Information Visualization
Data, Tasks, Nested Model
Ex: Abstractions

Tamara Munzner
Department of Computer Science
University of British Columbia

Week 2: 15 September 2021

http://www.cs.ubc.ca/~tmm/courses/547-21

Course Logistics

Async so far

- last week
  - async read only
    - Course Logistics (no comments, no responses)
  - async read & comment
    - VAD Ch 1: Why Visualization? (comments only, no responses)
  - async discuss
    - self-intros
- this week
  - async read & comment & respond
    - VAD Ch 2: Data Abstraction
    - VAD Ch 3: Task Abstraction
    - paper: Nested Model [basis for VAD Ch 4]

Updates

- All students moved from waitlist to registered
- Official enrolment now 38
- Very likely to move to Forestry (FSC) 2330 starting next week
  - especially if ventilation here in SWNG 207 remains terrible!
- Stay tuned for Canvas marks updates

Discussion: Round 1

Exercise: Abstractions

Now: In-class design exercise, in small groups

- Abstractions
  - practice with data & task abstractions, on concrete example: Aid to Countries
  - crucial ideas: determine cardinalities/ranges
    - precondition for all decisions about visual encoding
- Small-group exercise: 60-ish min
  - breakout groups (4 people/group)
  - googledoc worksheets, as before
  - document in your group's googledoc w/ text as you go!
  - reportbacks, as before (intermediate and final)
    - I'll flip through googledocs, some questions for group spokesperson

Discussion: Round 2

Next week

- to read & discuss (async, before next class)
  - VAD book, Ch 5: Marks & Channels
  - VAD book, Ch 6: Rules of Thumb
  - paper: Design Study Methodology

Backup/Reference Slides

Ch 1. What's Vis, and Why Do It?

Visualization defined & motivated

short version: alternate to next 3 slides

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

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

- human in the loop needs the details
  - doesn't know exactly what questions to ask in advance
  - longterm exploratory analysis
    - ***speed up*** through human-in-the-loop visual data analysis
  - presentation of known results
  - stepping stone towards automation: refining, trustbuilding
  - interplay between human judgement and automatic computation
- intended task, measurable definitions of effectiveness

Defining visualization (vis)

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

Why?...

Visualization (vis) defined & motivated

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

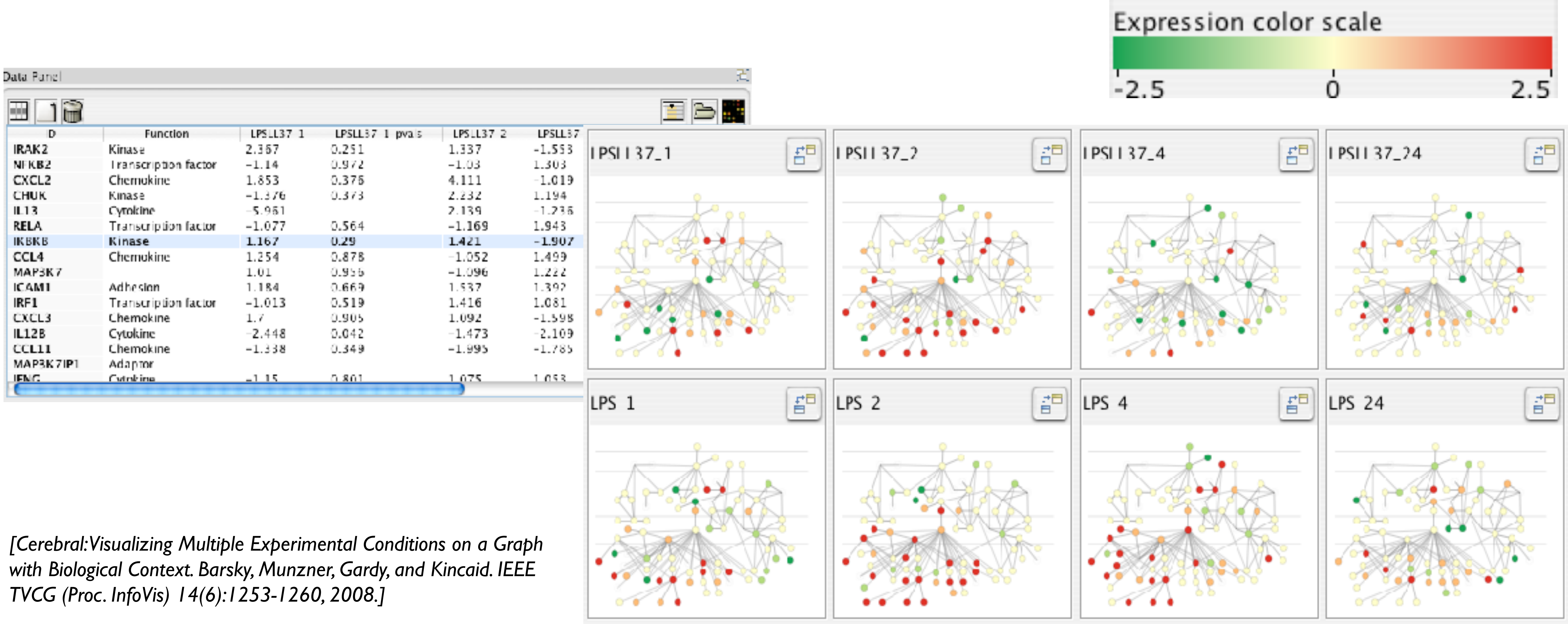
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- human in the loop needs the details & no trusted automatic solution exists
  - doesn't know exactly what questions to ask in advance
  - exploratory data analysis
    - ***speed up*** through human-in-the-loop visual data analysis
  - present known results to others
  - stepping stone towards automation
    - before model creation to provide understanding
    - during algorithm creation to refine, debug, set parameters
    - before or during deployment to build trust and monitor

Why use an external representation?

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

- external representation: replace cognition with perception



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

Why depend on vision?

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

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

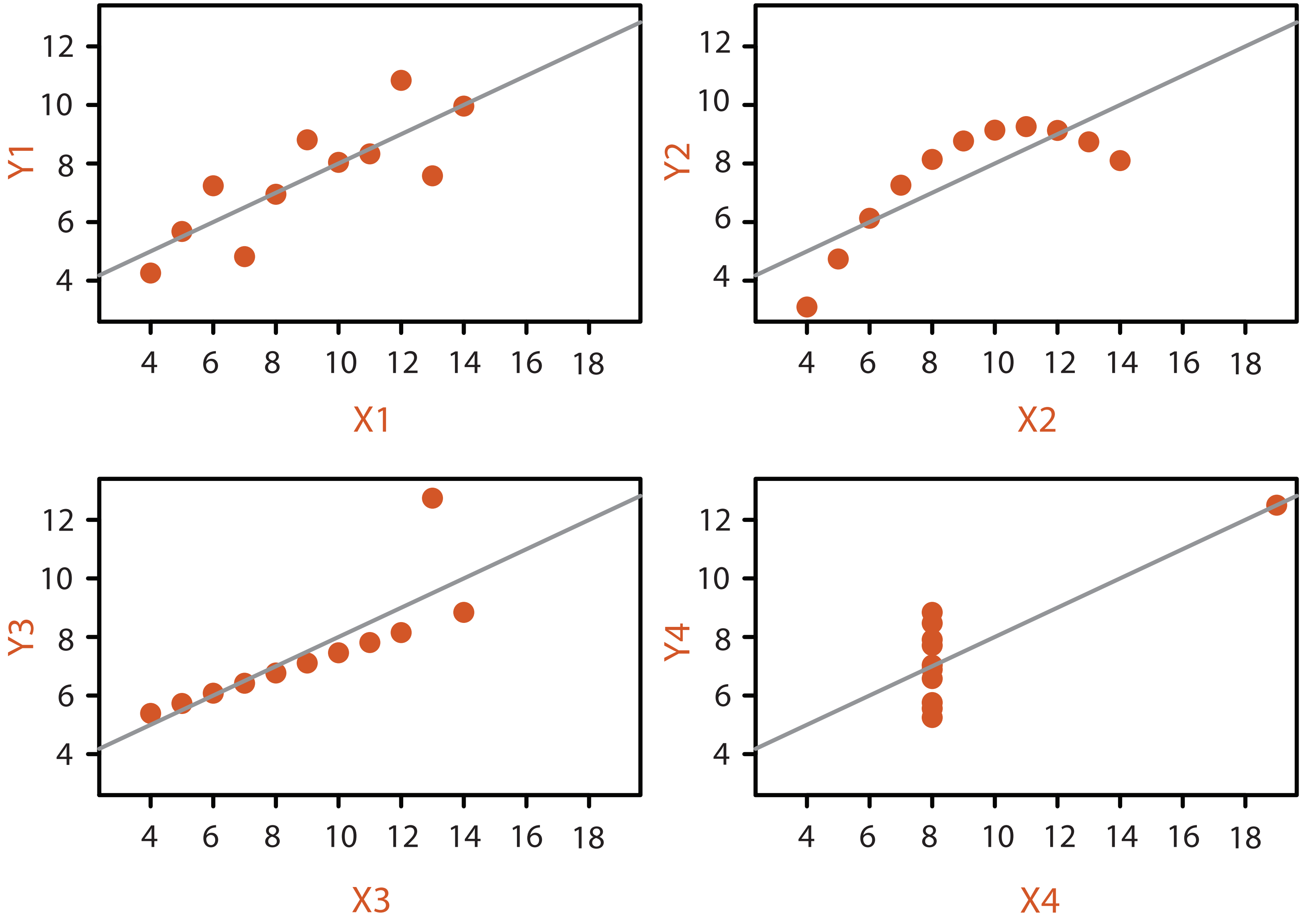
Why represent all the data?

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

- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

Anscombe's Quartet

Identical statistics	
x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816



https://www.youtube.com/watch?v=DbJyPELmhJc

Same Stats, Different Graphs

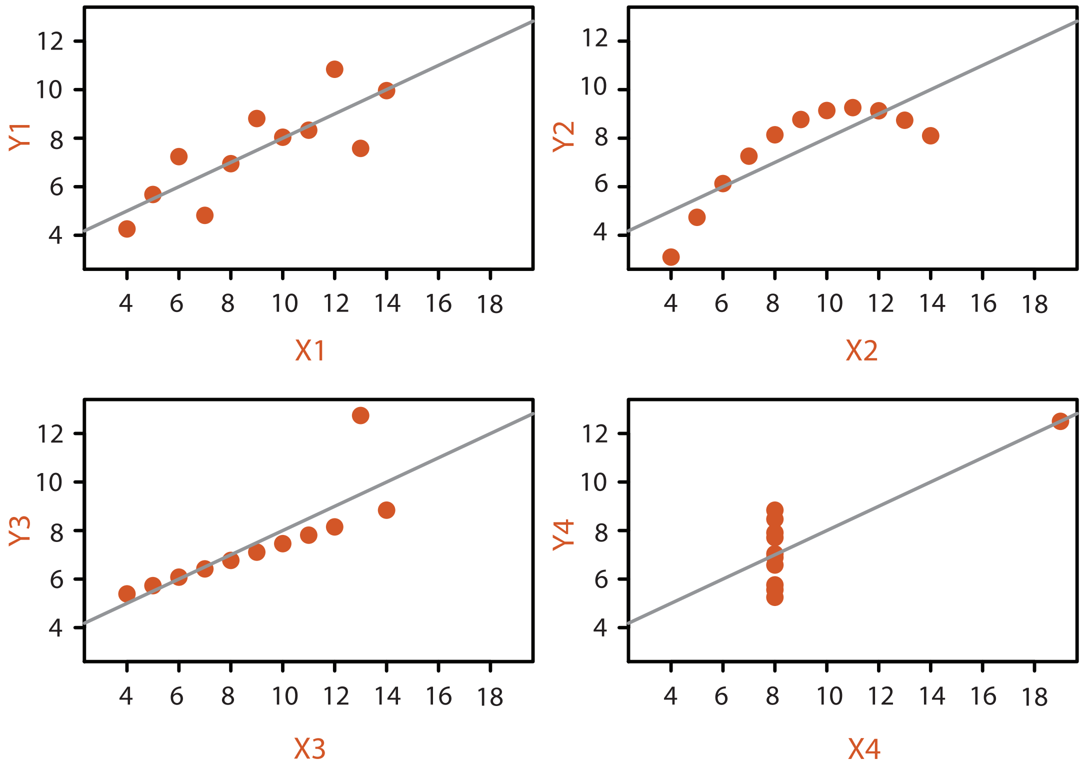
Visualization defined & motivated

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

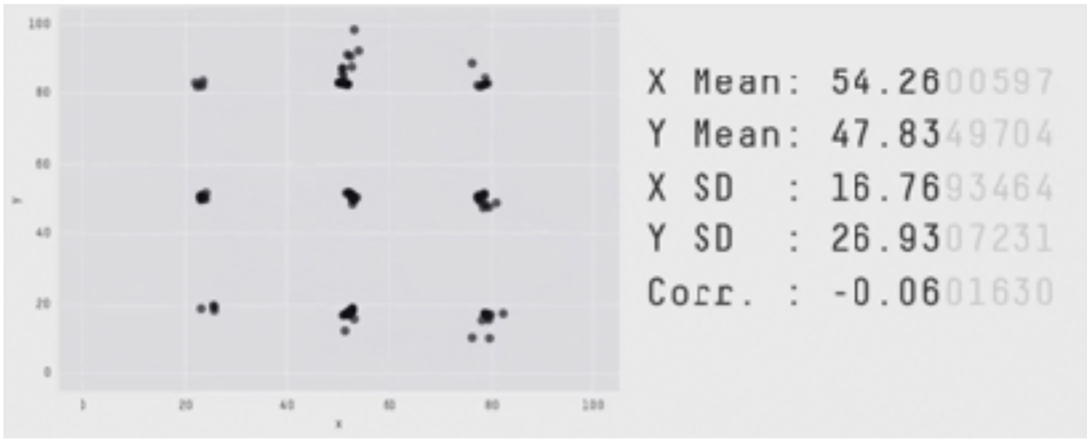
- suitable when human in the loop needs details
  - interplay between human judgement and automatic computation

Anscombe's Quartet

Identical statistics	
x mean	9
x variance	10
y mean	7.5
y variance	3.75
x/y correlation	0.816



Datasaurus Dozen



Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. CHI 2017. Matejka & Fitzmaurice

Why focus on tasks and effectiveness?

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

- effectiveness requires match between data/task and representation
  - set of representations is huge
  - many are ineffective mismatch for specific data/task combo
  - increases chance of finding good solutions if you understand full space of possibilities
- what counts as effective?
  - novel: enable entirely new kinds of analysis
  - faster: speed up existing workflows
- how to validate effectiveness
  - many methods, must pick appropriate one for your context

What resource limitations are we faced with?

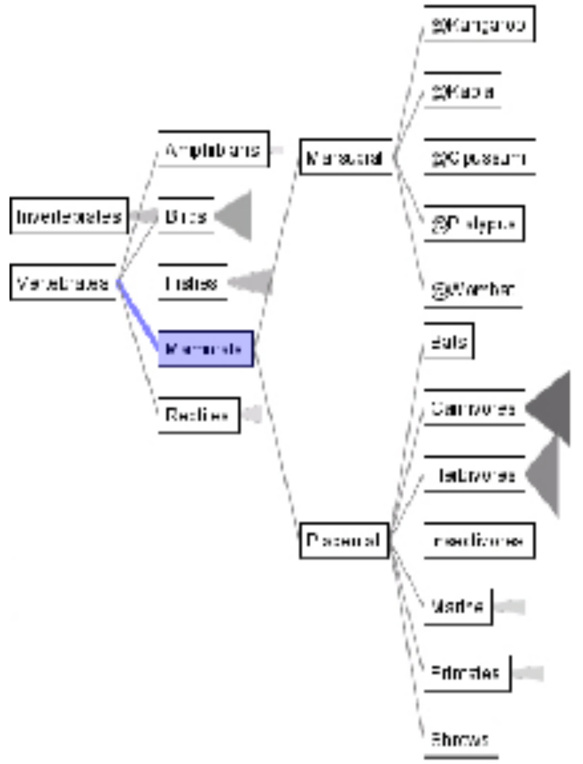
Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

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

Why analyze?

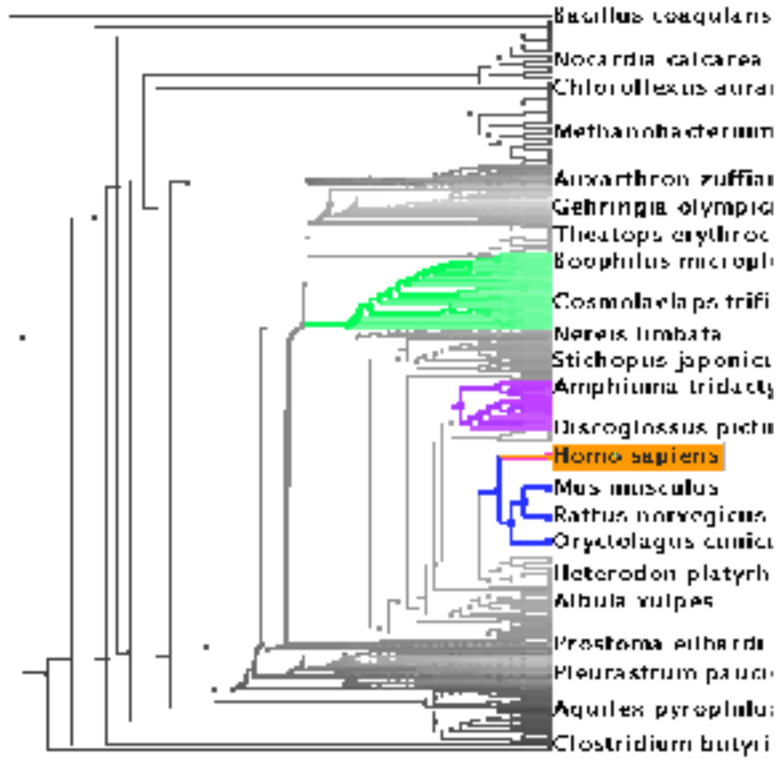
- imposes structure on huge design space
  - scaffold to help you think systematically about choices
  - analyzing existing as stepping stone to designing new
  - most possibilities ineffective for particular task/data combination

SpaceTree



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

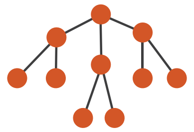
TreeJuxtaposer



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

What?

➔ Tree



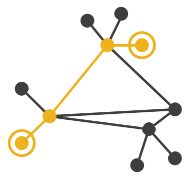
Why?

➔ Actions

→ Present → Locate → Identify

➔ Targets

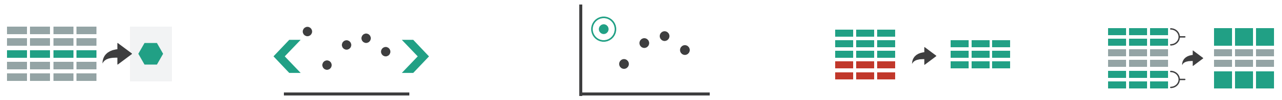
→ Path between two nodes



How?

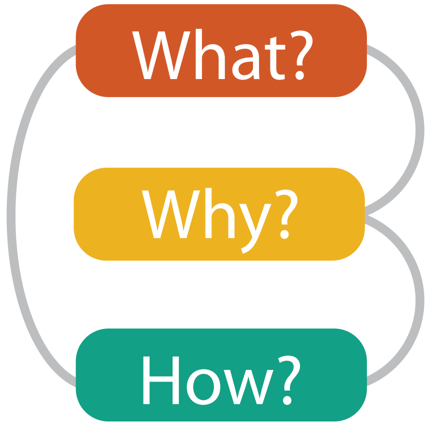
➔ SpaceTree

→ Encode → Navigate → Select → Filter → Aggregate



➔ TreeJuxtaposer

→ Encode → Navigate → Select → Arrange



How?

Encode

Arrange
- Express
- Separate
- Order
- Align
- Use

Map
from categorical and ordered attributes

- Color
  - *Hue*
  - *Saturation*
  - *Luminance*
- Size, Angle, Curvature, ...
- Shape
- Motion
  Direction, Rate, Frequency, ...

Manipulate

Change
Select
Navigate

Facet

Juxtapose
Partition
Superimpose

Reduce

Filter
Aggregate
Embed

What?

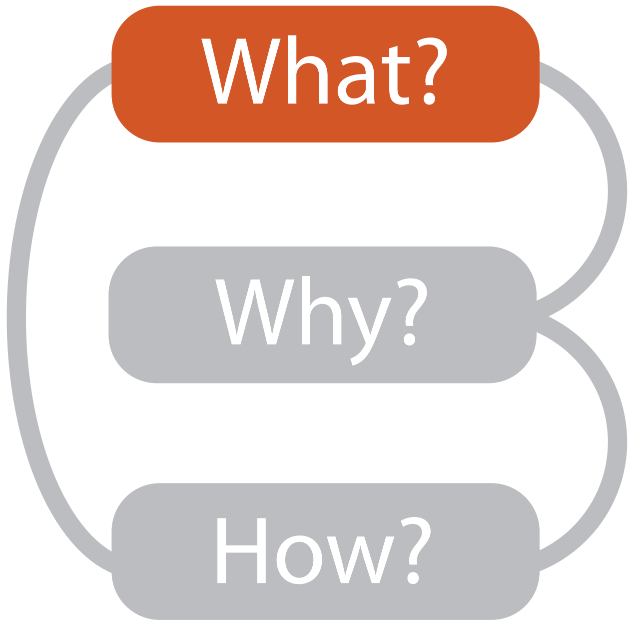
Why?

How?

Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - Chap 1: What's Vis, and Why Do It?
- The Nature of External Representations in Problem Solving. Jiajie Zhang. Cognitive Science 21:2 (1997), 179-217.
- A Representational Analysis of Numeration Systems. Jiajie Zhang and Donald A. Norman. Cognition 57 (1995), 271-295.
- Why a Diagram Is (Sometimes) Worth Ten Thousand Words.. Jill H. Larkin and Herbert A. Simon. Cognitive Science 11:1 (1987), 65-99.
- Graphs in Statistical Analysis.F.J. Anscombe. American Statistician 27 (1973), 17-21.
- Design Study Methodology: Reflections from the Trenches and the Stacks. Michael Sedlmair, Miriah Meyer, and Tamara Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2012), 18(12):2431-2440, 2012.
- Information Visualization: Perception for Design, 3rd edition, Colin Ware, Morgan Kaufmann, 2013.
- Current approaches to change blindness Daniel J. Simons. Visual Cognition 7, 1/2/3 (2000), 1-15.
- Semiology of Graphics, Jacques Bertin, Gauthier-Villars 1967, EHESS 1998
- The Visual Display of Quantitative Information. Edward R. Tufte. Graphics Press, 1983.

Ch 2. What: Data Abstraction



What?

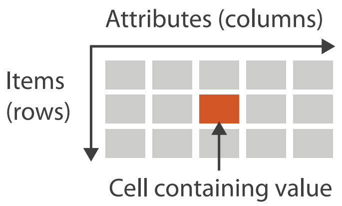
Datasets

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

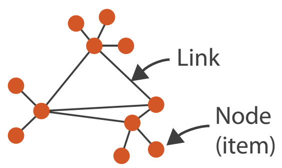
Data and Dataset Types

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

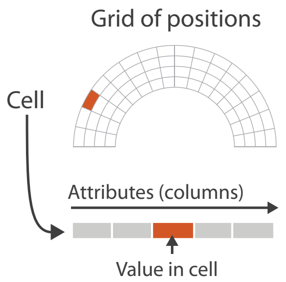
Dataset Types
- Tables



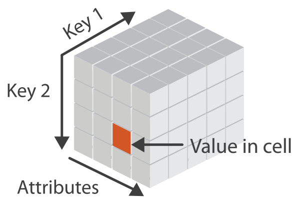
- Networks



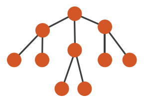
- Fields (Continuous)



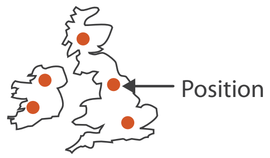
- *Multidimensional Table*



- *Trees*



- Geometry (Spatial)



Dataset Availability
- Static

- Dynamic

Attributes

Attribute Types
- Categorical

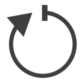
- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction
- Sequential

- Diverging

- Cyclic

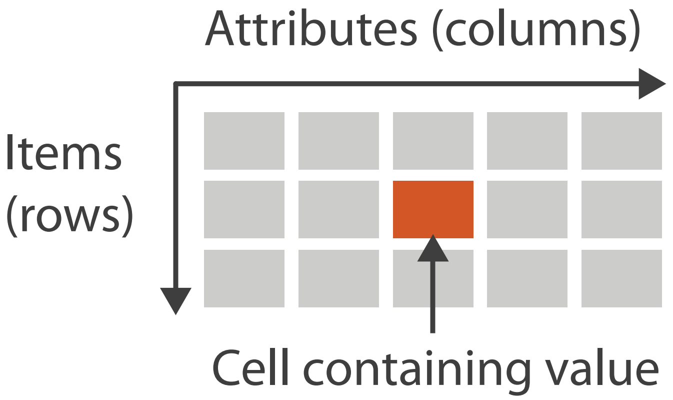


Types: Datasets and data

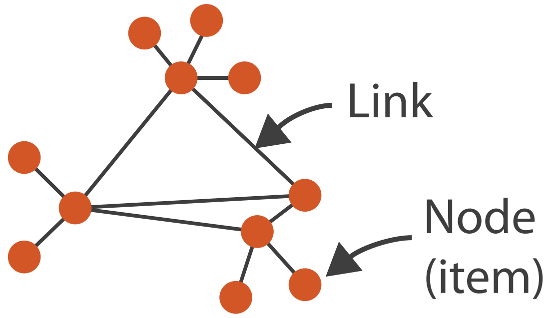
short version: alternate to next 3 slides

Dataset Types

→ Tables

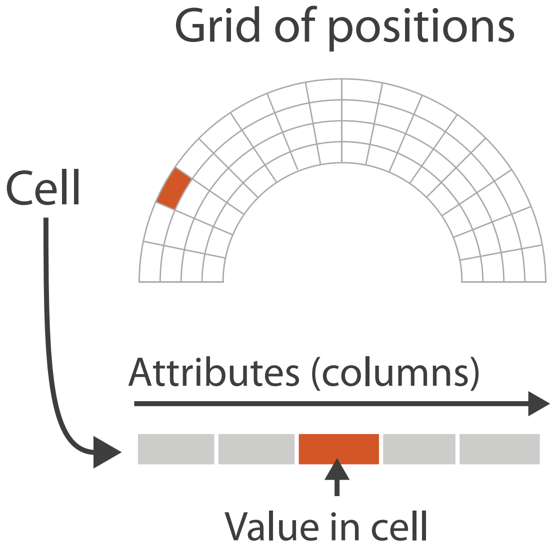


→ Networks

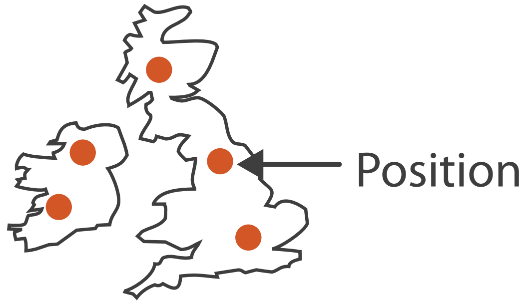


→ Spatial

→ Fields (Continuous)



→ Geometry (Spatial)

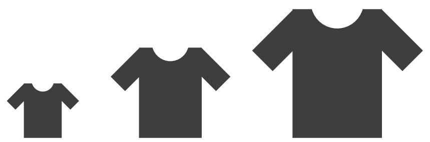


Attribute Types

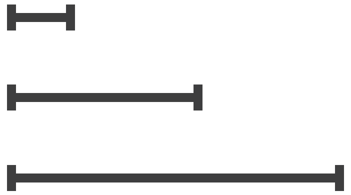
→ Categorical

→ Ordered

→ *Ordinal*



→ *Quantitative*

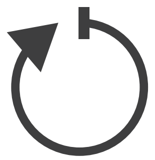


Ordering Direction

→ Sequential

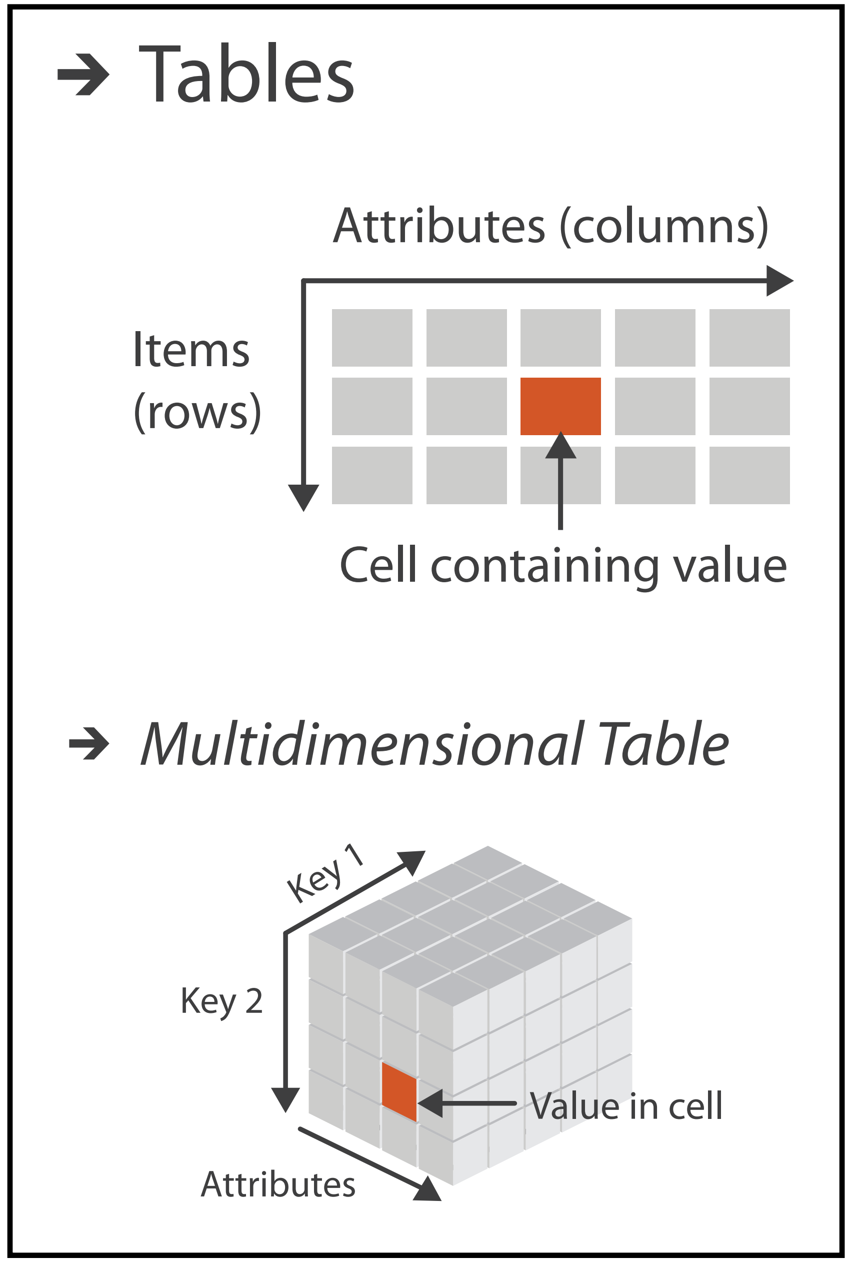
→ Diverging

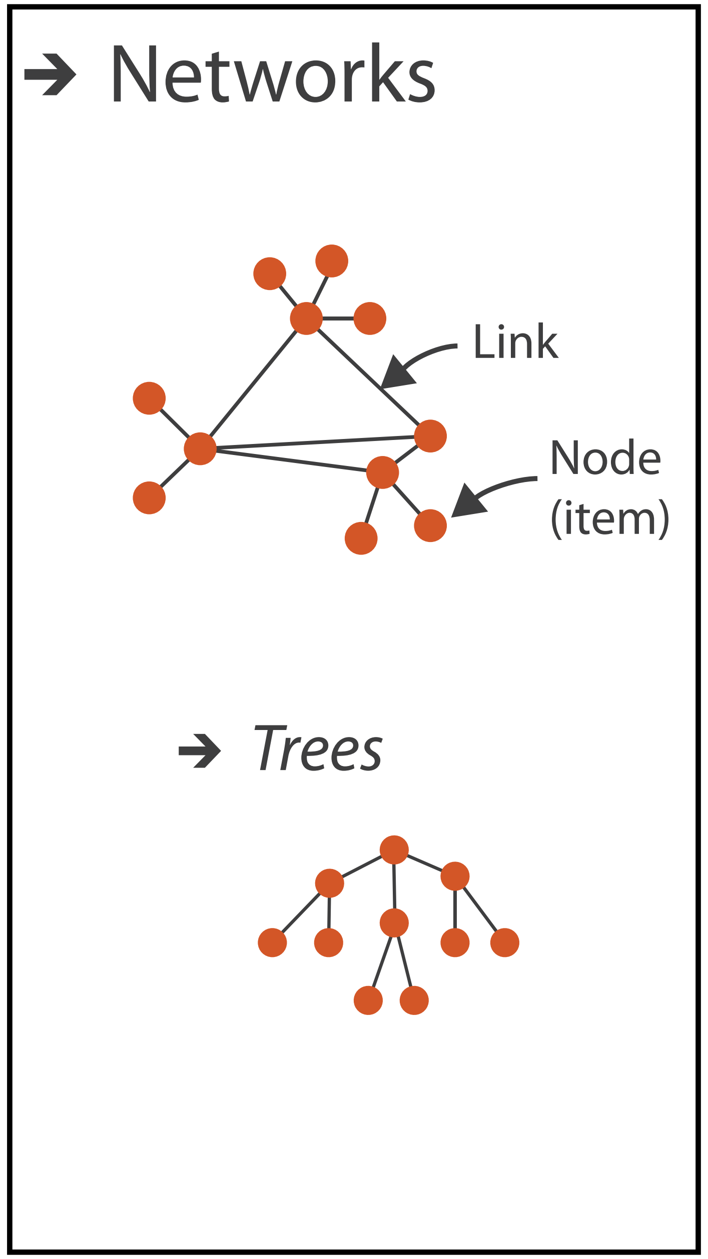
→ Cyclic

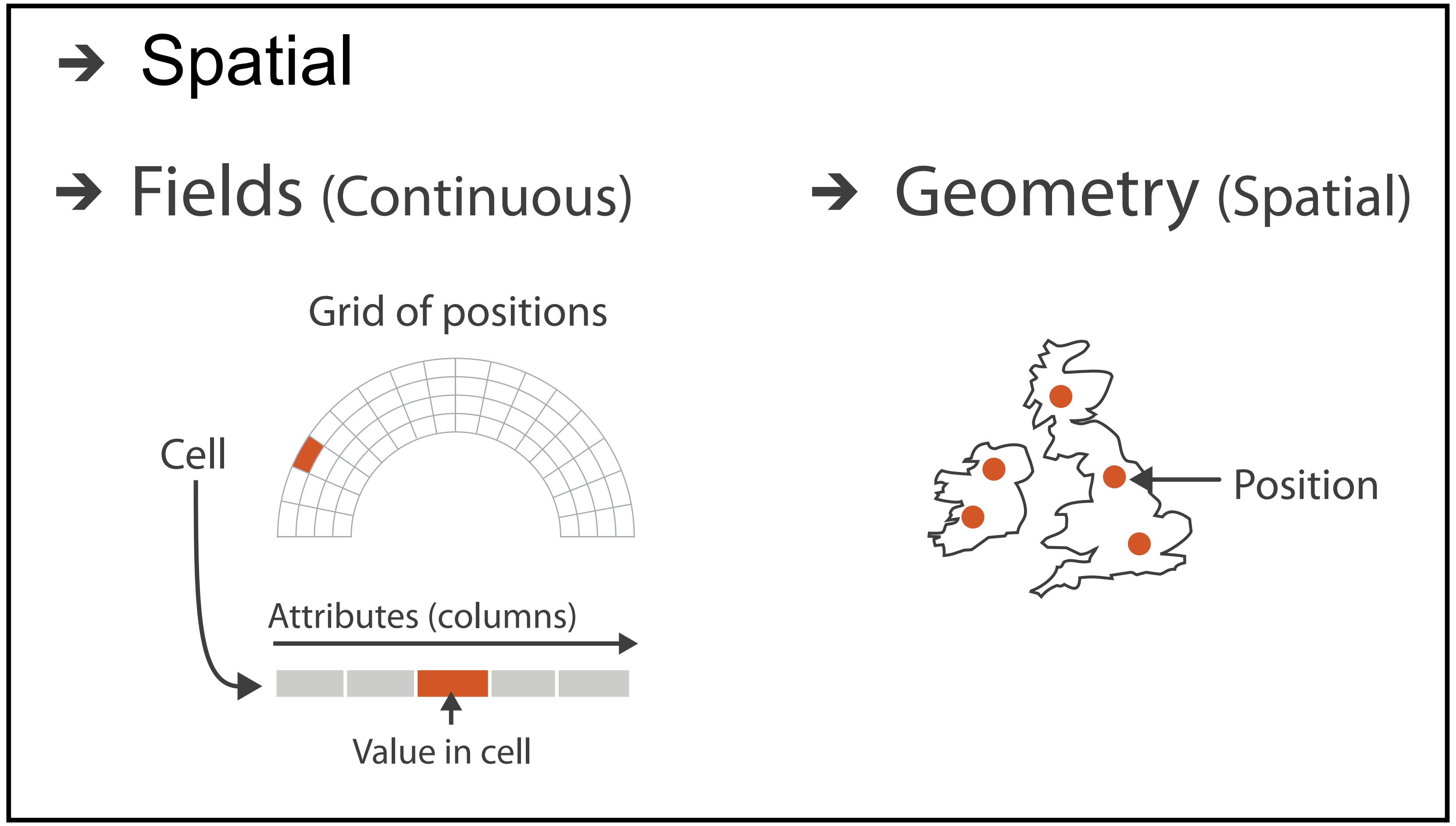


Three major datatypes

Dataset Types







- visualization vs computer graphics
  - geometry is design decision

Dataset and data types

➔ Data and Dataset Types

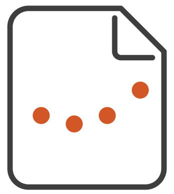
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

➔ Data Types

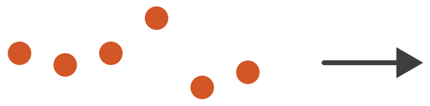
➔ Items ➔ Attributes ➔ Links ➔ Positions ➔ Grids

➔ Dataset Availability

➔ Static



➔ Dynamic

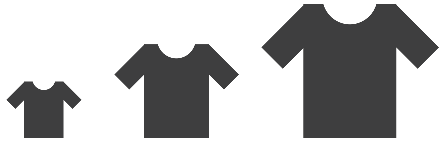


Attribute types

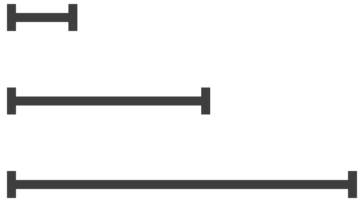
➔ Attribute Types

- ➔ Categorical

- ➔ Ordered
  - ➔ *Ordinal*



  - ➔ *Quantitative*

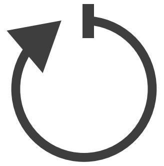


➔ Ordering Direction

- ➔ Sequential

- ➔ Diverging

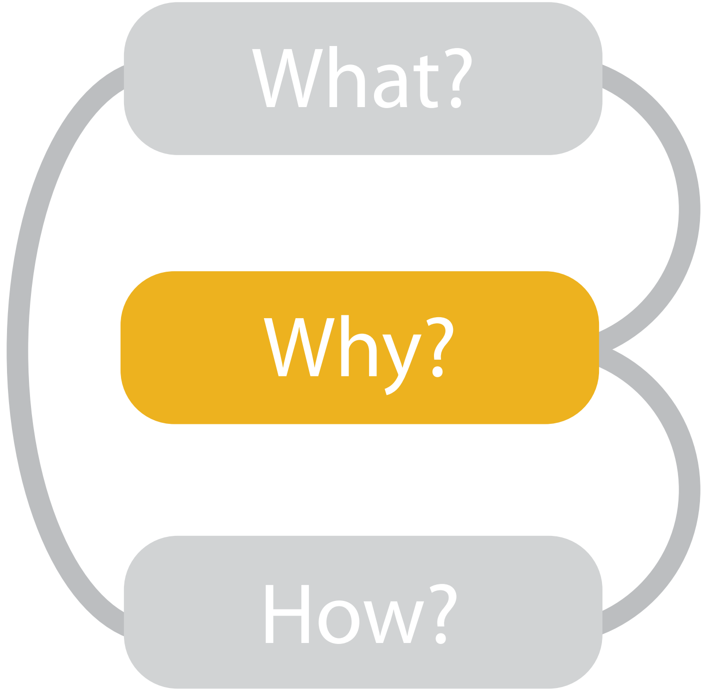
- ➔ Cyclic



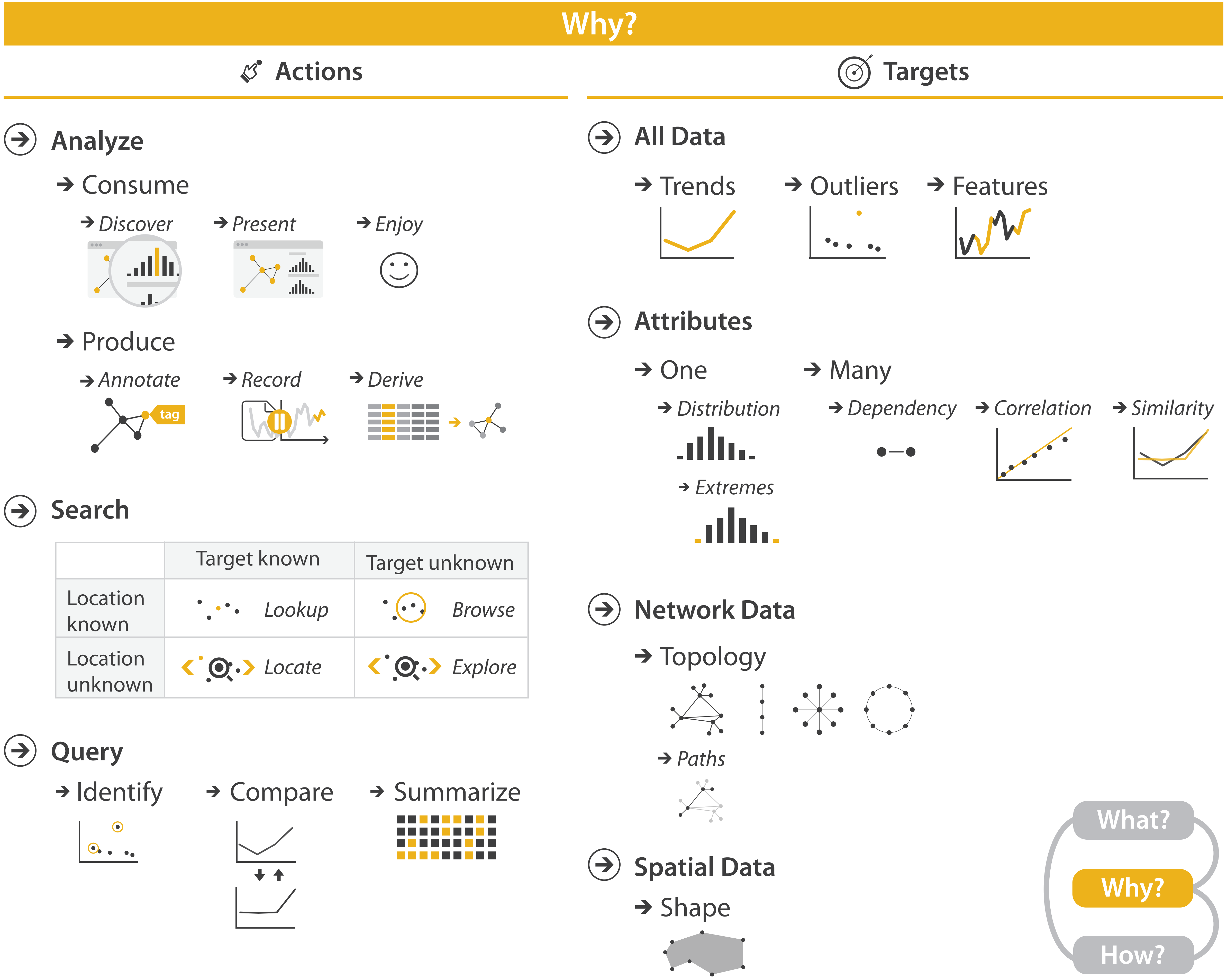
Further reading, full Ch 2

- Readings in Information Visualization: Using Vision To Think, Chapter 1. Stuart K. Card, Jock Mackinlay, and Ben Shneiderman. Morgan Kaufmann, 1999.
- Rethinking Visualization: A High-Level Taxonomy. InfoVis 2004, p 151-158, 2004.
- The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations Ben Shneiderman, Proc. 1996 IEEE Visual Languages
- Data Visualization: Principles and Practice, 2nd ed. Alexandru Telea, CRC Press, 2014.
- Interactive Data Visualization: Foundations, Techniques, and Applications, 2nd ed. Matthew O. Ward, Georges Grinstein, Daniel Keim. CRC Press, 2015.
- The Visualization Handbook. Charles Hansen and Chris Johnson, eds. Academic Press, 2004.
- Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, 4th ed. Will Schroeder, Ken Martin, and Bill Lorensen. Kitware 2006.
- The Structure of the Information Visualization Design Space. Stuart Card and Jock Mackinlay, Proc. InfoVis 97.
- Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases (extended paper) Chris Stolte, Diane Tang and Pat Hanrahan. IEEE TVCG 8(1):52-65 2002.
- Visualization of Time-Oriented Data. Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, Chris Tominski. Springer 2011.

Ch 3. Why: Task Abstraction



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



Actions: Analyze, Query

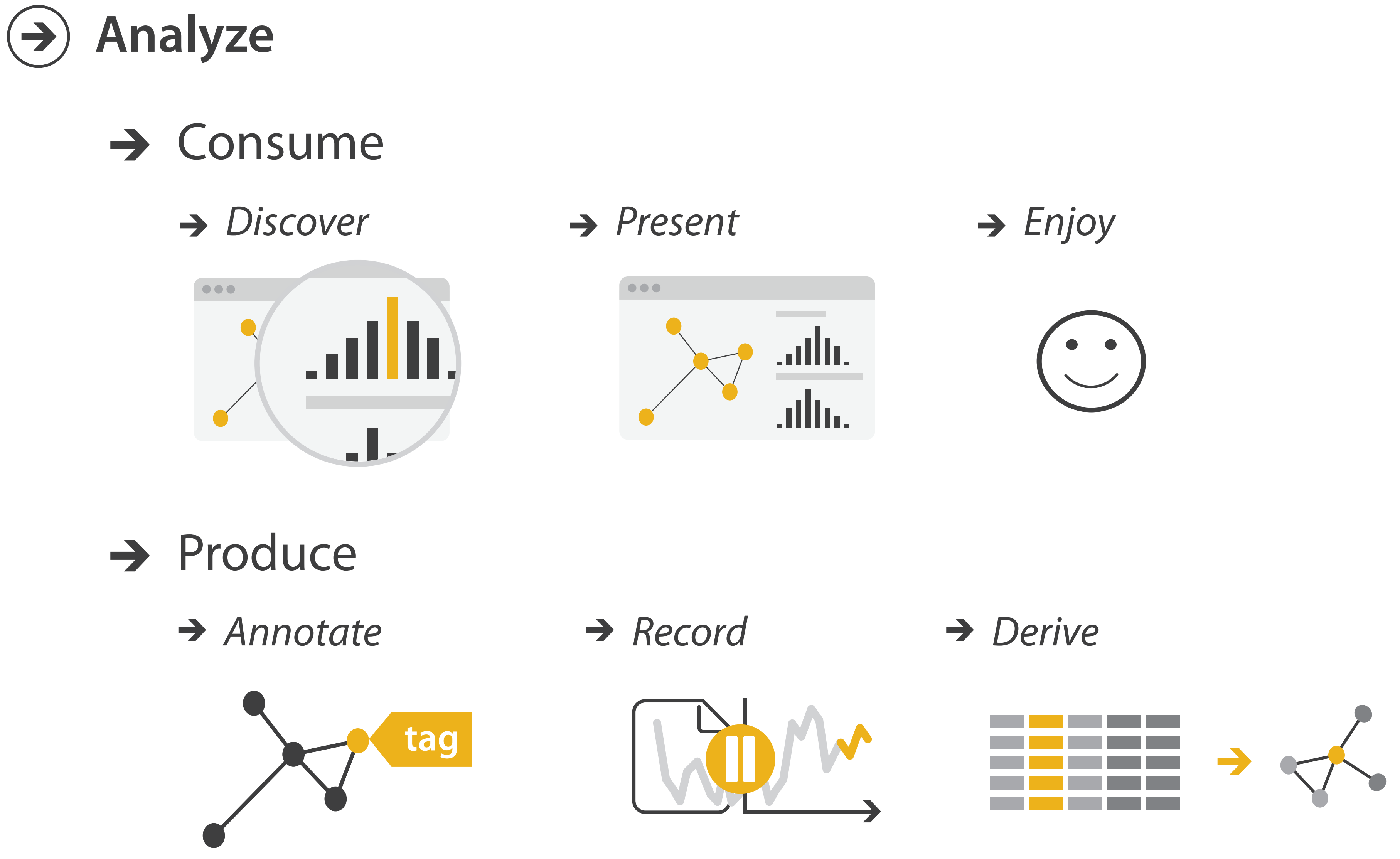
short version: alternate to next 4 slides

- analyze
  - consume
    - discover vs present
      - aka explore vs explain
    - enjoy
      - aka casual, social
  - produce
    - annotate, record, derive
- query
  - how much data matters?
    - one, some, all
- independent choices
  - analyze, query, (search)

Actions: Analyze

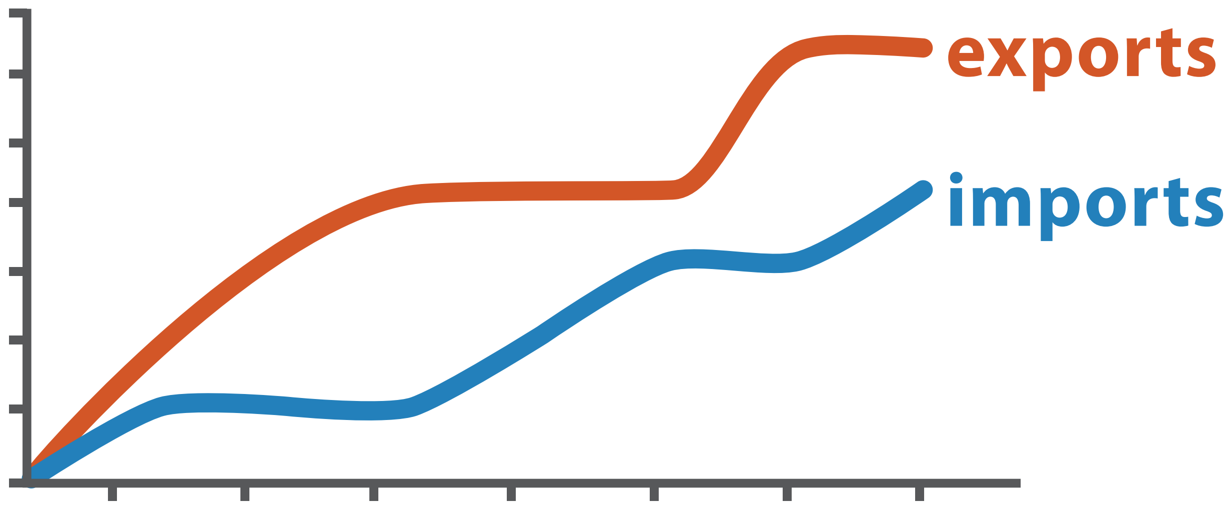
- consume
  - discover vs present
    - classic split
    - aka explore vs explain
  - enjoy
    - newcomer
    - aka casual, social

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

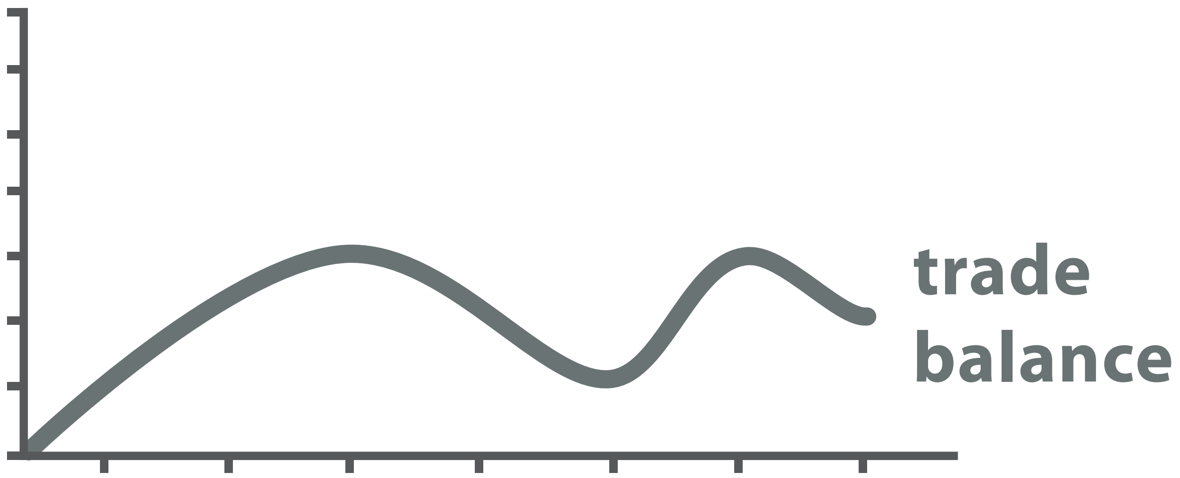


Derive

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



Original Data



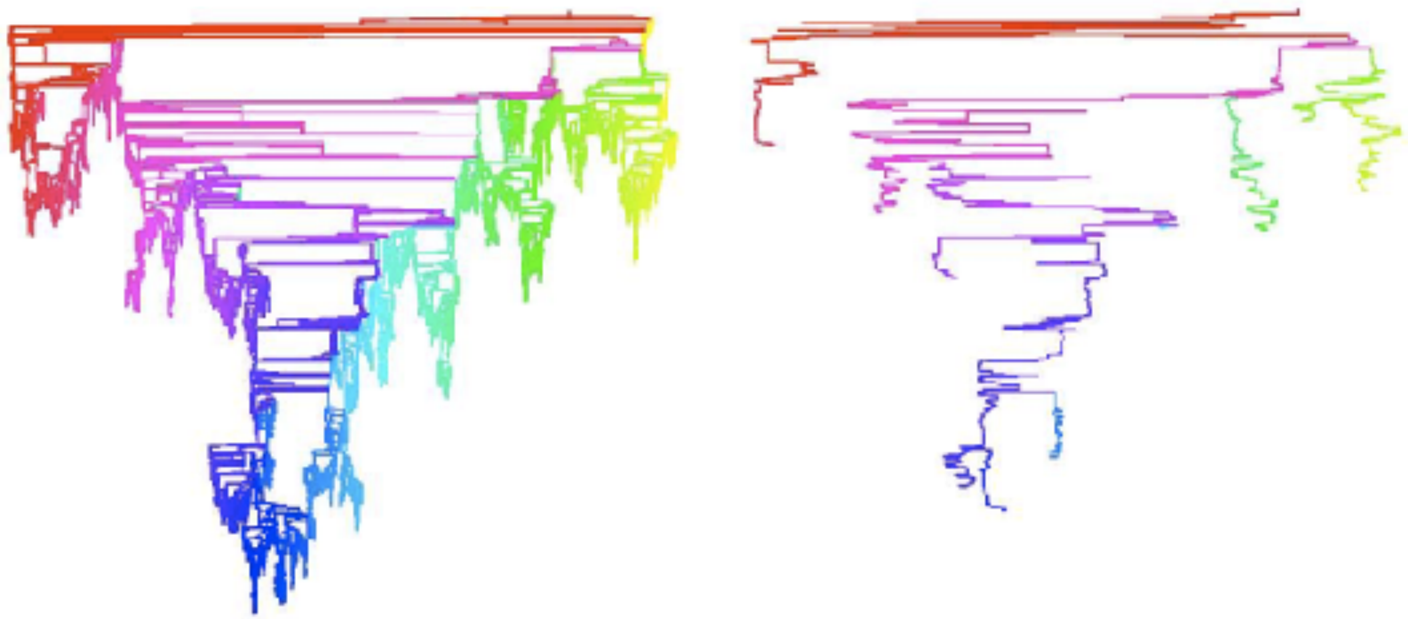
trade balance = exports − imports

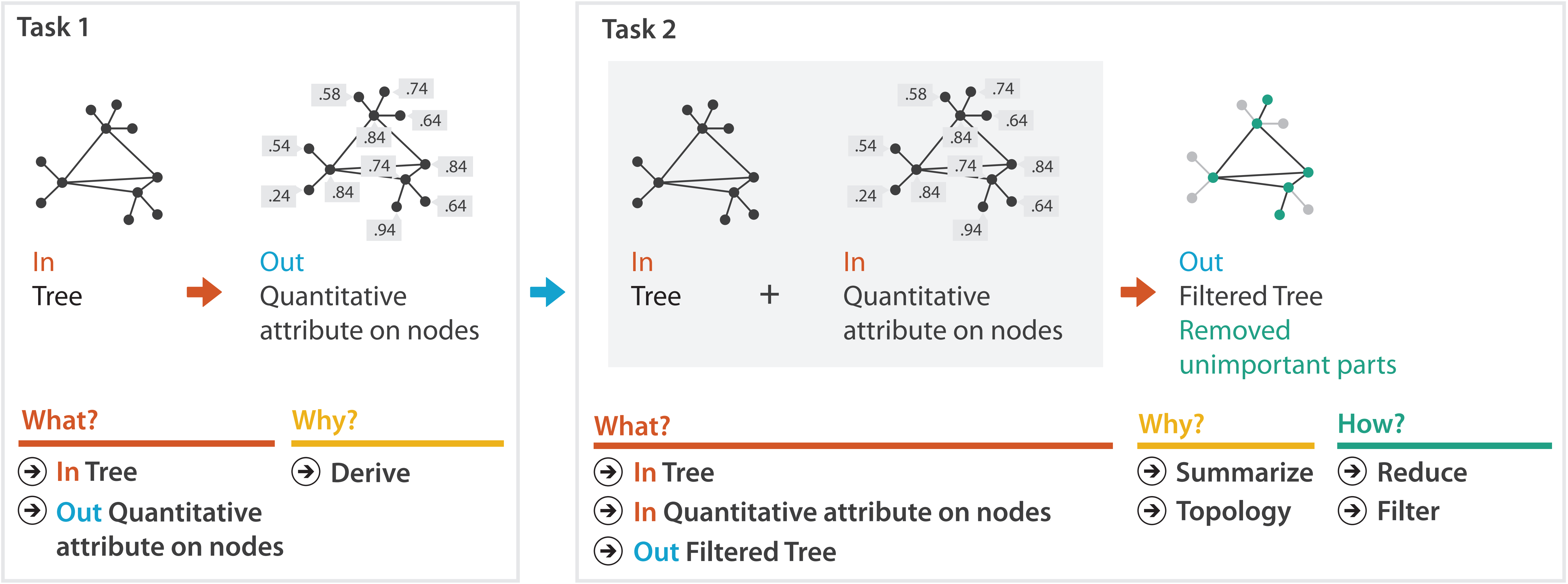
Derived Data

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.]





Actions: Search, query

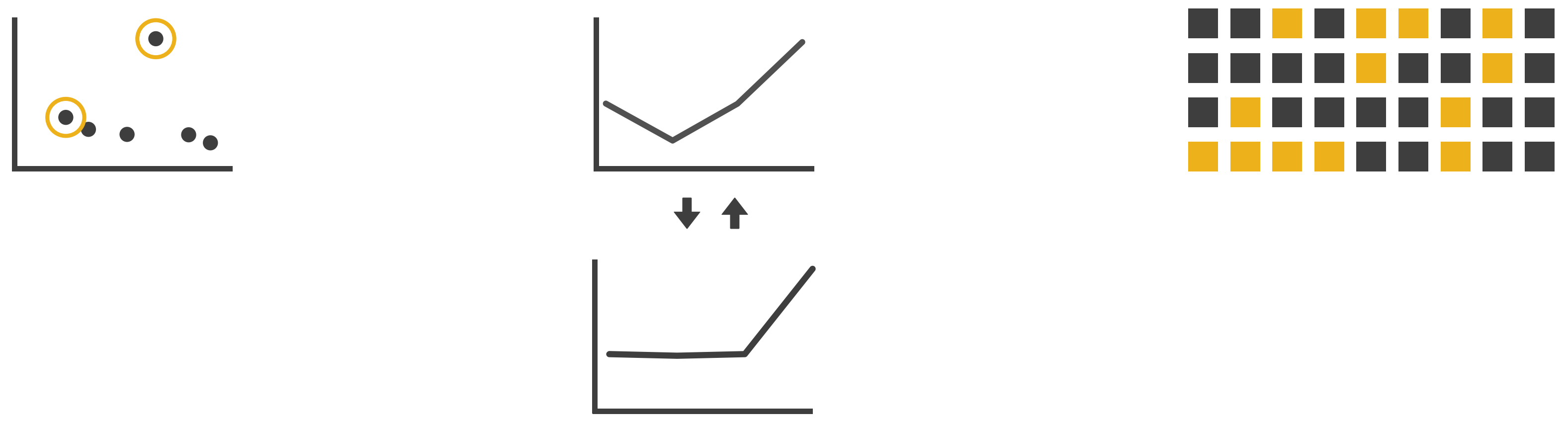
- what does user know?
  - target, location
- how much of the data matters?
  - one, some, all
- independent choices for each of these three levels
  - analyze, search, query
  - mix and match

Search

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

Query

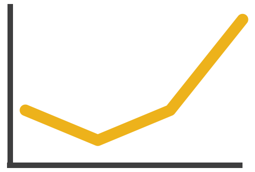
→ Identify → Compare → Summarize



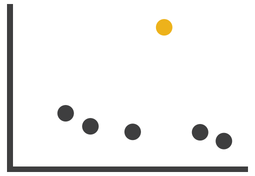
Why: Targets

All Data

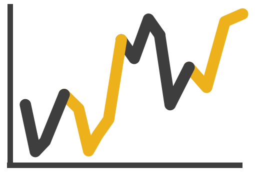
- Trends



- Outliers

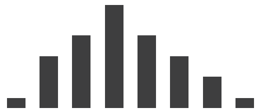


- Features

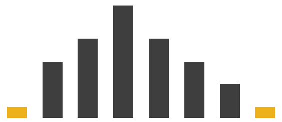


Attributes

- One
  - *Distribution*

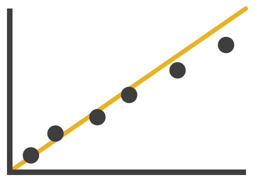


  - *Extremes*

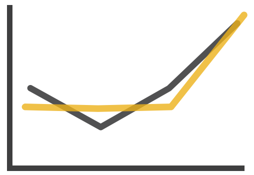


- Many
  - *Dependency*

  - *Correlation*

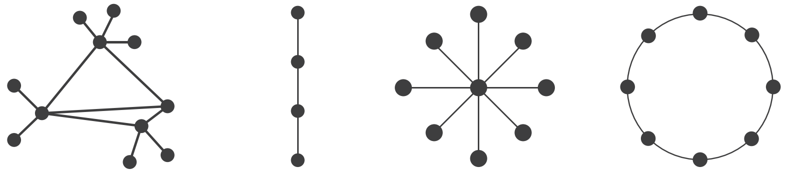


  - *Similarity*

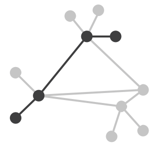


Network Data

- Topology

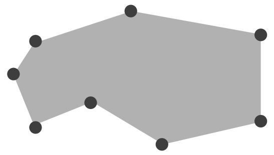


  - *Paths*



Spatial Data

- Shape



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.

Further reading, full Ch 3

- A Multi-Level Typology of Abstract Visualization Tasks.. Matthew Brehmer and Tamara Munzner. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 13) 19:12 (2013), 2376-2385.
- A characterization of the scientific data analysis process. Rebecca R. Springmeyer, Meera M. Blattner, and Nelson M. Max. Proc. Vis 1992, p 235-252.
- Low-Level Components of Analytic Activity in Information Visualization. Robert Amar, James Eagan, and John Stasko. Proc. InfoVis 05, pp. 111-117.
- Task taxonomy for graph visualization. Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. Proc. BELIV 2006.
- Interactive Dynamics for Visual Analysis. Jeffrey Heer and Ben Shneiderman. Communications of the ACM, 55(4), pp. 45-54, 2012.
- What does the user want to see?: what do the data want to be? A. Johannes Pretorius and Jarke J. van Wijk. Information Visualization 8(3):153-166, 2009.
- Chapter 1, Readings in Information Visualization: Using Vision to Think. Stuart Card, Jock Mackinlay, and Ben Shneiderman, Morgan Kaufmann 1999.
- An Operator Interaction Framework for Visualization Systems. Ed H. Chi and John T. Riedl. Proc. InfoVis 1998, p 63-70.
- Nominal, Ordinal, Interval, and Ratio Typologies are Misleading. Paul F. Velleman and Leland Wilkinson. The American Statistician 47(1):65-72, 1993.
- Rethinking Visualization: A High-Level Taxonomy. Melanie Tory and Torsten Möller, Proc. InfoVis 2004, pp. 151-158.
- SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Catherine Plaisant, Jesse Grosjean, and Ben B. Bederson. Proc. InfoVis 2002.
- TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with Guaranteed Visibility Tamara Munzner, Francois Guimbretiere, Serdar Tasiran, Li Zhang, and Yunhong Zhou. SIGGRAPH 2003.
- Feature detection in linked derived spaces. Chris Henze. Proc. Visualization (Vis) 1998, p 87-94.
- Using Strahler numbers for real time visual exploration of huge graphs. David Auber. Intl Conf. Computer Vision and Graphics, 2002, p 56-69.

Ch 4. Analysis: Four Levels for Validation

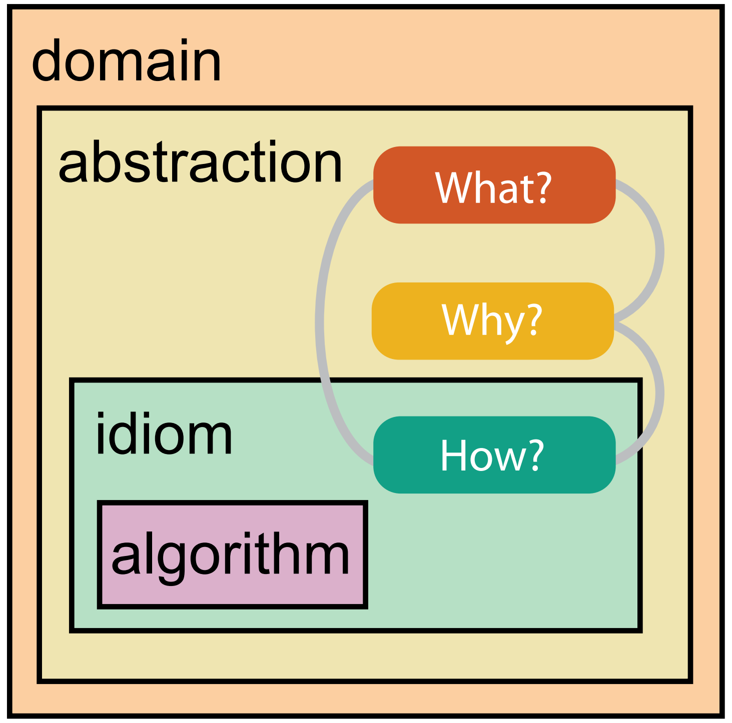
How to evaluate a visualization: So many methods, how to pick?

- Computational benchmarks?
  - quant: system performance, memory
- User study in lab setting?
  - quant: (human) time and error rates, preferences
  - qual: behavior/strategy observations
- Field study of deployed system?
  - quant: usage logs
  - qual: interviews with users, case studies, observations
- Analysis of results?
  - quant: metrics computed on result images
  - qual: consider what structure is visible in result images
- Justification of choices?
  - qual: perceptual principles, best practices

Nested model: Four levels of visualization design

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

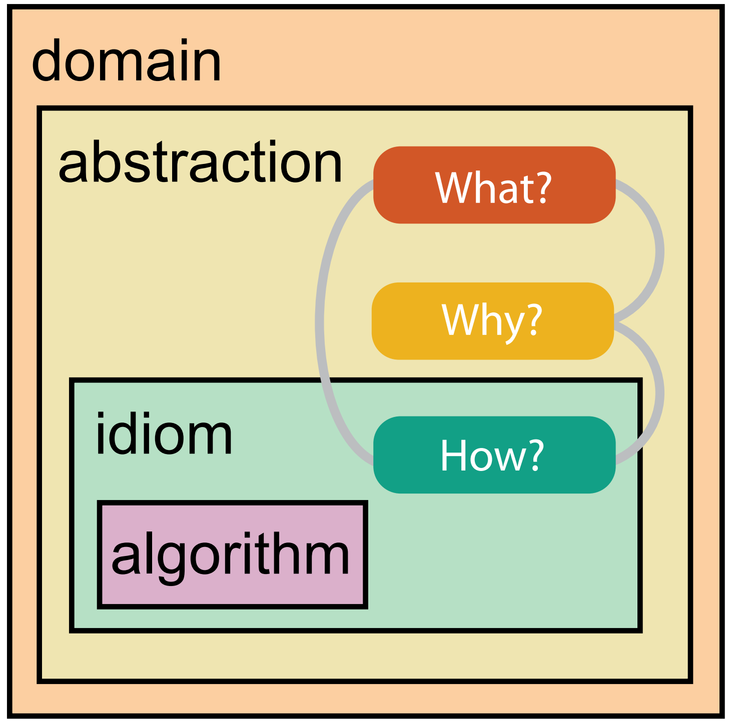
- *domain situation*
  - who are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of visualization
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
- *idiom*
  - **how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate
- *algorithm*
  - efficient computation



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

Nested model: Four levels of visualization design

- *domain situation*
  - who are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of visualization
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
      - often must transform data, guided by task
- *idiom*
  - **how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate
- *algorithm*
  - efficient computation

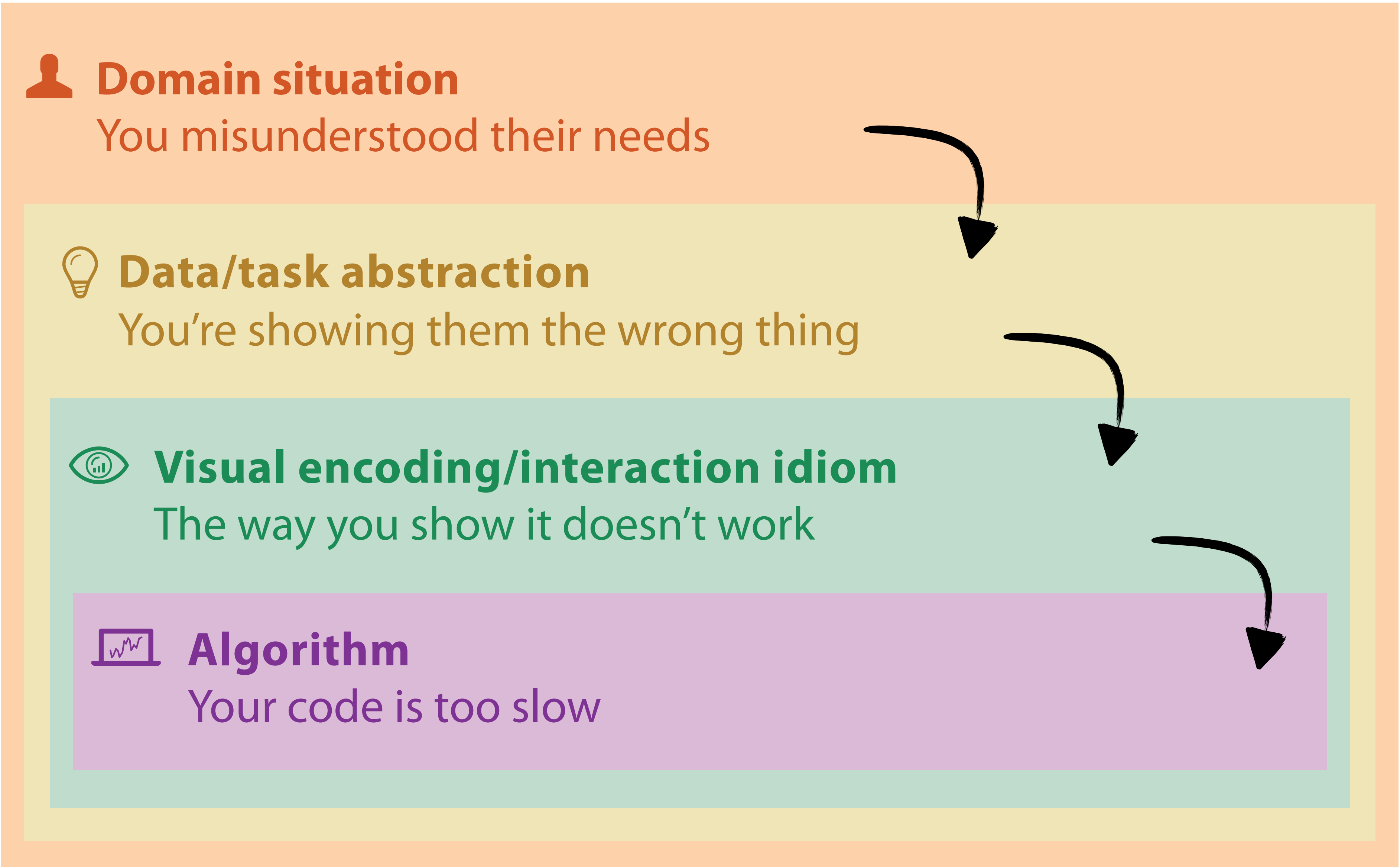


[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009
(Proc. InfoVis 2009).]

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

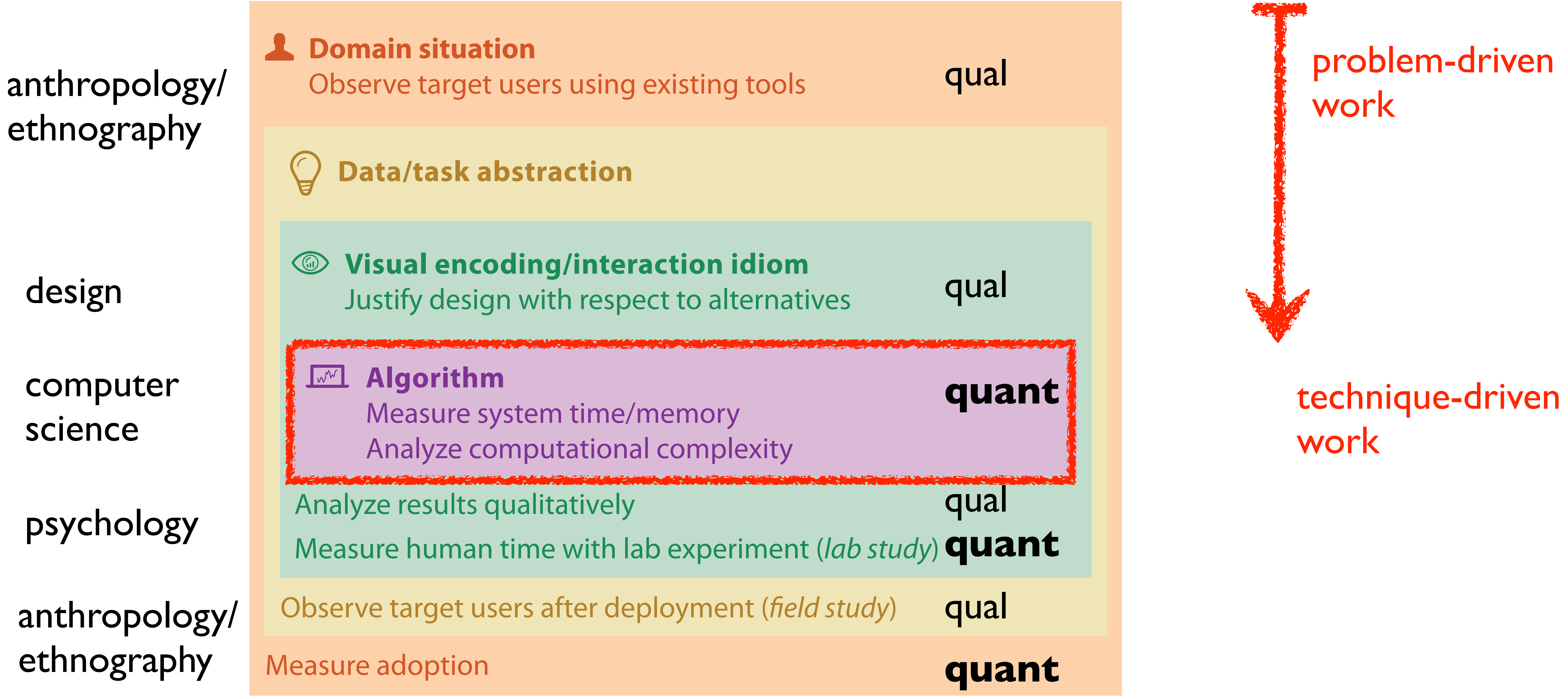
Different threats to validity at each level

- cascading effects downstream



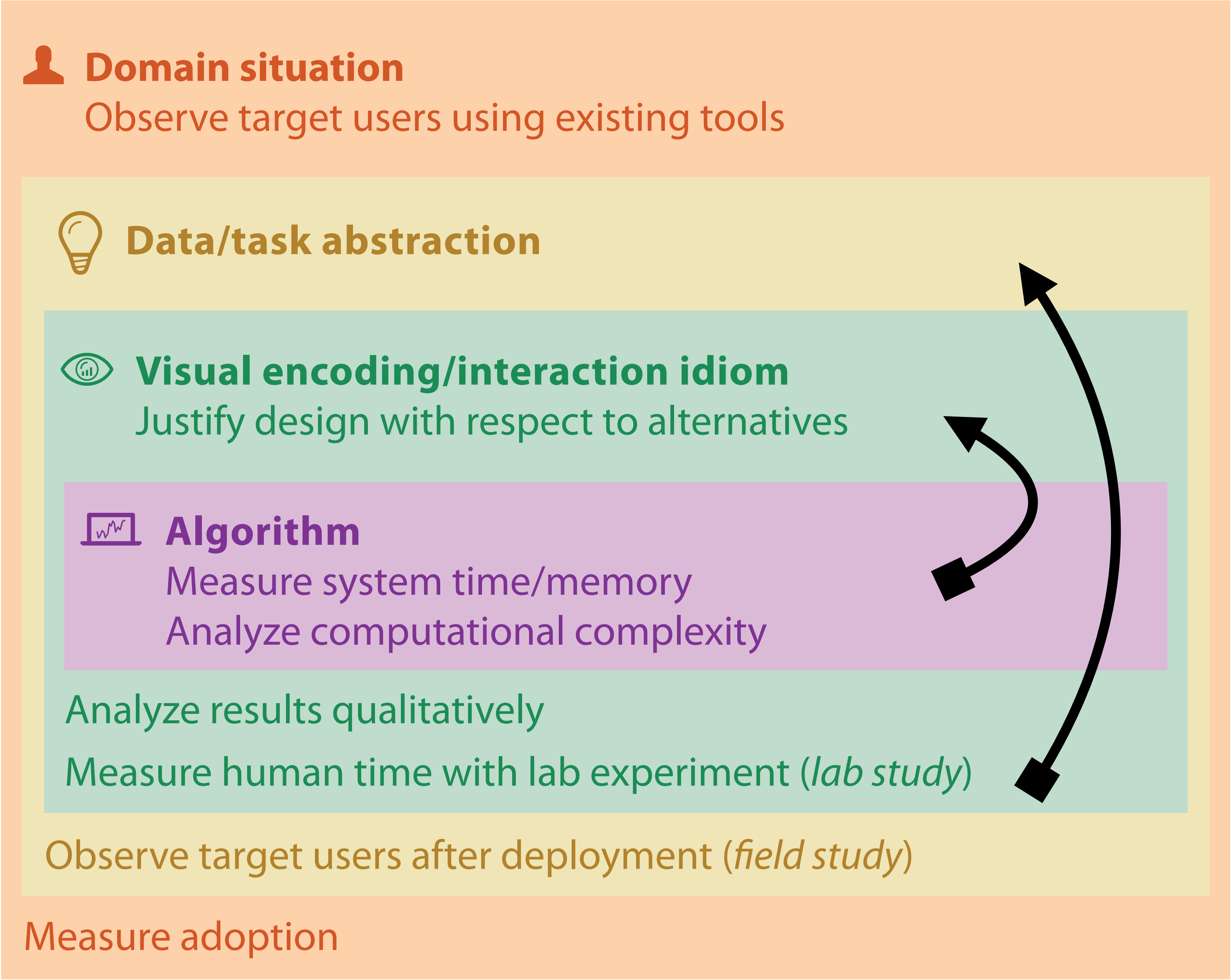
Interdisciplinary: need methods from different fields at each level

- mix of qual and quant approaches (typically)



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

Mismatches: Common problem



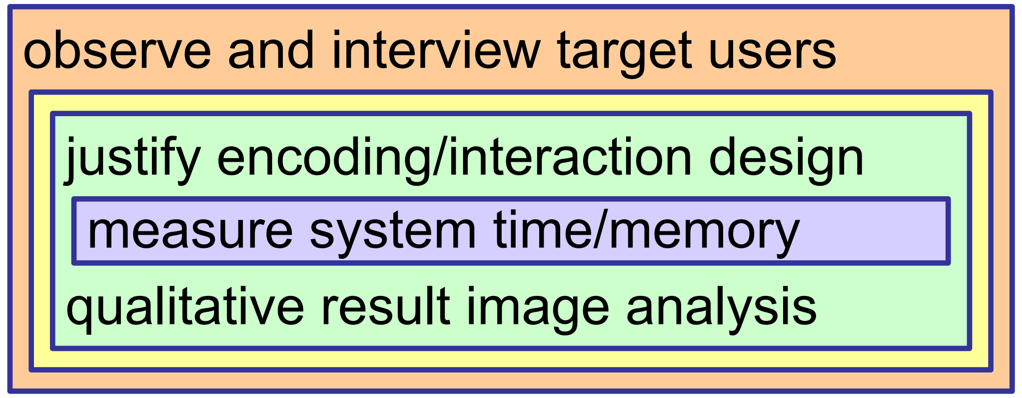
benchmarks can't confirm design

lab studies can't confirm task abstraction

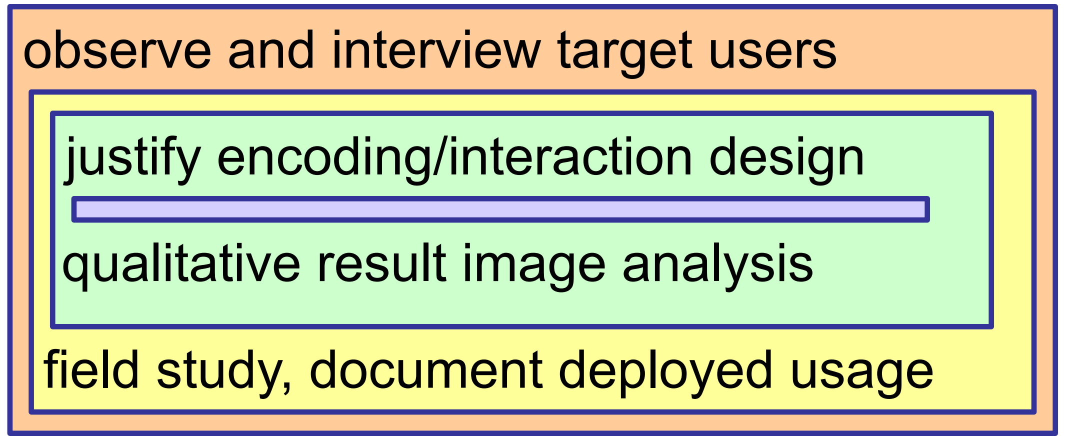
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis examples: Single paper includes only subset of methods

MatrixExplorer. Henry and Fekete. InfoVis 2006.

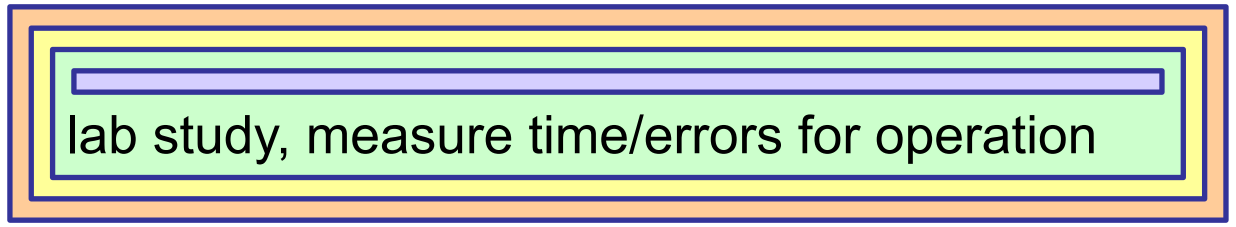


LiveRAC. McLachlan, Munzner, Koutsofios, and North. CHI 2008.

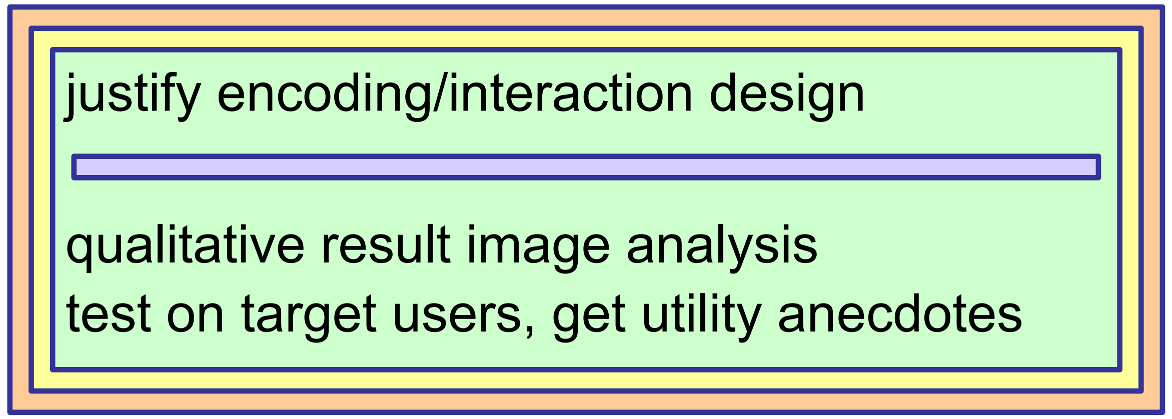


An energy model for visual graph clustering. (LinLog) Noack. Graph Drawing 2003

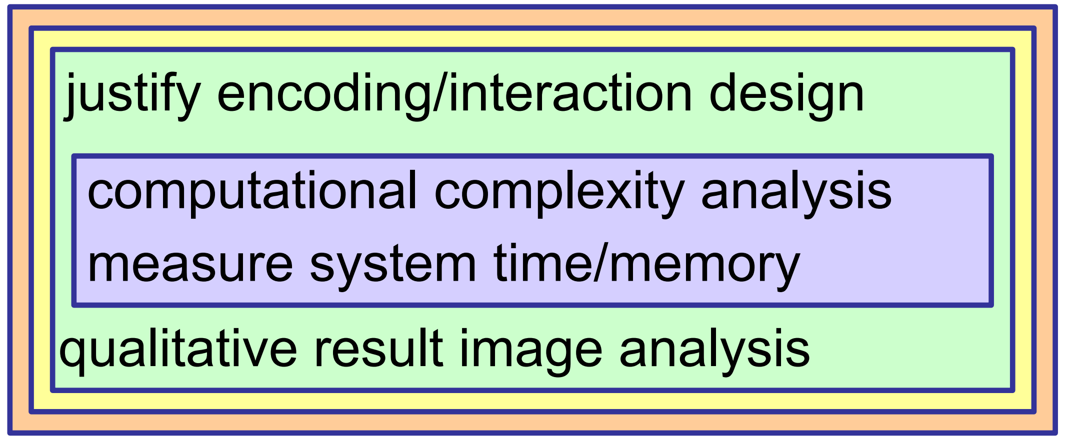
Effectiveness of animation in trend visualization. Robertson et al. InfoVis 2008.



Interactive visualization of genealogical graphs. McGuffin and Balakrishnan. InfoVis 2005.



Flow map layout. Phan et al. InfoVis 2005.



Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - *Chap 4: Analysis: Four Levels for Validation*
- Storks Deliver Babies (p= 0.008). Robert Matthews. Teaching Statistics 22(2):36-38, 2000.
- The Earth is spherical (p < 0.05): alternative methods of statistical inference. Kim J. Vicente and Gerard L. Torenvliet. Theoretical Issues in Ergonomics Science, 1(3):248-271, 2000.
- The Prospects for Psychological Science in Human-Computer Interaction. Allen Newell and Stuart K. Card. Journal Human-Computer Interaction 1(3):209-242, 1985.
- How to do good research, get it published in SIGKDD and get it cited!, Eamonn Keogh, SIGKDD Tutorial 2009.
- False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Joseph P. Simmons, Leif D. Nelson and Uri Simonsohn. Psychological Science 22(11):1359-1366, 2011.
- Externalisation - how writing changes thinking.. Alan Dix. Interfaces, Autumn 2008.

Guerilla/Discount Usability

- grab a few people and watch them use your interface
  - even 3-5 gives substantial coverage of major usability problems
  - agile/lean qualitative, vs formal quantitative user studies
    - goal is not statistical significance!
- think-aloud protocol
  - contextual inquiry (conversations back and forth) vs fly on the wall (you're silent)

Further reading, usability

- 7 Step Guide to Guerrilla Usability Testing, Markus Piper
  - https://userbrain.net/blog/7-step-guide-guerrilla-usability-testing-diy-usability-testing-method
- The Art of Guerrilla Usability Testing, David Peter Simon
  - http://www.uxbooth.com/articles/the-art-of-guerrilla-usability-testing/
- Discount Usability: 20 Years, Jakob Nielsen
  - https://www.nngroup.com/articles/discount-usability-20-years/
- Interaction Design: Beyond Human-Computer Interaction
  - Preece, Sharp, Rogers. Wiley, 4th edition, 2015.
- About Face: The Essentials of Interaction Design
  - Cooper, Reimann, Cronin, Noessel. Wiley, 4th edition, 2014.
- Task-Centered User Interface Design. Lewis & Rieman, 1994
  - http://hcibib.org/tcuid/
- Designing with the Mind in Mind. Jeff Johnson. Morgan Kaufmann, 2nd, 2014.