US Suicidal Rates Visualization

Team Members: Xujun Lian (xl2595), Tianyu Zhang (tz1180), Peiyang Wen (pw1179)

1. Overview

1.1 Summary of our project goals

Because of the development of society and capitalism, people nowadays are facing more stress than before. Therefore, the government and psychologists started to focus on people's mental health. Among those factors that can reflect the psychological conditions of people in society, we choose to visualize the suicide rate. In our visualization system, we would like to present the suicide rate in different states of different periods in the United States of America. By doing this, we can study people in which areas tend to have more severe mental problems.

1.2 Screenshot of the visualization

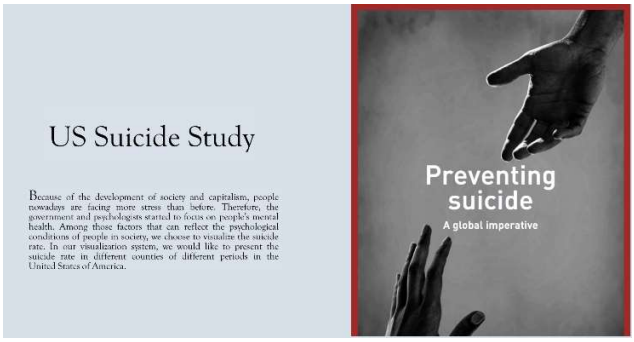


Figure 1: The homepage

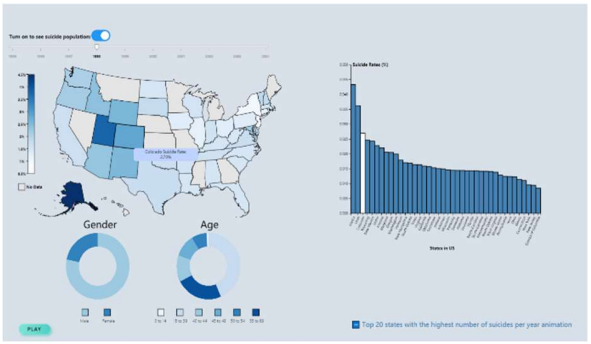


Figure 2: The main page of our visualization

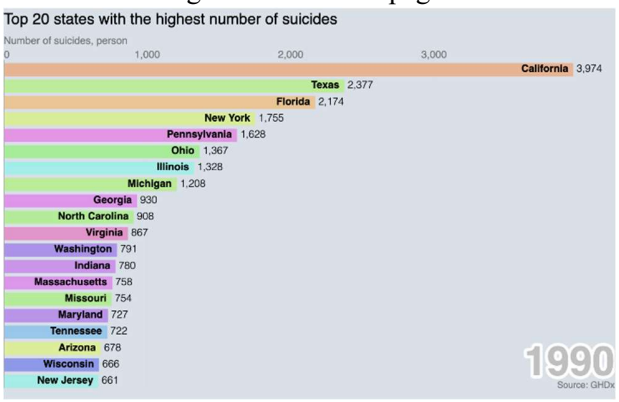


Figure 3: The bar chart race

2. Data

- **Name:** The United States Suicidal Rates Dataset
- **Link:** http://ghdx.healthdata.org/gbd-results-tool
- **Description:**
  The dataset provides suicidal rates and population in each state of the United States from 1990 to 2018. Detailed information on individuals' age, gender, and the cause of suicidal behavior are also available.
- **Attributes Used in Visualization:**

Attribute	Description	Scale and cardinality
Year	The year when data was collected	Ordinal data from 1990 to 2018
location_name	State name in the United States	Categorical data with 52 levels
sex_name and id	1 means male, 2 means female, 3 means both	Categorical data with 3 levels
age_name and id	Different age groups	Categorical data with 6 levels
val	Number of suicide cases or rate	Quantitative data with different range between number or rate
metric_id	1 means the number of suicides, 2 means rate	Categorical data with 2 levels
other attributes	Will not be used in data processing	

- **Data Pre-processing:**
  The raw data set contains all the information we need, which helps us a lot when we process the data. Since everyone in our team is responsible for a unique part of the visualization, so one may only need data with certain attributes while other members may not use them. The original dataset includes all the attributes and each of us selects data one needs and saves it as a new file. This helps us to process our data efficiently.

3. Goals and tasks

● Domain language:

Task1: We want to know the geometrical distribution of suicide rates or suicide numbers of different states in the US in a certain year.

Task2: We want to know the suicide rate or suicide numbers of a certain state in a certain year.

Task3: We want to know the gender and age breakdown of suicide cases of a certain state in a certain year.

Task4: We want to know how the suicide numbers of different states change over time

●Abstract language

Task1: Action: Explore in the search category. Target: all data about suicide rate or suicide numbers of different states in the US in a certain year.

Task2: Action: Lookup in the search category. Target: attributes including suicide rate or suicide numbers of a certain state in the US in a certain year.

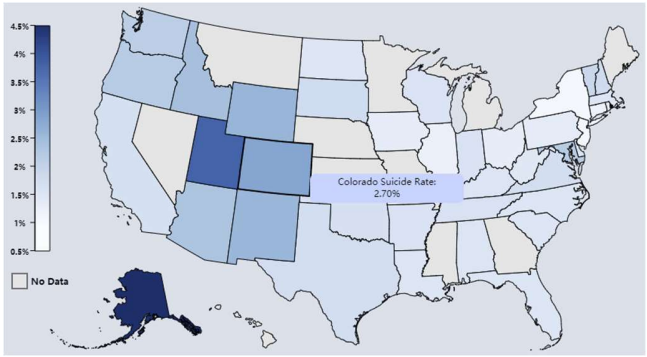
Task3: Action: Lookup in the search category. Target: the distribution of attributes including gender and age.

Task4: Action: Present in the analyze category. Target: the trends of suicide numbers of different states change over time.

4. Visualization & Sketch

In our project, we expect the visualization to provide users an interactive and sophisticated view into the US suicidal phenomenon. To achieve this, we designed three different views and two distinct UI widgets as our must-have features. Besides, we design a 'play' widget to play data from 1995 to 2004 automatically, which allows users to see how suicide numbers or rate changes from year to year. A bar chart race animation is also designed to show the top 20 states with the highest number of suicides in different years.

- **View 1: Heated Map: Suicidal Rates or Numbers of the United States**



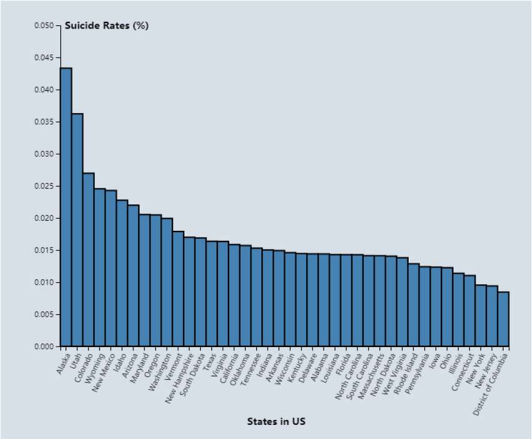
This is the main view to provide users an instant understanding of the US suicidal phenomenon. The area of each state in the map will be colored with different degrees of saturation to show the level of suicidal rates in that state in a selected year. The darker color represents higher suicidal rates/population in the corresponding state.

We use an area mark to visualize the geometrical location of each state in the US. Channel 1: Color Saturation showing the size of suicide rates/population. The geometrical shape colored with different saturation levels enables the users to recognize how the suicide phenomenon was distributed across the country. Besides, the state with the highest/lowest suicide rates/population could be easily identified.

When users hover on a state, the state will be highlighted with a thicker stroke. A tooltip describing the state's name and suicide number will be presented at the same time. Since this view is linked with the other two views, the mouseover event will also generate two corresponding donut charts (age & gender) and a highlight effect in the bar-chart.

- **View 2: Bar Chart: Suicidal Rates or Numbers of each State in a Selected Year**

In this view, a bar-chart will reveal and compare the suicidal rates/numbers of each state in a selected year. The x-axis is “state name” and the y-axis is “suicidal rates/numbers”. The bars will be presented in descending order so that users can identify the state with the highest and lowest suicidal rates/numbers and compare between them.

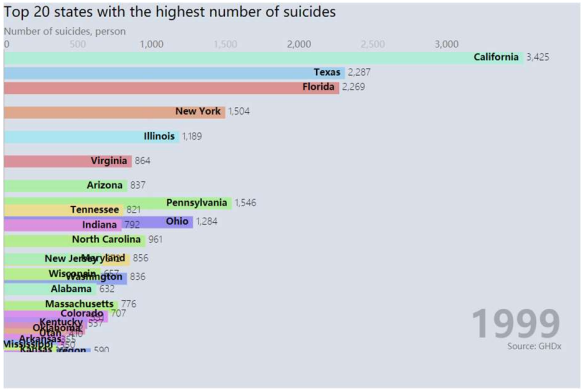


We use a line mark to represent the suicide rate or number of a state. And we encode the mark by 2 channels. Channel 1: length of the line showing the amount of rates or numbers. Channel 2: horizontal position showing the rank of rates or numbers of a state. It is reasonable to use this view as it makes the difference between suicide rates or numbers in different states clearer.

When moving your mouse on a certain bar, it will be highlighted, and the corresponding state will also be highlighted in the map view. The gender and age breakdown of this year will be shown.

- **View 3: Bar Chart Race: Top 20 States with the Highest Number of Suicides in Different Years**

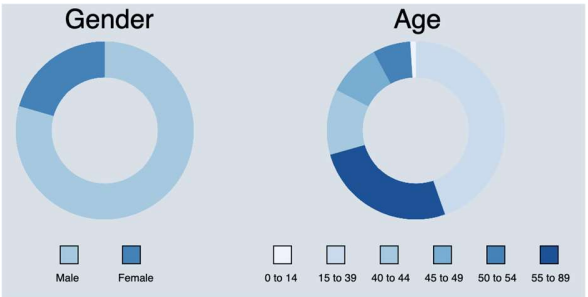
This is an interesting function we designed in our visualization as an extension of the bar chart view. Since the bar chart is static. A bar chart race is a nice solution to show the data dynamically. The bar chart race will be played automatically after entering its web page. Please notice that there will be about 10 – 20 seconds delay to make sure our user will not miss the whole animation.



We use a line mark to represent the suicide number of a state. And we encode the mark by 3 channels. Channel 1: horizontal position showing suicide numbers. Channel 2: length of the line showing the amount of numbers. Channel 3: color hue used to distinguish a state from others.

- **View 4: Donut Chart: Age & Gender Composition of a State in a Selected Year**

This view will only appear once the users select a state by moving the cursor towards a state on the map or a bar in the bar chart. Two donut charts will be presented respectively to show the age and gender composition in that state. Through this view, users can clearly identify which age group or gender accounts for the majority of suicide population in that state in a selected year.



We use the area mark to represent the suicide proportion of each gender or age group for each state in a selected year. Channel 1: angle showing the proportion. Channel 2: color hue showing different gender and age groups. Angle enables users to have an intuitive feeling of age and gender components and color hue helps to distinguish different groups. Besides, we set up color for each age group from young to old according to color saturation which enables users to identify each group more conveniently.

- **UI Widget 1: Suicidal Rate-Population Switcher** 

This widget enables the users to switch the suicidal data used in view1 and view2 between the rate and actual population. The default map and bar-chart will be presented using suicidal rates data. Once the users click on the switcher, the data used in these two views will be switched to the actual suicidal population and the views will change at the same time. Considering that some states with lower rates could actually have higher suicidal populations due to the higher size of the total population, we give users the option to directly see the population number instead of rates. Both rates and population data provide users with meaningful but distinct insights.

- **UI Widget 2: Year Slider** 

This widget enables users to change between years and observe the changes happening over time. Once the slider is moved, views in view1, view2 will change correspondingly and view3 will be temporarily eliminated until users click on a new state. The default year is set as 1998.

- **UI Widget 3: Play Button** 

This widget enables the views to be automatically played respectively from 1995 to 2004. Once users click on the "PLAY" button, this displaying procedure will automatically start. They can flexibly halt and continue the displaying procedure by alternatively clicking on the "PAUSE" and "PLAY" button. In the playing mode, the year slider will be temporarily hidden and the current year number will be presented to the users on the top left corner. Through this widget, users can observe the changes of suicide distribution across the country over time.

5 Reflection

We assigned each of our team members a view of our proposal. So according to it, we first did some research on how to realize certain visualizations based on d3.js and started our coding. After all three main views had already been realized, we combined our codes and functions to make them as a whole. Then we linked different views and implemented the two widgets. Then we had a rough visualization with no animation, but all the functions could work well. The next step was to add transition animation, to beautify the interface, and to add some functions that are extra but interesting. We added a bar chart race and another play widget to add more interactions and functions to our visualization.

In terms of visualization goals, we have mainly five changes. Firstly, we shrink the time scale of data visualization. The original dataset has a collection of data from 1990 to 2018 while the problem of missing values in the first and last few years are relatively serious, so we change the time scale from 1995 to 2004. Secondly, even if we reset the time scale, the problem of missing value is still inevitable, so we use grey to represent states with no data on the map. We also add an interaction to highlight all the states with no data on the map by pointing the cursor on the 'no data' legend. Thirdly, we use donut charts to display age and gender components instead of the pie charts that we initially planned. This is because the donut chart is not only more aesthetic but also makes the difference of components between each state more distinguished as it lets users only focus on the outer edge of the circle. Fourthly, we make donut charts displayed by pointing the cursor to a certain state on the map instead of clicking the state. In this way, donut charts are only shown when users move the cursor on the state and are removed automatically when users move the cursor out of the state, which is more convenient for users. Fifthly, we add a front page to provide the background information of our project.

In terms of technical goals, we didn't change much from our initial design. We mainly used Javascript and d3 version 6 as our coding language to process data and design the views. CSS and HTML help us to beautify our interface. But there is one change from our initial design when designing the bar chart race. We did not find a tutorial based on d3v6, but there are a few based on d3v5. So we used d3 v5 to code the bar chart race.