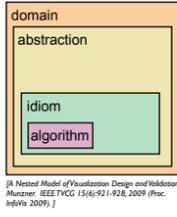


# Problem-Driven Visualization Through Design Studies

**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>



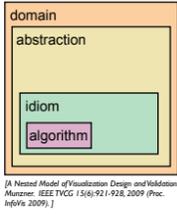
## Nested model: Four levels of visualization concerns



[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

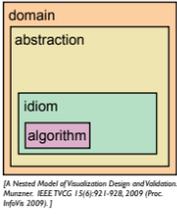
- **domain** situation
  - **who** are the target users?



[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

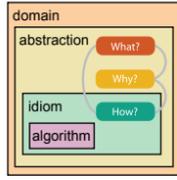
- **domain** situation
  - **who** are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis



[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

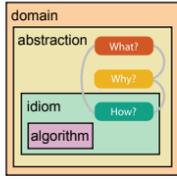
- **domain** situation
  - **who** are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**



[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. Behmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

## Nested model: Four levels of visualization concerns

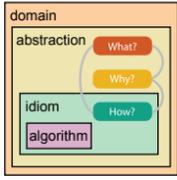
- **domain** situation
  - **who** are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form



[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. Behmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

## Nested model: Four levels of visualization concerns

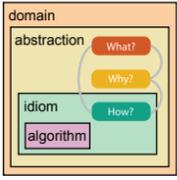
- **domain** situation
  - **who** are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**



[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. Behmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

## Nested model: Four levels of visualization concerns

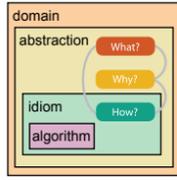
- **domain** situation
  - **who** are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- **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. Behmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

## Nested model: Four levels of visualization concerns

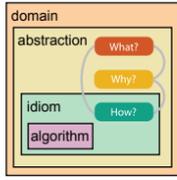
- **domain** situation
  - **who** are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- **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. Behmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]

## Nested model: Four levels of visualization concerns

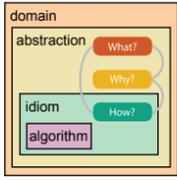
- **domain** situation
  - **who** are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- **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. Behmer 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. Behmer 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

## Why is validation difficult?

- different ways to get it wrong at each level

**Domain situation**  
 You misunderstood their needs

**Data/task abstraction**  
 You're showing them the wrong thing

[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

**Data/task abstraction**  
 You're showing them the wrong thing

**Visual encoding/interaction idiom**  
 The way you show it doesn't work

[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

**Data/task abstraction**  
 You're showing them the wrong thing

**Visual encoding/interaction idiom**  
 The way you show it doesn't work

**Algorithm**  
 Your code is too slow

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

[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).]

17

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).]

18

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).]

19

Validation solution: use methods from appropriate fields at each level

design

Visual encoding/interaction idiom  
Justify design with respect to alternatives

computer science

Algorithm  
Measure system time/memory  
Analyze computational complexity

technique-driven work

cognitive psychology

Analyze results qualitatively  
Measure human time with lab experiment (lab study)

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

20

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

technique-driven work

cognitive psychology

Analyze results qualitatively  
Measure human time with lab experiment (lab study)

anthropology/ethnography

Observe target users after deployment (field study)  
Measure adoption

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

21

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

technique-driven work

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

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

22

Validation solution: use methods from appropriate fields at each level

- avoid mismatches between level and validation

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

technique-driven work

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

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

23

Visualization: Angles of attack

problem-driven work

24

Visualization: Angles of attack

problem-driven work ↔ technique-driven work

25

Visualization: Angles of attack

problem-driven work ↔ technique-driven work

↙ ↘

evaluation

↖ ↗

26

Visualization: Angles of attack

problem-driven work ↔ technique-driven work

↙ ↘

theoretical foundations

↕

evaluation

↖ ↗

27

Visualization: Angles of attack

problem-driven work ↔ technique-driven work

↙ ↘

theoretical foundations

↕

evaluation

↖ ↗

28

Problem-driven visualization: Design studies

29

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...”

30

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...”

31

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).]

32

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

## 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).]

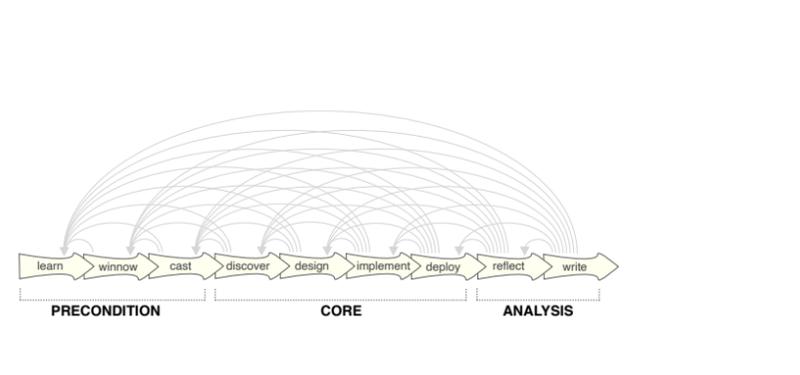
## Design study methodology: definitions

[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).]

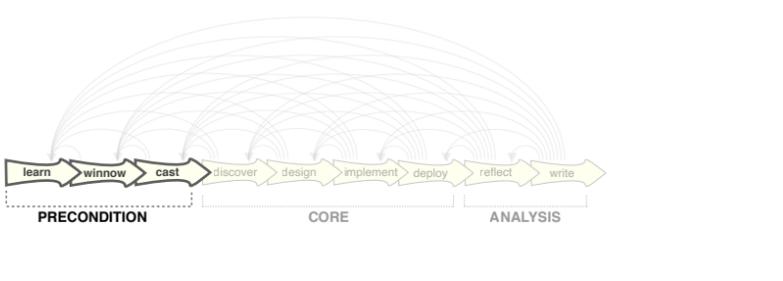
## 9-stage framework



[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

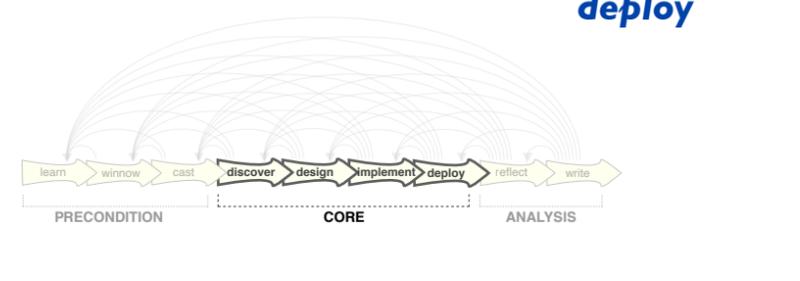
**learn  
winnow  
cast**



[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

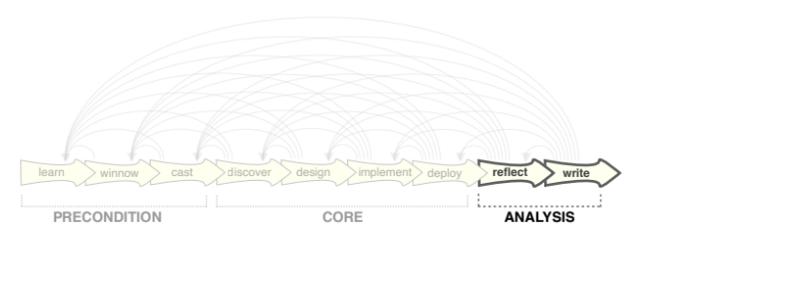
**discover  
design  
implement  
deploy**



[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

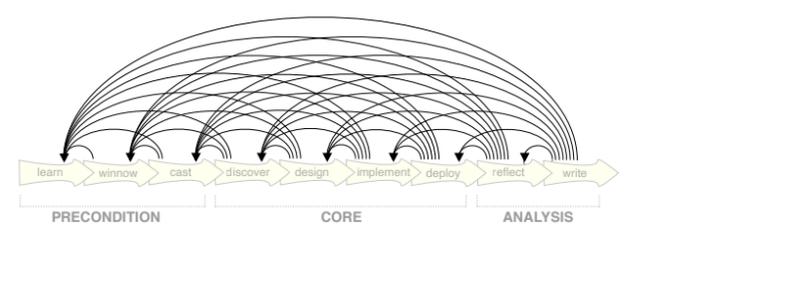
**reflect  
write**



[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

## 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-21	mistaking technique-driven for problem-driven work	design
PF-2	premature start: insufficient knowledge of vis literature	learn	PF-22	nonrapid prototyping	implement
PF-3	premature commitment: collaboration with wrong people	winnow	PF-23	usability: too little / too much	implement
PF-4	no real data available (yet)	winnow	PF-24	premature end: insufficient deploy time built into schedule	deploy
PF-5	insufficient time available from potential collaborators	winnow	PF-25	usage study not case study: non-real task/data/user	deploy
PF-6	no need for visualization: problem can be automated	winnow	PF-26	liking necessary but not sufficient for validation	deploy
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			

[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

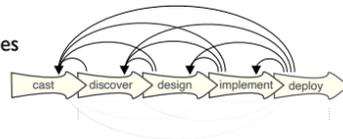


PF-1	premature advance: jumping forward over stages	general	PF-21	mistaking technique-driven for problem-driven work	design
PF-2	premature start: insufficient knowledge of vis literature	learn	PF-22	nonrapid prototyping	implement
PF-3	premature commitment: collaboration with wrong people	winnow	PF-23	usability: too little / too much	implement
PF-4	no real data available (yet)	winnow	PF-24	premature end: insufficient deploy time built into schedule	deploy
PF-5	insufficient time available from potential collaborators	winnow	PF-25	usage study not case study: non-real task/data/user	deploy
PF-6	no need for visualization: problem can be automated	winnow	PF-26	liking necessary but not sufficient for validation	deploy
PF-7	researcher expertise does not match domain problem	winnow	PF-27	failing to improve guidelines: confirm, refine, reject, propose	reflect
PF-8	no need for research: engineering vs. research project	winnow	PF-28	insufficient writing time built into schedule	write
PF-9	no need for change: existing tools are good enough	winnow	PF-29	no technique contribution / good design study	write
PF-10	no real/important/recurring task	winnow	PF-30	too much domain background in paper	write
PF-11	no rapport with collaborators	winnow	PF-31	story told chronologically vs. focus on final results	write
PF-12	not identifying front line analyst and gatekeeper before start	cast	PF-32	premature end: win race vs. practice music for debut	write
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			

[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

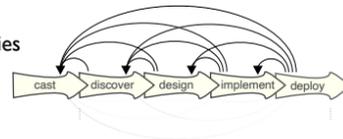
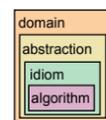
- user-centered design: well-known HCI methodology
  - iterative refinement & deployment
  - evaluation through case studies & field studies



49

## 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



50

## 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

52

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

54

## 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

56

## What is Clickstream Data?

## Data: Actions

58

## Data: Action Attributes

## Data: Action Types

60

## Action Hierarchy

## Action Hierarchy

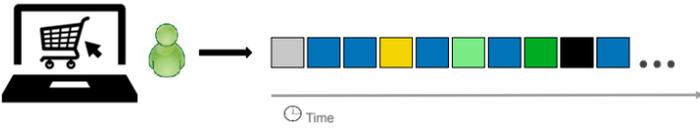
62

## Data: Sequences

## Data: Sequences

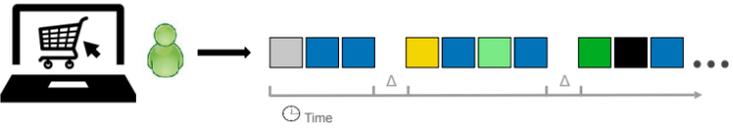
64

Data: Client Sequences



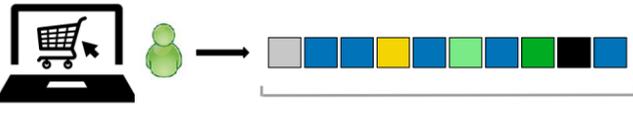
**Client Sequences:** all actions performed by a single user

Data: Session Sequences



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

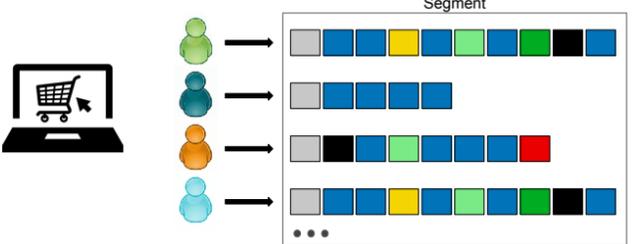
Data: Sequence Attributes



Start time   End time   Duration   Action Counts

- : 1
- : 1
- : 5
- : 1
- : 1

Data: Segments



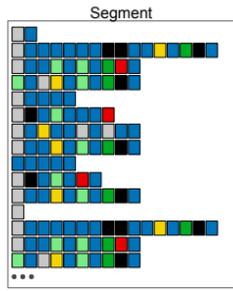
**Segment:** any set of sequences

Data: Segment Attributes

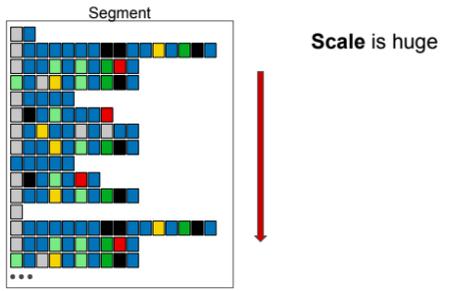
Segment

- Size**   Counts of sequences: Absolute, Relative
- Sequence Related**   Sequence Distributions: Start Time, Duration, Action Counts
- Action Related**   Action Distributions: Action Transitions: action before, action after

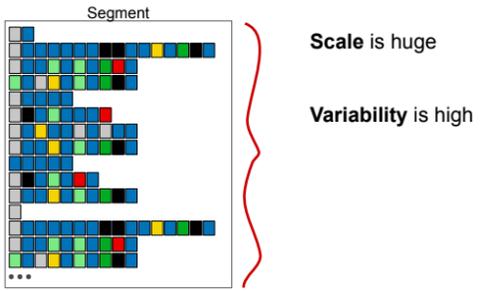
Real-world Clickstream Data



Real-world Clickstream Data



Real-world Clickstream Data



Real-world Clickstream Data



What are **Clickstream Data Analysis Tasks?**

Tasks: Segment Behavior

Segment

**Behavior:** set of attribute constraints

Behavior

- Viewed 4 pages
- Purchased
- Between 9 - 10 am

Tasks: Segment Behavior

Segment

**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

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

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

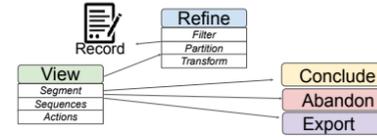
**Frequency:** Determine how many sequences are in the segment defined by behavior

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

# High-Level Segmentifier Analysis Model

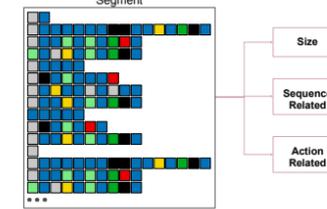
## High-Level Segmentifier Analysis Model

- 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



81

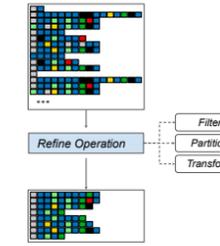
## High-Level Segmentifier Analysis Model



- 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*

82

## High-Level Segmentifier Analysis Model

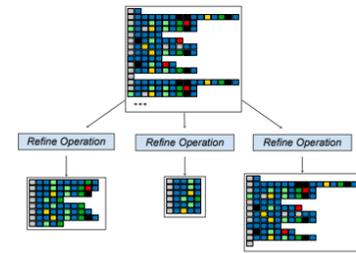


- Apply operation to create new segments
- Type of Refinements
  - Filter
  - Partition
  - Transform

83

84

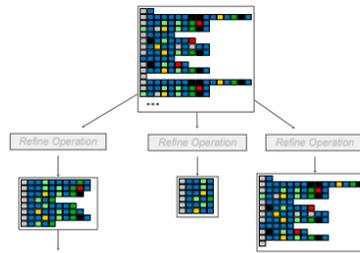
## High-Level Segmentifier Analysis Model



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

85

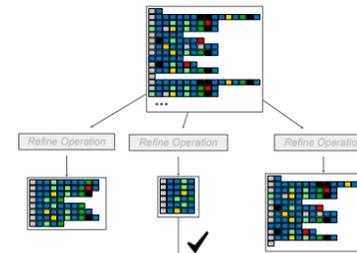
## High-Level Segmentifier Analysis Model



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

86

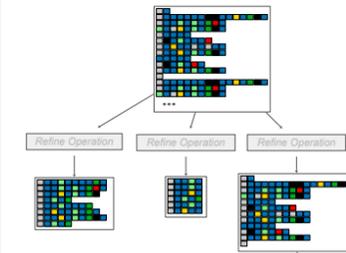
## High-Level Segmentifier Analysis Model



- Discover actionable insight by *viewing* segment

87

## High-Level Segmentifier Analysis Model



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

88

## Why Visual Analytics?



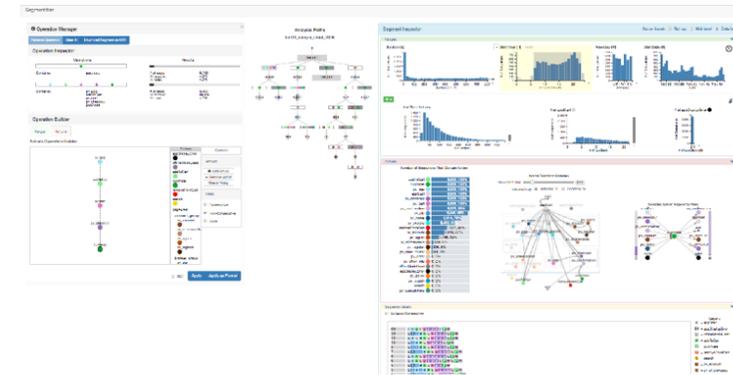
- Automation would be nice...
  - Put data in, actionable results appear
- ... but it is not realistic
  - Many possible questions, data-driven interplay between finding answers and generating new questions
- Human-in-the-loop visual data analysis
  - Integrate computing power of machine with intuition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

89

## Solution

90

## The Segmentifier Interface



91

## Video

### Segmentifier: Interactively Refining Clickstream Data into Actionable Segments

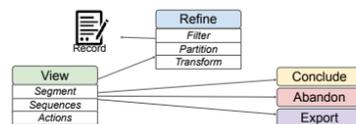


<https://www.youtube.com/watch?v=TobYDFeISQg&t=20s>

92

## Segmentifier Contributions

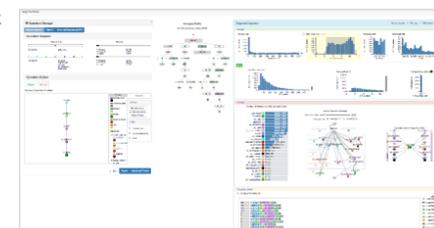
- Thorough characterization of task and data abstraction for clickstream data analysis



93

## Segmentifier Contributions

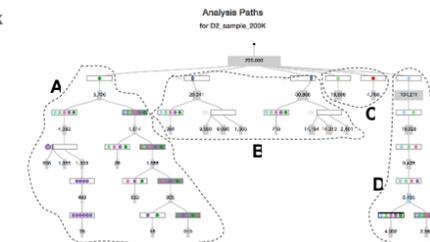
- Thorough characterization of task and data abstraction for clickstream data analysis
- Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis



94

## 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



95

## Three case studies of problem-driven work

- e-commerce



- facilities management



- biology



e-commerce by shashank singh from the Noun Project Biology by lezar tantular from the Noun Project Business by Colourswatch from the Noun Project

96



## 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.

# Ocupado

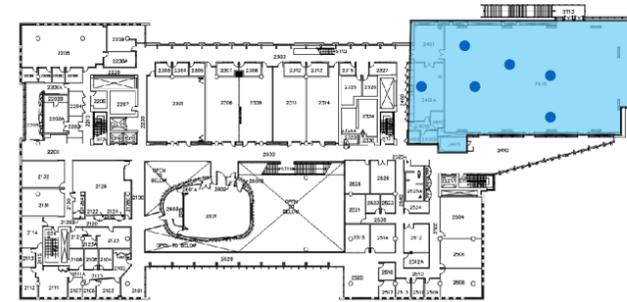
Visualizing Location-Based Counts Over Time Across Buildings

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

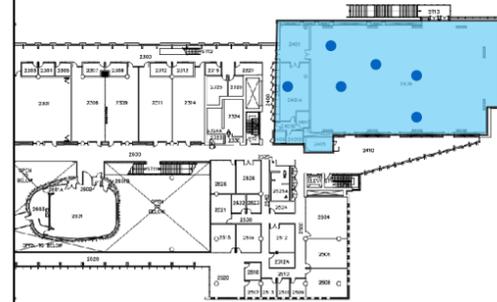
Ocupado: Visualizing Location-Based Counts Over Time Across Buildings.  
Oppermann and Munzner. Computer Graphics Forum (Proc. EuroVis 2020) 39(3):127-138 2020.



## WiFi Connections: Location-Based Counts



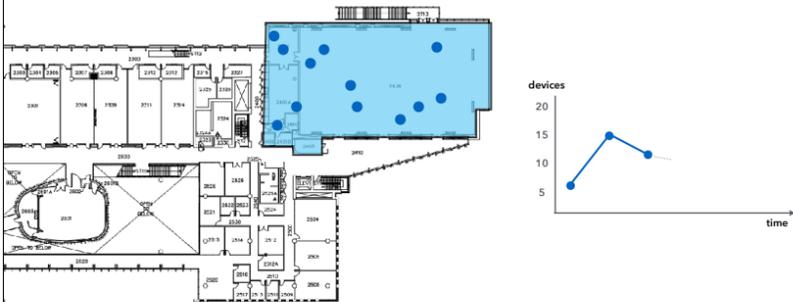
## WiFi Connections: Location-Based Counts



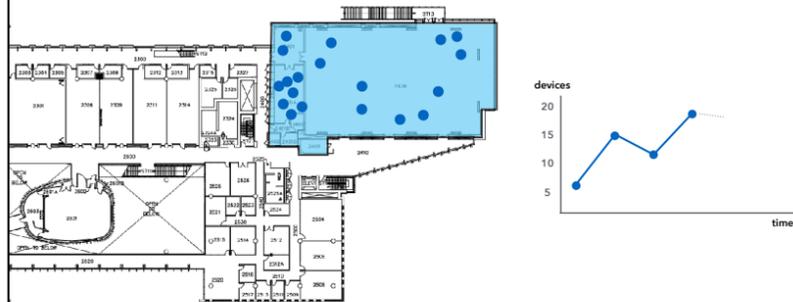
## WiFi Connections: Location-Based Counts



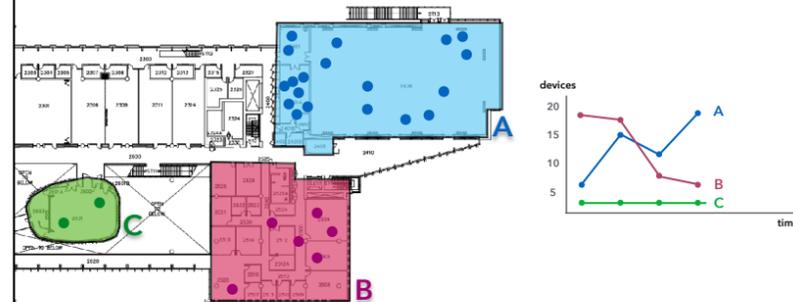
## WiFi Connections: Location-Based Counts



## 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

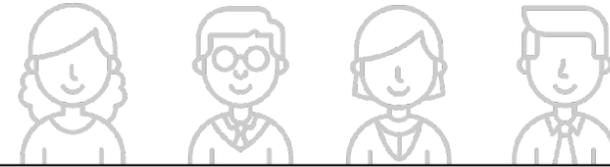
113



### WiFi connections as a proxy for occupancy

114

### Interviews with potential stakeholders



115

### Focus Domains

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

116

### Focus Domains

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



### Semi-structured discussions and live demos

117

### Tasks

- ✓ Confirm assumptions or previous observations.  
Do students occupy room x in evenings or on weekends?

118

### Tasks

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

119

### Tasks

- ✓ Confirm assumptions or previous observations.
- ✓ Monitor the current/recent utilization rate.
- ✓ Communicate space usage and justify decisions.  
Space usage improved after renovation.

120

### Tasks

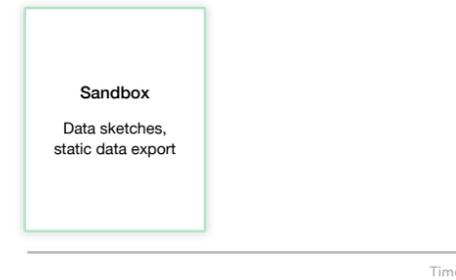
- ✓ Confirm assumptions or previous observations.
- ✓ Monitor the current/recent utilization rate.
- ✓ Communicate space usage and justify decisions.
- ✓ Validate the data (quality control).  
Check minimum size of a room that can be captured.

121

### Spatial and Temporal Data Granularities

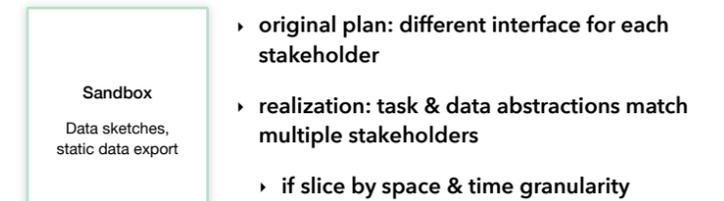
122

### Visualization Prototypes



123

### Visualization Prototypes



124

### Spatial and Temporal Data Granularities

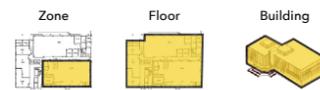
### Regions of interest



125

### Spatial and Temporal Data Granularities

### Regions of interest

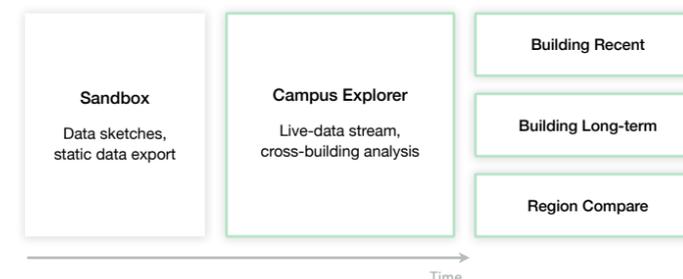


### Periods of interest

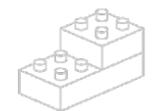


126

### Visualization Prototypes



127



### Reusable Visualization Components

128

### Reusable Visualization Components

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

129

### Reusable Visualization Components

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

130

### Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
		Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Juxtaposition	Repeating patterns, trends, outliers (non-contiguous)
		Aggregation	Typical utilization profiles

131

### Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
		Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Juxtaposition	Repeating patterns, trends, outliers (non-contiguous)
		Aggregation	Typical utilization profiles
		Superposition	Within-session patterns, outliers

132

### Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
		Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Juxtaposition	Repeating patterns, trends, outliers (non-contiguous)
		Aggregation	Typical utilization profiles
		Superposition	Within-session patterns, outliers

Temporal

133

### Reusable Visualization Components

Layout	Visual Encoding	Facet	Comparisons
		Juxtaposition	Repeating patterns, trends, outliers (contiguous)
		Juxtaposition	Repeating patterns, trends, outliers (non-contiguous)
		Aggregation	Typical utilization profiles
		Superposition	Within-session patterns, outliers
		Superposition	Within local spatial neighborhood
		Containment (nested)	Across distributed regions

Temporal

Spatial

134

### Ocupado Interfaces

135

- ### 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
- 136

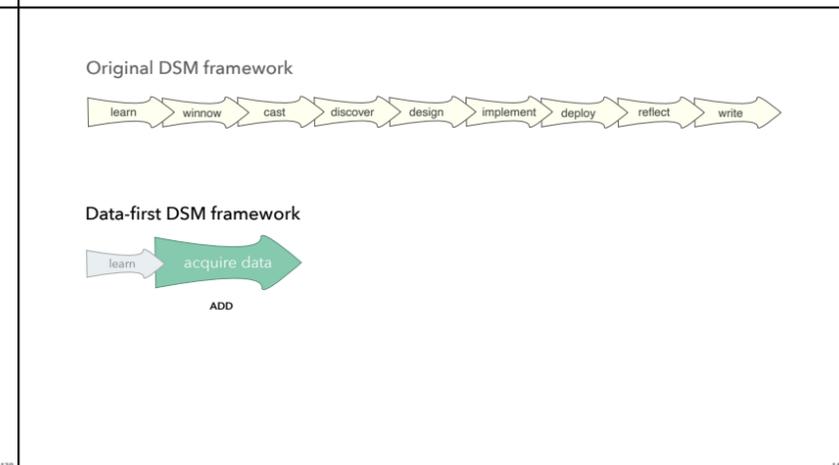
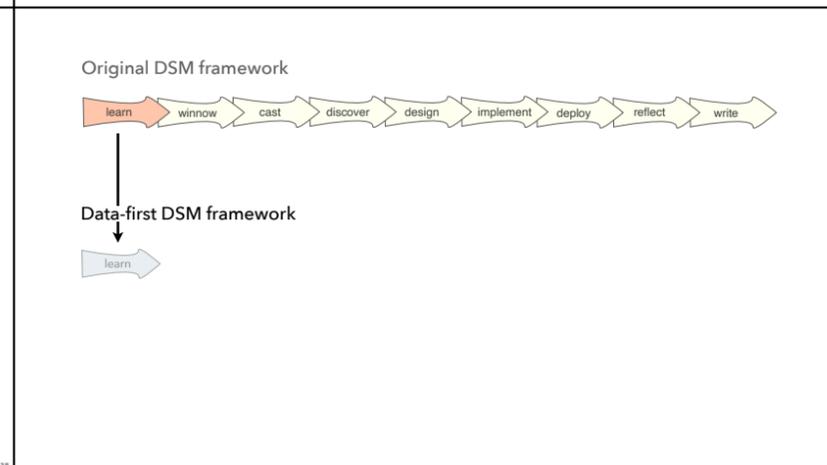
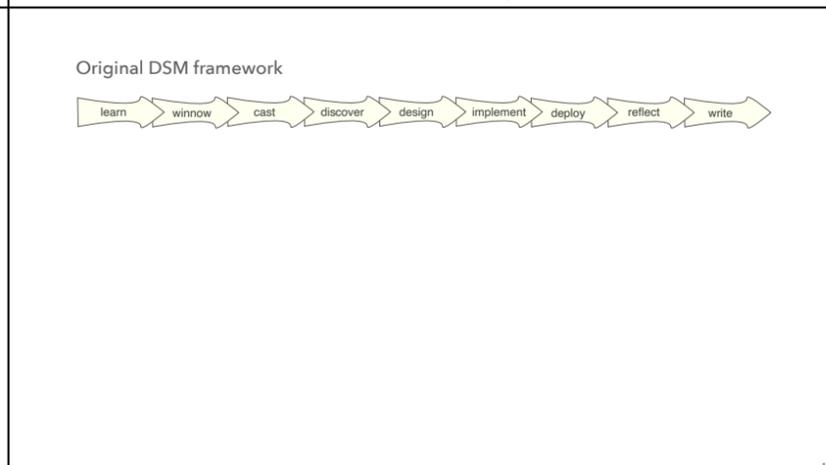
Michael Oppermann

## 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.

137



### Original DSM framework

### Data-first DSM framework

ADD

- What type of data am I working with?

141

### Original DSM framework

### Data-first DSM framework

ADD

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

142

### 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?

143

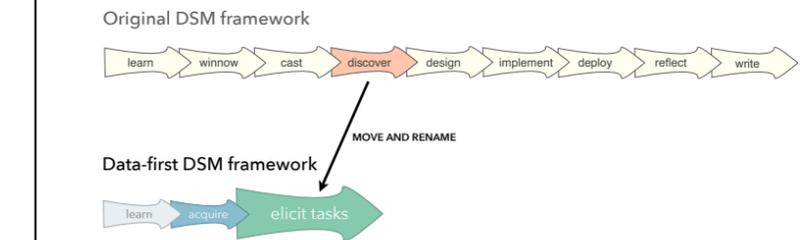
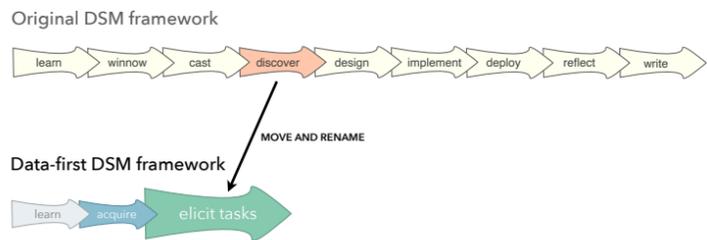
### Original DSM framework

### Data-first DSM framework

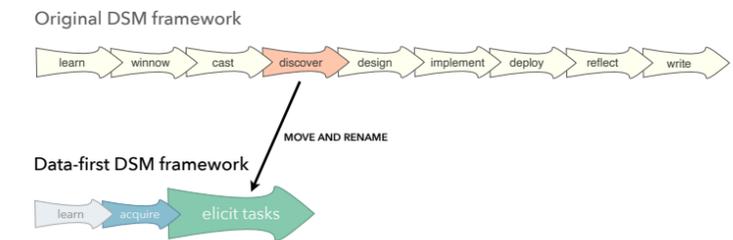
ADD

- 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?

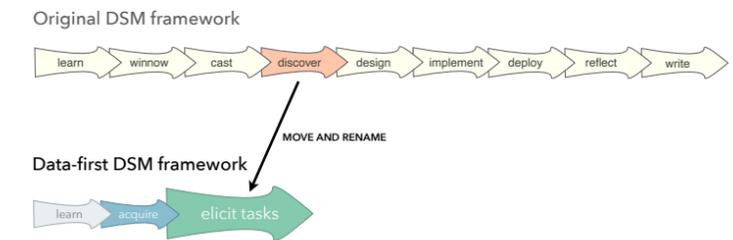
144



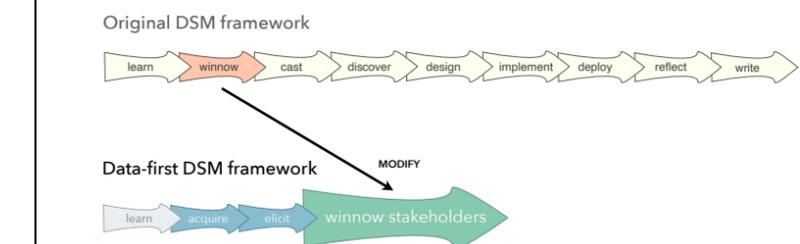
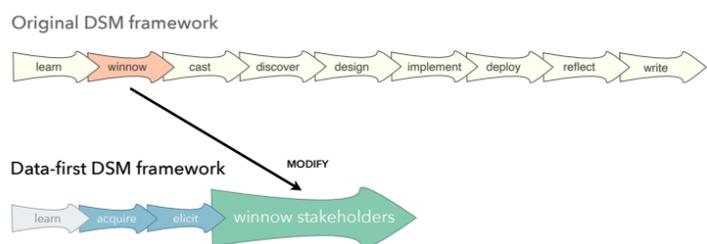
- Multiple potential stakeholders



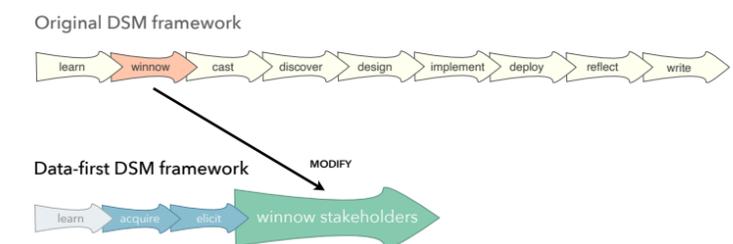
- Multiple potential stakeholders
- Explain initial data abstractions



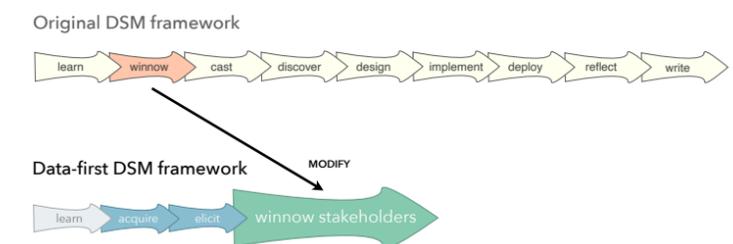
- Multiple potential stakeholders
- Explain initial data abstractions
- Learn about unsolved stakeholder needs



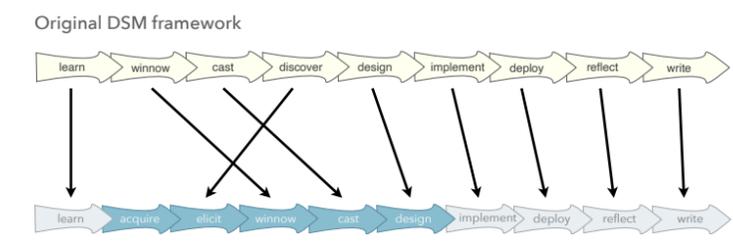
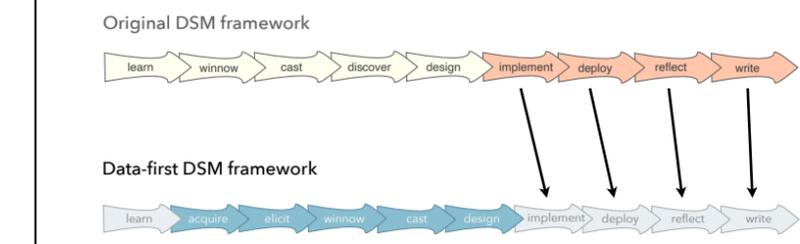
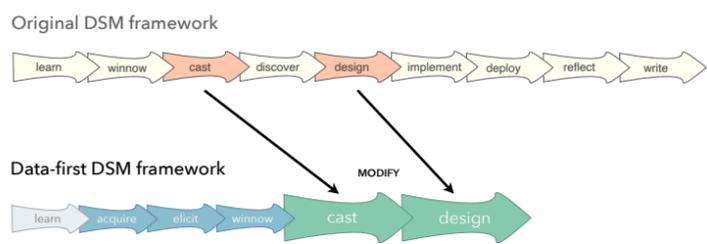
- How frequent are their data-relevant tasks?



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



- 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?



### 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 Colourtype 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

Human	A	T	G	A	C	A
Chimpanzee	A	T	C	G	A	C
Macaque	A	C	G	A	C	A

Genetic information

Computational workflow

Phylogenetic tree

### Many phylogenetic trees

Understand relationships between genes and species trees

Explore trees generated with different methods and data

Computational workflow

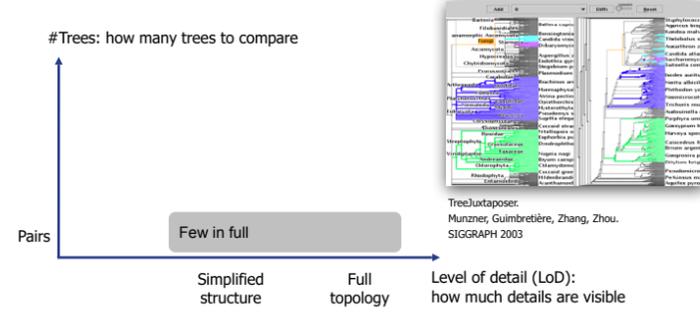
Phylogenetic tree

### Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare

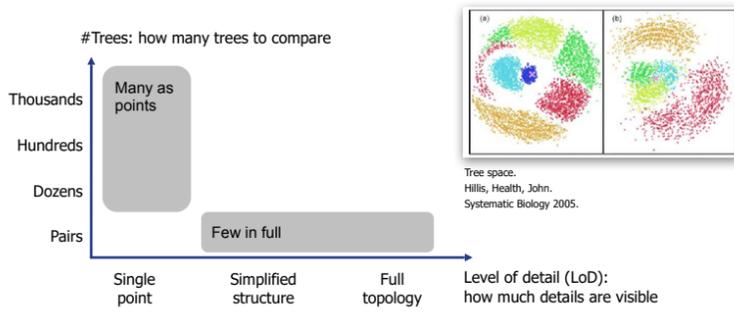
Level of detail (LoD): how much details are visible

# Scalability of Existing Tree Comparison Systems



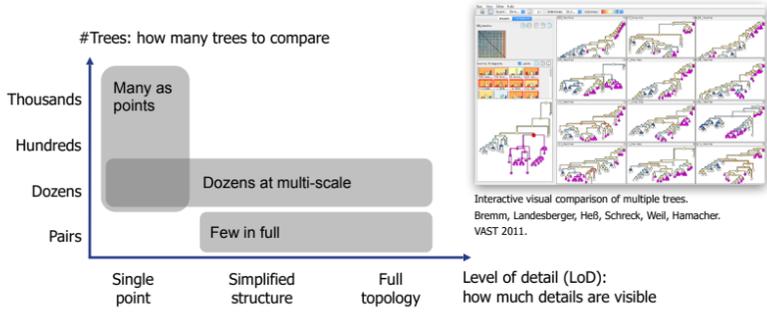
161

# Scalability of Existing Tree Comparison Systems



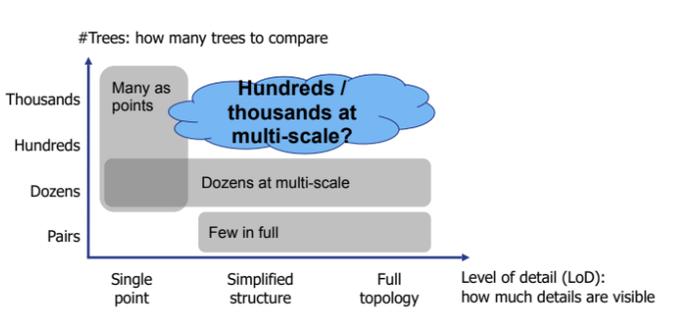
162

# Scalability of Existing Tree Comparison Systems



163

# Comparing many phylogenetic trees



164

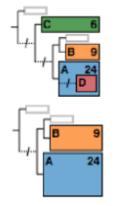
## Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees

165

## Contributions include idiom & algorithm levels

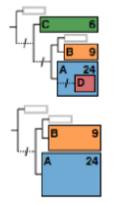
- Data and task abstractions for comparison of phylogenetic trees
- A new visual encoding: **Aggregated Dendrogram**
  - Compact tree representation that focuses on selected subtrees
  - Adapts to available screen space



166

## 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

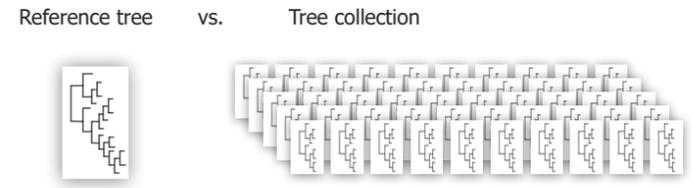


167

## Data & Tasks

- Tree data
- Two crucial tasks

## Tree data



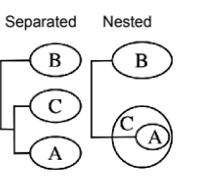
168

## Two crucial tasks

**Topological** relationships between subtrees / leaf nodes

## Two crucial tasks

**Topological** relationships between subtrees / leaf nodes

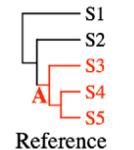
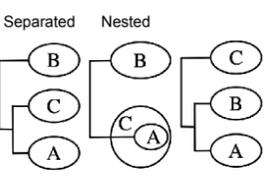


## Two crucial tasks

**Topological** relationships between subtrees / leaf nodes

**Leaf** node memberships compared to reference tree

- Topological distance



169

170

171

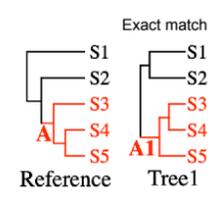
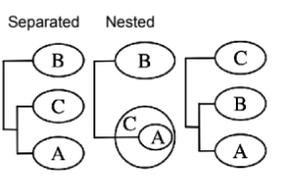
172

## Two crucial tasks

**Topological** relationships between subtrees / leaf nodes

- Topological distance

**Leaf** node memberships compared to reference tree

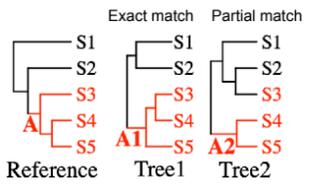
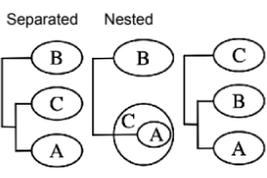


## Two crucial tasks

**Topological** relationships between subtrees / leaf nodes

- Topological distance

**Leaf** node memberships compared to reference tree



173

174

## Aggregated Dendrogram (AD)

- Intuition
- Visual design

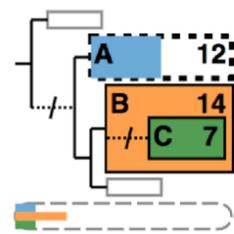
Intuition

Use glyphs to compress a tree according to user selections

175

176

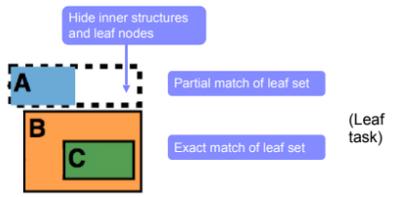
### Visual design: focus + context



177

### Visual design: focus + context

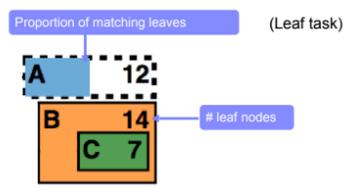
- Focus
  - Selected subtrees



178

### Visual design: focus + context

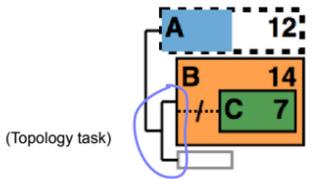
- Focus
  - Selected subtrees



179

### Visual design: focus + context

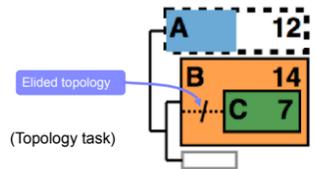
- Focus
  - Selected subtrees
  - Topological relationships between them



180

### Visual design: focus + context

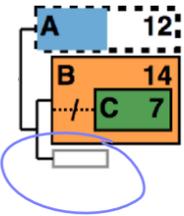
- Focus
  - Selected subtrees
  - Topological relationships between them



181

### Visual design: focus + context

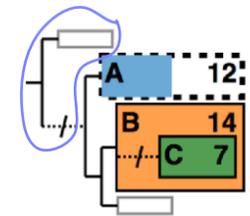
- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees



182

### Visual design: focus + context

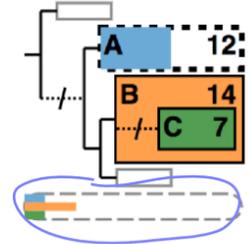
- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees
  - Upstream topology and root



183

### Visual design: focus + context

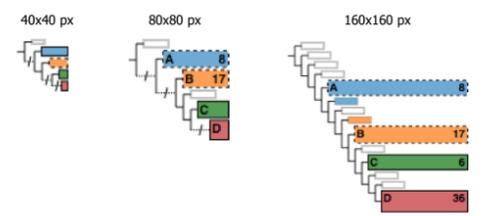
- Focus
  - Selected subtrees
  - Topological relationships between them
- Context
  - Neighboring subtrees
  - Upstream topology and root
  - Missing leaf nodes



184

### Visual design: algorithm adapts to space

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



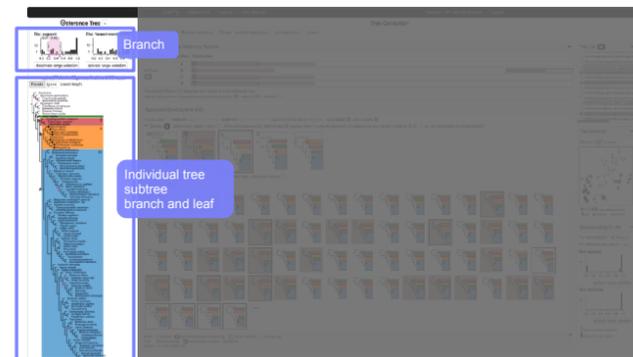
185

### ADView Interface: Multi-level structure across views



186

### Multi-level structure across views



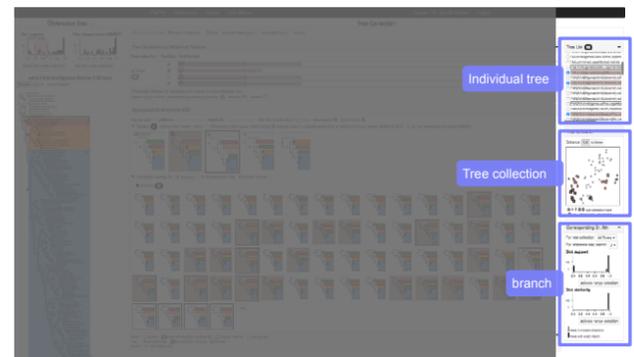
187

### Interface walkthrough: tree collection main views



188

### Interface walkthrough: tree collection aux. views



189

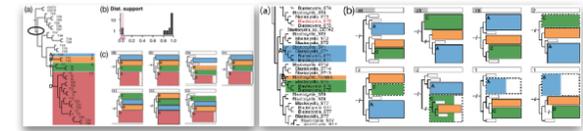
### Validation with many biologists

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

190

### Validation with many biologists

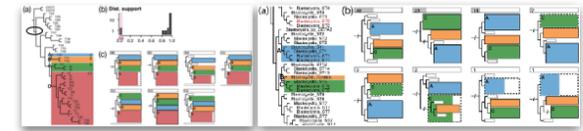
- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
  - 10 biologists at different times throughout project
- User study sessions
  - 5 biologists
  - Using their own datasets



191

### Validation with many biologists

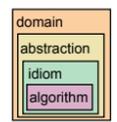
- Work closely with a biology PhD student (second author)
- Demos, interviews and discussions
  - 10 biologists at different times throughout project
- User study sessions
  - 5 biologists
  - Using their own datasets
- Biologists confirmed
  - Validity of data and task abstractions
  - Utility of ADView



192

## Problem-driven visualization through design studies

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



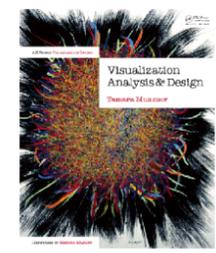
- three examples
  - different domains
  - different methods



## 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.



- 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>

