Design Study Methodology

Cmpt 767 - Visualization

Steven Bergner

sbergner@sfu.ca

[based on slides by Torsney-Weir / Möller / Munzner]

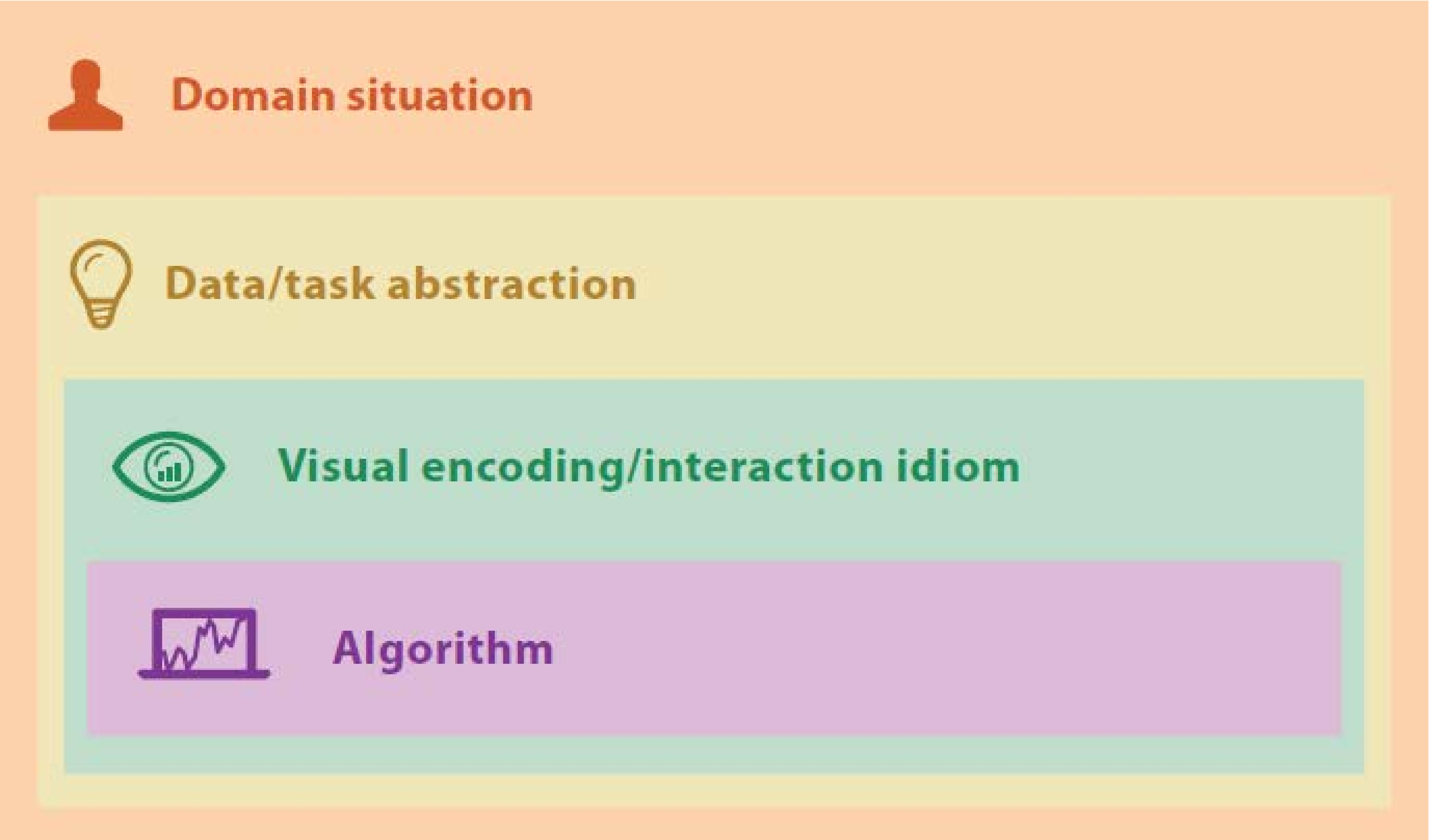
Acknowledgements

- Tamara's vis course slides

- Design study methodology: Reflections from the trenches and the stacks. Michael Sedlmair, Mariah Meyer, and Tamara Munzner. IEEE Trans. Visualization and Computer Graphics 18(12):2431-2440, 2012.

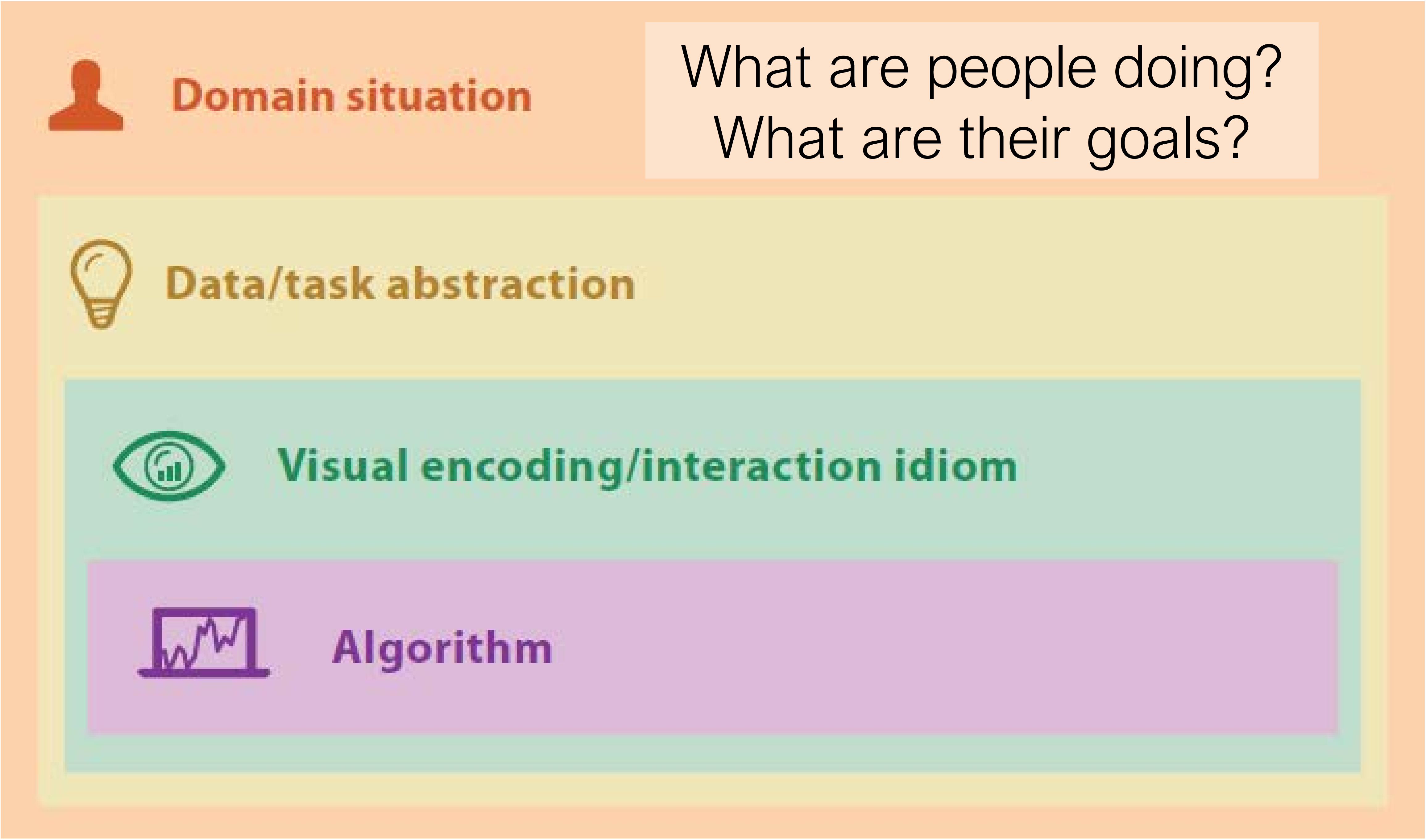
- Cluster and Calendar based Visualization of Time Series Data. Jarke J. van Wijk and Edward R. van Selow. Proc. InfoVis 1999, p 4-9.

Nested model for vis design: Four Levels for Validation

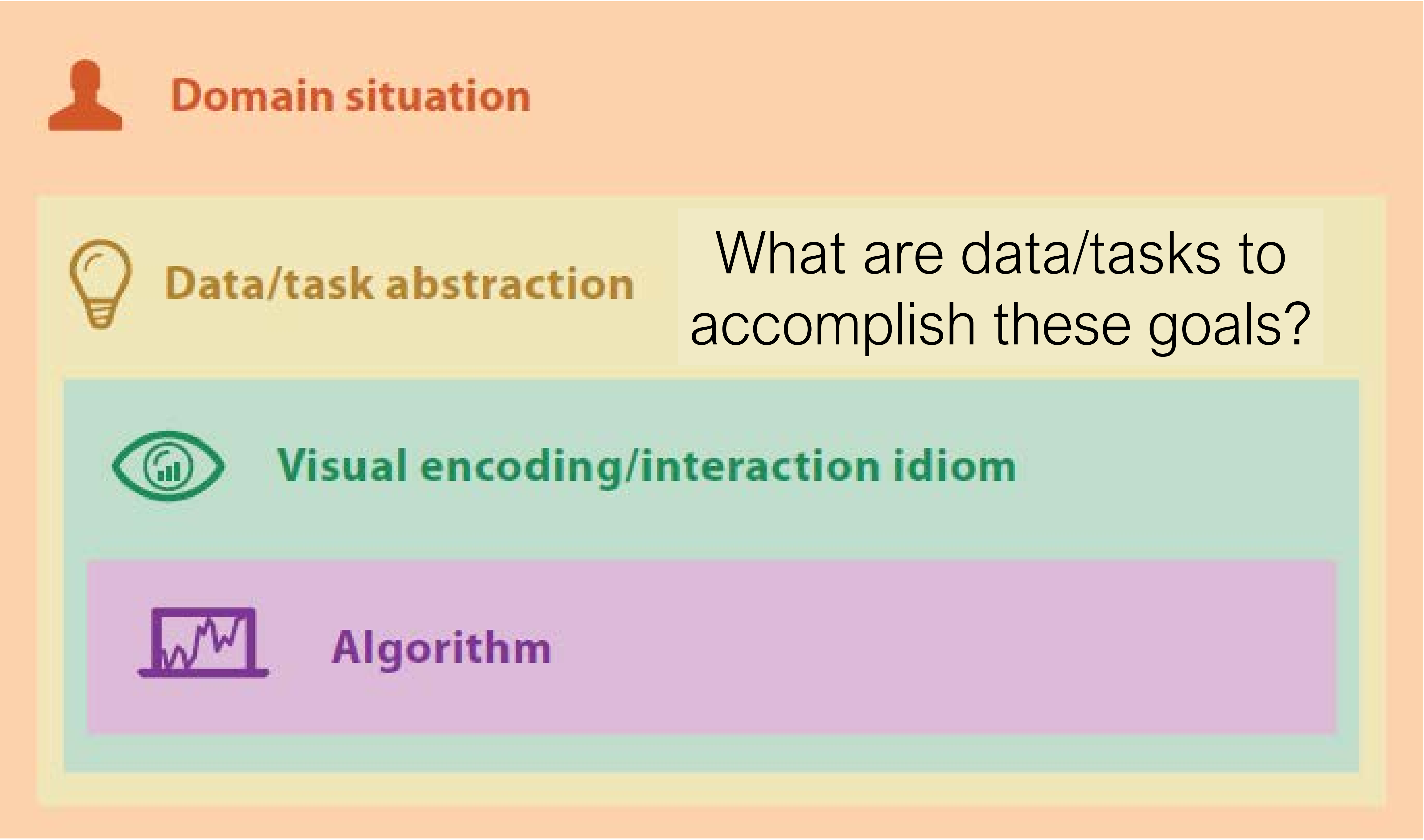


[Munzner]

Nested model for vis design

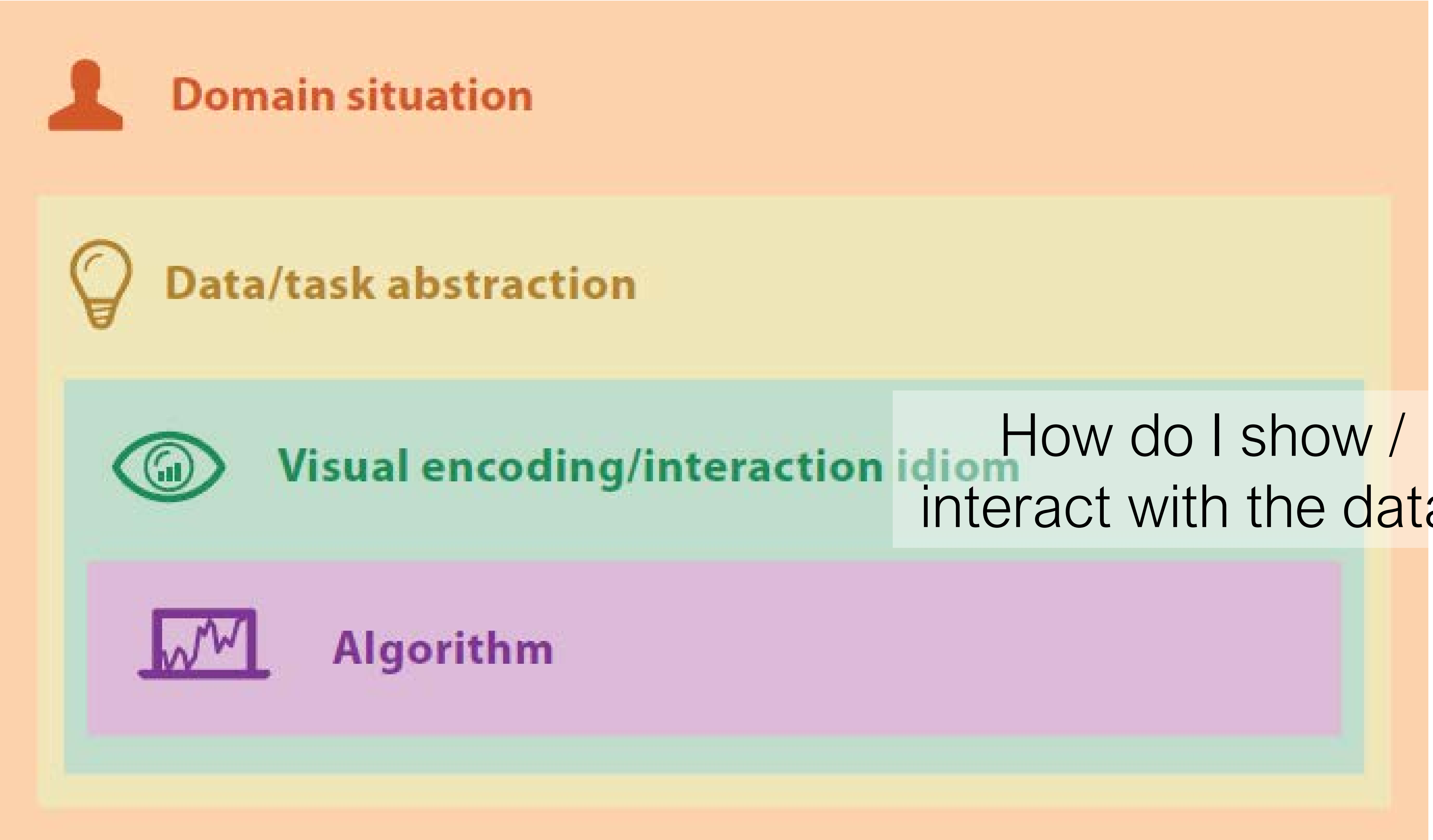


Nested model for vis design



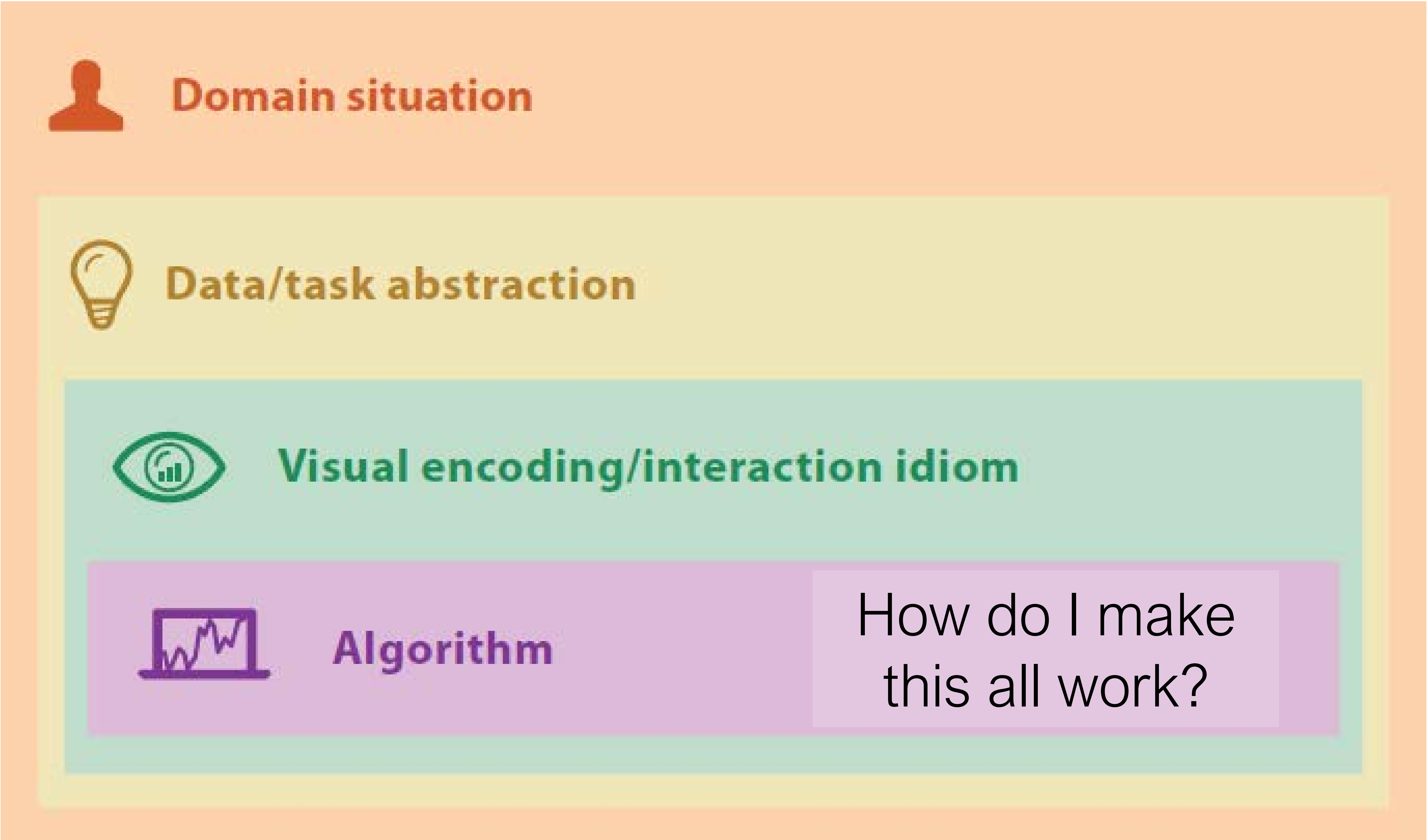
[Munzner]

Nested model for vis design



How do I show / interact with the data?

Nested model for vis design

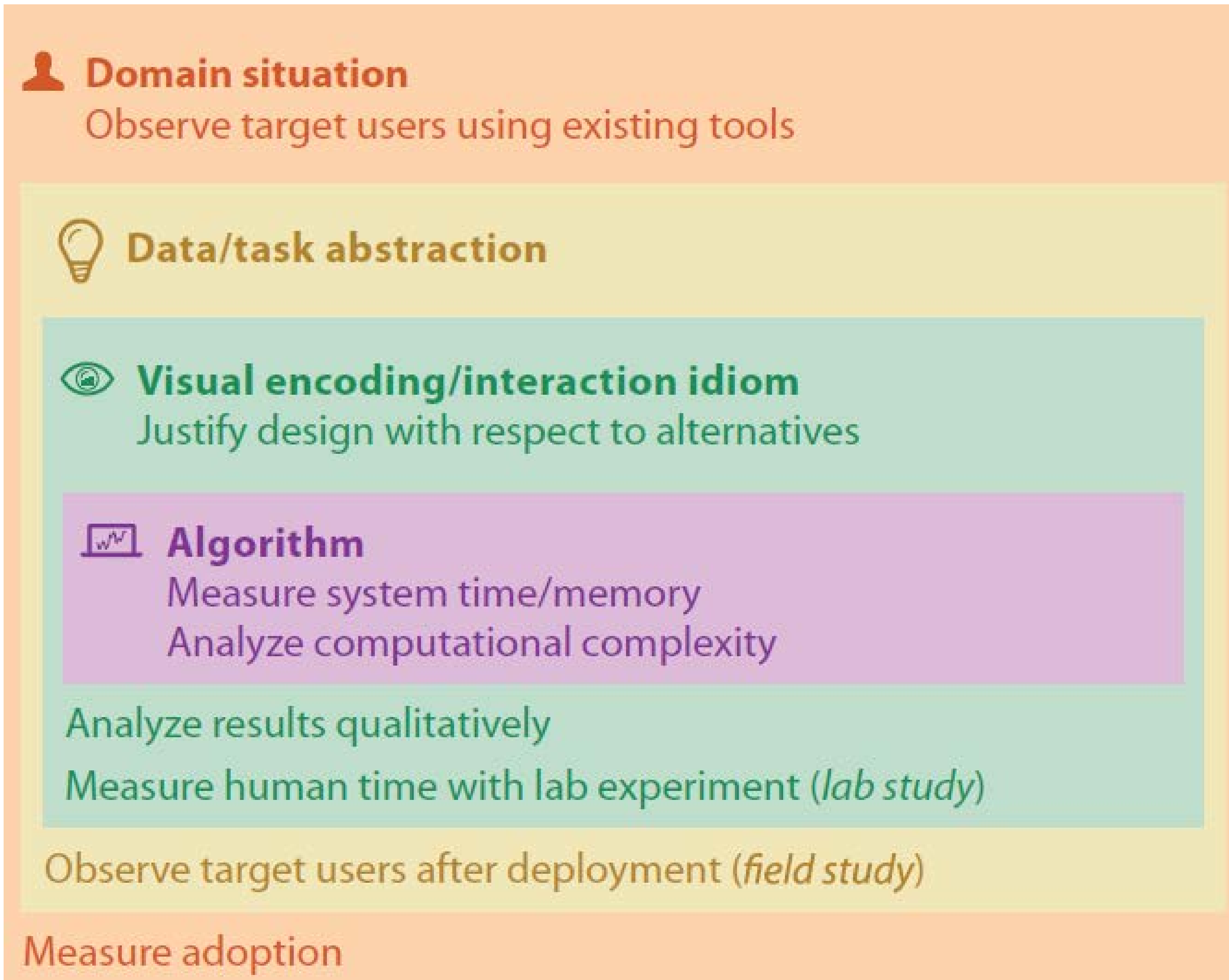


[Munzner]

Nested model for vis design: **Threats**

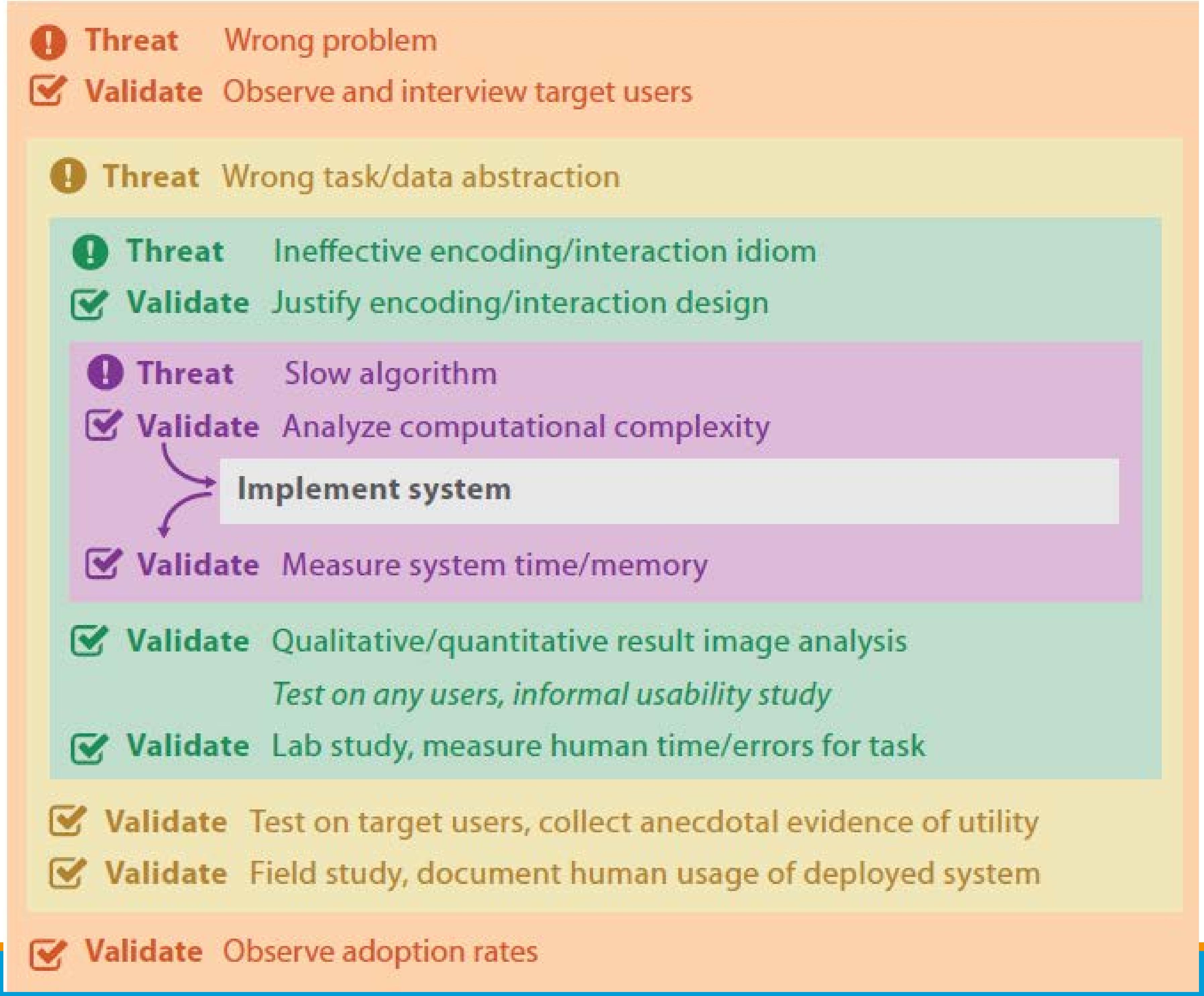
[Munzner]

Nested model for vis design: **Validation steps**



[Munzner]

Nested model for vis design



[Munzner]

Workflow to design a tool

11

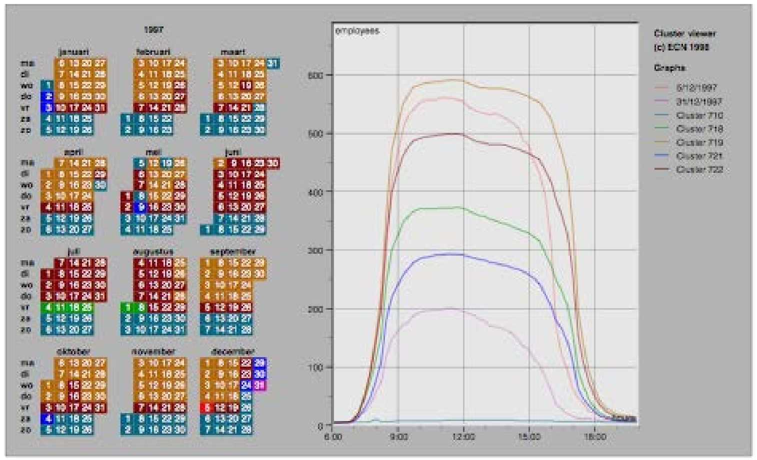
Make the right tool

Questions

Data

Tasks

Vis researcher



van Wijk:1999

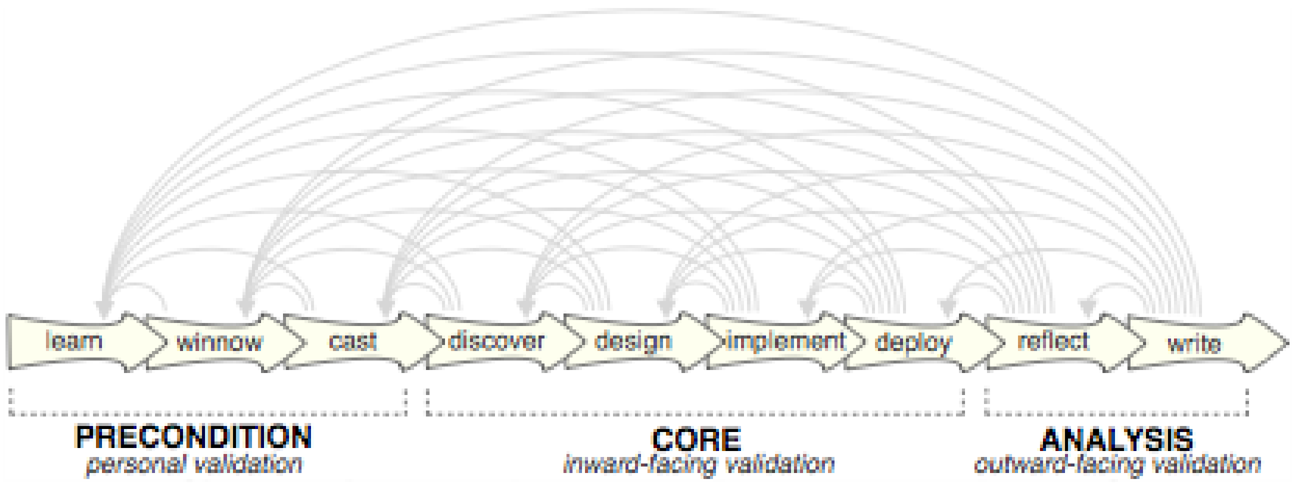
http://halalfocus.net/wp-content/uploads/2014/05/question-mark-nothing.jpg

Make the right tool

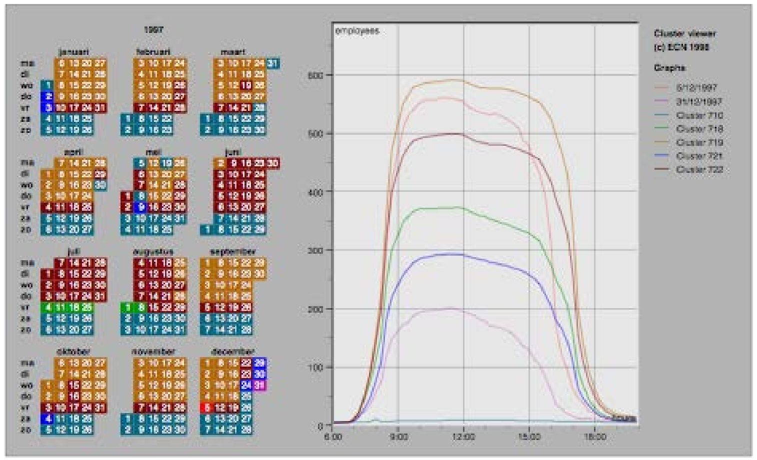
Questions

Data

Tasks

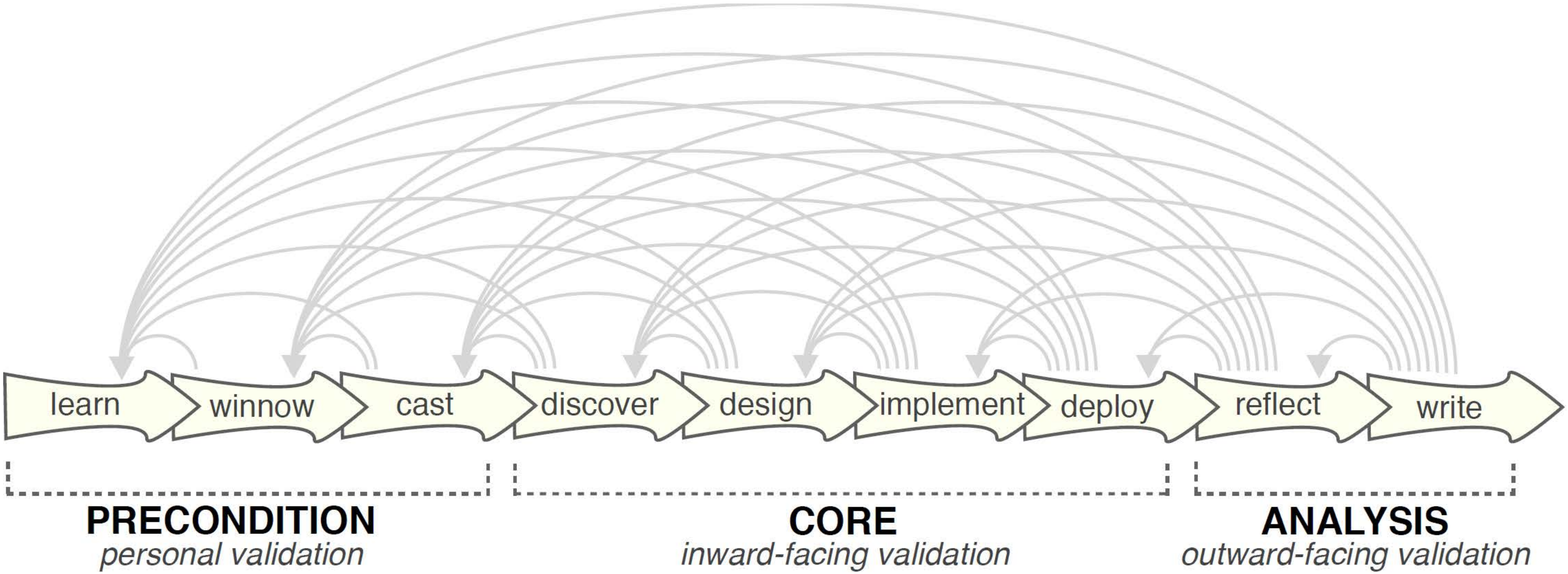


Design study methodology



van Wijk:1999

Design study methodology



Sedlmair:2012

Design study definition

Design study papers explore the choices made when applying infovis techniques in an application area, for example relating the visual encodings and interaction techniques to the requirements of the target task. Although a limited amount of application domain background information can be useful to provide a framing context in which to discuss the specifics of the target task, the primary focus of the case study must be the infovis content. Describing new techniques and algorithms developed to solve the target problem will strengthen a design study paper, but the requirements for novelty are less stringent than in a Technique paper.

[InfoVis03 CFP, infovis.org/infovis2003/CFP]

Munzner

Design study definition

Design study papers **explore the choices made when applying infovis techniques in an application area**, for example relating the visual encodings and interaction techniques to the requirements of the target task. Although a limited amount of application domain background information can be useful to provide a framing context in which to discuss the specifics of the target task, the primary focus of the case study must be the infovis content. Describing new techniques and algorithms developed to solve the target problem will strengthen a design study paper, but the requirements for novelty are less stringent than in a Technique paper.

[InfoVis03 CFP, infovis.org/infovis2003/CFP]

Munzner

Design study definition

Design study papers explore the choices made when applying infovis techniques in an application area, for example relating the visual encodings and interaction techniques to the requirements of the target task. Although a limited amount of application domain background information can be useful to provide a framing context in which to discuss the specifics of the target task, **the primary focus of the case study must be the infovis content**. Describing new techniques and algorithms developed to solve the target problem will strengthen a design study paper, but the requirements for novelty are less stringent than in a Technique paper.

[InfoVis03 CFP, infovis.org/infovis2003/CFP]

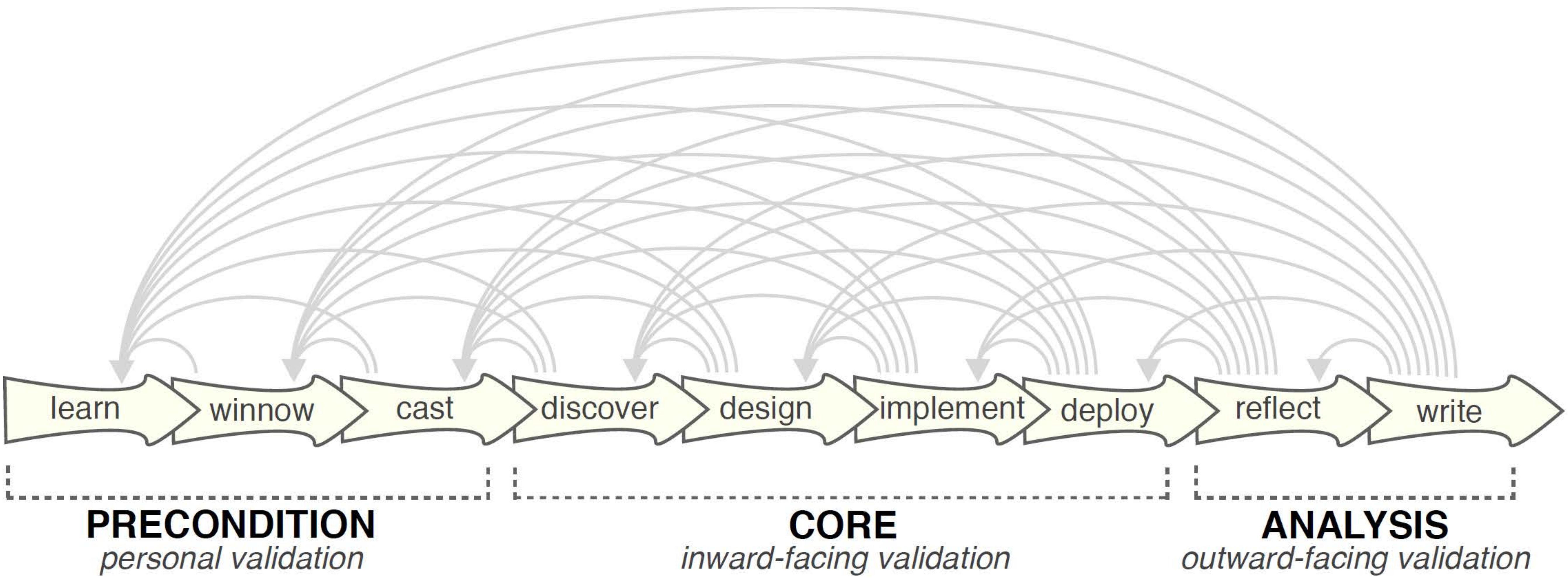
Design study definition

Design study papers explore the choices made when applying infovis techniques in an application area, for example relating the visual encodings and interaction techniques to the requirements of the target task. Although a limited amount of application domain background information can be useful to provide a framing context in which to discuss the specifics of the target task, the primary focus of the case study must be the infovis content. Describing new techniques and algorithms developed to solve the target problem will strengthen a design study paper, but **the requirements for novelty are less stringent than in a Technique paper**.

[InfoVis03 CFP, infovis.org/infovis2003/CFP]

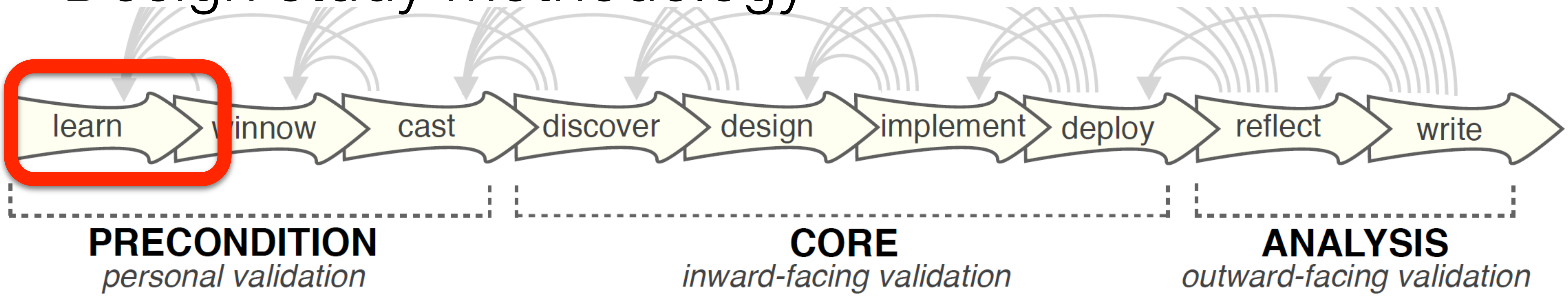
Munzner

Design study methodology



Sedlmair:2012

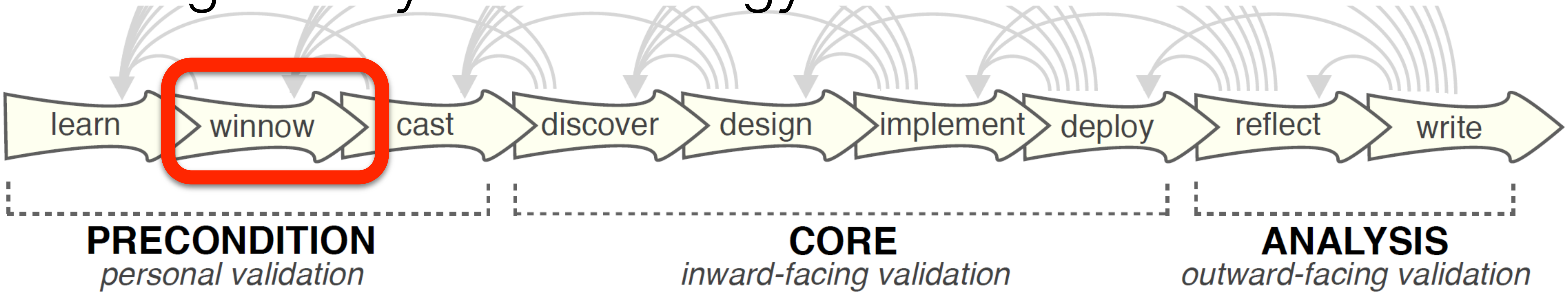
Design study methodology



What tools/techniques are available?

- Read vis papers
- Read vis books
- Talk to vis practitioners
- This course!

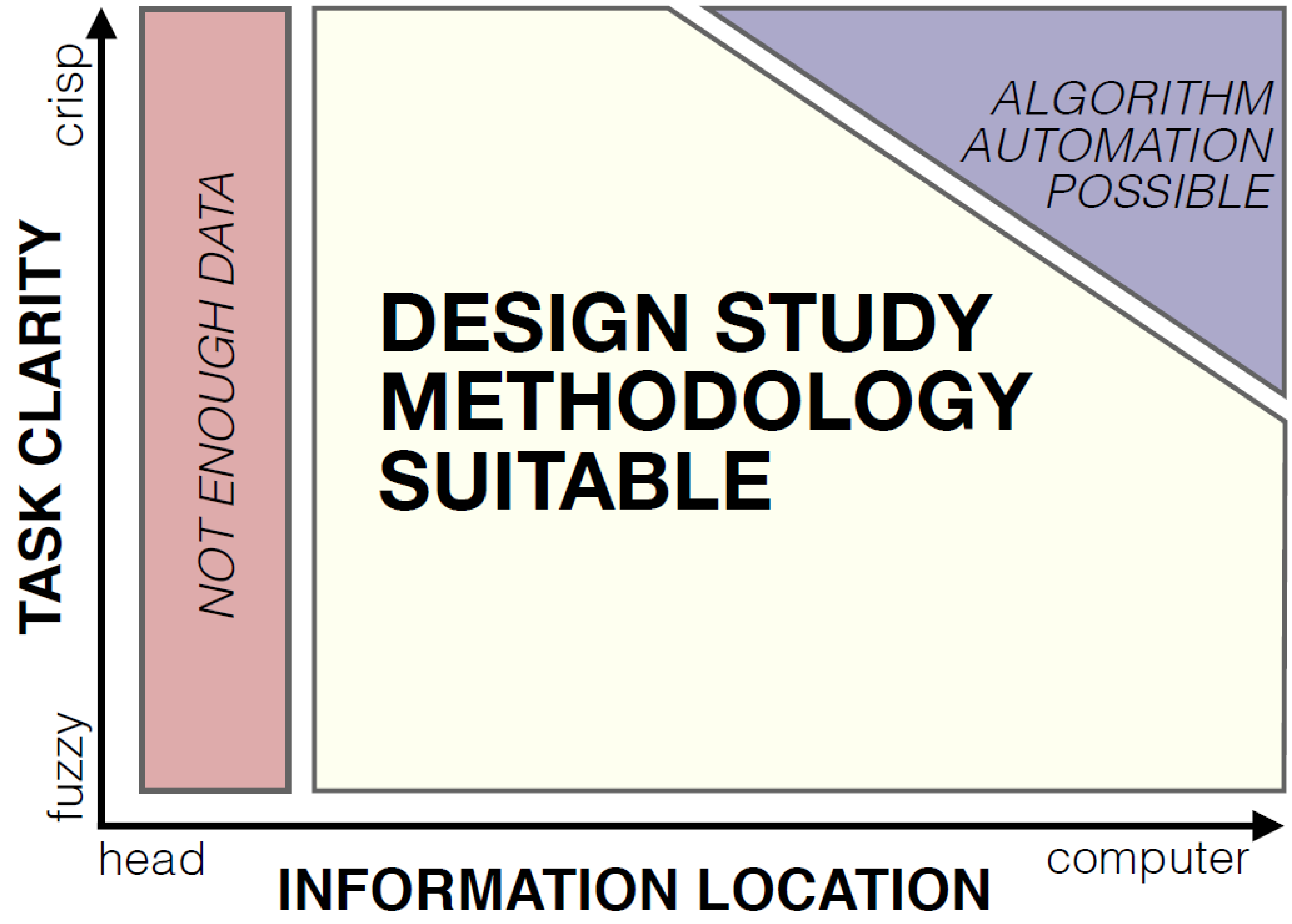
Design study methodology



Are these good collaborators?

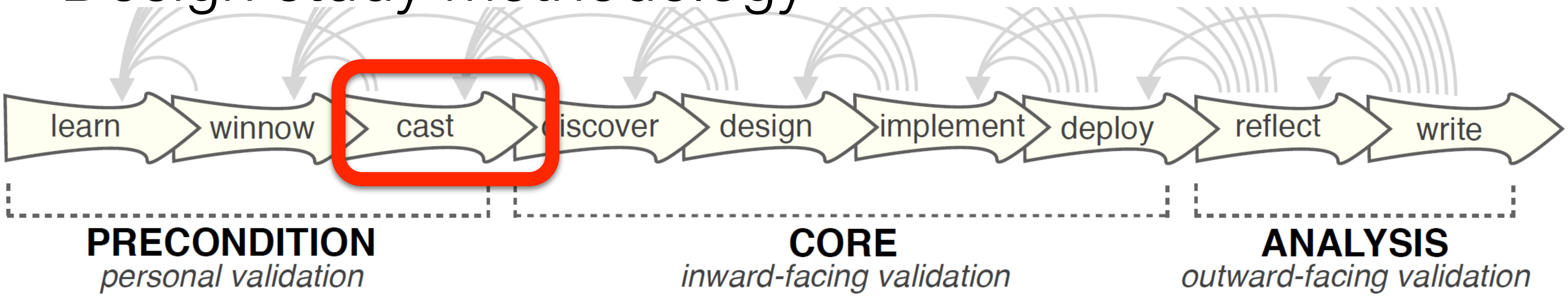
- Do they have interesting problems?
- Do they need novel solutions?
- Is there data?
- Can I work with these people?

When can you do a design study?



Sedlmair:2012

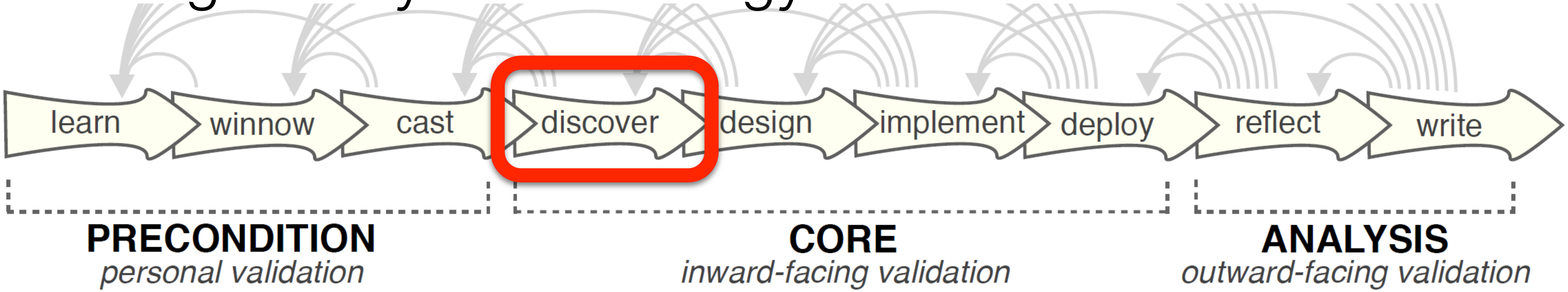
Design study methodology



Who's who?

- Do people have time for a new project?
- "Front-line analyst" is the domain expert
- Are there false "front-line analysts"?
- Do you need a "translator"?

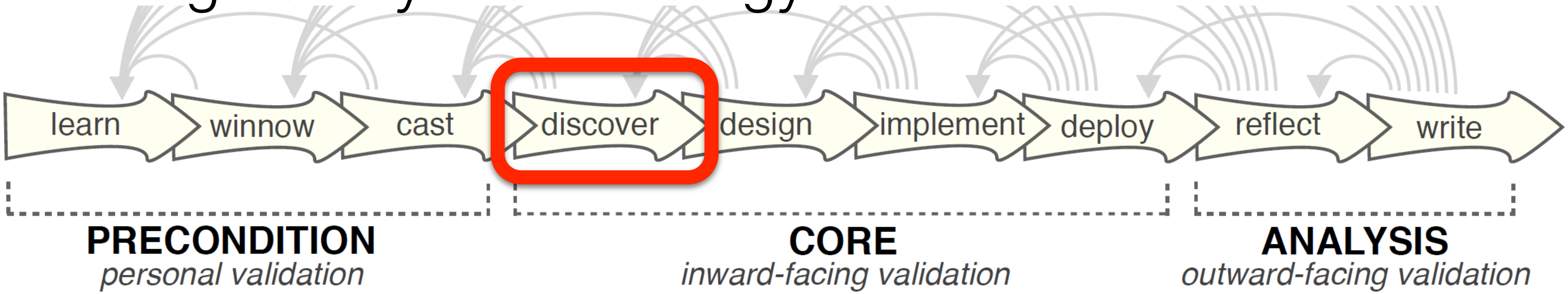
Design study methodology



Problem characterization and abstraction

- Requirements analysis
- Critical reflection on requirements!
- Abstraction is important for transferability
- Need some domain-expert knowledge

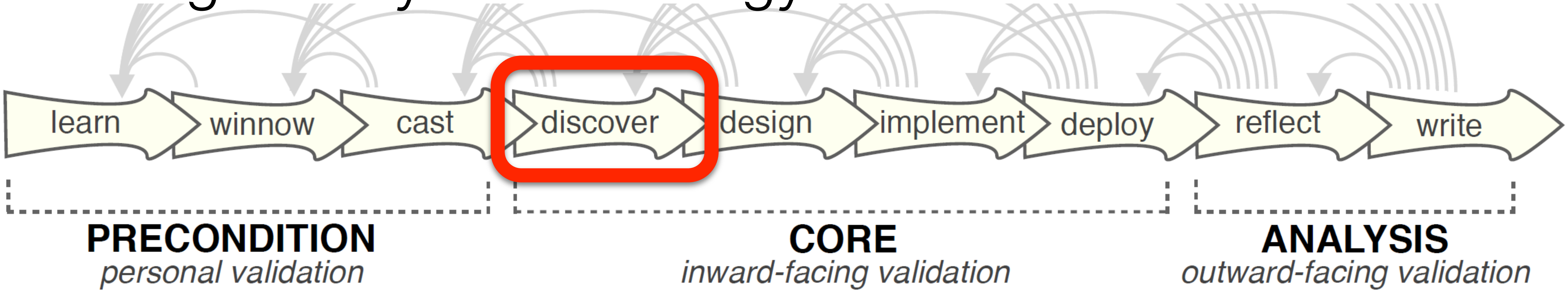
Design study methodology



example: Cluster-Calendar, van Wijk and van Selow

- Overall goal: are there temporal patterns in power consumption?

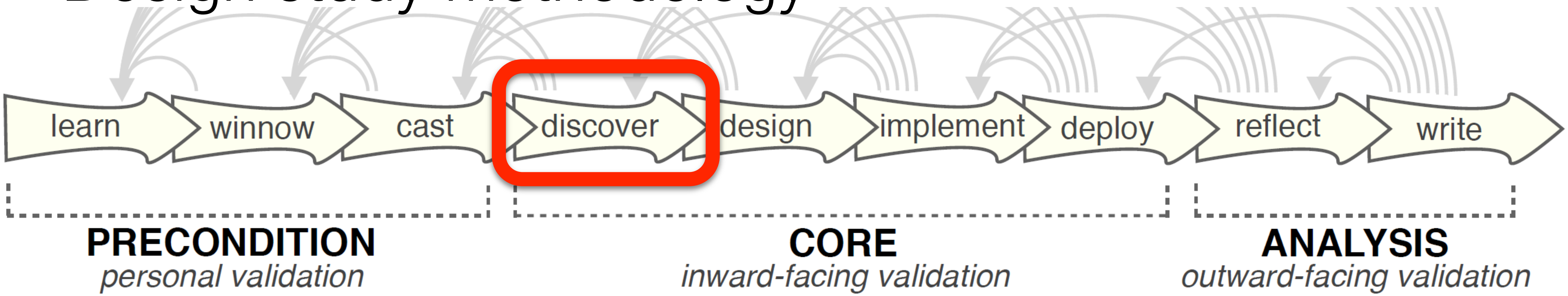
Design study methodology



example: Cluster-Calendar, van Wijk and van Selow

- Data: ~50K pairs of (value, time)
- Tasks
  - Find standard day patterns
  - Find out how patterns are distributed over year, week, season
  - Find outliers from standard daily patterns
  - Want overview first, details on demand

Design study methodology

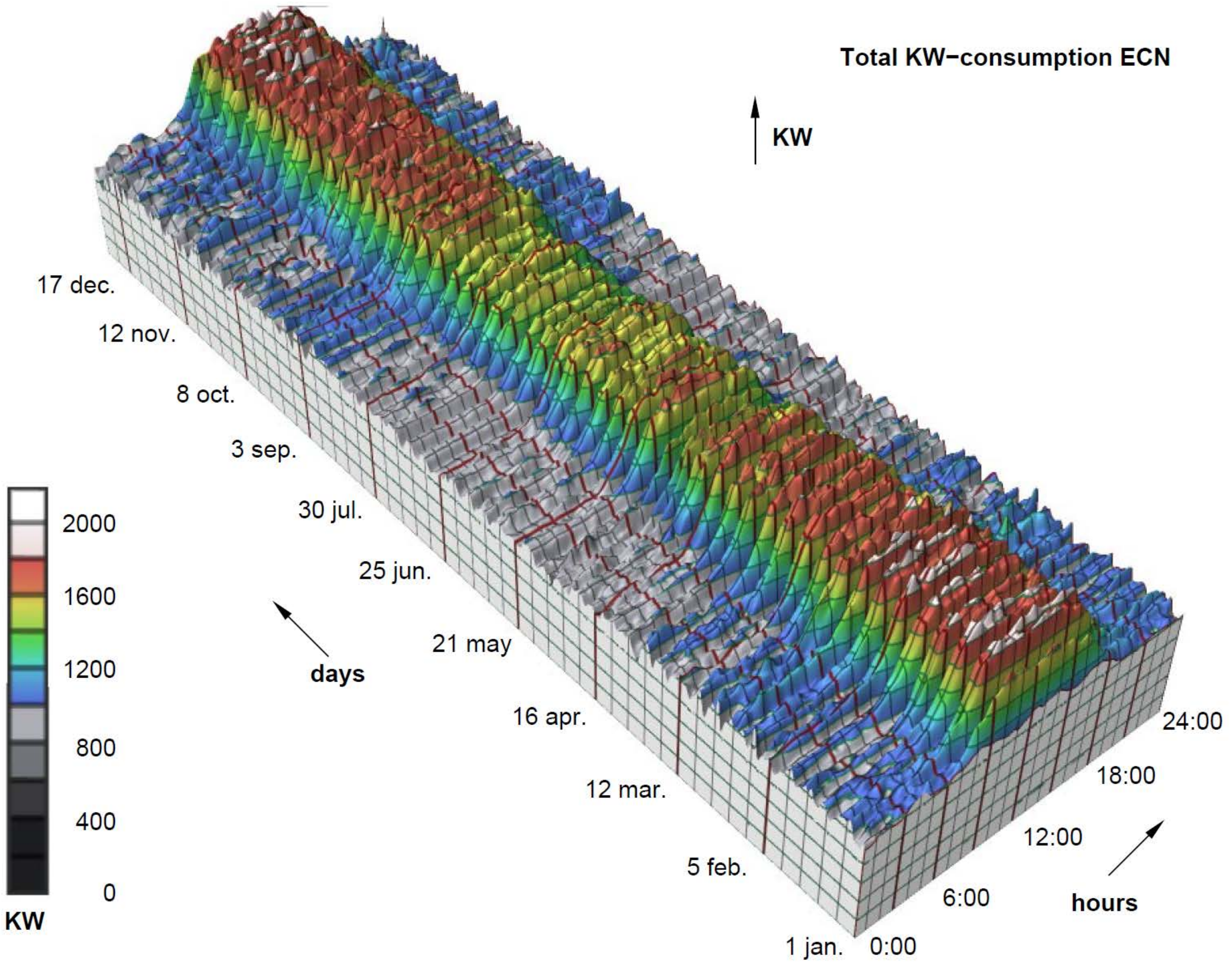


example: Cluster-Calendar, van Wijk and van Selow

- Limitations of previous work:
  - predictive mathematical models: details lost
  - scale-space approaches (wavelet, fourier, fractal): hard to interpret, known scales lost
  - 3D mountain diagram (x: hours, y: value, z: days)

Design study methodology

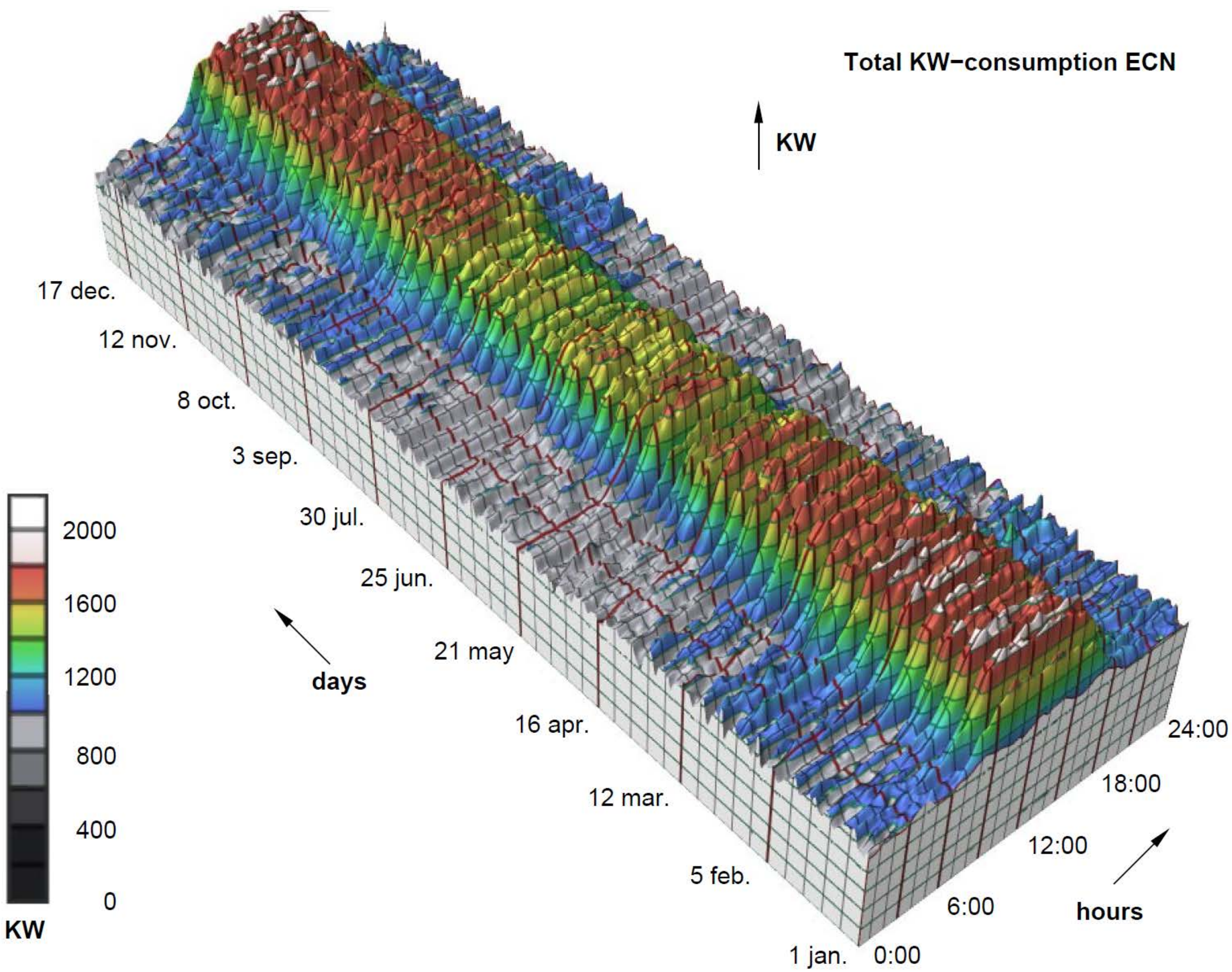
Power demand by ECN, displayed as a function of hours and days



van Wijk:1999

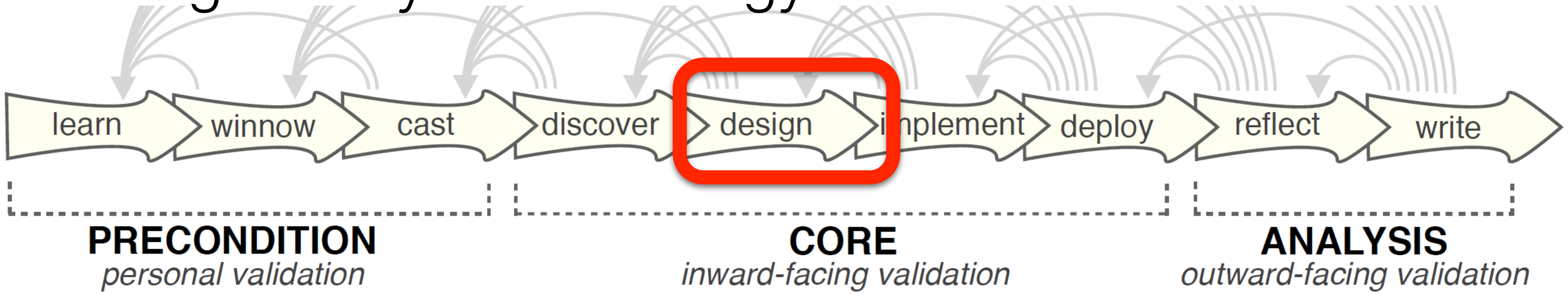
Design study methodology

- Pretty, not so useful
- Daily, weekly patterns are hard to see



van Wijk:1999

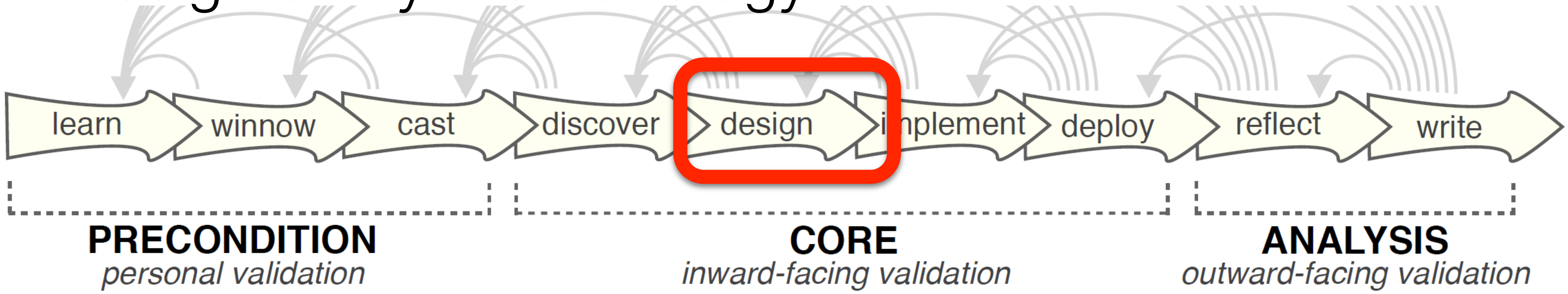
Design study methodology



Data abstraction, visual encoding, interaction

- What data transformations are needed?
- What visual designs to use?
- How to tie this together with interaction?
- Don't code!

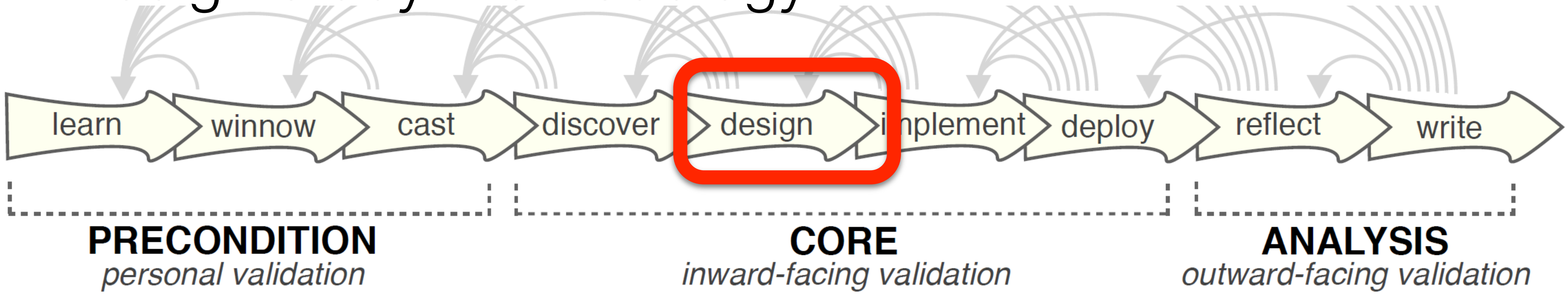
Design study methodology



example: Cluster-Calendar, van Wijk and van Selow

- Data transform: hierarchical clustering

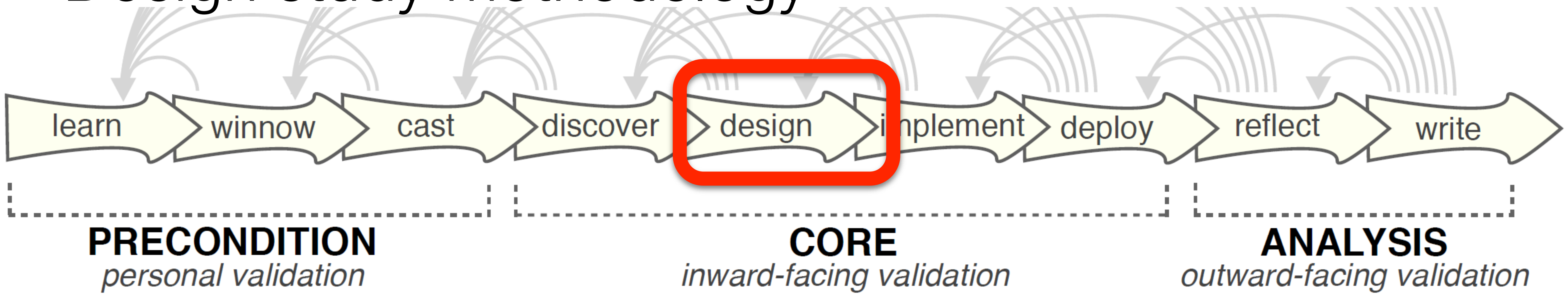
Design study methodology



example: Cluster-Calendar, van Wijk and van Selow

- Data transform: hierarchical clustering
- start with M day patterns
  - compute pair-wise differences, merge most similar
  - now we have M-1 patterns
  - repeat until we have 1 root cluster
- result: binary hierarchy of clusters

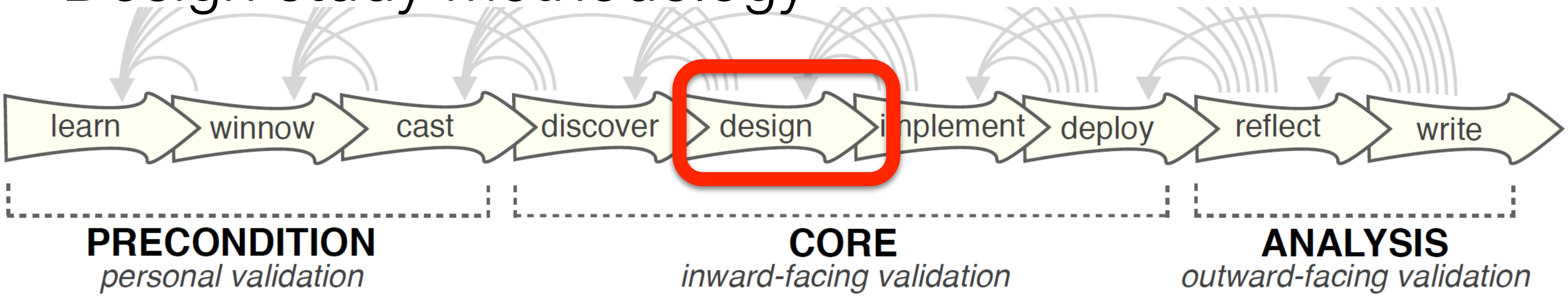
Design study methodology



example: Cluster-Calendar, van Wijk and van Selow

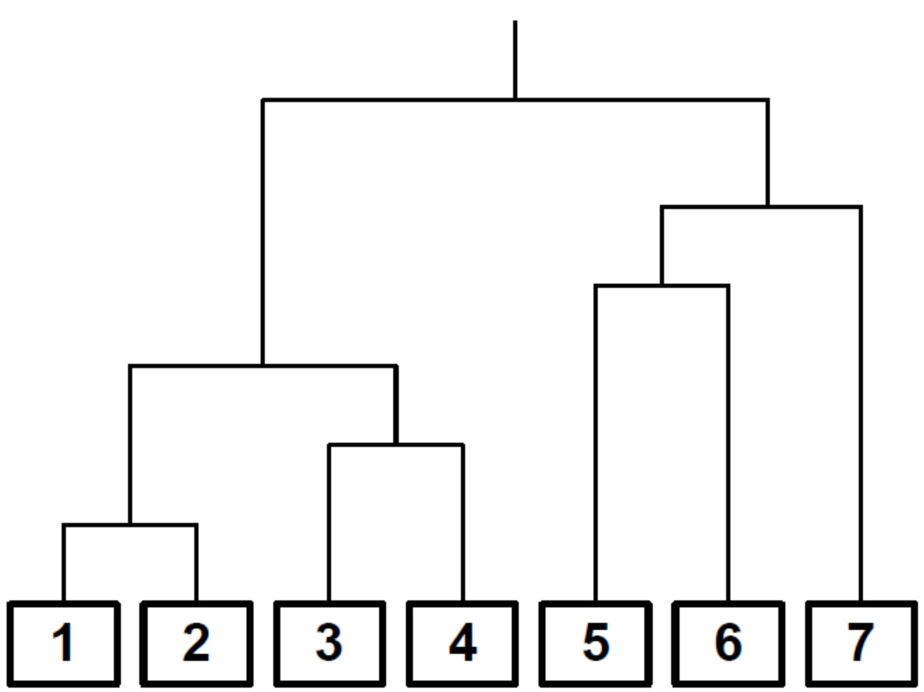
- Data transform: hierarchical clustering
- issues:
  - distance metric to use?
  - how to display the cluster?

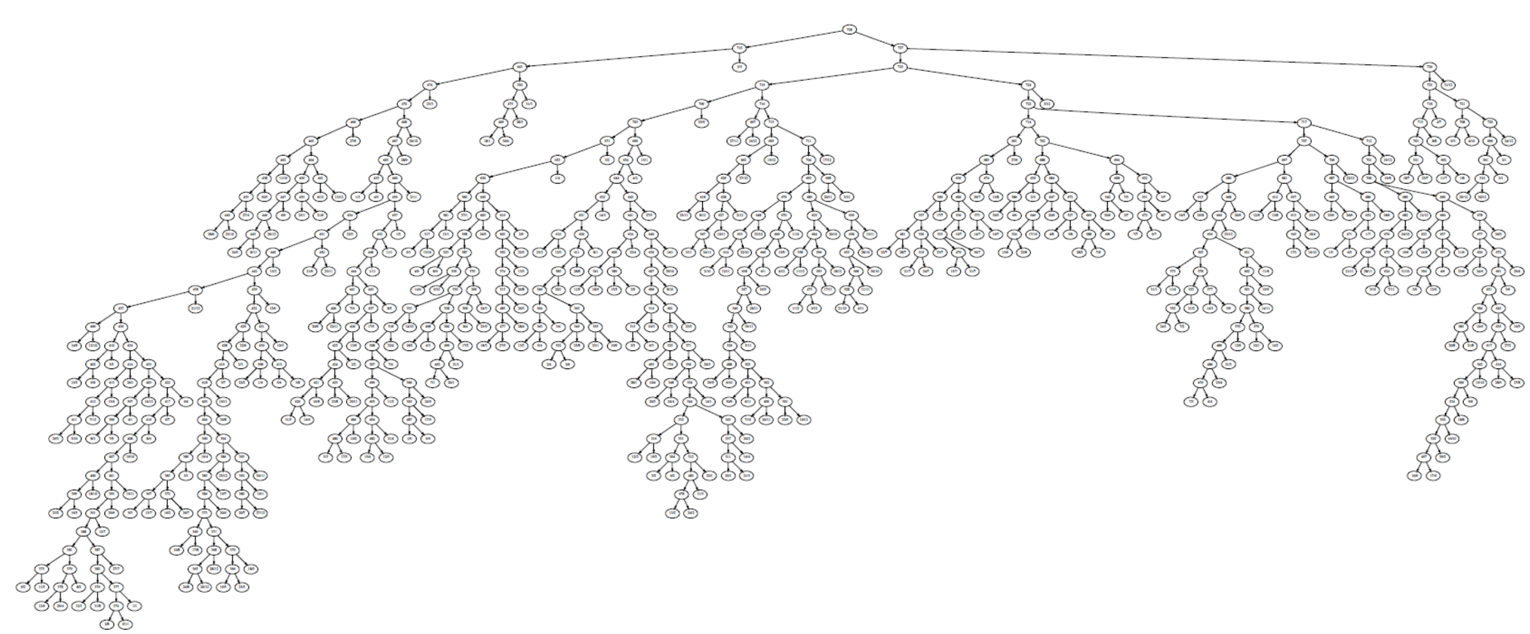
Design study methodology



example: Cluster-Calendar, van Wijk and van Selow

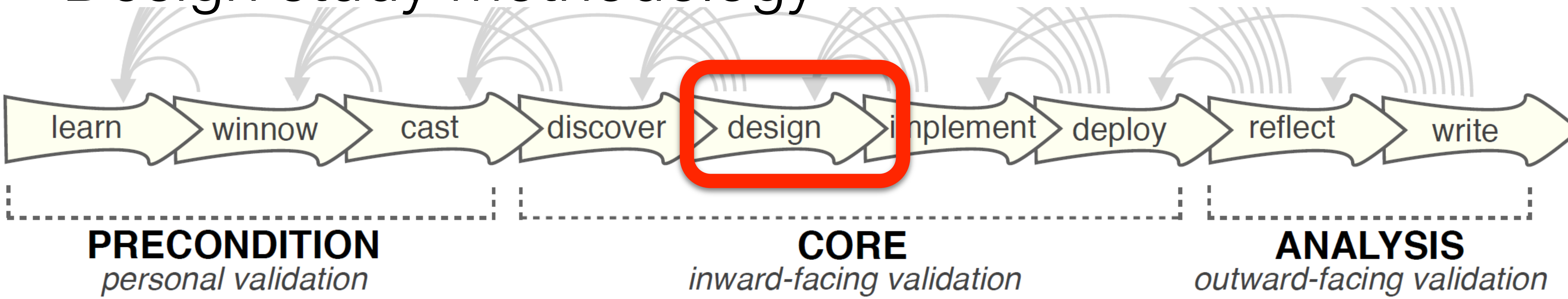
- dendrogram





van Wijk:1999

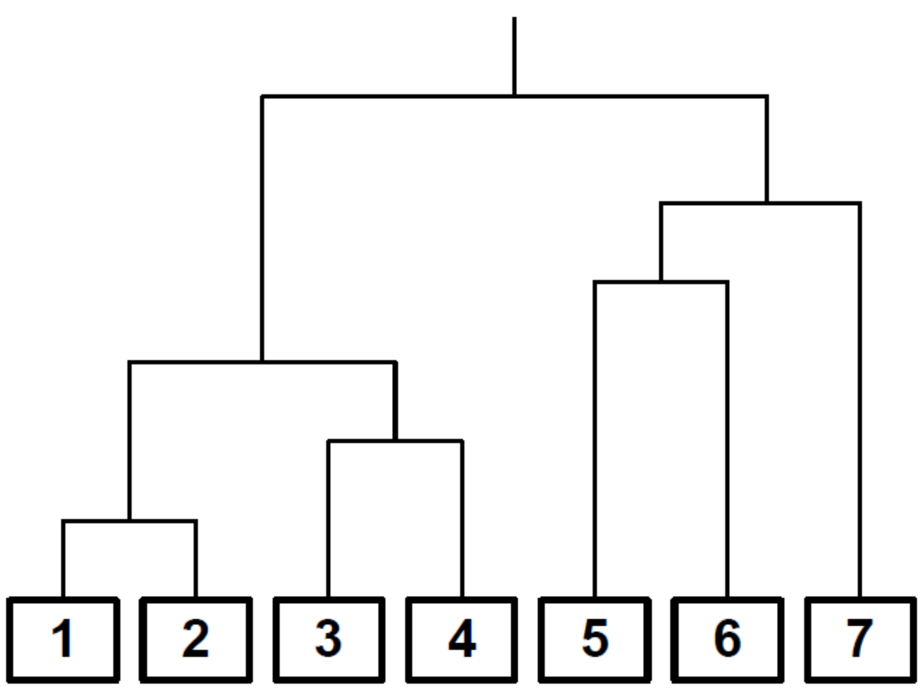
Design study methodology

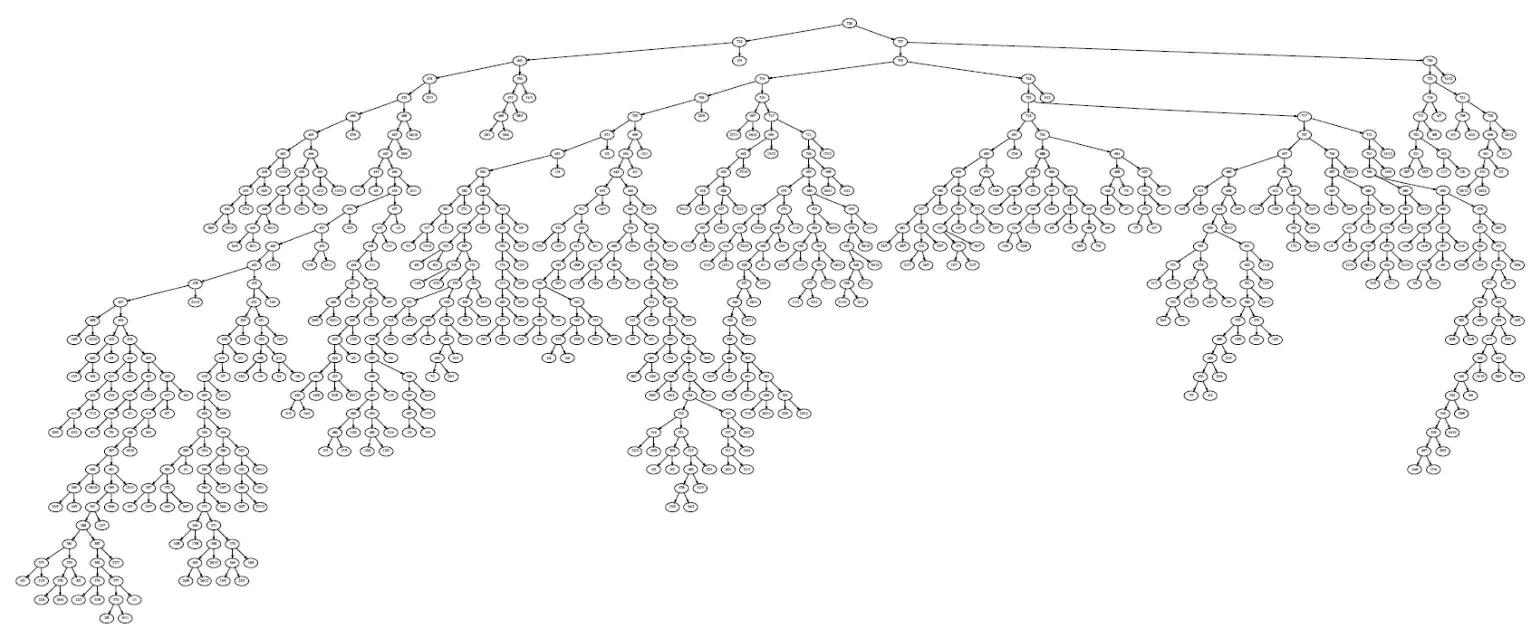


example: Cluster-Calendar, van Wijk and van Selow

- dendrogram

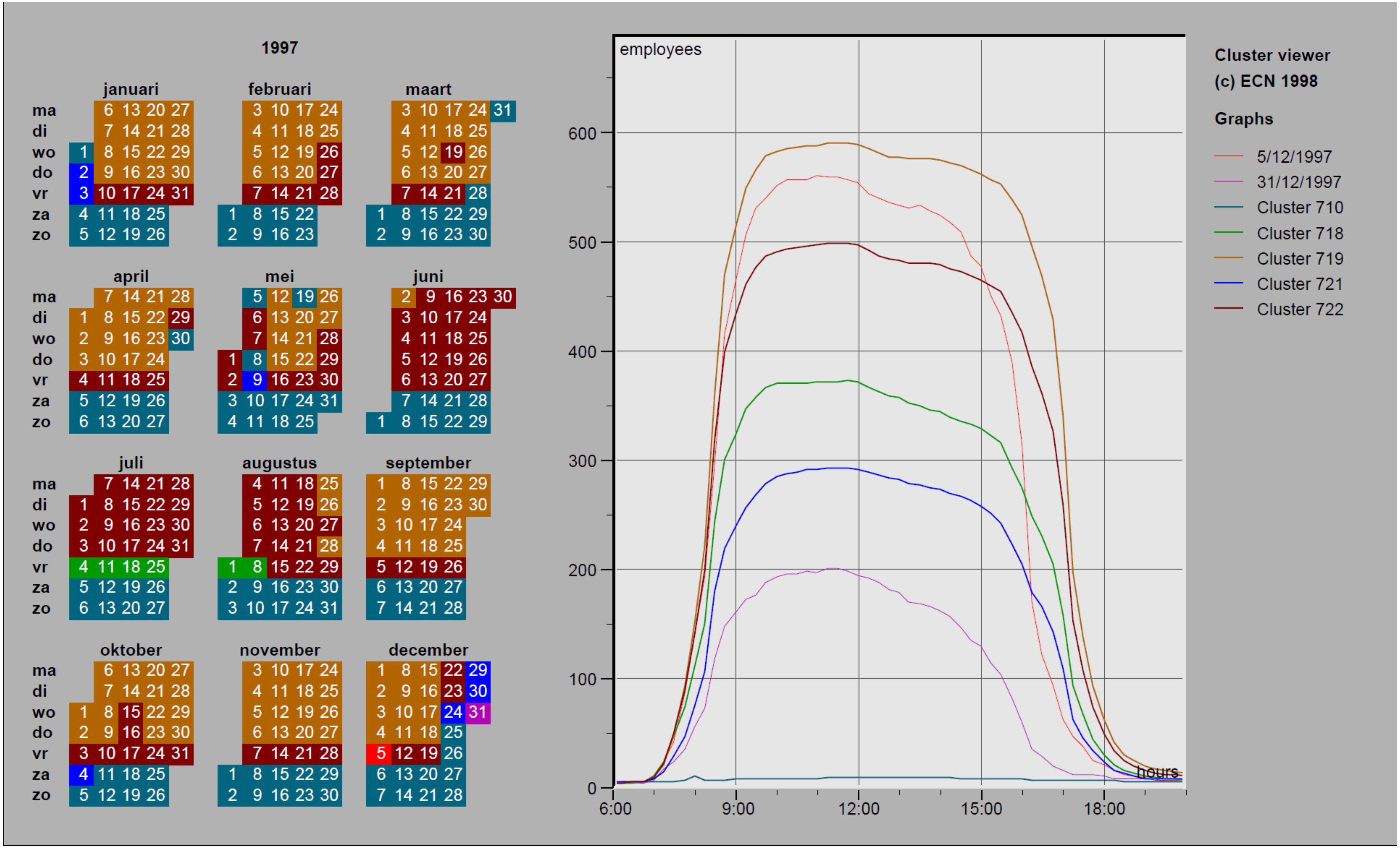
Shows hierarchical structure but not time distribution!





van Wijk:1999

Design study methodology

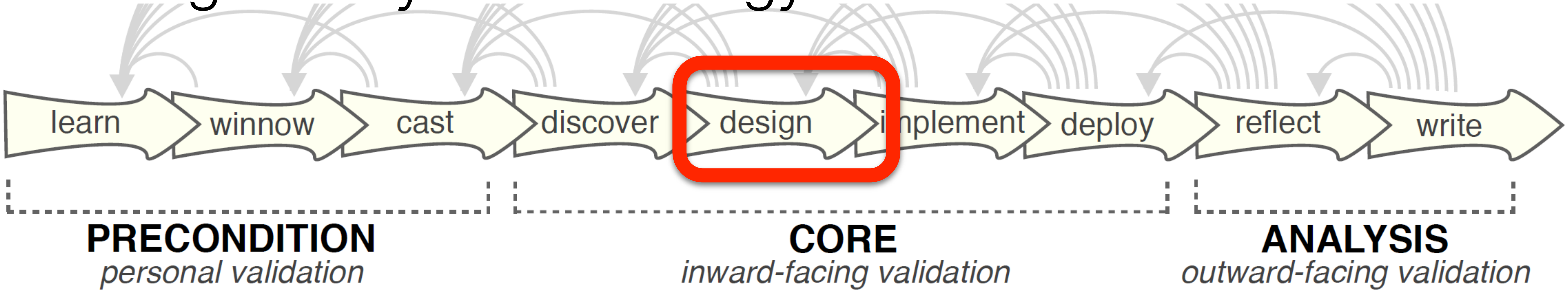


Overview

Detail

van Wijk:1999

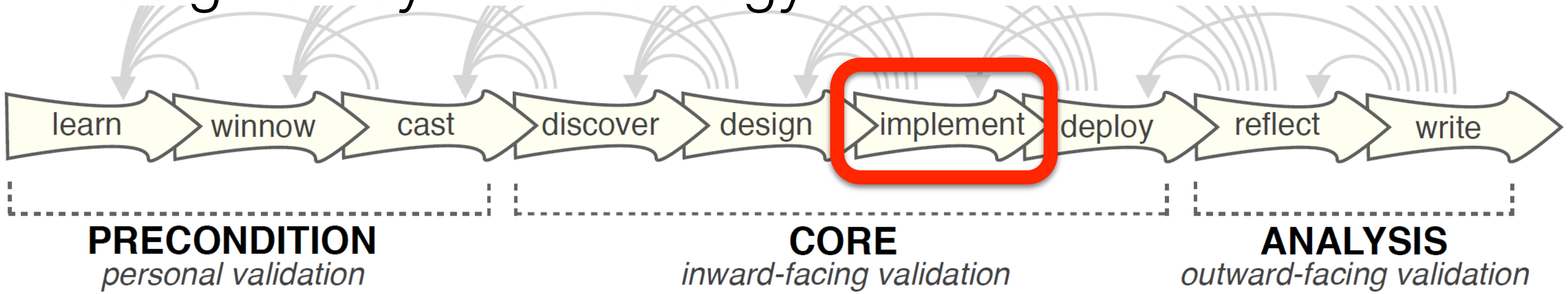
Design study methodology



example: Cluster-Calendar, van Wijk and van Selow

- clusters: data transformation to aggregate data
- calendar: familiar visual representation for time
- linking: interactive exploration of the data
- task analysis guided choices: 3D extrusion and dendrogram don't work

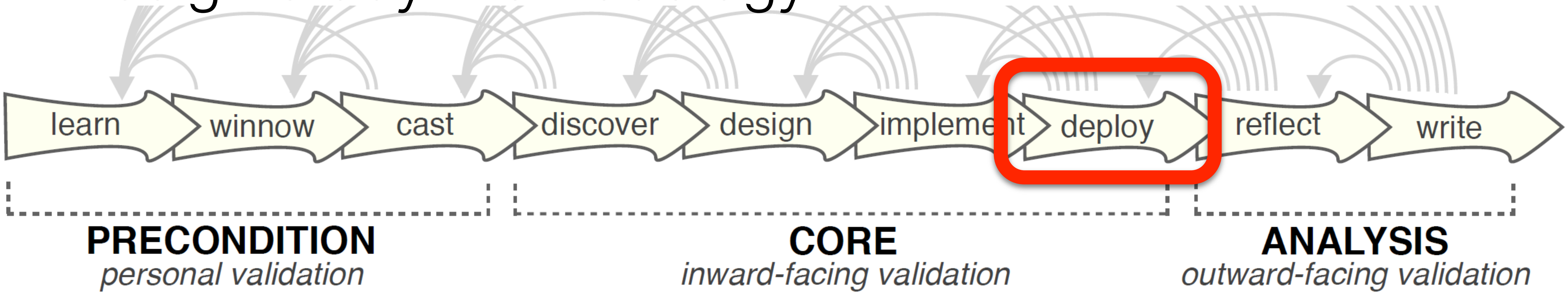
Design study methodology



Yay coding!

- Need to test design hypotheses
- Rapid prototyping (will probably throw away alot of code)
- Breaking bugs vs annoying bugs
- Fast usability testing

Design study methodology

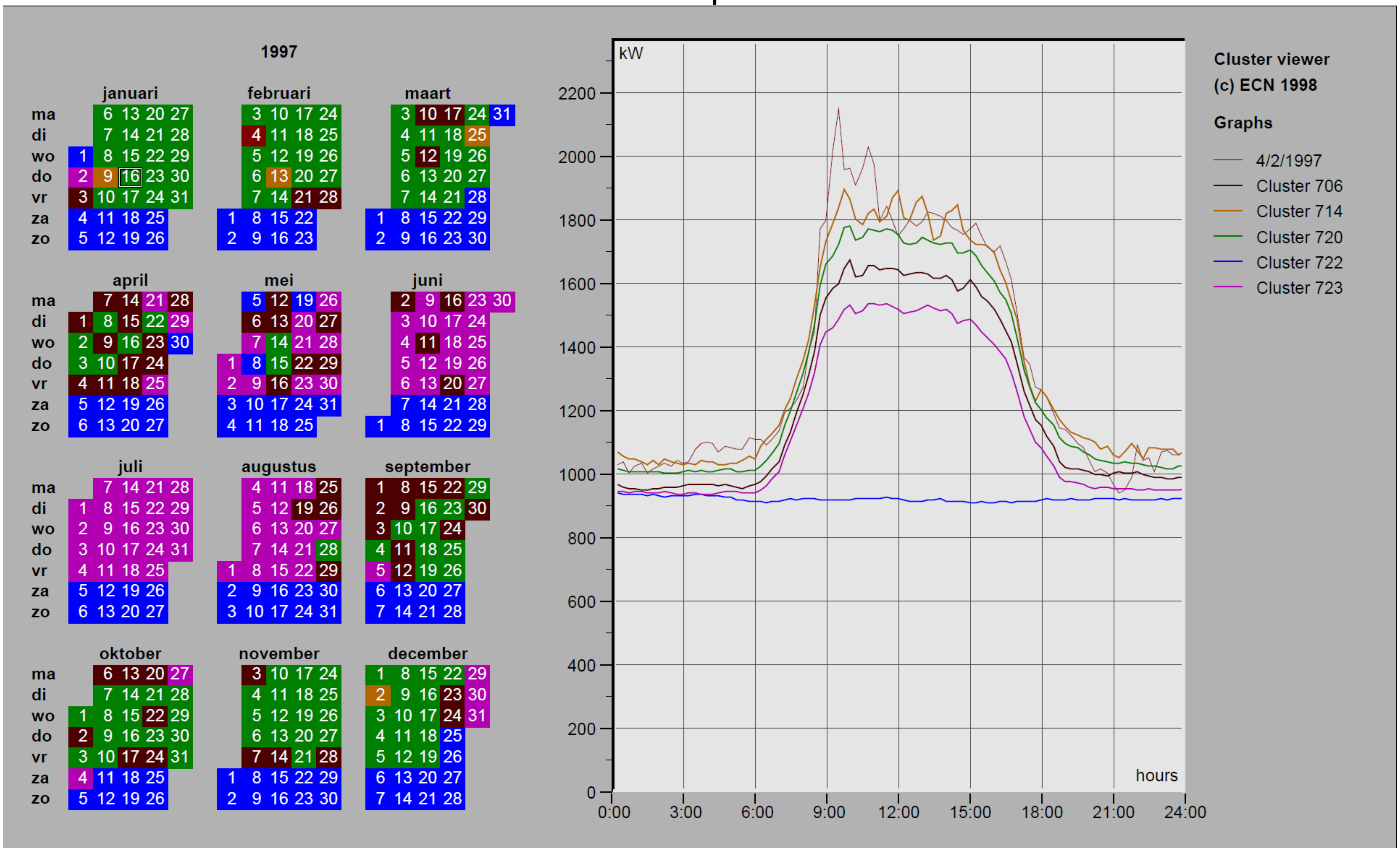


Hand-off to the users

- Domain experts need to play with software
- What works, what doesn't?
- How to evaluate?
- May need to redesign/reimplement a lot

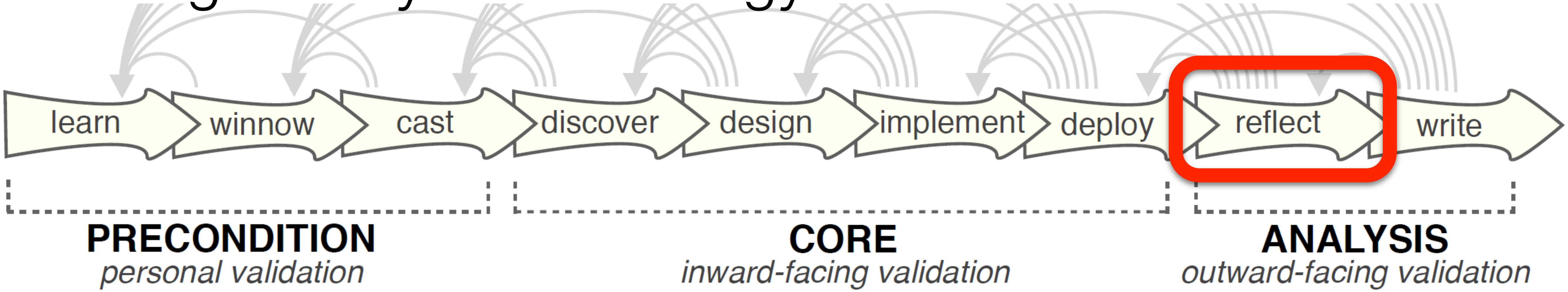
Design study methodology

Critique?



van Wijk:1999

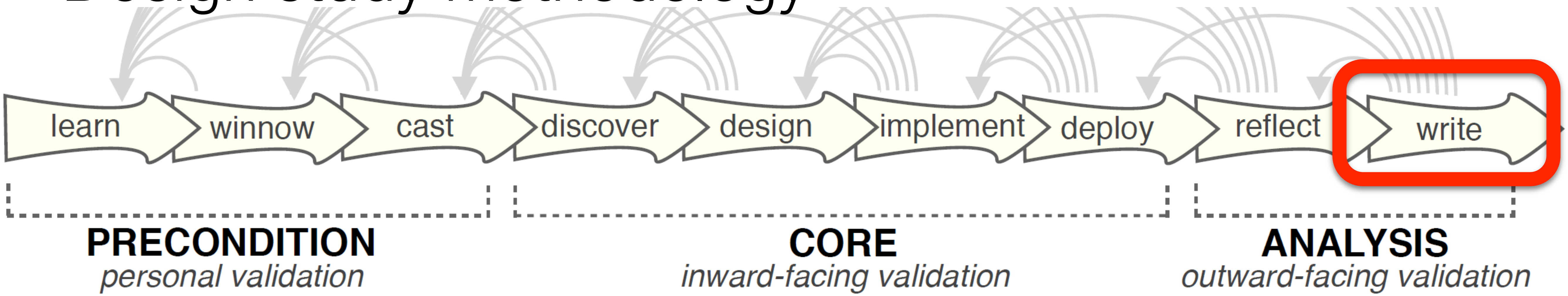
Design study methodology



Refine, reject, propose guidelines

- Compare to existing design guidelines
- Confirm which ones worked
- Reject which ones didn't work
- Come up with new guidelines

Design study methodology



Yay words!

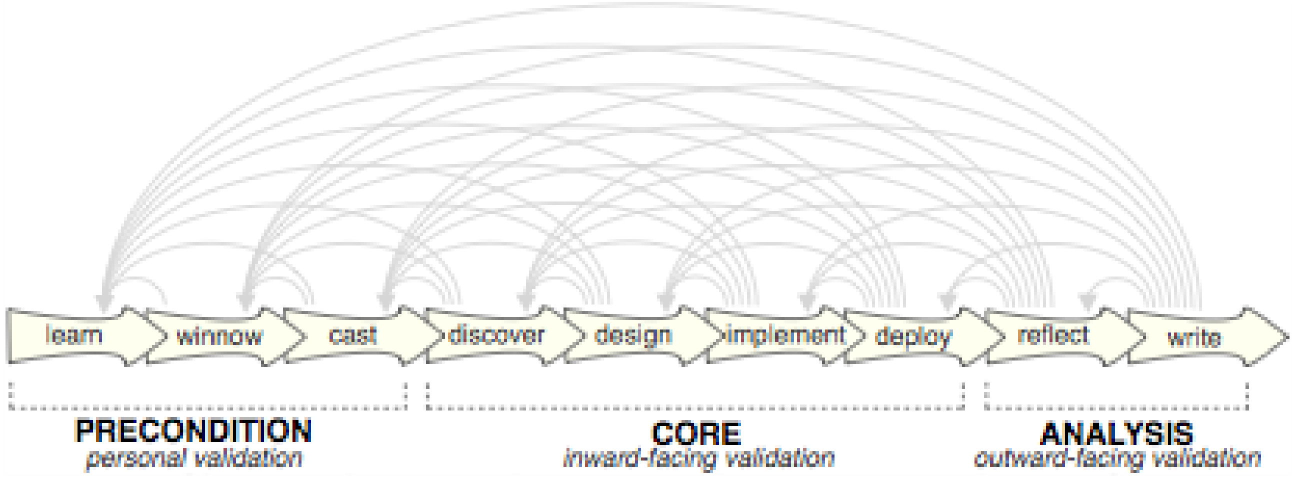
- Forces clear articulation of problem, tasks, solution
- Who else does my study help? - transferability!
- Think carefully about what readers will care about
- This takes time to do well!

Make the right tool

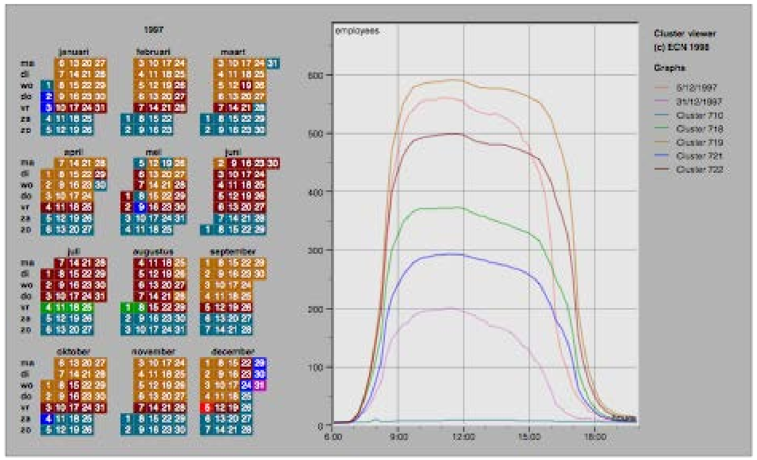
Questions

Data

Tasks

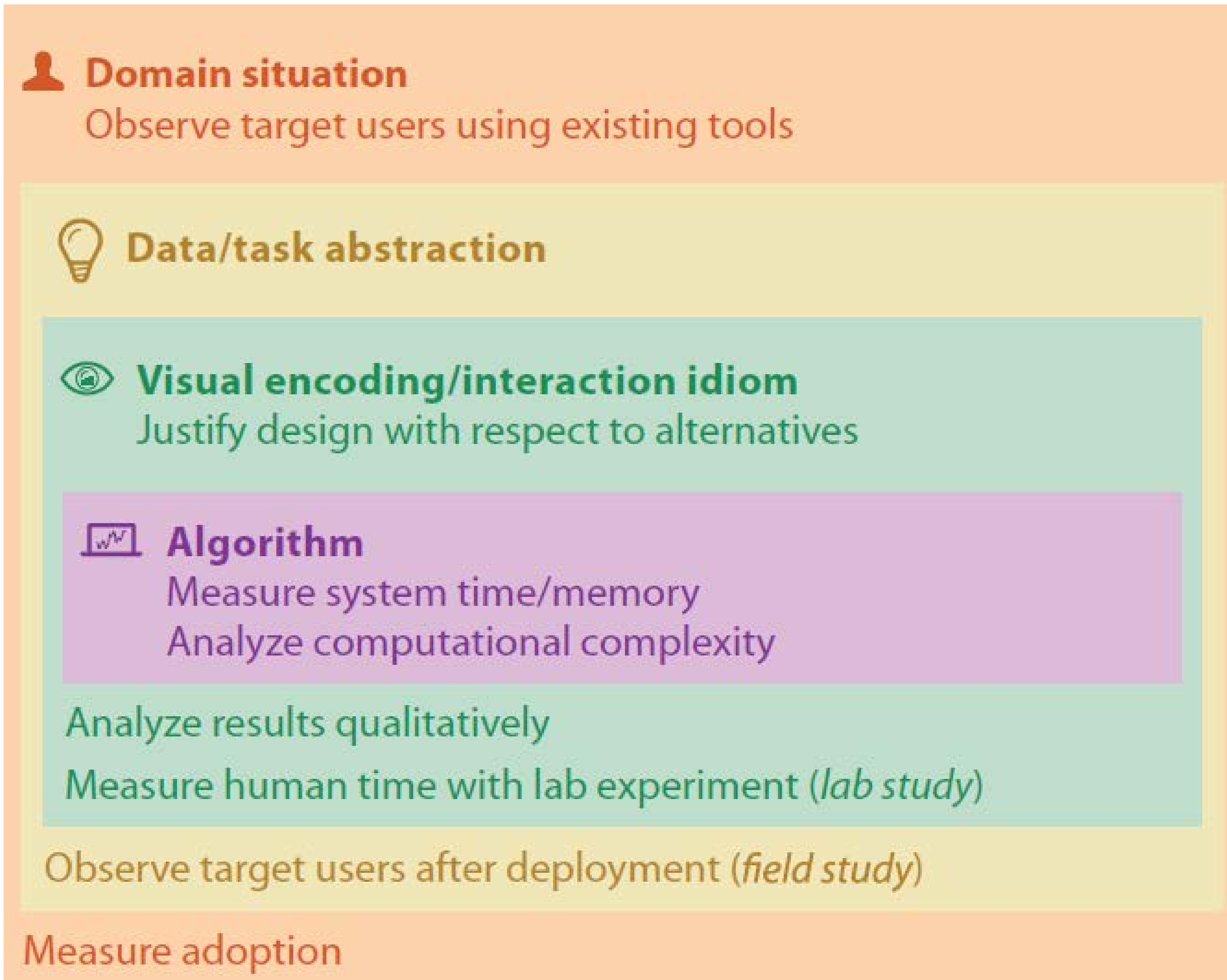


Design study methodology



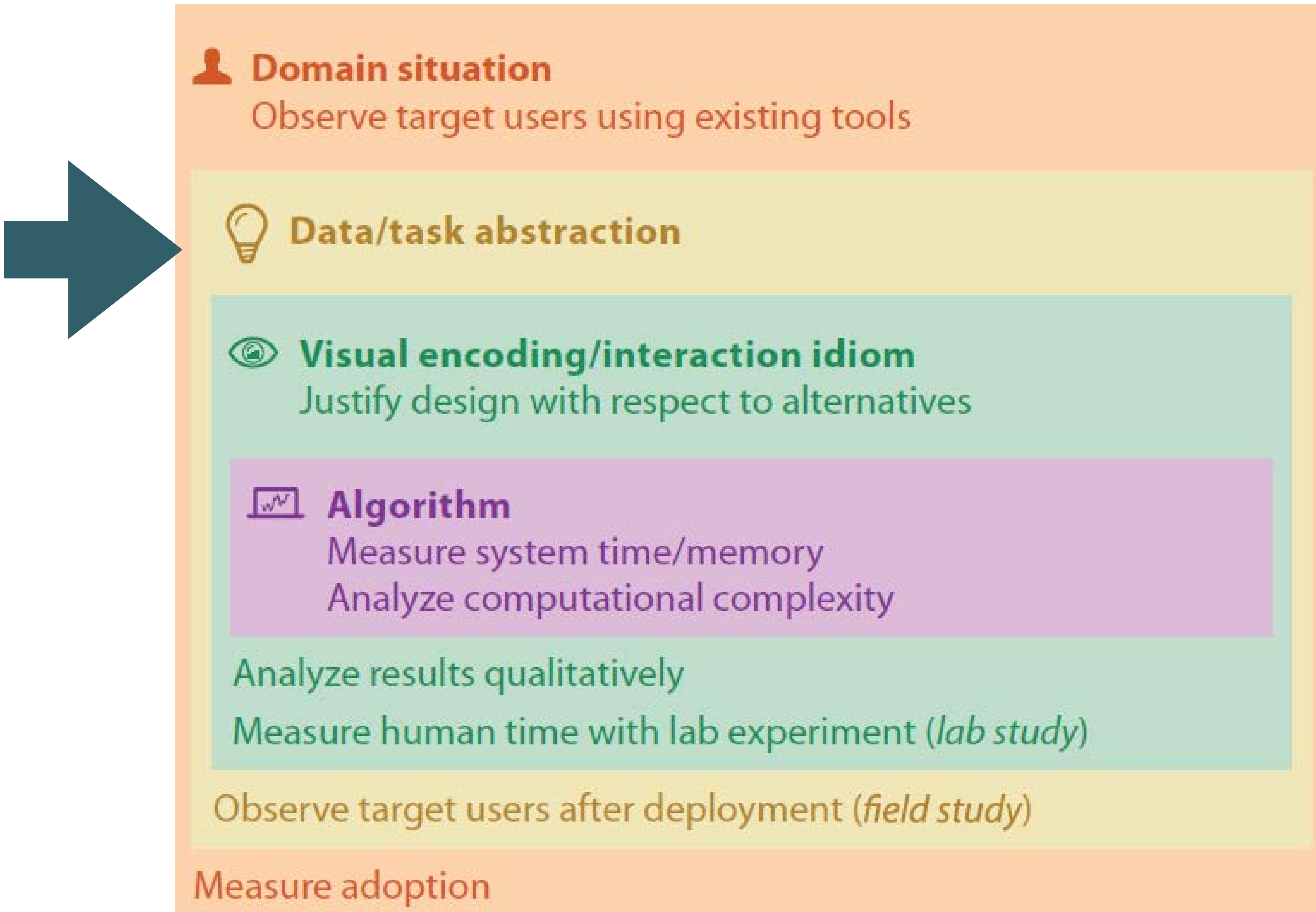
van Wijk:1999

Where are design studies?



[Munzner]

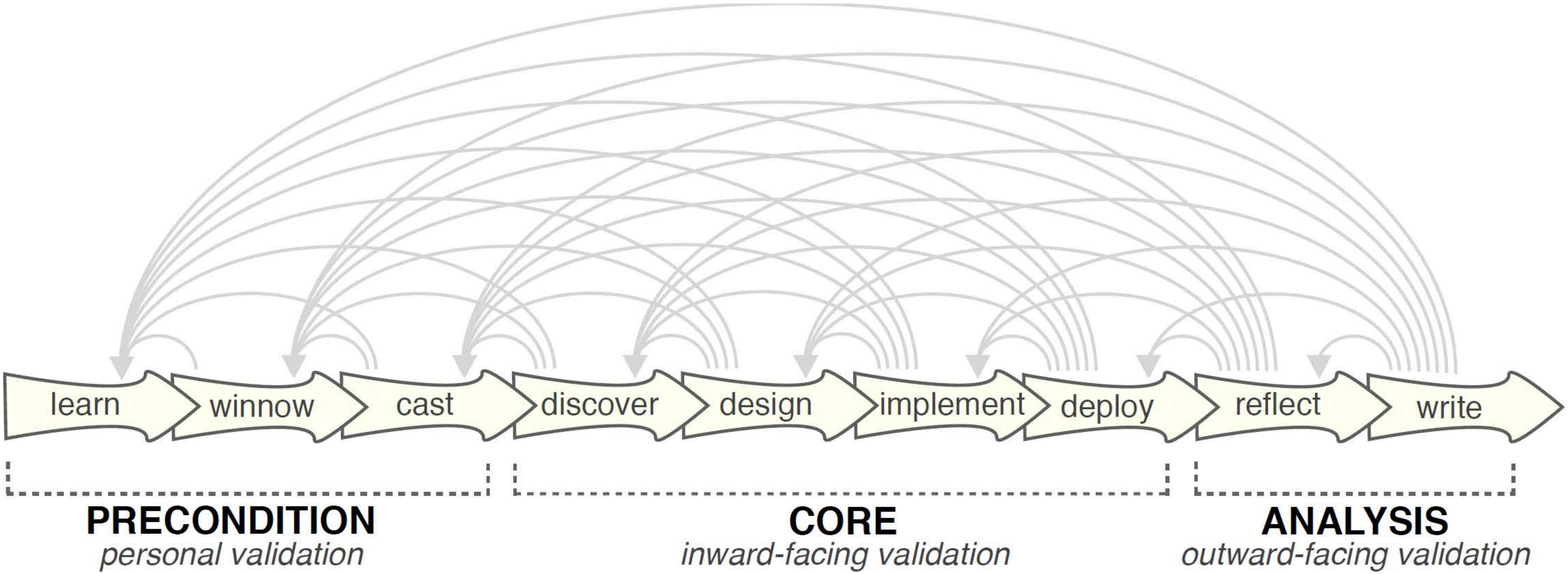
Where are design studies?



[Munzner]

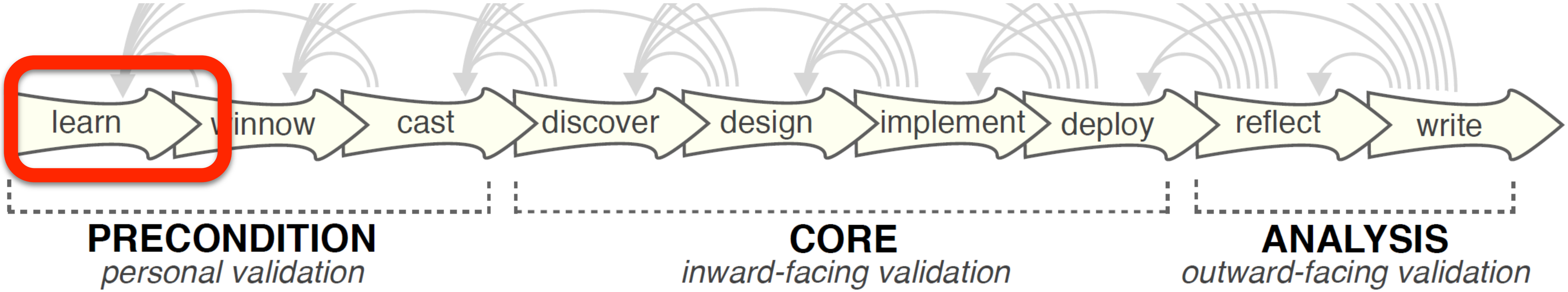
Pitfalls

Pitfalls



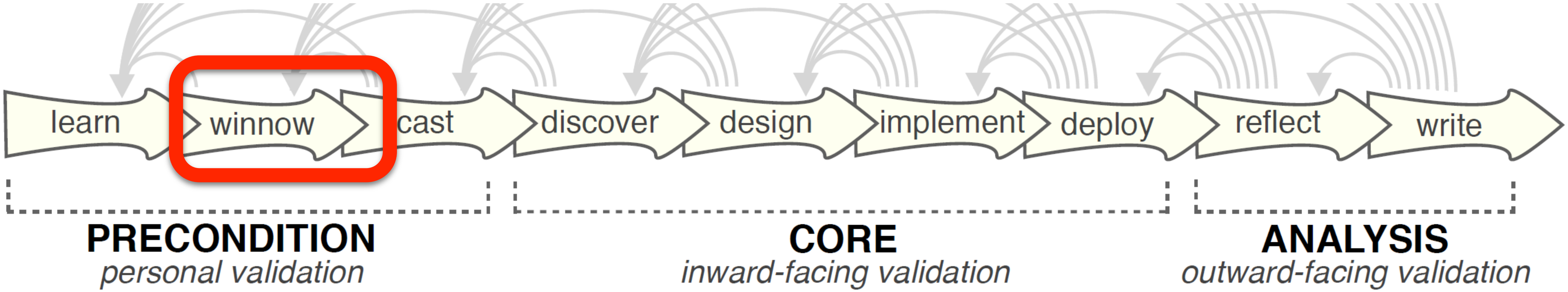
#1: Don’t skip steps!

Pitfalls



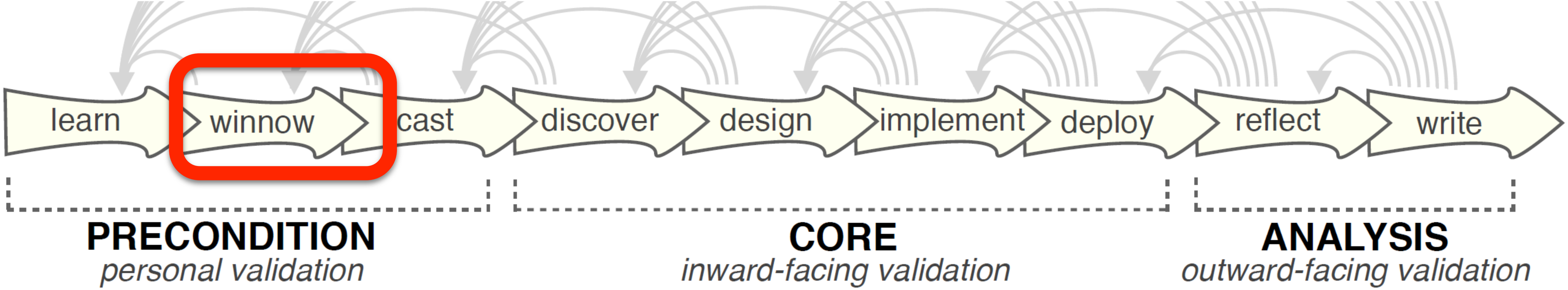
- insufficient knowledge of literature

Pitfalls



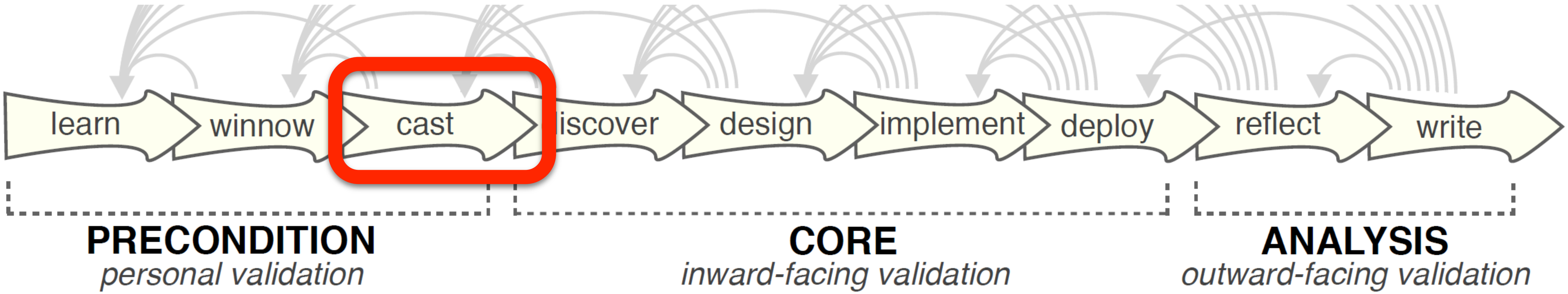
- collaboration with the wrong people
- **no real data available**
- insufficient time available from collaborators
- **no need for visualization: automate**
- no need for research: engineering project

Pitfalls



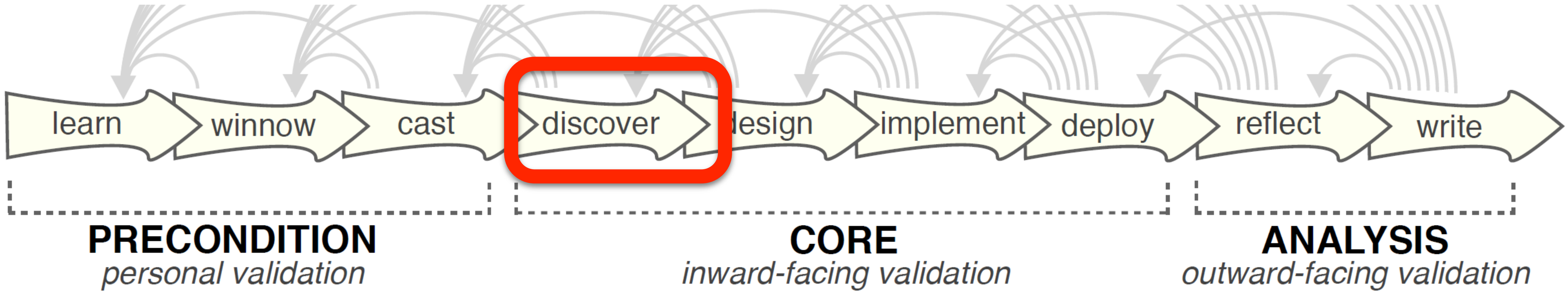
- **is this interesting to me?**
- existing tools are good enough
- **not an important/recurring task**
- no rapport with collaborators

Pitfalls



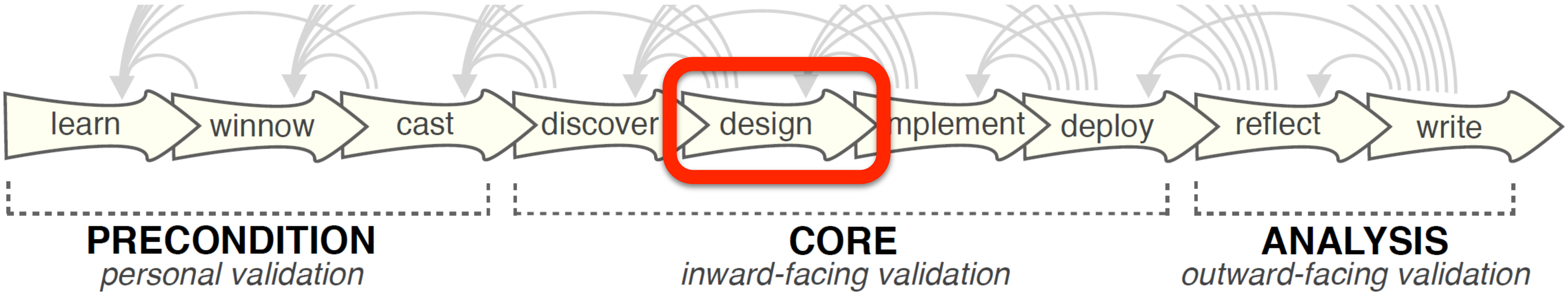
- **not identifying front-line analyst and gatekeeper**
- assuming same role distribution across projects
- mistaking tool-builders for real end users

Pitfalls



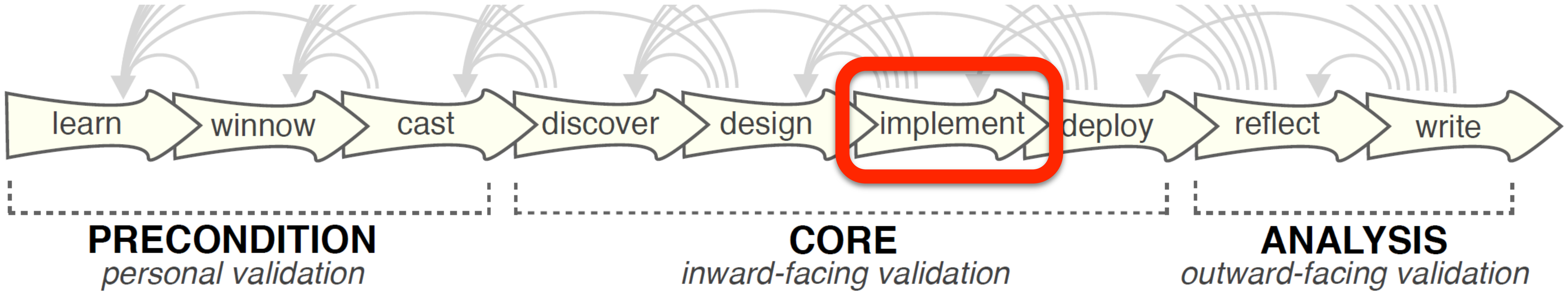
- **ignoring practices that currently work well**
- **expecting *just talking* or *fly on the wall* to work**
- **domain experts design the visualizations**
- too much/too little domain knowledge

Pitfalls



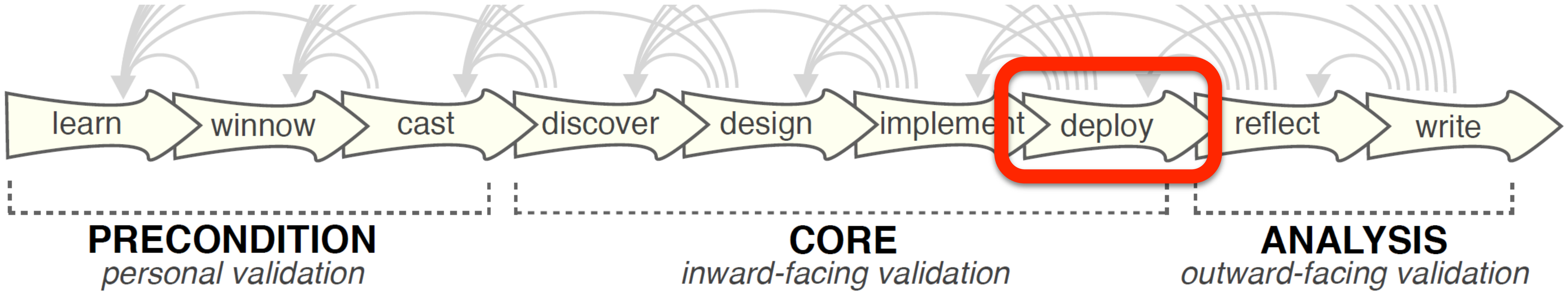
- too little abstraction
- **design consideration space too small**
- mistaking technique-driven and problem-driven work

Pitfalls



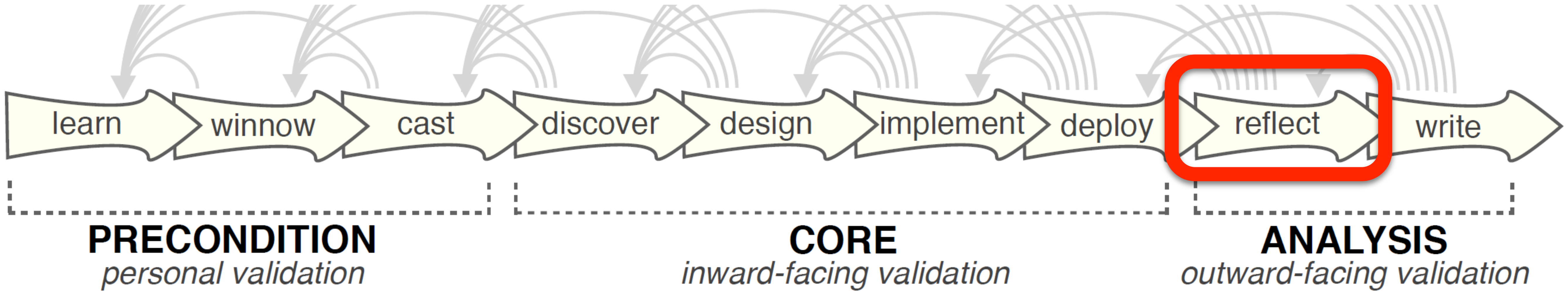
- **non-rapid prototyping**
- **usability: too little/too much**

Pitfalls



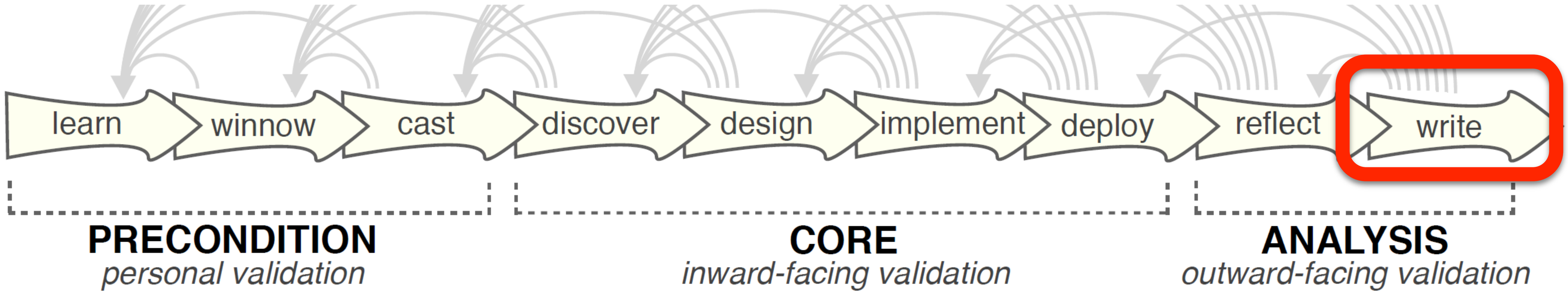
- **insufficient deploy time**
- non-real task/data/user
- *liking* a tool is not validation!

Pitfalls



- failing to improve guidelines

Pitfalls



- **not enough writing time**
- **no technique contribution ≠ write a design study**
- too much domain background
- chronological story vs concentrating on results
- **premature end to the project**

Additional reading

- Design study methodology: Reflections from the trenches and the stacks. Michael Sedlmair, Mariah Meyer, and Tamara Munzner. IEEE Trans. Visualization and Computer Graphics 18(12):2431-2440, 2012.

- Cluster and Calendar based Visualization of Time Series Data. Jarke J. van Wijk and Edward R. van Selow. Proc. InfoVis 1999, p 4-9.

Evaluating Information Visualisations

Sources

- **Evaluating Information Visualizations.** Sheelagh Carpendale. Chapter in *Information Visualization: Human-Centered Issues and Perspectives*, Springer LNCS 4950, 2008, p 19-45.
- Tamara Munzner's Course Slides on Evaluation

Psychophysics

- method of limits
  - find limitations of human perceptions
- error detection methods
  - find threshold of performance degradation
  - staircase procedure to nd threshold faster
- method of adjustment
  - find optimal level of stimuli by letting subjects control the level

Cognitive Psychology

- repeating simple, but important tasks, and measure reaction time or error
  - Miller's 7+/- 2 short-term memory experiments
  - Fitts' Law (target selection)
  - Hick's Law (decision making given n choices)
- interference between channels
- multi-modal studies
  - MacLean 2005, Perceiving Ordinal Data Haptically Under Workload
- using haptic feedback for interruption when the participants were visually (and cognitively) busy

Structural Analysis

- requirement analysis, task analysis
- structured interviews
  - can be used almost anywhere, for open-ended questions and answers
- rating/Likert scales
  - commonly used to solicit subjective feedback
  - ex: NASA-TLX (Task Load Index) to assess mental workload
    - “it is frustrating to use the interface”
      Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree

Comparative User Studies

- study design: factors and levels
- factors
  - independent variables
  - ex: interface, task, participant demographics
- levels
  - number of values in each factor
  - limited by length of study and number of participants

Comparative User Studies

- study design: within, or between?
- within
  - everybody does all the conditions
  - can lead to ordering effects
  - can account for individual differences and reduce noise
  - thus can be more powerful and require fewer participants
  - combinatorial explosion
    - severe limits on number of conditions
  - possible workaround is multiple sessions
- between
  - divide participants into groups
- each group does only some conditions

Comparative User Studies

- measurements (dependent variables)
  - performance indicators: task completion time, error rates, mouse movement
  - subjective participant feedback: satisfaction ratings, closed-ended questions, interview
  - observations: behaviors, signs of frustration
- number of participants
  - depends on effect size and study design: power of experiment
- possible confounds?
  - learning effect: did everybody use interfaces in a certain order?
- if so, are people faster because they are more practiced, or because of true interface effect?

Comparative User Studies

- result analysis
  - should know how to analyze the main results/hypotheses BEFORE study
  - hypothesis testing analysis (using ANOVA or t-tests) tests how likely observed differences between groups are due to chance alone
  - ex: a p-value of 0.05 means there is a 5% probability the difference occurred by chance
- usually good enough for HCI studies
- pilots!
  - should have good idea of forthcoming results of the study BEFORE running actual study trials

Evaluation Throughout Design Cycle

- user/task centered design cycle
  - initial assessments
  - iterative design process
  - benchmarking
  - deployment
- identify problems, go back to previous step

Task-Centered User Interface Design, Clayton Lewis and John Rieman, Chapters 0-5.

Initial Assessments

- what kind of problems are the system aiming to address?
  - analyze a large and complex dataset
- who are your target users?
  - data analysts
- what are the tasks? what are the goals?
  - find trends and patterns in the data via exploratory analysis
- what are their current practices
  - statistical analysis
- why and how can visualization be useful?
  - visual spotting of trends and patterns
- talk to the users, and observe what they do
- task analysis

Iterative Design Process

- does your design address the users' needs?
- can they use it?
- where are the usability problems?
- evaluate without users
  - cognitive walkthrough
  - action analysis
  - heuristics analysis
- evaluate with users
  - usability evaluations (think-aloud)
- bottom-line measurements

Benchmarking

- how does your system compare to existing ones?
- empirical, comparative studies
  - ask specific questions
  - compare an aspect of the system with specific tasks
    - Amar/Stasko task taxonomy paper
  - quantitative, but limited
    - The Challenge of Information Visualization Evaluation, Catherine Plaisant, Proc. AVI 2004

Deployment

- how is the system used in the wild?
- how are people using it?
- does the system fit into existing work flow? environment?

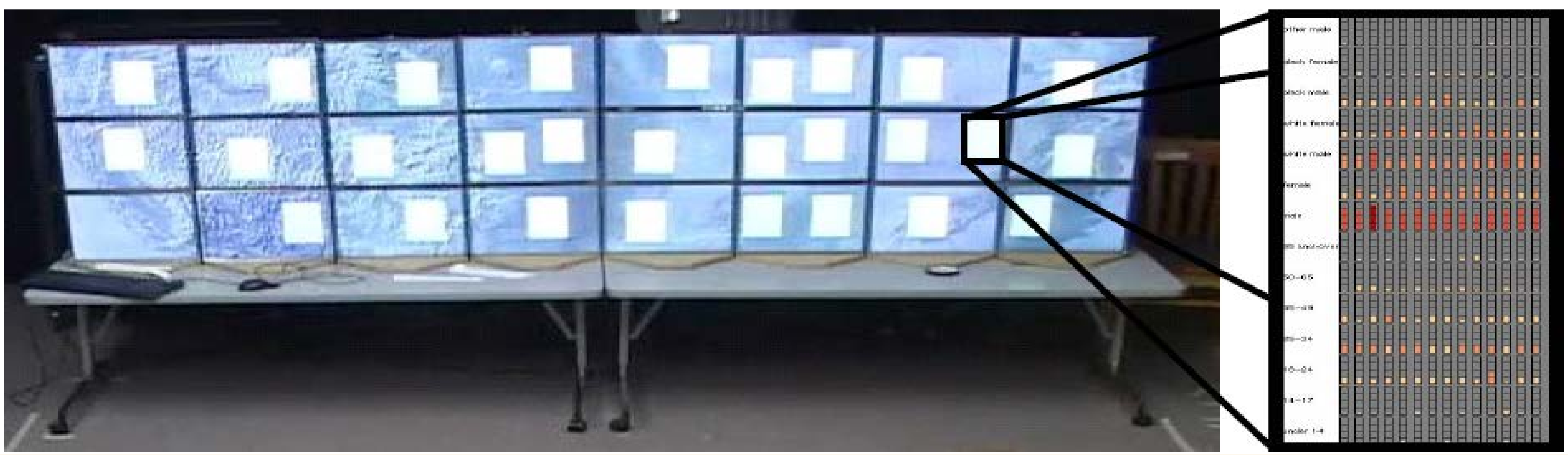
- contextual studies, field studies

Compare Systems vs. Characterize Usage

- user/task centered design cycle:
  - initial assessments
  - iterative design process
  - benchmarking: head-to-head comparison
  - deployment
  - (identify problems, go back to previous step)
- understanding/characterizing techniques
  - tease apart factors
  - when and how is technique appropriate
- line is blurry: intent

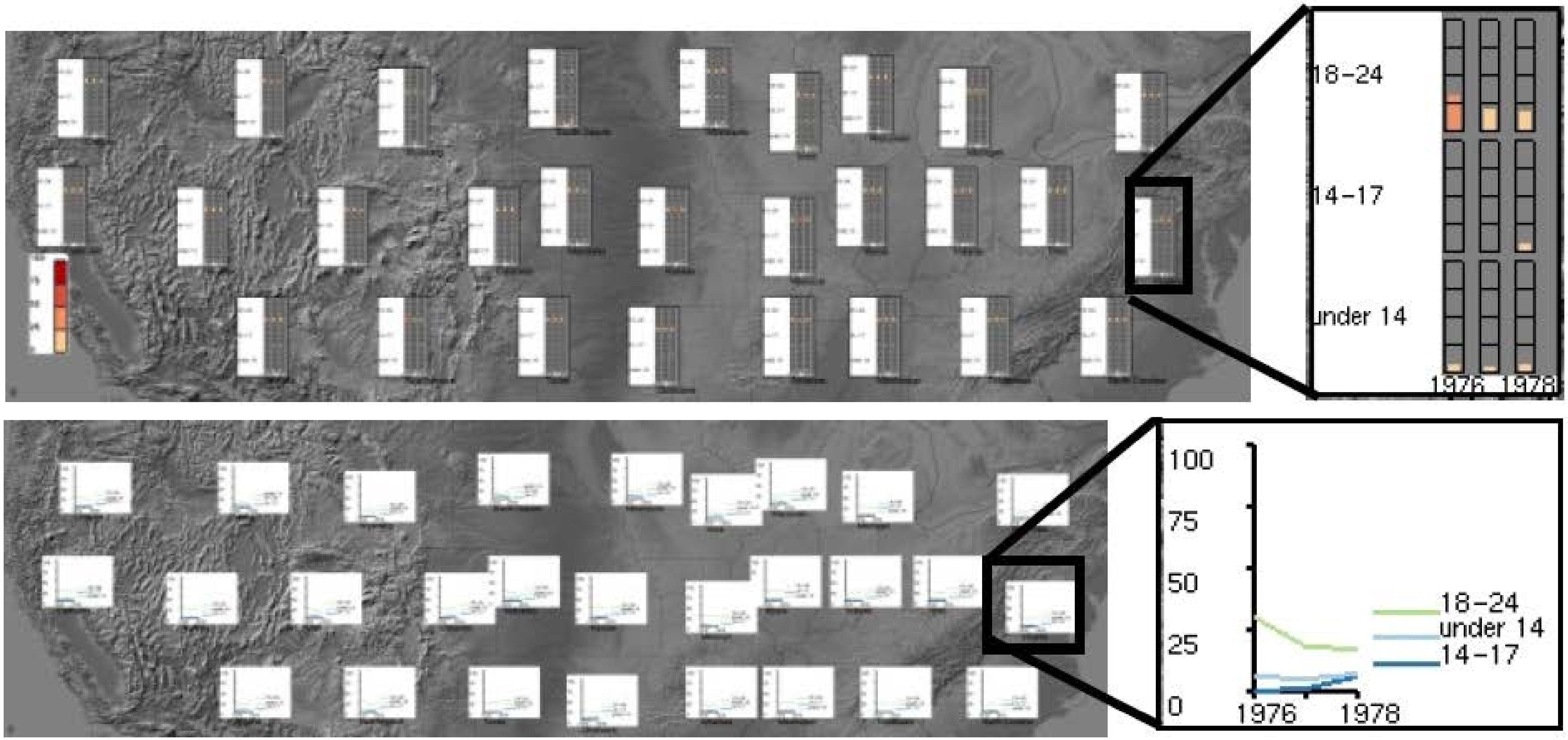
Perceptual Scalability

- what are perceptual/cognitive limits when screen-space constraints lifted?
  - 2 vs. 32 M pixel display
  - macro/micro views
- perceptually scalable
  - no increase in task completion times when normalize to amount of data



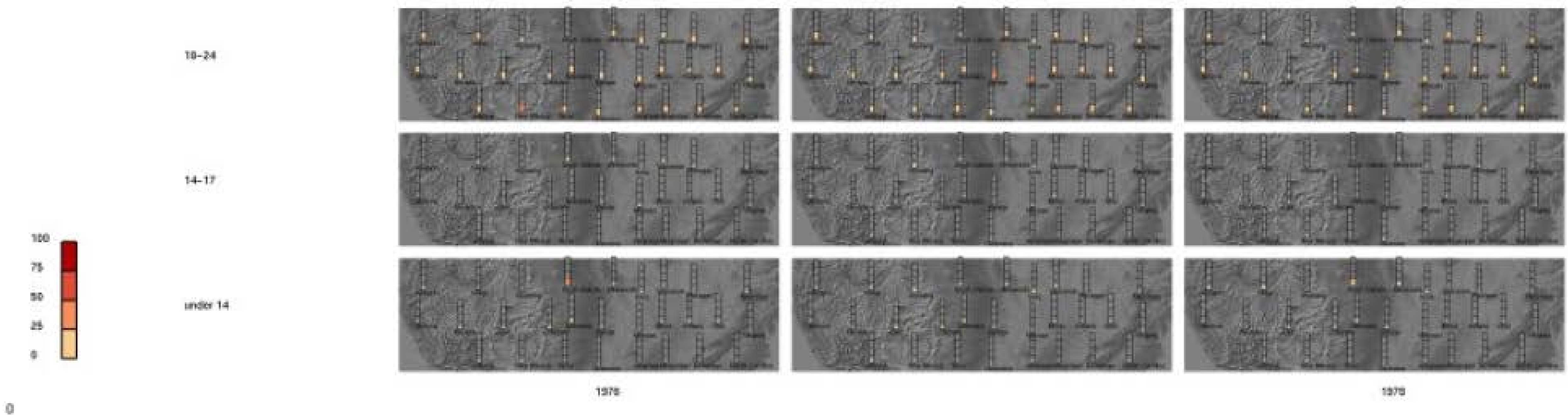
[**The Perceptual Scalability of Visualization.** Beth Yost and Chris North. IEEE TVCG 12(5) (Proc. InfoVis 06), Sep 2006, p 837-844.]

Embedded Visualizations



[**The Perceptual Scalability of Visualization.** Beth Yost and Chris North. IEEE TVCG 12(5) (Proc. InfoVis 06), Sep 2006, p 837-844.]

Small Multiples Visualization



- Attribute-centric instead of space-centric

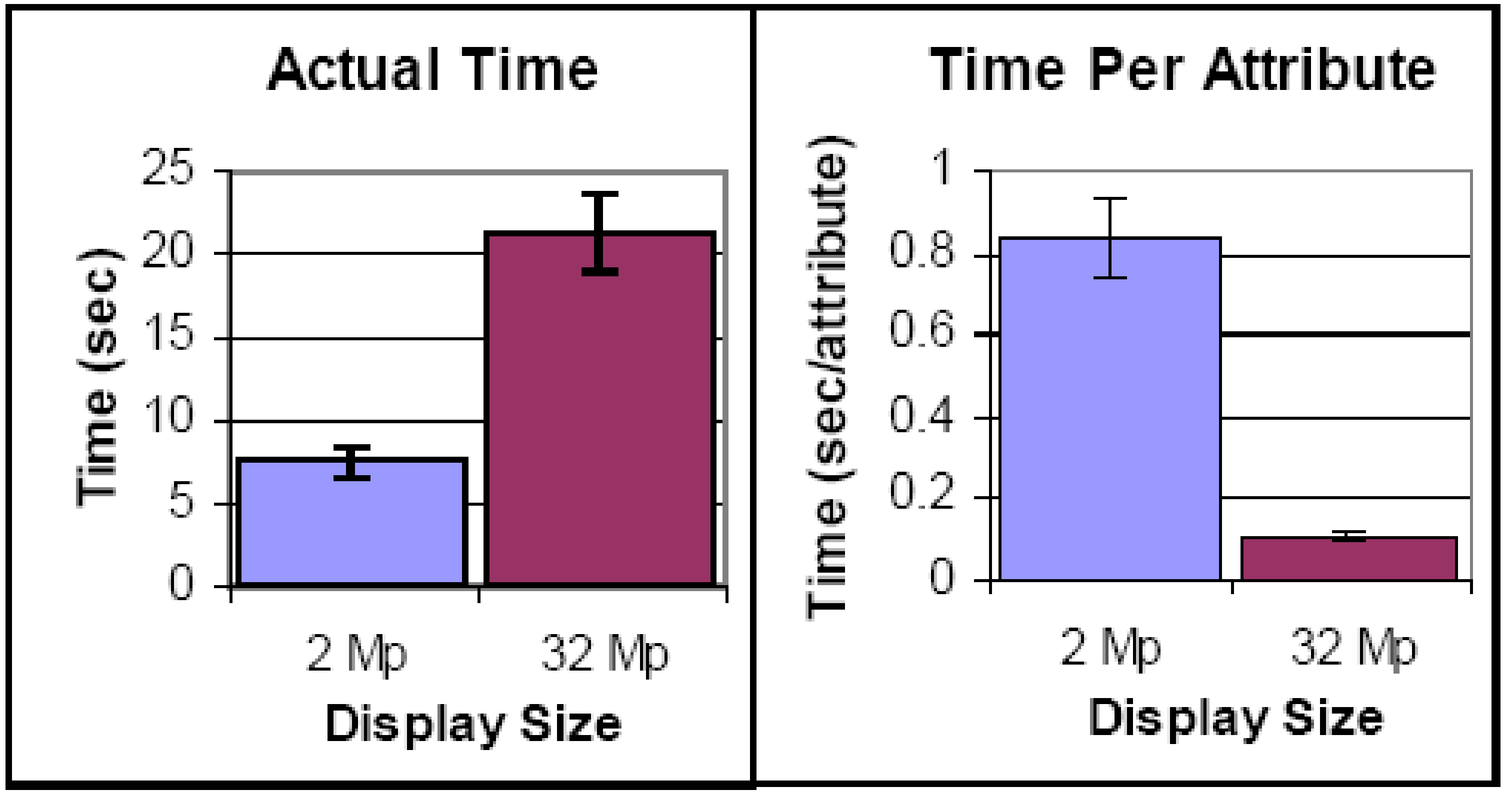
[**The Perceptual Scalability of Visualization.** Beth Yost and Chris North. IEEE TVCG 12(5) (Proc. InfoVis 06), Sep 2006, p 837-844.]

Perceptual Scalability

- design
  - 2 display sizes, between-subjects
    - (data size also increased proportionally)
  - 3 visualization designs, within
    - small multiples: bars
    - embedded graphs
    - embedded bars
  - 7 tasks, within
  - 42 tasks per participant
    - 3 vis x 7 tasks x 2 trials

[**The Perceptual Scalability of Visualization.** Beth Yost and Chris North. IEEE TVCG 12(5) (Proc. InfoVis 06), Sep 2006, p 837-844.]

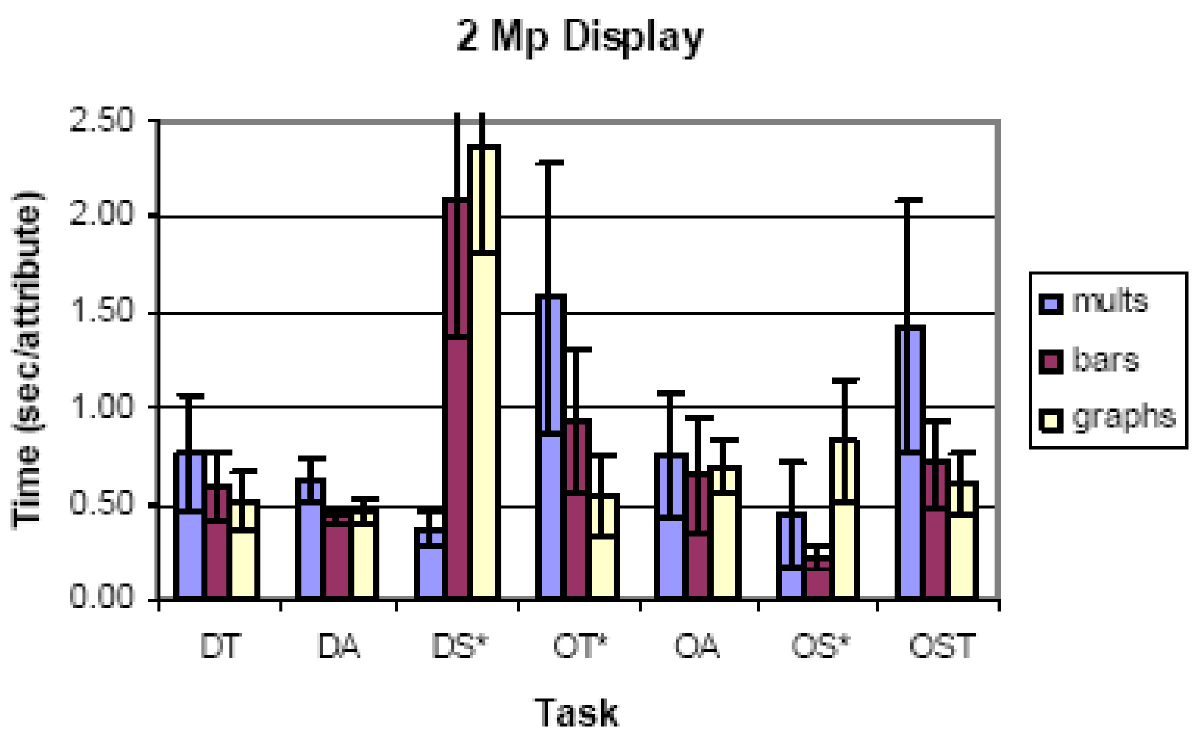
Results

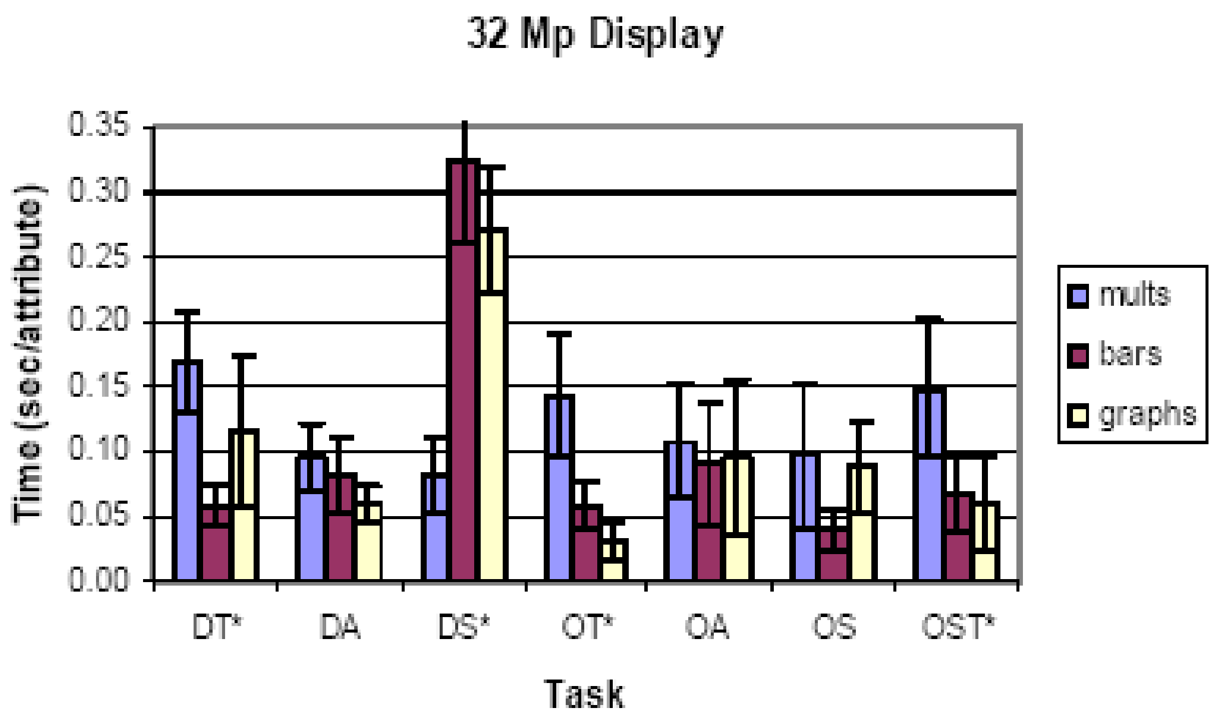


- 20x increase in data, but only 3x increase in absolute task times

[**The Perceptual Scalability of Visualization.** Beth Yost and Chris North. IEEE TVCG 12(5) (Proc. InfoVis 06), Sep 2006, p 837-844.]

Results





- 20x increase in data,
  but only 3x increase in absolute task times

[**The Perceptual Scalability of Visualization.** Beth Yost and Chris North. IEEE TVCG 12(5) (Proc. InfoVis 06), Sep 2006, p 837-844.]

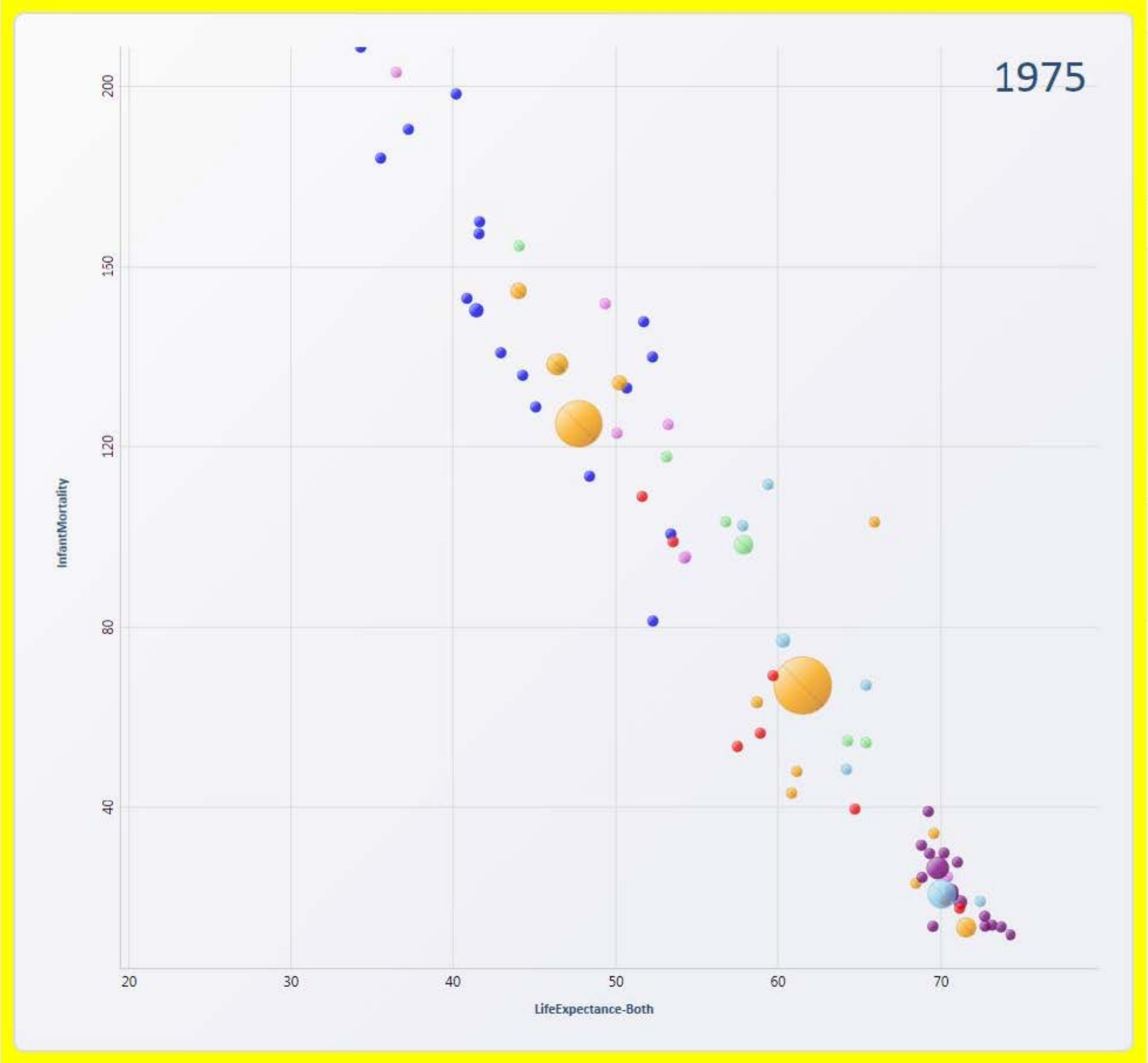
Results

- visual encoding important on small displays
  - DS: mults sig slower than graphs on small
  - DS: mults sig slower than embedded on large
  - OS: bars sig faster than graphs for small
  - OS: no sig dierence bars/graphs for large
- spatial grouping important on large displays
  - embedded sig faster+preferred over small mult
  - no bar/graph differences

[**The Perceptual Scalability of Visualization.** Beth Yost and Chris North. IEEE TVCG 12(5) (Proc. InfoVis 06), Sep 2006, p 837-844.]

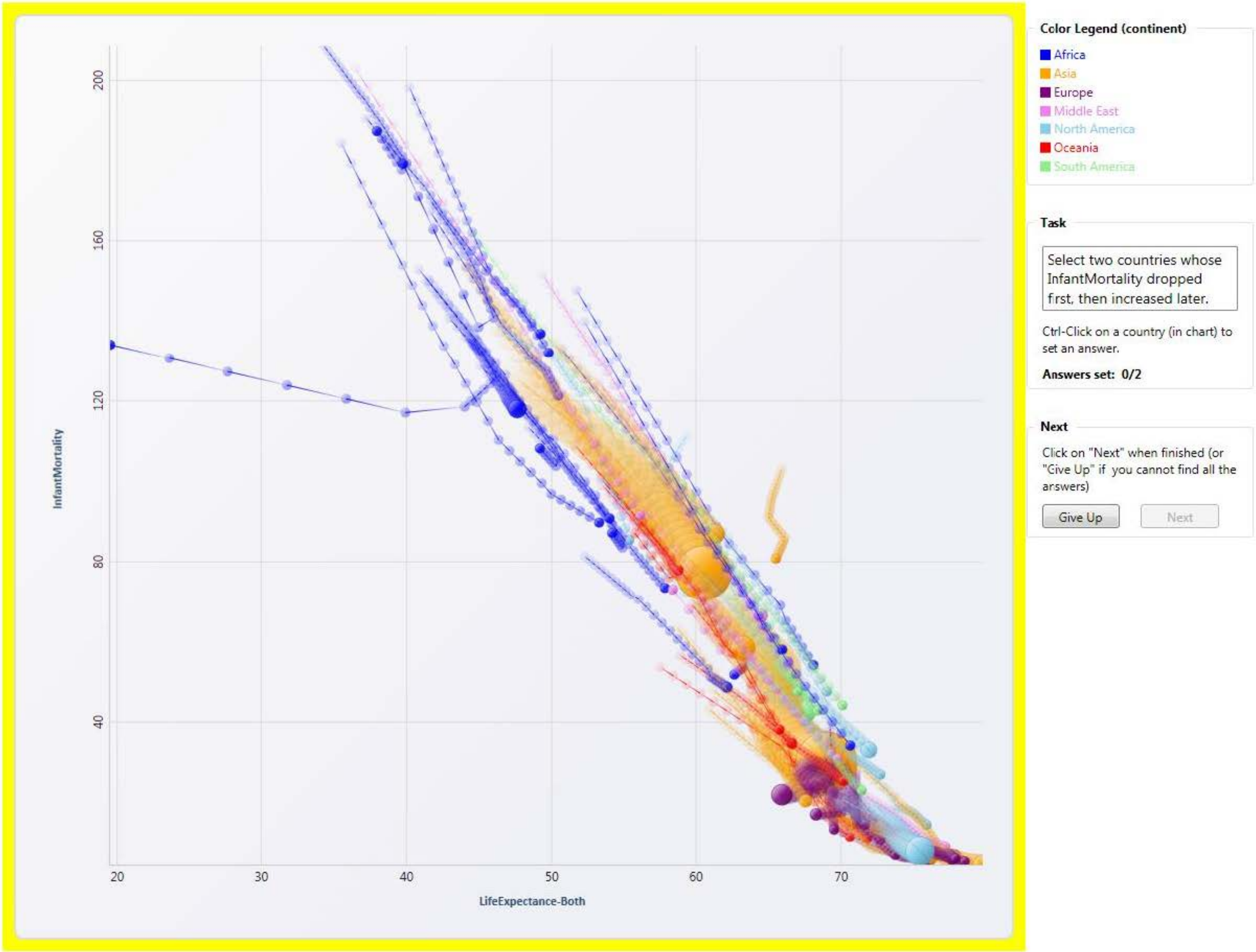
Trends: Animation, Trails, Small Multiples

- Gapminder: animated bubble charts + human
  - x/y position, size, color, animation
  - is animation effective?
    - presentation vs analysis
    - trend vs transitions



[**Effectiveness of Animation in Trend Visualization.** Robertson et al. IEEE TVCG (Proc. InfoVis 2008). 14(6): 1325-1332 (2008)]

Trends



- many countertrends lost in clutter

[**Effectiveness of Animation in Trend Visualization.** Robertson et al. IEEE TVCG (Proc. InfoVis 2008). 14(6): 1325-1332 (2008)]

Small Multiples

- individual plots get small

[**Effectiveness of Animation in Trend Visualization.** Robertson et al. IEEE TVCG (Proc. InfoVis 2008). 14(6): 1325-1332 (2008)]

Design

- 2 use: presentation vs. analysis (between-subjects)
  - 3 vis encodings: animation vs. traces vs. small mults
- 2 dataset size: small vs. large
  - 3 encoding x 2 size: within-subjects
- 24 tasks per participant
  - 4 tasks x 3 encodings x 2 sizes

[**Effectiveness of Animation in Trend Visualization.** Robertson et al. IEEE TVCG (Proc. InfoVis 2008). 14(6): 1325-1332 (2008)]

Results

- small multiples more accurate than animation
- animation faster for presentation, slower for analysis than small multiples and trends
- dataset size matters (unsurprisingly)

[**Effectiveness of Animation in Trend Visualization.** Robertson et al. IEEE TVCG (Proc. InfoVis 2008). 14(6): 1325-1332 (2008)]

User Study Goals

- compare systems
- characterize methods
- formative feedback
- summative judgement
- convince stakeholders

Thank you for your attention!

Notes from S. Carpendale's *Evaluating Information Visualizations*

Supplementary Slides

Notes from the paper

- current evaluations not convincing enough to encourage widespread adoption of information visualization tools
  - Small datasets
  - University participants only
  - Simple tasks
- Rather: real users & tasks, complex datasets

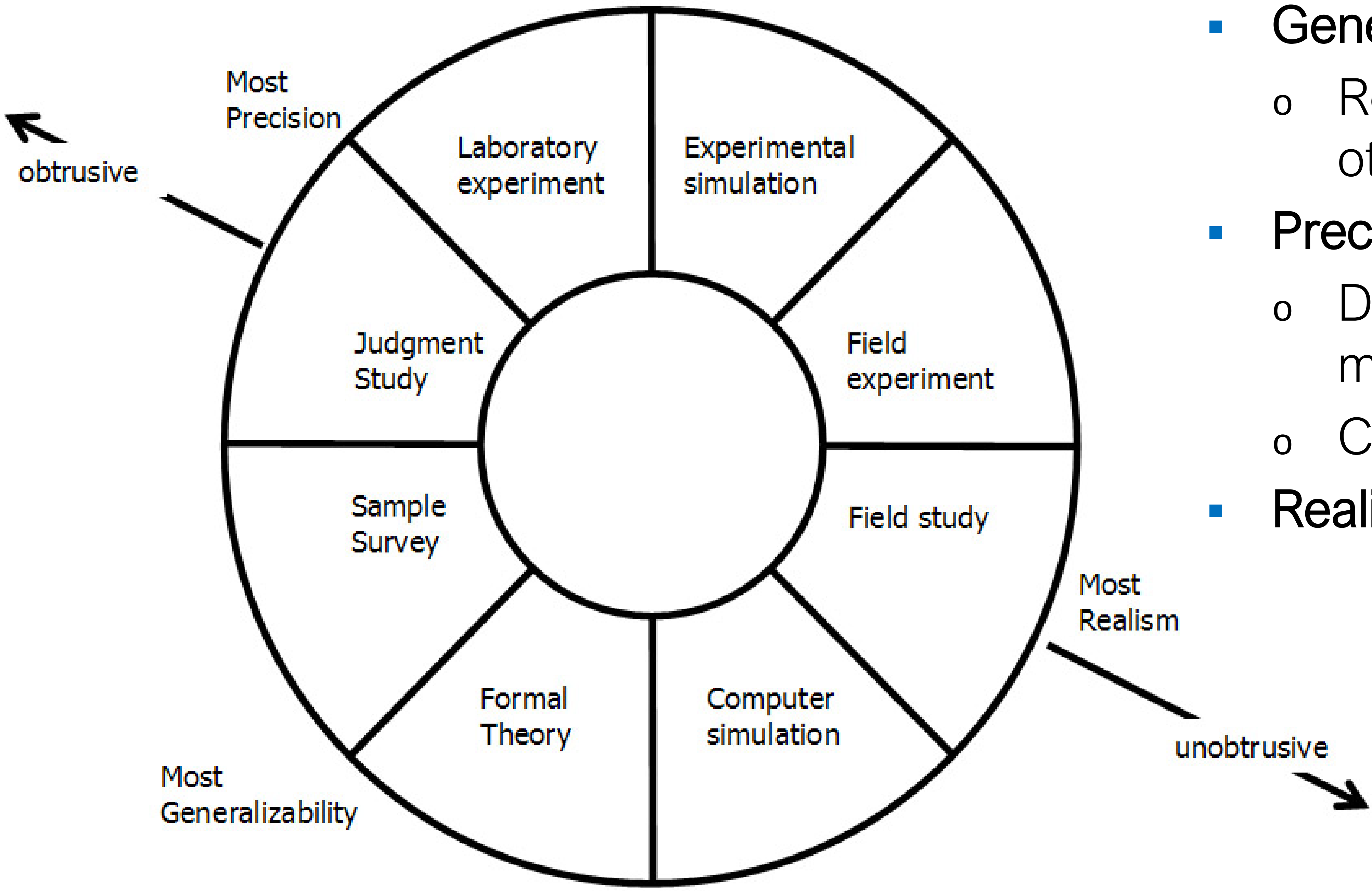
Challenges of Evaluations

- Difficult to pick right **focus** and ask the right **questions**
- Choose right **methodology**, sufficiently rigorous in procedure and data collection
- Info vis research relates to other empirical research
  - **Human computer interaction** (HCI)
    - Many tasks are interface interaction
    - Usability aspects: access to visual representation and underlying dataset
    - Appropriate sample of participants (domain experts)
    - Results due to particular technique or overall solution?
    - Participants familiarity with existing approaches
  - **Perceptual psychology**
    - Appropriate representational encoding, readability of visuals
  - **Cognitive reasoning -** Tasks vary with data type
    - Low- level: compare, associate, rank, cluster, correlate, categorize
    - High-level: *understand* data trends, uncertainties, causal relationships, predicting the future, or learning a domain

Challenge: Does vis promote insight into the data?

- Discover the unexpected
  - Often long-term, on-going
  - “answering questions you didn’t know you had”
  - Depends on participant – motivation, knowledge, experience

Choose an Evaluation Approach



- **Generalizability**
  - Results apply to other people?
- **Precision**
  - Definite measurements
  - Controlled factors
- **Realism**

[Carpendale ‘08, adapted from McGrath *Methodology Matters*]

Approaches glossary

- Field study
  - conducted in actual situation, unobtrusive observer
- Field experiment
  - ask participants to perform specific task
  - more precision, less realism
- Lab experiment
  - may add ‘thinkaloud’ protocol
- Experimental Simulation
  - reintroduce some realism via sim, avoid risky or unethical situations
- Judgement study
  - person’s response to a set of stimuli under ‘neutral conditions’
  - e.g. what surface texture (or shading) is better to interpret shape?
- Sample survey
  - discover relationships between variables in a given population
- Formal theory
  - possibly meta-study to further infovis theory
- Computer simulation
  - May not involve participants

Quantitative Methodology

- Hypothesis development
  - Precise questions of broad interest
- Identification of independent vars
  - Factors that may affect hypothesis, ideally few
- Control of independent vars
  - How indep. vars are changed, experimental design
- Elimination of complexity
  - Controlled environment
- Measurement of dependent vars
  - Common metrics: speed, accuracy, error rate, satisfaction
- Statistical analysis

Qualitative Methods

- Observation techniques
  - Take notes unobtrusively, maybe during break
  - Note overt and covert in activities and communications
  - Use only one side of the note-paper
- Interview techniques
  - Make sure you understand
  - Limit your inclination to talk
  - Listen for and encourage the less formal, less guarded expression of participant’s thoughts
  - Avoid leading Q’s, ask open Q’s, ask for concrete details
  - Humanity of interview, be present, aware, sensitive

Qualitative Methodologies

- **Nested qualitative methods**, part of quantitative study
  - Log Experimenter Observations during qualitative study
  - Think-aloud protocol (may be unnatural to participants)
  - Collect participants' opinions

Qualitative methodologies (cont'd)

- **Inspection Evaluation Methods** conducted by experts
  - Usability heuristics, e.g.
    - visibility of system status
    - match between system and real world
    - personal control and freedom
    - consistency and standards
    - error prevention and handling, help and documentation
    - recognition rather than recall
    - aesthetic and minimalist design
  - Collaboration heuristics
    - communication and coordination
    - distributed or co-located

Primarily qualitative study

- In situ observational study
  - "fly on the wall"
- Participatory observation
  - Collaborate with domain experts
- Laboratory observational study
- Contextual interview
  - Ask about task, setting, or application of interest

- Lessen the task and data comprehension divide between visualization experts and the domain experts

Challenges of qualitative study

- Sample size
- Subjectivity
- Analysing qualitative data

"Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted" – Albert Einstein