

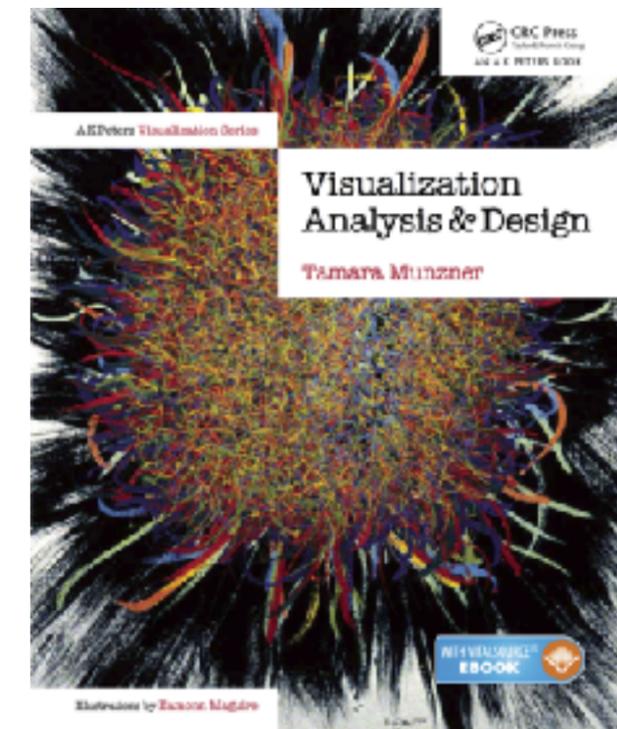
Visualization Analysis & Design

Tamara Munzner

Department of Computer Science
University of British Columbia

*Bio+Med+Vis Spring School
May 19 2021, virtual*

<http://www.cs.ubc.ca/~tmm/talks.html#vad21biomedvis>



[@tamaramunzner](https://twitter.com/tamaramunzner)

Visualization: definition & motivation

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

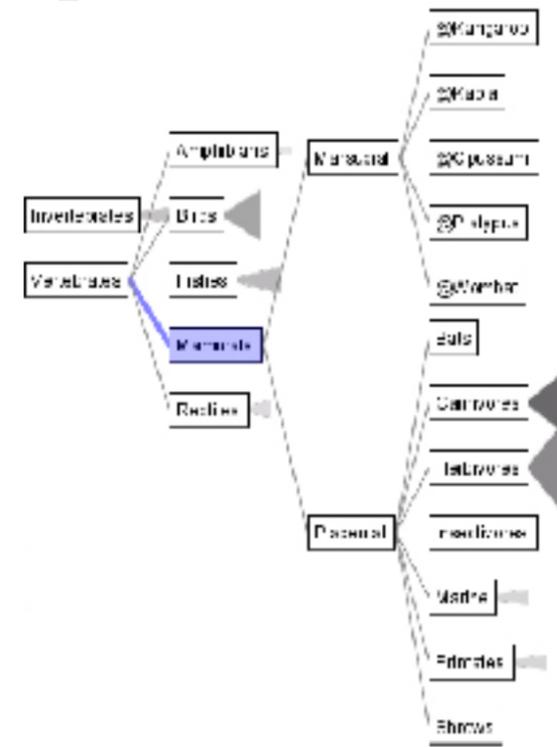
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- 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

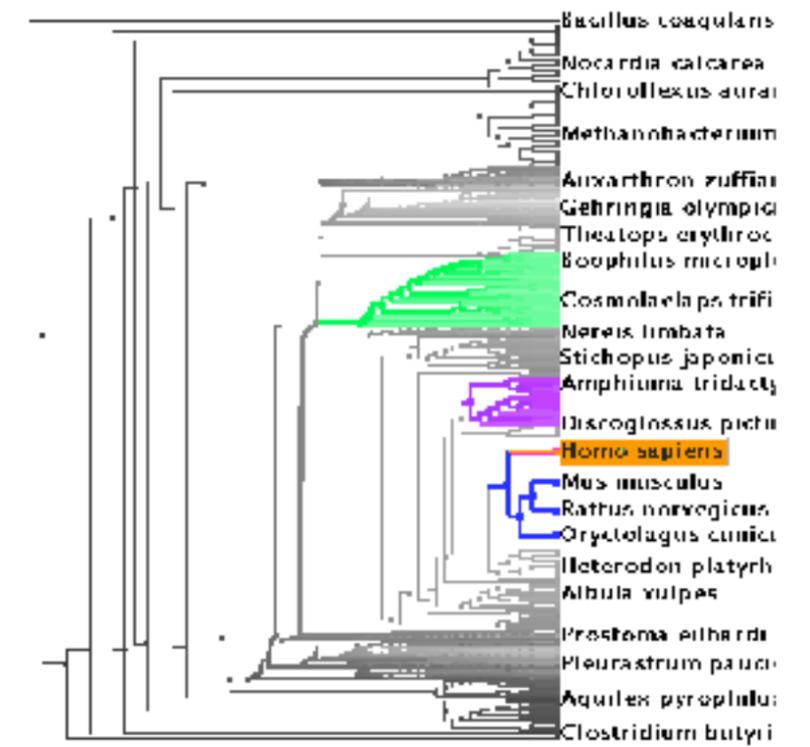
Why analyze?

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

SpaceTree



TreeJuxtaposer



[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



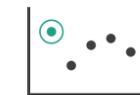
→ Targets

- Path between two nodes



→ SpaceTree

- Encode
- Navigate
- Select
- Filter
- Aggregate



→ TreeJuxtaposer

- Encode
- Navigate
- Select
- Arrange



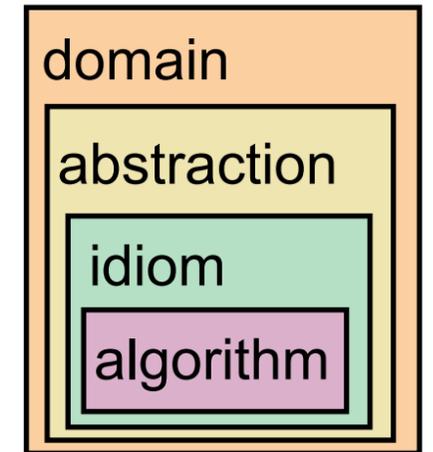
What?

Why?

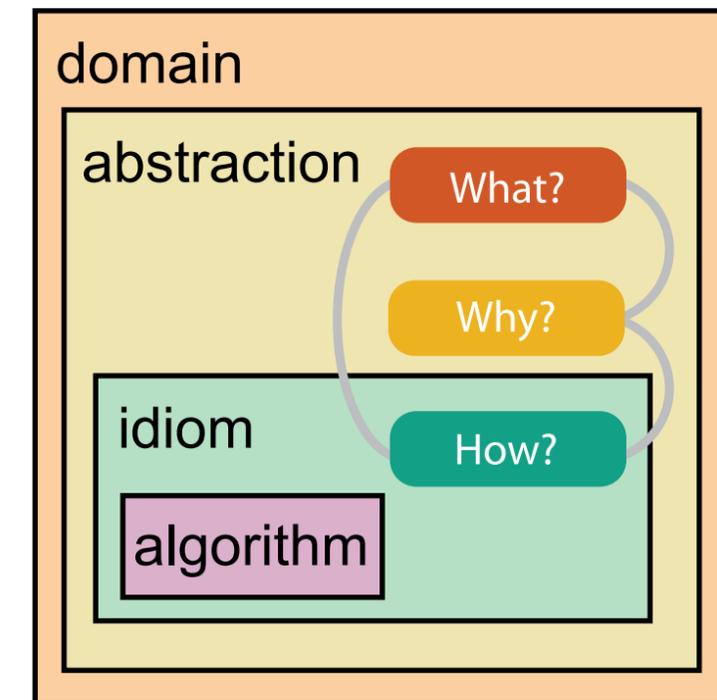
How?

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**
- **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



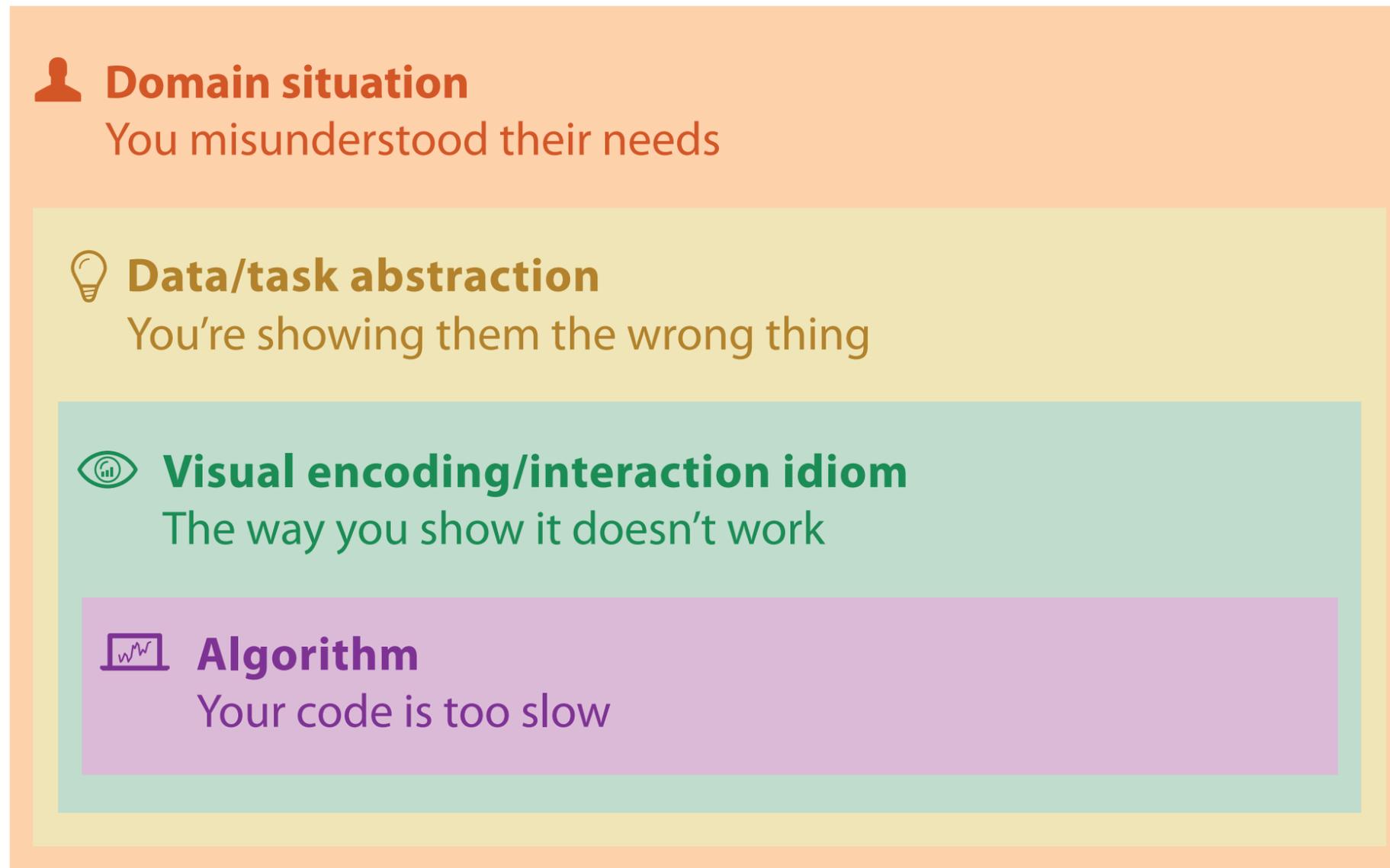
[A Nested Model of Visualization Design and Validation. Munzner. *IEEE TVCG* 15(6):921-928, 2009 (Proc. InfoVis 2009).]



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

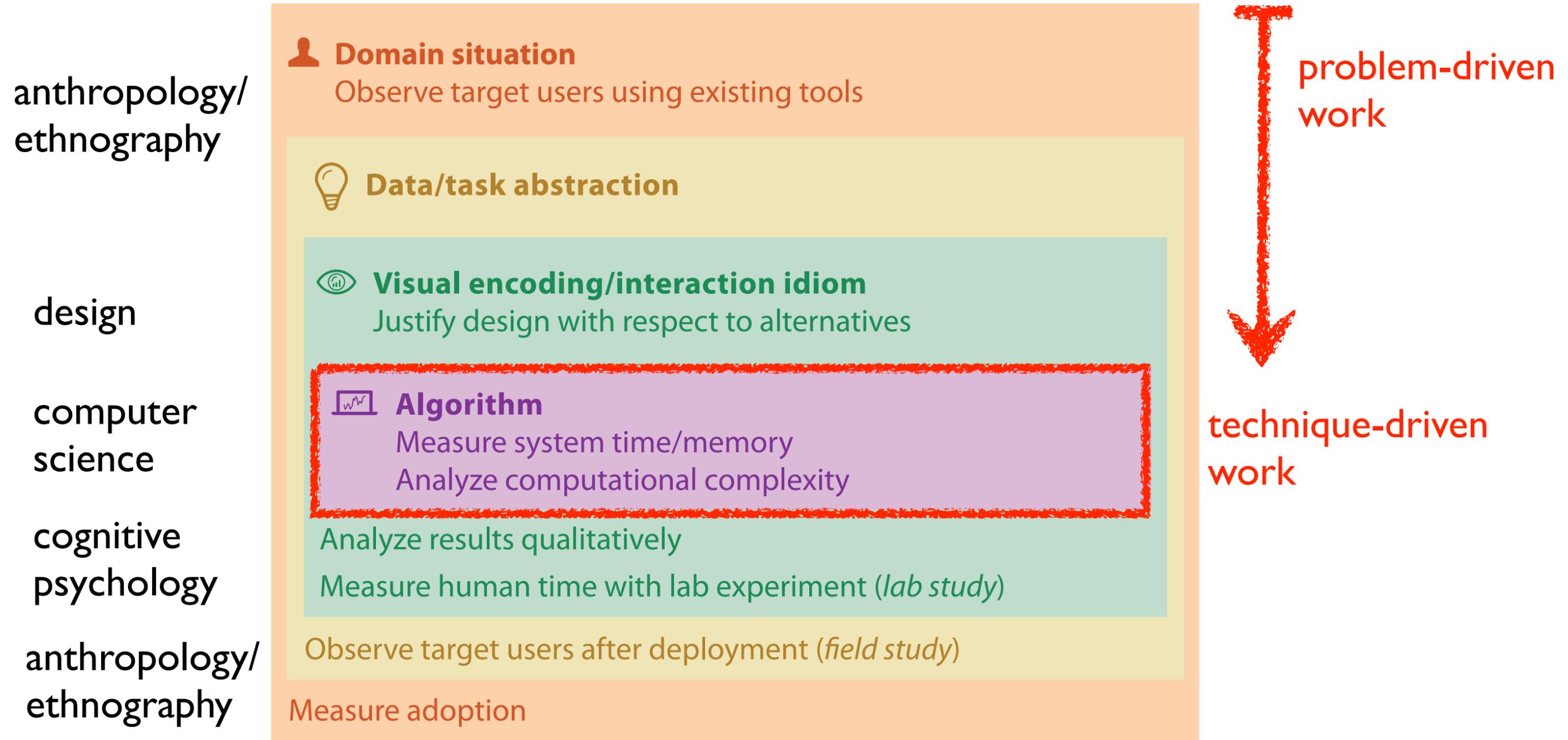
Why is validation difficult?

- different ways to get it wrong at each level

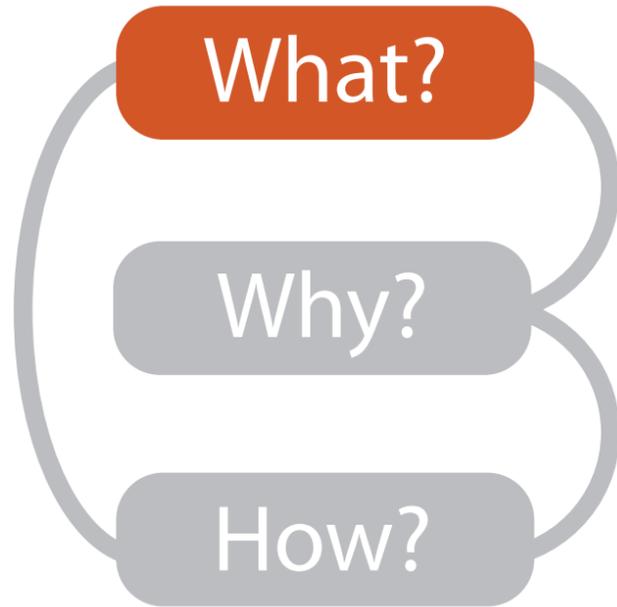


Why is validation difficult?

- 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).]



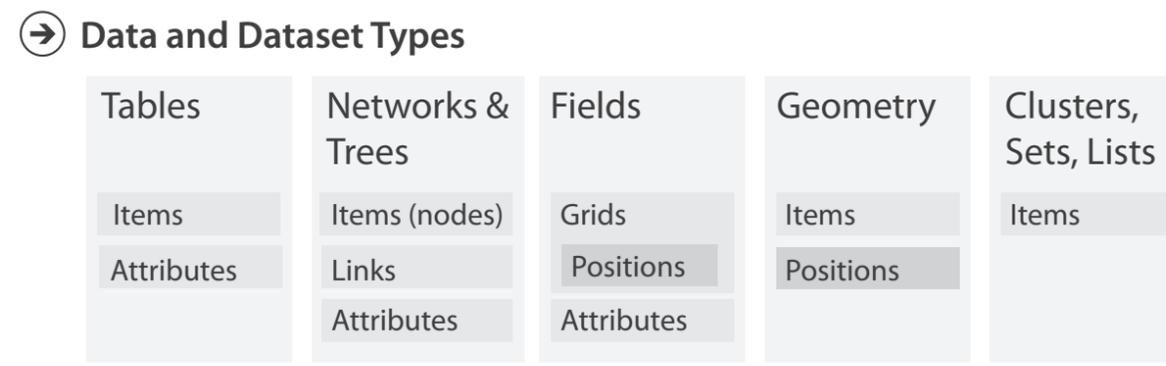
What?

Datasets

Attributes

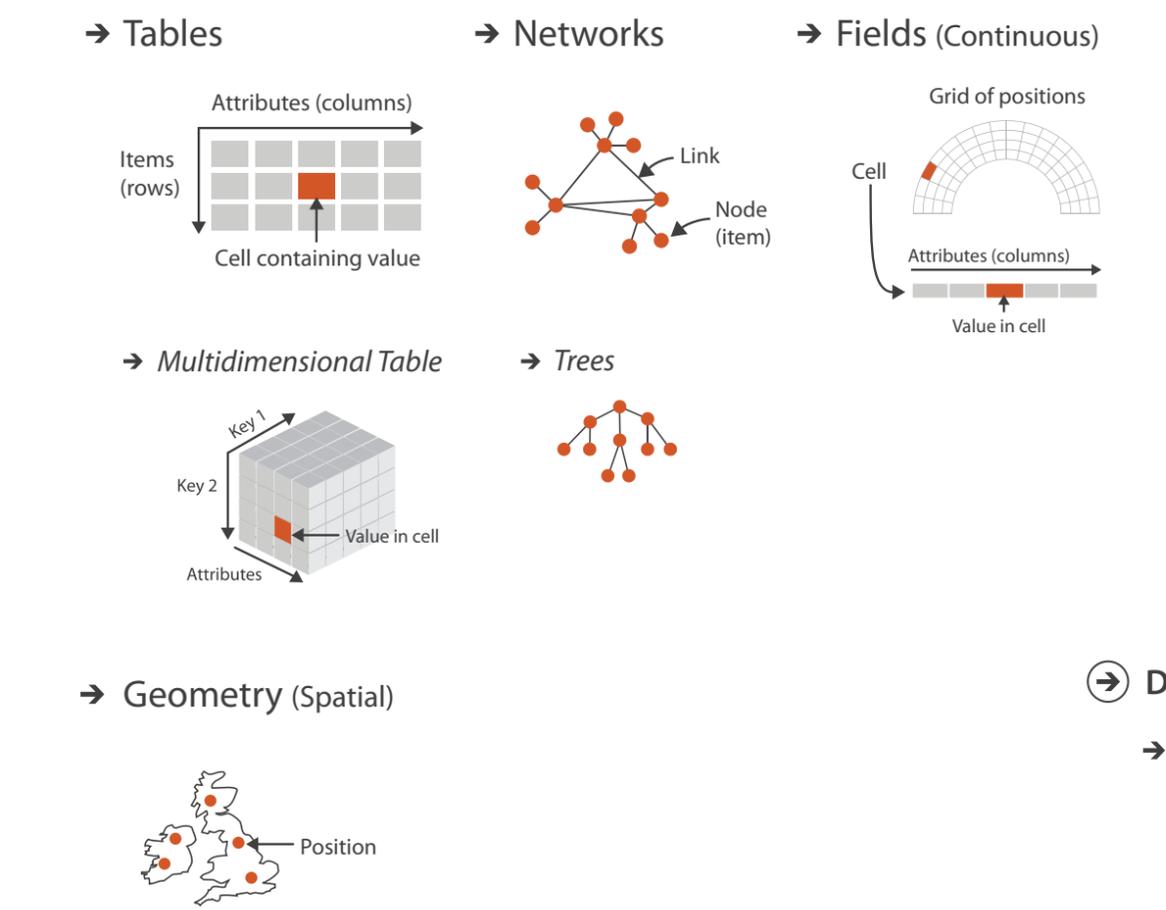
- Data Types
 - Items
 - Attributes
 - Links
 - Positions
 - Grids

- Attribute Types
 - Categorical

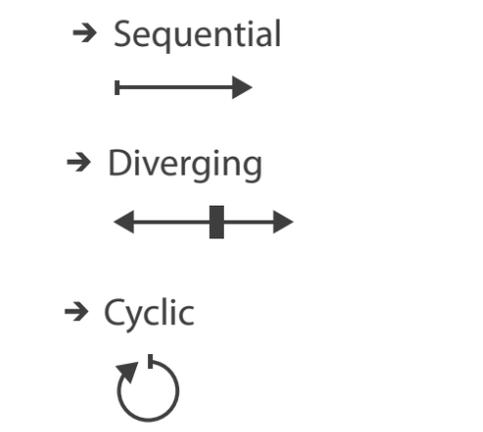


- Categorical
 - + ● ■ ▲
- Ordered
 - Ordinal
 - ↑ ↑↑ ↑↑↑
 - Quantitative
 - | —|—| —|—|—|

Dataset Types



Ordering Direction



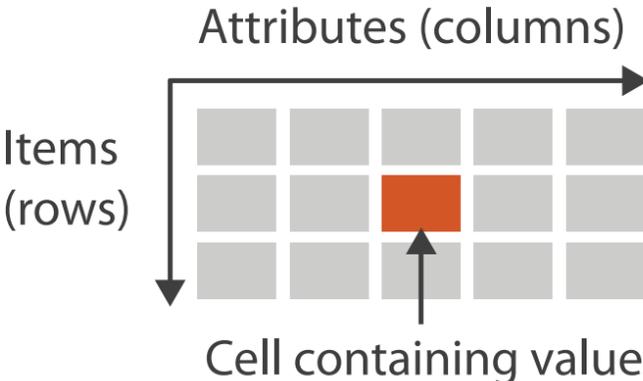
Dataset Availability



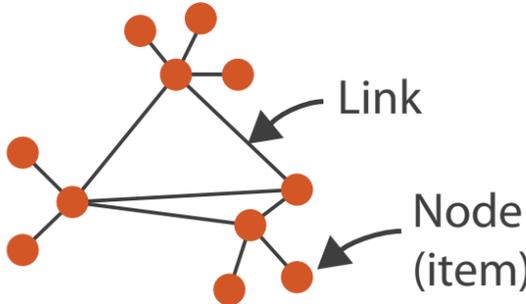
Types: Datasets and data

→ Dataset Types

→ Tables

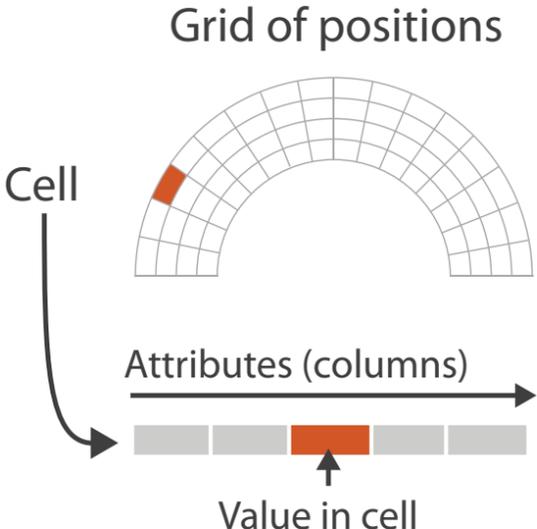


→ Networks

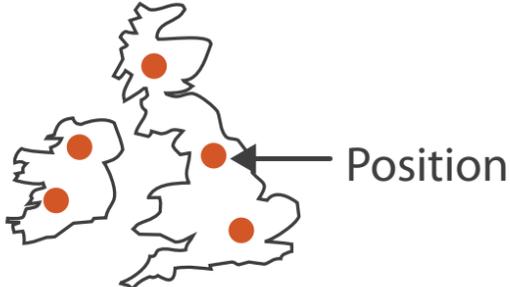


→ Spatial

→ Fields (Continuous)



→ Geometry (Spatial)



→ Attribute Types

→ Categorical

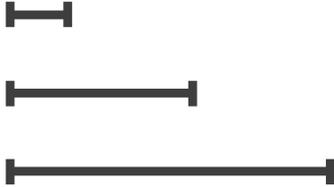


→ Ordered

→ Ordinal



→ Quantitative



👉 **Actions**

🎯 **Targets**

➔ **Analyze**

- ➔ Consume
 - ➔ Discover
 - ➔ Present
 - ➔ Enjoy
- ➔ Produce
 - ➔ Annotate
 - ➔ Record
 - ➔ Derive

➔ **All Data**

- ➔ Trends
- ➔ Outliers
- ➔ Features

➔ **Attributes**

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

➔ **Search**

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

➔ **Network Data**

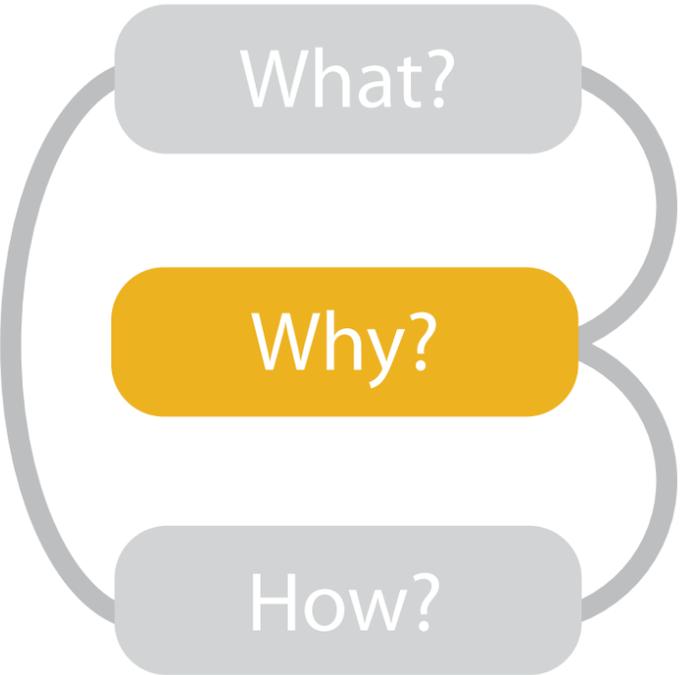
- ➔ Topology
- ➔ Paths

➔ **Query**

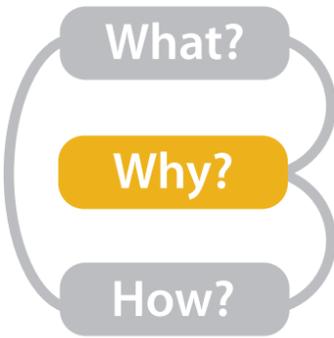
- ➔ Identify
- ➔ Compare
- ➔ Summarize

➔ **Spatial Data**

- ➔ Shape



- {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

- analyze, query, (search)

→ Analyze

- Consume

- Discover



- Present

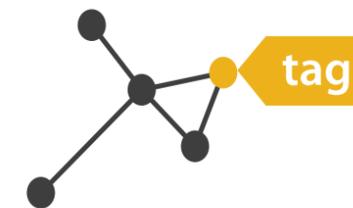


- Enjoy



- Produce

- Annotate



- Record

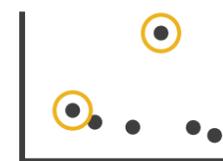


- Derive



→ Query

- Identify



- Compare

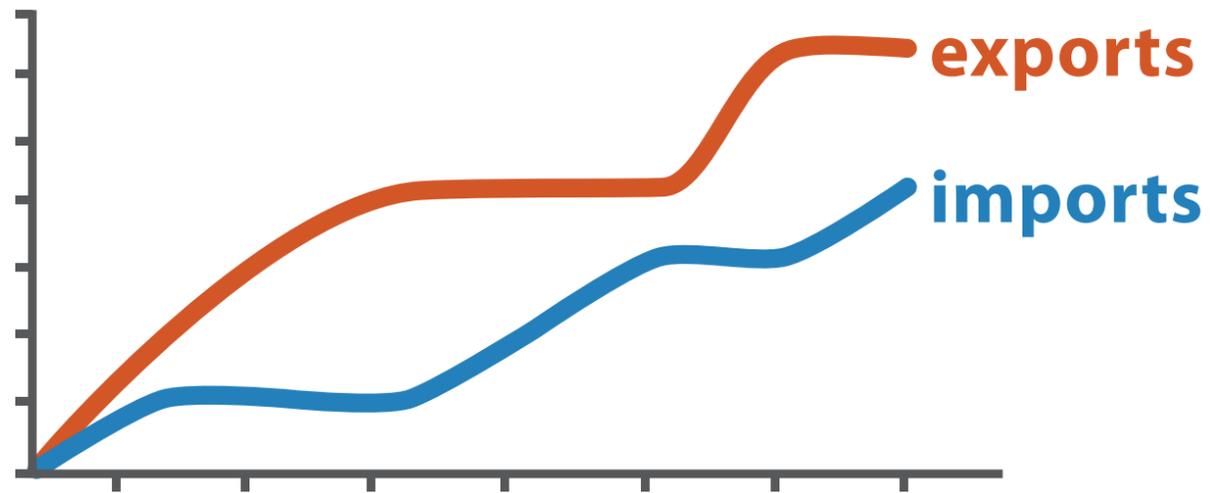


- Summarize

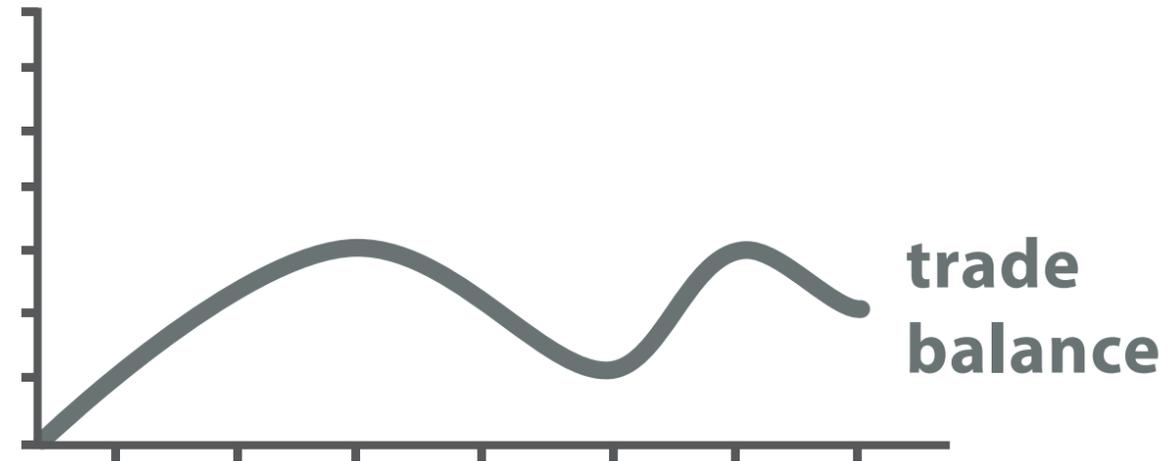


Derive: Crucial Design Choice

- 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



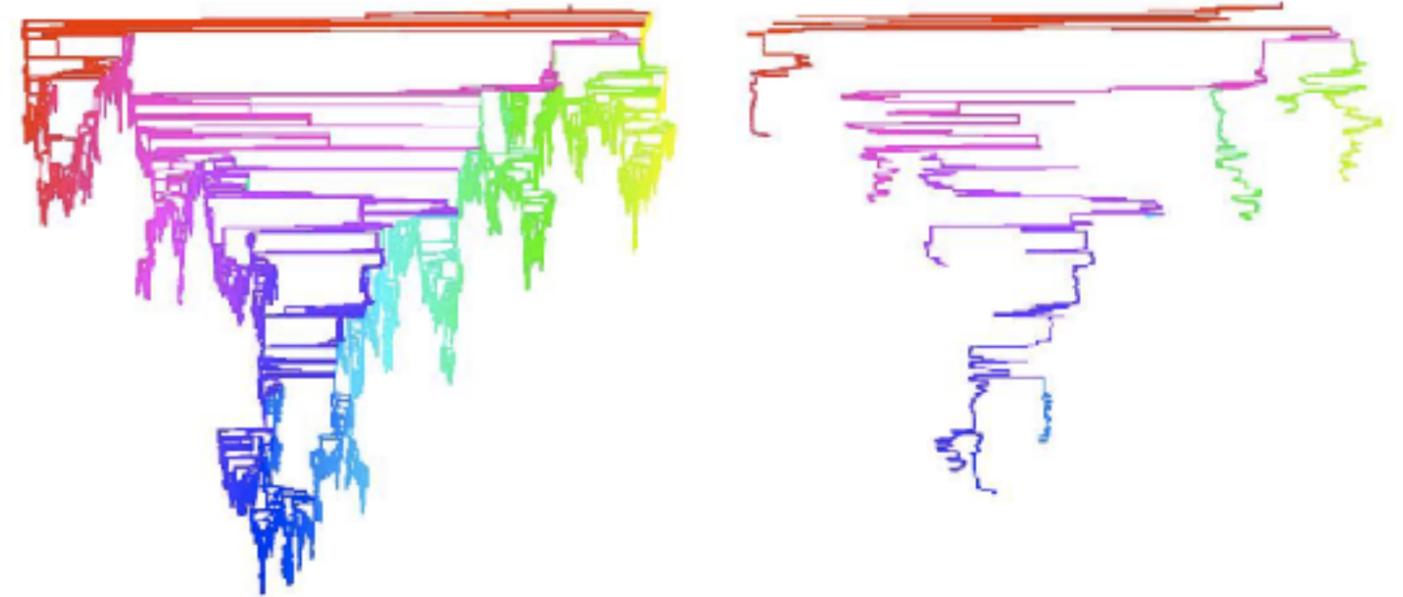
$$\text{trade balance} = \text{exports} - \text{imports}$$

Derived Data

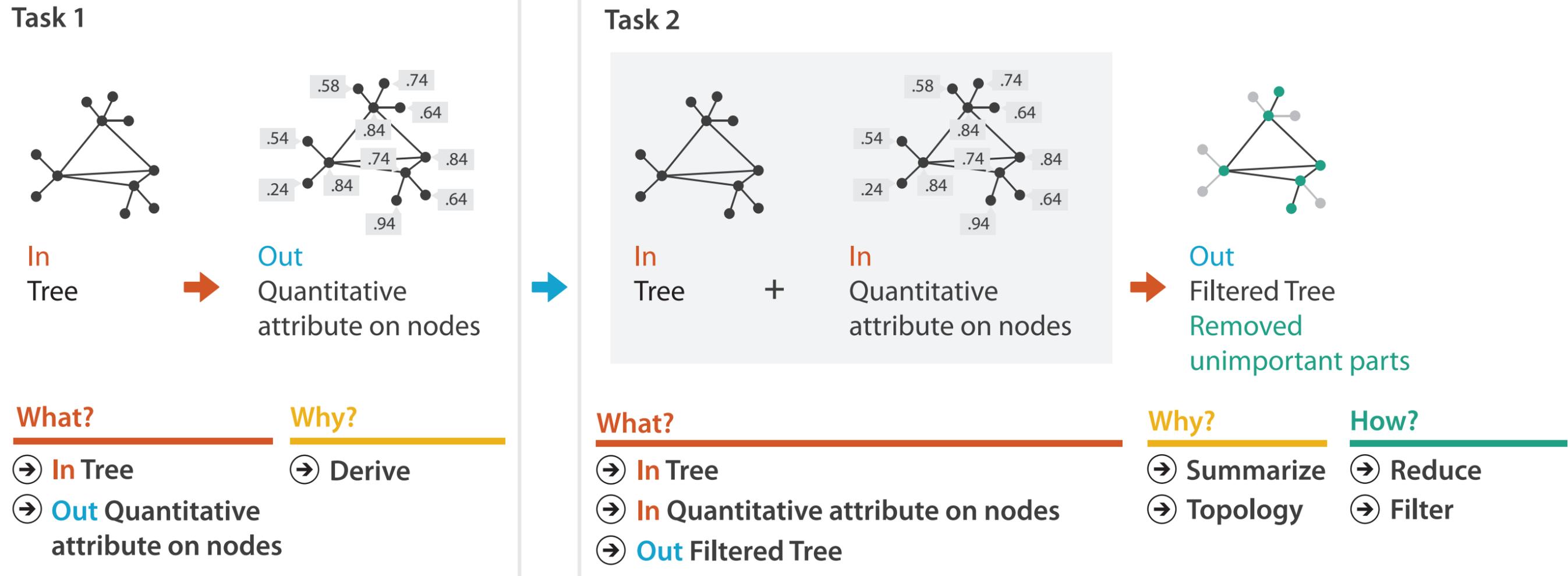
Analysis example: Derive one attribute

- 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

→ All Data

→ Trends



→ Outliers



→ Features



→ Attributes

→ One

→ *Distribution*



→ *Extremes*



→ Many

→ *Dependency*



→ *Correlation*

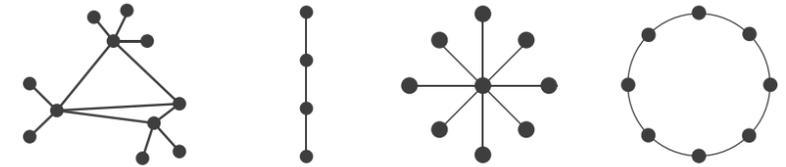


→ *Similarity*



→ Network Data

→ Topology

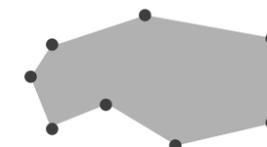


→ *Paths*



→ Spatial Data

→ Shape



How?

Encode

→ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



→ Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



→ Motion

Direction, Rate, Frequency, ...



Manipulate

→ Change



→ Select



→ Navigate

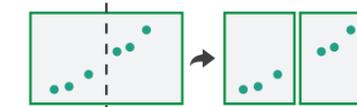


Facet

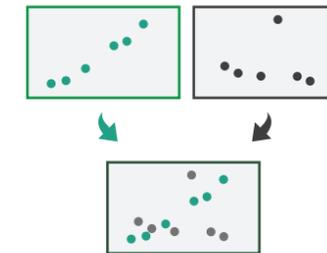
→ Juxtapose



→ Partition

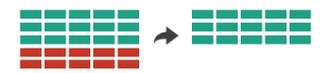


→ Superimpose



Reduce

→ Filter



→ Aggregate



→ Embed



What?

Why?

How?

How to encode: Arrange space, map channels

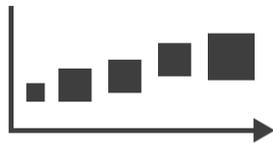
Encode

② Arrange

→ Express



→ Order



→ Use



→ Separate



→ Align



② Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



→ Motion

Direction, Rate, Frequency, ...

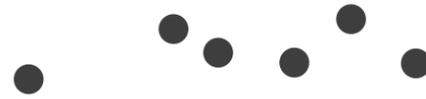


Definitions: Marks and channels

- marks

- geometric primitives
- one per item

→ Points



→ Lines



→ Areas



- channels

- control appearance of marks

→ Position

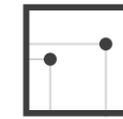
→ Horizontal



→ Vertical



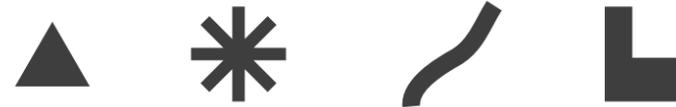
→ Both



→ Color



→ Shape



→ Tilt



→ Size

→ Length

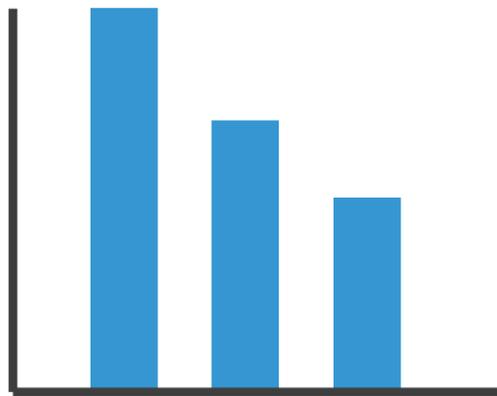
→ Area

→ Volume



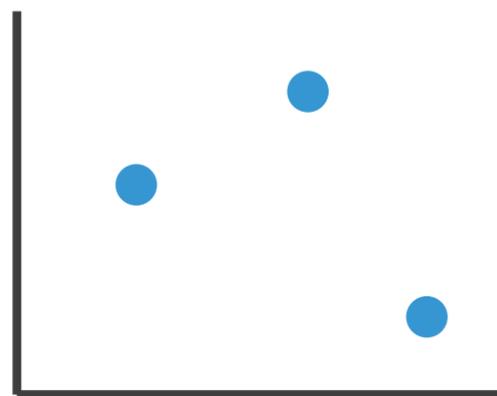
Encoding visually with marks and channels

- analyze idiom structure
 - as combination of marks and channels



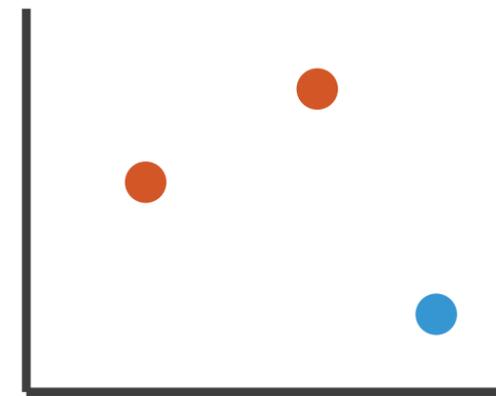
1:
vertical position

mark: line



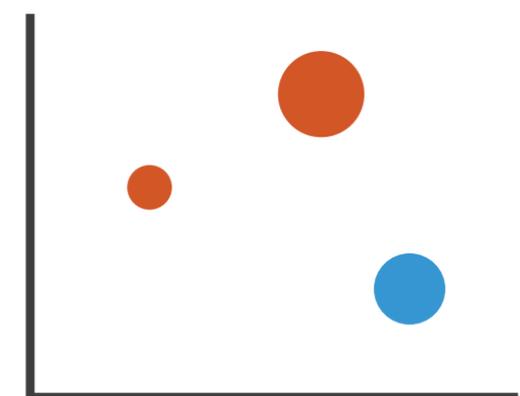
2:
vertical position
horizontal position

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



Same

Same

Channels: Matching Types

➔ Magnitude Channels: Ordered Attributes

Position on common scale 

Position on unaligned scale 

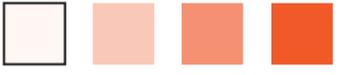
Length (1D size) 

Tilt/angle 

Area (2D size) 

Depth (3D position) 

Color luminance 

Color saturation 

Curvature 

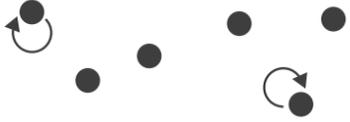
Volume (3D size) 

Same
Same

➔ Identity Channels: Categorical Attributes

Spatial region 

Color hue 

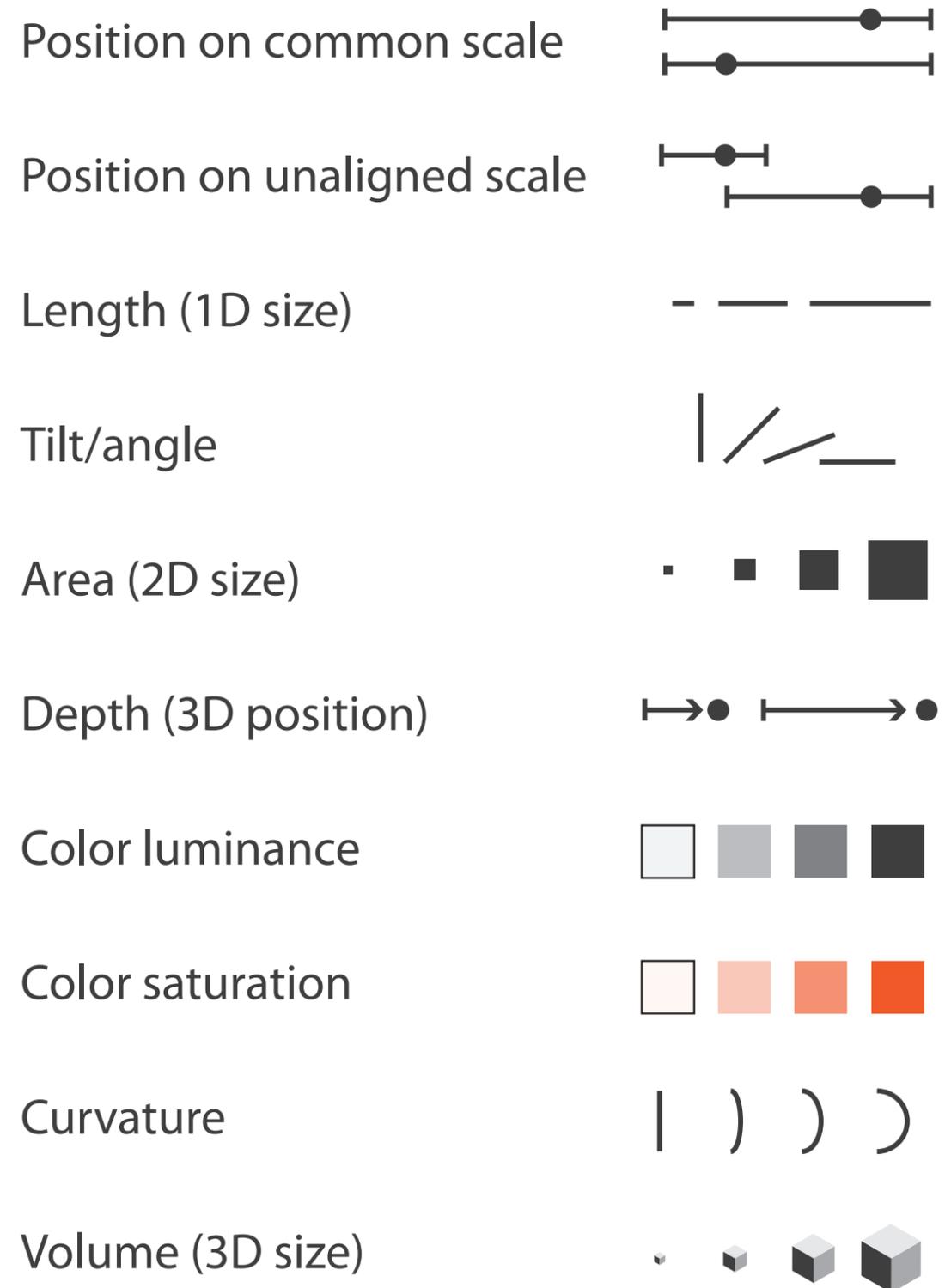
Motion 

Shape 

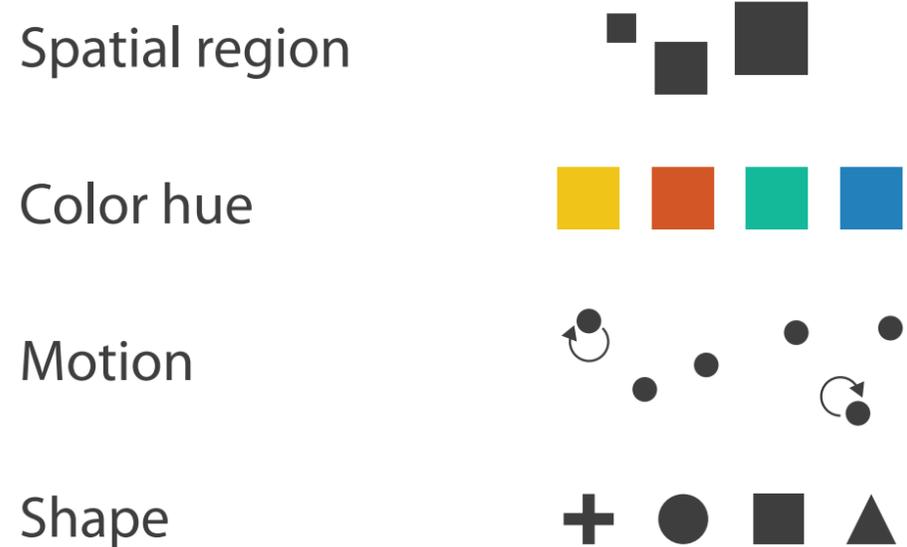
- **expressiveness principle**
 - match channel and data characteristics

Channels: Rankings

➔ Magnitude Channels: Ordered Attributes



➔ Identity Channels: Categorical Attributes



Best

Effectiveness

Least

Same

Same

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

How?

Encode

→ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



→ Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



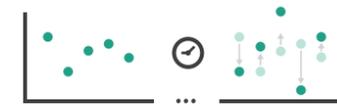
→ Motion

Direction, Rate, Frequency, ...



Manipulate

→ Change



→ Select



→ Navigate

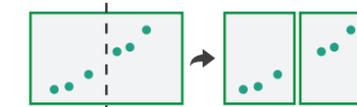


Facet

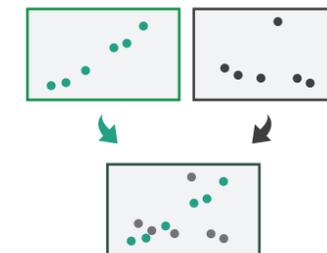
→ Juxtapose



→ Partition

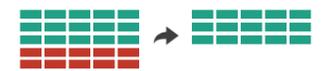


→ Superimpose



Reduce

→ Filter



→ Aggregate



→ Embed

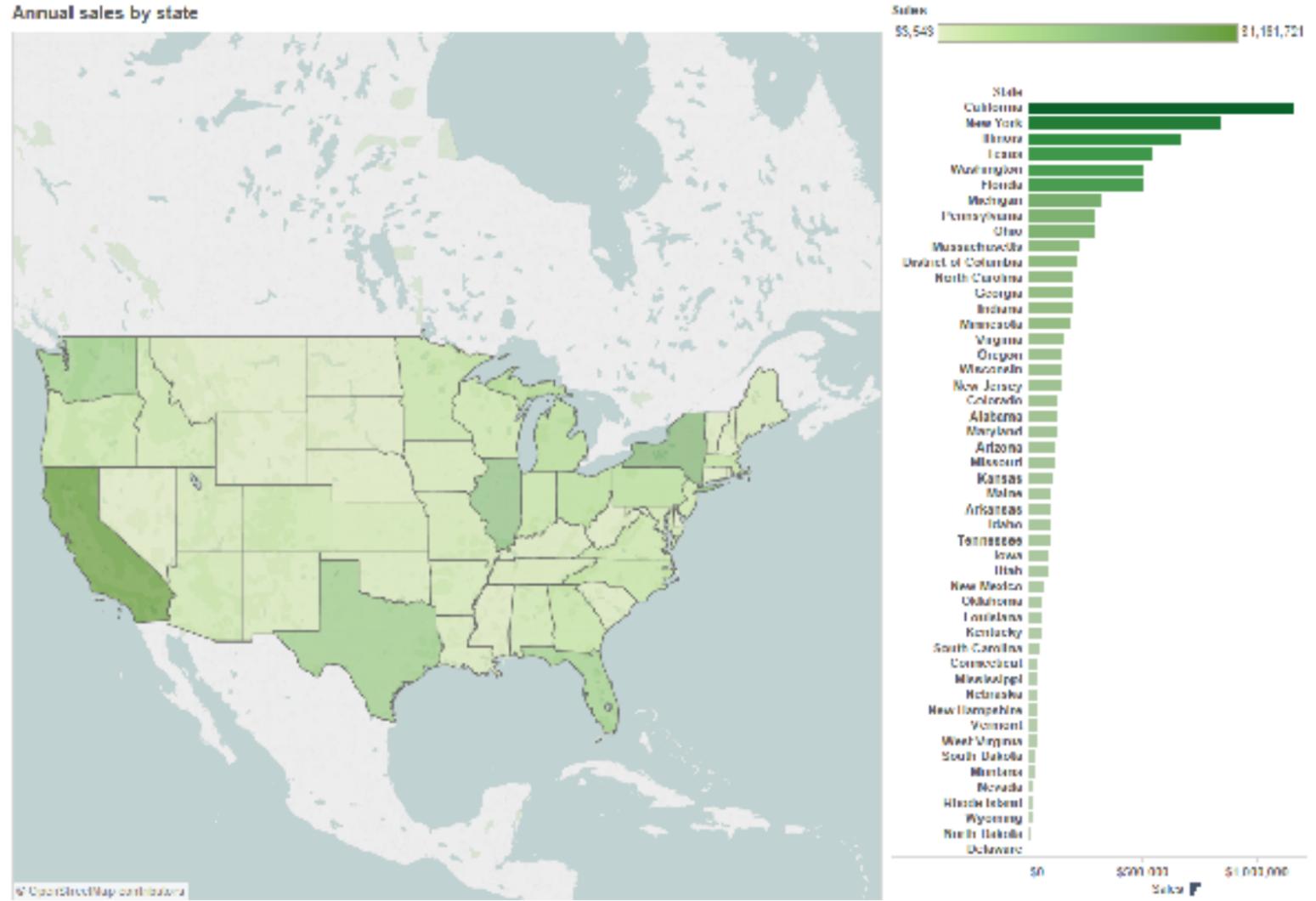
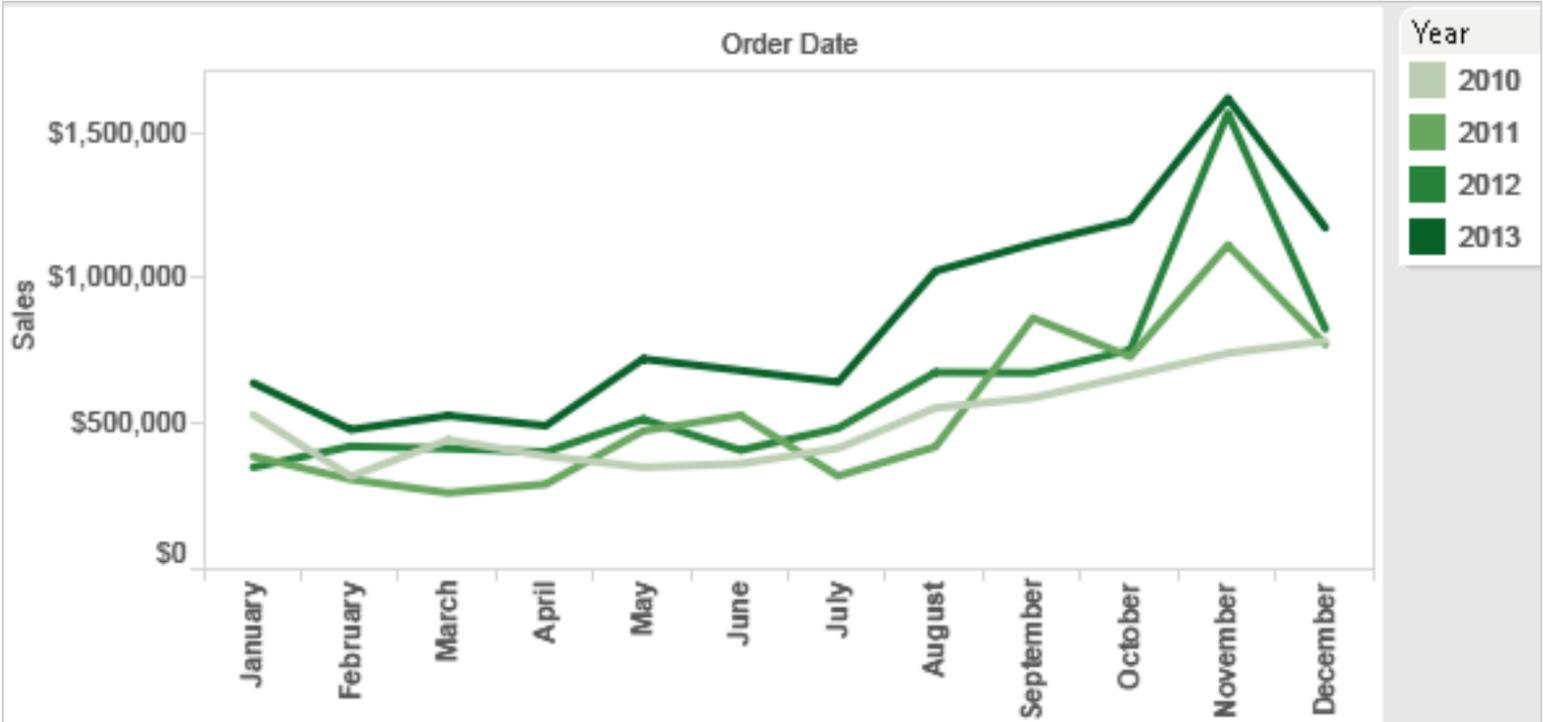
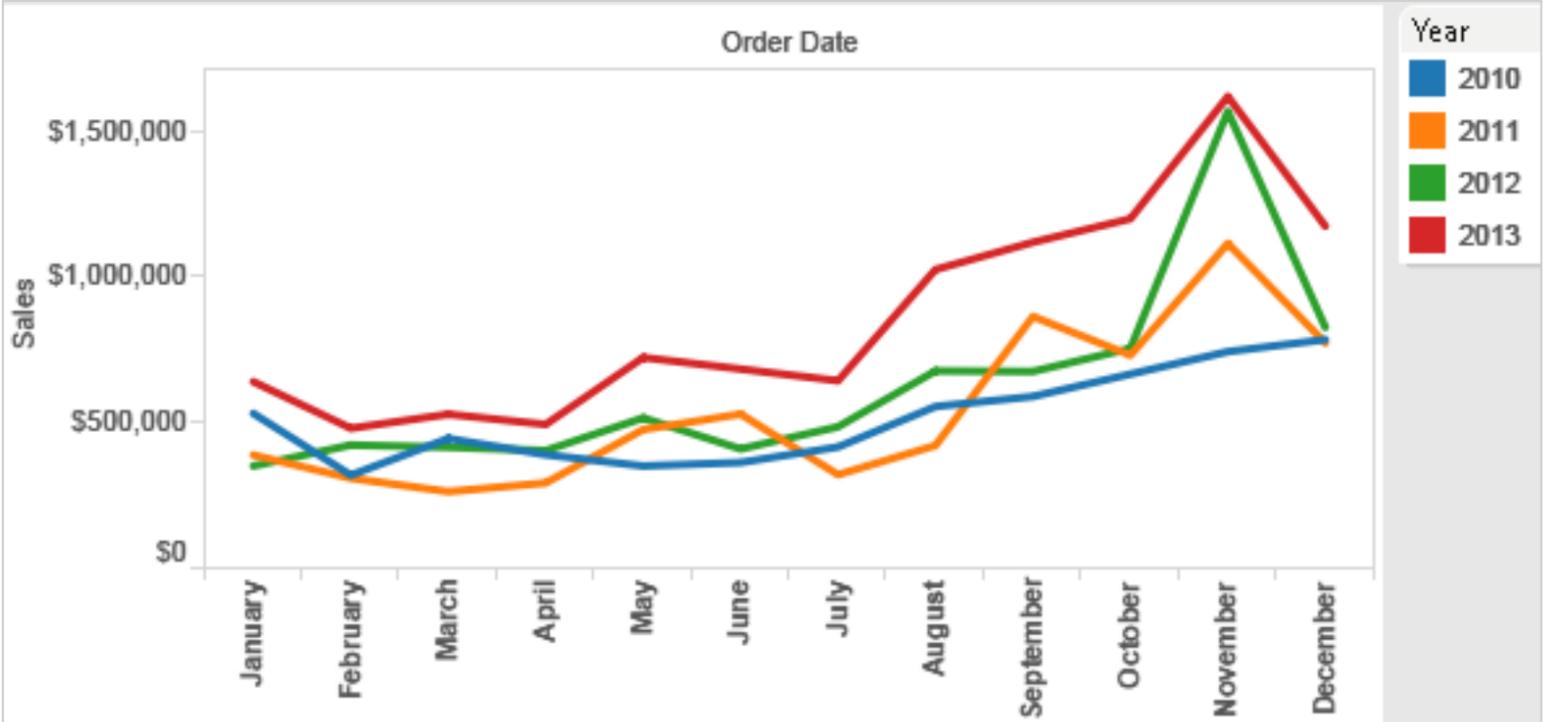


What?

Why?

How?

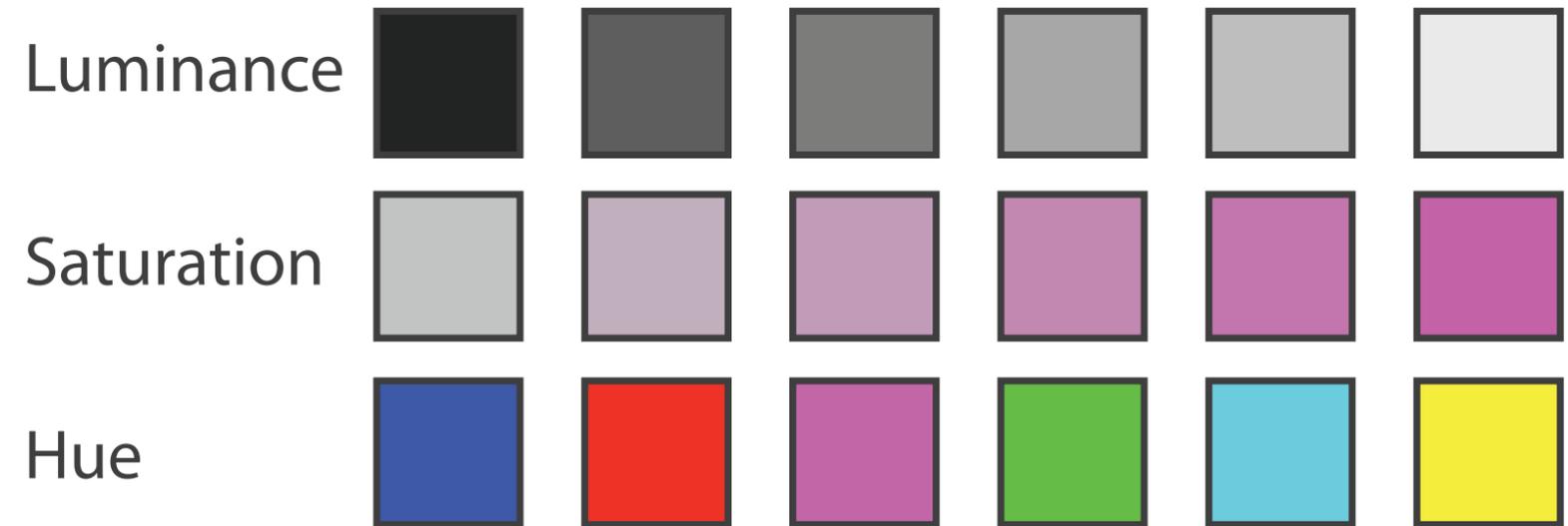
Categorical vs ordered color



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

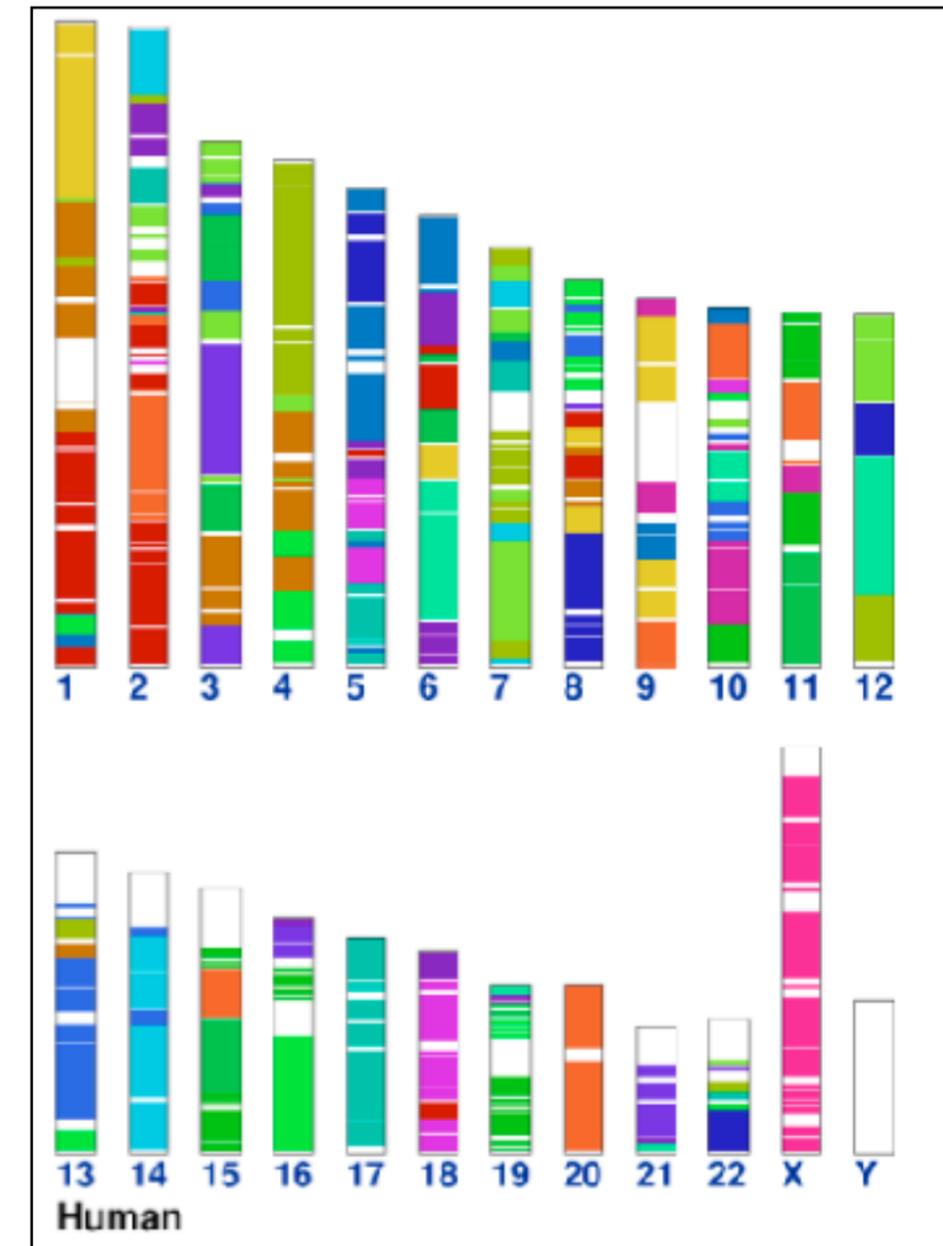
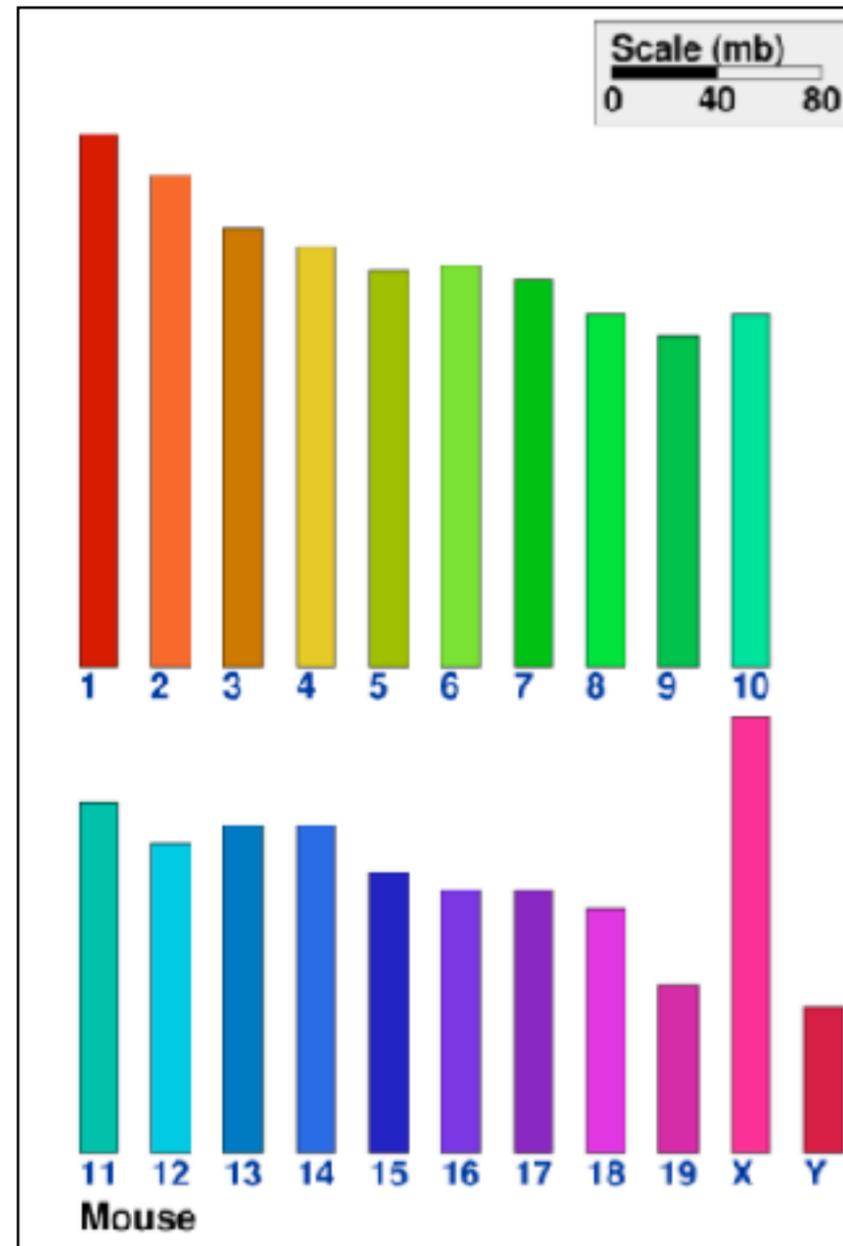
Decomposing color

- 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

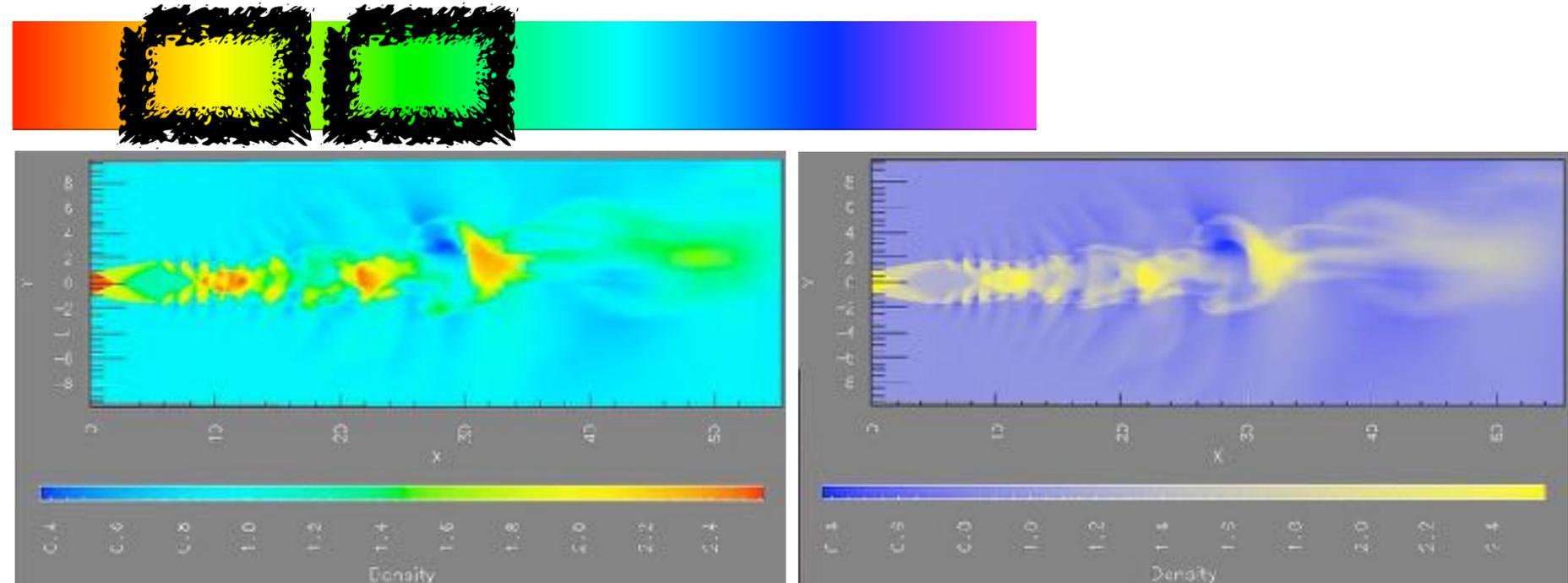
- human perception built on relative comparisons
 - great if color contiguous
 - surprisingly bad for absolute comparisons
- noncontiguous small regions of color
 - fewer bins than you want
 - rule of thumb: 6-12 bins, including background and highlights
- alternatives? other talks!



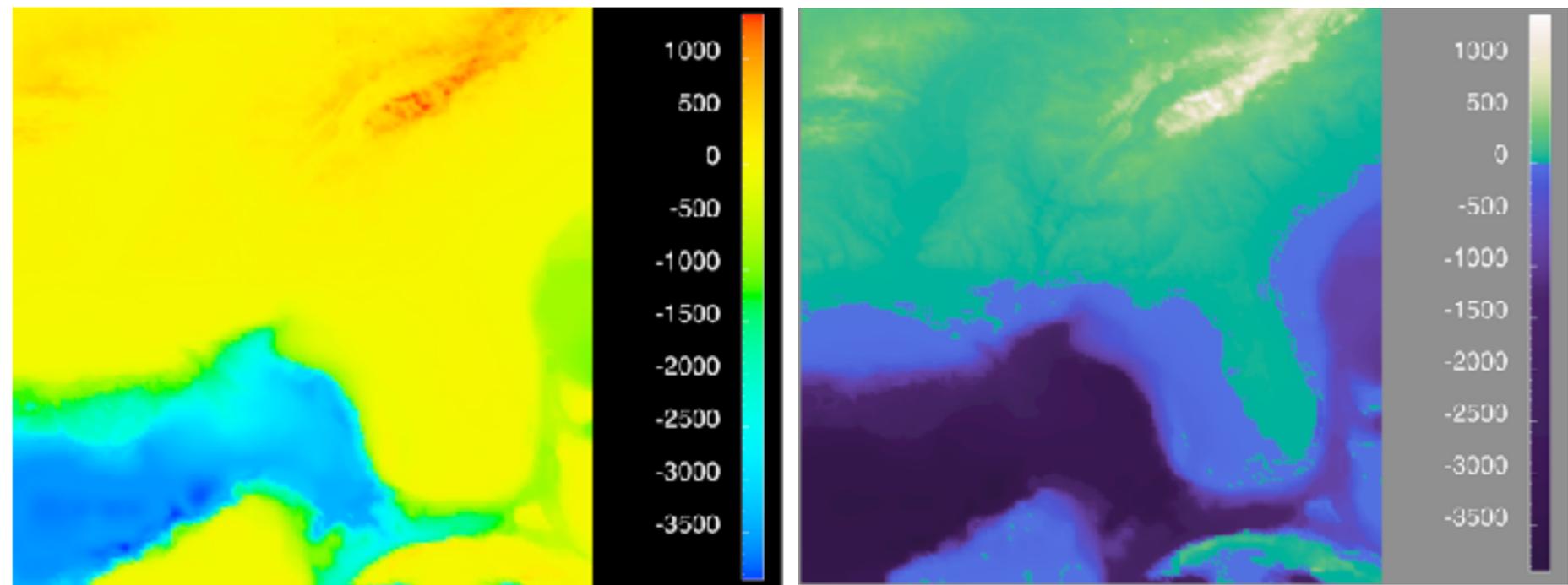
[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

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



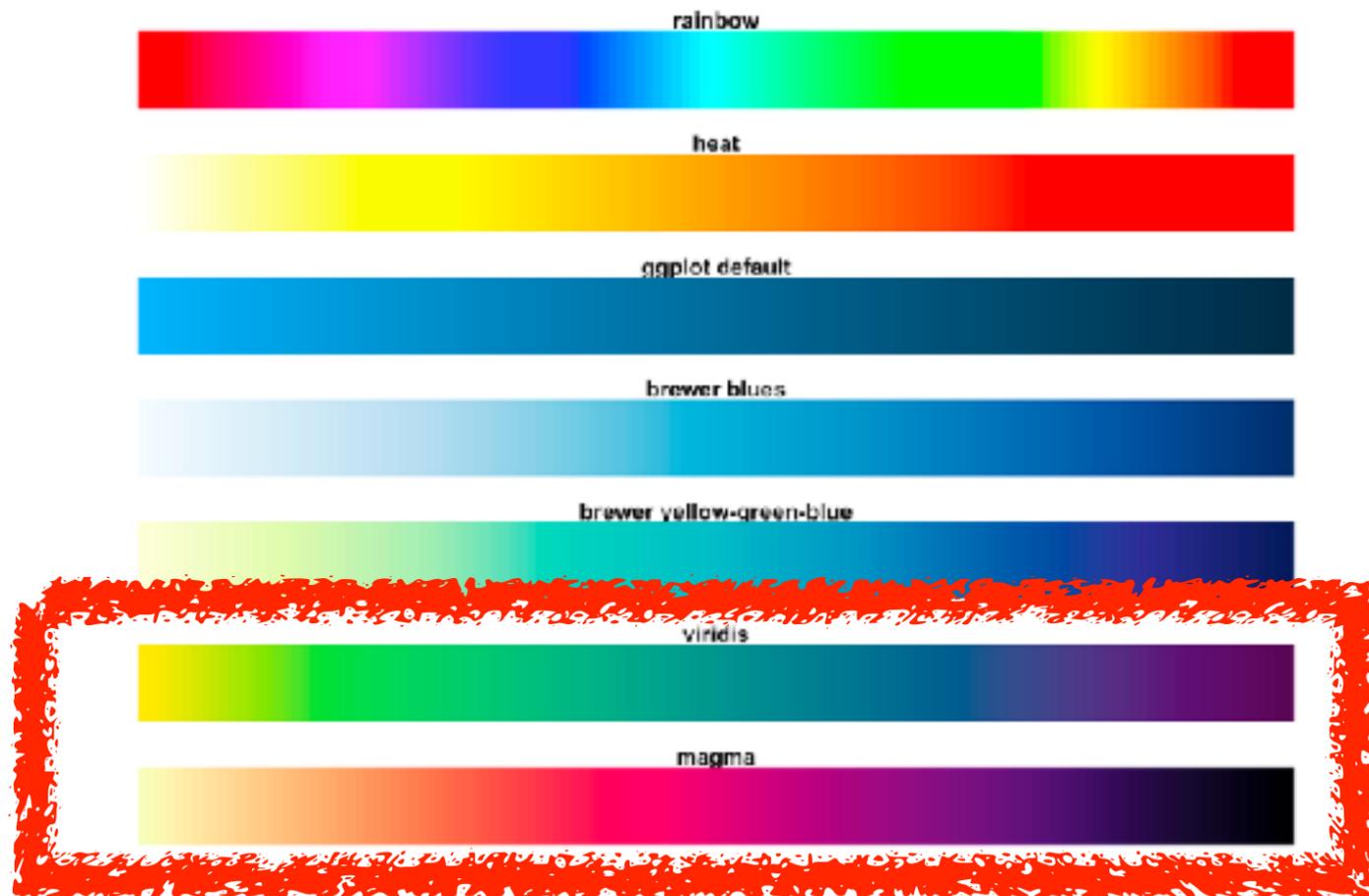
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



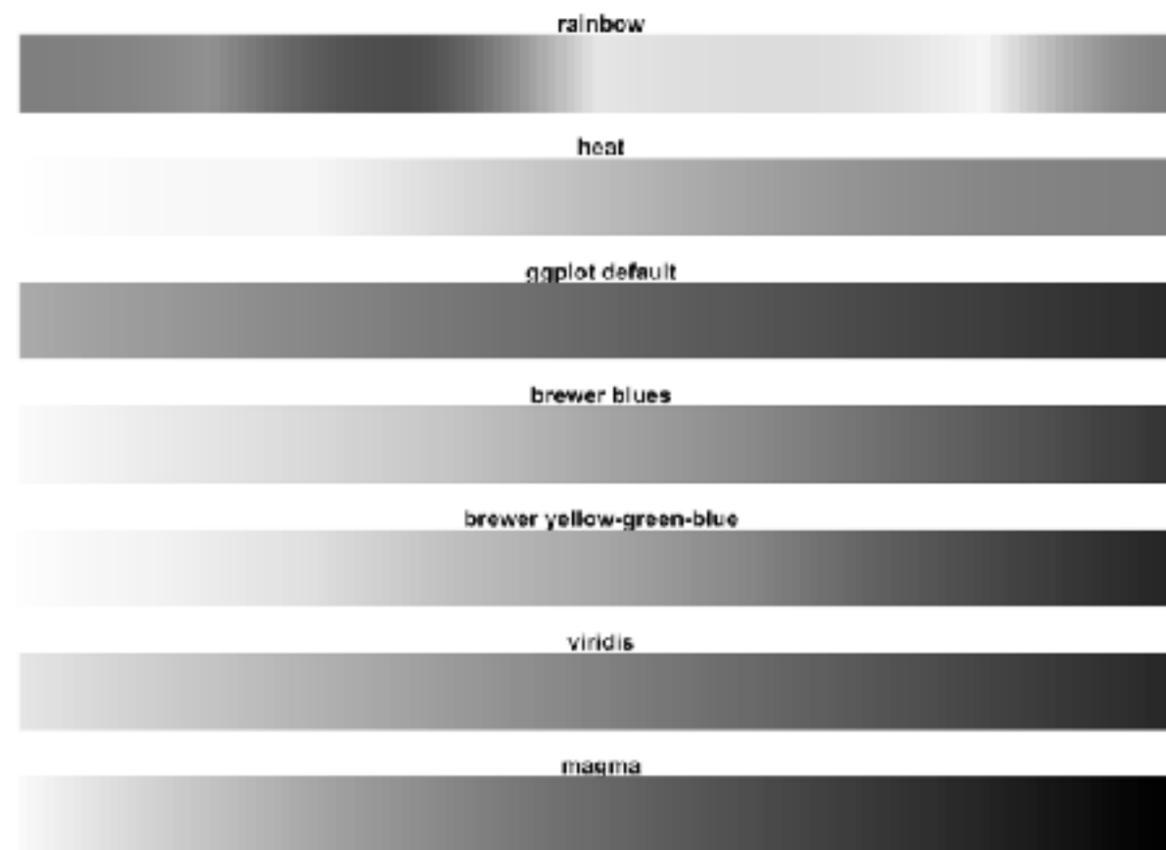
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. <http://www.research.ibm.com/people/llloyd/color/color.HTM>]

Viridis / Magma

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



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



How?

Encode

→ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



→ Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



→ Motion

Direction, Rate, Frequency, ...



Manipulate

→ Change



→ Select



→ Navigate



Facet

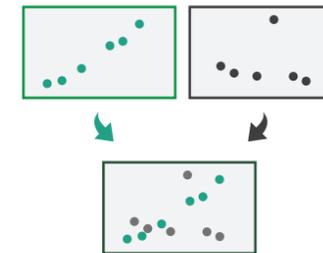
→ Juxtapose



→ Partition

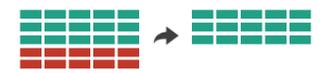


→ Superimpose



Reduce

→ Filter



→ Aggregate



→ Embed



What?

Why?

How?

How to handle complexity: 3 more strategies + 1 previous

Manipulate

➔ Change



➔ Select

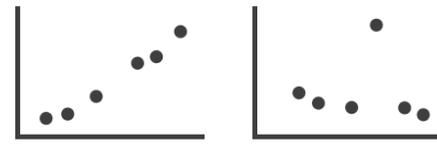


➔ Navigate

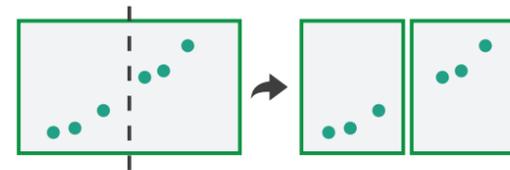


Facet

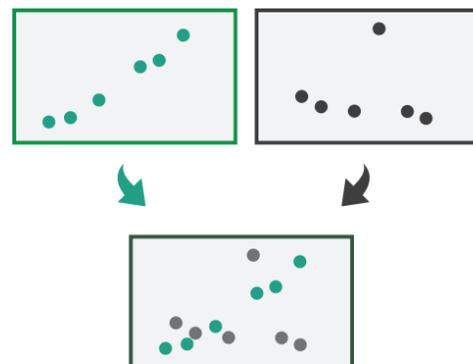
➔ Juxtapose



➔ Partition



➔ Superimpose



Reduce

➔ Filter



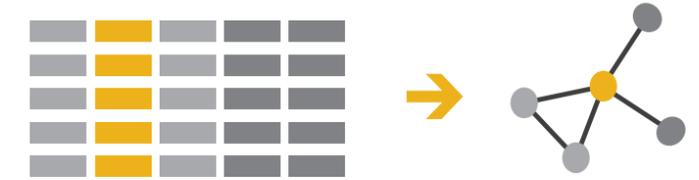
➔ Aggregate



➔ 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

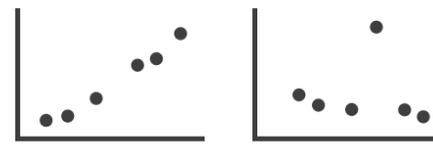
Manipulate

→ Change



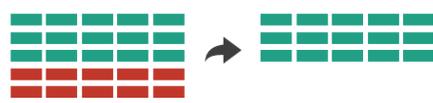
Facet

→ Juxtapose

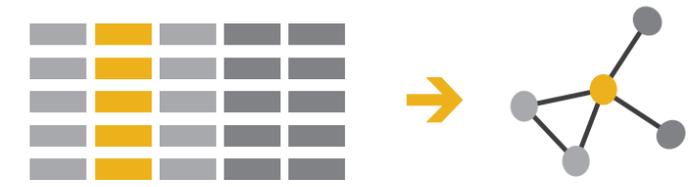


Reduce

→ Filter



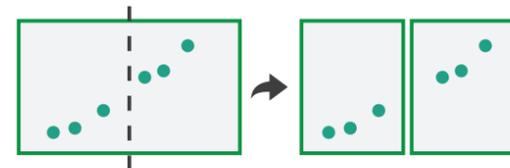
→ *Derive*



→ Select



→ Partition



→ Aggregate



→ Navigate



→ Superimpose



→ Embed



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

How to handle complexity: 3 more strategies

+ 1 previous

Manipulate

→ Change



→ Select

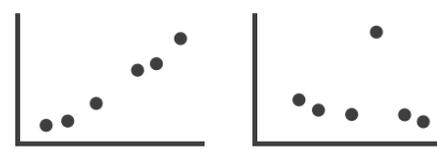


→ Navigate

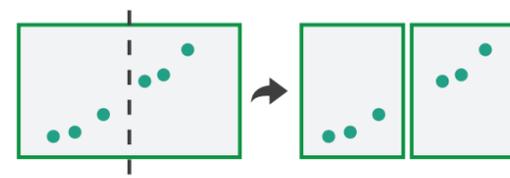


Facet

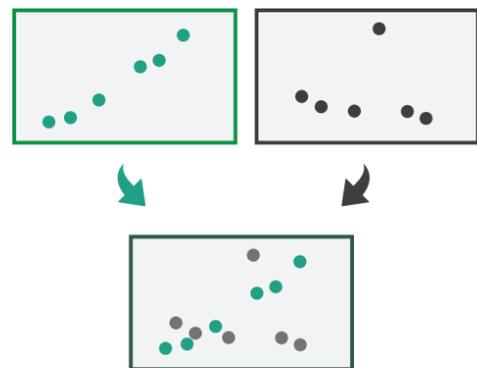
→ Juxtapose



→ Partition



→ Superimpose



Reduce

→ Filter



→ Aggregate



→ Embed



→ *Derive*

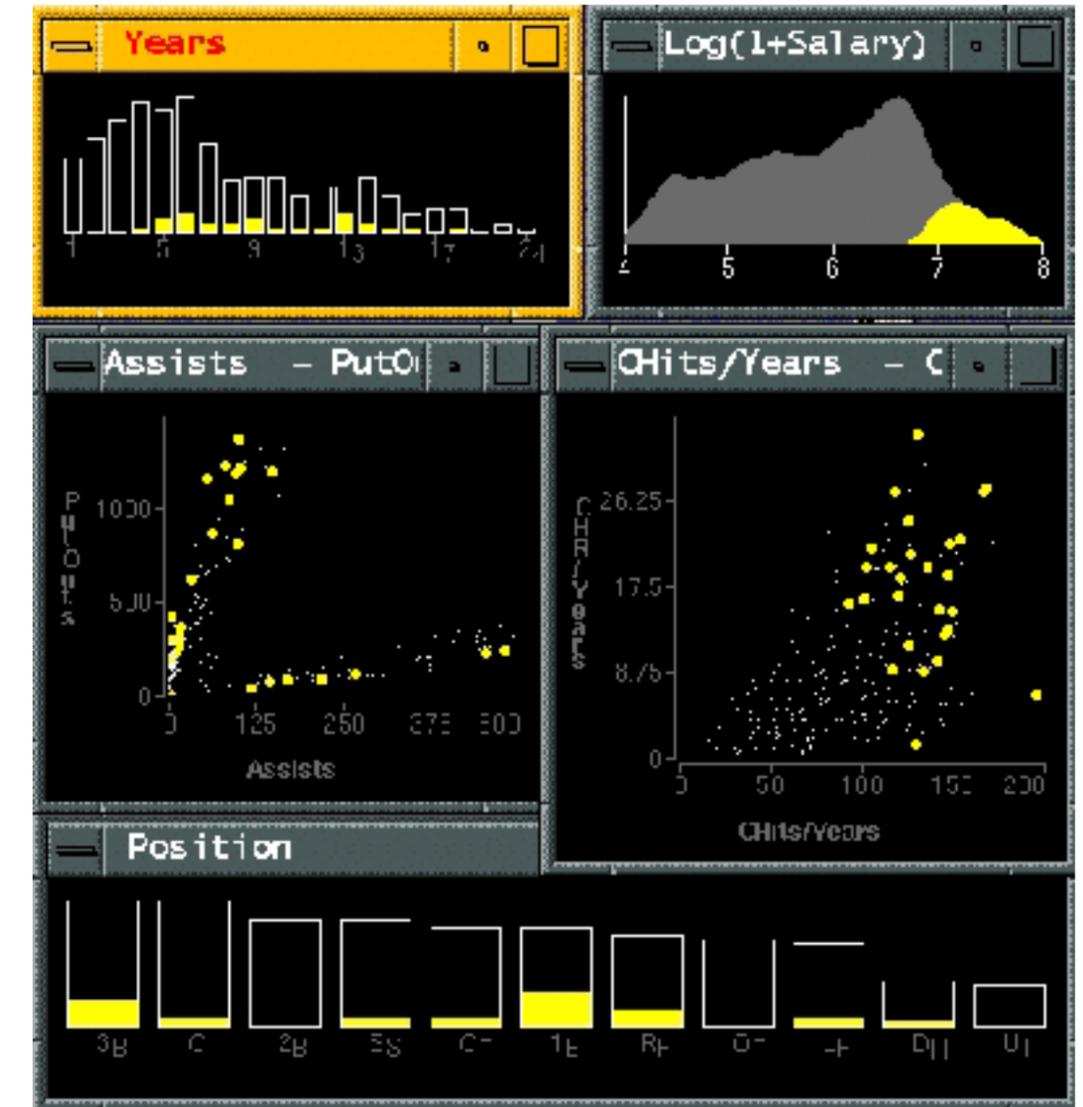


- facet data across multiple views

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)

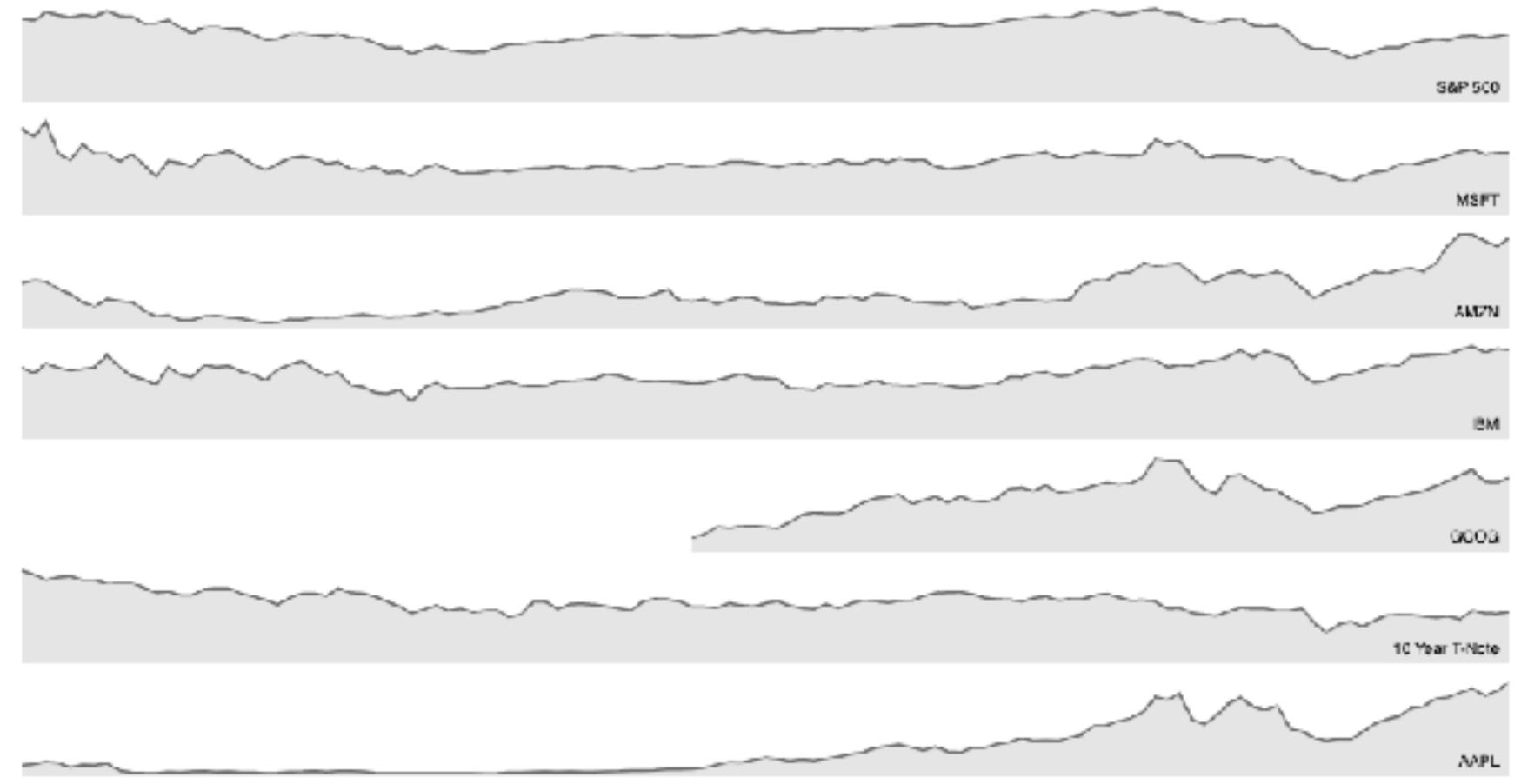


• **overview-detail**

[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**

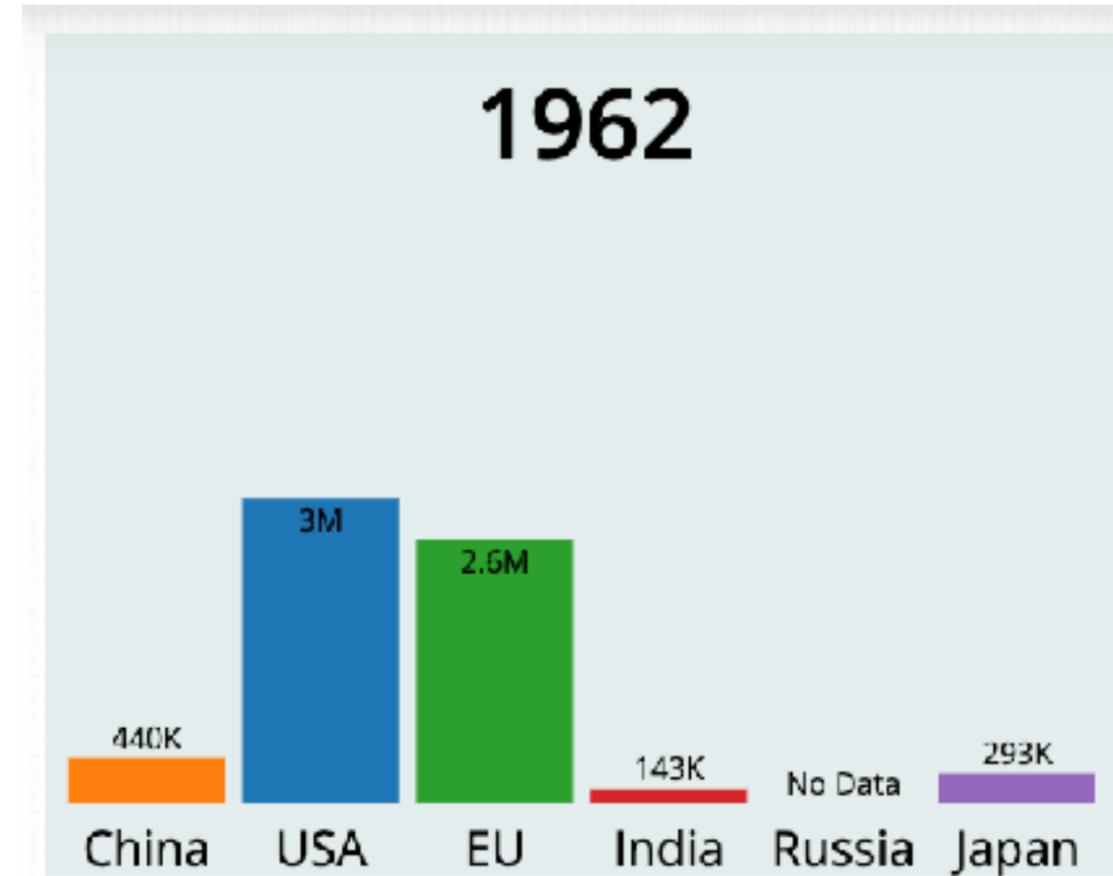
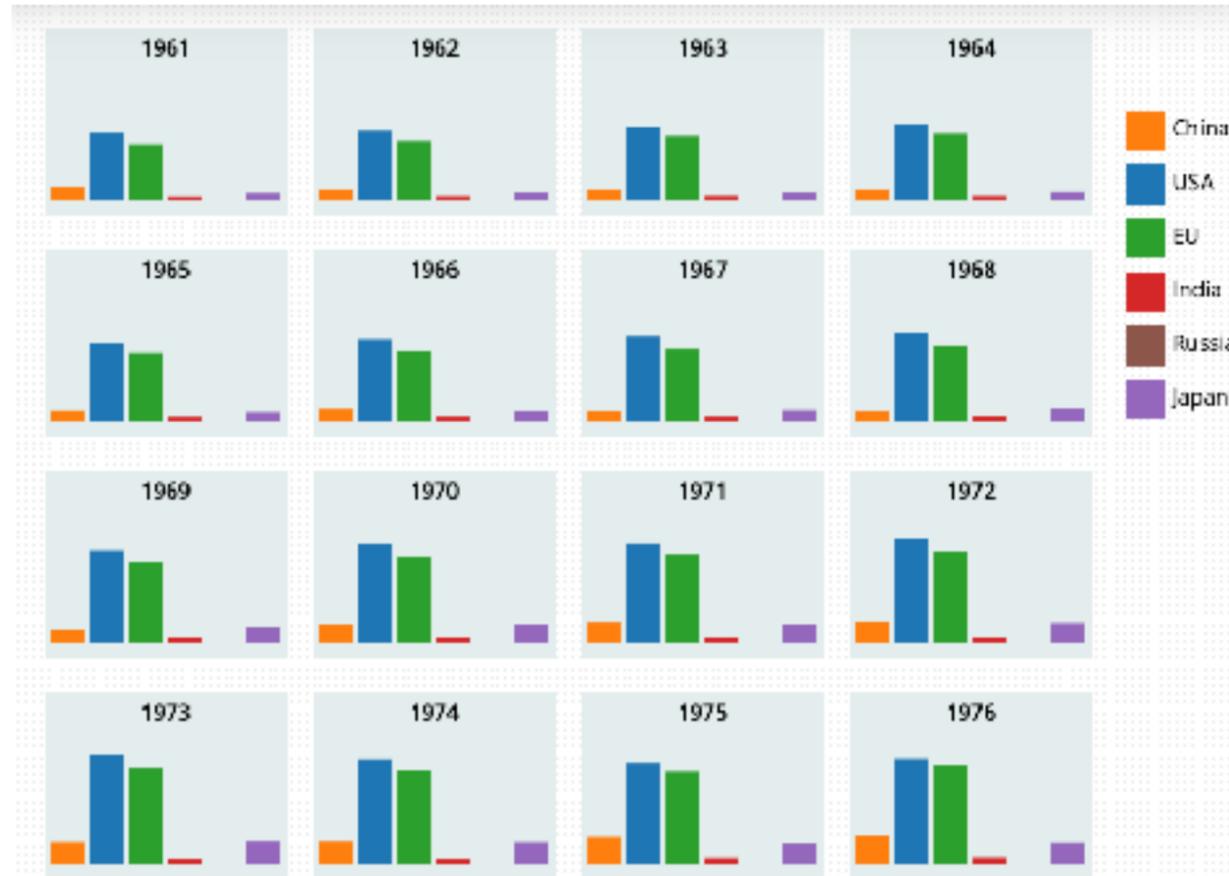
- encoding: same
 - ex: line charts
- data: none shared
 - different slices of dataset
 - items or attributes
 - ex: stock prices for different companies



[\[https://bl.ocks.org/mbostock/1157787\]](https://bl.ocks.org/mbostock/1157787)

Idiom: **Small multiples** + **details on demand**

- combining idioms



[\[http://vallandingham.me/co2_small_multiple\]](http://vallandingham.me/co2_small_multiple)

[\[https://vallandingham.me/small_multiples_with_details.html\]](https://vallandingham.me/small_multiples_with_details.html)

Interactive small multiples

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

The Rise and Decline of Ask MetaFilter

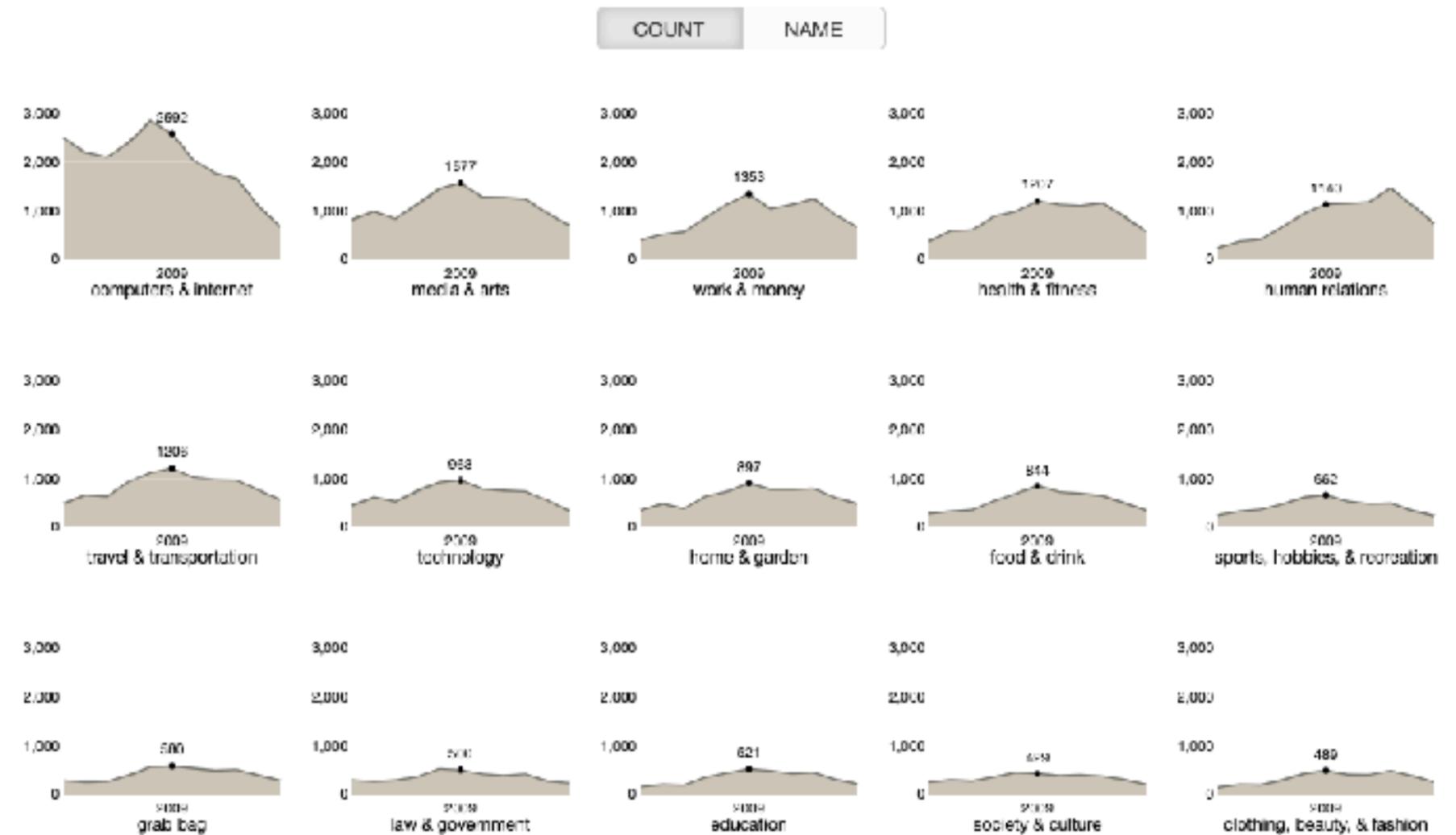
Metafilter's revenue has been on the decline, but has its content dried up as well?

Here we look at new posts on Ask Metafilter by category.

Categories like computers & internet have been dropping in use for a long time, most likely due to competition like Stack Overflow.

Other smaller categories have had consistent use patterns until more recently.

Disclaimer: 2014 is included, even though the year is not over yet.



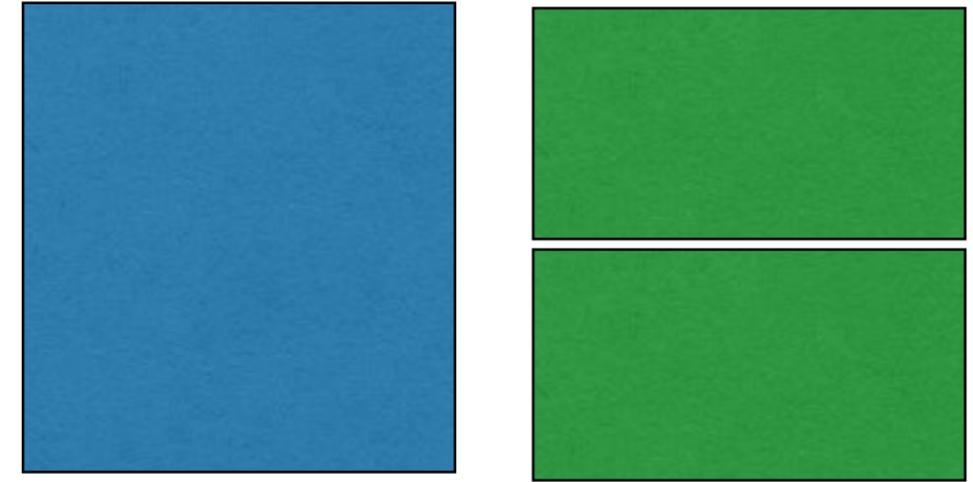
[<https://bl.ocks.org/ColinEberhardt/3c780088c363d1515403f50a87a87121>]

[<https://blog.scottlogic.com/2017/04/05/interactive-responsive-small-multiples.html>]

[http://projects.flowingdata.com/tut/linked_small_multiples_demo/]

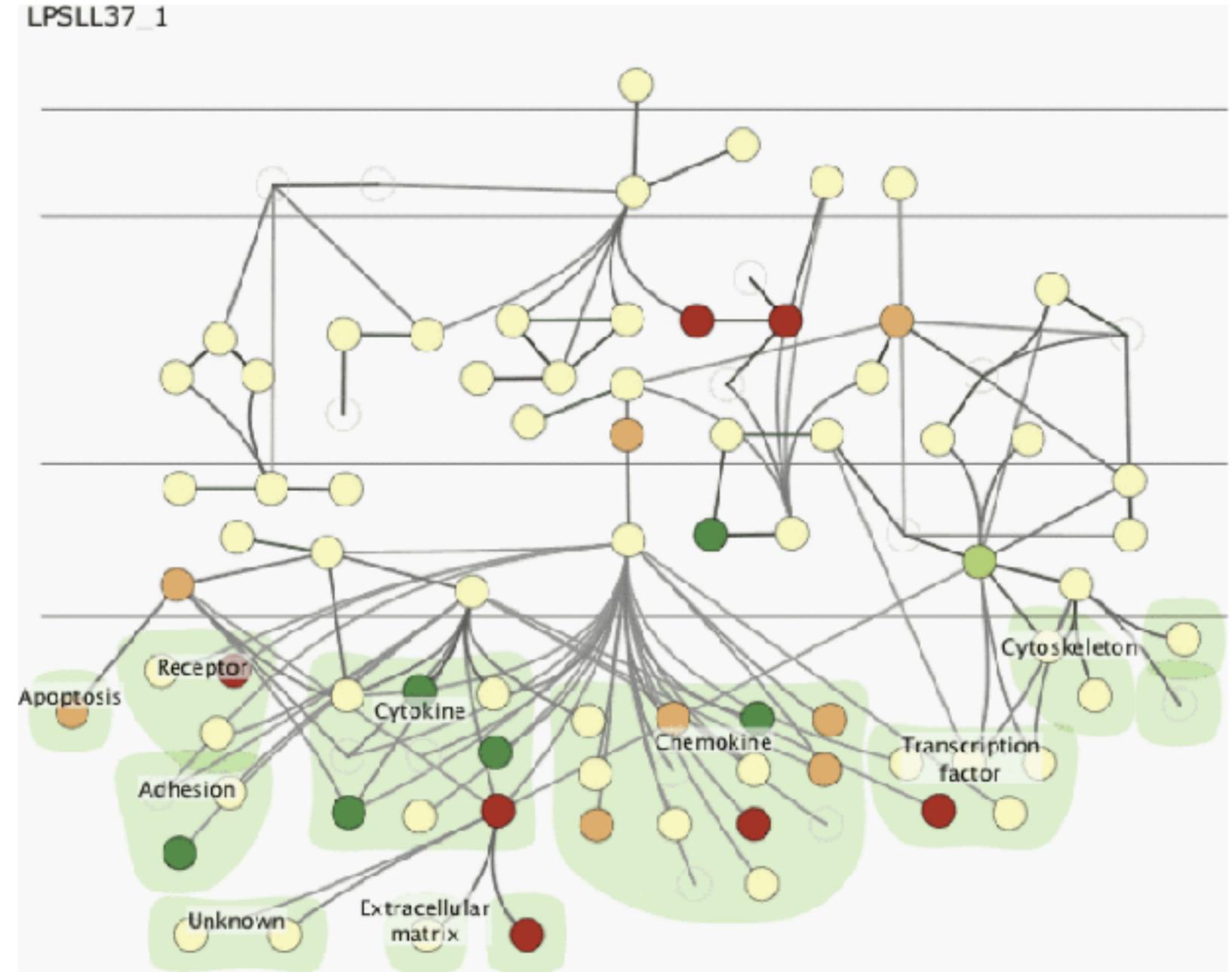
Juxtapose views: tradeoffs

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



Juxtapose vs animate

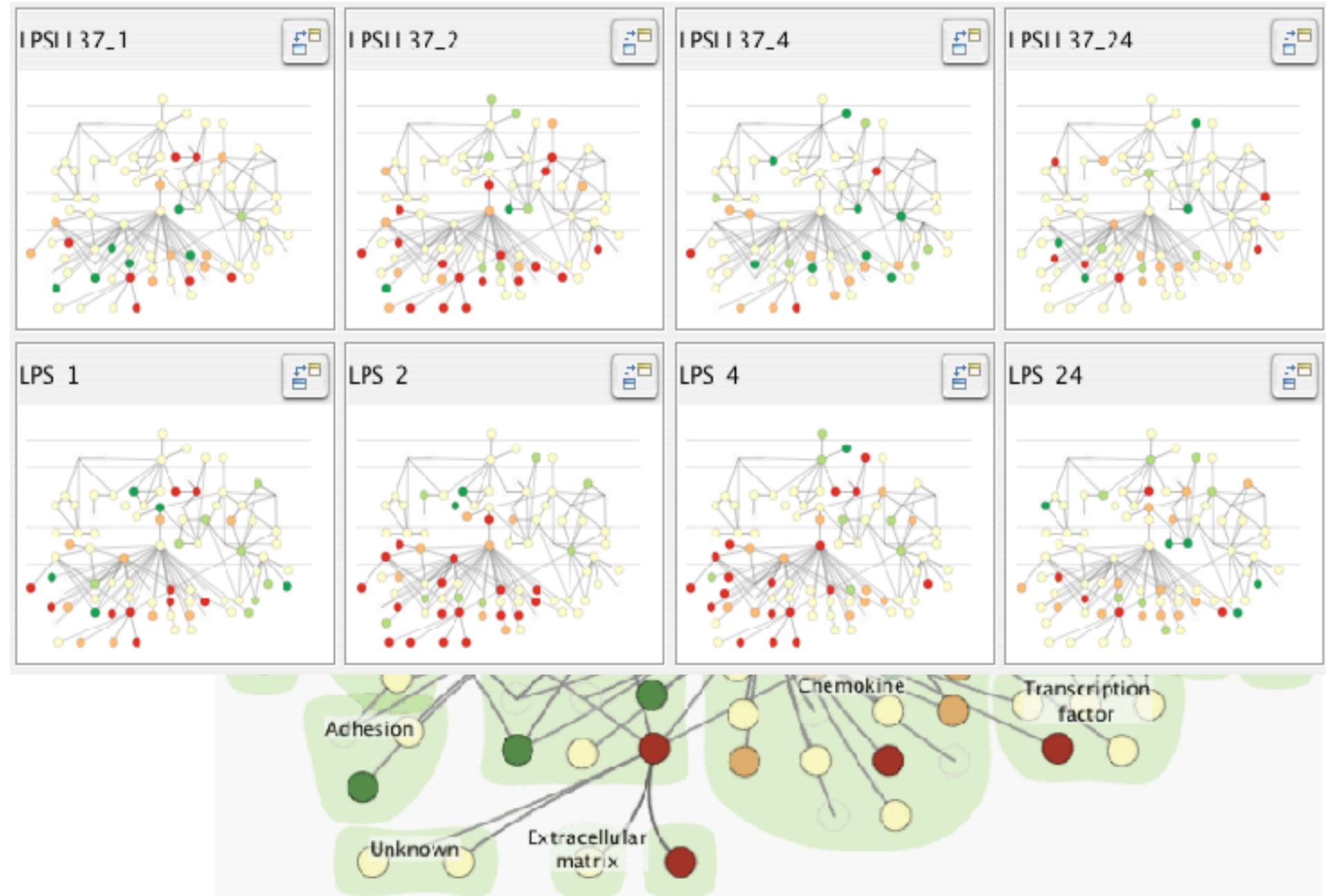
- animate: hard to follow if many scattered changes or many frames
 - vs easy special case: animated transitions



[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.]

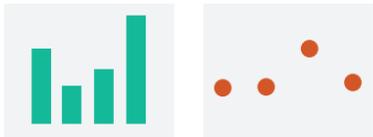
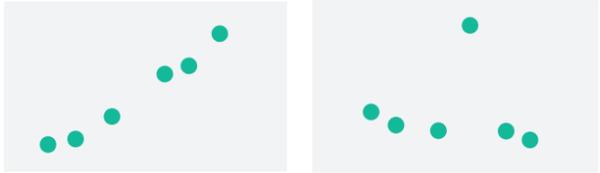
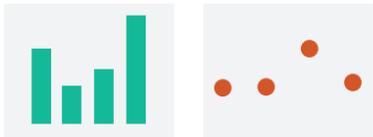
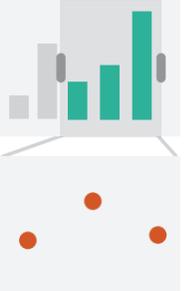
Juxtapose vs animate

- animate: hard to follow if many scattered changes or many frames
 - vs easy special case: animated transitions
- juxtapose: easier to compare across small multiples
 - different conditions (color), same gene (layout)



[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.]

View coordination: Design choices

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

How to handle complexity: 3 more strategies

+ 1 previous

Manipulate

➔ Change



➔ Select



➔ Navigate

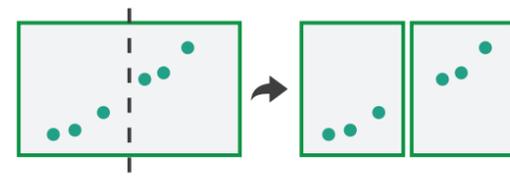


Facet

➔ Juxtapose



➔ Partition



➔ Superimpose



Reduce

➔ Filter



➔ Aggregate



➔ Embed



➔ *Derive*



- reduce what is shown within single view

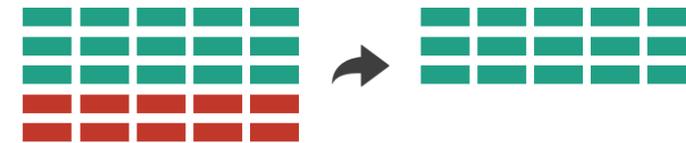
Reduce items and attributes

- 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

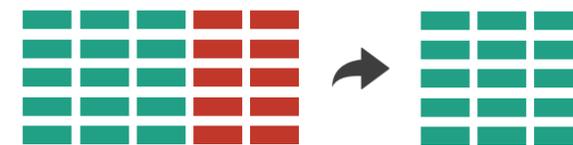
Reducing Items and Attributes

→ Filter

→ Items

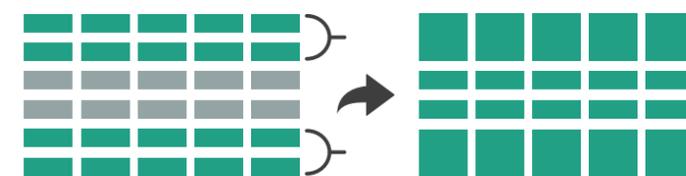


→ Attributes

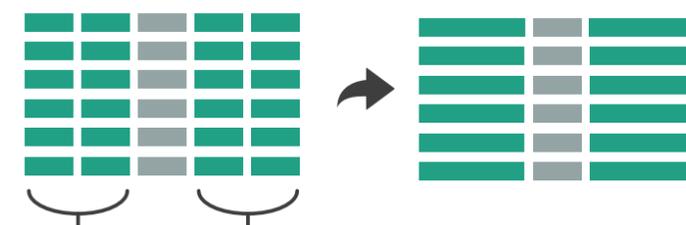


→ Aggregate

→ Items



→ Attributes



Reduce

→ Filter



→ Aggregate

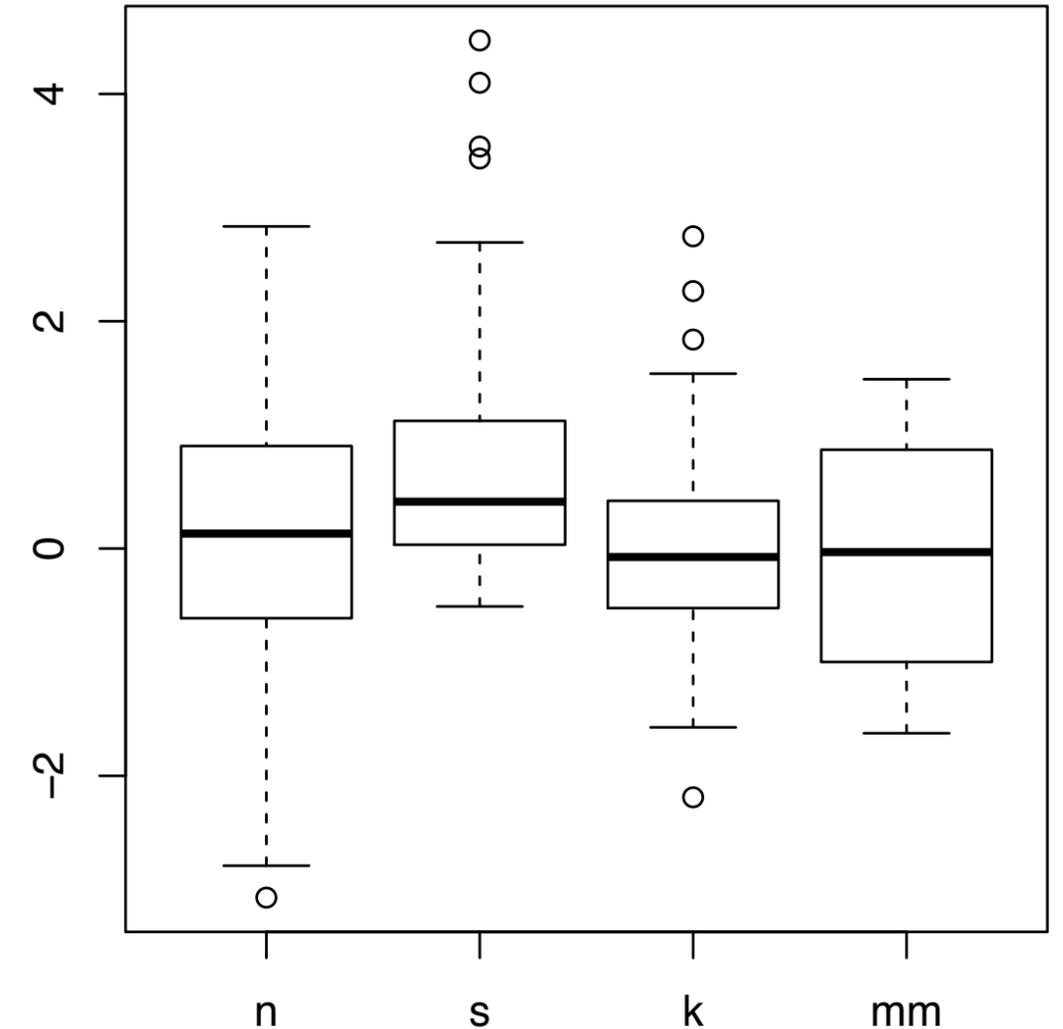


→ Embed



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



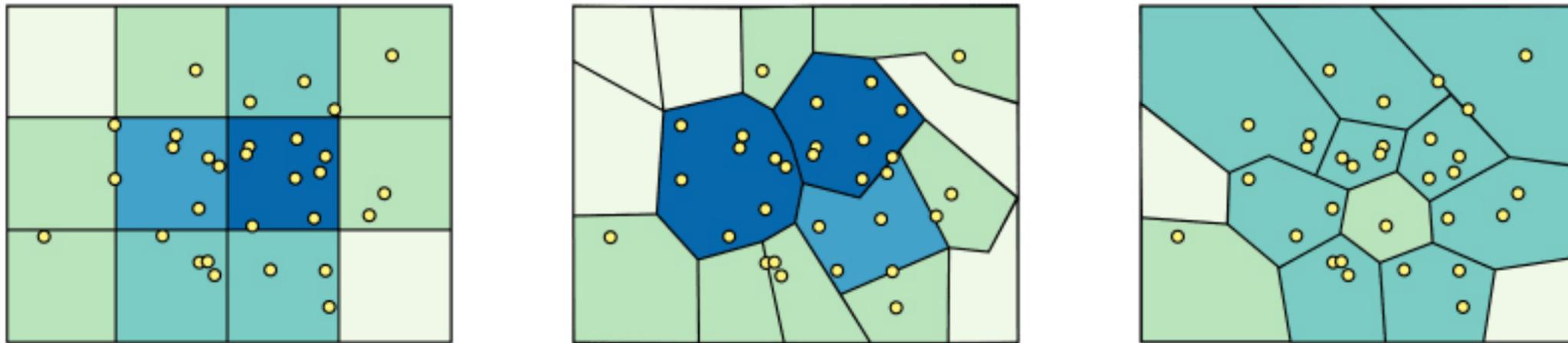
[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

Spatial aggregation

- MAUP: Modifiable Areal Unit Problem

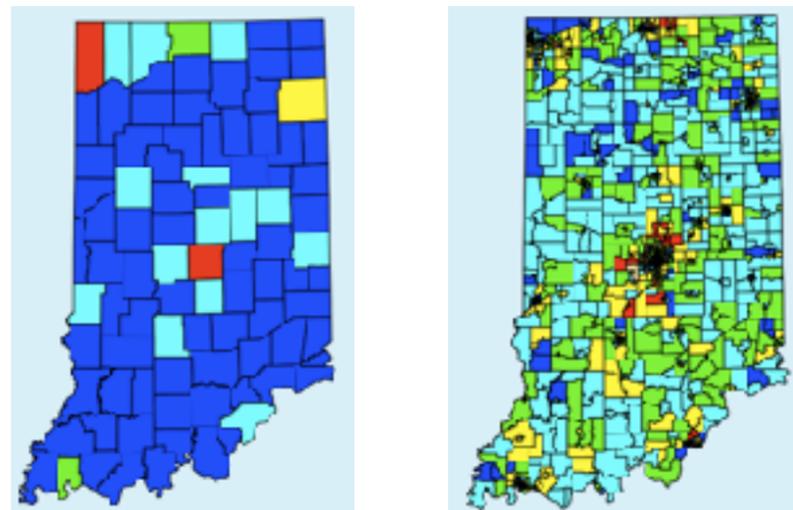
- changing boundaries of cartographic regions can yield dramatically different results

- zone effects



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

- scale effects



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

Dimensionality reduction

- attribute aggregation
 - derive low-dimensional target space from high-dimensional measured space
 - capture most of variance with minimal error
 - use when you can't directly measure what you care about
 - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
 - latent factors, hidden variables

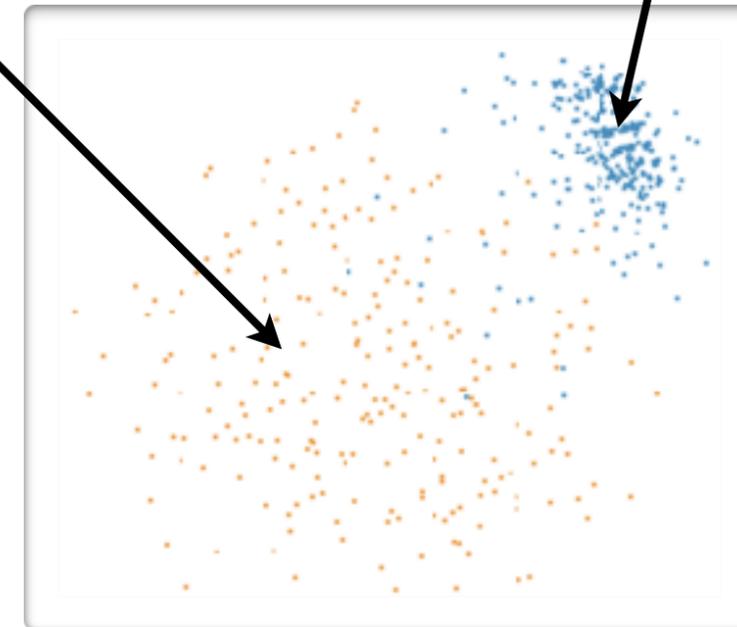
Tumor
Measurement Data

data: 9D measured space

→ DR →

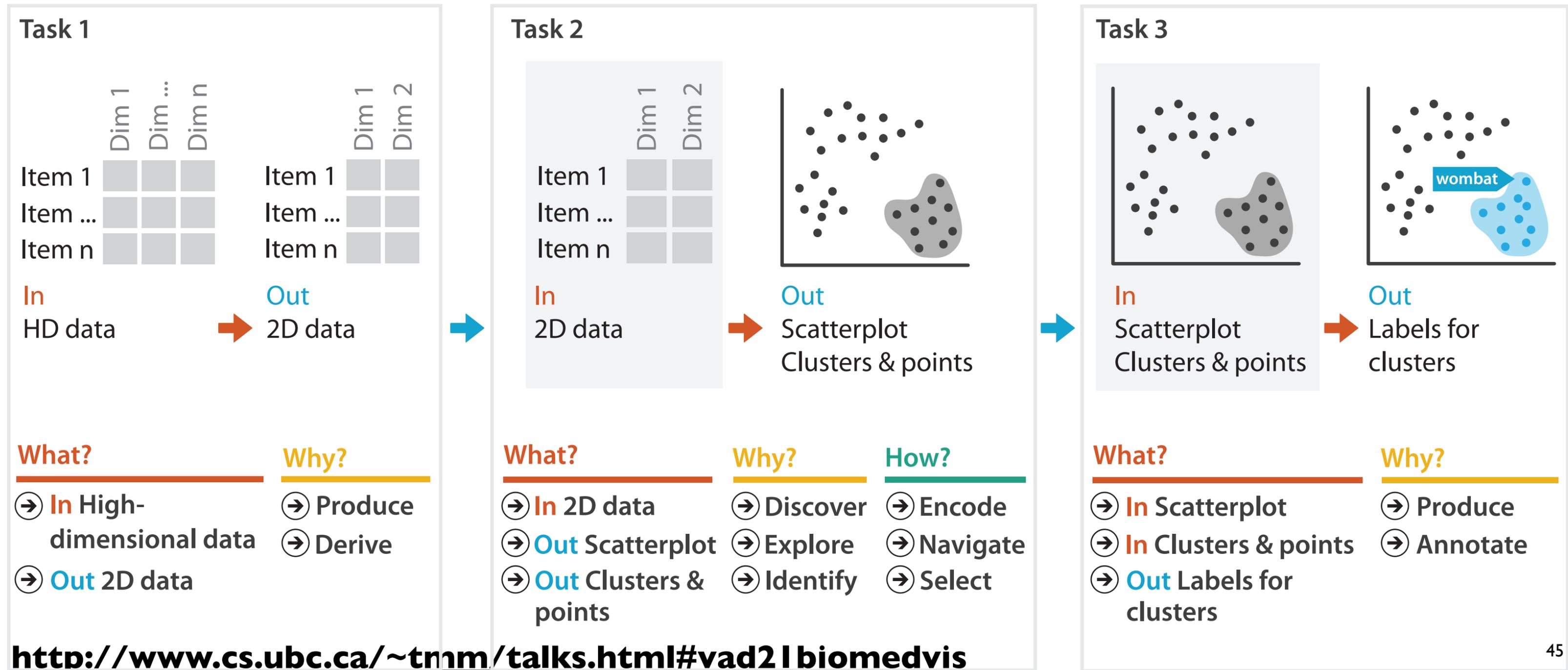
Malignant

Benign



derived data: 2D target space

Idiom: Dimensionality reduction for documents



What?

Datasets

Attributes

domain

abstraction

What?

Why?

idiom

How?

algorithm

Why?

→ Data Types

→ Items

👉 Actions

🎯 Targets

→ Data and D

Tables

Items

Attributes

→ Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ All Data

→ Trends



→ Outliers

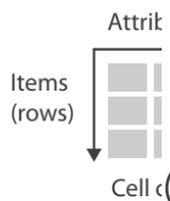


→ Features



→ Dataset Typ

→ Tables



→ Produce

→ Annotate



→ Search

	Tar
Location known	••
Location unknown	<••

→ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



→ Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



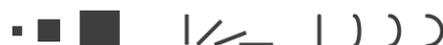
→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...

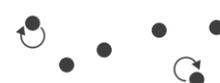


→ Shape



→ Motion

Direction, Rate, Frequency, ...



How?

Encode

Manipulate

Facet

Reduce

→ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



→ Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...

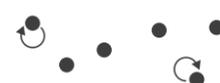


→ Shape



→ Motion

Direction, Rate, Frequency, ...



→ Change



→ Select



→ Navigate



→ Juxtapose



→ Partition



→ Superimpose



→ Filter



→ Aggregate



→ Embed



What?

Why?



More Information

[@tamaramunzner](https://twitter.com/tamaramunzner)

- this talk

<http://www.cs.ubc.ca/~tmm/talks.html#vad2|biomedvis>

- book page (including tutorial lecture slides)

<http://www.cs.ubc.ca/~tmm/vadbook>

– 20% promo code for book+ebook combo: HVN17

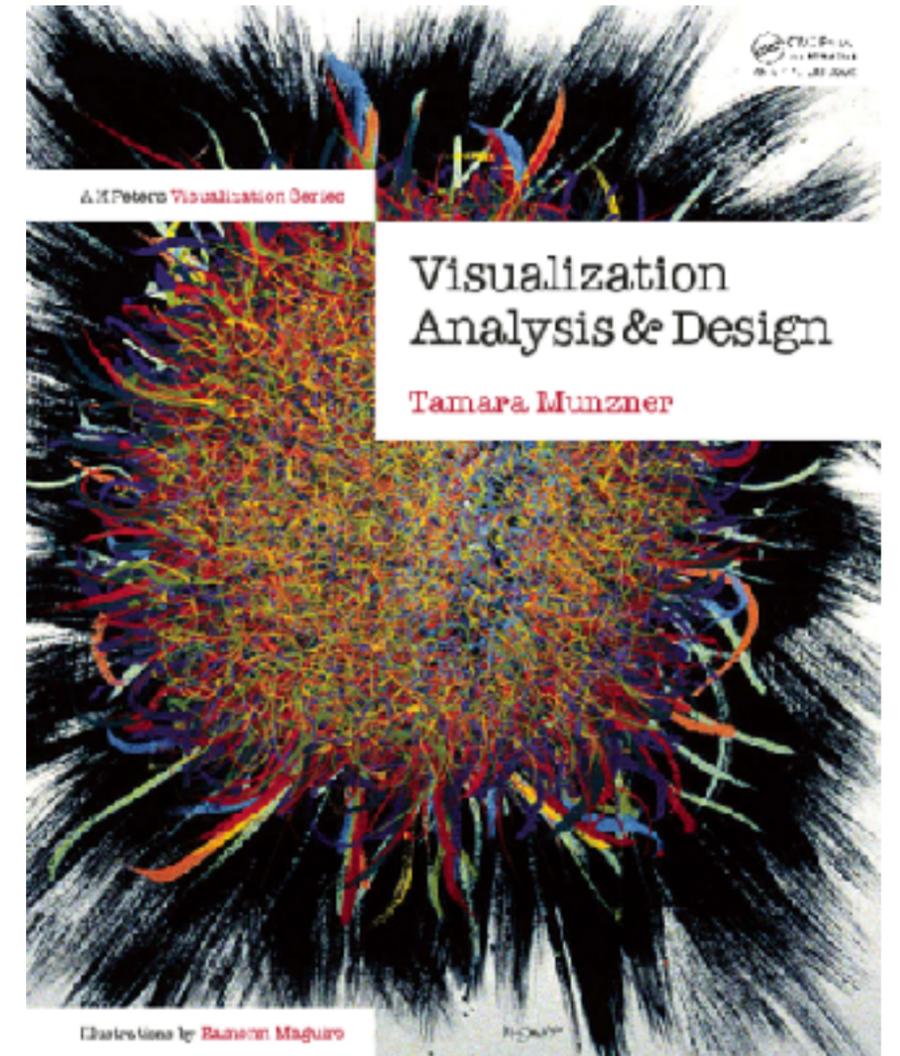
– <http://www.crcpress.com/product/isbn/9781466508910>

– 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.

<http://www.cs.ubc.ca/~tmm/talks.html#vad2|biomedvis>