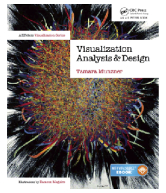
Visualization Analysis & Design



Tamara Munzner
Department of Computer Science
University of British Columbia

UBC Alumni/Industry Lecture
Feb 27 2020, Vancouver BC

DESIGNING for PEOPLE
CAIDA
UBC Data Science Institute

http://www.cs.ubc.ca/~tmm/talks.html#vad20alum **@tamaramunzner**

Visualization: definition & motivation

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.

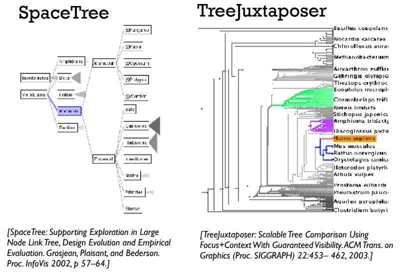
- human in the loop needs the details & no trusted automatic solution exists
  - doesn't know exactly what questions to ask in advance
  - exploratory data analysis
    - ***speed up*** through human-in-the-loop visual data analysis
  - present known results to others
  - stepping stone towards automation
    - before model creation to provide understanding
    - during algorithm creation to refine, debug, set parameters
    - before or during deployment to build trust and monitor

more at: Visualization Analysis and Design. Munzner. CRC Press, 2014.

www.cs.ubc.ca/~tmm/talks.html#vad20alum

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.]

What?	Why?	How?
Tree	Actions: Present, Locate, Identify	SpaceTree: Encode, Navigate, Select, Filter, Aggregate
	Targets: Path between two nodes	TreeJuxtaposer: Encode, Navigate, Select, Arrange

www.cs.ubc.ca/~tmm/talks.html#vad20alum

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

domain / abstraction / idiom / algorithm

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

domain / abstraction (What? Why?) / idiom (How?) / algorithm

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

www.cs.ubc.ca/~tmm/talks.html#vad20alum

Why is validation difficult?

- different ways to get it wrong at each level

Domain situation — You misunderstood their needs
Data/task abstraction — You're showing them the wrong thing
Visual encoding/interaction idiom — The way you show it doesn't work
Algorithm — Your code is too slow

www.cs.ubc.ca/~tmm/talks.html#vad20alum

Why is validation difficult?

- solution: use methods from different fields at each level

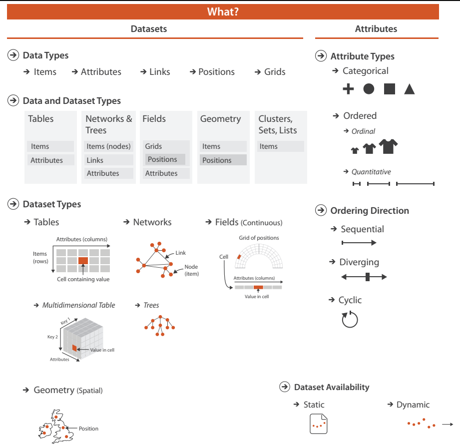
anthropology/ ethnography	**Domain situation** — Observe target users using existing tools
	Data/task abstraction
design	**Visual encoding/interaction idiom** — Justify design with respect to alternatives
computer science	**Algorithm** — Measure system time/memory; Analyze computational complexity
cognitive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

problem-driven work
technique-driven work

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

www.cs.ubc.ca/~tmm/talks.html#vad20alum

What? Why? How?



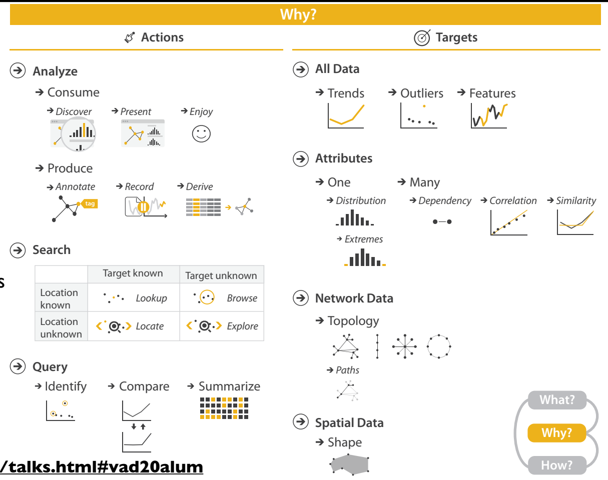
www.cs.ubc.ca/~tmm/talks.html#vad20alum

Types: Datasets and data

Dataset Types
- Tables — Attributes (columns), Items (rows), Cell containing value
- Networks — Link, Node (item)
- Spatial
  - Fields (Continuous) — Grid of positions, Cell, Attributes (columns), Value in cell
  - Geometry (Spatial) — Position

Attribute Types
- Categorical
- Ordered
  - *Ordinal*
  - *Quantitative*

www.cs.ubc.ca/~tmm/talks.html#vad20alum

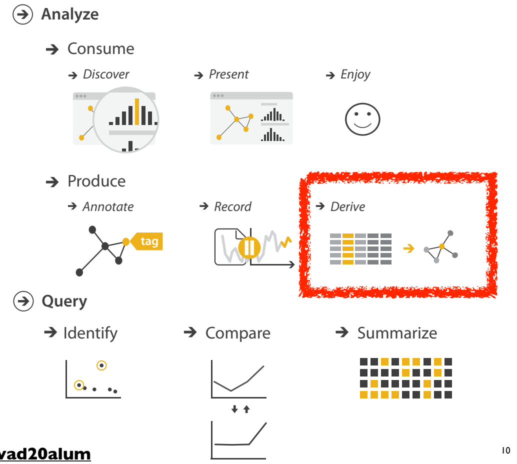


- {action, target} pairs
  - discover distribution
  - compare trends
  - locate outliers
  - browse topology

www.cs.ubc.ca/~tmm/talks.html#vad20alum

Actions: Analyze, Query

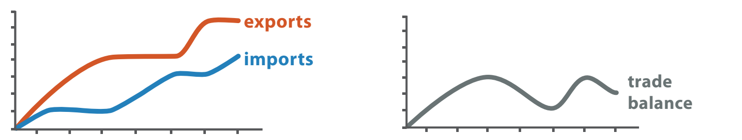
- analyze
  - consume
    - discover vs present
      - aka explore vs explain
    - enjoy
      - aka casual, social
  - produce
    - annotate, record, derive
- query
  - how much data matters?
    - one, some, all
- independent choices
  - analyze, query, (search)



www.cs.ubc.ca/~tmm/talks.html#vad20alum

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



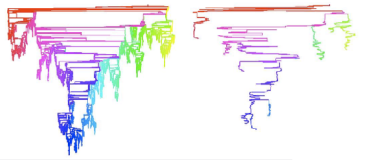
trade balance = *exports* − *imports*

Original Data — Derived Data

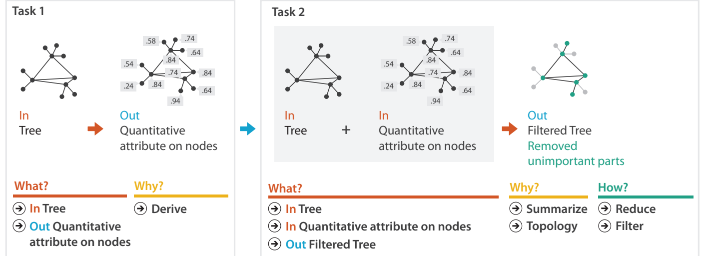
www.cs.ubc.ca/~tmm/talks.html#vad20alum

Analysis example: Derive one attribute

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

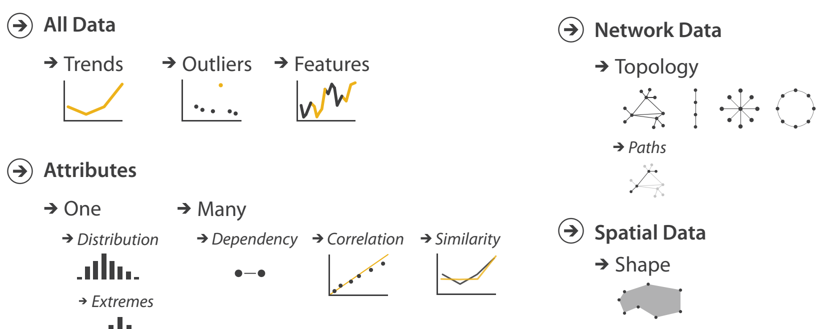


[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]

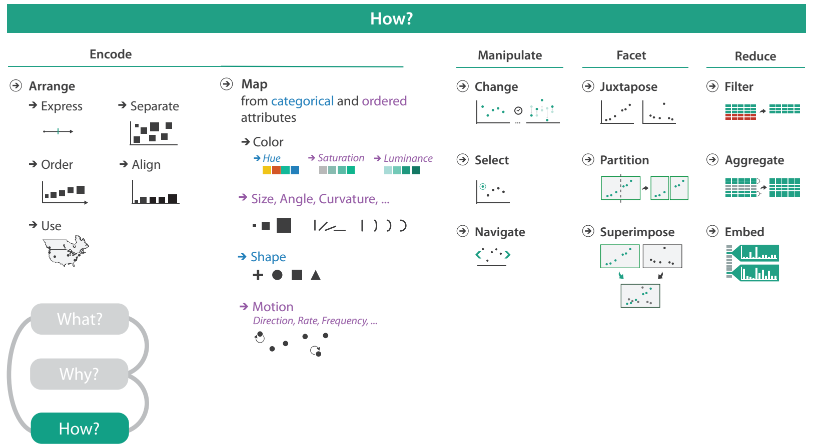


www.cs.ubc.ca/~tmm/talks.html#vad20alum

Targets

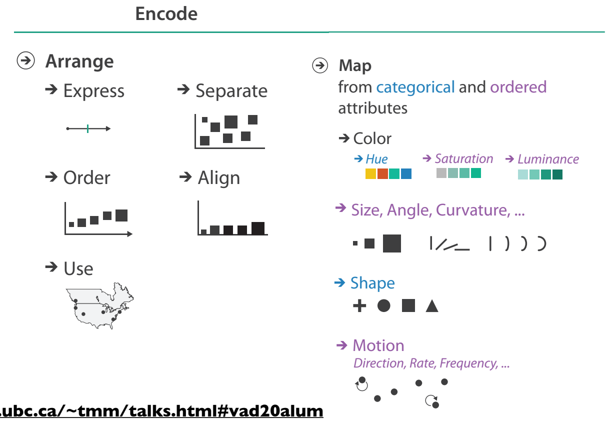


www.cs.ubc.ca/~tmm/talks.html#vad20alum



www.cs.ubc.ca/~tmm/talks.html#vad20alum

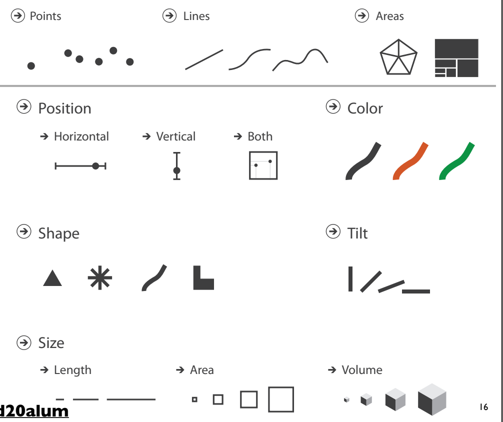
How to encode: Arrange space, map channels



www.cs.ubc.ca/~tmm/talks.html#vad20alum

Definitions: Marks and channels

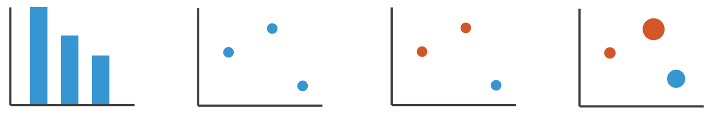
- marks
  - geometric primitives
  - one per item
- channels
  - control appearance of marks



www.cs.ubc.ca/~tmm/talks.html#vad20alum

Encoding visually with marks and channels

- analyze idiom structure
  - as combination of marks and channels



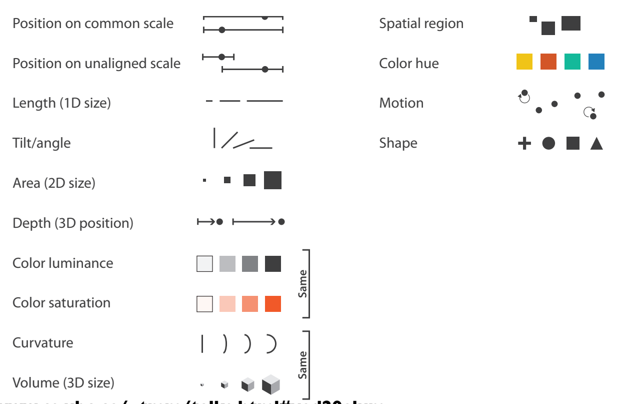
1: vertical position — mark: line

2: vertical position, horizontal position — mark: point

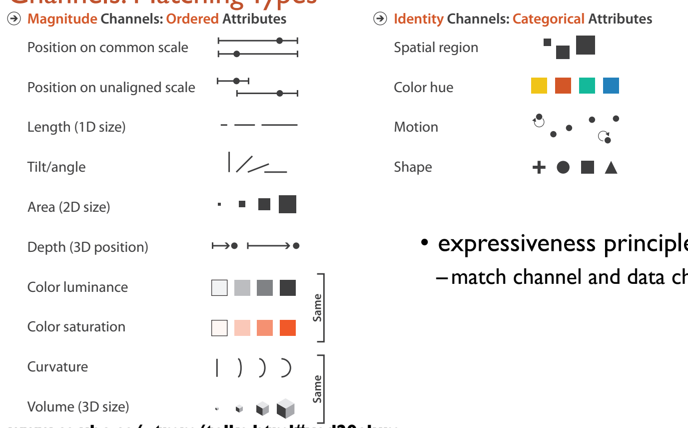
3: vertical position, horizontal position, color hue — mark: point

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

Channels

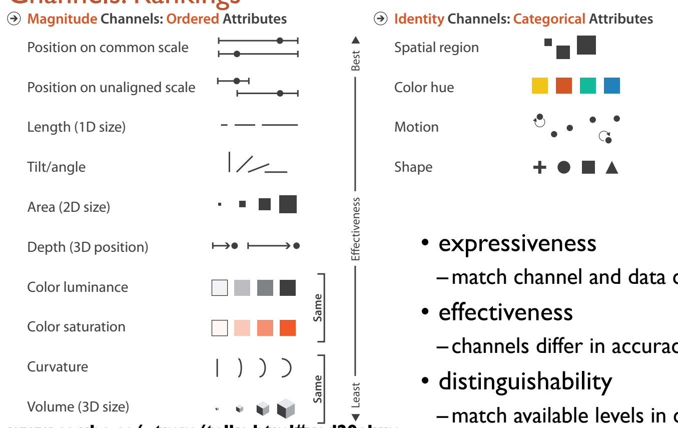


Channels: Matching Types

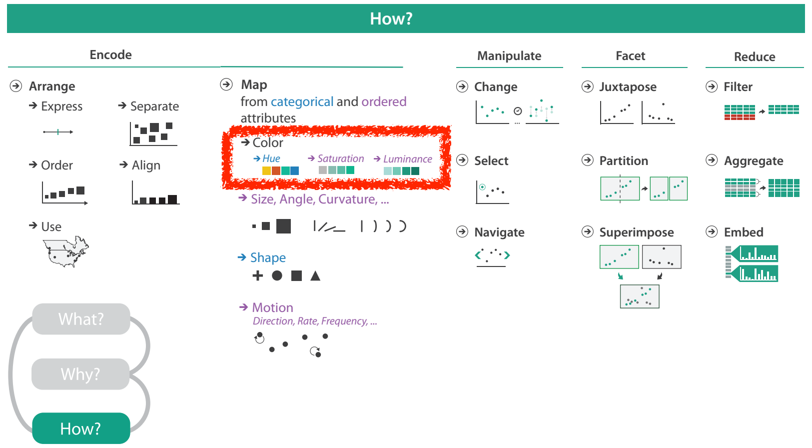


- expressiveness principle
  - match channel and data characteristics

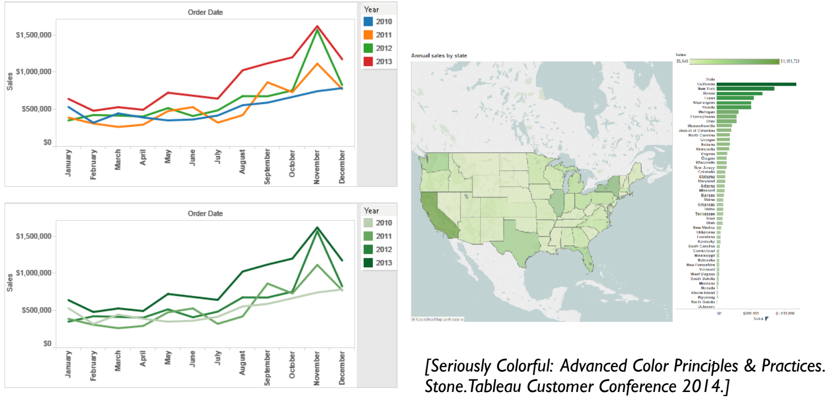
Channels: Rankings



- expressiveness
  - match channel and data characteristics
- effectiveness
  - channels differ in accuracy of perception
- distinguishability
  - match available levels in channel w/ data



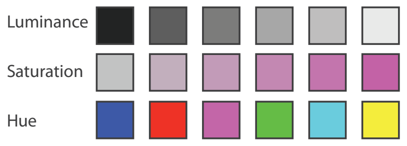
Categorical vs ordered color



[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

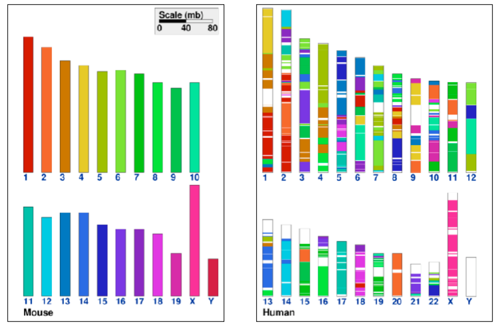
Decomposing color

- first rule of color: do not talk about color!
  - color is confusing if treated as monolithic
- decompose into three channels
  - ordered can show magnitude
    - luminance: how bright
    - saturation: how colorful
  - categorical can show identity
    - hue: what color
  - caveat: not well supported by current tools
- channels have different properties
  - what they convey directly to perceptual system
  - how much they can convey: how many discriminable bins can we use?



Categorical color: limited number of discriminable bins

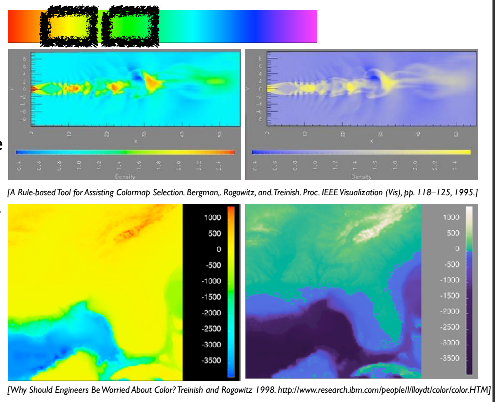
- human perception built on relative comparisons
  - great if color contiguous
  - surprisingly bad for absolute comparisons
- noncontiguous small regions of color
  - fewer bins than you want
  - rule of thumb: 6-12 bins, including background and highlights
- alternatives? other talks!



[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

Ordered color: Rainbow is poor default

- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - small-scale structure: see & name
- alternatives
  - large-scale structure: fewer hues
  - known structure: segmented
  - have it both ways, small+large:
    - multiple hues
    - monotonically increasing luminance



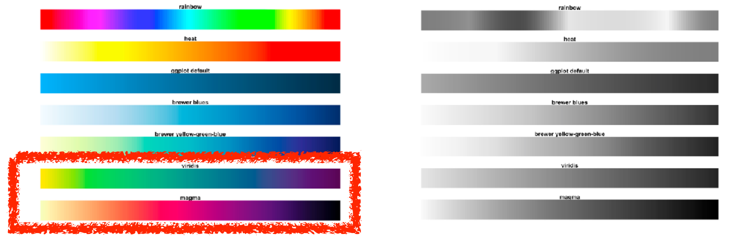
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

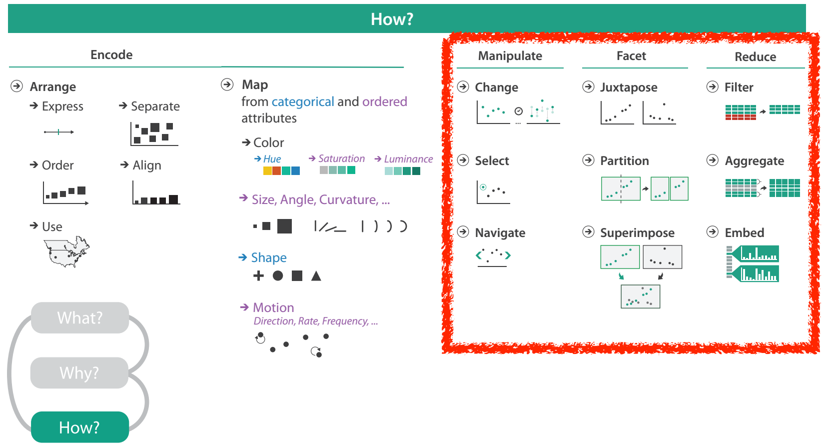
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

Viridis / Magma

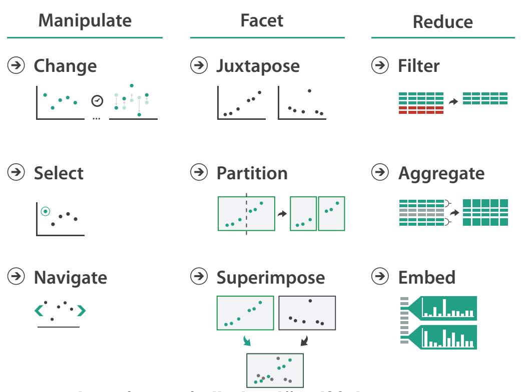
- colorful, perceptually uniform, colorblind-safe, monotonically increasing luminance

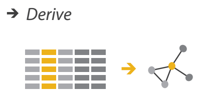
https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html





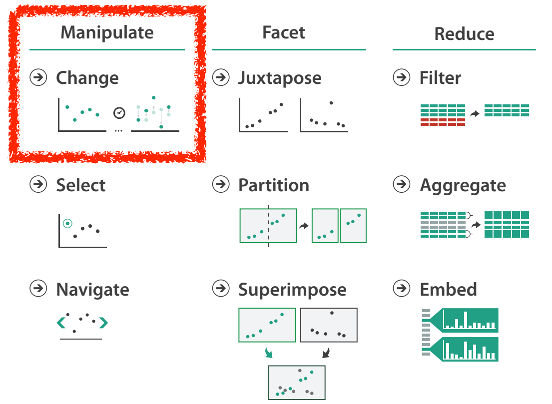
How to handle complexity: 3 more strategies + 1 previous

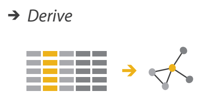




- change view over time
- facet across multiple views
- reduce items/attributes within single view
- derive new data to show within view

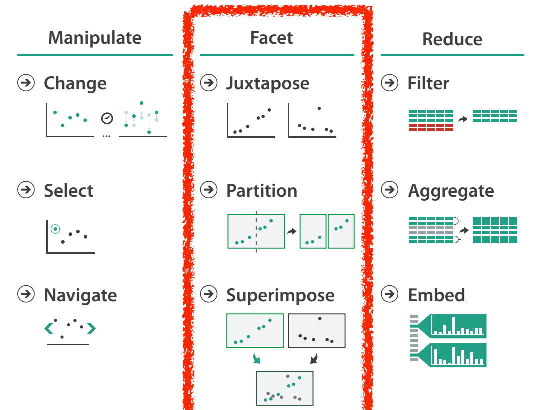
How to handle complexity: 3 more strategies + 1 previous

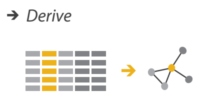




- change over time
  - most obvious & flexible of the 4 strategies

How to handle complexity: 3 more strategies + 1 previous

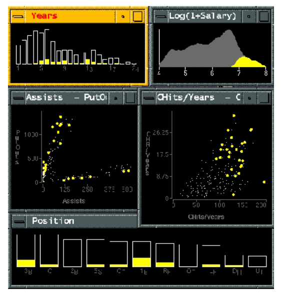




- facet data across multiple views

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.]

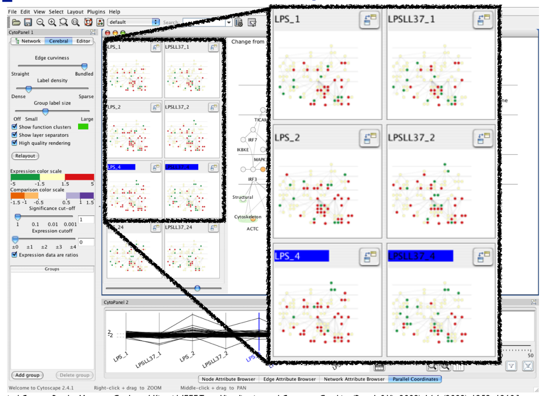
Idiom: **bird's-eye maps** System: **Google Maps**

- 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
  - nodes colored differently for each time/condition case
  - (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.]

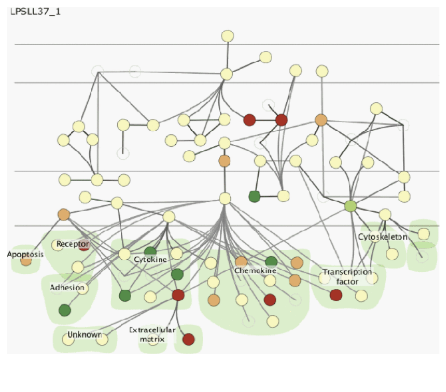
Coordinate views: Design choice interaction

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

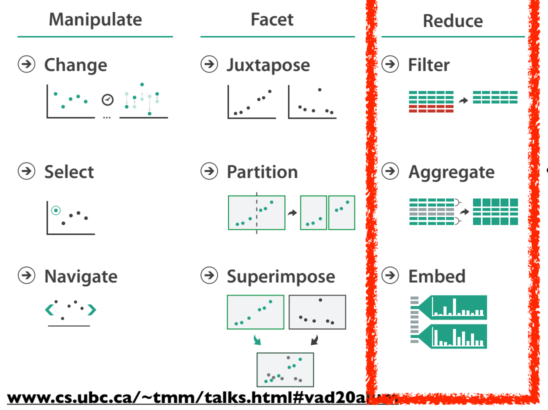
- why juxtapose views?
  - benefits: eyes vs memory
    - lower cognitive load to move eyes between 2 views than remembering previous state with single changing view
  - costs: display area, 2 views side by side each have only half the area of one view

Idiom: Animation (change over time)

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



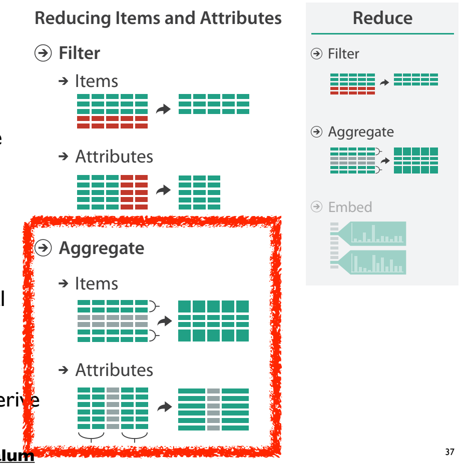
How to handle complexity: 3 more strategies + 1 previous



- reduce what is shown within single view

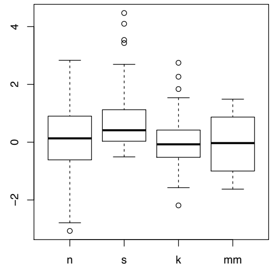
Reduce items and attributes

- 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**

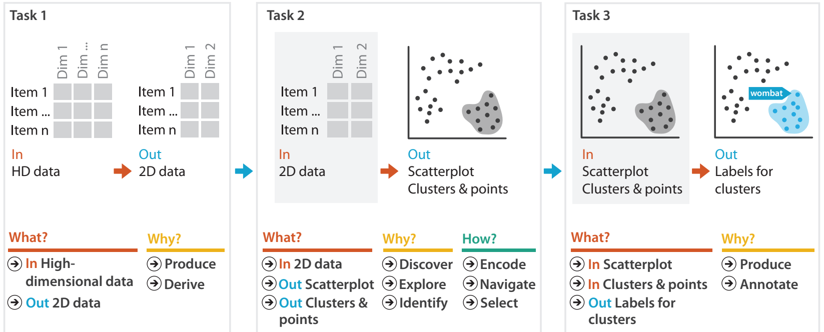
- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown

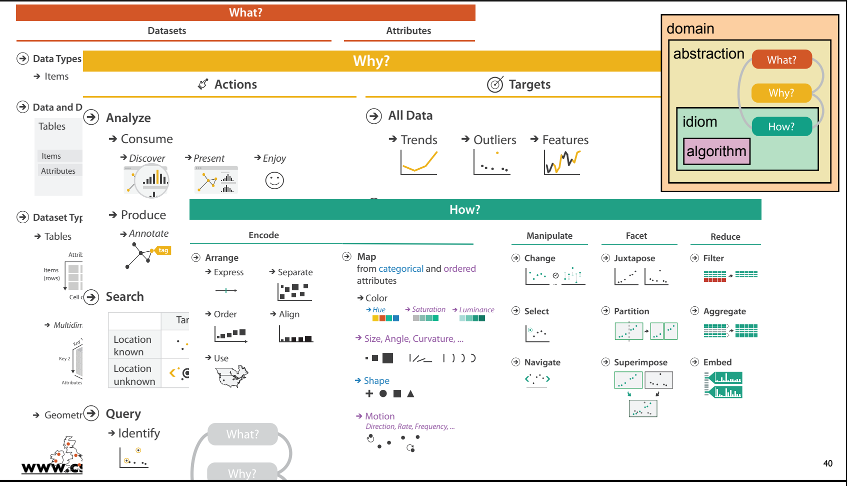


[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

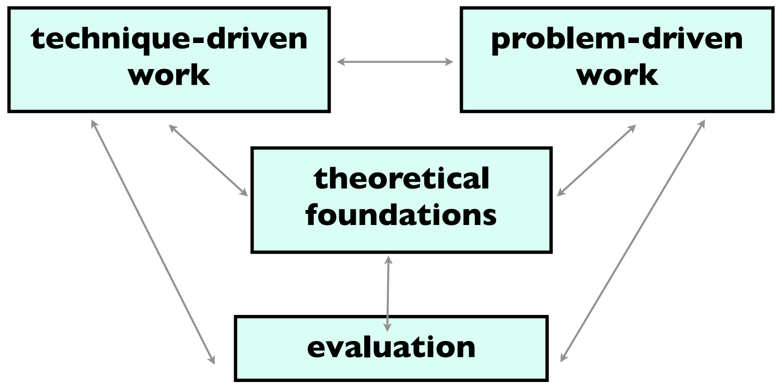
Idiom: **Dimensionality reduction for documents**

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





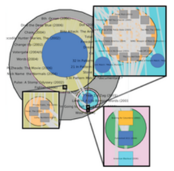
A quick taste of my own work!



Technique-driven: Graph/network drawing

T P F E

Daniel Archambault David Auber (Bordeaux)



https://youtu.be/AWXAe8zvkt8

TopoLayout
SPF
Grouse
GrouseFlocks
TugGraph

Benjamin Renoust Guy Melançon (Bordeaux)

Detangler
https://youtu.be/QOtnHSsUV6k

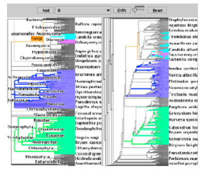
Technique-driven: Tree drawing

T P F E

Zipeng Liu

Shing Hei Zhan

Aggregated Dendrograms
https://youtu.be/2SLcz7KNLJw

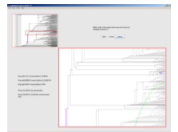


TreeJuxtaposer
https://youtu.be/GdaPj8a9QEo

Evaluation experiments: Graph/tree drawing

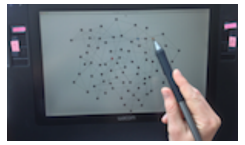
T P F E

Dmitry Nekrasovski Adam Bodnar Joanna McGrenere



Stretch and squish navigation

Jessica Dawson Joanna McGrenere



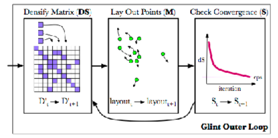
Search set model of path tracing

Technique-driven: Dimensionality reduction

T P F E

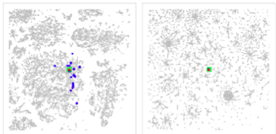
Stephen Ingram

Glimmer



Glint

DimStiller

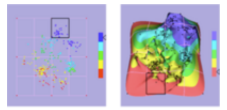


QSNE

Evaluation experiments: Dimensionality reduction

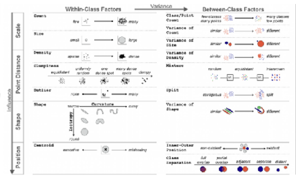
T P F E

Melanie Tory

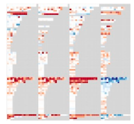


Points vs landscapes for dimensionally reduced data

Michael Sedlmair Melanie Tory



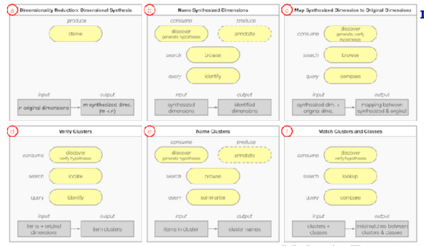
Taxonomy of cluster separation factors



Guidance on DR & scatterplot choices

Evaluation in the field: Dimensionality reduction

T P F E



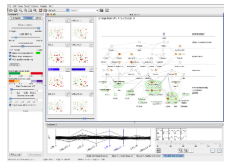
DR in the Wild

Matt Brehmer Michael Sedlmair Melanie Tory Stephen Ingram

Problem-driven: Genomics

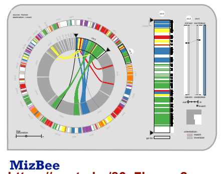
T P F E

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

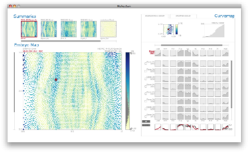


Cerebral
https://youtu.be/76HhG1FQnql

Miriah Meyer Hanspeter Pfister (Harvard)



MizBee
https://youtu.be/86p7brwuz2g



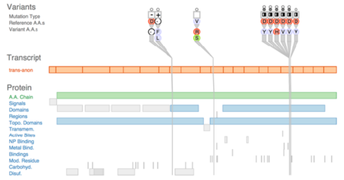
MulteeSum, Pathline

Problem-driven: Genomics, fisheries

T F E P

Joel Ferstay

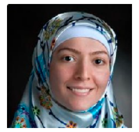
Cydney Nielsen (BC Cancer)



Variant View
https://youtu.be/AHDnv_qMXxQ

Vismon https://youtu.be/h0kHoS4VYmk

Maryam Booshehrian



Torsten Moeller (SFU)

Problem-driven: Tech industry

T F E P

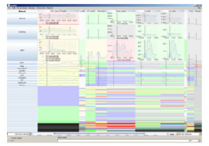
Heidi Lam

Diane Tang (Google)

SessionViewer: web log analysis
https://youtu.be/T4MaTZd56G4

Peter McLachlan

Stephen North (AT&T Research)



LiveRAC: systems time-series https://youtu.be/ld0c3H0VSkw

Problem-driven: Building energy mgmt, journalism

T F E P

Matt Brehmer

Kevin Tate (Pulse/EnerNOC)

Energy Manager

redesign success:
industrial swdev
resources committed

Matt Brehmer

Stephen Ingram

Jonathan Stray (Assoc Press)

Overview
https://vimeo.com/71483614

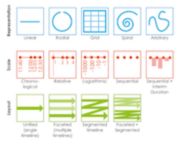
Curation & Presentation: Timelines

T F E P

TimeLineCurator
https://vimeo.com/123246662

Matt Brehmer

Johanna Fulda (Sud. Zeitung)



Timelines Revisited
timelinesrevisited.github.io/

Matt Brehmer

Bongshin Lee (Microsoft)

Benjamin Bach (Microsoft)

Nathalie Henry-Riche

Problem-driven: Current data science

T F E P

Kimberly Dextras-Romagnino

recent work:
Segmentifier
(Mobify)

e-commerce clickstreams

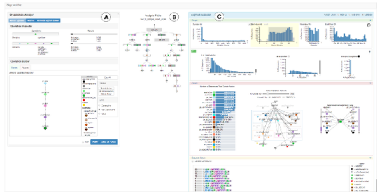
build tools for human-in-the-loop visual data analysis

Michael Oppermann

recent work:
Ocupado
(Sensible Building Science)

wifi proxy for real-time building occupancy

visual analytics for facilities management

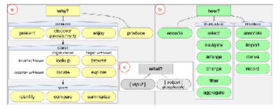


https://youtu.be/TobYDFelSOg

Theoretical foundations: Typologies

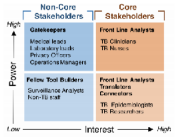
T F E P

Matt Brehmer

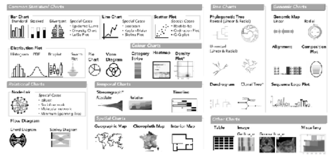


Abstract Tasks

Anamaria Crisan



Regulatory & Organizational Constraints



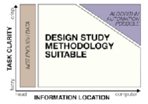
GEViT: Genomic Epidemiology Visualization Typology

Theoretical foundations

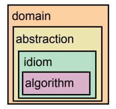
T F E P

- 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



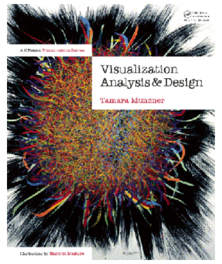
Design Study Methodology



Nested Model

Michael Sedlmair

Miriah Meyer

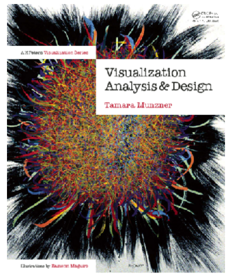


Visualization Analysis & Design

More Information

@tamaramunzner

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#vad20alum
- 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.