



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](mailto:ycwu@zju.edu.cn)

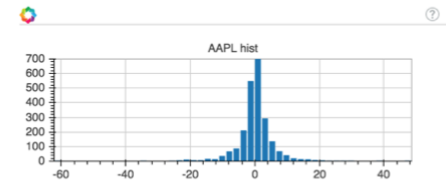
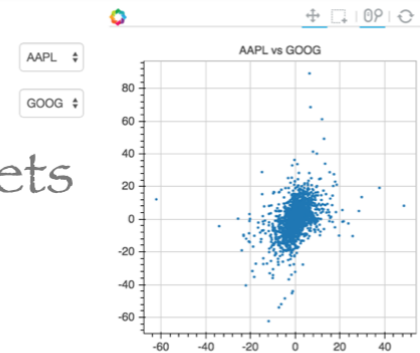
Wei Chen

Professor

State Key Lab of CAD&CG

Email: [chenvis@zju.edu.cn](mailto:chenvis@zju.edu.cn)

Widgets

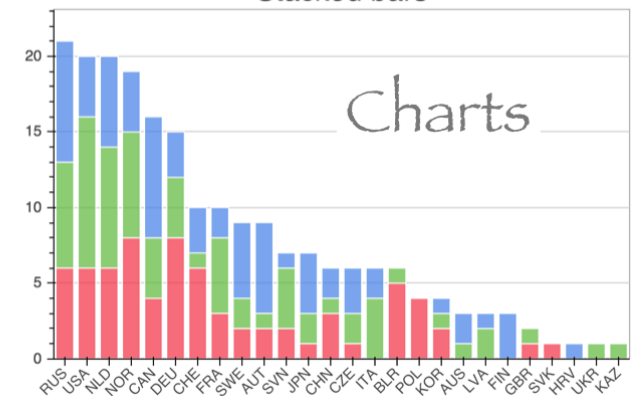


Bokeh

Tools



Stacked bars



Charts

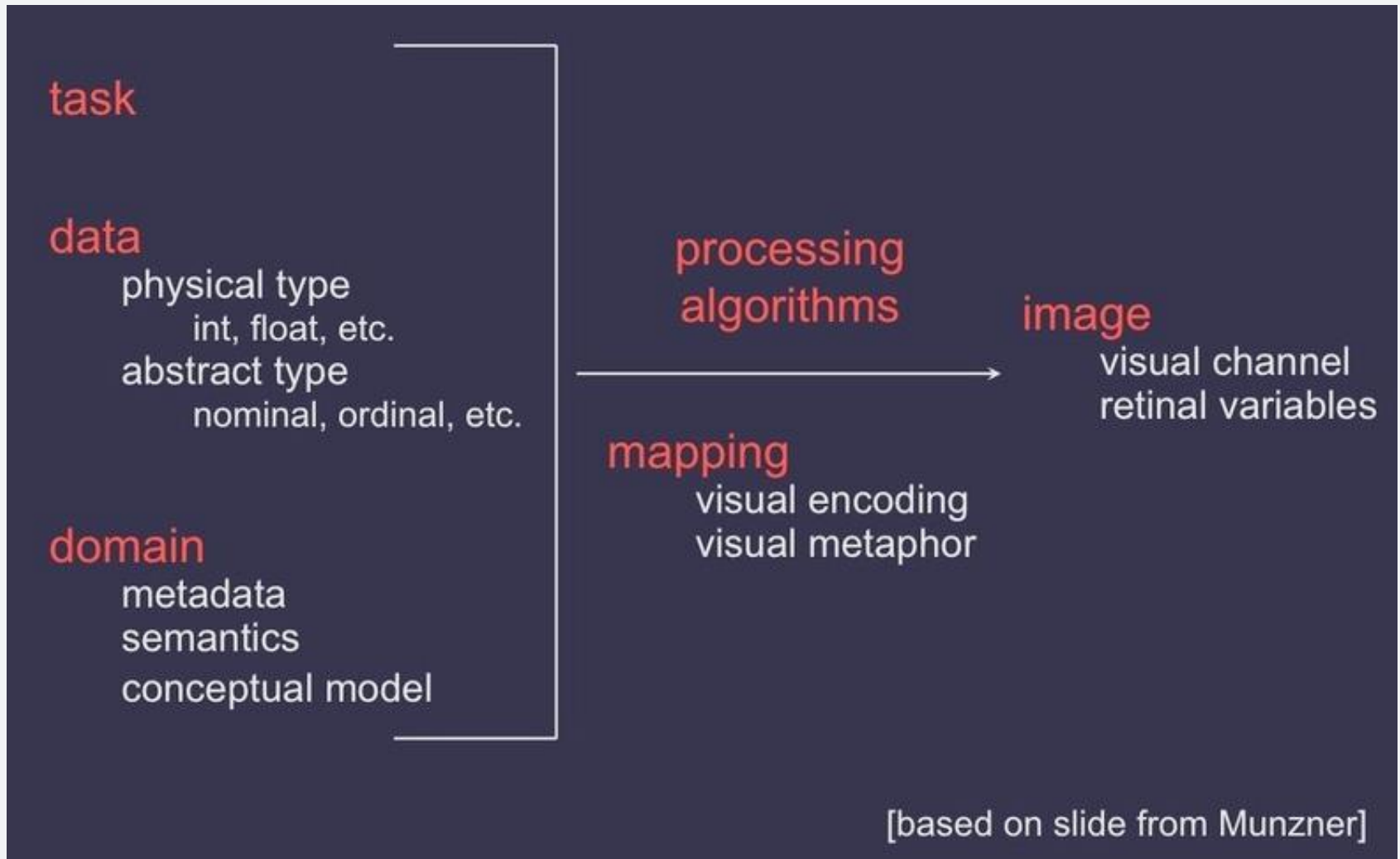
# OUTLINE

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

# OUTLINE

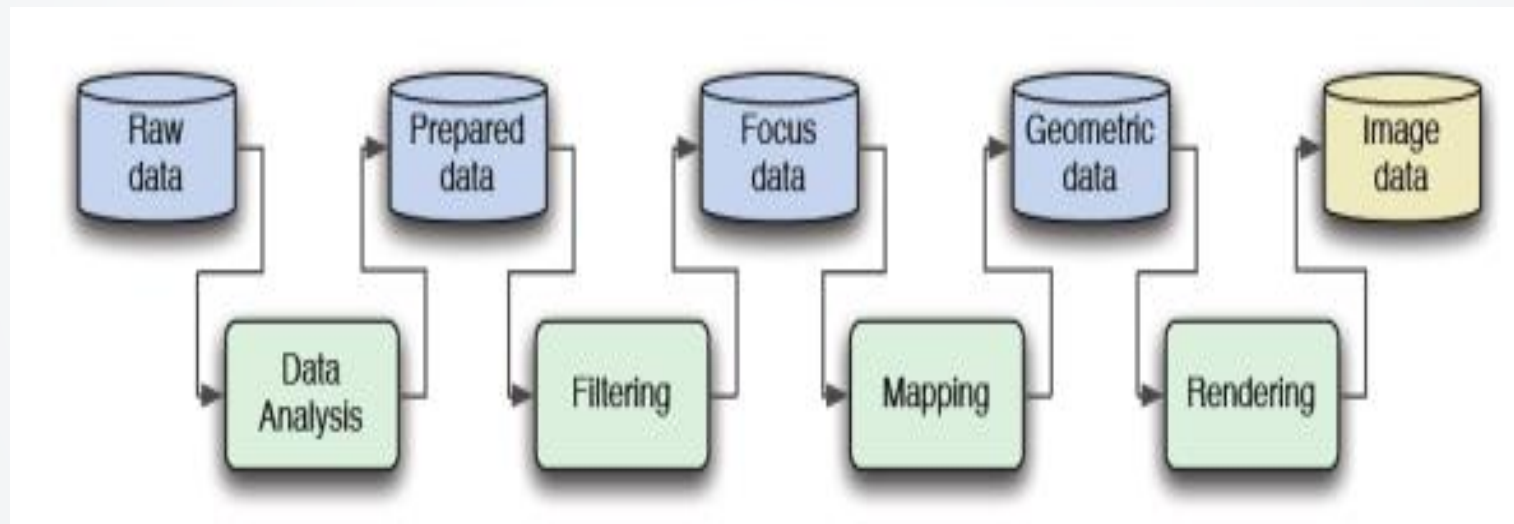
1. Visualization Process Model
2. Visual Encoding Principles
3. Visual Analysis Model

# Munzner's Model



# A Conceptual Model

Visualization idioms by Haber and McNabb

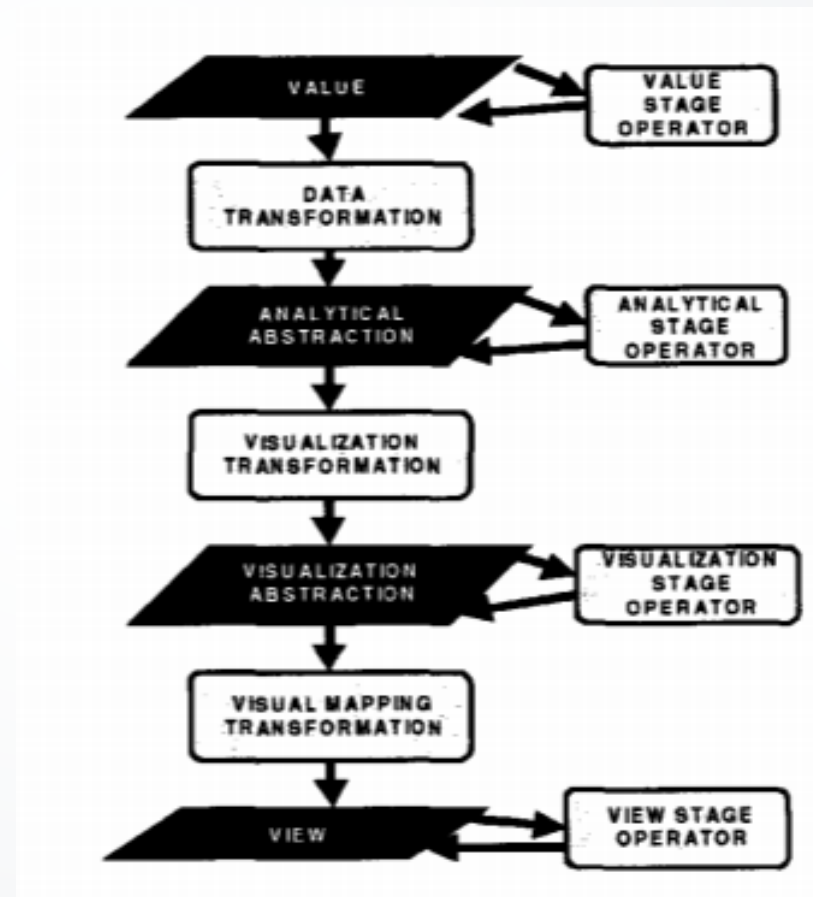


Haber, R. B. and McNabb, D. A. Visualization idioms:  
A conceptual model for scientific visualization systems,  
1990.

# Data State Reference Model

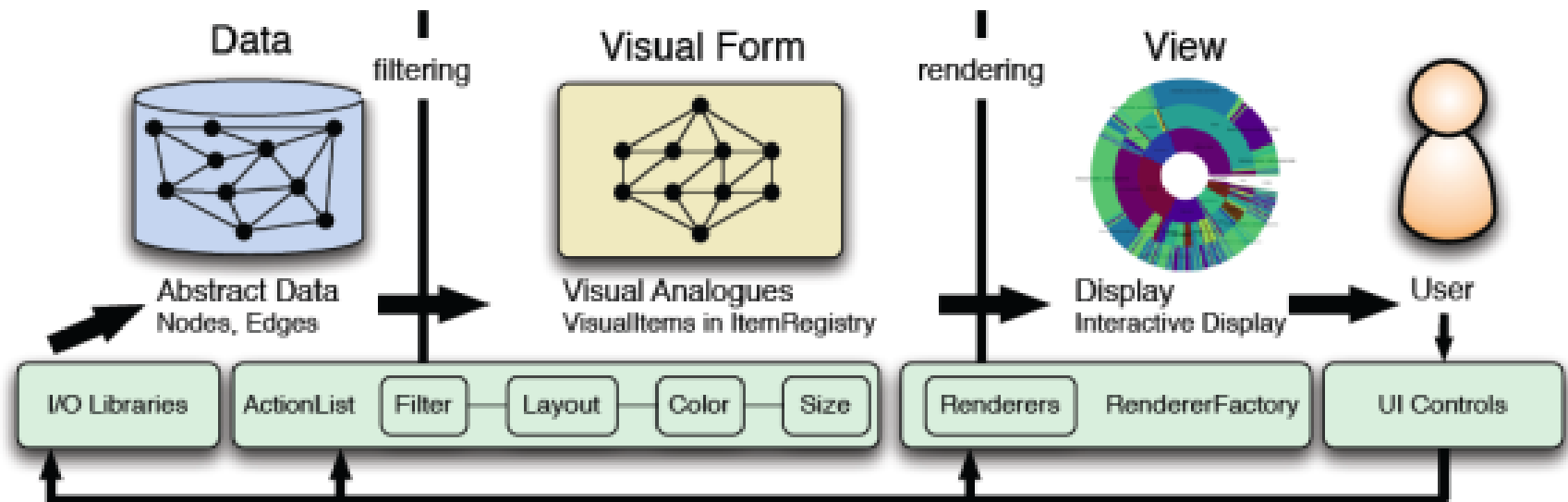
Date state reference model by Ed H. Chi

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.

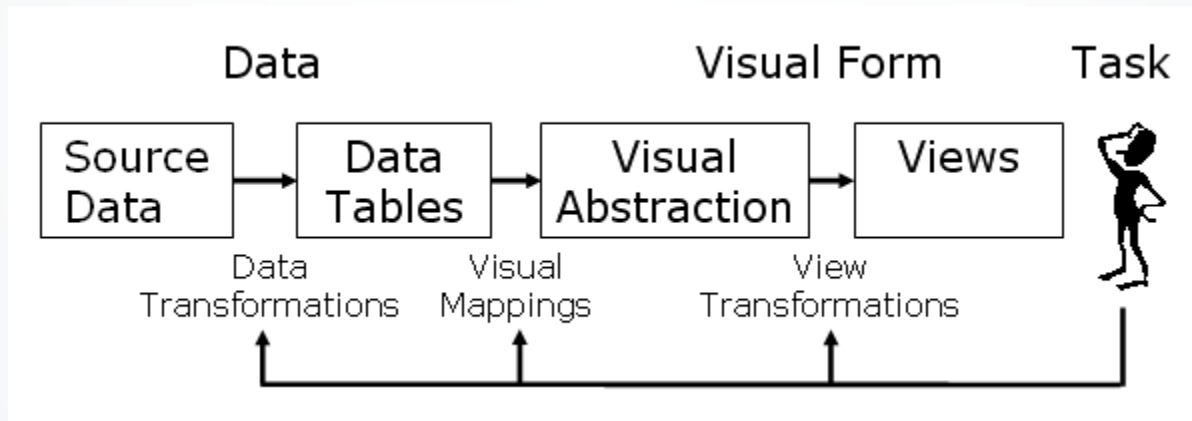


Diagram depicting the information visualization reference model.

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



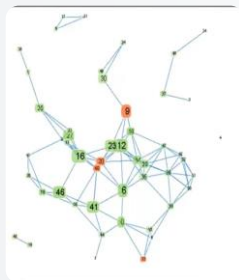
Raw data

Data  
Transformation

Visual  
Mapping

Visualization

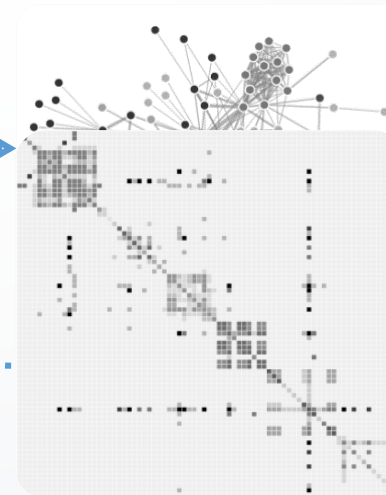
Change layout



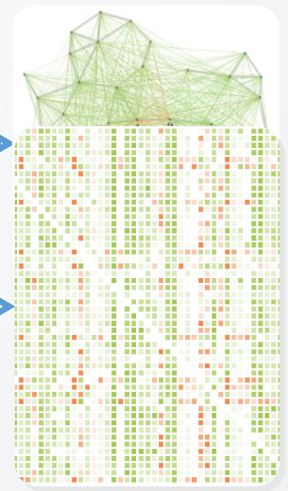
Ad-hoc network

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20	x21	-
x1	3	72	99	96	78	42	44	97	2	39	85	50	52	77	32	29	38	2	14	96	22	
x2	101	61	62	60	63	6	52	19	18	40	28	51	30	94	64	80	66	70	71	83	76	
x3	60	98	29	74	32	40	96	26	83	37	95	27	89	21	68	41	93	15	25	2	84	
x4	12	55	76	97	4	37	9	40	40	16	32	82	12	26	29	96	100	94	97	97	96	
x5	44	63	46	30	44	24	93	7	50	44	30	25	12	8	2	100	18	5	9	30	14	
x6	4	87	32	9	9	29	6	87	100	58	22	40	77	23	52	81	59	40	7	73	70	
x7	12	76	94	90	16	29	43	43	88	32	14	96	47	33	94	18	83	5	74	11	50	
x8	46	39	75	59	60	39	56	39	24	1	83	1	46	87	12	45	33	13	34	77	9	
x9	11	47	19	8	51	4	96	23	95	4	32	43	21	43	49	90	76	97	95	50	30	
x10	23	35	84	94	27	26	29	25	59	33	32	94	77	47	44	52	69	93	69	12	14	
x11	85	13	38	3	56	29	92	45	89	73	10	76	30	12	83	101	21	80	46	19	82	
x12	49	33	36	1	89	85	66	76	11	11	30	41	59	74	11	52	38	19	1	39	95	
x13	92	46	37	63	44	12	64	66	6	87	32	35	20	43	135	35	77	84	17	58	42	
x14	99	7	46	69	93	96	65	58	54	14	8	30	35	89	77	64	53	87	78	11	47	
x15	61	46	17	71	81	84	19	63	20	101	39	74	53	57	19	91	28	63	65	94	62	
x16	58	80	45	29	60	85	59	41	75	36	84	43	77	7	23	55	15	86	95	74	11	
x17	19	95	23	47	88	20	11	82	60	12	97	95	90	90	7	71	35	41	99	68	7	
x18	7	72	5	83	4	88	87	69	47	19	52	80	85	16	93	90	94	92	89	10	71	
x19	58	50	41	87	32	50	90	1	69	83	78	98	77	74	29	60	6	29	33	14	38	
x20	70	26	34	41	56	100	93	27	61	97	39	19	95	10	24	96	1	62	90	18	63	
x21	66	80	36	25	91	2	8	6	55	71	17	84	42	59	84	90	71	95	65	56	74	

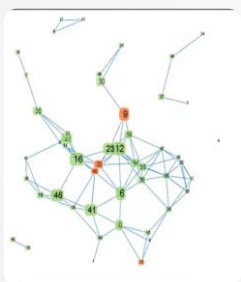
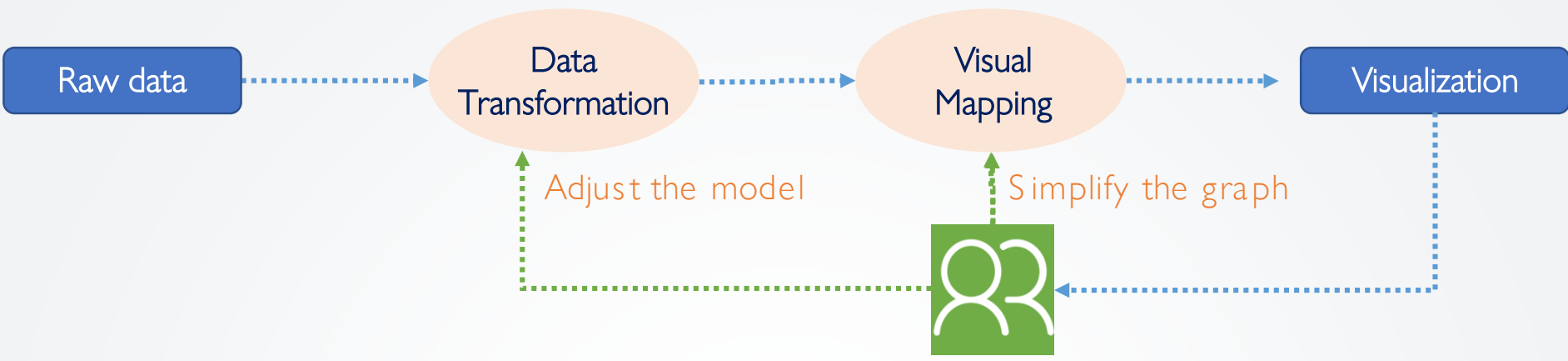
Inter-node trust  
relation



Matrix



Matrix



Ad-hoc network

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20	x21	-
x1	3	72	89	96	78	42	44	97	2	39	85	50	52	77	32	29	38	2	14	96	22	
x2	101	61	62	60	63	6	52	19	18	40	28	51	30	94	64	80	66	70	71	83	76	
x3	60	98	29	74	32	40	96	26	83	37	95	27	89	21	68	41	93	15	25	2	84	
x4	12	55	75	87	4	37	9	40	16	32	82	12	26	29	86	100	94	87	87	86		
x5	44	63	46	30	44	24	93	7	50	44	30	35	12	8	2	100	18	5	9	30	14	
x6	4	87	32	9	9	29	6	87	100	58	22	40	77	23	32	81	39	60	7	73	70	
x7	12	76	94	94	16	29	43	63	88	32	14	98	47	33	94	18	83	5	74	11	50	
x8	46	39	75	59	60	39	56	39	24	1	83	1	46	87	12	45	33	13	34	77	9	
x9	11	47	19	8	51	4	96	23	95	4	32	43	21	43	49	90	76	97	95	50	30	

$$f(j) = \left( \frac{Trustee(j)_{n-1} + 1}{2} \right)^2$$

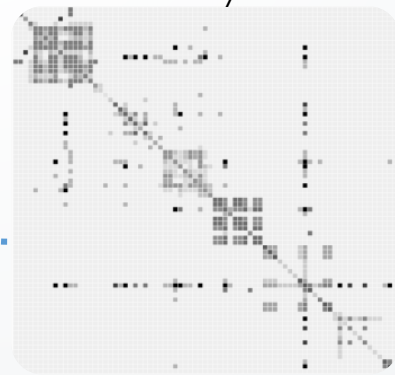
$$Trustee(i)_n = \sum_j^{j \neq i} \frac{f(j) * Trust[j][i]}{N - 1}$$

$$F_{size}(i) = \left( \frac{Trustee(i)_n - 1}{2} \right)^2$$

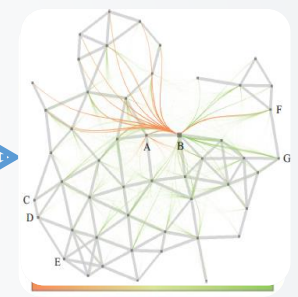
Relation-based reordering, clustering and



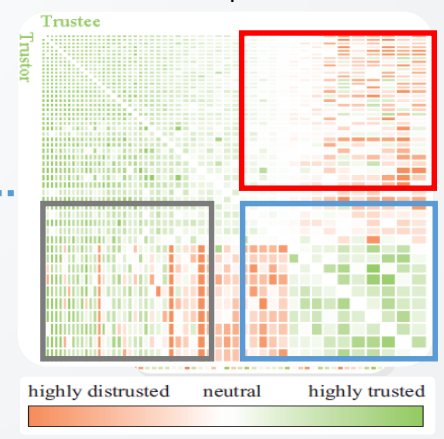
Graph layout



Matrix



Node-based simplification

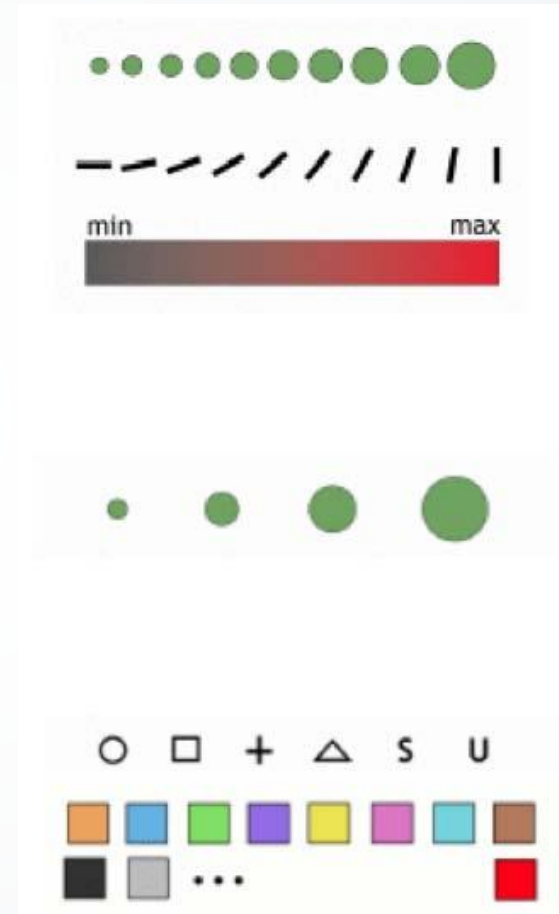


# OUTLINE

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 Pack	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
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

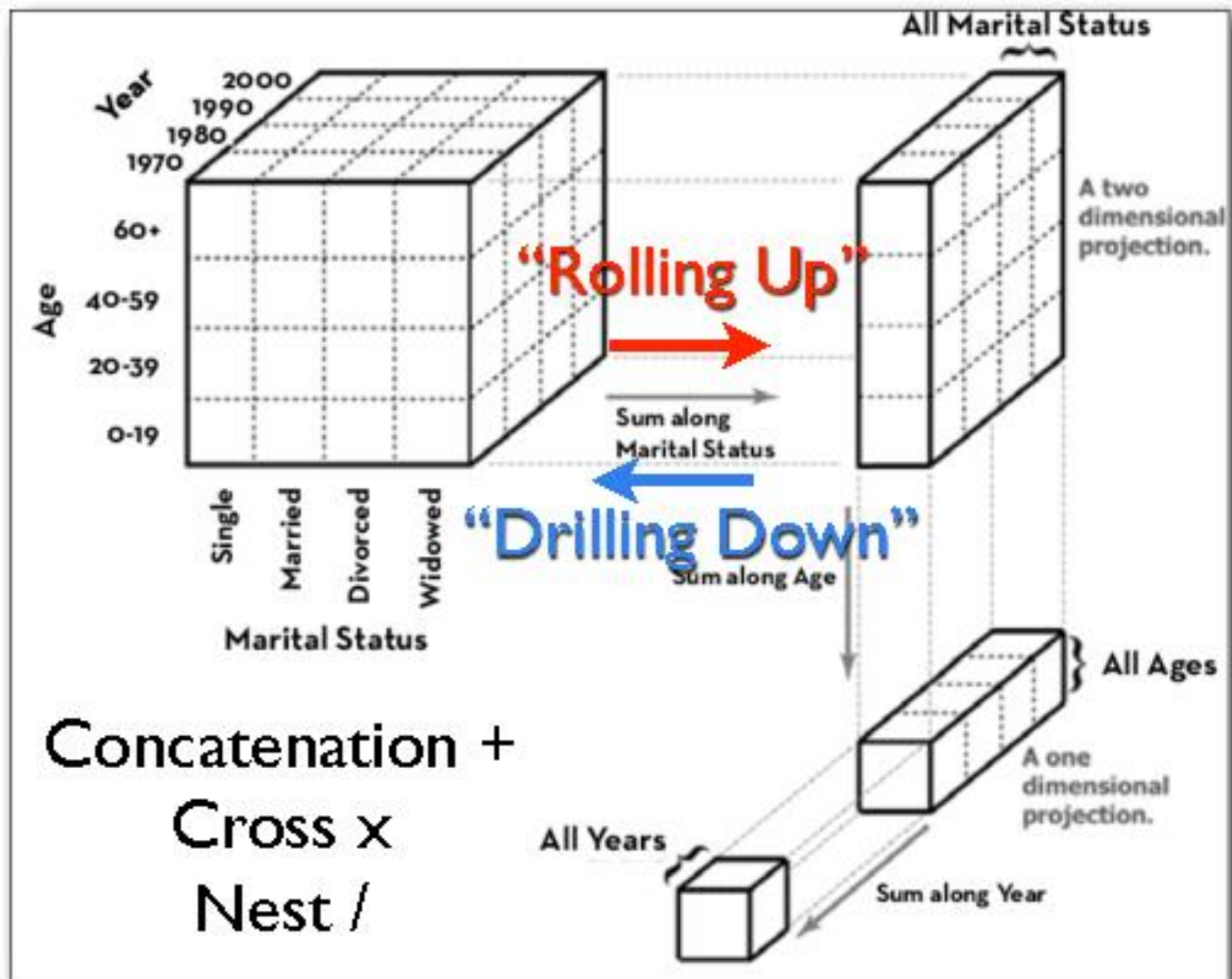
1 = Quantitative

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](http://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 anomalies in local weather patterns
  - classifying showers
  - making toast

# Jacques Bertin

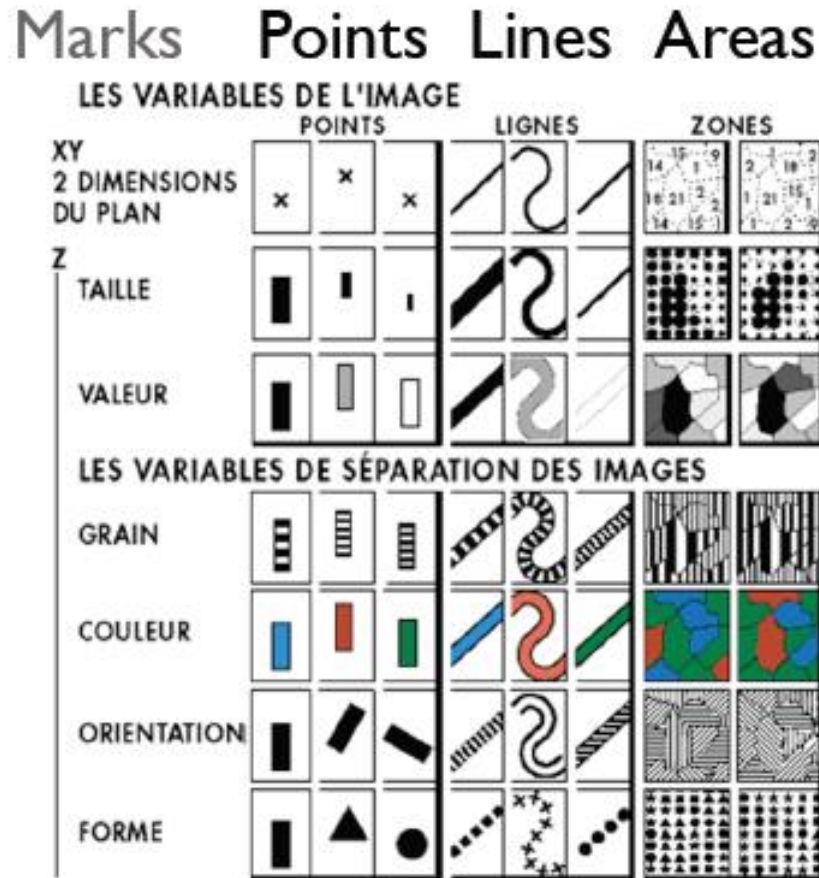


- French cartographer [1918]
- Semiology of Graphics [1967]
- Theoretical principles for visual encodings

# Visual Encodings

Channels

- Position
- Size
- (Grey)Value
- Texture
- Color
- Orientation
- Shape



# Information in Position

Good to encode quantitative variables (Q)

C+

B+

A+

BC twice as long as AB

# Information in Color

(Grey)Value is perceived as ordered (O)



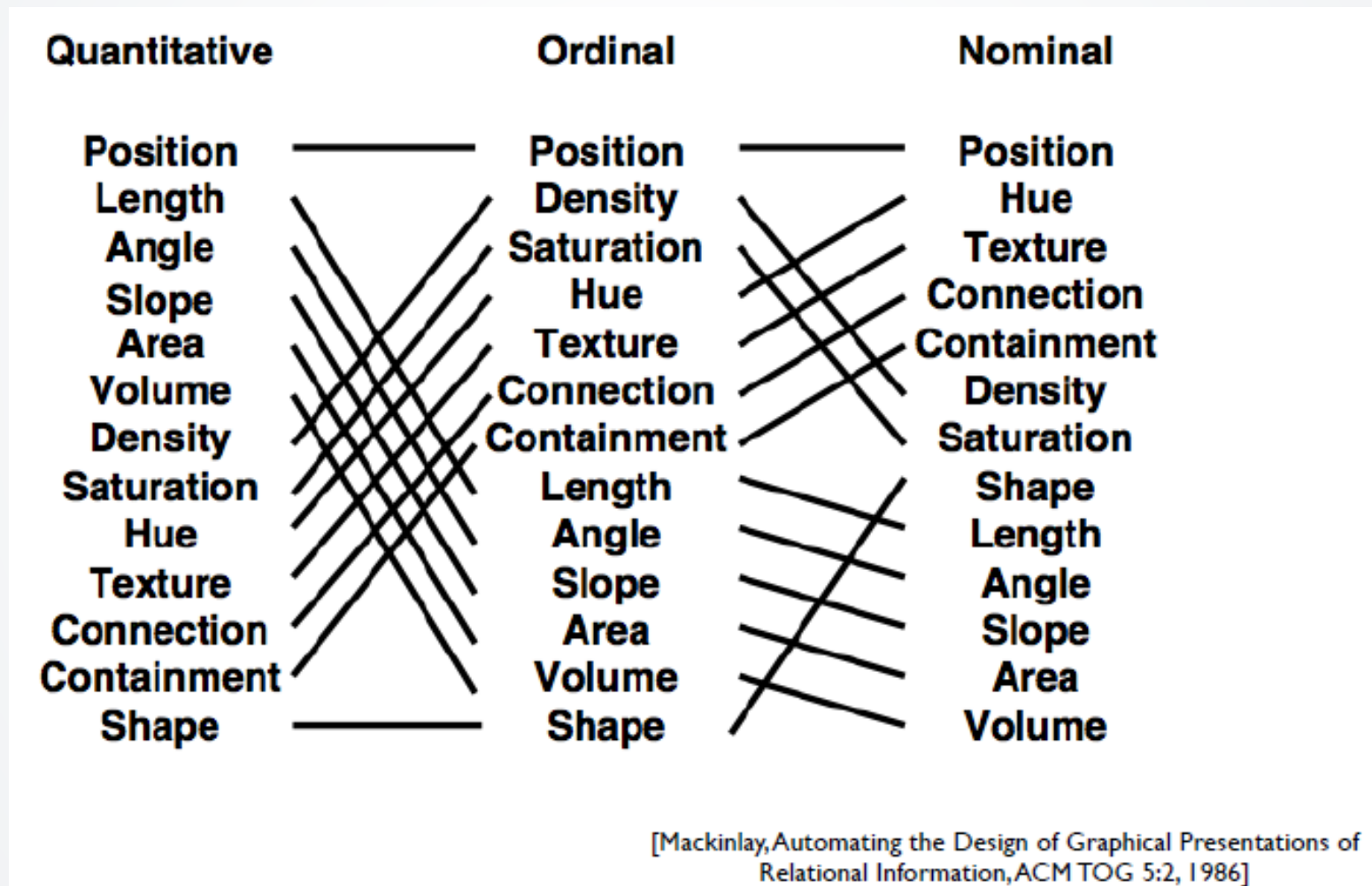
Can encode quantitative values (Q) [not as well]



Hue is normally perceived as unordered (N)





# Mackinlay's Retinal Variables





# 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)
- Tasks
  - overview, zoom, filter, details-on-demand
  - relate, history, extract
  - data alone not enough
  - what do you need to do?
- Mantra:
  - overview first
  - zoom and filter
  - details on demand

# Definitions: Marks and channels

## Marks

- Geometric primitives

### → Points



### → Lines



### → Areas



## Channels

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

### → Position

→ Horizontal



→ Vertical



→ Both



### → Color



### → Shape



### → Tilt

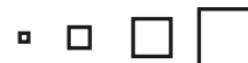


### → Size

→ Length



→ Area



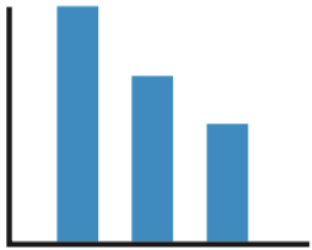
→ Volume



# Visual encoding

## Analyze idiom structure

- As combination of marks and channels



1:  
vertical position

mark: line



2:  
vertical position  
horizontal position

mark: point



3:  
vertical position  
horizontal position  
color hue

mark: point



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

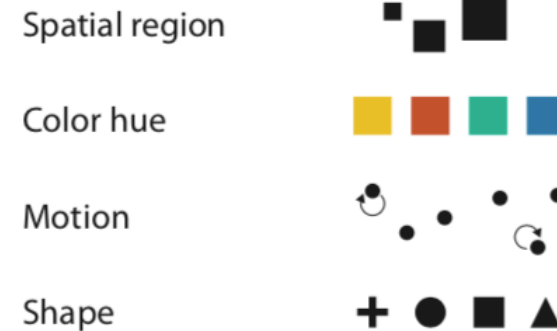
mark: point

# Channels: expressiveness types and effectiveness rankings

## ➔ Magnitude Channels: Ordered Attributes

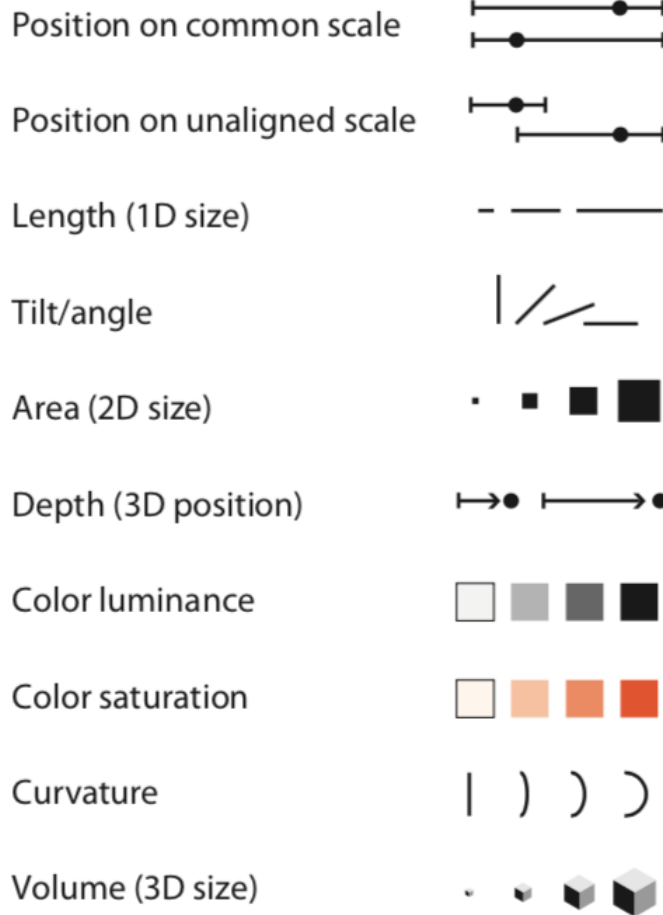


## ➔ Identity Channels: Categorical Attributes

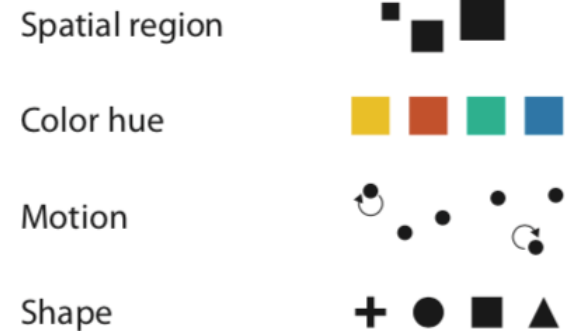


# Channels: Matching Types

## ➔ Magnitude Channels: **Ordered** Attributes



## ➔ Identity Channels: **Categorical** Attributes



## Expressiveness principle

- match channel and data characteristics

# Channels: Rankings

## ➔ Magnitude Channels: Ordered Attributes



## ➔ Identity Channels: Categorical Attributes



## Expressiveness principle

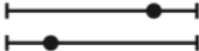
- match channel and data characteristics

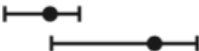
## Effectiveness principle

- encode most important attributes with highest ranked channels

# Channels: Expressiveness types and effectiveness rankings

## ➔ Magnitude Channels: Ordered Attributes

Position on common scale 

Position on unaligned scale 

Length (1D size) 

Tilt/angle 

Area (2D size) 

Depth (3D position) 

Color luminance 

Color saturation 

Curvature 

Volume (3D size) 

Same

Same

Most

Effectiveness

Least

## ➔ Identity Channels: Categorical Attributes

Spatial region 

Color hue 

Motion 

Shape 

## Expressiveness principle

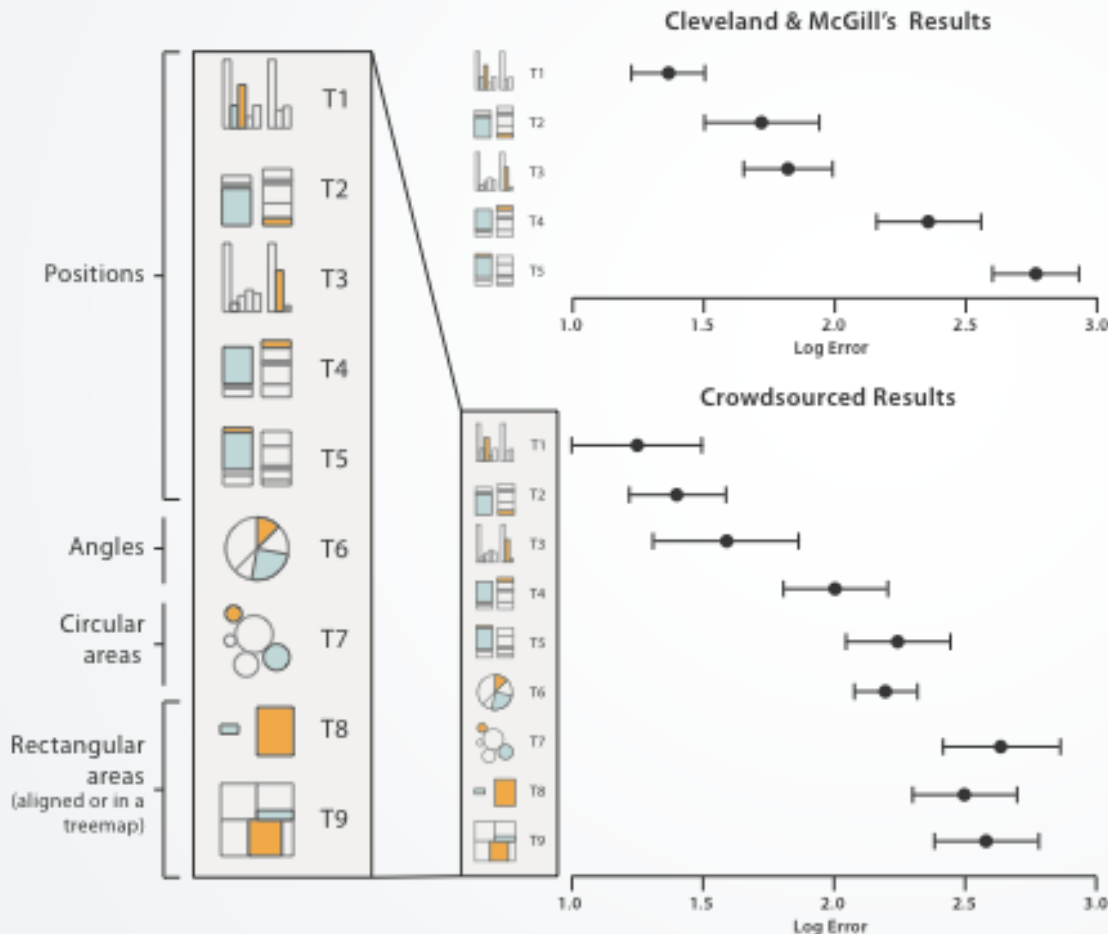
- match channel and data characteristics

## Effectiveness principle

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



# Accuracy: Vis experiments

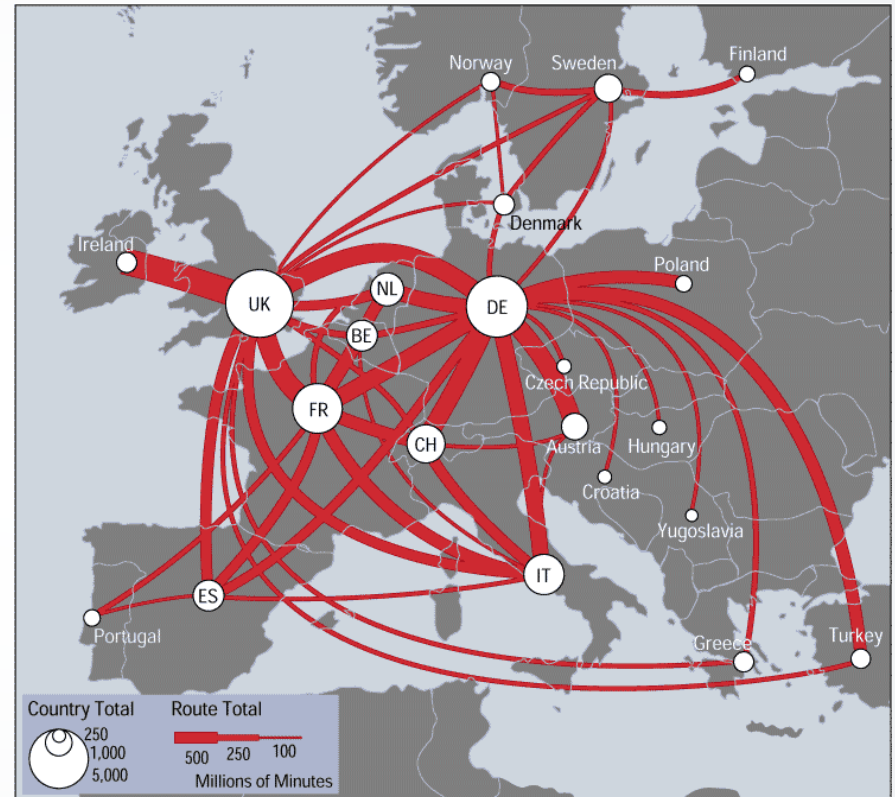


[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010, p. 203–212.]

# Discriminability: How many usable steps?

Must be sufficient for the number of attribute levels to show

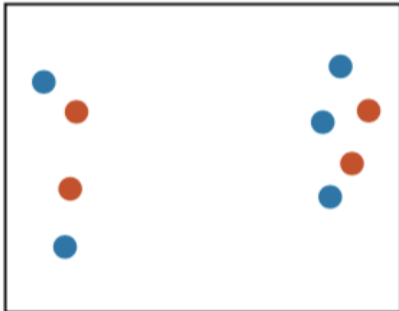
- **Linewidth**: a few bins



Linewidth has a limited number of discriminable bins

# Separability vs. Integrality

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

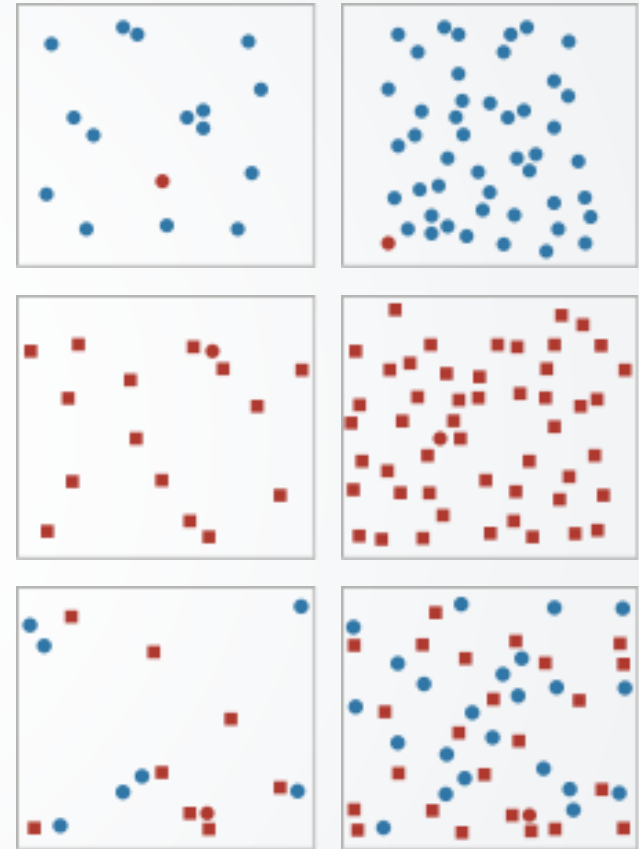


Major interference

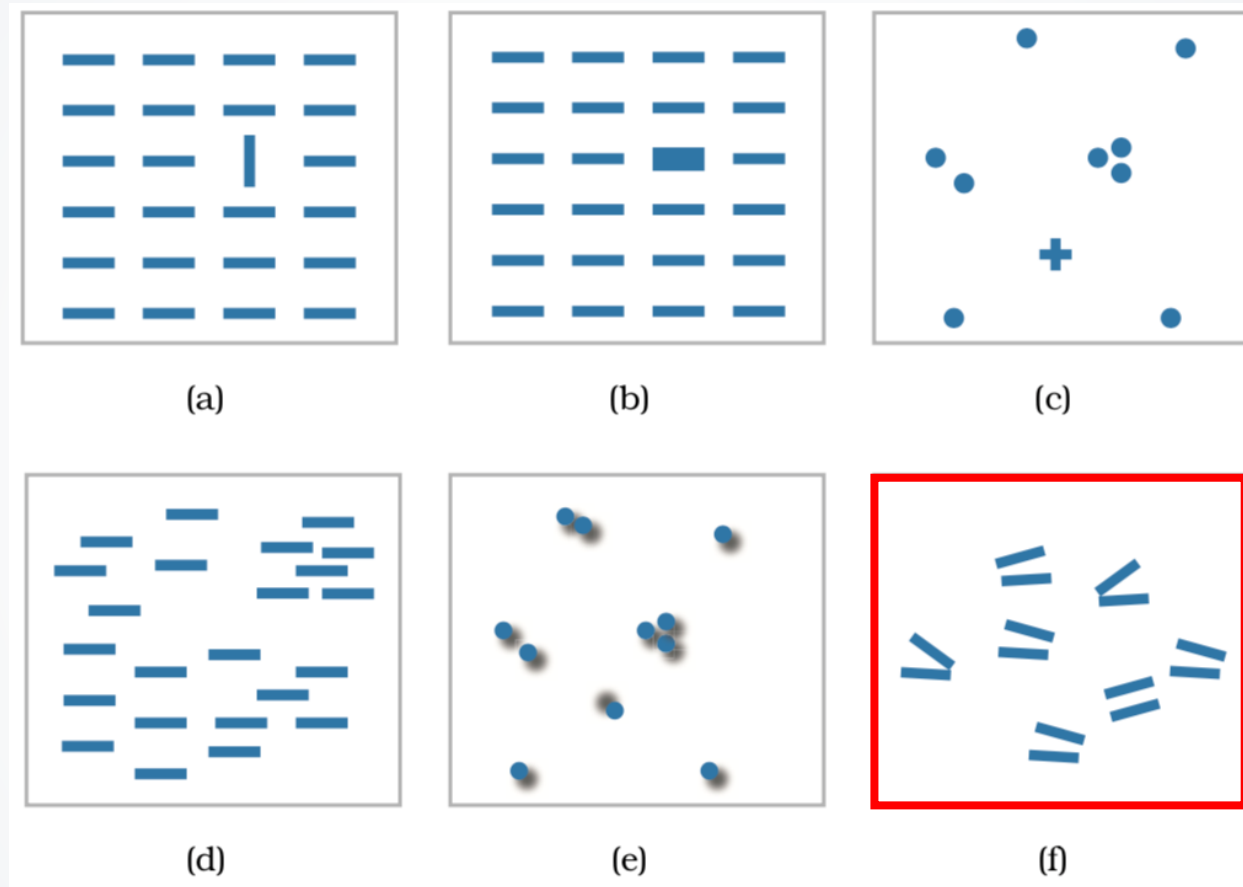
4 groups total:  
integral hue

# 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: (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

## Marks as Links

### ➔ Containment



### ➔ Connection



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

## ➔ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion

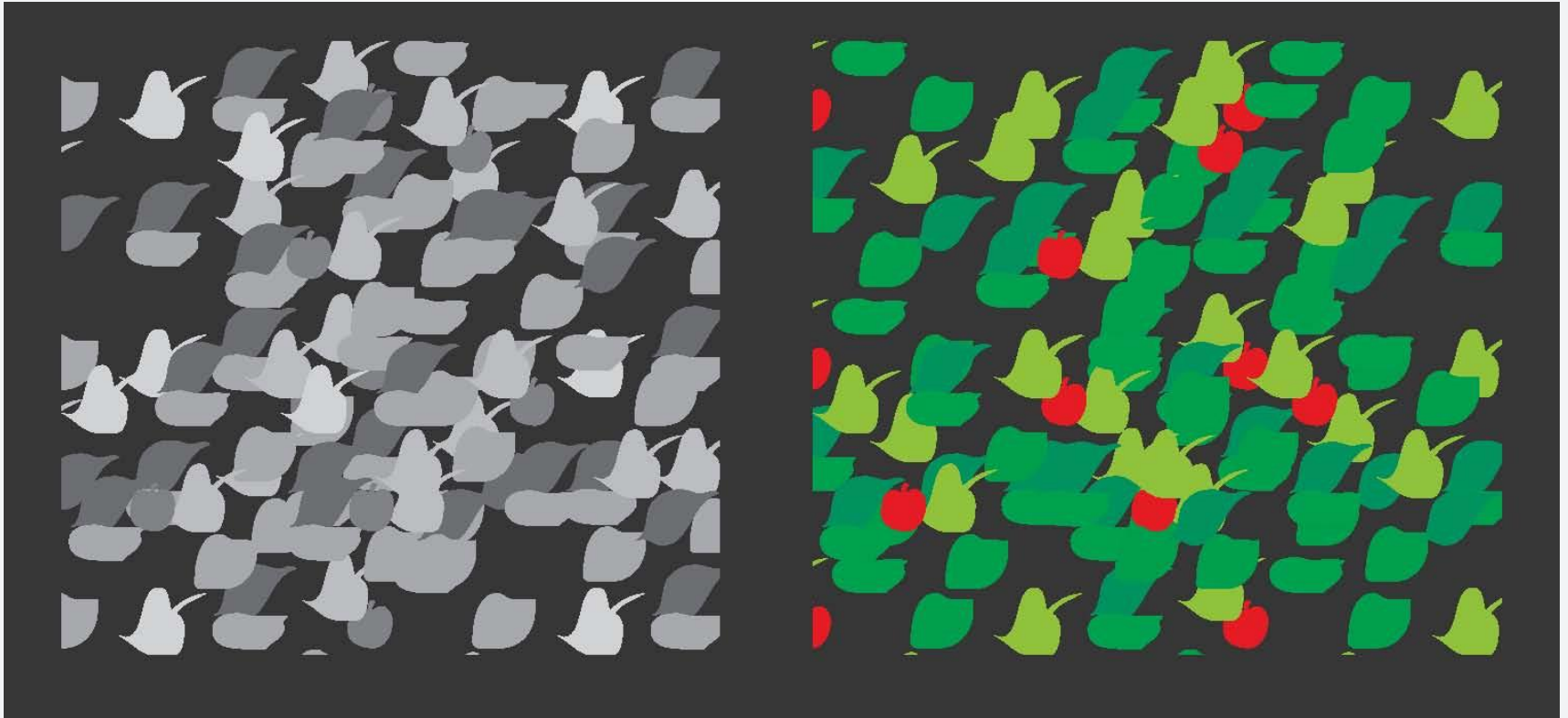


Shape



Color: small areas

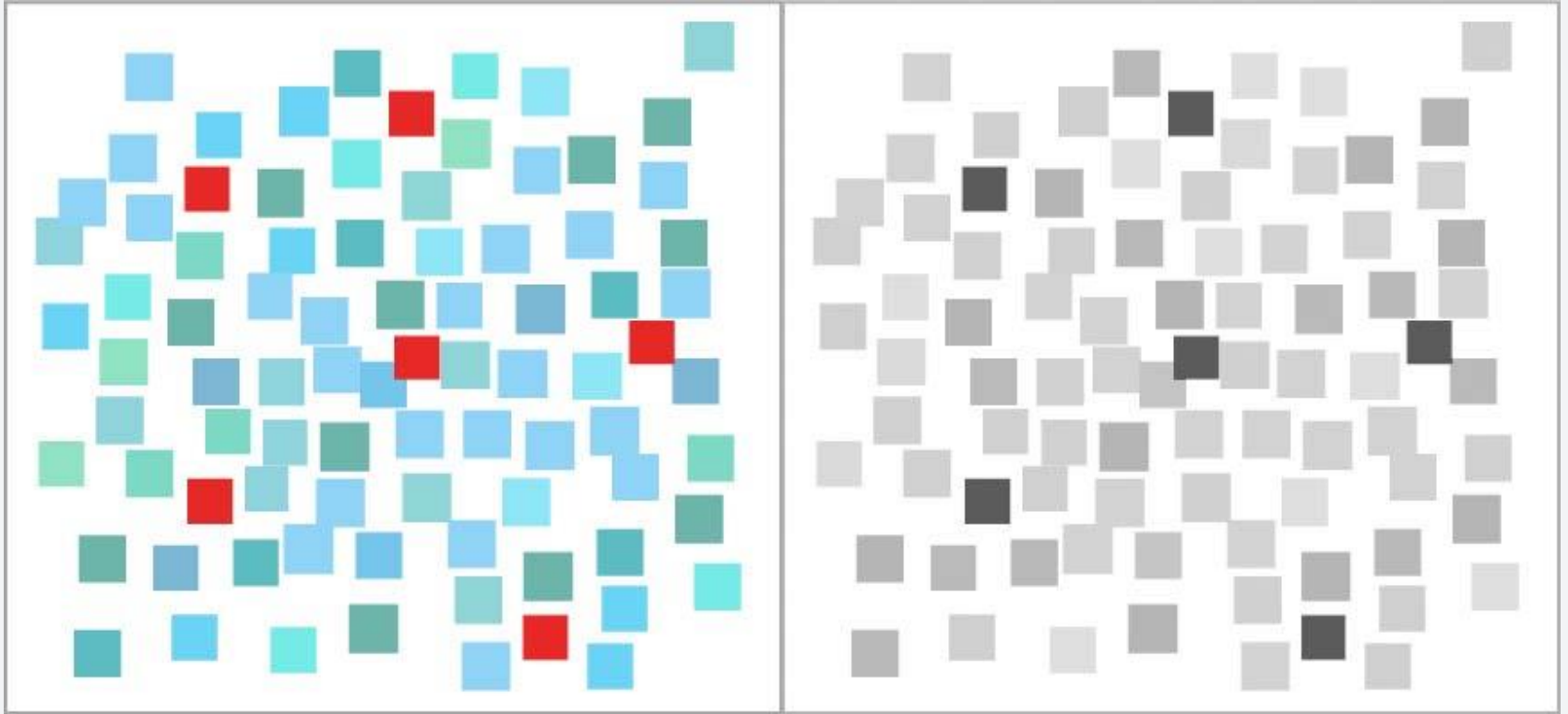
# Small Areas



Ware, “Information Visualization”



# Pop-Out



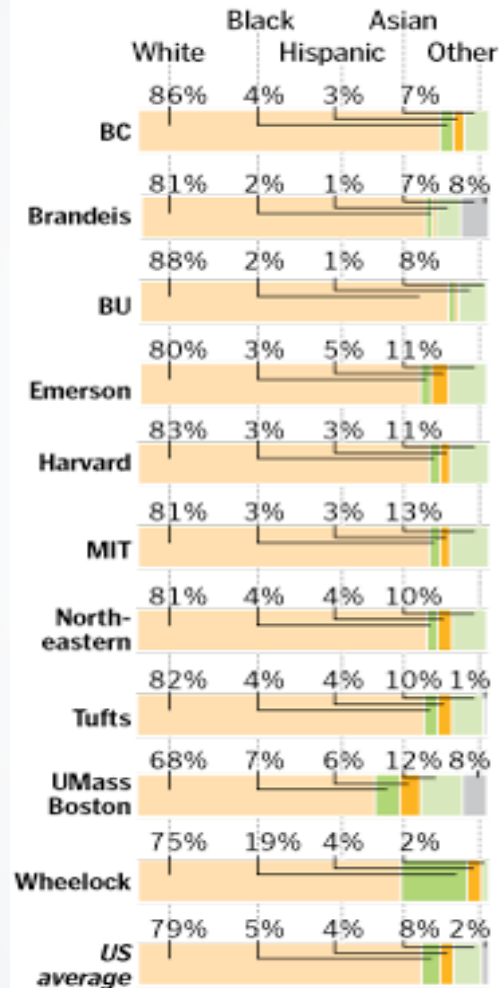
Hue and lightness

Lightness only

Based on slide from Stone

## FACULTY DIVERSITY AT BOSTON-AREA COLLEGES

Percent of tenured and tenure-track professors who belong to various racial and ethnic groups



SOURCE: American Council on Education, individual colleges and universities

GLOBE STAFF

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.














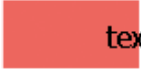









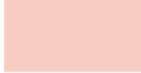







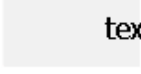



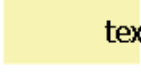



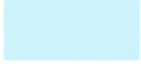
# Colors for Categories



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



# Tableau Colors

	Regular	Medium	Light	Ultra-light
Blue				
Orange				
Green				
Red				
Purple				
Brown				
Pink				
Gray				
Gold				
Teal				

# Facts

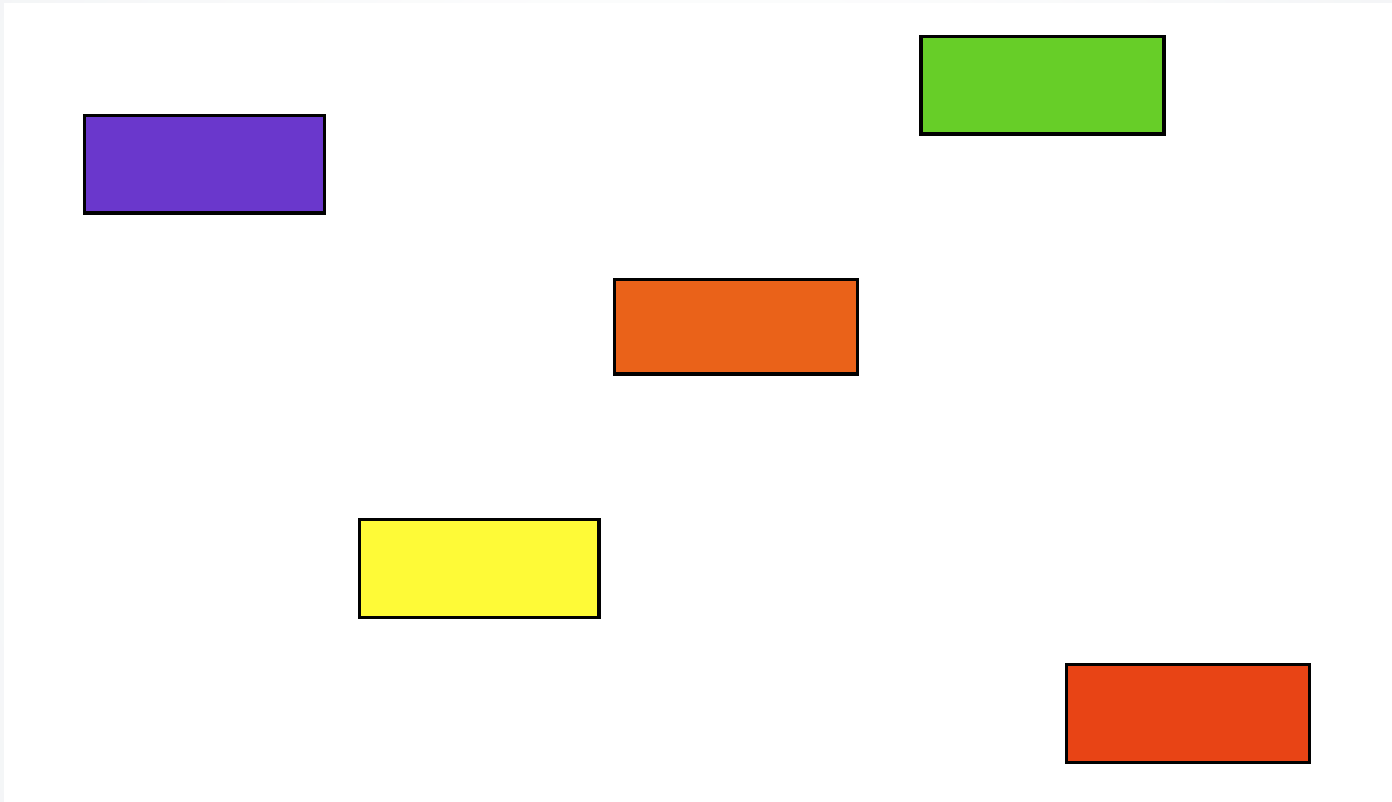
Only a small number of colors can be used effectively as nominal labels.

Keep the number of colors for nominal data to less than eight.

Use quiet medium grey backgrounds.

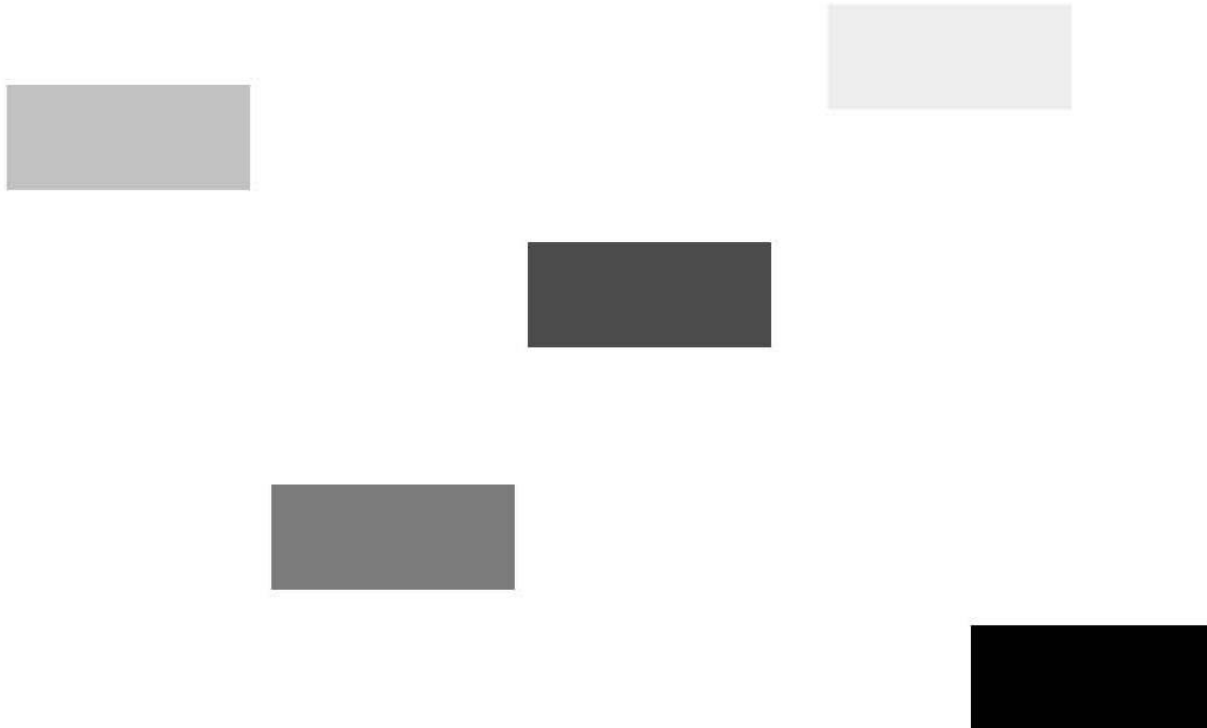
Color: Ordinal

# Order These Colors



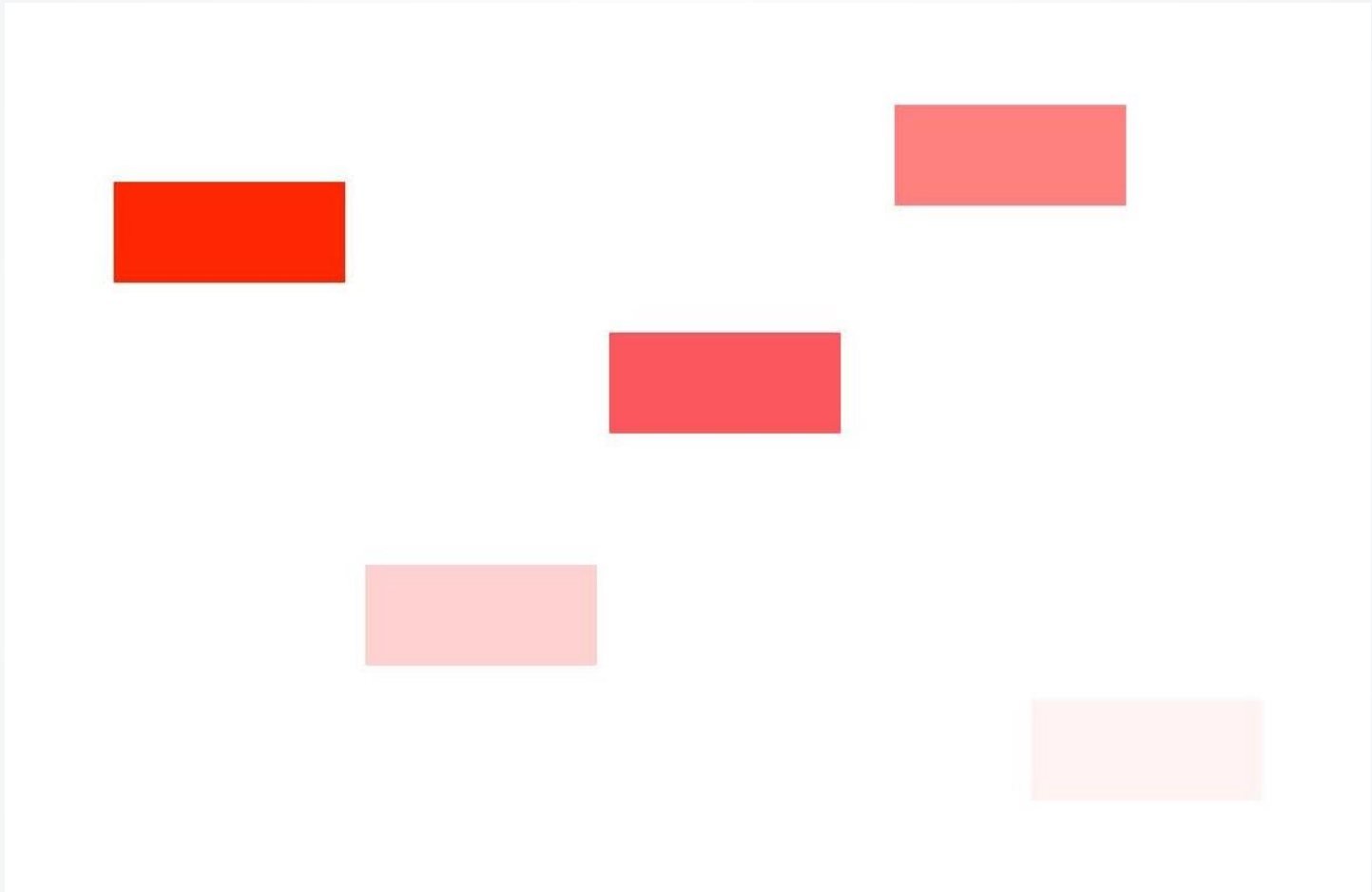
Based on slide from Stasko

# Order These Colors



Based on slide from Stasko

# Order These Colors

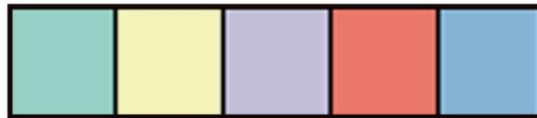


Based on slide from Stasko

# Brewer Scales

Nominal

Qualitative Scale



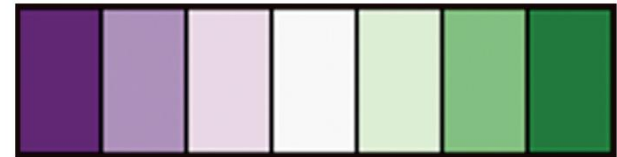
Ordinal

Sequential Scale



0 → Max


Diverging Scale



Max ← 0 → Max

Cynthia Brewer, Color Use Guidelines for Data Representation

**COLORBREWER 2.0**  
color advice for cartography

**sequential**  [learn more >](#)

single hue

☒ colorblind safe    ☐ print friendly  
☐ photocopy-able    [learn more >](#)

☒ RGB ☐ CMYK ☐ HEX

adjust map context

☐ roads


cities

☒ borders

select a background

☒ solid color

☐ terrain

 color transparency

**EXPORT YOUR COLORS >>**

## SCORE CARD

© Cynthia Brewer, Mark Harrower and The Pennsylvania State University  
[Support](#)  
[Back to ColorBrewer 1.0](#)





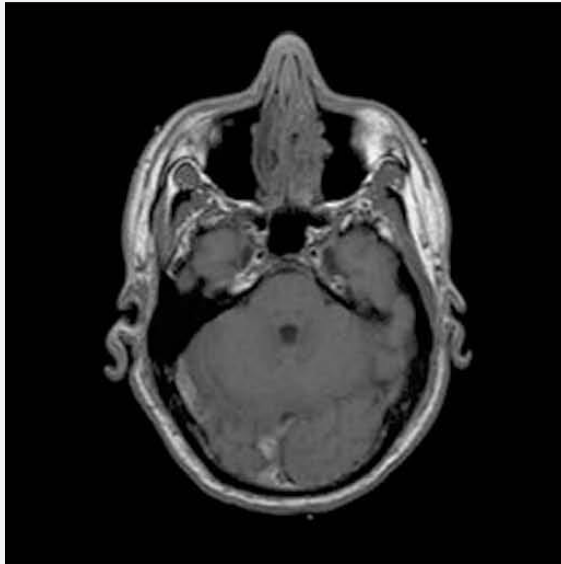
# Facts

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

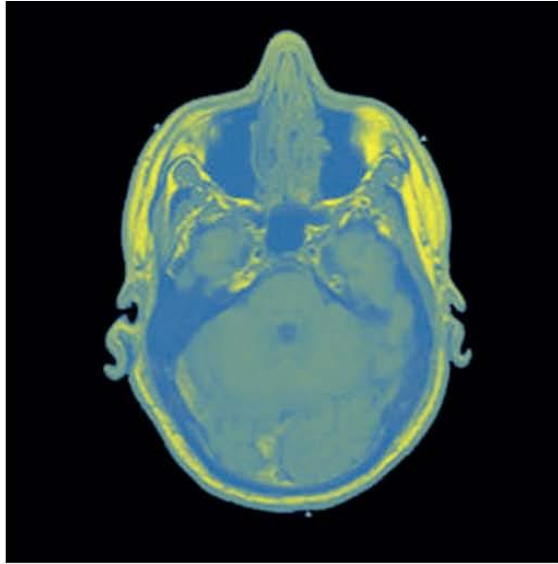
Show ordinal data with a discrete set of color values that change in lightness or saturation

Color: Quantitative

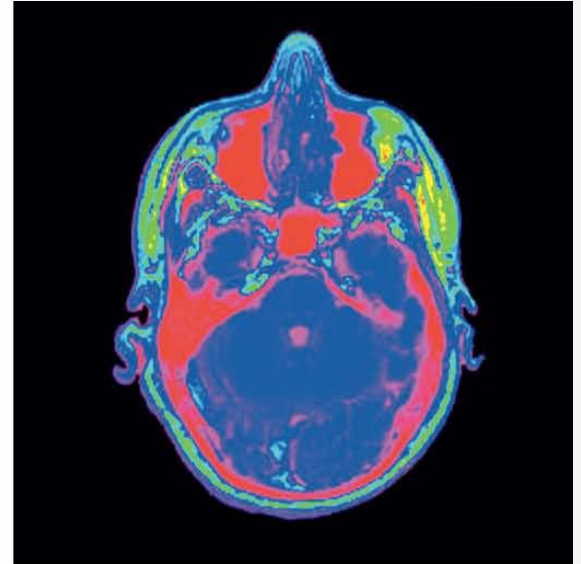
# Colormaps



Lightness scale



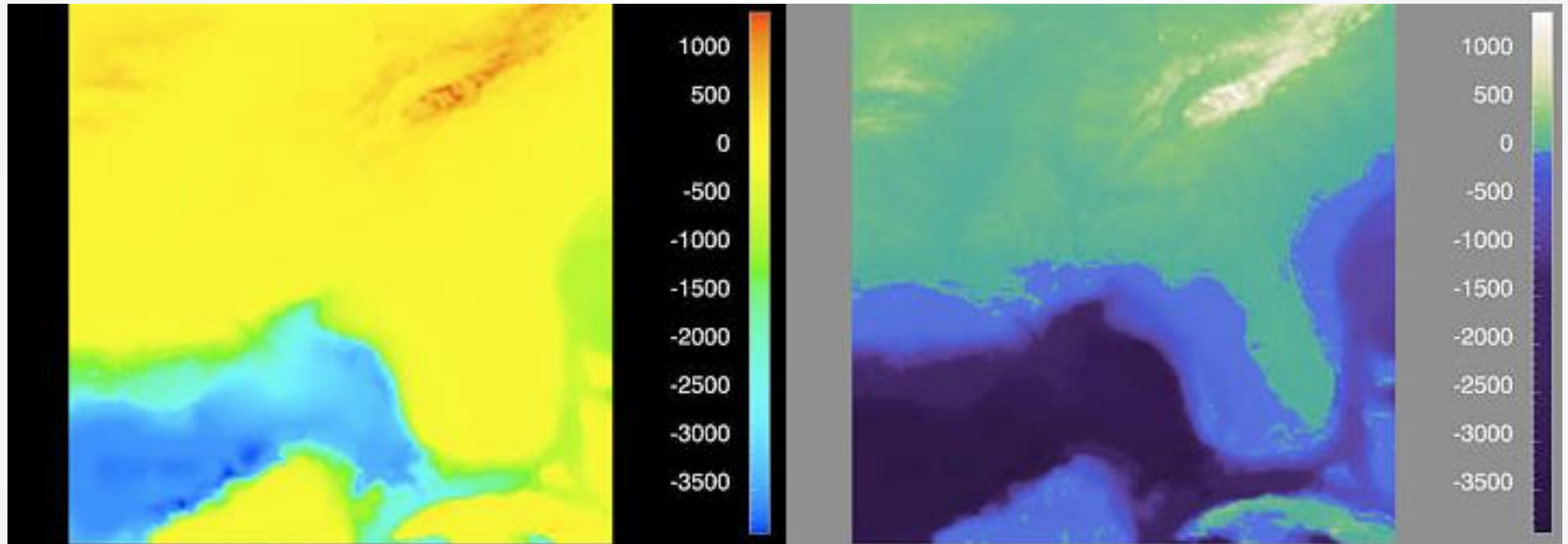
Lightness scale  
with hue and  
chroma variation



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

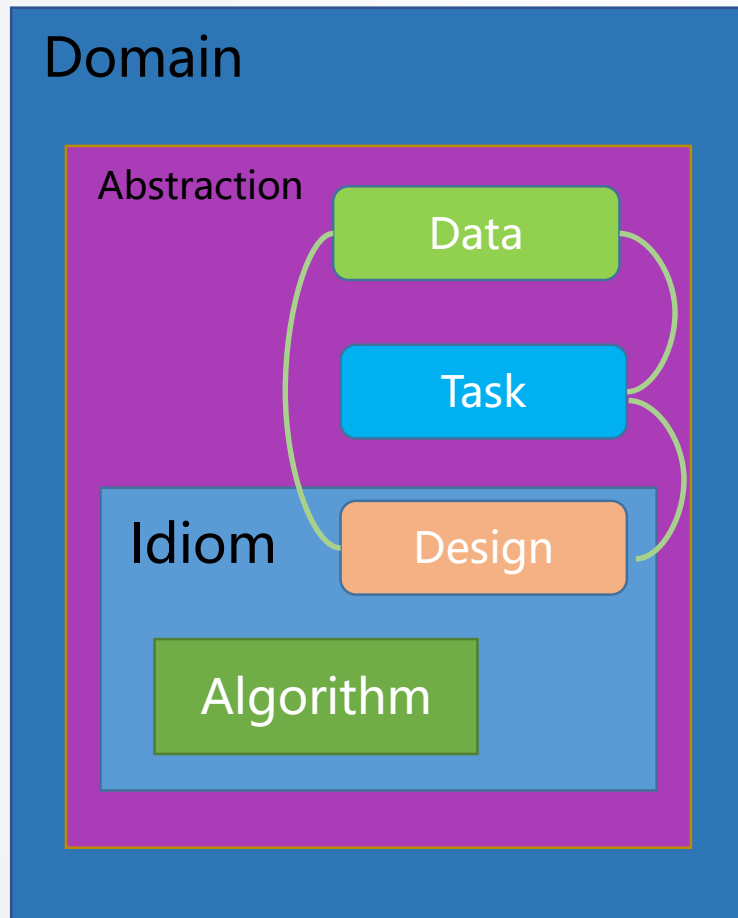
Use colormaps with a limited hue palette and redundantly vary lightness and saturation.

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

threat: wrong problem

validate: observe and interview target users

threat: bad data/operation abstraction

threat: ineffective encoding/interaction technique

validate: justify encoding/interaction design

threat: slow algorithm

validate: analyze computational complexity

implement system

validate: measure system time/memory

validate: qualitative/quantitative result image analysis

[test on any users, informal usability study]

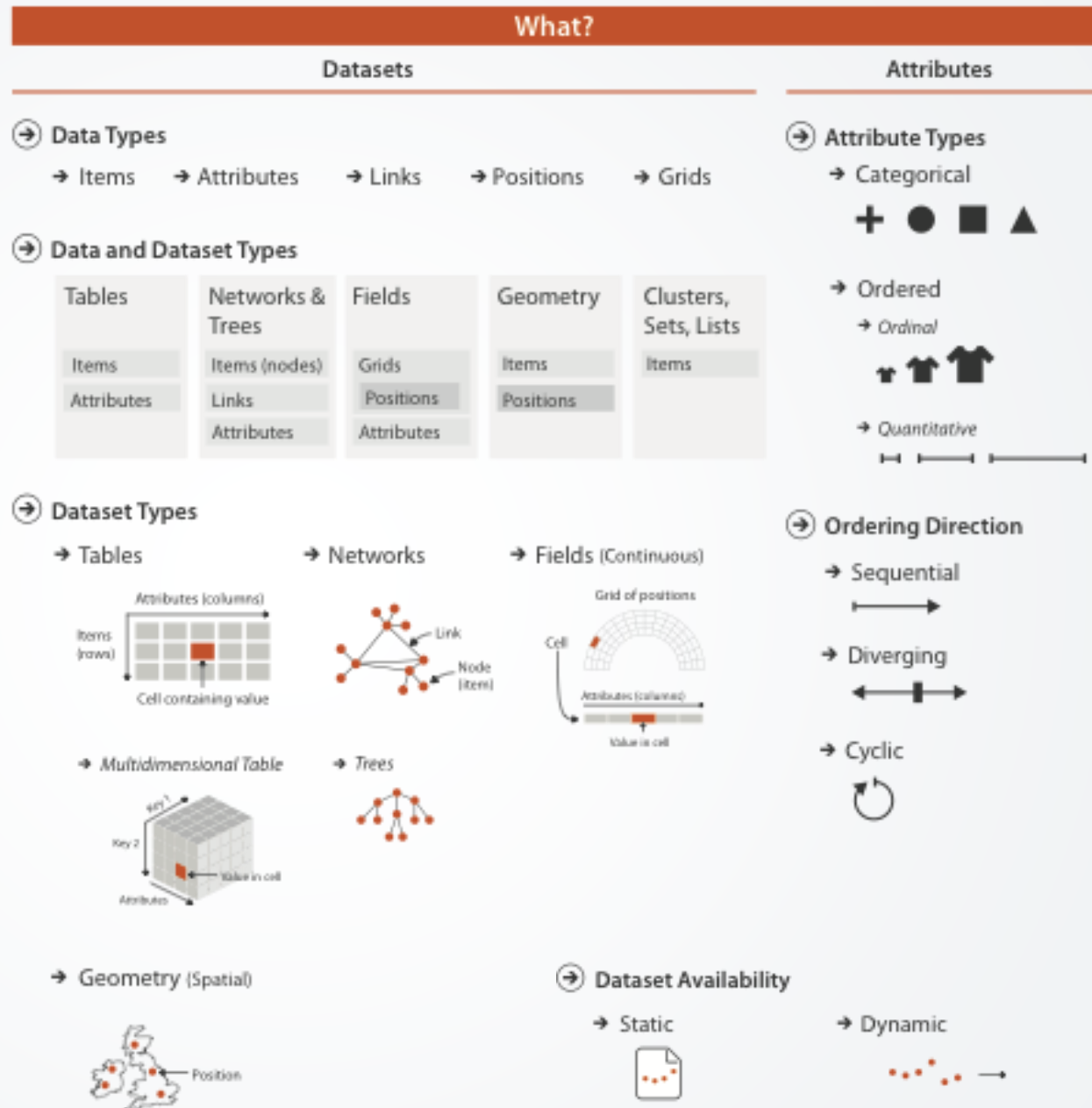
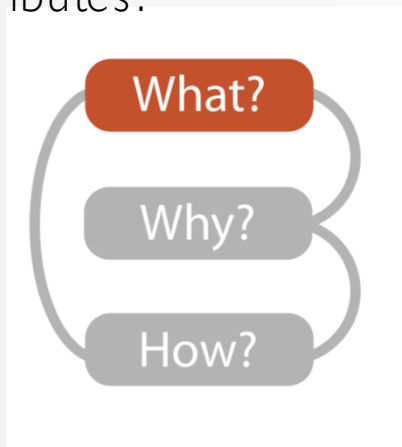
validate: lab study, measure human time/errors for operation

validate: test on target users, collect anecdotal evidence of utility

validate: field study, document human usage of deployed system

validate: observe adoption rates

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



# Data types

## → Data Types

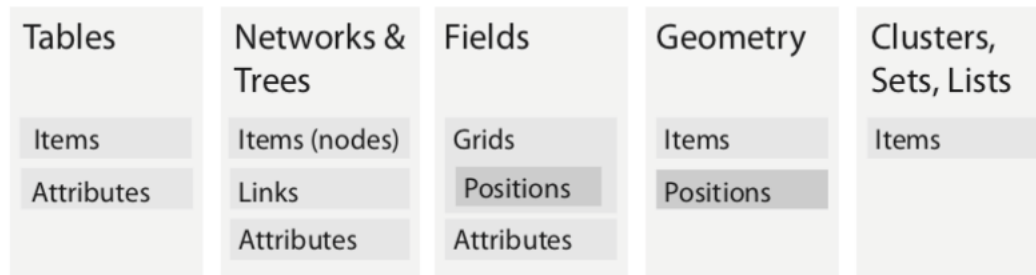
→ Items    → Attributes    → Links    → Positions    → Grids

# Data and Dataset Types

## → Data Types

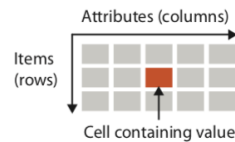
→ Items → Attributes → Links → Positions → Grids

## → Data and Dataset Types

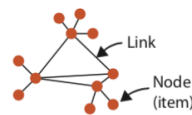


## → Dataset Types

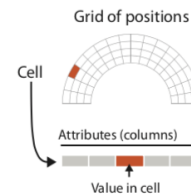
→ Tables



→ Networks



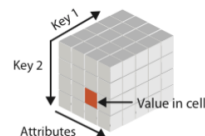
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



→ Trees



## → Dataset Availability

→ Static



→ Dynamic



# Attribute types

## Attributes

---

### ➔ Attribute Types

➔ Categorical

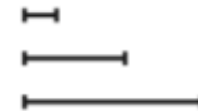


➔ Ordered

➔ Ordinal



➔ Quantitative



### ➔ Ordering Direction

➔ Sequential



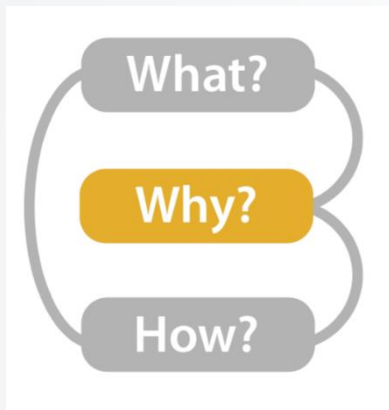
➔ Diverging



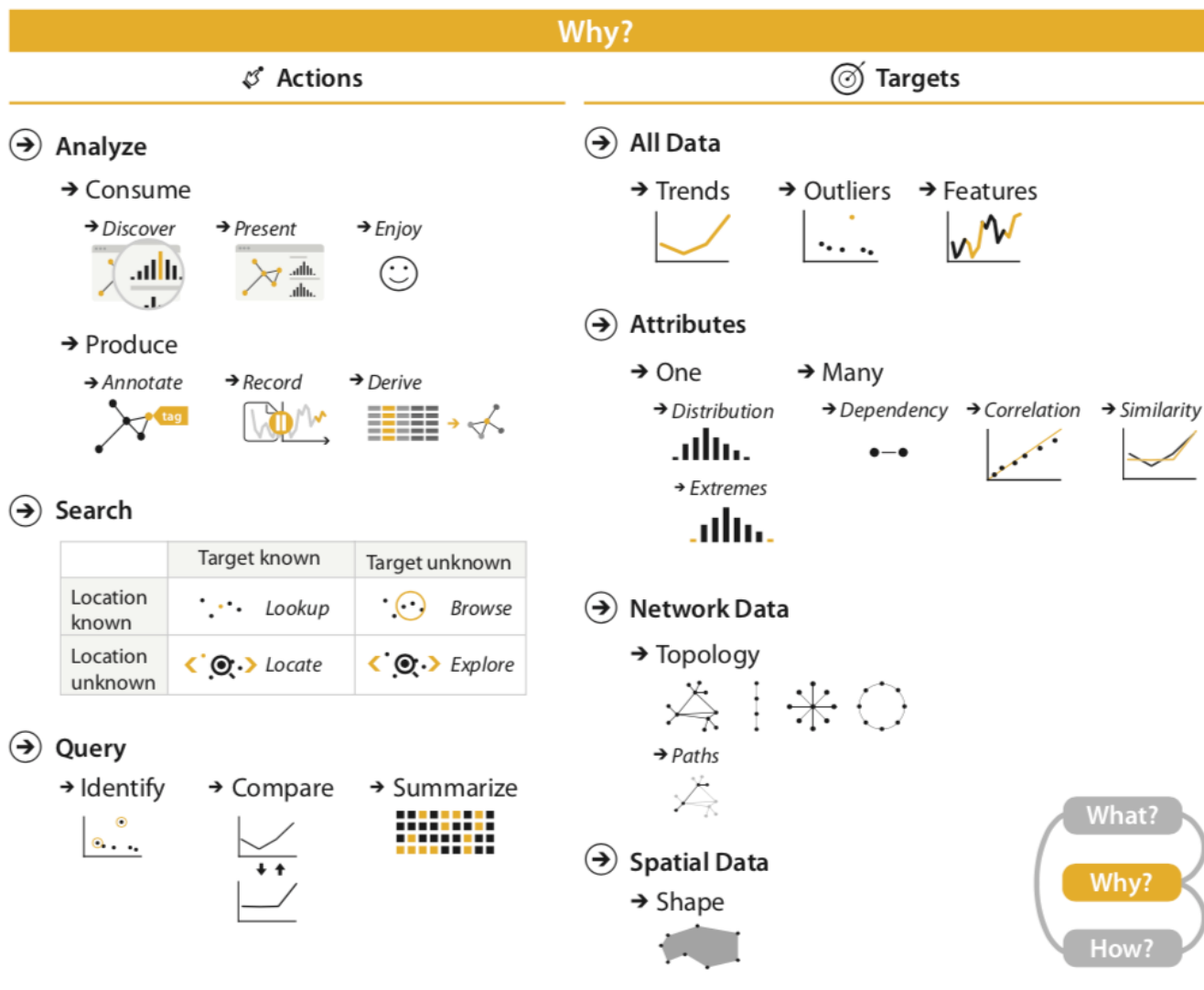
➔ Cyclic



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



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



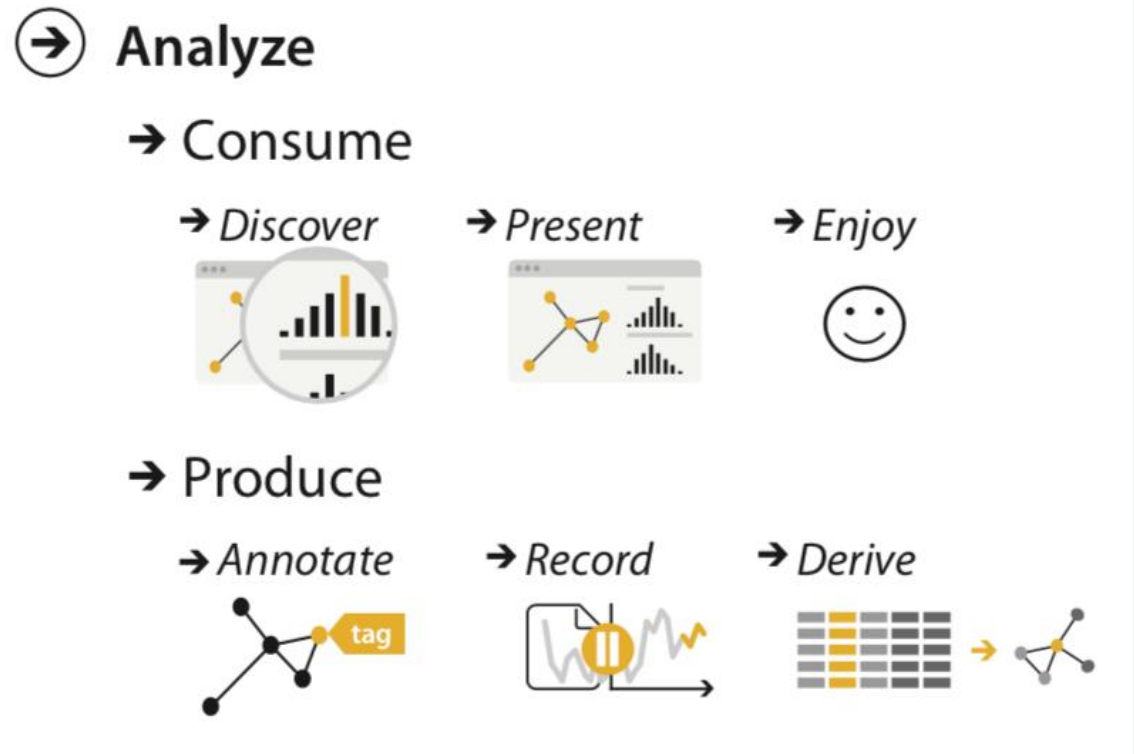
# Actions: Analyze

- Consume

- Discover vs present
  - classic split
  - aka explore vs explain
- Enjoy
  - newcomer
  - aka casual, social

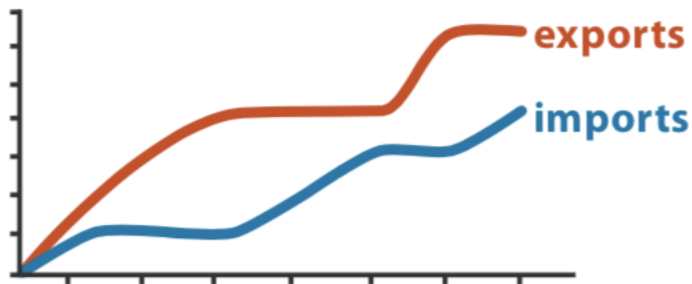
- Produce

- Annotate, record
- Derive
  - crucial design choice

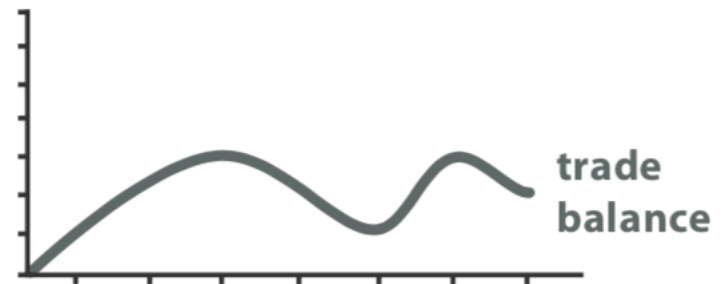


# Derive

- 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



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

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

## ➔ Query

### ➔ Identify



### ➔ Compare



### ➔ Summarize



# Why: Targets

## → All Data

→ Trends



→ Outliers



→ Features



## → Attributes

→ One

→ Distribution



→ Extremes

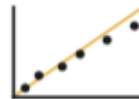


→ Many

→ Dependency



→ Correlation

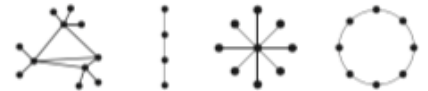


→ Similarity



## → Network Data

→ Topology



→ Paths



## → Spatial Data

→ Shape



# How?

## Encode

### ➔ Arrange

➔ Express



➔ Order



➔ Use



➔ Separate



➔ Align



### ➔ Map

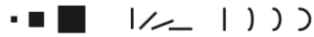
from **categorical** and **ordered** attributes

➔ Color

➔ Hue      ➔ Saturation      ➔ Luminance



➔ Size, Angle, Curvature, ...



➔ Shape



➔ Motion

Direction, Rate, Frequency, ...



## Manipulate

### ➔ Change



### ➔ Select

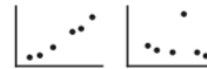


### ➔ Navigate



## Facet

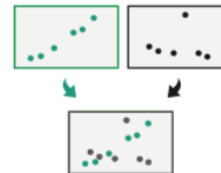
### ➔ Juxtapose



### ➔ Partition



### ➔ Superimpose



## Reduce

### ➔ Filter



### ➔ Aggregate



### ➔ Embed



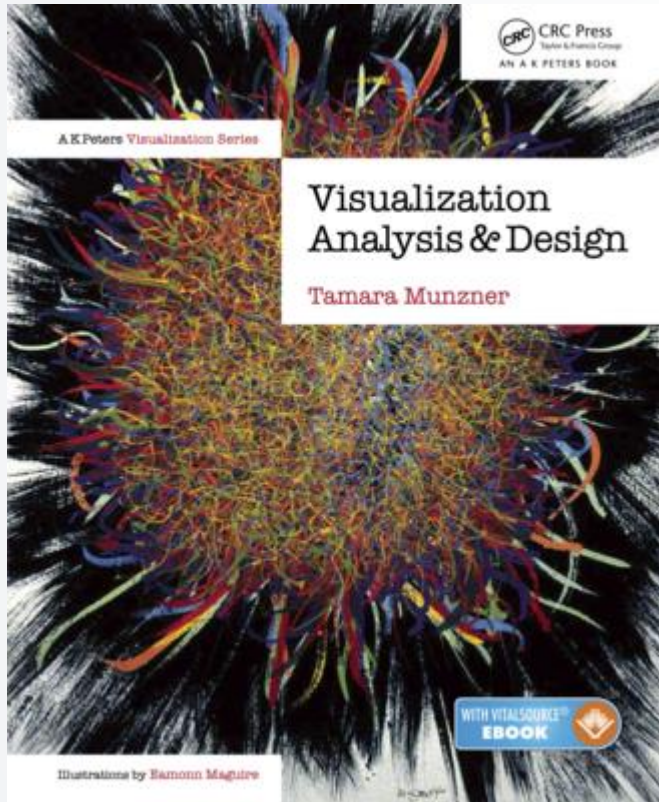
What?

Why?

How?

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

# Suggest Readings



## Wilkinson's GoG

