Visualization: Abstractions & Idioms

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

University of Coimbra Guest Lecture
9 Mar 2022, virtual / Coimbra, Portugal

http://www.cs.ubc.ca/~tmm/talks.html#coimbra22

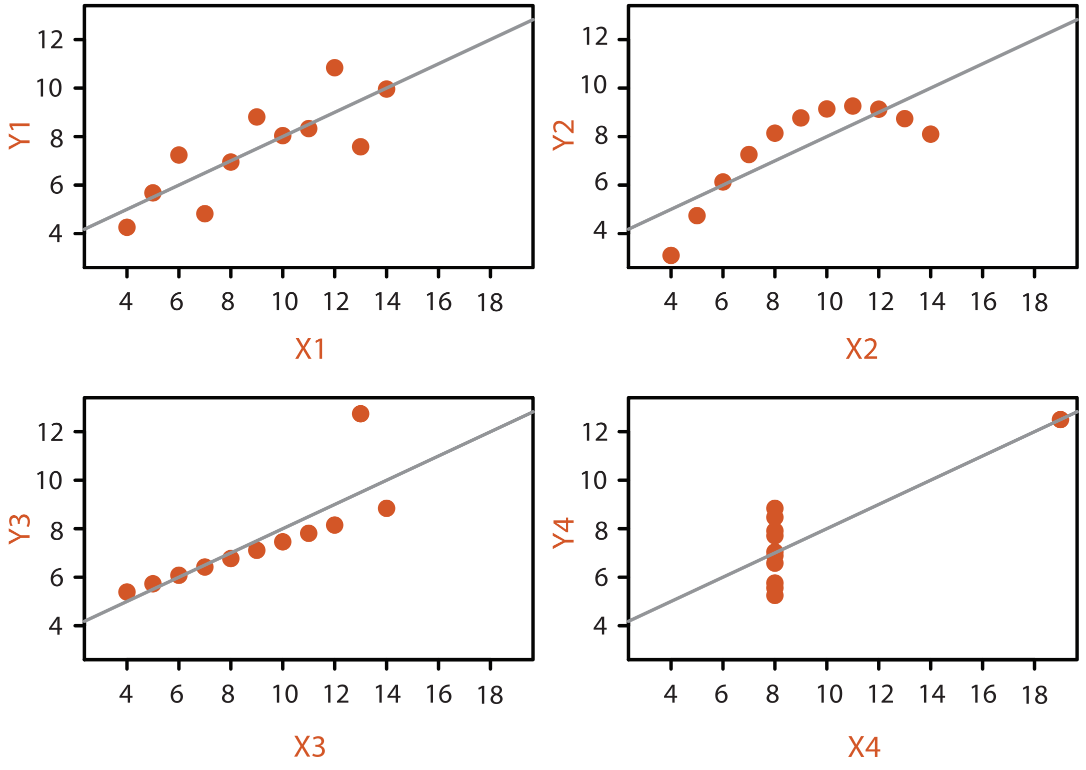
Visualization defined & motivated

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

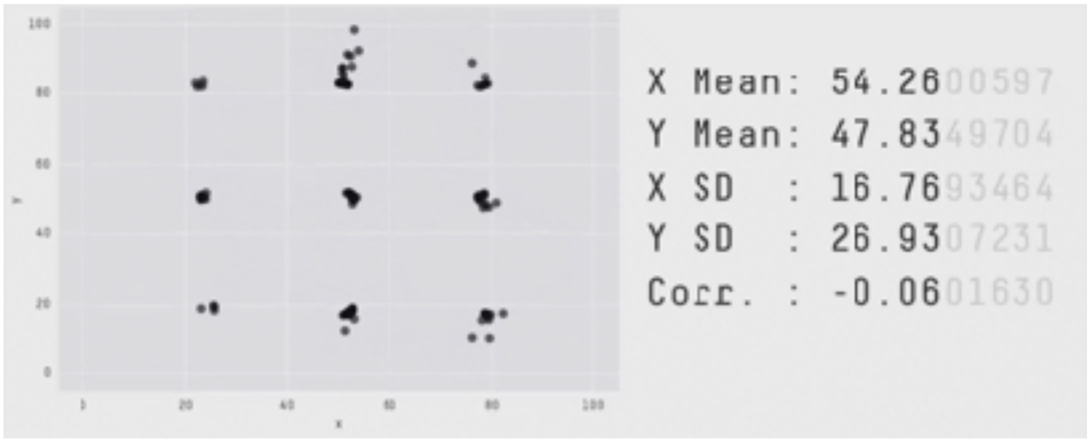
- suitable when human in the loop needs details
  - interplay between human judgement and automatic computation

Anscombe's Quartet

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



Datasaurus Dozen



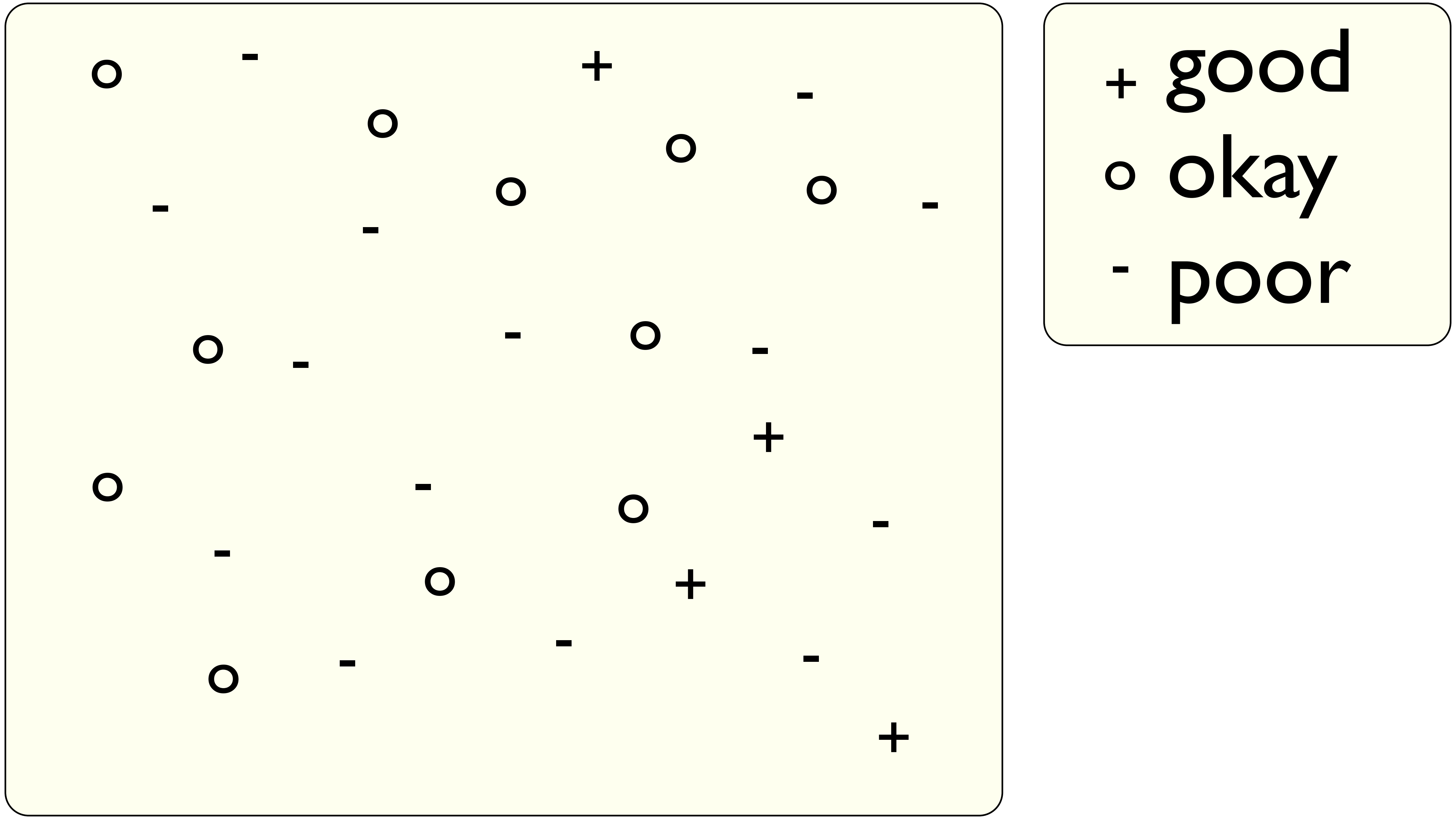
Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. CHI 2017. Matejka & Fitzmaurice

Why focus on tasks and effectiveness?

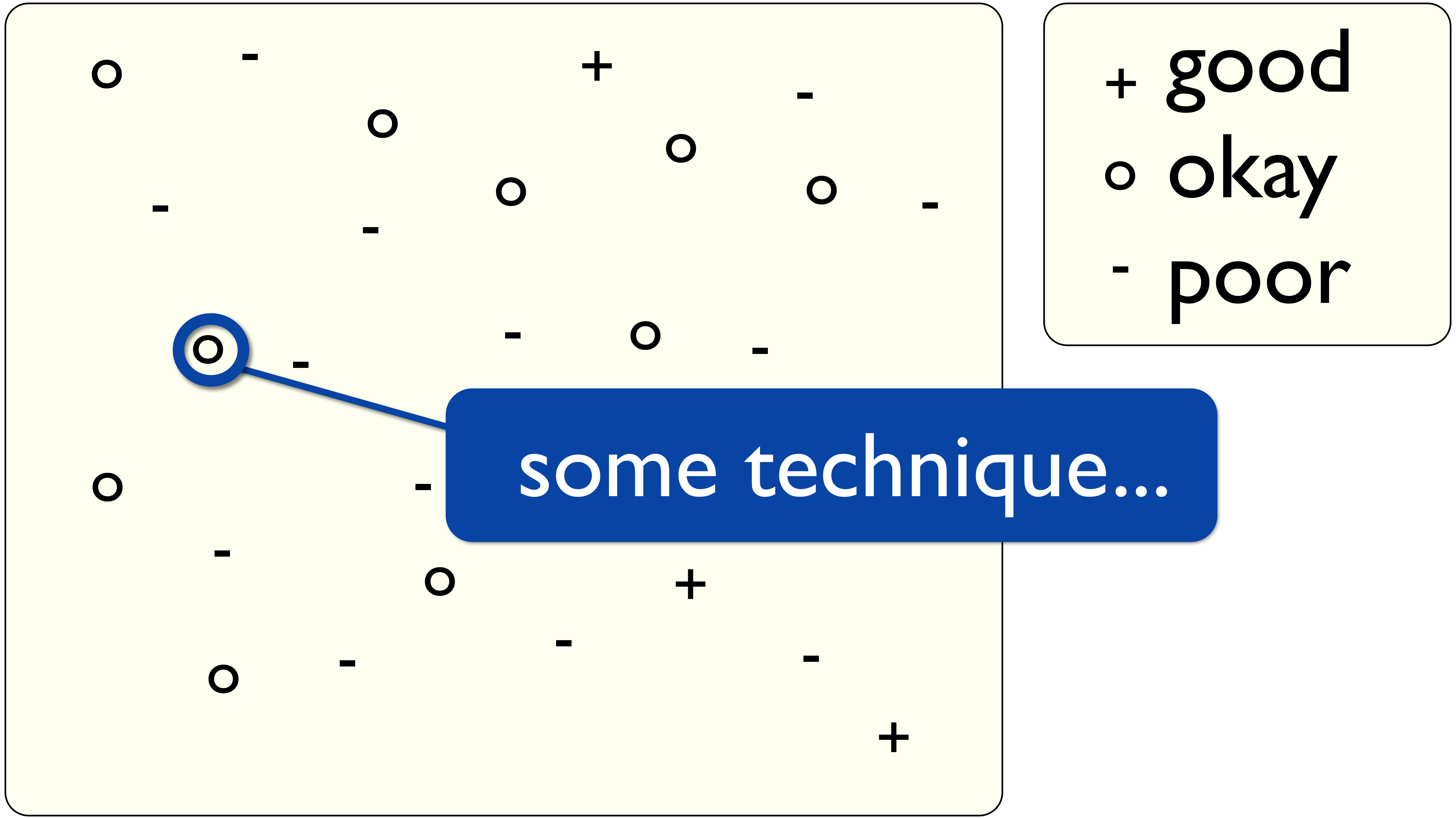
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- effectiveness requires match between data/task and representation
  - set of representations is huge
  - many are ineffective mismatch for specific data/task combo
  - increases chance of finding good solutions if you understand full space of possibilities

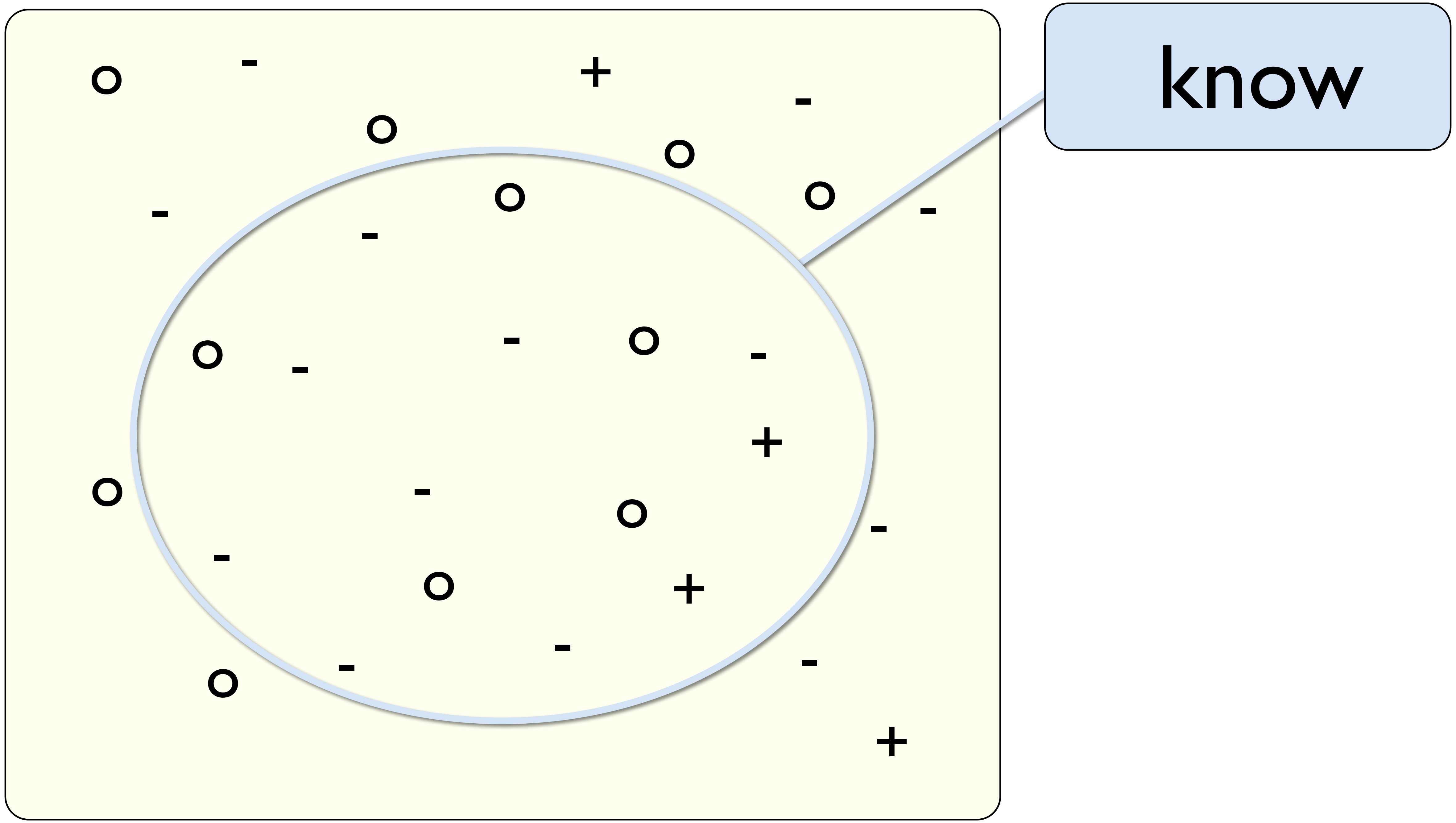
Metaphor: Design space



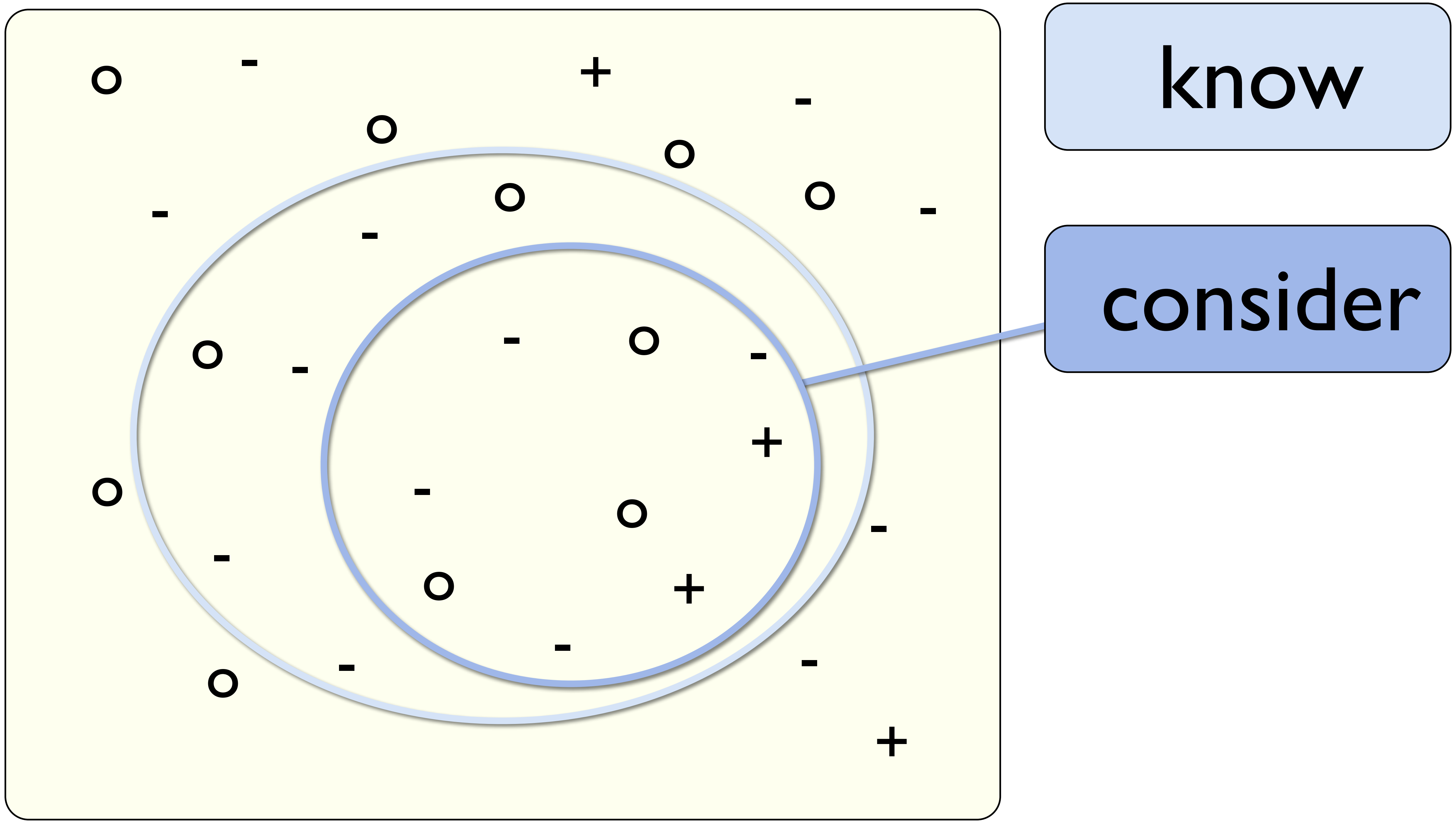
Metaphor: Design space



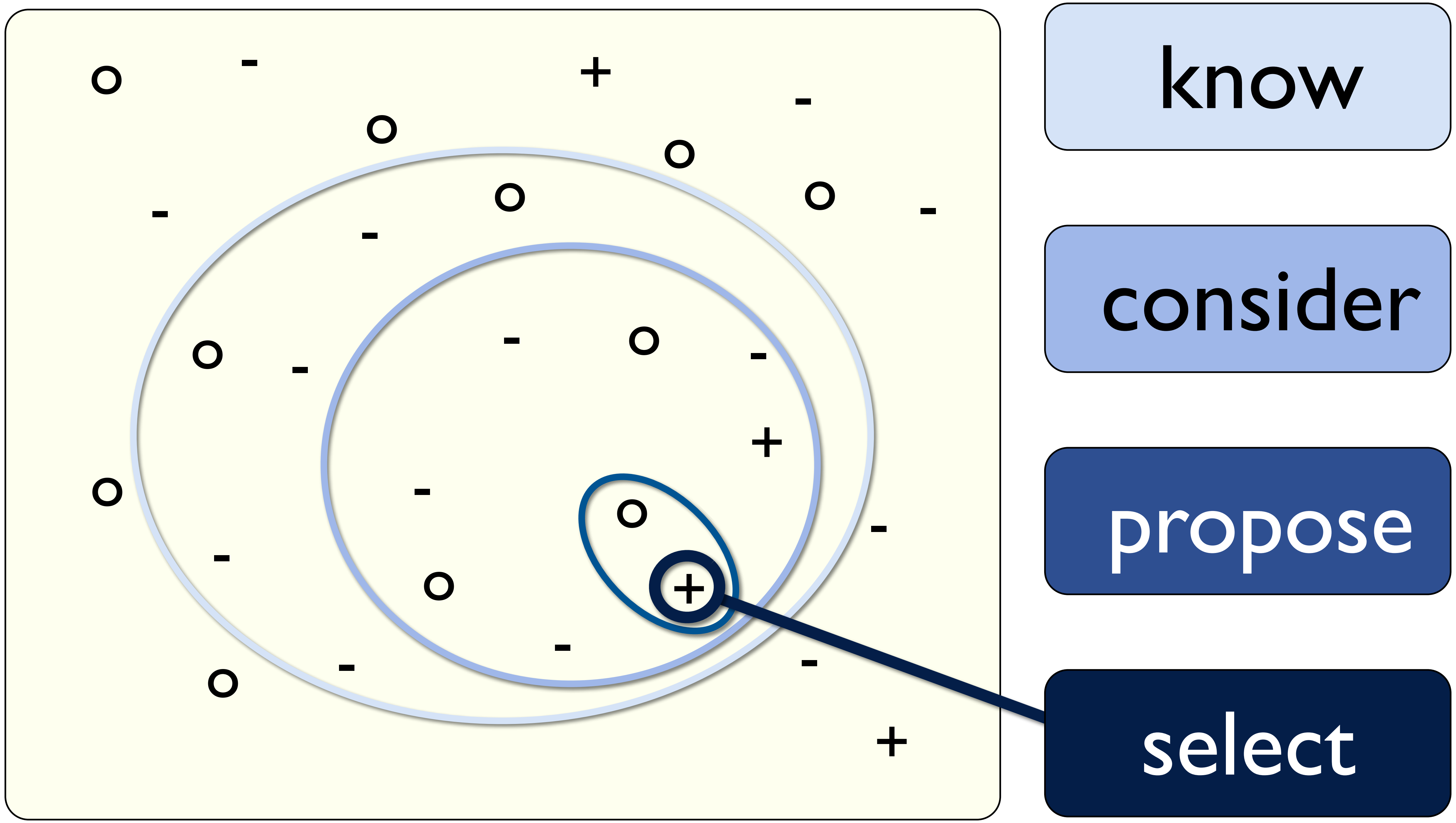
Metaphor: Design space



Metaphor: Design space



Metaphor: Design space



Metaphor: Design space

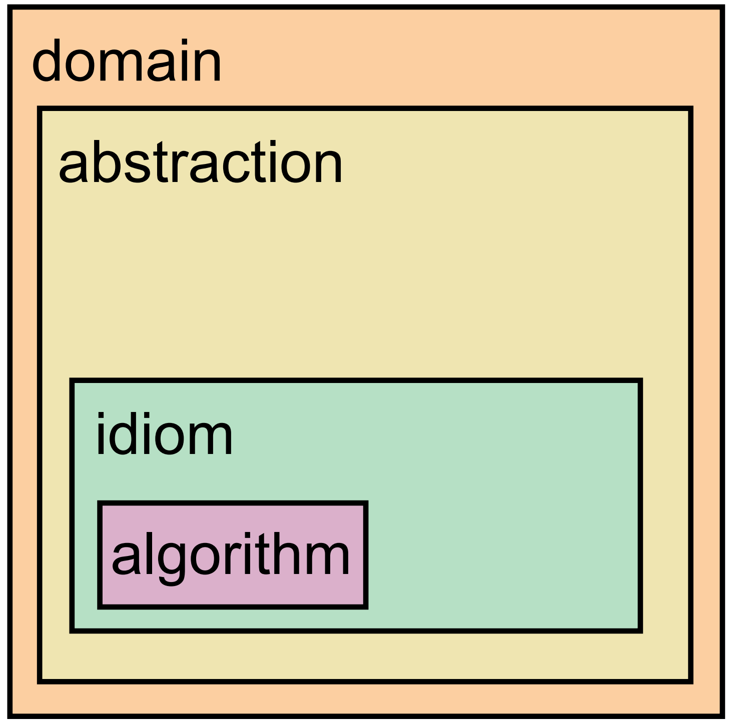
Think broad!

Why focus on tasks and effectiveness?

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

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

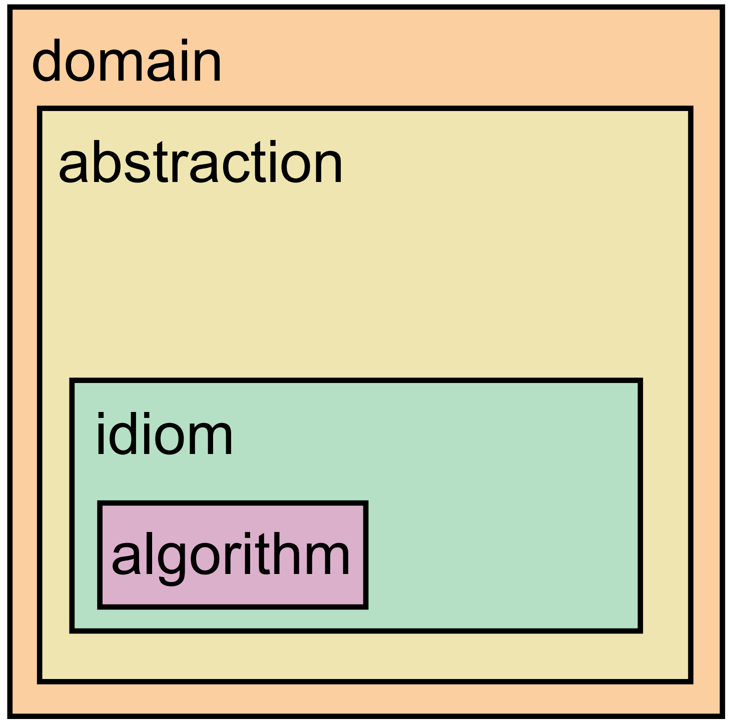
Nested model: Four levels of visualization concerns



[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

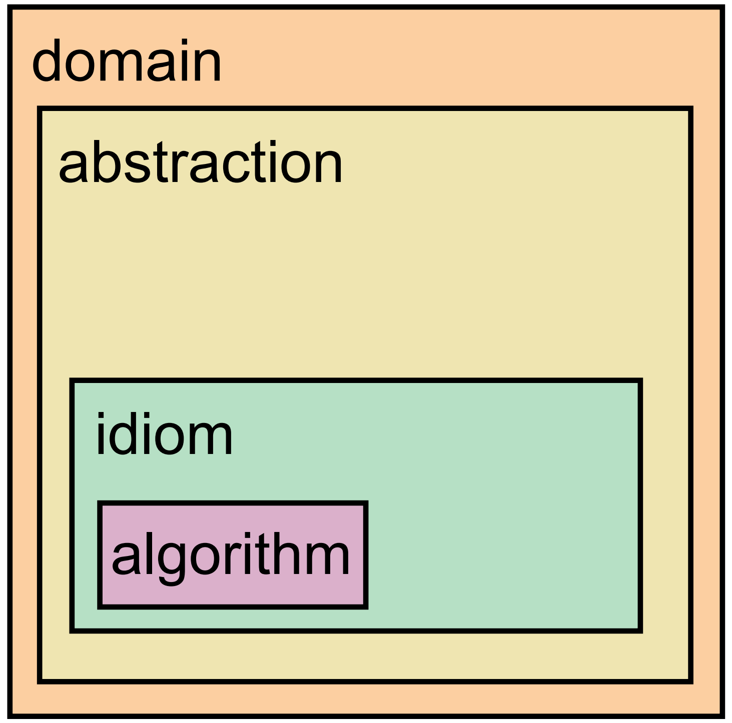
- *domain* situation
  - **who** are the target users?



[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

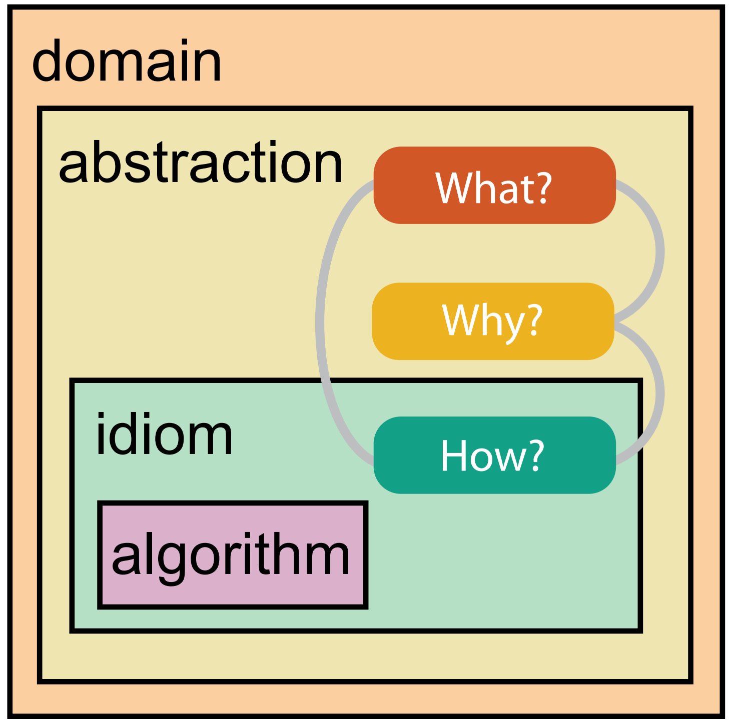
- *domain* situation
  - **who** are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of vis



[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**

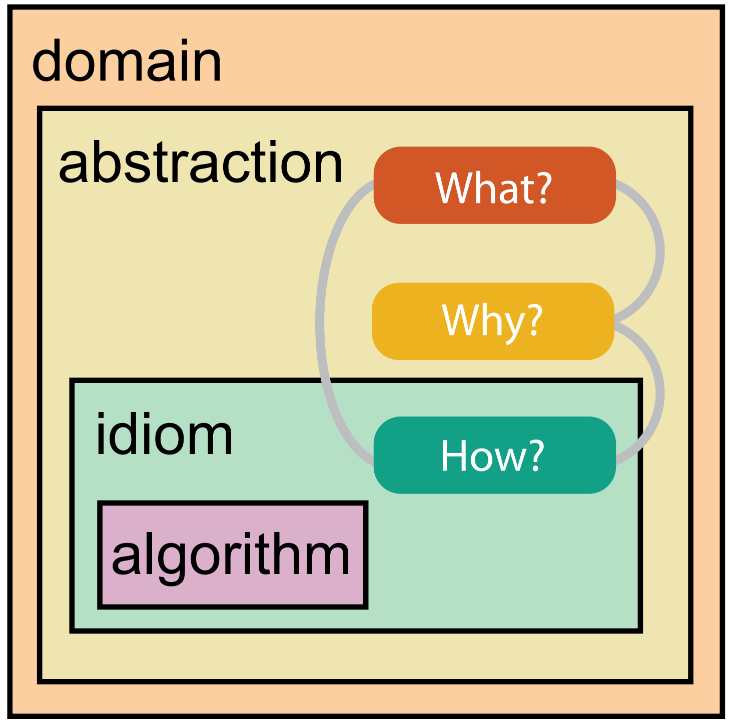


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[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

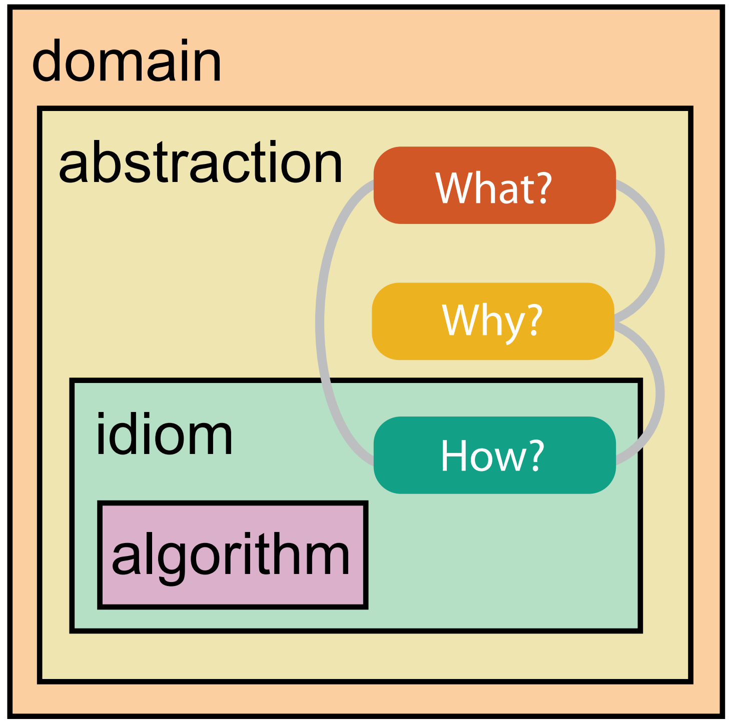


[*A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).*]

[*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**

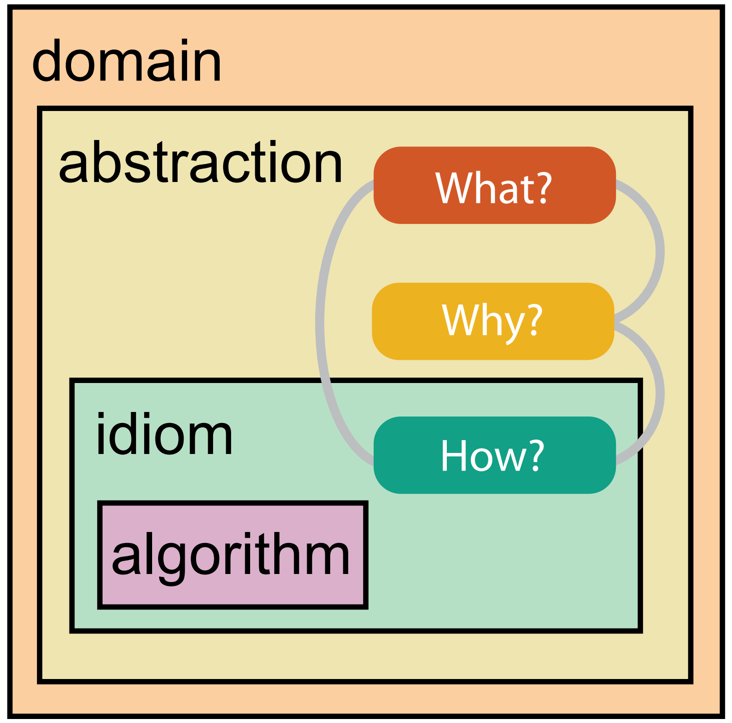


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[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?

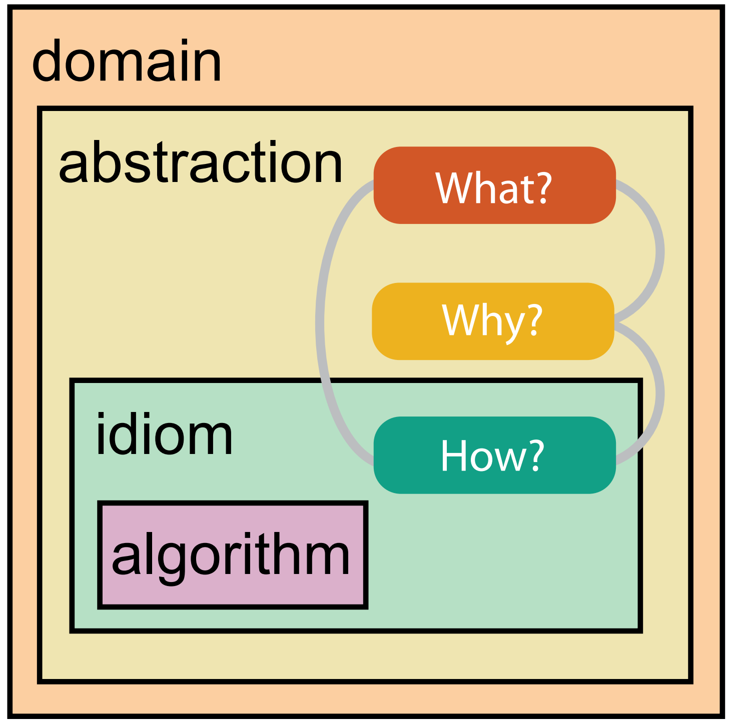


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[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

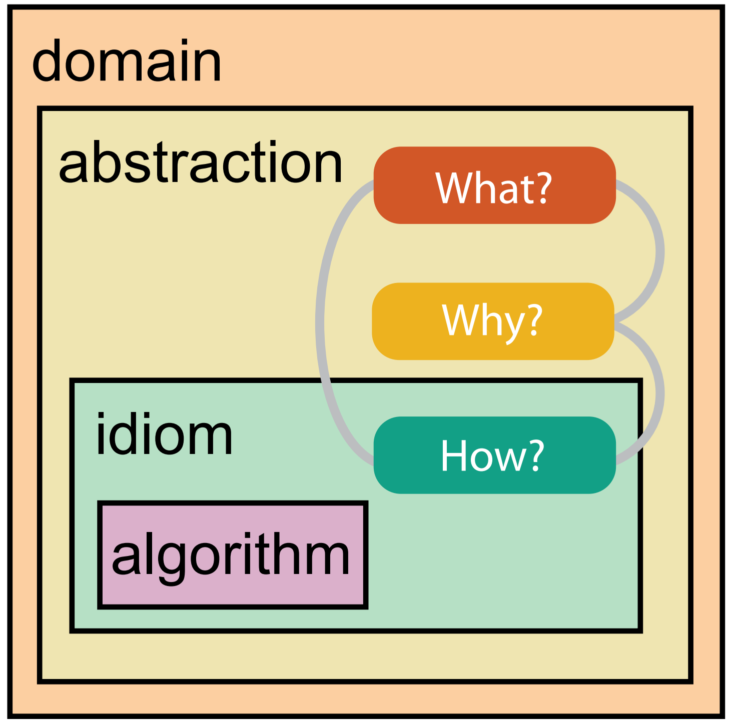


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[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

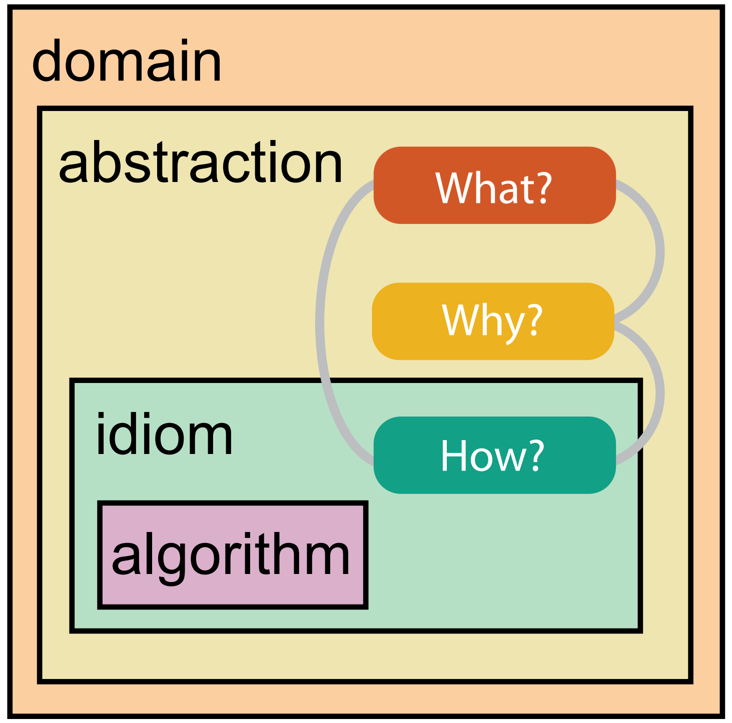


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

[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?

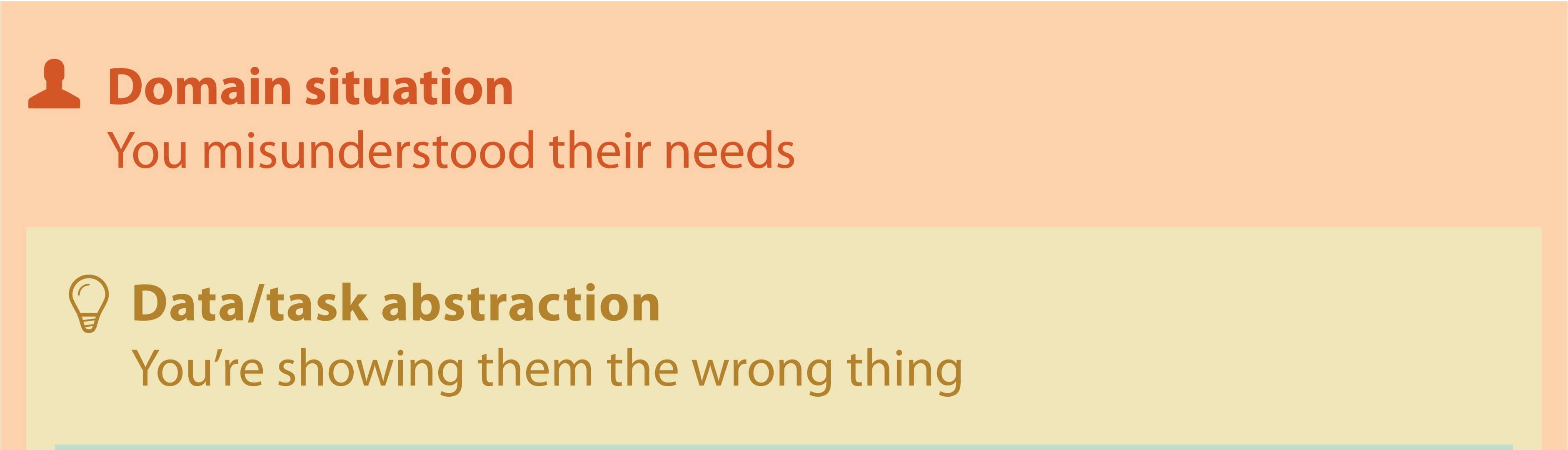
- different ways to get it wrong at each level

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?

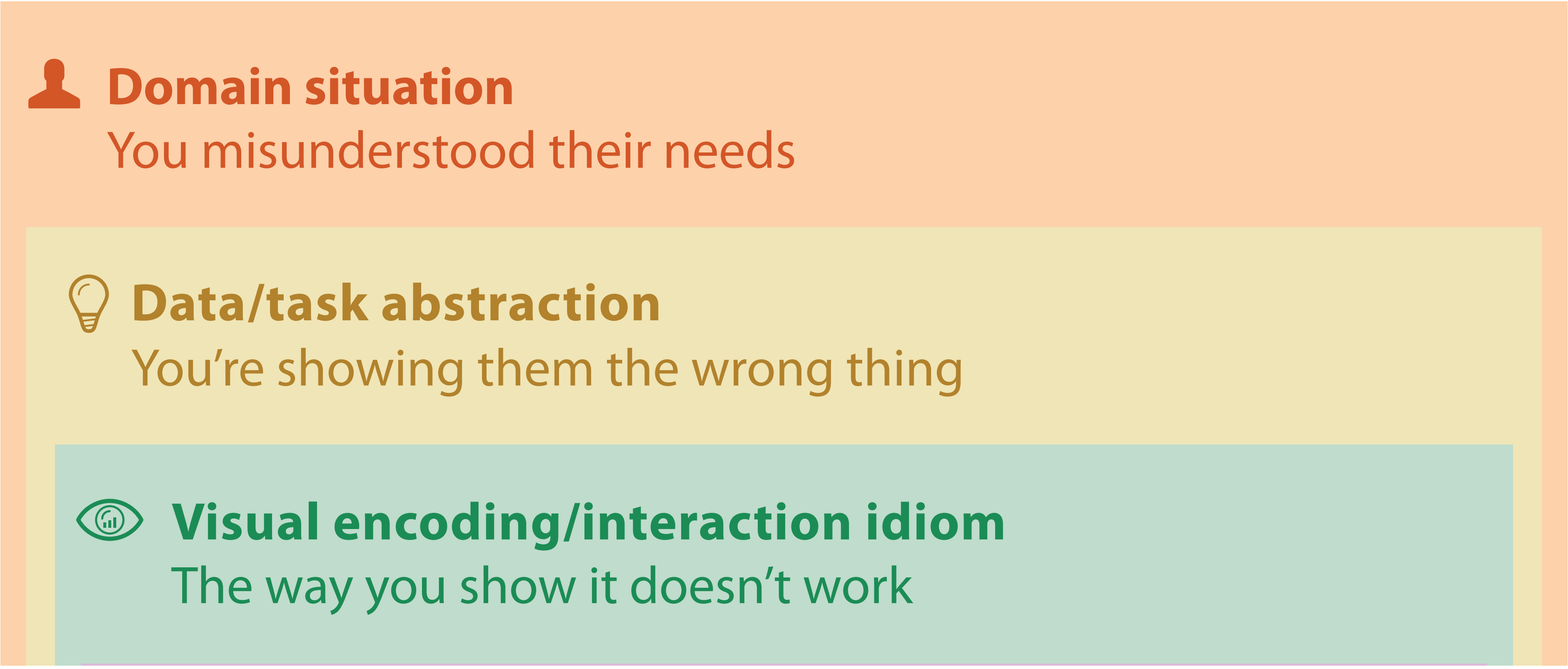
- different ways to get it wrong at each level



[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

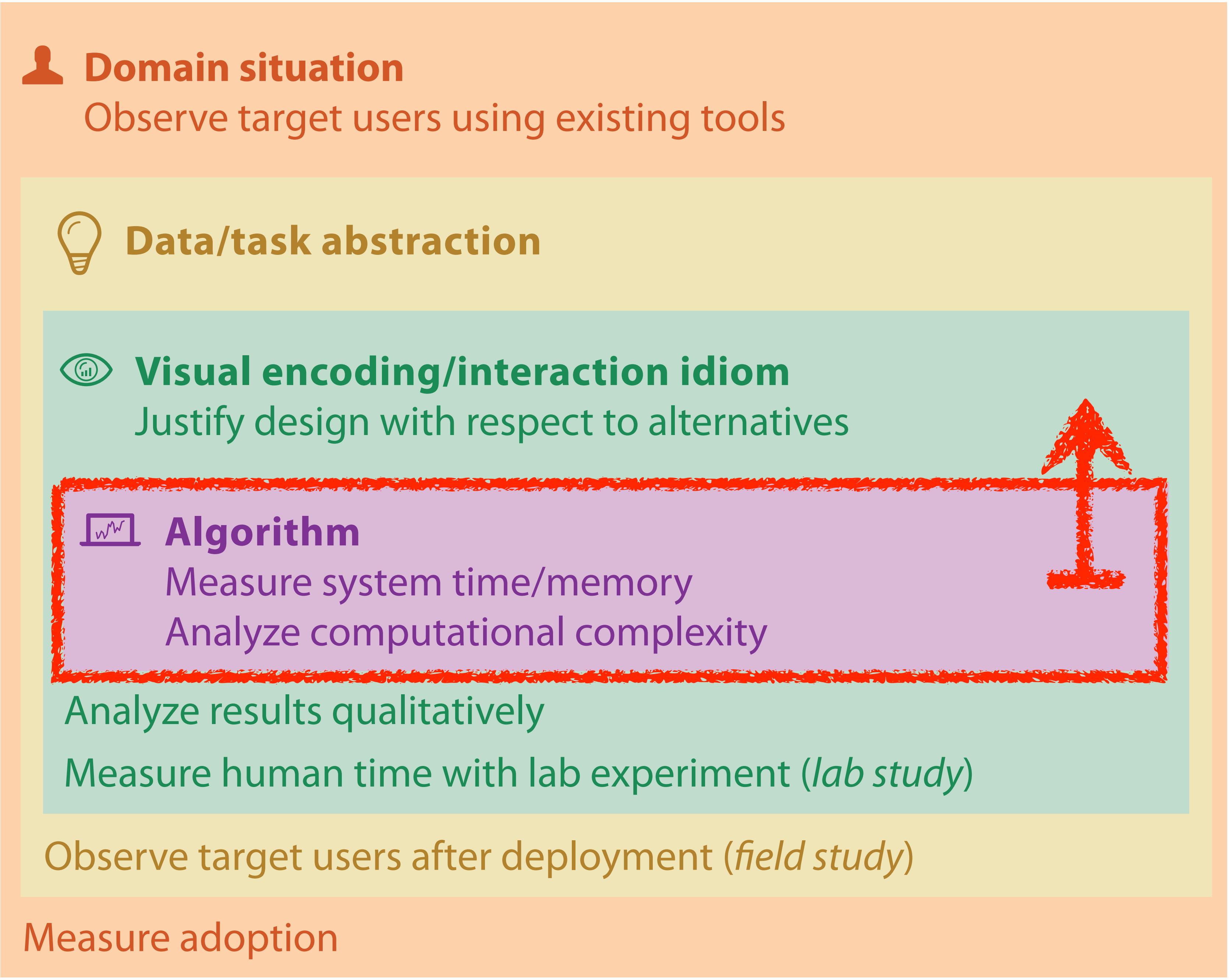
anthropology/
ethnography

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

anthropology/ ethnography

Domain situation
Observe target users using existing tools

Data/task abstraction

design

Visual encoding/interaction idiom
Justify design with respect to alternatives

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

cognitive psychology

Analyze results qualitatively

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

anthropology/ ethnography

Observe target users after deployment (*field study*)

Measure adoption

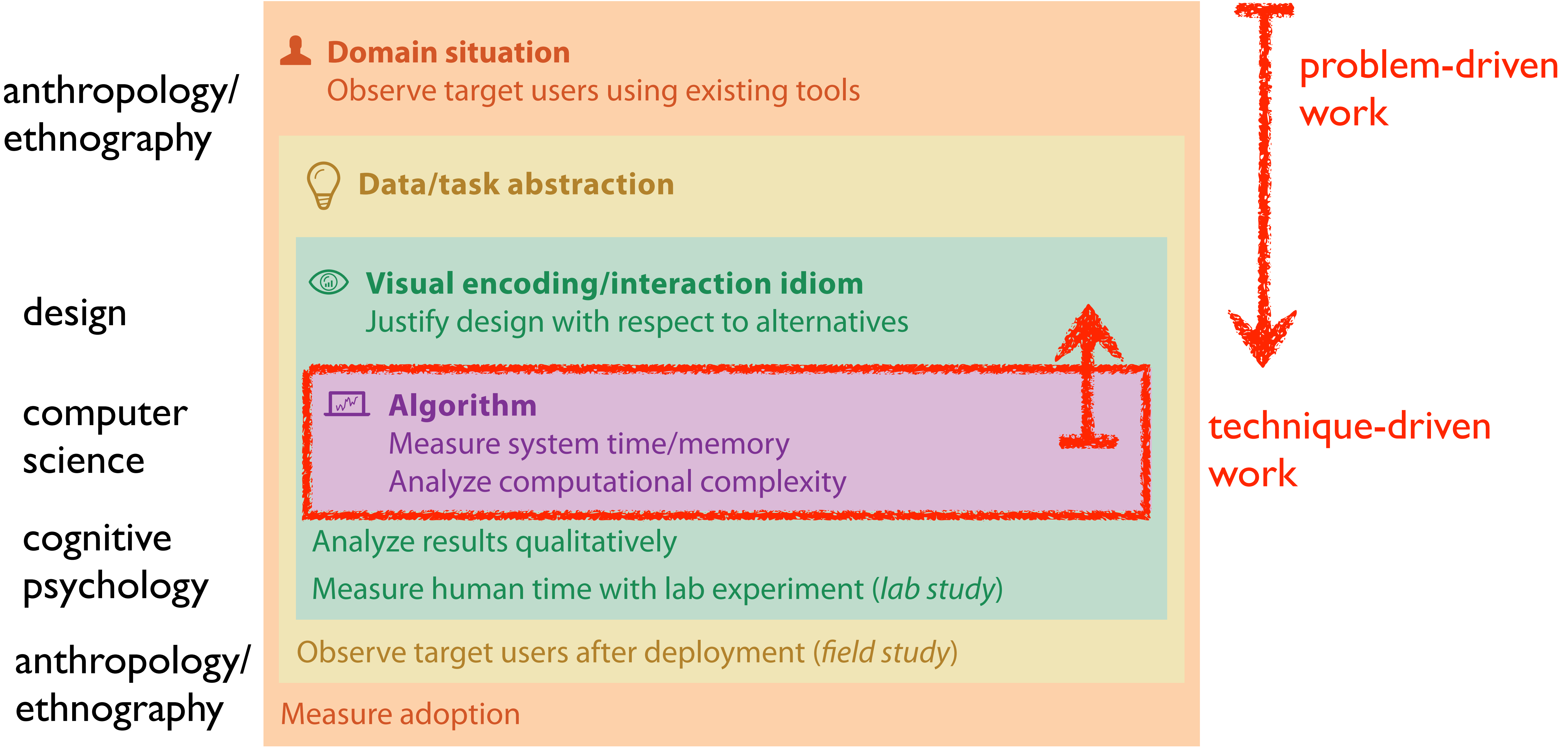
problem-driven work

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

- 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).]

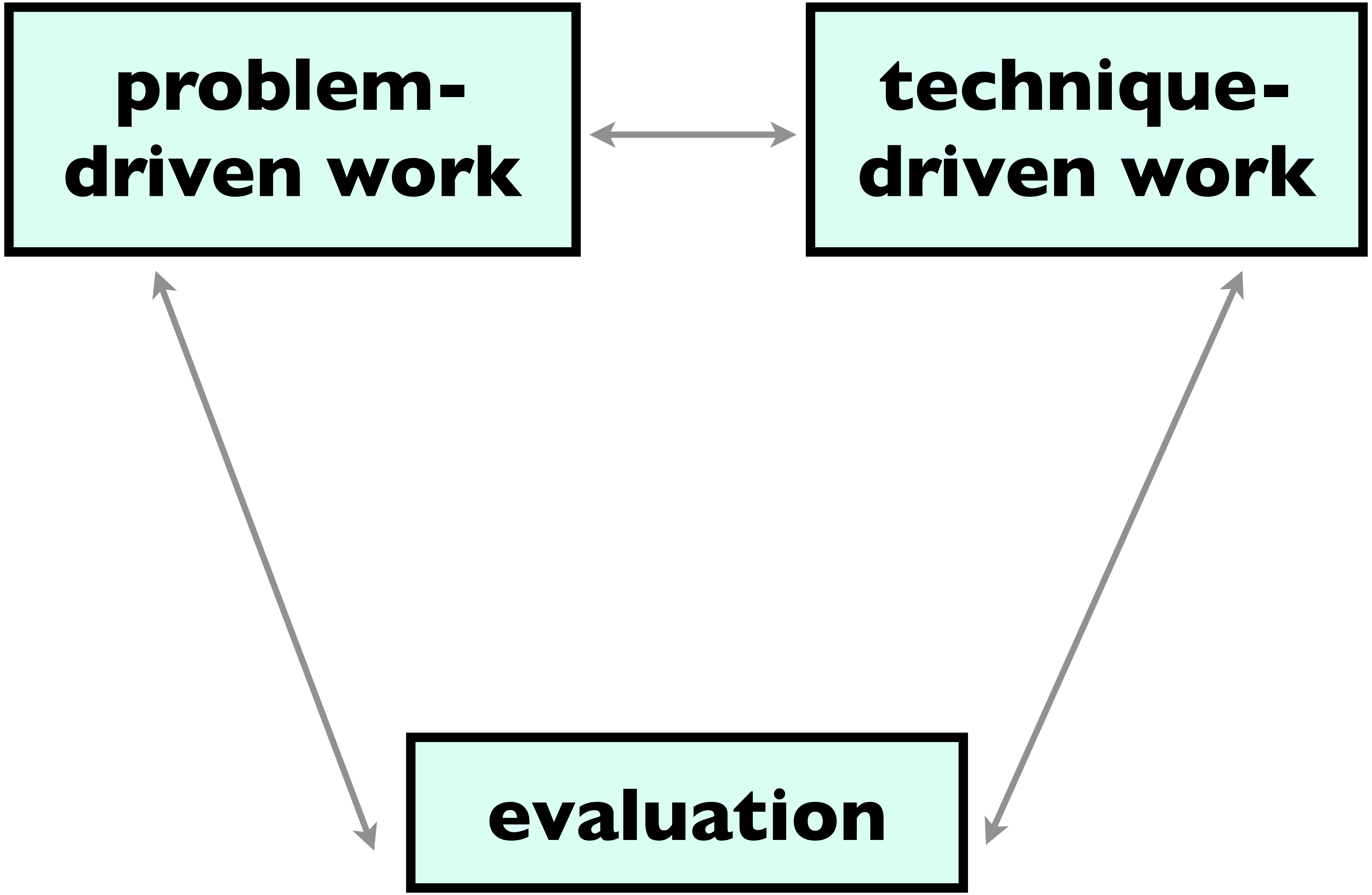
Visualization: Angles of attack

problem-driven work

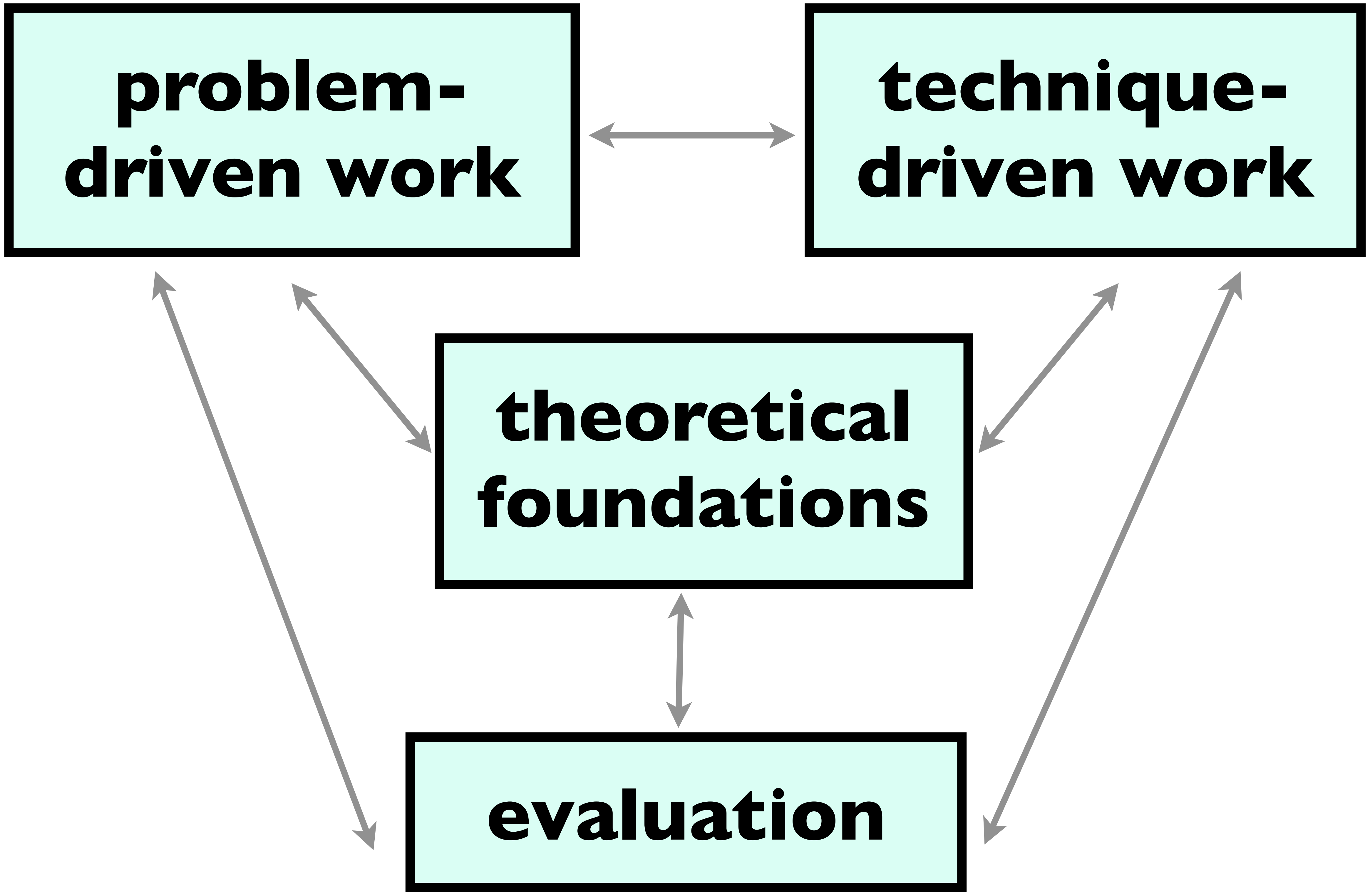
Visualization: Angles of attack

problem-driven work ⟷ **technique-driven work**

Visualization: Angles of attack



Visualization: Angles of attack



Three case studies: Abstractions & idioms

- e-commerce

- facilities management

- biology

Three case studies: Abstractions & idioms

- 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

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

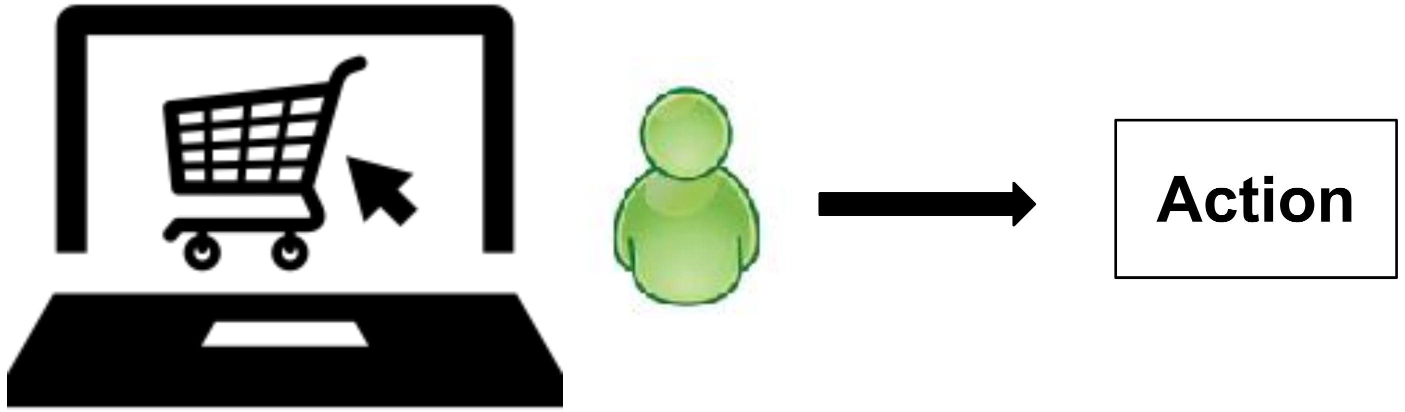
Clickstream Data

Clickstream Analysis Tasks

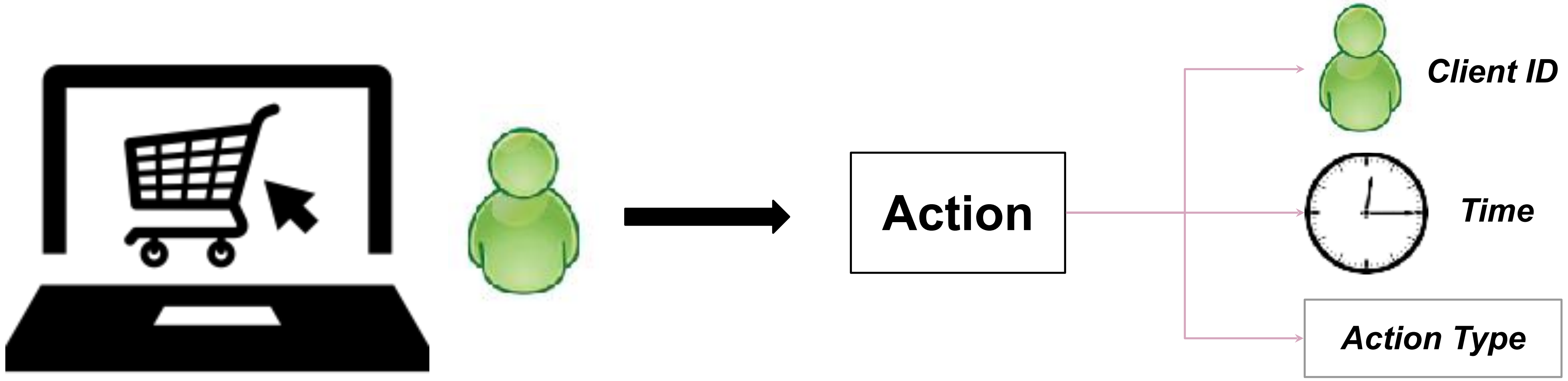
Segmentifier Analysis Model

What is ***Clickstream Data***?

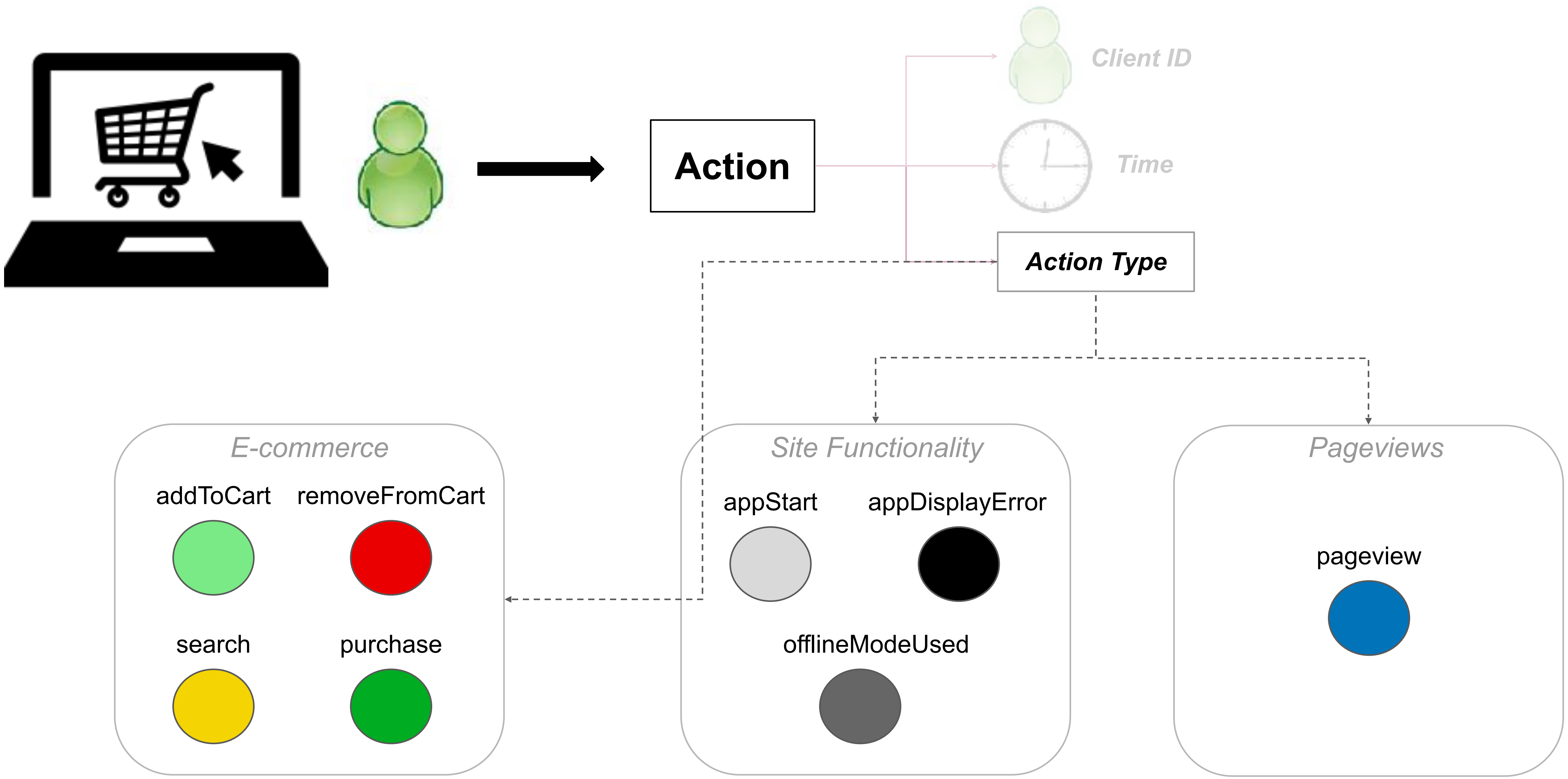
Data: *Actions*



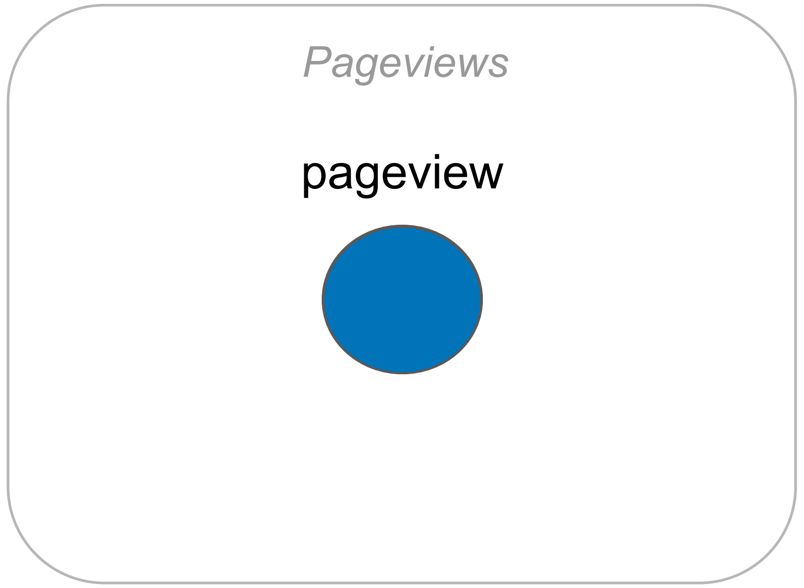
Data: *Action Attributes*



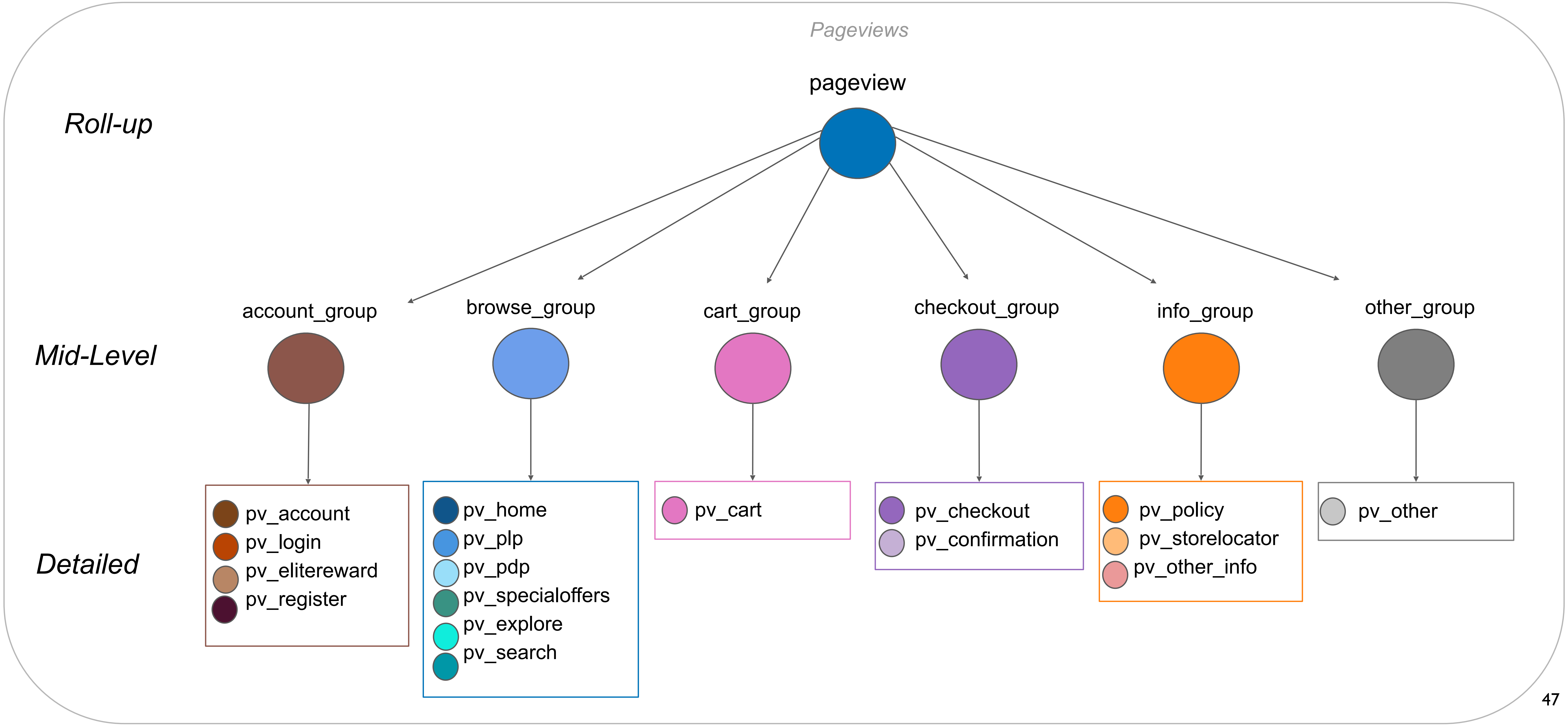
Data: *Action Types*



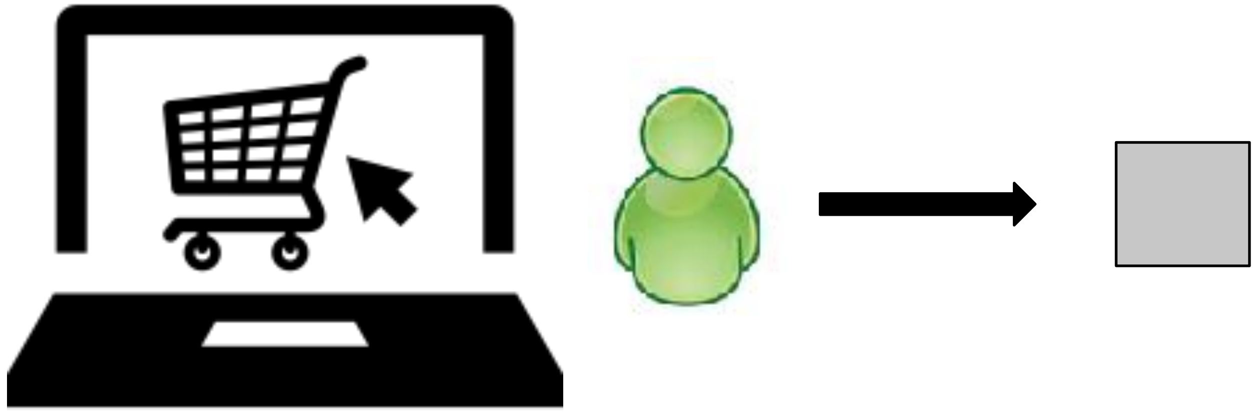
Action Hierarchy

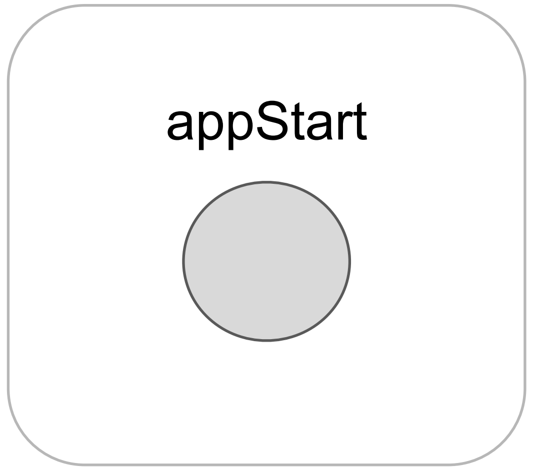


Action Hierarchy

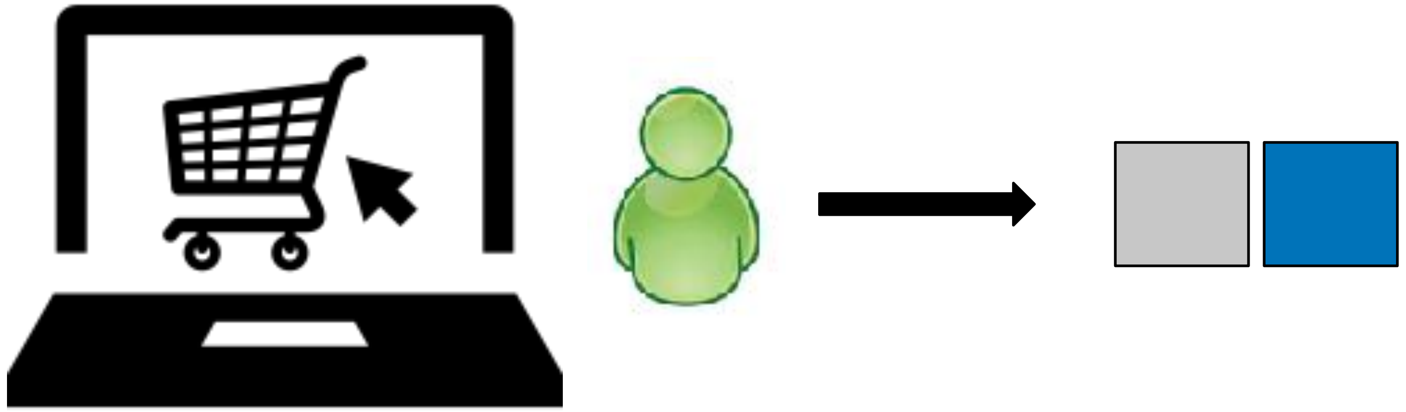


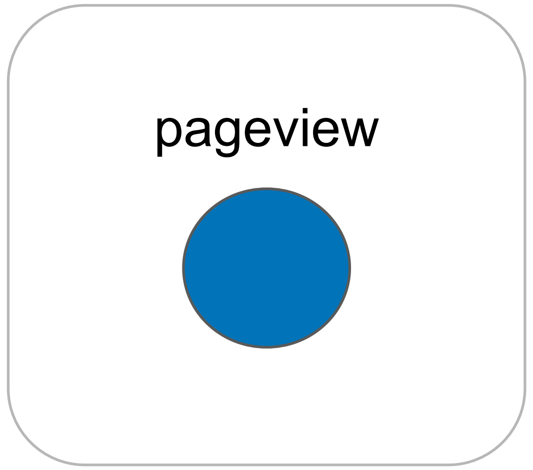
Data: *Sequences*



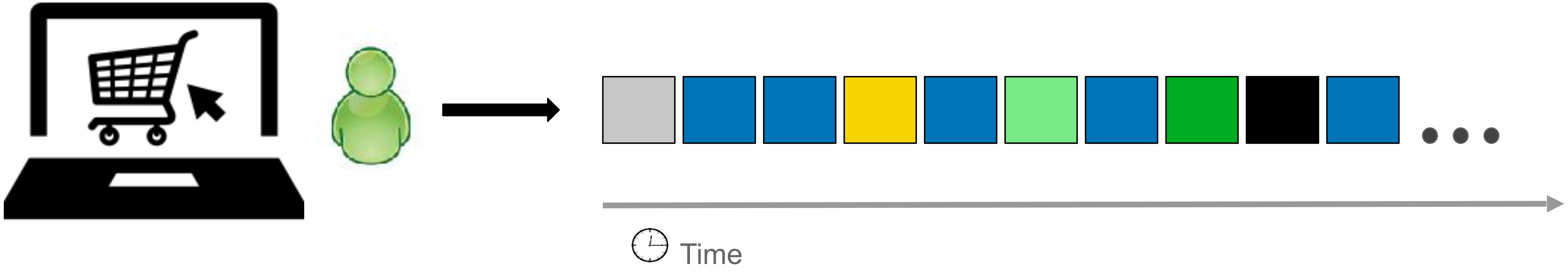


Data: *Sequences*



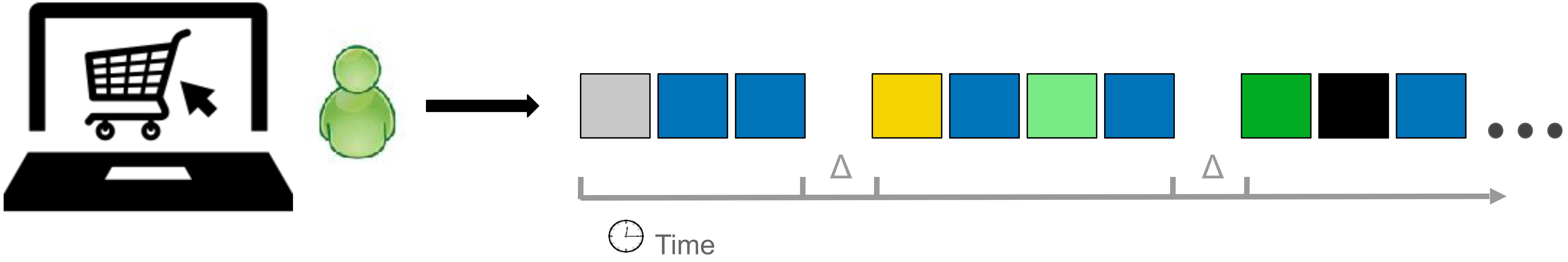


Data: *Client Sequences*



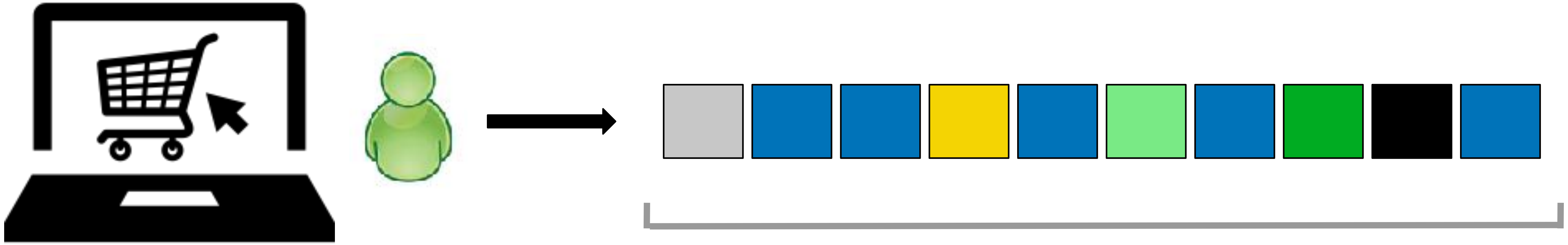
Client Sequences: all actions performed by a single user

Data: *Session Sequences*

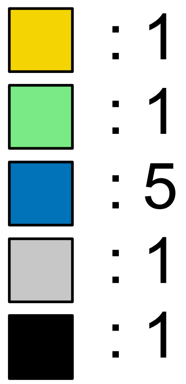


Session Sequences: all actions performed by a single user within a defined amount of time (Δ) from each other.
Δ is usually 30 min.

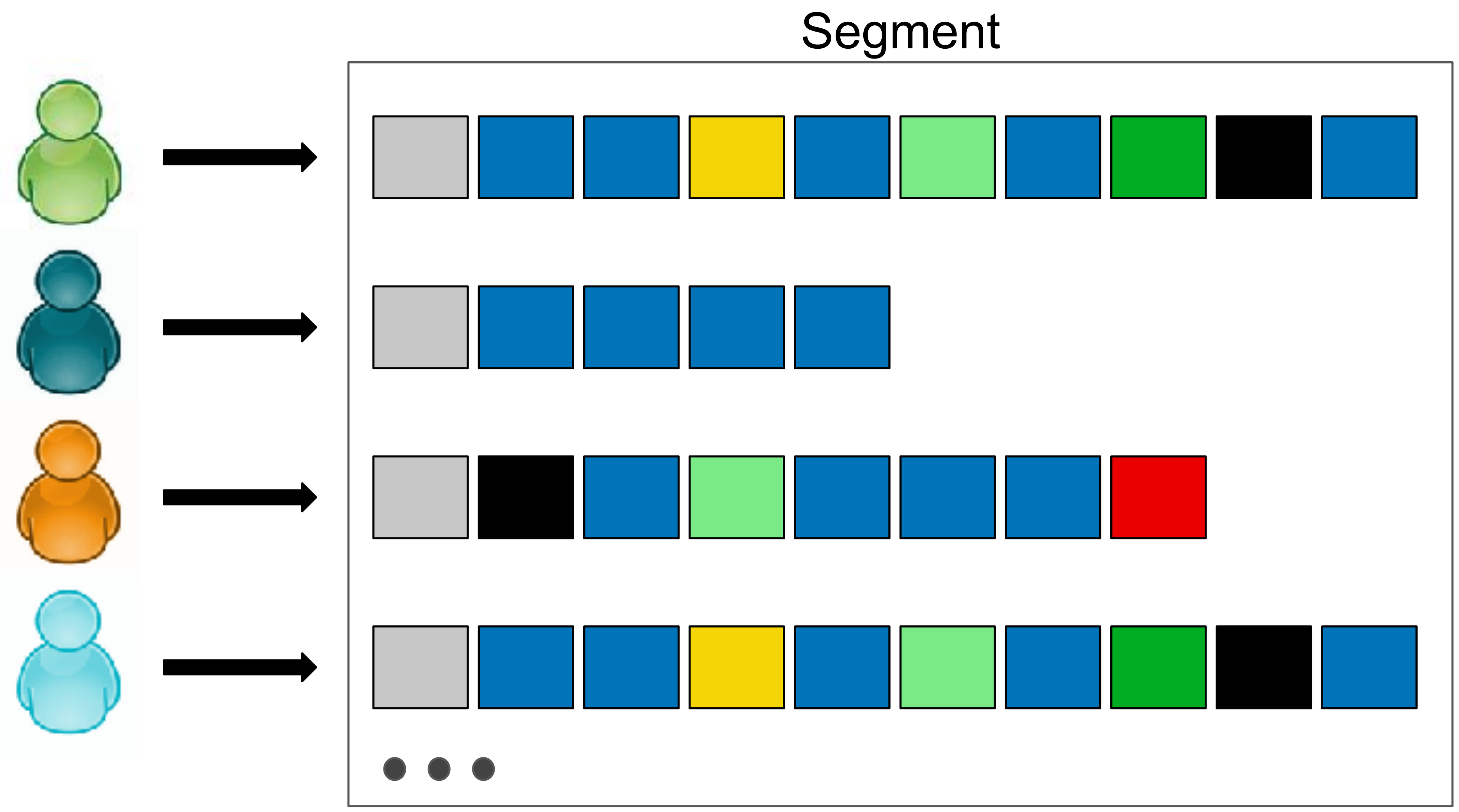
Data: *Sequence Attributes*



Start time | End time | Duration | Action Counts

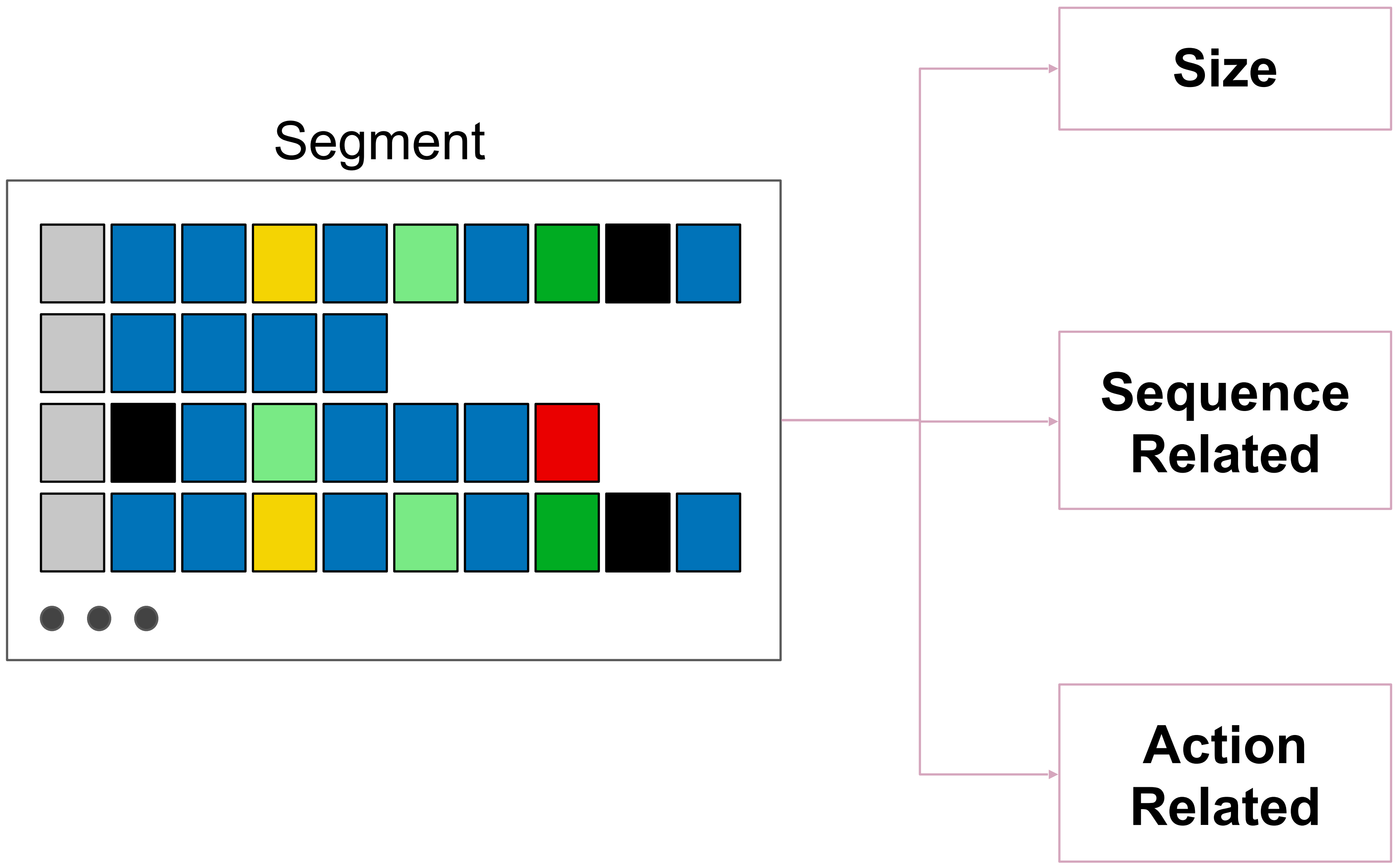


Data: *Segments*



Segment: any set of sequences

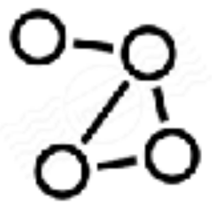
Data: *Segment Attributes*



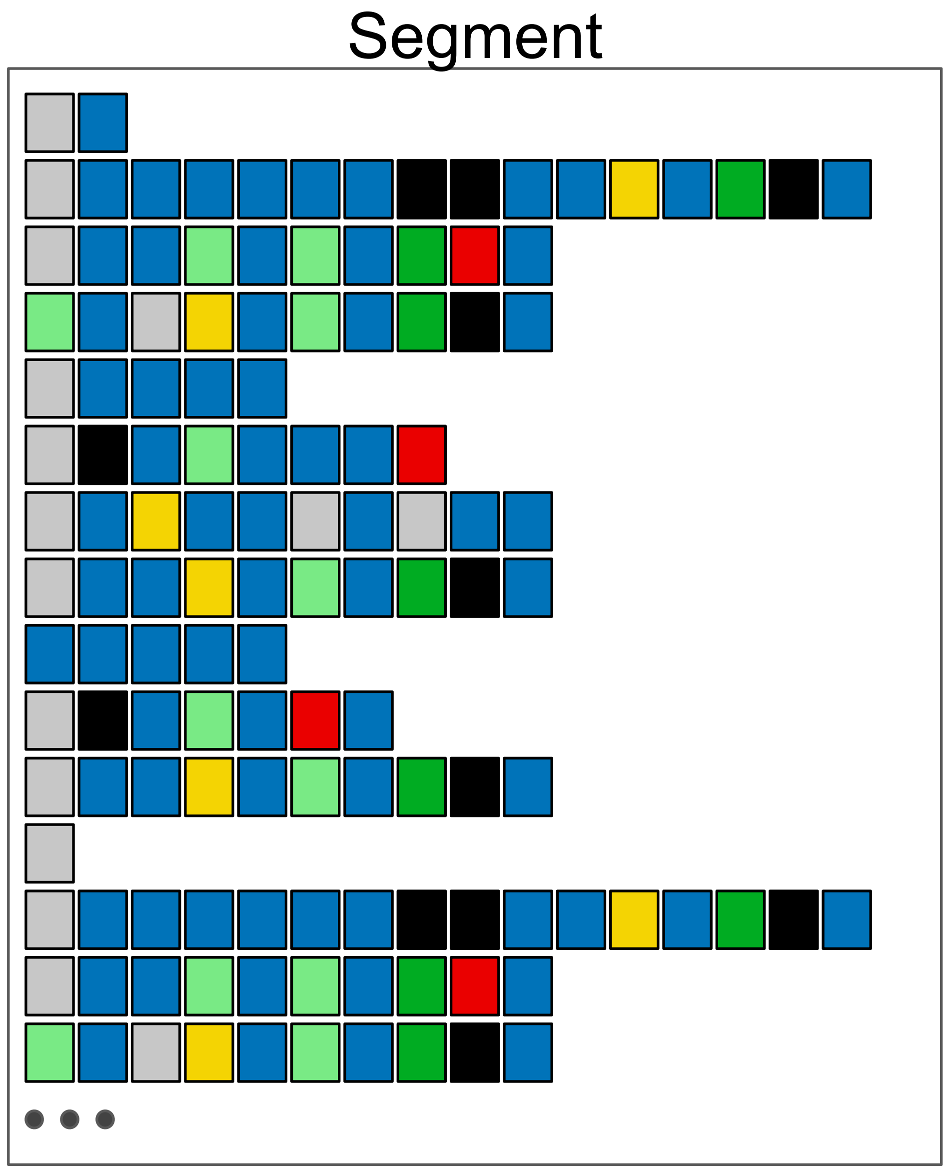
Counts of sequences:
Absolute, Relative

Sequence Distributions:
Start Time, Duration, Action Counts

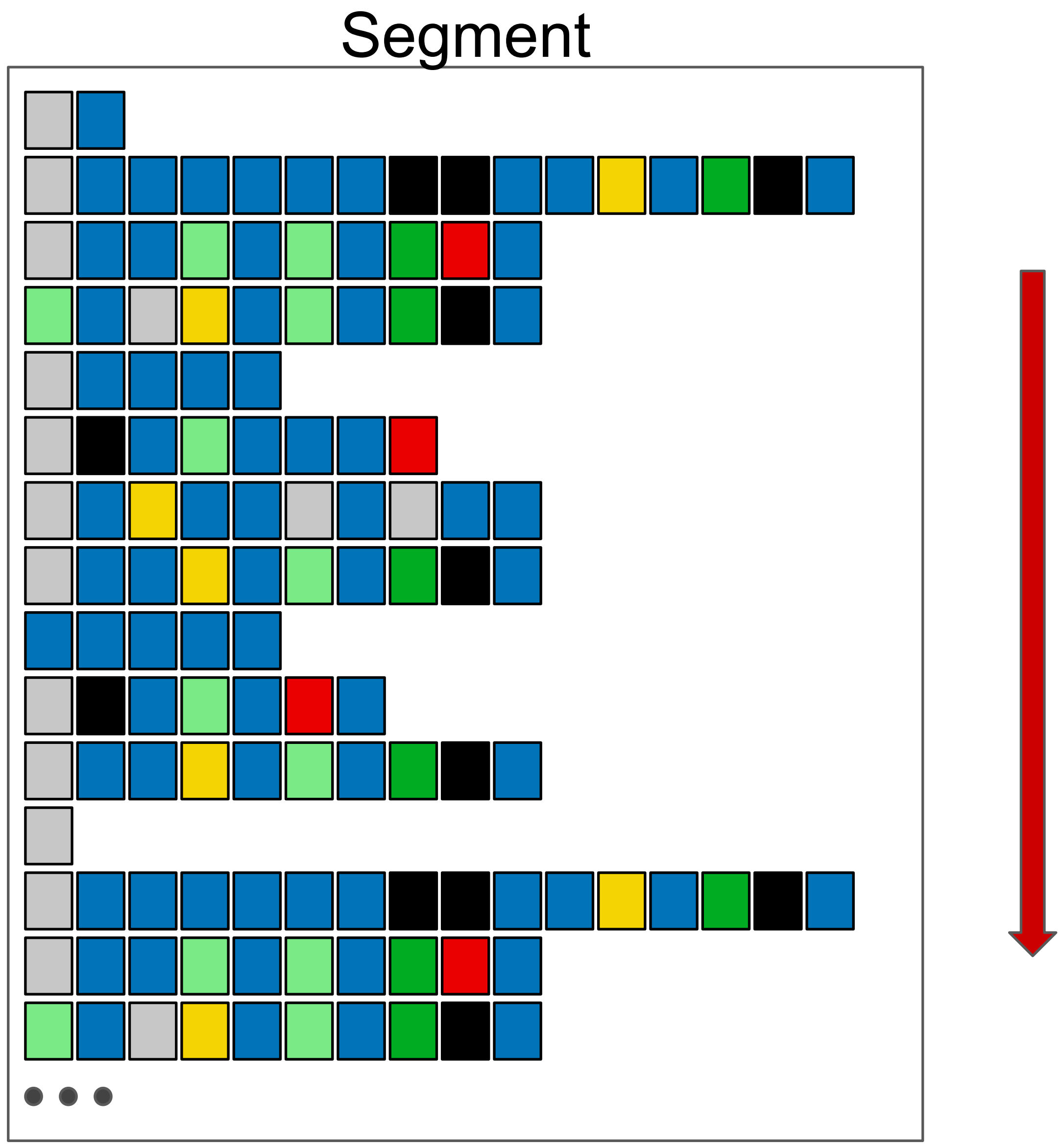
Action Distributions:
Action Transitions:
action before, action after



Real-world Clickstream Data

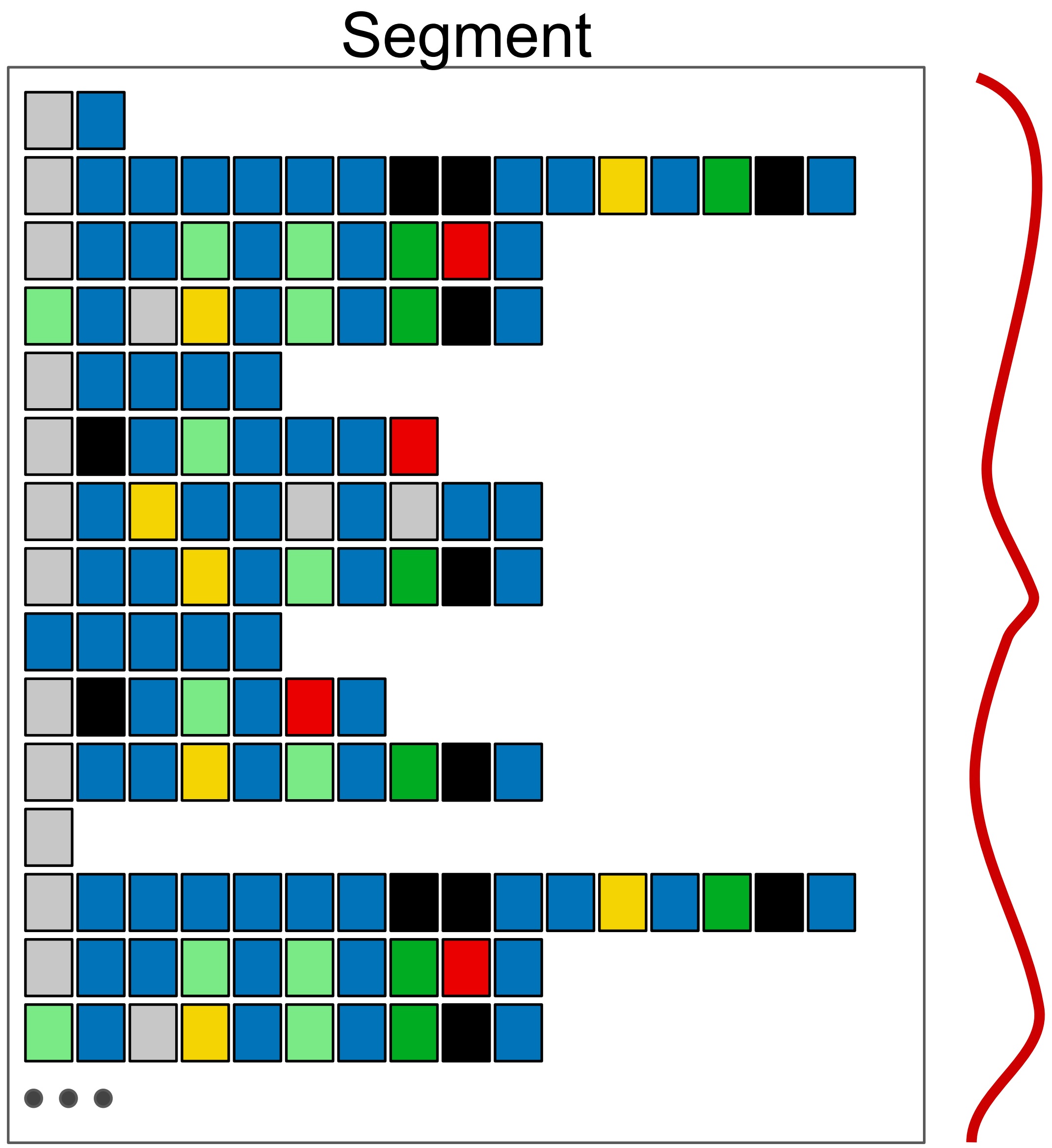


Real-world Clickstream Data



Scale is huge

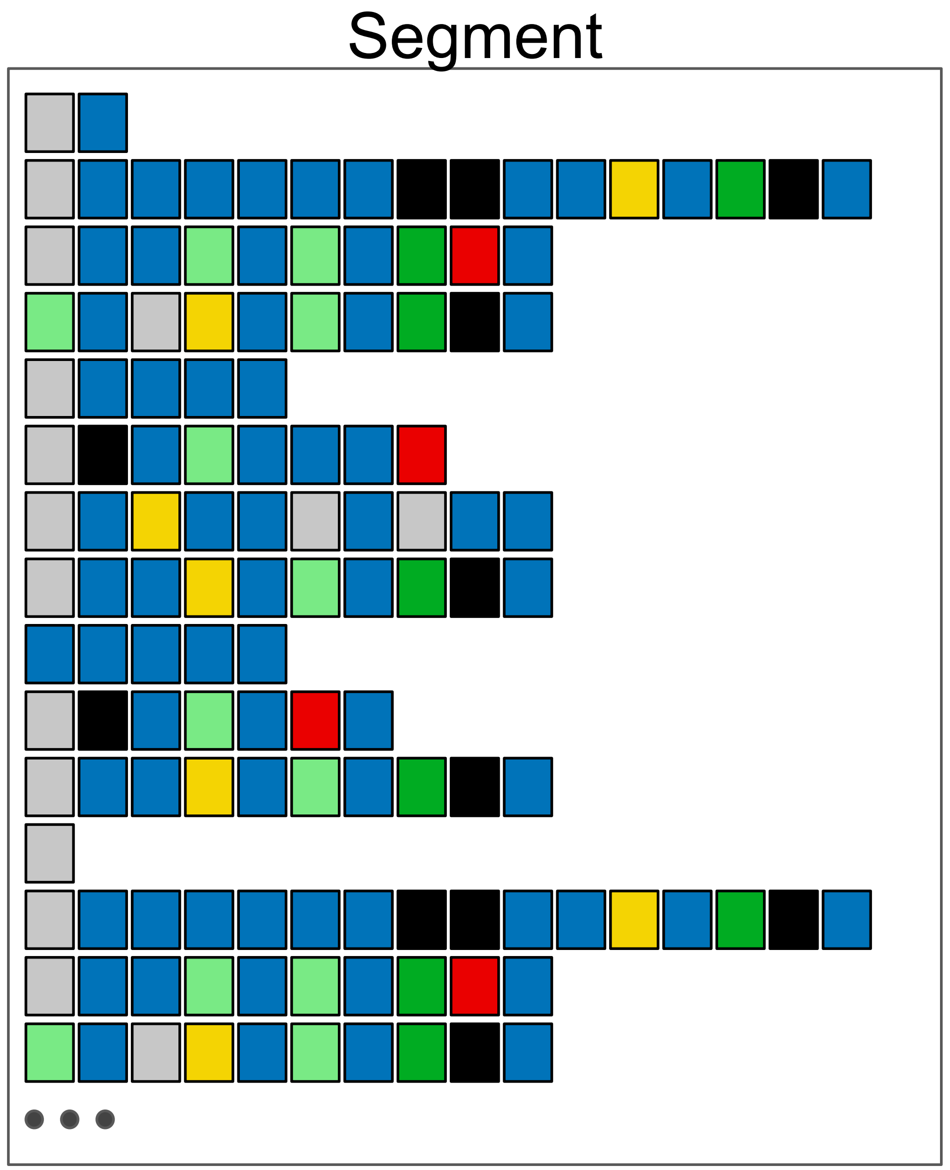
Real-world Clickstream Data



Scale is huge

Variability is high

Real-world Clickstream Data



Scale is huge

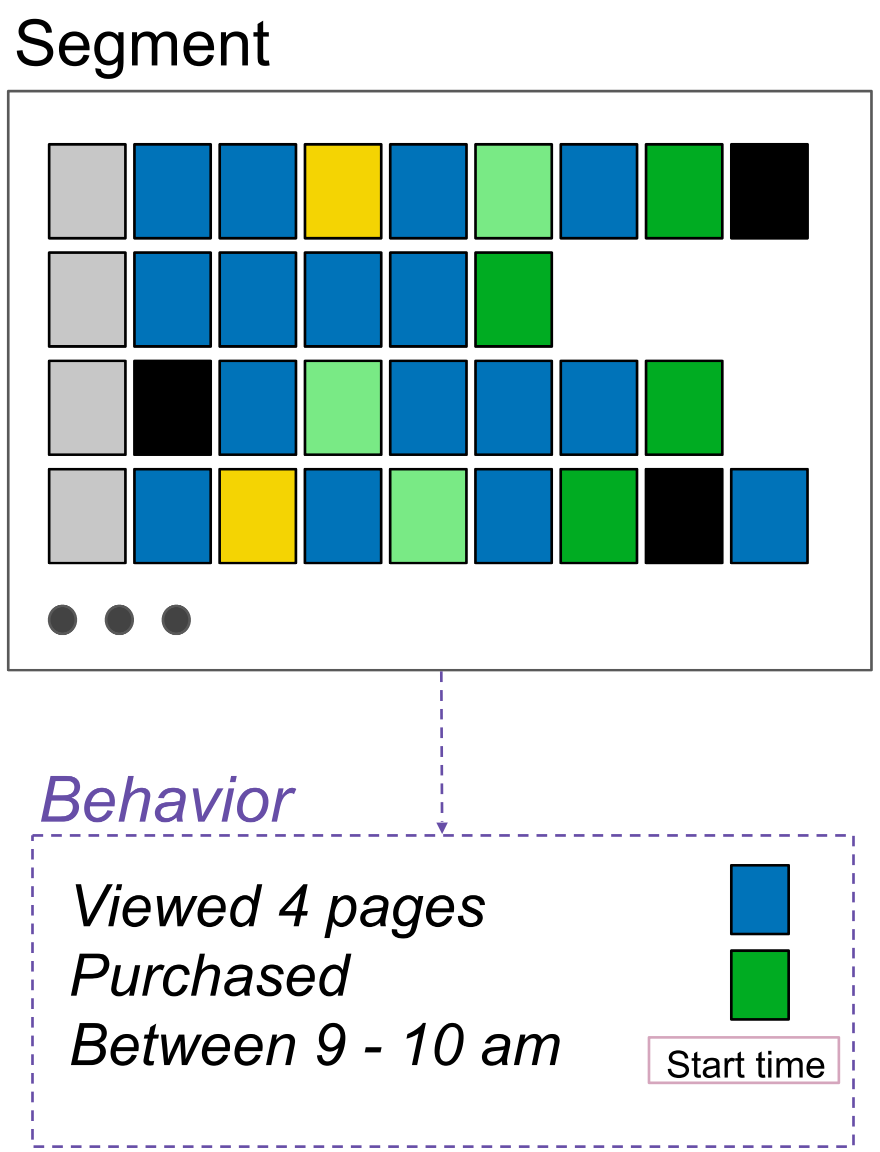
Variability is high

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

What are

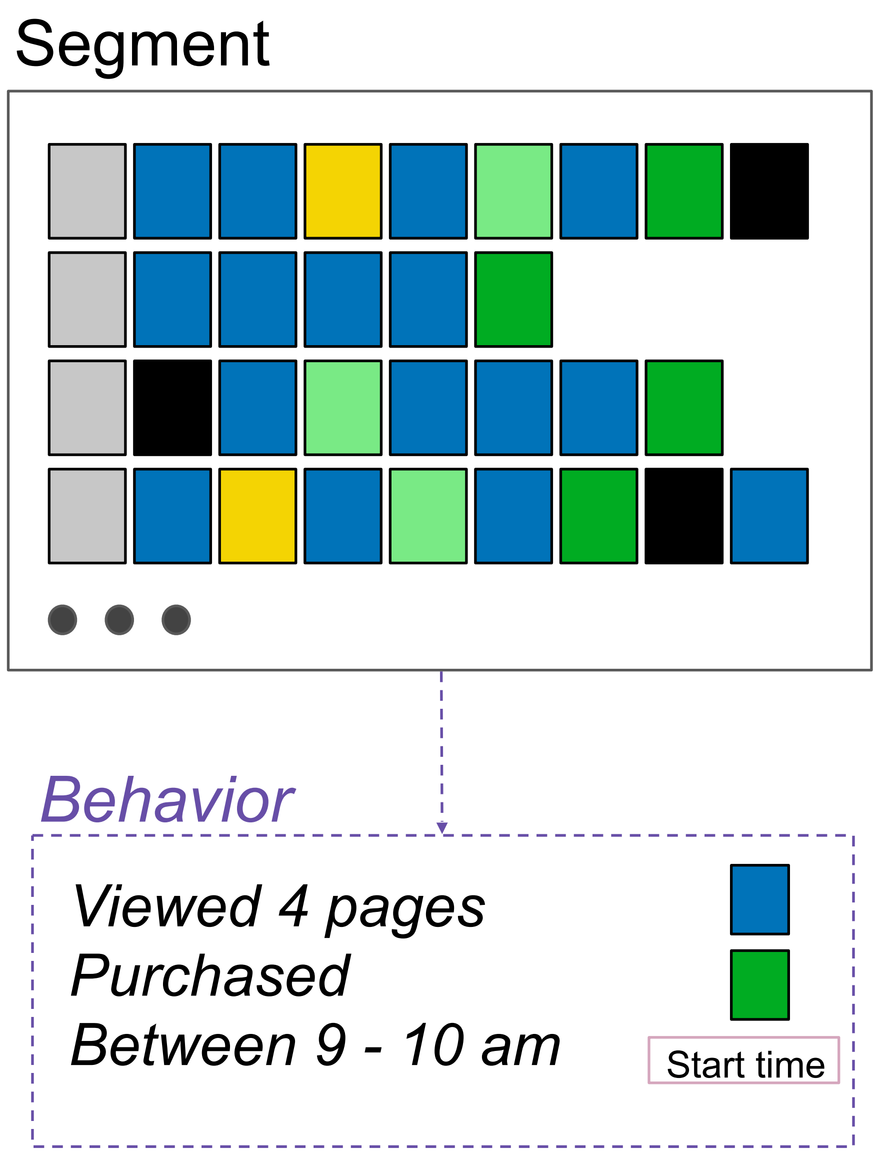
Clickstream Data Analysis Tasks?

Tasks: Segment Behavior



Behavior: set of attribute constraints

Tasks: Segment Behavior

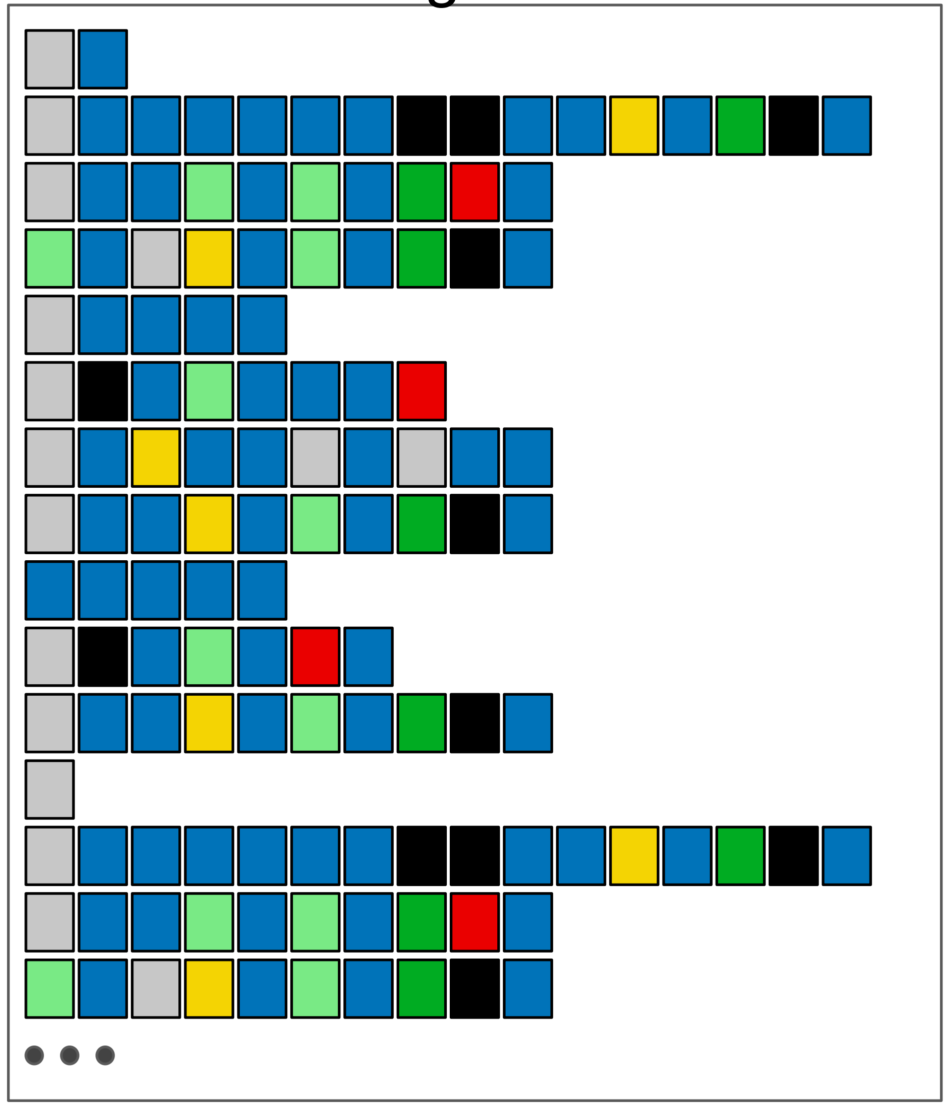


Behavior: set of attribute constraints

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

Tasks: Task Abstraction

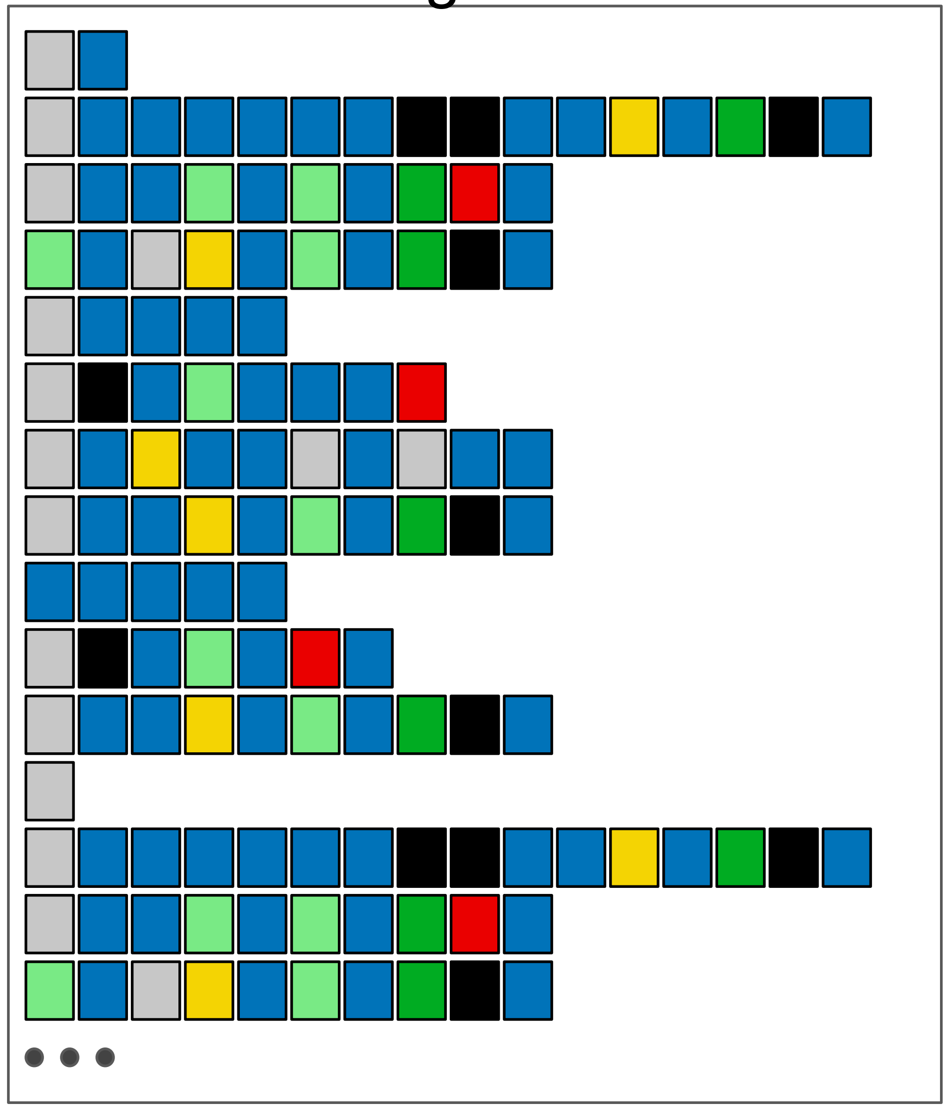
Segment



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

Tasks: Task Abstraction

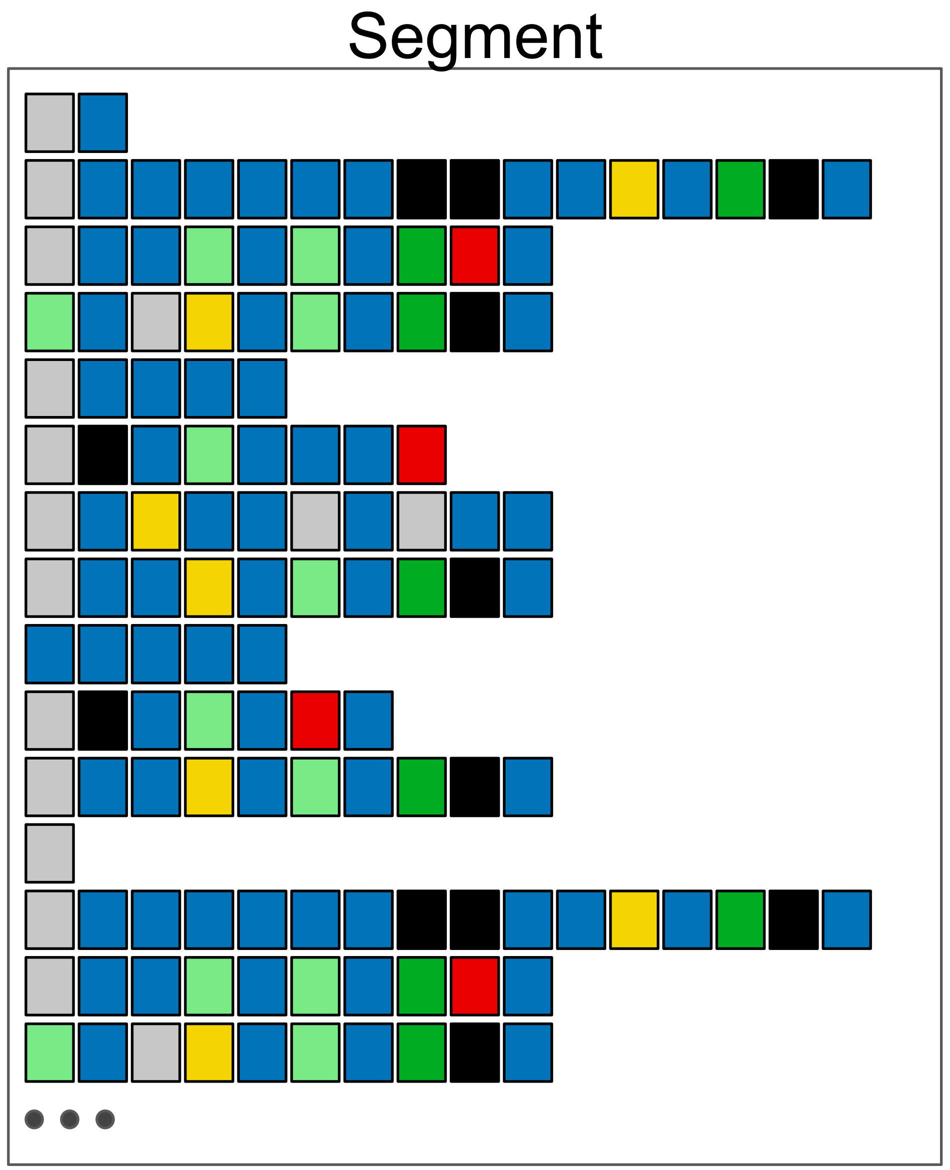
Segment



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



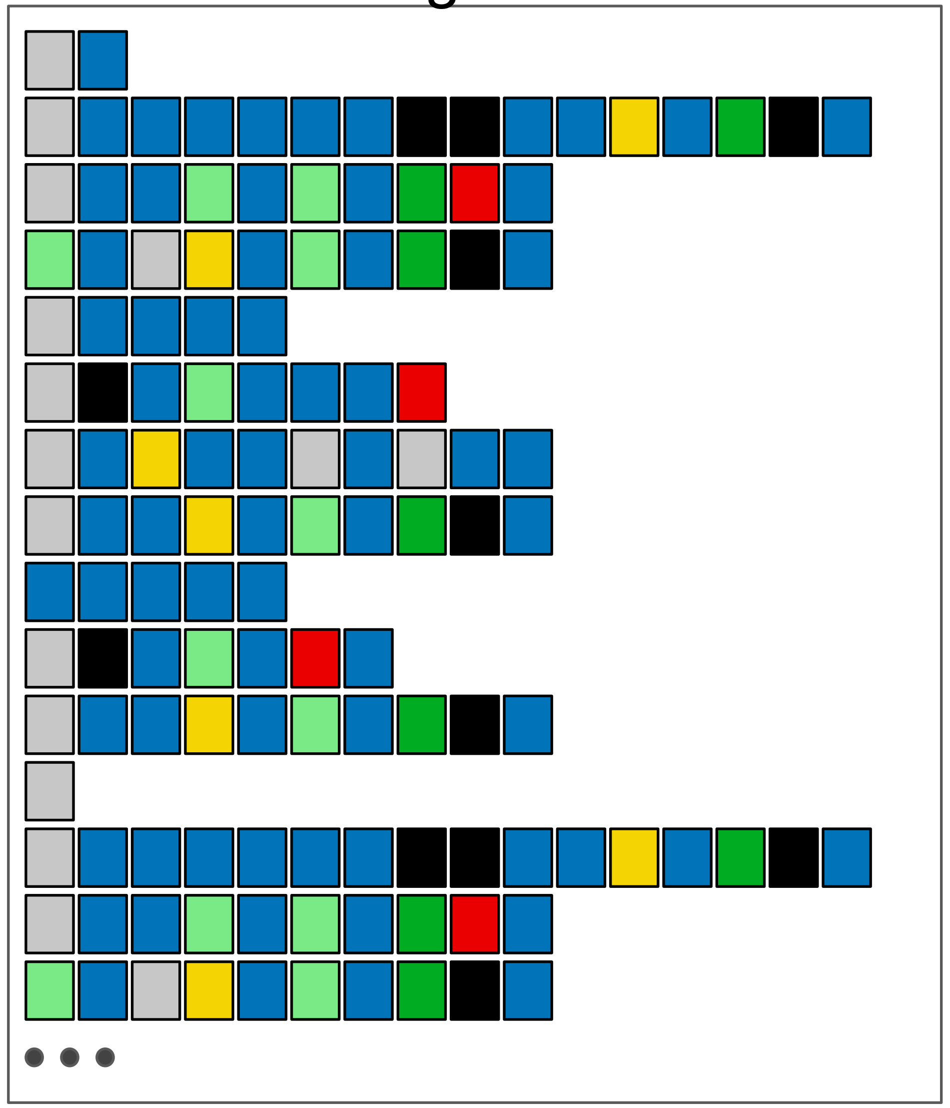
Identify: Find some set of sequences that constitutes interesting *behavior*

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

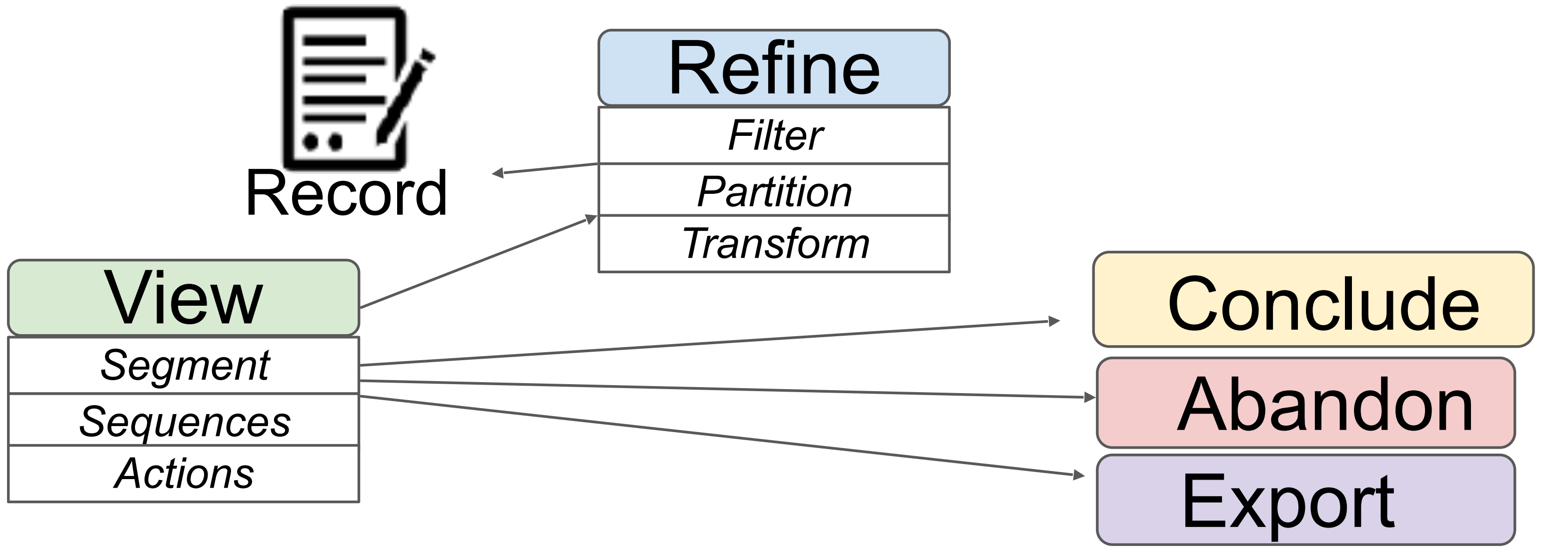
High-Level Segmentifier Analysis Model

High-Level Segmentifier Analysis Model

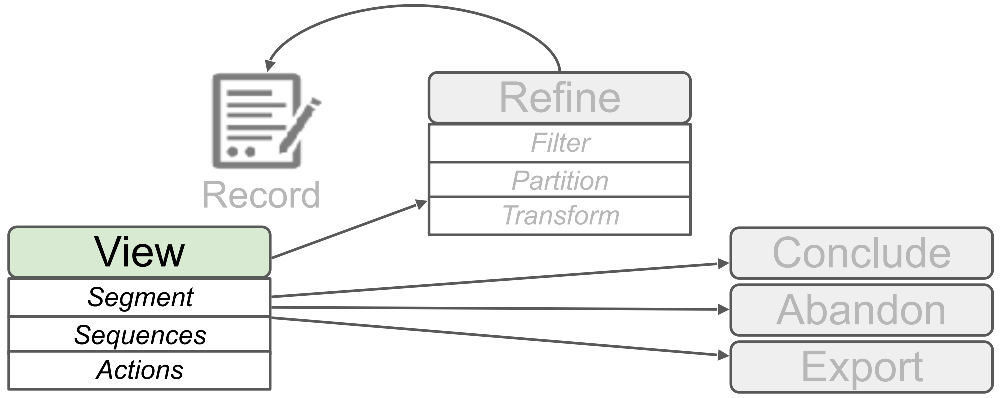
- Abstraction above task/data level to provide design rationale

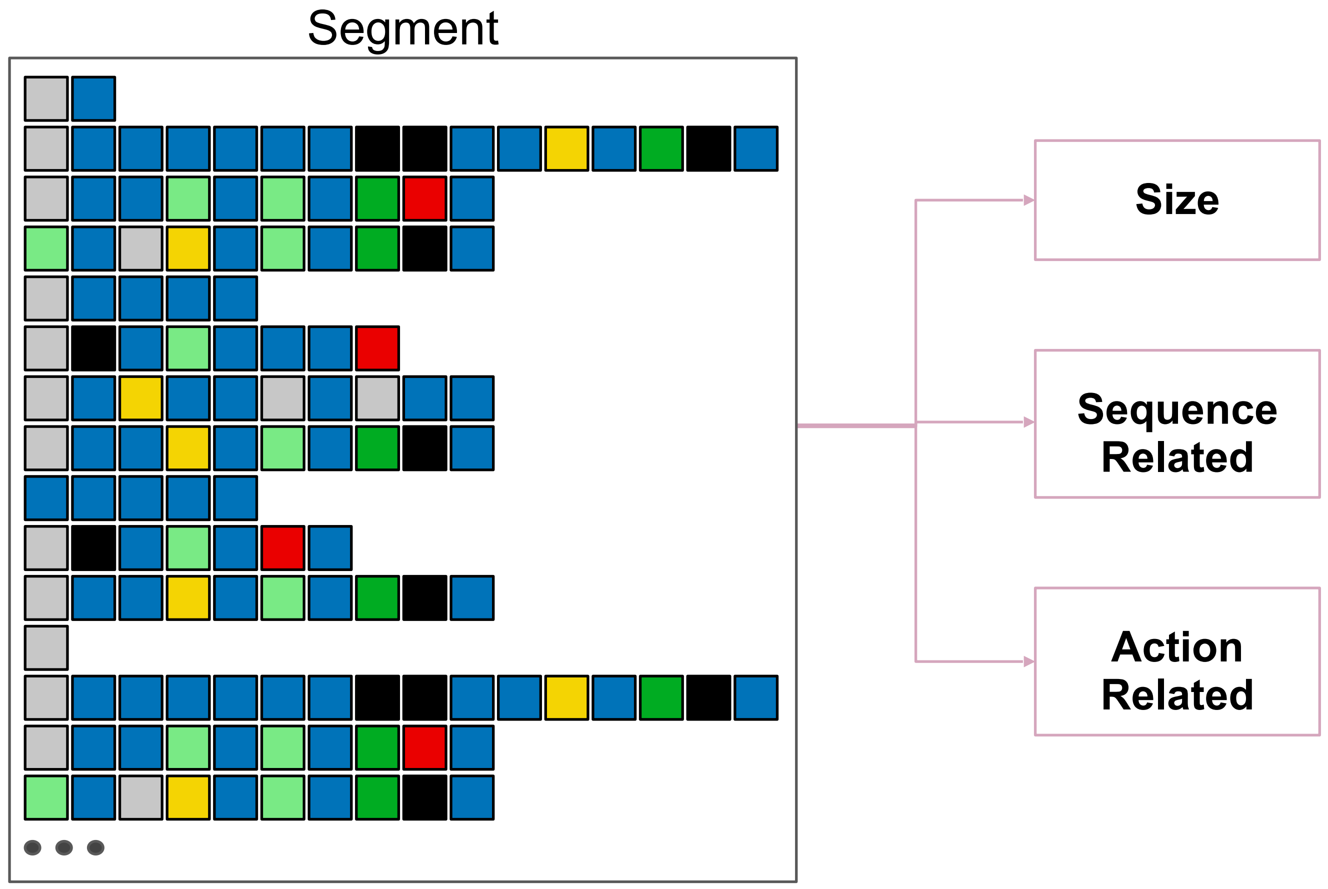
- Take a *giant, noisy dataset* and refine it into *small, clean segments* for
  - actionable insights
  - downstream analysis

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



High-Level Segmentifier Analysis Model

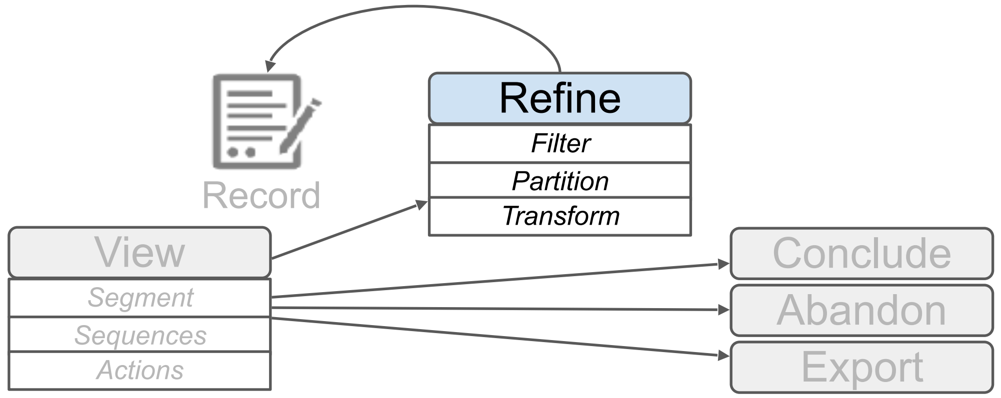


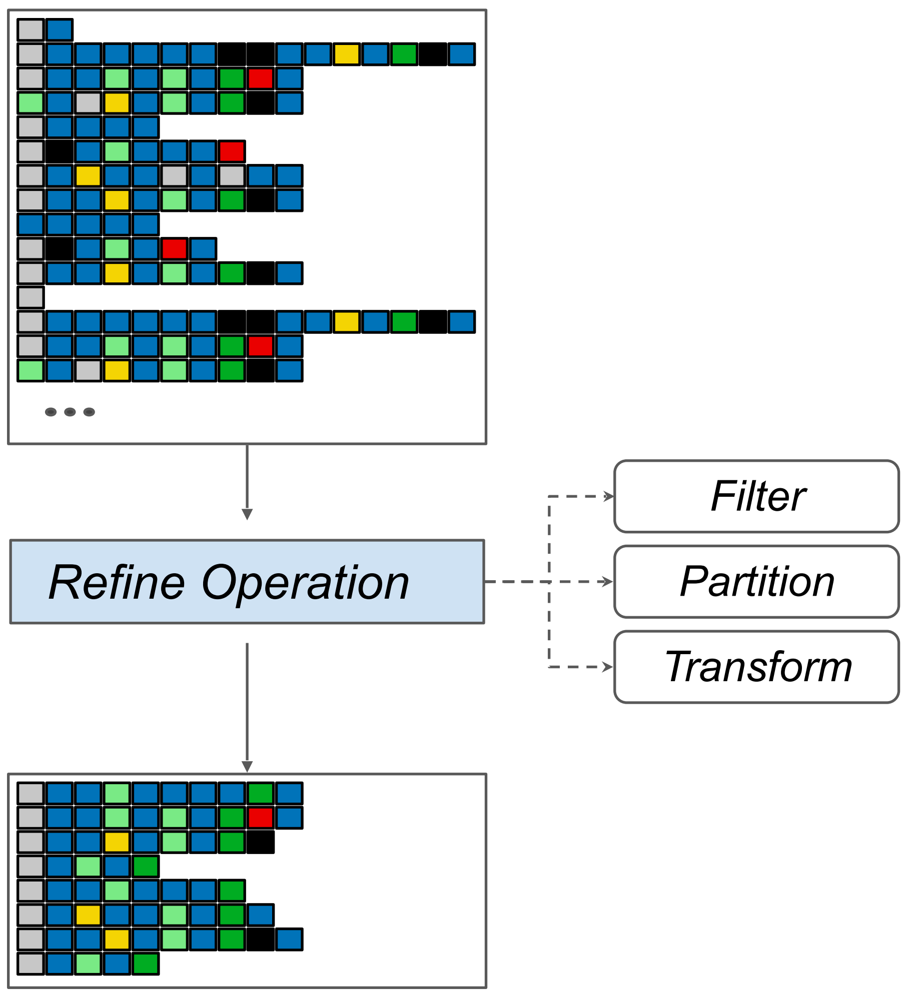


View

- 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

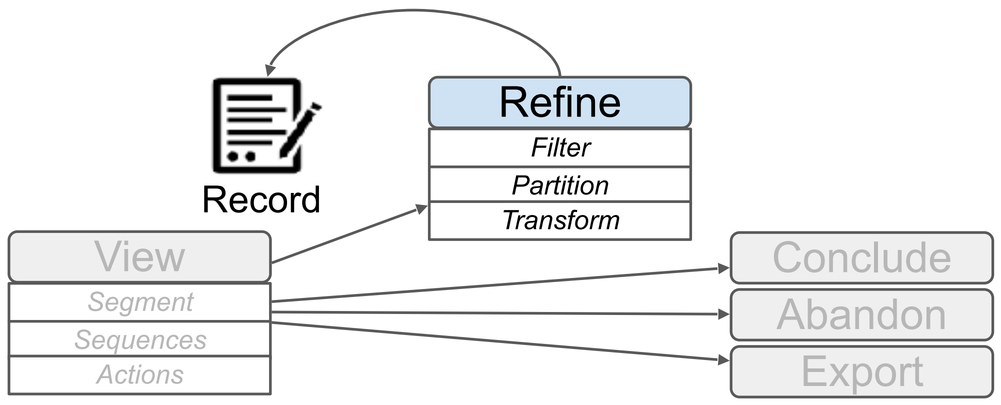


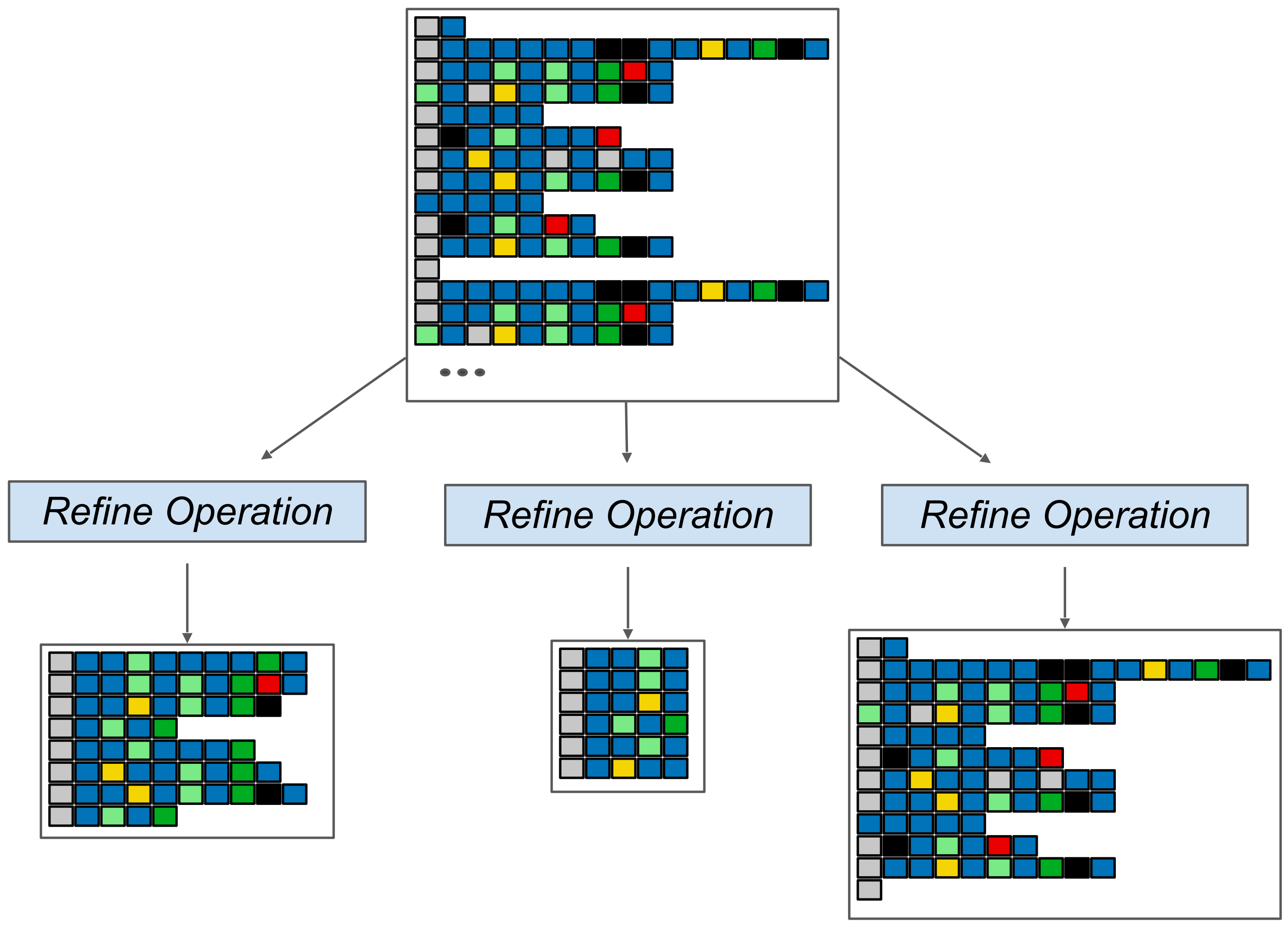


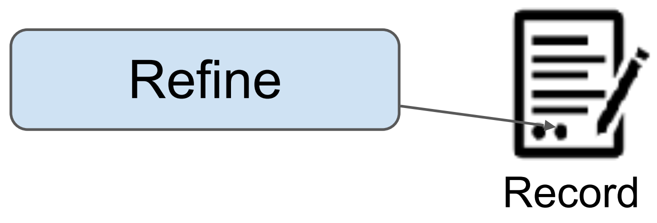
Refine

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

High-Level Segmentifier Analysis Model

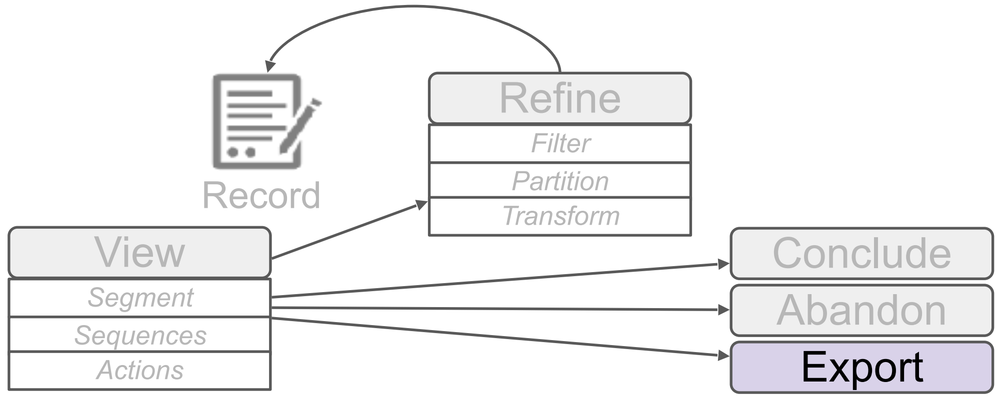


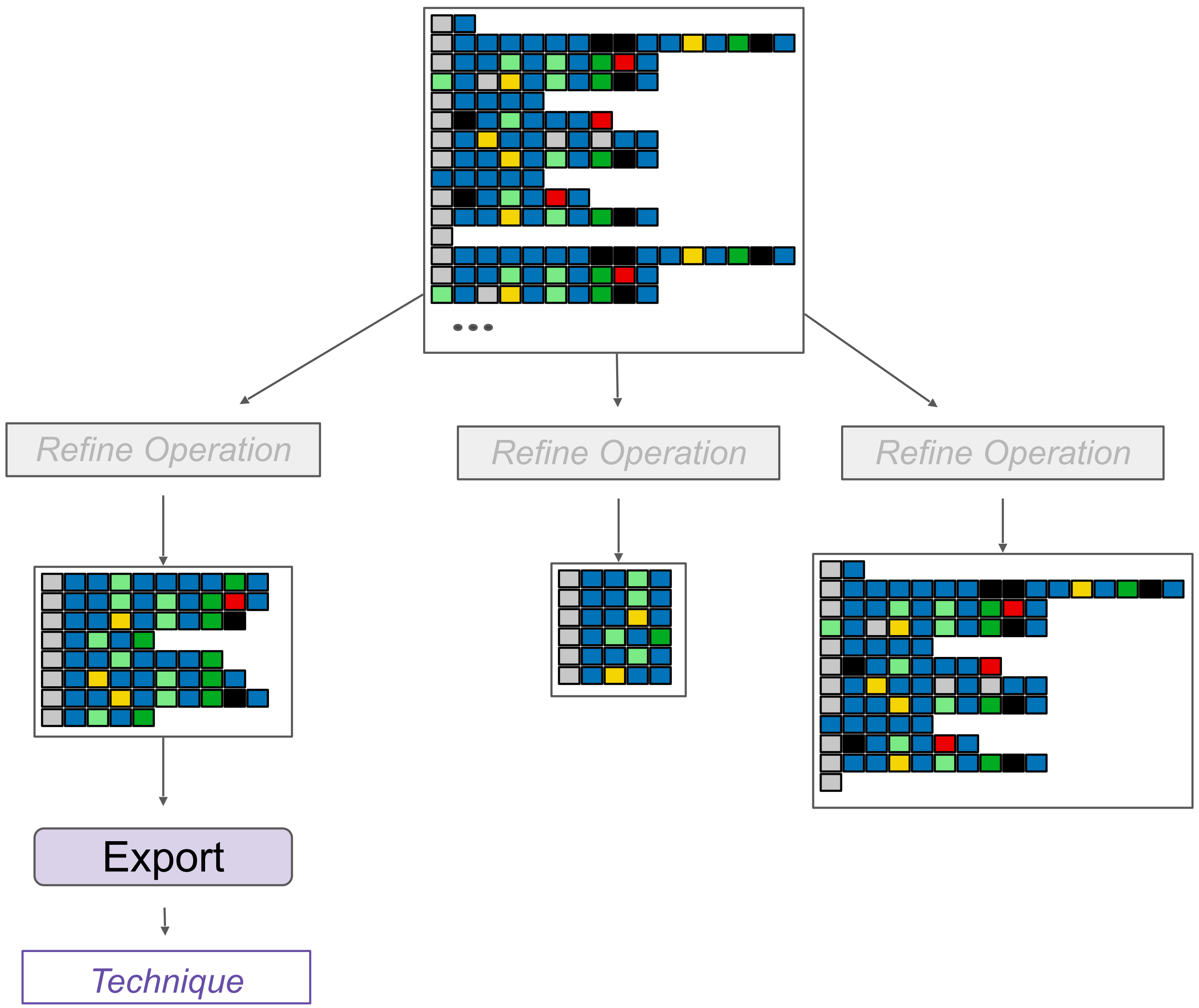




- Record all refinement steps automatically
- Keep track of questions asked and hypotheses tested
- Ability to create and view multiple segments from the same segment

High-Level Segmentifier Analysis Model

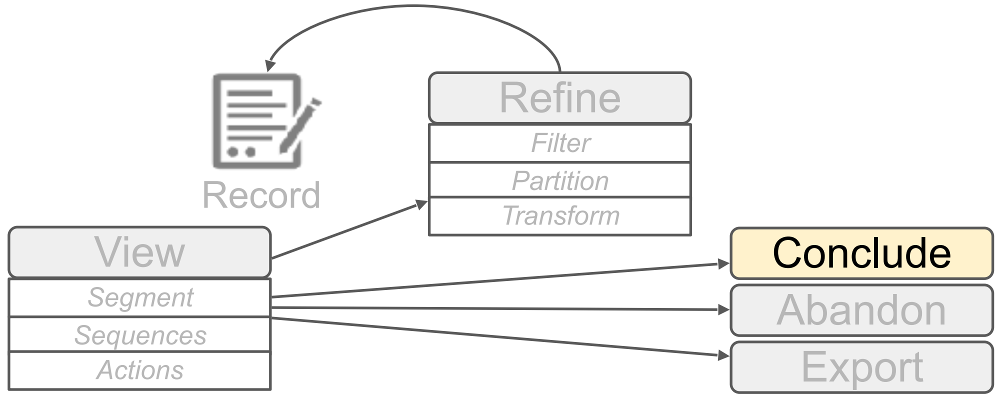


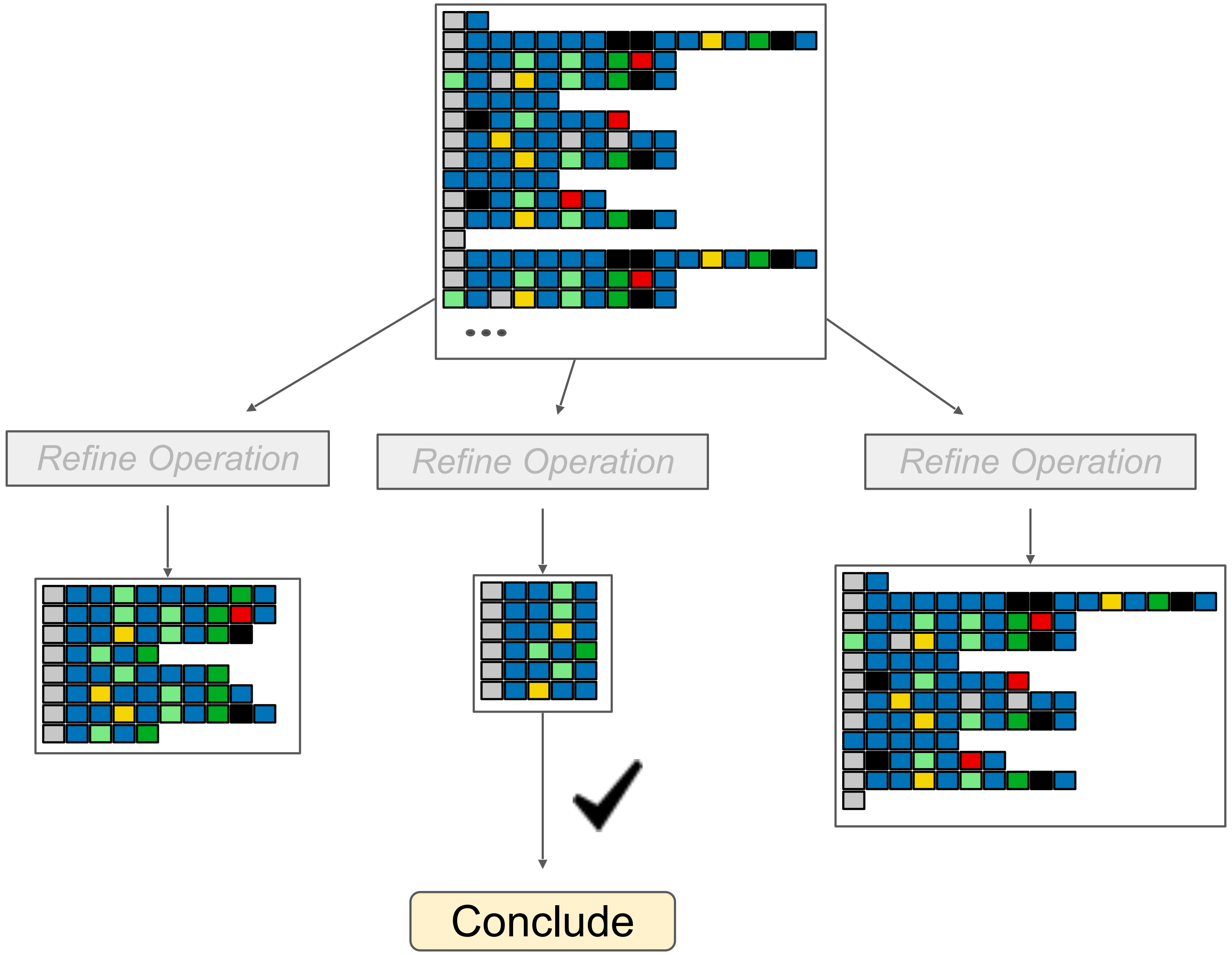


Export

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

High-Level Segmentifier Analysis Model

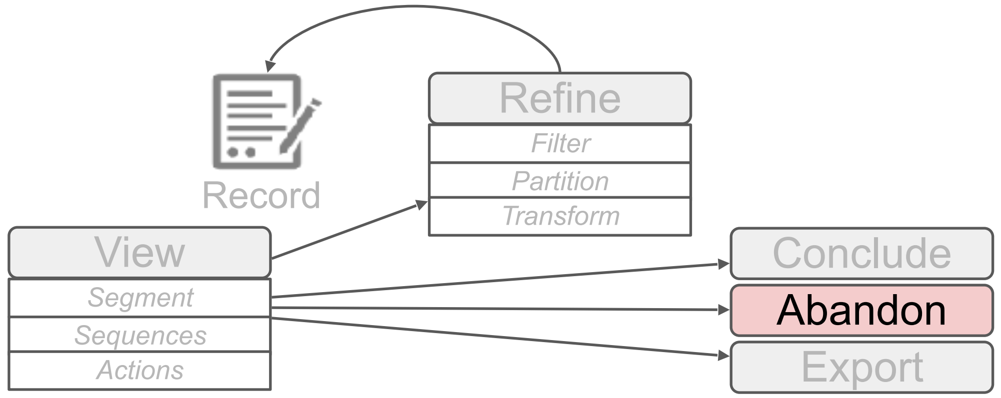


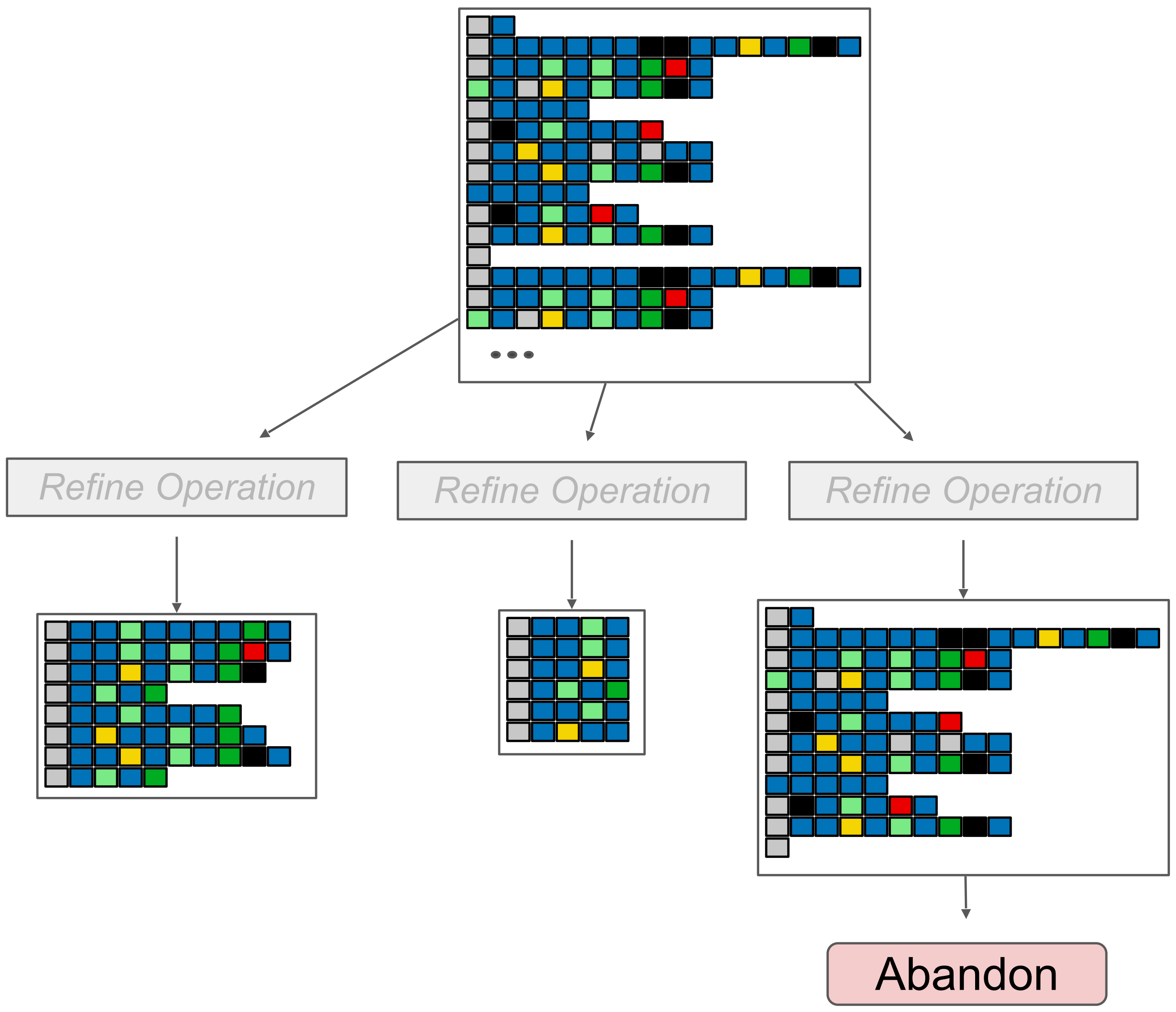


Conclude

- Discover actionable insight by *viewing* segment

High-Level Segmentifier Analysis Model

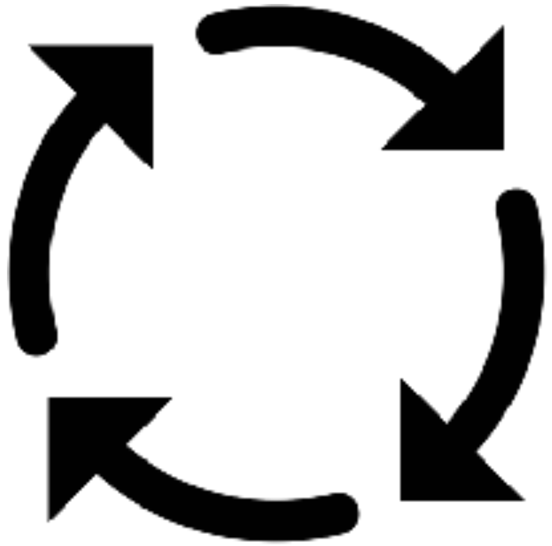




Abandon

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

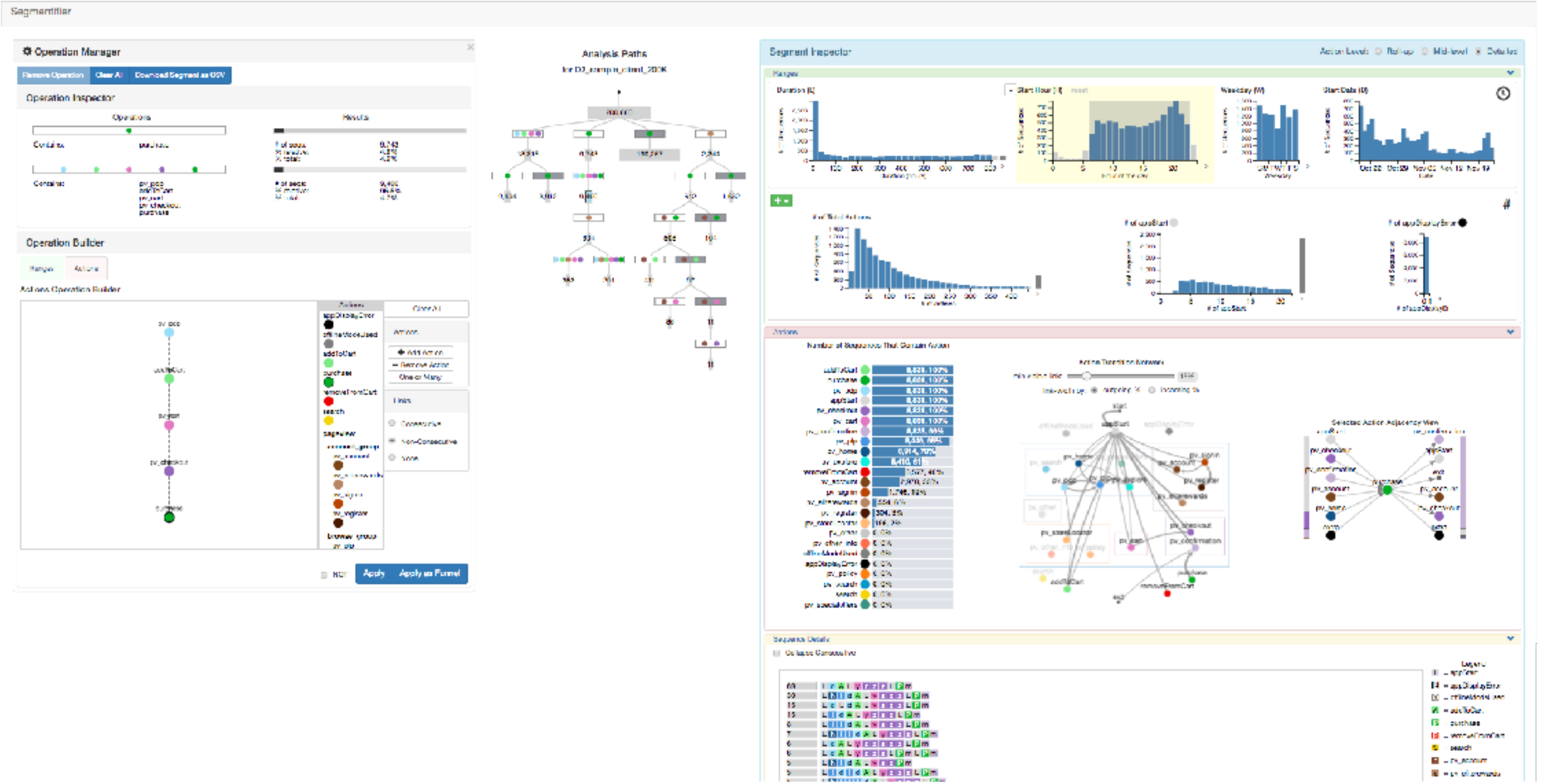
Why Visual Analytics?



- 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

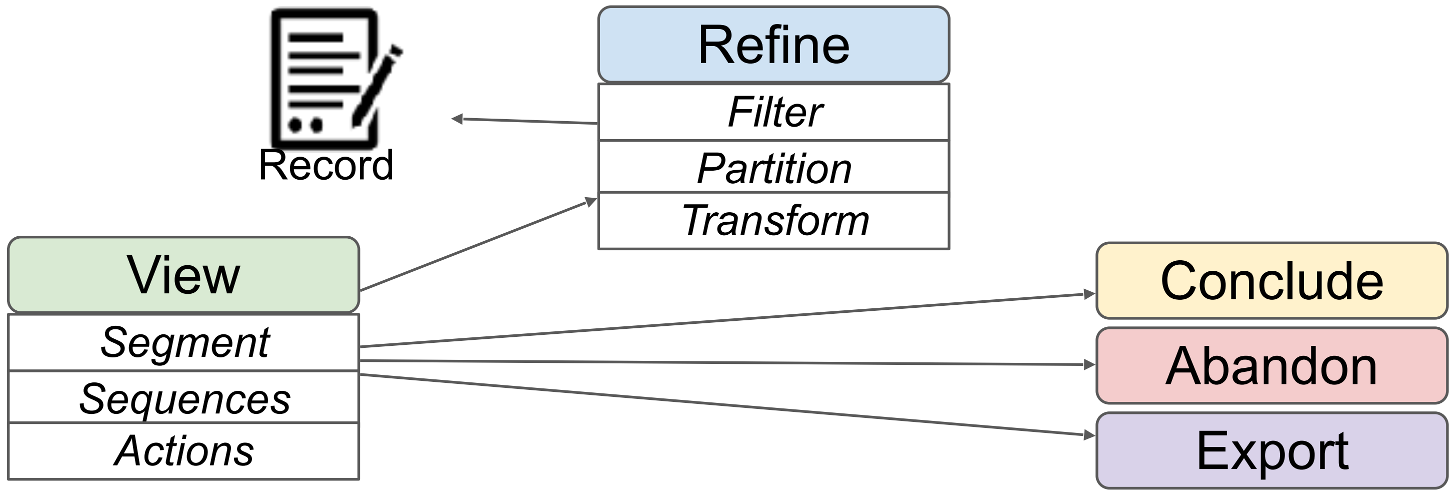


Video

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

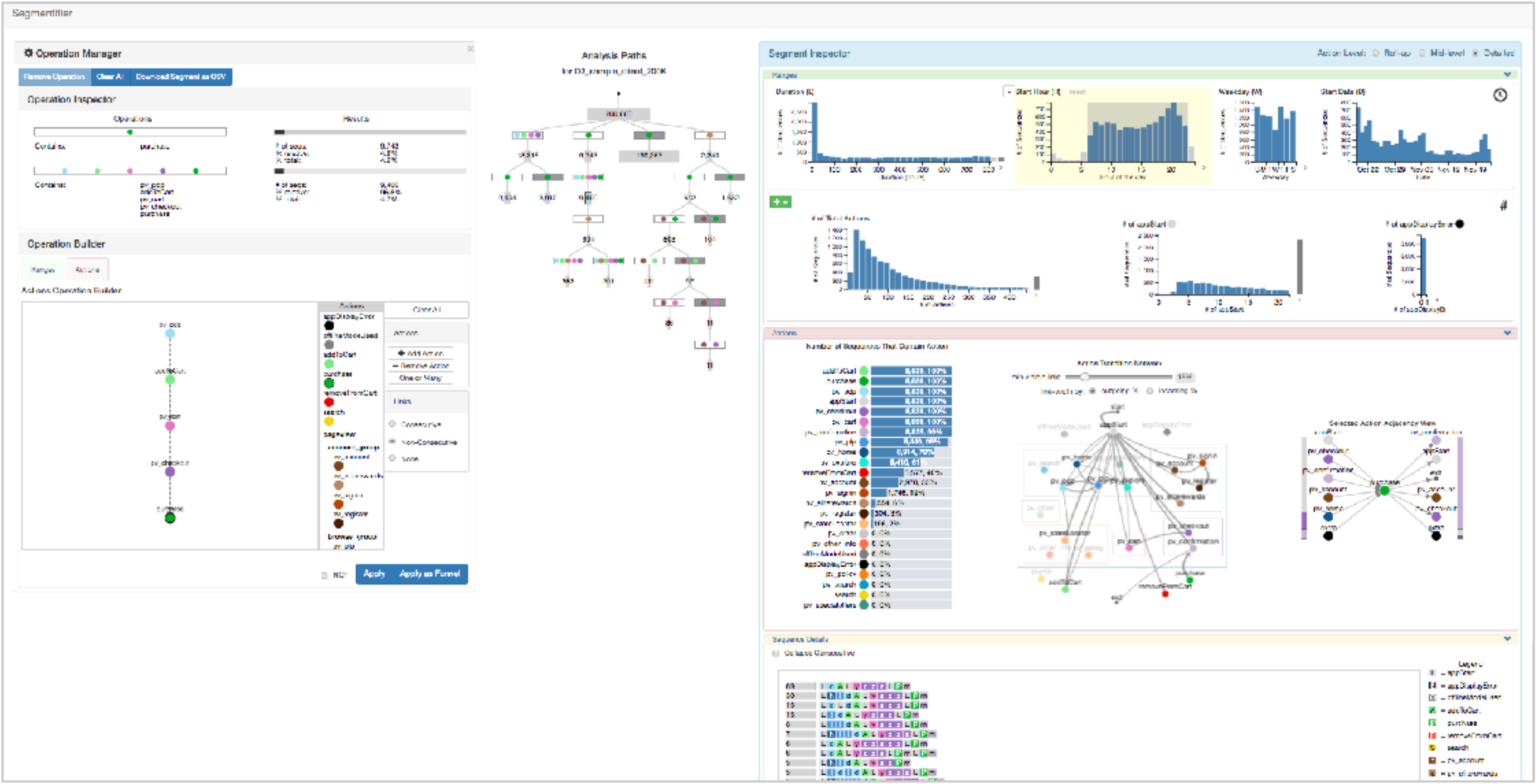
Segmentifier Contributions

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



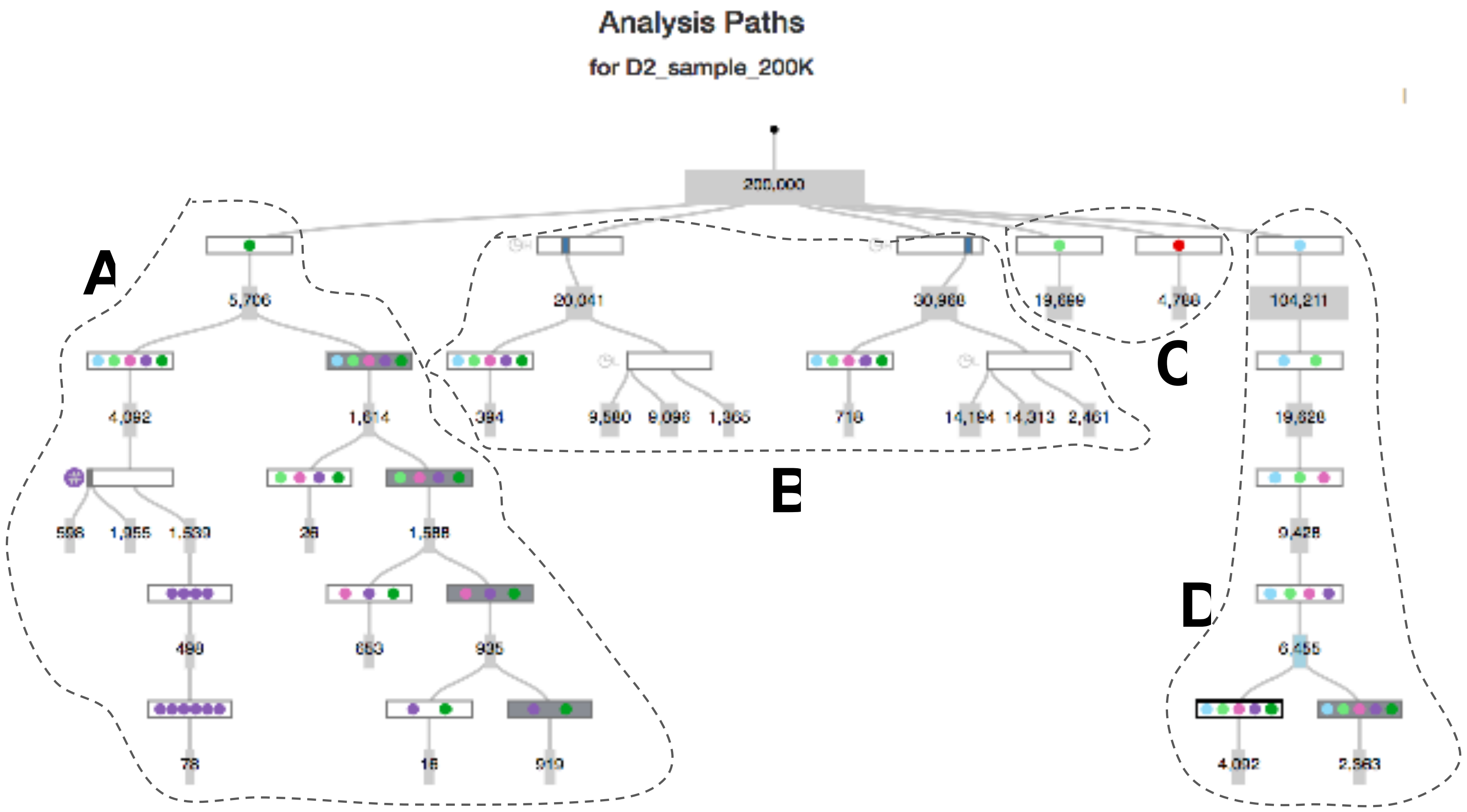
Segmentifier Contributions

- ➢ 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.

Video

https://www.youtube.com/watch?v=KcwjVK8eUdw

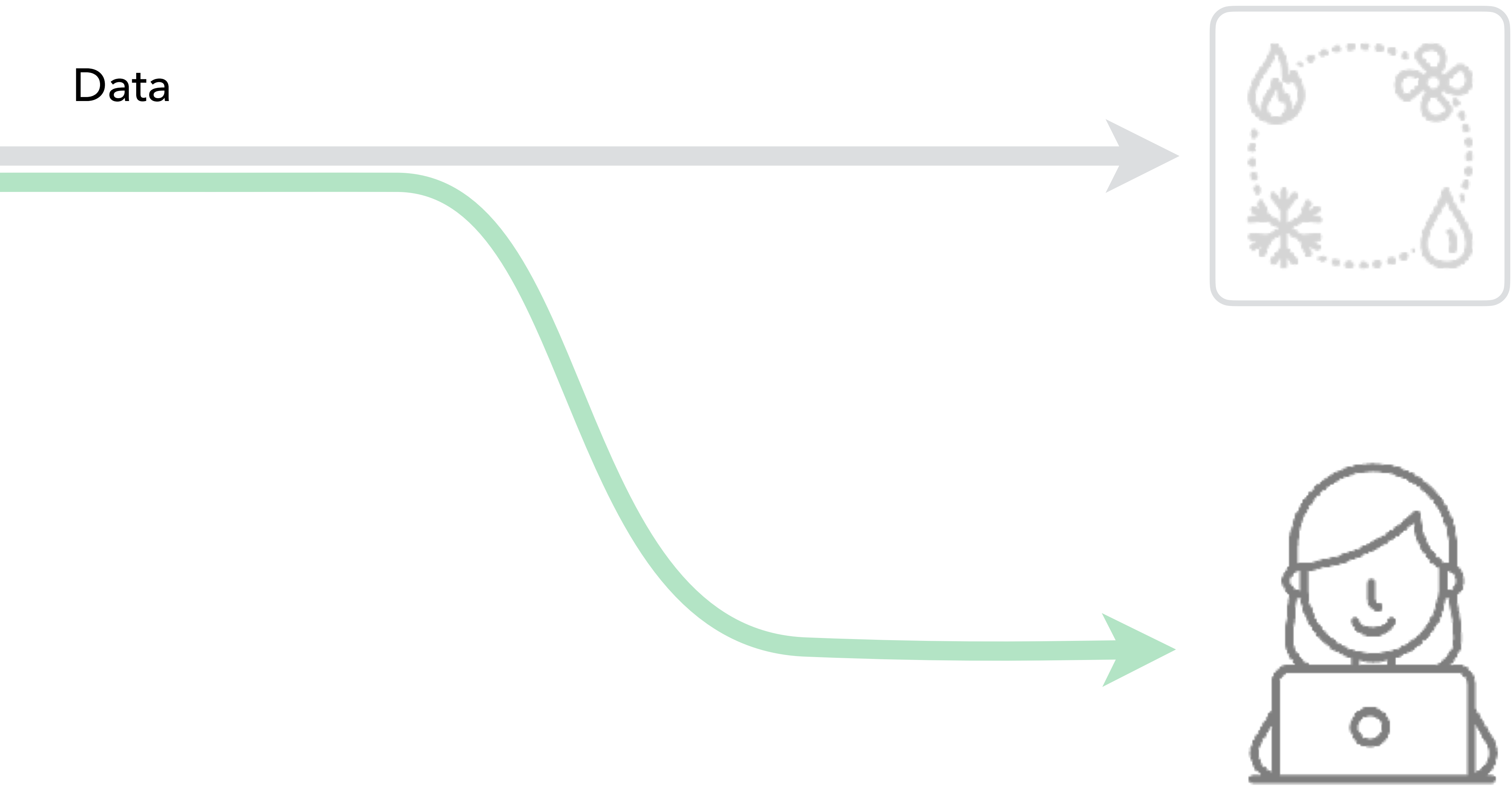
Location-Based Counts

- 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

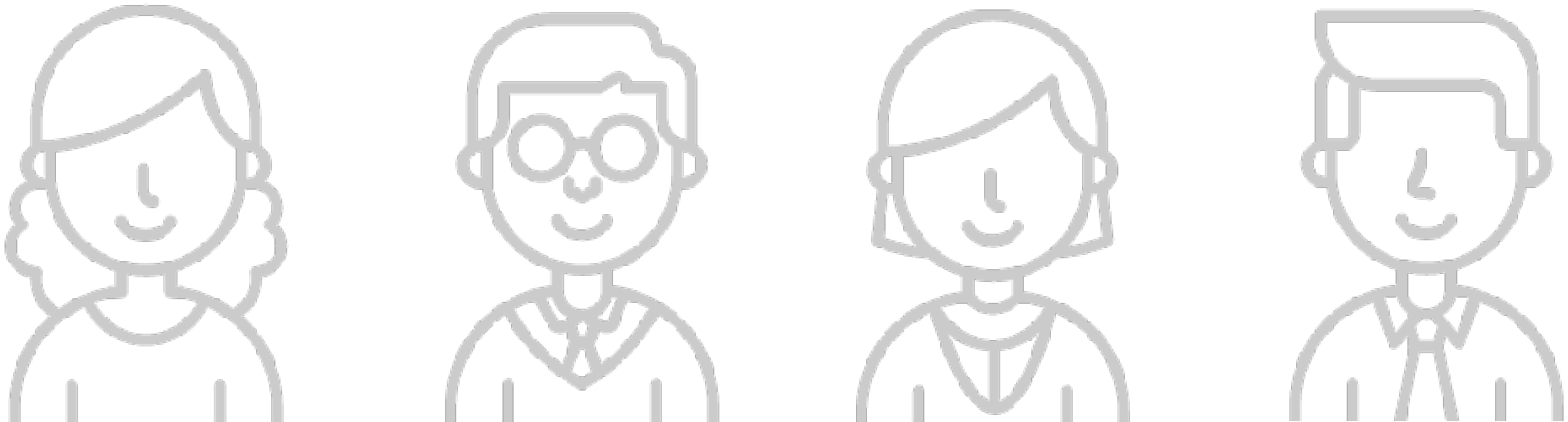


Decision making

WiFi connections as a proxy for occupancy

WiFi connections as a proxy for occupancy

Interviews with potential stakeholders



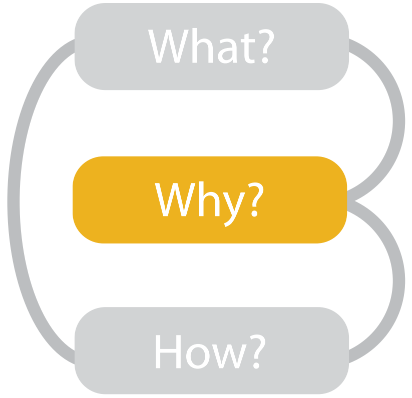
Focus Domains

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

Focus Domains

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

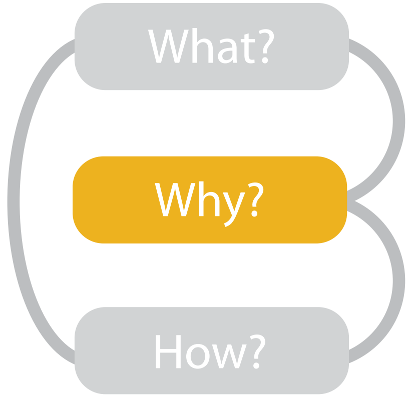
Semi-structured discussions and live demos



Tasks

Confirm assumptions or previous observations.

Do students occupy room x in evenings or on weekends?



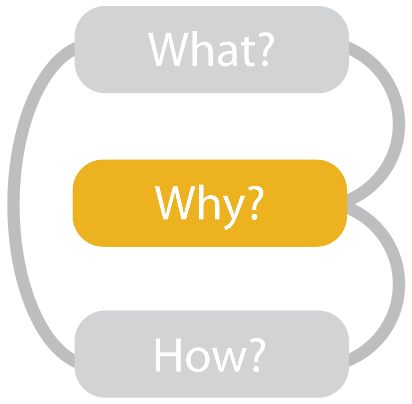
Tasks

- **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

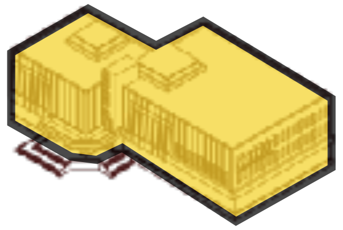
Spatial and Temporal Data Granularities

Regions of interest

Zone

Floor

Building



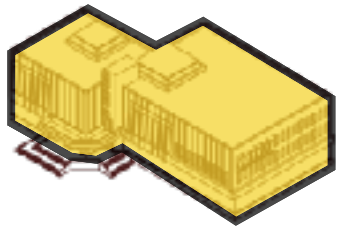
Spatial and Temporal Data Granularities

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

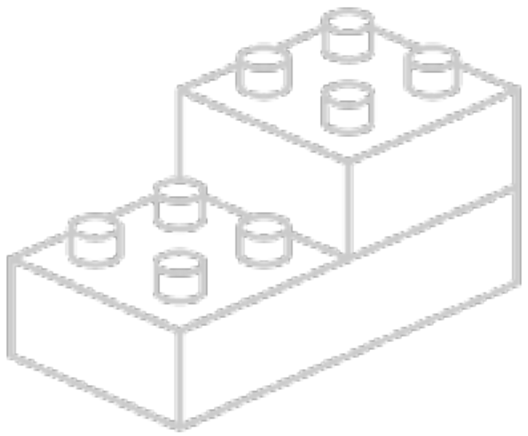
Live-data stream,
cross-building analysis

Building Recent

Building Long-term

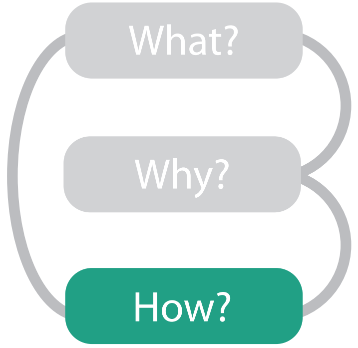
Region Compare

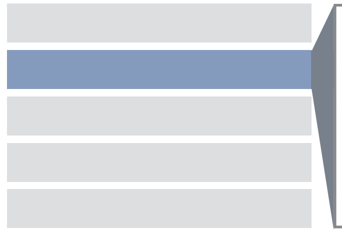
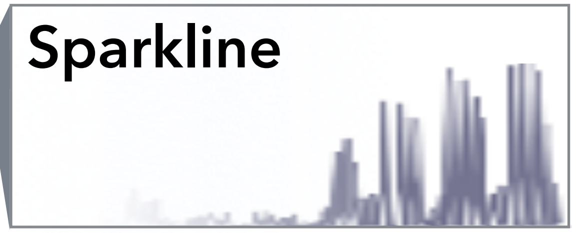
Time



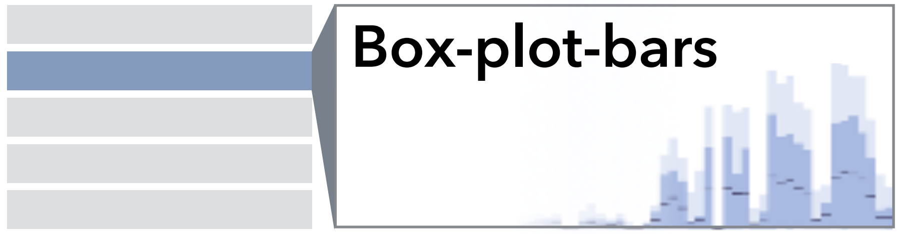
Reusable Visualization Components

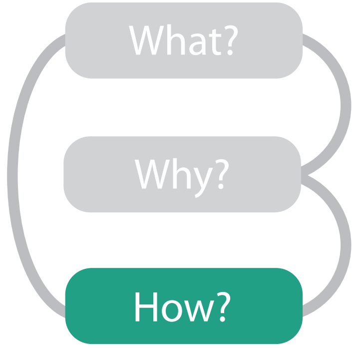
Reusable Visualization Components



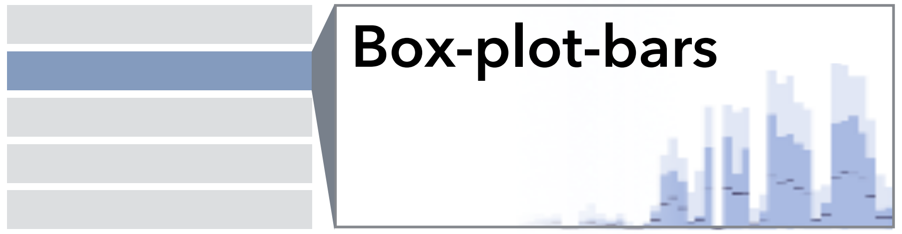
Layout	Visual Encoding	Facet	Comparisons
	**Sparkline** 	Juxtaposition	Repeating patterns, trends, outliers (contiguous)

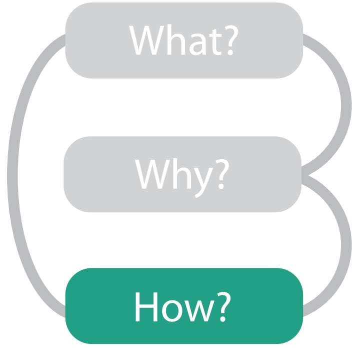
Reusable Visualization Components

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

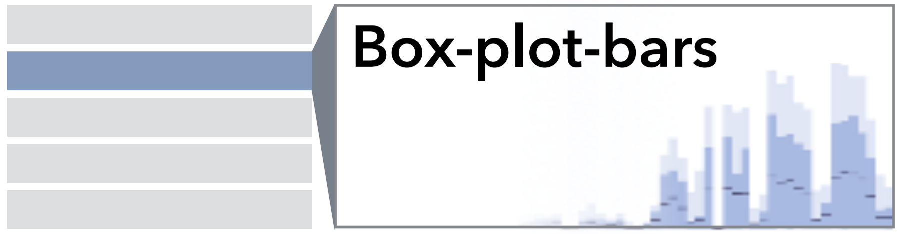
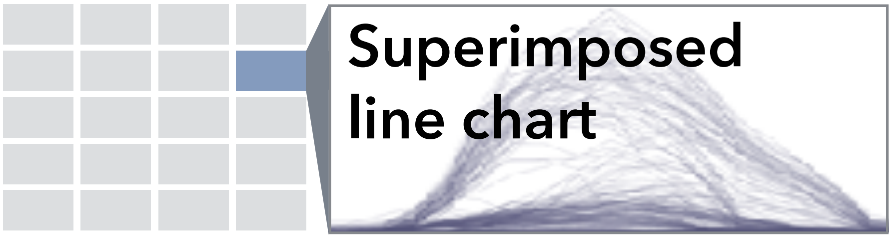


Reusable Visualization Components

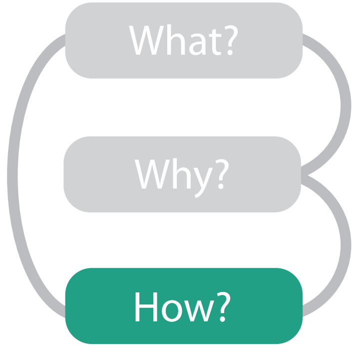
Layout	Visual Encoding	Facet	Comparisons
	**Sparkline**	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
	**Box-plot-bars**	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
	**Confidence band line chart**	Aggregation	Typical utilization profiles

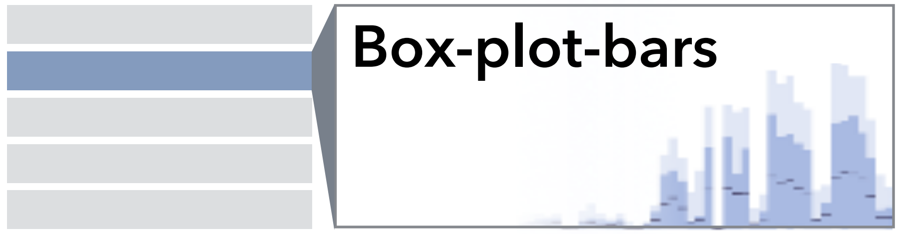
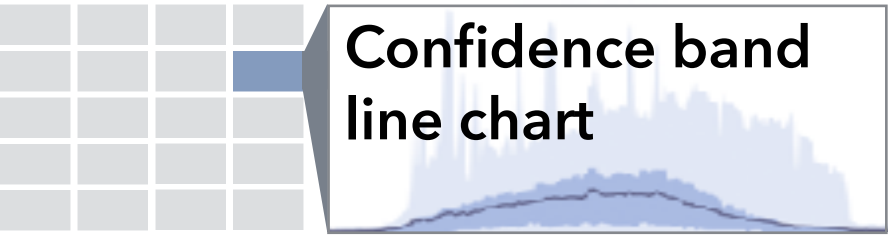
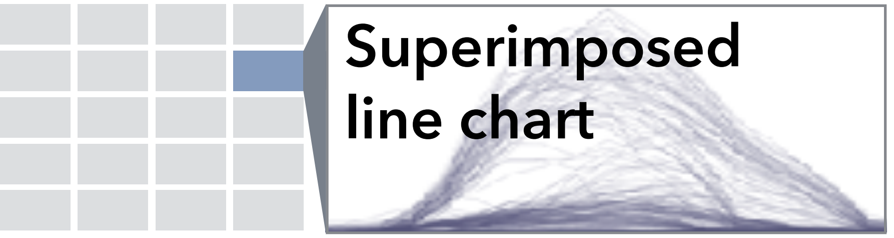


Reusable Visualization Components

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

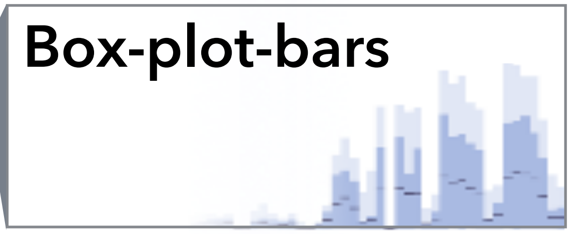
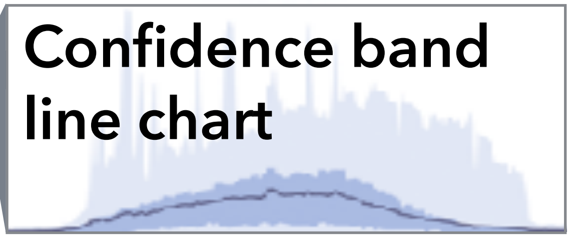
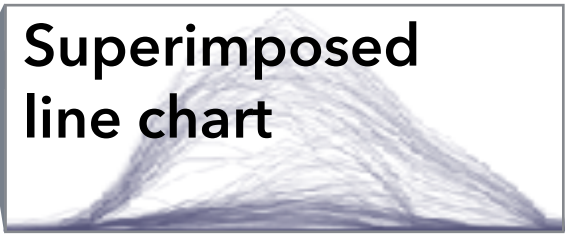
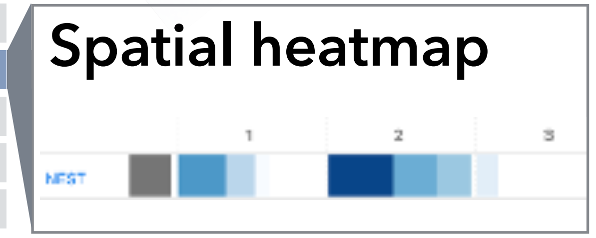
Reusable Visualization Components



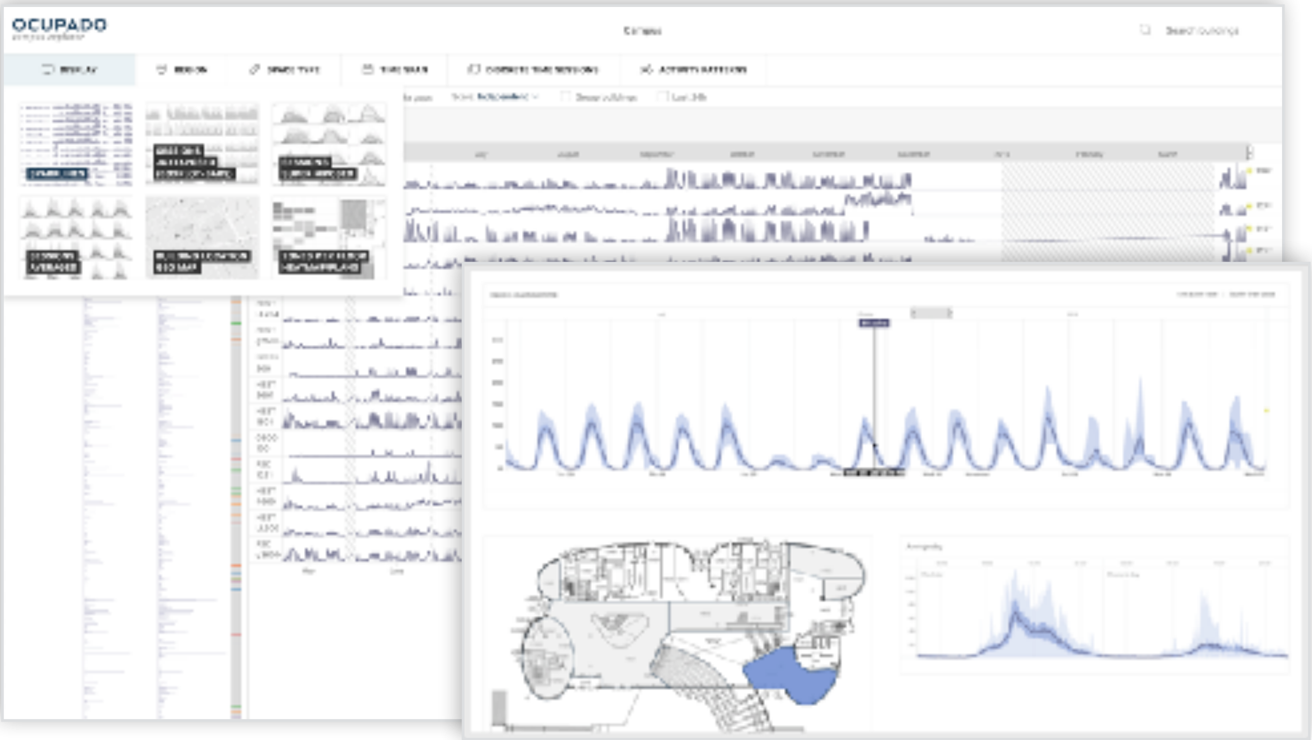
	Layout	Visual Encoding	Facet	Comparisons
Temporal		Sparkline	Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Box-plot-bars	Juxtaposition	Repeating patterns, trends, outliers *(non-contiguous)*
		Confidence band line chart	Aggregation	Typical utilization profiles
		Superimposed line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

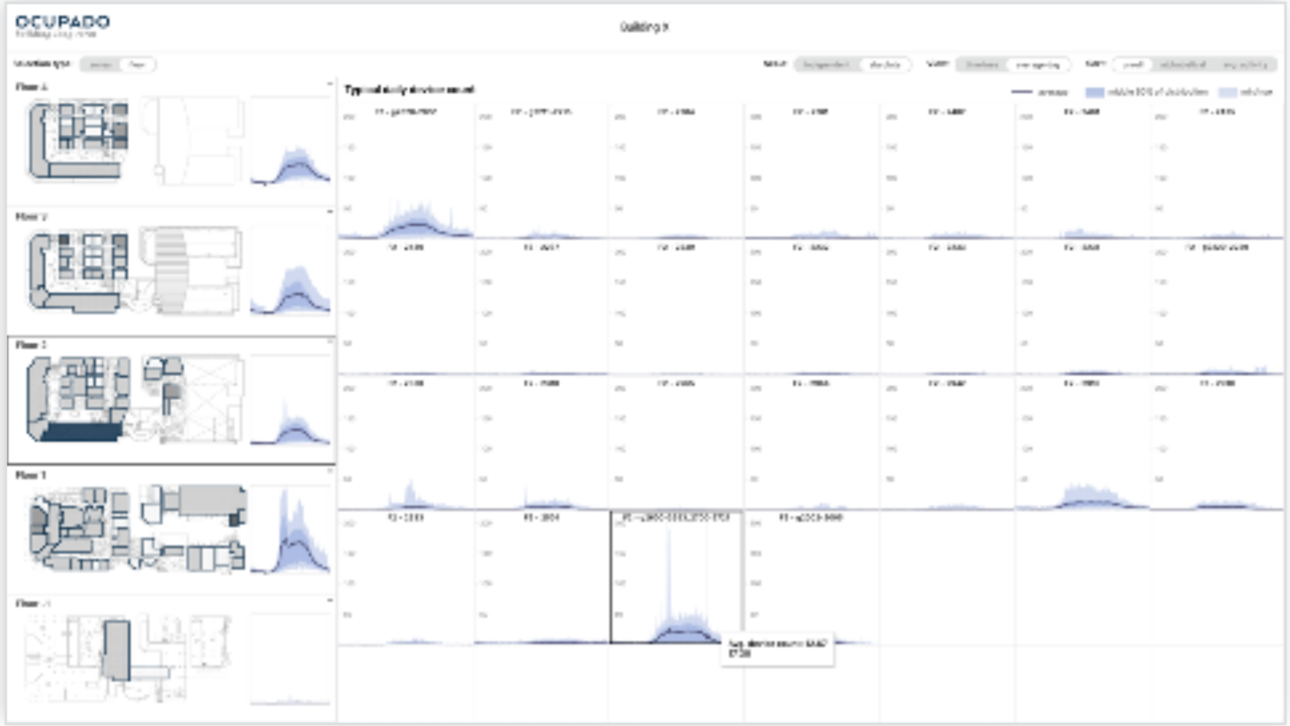
What?
Why?
How?

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



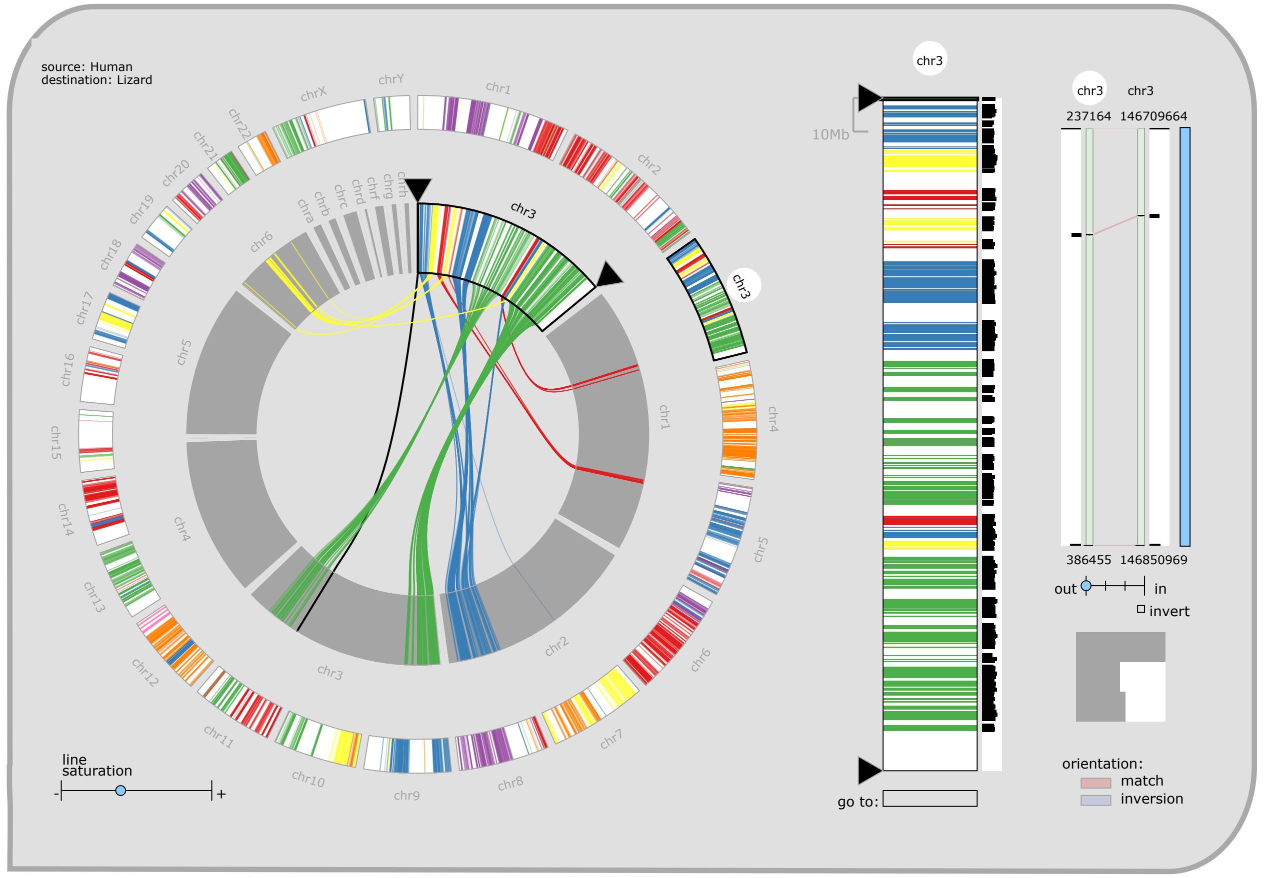


Ocupado Contributions

- 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

MizBee

A Multiscale Synteny Browser



joint work with:

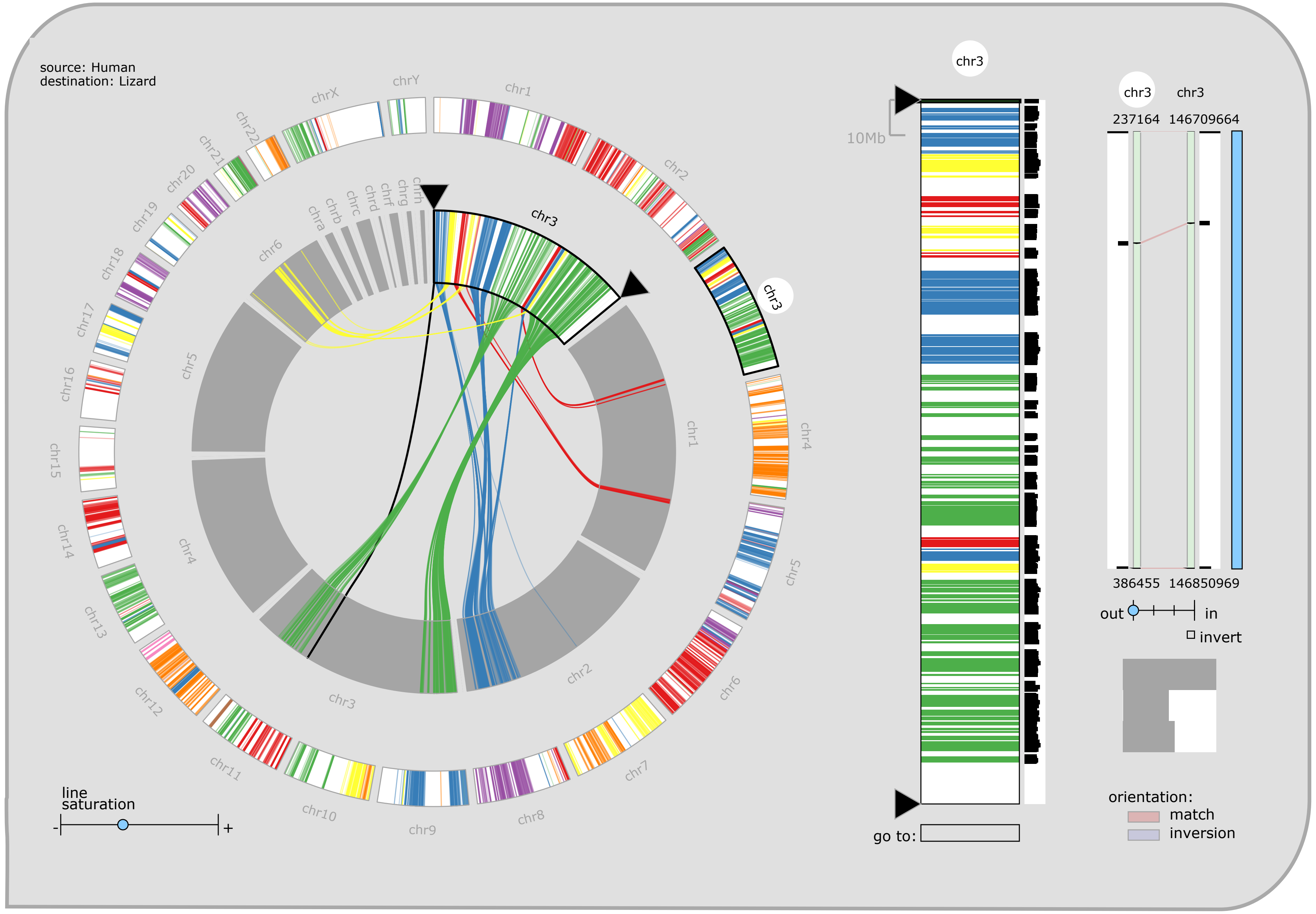
Miriah Meyer, Hanspeter Pfister

http://www.cs.utah.edu/~miriah/mizbee

MizBee: A Multiscale Synteny Browser.

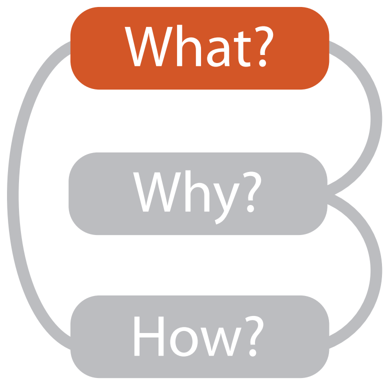
Meyer, Munzner, Pfister. IEEE Trans. Visualization and Computer Graphics 15(6):897-904, 2009 (Proc. InfoVis 2009).

Video

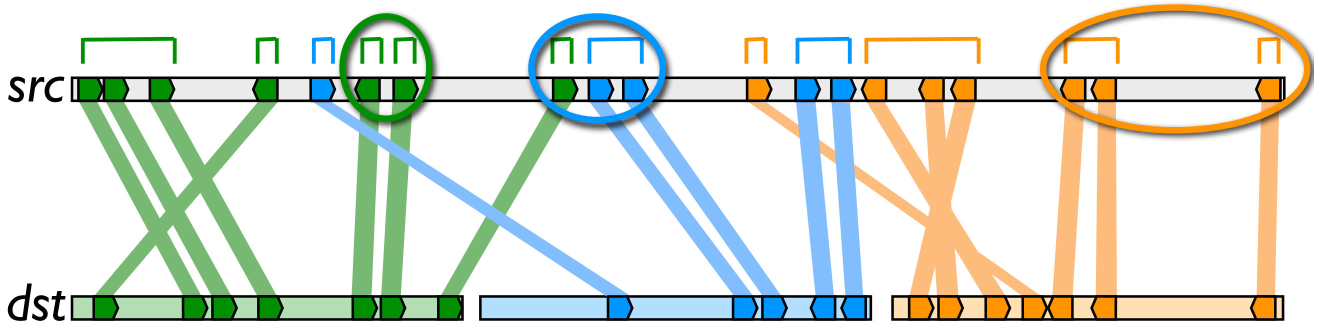


https://www.youtube.com/watch?v=86p7brwuz2g

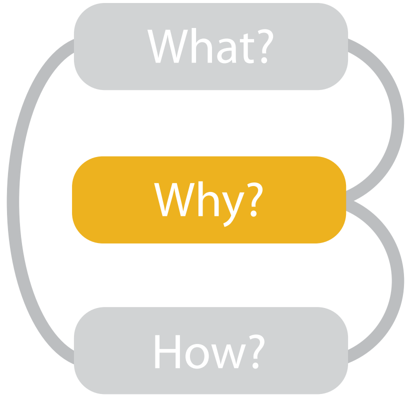
What: Data abstraction



- data: multiscale lists
  - features: hundreds of thousands
    - ordered attribute: position in chromosome sequence coordinates
    - categorical attributes: orientation, chromosome of matching feature
    - quantitative attributes: length, similarity score
  - syntenic blocks: thousands
    - contiguous sets of features on same chromosome
    - combine thresholded features if
      - destination chromosome and orientation match
      - close together
  - chromosomes: dozens
  - genomes: two

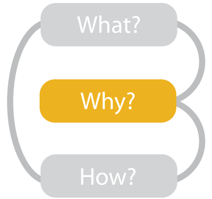


Why: Tasks in domain language



- analyze conservation (similarity) relationships between genomic features
  - high-level biology questions
    - evolution
      - how long ago did two species share common ancestor?
    - function
      - which segment of the genome is responsible for specific function in the cell?
    - ...
  - low-level data-centric questions
    - algorithm refinement
      - are paired features within a block contiguous?
      - which chromosomes share conserved blocks?
      - are similarity scores alike within block?
      - ...

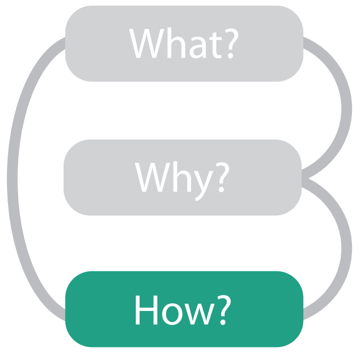
Why: Tasks abstraction



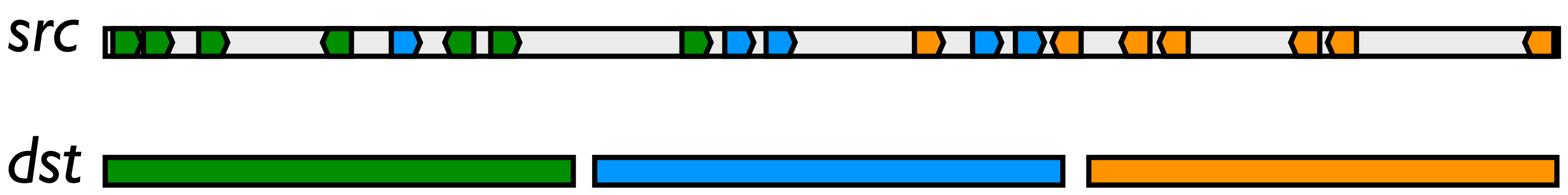
- relationship types: proximity, size, orientation, similarity
- data scales: genome, chromosome, block, feature
- topics: algorithm in/out, block reliability, high-level science

	relationship				scale			
	genome	chromosome	block	feature	proximity / location	size	orientation	similarity
Which chromosomes share conserved blocks?	x				x			
For one chromosome, how many other chromosomes does it share blocks with?	x	x			x			
What is the density of coverage and where are the gaps on: chromosomes? blocks?	x	x	x		x			
Where are the blocks: on chromosomes? around a specific location on a chromosome?	x	x			x			
What are the sizes and locations of other genomic features near a block?		x			x	x		
How large are the blocks?		x				x		
Do neighboring blocks go to the same: chromosomes? relative location on a chromosome?	x	x			x			
Are the orientations matched or inverted for: block pairs? feature pairs?		x	x				x	
Do the orientations match for pairs of: neighboring blocks? features within a block?		x	x				x	
Are similarity scores alike: with respect to neighboring blocks? within a block?		x	x					x
Are the paired features within a block contiguous?			x		x			
How large is a feature relative to other genes within a block?			x			x		
What are the sizes, locations, and names of features within a block?			x		x	x		
What are the differences between individual nucleotides of feature pairs?				x				x

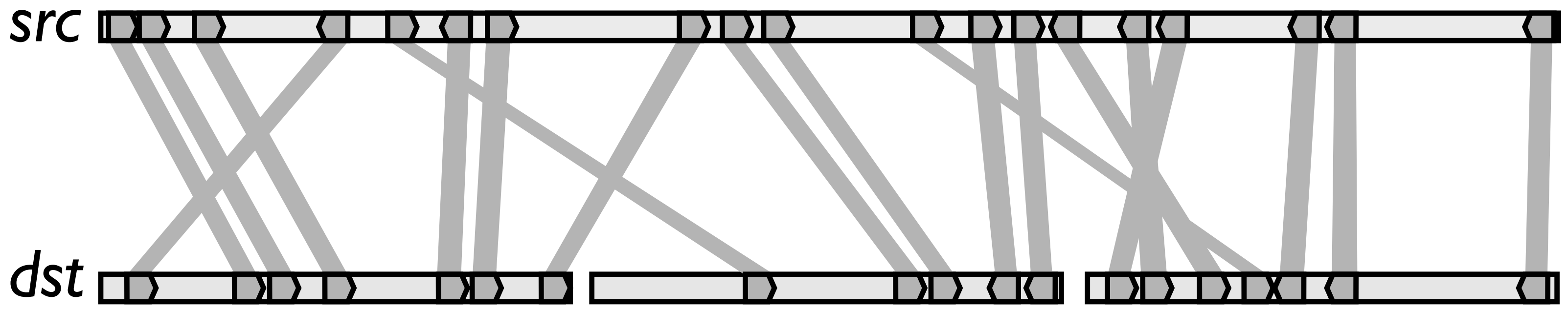
How: Idiom design choices

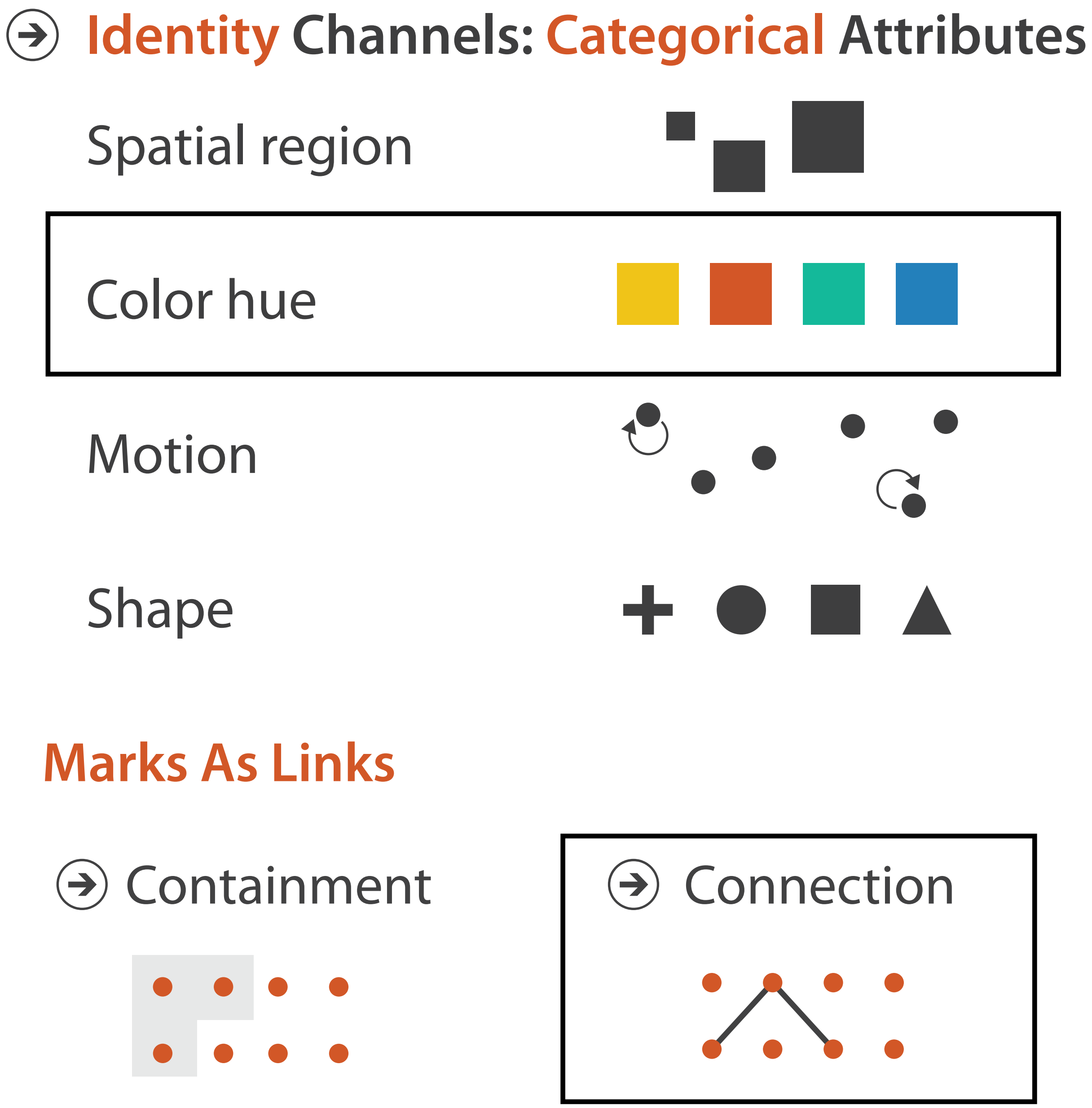


- encode match relationships between chromosome segments with both
  - color



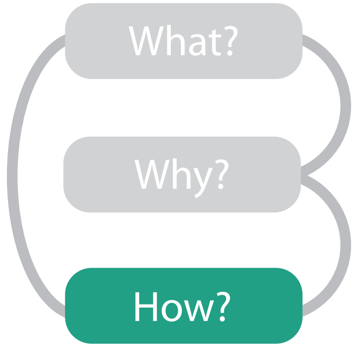
  - connection marks

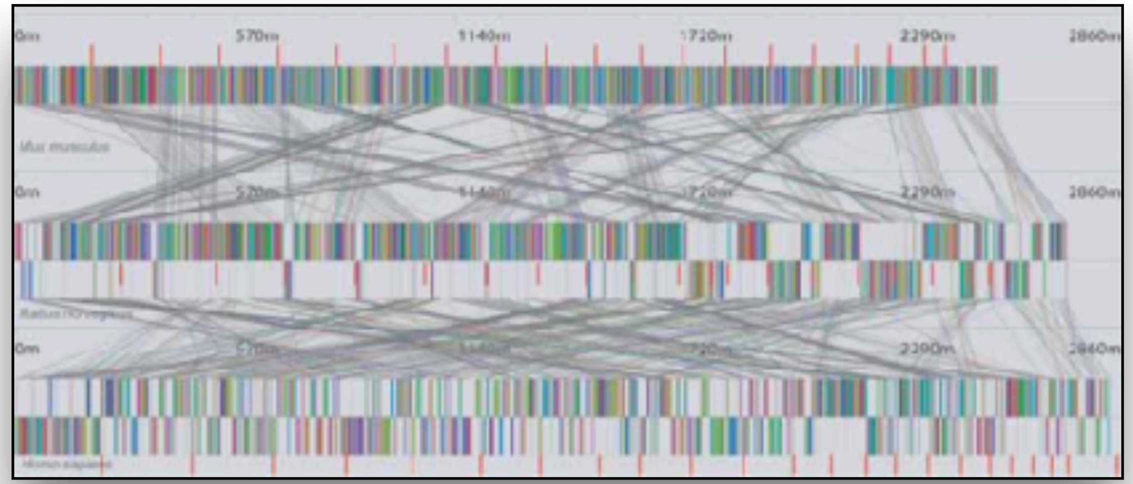




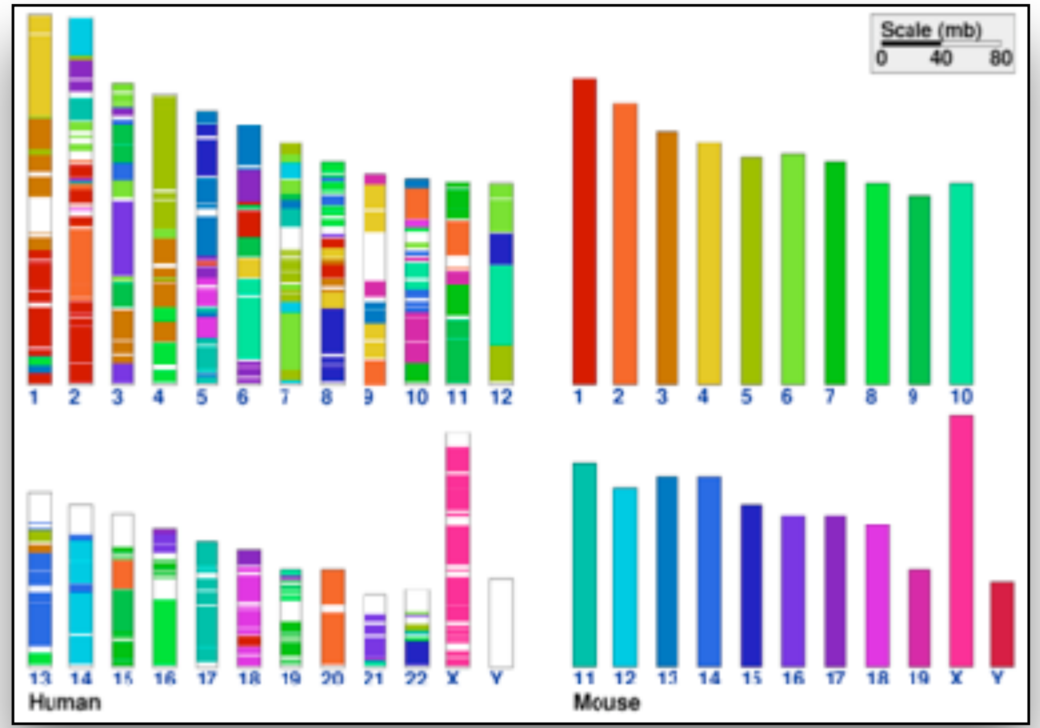
How: Arrange space

- design space of arrangements

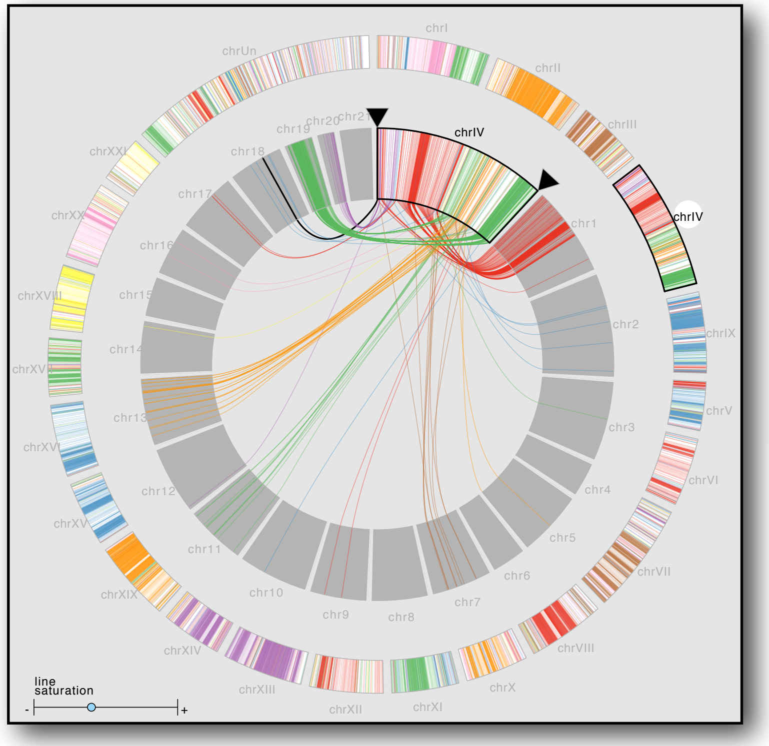




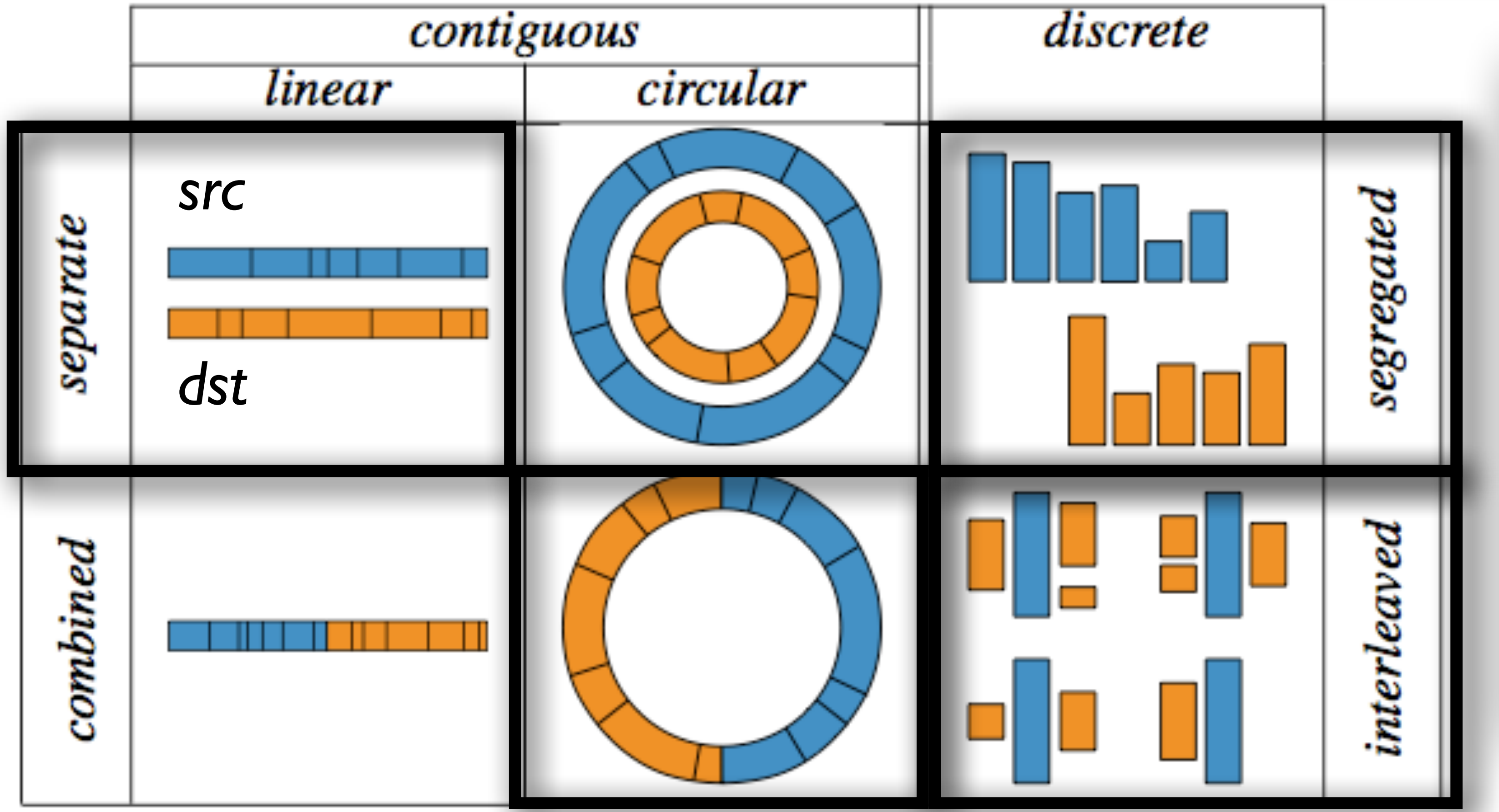
Mauve [Darling04]

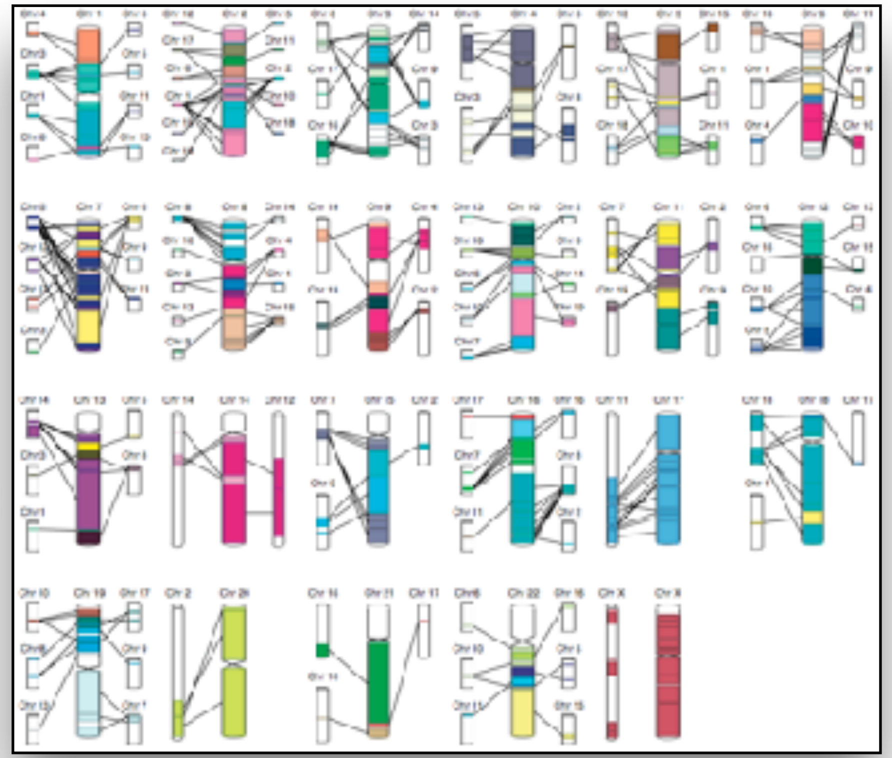


Cinteny



MizBee

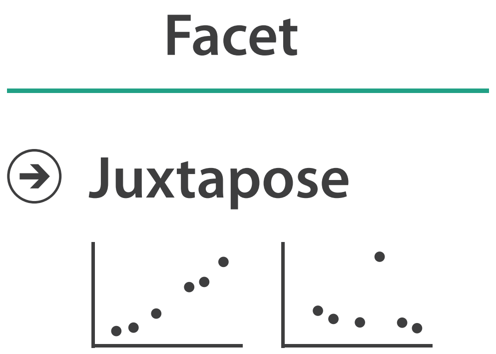


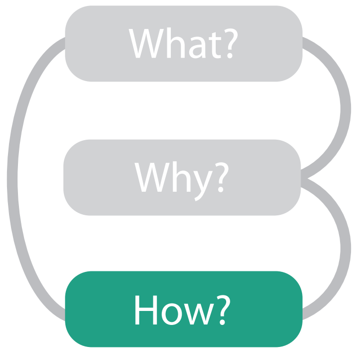


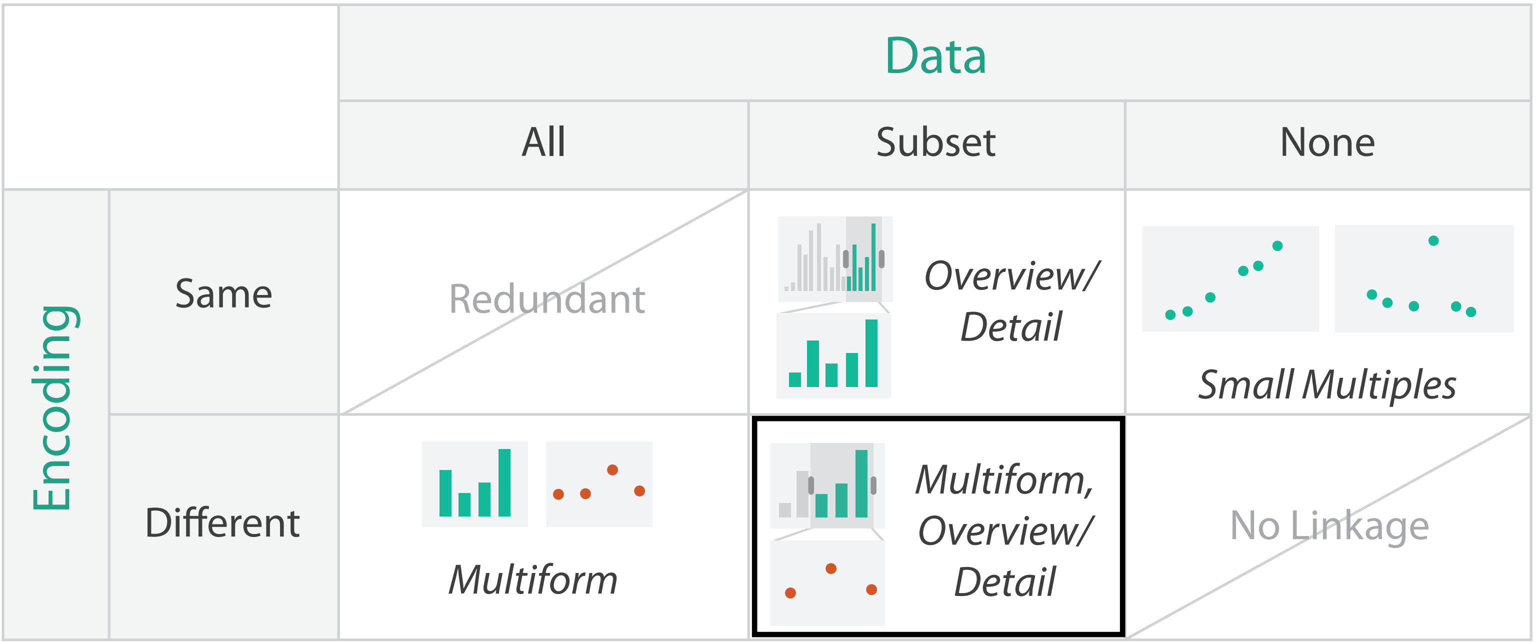
Apollo [Lewis02]

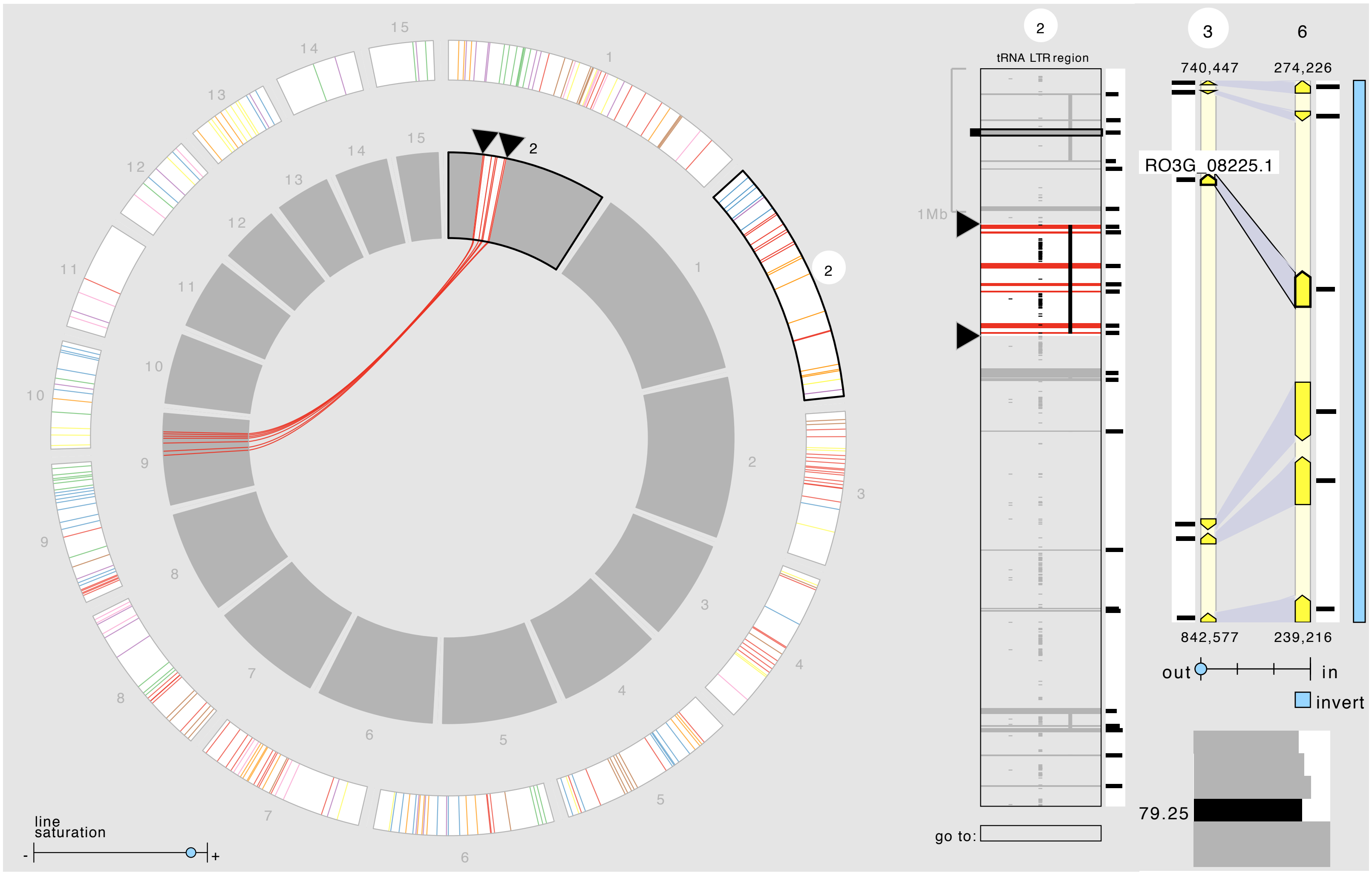
How: Idiom design choices

- juxtapose linked views
  - *multiform overview-detail*
    - three views: genome, chromosome, block
    - different visual encoding in each

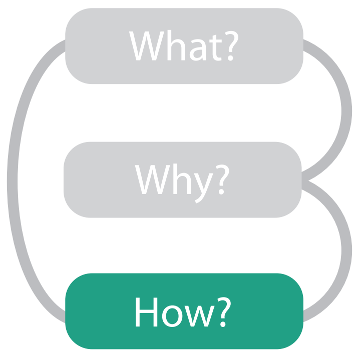




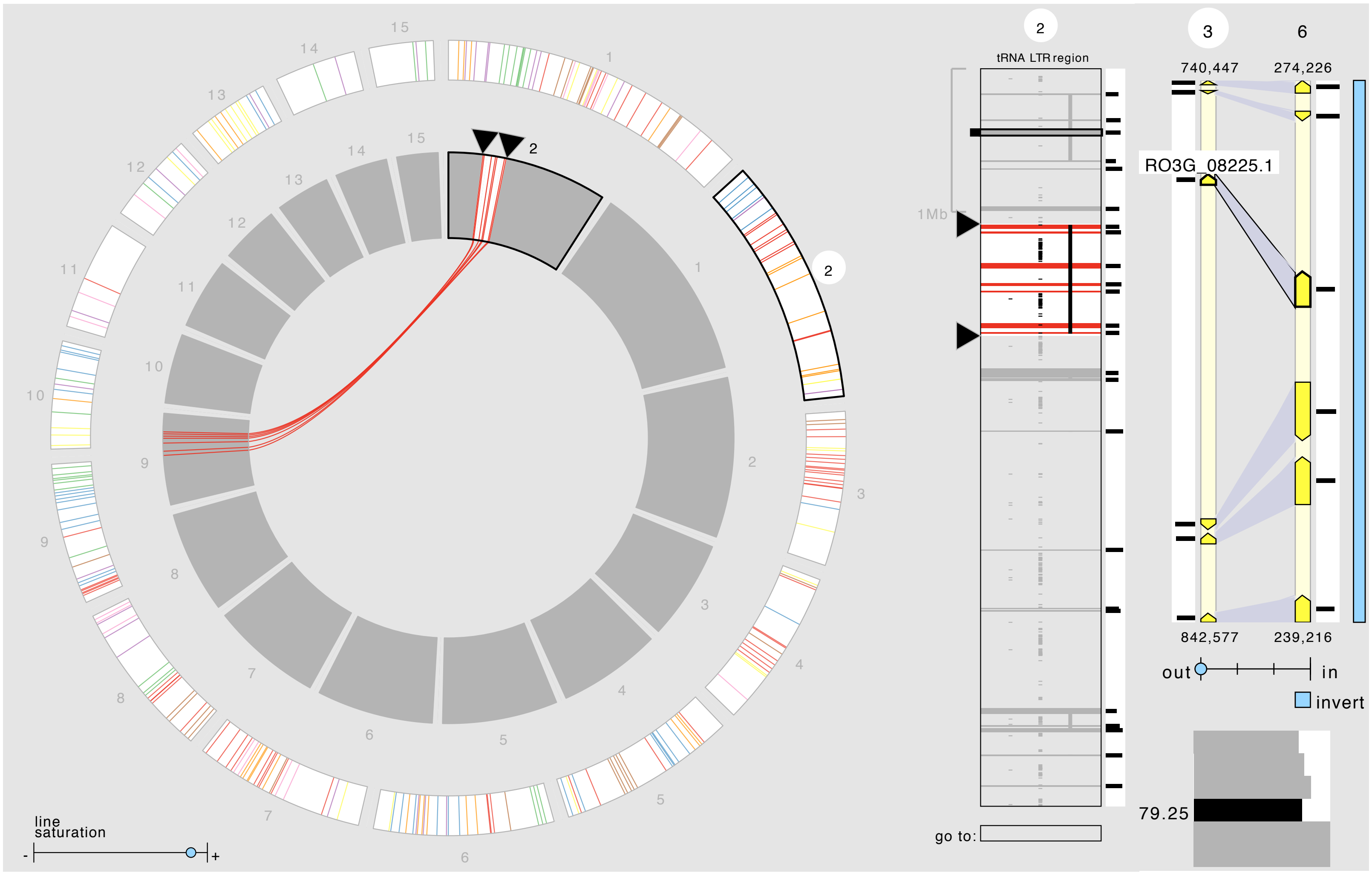


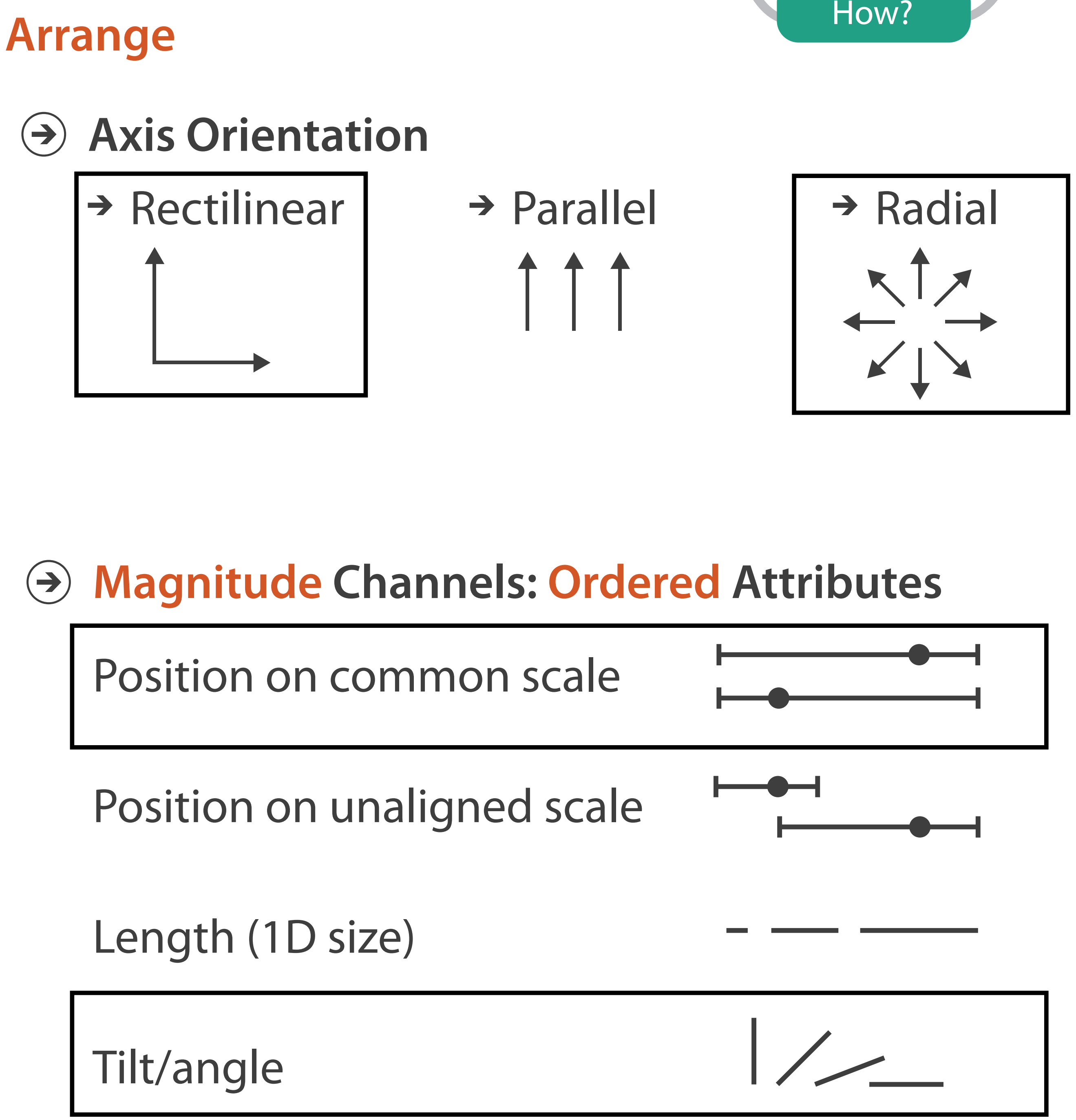


How: Idiom design choices



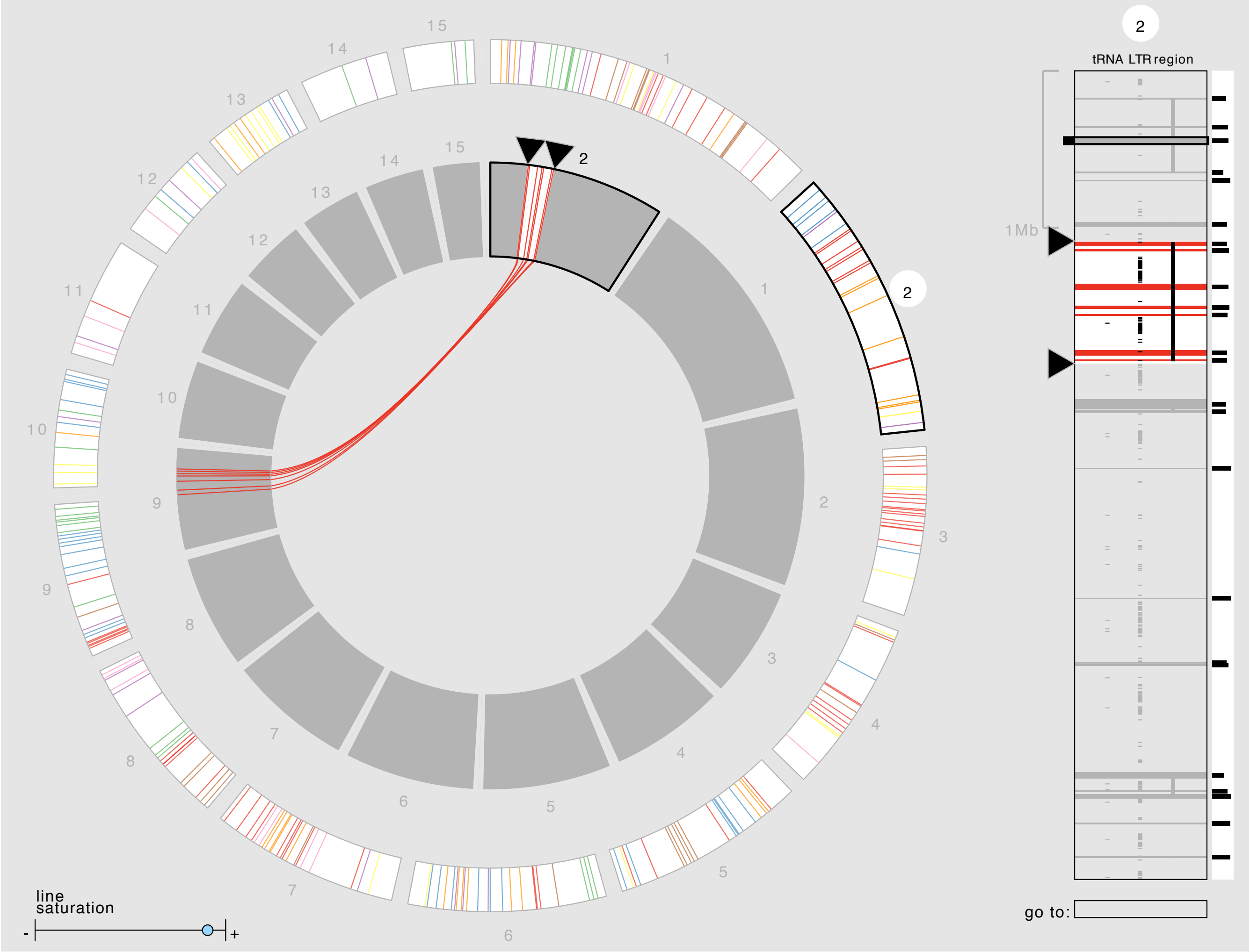
- axis orientation
  - radial: genome
  - rectilinear: chromosome, block
    - aligned position more accurate than angle

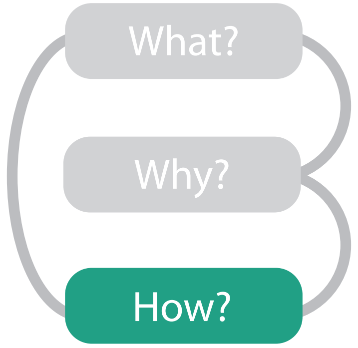




How: Idiom design choices

- filter

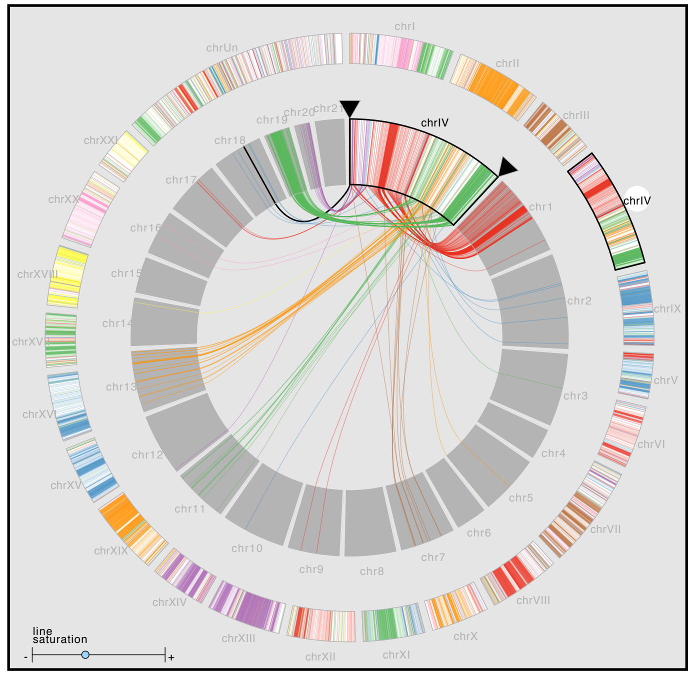


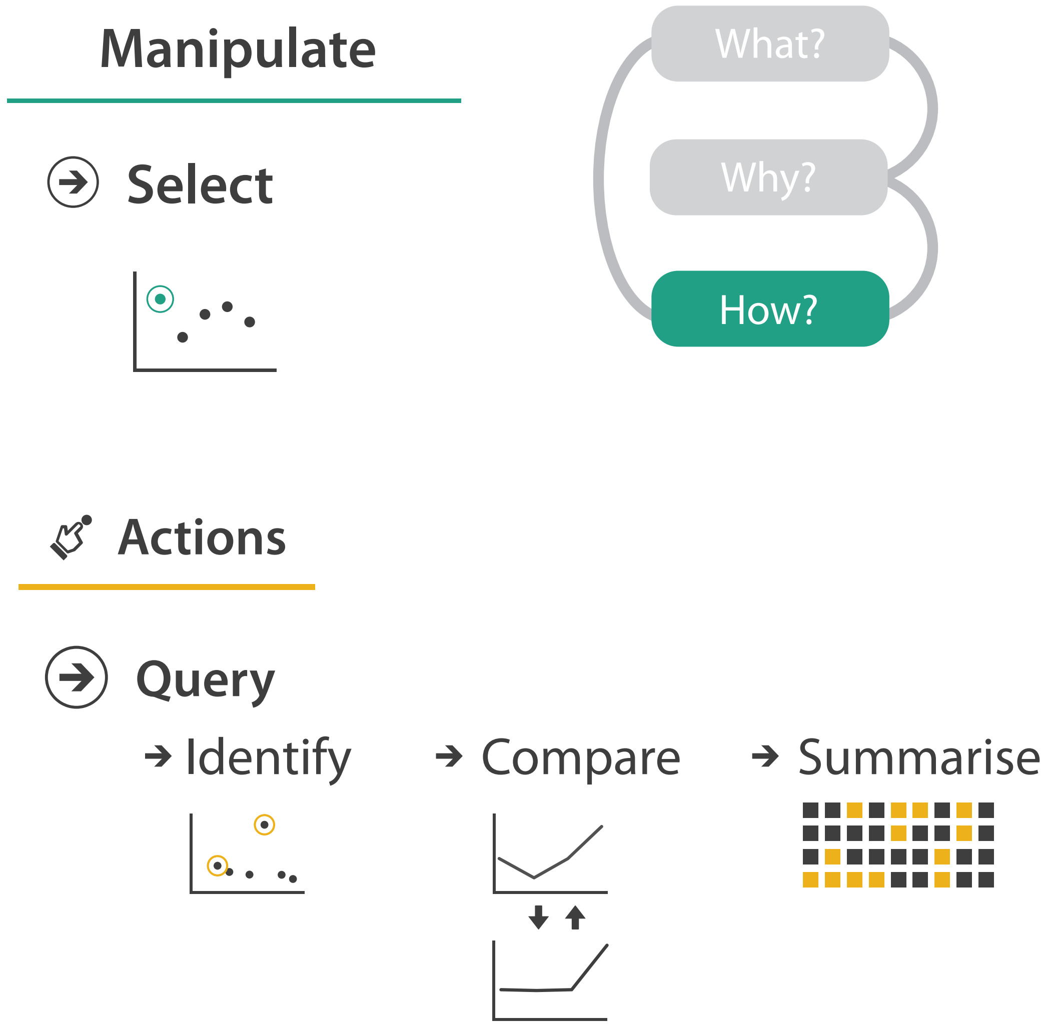


Reduce

How: Idiom design choices

- outer ring: summarize relationships with color
  - select one chromosome from set of source chromosomes
- inner ring:
  - destination chromosomes around copy of selected source chromosome
  - show relationship details with connection marks as well as color





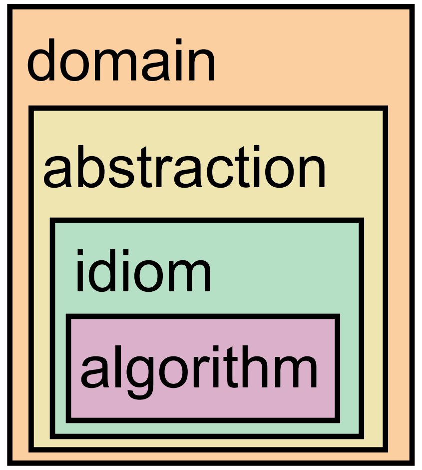
MizBee contributions

- first synteny browser with side-by-side linked views
  - across the range of scales
  - encoding all four conservation relationship types
    - proximity, size, orientation, similarity

- open source
  http://www.cs.utah.edu/~miriah/mizbee

Visualization: Abstractions & idioms

- levels of design
  - identify abstractions
    - crucial & difficult, iterative process
  - select appropriate idioms
    - or create new ones if necessary

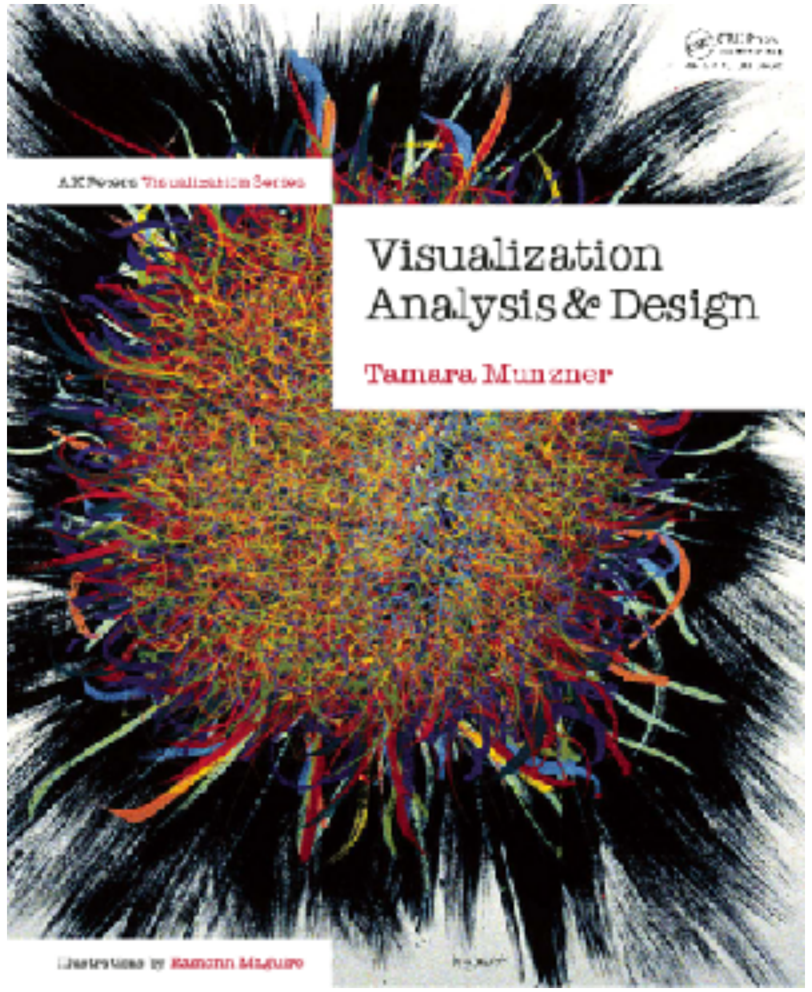


- three examples
  - different domains
  - different abstractions
  - different idioms



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#coimbra22