

Section I

◀

1,853

2,967

...

Scientific visualisation

Repository (will be updated during course)

<https://github.com/fercook/dataviz-notes>

Bibliography:

Visualisation Analysis and Design (Tamara Munzner)

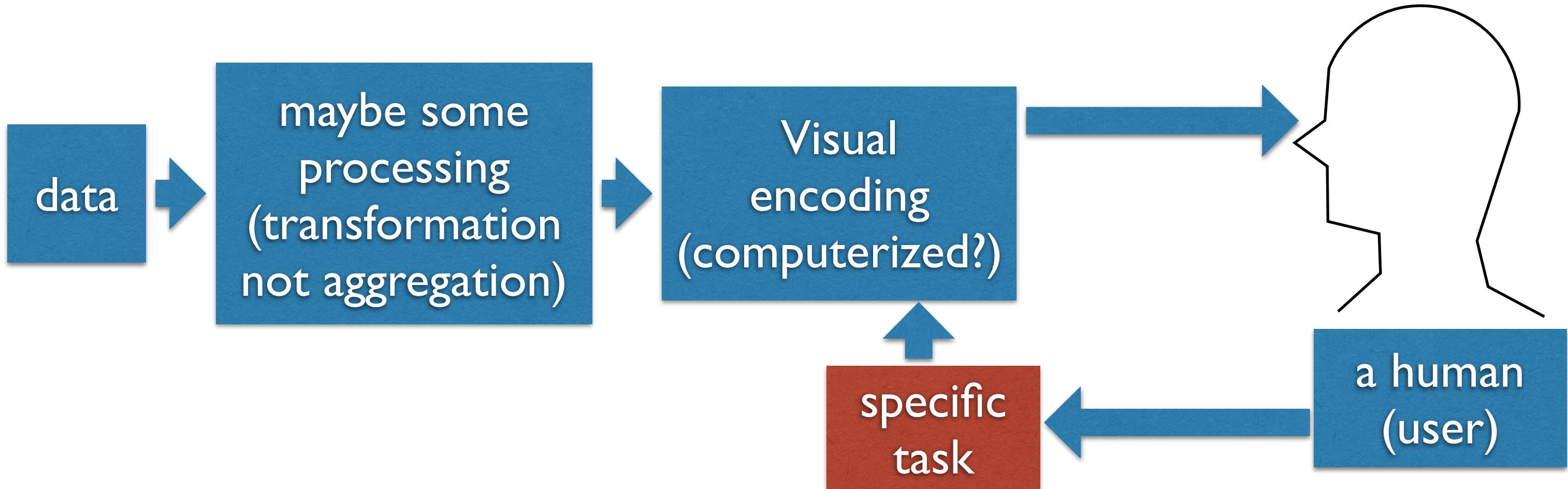
The Functional Art (Alberto Cairo)

The Truthful Art (Alberto Cairo)

Data Visualisation (Andy Kirk)

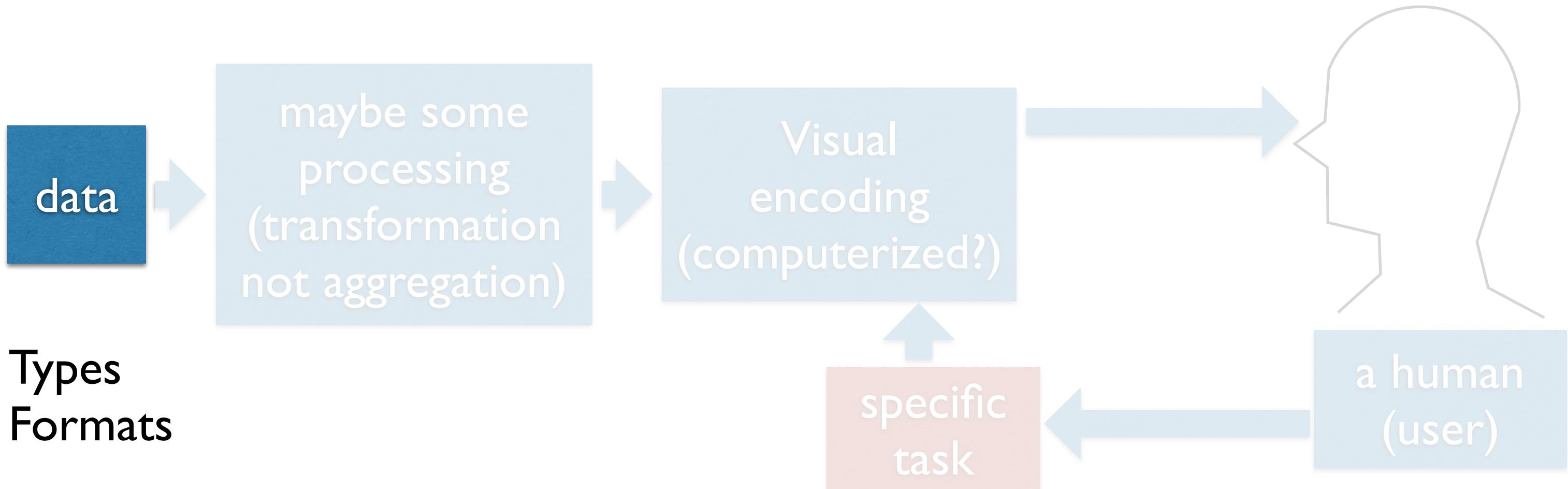
Data Visualisation

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



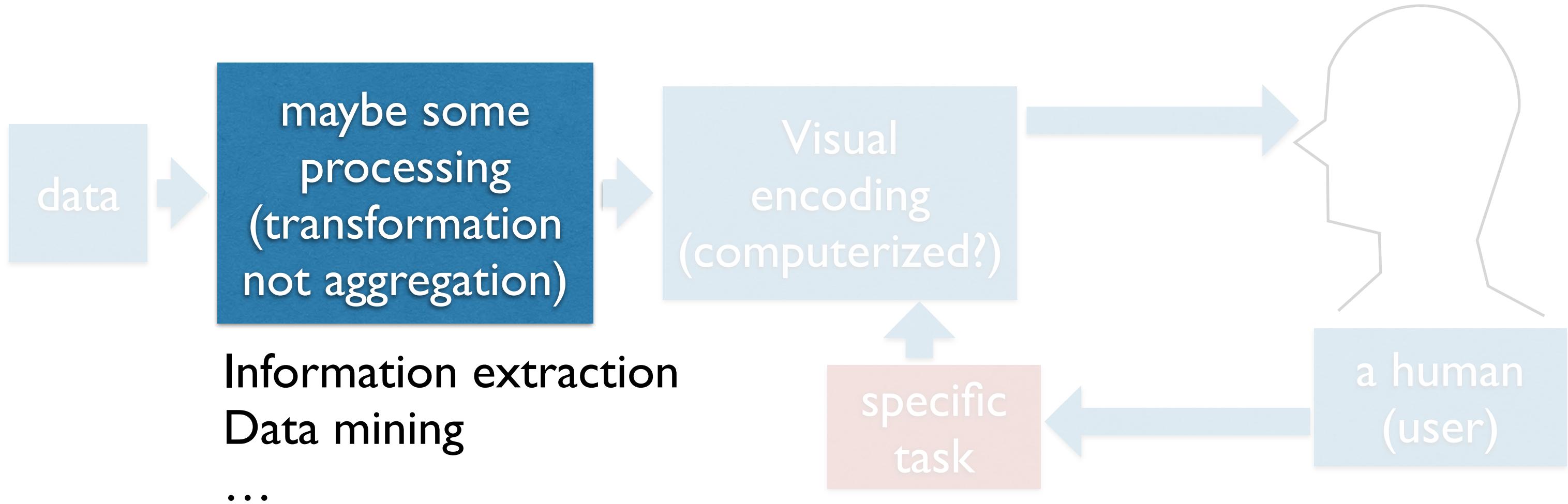
Data Visualisation

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



Data Visualisation

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



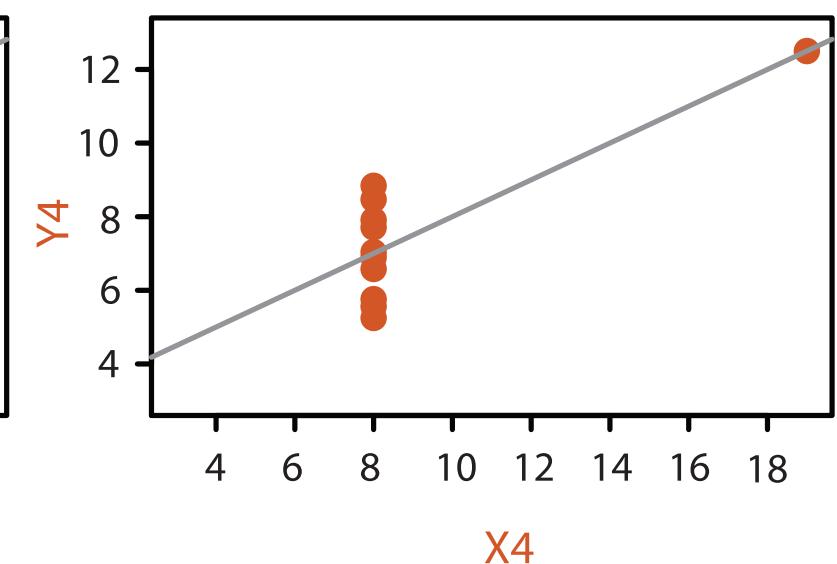
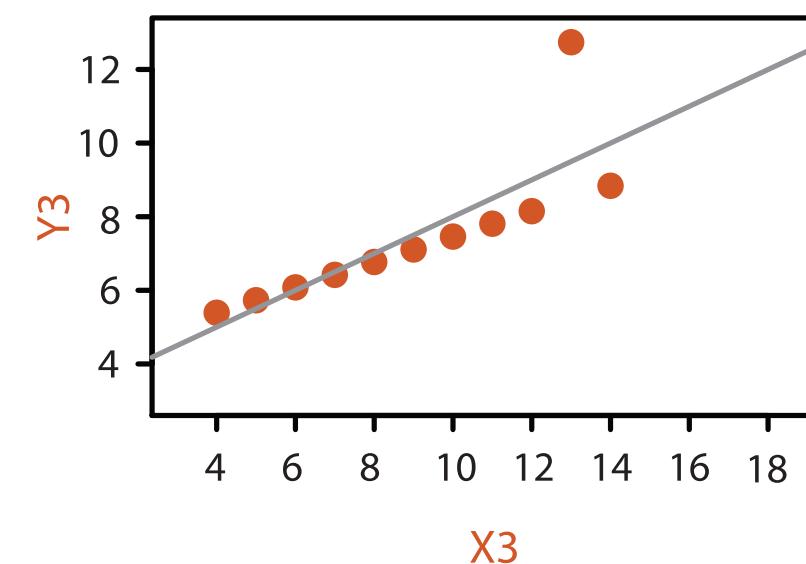
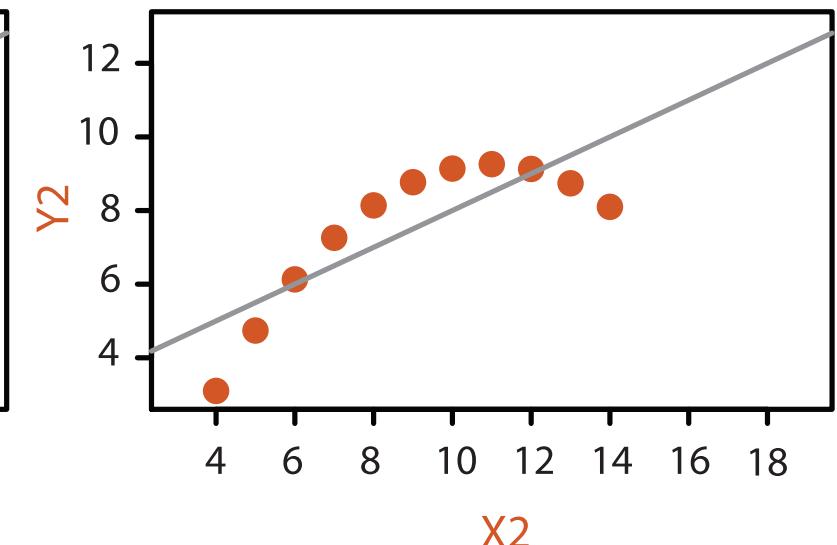
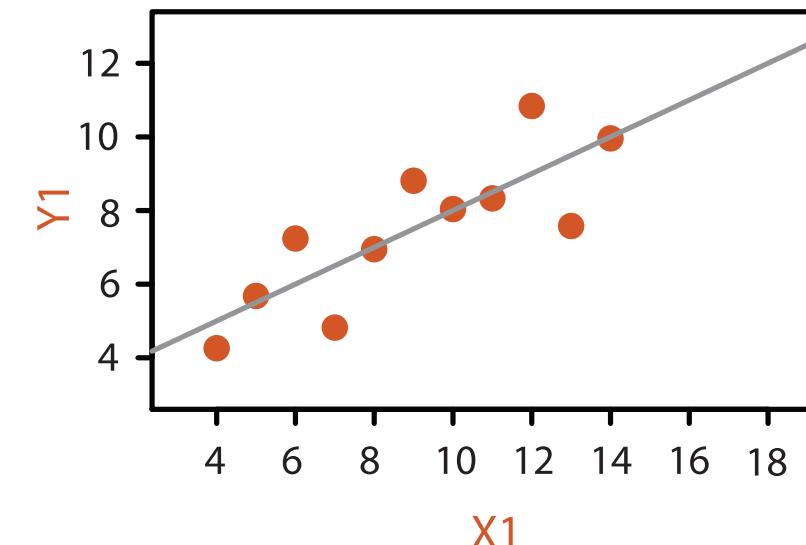
Why represent all the data?

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

Anscombe's Quartet

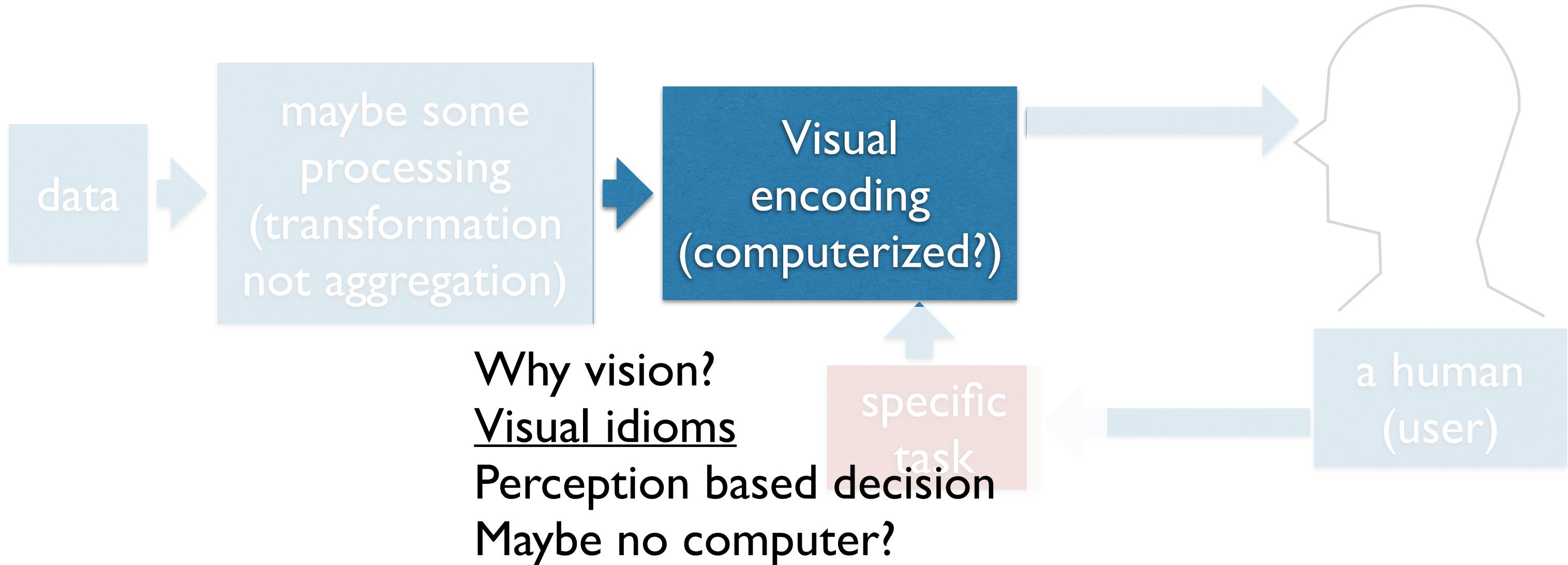
Identical statistics

x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816



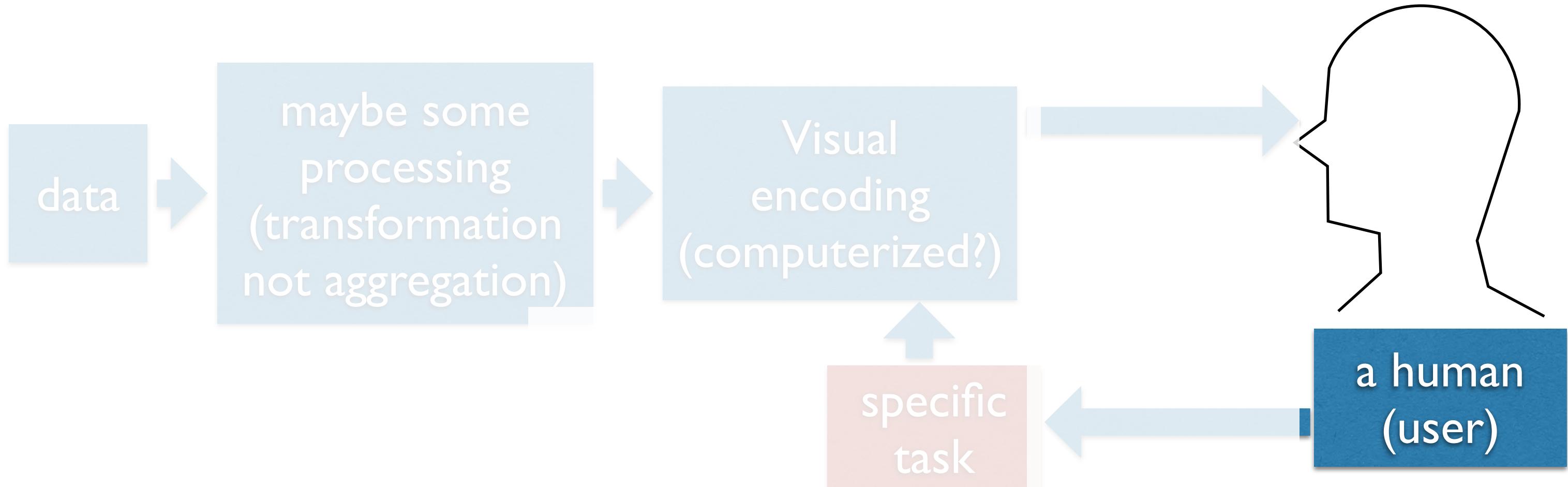
Data Visualisation

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



Data Visualisation

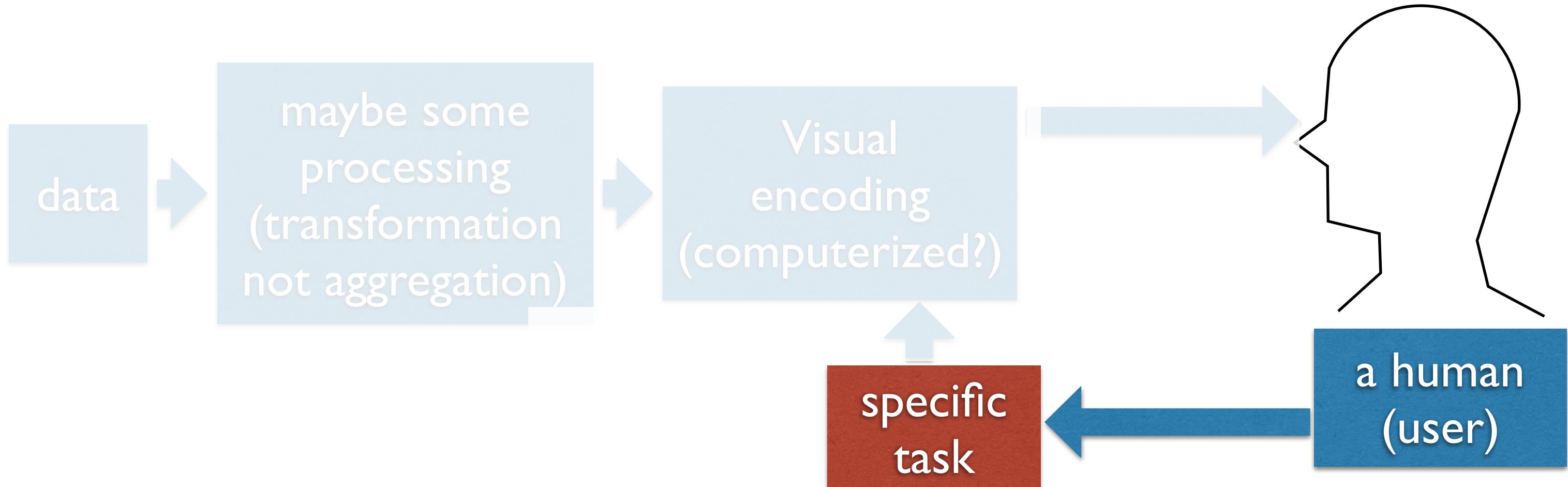
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively (T.Munzner)



No vis if algorithm is better
Good at vague tasks

Data Visualisation

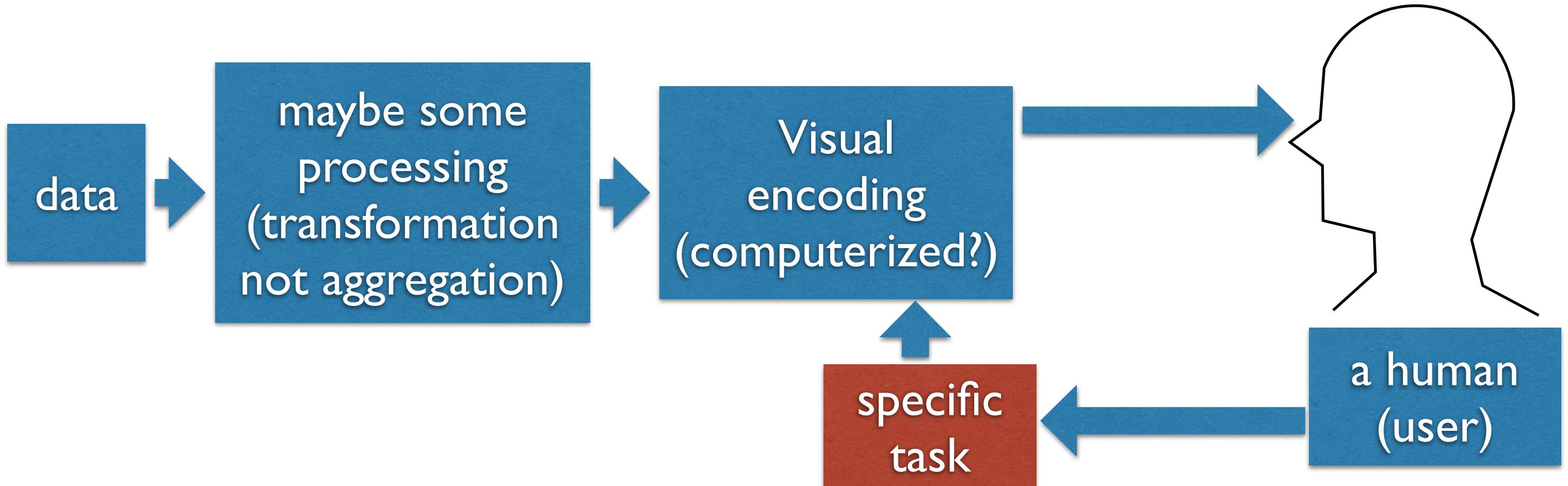
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively (T.Munzner)



Constrain improves efficiency
Filters out possibilities

Data Visualisation

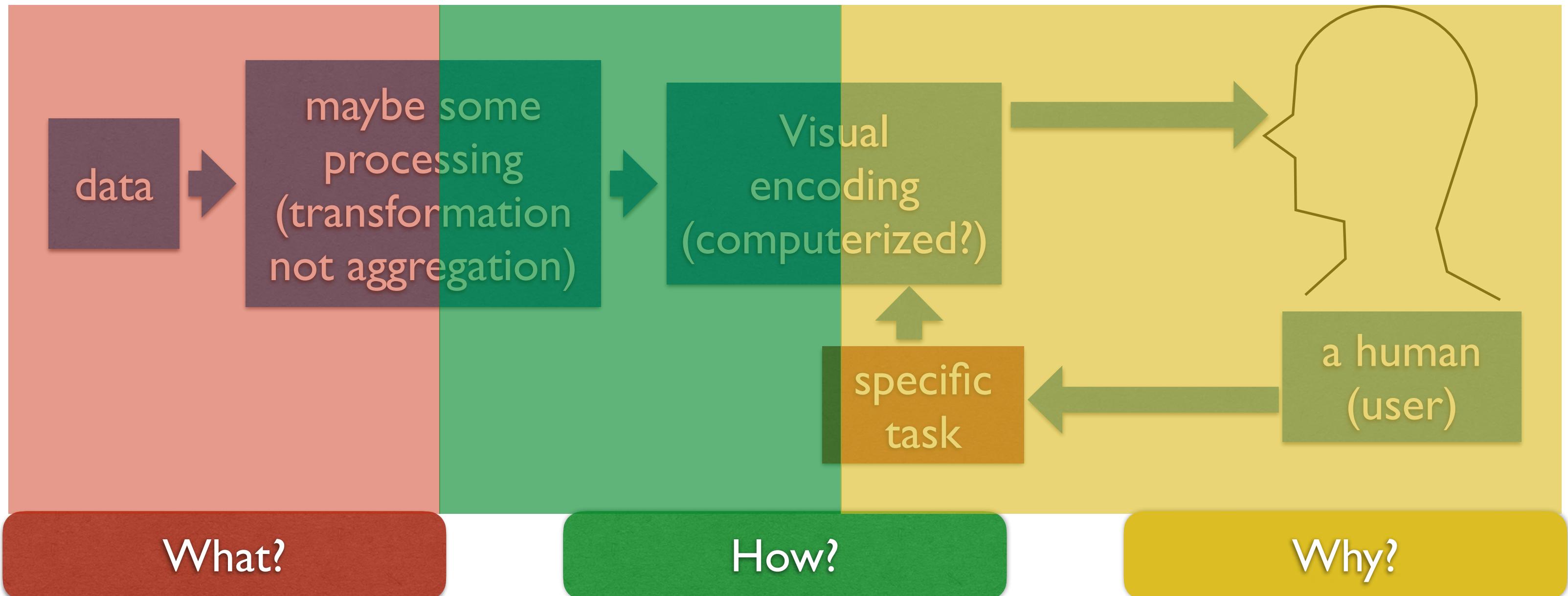
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively (T.Munzner)



Charts!!!

Data Visualisation

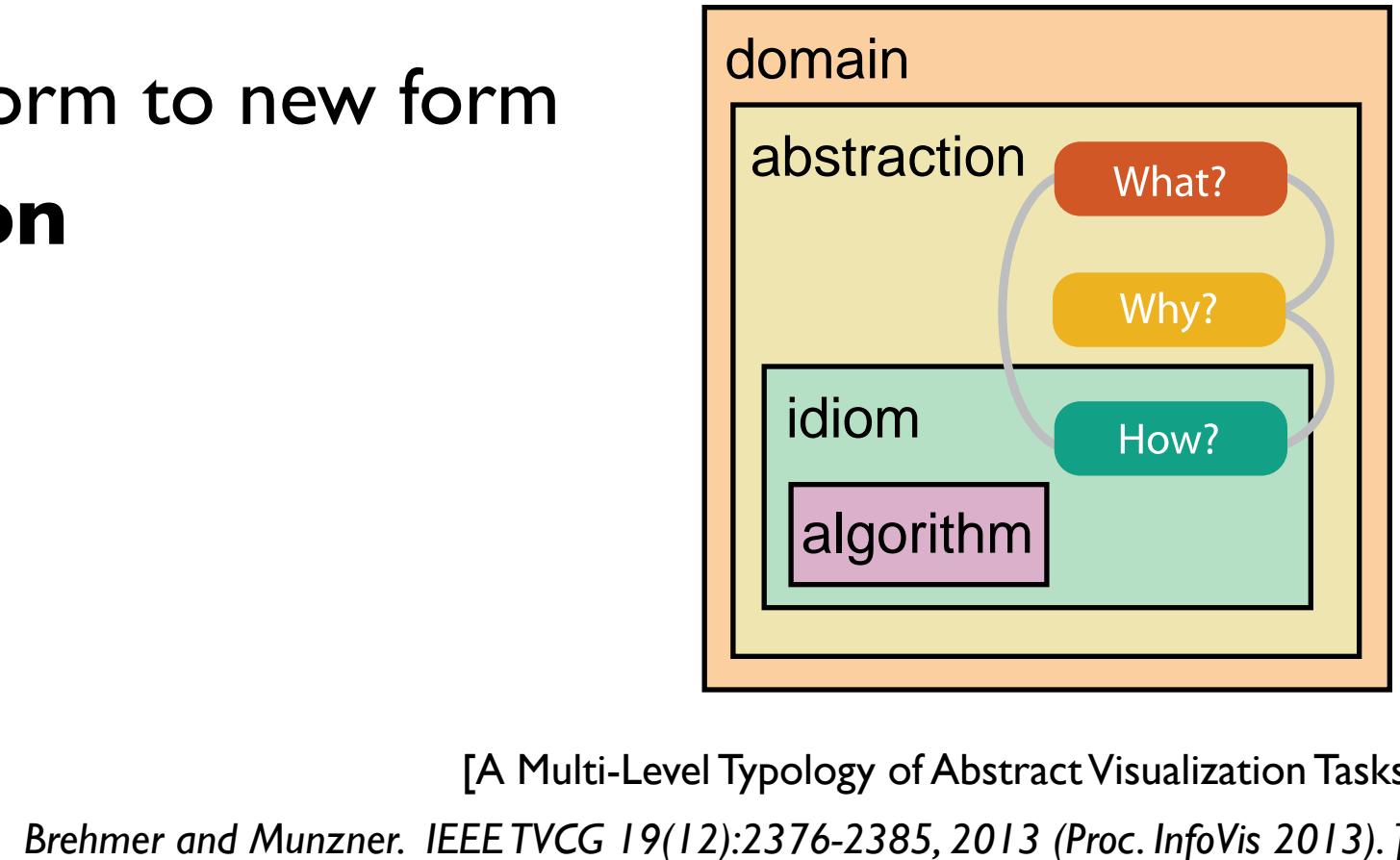
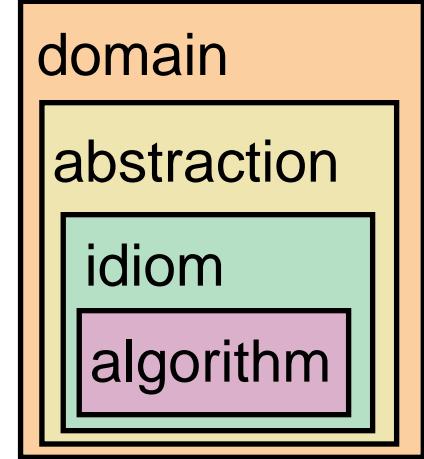
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively (T.Munzner)



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).]



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

Why is validation difficult?

- solution: use methods from different fields at each level

anthropology/
ethnography

design

computer
science

cognitive
psychology

anthropology/
ethnography

👤 Domain situation

Observe target users using existing tools

💡 Data/task abstraction

👁️ Visual encoding/interaction idiom

Justify design with respect to alternatives

📈 Algorithm

Measure system time/memory

Analyze computational complexity

Analyze results qualitatively

Measure human time with lab experiment (*lab study*)

Observe target users after deployment (*field study*)

Measure adoption



problem-driven
work

technique-driven
work

Question:
What types of data do you know?

What?

Datasets

Attributes

→ Data Types

→ Items → Attributes → Links → Positions → Grids

→ Attribute Types

→ Categorical



→ Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Clusters, Sets, Lists
Attributes	Links	Positions	Positions	Items

→ Ordered

→ Ordinal

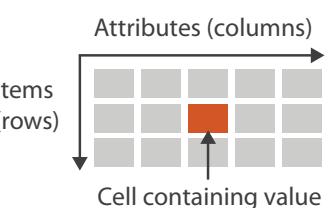


→ Quantitative

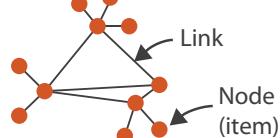


→ Dataset Types

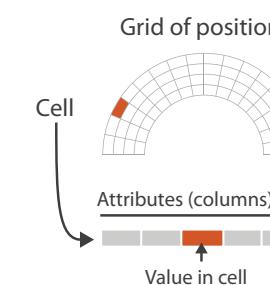
→ Tables



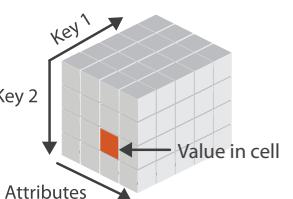
→ Networks



→ Fields (Continuous)



→ Multidimensional Table



→ Geometry (Spatial)



→ Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



→ Dataset Availability

→ Static



→ Dynamic



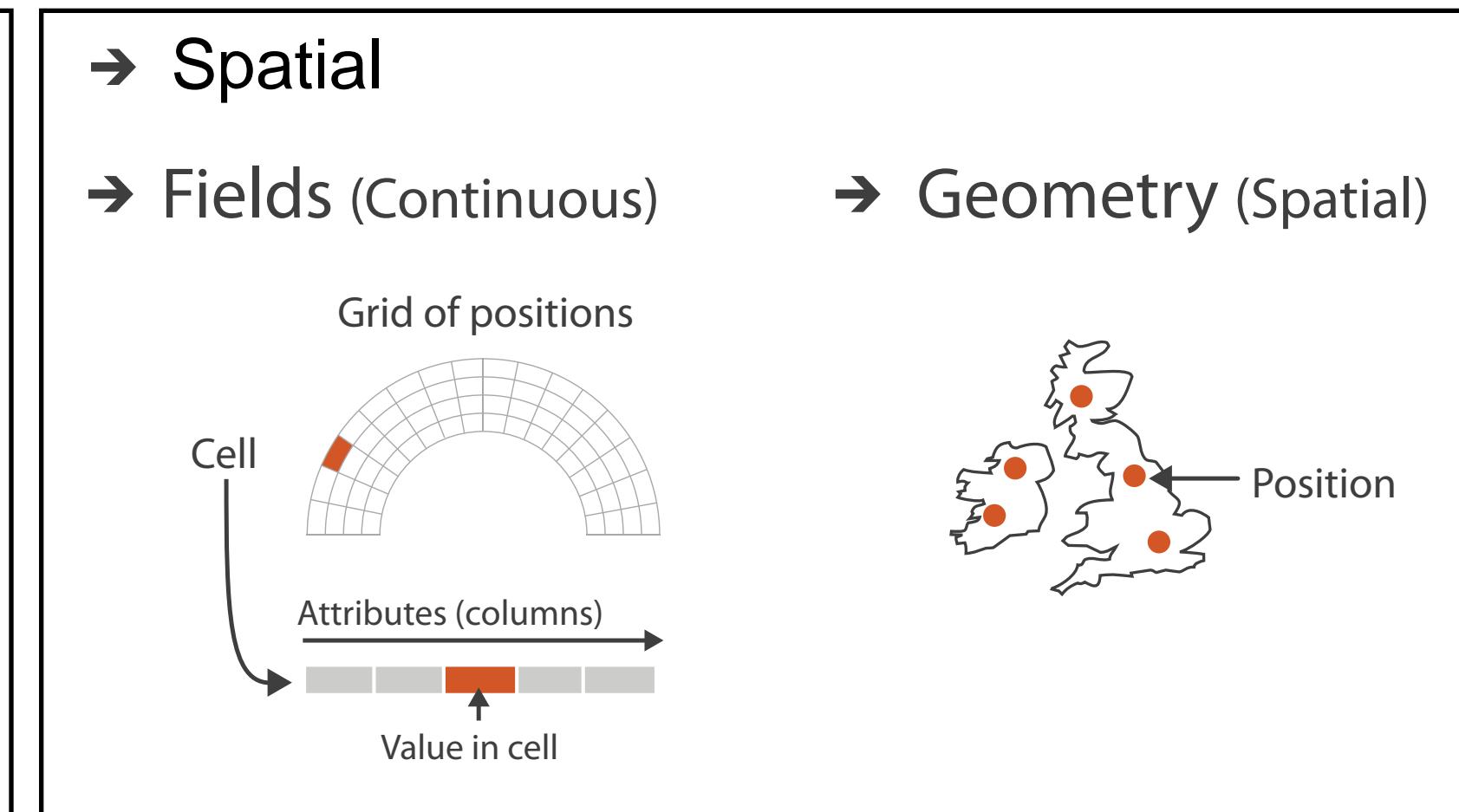
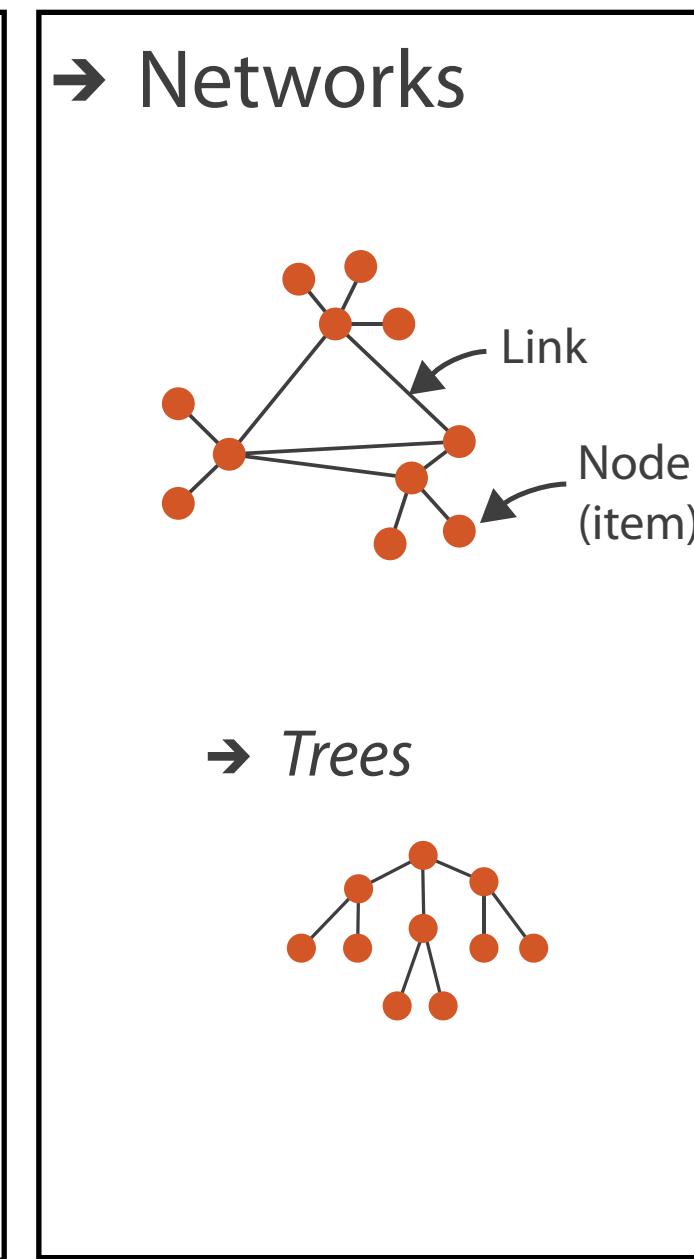
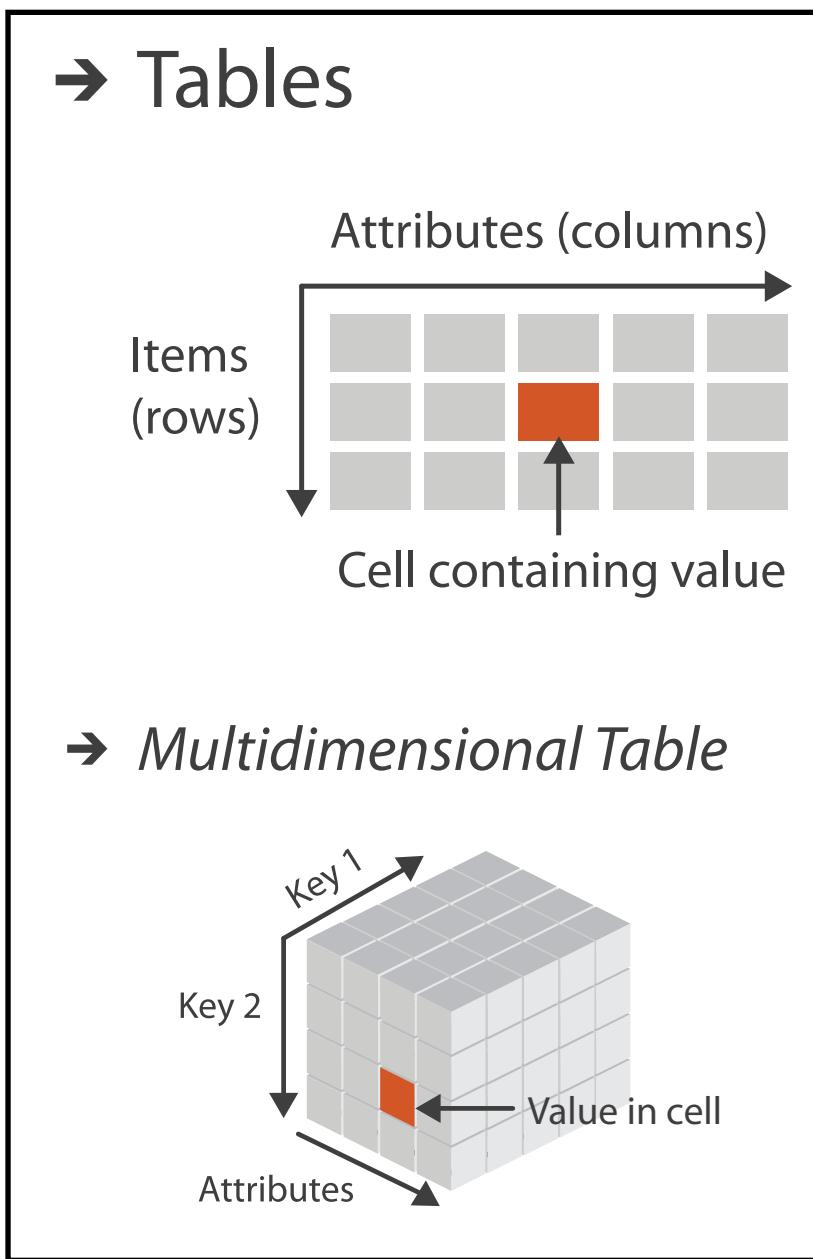
What?

Why?

How?

Three major datatypes

→ Dataset Types



- visualization vs computer graphics
 - geometry is design decision

Dataset and data types

→ Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	

→ Data Types

→ Items → Attributes → Links → Positions → Grids

→ Dataset Availability

→ Static → Dynamic



Attribute types

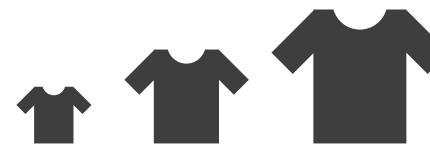
→ Attribute Types

→ Categorical

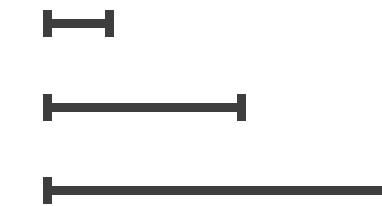


→ Ordered

→ *Ordinal*



→ *Quantitative*



→ Ordering Direction

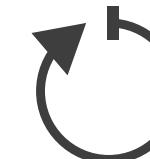
→ Sequential



→ Diverging

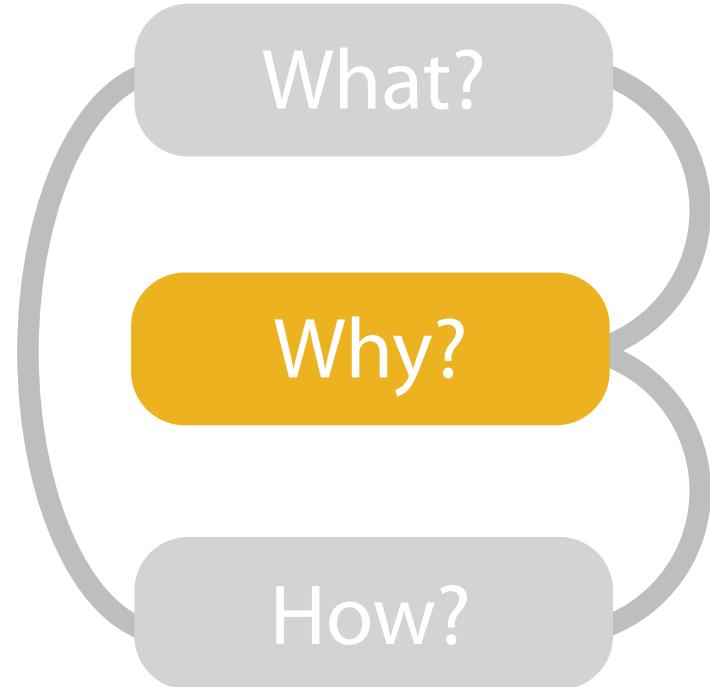


→ Cyclic

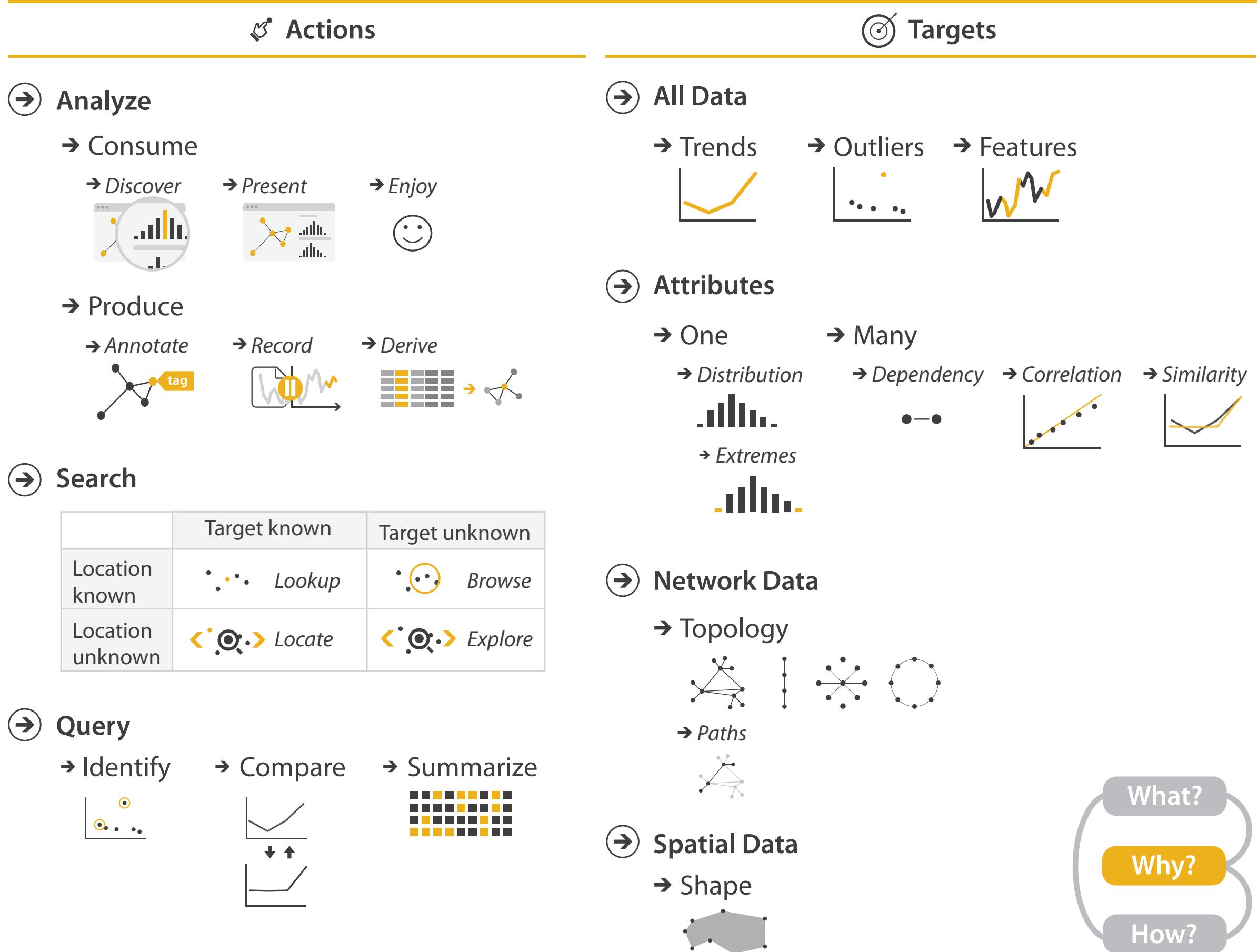


Question:

What do you want to do with your data?

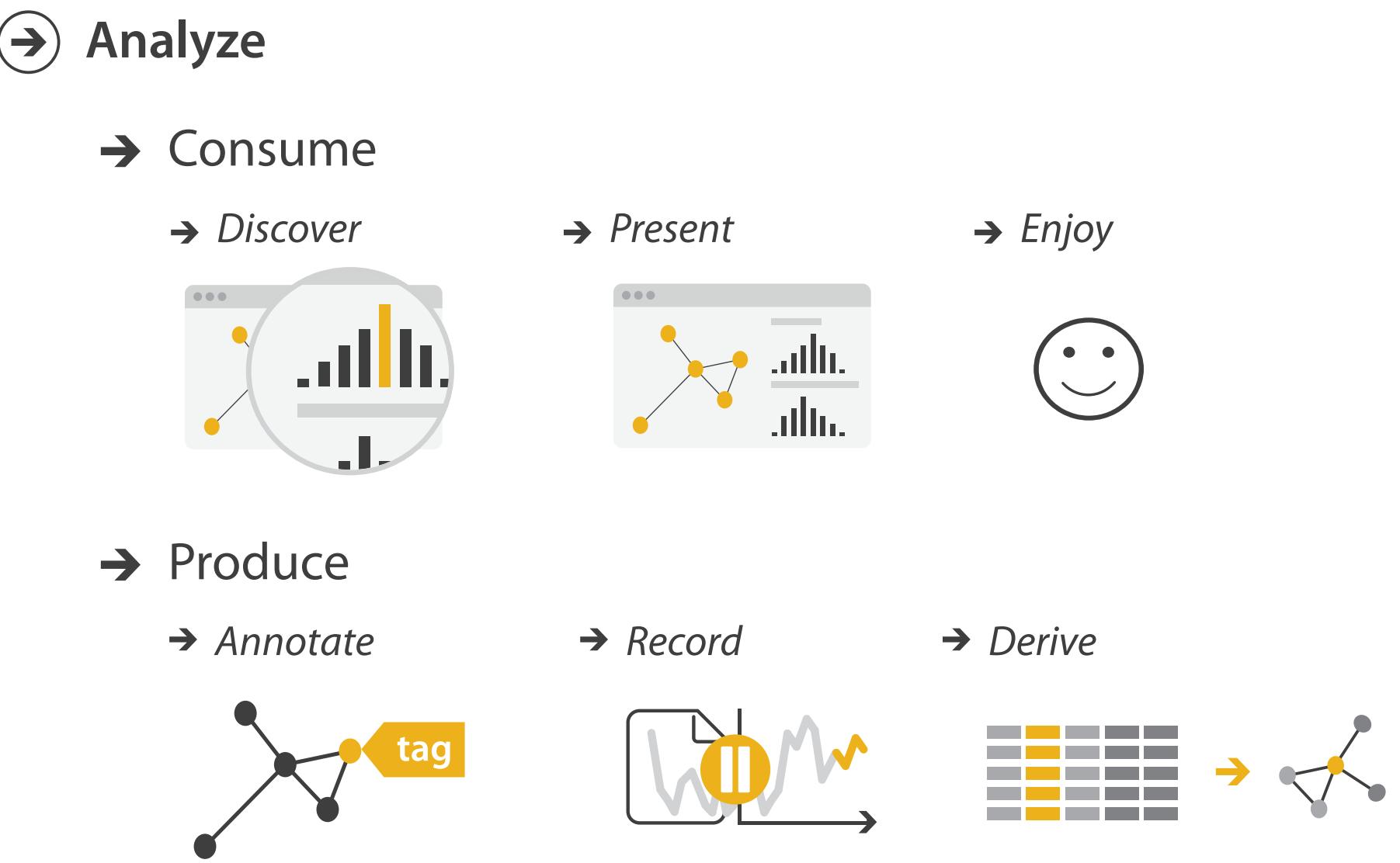


- {action, target} pairs
 - discover distribution
 - compare trends
 - locate outliers
 - browse topology



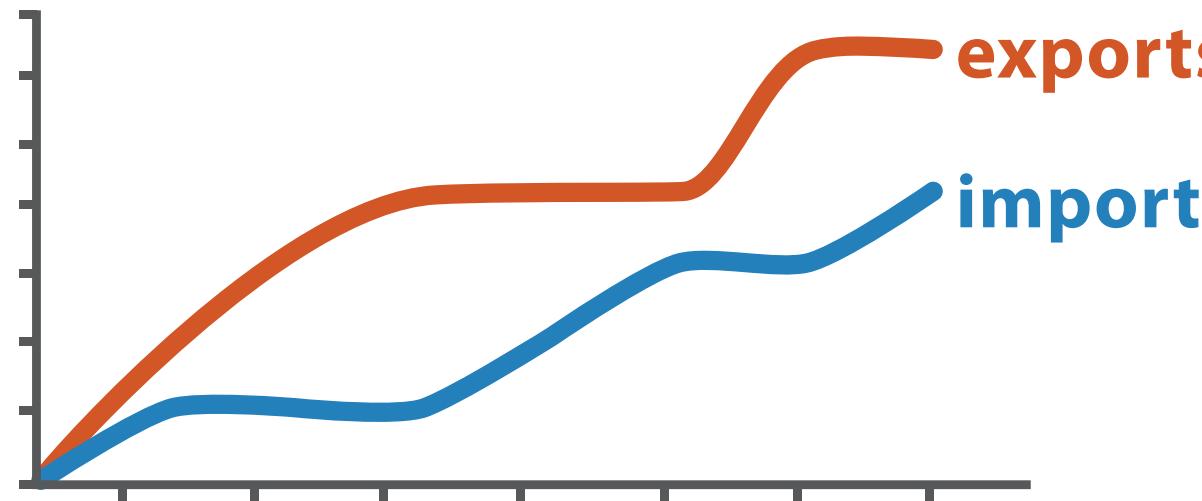
Actions: Analyze

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

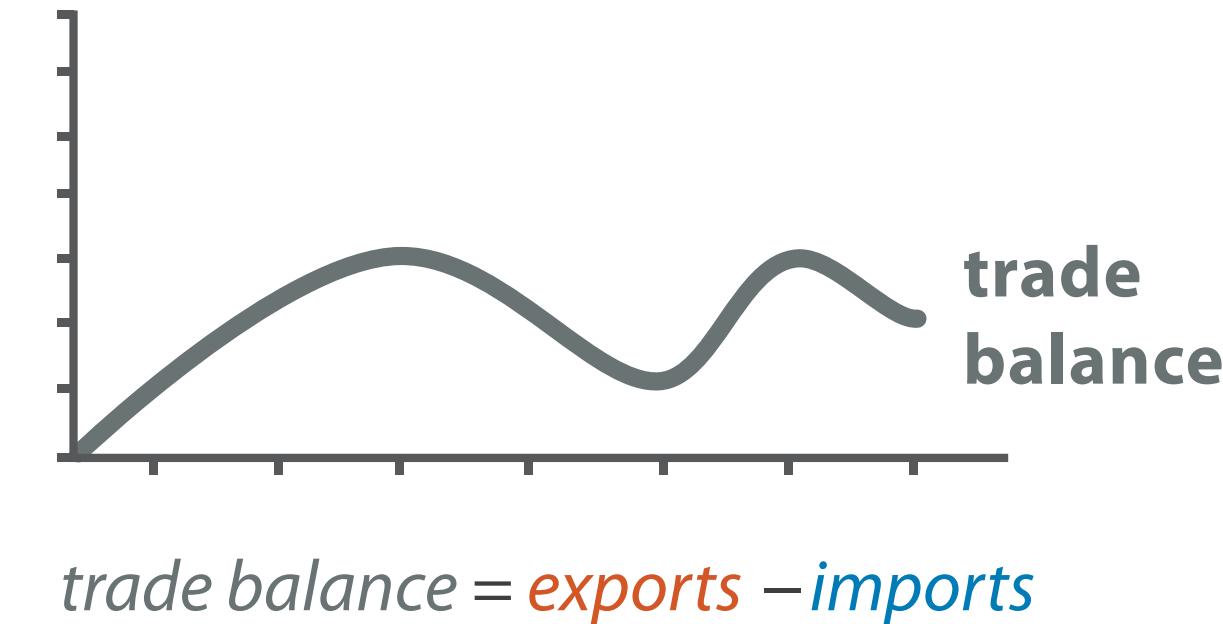


Derive

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



Original Data



Derived Data

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

Actions: Search, query

- what does user know? → Search

– target, location

- how much of the data matters?

– one, some, all

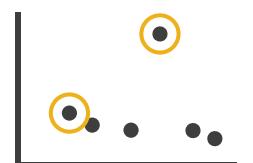
	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

- independent choices for each of these three levels

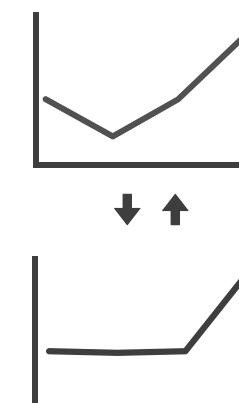
– analyze, search, query
– mix and match

→ Query

→ Identify



→ Compare



→ Summarize



Why: Targets

→ All Data

→ Trends



→ Outliers

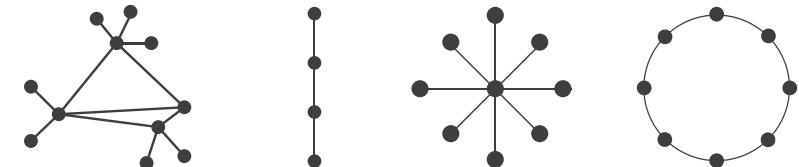


→ Features

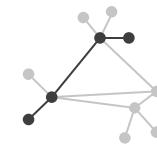


→ Network Data

→ Topology



→ Paths



→ Attributes

→ One

→ Distribution

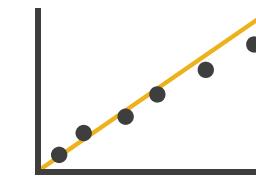


→ Many

→ Dependency



→ Correlation

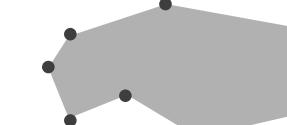


→ Similarity



→ Spatial Data

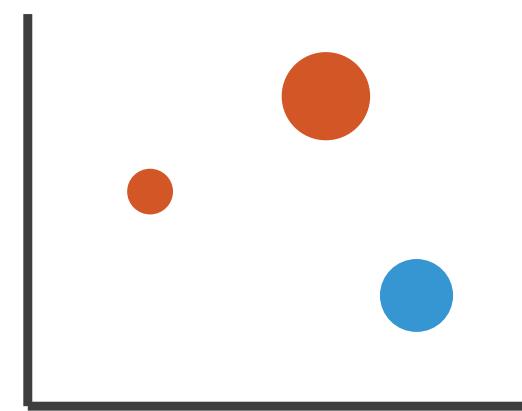
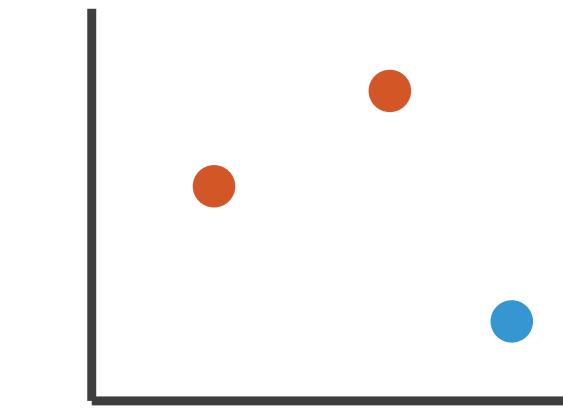
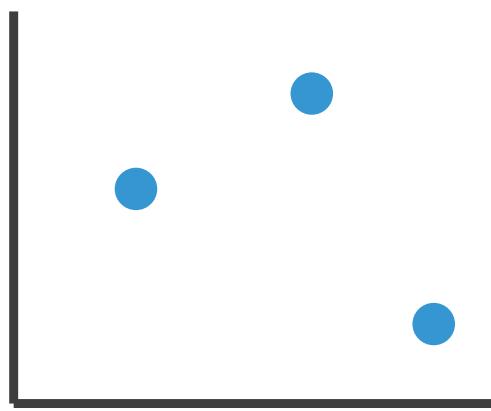
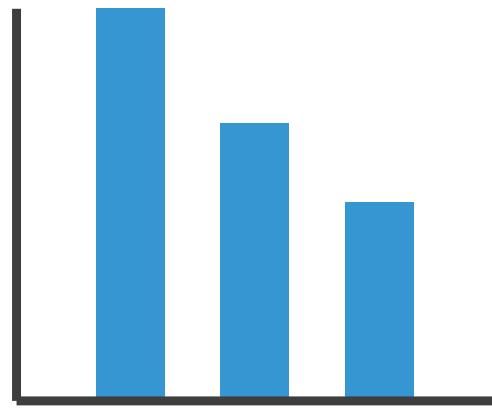
→ Shape



Excercise:
Draw 73 and 34

Visual encoding

- analyze idiom structure

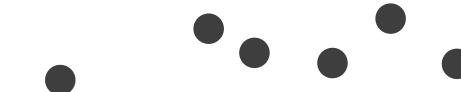


Definitions: Marks and channels

- marks

- geometric primitives

→ Points



→ Lines



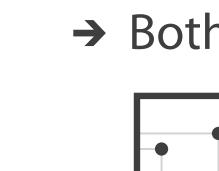
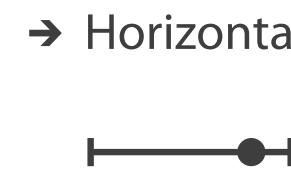
→ Areas



- channels

- control appearance of marks
 - can redundantly code with multiple channels

→ Position



→ Color



→ Shape



→ Tilt



→ Size

→ Length



→ Area

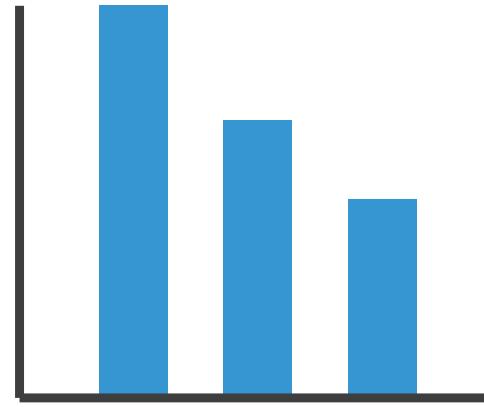


→ Volume



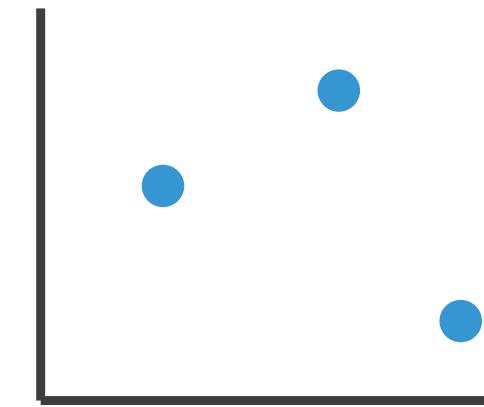
Visual encoding

- analyze idiom structure
 - as combination of marks and channels



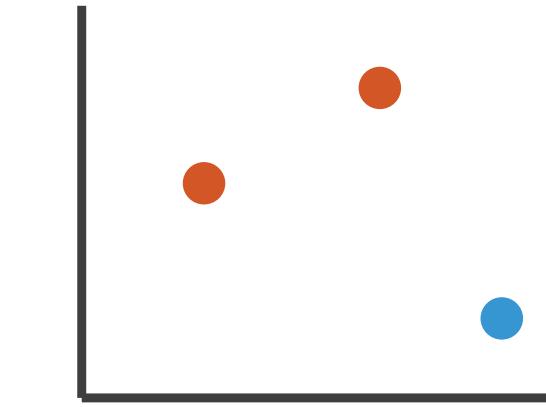
1:
vertical position

mark: line



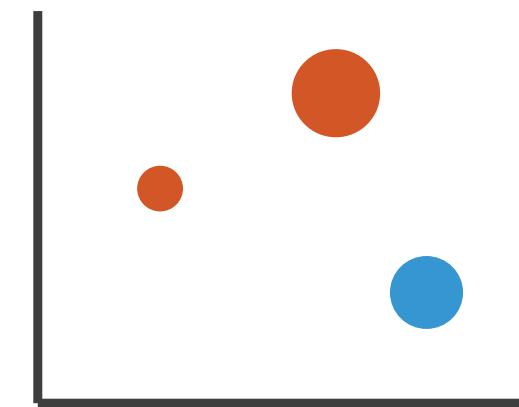
2:
vertical position
horizontal position

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



Same

Color saturation



Same

Curvature



Volume (3D size)



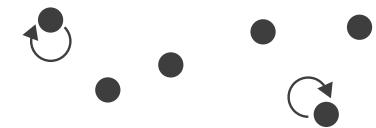
Spatial region



Color hue



Motion



Shape



Channels: Matching Types

→ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt angle



Area (2D size)



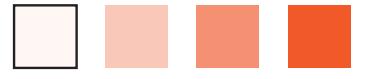
Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



→ Identity Channels: Categorical Attributes

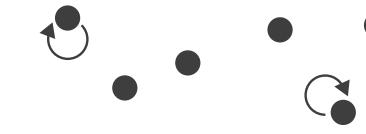
Spatial region



Color hue



Motion



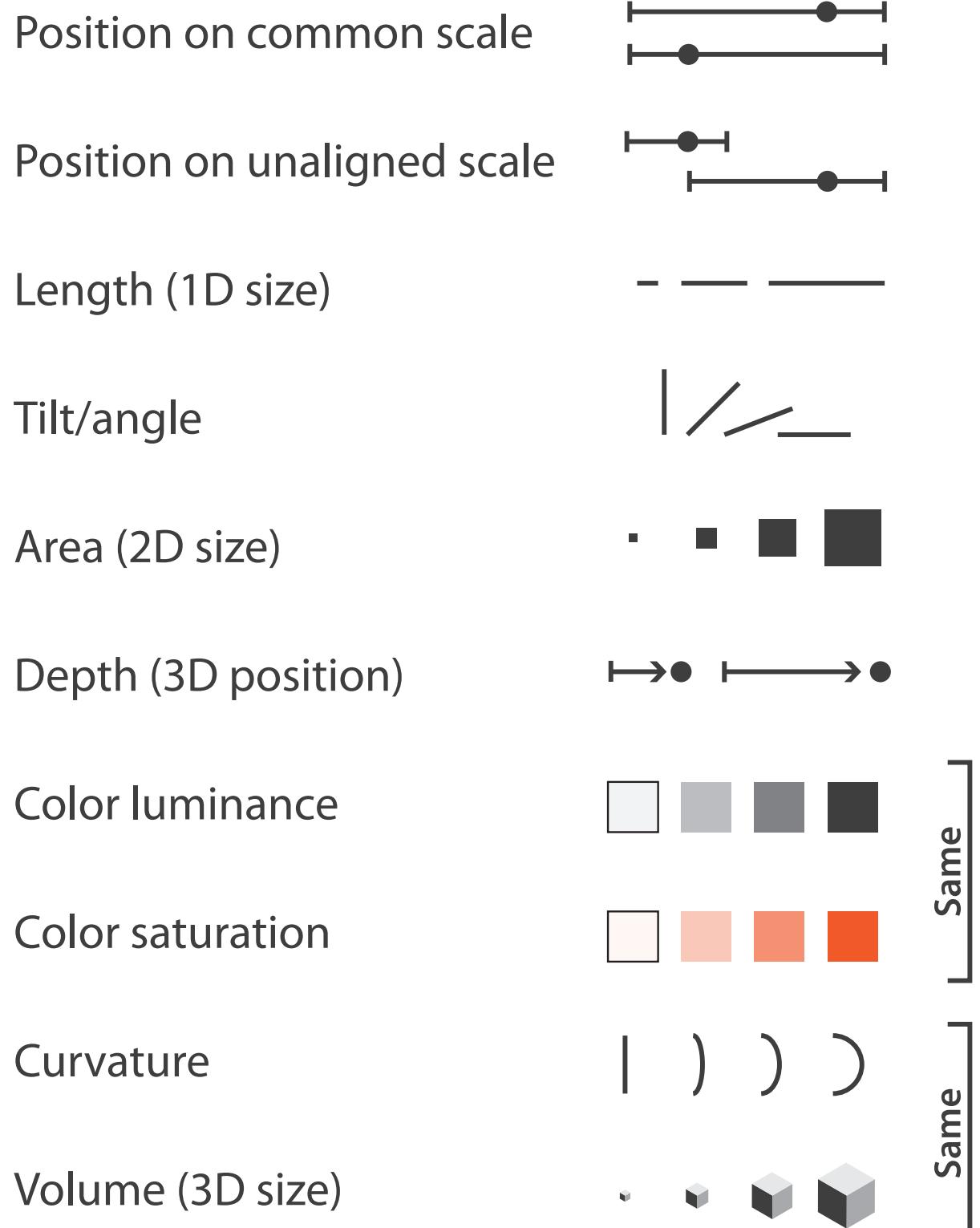
Shape



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

Channels: Rankings

→ Magnitude Channels: Ordered Attributes



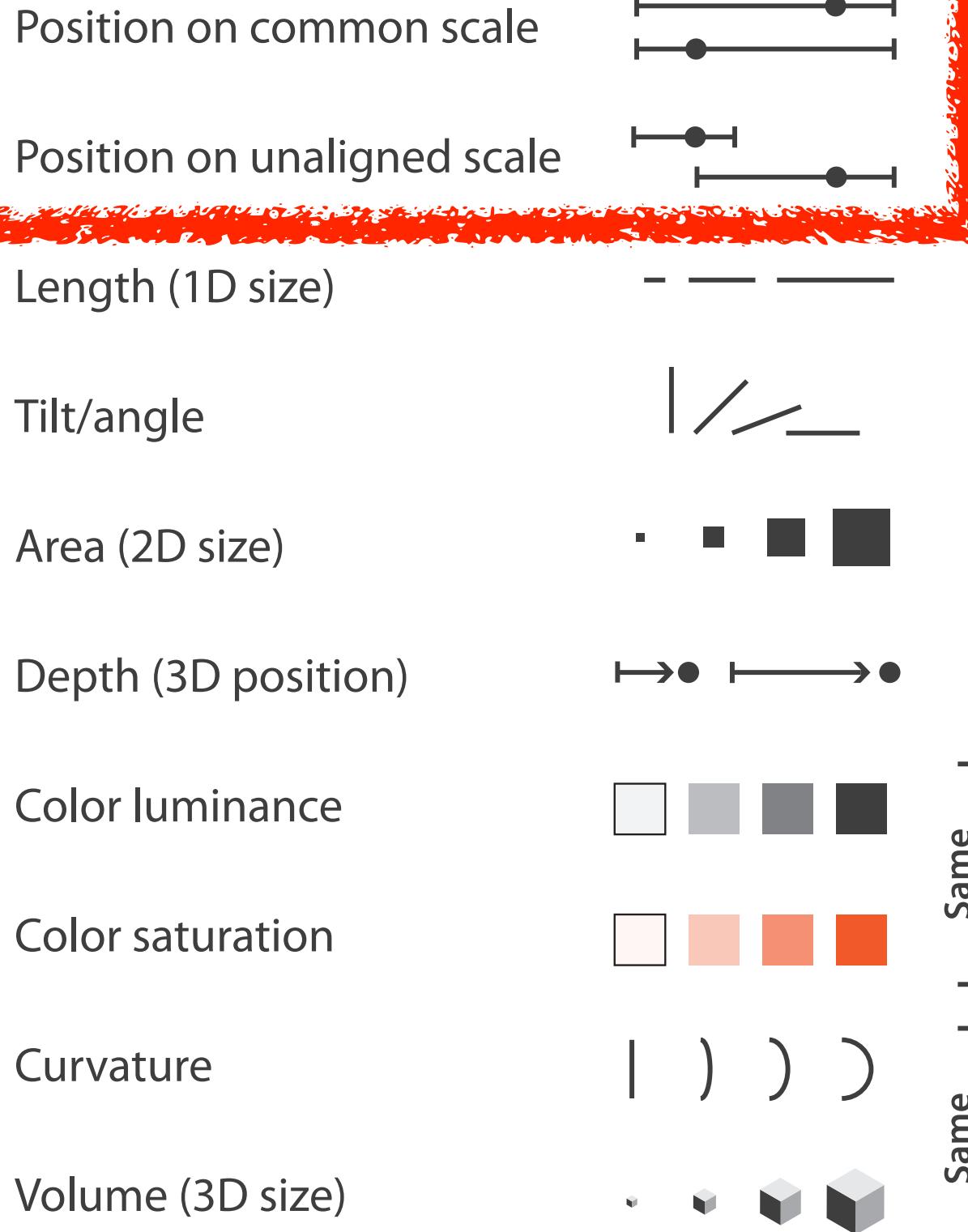
→ Identity Channels: Categorical Attributes



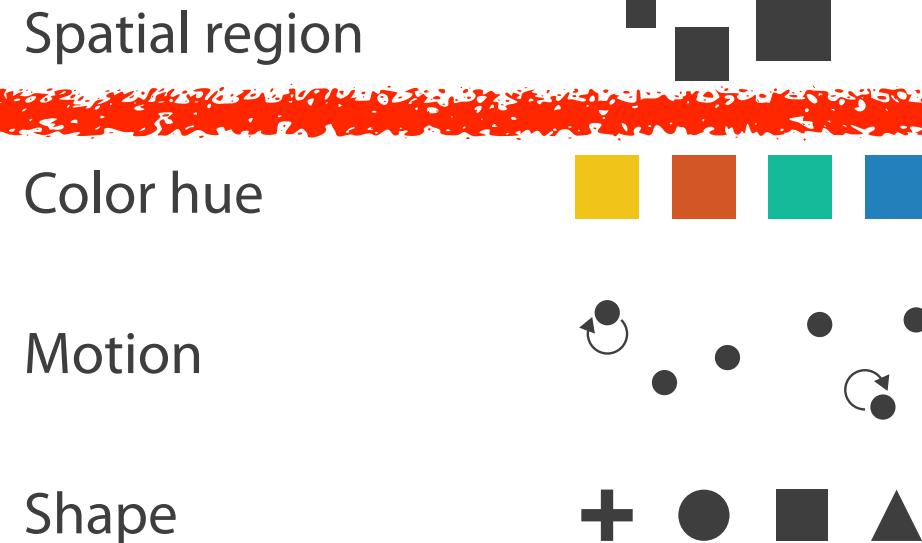
- **expressiveness principle**
 - match channel and data characteristics
- **effectiveness principle**
 - encode most important attributes with highest ranked channels

Channels: Expressiveness types and effectiveness rankings

→ Magnitude Channels: Ordered Attributes



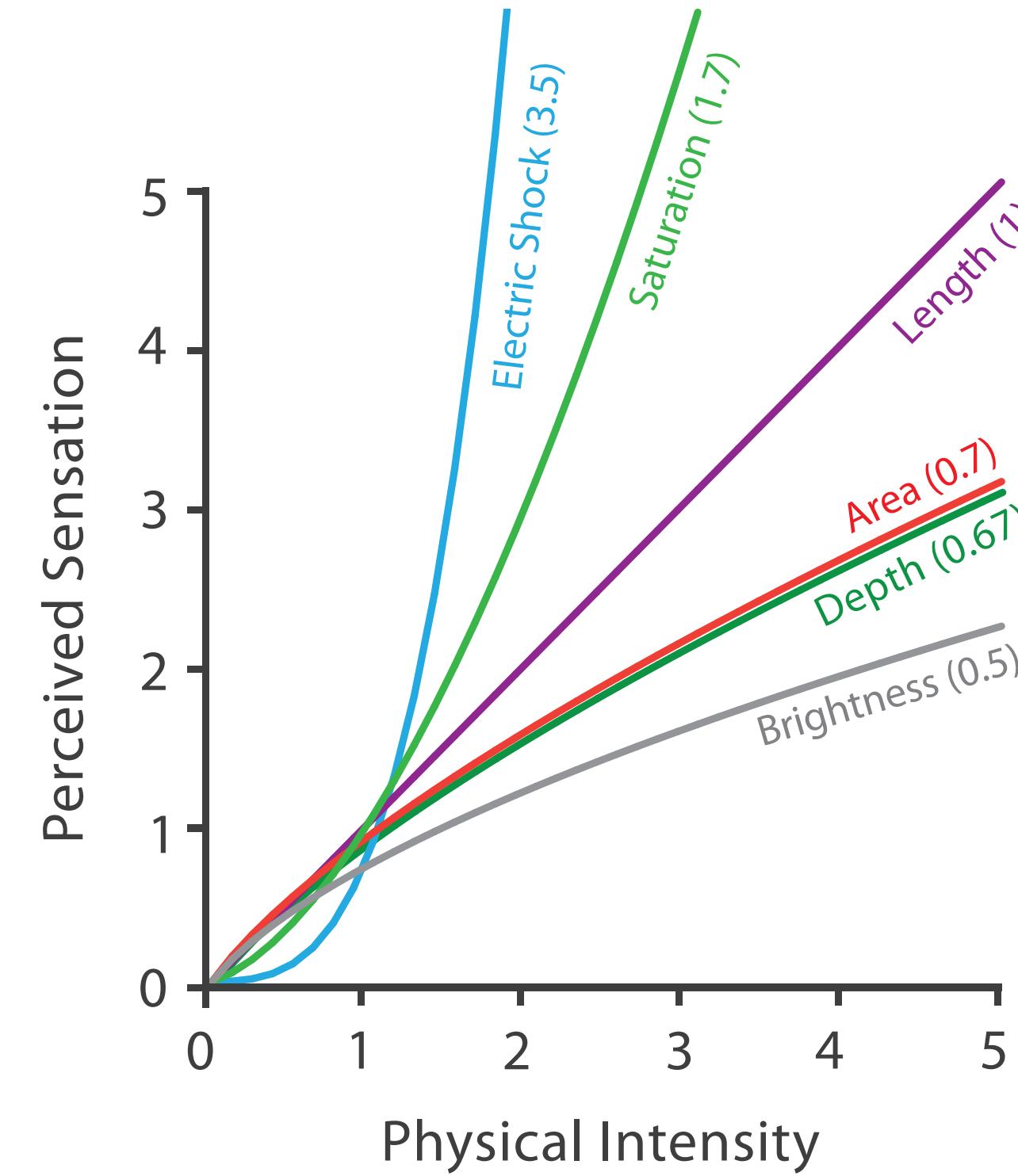
→ Identity Channels: Categorical Attributes



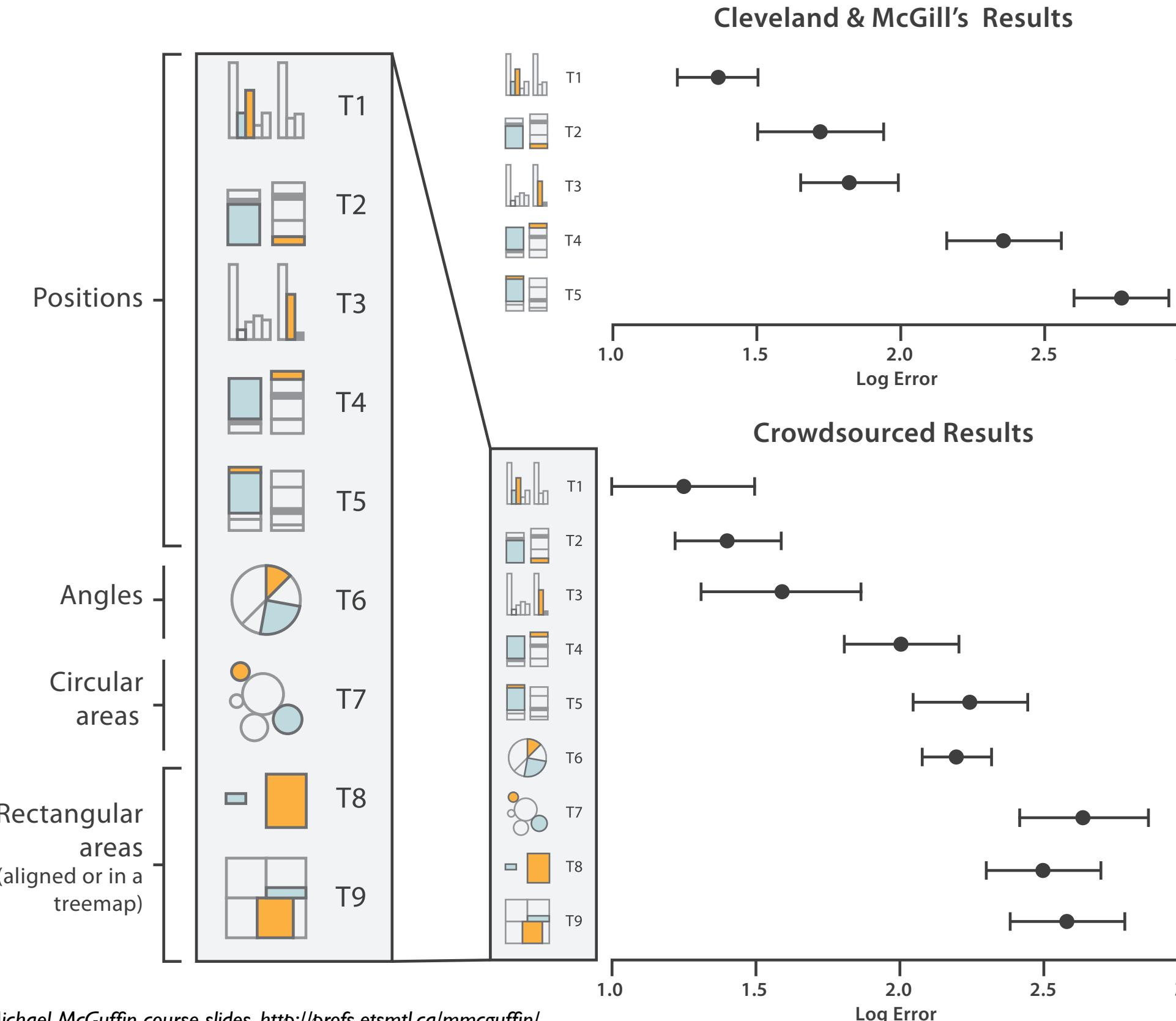
- **expressiveness principle**
 - match channel and data characteristics
- **effectiveness principle**
 - encode most important attributes with highest ranked channels
 - spatial position ranks high for both

Accuracy: Fundamental Theory

Steven's Psychophysical Power Law: $S = I^n$



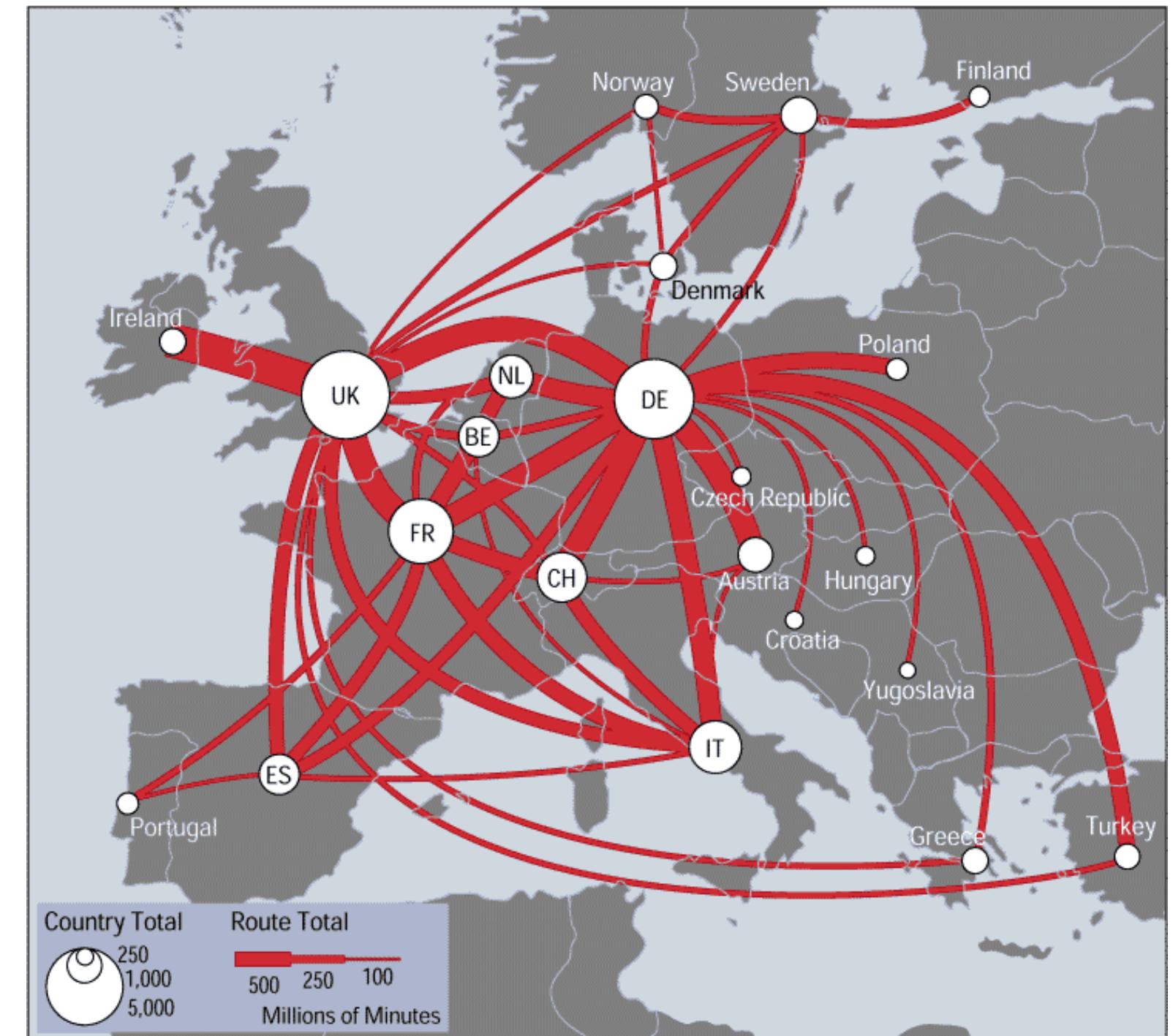
Accuracy: Vis experiments



[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design.
Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.]

Discriminability: How many usable steps?

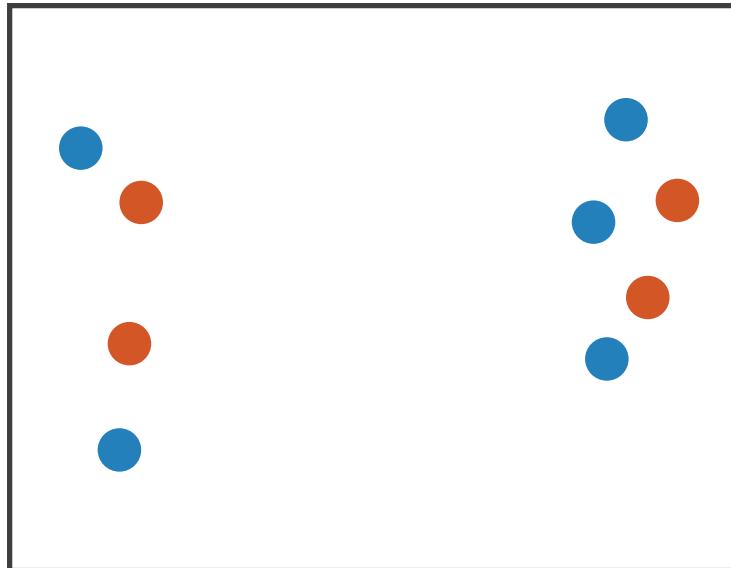
- must be sufficient for number of attribute levels to show
 - linewidth: few bins



[mappa.mundi.net/maps/maps_014/telegeography.html]

Separability vs. Integrality

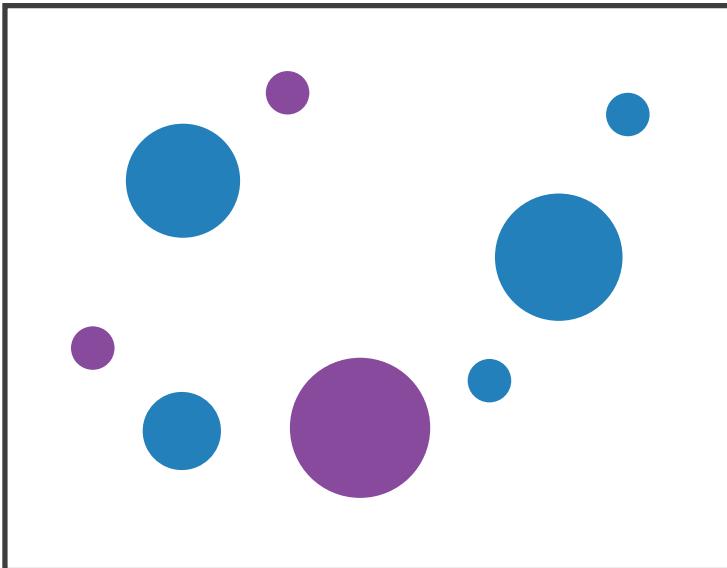
Position
+ Hue (Color)



Fully separable

2 groups each

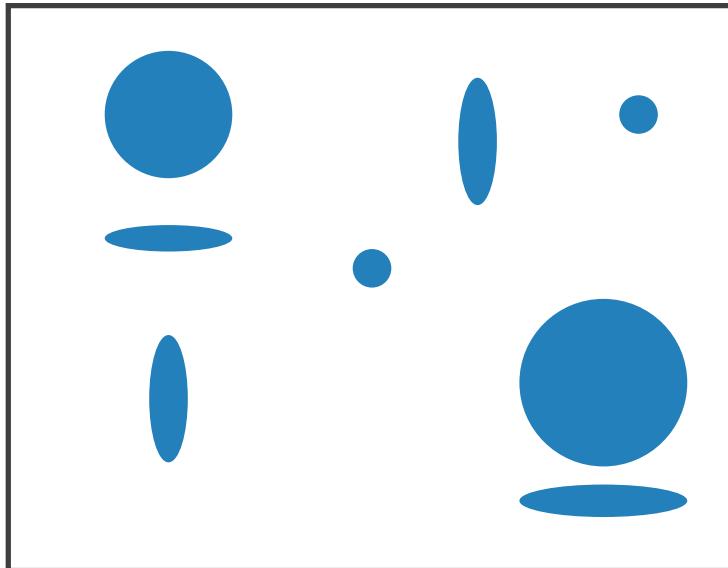
Size
+ Hue (Color)



Some interference

2 groups each

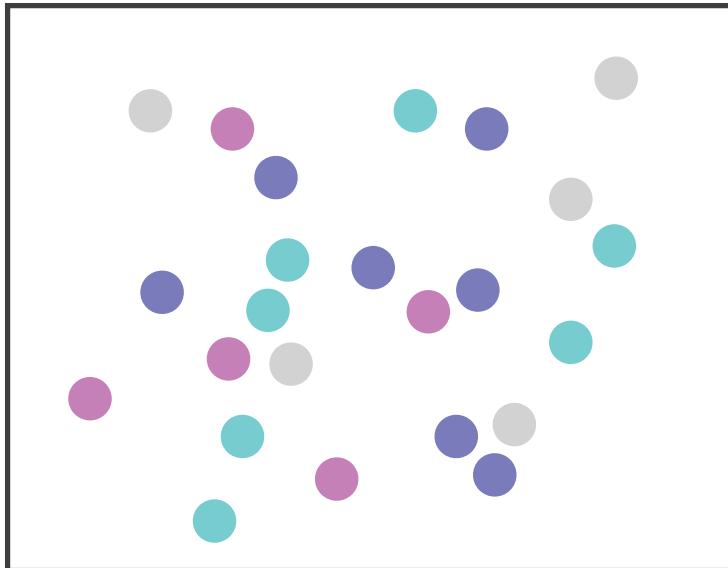
Width
+ Height



Some/significant
interference

3 groups total:
integral area

Red
+ Green

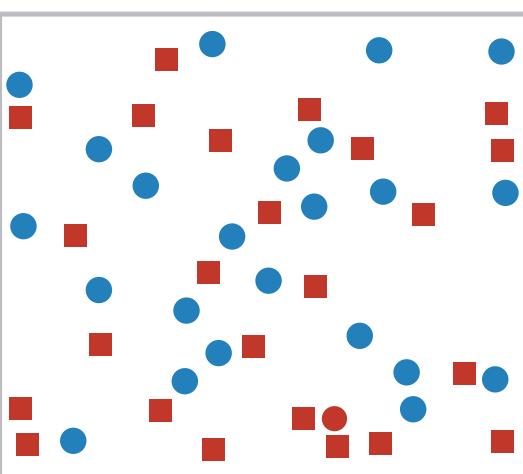
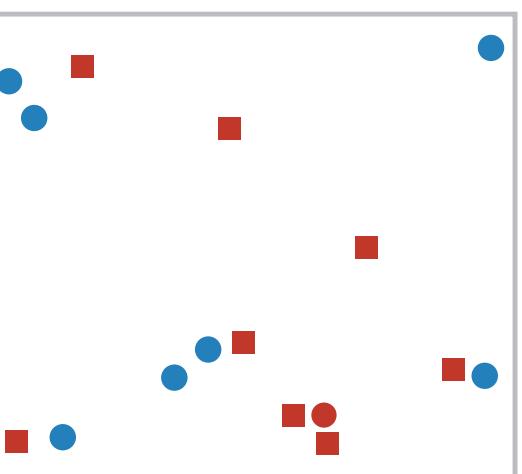
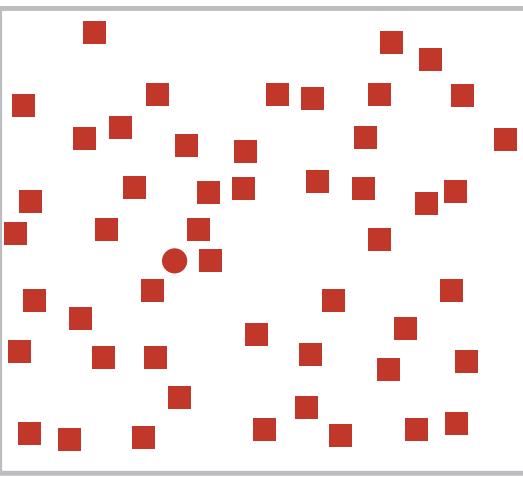
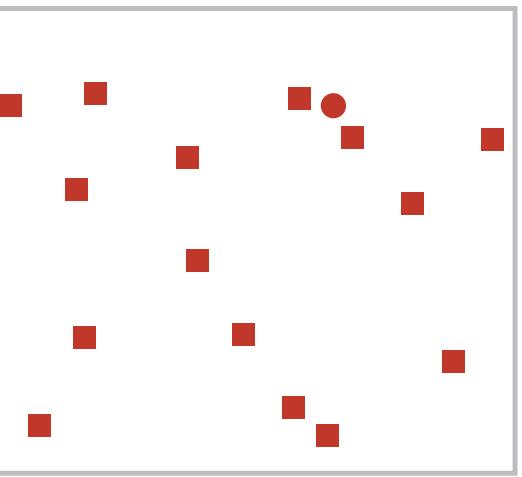
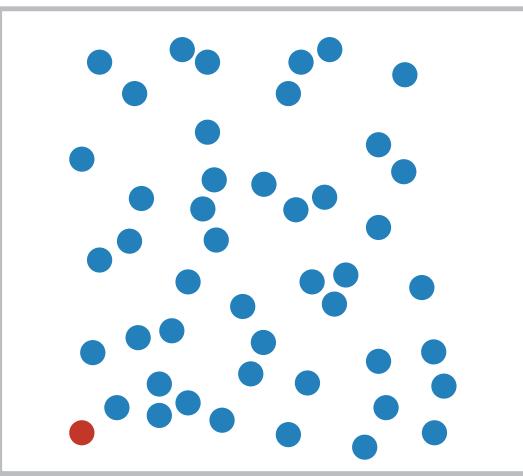
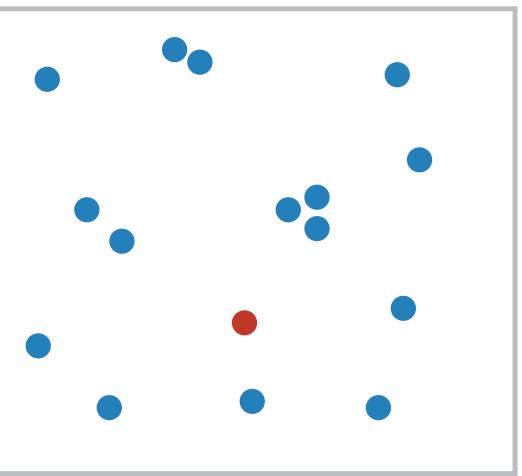


Major interference

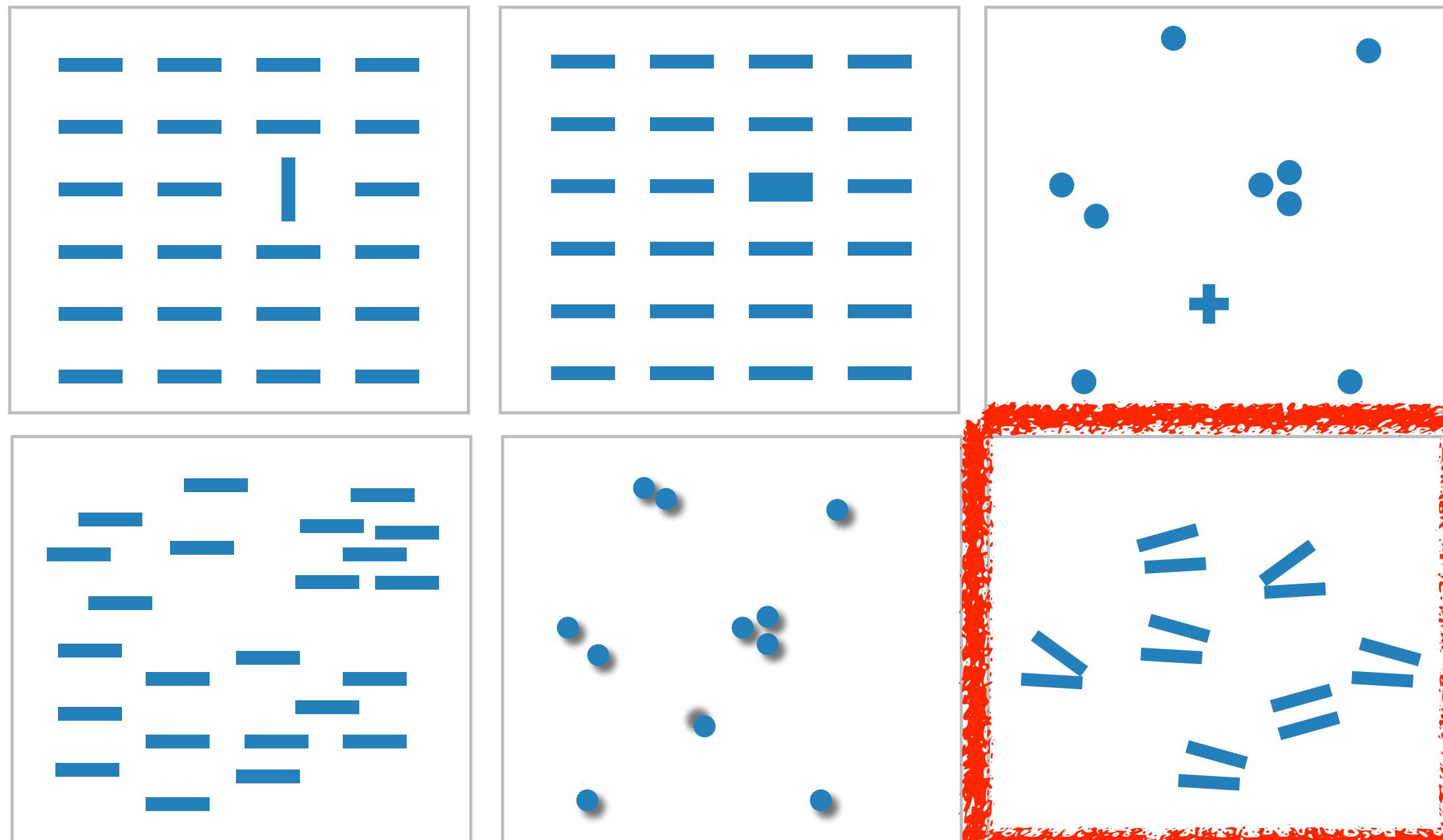
4 groups total:
integral hue

Popout

- find the red dot
 - how long does it take?
- parallel processing on many individual channels
 - speed independent of distractor count
 - speed depends on channel and amount of difference from distractors
- serial search for (almost all) combinations
 - speed depends on number of distractors



Popout



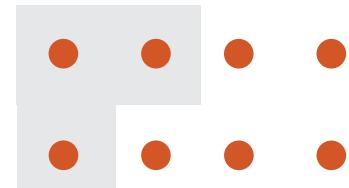
- many channels: tilt, size, shape, proximity, shadow direction, ...
- but not all! parallel line pairs do not pop out from tilted pairs

Grouping

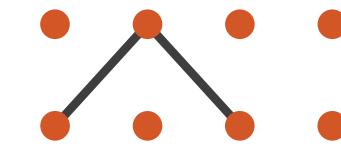
- containment
- connection
- proximity
 - same spatial region
- similarity
 - same values as other categorical channels

Marks as Links

→ Containment



→ Connection



→ Identity Channels: Categorical Attributes

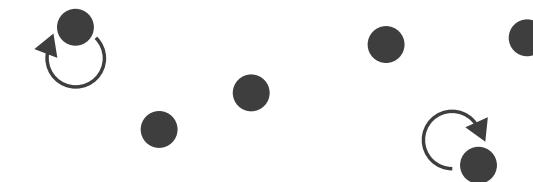
Spatial region



Color hue



Motion



Shape

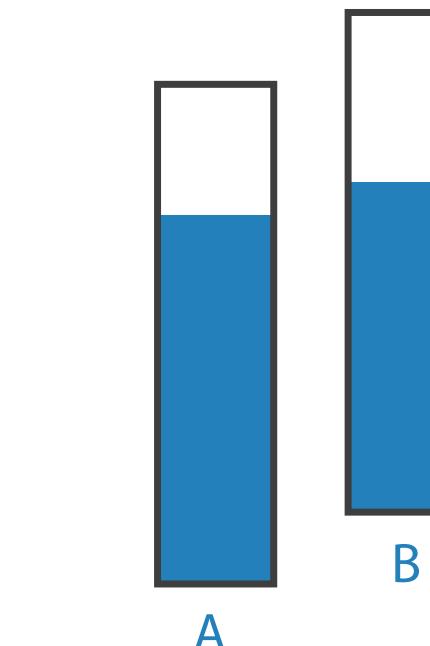


Relative vs. absolute judgements

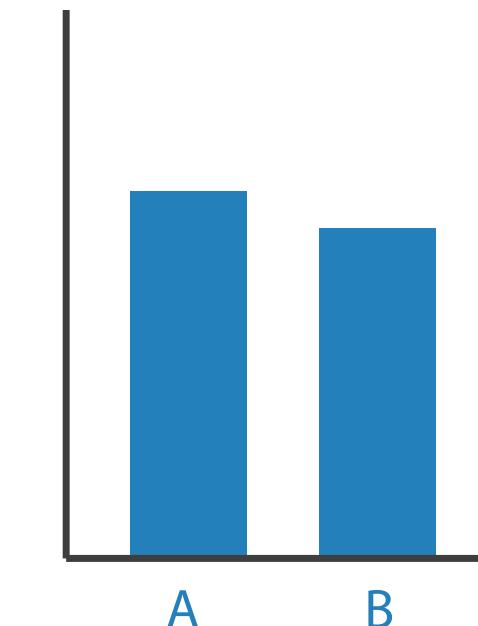
- perceptual system mostly operates with relative judgements, not absolute
 - that's why accuracy increases with common frame/scale and alignment
 - Weber's Law: ratio of increment to background is constant
 - filled rectangles differ in length by 1:9, difficult judgement
 - white rectangles differ in length by 1:2, easy judgement



length



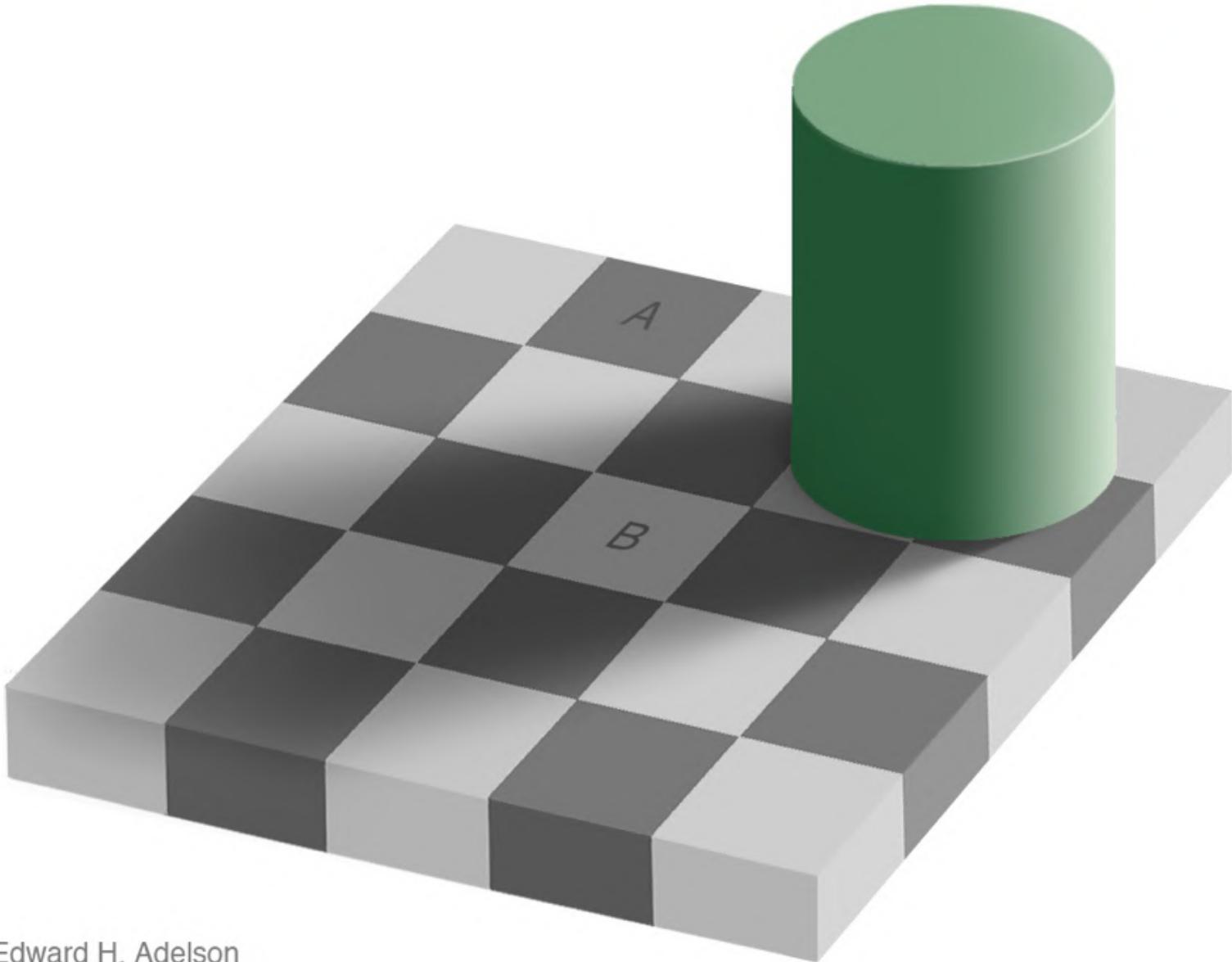
position along
unaligned
common scale



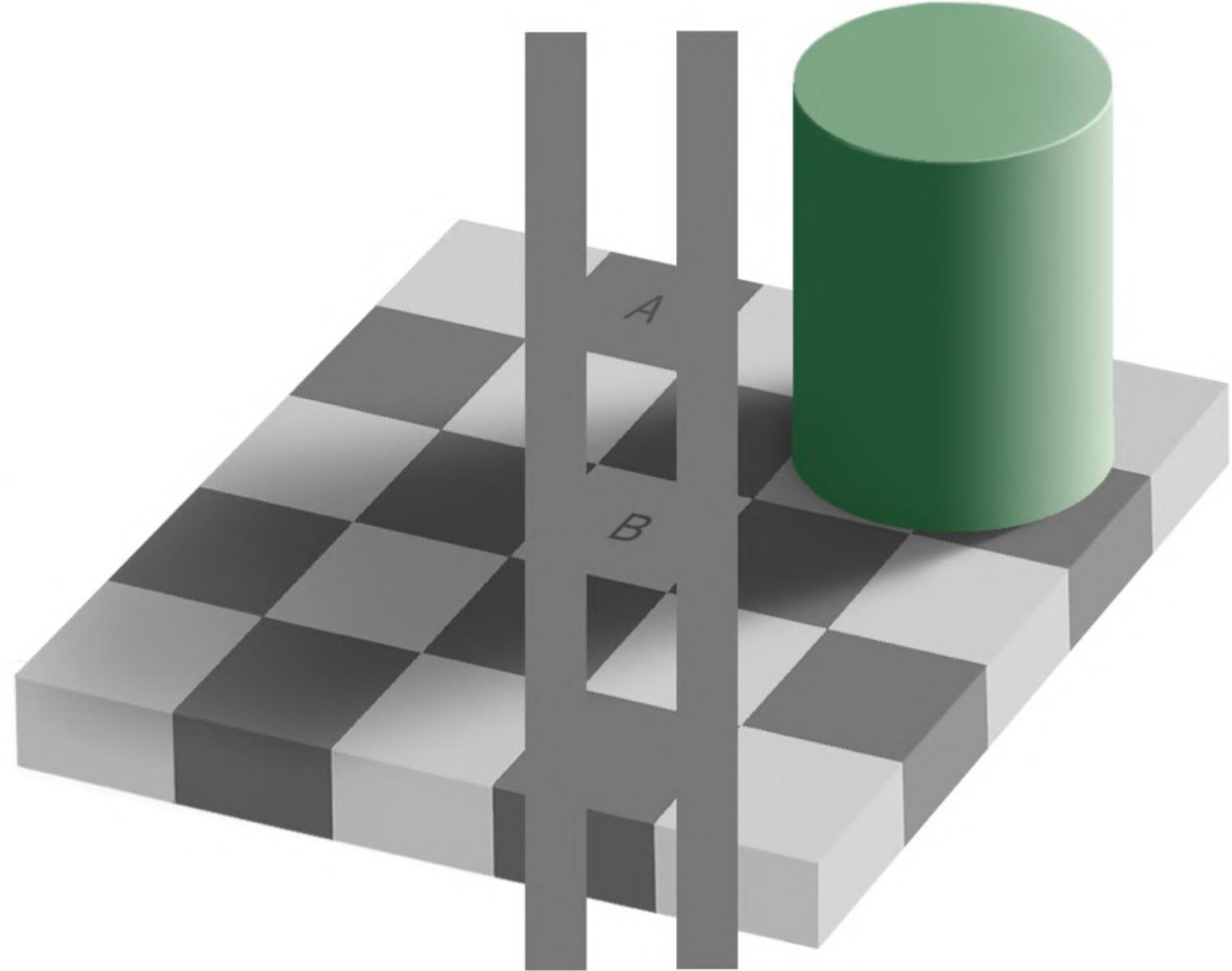
position along
aligned scale

Relative luminance judgements

- perception of luminance is contextual based on contrast with surroundings

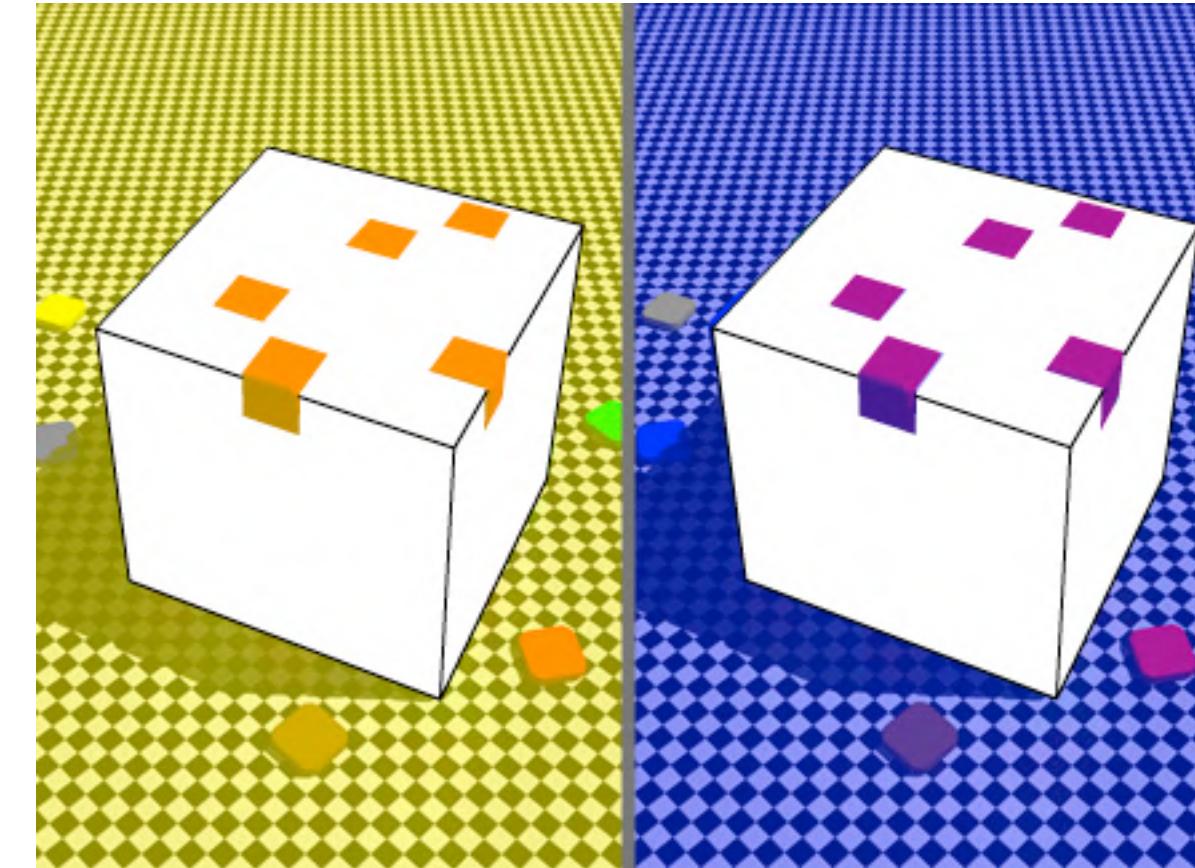
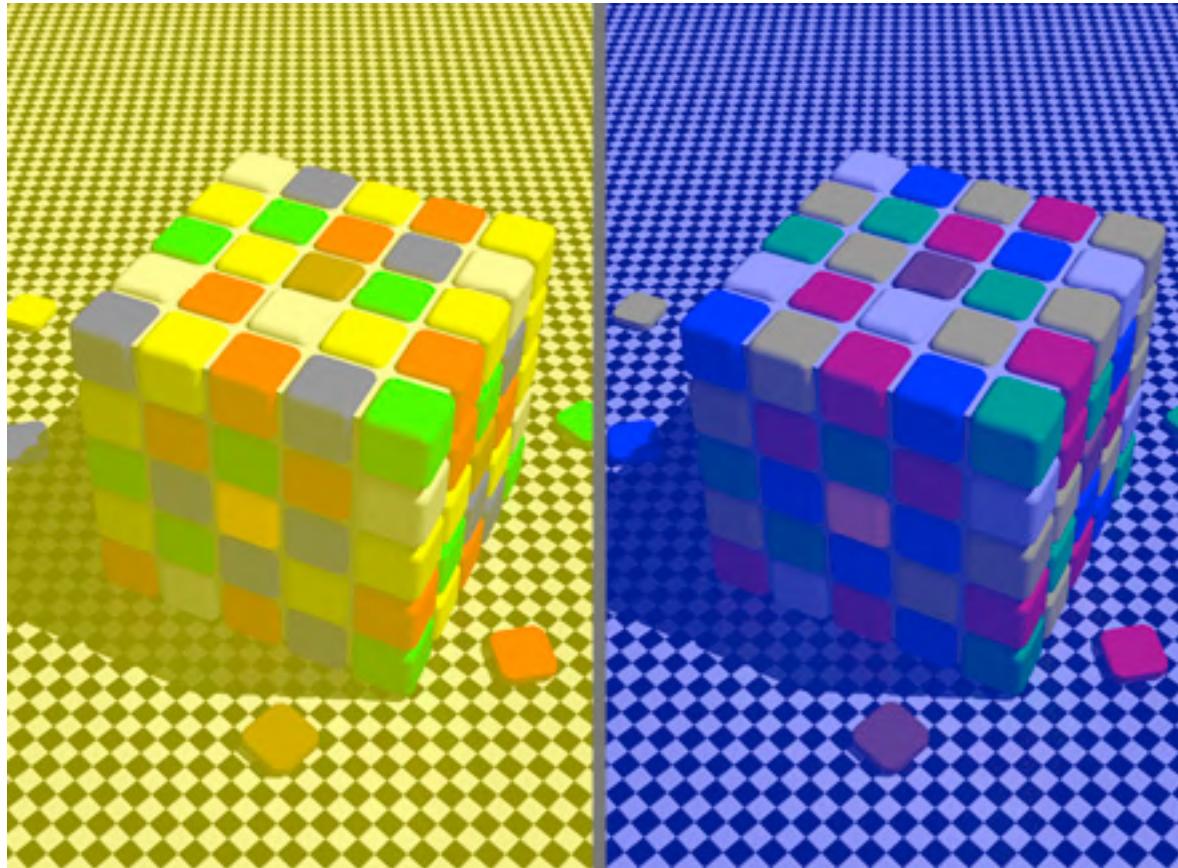


Edward H. Adelson

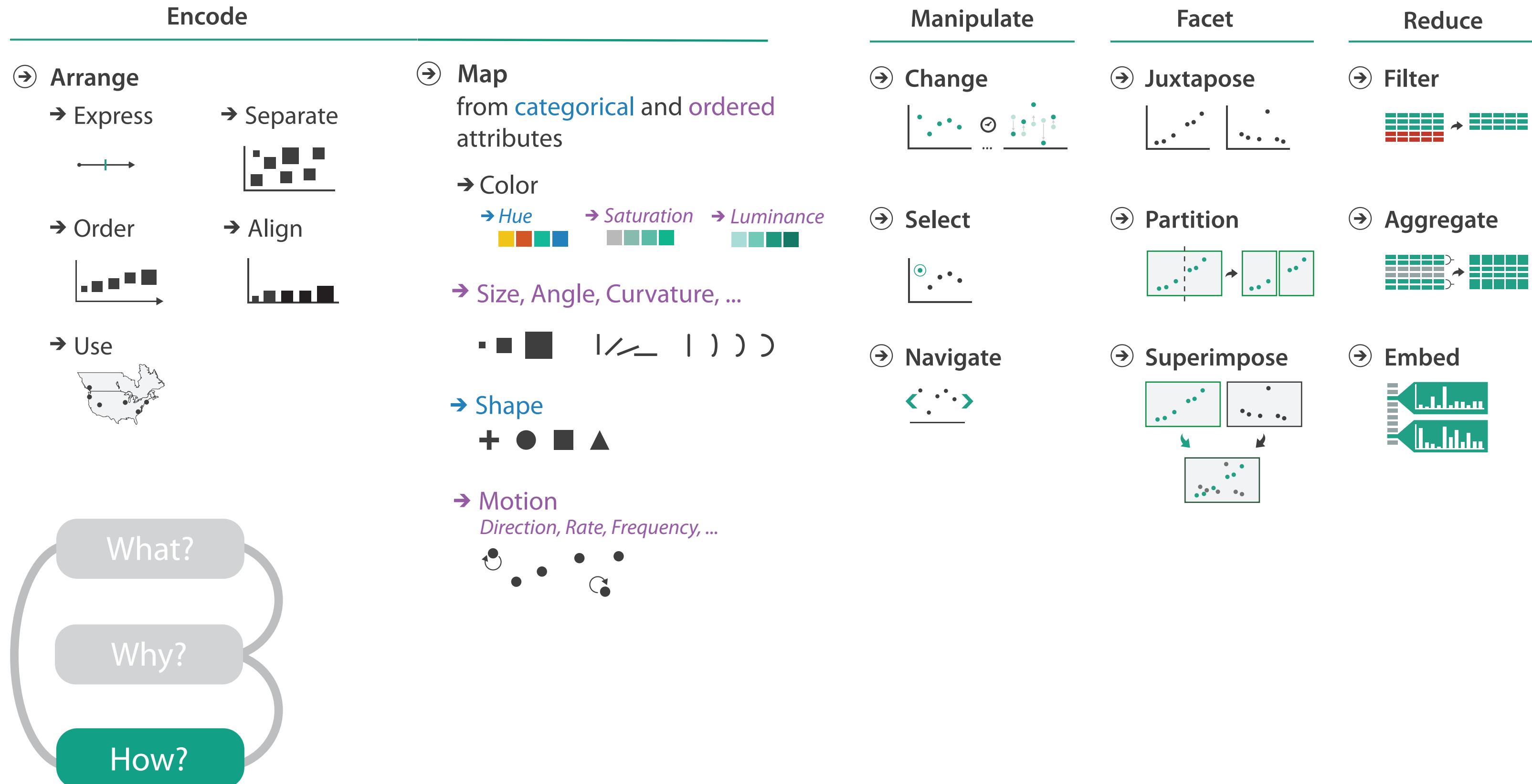


Relative color judgements

- color constancy across broad range of illumination conditions



How?



Encode tables: Arrange space

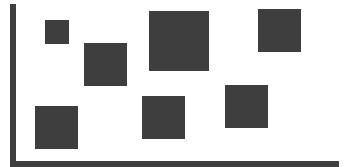
Encode

→ Arrange

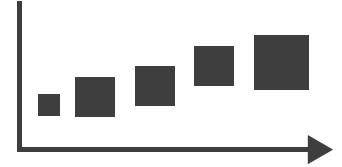
→ Express



→ Separate



→ Order



→ Align



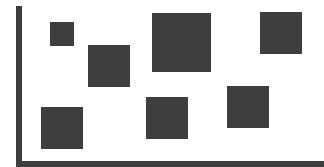
Arrange tables

→ Express Values

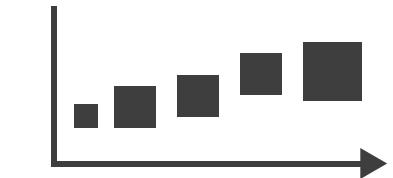


→ Separate, Order, Align Regions

→ Separate



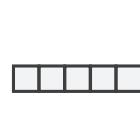
→ Order



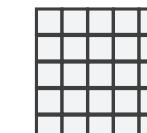
→ Align



→ 1 Key
List



→ 2 Keys
Matrix

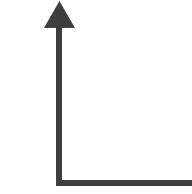


→ 3 Keys
Volume

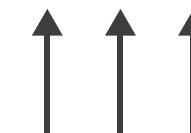


→ Axis Orientation

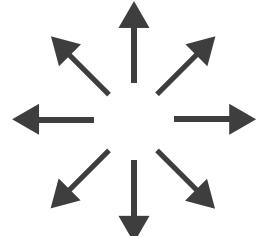
→ Rectilinear



→ Parallel

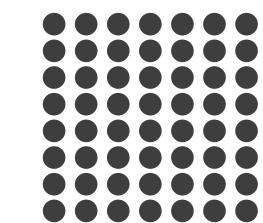


→ Radial



→ Layout Density

→ Dense



→ Space-Filling

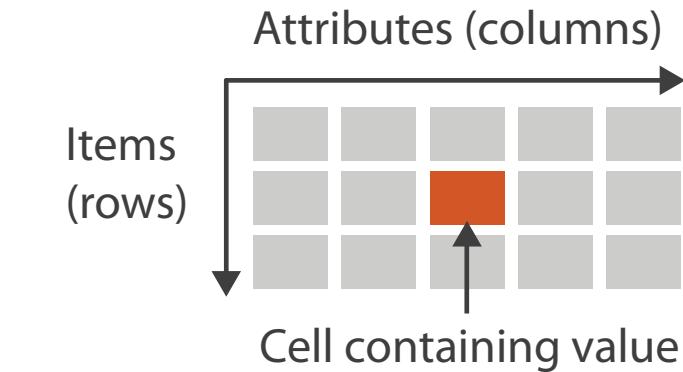


→ Many Keys
Recursive Subdivision

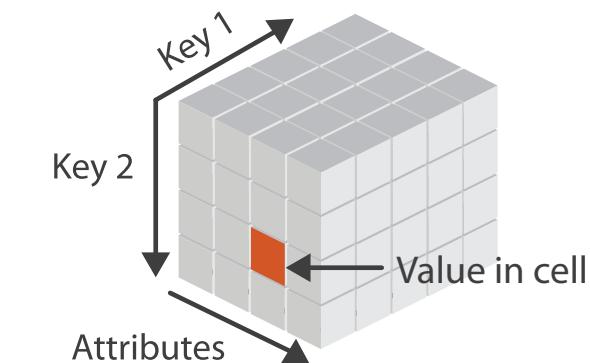


Keys and values

→ Tables



→ *Multidimensional Table*



- **key**
 - independent attribute
 - used as unique index to look up items
 - simple tables: 1 key
 - multidimensional tables: multiple keys
- **value**
 - dependent attribute, value of cell
- **classify arrangements by key count**
 - 0, 1, 2, many...

→ Express Values

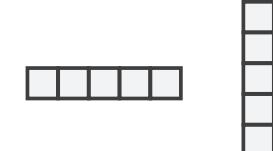
→ 1 Key

List



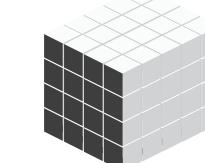
→ 2 Keys

Matrix



→ 3 Keys

Volume



→ Many Keys

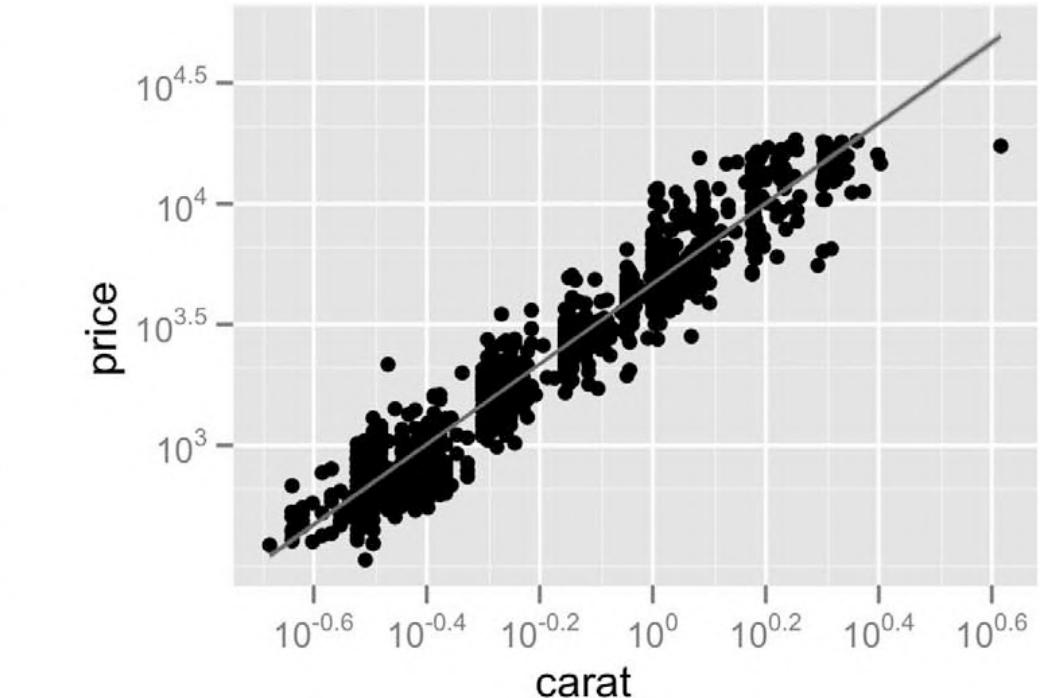
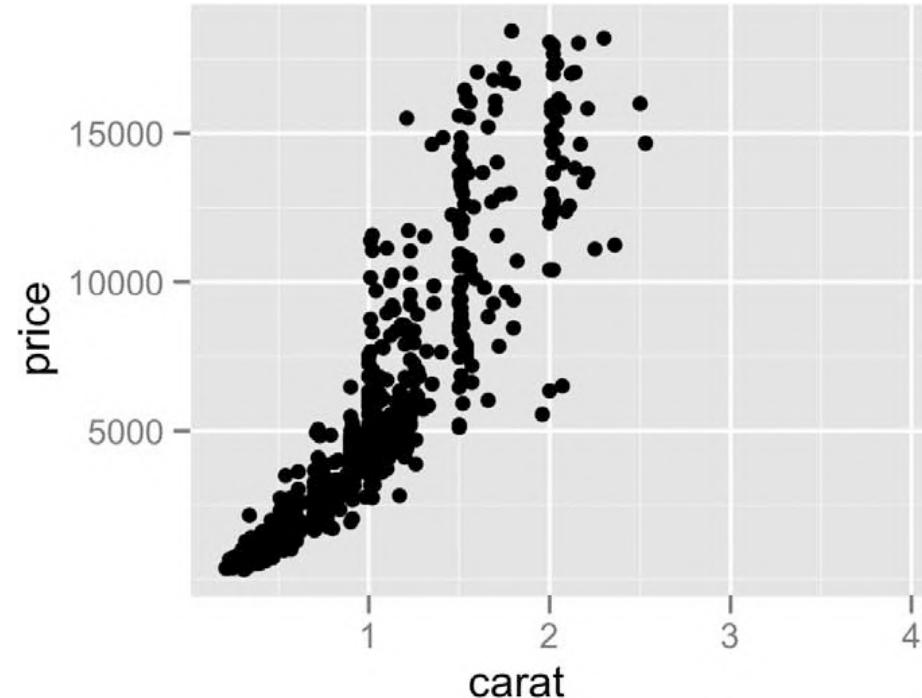
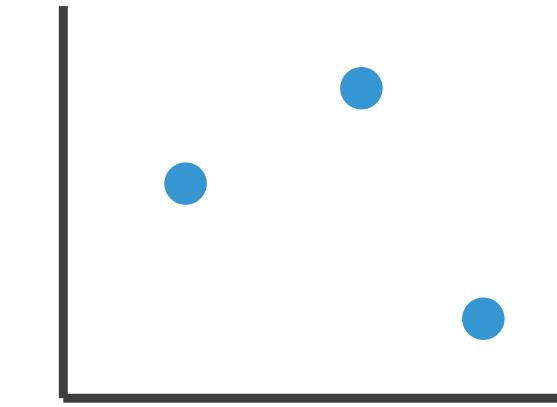
Recursive Subdivision



Idiom: scatterplot

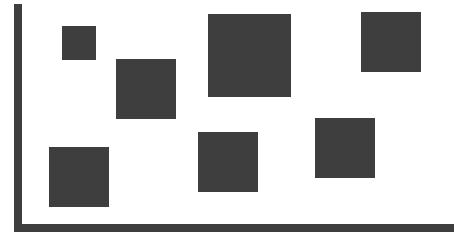
→ Express Values

- **express** values
 - quantitative attributes
- no keys, only values
 - data
 - 2 quant attrs
 - mark: points
 - channels
 - horiz + vert position
 - tasks
 - find trends, outliers, distribution, correlation, clusters
 - scalability
 - hundreds of items

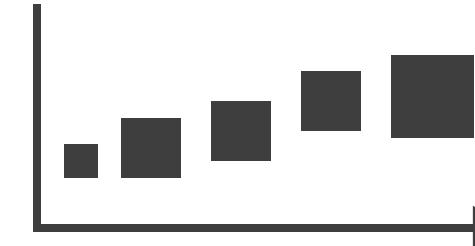


Some keys: Categorical regions

→ Separate



→ Order



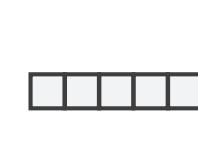
→ Align



- **regions**: contiguous bounded areas distinct from each other
 - using space to **separate** (proximity)
 - following expressiveness principle for categorical attributes
- use ordered attribute to **order** and **align** regions

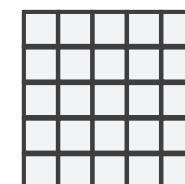
→ 1 Key

List



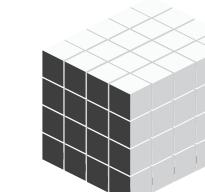
→ 2 Keys

Matrix



→ 3 Keys

Volume



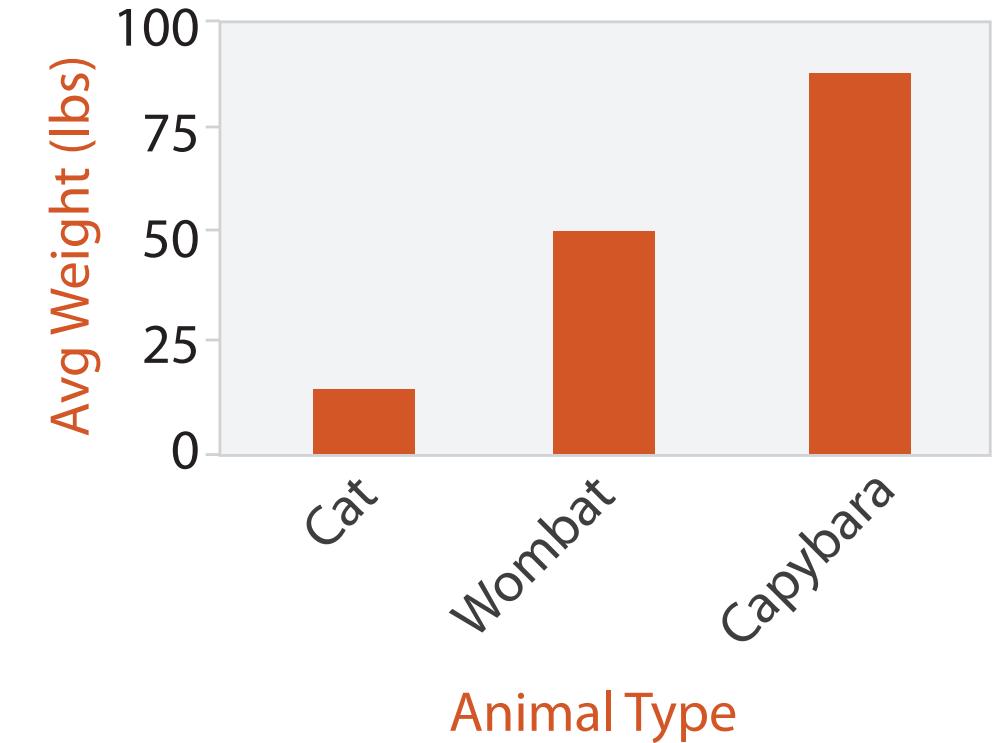
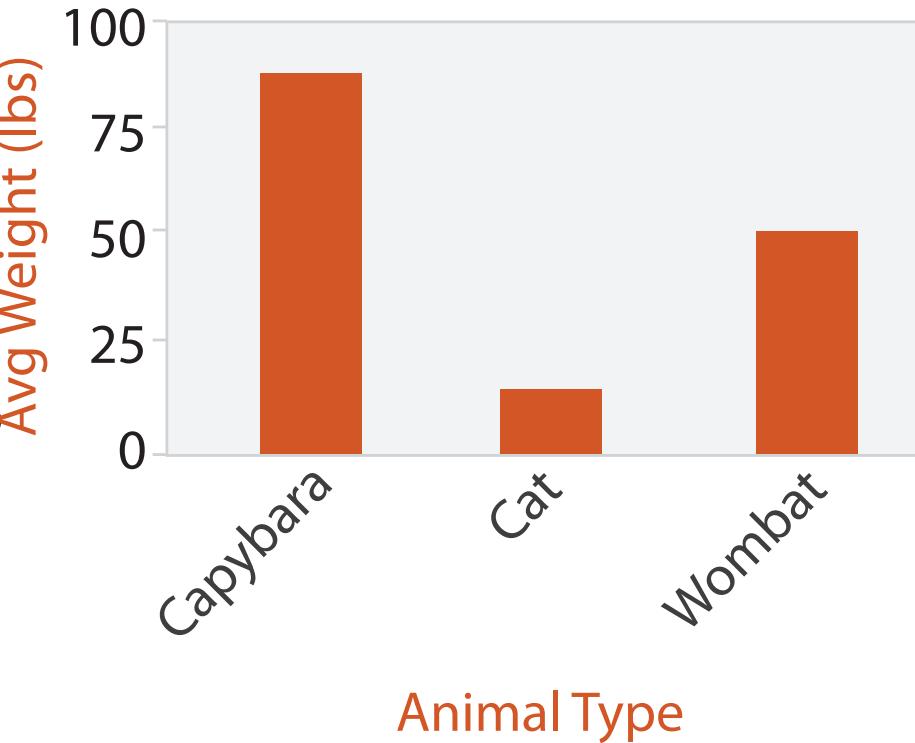
→ Many Keys

Recursive Subdivision



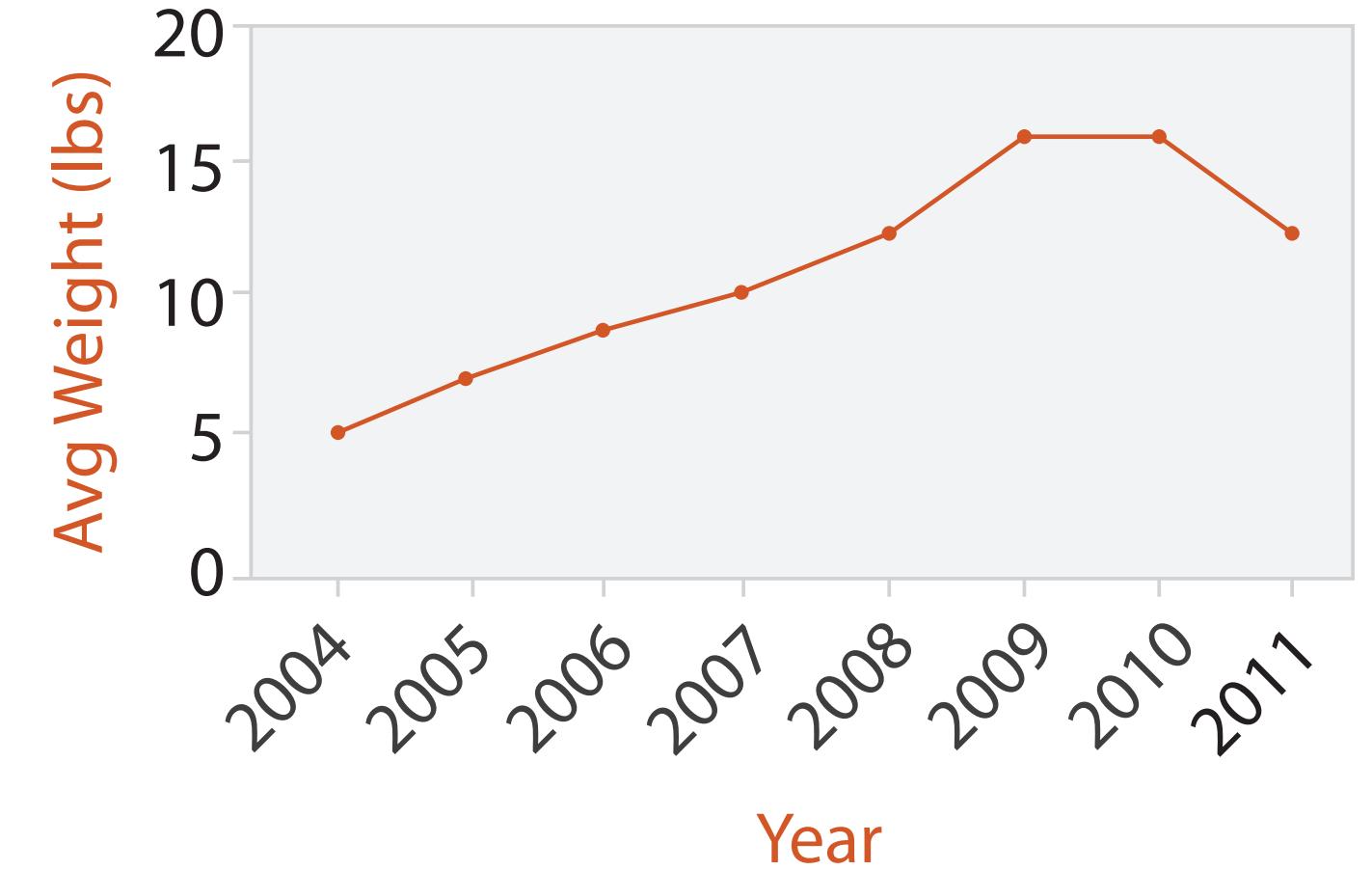
Idiom: bar chart

- one key, one value
 - data
 - 1 categor attrib, 1 quant attrib
 - mark: lines
 - channels
 - length to express quant value
 - spatial regions: one per mark
 - separated horizontally, aligned vertically
 - ordered by quant attrib
 - » by label (alphabetical), by length attrib (data-driven)
 - task
 - compare, lookup values
 - scalability
 - dozens to hundreds of levels for key attrib



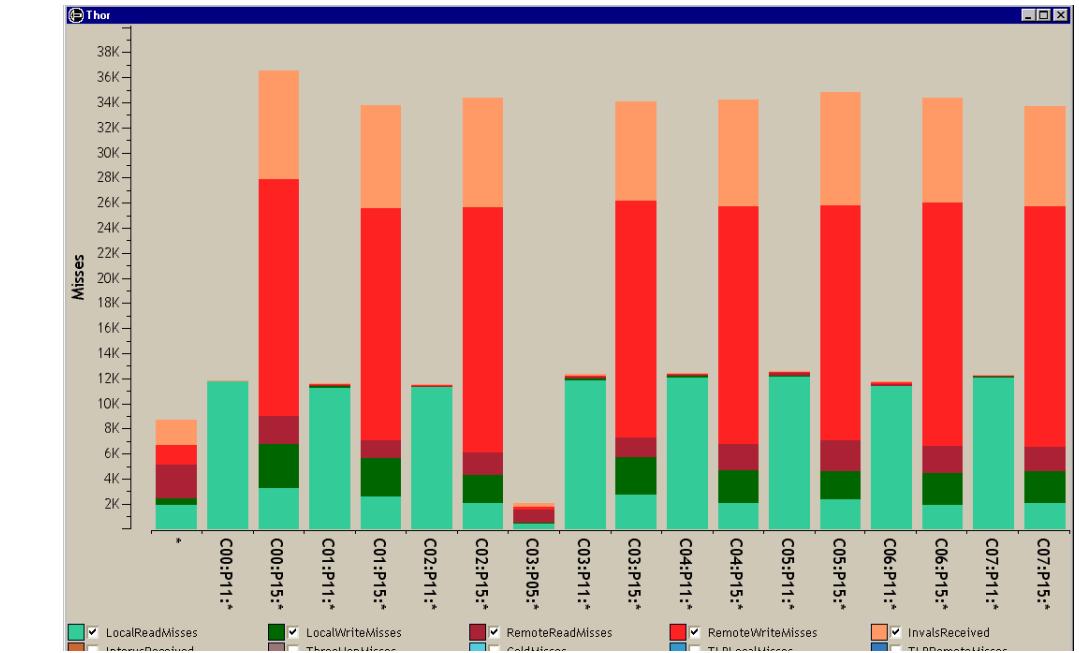
Idiom: line chart

- one key, one value
 - data
 - 2 quant attrs
 - mark: points
 - line connection marks between them
 - channels
 - aligned lengths to express quant value
 - separated and ordered by key attrib into horizontal regions
 - task
 - find trend
 - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next



Idiom: stacked bar chart

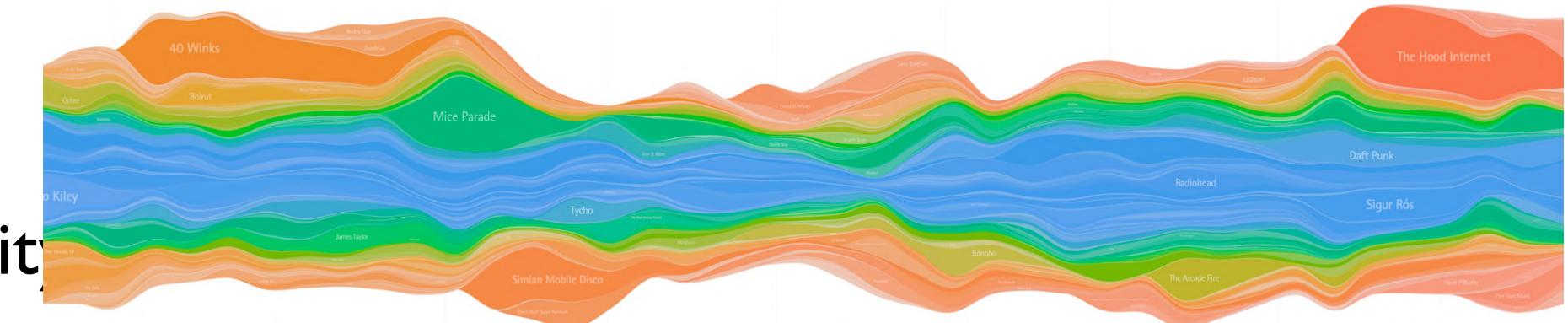
- one more key
 - data
 - 2 categ attrib, 1 quant attrib
 - mark: vertical stack of line marks
 - **glyph**: composite object, internal structure from multiple marks
 - channels
 - length and color hue
 - spatial regions: one per glyph
 - aligned: full glyph, lowest bar component
 - unaligned: other bar components
 - task
 - part-to-whole relationship
 - scalability
 - several to one dozen levels for stacked attrib



[Using Visualization to Understand the Behavior of Computer Systems. Bosch. Ph.D. thesis, Stanford Computer Science, 2001.]

Idiom: streamgraph

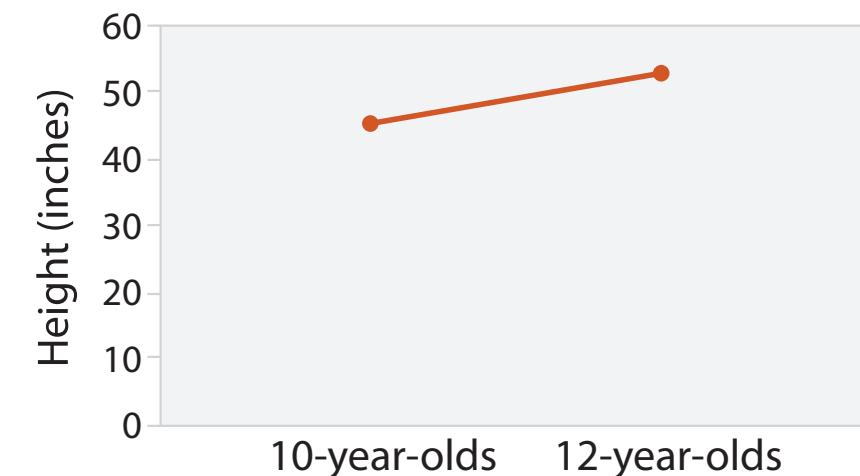
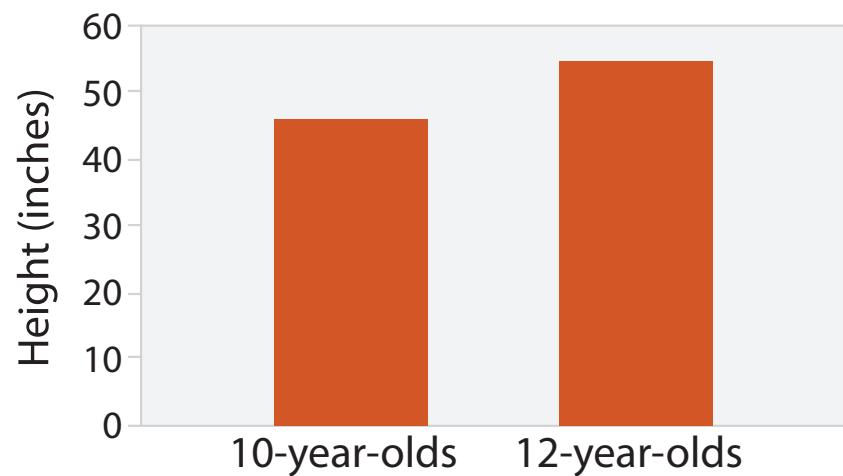
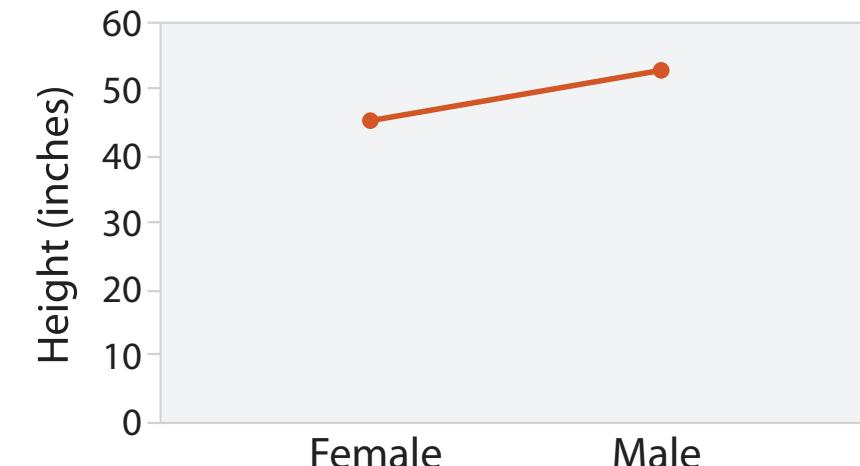
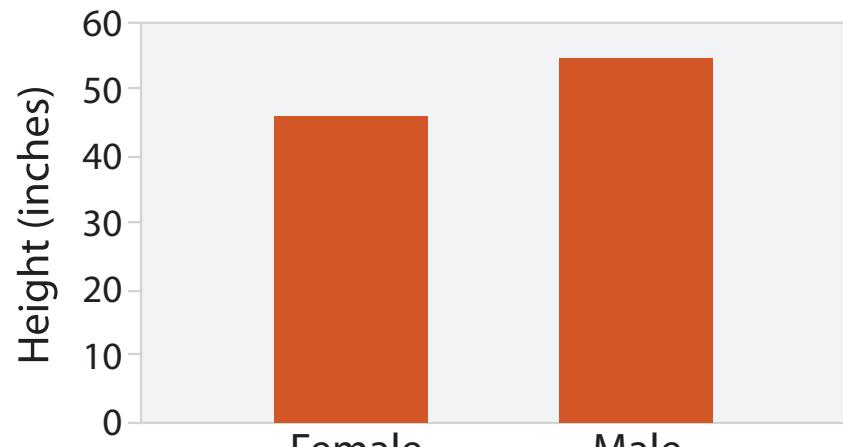
- generalized stacked graph
 - emphasizing horizontal continuity
 - vs vertical items
 - data
 - 1 categ key attrib (artist)
 - 1 ordered key attrib (time)
 - 1 quant value attrib (counts)
 - derived data
 - geometry: layers, where height encodes counts
 - 1 quant attrib (layer ordering)
 - scalability
 - hundreds of time keys
 - dozens to hundreds of artist keys
 - more than stacked bars, since most layers don't extend across whole chart



[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

Choosing bar vs line charts

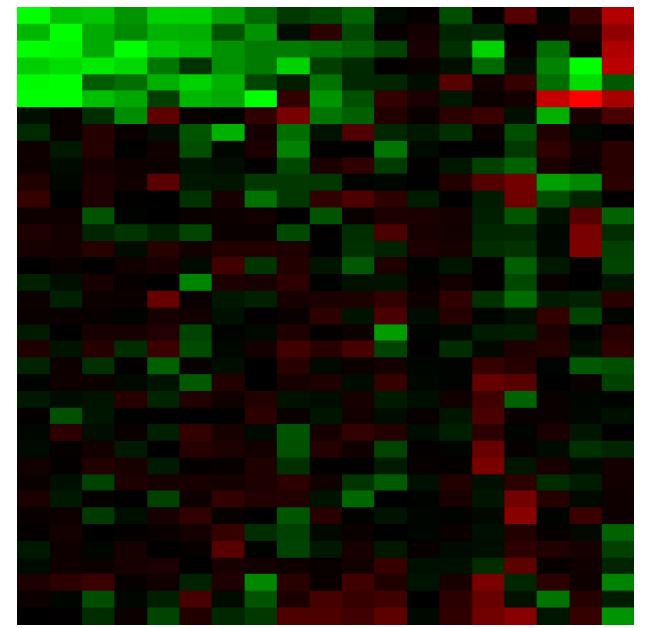
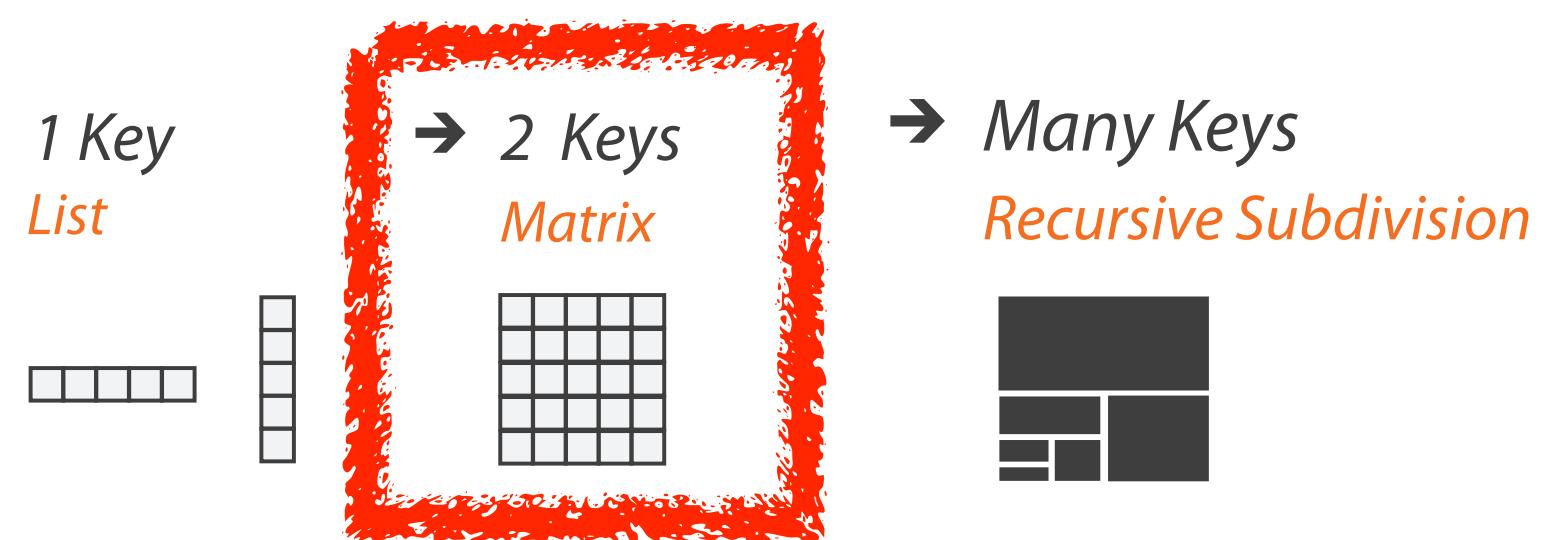
- depends on type of key attrib
 - bar charts if categorical
 - line charts if ordered
- do not use line charts for categorical key attrs
 - violates expressiveness principle
 - implication of trend so strong that it overrides semantics!
 - “The more male a person is, the taller he/she is”



after [Bars and Lines: A Study of Graphic Communication.
Zacks and Tversky. Memory and Cognition 27:6 (1999),
1073–1079.]

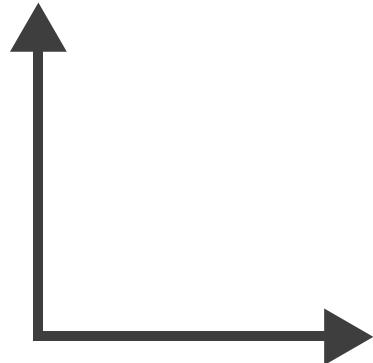
Idiom: heatmap

- two keys, one value
 - data
 - 2 categ attrs (gene, experimental condition)
 - 1 quant attrib (expression levels)
 - marks: area
 - separate and align in 2D matrix
 - indexed by 2 categorical attributes
 - channels
 - color by quant attrib
 - (ordered diverging colormap)
 - task
 - find clusters, outliers
 - scalability
 - 1M items, 100s of categ levels, ~10 quant attrib levels

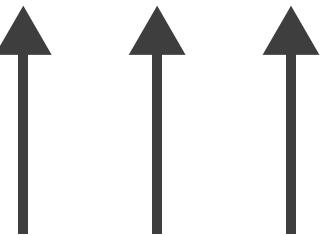


→ Axis Orientation

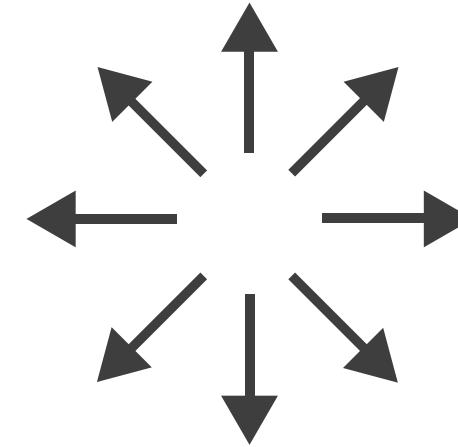
→ Rectilinear



→ Parallel



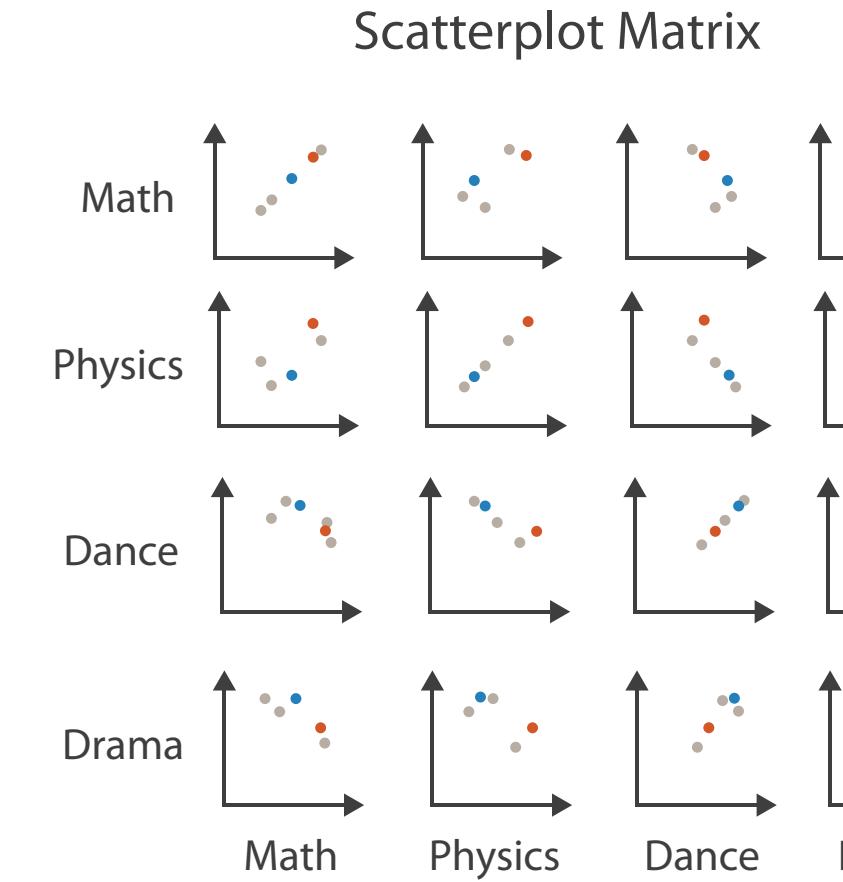
→ Radial



Idioms: scatterplot matrix, parallel coordinates

- scatterplot matrix (SPLOM)

- rectilinear axes, point mark
- all possible pairs of axes
- scalability
 - one dozen attrs
 - dozens to hundreds of items



- parallel coordinates

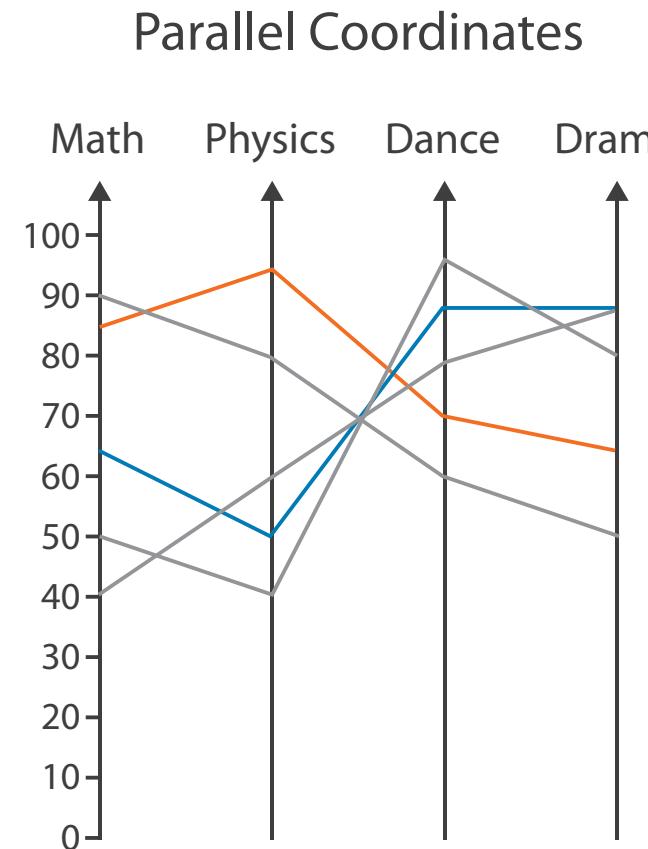
- parallel axes, jagged line representing item

- rectilinear axes, item as point

- axis ordering is major challenge

- scalability

- dozens of attrs
 - hundreds of items



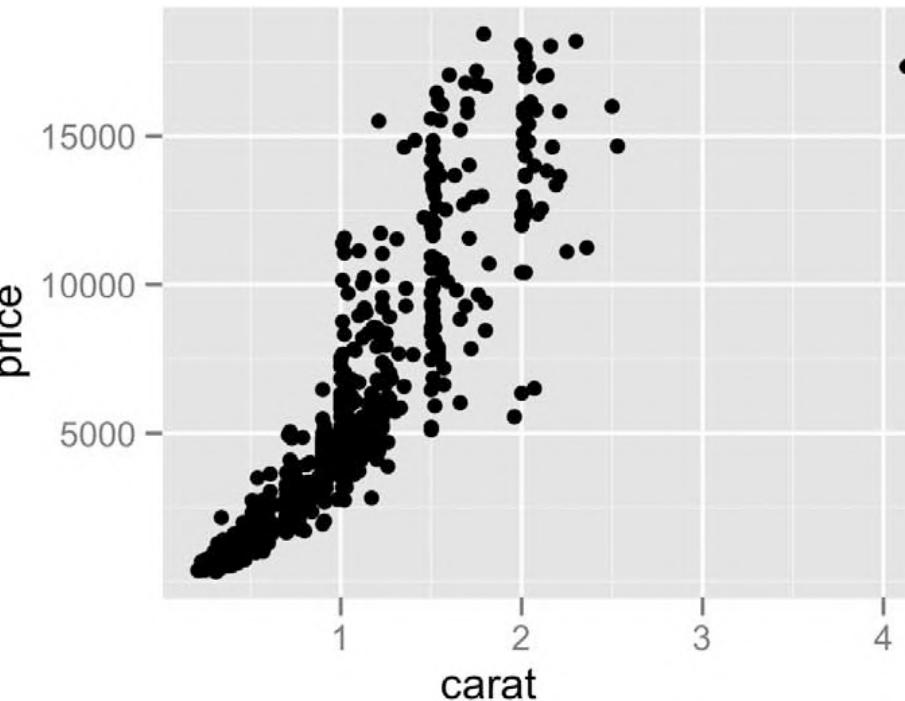
Table

	Math	Physics	Dance	Drama
1	85	95	70	65
2	90	80	60	50
3	65	50	90	90
4	50	40	95	80
5	40	60	80	90

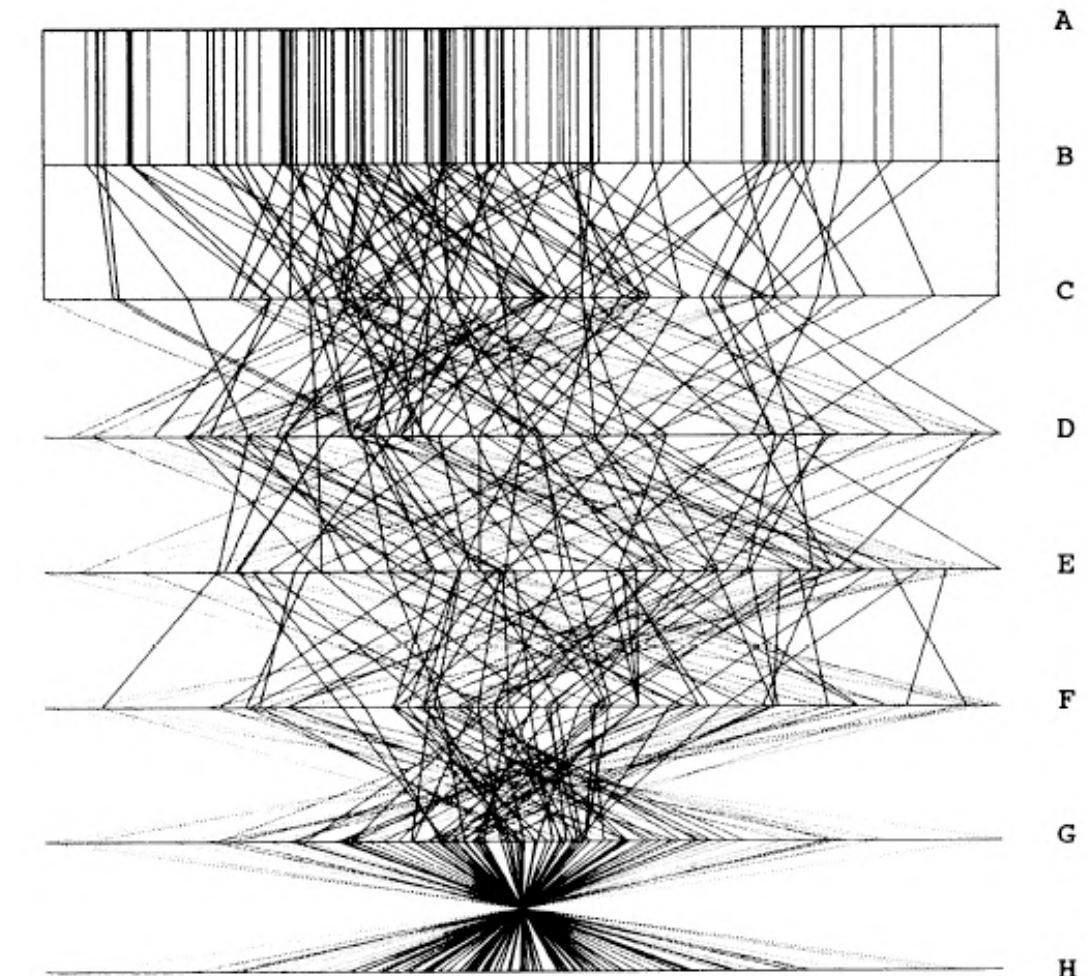
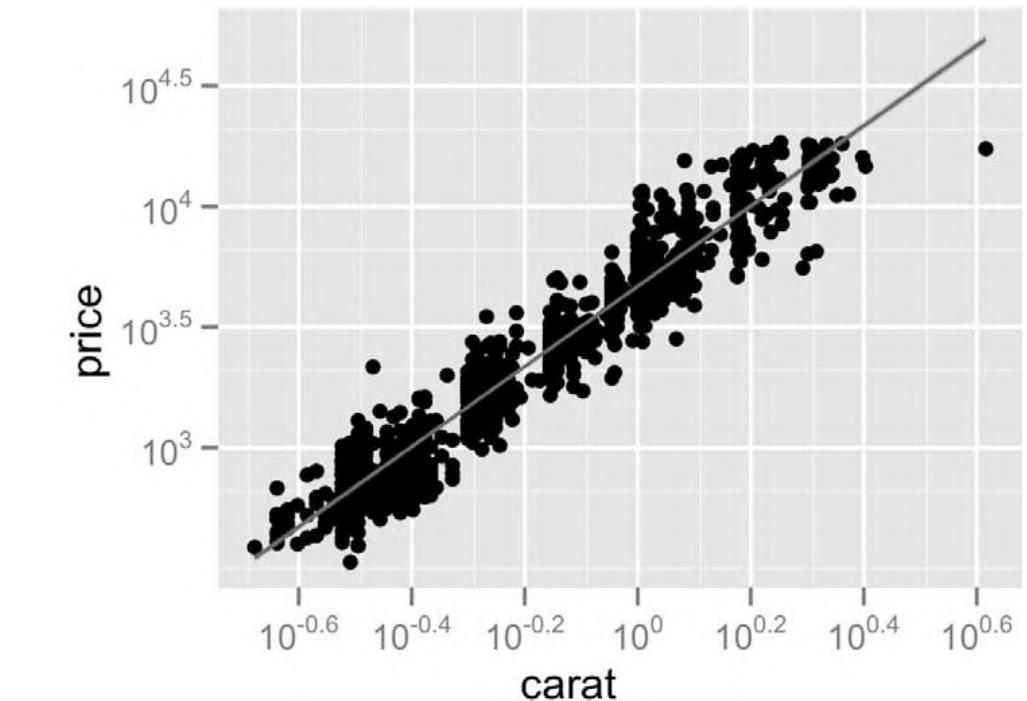
after [Visualization Course Figures. McGuffin, 2014. <http://www.michaelmcguffin.com/courses/vis/>]

Task: Correlation

- scatterplot matrix
 - positive correlation
 - diagonal low-to-high
 - negative correlation
 - diagonal high-to-low
 - uncorrelated
- parallel coordinates
 - positive correlation
 - parallel line segments
 - negative correlation
 - all segments cross at halfway point
 - uncorrelated
 - scattered crossings



[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

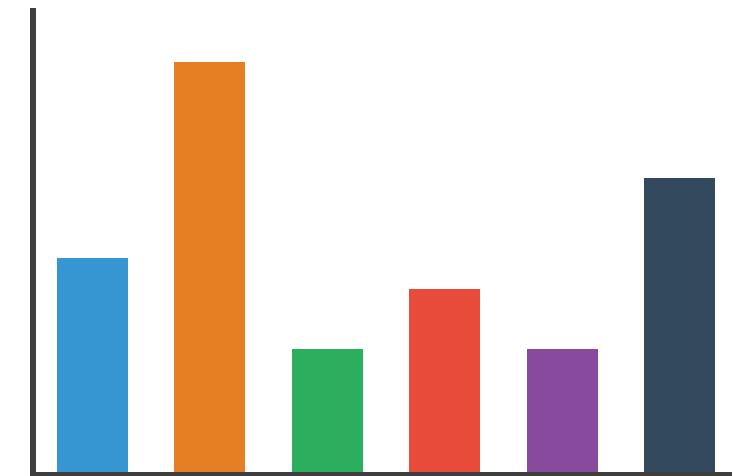
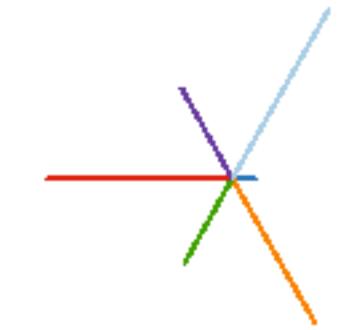
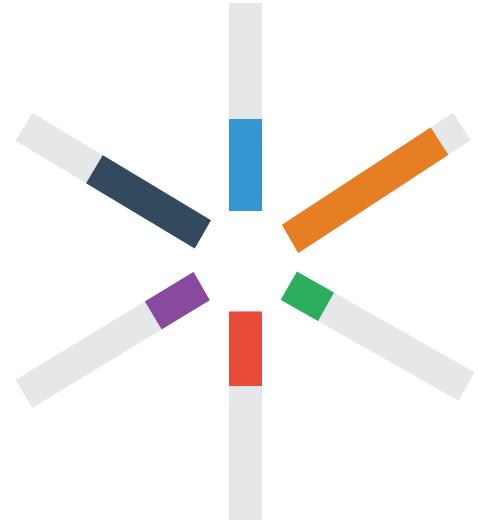


[Hyperdimensional Data Analysis Using Parallel Coordinates. Wegman. Journ. American Statistical Association 85:411 (1990), 664–675.]

Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8$, and -1 .

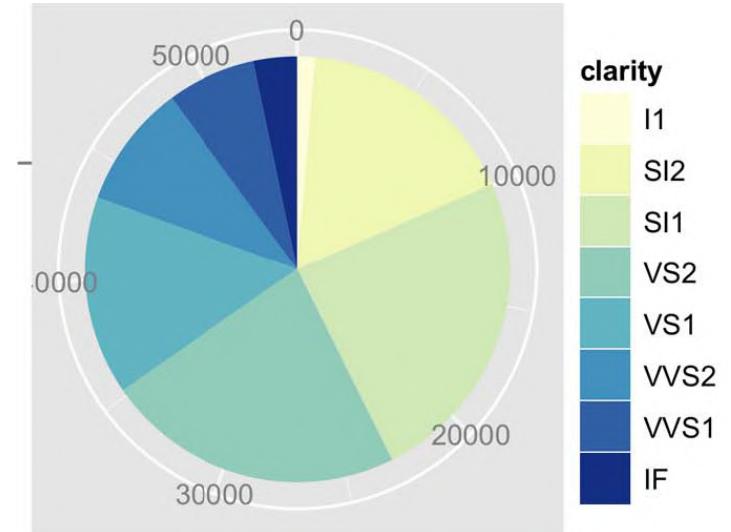
Idioms: **radial bar chart, star plot**

- radial bar chart
 - radial axes meet at central ring, line mark
- star plot
 - radial axes, meet at central point, line mark
- bar chart
 - rectilinear axes, aligned vertically
- accuracy
 - length unaligned with radial
 - less accurate than aligned with rectilinear

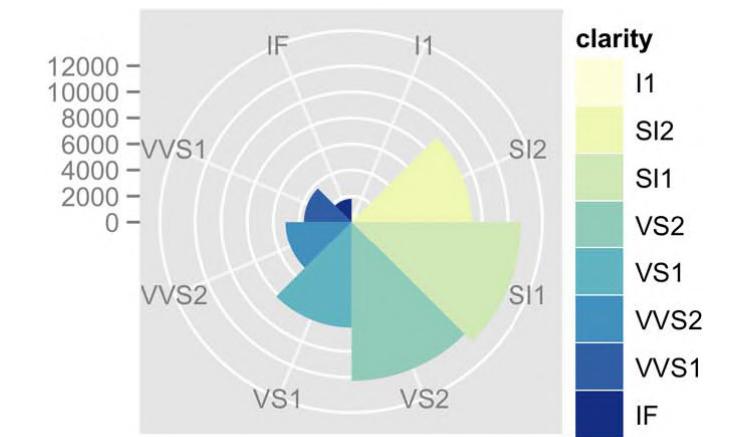


Idioms: pie chart, polar area chart

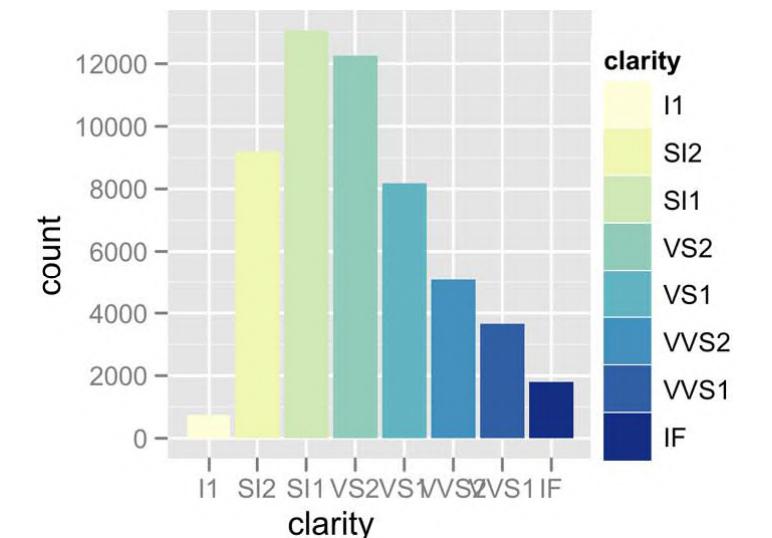
- pie chart
 - area marks with angle channel
 - accuracy: angle/area less accurate than line length
 - arclength also less accurate than line length



- polar area chart
 - area marks with length channel
 - more direct analog to bar charts

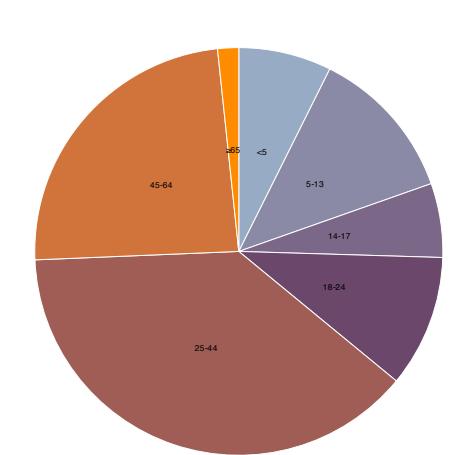
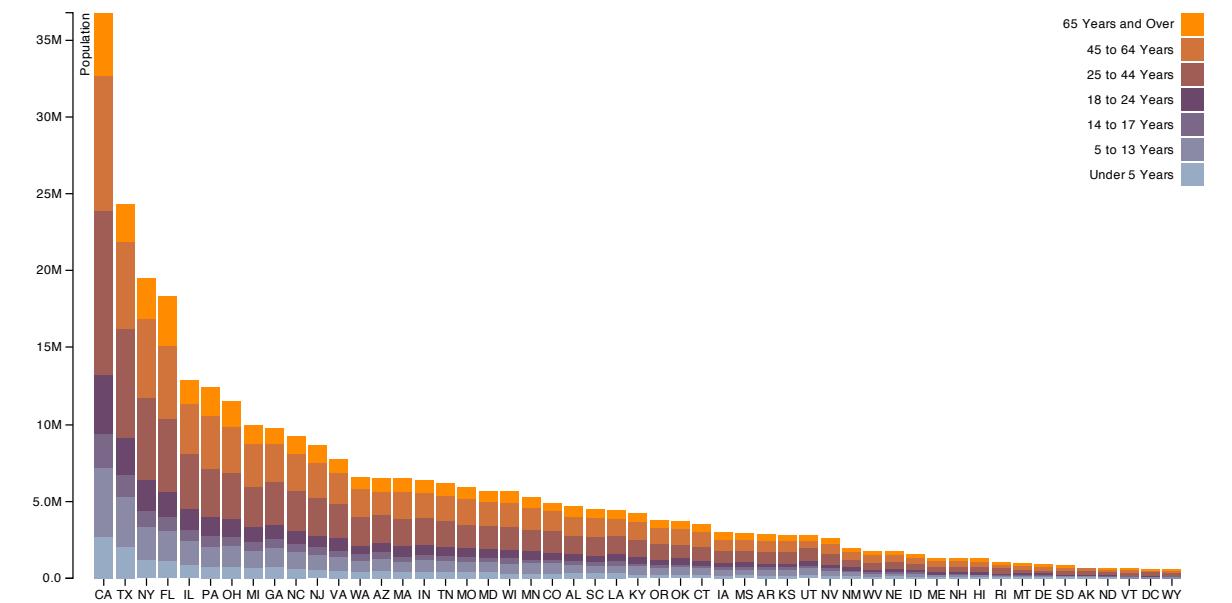
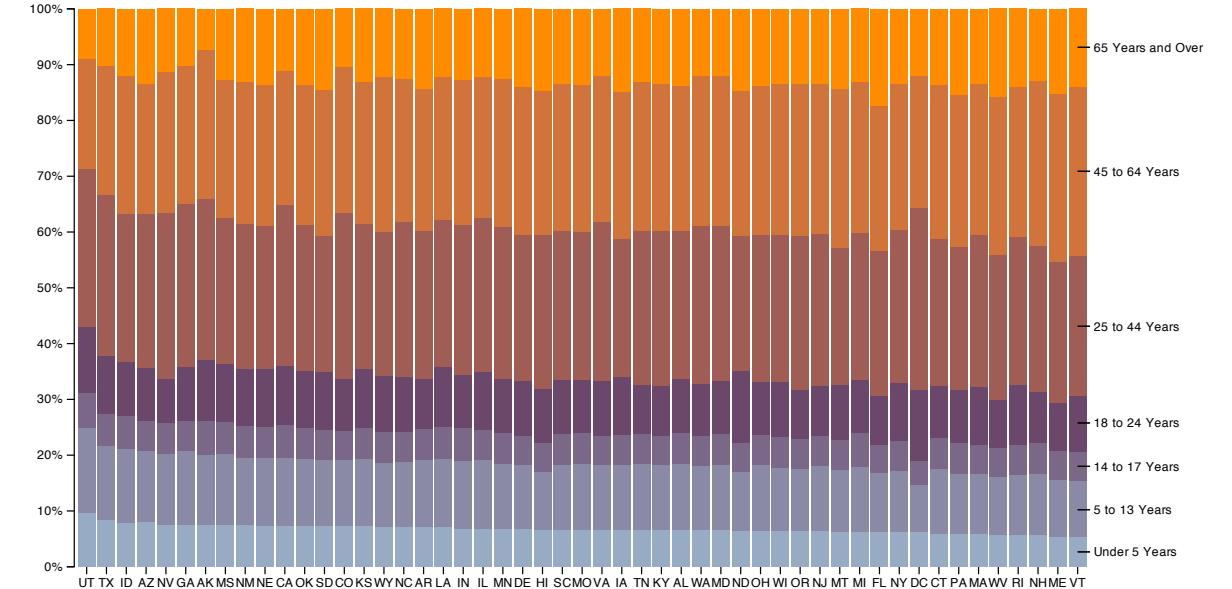


- data
 - 1 categor key attrib, 1 quant value attrib
- task
 - part-to-whole judgements



Idioms: normalized stacked bar chart

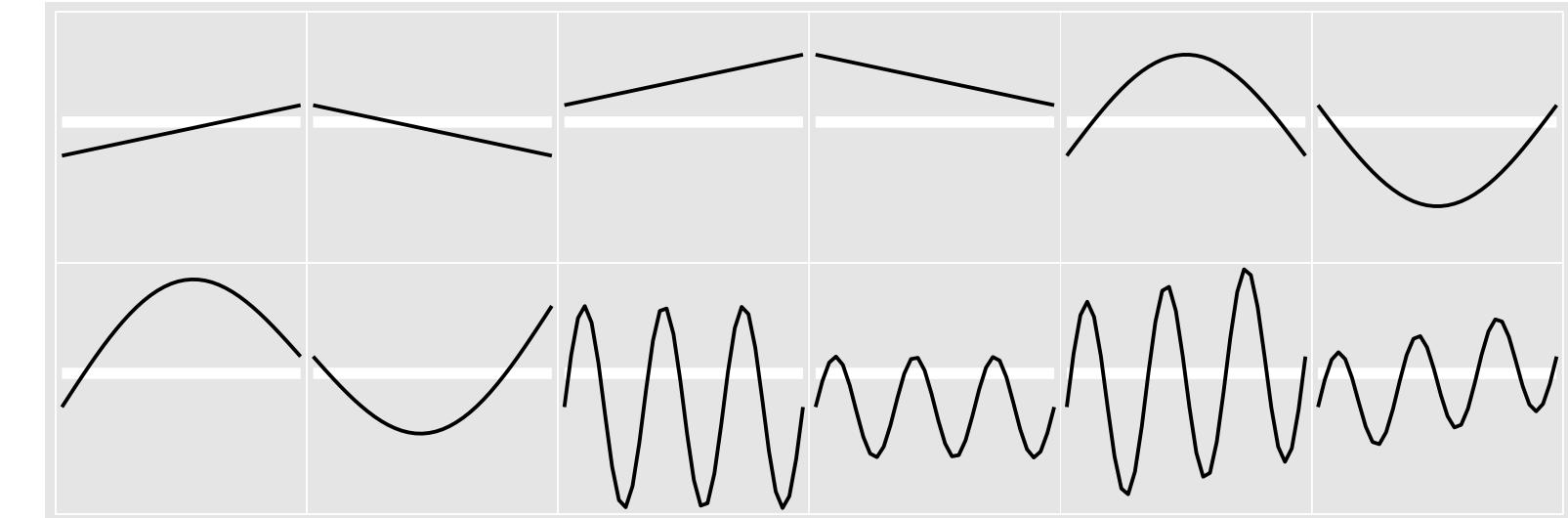
- task
 - part-to-whole judgements
- normalized stacked bar chart
 - stacked bar chart, normalized to full vert height
 - single stacked bar equivalent to full pie
 - high information density: requires narrow rectangle
- pie chart
 - information density: requires large circle



<http://bl.ocks.org/mbostock/3887235>,
<http://bl.ocks.org/mbostock/3886208>,
<http://bl.ocks.org/mbostock/3886394>.

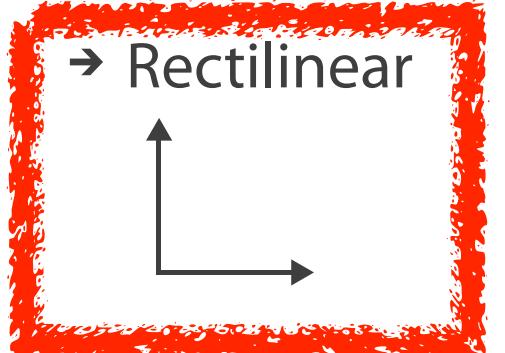
Idiom: glyphmaps

- rectilinear good for linear vs nonlinear trends

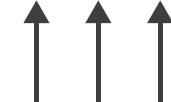


- radial good for cyclic patterns

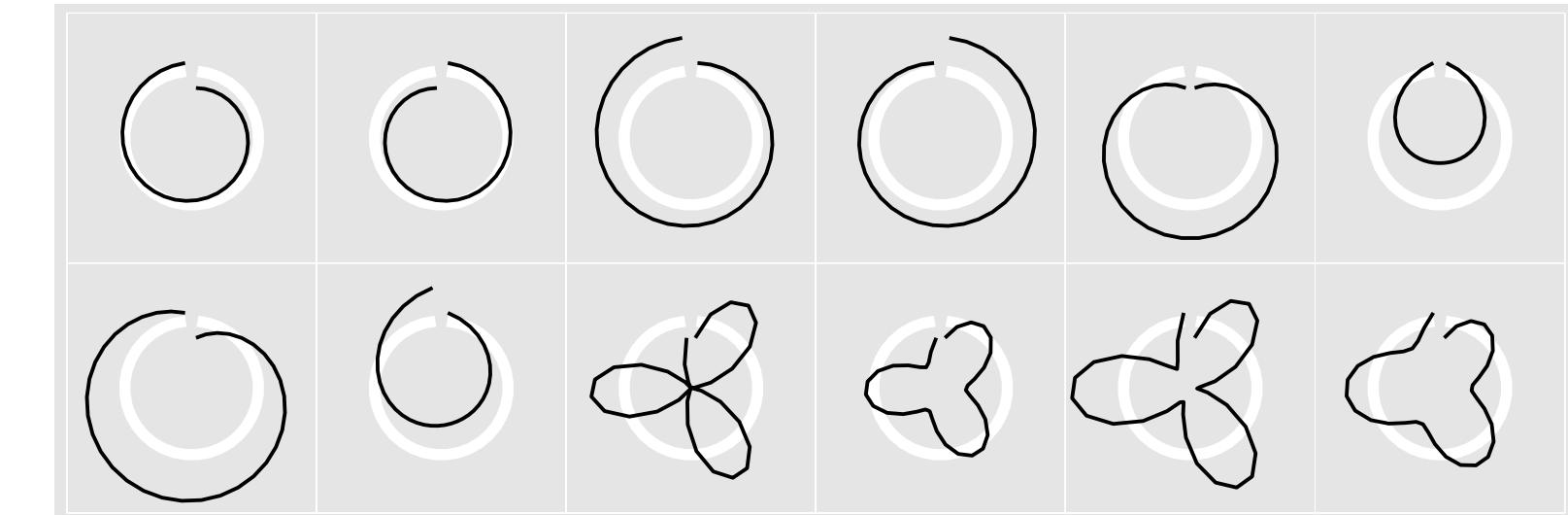
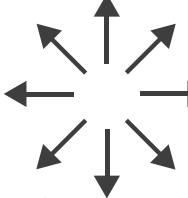
→ Axis Orientation



→ Parallel



→ Radial



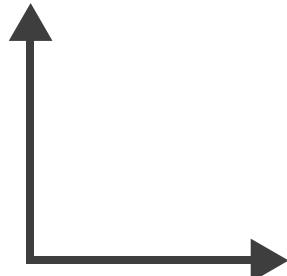
[*Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models*. Wickham, Hofmann, Wickham, and Cook. *Environmetrics* 23:5 (2012), 382–393.]

Orientation limitations

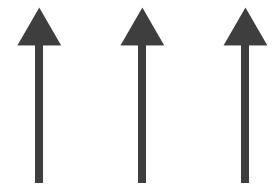
- rectilinear: scalability wrt #axes
 - 2 axes best
 - 3 problematic
 - 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
 - angles lower precision than lengths
 - asymmetry between angle and length
 - can be exploited!

→ Axis Orientation

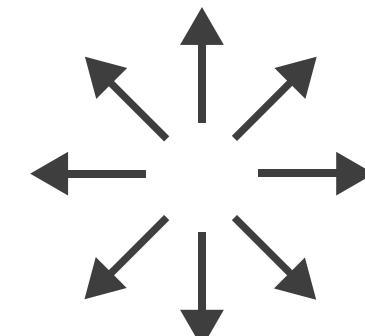
→ Rectilinear



→ Parallel



→ Radial



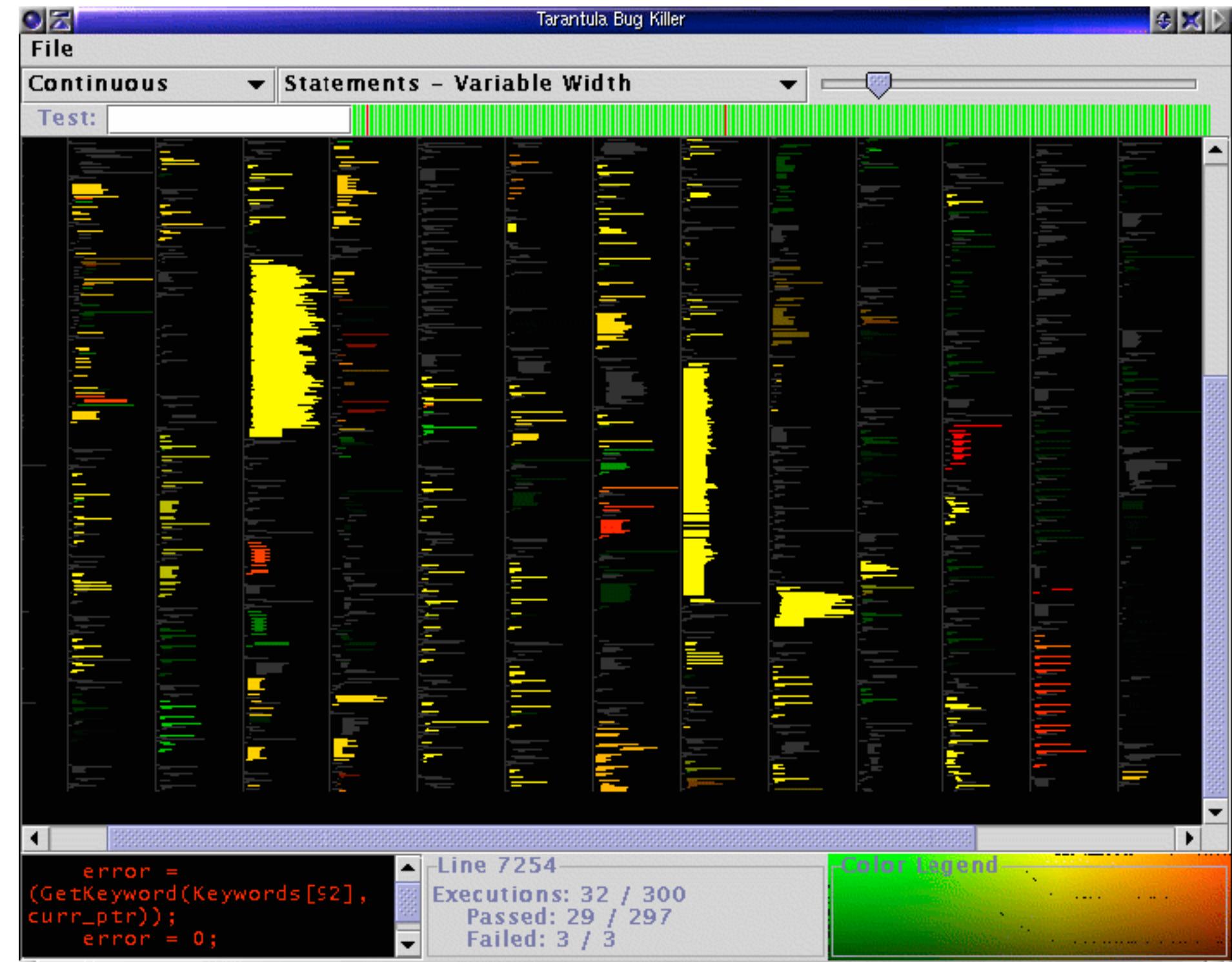
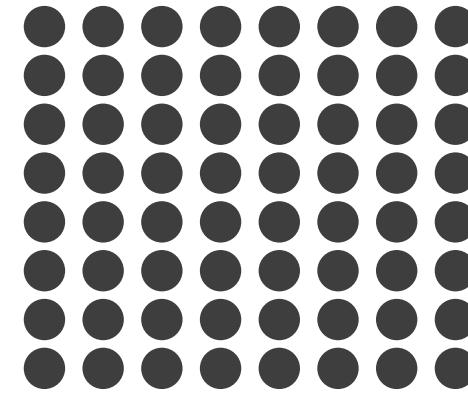
[*Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach.* Diehl, Beck and Burch. *IEEE TVCG (Proc. InfoVis)* 16(6):935–942, 2010.]



Layout Density

dense software

→ Dense



Arrange spatial data

④ Use Given

→ Geometry

→ *Geographic*

→ *Other Derived*

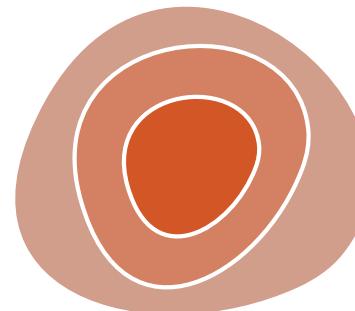


→ Spatial Fields

→ *Scalar Fields (one value per cell)*

→ *Isocontours*

→ *Direct Volume Rendering*



→ *Vector and Tensor Fields (many values per cell)*

→ *Flow Glyphs (local)*



→ *Geometric (sparse seeds)*



→ *Textures (dense seeds)*

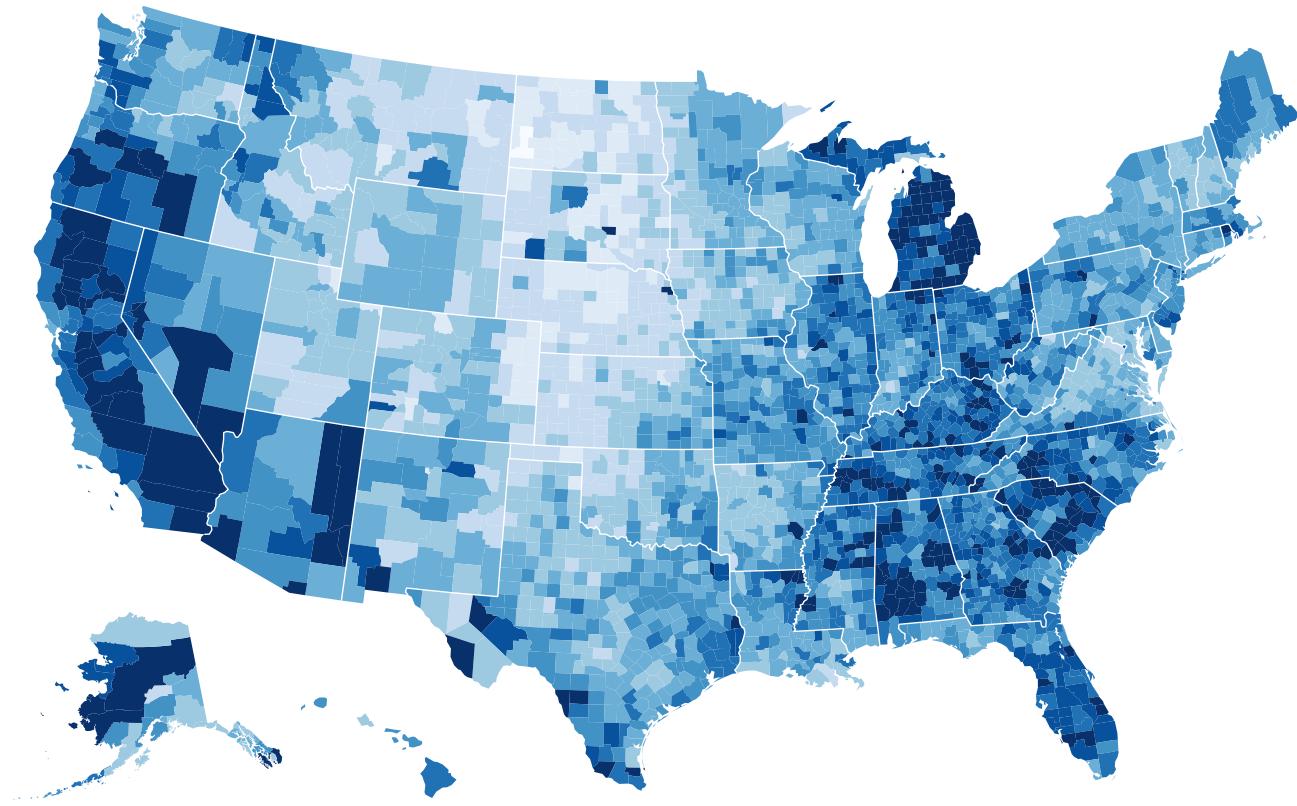


→ *Features (globally derived)*



Idiom: choropleth map

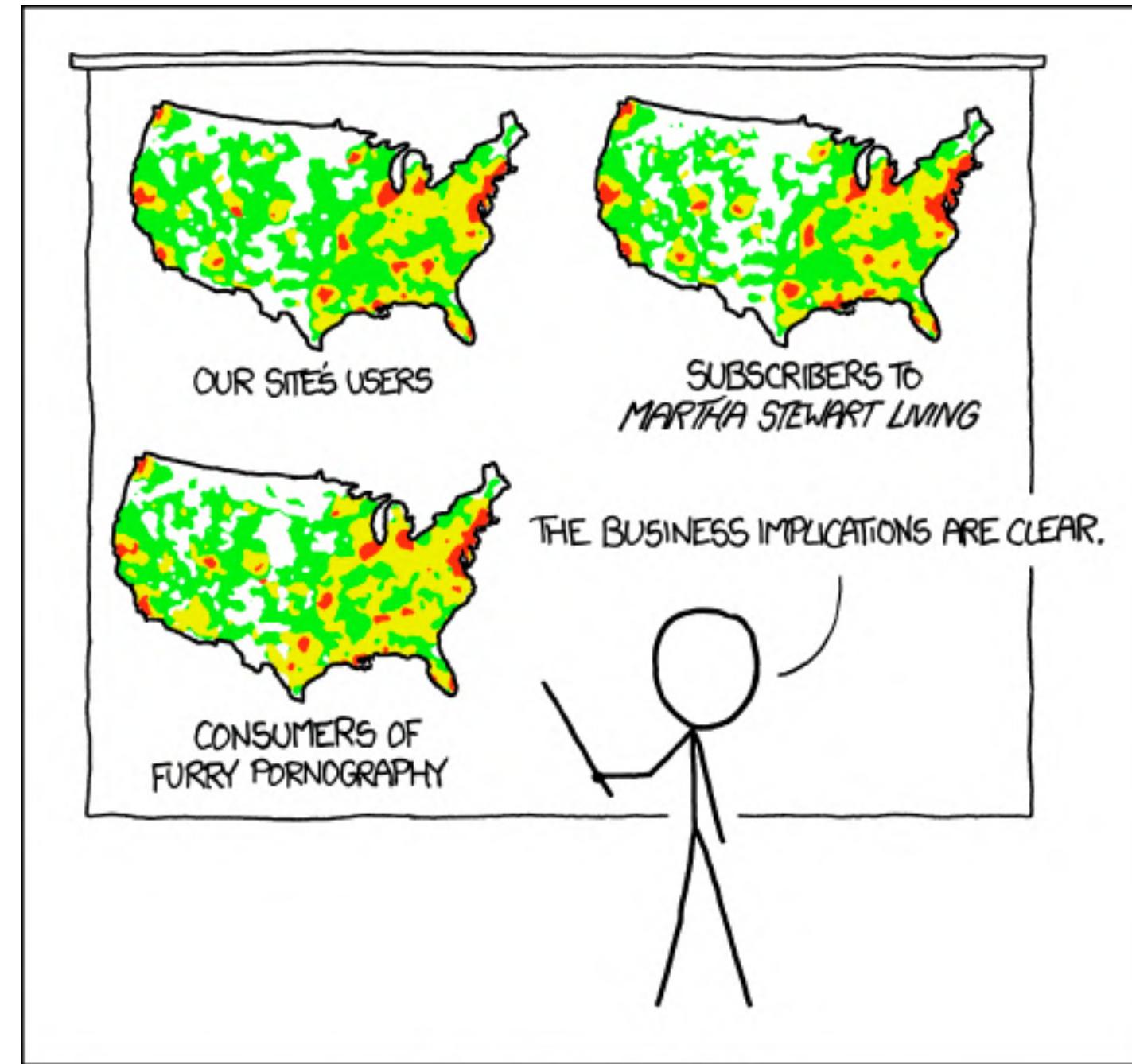
- **use given spatial data**
 - when central task is understanding spatial relationships
- **data**
 - geographic geometry
 - table with 1 quant attribute per region
- **encoding**
 - use given geometry for area mark boundaries
 - sequential segmented colormap [more later]



<http://bl.ocks.org/mbostock/4060606>

Beware: Population maps trickiness!

[<https://xkcd.com/1138>]



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

Idiom: topographic map

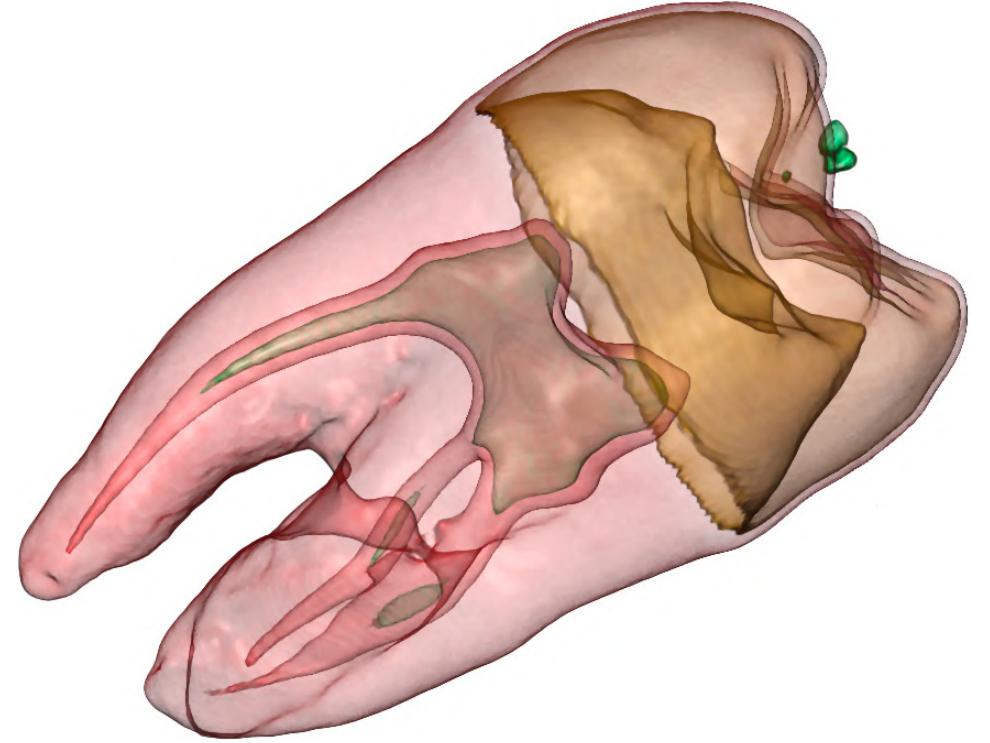
- data
 - geographic geometry
 - scalar spatial field
 - 1 quant attribute per grid cell
- derived data
 - isoline geometry
 - isocontours computed for specific levels of scalar values



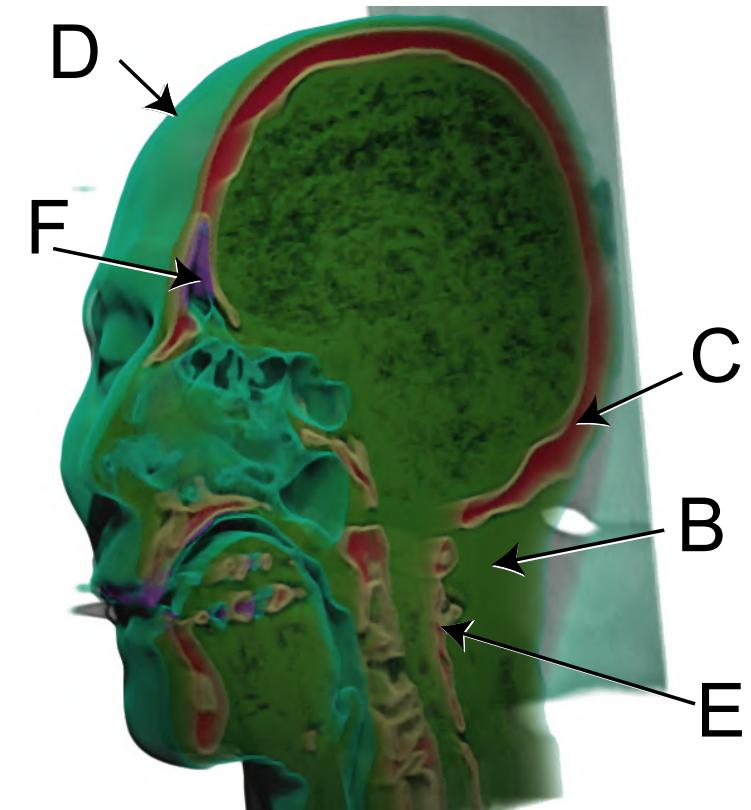
Land Information New Zealand Data Service

Idioms: **isosurfaces**, direct volume rendering

- data
 - scalar spatial field
 - 1 quant attribute per grid cell
- task
 - shape understanding, spatial relationships
- isosurface
 - derived data: isocontours computed for specific levels of scalar values
- direct volume rendering
 - transfer function maps scalar values to color, opacity



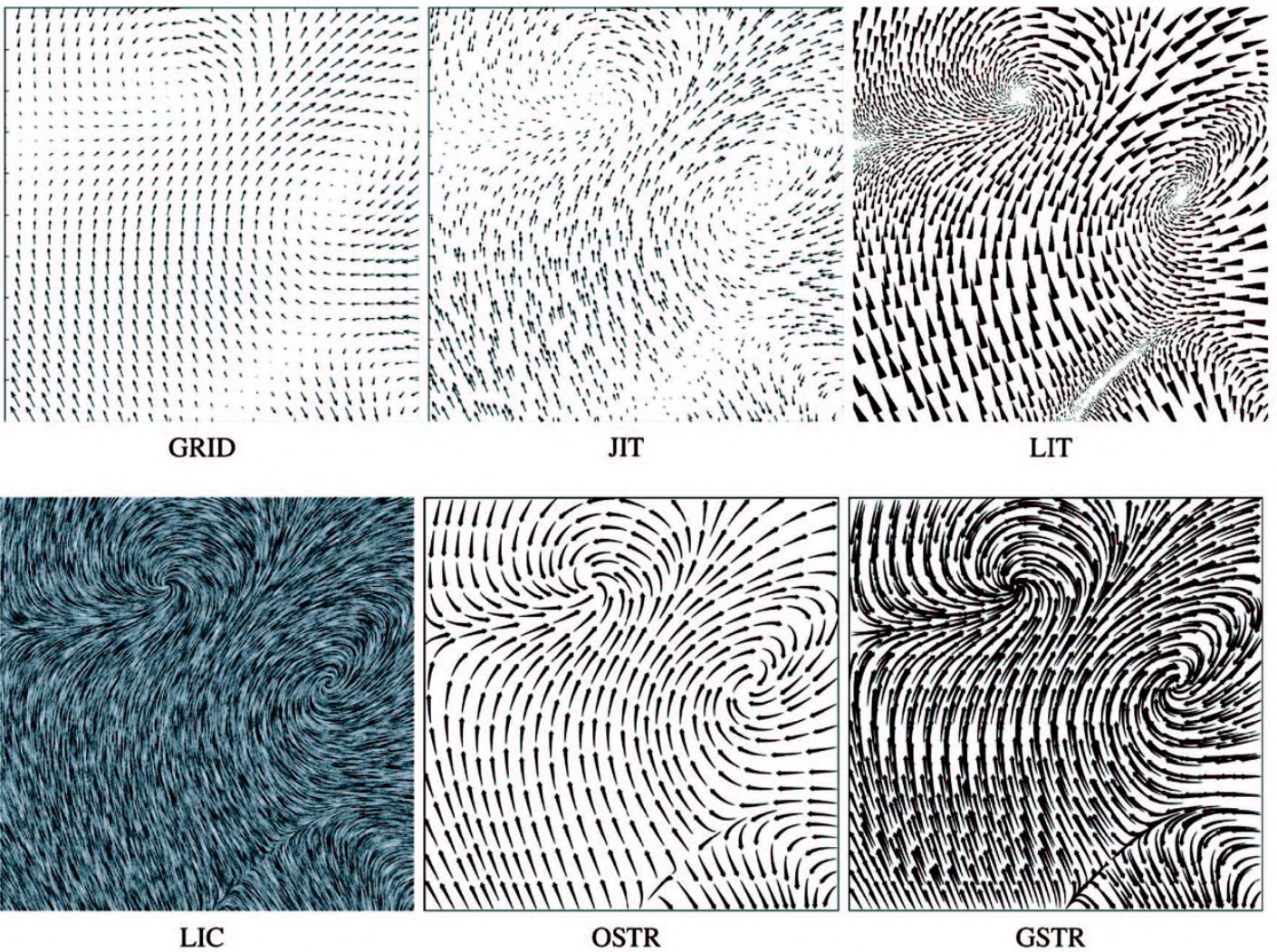
[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]



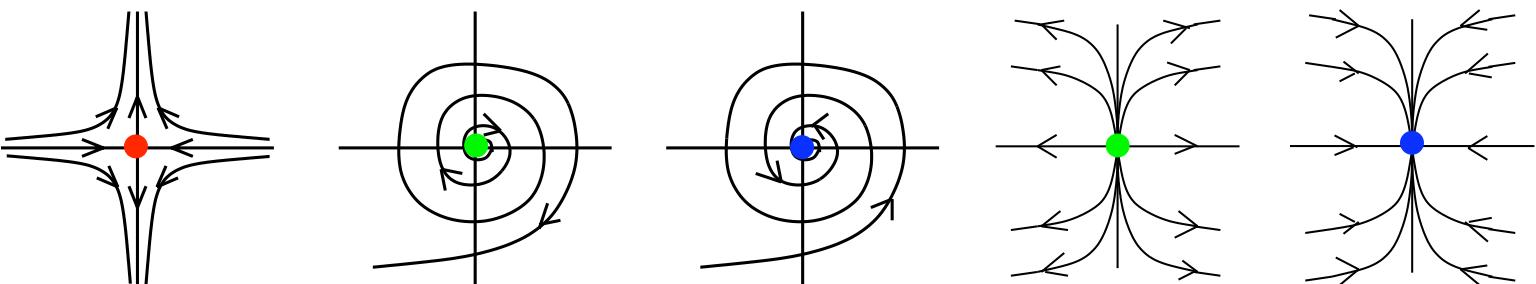
[Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]

Vector and tensor fields

- data
 - many attrs per cell
- idiom families
 - **flow glyphs**
 - purely local
 - **geometric flow**
 - derived data from tracing particle trajectories
 - sparse set of seed points
 - **texture flow**
 - derived data, dense seeds
 - **feature flow**
 - global computation to detect features
 - encoded with one of methods above



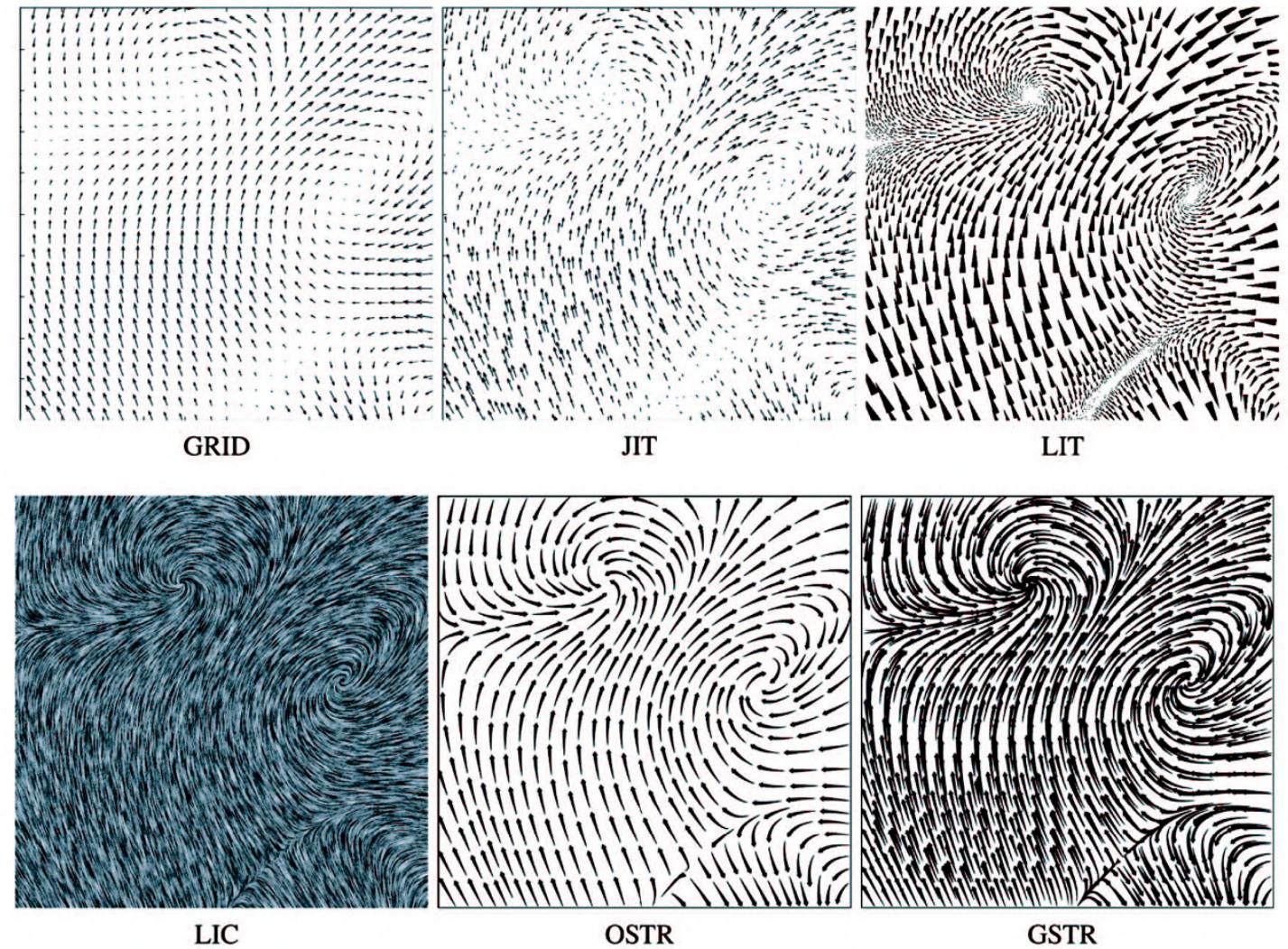
[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]



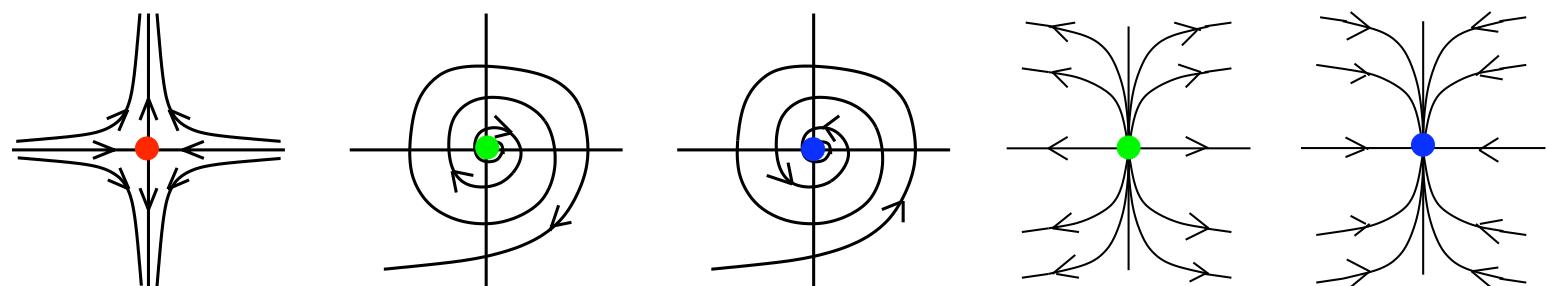
[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257.]

Vector fields

- empirical study tasks
 - finding critical points, identifying their types
 - identifying what type of critical point is at a specific location
 - predicting where a particle starting at a specified point will end up (advection)



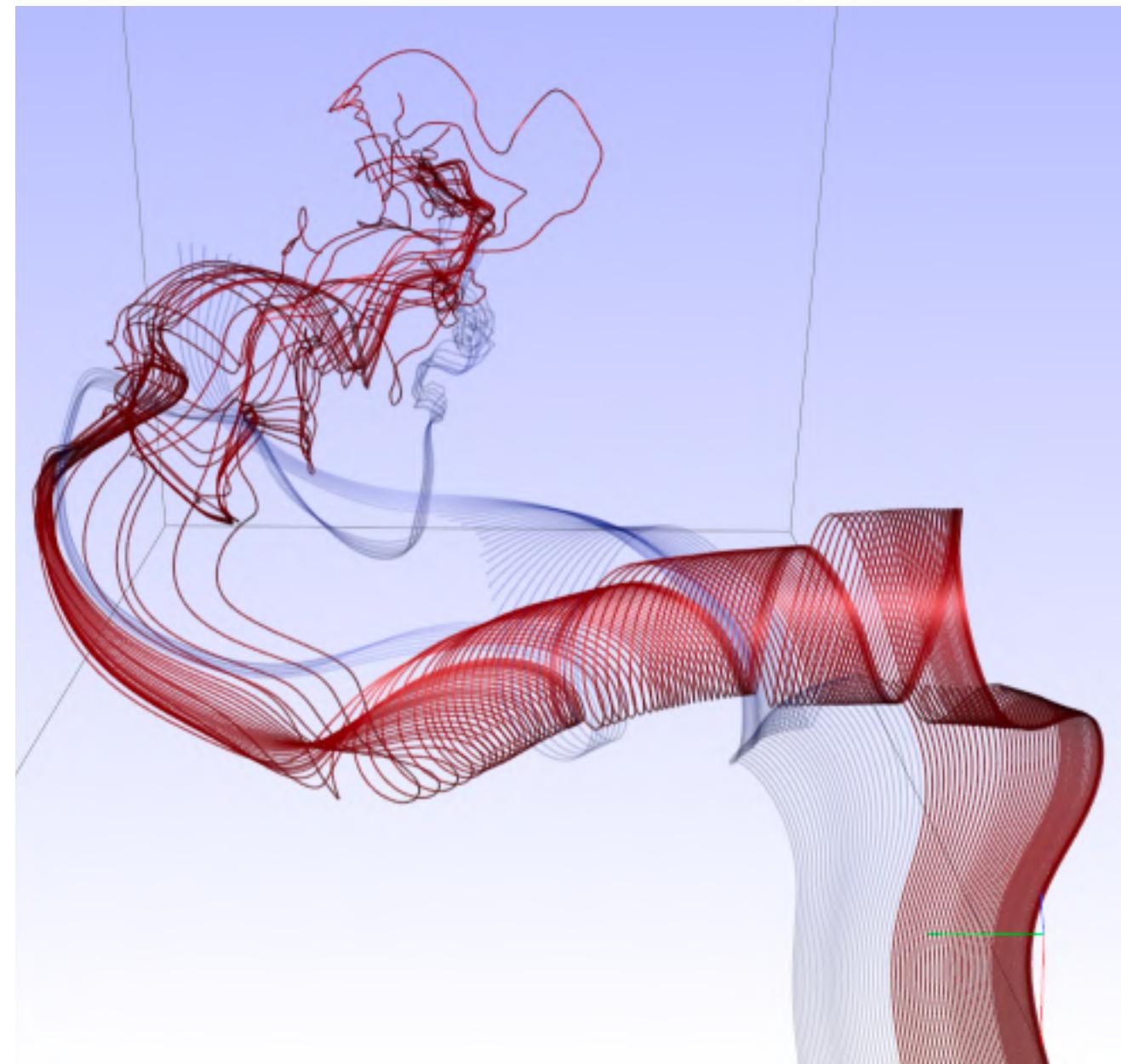
[Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.]



[Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche, Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257.]

Idiom: similarity-clustered streamlines

- data
 - 3D vector field
- derived data (from field)
 - streamlines: trajectory particle will follow
- derived data (per streamline)
 - curvature, torsion, tortuosity
 - signature: complex weighted combination
 - compute cluster hierarchy across all signatures
 - encode: color and opacity by cluster
- tasks
 - find features, query shape
- scalability
 - millions of samples, hundreds of streamlines



[*Similarity Measures for Enhancing Interactive Streamline Seeding.*
McLoughlin, Jones, Laramee, Malki, Masters, and Hansen. IEEE Trans.
Visualization and Computer Graphics 19:8 (2013), 1342–1353.]

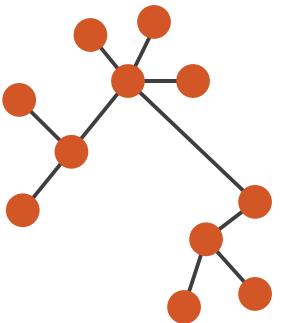
Arrange networks and trees

→ Node–Link Diagrams

Connection Marks

NETWORKS

TREES



→ Adjacency Matrix

Derived Table

NETWORKS

TREES

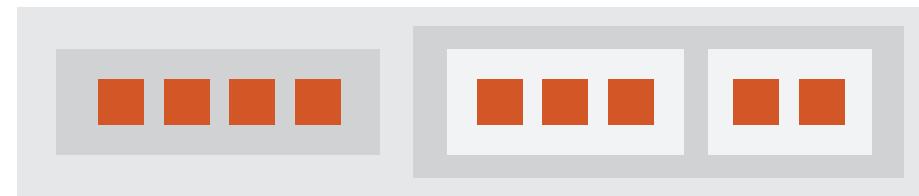
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■

→ Enclosure

Containment Marks

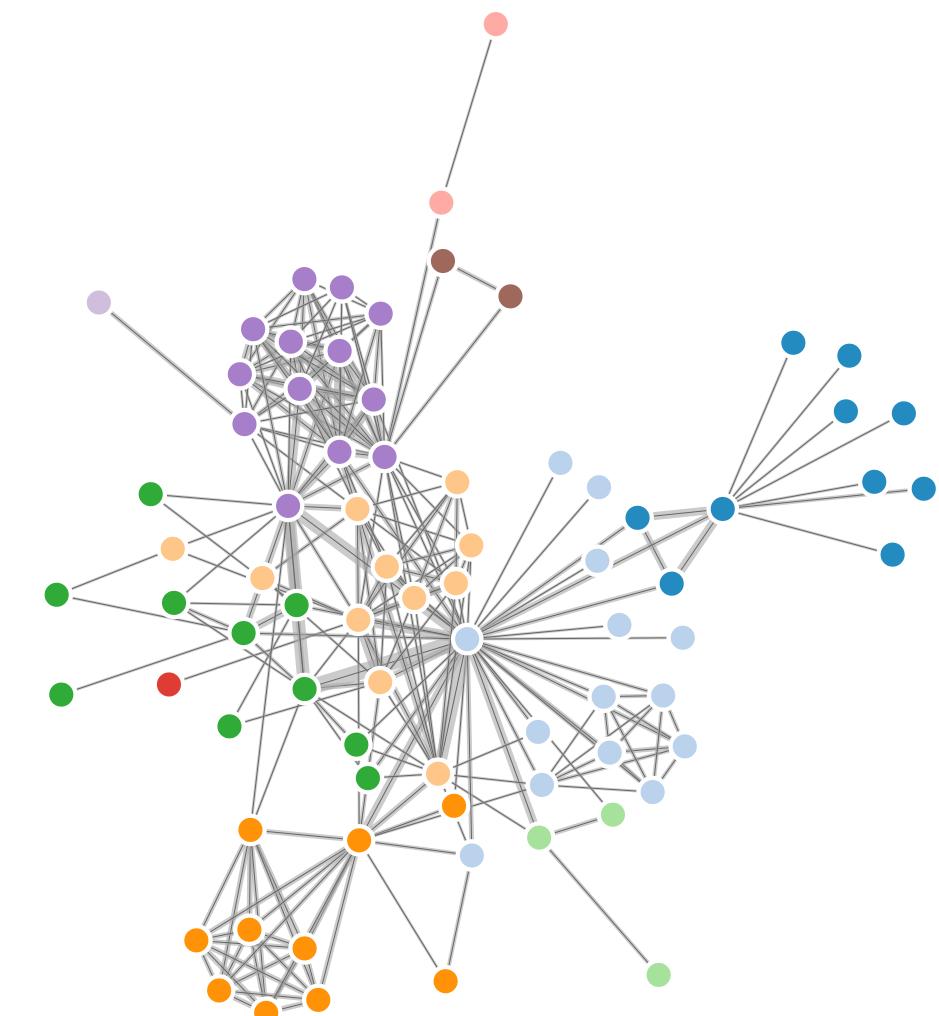
NETWORKS

TREES



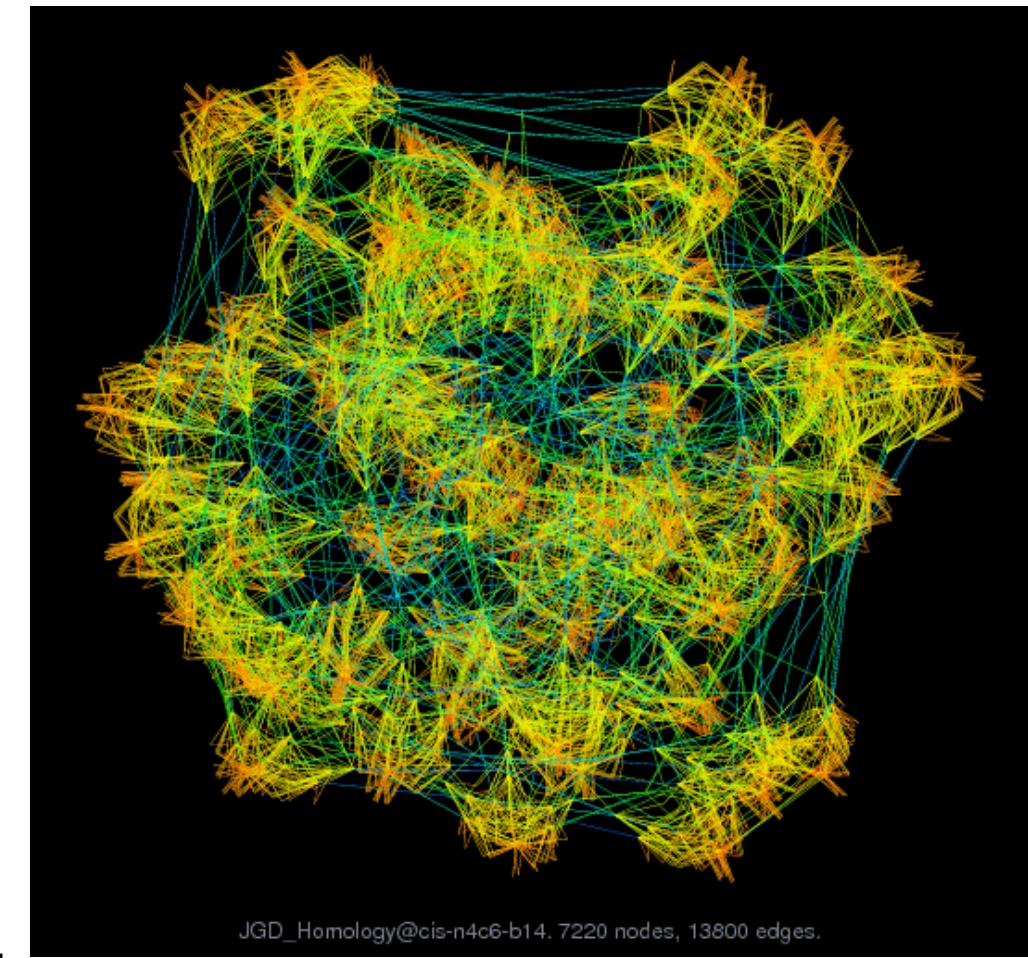
Idiom: force-directed placement

- visual encoding
 - link connection marks, node point marks
- considerations
 - spatial position: no meaning directly encoded
 - left free to minimize crossings
 - proximity semantics?
 - sometimes meaningful
 - sometimes arbitrary, artifact of layout algorithm
 - tension with length
 - long edges more visually salient than short
- tasks
 - explore topology; locate paths, clusters
- scalability
 - node/edge density $E < 4N$

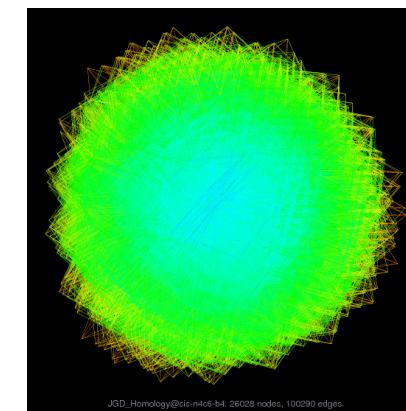


Idiom: **sfdp** (multi-level force-directed placement)

- data
 - original: network
 - derived: cluster hierarchy atop it
- considerations
 - better algorithm for same encoding technique
 - same: fundamental use of space
 - hierarchy used for algorithm speed/quality but not shown explicitly
 - (more on algorithm vs encoding in afternoon)
- scalability
 - nodes, edges: 1K-10K
 - hairball problem eventually hits



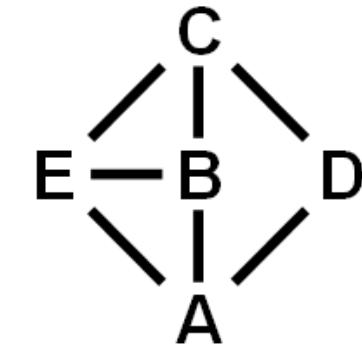
[Efficient and high quality force-directed graph drawing.
Hu. *The Mathematica Journal* 10:37–71, 2005.]



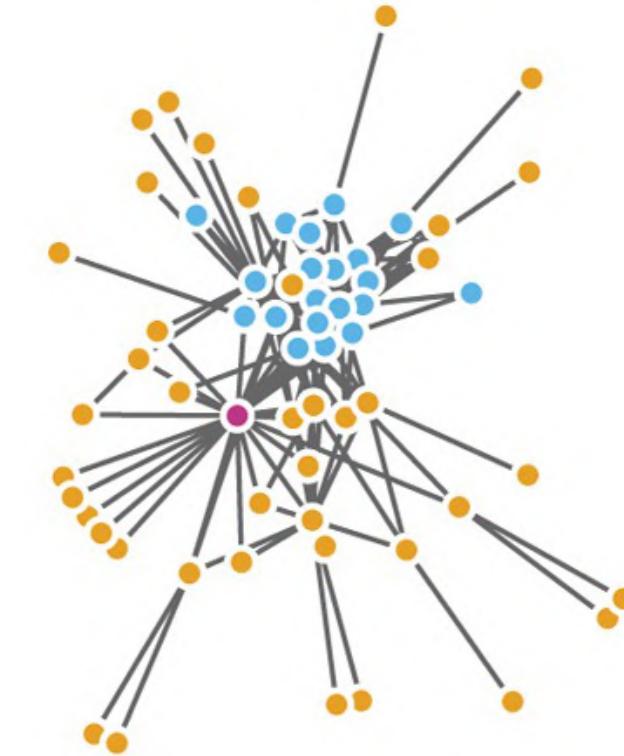
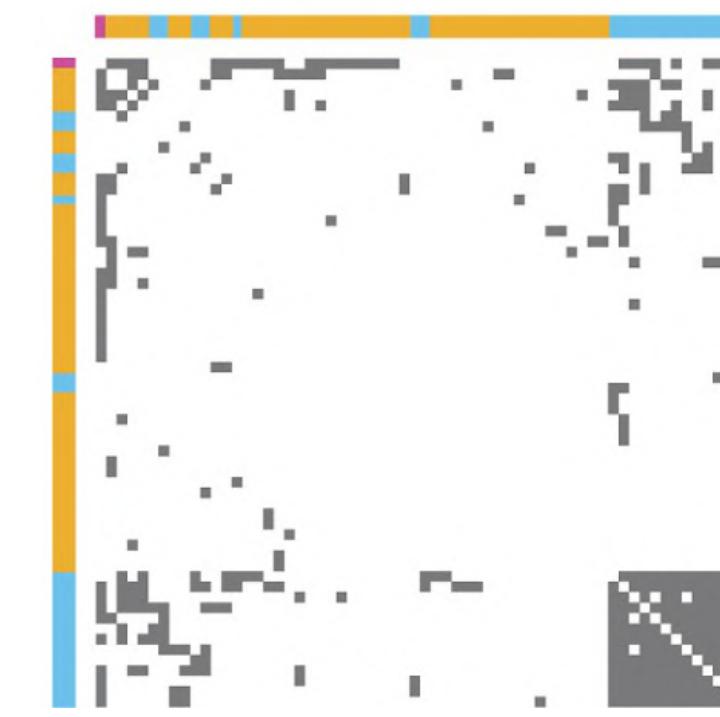
Idiom: adjacency matrix view

- data: network
 - transform into same data/encoding as heatmap
- derived data: table from network
 - 1 quant attrib
 - weighted edge between nodes
 - 2 categ attribs: node list x 2
- visual encoding
 - cell shows presence/absence of edge
- scalability
 - 1K nodes, 1M edges

	A	B	C	D	E
A	A				
B		B			
C			C		
D				D	
E					E



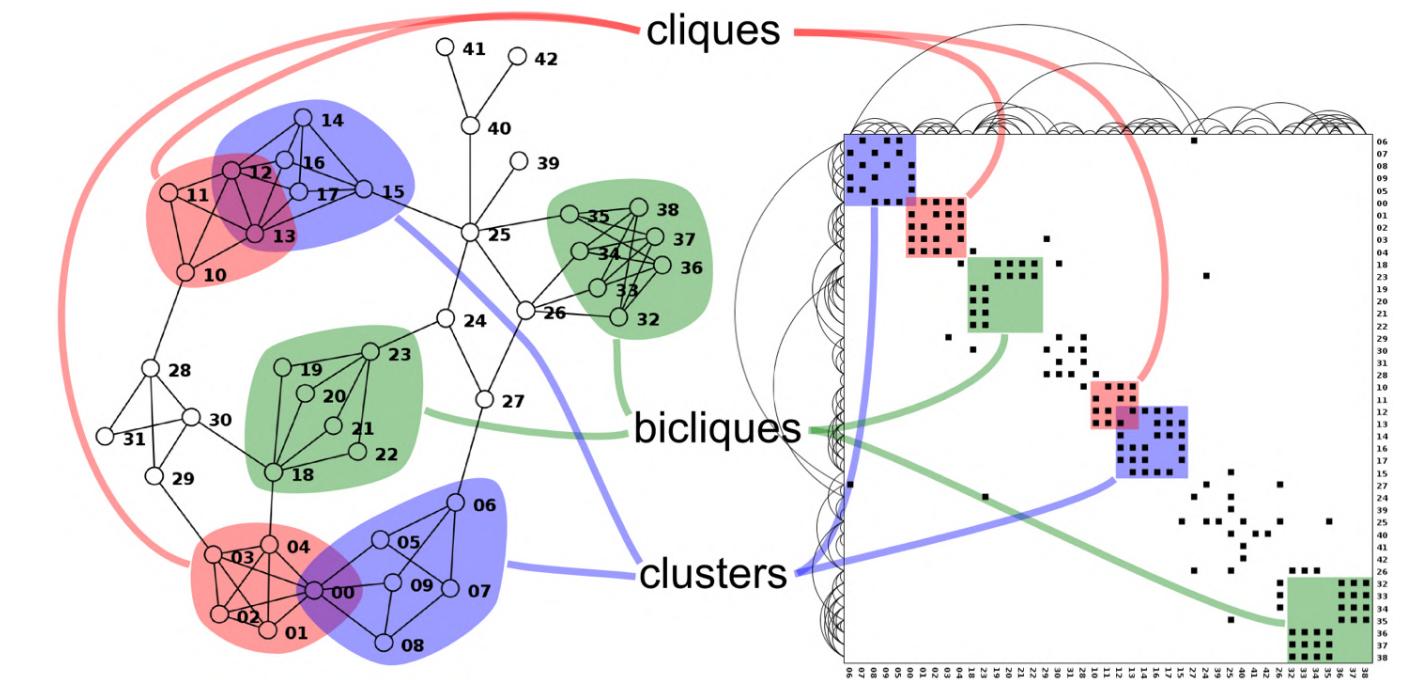
[NodeTrix: a Hybrid Visualization of Social Networks.
Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis)
13(6):1302-1309, 2007.]



[Points of view: Networks. Gehlenborg and Wong. Nature Methods 9:115.]

Connection vs. adjacency comparison

- adjacency matrix strengths
 - predictability, scalability, supports reordering
 - some topology tasks trainable
- node-link diagram strengths
 - topology understanding, path tracing
 - intuitive, no training needed
- empirical study
 - node-link best for small networks
 - matrix best for large networks
 - if tasks don't involve topological structure!

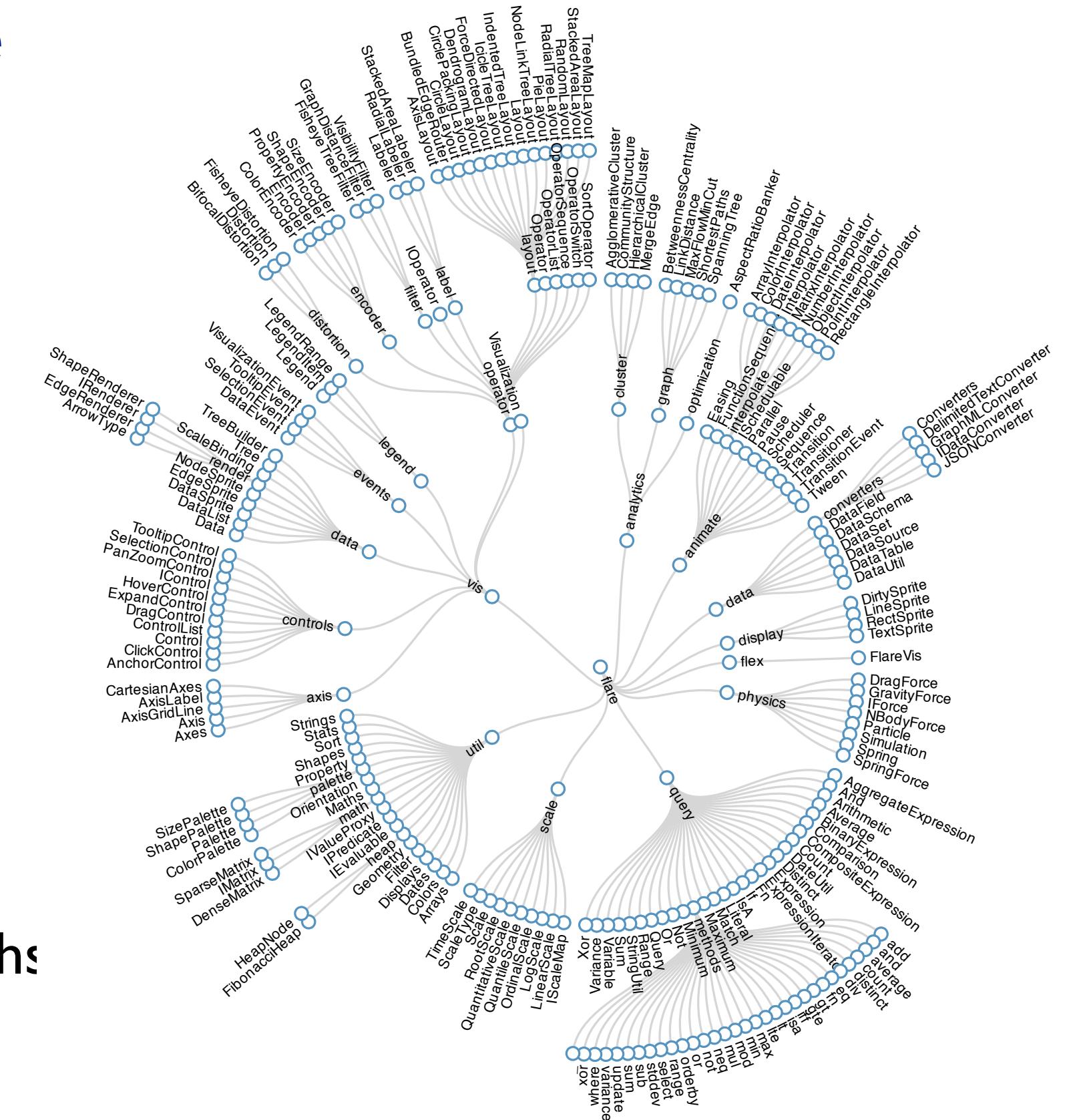


<http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png>

[On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis.
Ghoniem, Fekete, and Castagliola. *Information Visualization* 4:2 (2005), 114–135.]

Idiom: radial node-link tree

- data
 - tree
 - encoding
 - link connection marks
 - point node marks
 - radial axis orientation
 - angular proximity: siblings
 - distance from center: depth in tree
 - tasks
 - understanding topology, following paths
 - scalability
 - 1K - 10K nodes



Idiom: treemap

- data
 - tree
 - 1 quant attrib at leaf nodes
- encoding
 - area containment marks for hierarchical structure
 - rectilinear orientation
 - size encodes quant attrib
- tasks
 - query attribute at leaf nodes
- scalability
 - 1M leaf nodes

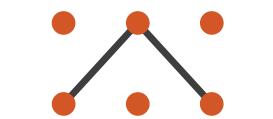


http://tulip.labri.fr/Documentation/3_7/userHandbook/html/ch06.html

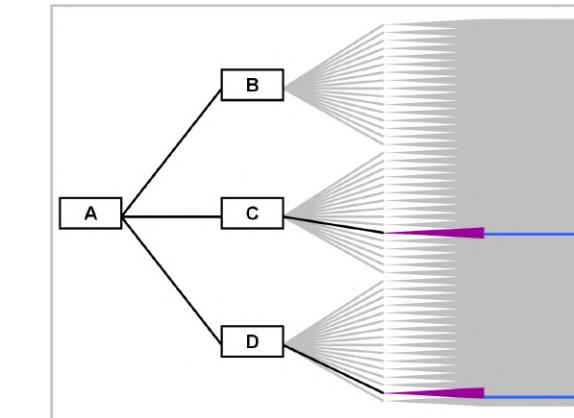
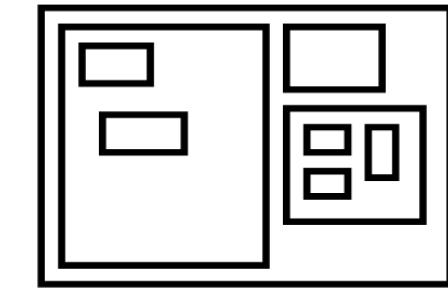
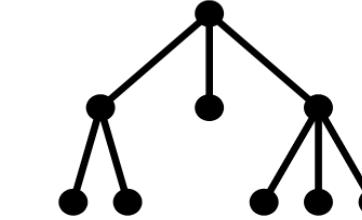
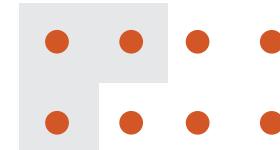
Link marks: Connection and containment

- marks as links (vs. nodes)
 - common case in network drawing
 - 1D case: connection
 - ex: all node-link diagrams
 - emphasizes topology, path tracing
 - networks and trees
 - 2D case: containment
 - ex: all treemap variants
 - emphasizes attribute values at leaves (size coding)
 - only trees

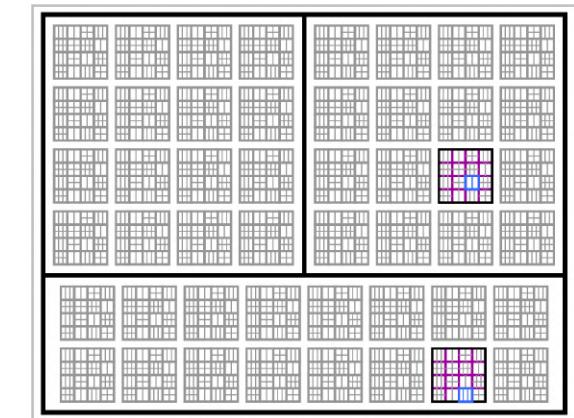
→ Connection



→ Containment



Node-Link Diagram

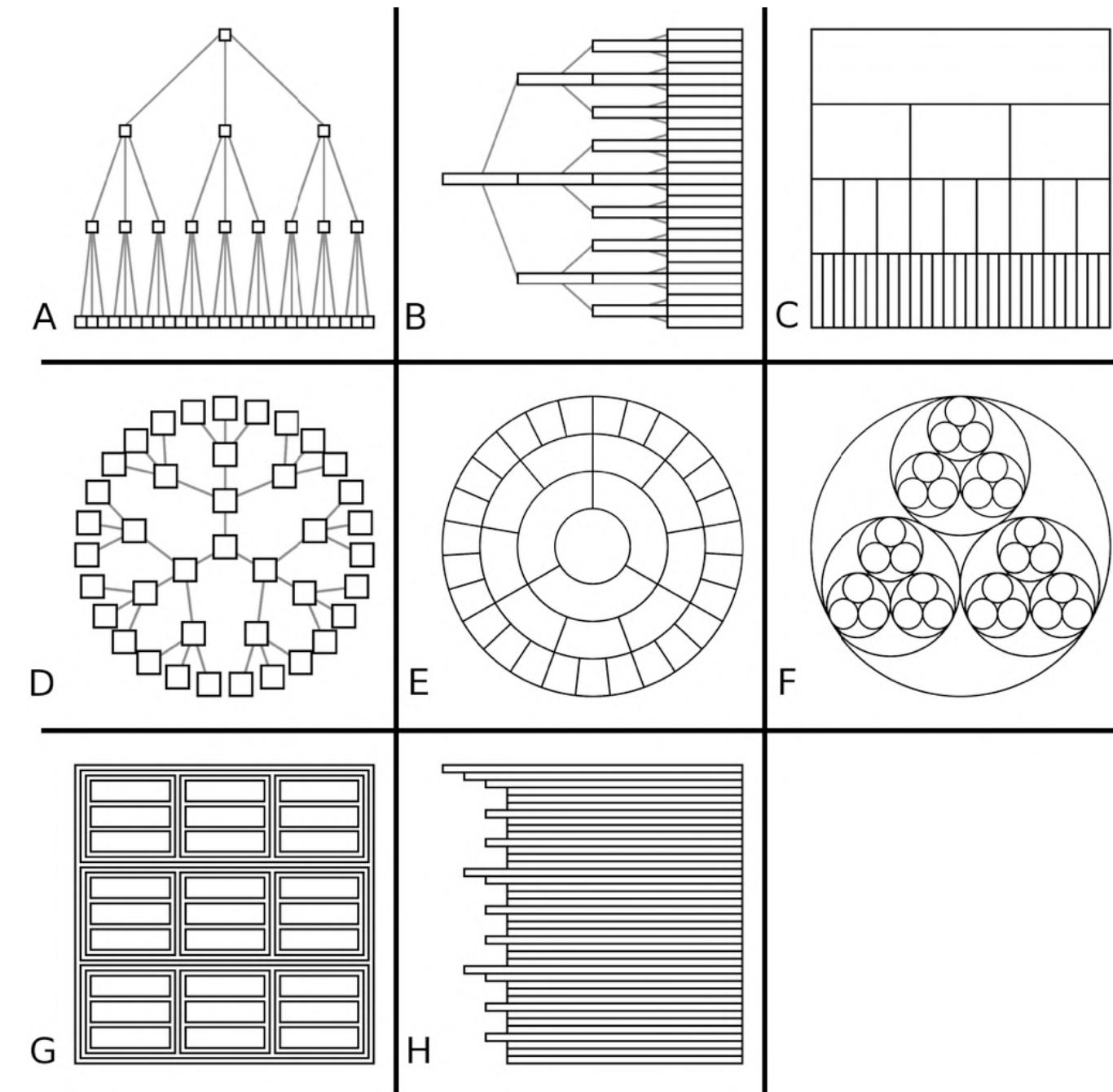


Treemap

[*Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams*. Dong, McGuffin, and Chignell. Proc. InfoVis 2005, p. 57-64.]

Tree drawing idioms comparison

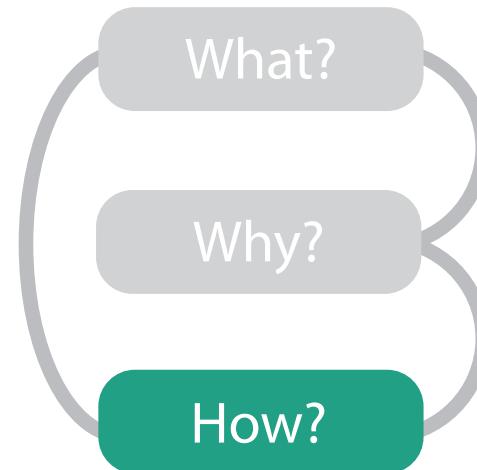
- data shown
 - link relationships
 - tree depth
 - sibling order
- design choices
 - connection vs containment link marks
 - rectilinear vs radial layout
 - spatial position channels
- considerations
 - redundant? arbitrary?
 - information density?
 - avoid wasting space



[Quantifying the Space-Efficiency of 2D Graphical Representations of Trees. McGuffin and Robert. Information Visualization 9:2 (2010), 115–140.]

Idiom design choices: Encode

Encode



→ Arrange

→ Express



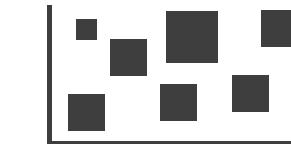
→ Order



→ Use



→ Separate



→ Align



→ Map

from categorical and ordered attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...

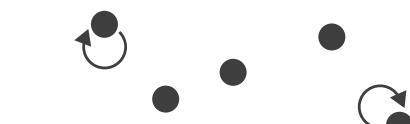


→ Shape

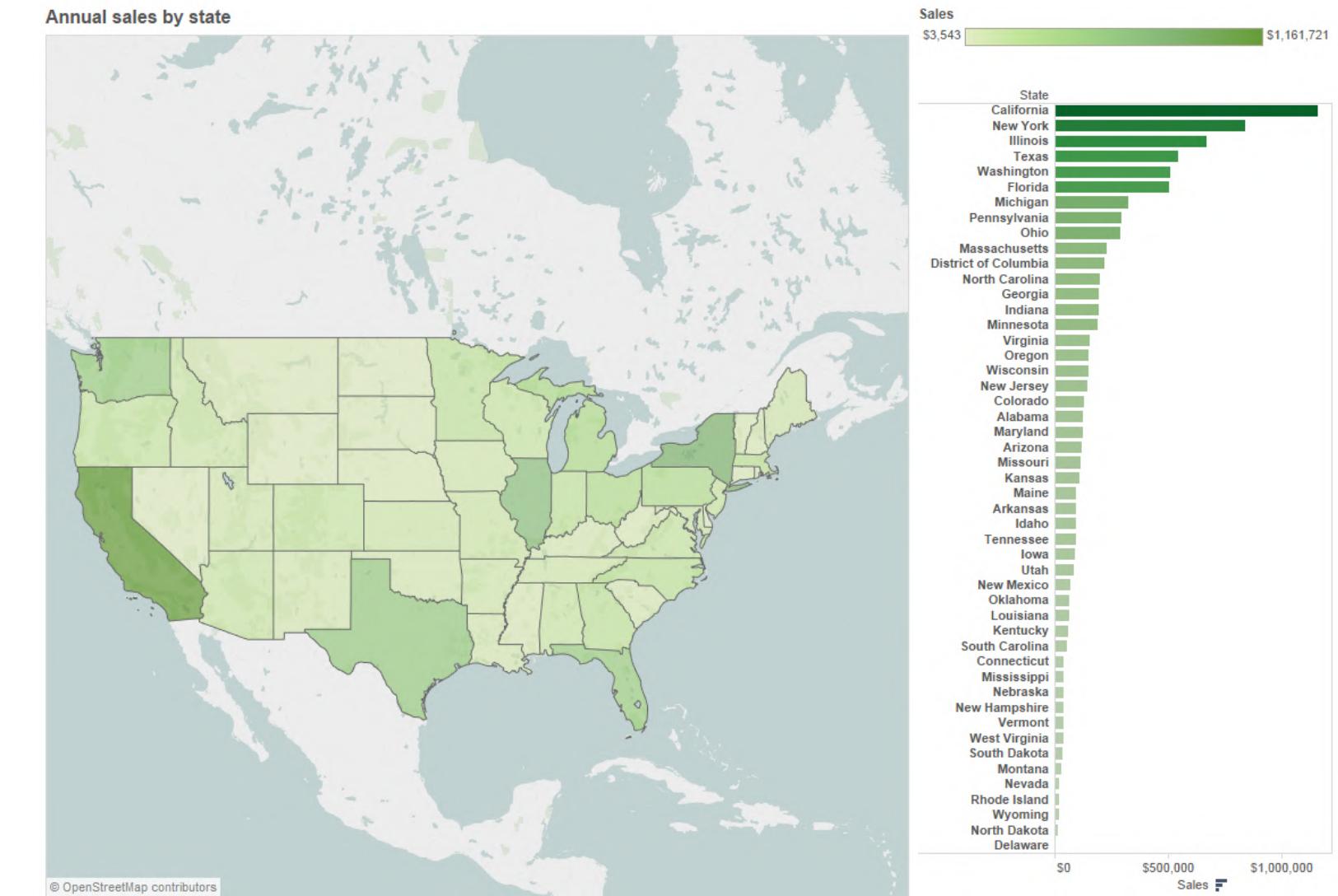
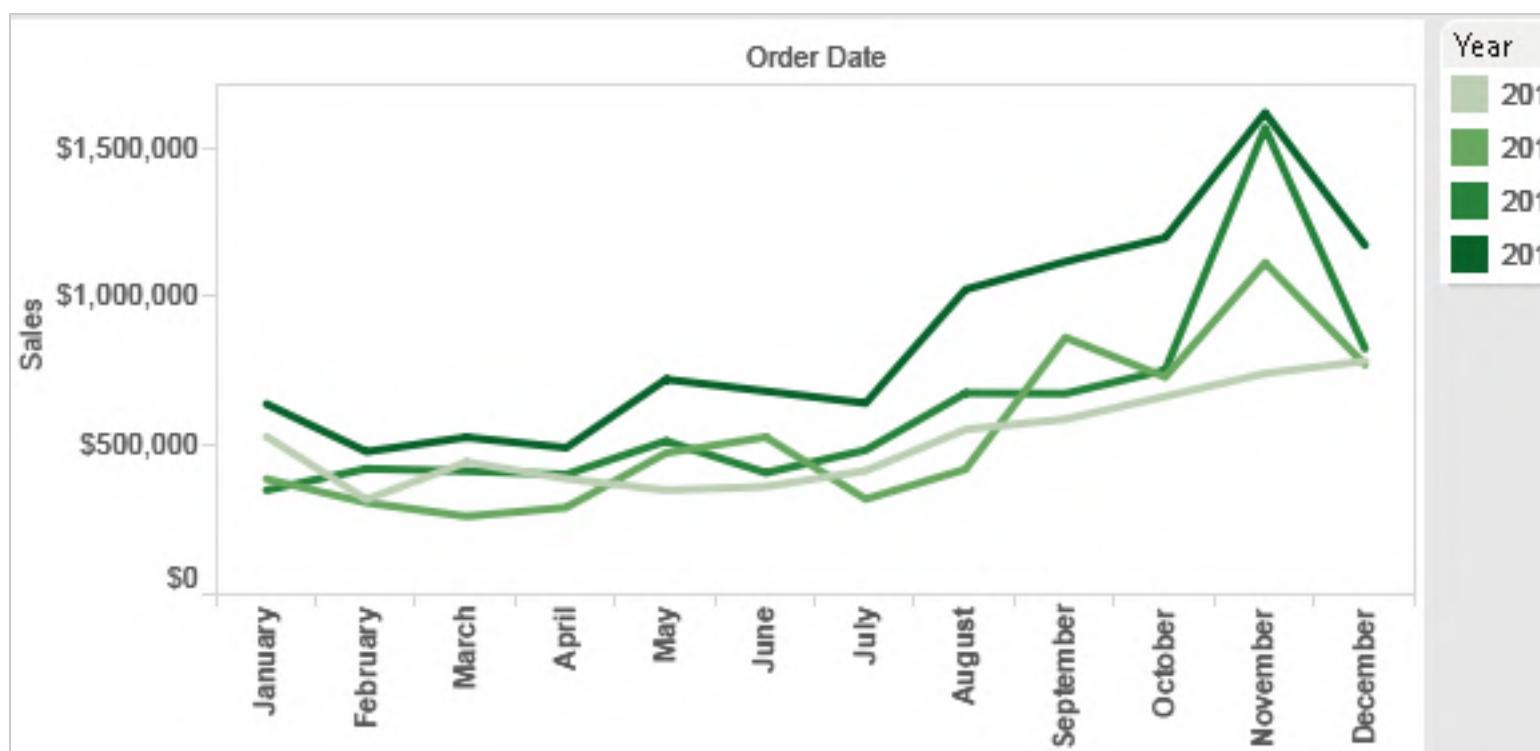
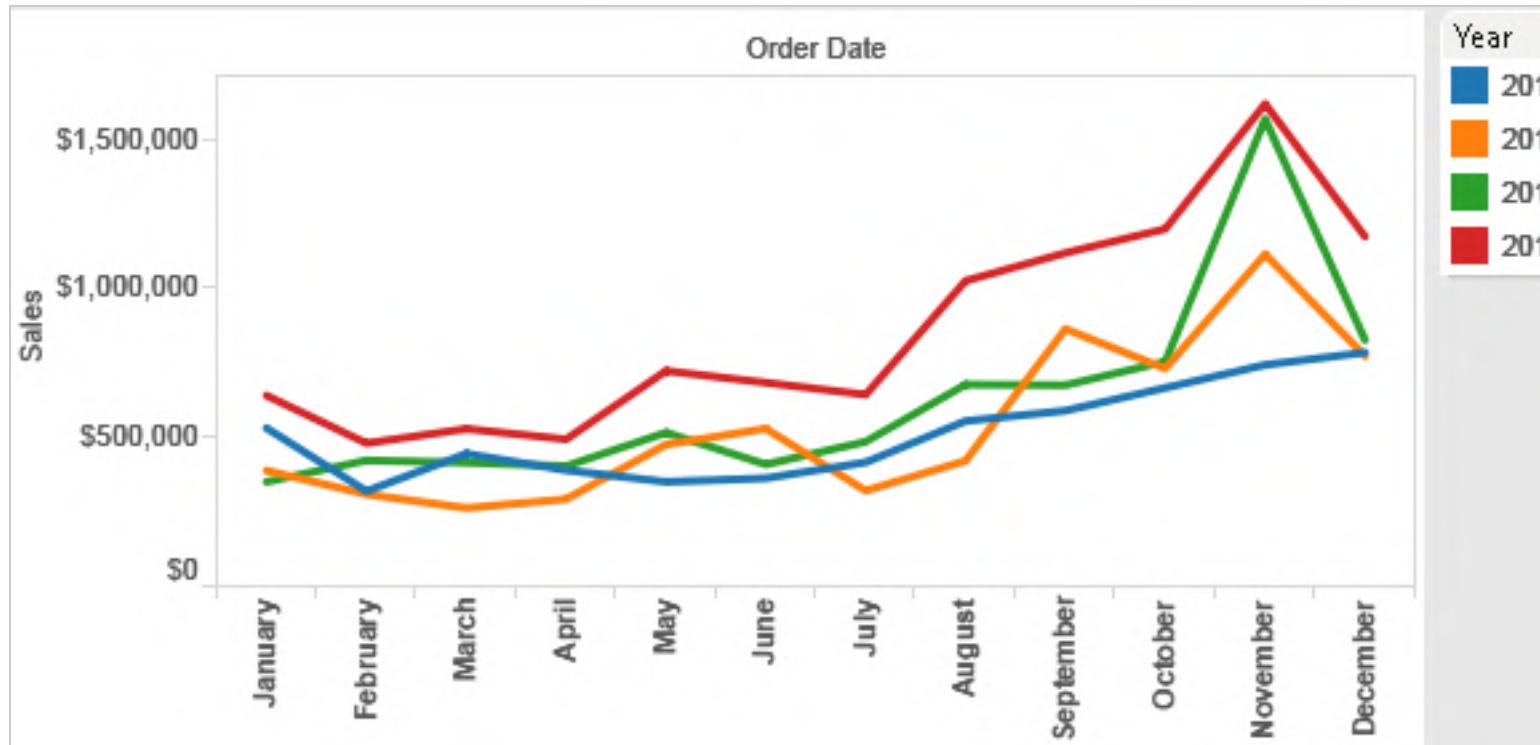


→ Motion

Direction, Rate, Frequency, ...



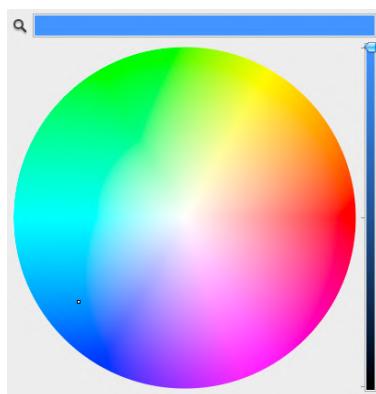
Categorical vs ordered color



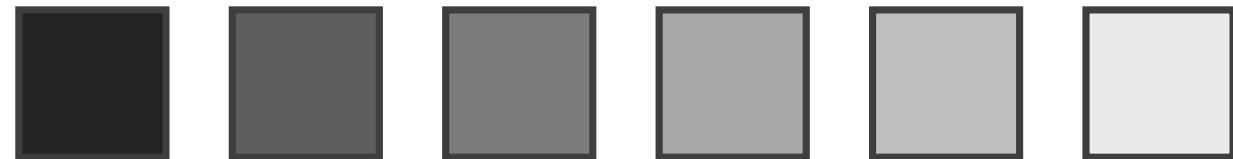
[Seriously Colorful: Advanced Color Principles & Practices.
Stone.Tableau Customer Conference 2014.]

Color: Luminance, saturation, hue

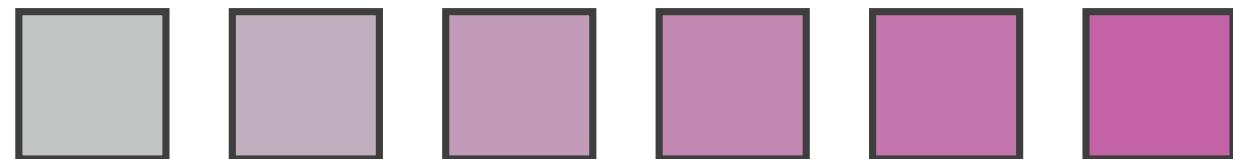
- 3 channels
 - identity for categorical
 - hue
 - magnitude for ordered
 - luminance
 - saturation
- RGB: poor for encoding
- HSL: better, but beware
 - lightness \neq luminance



Luminance



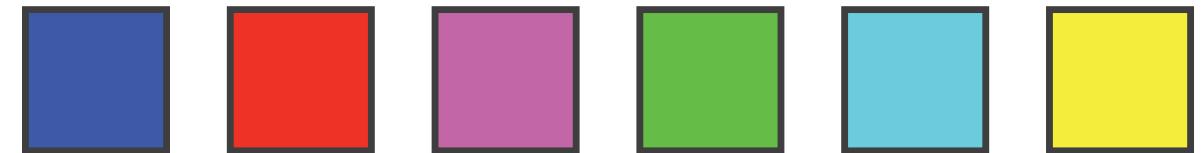
Saturation



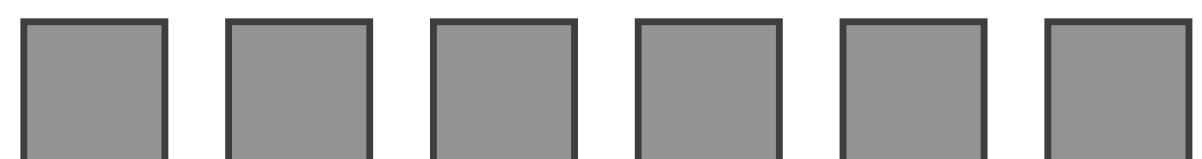
Hue



Corners of the RGB
color cube



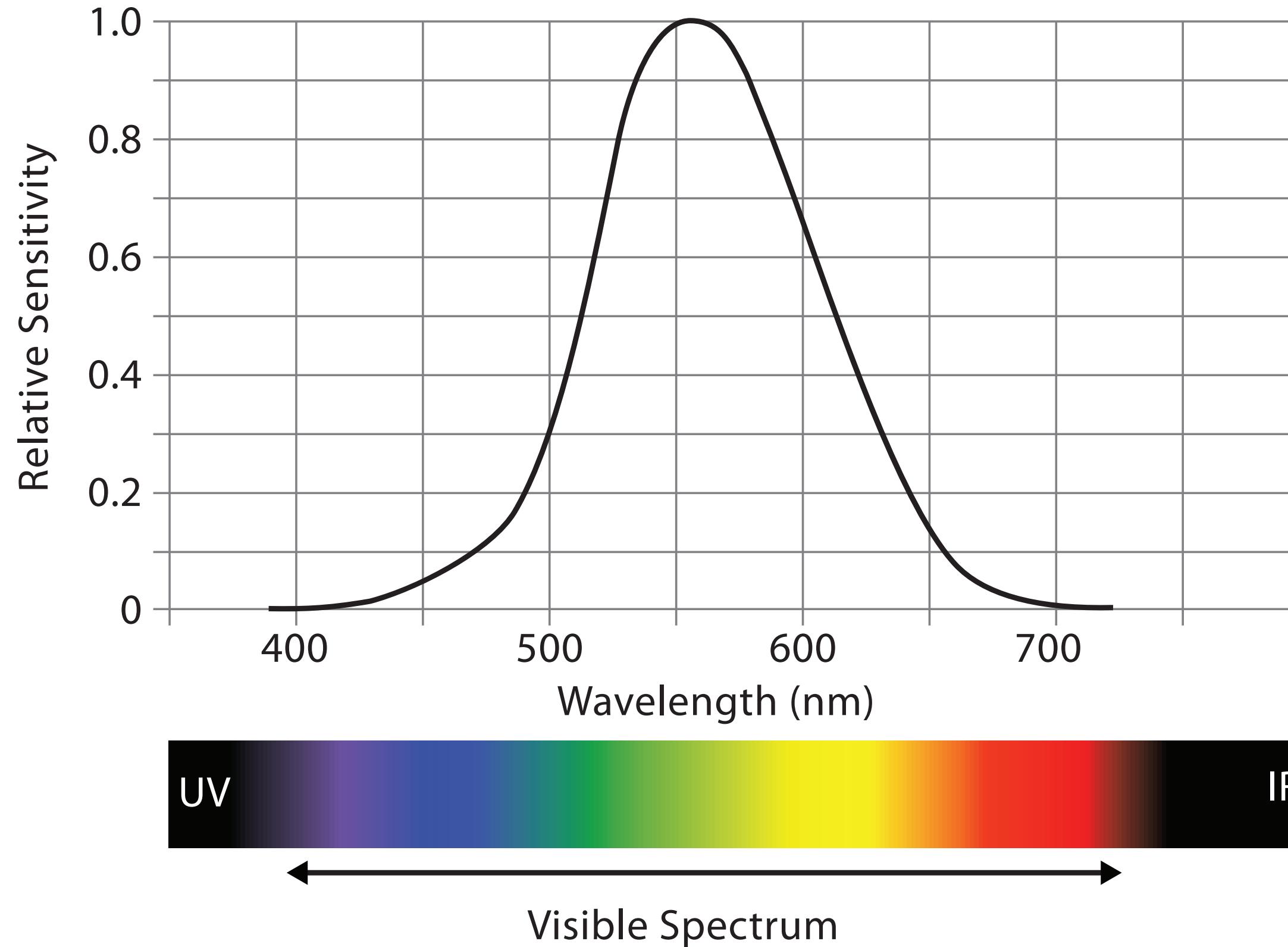
L from HLS
All the same



Luminance values

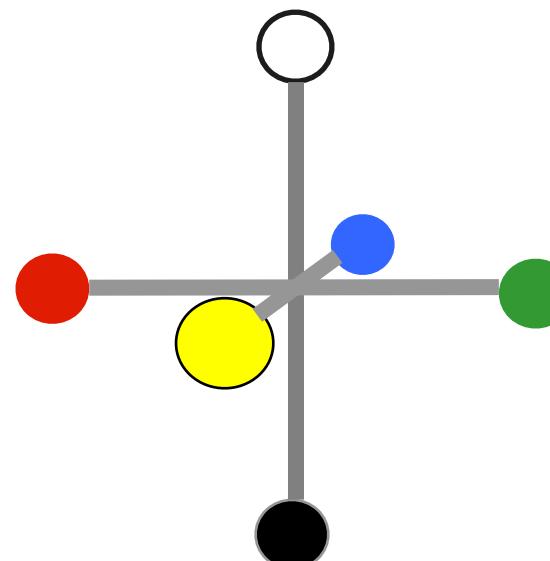
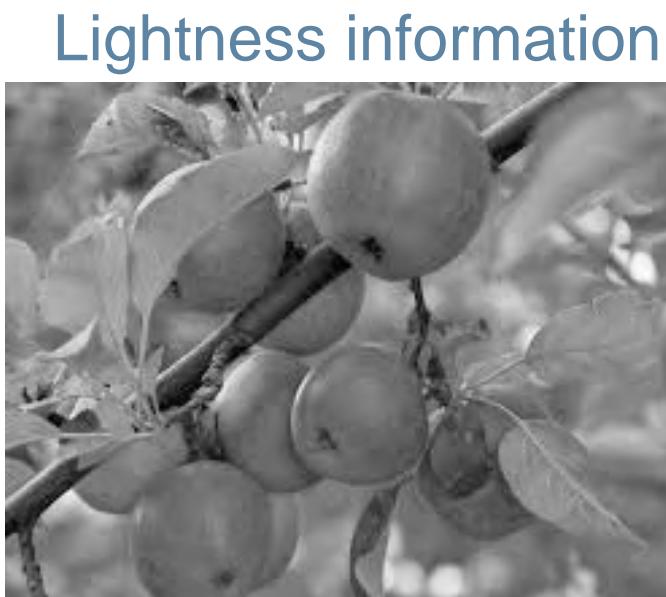


Spectral sensitivity



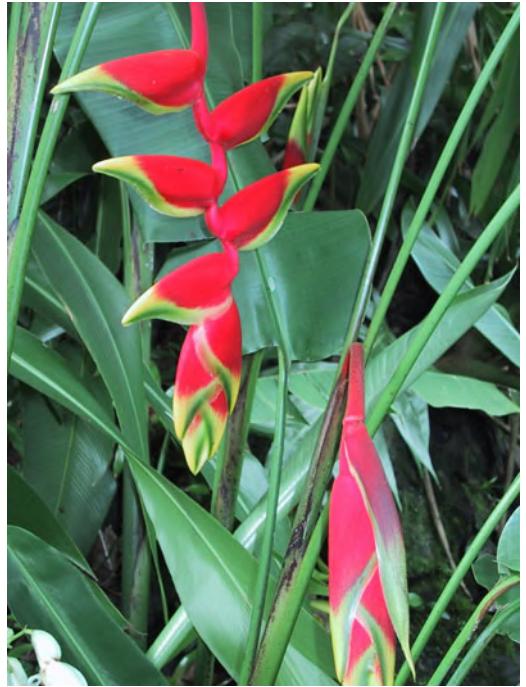
Opponent color and color deficiency

- 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



Normal
vision



Deutanope

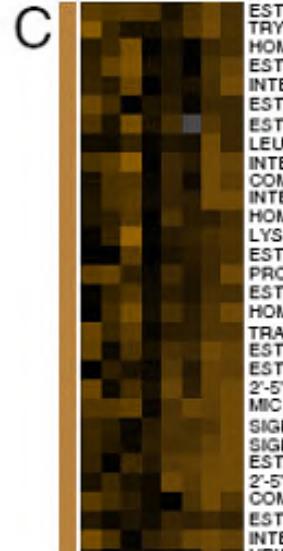
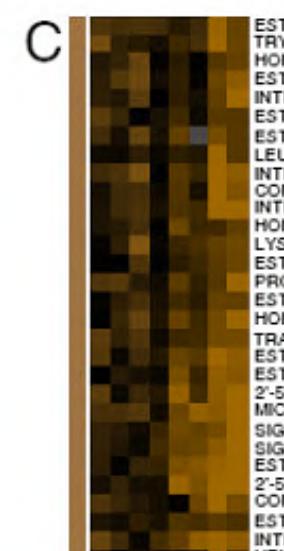


Protanope



Tritanope

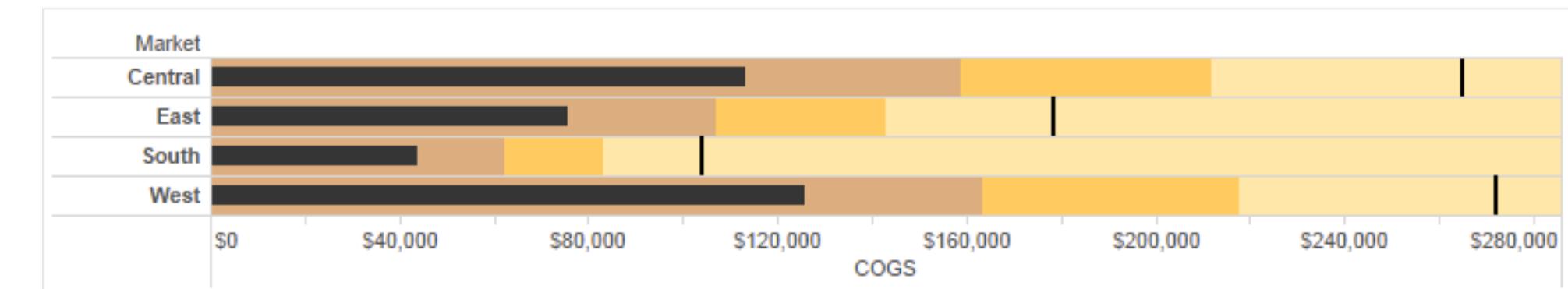
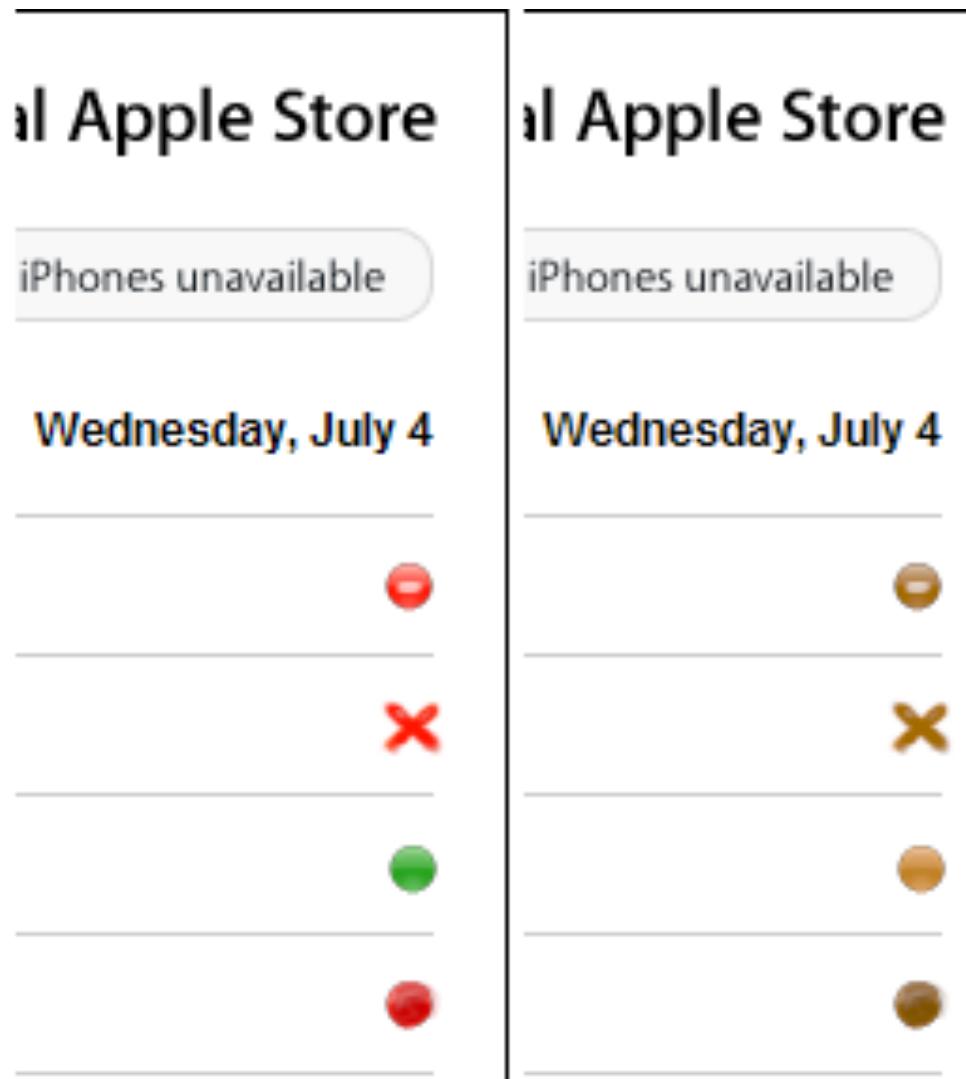
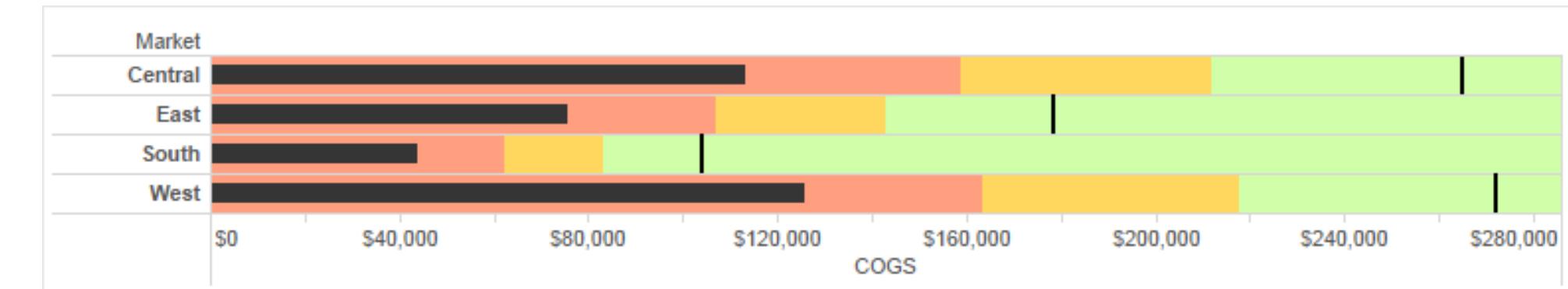
<http://rehue.net>



[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

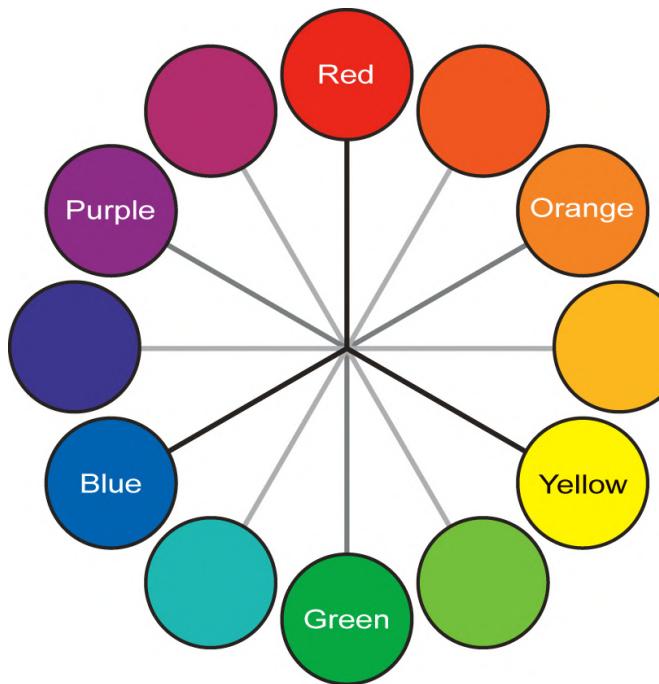


Deutanope simulation

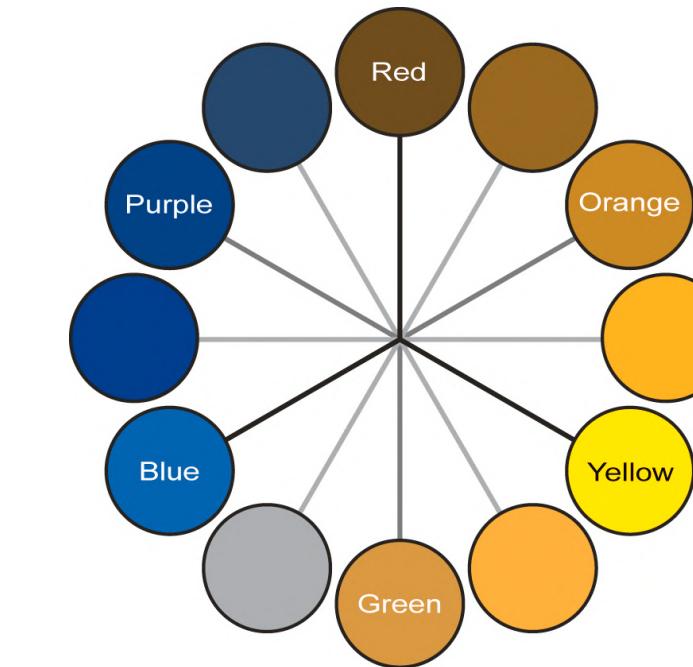
Change the shape

Vary luminance

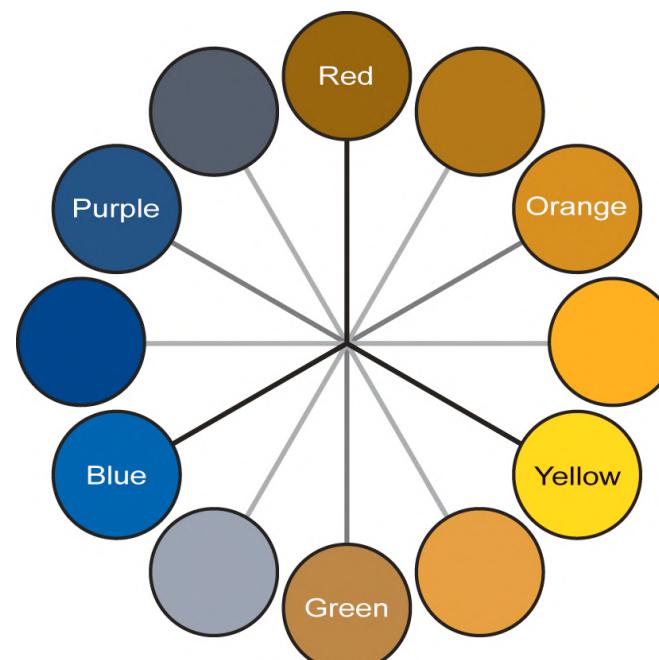
Color deficiency: Reduces color to 2 dimensions



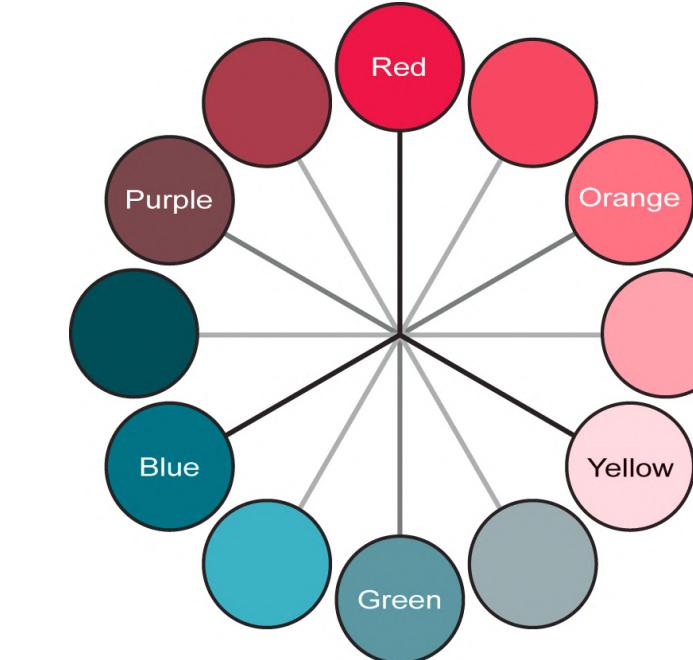
Normal



Protanope



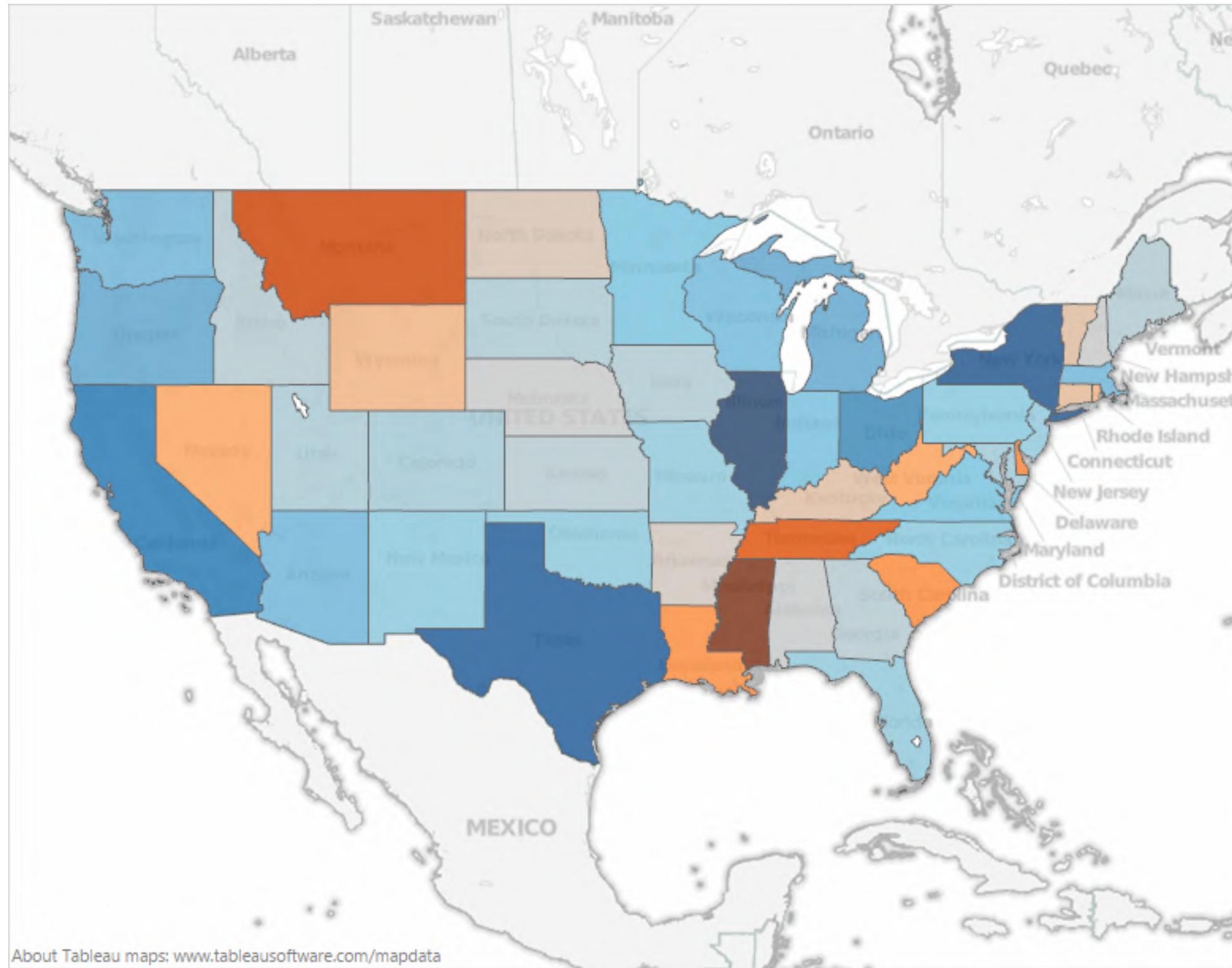
Deuteranope



Tritanope

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

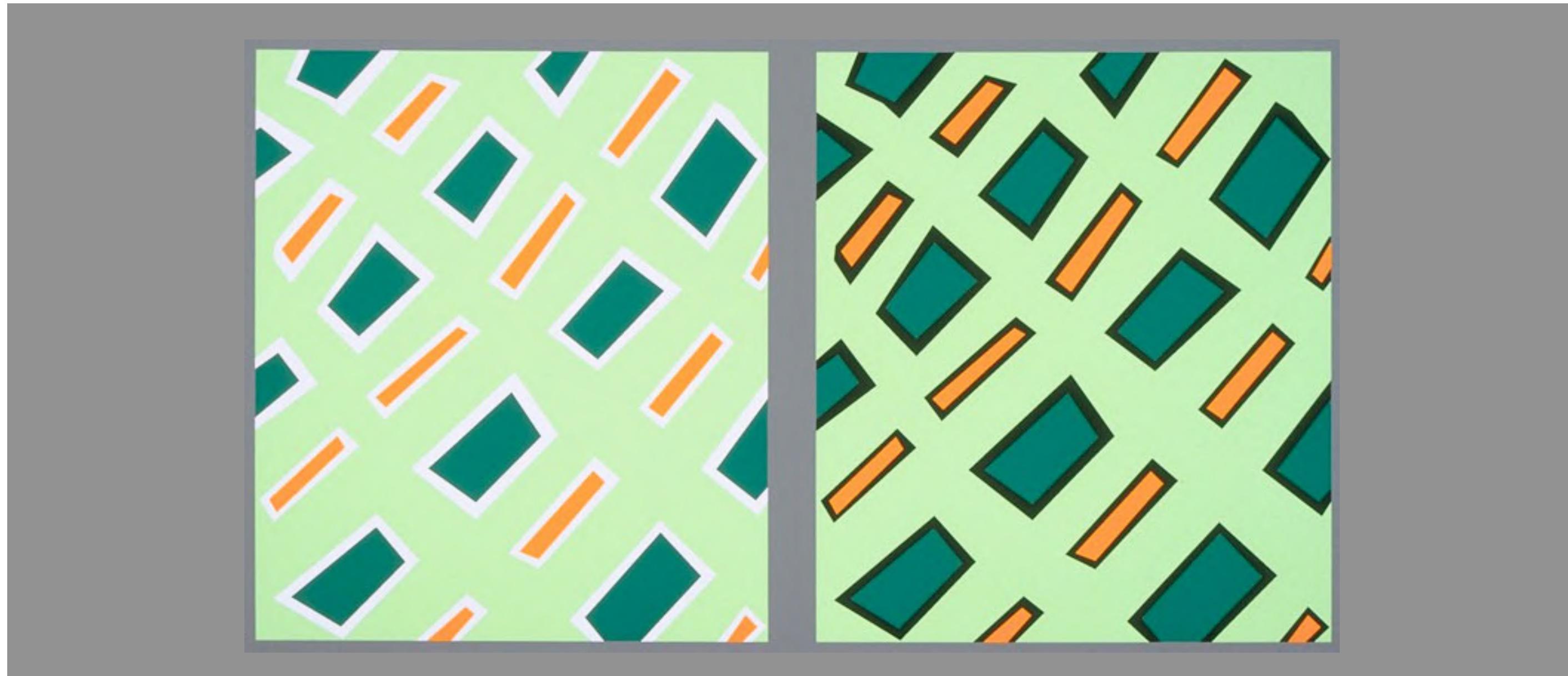
Designing for color deficiency: Blue-Orange is safe



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

Bezold Effect: Outlines matter

- color constancy: simultaneous contrast effect



Color/Lightness constancy: Illumination conditions

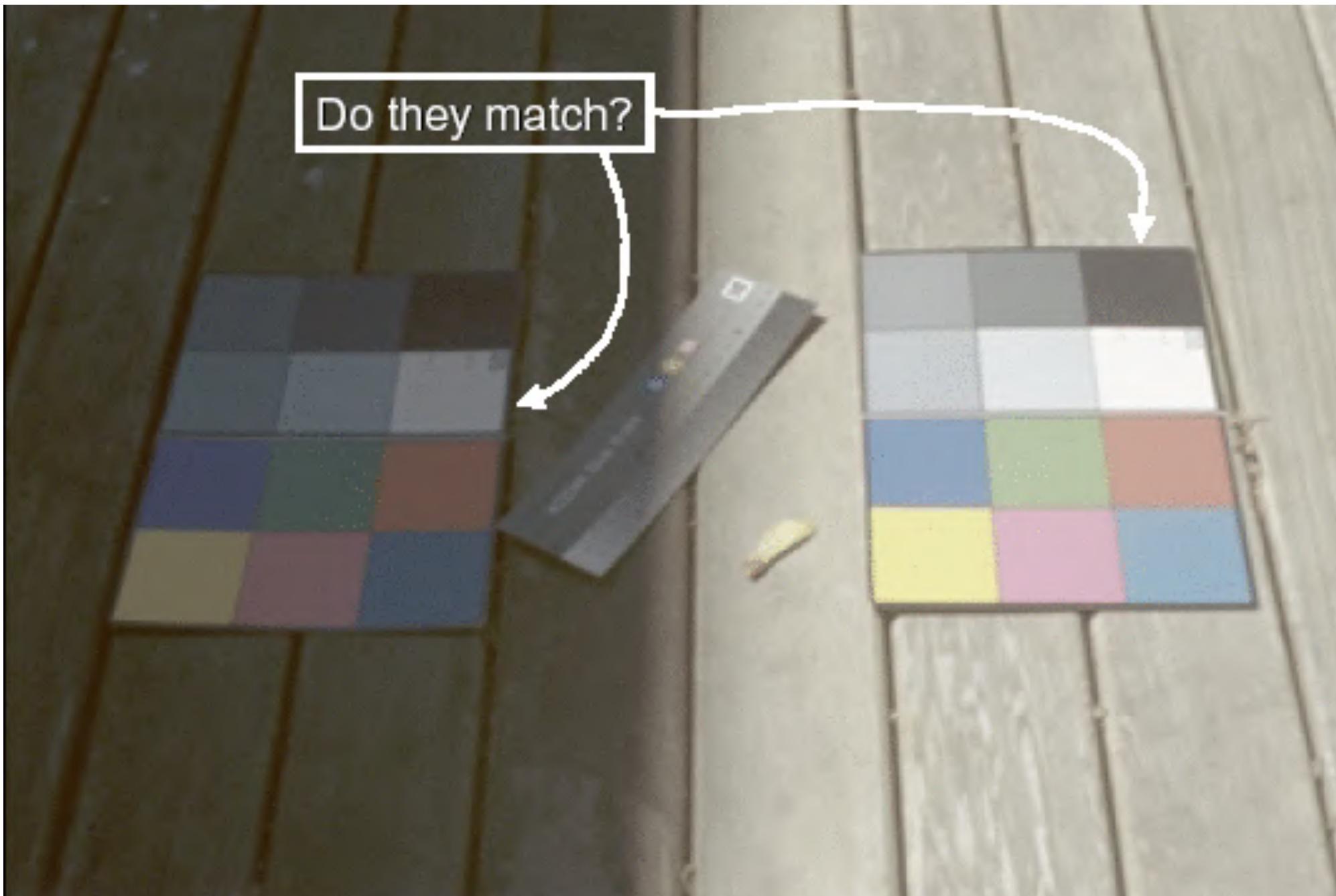


Image courtesy of John McCann

Color/Lightness constancy: Illumination conditions

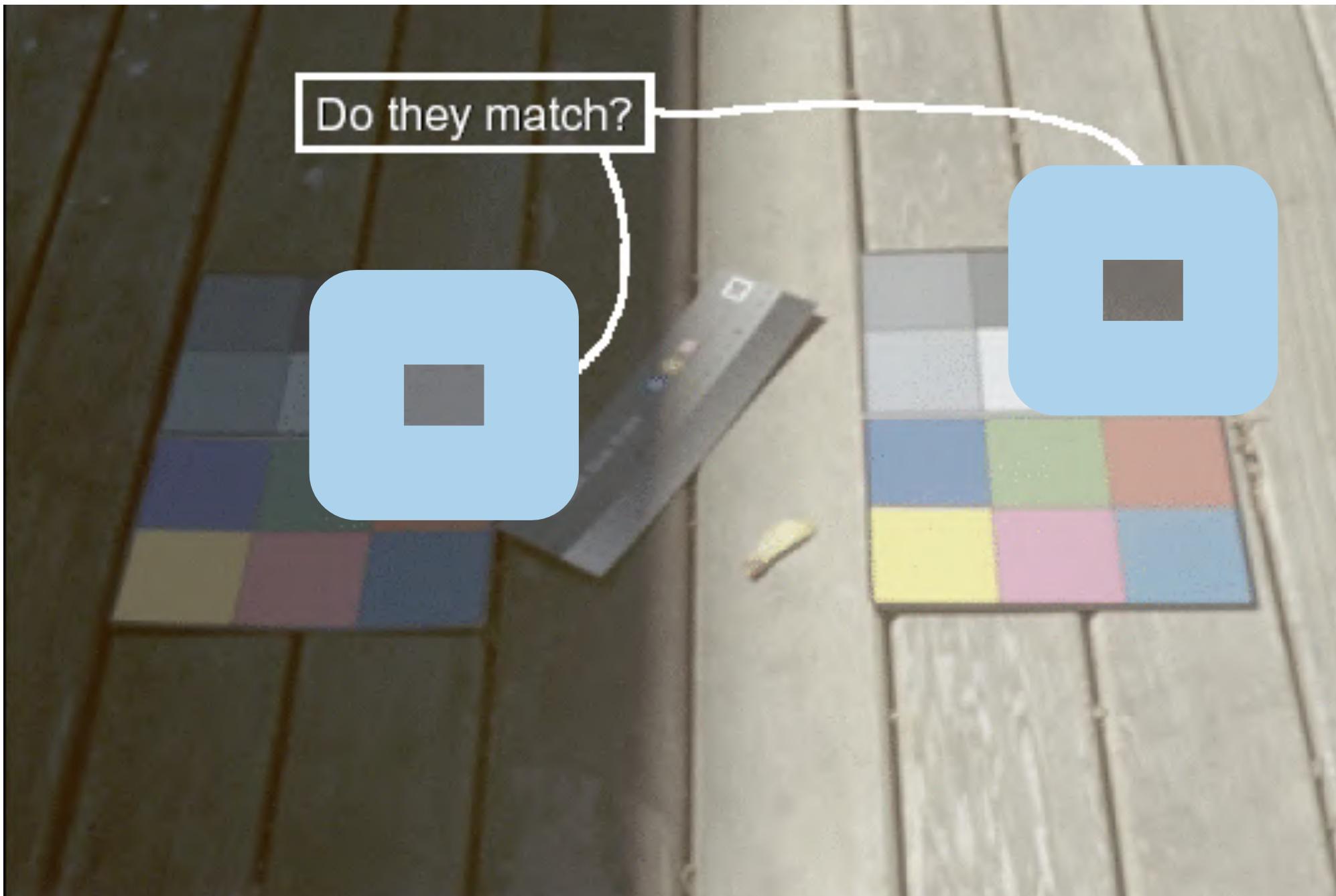


Image courtesy of John McCann

Colormaps

→ Categorical



→ Ordered

→ *Sequential*



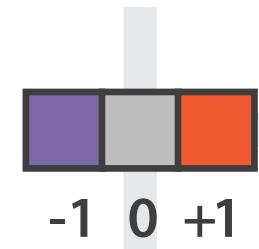
→ *Diverging*



Binary



Diverging



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

Colormaps

→ Categorical

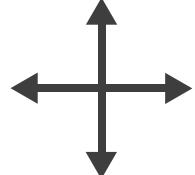


→ Ordered

→ Sequential



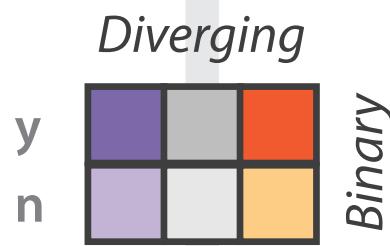
→ Bivariate



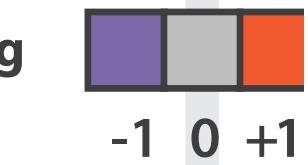
→ Diverging



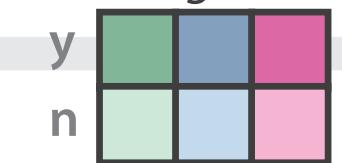
Binary



Diverging



Categorical



T F A

Binary

y n

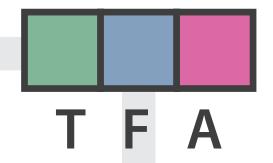
y n

T F A

Binary

y n

Categorical



T F A

Categorical

3 2 1

Sequential



T F A

Sequential

3 2 1

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

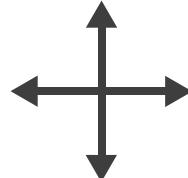
→ Sequential



→ Diverging

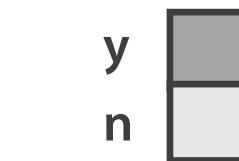


→ Bivariate

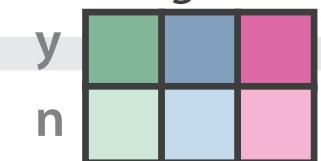


use with care!

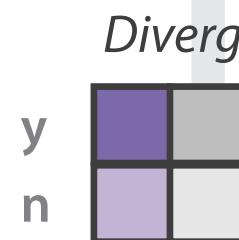
Binary



Categorical



Binary



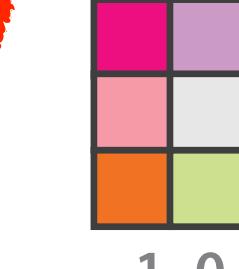
Diverging



Diverging



Diverging



Diverging

-1 0 +1

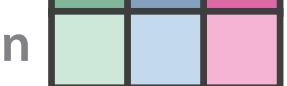
-1 0 +1

-1 0 +1

Diverging

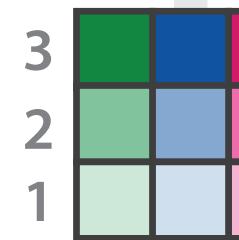
Binary

Binary



T F A

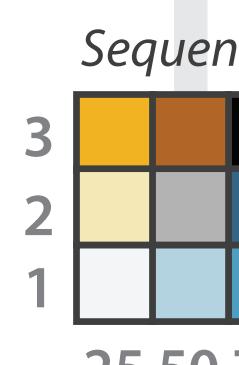
Categorical



Sequential

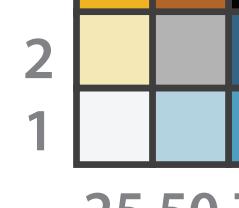
T F A

Categorical



Sequential

Sequential



25 50 75

25 50 75

25 50 75

Categorical

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

Colormaps

→ Categorical



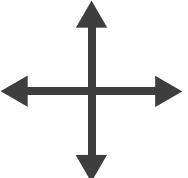
→ Ordered

→ Sequential

→ Diverging



→ Bivariate



- color channel interactions

- size heavily affects salience

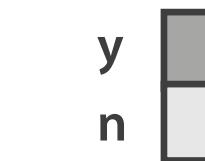
- small regions need high saturation

- large need low saturation

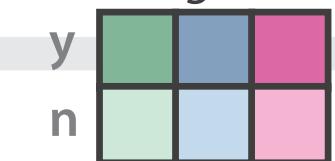
- saturation & luminance: 3-4 bins max

- also not separable from transparency

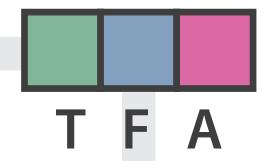
Binary



Categorical

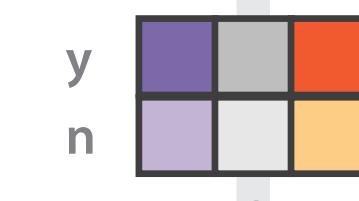


Binary

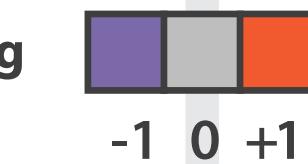


Categorical

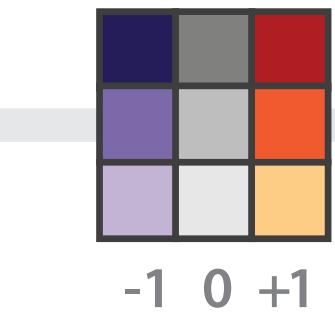
Diverging



Diverging

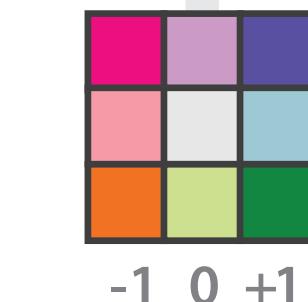


Diverging

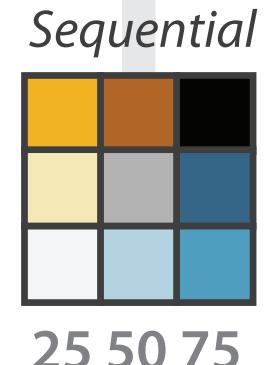


Sequential

Diverging



Diverging

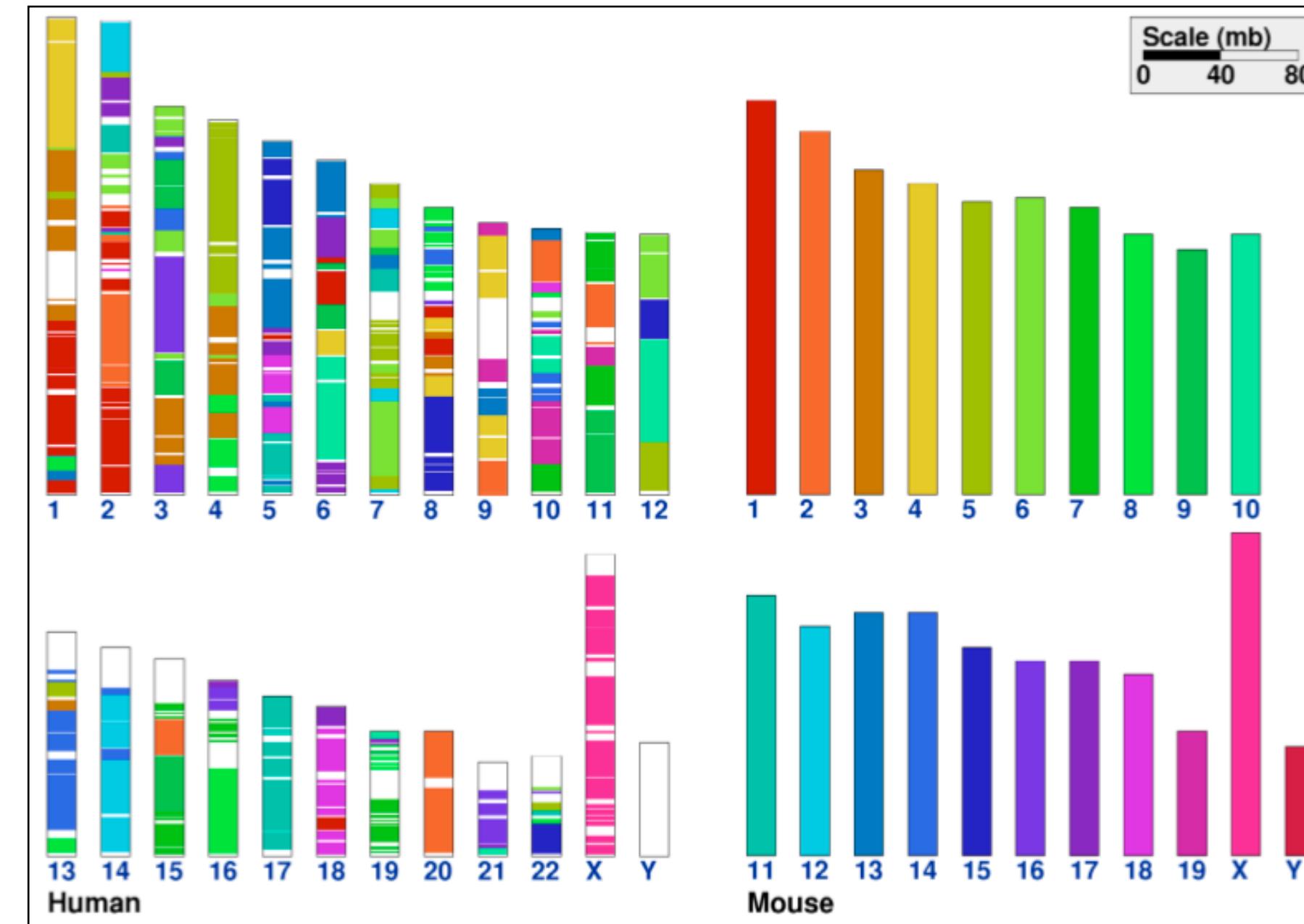


Sequential

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

Categorical color: Discriminability constraints

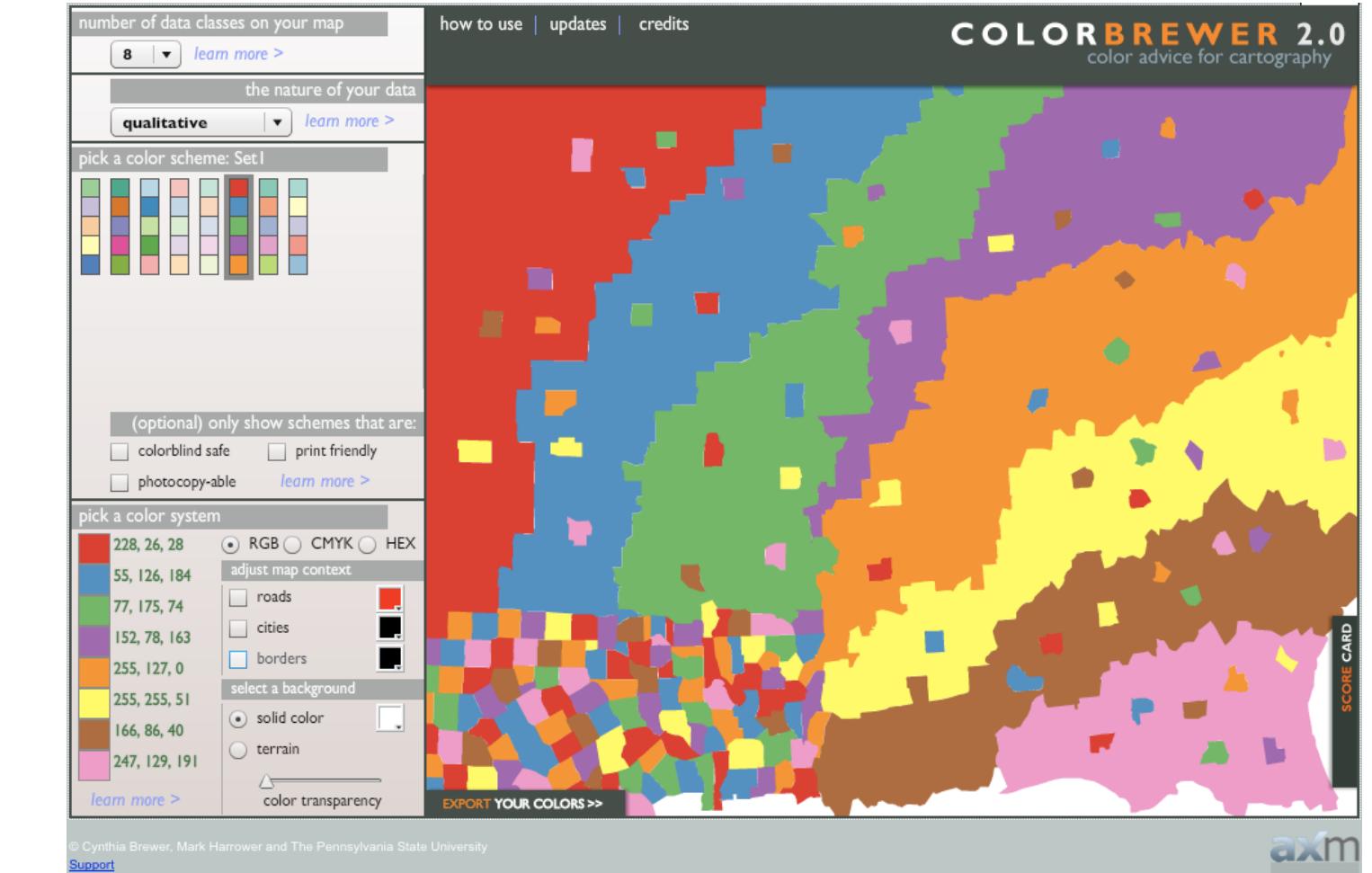
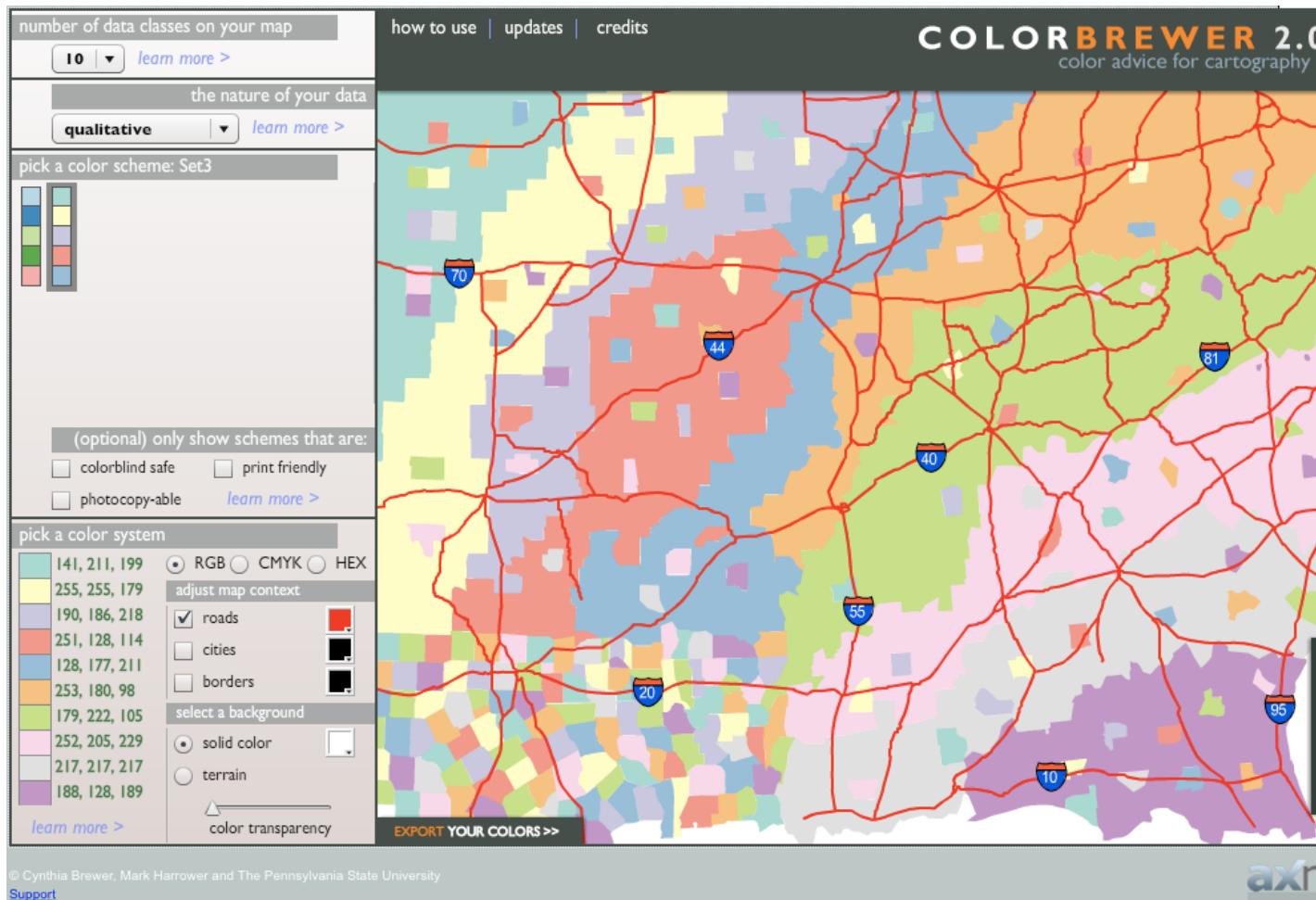
- noncontiguous small regions of color: only 6-12 bins



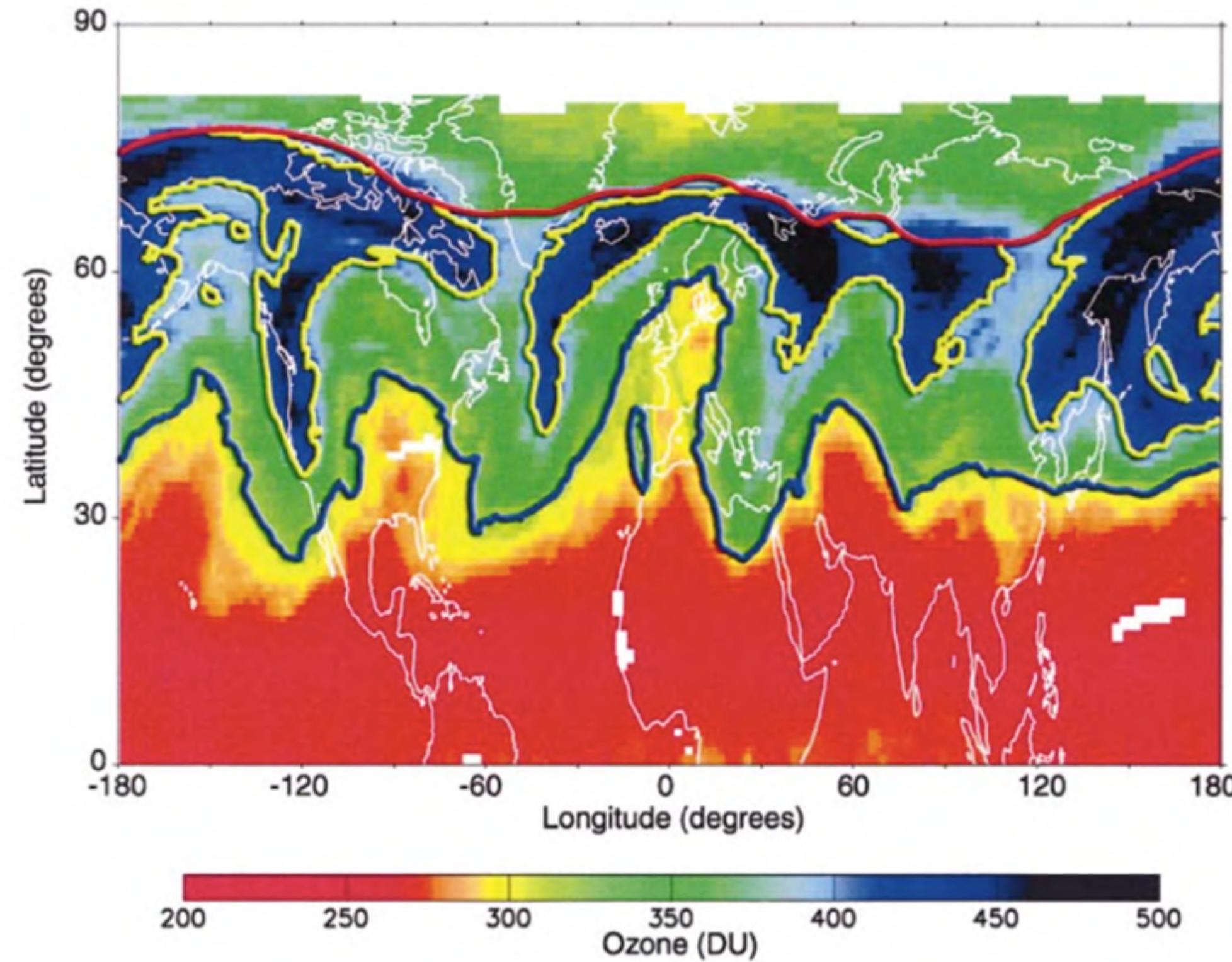
[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

ColorBrewer

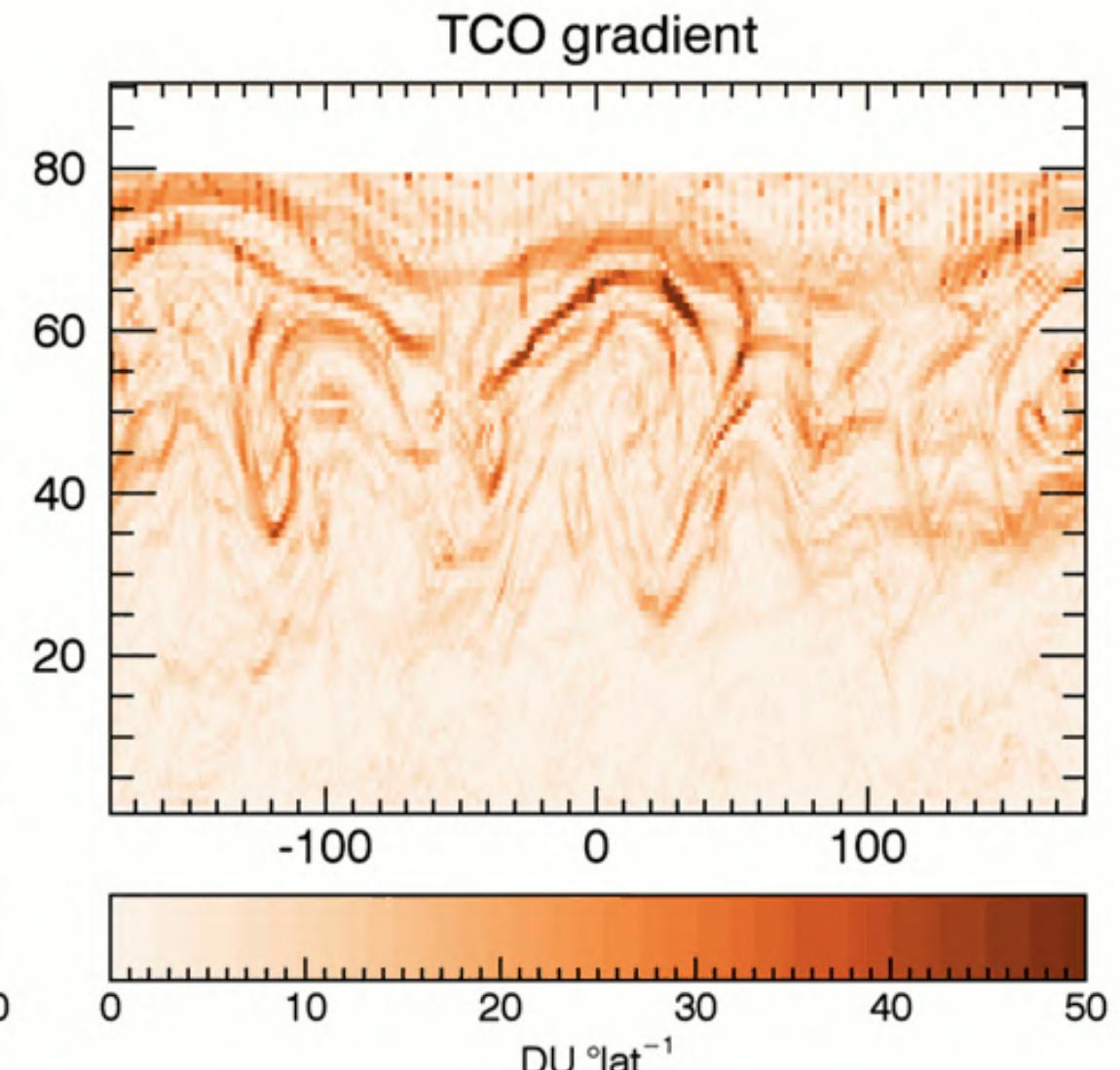
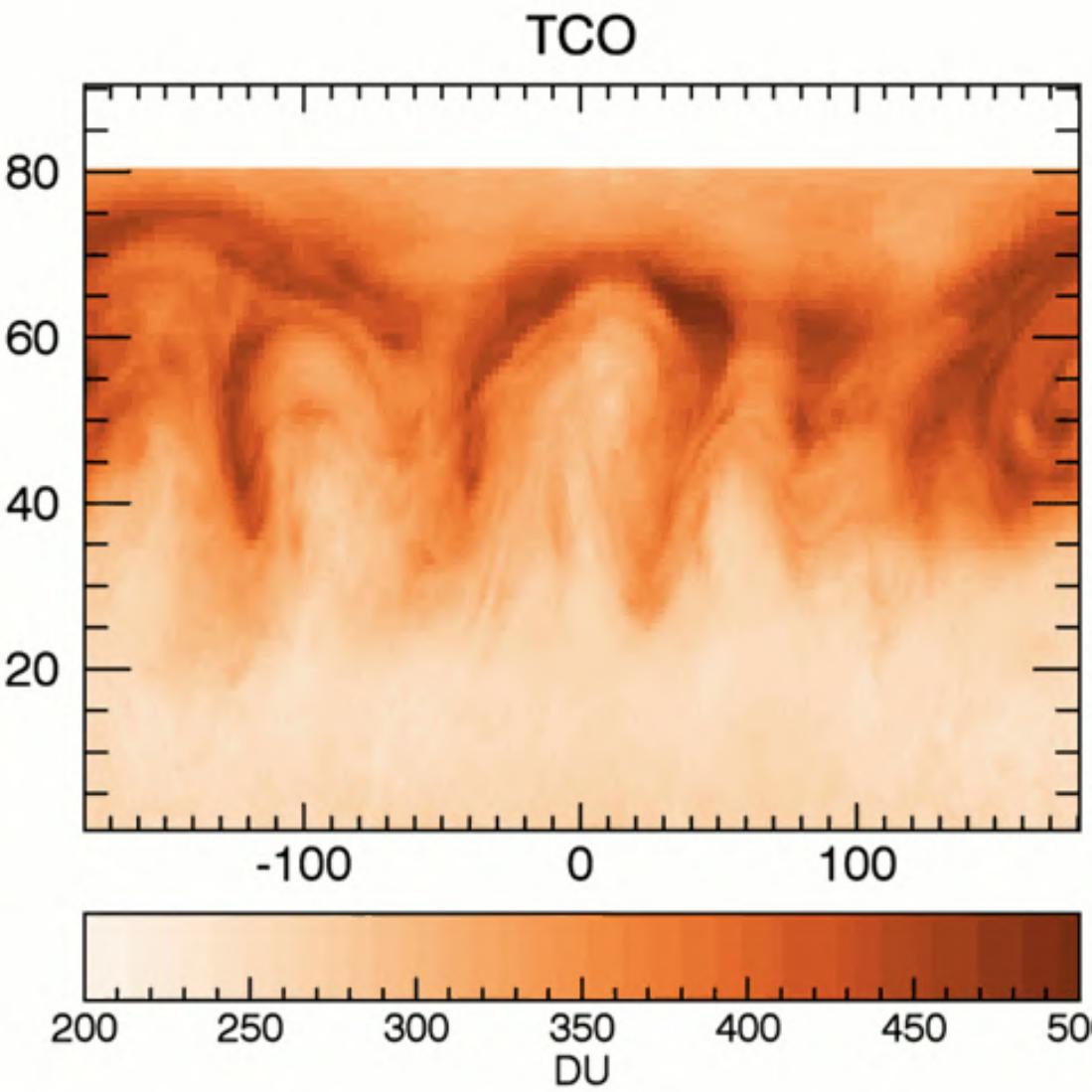
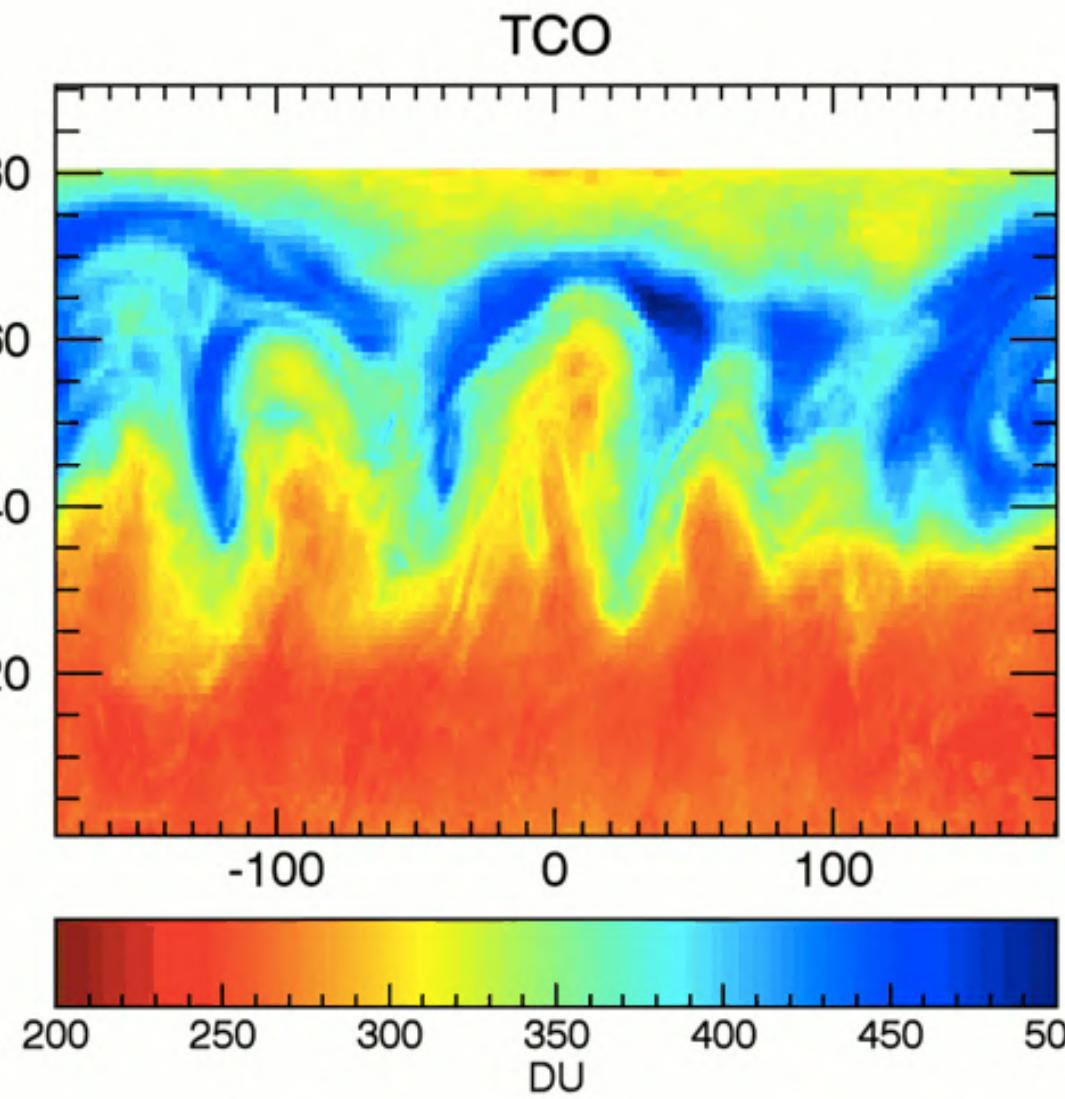
- <http://www.colorbrewer2.org>
- saturation and area example: size affects salience!



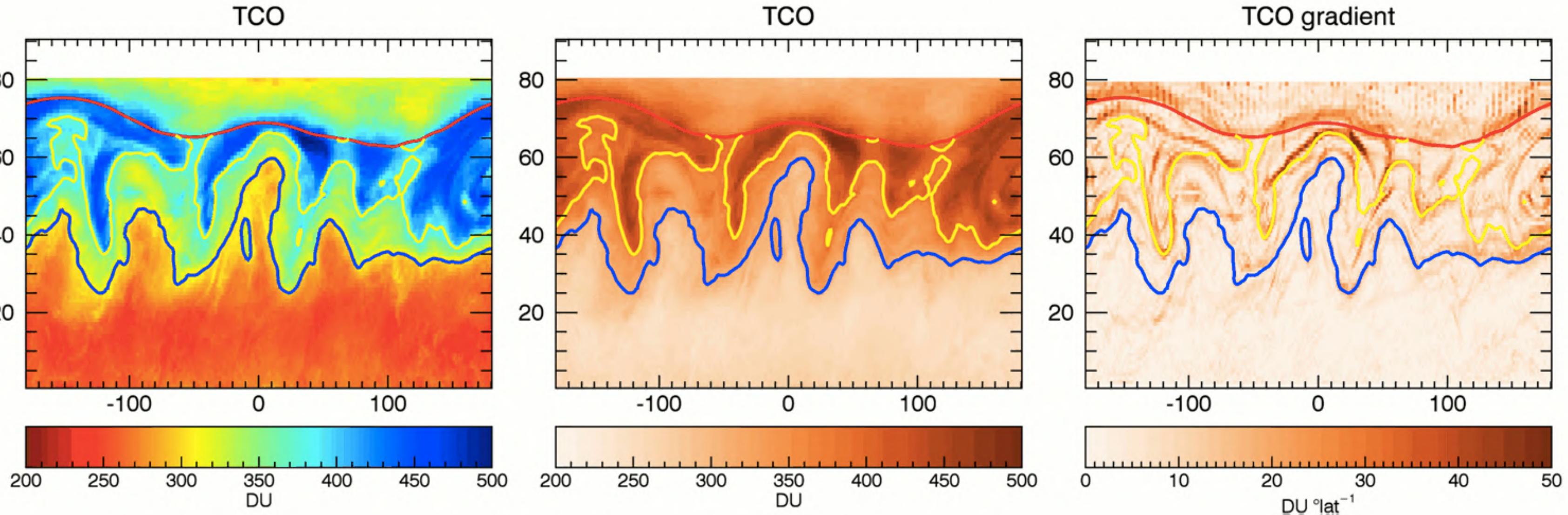
Ordered color: Rainbow is poor default



Ordered color: Rainbow is poor default

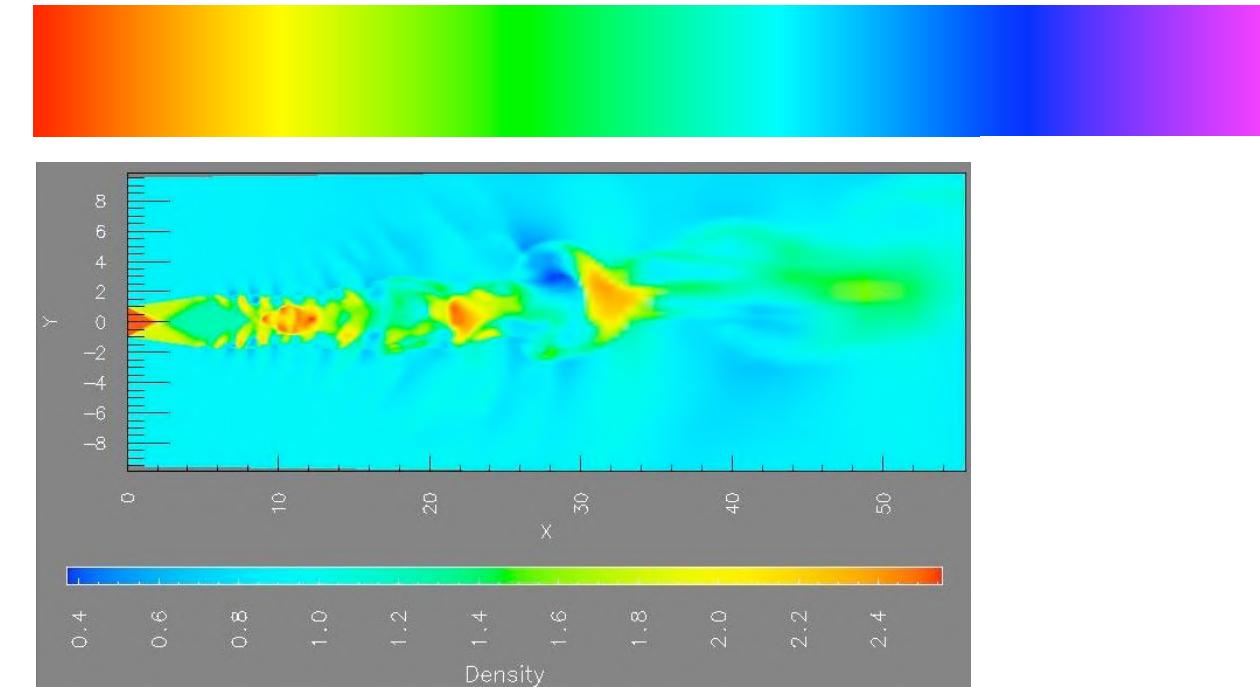


Ordered color: Rainbow is poor default

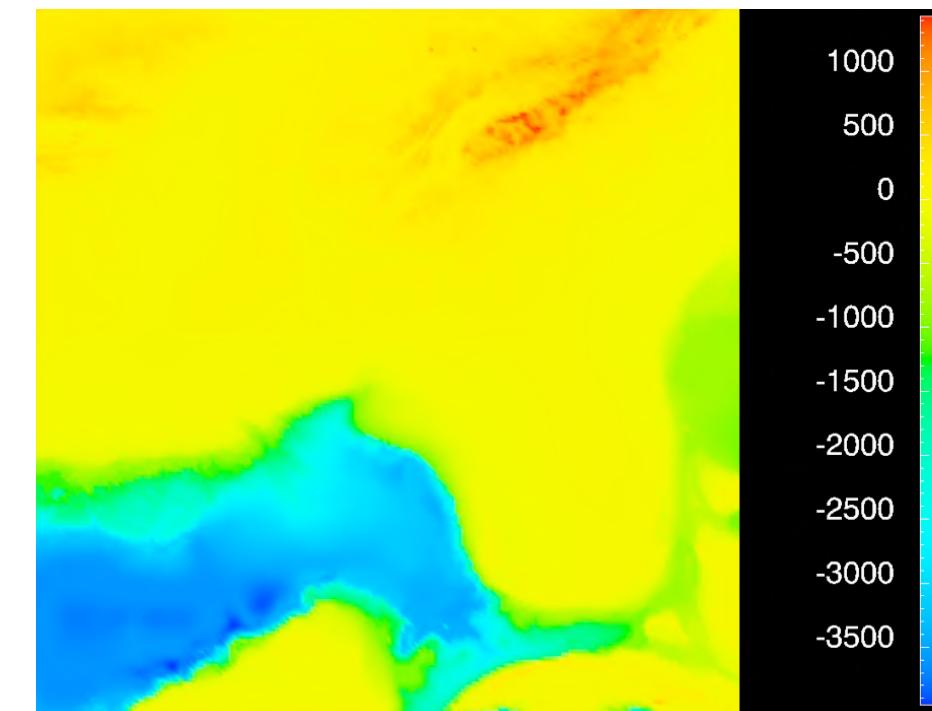


Ordered color: Rainbow is poor default

- 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.]

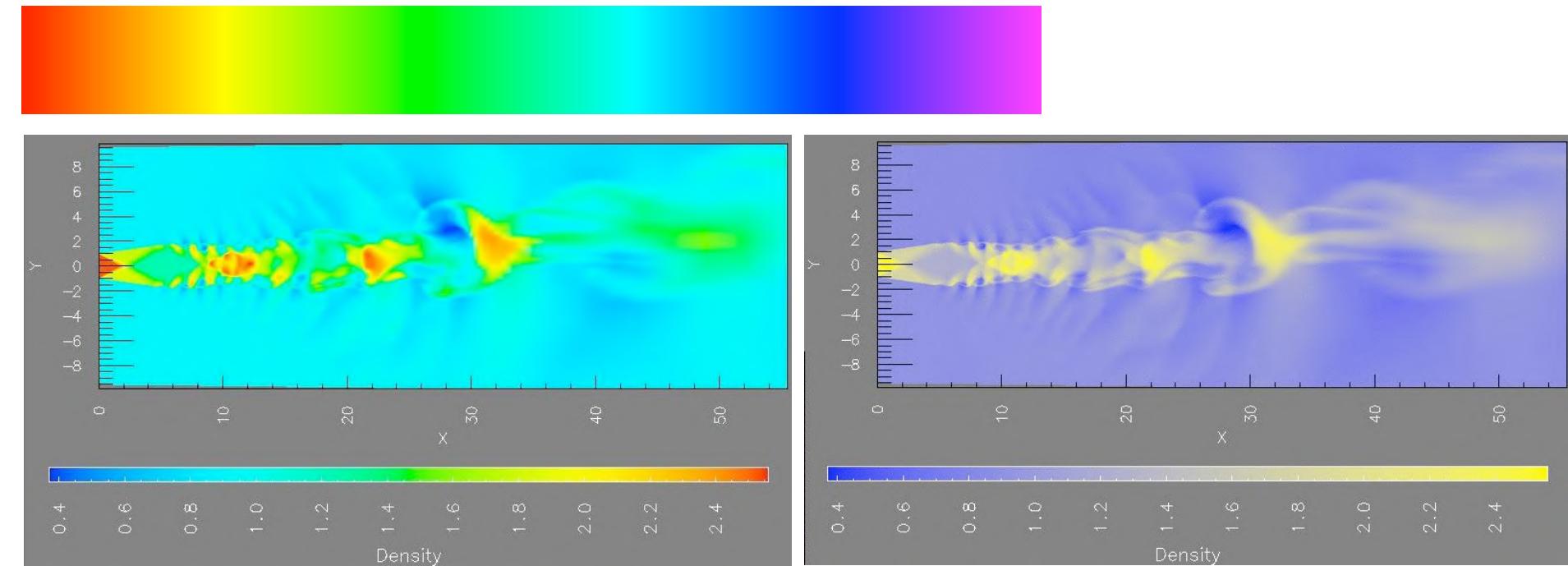


[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. <http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]

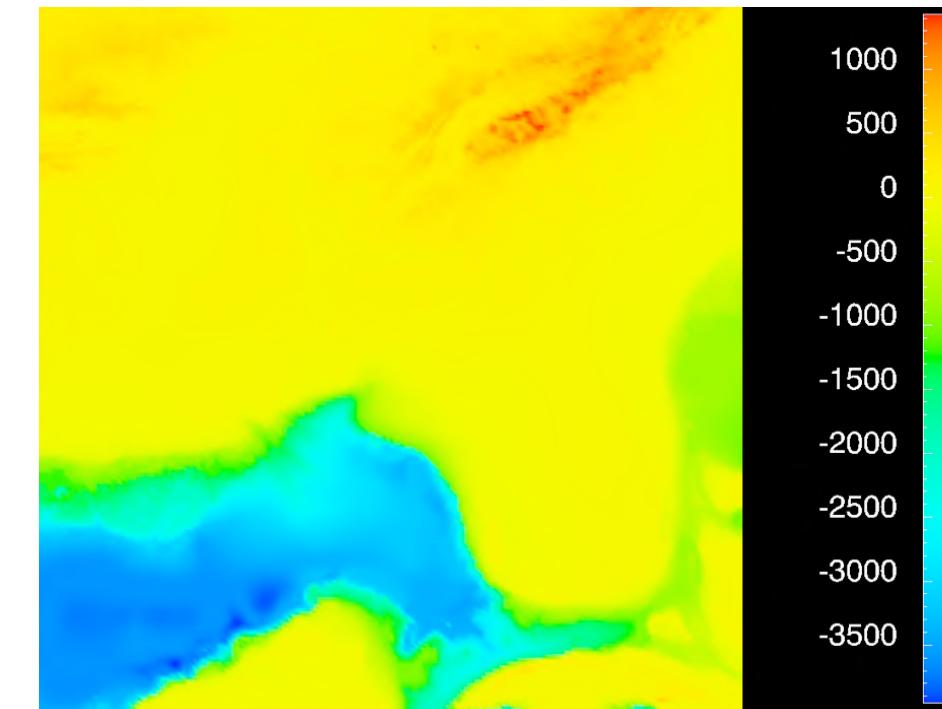
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

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.]

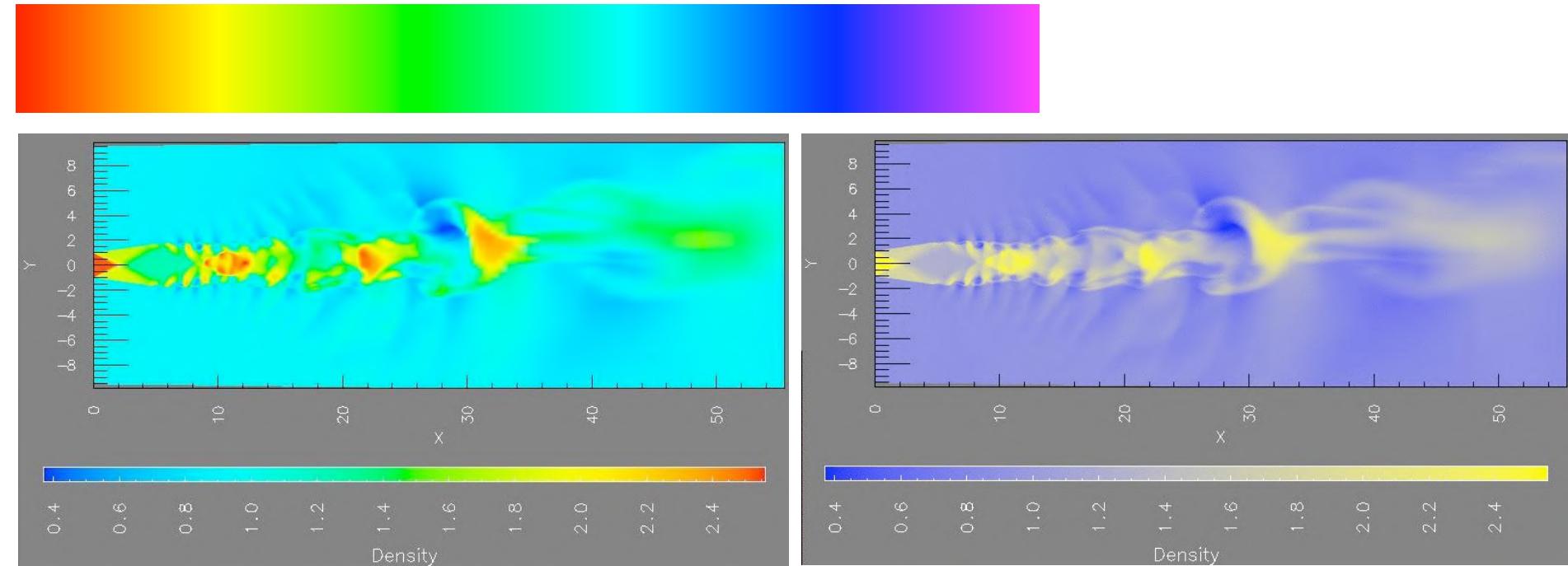


[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. <http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]

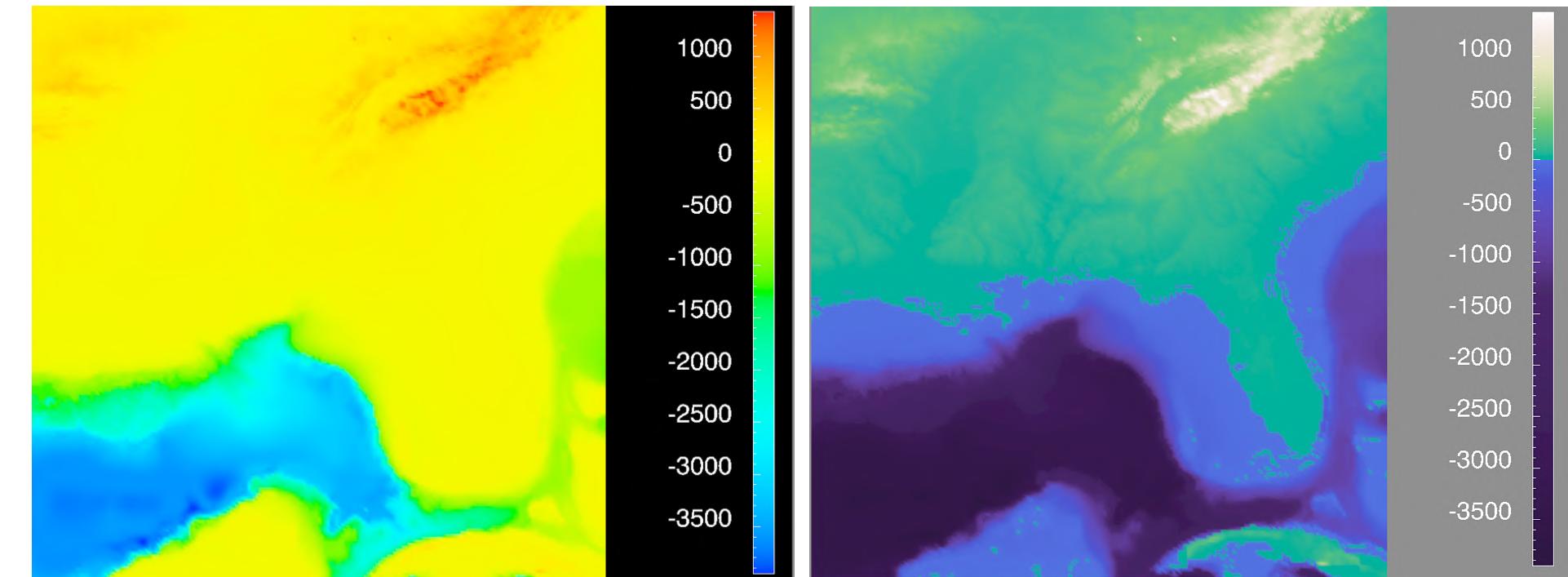
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

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.]

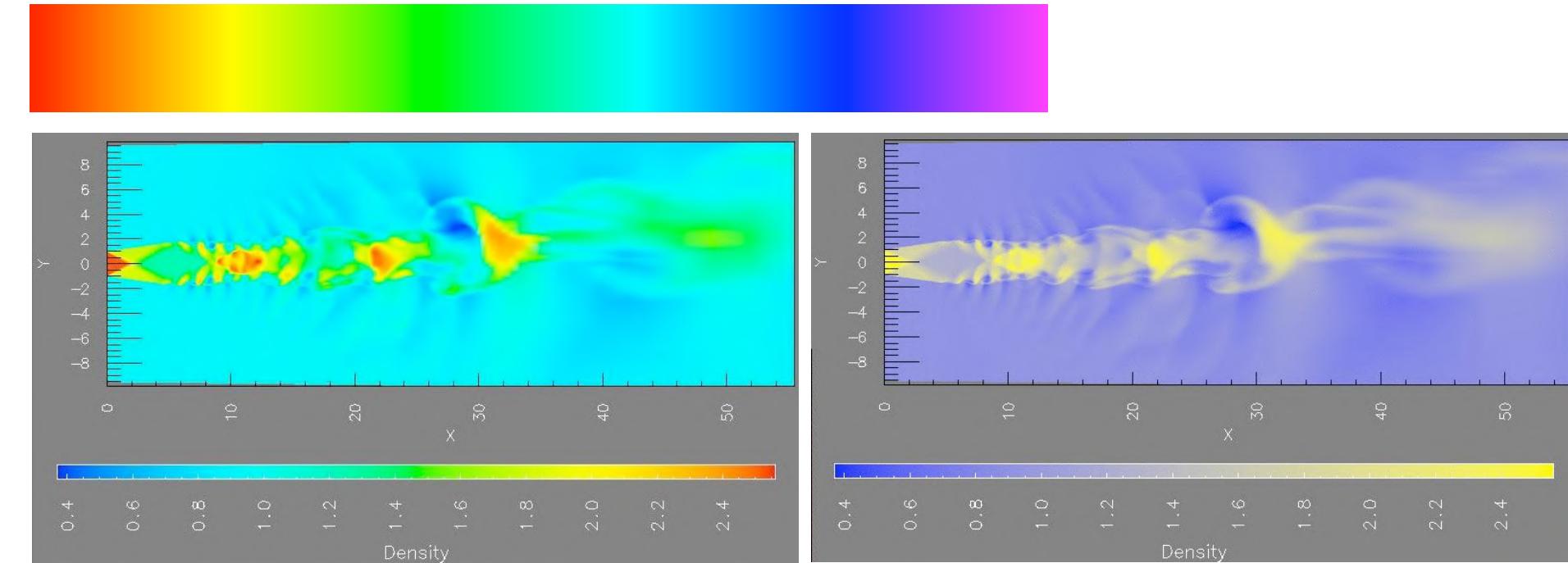


[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. <http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]

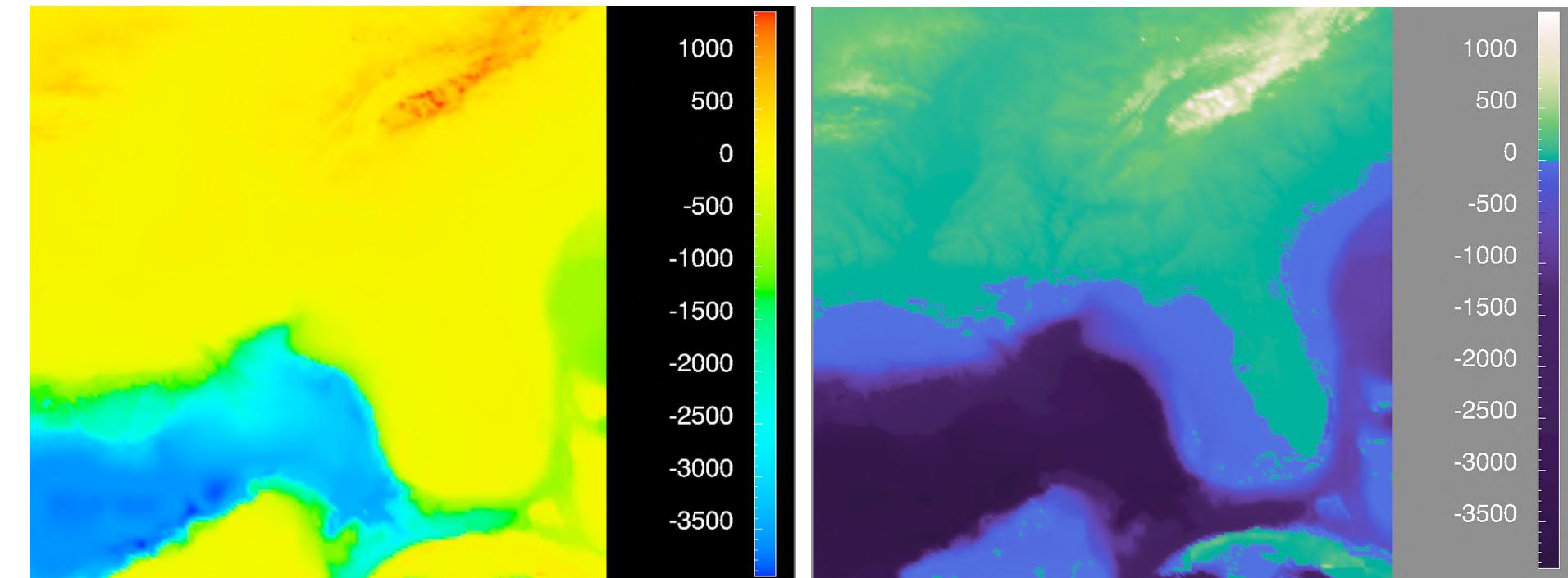
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

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]
 - segmented rainbows for binned

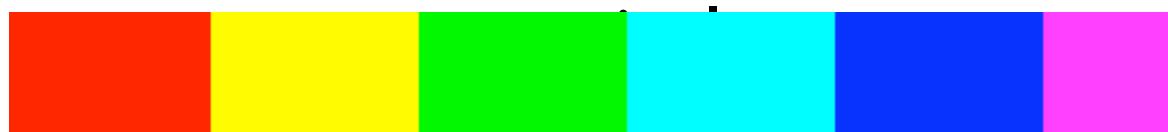


[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



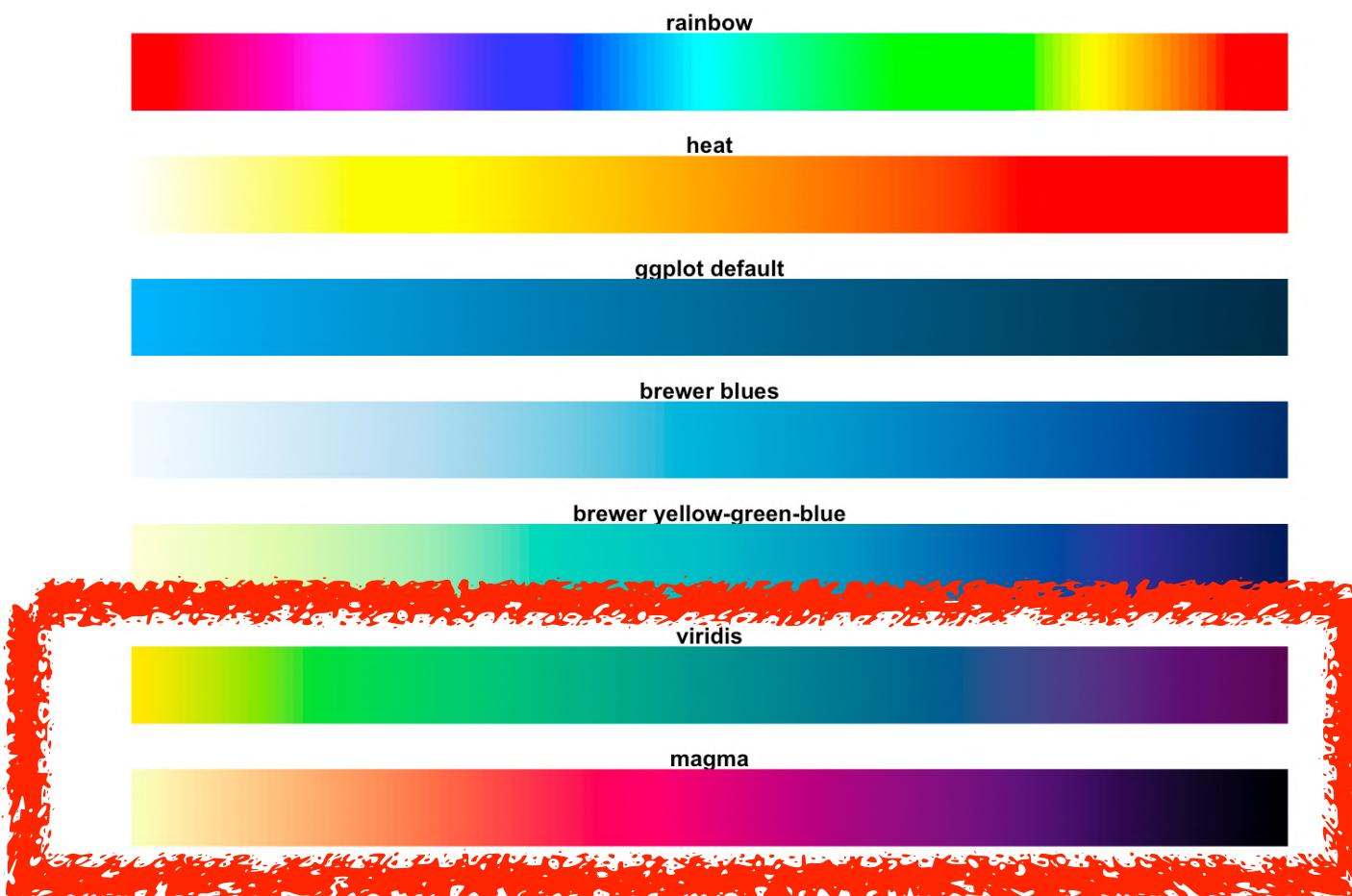
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. <http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]

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

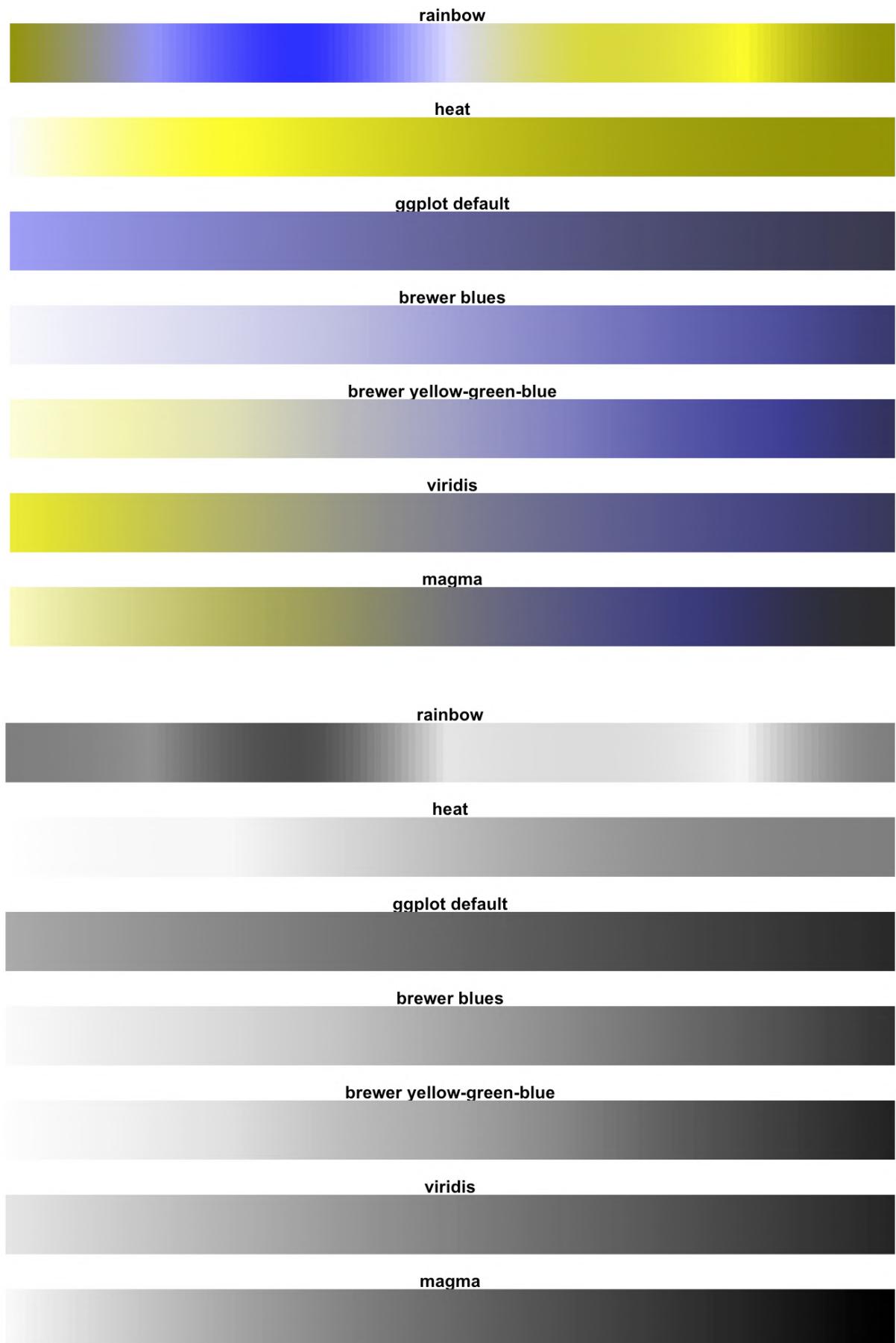


Viridis

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



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



Map other channels

- size
 - length accurate, 2D area ok, 3D volume poor
- angle
 - nonlinear accuracy
 - horizontal, vertical, exact diagonal
- shape
 - complex combination of lower-level primitives
 - many bins
- motion
 - highly separable against static
 - binary: great for highlighting
 - use with care to avoid irritation

→ Size, Angle, Curvature, ...

→ Length



→ Angle



→ Area



→ Curvature



→ Volume



→ Shape

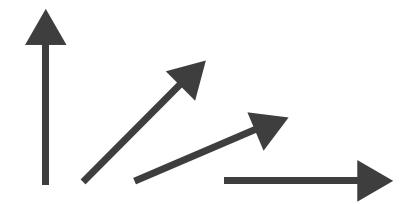
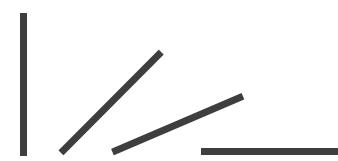


→ Motion

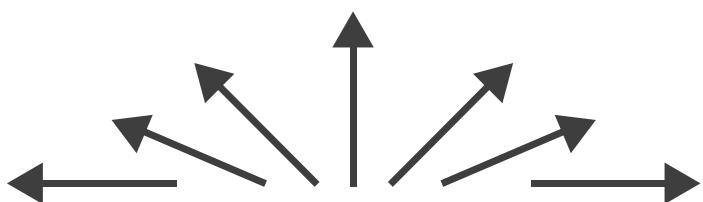
→ Motion
Direction, Rate,
Frequency, ...



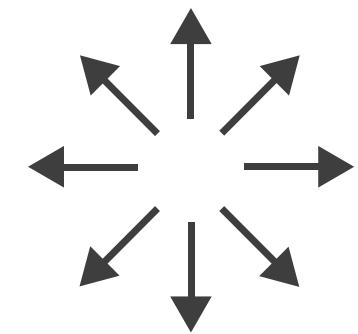
Angle



Sequential ordered
line mark or arrow glyph



Diverging ordered
arrow glyph



Cyclic ordered
arrow glyph

Homework I:
take a dataset
and create as
many visualisations
as you can.