

# Process Book for Mental Health In India Project

**Nikhil Ramesh**  
u1266557  
u1266557@umail.utah.edu

**Madhura Mathkar**  
u1266039  
u1266039@umail.utah.edu

**Monesha Murdeshwar**  
u1266063  
u1266063@umail.utah.edu

## 1 Overview

In this project, we hope to identify the most common reasons for suicides and their geographical and demographic distribution in the country. 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 insights obtained from the visualization will help people and authorities to 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 commits 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 states 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 question:

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 Specify)":
        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.

Then 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 on 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. The name of the states in our dataset did not match with that of the geo-Json hence we had to make changes so that both matched.

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. Creating the heatmap using this dataset meant we did not have to count the total per state per year every time the Year scroll was moved, saving a lot on processing time and preventing lag.
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
3. **Project.data4.csv** : This dataset has the structure shown

```
data > Project_data2.csv
1 ,Year,State,Total
2 0,2001,Andaman and Nicobar,129
3 1,2001,Andhra Pradesh,10522
4 2,2001,Arunachal Pradesh,111
5 3,2001,Assam,2647
6 4,2001,Bihar,603
7 5,2001,Chandigarh,70
8 6,2001,Chhattisgarh,4025
9 7,2001,Dadra and Nagar Haveli,50
10 8,2001,Daman & Diu,14
11 9,2001,Delhi,1239
12 10,2001,Goa,256
13 11,2001,Gujarat,4791
14 12,2001,Haryana,2007
```

Fig. 3. Project.data2.csv

```
data > Project_data3.csv
1 ,State,Year,Categories,Gender,Total
2 0,Andaman and Nicobar,2001,Causes Not known,Female,12
3 1,Andaman and Nicobar,2001,Causes Not known,Male,27
4 2,Andaman and Nicobar,2001,Others Causes,Female,8
5 3,Andaman and Nicobar,2001,Others Causes,Male,9
6 4,Andaman and Nicobar,2001,Pressures of Society,Female,2
7 5,Andaman and Nicobar,2001,Pressures of Society,Male,2
8 6,Andaman and Nicobar,2001,Relation Troubles,Female,9
9 7,Andaman and Nicobar,2001,Relation Troubles,Male,14
10 8,Andaman and Nicobar,2001,Sickness,Female,19
11 9,Andaman and Nicobar,2001,Sickness,Male,27
12 10,Andaman and Nicobar,2002,Causes Not known,Female,11
13 11,Andaman and Nicobar,2002,Causes Not known,Male,24
```

Fig. 4. Project.data3.csv

```
data > Project_data4.csv
1 ,State,Year,Categories,Age_group,Total
2 0,Andaman and Nicobar,2001,Causes Not known,15-29,17
3 1,Andaman and Nicobar,2001,Causes Not known,30-44,10
4 2,Andaman and Nicobar,2001,Causes Not known,45-59,11
5 3,Andaman and Nicobar,2001,Causes Not known,60+,1
6 4,Andaman and Nicobar,2001,Others Causes,0-14,1
7 5,Andaman and Nicobar,2001,Others Causes,15-29,10
8 6,Andaman and Nicobar,2001,Others Causes,30-44,3
9 7,Andaman and Nicobar,2001,Others Causes,45-59,1
10 8,Andaman and Nicobar,2001,Others Causes,60+,2
11 9,Andaman and Nicobar,2001,Pressures of Society,0-14,1
12 10,Andaman and Nicobar,2001,Pressures of Society,15-29,3
13 11,Andaman and Nicobar,2001,Relation Troubles,0-14,1
```

Fig. 5. Project.data4.csv

in Fig 5. The data sets holds 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.

## 5 Design Evolution

### 5.1 Proposal design

The proposal design is shown above. The design Choices made are as follows.

The heatmap combined with the year scroller allowed the viewer to see the changes in the states clearly over time and any changes would clearly be visible. The two main attributes were state and number of suicides and a Heat map displays them in accordance to the effectiveness principle.

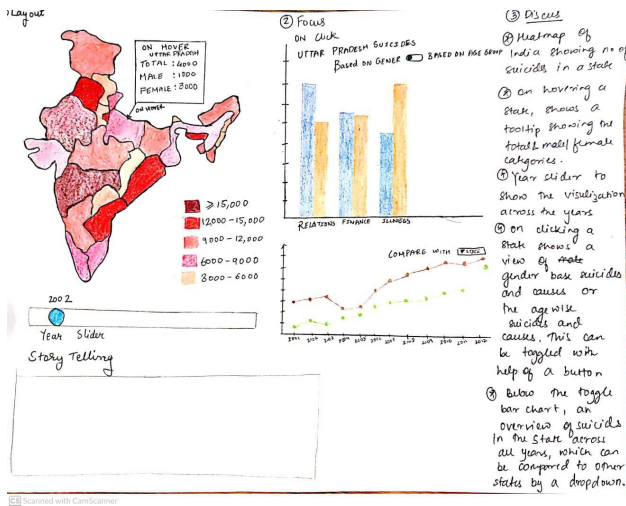


Fig. 6. Final proposal Viz

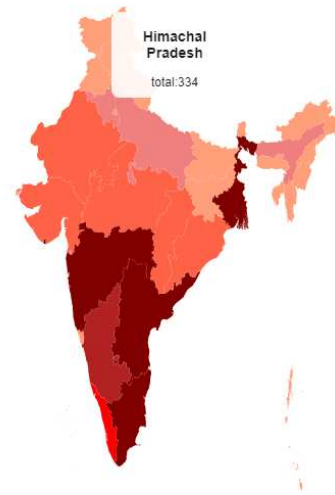


Fig. 7. Heatmap and tool-tip

Further we used 7 color brackets for the heatmap to avoid any Perceptual errors.

The on Hover function was changed to only show the State name and total. This was decided since the split between Gender is shown on selection in the detailed view and thus felt redundant.

On click of a state allows to view a bar chart showing the numbers per category.

A design choice made, for the age-group grouped bars, was to allow the bars to overlap because we felt it's easier to compare within each category and distinguish between categories. Further overlapping looked aesthetically pleasing and hence was used.

Another design decision made on the linked view is , If another state is picked , the toggle is reset to default Gender based split within the categories. The same occurs to the toggle if the Year scroller is shifted after a State is selected. This we felt will make it clear to the user that the detailed view is changing

## 6 Implementation

We have implemented five of our must-have features. In the points further, we explain their functionalities.

### 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 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. The heatmap we felt creates an immediate impact with the viewer and it makes it easy to clearly distinguish how the states are doing with respect to one another. Fig 7. shows our visualized Heatmap

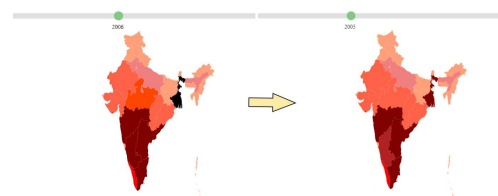


Fig. 8. Year slider and its affect on heatmap.csv

### 6.2 Tooltip that gives information about the suicides in the state in that year

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 name of the state we decided to have on Hover since people may not be familiar with state names in the country but having them printed on the state would make the heatmap less effective. Further displaying the exact total meant people interested in exact numbers could access the data by simply hovering over the states. Fig 7. shows our visualized Tooltip.

### 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 changes in the heatmap help us analyze suicide rates in different states over time.

### 6.4 Bar Chart of suicides of the state by categories

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. We chose to use a bar chart as it makes it easy to compare data and is aesthetically pleasing. The names of the categories are listed below the bars.

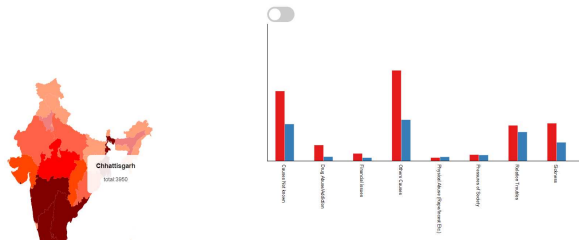


Fig. 9. Bar chart comparing suicide rates by cause categories and gender

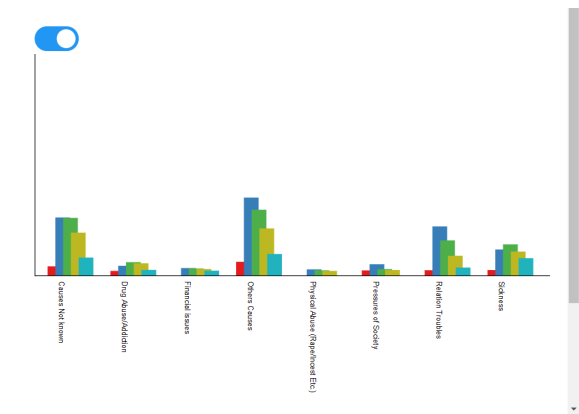


Fig. 10. Bar chart comparing suicide rates by cause categories and age-group on toggling

### 6.5 Toggling between suicides by gender and suicides by age-groups

A toggle switch is used to toggle 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. The reason we use this toggle is to understand the relationship between gender and the causes as well as age-group and causes without adding to any clutter in the Visualization. Fig 10. shows our visualized toggle switch .

## 7 References

Links to the mentioned sources,

- [1] Suicide Statistics.
- [2] Related work
- [3] Dataset Kaggle source.