**CSCE 5320 SCEINTIFIC DATA VISUALIZATION**

**Project: Analysis of Suicide Rates in India**

**Team members:**

Prathima Chowdary Sudalagunta-11662604

Prathyusha Jamapala-11648909

Manasa Karumanchi-

Siva Manikanta Lakumarapu-

**Goals and Objectives:**

## Motivation

To acquire an improved awareness of the trends and patterns relating to suicide in India through time, the Suicide in India dataset is being visually represented. The ability to visualize data allows us to spot trends that may not be immediately apparent when examining the raw data, such as whether suicide rates have been rising or falling over time, which states or populations have the highest rates of suicide, and whether there are any associations between suicide rates and other variables like age, gender, or educational attainment. Because they may present complicated information in a manner that is simple to comprehend and analyze, visualizations can also aid in the more effective dissemination of the data to a larger audience. Clear and eye-catching visualizations may be made to assist spread the word about the suicide problem in India and encourage dialogue and action to solve this crucial public health issue.

## Significance

**The statistics on suicide in India is important for a number of reasons**

First off, it offers insightful information about the suicide problem in India, which is a major public health issue. Suicide is a prominent cause of mortality for young people in India, which has one of the highest suicide rates in the world, according to the World Health Organization. We can better understand the causes of suicide in India and create preventative and intervention plans that are more successful by analyzing this information(Bhatnagar *et al.* 2021).

The second reason the Suicide in India dataset is important is that it gives a thorough and in-depth overview of suicide trends and patterns in India across various demographic groups and geographical locations. The dataset includes data on a variety of factors, including age, gender, education level, marital status, and employment, in addition to details on suicide reasons and methods. Researchers may do more thorough analysis and pinpoint particular risk factors for suicide among various populations because to this extensive information.

Last but not least, the Suicide in India dataset is important because it is openly accessible and may be utilised by academics, decision-makers, and activists to support action to solve the suicide epidemic in India. It can promote more openness and cooperation in our efforts to solve this significant public health concern by making this data available.

## Objective

The Suicide in India dataset seeks to provide a complete and in-depth picture of India's suicide trends and patterns. The dataset intends to record the prevalence of suicide across various demographic groups and geographical locations in India and to pinpoint the causes of suicide in the nation.

The Suicide in India dataset has the following precise goals:

* To monitor changes in India's suicide rates over time, across various racial and ethnic groupings, and in various geographical areas(Shaik *et al.* 2022).
* To determine the characteristics that affect suicide risk in India, such as age, gender, education level, marital status, and employment.
* To compile information on India's suicide rates, causes, and techniques.
* To advance a better understanding and awareness of the suicide problem in India and to inspire the creation of more efficient preventive and intervention techniques.

The Suicide in India dataset's overall goal is to provide a trustworthy and thorough source of data on suicide in India that can be used to drive research efforts, explain policy choices, and raise public awareness of this significant public health problem.

## 

## Feature

The Suicide in India dataset includes a number of factors that provide important new perspectives on the suicide problem in India. The dataset's primary characteristics include:

* Year: The dataset enables researchers to analyze changes in suicide rates over time since it has data on India's suicide rates for every year between 1967 and 2017.
* Gender: The dataset contains data on the gender of the suicide victim, which may be used to spot gender-specific trends and suicide risk factors.
* Age: The dataset contains information on the age of the suicide victim, which may be used to spot age-specific trends and suicide risk factors.
* State: The dataset contains data on the Indian state where the suicide happened, allowing researchers to spot geographic trends and suicide risk factors(Shaik *et al.* 2022).
* Education level: The dataset contains data on the educational background of the suicide victim, which may be used to analyze educational trends and suicide risk factors.
* Marital status: The dataset contains information on the marital status of the suicide victim, which may be utilized to uncover patterns particular to a certain marital status as well as suicide risk factors.
* Occupation: The dataset contains information on the occupation of the suicide victim, which may be used to spot trends and risk factors unique to that employment.
* Suicide method: The dataset contains information on the suicide technique, which may be used to find trends and risk factors connected to various suicide procedures.
* To determine the underlying causes and risk factors for suicide, it is possible to utilize the dataset's information on the cause of suicide.

Overall, the Suicide in India dataset's characteristics provide a comprehensive and complete picture of suicide trends and patterns in India. These characteristics may be utilized to pinpoint particular risk factors and create specialized prevention and intervention plans.

**For Increment 1:**

**Domain: Life:**

In general, those who are in need of assistance are those who belong to marginalised or vulnerable groups, such as those who are poor, subject to prejudice, or isolated from others. Lack of access to healthcare or education, discrimination or oppression, economic hardship, environmental catastrophes or crises, and social or political instability are just a few examples of the issues that people or communities may encounter. Depending on the individual problem that has to be solved, the timing will change. Others could be more recent in nature, while other issues might have existed for years or even decades(Shaik *et al.* 2022). Depending on the exact problem, the environment and place where it occurs might also differ greatly. For instance, a natural catastrophe could affect a certain area while a community or neighbourhood might suffer from economic hardship or social isolation(Bhaduri *et al.* 2020). Depending on the individual issue being addressed, the causes of a problem may be complicated, multidimensional, or a combination of these. Systemic inequality, political or economic unpredictability, a lack of infrastructure or resources, cultural attitudes and beliefs, or lack of resources are all potential causes of issues. In certain circumstances, the details offered in response to the other questions (what, when, where, and why) may be sufficient to answer the how question(Singh *et al.* 2021). Other times, a deeper study or inquiry may be necessary to determine how an issue came about.

**Data:**

To guarantee that any inferences made from the data are correct and representative, it is crucial to carefully analyze the sampling technique and the characteristics of the sample population. In order to preserve the privacy of the people who have their information included in the dataset, it is also crucial to make sure that any identifying information is correctly anonymized. The precise dataset being discussed will determine the answer to this question. To make sure that the variables in the dataset are relevant to the research topic being addressed, it is crucial to carefully assess their inclusion(Basu *et al.* 2022).

The GIS software has not been used here because there is no longitude latitude in the dataset. But in the dataset the state and the suicide data is covered so the R studio software has been used to visualize the data.

**Task Abstraction(Target and Actions):**

**Users:**

We are analyzing the suicide data in India over the year 2001 to 2012 which clearly shows the analysis of types of suicides, reasons for the suicides, which age category of people are committing what type of suicides and which gender people are committing what kind of suicide. So, this project helps the user to analyze the reasons for the problem which makes it easier to predict the solution for the problem as well.

Not only for the user, this visualizations on suicide data helps the government to identify, analyze and take measures to reduce or prevent t the problem. By viewing the visualizations more interactively through our project, government will be able to identify the reasons for committing suicides, what age group of people are committing more suicides, which state or province has more suicide rates or which gender people are committing what kind of suicides. This analysis helps the government to take measures and implement actions to reduce the suicide rate in India. Our system shows a clear and thorough visualizations of all the data regarding suicides which makes it easier to view and understand the data more accurately.

**Implementation using tools:**

* **Describing the tools:**

**Python:**

Python is the general purpose programming language which is commonly used for visualizations.

**R language:**

R language is used for statistical computing and graphics. Our dataset contains the numerical data so we prefer this language to analyze the data which provides the best results in representing the statistical data.

**Rstudio:**

It is used in data analysis to import access ,transform , explore ,plot and model data

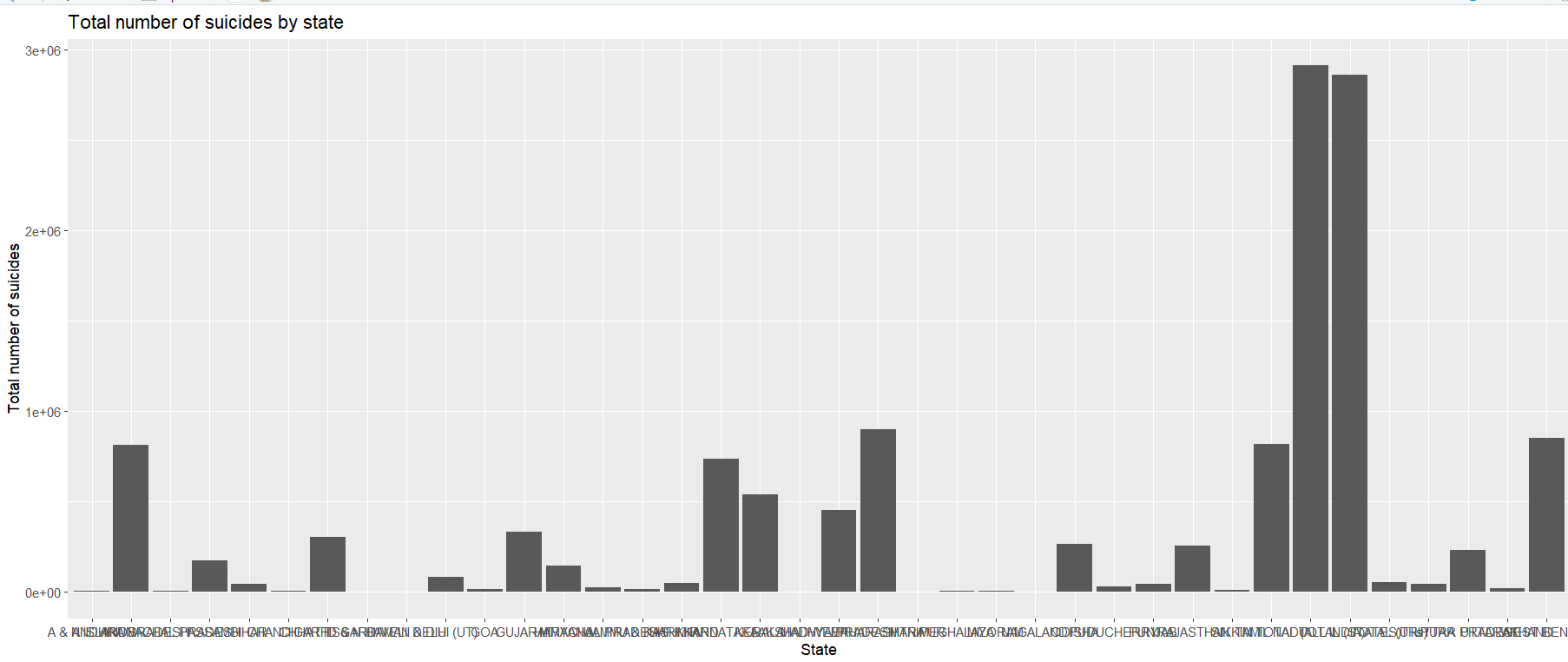
in order to make predictions on the data**.**

**Preliminary results for Analysis:**

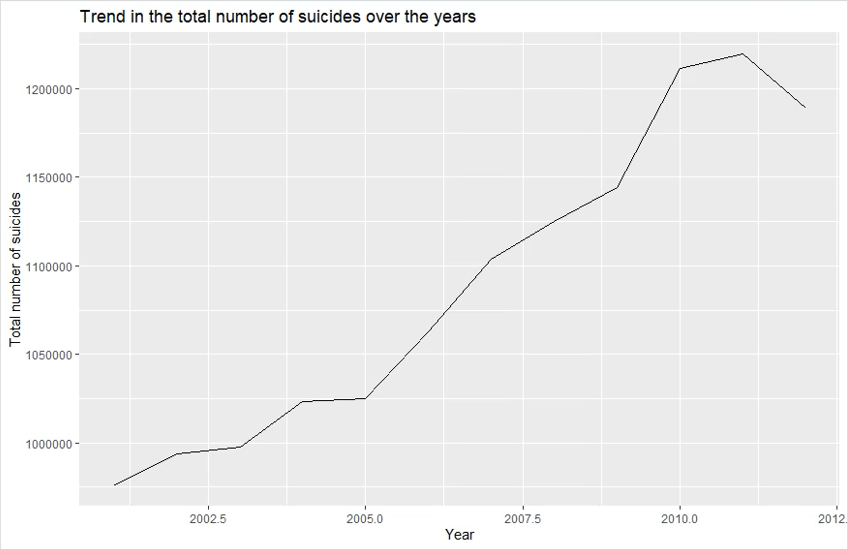
The following are the visualizations which shows the analysis of different aspects in related to the suicide data in India.

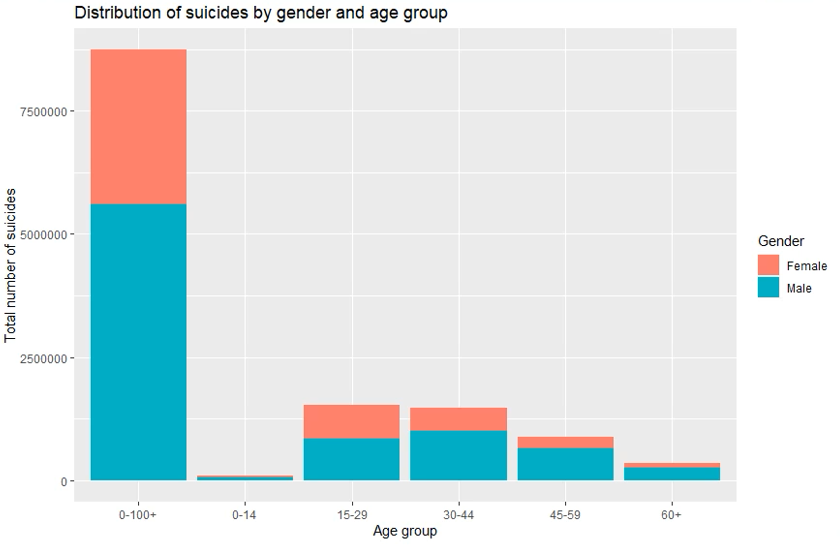
* **Visualization graphs:**

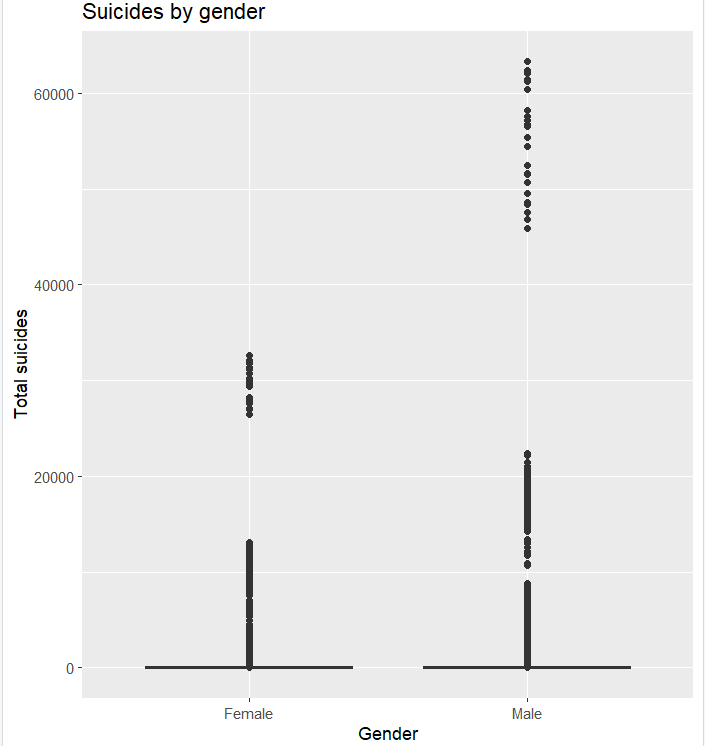
The first visualization is the bar chart visualization which shows the total no of suicide rate in each state in India.

****

The second visualization is the line chart visualization which shows the trends in total numbers of suicides over the years in India.





****

**Work completed:**

**Description:**

For this increment we have showed four visualizations namely, bar graph, line chart, stacked bar chart and box plot which are related to the suicides in India .These four visualizations helps us to analyze the data in four different aspects related to the suicides rate in India.

Through bar graph, we show the total number of suicides in each state over the years. By viewing at the bar graph, it is clear to understand which state has highest suicide rate and which state has the lowest.

Through line chart,we show the total number of suicides over the years. By the graph,we can see that the total number of suicides are increased drastically over the year until the year 2011 but decreased slightly in the year 2012.

Our next visualization is a stacked bar chart which shows the suicide rates of both sexes individually over the years of different age groups. By viewing the stacked bar chart it us clear that suicide rate in male is highest than females and the 15-29 age group has highest suicide rates.

Our last visualization is box plot which shows the total suicides of each sexes individually and it clearly shows that male has highest suicide rate over the years than females.

Responsibility:

Prathima:

For this increment I helped my team to visualize the data with different attributes. I also helped in identifying the attributes that need to be represented in different visualizations.

Prathyusha:

My task is to help the the team in providing ideas to develop different types of visualizations through which data can be analyzed more clearly and thoroughly.

Manasa: My task is search for the data and create visualizations for the attributes accordingly by doing research about tools and libraries that need to be used to provide interactive visualizations.

Siva Manikanta: My task is to help with data modifications, and also research about interactions that needed to be provided while users are on the web page.

**Work to be completed:**

**Description:**

For this increment we have showed four visualizations which are related to the suicides in India .These four visualizations helps us to analyze the data in four different aspects related to the suicides rate in India.

For the next increment we want to show some more visualizations possible for the data which helps the users to explore more and analyze the data more.On top of that we also want to make some of the visualizations more interactive so that user can feel comfortable while analyzing the data.

Finally we want to make the web page which is interactive that depicts all the visualizations at once in a single web page that provides the complete information about the suicides rates in India.

The above are assumptions which we want to fulfill for the next increment.

Responsibility:

Prathima:

For the next increment I want to help the team in making the interactive web page that shows all the visualizations.Currently I am learning and working on developing the web page .

Prathyusha:

My task is to help the the team in providing ideas to add more visualizations and add more interactive visualizations into the web page.

Manasa:

My task is search the data and create the web page to view the data interactively by the user and make them understand and analyze the data thoroughly.

Siva Manikanta: My task is to help with data modifications, and also research about interactions that needed to be provided while users are on the web page.

Issues and Concerns:

Since,we are trying to encode all these visualizations in web page, our main concern is whether we can fit these visualizations into the web page. Also, we want to include more interactive visualizations into our web page in the next increment which is our main concern about the next increment. We are currently working on designing the web page.

**References:**

https://www.medrxiv.org/content/10.1101/2020.05.25.20113043v1.full

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7523173/https://europepmc.org/article/ppr/ppr488560

https://cegh.net/article/S2213-3984%2820%2930226-8/fulltext

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7523173/