Information Visualization
Facet, Reduce, ScalableInsets
Ex: Complexity Families

Tamara Munzner
Department of Computer Science
University of British Columbia

Week 5, 2 Oct 2025

https://www.cs.ubc.ca/~tmm/courses/547-25

Plan for today

- small group exercises
  - Complexity Families
- break
- mini-lecture / Q&A responses
  - Facet
  - Reduce
  - Scalable Insets

Upcoming

- tomorrow (Fri Oct 3): Groups finalized by noon, tell me via Piazza post
- next week (W6)
  - to read & discuss (async, before next class)
    - VAD book, Ch 14: Embed
    - paper: TensorFlowGraph [design study]
    - paper: TBD (see Piazza)
  - in class
    - pre-proposal meetings with each team
    - project work together when I'm not meeting with your team

In-Class Exercise

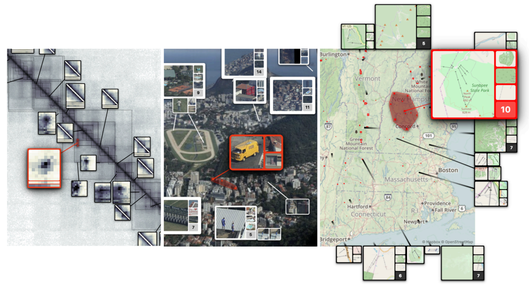
Break: 4-4:10

Mini-Lecture

Scalable Insets

Scalable Insets

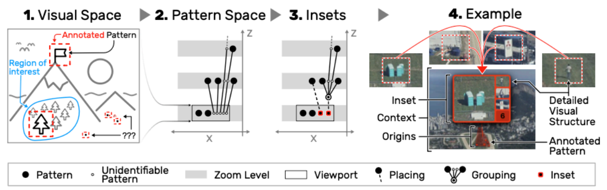
- Pattern-Driven Navigation in 2D Multiscale Visualizations with Scalable Insets.
  Lekschas, Behrisch, Bach, Kerpedjiev, Gehlenborg, Pfister. IEEE TVCG (Proc. VIS 2019) 26(1):611 - 621, 2020
- good example for
  - multiple views
  - aggregation
- technique paper
  - idiom combines encoding & interaction



Pattern-Driven Navigation in 2D Multiscale Visualizations with Scalable Insets. Lekschas, Behrisch, Bach, Kerpedjiev, Gehlenborg, Pfister. IEEE TVCG (Proc. VIS 2019) 26(1):611 - 621, 2020. Fig 1

Key ideas

- multiscale problem: things of interest too small to see from overview
  - pan & zoom (navigation), overview & detail (multiple windows), detail-in-context
  - focus+context: more on this next week (Embed)
  - highlighting
  - aggregation & simplification



Pattern-Driven Navigation in 2D Multiscale Visualizations with Scalable Insets. Lekschas, Behrisch, Bach, Kerpedjiev, Gehlenborg, Pfister. IEEE TVCG (Proc. VIS 2019) 26(1):611 - 621, 2020. Fig 2

Use cases

- technique shown in three applications
  - **large-scale structural genomics matrices**
    - underlying motivation...
  - gigapixel images
  - geographic maps

Key ideas

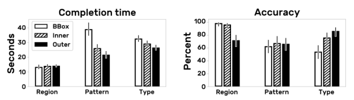
- framing tradeoff: locality, context, details

- algorithms
  - inset placement with simulated annealing
    - cost function as main contribution
  - aggregation
    - derived data: clustering (density-based dynamic) - as discussed in chapter
    - encoding: pile-based gallery

Pattern-Driven Navigation in 2D Multiscale Visualizations with Scalable Insets. Lekschas, Behrisch, Bach, Kerpedjiev, Gehlenborg, Pfister. IEEE TVCG (Proc. VIS 2019) 26(1):611 - 621, 2020. Fig 2 & 3

Evaluation

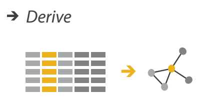
- quant confirmatory study (N=18)
  - 3 methods: compare baseline (bbox) to proposed (SI inside, SI outside)
  - 3 tasks: region, pattern type
  - 5 trials per condition
  - measured time & errors
  - mixed results, some tradeoffs



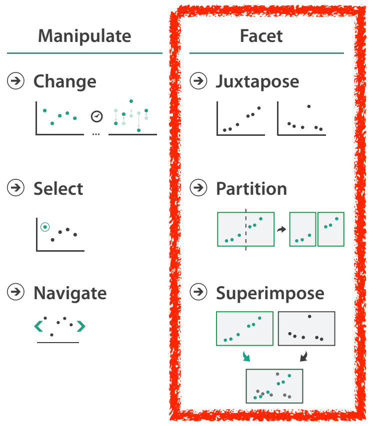
- qual exploratory study with genomics domain experts (N=6)
  - how did people use it
- computational benchmarks
  - frame rate as # insets increases

Multiple Views

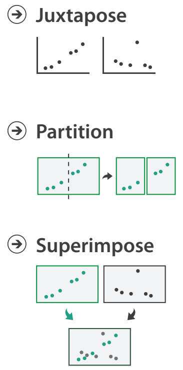
How to handle complexity: 1 previous strategy + 2 more



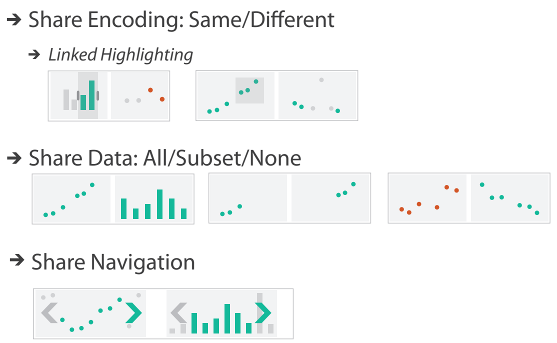
- derive new data to show within view
- change view over time
- facet across multiple views



Facet

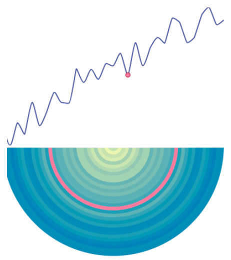


Juxtapose and coordinate views



Linked views: Directionality

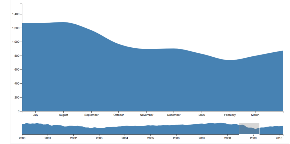
- unidirectional vs bidirectional linking
  - bidirectional almost always better!



http://pbeshai.github.io/linked-highlighting-react-vega-redux/
https://medium.com/@pbesh/linked-highlighting-with-react-d3-js-and-reflux-16e9c0b2210b

Idiom: **Overview-detail navigation**

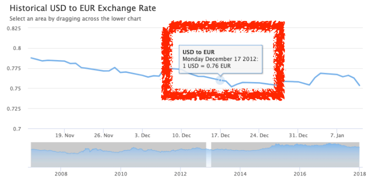
- encoding: same or different
- data: subset shared
- navigation: shared
  - unidirectional linking
  - select in small overview, change extent in large detail view



https://observablehq.com/@uwdata/interaction

Idiom: **Tooltips**

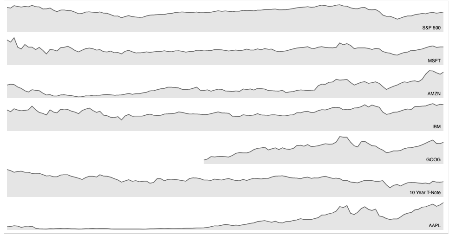
- popup information for selection
  - hover or click
  - specific case of detail view: provide useful additional detail on demand
  - beware: does not support overview!
    - always consider if there's a way to visually encode directly to provide overview
    - "If you make a rollover or tooltip, assume nobody will see it. If it's important, make it explicit."
      – Gregor Aisch, NYTimes



[https://www.highcharts.com/demo/dynamic-master-detail]

Idiom: **Small multiples**

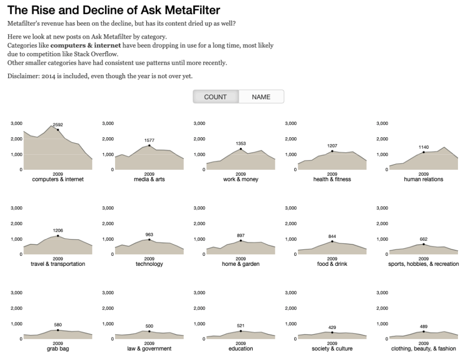
- encoding: same
  - ex: line charts
- data: none shared
  - different slices of dataset
    - items or attributes
    - ex: stock prices for different companies



[https://bl.ocks.org/mbostock/1157787]

Interactive small multiples

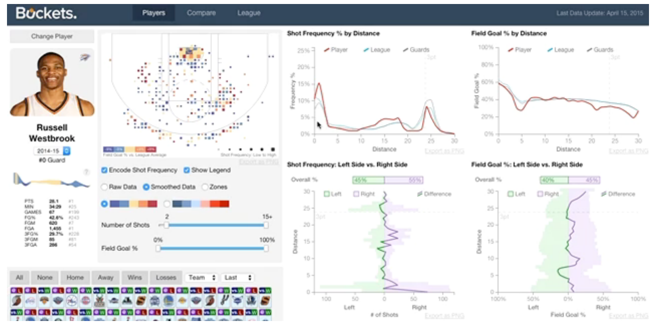
- linked highlighting: analogous item/attribute across views
  - same year highlighted across all charts if hover within any chart



[https://bl.ocks.org/ColinEberhardt/3c780088c363d1515403f50a87a87121]
[https://blog.scottlogic.com/2017/04/05/interactive-responsive-small-multiples.html]
[http://projects.flowingdata.com/tut/linked_small_multiples_demo/]

Example: Combining many interaction idioms

System: **Buckets**

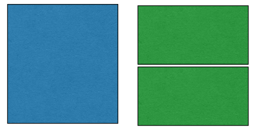


- multiform
- multidirectional linked highlighting of small multiples
- tooltips

http://buckets.peterbeshai.com/

Juxtapose views: tradeoffs

- juxtapose costs
  - display area
    - 2 views side by side: each has only half the area of one view
- juxtapose benefits
  - cognitive load: eyes vs memory
    - lower cognitive load: move eyes between 2 views
    - higher cognitive load: compare single changing view to memory of previous state

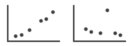


View coordination: Design choices

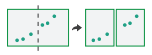
		Data		
		All	Subset	None
Encoding	Same	Redundant	Overview/Detail	Small Multiples
	Different	Multiform	Multiform, Overview/Detail	No Linkage

Facet

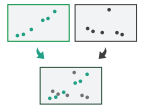
Juxtapose



Partition



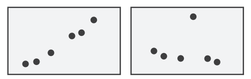
Superimpose



Partition into views

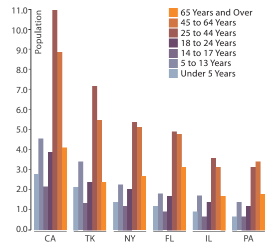
- how to divide data between views
  - split into regions by attributes
  - encodes association between items using spatial proximity
  - order of splits has major implications for what patterns are visible

Partition into Side-by-Side Views



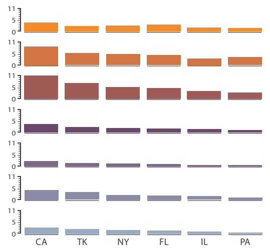
Partitioning: Grouped vs small-multiple bars

- single bar chart with grouped bars
  - split by state into regions
    - complex glyph within each region showing all ages
  - compare: easy within state, hard across ages



[https://observablehq.com/@d3/grouped-bar-chart]

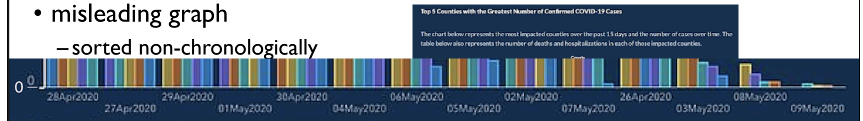
- small-multiple bar charts
  - split by age into regions
    - one chart per region
  - compare: easy within age, harder across states



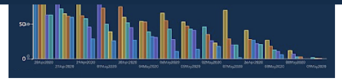
[https://bl.ocks.org/mbostock/4679202]

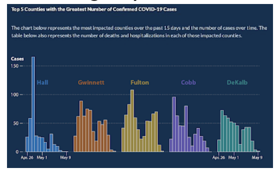
Partioning: Grouped vs small multiples

- misleading graph
  - sorted non-chronologically



- Alberto Cairo redesign:
  - separate by county then sort chronologically



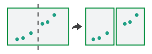


https://thefunctionalart.blogspot.com/2020/05/about-that-weird-georgia-chart.html

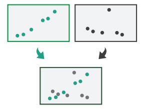
Facet

Juxtapose

Partition



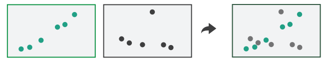
Superimpose



Superimpose layers

- layer: set of objects spread out over region
  - each set is visually distinguishable group
  - extent: whole view
- design choices
  - how many layers, how to distinguish?
    - encode with different, nonoverlapping channels
    - two layers achievable, three with careful design
  - small static set, or dynamic from many possible?

Superimpose Layers



Static visual layering

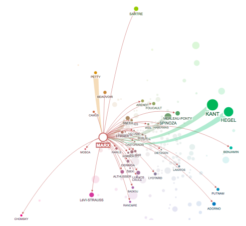
- foreground layer: roads
  - hue, size distinguishing main from minor
  - high luminance contrast from background
- background layer: regions
  - desaturated colors for water, parks, land areas
- user can selectively focus attention

[Get it right in black and white. Stone. 2010.
http://www.stonesc.com/wordpress/2010/03/get-it-right-in-black-and-white]

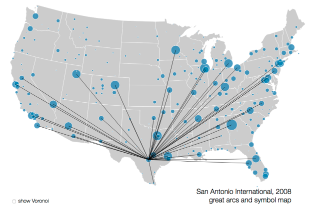
Dynamic visual layering

- interactive, based on selection
- one-hop neighbour highlighting

click (heavyweight)

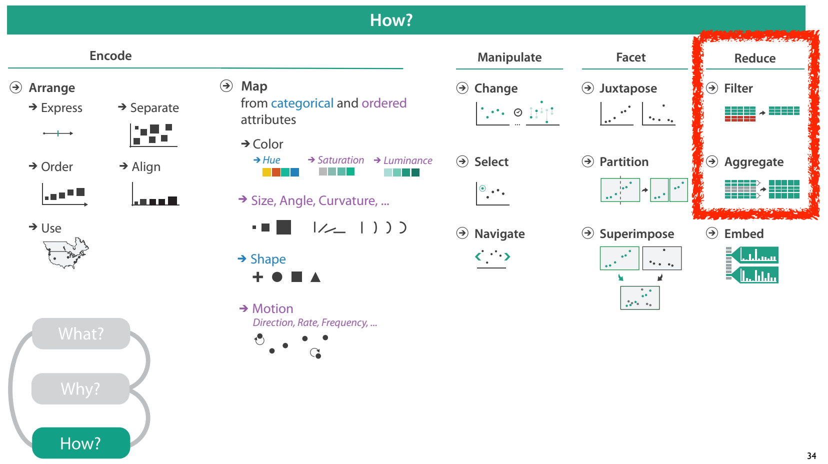


hover (fast)



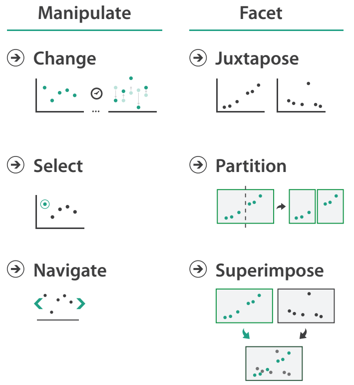
https://mariandoerk.de/edgemaps/demo/
http://mbostock.github.io/d3/talk/20111116/airports.html

Reduce: Aggregation & Filtering



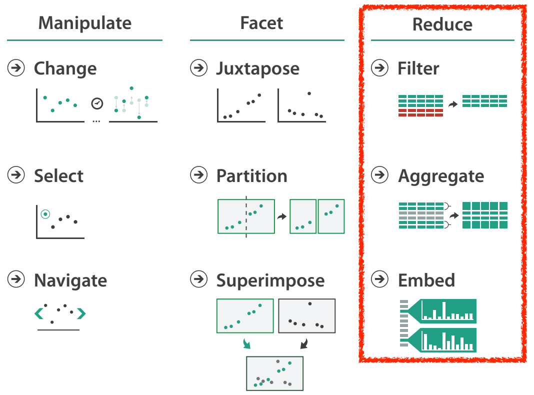
How to handle complexity: 3 previous strategies

- derive new data to show within view
- change view over time
- facet across multiple views



How to handle complexity: 3 previous strategies + 1 more

- derive new data to show within view
- change view over time
- facet across multiple views
- reduce items/attributes within single view

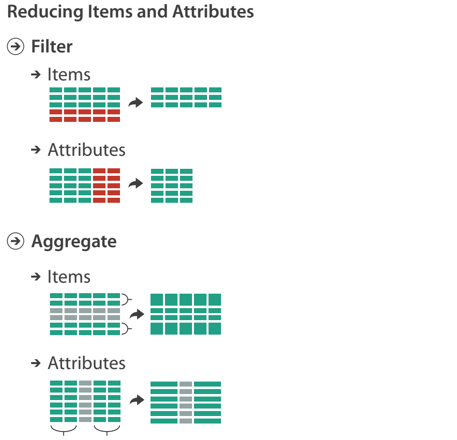


Reduce items and attributes

- reduce/increase: inverses
- filter
  - pro: straightforward and intuitive
    - to understand and compute
  - con: out of sight, out of mind

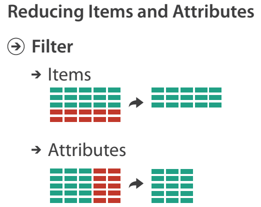
Reduce items and attributes

- 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



Filter

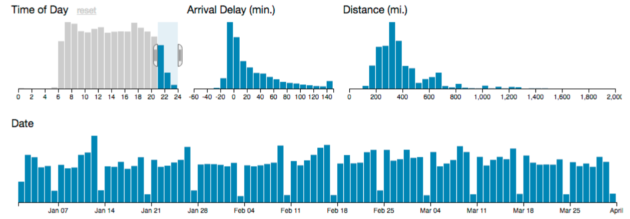
- eliminate some elements
  - either items or attributes
- according to what?
  - any possible function that partitions dataset into two sets
    - attribute values bigger/smaller than x
    - noise/signal
- filters vs queries
  - query: start with nothing, add in elements
  - filters: start with everything, remove elements
  - best approach depends on dataset size



Idiom: **cross filtering**

System: **Crossfilter**

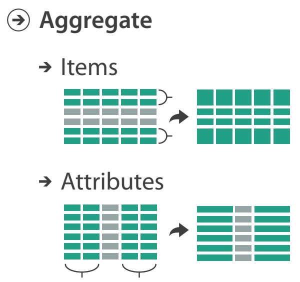
- item filtering
- coordinated views/controls combined
  - all scented histogram bisliders update when any ranges change



http://square.github.io/crossfilter/
https://observablehq.com/@uwdata/interaction

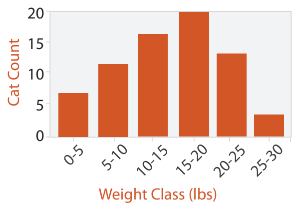
Aggregate

- a group of elements is represented by a smaller number of derived elements



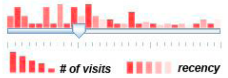
Idiom: **histogram**

- 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**

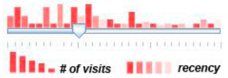
- augmented widgets show *information scent*
  - better cues for *information foraging*: show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider



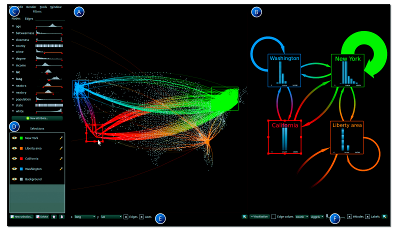
[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]

Idiom: **scented widgets**

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



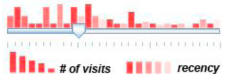
[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]



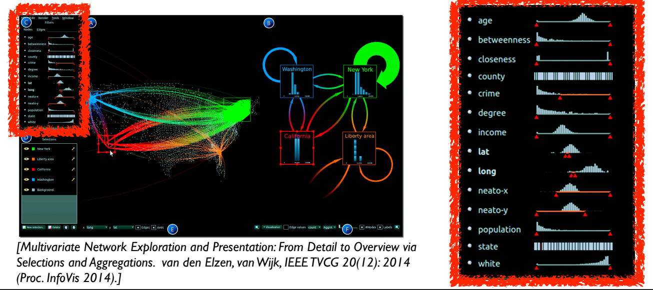
[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen, van Wijk, IEEE TVCG 20(12): 2014 (Proc. InfoVis 2014).]

Idiom: **scented widgets**

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

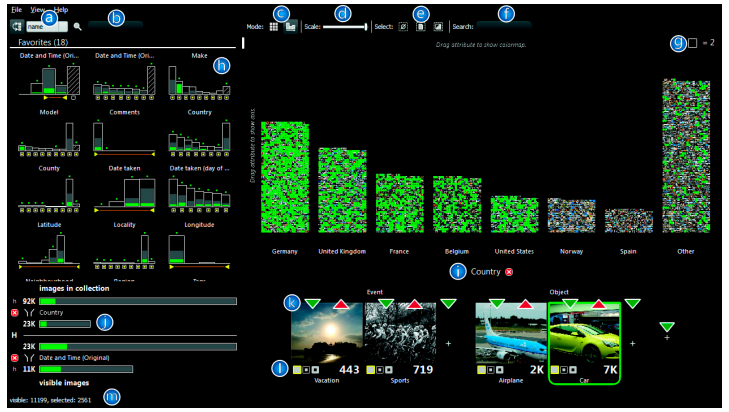


[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE TVCG (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]



[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen, van Wijk, IEEE TVCG 20(12): 2014 (Proc. InfoVis 2014).]

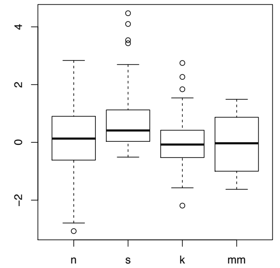
Scented histogram bisliders: detailed



[ICLIC: Interactive categorization of large image collections. van der Corput and van Wijk. Proc. PacificVis 2016.]

Idiom: **boxplot**

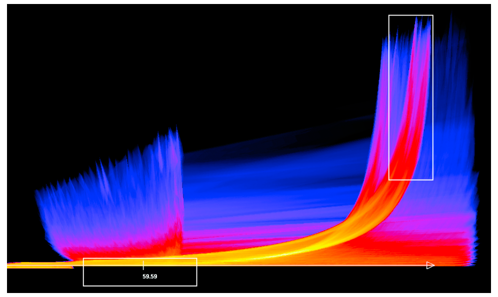
- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown
-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



[40 years of boxplots. Wickham and Stryjewski. 2012]

Idiom: **Continuous scatterplot**

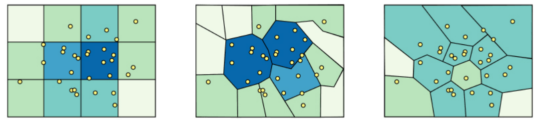
- static item aggregation
- data: table
- derived data: table
  - key attribs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color:
  sequential categorical hue + ordered luminance colormap
- scalability
  - no limits on overplotting: millions of items



[Continuous Scatterplots. Bachthaler and Weiskopf. IEEE TVCG (Proc. Vis 08) 14:6 (2008), 1428–1435. 2008.]

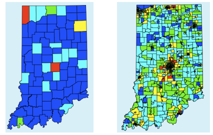
Spatial aggregation

- MAUP: Modifiable Areal Unit Problem
  - changing boundaries of cartographic regions can yield dramatically different results
  - zone effects



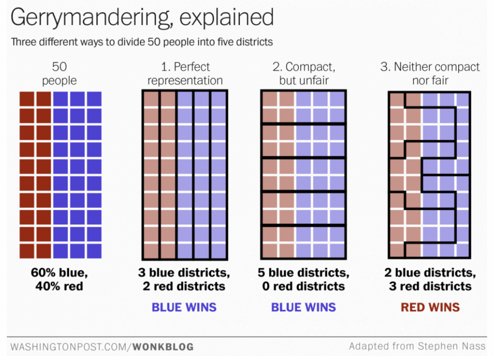
[http://www.e-education.psu/edu/geog486/l4_p7.html, Fig 4.cg.6]

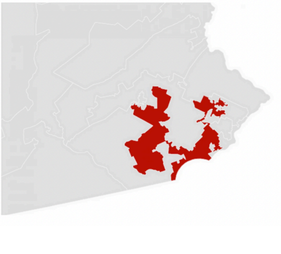
  - scale effects



https://blog.cartographica.com/blog/2011/5/19/the-modifiable-areal-unit-problem-in-gis.html

Gerrymandering: MAUP for political gain





A real district in Pennsylvania: Democrats won 51% of the vote but only 5 out of 18 house seats

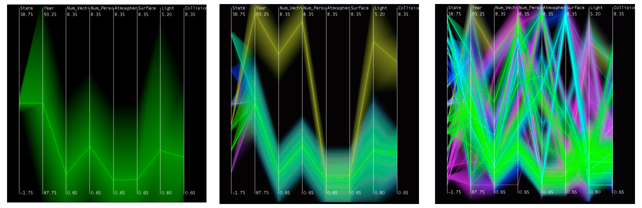
https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see/

Dynamic aggregation: Clustering

- clustering: classification of items into similar bins
  - based on similiarity measure
  - hierarchical algorithms produce "similarity tree": cluster hierarchy
    - agglomerative clustering: start w/ each node as own cluster, then iteratively merge
- cluster hierarchy: derived data used w/ many dynamic aggregation idioms
  - cluster more homogeneous than whole dataset
    - statistical measures & distribution more meaningful

Idiom: **Hierarchical parallel coordinates**

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



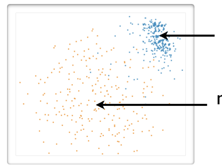
[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua, Ward, and Rundensteiner. Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

Dimensionality Reduction

Attribute aggregation: 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 → DR →



benign

malignant

data: 9D measured space

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

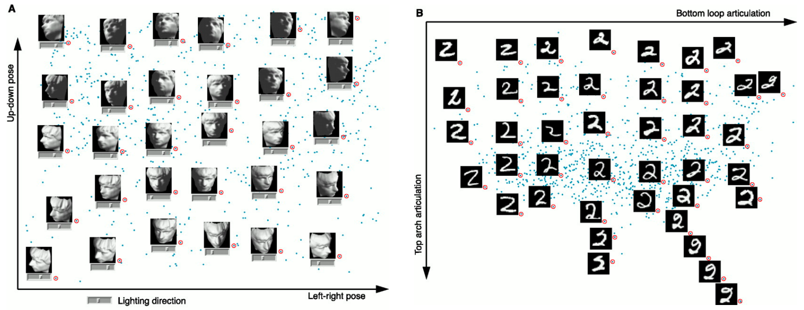
Dimensionality reduction & visualization

- 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

[Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences. Brehmer, Sedlmair, Ingram, and Munzner. Proc. BELIV 2014.]

Dimension-oriented tasks

- naming synthesized dims: inspect data represented by lowD points

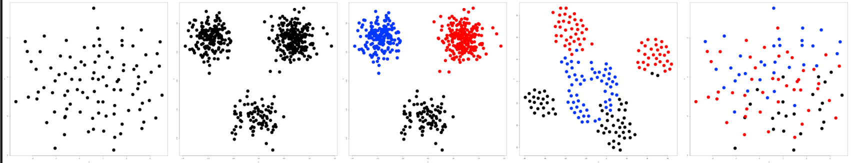


[A global geometric framework for nonlinear dimensionality reduction. Tenenbaum, de Silva, and Langford. Science, 290(5500):2319–2323, 2000.]

Cluster-oriented tasks

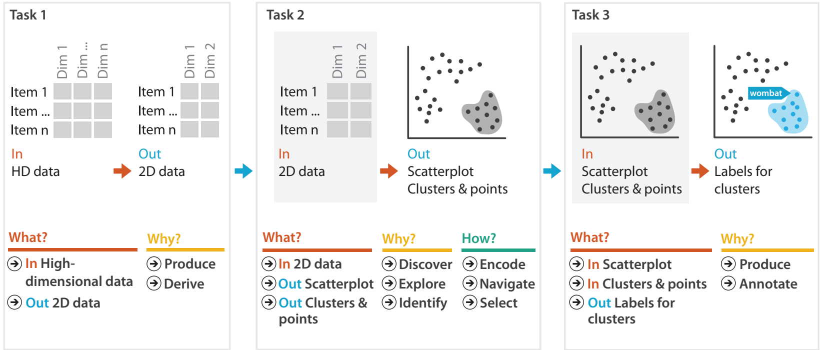
- verifying, naming, matching to classes

no discernable clusters | clearly discernable clusters | clear match cluster/class | partial match cluster/class | no match cluster/class



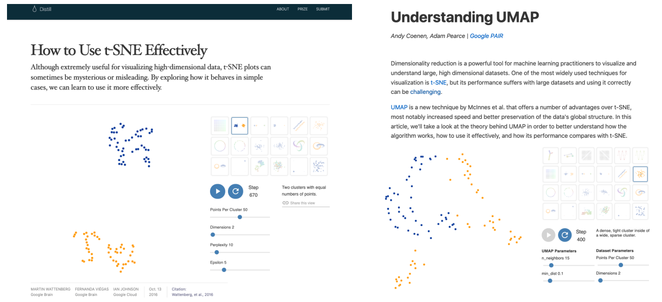
[Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences. Brehmer, Sedlmair, Ingram, and Munzner. Proc. BELIV 2014.]

Idiom: **Dimensionality reduction for documents**

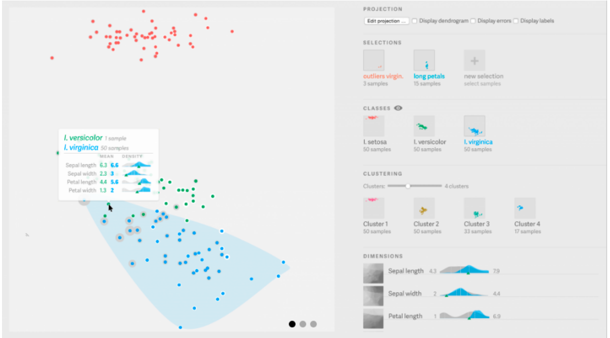


Latest algorithms: t-SNE, UMAP

- t-SNE https://distill.pub/2016/misread-tsne/
- UMAP https://pair-code.github.io/understanding-umap/



Interacting with dimensionally reduced data

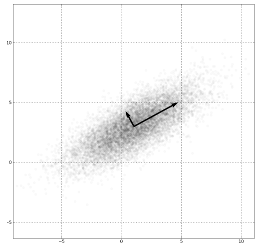


[https://uclab.fh-potsdam.de/projects/probing-projections/]

[Probing Projections: Interaction Techniques for Interpreting Arrangements and Errors of Dimensionality Reductions. Stahnke, Dörk, Müller, and Thom. IEEE TVCG (Proc. InfoVis 2015) 22(1):629-38 2016.]

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



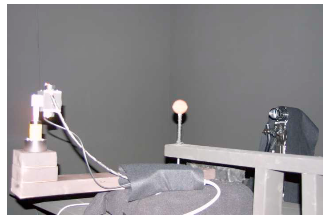
[http://en.wikipedia.org/wiki/File:GaussianScatterPCA.png]

Nonlinear dimensionality reduction

- pro: can handle curved rather than linear structure
- cons: lose all ties to original dims/attribs
  - 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

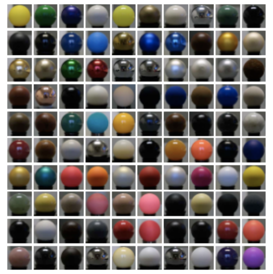
- DR for computer graphics reflectance model
  - goal: simulate how light bounces off materials to make realistic pictures
    - computer graphics: BRDF (reflectance)
  - idea: measure what light does with real materials



[Fig 2. Matusik, Pfister, Brand, and McMillan. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Capturing & using material reflectance

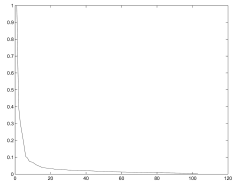
- reflectance measurement: interaction of light with real materials (spheres)
- result: 104 high-res images of material
  - each image 4M pixels
- goal: image synthesis
  - simulate completely new materials
- need for more concise model
  - 104 materials * 4M pixels = 400M dims
  - want concise model with meaningful knobs
    - how shiny/greasy/metallic
    - DR to the rescue!

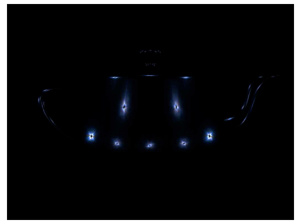


[Figs 5/6. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

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!

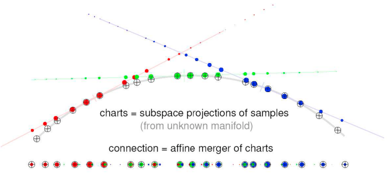


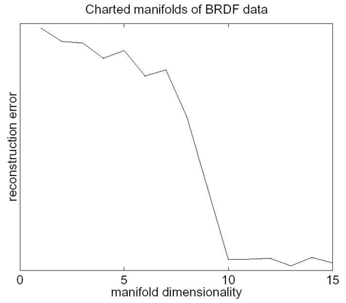


[Figs 6/7. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Nonlinear DR

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

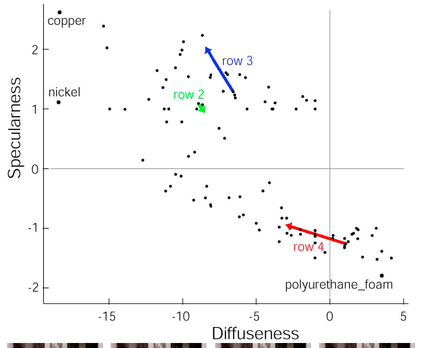




[Fig 10/11. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

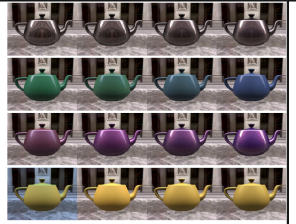
Finding semantics for synthetic dimensions

- 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

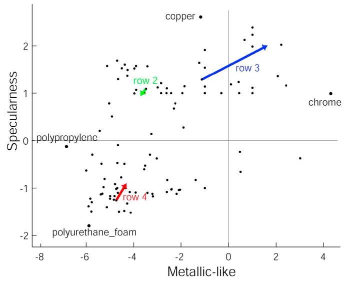


[Fig 12/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

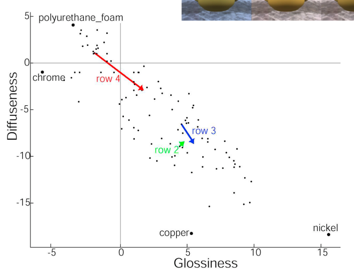
Understanding synthetic dimensions



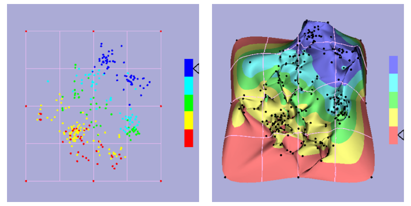
Specular-Metallic



Diffuseness-Glossiness



[Fig 13/14/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]



Spatialization Design

Comparing Points and Landscapes

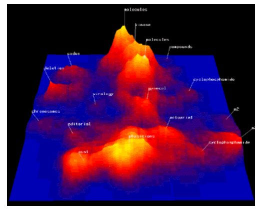
joint work with:
Melanie Tory, David W. Sprague, Fuqu Wu, Wing Yan So

http://webhome.cs.uvic.ca/~mtory/publications/infovis2007.pdf

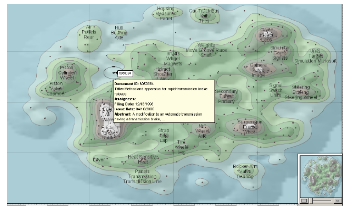
Spatialization Design: Comparing Points and Landscapes.
Tory, Sprague, Wu, So, and Munzner.
IEEE TVCG 13(6):1262–1269, 2007 (Proc. InfoVis 07).

Information Landscapes

- 2D or 3D landscape from set of DR points
  - height based on density
- oddly popular choice in DR
  - despite known occlusion/distortion problems with 3D
  - assertions: pattern recognition, spatial reasoning, familiar



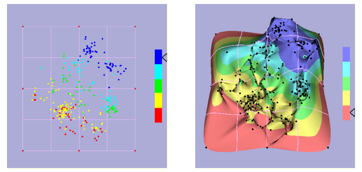
Themescape:
[http://www.k-n-o-r-z.de/publ/example/retriev1.htm]



[Guide to MicroPatent Aureka 9 ThemeScape]

Understanding User Task

- abstract: search involving spatial areas and estimation
  Estimate which grid cell has the most points of the target color



- domain-specific examples
  "Where in the display are people with high incomes?"
  "Does this area also have high education levels?"
  "Does this area correspond to a particular work sector?"
- non-trivial complexity yet fast response time
- frequent subtask in pilot test of real data analysis

Lab Study: Test Human Response Time and Error

- hypotheses
  - points are better than landscapes
    - result: yes!
    - much better: 2-4 × faster, 5-14 × more accurate
  - 2D landscapes (color only) better than 3D landscapes (color + height redundantly encoded)
    - result: yes
    - significantly faster, no significant difference in accuracy