浙江大学
ZHEJIANG UNIVERSITY

DATA VISUALIZATION SERIES

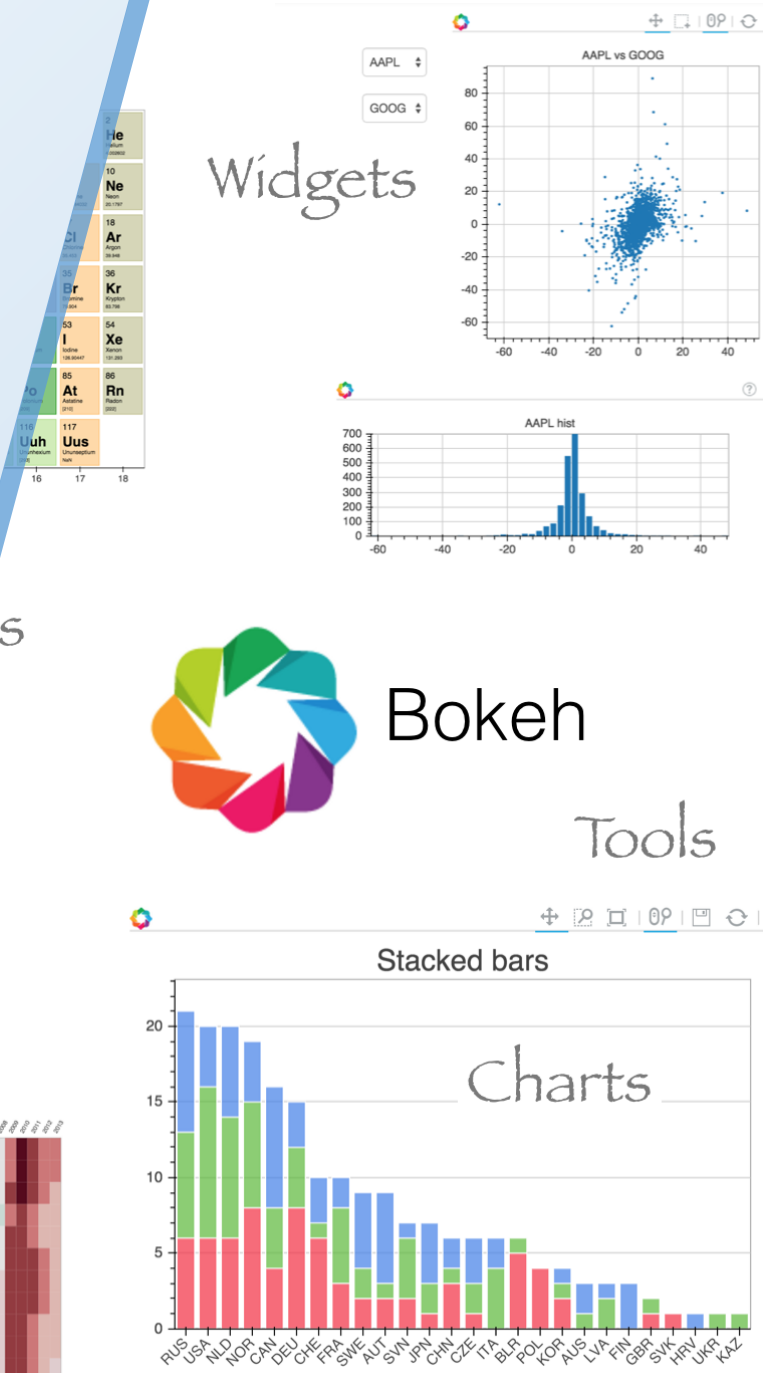
Basics of Data Visualization 1

Yingcai Wu

Professor
State Key Lab of CAD&CG
Email: ycwu@zju.edu.cn

Wei Chen

Professor
State Key Lab of CAD&CG
Email: chenvis@zju.edu.cn



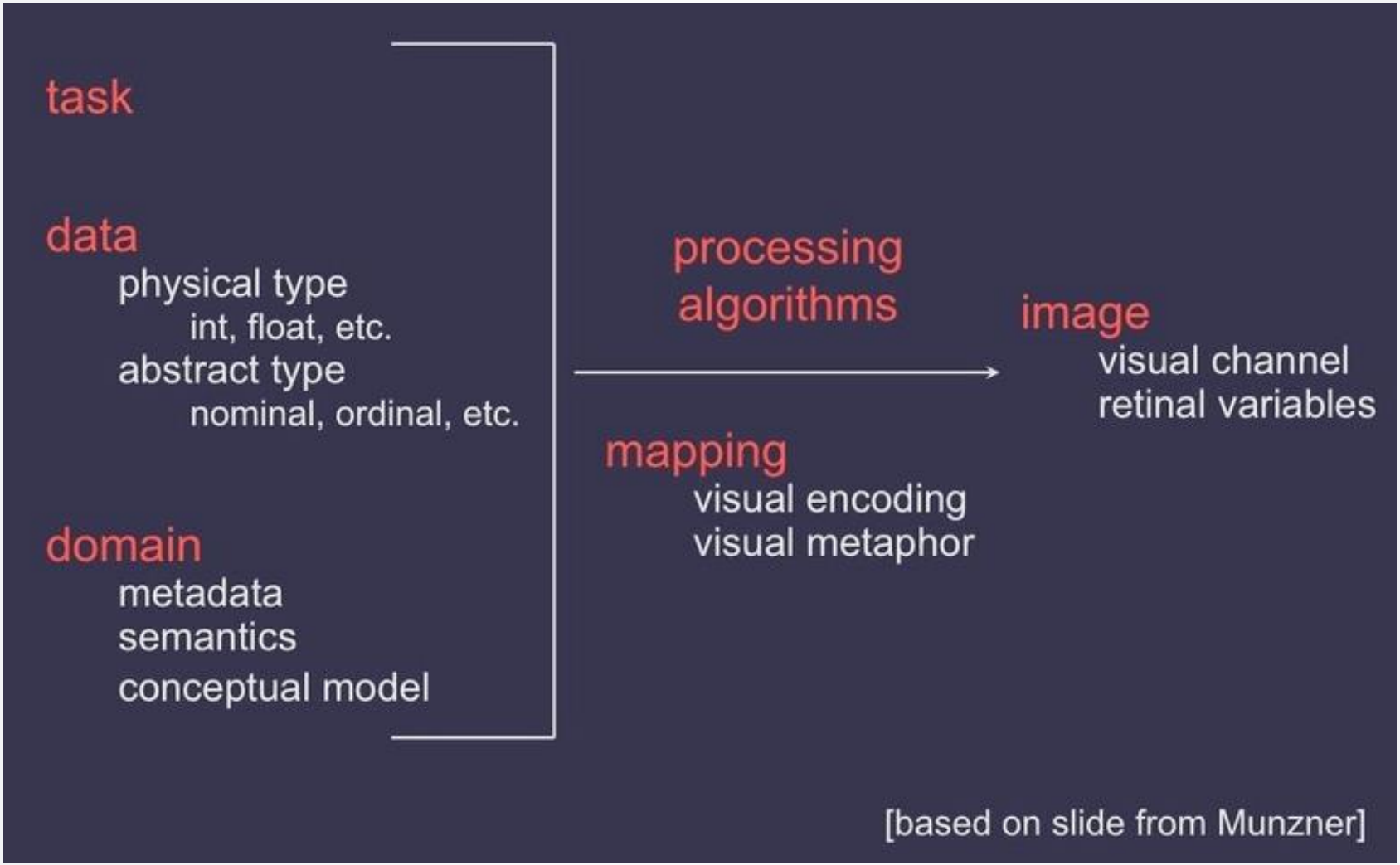
OUTLINE

1. Visualization Process Model
2. Visual Encoding Principles
3. Visual Analysis Model

OUTLINE

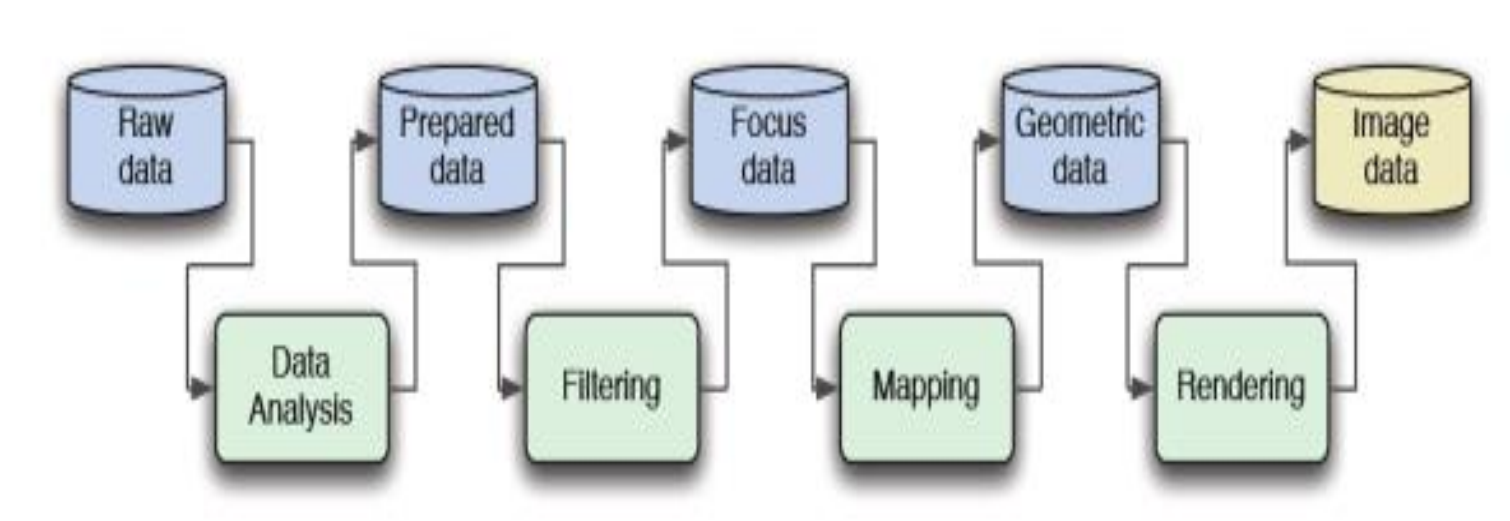
1. Visualization Process Model
2. Visual Encoding Principles
3. Visual Analysis Model

Munzner's Model



A Conceptual Model

Visualization idioms by Haber and McNabb

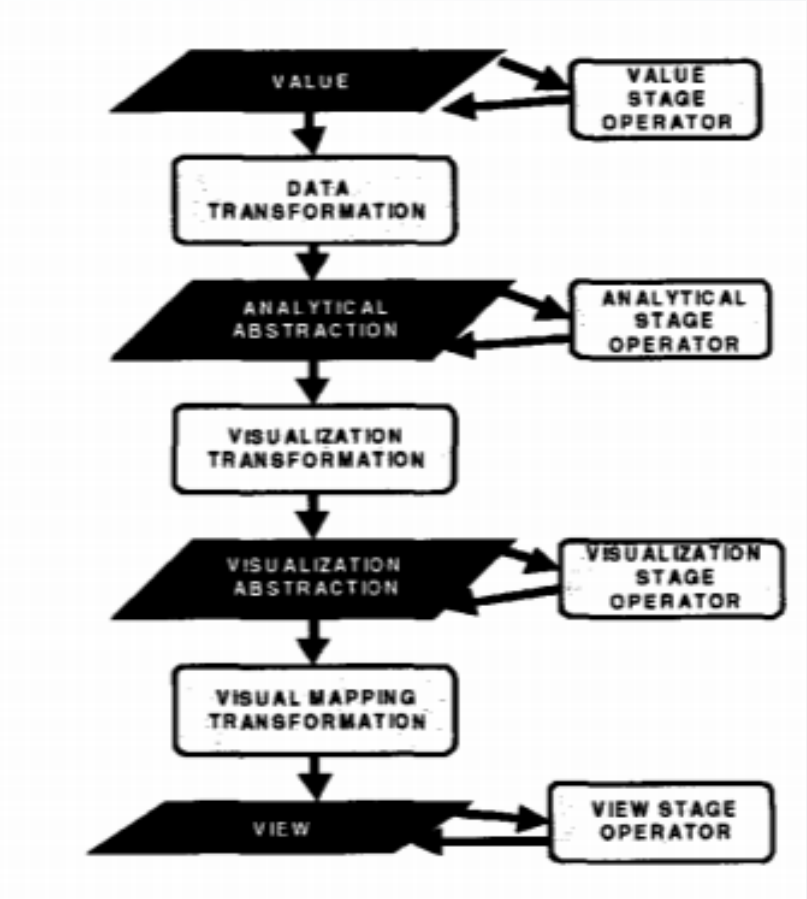


Haber, R. B. and McNabb, D. A. Visualization idioms: A conceptual model for scientific visualization systems, 1990.

Data State Reference Model

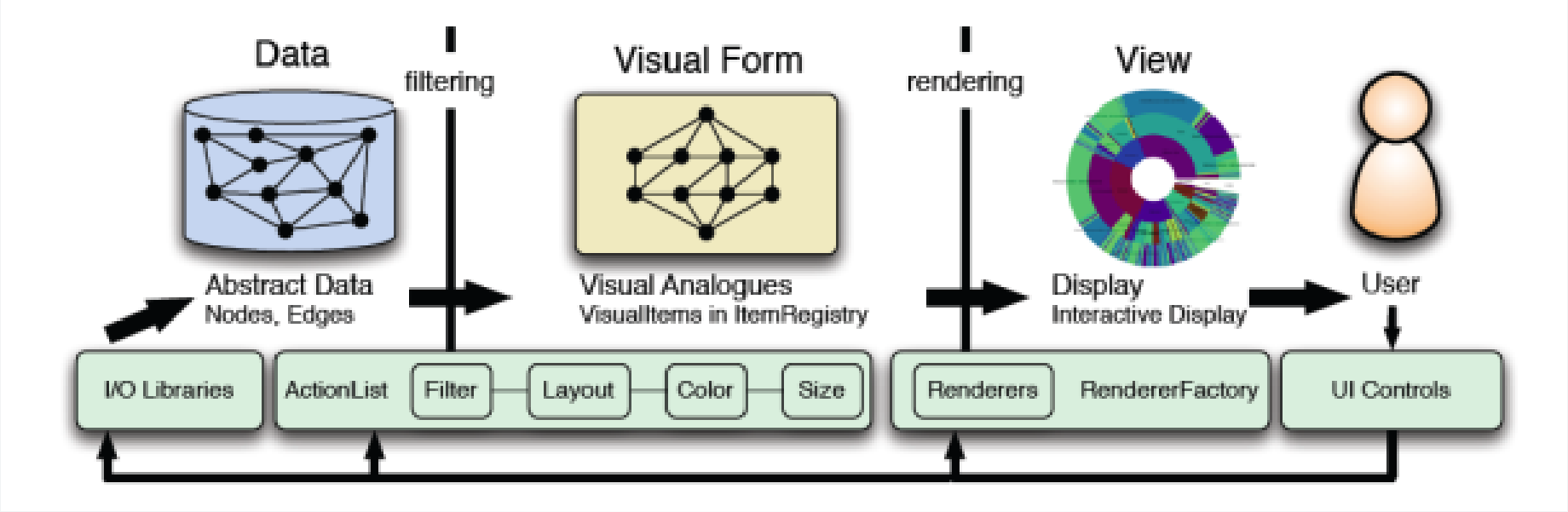
Date state reference model by Ed H. Chi

Ed H. Chi, A Taxonomy of Visualization Techniques using Data State Reference Model. 2000



Visualization Reference Model

Visualization reference model by Card, Mackinlay & Shneiderman



Card, S. K., Mackinlay, J. D., and Shneiderman, B., editors. Readings in Information Visualization: Using Vision to Think, 1999.

Visualization Reference Model

Visualization reference model used in Prefuse, a Java visualization toolkit which had been widely used.

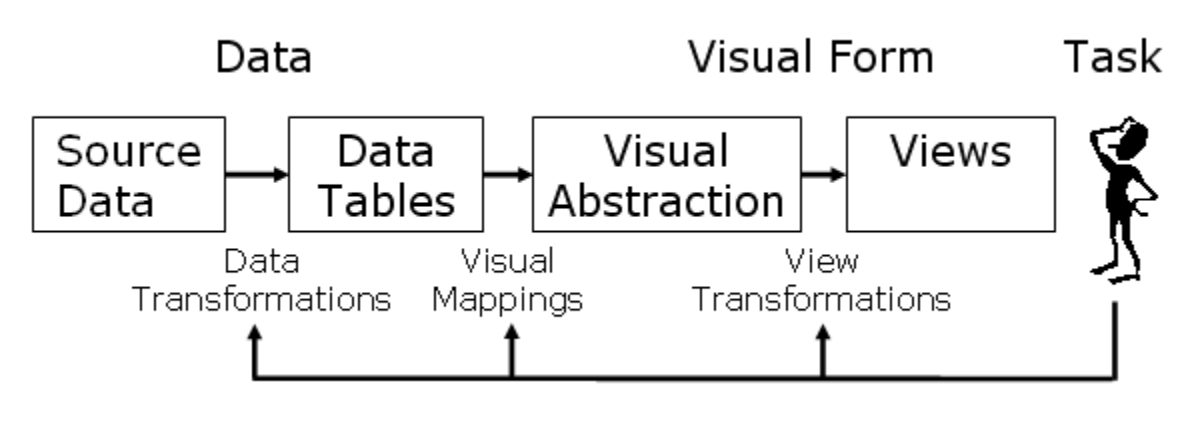
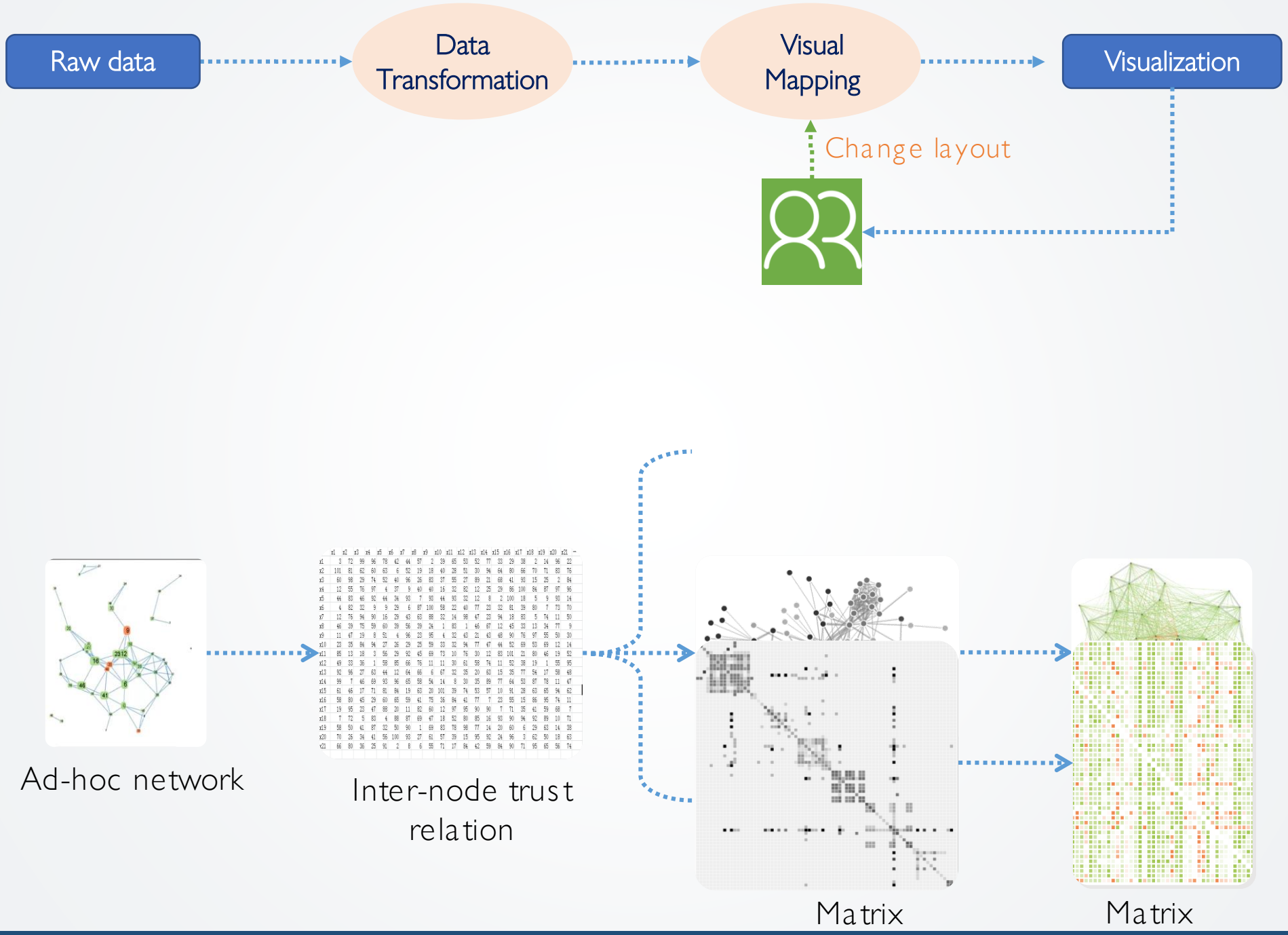
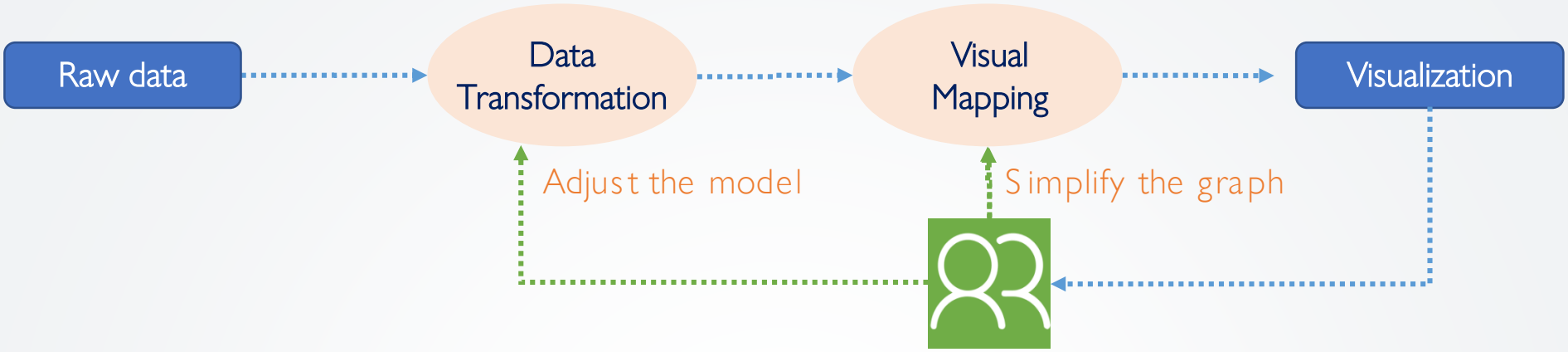
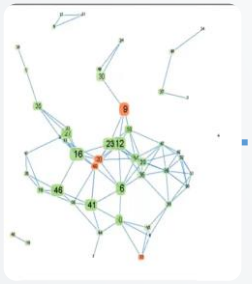


Diagram depicting the information visualization reference model.

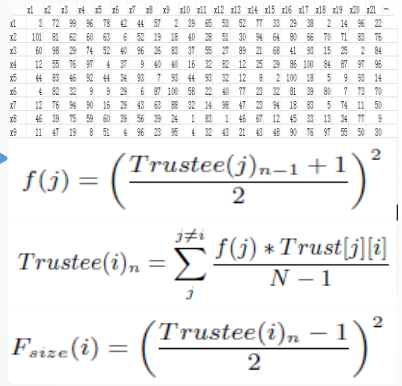
http://prefuse.org/doc/manual/introduction/structure/


Raw data
Data Transformation
Visual Mapping
Visualization
Change layout
Ad-hoc network
Inter-node trust relation
Matrix
Matrix



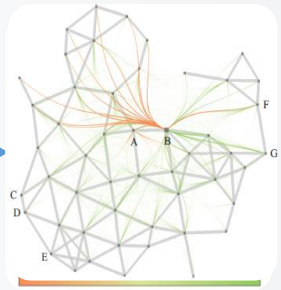


Ad-hoc network

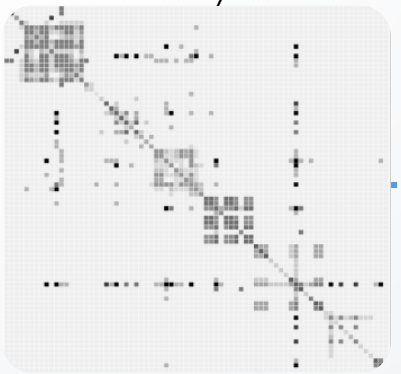


Relation-based reordering, clustering and relation

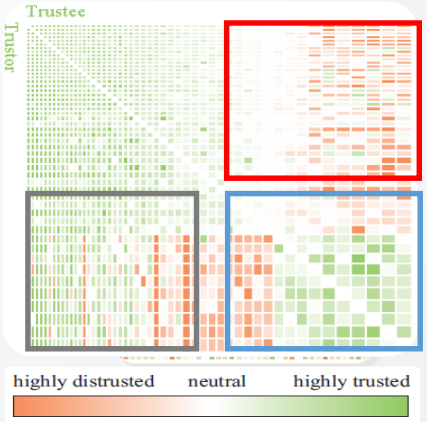
Graph layout



Node-based Graph simplification



Matrix



OUTLINE

1. Visualization Process Model
2. Visual Encoding Principles
3. Visual Analysis Model

Data Types

- Continuous (quantitative)
  - 10 inches, 17 inches, 23 inches

- Ordered (ordinal)
  - small, medium, large
  - days: Sun, Mon, Tue, ...

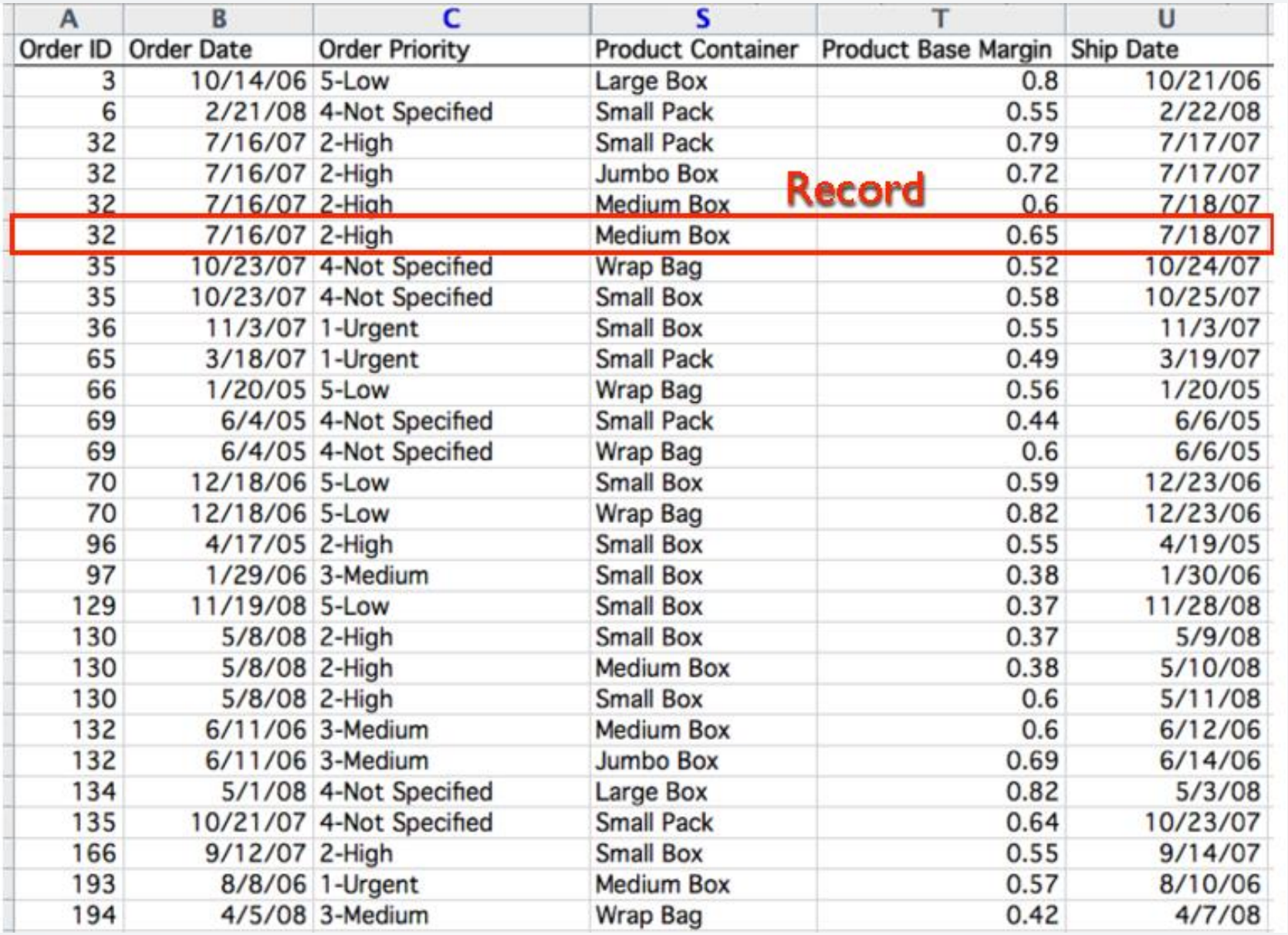
- Categorical (nominal)
  - apples, oranges, bananas

Quantitative

- Interval (location of zero arbitrary)
  - **Dates**: Jan 19; **Location**: (Lat, Long)
  - Only differences (i.e., intervals) can be compared

- Ratio (zero fixed)
  - Measurements: **Length**, **Mass**, **Temp**, ...
  - Origin is meaningful, can measure ratios & proportions

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

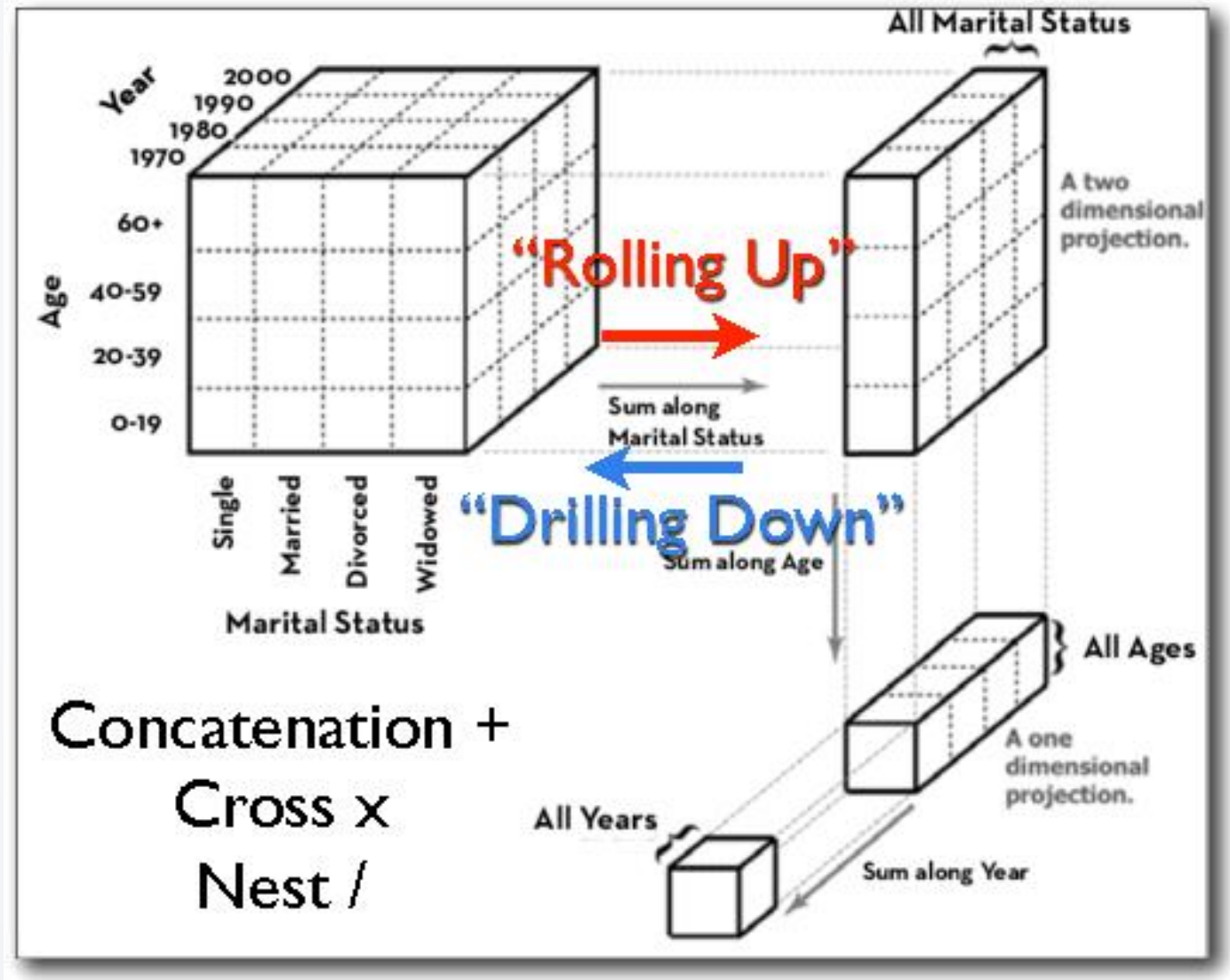


A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Small Box	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Small Box	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

1 = Quantitative
2 = Nominal
3 = Ordinal

Relational Data Cubes



Data Models vs. Conceptual Models

Data model: mathematical abstraction

- Set with operations
- E.g. integers or floats with +, -, *, or /

Conceptual model: mental construction

- Includes semantics, support data
- e.g. navigating through city using landmarks

[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding/walk005.html]
[Rethinking Visualization: A High-Level Taxonomy. Melanie Tory and Torsten Moller, Proc. InfoVis 2004, pp. 151-158.]

Models Example

- From data model
  - 17, 25, -4, 28.6
  - (floats)
- Using conceptual model
  - (temperature)
  - to data type
- Continuous to 4 sig figures (Q)
  - hot, warm, cold (O)
  - burned vs. not burned (N)
- Using task
  - finding anamolies in local weather patterns
  - classifying showers
  - making toast

Jacques Bertin

- French cartographer [1918]
- Semiology of Graphics [1967]
- Theoretical principles for visual encodings

Visual Encodings

Channels

Position

Size

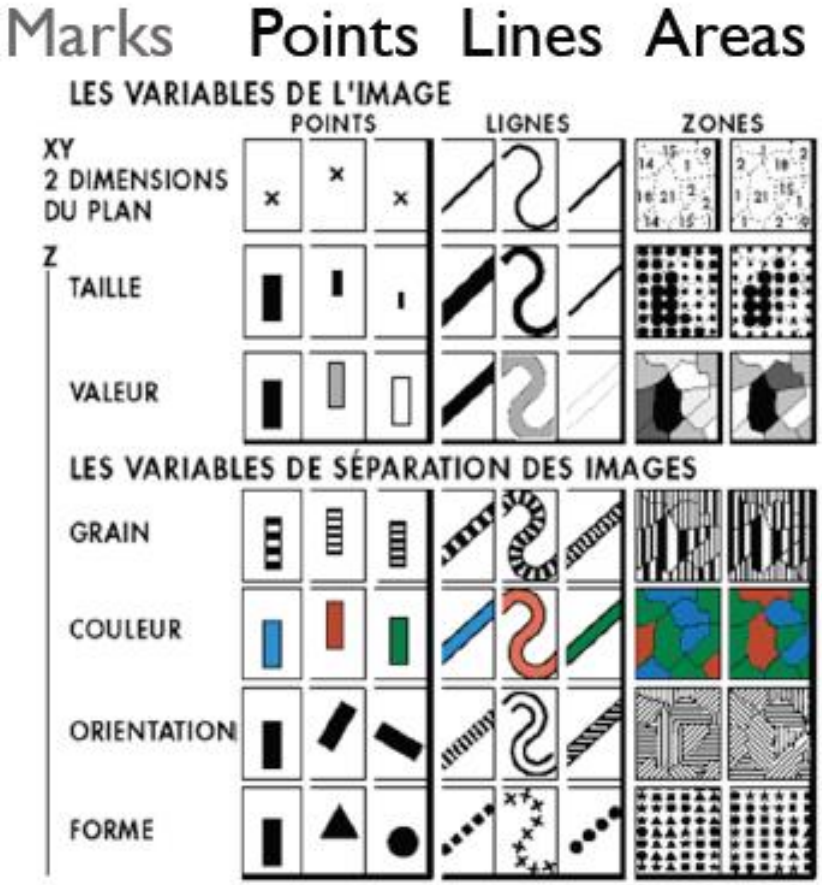
(Grey)Value

Texture

Color

Orientation

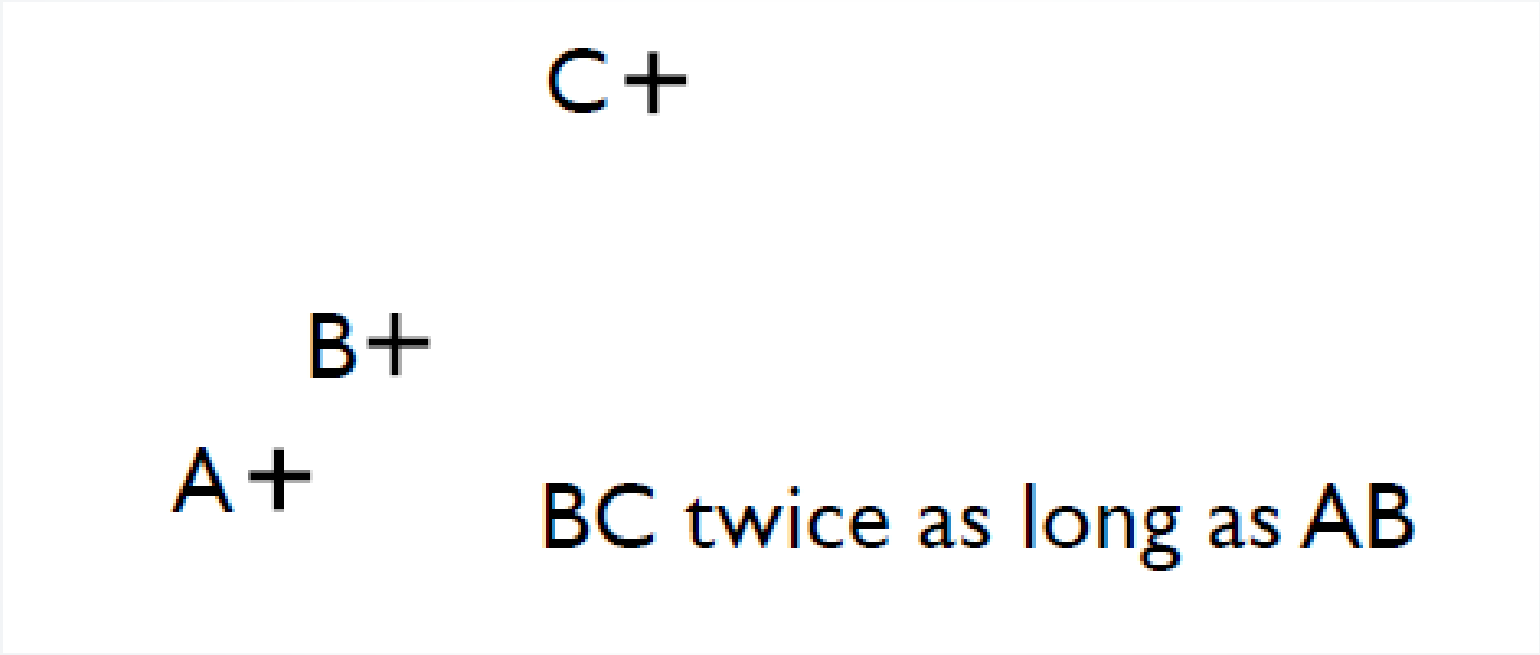
Shape



Semiology of Graphics [J. Bertin, 67]

Information in Position

Good to encode quantitative variables (Q)



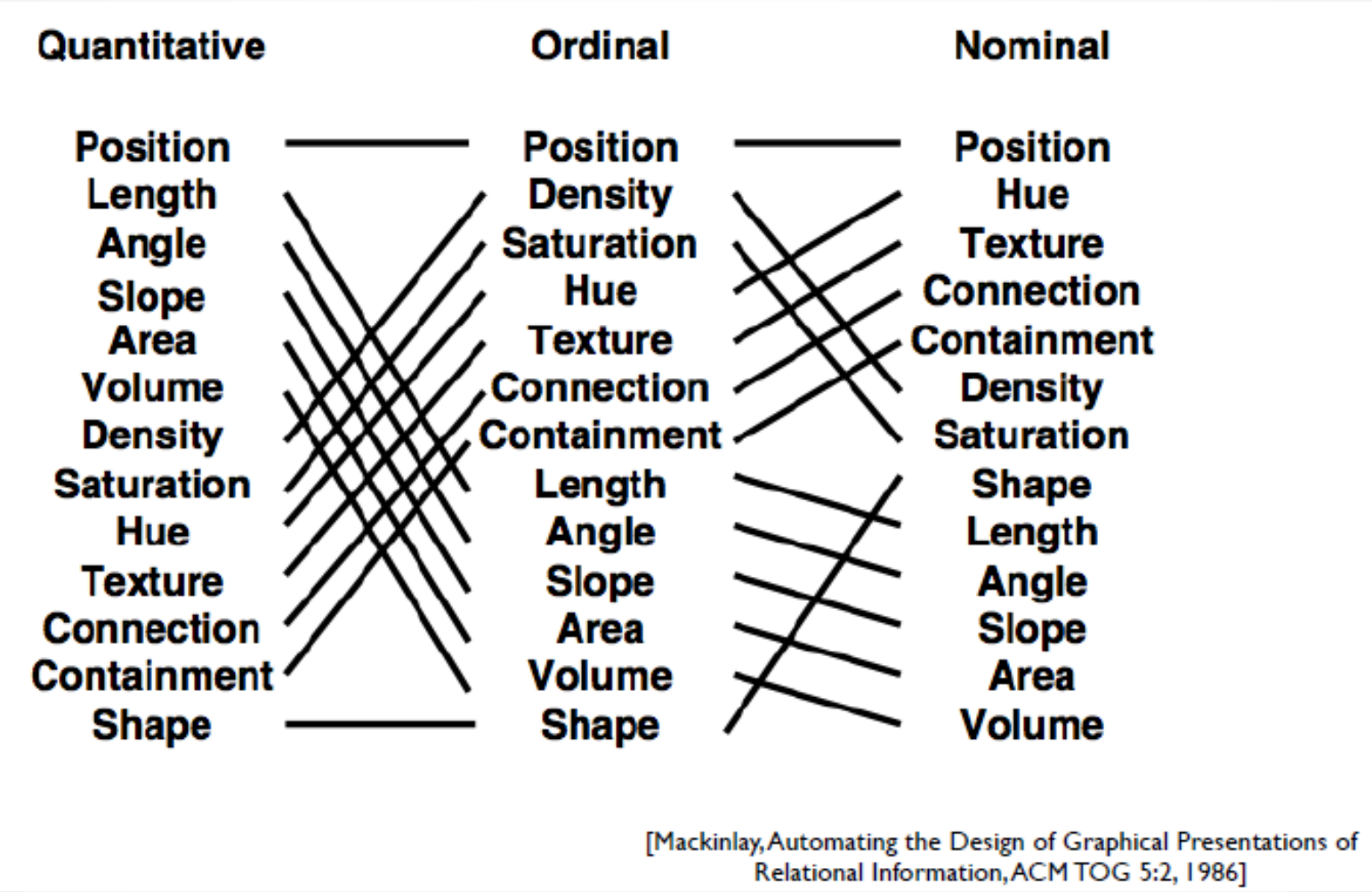
Information in Color

(Grey)Value is perceived as ordered (O)

Can encode quantitative values (Q) [not as well]

Hue is normally perceived as unordered (N)

Mackinlay's Retinal Variables



[Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]

Tableau's Retinal Variables

property	marks	ordinal/nominal mapping	quantitative mapping
shape	glyph	○ □ + △ S U	
size	rectangle, circle, glyph, text		
orientation	rectangle, line, text		
color	rectangle, circle, line, glyph, y-bar, x-bar, text, gantt bar		min … max

["Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases" Chris Stolte, Diane Tang, and Pat Hanrahan, 2002]

Shneiderman’s Data & Tasks Taxonomy

- Data
  - 1D, 2D, 3D, temporal, nD, trees, networks
  - text and documents (Hanrahan)
- Mantra:
  - overview first
  - zoom and filter
  - details on demand
- Tasks
  - overview, zoom, filter, details-on-demand
  - relate, history, extract
  - data alone not enough
  - what do you need to do?

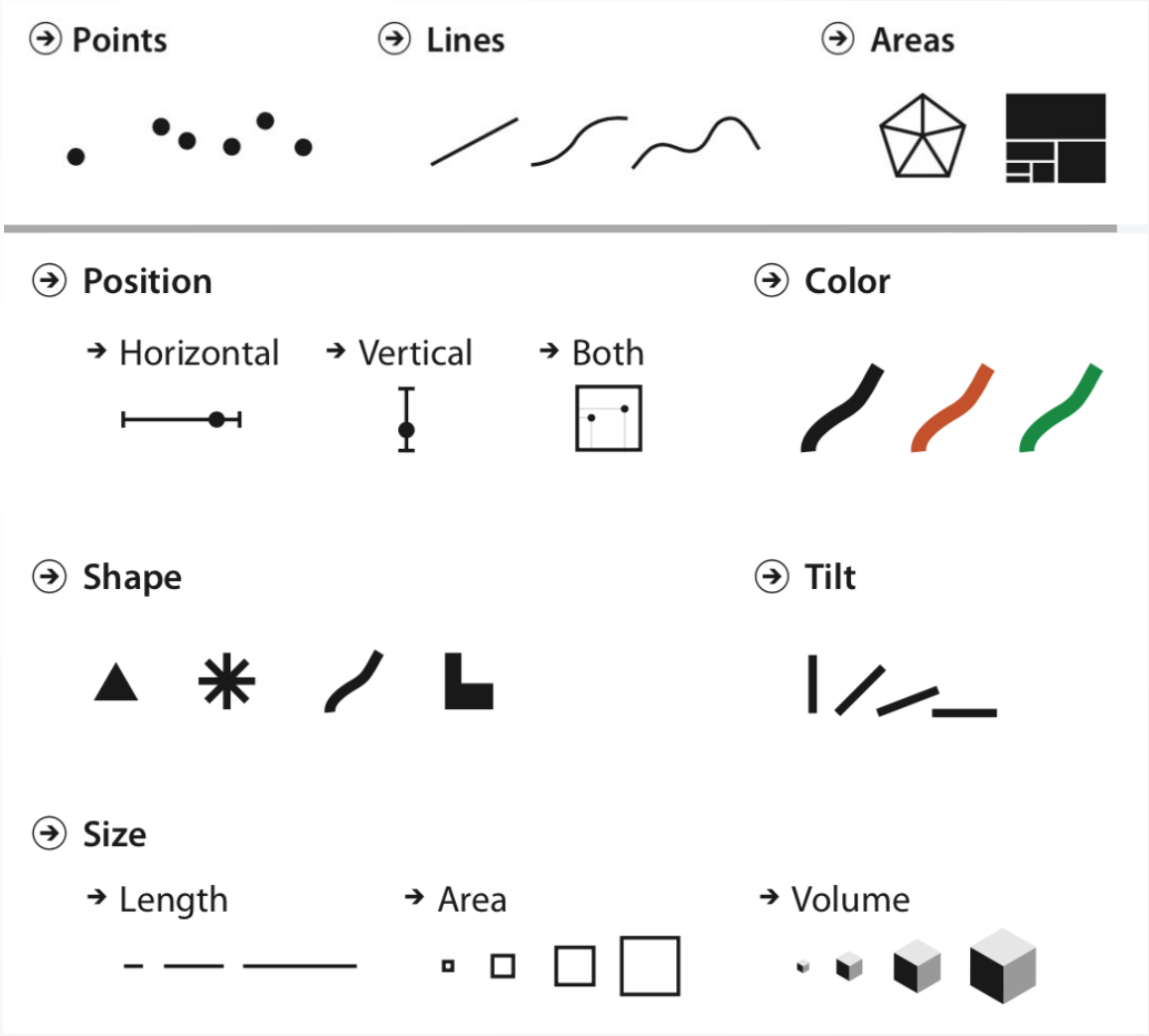
Definitions: Marks and channels

Marks

- Geometric primitives

Channels

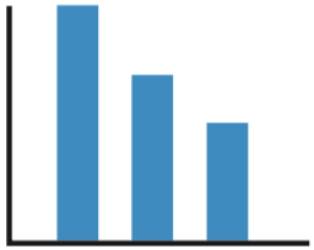
- Control appearance of marks
- Can redundantly code with multiple channels



Visual encoding

Analyze idiom structure

- As combination of marks and channels



1:
vertical position

mark: line

2:
vertical position
horizontal position

mark: point

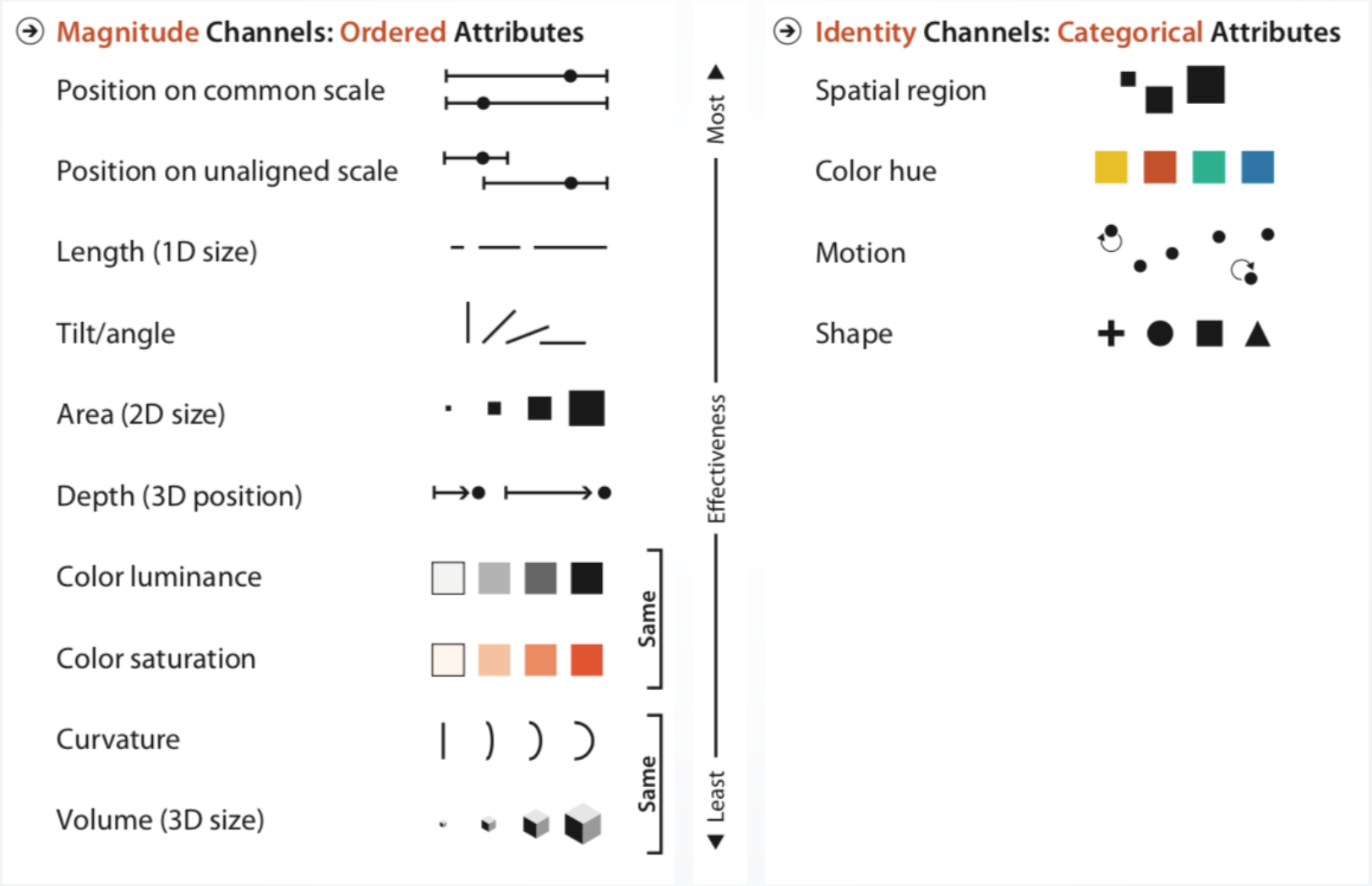
3:
vertical position
horizontal position
color hue

mark: point

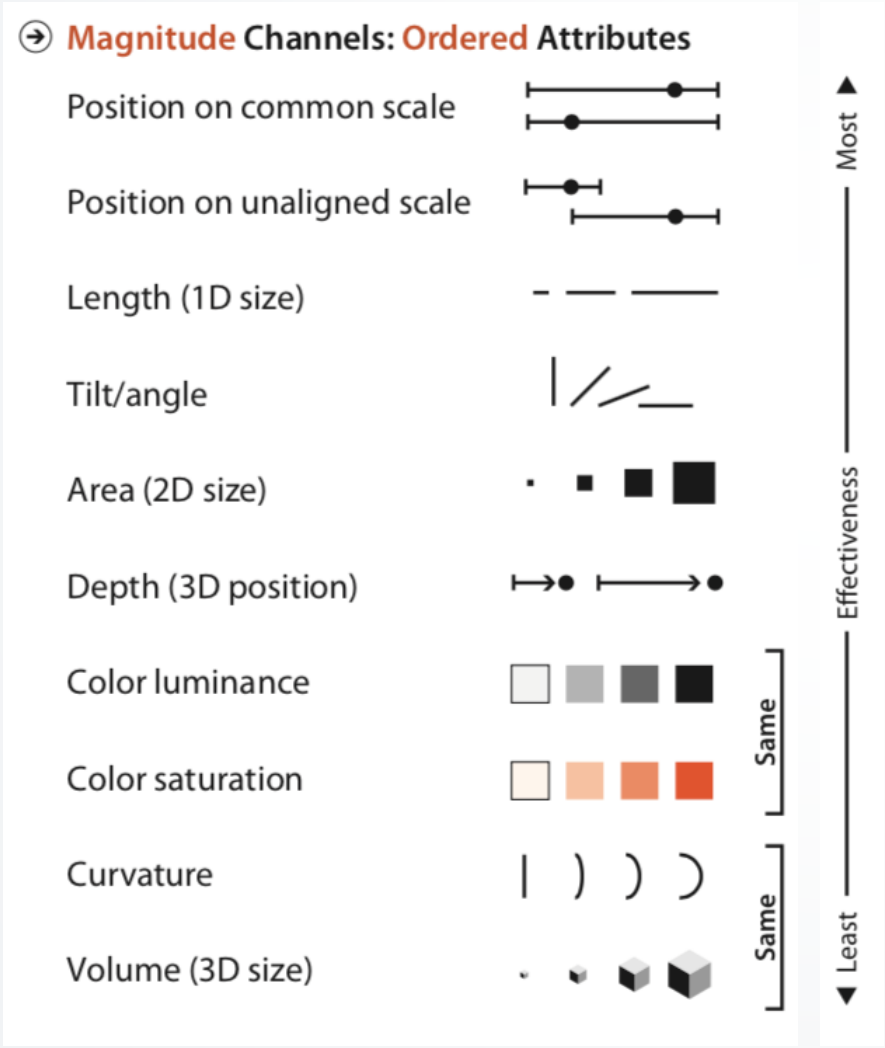
4:
vertical position
horizontal position
color hue
size (area)

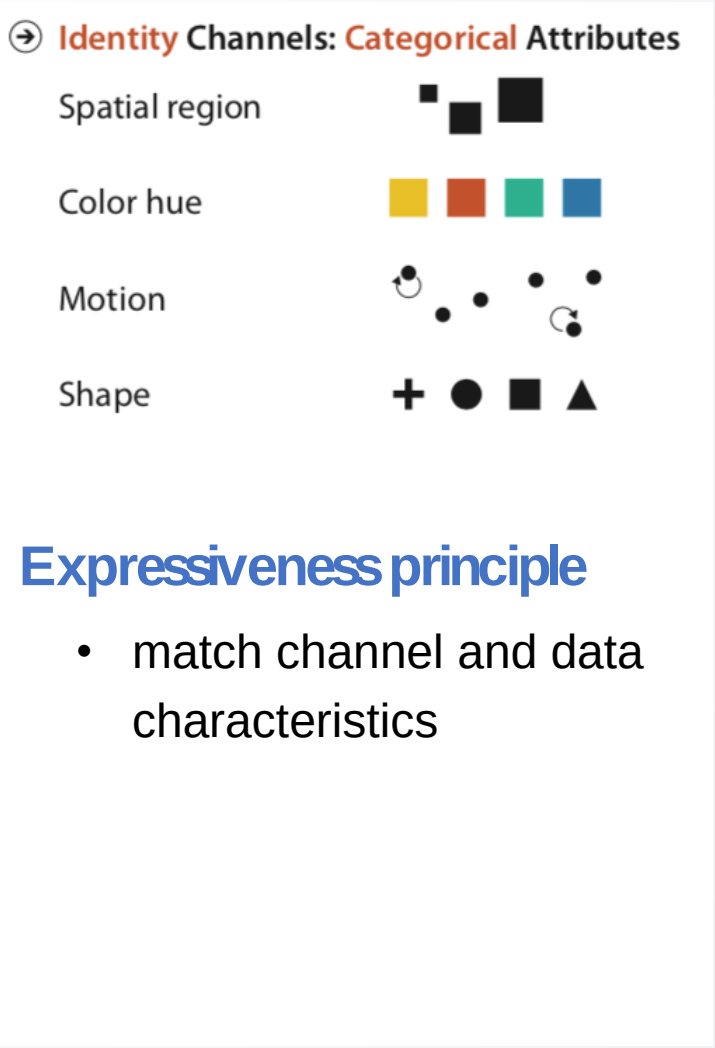
mark: point

Channels: expressiveness types and effectiveness rankings



Channels: Matching Types

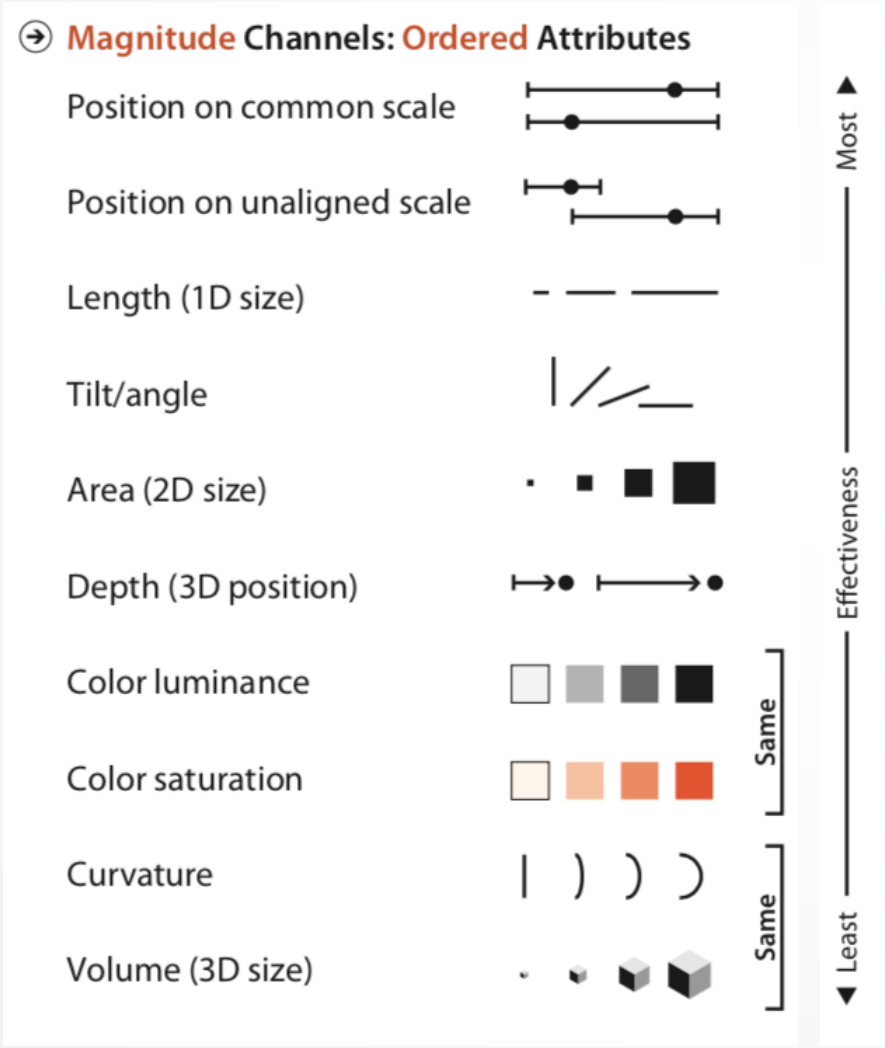


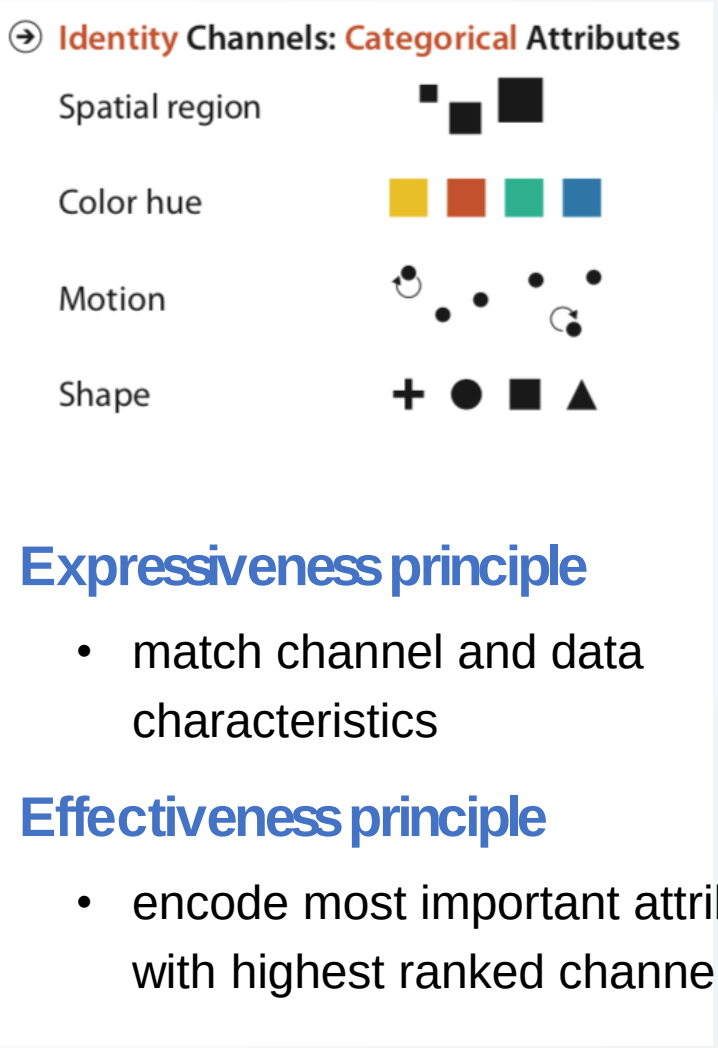


Expressiveness principle

- match channel and data characteristics

Channels: Rankings





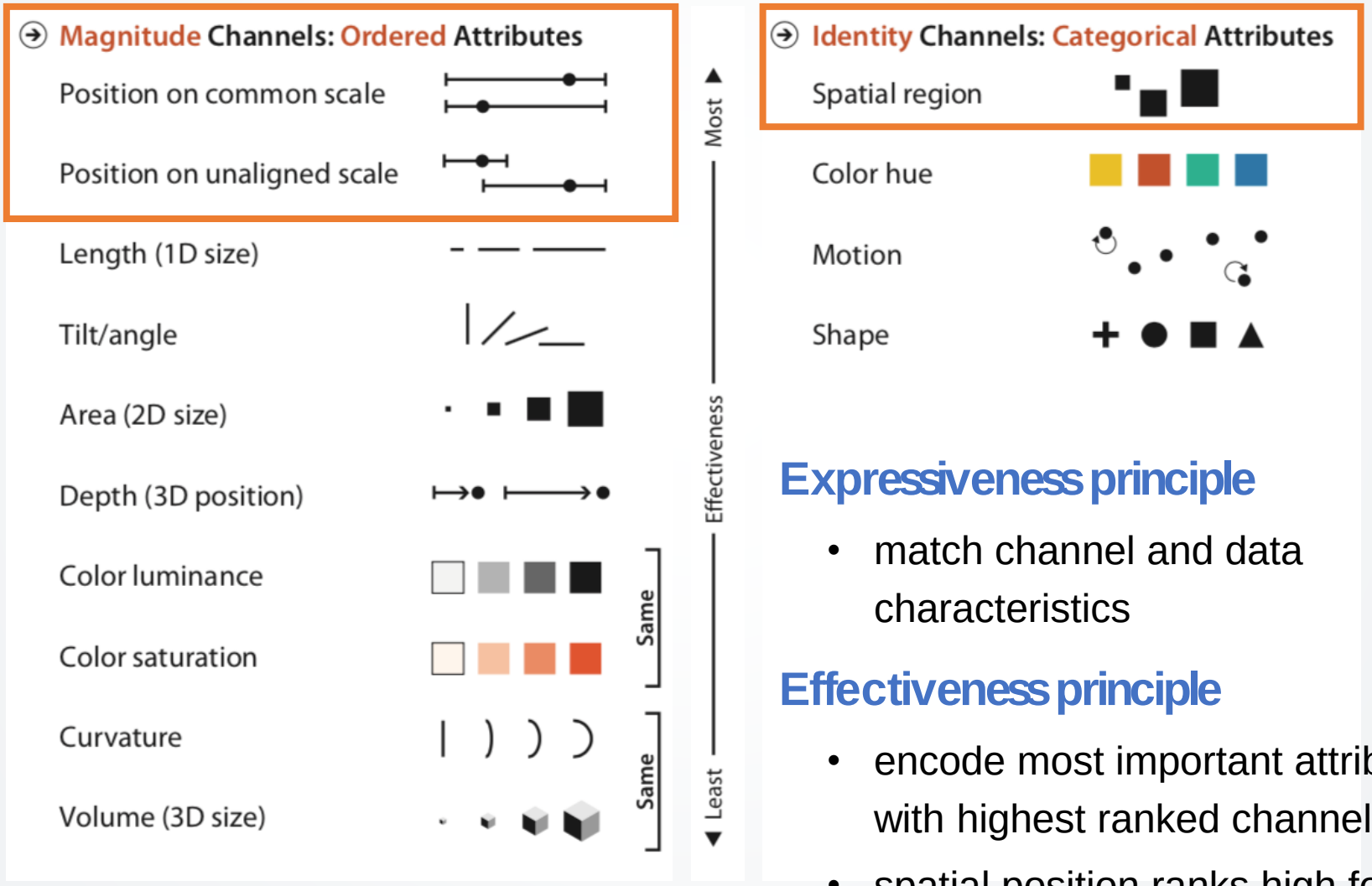
Expressiveness principle

- match channel and data characteristics

Effectiveness principle

- encode most important attributes with highest ranked channels

Channels: Expressiveness types and effectiveness rankings



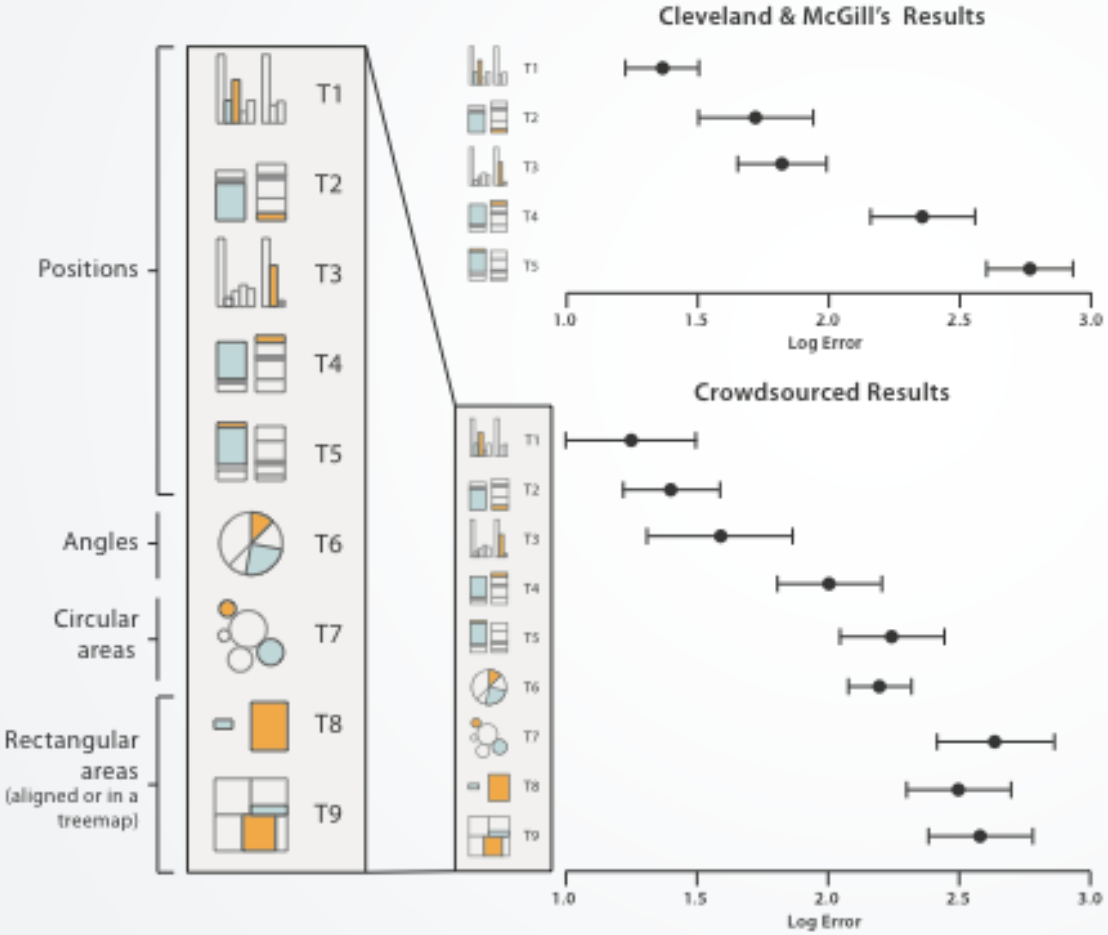
Expressiveness principle

- match channel and data characteristics

Effectiveness principle

- encode most important attributes with highest ranked channels
- spatial position ranks high for both

Accuracy: Vis experiments

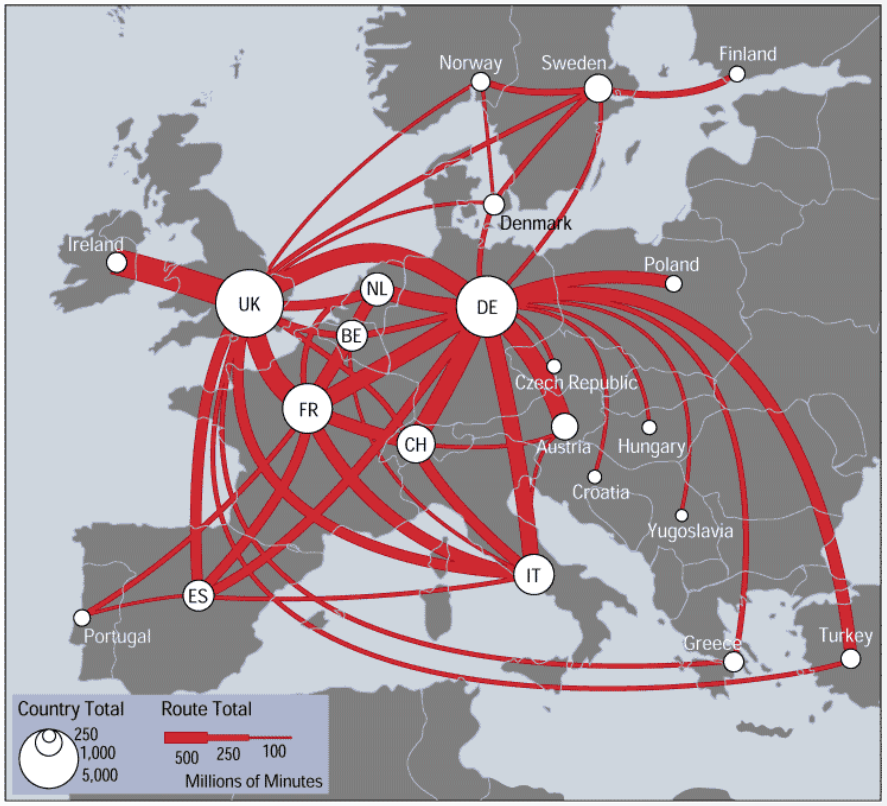


[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.]

Discriminability: How many usable steps?

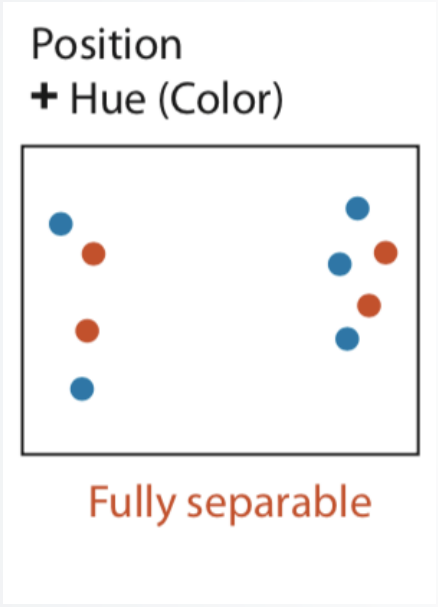
Must be sufficient for the number of attribute levels to show

- Linewidth: a few bins

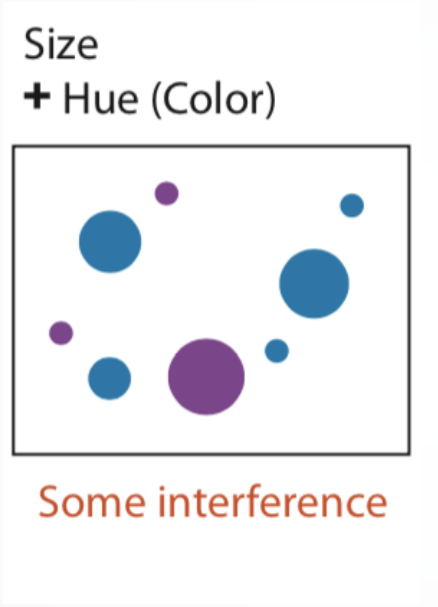


Linewidth has a limited number of discriminable bins

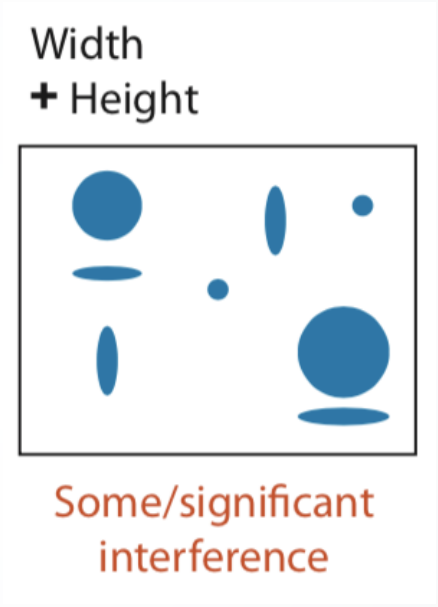
Separability vs. Integrality



2 groups each



2 groups each

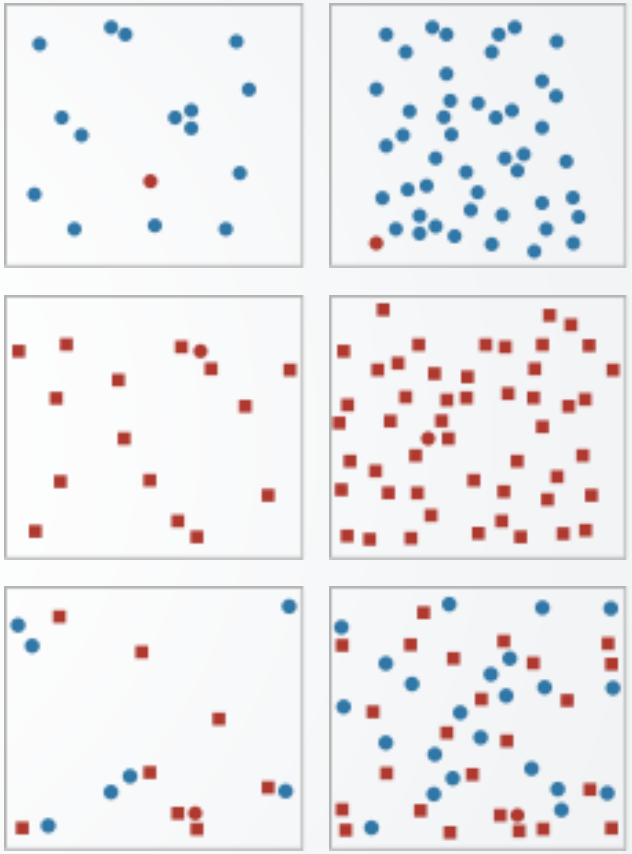


3 groups total:
integral area

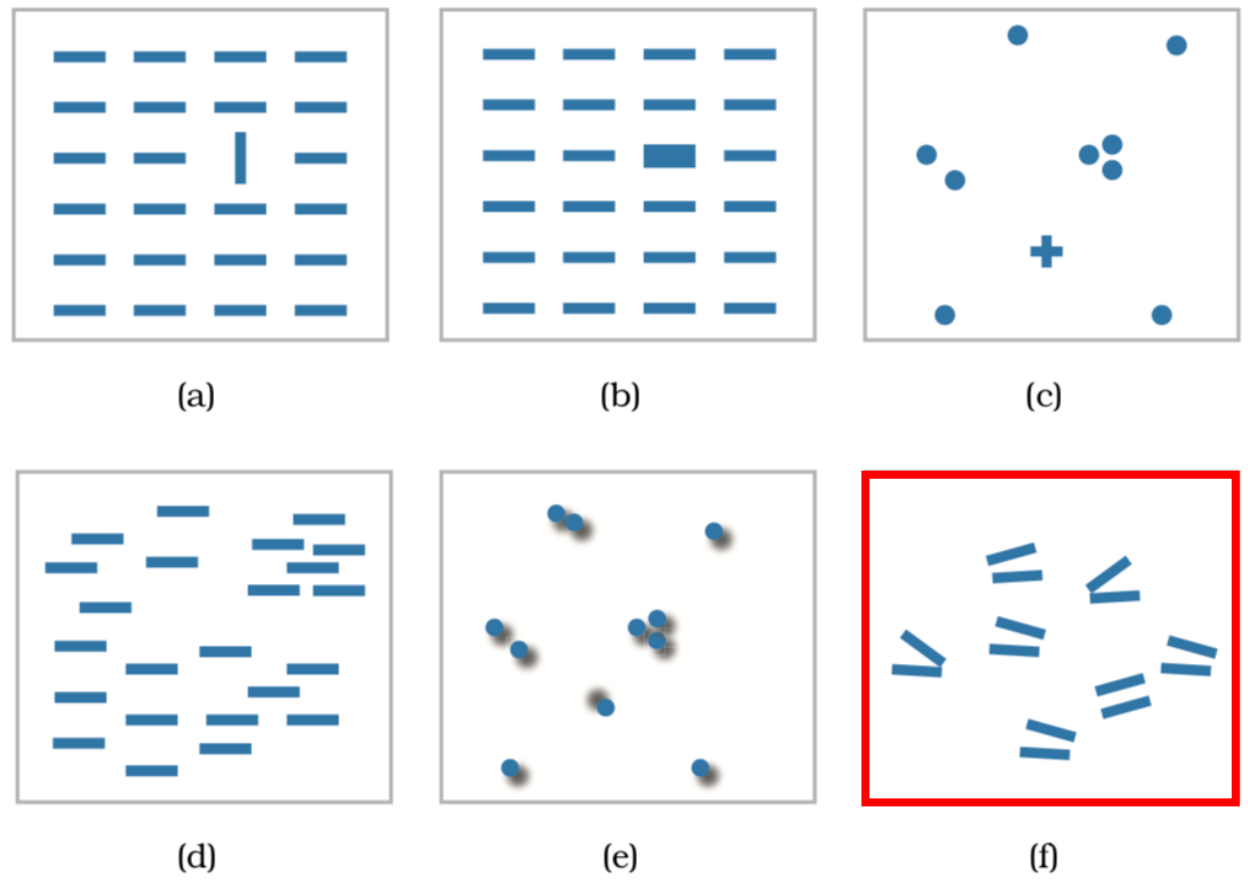
4 groups total:
integral hue

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

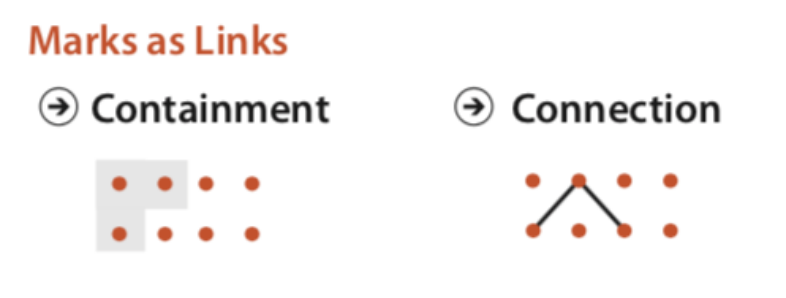


Many channels: (a)tilt, (b)size, (c)shape, (d)proximity, (e)shadow direction,

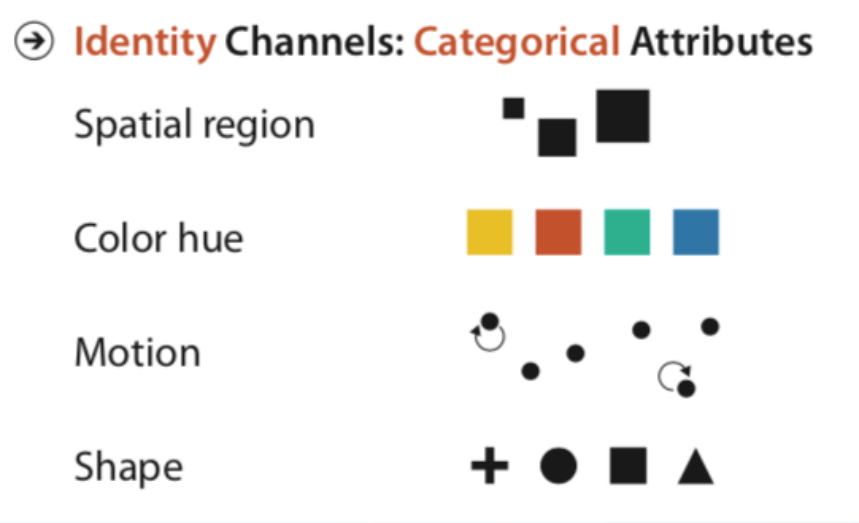
(f) but not all! parallel line pairs do not pop out from tilted pairs.

Grouping

- Containment
- Connection

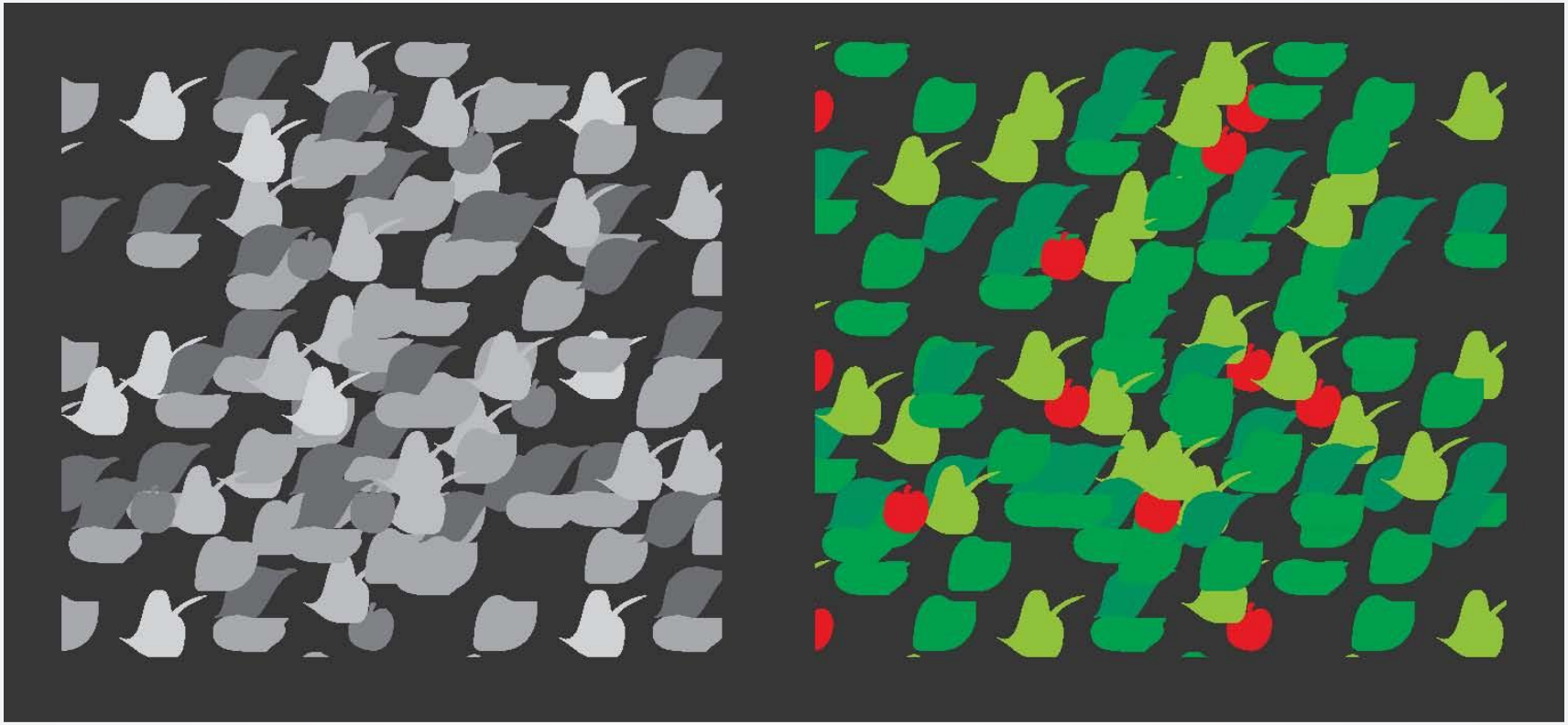


- Proximity
  - Same spatial region
- Similarity
  - Same values as other categorical channels



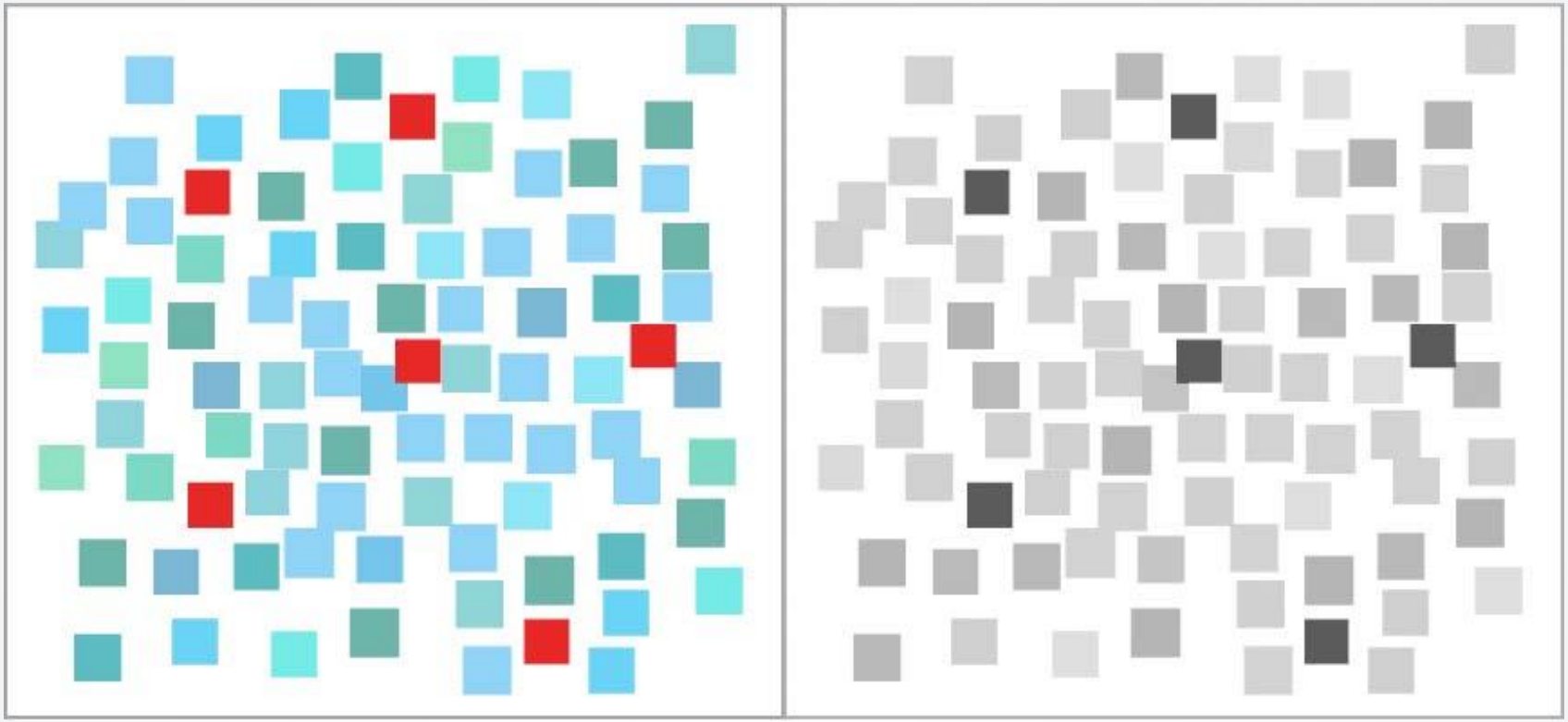
Color: small areas

Small Areas



Ware, “Information Visualization”

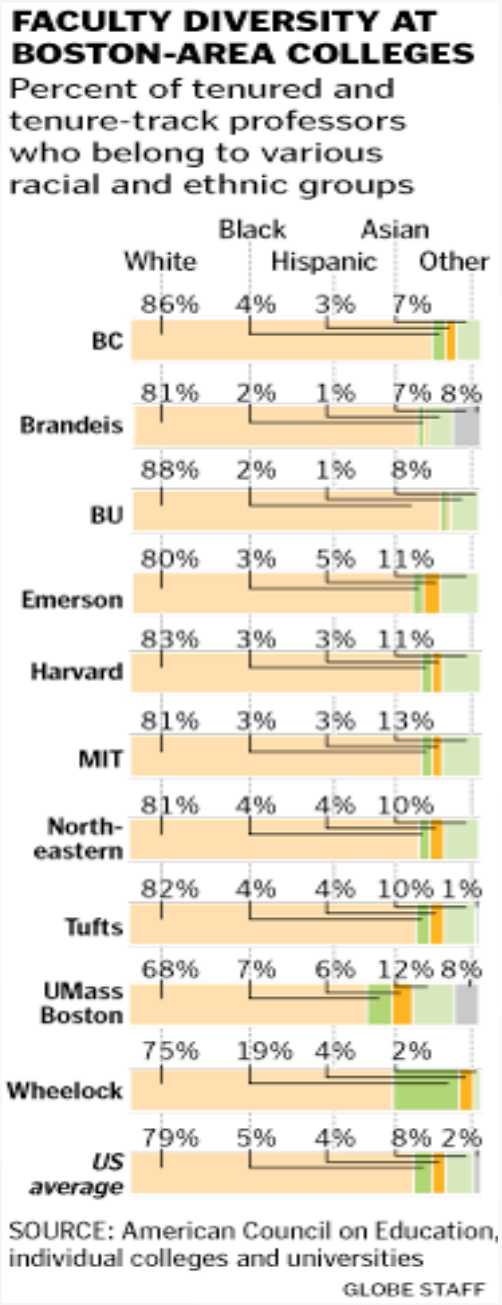
Pop-Out



Hue and lightness

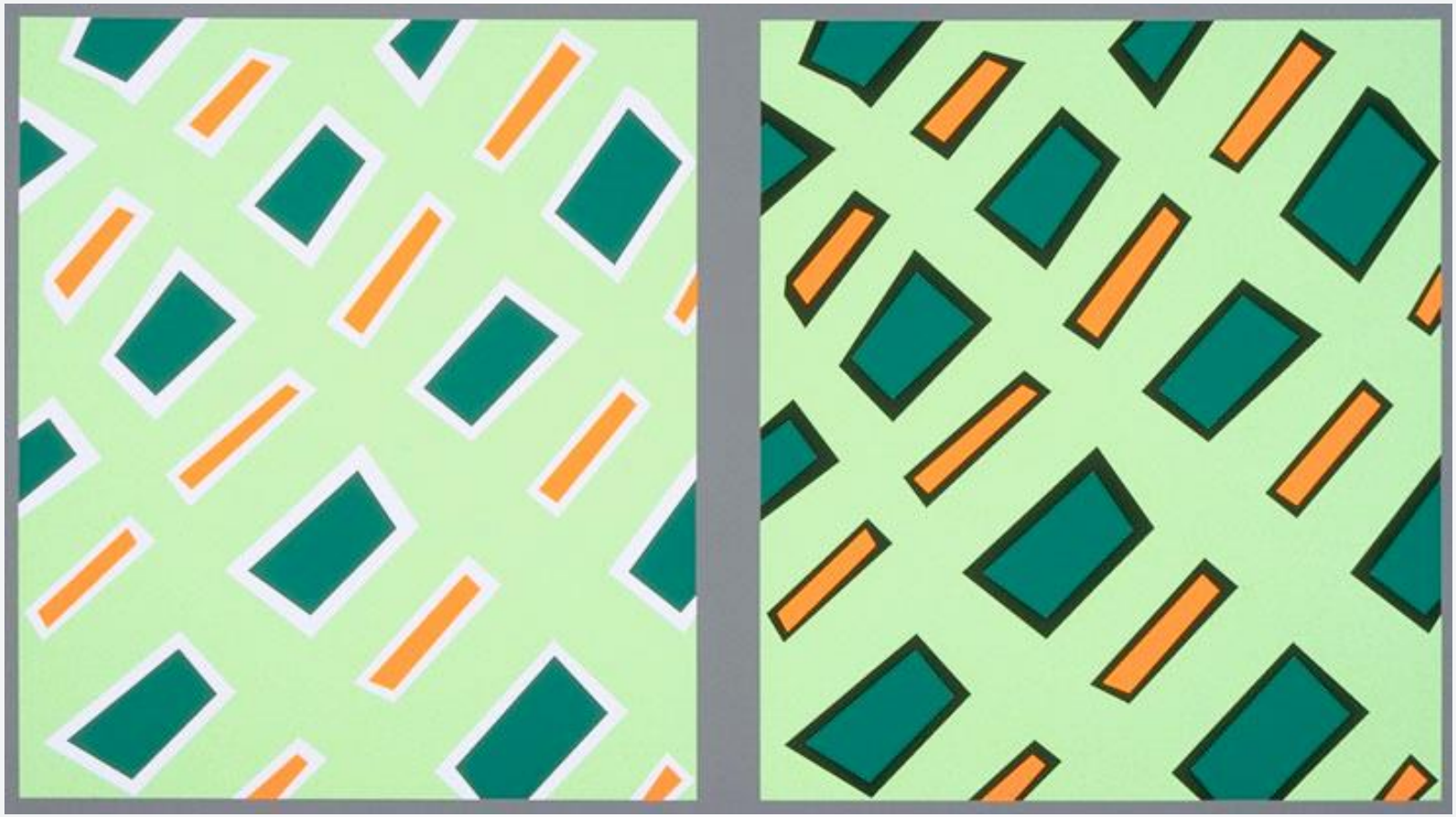
Lightness only

Based on slide from Stone



Boston Globe, Feb 16, 2010

Bezold Spreading Effect



Based on slide from Stone

Highlighting

	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
red	25.37	13.70	0.05	26.27	14.13	0.04	18.41	10.16	0.05	17.43	9.30	0.00
green	22.14	51.24	0.35	20.68	49.17	0.44	21.11	46.00	0.20	16.36	37.95	0.12
blue	13.17	3.71	74.89	15.38	5.20	86.83	11.55	3.37	65.53	9.96	3.44	56.14
gray	63.46	73.30	78.05	64.66	71.99	90.08	52.96	62.49	67.99	45.54	53.65	58.14
black	0.66	0.70	0.77	0.63	0.66	1.09	0.47	0.58	0.70	0.44	0.54	0.71

	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
red	25.37	13.70	0.05	26.27	14.13	0.04	18.41	10.16	0.05	17.43	9.30	0.00
green	22.14	51.24	0.35	20.68	49.17	0.44	21.11	46.00	0.20	16.36	37.95	0.12
blue	13.17	3.71	74.89	15.38	5.20	86.83	11.55	3.37	65.53	9.96	3.44	56.14
gray	63.46	73.30	78.05	64.66	71.99	90.08	52.96	62.49	67.99	45.54	53.65	58.14
black	0.66	0.70	0.77	0.63	0.66	1.09	0.47	0.58	0.70	0.44	0.54	0.71

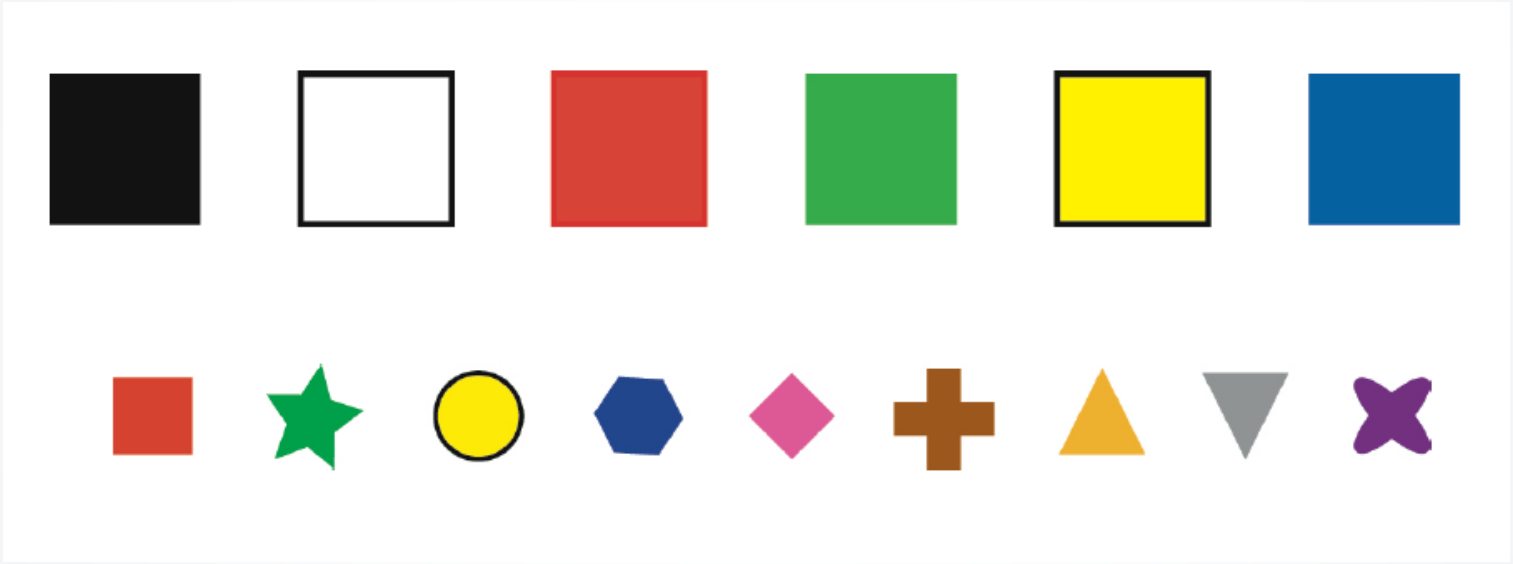
Based on slide from Stone

Facts

Color in small regions is difficult to perceive, and bright colors in large areas appear bigger

Use bright, saturated colors for small regions, and use low saturation pastel colors for large regions and backgrounds.

Colors for Categories



Trail Rating	Symbol	Level of difficulty
Green circle		Easiest
Blue square		Intermediate
Black diamond		Difficult
Double black diamond		Expert
Variations		Various
Terrain parks		Various

Tableau Colors

www.tableausoftware.com

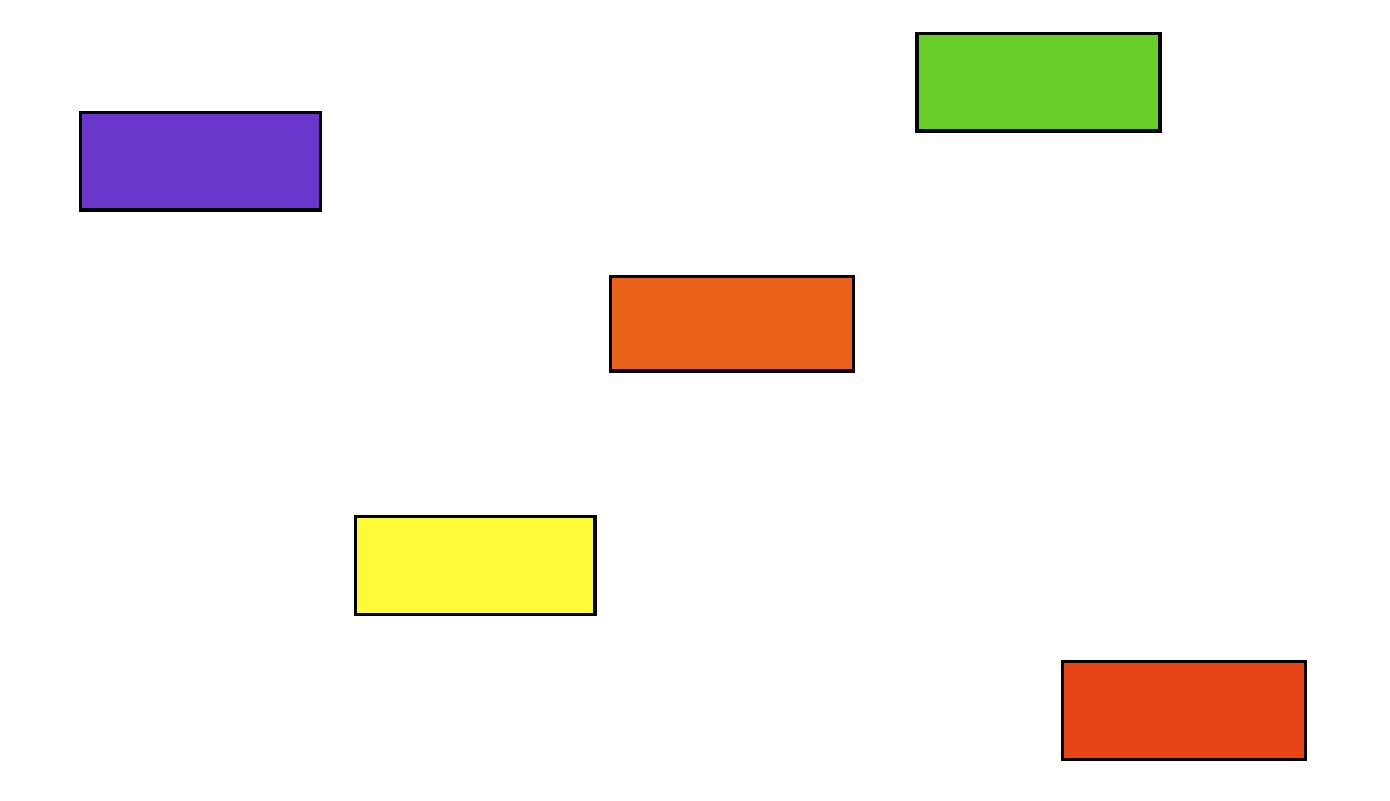
Facts

Only a small number of colors can be used effectively as nominal labels.

- Keep the number of colors for nominal data to less than eight.
- Use quiet medium grey backgrounds.

Color: Ordinal

Order These Colors

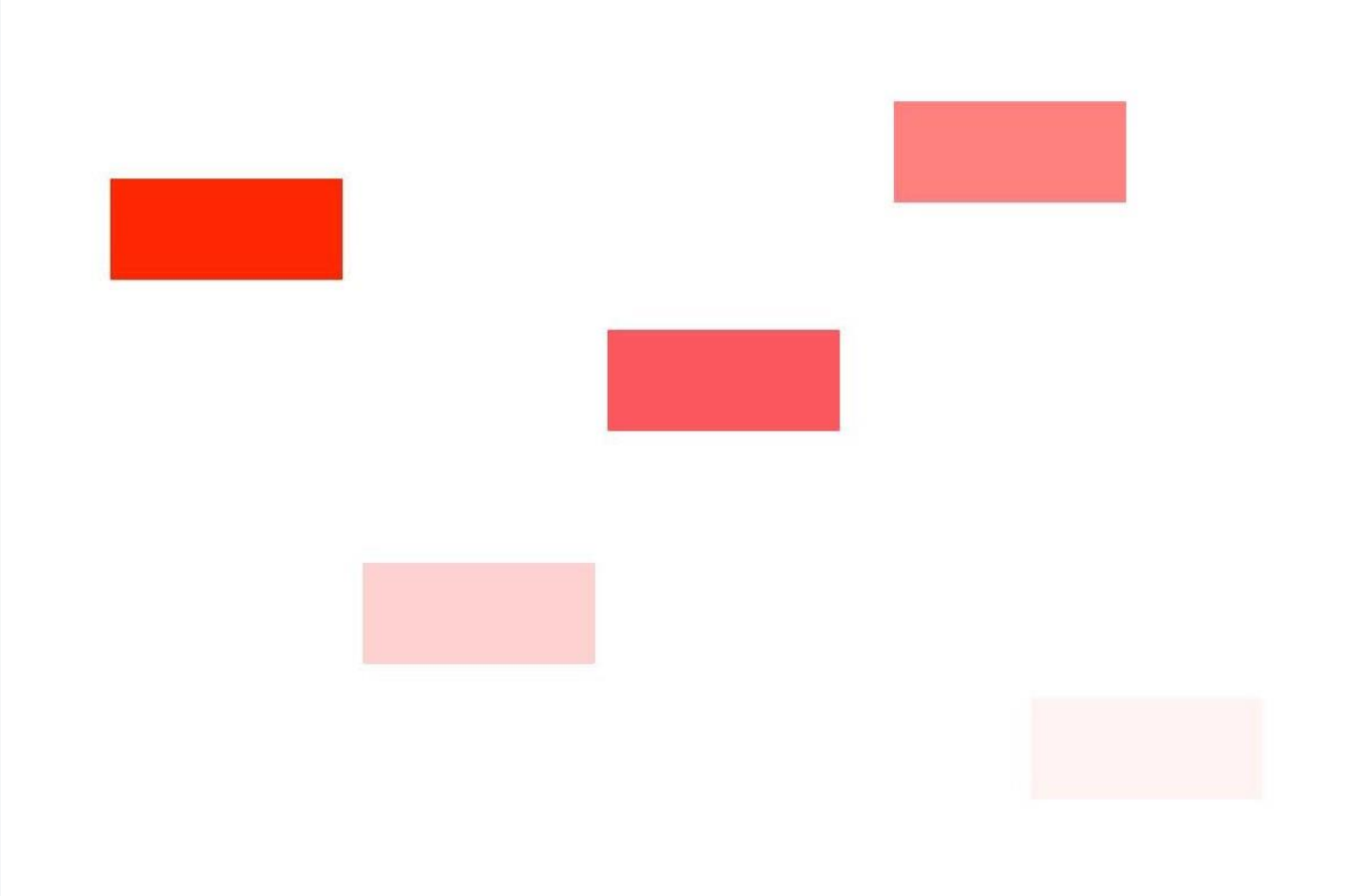


Based on slide from Stasko

Order These Colors

Based on slide from Stasko

Order These Colors

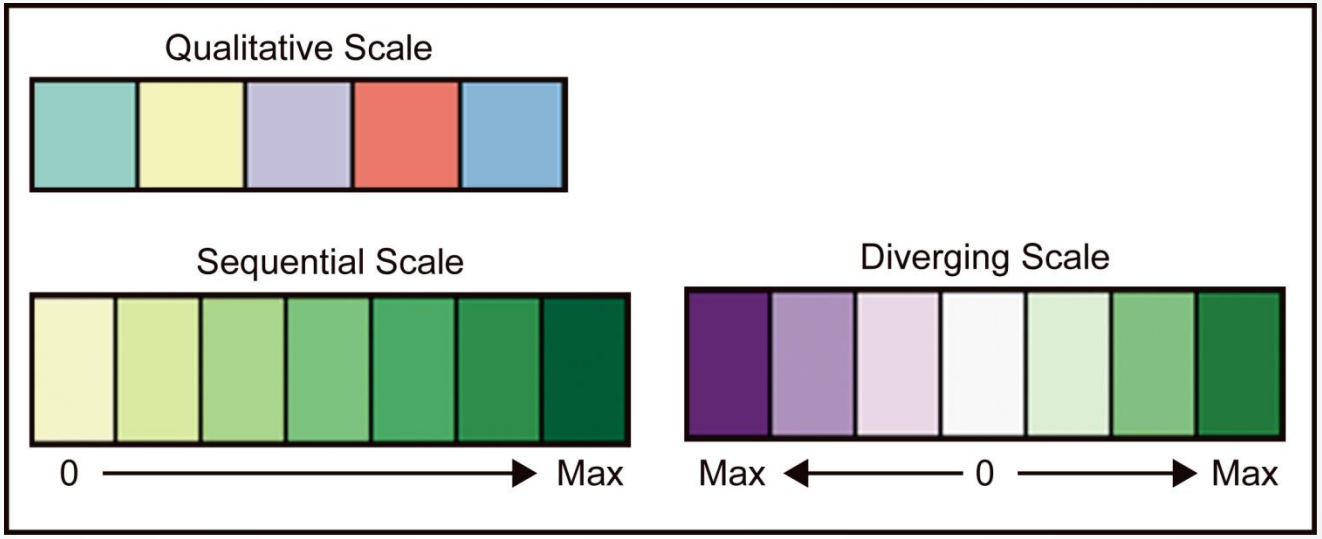


Based on slide from Stasko

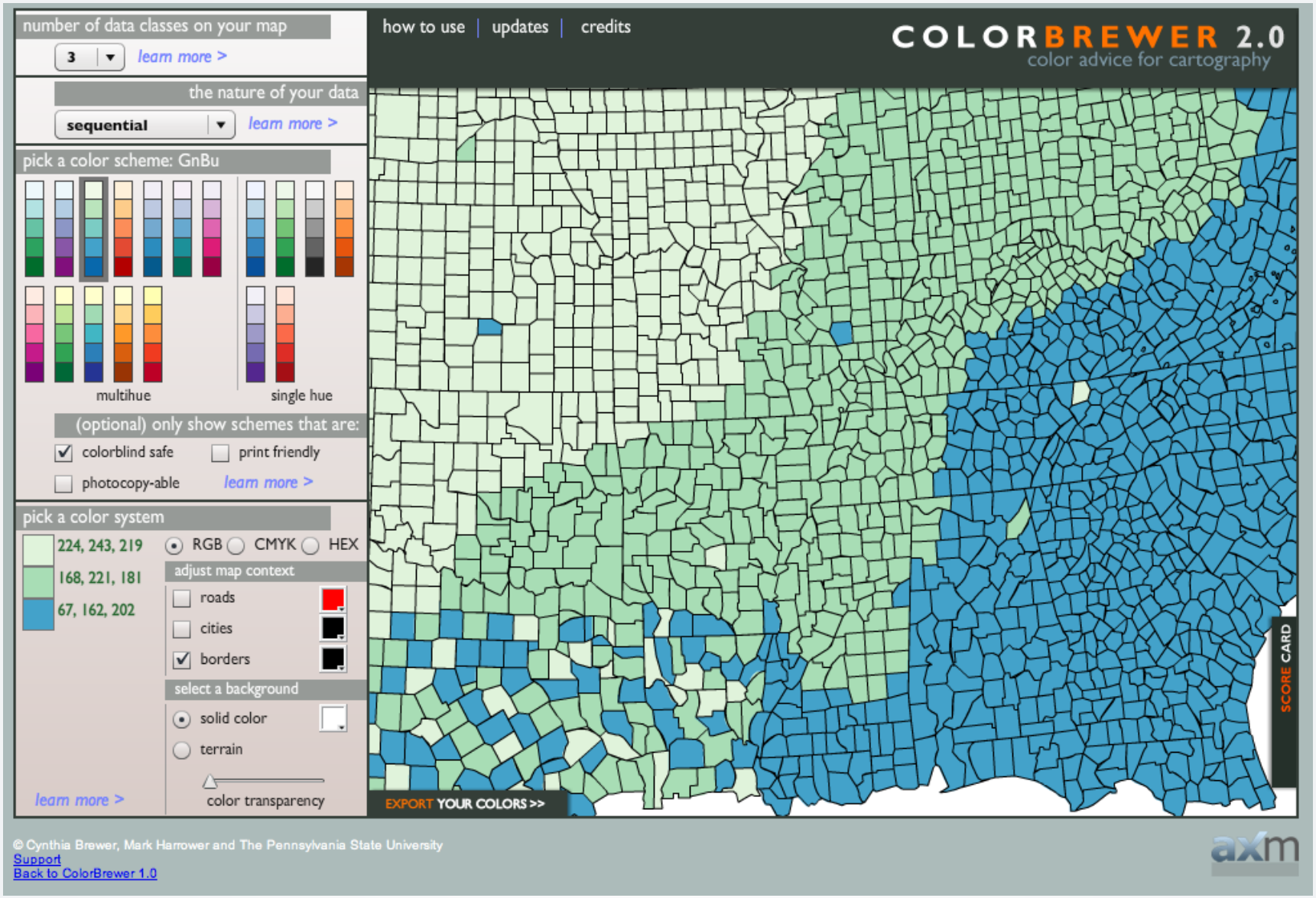
Brewer Scales

Nominal

Ordinal



Cynthia Brewer, Color Use Guidelines for Data Representation



© Cynthia Brewer, Mark Harrower and The Pennsylvania State University
Support
Back to ColorBrewer 1.0

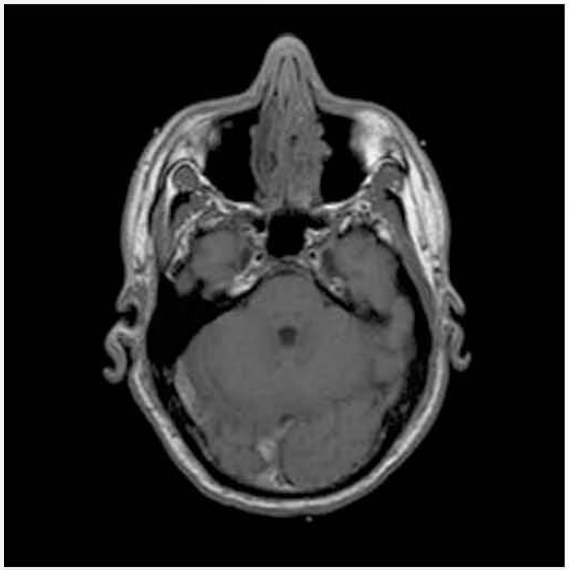
Facts

Lightness and saturation are effective for ordinal data because they have an implicit perceptual ordering

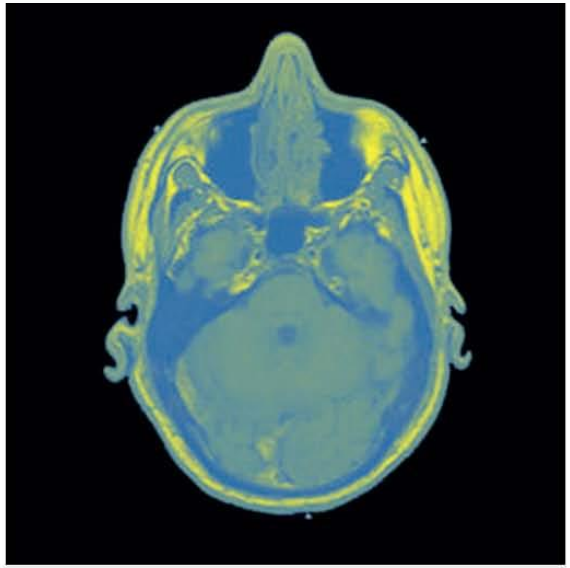
Show ordinal data with a discrete set of color values that change in lightness or saturation

Color: Quantitative

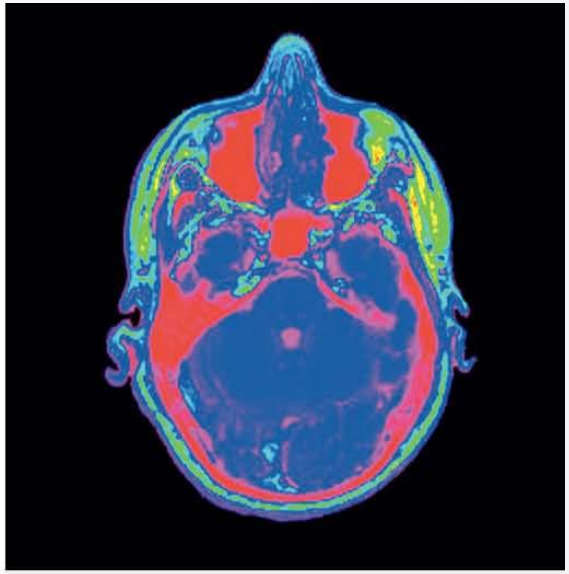
Colormaps



Lightness scale



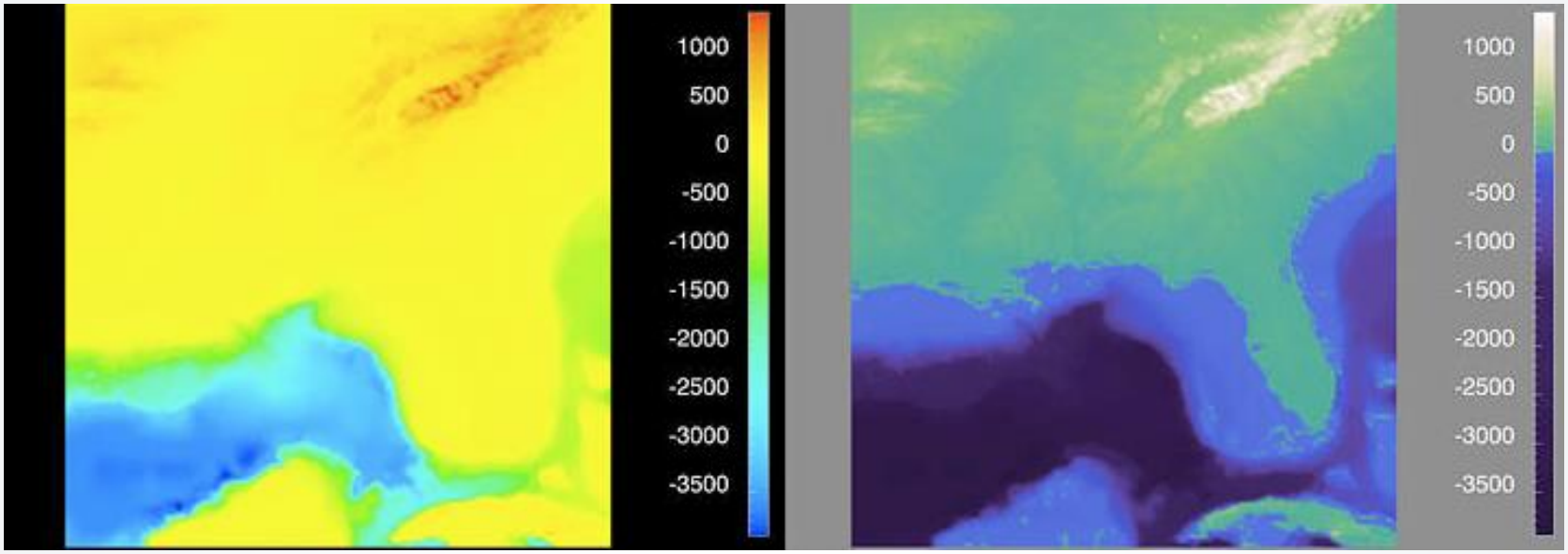
Lightness scale
with hue and
chroma variation



Hue scale with
lightness variation

After slide from M. stone

Rainbow Colormap



Rogowitz and Treinish, Why should engineers and scientists be worried about color?

Rainbow Colormap

- Hue is used to show ordinal data
- Not perceptually linear:
  - Equal steps in the continuous range are not perceived as equal steps
- Not good for colorblind people

Facts

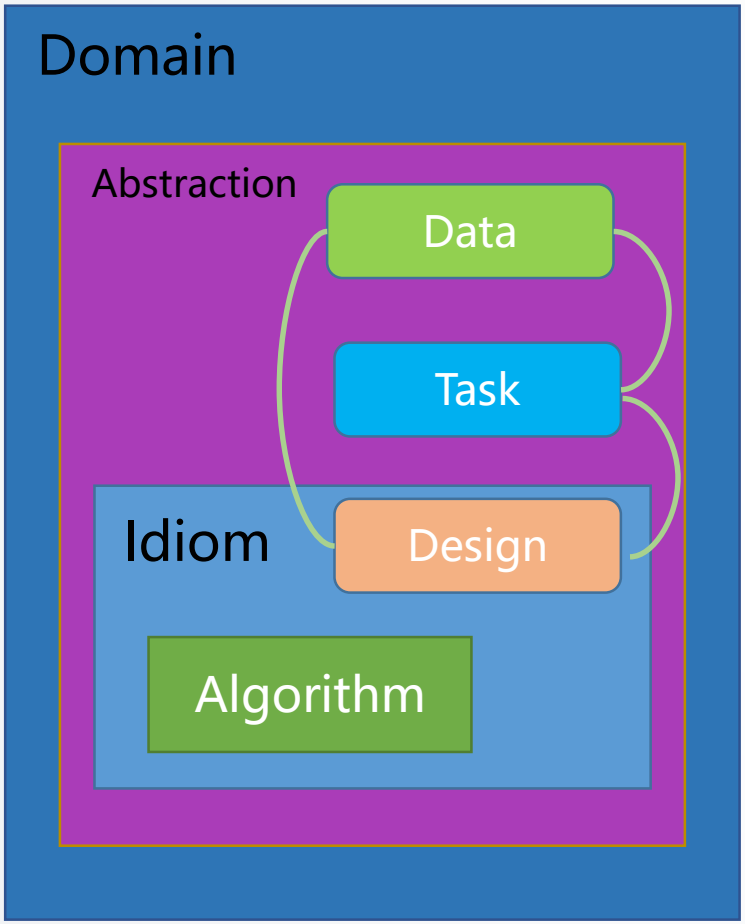
Quantitative data can be shown with a discrete or continuous colormap

- Use colormaps with a limited hue palette and redundantly vary lightness and saturation.
- Use discrete colormaps for accuracy.

OUTLINE

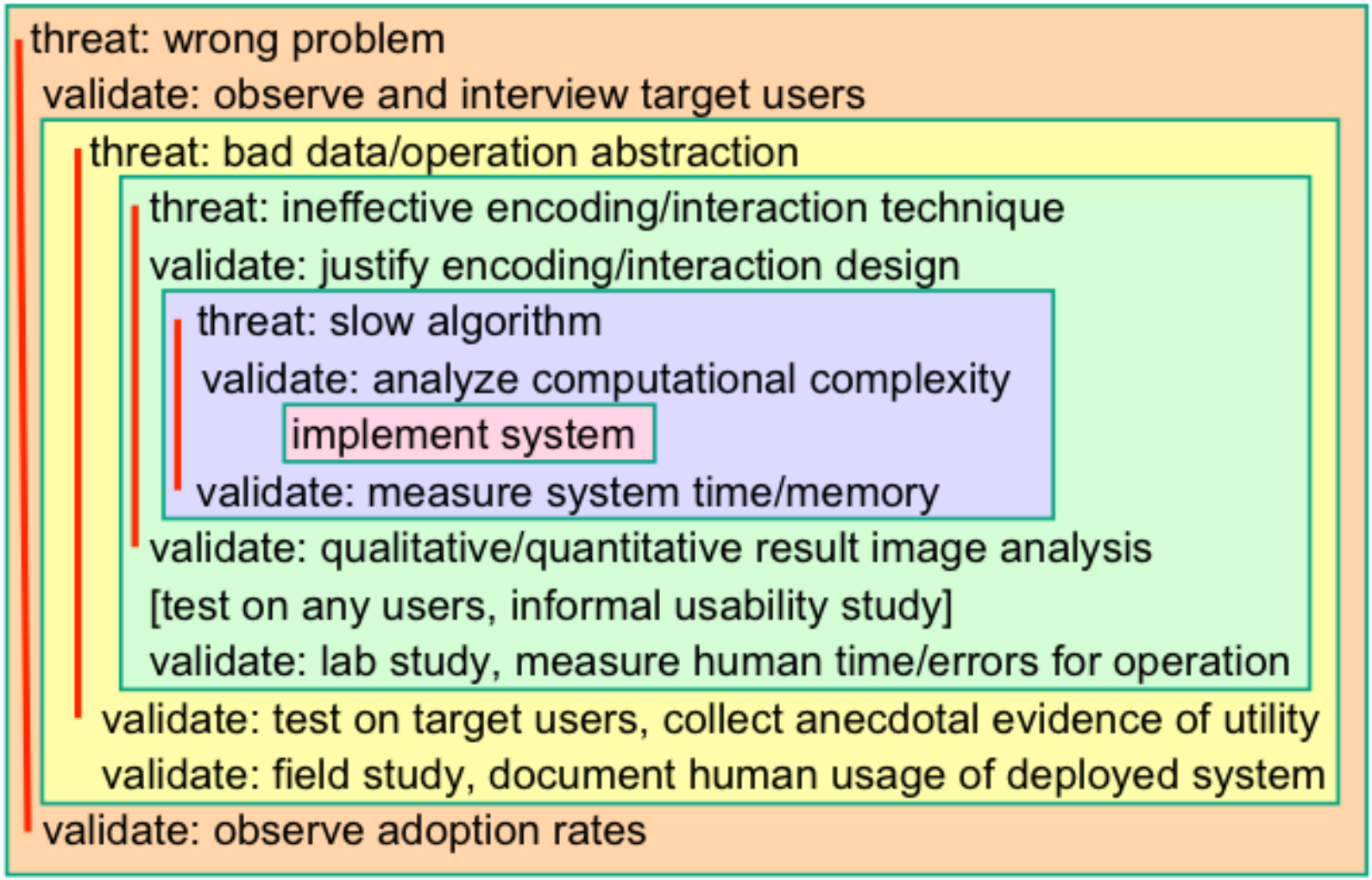
1. Visualization Process Model
2. Visual Encoding Principles
3. Visual Analysis Model

The Nested Model

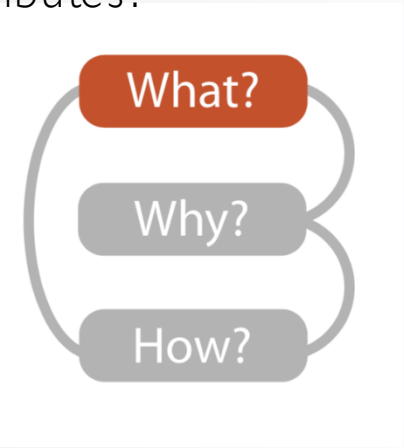


- Domain – who will use it?
- Abstract – translate to vis
  - What? Abstraction of data
  - Why? Abstraction of tasks
- Idiom – How?
  - Visual encoding
  - Visual interaction
- Algorithm – efficiency

The Nested Model



What can be visualized: data, datasets, and attributes.



What?

Datasets

Data Types

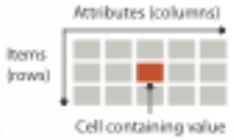
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

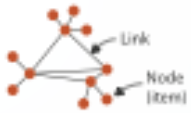
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

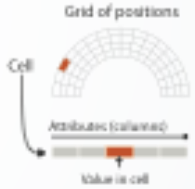
- Tables



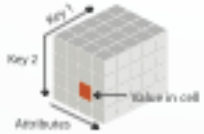
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic

Dataset Availability

- Static

- Dynamic

Data types

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

Data and Dataset Types

Data Types

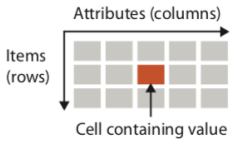
→ Items → Attributes → Links → Positions → Grids

Data and Dataset Types

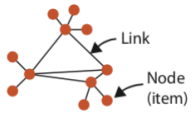
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

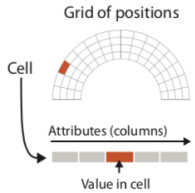
→ Tables



→ Networks

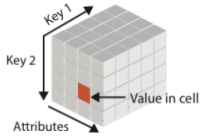


→ Fields (Continuous)



→ Geometry (Spatial)

→ *Multidimensional Table*



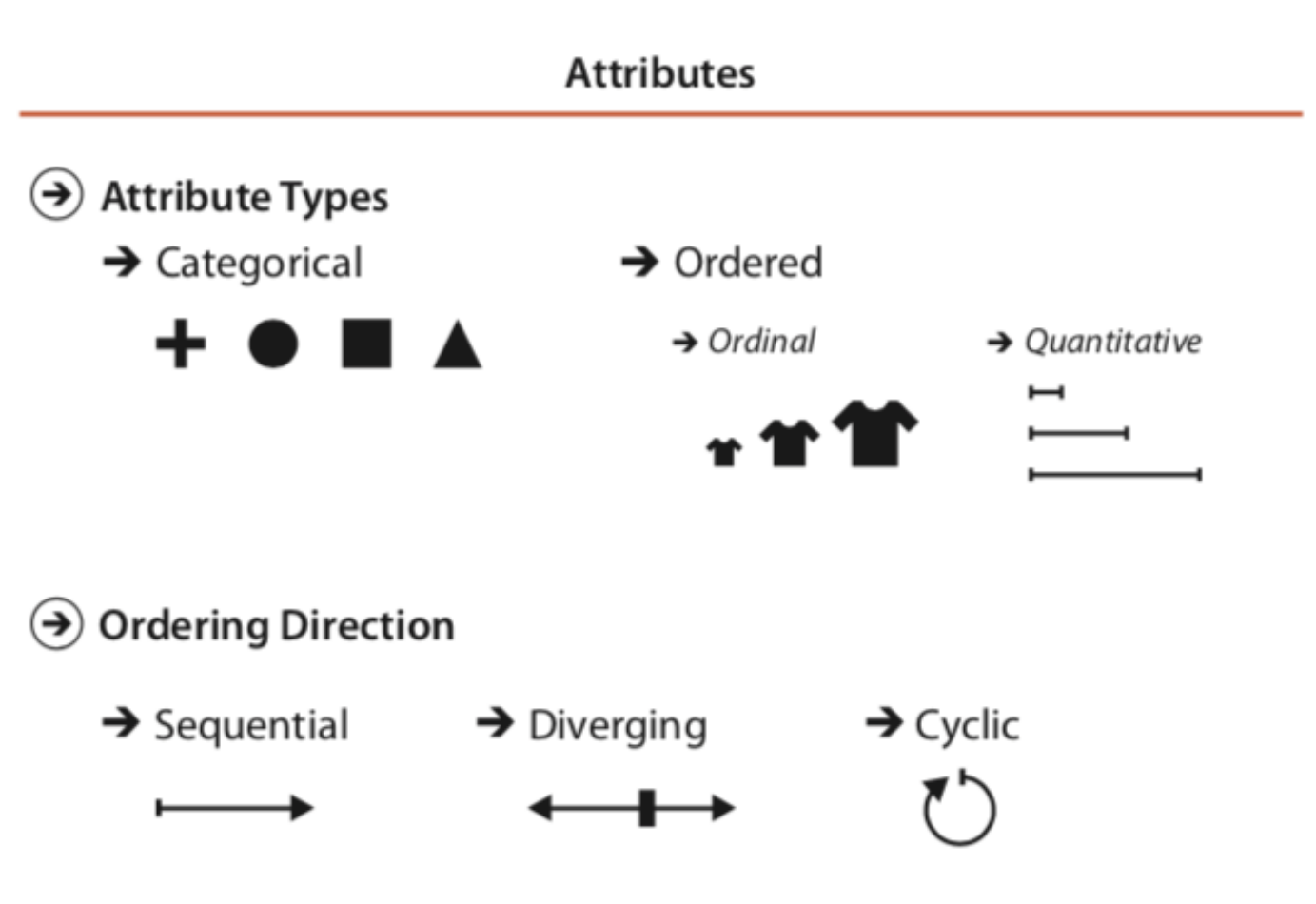
→ *Trees*

Dataset Availability

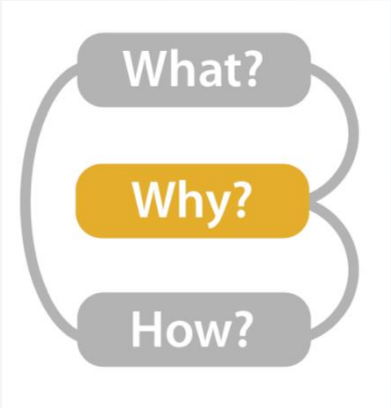
→ Static

→ Dynamic

Attribute types



Why people are using vis in terms of actions and targets.



- {action, target} pair
  - discover distributior
  - compare trends
  - locate outliers
  - browse topology

Why?

Actions

Analyze

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

Search

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

Query

- Identify
- Compare
- Summarize

Targets

All Data

- Trends
- Outliers
- Features

Attributes

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

Network Data

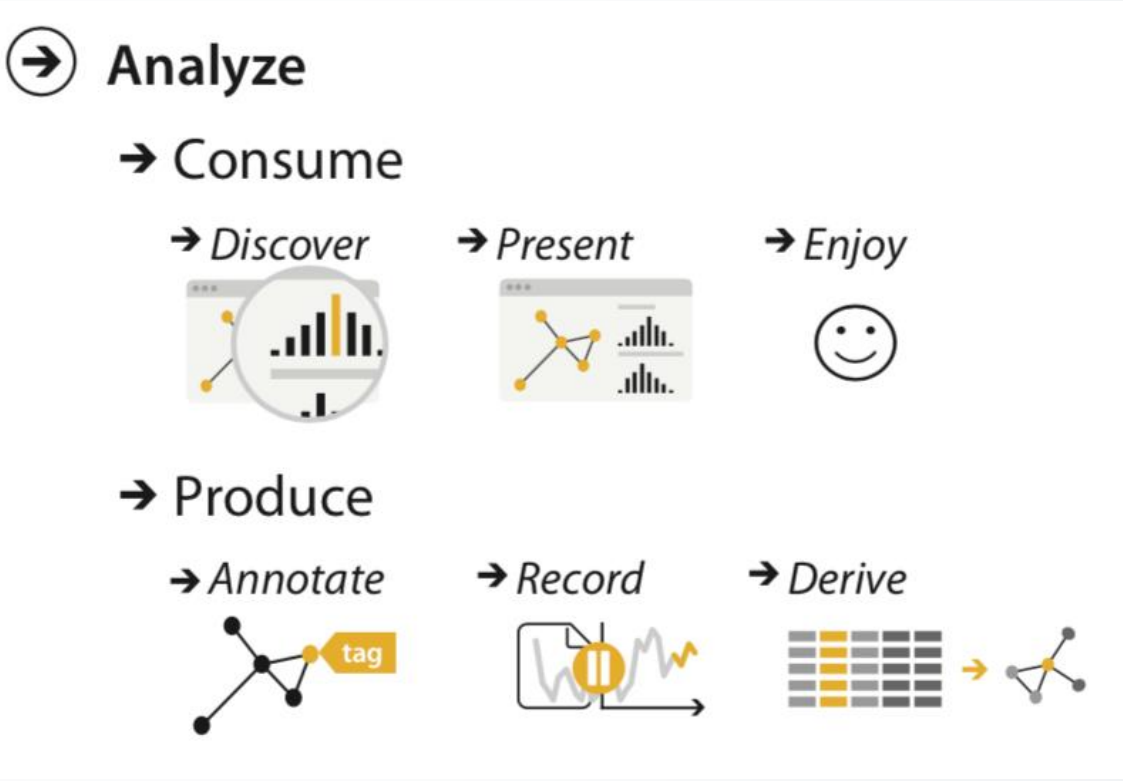
- Topology
- Paths

Spatial Data

- Shape

Actions: Analyze

- Consume
  - Discover vs present
    - classic split
    - aka explore vs explain
  - Enjoy
    - newcomer
    - aka casual, social
- Produce
  - Annotate, record
  - Derive
    - crucial design choice

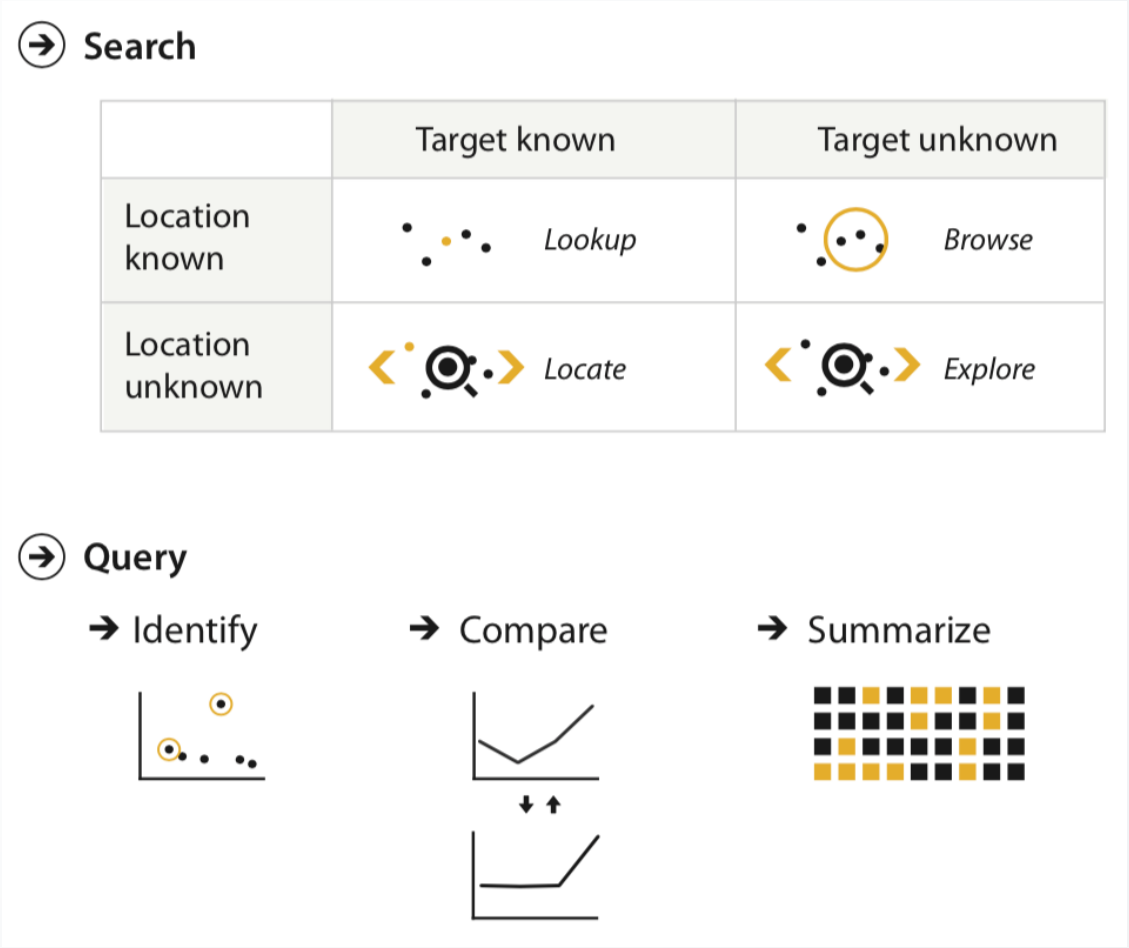


Derive

- 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

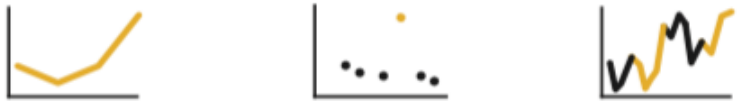
Actions: Search, query

- 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



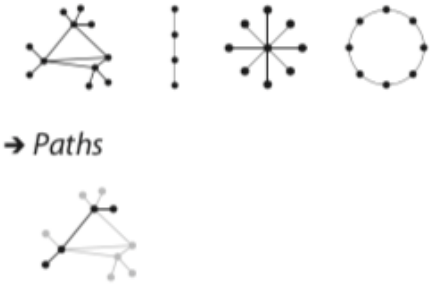
Why: Targets

- **All Data**
  - Trends
  - Outliers
  - Features

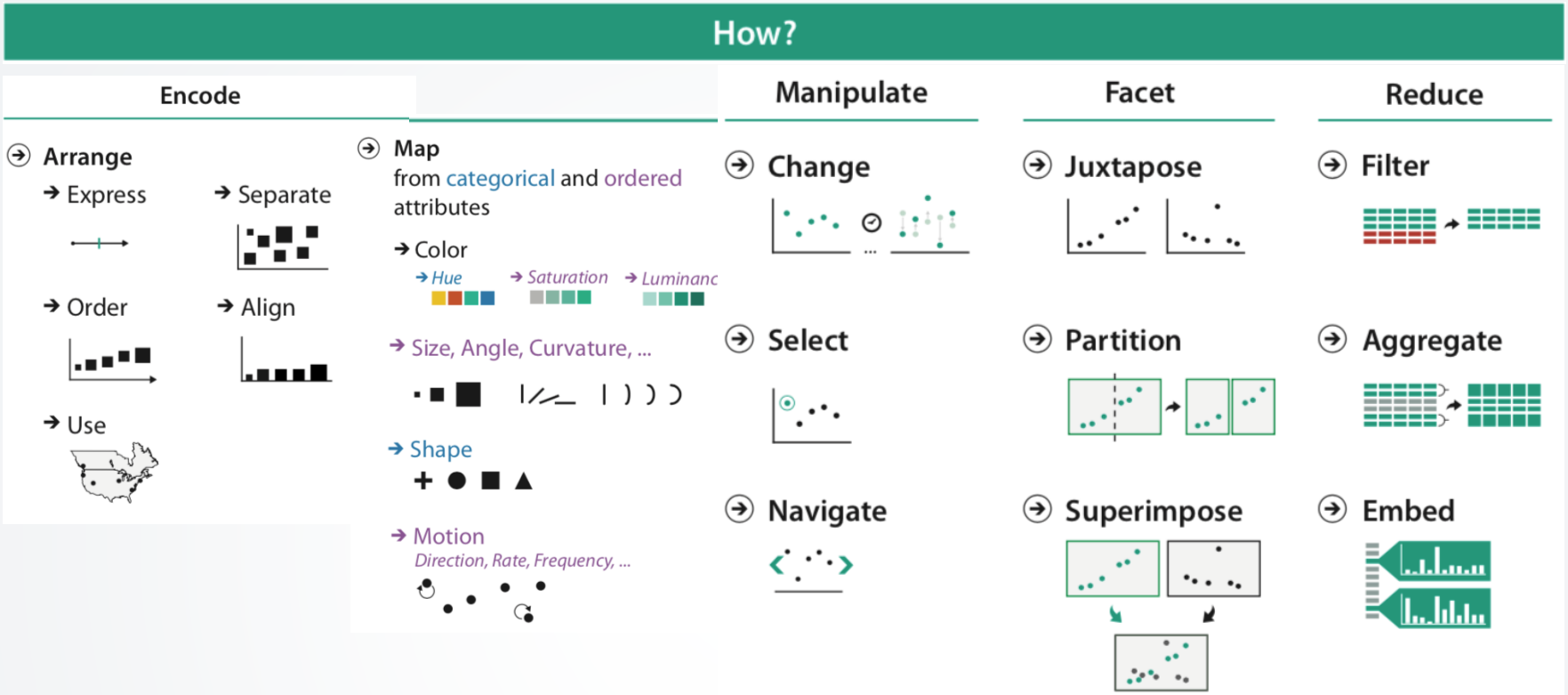


- **Attributes**
  - One
    - *Distribution*
    - *Extremes*
  - Many
    - *Dependency*
    - *Correlation*
    - *Similarity*

- **Network Data**
  - Topology
    - *Paths*

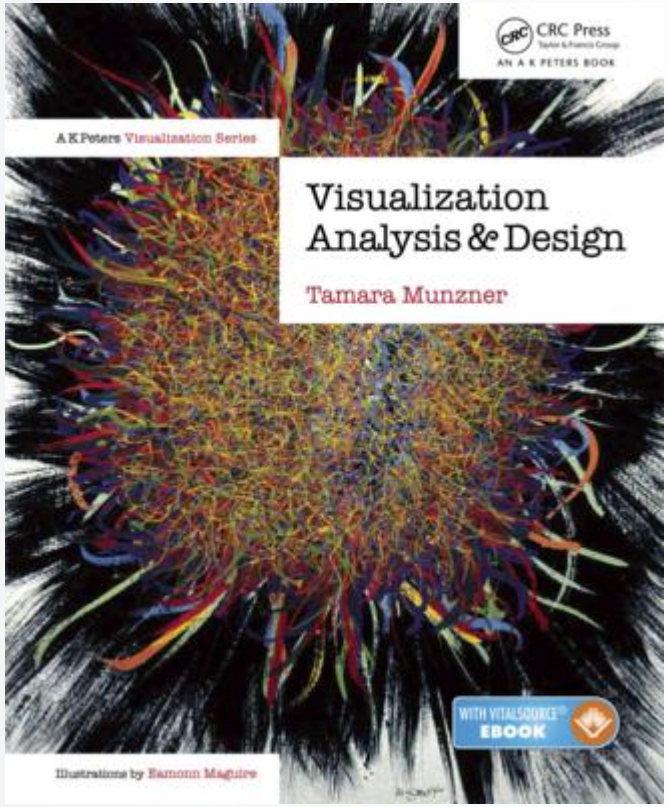
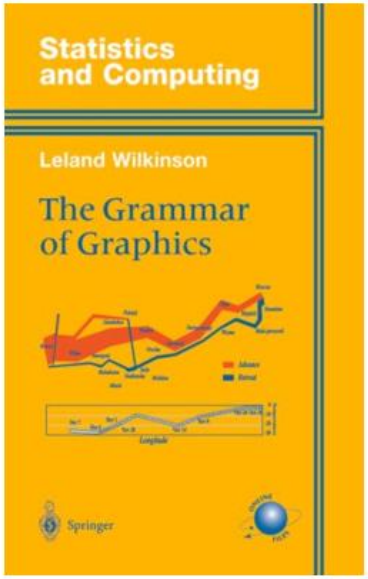


- **Spatial Data**
  - Shape



How to design vis idioms: encode, manipulate, facet, and reduce

Suggest Readings



Wilkonson's GoG