



Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India

Lab Manual: Design Big Data Dashboards using Tableau on Women
Empowerment/Gender Participation Dataset

1. Aim

To design and develop interactive Big Data dashboards in Tableau using the dataset on Women Empowerment and Gender Participation. This project focuses on creating visualizations that represent various facets of gender participation, using basic and advanced charts to derive insights.

2. Objectives

1. Develop a Tableau dashboard that visualizes different aspects of women empowerment and gender participation.
2. Create basic charts (Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot) to represent fundamental data distributions and relationships.
3. Design advanced charts (Word chart, Box and Whisker plot, Violin plot, Regression plot, 3D chart, Jitter) to analyze complex relationships and distributions.
4. Derive insights and observations from each visualization to better understand patterns and trends within the dataset.
5. Document a comprehensive analysis with observations for each chart and suggest potential areas for further research or application.

3. Expected Outcomes

- A set of interactive, visually rich Tableau dashboards that demonstrate different techniques for visualizing data.
- Insights and observations on gender participation trends, disparities, and distributions across various dimensions.
- Enhanced understanding of the impact of women empowerment initiatives through interactive data visualization.

4. Theory

Big Data visualization using Tableau provides a powerful way to gain insights into extensive

datasets. In this lab, we explore how to use Tableau to visualize gender-based data effectively.

- Basic Charts allow for foundational analysis and comparison, helping to display straightforward trends and differences across categories.
- Advanced Charts enable deeper insights into the data by visualizing distributions, variations, and more intricate relationships.

Each chart has a specific use:

- Bar and Pie charts are effective for categorical comparisons.
- Histograms help understand data distributions.
- Timeline and Scatter plots show trends over time and relationships between variables.
- Bubble plots add another dimension to scatter plots, useful for displaying data points with varying magnitudes.
- Box and Whisker, Violin plots display statistical distributions.
- Regression plots analyze relationships between variables.
- 3D charts provide a multi-dimensional view.
- Jitter plots prevent overplotting by slightly shifting data points.

5. Python Preprocessing (Optional)

If the dataset is too large or requires preprocessing before importing into Tableau, Python can help clean and structure the data.

6. Tableau Implementation

1. Data Import: Import the file into Tableau.

2. Chart Creation:

Basic Charts

- Bar Chart: Compare participation levels across countries.
- Pie Chart: Show the distribution of participation by region or sector.
- Histogram: Analyze the distribution of empowerment scores across different demographics.
- Timeline Chart: Display trends in empowerment scores over time for various countries.
- Scatter Plot: Show the relationship between participation levels and empowerment scores.
- Bubble Plot: Visualize participation levels with bubble size indicating population count.

3. Advanced Charts

- Word Chart: Represent common themes or terms from survey responses or empowerment program descriptions.

- Box and Whisker Plot: Show variations in empowerment scores across regions.
 - Violin Plot: Visualize the density and distribution of empowerment scores.
 - Regression Plot: Show linear and nonlinear relationships between age and empowerment score or education level and empowerment.
 - 3D Chart: Combine three dimensions, such as age, score, and participation level, to analyze multidimensional trends.
 - Jitter Plot: Display data points spread out slightly to avoid overlap, especially useful for dense regions.
4. Observations and Insights:
- Analyze each chart for patterns, trends, and anomalies.
 - Compare results across regions, age groups, or years to identify significant insights.

Dataset:

The Women Empowerment Index (WEI) is a multifaceted tool designed to assess and track the progress of women's empowerment within societies. Drawing from diverse metrics and indicators, the WEI offers a nuanced understanding of the status of women across various domains. It builds upon the foundation laid by existing indices like the Gender Inequality Index (GII) but focuses specifically on aspects related to women's empowerment. The WEI encompasses several key dimensions, including economic participation, political representation, access to education and healthcare, and social inclusivity. By analyzing these dimensions, the index sheds light on the extent to which women are able to exercise agency, access resources, and participate fully in societal processes.

Content

This dataset provides essential information on gender development indicators, facilitating comprehensive analysis and comparison across countries and regions. Here are the key columns included:

- **Country:** The name of the country or territory included in the dataset.
- **Women's Empowerment Index (WEI) - 2022:** Quantifies women's empowerment within each country for the year 2022, offering insights into gender dynamics and opportunities for women.
- **Women's Empowerment Group - 2022:** Categorizes countries based on their performance and status regarding women's empowerment, facilitating comparative analysis.
- **Global Gender Parity Index (GGPI) - 2022:** Measures gender parity at a global level for the year 2022, indicating the degree of equality between men and women.
- **Gender Parity Group - 2022:** Classifies countries into groups based on their level of gender parity, aiding in the identification of trends and disparities.
- **Human Development Group - 2021:** Classifies countries into specific groups based on their Human Development Index (HDI) scores and development levels for the year 2021.
- **Sustainable Development Goal regions:** Aligns countries with specific regions as defined

by the Sustainable Development Goals (SDGs), providing context for development efforts and initiatives.

Dashboard:

The dashboard you provided appears to be focused on analyzing the relationship between various socioeconomic indicators, particularly those related to women's empowerment and human development. Let's break down the different elements and their potential insights:

Overall Visualization:

- The dashboard seems to be comparing the **Women's Empowerment Index (WEI)**, the **Global Gender Parity Index (GGPI)**, and the **Human Development Index (HDI)** across different countries.

Individual Elements:

1. WEI for each country:

- This bar chart likely displays the WEI scores for each country, allowing for a visual comparison of women's empowerment levels across the globe.
- The countries are categorized into different Human Development Groups (High, Medium, etc.), possibly indicating how WEI scores correlate with overall human development.

2. Proportions by Women's Empowerment Groups:

- This mosaic plot appears to show the distribution of countries across different WEI groups (Low, Lower-Middle, etc.).
- The size of each tile likely represents the number of countries in that group.

3. Countries by Women Empowerment Groups:

- This world map probably displays countries categorized by their WEI group.
- The color of each country likely corresponds to its WEI group, providing a

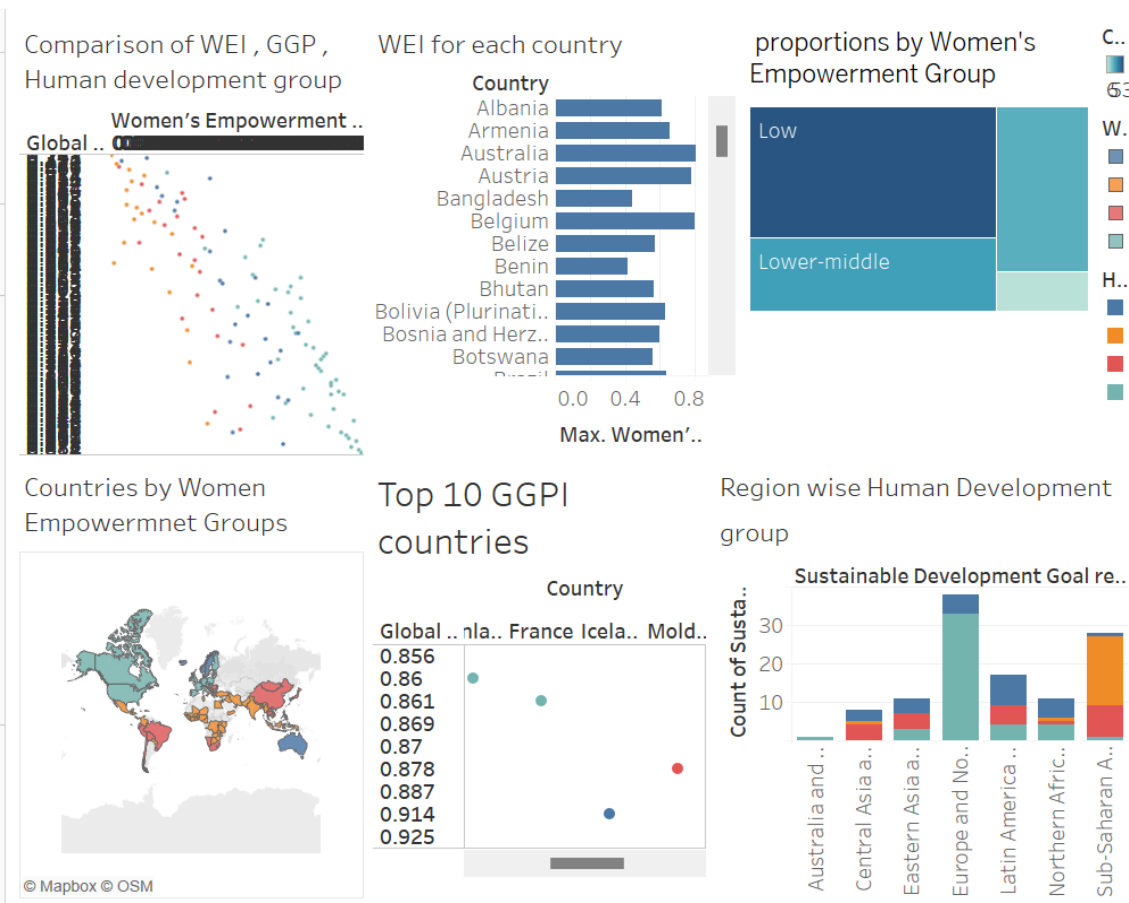
geographic perspective on women's empowerment.

4. Top 10 GGPI countries:

- This scatter plot likely shows the top 10 countries with the highest GGPI scores.
- The x-axis might represent the GGPI score, and the y-axis could be another relevant metric (e.g., WEI score).

5. Region-wise Human Development:

- This stacked bar chart seems to compare the distribution of countries across different Human Development Groups (High, Medium, etc.) for each region.
- The height of each bar represents the total number of countries in that region, and the different colored segments within each bar show the proportion of countries in each HDG.



7. Conclusion

This lab demonstrates how Big Data visualizations can be designed to gain insights into gender participation and women empowerment. Tableau's capabilities allow for interactive and visually rich dashboards, making it easier to analyze large datasets and identify trends or gaps. The visualizations offer a valuable perspective for decision-makers to understand the effectiveness of empowerment programs.

Lab Manual: Design Big Data Dashboards using Tableau on Education Sector Dataset

1. Aim

To design interactive Big Data dashboards in Tableau using datasets from the Education sector, focusing on enrollment rates, academic performance, funding, and demographics, among other factors. This project aims to reveal insights on trends, disparities, and key metrics in education.

2. Objectives

1. Create basic and advanced charts in Tableau to visualize data relevant to the Education sector.
2. Develop dashboards that convey meaningful insights on various education metrics.
3. Derive observations and trends from each visualization to understand the educational landscape better.
4. Document findings for each visualization and provide a basis for future research or recommendations.

3. Expected Outcomes

- A suite of interactive Tableau dashboards that highlight trends, distributions, and relationships within the Education dataset.
- Clear observations on trends in enrollment, performance, funding allocation, and demographic patterns.
- An enhanced understanding of educational disparities and patterns across different demographics and regions.

4. Theory

Big Data dashboards enable the visualization of extensive datasets, turning complex educational data into insights.

Basic Charts focus on fundamental distributions and comparisons, helping illustrate

5. Tableau Implementation

1. Data Import: Import the cleaned_education_data.csv file into Tableau.

2. Chart Creation:

Basic Charts

- Bar Chart: Show the total enrollment across different regions.

Pie Chart: Display the distribution of educational funding by region or program type.

- Histogram: Examine the distribution of student scores or other performance metrics.

- Timeline Chart: Analyze trends in enrollment or funding over time.

- Scatter Plot: Visualize the relationship between funding and student performance.
- Bubble Plot: Represent multiple dimensions, such as enrollment size with bubble size indicating funding.

3. Advanced Charts

- Word Chart: Display frequently occurring themes or keywords in student or educator feedback.
- Box and Whisker Plot: Show the distribution of student scores across different regions.
- Violin Plot: Display density and distribution of enrollment numbers by demographic.
- Regression Plot: Use linear and nonlinear regression to analyze relationships, such as between funding and average test scores.
- 3D Chart: Combine dimensions, such as funding, enrollment, and performance, to analyze multi-dimensional trends.
- Jitter Plot: Spread out dense data points in enrollment numbers across demographics for clearer visualization.

4. Observations and Insights:

- Document findings and insights from each chart to provide a deeper understanding of the data.

Observations:

Here are some observations based on each section of the dashboard:

1. Schools Having Drinking Water (Line Chart):

- The availability of drinking water in schools varies significantly across states.
- Some states, such as Punjab and Tripura, have a noticeably higher number of schools with drinking water facilities compared to others.
- States like Dadra and Nagar Haveli show a lower count, indicating a disparity in access to drinking water facilities across different regions.

2. Distribution of SC Students (Box Plot):

- The box plot shows the projected population of Scheduled Caste (SC) students, likely across different states.
- There is a considerable range in SC student population across states, with a few outliers above 1.5 million.
- The distribution shows that most states have a similar SC population range, but some states have significantly higher populations, which could represent more densely populated areas or areas with a high concentration of SC students.

3. State-Wise Student Enrollment (Map):

- The map highlights student enrollment by state, with darker shades indicating higher enrollment numbers.
- States like Uttar Pradesh and Bihar have the highest student enrollment, as indicated by the darker shades, likely due to their larger populations.
- Northeastern states and smaller regions tend to have lower enrollment, reflecting population and school density differences across India.

4. Number of Teachers vs. Students (Scatter Plot):

- The scatter plot shows the relationship between the number of teachers and the number of students across various states.
- Some states have a high number of students but a relatively lower number of teachers, suggesting a high student-to-teacher ratio in these regions.
- States with more balanced points indicate better student-to-teacher ratios, potentially reflecting better resource allocation.

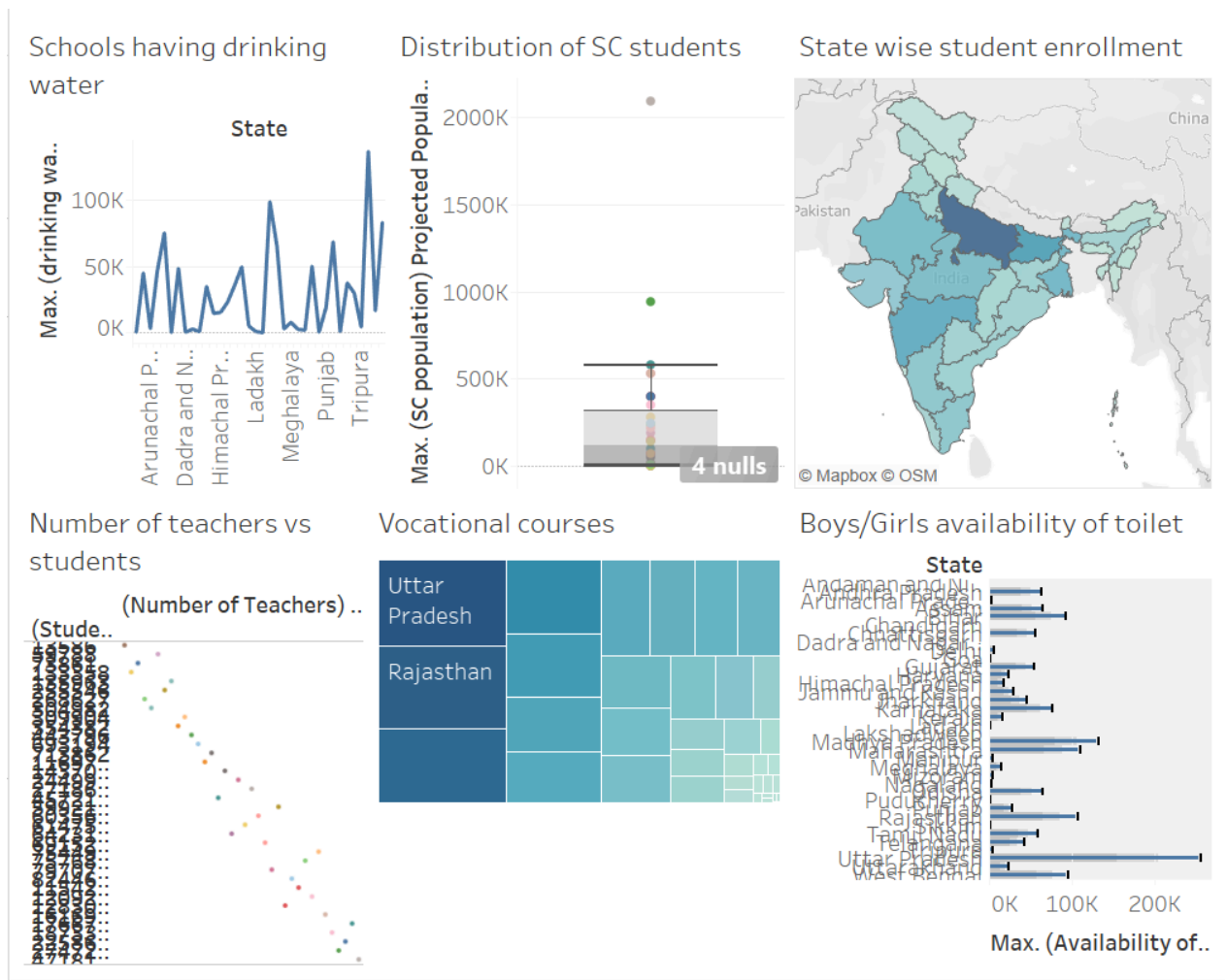
5. Vocational Courses (Treemap):

- This treemap shows the distribution of vocational courses across states.
- Larger squares, such as Uttar Pradesh and Rajasthan, indicate a higher number of vocational courses in those states, suggesting they may prioritize vocational education more heavily.
- Smaller squares indicate fewer vocational course offerings in other states, possibly due to resource limitations or lower demand.

6. Boys/Girls Availability of Toilet (Bar Chart):

- The bar chart compares the availability of toilets for boys and girls across states.
- Most states have similar availability for boys' and girls' toilets, with some differences.
- States like Uttar Pradesh and Bihar show high toilet availability, reflecting better infrastructure, whereas smaller states have lower counts, indicating a gap in sanitation facilities.

Overall, this dashboard shows disparities in facilities and resources (like drinking water, toilets, and vocational courses) across states, with populous states like Uttar Pradesh and Bihar generally having higher counts. The data also suggests that some regions have more challenges in providing adequate infrastructure and teaching staff relative to student enrollment.



Conclusion

The lab exercise demonstrates the effectiveness of using Tableau for educational data visualization. The visualizations allow insights into trends, disparities, and areas of improvement, facilitating a comprehensive understanding of the educational landscape.