IT in Domains - HealthCare Project Report

Indian Suicide Statistics - Data Visualisation

By: G Neha (IMT2014018), K C Pavithira (IMT2014025), Meghana Kotagiri (IMT2014034)

Objective

We had a dataset of suicides in each state in India, classified according to various parameters (suicide causes, education status, by means adopted, professional profile, social status) from 2001-2012 by the National Crime Records Bureau (NCRB), Govt of India. Our goal was to analyse and visualise the suicide data based on causes, demography etc. chronologically to find interesting and useful insights. The analysis and visualisations can help the medical institutions and Governments in understanding the existing scenario and in devising better suicide preventive interventions.

Exploratory Analysis & Correlation with Google Trends

1) To analyse and visualise the statewise suicide data based on :
    a) Demography
    b) Causes
    c) Chronologically
    d) Geographically

to find interesting and useful insights.

2) To analyse the situation in India, as compared to other countries and find parameters that it can improve upon, to reduce the suicide rates.

3) To find whether there is any correlation between google searches related to ‘suicide’, ‘depression’ and the actual incidents of suicides in a place, in a particular time period.

Target

Every suicide is a tragedy. It is an extremely complex issue, and while it is not possible to pin down its causes, There are some risk factors that can be identified with suicide, mainly through correlations and analysis of past data. Our project helped in identifying some of these factors, which shed light on appropriate interventions that can be taken to prevent suicide.

Our project workflow involved exploratory analysis followed by visualisations, which were embedded to build a dashboard.

Dataset

1) Data Source: National Crime Records Bureau (NCRB), Govt of India
2) Google Trends
3) Global Health Observatory data repository of World Health Organisation (WHO)

Dataset Description

- Time Period: 2001-2012
- Granularity:
  - Yearly
  - State and UT level
- Size: 2,37,520 data points
- Parameters:
  - State
  - Year
  - Type Code
  - Gender
  - Age Group

Columns & Values

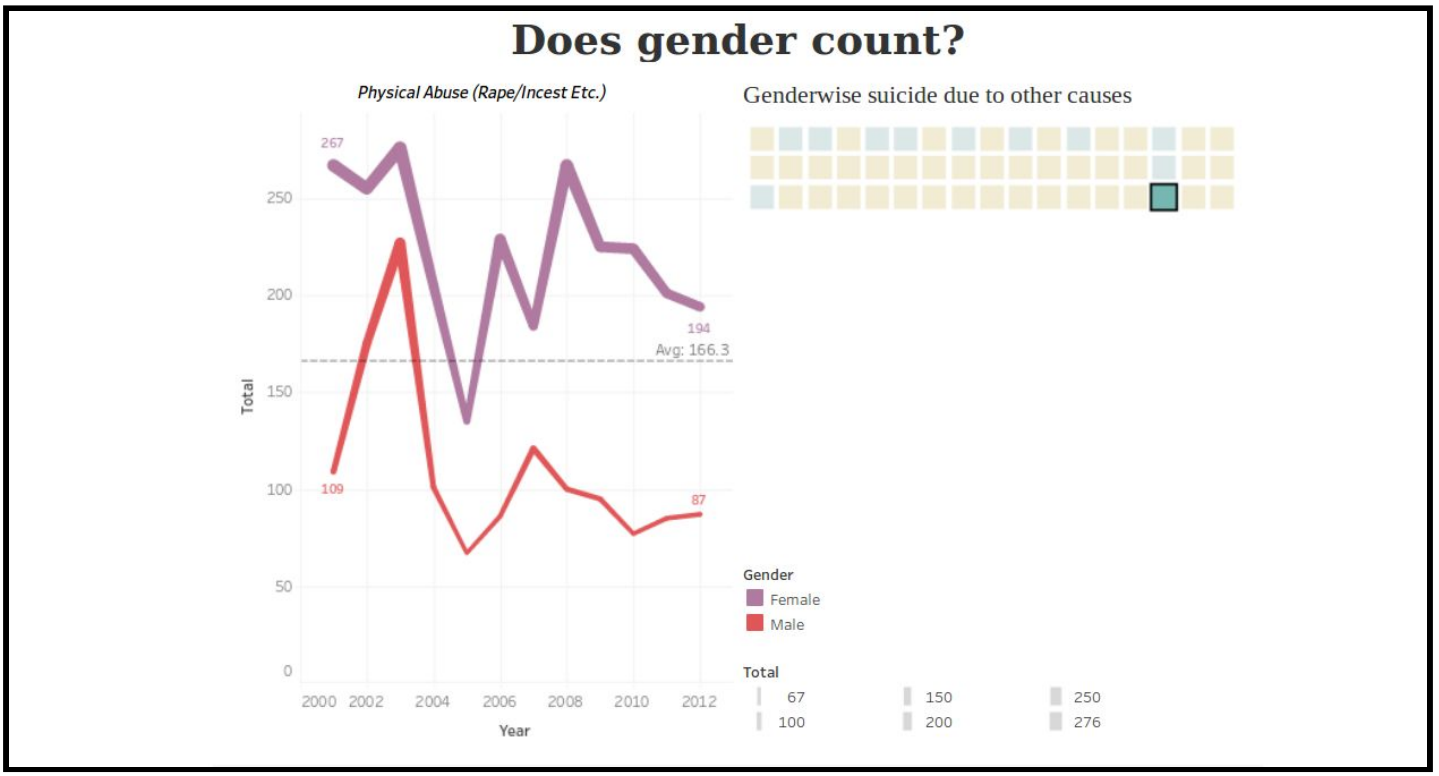
Causes: Unemployment, Love Affairs, Failure in Examination, Dowry Dispute, Death of Dear Person etc….
Means Adopted: By consuming insecticides, By fire/self-immolation, By over-alcoholism, By hanging, By drowning etc…..
Professional Profile: Farming/agricultural activity, student, unemployed, retired person, house-wife etc….
Education Status: Matriculate/Secondary, Graduate, No Education etc….
Social Status: Married, Divorced, Never Married etc…..

Google Trends

It is a freely accessible tool that allows users to interact with internet search data which can be used to get deep insights into population behaviour and health related phenomenon. We used this tool to find internet searches for a particular keywords like suicide, depression, exams etc. and found data region wise chronologically.

Results

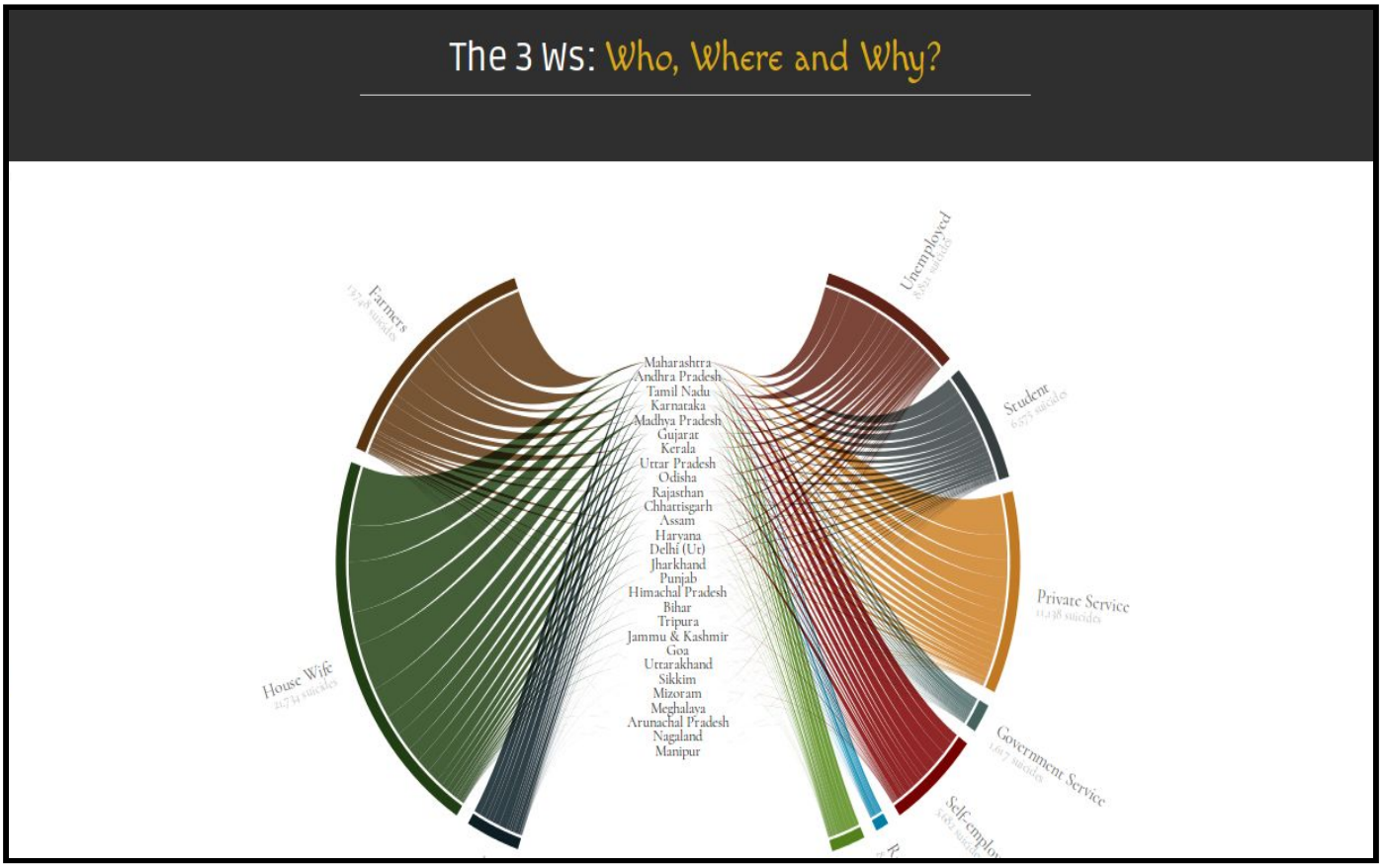
Reading the plot: In this visualization, we capture gender bias for different factors: causes, means adopted for suicide and education profile of suicide victims. The blocks on the right represent these factors. The line plot on the left gives the number of male and female suicides committed in years 2000-2012 for the selected factor along with national average. The thickness of line is based on the value.



Insights observed: For factors such as physical abuse, self immolation etc. more number of female suicides were observed compared to male suicides. Though in totally, male suicides are way higher. For the factor drug abuse, number of male suicides are chronologically increasing and are way higher than females.

Tools used: Tableau for visualization, R script for preprocessing

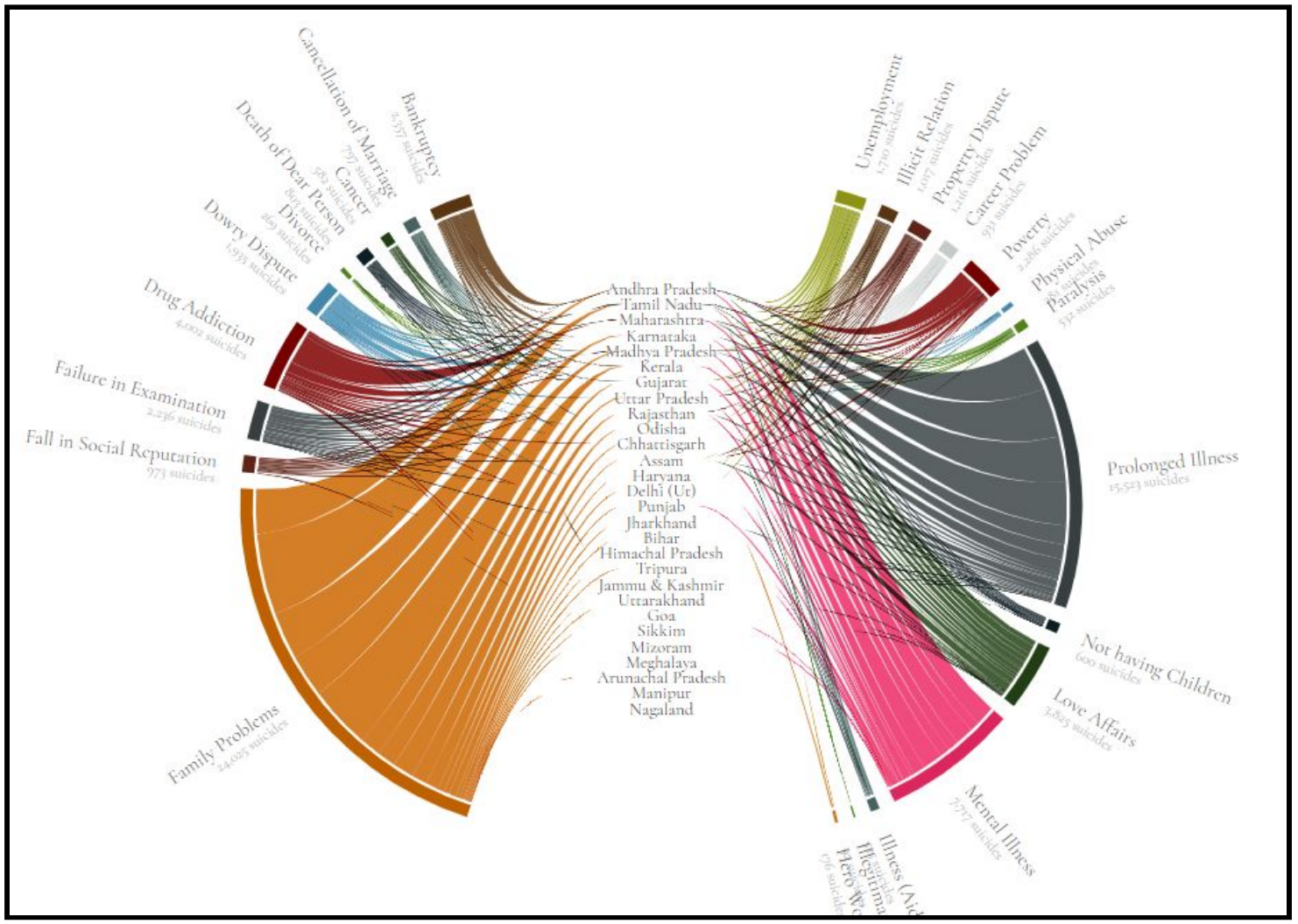
Reading the plot: This visualization captures the relationship between states and professional profile of suicide victims(who and where). The outer arc length indicates the number of suicides due to particular factor. The states are arranged in decreasing order of number of suicide victims.
On hovering over a state, we can see total number of suicide victims for that state. It also highlights the factors relevant only for that state and displays the number of suicides due to that factor.



Insights observed: Surprisingly, maximum number of people committing suicides were housewives followed by those in farming/agricultural activities.

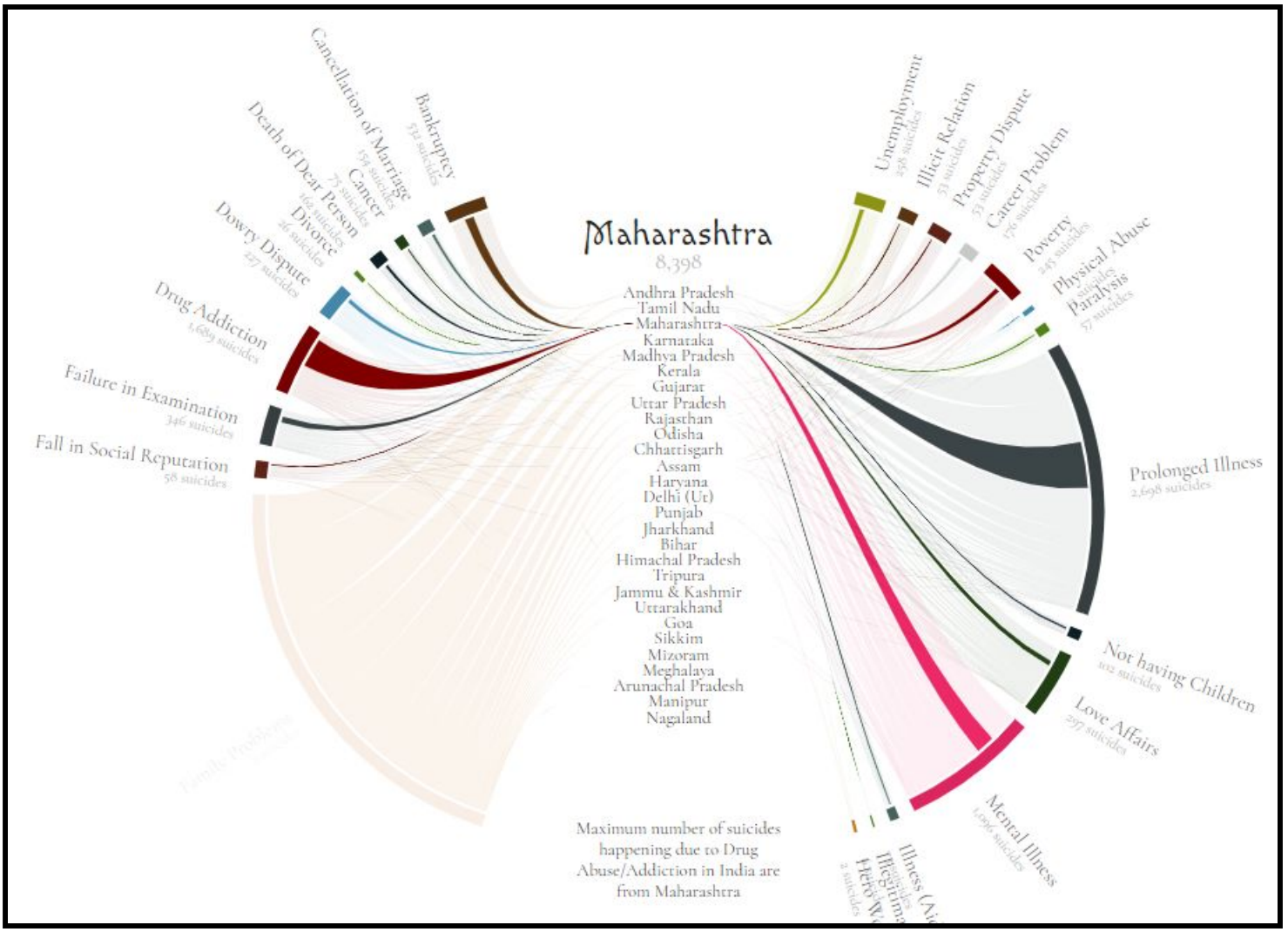
Tools used: d3.js for visualization, python script for preparing json files and preprocessing

Reading the plot: This visualization captures the relationship between states and causes of suicide(why and where). The outer arc length indicates the number of suicides due to particular cause. The states are arranged in decreasing order of number of suicide victims.
On hovering over a state, we can see total number of suicide victims for that state. It also highlights the causes relevant only for that state and displays the number of suicides due to that cause.



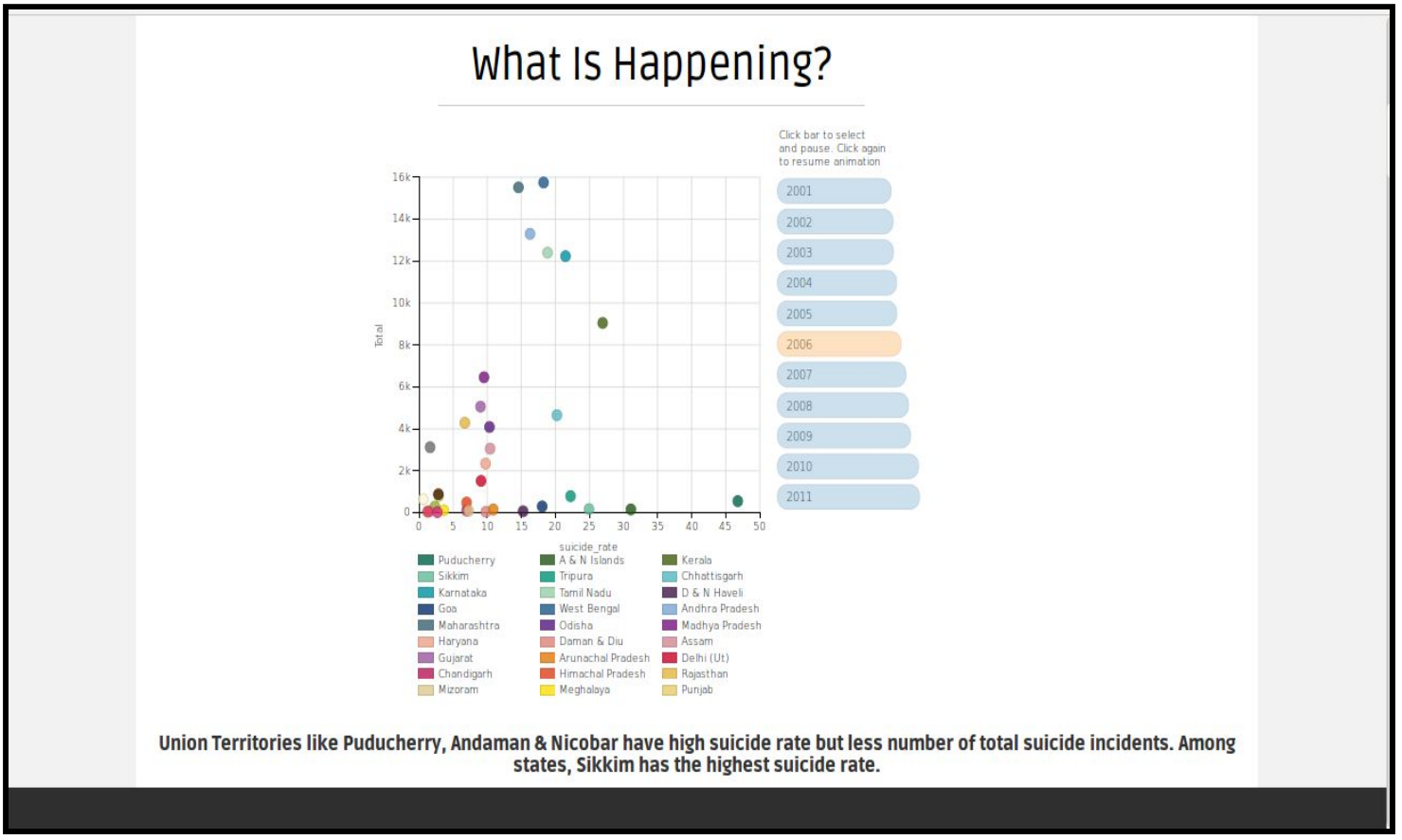
Insights observed: Overall, one of the major causes of suicide has been reported to be FAMILY PROBLEMS. It is especially higher in the four Southern states. Other major causes of suicide have been mental illness and other prolonged illness.

Tools used: d3.js for visualization, python script for preparing json files and preprocessing



Insights observed: For Maharashtra, major causes of suicide are prolonged illness, mental illness and drug abuse. Maximum number of suicides committed due to drug abuse/ drug addiction are from Maharashtra. So, government can accordingly implement preventive measures to specifically tackle this cause.

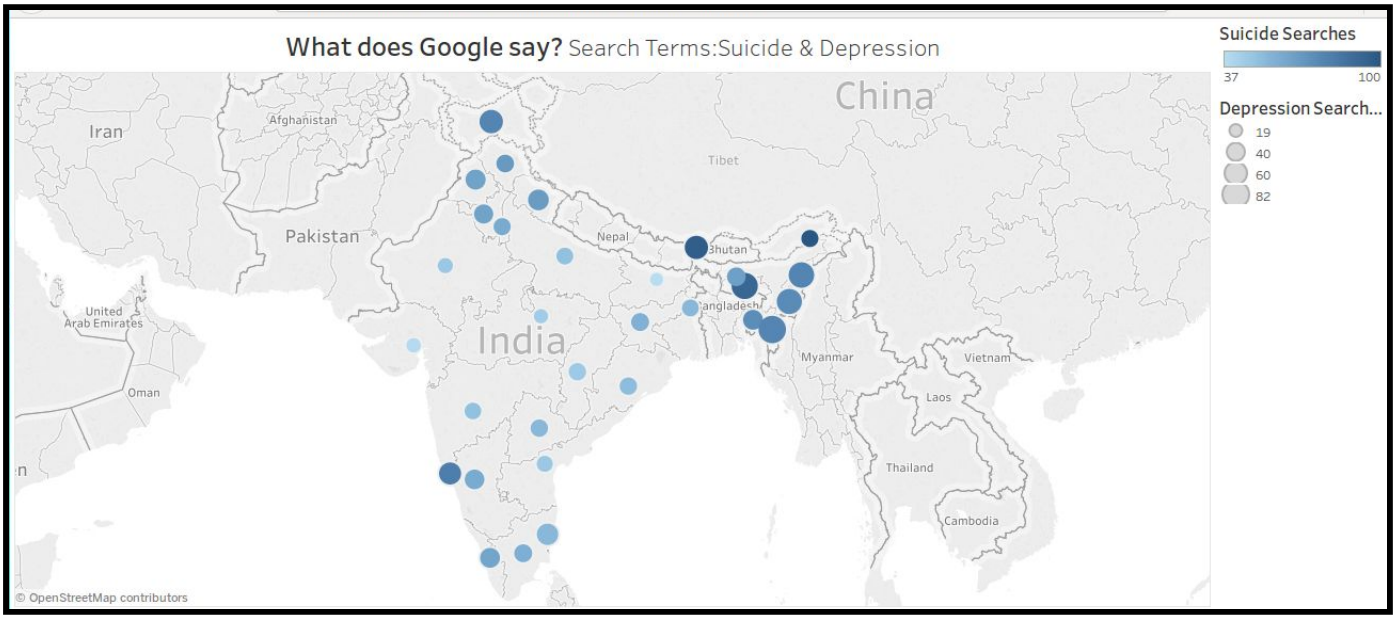
Reading the plot: This plot capture the total number of suicides vs suicide rate for different states and union territories of India for the years 2001 to 2011. The bars on the right indicate the total number of suicide deaths in India. Here, the size of the bar indicates the value.



Insights observed: As we can observe, the bar size is increasing with time. This indicates that the total number of suicides in India have been increasing. In the scatter plot, the bottom right quadrant identified as region of interest because here absolute number of suicides is less but the suicide rate is high. For example, Puducherry, A & N islands, Sikkim lie in this region.

Tools used: dimple.js for visualization, tableau for preprocessing and integration of different data source files.

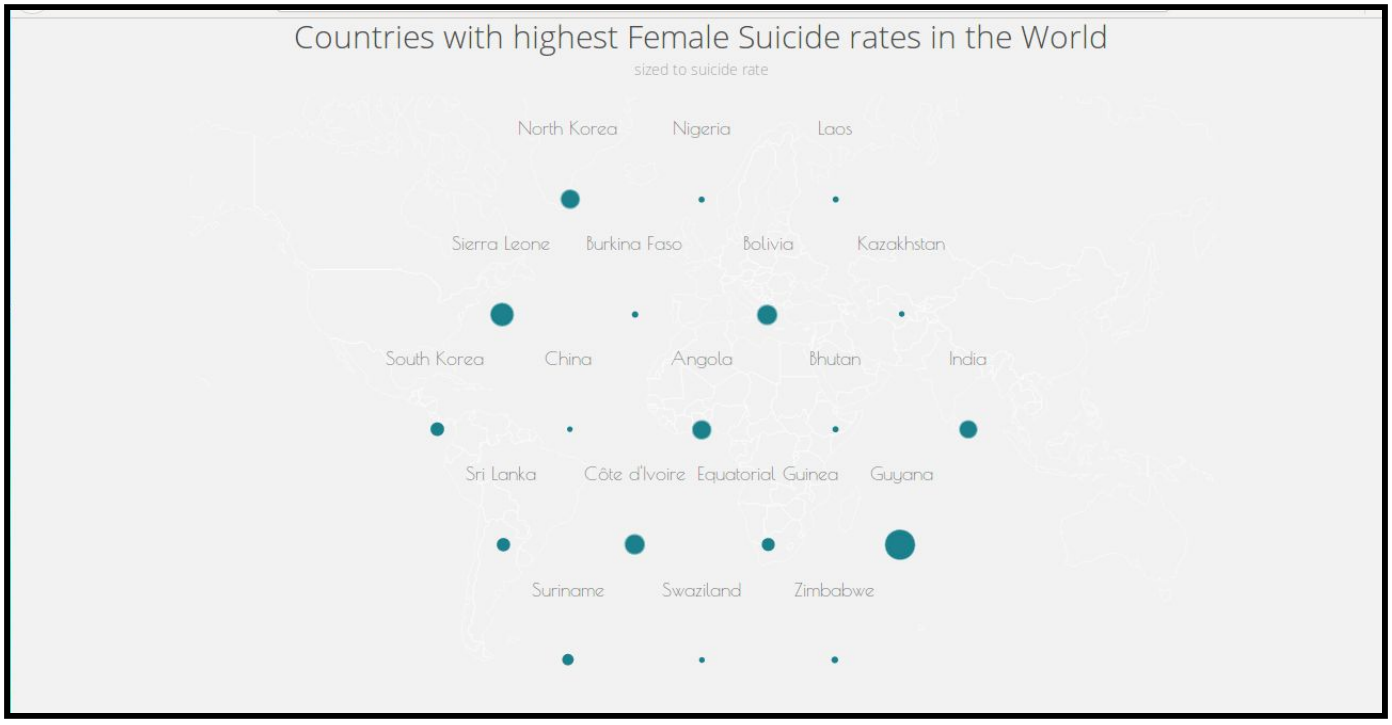
Reading this plot: In this plot, we can see the number of searches related to suicide and depression for different states. Here, the size of the dot indicates the popularity of depression search term and the darkness of the color indicates the popularity of suicide search term in a particular region.

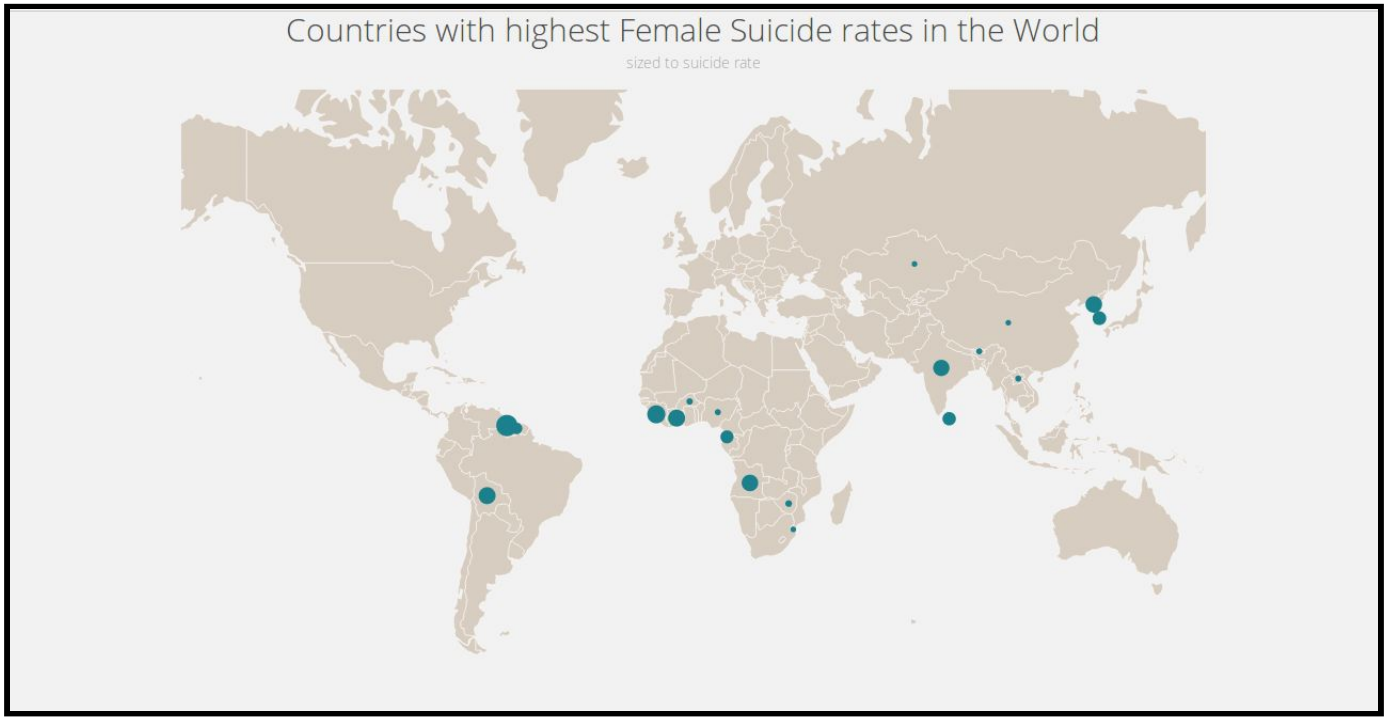


Insights observed: Surprisingly, the plot shows that the states in north-east states like Meghalaya, Mizoram, Manipur, Sikkim, Nagaland along with J&K and Goa search for depression way more than Maharashtra, AP, TN which show high suicide occurrence.

Tools used: Tableau for visualization

Reading the plot: This animated visualization shows the top 20 countries with highest female suicide rate. India ranks 7th in the same. The size of the blob indicates the suicide rate in that country.



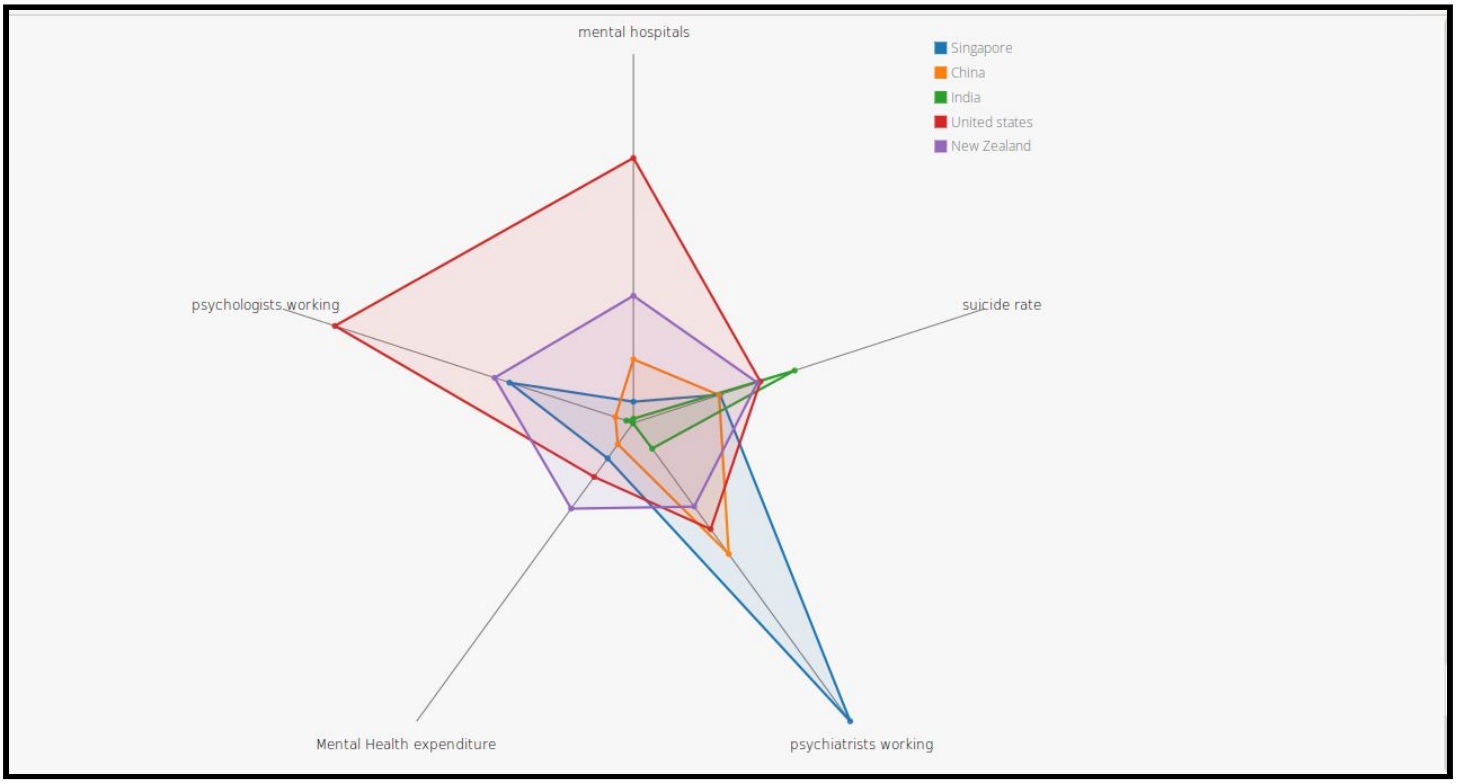


Tools used: d3.js for visualization

Reading the plot: This visualization compares the mental health infrastructure of countries with suicide rate lower than that of India. The data about these normalised parameters of various countries was acquired from the Global Health Observatory data repository of World Health Organisation (WHO). Different axes in the plot represent different dimensions

- number of mental hospitals per 100000
- psychiatrists working per 100000 people
- psychologists working per 100000 people
- mental health expenditure as percentage of GDP

of mental health infrastructure. Each polygon in the plot represent one country.



Insights observed: Other countries with lower suicide rate than India have significantly higher mental health expenditure, more psychiatrists & psychologists working and more mental health hospitals.

Government should invest more in mental health expenditure, Train more psychologists & psychiatrists, Build more mental health hospitals to reduce the suicide rate. Mental health education and resources for dealing with symptoms of mental health should be taught and promoted from an early age. Doing so will provide an avenue for the maintenance of dignity and help-seeking behavior

Tools used: d3.js for visualization

Limitations & Technical Challenges:

There is stigma associated with suicide in our society because of which many cases of suicide might not get reported. Hence, our results may not be accurate representation of what is happening in the country. The conclusions which we drew using google trends is applicable to only a subset of population (who are digitally literate, financially sound etc). Extrapolating those conclusions to entire population can be misleading.

Some technical challenges included absence of population data corresponding to different states for different demographics, which was needed for normalising the data. We had to find the population data from other sources and integrate with ours to get suicide rate etc. Also, additional scripts had to be written to integrate the data with D3 visualizations and external datasets like Global Health Observatory data repository of World Health Organisation (WHO). Converting the data in the format compatible for our visualizations required us to write scripts in R and python. While visualizing, one of the major challenges was to represent all the data point effectively: some states have very low suicide incidents as compared to other states. Coming up with the right scale for the plot was difficult.