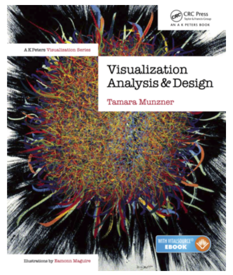
# Visualization Analysis & Design




Tamara Munzner
Department of Computer Science
University of British Columbia

*NASA Goddard Information Science and Technology Colloquium*
*December 14 2016, Greenbelt MD*

http://www.cs.ubc.ca/~tmm/talks.html#vad16nasa

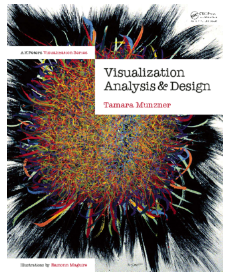
@tamaramunzner

## Visualization (vis) defined & motivated

**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.**

- human in the loop needs the details
  - doesn't know exactly what questions to ask in advance
  - longterm exploratory analysis
  - presentation of known results
  - stepping stone towards automation: refining, trustbuilding
- external representation: perception vs cognition
- intended task, measurable definitions of effectiveness

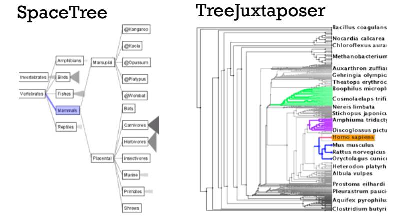



more at:
Visualization Analysis and Design, Chapter 1.
*Munzner. AK Peters Visualization Series, CRC Press, 2014.*



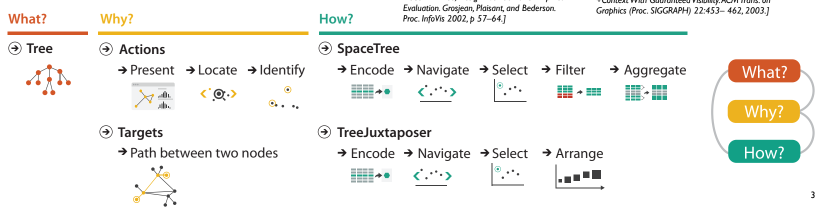
## Why analyze?

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




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

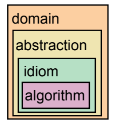
*[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453–462, 2003.]*



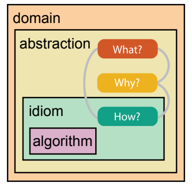



## Analysis framework: Four levels, three questions

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




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




[A Multi-Level Typology of Abstract Visualization Tasks Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]



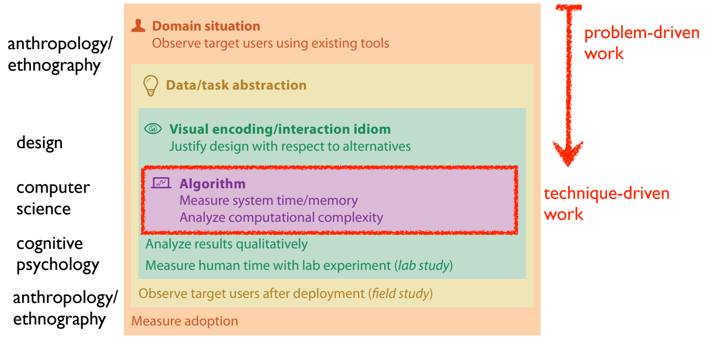
## Why is validation difficult?

- different ways to get it wrong at each level



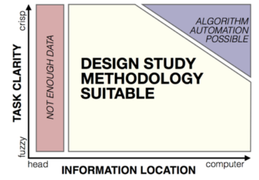

## Why is validation difficult?

- solution: use methods from different fields at each level




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






Michael Sedlmair

Miriah Meyer

# Design Study Methodology

*Reflections from the Trenches and from the Stacks*

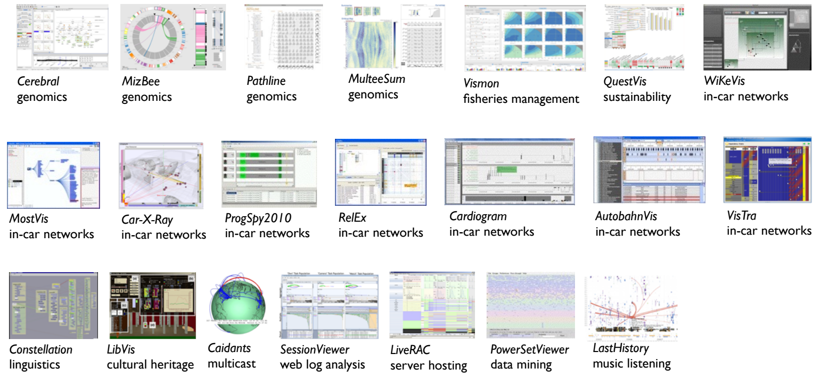
http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/

Tamara Munzner
@tamaramunzner

Design Study Methodology: Reflections from the Trenches and from the Stacks.
*Sedlmair, Meyer, Munzner. IEEE Trans. Visualization and Computer Graphics 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).*



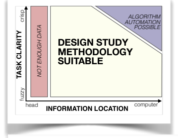
## Design Studies: Lessons learned after 21 of them



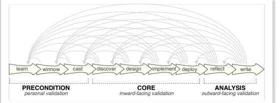



## Methodology for Problem-Driven Work

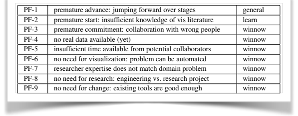
- definitions



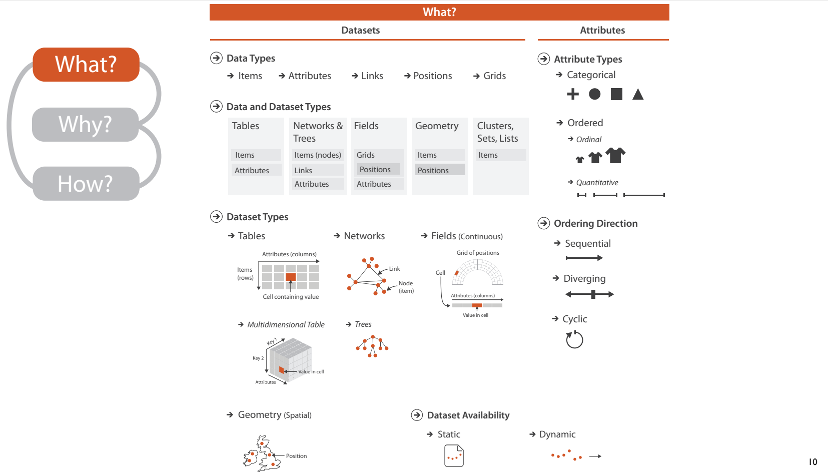
- 9-stage framework




- 32 pitfalls
  and how to avoid them

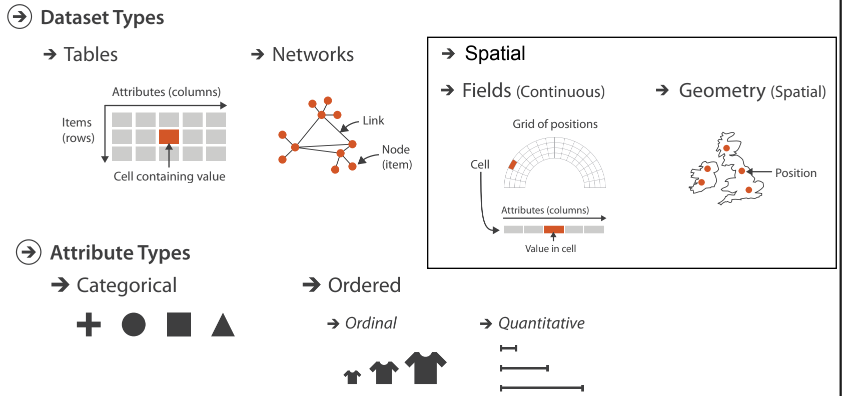




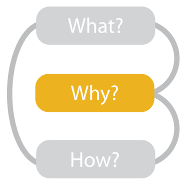




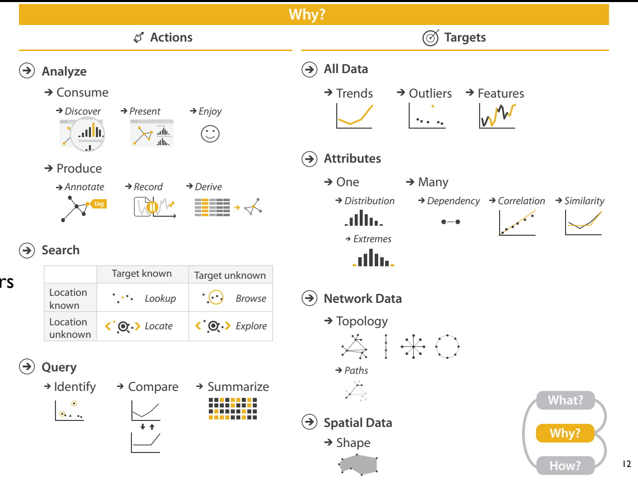

## Types: Datasets and data





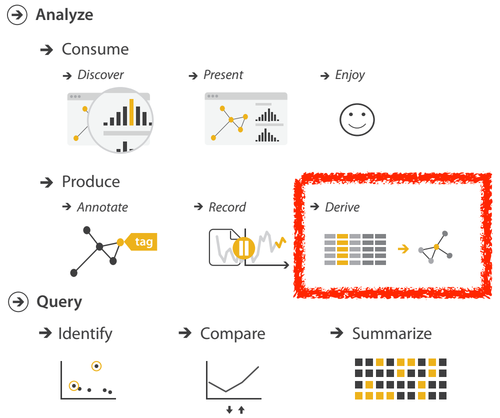

- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*






## Actions: Analyze, Query

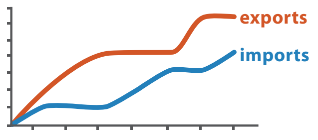
- analyze
  - consume
    - discover vs present
      - aka explore vs explain
    - enjoy
      - aka casual, social
  - produce
    - annotate, record, derive
- query
  - how much data matters?
    - one, some, all
- independent choices



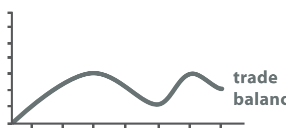



## Derive: Crucial Design Choice

- 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




Original Data




*trade balance = exports − imports*

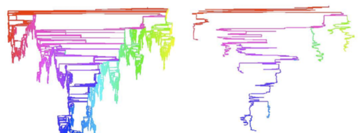
Derived Data

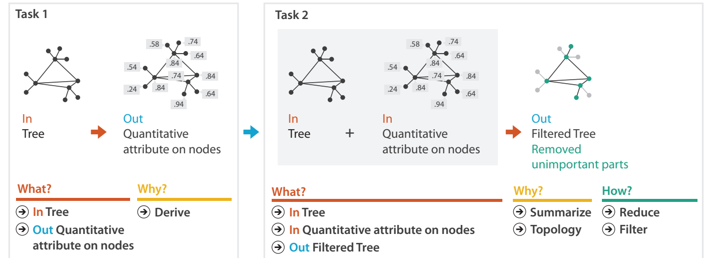


## Analysis example: Derive one attribute

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

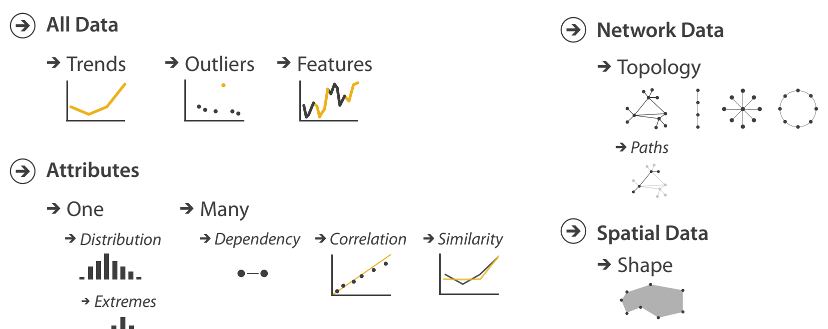
*[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]*

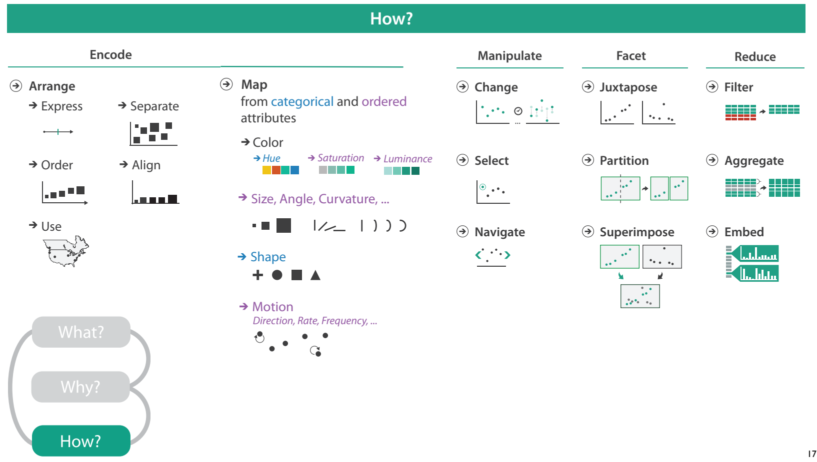






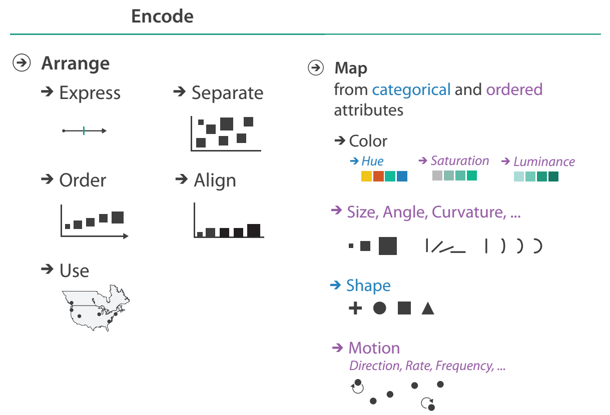

## Targets







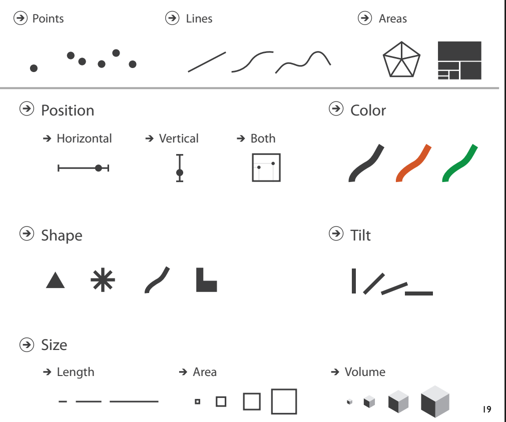

# How to encode: Arrange space, map channels






# Definitions: Marks and channels

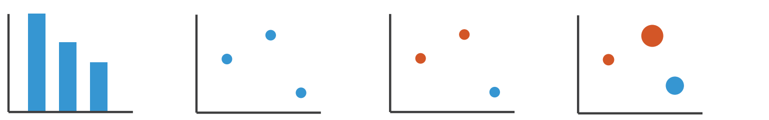
- marks
  - geometric primitives
- channels
  - control appearance of marks






# Encoding visually with marks and channels

- analyze idiom structure
  - as combination of marks and channels



1: vertical position — mark: line

2: vertical position, horizontal position — mark: point

3: vertical position, horizontal position, color hue — mark: point

4: vertical position, horizontal position, color hue, size (area) — mark: point



# 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)

Spatial region
Color hue
Motion
Shape



# 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**
Spatial region
Color hue
Motion
Shape

- expressiveness principle
  - match channel and data characteristics



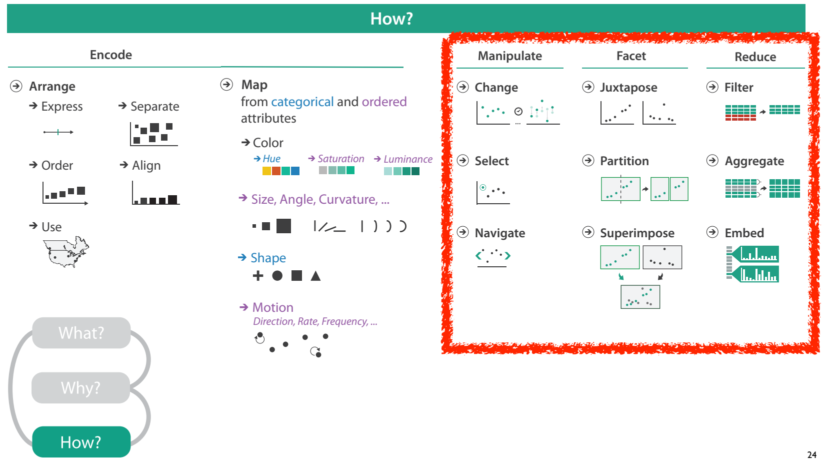
# Channels: Rankings

**Magnitude Channels: Ordered Attributes** (Best → Least Effectiveness)
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**
Spatial region
Color hue
Motion
Shape

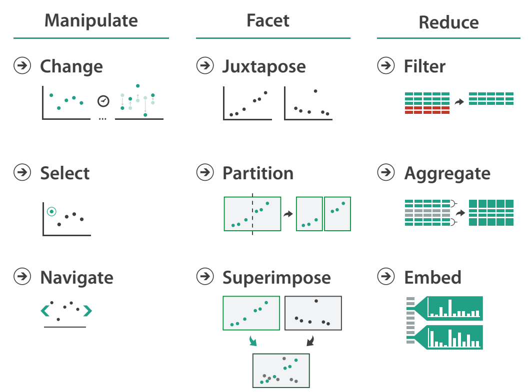
- expressiveness principle
  - match channel and data characteristics
- effectiveness principle
  - encode most important attributes with highest ranked channels







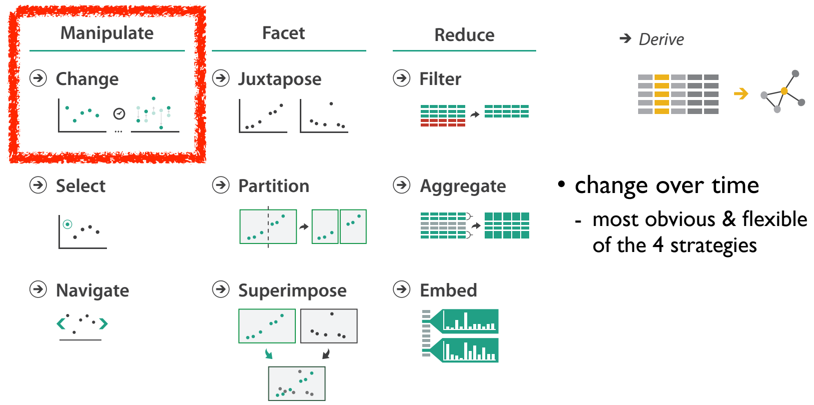

# How to handle complexity: 3 more strategies + 1 previous




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



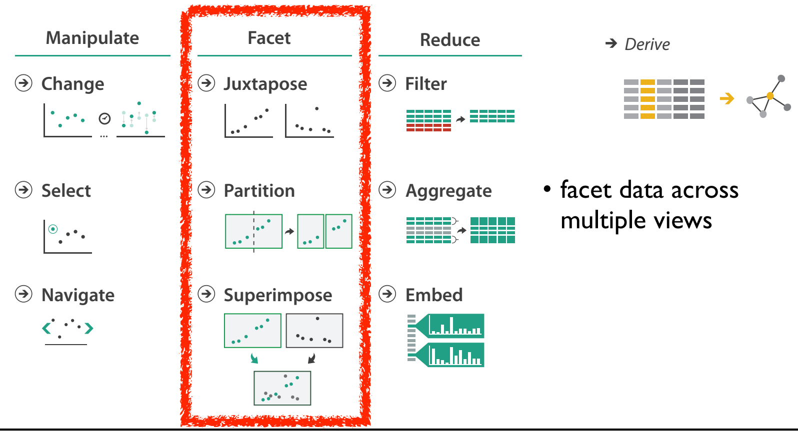
# How to handle complexity: 3 more strategies + 1 previous



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



# How to handle complexity: 3 more strategies + 1 previous

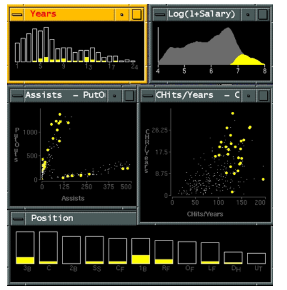


- facet data across multiple views



# Idiom: **Linked highlighting** System: **EDV**

- see how regions contiguous in one view are distributed within another
  - powerful and pervasive interaction idiom
- encoding: different
- data: all shared



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



# Idiom: **bird's-eye maps** System: **Google Maps**

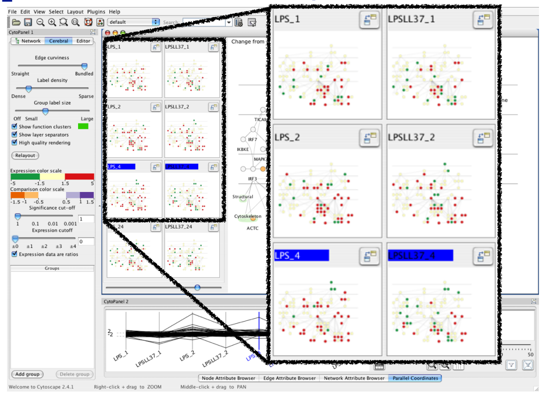
- encoding: same
- data: subset shared
- navigation: shared
  - bidirectional linking
- differences
  - viewpoint
  - (size)
- *overview-detail*

[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

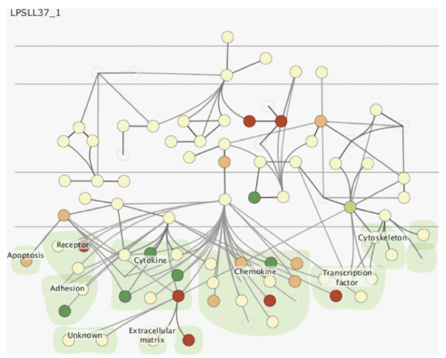
| Encoding | Data: All | Data: Subset | Data: None |
|---|---|---|---|
| Same | Redundant | Overview/Detail | Small Multiples |
| Different | Multiform | Multiform, Overview/Detail | No Linkage |

- 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



# Idiom: Animation (change over time)

- weaknesses
  - widespread changes
  - disparate frames
- strengths
  - choreographed storytelling
  - localized differences between contiguous frames
  - animated transitions between states



## How to handle complexity: 3 more strategies + 1 previous

| Manipulate | Facet | Reduce |
|---|---|---|
| Change | Juxtapose | Filter |
| Select | Partition | Aggregate |
| Navigate | Superimpose | Embed |

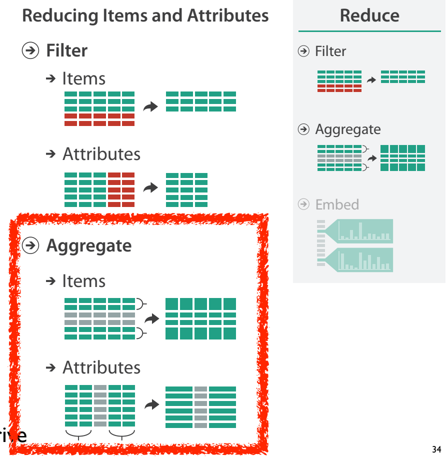
→ Derive

- reduce what is shown within single view



## Reduce items and attributes

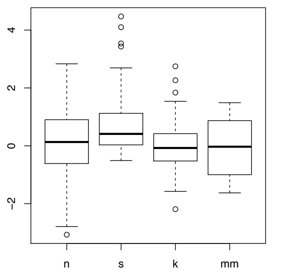
- reduce/increase: inverses
- filter
  - pro: straightforward and intuitive
    - to understand and compute
  - con: out of sight, out of mind
- aggregation
  - pro: inform about whole set
  - con: difficult to avoid losing signal
- not mutually exclusive
  - combine filter, aggregate
  - combine reduce, facet, change, derive






## 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

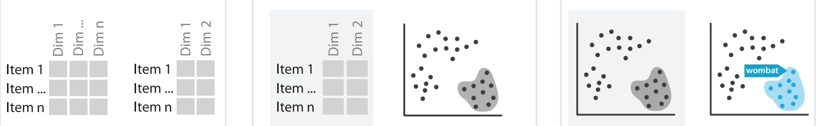



*[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]*



## Idiom: **Dimensionality reduction for documents**

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

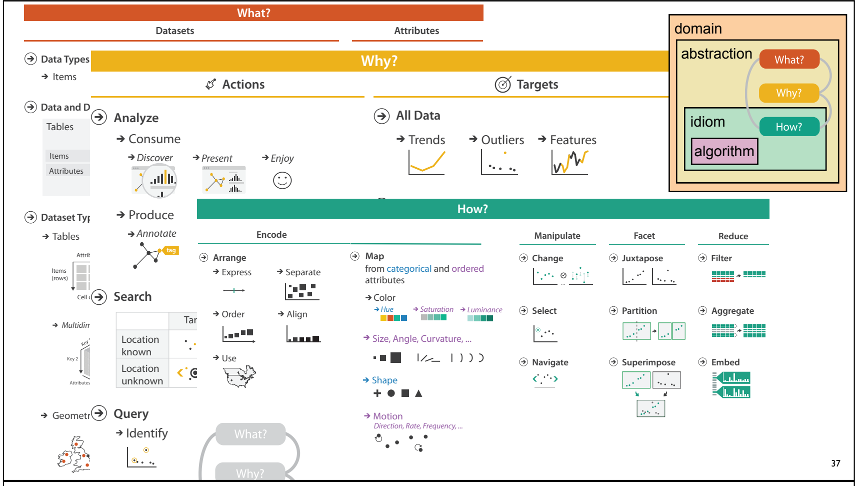


Task 1: In HD data → Out 2D data. What? In High-dimensional data, Out 2D data. Why? Produce, Derive.

Task 2: In 2D data → Out Scatterplot, Clusters & points. What? In 2D data, Out Scatterplot, Out Clusters & points. Why? Discover, Explore, Identify. How? Encode, Navigate, Select.

Task 3: In Scatterplot, Clusters & points → Out Labels for clusters. What? In Scatterplot, In Clusters & points, Out Labels for clusters. Why? Produce, Annotate.







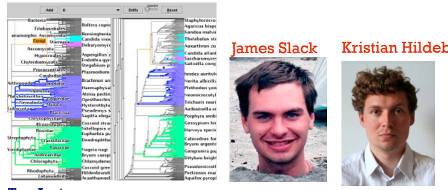

## A quick taste of my own work!

- technique-driven work
- problem-driven work
- theoretical foundations
- evaluation



## Technique-driven: Graph drawing

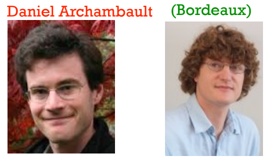
T F E P

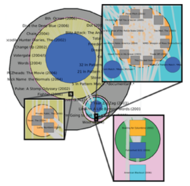


James Slack, Kristian Hildebrand

TreeJuxtaposer

Daniel Archambault, David Auber (Bordeaux)





TopoLayout, SPF, Grouse, GrouseFlocks, TugGraph



## Evaluation: Graph drawing

T F E P

Dmitry Nekrasovski, Adam Bodnar, Joanna McGrenere (UBC)

Stretch and squish navigation

Jessica Dawson, Joanna McGrenere (UBC)

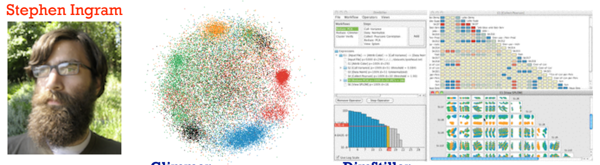
Search set model of path tracing



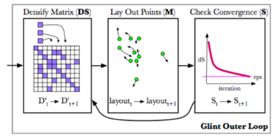
## Technique-driven: Dimensionality reduction

T F E P

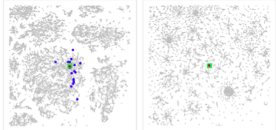
Stephen Ingram



Glimmer, DimStiller



Glint



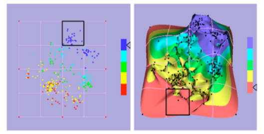
QSNE



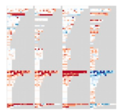
## Evaluation: Dimensionality reduction

T F E P

Melanie Tory



Points vs landscapes for dimensionally reduced data



Guidance on DR & scatterplot choices

Michael Sedlmair, Melanie Tory (UVic)

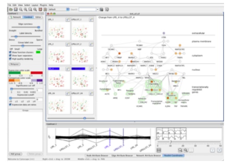
Taxonomy of cluster separation factors



## Problem-driven: Genomics

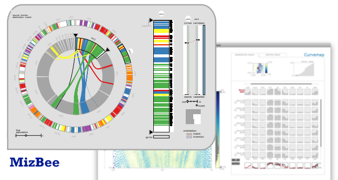
T F E P

Aaron Barsky, Jenn Gardy (Microbio), Robert Kincaid (Agilent)



Cerebral

Miriah Meyer, Hanspeter Pfister (Harvard)



MizBee

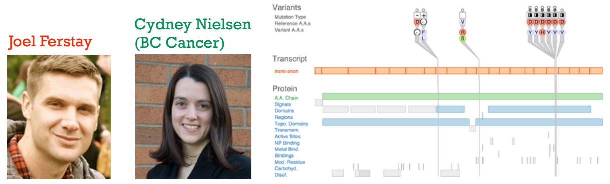
MulteeSum, Pathline



## Problem-driven: Genomics, fisheries

T F E P

Joel Ferstay, Cydney Nielsen (BC Cancer)



Variant View

Vismon

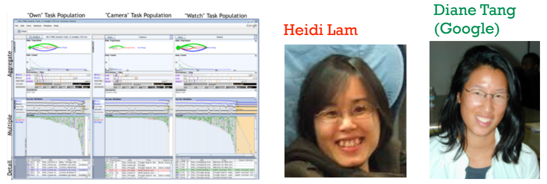
Maryam Booshehrian, Torsten Moeller (SFU)



## Problem-driven: Many domains

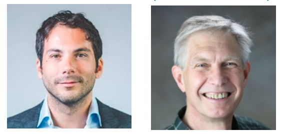
T F E P

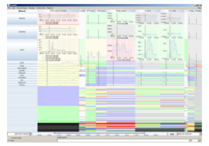
Heidi Lam, Diane Tang (Google)



SessionViewer: web log analysis

Peter McLachlan, Stephen North (AT&T Research)





LiveRAC: systems time-series



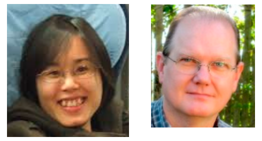
## Evaluation: Focus+Context

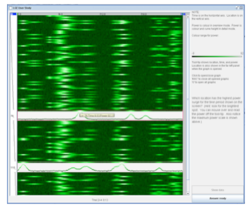
T F E P

Heidi Lam, Ron Rensink (UBC)

Distortion impact on search/memory

Heidi Lam, Robert Kincaid (Agilent)





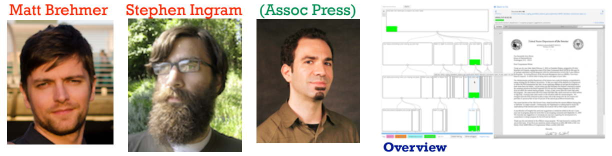
Separate vs integrated views



## Journalism

T F E P

Matt Brehmer, Stephen Ingram, Jonathan Stray (Assoc Press)



Overview

Johanna Fulda (Sud. Zeitung), Matt Brehmer

TimeLineCurator

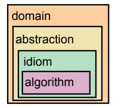


## Theoretical foundations

T F E P

- 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

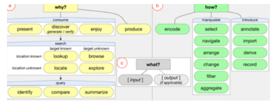
Papers Process & Pitfalls




Nested Model

Design Study Methodology

Michael Sedlmair, Miriah Meyer

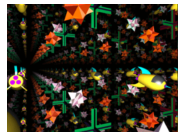


Abstract Tasks

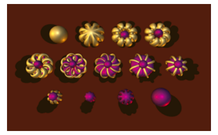
Matt Brehmer

# Geometry Center 1990-1995

Geomview



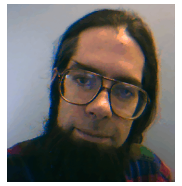
The Shape of Space



Outside In

Charlie Gunn

Stuart Levy



Mark Phillips

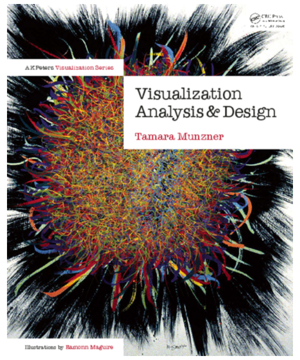
Delle Maxwell



# More Information

@tamaramunzner

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#vad16nasa
- 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, courses
  http://www.cs.ubc.ca/group/infovis
  http://www.cs.ubc.ca/~tmm




Visualization Analysis and Design.
*Munzner. A K Peters Visualization Series, CRC Press, Visualization Series, 2014.*