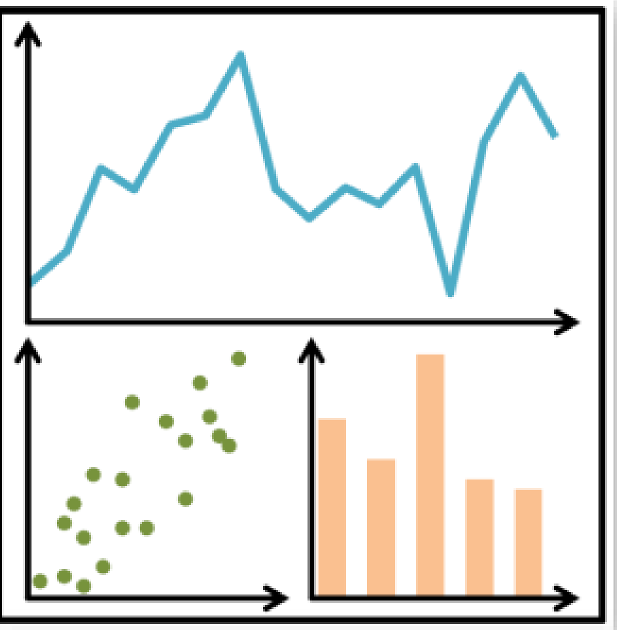
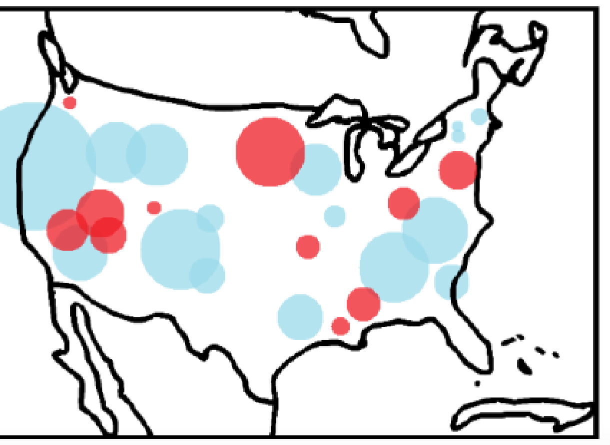
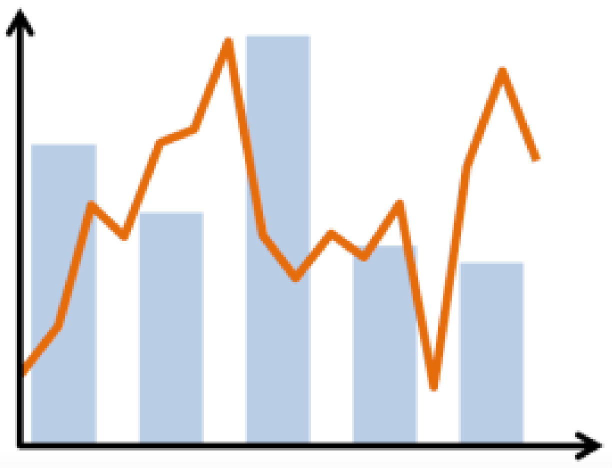
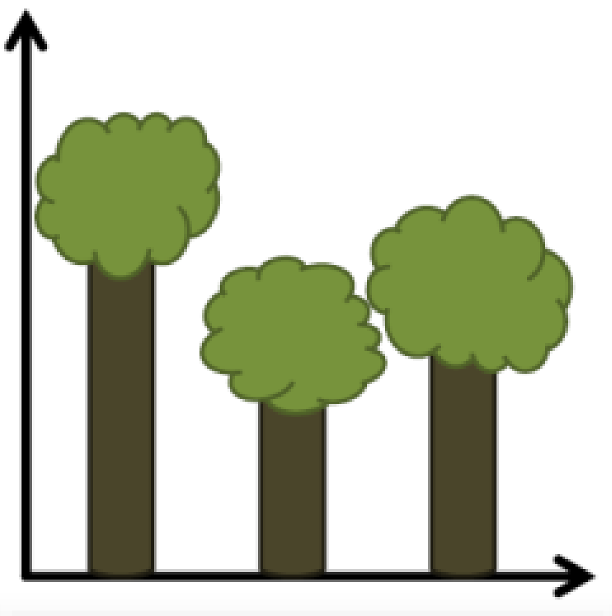
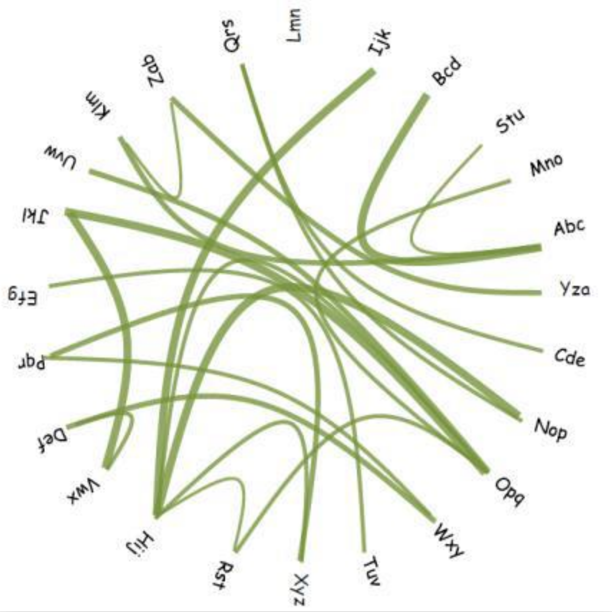






Validation & Evaluation

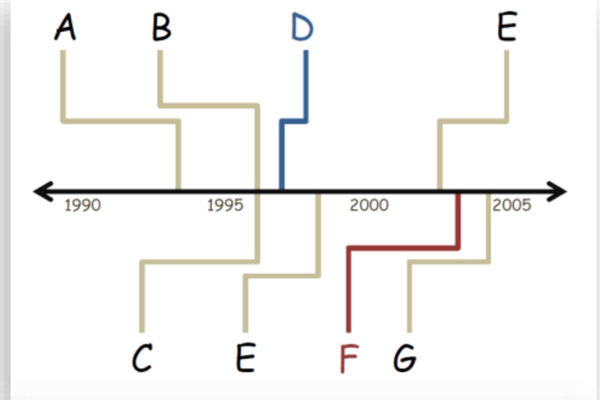
CS 7250

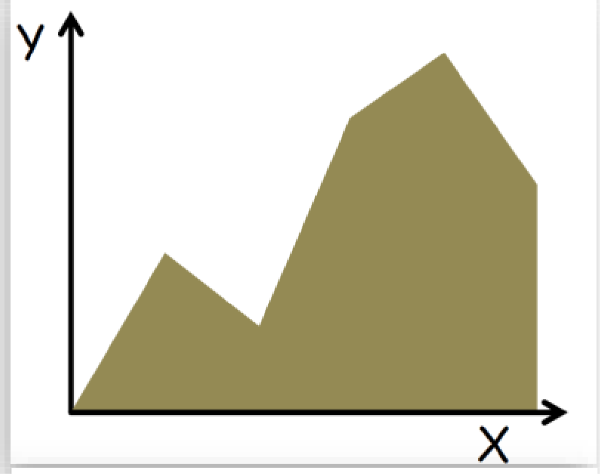
SPRING 2020

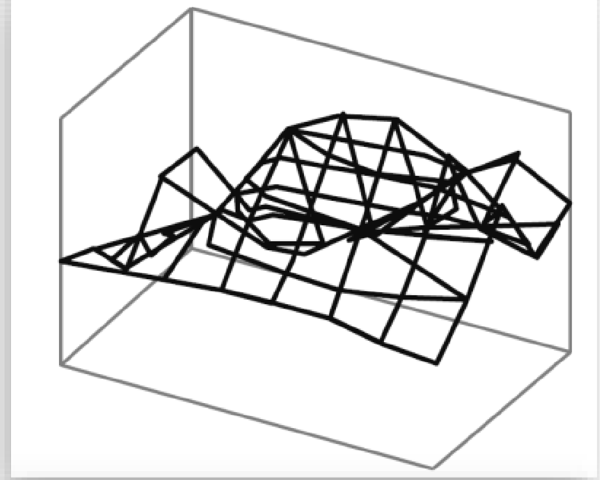
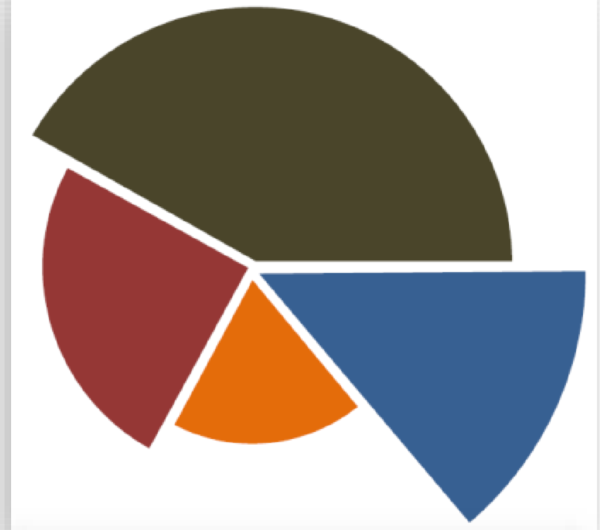
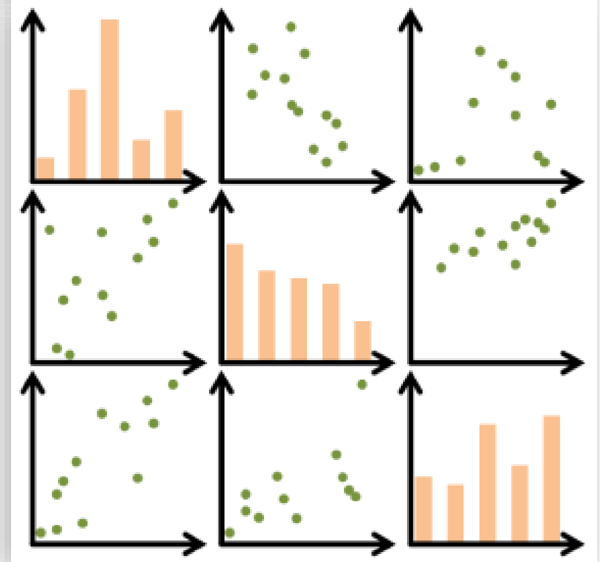
Prof. Cody Dunne

NORTHEASTERN UNIVERSITY

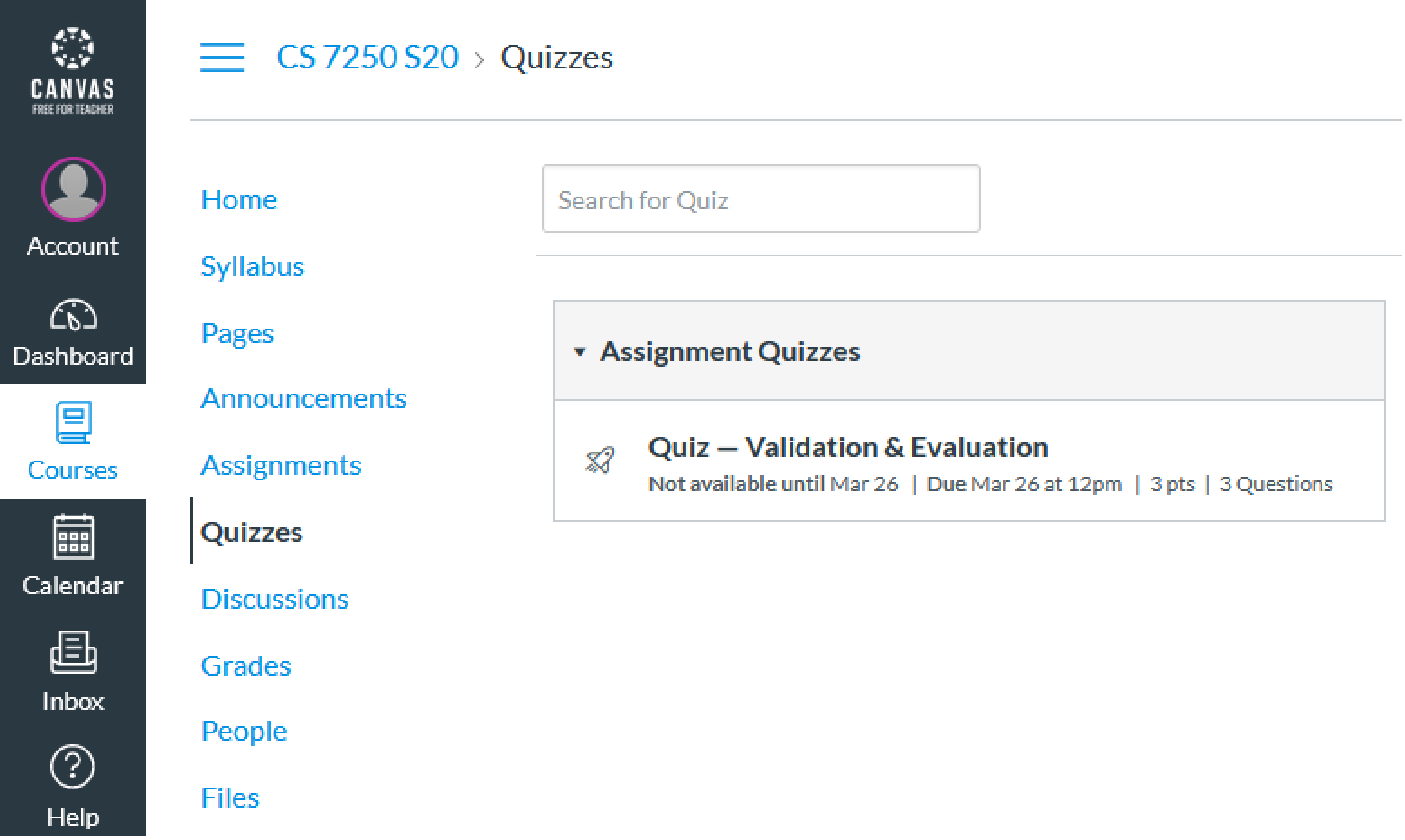
Slides and inspiration from Michelle Borkin, Krzysztof Gajos, Hanspeter Pfister, Miriah Meyer, Jonathan Schwabish, and David Sprague





Burning Questions?



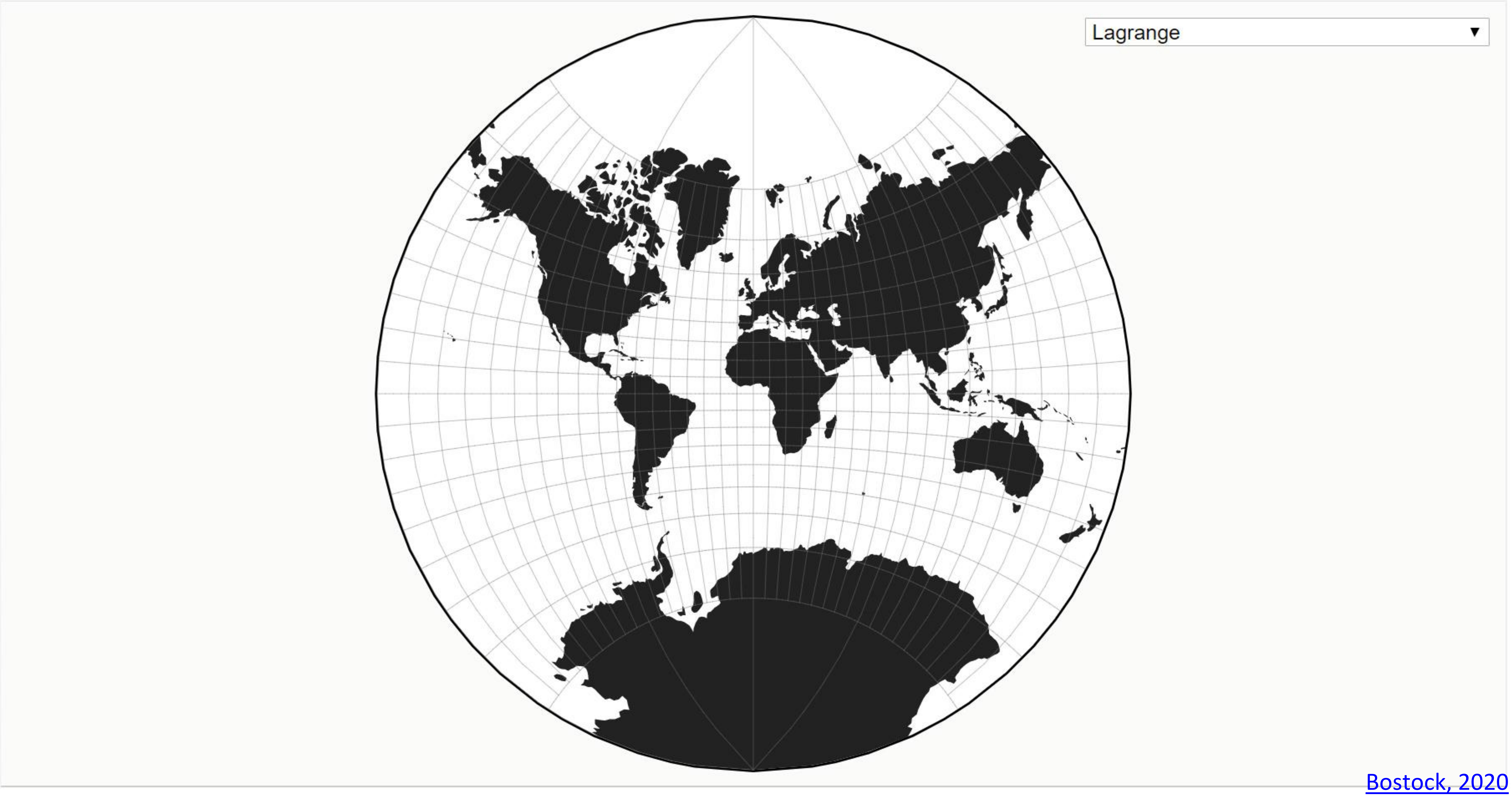
READING QUIZ

Quiz — Validation & Evaluation

~6 min

Previously, on CS 7250…

Projection Transitions

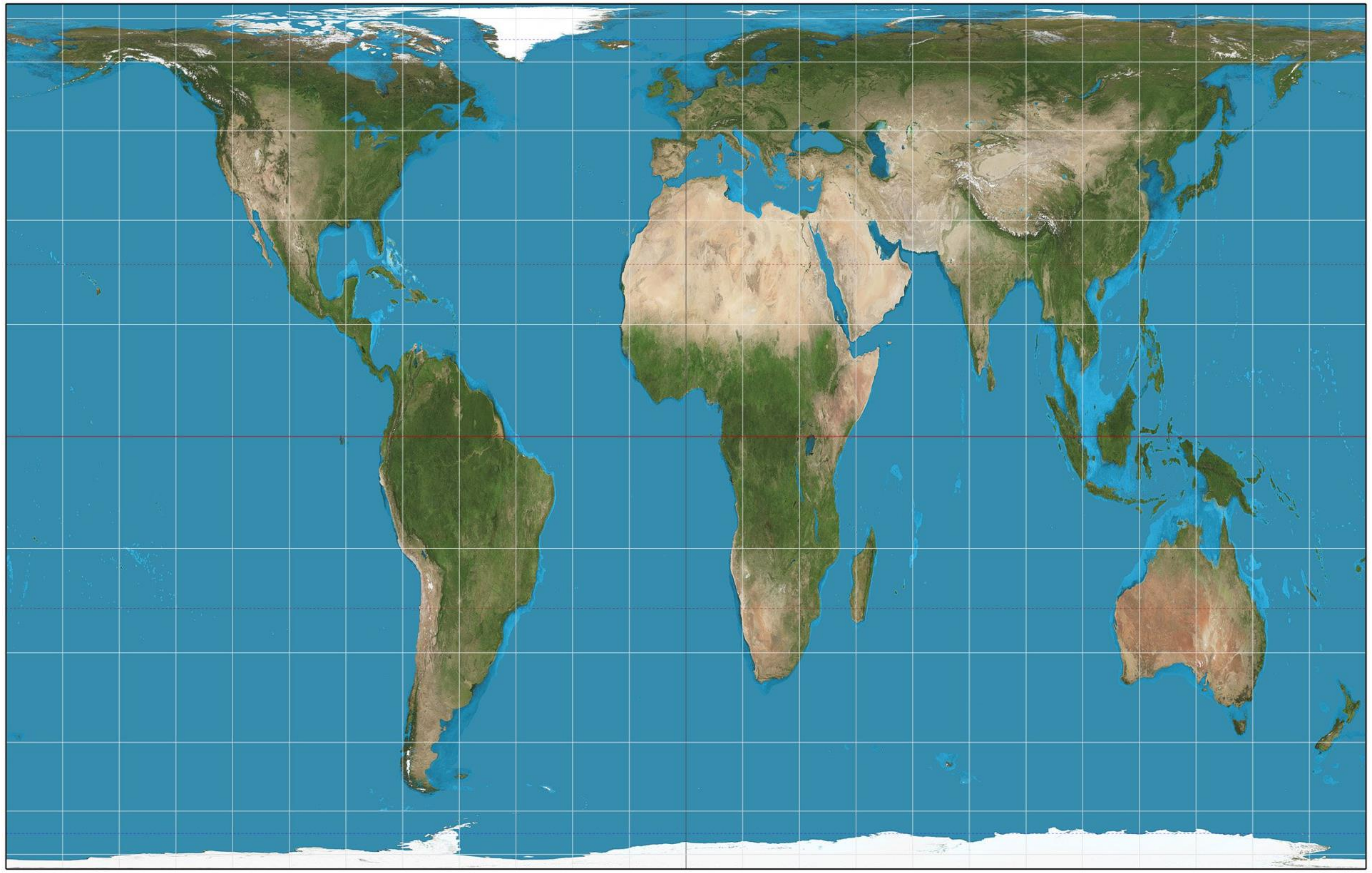


Bostock, 2020

Mercator Projection

Great for ocean navigation, but dramatically exaggerates poles.

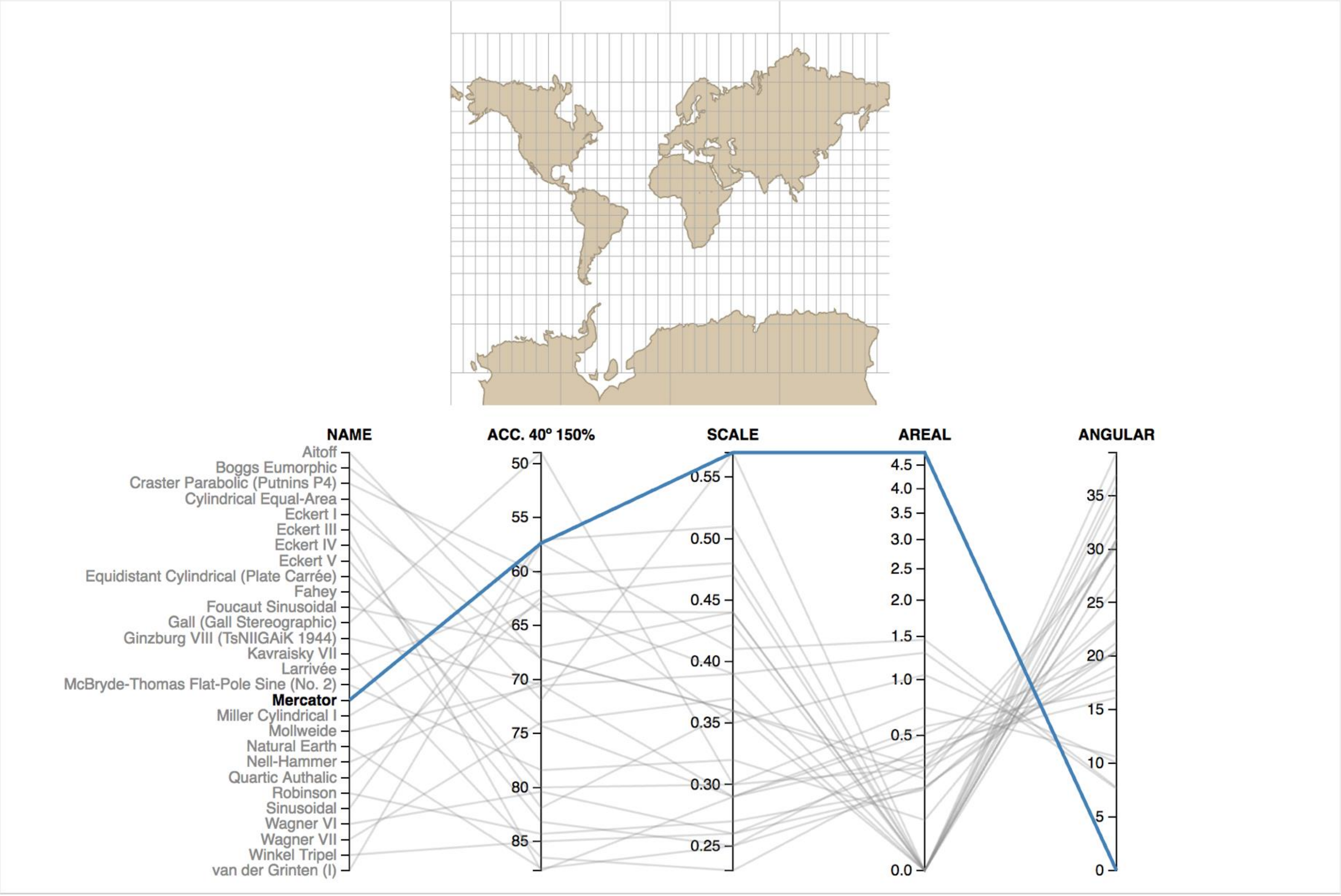
Gall-Peters Projection



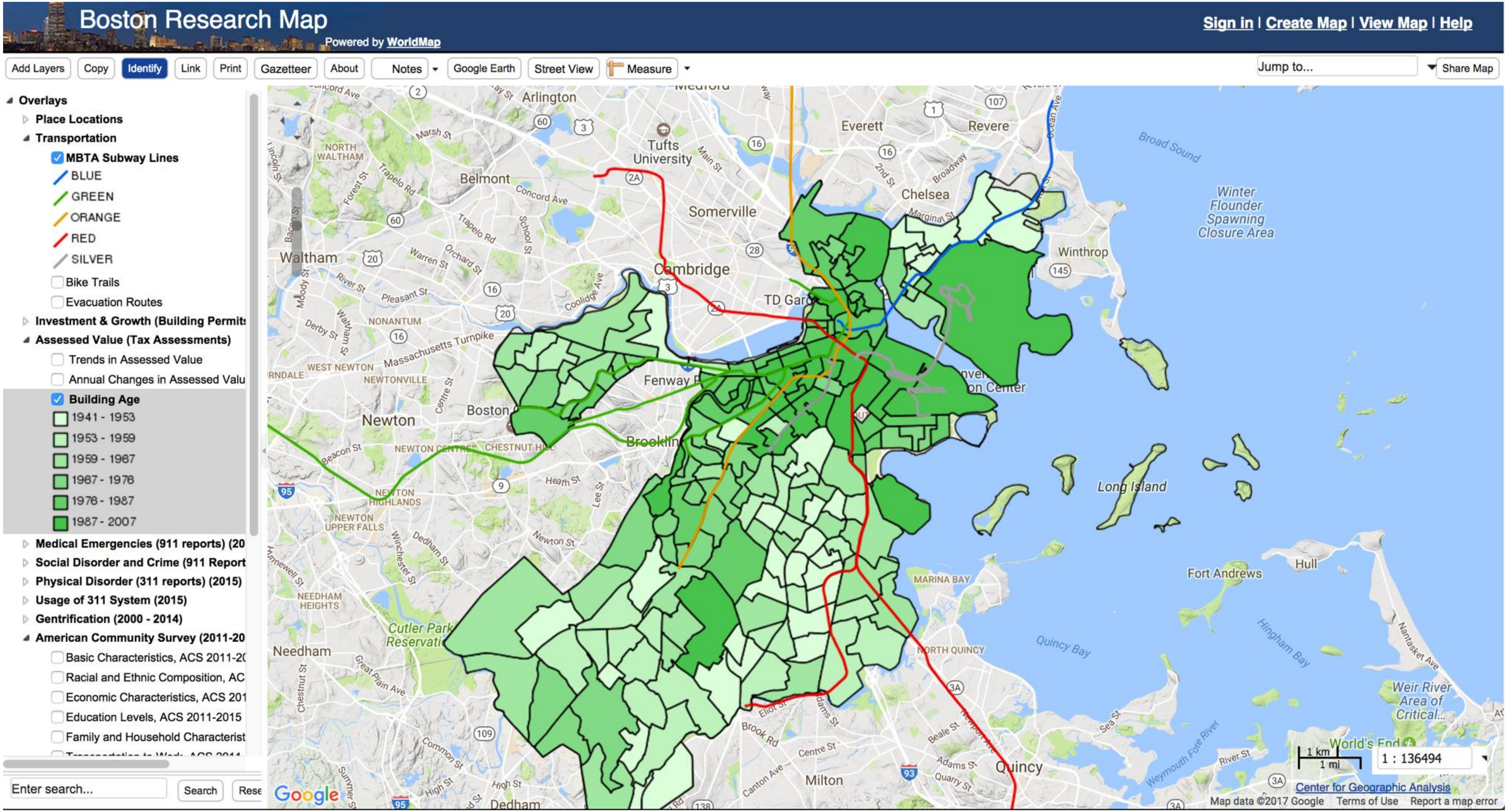
More accurate land areas. (Officially endorsed by the UN.)

Bec Crew, 2017

Comparing Map Projections



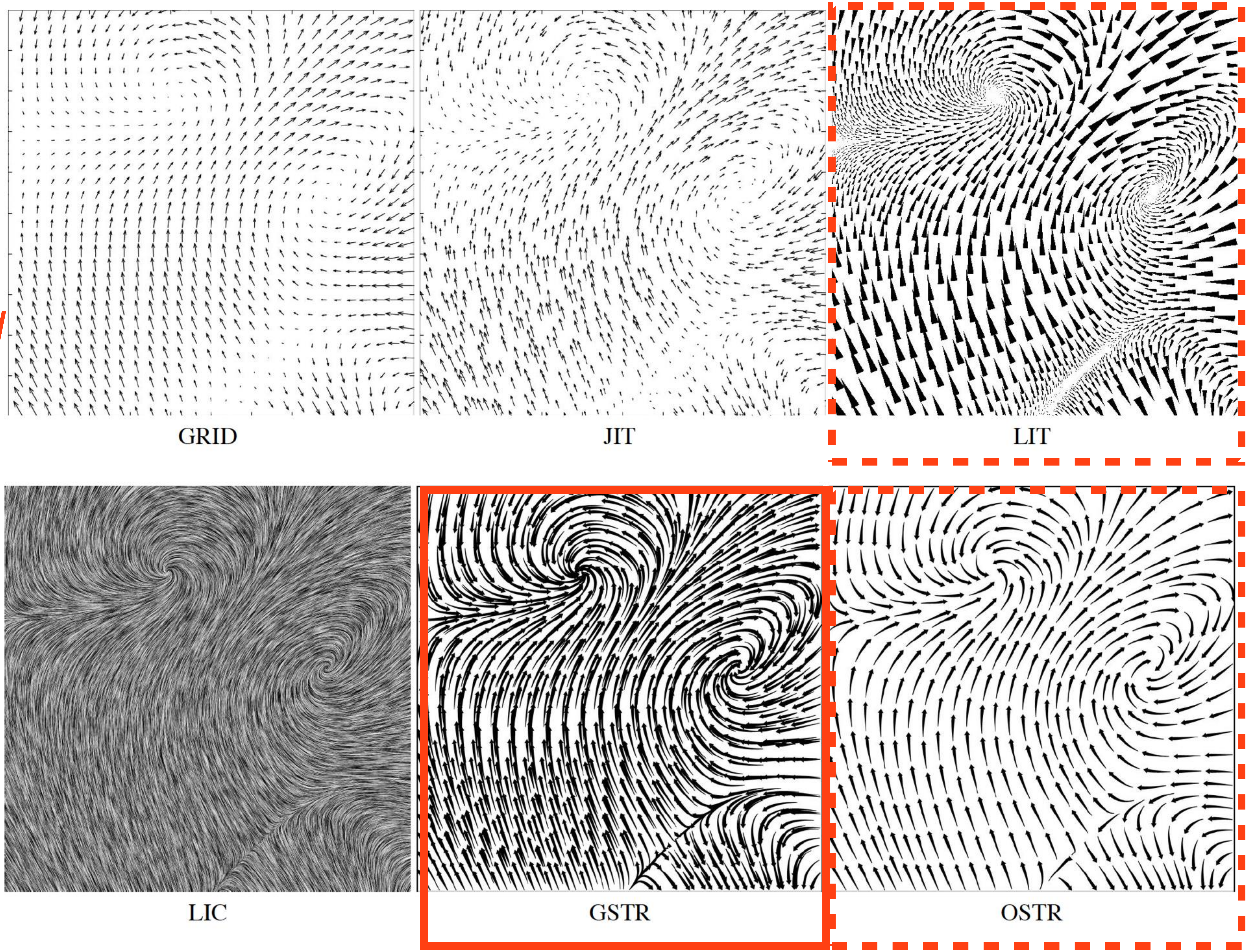
[Kai/syntagmatic, 2017]



Bari/Worldmap, 2011

Vector Field Encoding Examples:

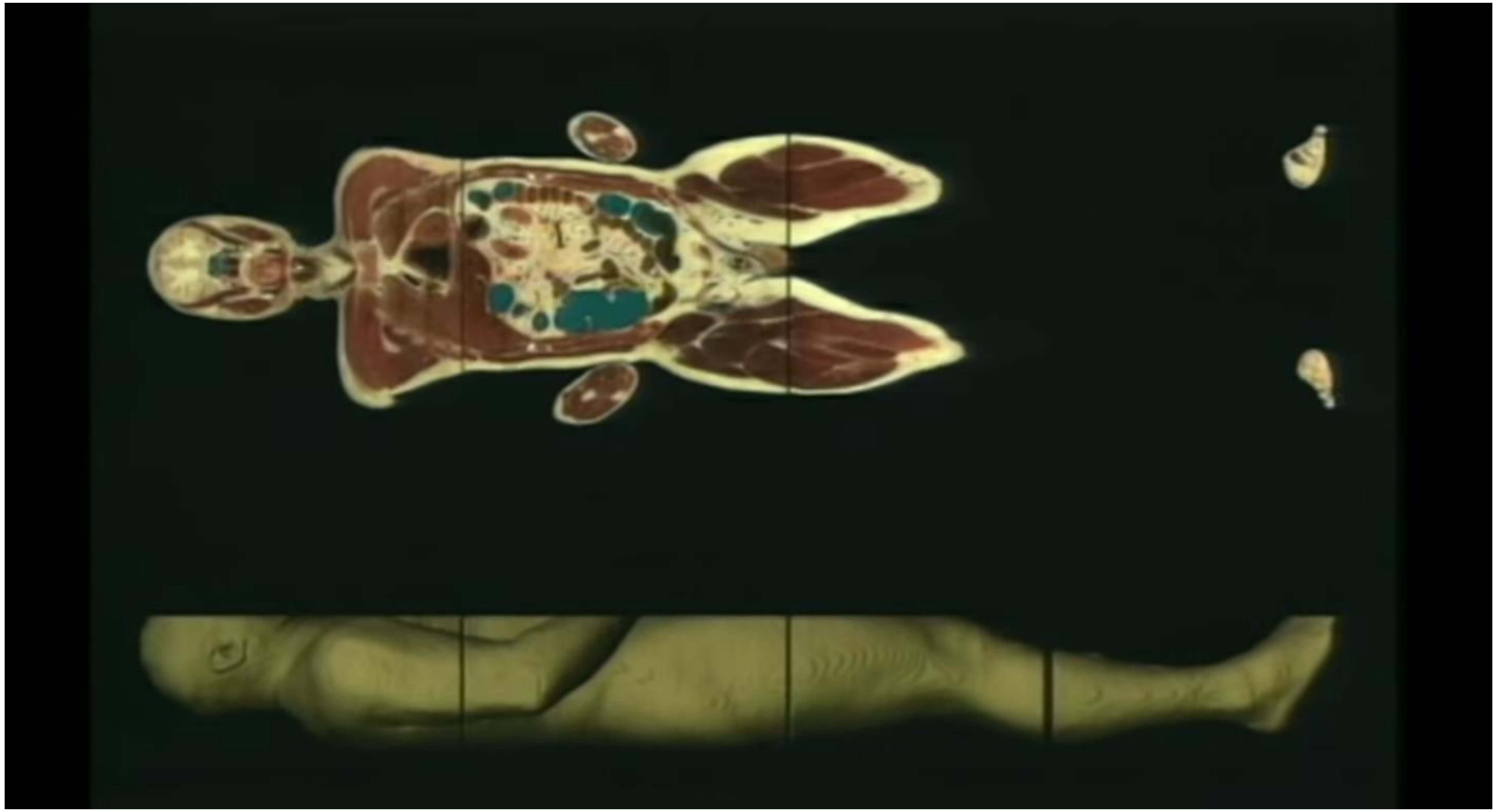
Most accurate and efficient for certain spatial tasks



Laidlaw, et al. 2001

Isosurfaces & Volume Rendering

[Visible Human Project]

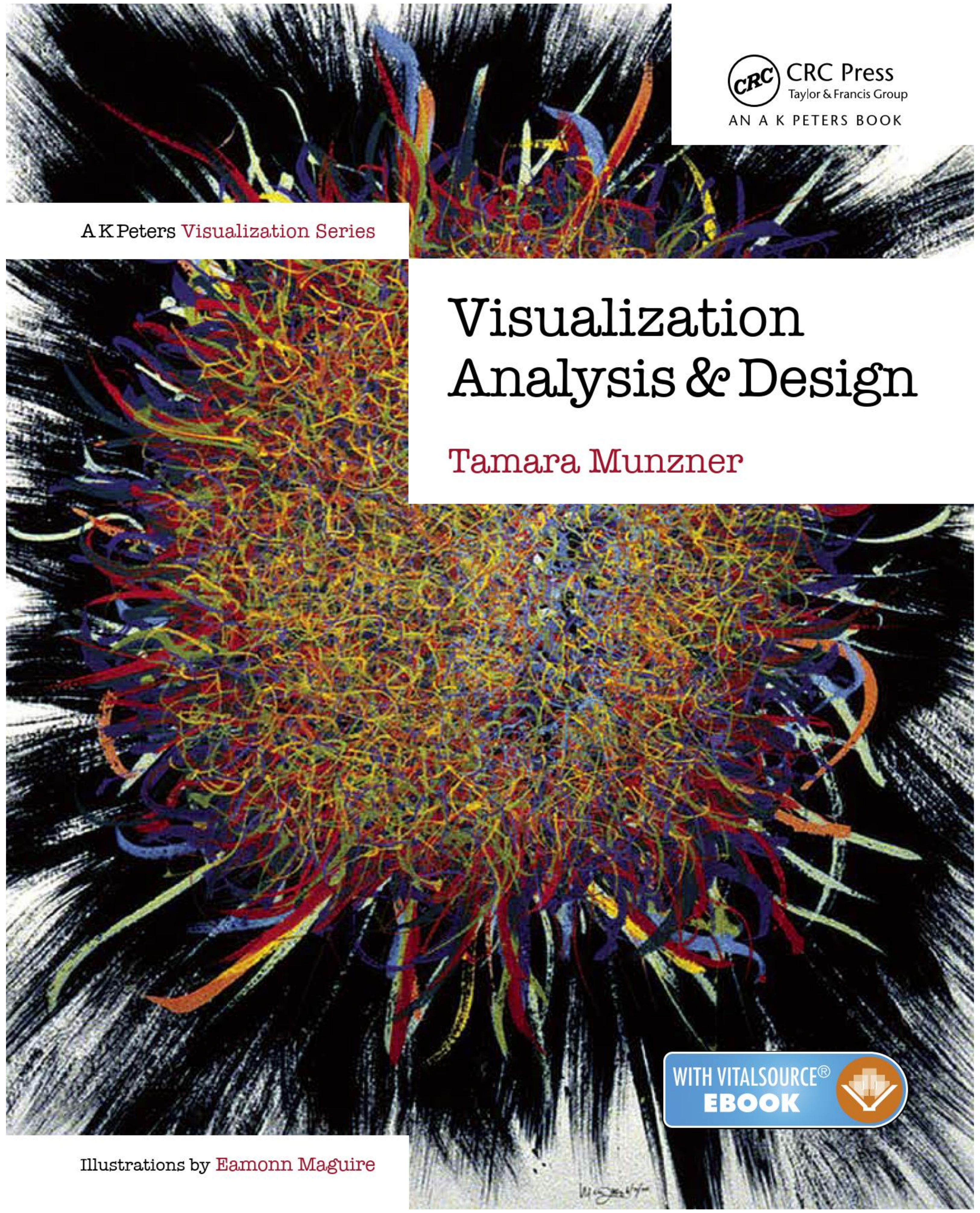


[https://www.youtube.com/watch?v=7GPB1sjEhIQ]

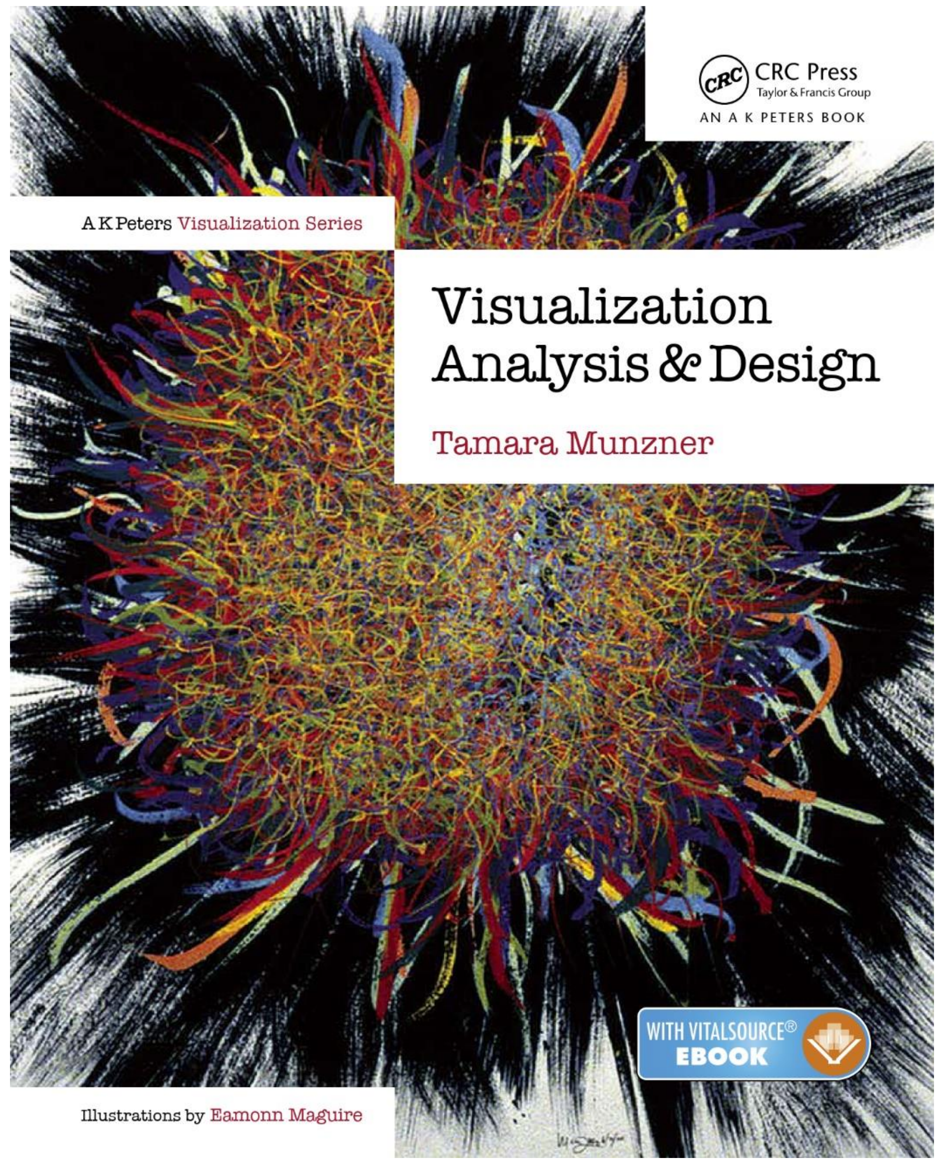
Now, on CS 7250…

The Nested Model for Visualization Validation

TEXTBOOK



Additional "recommended" books as resources in syllabus



"Nested Model"

Domain situation
Observe target users using existing tools

Example

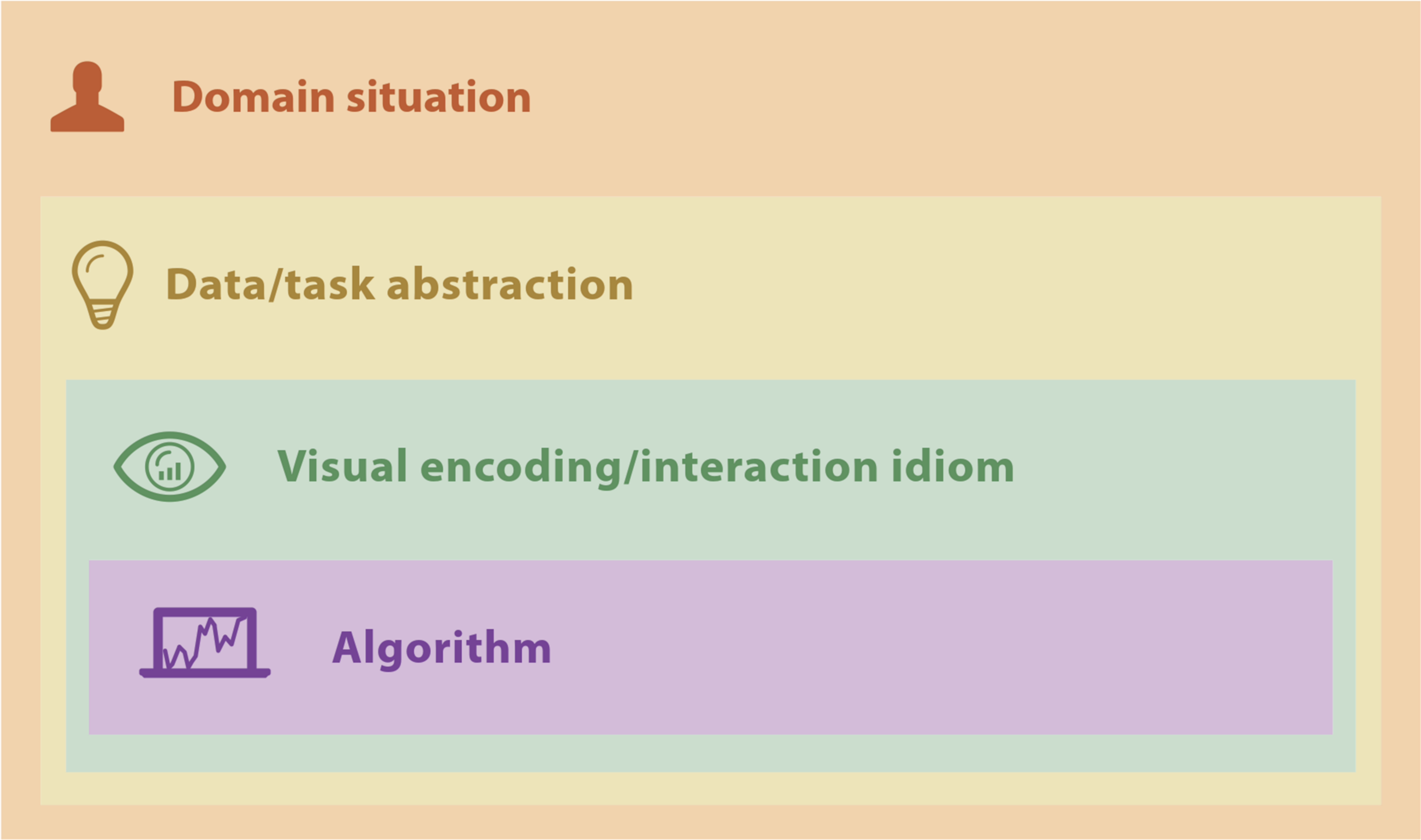
FAA (aviation)

What is the busiest time of day at Logan Airport?

Map vs. Scatter Plot vs. Bar


[Tamara Munzner]

Nested Model



Nested Model

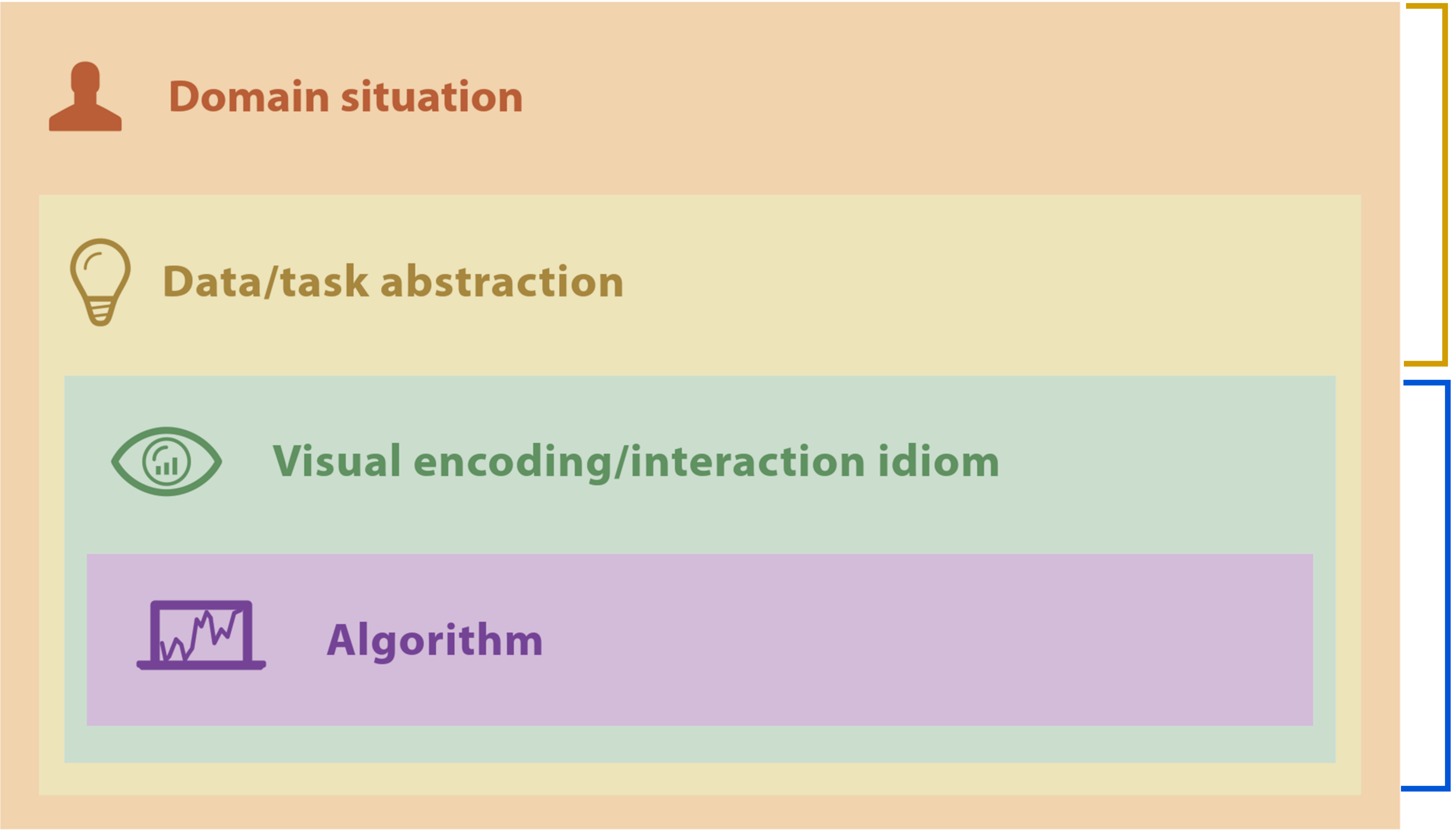
Human-centered design

Designer understands user

Abstract domain tasks

Visualization design

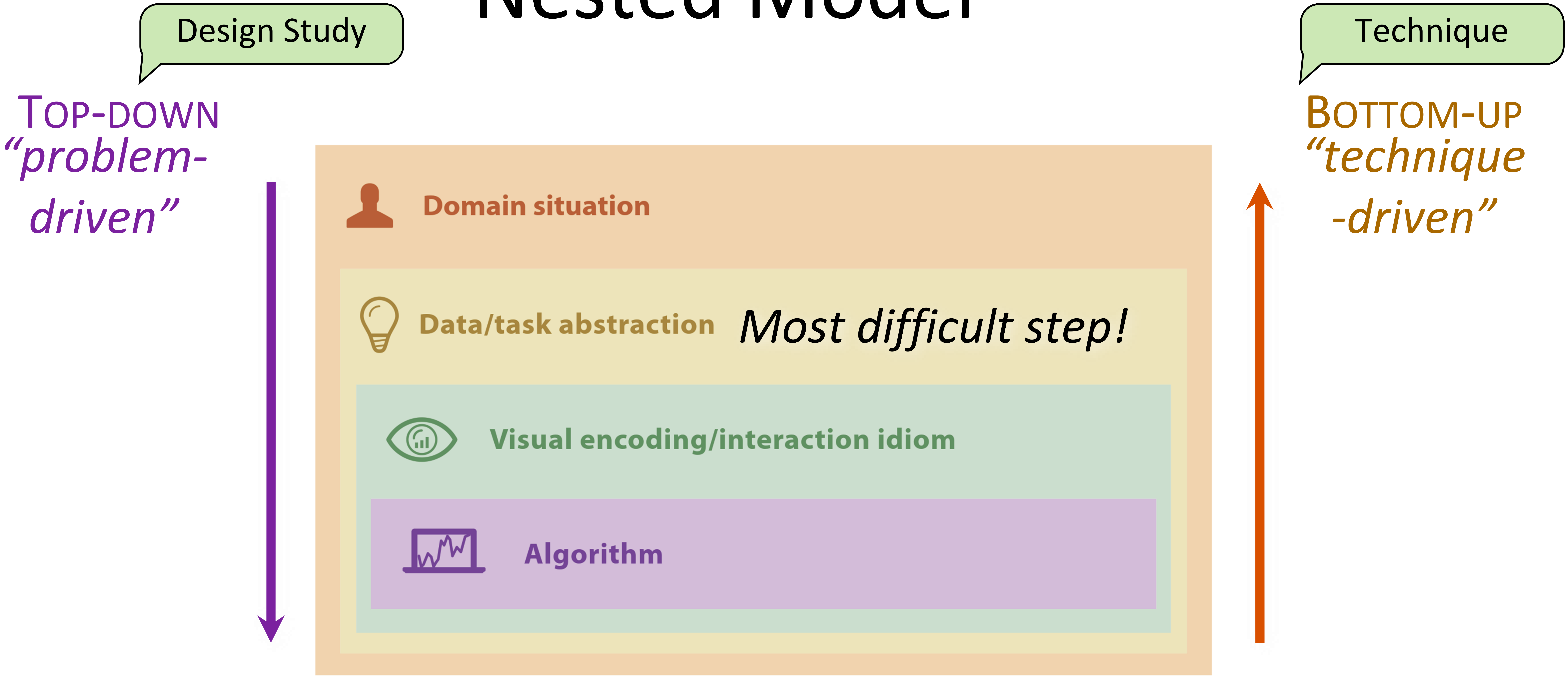
Implementation



Identified

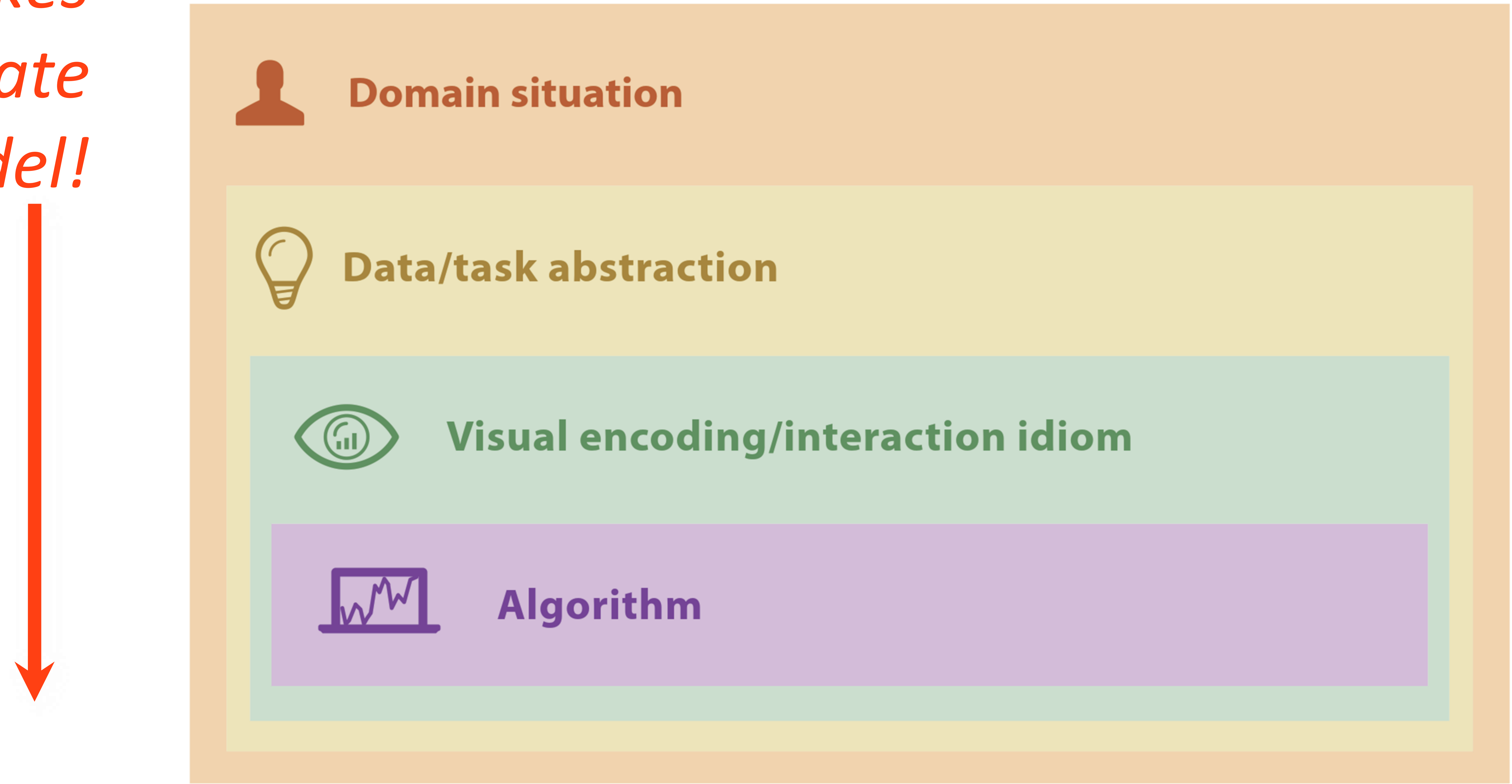
Designed

Nested Model

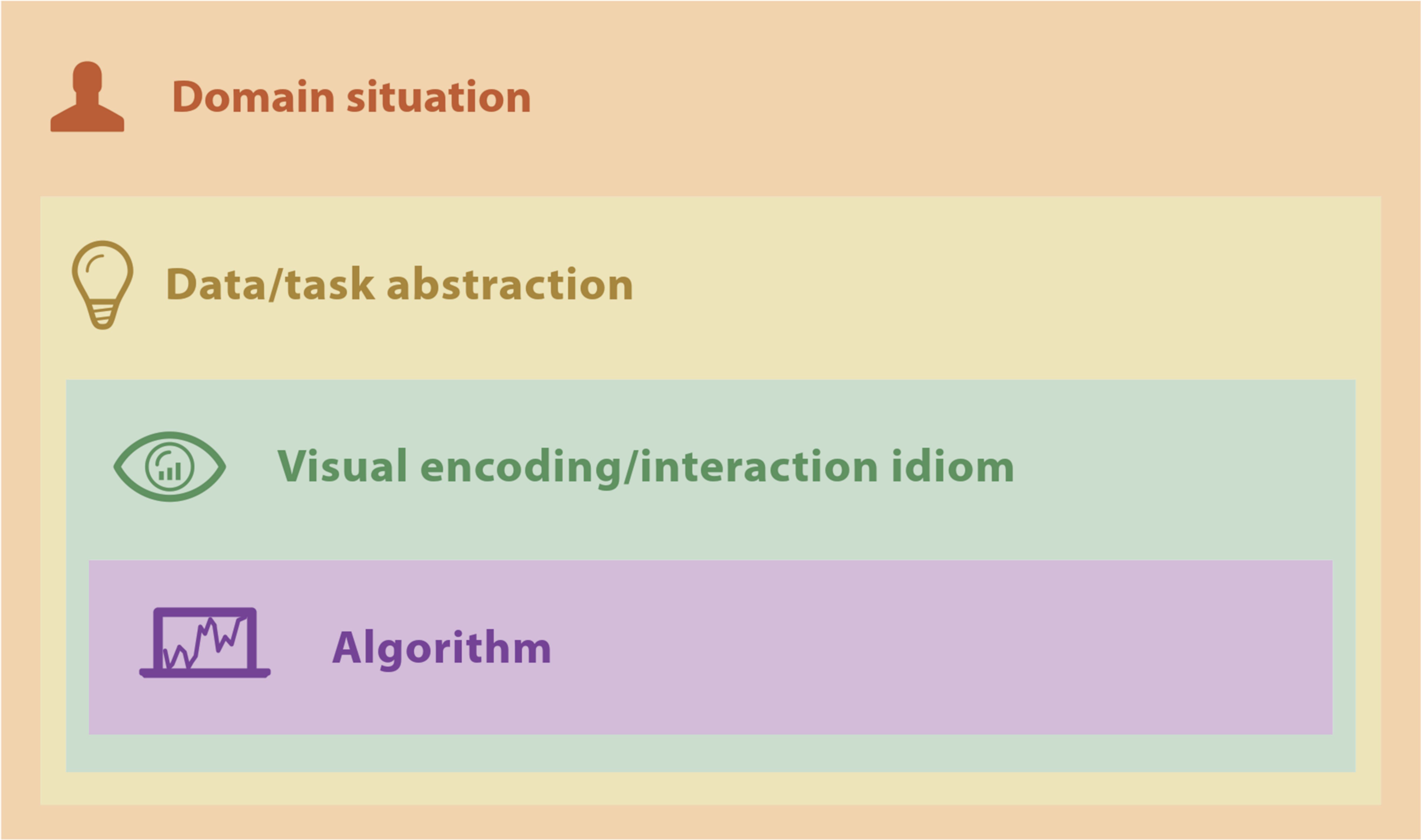


Nested Model

Mistakes propagate through model!



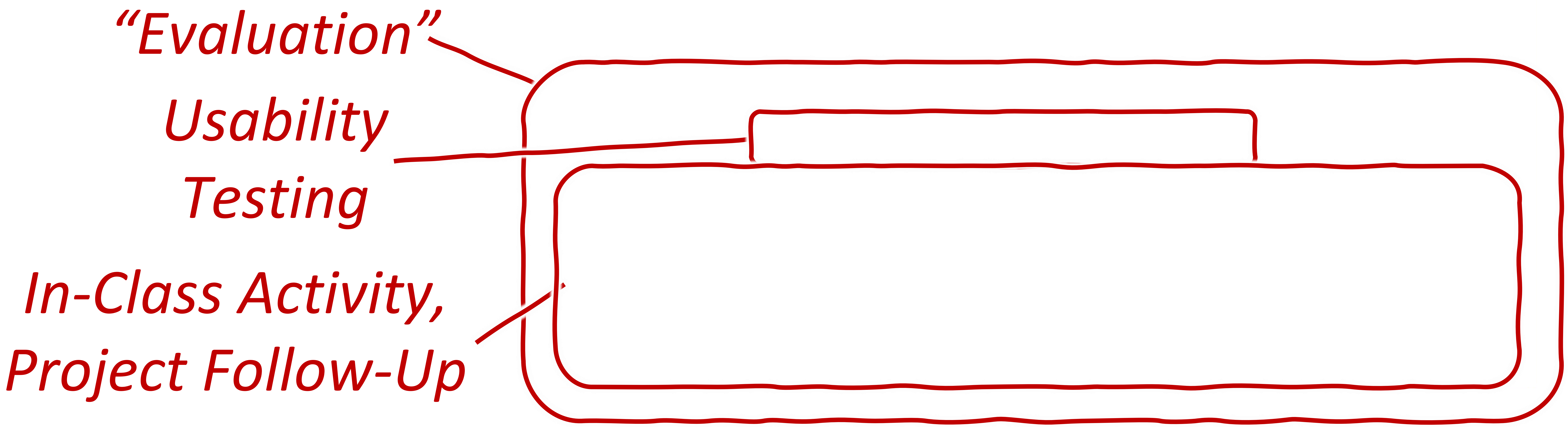
Threats to Validity



Threats to Validity

✓ *Final Project validation*

- Domain situation
- Data/task abstraction
- Visual encoding/interaction idiom
- Algorithm



Empirical Studies in Information Visualization: Seven Scenarios

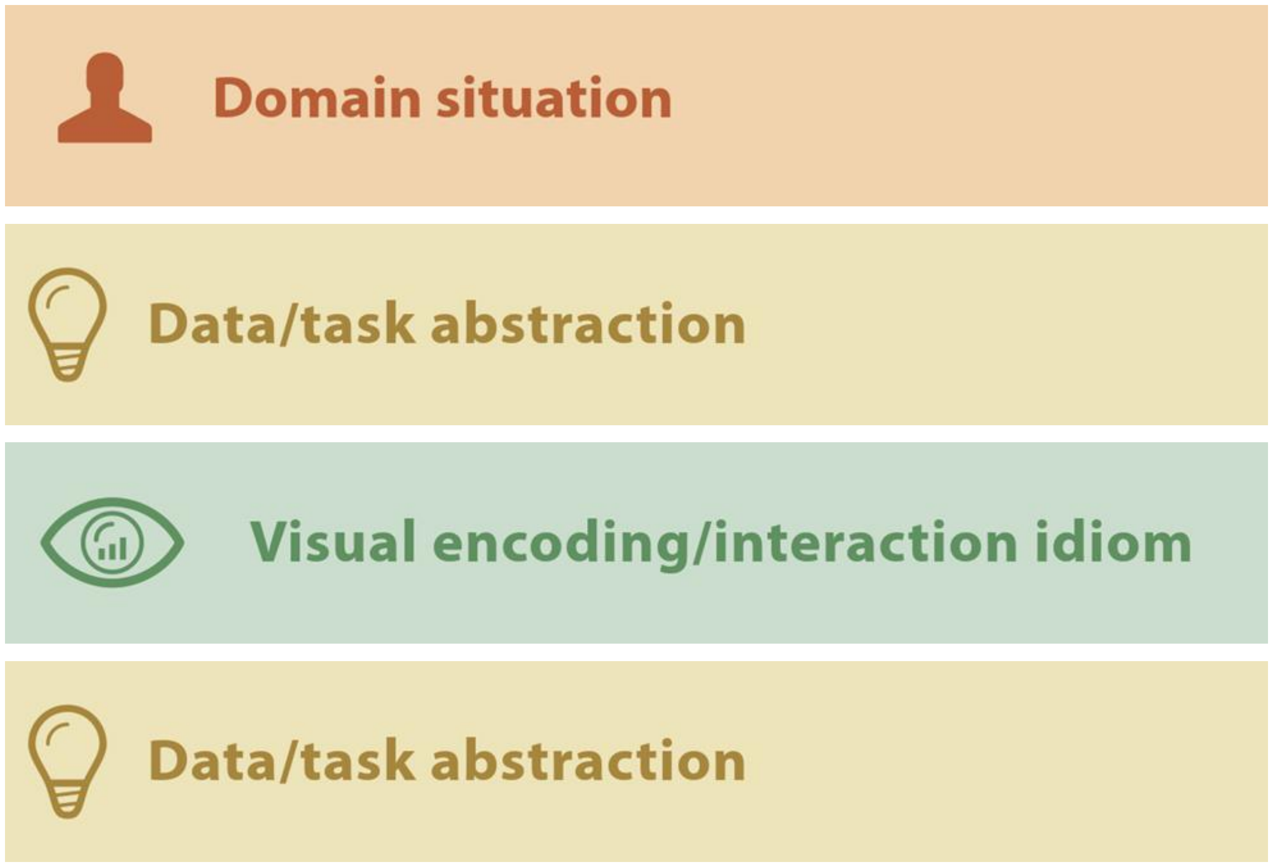
Empirical Studies in Information Visualization: Seven Scenarios

Lam et al., 2012

7 Evaluation Scenarios

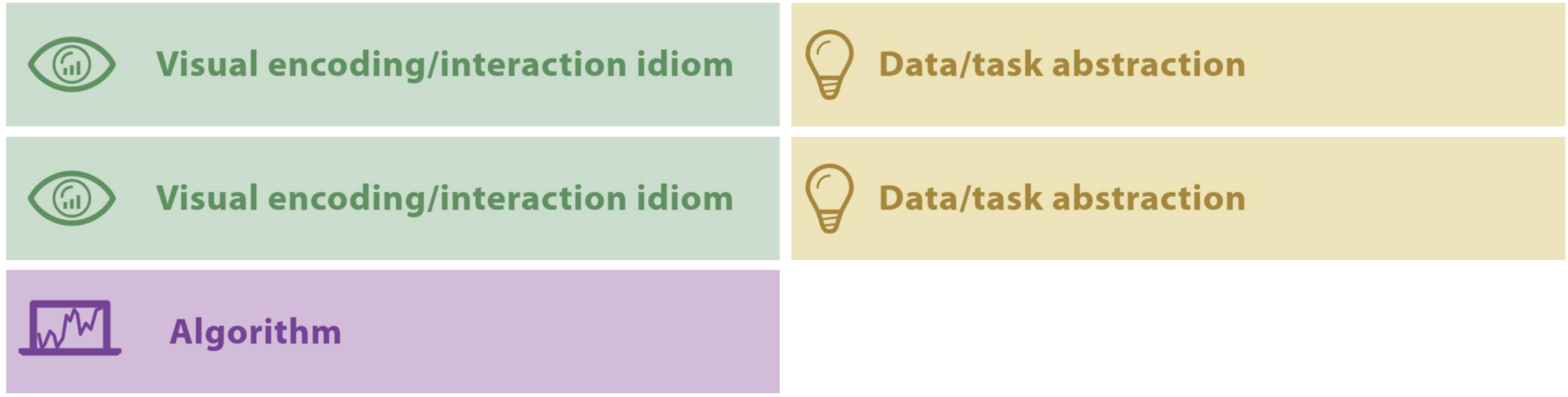
How to understand your data:

- Understanding Environments and Work Practices
- Evaluating Visual Data Analysis and Reasoning
- Evaluating Communication Through Visualization
- Evaluating Collaborative Data Analysis



How to understand your visualization:

- Evaluating User Performance
- Evaluating User Experience
- Evaluating Visualization Algorithms



[Lam et al., 2012](#)

Understanding environments and work practices

- Goals & outputs
  - Understand work, analysis, or information processing practices of people
  - Without software in use: inform design
  - With software in use: assess factors for adoption, how appropriated for future design
- Evaluation Questions
  - Context of use?
  - Integrate into which daily activities?
  - Supported analyses?
  - Characteristics of user group and environment?
  - What data & tasks?
  - What visualizations/tools used?
  - How current tools solve tasks?
  - Challenges and usage barrier?

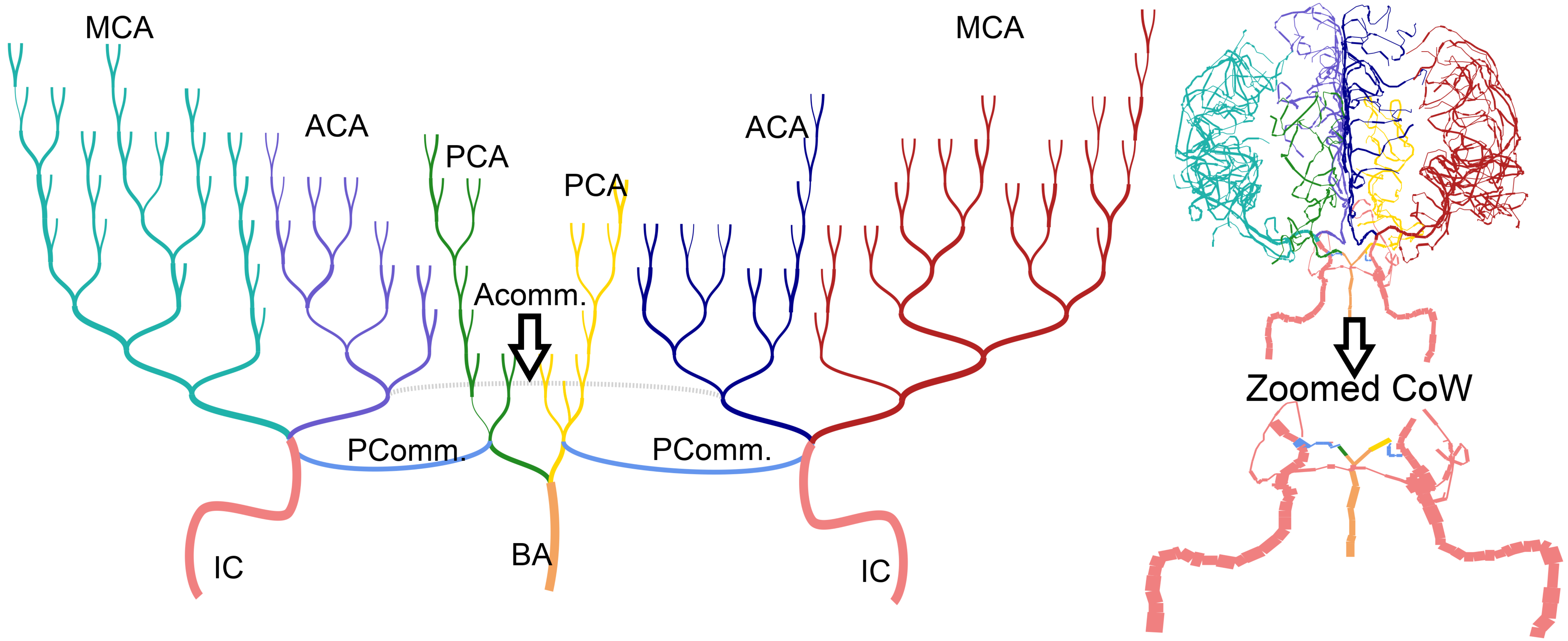
Lam et al., 2012

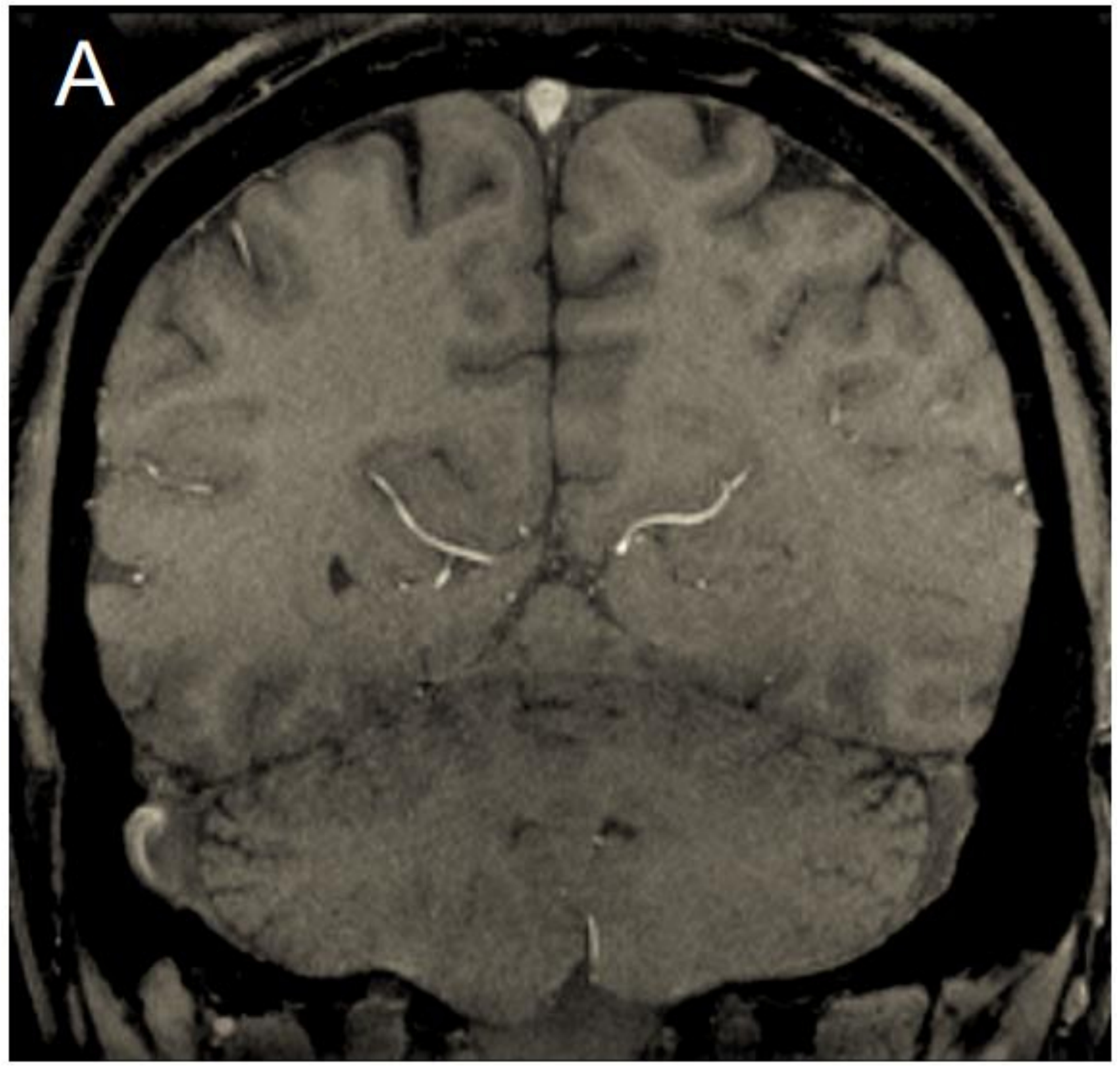
Understanding environments and work practices

- Methods
  - Field Observation
    - Real world, free use of tool
    - Derive requirements
  - Interviews
    - Contextual inquiry: interview then observe in routines, with little interference
    - Pick the right person
    - Laboratory context w/domain expert
  - Laboratory Observation
    - How people interact with each other, tools
    - More control of situation

Lam et al., 2012

Understanding environments and work practices: Example





Pandey, Dunne, et al., 2019

Evaluating visual data analysis and reasoning

Data/task abstraction

- Goals & outputs
  - Assess visualization tool's ability to support visual analysis and reasoning
    - As a whole! Not just a technique
  - Quantifiable metrics or subjective feedback
- Evaluation Questions: Does it support…
  - Data exploration?
  - Knowledge discovery?
  - Hypothesis generation?
  - Decision making?

Lam et al., 2012

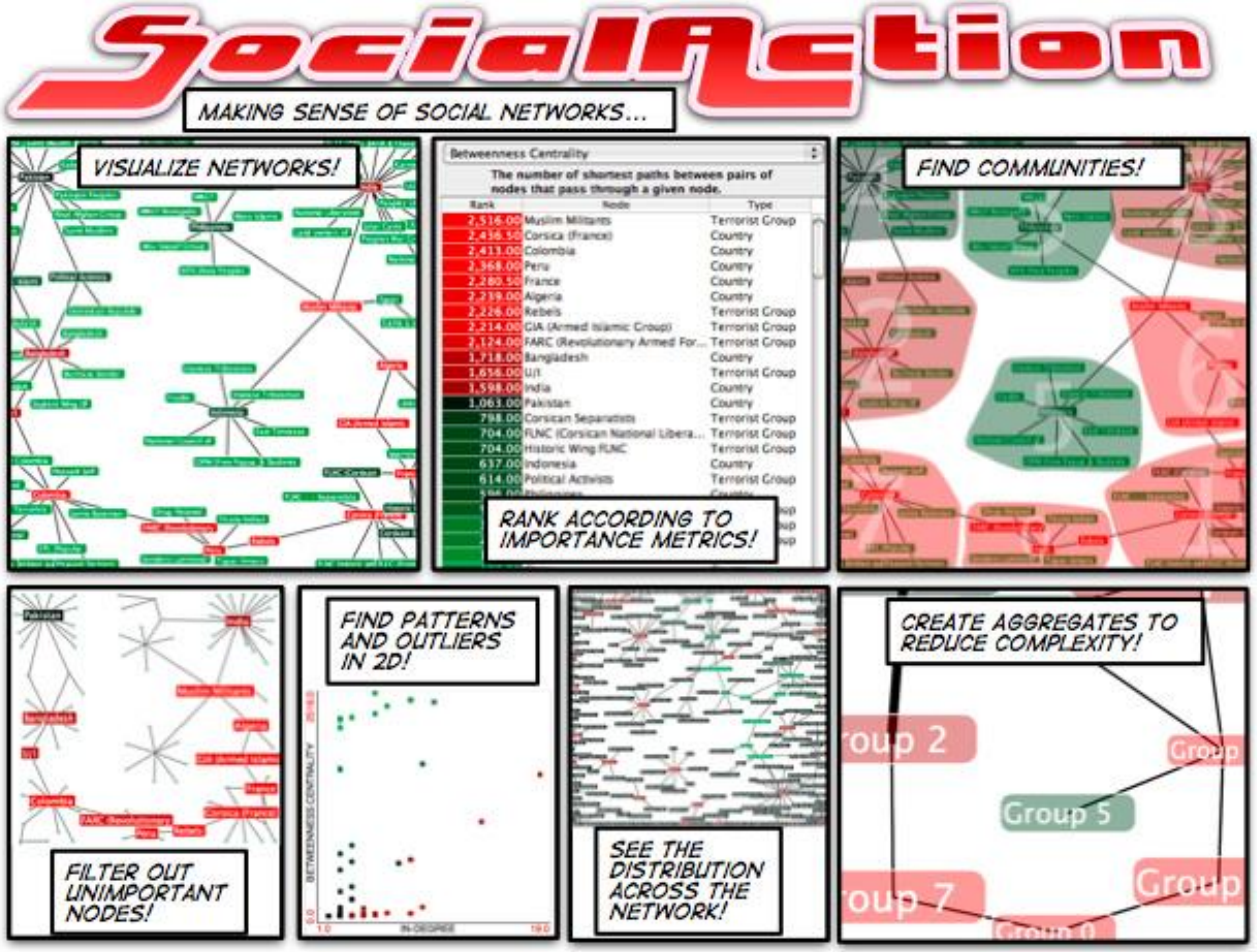
Evaluating visual data analysis and reasoning

Data/task abstraction

- Methods
  - Case studies
    - Motivated experts with own data in own environment
    - Can be longitudinal
    - Insight-Based (Saraiya et al., 2004)
      - Unguided, diary, debriefing meetings
    - MILCS: Multidimensional In-depth Long-term Case studies (Shneiderman & Plaisant, 2006)
      - Guided, observations, interviews, surveys, automated logging
      - Assess interface efficacy, user performance, interface utility
      - Improve system during
  - Lab observations and interviews
    - Code results
    - Think aloud
  - Controlled Experiment
    - Isolate important factors

Lam et al., 2012

Evaluating visual data analysis and reasoning



Perer et al., 2006

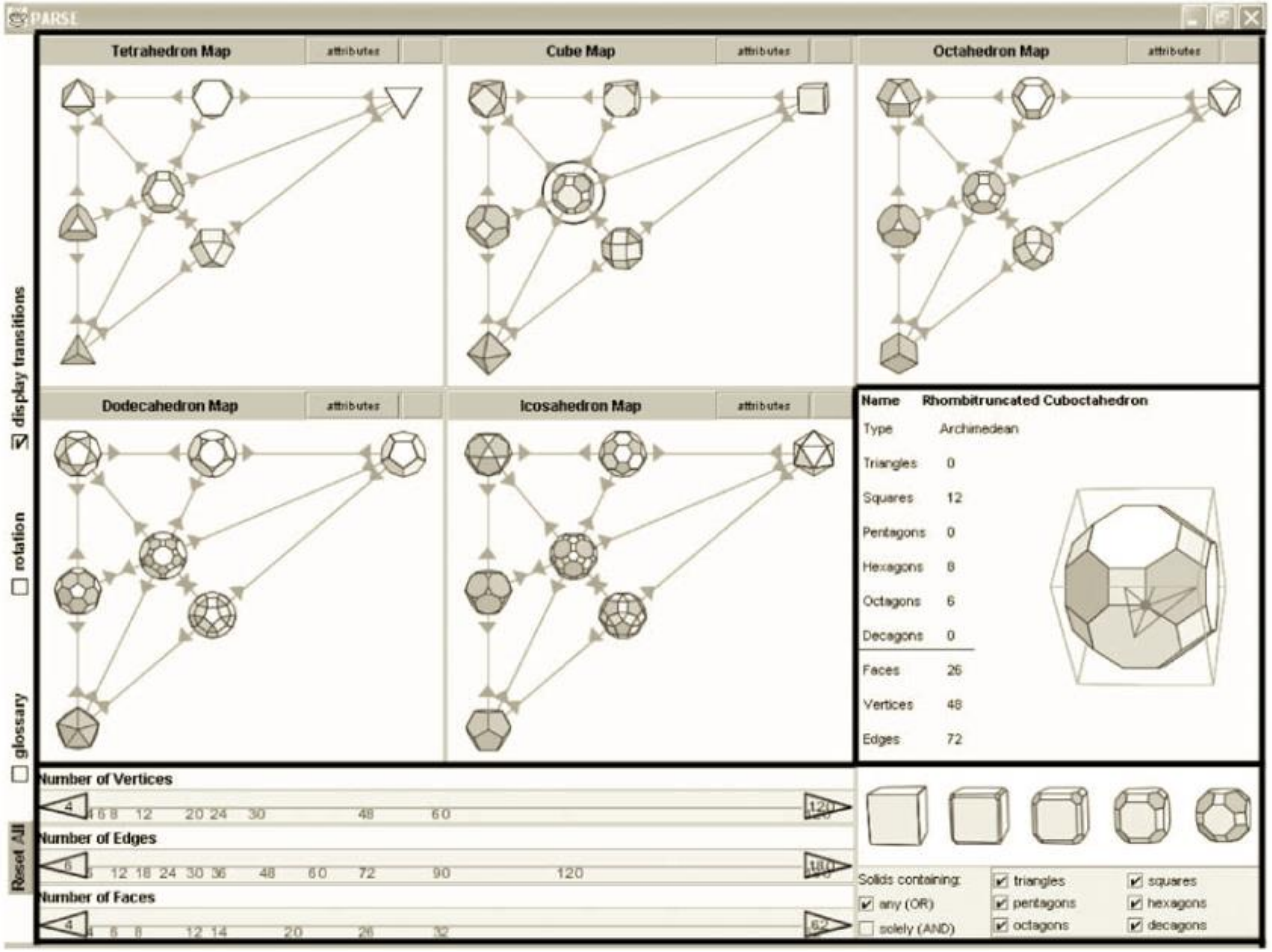
Evaluating communication through visualization

Visual encoding/interaction idiom

- Goals & outputs
  - How effectively is a message delivered and acquired
- Evaluation Questions
  - Quantitative: learning rate, information retention and accuracy
  - Qualitative: interaction patterns
- Methods
  - Controlled experiments
  - Field observation & interviews

Lam et al., 2012

Evaluating communication through visualization: Example



Sedig et al., 2003

Evaluating Collaborative Data Analysis

Data/task abstraction

- Goals & outputs
  - Evaluate support for taskwork and teamwork
  - Holistic understanding of group work processes or tool use
  - Derive design implications
- Evaluation Questions
  - Effective and efficient?
  - Satisfactorily support or stimulate group sensemaking?
  - Support group insight?
  - Is social exchange and communication facilitated?
  - How is the tool used? Features, patterns…
  - What is the process? User requirements?

Lam et al., 2012

Evaluating Collaborative Data Analysis

Data/task abstraction

- Methods
  - Context critical, but early formative studies less dependant
  - Heuristic evaluation
    - Heuristics: actions, mechanics, interactions, locales needed
  - Log analysis
    - Distributed or web-based tools
    - Combine with questionnaire or interview
    - Hard to evaluate unlogged & qualitative aspects
  - Field or laboratory observation
    - Involve group interactions and harmony/disharmony
    - Combine with insight-based?

Lam et al., 2012

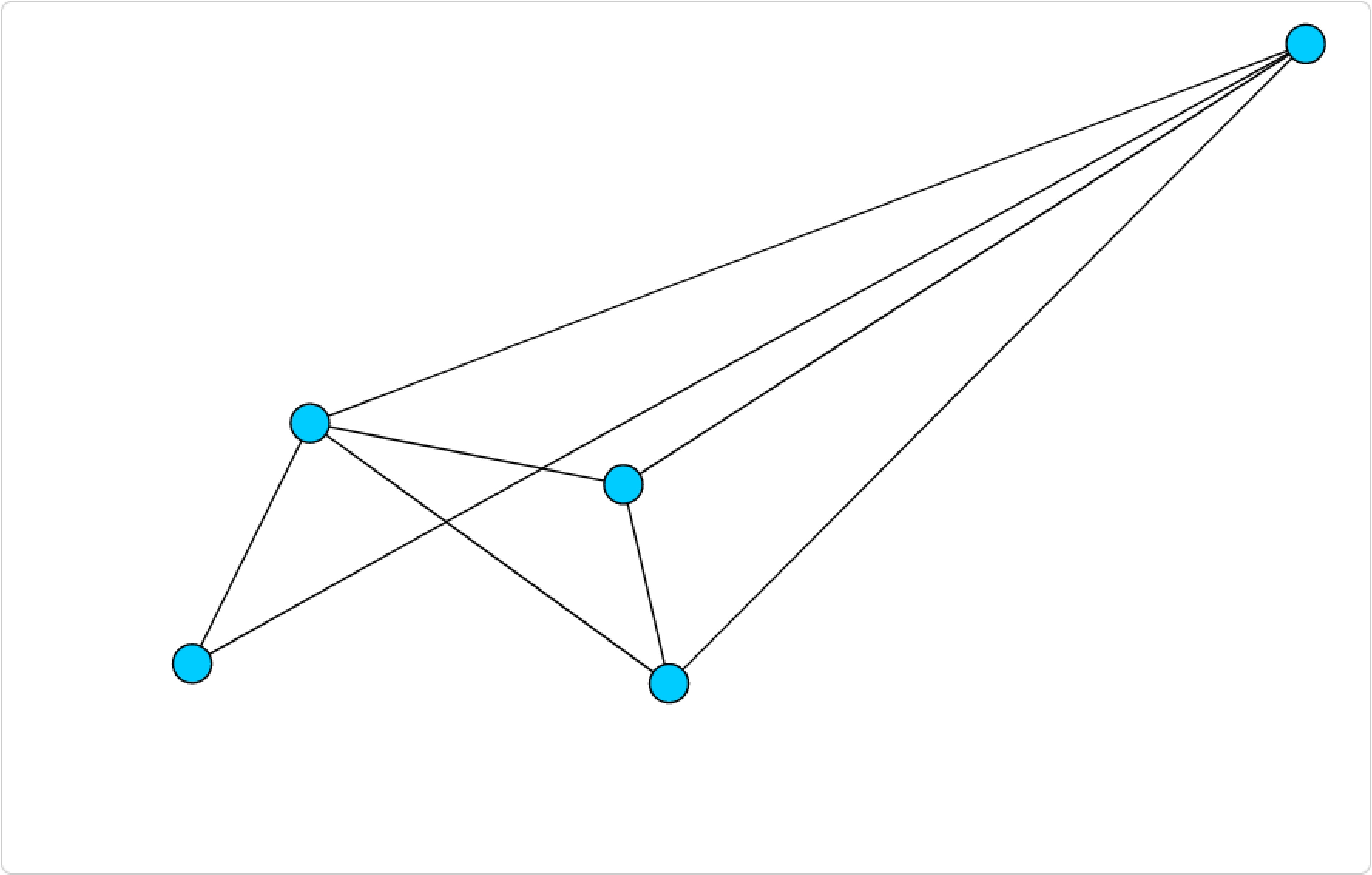
Evaluating Collaborative Data Analysis: Examples

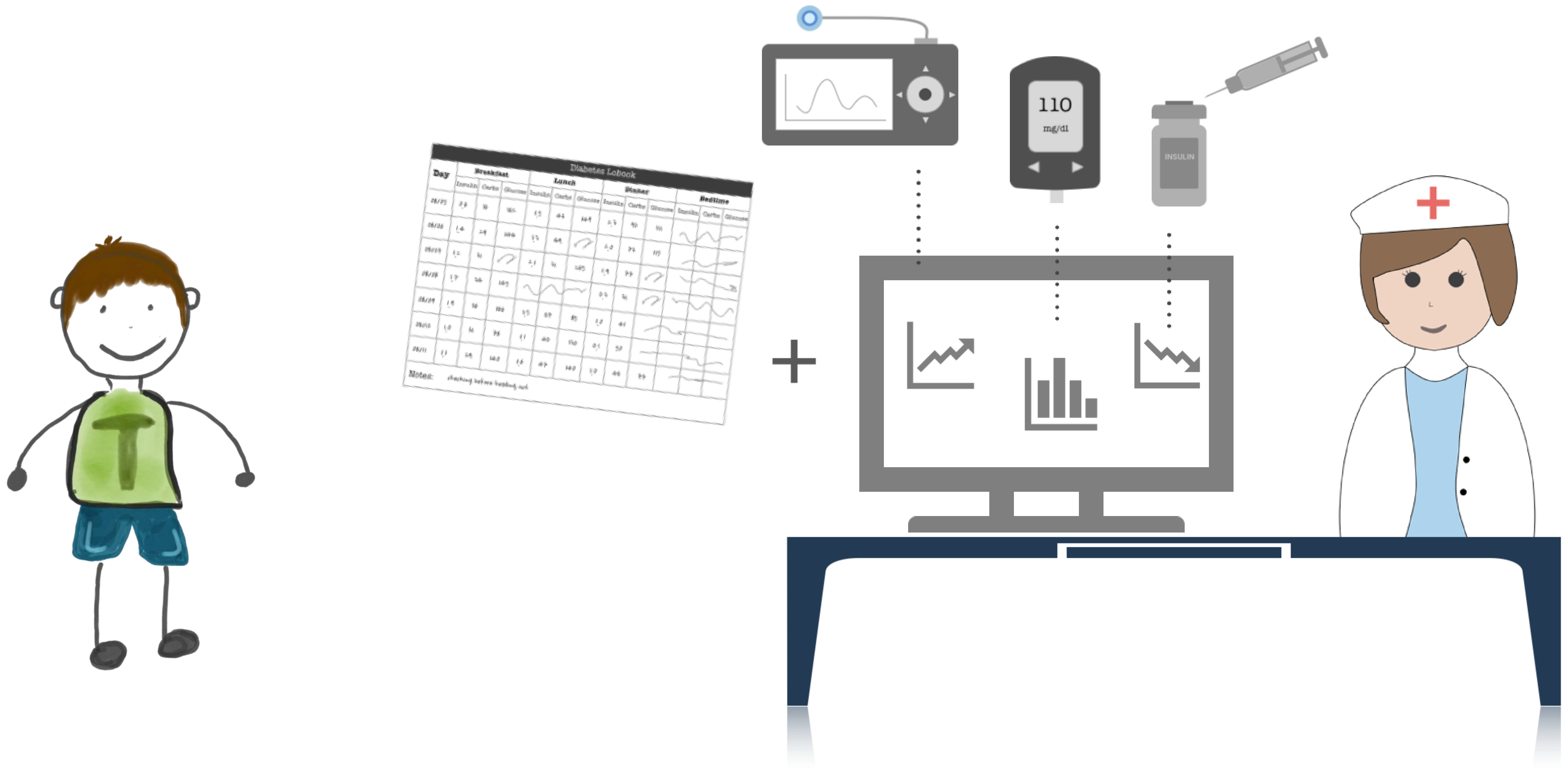
Planarity Party

Can you untangle the graph? See if you can position the vertices so that no two lines cross.

Level 1. Number of line crossings detected: 2.

0 moves. Next Level





Schwab, … Dunne, … et al., 2020

Zhang, … Dunne, … et al., 2018

Evaluating User Performance

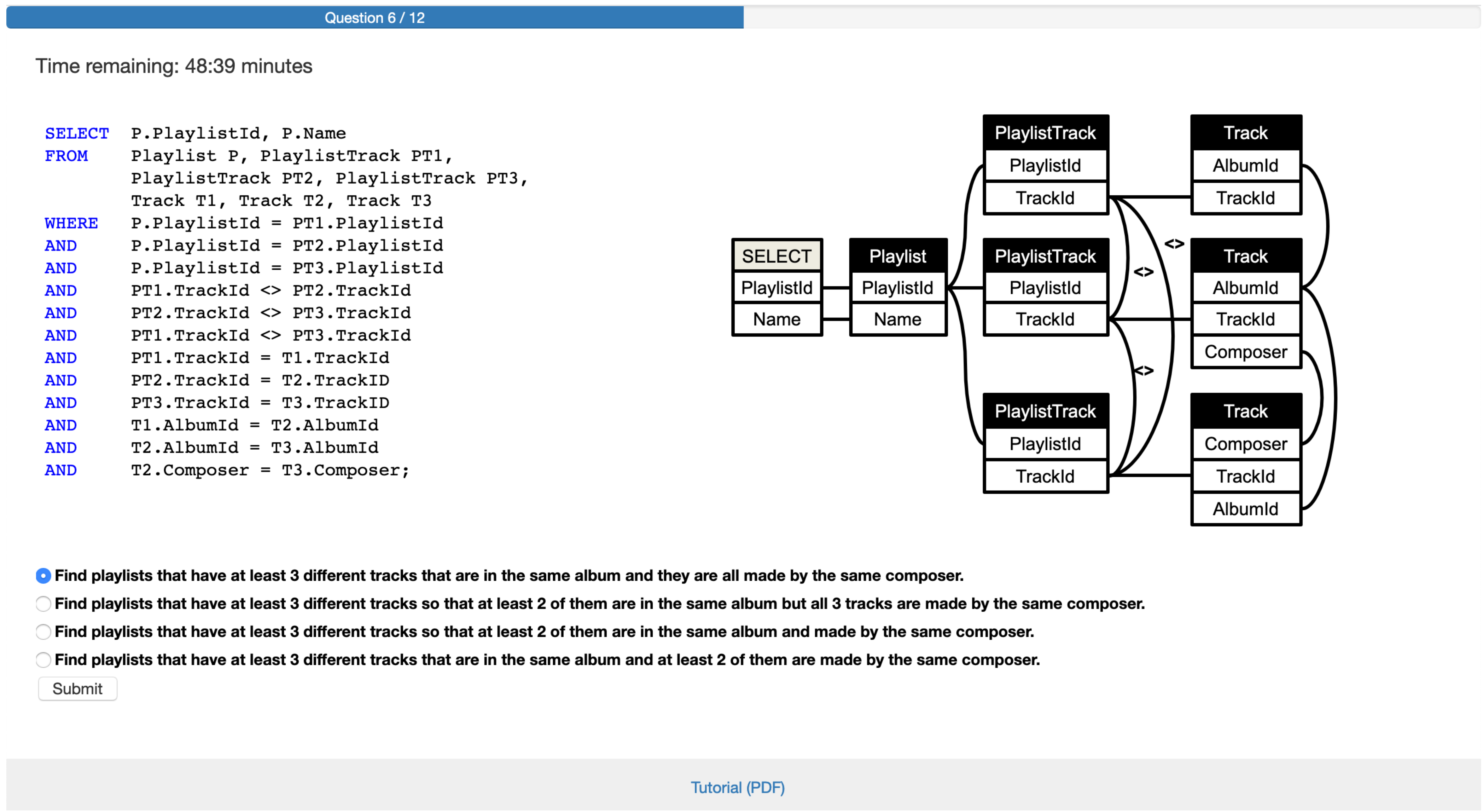
Data/task abstraction

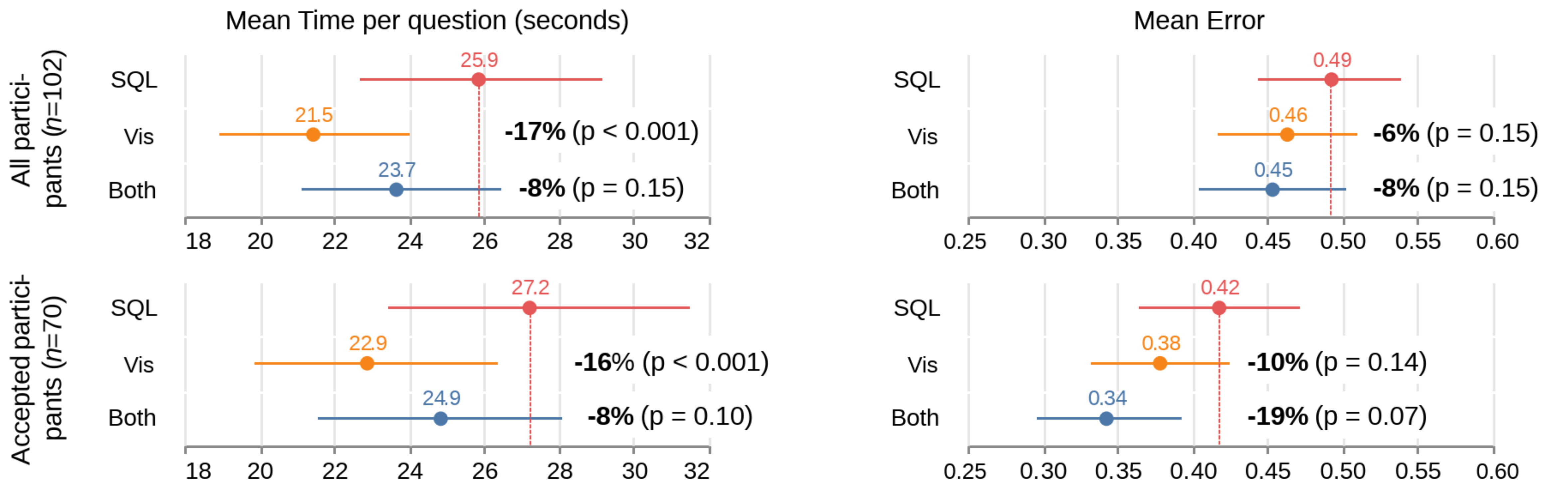
Visual encoding/interaction idiom

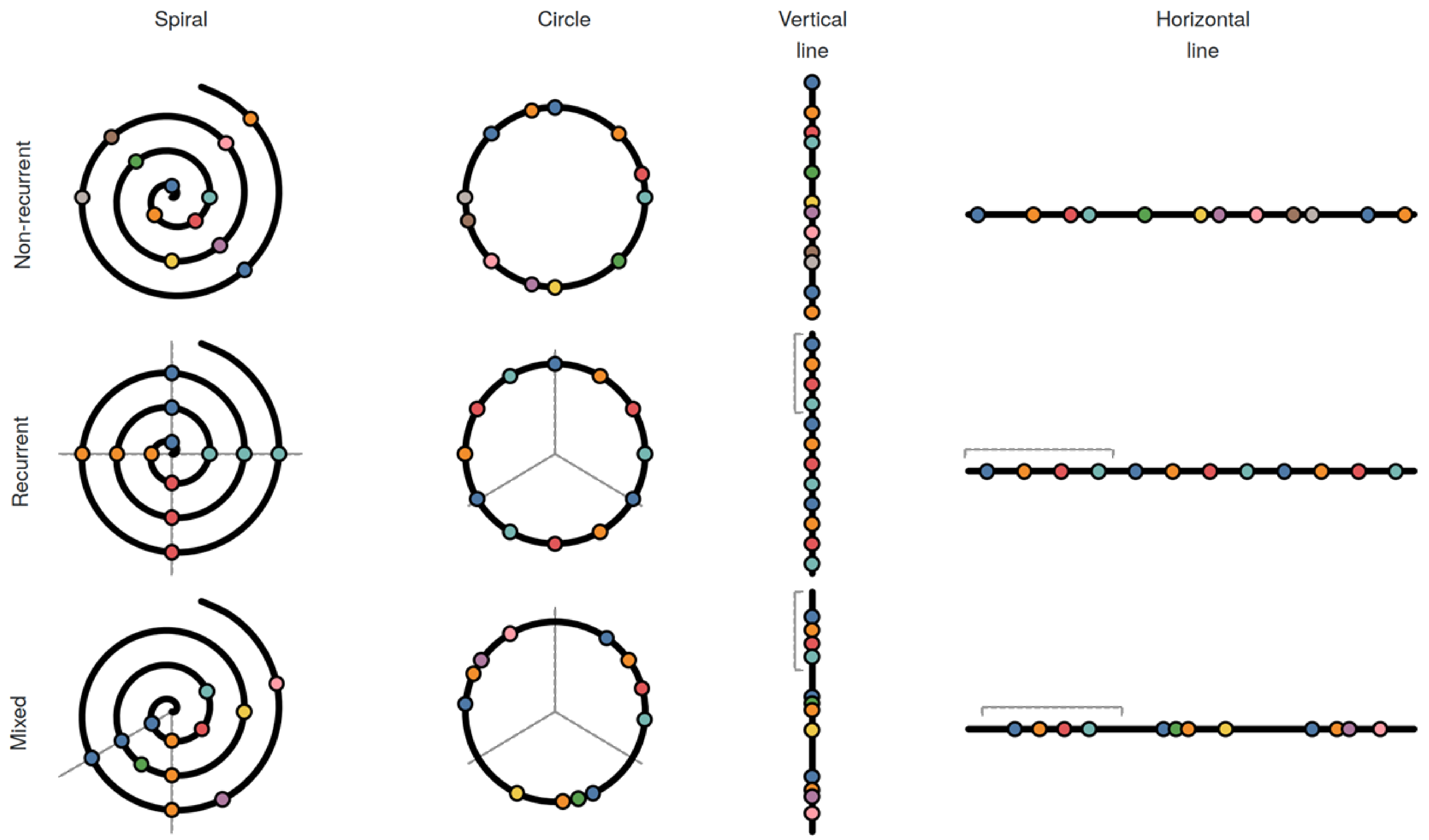
- Goals & outputs
  - Measure specific features
  - Time, accuracy, and error; work quality (if quantifiable); memorability
  - Descriptive statistics results
- Evaluation Questions
  - What are the limits of human perception and cognition?
  - How do techniques compare?
- Methods
  - Controlled experiment → design guideline, model, head-to-head
    - Few variables
    - Simple tasks
    - Individual differences matter
  - Field logs
    - Suggest improvements, recommendation systems

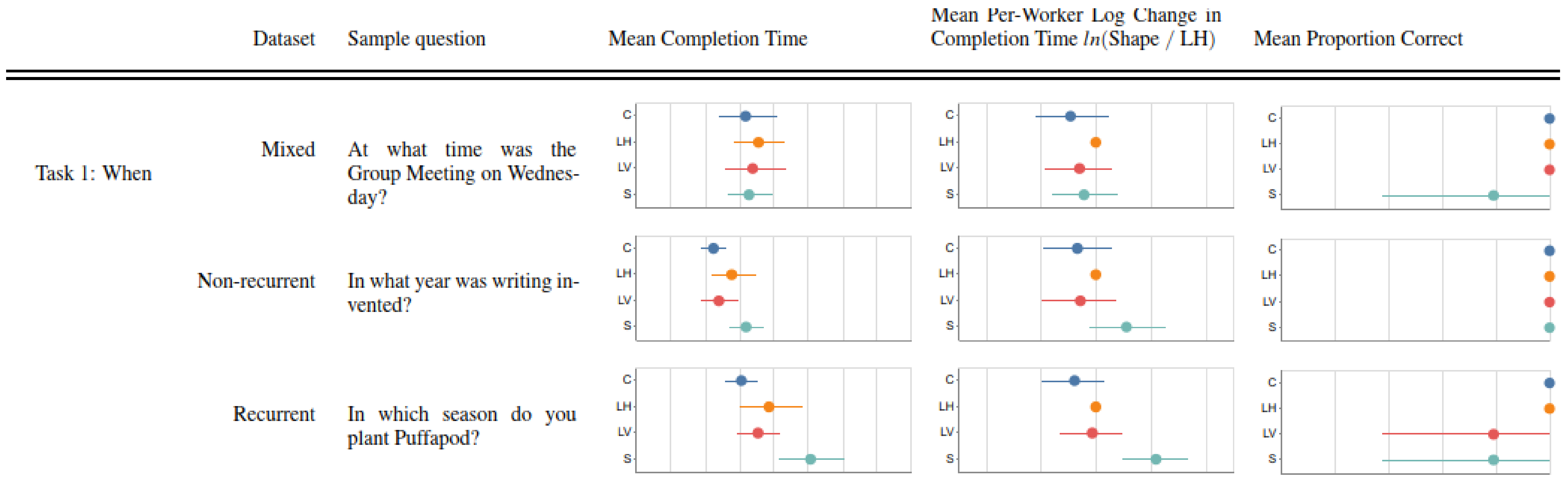
Lam et al., 2012

Evaluating User Performance: Examples









[Leventidis, Dunne, et al., 2020](#) [Di Bartolomeo, Dunne, et al., 2020](#)

Evaluating User Experience

Data/task abstraction

Visual encoding/interaction idiom

- Goals & outputs
  - Inform design: uncover gaps in functionality, limitations, directions for improvement
  - Subjective: user responses
    - Effectiveness, efficiency, correctness, satisfaction, trust, features liked/disliked
  - Objective: body sensors, eye tracking
- Evaluation Questions
  - Features: useful, missing, to rework?
  - Are there limitations that hinder adoption?
  - Is the tool understandable/learnable?

Lam et al., 2012

Evaluating User Experience

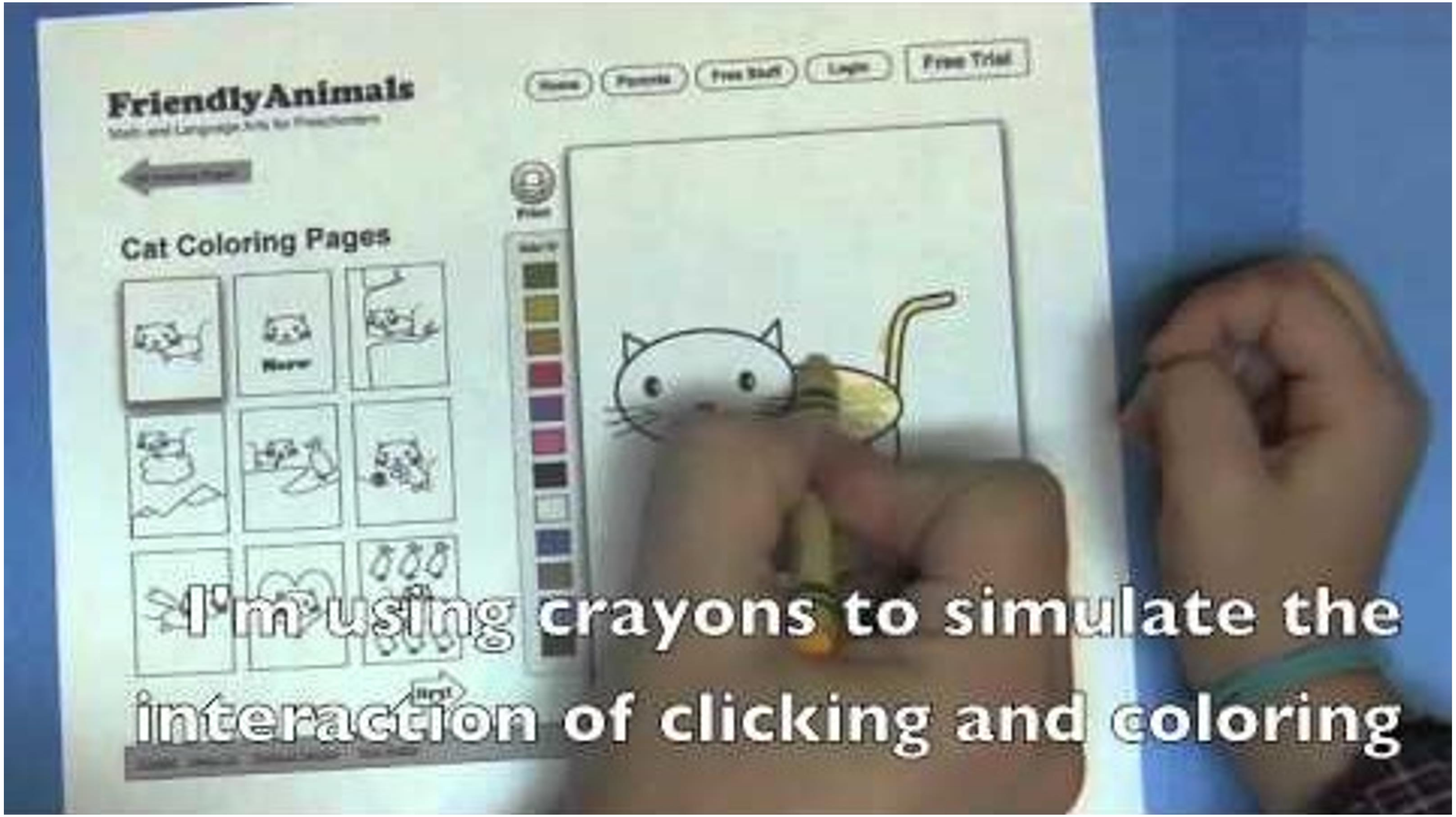
Data/task abstraction

Visual encoding/interaction idiom

- Methods
  - Informal evaluation
    - Demo for domain experts (usually) and collect feedback
  - Usability test
    - Watch (video) how participants perform set of tasks to perfect design
    - Take note of behaviors, remarks, problems
    - Carefully prepare tasks, interview script, questionnaires
  - Field observation
    - Understand interaction in real setting
  - Laboratory questionnaire
    - Likert scale
    - Open ended

Lam et al., 2012

Evaluating User Experience: Example



BlueDuckLabs, 2010

Evaluating Visualization Algorithms

Data/task abstraction

Visual encoding/interaction idiom

- Goals & outputs
  - Quantitatively or qualitatively judge generated output quality (metrics) & performance
  - How scores vs. alternatives
  - Explore limits & behavior
- Evaluation Questions
  - Which shows interesting patterns best?
  - Which is more truthful?
  - Which is less cluttered?
  - Faster, less memory, less money?
  - How does it scale?
  - Extreme cases?

Lam et al., 2012

Evaluating Visualization Algorithms

- Methods
  - Visualization quality assessment
    - Readability metrics, image quality measures
  - Algorithmic performance
    - Varied data, size, complexity, corner cases
    - Benchmark data sets

Data/task abstraction

Visual encoding/interaction idiom

Lam et al., 2012

Evaluating Visualization Algorithms: Example

Type	Name	$recn_\Gamma$				
		GVA	FM^3	FMS	ACE	HDE
Kind Arti-ficial	rnd_grid_032	3.82	**0**	**0**	**0**	<0.01
	rnd_grid_100	14.75	**0**	**0**	**0**	<0.01
	rnd_grid_320	181.51	**0**	(N)	<0.01	<0.01
	sierpinski_06	2.00	0.05	<0.01	**0**	0.02
	sierpinski_08	9.49	0.07	**0.01**	0.02	0.08
	sierpinski_10	99.97	0.09	(N)	0.27	**0.01**
Kind Real World	crack	30.82	<0.01	(N)	**0**	0.07
	fe_pwt	150.70	2.45	(N)	(N)	**1.61**
	finan_512	301.25	18.81	(N)	**12.27**	21.27
	fe_ocean	622.48	**7.13**	(N)	9.07	8.24
Chal-lenging Arti-ficial	tree_06_04	2.21	1.16	7.89	0.01	**0**
	tree_06_05	9.33	1.89	11.48	**0**	22.92
	tree_06_06	70.68	**3.31**	(N)	4.16	128.82
	snowflake_A	0.63	**0**	0.10	<0.01	0.62
	snowflake_B	1.46	**0**	8.18	(N)	6.92
	snowflake_C	15.53	**0**	(N)	(N)	195.87
	spider_A	15.62	16.55	**1.17**	6.60	1.25
	spider_B	154.70	132.96	1.64	**0**	**0**
	spider_C	2522.89	1029.64	(N)	**0**	**0**
	flower_A	46.71	49.08	5.63	**0.26**	0.55
	flower_B	64.90	51.57	1.90	**0.06**	0.34
	flower_C	578.22	53.39	(N)	(N)	**0.30**
Chal-lenging Real World	ug_380	22.93	19.55	13.67	20.99	**1.35**
	esslingen	47.52	23.71	28.42	20.81	**3.89**
	add_32	8.65	1.69	5.75	**0.89**	5.80
	dg_1087	1.74	< **0.01**	37.07	5.92	6.49
	bcsstk_33	720.94	376.18	4171.05	413.56	**113.86**
	bcsstk_31	708.69	94.26	(N)	**63.00**	611.21

Table 3: The relative edge-crossing numbers ($recn_\Gamma$) of the drawings Γ computed by the tested algorithms. The entry (N) indicates that no drawing was computed. Best values are printed bold. Worst values are underlined.

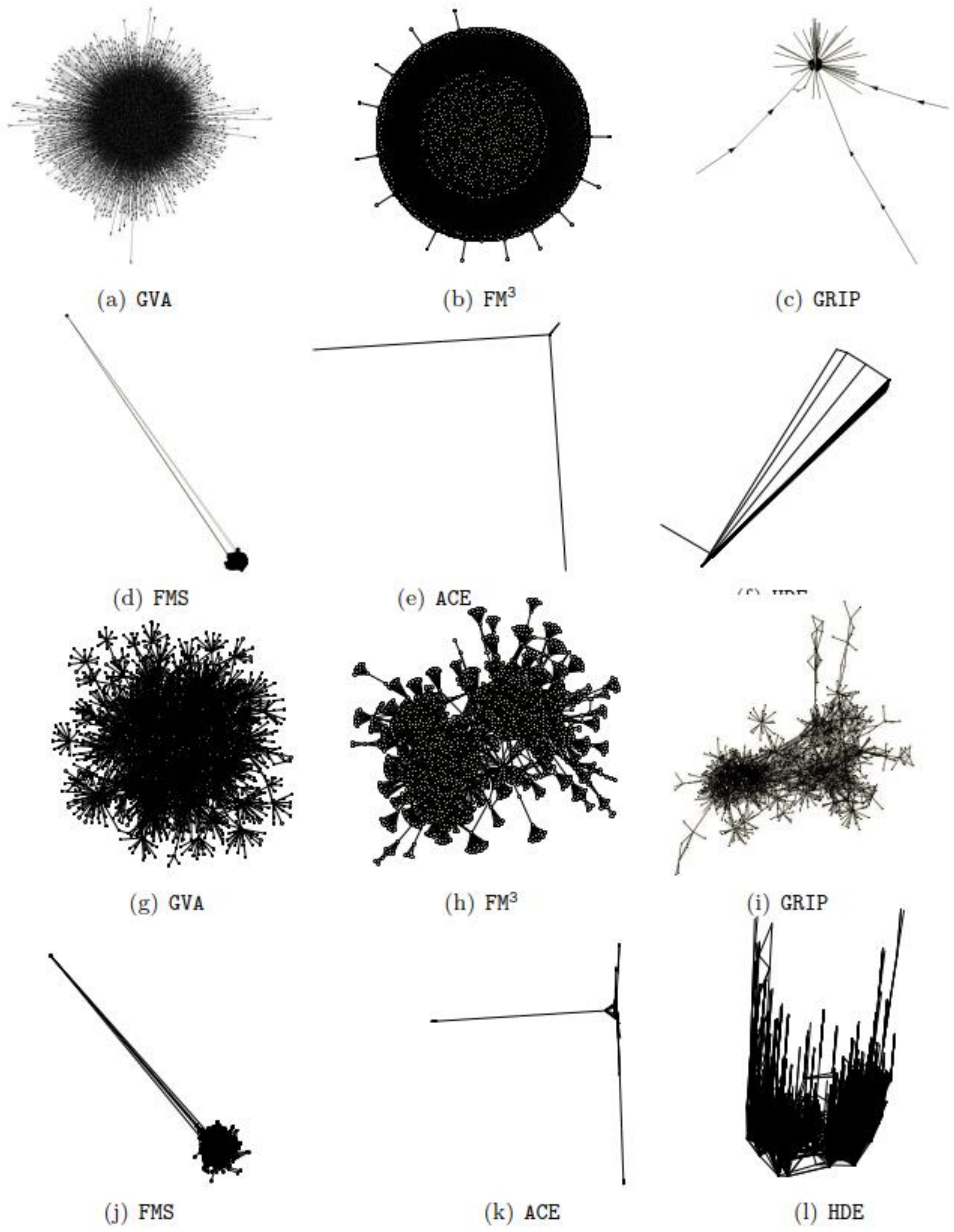


Figure 8: (a)-(f) Drawings of dg_1087 and (g)-(l) esslingen generated by different algorithms

7 Evaluation Scenarios

How to understand your data:

- Understanding Environments and Work Practices — Domain situation
- Evaluating Visual Data Analysis and Reasoning — Data/task abstraction
- Evaluating Communication Through Visualization — Visual encoding/interaction idiom
- Evaluating Collaborative Data Analysis — Data/task abstraction

How to understand your visualization:

- Evaluating User Performance — Visual encoding/interaction idiom, Data/task abstraction
- Evaluating User Experience — Visual encoding/interaction idiom, Data/task abstraction
- Evaluating Visualization Algorithms — Algorithm

Lam et al., 2012

7 Evaluation Scenarios

How to understand your data:

- Understanding Environments and Work Practices — Field Observations, Interviews
- Evaluating Visual Data Analysis and Reasoning — Case Studies, Controlled Experiment
- Evaluating Communication Through Visualization — Field Observation, Controlled Experiment
- Evaluating Collaborative Data Analysis — Field Observation, Heuristic Evaluation, Log Analysis

How to understand your visualization:

- Evaluating User Performance — Controlled Experiment, Log Analysis
- Evaluating User Experience — Informal Evaluation, Usability Test, Field Observation
- Evaluating Visualization Algorithms — Visualization Quality Assessment, Algorithm Performance

Lam et al., 2012

In-Class Validation — Final Project Evaluation

~35 min