

Process Book for Mental Health In India

Madhura Mathkar
u1266039
u1266557@uemail.utah.edu

Monesha Murdeshwar
u1266063
u1266063@uemail.utah.edu

Nikhil Ramesh
u1266557
u1266557@uemail.utah.edu

1. Overview

This project hopes to identify the most common reasons for suicides and their geographical and demographic distribution. Our visualization aims to find patterns in different causes, age-groups, and gender orientations and analyze how various Indian states differ in their suicide rates from 2001 to 2012. We hope the visualization insights will help people and authorities create awareness and develop better policies accordingly.

1.1 Motivation

According to the Indian Ministry of Statistics and Program Implementation, just in 2020, about 220,481 people committed suicides, the highest in the world. Attributing to the large population of the country, the suicide rate per person is relatively low. In Indian Society, the subject of mental health is often a taboo topic of discussion. The importance given to a person's mental well being is significantly less as compared to their physical well being. There is much stigma around receiving professional help regarding mental health problems, and these problems are often shunned away as trivial. This stigma often leads to people with mental health issues feeling isolated and without options. In this project, we hope to show the importance of open discussion on mental health by highlighting the numbers and the reasons that have lead people to commit suicide in the country. Since each state in India is entirely different from the others in terms of culture and socio-economics, the suicide causes and demographics that commit them also change from State to State. These factors are also something that we want to analyze in this project

1.2 Goals

There are a lot of underlying patterns that we hope to discover through our visualization. They are listed below:

1. An overview of the state's suicide rates over the time period
2. How do regions compare with one another?
3. The leading causes for suicide within each region
4. The relationship between gender and cause
5. The relationship between age group and cause

We believe identifying the various causes and their relationship with gender and age groups will provide people with the information required to identify and help people in similar situations.

2. Related Work

In recent times the topic of mental health has garnered a lot of attention in India. Even so there continues to be a lot of stigma surrounding the topic in the country. There is a need for a better understanding and open discussion on the topic. In line with this we found some work on World suicide data in [2]. We realized we could create Suicide awareness and study some of the common causes for suicide in India to contribute to the efforts currently beginning on the subject of mental health in the country

3. Questions

We were trying to answer the following questions:

1. We wanted to explore what causes people to take such an grave step such as suicide.
2. We also wanted to know in which demographics suicides are more prevalent and how it has changed over the years.
3. To see if there were underlying patterns between State, gender, age-groups with causes of suicide.

4. Data

4.1 Source

The data was obtained from kaggle [3]. It consists of the number of suicides per state per year for the years 2001 to 2012. The data also consists columns that detail the causes for the suicide , the age group as well the gender of the people committing suicide.

```
[ ] display(df)
```

	State	Year	Type_code	Type	Gender	Age_group	Total
0	A & N Islands	2001	Causes	Illness (Aids/STD)	Female	0-14	0
1	A & N Islands	2001	Causes	Bankruptcy or Sudden change in Economic	Female	0-14	0
2	A & N Islands	2001	Causes	Cancellation/Non-Settlement of Marriage	Female	0-14	0
3	A & N Islands	2001	Causes	Physical Abuse (Rape/Incest Etc.)	Female	0-14	0
4	A & N Islands	2001	Causes	Dowry Dispute	Female	0-14	0

Fig. 1. Dataset with only Type code = Causes entries

```
[ ]
Sickness = ['Other Prolonged Illness', 'Illness (Aids/STD)', 'Cancer', 'Paralysis', 'Insanity/Mental Illness']
Fin_issue = [ 'Poverty', 'Professional/Career Problem', 'Bankruptcy or Sudden change in Economic',
              'Bankruptcy or Sudden change in Economic Status', 'Property Dispute', 'Unemployment']
Rel_trb = ['Family Problems', 'Death of Dear Person', 'Not having Children (Barrenness/Impotency',
           'Not having Children(Barrenness/Impotency', 'Love Affairs', 'Divorce', 'Suspected/Illicit Relation',
           'Dowry Dispute', 'Illegitimate Pregnancy', 'Cancellation/Non-Settlement of Marriage']
Societal_pressure = [ 'Ideological Causes/Hero Worshipping', 'Failure in Examination', 'Fall in Social Reputation']
```

```
def new_col(row):
    s = row["Type"]
    if s in Sickness:
        return "Sickness"
    elif s in Fin_issue:
        return "Financial issues"
    elif s in Rel_trb:
        return "Relation Troubles"
    elif s in Societal_pressure:
        return "Pressures of Society"
    elif s == "Other Causes (Please Specity)":
        return "Others Causes"
    else:
        return s
```

Fig. 2. Code block used to segregate data into categories

Post cleaning the data has 39,531 rows with 1,440,974 cases of suicides over the 10 years.

4.2 Preprocessing

We used python on google colab to pre-process the data and store the result into three different CSV files. Different parts of the visualization use these CSV files. First, we removed the rows where the number of suicides was zero. As discussed in the proposal, we decided to retain only those rows where the “Type code” was “Causes.” This is because the remaining type codes had lesser data and had errors within the data.

Further, the data had no clear explanation of the relation between the different ”Type code” values. Further, we decided to bin similar reasons for suicide into six main categories. This was done to reduce clutter and make the understanding of the visualization easier. Fig 2 shows the code used for the above process. We used a geo-Json of Indian states to draw the map of India. State names in our dataset did not match with the geo-Json hence we made changes to match both.

Separate CSV files were created for different visualizations.

1. Project.data2.csv : This dataset has the structure shown in Fig 3. We calculated the total suicides in the states for a year. Then using the 2011 census data, we calculated the suicide rates in each state over the years. We chose the 2011 census data since a census is conducted every ten years, and this was the more recent data available within the period of our dataset. Creating the Heatmap using this dataset meant we did not have to calculate the suicide rate per state per year every time the Year scroll was moved , saving a lot on processing time and preventing lag

	Year	State	Total	Percentage
0	2001	Andaman and Nicobar	129	33.858268
1	2001	Andhra Pradesh	10522	12.440146
2	2001	Arunachal Pradesh	111	8.020231
3	2001	Assam	2647	8.482343
4	2001	Bihar	603	0.579256

Fig. 3. Project_data2.csv

2. Project.data3.csv : This dataset has the structure shown in Fig 4. The data sets holds the suicides per State per Year Per Category per Gender. This dataset was used to create the bar chart when toggling based on gender. This is done to cut down processing time and click and toggle when the bar charts are drawn

	State	Year	Categories	Gender	Total
0	Andaman and Nicobar	2001	Causes Not known	Female	12
1	Andaman and Nicobar	2001	Causes Not known	Male	27
2	Andaman and Nicobar	2001	Others Causes	Female	8
3	Andaman and Nicobar	2001	Others Causes	Male	9
4	Andaman and Nicobar	2001	Pressures of Society	Female	2

Fig. 4. Project_data3.csv

3. Project.data4.csv : This dataset has the structure shown in Fig 5. The data sets hold the suicides per State per Year Per Category per Age group. This dataset was used to create the bar chart when toggling based on age groups. This is done to cut down processing time and click and toggle when the bar charts are drawn.

	State	Year	Categories	Age_group	Total
0	Andaman and Nicobar	2001	Causes Not known	15-29	17
1	Andaman and Nicobar	2001	Causes Not known	30-44	10
2	Andaman and Nicobar	2001	Causes Not known	45-59	11
3	Andaman and Nicobar	2001	Causes Not known	60+	1
4	Andaman and Nicobar	2001	Others Causes	0-14	1

Fig. 5. Project_data4.csv

The original dataset did not have 2012 data for the state West Bengal, hence we decided to add 0, as the total number of suicides for West Bengal.

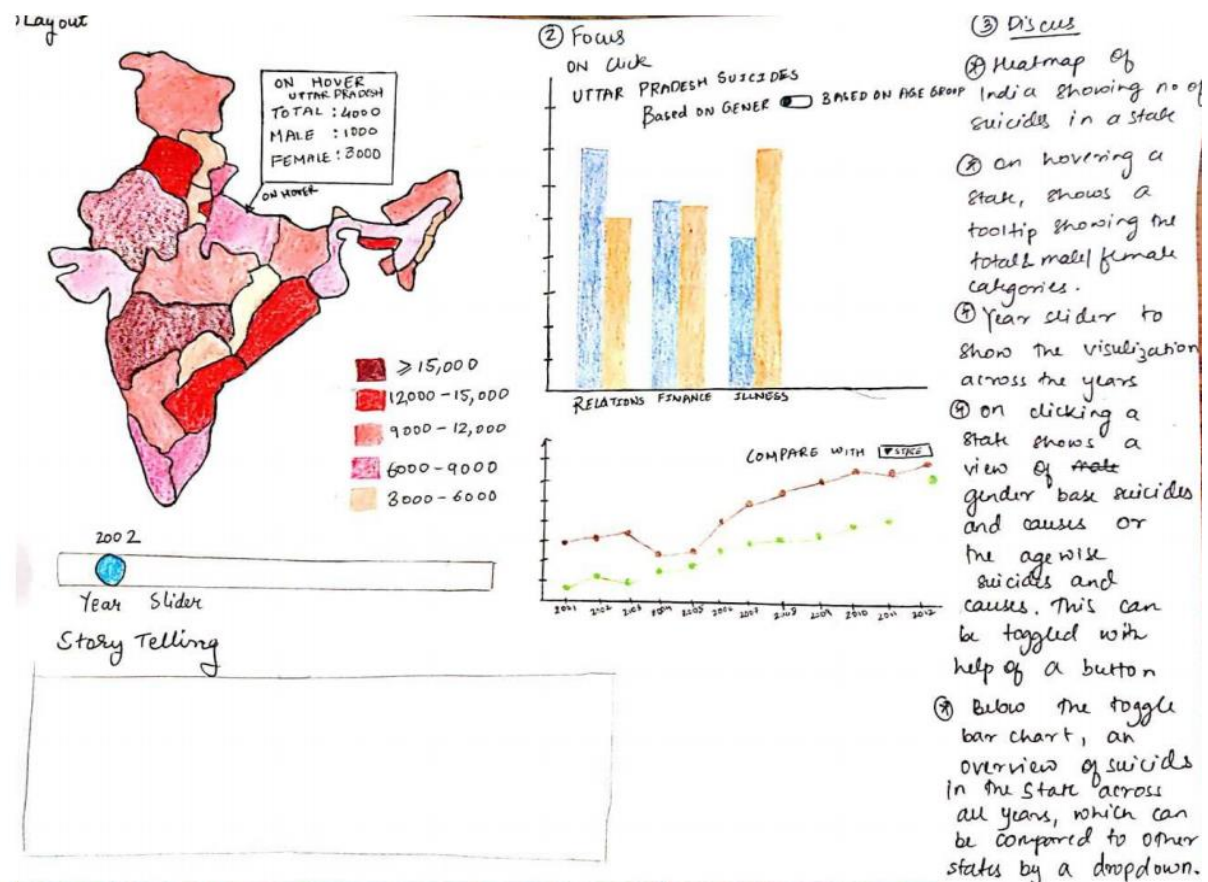


Fig. 6. Final Proposal Diagram

5. Design Evolution

The proposed design is shown in Figure 6. The design Choices made are as follows.

For our main view, we chose a Heatmap. We changed the Heatmap to represent the percentage of deaths(suicide rates) instead of absolute numbers on our project advisor's recommendation. This we felt would allow for a more like to like comparison between states that have varied populations. The two main attributes were state and suicide rates, and a Heat map displays them according to the effectiveness principle. Further, we used seven color brackets for the Heatmap to avoid any perceptual errors. The Heatmap, combined with the year scroller, allows the user to see the changes taking place in the suicide rates over the years.

On hover, we display the state name and the exact total suicides for the currently selected year. The name allows people who are not familiar with the Indian map to navigate it easily and provides the exact number of suicides instead of just the percentage for those interested.

One of the design choices we made was to have a small writeup show when the webpage is initially loaded. This we felt would act as an introduction to the visualization, some important suicide hotlines that will allow the viewers to help others in the future and summarize the possible interactions available in the visualization.

The secondary view, which shows up on a particular state's click, has two parts: a bar chart and a line chart.

The bar chart allows for studying the reasons behind the suicides split by gender or Age groups. This choice was made to address our objectives to find any underlying patterns between causes and gender, age. Post the midpoint; we chose to separate the bar chart bars to allow for a more clean and accurate visualization and avoid any perceptual errors.

The line chart captures in one view the performance of the selected state over the period. We felt it was essential to capture the changes in a single view following the "eyes beat memory" rule. Further, a comparison option is provided to compare the suicide percentages over the years for two states. We chose to use absolute numbers for individual line charts but percentages for the comparison to ensure we account for the state populations.

6 Implementation

We implemented all our target features for the visualization. Listed below are the features and their implementation details.

6.1 Heat map of suicides rates in each state

Having looked at other designs, we decided a heatmap would best capture the data of all states for a particular year. One of our objectives was to give an overview of the states' suicide rates over a particular period and compare which states are doing better. These objectives were well met by having a heatmap. Using **d3.geoPath().projection** and the geoJson file, we implemented the map. For the color coding we used the percentages from **Project_data2.csv** to fill in the

colors. Everytime a state is selected it is highlighted using dark blue to signify the state is selected. Fig 7. shows our visualized Heatmap.

6.2 Tooltip on hovering over a state

On hovering the mouse over a state, it gives the state's name and the total suicides in the state in that given year. The tooltip allowed us to provide the state name without disturbing the heatmap and also provide the exact number of suicides per state. This was implemented using the events **mouseover** and **mouseout** and the data from **Project_data2.csv** was used for the value. Fig 7. shows our visualized Tooltip.

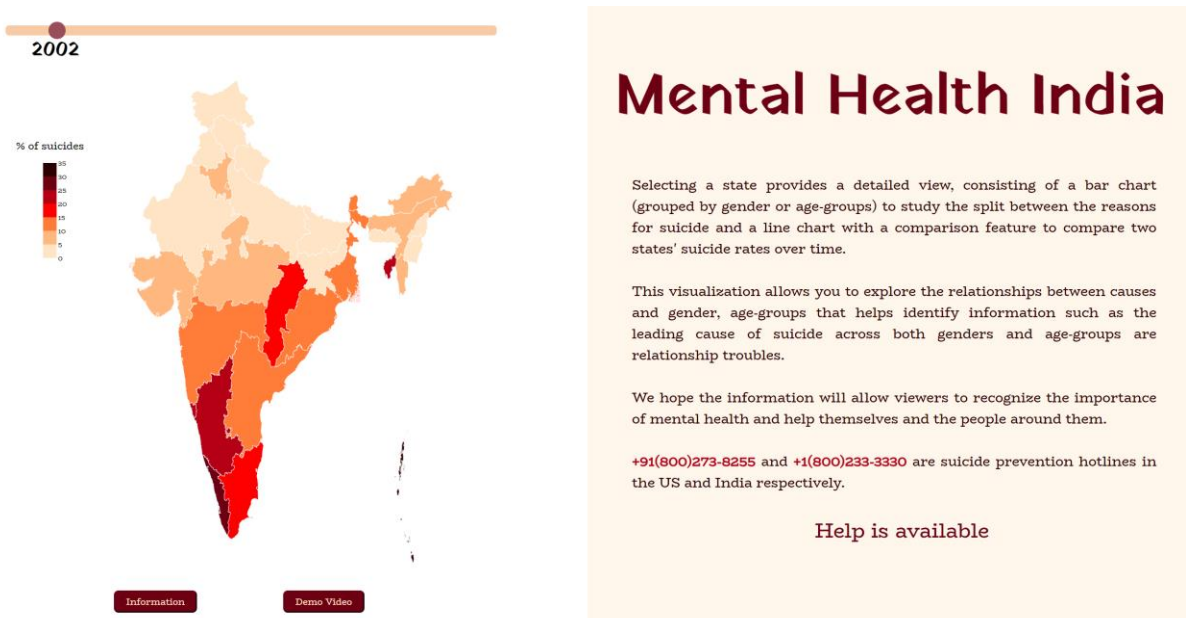


Fig. 7. Overview Diagram

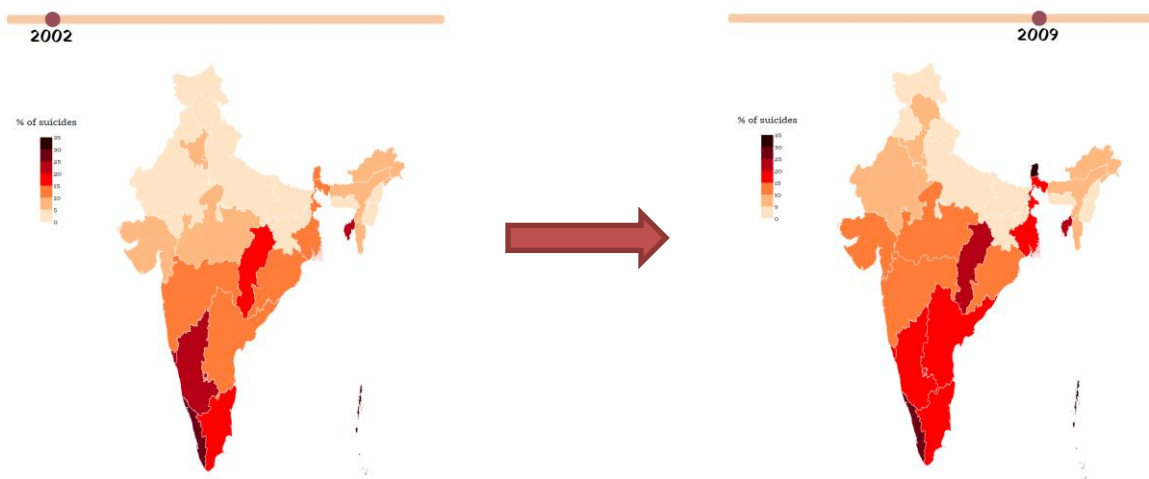


Fig. 8. Year slider and its effect on heatmap

6.3 Year Slider

We have a year slider for the years 2001 - 2012. Moving the slider allows us to visualize the data over the years. Fig 8. shows our implemented year slider, and the changes are reflected in the Heatmap. We chose a year slider as it helps communicate much data without cluttering the screen, and the Heatmap changes help us analyze suicide rates in different states over time. This was implemented using **-webkit-transition** in CSS and the input event, similar to the homework assignments. Figure 8 shows the implementation of our year slider.

6.4 Bar Chart

When we click on a state, it shows more details about the causes of suicides in the state grouped by categories. Within each category the values can be split by either gender or age groups. Fig 9. shows our visualized Bar chart. The bars are implemented using **d3 rect** and the data available from the **Project.data3.csv** and **Project.data4.csv**. Further animations that show the bars moving into position are added using the **d3 transition** function. Figure 9 shows the implementation of the bar chart.

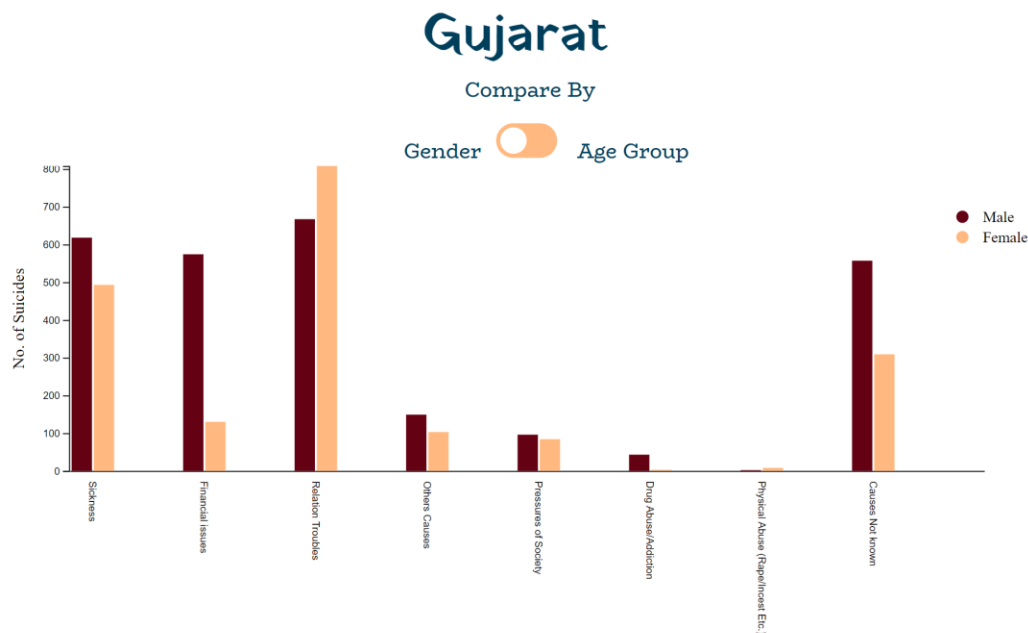


Fig. 9. Bar Chart

6.5 Toggling The Bar chart

A toggle switch is used to switch between visualization based on gender and visualization based on age-group. This toggle helps reduce clutter as we can show two different data in the same space. This is implemented using the **input** event and redrawing the bar chart whenever the toggle is used. Figure 9.1 shows our implementation of the toggle.

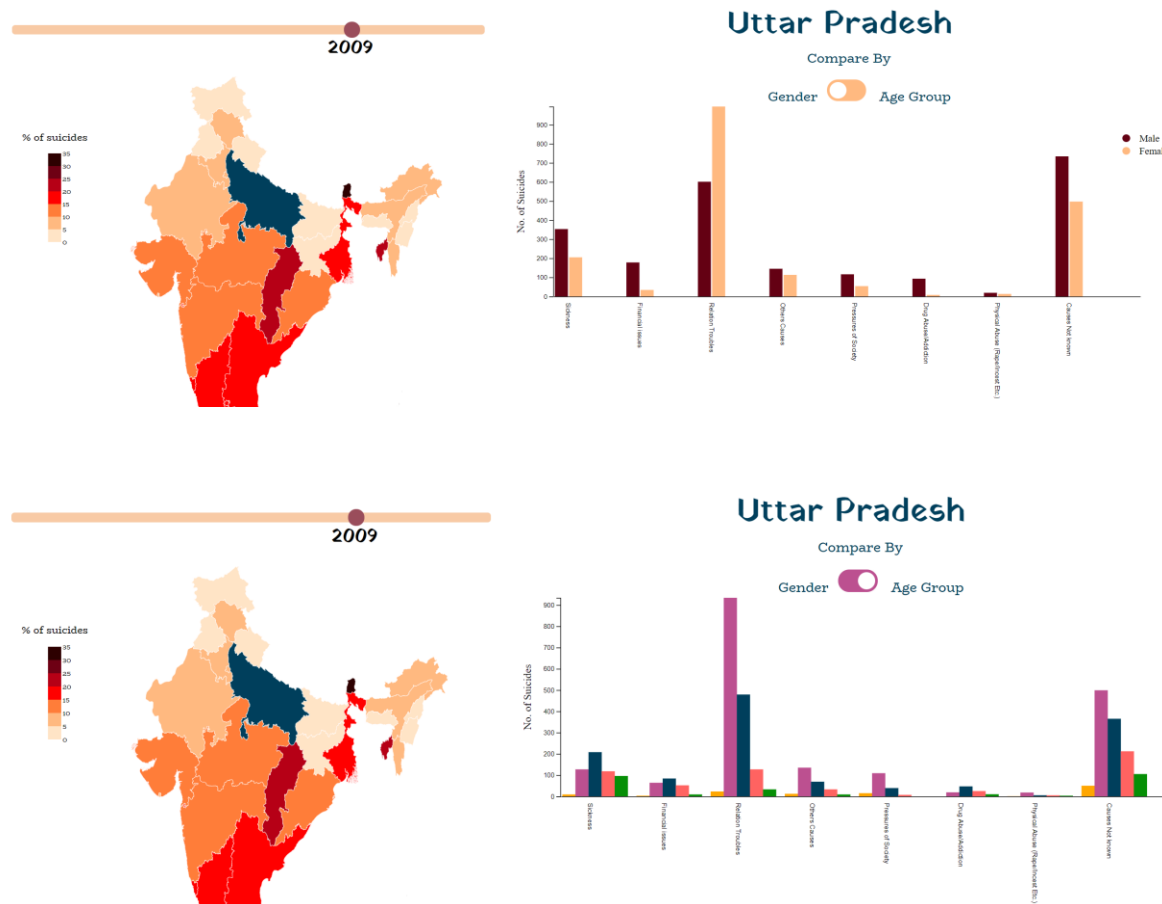


Fig. 9.1 Bar Chart and the effect after toggle

6.6 Line Chart

We wanted to show how the suicide numbers have transitioned over the years for a state in one view. The line chart shows the transition over the years, and it is easy to see the rise or fall in the numbers because of the added animation as the line chart transitions from the start year to the end year. The linechart is implemented using the **d3.line()** and data for total suicides in a state comes from **Project_data2.csv**. The lines are animated to grow from the year 2001 to 2012 using the **stroke-dasharray** and **transition** functions.

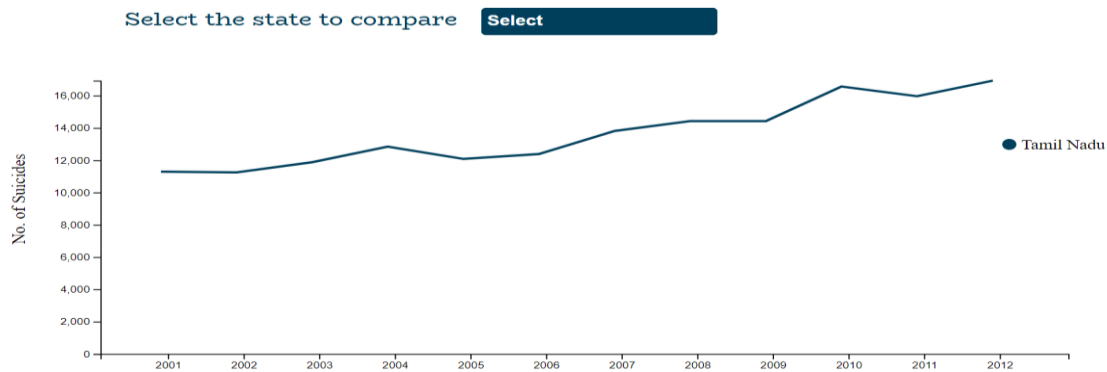


Fig. 10. Year slider and its effect on heatmap

6.6 Comparison in the Line Chart

We added a comparison of percentage suicides over the years for two states using the line charts. A dropdown is provided to select the state against which the comparison of the selected state should occur. The **change** event is used to check for any selections and redraw the lines accordingly.

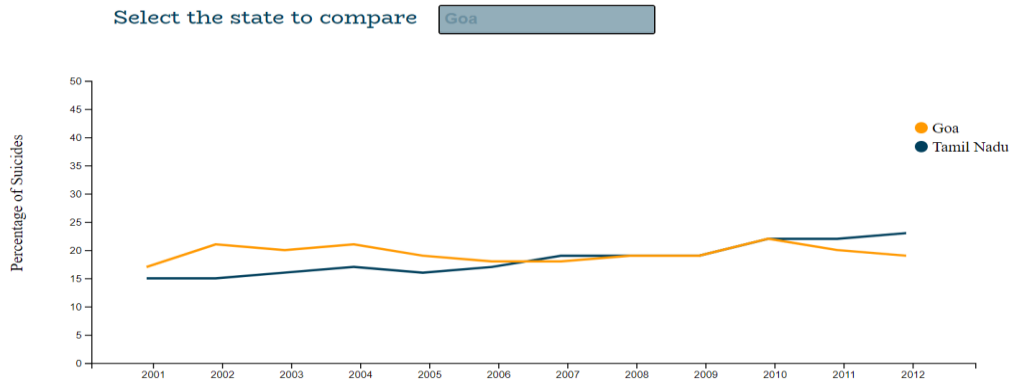


Fig. 11. Year slider and its effect on heatmap

6.7 Story Tell

We added a storytelling feature; when the story button is clicked, we show the states where the average percentage of suicides across all the years is maximum and minimum, respectively. We do this with the help of division boxes, which point to the respective states; we used transition to animate the lines which point to the respective state.

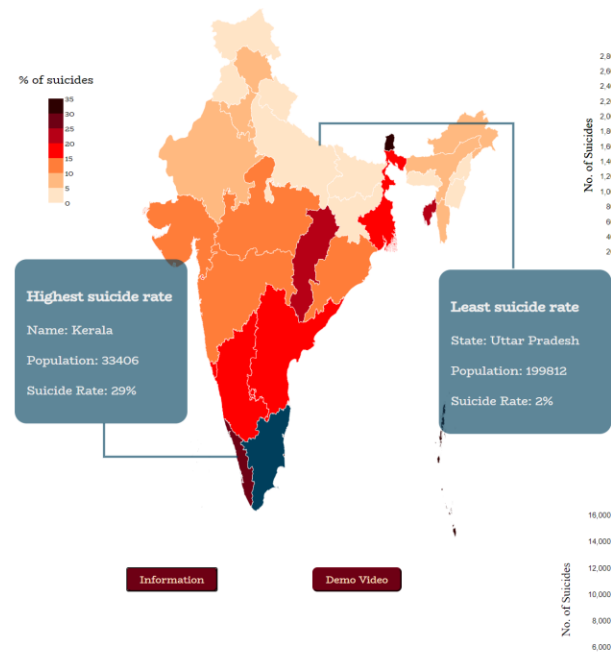


Fig. 12. Story Telling

7 Evaluation

The visualization enables us to draw out many insightful inferences. We can understand better what reasons lead to suicides and how are various age groups or genders affected by it.

Here are the few conclusions that we could draw

1. The state that did the worst in term of suicide rates is Kerela, while Uttar Pradesh had the least suicide rates
2. Relationship troubles are the leading causes of suicides for both males and females.
3. One significant difference in causes between male and female is that in males, financial issues lead to more suicides whereas for females, dowry dispute is a leading cause for suicides
4. The age group, 15-29, had the highest number of suicides.
5. Almost 10% of the Indian population commit suicide every year.

Information on how the states performed can be drawn from the Heatmap as well as the line charts. The Bar chart allows us to answer the causes of suicides and how they link to gender and age groups. We were also able to understand the leading causes of suicide statewide from the bar chart.

We believe that the visualization was able to meet all the objectives we identified while exploring the data. We believe the information from this can help people who might be in similar situations like those found in the data.

The one shortcoming we felt was the data is pretty dated. Though the conclusions may remain relevant, we feel newer data might interest people more, and there may be some more interesting facts and patterns we can uncover. The one shortcoming we felt was the data is pretty dated. Though the conclusions may remain relevant we feel newer data might interest people more and there may be some more interesting facts and patterns we can uncover.

7 References Links

[1] [Download Reports | Ministry of Statistics and Program Implementation | Government Of India \(mospi.nic.in\)](https://mospi.nic.in/)

[2] [World Suicide Data Visualization with Tableau | by Lulu Ilmaknun Qurotaini | Medium](#)

[3] [Suicides in India | Kaggle](#)

