Lecture 1:
Intro, Data and Tasks
Marks and Channels

Tamara Munzner
Department of Computer Science
University of British Columbia

DSCI 531: Data Visualization 1
Lecture 1: 16 November 2016

https://github.ubc.ca/ubc-mds-2016/DSCI_531_viz-1_students

What's when

- 8 lectures in 4 weeks
  - Wed & Mon, 11am-12:20pm (80 min), Nov 16 - Dec 12, SPPH 143
- 4 labs
  - Thu, 2-4pm, Nov 17 - Dec 8, ESB 1042
  - start work Thu 2pm, due next Wed 9am, 12.5% each
- 2 quizzes: Week 3 (Dec 1) & week 5 (Dec 15)
  - Thu 2-2:30pm, 25% each
- my (optional) office hrs are in ICICS/CS X661
  - Thu Nov 17, 5-6pm
  - Thu Nov 24, 5-6pm
  - Thu Dec 1, 5-6pm
  - Wed Dec 7, 6-7pm (note outer building doors close at 6:30)

Reading

- core foundational material covered in lectures

- textbook as backup to lectures
  - Tamara Munzner. Visualization Analysis and Design. CRC Press, 2014.
    - library has multiple ebook copies for free
    - to buy yourself, see http://www.cs.ubc.ca/~tmm/vadbook/

Topics

- Lecture 1
  - Intro, Data and Tasks
  - Marks and Channels
- Lecture 2
  - In-Class Vis Design Exercise
- Lecture 3
  - Arrange Table Data, part I
- Lecture 4
  - Arrange Table Data, part I
- Lecture 5
  - Arrange Spatial Data
- Lecture 6
  - Color
- Lecture 7
  - Arrange Network Data
- Lecture 8
  - Rules of Thumb
  - Graphic Design Principles

Introduction: Defining visualization (vis)

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

Why?...

Why have a human in the loop?

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

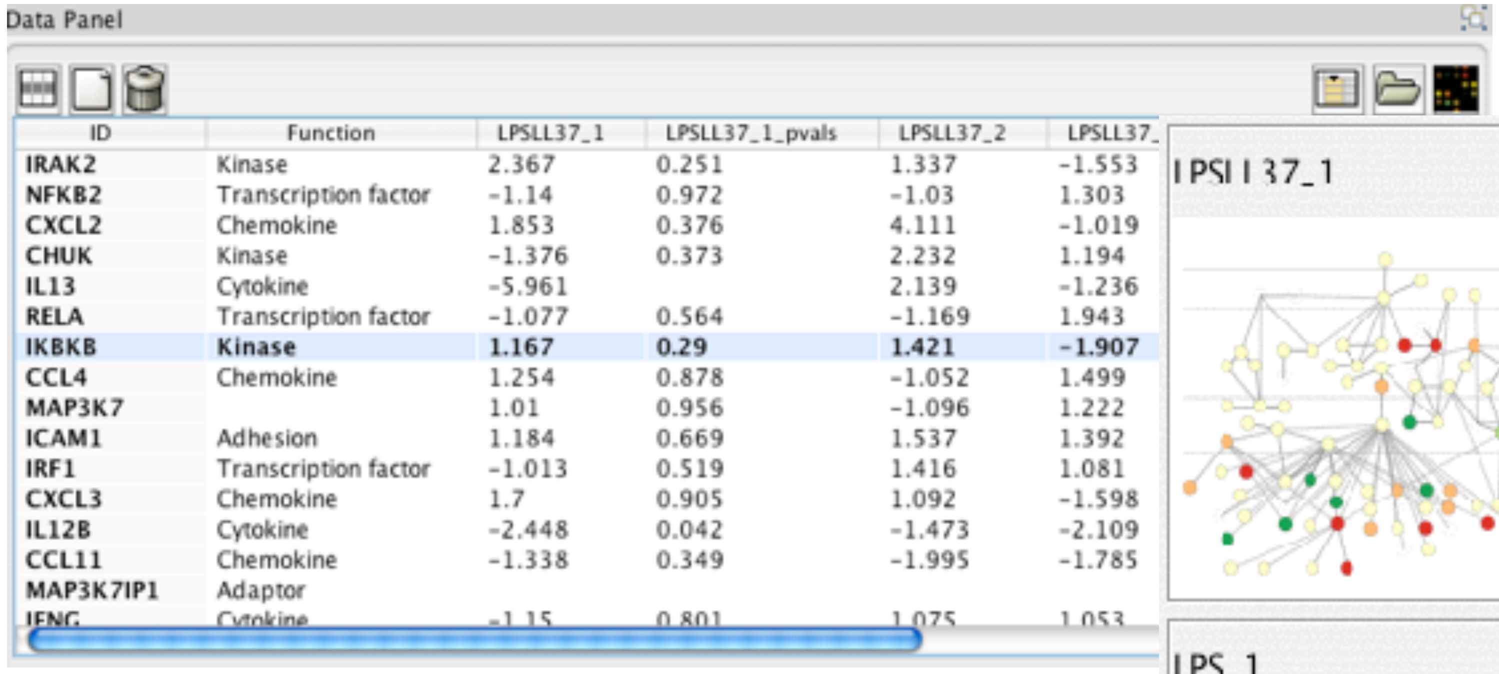
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- don't need vis when fully automatic solution exists and is trusted
- many analysis problems ill-specified
  - don't know exactly what questions to ask in advance
- possibilities
  - long-term use for end users (e.g. exploratory analysis of scientific data)
  - *presentation of known results*
  - stepping stone to better understanding of requirements before developing models
  - help developers of automatic solution refine/debug, determine parameters
  - help end users of automatic solutions verify, build trust

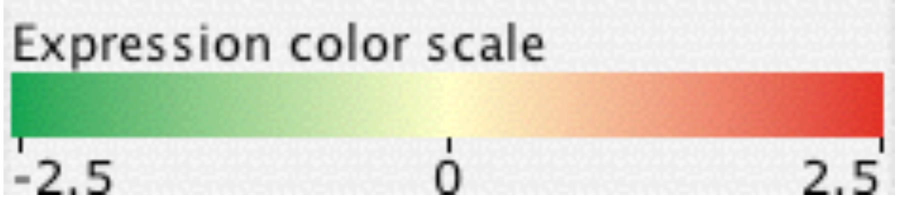
Why use an external representation?

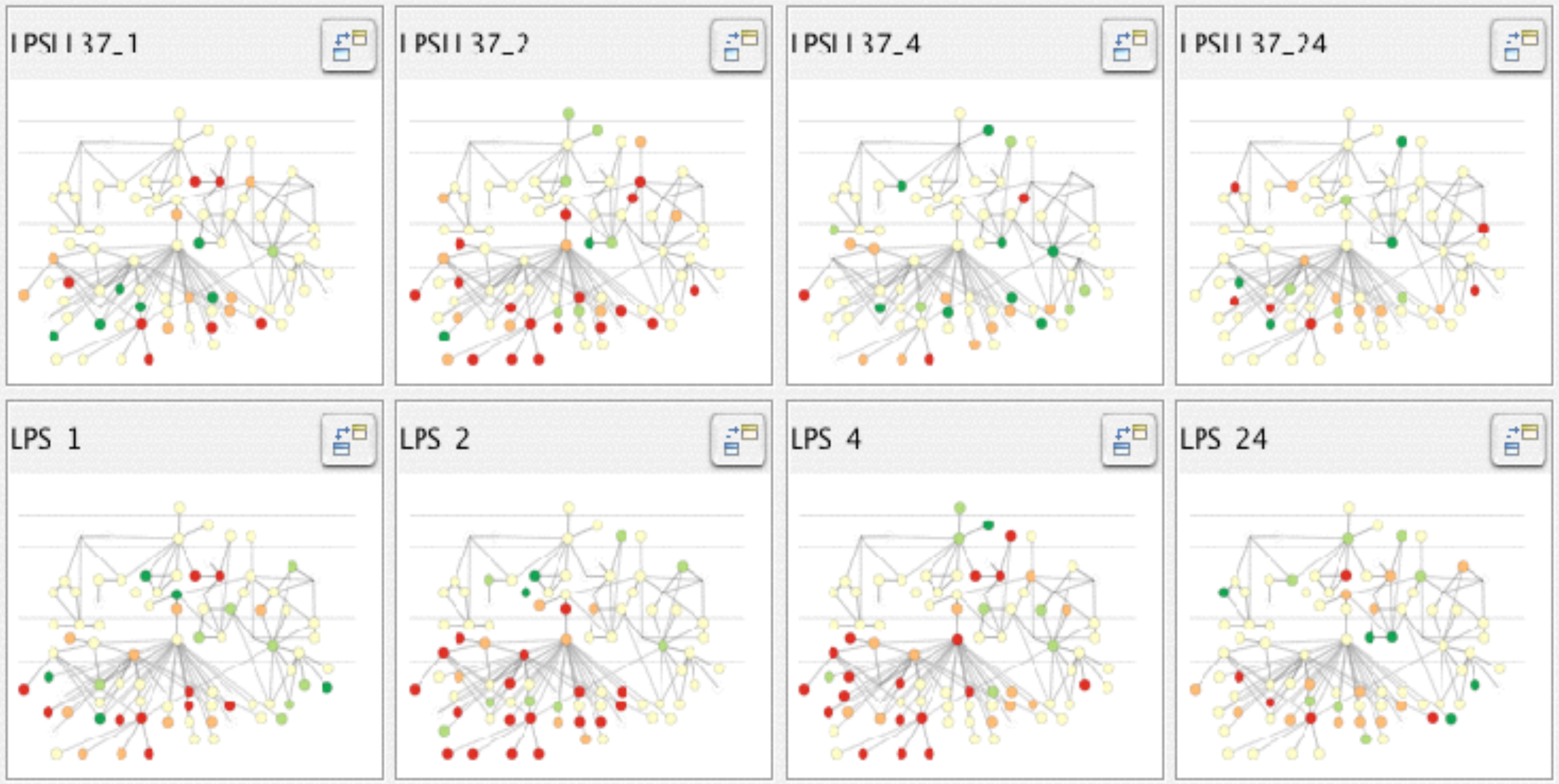
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- external representation: replace cognition with perception



ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_
IRAK2	Kinase	2.367	0.251	1.337	-1.553
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303
CXCL2	Chemokine	1.853	0.376	4.111	-1.019
CHUK	Kinase	-1.376	0.373	2.232	1.194
IL13	Cytokine	-5.961		2.139	-1.236
RELA	Transcription factor	-1.077	0.564	-1.169	1.943
IKBKB	Kinase	1.167	0.29	1.421	-1.907
CCL4	Chemokine	1.254	0.878	-1.052	1.499
MAP3K7		1.01	0.956	-1.096	1.222
ICAM1	Adhesion	1.184	0.669	1.537	1.392
IRF1	Transcription factor	-1.013	0.519	1.416	1.081
CXCL3	Chemokine	1.7	0.905	1.092	-1.598
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785
MAP3K7IP1	Adaptor				
IFNG	Cytokine	-1.15	0.801	1.075	1.053





[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why depend on vision?

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

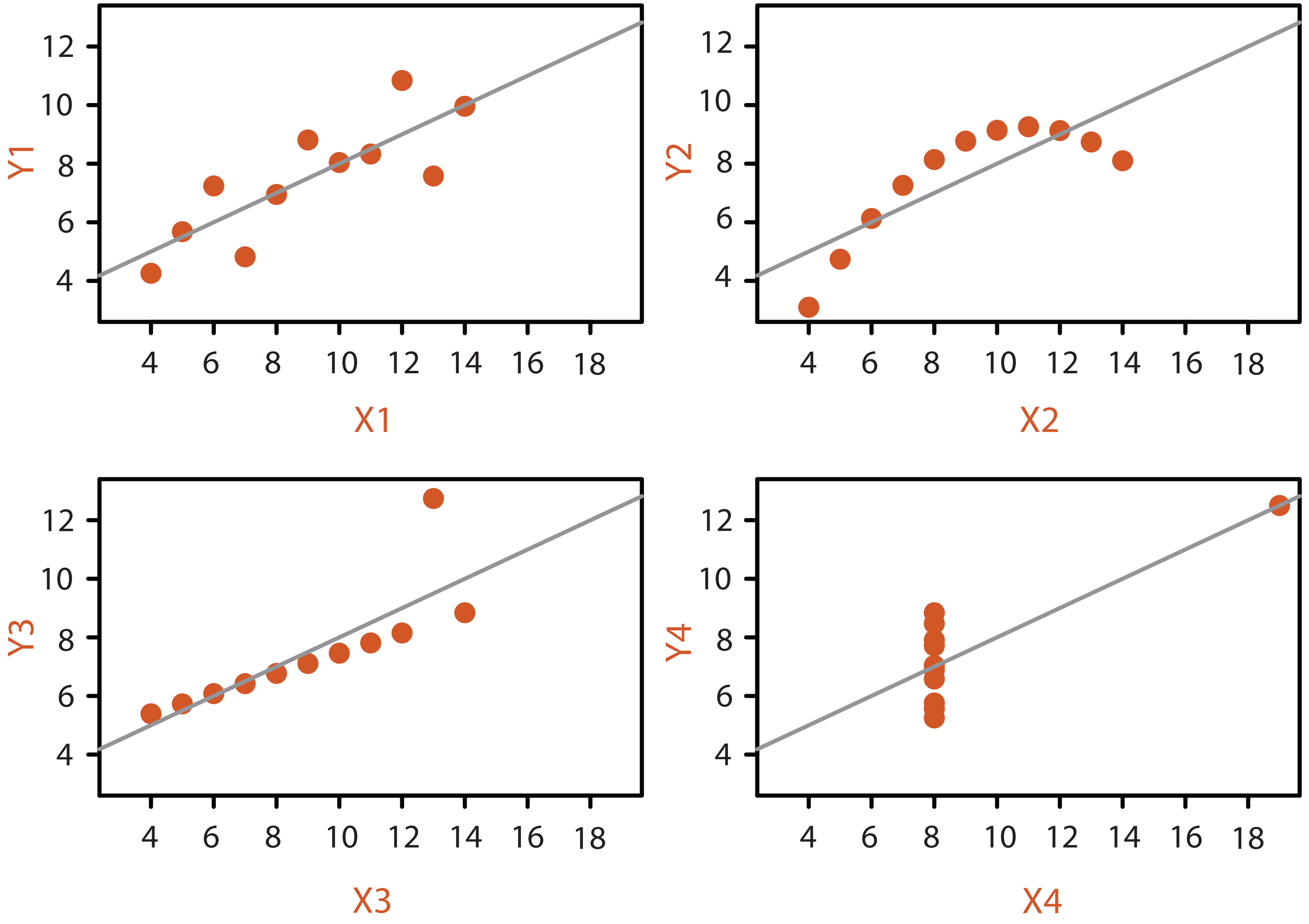
- human visual system is high-bandwidth channel to brain
  - overview possible due to background processing
    - subjective experience of seeing everything simultaneously
    - significant processing occurs in parallel and pre-attentively
- sound: lower bandwidth and different semantics
  - overview not supported
    - subjective experience of sequential stream
- touch/haptics: impoverished record/replay capacity
  - only very low-bandwidth communication thus far
- taste, smell: no viable record/replay devices

Why show the data in detail?

- summaries lose information
  - 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



Why focus on tasks and effectiveness?

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

- tasks serve as constraint on design (as does data)
  - idioms do not serve all tasks equally!
  - challenge: recast tasks from domain-specific vocabulary to abstract forms
- most possibilities ineffective
  - validation is necessary, but tricky
  - increases chance of finding good solutions if you understand full space of possibilities
- what counts as effective?
  - novel: enable entirely new kinds of analysis
  - faster: speed up existing workflows

What resource limitations are we faced with?

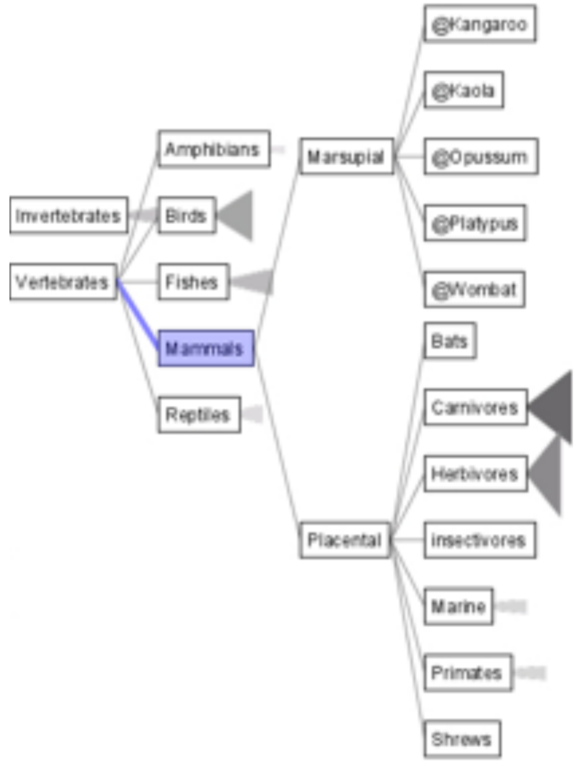
Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

- computational limits
  - processing time
  - system memory
- human limits
  - human attention and memory
- display limits
  - pixels are precious resource, the most constrained resource
  - **information density**: ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space, find sweet spot between dense and sparse

Why analyze?

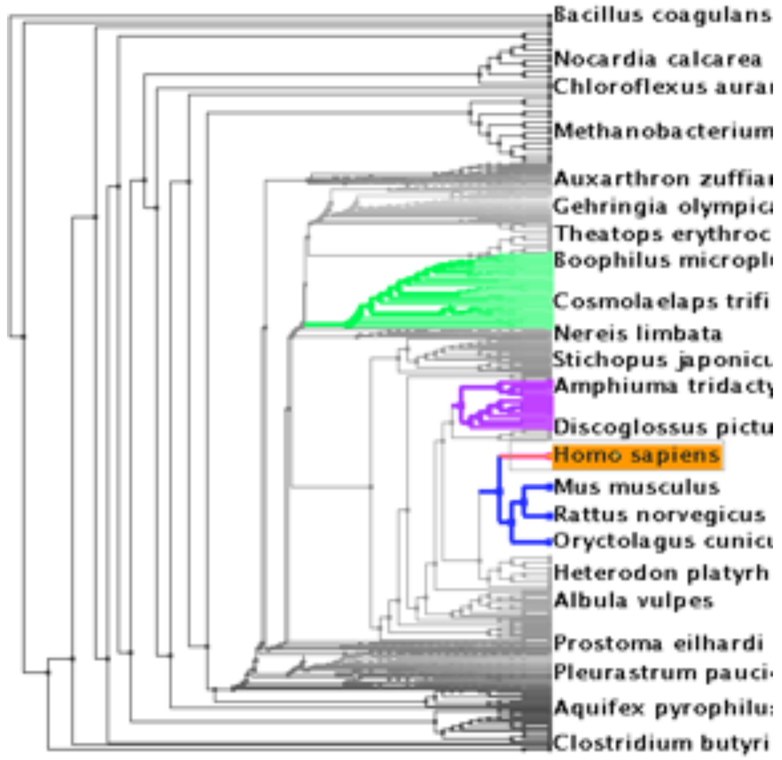
- imposes structure on huge design space
  - scaffold to help you think systematically about choices
  - analyzing existing as stepping stone to designing new
  - most possibilities ineffective for particular task/data combination

SpaceTree

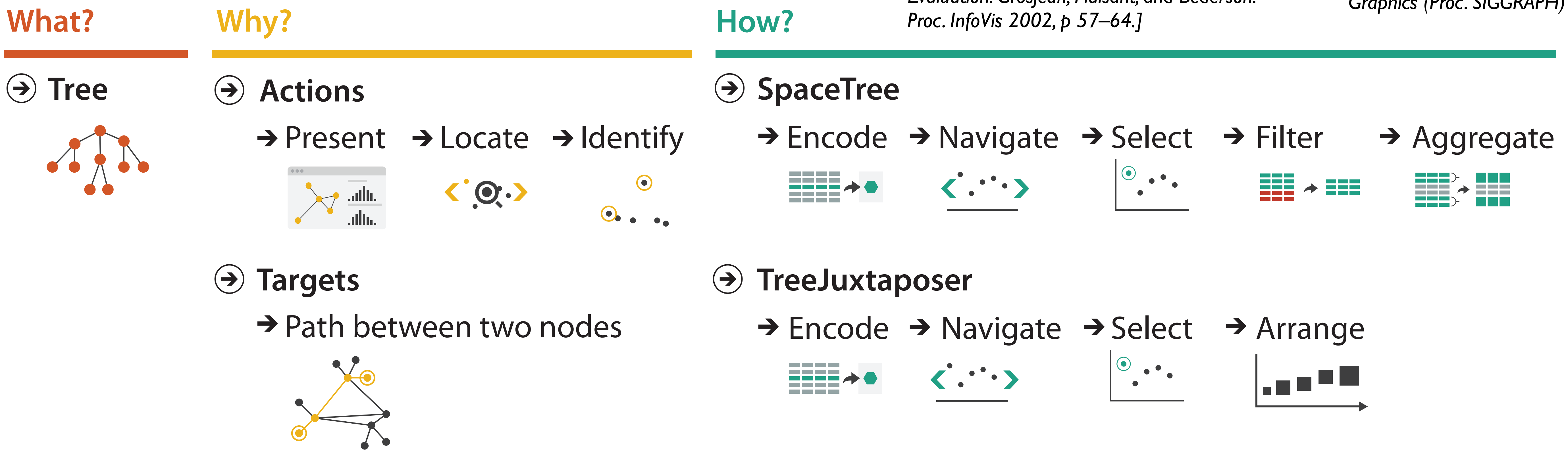


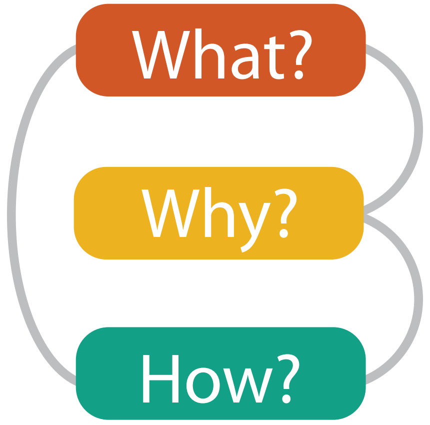
[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57–64.]

TreeJuxtaposer

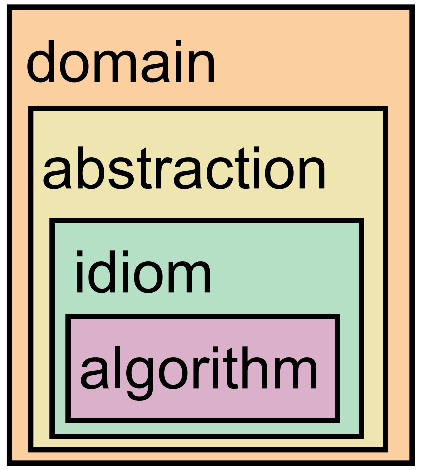


[TreeJuxtaposer: Scalable Tree Comparison Using Focus +Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]



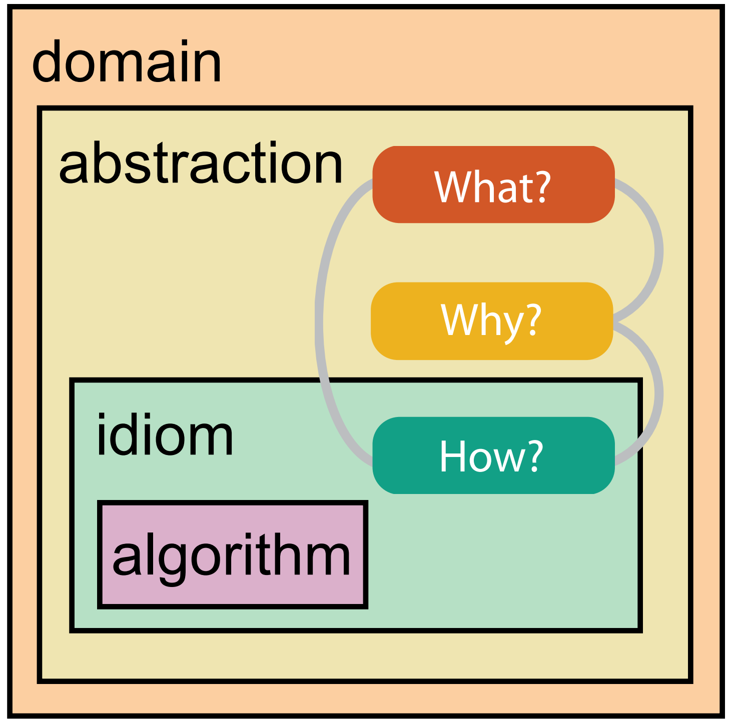


Analysis framework: Four levels, three questions



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

- *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



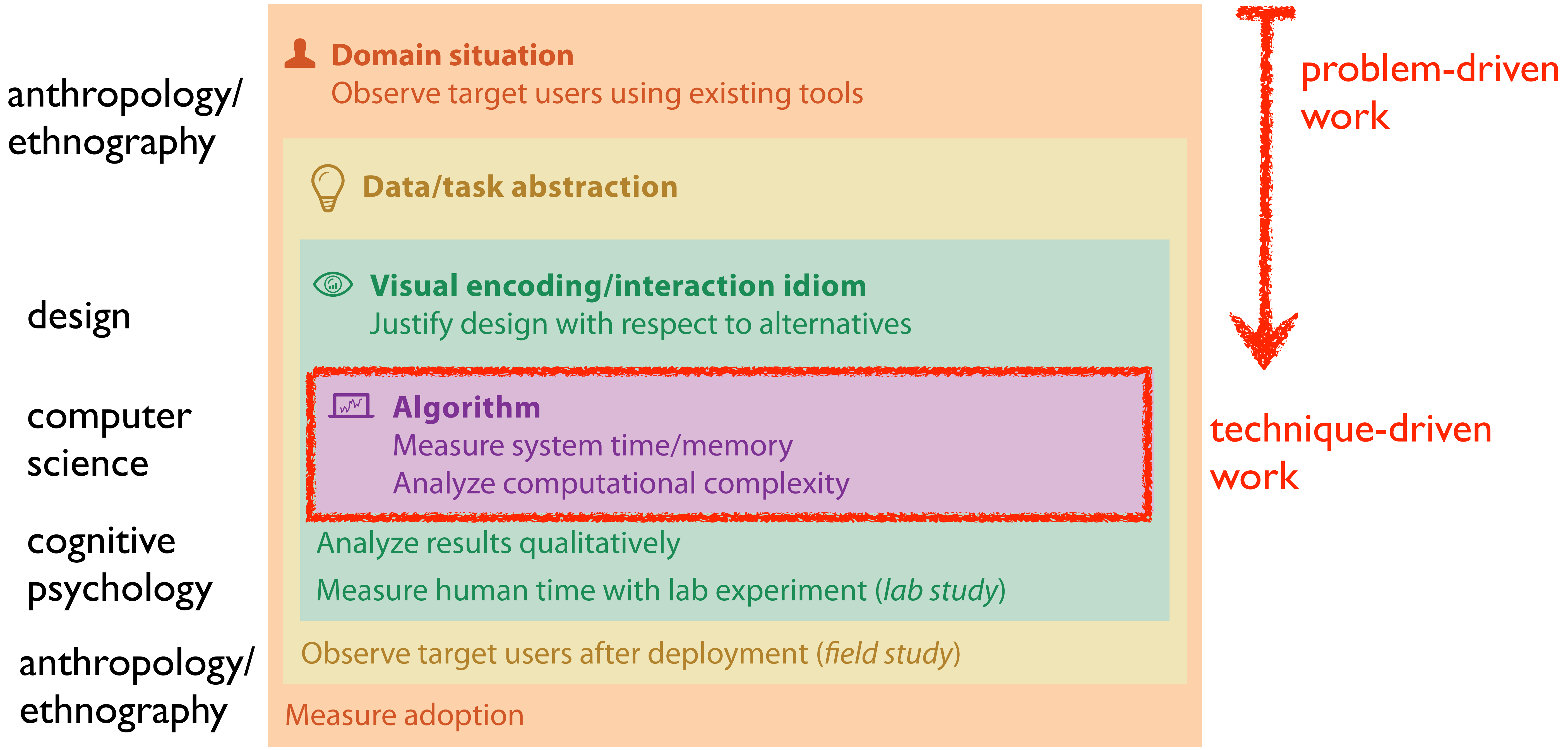
[A Multi-Level Typology of Abstract Visualization Tasks
Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Why is validation difficult?

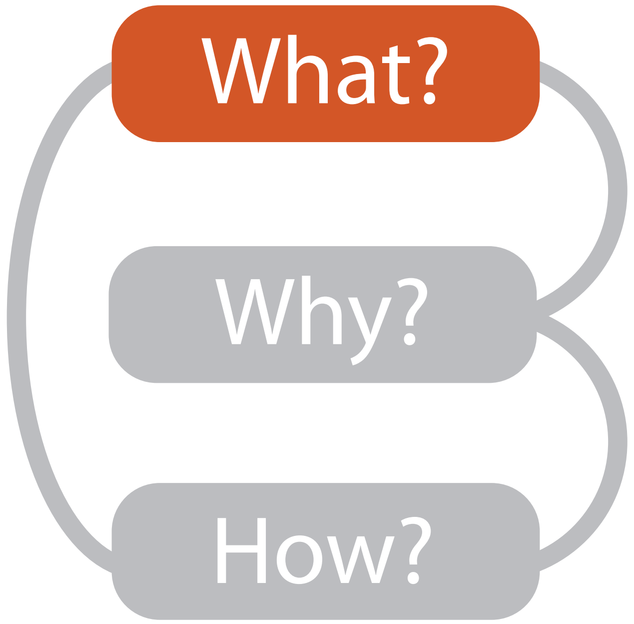
- different ways to get it wrong at each level

Why is validation difficult?

- solution: use methods from different fields at each level



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



What?

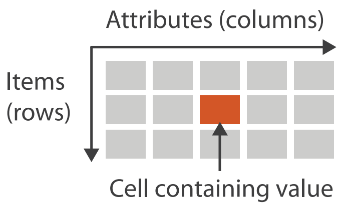
Datasets

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

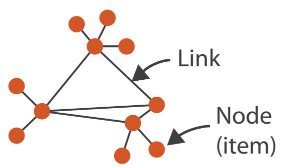
Data and Dataset Types

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

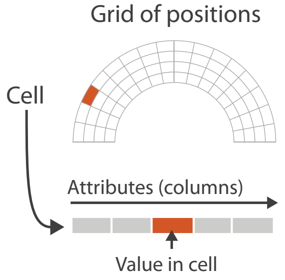
Dataset Types
- Tables



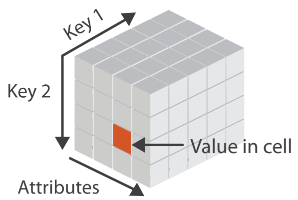
- Networks



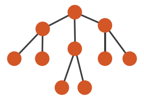
- Fields (Continuous)



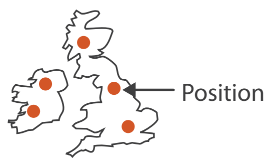
- *Multidimensional Table*



- *Trees*



- Geometry (Spatial)



Dataset Availability
- Static

- Dynamic

Attributes

Attribute Types
- Categorical

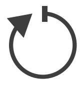
- Ordered
  - *Ordinal*

  - *Quantitative*

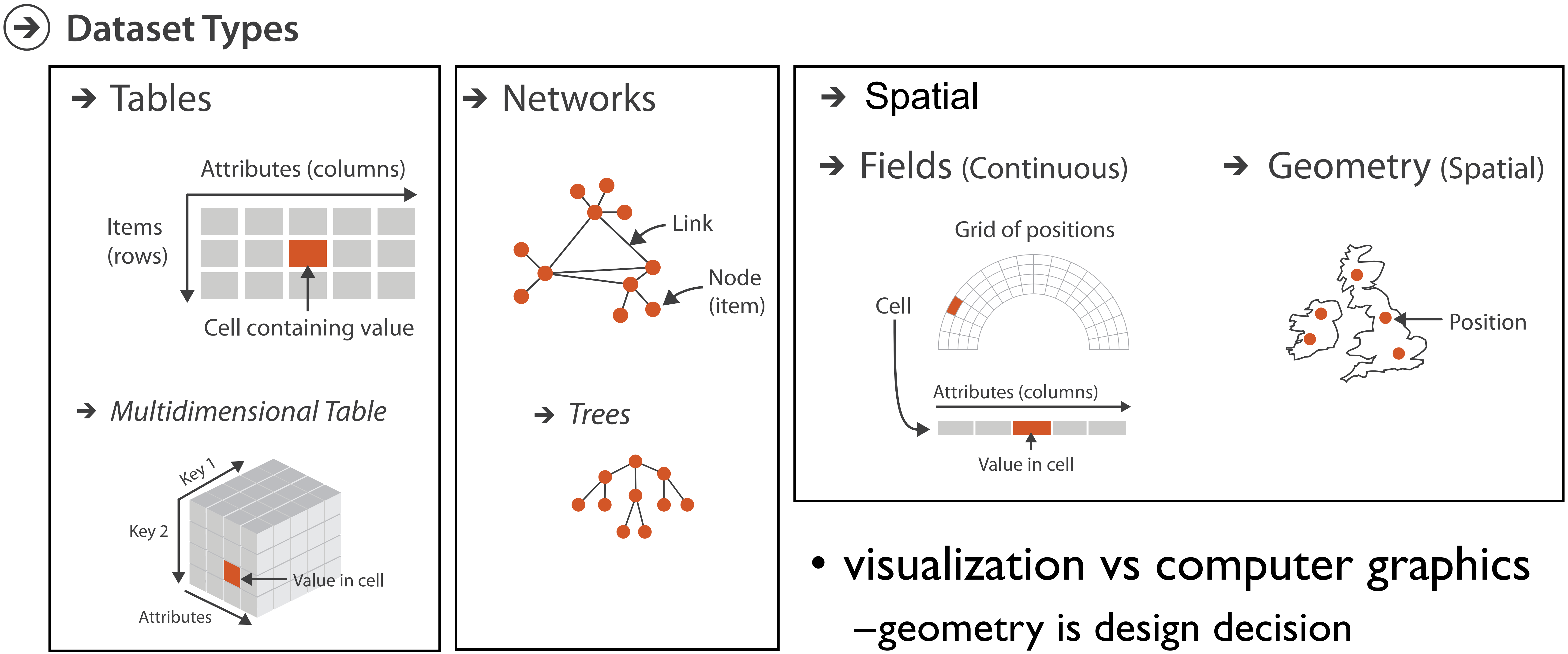
Ordering Direction
- Sequential

- Diverging

- Cyclic



Three major datatypes



- 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	
	Attributes	Attributes		

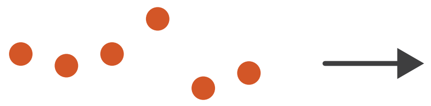
➔ Data Types

- ➔ Items
- ➔ Attributes
- ➔ Links
- ➔ Positions
- ➔ Grids

➔ Dataset Availability

- ➔ Static

- ➔ Dynamic

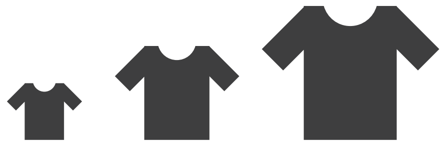


Attribute types

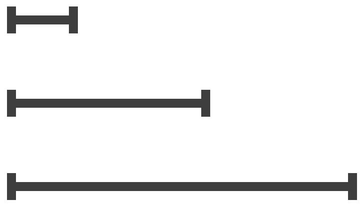
Attribute Types

- Categorical

- Ordered
  - *Ordinal*



  - *Quantitative*

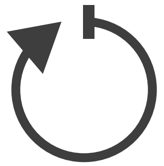


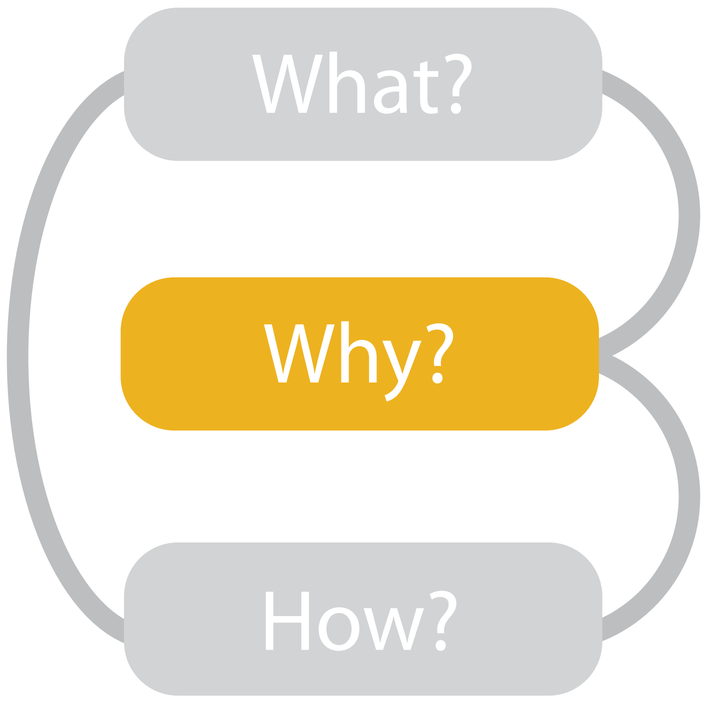
Ordering Direction

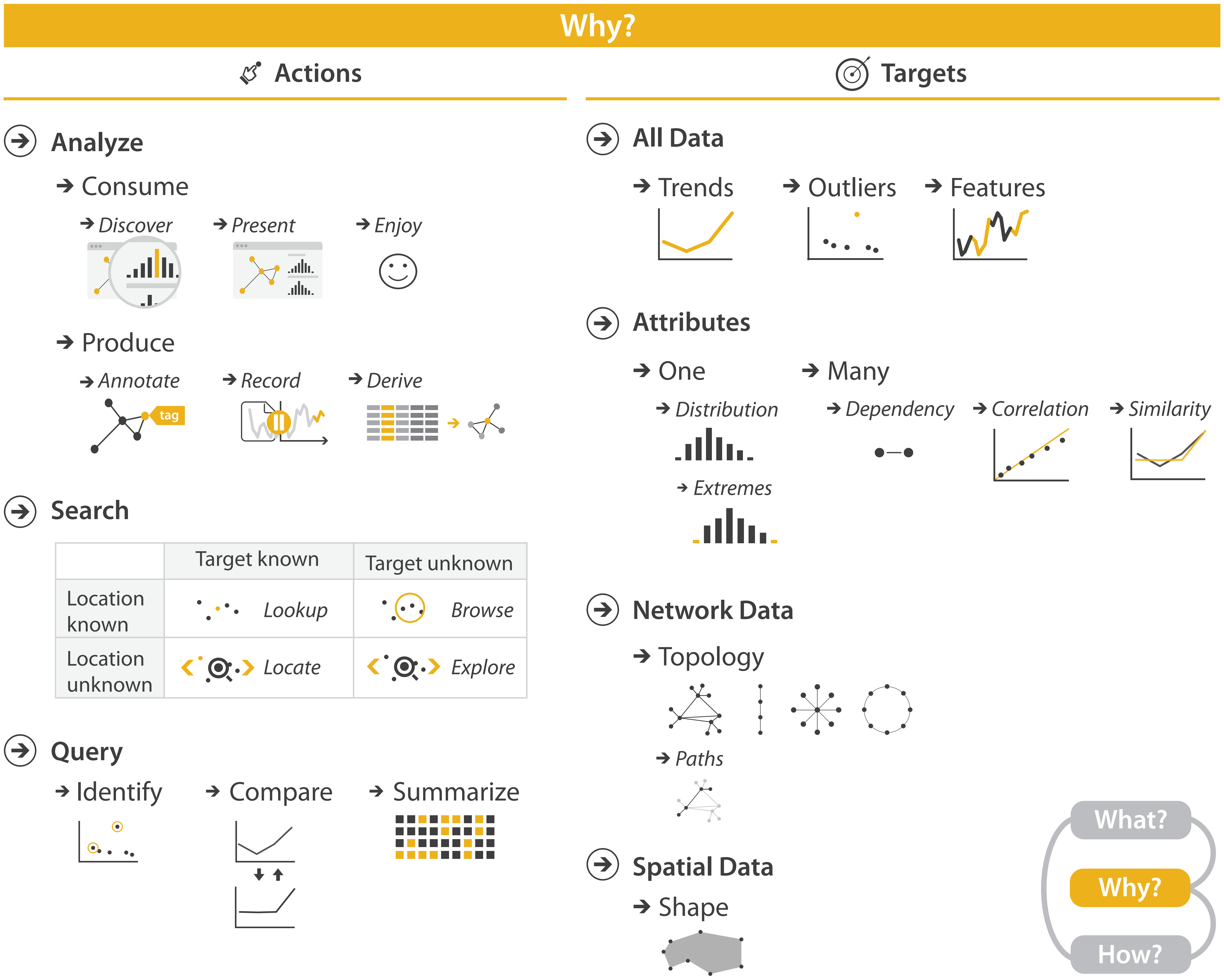
- Sequential

- Diverging

- Cyclic



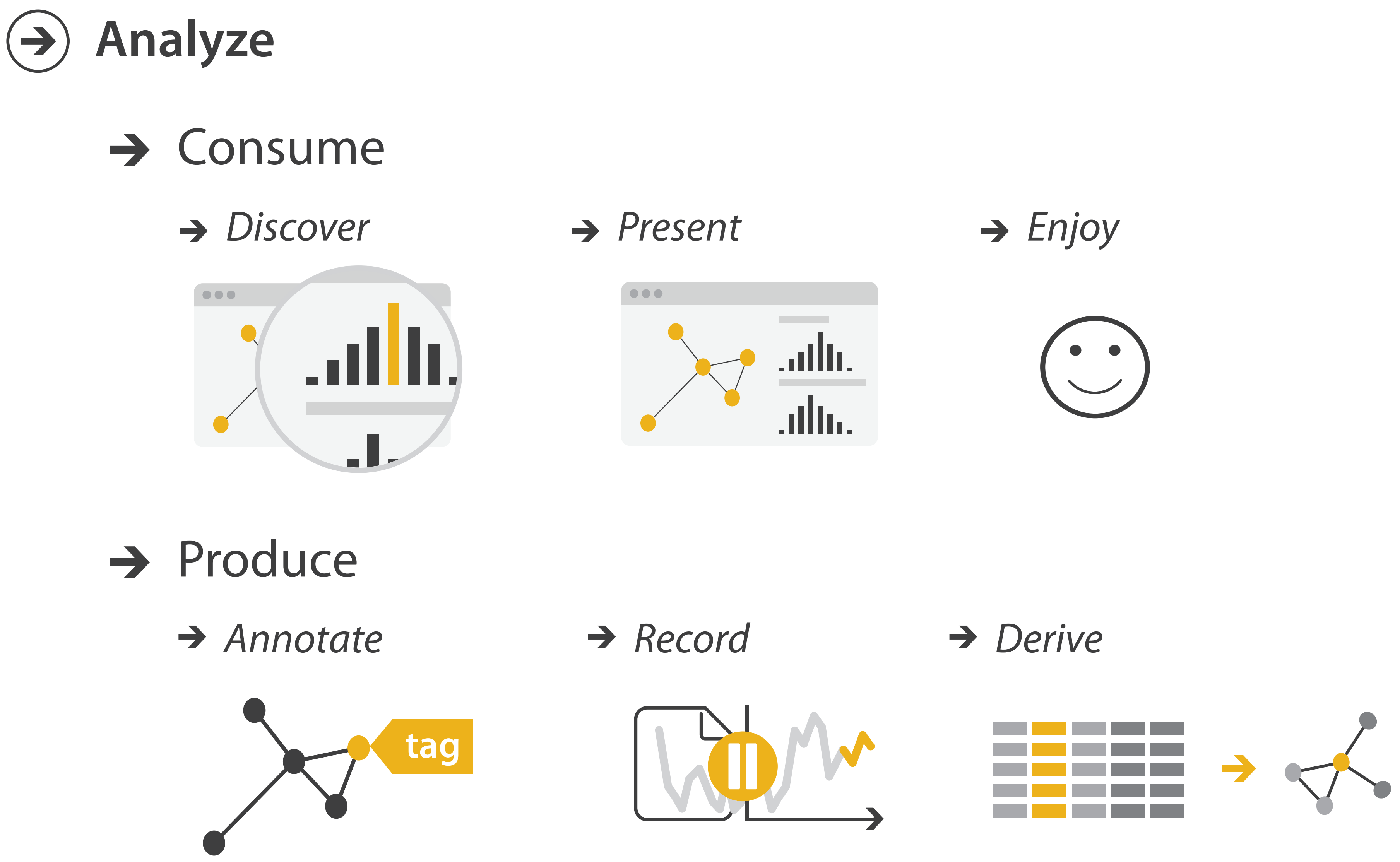




- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

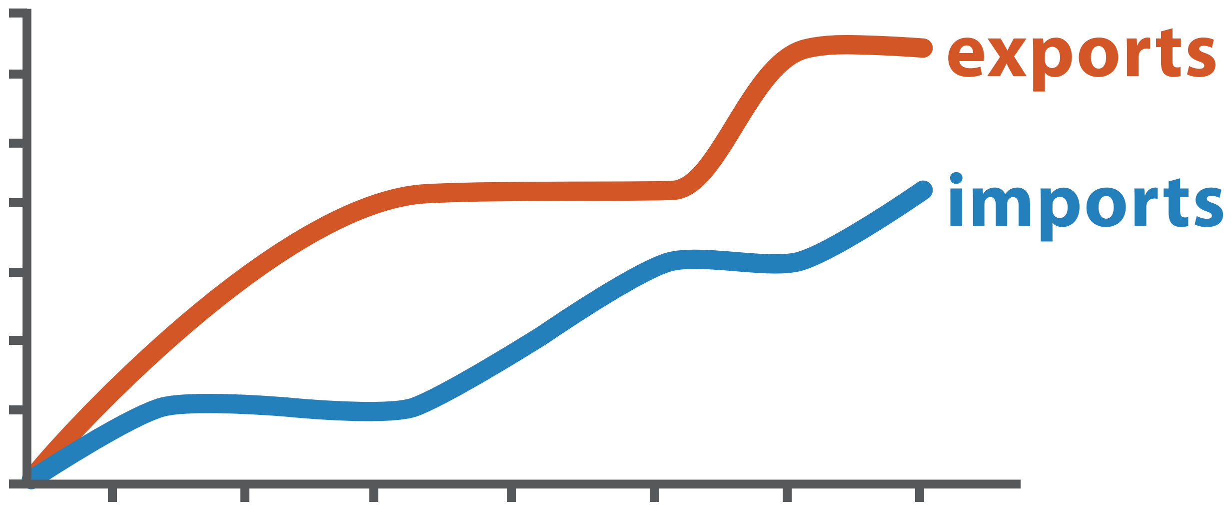
Actions: Analyze

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

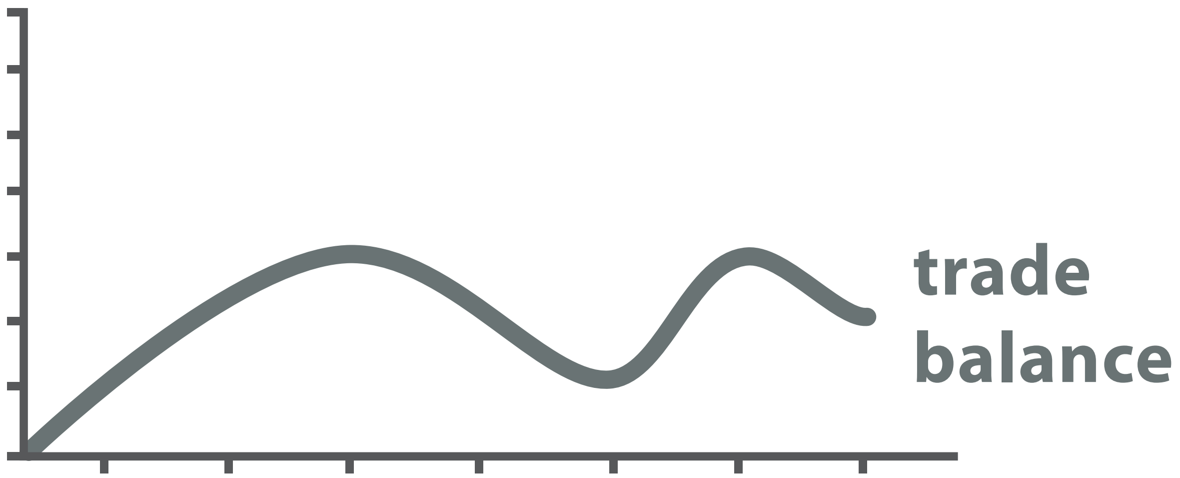


Derive

- 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



trade balance = exports − imports

Derived Data

Actions: Search, query

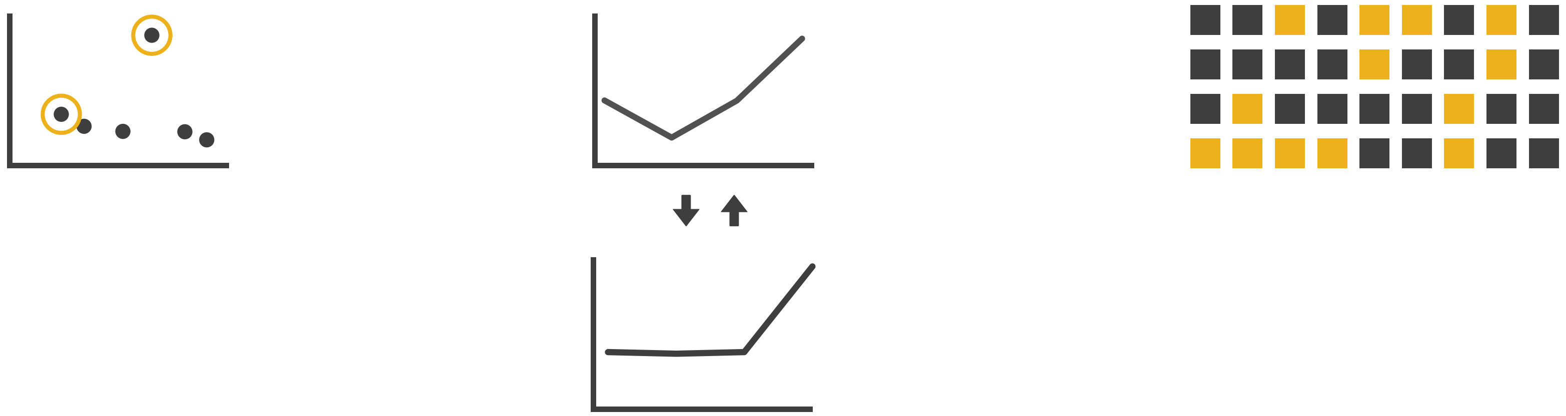
- what does user know?
  - –target, location
- how much of the data matters?
  - –one, some, all
- independent choices for each of these three levels
  - –analyze, search, query
  - –mix and match

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

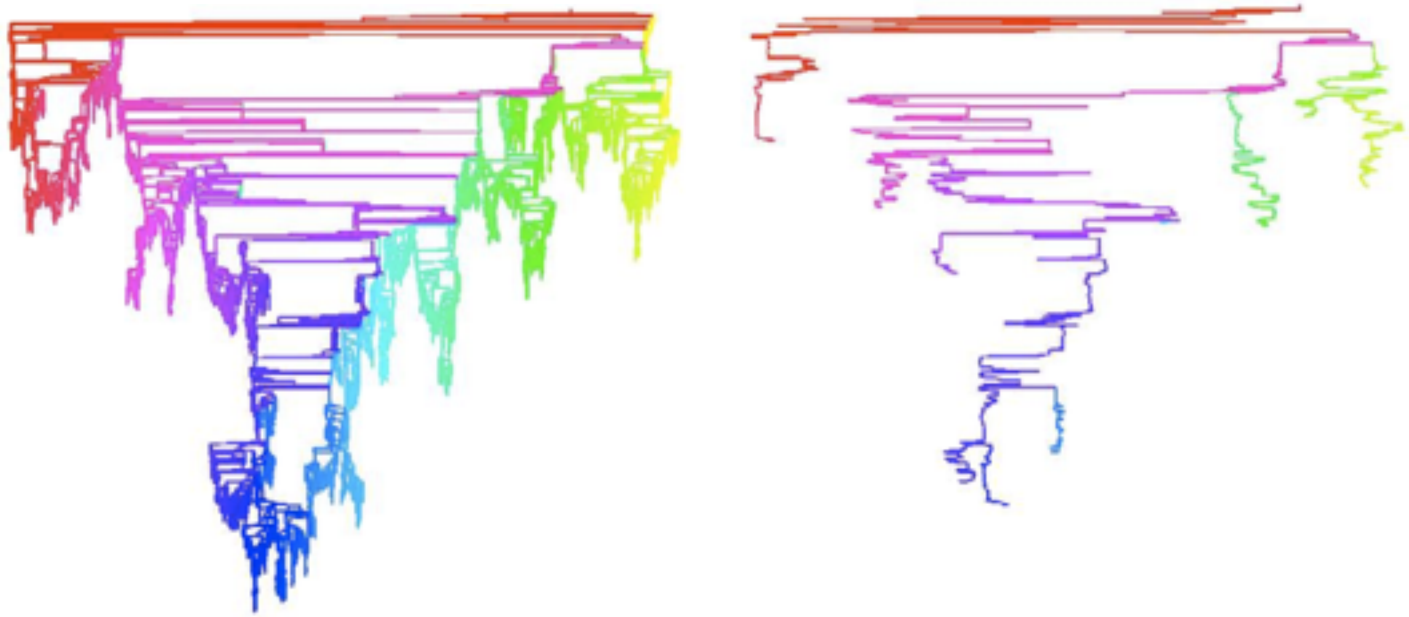
→ Identify → Compare → Summarize



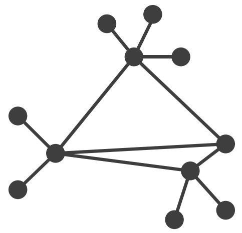
Analysis example: Derive one attribute

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

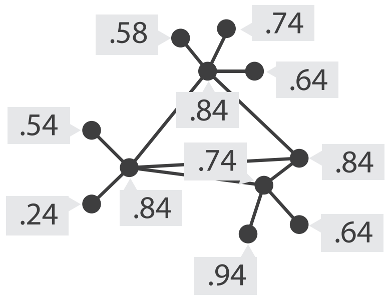
[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]



Task 1



In
Tree

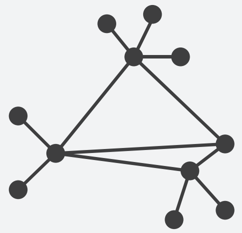


Out
Quantitative attribute on nodes

What?
- In Tree
- Out Quantitative attribute on nodes

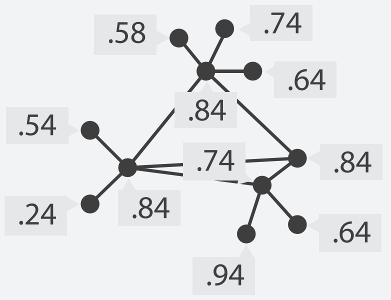
Why?
- Derive

Task 2

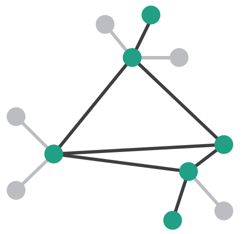


In
Tree

+



In
Quantitative attribute on nodes



Out
Filtered Tree
Removed unimportant parts

What?
- In Tree
- In Quantitative attribute on nodes
- Out Filtered Tree

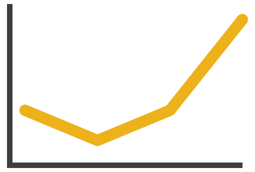
Why?
- Summarize
- Topology

How?
- Reduce
- Filter

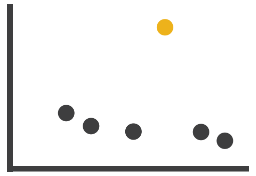
Why: Targets

➔ All Data

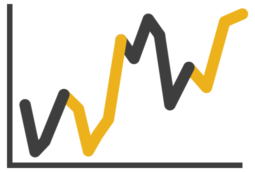
➔ Trends



➔ Outliers



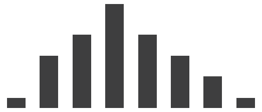
➔ Features



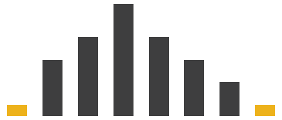
➔ Attributes

➔ One

➔ *Distribution*



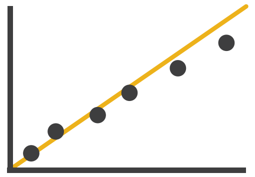
➔ *Extremes*



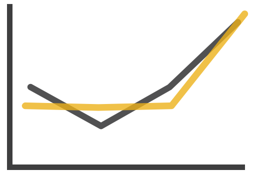
➔ Many

➔ *Dependency*

➔ *Correlation*

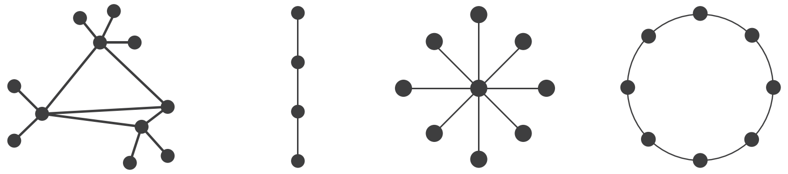


➔ *Similarity*

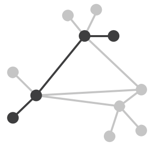


➔ Network Data

➔ Topology

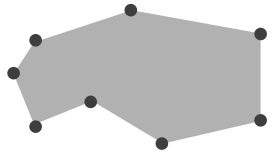


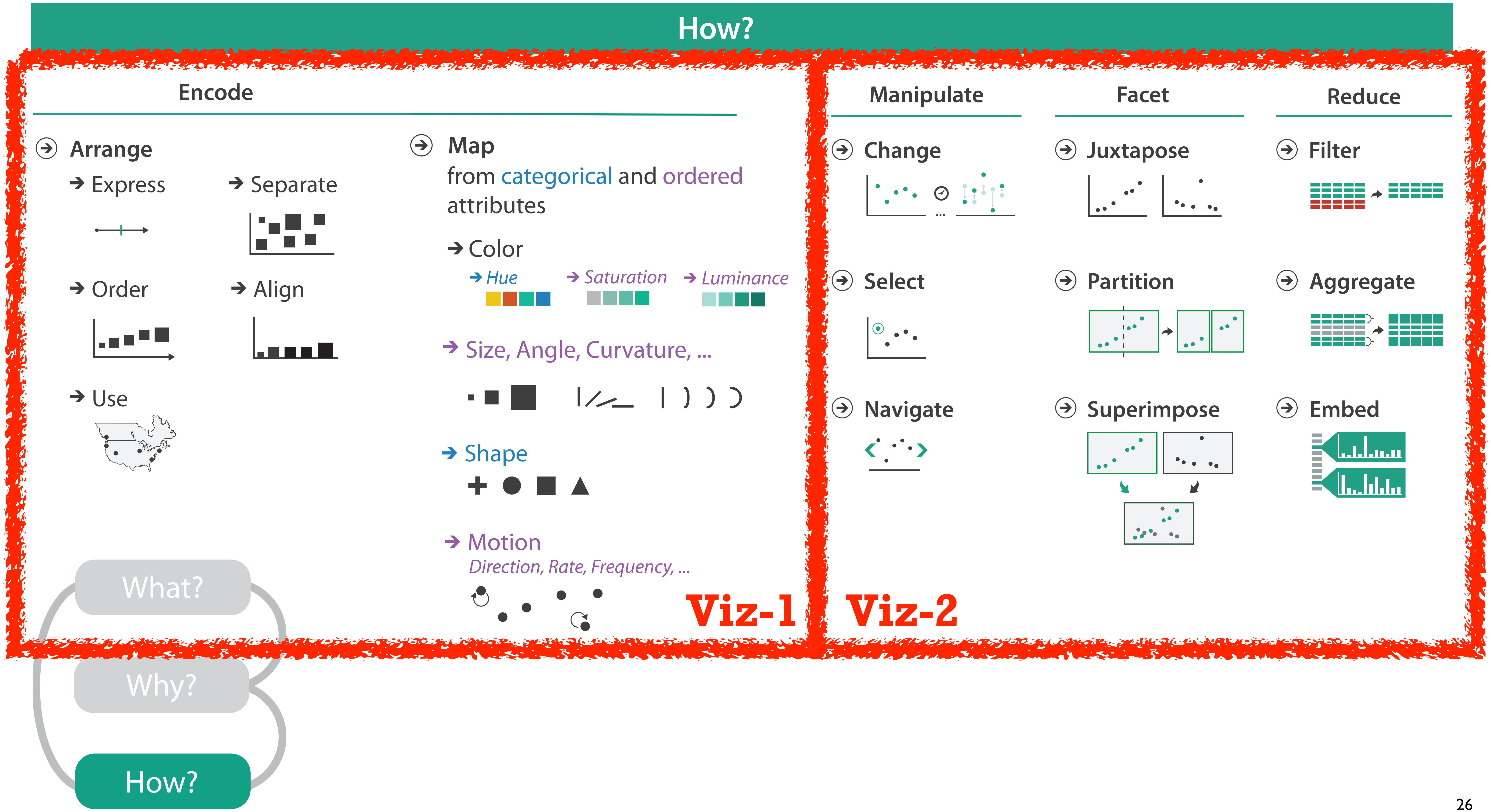
➔ *Paths*



➔ Spatial Data

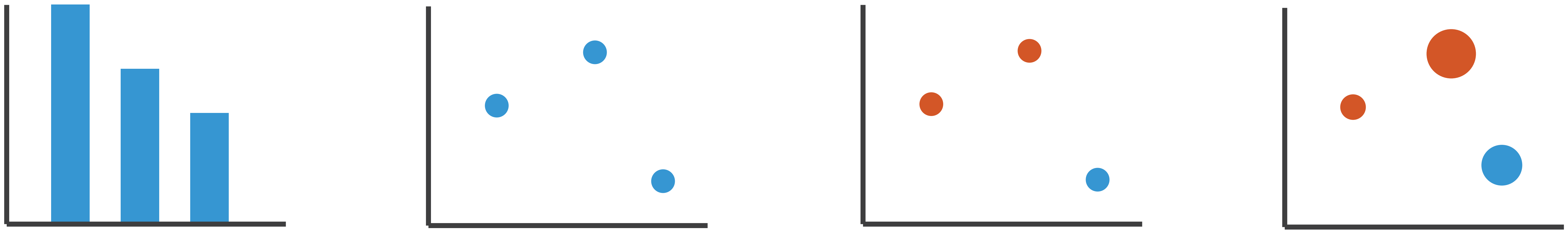
➔ Shape




How?
Encode
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
Viz-1
Manipulate
Change
Select
Navigate
Facet
Juxtapose
Partition
Superimpose
Reduce
Filter
Aggregate
Embed
Viz-2
What?
Why?
How?

Encoding visually

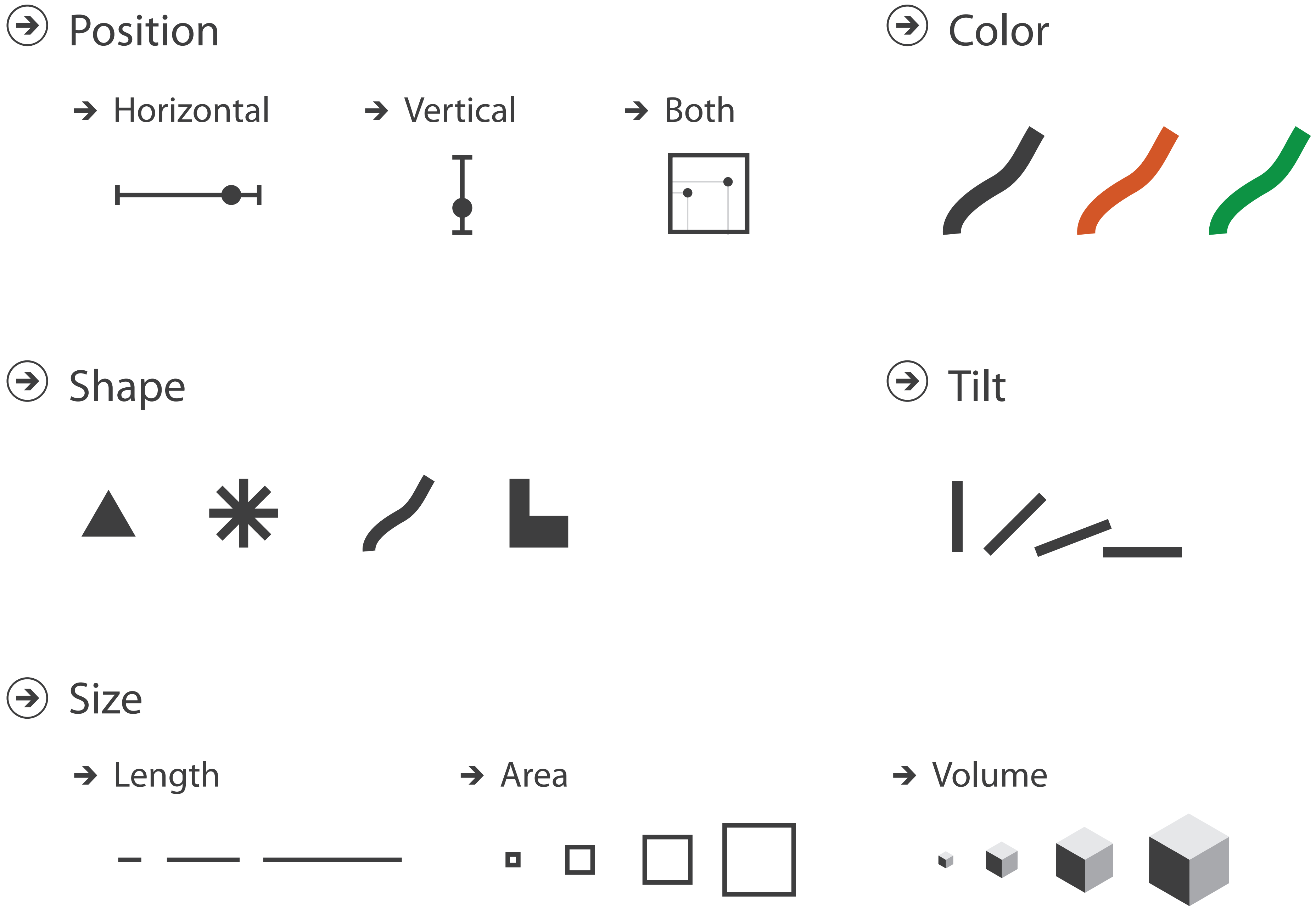
- analyze idiom structure



Definitions: Marks and channels

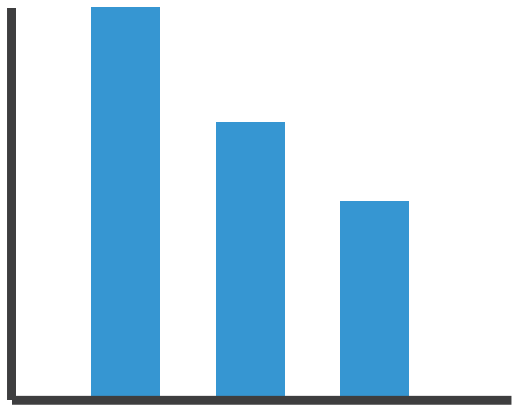
- marks
  - geometric primitives

- channels
  - control appearance of marks



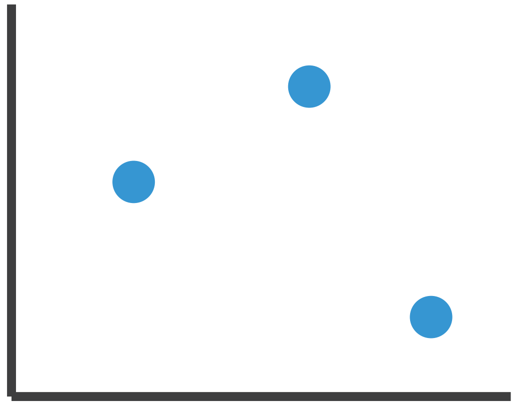
Encoding visually with marks and channels

- analyze idiom structure
  - as combination of marks and channels



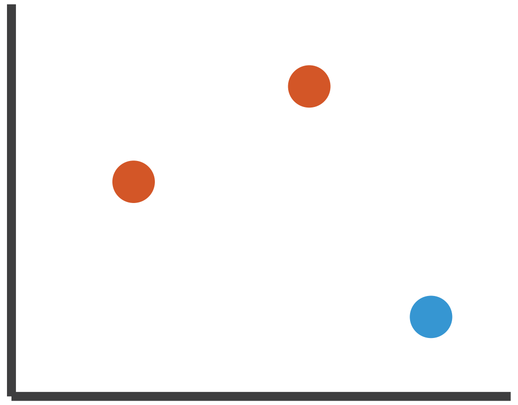
1:
vertical position

mark: line



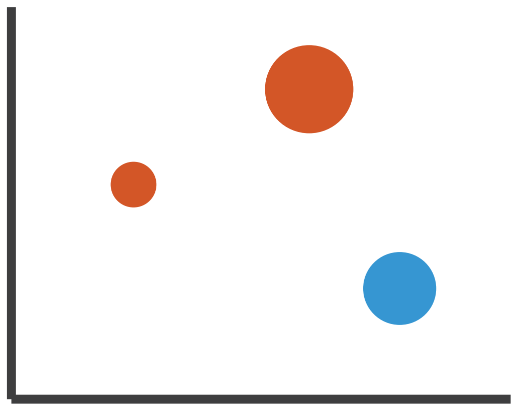
2:
vertical position
horizontal position

mark: point



3:
vertical position
horizontal position
color hue

mark: point

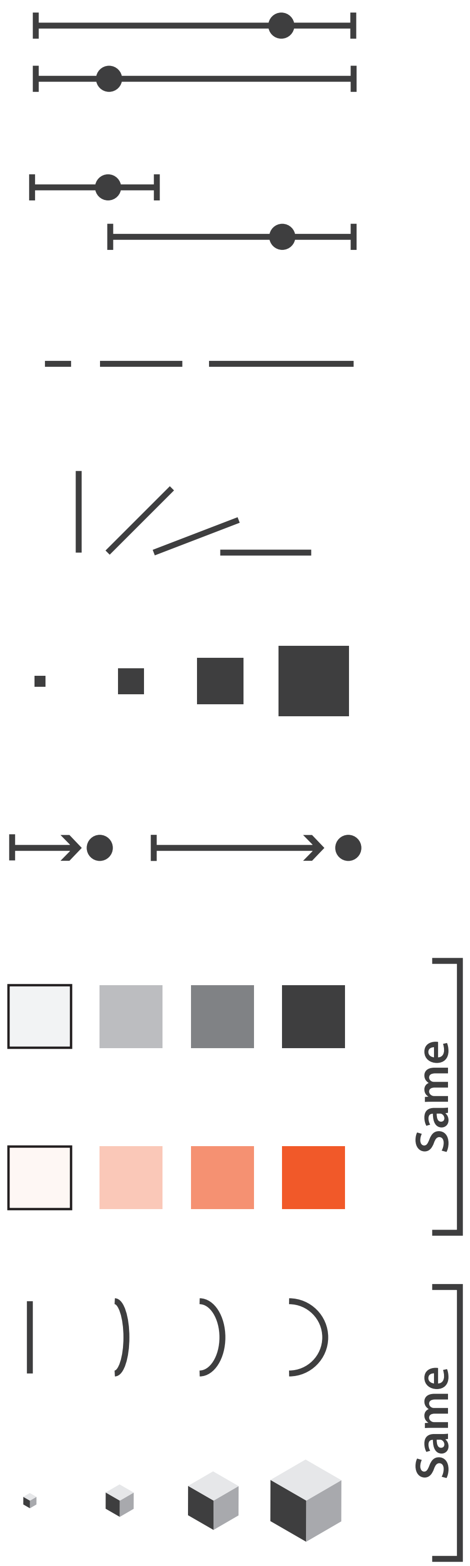


4:
vertical position
horizontal position
color hue
size (area)

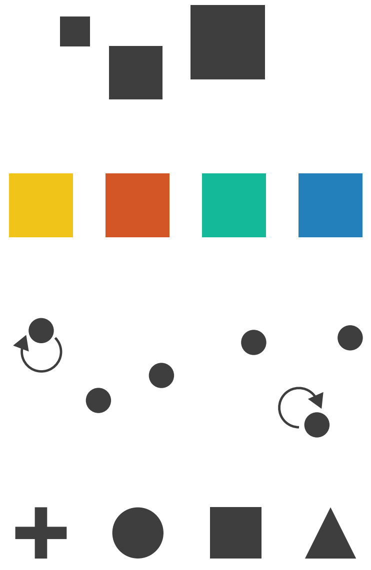
mark: point

Channels

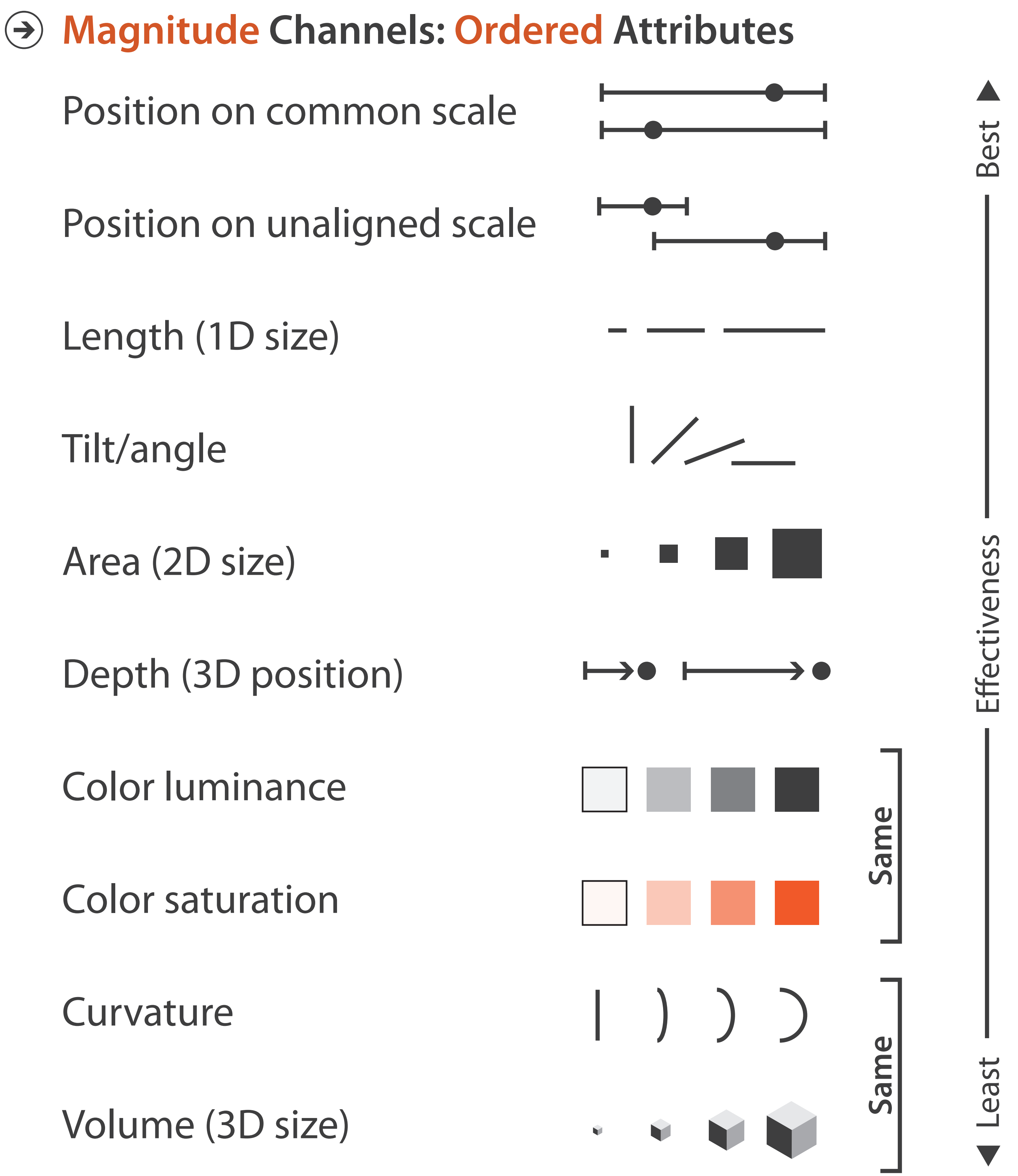
Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)

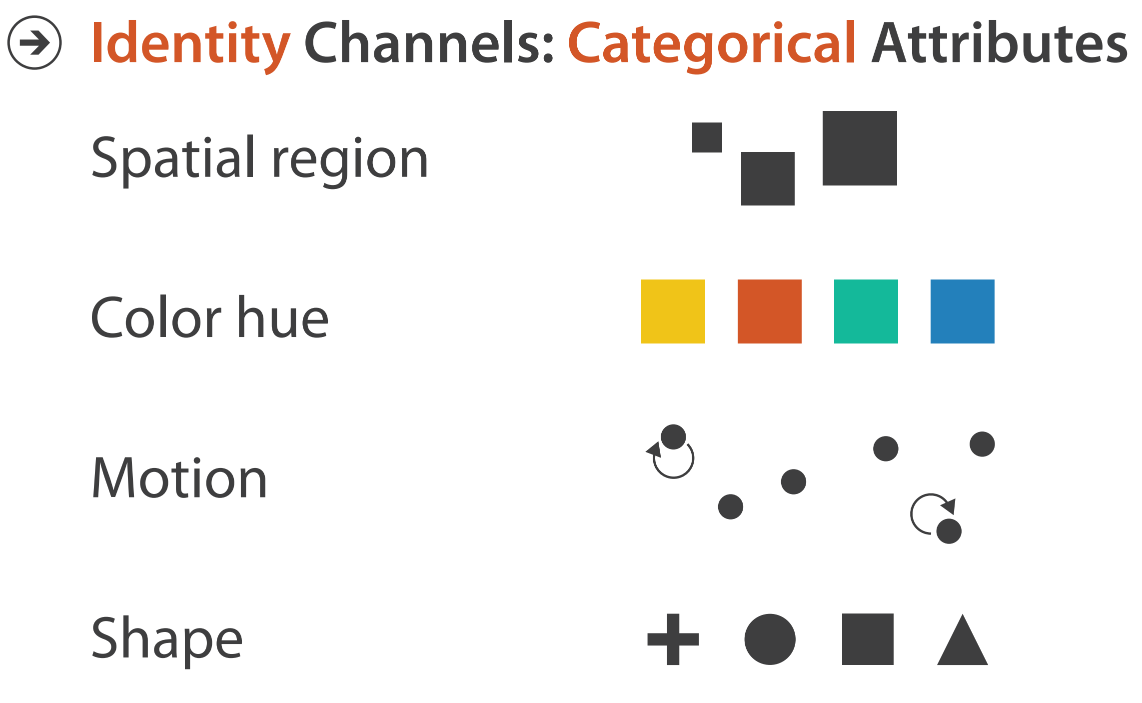


Spatial region
Color hue
Motion
Shape



Channels: Rankings

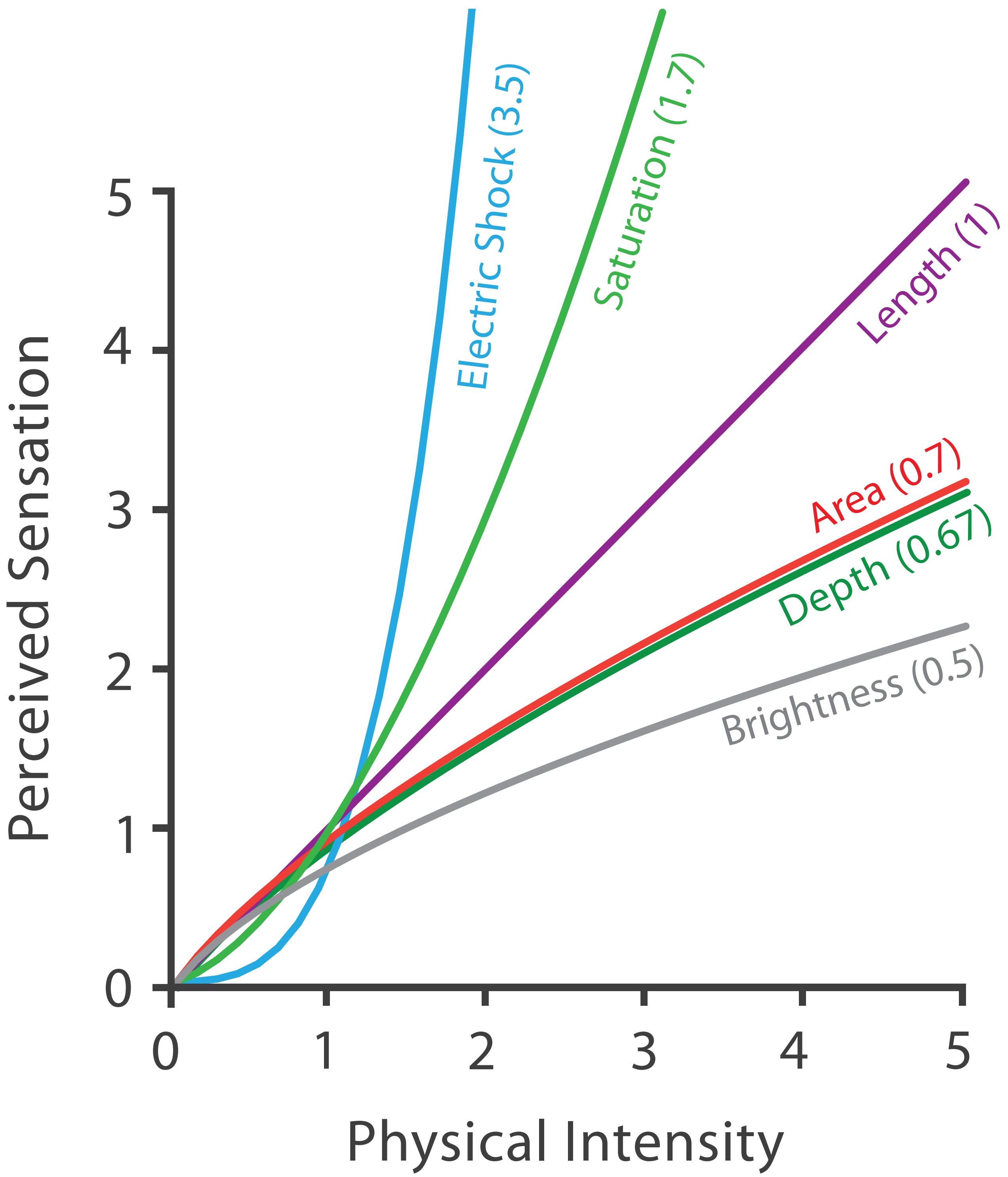




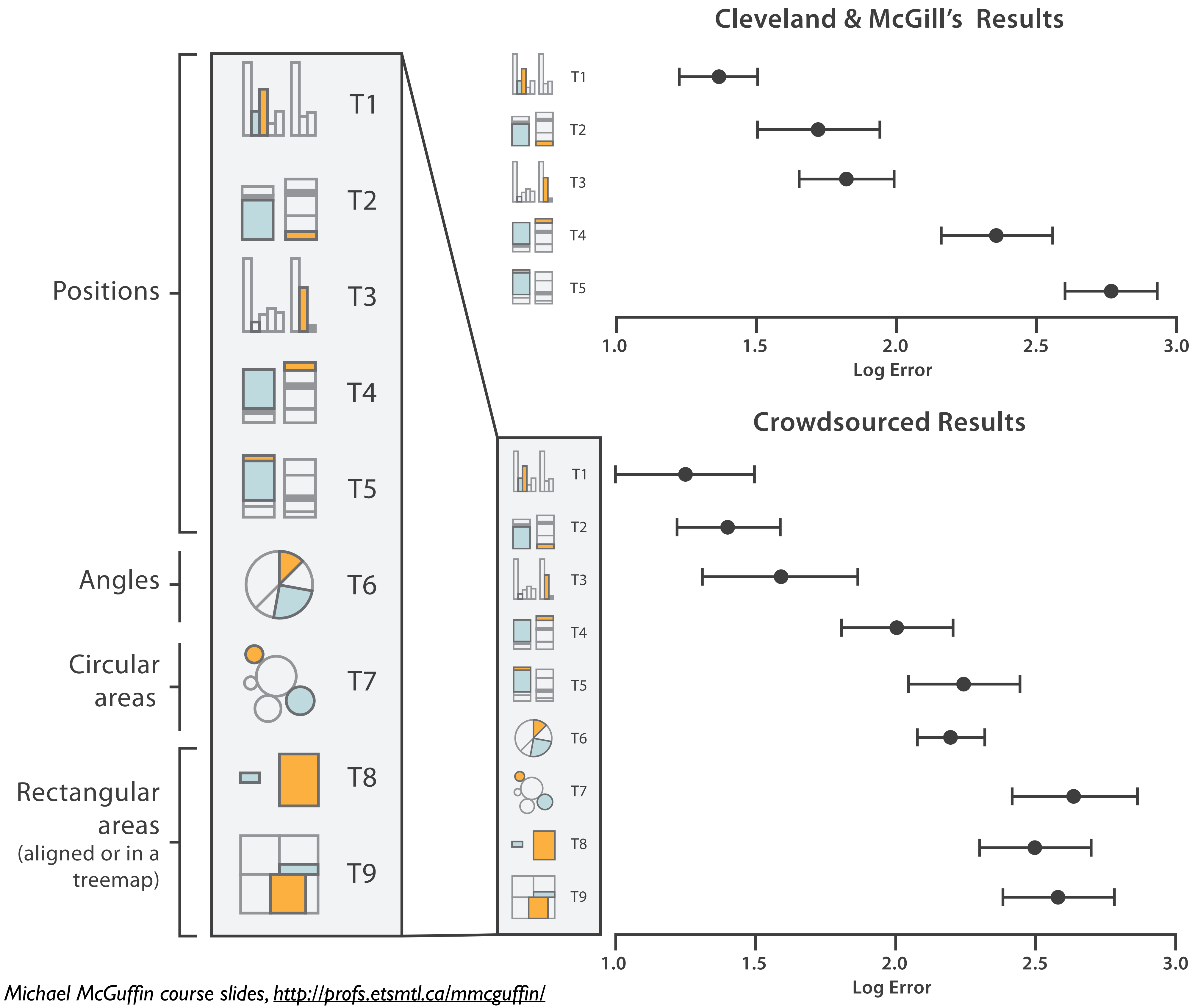
- effectiveness principle
  - encode most important attributes with highest ranked channels
- expressiveness principle
  - match channel and data characteristics

Accuracy: Fundamental Theory

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



Accuracy: Vis experiments

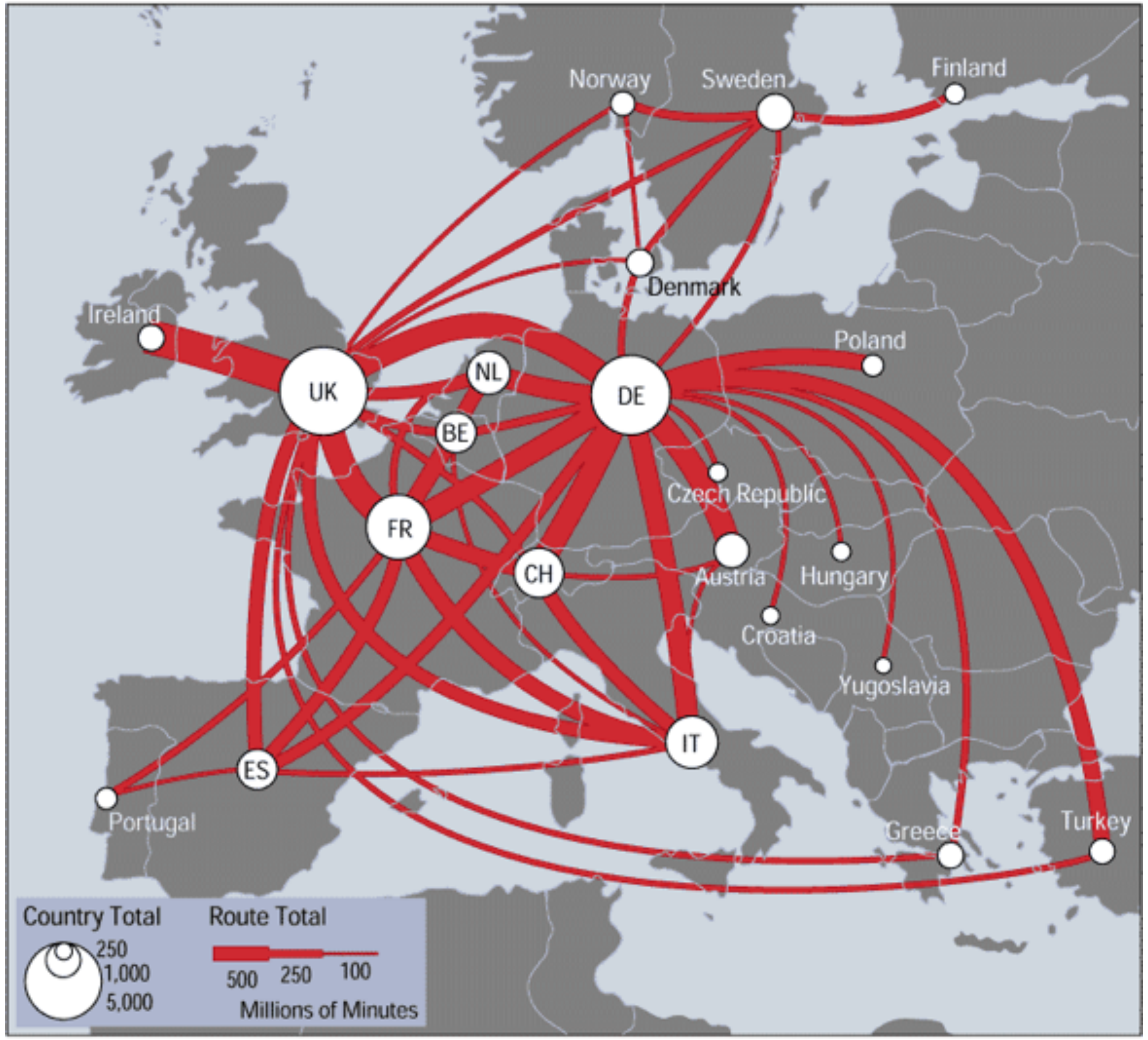


[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.]

after Michael McGuffin course slides, http://profs.etsmtl.ca/mmcguffin/

Discriminability: How many usable steps?

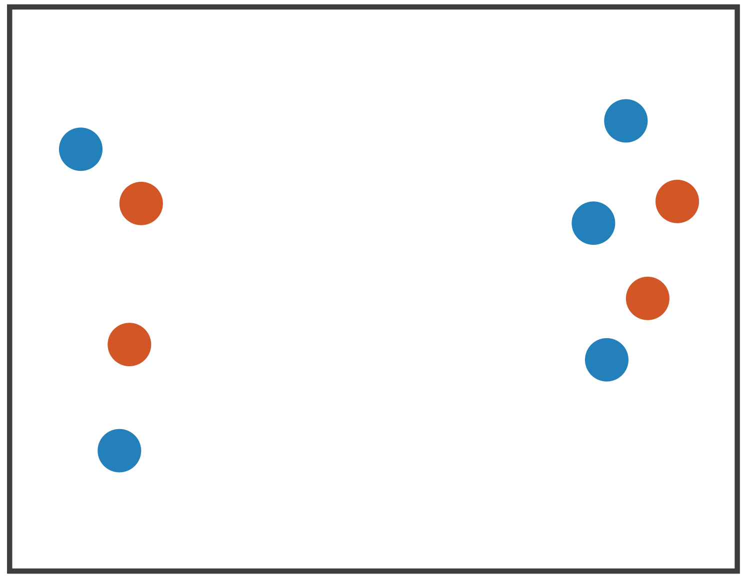
- must be sufficient for number of attribute levels to show
  - linewidth: few bins



[mappa.mundi.net/maps/maps 014/telegeography.html]

Separability vs. Integrality

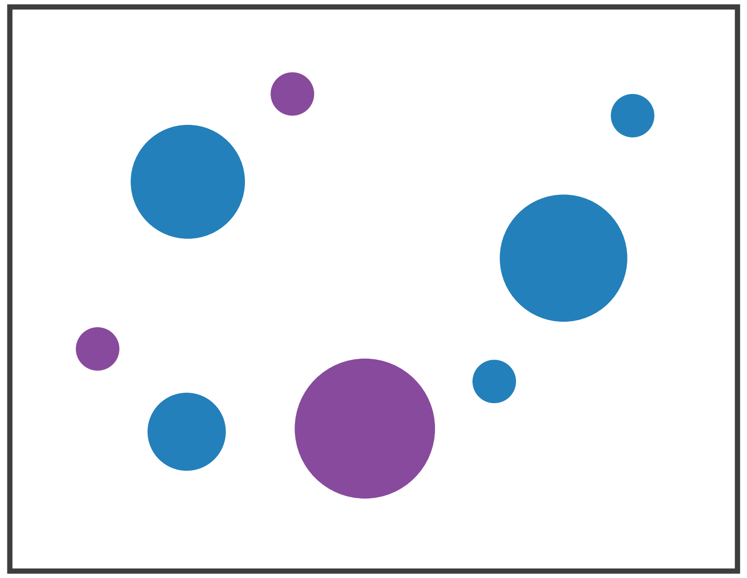
Position
+ Hue (Color)



Fully separable

2 groups each

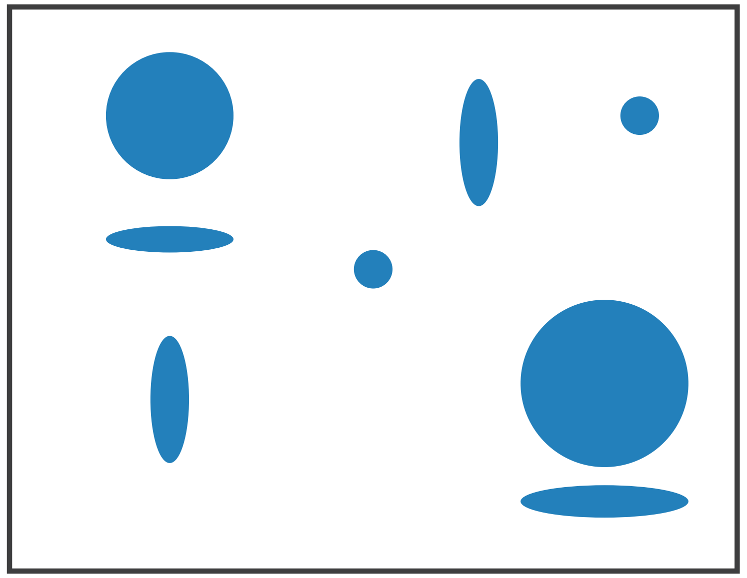
Size
+ Hue (Color)



Some interference

2 groups each

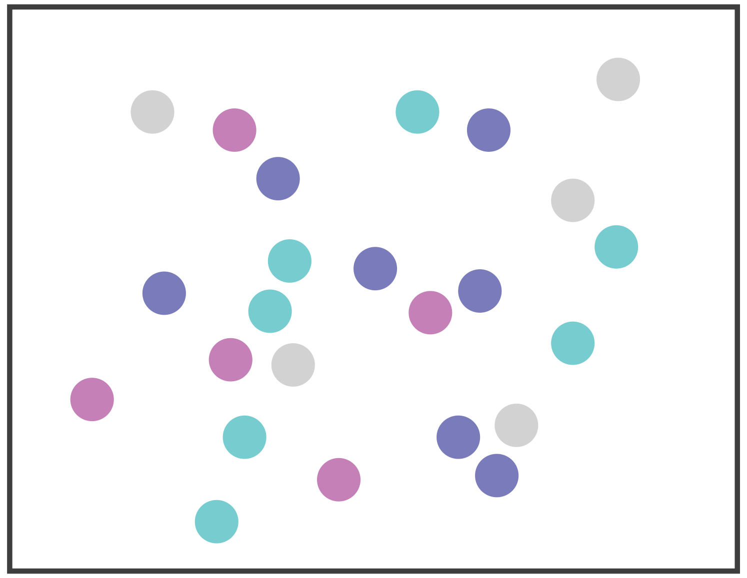
Width
+ Height



Some/significant interference

3 groups total: integral area

Red
+ Green

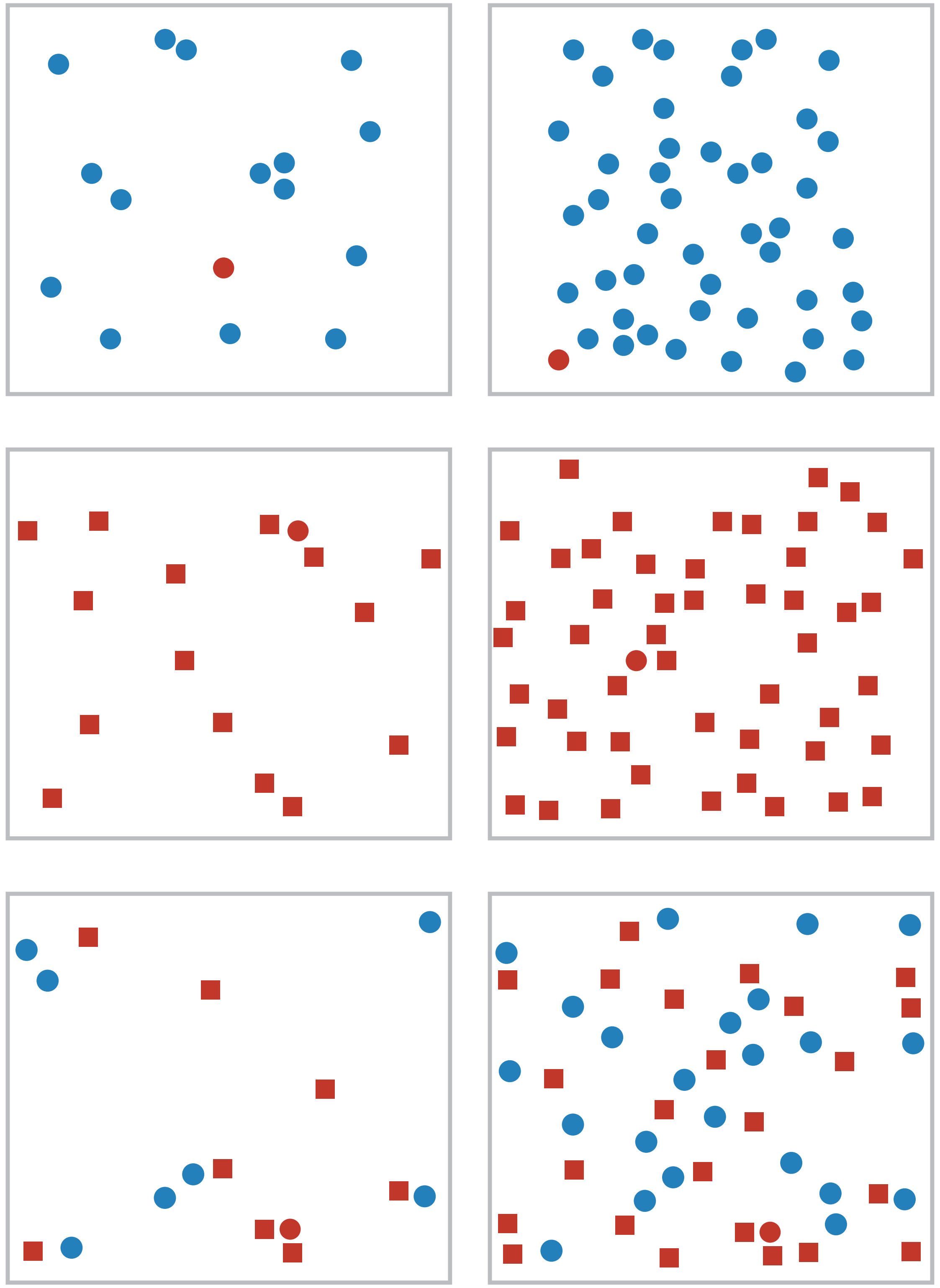


Major interference

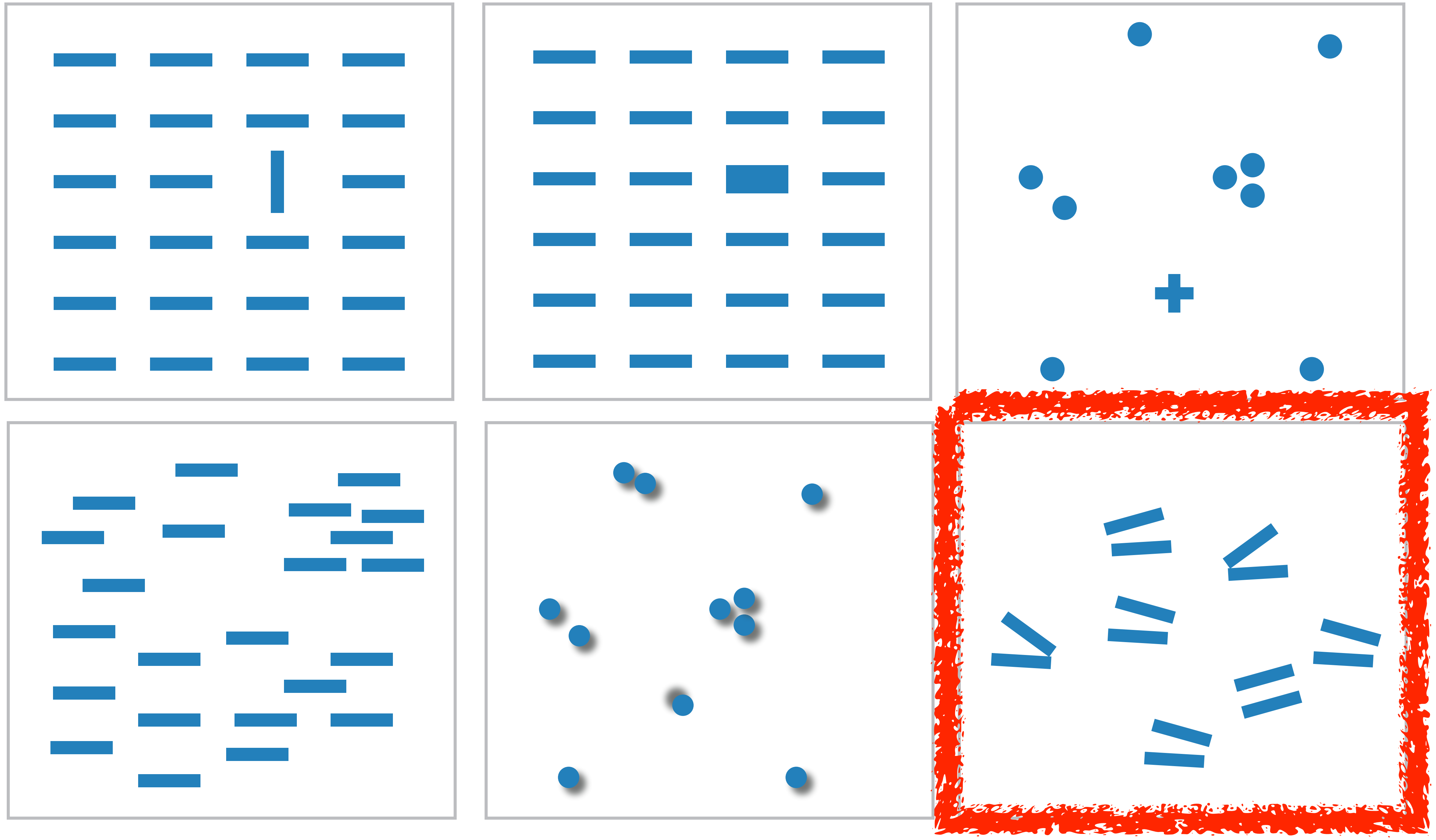
4 groups total: integral hue

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

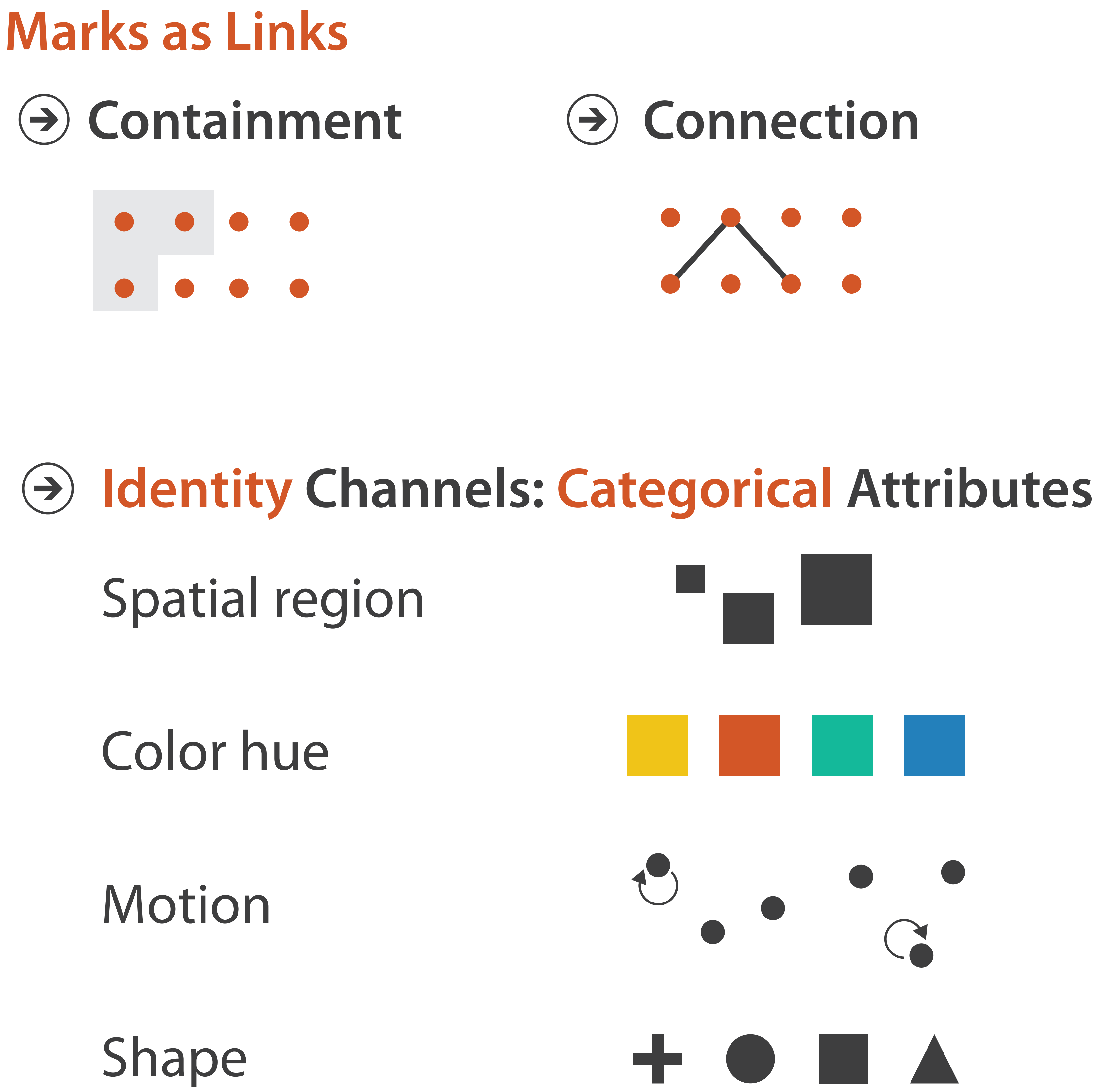


- many channels: tilt, size, shape, proximity, shadow direction, ...
- but not all! parallel line pairs do not pop out from tilted pairs

Grouping

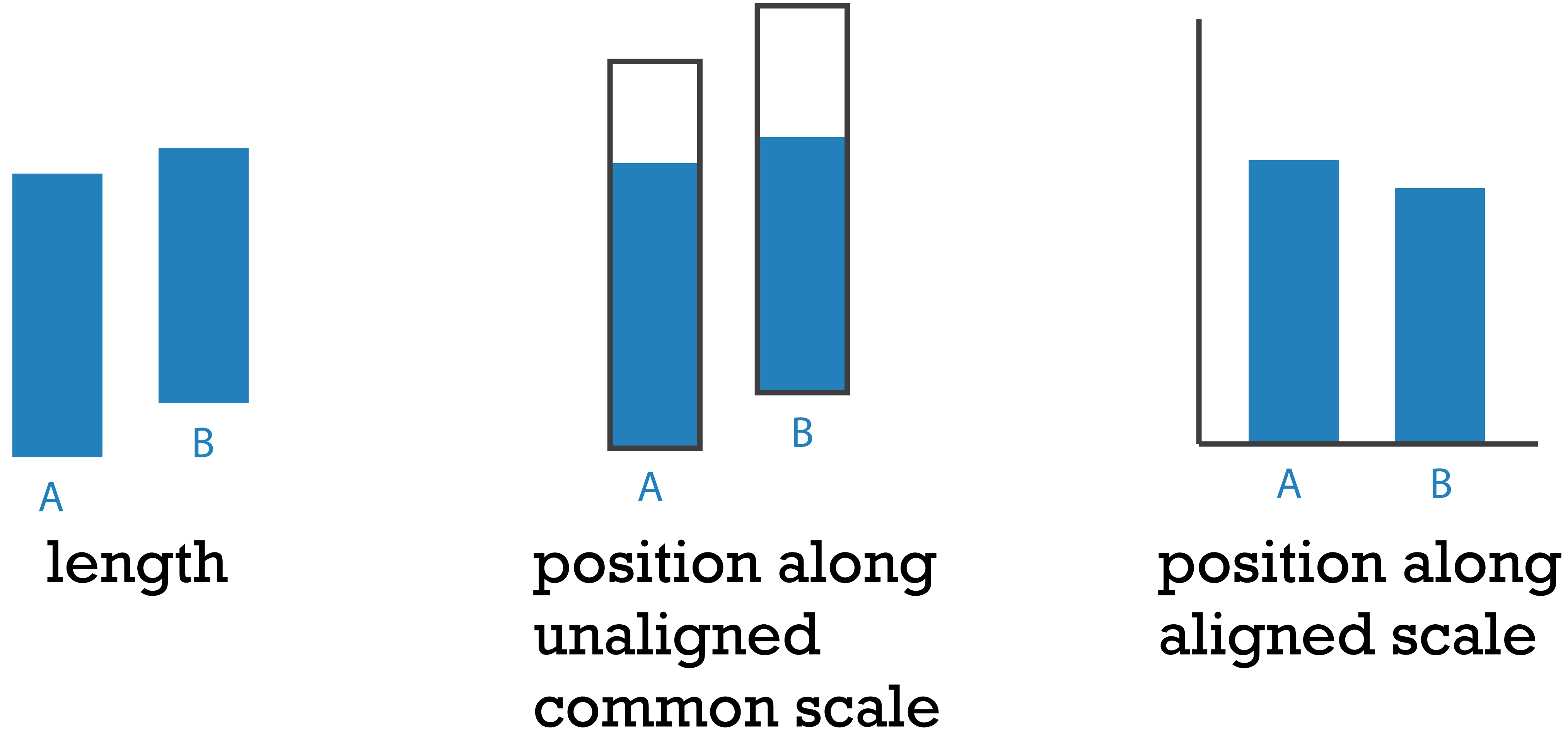
- containment
- connection

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



Relative vs. absolute judgements

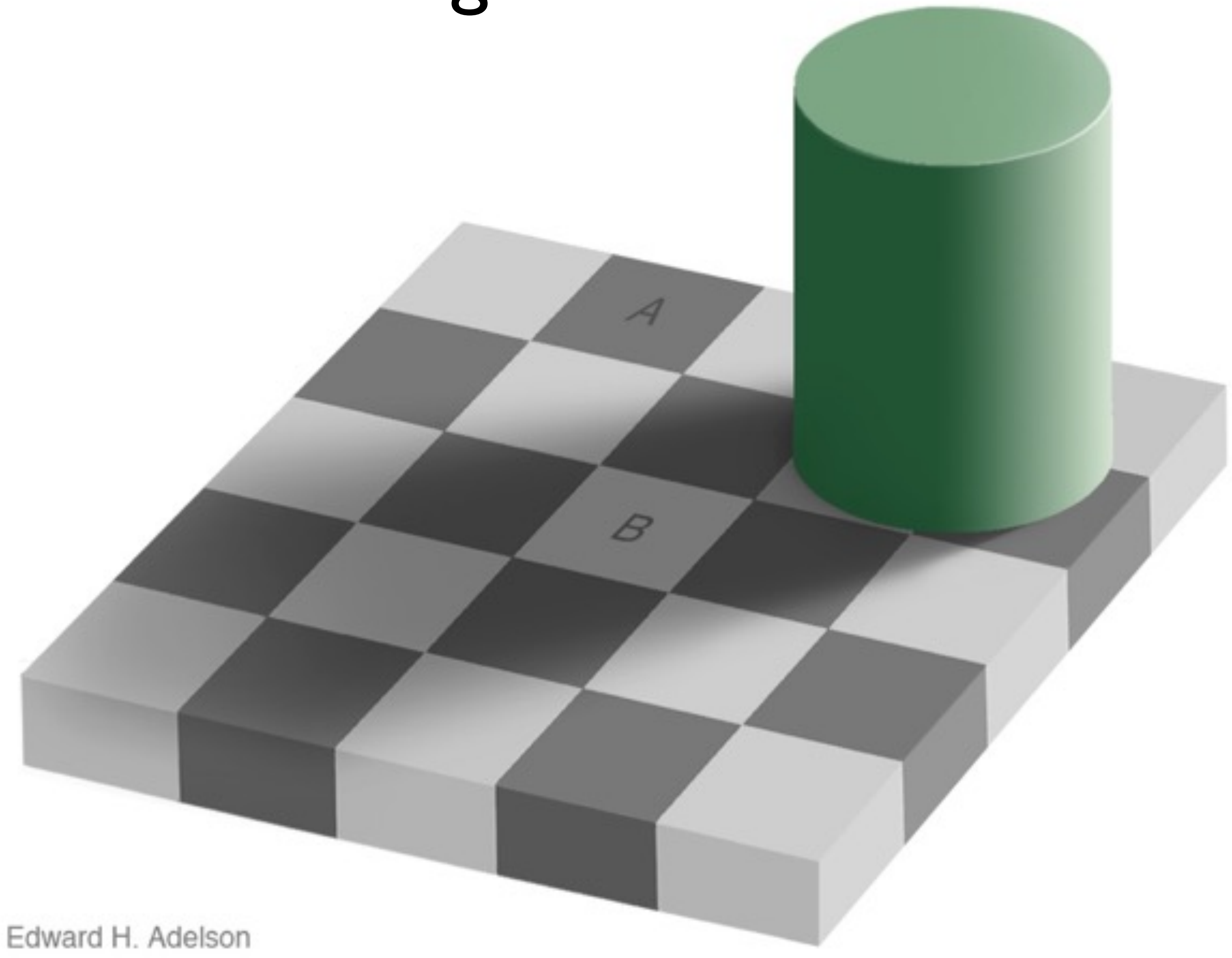
- perceptual system mostly operates with relative judgements, not absolute
  - that's why accuracy increases with common frame/scale and alignment
  - Weber's Law: ratio of increment to background is constant
    - filled rectangles differ in length by 1:9, difficult judgement
    - white rectangles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

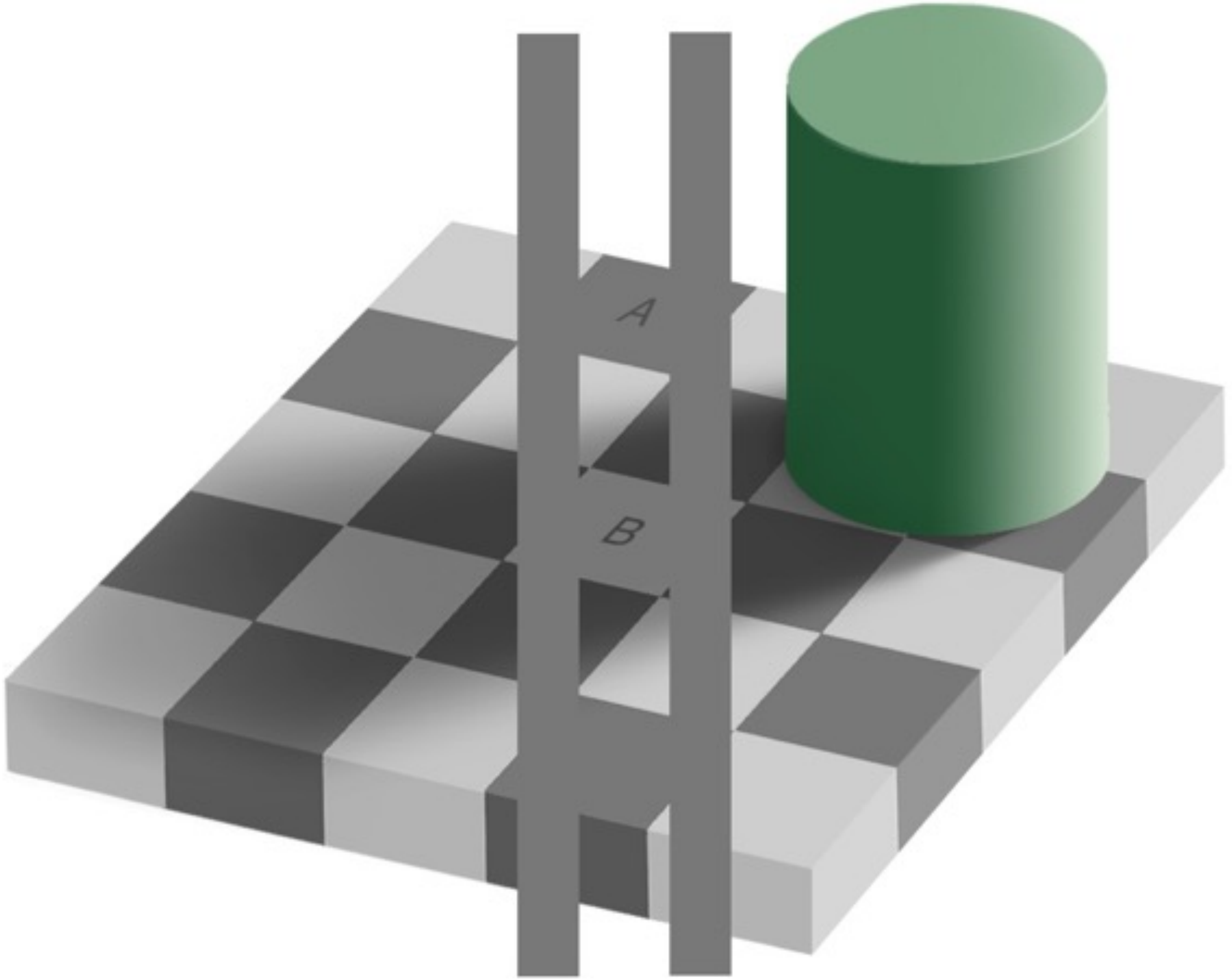


after [Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Cleveland and McGill. Journ. American Statistical Association 79:387 (1984), 531–554.]

Relative luminance judgements

- perception of luminance is contextual based on contrast with surroundings

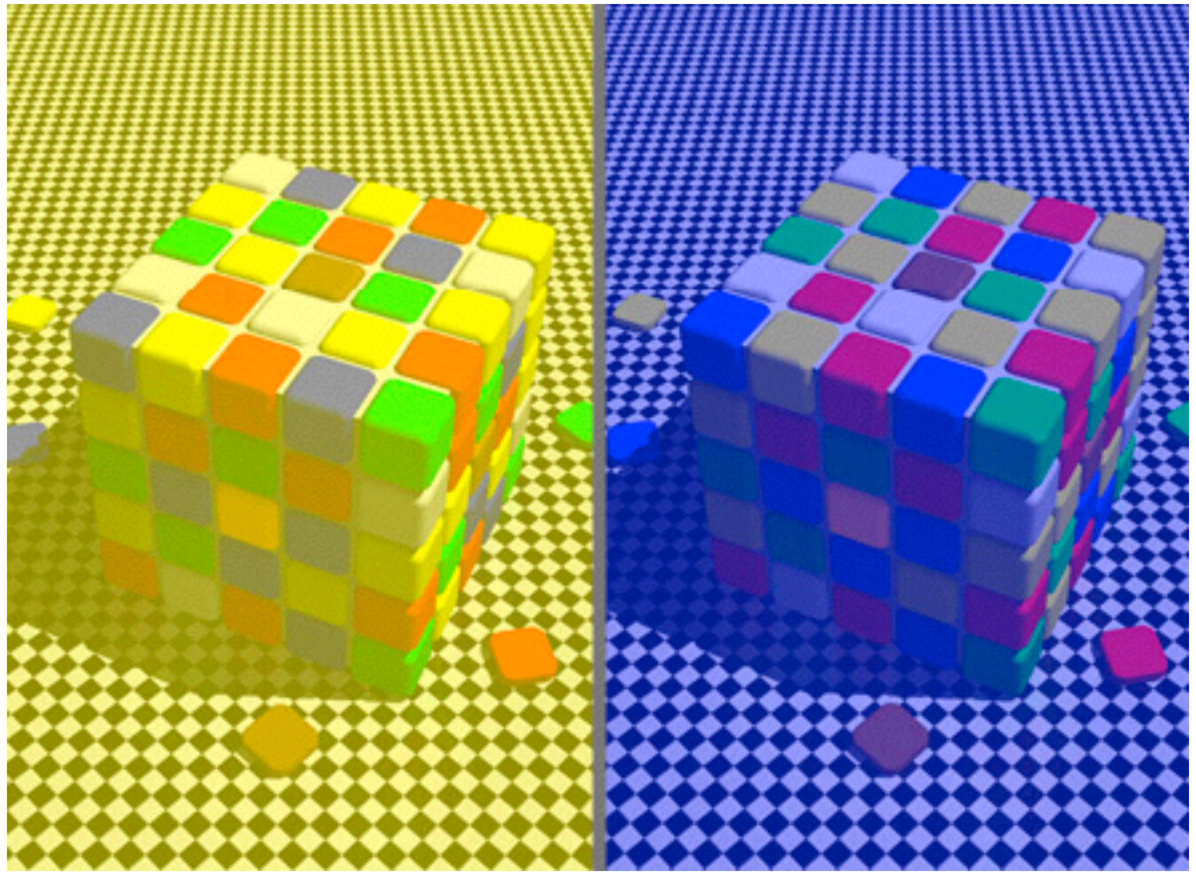


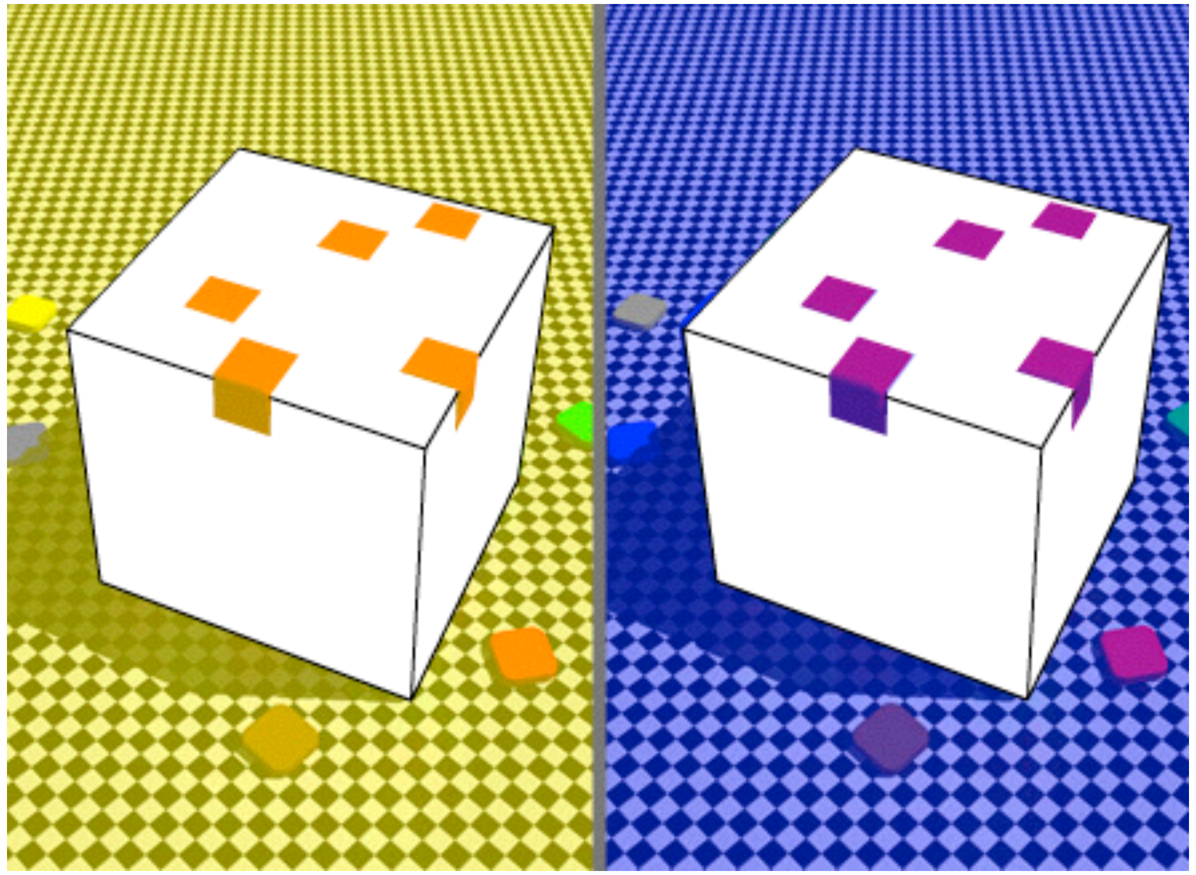


http://persci.mit.edu/gallery/checkershadow

Relative color judgements

- color constancy across broad range of illumination conditions





http://www.purveslab.net/seeforyourself/

- Visualization Analysis and Design. Tamara Munzner. CRC Press, 2014.
  - *Chap 1, What's Vis, and Why Do It?*
  - *Chap 2, What: Data Abstraction*
  - *Chap 3, Why: Task Abstraction*
  - *Chap 4, Analysis: Four Levels for Validation*
  - *Chap 5, Marks and Channels*

- Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Jeffrey Heer and Michael Bostock. Proc. CHI 2010
- Perception in Vision web page with demos, Christopher Healey.
- Visual Thinking for Design. Colin Ware. Morgan Kaufmann, 2008.