Guest Lectures:
Bettina Speckmann, Cartography & Flow;
Yang Wang, Architectures for Scale
Paper: TopoFisheye
Example Present: Biomechanical Motion
Ch 13/14/15: Reduce, Embed, Case Studies

Tamara Munzner
Department of Computer Science
University of British Columbia

CPSC 547, Information Visualization
Week 8: 29 Oct 2019

http://www.cs.ubc.ca/~tmm/courses/547-19
News

• presentation days assigned
  – both times and papers; still need topics from two of you!

• today
  – guest lecture: Bettina Speckmann
    • Necklaces and Flows: Algorithms for Automated Cartography;
  – guest lecture: Yang Wang
    • Architecting Visualizations at Scale
  – break
  – example presentation
  – topo fisheye views paper
  – chapters: reduce, embed, case studies
Example Presentation: Biomechanical Motion
Presentation expectations

• 20 minute time slots for presentations
  – aim for 18 min presenting and 2 min discussion

• slides required
  – if you’re using my laptop, send to me by 12pm
  – if you’re using your own, send to me by 6pm (right after class)

• three goals: up to you whether sequential or interleaved
  – explain core technical content to audience
  – analyze with doing what/why/how framework
  • do scale analysis of data for this system in specific, not for technique in general
  – critique strengths/weaknesses of technical paper

• marking criteria
  – Summary 40%, Analysis 15%, Critique 15%
  – Presentation Style 15%, Materials Preparation 15%
Analysis & critique

• paper type dependent
  – required for design studies and technique papers
  – some possible for algorithm papers
    • but more emphasis on presenting algorithm clearly
  – minimal for evaluation papers
    • but can discuss study design and statistical analysis methods

• please distinguish: their analysis (future work, limitations) from your own thoughts/critiques
  – good to present both
Beyond paper itself

• check for author paper page
  – may have video
  – may have talk slides you could borrow as a base
    • do acknowledge if so!
  – may have demo or supplemental material
  – include paper page URL in slides if it exists

• if using video, consider when it’s most useful to show
  – at very start for overview of everything
  – after you’ve explained some of background
  – after you’ve walked us through most of interface, to show interaction in specific
Slides

• do include both text and images

• text
  – font must be readable from back of room
    • 24 point as absolute minimum
    • use different type sizes to help guide eye, with larger title font
    • avoid micro text with macro whitespace
  – bullet style not sentences
    • sub-bullets for secondary points
    • Compare what it feels like to read an entire long sentence on a slide; while complex structure is a good thing to have for flow in writing, it’s more difficult to parse in the context of a slide where the speaker is speaking over it.

• legibility
  – remember luminance contrast requirements with colors!
Slide images

• figures from paper
  – good idea to use figures from paper, especially screenshots
    • judgement call about some/many/all

• new images
  – you might make new diagrams
  – you might grab other images, especially for background or if comparing to prev work
  – avoid random clip art

• images alone often hard to follow
  – images do not speak for themselves, you must walk us through them
    • text bullets to walk us through your highest-level points
      – hard to follow if they’re only made verbally
    • judgement call on text/image ratio, avoid extremes
Style

• face audience, not screen
  – pro tip: your screen left/right matches audience left/right in this configuration

• project voice so we can hear you
  – avoid muttered comments to self, volume drop-off at end of slide
  – avoid robot monotone, variable emphasis helps keep us engaged

• avoid reading exactly what the slide says
  – judgement call: how much detail to have in presenter notes

• use laser pointer judiciously
  – avoid constant distracting jiggle

• practice, practice, practice
  – for flow of words and for timing

• question handling: difficult to practice beforehand…
Technical talks advice

- **How To Give An Academic Talk**
  - Paul N. Edwards

- **How To Give a Great Research Talk**
  - Simon L Peyton Jones, John Hughes, and John Launchbury

- **How To Present A Paper**
  - Leslie Lamport

- **Things I Hope Not To See or Hear at SIGGRAPH**
  - Jim Blinn

- **Scientific Presentation Planning**
  - Jason Harrison
Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data


Biomechanical motion design study

• large DB of 3D motion data
  – pigs chewing: high-speed motion at joints, 500 FPS w/ sub-mm accuracy

• domain tasks
  – functional morphology: relationship between 3D shape of bones and their function
  – what is a typical chewing motion?
  – how does chewing change over time based on amount/type of food in mouth?

• abstract tasks
  – trends & anomalies across collection of time-varying spatial data
  – understanding complex spatial relationships

• pioneering design study integrating infovis+scivis techniques

• let’s start with video showing system in action
  
  https://youtu.be/OUNezRNtE9M
Multiple linked spatial & non-spatial views

- data: 3D spatial, multiple attribs (cyclic)
- encode: 3D spatial, parallel coords, 2D line (xy) plots
- facet: few large multiform views, many small multiples (~100)
  - encode: color by trial for window background
  - view coordination:
    line in parcoord ==
    frame in small mult

3D+2D

- change
  - 3D navigation
    - rotate/translate/zoom
- filter
  - zoom to small subset of time
- facet
  - select for one large detail view
  - linked highlighting
  - linked navigation
    - between all views
    - driven by large detail view

Derived data: traces/streamers

- derived data: 3D motion tracers from interactively chosen spots
  - generates x/y/z data over time
  - streamers
  - shown in 3D views directly
  - populates 2D plots

Small multiples for overview

- facet: small multiples for overview
  - aggressive/ambitious, 100+ views

- encode: color code window bg by trial

- filter:
  - full/partial skull
  - streamers
    - simple enough to be useable at low information density

Derived data: surface interactions

- derived data
  - 3D surface interaction patterns
- facet
  - superimposed overlays in 3D view
- encoding
  - color coding

Side by side views demonstrating tooth slide

- facet: linked navigation w/ same 3D viewpoint for all
- encode: coloured by vertical distance separating teeth (derived surface interactions)
  - also 3D instantaneous helical axis showing motion of mandible relative to skull

Cluster detection

• identify clusters of motion cycles
  – from combo: 2D xy plots & parcoords
  – show motion itself in 3D view

• facet: superimposed layers
  – foreground/background layers in parcoord view itself

Analysis summary

• what: data
  – 3D spatial, multiple attributes (cyclic)
• what: derived
  – 3D motion traces
  – 3D surface interaction patterns
• how: encode
  – 3D spatial, parallel coords, 2D plots
  – color views by trial, surfaces by interaction patterns
• how: change
  – 3D navigation
• how: facet
  – few large multiform views
  – many small multiples (~100)
  – linked highlighting
  – linked navigation
  – layering
• how: reduce
  – filtering

Critique

• many strengths
  – carefully designed with well justified design choices
  – explicitly followed mantra “overview first, zoom and filter, then details-on-demand”
  – sophisticated view coordination
  – tradeoff between strengths of small multiples and overlays, use both
    – informed by difficulties of animation for trend analysis
    – derived data tracing paths

• weaknesses/limitations
  – (older paper feels less novel, but must consider context of what was new)
  – scale analysis: collection size of $\leq 100$, not thousands (understandably)
  – aggressive about multiple views, arguably pushing limits of understandability
Paper: TopoFisheye
Topological Fisheye Views

• derived data
  – input: laid-out network (spatial positions for nodes)
  – output: multilevel hierarchy from graph coarsening

• interaction
  – user changed selected focus point

• visual encoding
  – hybrid view made from cut through several hierarchy levels

[Fig 4,8. Topological Fisheye Views for Visualizing Large Graphs. Gansner, Koren and North, IEEE TVCG 11(4), p 457-468, 2005]
Coarsening requirements

- uniform cluster/metanode size
- match coarse and fine layout geometries
- scalable

Coarsening strategy

- must preserve graph-theoretic properties
- use both topology and geometry
  - topological distance (hops away)
  - geometric distance - but not just proximity alone!
    - just contracting nodes/edges could create new cycles
- derived data: proximity graph

what not to do!

[Fig 10, 12. Topological Fisheye Views for Visualizing Large Graphs. Gansner, Koren and North, IEEE TVCG 11(4), p 457-468, 2005]
Candidate pairs: neighbors in original and proximity graph

• proximity graph: compromise between larger DT and smaller RNG
  – better than original graph neighbors alone
    • slow for cases like star graph

• maximize weighted sum of
  – geometric proximity
    • goal: preserve geometry
  – cluster size
    • goal: keep uniform cluster size
  – normalized connection strength
    • goal: preserve topology
  – neighborhood similarity
    • goal: preserve topology
  – degree
    • goal: penalize high-degree nodes to avoid salient artifacts and computational problems
Hybrid graph creation

- cut through coarsening hierarchy to get active nodes
  – animated transitions between states

[Fig 10, 12. Topological Fisheye Views for Visualizing Large Graphs. Gansner, Koren and North, IEEE TVCG 11(4), p 457-468, 2005]
Final distortion

- geometric distortion for uniform density
- (colorcoded by hierarchy depth just to illustrate algorithm)
  - compare to original
  - compare to simple topologically unaware fisheye distortion

[Fig 2,15. Topological Fisheye Views for Visualizing Large Graphs. Gansner, Koren and North, IEEE TVCG 11(4), p 457-468, 2005]
Ch 13: Reduce
Reduce items and attributes

• reduce/increase: inverses

• filter
  – pro: straightforward and intuitive
    • to understand and compute
  – con: out of sight, out of mind

• aggregation
  – pro: inform about whole set
  – con: difficult to avoid losing signal

• not mutually exclusive
  – combine filter, aggregate
  – combine reduce, change, facet
**Idiom:** cross filtering

- item filtering
- coordinated views/controls combined
  - all scented histogram bisliders update when any ranges change

[System: Crossfilter](http://square.github.io/crossfilter/)
Idiom: cross filtering

[https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0]
Idiom: histogram

• static item aggregation
• task: find distribution
• data: table
• derived data
  – new table: keys are bins, values are counts
• bin size crucial
  – pattern can change dramatically depending on discretization
  – opportunity for interaction: control bin size on the fly
Idiom: **scented widgets**

- augmented widgets show information scent
  - cues to show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider


Scented histogram bisliders: detailed

Idiom: **Continuous scatterplot**

- static item aggregation
- data: table
- derived data: table
  - key attribs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color: sequential categorical hue + ordered luminance colormap

Spatial aggregation

• MAUP: Modifiable Areal Unit Problem
  – gerrymandering (manipulating voting district boundaries) is only one example!
  – zone effects

[http://www.e-education.psu.edu/geog486/l4_p7.html, Fig 4.cg.6]

– scale effects

**Idiom:** boxplot

- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown

[pods and the rug plot looks like the seeds within] Kampstra nyww8o also suggests a way of comparing two groups more easily: use the left and right sides of the bean to display different distributions. A related idea is the raindrop plot nBarrowman and Myers] ywwzos but its focus is on the display of error distributions from complex models.

Figure fi demonstrates these density boxplots applied to numbers drawn from each of four distributions with mean and standard deviation:

- a standard normal distribution
- a skew-right distribution
- a leptokurtic distribution
- a bimodal distribution

Richer displays of density make it much easier to see important variations in the distribution: multimodality is particularly important and yet completely invisible with the boxplot.

[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]
Idiom: **Hierarchical parallel coordinates**

- dynamic item aggregation
- derived data: **hierarchical clustering**
- encoding:
  - cluster band with variable transparency, line at mean, width by min/max values
  - color by proximity in hierarchy

Idiom: aggregation via hierarchical clustering (visible)

System:
Hierarchical Clustering Explorer

[http://www.cs.umd.edu/hcil/hce/]
Dimensionality reduction

• attribute aggregation
  – derive low-dimensional target space from high-dimensional measured space
    • capture most of variance with minimal error
  – use when you can’t directly measure what you care about
    • true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
    • latent factors, hidden variables

Tumor Measurement Data
  data: 9D measured space
  \[\rightarrow\text{DR}\rightarrow\text{Benign}\]
  derived data: 2D target space
Dimensionality vs attribute reduction

- vocab use in field not consistent
  - dimension/attribute

- attribute reduction: reduce set with filtering
  - includes orthographic projection

- dimensionality reduction: create smaller set of new dims/attribs
  - typically implies dimensional aggregation, not just filtering
  - vocab: projection/mapping
Dimensionality reduction & visualization

• why do people do DR?
  – improve performance of downstream algorithm
    • avoid curse of dimensionality
  – data analysis
    • if look at the output: visual data analysis

• abstract tasks when visualizing DR data
  – dimension-oriented tasks
    • naming synthesized dims, mapping synthesized dims to original dims
  – cluster-oriented tasks
    • verifying clusters, naming clusters, matching clusters and classes

Dimension-oriented tasks

- naming synthesized dims: inspect data represented by lowD points

Cluster-oriented tasks

• verifying, naming, matching to classes

Idiom: Dimensionality reduction for documents

Task 1

<table>
<thead>
<tr>
<th>In</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD data</td>
<td>2D data</td>
</tr>
</tbody>
</table>

What?
- In High-dimensional data
- Out 2D data

Why?
- Produce
- Derive

Task 2

<table>
<thead>
<tr>
<th>In</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D data</td>
<td>Scatterplot Clusters &amp; points</td>
</tr>
</tbody>
</table>

What?
- In 2D data
- Out Scatterplot
- Out Clusters & points

Why?
- Discover
- Explore
- Identify

How?
- Encode
- Navigate
- Select

Task 3

<table>
<thead>
<tr>
<th>In</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scatterplot Clusters &amp; points</td>
<td>Labels for clusters</td>
</tr>
</tbody>
</table>

What?
- In Scatterplot Clusters & points
- Out Labels for clusters

Why?
- Produce
- Annotate
Interacting with dimensionally reduced data

[https://uclab.fh-potsdam.de/projects/probing-projections/]
Linear dimensionality reduction

• principal components analysis (PCA)
  – finding axes: first with most variance, second with next most, …
  – describe location of each point as linear combination of weights for each axis
  • mapping synthesized dims to original dims

Nonlinear dimensionality reduction

• pro: can handle curved rather than linear structure
• cons: lose all ties to original dims/attrs
  – new dimensions often cannot be easily related to originals
    – mapping synthesized dims to original dims task is difficult

• many techniques proposed
  – many literatures: visualization, machine learning, optimization, psychology, ...
  – techniques: t-SNE, MDS (multidimensional scaling), charting, isomap, LLE,…
    – t-SNE: excellent for clusters
      – but some trickiness remains: http://distill.pub/2016/misread-tsne/
    – MDS: confusingly, entire family of techniques, both linear and nonlinear
      – minimize stress or strain metrics
      – early formulations equivalent to PCA
VDA with DR example: nonlinear vs linear

- DR for computer graphics reflectance model
  - goal: simulate how light bounces off materials to make realistic pictures
    - computer graphics: BRDF (reflectance)
  - idea: measure what light does with real materials

[Fig 2. Matusik, Pfister, Brand, and McMillan. A Data-Driven Reflectance Model. SIGGRAPH 2003]
Capturing & using material reflectance

• reflectance measurement: interaction of light with real materials (spheres)
  – each image 4M pixels

• result: 104 high-res images of material
  – each image 4M pixels

• goal: image synthesis
  – simulate completely new materials

• need for more concise model
  – 104 materials * 4M pixels = 400M dims
  – want concise model with meaningful knobs
    • how shiny/greasy/metallic
    • DR to the rescue!

[Figs 5/6. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]
Linear DR

• first try: PCA (linear)
• result: error falls off sharply after ~45 dimensions
  – scree plots: error vs number of dimensions in lowD projection
• problem: physically impossible intermediate points when simulating new materials
  – specular highlights cannot have holes!

[Fig 6/7. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]
Nonlinear DR

• second try: charting (nonlinear DR technique)
  – scree plot suggests 10-15 dims
  – note: dim estimate depends on technique used!

[Fig 10/11. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]
Finding semantics for synthetic dimensions

• look for meaning in scatterplots
  – synthetic dims created by algorithm but named by human analysts
  – points represent real-world images (spheres)
  – people inspect images corresponding to points to decide if axis could have meaningful name

• cross-check meaning
  – arrows show simulated images (teapots) made from model
  – check if those match dimension semantics

[Fig 12/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]
Understanding synthetic dimensions

Specular-Metallic

Diffuseness-Glossiness

[Fig 13/14/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]
Ch 14: Embed
Embed: Focus+Context

- combine information within single view
- elide
  - selectively filter and aggregate
- superimpose layer
  - local lens
- distortion design choices
  - region shape: radial, rectilinear, complex
  - how many regions: one, many
  - region extent: local, global
  - interaction metaphor
Idiom: **DOITrees Revisited**

- elide
  - some items dynamically filtered out
  - some items dynamically aggregated together
  - some items shown in detail

Idiom: **Fisheye Lens**

- distort geometry
  - shape: radial
  - focus: single extent
  - extent: local
  - metaphor: draggable lens

http://tulip.labri.fr/TulipDrupal/?q=node/351
http://tulip.labri.fr/TulipDrupal/?q=node/371
Idiom: **Fisheye Lens**

System: **D3**

[D3 Fisheye Lens](https://bost.ocks.org/mike/fisheye/)
Idiom: *Stretch and Squish Navigation*

- distort geometry
  - shape: rectilinear
  - foci: multiple
  - impact: global
  - metaphor: stretch and squish, borders fixed

System: *TreeJuxtaposer*

Distortion costs and benefits

• benefits
  – combine focus and context information in single view

• costs
  – length comparisons impaired
    • network/tree topology comparisons unaffected: connection, containment
  – effects of distortion unclear if original structure unfamiliar
  – object constancy/tracking maybe impaired

Ch 15: Case Studies
Analysis Case Studies

Scagnostics

VisDB

InterRing

HCE

PivotGraph

Constellation
Graph-Theoretic Scagnostics

• scatterplot diagnostics
  – scagnostics SPLOM: each point is one original scatterplot

[Graph-Theoretic Scagnostics Wilkinson, Anand, and Grossman. Proc InfoVis 05.]
## Scagnostics analysis

<table>
<thead>
<tr>
<th>System</th>
<th>Scagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What: Data</strong></td>
<td>Table.</td>
</tr>
<tr>
<td><strong>What: Derived</strong></td>
<td>Nine quantitative attributes per scatterplot (pairwise combination of original attributes).</td>
</tr>
<tr>
<td><strong>Why: Tasks</strong></td>
<td>Identify, compare, and summarize; distributions and correlation.</td>
</tr>
<tr>
<td><strong>How: Encode</strong></td>
<td>Scatterplot, scatterplot matrix.</td>
</tr>
<tr>
<td><strong>How: Manipulate</strong></td>
<td>Select.</td>
</tr>
<tr>
<td><strong>How: Facet</strong></td>
<td>Juxtaposed small-multiple views coordinated with linked highlighting, popup detail view.</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>Original attributes: dozens.</td>
</tr>
</tbody>
</table>
VisDB

- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

VisDB Results

• partition into many small regions: dimensions grouped together
VisDB Results

- partition into small number of views
  - inspect each attribute

## VisDB Analysis

<table>
<thead>
<tr>
<th>System</th>
<th>VisDB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What: Data</strong></td>
<td>Table (database) with $k$ attributes; query returning table subset (database query).</td>
</tr>
<tr>
<td><strong>What: Derived</strong></td>
<td>$k + 1$ quantitative attributes per original item: query relevance for the $k$ original attributes plus overall relevance.</td>
</tr>
<tr>
<td><strong>Why: Tasks</strong></td>
<td>Characterize distribution within attribute, find groups of similar values within attribute, find outliers within attribute, find correlation between attributes, find similar items.</td>
</tr>
<tr>
<td><strong>How: Encode</strong></td>
<td>Dense, space-filling; area marks in spiral layout; colormap: categorical hues and ordered luminance.</td>
</tr>
<tr>
<td><strong>How: Facet</strong></td>
<td>Layout 1: partition by attribute into per-attribute views, small multiples. Layout 2: partition by items into per-item glyphs.</td>
</tr>
<tr>
<td><strong>How: Reduce</strong></td>
<td>Filtering</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>Attributes: one dozen. Total items: several million. Visible items (using multiple views, in total): one million. Visible items (using glyphs): 100,000</td>
</tr>
</tbody>
</table>
Hierarchical Clustering Explorer

- heatmap, dendrogram
- multiple views

[Interactively Exploring Hierarchical Clustering Results. Seo and Shneiderman, IEEE Computer 35(7): 80-86 (2002)]
<table>
<thead>
<tr>
<th><strong>System</strong></th>
<th>Hierarchical Clustering Explorer (HCE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What: Data</strong></td>
<td>Multidimensional table: two categorical key attributes (genes, conditions); one quantitative value attribute (gene activity level in condition).</td>
</tr>
<tr>
<td><strong>What: Derived</strong></td>
<td>Hierarchical clustering of table rows and columns (for cluster heatmap); quantitative derived attributes for each attribute and pairwise attribute combination; quantitative derived attribute for each ranking criterion and original attribute combination.</td>
</tr>
<tr>
<td><strong>Why: Tasks</strong></td>
<td>Find correlation between attributes; find clusters, gaps, outliers, trends within items.</td>
</tr>
<tr>
<td><strong>How: Encode</strong></td>
<td>Cluster heatmap, scatterplots, histograms, boxplots. Rank-by-feature overviews: continuous diverging colormaps on area marks in reorderable 2D matrix or 1D list alignment.</td>
</tr>
<tr>
<td><strong>How: Reduce</strong></td>
<td>Dynamic filtering; dynamic aggregation.</td>
</tr>
<tr>
<td><strong>How: Manipulate</strong></td>
<td>Navigate with pan/scroll.</td>
</tr>
<tr>
<td><strong>How: Facet</strong></td>
<td>Multiform with linked highlighting and shared spatial position; overview–detail with selection in overview populating detail view.</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>Genes (key attribute): 20,000. Conditions (key attribute): 80. Gene activity in condition (quantitative value attribute): 20,000 \times 80 = 1,600,000.</td>
</tr>
</tbody>
</table>
InterRing

## InterRing Analysis

<table>
<thead>
<tr>
<th>System</th>
<th>InterRing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What: Data</strong></td>
<td>Tree.</td>
</tr>
<tr>
<td><strong>Why: Tasks</strong></td>
<td>Selection, rollup/drilldown, hierarchy editing.</td>
</tr>
<tr>
<td><strong>How: Facet</strong></td>
<td>Linked coloring and highlighting.</td>
</tr>
<tr>
<td><strong>How: Reduce</strong></td>
<td>Embed: distort; multiple foci.</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>Nodes: hundreds if labeled, thousands if dense. Levels in tree: dozens.</td>
</tr>
</tbody>
</table>
PivotGraph

• derived rollup network

[Visual Exploration of Multivariate Graphs, Martin Wattenberg, CHI 2006.]
[Visual Exploration of Multivariate Graphs, Martin Wattenberg, CHI 2006.]
## PivotGraph Analysis

<table>
<thead>
<tr>
<th>Idiom</th>
<th>PivotGraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Network.</td>
</tr>
<tr>
<td>What: Derived</td>
<td>Derived network of aggregate nodes and links by roll-up into two chosen attributes.</td>
</tr>
<tr>
<td>Why: Tasks</td>
<td>Cross-attribute comparison of node groups.</td>
</tr>
<tr>
<td>How: Encode</td>
<td>Nodes linked with connection marks, size.</td>
</tr>
<tr>
<td>How: Reduce</td>
<td>Aggregation, filtering.</td>
</tr>
<tr>
<td>Scale</td>
<td>Nodes/links in original network: unlimited. Roll-up attributes: 2. Levels per roll-up attribute: several, up to one dozen.</td>
</tr>
</tbody>
</table>
Analysis example: Constellation

• data
  – multi-level network
    • node: word
    • link: words used in same dictionary definition
  – subgraph for each definition
    – not just hierarchical clustering
  – paths through network
    • query for high-weight paths between 2 nodes
      – quant attrib: plausibility

Using space: Constellation

- visual encoding
  - link connection marks between words
  - link containment marks to indicate subgraphs
  - encode plausibility with horiz spatial position
  - encode source/sink for query with vert spatial position

- spatial layout
  - curvilinear grid: more room for longer low-plausibility paths

Using space: Constellation

- edge crossings
  - cannot easily minimize instances, since position constrained by spatial encoding
  - instead: minimize perceptual impact

- views: superimposed layers
  - dynamic foreground/background layers onmouseover, using color
  - four kinds of constellations
    - definition, path, link type, word
      - not just 1-hop neighbors

https://youtu.be/7sjC3QVpSkQ

<table>
<thead>
<tr>
<th>System</th>
<th>Constellation</th>
</tr>
</thead>
<tbody>
<tr>
<td>What: Data</td>
<td>Three-level network of paths, subgraphs (definitions), and nodes (word senses).</td>
</tr>
<tr>
<td>Why: Tasks</td>
<td>Discover/verify: browse and locate types of paths, identify and compare.</td>
</tr>
<tr>
<td>How: Encode</td>
<td>Containment and connection link marks, horizontal spatial position for plausibility attribute, vertical spatial position for order within path, color links by type.</td>
</tr>
<tr>
<td>How: Reduce</td>
<td>Superimpose dynamic layers.</td>
</tr>
</tbody>
</table>
What-Why-How Analysis

• this approach is not the only way to analyze visualizations!
  – one specific framework intended to help you think
  – other frameworks support different ways of thinking
    • following: one interesting example
Algebraic Process for Visualization Design

- which mathematical structures in data are preserved and reflected in visualization — negation, permutation, symmetry, invariance

Algebraic process: Vocabulary

• **invariance** violation: single dataset, many visualizations
  – **hallucinator**

• **unambiguity** violation: many datasets, same vis
  – data change invisible to viewer
    • **confuser**

• **correspondence** violation:
  – can’t see change of data in vis
    • **jumbler**
      – salient change in vis not due to significant change in data
    • **misleader**
      – match mathematical structure in data with visual perception

• **we can X the data; can we Y the image?**
  – are important data changes well-matched with obvious visual changes?
Next time

• deadlines
  – meetings due by Fri Nov 2, 6pm
    • several of the projects are not yet signed off, slots filling up fast
  – proposals due by Mon Nov 5, 10pm

• next week
  – presentations 1
  – finishing up chapters discussions