Data Visualization and interpretation

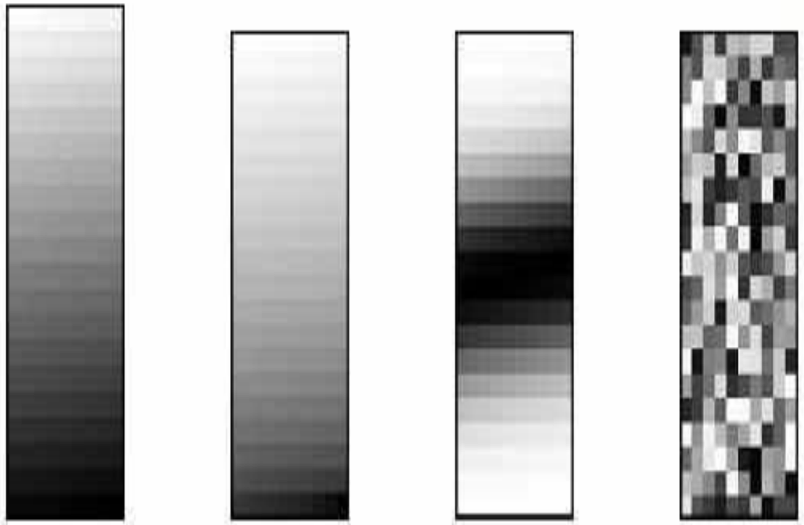
VISUALIZATION TECHNIQUES

Data visualization aims to communicate data clearly and effectively through graphical representation.

Visualization techniques are of increasing importance in exploring and analyzing large amounts of multidimensional information.

Pixel Oriented visualization techniques

- Visualize the value of a dimension using 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: ***AllElectronics*** maintains a customer information table, which consists of 4 dimensions: ***income, credit_limit, transaction_volume*** and ***age***.
- To analyze the correlation between ***income*** and other attributes by visualization, sort all customers in **income-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.

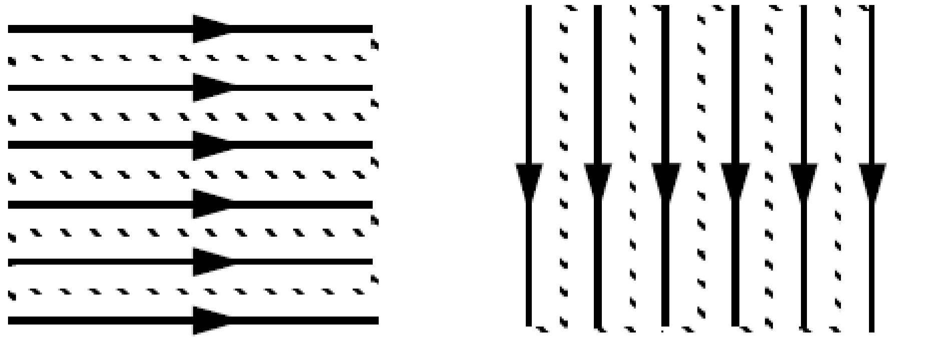


(a) income (b) credit_limit (c) transaction_volume (d) age

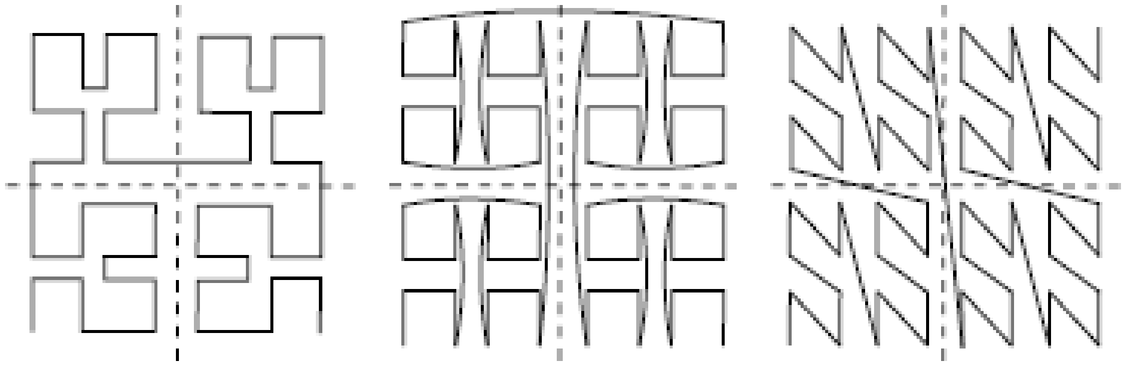
credit_limit increases as ***income*** increases. Customers whose ***income*** is in the middle range are more likely to purchase more from ***AllElectronics***. There is no clear correlation between ***income*** and ***age***.

Pixel Oriented visualization techniques

- **How to find meaningful arrangements of the pixels on the screen?**
- To save space and show the connections among multiple dimensions, space filling is often done in a circle segment. Lay out the data records in a **space-filling curve** to fill the windows.
- The basic idea of spacefilling curves is to provide a continuous curve which passes through every point of a regular spatial region (e.g., a square)
- A **space-filling curve** is a curve with a range that covers the entire n-dimensional unit hypercube. Since the visualization windows are 2-D, we can use any 2-D space-filling curve.



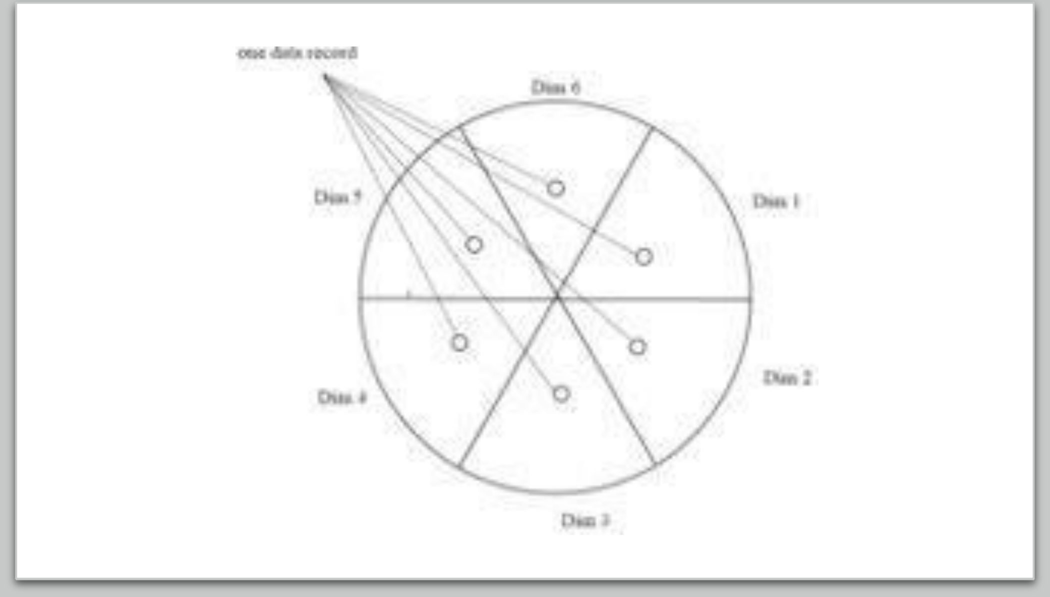
a. line-by-line b. column-by-column

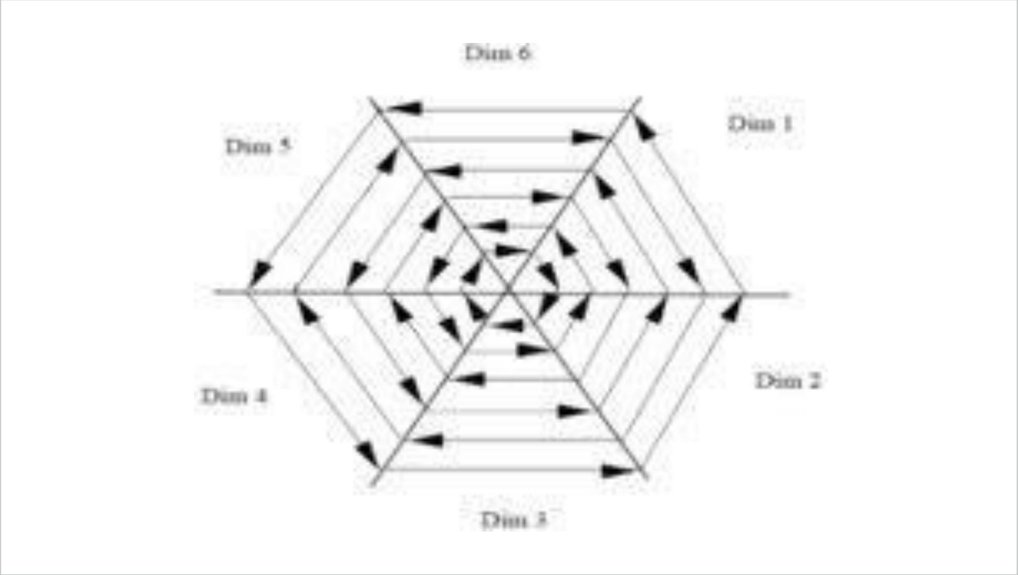


(a) Hilbert curve (b) Gray code (c) Z-curve

Pixel Oriented visualization techniques

- Pixel oriented techniques may be divided into **query-independent techniques** which directly visualize the data (or a certain portion of it) and **query-dependent techniques** which visualize the data in the context of a specific query
- **DRAWBACK:** cannot help us much in understanding the distribution of data in a multidimensional space.

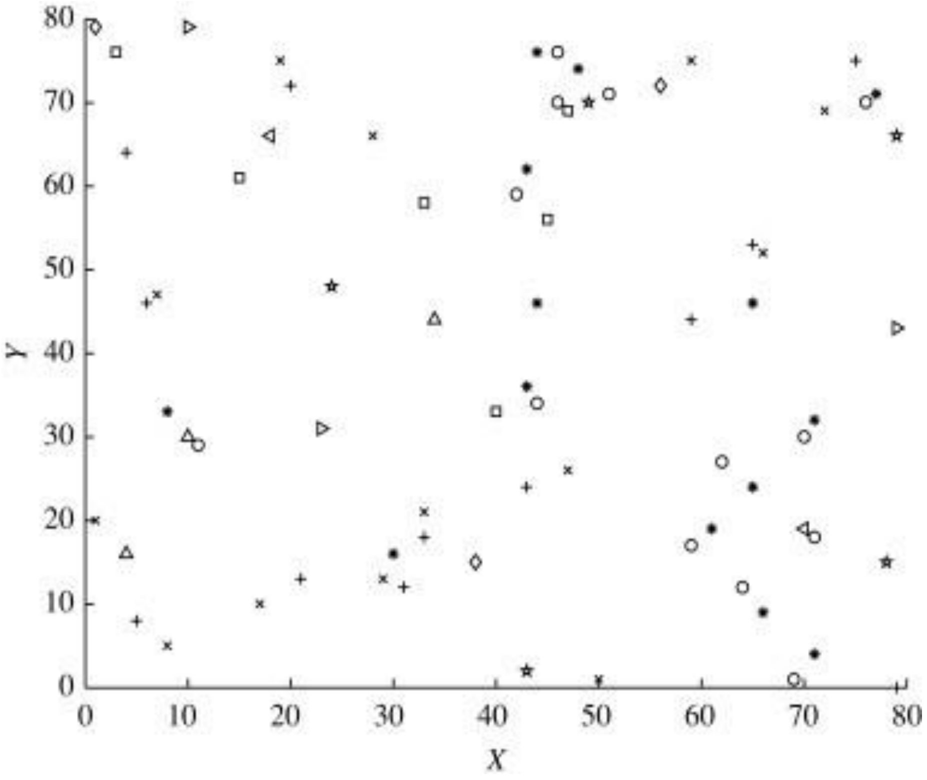




Geometric projection visualization techniques

- Geometric projection techniques help users find **interesting projections of multidimensional data sets.**
- It tries to visualize a high-dimensional space on a 2-D display.
- Methods
  - Direct visualization
  - Scatterplot and scatterplot matrices
  - Landscapes
  - Projection pursuit technique: Help users find meaningful projections of multidimensional data
  - Prosection views
  - Hyperslice
  - Parallel coordinates

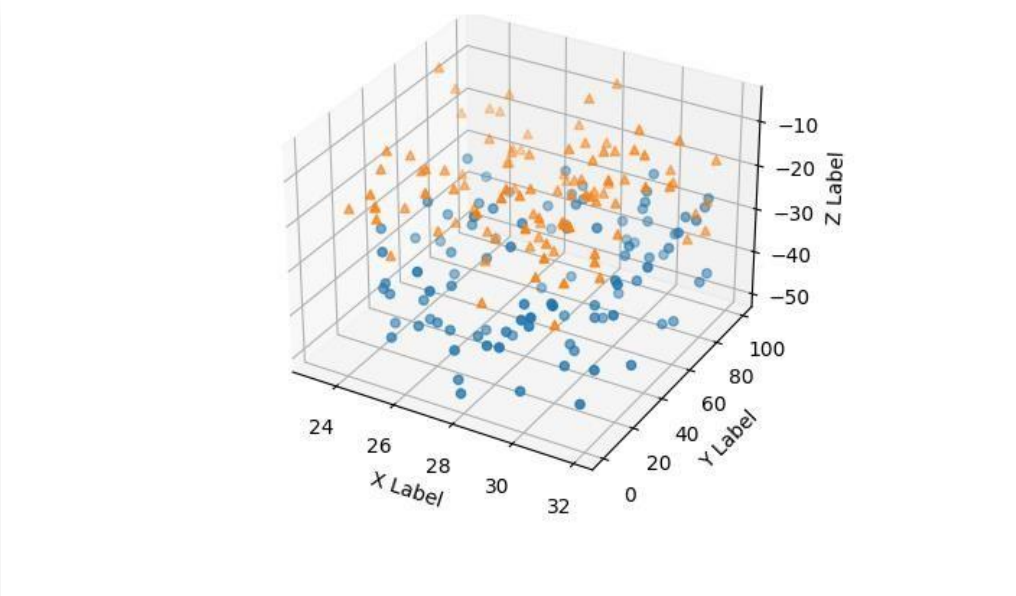
A **scatter plot** displays 2-D data point using Cartesian co-ordinates. A third dimension can be added using different colors of shapes to represent different data points.

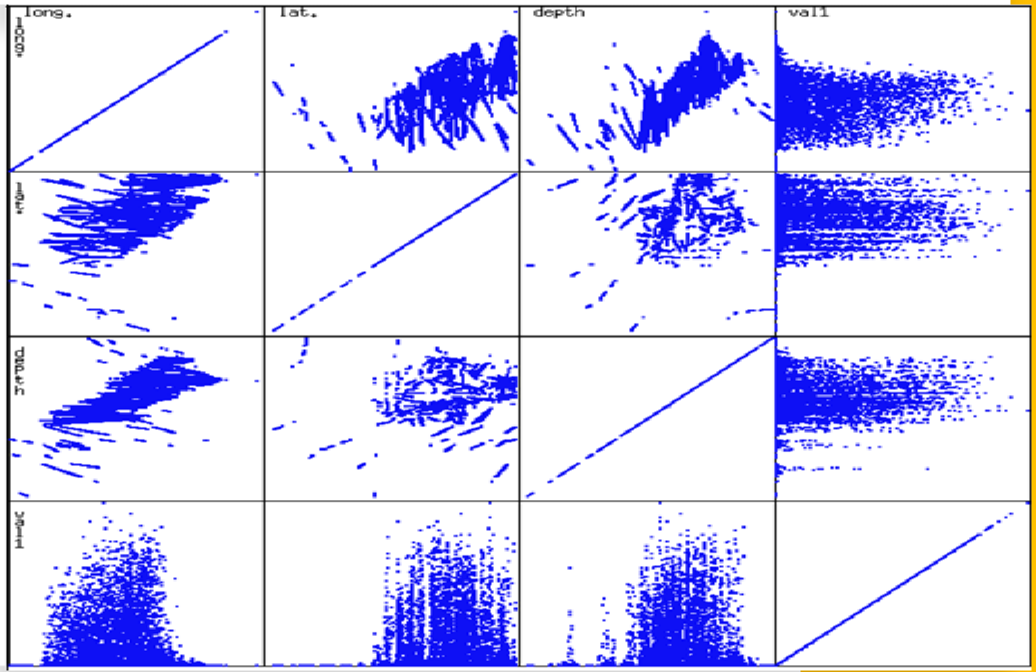


x and y are two spatial attributes and the third dimension is represented by different shapes

Geometric projection visualization techniques: Scatter Matrix

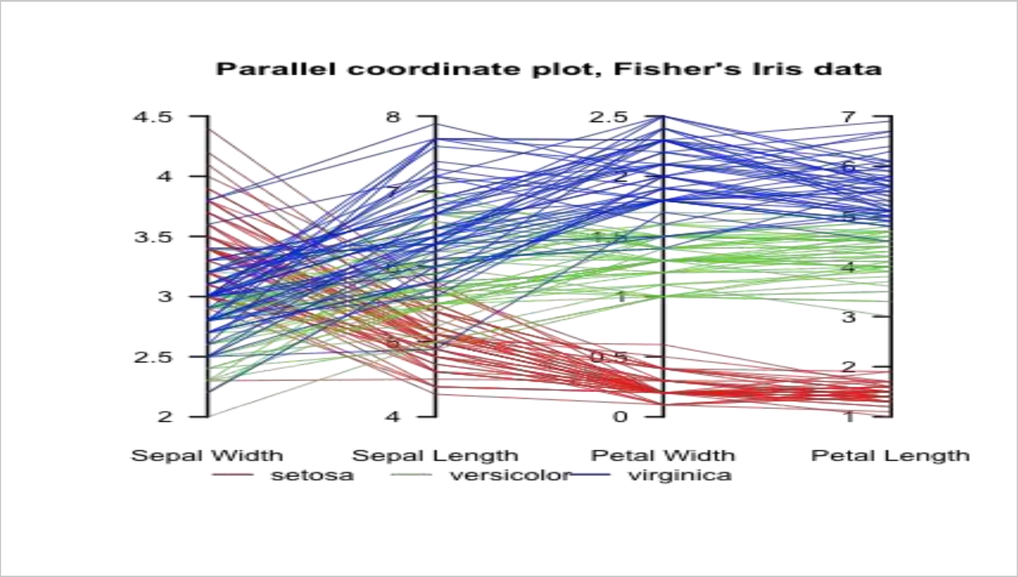
- A 3-D scatter plot uses three axes in a Cartesian coordinate system. If it also uses color, it can display up to 4-D data points.
- For data sets with **more than four dimensions, scatter plots are usually ineffective.**
- The **scatter-plot matrix technique** is a useful extension to the scatter plot. For an ndimensional data set, a scatter-plot matrix is an ***nXn*** grid of 2-D scatter plots that provides a visualization of each dimension with every other dimension.
- The scatter-plot matrix becomes less effective as the dimensionality increases.

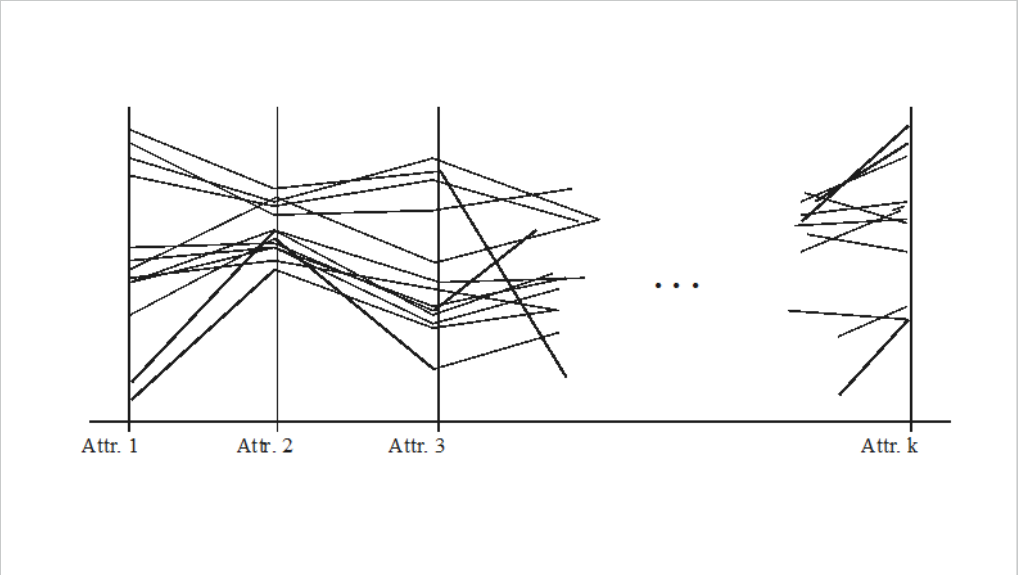




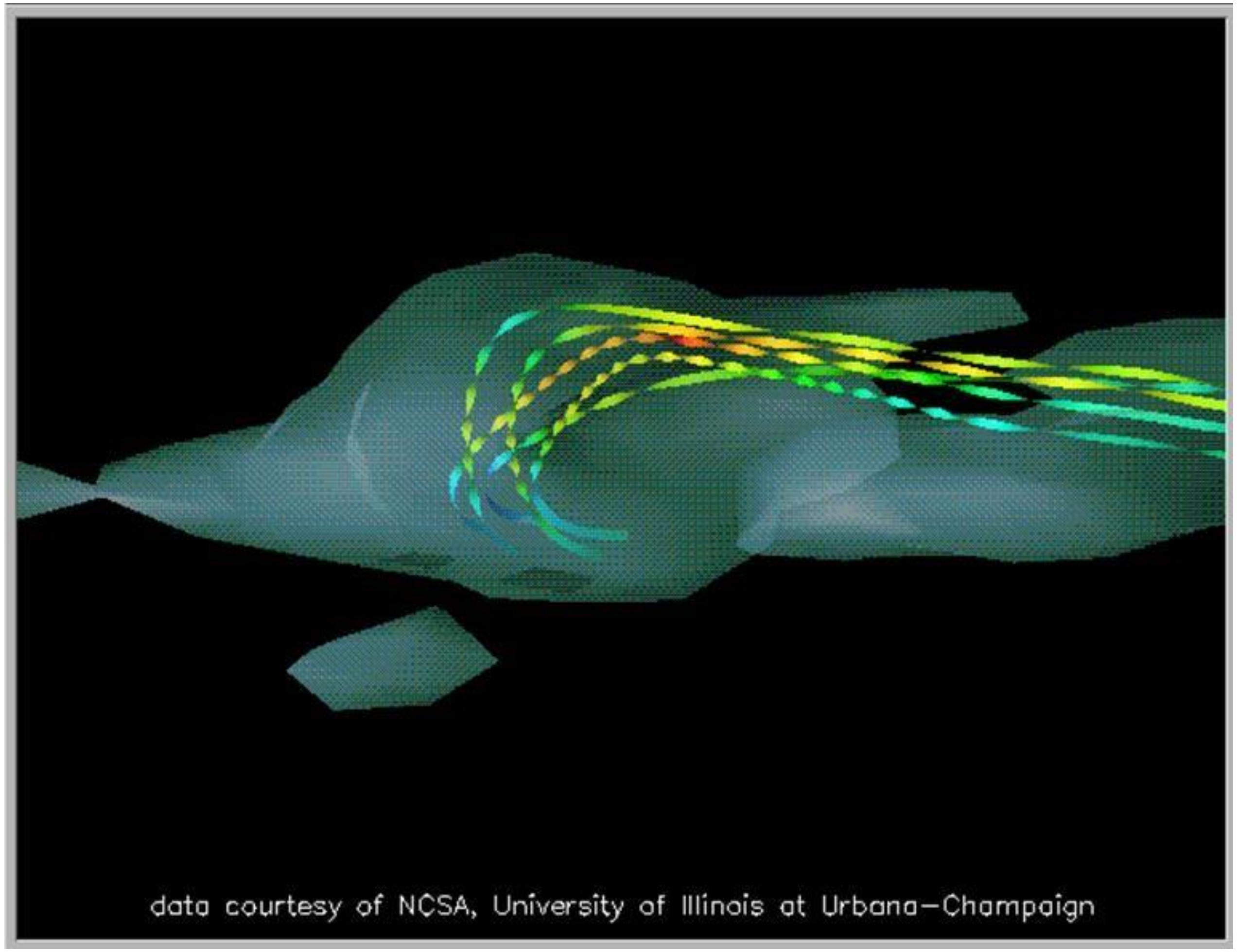
Geometric projection visualization techniques: Parallel Cordinates

- **Parallel coordinates** technique draws n equally spaced axes, one for each dimension, parallel to one of the display axes.
- The axes are scaled to the [minimum, maximum]: range of the corresponding attribute
- A data record is represented by a polygonal line that intersects each axis at the point corresponding to the associated dimension value
- Parallel coordinates, can handle higher dimensionality.
- A major limitation of the parallel coordinates technique is that it **cannot effectively show a data set of many records.** Even for a data set of several thousand records, visual clutter and overlap often reduce the readability of the visualization and make the patterns hard to find.



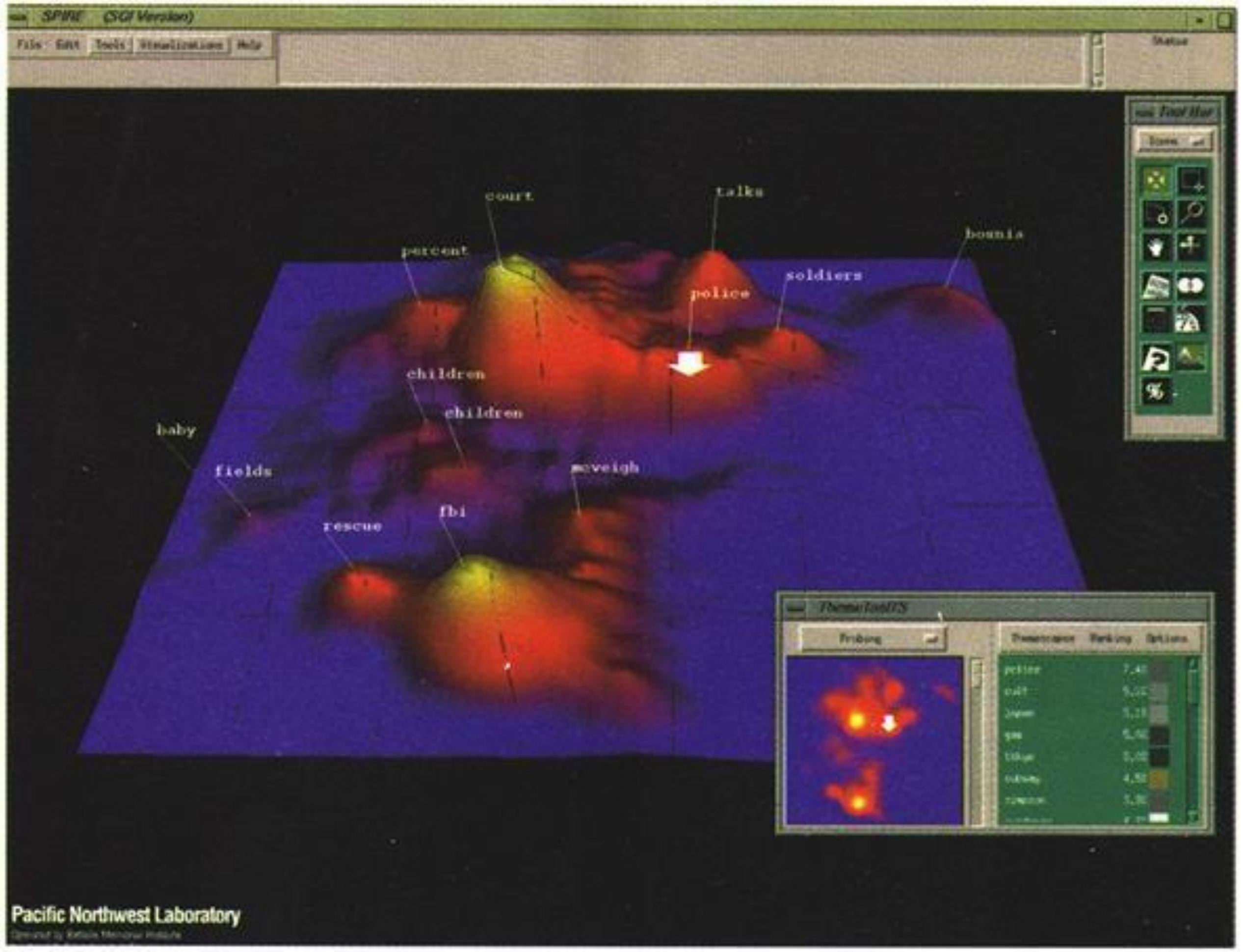


Geometric projection visualization techniques: Direct Data Visaulisation

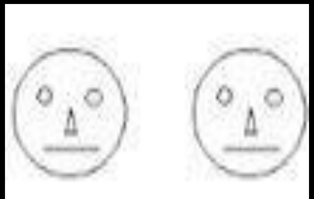


Geometric projection visualization techniques: Landscapes

- Visualization of the data as perspective landscape
- The data needs to be transformed into a (possibly artificial) 2D spatial representation which preserves the characteristics of the data



Icon based visualization techniques



- Icon-based visualization techniques uses small icons to represent multidimensional data values.
- They map data variables onto geometric (e.g., shape, size, orientation) and non-geometric (e.g., color and texture) visual attributes
- Here each dimension is mapped to the properties of a face icon, i.e, shape of face, nose, mouth and eyes. This technique can deal with multidimensional data, but its interpretation require training and each dimension may be treated differently by the human visual perception.
- 2 popular icon based techniques:-
  1. **Chern off faces: -** They display multidimensional data of up to 18 variables as a cartoon human face. data is mapped to Components of the face, such as the eyes, ears, mouth, and nose, represent values of the dimensions by their shape, size, placement, and orientation.
  2. **Stick figures:** icon-based techniques generating textures. It maps multidimensional data to five –piece stick figure, where each figure has 4 limbs and a body.2 dimensions are mapped to the display axes and the remaining dimensions are mapped to the angle and/ or length

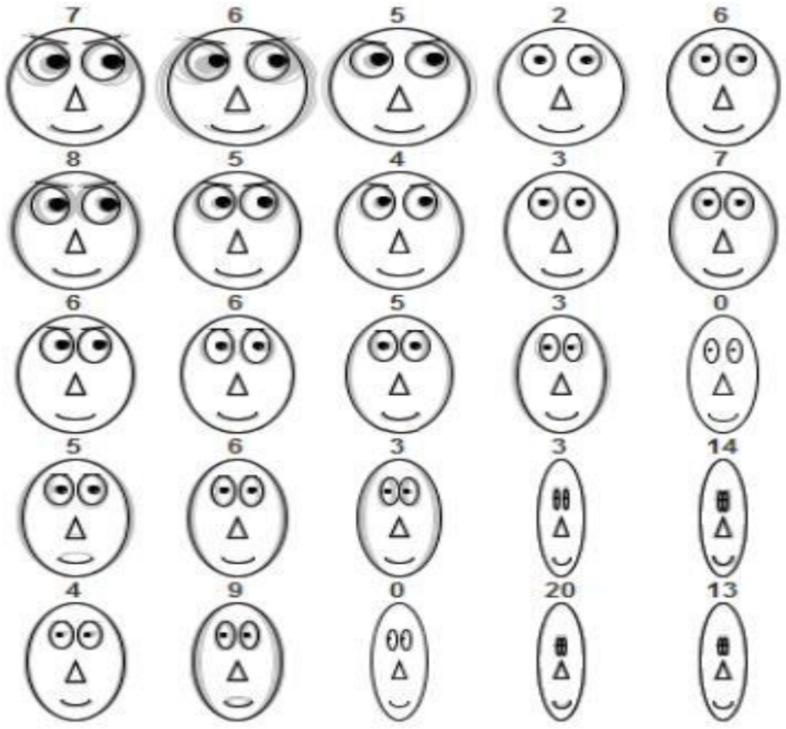
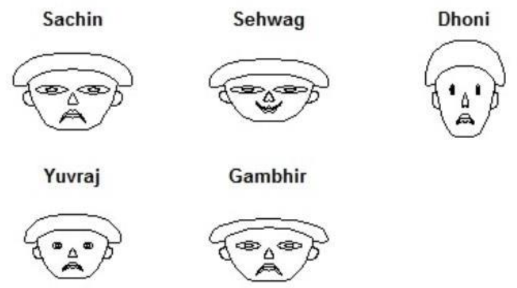


Fig. 4. Iris data set visualized by Chernoff faces over a SOM grid

CHERNOFFF FACES

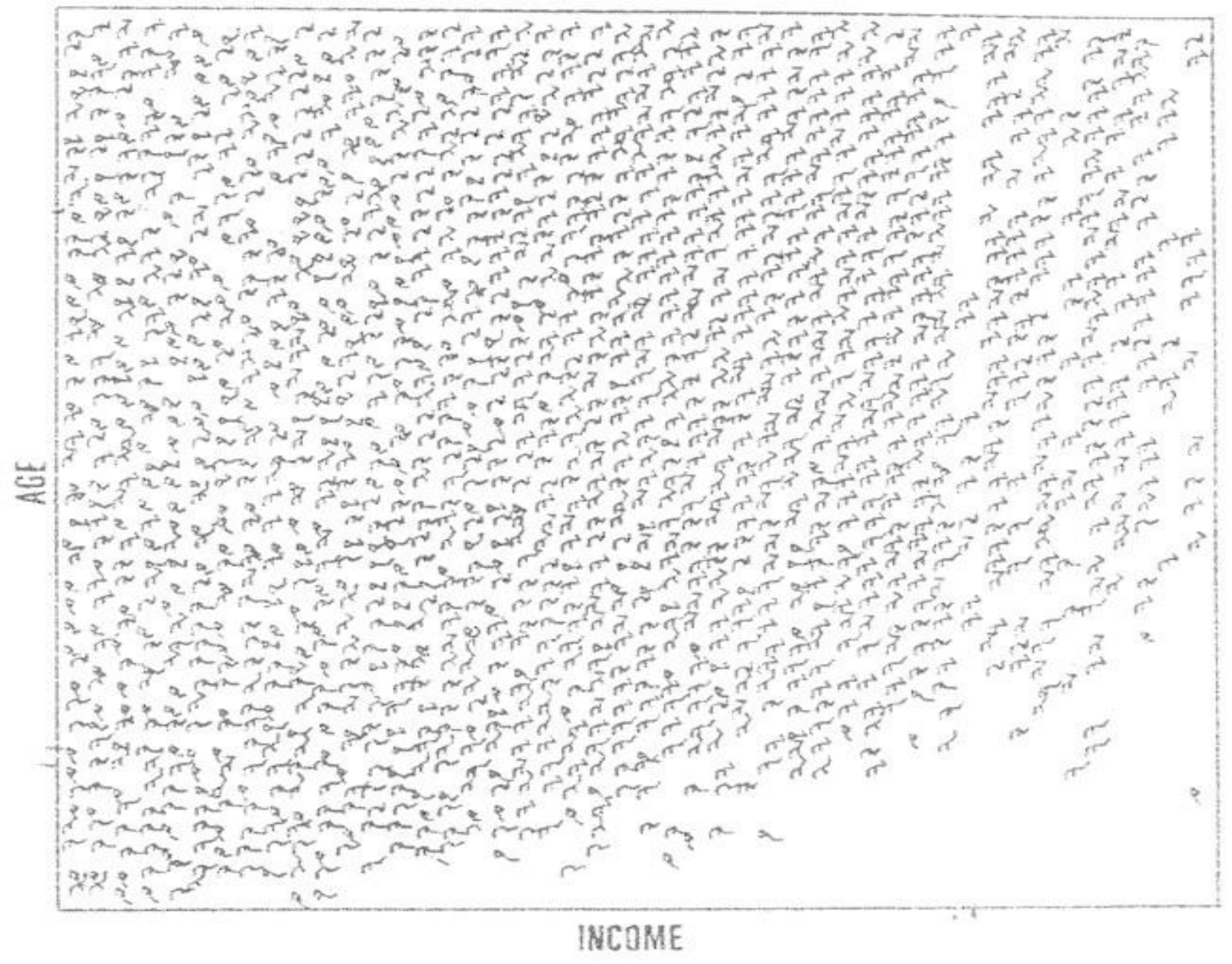
- **Chern off faces is a** way to display variables on a two-dimensional surface, e.g., let x be eyebrow slant, y be eye size, z be nose length, etc.
- The figure shows faces produced using 10 characteristics--head eccentricity, eye size, eye spacing, eye eccentricity, pupil size, eyebrow slant, nose size, mouth shape, mouth size, and mouth opening): Each assigned one of 10 possible values, generated using *Mathematica*
- *DISADVANTAGES*
  - Viewing large tables of data can be tedious
  - specific data values are not shown.
  - Furthermore, facial features vary in perceived importance

Metrics	Facial Features
Batting average	Height of face
Strike rate	Curve of smile
Number of fours per match	Width of eyes
Number of sixers per match	Height of eyes
Ratio of Innings to total matches played	Width of face



Stick Figures

- The stick figure visualization technique maps multidimensional data to five-piece stick figures, where each figure has four limbs and a body.
- Two dimensions are mapped to the display (x and y) axes and the remaining dimensions are mapped to the angle or length of the limbs.
- General techniques
    1. **Shape coding:** Use shape to represent certain information encoding
    2. **Color icons:** Use color icons to encode more information
    3. **Tile bars:** Use small icons to represent the relevant feature vectors in document retrieval



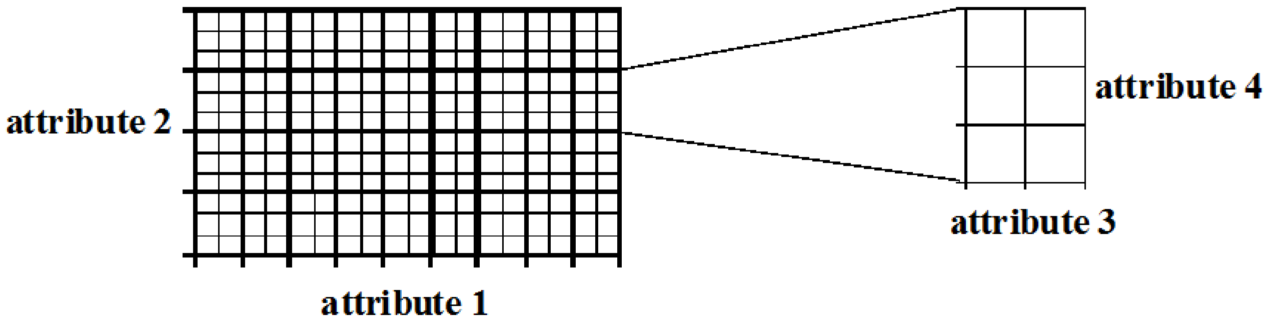
A census data figure showing age, income, gender, education, etc. A 5-piece stick figure (1 body and 4 limbs w. different angle/length)

Hierarchical visualization techniques

- Hierarchical visualization techniques partition all dimensions into subsets (i.e., subspaces).
- The subspaces are visualized in a hierarchical manner.
- Methods
  1. Dimensional Stacking
  2. Worlds-within-Worlds
  3. Tree-Map
  4. Cone Trees
  5. InfoCube

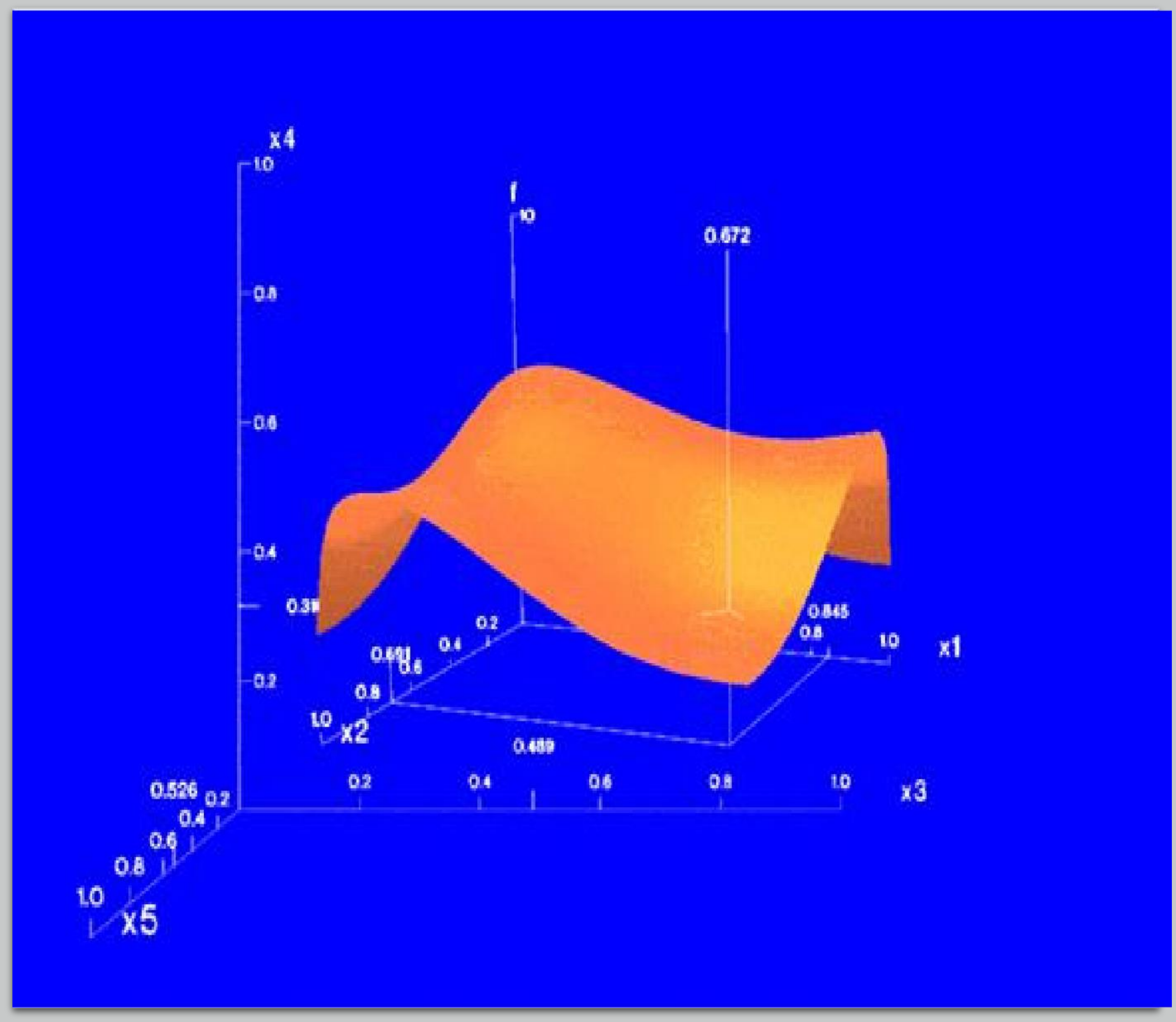
Dimensional Stacking

- Partitioning of the n-dimensional attribute space in 2-D subspaces, which are ‘stacked’ into each other
- Partitioning of the attribute value ranges into classes. The important attributes should be used on the outer levels.
- Adequate for data with ordinal attributes of low cardinality
- But, difficult to display more than nine dimensions
- Important to map dimensions appropriately



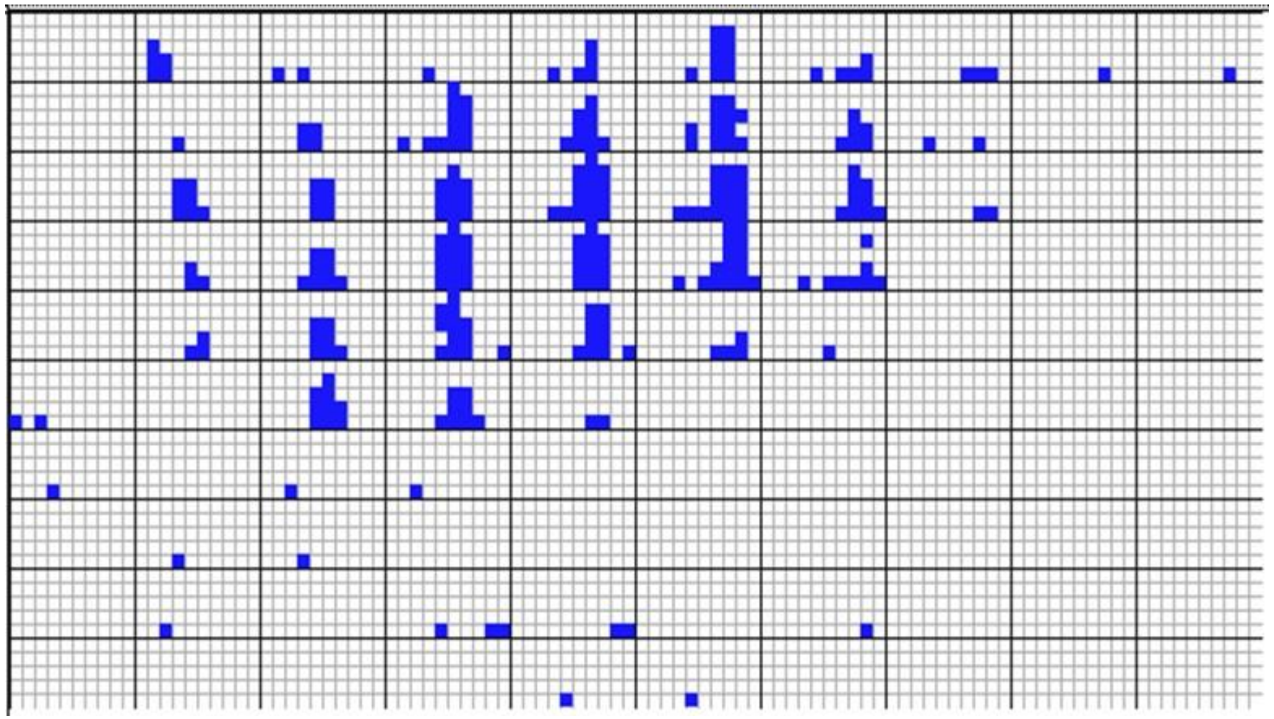
1. Worlds-within-Worlds

- Assign the function and two most important parameters to innermost world
- Fix all other parameters at constant values - draw other (1 or 2 or 3 dimensional worlds choosing these as the axes)
- Software that uses this paradigm
- **N–vision**: Dynamic interaction through data glove and stereo displays, including rotation, scaling (inner) and translation (inner/outer)
- **Auto Visual:** Static interaction by means of queries



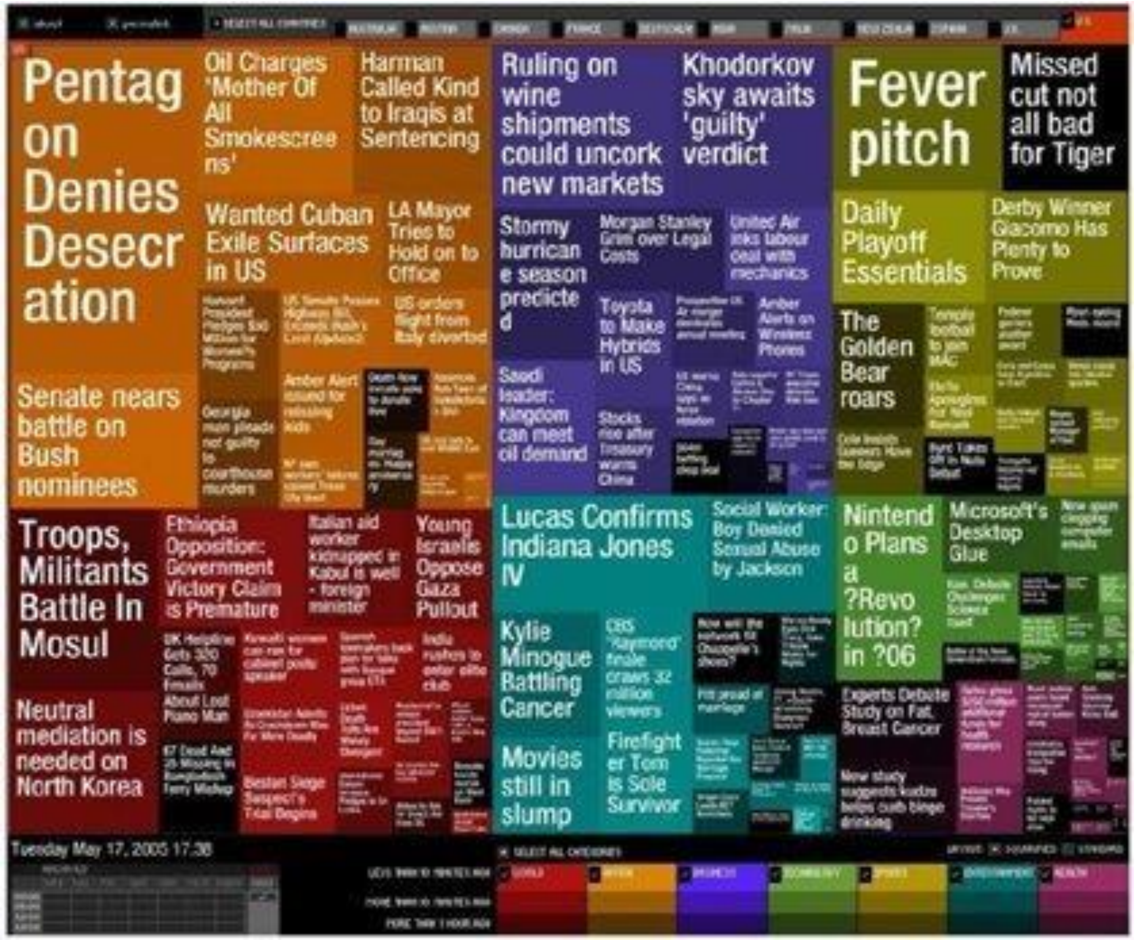
2. Dimensional Stacking

- Visualization of oil mining data with longitude and latitude mapped to the outer x-, y-axes and ore grade and depth mapped to the inner x-, y-axes

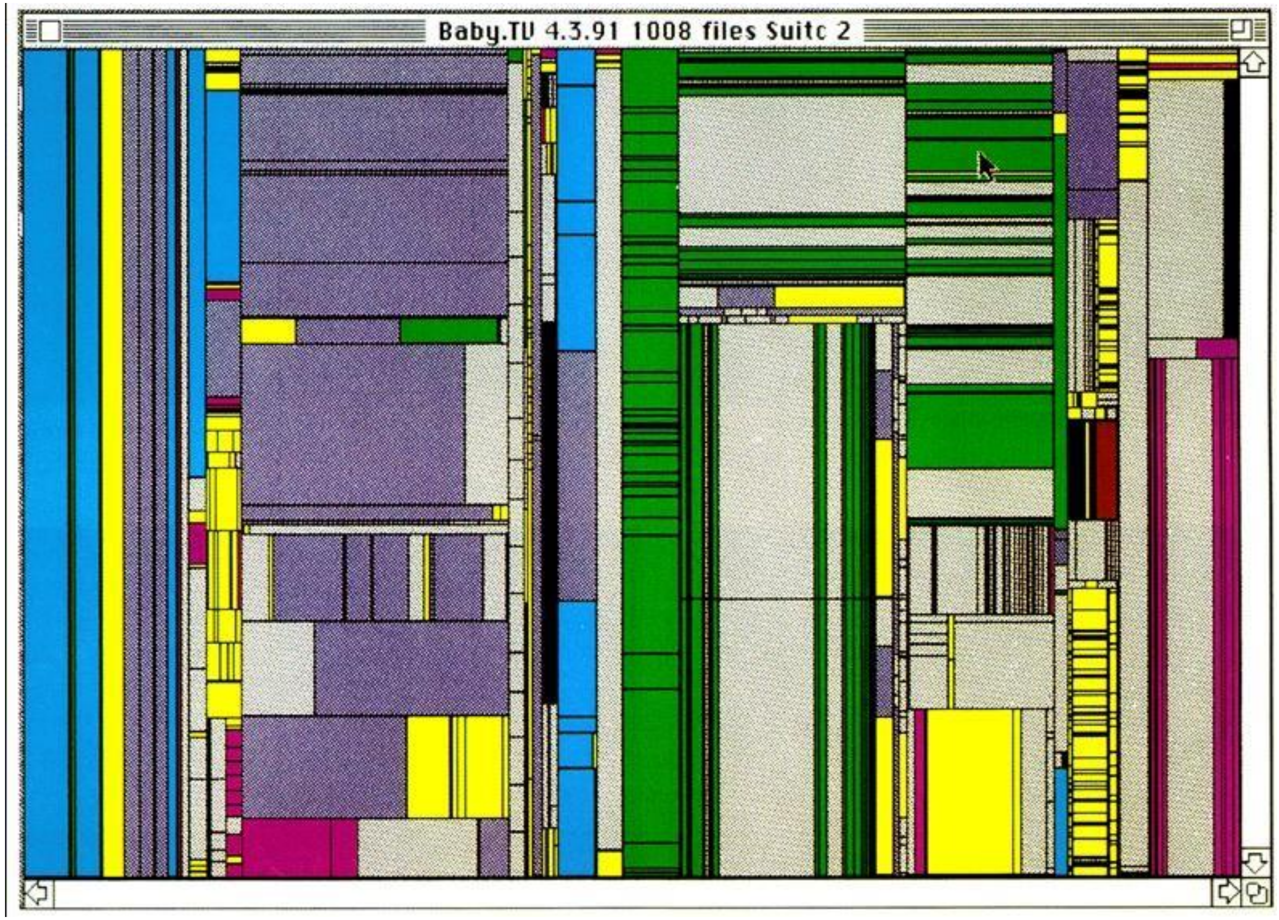


3. Tree-Map

- tree-maps display hierarchical data as a set of nested rectangles

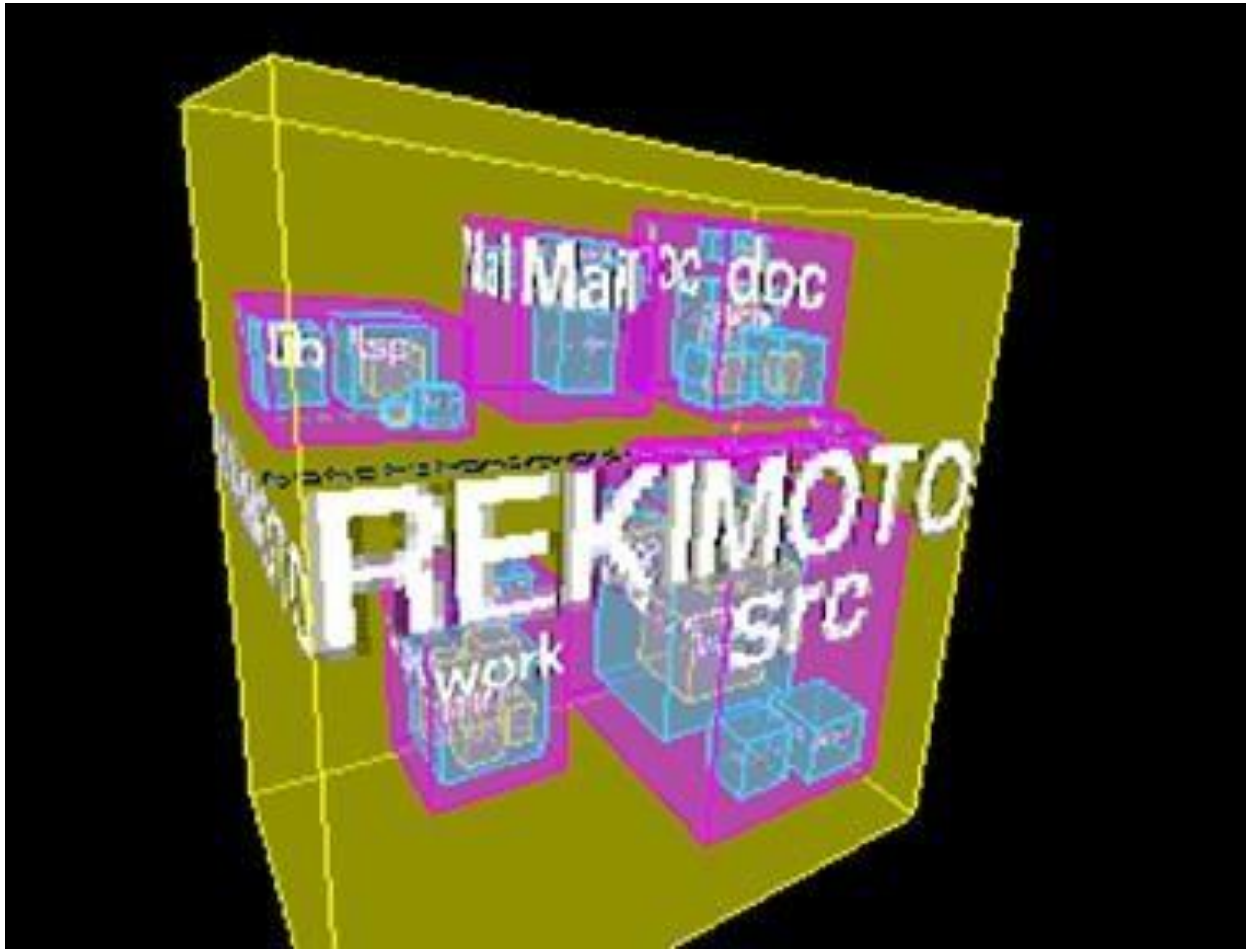


Tree-Map of a File System (Schneiderman)



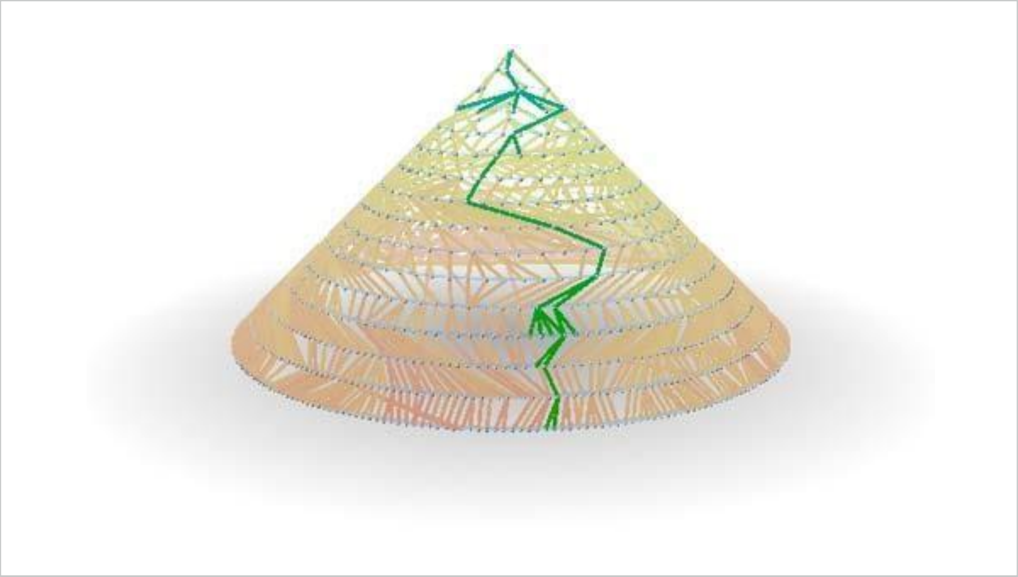
4. InfoCube

- A 3-D visualization technique where hierarchical information is displayed as nested semi-transparent cubes
- The **outermost cubes correspond to the top level data**, while the **subnodes or the lower level data are represented as smaller cubes inside the outermost cube**s, and so on



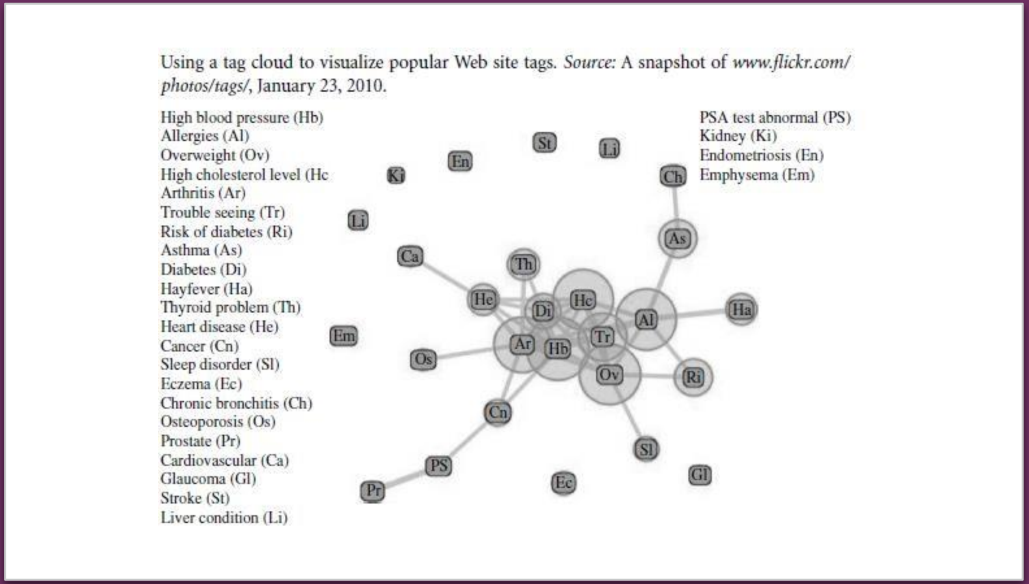
5. Three-D Cone Trees

- *3D cone tree* visualization technique works well for up to a thousand nodes or so
- First build a *2D circle tree* that arranges its nodes in concentric circles centered on the root node
- Cannot avoid overlaps when projected to 2D
- G. Robertson, J. Mackinlay, S. Card. “Cone Trees: Animated 3D Visualizations of Hierarchical Information”, *ACM SIGCHI'91*
- Graph from Nadeau Software Consulting website: Visualize a social network data set that models the way an infection spreads from one person to the next



Visualizing Complex Data and Relations

- Visualizing non-numerical data: text and social networks
- **Tag cloud:** visualizing user-generated tags
- The importance of tag is represented by font size/color
- Size of a tag is used to represent the number of times that the tag is applied to this item by different users. Also it can be used to visualize the tag statistics on multiple items, to represent the number of items that the tag has been applied to, that is, the popularity of the tag.
- Besides text data, there are also methods to visualize relationships, such as visualizing social networks.



Scoreboard Vs Dashboard

Scoreboard Vs Dashboard

- A dashboard or scorecard interface finally makes it easy for users to quickly find, analyze, and explore the information they need to perform their jobs.

SCOREBOARD

- A scorecard is a type of report that measures and compares your performance against your projections and goals.
- Scorecards are the performance management tool that compares strategic goals with results. This tool is typically a top-down approach that allows management to implement its strategy by aligning performance with goals.
- A scorecard is a type of report that measures and compares your performance against your projections and goals. It evaluates the success and failure of your efforts, based on key performance indicators (KPIs). These KPI's must be determined early which allows management to successfully evaluate progress compared to expectations.

DASHBOARD

- Dashboards are a collection of graphs, charts, gauges or other visual representations that are used to monitor the levels of the selected KPIs.
- Dashboards are made up of multiple report types, allowing users to easily compare and contrast different

Scoreboard Vs Dashboard

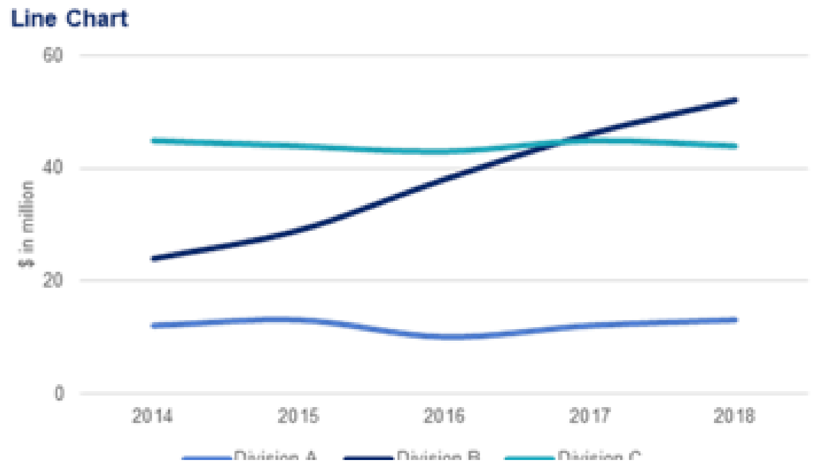
Dashboard	Scorecard
Tactical – focused on short-term decision making	**Strategic** – focused on long-term decision making
Provides a snapshot of business performance	Represents trends / changes in business activity over time
Operationally focused and supported by individual managers	Supported by clearly defined management strategy
Change in performance evaluated by primary decision-makers / stakeholders	Changes in performance measured against business goals
Display performance (metric)	Display progress (metrics + target)
Real time feed	Monthly snapshots
Visualize the performance to understand the current state	Align KPIs, Objectives and actions to see the connections between them

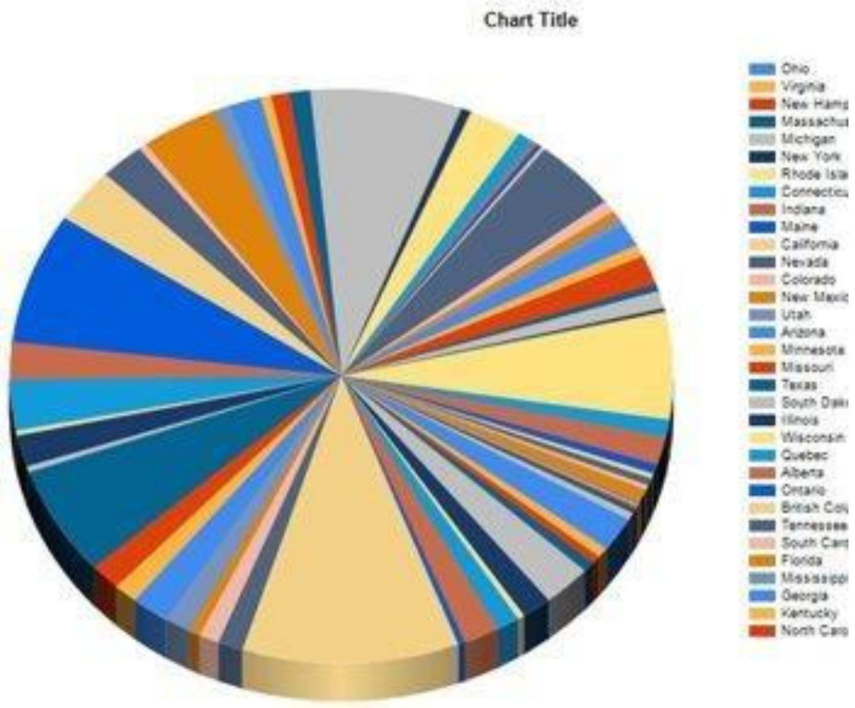
Dashboard vs. Scorecard: A Quick Look at the Key Differences

	Dashboard	Scorecard
Purpose	Performance at a glance	Managing performance
Timeliness of data	Current (often real time)	Periodic (often monthly)
Users	Lower level managers	Upper level managers
Linked to systems	Almost always	Sometimes

Charts and Purposes

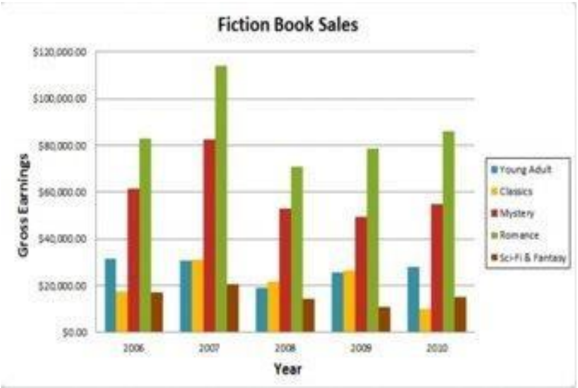
- Use bar charts to: • Compare data across categories. Bar charts are best suited for data that can be split into several groups. Bar charts are great when we want to track the development of one or two variables over time
- Use line charts to: • View trends in data, usually over time. Line charts are most often used to show a value evolving over time, but they aren't limited to that. Any dimension—like date types, time intervals, and other ordinal data—can be used as the horizontal axis. For example, stock price change over five years, website page views during a month, and revenue growth by quarte
- use pie charts: • Alongside other charts. A pie chart is most impactful when used to add information to other types of charts and graphs.
- Use scatter plots to: • Investigate the relationship between different variables. Scatter plots are used to show if one variable is a good predictor of another, or if they tend to change independently.
- Use bubble charts to: • Show the relationship between three or more measures. Bubbles can add more detail to the traditional two axis chart, highlighting the relationship between three or more variables, without overwhelming the viewer

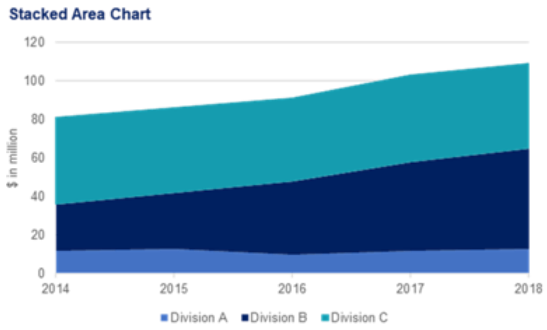




Charts and Purposes

- Use histograms to: • Understand the distribution of your data. The histogram is your best option for visualizing how data falls into categories
- Use heat maps to: • Compare categorical data using color. Heat maps are best for presenting data sets with lots of categories. This chart type can pack hundreds of comparisons into a small area, and still be easy to understand at a glance
- Use box-and-whisker plots to: • Show the distribution of a data set: Use box-and-whisker diagrams to understand your data at a glance. See how data is skewed towards one end and identify outliers in your data





To create a comparison of values:	to show the composition of something	to understand the distribution of your data : **To** understand outliers, the normal tendency, and the range of information in your values	**To** analyzing trends in your data set: information about how a data set performed during a specific time period	understand the relationship between value sets how one variable relates to one or numerous different variables.
Bar	Pie	Scatter Plot	Line	Scatter
Pie	Stacked Bar	Line		bubble
Line	Area	Bar		Line
Scatter Plot				

Thank You