Visualization Principles

Tamara Munzner
Department of Computer Science
University of British Columbia

VIZBI 2011: Workshop on Visualizing Biological Data
16 Mar 2011

http://www.cs.ubc.ca/~tmm/talks.html#vizbi11

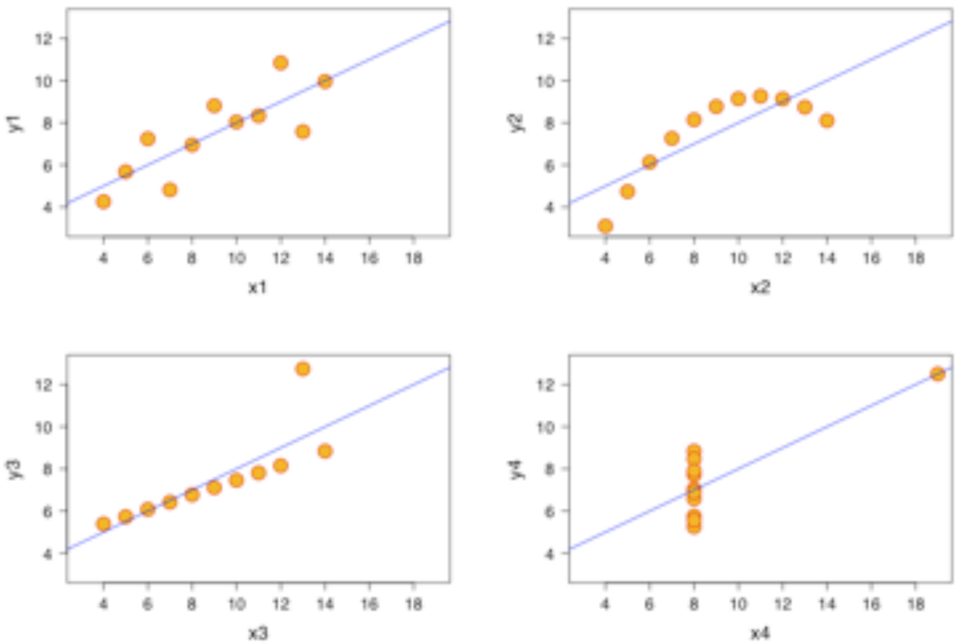
Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively

Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively

- human in the loop needs the details



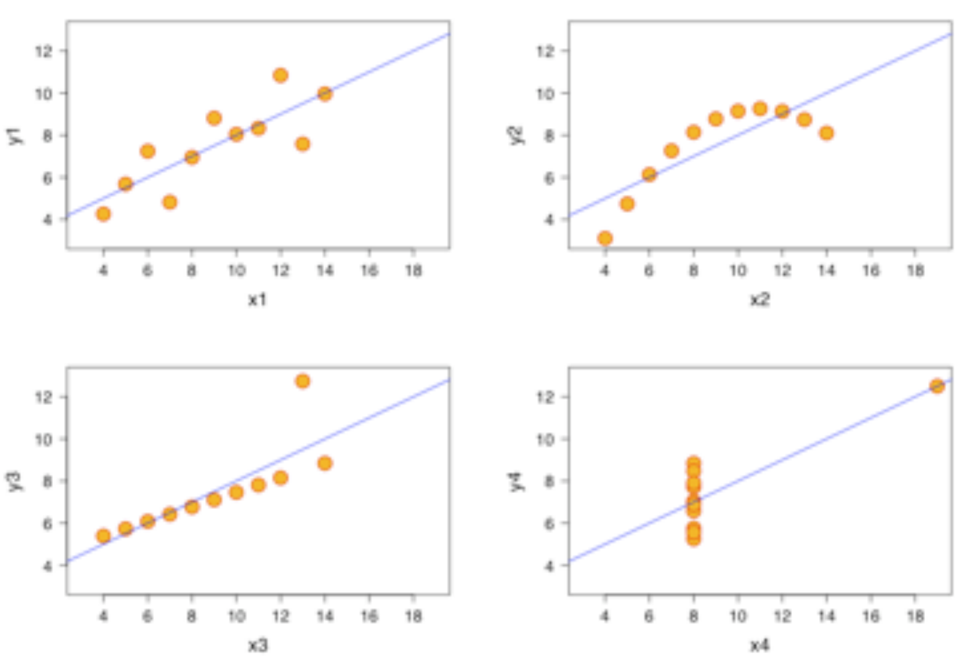
http://upload.wikimedia.org/wikipedia/commons/b/b6/Anscombe.svg

Identical statistics	
x mean	9.0
x variance	10.0
y mean	7.50
y variance	3.75
x/y correlation	0.816

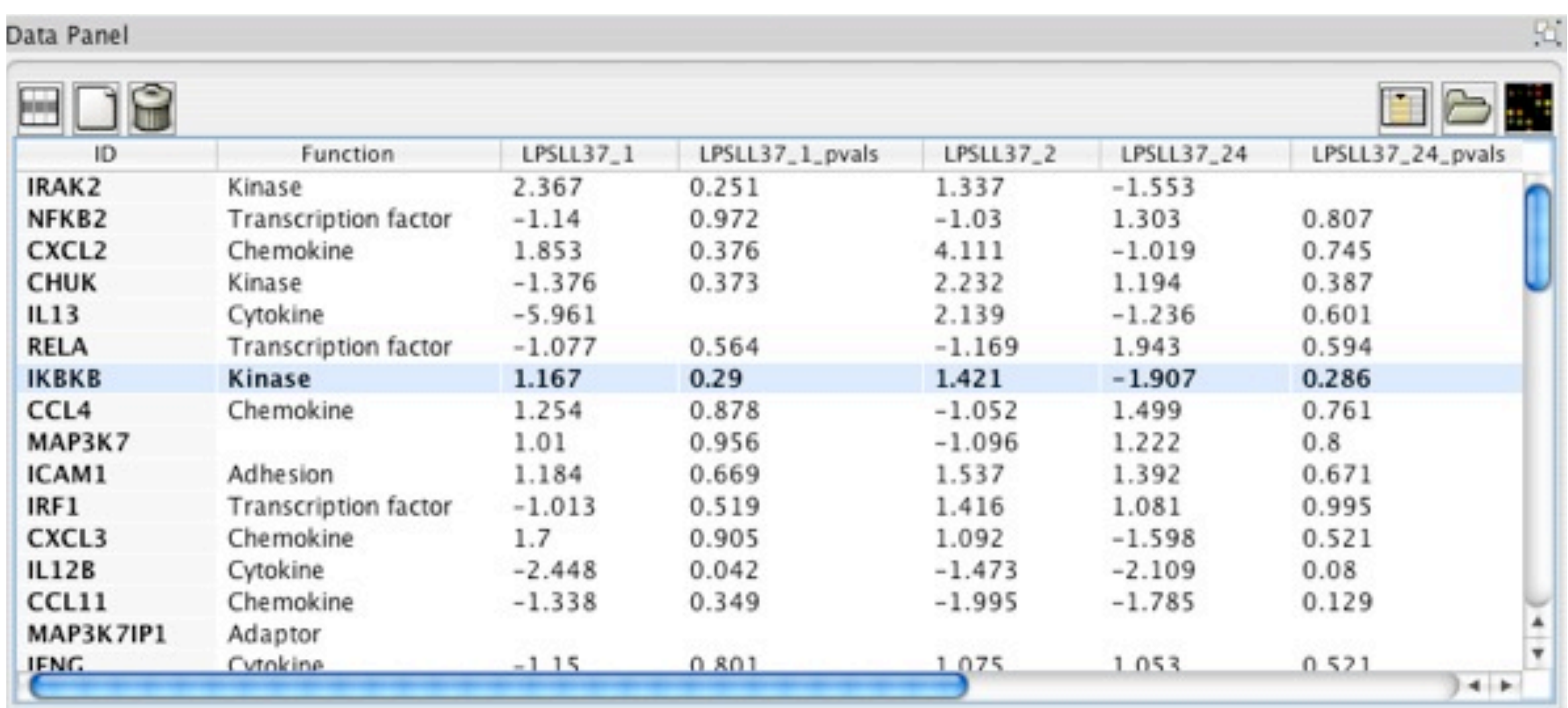
Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively

- human in the loop needs the details
- external representation: perception vs cognition



http://upload.wikimedia.org/wikipedia/commons/b/b6/Anscombe.svg

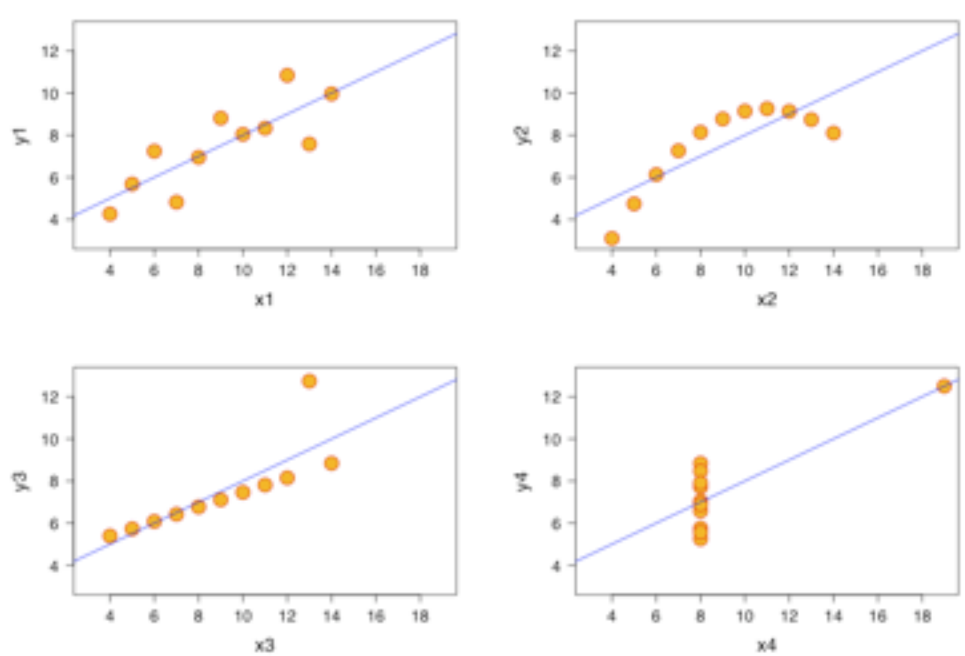


ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521

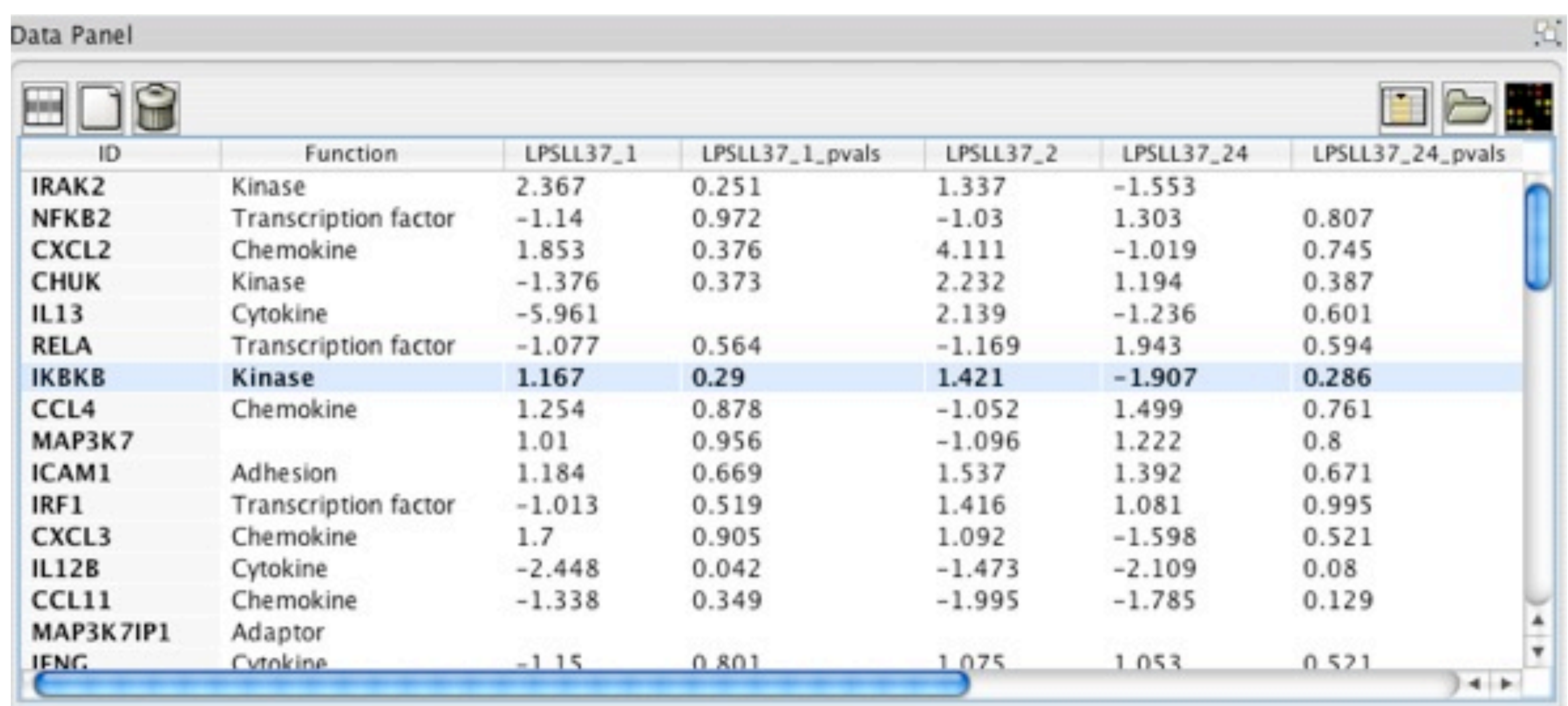
Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively

- human in the loop needs the details
- external representation: perception vs cognition
- intended task



http://upload.wikimedia.org/wikipedia/commons/b/b6/Anscombe.svg



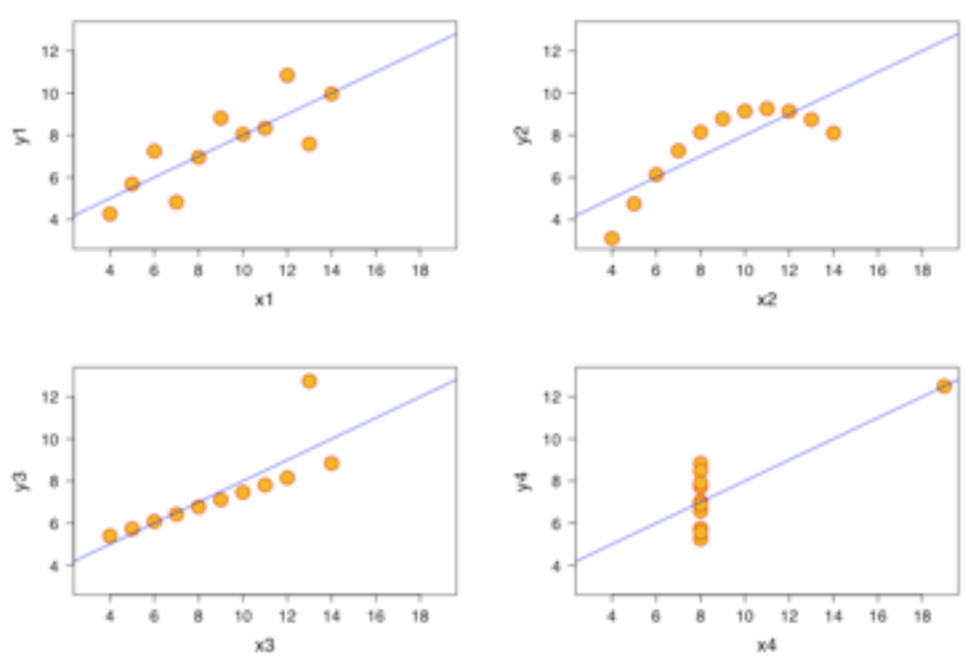
Data Panel

ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521

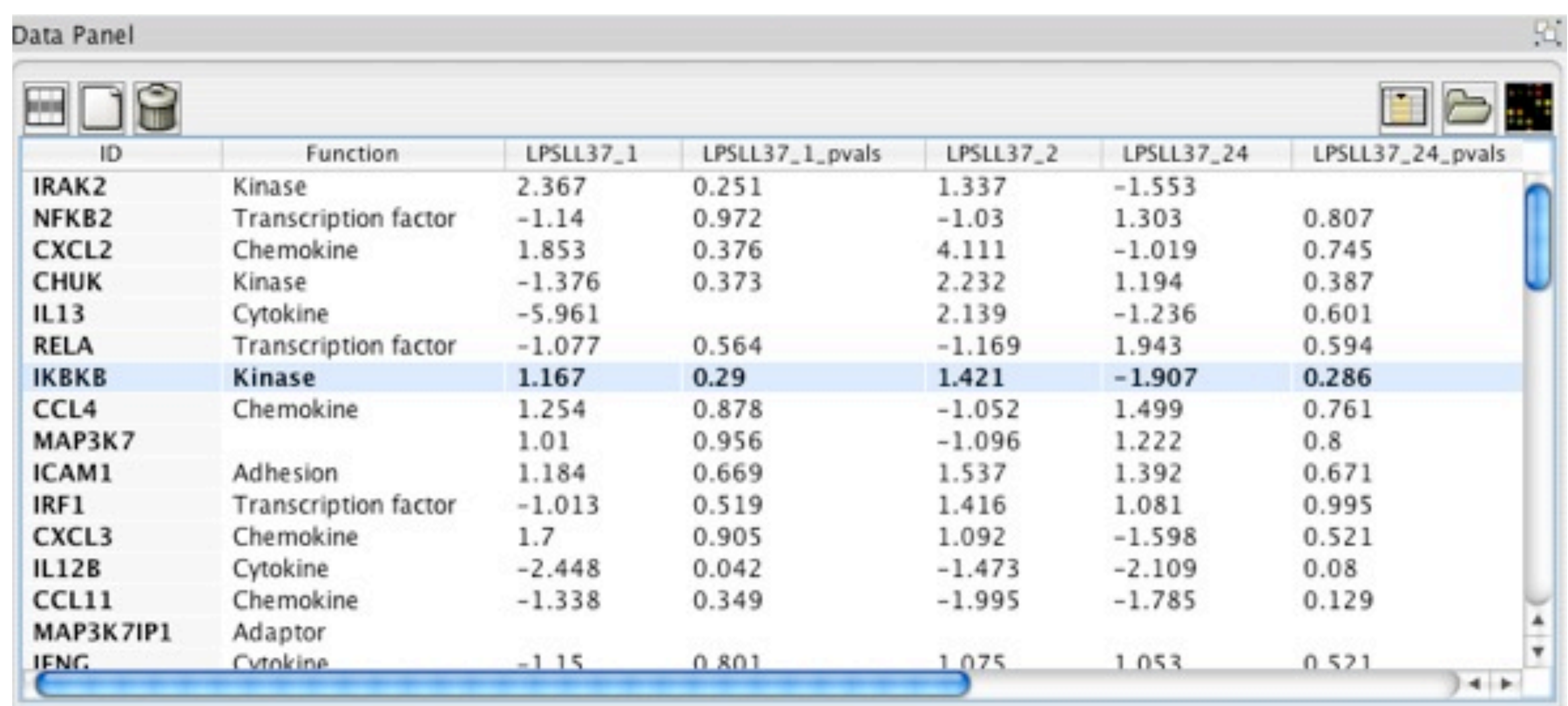
Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively

- human in the loop needs the details
- external representation: perception vs cognition
- intended task
- measureable definitions of effectiveness



http://upload.wikimedia.org/wikipedia/commons/b/b6/Anscombe.svg



Data Panel

ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	**Kinase**	**1.167**	**0.29**	**1.421**	**-1.907**	**0.286**
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521

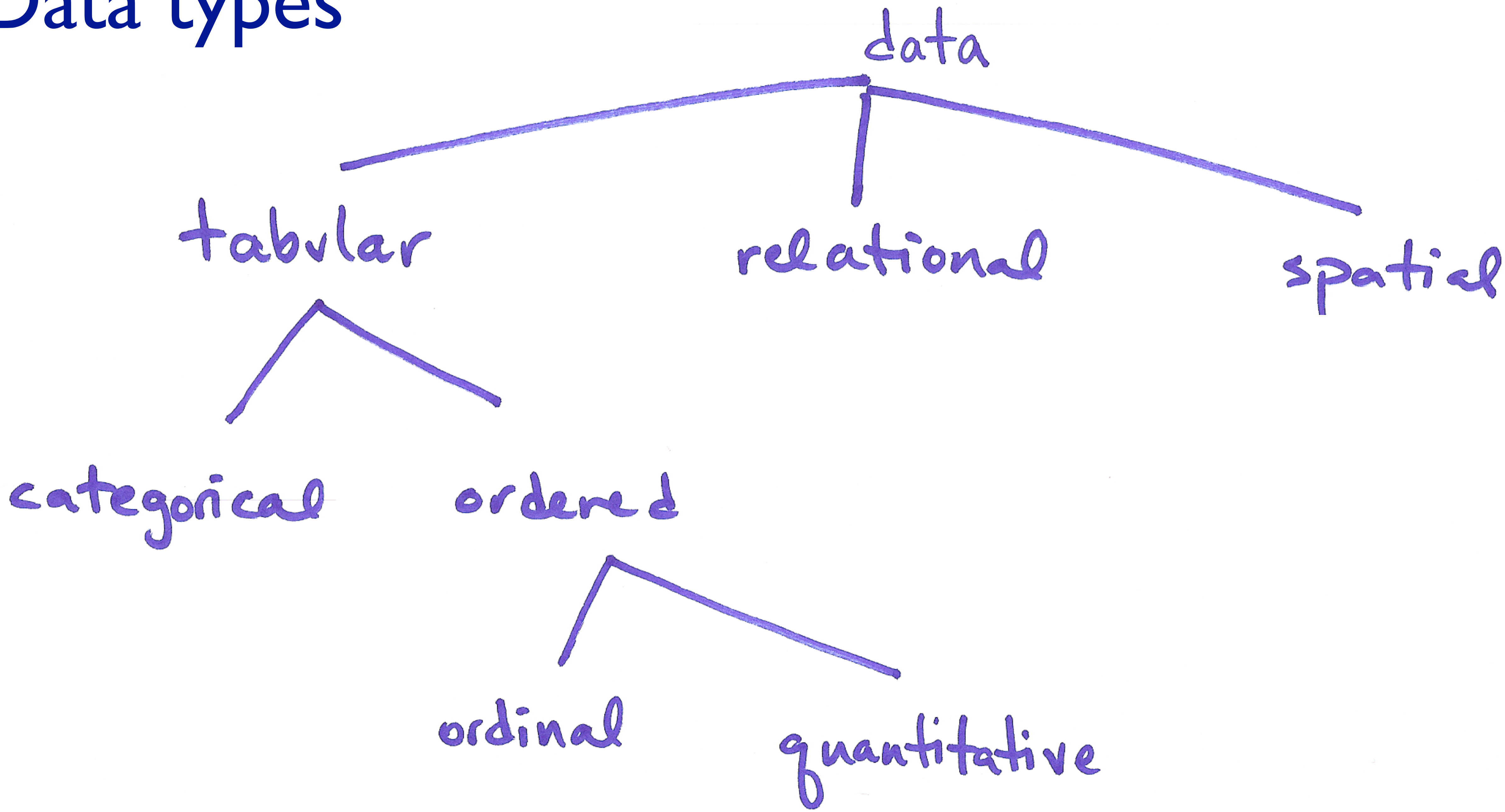
Visualization design space

- huge space of design alternatives
  - tradeoffs abound
- many possibilities now known to be ineffective
  - avoid random walk through parameter space
  - avoid some of our past mistakes
  - extensive experimentation has already been done
- guidelines continue to evolve
  - we reflect on lessons learned in design studies
  - iterative refinement usually wise

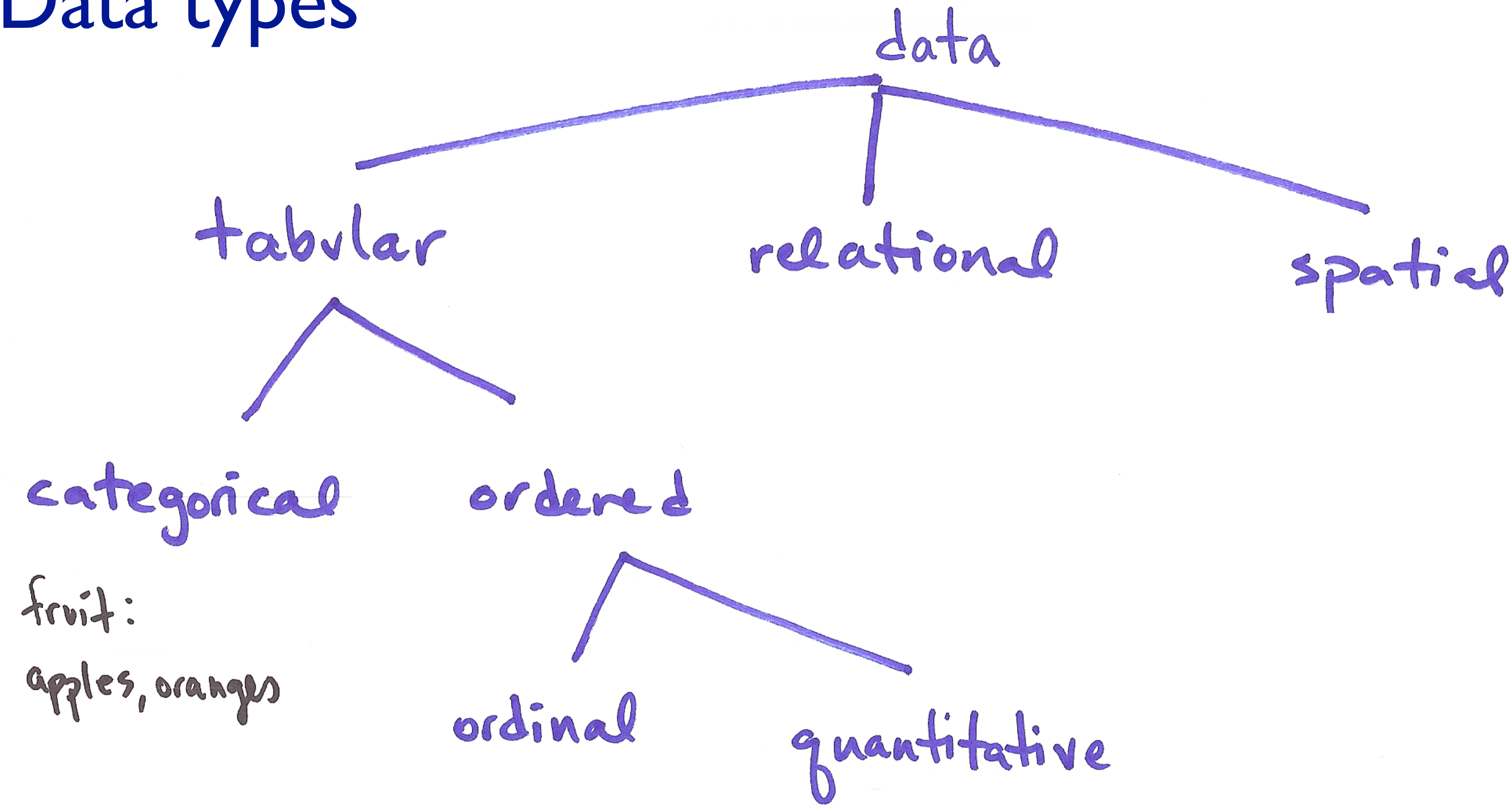
Principles

- know your visual channel types and ranks
- categorical color constraints
- power of the plane
- danger of depth
- resolution beats immersion
- eyes beat memory
- validate against the right threat

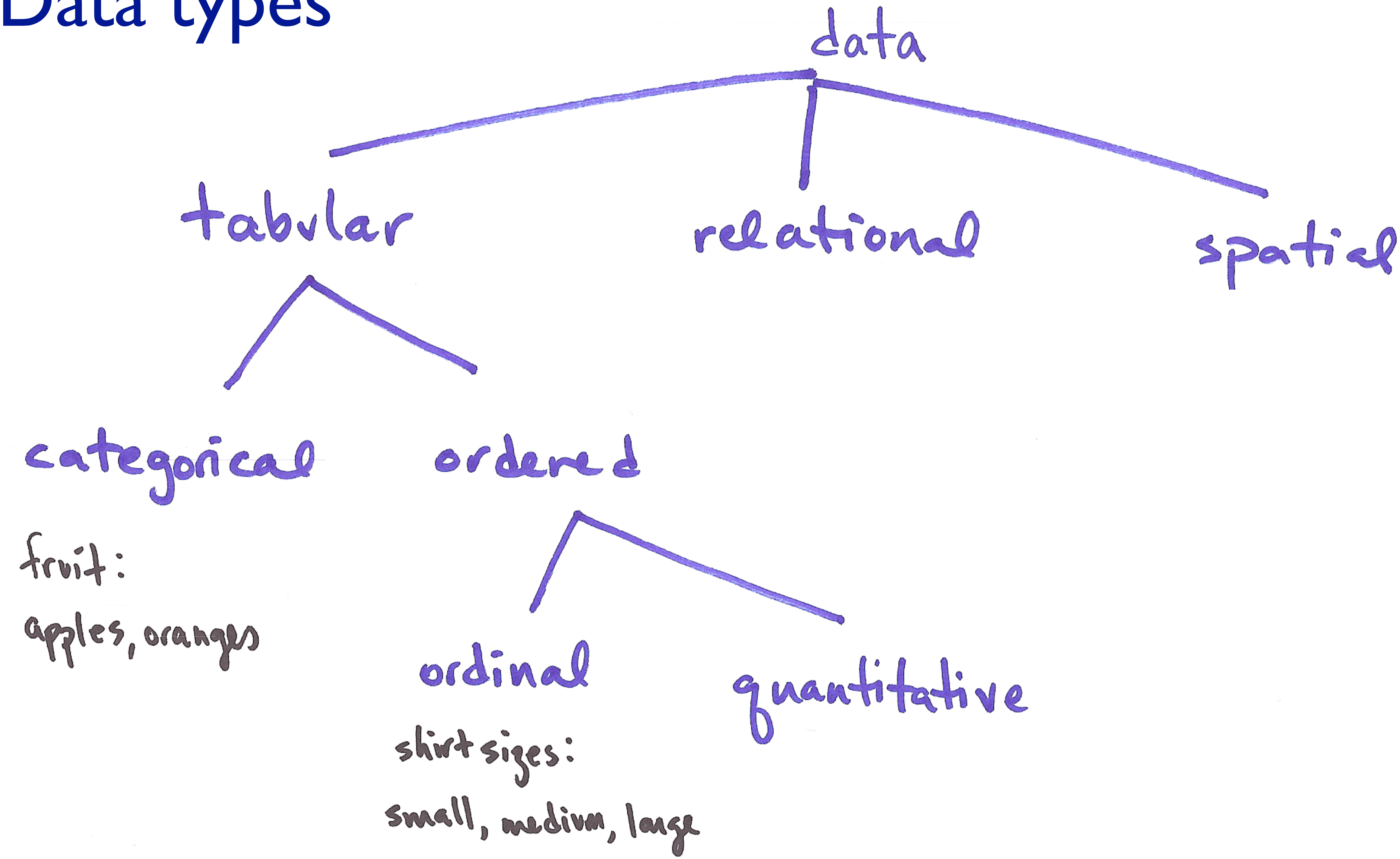
Data types


data
tabular
relational
spatial
categorical
ordered
ordinal
quantitative

Data types

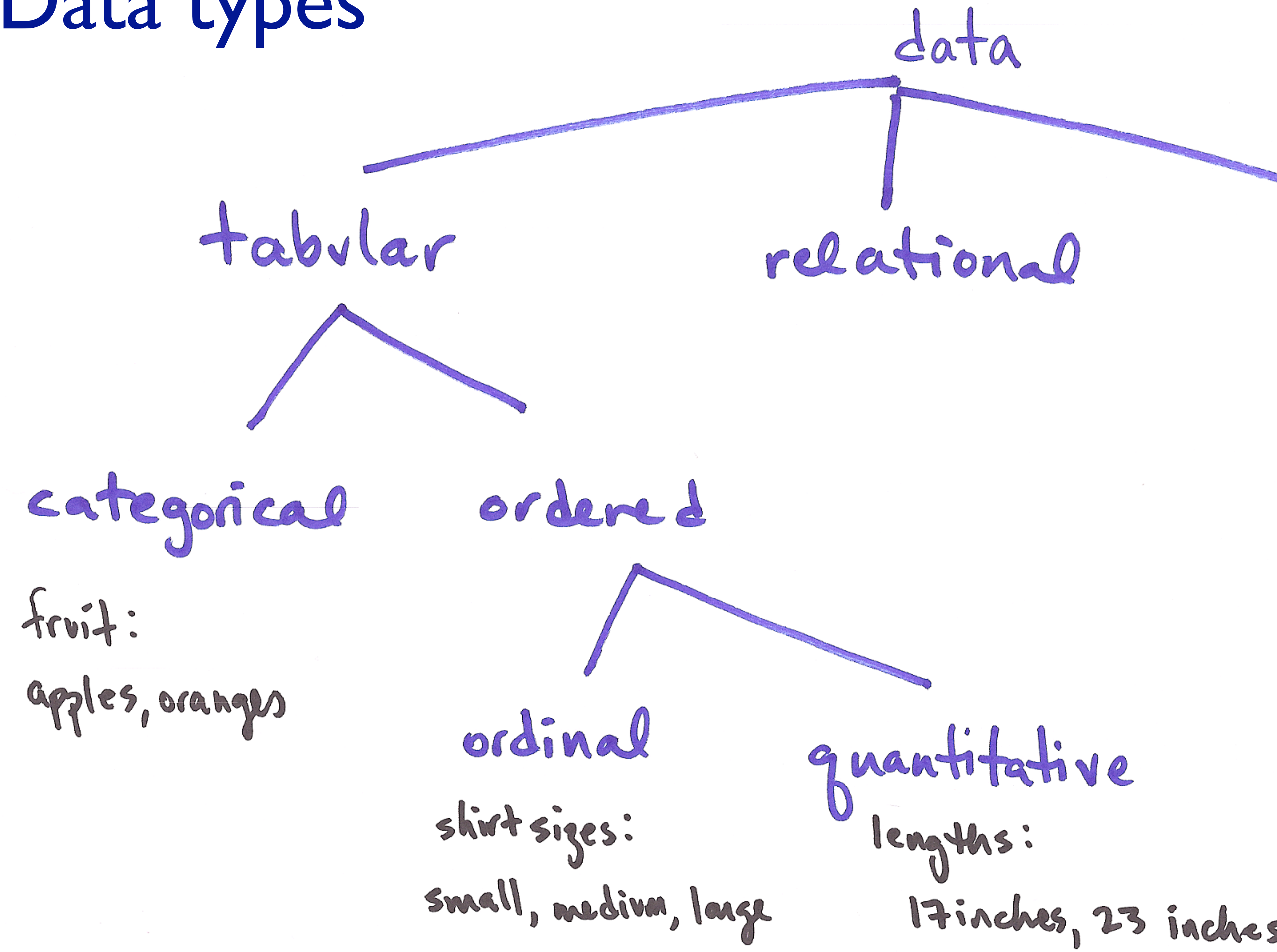

data
tabular
relational
spatial
categorical
ordered
fruit:
apples, oranges
ordinal
quantitative

Data types

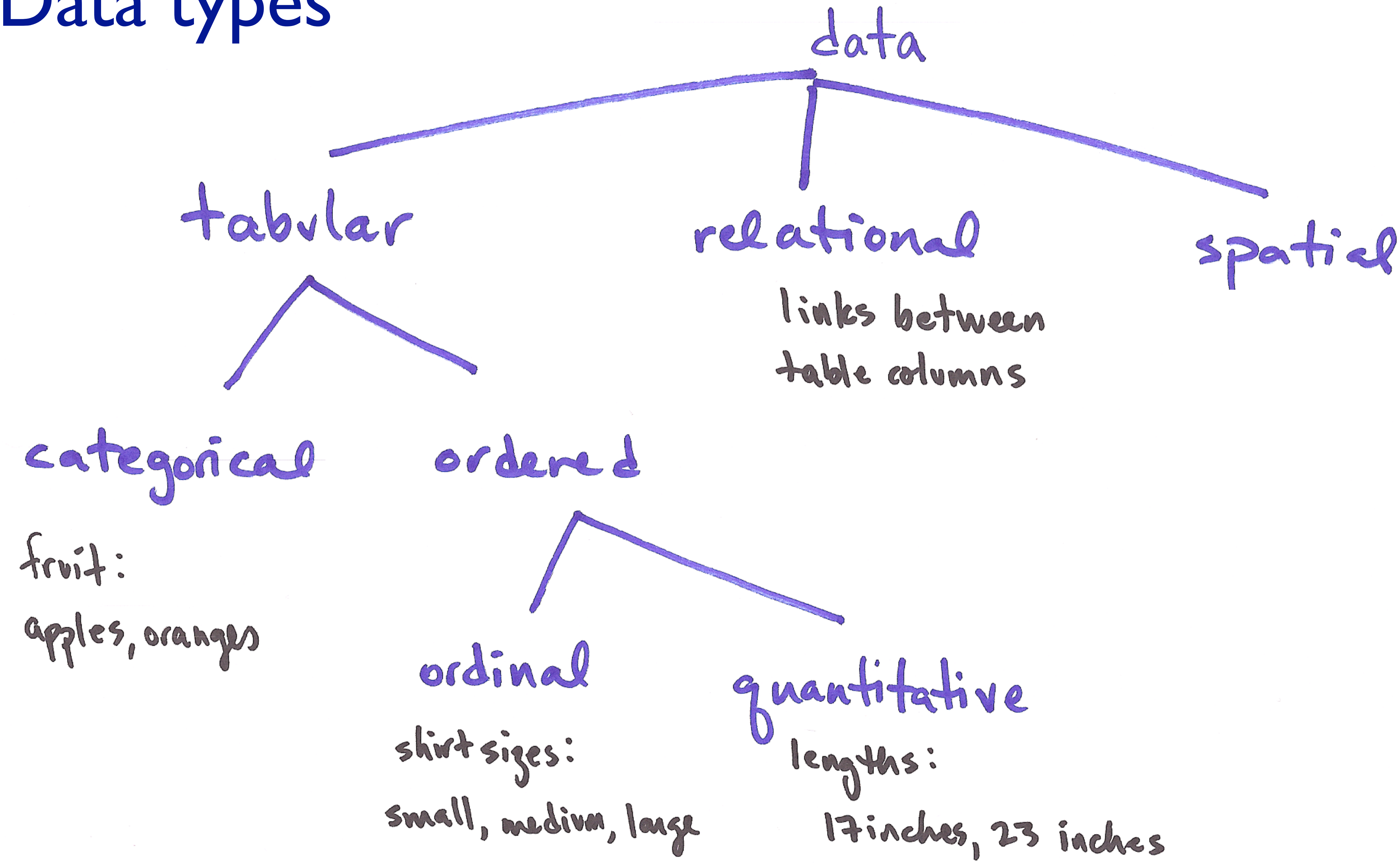


data
tabular
relational
spatial
categorical
ordered
fruit:
apples, oranges
ordinal
quantitative
shirt sizes:
small, medium, large

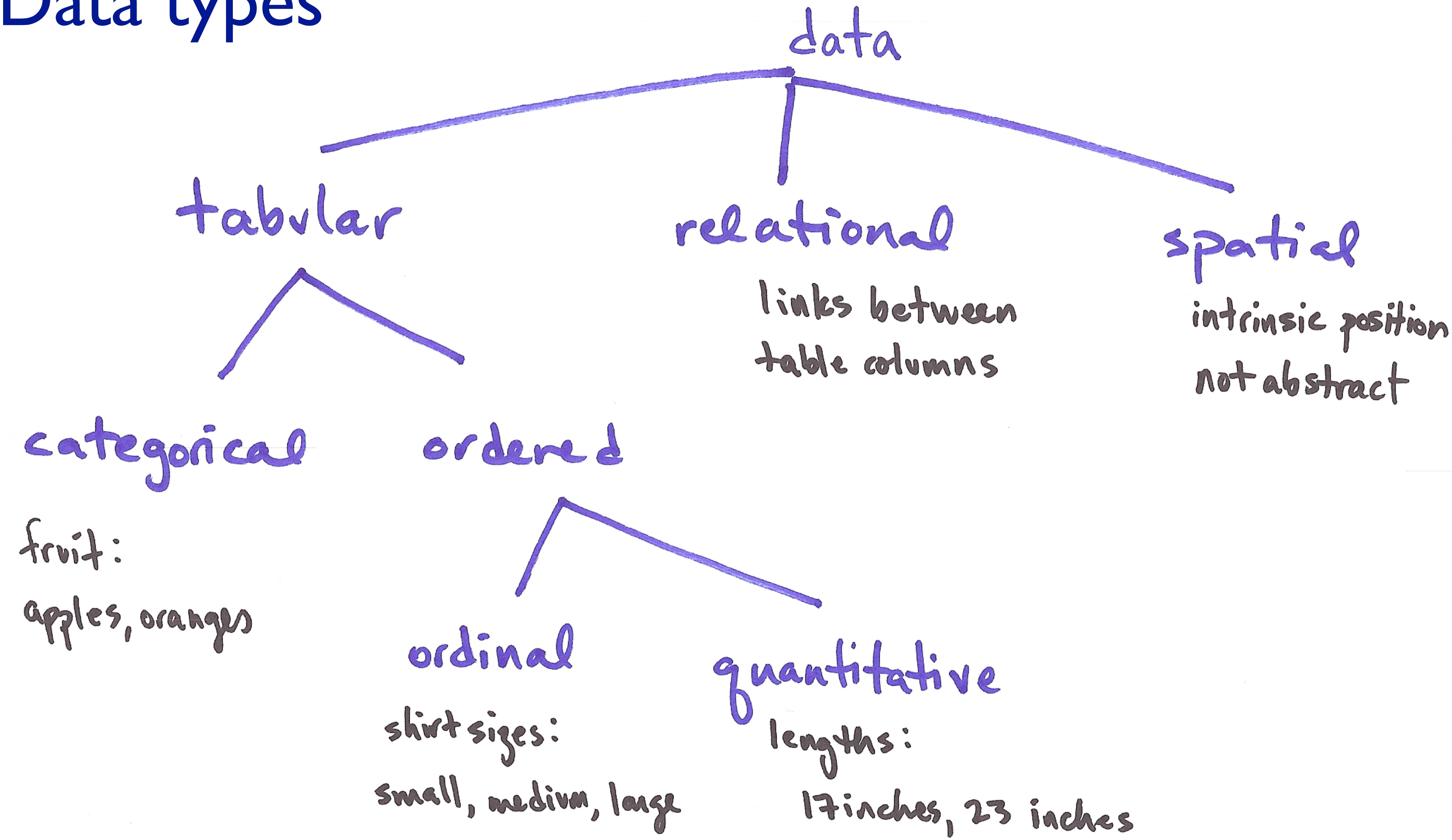
Data types


data
tabular
relational
spatial
categorical
ordered
fruit:
apples, oranges
ordinal
quantitative
shirt sizes:
small, medium, large
lengths:
17inches, 23 inches

Data types


data
tabular
relational
spatial
links between table columns
categorical
ordered
fruit: apples, oranges
ordinal
quantitative
shirt sizes: small, medium, large
lengths: 17inches, 23 inches

Data types


data
tabular
relational
links between table columns
spatial
intrinsic position; not abstract
categorical
fruit: apples, oranges
ordered
ordinal
shirt sizes: small, medium, large
quantitative
lengths: 17 inches, 23 inches

Visual encoding

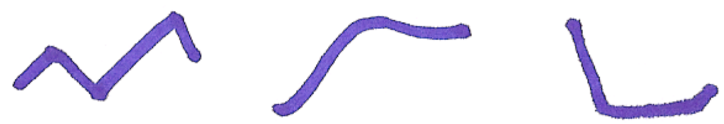
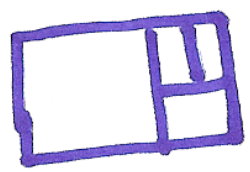
- analyze

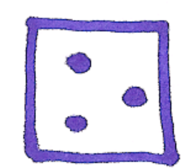
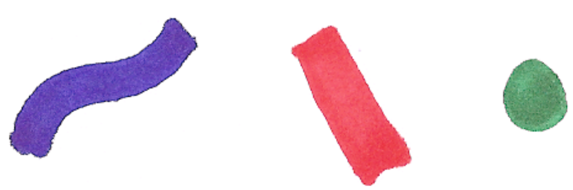
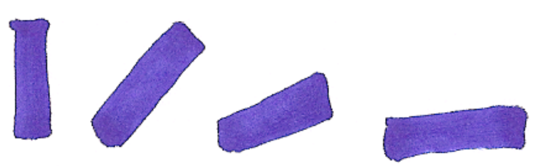
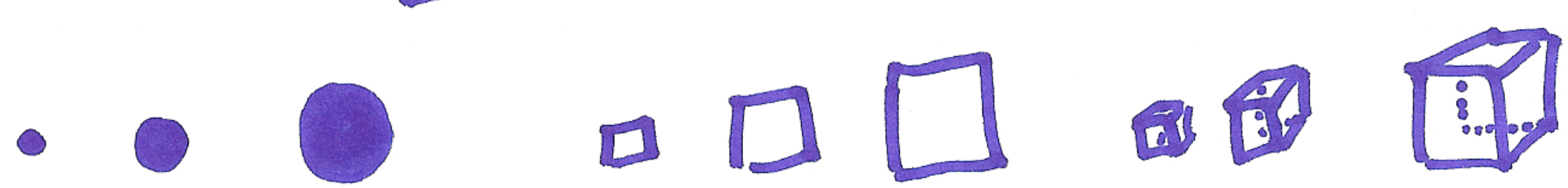
  showing abstract data dimensions

Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions

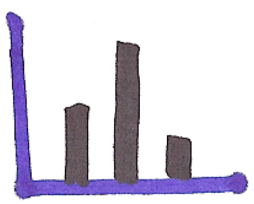
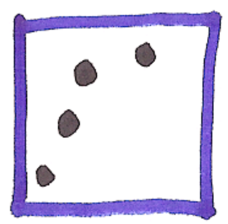
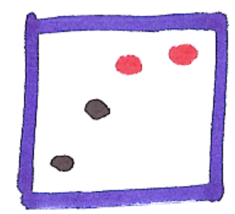
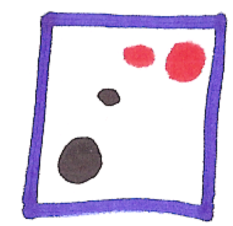
Image theory

- marks: geometric primitives
  - points 
  - lines 
  - areas 

- visual channels: control appearance of marks
  - position
    horizontal , vertical , both 
  - color 
  - shape 
  - tilt 
  - size 

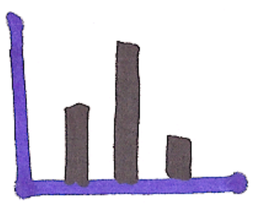
Visual encoding

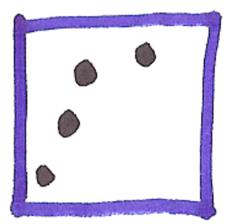
- analyze as combination of marks and channels showing abstract data dimensions

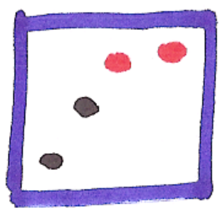





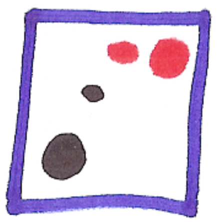
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions







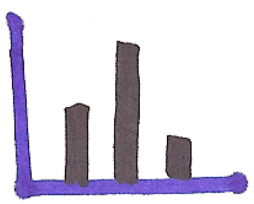


1: vertical position

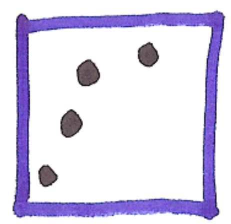
mark: line

Visual encoding

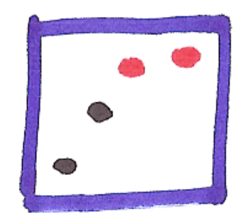
- analyze as combination of marks and channels showing abstract data dimensions

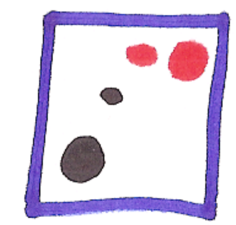


1: vertical position



2: vertical position, horizontal position



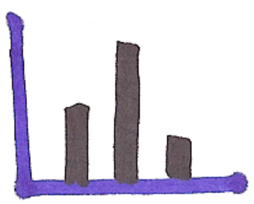


mark: line

mark: point

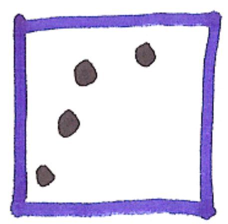
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions



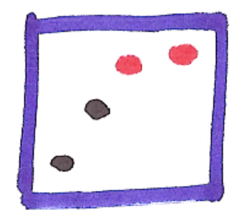
1: vertical position

mark: line



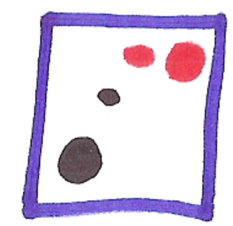
2: vertical position, horizontal position

mark: point



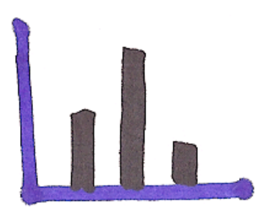
3: vertical position, horizontal position, color

mark: point



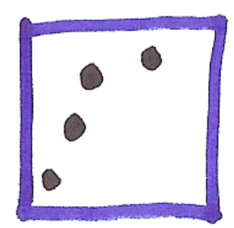
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions



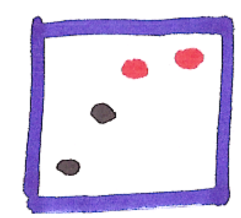
1: vertical position

mark: line



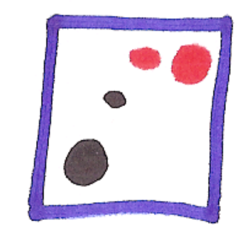
2: vertical position, horizontal position

mark: point



3: vertical position, horizontal position, color

mark: point



4: vertical position, horizontal position, color, size

mark: point

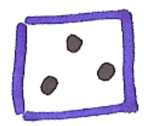
Visual channel types and rankings

what / where

How much

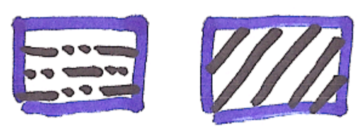
Visual channel types and rankings

what / where

planar position 

color hue 

shape 

stipple pattern 

How much

Visual channel types and rankings

what/where

- planar position
- color hue
- shape
- stipple pattern

How much

- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness black/white
- color saturation
- stipple density

Visual channel types and rankings

Categorical

what/where

- planar position
- color hue
- shape
- stipple pattern

How much

- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness black/white
- color saturation
- stipple density

Visual channel types and rankings

Categorical
what/where

- planar position
- color hue
- shape
- stipple pattern

Ordered: ordinal/Quantitative
How much

- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness black/white
- color saturation
- stipple density

Visual channel types and rankings

Categorical

what/where

- planar position
- color hue
- shape
- stipple pattern

Grouping

- Containment (2D)
- Connection (1D)
- Similarity (other channels)
- Proximity (position)

Ordered: Ordinal/Quantitative

How much

- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness black/white
- color saturation
- stipple density

Visual channel types and rankings

Categorical
what/where

- planar position
- color hue
- shape
- stipple pattern

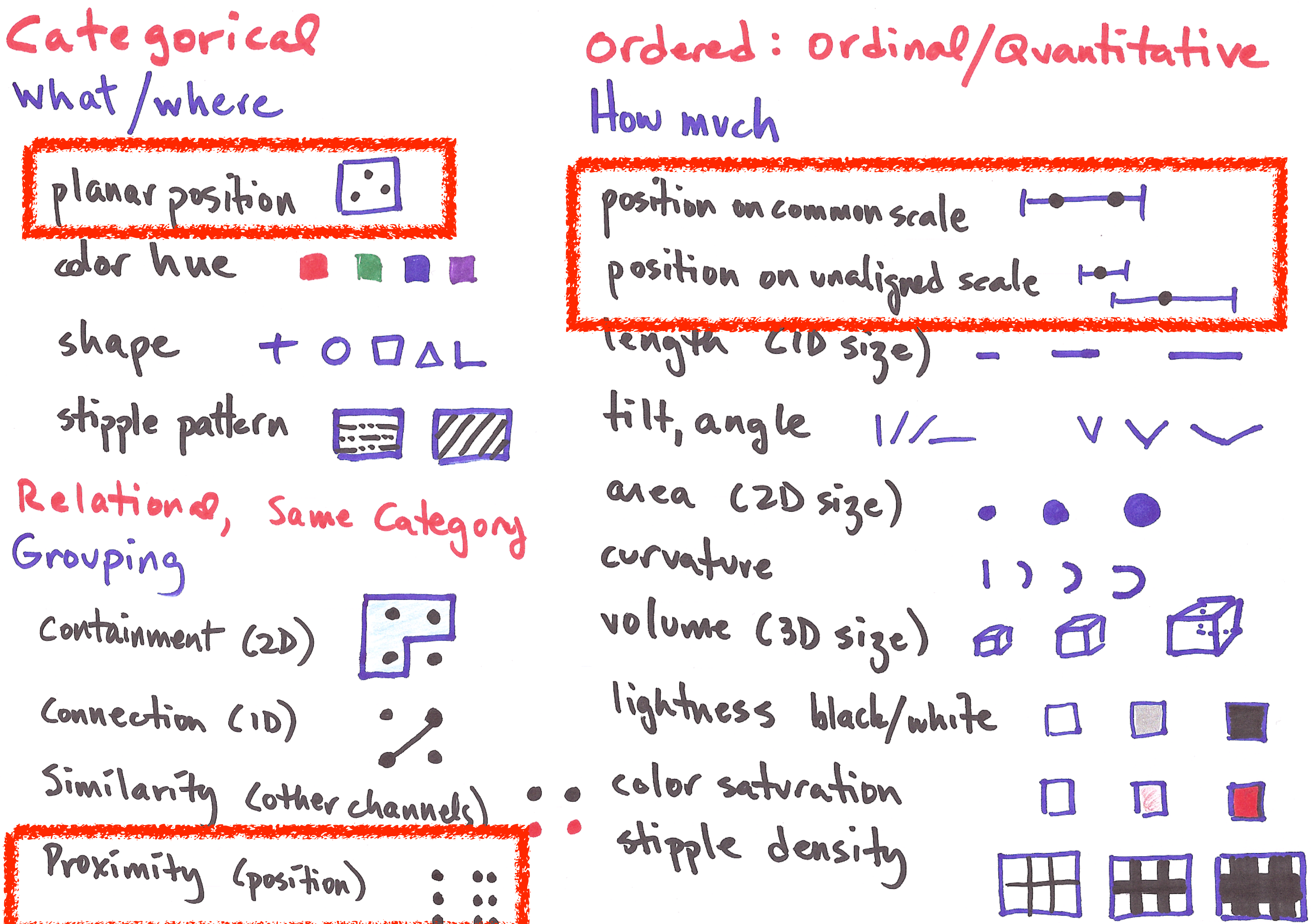
Relational, Same Category
Grouping

- Containment (2D)
- Connection (1D)
- Similarity (other channels)
- Proximity (position)

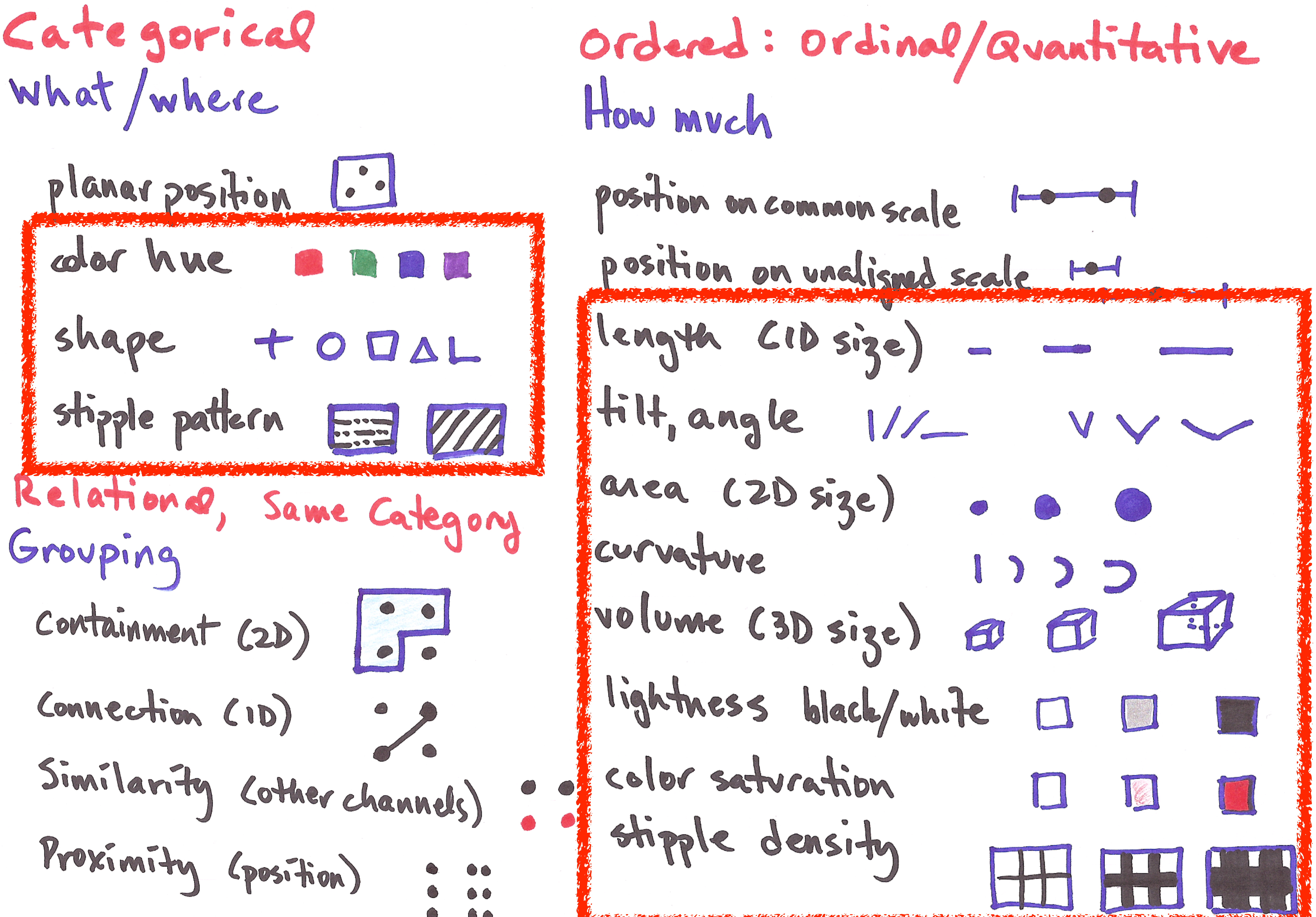
Ordered: Ordinal/Quantitative
How much

- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness black/white
- color saturation
- stipple density

Power of the plane: only position works for all!



Ranking differs for all other channels

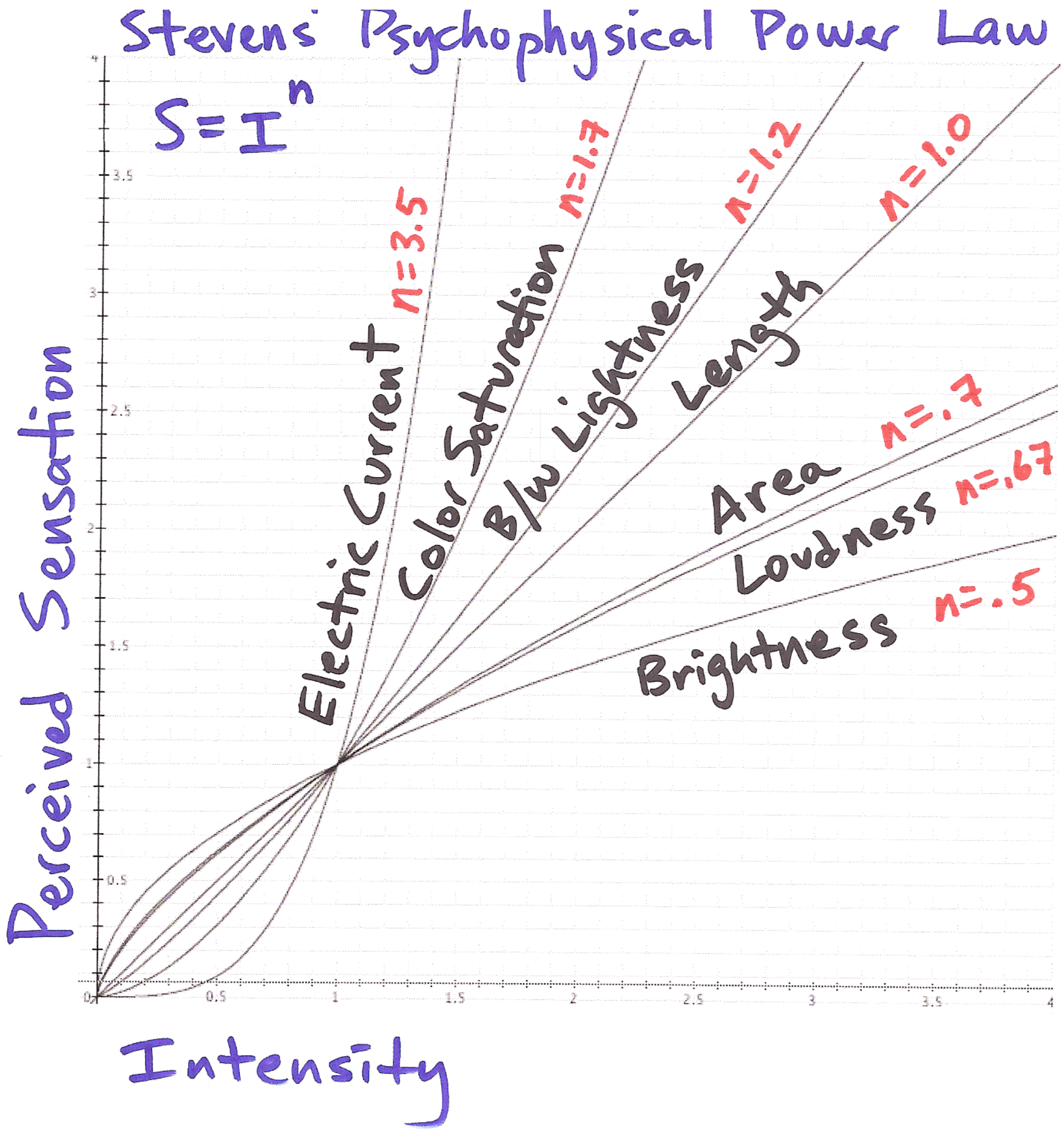


Channel rankings

- effectiveness principle: encode most important attributes with highest ranked channels [Mackinlay 86]

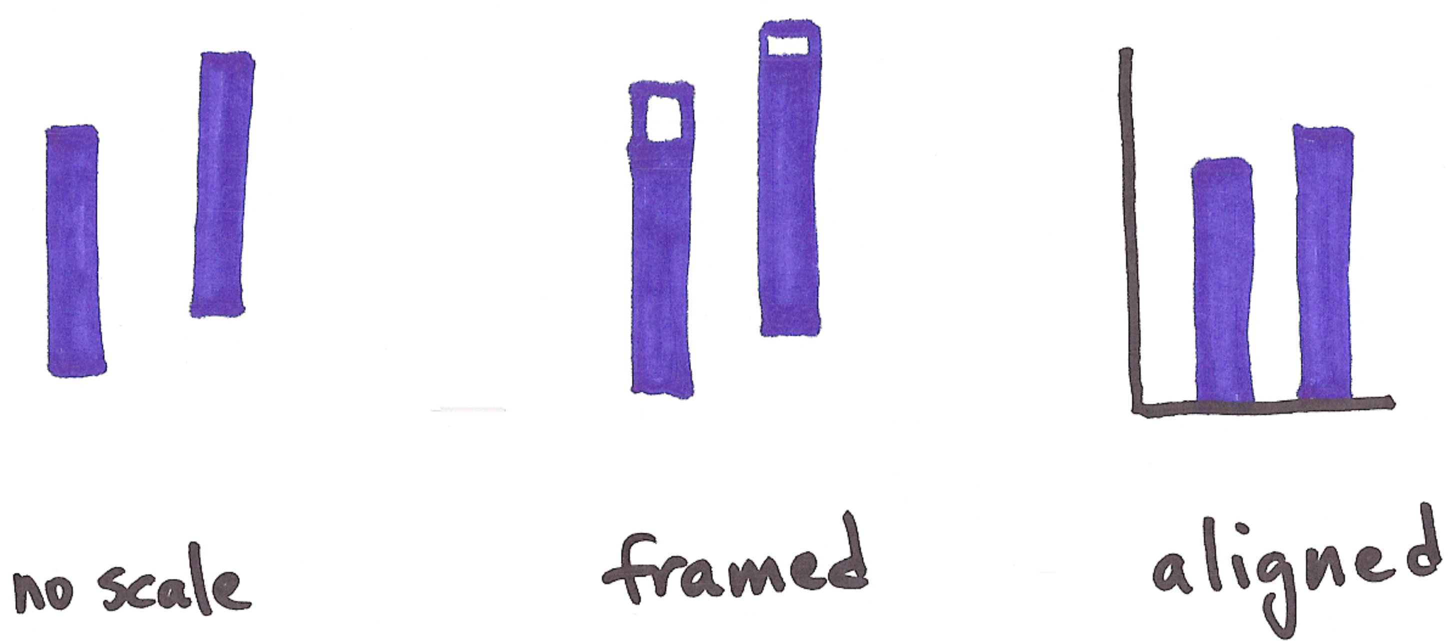
- where do rankings come from?
  - accuracy, discriminability, separability, popout

Accuracy


Stevens' Psychophysical Power Law
S=I^n
Perceived Sensation
Intensity
Electric Current n=3.5
Color Saturation n=1.7
B/w Lightness n=1.2
Length n=1.0
Area n=.7
Loudness n=.67
Brightness n=.5
0
0.5
1
1.5
2
2.5
3
3.5
4

Accuracy

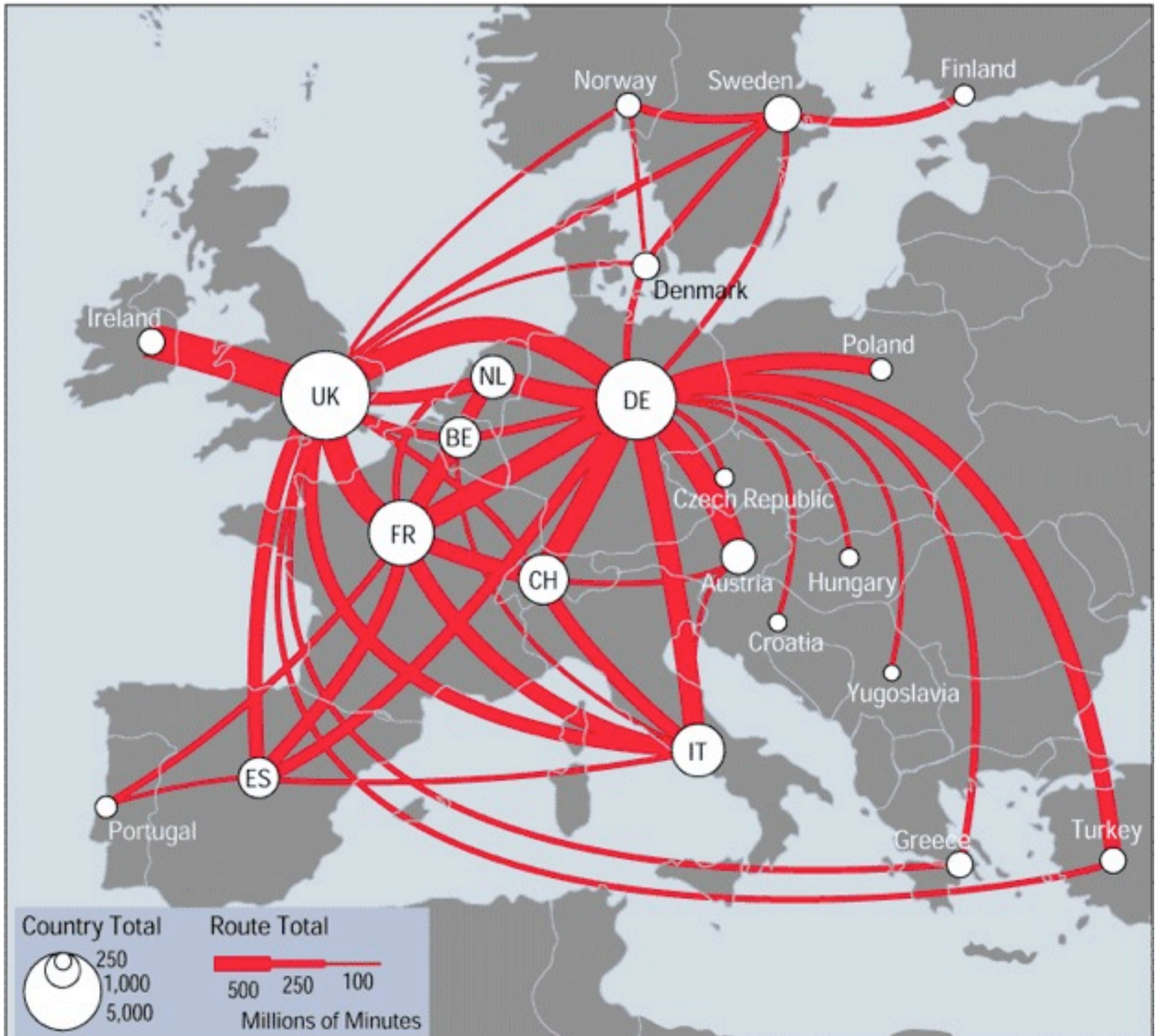
- position along common scale



  - frame increases accuracy [Cleveland 84]
  - Weber's Law: relative judgements
    - filled rectangles differ by 1:9
    - white rectangles differ by 1:2

Discriminability: How many usable steps?

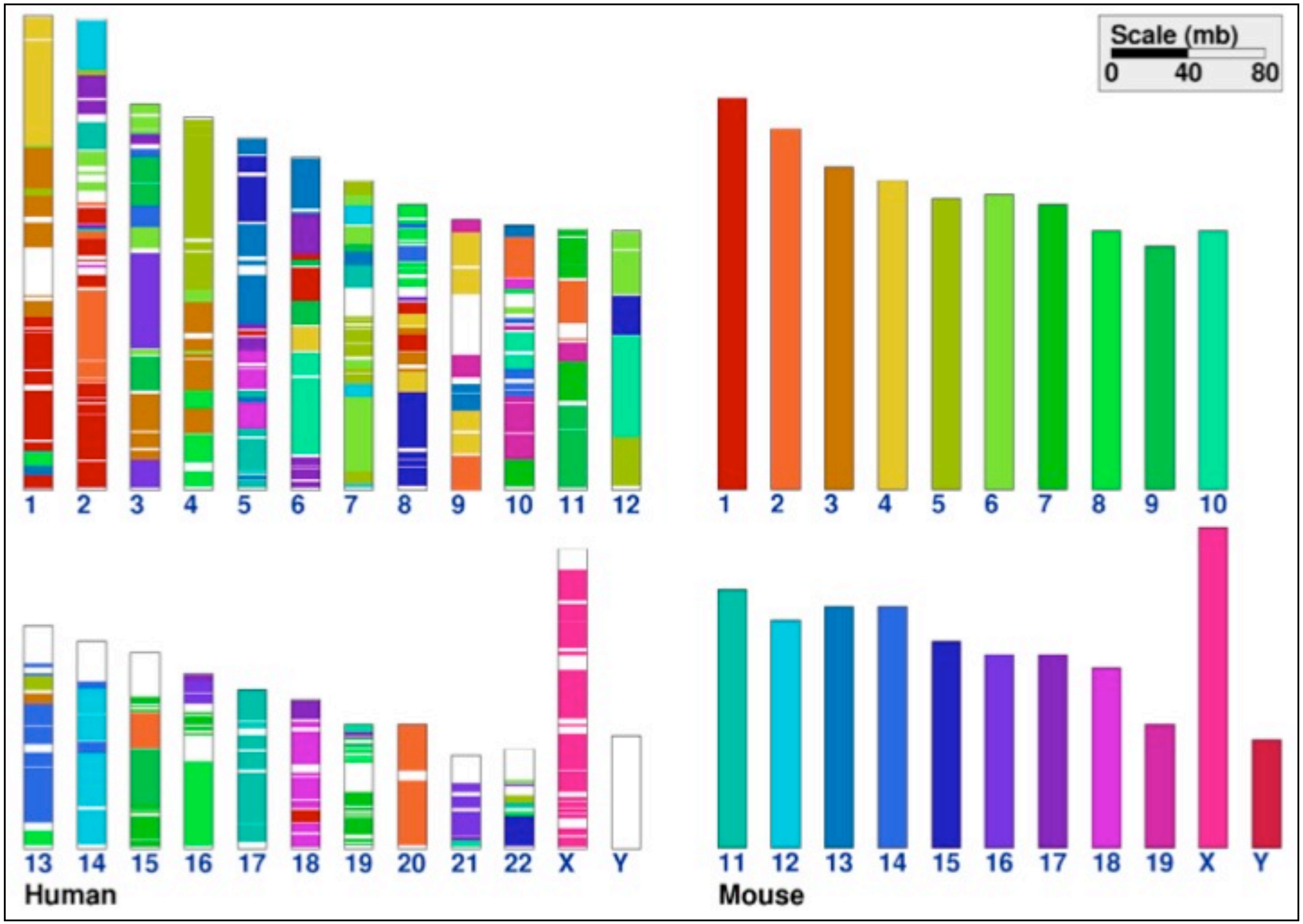
- linewidth: only a few



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

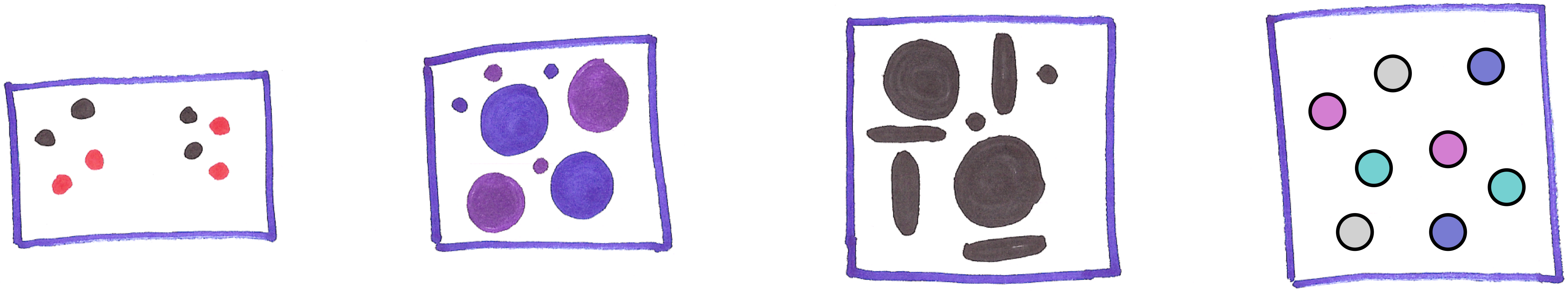
Discriminability: Categorical color constraints

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

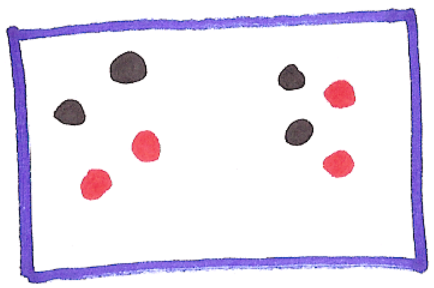


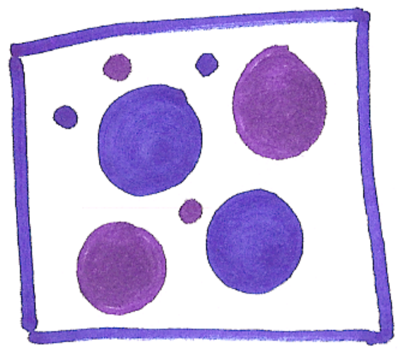
Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. Bioinformatics 2007

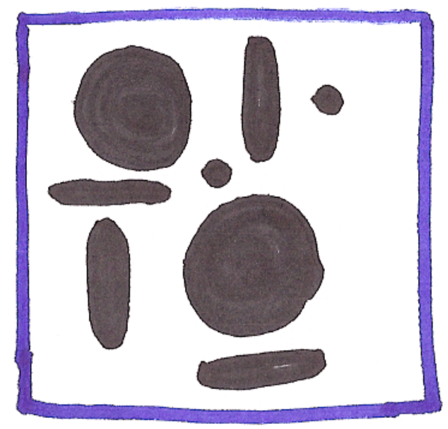
Separability vs. integrality

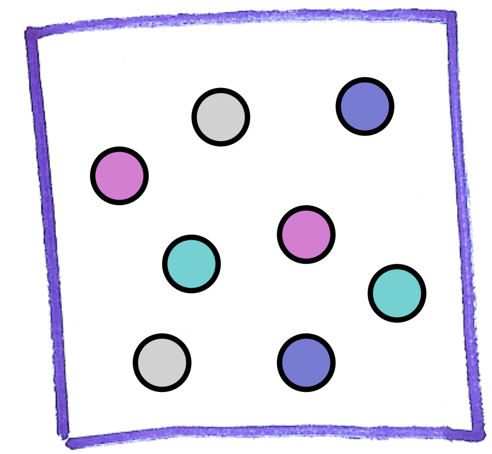


Separability vs. integrality







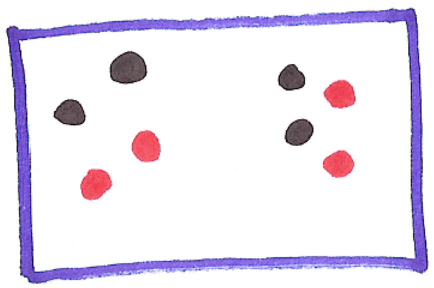


position
hue (color)

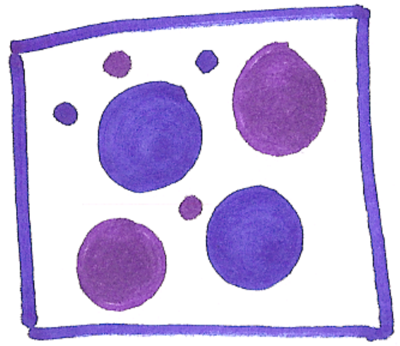
fully separable

2 groups each

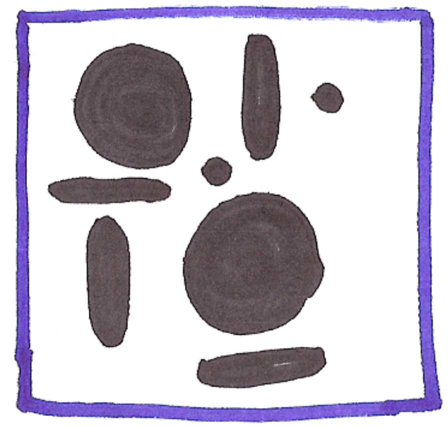
Separability vs. integrality

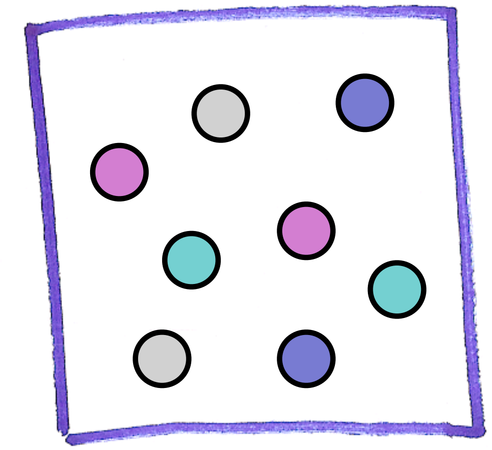


position
hue (color)



size
hue (color)





fully separable

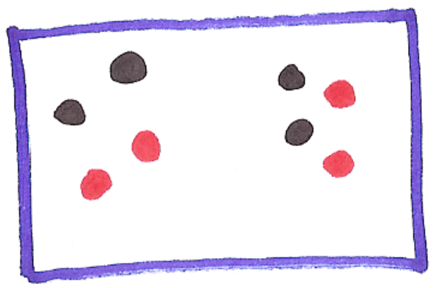
some interference

difficult to discriminate small items

2 groups each

{2 groups each}

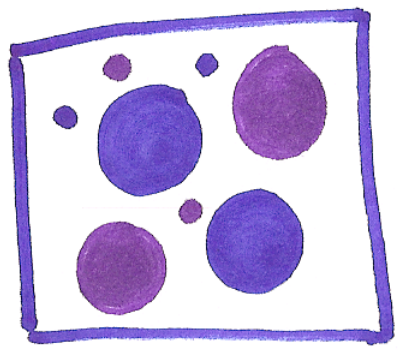
Separability vs. integrality



position
hue (color)

fully separable

2 groups each

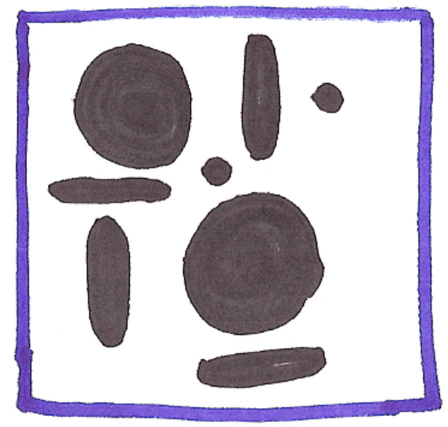


size
hue (color)

some interference

difficult to discriminate small items

(2 groups each)

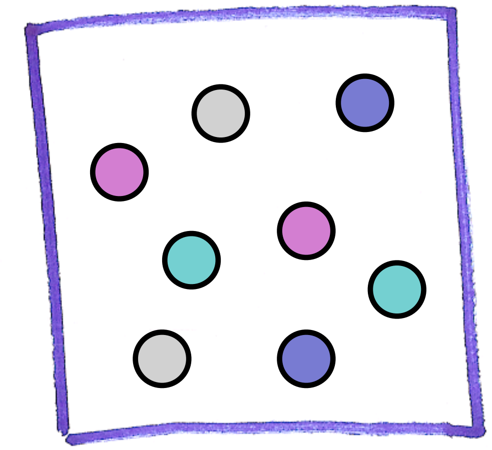


size: width
size: height

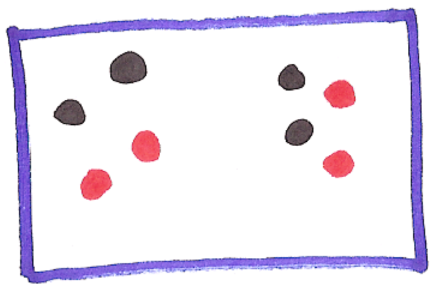
some/significant interference

integral percept: area (planar size)

3 groups



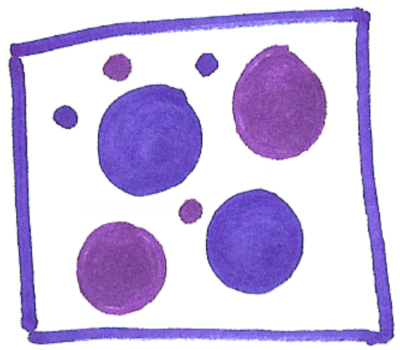
Separability vs. integrality



position
hue (color)

fully separable

2 groups each

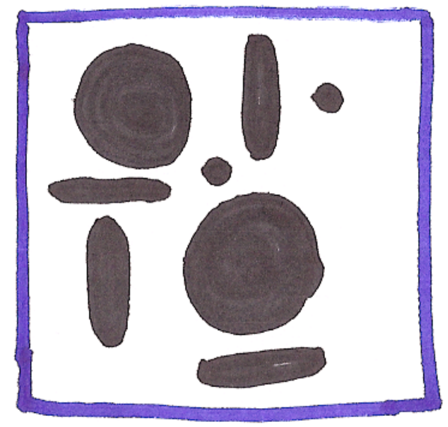


size
hue (color)

some interference

difficult to discriminate small items

{2 groups each}

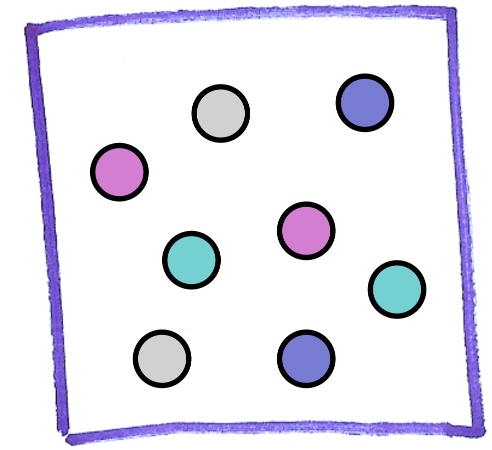


size: width
size: height

some/significant interference

integral percept: area (planar size)

3 groups



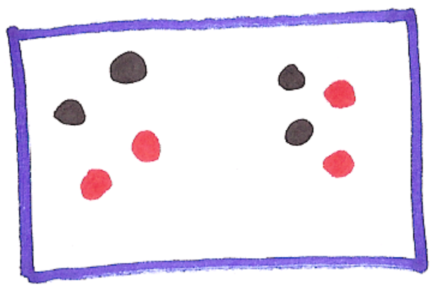
red
green

major interference

integral percept: color/hue

4 groups

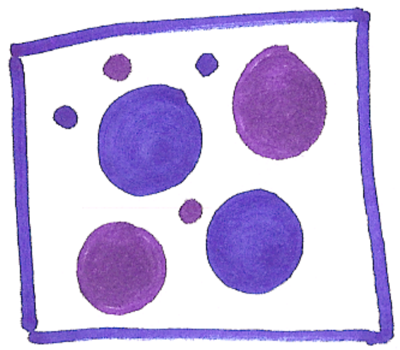
Separability vs. integrality



position
hue (color)

fully separable

2 groups each

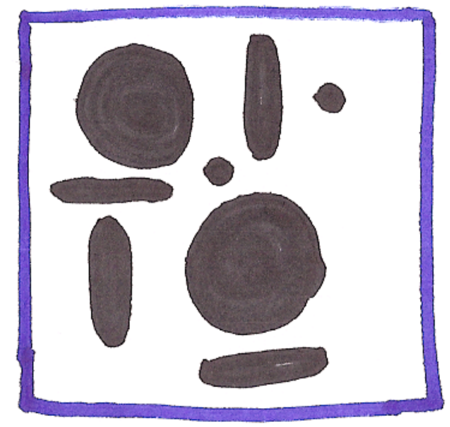


size
hue (color)

some interference

difficult to discriminate small items

{2 groups each}

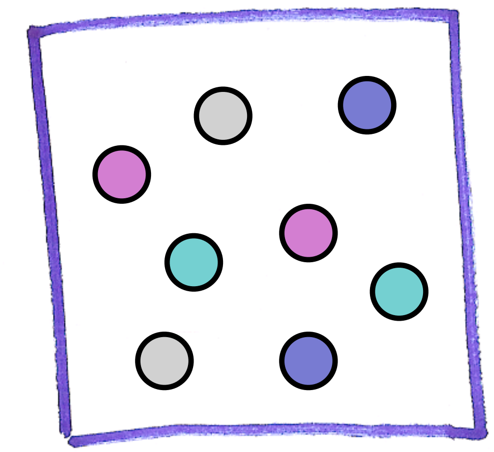


size: width
size: height

some/significant interference

integral percept:
area
(planar size)

3 groups



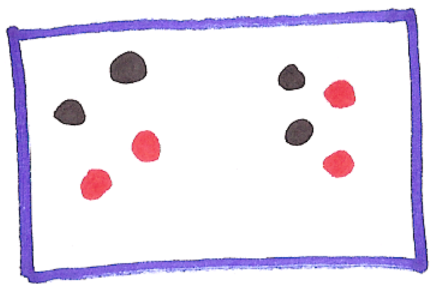
red
green

major interference

integral percept:
color/hue

4 groups

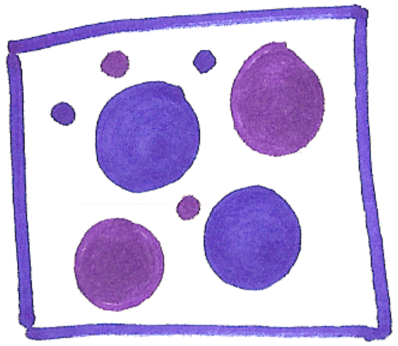
Separability vs. integrality



position
hue (color)

fully separable

2 groups each

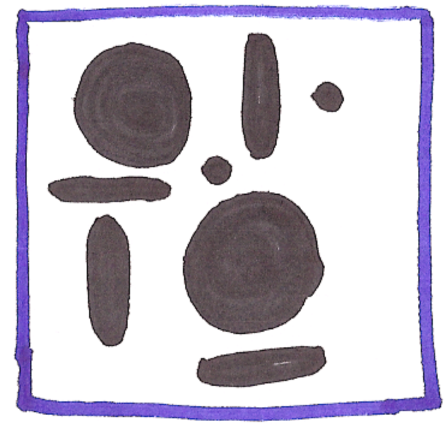


size
hue (color)

some interference

difficult to discriminate small items

2 groups each

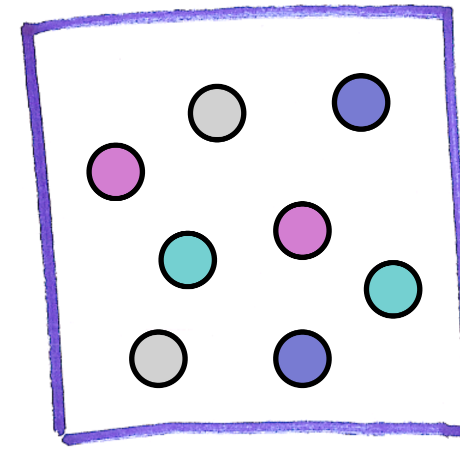


size: width
size: height

some/significant interference

integral percept: area (planar size)

3 groups



red
green

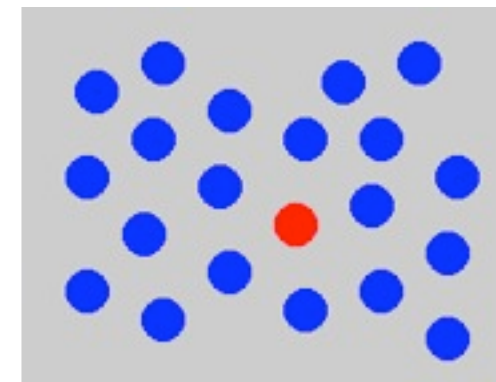
major interference

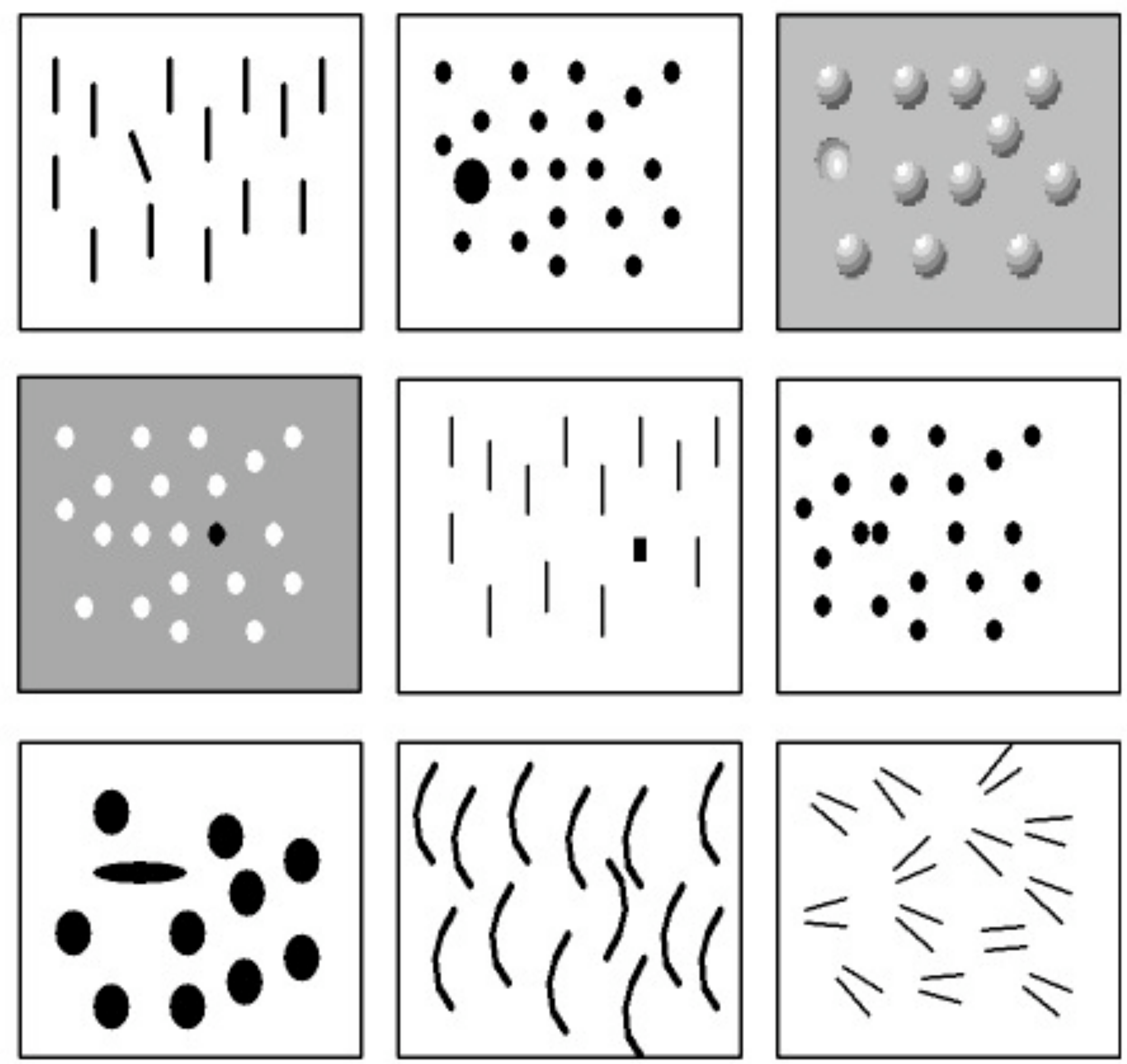
integral percept: color/hue

4 groups

Popout: Most channels

- parallel processing on most channels
  - sufficiently different item noticed immediately, independent of distractor count



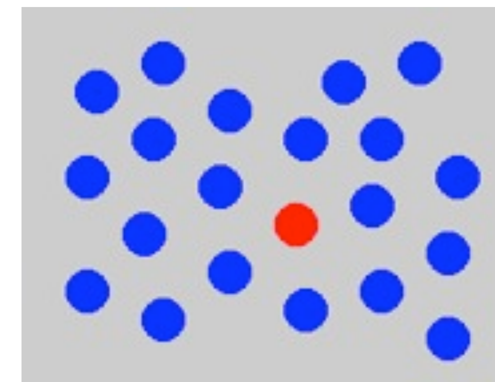


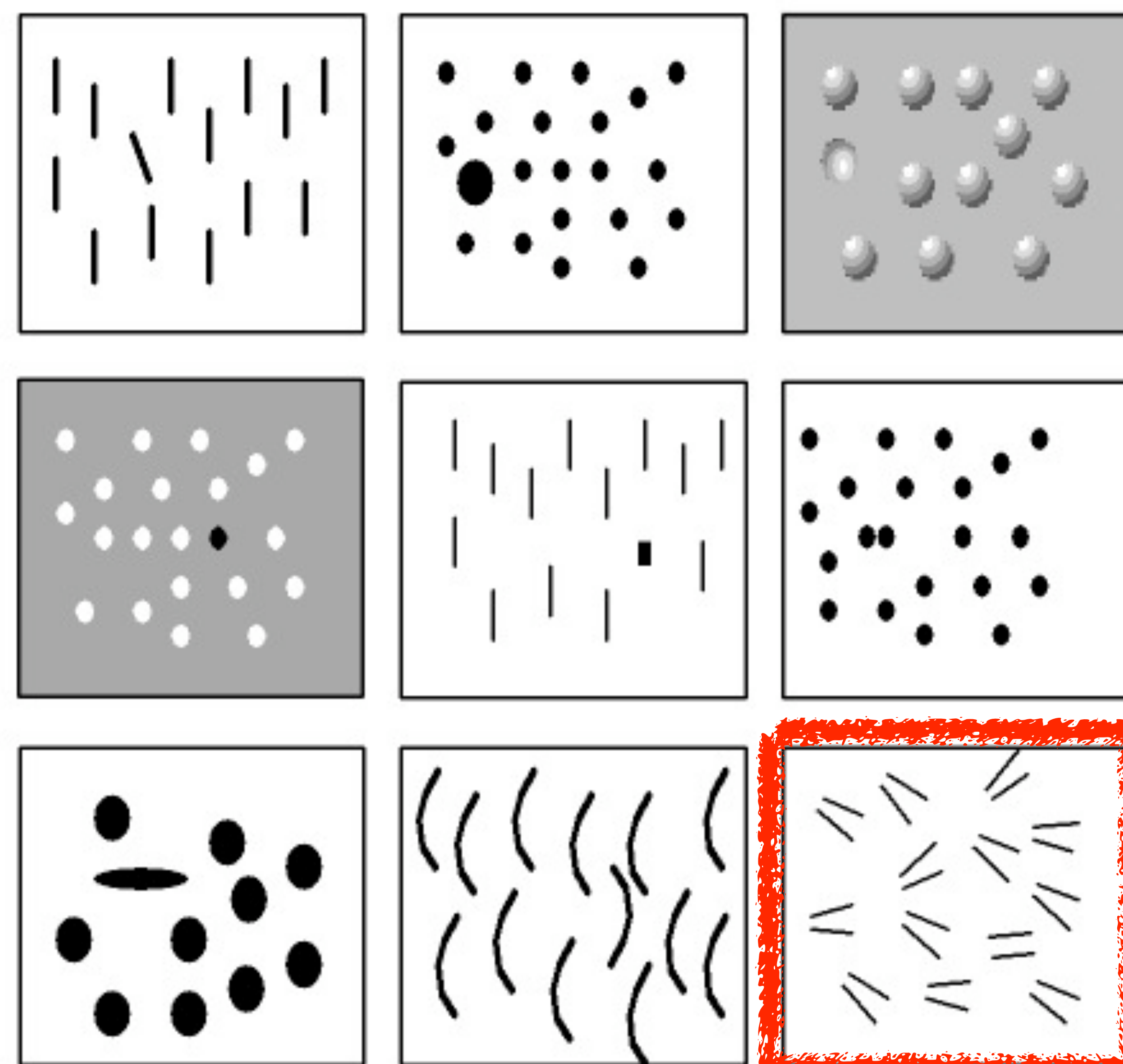
- some channels have no popout: serial search required

Healey. Perception in Visualization
http://www.csc.ncsu.edu/faculty/healey/PP/

Popout: Most channels

- parallel processing on most channels
  - sufficiently different item noticed immediately, independent of distractor count



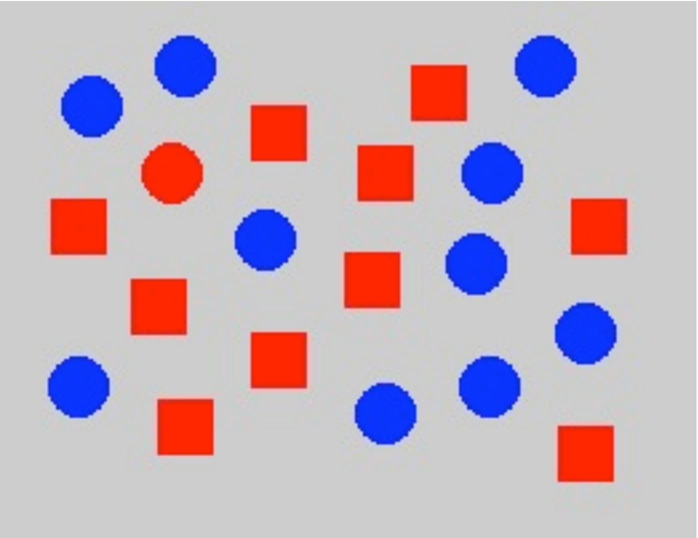


- some channels have no popout: serial search required

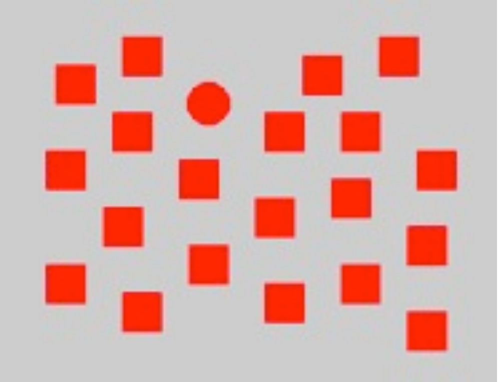
Healey. Perception in Visualization
http://www.csc.ncsu.edu/faculty/healey/PP/

Popout limits

- only one channel at a time
  - combination searches are serial
    - most channel pairs
    - all channel triplets, etc



- within channel, speed depends on which channel and how different item is from surroundings
  - 'sufficiently different': context dependent

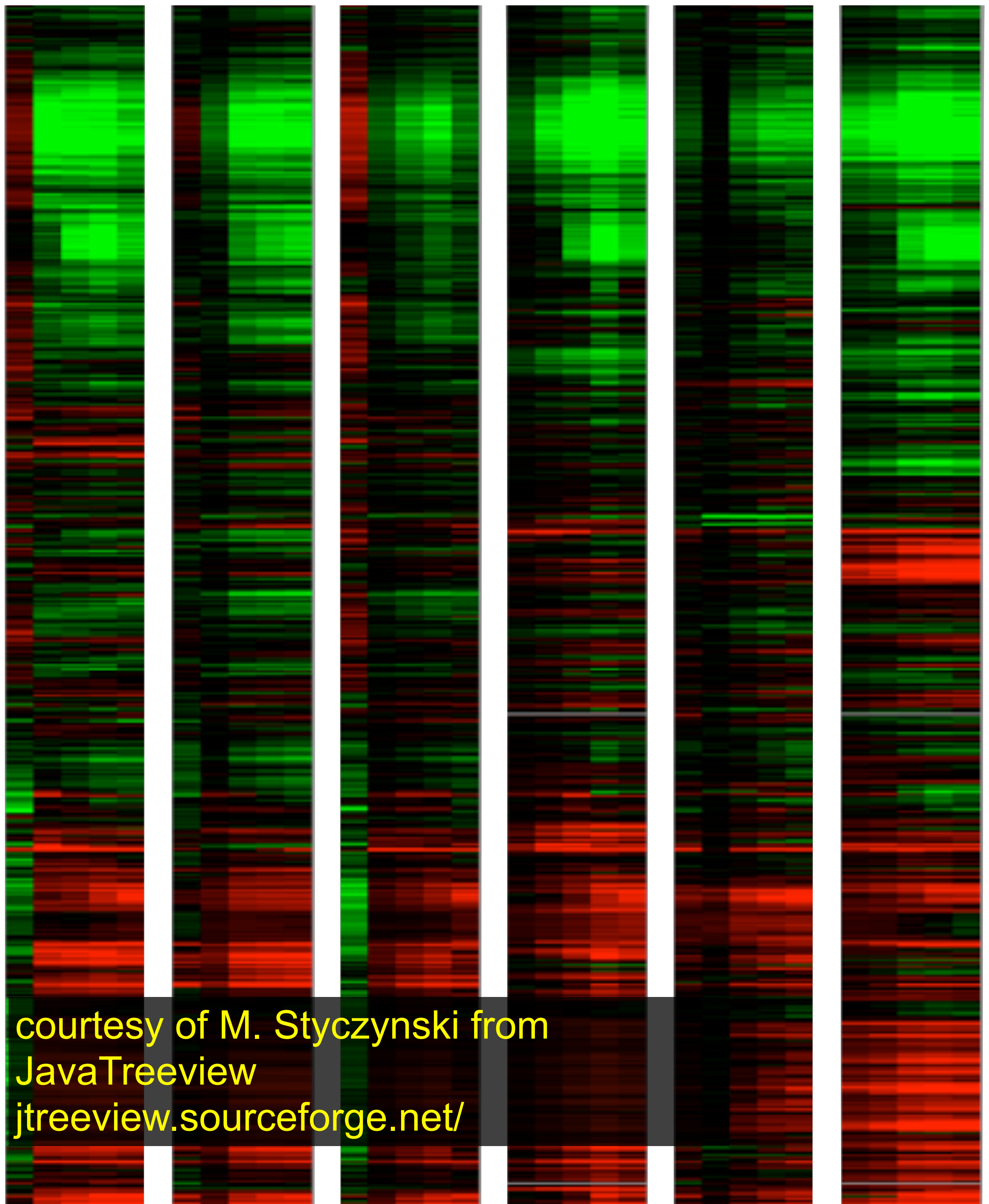


Healey. Perception in Visualization
http://www.csc.ncsu.edu/faculty/healey/PP/

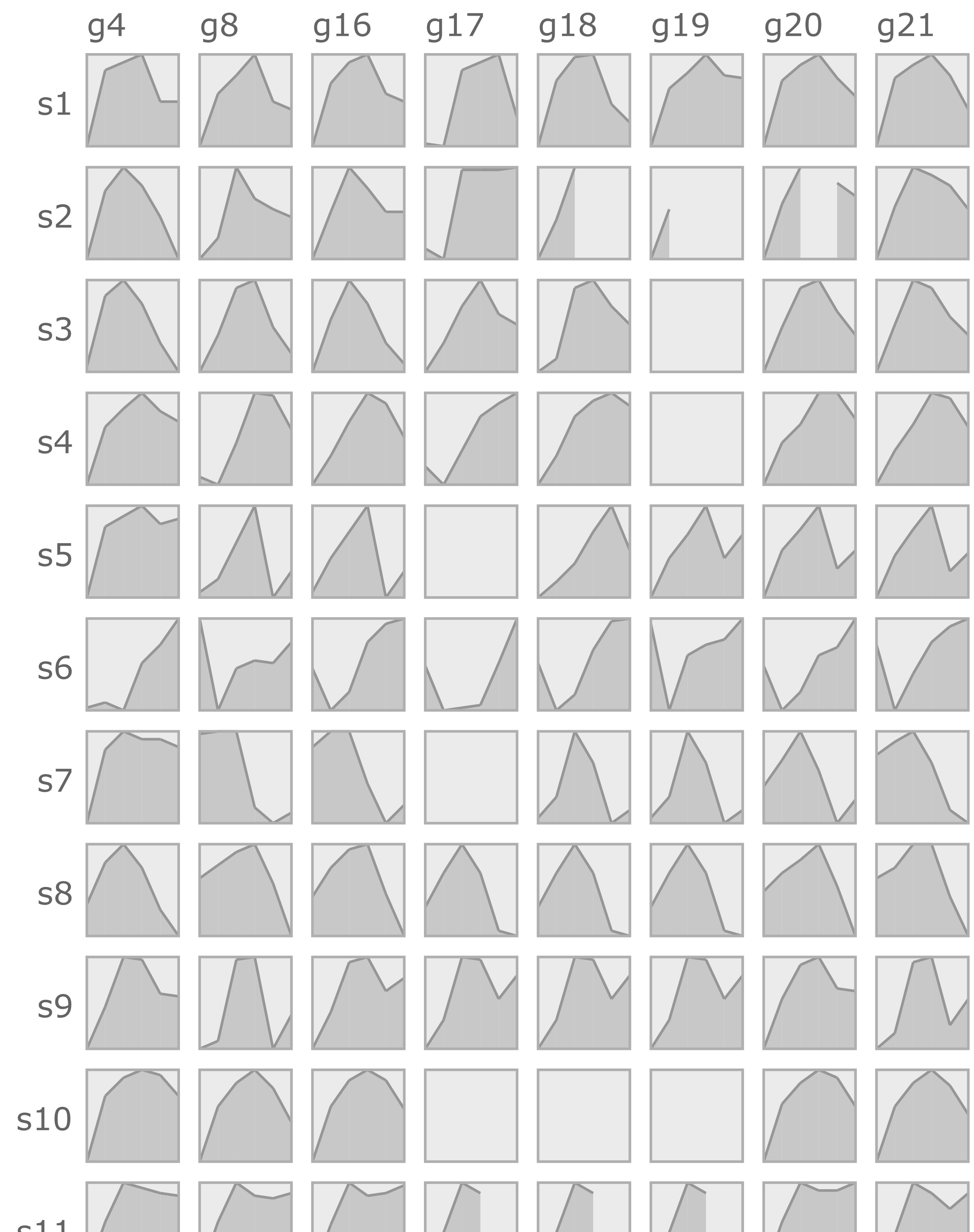
Encoding example: Heatmaps vs. curvemaps

- color traditional, but spatial position outranks it

heatmap

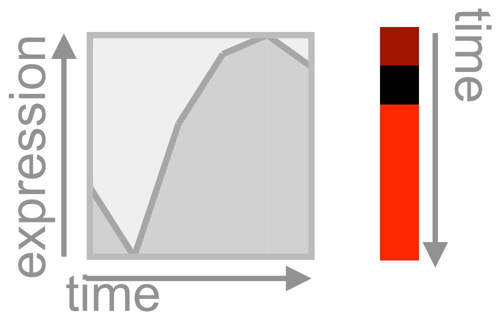


curvemap



Curvemap

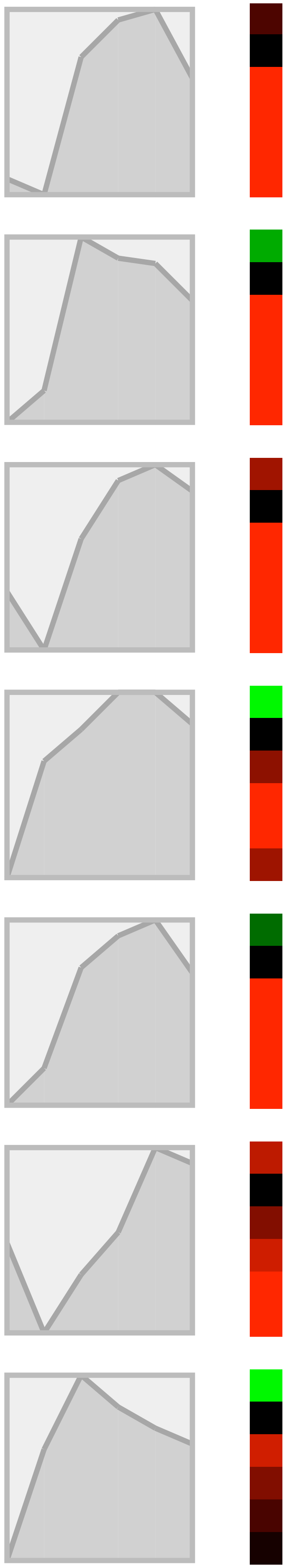
- shape perception easier for filled framed line charts than colored boxes



Pathline: A Tool for Comparative Functional Genomics. Meyer, Wong, Styczynski, Munzner, Pfister. EuroVis 2010.

Curvemap

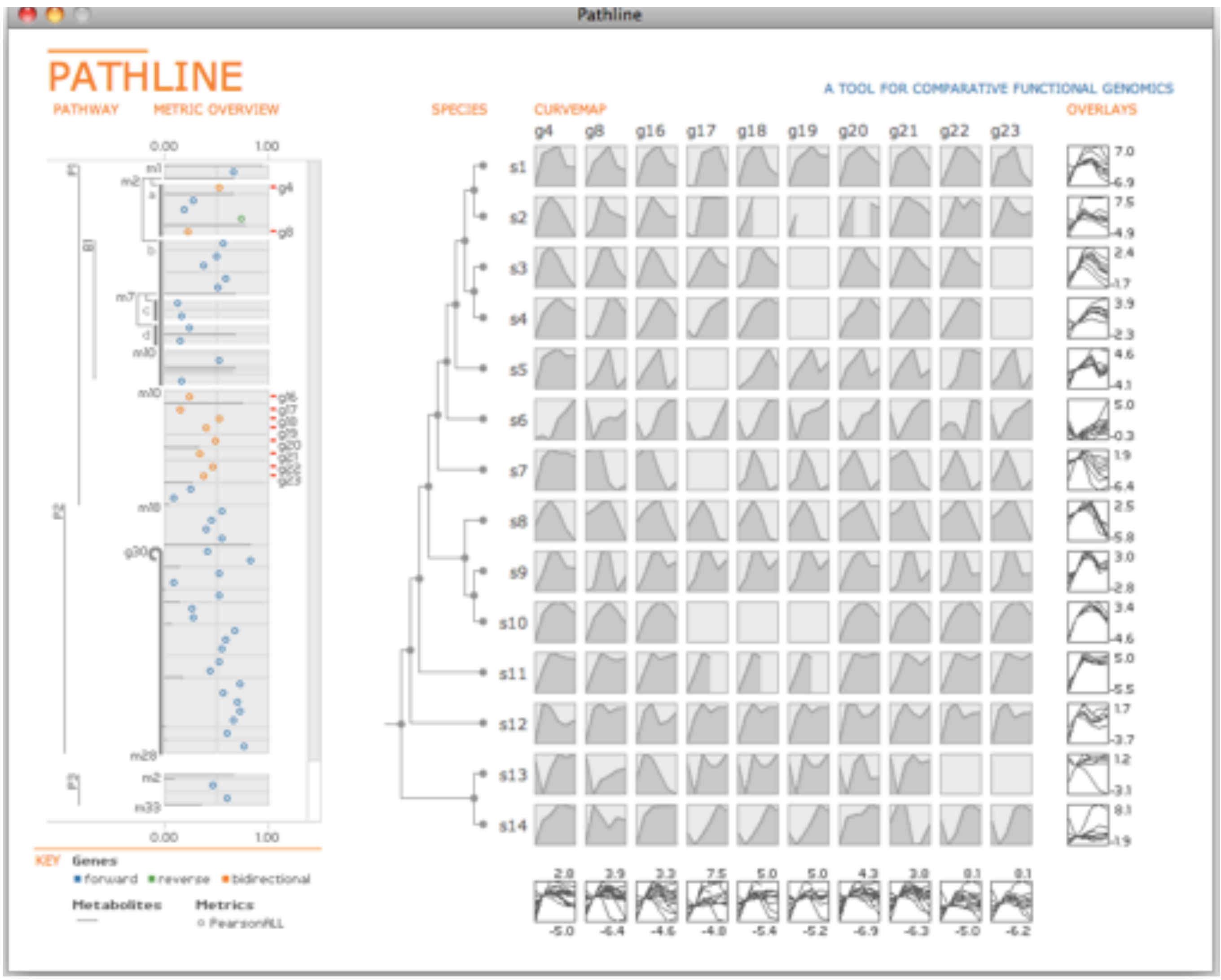
- shape perception easier for filled framed line charts than colored boxes



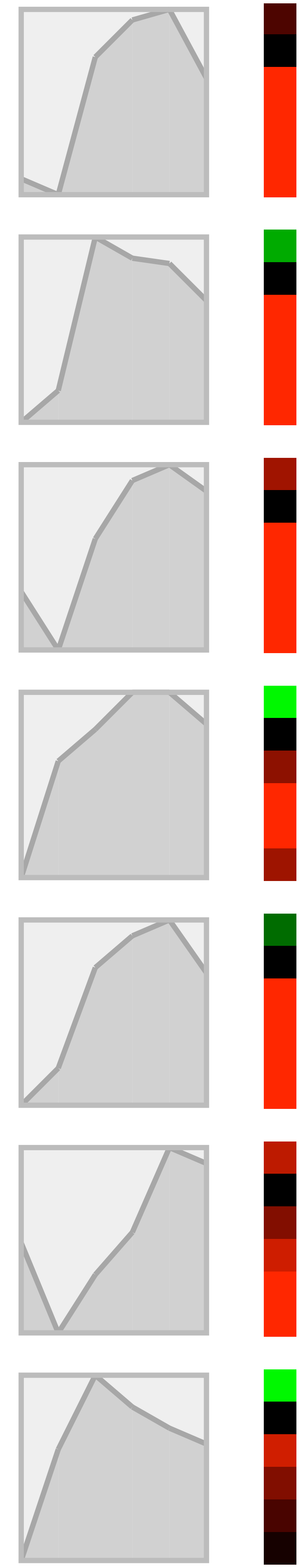
Pathline: A Tool for Comparative Functional Genomics. Meyer, Wong, Styczynski, Munzner, Pfister. EuroVis 2010.

Curvemap

- shape perception easier for filled framed line charts than colored boxes

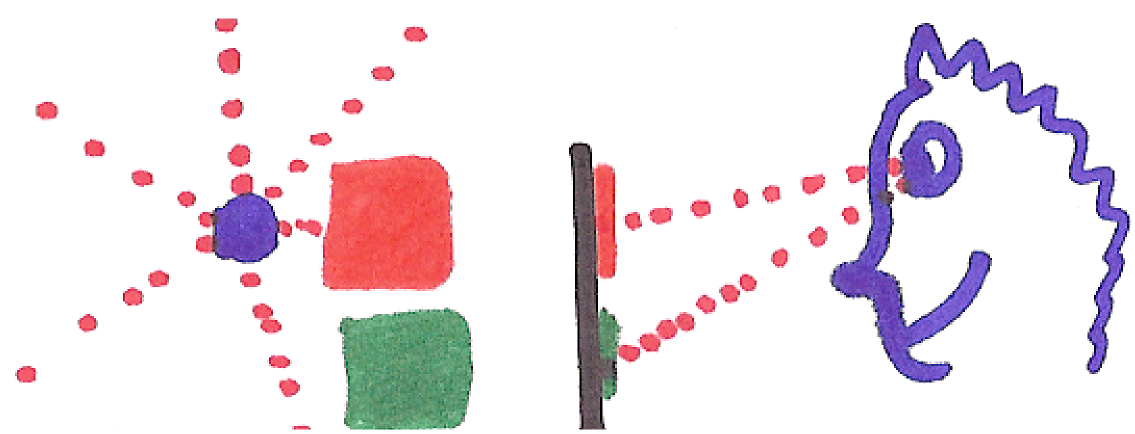


Pathline: A Tool for Comparative Functional Genomics. Meyer, Wong, Styczynski, Munzner, Pfister. EuroVis 2010.



Dangers of depth

- rankings for **planar** spatial position, not depth!
- we don't really live in 3D: we **see** in 2.05D
  - up/down and sideways: image plane
    - acquire more info quickly from eye movements
  - away: depth into scene
    - only acquire more info from head/body motion

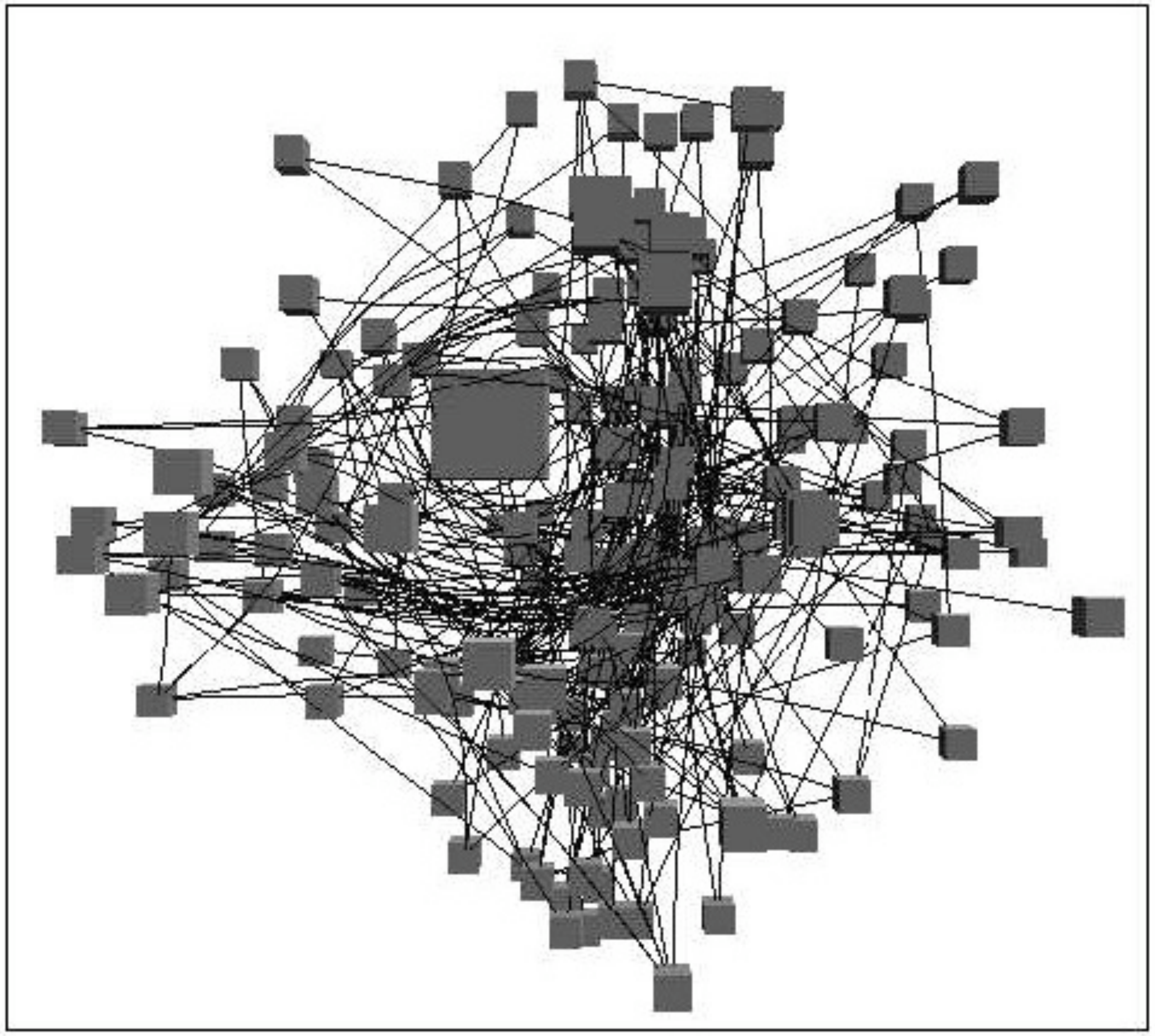


- further reading

  Visual Thinking for Design (Chap 5). Colin Ware. 2008

Dangers of depth: difficulties of 3D

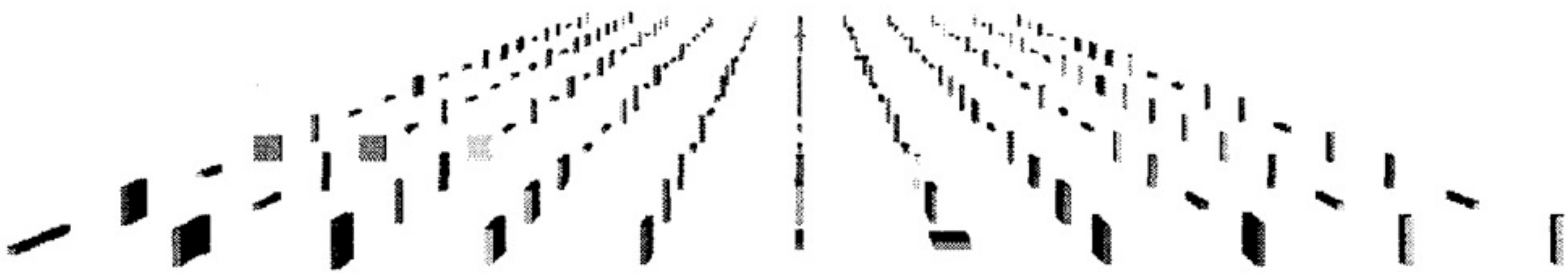
- occlusion
- interaction complexity



Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis 1996.

Dangers of depth: difficulties of 3D

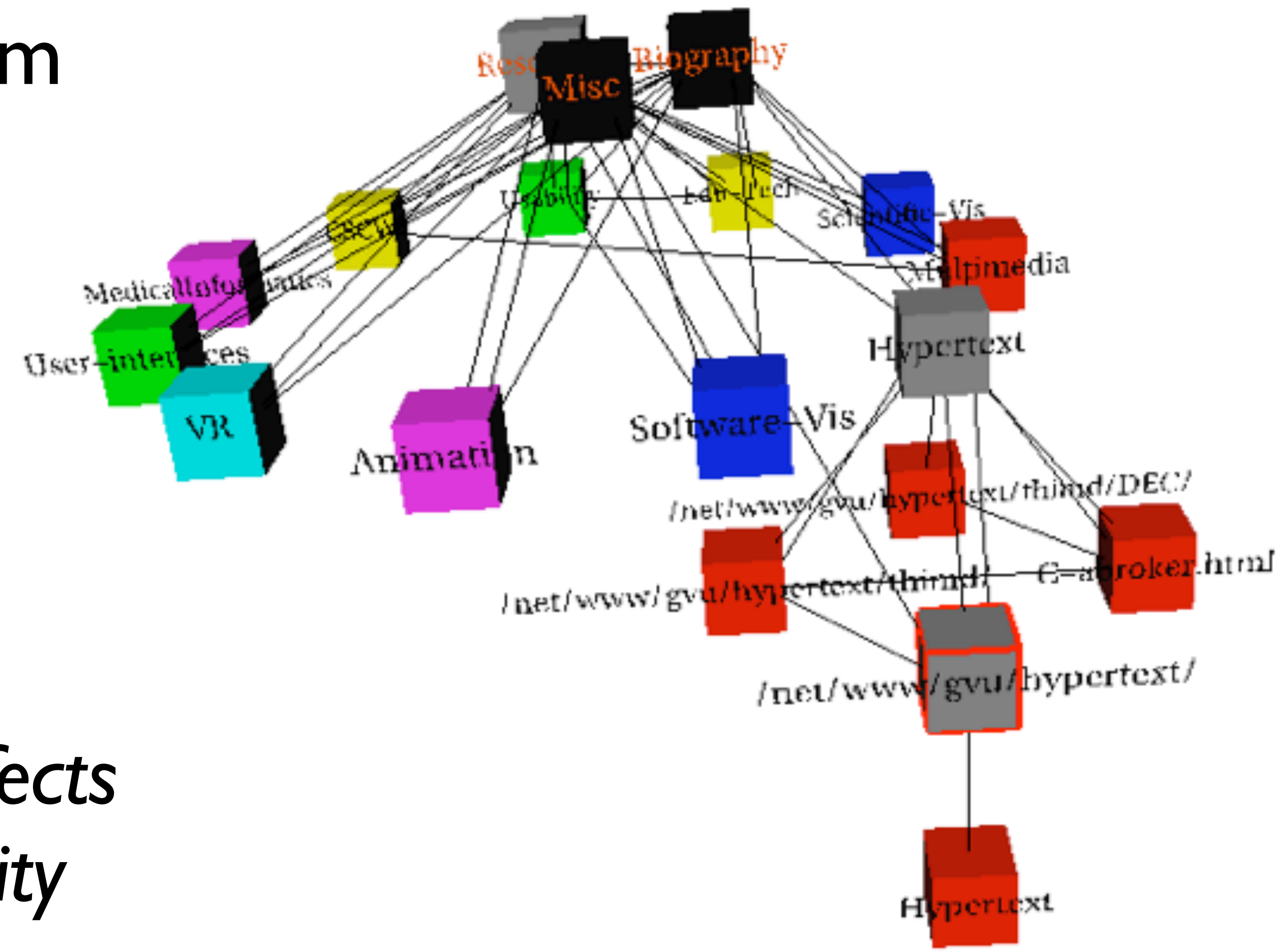
- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!



Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96

Dangers of depth: difficulties of 3D

- text legibility
  - far worse when tilted from image plane



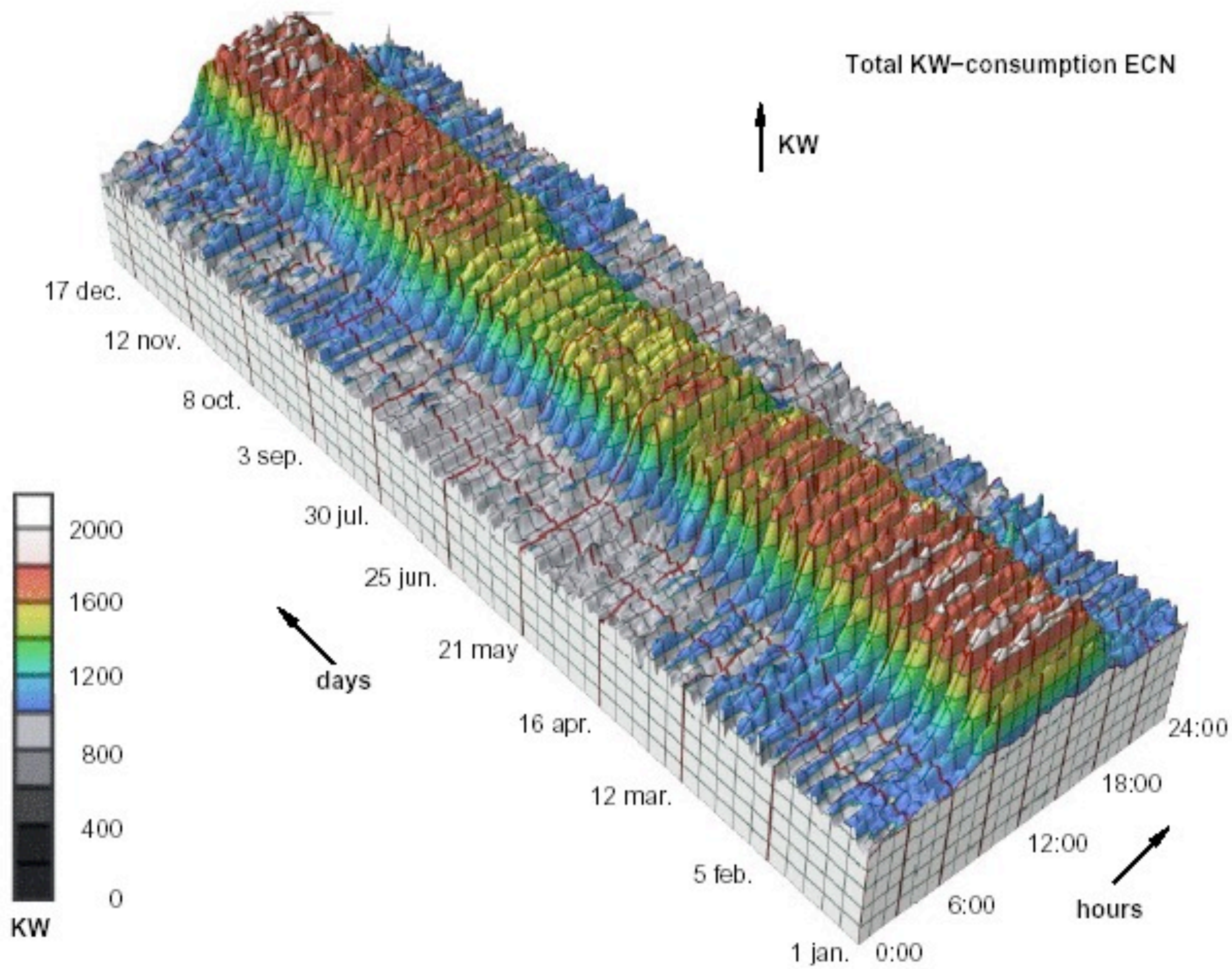
- further reading

  Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. Grossman et al. CHI 2007

Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems, 1995.

Dangers of depth example

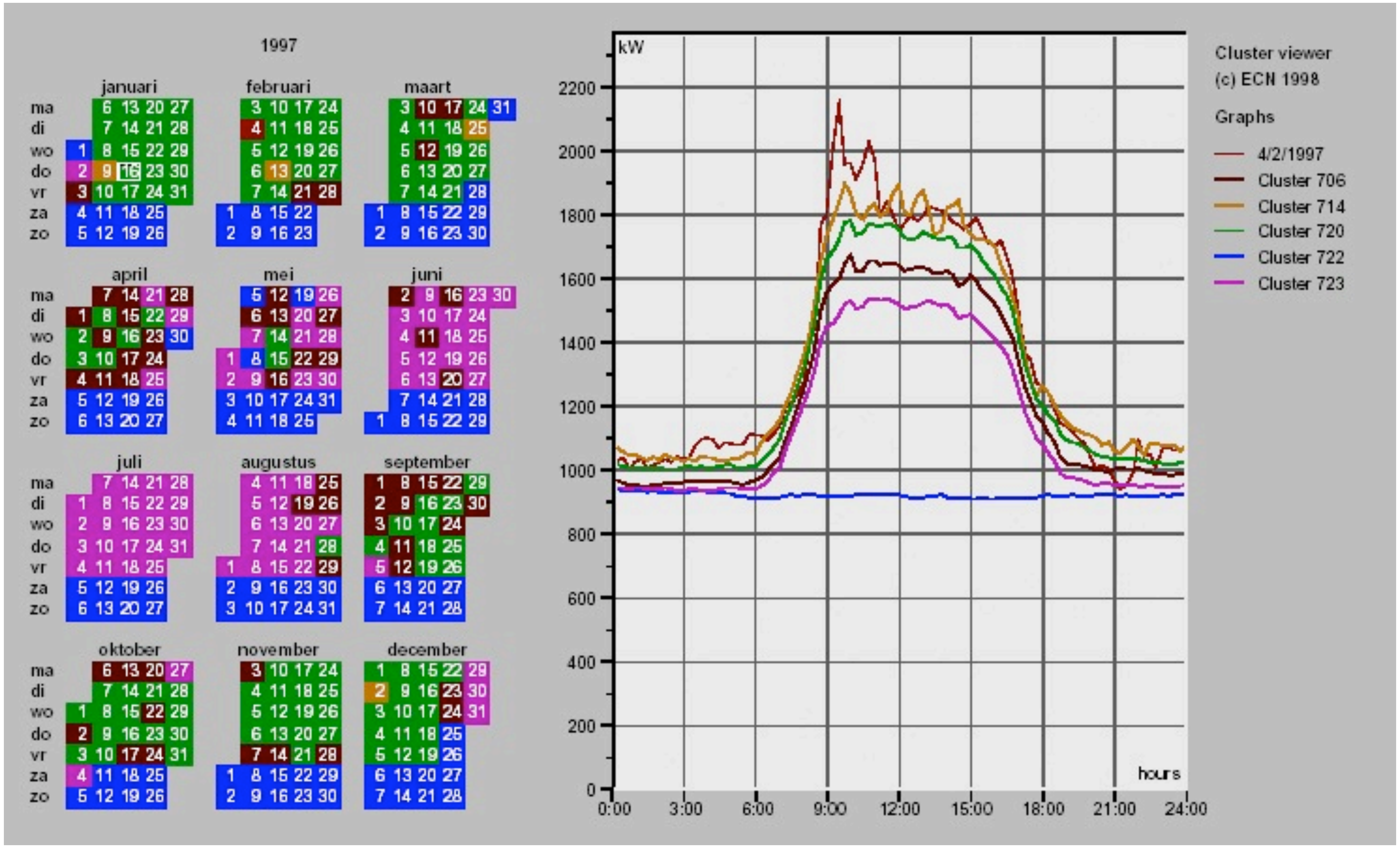
- extruded curves: detailed comparisons impossible



Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc InfoVis 99.

Transformation to suitable abstraction

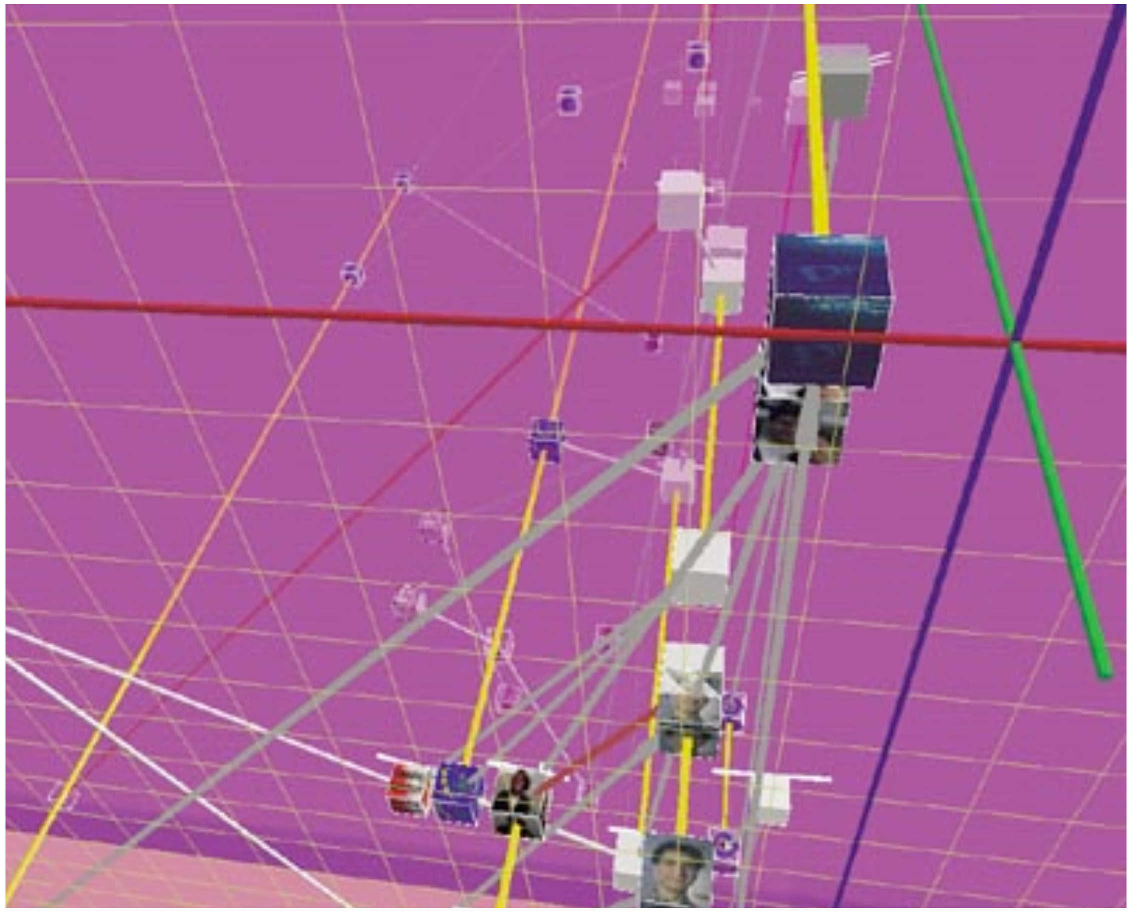
- derived data: clusters
- multiple views: calendar, superimposed 2D curves



Cluster and Calendar based Visualization of Time Series Data.
van Wijk and van Selow, Proc InfoVis 99.

Dangers of depth: must justify

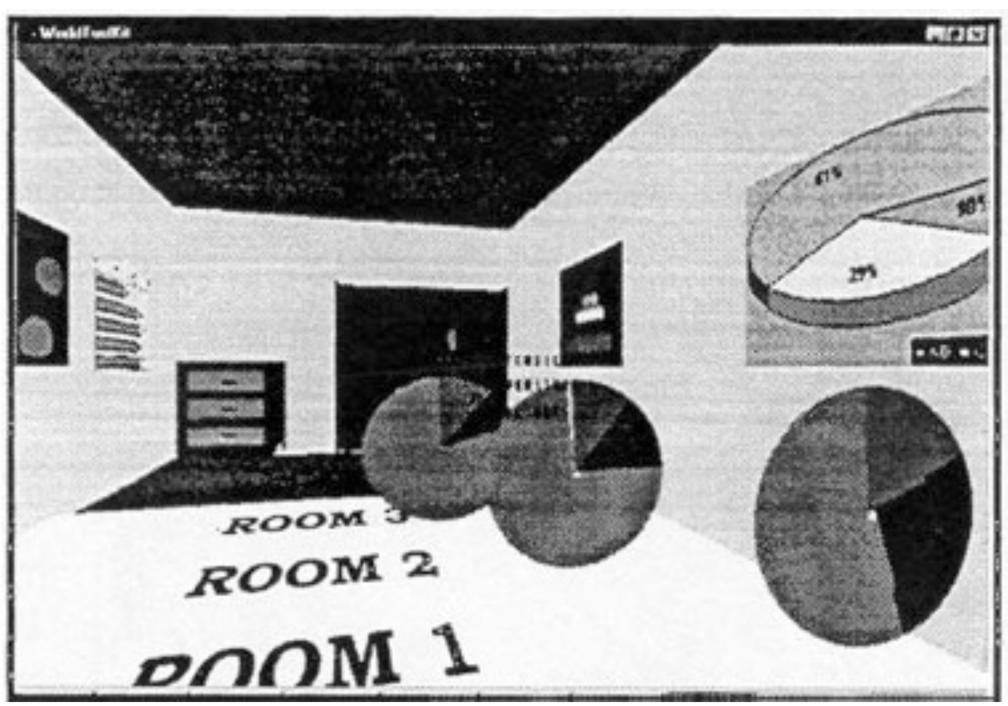
- 3D legitimate for true 3D spatial data
- 3D needs very careful justification for abstract data
  - enthusiasm in 1990s, but now skepticism
  - be especially careful with 3D for point clouds or networks

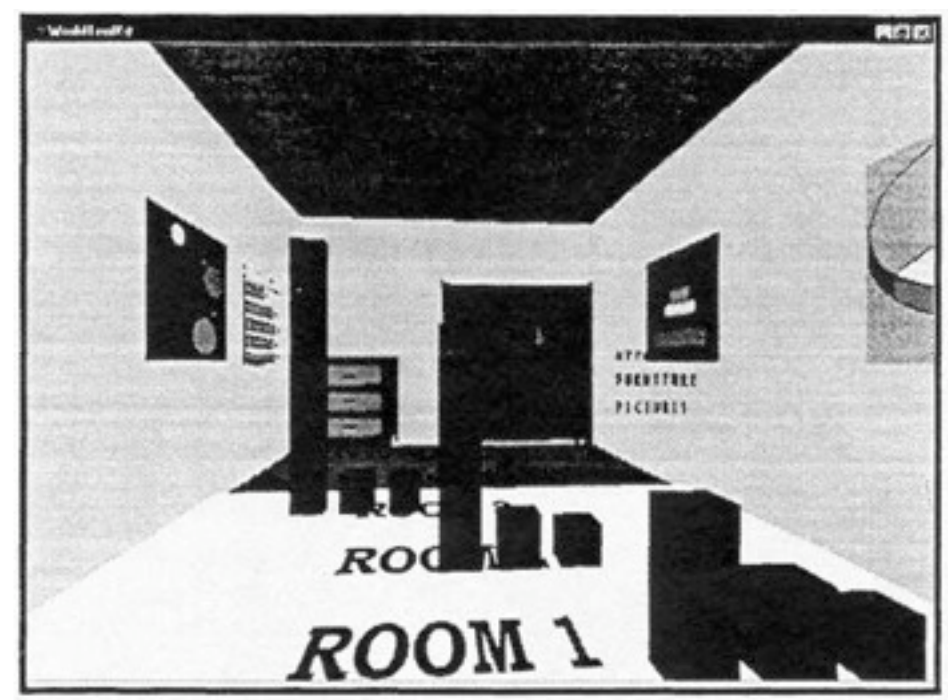


WEBPATH-a three dimensional Web history. Frecon and Smith. InfoVis 1999

Resolution beats immersion

- immersion typically not helpful for abstract data
  - do not need sense of presence or stereoscopic 3D
- resolution much more important
  - pixels are the scarcest resource
  - desktop also better for workflow integration
- virtual reality for abstract data very difficult to justify





Development of an information visualization tool using virtual reality. Kirner and Martins. Symp Applied Computing 2000

Eyes beat memory

- principle: external cognition vs. internal memory
  - easy to compare by moving eyes between side-by-side views
  - harder to compare visible item to memory of what you saw
- implications for animation
  - great for choreographed storytelling
  - great for transitions between two states
  - poor for many states with changes everywhere
    - consider small multiples instead

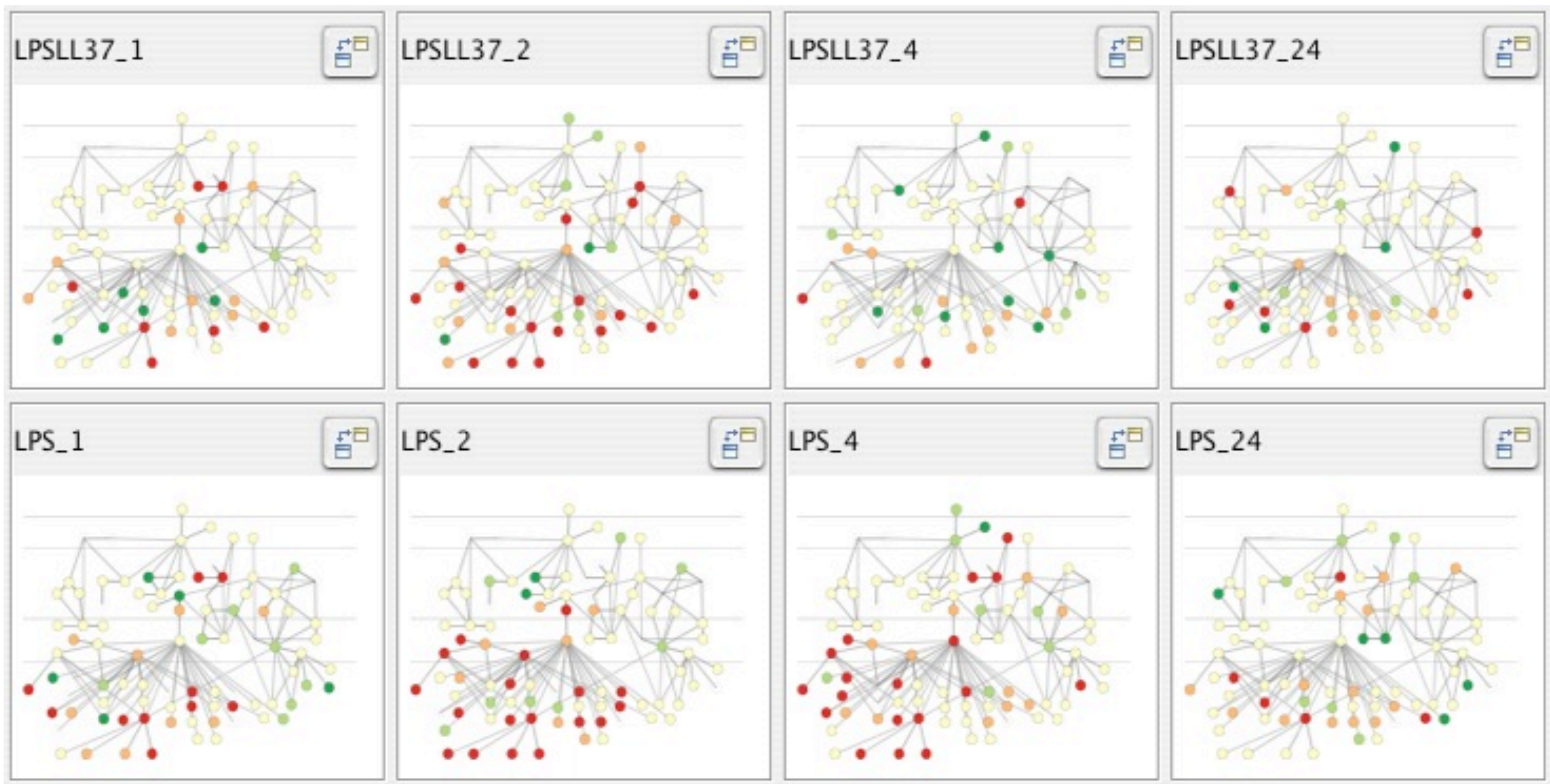
literal | abstract

animation | small multiples

show time with time ⟷ show time with space

Small multiples example: Cerebral

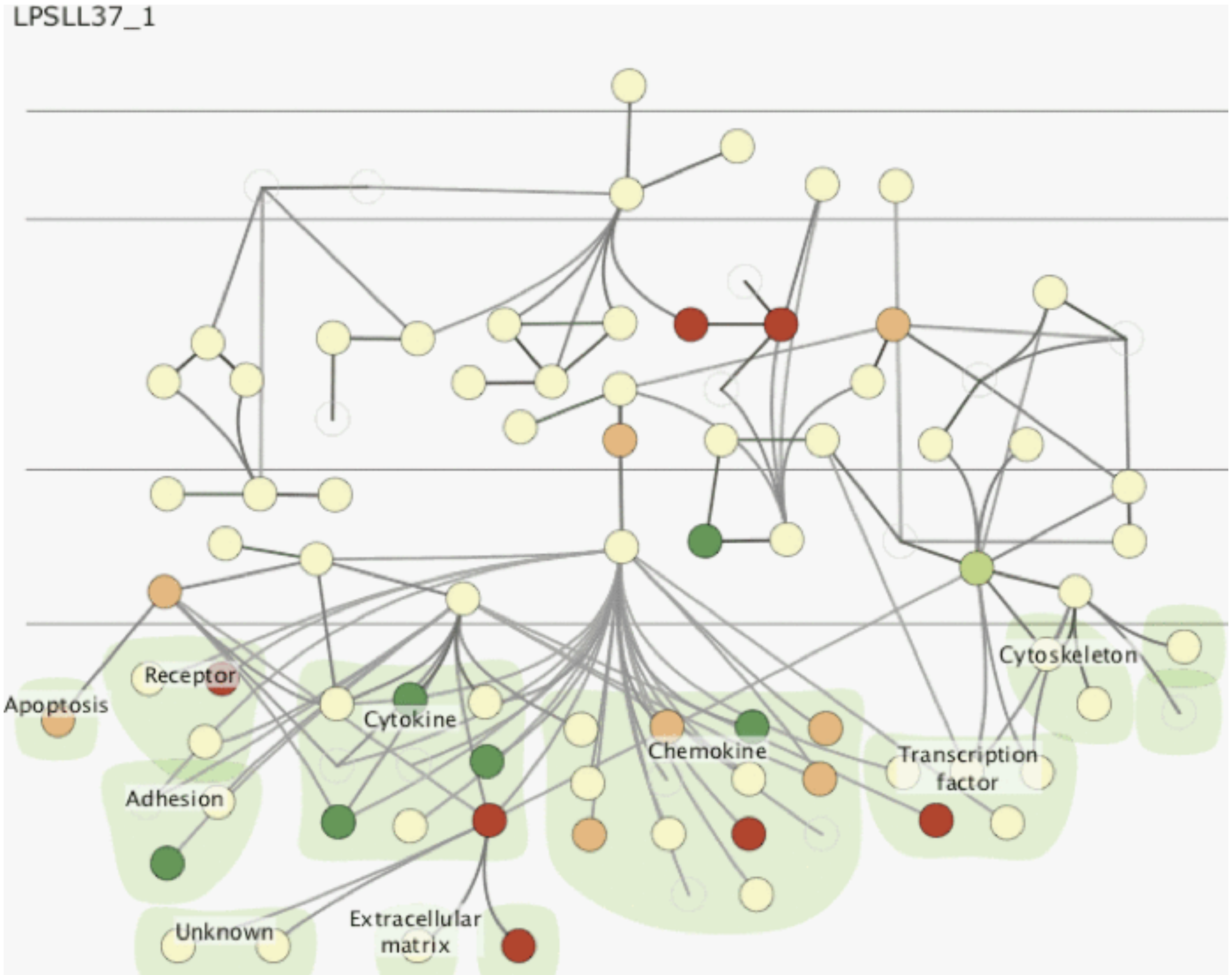
- small multiples: one graph instance per experimental condition
  - same spatial layout
  - color differently, by condition



Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, Kincaid. IEEE InfoVis 2008.

Why not animation?

- global comparison difficult



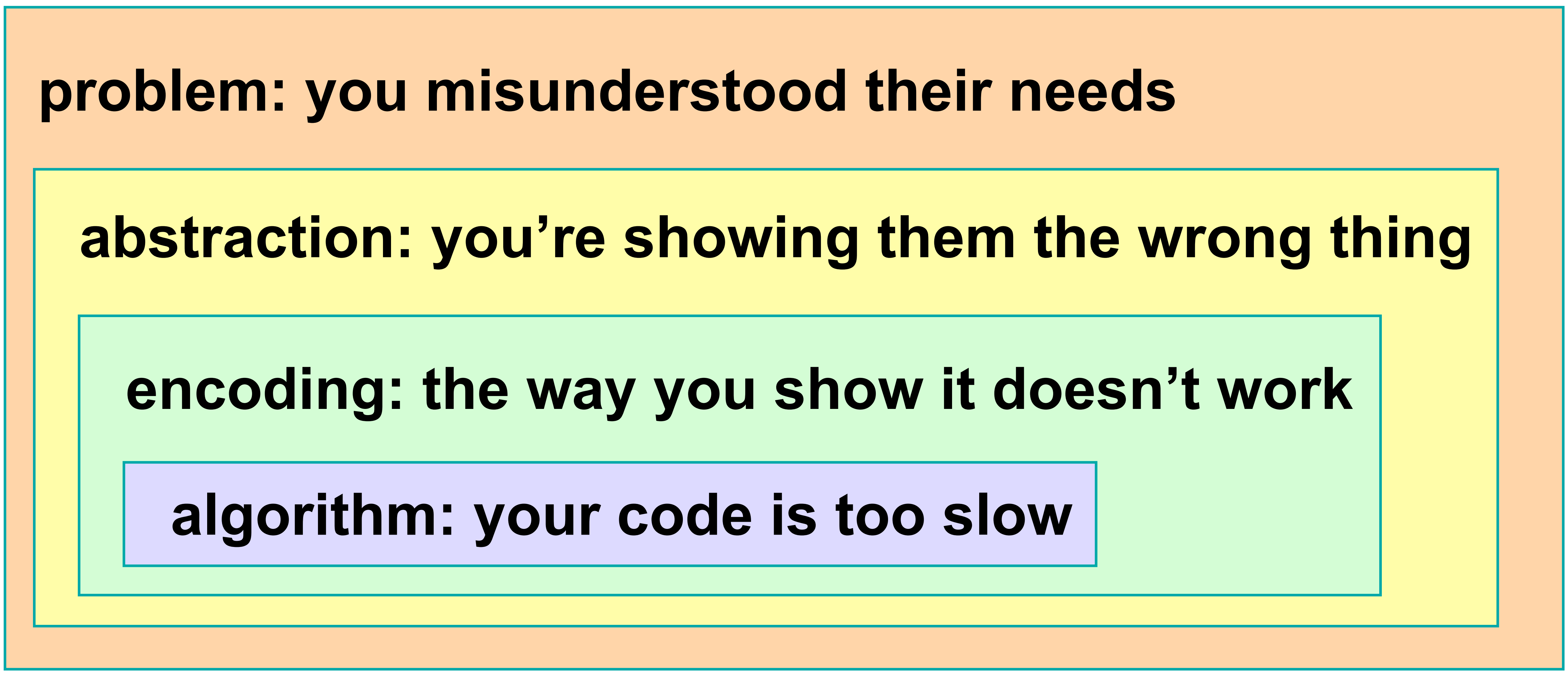
Why not animation?

- further reading

  Animation: can it facilitate? Tversky et al.
  Intl Journ Human-Computer Studies, 57(4):247-262, 2002.

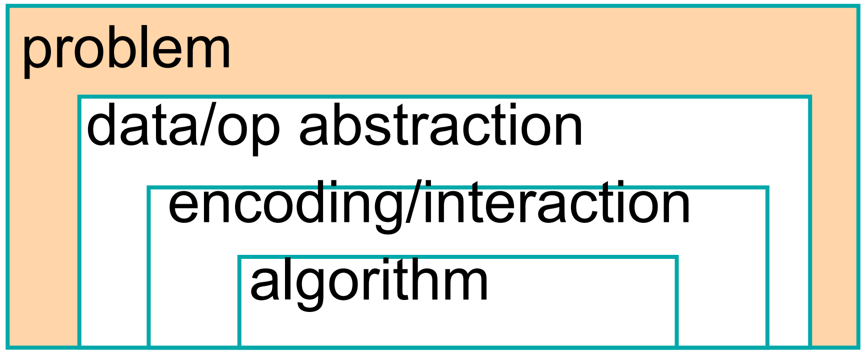
Beyond encoding and interaction

- three more levels of design questions
  - different threats to validity at each level
- validate against the right threat



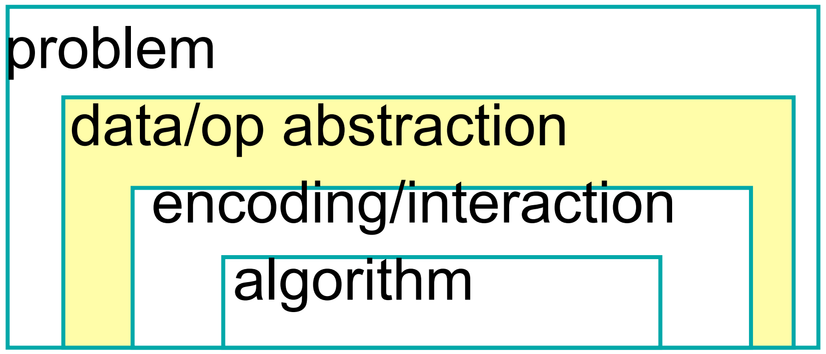
A Nested Model for Visualization Design and Validation. Munzner. IEEE InfoVis 2009.

Characterizing problems of real-world users



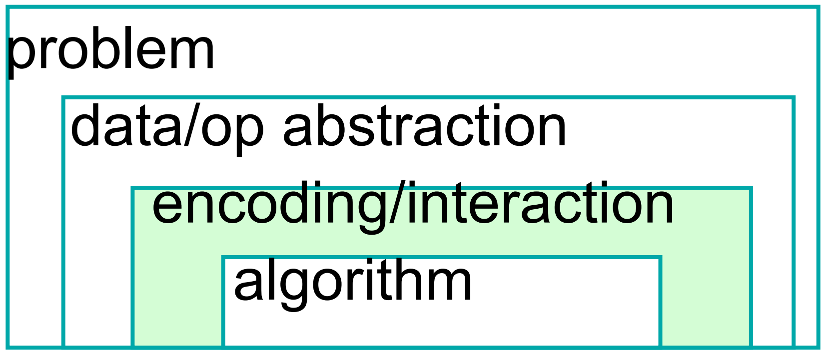
- identify a problem amenable to vis
  - provide novel capabilities
  - speed up existing workflow
- validation
  - immediate: interview and observe target users
  - downstream: notice adoption rates

Abstracting into operations on data types



- abstract from domain-specific to generic
- operations
  - sorting, filtering, browsing, comparing, finding trend/outlier, characterizing distributions, finding correlation...
- data types
  - tables of numbers, relational networks, spatial
  - transform into useful configuration: derived data
- validation
  - deploy in the field and observe usage

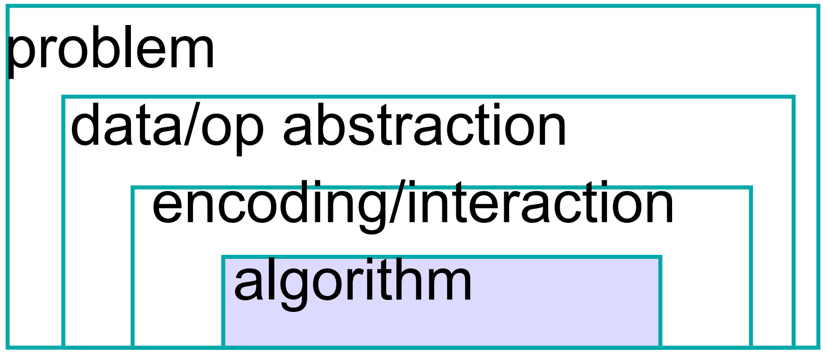
Designing visual encoding, interaction techniques



- visual encoding: drawings they are shown
- interaction: how they manipulate drawings
- validation
  - immediate: careful justification wrt known principles
  - downstream: qualitative or quantitative analysis of results
  - downstream: lab study measuring time/error on given task

- focus of this talk

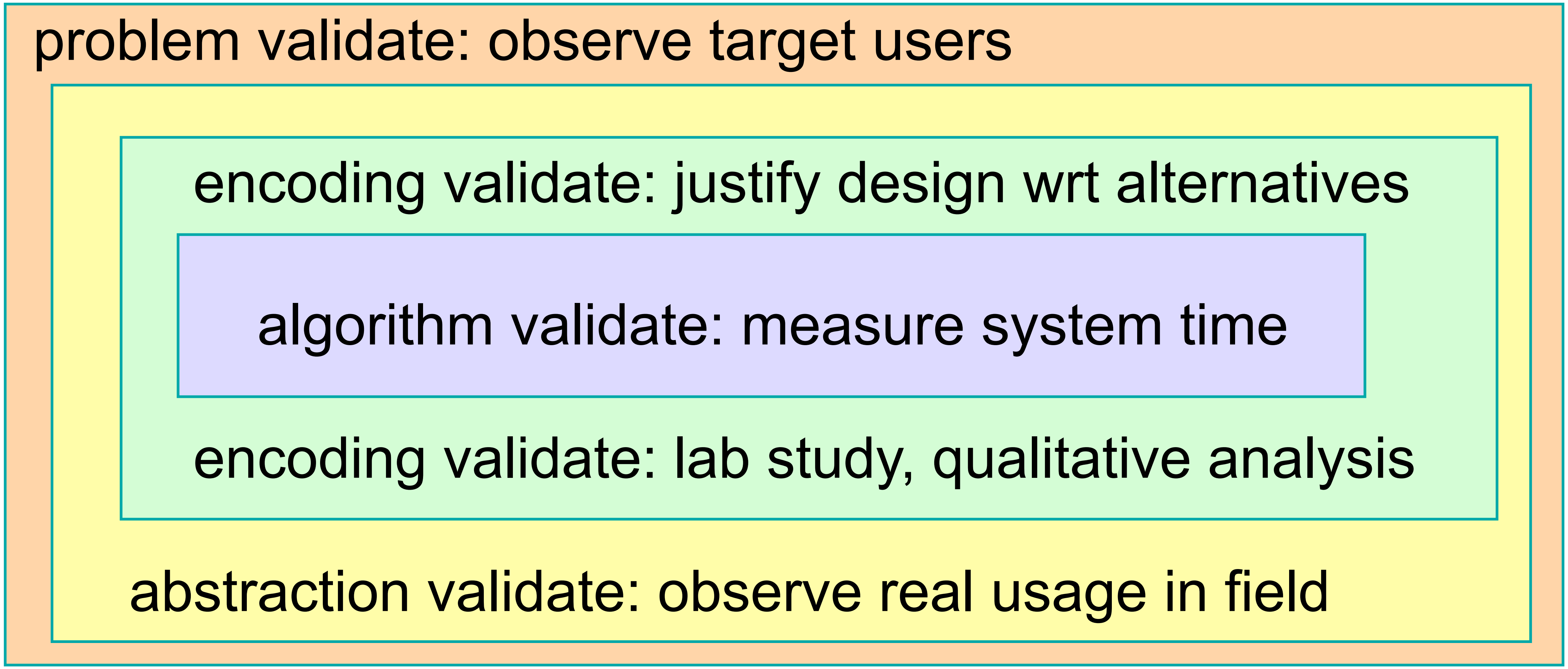
Creating algorithms to execute techniques



- automatically carry out specification
- validation
  - immediate: complexity analysis
  - downstream: benchmarks for system time, memory

Danger of validation mismatch

- cannot show encoding good with system timings
- cannot show abstraction good with lab study



Principles recap

- know your visual channel types and ranks
- categorical color constraints
- power of the plane
- danger of depth
- resolution beats immersion
- eyes beat memory

- validate against the right threat

More information

- vis intro book chapter
  - principles in more depth
  - also, techniques!

    http://www.cs.ubc.ca/~tmm/papers.html#akpchapter

- papers, videos, software, talks, courses
  http://www.cs.ubc.ca/~tmm

- this talk
  http://www.cs.ubc.ca/~tmm/talks.html#vizbi11