**UNIT-V**

**Data Visualization**

**Pixel-Oriented Visualization Techniques,**

**Geometric Projection Visualization Techniques,**

**Icon-Based Visualization Techniques,**

**Hierarchical Visualization Techniques,**

**Visualizing Complex Data and Relations**

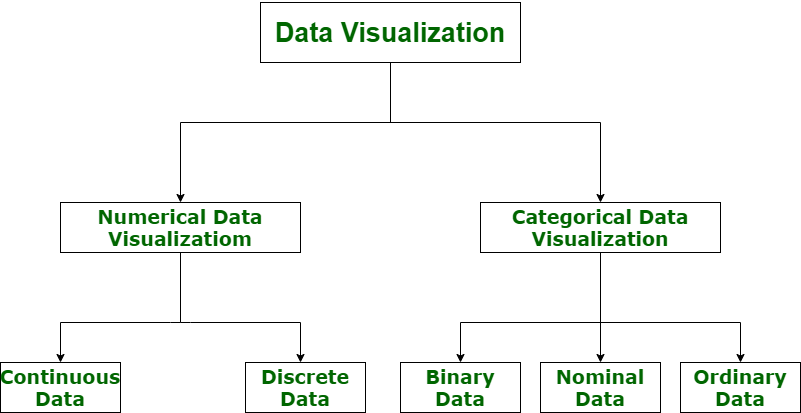
**Data Visualization:**

Data visualization is actually a set of data points and information that are represented graphically to make it easy and quick for user to understand. Data visualization is good if it has a clear meaning, purpose, and is very easy to interpret, without requiring context. Tools of data visualization provide an accessible way to see and understand trends, outliers, and patterns in data by using visual effects or elements such as a chart, graphs, and maps.

**Characteristics of Effective Graphical Visualization**

* It shows or visualizes data very clearly in an understandable manner.
* It encourages viewers to compare different pieces of data.
* It closely integrates statistical and verbal descriptions of data set.
* It grabs our interest, focuses our mind, and keeps our eyes on message as human brain tends to focus on visual data more than written data.
* It also helps in identifying area that needs more attention and improvement.
* Using graphical representation, a story can be told more efficiently. Also, it requires less time to understand picture than it takes to understand textual data.

**Categories of Data Visualization**   
Data visualization is very critical to market research where both numerical and categorical data can be visualized that helps in an increase in impacts of insights and also helps in reducing risk of analysis paralysis. So, data visualization is categorized into following categories :



1. **Numerical Data :**  
   Numerical data is also known as Quantitative data. Numerical data is any data where data generally represents amount such as height, weight, age of a person, etc. Numerical data visualization is easiest way to visualize data. It is generally used for helping others to digest large data sets and raw numbers in a way that makes it easier to interpret into action. Numerical data is categorized into two categories :
   * **Continuous Data –**  
     It can be narrowed or categorized (Example: Height measurements).
   * **Discrete Data –**  
     This type of data is not “continuous” (Example: Number of cars or children’s a household has).

The type of visualization techniques that are used to represent numerical data visualization is Charts and Numerical Values. Examples are Pie Charts, Bar Charts, Averages, Scorecards, etc.

1. **Categorical Data :**  
   Categorical data is also known as Qualitative data. Categorical data is any data where data generally represents groups. It simply consists of categorical variables that are used to represent characteristics such as a person’s ranking, a person’s gender, etc. Categorical data visualization is all about depicting key themes, establishing connections, and lending context. Categorical data is classified into three categories :
   * **Binary Data –**  
     In this, classification is based on positioning (Example: Agrees or Disagrees).
   * **Nominal Data –**  
     In this, classification is based on attributes (Example: Male or Female).
   * **Ordinal Data –**  
     In this, classification is based on ordering of information (Example: Timeline or processes).

The type of visualization techniques that are used to represent categorical data is Graphics, Diagrams, and Flowcharts. Examples are Word clouds, Sentiment Mapping, Venn Diagram, etc.

**Data visualization** is the representation of data through use of common graphics, such as charts, plots, infographics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand.

**Dashboards** are effective data visualization tools for tracking and visualizing data from multiple data sources, providing visibility into the effects of specific behaviors by a team or an adjacent one on performance. Dashboards include common visualization techniques, such as:

* **Tables:**This consists of rows and columns used to compare variables. Tables can show a great deal of information in a structured way, but they can also overwhelm users that are simply looking for high-level trends.
* **Pie charts and stacked bar charts:** These graphs are divided into sections that represent parts of a whole. They provide a simple way to organize data and compare the size of each component to one other.
* **Line charts and area charts:** These visuals show change in one or more quantities by plotting a series of data points over time and are frequently used within predictive analytics. Line graphs utilize lines to demonstrate these changes while area charts connect data points with line segments, stacking variables on top of one another and using color to distinguish between variables.
* **Histograms:** This graph plots a distribution of numbers using a bar chart (with no spaces between the bars), representing the quantity of data that falls within a particular range. This visual makes it easy for an end user to identify outliers within a given dataset.
* **Scatter plots:** These visuals are beneficial in reveling the relationship between two variables, and they are commonly used within regression data analysis. However, these can sometimes be confused with bubble charts, which are used to visualize three variables via the x-axis, the y-axis, and the size of the bubble.
* **Heat maps:** These graphical representation displays are helpful in visualizing behavioral data by location. This can be a location on a map, or even a webpage.
* **Tree maps,** which display hierarchical data as a set of nested shapes, typically rectangles. Treemaps are great for comparing the proportions between categories via their area size.

**Open source visualization tools**

Access to data visualization tools has never been easier. Open source libraries, such as D3.js, provide a way for analysts to present data in an interactive way, allowing them to engage a broader audience with new data. Some of the most popular open source visualization libraries include:

* **D3.js:** It is a front-end JavaScript library for producing dynamic, interactive data visualizations in web browsers. [D3.js](https://d3js.org/) (link resides outside IBM) uses HTML, CSS, and SVG to create visual representations of data that can be viewed on any browser. It also provides features for interactions and animations.
* **ECharts:** A powerful charting and visualization library that offers an easy way to add intuitive, interactive, and highly customizable charts to products, research papers, presentations, etc. [Echarts](https://echarts.apache.org/en/index.html" \t "_blank) (link resides outside IBM) is based in JavaScript and ZRender, a lightweight canvas library.
* **Vega:** [Vega](https://vega.github.io/vega/) (link resides outside IBM) defines itself as “visualization grammar,” providing support to customize visualizations across large datasets which are accessible from the web.
* **deck.gl:** It is part of Uber's open source visualization framework suite. [deck.gl](https://deck.gl/) (link resides outside IBM) is a framework, which is used for [exploratory data analysis](https://www.ibm.com/topics/exploratory-data-analysis) on big data. It helps build high-performance GPU-powered visualization on the web.

**pixel oriented visualization techniques**

**What is pixel oriented visualization?**

The basic idea of pixel-oriented visualization techniques is to represent as many data objects as possible on the screen at the same time by mapping each data value to a pixel of the screen and arranging the pixels adequately.

Pixel-oriented visualization techniques in data analytics refer to approaches that use pixels as the fundamental units for representing and displaying data. These techniques leverage the properties of individual pixels, such as color, brightness, and position, to encode and communicate information effectively. Here are a few examples of pixel-oriented visualization techniques commonly used in data analytics:

Heatmaps: Heatmaps use variations in color intensity to represent data values. Each pixel in the heatmap corresponds to a specific data point, and its color intensity reflects the value of that data point. Heatmaps are often used to visualize patterns, correlations, and density distributions in large datasets.

Pixel Bar Charts: Pixel bar charts represent data using horizontal or vertical bars made up of individual pixels. Each pixel within the bar corresponds to a specific data value, and its color or brightness indicates the magnitude of that value. Pixel bar charts are useful for visualizing categorical or discrete data in a compact and informative way.

Pixel Matrix: A pixel matrix visualization technique represents data as a grid of individual pixels. Each pixel in the grid corresponds to a single data point, and its color or brightness represents the value of that point. Pixel matrices are commonly used to visualize images, time series data, or multi-dimensional datasets.

Pixel Plots: Pixel plots are scatterplots where each data point is represented by a single pixel. The position of the pixel indicates the x and y coordinates of the data point, while the color or brightness represents an additional dimension of the data. Pixel plots are useful for visualizing large datasets and identifying clusters or patterns.

Pixel-based Geospatial Visualization: In geospatial data analysis, pixel-oriented visualization techniques can be used to represent geographic information. Each pixel on a map corresponds to a specific location, and its color or brightness can represent attributes such as population density, temperature, or elevation. This approach allows for the visualization of complex spatial patterns and relationships.

**Geometric projection techniques**

Geometric projection techniques are a set of visualization methods that transform three-dimensional (3D) geometric objects or scenes onto a two-dimensional (2D) plane. These techniques help represent complex spatial relationships and structures in a more manageable and interpretable format. Here are some commonly used geometric projection visualization techniques:

1. Perspective Projection: Perspective projection mimics how the human eye perceives objects in 3D space. It creates the illusion of depth by projecting points from a 3D scene onto a 2D plane using a vanishing point and a projection center. This technique is often used in architectural renderings, computer graphics, and virtual reality applications.
2. Orthographic Projection: Orthographic projection is a type of parallel projection that preserves the relative sizes and shapes of objects. It projects each point of a 3D object onto a 2D plane along parallel lines perpendicular to the plane. Orthographic projections are widely used in technical drawings, engineering, and industrial design.
3. Axonometric Projection: Axonometric projections are a family of parallel projections that maintain equal scaling along each axis. These projections include isometric, dimetric, and trimetric projections. Isometric projection, in particular, is commonly used to represent 3D objects in technical drawings and video games.
4. Oblique Projection: Oblique projection is a type of parallel projection that combines orthographic projection with a skew transformation. It allows for the inclusion of depth cues by tilting the object's faces relative to the projection plane. Oblique projection can be useful for emphasizing specific features of an object or for creating artistic renderings.
5. Stereographic Projection: Stereographic projection maps points from a sphere onto a plane. It is often used to visualize spherical objects or data, such as celestial bodies or crystal structures. Stereographic projections provide a way to view and analyze the spatial relationships of 3D data in a 2D format.
6. Cylindrical Projection: Cylindrical projection is a type of map projection that maps points on a sphere onto a cylinder, which is then unrolled onto a flat surface. This projection preserves angles but introduces distortions in scale and shape. Cylindrical projections, such as the Mercator projection, are commonly used for world maps and navigation purposes.
7. Conic Projection: Conic projection maps points on a sphere onto a cone, which is then unwrapped onto a flat surface. It preserves distances along certain lines (usually along a cone's meridians or parallels) but introduces distortions elsewhere. Conic projections are commonly used for regional or country-specific maps.

These geometric projection techniques provide different ways to represent and understand 3D objects or spatial relationships in 2D formats. The choice of projection depends on the specific requirements of the visualization task and the type of data being represented.

**Icon-based visualization techniques**

Icon-based visualization techniques utilize graphical icons or symbols to represent data or concepts in a visual representation. These techniques are effective for conveying information quickly and intuitively. Here are some commonly used icon-based visualization techniques:

1. Icon Arrays: Icon arrays use individual icons or pictograms to represent discrete quantities or values. The icons are arranged in a grid or matrix, with each icon representing a specific quantity. Icon arrays are useful for visualizing countable data or displaying proportions.
2. Bar Icons: Bar icons are similar to bar charts but use graphical icons instead of traditional bars. Each icon represents a certain value or category, and their lengths or sizes are proportional to the corresponding values. Bar icons are helpful for comparing values across different categories.
3. Progress Icons: Progress icons depict the progress or completion of a task or goal using visual icons. They often consist of a series of icons or symbols that gradually fill up or change appearance to indicate progress. Progress icons are commonly used in project management or task tracking applications.
4. Heatmap Icons: Heatmap icons combine icons with color gradients to represent data values or intensities. Icons are arranged in a grid, and their colors or sizes vary based on the data values they represent. Heatmap icons are useful for visualizing data patterns, clustering, or intensity levels.
5. Treemaps Icons: Treemap icons combine the concept of treemaps with icons. Treemaps use nested rectangles to represent hierarchical data, and icons can be incorporated into these rectangles to provide additional information or visual cues. Treemap icons are effective for displaying hierarchical data with associated icons.
6. Glyphs: Glyphs are simple, abstract icons or symbols that represent specific concepts or data points. They can be used to visualize categorical or qualitative data, such as representing different types of objects or attributes. Glyphs are commonly used in information visualization, data dashboards, and data encoding.
7. Thematic Icons: Thematic icons use specific graphical symbols or icons to represent thematic concepts or categories. These icons are carefully designed to visually represent the intended meaning or concept. Thematic icons are often used in maps, signage systems, or data visualizations to convey specific information or attributes.
8. Flow Icons: Flow icons visualize the flow or movement of data, resources, or processes. These icons depict the direction and intensity of the flow using graphical elements, such as arrows or connectors. Flow icons are useful for illustrating data flows, network traffic, or process diagrams.

These icon-based visualization techniques offer concise and intuitive ways to represent data or concepts. By leveraging the power of visual symbols, they enhance the understanding and interpretation of information in a visually engaging manner. The choice of technique depends on the nature of the data, the desired level of detail, and the intended message to be conveyed.

**Hierarchical visualization techniques**

Hierarchical visualization techniques are used to represent and explore hierarchical structures or relationships within data. These techniques help visualize nested or hierarchical data in a way that facilitates understanding and analysis. Here are some commonly used hierarchical visualization techniques:

1. Tree Diagrams: Tree diagrams, also known as dendrograms or hierarchy diagrams, use a tree-like structure to represent hierarchical relationships. Each node represents a category or subcategory, and the branches depict the parent-child relationships. Tree diagrams are widely used in fields like organizational charts, file systems, and biological classifications.
2. Sunburst Charts: Sunburst charts are radial diagrams that represent hierarchical data in a circular format. The innermost circle represents the root category, and subsequent rings represent subcategories. The size of each ring or arc is proportional to the data values associated with the category or subcategory. Sunburst charts are effective for visualizing hierarchical data with multiple levels.
3. Treemaps: Treemaps divide the screen or a designated area into nested rectangles, with each rectangle representing a category or subcategory. The size or color of each rectangle represents a data attribute, allowing for easy comparison and identification of patterns. Treemaps are commonly used for visualizing hierarchical data with varying attribute values.
4. Nested Pie Charts: Nested pie charts are a hierarchical variation of traditional pie charts. Each level of the hierarchy is represented by a nested pie, with slices within each pie representing subcategories. The size or angle of each slice corresponds to the proportion or value associated with the category or subcategory. Nested pie charts provide an intuitive representation of hierarchical relationships.
5. Collapsible/Expandable Nodes: This technique allows users to interactively explore hierarchical data by collapsing or expanding nodes. The initial view shows only the top-level categories, and users can expand a node to reveal its subcategories or collapse it to hide the details. This technique is often used in interactive tree diagrams or folder structures.
6. Indented Lists: Indented lists represent hierarchical relationships through indentation. Each level is indented further to the right, making it visually apparent which categories are nested within others. Indented lists are commonly used for displaying hierarchical data in text-based formats, such as outlines or file directories.
7. Network Graphs: Network graphs can also be used to visualize hierarchical relationships. In this case, nodes represent categories or entities, and edges represent the connections or relationships between them. By using different colors or shapes for nodes at different levels, hierarchical relationships can be visually represented in the network graph.
8. Radial Tree Diagrams: Radial tree diagrams are similar to traditional tree diagrams, but they use a radial layout instead of a horizontal or vertical layout. The root category is placed at the center, and the subsequent levels branch out in a circular pattern. Radial tree diagrams provide a compact and visually appealing representation of hierarchical structures.

These hierarchical visualization techniques offer different ways to represent and explore hierarchical relationships within data. The choice of technique depends on factors such as the complexity of the hierarchy, the amount of data, and the desired visual representation.

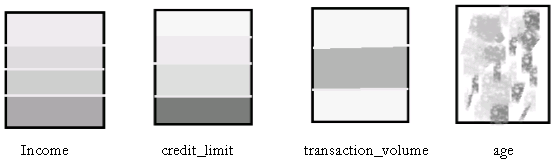
**EXPLAIN DIFFERENT TYPES OF DATA VISUALIZATION TECHNIQUES:**

**Pixel oriented visualization techniques:**

* A simple way to visualize the value of a dimension is to use a pixel where the color of the pixel reflects the dimension’s value.
* For a data set of m dimensions pixel oriented techniques create m windows on the screen, one for each dimension.
* The m dimension values of a record are mapped to m pixels at the corresponding

position in the windows.

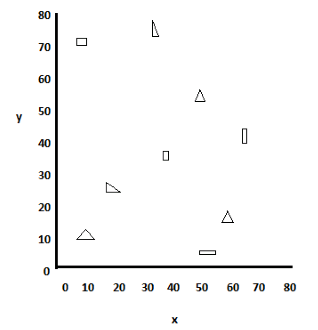
* The color of the pixel reflects other corresponding values.
* Inside a window, the data values are arranged in some global order shared by all windows
* Eg: All Electronics maintains a customer information table, which consists of 4 dimensions: income, credit\_limit, transaction\_volume and age. We analyze the correlation between income and other attributes by visualization.
* We sort all customers in income in ascending order and use this order to layout the customer data in the 4 visualization windows as shown in fig.
* The pixel colors are chosen so that the smaller the value, the lighter the shading.
* Using pixel based visualization we can easily observe that credit\_limit increases as income increases customer whose income is in the middle range are more likely to purchase more from All Electronics, these is no clear correlation between income and age.



**Fig: Pixel oriented visualization of 4 attributes by sorting all customers in income Ascending order**.

**Geometric Projection visualization techniques**

* A drawback of pixel-oriented visualization techniques is that they cannot help us much in understanding the distribution of data in a multidimensional space.
* Geometric projection techniques help users find interesting projections of multidimensional data sets.
* A scatter plot displays 2-D data point using Cartesian co-ordinates. A third dimension can be added using different colors of shapes to represent different data points.
* Eg. Where x and y are two spatial attributes and the third dimension is represented by different shapes
* Through this visualization, we can see that points of types “+” &”X” tend to be collocated.



**Fig: visualization of 2D data set using scatter plot**

**Icon based visualization techniques:-**

Visualization of the data values as features of icons

Typical visualization methods

* Chernoff Faces
* Stick Figures

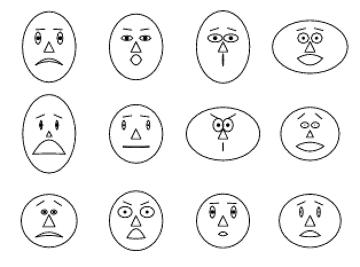
General techniques

* Shape coding: Use shape to represent certain information encoding
* Color icons: Use color icons to encode more information
* Tile bars: Use small icons to represent the relevant feature vectors in document retrieval

Chernoff Faces

A way to display variables on a two-dimensional surface, e.g., let x be eyebrow slant, y be eye size, z be nose length, etc.

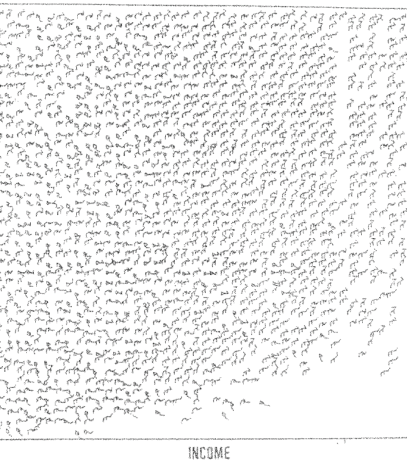
The figure shows faces produced using 10 characteristics–head eccentricity, eye size, eye spacing, eye eccentricity, pupil size, eyebrow slant, nose size, mouth shape, mouth size, and mouth opening): Each assigned one of 10 possible values.



Chernoff Faces

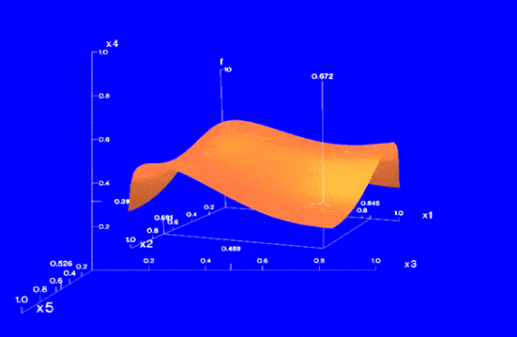
Stick Figure

* A census data figure showing age, income, gender, education
* A 5-piece stick figure (1 body and 4 limbs w. different angle/length)
* Age, income are indicated by position of the figure.
* Gender, education are indicated by angle/length.



**Hierarchical visualization techniques** (i.e. subspaces)

The subspaces are visualized in a hierarchical manner.

For a large data set of high dimensionality, it would be difficult to visualize all dimensions at the same time. Hierarchical visualization techniques partition all dimensions into subsets (i.e., subspaces). The subspaces are visualized in a hierarchical manner “Worlds-within-Worlds,” also known as n-Vision, is a representative hierarchical visualization method. To visualize a 6-D data set, where the dimensions are F,X1,X2,X3,X4,X5. We want to observe how F changes w.r.t. other dimensions. We can fix X3,X4,X5 dimensions to selected values and visualize changes to F w.r.t. X1, X2  
  


**Visualizing Complex Data and Relations**

Most visualization techniques were mainly for numeric data. Recently, more and more non-numeric data, such as text and social networks, have become available. Many people on the Web tag various objects such as pictures, blog entries, and product reviews.

A tag cloud is a visualization of statistics of user-generated tags. Often, in a tag cloud, tags are listed alphabetically or in a user-preferred order.

The importance of a tag is indicated by font size or color.

  
  
Visualizing complex data and relations can be a challenging task, but it is crucial for gaining insights and understanding patterns within the data. Here are some approaches and techniques for visualizing complex data and relations:

1. Scatter plots: Scatter plots are useful for visualizing the relationship between two variables. Each data point is represented by a point on the plot, with one variable mapped to the x-axis and another to the y-axis. By examining the distribution and clustering of points, you can identify patterns or correlations.
2. Network diagrams: Network diagrams, also known as graph visualizations, are ideal for representing relationships between entities. Nodes represent entities, such as people or objects, and edges represent connections or interactions between them. Network diagrams can be used to visualize social networks, organizational structures, or any other interconnected system.
3. Heatmaps: Heatmaps are effective for displaying large amounts of data in a tabular format. They use color gradients to represent values, allowing you to identify patterns or clusters within the data. Heatmaps are commonly used in fields like genomics, finance, and weather forecasting.
4. Sankey diagrams: Sankey diagrams visualize flow or movement between different states or categories. They are particularly useful for understanding processes or systems involving inputs, outputs, and transitions. Sankey diagrams show the flow of data, energy, money, or any other quantifiable resource.
5. Tree maps: Tree maps display hierarchical data structures by representing each level of the hierarchy as nested rectangles. The size of each rectangle can be proportional to a certain attribute or value, making it easy to compare and visualize different levels or categories.
6. Parallel coordinates: Parallel coordinates plots are useful for visualizing multivariate data. Each variable is represented by a vertical axis, and data points are connected by lines that intersect these axes. Parallel coordinates plots allow you to identify patterns, relationships, or clusters in high-dimensional data.
7. Chord diagrams: Chord diagrams illustrate relationships or flows between entities in a circular layout. The thickness of the chords represents the strength or magnitude of the relationship, while the colors can indicate different categories or groups. Chord diagrams are commonly used to visualize network traffic, migration patterns, or trade flows.
8. Geospatial visualizations: Geospatial visualizations use maps to represent data in a spatial context. They are effective for displaying location-based data, such as demographics, sales figures, or disease outbreaks. Techniques like choropleth maps, dot maps, or heat maps can be employed to visualize geospatial data effectively.
9. Time series plots: Time series plots are ideal for visualizing data that changes over time. They display data points along a timeline, allowing you to observe trends, seasonality, or patterns. Line charts, area charts, or candlestick charts are commonly used to represent time series data.