Problem-Driven Visualization Through Design Studies

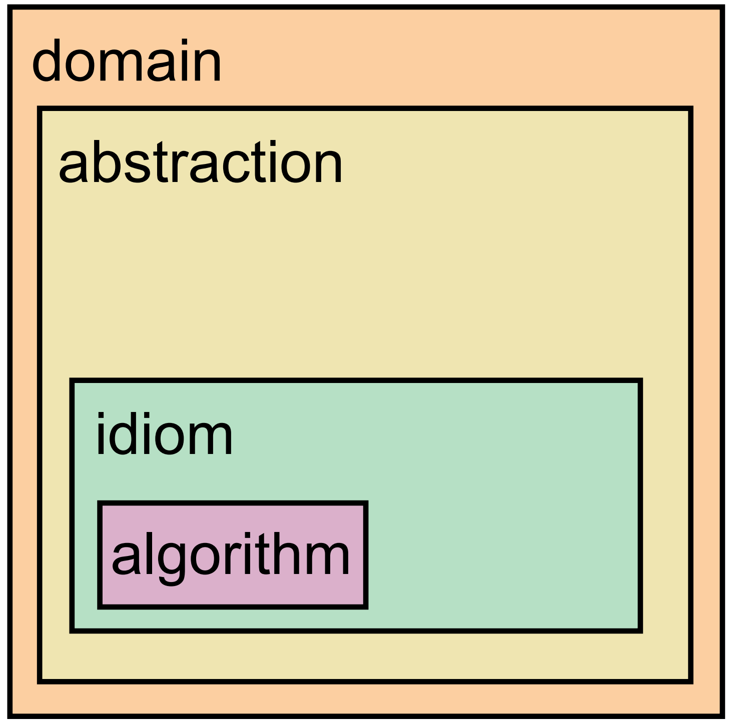
Tamara Munzner
Department of Computer Science
University of British Columbia
@tamaramunzner

VINCI 2021 Keynote
Sep 7 2021, virtual / Potsdam

http://www.cs.ubc.ca/~tmm/talks.html#vinci21

UBC
DESIGNING for PEOPLE
UBC Data Science Institute
CAIDA

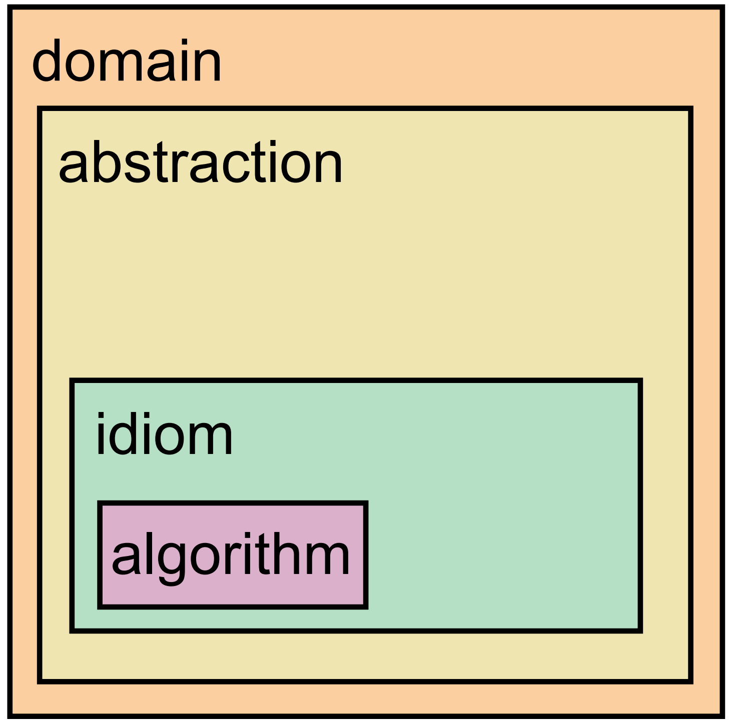
Nested model: Four levels of visualization concerns



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

Nested model: Four levels of visualization concerns

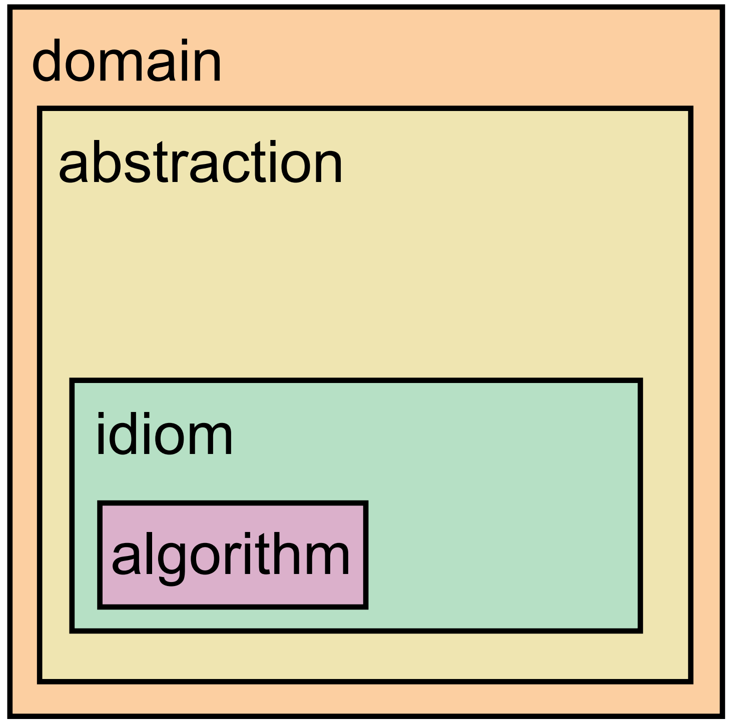
- *domain* situation
  - **who** are the target users?



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

Nested model: Four levels of visualization concerns

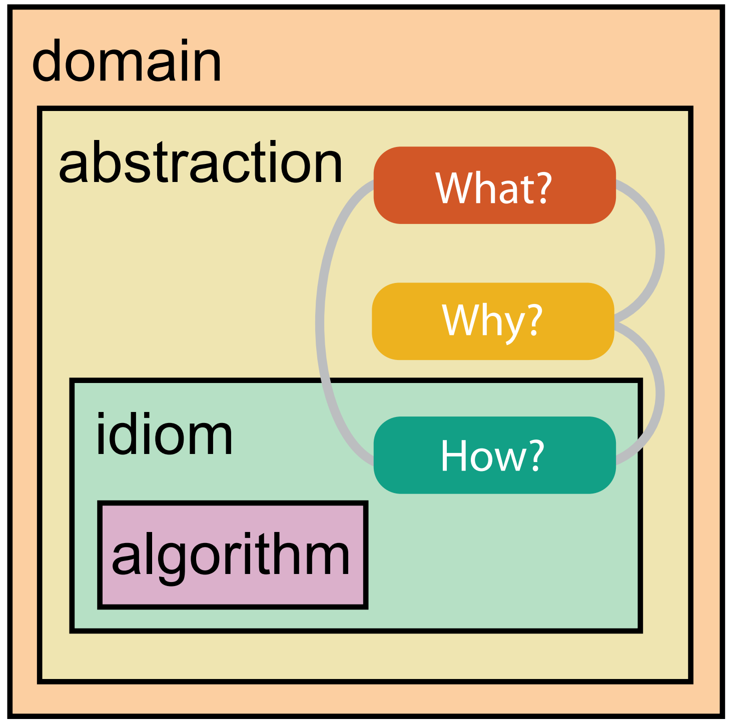
- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis



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

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**

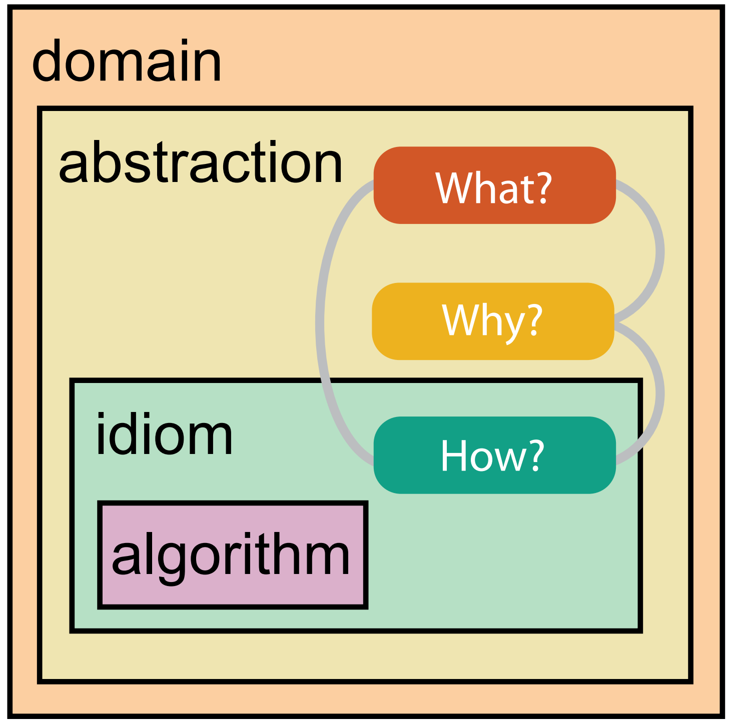


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

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

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - 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

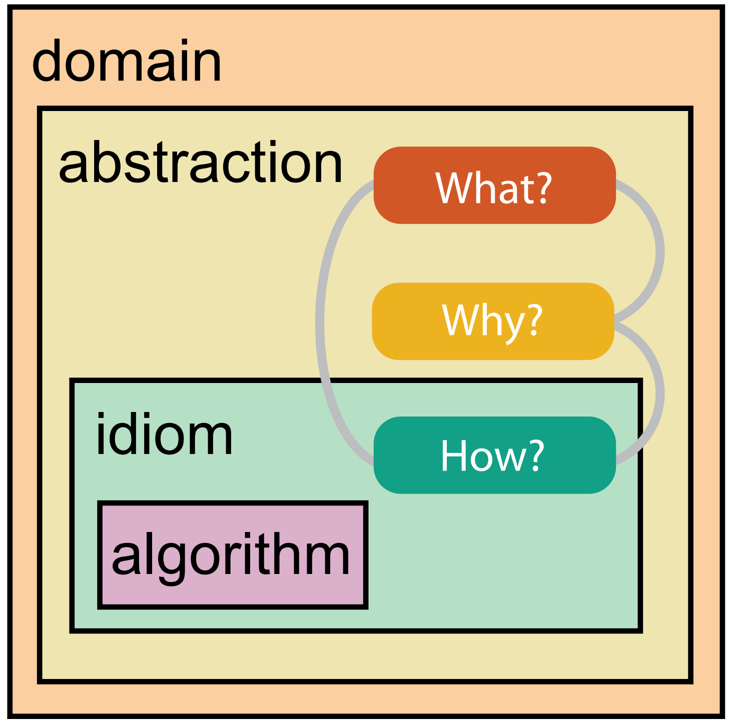


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

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

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - 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**

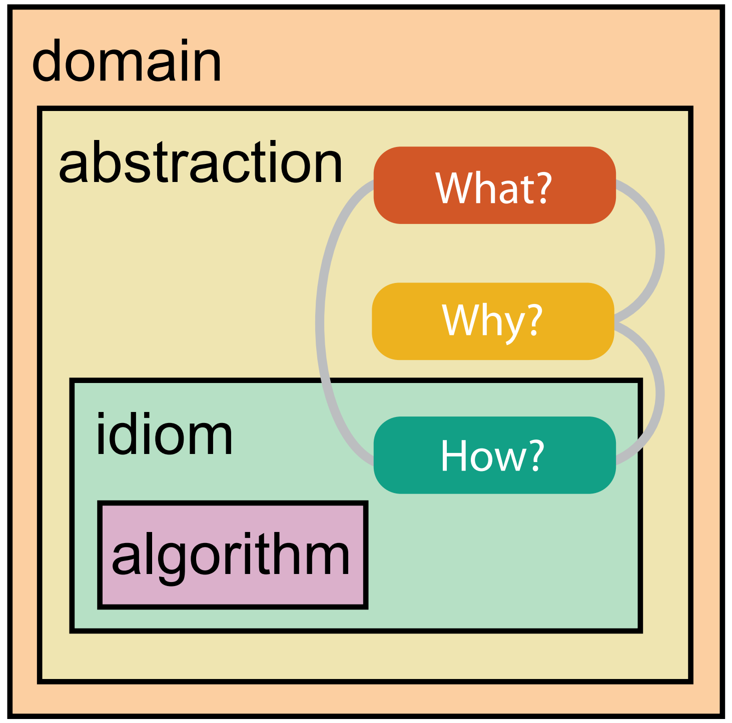


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

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

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - 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?

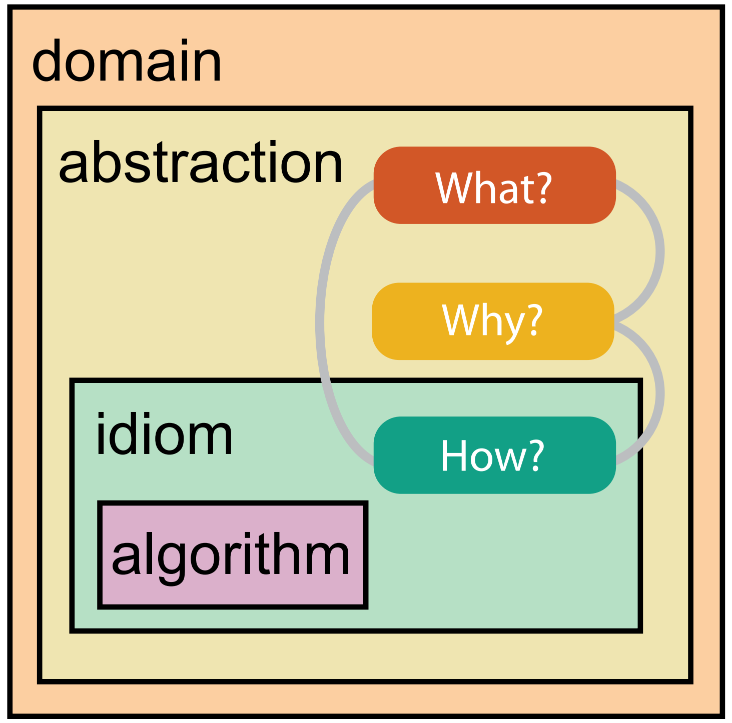


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

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

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - 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

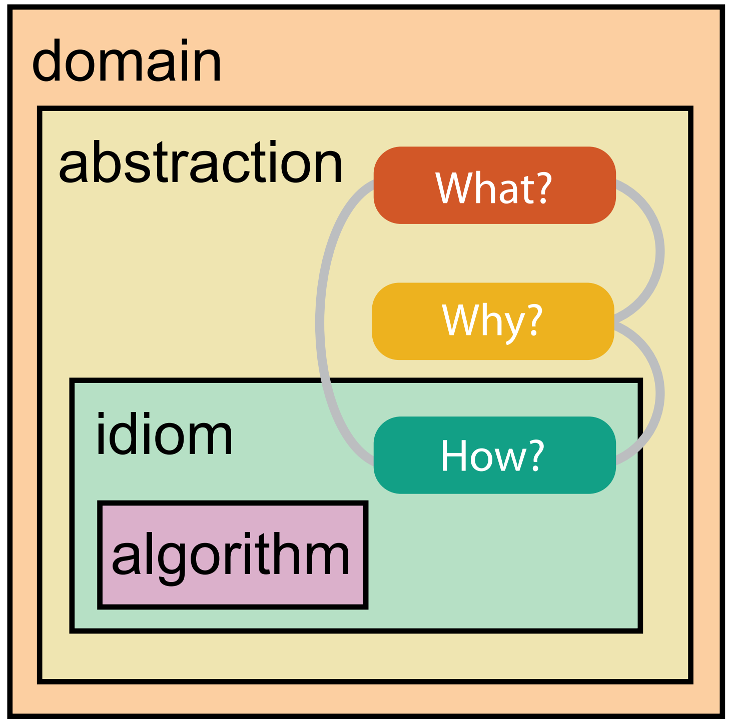


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

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

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - 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

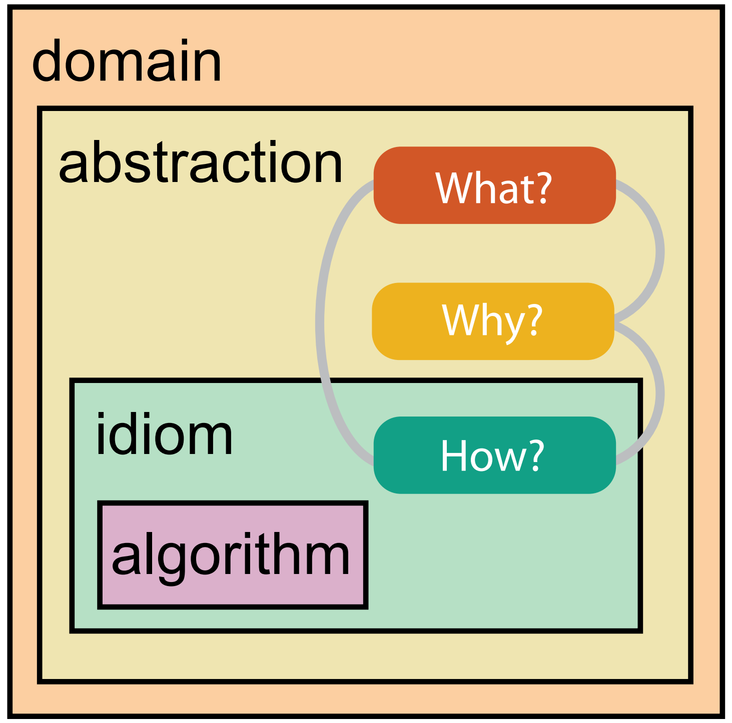


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

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

Nested model: Four levels of visualization concerns

- *domain* situation
  - **who** are the target users?
- *abstraction*
  - 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

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

Why is validation difficult?

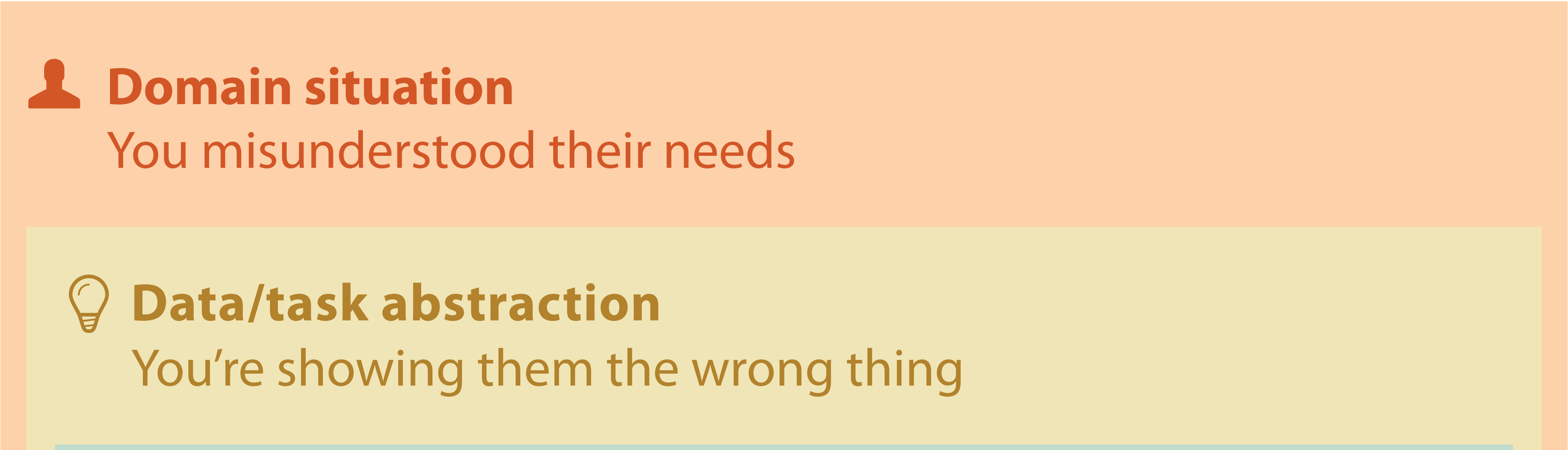
- different ways to get it wrong at each level

Domain situation
You misunderstood their needs

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

Why is validation difficult?

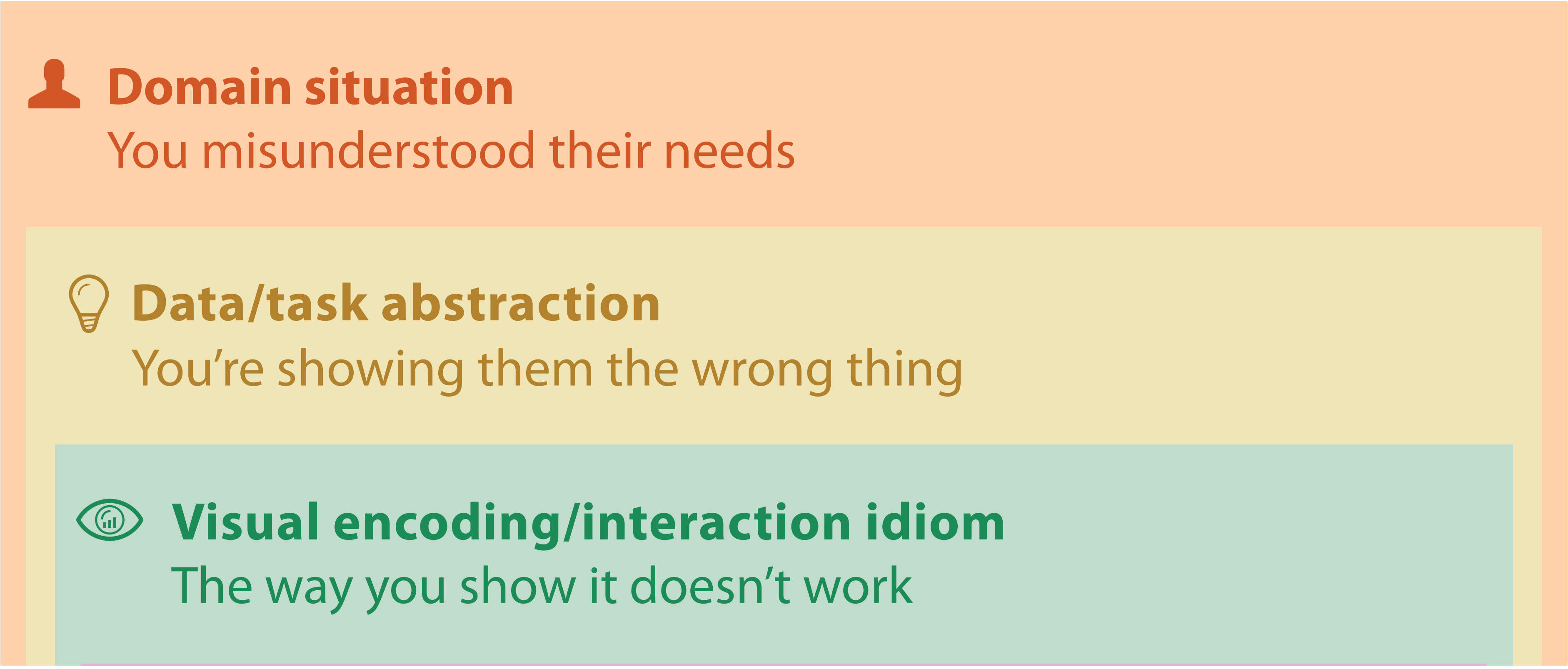
- different ways to get it wrong at each level



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

Why is validation difficult?

- different ways to get it wrong at each level



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

Why is validation difficult?

- different ways to get it wrong at each level

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

Validation solution: use methods from appropriate fields at each level

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

Validation solution: use methods from appropriate fields at each level

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

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

Validation solution: use methods from appropriate fields at each level

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

technique-driven work

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

Validation solution: use methods from appropriate fields at each level

design

computer science

cognitive psychology

Visual encoding/interaction idiom
Justify design with respect to alternatives

Algorithm
Measure system time/memory
Analyze computational complexity

Analyze results qualitatively

Measure human time with lab experiment (*lab study*)

technique-driven work

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

Validation solution: use methods from appropriate fields at each level

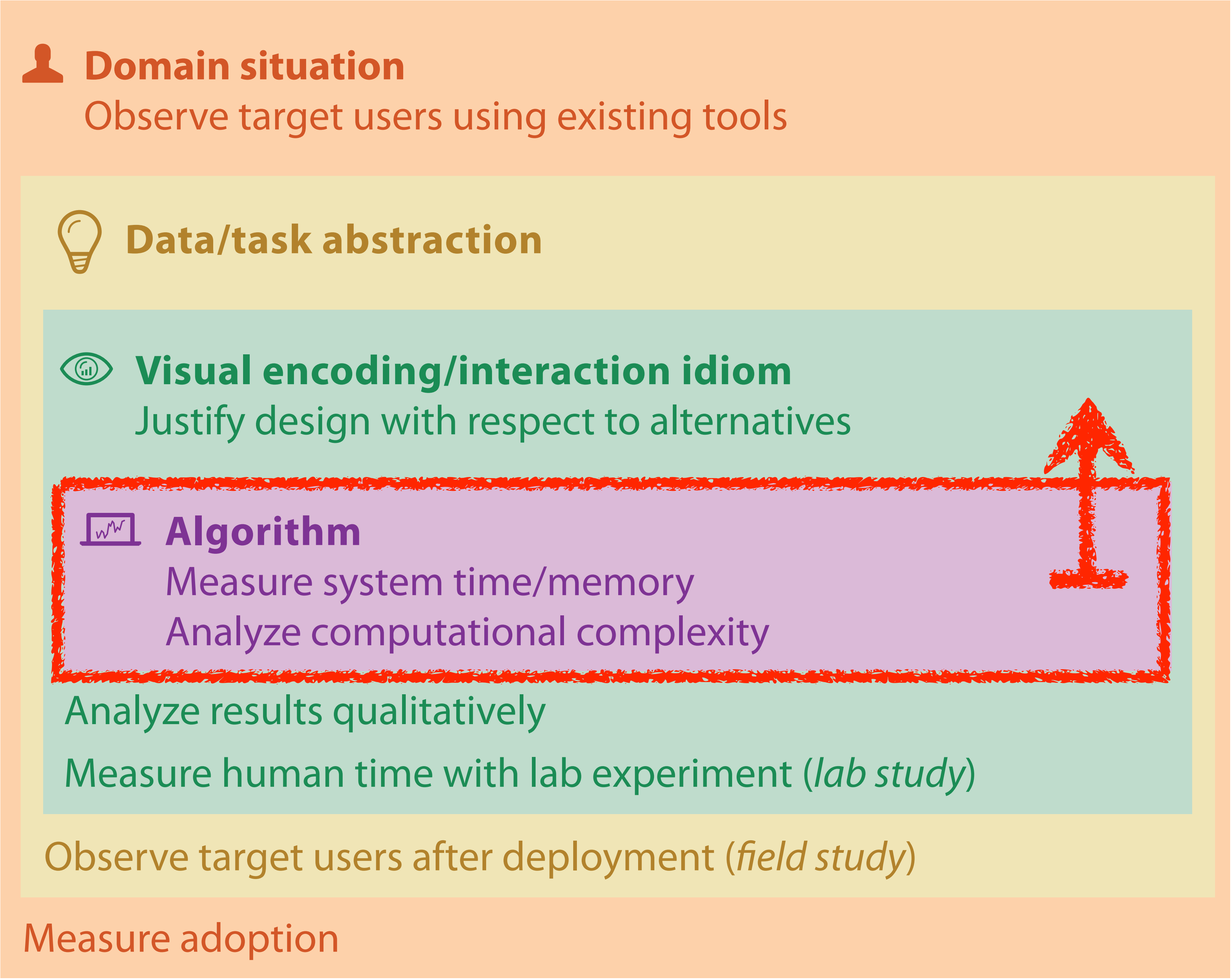
anthropology/
ethnography

design

computer
science

cognitive
psychology

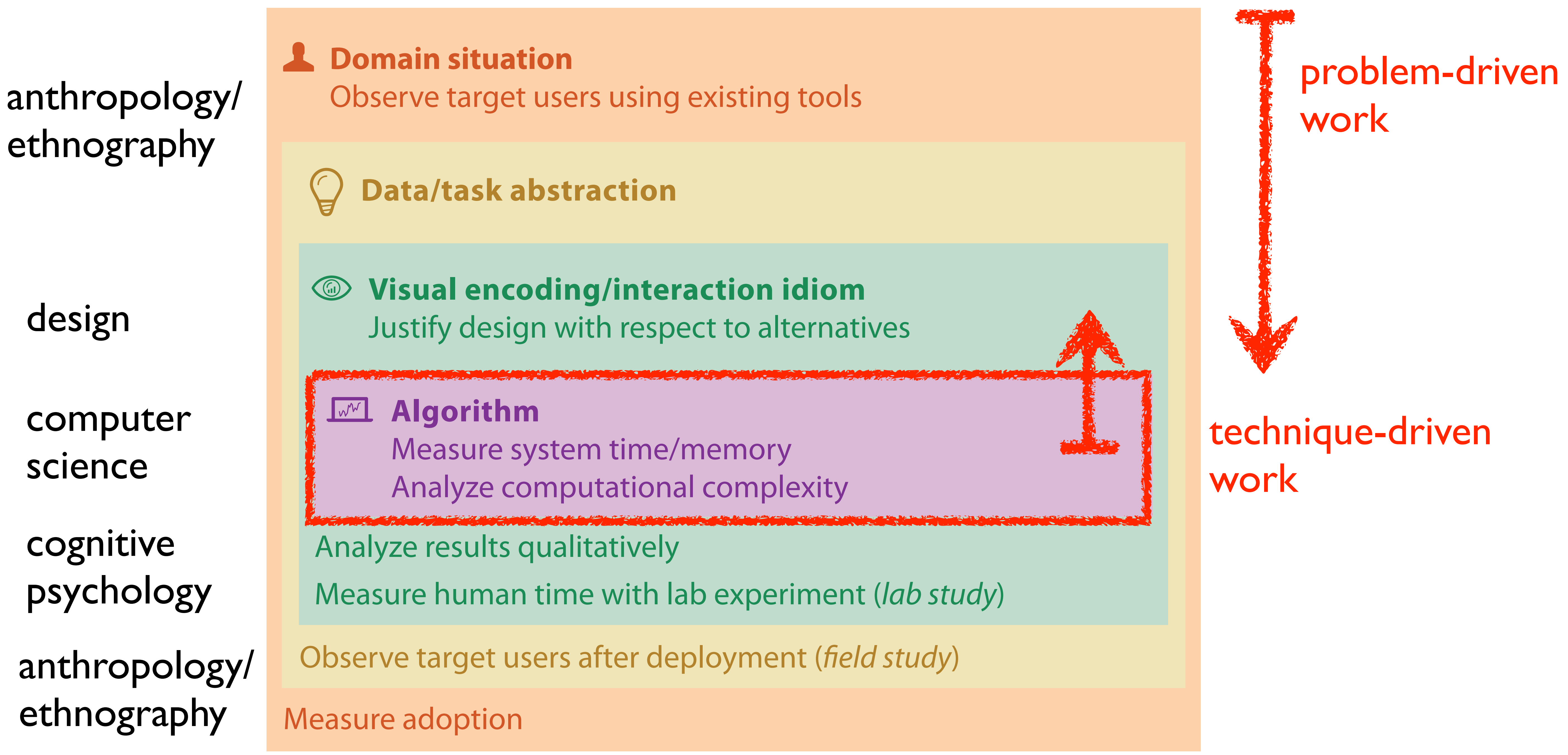
anthropology/
ethnography



technique-driven
work

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

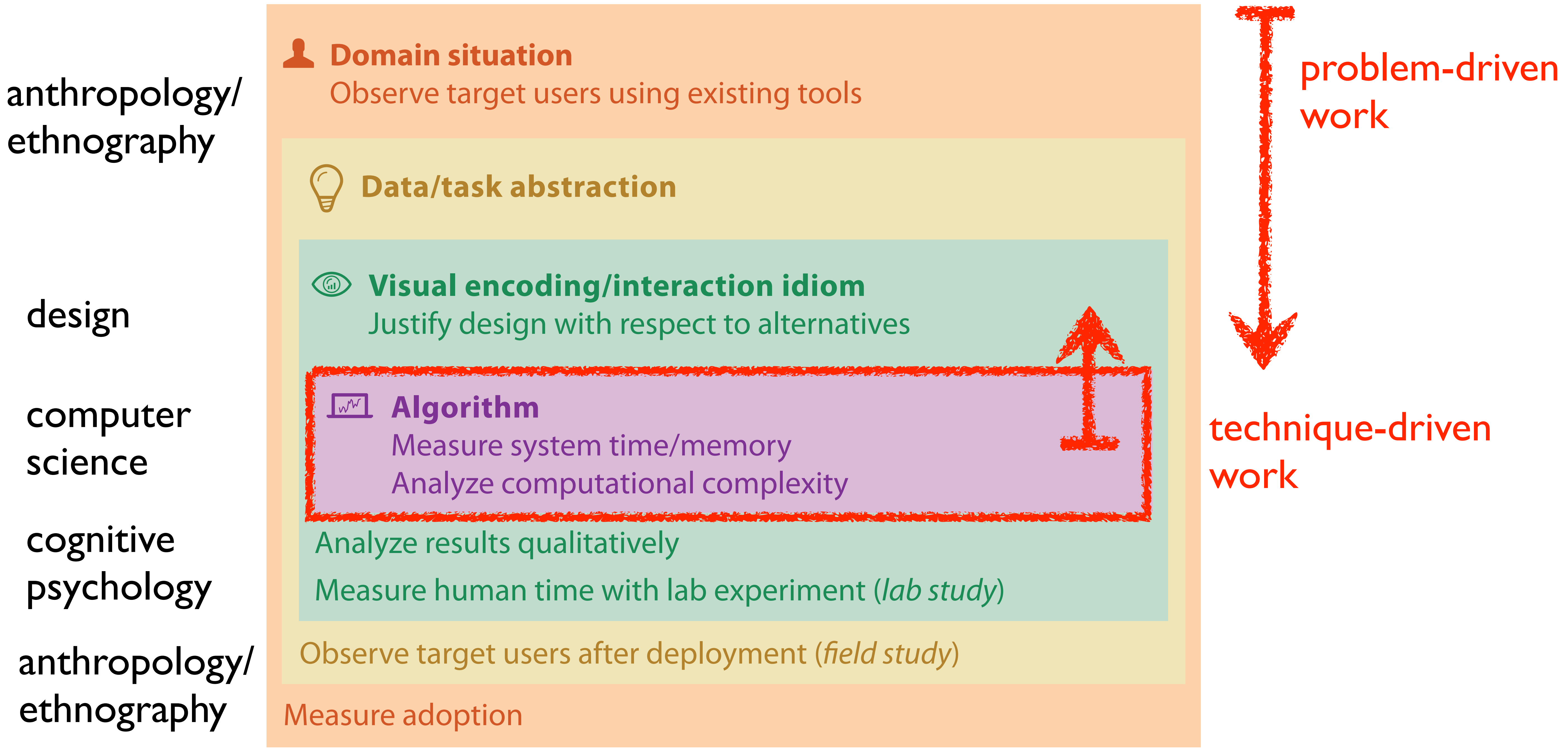
Validation solution: use methods from appropriate fields at each level



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

Validation solution: use methods from appropriate fields at each level

- avoid mismatches between level and validation

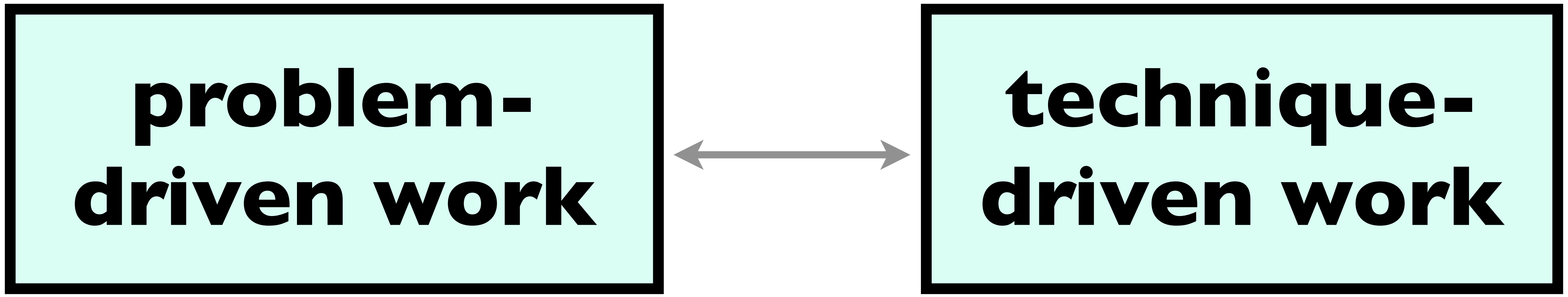


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

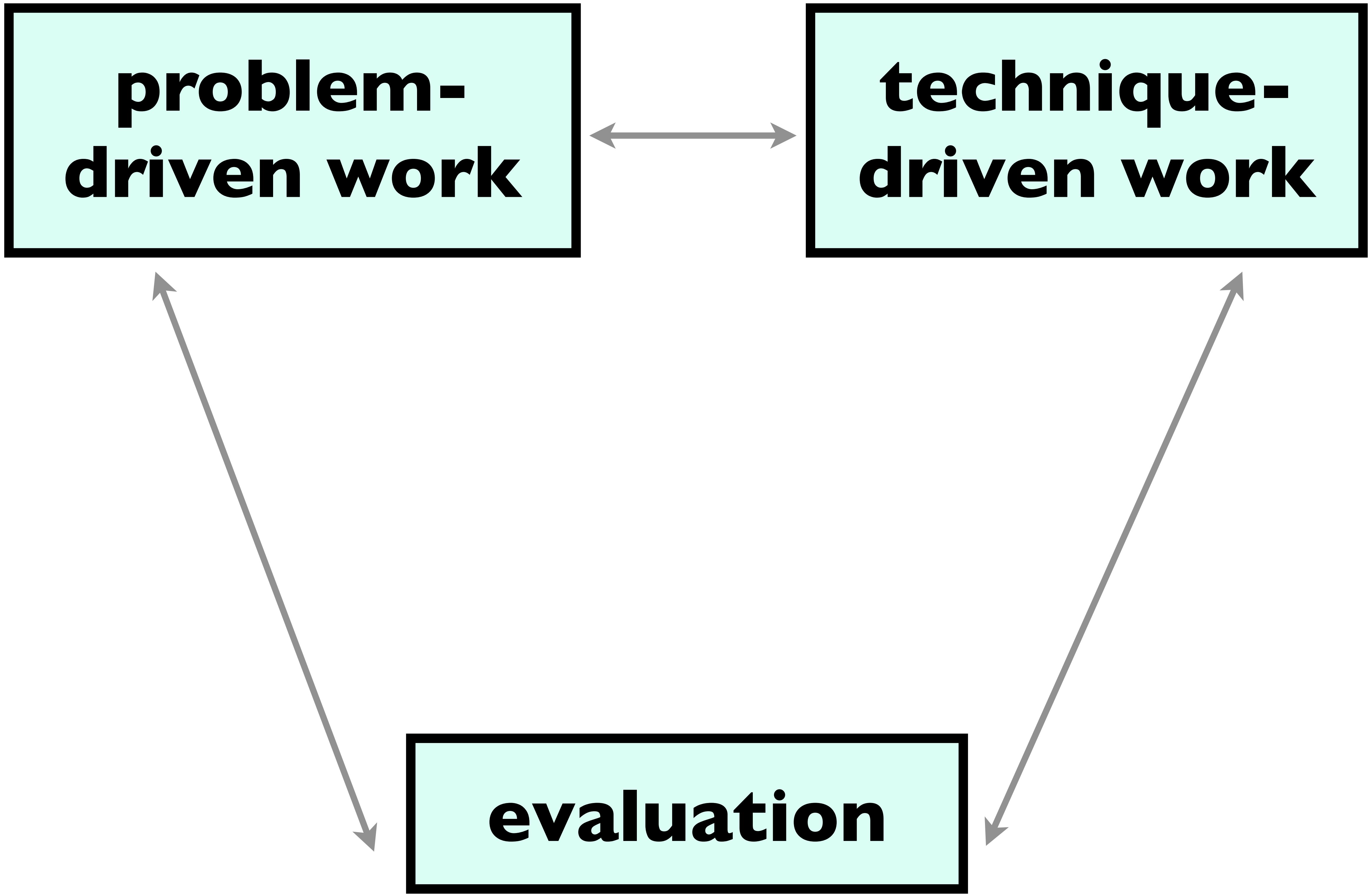
Visualization: Angles of attack

problem-driven work

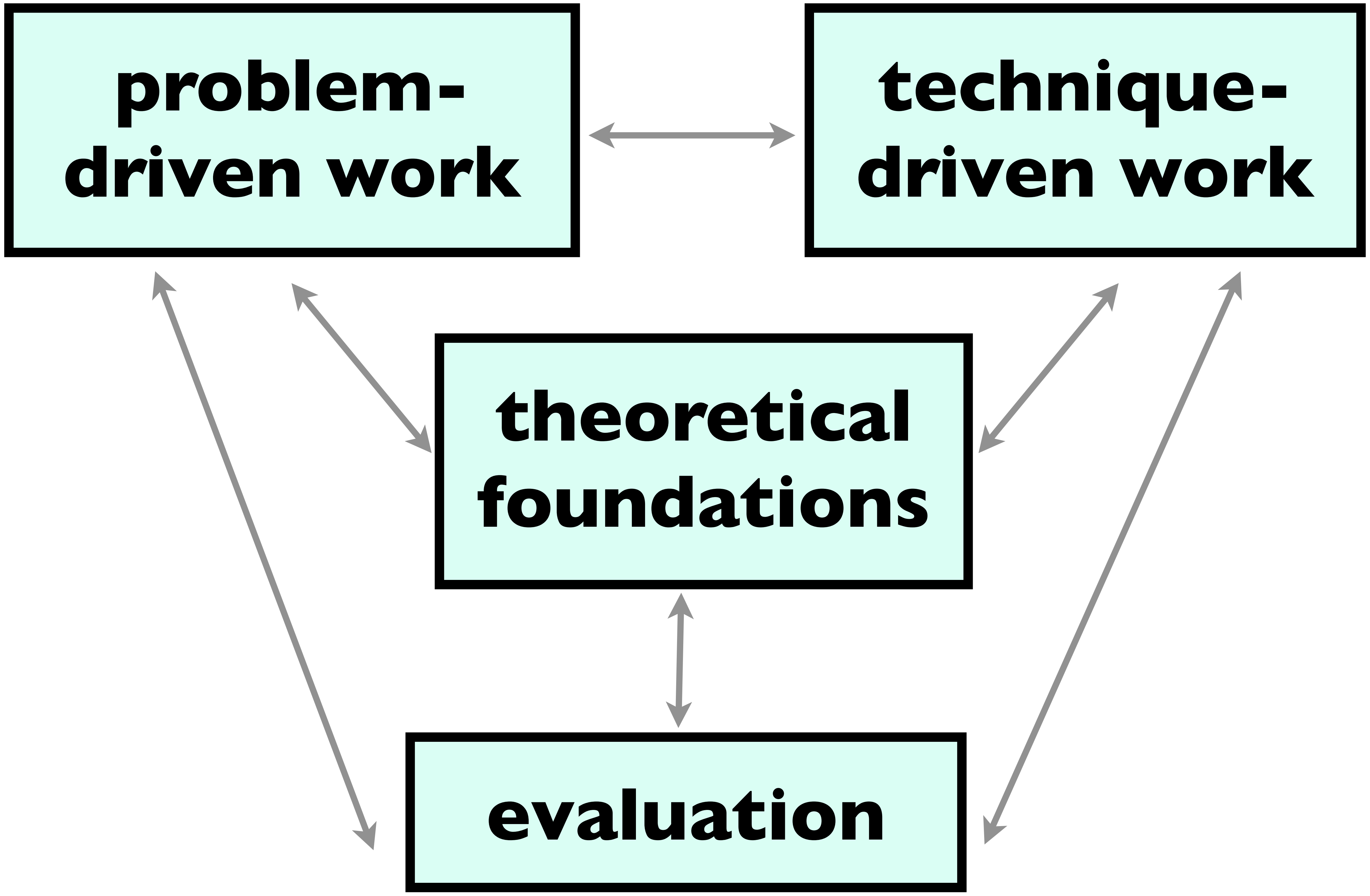
Visualization: Angles of attack



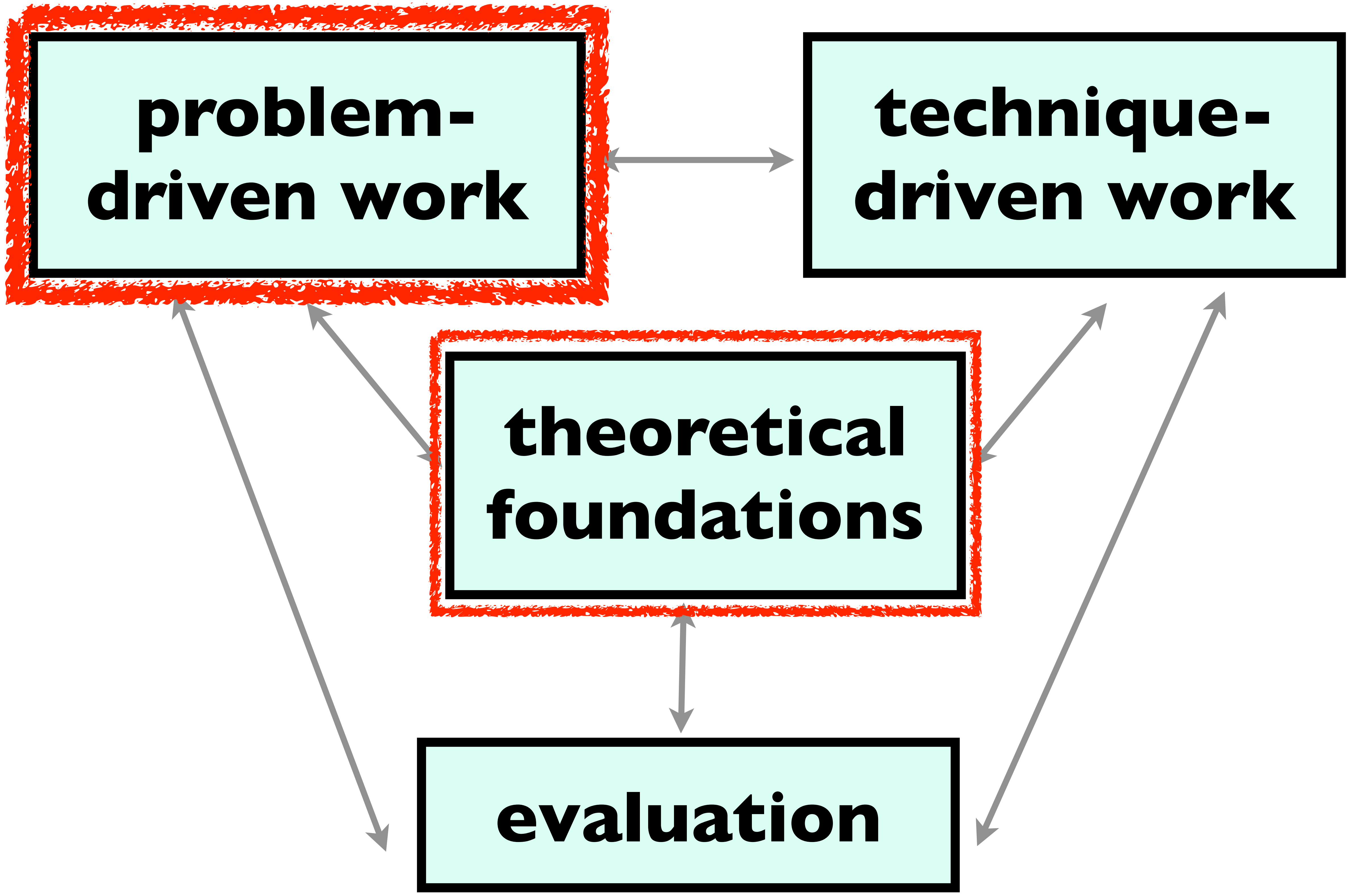
Visualization: Angles of attack



Visualization: Angles of attack



Visualization: Angles of attack



Problem-driven visualization: Design studies

Problem driven visualization: Design studies

"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts…"

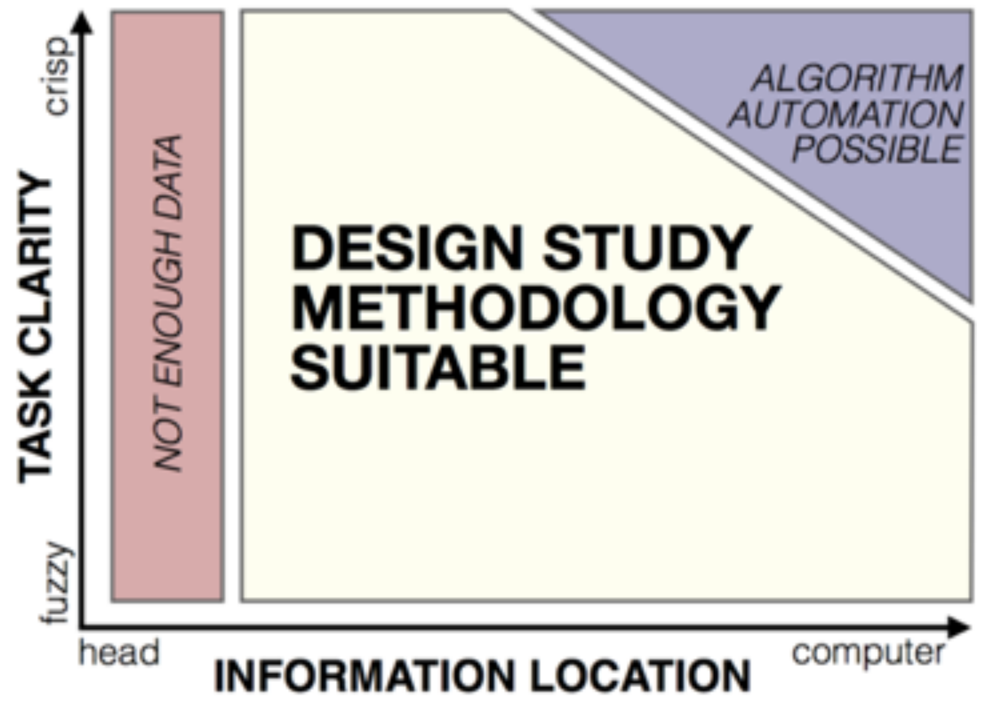
Problem driven visualization: Design studies

"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem…"

Problem driven visualization: Design studies

"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines."

[Design Study Methodology: Reflections from the Trenches and the Stacks.
Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]



Michael Sedlmair

Miriah Meyer

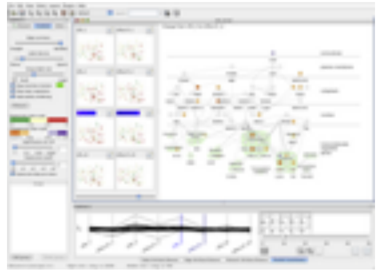
Design Study Methodology

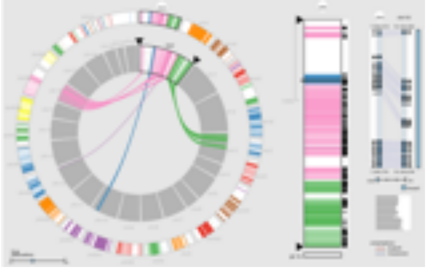
Reflections from the Trenches and from the Stacks

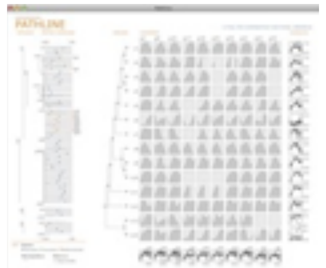
http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/

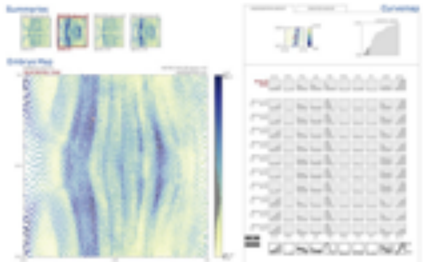
Design Study Methodology: Reflections from the Trenches and from the Stacks.
Sedlmair, Meyer, Munzner. IEEE Trans. Visualization and Computer Graphics 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

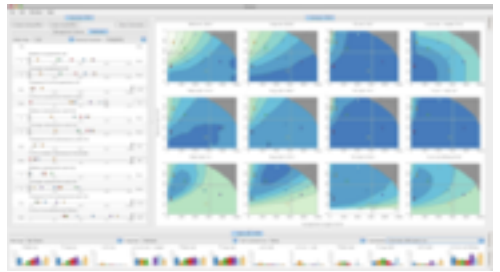
Lessons learned from the trenches: 20+ between us

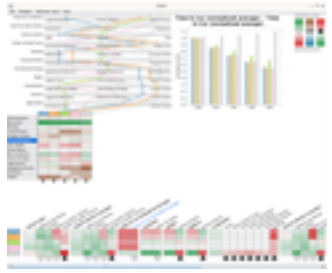

Cerebral
genomics


MizBee
genomics


Pathline
genomics

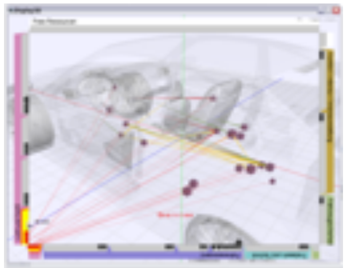

MulteeSum
genomics


Vismon
fisheries management


QuestVis
sustainability

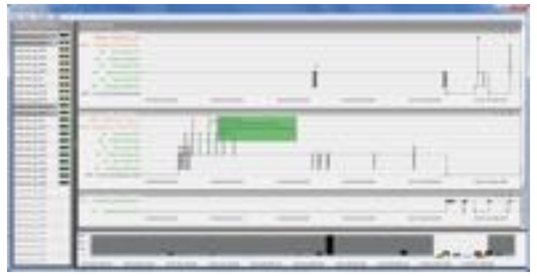

WiKeVis
in-car networks


MostVis
in-car networks


Car-X-Ray
in-car networks


ProgSpy2010
in-car networks

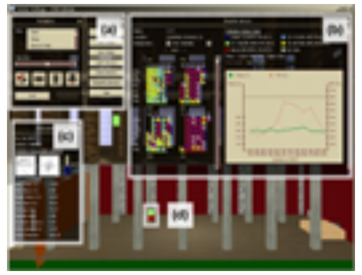

RelEx
in-car networks

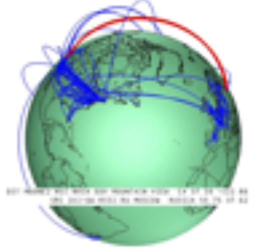

Cardiogram
in-car networks

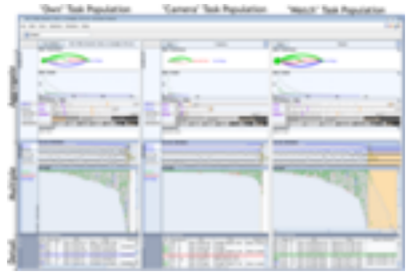

AutobahnVis
in-car networks

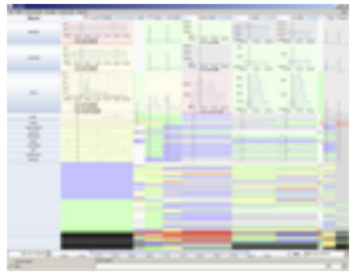

VisTra
in-car networks

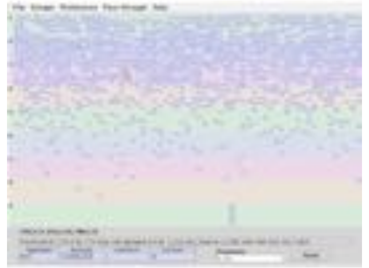

Constellation
linguistics


LibVis
cultural heritage

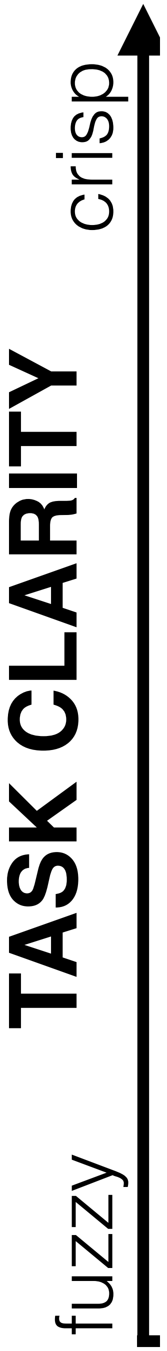

Caidants
multicast


SessionViewer
web log analysis


LiveRAC
server hosting

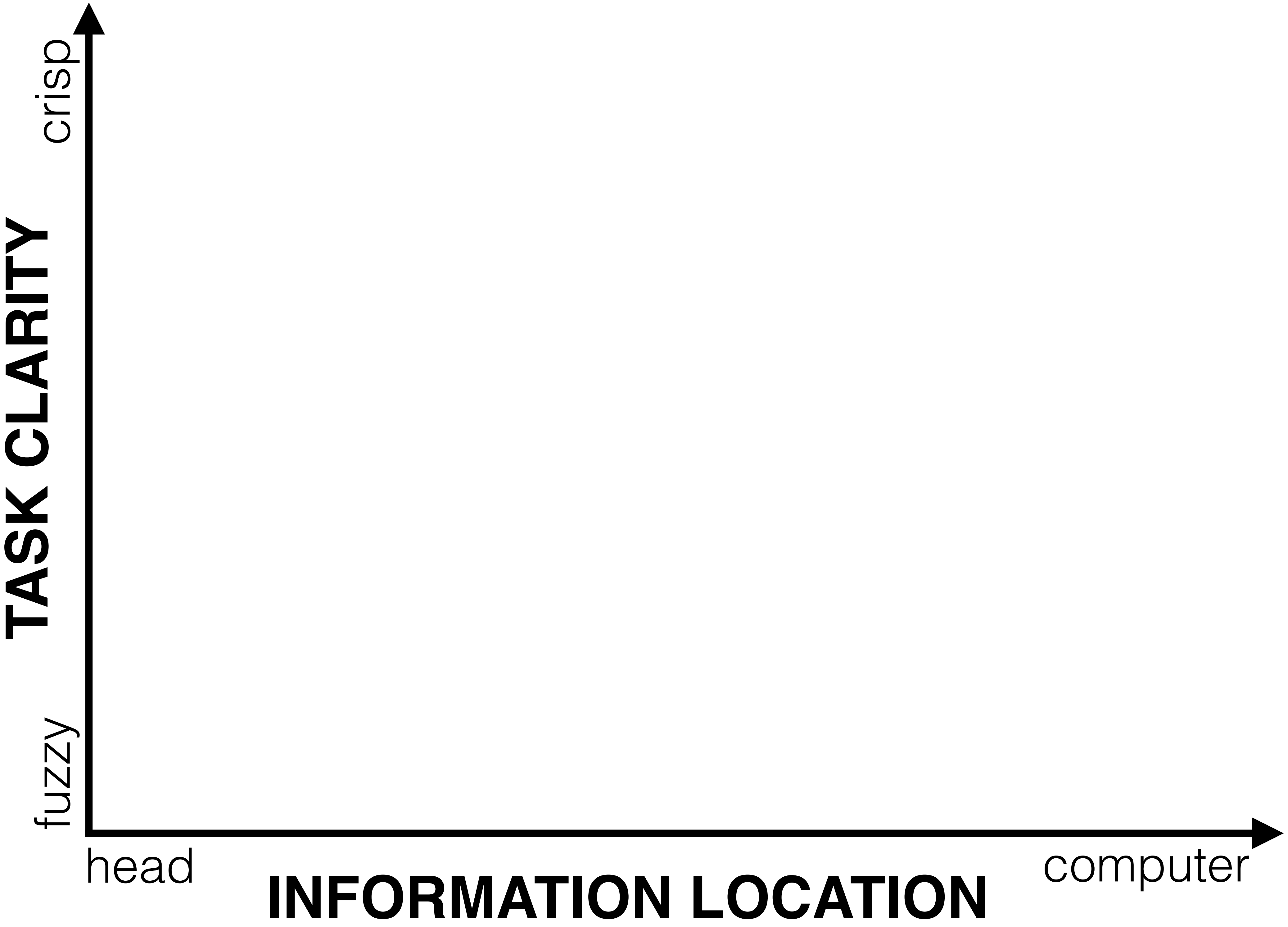

PowerSetViewer
data mining

Design study methodology: definitions



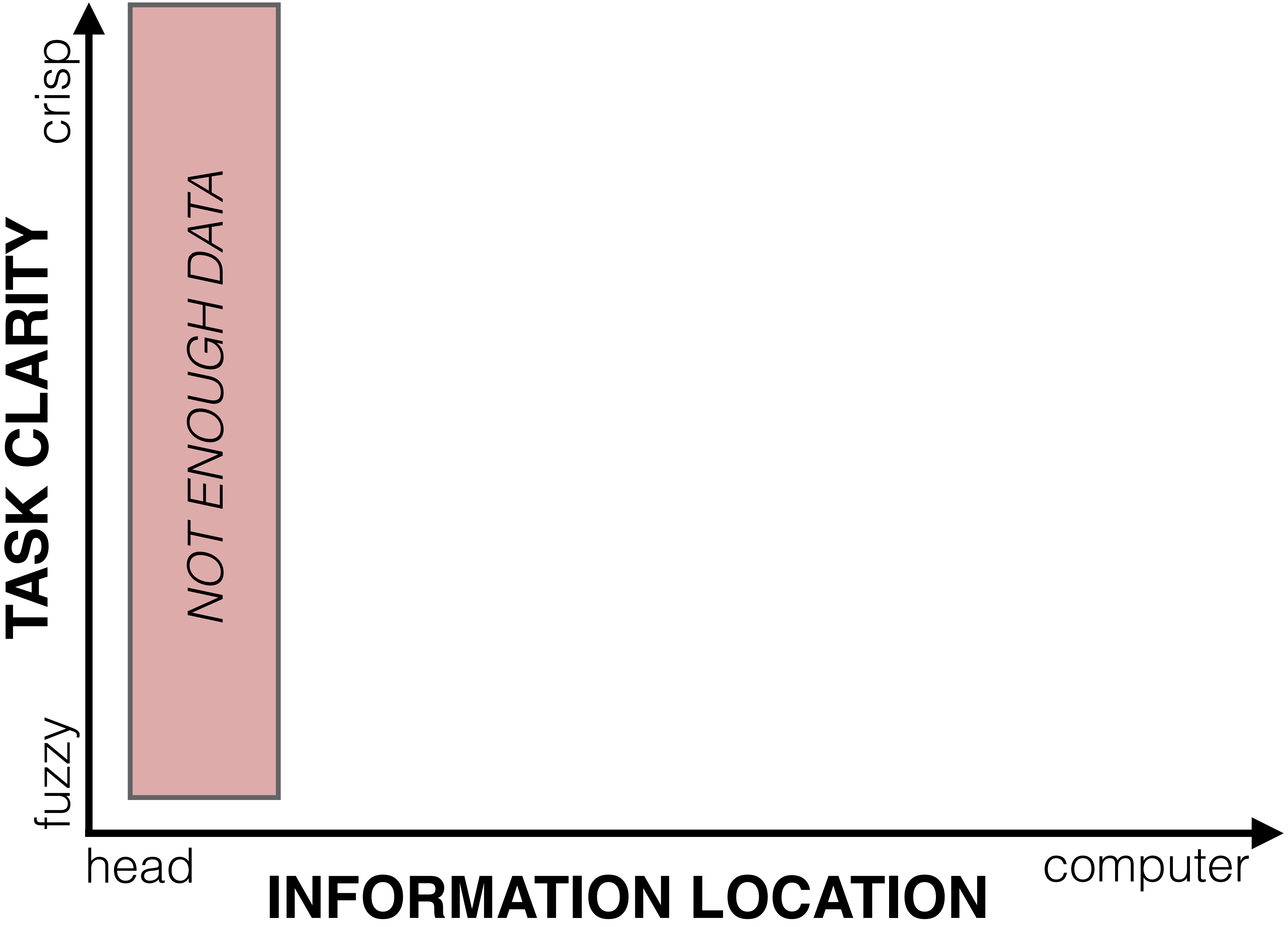
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



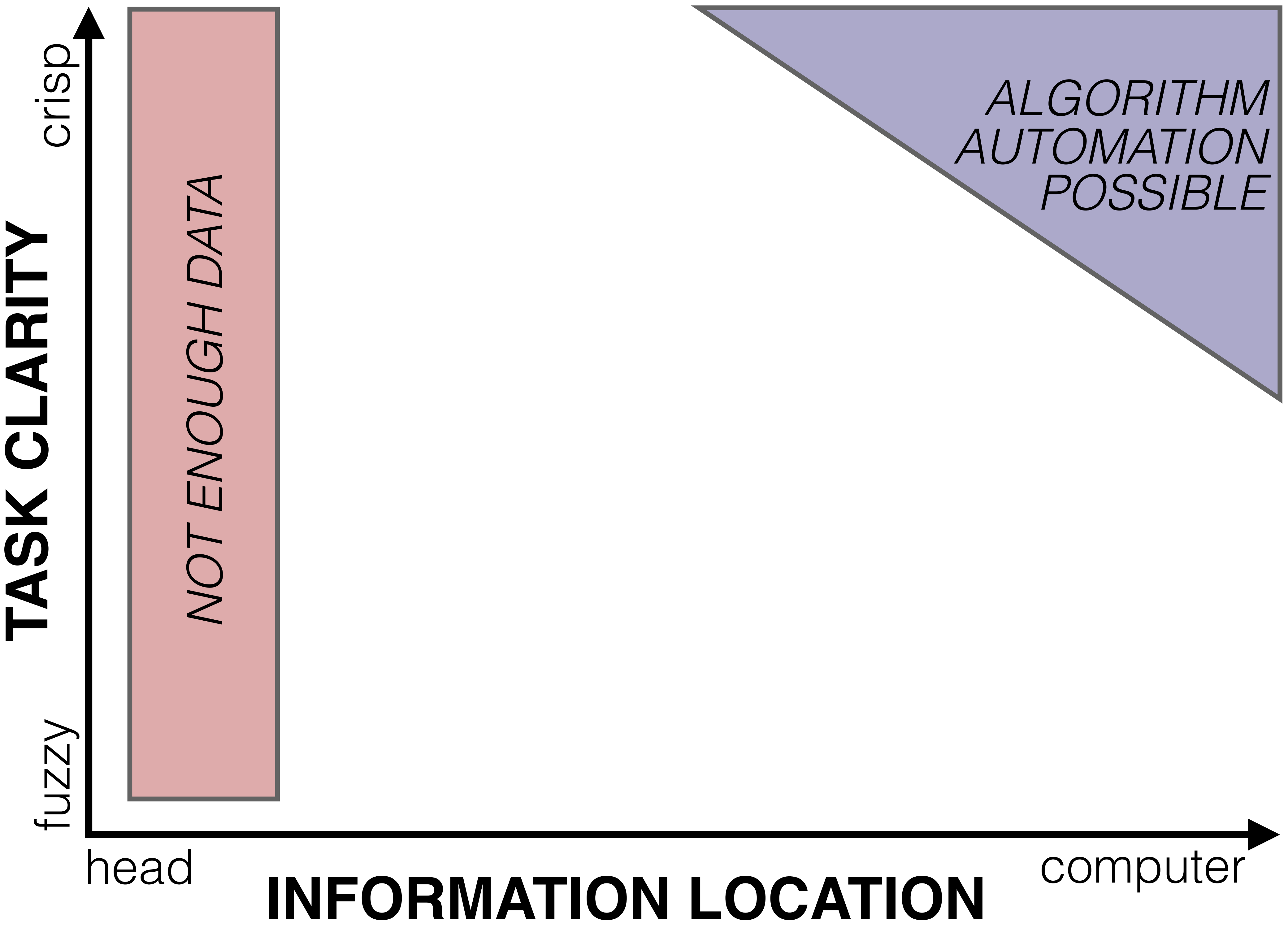
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



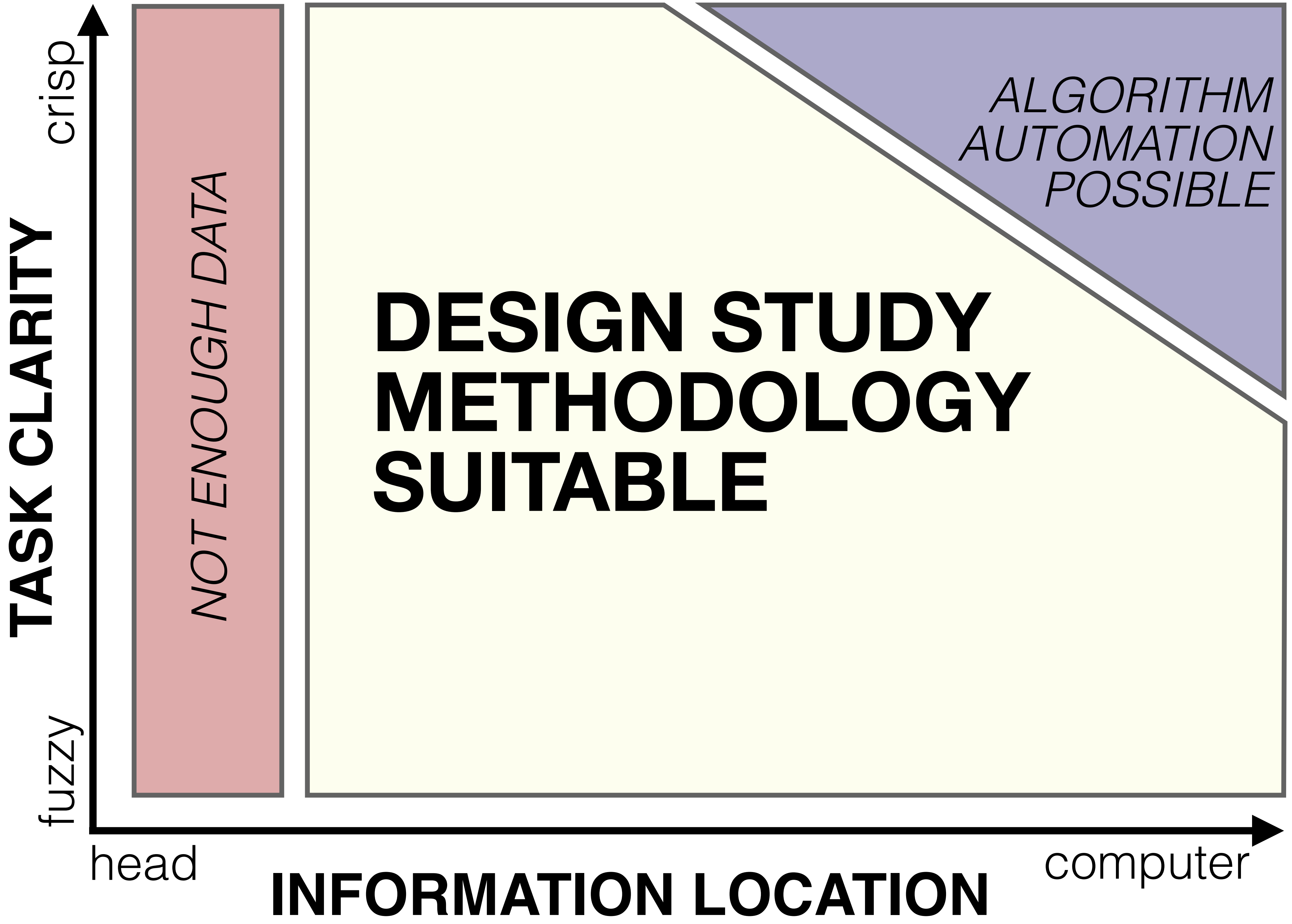
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



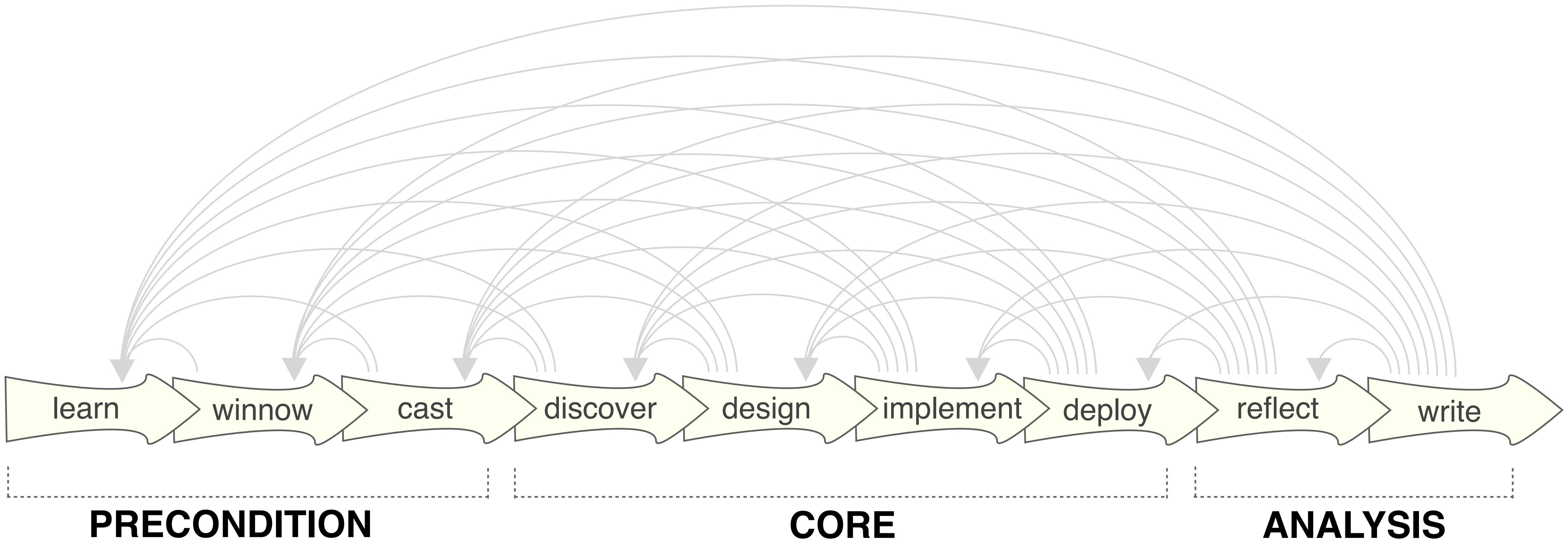
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

Design study methodology: definitions



[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

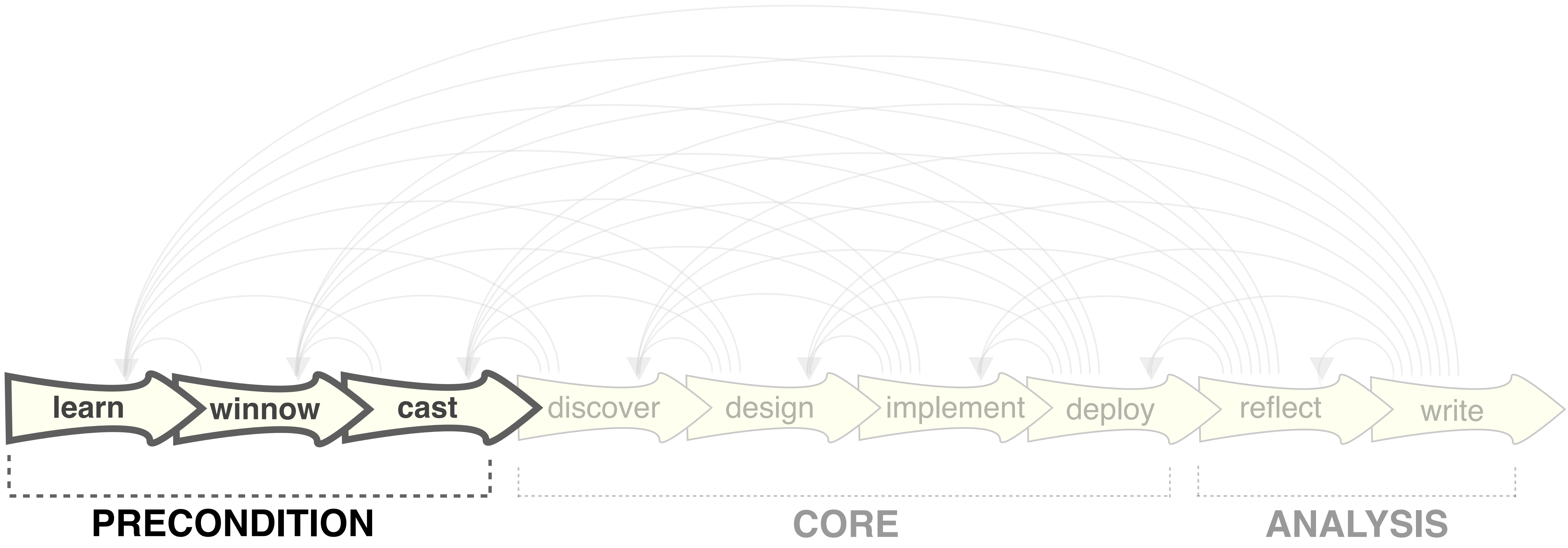
9-stage framework



[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

9-stage framework

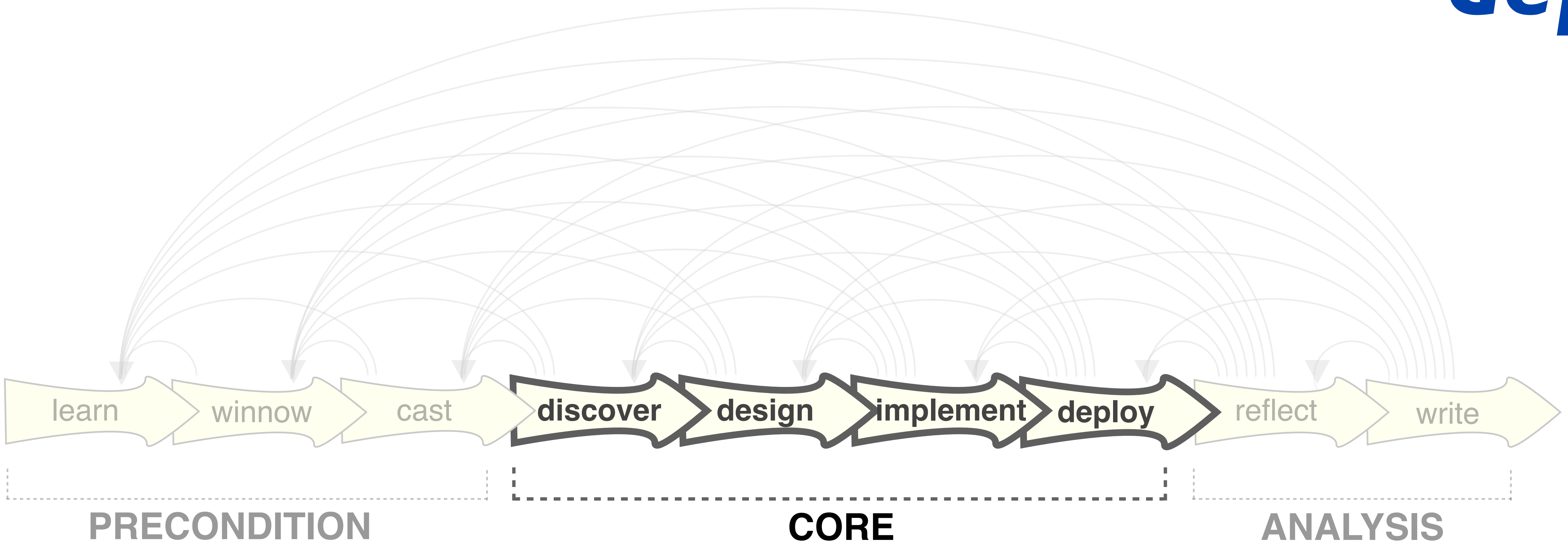
learn
winnow
cast



[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

9-stage framework

discover
design
implement
deploy

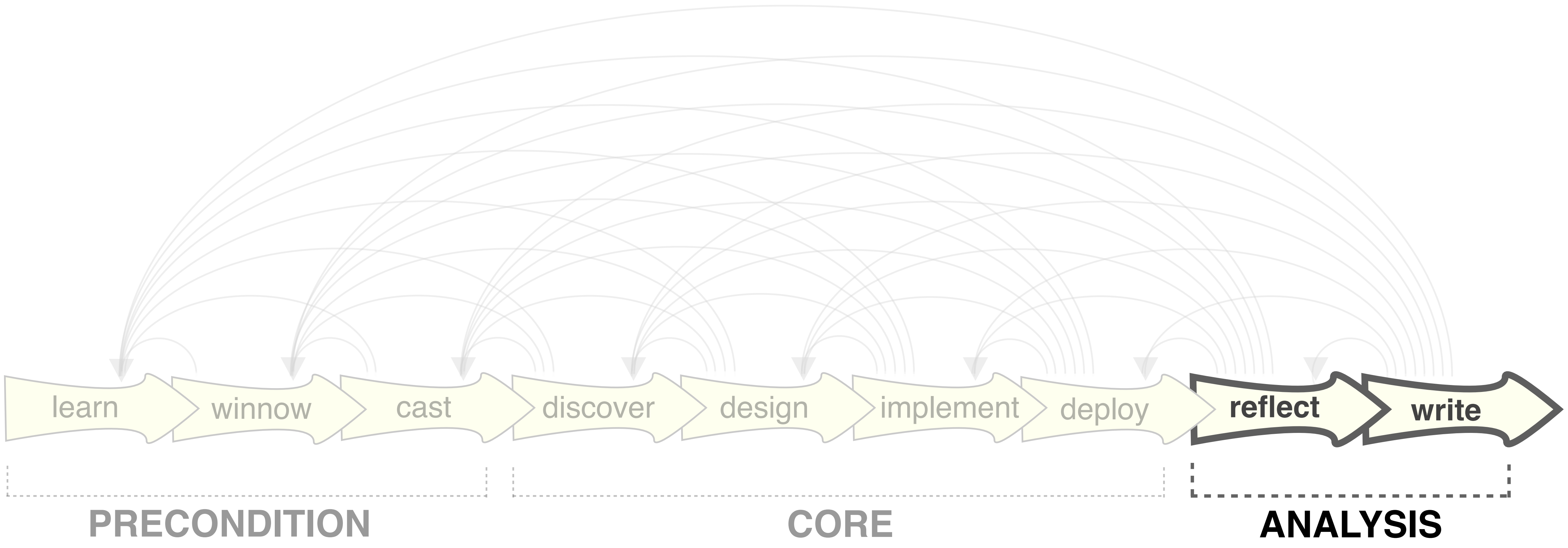


[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

9-stage framework

reflect
write

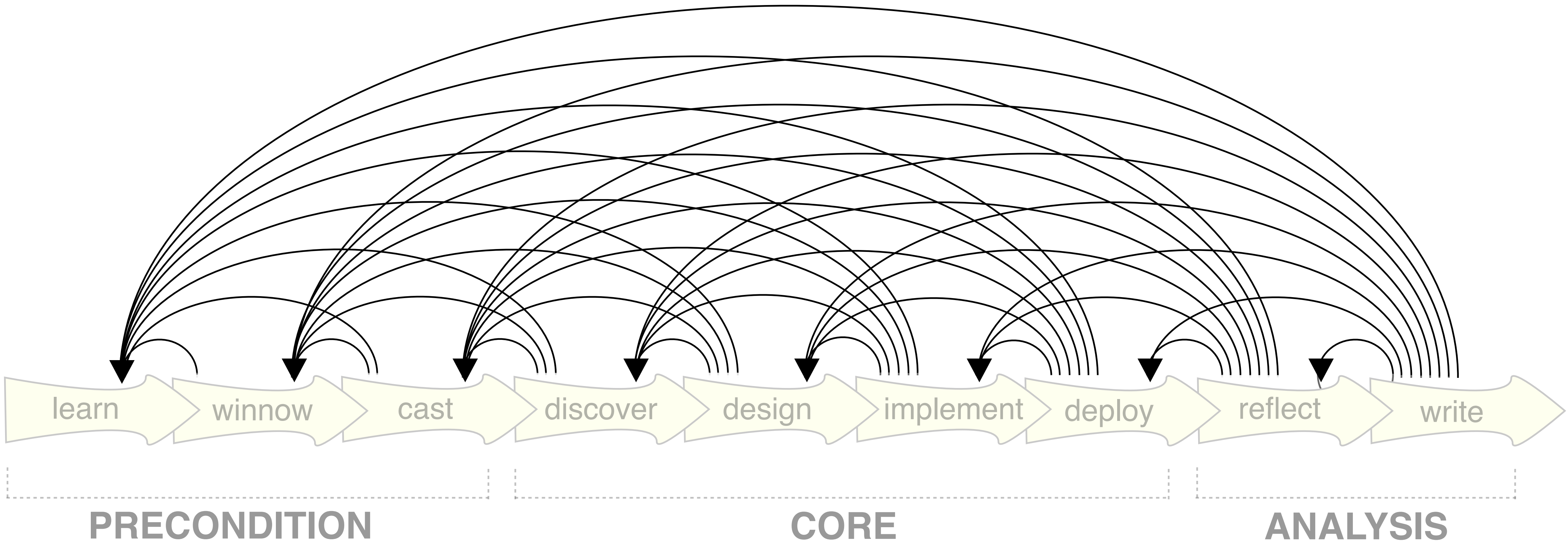
- guidelines: confirm, refine, reject, propose



[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

9-stage framework

iterative



[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

32 pitfalls & how to avoid them

[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

32 pitfalls & how to avoid them

PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaboration with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow
PF-10	no real/important/recurring task	winnow
PF-11	no rapport with collaborators	winnow
PF-12	not identifying front line analyst and gatekeeper before start	cast
PF-13	assuming every project will have the same role distribution	cast
PF-14	mistaking fellow tool builders for real end users	cast

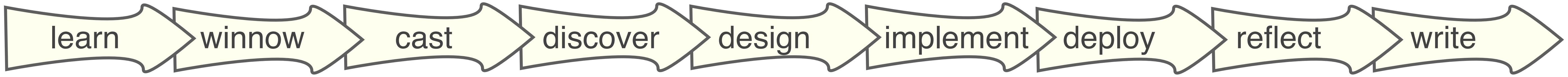
[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

32 pitfalls & how to avoid them

PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaboration with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow
PF-10	no real/important/recurring task	winnow
PF-11	no rapport with collaborators	winnow
PF-12	not identifying front line analyst and gatekeeper before start	cast
PF-13	assuming every project will have the same role distribution	cast
PF-14	mistaking fellow tool builders for real end users	cast
PF-15	ignoring practices that currently work well	discover
PF-16	expecting *just talking* or *fly on wall* to work	discover
PF-17	experts focusing on visualization design vs. domain problem	discover
PF-18	learning their problems/language: too little / too much	discover
PF-19	abstraction: too little	design
PF-20	premature design commitment: consideration space too small	design
PF-21	mistaking technique-driven for problem-driven work	design
PF-22	nonrapid prototyping	implement
PF-23	usability: too little / too much	implement
PF-24	premature end: insufficient deploy time built into schedule	deploy
PF-25	usage study not case study: non-real task/data/user	deploy
PF-26	*liking* necessary but not sufficient for validation	deploy

[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

32 pitfalls & how to avoid them

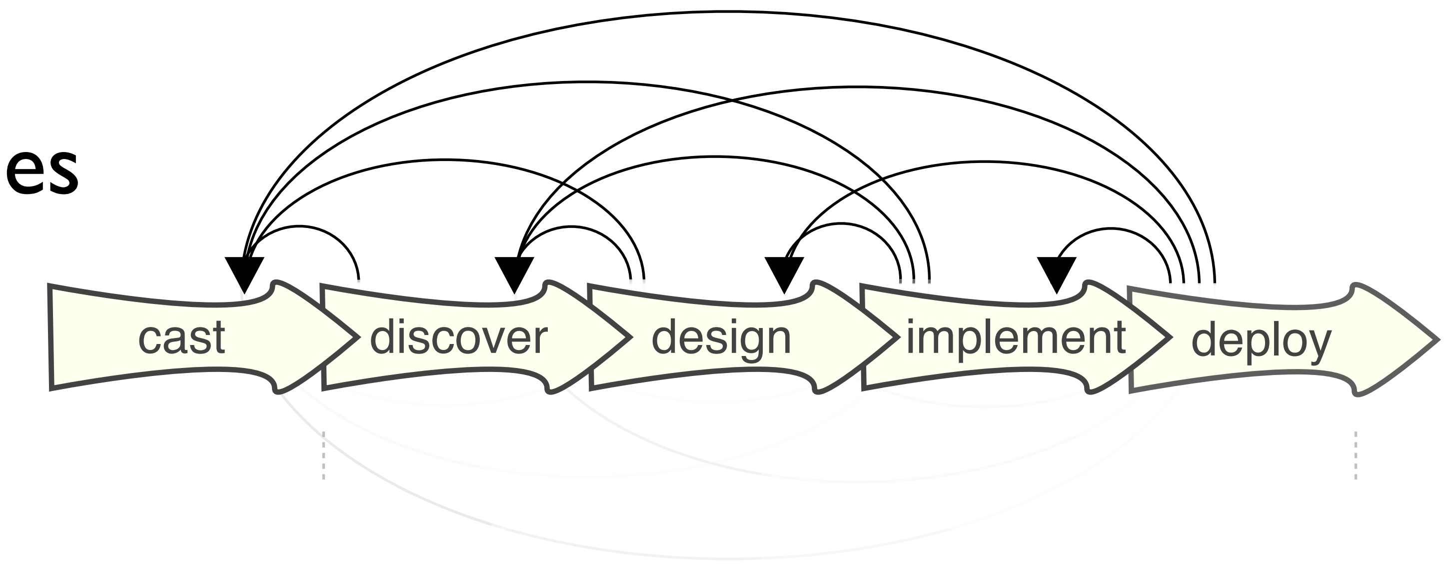


PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaboration with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow
PF-10	no real/important/recurring task	winnow
PF-11	no rapport with collaborators	winnow
PF-12	not identifying front line analyst and gatekeeper before start	cast
PF-13	assuming every project will have the same role distribution	cast
PF-14	mistaking fellow tool builders for real end users	cast
PF-15	ignoring practices that currently work well	discover
PF-16	expecting *just talking* or *fly on wall* to work	discover
PF-17	experts focusing on visualization design vs. domain problem	discover
PF-18	learning their problems/language: too little / too much	discover
PF-19	abstraction: too little	design
PF-20	premature design commitment: consideration space too small	design
PF-21	mistaking technique-driven for problem-driven work	design
PF-22	nonrapid prototyping	implement
PF-23	usability: too little / too much	implement
PF-24	premature end: insufficient deploy time built into schedule	deploy
PF-25	usage study not case study: non-real task/data/user	deploy
PF-26	*liking* necessary but not sufficient for validation	deploy
PF-27	failing to improve guidelines: confirm, refine, reject, propose	reflect
PF-28	insufficient writing time built into schedule	write
PF-29	no technique contribution ≠ good design study	write
PF-30	too much domain background in paper	write
PF-31	story told chronologically vs. focus on final results	write
PF-32	premature end: win race vs. practice music for debut	write

[Design Study Methodology: Reflections from the Trenches and the Stacks. Sedlmair, Meyer & Munzner. IEEE TVCG 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).]

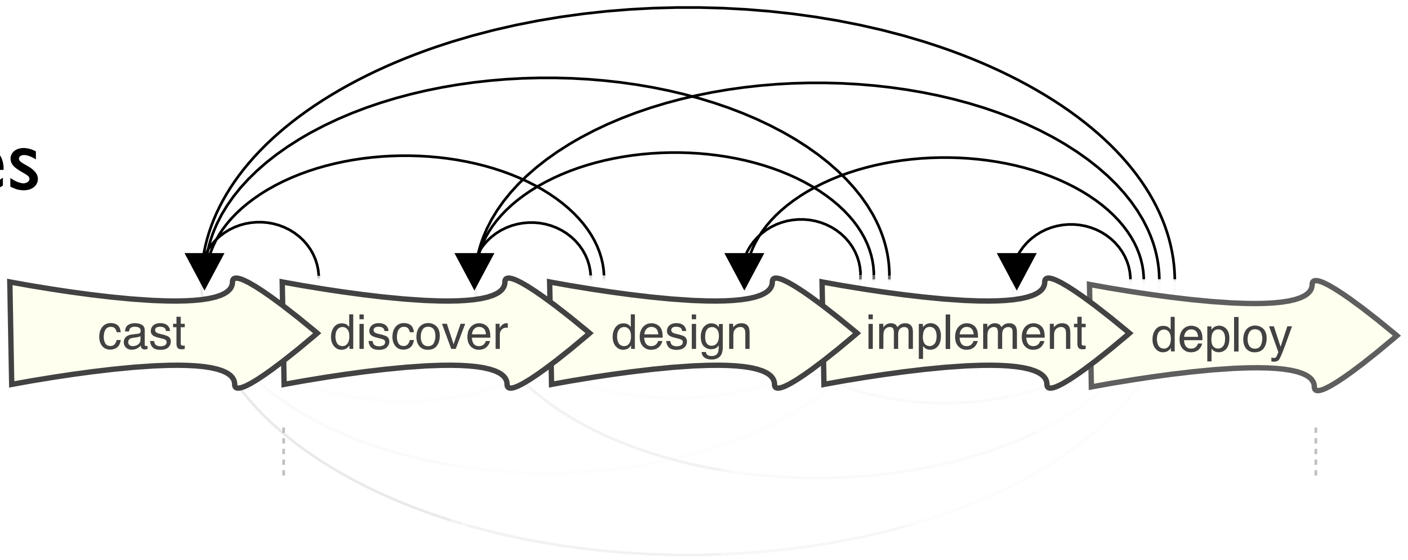
Design studies & user-centered design

- user-centered design: well-known HCI methodology
  - iterative refinement & deployment
  - evaluation through case studies & field studies

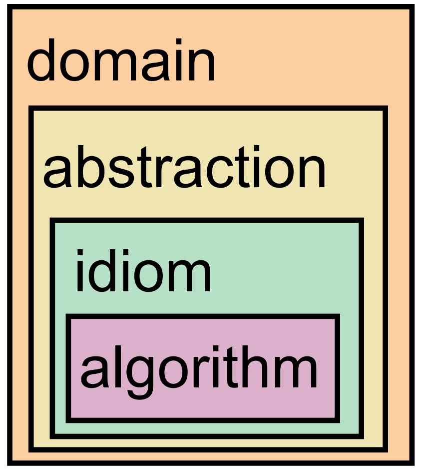


Design studies & user-centered design

- user-centered design: well-known HCI methodology
  - iterative refinement & deployment
  - evaluation through case studies & field studies



- what's specific to visualization?
  - discovering task and data **abstractions**
  - designing visual encoding & interaction **idioms** that map to abstractions



Three case studies of problem-driven work

- e-commerce

- facilities management

- biology

e commerce by shashank singh from the Noun Project Biology by tezar tantular from the Noun Project Business by Colourcreatype from the Noun Project

Three case studies of problem-driven work

- e-commerce

- facilities management

- biology

e commerce by shashank singh from the Noun Project Biology by tezar tantular from the Noun Project Business by Colourcreatype from the Noun Project

Kim
Dextras-Romagnino

Segmentifier

Interactive Refinement of Clickstream Data

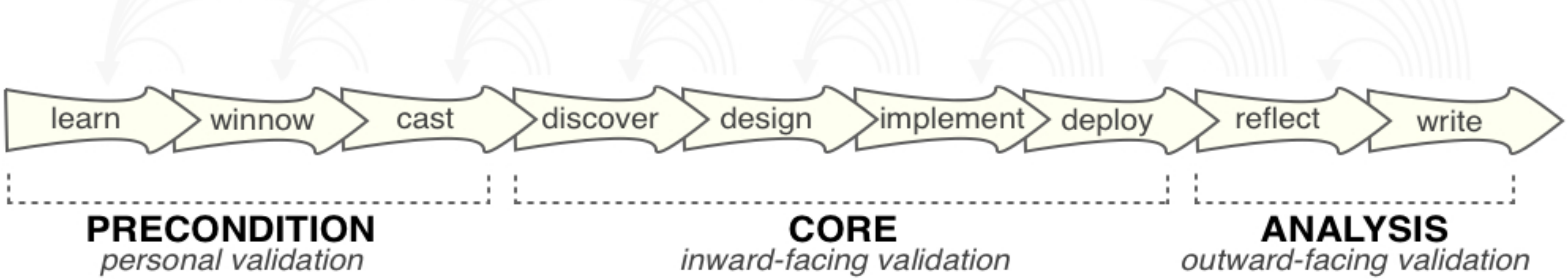
http://www.cs.ubc.ca/labs/imager/tr/2019/segmentifier

Segmentifier: Interactive Refinement of Clickstream Data.
Dextras-Romagnino and Munzner. Computer Graphics Forum (Proc. EuroVis 2019) 38(3):623–634 2019

E-commerce: mobile apps for large companies

Process: Design Study Methodology

- **Precondition Phase** (5 months) : interviews with 12 employees
- **Core Phase** (11 months): Iterative design and implementation
- **Analysis Phase** (3 months): Reflect and write



What are the **Data and Task Abstractions** for *Clickstream Data Analysis?*

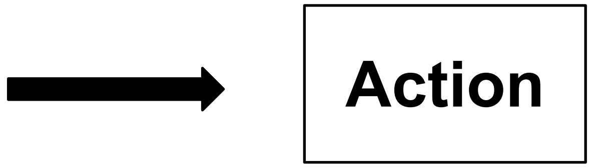
Clickstream Data

Clickstream Analysis Tasks

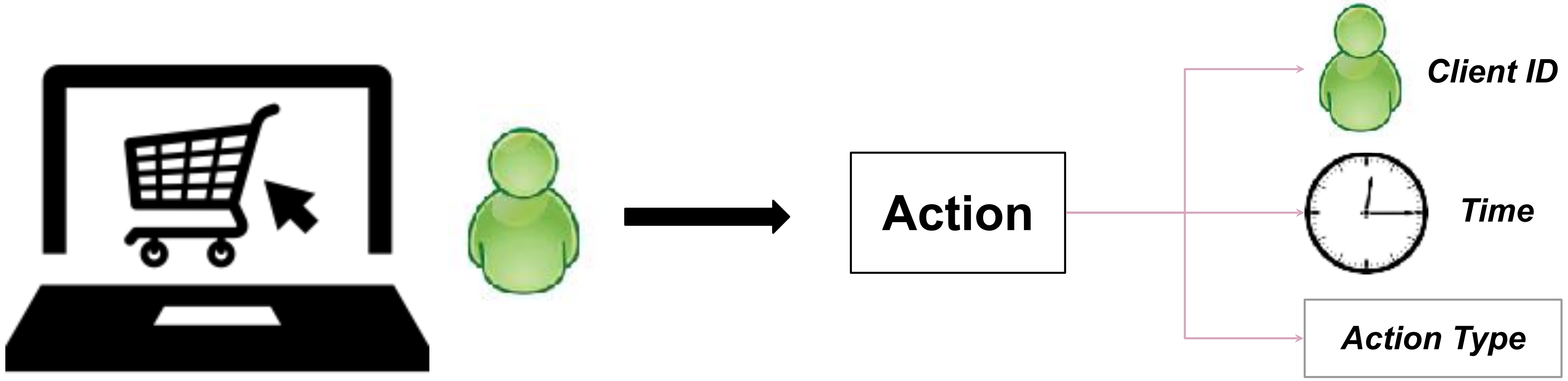
Segmentifier Analysis Model

What is ***Clickstream Data***?

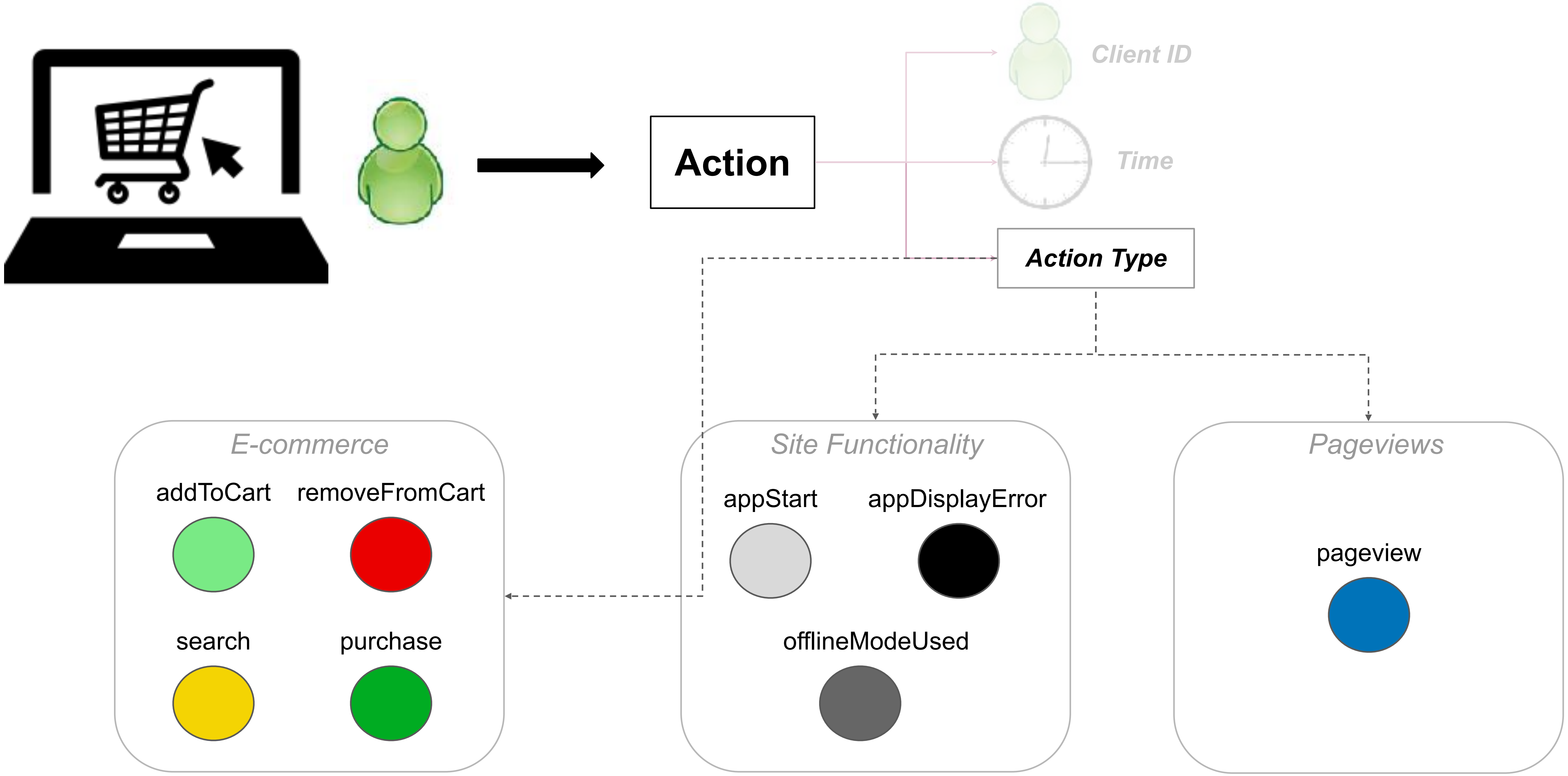
Data: *Actions*



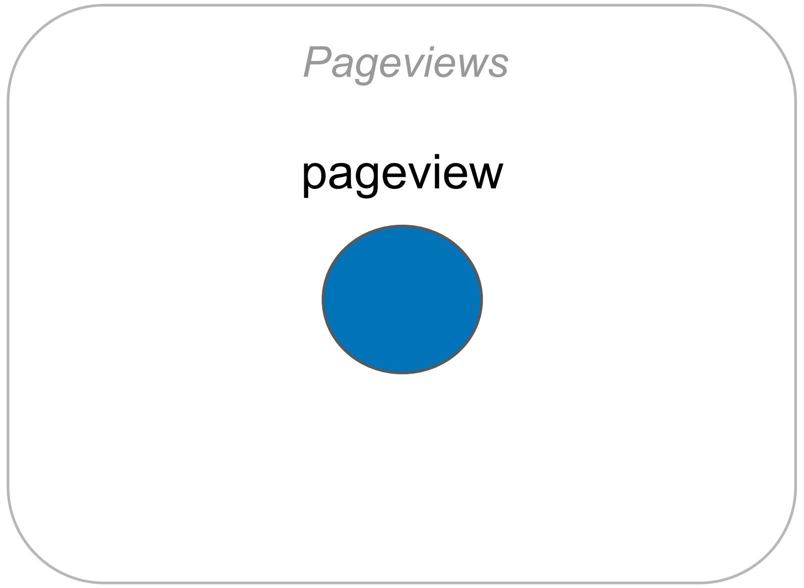
Data: *Action Attributes*



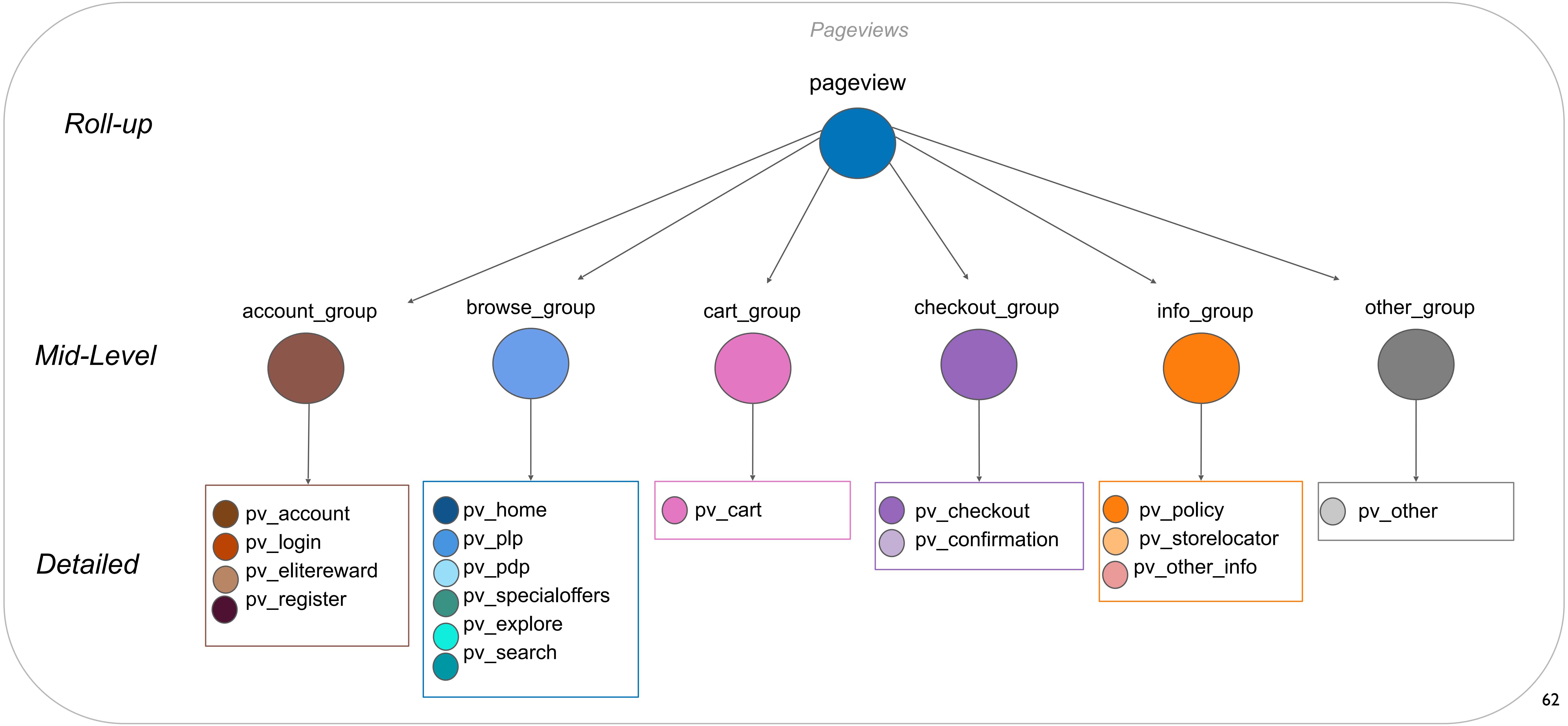
Data: *Action Types*



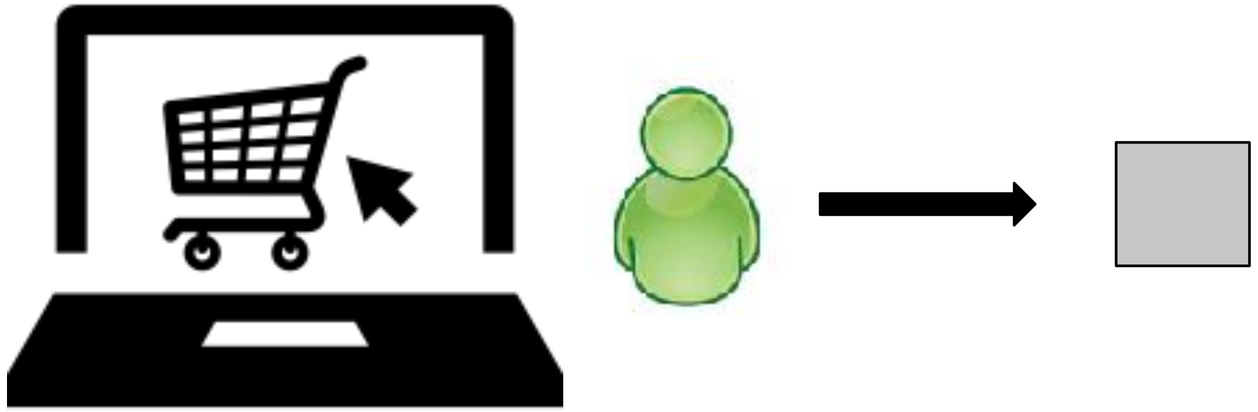
Action Hierarchy

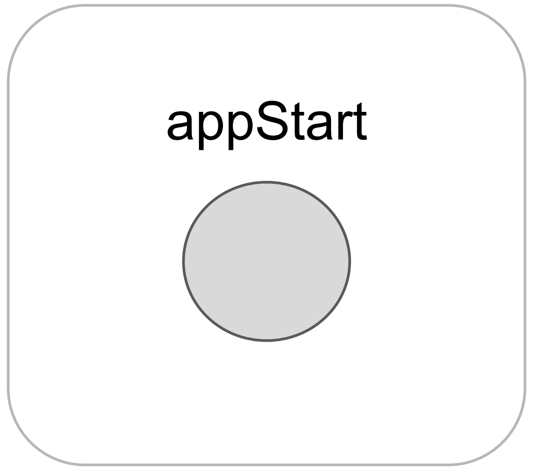


Action Hierarchy

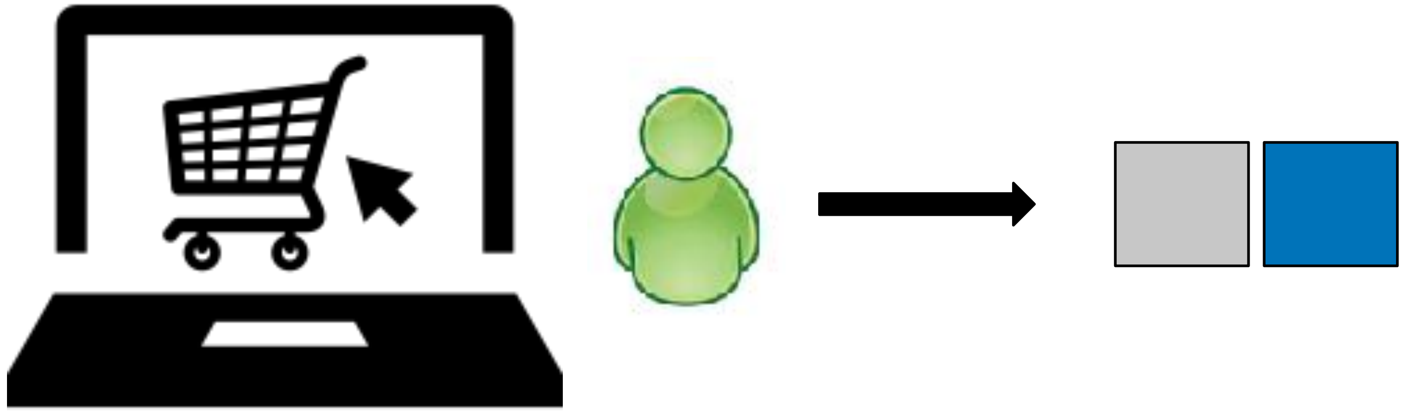


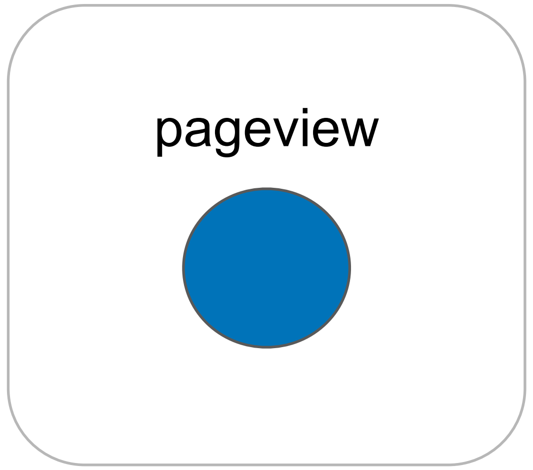
Data: *Sequences*



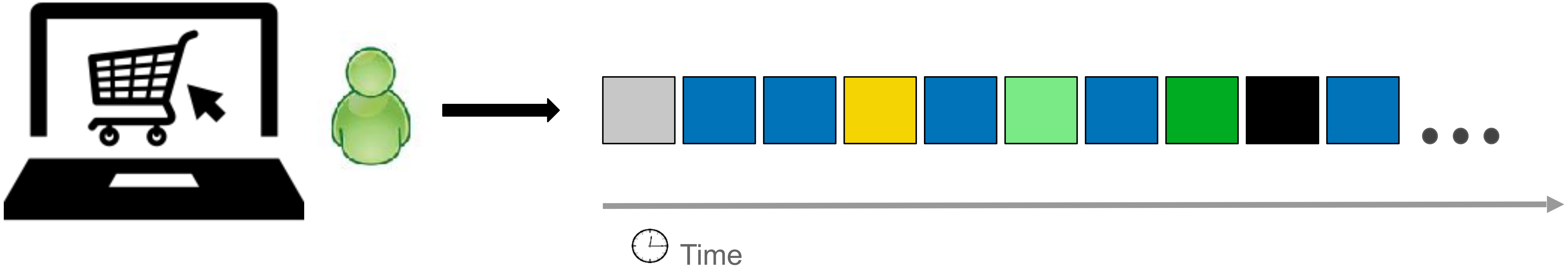


Data: *Sequences*



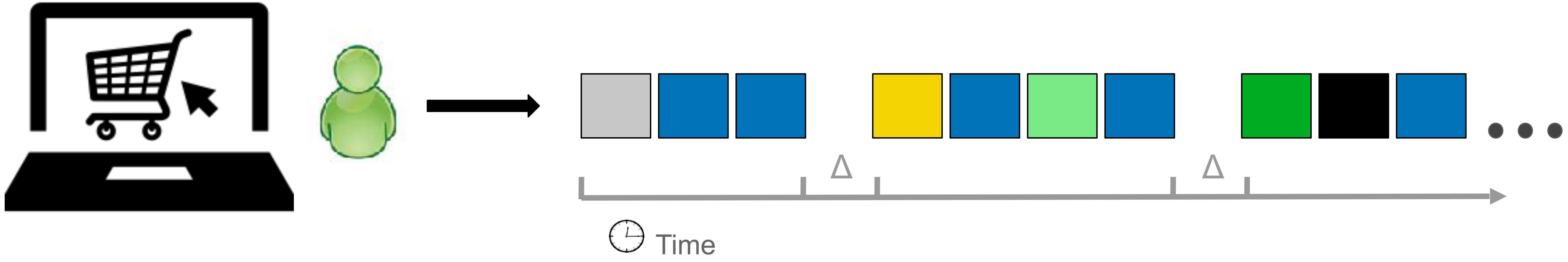


Data: *Client Sequences*



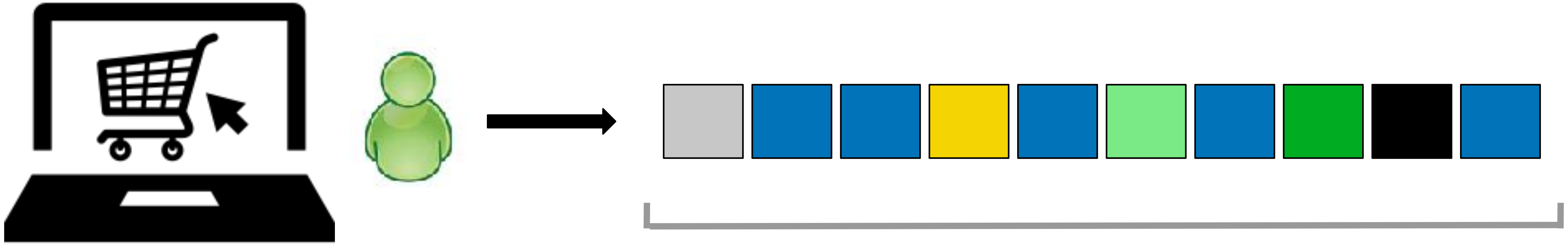
Client Sequences: all actions performed by a single user

Data: *Session Sequences*



Session Sequences: all actions performed by a single user within a defined amount of time (Δ) from each other.
Δ is usually 30 min.

Data: *Sequence Attributes*

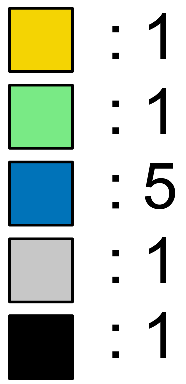


Start time

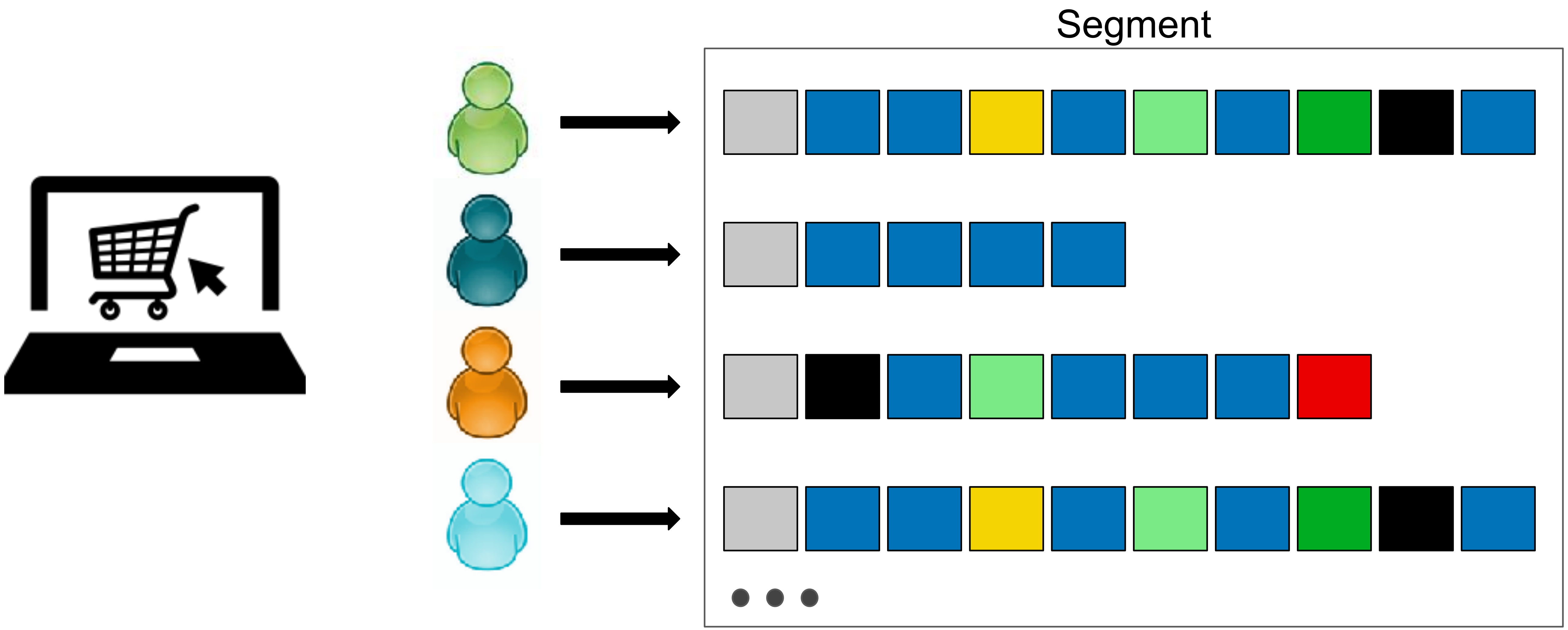
End time

Duration

Action Counts

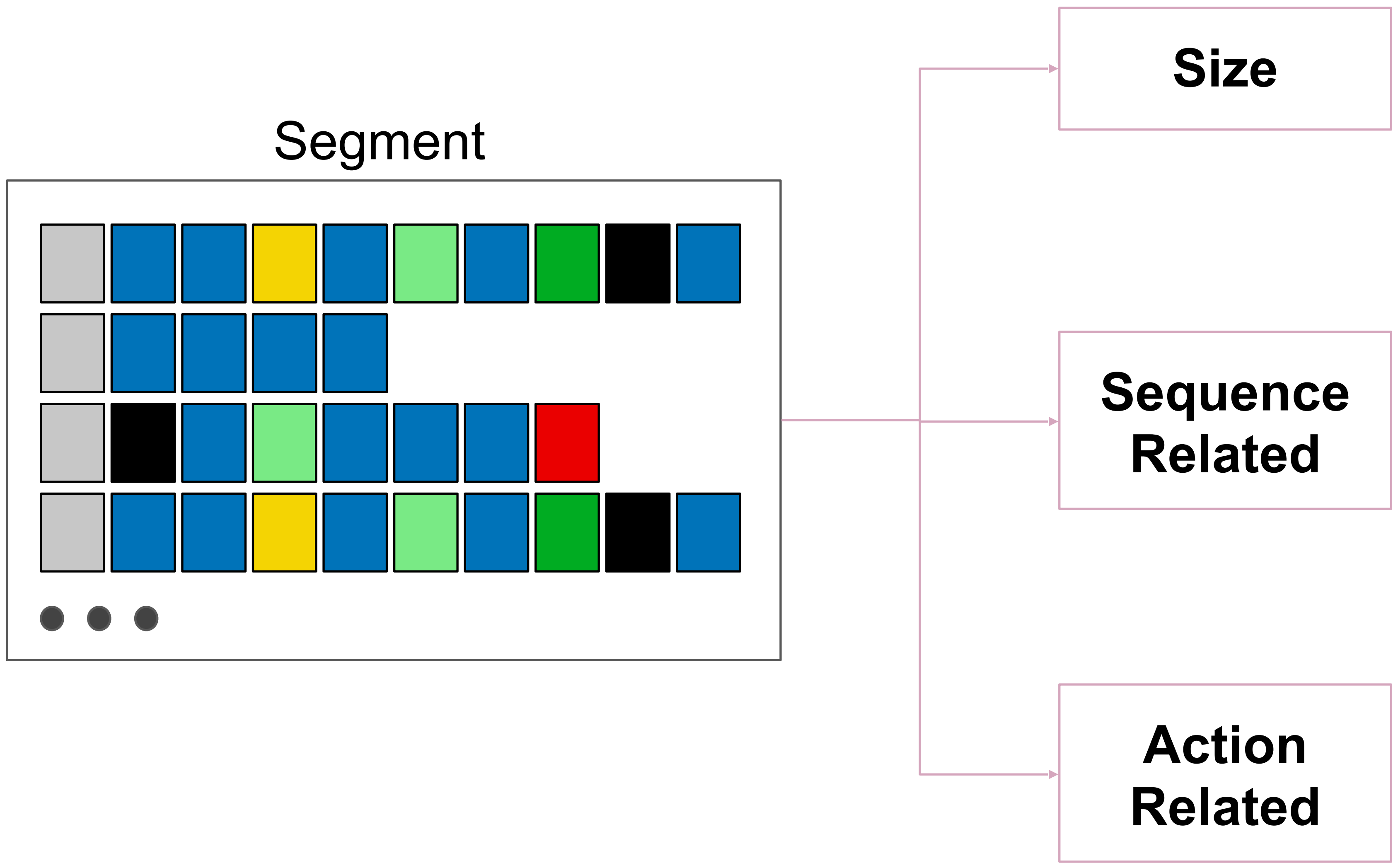


Data: *Segments*



Segment: any set of sequences

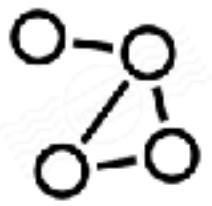
Data: *Segment Attributes*



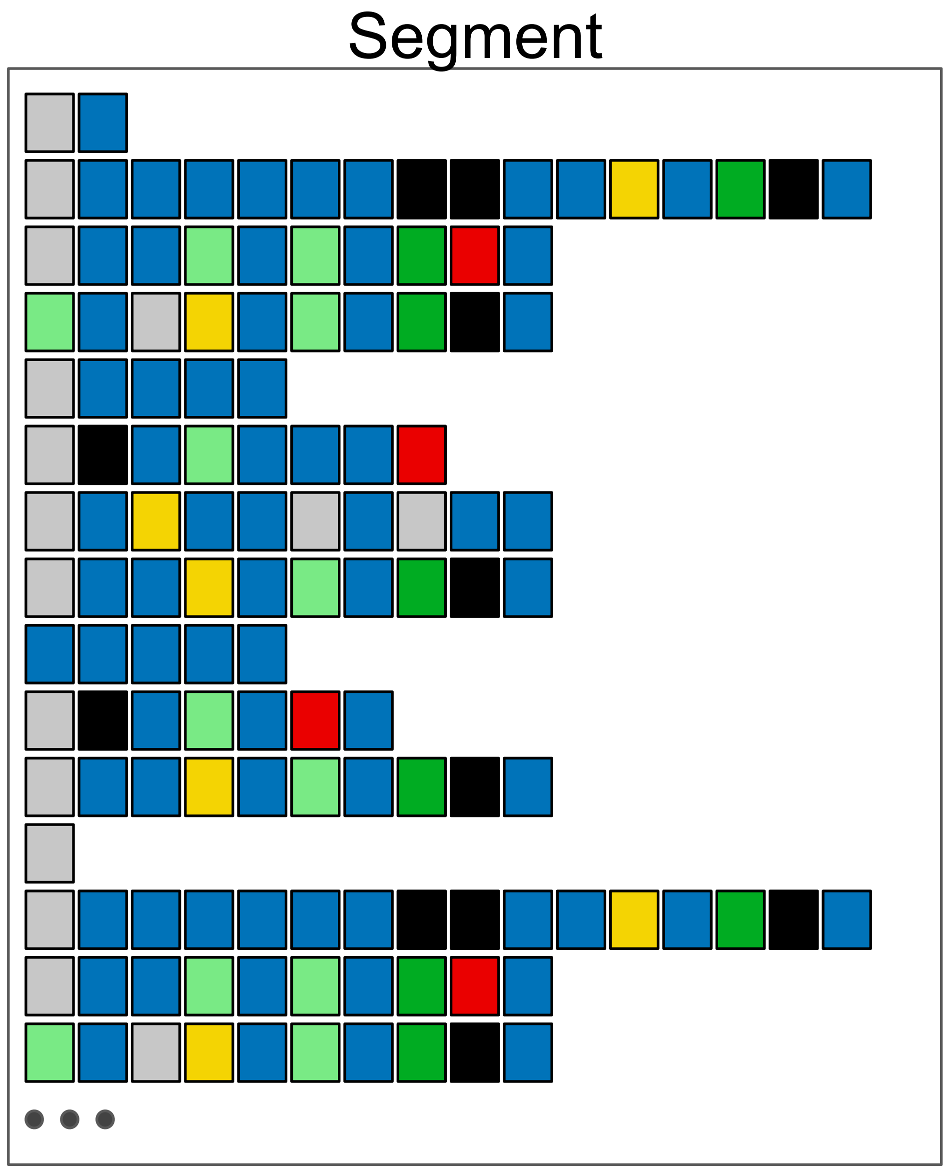
Counts of sequences:
Absolute, Relative

Sequence Distributions:
Start Time, Duration, Action Counts

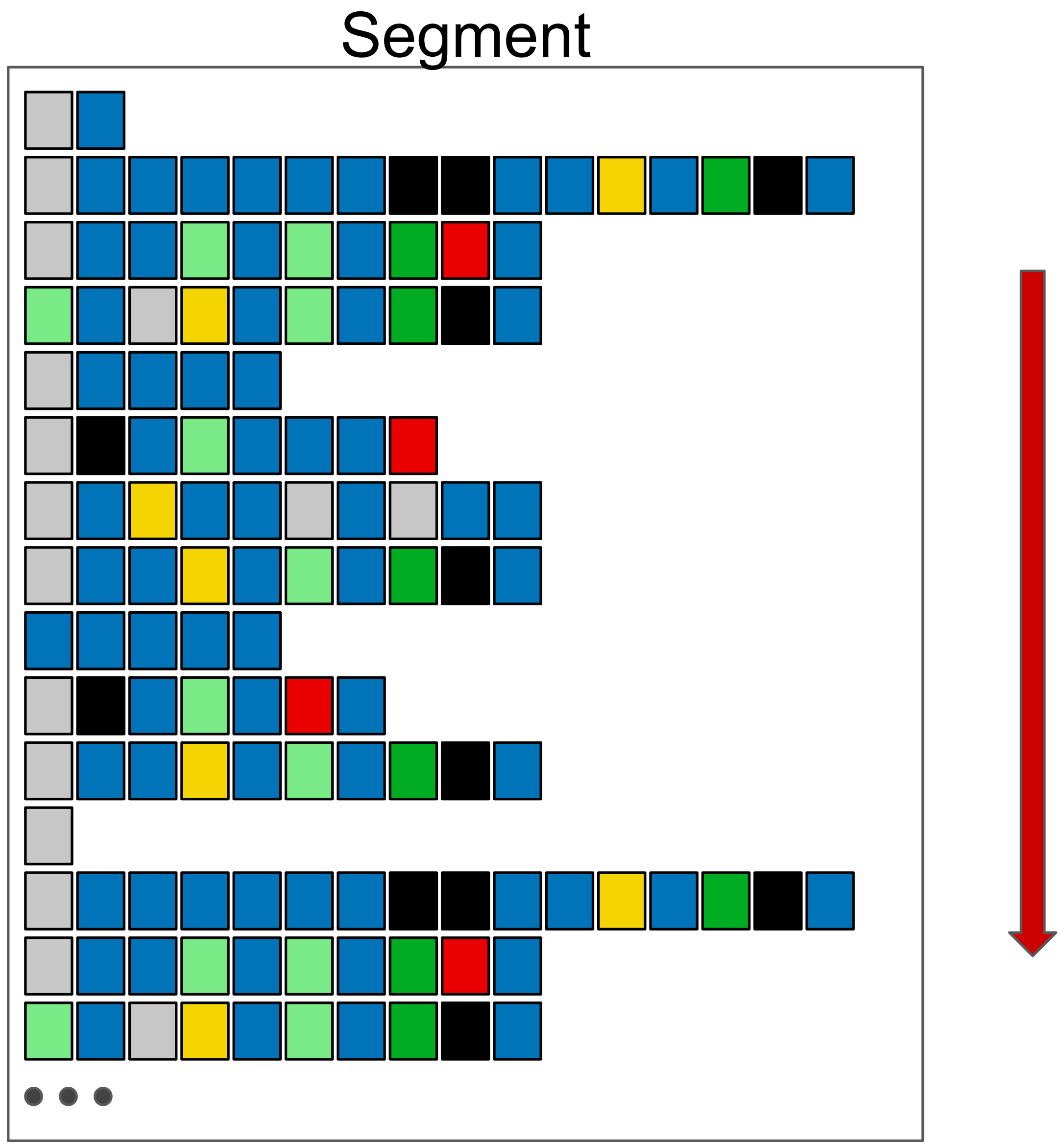
Action Distributions:
Action Transitions:
action before, action after



Real-world Clickstream Data

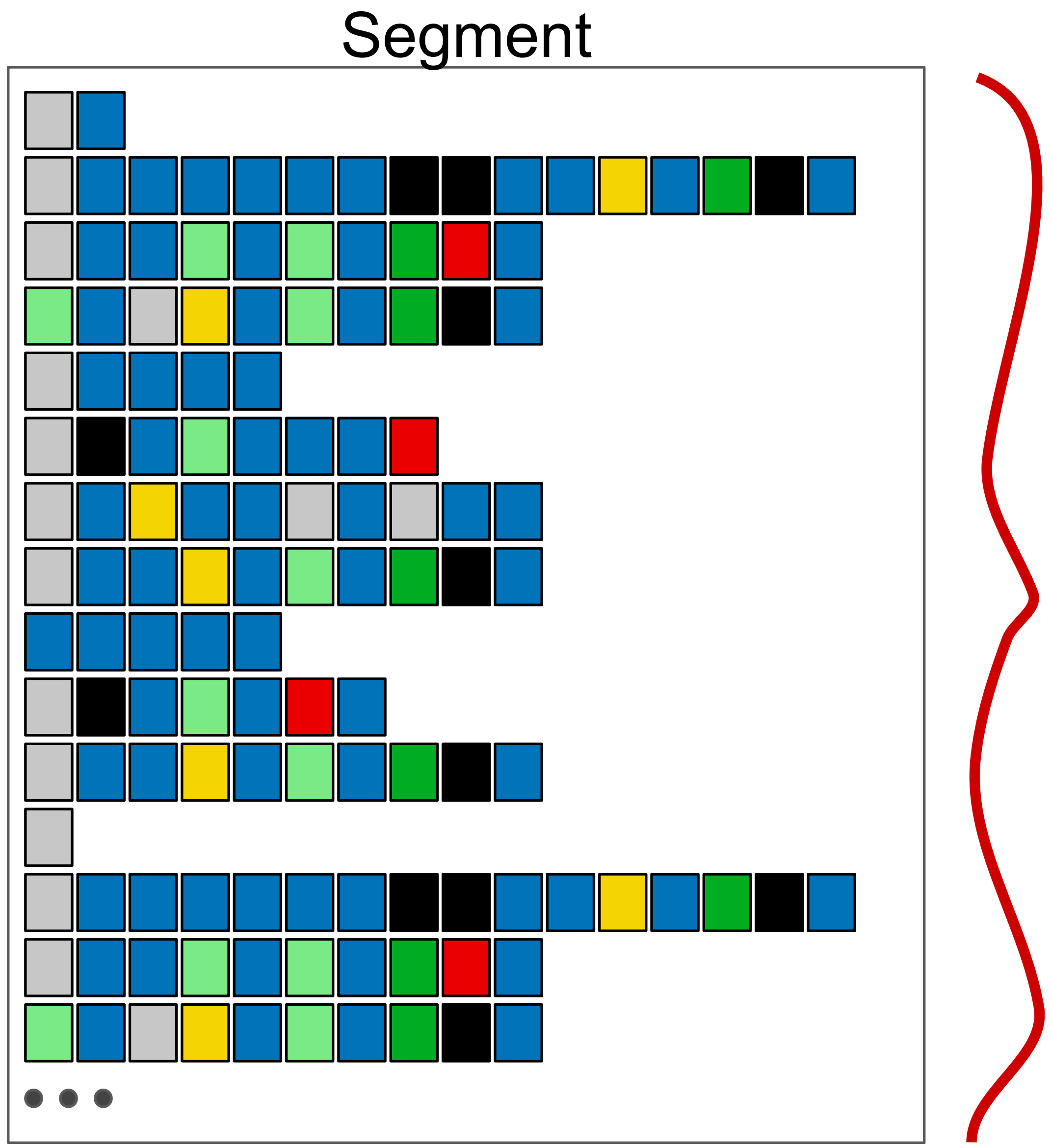


Real-world Clickstream Data



Scale is huge

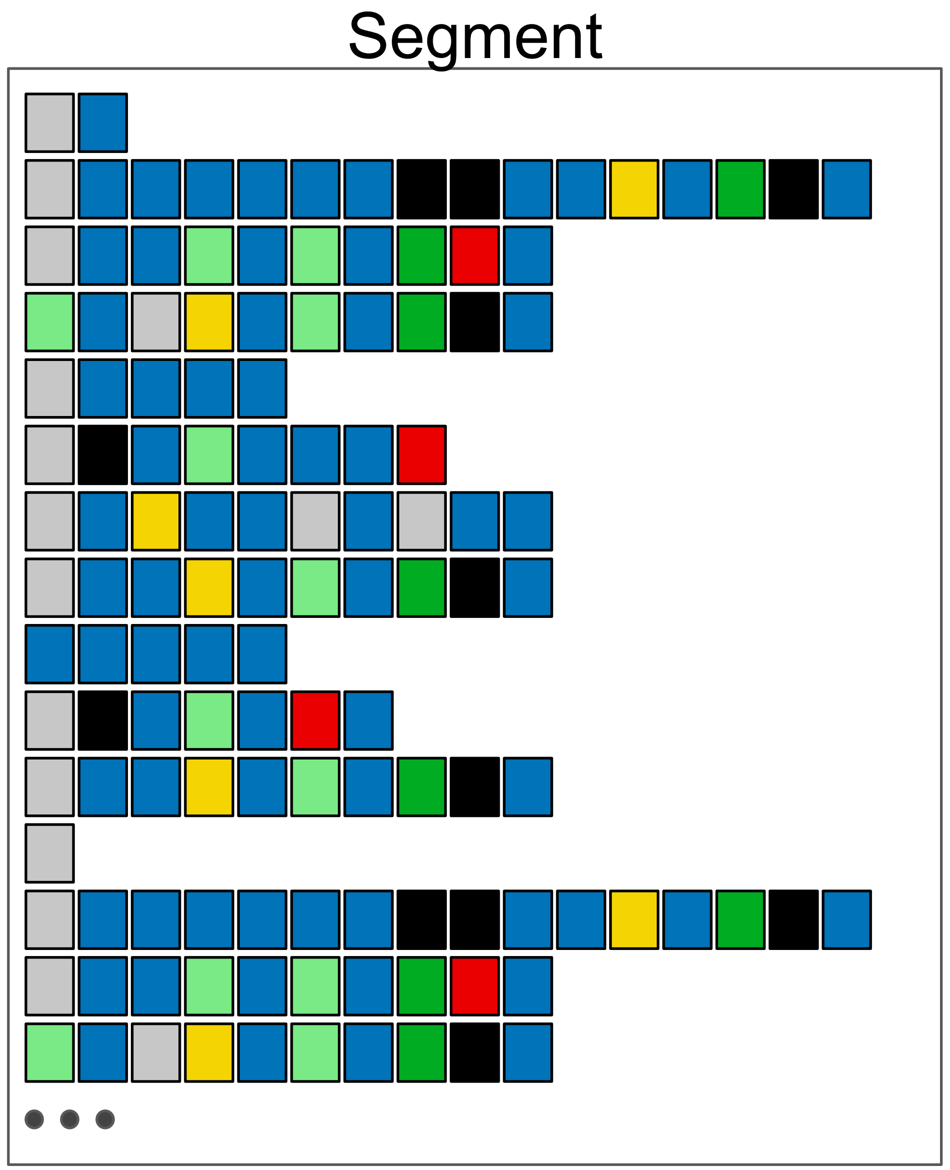
Real-world Clickstream Data



Scale is huge

Variability is high

Real-world Clickstream Data



Scale is huge

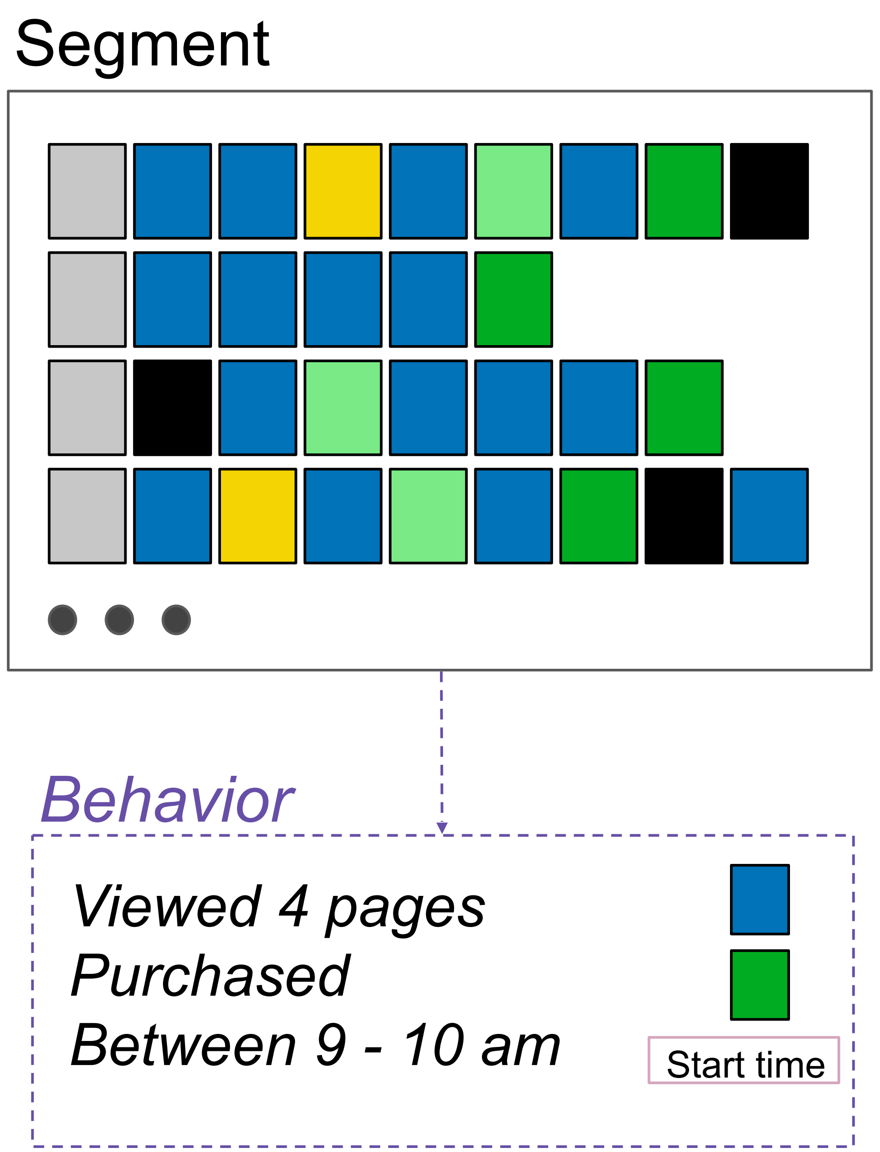
Variability is high

Most work **fails** when applied to real-world data

What are

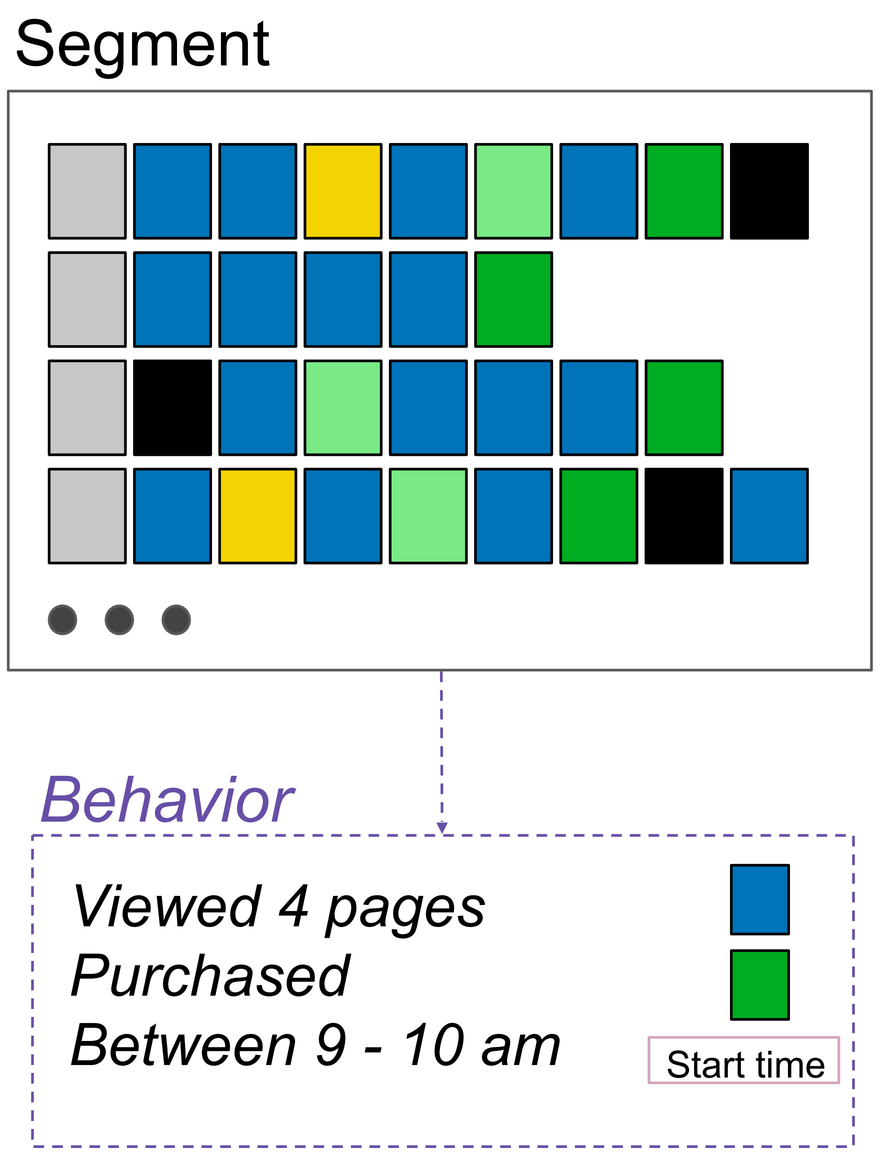
Clickstream Data Analysis Tasks?

Tasks: Segment Behavior



Behavior: set of attribute constraints

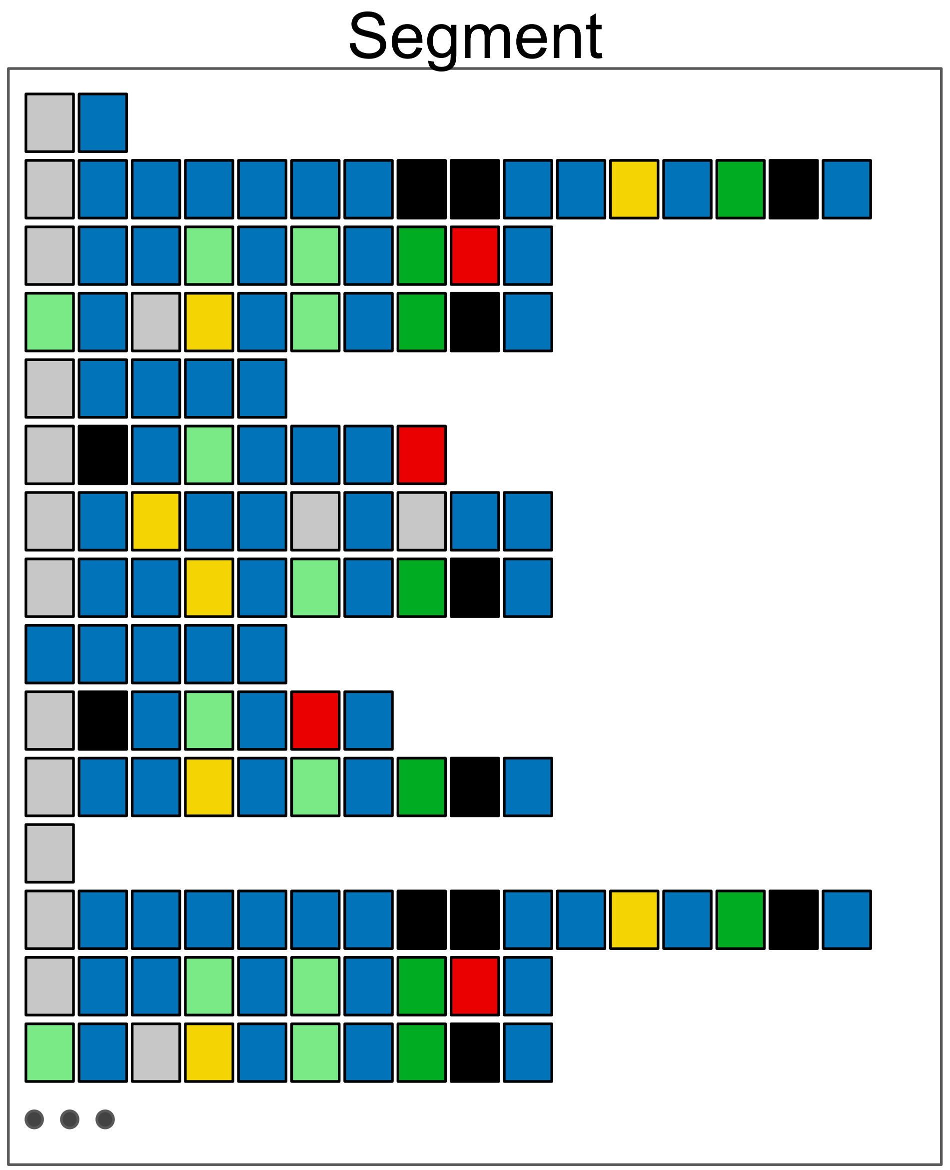
Tasks: Segment Behavior



Behavior: set of attribute constraints

- **Expected**
  Users add to cart before purchasing
- **Unexpected**
  No purchases on a certain month
- **Favorable**
  Purchased
- **Unfavorable**
  Bounced

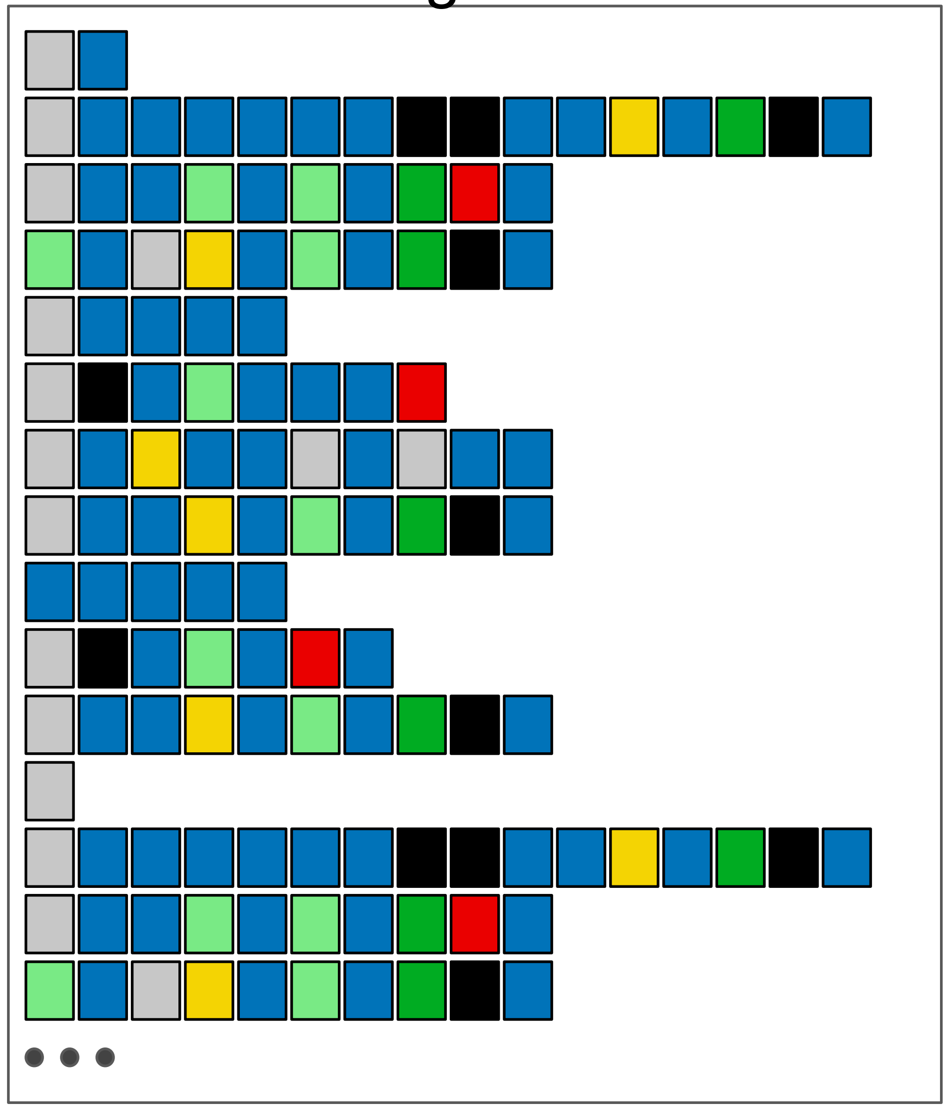
Tasks: Task Abstraction



Identify: Find some set of sequences that constitutes interesting *behavior*

Tasks: Task Abstraction

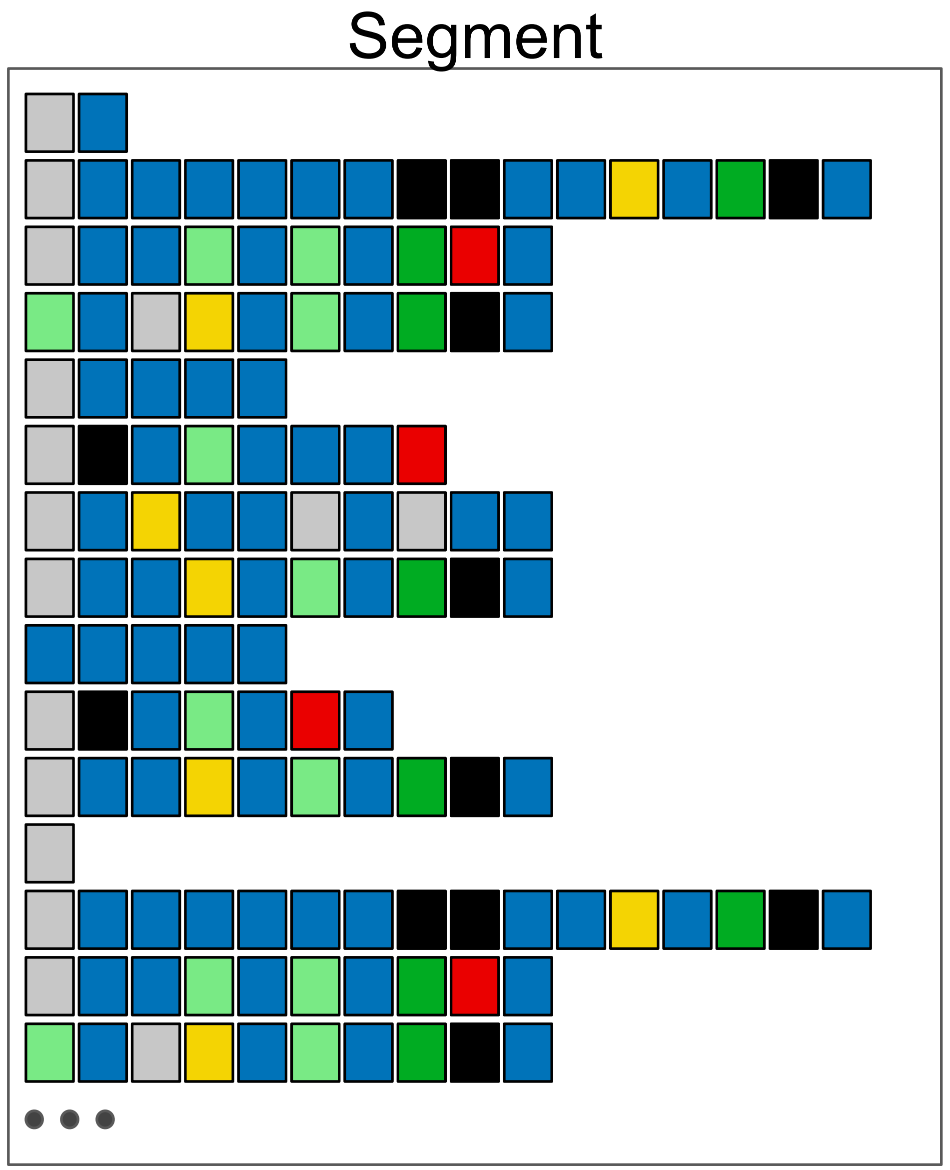
Segment



Identify: Find some set of sequences that constitutes interesting *behavior*

Drilldown: Distinguish more specific *behaviors* to further partition a segment previously defined by looser constraints

Tasks: Task Abstraction



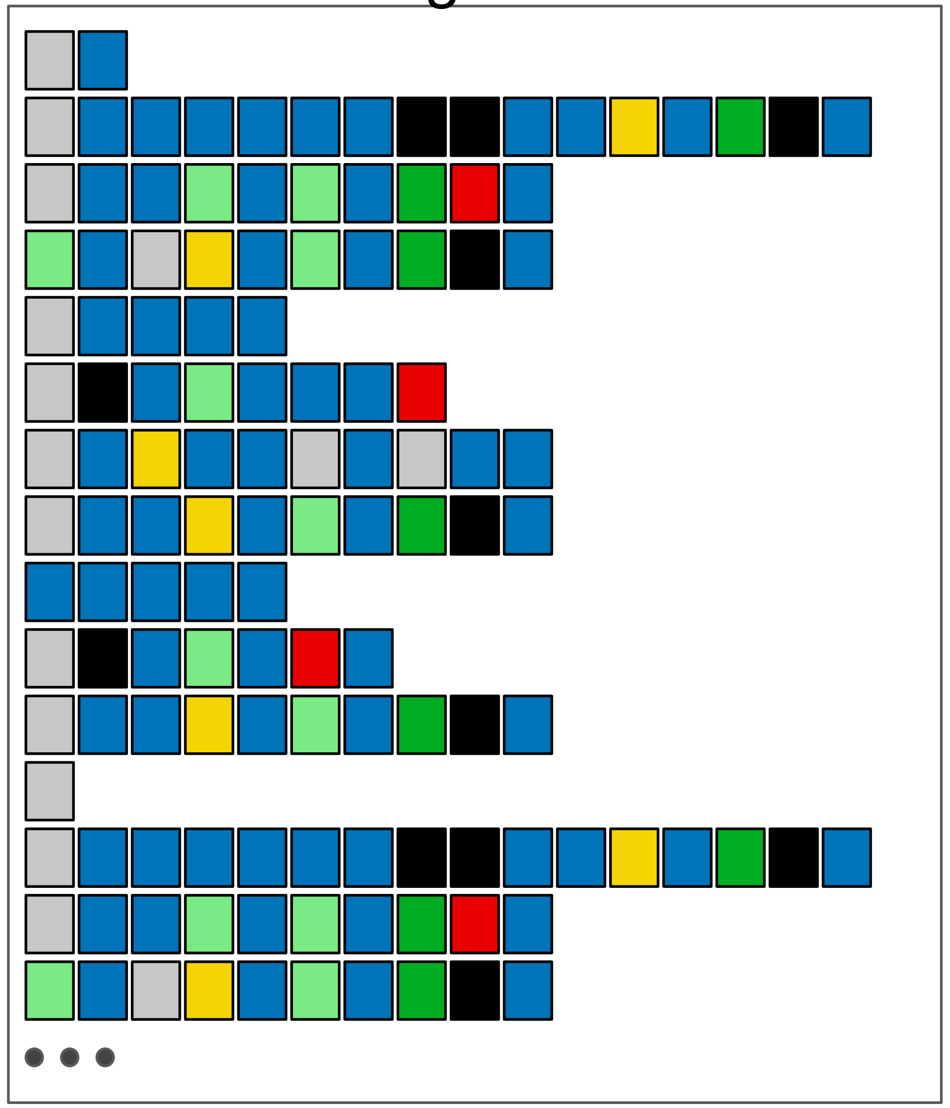
Identify: Find some set of sequences that constitutes interesting *behavior*

Drilldown: Distinguish more specific *behaviors* to further partition a segment previously defined by looser constraints

Frequency: Determine how many sequences are in the segment defined by *behavior*

Tasks: Task Abstraction

Segment



Identify: Find some set of sequences that constitutes interesting *behavior*

Drilldown: Distinguish more specific *behaviors* to further partition a segment previously defined by looser constraints

Frequency: Determine how many sequences are in the segment defined by *behavior*

Ordering within sequence: Match if one action subsequence occurs before (or after) another action subsequence in a sequence

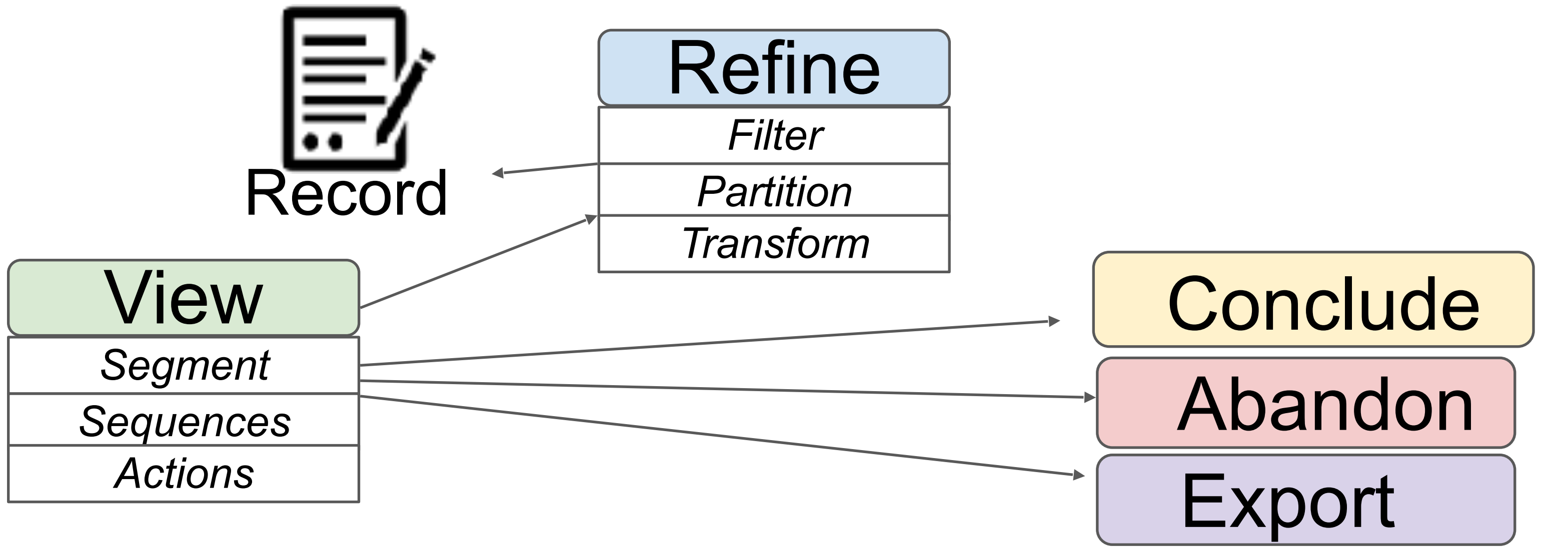
High-Level Segmentifier Analysis Model

High-Level Segmentifier Analysis Model

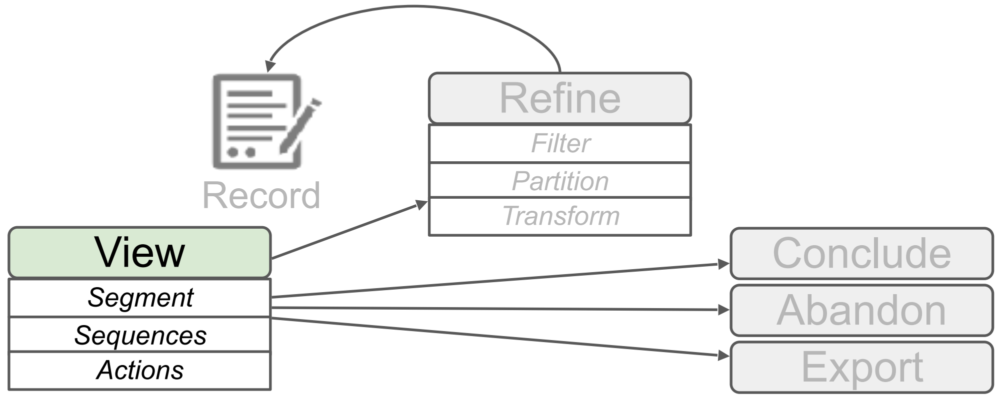
- Abstraction above task/data level to provide design rationale

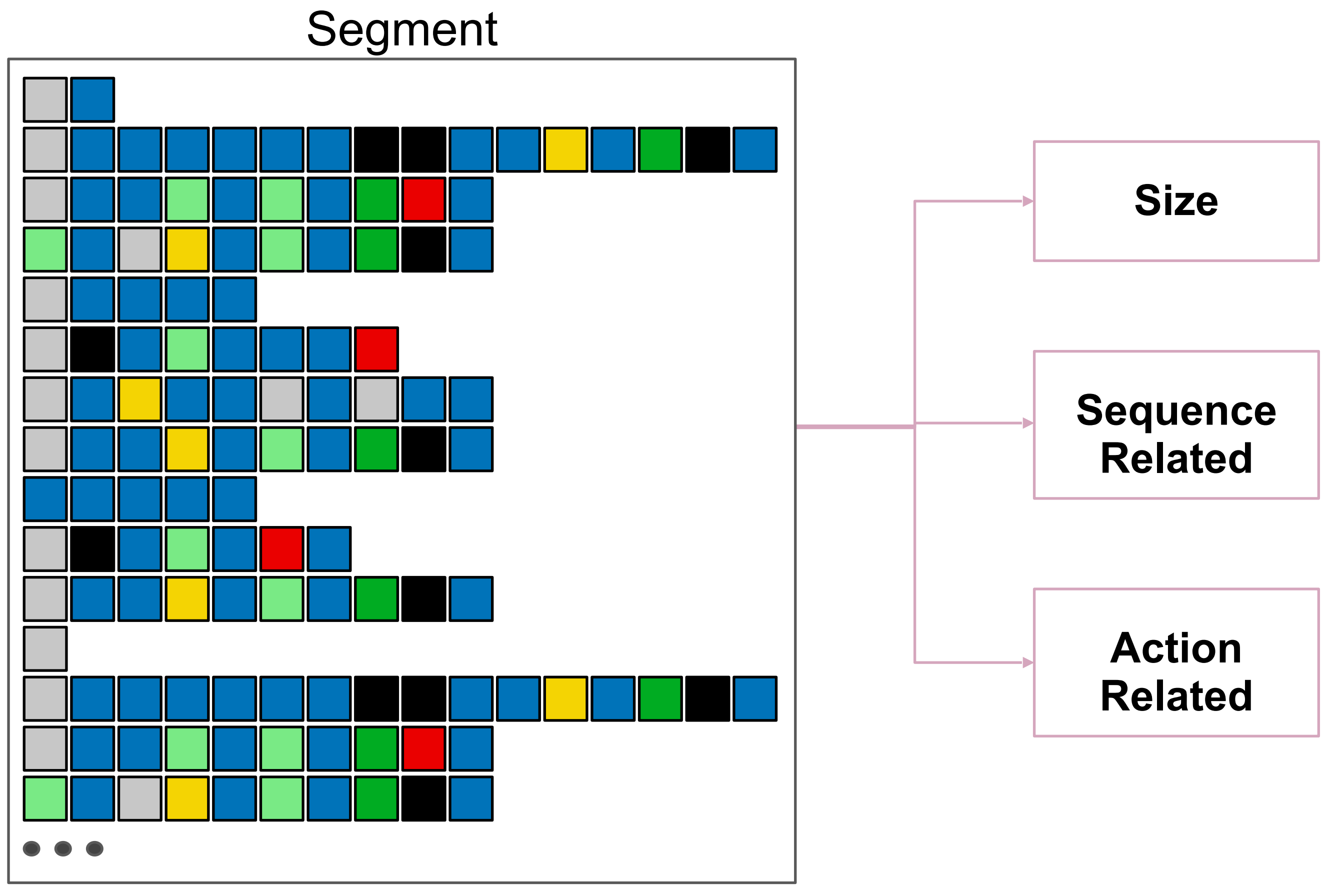
- Take a *giant, noisy datase*t and refine it into *small, clean segments* for
  - actionable insights
  - downstream analysis

- Bridge the gap between *real-world data* and other techniques



High-Level Segmentifier Analysis Model

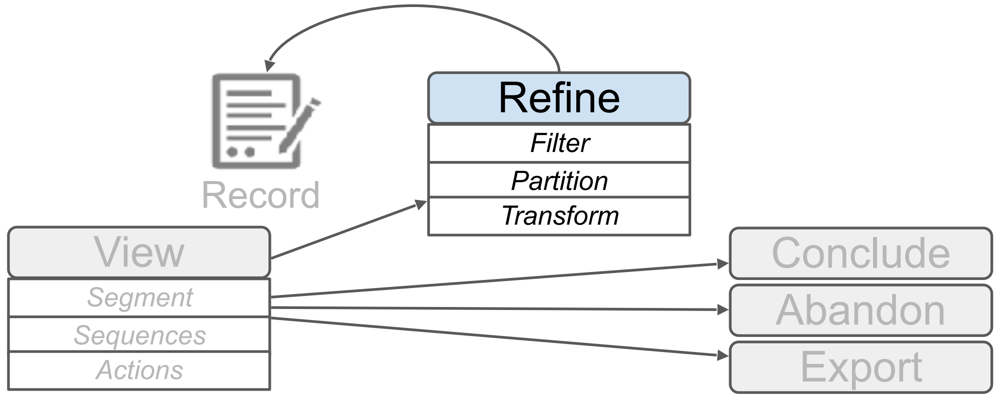


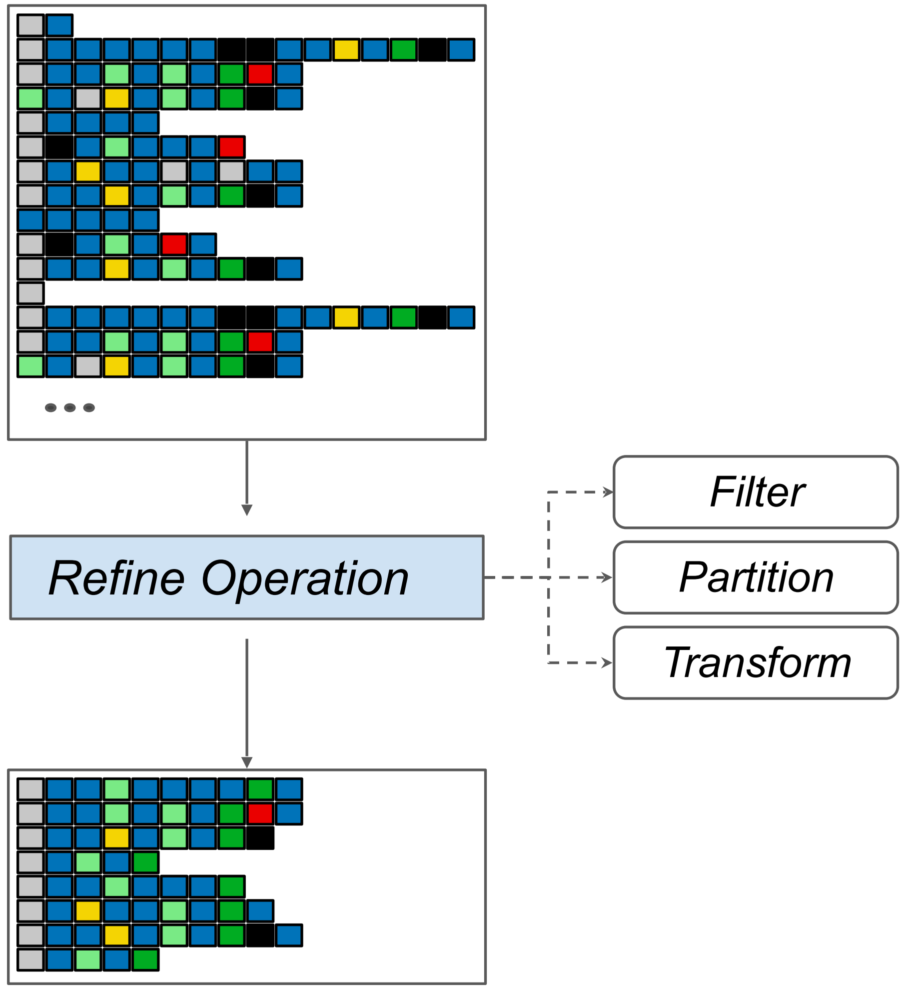


View

- Gives Insight into underlying data of segment
  - Action Attributes
  - Sequence Attributes
  - Segment Attributes
- Leads to:
  - Insights
  - New ways on how to *refine*
  - Whether segment should be *abandoned*
  - Whether segment should be *exported*

High-Level Segmentifier Analysis Model

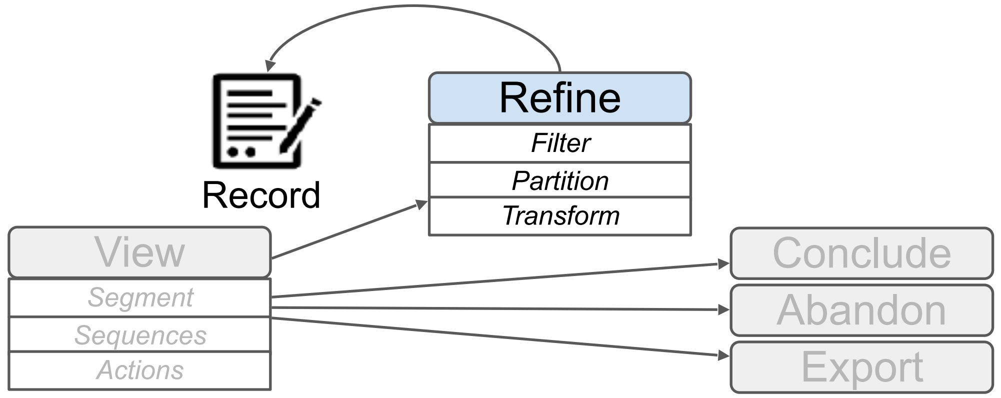


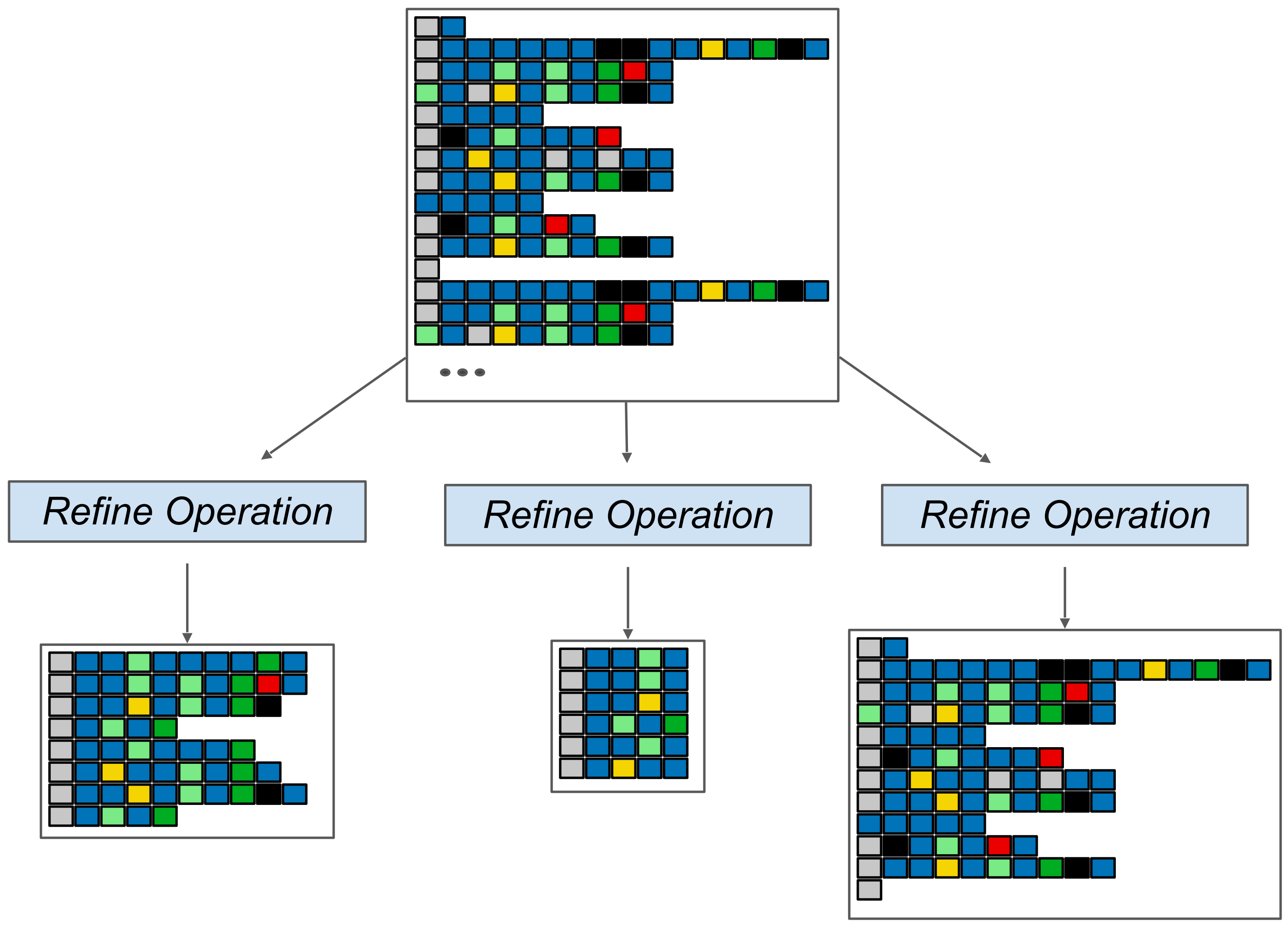


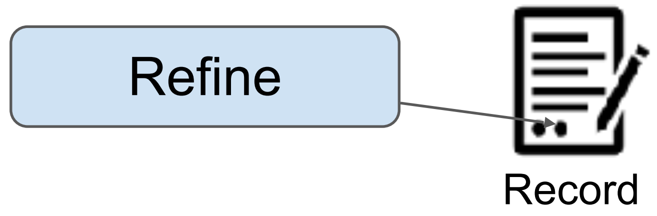
Refine

- Apply operation to create new segments
- Type of Refinements
  - *Filter*
  - *Partition*
  - *Transform*

High-Level Segmentifier Analysis Model

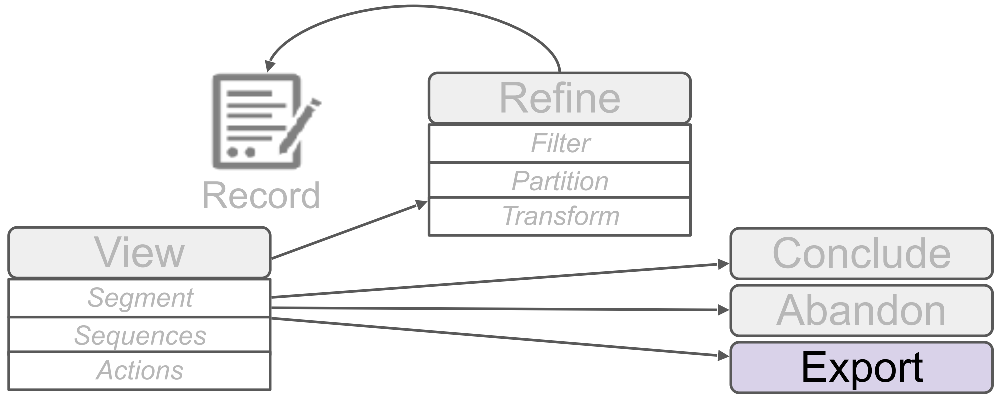


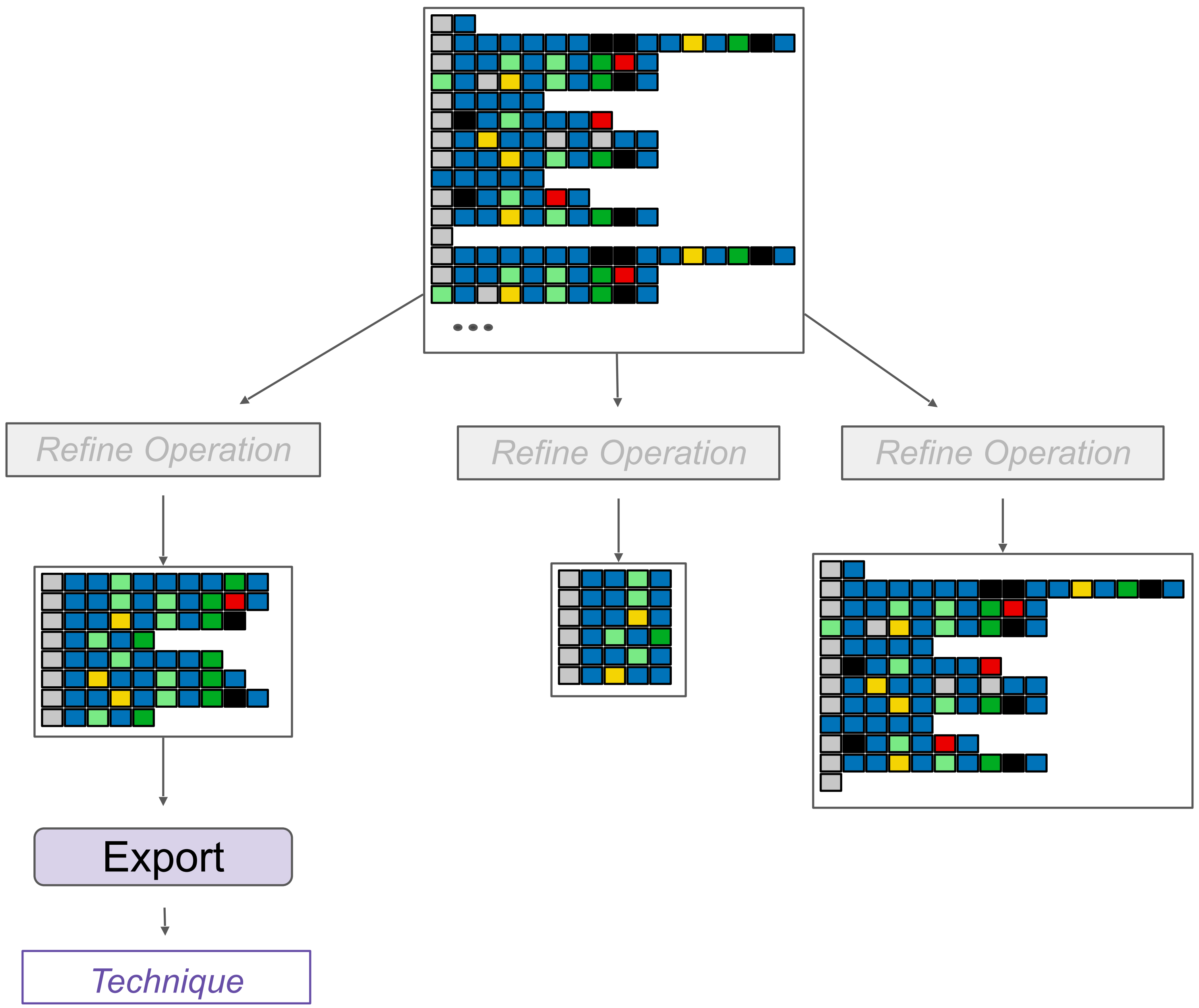




- Record all refinement steps automatically
- Keep track of questions asked and hypotheses tested
- Ability to create and view multiple segments from the same segment

High-Level Segmentifier Analysis Model

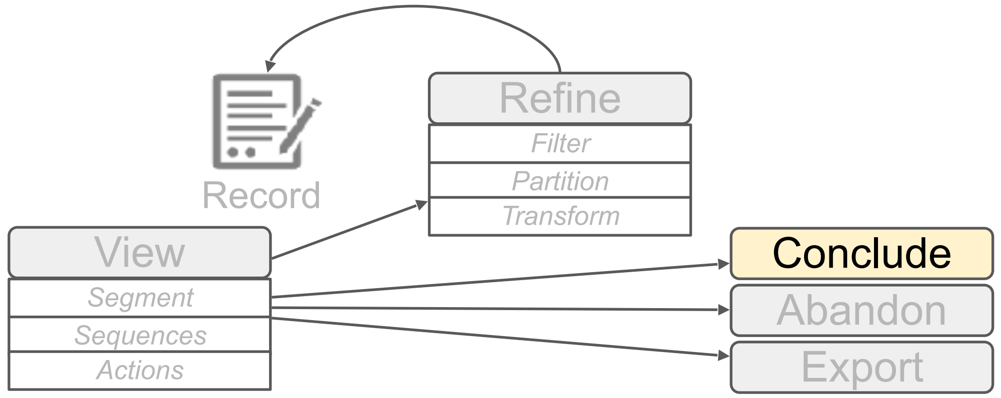




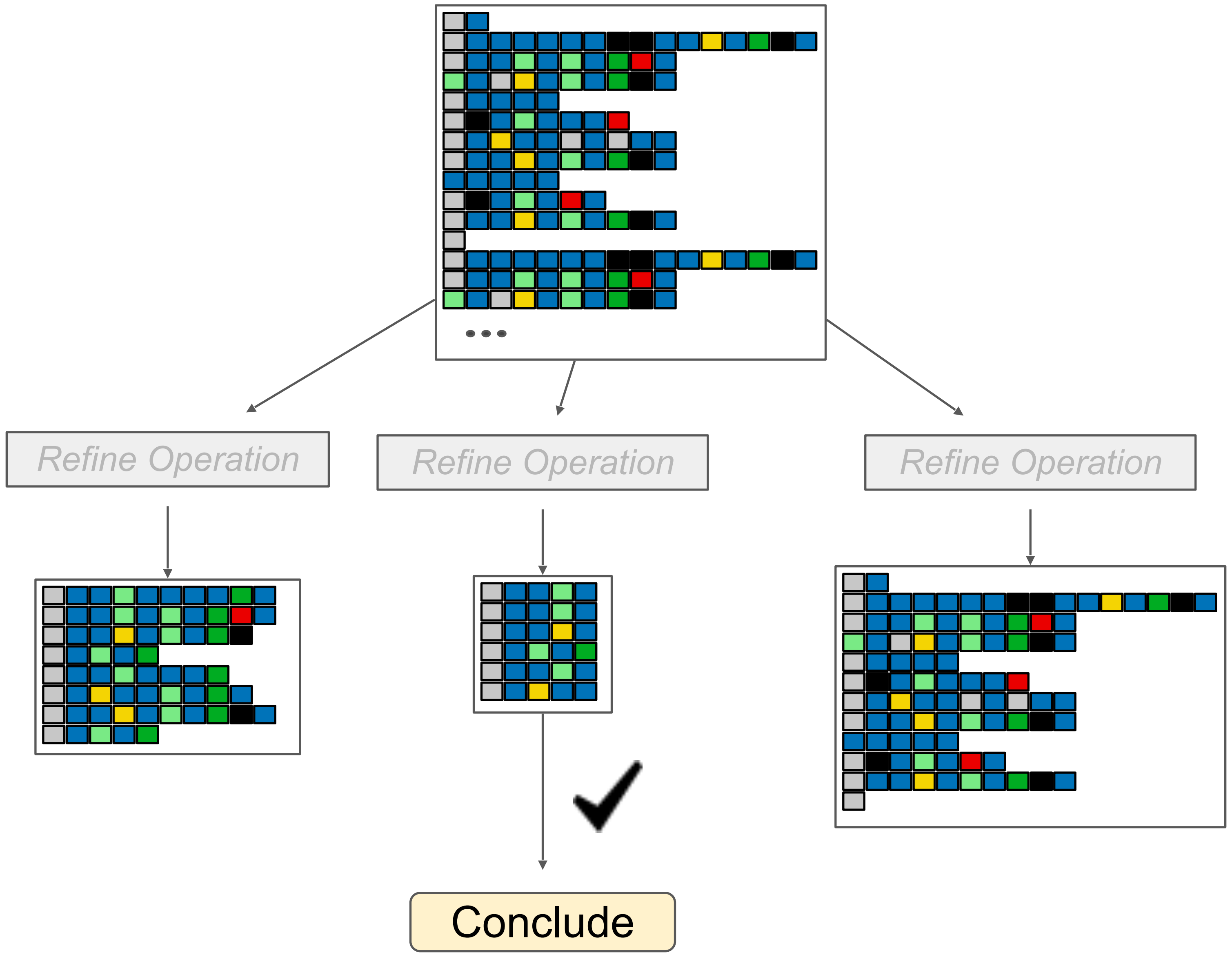
Export

- Export refined segments for further downstream analysis, to more specific tools:
    - Pattern mining
    - Clustering

High-Level Segmentifier Analysis Model

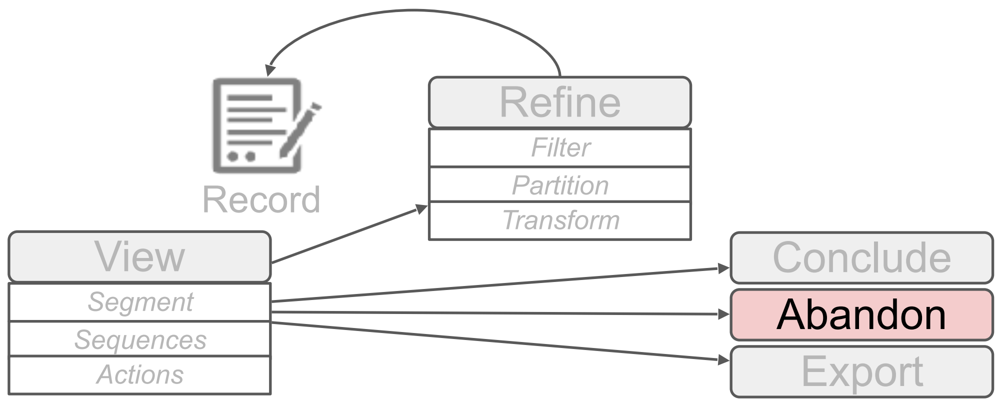


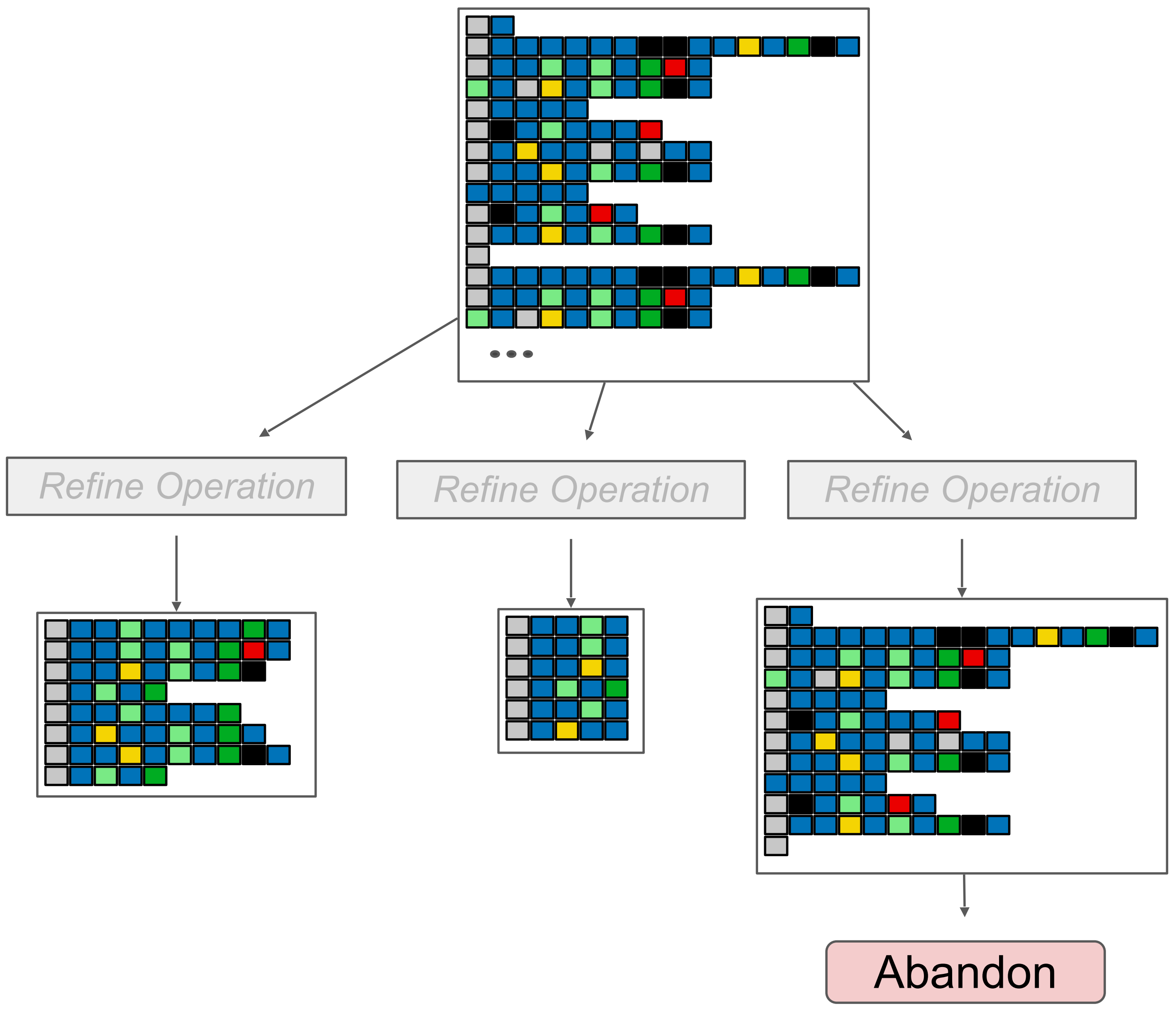
Conclude



- Discover actionable insight by *viewing* segment

High-Level Segmentifier Analysis Model

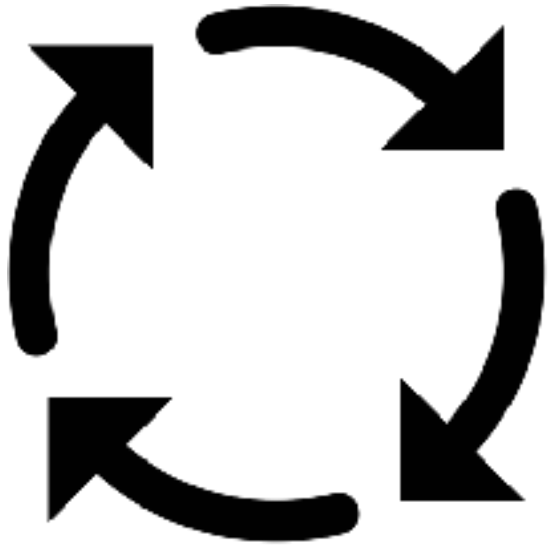




Abandon

- By *viewing* the segment, analyst *abandons* if:
  - No actionable insights
  - No further ways to *refine*
  - Not suitable for *export*

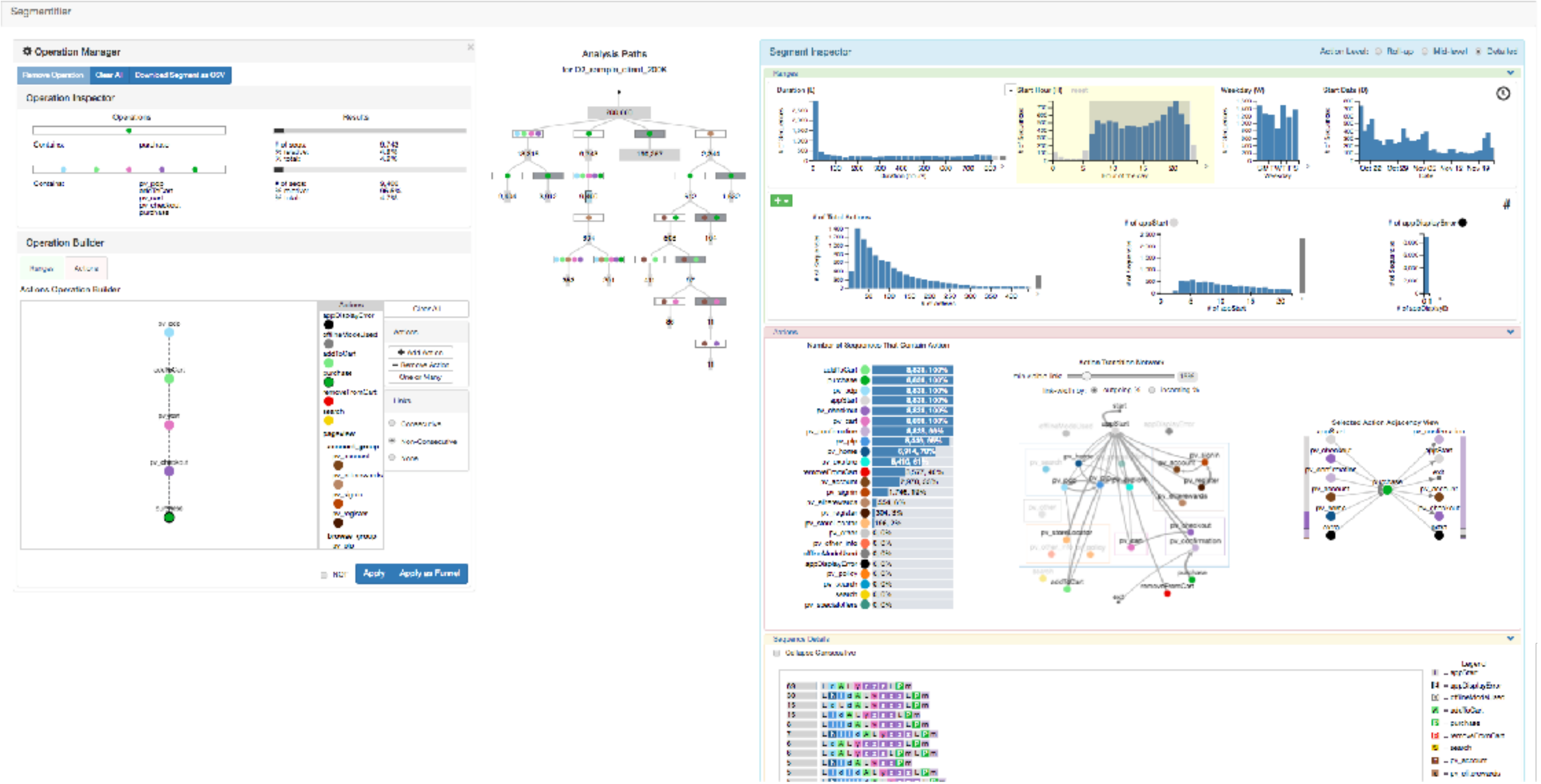
Why Visual Analytics?



- Automation would be nice...
  - Put data in, actionable results appear
- … but it is not realistic
  - Many possible questions, data-driven interplay between finding answers and generating new questions
- Human-in-the-loop visual data analysis
  - Integrate computing power of machine with intuition of domain experts

Solution

The Segmentifier Interface

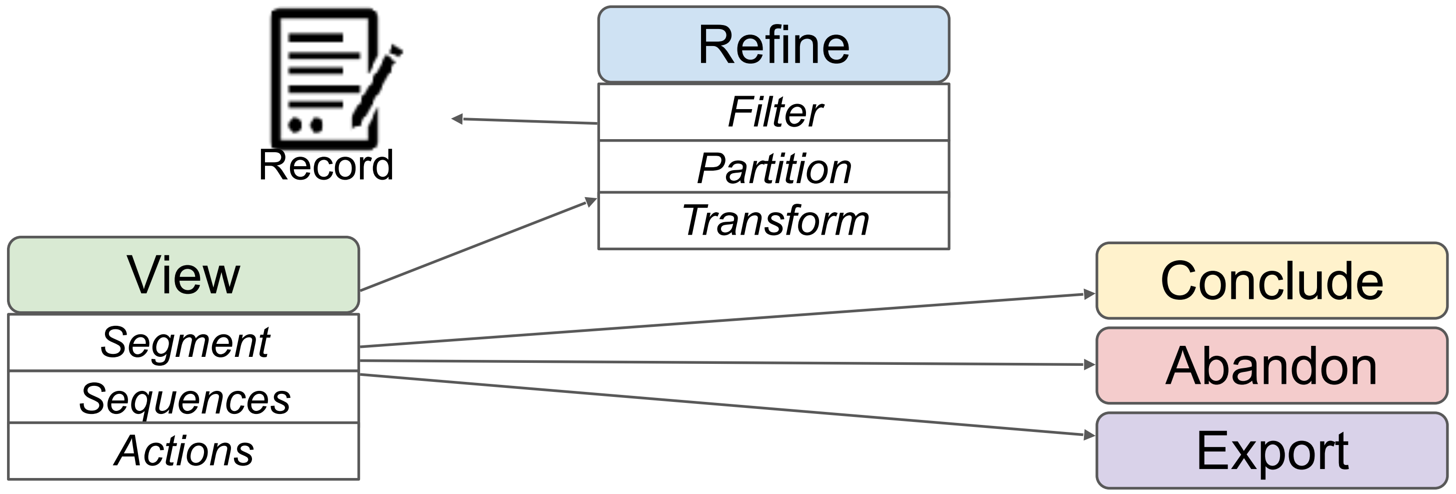


Video

https://www.youtube.com/watch?v=TobYDFeISOg&t=20s

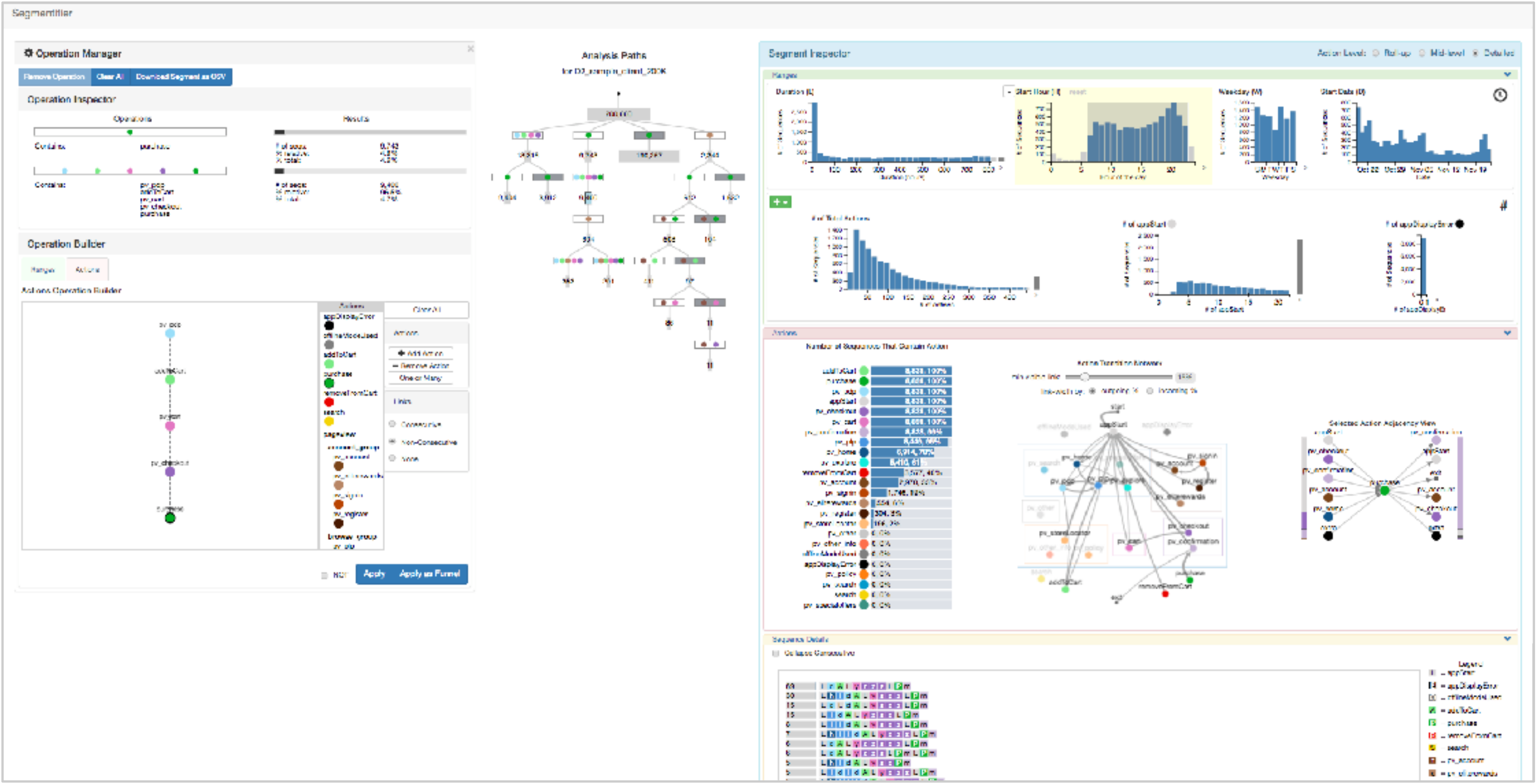
Segmentifier Contributions

➢ Thorough **characterization of task and data abstraction** for clickstream data analysis



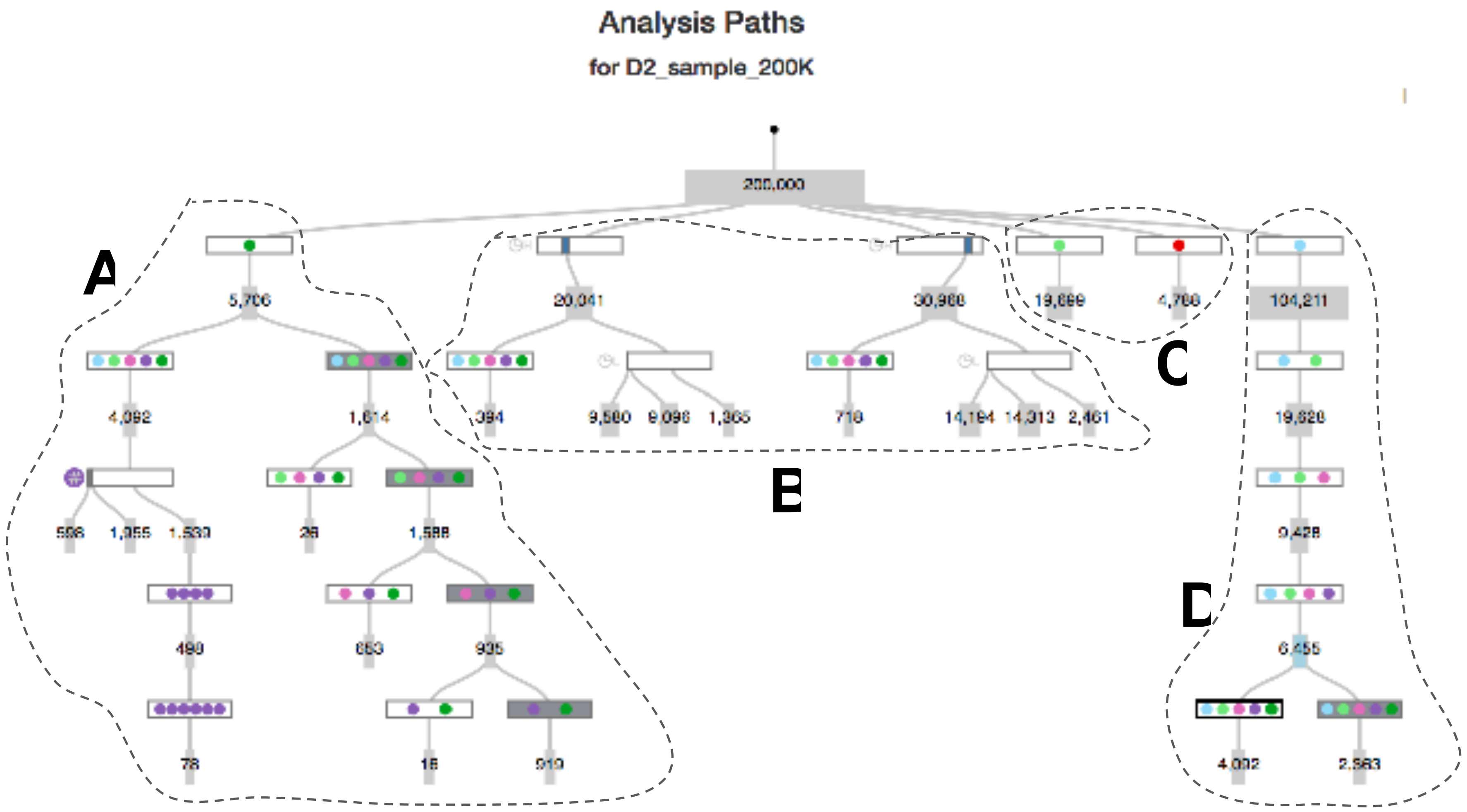
Segmentifier Contributions

- Thorough **characterization of task and data abstraction** for clickstream data analysis
- **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis



Segmentifier Contributions

- ➢ Thorough **characterization of task and data abstraction** for clickstream data analysis
- ➢ **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis
- ➢ Preliminary **evidence of utility**



Three case studies of problem-driven work

- e-commerce

- facilities management

- biology

e commerce by shashank singh from the Noun Project   Biology by tezar tantular from the Noun Project   Business by Colourcreatype from the Noun Project

Michael Oppermann

Ocupado

Visualizing Location-Based Counts Over Time Across Buildings

http://www.cs.ubc.ca/labs/imager/tr/2020/ocupado/

Ocupado: Visualizing Location-Based Counts Over Time Across Buildings.

Oppermann and Munzner. Computer Graphics Forum (Proc. EuroVis 2020) 39(3):127-138 2020.

Location-Based Counts

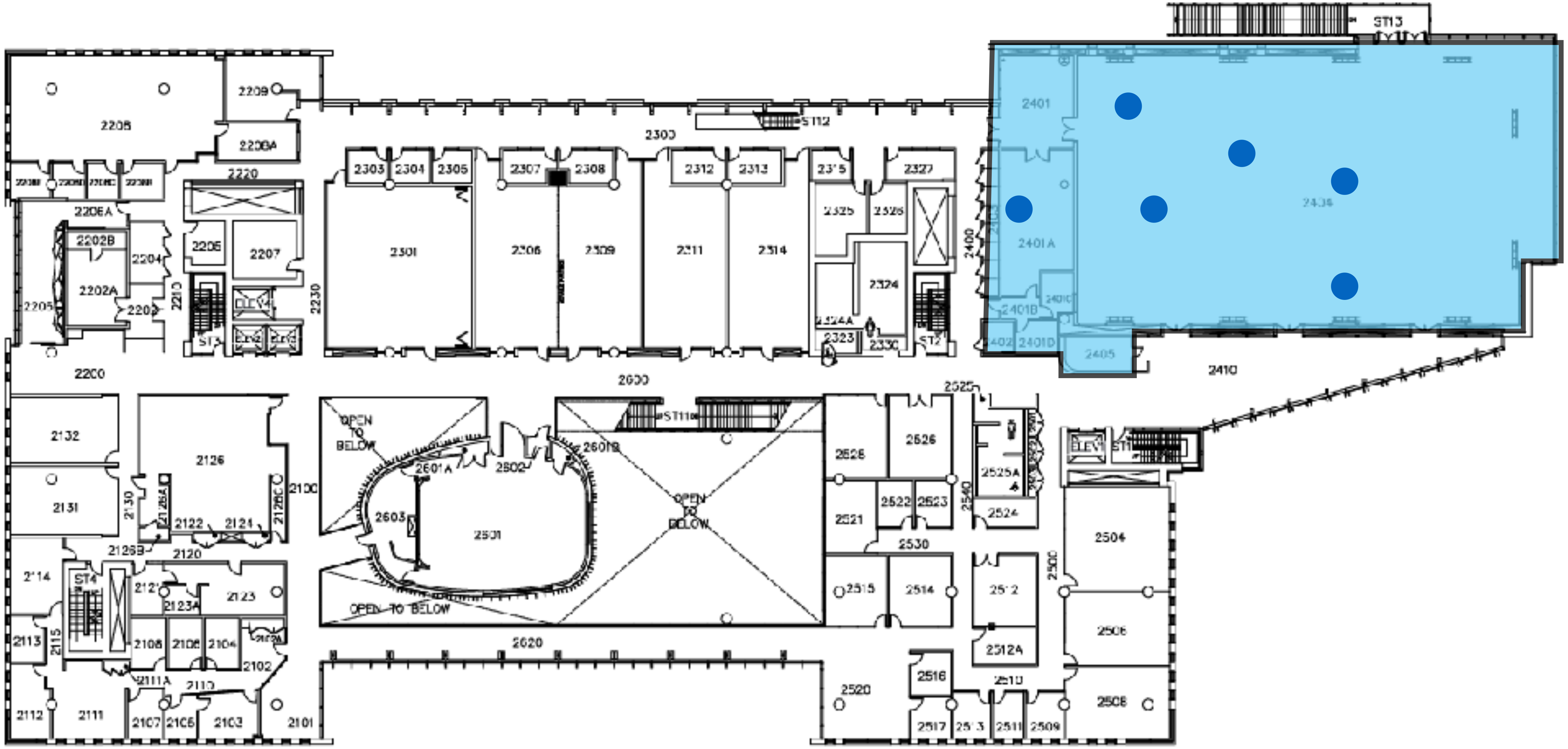
Previous measurement required physical counting or installation of additional hardware.

Previous measurement required physical counting or installation of additional hardware.

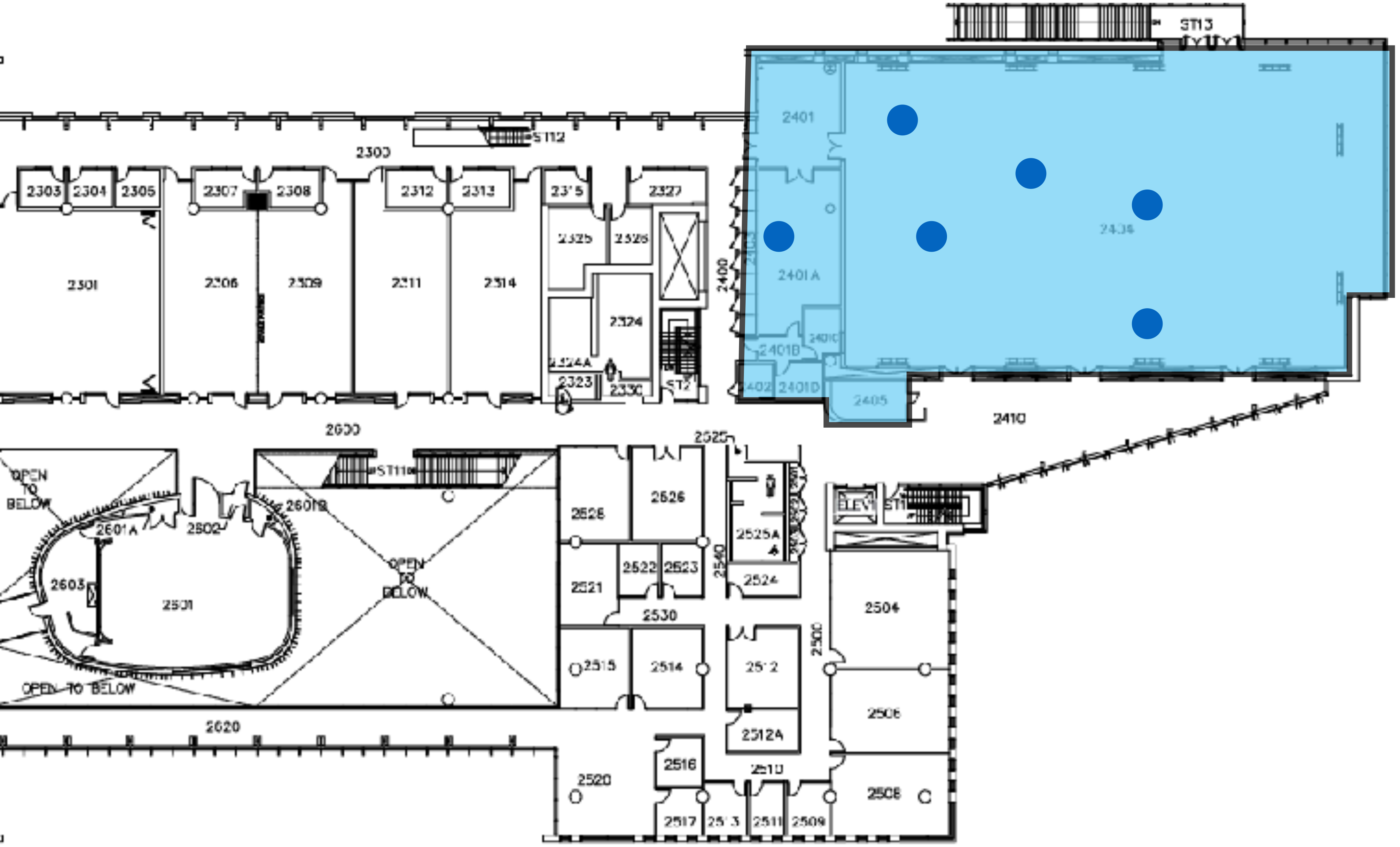
Previous visualization attempts were limited in space and time.

Design Study

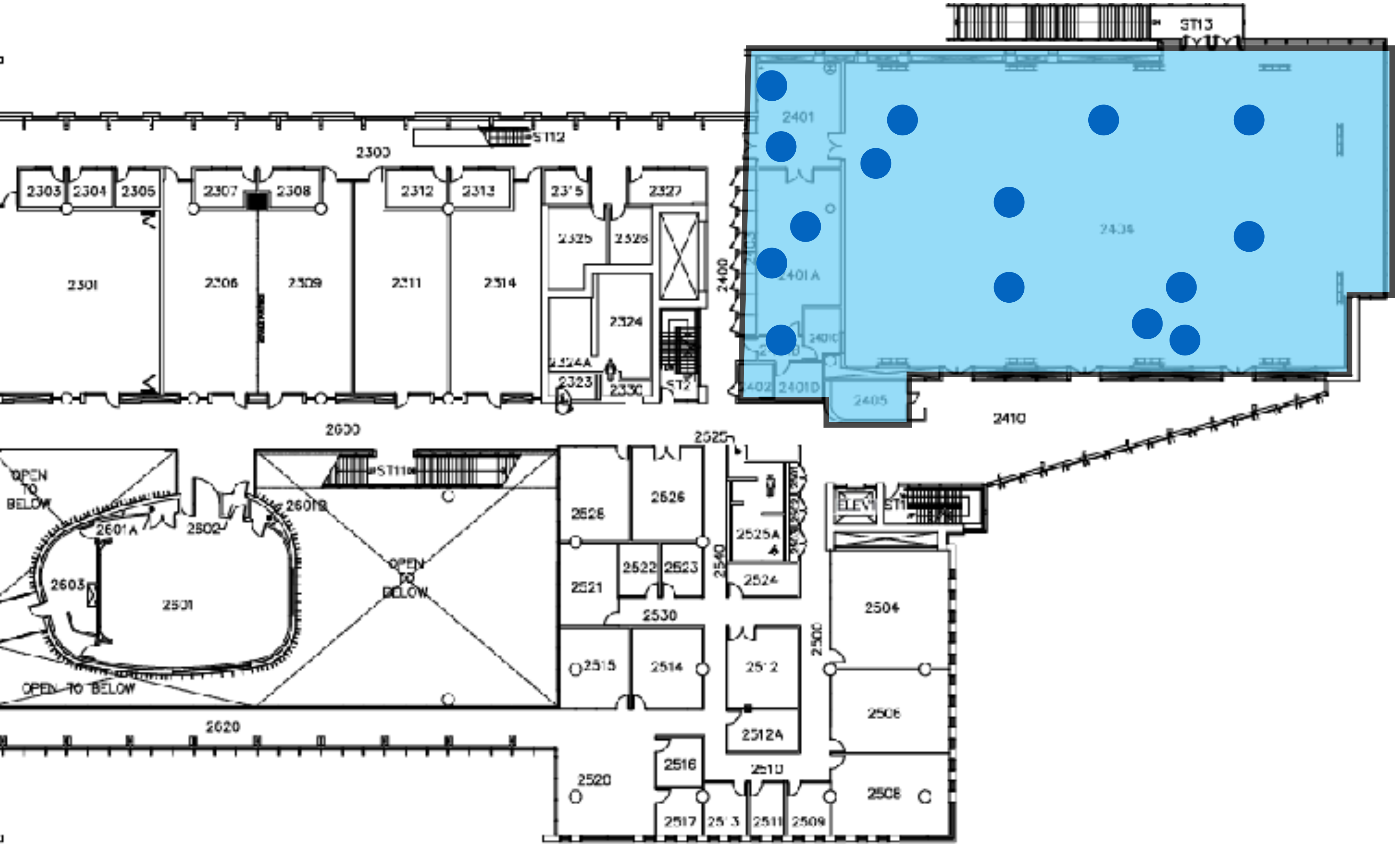
WiFi Connections: Location-Based Counts

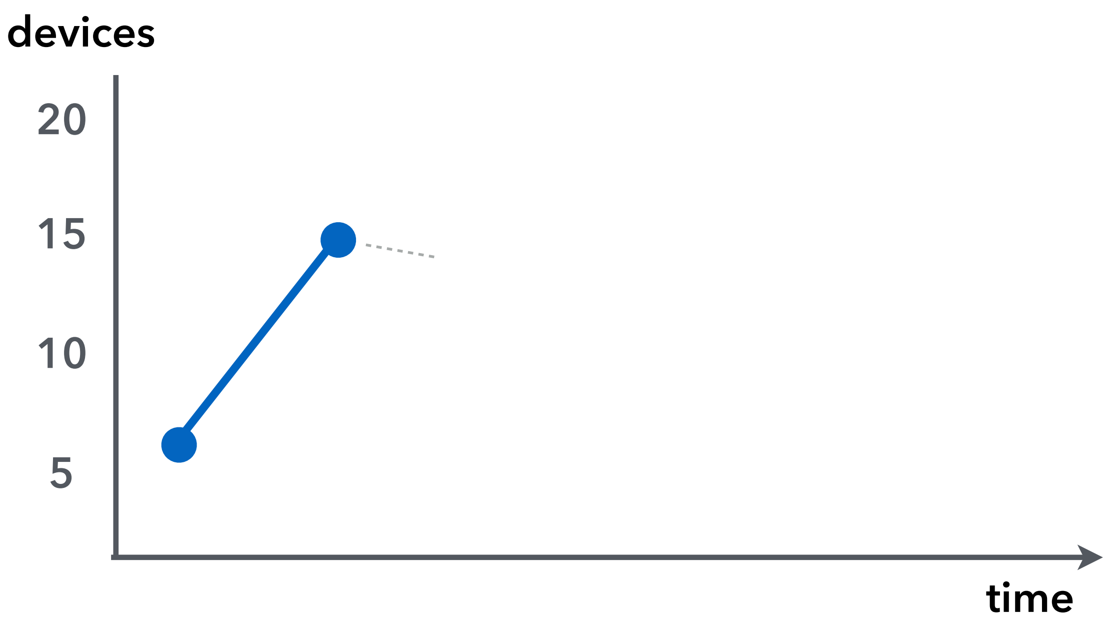


WiFi Connections: Location-Based Counts

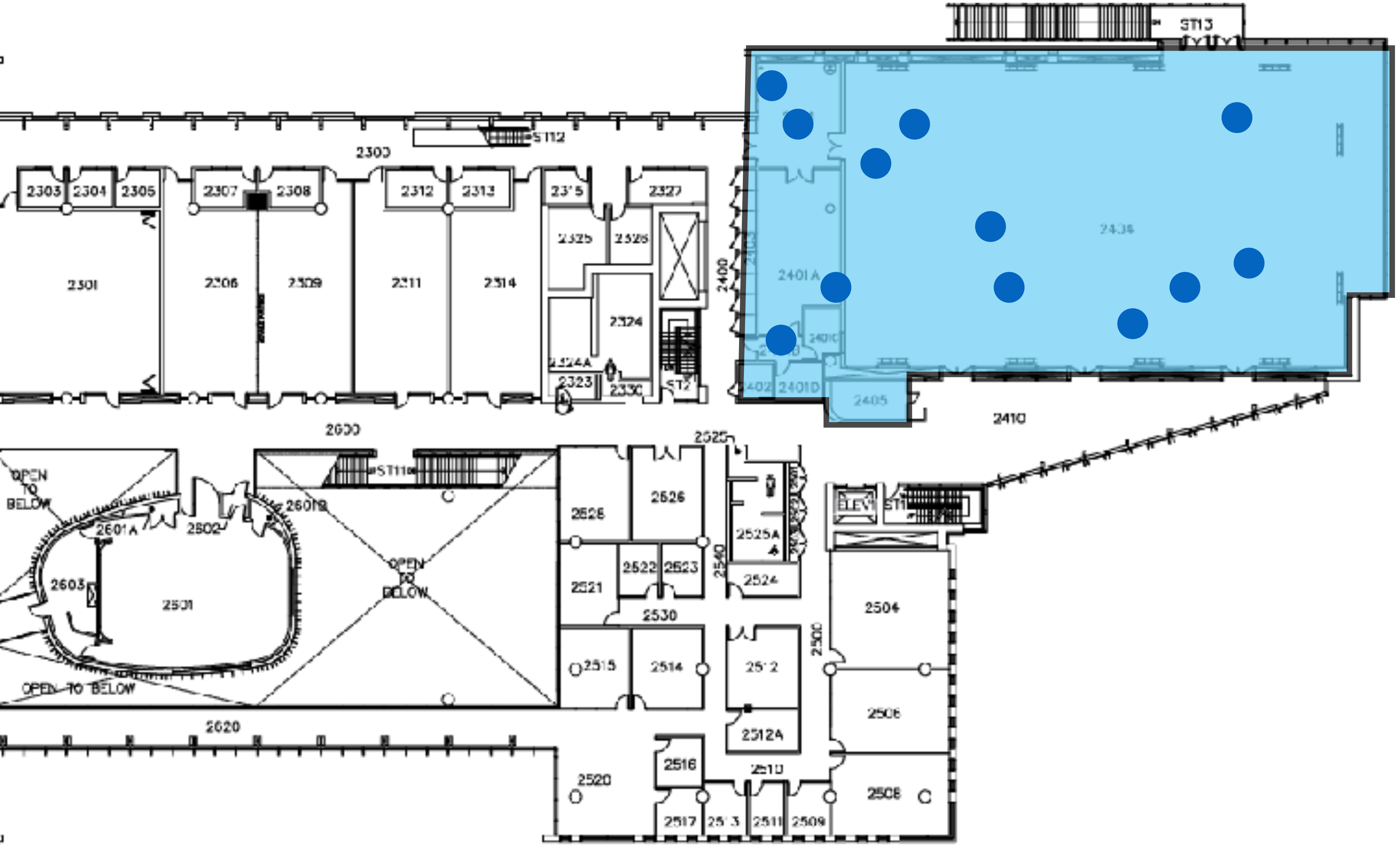


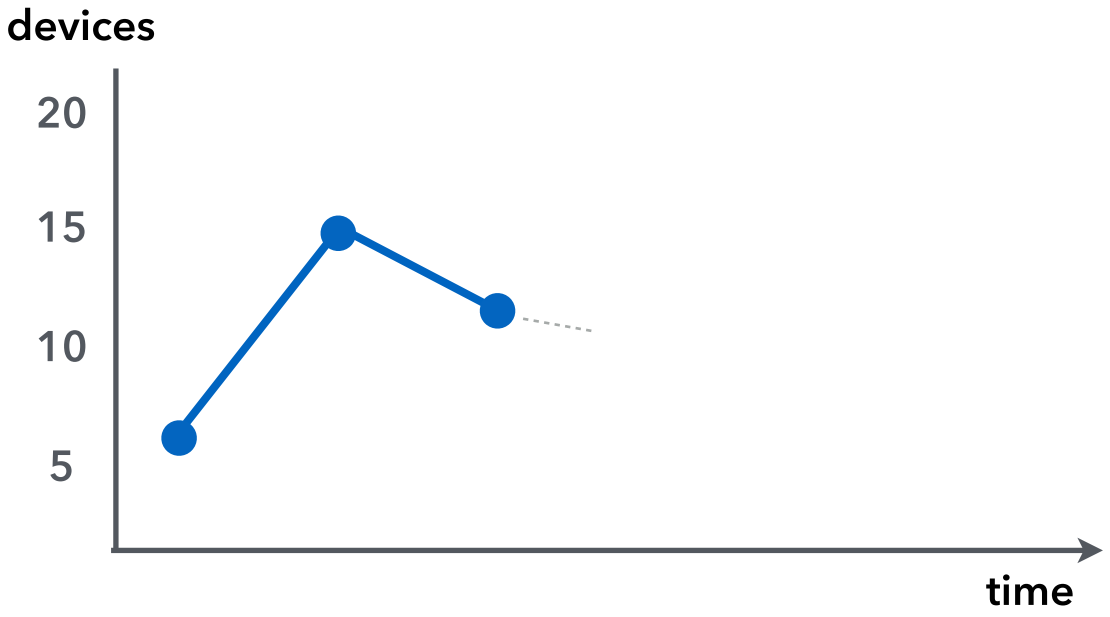
WiFi Connections: Location-Based Counts



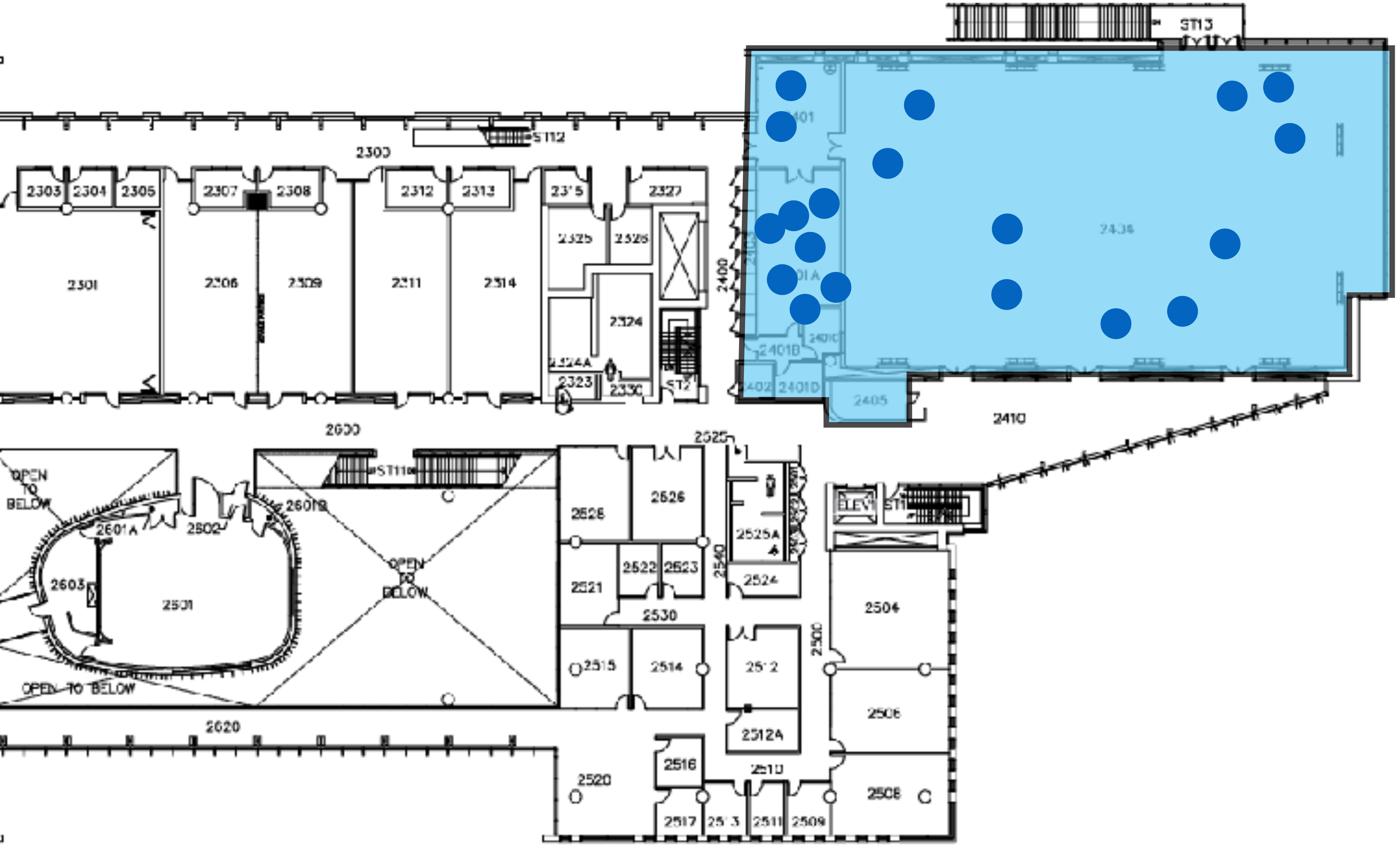


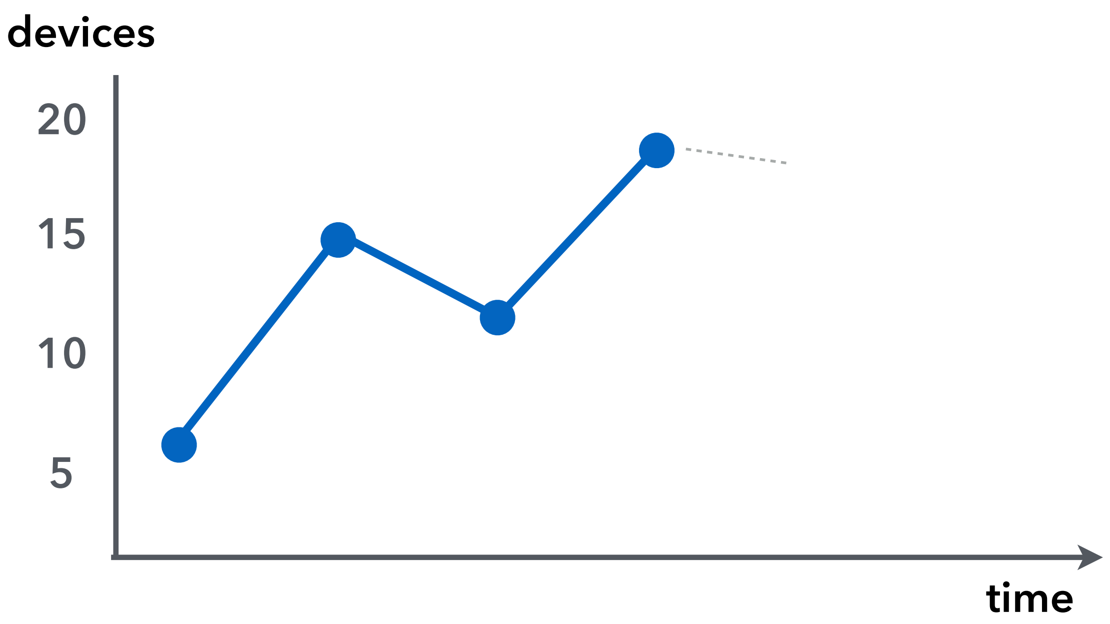
WiFi Connections: Location-Based Counts



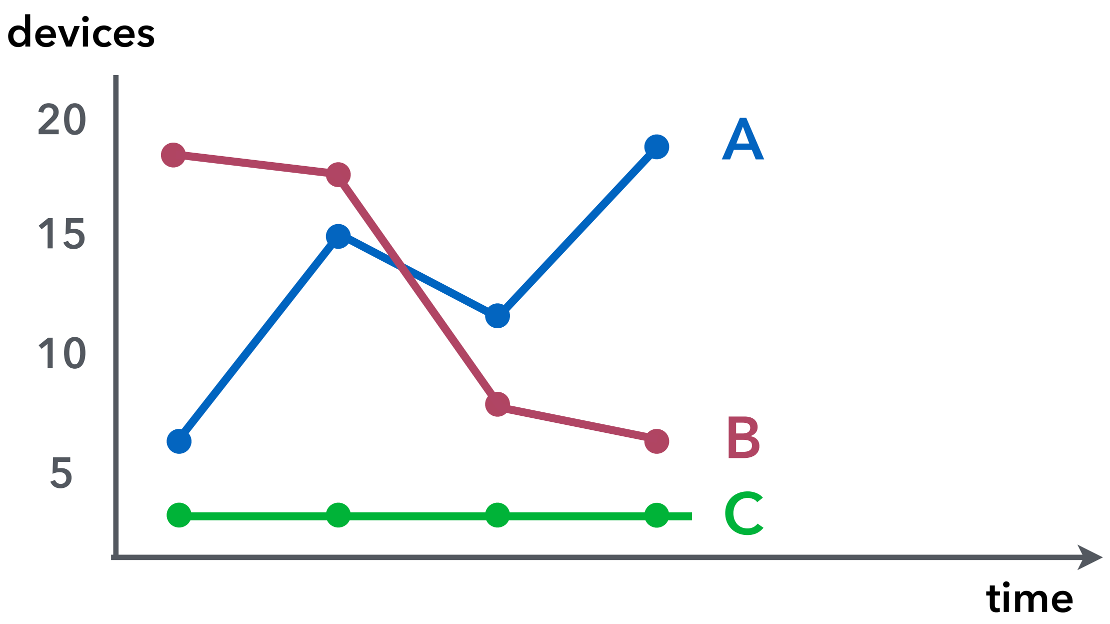


WiFi Connections: Location-Based Counts

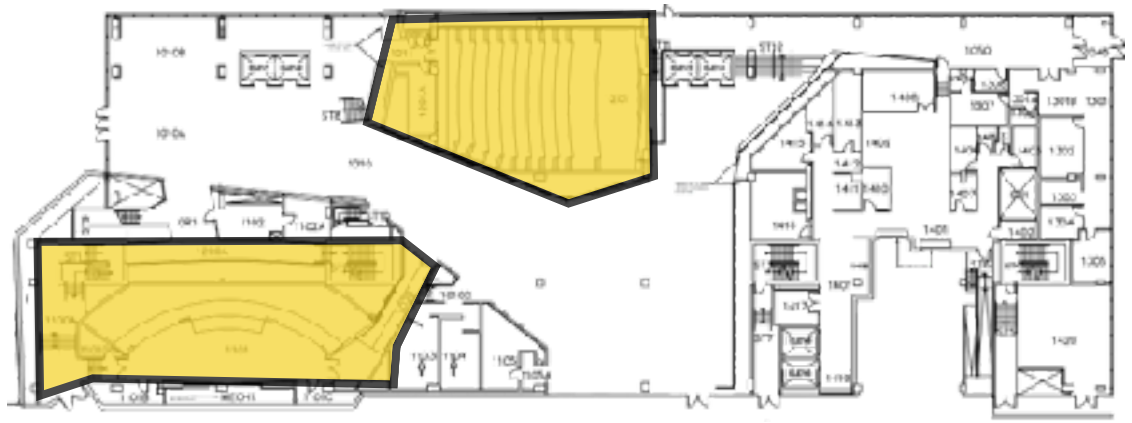


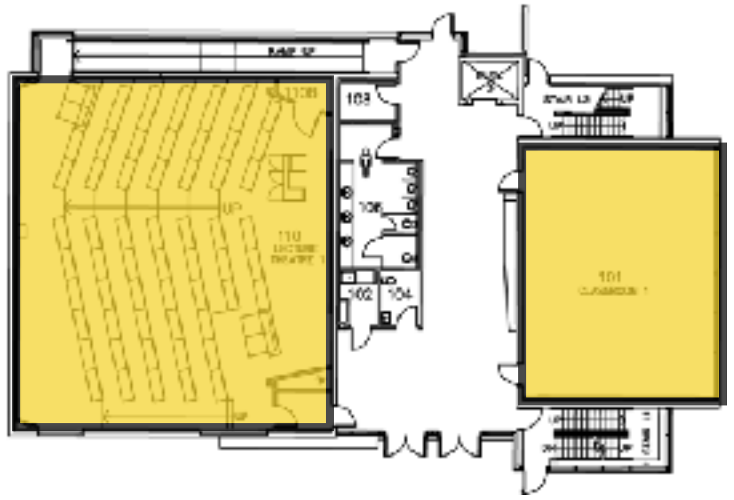


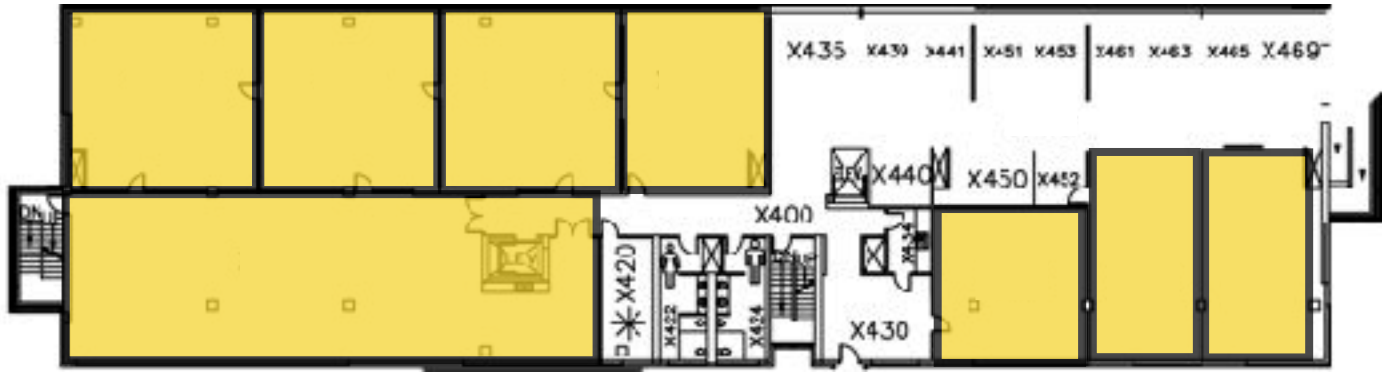
WiFi Connections: Location-Based Counts

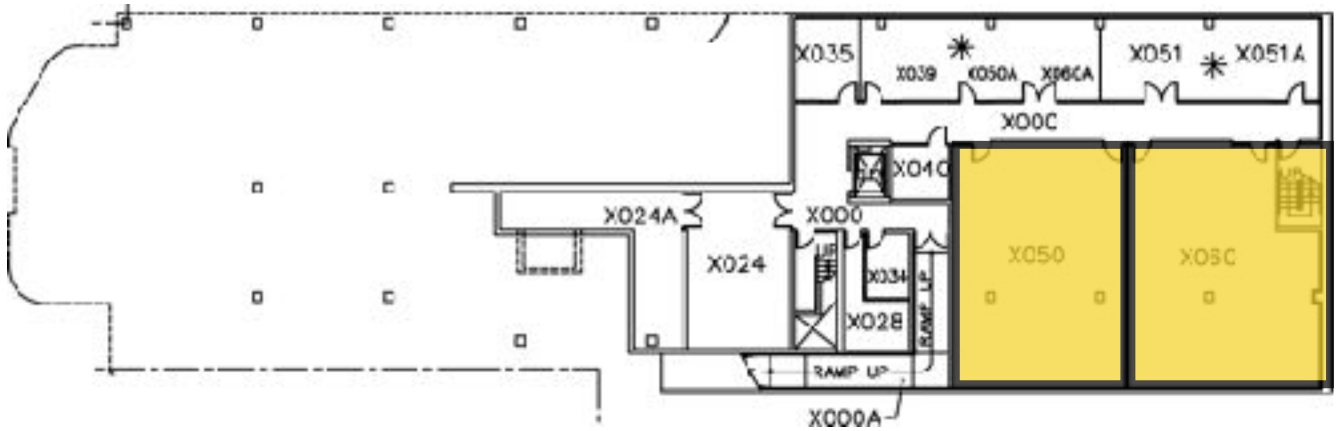


WiFi Connections: Location-Based Counts









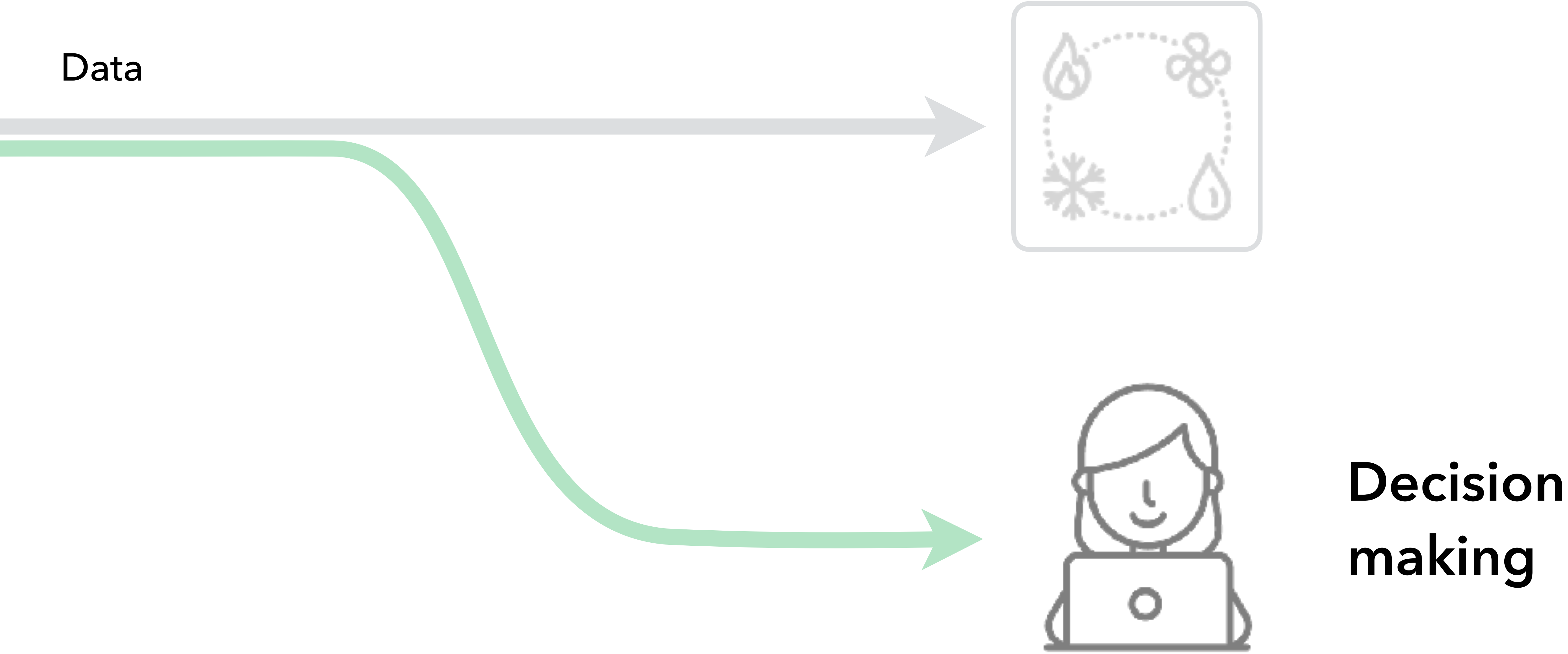
Location-Based Counts

- Regular intervals (e.g., every 5 minutes)
- Spatial hierarchy (Zone → Floor → Building → Campus)
- No trajectories or device identifiers are recorded
- Intrinsic privacy advantages

Data

Automated HVAC control

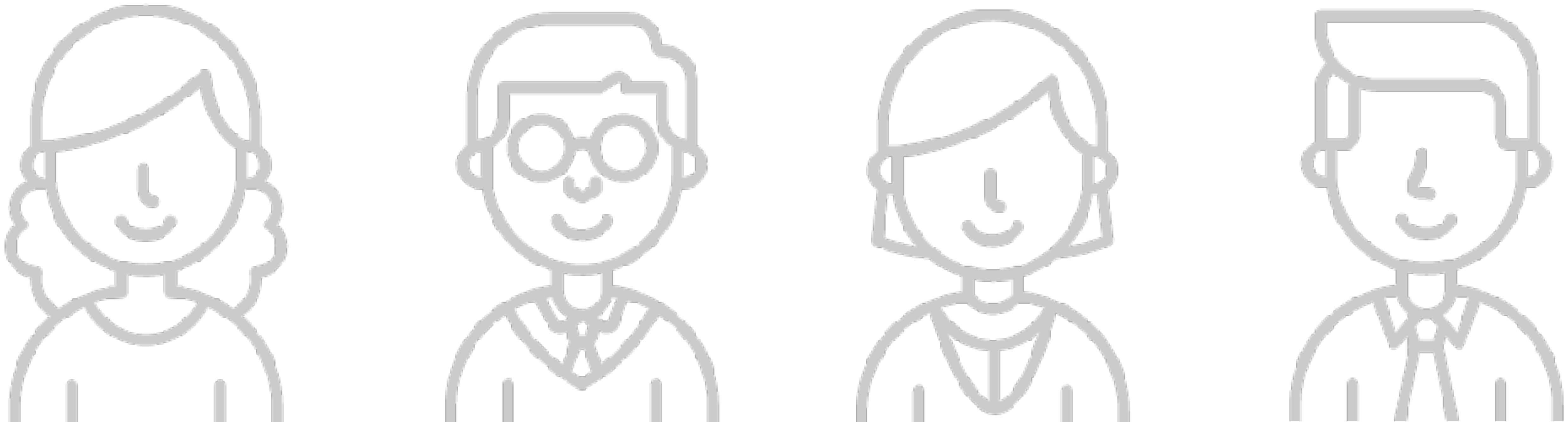
Data


Data
Decision
making

WiFi connections as a proxy for occupancy

WiFi connections as a proxy for occupancy

Interviews with potential stakeholders



Focus Domains

- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control

Focus Domains

- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control

Semi-structured discussions and live demos

Tasks

☑ **Confirm** assumptions or previous observations.

Do students occupy room x in evenings or on weekends?

Tasks

- ☑ **Confirm** assumptions or previous observations.
- ☑ **Monitor** the current/recent utilization rate.
  Which rooms are empty/busy?

Tasks

- **Confirm** assumptions or previous observations.
- **Monitor** the current/recent utilization rate.
- **Communicate** space usage and justify decisions.

  Space usage improved after renovation.

Tasks

- **Confirm** assumptions or previous observations.
- **Monitor** the current/recent utilization rate.
- **Communicate** space usage and justify decisions.
- **Validate** the data (quality control).
  Check minimum size of a room that can be captured.

Spatial and Temporal Data Granularities

Visualization Prototypes

Sandbox

Data sketches,
static data export

Time

Visualization Prototypes

Sandbox

Data sketches,
static data export

- original plan: different interface for each stakeholder
- realization: task & data abstractions match multiple stakeholders
  - if slice by space & time granularity

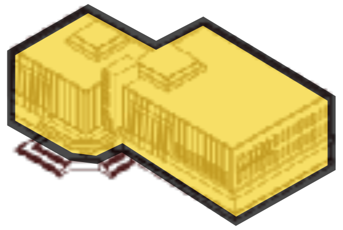
Spatial and Temporal Data Granularities

Regions of interest

Zone

Floor

Building



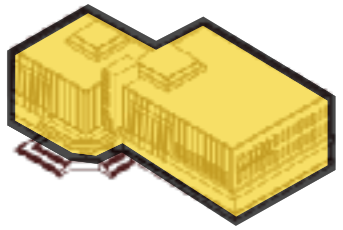
Spatial and Temporal Data Granularities

Regions of interest

Zone

Floor

Building



Periods of interest

Mondays

Weekdays

last 12 hours

Summer term

Fr 8-10am

Weekends

Visualization Prototypes

Sandbox

Data sketches,
static data export

Campus Explorer

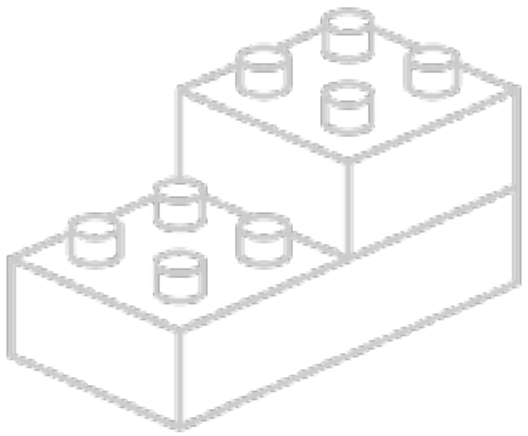
Live-data stream,
cross-building analysis

Building Recent

Building Long-term

Region Compare

Time

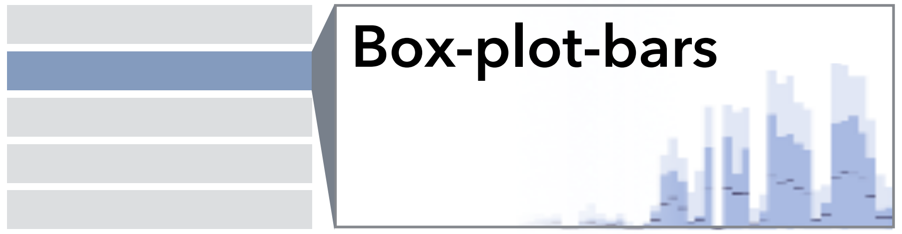


Reusable Visualization Components

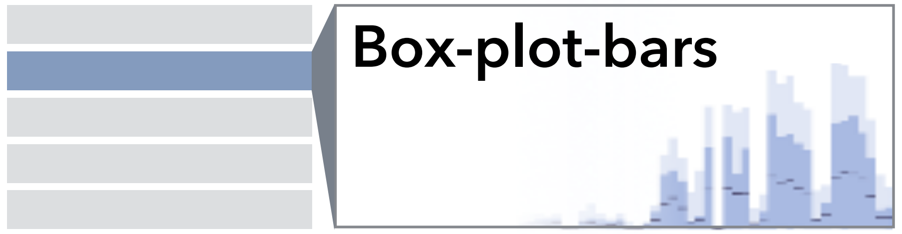
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)

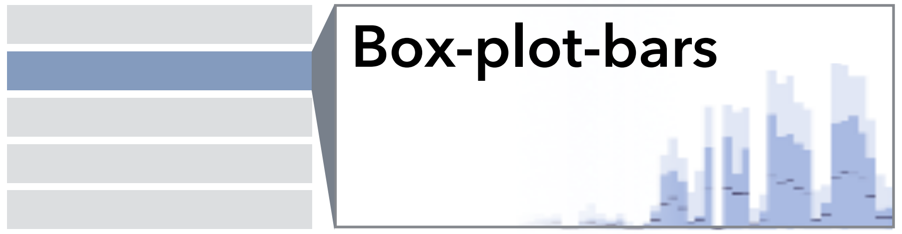
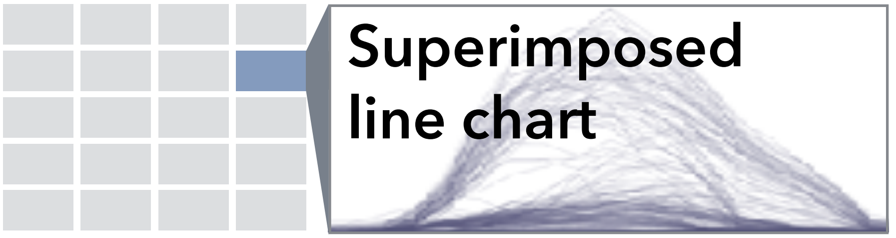
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	**Sparkline**	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	**Box-plot-bars**	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*

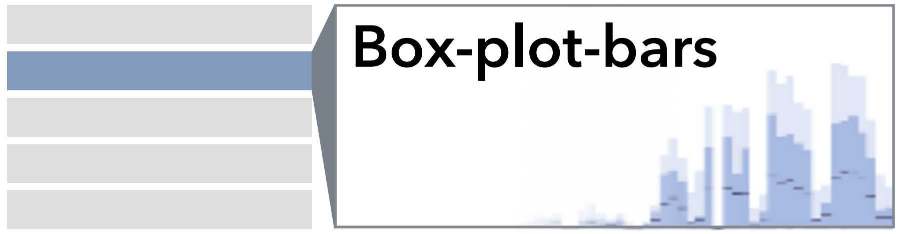
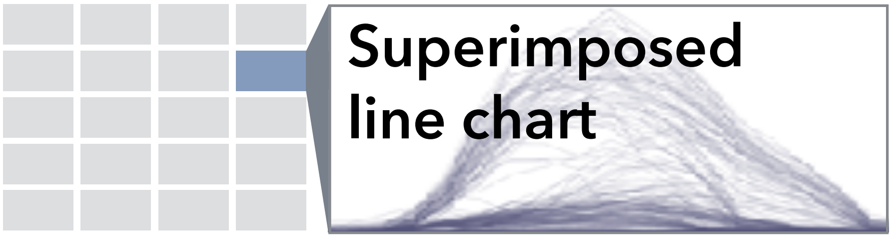
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	**Sparkline**	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	**Box-plot-bars**	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
	**Confidence band line chart**	Aggregation	Typical utilization profiles

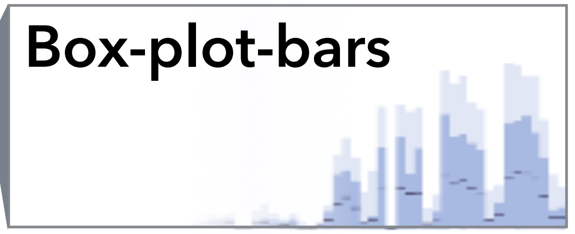
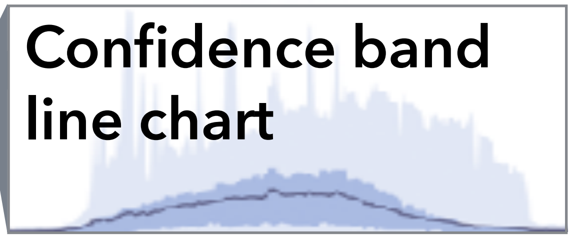
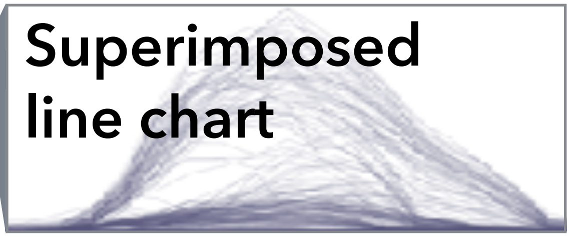
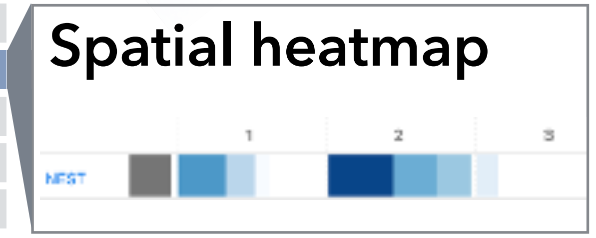
Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
	**Sparkline**	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	**Box-plot-bars**	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
	**Confidence band line chart**	Aggregation	Typical utilization profiles
	**Superimposed line chart**	Superposition	Within-session patterns, outliers

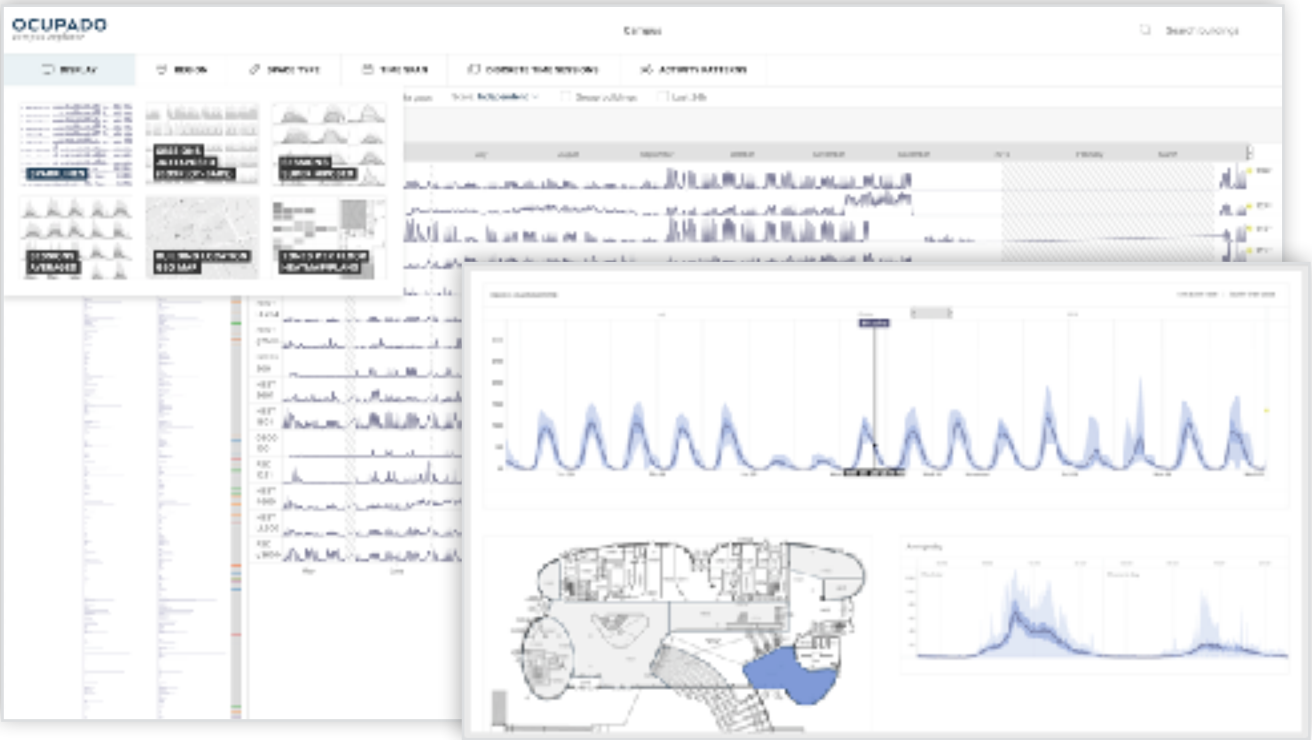
Reusable Visualization Components

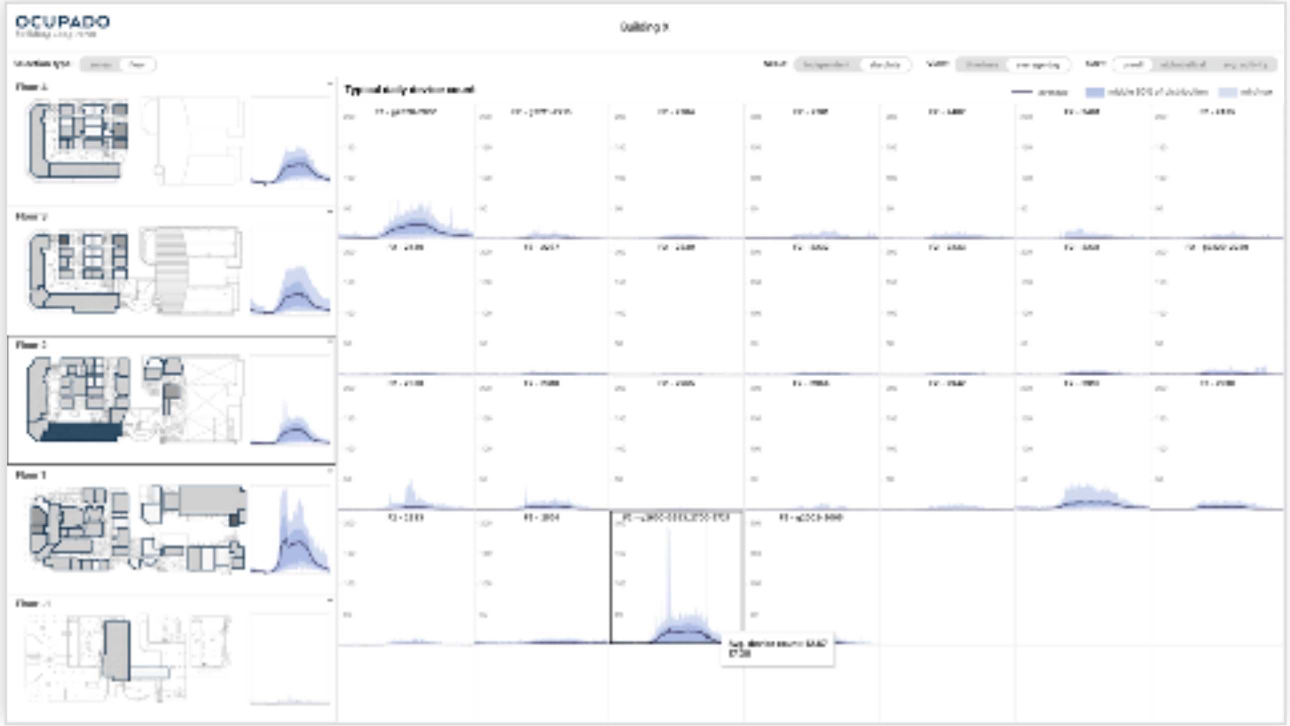
	Layout	Visual Encoding	Facet	Comparisons
Temporal		Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Box-plot-bars	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
		Confidence band line chart	Aggregation	Typical utilization profiles
		Superimposed line chart	Superposition	Within-session patterns, outliers

Reusable Visualization Components

	Layout	Visual Encoding	Facet	Comparisons
Temporal		**Sparkline** 	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Box-plot-bars 	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
		Confidence band line chart 	Aggregation	Typical utilization profiles
		Superimposed line chart 	Superposition	Within-session patterns, outliers
Spatial		**Floor plan with symbols** 	Superposition	Within local spatial neighborhood
		Spatial heatmap 	Containment (nested)	Across distributed regions

Ocupado Interfaces

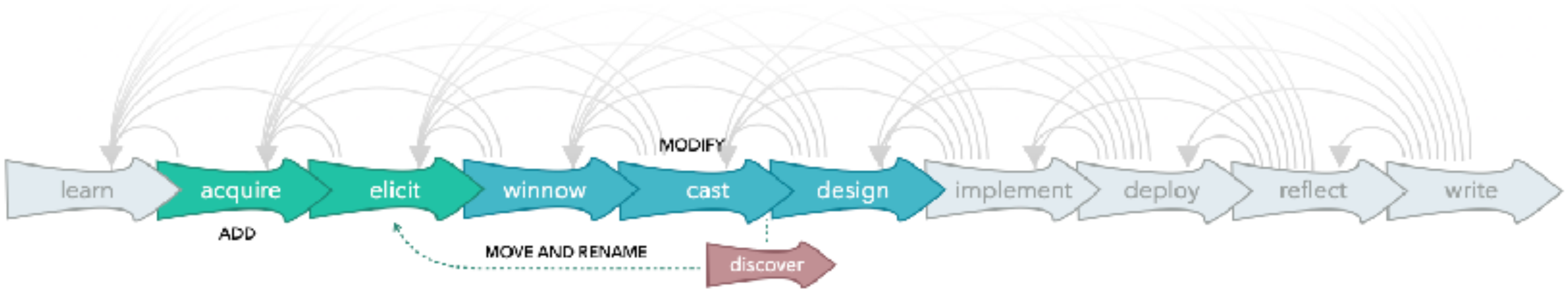




Ocupado Contributions

- Analysis and abstraction of data and tasks for studying space utilization
- Ocupado, a set of visual decision support tools
- Generalizable design choices for visualizing non-trajectory spatiotemporal data relating to large-scale indoor environments

Michael
Oppermann

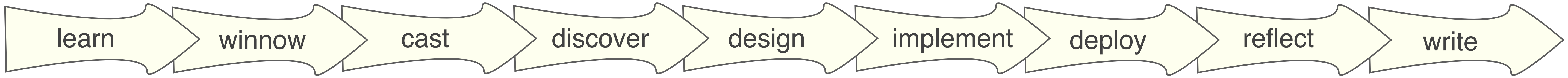


Data-First Design Studies

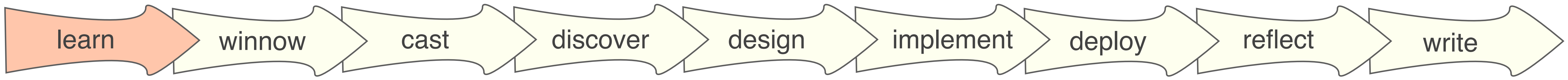
http://www.cs.ubc.ca/group/infovis/pubs/2020/data-first/

Data-First Design Studies.
Oppermann and Munzner. Proc. IEEE VIS BELIV Workshop 2020.

Original DSM framework

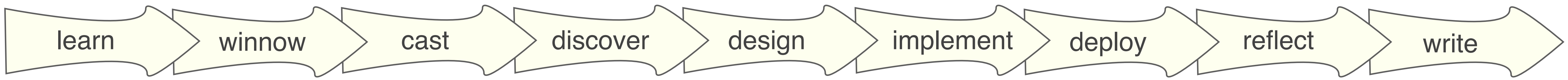


Original DSM framework

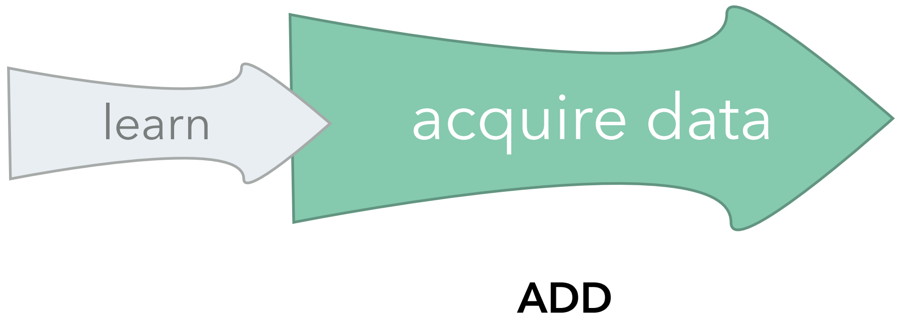


Data-first DSM framework

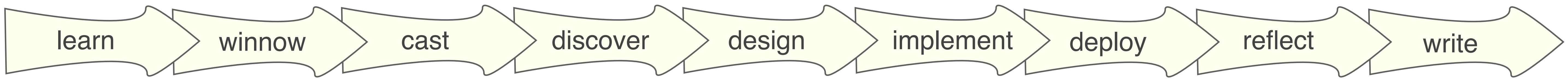
Original DSM framework



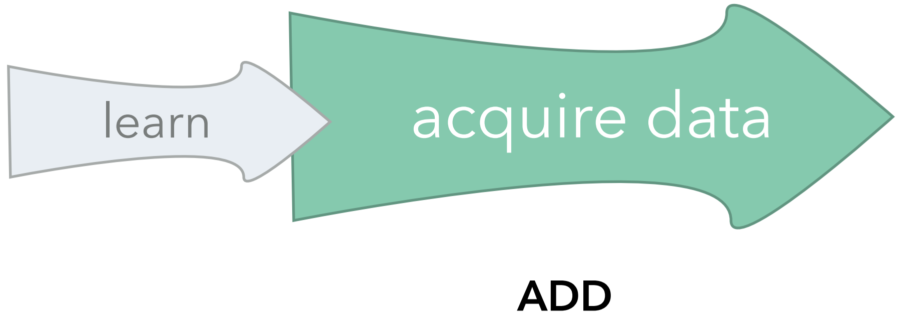
Data-first DSM framework



Original DSM framework

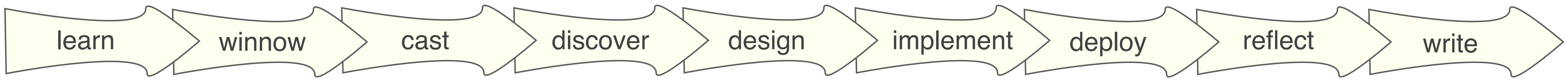


Data-first DSM framework



- What type of data am I working with?

Original DSM framework

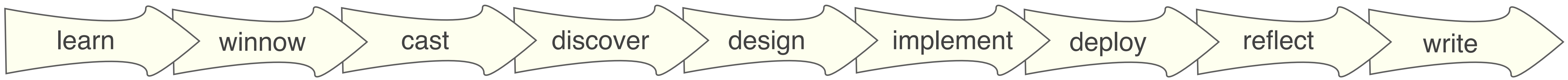


Data-first DSM framework

ADD

- What type of data am I working with?
- Are there any data quality challenges?

Original DSM framework

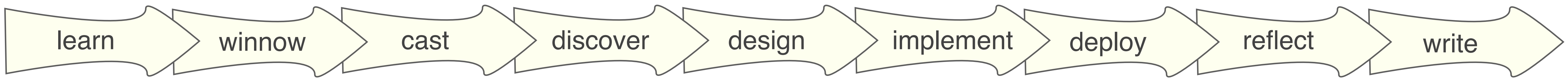


Data-first DSM framework

ADD

- What type of data am I working with?
- Are there any data quality challenges?
- What is special about this data?

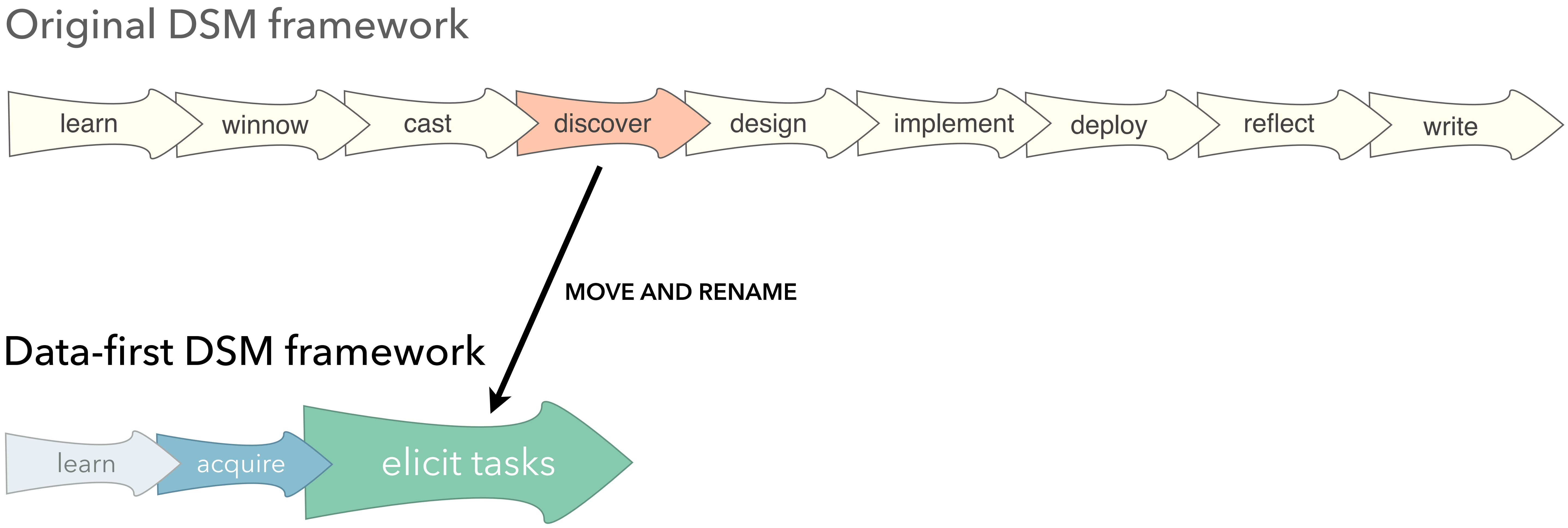
Original DSM framework



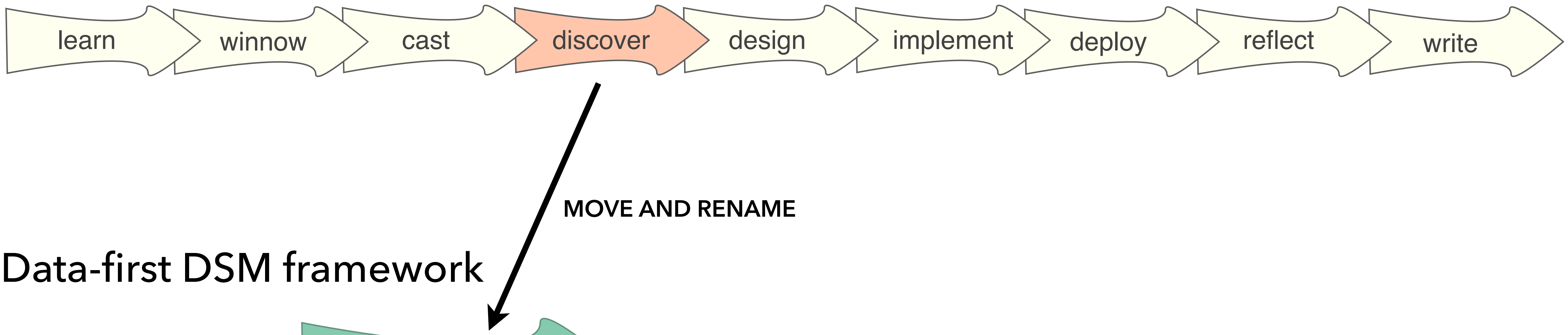
Data-first DSM framework

ADD

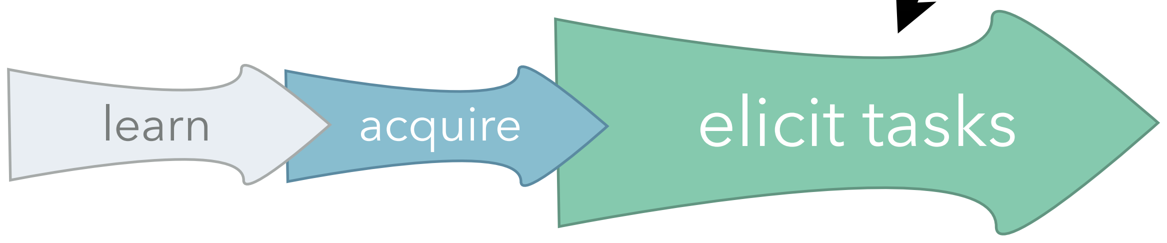
- What type of data am I working with?
- Are there any data quality challenges?
- What is special about this data?
- Who would benefit from seeing and exploring it?


Original DSM framework
learn
winnow
cast
discover
design
implement
deploy
reflect
write
MOVE AND RENAME
Data-first DSM framework
learn
acquire
elicit tasks

Original DSM framework

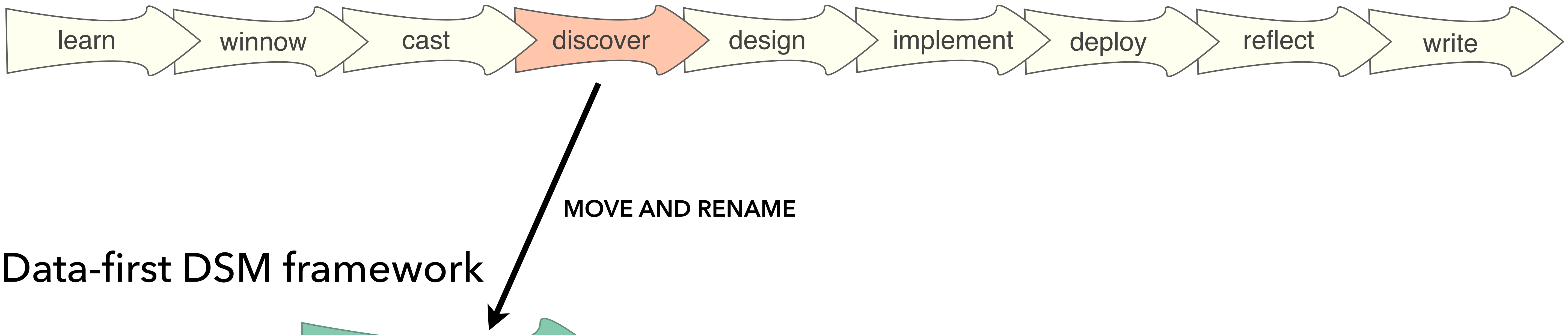


Data-first DSM framework

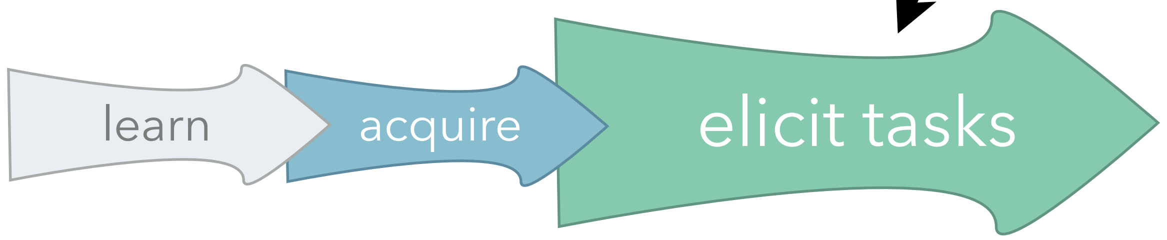


- Multiple potential stakeholders

Original DSM framework

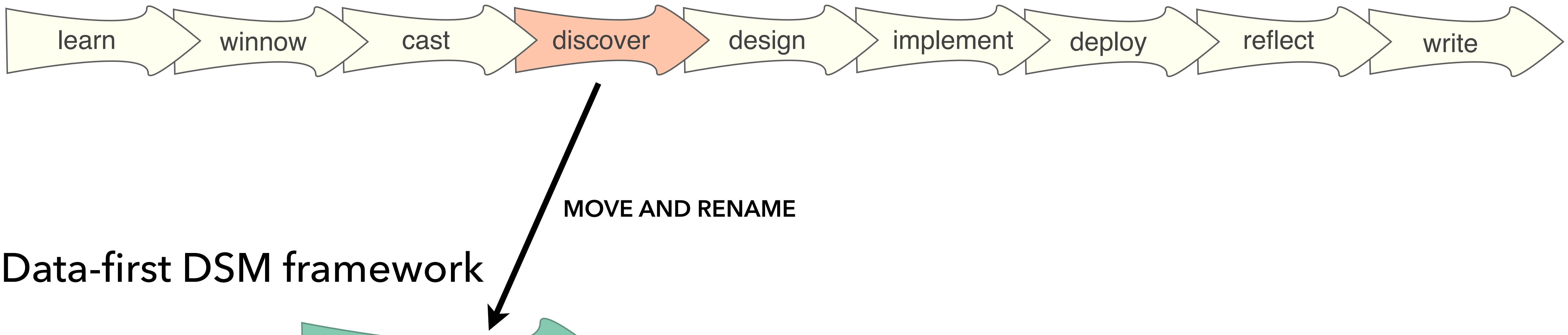


Data-first DSM framework

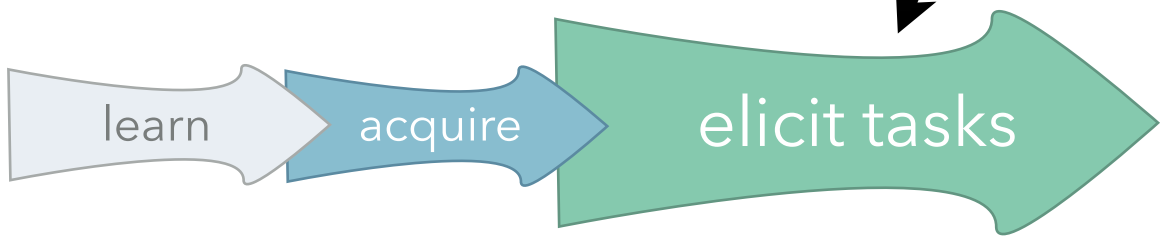


- Multiple potential stakeholders
- Explain initial data abstractions

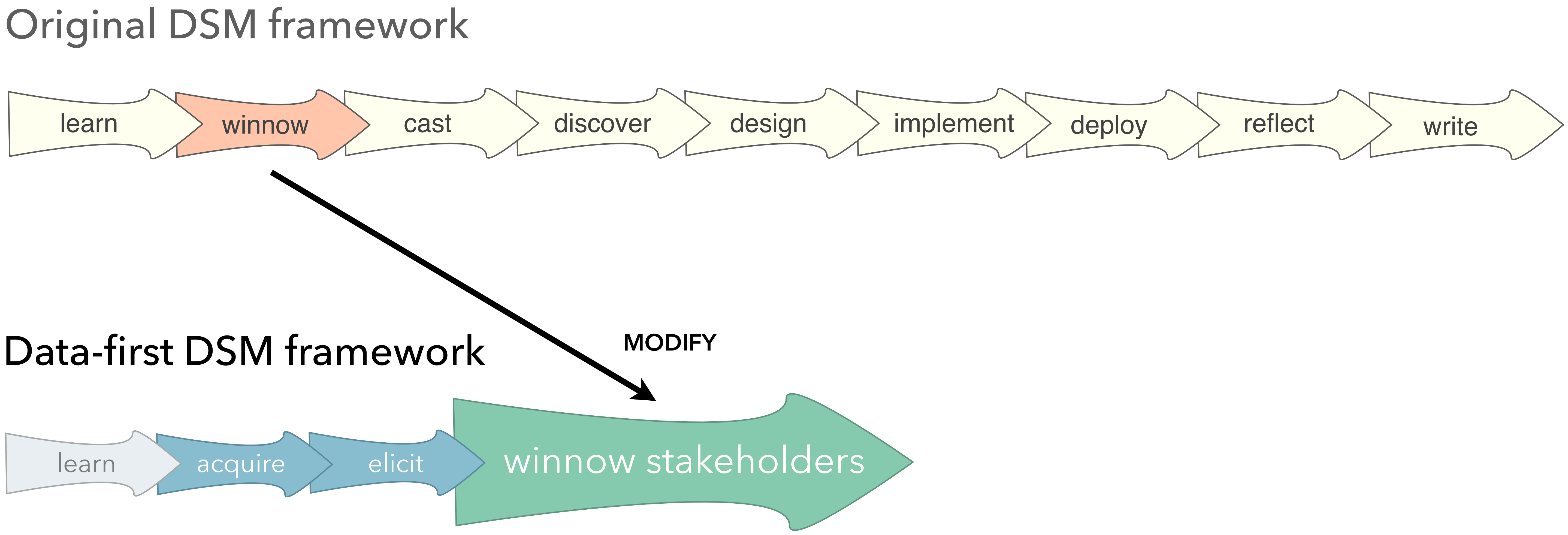
Original DSM framework



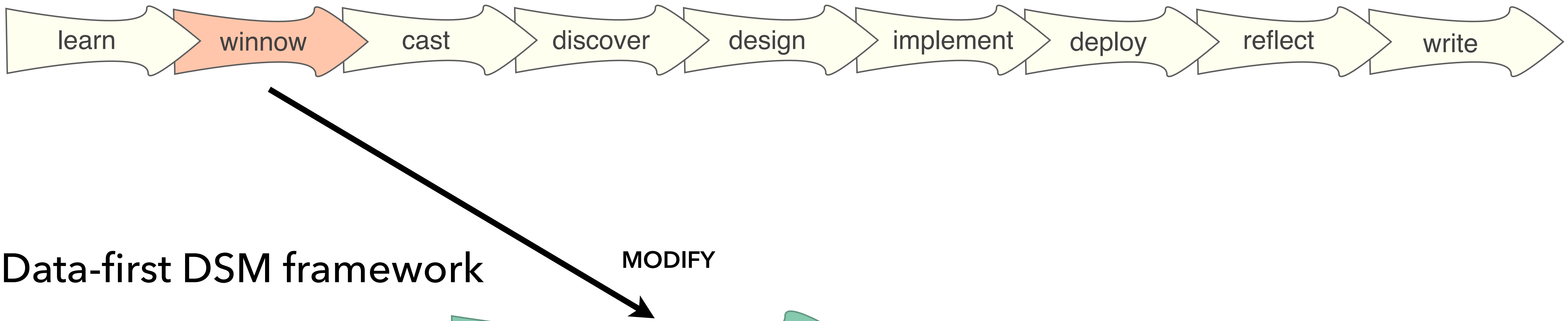
Data-first DSM framework



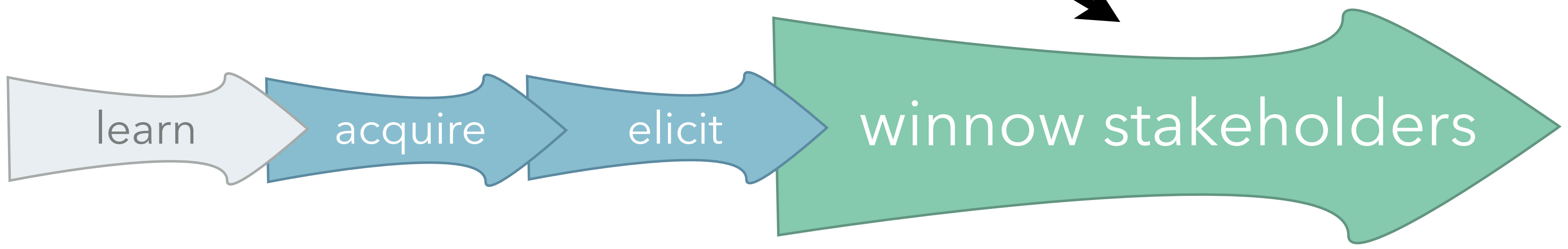
- Multiple potential stakeholders
- Explain initial data abstractions
- Learn about unsolved stakeholder needs


Original DSM framework
learn
winnow
cast
discover
design
implement
deploy
reflect
write
MODIFY
Data-first DSM framework
learn
acquire
elicit
winnow stakeholders

Original DSM framework

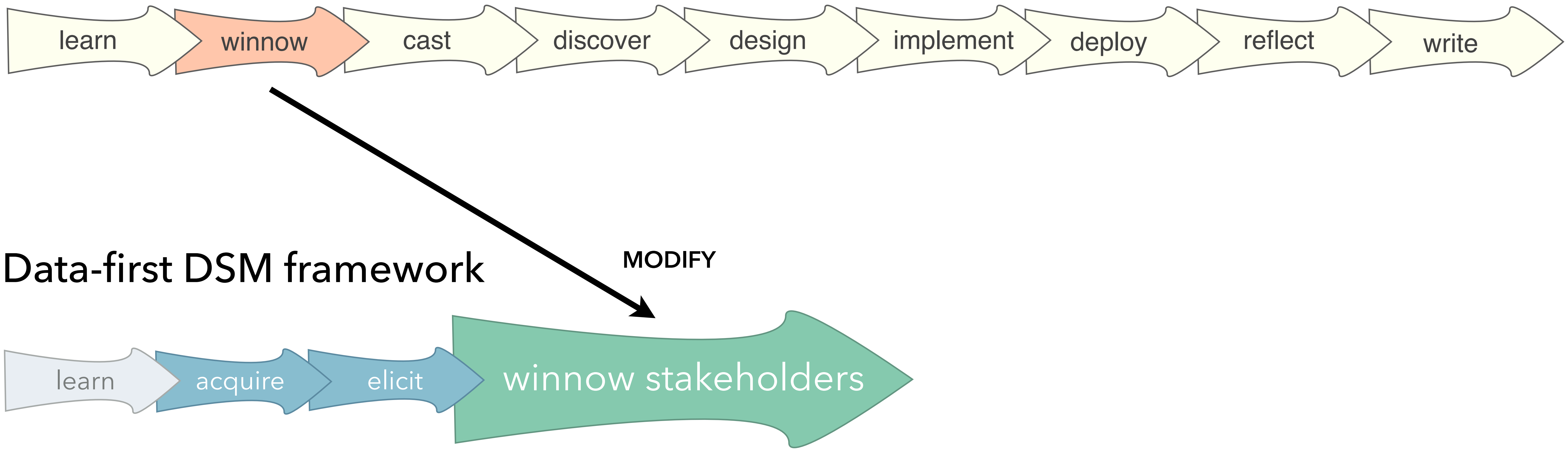


Data-first DSM framework



- How frequent are their data-relevant tasks?

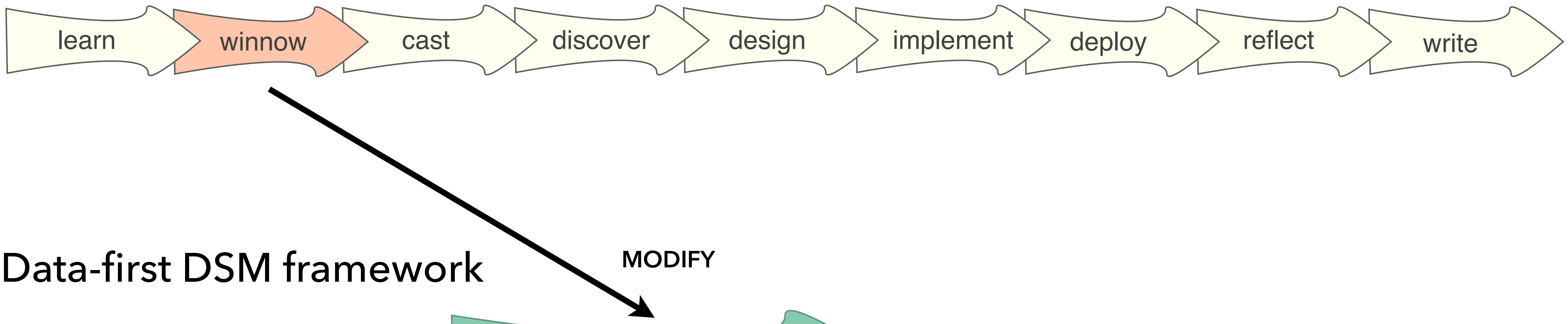
Original DSM framework



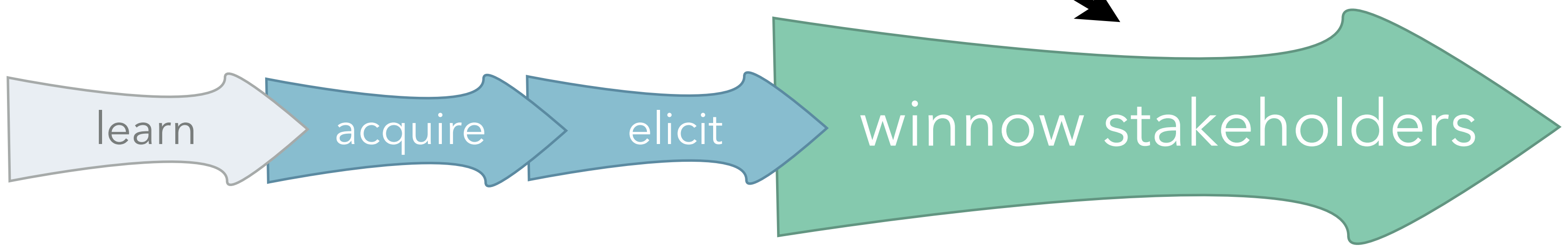
Data-first DSM framework

- How frequent are their data-relevant tasks?
- How central are these tasks to the stakeholder's primary mission?

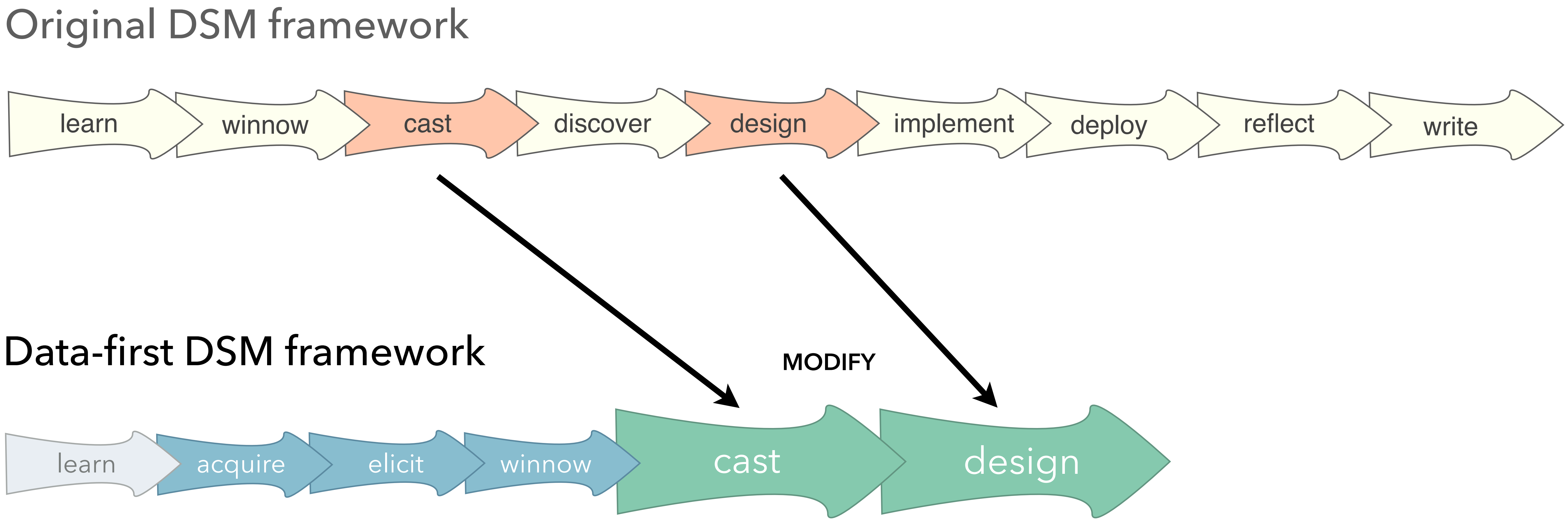
Original DSM framework



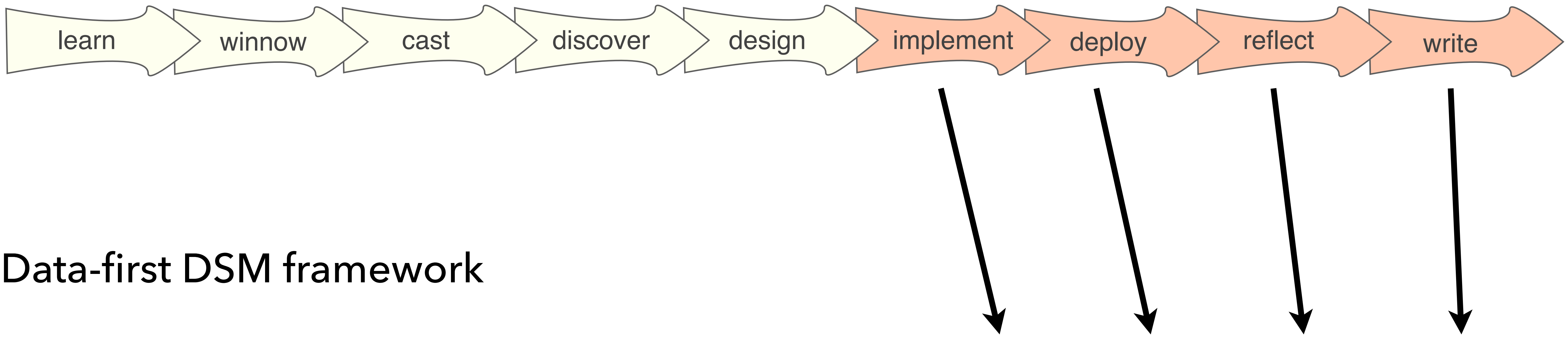
Data-first DSM framework



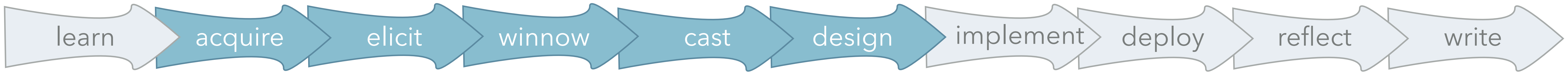
- How frequent are their data-relevant tasks?
- How central are these tasks to the stakeholder's primary mission?
- How many people in the organization deal with these tasks?


Original DSM framework
learn
winnow
cast
discover
design
implement
deploy
reflect
write
Data-first DSM framework
MODIFY
learn
acquire
elicit
winnow
cast
design

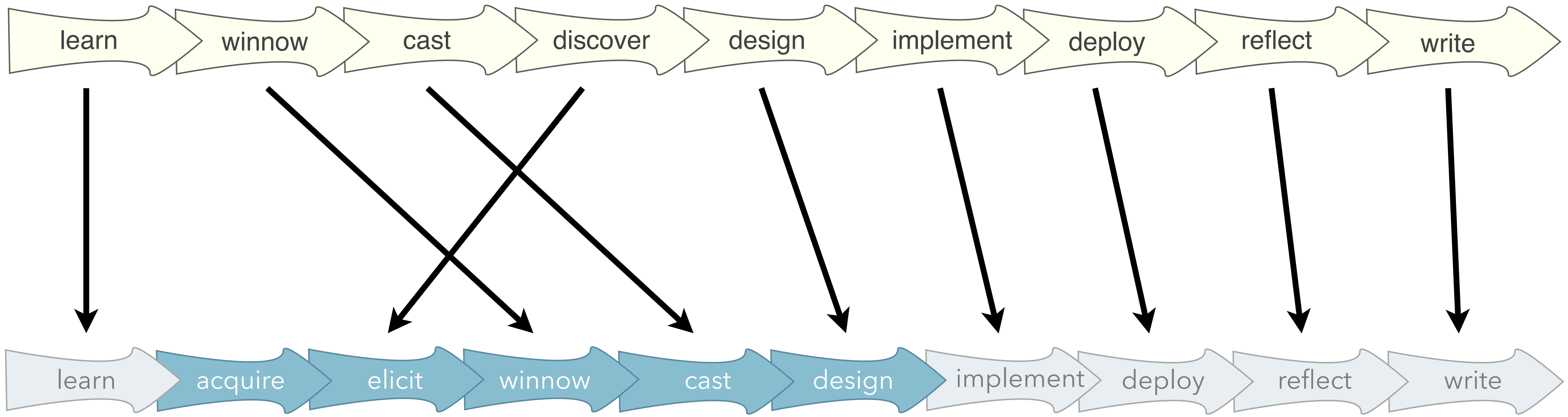
Original DSM framework



Data-first DSM framework



Original DSM framework



Data-first DSM framework

Three case studies of problem-driven work

- e-commerce

- facilities management

- biology

e commerce by shashank singh from the Noun Project   Biology by tezar tantular from the Noun Project   Business by Colourcreatype from the Noun Project

Zipeng Liu

Shing Hei Zhan

Aggregated Dendrograms

for Visual Comparison Between Many Phylogenetic Trees

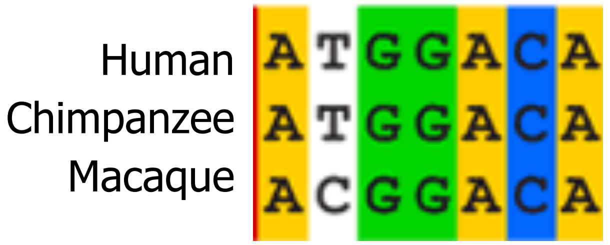
http://www.cs.ubc.ca/labs/imager/tr/2019/adview

Aggregated Dendrograms for Visual Comparison Between Many Phylogenetic Trees.

Liu, Zhan, Munzner. IEEE Trans. Visualization and Computer Graphics (TVCG) 26(9):2732-2747, 2019.

Phylogenetic tree

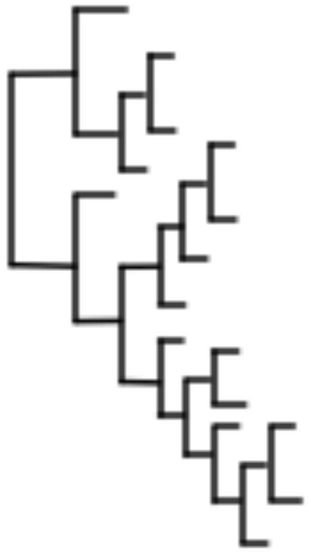
Evolutionary relationships of organisms



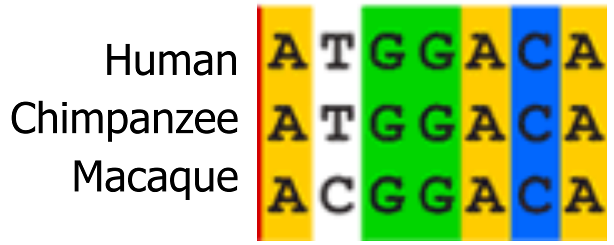
Genetic information

Computational workflow

Phylogenetic tree



Many phylogenetic trees

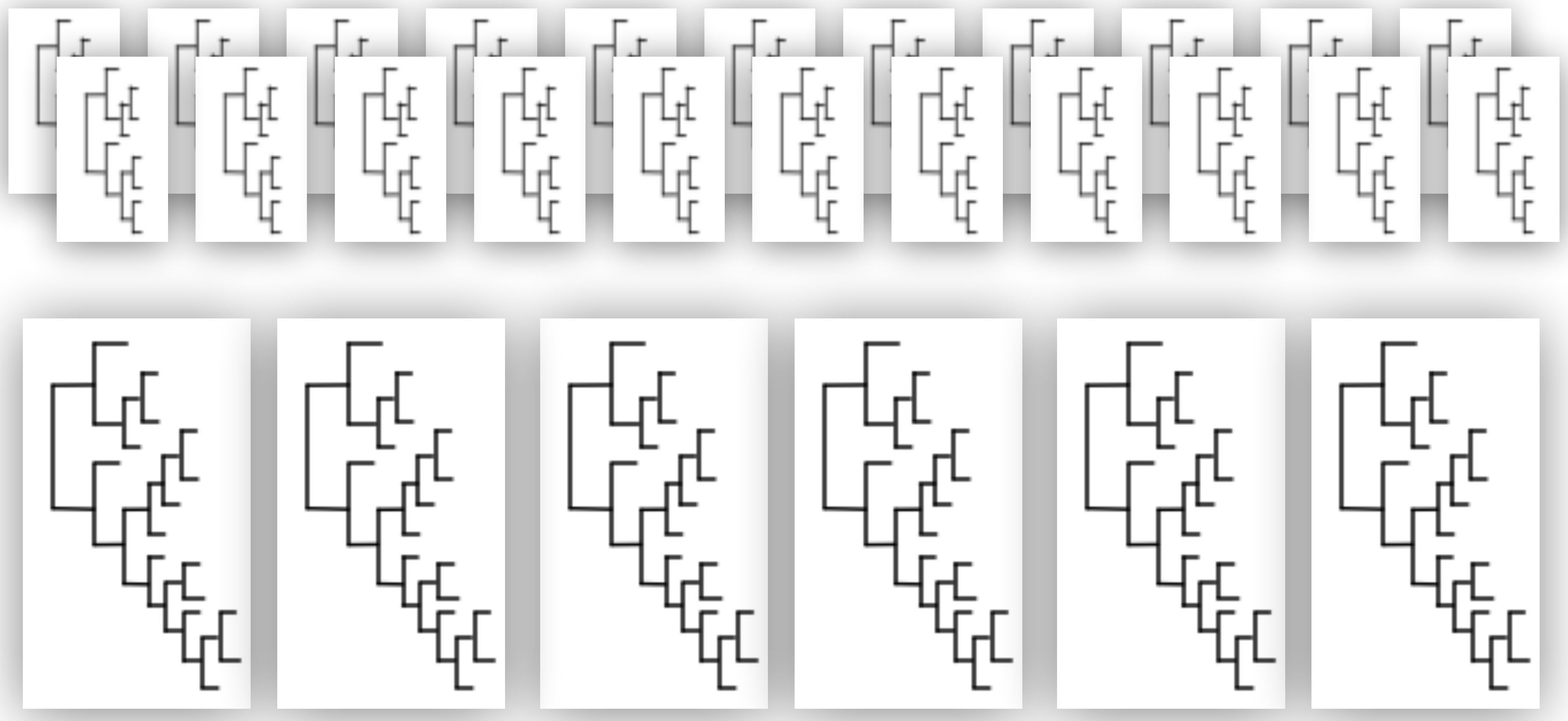


Genetic information

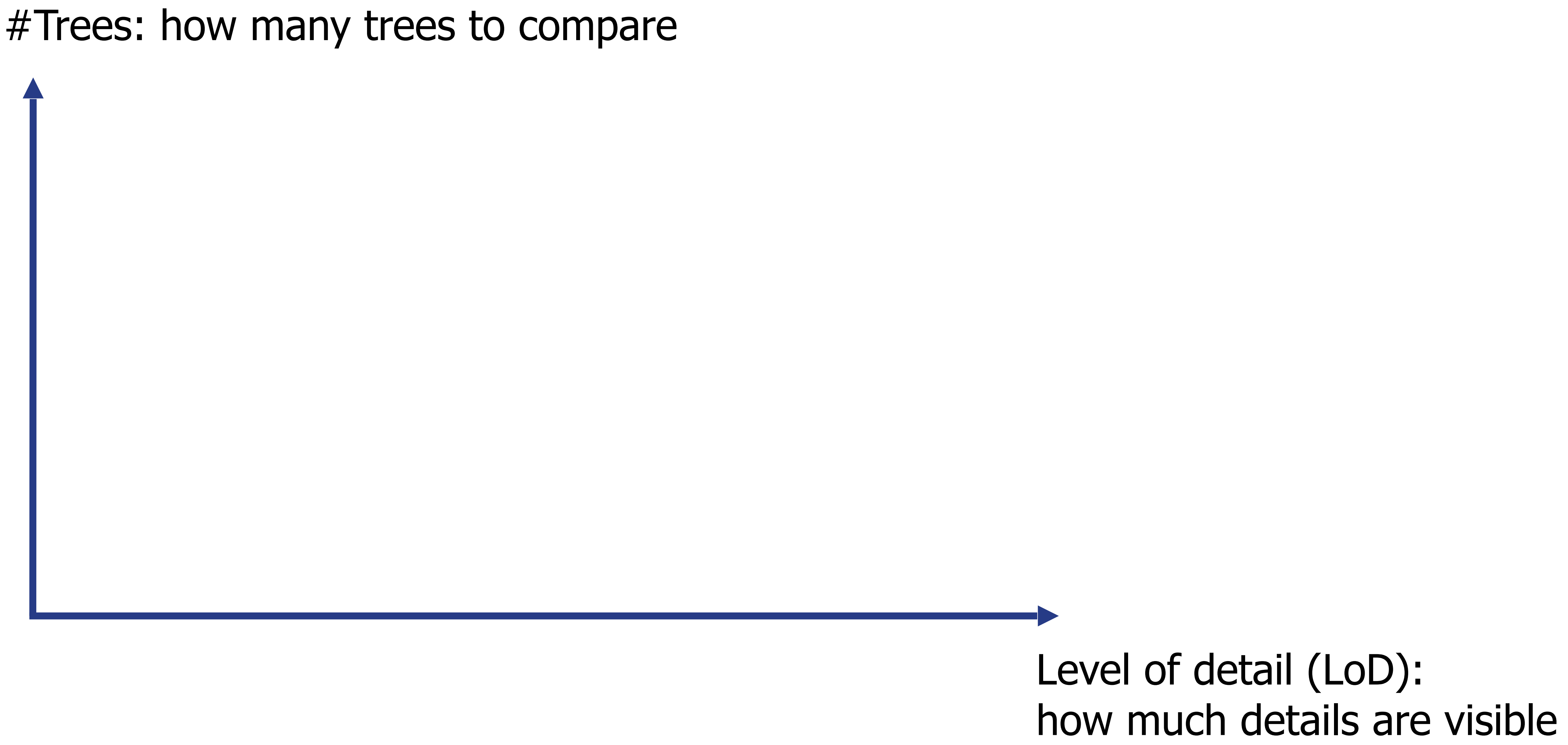
Computational workflow

Phylogenetic tree

- Understand relationships between genes and species trees
- Explore trees generated with different methods and data



Scalability of Existing Tree Comparison Systems



Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare

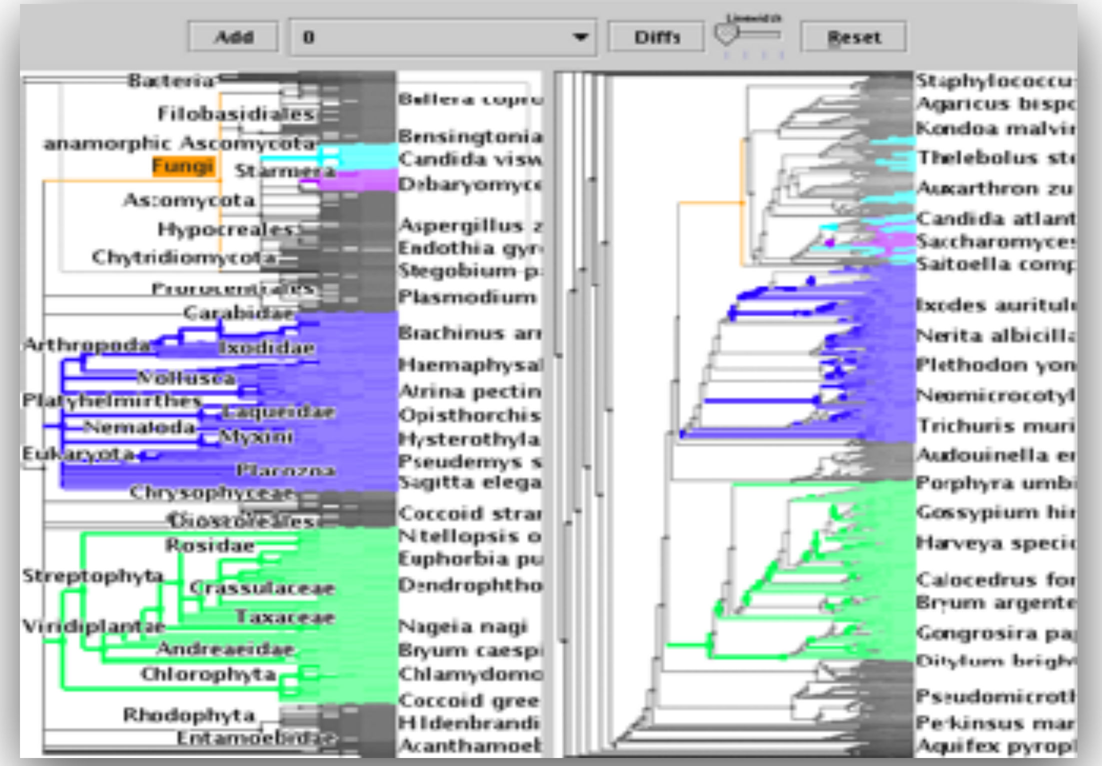
Pairs

Few in full

Simplified structure

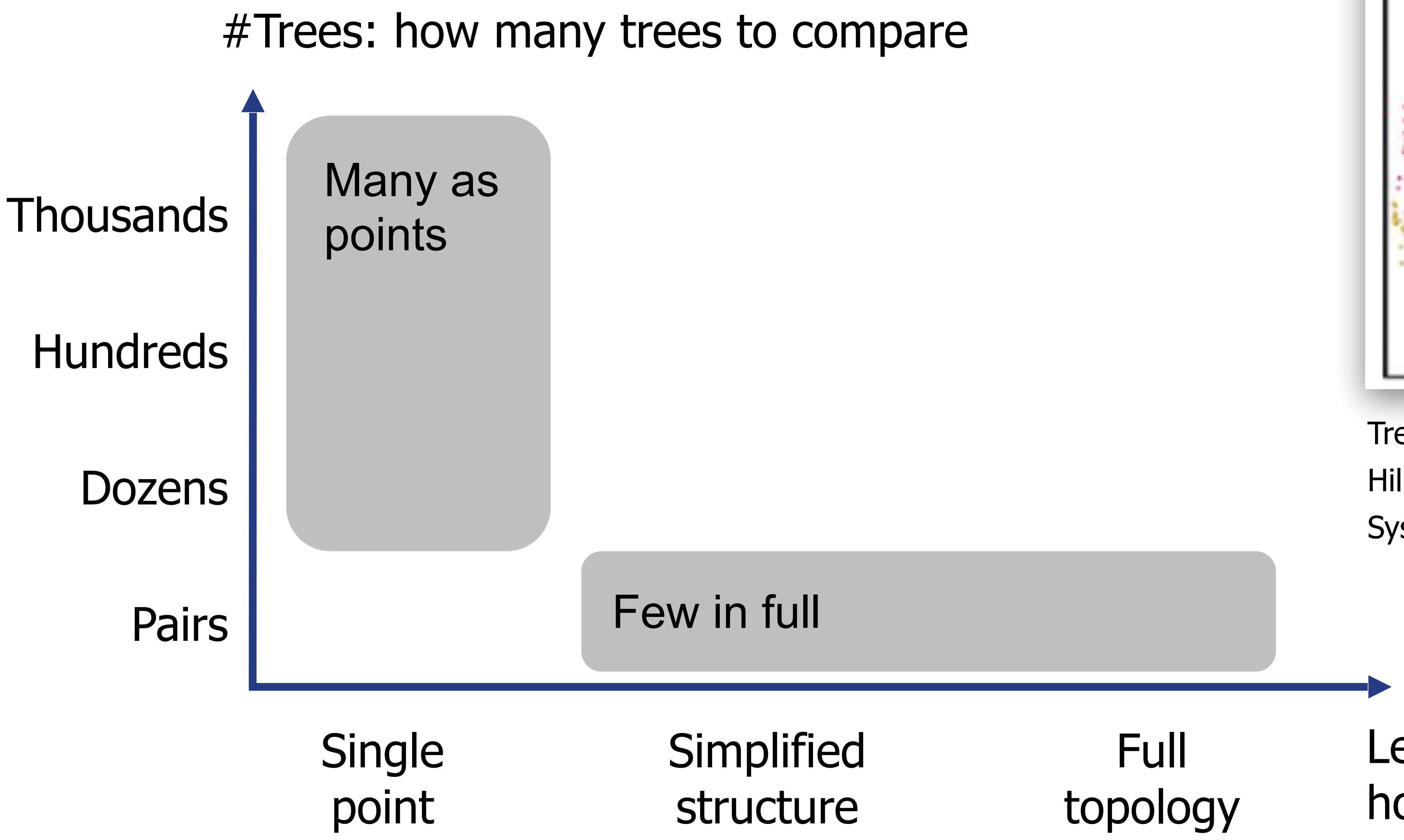
Full topology

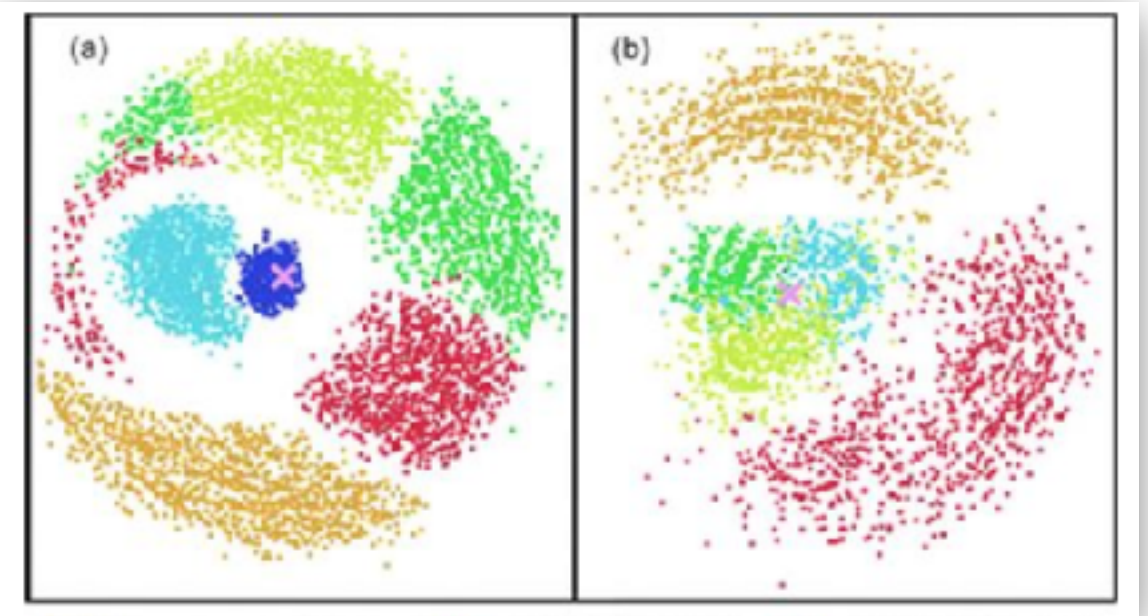
Level of detail (LoD):
how much details are visible



TreeJuxtaposer.
Munzner, Guimbretière, Zhang, Zhou.
SIGGRAPH 2003

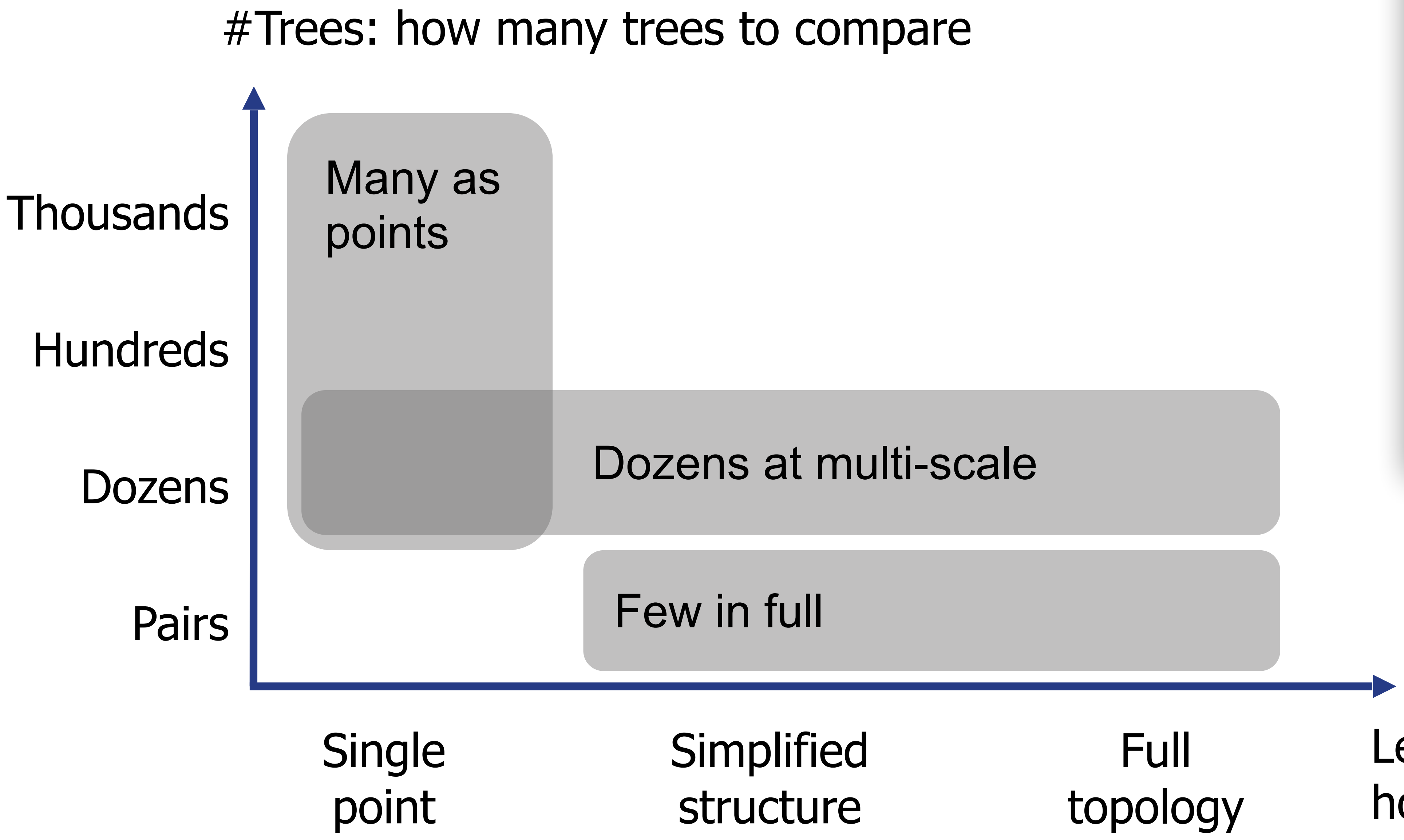
Scalability of Existing Tree Comparison Systems

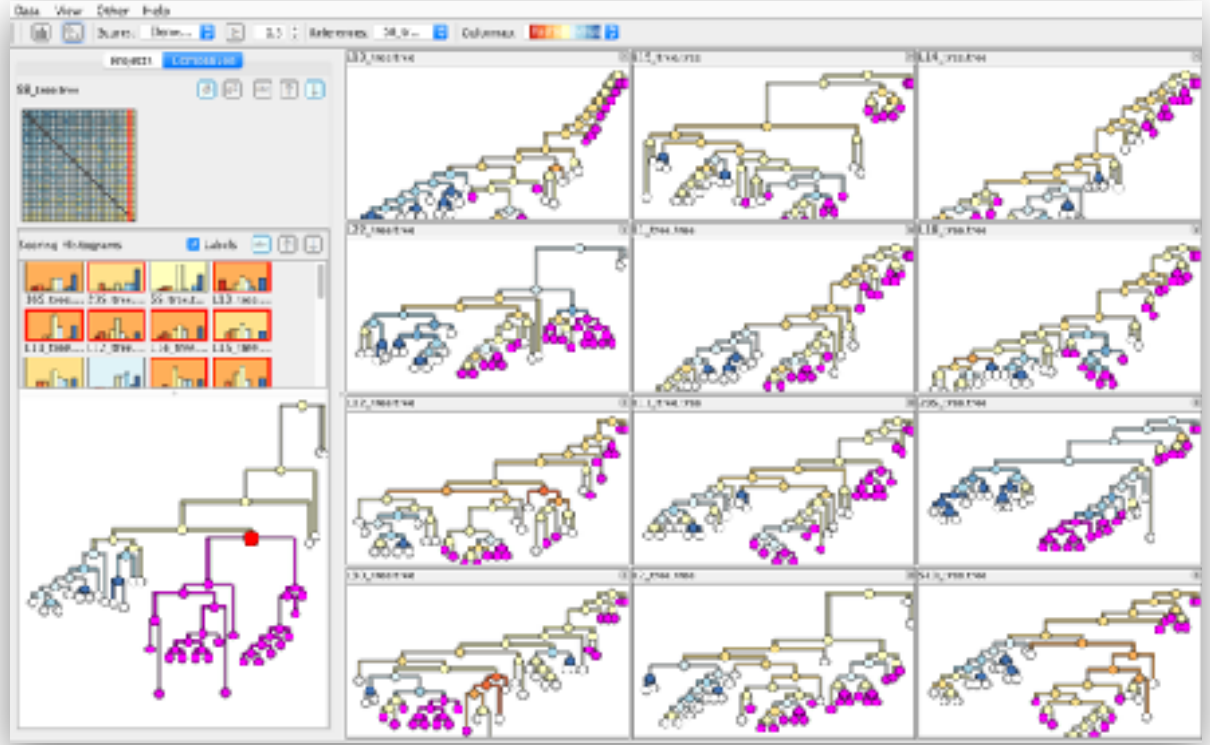




Tree space.
Hillis, Health, John.
Systematic Biology 2005.

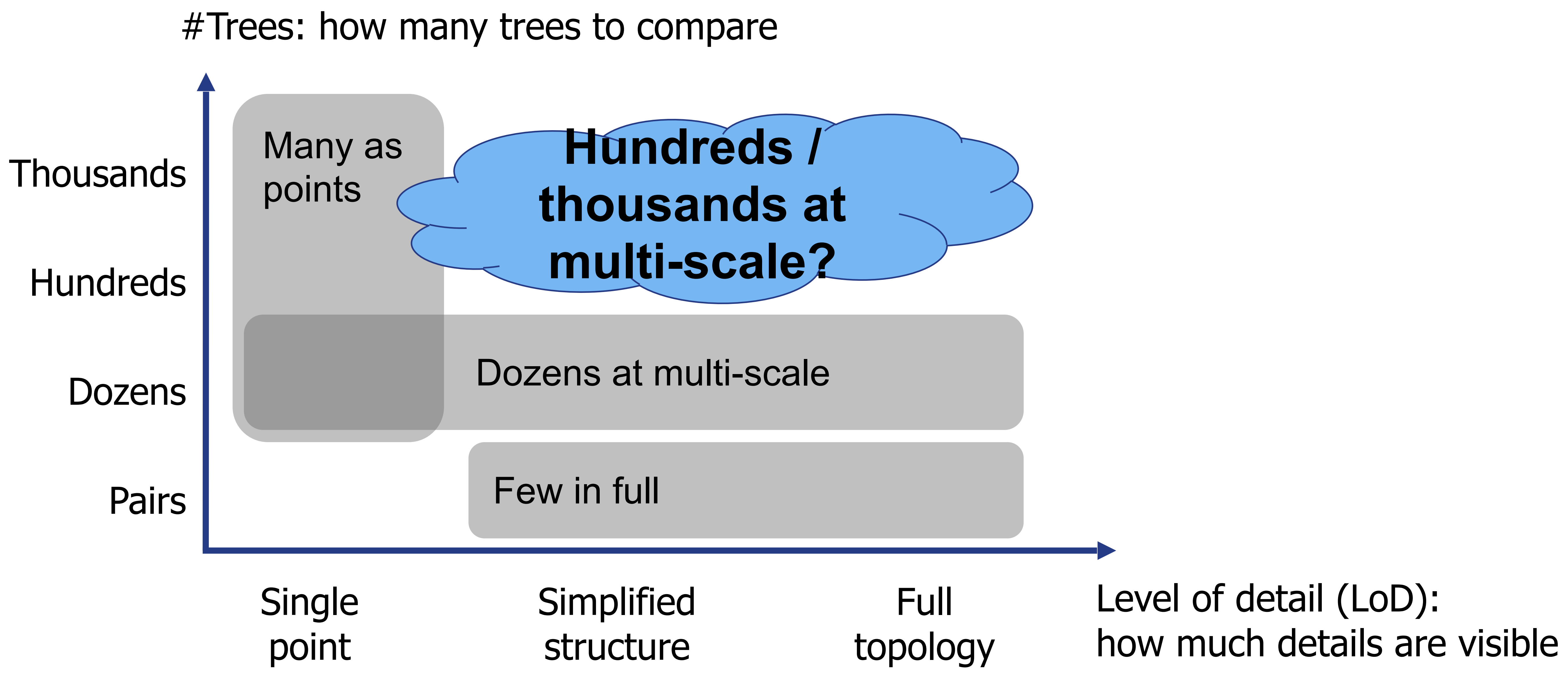
Scalability of Existing Tree Comparison Systems





Interactive visual comparison of multiple trees.
Bremm, Landesberger, Heß, Schreck, Weil, Hamacher.
VAST 2011.

Comparing many phylogenetic trees

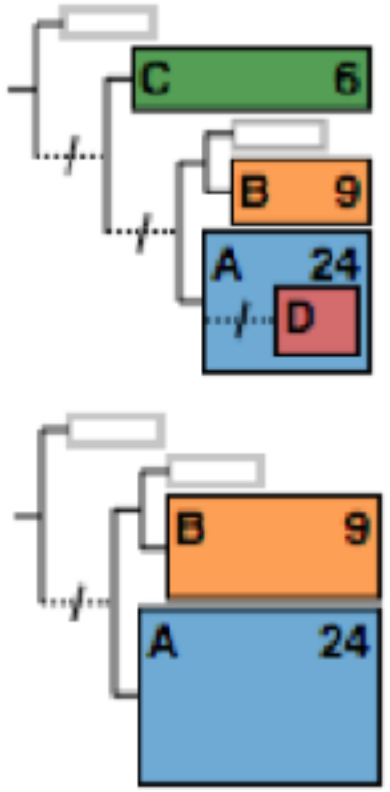


Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees

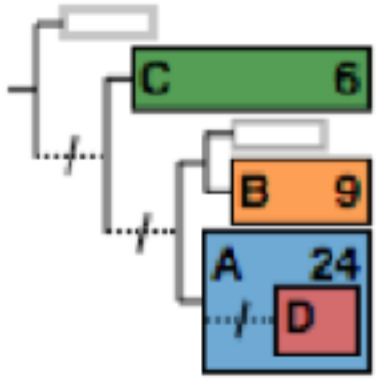
Contributions include idiom & algorithm levels

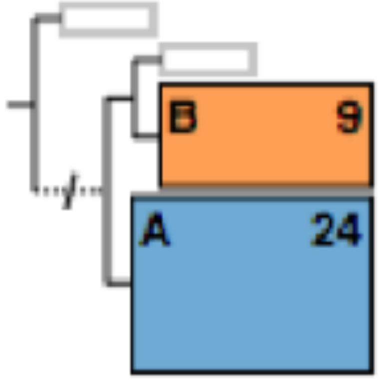
- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
  - Compact tree representation that focuses on selected subtrees
  - Adapts to available screen space



Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
  - Compact tree representation that focuses on selected subtrees
  - Adapts to available screen space
- A multi-view interactive tool: **ADView**
  - Covers multiple levels of details for tree comparison



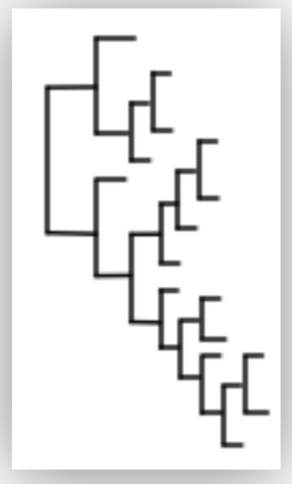


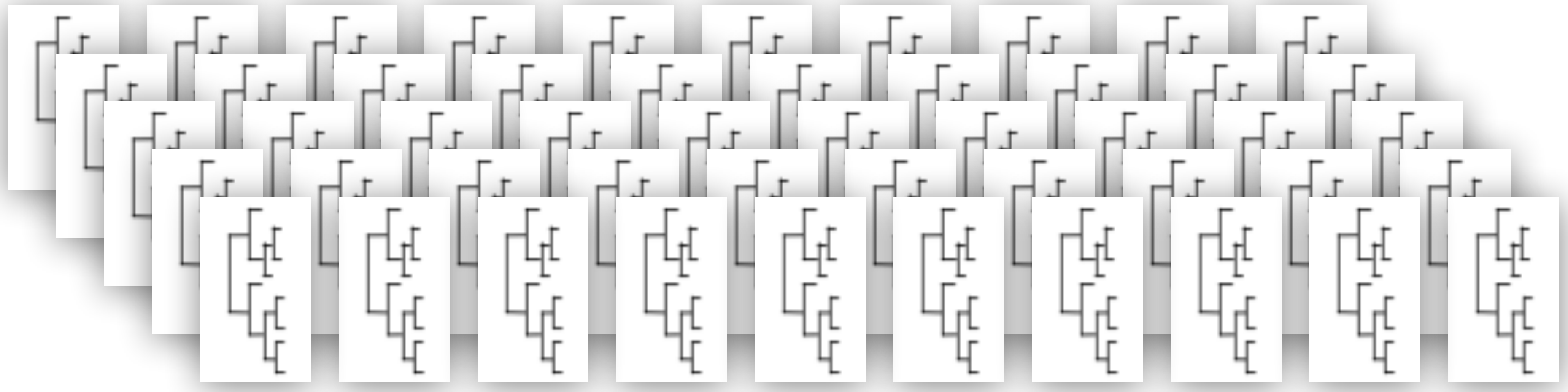
Data & Tasks

- Tree data
- Two crucial tasks

Tree data

Reference tree vs. Tree collection



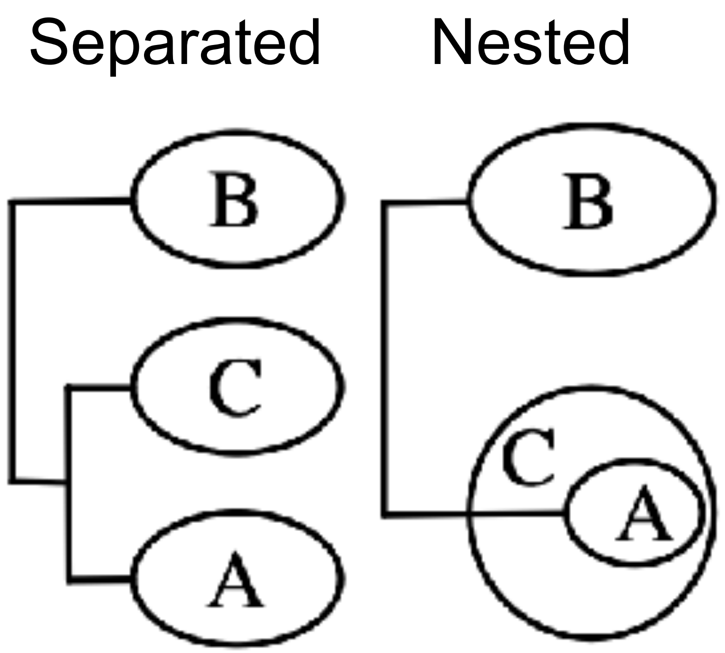


Two crucial tasks

Topological relationships between subtrees / leaf nodes

Two crucial tasks

Topological relationships between subtrees / leaf nodes

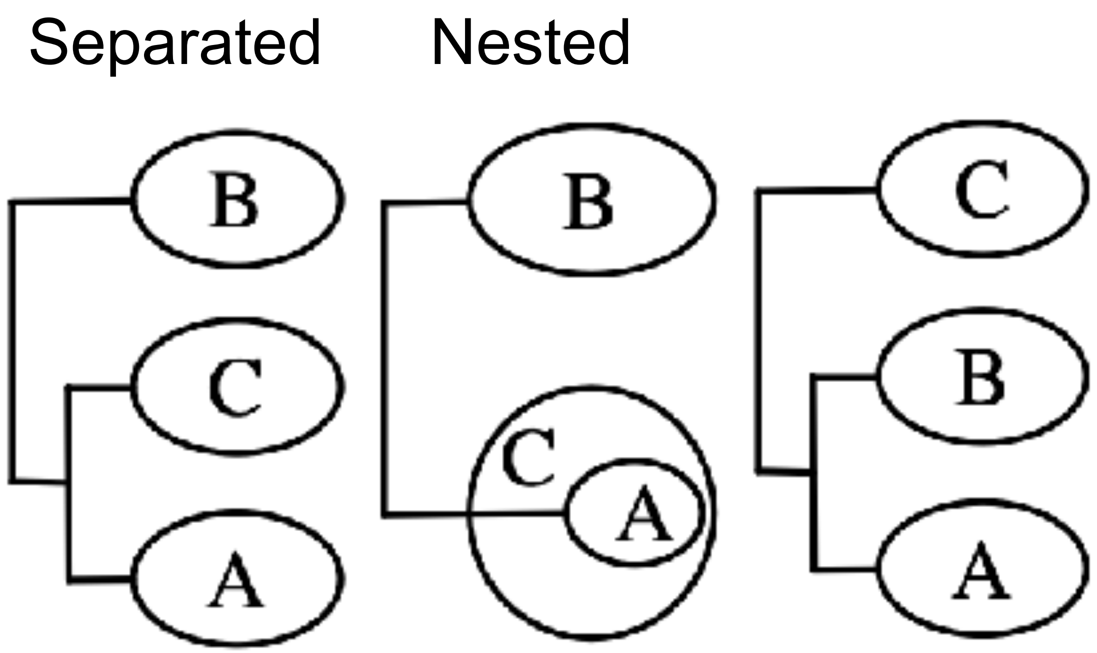


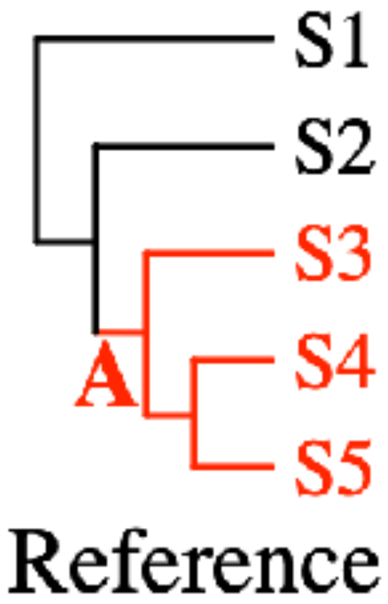
Two crucial tasks

Topological relationships between subtrees / leaf nodes

- Topological distance

Leaf node memberships compared to reference tree



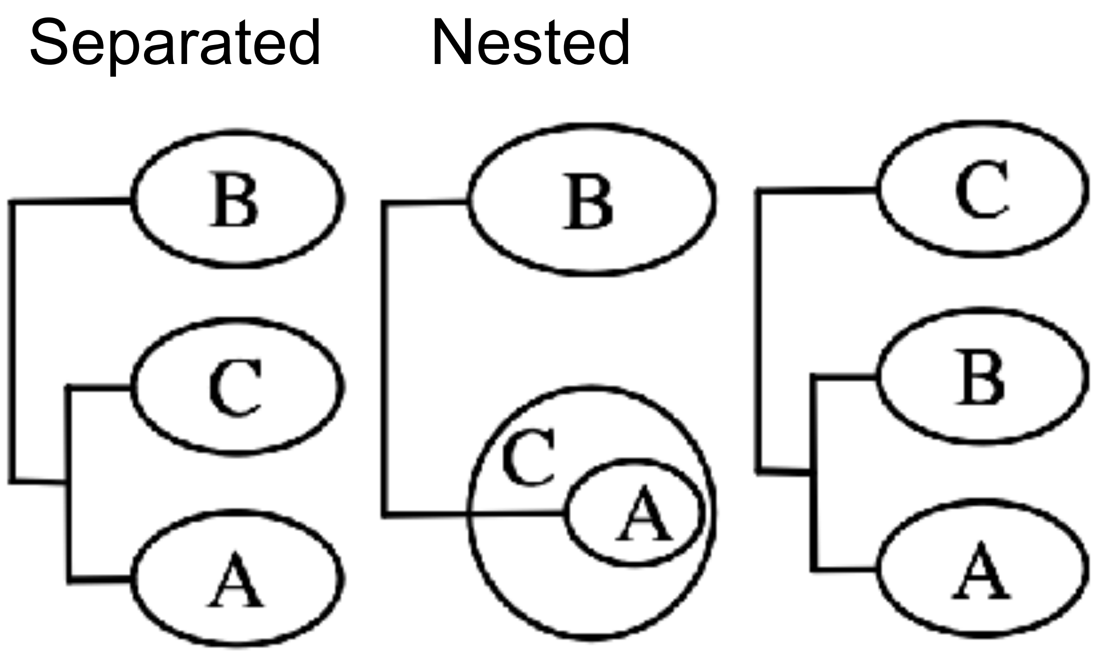


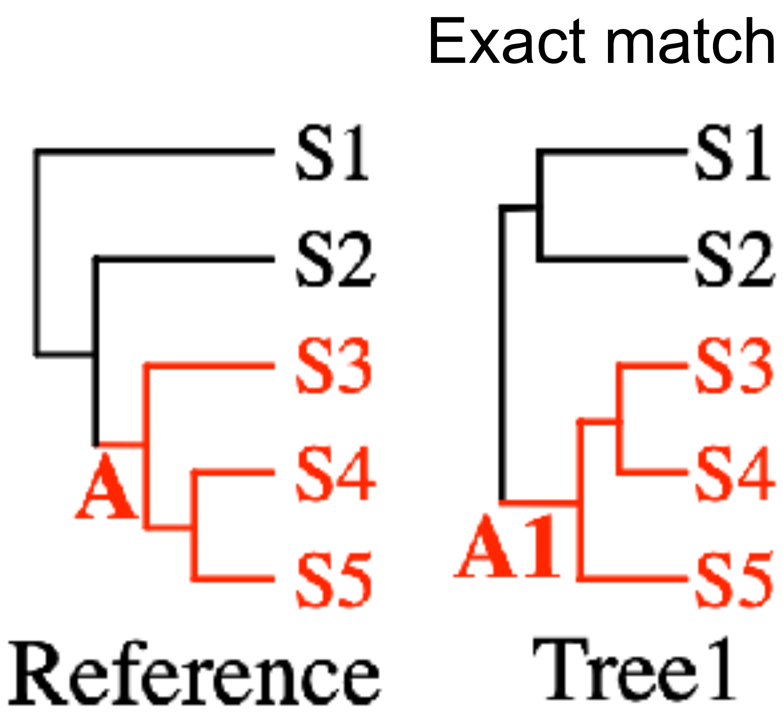
Two crucial tasks

Topological relationships between subtrees / leaf nodes

- Topological distance

Leaf node memberships compared to reference tree



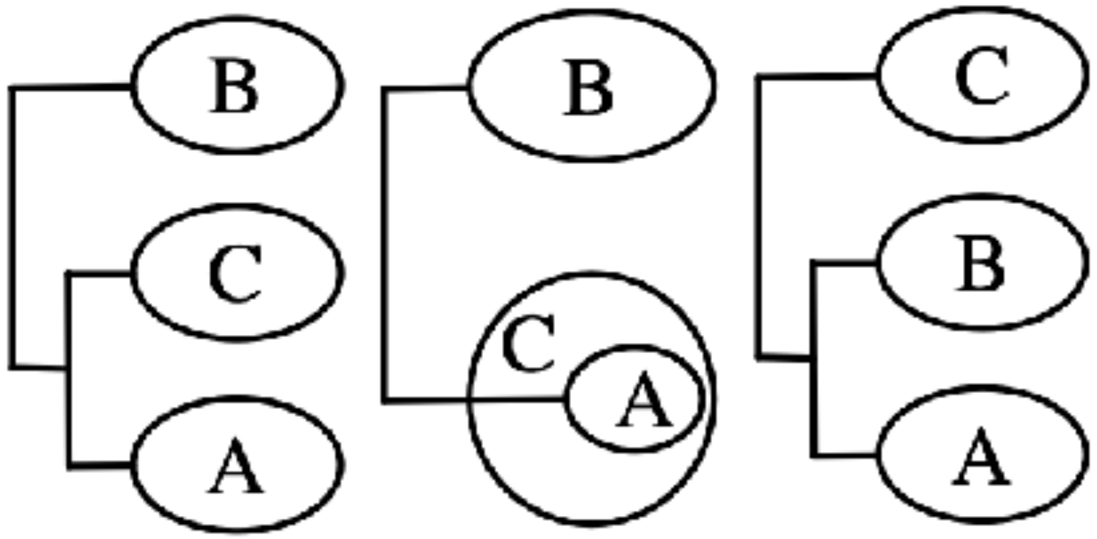


Two crucial tasks

Topological relationships between subtrees / leaf nodes

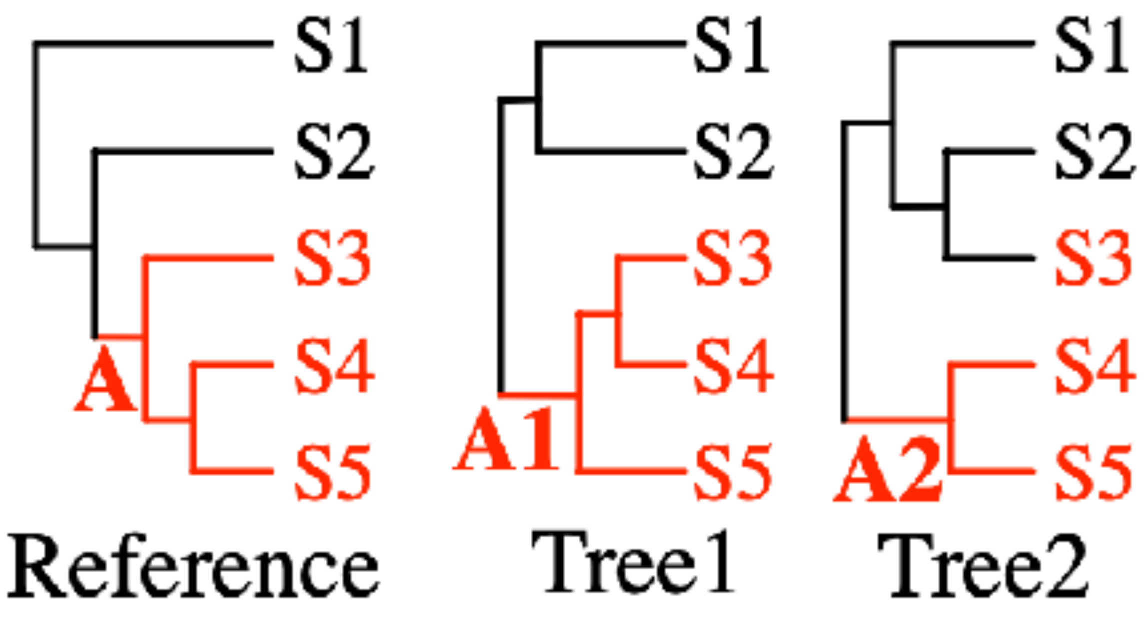
- Topological distance

Separated Nested



Leaf node memberships compared to reference tree

Exact match Partial match

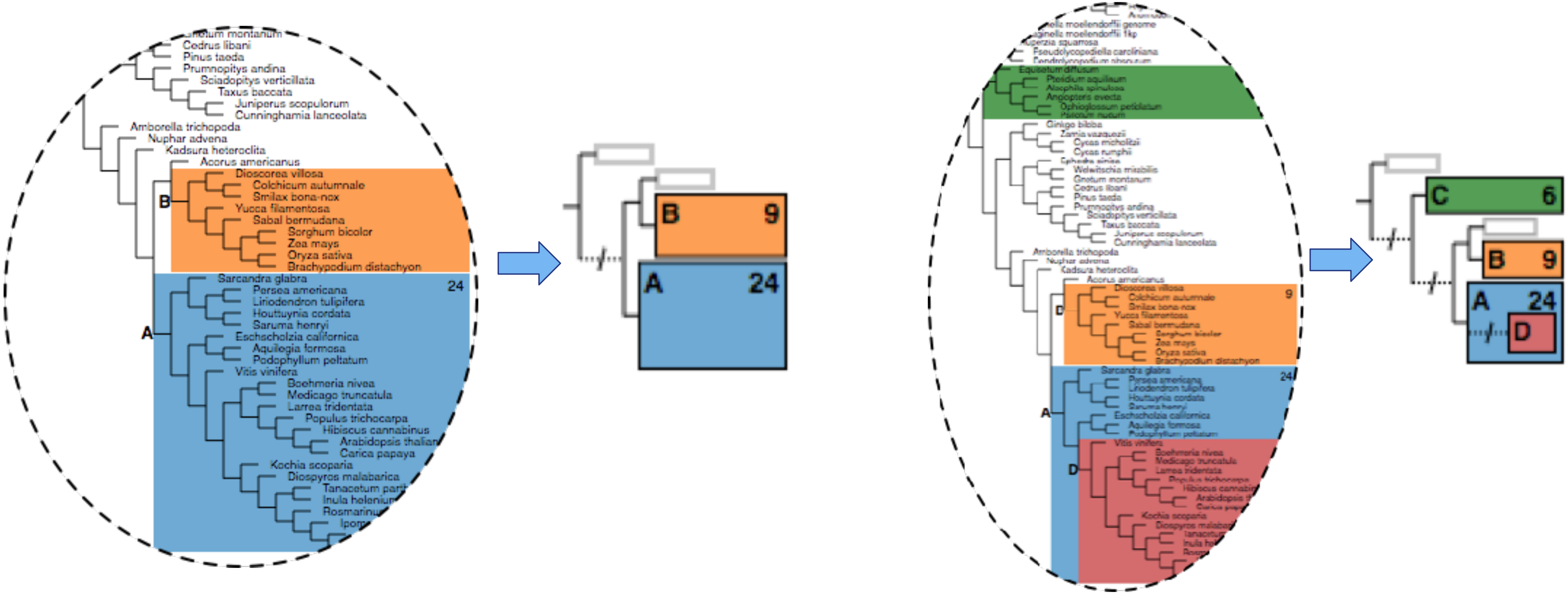


Aggregated Dendrogram (AD)

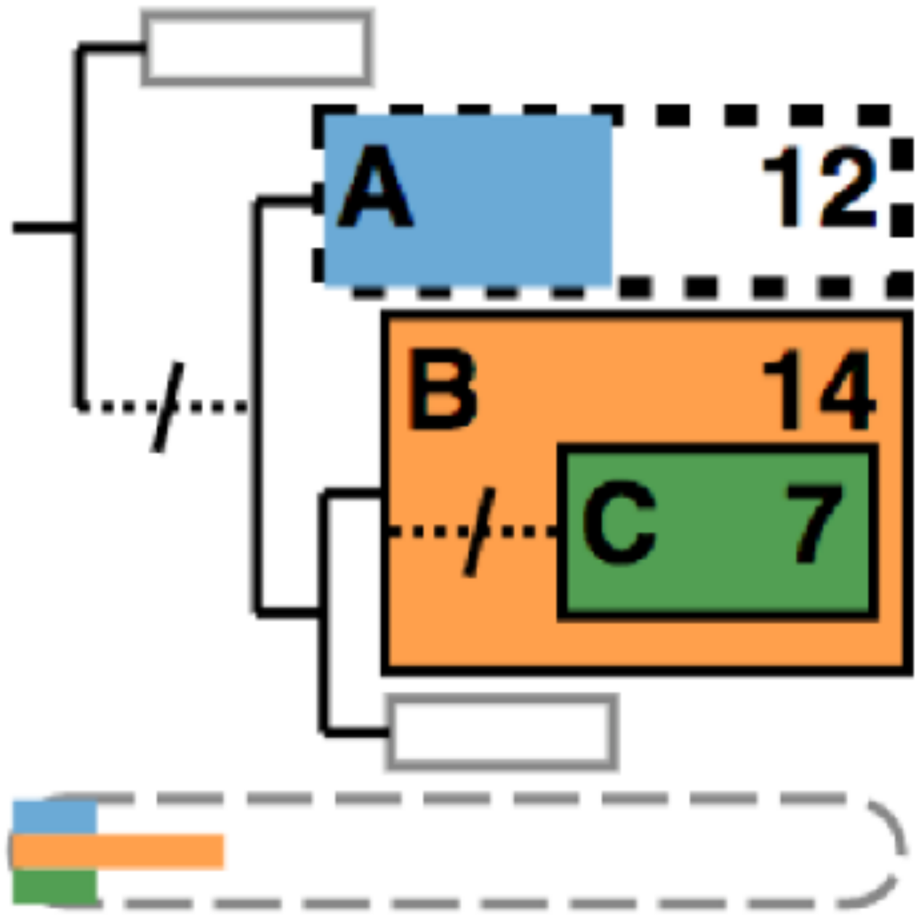
- Intuition
- Visual design

Intuition

Use glyphs to compress a tree according to user selections

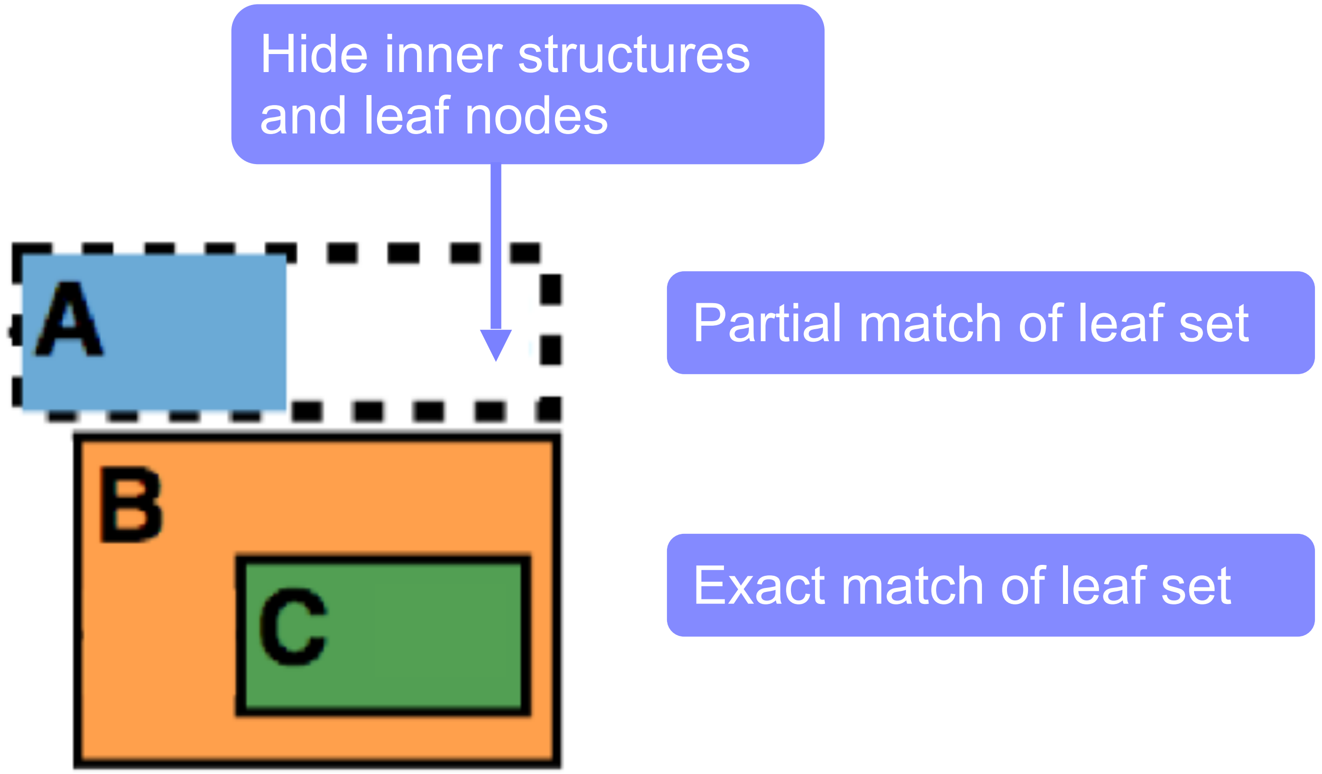


Visual design: focus + context



Visual design: focus + context

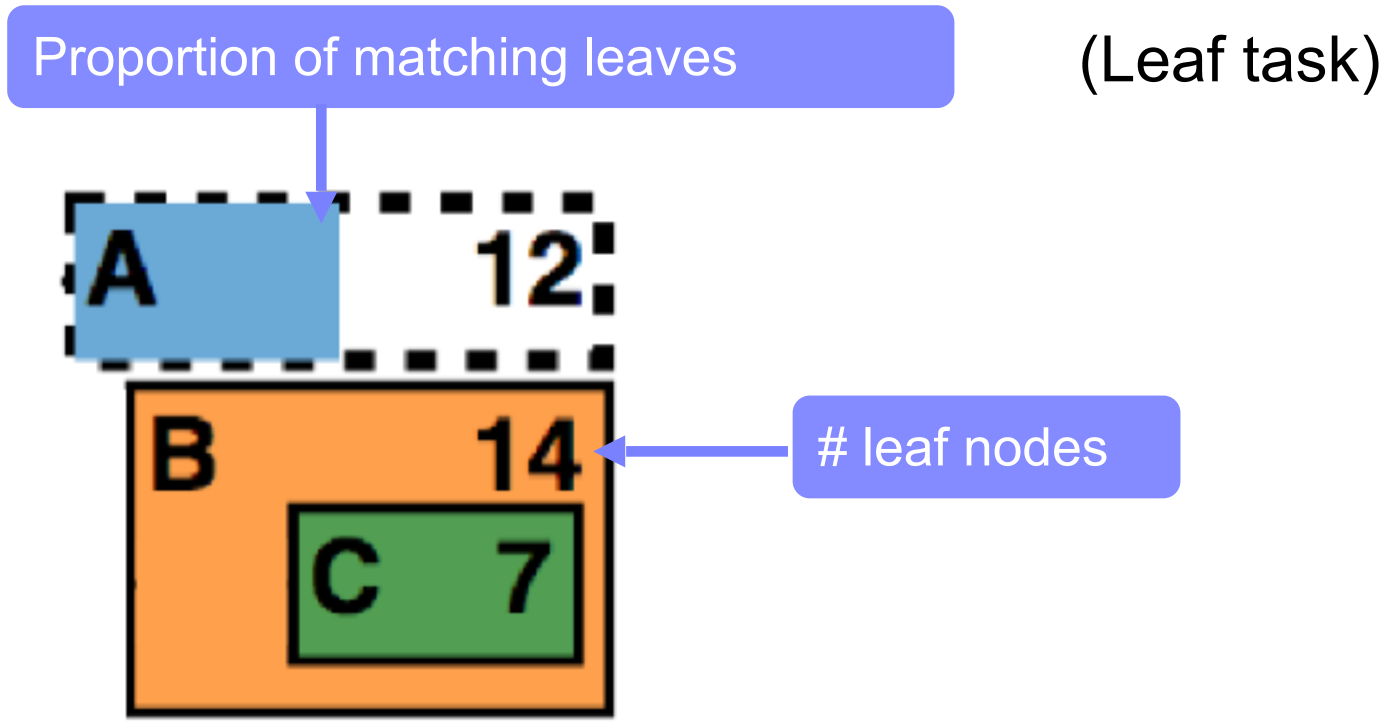
- Focus
  - Selected subtrees



(Leaf task)

Visual design: focus + context

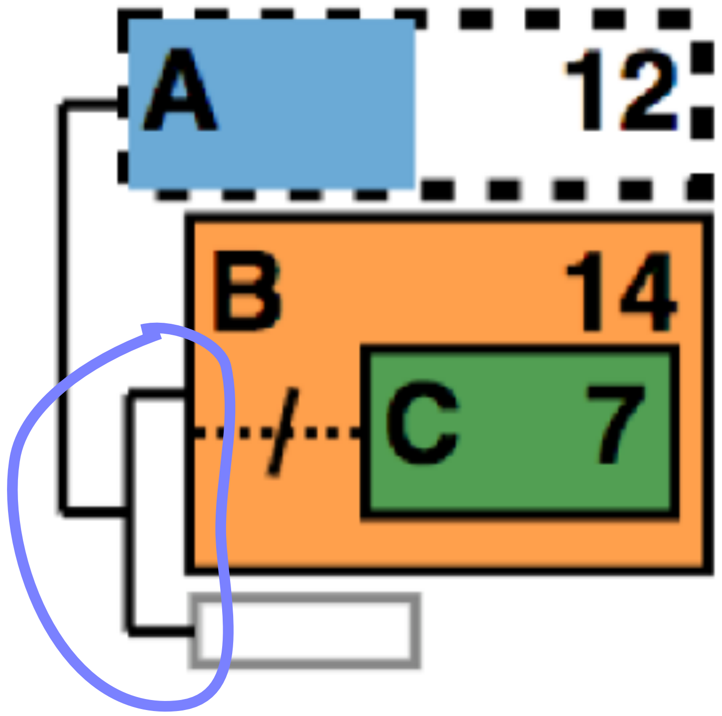
- Focus
  - Selected subtrees



Visual design: focus + context

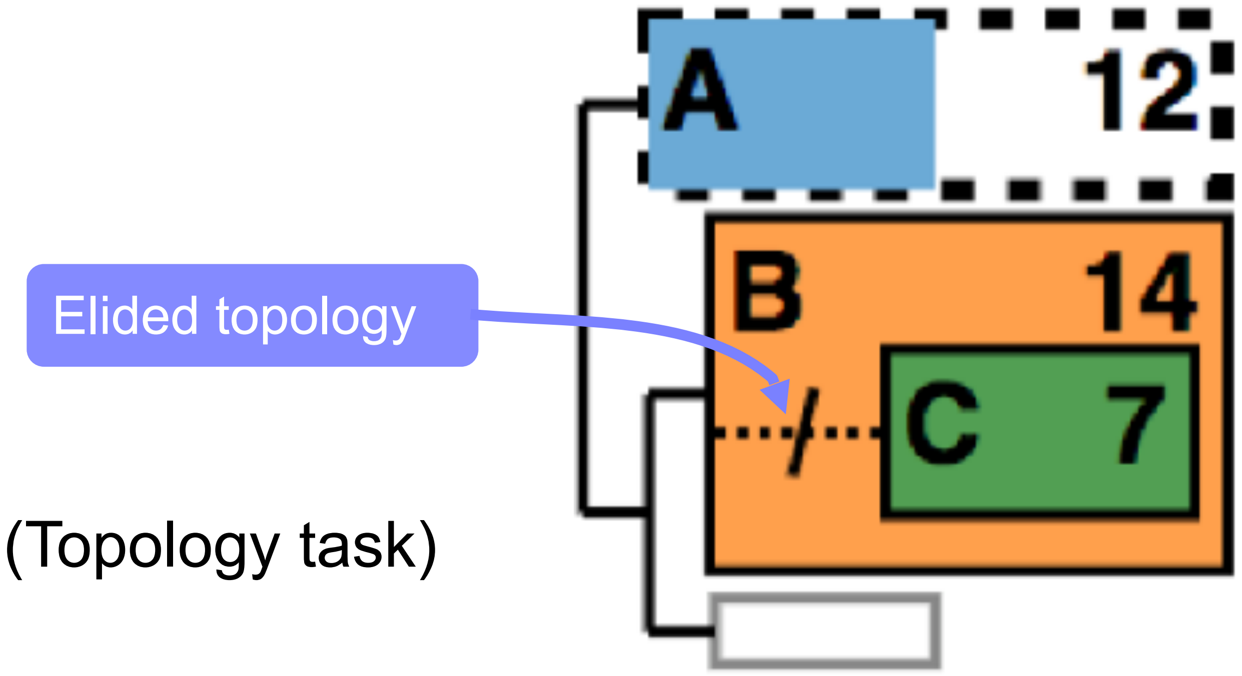
- Focus
  - Selected subtrees
  - Topological relationships between them

(Topology task)



Visual design: focus + context

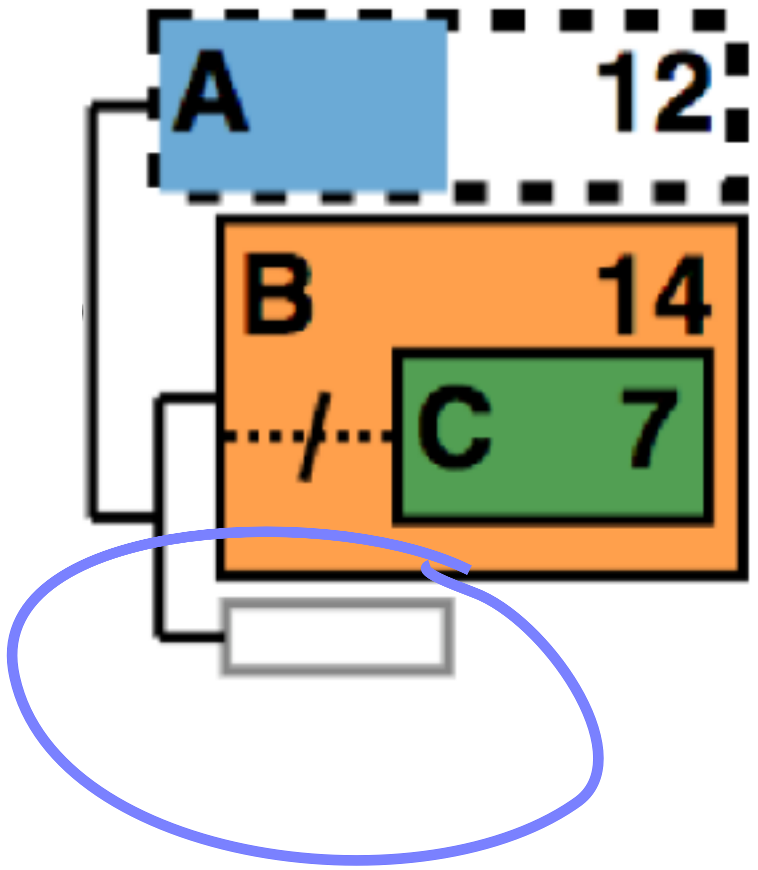
- Focus
  - Selected subtrees
  - Topological relationships between them



(Topology task)

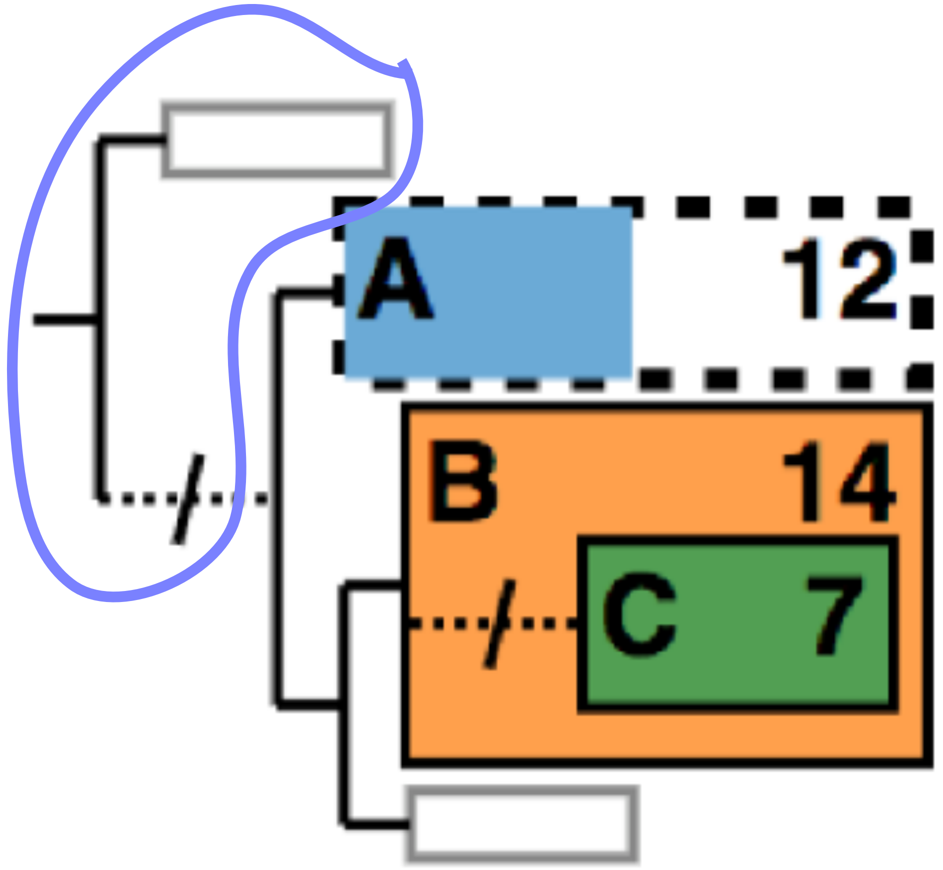
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees



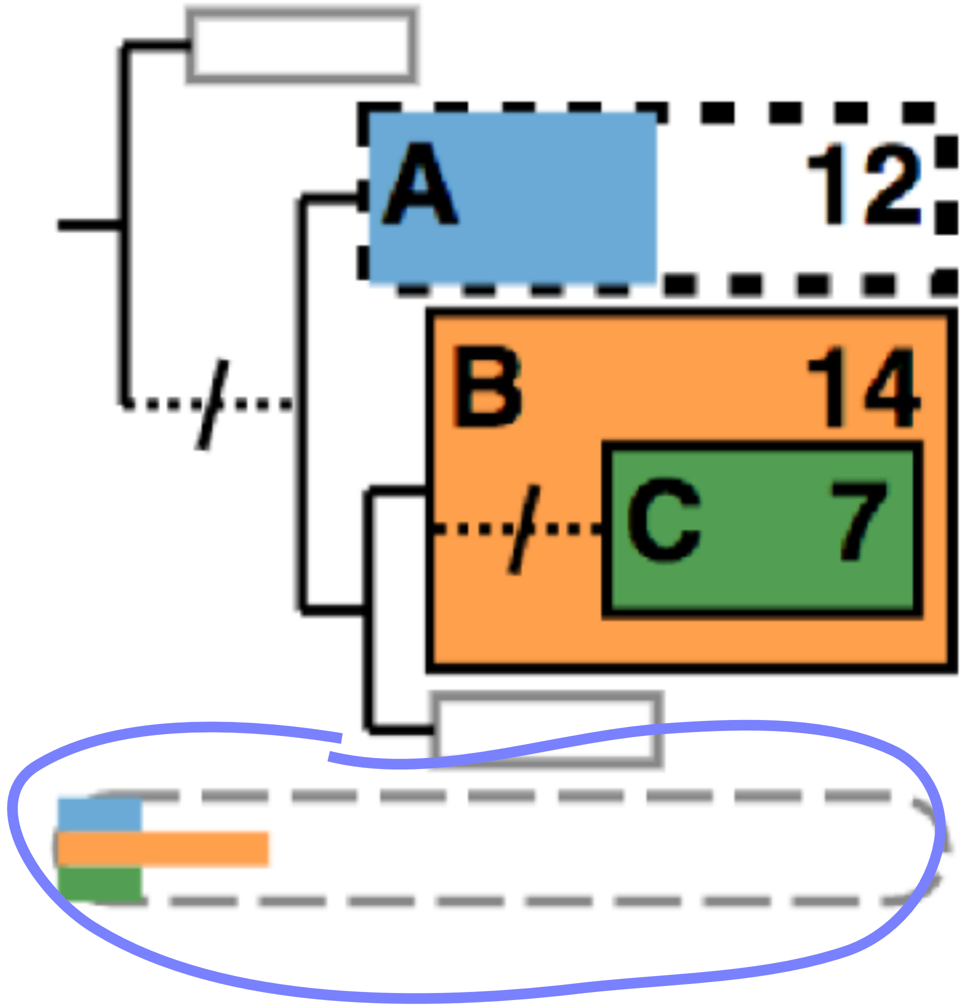
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees
  - Upstream topology and root



Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees
  - Upstream topology and root
  - Missing leaf nodes

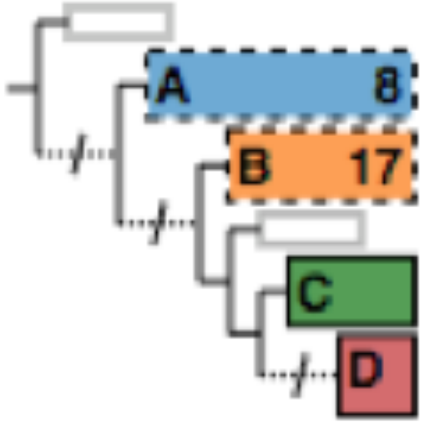


Visual design: algorithm adapts to space

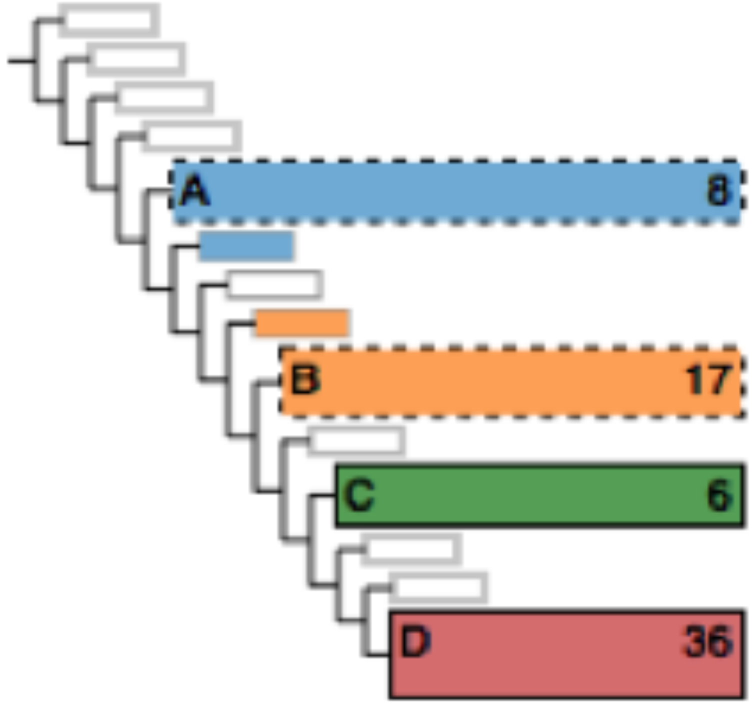
- Show more info when space permitted
  - Labels
  - #leaf nodes
  - Neighboring blocks

40x40 px

80x80 px

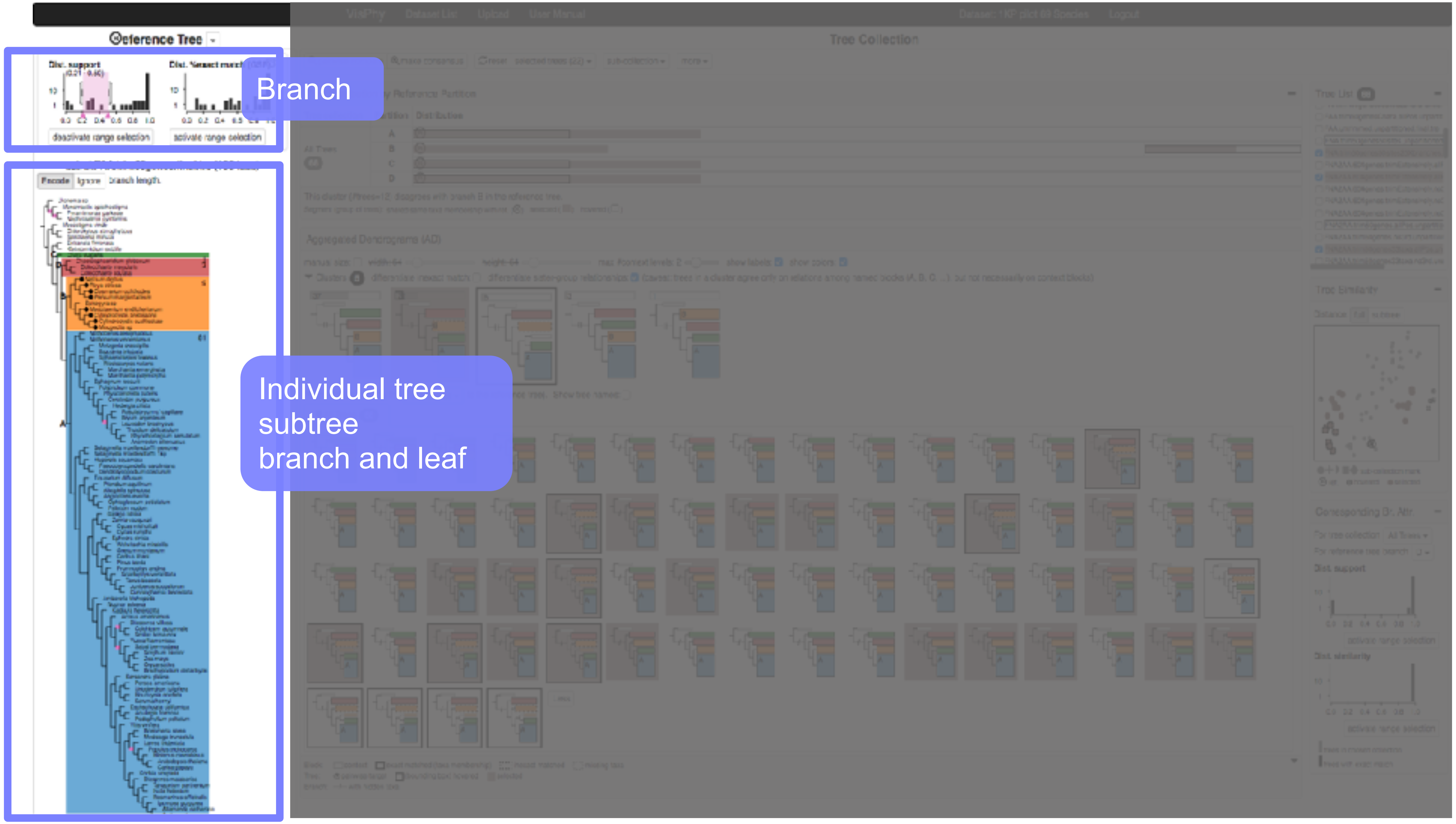


160x160 px

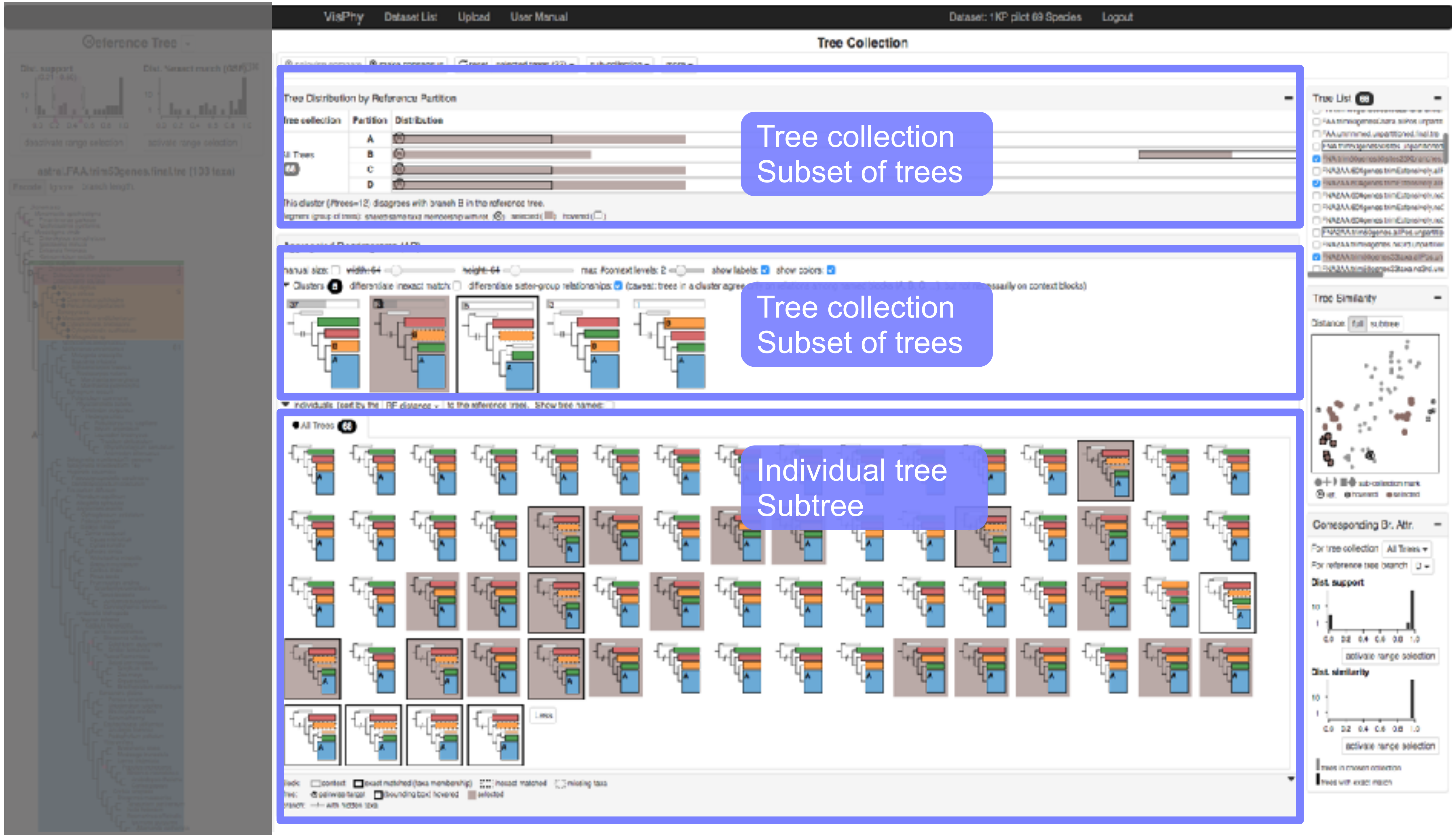


ADView Interface: Multi-level structure across views

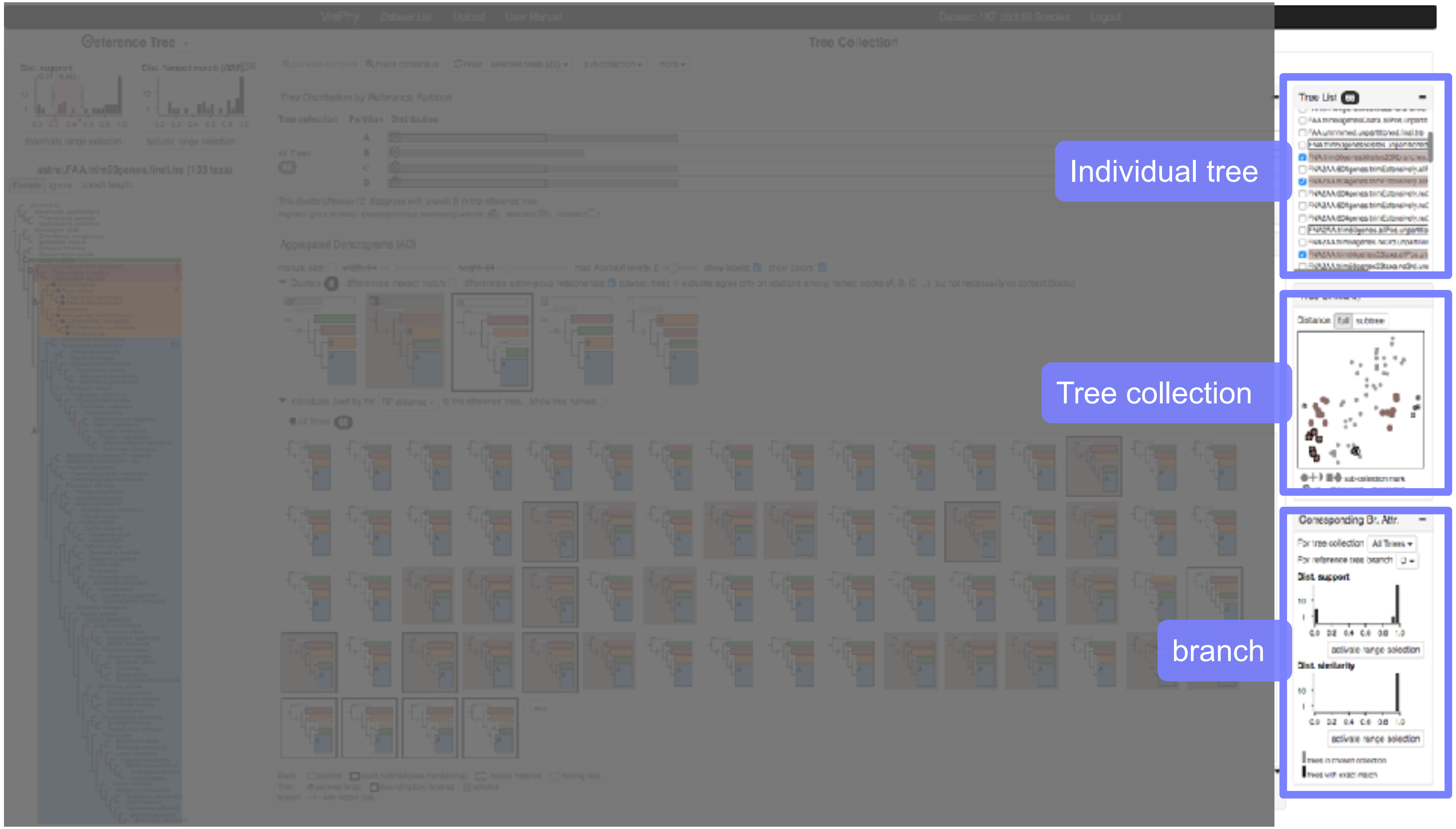
Multi-level structure across views



Interface walkthrough: tree collection main views



Interface walkthrough: tree collection aux. views

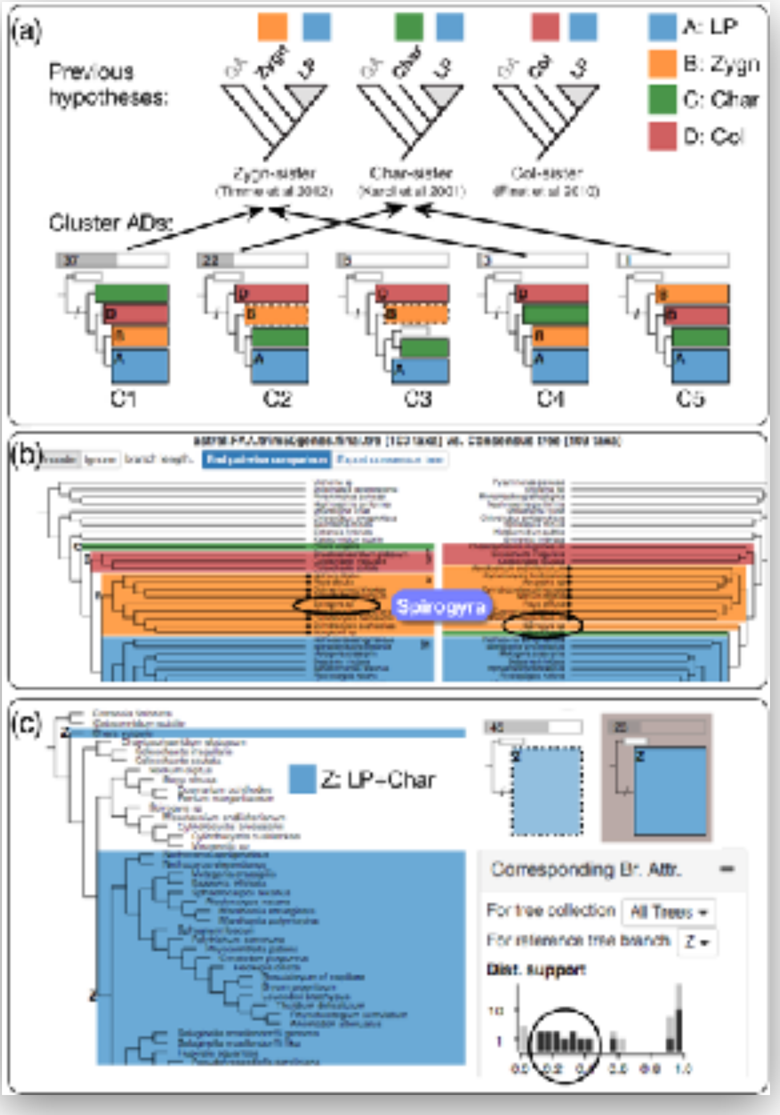


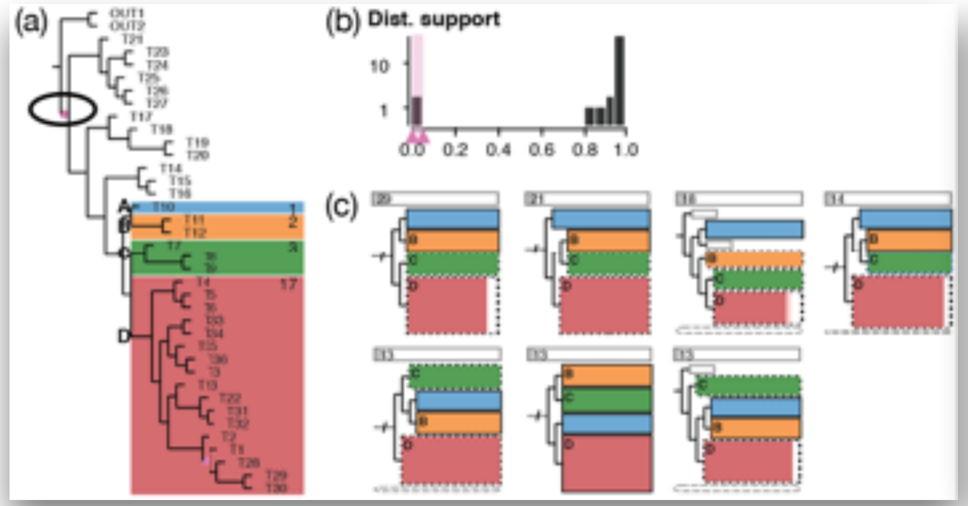
Validation with many biologists

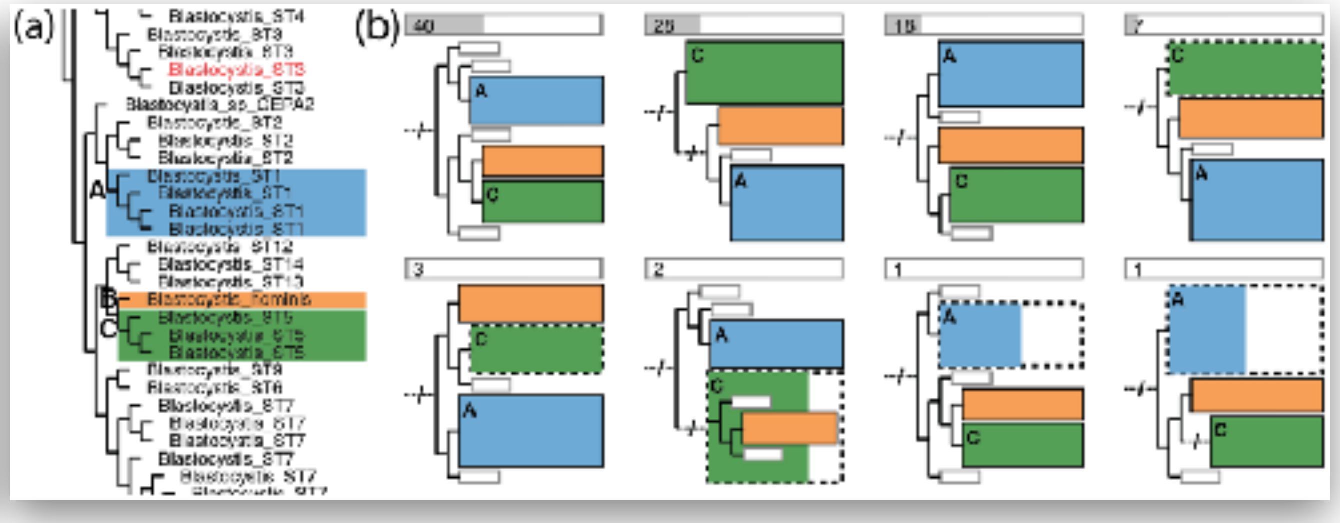
- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
    - 10 biologists at different times throughout project

Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
  - 10 biologists at different times throughout project
- User study sessions
  - 5 biologists
  - Using their own datasets

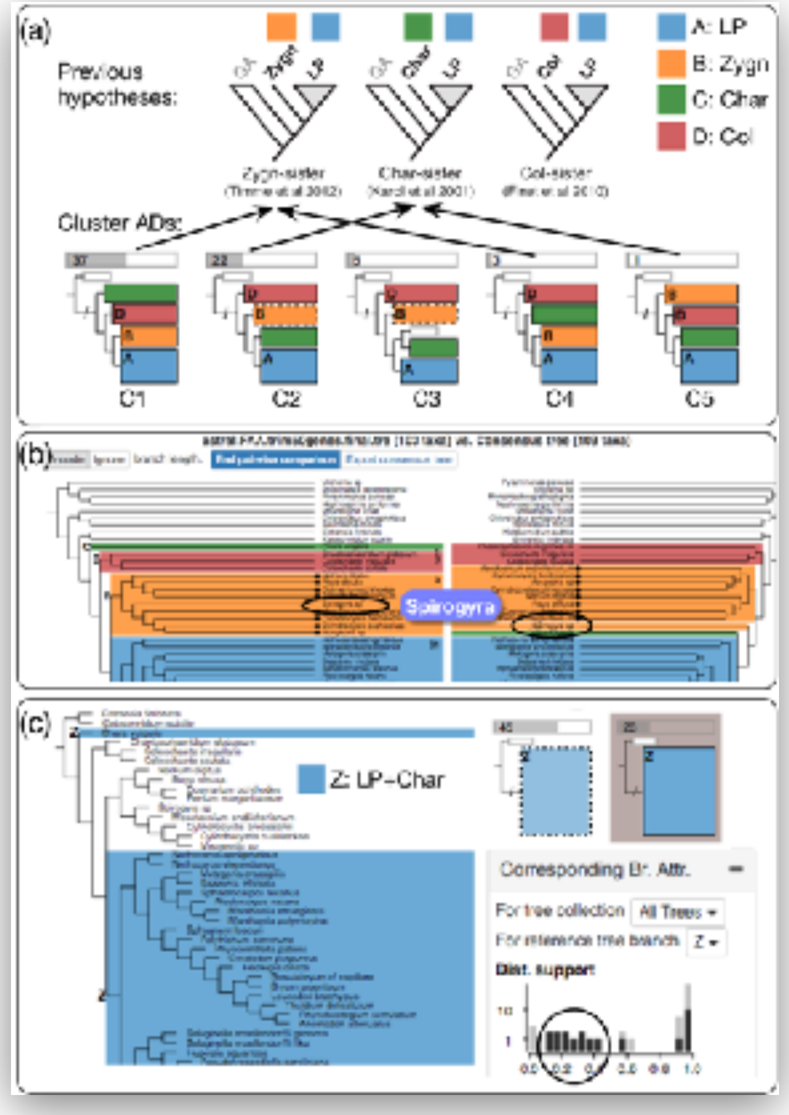


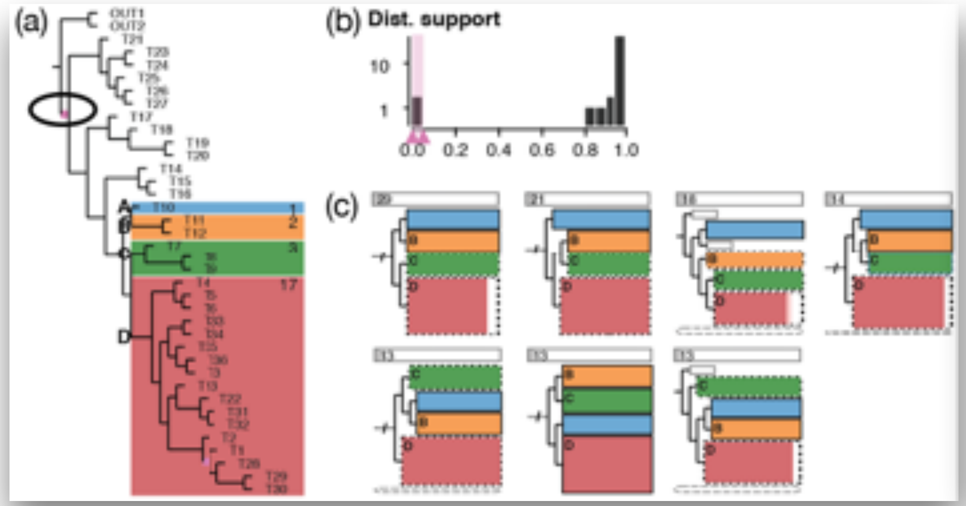


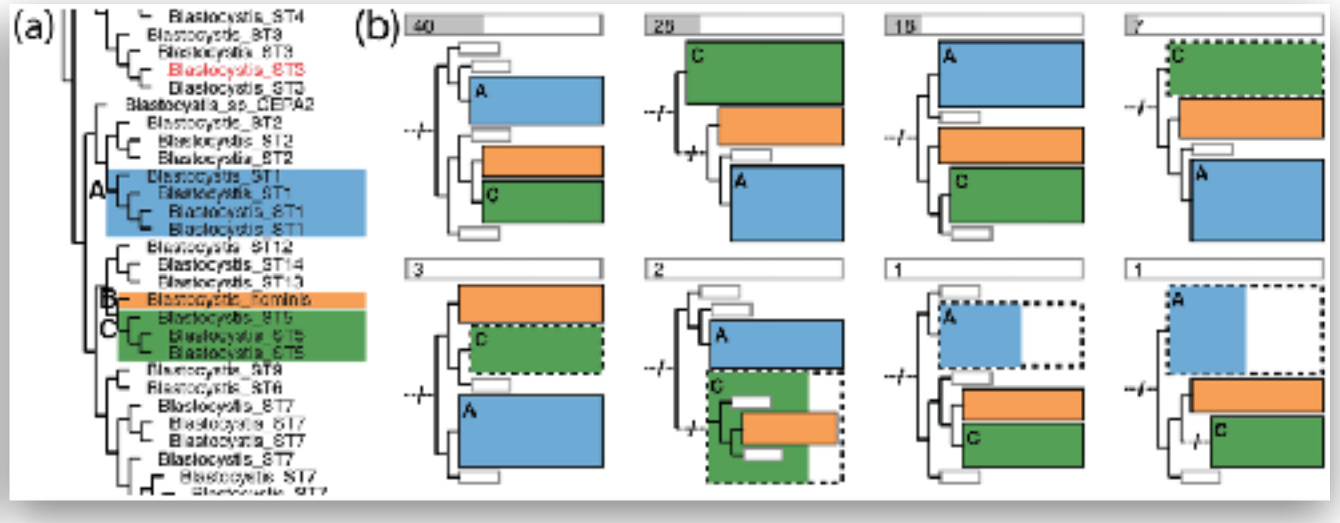


Validation with many biologists

- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
  - 10 biologists at different times throughout project
- User study sessions
  - 5 biologists
  - Using their own datasets
- Biologists confirmed
  - Validity of data and task abstractions
  - Utility of ADView

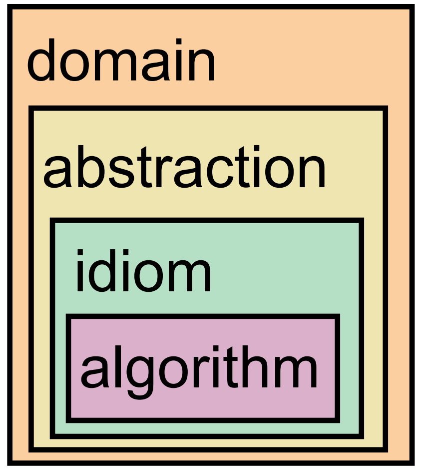






Problem-driven visualization through design studies

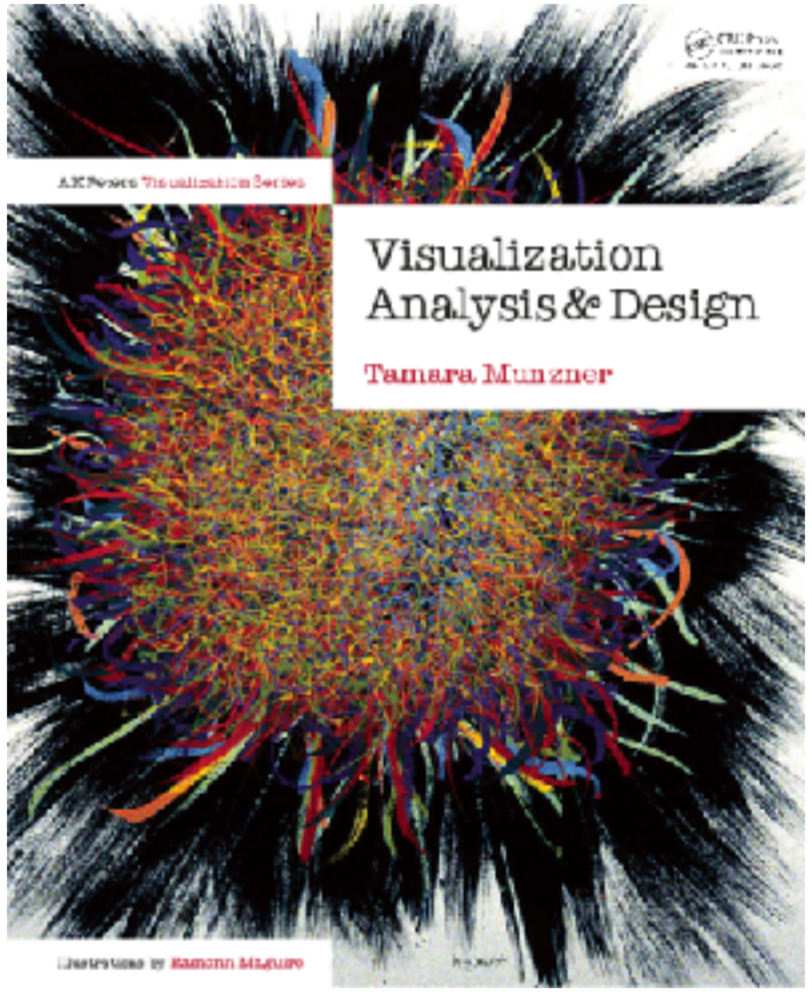
- methodology matters
  - identify abstractions
    - crucial & difficult, iterative process
  - select appropriate idioms
    - or create new ones if necessary



- three examples
  - different domains
  - different methods

More information

- theoretical foundations: book
  (+ tutorial/course lecture slides)
  http://www.cs.ubc.ca/~tmm/vadbook



Visualization Analysis and Design.
Munzner.
AK Peters Visualization Series.
CRC Press, 2014.

@tamaramunzner

UBC

UBC Data Science Institute

DESIGNING for PEOPLE

CAIDA

- papers, videos, software, talks, courses
  http://www.cs.ubc.ca/group/infovis
  http://www.cs.ubc.ca/~tmm

- this talk

http://www.cs.ubc.ca/~tmm/talks.html#vinci21