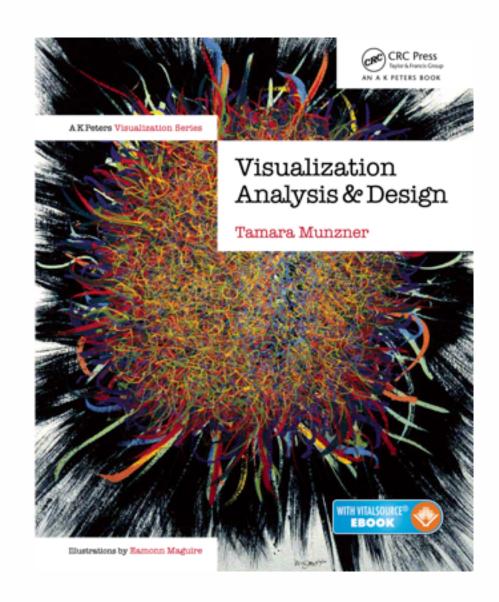
Visualization Analysis & Design

Tamara Munzner

Department of Computer Science University of British Columbia

UBC CS Faculty Lecture Series December 3 2015, Vancouver BC



http://www.cs.ubc.ca/~tmm/talks.html#vad15fls

<u>@tamaramunzner</u>

Defining visualization (vis)

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why?...

Why have a human in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

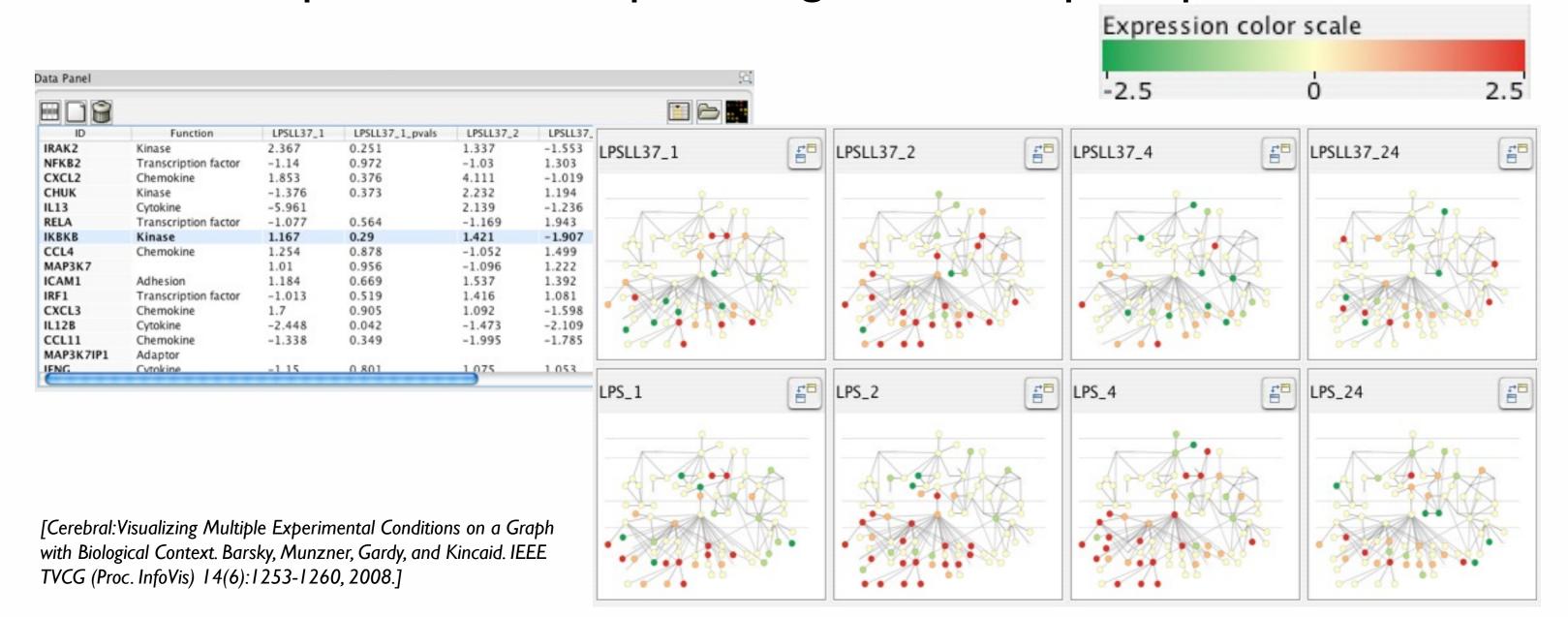
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- don't need vis when fully automatic solution exists and is trusted
- many analysis problems ill-specified
 - -don't know exactly what questions to ask in advance
- possibilities
 - -long-term use for end users (e.g. exploratory analysis of scientific data)
 - -presentation of known results
 - stepping stone to better understanding of requirements before developing models
 - help developers of automatic solution refine/debug, determine parameters
 - -help end users of automatic solutions verify, build trust

Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

external representation: replace cognition with perception



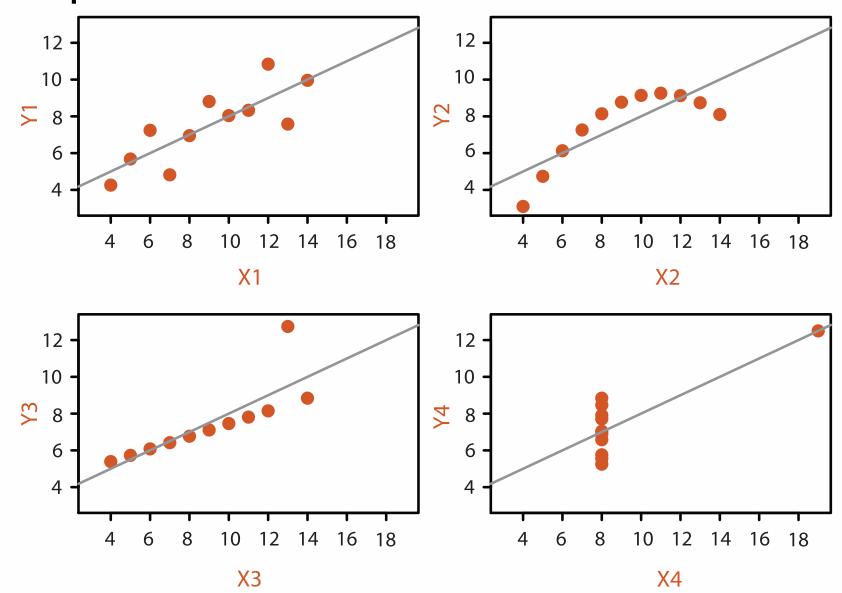
Why represent all the data?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, details matter
 - -confirm expected and find unexpected patterns
 - -assess validity of statistical model

Anscombe's Quartet

Identical statistics		
x mean	9	
x variance	10	
y mean	8	
y variance	4	
x/y correlation	1	

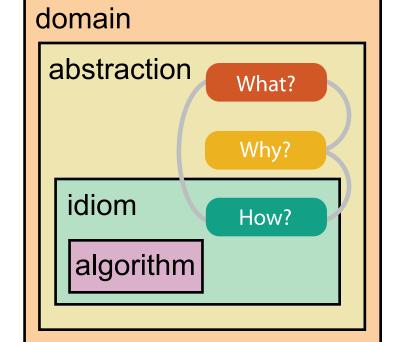


Analysis framework: Four levels, three questions

- domain situation
 - who are the target users?
- abstraction
 - translate from specifics of domain to vocabulary of vis

[A Nested Model of Visualization Design and Validation. *Munzner. IEEETVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).*]

- what is shown? data abstraction
 - often don't just draw what you're given: transform to new form
- why is the user looking at it? task abstraction
- idiom
 - how is it shown?
 - visual encoding idiom: how to draw
 - interaction idiom: how to manipulate
- algorithm
 - efficient computation



domain

abstraction

algorithm

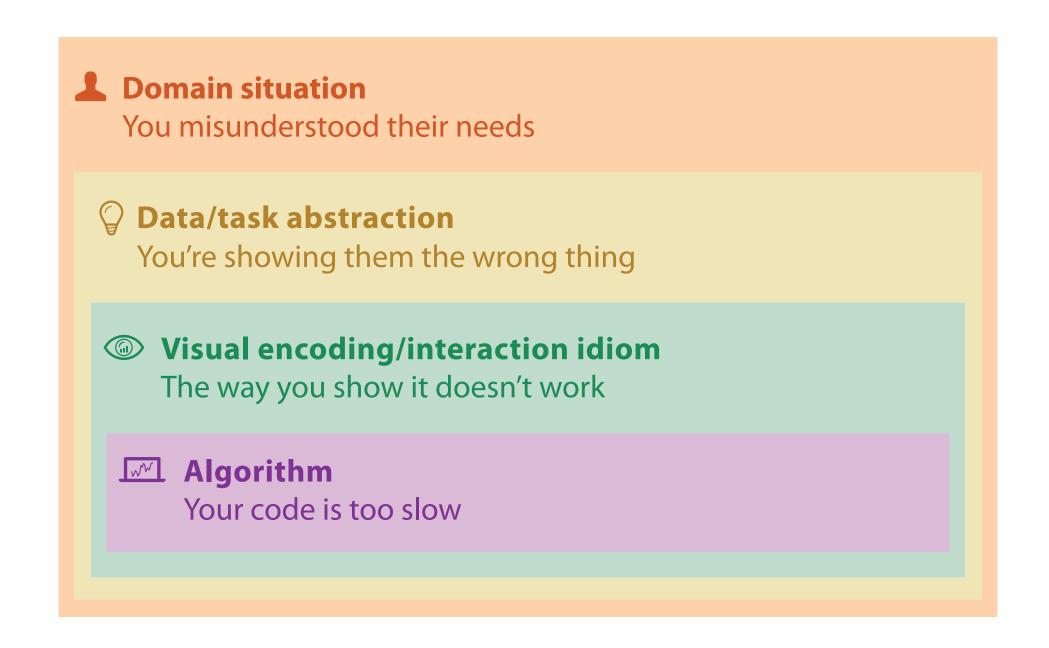
idiom

[A Multi-Level Typology of Abstract Visualization Tasks

Brehmer and Munzner. IEEETVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Why is validation difficult?

different ways to get it wrong at each level



Why is validation difficult?

solution: use methods from different fields at each level

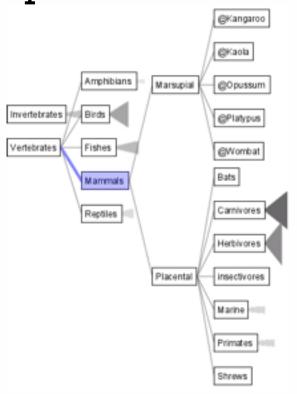
Domain situation anthropology/ Observe target users using existing tools ethnography **Data/task abstraction** Wisual encoding/interaction idiom design Justify design with respect to alternatives **Algorithm** computer Measure system time/memory science Analyze computational complexity cognitive Analyze results qualitatively psychology Measure human time with lab experiment (*lab study*) Observe target users after deployment (*field study*) anthropology/ ethnography Measure adoption

technique-driven work

Why analyze?

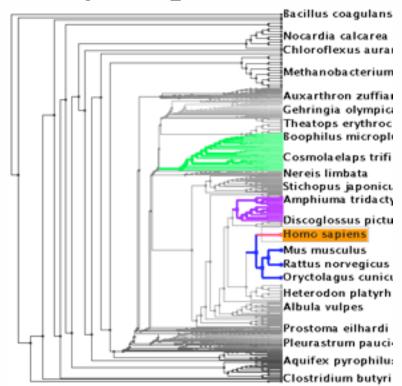
- imposes a structure on huge design space
 - -scaffold to help you think systematically about choices
 - -analyzing existing as stepping stone to designing new

SpaceTree



[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57-64.]

TreeJuxtaposer



[Tree]uxtaposer: Scalable Tree Comparison Using Focus +Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453-462, 2003.]

What?

Tree



Why?

- Actions
 - → Present → Locate → Identify







- **→** Targets
 - → Path between two nodes



How?

→ SpaceTree

→ Encode → Navigate → Select → Filter













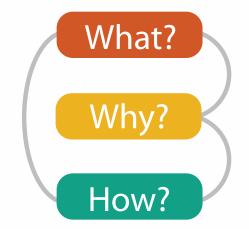
→ Encode → Navigate → Select → Arrange











What? Why? How?



Datasets

Attributes

→ Data Types

→ Items ·

→ Attributes → Links

→ Positions

→ Grids

Data and Dataset Types



→ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative

Dataset Types

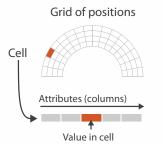
→ Tables



→ Networks

→ Trees





Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



→ Multidimensional Table

Cell containing value



→ Geometry (Spatial)

Key 2

Attributes



→ Dataset Availability



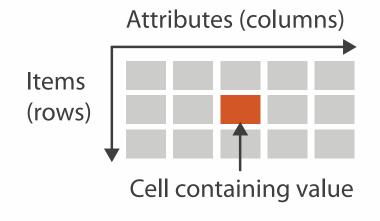


→ Dynamic

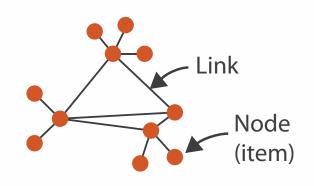


Types: Datasets and data

- Dataset Types
 - → Tables
 - Tables



→ Networks



→ Spatial
 → Fields (Continuous)
 → Geometry (Spatial)
 Cell
 Attributes (columns)

- Attribute Types
 - → Categorical









- → Ordered
 - → Ordinal
 - ***
- → Quantitative



Value in cell

What? Why? How?

• {action, target} pairs

- discover distribution
- compare trends
- locate outliers
- browse topology

Why?

- Analyze
 - → Consume







- → Produce
 - → Annotate







Search

	Target known	Target unknown
Location known	·.··· Lookup	*. Browse
Location unknown	₹ • Locate	< ∙ Explore

- Query
 - → Identify













All Data







Attributes



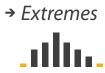




→ Many







- **Network Data**
 - → Topology







→ Paths



- **Spatial Data**
 - → Shape





Actions I:Analyze

- consume
 - -discover vs present
 - classic split
 - aka explore vs explain
 - -enjoy
 - newcomer
 - aka casual, social
- produce
 - -annotate, record
 - -derive
 - crucial design choice



→ Consume







- → Produce
 - → Annotate



→ Record

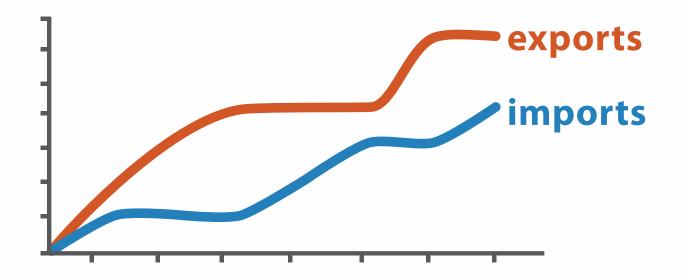


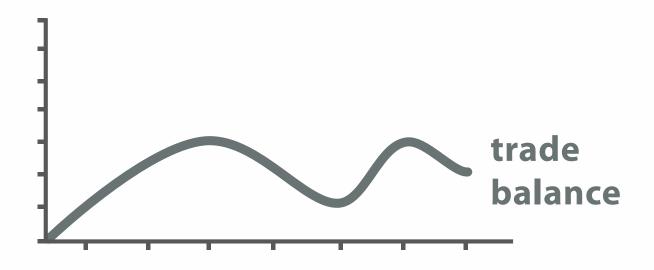
→ Derive



Derive

- don't just draw what you're given!
 - -decide what the right thing to show is
 - -create it with a series of transformations from the original dataset
 - -draw that
- one of the four major strategies for handling complexity





 $trade\ balance = exports - imports$

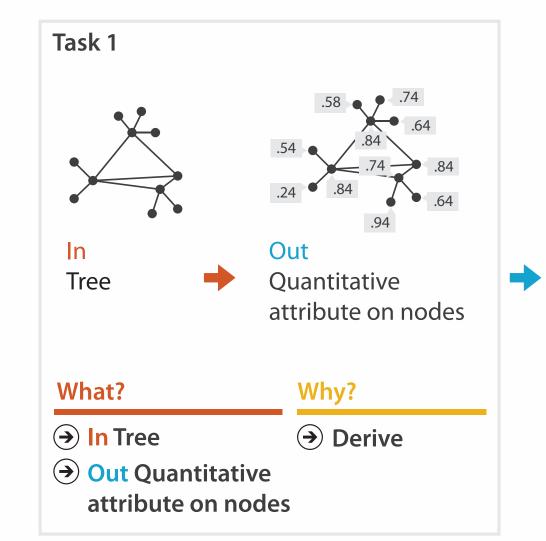
Derived Data

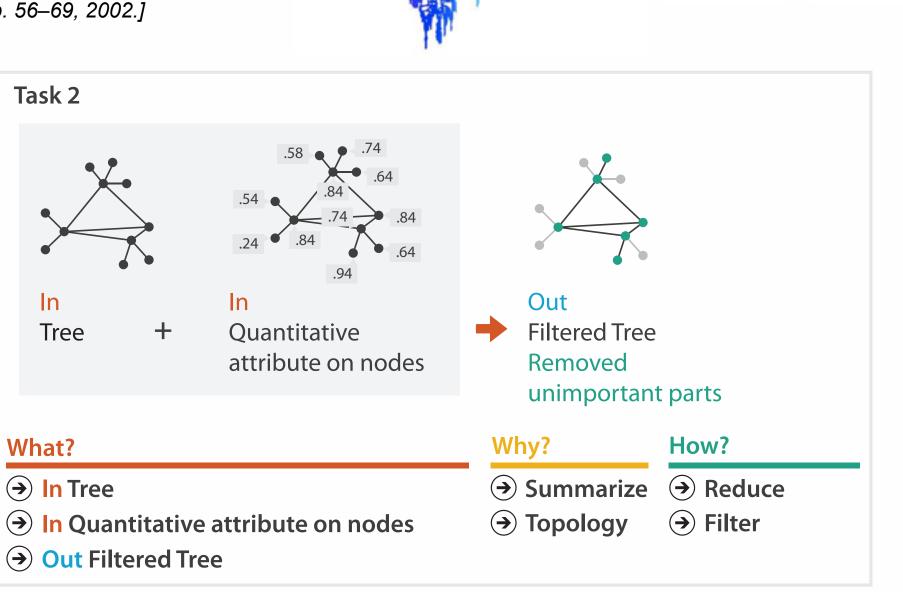
Analysis example: Derive one attribute

- Strahler number
 - centrality metric for trees/networks
 - derived quantitative attribute
 - draw top 5K of 500K for good skeleton

[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]







Actions II: Search

- what does user know?
 - -target, location

→ Search

	Target known	Target unknown
Location known	• • • Lookup	Browse
Location unknown	Cipi. Locate	< Explore

Actions III: Query

- what does user know?
- → Search

- -target, location
- how much of the data matters?

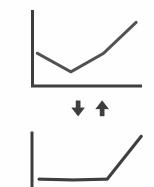
• analyze, search, query

- one, some, all

	Target known	Target unknown
Location known	• • • Lookup	• • • Browse
Location unknown	C Locate	< Explore

- () Query
- -independent choices for each
- → Identify

→ Compare



→ Summarize



Targets

- **All Data**
 - → Trends
- → Outliers
- → Features





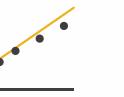
- **Attributes**
 - → One

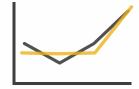
- → Many
- → Distribution

 - → Extremes

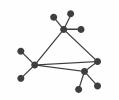


- → Dependency → Correlation → Similarity

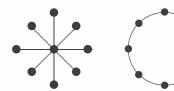




- **Network Data**
 - → Topology



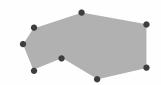




→ Paths



- **Spatial Data**
 - → Shape



Encode



→ Express



→ Separate

→ Order







→ Use



How?

Map

from categorical and ordered attributes

→ Color



→ Size, Angle, Curvature, ...













→ Motion Direction, Rate, Frequency, ...



Manipulate

Facet

Reduce

→ Change







→ Filter



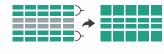
→ Select



→ Partition



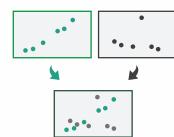
Aggregate



→ Navigate



→ Superimpose

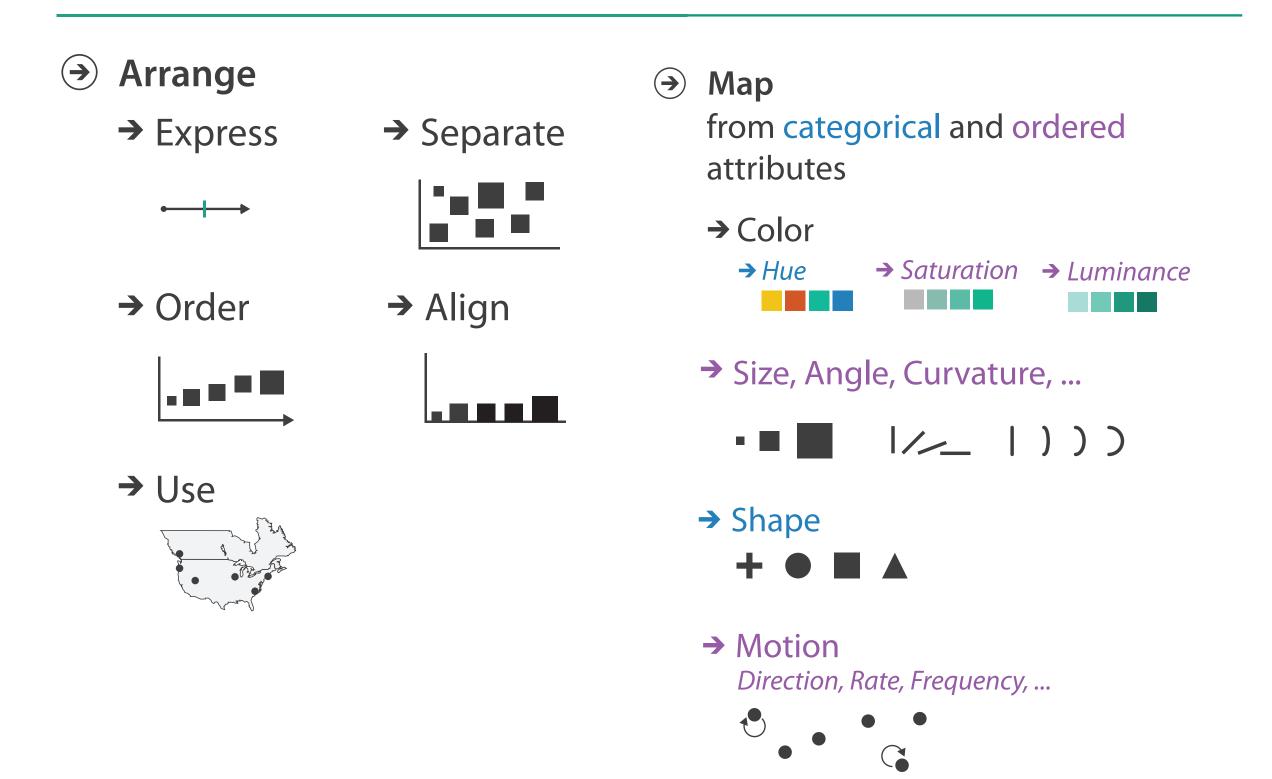


→ Embed



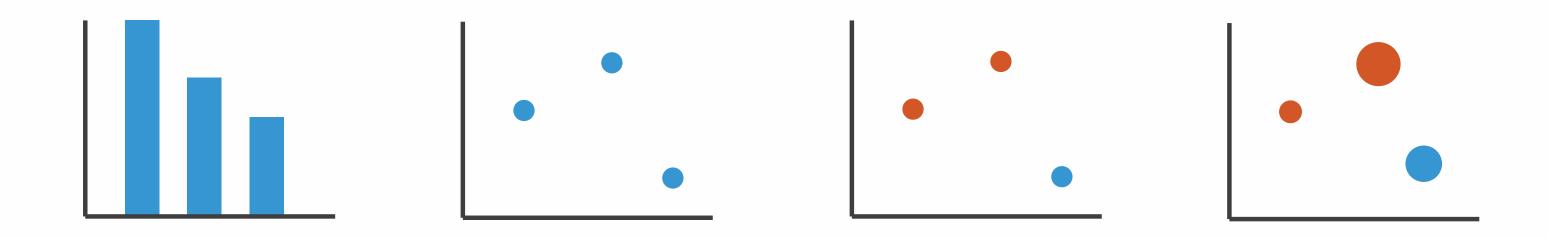
How to encode: Arrange space, map channels

Encode



Encoding visually

• analyze idiom structure



Definitions: Marks and channels

• marks

channels

-geometric primitives















- -control appearance of marks













<a>Shape









Color



Size





→ Volume





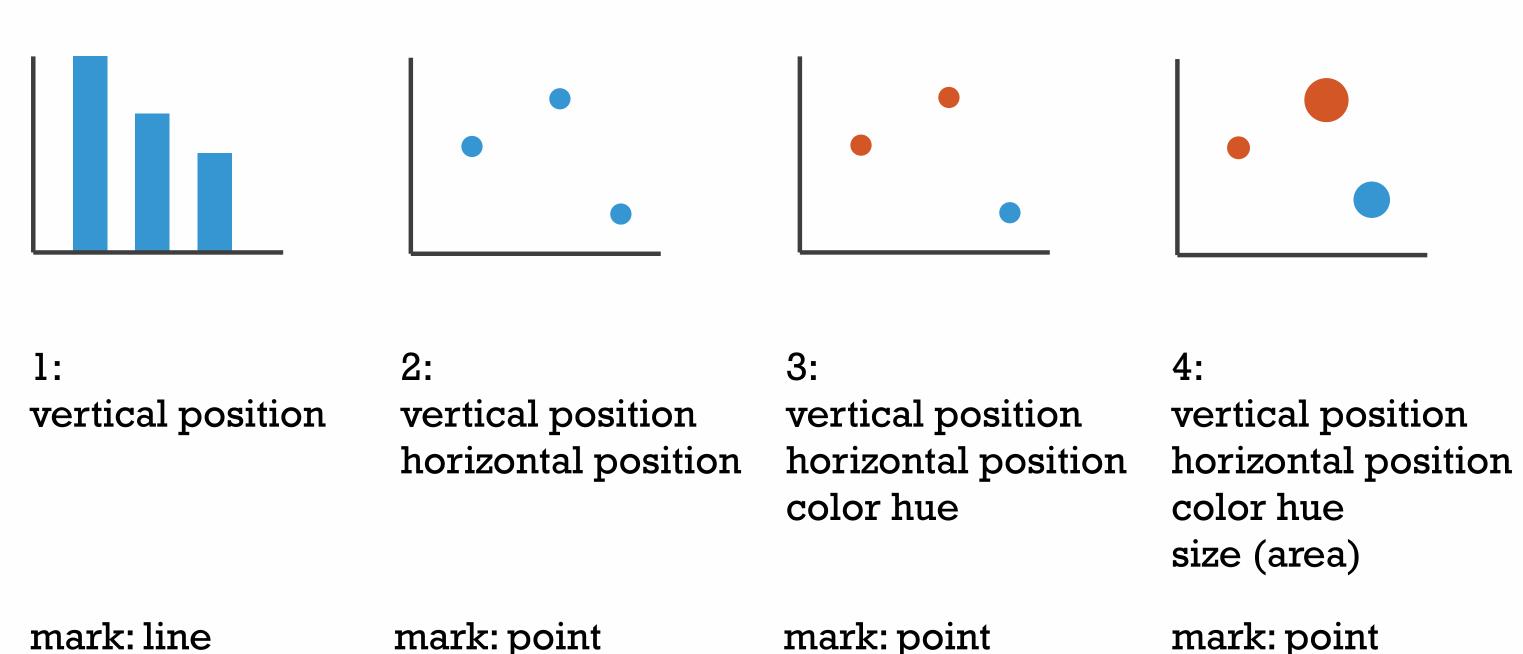






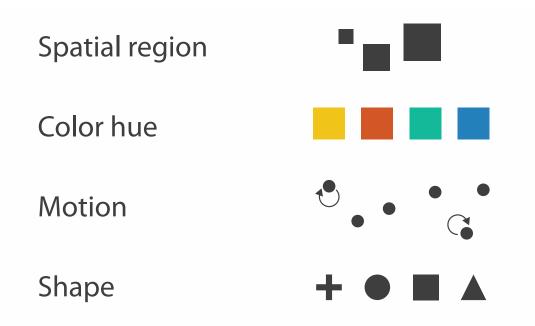
Encoding visually with marks and channels

- analyze idiom structure
 - -as combination of marks and channels



Channels

Position on common scale Position on unaligned scale Length (1D size) Tilt/angle Area (2D size) Depth (3D position) Color luminance Color saturation Curvature Volume (3D size)



Channels: Matching Types

Magnitude Channels: Ordered Attributes Position on common scale Position on unaligned scale Length (1D size) Tilt/angle Area (2D size) Depth (3D position) Color luminance Color saturation Curvature

Volume (3D size)

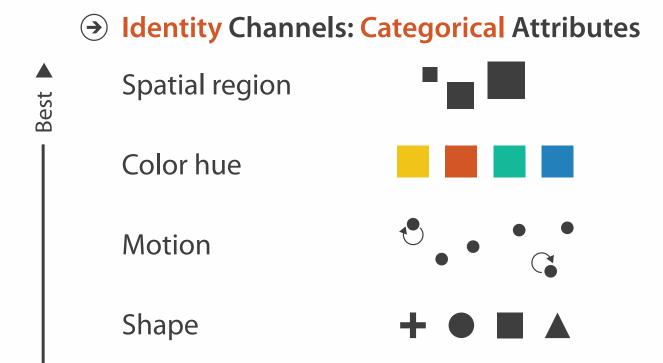
→ Identity Channels: Categorical Attributes



- expressiveness principle
 - match channel and data characteristics

Channels: Rankings

Magnitude Channels: Ordered Attributes Position on common scale Position on unaligned scale Length (1D size) Tilt/angle Area (2D size) Depth (3D position) Color luminance Color saturation Curvature Volume (3D size)



- expressiveness principle
 - -match channel and data characteristics
- effectiveness principle
 - encode most important attributes with highest ranked channels

more on channels: slides 30-38 of http://www.cs.ubc.ca/~tmm/talks.html#halfdaycourse15

Encode



→ Express

→ Separate





→ Order

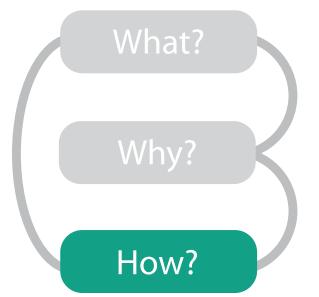






→ Use





→ Map

from categorical and ordered attributes

→ Color



→ Size, Angle, Curvature, ...



→ Shape



→ Motion

Direction, Rate, Frequency, ...



Manipulate

Facet

Reduce

→ Change



Juxtapose



→ Filter



→ Select



→ Partition



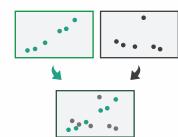
Aggregate



→ Navigate



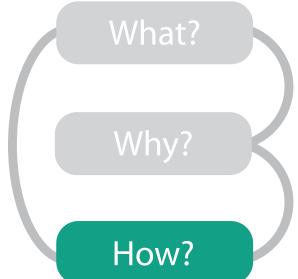
Superimpose



→ Embed



Encode Manipulate **Facet** Reduce → Arrange Map Juxtapose **→** Change **→** Filter from categorical and ordered → Express Separate attributes → Color → Hue **→** Select → Saturation → Luminance **→** Partition Aggregate → Order → Align e, Curvature, ... → Use **→** Navigate **→** Superimpose **→** Embed → Shape → Motion Direction, Rate, Frequency, ...



much more on visual encoding: slides 39-91 of http://www.cs.ubc.ca/~tmm/talks.html#halfdaycourse15

Encode

- Arrange
 - → Express
- → Separate





- → Order
- → Align





→ Use



Map

from categorical and ordered attributes

→ Color



→ Size, Angle, Curvature, ...



→ Shape



→ Motion Direction, Rate, Frequency, ...



Manipulate

Facet

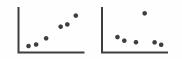
To Take The State of the Antique of the State of the Stat

Reduce

→ Change



→ Juxtapose



→ Filter



→ Select



→ Partition



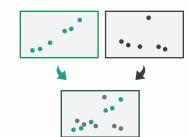
Aggregate



→ Navigate



→ Superimpose



→ Embed



How?

How to handle complexity: 3 more strategies + 1 previous

Manipulate

Facet

Reduce

→ Derive

Change







→ Filter





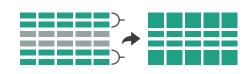
Select



Partition



Aggregate



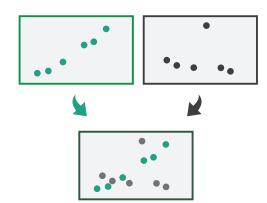
change view over time

 facet across multiple views

Navigate



Superimpose



→ Embed



- reduce items/attributes
 within single view
- derive new data to show within view

How to handle complexity: 3 more strategies

+ I previous

Manipulate

ANATON STORESTORE OF THE PROPERTY OF

Facet

Reduce

→ Derive

Change



Juxtapose



→ Filter



→

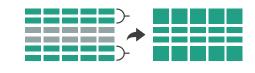
→ Select



Partition



Aggregate

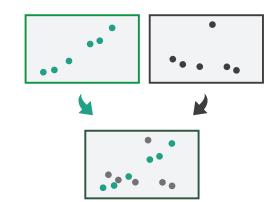


- change over time
 - most obvious & flexible of the 4 strategies

Navigate



Superimpose

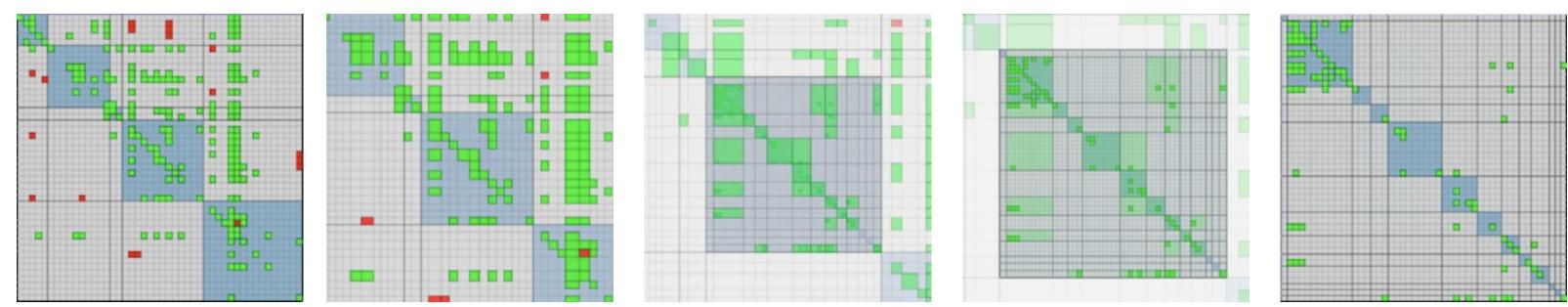


→ Embed



Idiom: Animated transitions

- smooth transition from one state to another
 - -alternative to jump cuts
 - -support for item tracking when amount of change is limited
- example: multilevel matrix views
 - -scope of what is shown narrows down
 - middle block stretches to fill space, additional structure appears within
 - other blocks squish down to increasingly aggregated representations



[Using Multilevel Call Matrices in Large Software Projects. van Ham. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 227–232, 2003.]

How to handle complexity: 3 more strategies

+ I previous

Manipulate

→ Change



→ Select



Navigate



Facet

Juxtapose



Reduce





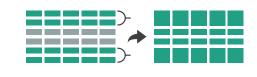
→ Derive



Partition

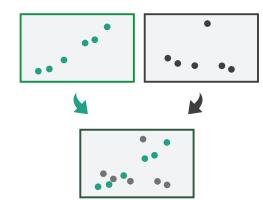


Aggregate



 facet data across multiple views



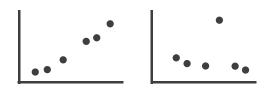


→ Embed

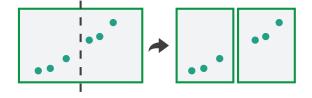


Facet

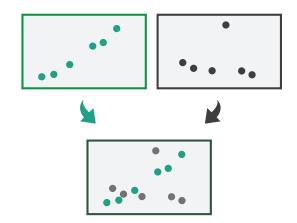
Juxtapose



Partition

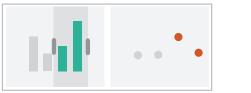


Superimpose



→ Coordinate Multiple Side By Side Views

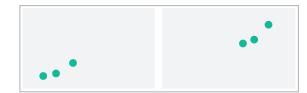
- → Share Encoding: Same/Different
 - → Linked Highlighting





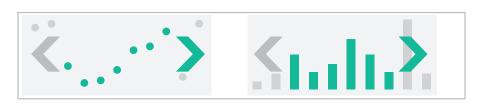
→ Share Data: All/Subset/None







→ Share Navigation

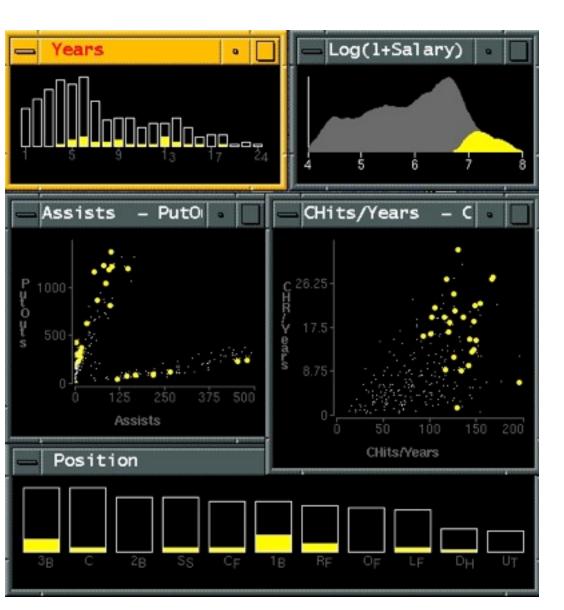


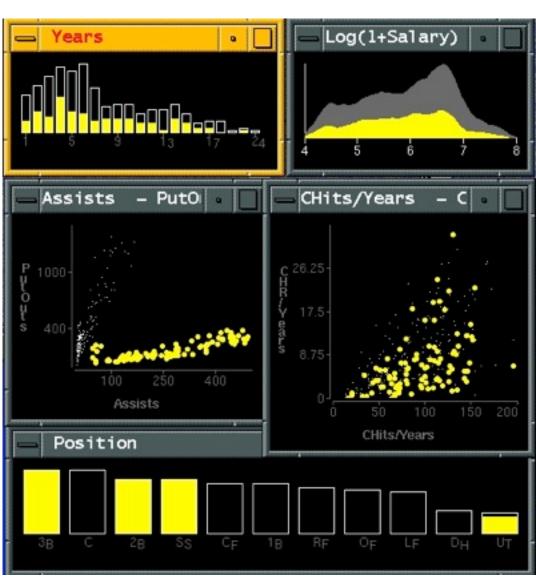
ldiom: Linked highlighting

System: **EDV**

- see how regions contiguous in one view are distributed within another
 - powerful and pervasive interaction idiom

- encoding: different
 - multiform
- data: all shared





[Visual Exploration of Large Structured Datasets.Wills. Proc. New Techniques and Trends in Statistics (NTTS), pp. 237–246. IOS Press, 1995.]

ldiom: bird's-eye maps

System: Google Maps

- encoding: same
- data: subset shared
- navigation: shared
 - -bidirectional linking
- differences
 - -viewpoint
 - -(size)
- overview-detail

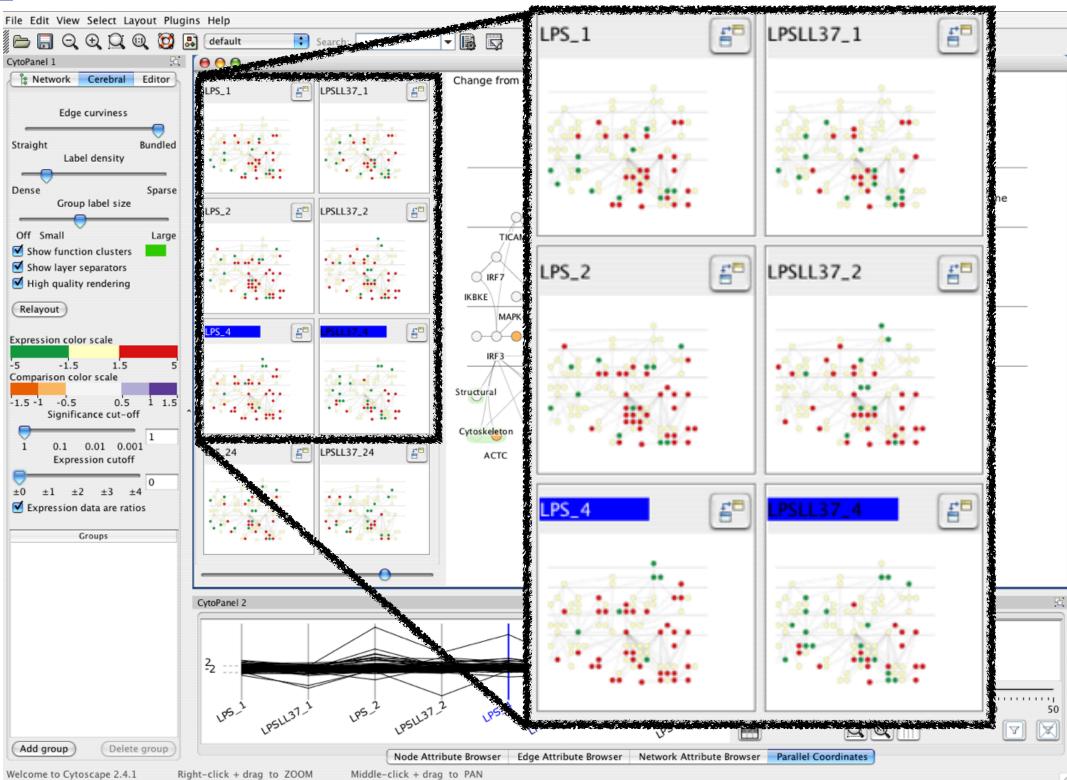


[A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.]

Idiom: Small multiples

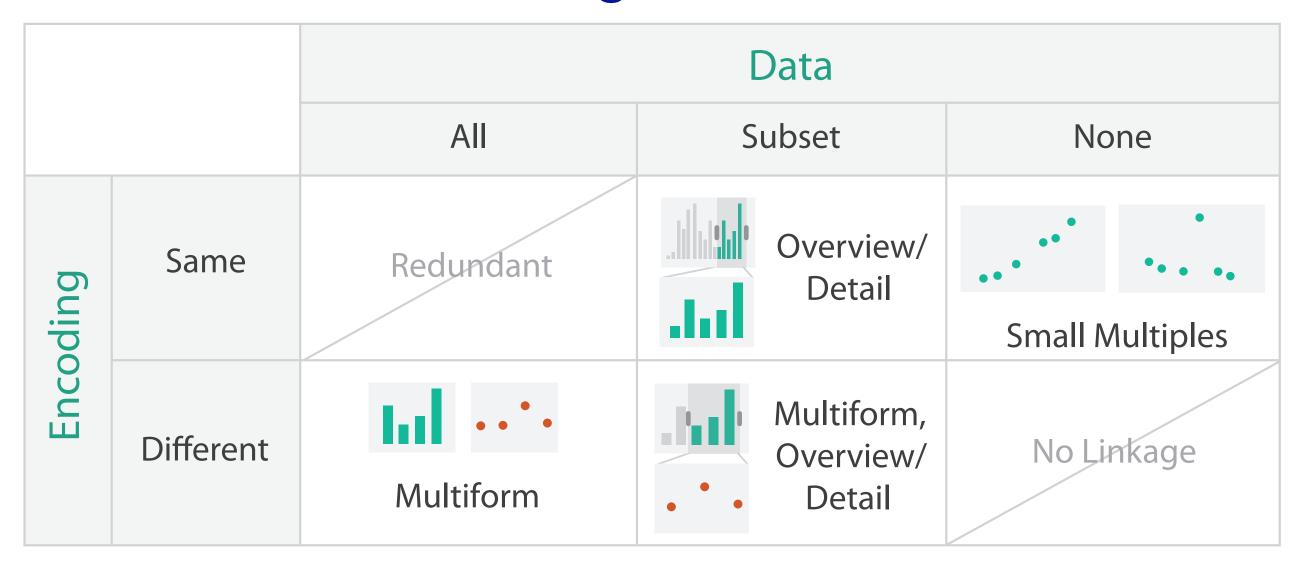
System: Cerebral

- encoding: same
- data: none shared
 - different attributes for node colors
 - -(same network layout)
- navigation: shared



[Cerebral:Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14:6 (2008), 1253–1260.]

Coordinate views: Design choice interaction



- why juxtapose views?
 - -benefits: eyes vs memory
 - lower cognitive load to move eyes between 2 views than remembering previous state with single changing view
 - -costs: display area, 2 views side by side each have only half the area of one view

Partition into views

- how to divide data between views
 - encodes association between items using spatial proximity
 - -major implications for what patterns are visible
 - -split according to attributes
- design choices
 - how many splits
 - all the way down: one mark per region?
 - stop earlier, for more complex structure within region?
 - order in which attribs used to split
 - -how many views

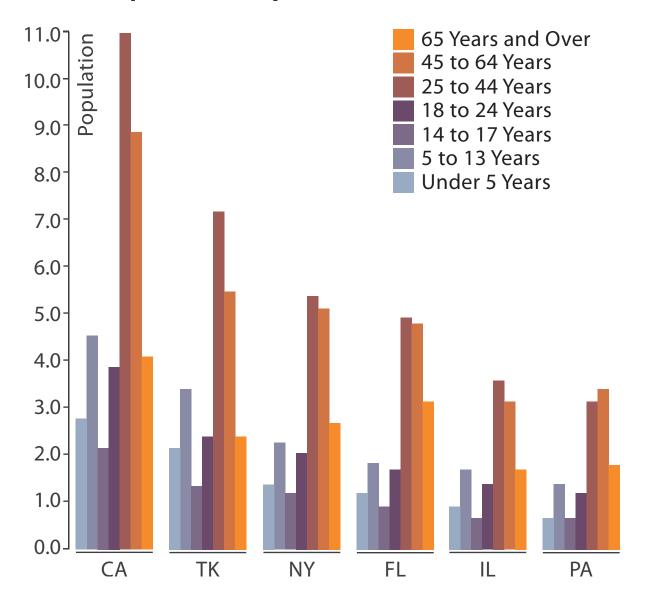






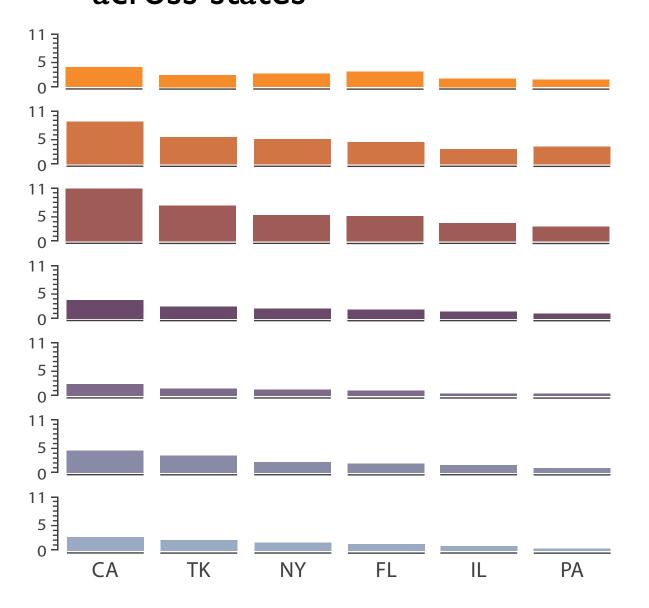
Partitioning: List alignment

- 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



http://bl.ocks.org/mbostock/3887051 http://bl.ocks.org/mbostock/4679202

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

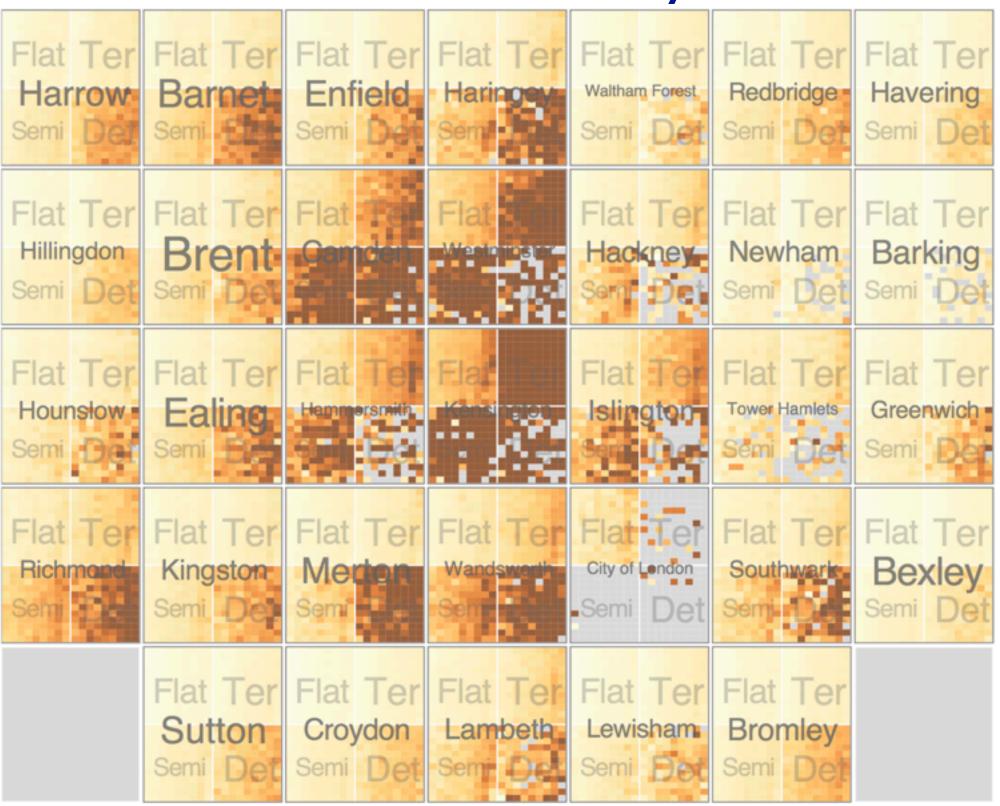


Partitioning: Recursive subdivision

System: **HIVE**

- split by neighborhood
- then by type
- then time
 - -years as rows
 - -months as columns
- color by price

- neighborhood patterns
 - -where it's expensive
 - where you pay much more for detached type



[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

Partitioning: Recursive subdivision

- switch order of splits
 - -type then neighborhood
- switch color
 - -by price variation

- type patterns
 - within specific type, which neighborhoods inconsistent

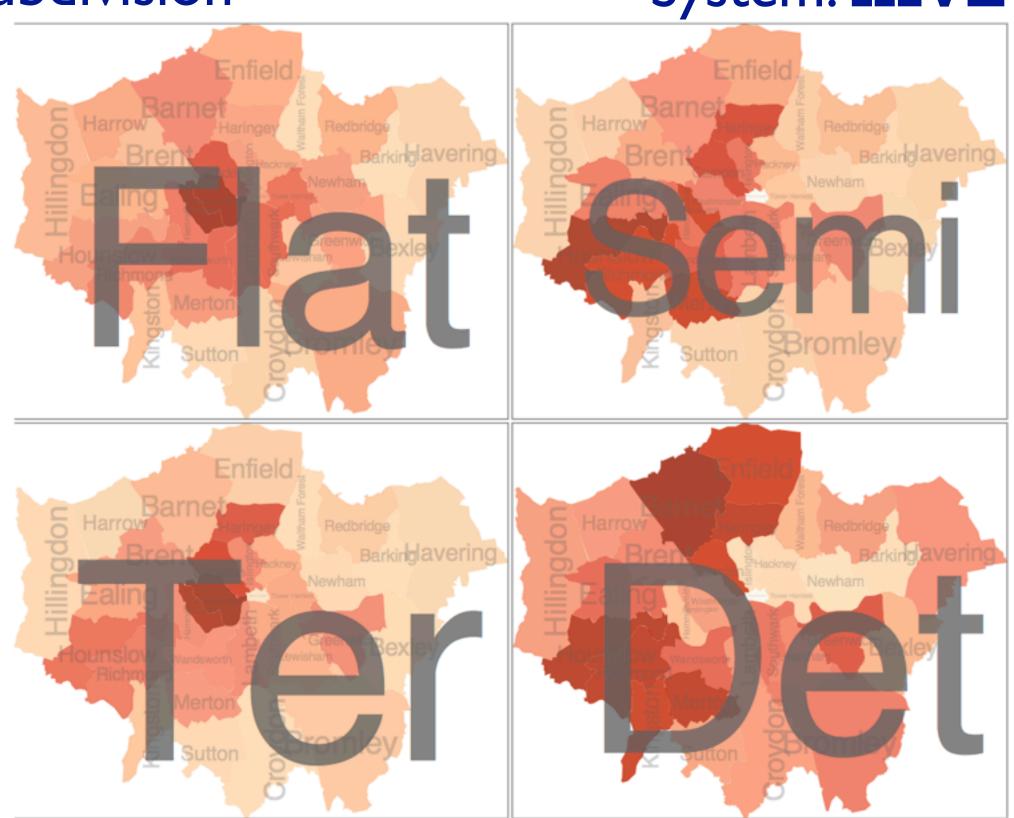


System: **HIVE**

Partitioning: Recursive subdivision

System: **HIVE**

- different encoding for second-level regions
 - -choropleth maps



How to handle complexity: 3 more strategies

+ I previous

Manipulate

Facet

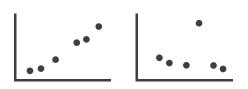
Reduce

→ Derive

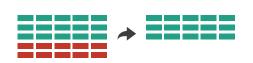














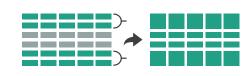
→ Select



Partition



₽ Aggregate

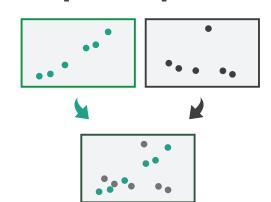


reduce what is shown within single view

Navigate



Superimpose



→ Embed



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, facet, change, derive

Reducing Items and Attributes



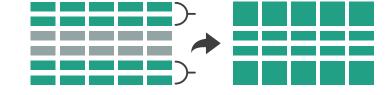


→ Attributes

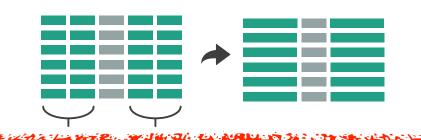


Aggregate

→ Items



→ Attributes



Reduce

→ Filter



Aggregate

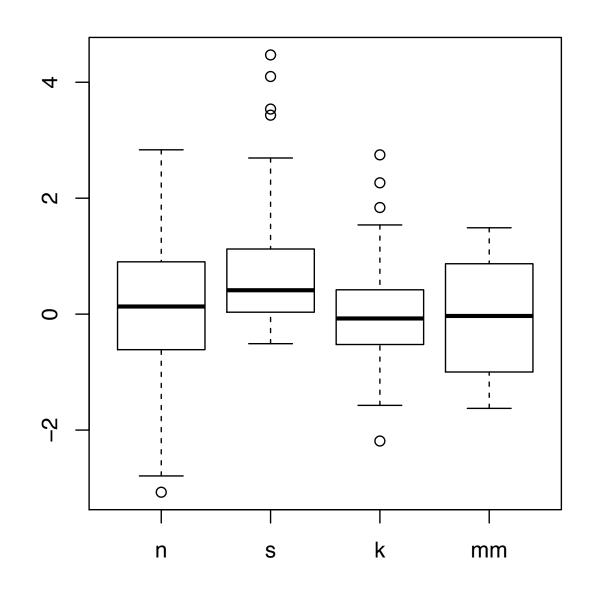


Embed



Idiom: **boxplot**

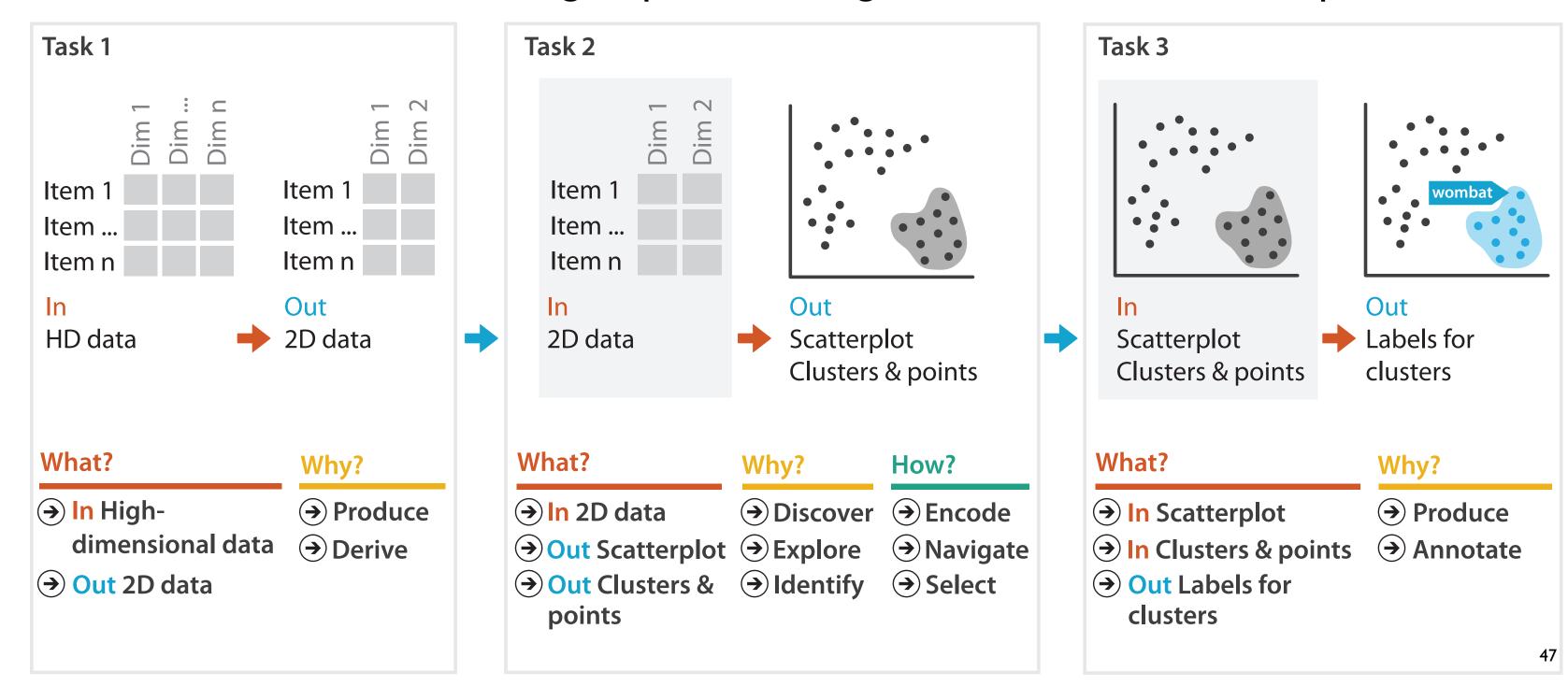
- static item aggregation
- task: find distribution
- data: table
- derived data
 - -5 quant attribs
 - median: central line
 - lower and upper quartile: boxes
 - lower upper fences: whiskers
 - values beyond which items are outliers
 - outliers beyond fence cutoffs explicitly shown

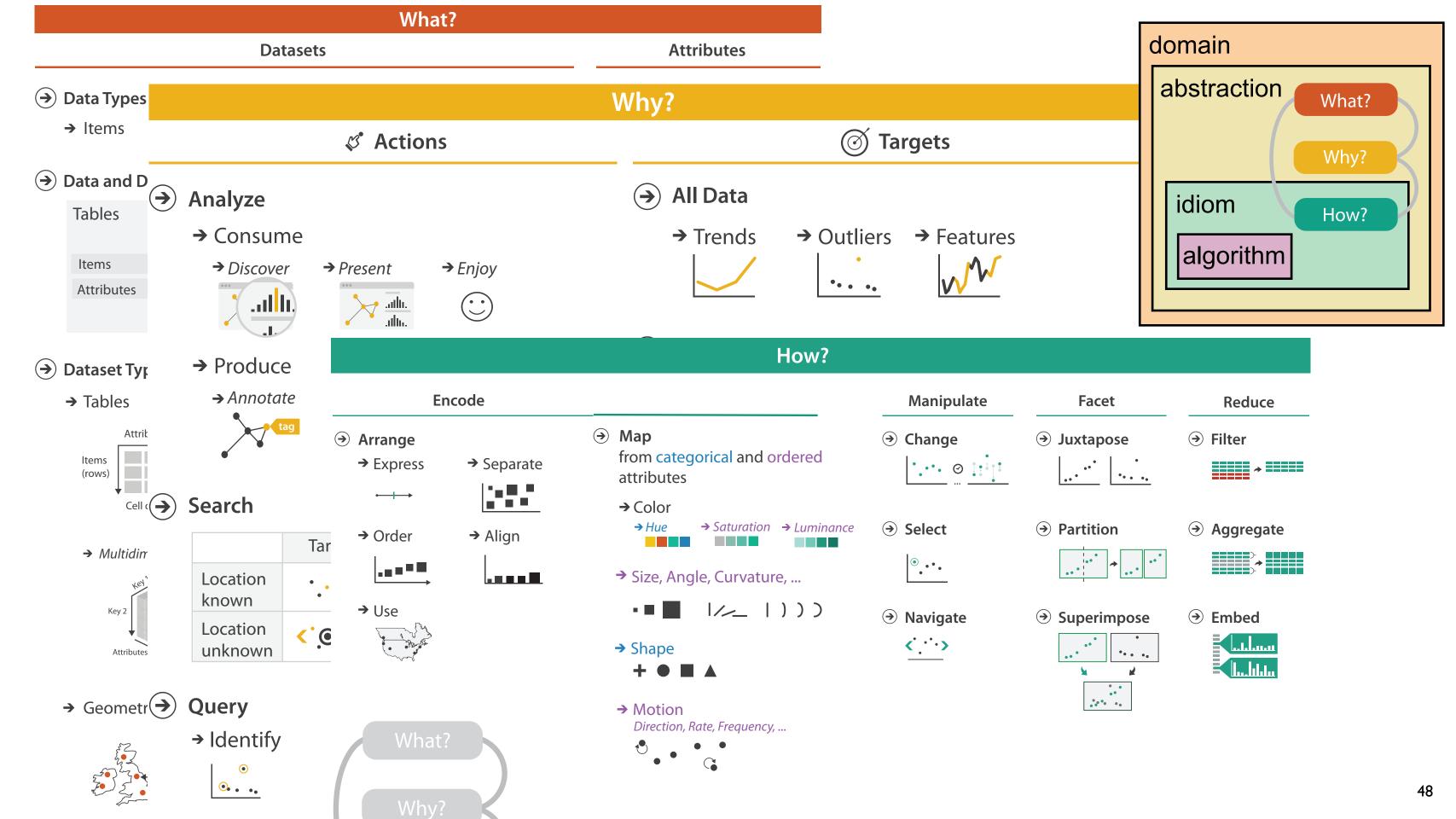


[40 years of boxplots.Wickham and Stryjewski. 2012. had.co.nz]

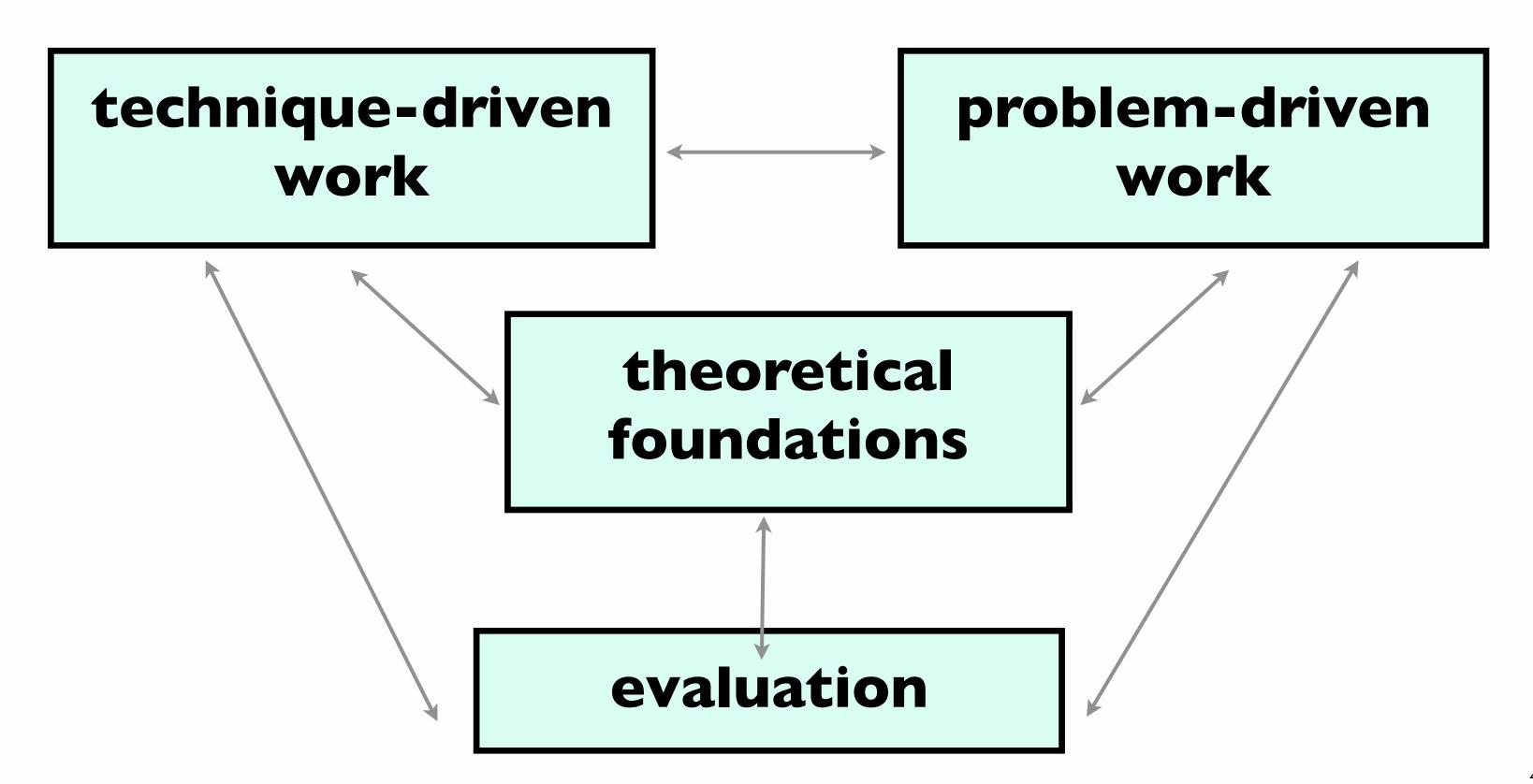
Idiom: Dimensionality reduction for documents

- attribute aggregation
 - derive low-dimensional target space from high-dimensional measured space

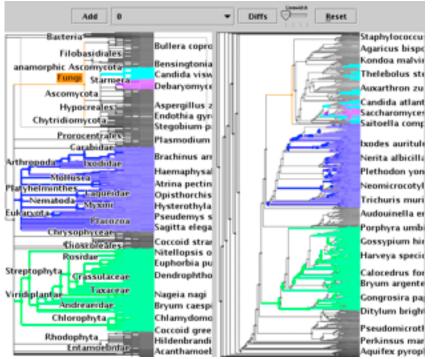




A quick taste of my own work!



Technique-driven: Graph drawing







Kristian Hildebrand



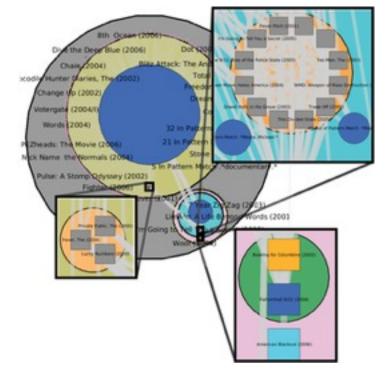
TreeJuxtaposer

Daniel Archambault



David Auber (Bordeaux)





TopoLayout
SPF
Grouse
GrouseFlocks
TugGraph



F

Evaluation: Graph drawing

















Joanna McGrenere

(UBC)

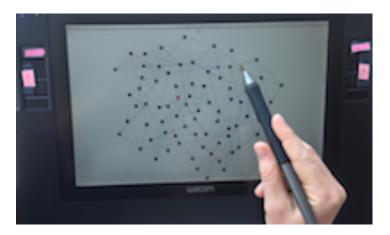
Stretch and squish navigation

Jessica Dawson



Joanna McGrenere (UBC)





Search set model of path tracing

Technique-driven: Dimensionality reduction



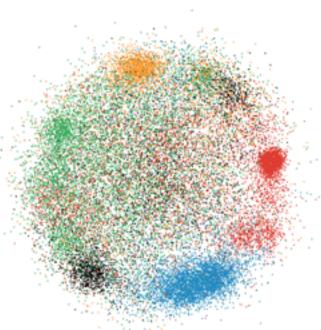
F

P

E







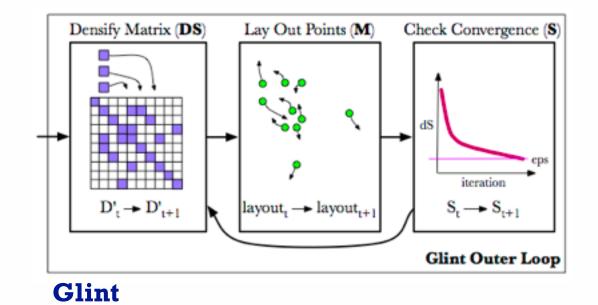
File Starkflow Operation Views

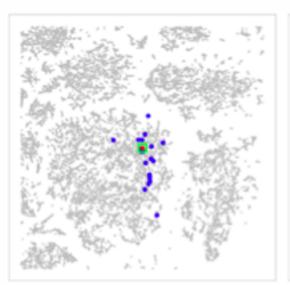
Markflows Stage

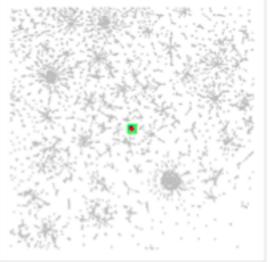
Markflow

Glimmer

DimStiller







QSNE

Evaluation: Dimensionality reduction



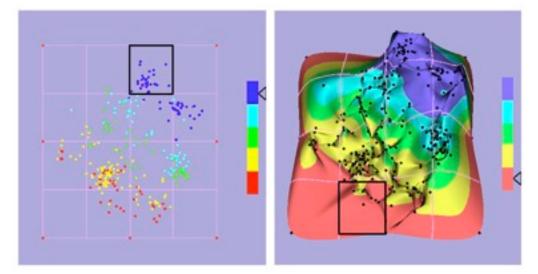












Points vs landscapes for dimensionally reduced data



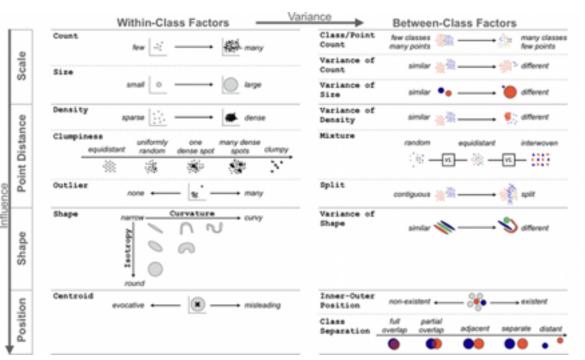
Guidance on DR & scatterplot choices

Michael Sedlmair





Melanie Tory



Taxonomy of cluster separation factors

Problem-driven: Genomics









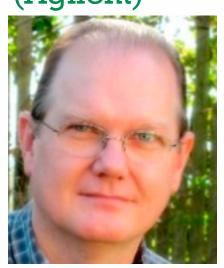
Aaron Barsky



Jenn Gardy (Microbio)



Robert Kincaid (Agilent)



The state of the s

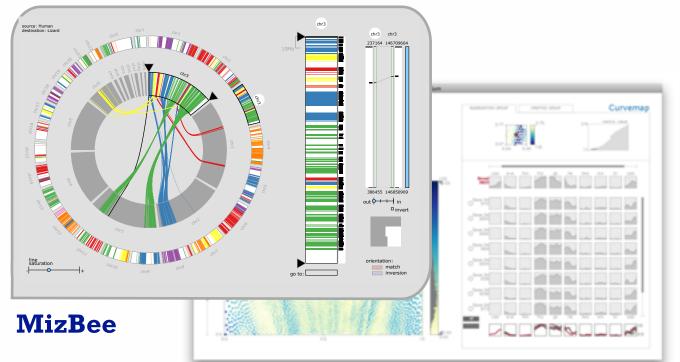
Cerebral

Miriah Meyer



Hanspeter Pfister (Harvard)





Problem-driven: Genomics, fisheries







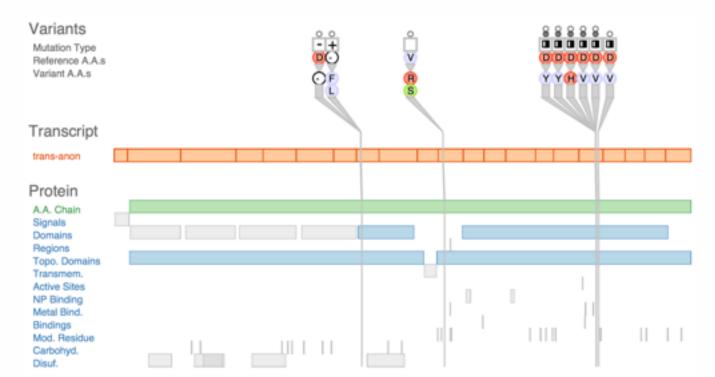






Cydney Nielsen (BC Cancer)





Variant View



Maryam Booshehrian

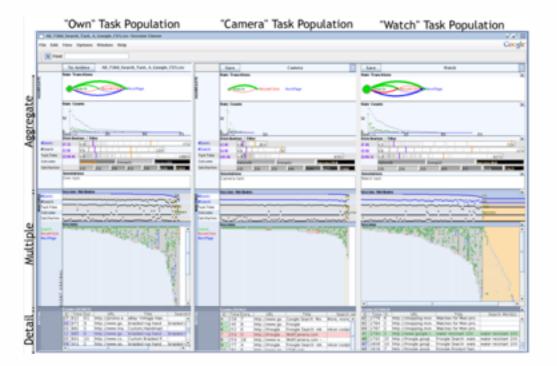


Torsten Moeller (SFU)



Vismon

Problem-driven: Many domains



Heidi Lam



Diane Tang (Google)

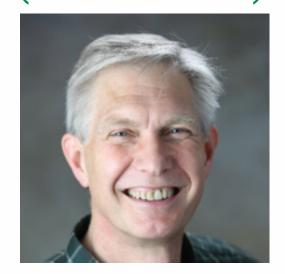


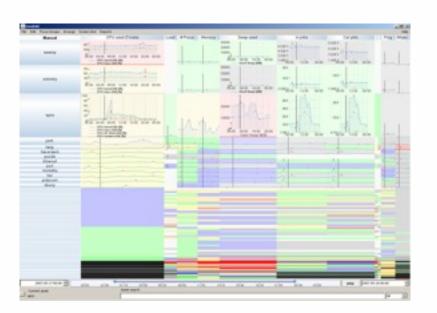
SessionViewer: web log analysis

Peter McLachlan



Stephen North (AT&T Research)





LiveRAC: systems time-series

P

F

E

Evaluation: Focus+Context

T









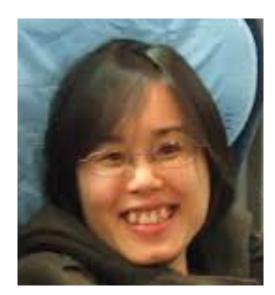


Ron Rensink (UBC)



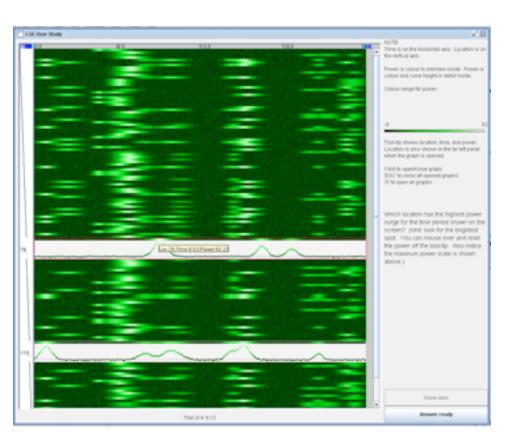
Distortion impact on search/memory

Heidi Lam



Robert Kincaid (Agilent)





Separate vs integrated views

Journalism

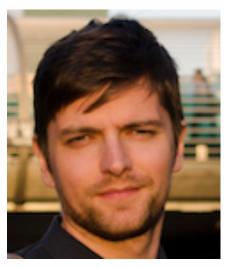












Stephen Ingram



Jonathan Stray (Assoc Press)



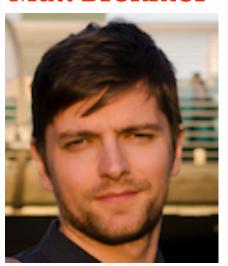
See Advanced to the process of the format of

Overview

Johanna Fulda (Sud. Zeitung)



Matt Brehmer



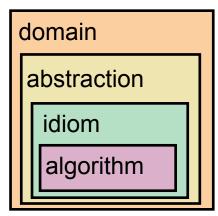


TimeLineCurator

Theoretical foundations

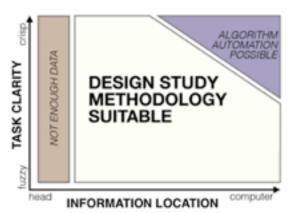
- Visual Encoding Pitfalls
 - Unjustified Visual Encoding
 - Hammer In Search Of Nail
 - 2D Good, 3D Better
 - Color Cacophony
 - Rainbows Just Like In The Sky

- Strategy Pitfalls
- What I Did Over My Summer
- Least Publishable Unit
- Dense As Plutonium
- Bad Slice and Dice



Nested Model

Papers Process & Pitfalls



Design Study Methodology

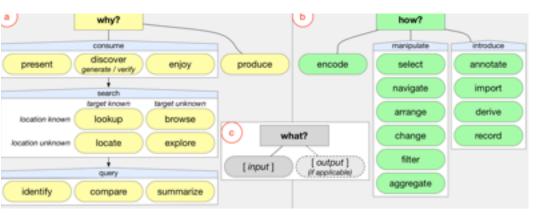
Michael Sedlmair



Miriah Meyer



Matt Brehmer



Abstract Tasks



P

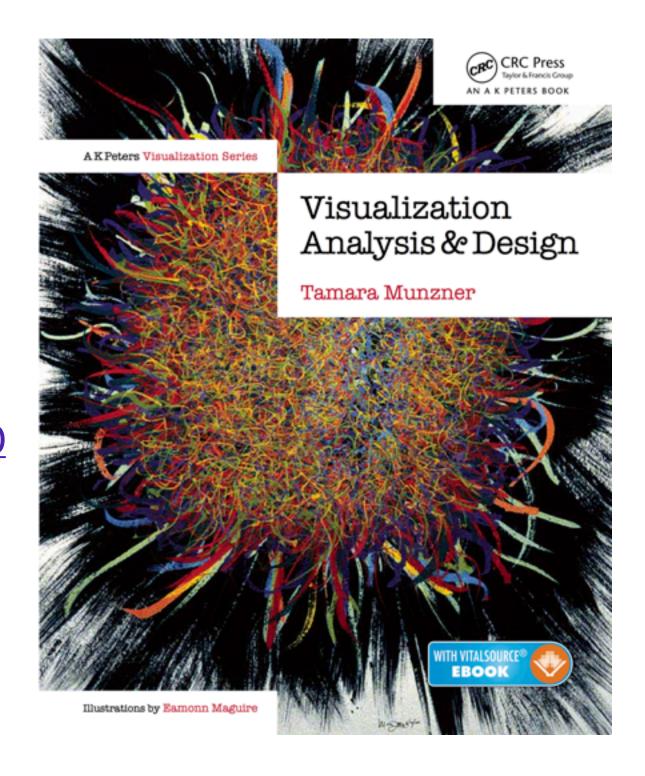
F

E

More Information

@tamaramunzner

- this talk
 http://www.cs.ubc.ca/~tmm/talks.html#vad15fls
- book page (including tutorial lecture slides) http://www.cs.ubc.ca/~tmm/vadbook
 - –20% promo code for book+ebook combo: HVN17
 - http://www.crcpress.com/product/isbn/9781466508910
 - illustrations: Eamonn Maguire
- papers, videos, software, talks, full courses
 http://www.cs.ubc.ca/group/infovis
 http://www.cs.ubc.ca/~tmm



Visualization Analysis and Design.