India. For this I shortlisted a list of charts which I can use for visualizing the data, namely Bubble chart, Word cloud, Line graph, Dot Matrix plot, Bubble map and Radial Bar graph. After brainstorming, I filtered the ideas and removed graphs which are redundant to answer the questions proposed in the project. Next, I categorized the ideas based on questions the graph answers. Refined the ideas to have a final list of plot that are going to be implemented in the visualization project.

Sheet 2

In design sheet 2, the first proposed layout of the selected graphs that answers all the questions are discussed. In this design, the graphs selected are Wordcloud, Bar graph with dropdown, Line graph and Bubble chart.

Wordcloud will answer the question of most crimes committed in India. Size is the main factor in this plot. Size here signifies the number of crime cases. Also, for better understanding of the data the top two committed crime in India is show in red. Other crime names have a color of black.

Line graph along with drop down list to select the crime type is used to show the data based on crime type. User can select the Crime type from dropdown and the Line graph changes accordingly, showing the data for each year from 2001 to 2012.

Bubble chart is used to show the relation between Crime cases, poverty, and literacy. Size of the Bubble signifies the number of cases for a state. State with the greatest number of crime is shown in red.

Tooltip is implemented for all the charts showing detail like number of cases, crime type, state and year.

Sheet 3

In this sheet, the second proposed layout of the graphs is discussed. In this design, the graphs selected are Wordcloud, Dot matrix plot, Donut chart and Bubble chart.

Wordcloud will answer the question of most crimes committed in India. Size is the main factor in this plot. Size here signifies the number of crime cases. Also, for better understanding of the data the top two committed crime in India is show in red. Other crime names have a color of black.

Dot matrix plot the distribution of cases for each state. Each state is represented using a different color. The number of dots is selected based on the cases in a state.

Donut chart shows the number of cases for each crime type.

Bubble chart is used to show the relation between Crime cases, poverty and literacy. Size of the Bubble signifies the number of caes for a state. State with the greatest number of crime is shown in red. Tooltip for bubble chart shows the Literacy rate, Poverty rate and number of cases for a state.

Tooltip is implemented for all the charts showing detail like number of cases, crime type, state and year.

Sheet 4
In this sheet, the third and final proposed layout of the graphs is discussed. In this design, the graphs selected are Bubble plot, Bar graph with dropdown, Radial bar graph and Scatter plot.

Bubble chart will answer the question of most crimes committed in India. Size is the main factor in this plot. Size here signifies the number of crime cases. Also, for better understanding of the data the top two committed crime in India is show in red. Other crime types are shown in black. Tooltip shows crime name and the number of cases for that crime type.

Bar graph along with group of buttons, one for each year, is used to show the data based on year. User can select the year from button group and the Bar graph changes accordingly, showing the data for selected year. Also, for each year selected the number of crime cases for different types are updated dynamically in a list adjoining the Bar graph to easily follow the data.

Sheet 5
This is the final sheet, which shows the layout of the visualization project. It incorporates the main features from the three designs proposed in sheet 2,3,4 and a final design is prepared for the project.

Wordcloud, Bar graph with button group, Line graph with dropdown selection, Bubble chart and Scatter plot are the final charts selected for final implementation.

This sheet also discusses the technology to use for implementing the project, cost involved and the schedule for the project.

Justification for selected design
Wordcloud is selected from design 1, as it easily shows the top crime committed in India. User need not hover to get the answer. On the other hand, if a Bubble chart is used, although the size will tell the top crime, but the user will need to make extra effort to hover on the bubble to see the name.

Bar graph along with group of buttons, one for each year, is selected from design 3 because it adds interactivity to the visualization. Also, user gets the data dynamically shown in a list for all crime type, of a year selected. Also, bar graph is easy to follow for majority of the user than a Radial/Circular bar graph.

Line graph along with drop down list to select the crime type is selected because it gives more details of the cases for a particular crime for each year. It helps see the change in numbers for different years.

Bubble chart is used to show the relation between Crime cases, poverty and literacy. Size of the Bubble signifies the number of cases for a state. This chart was selected because it allows me to show three variable in a single chart namely, Crime cases, Poverty rate and Literacy rate.

Finally, a scatter plot is used to see states with high crime rate and cases. This chart was used because it helps to group the states that have close characteristics like cases and crime type. Also, color element makes it easy to identify the group to which the dot belong.

An overall colour theme of Red, black and grey is used in the project as red is color is mostly associated with danger by vast majority of public around the world.

Alternative design
Alternative design would comprise of a Choropleth map to show cases in different districts of India. Because of non-availability of clean GeoJson file for different districts of India and implementation overhead, this option was eliminated. In future this could be implemented at State level.

Adding a Pictogram chart for better visual appeal could be implemented.

Also, a stacked area chart for each state can be good option to compare the cases for a single state.

A Treemap consisting of state at high level and type of cases as child, can be a good way to show the cases for each type.

Implementation

For the project, D3.js (also known as D3, short for Data driven Documents) a lightweight JavaScript library for making dynamic and interactive visualizations for web browsers was used. The project has separate blocks of HTML structure code, CSS code and JavaScript code which makes it easy to scale and troubleshoot issues. The external CSV files need to load the data are stored in the project folder. For each plot a separate SVG is created and data is loaded for that plot from CSV file. For better understanding of code comments are added.

Technology

- D3.js – To design interactive Scalable Vector Graphics for web browsers.

- HTML5 – Markup language used to make web pages.

- CSS – Add style to HTML elements.

- JavaScript – For processing the dataset.

- Bootstrap – To design overall layout of website.

Why D3?

Following are the reason of Choosing D3 over a Shiny app in R

As D3.js is an open source java script library, it can be used with any JS framework like Angular, ReactJS or Ember.js.

D3 is data driven which makes it suitable tool for Data visualizations.

Powerful library which need no plugin or other technology to run.

Anyone familiar with we standards like HTML, CSS and Java script can make visualizations. No new learning or tools required.

No fixed rules for implementing a visualization giving the developers the complete freedom and control of the visualizations.

It is lightweight library and works well with data sets of different forms.

Dataset

1. District wise crimes committed from year 2001 to 2012

Tabular data: ~9K rows x 33 columns. It has district wise data of various crimes committed for year 200112. The columns include name of State/UT, year, type of criminal activity and the total IPC cases in that district for that year. The dataset has numbers for 28 different types of crimes.

URL - https://data.world/rajanand/crime-in-india

2. Census of India 2011

Tabular data: 640 rows x 118 columns.

Provides data of population of various districts, religions, urban and rural households, age group of people, education, household size, no of married couples in a household, house ownership etc.

URL - https://www.kaggle.com/danofer/india-census/data#india-districts-census-2011.csv

3. State wise percentage and number of people below poverty line

Tabular data: 35 rows x 8 columns. Also, JSON format available.

Provide state wise data of population below poverty line in lakhs. It also gives data for no of people below poverty line in rural and urban areas.

URL - https://data.gov.in/resources/state-wise-percentage-andnumber-persons-below-poverty-lineyear-2009-10

User Guide

This is an interactive web-based project which uses a static web server for running. Following are the steps that can help you set up the static web server.

Steps to setup a static web server

For security reason, browsers cannot access local files directly, so a static web server needs to be set up to allow the D3 script to load data. The **easiest way** to do this is to download http://brackets.io/. This is a free text editor with live preview feature which runs a static web server at the back end. However, you can use your favourite text editor and skip this step if you already know how to set up a static web server.

(Optional)If you prefer to launch a static server, you can do this with:

- Install Python (on Windows, Python should be installed by default for most Linux and Mac OS)

- Adding Python to your system path (on Windows)

- Use a terminal (command line) to enter your project directory/folder (with cd command)

- Running the command:
  for python 2:
  python -m SimpleHTTPServer 8000
  for python 3:
  python -m http.server 8000

- Then you can open http://localhost:8000/ in your browser to access your D3 visualisations.

Navigation

The project consists of different sections having plots which help answers the proposed questions. For easy navigation, a navbar with tabs is implemented. Clicking on a tab takes the focus to different section. A scroll down button has been added at the bottom of home section to take the focus to next page. Scrolling effect for smooth transitioning is also added. By clicking on 'Crime in India' tab reload the Wordcloud to generate a different layout.

Figure 1: Wordcloud showing the maximum cases in India

On hovering the name of a crime type the name of Crime and total of cases from 2001 to 2012 is shown.
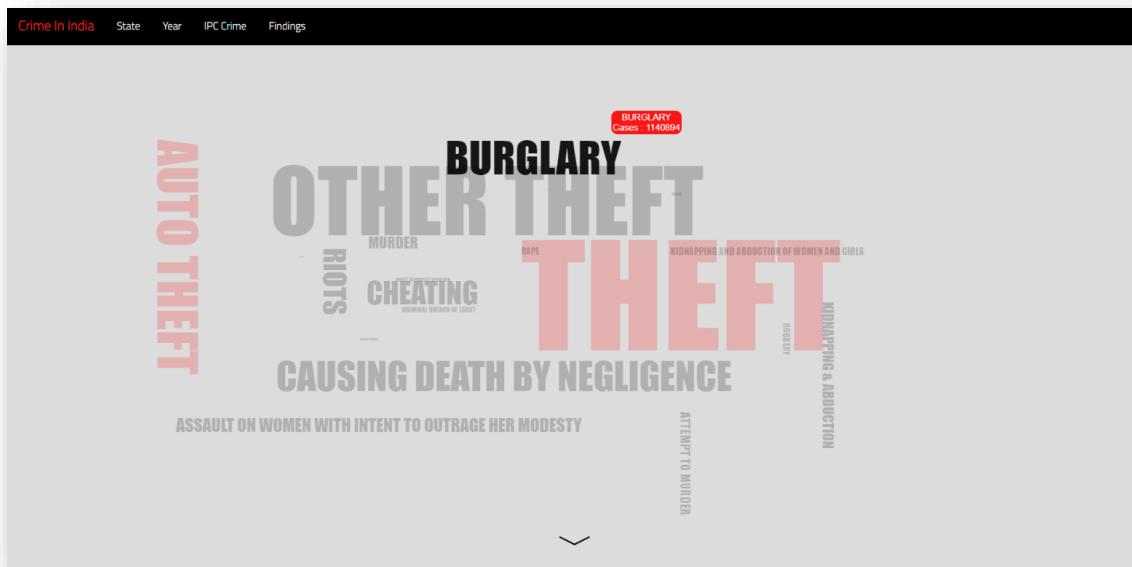
Figure 2: On hover shows total cases from year 2001 to 2012

On clicking the State tab, the focus shifts to chart having selection for state.
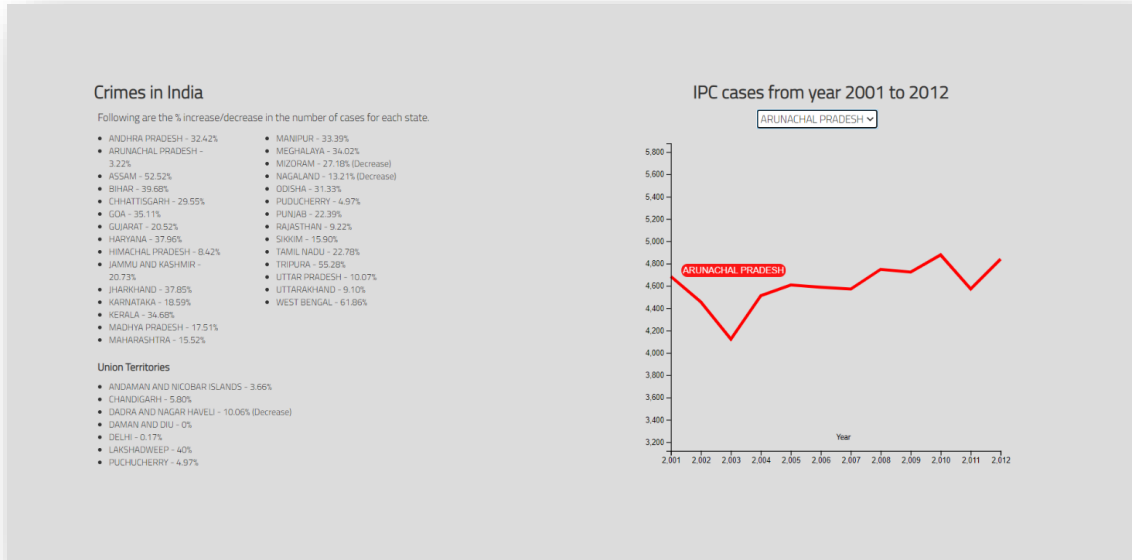
Figure 3: Line chart with dropdown to select State

On clicking the Year tab, the focus shifts to chart having selection for Year. The selected year is shown on the right heading with the breakdown of cases for each crime type.
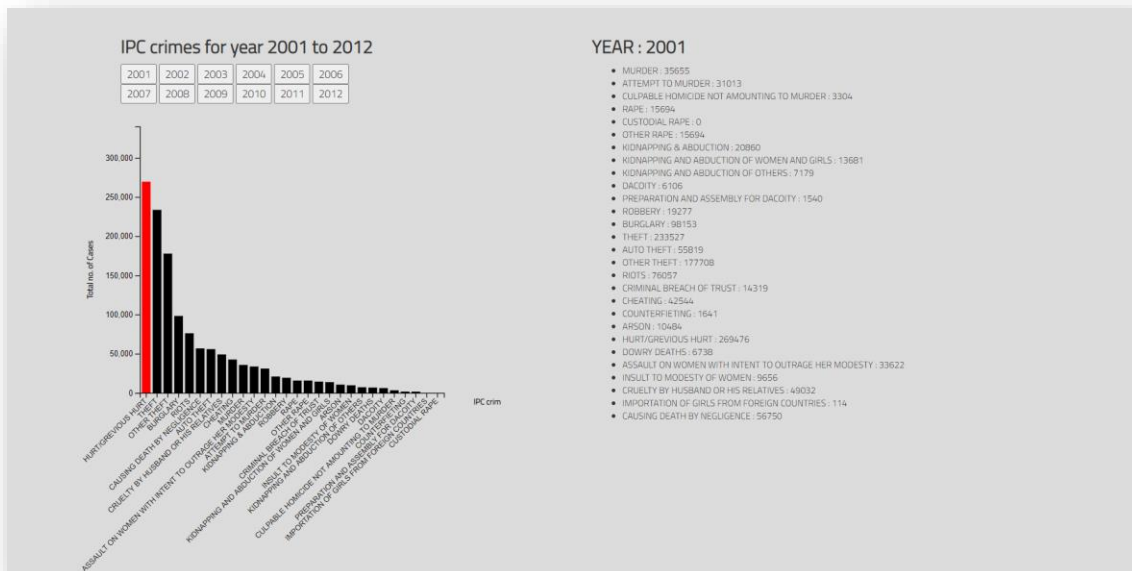
Figure 4: Bar graph with Button group to select a year

On hover the bar chart shows the Crime name and corresponding cases
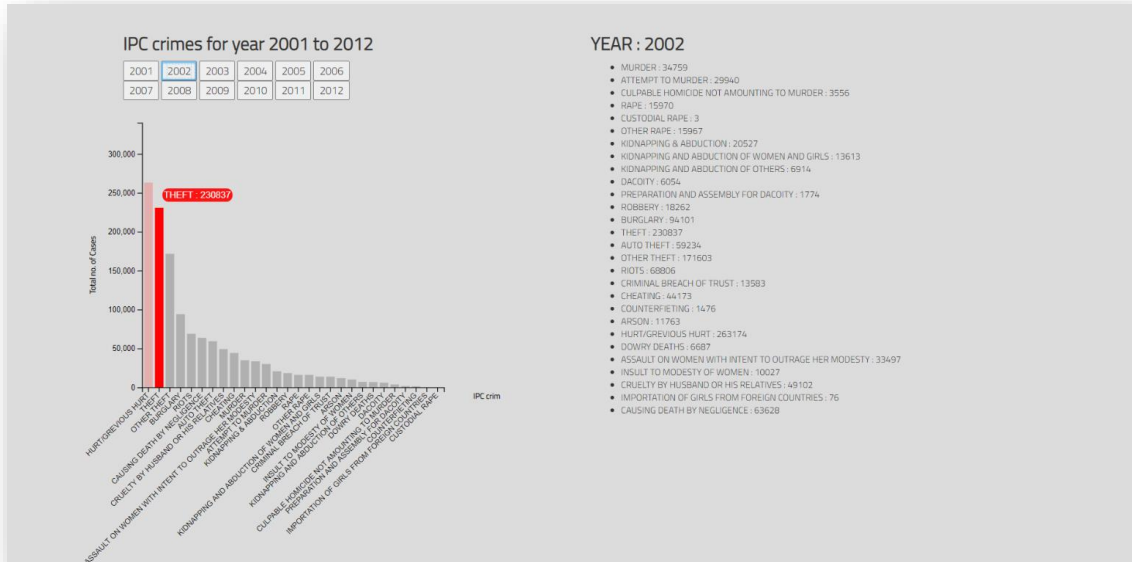
Figure 5: On hover shows highlights the Bar and shows the cases

On clicking the Findings tab, takes the focus to Section having Bubble chart. On hovering a Bubble, gives more details like State name, Cases, Poverty rate and Literacy rate.
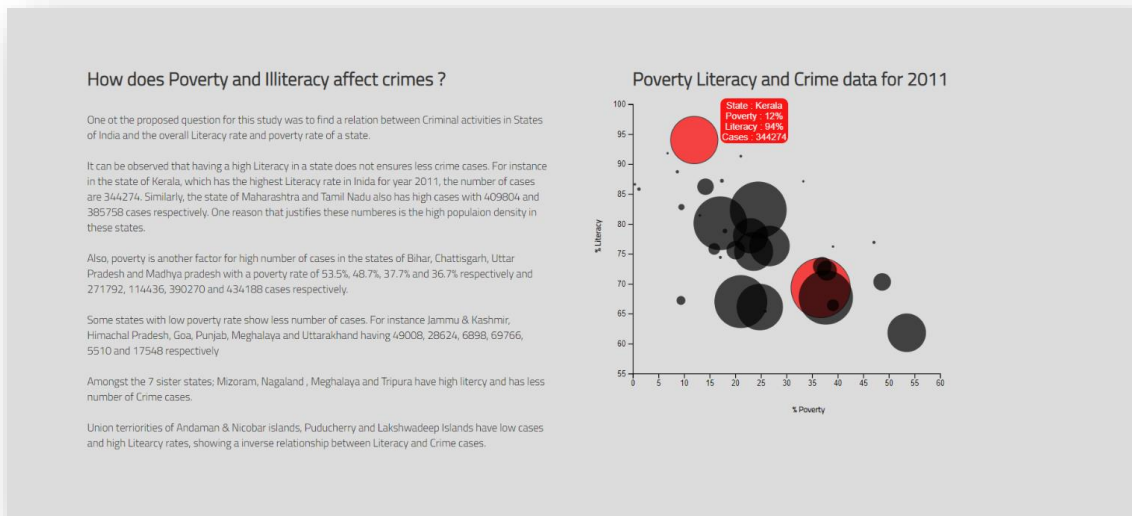
Figure 6: Bubble chart showing relation between poverty, literacy and crime for data of year 2011

Other sections of the visualisation.
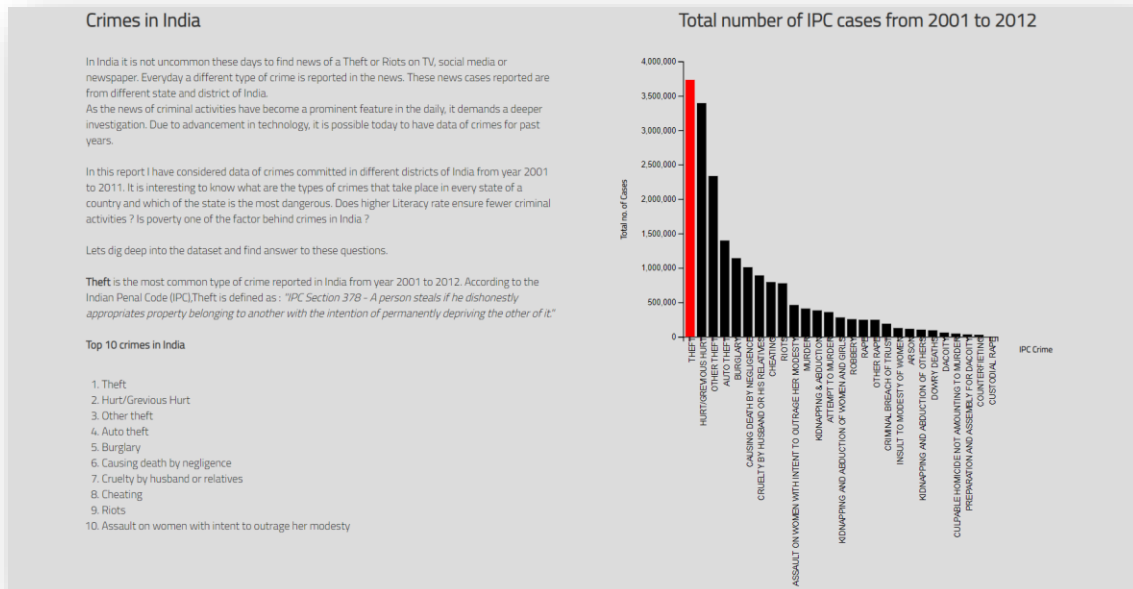
Figure 7: Bar graph showing total cases for each crime type
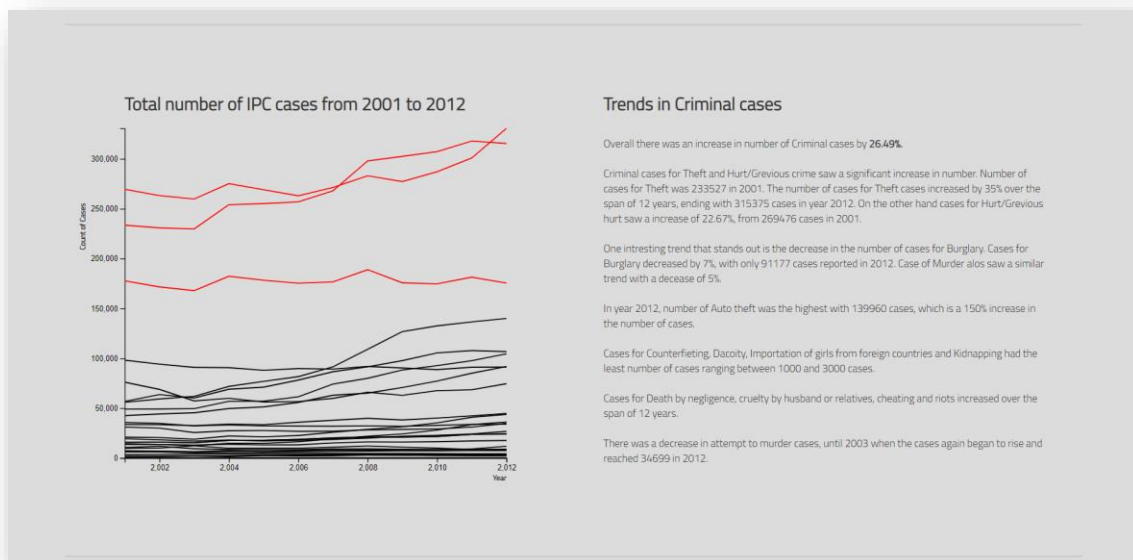
Figure 8: Multiple line graph showing crime for different year and type

13
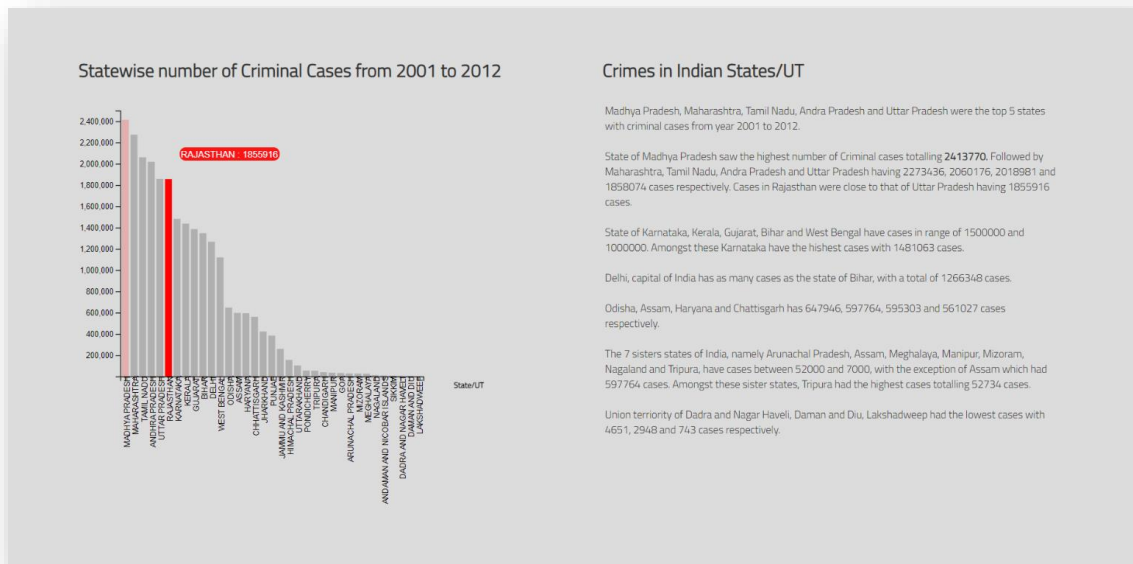
Figure 9:Bar graph showing state wise total number of cases
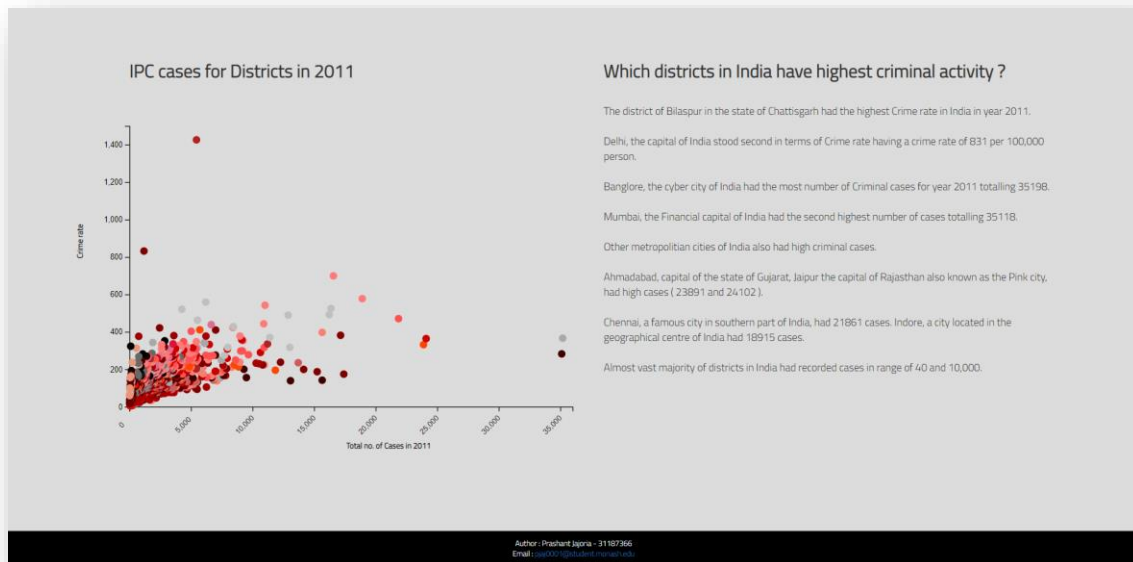
Figure 10: Scatter plot showing distribution of cases for district and the corresponding Crime rate

Conclusion

This interactive visualization helps better know the IPC crime dataset, census dataset and Dataset having Poverty and Literacy values using visually appealing and interactive plots, that keep the audience engaged in the topic. It helps in answering the proposed questions easily and also give insights that otherwise are difficult to get using traditional ways of Data representations.

The design phase using Five design sheets make the process of designing the final layout easy, and also allows to explore all the ideas for visualization.

In future a Choropleth map to show cases in different districts of India can be implemented to improve the visualization.

Reflection

By implementing this project in D3.JS I have learned a new Data visualization framework, which earlier I was not familiar with. The learning of D3.JS also improved my JavaScript abilities.

By doing the project single handedly, I have learned to organize my work by maintaining separate copies of code versions, work within set deadlines and document my entire work for future referencing, which is an important part every developer need to do when working professionally. I learned to debug the JavaScript code in the web browser, capture the data requested by webpages using Chrome Developer tools.

I am confident to implement a Data driven visualization if needed for future projects.

Bibliography

- Sheet 2, 3, 4 – Initial designs | Five Design Sheet. (2020). Retrieved 7 June 2020, from http://fds.design/index.php/2015/06/25/sheet-2-3-4-initial-designs/

- Data Visualization Salary in Australia | PayScale. (2020). Retrieved 7 June 2020, from https://www.payscale.com/research/AU/Skill=Data_Visualization/Salary

- Host a Personal Website - Amazon Web Services (AWS). (2020). Retrieved 7 June 2020, from https://aws.amazon.com/getting-started/hands-on/host-static-website/

- D3.js. (2020). Retrieved 21 June 2020, from https://en.wikipedia.org/wiki/D3.js

- What is D3.js?. (2020). Retrieved 21 June 2020, from https://www.tutorialsteacher.com/d3js/what-is-d3js

• Holtz, Y. (2020). The D3 Graph Gallery – Simple charts made in d3.js. Retrieved 21 June 2020, from https://www.d3-graph-gallery.com

Appendix

Five Design Sheets
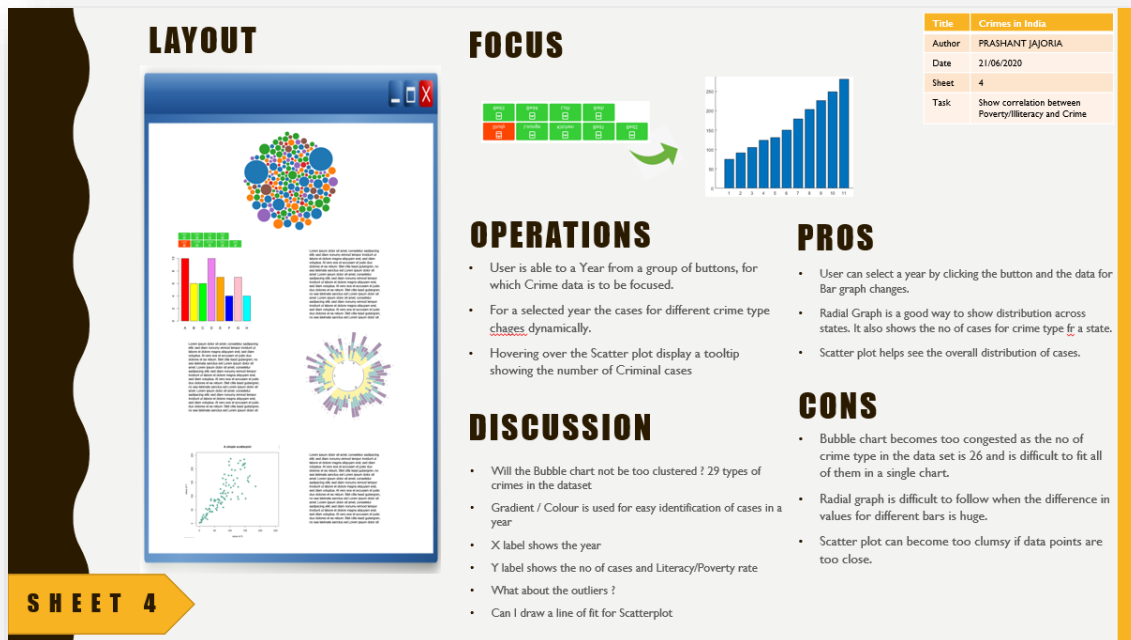

Figure 11: Sheet 1

Figure 12: Sheet 2

Figure 13: Sheet 3

Figure 14: Sheet 4

Figure 15: Sheet 5