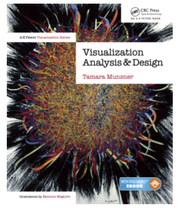
Visualization Analysis & Design

Tamara Munzner
Department of Computer Science
University of British Columbia

Bedford Lab, Hutch Cancer Research Centre
July 28 2017, Seattle WA

www.cs.ubc.ca/~tmm/talks.html#vad17bedford

@tamaramunzner



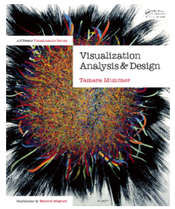
Visualization (vis) defined & motivated

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.



2

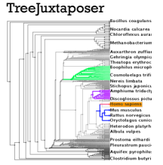
Why analyze?

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

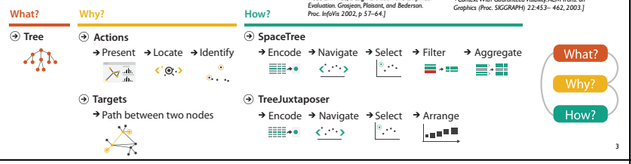
SpaceTree

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

TreeJuxtaposer



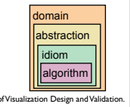
[TreeJuxtaposer: Scalable Tree Comparison Using Focus +Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]



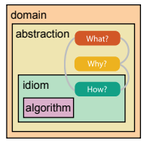
3

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).*]

4

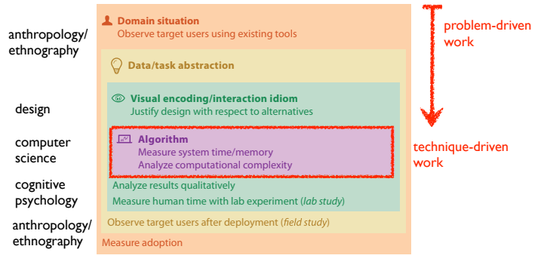
Why is validation difficult?

- different ways to get it wrong at each level

5

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).]

6

Michael Sedlmair

Miriah Meyer

Design Study Methodology

Reflections from the Trenches and from the Stacks

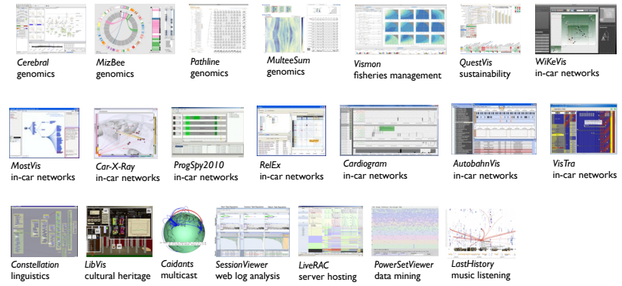
http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/

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).

7

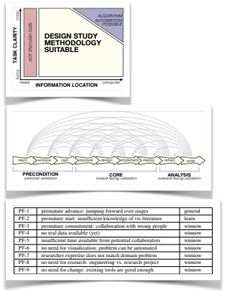
Design Studies: Lessons learned after 21 of them



8

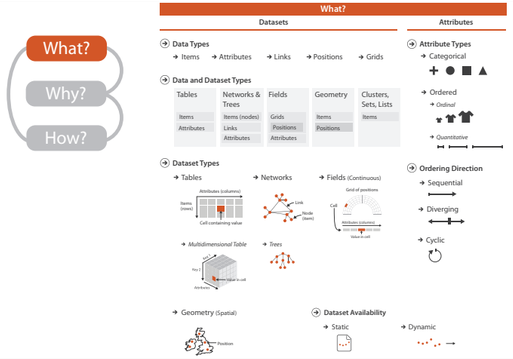
Methodology for Problem-Driven Work

- definitions
- 9-stage framework
- 32 pitfalls and how to avoid them
  - some on collaboration
  - some still apply even when designer == domain expert



PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaborators with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow

9

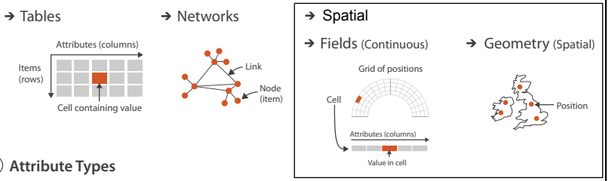


10

Types: Datasets and data

Dataset Types

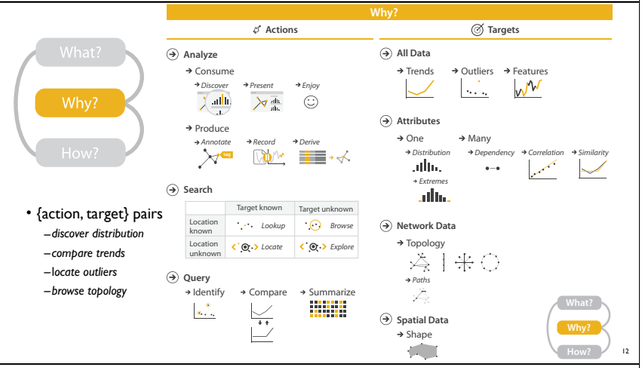
- Tables
- Networks
- Spatial
  - Fields (Continuous)
  - Geometry (Spatial)



Attribute Types

- Categorical
- Ordered
  - Ordinal
  - Quantitative

11

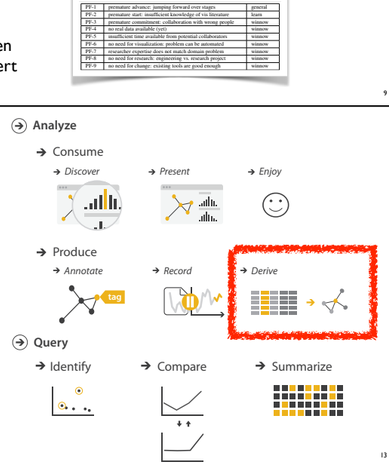


- {action, target} pairs
  - –*discover distribution*
  - –*compare trends*
  - –*locate outliers*
  - –*browse topology*

12

Actions: Analyze, Query

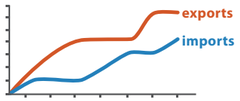
- analyze
  - –consume
    - discover vs present
      - –aka explore vs explain
    - enjoy
      - –aka casual, social
  - –produce
    - annotate, record, derive
- query
  - –how much data matters?
    - one, some, all
- independent choices
  - –analyze, query, (search)



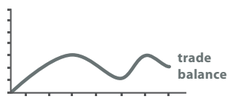
13

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

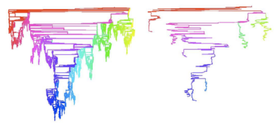
Derived Data

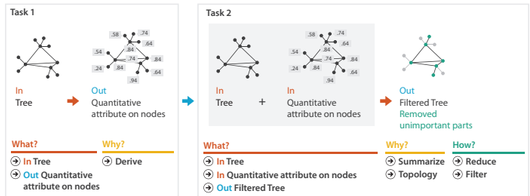
14

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.]





15

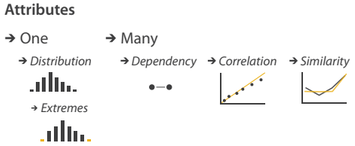
Targets

All Data

- Trends
- Outliers
- Features

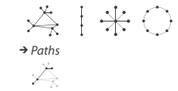
Attributes

- One
- Many
  - Distribution
  - Dependency
  - Correlation
  - Similarity
  - Extremes



Network Data

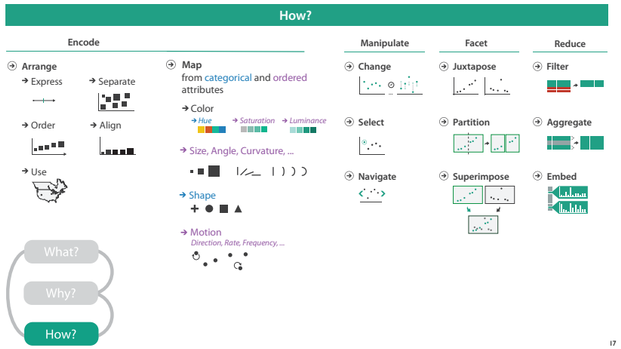
- Topology
- Paths



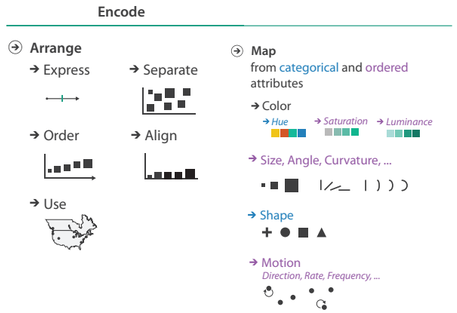
Spatial Data

- Shape

16



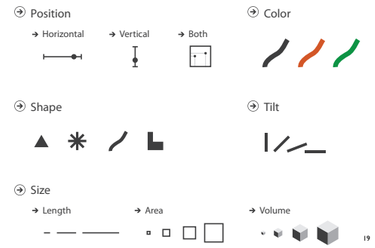
How to encode: Arrange space, map channels



Definitions: Marks and channels

- marks
  - geometric primitives

- channels
  - control appearance of marks

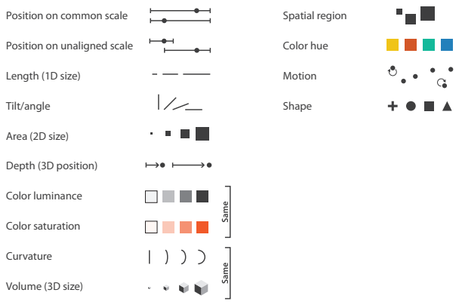


Encoding visually with marks and channels

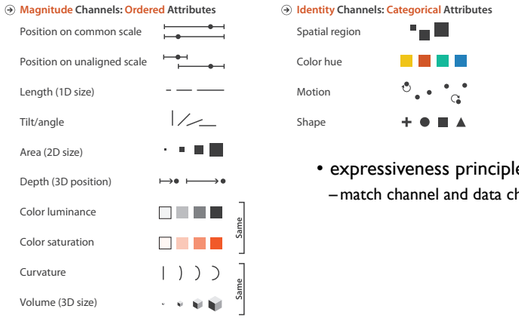
- analyze idiom structure
  - as combination of marks and channels

1:	2:	3:	4:
vertical position	vertical position horizontal position	vertical position horizontal position color hue	vertical position horizontal position color hue size (area)
mark: line	mark: point	mark: point	mark: point

Channels

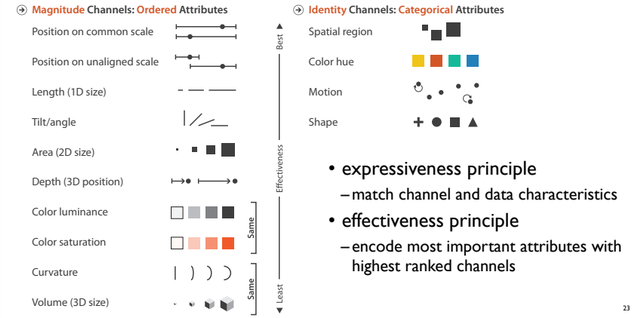


Channels: Matching Types

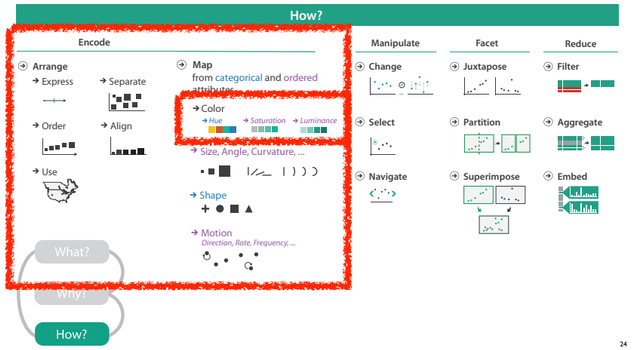


- expressiveness principle
  - match channel and data characteristics

Channels: Rankings

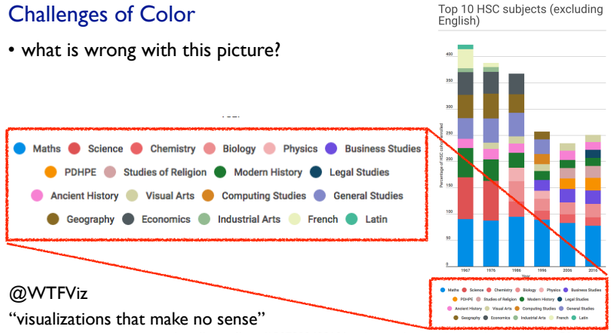


- expressiveness principle
  - match channel and data characteristics
- effectiveness principle
  - encode most important attributes with highest ranked channels



Challenges of Color

- what is wrong with this picture?

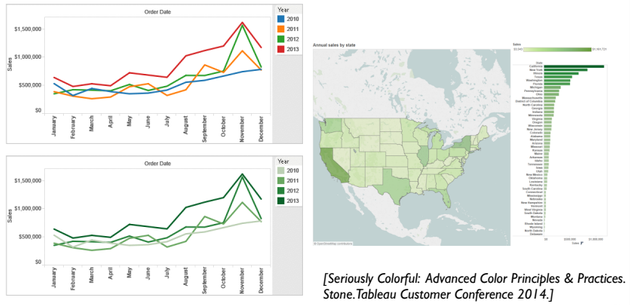


@WTFViz

"visualizations that make no sense"

http://viz.wtf/post/150780948819/maths-enrolments-drop-to-lowest-rate-in-50-years

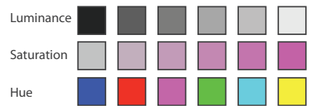
Categorical vs ordered color



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

Decomposing color

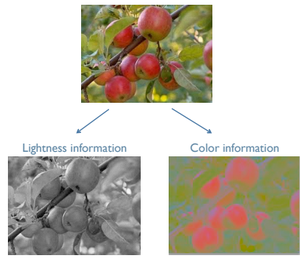
- first rule of color: do not talk about color!
  - color is confusing if treated as monolithic
- decompose into three channels
  - ordered can show magnitude
    - luminance
    - saturation
  - categorical can show identity
    - hue



- channels have different properties
  - what they convey directly to perceptual system
  - how much they can convey: how many discriminable bins can we use?

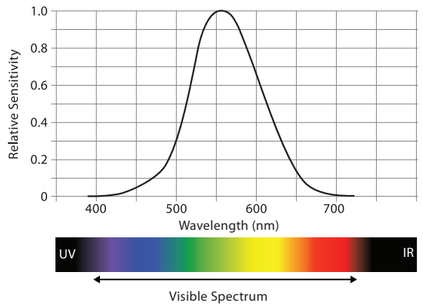
Luminance

- need luminance for edge detection
  - fine-grained detail only visible through luminance contrast
  - legible text requires luminance contrast!
- intrinsic perceptual ordering



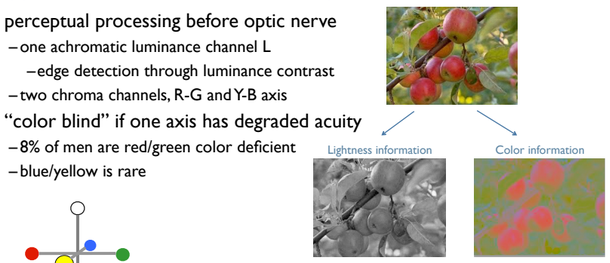
[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

Spectral sensitivity



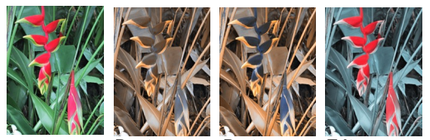
Opponent color and color deficiency

- perceptual processing before optic nerve
  - one achromatic luminance channel L
    - edge detection through luminance contrast
  - two chroma channels, R-G and Y-B axis
- "color blind" if one axis has degraded acuity
  - 8% of men are red/green color deficient
  - blue/yellow is rare

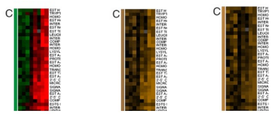


[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

Designing for color deficiency: Check with simulator



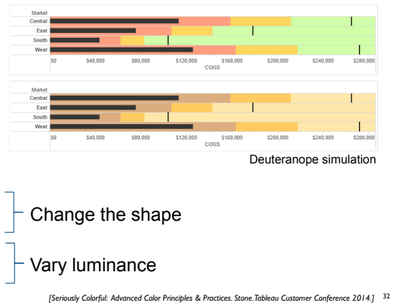
http://rehue.net



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

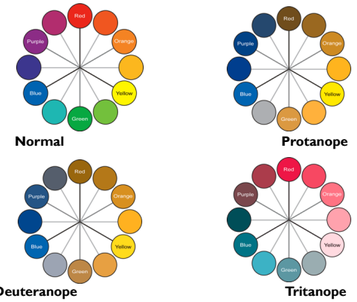
Designing for color deficiency: Avoid encoding by hue alone

- redundantly encode
  - vary luminance
  - change shape



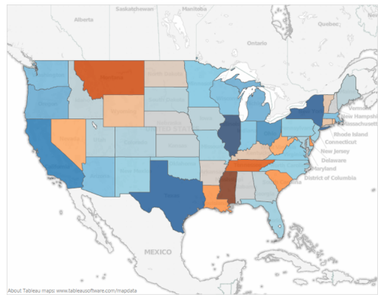
[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]

Color deficiency: Reduces color to 2 dimensions



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

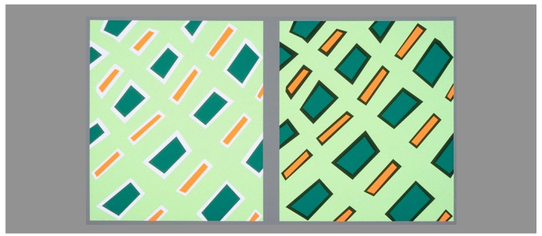
Designing for color deficiency: Blue-Orange is safe



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

Bezold Effect: Outlines matter

- color constancy: simultaneous contrast effect



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

Color/Lightness constancy: Illumination conditions

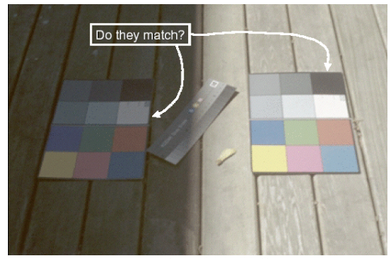


Image courtesy of John McCann

Color/Lightness constancy: Illumination conditions

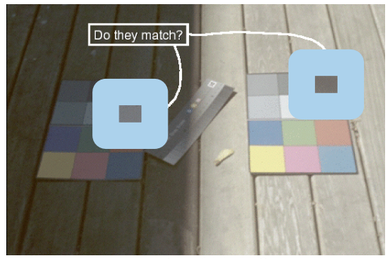
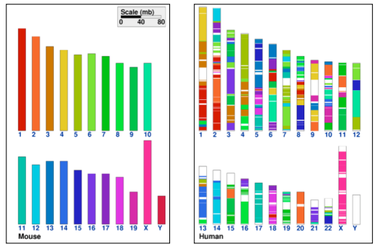


Image courtesy of John McCann

Categorical color: limited number of discriminable bins

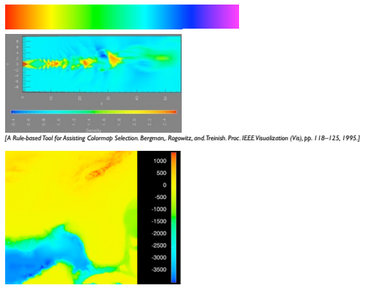
- human perception built on relative comparisons
  - great if color contiguous
  - surprisingly bad for absolute comparisons
- noncontiguous small regions of color
  - fewer bins than you want
  - rule of thumb: 6-12 bins, including background and highlights
  - alternatives? this afternoon!



[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
  - fine-grained structure visible and nameable



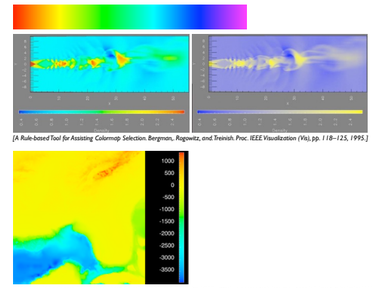
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

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

[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Ordered color: Rainbow is poor default

- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - fine-grained structure visible and nameable
- alternatives
  - large-scale structure: fewer hues



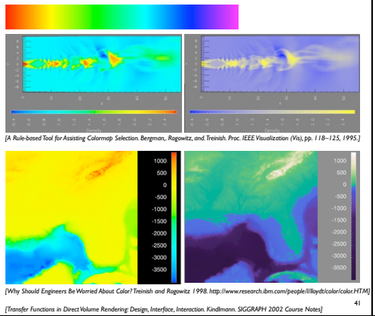
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

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

[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Ordered color: Rainbow is poor default

- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - fine-grained structure visible and nameable
- alternatives
  - large-scale structure: fewer hues
  - fine structure: multiple hues with monotonically increasing luminance [eg viridis R/python]



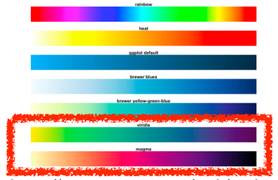
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

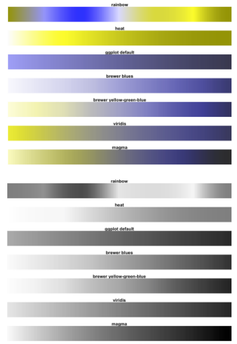
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Viridis

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

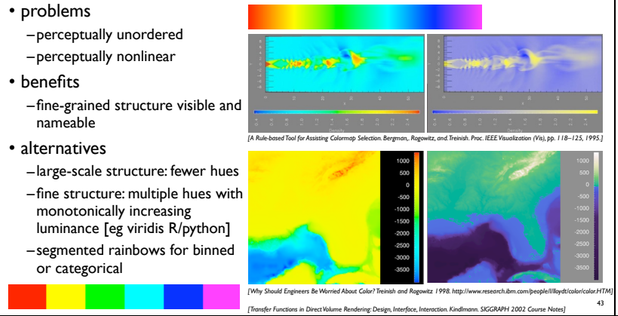




https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html

Ordered color: Rainbow is poor default

- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - fine-grained structure visible and nameable
- alternatives
  - large-scale structure: fewer hues
  - fine structure: multiple hues with monotonically increasing luminance [eg viridis R/python]
  - segmented rainbows for binned or categorical



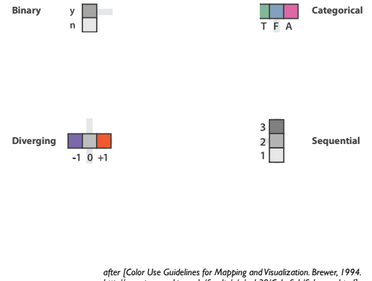
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

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

[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Colormaps

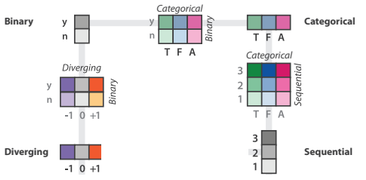
- Categorical
- Ordered
  - Sequential
  - Diverging
- Bivariate



after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]

Colormaps

- Categorical
- Ordered
  - Sequential
  - Diverging
- Bivariate

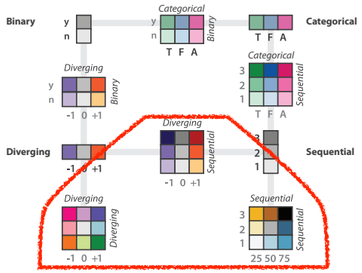


after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]

Colormaps

- Categorical
- Ordered
  - Sequential
  - Diverging
- Bivariate

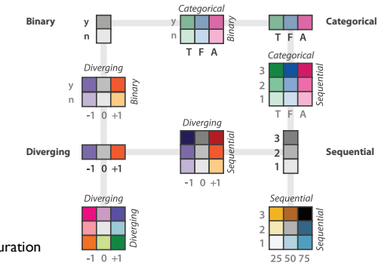
use with care!



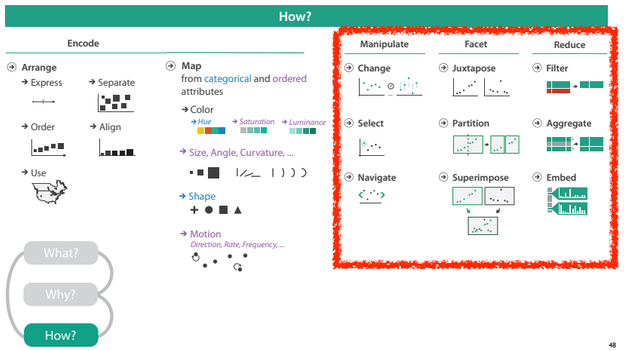
after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]

Colormaps

- Categorical
- Ordered
  - Sequential
  - Diverging
- Bivariate
  - color channel interactions
    - size heavily affects salience
      - small regions need high saturation
      - large need low saturation
    - saturation & luminance: 3-4 bins max
      - also not separable from transparency



after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]



How to handle complexity: 3 more strategies + 1 previous

Manipulate: Change, Select, Navigate
Facet: Juxtapose, Partition, Superimpose
Reduce: Filter, Aggregate, Embed
Derive

- 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

Manipulate: Change, Select, Navigate
Facet: Juxtapose, Partition, Superimpose
Reduce: Filter, Aggregate, Embed
Derive

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

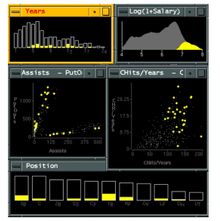
How to handle complexity: 3 more strategies + 1 previous

Manipulate: Change, Select, Navigate
Facet: Juxtapose, Partition, Superimpose
Reduce: Filter, Aggregate, Embed
Derive

- 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.]

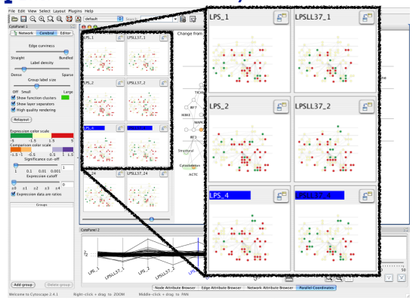
Idiom: **bird's-eye maps** System: **Google Maps**

- 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.]

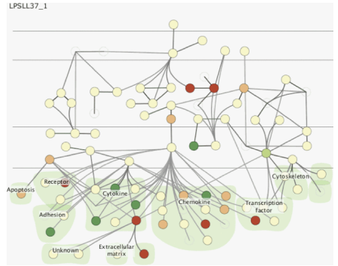
Coordinate views: Design choice interaction

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



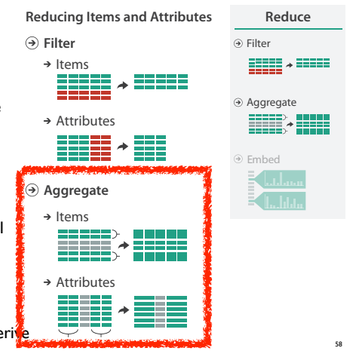
How to handle complexity: 3 more strategies + 1 previous

Manipulate: Change, Select, Navigate
Facet: Juxtapose, Partition, Superimpose
Reduce: Filter, Aggregate, Embed
Derive

- reduce what is shown within single view

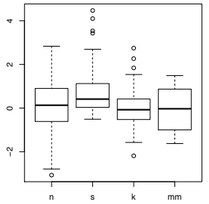
Reduce items and attributes

- 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**

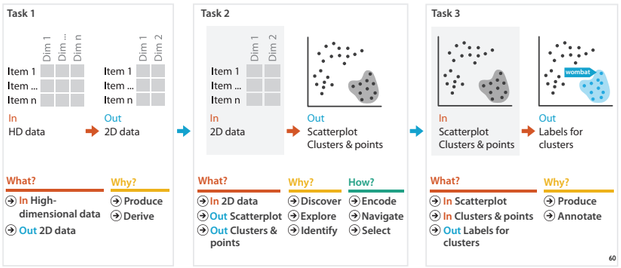
- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown

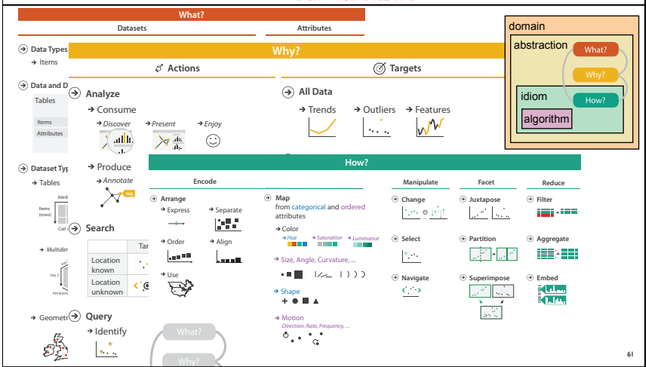


[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

Idiom: **Dimensionality reduction for documents**

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





A quick taste of other work!

technique-driven work

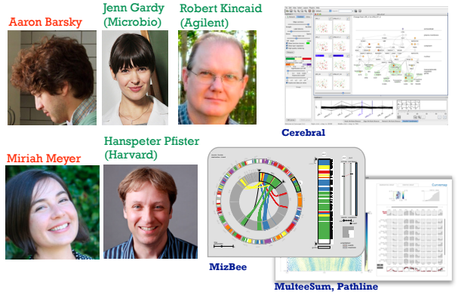
problem-driven work

theoretical foundations

evaluation

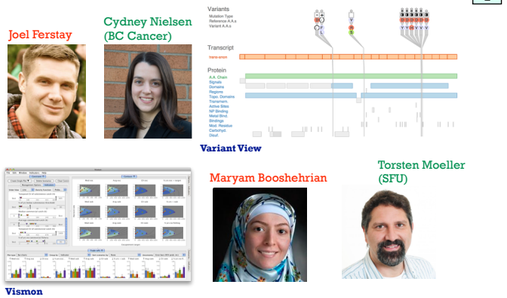
Problem-driven: Genomics

T F E P



Problem-driven: Genomics, fisheries

T F E P



Problem-driven: Tech industry

T F E P

SessionViewer: web log analysis

Heidi Lam

Diane Tang (Google)

Peter McLachlan

Stephen North (AT&T Research)

LiveRAC: systems time-series

Problem-driven: Journalism

T F E P

Matt Brehmer

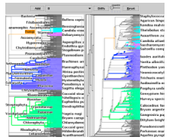
Stephen Ingram

Jonathan Stray (Assoc Press)

Overview

Technique-driven: Graph drawing

T F E P



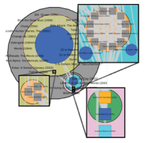
TreeJuxtaposer

James Slack

Kristian Hildebrand

Daniel Archambault

David Auber (Bordeaux)



TopoLayout
SPF
Grouse
GrouseFlocks
TugGraph

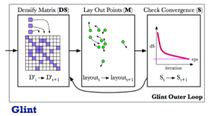
Technique-driven: Dimensionality reduction

T F E P

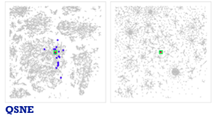
Stephen Ingram

Glimmer

DimStiller



Glint

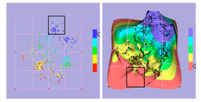


QSNE

Evaluation: Dimensionality reduction

T F E P

Melanie Tory

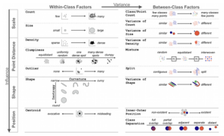


Points vs landscapes for dimensionally reduced data

Guidance on DR & scatterplot choices

Michael Sedlmair

Melanie Tory (UVic)



Taxonomy of cluster separation factors

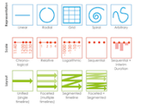
Curation & Presentation: Timelines

T F E P

TimeLineCurator
https://vimeo.com/123246662

Johanna Fulda (Sud. Zeitung)

Matt Brehmer



Timelines Revisited
timelinesrevisited.github.io/

Bongshin Lee (Microsoft)

Benjamin Bach (Microsoft)

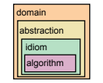
Nathalie Henry-Riche (Microsoft)

Theoretical foundations

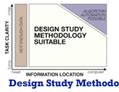
T F E P

- 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

Visualization Analysis & Design

Geometry Center 1990-1995

Geomview

The Shape of Space

Outside In

Charlie Gunn

Stuart Levy

Mark Phillips

Delle Maxwell

More Information

@tamaramunzner

- this talk
  www.cs.ubc.ca/~tmm/talks.html#vad17bedford
- 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.