Visualization of Complex Data

DATS 6401

Final Term Project- Proposal

AVANI RAO
G45297503

This dataset is particularly interesting for visualization due to its comprehensive coverage of health-related factors, including lifestyle behaviors, physical health metrics, and medical history. The ability to visualize such data can reveal critical insights into public health trends, identify potential risk factors for heart disease, and uncover the effects of various lifestyle choices on overall well-being. The inclusion of both numerical and categorical variables allows for a multifaceted analysis, enabling the exploration of complex relationships within the data.

The selected dataset for the term project is from a public health survey conducted in 2022, focusing on heart health and related lifestyle factors. It comprises over 50,000 observations (exact count=4,45,133 rows), with a rich mix of numerical and categorical data. Numerical data includes variables like PhysicalHealthDays, MentalHealthDays, SleepHours, and Body Mass Index (BMI), while categorical data encompasses State, Sex, GeneralHealth status, and various health behaviors and conditions (e.g., PhysicalActivities, AlcoholDrinkers, HadHeartAttack). The dataset contains a wide range of variables related to health, including both categorical (e.g., State, Sex, GeneralHealth) and numerical data (e.g., PhysicalHealthDays, MentalHealthDays, SleepHours). It spans various aspects of health and lifestyle, such as general health status, physical activities, sleep hours, and medical history (e.g., heart attack, stroke), along with demographic information like state and sex.

Important Features for Static Plots

For static plots, the following features and plot types are essential to revealing underlying information:

- GeneralHealth Status (Categorical): Bar charts to show the distribution of general health perceptions among the population.
- PhysicalHealthDays and MentalHealthDays (Numerical): Histograms and box plots to visualize the distribution of days people report being in poor physical or mental health.
- SleepHours (Numerical): Density plots to examine the distribution of sleep hours and its correlation with health outcomes.
- BMI (Numerical): Scatter plots to explore the relationship between BMI and health days, highlighting potential obesity-related health issues.
- State (Categorical): Bar charts or pie charts to display health metrics variations across different states, identifying regions with higher incidences of health issues.
- HadHeartAttack (Categorical): Bar charts to compare the proportion of individuals who have had a heart attack across different demographic groups.

Features for Interactive Plots

For the interactive dashboard, leveraging dynamic visualizations will enhance user engagement and provide deeper insights:

- Interactive Maps: Utilize the State variable to create interactive maps that show health metrics by region, allowing users to explore geographical health trends.
- Dynamic Scatter Plots: Offer users the ability to select variables (e.g., BMI vs. SleepHours) and demographic filters (e.g., Sex, Age) to investigate relationships between different health indicators.
- Time Filters for Trends Analysis: If temporal data can be derived through feature engineering or additional data linkage, include time sliders to allow users to observe changes over time in health indicators or outcomes.
- Customizable Histograms and Box Plots: Enable users to select specific variables (e.g., PhysicalHealthDays) and demographic segments (e.g., by GeneralHealth status or Sex) for customized distribution visualizations.

Dashboard Usability Plan

To ensure the dashboard is practical and user-friendly, the following elements can be incorporated:

- Intuitive Navigation: A clear, straightforward layout with a sidebar for selecting visualizations, filters, and settings.
- Responsive Design: Ensure the dashboard is accessible and performs well on a variety of devices, including desktops, tablets, and smartphones.
- Interactive Elements: Sliders, dropdown menus, and checkboxes to allow users to dynamically adjust the data displayed according to specific criteria or interests.
- Informative Tooltips: Hover-over tooltips and click-based pop-ups that provide definitions, additional information, or insights about the data points and variables to educate and inform users.
- Data Filtering and Comparison Features: Facilitate direct comparisons between different demographic groups, states, or health metrics, enhancing the analytical value of the visualizations.

This approach aims to make the dashboard not only a powerful tool for data exploration and analysis but also an engaging platform for users to learn about public health trends and how various factors impact heart health and overall well-being.