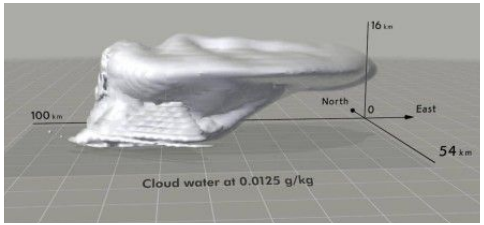
Theory Recap

IFSC 54503 Data Visualization

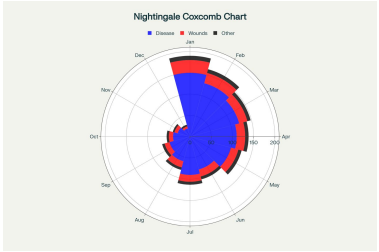
Submitted to - Prof. Mihail Tudoreanu

Introduction to Information Visualization

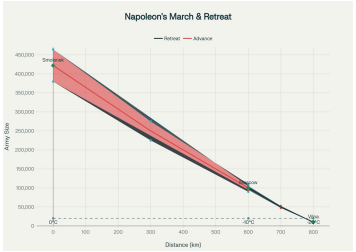
- **Definition**: Information Visualization is the graphical representation of abstract data to amplify cognition and help users think, discover, and make decisions.
- **Types**: Includes both Scientific Visualization (physical data with geometry) and Information Visualization (abstract data without physical shape, e.g., text, financial, internet datasets).
- **Visualization** is tailored for tasks, often relying on both graphics and text.
- **Famous examples**: Minard's chart on Napoleon's march and Nightingale's diagram on hospital mortality rates illustrate the power of visual storytelling.
- **Key principle**: No single visualization fits all tasks, data types, or user needs.



Scientific Visualization VS Information Visualization



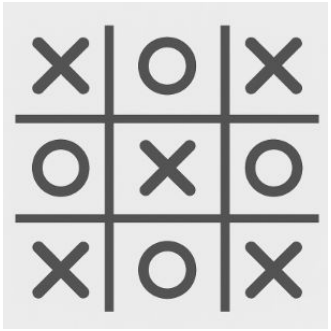
Florence Nightingale's Coxcomb Diagram — Causes of Hospital Mortality

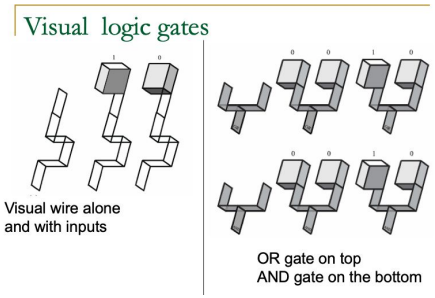


Minard's Chart — Napoleon's March to Moscow and Retreat

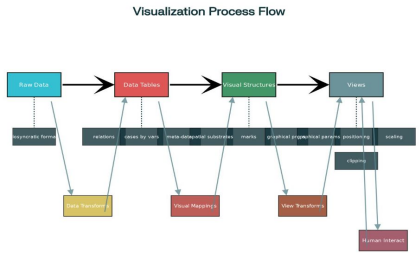
How Visualization works and best practices

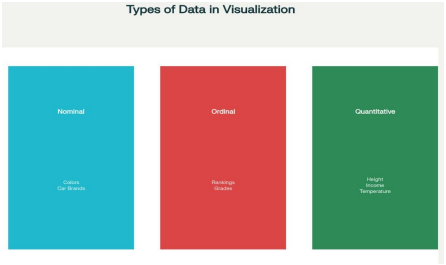
- **Visualization Process**: Turns raw data into visual forms through mapping and transformation—think of logic gates, multiplication, or abstract games like tic-tac-toe and chessboard problems.
- **Core benefits**: Improves memory capacity, supports discovery and learning, and reveals patterns using perceptual inference and cognitive maps.
- **Key data concepts**: Recognize types (nominal, ordinal, quantitative), size, complexity, and user context when designing visualizations.
- **Visualizations** must be customized for data and task. Interaction often increases effectiveness (e.g., allowing users to manipulate game states).
- **Real-world applications**: Used across scientific, financial, and internet information domains.





From Data to Insight



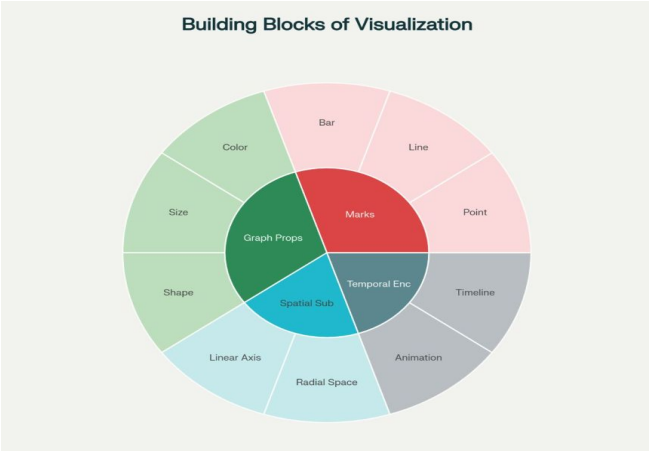


Core building blocks of Data Visualization

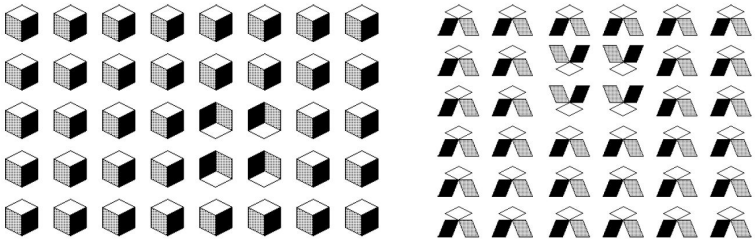
- **Spatial Substrate**: The foundational "space" or axes (linear, radial, 2D/3D) on which information is placed.
- **Marks**: Visual symbols representing data points (e.g., points, lines, bars, regions).
- **Graphical Properties**: Attributes like color, size, shape, orientation, and texture encode data values and categories.
- **Temporal Encoding**: Animation or sequence to display changes over time and enhance understanding.

Key Principles:

- **Gestalt** laws help us perceive patterns (proximity, similarity, closure).
- Pre-attentive processing allows near-instant distinction of important features (within 250 ms).
- Choosing the right structure and marks for your data & task is fundamental.



Same 3 types of diamonds used on both side;
Pre-attentive detection works only on the left



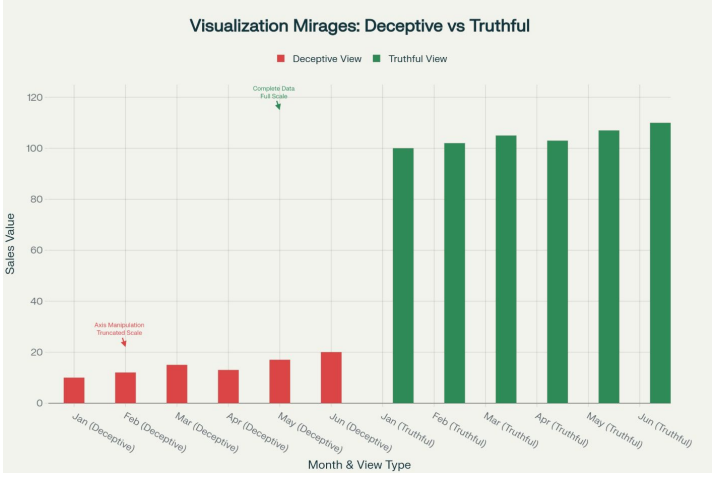
Detecting Visualization Mirages & Upholding Excellence

Visualization Mirages:

- Charts can mislead via axis tricks, omission, biased aggregation, or excessive "**chartjunk**.“

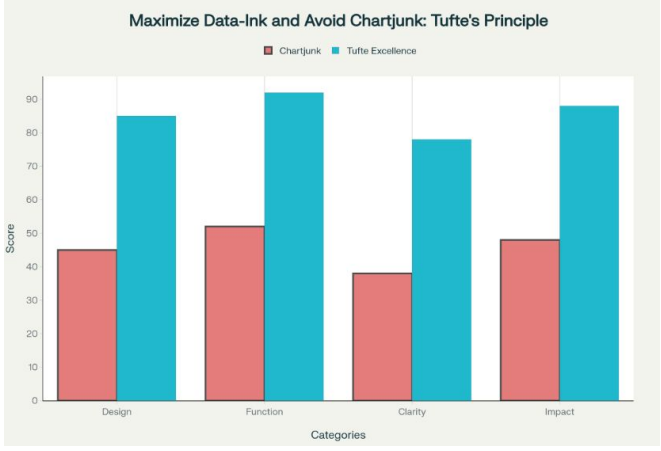
Deception Examples:

- Manipulating scales, hiding data, using similar mark colors for different categories, overplotting/hiding uncertainty.



Tufte’s Excellence Principles:

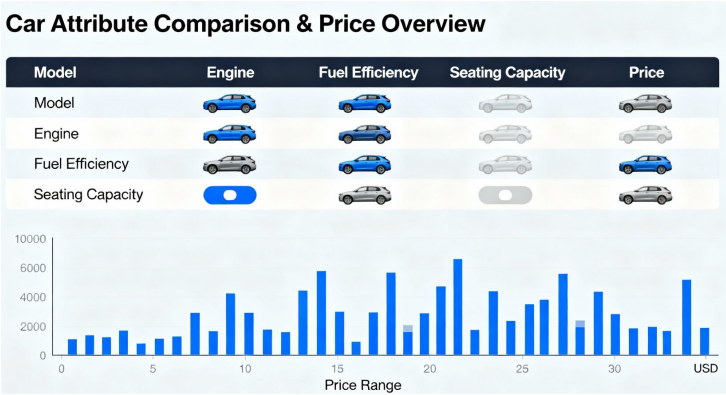
- Maximize **data-ink ratio**.
- Avoid “**chartjunk**” (non-informative visuals).
- Use **Multifunctional** graphical elements
- Use “**small multiples**” and repeatable, clear designs.
- Show mechanism, process, **dynamics and casuality**



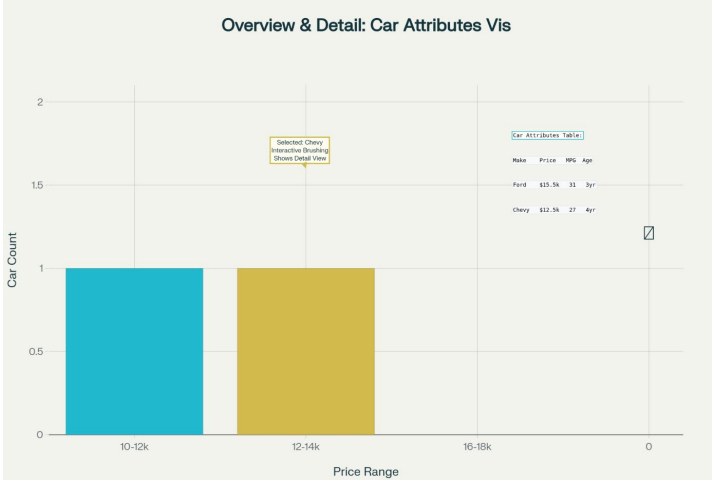
Visualization Overview vs Details (Cars Dataset Example)

Key idea: Effective visualizations give both a broad overview and detail-on-demand.

- Starts with a table of car attributes (Make, Price, MPG, Age) for comprehensive review.
- **Bargram** summarizes prices: wider bars mean more cars in that price range.
- **Detail and attention**: Car icons above bars can draw focus; appearance matters even if not strictly quantifiable.
- Selective **brushing** or highlighting enables interaction, helping users connect images and data for further exploration.



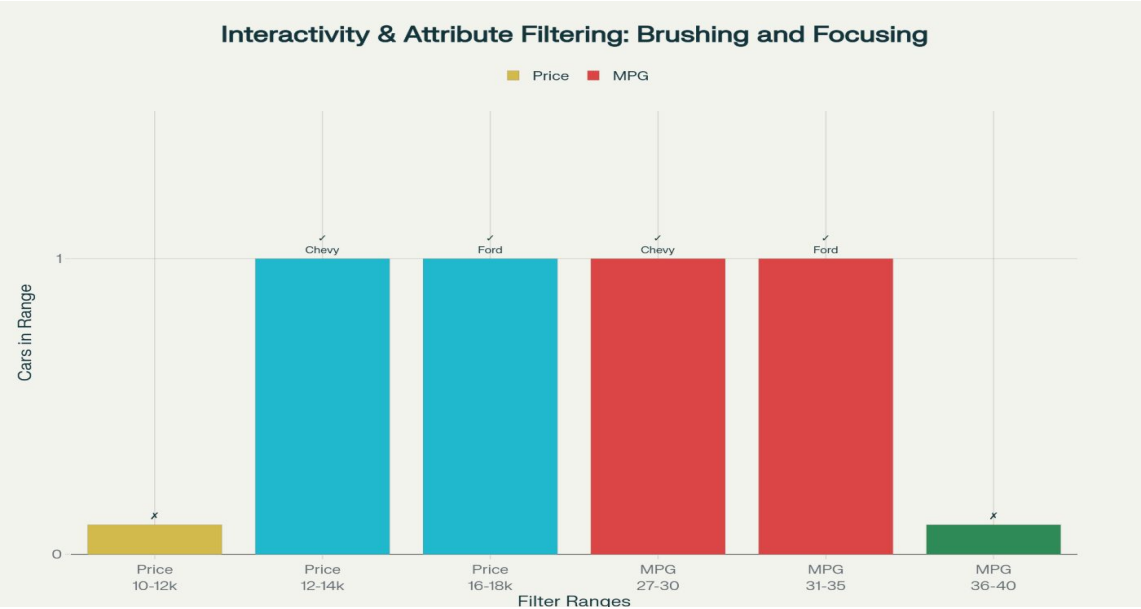
Two bargram charts for Price and MPG with color highlights showing brushing and multivariate filtering interaction, cars that meet both criteria are highlights



Interactive Filtering and Multivariate Selection

Key idea: Filtering makes sense of complex, multi-attribute datasets and refines choices.

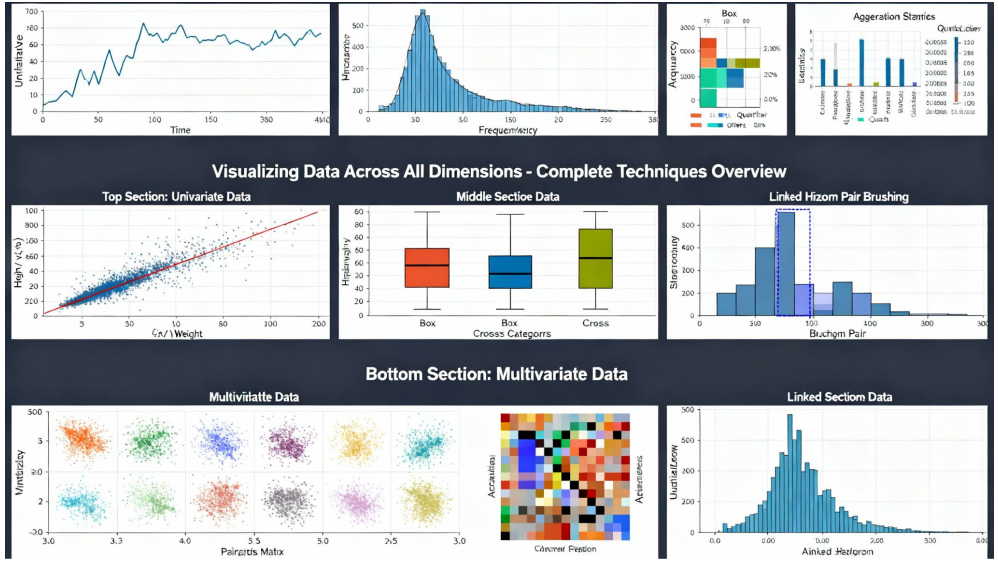
- **Bargrams** represent Price and MPG; colored highlights show active filters.
- Interactivity (e.g., “**brushing**” or “**focusing**”): Selecting a price range and then MPG instantly shows which cars match both, while icons above bars help visualize which cars are included or excluded.
- Space limitations? Dynamic **layouts**, **scrolling**, or **shrinking** bars preserve overall structure and clarity.



Visualizing Different Dimensions of Data

Key Ideas:

- Data can be **uni-dimensional** (one attribute), **bi-dimensional** (two), or **multi-dimensional** (many attributes).
- Classic techniques: Line plots for single variables, **scatter plots** for two, scatterplot matrices for higher dimensions.
- **Aggregation** and **linking** (histograms, box plots, brushing) reveal insight for larger datasets.

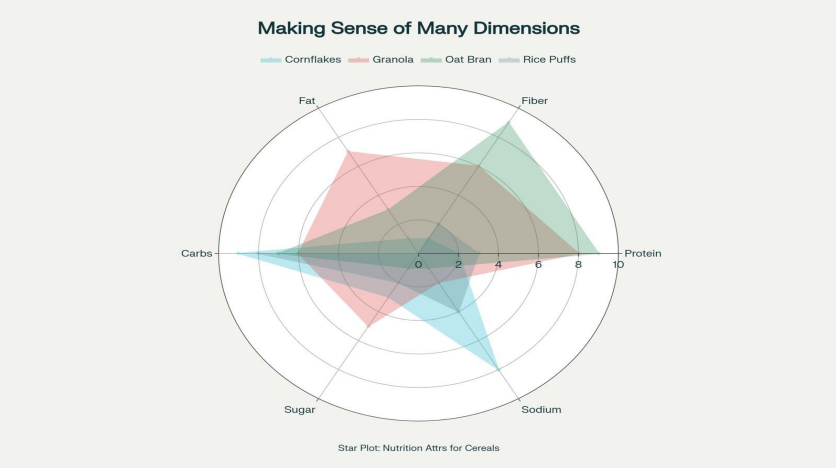


Making Sense of Many Dimensions—Multivariate Visualization

Techniques

Key Ideas

- Explore advanced tools: **Star plots** summarize objects (e.g., cereal nutrition), parallel coordinates uncover clusters and trends, **Chernoff faces** turn attributes into expressive icons.
- **Attribute explorer** and **brushing** help users filter and zoom in on relevant variables.
- **Semantic zooming** and l**inked displays** reveal hidden detail—crucial for tackling complex, high-dimensional data.

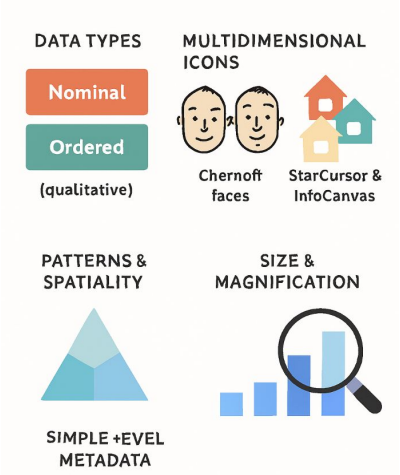


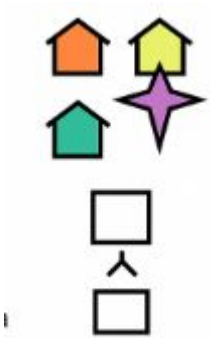
Encoding Non-Quantitative Data

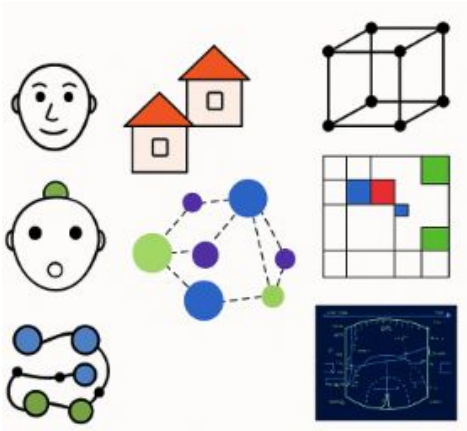
Data types: nominal, ordered (qualitative vs quantitative display)
Challenges: quantitative techniques can mislead when applied to qualitative data
Multidimensional icons:

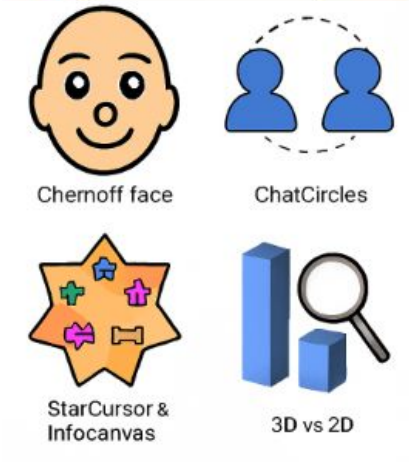
- **Chernoff faces** – customizable facial features to encode variables
- **StarCursor & Infocanvas** – domain-specific icon systems (e.g., real estate)
- **ChatCircles** – spatial chat visualization, showing interaction patterns

Patterns & spatiality: simple data forms higher-level metadata; 3D vs 2D representations
Size & magnification: length, width, adjacency relations, cockpit views, selective magnification









Color and Sound in Visualization

Color encoding:

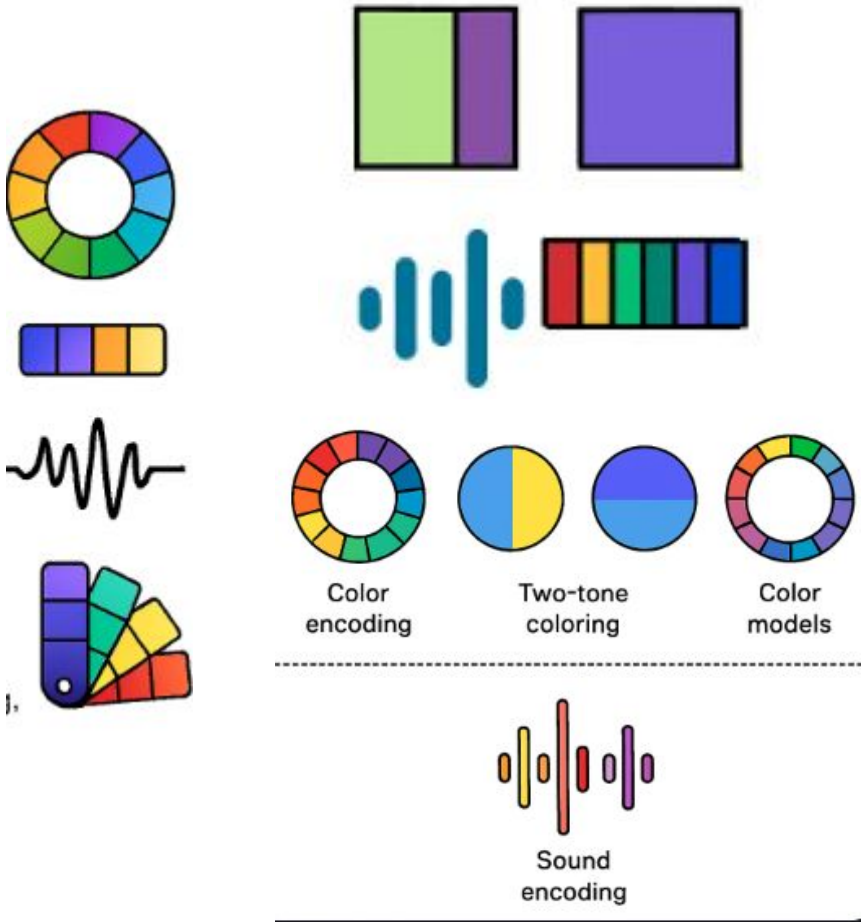
- Distinguish 5–10 categories (nominal)
- Ordered data harder to represent
- **Color models**: HSL (Hue, Saturation, Lightness)
- **Schemes**: grayscale, spectral, single hue, double-ended scales
- **Effective use**: limit colors, avoid vibration, consider brightness & color-blindness

Two-tone coloring: compact visualization for 1D data (discrete vs continuous)

Sound encoding: categorical data via auditory cues (e.g., brain tumor exploration, design listening)

Design tools: palettes of techniques (visibility, brushing, retinal variables)

Key takeaway: Best technique depends on task; empirical evaluation needed



Graph and Node-Link Visualization

Graphs: sets of nodes connected by links (directed/undirected)

- Represent connections: phone calls, social networks, criminal activities, transport
- Structures: adjacency lists & adjacency matrices

Node-link diagrams:

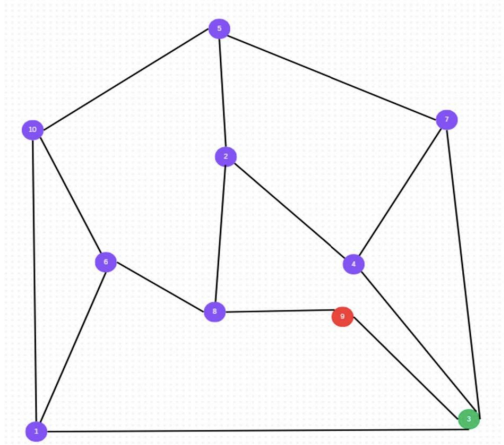
- Transform textual data into visual connections
- Reveal disconnected subsets, clusters, and social choices (e.g., Moreno's sociograms, Freeman's employee networks)

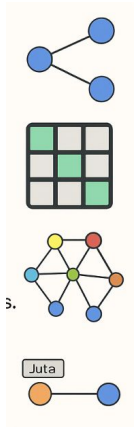
Graph drawing basics:

- Nodes: position, size, shape, color, labels
- Links: path (straight/curved), thickness, color, labels

Goals: reduce intersections, occlusion, link length variation, and bends

1	3,6,10
2	4,5,8
3	1,4,7,9
4	2,3,7
5	2,7,10
6	1,8,10
7	3,4,5
8	2,6,9
9	3,8
10	1,5,6





Layout Techniques and Tools

Layouts:

- **Case 1**: known positions (e.g., subway maps, telephone networks)
- **Case 2**: unconstrained positions → group nodes, circular layouts, step-link focus (NetMap)

Advanced techniques:

- **Green Arrow**: extended labels
- **Edge bundling**: cost/image-based aggregation to reduce clutter
- **Stretch-out graphs**: historical/regulatory networks
- **Interactive LOD rendering**: zoom from overview to detail with density-based aggregation

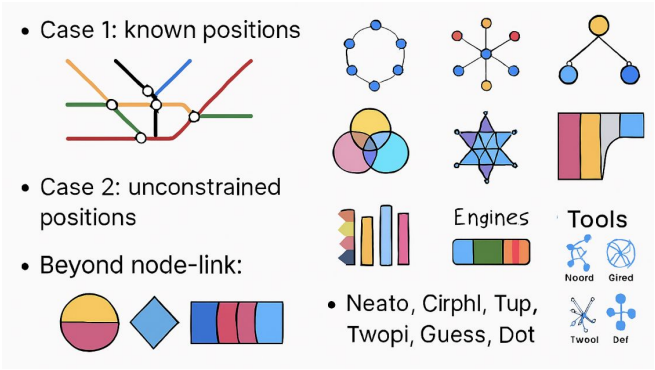
Beyond node-link:

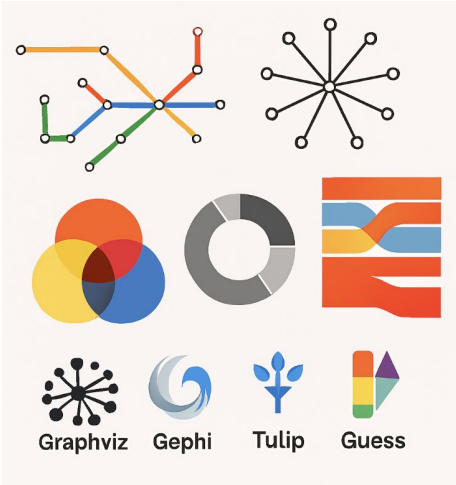
- Venn diagrams, InfoCrystal, Upset plots, Parallel Sets, Sankey/Alluvial diagrams

Tools: Graphviz (.dot language), Gephi, Tulip, Guess

- Layout engines: **neato** (spring), **circo** (circular), **twopi** (radial), **dot** (directed)

.





Tree Visualization Techniques

Definition of a Tree

- A connected, acyclic graph
- Rooted hierarchy: root → children → grandchildren

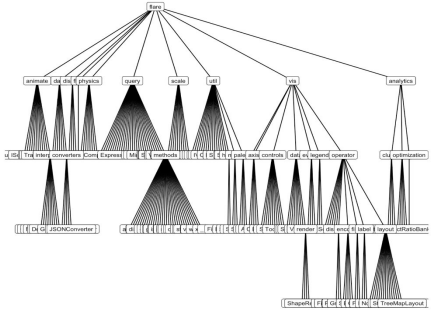
Node-Link Diagrams

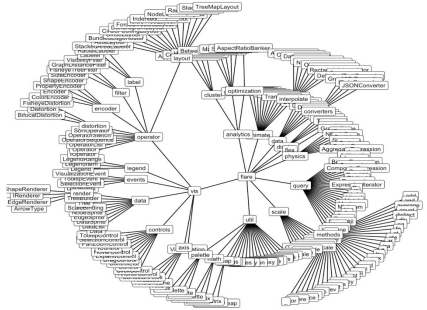
- Top-down, circular, and 3D layouts (e.g., Cone Trees)
- Hyperbolic browsers (2D/3D): focus+context, scalable to thousands of nodes
- SpaceTree: interactive node-link layout from HCIL

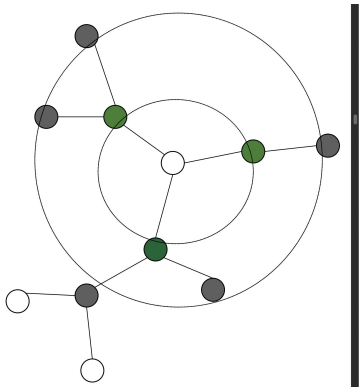
Challenges

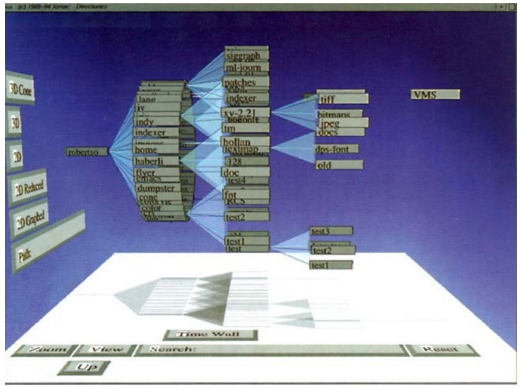
- Exponential growth → clutter
- Occlusion in 3D
- Hard to show media/text in small nodes

.









Space-Filling Tree Visualizations

TreeMaps (Shneiderman & Johnson)

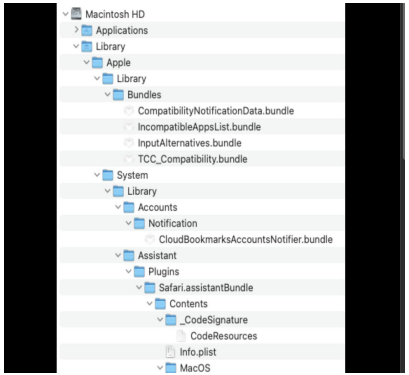
- Use area to represent data dimensions
- Alternate horizontal/vertical slicing
- Color and labels for additional properties

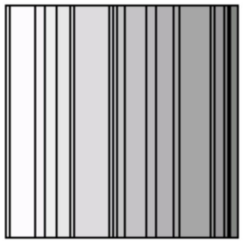
Variants

- Squarified TreeMaps: better aspect ratios
- Cushion TreeMaps: shading for hierarchy depth
- Voronoi TreeMaps: polygonal layout
- Radial Space-Filling & SunBurst: circular containment

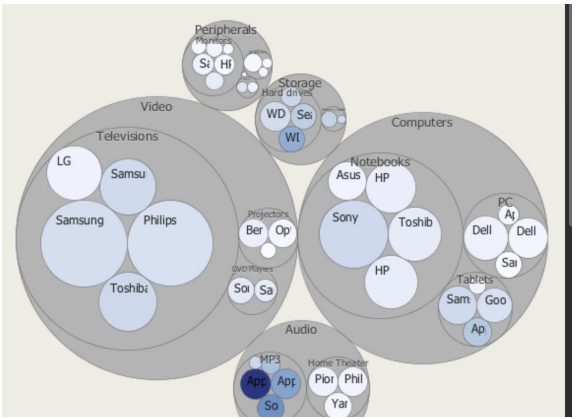
Applications

- File systems (SequoiaView)
- Stock market maps (SmartMoney)
- Entity resolution tools (D-Dupe)





Slice-and-dice: ordered white to black; rectangles are very thin



Re-arrangement & Scrolling Challenges

Re-arrangement Techniques

- Change visual order without altering data
- Examples: scatterplots of couple data, TableLens (SAP/Inxight)

Large Data Set Issues

- Too many items to fit on one screen
- Multidimensional ≠ large (size = #objects, dimensionality = #properties)

Scrolling Limitations

- Poor context, slow interaction, hard to detect patterns
- Efficiency:
  - Simple search → diameter NN
  - Binary search → log(N)\log(N)
  - 2D rearrangement → N\sqrt{N}

Presentation Problem

- Small screens can't show all relevant info
- Context loss during scrolling (e.g., maps, microfilm)

Scrolling

Age of Empire Game - too many items to fit in one screen

Overview + Detail & Mural Techniques

Overview + Detail Models

- Detached views → switching back and forth
- Zoom factor, multiple detail views, transparency/superposition
- DragMag: dynamic magnifying glass
- Tiled multilevel views: global, intermediate, detailed

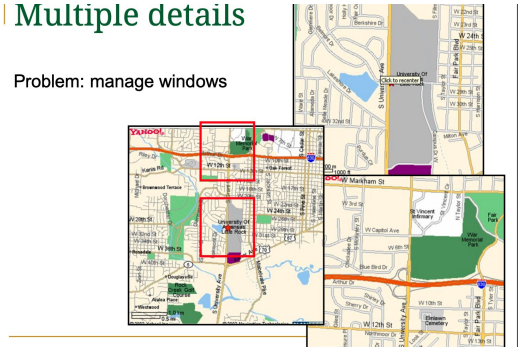
Information Mural

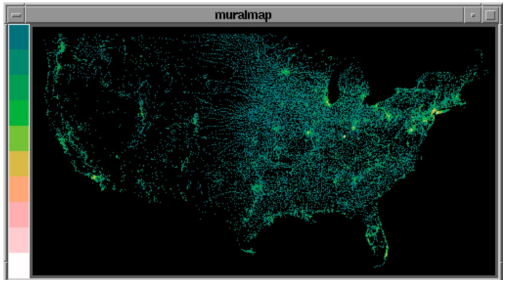
- Don’t eliminate data — compress it visually
- Use antialiasing to smooth pixel bins
- Think of data points as ink, pixels as bins
- Examples: muralized parallel coordinates, sunspot activity, execution traces

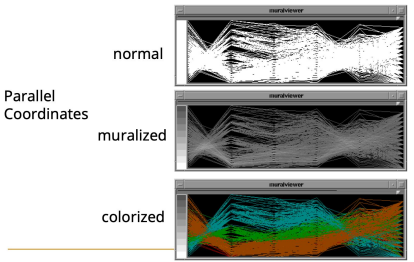
Task-Driven Design

- CAD → detail-heavy
- Exploration → fast navigation
- Monitoring → zoom on multiple areas
- Navigation → current position + destination context

.





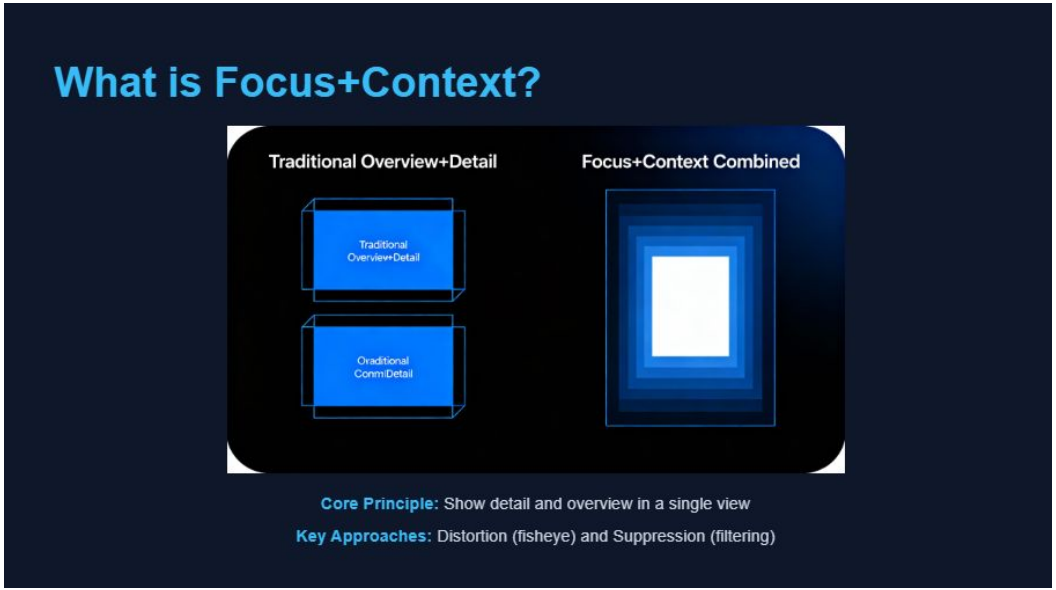


What is Focus+Context Visualization?

Key Points

- Helps users see both detail and overview in large data sets
- Mixes *focus* (detailed area) with *context* (general overview)
- Three principles:
  - Both focus and context needed for the task
  - Info in overview ≠ info in context
  - One combined display

.

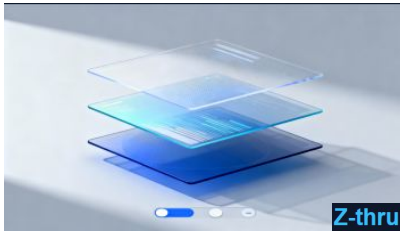


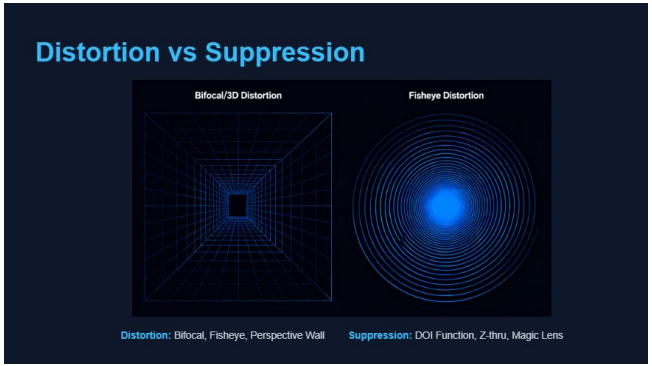
Techniques: Distortion and Suppressions

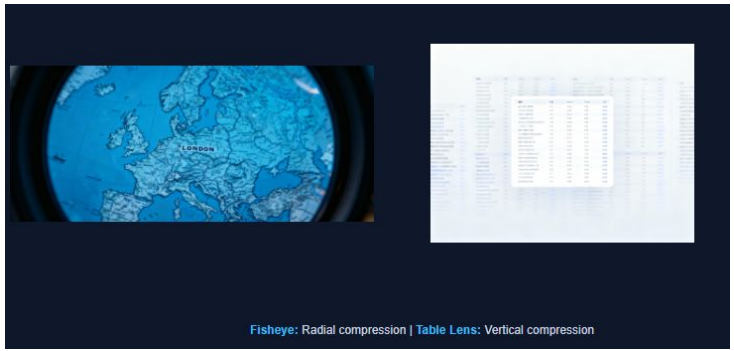
Key Points

- Distortion:
  - Fisheye views, bifocal view, cone trees, Table Lens
  - Manipulate X & Y axes or use 3D to magnify important regions, compress others
- Suppression:
  - Filter points far from focus (generalized fisheye, Z-thru)
  - Semantic zooming, micro-macro readings

.



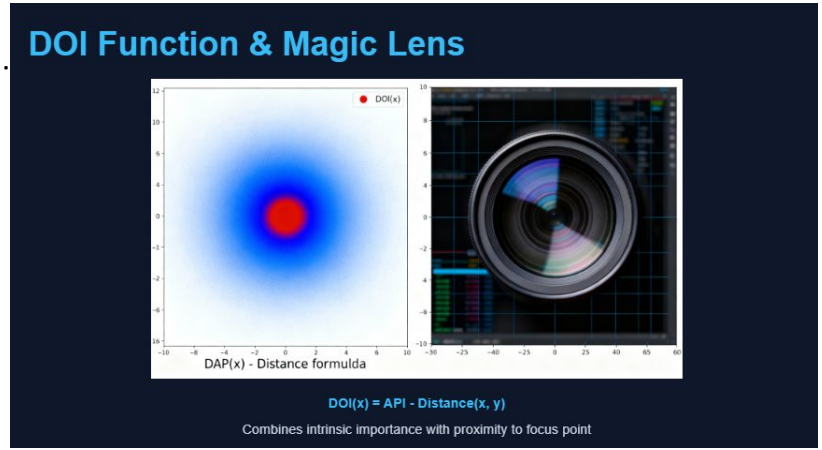




Degree of Interest (DOI) and Magic Lens

Key Points

- Degree Of Interest (DOI):
  - DOI(x) = API - Distance(x, y)
  - Filters or distorts data based on relevance to the focus
- Magic Lens, Implicit Context, Halos:
  - Overlay tools (lens, halo) add/removes info layers interactively



References

Lecture Slides from course IFSC 54503

Pre - recorded videos from course IFSC 54503

Images from Google and Lecture slides from course IFSC 54503

Googled some concepts for more details

Canva tool to join smaller videos canva.com