Multivariate Data Visualization

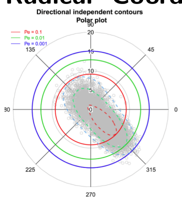
Geometric Projection

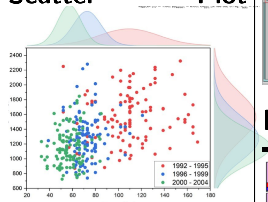
1. Data: 2 quantitative value attribute
2. Mark: point + (regression) line/area
3. Channels: position, color, size(shape)
4. Task: identify trend, outlier, distribution, correlation, cluster(enclosure)

isoline map for visual clutter, lose individual information. Variation: Scatter Plot Matrix(SPLOM). Diagonal for distribution of each attribute. Scaling and saliency problem.

1. Transform layout of scatter plot
2. usually for navigation or time
3. Variant: Star coor for higher dimension. spatial distortion and occlusion problem

Radical Coordinate **Scatter Plot**





Parallel Coordinate

line mark, position/color/size channel. To summarize, explore, cluster, correlate.

Reduce visual clutter: Scale /reorder axis /highlight lines. add baseline, brush, change density, color hue+ bundling=combo. Hierarchical, add scattered points, add histogram.

Circular Parallel coordinate

Transform layout of parallel coordinate with higher dimension. For example, radar graph.

Layout Density

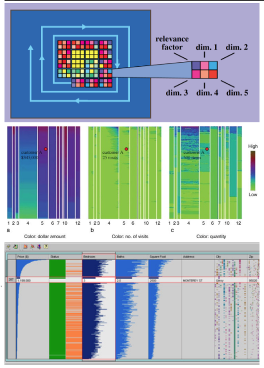
Space-filling layout: Heatmap

area mark to achieve high information density(data-ink ratio), position for key attributes(categorical), color for quantitative attribute. To find cluster, outlier.

Space-filling layout: Table lens

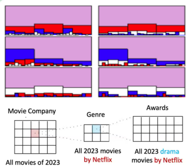
area mark in table cell, multi-key attributes separated by color and encoded by length(stacked bar chart)

Dense layout



1. M: point/line as pixel.n colored pixel for n key attributes of 1 data item.
2. C: position/color (**NO SIZE/SHAPE**)
3. E.g: Pixel-Oriented Layout, Pixel Bar Chart, Dense Table Lens

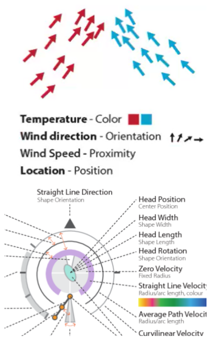
Hierarchical Display



subdivides the attributes of different importance(hierarchical data) into hierarchical fashion.

1. Hierarchical Axis
2. Dimensional Stacking

Iconography



glyph: a graphic symbol that provides the appearance or form for an item.

1. Star Glyph
2. Face Glyph
3. Flower Glyph

Dimensionality Reduction(DR)

Linear: Principle components analysis(PCA)

PCA: project alone axis with most variance.

Nonlinear DR

Mapping synthesized dims to original dims is difficult. But DR result still keep cluster and distance information.

1. t-SNE(local features): KL divergence for loss training, need tuning perplexity
2. UMAP(local and global): entropy as loss and easier for tuning

Text Visualization

get the gist of doc, cluster/compare/correlate.

Text Mining

1. Info Extraction: to find structured information or knowledge(named entity/relation tree)
2. Text Summarization: extract/abstract
3. Opinion Mining & Sentiment Analysis
4. Text Clustering: find similarity (vectors) and do Topic modeling(LSA/LDA/Topic embedding)
5. Text Classification

Word-level Design

Lexical Analysis: token distribution in doc

Word Cloud⇒ ManiWordle(reordered for saliency)⇒ Typography(bold original text, redundant), scatter plot

Idiom	Mark & Channel	Meaning	Pros	Cons
Cloud	text(M), size/ color/ position/ order(C)	frequency, semantics, temporal/ spatial pattern	show gist, initial query info, easy to compute	sub-optimal encoding; inaccurate size; unstable/ meaningless layout;
Node-link	text/line(M), text color/ size/ opacity, line size/ opacity (C)	occurrence by node size, pattern by edge, struct by concordance		readability & scalability
Pixel	point(M), color/ size/ opacity(C)	occurrence by density of pixel, structure by color patterns	info density	resolution & vis saliency

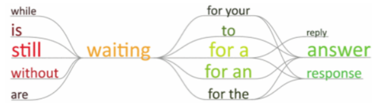
Topic and Trend Analysis

Clustered(by color), Juxtapose, Morphable (plant topic in leaf shape),Parallel,Matrix word clouds, SparkClouds(for trend)

Sentence-Level Design

Phrase syntactic: node-link graph (Word/line, Size/Color/line size for co-occurrence)

Sentence syntactic: Aligned Node-link tree



Document-Level Design

sentence func(syntactic) analysis, topic analysis, sentiment analysis, coherence in document(discourse analysis).

Syntactics: Fingerprinting

1. Area: cell for each sentence(M), color for sentence value and position for order(C).
2. Pixel: pixel for sentence(M)

Typography for Topic or Sentiment

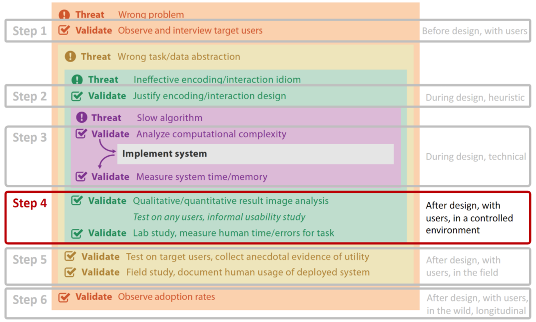
Colored text for each topic/sentiment. Problem: lack overview→ combine with tree map.

Documents

1. line graph to connect coherent sequence
2. Dendrogram(tree): show hierarchy
3. fingerprinting

Corpus: topic, sentiment, discourse, trend, relation

Evaluation



Graph Visualization

Advantages: solid theory as fundamental model, explanatory in vis(connectedness, proximity, color, size, shape)

Tasks: seek info, analytics, communication

Challenges: complex algo, display clutter, complex data struct, cognitive issue, spatial layout.

How-information-seeking Mantra:overview first, zoom and filter, then details-on-demand.

Graph Encoding Idioms

Node-link Layouts

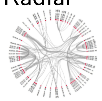
Node points and Lines(M), encoded by color and size(C). Challenges: scalability, cluttering.

For general graph: add clustering by enclosure/color, minimize edge crossing, uniform distribution of nodes and length, forth-directed layout, arc diagram(improve stability), radial graph, chord diagram.

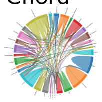
For bipartite graph: Parallel Layout, Sankey Diagram(Parallel Set/Generic)

For tree: rooted trees, indented tree layout, cluster dendrograms, radial tree(space efficiency)

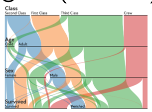
Radial



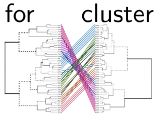
Chord



Sankey Diagram(Parallel)



Dendrogram for cluster



Other layouts for trees: enclosure, division

Enclosure Layout

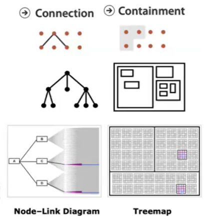
Children nodes inside parent node, use color and size to encode area.

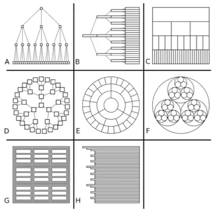
- Tree-map: rectangle area correspond to value. (Slice and Dice Layout, Squarified Layout to avoid extreme aspect ratio)
- Voronoi Tree-map: accurate leave-node
- Radial Tree-map: perception, wastespace

Division Layout

Children stacked on parents. Variation: sunburst diagram. low data-ink ratio and little space for children nodes.

Matrix Layout



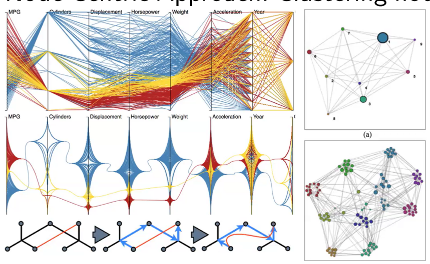


3D Layout more information due to extra dimension, while might cause occlusion problem.

Design Techniques

Visual Clutter Reduction

- Edge-Centric Approach: confluent drawing, Hierarchical trees, edge bundles.
- Node Centric Approach: Clustering nodes



- Appearance-Centric Approach: Sampling, filtering, clustering, changing point size and opacity, animation.
- Space-centric approaches: point/line displacement, Topological distortion, space filling, pixel-plotting, dimension reorder.

Interaction & Navigation

Common interactions:

- Select with focus+context
- Zoom and pan(fisheye view, magic lens)
- Animation

IMPORTANT things before midterm

nominal, ordinal, interval and ratio data

1. **Proximity**: spatial concentration
2. **Similarity**: same color/shape/size
3. **Connectedness**
4. **Enclosure**: objects enclosed by lines or placed in a common container appear as a collective. **strongest pattern!**
5. **Continuity**: smooth and continuous.
6. **Common Fate**: same moving direction
7. **Symmetry**: compare time series data.
8. **Closure**
9. **Figure and Ground**

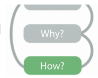
- Graphical Integrity: Telling the truth
  - lie factor

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

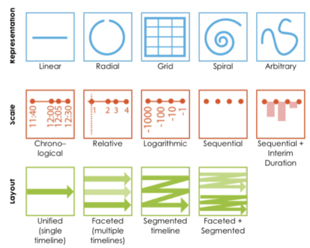
$$\text{size of effect} = \frac{|\text{second value - first value}|}{\text{first value}}$$

  - express data in context
  - correct(consistent, standardized, meaningful) scale
- Data-Ink Ratio:
- Chart Junk Avoidance
- No unjustified 3D
- No unjustified 2D
- Eyes beat memory
- Resolution ever immersion
- Overview first, zoom and filter, detail on demand
- Responsiveness is required
- Get it right in black and white
- Function first, form next

Recap: Visualization of Time Series Data

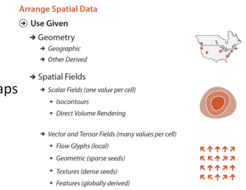


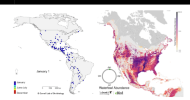
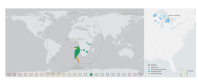
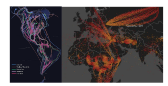
- Line graph
  - Others, e.g., connected scatterplot
- Area graph
  - Stacked graph
    - Horizon graph
  - Streamgraph
- Heatmap
- Bar graph
  - (Stacked) bar chart
  - Gantt chart
- Metaphorical idioms
  - Radial timeline
  - Calendar
  - Theme river



Recap

- Cartography: Map Basics
  - Projection: Azimuthal, Equal-Area, Conformal
  - Scale
- Mapping: Visualizing Geospatial Data
  - Choropleth maps
  - Symbol maps
    - Dot maps
    - Bubble maps
    - Spatial heatmaps and isoline maps
    - Others
  - Cartograms
  - Flow maps
  - Others



	Techniques	Visualizations	Features
Spatial	Points Lines (Curves) Lines + Areas		1. Simple 2. Visual clutter
Temporal	Animation Color Shape 3D		1. Clear overview 2. Instant display v durative display
Other (Speed, Density, Height)	Color Height + Illumination		Clear visual pattern