**Project Report: Interactive Data Visualization Dashboard for TB Using D3**



**Project Objective**

The goal of this project is to create an interactive dashboard that helps people explore and understand tuberculosis (TB) data. The dashboard will let users see patterns and connections in the data based on things like where and when the data was collected, and the type of relationships between different TB-related factors. By using different types of visual charts, the dashboard will make it easier for users to spot trends and make better decisions related to TB.

**Dataset Overview**

The dataset you are working with includes various data sources related to tuberculosis (TB) and HIV, including:

* **TB Portals Data Sharing**
* **TB-DIAH Data Hub**

The data contains multiple attributes essential for analysis:

* **Relationship Type:** The type of relationship between entities (e.g., collaboration, conflict).
* **Location**: Geographic identifiers (country, region).
* **Timestamp**: Date and time of the data entry.
* **Entity IDs:** Unique identifiers for entities.
* **Path**: Descriptive paths representing relationships or interactions.

**Data Preparation**

The dataset underwent several stages of cleaning and preprocessing to ensure readiness for analysis:

1. **Loading and Merging Datasets:**

The following datasets were loaded:

* tb\_data
* tbhiv\_data
* TBreport\_data
* tx\_data

These datasets were then merged along the columns to form a unified dataset.

1. **Duplicate Columns**:

Duplicate columns were identified using the value counts () function and dropped. The first occurrence of each column was retained, and the rest were removed.

1. **Data Cleaning**:

 The merged dataset was further cleaned by removing rows with Nan values, specifically those from rows 1 to 35.

 Missing values were filled with the mean of each respective column to maintain consistency and avoid any gaps in the dataset.

1. **Column Selection**:

The following key columns were selected for the final dataset:

* iso3, iso2, iso.numeric, year, country, g.whoregion, and others related to TB incidence, HIV test results, and SDG targets.

1. **Filtering**:

Data for the years 2020 and 2014 was extracted, which is likely intended to focus the analysis on recent and historical trends.

1. **Final Clean-Up**

The filtered dataset had its columns renamed for better clarity and to ensure consistency for later use in visualizations.

The final dataset is now saved as filtered\_data\_2020\_2014.csv.

**Key Visualizations and Features:**

**Force-Directed Graph:**

**I. Description:**

The force-directed graph is the main feature of the dashboard. It shows how countries and regions are connected by relationships, using circles (nodes) to represent the countries or regions. The lines (edges) between the nodes show how they are related, such as by TB incidence rates or the number of HIV-positive cases. The graph is interactive, so users can explore and understand how different countries and regions are linked based on these TB and HIV statistics. This helps to visually uncover patterns and connections between locations.

* **Features:**

**Nodes:** In the force-directed graph, countries and regions are shown as circles. The size of each circle changes dynamically based on TB incidence data, so larger circles represent countries or regions with a higher TB burden, while smaller circles indicate areas with a lower TB burden. This feature helps users quickly see which areas are most affected by TB, making it easier to compare the scale of the problem across different locations.

**Links:** In the force-directed graph, the lines connecting the circles (nodes) represent the relationships between countries and regions, such as how TB incidence is linked across locations. The thickness of these lines changes depending on the strength of the relationship—thicker lines indicate stronger connections (e.g., higher similarity in TB incidence), while thinner lines show weaker connections. This helps users quickly understand how closely related different regions are in terms of TB data, visually highlighting the strongest relationships.

* **Color Coding:**

**Nodes:** Countries are colored using a color scale based on TB incidence, while regions are assigned a distinct color using a predefined color scale

**Links:** The edges between nodes are colored and styled based on the type of relationship (e.g., regional connections).

* **Interactivity:**

**Hovering over nodes** shows a tooltip with detailed information, including TB incidence, HIV-positive statistics, and other relevant health data.

**Clicking on nodes** allows users to expand the nodes, providing a deeper look into the details and relationships of countries and regions.

**Zooming and panning**: The visualization supports zooming and panning to allow users to explore the data at different levels of detail.

1. **Tooltip:**

* **Functionality**: When users hover over any node, a tooltip appears, providing a summary of key information such as the country's TB incidence, HIV-positive rate, and the region to which the country belongs. If a region node is hovered, the tooltip shows summary statistics for that region, such as the average TB incidence and the number of countries in the region.

 **Design**: The tooltip appears near the mouse cursor and smoothly fades in and out to provide users with detailed information without cluttering the interface.

1. **Node Expansion:**

**Functionality:** Clicking on a node (whether a country or region) expands it, showing related detail nodes around it. For instance:

* For countries, additional nodes represent detailed health metrics like TB incidence, HIV-positive cases, etc.
* For regions, clicking on a region reveals the countries in that region arranged in a circular layout.

**Purpose:** This feature allows users to explore the data at a more granular level, drilling down into country-specific details or understanding the composition of regional statistics.

1. **Dynamic Filtering:**

 **Description:** The visualization includes dynamic filters that allow users to control the visibility of different types of relationships.

 **Functionality:**

* Users can check or uncheck relationship types (e.g., "regional connection") in the filter panel to control which links are displayed on the graph. This feature helps users focus on specific data and gain deeper insights.
* When filters are changed, the graph dynamically updates by hiding or showing relevant links and nodes.

1. **Zoom and Pan:**

 **Description:** The graph supports zooming and panning, allowing users to explore different parts of the network and view the relationships in detail.

** Functionality:**

* The zoom functionality is controlled using mouse gestures or touch interactions, providing users with flexibility in exploring the graph.
* Panning allows users to shift the focus of the graph to different regions or areas of interest.

1. **Node Size and Layout Adjustments:**

 **Dynamic Node Size:** The size of each node is dynamically adjusted based on its properties (e.g., TB incidence or HIV-positive rate). For instance, nodes representing countries with higher TB incidence will be larger than those with lower rates.

 Layout**:** The graph layout is optimized using a force-directed simulation, ensuring that nodes are evenly spaced and visually balanced while maintaining their relationships. This layout is recalculated during interactions like dragging and zooming.

**Map Chart for Geographic Representation**

1. **Description:**

The map visually displays the world, with pins representing entities (countries) and their respective TB-related data. The geographic layout of the map helps users to understand the global distribution of TB cases and relationships.

**Data Representation:**

**Pins:** Each country on the map is represented by a pin. The position of the pins corresponds to the geographic coordinates (latitude and longitude) of the respective country.

**Color Coding:** Pins are color-coded based on the type of relationship being visualized. The relationships include:

**Male Cases:** Represented by one color (e.g., blue).

**Female Cases:** Represented by another color (e.g., orange).

**HIV-positive Cases:** Represented by yet another color (e.g., green).

The color coding allows users to quickly identify and distinguish between different types of TB-related relationships across countries.

1. **Interactive Features:**

**Zoom and Pan:**

* Users can zoom in and out of the map using mouse scrolls or touch gestures, and pan around to focus on specific regions or countries.
* This feature is essential for exploring different geographic areas and seeing data from a macro (global) to micro (individual country) perspective.
* The zoom and pan interactions are supported by **D3.js** zoom functionality, which dynamically adjusts the map view.

**Tooltips:**

* When users hover over a pin (representing a country), a tooltip appears displaying detailed information about the TB relationship data for that country. The information shown includes:

Country Name

Region (e.g., Africa, Asia, Europe, etc.)

New TB Cases (Male, Female, HIV-positive)

* This tooltip offers a quick snapshot of the country's TB data and serves as an informative overlay without cluttering the map.
* The tooltip's design allows users to view the data without being distracted from the main visualization.

**Pin Highlighting:**

* When a user clicks on a pin, it is highlighted, and all other pins are dimmed (opacity reduced), providing a focused view on the clicked country.
* This feature helps in exploring a particular entity in-depth by reducing distractions from other pins on the map.

1. **Filtering Options:**

* **Region Filter:**
* Users can filter the map data by region. A dropdown allows the user to select a specific region (e.g., Europe, Africa, Asia, etc.) or view data for all regions.

**Functionality:** The map updates in real-time to show only the pins and data for the selected region, helping users zoom into specific geographic areas of interest.

* **Relationship Type Filter:**

This filter enables users to focus on specific relationships, such as:

* **Male TB cases**
* **Female TB cases**
* **HIV-positive TB cases**
* **Functionality:** Selecting a relationship type filter updates the map by showing only the pins that correspond to the selected relationship. This feature allows users to analyze gender or HIV-related TB cases specifically, thereby offering insights into these categories.
* **Combined Filters:** The user can apply both region and relationship filters simultaneously, enabling them to analyze specific TB relationships within a given geographic region. The map adjusts dynamically as both filters are changed.

#### **Legend:**

* A **legend** is displayed at the bottom-right corner of the map, which serves to explain the color coding used for the pins. Each relationship type (male, female, HIV-positive) has an associated color, which is explained in the legend.
* This provides users with clarity on what each color represents, ensuring that the map is easy to interpret.

#### **Geographic Projection:**

#### The map uses **Mercator projection** for accurate geographic representation, which is well-suited for world maps and provides a straightforward way to scale the map based on the container size.

#### **Pins and Filtered Data:**

* Based on the selected filters (region and relationship type), the data is processed to update the pins on the map. The pins' appearance (color, size, position) is updated dynamically based on the filtered data.
* The **pins** are circles that represent the countries, with their position on the map determined by the corresponding latitude and longitude coordinates from the CSV data.

#### **Update Mechanism:**

* The update Pins function is responsible for re-rendering the map based on the selected filters. It filters the data according to the user's choices and updates the pins accordingly. If a pin already exists, it is updated; if not, a new pin is added.

**Timeline Visualization with Animation**

**I. Description**:

This project creates an interactive timeline to show how tuberculosis (TB) diagnoses, treatments, and related data change over time. It focuses on how different regions and factors affect TB diagnosis and treatment. Built with D3.js, the visualization includes animations, interactive features, and options to filter data dynamically.

* **Animated Timeline**: The primary focus is on visualizing the evolution of TB-related data over time, with an interactive animation that plays through the years.
* **Date Range Slider**: Users can adjust the timeline by selecting a specific range of years using a slider. This allows for detailed analysis over a desired period.
* **Playback Controls**: Users can play or pause the animation to either observe the data flow or examine specific periods closely.
* **Dynamic Date Display**: The year is updated dynamically as the slider is adjusted or during the playback of the animation, helping users track the progression of TB-related metrics.

II. **Data Representation**:

* **Data Points as Circles (Dots)**: Each data point representing a year's TB incidence is shown as a circle on the timeline. The size of these circles can be adjusted dynamically based on the selected metric (e.g., treatment success rate, deaths, failures).
* **Interactive Dots**: Hovering over the dots displays a tooltip with additional information, including details such as the private vs. public sector diagnosis ratio or treatment success rate, depending on the chosen metric.

**III. Color-Coding by Region**:

* **Regional Color Scale**: Each data point is color-coded according to the region (WHO\_Region) using a predefined color scale. This helps differentiate between data points from different regions, making it easier to compare regional trends over time.

**IV. Filters and Interactivity**:

* **Size Metric Selection**: Users can choose from a list of metrics (e.g., sector ratio, treatment success, deaths) that control how the size of the data points (dots) is calculated. This allows for different views of the data based on the user’s interest.
* **Year Filter**: Users can specify a start and end year to filter the data shown in the timeline. Only data points within the selected date range are displayed, enhancing the focus on specific periods.

**V. Legend**:

* **Legend for Regions**: A legend is provided to map the color-coded regions to their corresponding names, helping users understand the data’s geographical context.

1. **Tooltips**:

* Tooltips are displayed when users hover over the data points. The tooltip content adapts based on the selected metric, offering different types of information, such as:
  + Private vs. public sector diagnosis ratio
  + Treatment success rate percentage
  + Number of deaths, treatment failures, or lost-to-follow-up cases.

1. **Technical Features:**

**D3.js Data Binding**: The project leverages D3.js to bind the TB data to the visualization elements. Data is read from a CSV file, and each data point (representing a year and its associated metrics) is mapped to a circle (dot) on the timeline.

**Scales and Axes**:

* **xScale (Linear Scale)**: Used for positioning the data points along the timeline based on the year.
* **Yscale (Linear Scale)**: Used for positioning the data points along the vertical axis based on the incidence of TB cases.
* **Axes**: The x-axis represents the years (from 1995 to 2024), and the y-axis represents the number of new TB incidences.

**Animation**:

* The #play-pause button toggles between play and pause states, allowing the user to animate the timeline or stop it to explore a specific period.
* The animation gradually progresses through the years, with data points becoming more visible as the timeline advances.

**Smooth Transitions**: The transition effects applied to the circles ensure that the changes in dot size, opacity, and position are visually appealing and easy to follow.

**Interactivity**:

* **Hover and Tooltip Interactivity**: When the user hovers over a data point, the circle enlarges, and a tooltip is displayed with detailed information about the selected metric.
* **Range Slider**: The #date-range slider allows users to manually control the timeline’s progress.

**Hierarchical Tree Map for Entity Categorization**

**I. Description:**

The hierarchical tree map shows relationships in a layered structure, such as WHO regions, countries, and their metrics. Users can interact with the map to explore data, filter by region, and dig deeper into specific levels. Colors and sizes help highlight key metrics, like the frequency of relationships or level of activity.

**II. Nested Rectangles**: The tree map uses rectangles inside rectangles to show different levels of relationships. It starts with the main "World" category, which splits into WHO regions, and then into individual countries. This setup helps users see how relationships are organized and spread across different levels.

**III. Size and Color Representation:**

**Size:** Each rectangle's size is proportional to the sum of New\_Incidence for that particular region or country. This metric highlights the intensity of tuberculosis (TB) cases at each hierarchical level.

**Color:** The rectangles are color-coded using a sequential blue scale, where darker shades indicate higher values for a metric (e.g., treatment success rates, relationship intensities). The color intensity helps users quickly identify areas with more significant data points.

**IV. Interactivity:**

**Hover Effects:** When users hover over a rectangle, a tooltip displays detailed information about the entity, such as its name, value, and the metric associated with the entity. This enhances the exploration and understanding of data at specific locations within the hierarchy.

**Drill-down Functionality:** Users can click on any rectangle to drill deeper into the hierarchy, revealing more detailed child nodes. This feature provides a deeper analysis of specific regions, countries, or categories.

**Breadcrumb Navigation:** As users drill down, a breadcrumb trail displays the current path, allowing users to trace their exploration history and easily navigate back to higher levels. This feature is essential for maintaining context within the hierarchical structure.

**V. Filters:**

* The visualization includes a filter that allows users to select specific WHO regions for analysis. This filter enables the tree map to update dynamically, showing data only for the selected region or for all regions combined if no filter is applied. This feature ensures that users can tailor the data presented according to their focus area.

**VII.Responsive Layout:**

* The tree map adjusts its layout to fit various screen sizes, ensuring that users can interact with the visualization effectively on different devices.

**Technical Aspects:**

* **Data loading and Processing:** The data is loaded from a CSV file, and the hierarchical structure is created using the d3.group method. The data is aggregated by WHO Region and Country, with the New Incidence and other relationship metrics being used to calculate the size and color of each node.
* **Tree map Layout:** The d3.treemap () function is used to create the hierarchical layout, which assigns each node a position within the SVG container. The padding between nodes is set to ensure clear distinctions between them.
* **Tooltip Functionality:** The tooltip is implemented using a simple HTML element that appears on hover. It is positioned based on the user's mouse location and displays information dynamically.
* **Responsive Interaction:** The on ("click") function listens for user clicks to navigate between hierarchical levels, while the breadcrumb updates accordingly.

**Sunburst Chart for Hierarchical Data**

**I.Description:**

The HTML file creates an interactive sunburst chart to show tuberculosis (TB) data in a hierarchy, such as WHO regions and countries. Users can explore the data by clicking on different layers, applying filters, and easily navigating through the various levels of information.

**II.Sunburst Chart Layout:**

* The chart uses a radial layout, with the innermost ring representing global-level data ("Global"), followed by regions and countries as outer rings.
* Each segment within the chart is a slice that represents either a region or a country.
* The segments are sized proportionally based on a metric (in this case, the sum of "New\_Relationship\_Male" values), and color-coded to distinguish different countries.

**III.Interactivity:**

* **Hover Effect:** When hovering over a segment, a tooltip appears showing detailed information about the selected region or country, including metrics such as value, relationship strength, and timestamp.
* **Click Effect:** Clicking on a segment zooms into the next level of the hierarchy, allowing deeper exploration of the data. Breadcrumb navigation at the top of the page updates accordingly to reflect the current position within the hierarchy.
* **Zooming:** The chart implements zoom functionality (Focus-Plus-Context) to allow users to zoom in and out of the sunburst chart, enhancing the focus on specific data levels.

**VI.Filters:**

* The chart features two dropdown menus that allow users to filter the data based on WHO Region and Country. When a region or country is selected, the chart updates to reflect only the relevant data.
* Filtered results are displayed dynamically in the "Filter Display" section below the chart, ensuring users are aware of their current selections.

**VII.Legend:**

A legend on the right panel provides a color-coded reference for the different countries in the dataset. Each country is represented by a unique color, and clicking on a country name in the legend toggles the visibility of that country's corresponding segments in the chart.

**VIII.Breadcrumb Navigation:**

Breadcrumb navigation helps users track their location in the data hierarchy. As users click on regions or countries, the breadcrumb dynamically updates to show the current path, allowing for easy backtracking.

**IX .Dynamic Data Rendering:**

The data is loaded dynamically from a CSV file. The dataset is grouped by WHO Region and Country, then hierarchical relationships are constructed using D3.js. If no data is available for the selected filters, an appropriate message is displayed to the user.

**X. Responsive Design:**

The layout is flexible, adjusting to different screen sizes. The sunburst chart and side panel (with the legend) are arranged to be viewable on various devices, ensuring a responsive user experience.

**XI.Detailed Tooltip Information:**

* When hovering over a segment, the tooltip displays detailed information such as the name of the region or country, the value (sum of "New\_Relationship\_Male"), relationship strength (sum of "New Incidence"), and the latest timestamp from the dataset.

**X.Color and Style:**

* The visual design uses a color palette with distinct hues for different countries. The inner nodes (regions) are filled with a neutral color, while countries are assigned different colors to enhance visual distinction.
* The layout also incorporates a dark background (#443f92) and light text for clear readability.

**CONCLUSION**

In conclusion, the interactive dashboard combines different visual tools—like force-directed graphs, maps, timelines, and a sunburst chart—to explore tuberculosis (TB) data in a dynamic way. Each tool has a special role: the graph shows connections between entities, the map shows TB spread by region, the timeline tracks TB changes over time, and the sunburst chart lets users explore data at global, regional, and country levels. Together, these features create an engaging and informative experience, helping users find insights and make better decisions. The dashboard also has interactive options like zooming, tooltips, and filters to improve its usefulness for analysis.

**Video link:**

**https://drive.google.com/file/d/1XNpMiceWyk-X1fOz9F2RT07V5IIZTnJP/view?usp=drive\_link**