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 the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
Uses of data visualization
¬ Powerful way to explore data with presentable results.
¬ Primary use is the preprocessing portion of the data mining process.
¬ Supports in data cleaning process by finding incorrect and missing values.
General Data visualization Techniques: ¬ Box plots
¬ Histograms ¬ Heat maps ¬ Charts
¬ Tree maps

Data visualization is one of the steps of the data science process, which states that after data has been collected, processed and modeled, it must be visualized for conclusions to be made.

Data visualization Techniques:
Data visualization aims to communicate data clearly and effectively through graphical representation. Data visualization has been used extensively in many applications—for example, at work for reporting, managing business operations, and tracking progress of tasks.
More popularly, we can take advantage of visualization techniques to discover data relationships that are otherwise not easily observable by looking at the raw data.

Data visualization Techniques:
1. Pixel Oriented Visualization technique
2. Geometric Projection Visualization technique-
a. Scatter plot matrices b. Hyper slice
c. Parallel coordinates
3. Icon based visualization techniques
4. Hierarchical 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 positions in the windows. The colors of the pixels reflect the corresponding values.
- ¬ Inside a window, the data values are arranged in some global order shared by all windows.

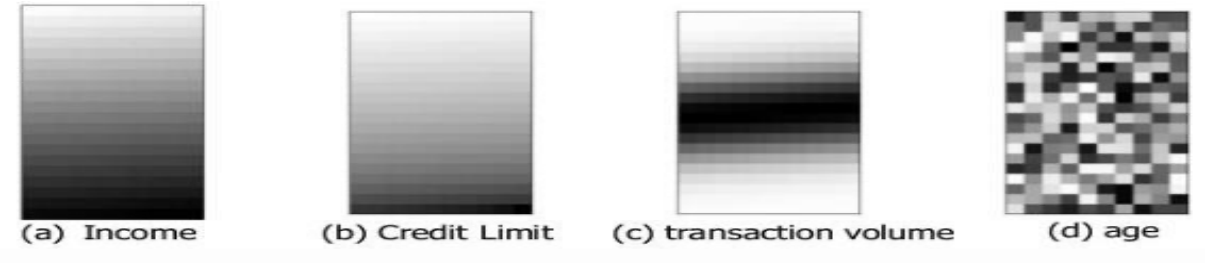
- ¬ The global order may be obtained by sorting all data records in a way that's meaningful for the task at hand.

Pixel-based visualizations use that approach and are capable of displaying large amounts of data on a single screen.

Case Study:

¬All Electronics maintains a customer information table, which consists of 4 dimensions: income, transaction_volume and age.

- ¬ We analyse 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.



Geometric Projection visualization techniques:

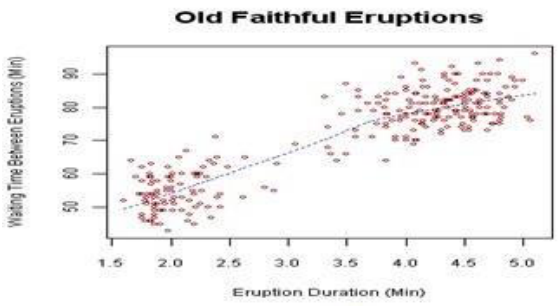
A drawback of pixel-oriented visualization techniques is that they cannot help us much in understanding the distribution of data in a multidimensional space.

Geometric projection techniques help users find interesting projections of multidimensional data sets.

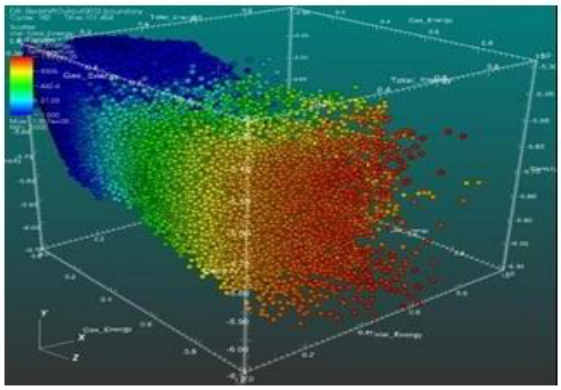
Geometric projection techniques are a good choice for finding outliers and correlation between attributes in multivariate data. A geometric projection technique does this by using transformations and projections of the data. When using large data sets a clustering algorithm is usually necessary to apply before the visualization technique to avoid cluttered and unclear data caused by the too much information. Some widely used geometric projection techniques are:

Scatter plots:

A scatter plot is one of the most common visualization techniques and can be visualized both in 3D and 2D. The scatter plot visualizes different attributes of the data on the x,y axis for 2D visualizations and also along the z-axis in 3D. Scatter plots are usable to find correlations between attributes in arbitrary small data sets. If the data set gets too big or contains too many attributes the scatter plot gets cluttered and hard to interpret.



2D Scatter Plot



3D Scatter Plot

HyperSlice:

HyperSlice is a new method for the visualization of scalar functions of many variables. With this method the multi-dimensional function is presented in a simple and easy to understand way in which all dimensions are treated identically. The central concept is the representation of a multi-dimensional function as a matrix of orthogonal two-dimensional slices. These two- dimensional slices lend themselves very well to interaction via direct manipulation, due to a one to one relation between screen space and variable space.

Parallel coordinates:

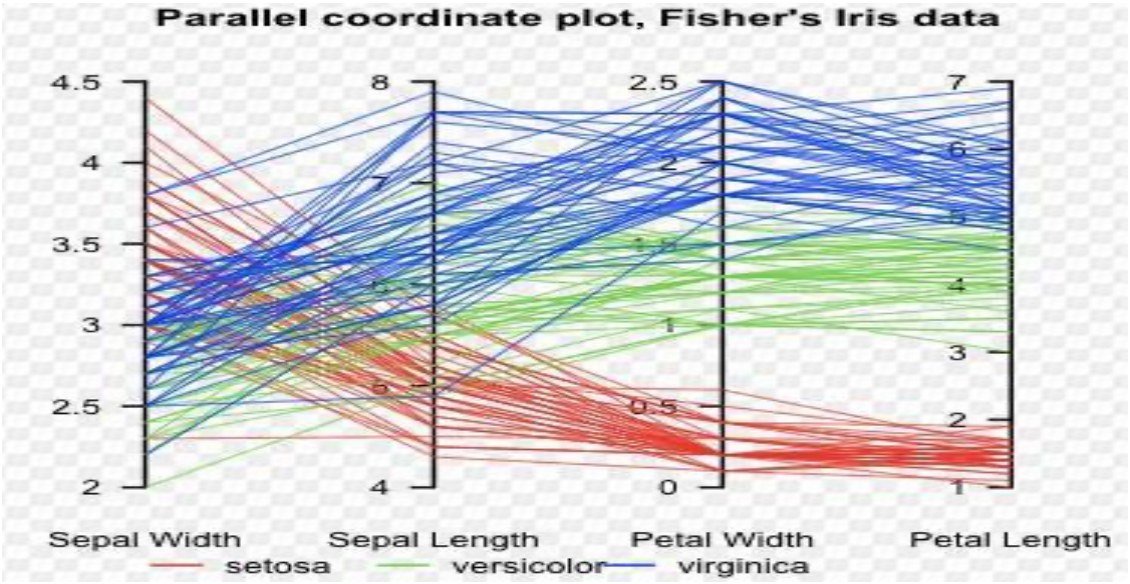
¬To visualize n-dimensional data points, the parallel coordinates technique draws n equally spaced axes, one for each dimension, parallel to one of the display axes.

- ¬ A data record is represented by a polygonal line that intersects each axis at the point corresponding to the associated dimension value.
- ¬ A major limitation of the parallel coordinate's technique is that it

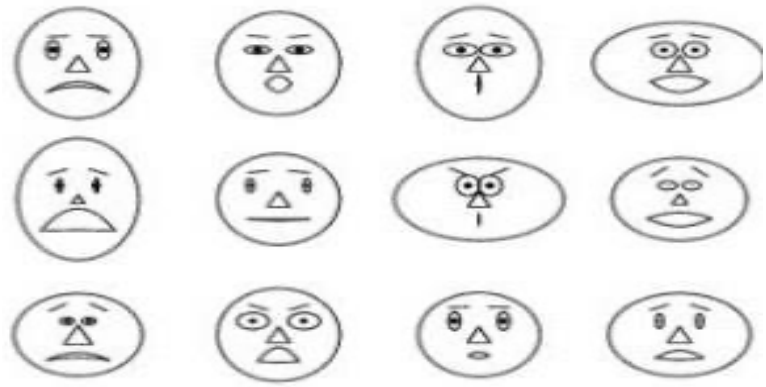
cannot effectively show a data set of many records.

Icon-based visualization techniques:

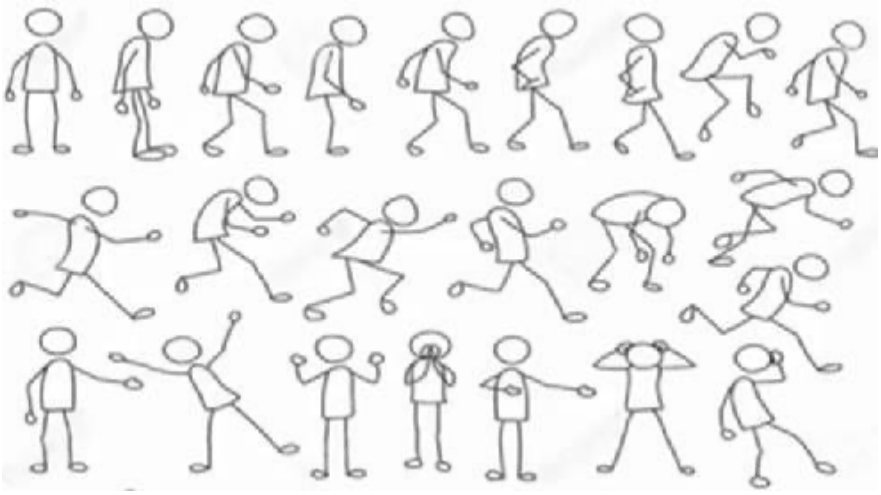
Icon-based techniques visualize data by changing the properties of an icon or glyph according to the data. An early version was Chernoff faces where data is mapped to different face parts as nose, mouth, eyes and more. For example how rich people are can be mapped to the mouth of the Chernoff



face. Rich people represented by a happy mouth and and poor people by a sad mouth. Other methods are:



Stick figures:
It maps multidimensional data to five –piece stick figure, where each figure has 4 limbs and a body.
Two dimensions are mapped to the display (x and y) axes and the remaining dimensions are mapped to the angle and/or length of the limbs.



Hierarchical Visualization Techniques:
Hierarchical visualization techniques are techniques, whose domain data structure and type of information are, respectively, tree and hierarchical information. There are two basic branches of visualization techniques for hierarchies. The first is based on a node-edge graph-layout approach which focuses attention on the structure and relationships, and the second on space-filling approaches, which focus attention on the relative sizes of nodes in the hierarchy.

- ¬ The visualization techniques discussed so far focus on visualizing multiple dimensions simultaneously.
- ¬ However, 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.

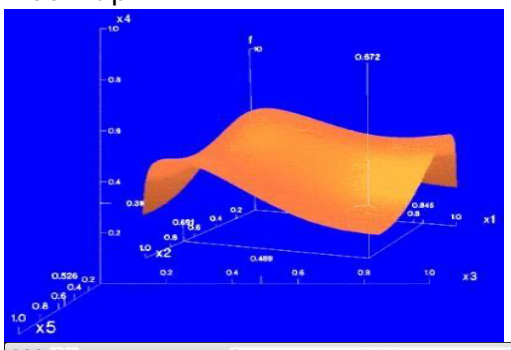
- ¬ Suppose we want to visualize a 6-D data set, where the dimensions are F,X1, : : : ,X5.

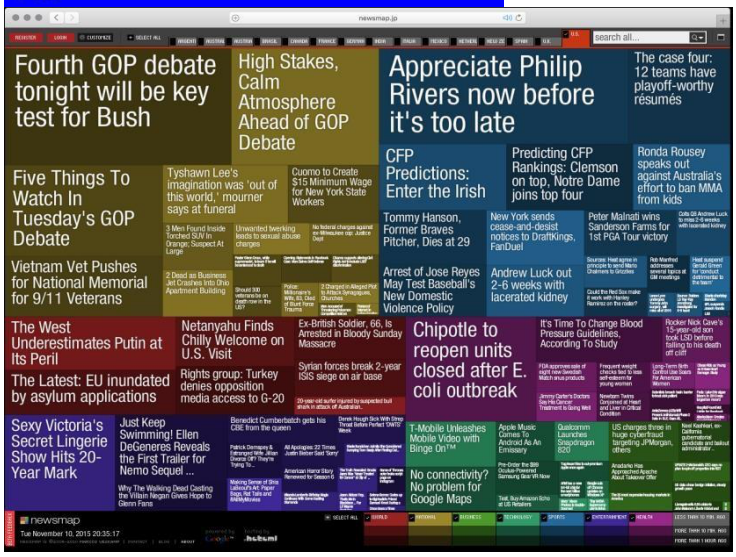
Given more dimensions, more levels of worlds can be used, which is why the method is called “worlds-within-worlds.”

Worlds - Within – Worlds

As another example of hierarchical visualization methods, tree-maps display hierarchical data as a set of nested rectangles. For example, a tree-map visualizing Google news stories.

Tree Map





All news stories are organized into seven categories, each shown in a large rectangle of a unique color. Within each category (i.e., each rectangle at the top level), the news stories are further partitioned into smaller subcategories.

Data visualization choices:

Five factors that influence data visualization choices:

Audience: It’s important to adjust data representation to the specific target audience.

Content: The type of data you are dealing with will determine the tactics. Context: You can use different data visualization approaches and read data depending on the context.

Dynamics: There are various types of data, and each type has a different rate of change.

Purpose: The goal of data visualization affects the way it is implemented. In order to make a complex analysis, visualizations are compiled into dynamic and controllable dashboards that work as visual data analysis techniques and tools.

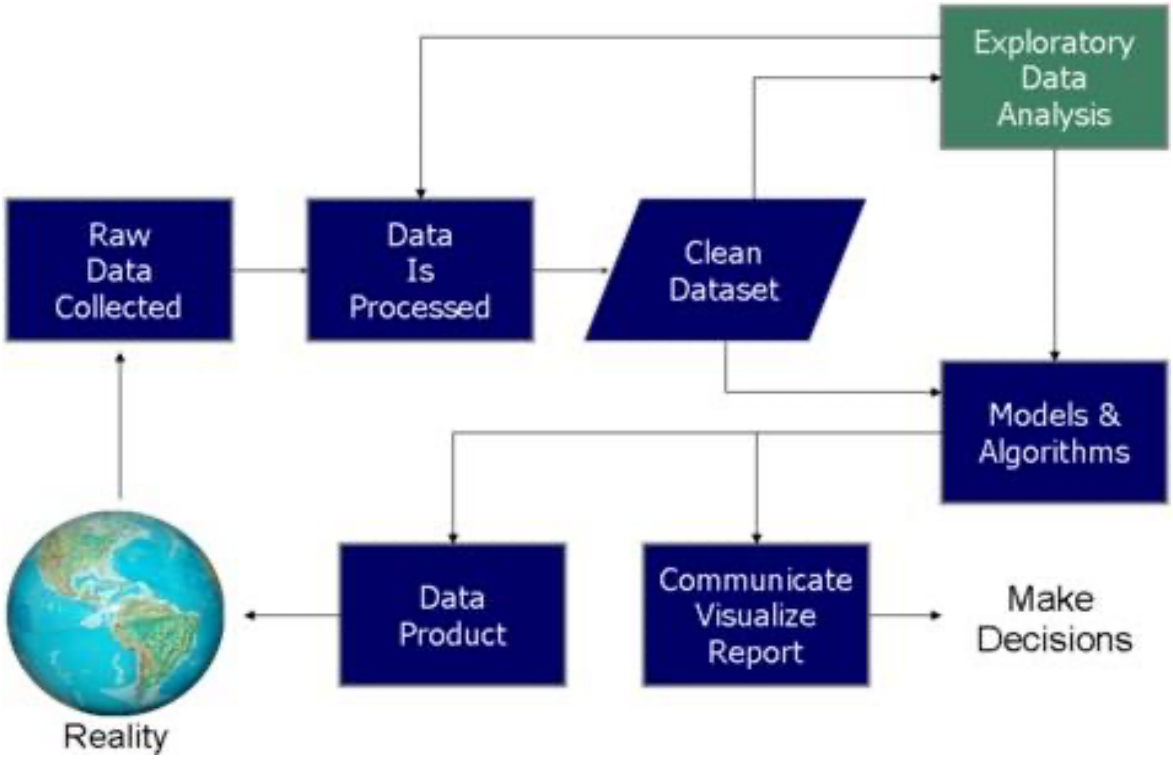
Tools for Data visualization:

Data visualization tools for different types of users and purposes.

Tableau is one of the leaders in this field. A user-friendly interface and a rich library of interactive visualizations, Tableau stands out for its powerful capabilities. The platform provides large integration options including My SQL, Teradata, Hadoop and Amazon Web Services. This platform to derive meaning from data and use insights for effective storytelling.
R and Python are well-equipped for data visualization. Customizing graphics is easier and more intuitive in R with the help of ggplot2 than in Python with Matplotlib. The Seaborn library helps to overcome this, and offers good standard solutions which get by with relatively few lines of code.

- Tableau
- R
- Python
- Plotly
- IBM Watson Analytics

Plotly is one of the most popular platforms in this category. It's more complex than Tableau, however, comes with analytics perks. With this visualization tool, you can create charts using R or Python, build custom data analytics
IBM Watson Analytics is known for its NLP capabilities. The platform literally supports conversational data control a longside strong dashboard building and data reporting tools.
Tools for complex data visualization:
The growing adoption of connected technology places a lot of opportunities before the companies and organizations. To deal with large volumes ofmulti- source often unstructured data, businesses search for more complex visualization and analytics solutions. This category includes Power BI, Kibana and Grafana.
Power BI is exceptional for its highly intuitive drag-and-drop interface, short learning curve and large integration capabilities, including Salesforce and MailChimp.
Kibana is the part of the Elastic Stack that turns data into visual insights. It's built on and designed to work on Elasticsearch data only. This exclusivity, however, does not prevent it from being one of the best data visualization tools for log data.
Grafana a professional data visualization and analytic tool that supports up to 30 data sources, including AWS, Elastic search and Prometheus. Grafana is more flexible in terms of integrations compared to Kibana, each of the systems works best with its own type of data.
Data Visualization Process:
Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets.
Fig: Data visualization Process


Exploratory Data Analysis
Raw Data Collected
Data Is Processed
Clean Dataset
Models & Algorithms
Reality
Data Product
Communicate Visualize Report
Make Decisions