Lecture:
Case Studies, Reproducibility

Tamara Munzner
Department of Computer Science
University of British Columbia

CPSC 547, Information Visualization
12 November 2020

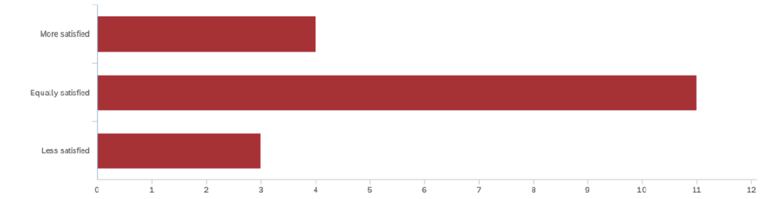
http://www.cs.ubc.ca/~tmm/courses/547-20

Survey feedback

- mixed responses
- Q4/Q5: best **and** worst
  - async online discussion
  - in-class group work exercises during sync class time

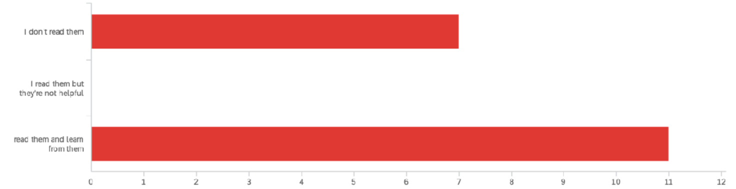
Survey: Q1

Q1 - This year for the virtual course we're doing discussion asynchronously in writing instead of live by talking. When you think about how it's working vs a more traditional live discussion are you (more, equal, less) satisfied?



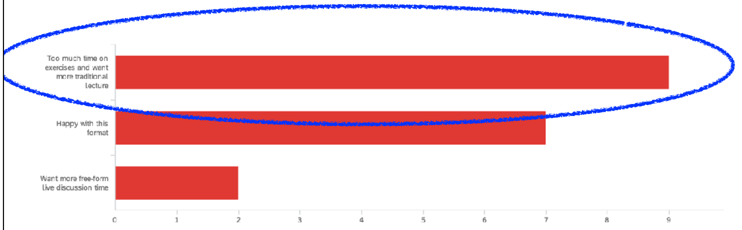
Survey: Q3

Q3 - How helpful are the instructor's contributions to the asynchronous discussion that are posted at the end of the week?



Survey: Q2

Q2 - This year for the virtual course we're mostly using the synchronous class time for in-class exercises (and reportbacks plus discussion) with very minimal traditional live lecture from the instructor. When you think about how it's working vs a more traditional live lecture, do you think the ratio is:



Today: Lecture

- case studies
  - Biomechanical Motion
  - VAD Ch 15 (not assigned as reading)
    - Scagnostics, VisDB, InterRing, HCE, PivotGraph, Constellation
- Algebraic Design
- replicability crisis / credibility revolution

Biomechanical Motion

Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data

Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.

http://ivlab.cs.umn.edu/generated/pub-Keefe-2009-MultiViewVis.php

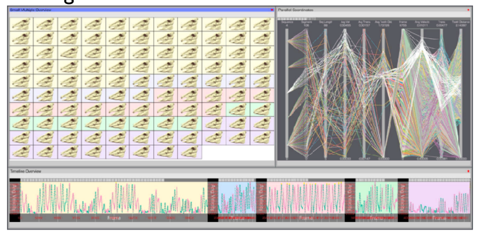
Biomechanical motion design study

- 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

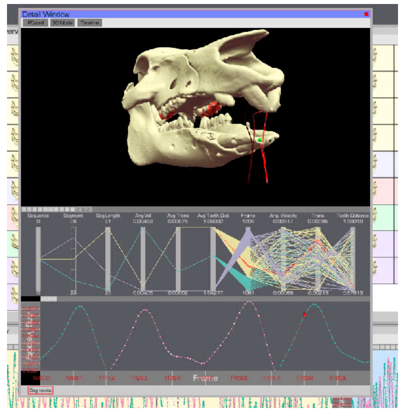
- data: 3D spatial, multiple attribs (cyclic)
- encode: 3D spatial, parallel coords, 2D line (xy) plots
- facet: few large multiform views, many small multiples (~100)
  - encode: color by trial for window background
  - view coordination: line in parcoord == frame in small mult



[Fig 1. Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

3D+2D

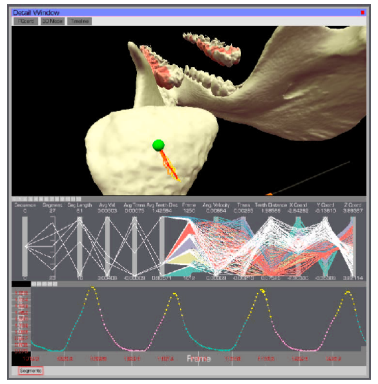
- change
  - 3D navigation
    - rotate/translate/zoom
- filter
  - zoom to small subset of time
- facet
  - select for one large detail view
  - linked highlighting
  - linked navigation
    - between all views
    - driven by large detail view



[Fig 3. Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

Derived data: traces/streamers

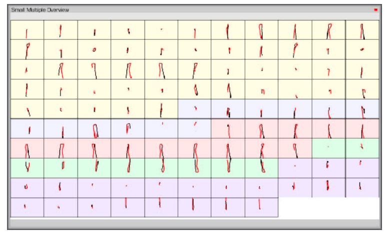
- derived data: 3D motion tracers from interactively chosen spots
  - generates x/y/z data over time
    - streamers
  - shown in 3D views directly
  - populates 2D plots



[Fig 4. Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

Small multiples for overview

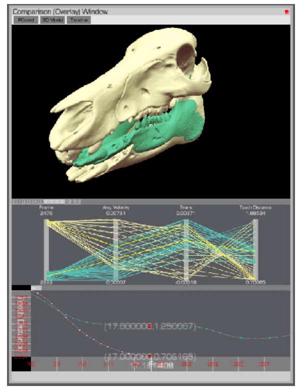
- facet: small multiples for overview
  - aggressive/ambitious, 100+ views
- encode: color code window bg by trial
- filter:
  - full/partial skull
  - streamers
    - simple enough to be useable at low information density



[Fig 2. Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

Derived data: surface interactions

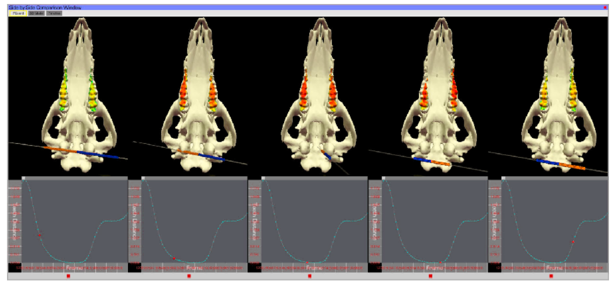
- derived data
  - 3D surface interaction patterns
- facet
  - superimposed overlays in 3D view
- encoding
  - color coding



[Fig 5. Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

Side by side views demonstrating tooth slide

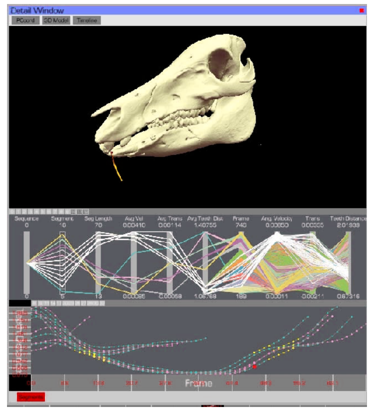
- facet: linked navigation w/ same 3D viewpoint for all
- encode: coloured by vertical distance separating teeth (derived surface interactions)
  - also 3D instantaneous helical axis showing motion of mandible relative to skull



[Fig 6. Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

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



[Fig 7. Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

Analysis summary

- what: data
  - 3D spatial, multiple attribs (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

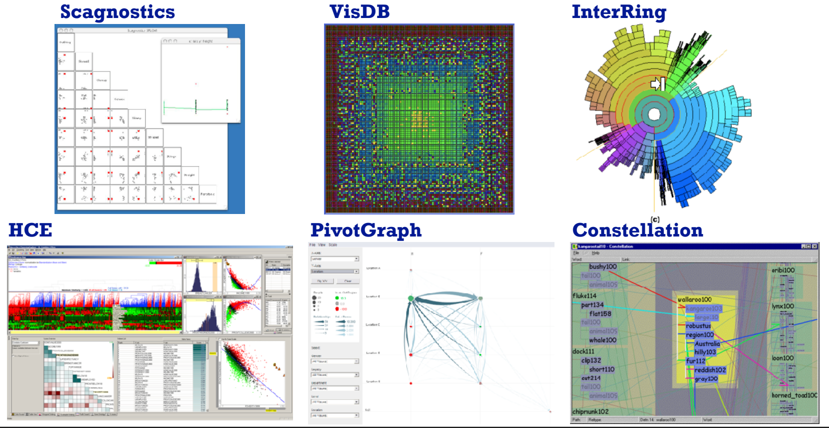
[Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data. Daniel F. Keefe, Marcus Ewert, William Ribarsky, Remco Chang. IEEE Trans. Visualization and Computer Graphics (Proc. Vis 2009), 15(6):1383-1390, 2009.]

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 <=100, not thousands (understandably)
  - aggressive about multiple views, arguably pushing limits of understandability

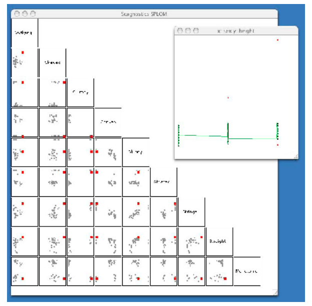
Case Studies

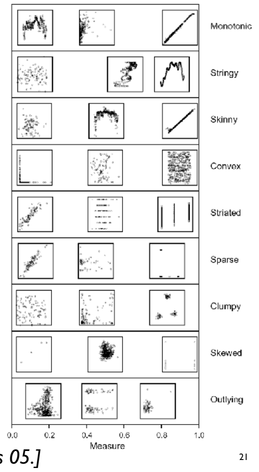
Analysis Case Studies



Graph-Theoretic Scagnostics

- scatterplot diagnostics
  - scagnostics SPLOM: each point is one original scatterplot





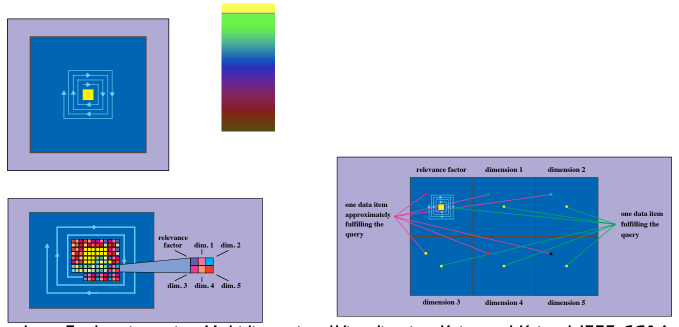
[Graph-Theoretic Scagnostics Wilkinson, Anand, and Grossman. Proc InfoVis 05.]

Scagnostics analysis

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

VisDB

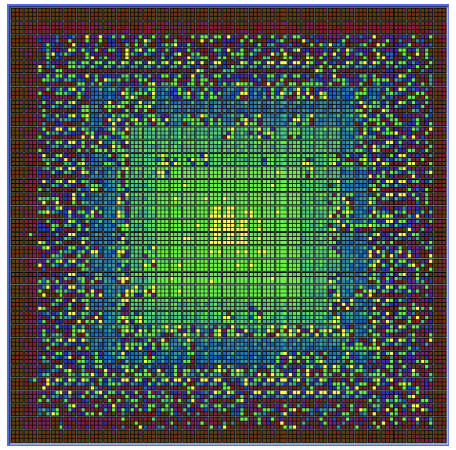
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views



[VisDB: Database Exploration using Multidimensional Visualization, Keim and Kriegel, IEEE CG&A, 1994]

VisDB Results

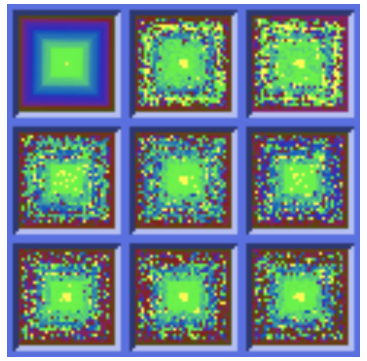
- partition into many small regions: dimensions grouped together



[VisDB: Database Exploration using Multidimensional Visualization, Keim and Kriegel, IEEE CG&A, 1994]

VisDB Results

- partition into small number of views
  - inspect each attribute



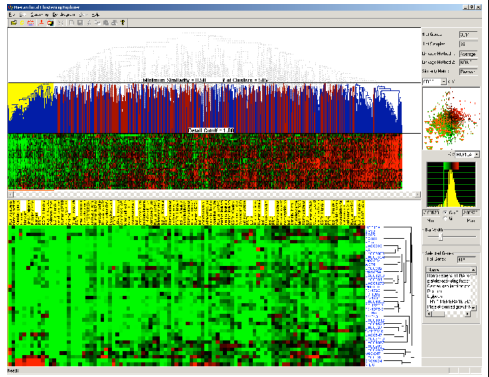
[VisDB: Database Exploration using Multidimensional Visualization, Keim and Kriegel, IEEE CG&A, 1994]

VisDB Analysis

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

Hierarchical Clustering Explorer

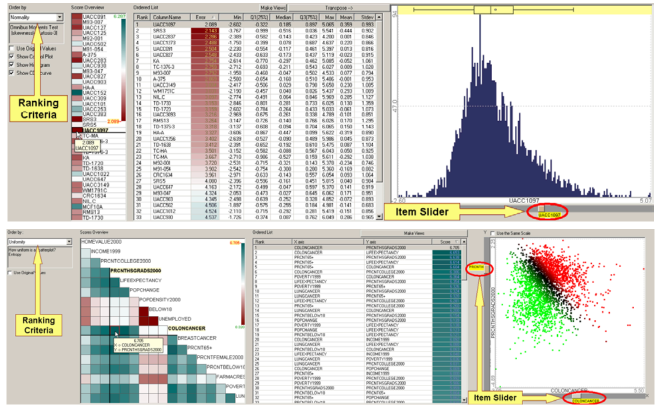
- heatmap, dendrogram
- multiple views



[Interactively Exploring Hierarchical Clustering Results. Seo and Shneiderman, IEEE Computer 35(7): 80-86 (2002)]

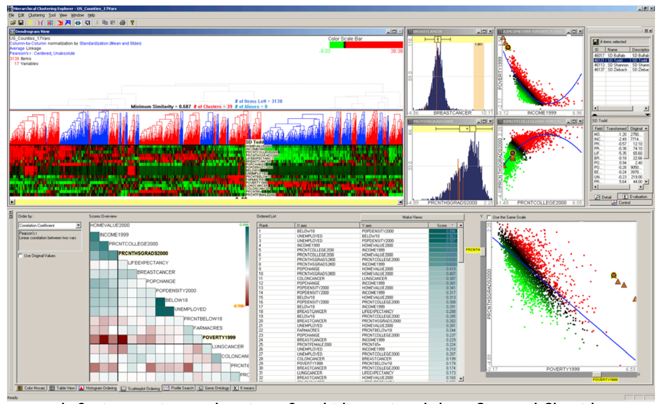
HCE

- rank by feature idiom
  - 1D list
  - 2D matrix



A rank-by-feature framework for interactive exploration of multidimensional data. Seo and Shneiderman. Information Visualization 4(2): 96-113 (2005)

HCE

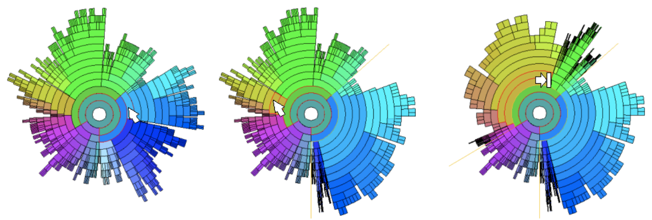


A rank-by-feature framework for interactive exploration of multidimensional data. Seo and Shneiderman. Information Visualization 4(2): 96-113 (2005)

HCE Analysis

System	Hierarchical Clustering Explorer (HCE)
What: Data	Multidimensional table: two categorical key attributes (genes, conditions); one quantitative value attribute (gene activity level in condition).
What: Derived	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.
Why: Tasks	Find correlation between attributes; find clusters, gaps, outliers, trends within items.
How: Encode	Cluster heatmap, scatterplots, histograms, boxplots. Rank-by-feature overviews: continuous diverging colormaps on area marks in reorderable 2D matrix or 1D list alignment.
How: Reduce	Dynamic filtering; dynamic aggregation.
How: Manipulate	Navigate with pan/scroll.
How: Facet	Multiform with linked highlighting and shared spatial position; overview–detail with selection in overview populating detail view.
Scale	Genes (key attribute): 20,000. Conditions (key attribute): 80. Gene activity in condition (quantitative value attribute): $20{,}000 \times 80 = 1{,}600{,}000$.

InterRing



original hierarchy   blue subtree expanded   tan subtree expanded

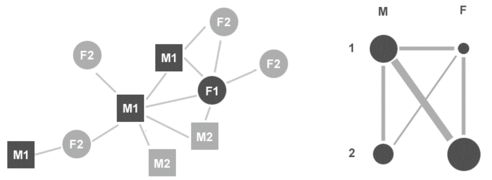
[InterRing: An Interactive Tool for Visually Navigating and Manipulating Hierarchical Structures. Yang, Ward, Rundensteiner. Proc. InfoVis 2002, p 77-84.]

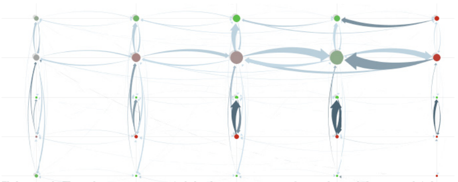
InterRing Analysis

System	InterRing
What: Data	Tree.
Why: Tasks	Selection, rollup/drilldown, hierarchy editing.
How: Encode	Radial, space-filling layout. Color by tree structure.
How: Facet	Linked coloring and highlighting.
How: Reduce	Embed: distort; multiple foci.
Scale	Nodes: hundreds if labeled, thousands if dense. Levels in tree: dozens.

PivotGraph

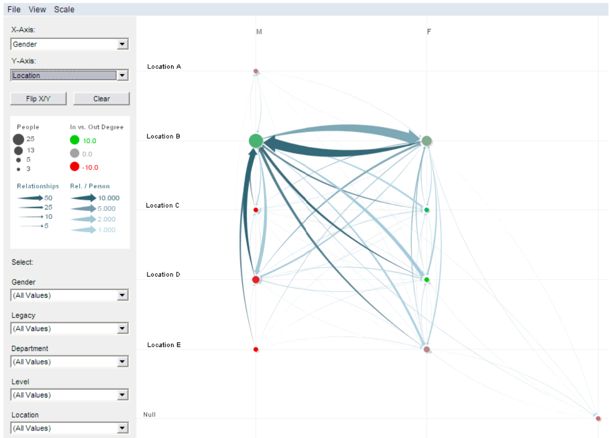
- derived rollup network





[*Visual Exploration of Multivariate Graphs*, Martin Wattenberg, CHI 2006.]

PivotGraph



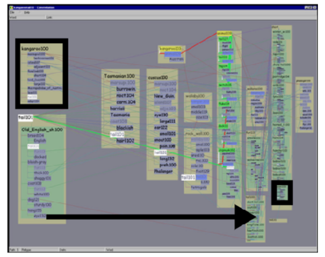
[*Visual Exploration of Multivariate Graphs*, Martin Wattenberg, CHI 2006.]

PivotGraph Analysis

Idiom	PivotGraph
What: Data	Network.
What: Derived	Derived network of aggregate nodes and links by roll-up into two chosen attributes.
Why: Tasks	Cross-attribute comparison of node groups.
How: Encode	Nodes linked with connection marks, size.
How: Manipulate	Change: animated transitions.
How: Reduce	Aggregation, filtering.
Scale	Nodes/links in original network: unlimited. Roll-up attributes: 2. Levels per roll-up attribute: several, up to one dozen.

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

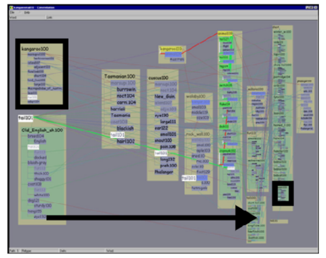


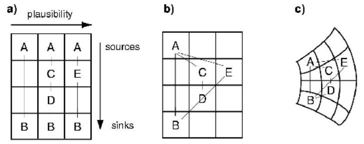
[Interactive Visualization of Large Graphs and Networks. Munzner. Ph.D. Dissertation, Stanford University, June 2000.]

[Constellation: A Visualization Tool For Linguistic Queries from MindNet. Munzner, Guimbretière and Robertson. Proc. IEEE Symp. InfoVis 1999, p.132-135.]

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

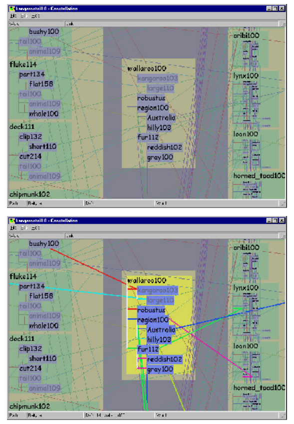




[Interactive Visualization of Large Graphs and Networks. Munzner. Ph.D. Dissertation, Stanford University, June 2000.]

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 on mouseover, using color
  - four kinds of constellations
    - definition, path, link type, word
      - not just 1-hop neighbors



[Interactive Visualization of Large Graphs and Networks. Munzner. Ph.D. Dissertation, Stanford University, June 2000.]

Constellation Analysis

System	Constellation
What: Data	Three-level network of paths, subgraphs (definitions), and nodes (word senses).
Why: Tasks	Discover/verify: browse and locate types of paths, identify and compare.
How: Encode	Containment and connection link marks, horizontal spatial position for plausibility attribute, vertical spatial position for order within path, color links by type.
How: Manipulate	Navigate: semantic zooming. Change: Animated transitions.
How: Reduce	Superimpose dynamic layers.
Scale	Paths: 10–50. Subgraphs: 1–30 per path. Nodes: several thousand.

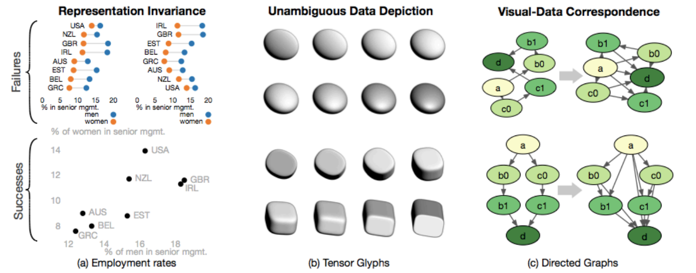
Algebraic Design

What-Why-How Analysis

- expected in your paper/topic presentations
  - in addition to content summarization and general reflection
- expected in your final projects
- 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
    - today's paper is interesting example!

Algebraic Process for Visualization Design

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



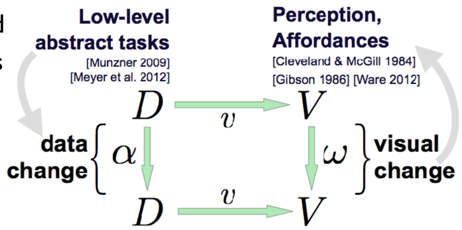
[Fig 1. An Algebraic Process for Visualization Design. Carlos Scheidegger and Gordon Kindlmann. IEEE TVCG (Proc. InfoVis 2014), 20(12):2181-2190.]

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?

Algebraic process: Model

- D: space of data to be visualized
- R: space of data representations
  - r: mapping from D to R
- V: space of visualizations
  - v: mapping from R to V
- α: data symmetries
- ω: visualization symmetries



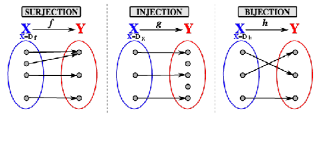
- commutative diagram
  - equality between paths

$$v \circ r_2 \circ \alpha = \omega \circ v \circ r_1$$

$$\begin{array}{ccccc} D & \xrightarrow{r_1} & R & \xrightarrow{v} & V \\ \alpha \downarrow & & & & \downarrow \omega \\ D & \xrightarrow{r_2} & R & \xrightarrow{v} & V \end{array}$$

Algebraic process: Previous work tie-in

- Stevens data types: categorical, ordinal, quant (interval & ratio)
  - defined by symmetry groups and invariances
- Ziemziewicz & Kosara surjective/injective/bijective
  - injectivity: unambiguity
- Mackinlay's Expressiveness Principle
  - convey all and only properties of data
    - invariance/hallucinator, correspondence/misleader
- Mackinlay's Effectiveness Principle
  - match important data attributes to salient visual channels
    - correspondence/jumbler, unambiguity/confuser
- Gibson/Ware affordances
  - perceivable structures show possibility of action
    - correspondence



Algebraic process: Previous work tie-in, cont.

- Tversky Congruence Principle & Apprehension Principle
  - congruence: visual external structure of graphic should correspond to mental internal representation of viewer
  - apprehension: graphics should be readily and easily perceived and comprehended
    - unambiguity and correspondence
- nested model
  - reason about mappings from abstraction to idiom
  - mathematical guidelines for abstraction layer

Reproducible and Replicable Research

Reproducible research

- 5: 15 minutes with free tools
- 4: 15 minutes with proprietary tools
- 3: considerable effort
- 2: extreme effort
- 1: cannot seem to be reproduced
- 0: cannot be reproduced

[Vandewalle, Kovacevic and Vetterli. Reproducible Research in Signal Processing - What, why, and how. IEEE Signal Processing Magazine, 26(3):37-47, May 2009.]

Why bother with reproducibility

- moral high ground
  - for Science!
- enlightened self-interest
  - make your own life easier
  - you'll be cited more often by academics
  - your work is more likely to be used by industry

Reproducibility: Levels to consider

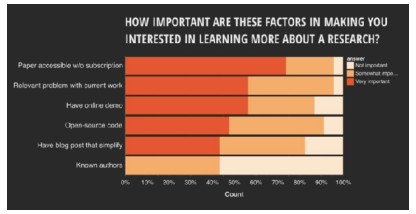
- paper
  - post it online
  - make sure it stays accessible when you move on to new place
  - external archives are better yet (arxiv.org)
- algorithm
  - well documented in paper itself
  - document further with supplemental materials
- code
  - make available as open source
  - pick right spot on continuum of effort involved, from minimal to massive
    - just put it up warts and all, minimal documentation
    - well documented and tested
    - (build a whole community - not the common case)

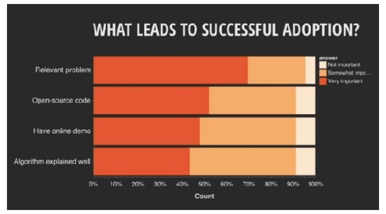
Reproducibility: Levels to consider, cont.

- data
  - make available
    - technique/algorithm: data used by system
      - tricky issue in visualization: data might not be yours to release!
    - evaluation: user study results
      - ethics approval possible if PII (personally identifiable information) sanitized, needs advance planning
- parameters
  - how exactly to regenerate/produce figures, tables
  - example: http://www.cs.utah.edu/~gk/papers/vis03/

View from industry

- Increasing the Impact of Visualization Research panel, VIS 2017
  - Krist Wongsuphasawat, Data Visualization Scientist, Twitter





https://www.slideshare.net/kristw/increasing-the-impact-of-visualization-research

Replication: crisis in psychology, medicine, etc

- early rumblings left me with (ignorable) qualms
  - papers: Is most published research false?, Storks Deliver Babies ($p= 0.008$), The Earth is spherical ($p < 0.05$), False-Positive Psychology
- groundswell of change for what methods are considered legitimate
  - out: QRPs (questionable research practices)
    - p-hacking / p-value fishing / data dredging
    - Hypothesizing After Results are Known (HARKing)
  - in
    - replication
    - pre-registration
  - brouhaha with bimodal responses
    - some people doubling down and defending previous work
    - many willing to repudiate (their own) earlier styles of working

Remarkable introspection on methods

- thoughtful willingness to change standards of field
  - Andrew Gelman's commentary on the Susan Fiske article
    - http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/
  - Simine Vazire's entire Sometimes I'm Wrong blog
    - http://sometimesimwrong.typepad.com/
    - especially posts on topic Scientific Integrity
  - Joe Simmons Data Colada blog post What I Want Our Field to Prioritize
    - http://datacolada.org/53/
  - Dana Carvey's brave statement on her previous power pose work
    - http://faculty.haas.berkeley.edu/dana_carney/pdf_My%20position%20on%20power%20poses.pdf

When and how will this storm hit visualization?

- they're ahead of us
  - they have some paper retractions
    - we don't (yet) have any retractions for methodological considerations
  - they agonize about difficulty of getting failure-to-replicate papers accepted
    - we hardly ever even try to do such work
  - they are a much older field
    - we're younger: might our power hierarchies thus be less entrenched??…
  - they are higher profile
    - we don't have vis research results appear regularly in major newspapers/magazines
  - they have rich fabric of blogs as major drivers of discussion
    - crosscutting traditional power hierarchies
    - we have far fewer active bloggers
- replication crisis was focus of BELIV 2018 workshop at IEEE VIS
  - evaluation and BEyond - methodoLogIcal approaches for Visualization
  - http://beliv.cs.univie.ac.at/