

**DATA VISUALIZATION**

**Sai Venkata Naga Jayaujwal Vura**

**S564579**

**Guided by Denise Case**

**Introduction:**

Tuberculosis (TB) is a serious disease caused by bacteria that mainly affects the lungs. It is a major health problem around the world, especially in poorer countries. Even though TB can be prevented and treated, many people still get sick and die from it each year. The number of people affected by TB varies greatly between different countries. This is due to factors like poverty, the quality of healthcare systems, the presence of HIV, and the spread of drug-resistant TB. Countries with high rates of TB often face difficulties such as lack of access to medical care, social stigma, and not enough funding for health programs.

Understanding how TB affects different countries is important for tackling the disease. This means looking at data on how many people get TB and how many die from it and understanding the reasons behind these numbers. By comparing this information across countries, we can learn what works well and what needs more attention and resources. This report aims to provide a clear and simple analysis of the TB burden in various countries. By doing this, we hope to help develop better health strategies and policies to reduce TB and work towards the World Health Organization's goal of ending the TB epidemic by 2030.

**Dataset Name:** Tuberculosis Burden among the Country

**Description:** The dataset provides estimates of tuberculosis prevalence and mortality by country, offering insights into the global burden of the disease. It includes data on various indicators related to tuberculosis burden, such as incidence rates, prevalence rates, mortality rates, treatment success rates, and drug resistance patterns.

Data Set in EXCEL format:



**Data Source and Link:** The dataset is sourced from the World Health Organization (WHO) and can be accessed via the following link: <https://www.who.int/data/gho/data/themes/tuberculosis>

**Number of Columns & rows:** The dataset comprises 51 columns and 5120 rows, encompassing a wide range of variables related to tuberculosis burden and associated factors.

**Data visualization tool:** Tableau

**Data Cleaning:**

Data cleaning, also known as data cleansing or data scrubbing, is the essential process of identifying and correcting errors, inconsistencies, and inaccuracies within a dataset to enhance its quality and reliability. This involves several key steps: removing duplicate records to prevent skewed analysis results, handling missing data by either imputing appropriate values or discarding records with excessive missing information, and correcting data entry errors such as typos and misplaced decimal points. Standardizing data formats ensures consistency in dates, numeric values, and text cases. Validation checks confirm that data values fall within expected ranges and make logical sense.

In this project, I have cleaned the data using tableau prep by removing unnecessary columns (i.e., columns contain completely null values and duplicate columns) in the data set.

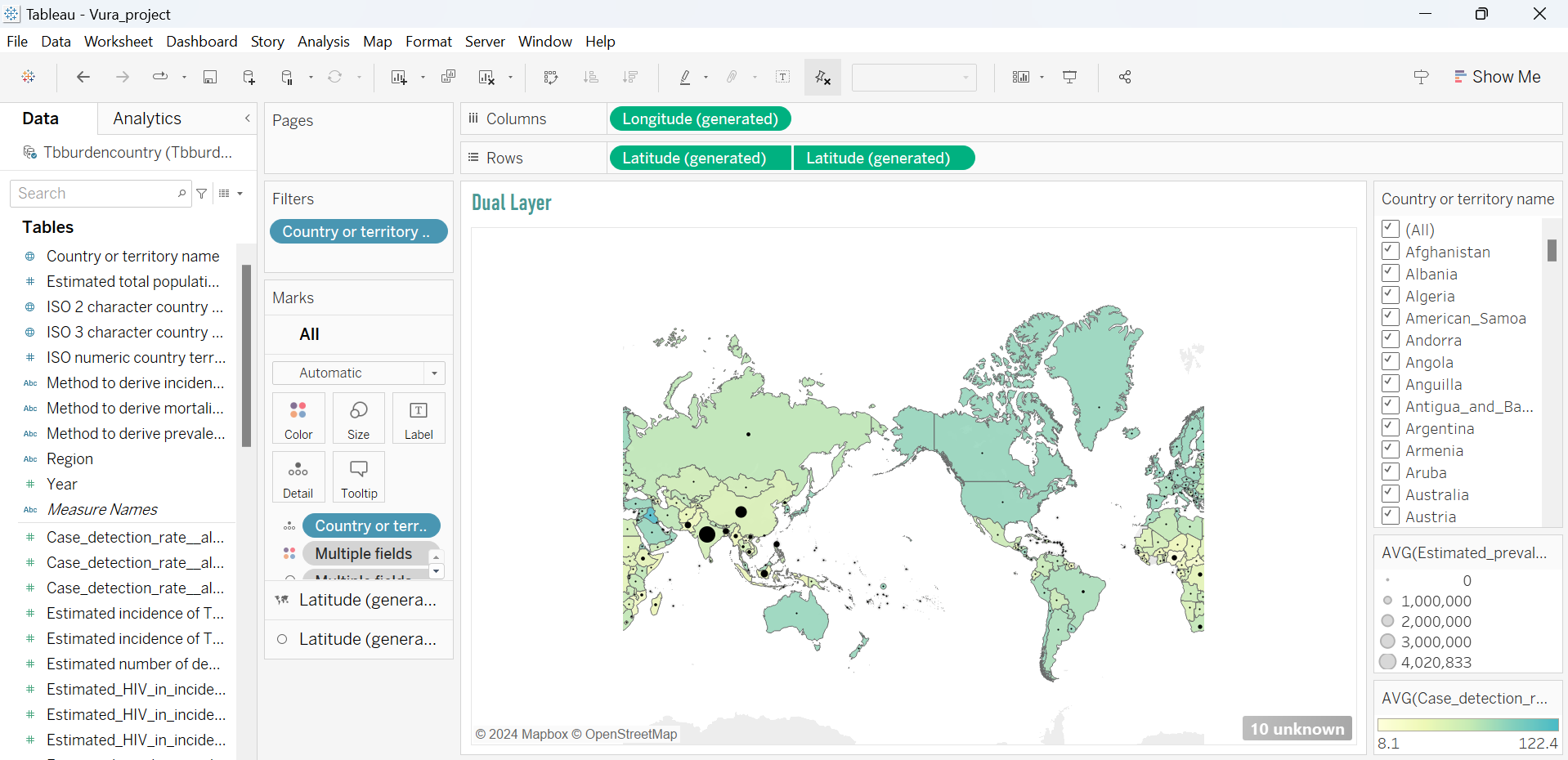
**Goals:**

1. Prioritize TB Control Efforts in High Burden Countries Using Incidence and Prevalence Data.
2. Analyze and Understand the Trend of TB Burden Over years Across Different Countries Using a Tree Map for Average Cases Detected
3. Investigate Socioeconomic Determinants Influencing TB Burden Using a Bar Graph Showing Average Number of People Affected by TB
4. Compare TB burden between developing vs. developed countries: This can reveal disparities in TB control and highlight where more resources/interventions are needed.
5. Examine TB in vulnerable groups: Assess TB burden among marginalized populations like prisoners, migrants, homeless, indigenous communities.
6. Analyze the difference between theTB deaths across different regions, highlighting the disparity in TB mortality rates.

|  |  |
| --- | --- |
| Goals | Charts Used |
| Goal 1 | Dual layer Map |
| Goal 2 | Tree Map |
| Goal 3 | Bar Chart |
| Goal 4 | Symbol Map |
| Goal 5 | Line Graph |
| Goal 6 | Lollipop Chart |

**Goal 1:** Prioritize TB Control Efforts in High Burden Countries Using Incidence and Prevalence Data.

**Dual layer Map (Incidence Rate VS Prevalence)**

****

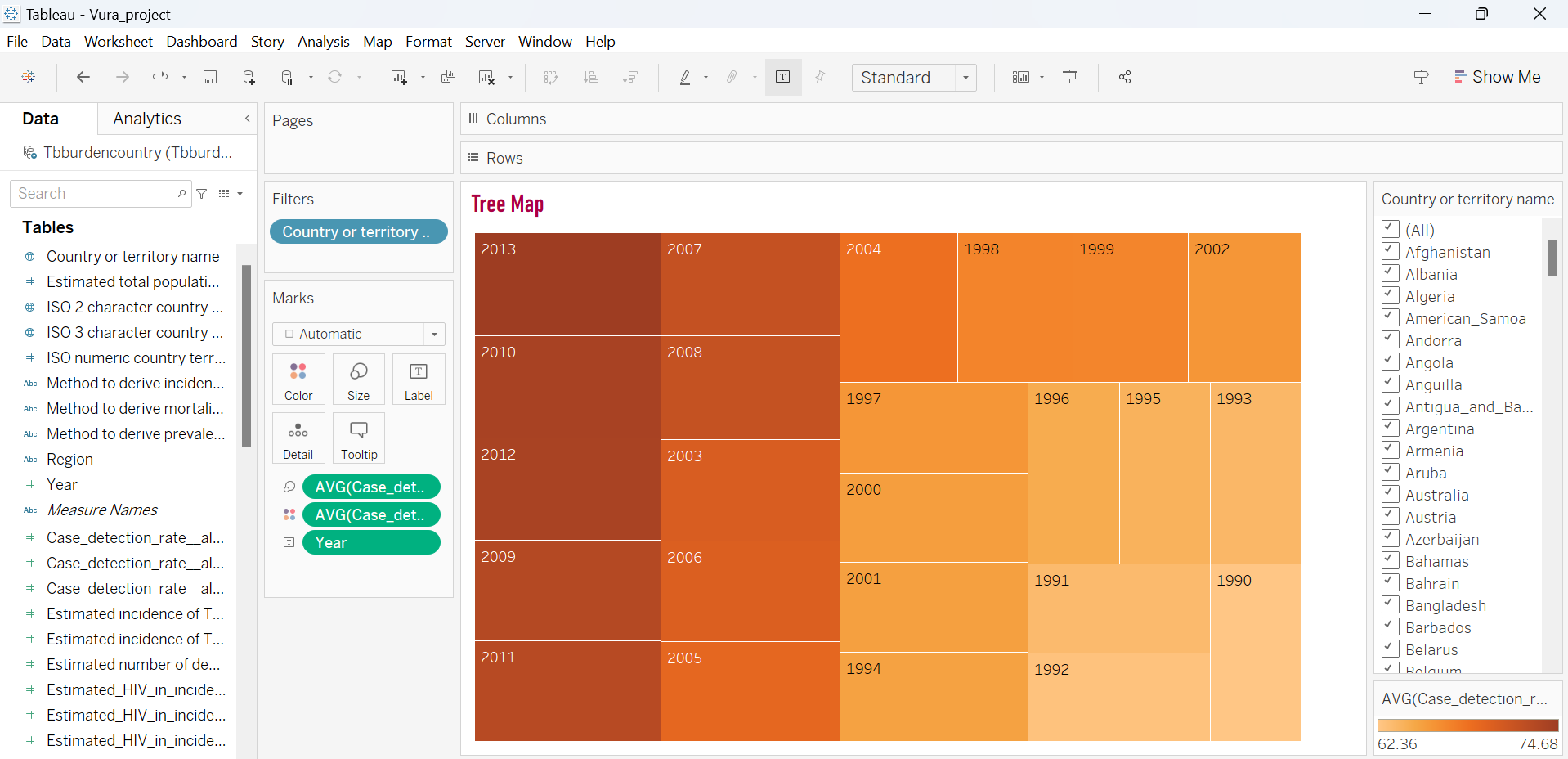
**Story:**

The figure displayed is a geographic visualization made with Tableau that shows how the burden of tuberculosis (TB) is distributed around the world among different nations. It represents TB-related statistics using proportional circles and color coding, with varied green hues denoting differing TB prevalence levels among the countries. Lighter hues indicate a lesser prevalence of tuberculosis, while darker tints indicate a higher prevalence. The map also has black circles overlaid on it, with the diameters of the circles representing the magnitude of various TB-related measures, including estimated prevalence or incidence. Users can browse statistics for all nations together or pick individual countries for more in-depth analysis using the filter on the right.

Areas in South Asia and Sub-Saharan Africa exhibit higher TB prevalence, denoted by larger circles and darker shades. This suggests that these regions face significant TB challenges because of factors such as high population density and inadequate healthcare infrastructure. On the other hand, the lighter shading of North America, Europe, and some regions of Australia denotes a lower prevalence of tuberculosis, which is frequently associated with improved healthcare systems and higher living standards.

**Goal 2:** Analyze and Understand the Trend of TB Burden Over years Across Different Countries Using a Tree Map for Average Cases Detected

**Tree Map (Detecting Average Cases)**

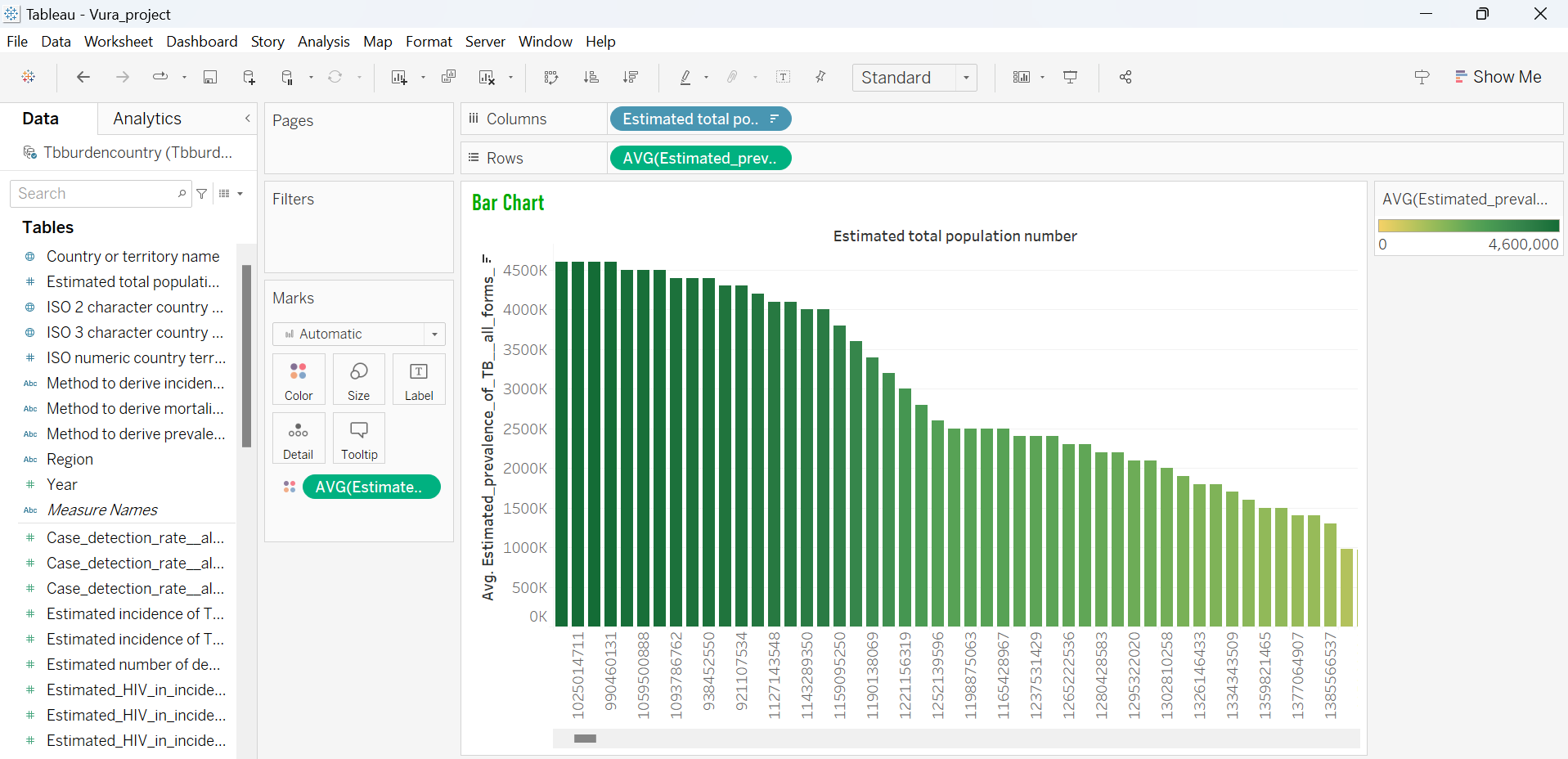
****

**Story:**

The figure on display is a Tableau tree map visualization that shows the average annual case detection rate of tuberculosis (TB). The size and color of each rectangle in the tree map indicate a particular year, and they also provide crucial information regarding the rates of TB case detection. Lighter orange hues suggest lower detection rates, while darker orange hues indicate higher rates. The average detection rates of the years with larger rectangles are greater, suggesting that TB was successfully detected during those times. On the other hand, years with lower detection rates are represented by smaller and lighter rectangles. This graphic makes it simple to compare the rates of tuberculosis case detection across time, emphasizing patterns and shifts in the efficiency of TB detection initiatives.

**Goal 3:** Investigate Socioeconomic Determinants Influencing TB Burden Using a Bar Graph Showing Average Number of People Affected by TB

**Bar Graph (Showing Average no of people effected by TB)**

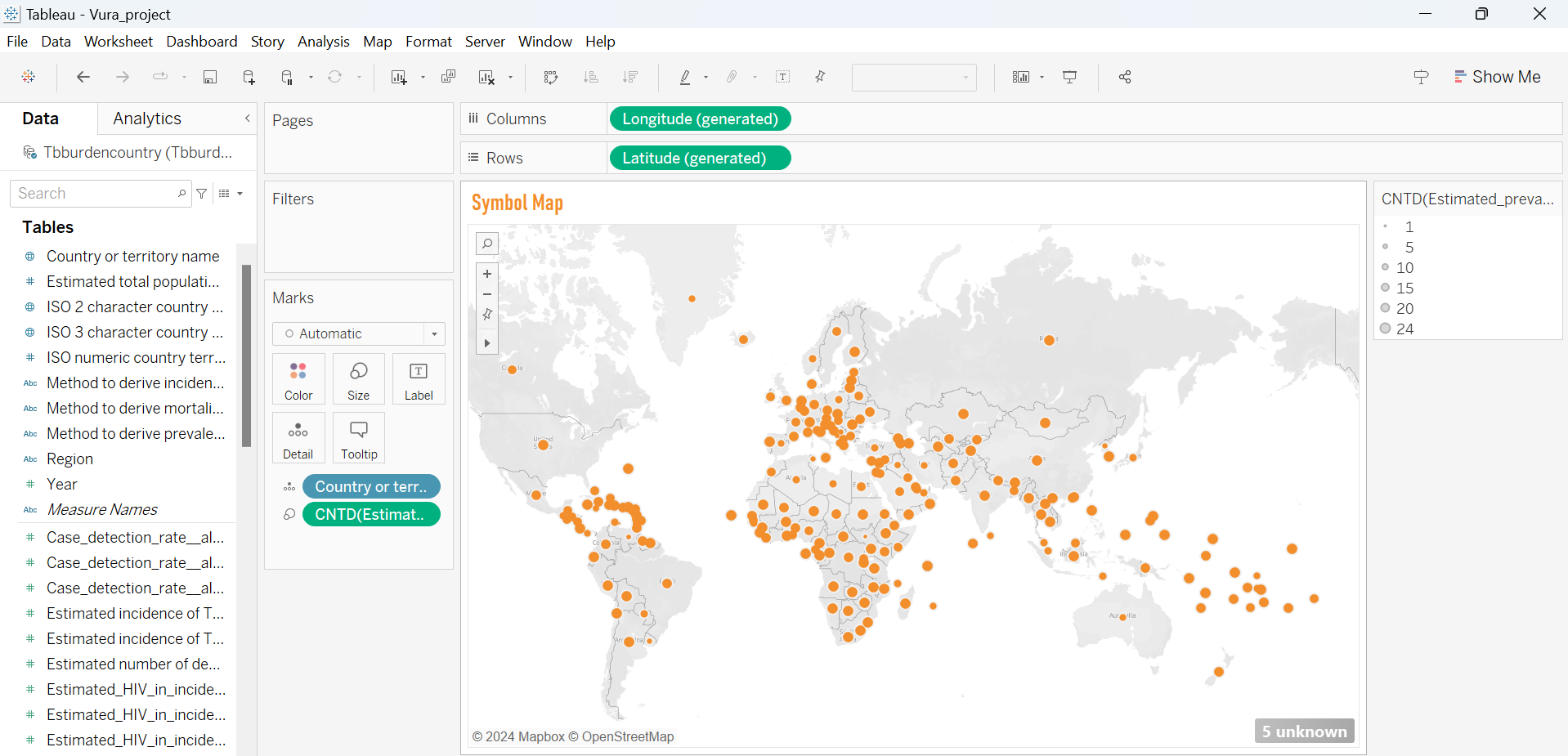
****

**Story:**

The graphic shows a bar chart with each bar representing a separate population group identified by its projected total population number. The bar chart shows the average estimated prevalence of a given metric (maybe a disease or health condition) across different population groups. Darker hues of the bars, which range from green to yellow, represent higher averages. The gradient is dependent on the prevalence figures. With the group with the largest projected total population on the left and progressively declining towards the right, the population groups are arranged in descending order. The more populated groups have the highest prevalence rates, as this graphic clearly illustrates the disparity in prevalence across various population sizes. This implies that the population size and the health indicator under study may be correlated.

**Goal 4:** Compare TB burden between developing vs. developed countries: This can reveal disparities in TB control and highlight where more resources/interventions are needed.

**Symbol Map (Showing Average estimated prevalence of countries)**

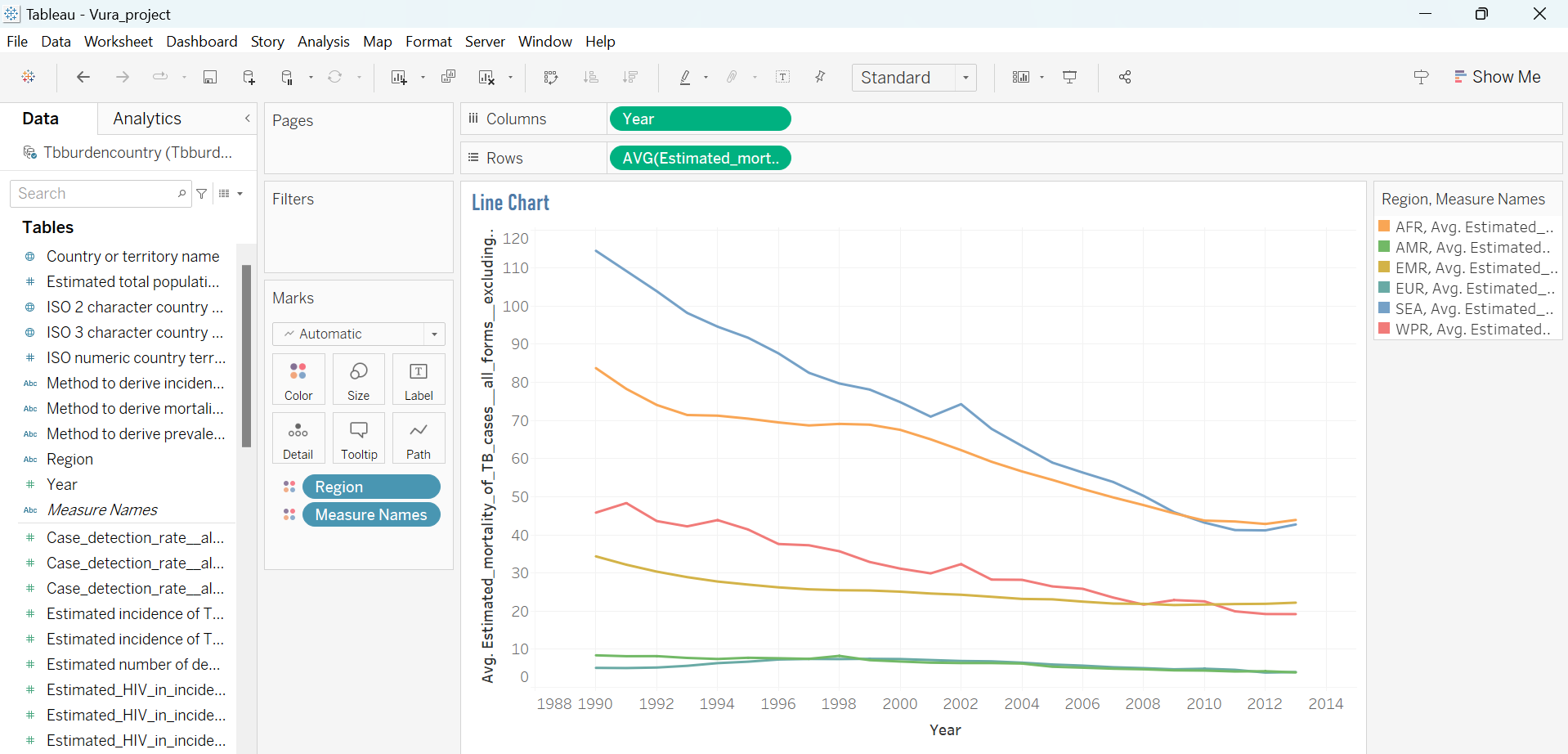


**Story:**

The image displays a symbol map that shows the global distribution of a certain metric possibly linked to population or health data across numerous nations and territories. Every orange dot indicates a location, and the number of estimated prevalence values is indicated by the size of the dot. Greater counts are indicated by larger dots, indicating areas with higher concentrations of the relevant metric. Important clusters are shown on the map in areas including South Asia, sub-Saharan Africa, and portions of Europe and South America. These clusters may indicate regions with greater reporting or incidence rates of the studied statistic. The map is a useful visual assistance for locating possible hotspots and geographic trends, which helps to clarify regional differences and direct focused activities or additional research.

**Goal 5**: Examine TB in vulnerable groups: Assess TB burden among marginalized populations like prisoners, migrants, homeless, indigenous communities.

**Line Graph (Estimating mortality rate of different regions)**

****

**Story:**

The image's line chart illustrates trends in average estimated mortality rates across several locations from 1988 to 2014 for a certain cause (perhaps tuberculosis or a comparable disease). AFR (Africa), AMR (Americas), EMR (Eastern Mediterranean), EUR (Europe), SEA (Southeast Asia), and WPR (Western Pacific) are the varied hues that each line symbolizes. The graph shows a distinct decreased trend in death rates in every location, underscoring the notable advancements in health outcomes over time.   
  
The AFR region is shown by the blue line, which begins with the greatest death rate and then noticeably declines to show significant progress while maintaining a relatively high rate in comparison to other locations. The lines that are red and yellow, signifying the SEA and AMR regions, respectively, exhibit notable declines in death rates, which are indicative of better disease control and efficient health interventions. The EUR and EMR regions are represented by the green and orange lines, which show lower beginning mortality rates and a continuous decline over time.  
  
Overall, the graph shows how much the world has progressed in lowering death rates from the cause, with regional variations that might be a result of disparities in public health programs, healthcare systems, and economic situations. The significance of consistent efforts and focused interventions to maintain the positive trend and address any lingering discrepancies is highlighted by this image.

**Goal 6:** Analyze the difference between theTB deaths across different regions, highlighting the disparity in TB mortality rates.

**Lollipop Chart (TB deaths vs TB Mortality rates)**

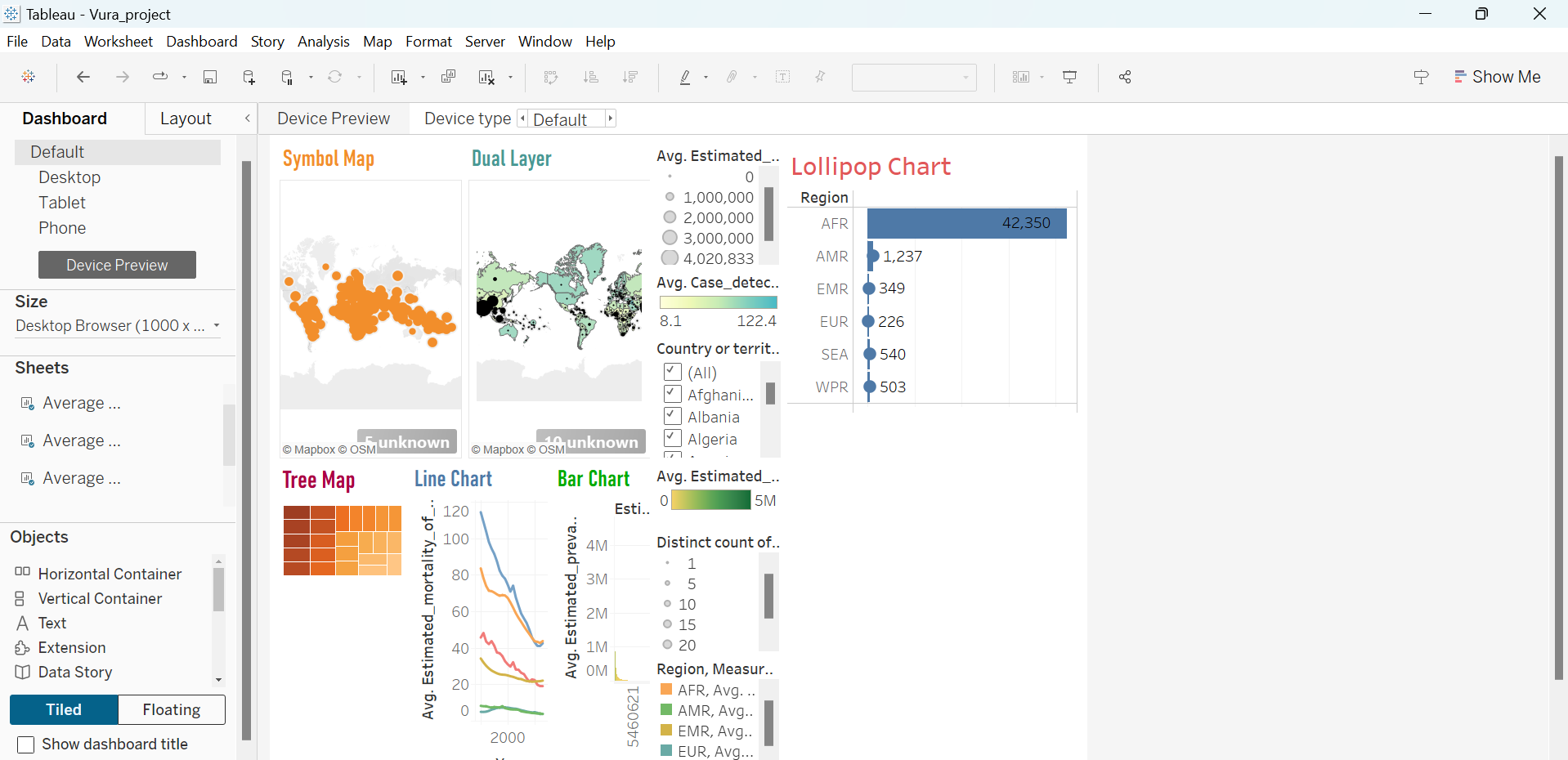
**A screenshot of a computer

Description automatically generated**

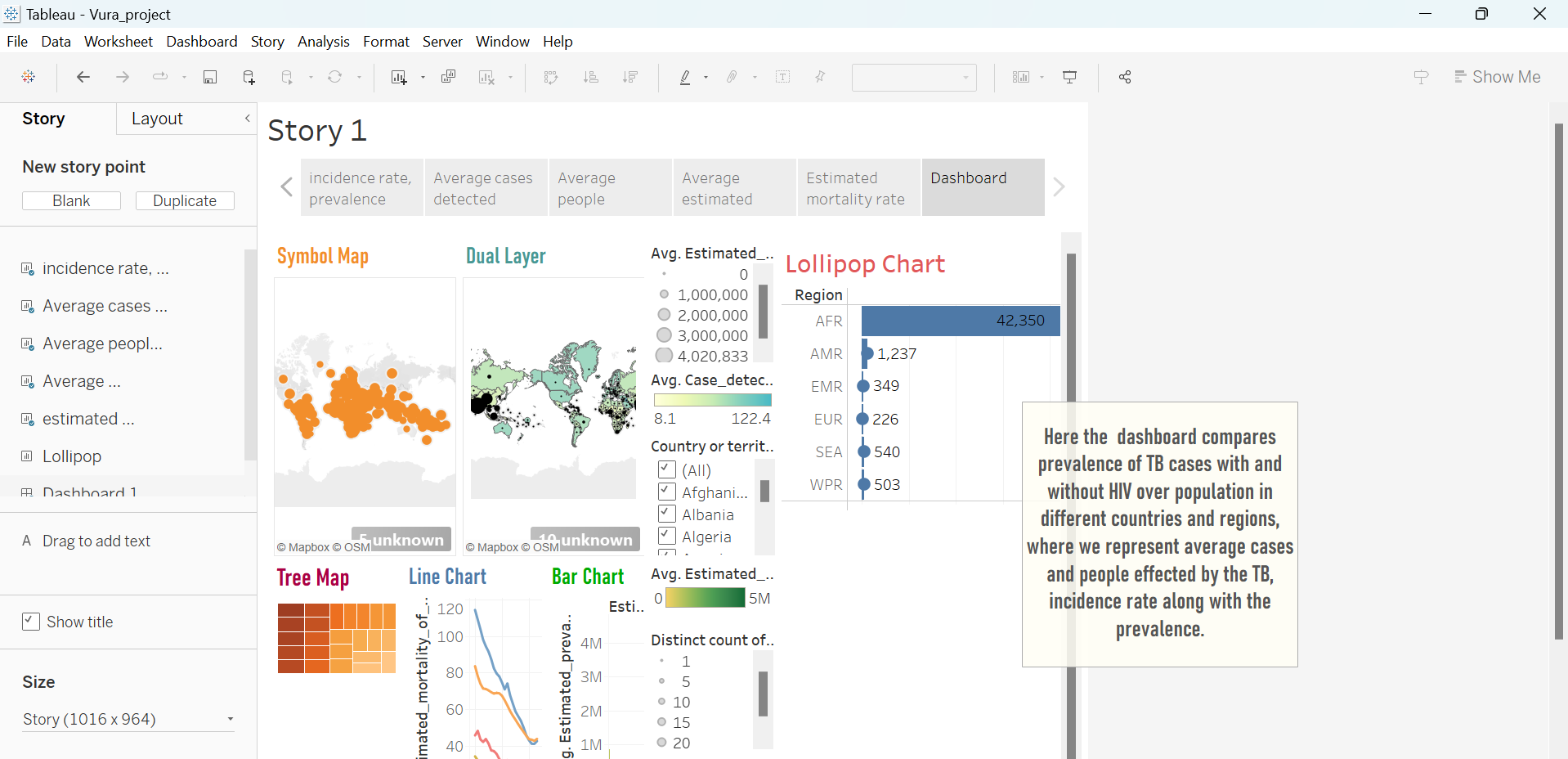
**Story:**

Significant differences may be seen in this lollipop chart, which graphically depicts the expected number of TB deaths in different parts of the world. With an estimated 42,350 fatalities, Africa (AFR) has an extremely high death rate that dwarfs all other regions. In comparison, the Eastern Mediterranean (EMR) and European (EUR) areas have 349 and 226 deaths, respectively, while the Americas (AMR) have a comparatively lower estimate of 1,237 deaths. With 540 and 503 deaths, respectively, Southeast Asia (SEA) and the Western Pacific (WPR) likewise have alarming statistics. To address this serious public health concern, focused TB interventions are urgently needed in the most affected areas, especially in Africa. This is highlighted by the striking visualization.

**Tableau Dashboard:**

****

**Tableau Storyboard:**

****

**Conclusion:**

* Summary of Findings: The project provided valuable insights into the global burden of tuberculosis, including trends, geographic disparities, and associated factors.
* Achievement of Goals: Through data analysis and visualization, the project successfully met its goals of understanding TB burden indicators and identifying areas for targeted interventions.
* Lessons Learned: The project highlighted the importance of data quality, collaborative teamwork, and iterative analysis in tackling complex public health challenges.
* Future Directions: Moving forward, the insights gained from this project can inform policy decisions, resource allocation strategies, and further research efforts aimed at TB control and prevention.