Unit V Big Data Visualization

5.1 Introduction to Data Visualization: Introduction to Data Visualization: Objective and challenges to Big data visualization, Conventional data visualization tools, techniques for visual data representation

📊 **Introduction to Data Visualization**

✅ **What is Data Visualization?**

Data visualization is the process of **turning raw data into visual formats** like graphs, charts, and maps so that it's easier to **understand, analyze, and communicate**.

💬 In simple words: It's like making a picture from numbers to **see trends and patterns**.

🎯 **Objectives of Data Visualization**

1. **Simplify complex data**

2. **Identify patterns, trends, and outliers**

3. **Support data-driven decision-making**

4. **Communicate findings clearly**

5. **Improve data storytelling**

🚧 **Challenges in Big Data Visualization**

Big data is **huge, fast-changing, and complex**, which makes it hard to visualize. Here are the key challenges:

Challenge	Explanation
Volume	Too much data to show clearly
Variety	Different types (text, images, numbers) hard to combine
Velocity	Data changes quickly, needs real-time visualization
Scalability	Tools may lag or crash with large data
Interpretability	Too much detail can confuse users
Interactivity	Users want to explore, filter, and zoom into data easily

🛠️ Conventional Data Visualization Tools

Tool	Use Case
MS Excel	Simple graphs and charts
Tableau	Drag-and-drop dashboards
Power BI	Interactive reports and visual analytics
Google Data Studio	Free tool for dashboards
R & ggplot2	Great for statistical graphics
Python & Matplotlib/Seaborn/Plotly	Custom visualizations for data science

🎨 Techniques for Visual Data Representation

Technique	Visual Format	Best For
Bar Chart	Bars	Comparing values across groups
Line Chart	Lines	Trends over time
Pie Chart	Circular segments	Proportions (use with caution!)
Histogram	Bars (with bins)	Data distribution
Scatter Plot	Dots	Relationship between variables
Heatmap	Colored grids	Correlation or density
Box Plot	Boxes with whiskers	Spread and outliers
Treemap	Nested rectangles	Hierarchical data
Word Cloud	Word size & color	Frequency of terms in text data

✅ Summary

Topic	Key Idea
What is Data Visualization?	Turning data into visuals to understand and communicate it
Objectives	Make data easy to understand, spot patterns, support decisions
Big Data Challenges	Volume, variety, speed, interactivity
Tools	Excel, Tableau, Power BI, Python, R, etc.
Techniques	Bar, Line, Scatter, Heatmaps, Word Clouds, etc.

5.2 types of data visualization, Tools used in data visualization, Open – source data visualization tools, Analytical techniques used in Big data visualization, Data Visualization using Tableau

📊 Introduction to Data Visualization

✅ What is Data Visualization?

Data visualization is the process of **turning raw data into visual formats** like graphs, charts, and maps so that it's easier to **understand, analyze, and communicate**.

💬 In simple words: It's like making a picture from numbers to **see trends and patterns**.

🎯 Objectives of Data Visualization

1. **Simplify complex data**

2. **Identify patterns, trends, and outliers**

3. **Support data-driven decision-making**

4. **Communicate findings clearly**

5. **Improve data storytelling**

🚧 Challenges in Big Data Visualization

Big data is **huge, fast-changing, and complex**, which makes it hard to visualize. Here are the key challenges:

Challenge	Explanation
Volume	Too much data to show clearly

Challenge	Explanation
Variety	Different types (text, images, numbers) hard to combine
Velocity	Data changes quickly, needs real-time visualization
Scalability	Tools may lag or crash with large data
Interpretability	Too much detail can confuse users
Interactivity	Users want to explore, filter, and zoom into data easily

🛠️ Conventional Data Visualization Tools

Tool	Use Case
MS Excel	Simple graphs and charts
Tableau	Drag-and-drop dashboards
Power BI	Interactive reports and visual analytics
Google Data Studio	Free tool for dashboards
R & ggplot2	Great for statistical graphics
Python & Matplotlib/Seaborn/Plotly	Custom visualizations for data science

🎨 Techniques for Visual Data Representation

Technique	Visual Format	Best For
Bar Chart	Bars	Comparing values across groups
Line Chart	Lines	Trends over time
Pie Chart	Circular segments	Proportions (use with caution!)
Histogram	Bars (with bins)	Data distribution
Scatter Plot	Dots	Relationship between variables
Heatmap	Colored grids	Correlation or density
Box Plot	Boxes with whiskers	Spread and outliers
Treemap	Nested rectangles	Hierarchical data
Word Cloud	Word size & color	Frequency of terms in text data

✅ Summary

Topic	Key Idea
What is Data Visualization?	Turning data into visuals to understand and communicate it
Objectives	Make data easy to understand, spot patterns, support decisions
Big Data Challenges	Volume, variety, speed, interactivity
Tools	Excel, Tableau, Power BI, Python, R, etc.
Techniques	Bar, Line, Scatter, Heatmaps, Word Clouds, etc.

✅ What is Data Visualization?

Data visualization is the process of **turning raw data into visual formats** like graphs, charts, and maps so that it's easier to **understand, analyze, and communicate**.

💬 In simple words: It's like making a picture from numbers to **see trends and patterns**.

🎯 Objectives of Data Visualization

1. **Simplify complex data**

2. **Identify patterns, trends, and outliers**

3. **Support data-driven decision-making**

4. **Communicate findings clearly**

5. **Improve data storytelling**

🚧 Challenges in Big Data Visualization

Big data is **huge, fast-changing, and complex**, which makes it hard to visualize. Here are the key challenges:

Challenge	Explanation
Volume	Too much data to show clearly
Variety	Different types (text, images, numbers) hard to combine
Velocity	Data changes quickly, needs real-time visualization
Scalability	Tools may lag or crash with large data

Challenge	Explanation
Interpretability	Too much detail can confuse users
Interactivity	Users want to explore, filter, and zoom into data easily

🛠️ Conventional Data Visualization Tools

Tool	Use Case
MS Excel	Simple graphs and charts
Tableau	Drag-and-drop dashboards
Power BI	Interactive reports and visual analytics
Google Data Studio	Free tool for dashboards
R & ggplot2	Great for statistical graphics
Python & Matplotlib/Seaborn/Plotly	Custom visualizations for data science

🎨 Techniques for Visual Data Representation

Technique	Visual Format	Best For
Bar Chart	Bars	Comparing values across groups
Line Chart	Lines	Trends over time
Pie Chart	Circular segments	Proportions (use with caution!)

Technique	Visual Format	Best For
Histogram	Bars (with bins)	Data distribution
Scatter Plot	Dots	Relationship between variables
Heatmap	Colored grids	Correlation or density
Box Plot	Boxes with whiskers	Spread and outliers
Treemap	Nested rectangles	Hierarchical data
Word Cloud	Word size & color	Frequency of terms in text data

✅ Summary

Topic	Key Idea
What is Data Visualization?	Turning data into visuals to understand and communicate it
Objectives	Make data easy to understand, spot patterns, support decisions
Big Data Challenges	Volume, variety, speed, interactivity
Tools	Excel, Tableau, Power BI, Python, R, etc.
Techniques	Bar, Line, Scatter, Heatmaps, Word Clouds, etc.

5.2 types of data visualization, Tools used in data visualization, Open – source data visualization tools, Analytical techniques used in Big data visualization, Data Visualization using Tableau

📊 **Types of Data Visualization**

Different visualizations are used depending on the type of data and purpose.

Type of Visualization	Description	Best Used For
Bar Chart	Bars of different heights	Compare categories
Line Chart	Points connected by lines	Show trends over time
Pie Chart	Circular chart with slices	Show proportions
Histogram	Bins of data values	Show distribution
Scatter Plot	Points on x and y axis	Show relationships between variables
Heatmap	Color-coded matrix	Show density or correlations
Box Plot	Box with median, quartiles, outliers	Show spread of data
Treemap	Nested rectangles	Show part-to-whole relationships
Word Cloud	Words sized by frequency	Visualize text data
Dashboard	Group of visuals on one screen	Summary of key metrics and trends

🛠️ Tools Used in Data Visualization

🔒 **Proprietary (Paid) Tools:**

- **Tableau** – Advanced, drag-and-drop visuals

- **Microsoft Power BI** – Business analytics dashboards

- **QlikView** – Data discovery and dashboards

- **Looker (Google)** – Cloud-based visualizations

🆓 **Open-Source Tools:**

Tool	Language	Use Case
Matplotlib	Python	Basic plots
Seaborn	Python	Beautiful statistical graphs
Plotly	Python/R/JS	Interactive plots
Bokeh	Python	Interactive web dashboards
ggplot2	R	Advanced graphics and plots
D3.js	JavaScript	Complex interactive web visualizations
Apache Superset	Web UI	Web-based BI dashboards (big data-ready)

📈 **Analytical Techniques in Big Data Visualization**

When dealing with big data, we need smart techniques to handle and visualize data:

Technique	Description
Sampling	Take a small representative part of the data
Aggregation	Summarize data (e.g., total, average)

Technique	Description
Filtering	Show only relevant or selected data
Dimensionality Reduction	Reduce features while keeping important info (e.g., PCA, t-SNE)
Real-Time Processing	Handle streaming data (e.g., dashboards that update live)
Drill-Down/Drill-Up	Zoom in or out of data levels

📊 Data Visualization Using Tableau

Tableau is one of the most popular tools for interactive and beautiful data visualizations.

🔧 Features of Tableau:

- **Drag-and-drop interface**

- Works with **Excel, CSV, databases, cloud data**

- Supports **filters, tooltips, and interactive dashboards**

- Allows **real-time data analysis**

🎯 Common Visualizations in Tableau:

- Bar, Line, Pie charts

- Maps and Geo plots

- Heatmaps and TreeMaps

- Dashboards and Stories

🛠️ Steps to Use Tableau:

1. **Connect to Data** (Excel, CSV, SQL, etc.)

2. **Drag Fields** to Columns and Rows

3. **Choose a Chart Type**

4. **Apply Filters** and **Colors**

5. **Create Dashboards** for multiple visualizations

✅ Summary

Topic	Key Points
Types of Visualization	Bar, Line, Pie, Scatter, Heatmap, Word Cloud, etc.
Tools	Tableau, Power BI, Matplotlib, Seaborn, ggplot2, Plotly, D3.js
Open-source Tools	Python (Matplotlib, Seaborn), R (ggplot2), Bokeh, Plotly, Superset
Big Data Techniques	Sampling, Filtering, Aggregation, Real-time, Drill-down
Tableau	Powerful drag-and-drop tool for interactive dashboards

5.3 Introduction to: Candela, D3.js, Google Chart API

🎨 Candela

✅ What is Candela?

- **Candela** is an open-source **Python-based** library designed for **creating interactive, web-based visualizations**.

- It is built on **Bokeh**, which is a popular Python visualization library.

🎯 **Main Features:**

- **Interactive Plots**: Allows for mouse interaction (hover, click).

- **Easy Integration**: Works well in Jupyter Notebooks and can generate visualizations directly for web-based dashboards.

- **Web-Based**: Generates HTML-based output that can be embedded in websites.

📊 **Examples of Use:**

- Candela allows you to build **interactive scatter plots**, **line graphs**, and more complex visualizations with minimal code.

🌐 **D3.js (Data-Driven Documents)**

✅ **What is D3.js?**

- **D3.js** is a **JavaScript library** for producing **dynamic, interactive data visualizations** on the web using **HTML, SVG, and CSS**.

- It gives you complete control over your visualizations, allowing you to manipulate the DOM based on the data.

🎯 **Main Features:**

- **Customization**: Highly flexible, you can build nearly any type of visualization, from bar charts to complex network graphs.

- **Data Binding**: D3 allows you to **bind data** to DOM elements (like SVG shapes) and then apply **data-driven transformations** to them.

- **Interactivity**: You can add **animations, zooming, and brushing** to make your visualizations interactive.

📊 **Examples of Use:**

- **Bar charts, line charts**, **heatmaps**, **treemaps**, and **network diagrams** are commonly created using D3.js.

- Popular for **custom dashboards** and **web-based visualizations**.

📊 **Google Chart API**

✅ **What is Google Chart API?**

- The **Google Chart API** is a **JavaScript library** that allows you to create a variety of **interactive charts and graphs**.

- It's very easy to integrate with **Google Sheets**, other **Google services**, and your own data.

🎯 **Main Features:**

- **Variety of Charts**: Offers charts like **line**, **pie**, **bar**, **scatter**, and **geographic maps**.

- **Interactive**: Many charts are interactive by default, allowing users to hover, zoom, and more.

- **Easy to Use**: Requires minimal coding and is great for quick web visualizations.

- **Integration with Google Services**: It integrates easily with **Google Sheets**, so you can visualize data without needing to import/export it.

📊 **Examples of Use:**

- **Pie charts**, **line charts**, and **geocharts**.

- Simple data visualization for websites or dashboards.

✅ Summary

Tool	Language	Key Features	Best For
Candela	Python	Interactive, web-based plots	Python-based interactive visualizations
D3.js	JavaScript	Highly customizable, dynamic	Complex and custom visualizations on the web
Google Chart API	JavaScript	Easy integration with Google, Interactive charts	Simple web-based charts, quick setups