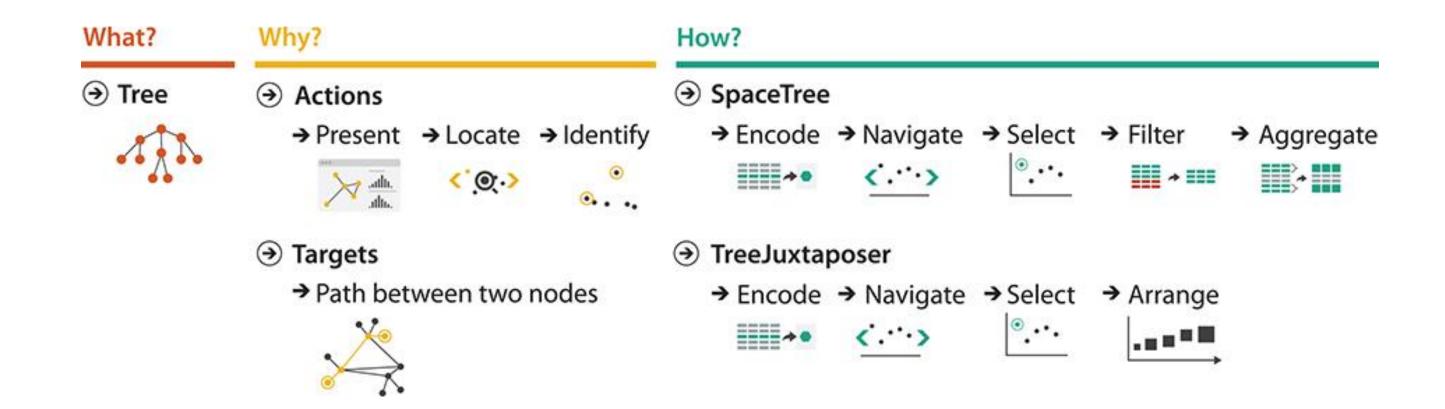
Analysis Framework

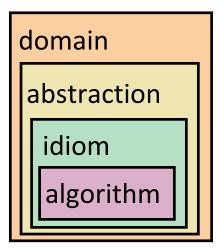
Yu-Shuen Wang, CS, NCTU

- What is shown?
 - Data abstraction
- Why is the user looking at it?
 - Task abstraction
- How is it shown?
 - Idiom: visual encoding and interaction



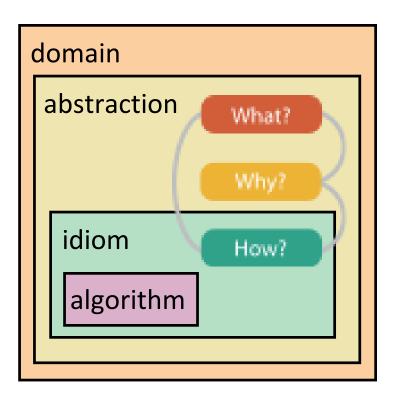
Analysis framework: Four levels, three questions

- domain situation
 - who are the target users?
- abstraction What do the target users need?
 - translate from specifics of domain to vocabulary of vis
- what is shown? data abstraction
 - often don't just draw what you're given: transform to new form
- why is the user looking at it? task abstraction
- idiom
- how is it shown?
 - visual encoding idiom: how to draw
 - interaction idiom: how to manipulate
- algorithm
 - efficient computation



[A Nested Model of Visualization Design and Validation.

Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

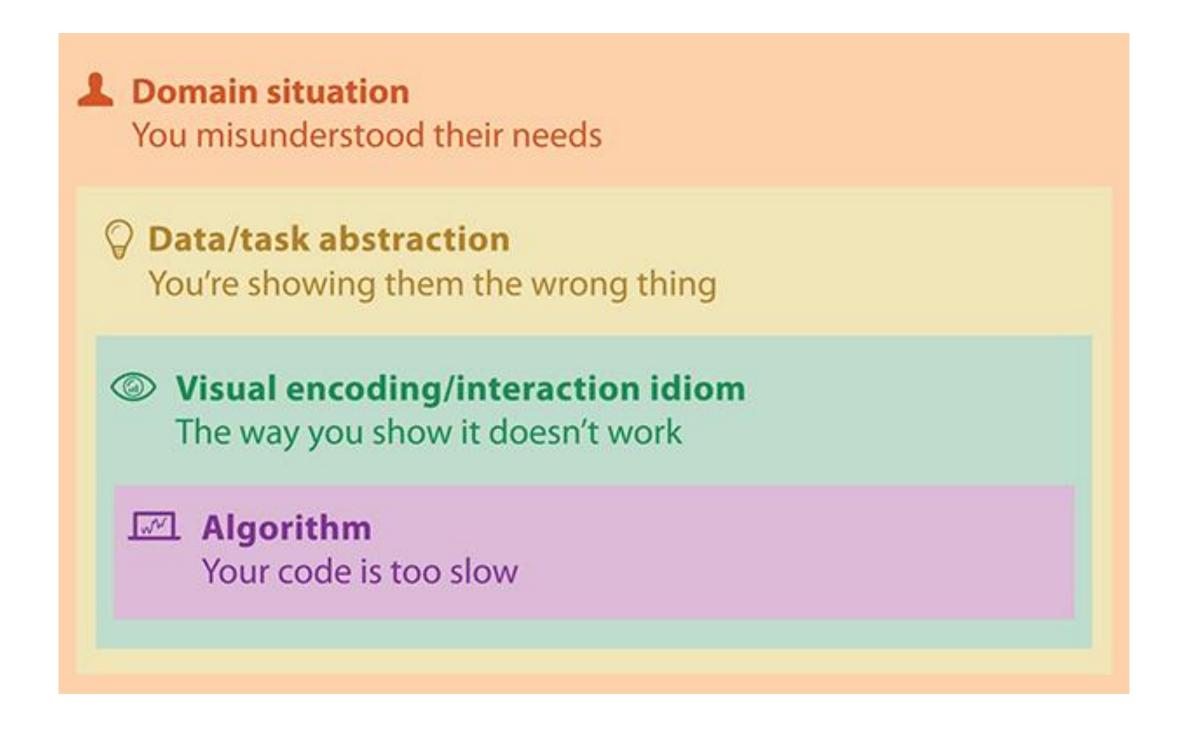


[A Multi-Level Typology of Abstract Visualization Tasks

Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

Why is validation difficult?

different ways to get it wrong at each level



Why is validation difficult?

solution: use methods from different fields at each level

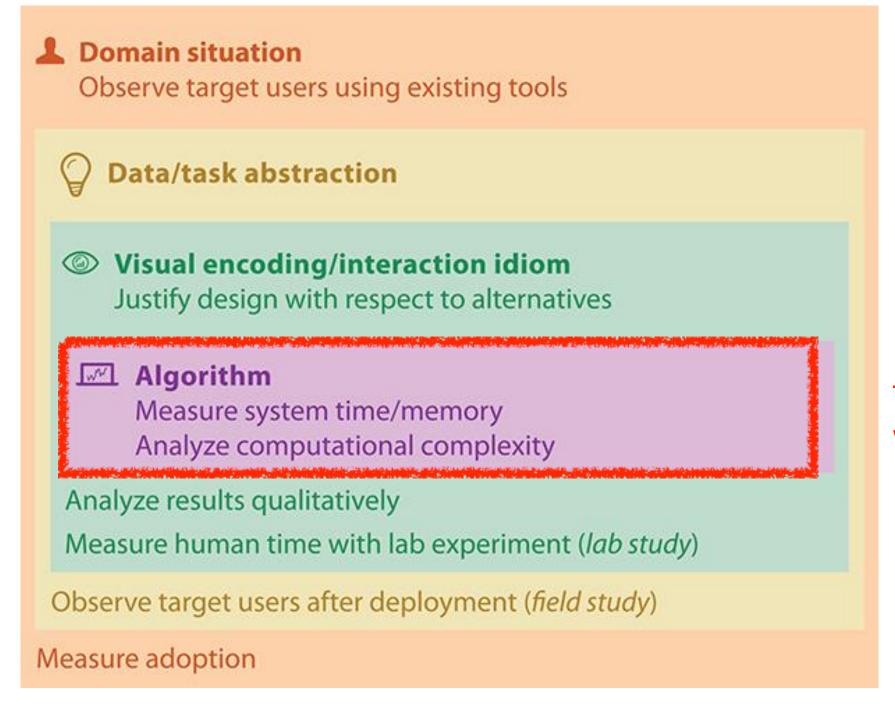
anthropology/ ethnography

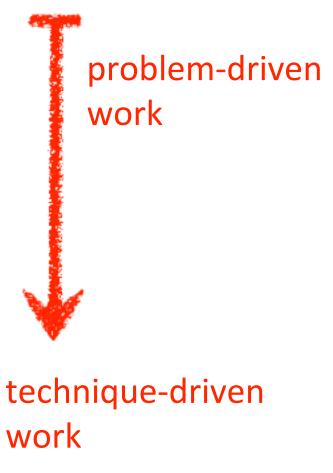
design

computer science

cognitive psychology

anthropology/ ethnography





What?

Dataset

Attribute

What?

Datasets

Attributes

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

→ Data and Dataset Types



- → Attribute Types
 - → Categorical



- → Ordered
 - → Ordinal

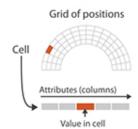


→ Quantitative

- → Dataset Types
 - → Tables

(rows)

- → Networks
- → Fields (Continuous)



→ Multidimensional Table

Attributes (columns)

Cell containing value

- → Trees



- Ordering Direction
 - → Sequential
 - → Diverging
 - → Cyclic



→ Geometry (Spatial)



- Dataset Availability
 - → Static

→ Dynamic

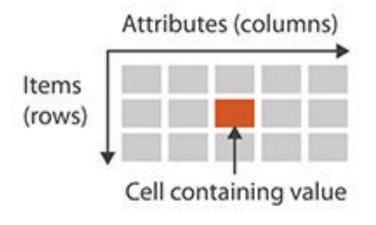




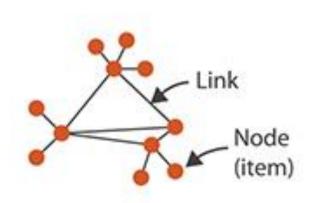


Dataset Types

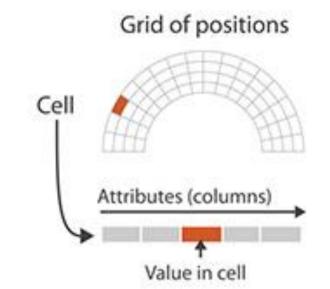
- → Tables
 - ables



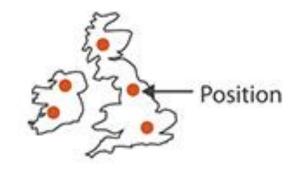
→ Networks



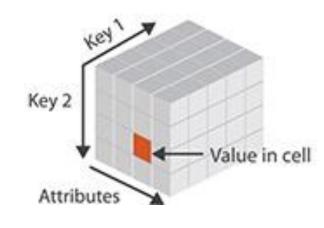
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table

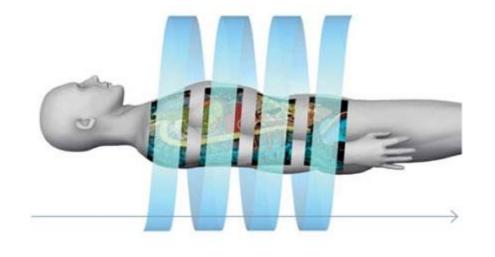


→ Trees

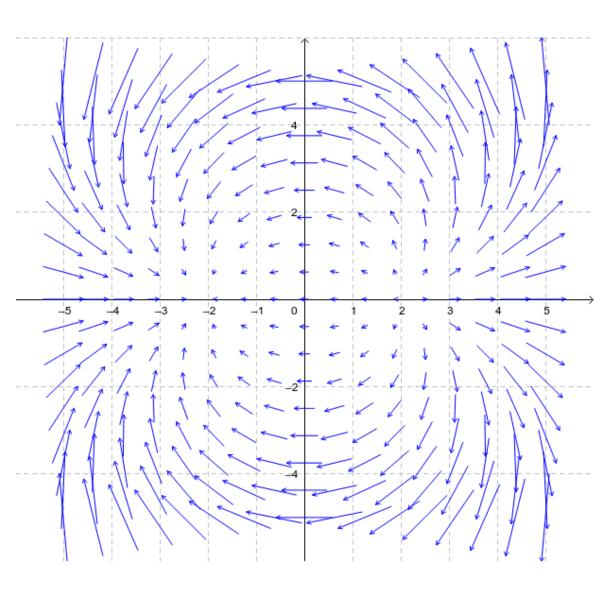


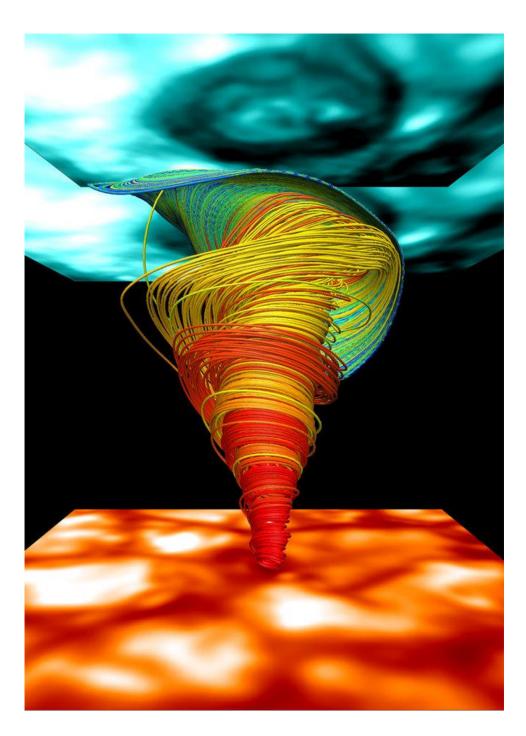
- <u>visualization</u> vs <u>computer graphics</u>
 - -geometry is design decision

Field data









Data and Dataset Types

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

Data Types

- → Items → Attributes → Links → Positions → Grids
- Dataset Availability
 - → Static → Dynamic

 ...

Attributes

- Attribute Types
 - → Categorical









- → Ordered
 - → Ordinal

→ Quantitative



- Ordering Direction
 - Sequential



→ Diverging



→ Cyclic



Key versus Value Semantics

Key as an index to look up value attributes

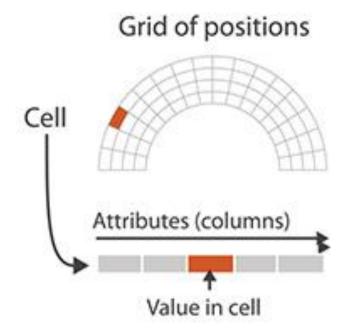
→ Tables

Attributes (columns)

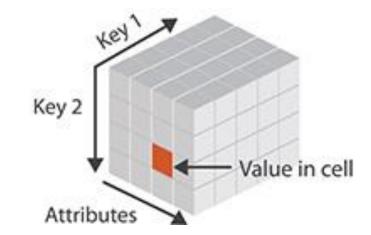
Items
(rows)

Cell containing value

→ Fields (Continuous)



→ Multidimensional Table



Temporal semantics

Time-varying

- Time is one of the key attributes (e.g., the location of each animal every second)
- Time-series
 - An ordered sequence of time-value pairs

Dynamic

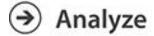
- Time varying
- Stream: change in the running session

Why?

Why?

S Actions





→ Consume



- → Produce
- → Derive → Record → Annotate
- Search

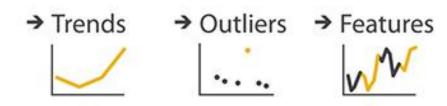
	Target known	Target unknown	
Location known	·.·· Lookup	•. Browse	
Location unknown	C O . ➤ Locate	€ ! Explore	

• {action, target} pairs

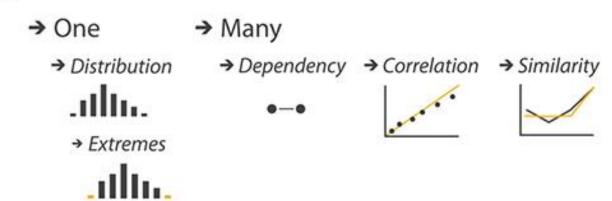
- -discover distribution
- -compare trends
- -locate outliers
- -browse topology



All Data

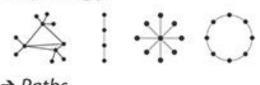


Attributes



Network Data





→ Paths



Spatial Data

→ Shape





Actions: Analyze

consume

- -discover vs present
 - classic split
 - aka explore vs explain
- -enjoy
 - newcomer
 - aka casual, social

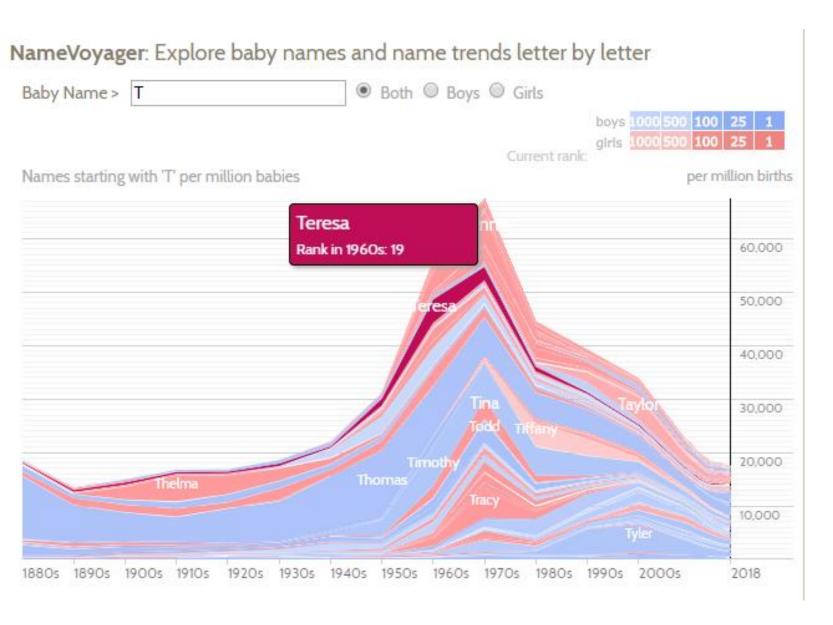
produce

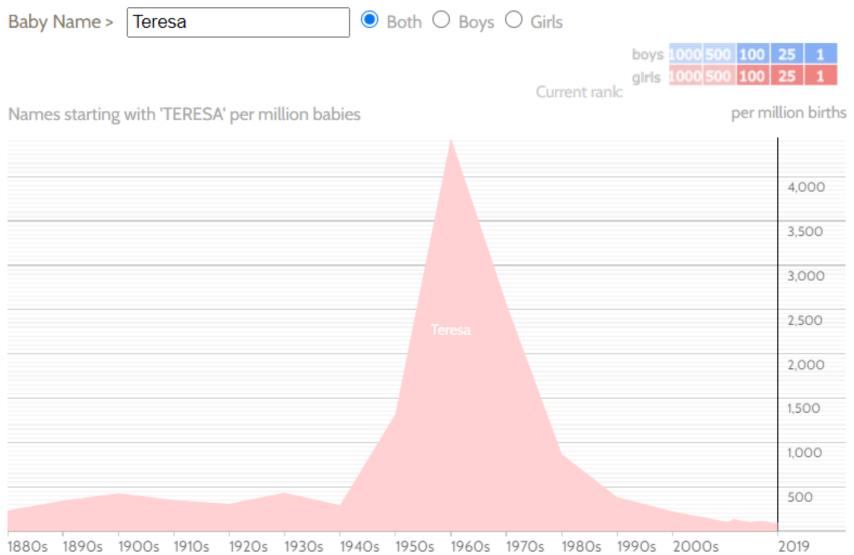
- -annotate, record
- -derive
 - crucial design choice



Enjoy

Name Voyager





Annotate



PERSON 1 ORG 2 PRODUCT 3 DATE 4

In a March 2014 DATE interview, Apple org designer Jonathan

Ive Person used the iPhone Product as an example of Apple org

's ethos of creating high - quality, life - changing products.

LabelMe

https://prodi.gy/

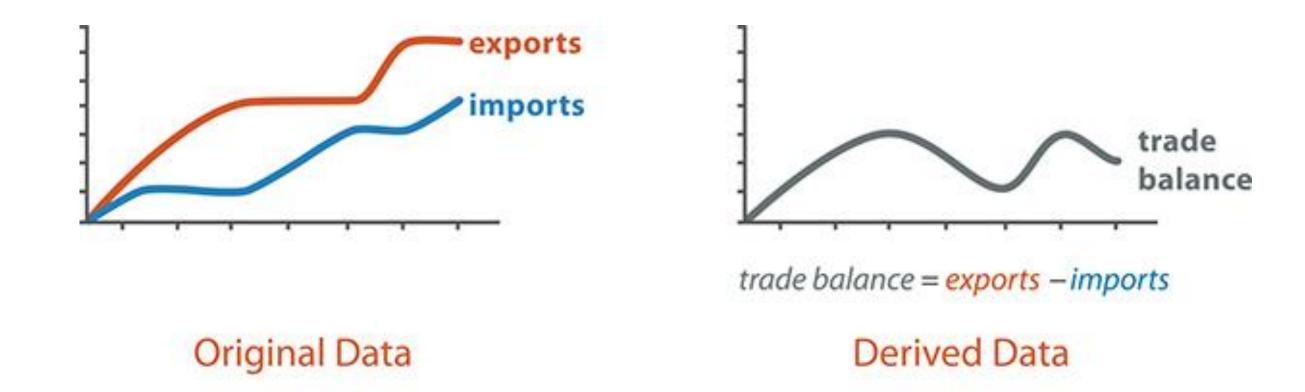
Record



• Working history. [Heer et al. 08]

Derive

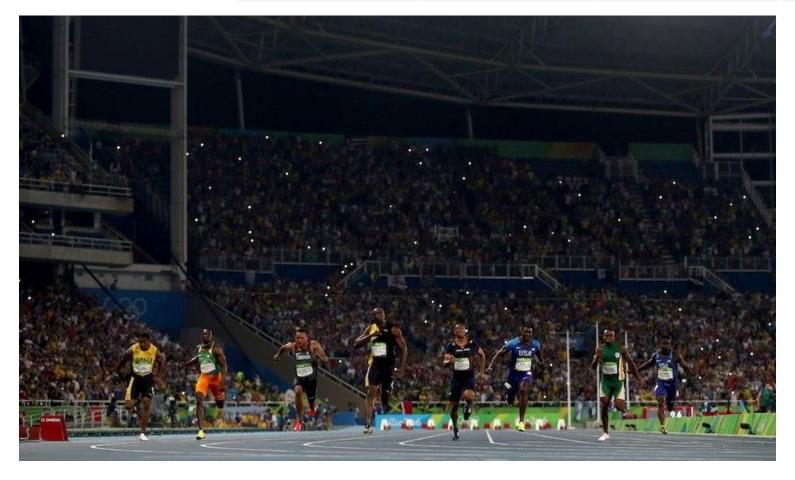
- don't just draw what you're given!
 - -decide what the right thing to show is
 - -create it with a series of transformations from the original dataset
 - -draw that
- one of the four major strategies for handling complexity



NY Times - Men's 100-Meter Sprint

• Olympic 2012

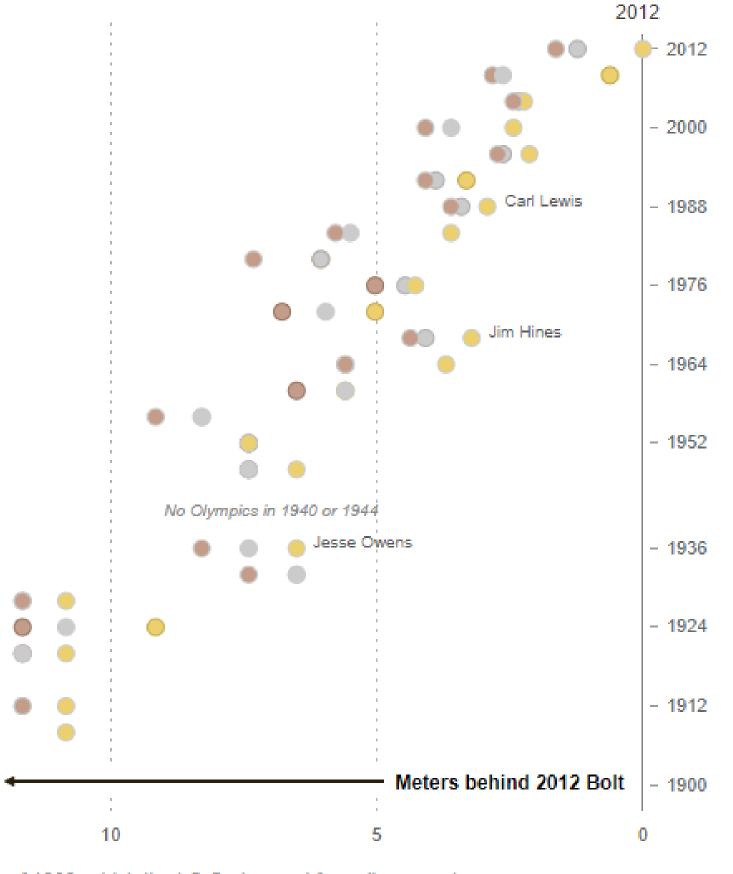
	Gold	Silver	Bronze
2012	9.63	9.75	9.79
2008	9.69	9.89	9.91
2004	9.85	9.86	9.87



Usain Bolt vs. 116 years of Olympic sprinters

Based on the athletes' average speeds, if every Olympic medalist raced each other, Usain Bolt (the London version) would win, with a wide distribution of Olympians behind him. Below, where each sprinter would be when Bolt finishes his race.





Usain Bolt

This chart includes medals for the United States and Australia in the "Intermediary" Games of 1906, which the I.O.C. does not formally recognize.

15

Thomas Burke

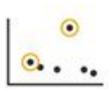
Actions: Search, query

- 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



	Target known	Target unknown	
Location known	• . Lookup	• Browse	
Location unknown	C.O. Locate	< O Explore	

- Query
 - → Identify



→ Compare



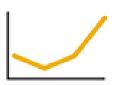
→ Summarize

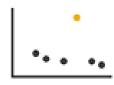


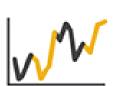
Why: Targets



- → All Data
 - → Trends
- → Outliers
- → Features



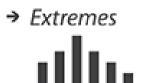




- → Attributes
 - → One

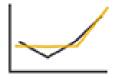
- → Many
- → Distribution
- → Dependency
- → Correlation
- → Similarity











- Network Data
 - → Topology













- Spatial Data
 - → Shape

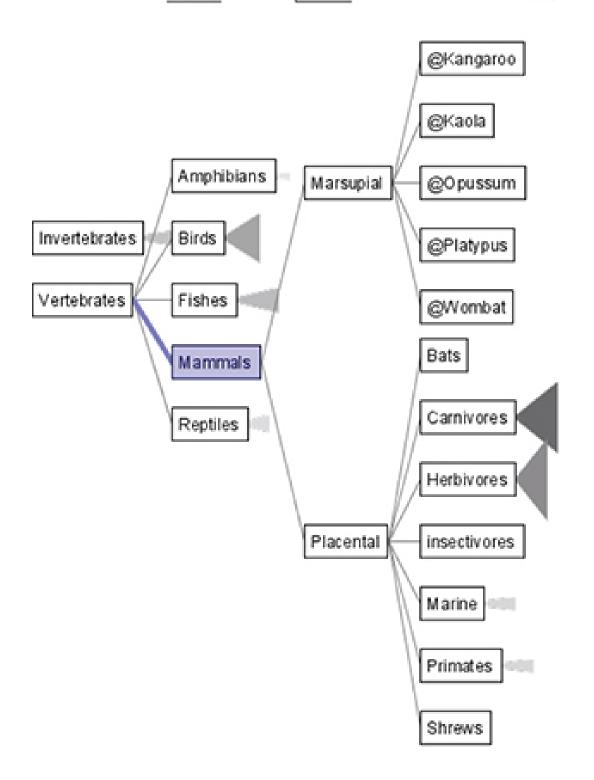


A preview of How

How?

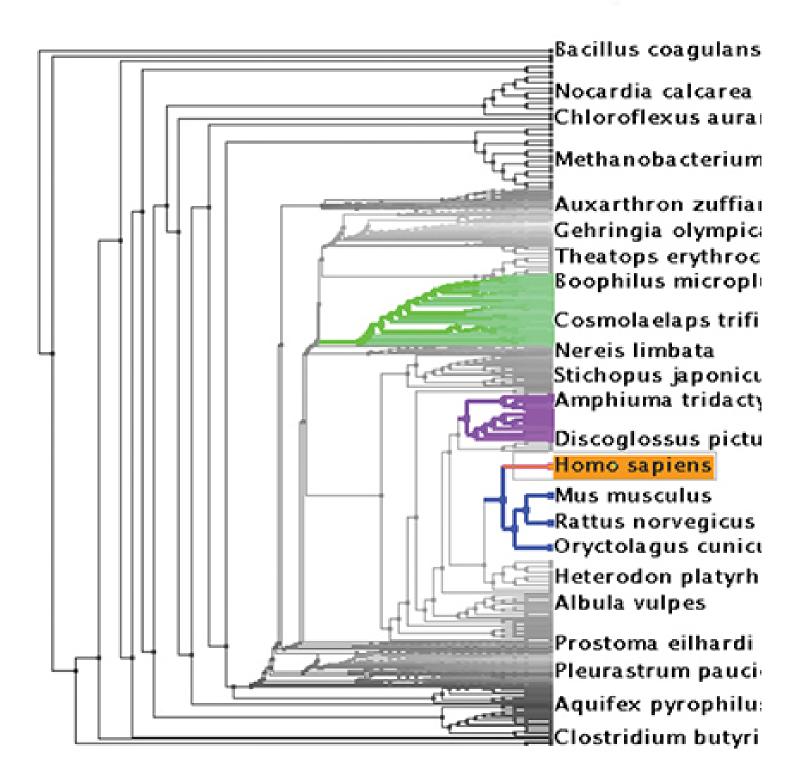
→ SpaceTree





→ TreeJuxtaposer

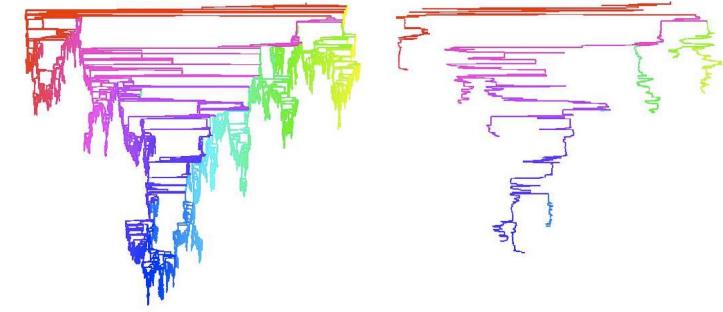


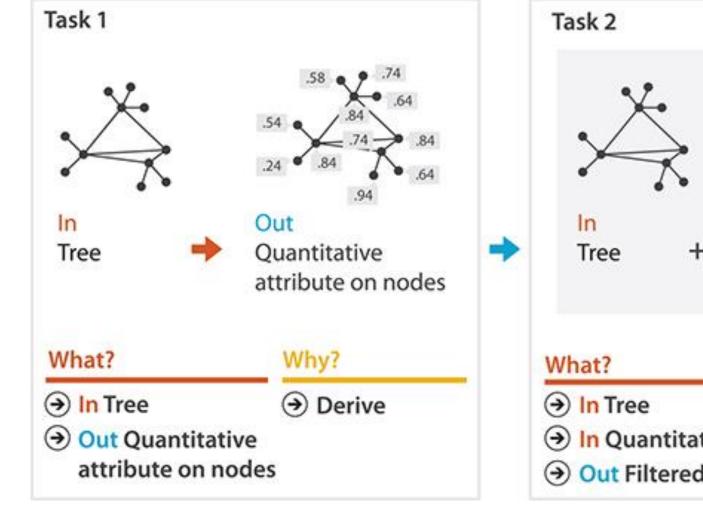


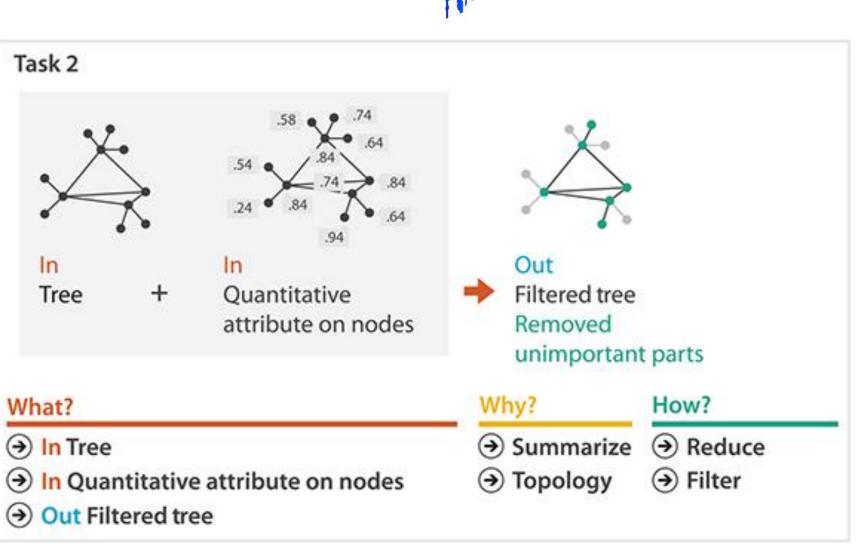
Analysis example: Derive one attribute

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

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

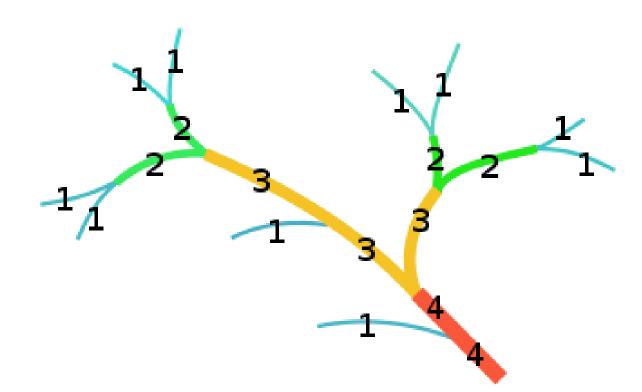






Strahler number

- If the node is a leaf (has no children), its Strahler number is one.
- If the node has one child with Strahler number *i*, and all other children have Strahler numbers less than *i*, then the Strahler number of the node is *i* again.
- If the node has two or more children with Strahler number *i*, and no children with greater number, then the Strahler number of the node is *i* + 1.



Encode Arrange → Express → Separate → Order → Align → Use → Map from categorical and ordered attributes → Color → Hue → Saturation → Luminance → Size, Angle, Curvature, ... 1/2 11)) → Shape → Motion

Direction, Rate, Frequency, ...

How? Manipulate **Facet** Reduce → Change → Juxtapose → Filter ... 0 1111 → Select Partition → Aggregate • • • → Navigate Superimpose → Embed سلىلىسا < >



Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
 - Chap 2: What: Data Abstraction
 - Chap 3: Why: Task Abstraction
- A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis) 19:12 (2013), 2376–2385.
- Low-Level Components of Analytic Activity in Information Visualization. Amar, Eagan, and Stasko. Proc. IEEE InfoVis 2005, p 111–117.
- A taxonomy of tools that support the fluent and flexible use of visualizations. Heer and Shneiderman. Communications of the ACM 55:4 (2012), 45–54.
- Rethinking Visualization: A High-Level Taxonomy. Tory and Möller. Proc. IEEE InfoVis 2004, p 151–158.
- Visualization of Time-Oriented Data. Aigner, Miksch, Schumann, and Tominski.
 Springer, 2011.

Reading visualization papers

- one strategy: multiple passes
 - title
 - abstract, authors/affiliation
 - flip through, glance at figures, notice structure from section titles
 - skim intro, results/discussion (maybe conclusion)
 - fast read to get big ideas
 - if you don't get something, just keep going
 - second pass to work through details
 - later parts may cast light on earlier parts for badly structured papers
 - third pass to dig deep
 - if it's highly relevant, or you're presenting it to class
- literature search
 - decide when to stop reading: is this relevant to my current concerns?

Marks and Channels

Definitions: Marks and channels

- marks
 - –geometric primitives















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

→ Position









→ Both





→ Shape









→ Tilt



→ Size







→ Volume

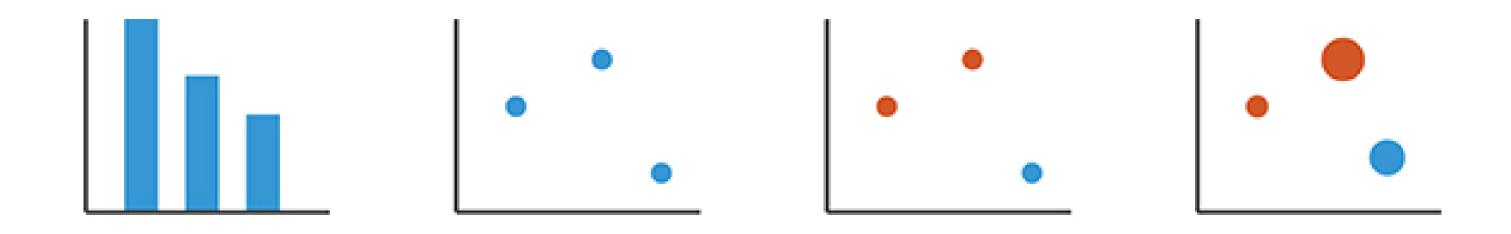






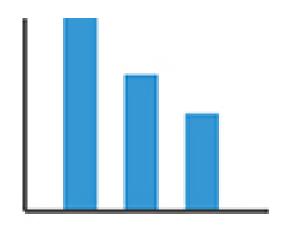
Visual encoding

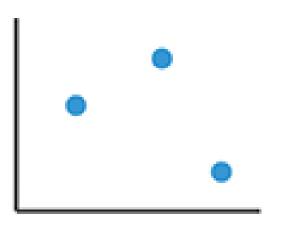
• analyze idiom structure

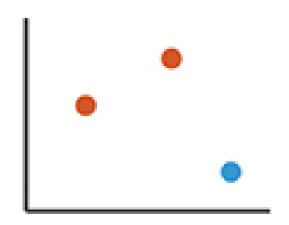


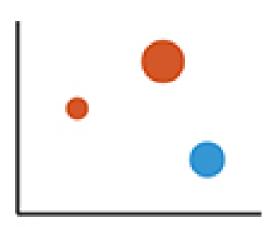
Visual encoding

- analyze idiom structure
 - -as combination of marks and channels









1: vertical position

2: vertical position horizontal position

3: vertical position horizontal position color hue

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

mark: line

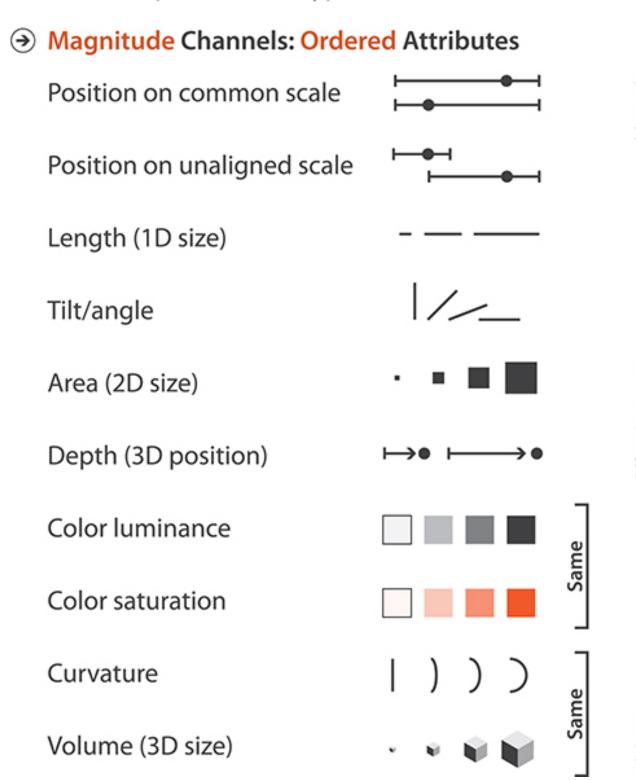
mark: point

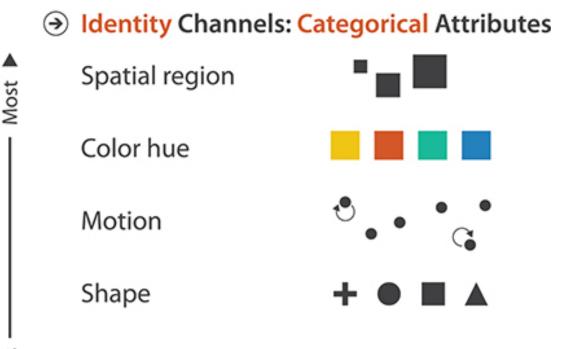
mark: point

mark: point

Channels: Expressiveness types and effectiveness rankings

Channels: Expressiveness Types and Effectiveness Ranks

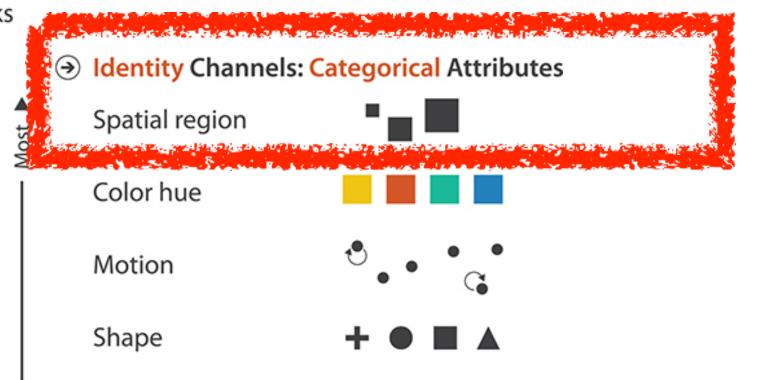




- expressiveness principle
 - match channel and data characteristics
- effectiveness principle
 - encode most important attributes with highest ranked channels

Channels: Expressiveness types and effectiveness rankings

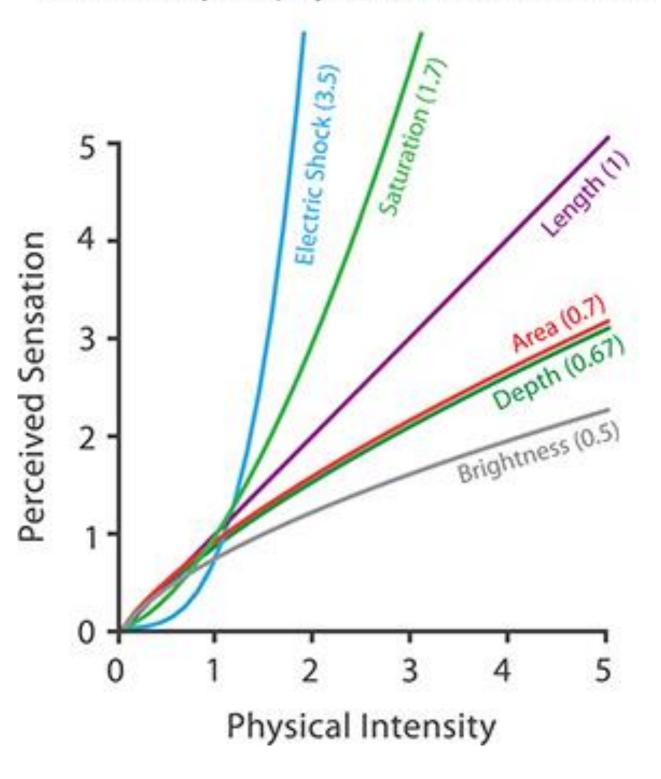
Channels: Expressiveness Types and Effectiveness Raks → Magnitude Channels: Ordered Attributes Position on common scale Position on unaligned scale Length (1D size) Tilt/angle Area (2D size) Depth (3D position) Color luminance Color saturation Curvature Volume (3D size)



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

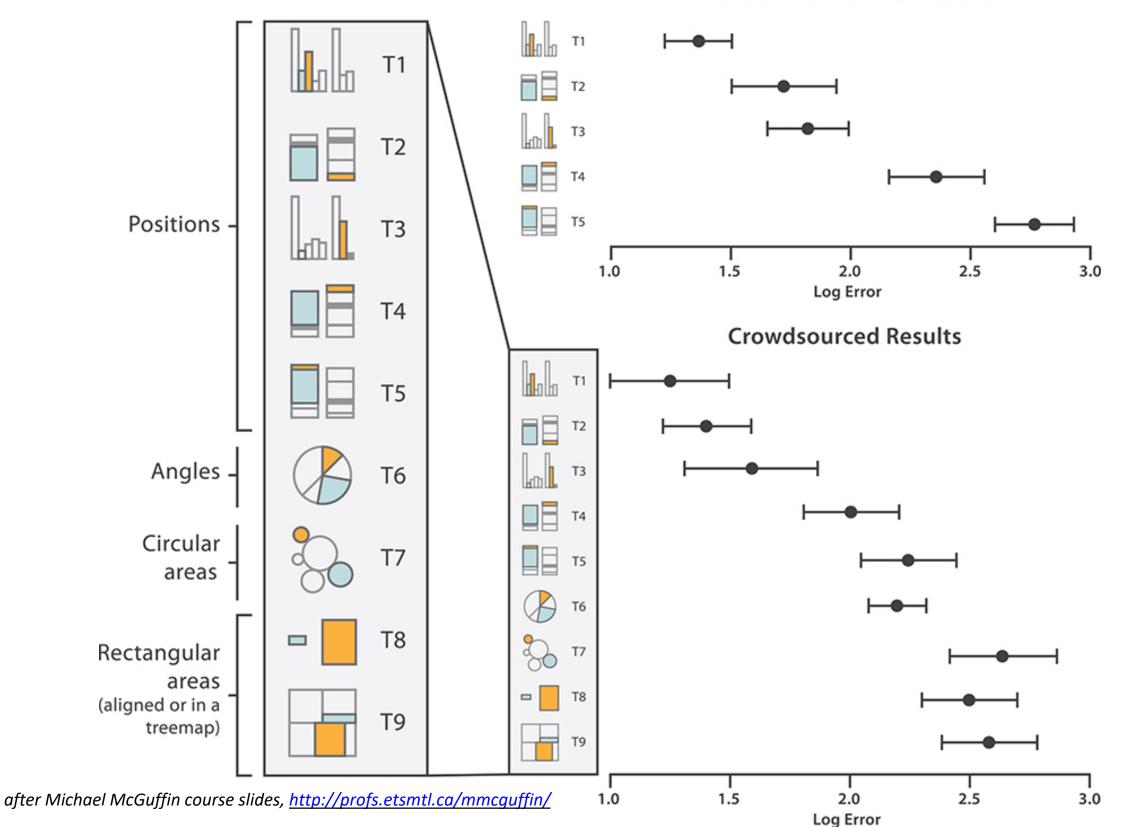
Accuracy: Fundamental Theory

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



Accuracy: Vis experiments

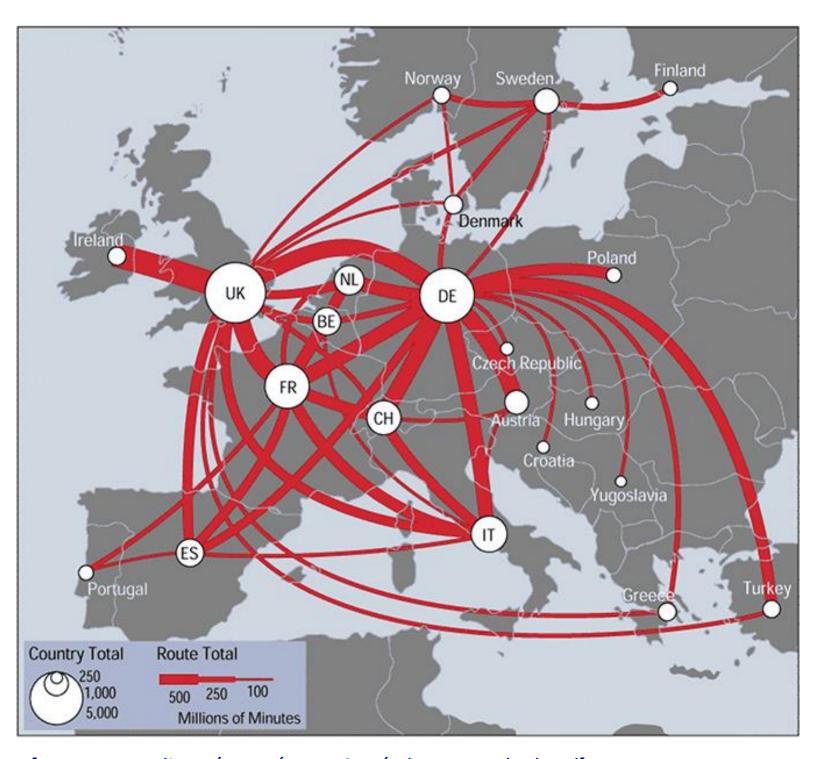
Cleveland & McGill's Results



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

Discriminability: How many usable steps?

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



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

Separability vs. Integrality

Position + Hue (Color) + Height + Green

Fully separable Some interference Some/significant Major interference

2 groups each

2 groups each

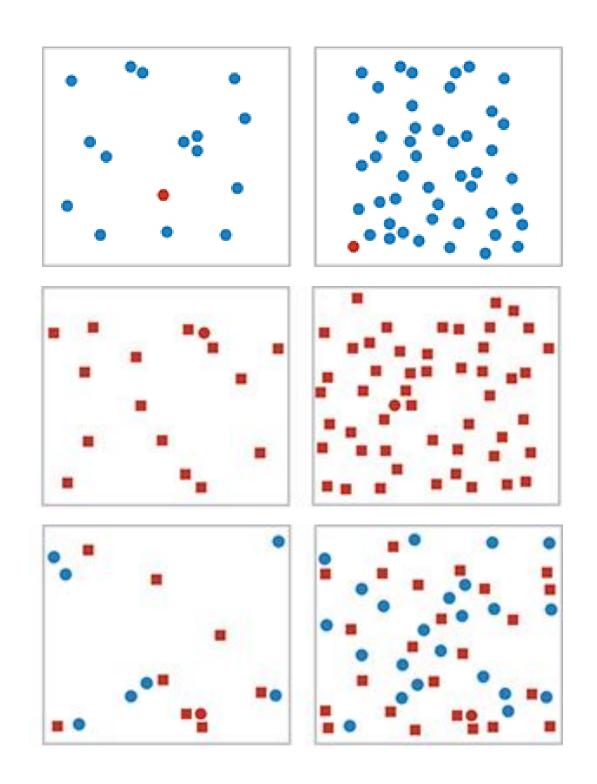
3 groups total: integral area

interference

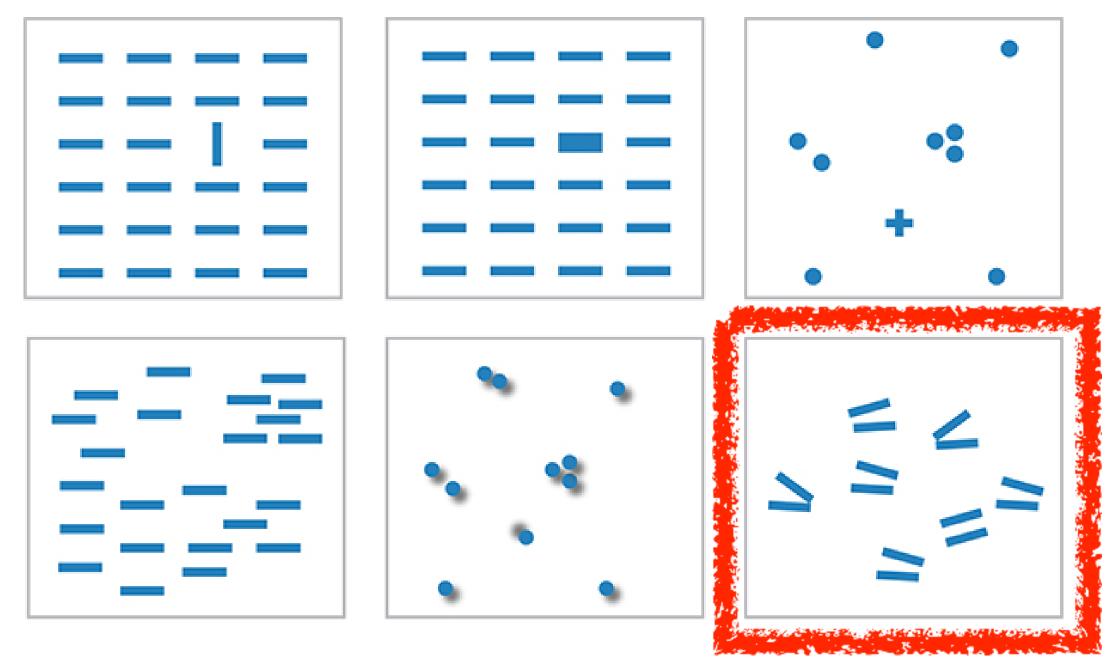
4 groups total: integral hue

Popout

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



Popout



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

Grouping

- containment
- connection

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

Marks as Links

→ Containment



Connection



→ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion

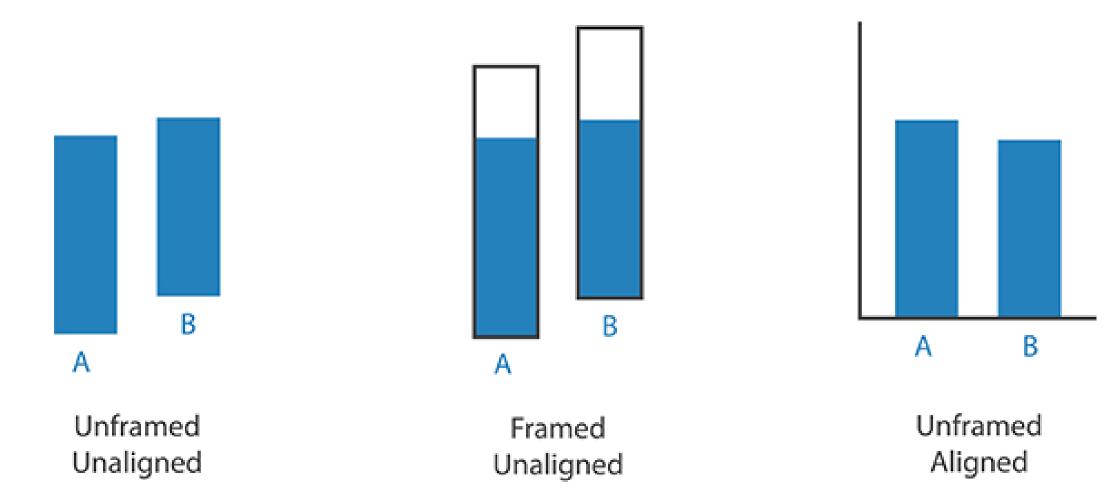


Shape



Relative vs. absolute judgements

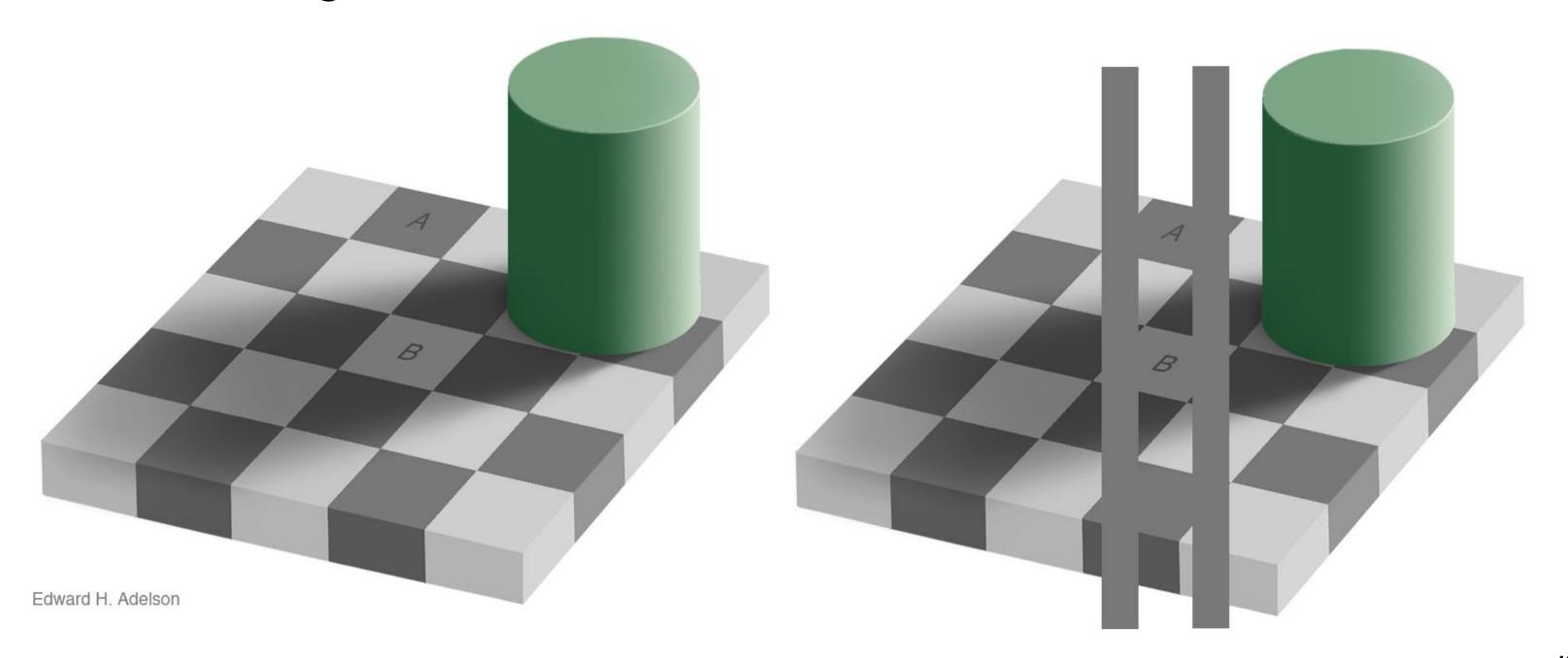
- perceptual system mostly operates with relative judgements, not absolute
 - that's why accuracy increases with common frame/scale and alignment
 - Weber's Law: ratio of increment to background is constant
 - filled rectangles differ in length by 1:9, difficult judgement
 - white rectangles differ in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



after [Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Cleveland and McGill. Journ. American Statistical Association 79:387 (1984),45 [1984]

Relative luminance judgements

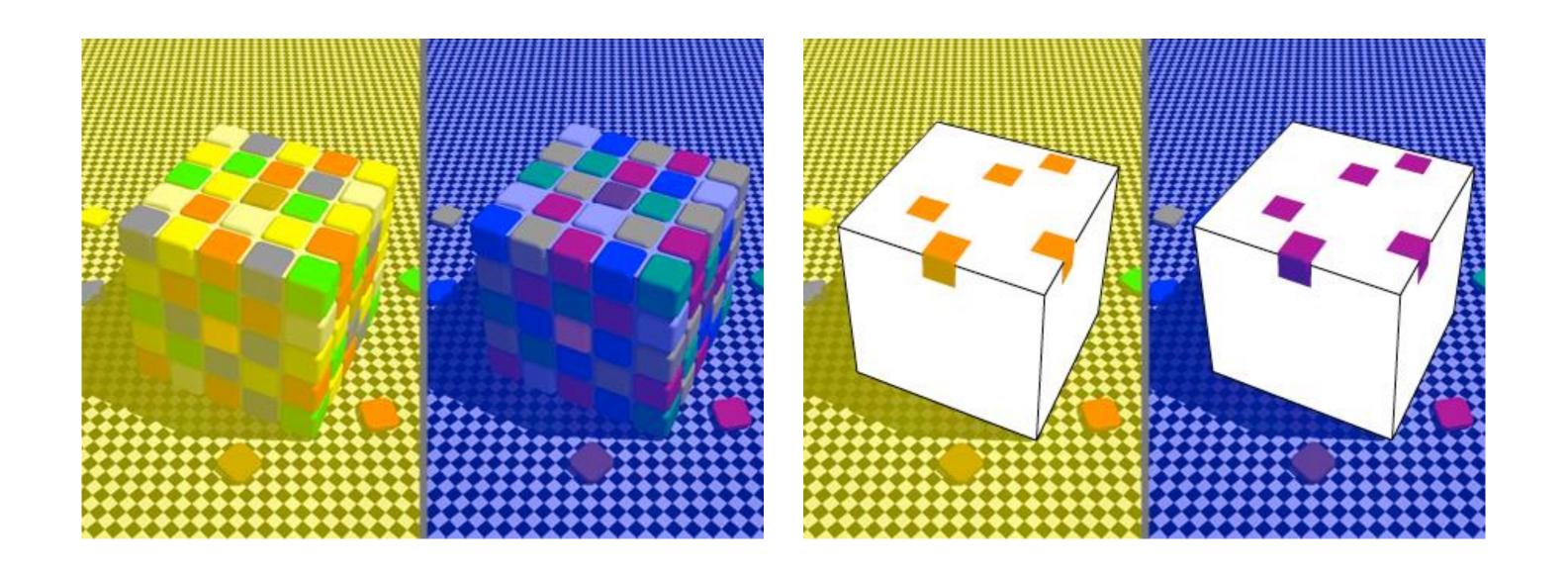
 perception of luminance is contextual based on contrast with surroundings



46

Relative color judgements

color constancy across broad range of illumination conditions



47

Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
 - Chap 5: Marks and Channels
- On the Theory of Scales of Measurement. Stevens. Science 103:2684 (1946), 677–680.
- Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. Stevens.
 Wiley, 1975.
- Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Cleveland and McGill. Journ. American Statistical Association 79:387 (1984), 531–554.
- Perception in Vision. Healey. http://www.csc.ncsu.edu/faculty/healey/PP
- Visual Thinking for Design. Ware. Morgan Kaufmann, 2008.
- Information Visualization: Perception for Design, 3rd edition. Ware. Morgan Kaufmann / Academic Press, 2004.