Jai Reddy Kukunuru

Geospatial Analysis of Suicide Trends in the United States: An Integrated Multi-View Visualization Study

Suicide, a public health crisis of alarming proportions in the United States, is a phenomenon greatly influenced by demographic and locational variables. Scrutinizing suicide rates across diverse demographic categories—sex, race, Hispanic origin, and age—illuminates the varying susceptibility across these groups. Additionally, geospatial patterns could be equally revealing, pointing towards socio-economic, cultural, or access-to-care influences that might be locally endemic or more prevalent in certain areas.

Integrating location data within this topic's analysis greatly augments our understanding of suicide trends, by bringing into focus the geographical dispersion and variances. States and regions with high suicide rates could be investigated for underlying commonalities. For instance, rural regions may exhibit higher suicide rates due to isolation, limited mental health services, or socio-economic struggles. Contrarily, urban areas might reflect a different set of stressors or protective factors. This geographical layering thereby helps in addressing the "how" of these patterns: how locational and demographic variables interplay in the manifestation of suicide rates.

Visualizing this geospatial data, along with demographic categorizations, allows us to see the correlations between location and suicide rates. It helps us comprehend the spatial distribution of suicide rates and to identify areas that need urgent interventions. The disparities in suicide rates across different geographical locations and demographic groups become readily apparent through such multi-layered visualization, leading to more targeted and effective prevention strategies.

As our society increasingly relies on data for understanding complex social issues, it's crucial that we harness the power of visualizations to present that data in accessible and meaningful ways. This is particularly the case when it comes to analyzing the troubling trend of suicide

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

rates in the United States. By utilizing geospatial visualizations, we can depict a clearer picture of how this public health crisis is distributed across different demographics and regions. Here we explore three geospatial visualization techniques: Choropleth Maps, Heat Maps, and Hexagonal Binning, all of which can significantly aid in the understanding and prevention of suicide trends.

Description of the comparison for 3 techniques:

Choropleth Maps: are widely used to represent geographically distributed data. In this map, geographical regions are filled with colors or patterns according to the variable's value in that area. For example, in a Choropleth Map of suicide rates in the U.S., each state is colored according to its average suicide rate over a certain period. Darker colors might represent higher rates, and lighter ones, lower rates. This approach allows the viewer to immediately discern spatial patterns and identify areas of concern. However, one limitation of Choropleth Maps is that they don't inherently account for population density, meaning a large state with a small population might overshadow a smaller but more populous state with a similar suicide rate.

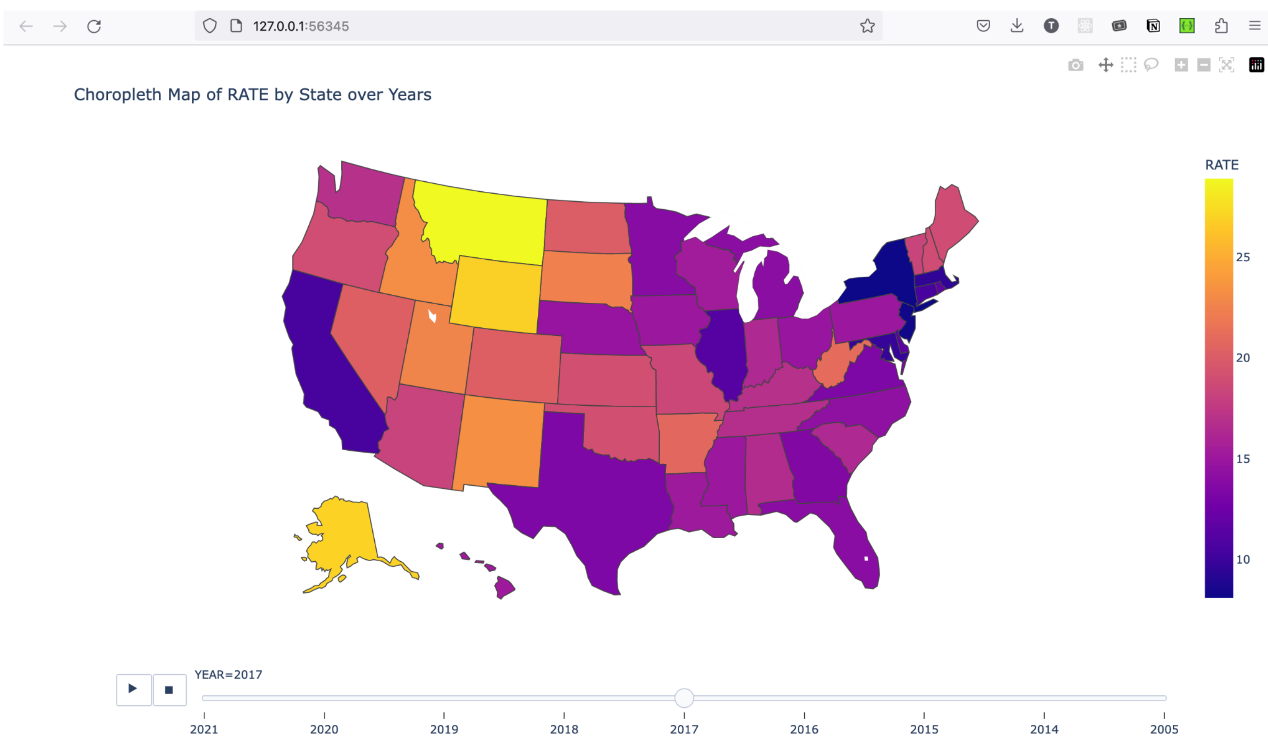
Heat Maps:, another type of geospatial visualization, use color intensity to represent the value of the variable in each area, similar to Choropleth Maps. However, instead of representing distinct geographical regions, Heat Maps often represent data on a grid where each cell's color indicates the value of the variable in that cell. Applied to the study of suicide trends, a Heat Map could provide a granular view of suicide rates, showing variations within states or cities. For instance, a Heat Map of a city could indicate suicide rates across different neighborhoods, highlighting areas that need more targeted prevention efforts. Nevertheless, one limitation of Heat Maps is that they may lack geographical context unless supplemented with additional map layers or geographical annotations.

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

Jai Reddy Kukunuru

Hexagonal Binning: is a technique used when dealing with large datasets, where data points are aggregated into hexagonal 'bins.' Each hexagonal bin encompasses a geographical area and contains the aggregated data of that area, with color used to represent the data's value. In the context of suicide rates, Hexagonal Binning could aggregate data for regions within states or counties. For instance, we could divide a state into hexagonal bins, with each bin representing the average suicide rate in that area. The result is a geospatial visualization that can represent detailed, local patterns within larger regions without being overwhelmed by individual data points. However, the granularity and accuracy of Hexagonal Binning maps depend on the chosen bin size, as larger bins could obscure local variations while smaller bins could appear noisy or erratic if data are sparse.

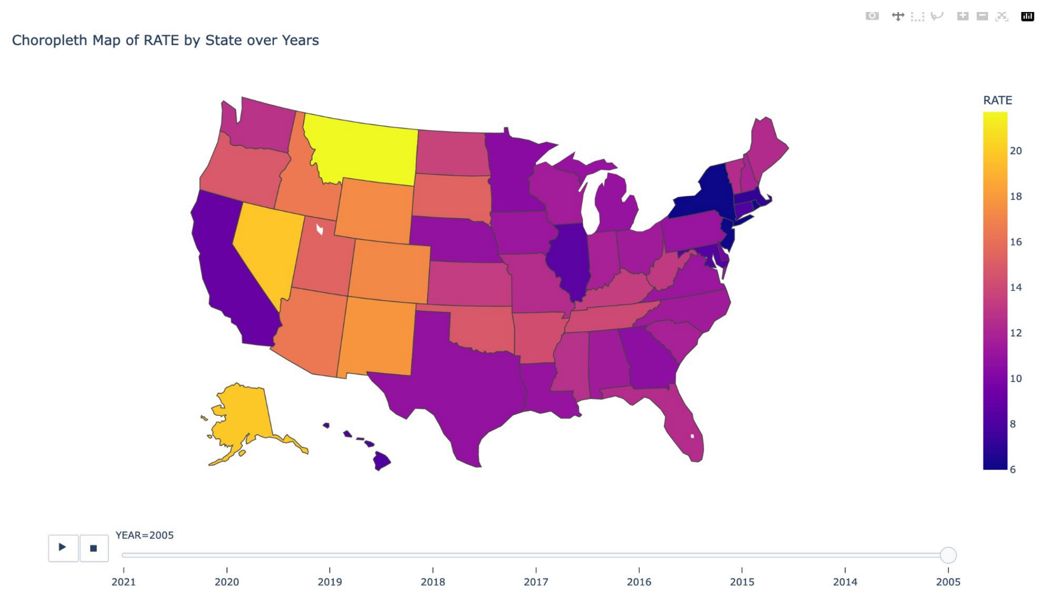
Chloropleth map by state for the year 2017 generated using plotly package in python:



Cholopleth map for the year 2005 statistics below:

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

Jai Reddy Kukunuru



HexagonBin Plot below for Suicide rate between 2006 and 2020:

Multiview Visualization :

Motivations: We aim to offer a holistic perspective of the suicide rates, allowing users to easily discern trends, disparities, correlations, and anomalies. This multi-dimensional issue cannot be fully grasped with a single visualization type. A combined approach will

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

present insights from different angles, supporting a more in-depth understanding of the trends, which can help policymakers, researchers, and community leaders devise better prevention strategies.

In the multiview visualization, the "what," "why," “when”, and "where" components interact with each other to provide a comprehensive understanding of the dataset. Let's discuss how these components are interconnected:

<u>What Component (Bar Chart):</u>

- The bar chart represents the "what" component of the visualization.
- It displays the death rates per state for the latest available year.

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

- Each bar represents a state, and the height of the bar represents the death rate.

- The states are sorted in descending order based on their death rates.

- This component provides an overview of the variation in death rates across different states.

Why Component (Line Chart):

- The line chart represents the "why" component of the visualization.

- It shows the temporal trend in death rates for the state with the highest number of deaths.

- The x-axis represents the years, and the y-axis represents the death rates.

- By analyzing the line chart, we can observe how the death rates have changed over time for the state with the maximum number of deaths.

- This component helps in identifying patterns, trends, or fluctuations in death rates over the years.

When Component (Line Chart):

- The bar graph represents the "when" component of the visualization.

- It shows the trend in death rates for the state with the highest number of deaths across various year.

- The x-axis represents the years, and the y-axis represents the death rates.

- By analyzing the bar graph, we can observe how the death rates have changed over time for the state with the maximum number of deaths.

Reference:

https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

- This component helps in identifying patterns, trends, or fluctuations in death rates over the years.

Where Component (Geospatial Map):

- The geospatial map represents the "where" component of the visualization.
- It visualizes the geographical distribution of the states using markers on the map.
- Each marker represents a state, and the size of the marker does not hold any significance.
- By hovering over the markers, you can see the state name as a popup.
- This component helps in understanding the spatial distribution of the states and their positions relative to each other.

The interaction between these components allows for a holistic exploration of the dataset. By examining the bar chart, we can identify states with higher or lower death rates. The line chart provides insights into the temporal patterns of the state with the maximum deaths. The geospatial map complements this information by visually representing the states' locations.

Analyzing the three components together enables us to identify states with high death rates, understand their temporal trends, and visualize their geographical positions. This comprehensive approach facilitates the exploration and understanding of the dataset from different perspectives, contributing to a more informed analysis of the data.

Reference:

https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

Jai Reddy Kukunuru

To create the multiview visualization, we used the following libraries and implemented the steps outlined below:

1. **Pandas**: Pandas library was used to load and manipulate the dataset.
2. **Plotly Express**: Plotly Express was used for creating interactive visualizations including bar charts, line charts, and geospatial maps.
3. **Dash**: Dash was used to create the web application framework for displaying the multiview visualization.

Below are the steps for creating the multiview visualization:

Loading the Data: The dataset was loaded using Pandas. The CSV file containing the data was read into a Pandas DataFrame.

Data Preprocessing: Data preprocessing steps were applied to the DataFrame as needed. This included converting data types, handling missing values, and transforming the data for specific visualizations.

Bar Chart (What Component): For the "what" component, a bar chart was created using Plotly Express. The DataFrame was grouped by the 'FullGeoName' column and the sum of 'DEATHS' for each state was calculated. The resulting DataFrame was used to create the bar chart with the 'FullGeoName' column as the x-axis and the 'DEATHS' column as the y-axis.

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

Line Chart (Why Component): For the "why" component, a line chart was created using Plotly Express. The DataFrame was filtered for a specific state (e.g., 'California') and grouped by the 'YEAR' column. The sum of 'DEATHS' for each year was calculated, resulting in a DataFrame representing the trend over the years. This DataFrame was used to create the line chart with the 'YEAR' column as the x-axis and the 'DEATHS' column as the y-axis.

Geospatial Map (Where Component): For the "where" component, a geospatial map was created using Plotly Express. The latitude and longitude coordinates of each state were obtained from a state-to-lat/lon dictionary. The DataFrame was used to create a choropleth map with the 'FullGeoName' column as the location, the 'DEATHS' column as the color scale, and the 'YEAR' column as the animation frame.

Layout and Styling: The Dash library was used to create a web application layout for displaying the multiview visualization. HTML and CSS were used to structure and style the different components of the layout, including the bar chart, line chart, and geospatial map. The components were arranged in a grid layout to ensure they are displayed side by side.

Interactive Functionality: Interactive functionality was implemented using Dash. For example, selecting a state in the line chart dynamically updates the geospatial map to show the selected state's data. Similarly, other interactive features such as hover tooltips, zoom, and pan were added to enhance the user experience.

Running the Application: The Dash application was launched, and the multiview visualization was displayed in a web browser. Users can interact with the visualizations and explore the data through the interactive features provided.

Reference:

https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

By leveraging the capabilities of Pandas, Plotly Express, and Dash libraries, the multiview visualization was created to provide an intuitive and interactive representation of the "what," "why," and "where" components of the dataset.

DESIGN PRINCIPLES

When creating the multiview visualization map, several design principles were employed to enhance its effectiveness and usability. Here are the design principles incorporated:

1. **Consistency:** The design elements, such as color schemes, fonts, and layout, are kept consistent across different views (bar chart, line chart, geospatial map). This ensures a cohesive visual experience and allows users to easily transition between different components without confusion.

2. **Clarity and Simplicity:** The visualization maintains a clean and uncluttered appearance. The use of clear labels, legends, and tooltips helps users understand the information presented. Unnecessary details or distractions are minimized to focus on the key insights.

3. **Visual Hierarchy:** The design employs visual hierarchy to guide users' attention to the most important information. The bar chart is given prominence as the primary focus of the "what" component, followed by the line chart for the "why" component, and the geospatial map for the "where" component. The use of color, size, and position emphasizes the hierarchy.

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

4. **Color Encoding:** Color is used effectively to encode information and create visual associations. In the bar chart, color is employed to distinguish different states or categories. In the line chart, a contrasting color is used for the line to make it stand out. In the geospatial map, a color palette is used to differentiate states or highlight specific data values.

5. **Interactivity:** The visualization incorporates interactive features to engage users and allow for exploration. Hover tooltips in the geospatial map provide additional information, and the ability to zoom and pan enables users to interact with the map. The line chart allows users to select different states and dynamically update the visualization.

6. **Responsive Design**: The visualization is designed to be responsive and adapt to different screen sizes or devices. This ensures that users can view and interact with the multiview visualization effectively, regardless of their device or screen size.

By adhering to these design principles, the multiview visualization map aims to enhance user understanding, facilitate exploration, and provide an engaging and informative experience.

Reference:
https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1