**REPORT**

**VISUALIZATION PROJECT**

**Abstract:**

The NCD Portal Diseases Data Set is a comprehensive collection of statistical data related to noncommunicable diseases (NCDs) from countries around the world. It encompasses a wide range of factors such as tobacco use, physical inactivity, harmful alcohol use, diabetes, chronic respiratory diseases, cardiovascular diseases, and cancer. These indicators provide valuable insights into the prevalence and impact of NCDs globally.

To effectively visualize and compare the NCD data across different countries and indicators, a choropleth map, a bar chart, and an interactive Pie chart are implemented. The choropleth map presents the NCD data on a world map, allowing users to explore the geographic distribution of various indicators. The map dynamically colors each country based on the corresponding indicator value, providing a visual representation of the severity or prevalence of a specific NCD.

In addition to the map and bar chart visualizations, the Pie chart offers an interactive way to explore the NCD data. It shows the values of various indicators categorized by Male, Female, and Total genders. Users can select specific regions or search for countries to view the Pie chart, or they can plot the chart by clicking on a country in the map view. This interactive Pie chart provides a clear representation of the indicator values for different genders, allowing users to compare and analyze the data in an intuitive manner.

The implementation includes a user-friendly interface that enables users to dynamically interact with the visualizations. Users can select specific NCD indicators, genders, and countries of interest through dropdown menus, triggering real-time updates of the visualizations based on the selected parameters. The combination of the choropleth map, bar chart, and interactive Pie chart offers a comprehensive and interactive exploration of the NCD data, facilitating a better understanding of the prevalence and impact of noncommunicable diseases worldwide

**Contributions:**

This visualization system makes significant contributions to the understanding of noncommunicable diseases (NCDs) by providing an intuitive and interactive interface to explore and analyze the NCD data set. Its key contributions include:

1. Comprehensive Global View: The choropleth map offers a comprehensive view of the global distribution of NCD indicators. By dynamically coloring each country based on the corresponding indicator value, users can easily identify regions with higher prevalence or severity of specific NCDs. This allows for a quick and intuitive understanding of the geographic patterns and hotspots of NCDs worldwide.
2. Comparative Analysis: The bar chart visualization enhances comparative analysis between countries. By presenting the indicator values for selected countries side by side, users can directly compare and identify variations and trends for different indicators and diseases. This facilitates the identification of disparities, similarities, and patterns between countries, contributing to a better understanding of the factors influencing NCD prevalence.
3. Interactive Exploration: The system provides interactive features for exploring the NCD data. Users can select specific NCD indicators, genders, and countries of interest through dropdown menus or by clicking on countries in the map view. Real-time updates of the visualizations based on the selected parameters enable users to dynamically explore and analyze the data, gaining valuable insights into the prevalence and impact of NCDs across different regions and population groups.
4. Pie Chart Representation: The addition of the interactive Pie chart further enriches the understanding of NCDs by showcasing indicator values categorized by Male, Female, and Total genders. Users can explore this chart by selecting regions or searching for specific countries. The Pie chart provides a clear and concise representation of the distribution of NCD indicator values across genders, facilitating a deeper understanding of gender-specific patterns and disparities in NCD prevalence.

**Challenges:**

* Inconsistent country names: Matching the country names between the GeoJSON file and the disease data file can be challenging due to different naming conventions or abbreviations used. Preprocessing and data cleaning techniques are necessary to ensure accurate mapping between the two datasets.
* Handling missing or incomplete data: Dealing with missing or incomplete data in the disease data file requires implementing strategies such as data interpolation or excluding countries with missing values to ensure the integrity and accuracy of the visualizations.
* Performance optimization: Optimizing the performance of visualizations, especially for large datasets, is crucial to ensure smooth and responsive user interactions.
* User experience design: Designing an intuitive and user-friendly interface that allows users to easily navigate and interpret the visualizations is essential for an engaging user experience.
* Scalability and flexibility: The implementation should be designed to accommodate future updates and additions to the NCD data set, requiring modular and extensible code structures and data handling techniques.

**Implementation Details:**

1. Technologies Used:
   * HTML, CSS, and JavaScript: These core web technologies are utilized for creating the user interface and defining the structure, style, and interactivity of the visualization system.
   * D3.js: D3.js, a powerful JavaScript library for data visualization, is leveraged to generate the choropleth map, bar chart, and pie chart components. D3.js provides a wide range of tools and functions for manipulating data and creating dynamic visualizations.
   * Node.js and Express: Node.js and Express are used to host the visualization system on a local server. They provide a lightweight and efficient server-side runtime environment.
2. Dataset and Techniques:
   * The NCD Portal Diseases Data Set is used as the primary dataset for the visualization system. This dataset encompasses statistical data related to various NCD indicators from countries worldwide.
   * Techniques such as data cleaning, filtering, and preprocessing are applied to ensure data accuracy and usability. For example, the dataset is compared with the GeoJSON data of world boundaries to match country names and make necessary adjustments for consistent mapping.
3. Data Cleaning Process:
   * The data cleaning process involves checking and reconciling the country names in the NCD dataset with the country names available in the GeoJSON data.
   * By comparing the names, any discrepancies or variations are identified and manually adjusted to ensure accurate mapping between the NCD data and the world map boundaries.
   * This data cleaning process ensures that the NCD indicators are correctly associated with the corresponding countries on the choropleth map.
4. Additional Implementation Details:
   * Bootstrap: The Bootstrap framework is used for creating a responsive and visually appealing user interface, providing a consistent layout and styling across different devices.
   * Dynamic Updates: The visualization system employs dynamic updates to reflect user selections and changes in real-time. When users interact with dropdown menus, click on countries, or change visualization options, the system responds by updating the map, bar chart, and pie chart accordingly.
   * Tooltip Display: Tooltips are implemented to display detailed information about specific countries and NCD indicators. These tooltips appear when users hover over the map or bar chart elements, providing additional context and data insights.
   * Responsive Design: The visualization system is designed to be responsive, adapting to different screen sizes and ensuring optimal user experience on desktop and mobile devices.

Features:

1. Choropleth Map Visualization:
   * The implementation includes a choropleth map that visually represents the selected NCD indicator data for different countries.
   * Each country is colored on the map based on its corresponding indicator value, providing a visual representation of the data distribution.
   * Tooltips are implemented to display additional information when hovering over the countries, such as the country name and indicator value.
2. Bar Chart Visualization:
   * The implementation includes a bar chart that allows for easy comparison of indicator values between different countries.
   * The bar chart displays the indicator values for the selected countries side by side, providing a clear visual comparison.
   * The bar chart can be sorted in ascending or descending order to facilitate ranking and comparison of countries based on the indicator values.
   * Different regions or groups of countries can be colored differently on the bar chart, providing additional insights into regional variations.
3. Interactive User Interface:
   * The implementation provides a user-friendly interface with dropdown menus for selecting the NCD indicator, gender, and countries of interest.
   * Users can select a specific NCD indicator and gender to display the corresponding data on the choropleth map and bar chart.
   * Buttons such as "Next" and "Previous" allow users to navigate through multiple pages of country data, enabling the plotting of all countries in a manageable manner.
4. Dynamic Data Updates:
   * The implementation dynamically updates the visualizations based on user selections, ensuring real-time display of relevant data.
   * When the user selects a specific NCD indicator, gender, or country, the visualizations are updated accordingly to reflect the new data.
   * The choropleth map and bar chart are updated with the filtered and sorted data, providing accurate and up-to-date visual representations.
5. Sorting and Pagination:
   * The bar chart includes sorting functionality that allows users to sort the indicator values in ascending or descending order.
   * Pagination is implemented to display multiple pages of country data in the bar chart, enabling the plotting of all countries in a manageable manner.
   * Buttons such as "Next" and "Previous" facilitate navigation between pages, providing a convenient way to explore the data across different countries.
6. Pie Chart Visualization:
   * The updated implementation now includes a pie chart visualization.
   * The pie chart presents the distribution of NCD indicator values based on gender (male, female, and total) for selected countries or regions.
   * Users can interact with the pie chart by selecting specific countries or regions through dropdown menus or by clicking on countries on the choropleth map.
   * The pie chart dynamically updates to display the proportion of indicator values for each gender category, providing a visual representation of gender-specific variations in NCD data.
   * The pie chart enhances the understanding of gender-related patterns in NCD indicators and allows for interactive exploration of the data based on regions or specific country selections.

Project Screenshots and Description:

1. Default Page

A map of the world

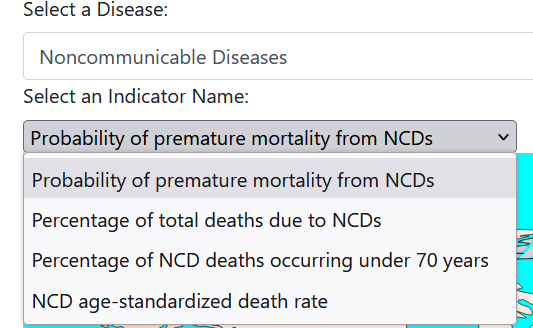
Description automatically generated

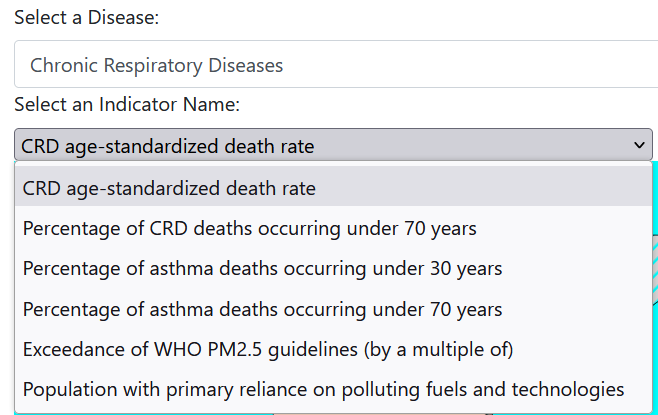
By default, the visualization is set to display data for "Non-Communicable Disease" (NCD), with the gender selected as "Total." The first indicator name available in the data will be used for analysis.

In the visualization, the oceans are colored using the "aqua" color scheme, providing a visually distinct representation. Additionally, countries that are not included in the selected disease data will be filled with striped lines, allowing for easy differentiation.

The legend accompanying the visualization dynamically shows the color scale with corresponding values. As the data changes, the legend adapts accordingly to reflect the updated information.

1. Indicator Name Drop Down Button





The Indicator Names in the visualization will dynamically change as the data changes, as the indicator name is a column in the dataset. This flexibility allows for specific information about different diseases for all countries to be accurately represented in the visualization.

1. Gender Drop down.

A picture containing text, line, screenshot

Description automatically generated

Unlike the Indicator Names Dropdown, it is Static and only 3 types of Genders are mentioned in the data set.

1. The color scale in the visualization represents the percentage or probability values (whichever is specified in the indicator name) for each country. It provides a clear visual representation of the data, allowing for easy comparison and analysis of the respective values across different countries.

A map of the world

Description automatically generated with medium confidence

1. In addition to the previous features mentioned, a tooltip has been implemented to provide detailed information about each country. When hovering over a specific country in the visualization, the tooltip will appear, displaying comprehensive information about that country. This includes relevant details and data points related to the selected disease, such as specific percentages, probabilities, or any other pertinent information available. The tooltip enhances the user experience by offering on-demand access to specific information while exploring the visualization.

A picture containing map, text, atlas

Description automatically generated

1. Pressing Toggle Chart Button will sow the Barchart view of the data. It will show the first 50 data Initially.

A screenshot of a computer

Description automatically generated with low confidence

1. Using Previous and Next Button we can navigate through the data

A screenshot of a graph

Description automatically generated with low confidence

1. Using Sort Button, we can Sort the data Sort will toggle between Ascending order and Descending order.

A picture containing text, screenshot, colorfulness, parallel

Description automatically generated

A picture containing text, screenshot, colorfulness, rectangle

Description automatically generated

1. After Sorting we can navigate through the Sorted data and axis will be dynamically scaled

A picture containing text, screenshot, colorfulness, parallel

Description automatically generated

This image is Next image of the previous sorted image.

1. We can plot only the selected region.

A screenshot of a graph

Description automatically generated with low confidence

1. After selecting a region, we can also sort it.

A screenshot of a graph

Description automatically generated with medium confidence

1. Pie Chart to visualize the relation ship between genders in a specific region or a country.

A screenshot of a computer

Description automatically generated with low confidence

1. We can search for countries in the search bar and It will suggest the countries from the data set.

A screenshot of a computer

Description automatically generated with medium confidence

1. After selecting the country and clicking the plot will plot the pie chart for that country.

A screenshot of a computer

Description automatically generated with low confidence

1. Also, by clicking a country from the chroplethMap we can plot that country data.

A picture containing map, text, atlas

Description automatically generated

A screenshot of a computer

Description automatically generated with low confidence

References

This [NCD Portal](https://ncdportal.org/), which focuses on noncommunicable diseases (NCDs), serves as a valuable resource for researchers, policymakers, and public health professionals, providing access to critical data to inform evidence-based interventions and strategies. And their key risk factors, aiming to address the global impact of NCDs on public health.

NCDs, including cardiovascular diseases (heart disease and stroke), cancer, diabetes, and chronic respiratory diseases, account for nearly three-quarters of global deaths. The factors contributing to these diseases are multifaceted, encompassing social, environmental, commercial, and genetic influences. Notably, NCDs disproportionately affect low- and middle-income countries, with 86% of the 17 million people who die from NCDs annually under the age of 70 residing in these regions.

The Noncommunicable Diseases Data Portal offers users the ability to explore comprehensive information on NCDs and their key risk factors, organized by country. I recreated everything from scratch never fully used any source from internet.

Data Set Used:

1. Cancer: Detailed data related to various types of cancer, including incidence, prevalence, mortality rates, and associated risk factors
2. Cardiovascular diseases (CVDs): Comprehensive information on heart disease and stroke, including prevalence, mortality rates, risk factors, and preventive measures.
3. Chronic respiratory diseases (CRDs): Data focusing on respiratory conditions such as asthma, chronic obstructive pulmonary disease (COPD), and other related ailments.
4. Diabetes: Detailed statistics and insights into diabetes, including prevalence, risk factors, management, and complications
5. Harmful alcohol use: Information on the detrimental effects of alcohol misuse and its impact on NCDs.
6. Obesity and Unhealthy diet: Data related to obesity rates, unhealthy dietary patterns, and their association with NCDs.
7. Physical inactivity: Information highlighting the importance of physical activity and the risks associated with a sedentary lifestyle.
8. Tobacco use: Comprehensive data on tobacco consumption, the prevalence of smoking, and its link to NCDs

Data Fields:

1. Year: 2019
2. Country Name: Names of various countries where the data was collected (e.g., Afghanistan, Albania, Algeria, Andorra, Angola, etc.)
3. Region: Different regions representing larger geographical areas (e.g., African Region, Eastern Mediterranean Region, European Region, Global Region of the Americas, South-East Asian Region, Western Pacific Region)
4. Gender: Not provided in the data
5. Area: Not provided in the data
6. Numeric: Value indicating the exceedance of PM2.5 guidelines
7. Lower Confidence Limit: Lower bound of the confidence interval for the numeric value
8. Upper Confidence Limit: Upper bound of the confidence interval for the numeric value