Why is validation difficult?
• different ways to get it wrong at each level

1. Domain situation
   - who are the target users?

2. Abstraction
   - translate from specifics of domain to vocabulary of visualization
   - 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

3. Visual encoding/interaction idiom
   - how is it shown?
   - visual encoding idiom: how to draw
   - interaction idiom: how to manipulate

4. Algorithm
   - efficient computation

Why is validation difficult?
• different ways to get it wrong at each level

Domain situation
You misunderstood the target users.

Abstraction
- translate from specifics of domain to vocabulary of visualization
- 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

Visual encoding/interaction idiom
- how is it shown?
- visual encoding idiom: how to draw
- interaction idiom: how to manipulate

Algorithm
- efficient computation

Next Slide
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..."
Design Study Methodology: Definitions

- Algorithm automation possible
- Design study methodology suitable
- Task clarity: not enough data
- Information location: computer

9-stage framework

1. Learn
2. Winnow
3. Cast
4. Discover
5. Implement
6. Design
7. Reflect
8. Write
9. Iterate

32 pitfalls & how to avoid them
**Real-world Clickstream Data**

- Scale is huge
- Variability is high

**Data: Client Sequences**

- Client Sequences: all actions performed by a single user

**Data: Session Sequences**

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

**Data: Segment Attributes**

- Size
- Sequence Attributes: Absolute, Relative
- Sequence Distribution: Start Time, Duration, Action Counts
- Action Distribution: Action Transitions: action before, action after

**Data: Segments**

- Segment: any set of sequences

**Tasks: Task Abstraction**

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

**Tasks: Task Abstraction**

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

**Tasks: Segment Behavior**

- Behavior: set of attribute constraints

**Tasks: Segment Behavior**

- Behavior set of attribute constraints
  - Expected
  - Unexpected
  - No purchases on a certain month
  - Favorable
  - Purchased
  - Unfavorable
  - Bounced

**Tasks: Task Abstraction**

- Identify: Find some set of sequences that constitutes interesting behavior
  - Drilldown: Distinguish more specific behaviors to further partition a segment previously defined by looser constraints
  - Frequency: Determine how many sequences are in the segment defined by behavior
  - Ordering: Determine when sequence occurs in the segment defined by behavior
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

Related Action

- 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

Why Visual Analytics?

Three case studies of problem-driven work
- e-commerce
- facilities management
- biology

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

Conclude

Solution

Video

Segmentifier: Interactively Refining Clickstream Data into Actionable Segments

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

Abandon

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

Visualizing Location-Based Counts Over Time Across Buildings

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

Previous measurement required physical counting or installation of additional hardware.

Previous visualization attempts were limited in space and time.
Focus Domains
- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control

Tasks
- Confirm assumptions or previous observations.
- Monitor the current/recent utilization rate.
- Communicate space usage and justify decisions.
- Validate the data (quality control).

Visualization Prototypes
Sandbox
- Data sketches, static data export
- Time
- Reusable Visualization Components

Spatial and Temporal Data Granularities
- Regions of interest
- Periods of interest
- Credit
- Building

Visualization Prototypes
- Campus Explorer
- Live-data stream, cross-building analysis
- Regions of interest
- Building Recent
- Building Long-term
- Region Compare

Evolutionary relationships of organisms

Computational workflow

Phylogenetic tree

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

Many phylogenetic trees

Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare

Level of detail (100): how much detail is visible
Scalability of Existing Tree Comparison Systems

- Comparing many phylogenetic trees
  - Levels of detail (LoD): how much detail is visible
  - How many trees to compare
  - Tree data
  - Topological relationships between subtrees / leaf nodes

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
  - Covers multiple levels of details for tree comparison

ADView
- Intuition
- Visual design
- Leaf node memberships compared to reference tree

Tree data
- Reference tree
- Tree collection

Two crucial tasks
- Topological relationships between subtrees / leaf nodes
- Leaf node memberships compared to reference tree

Intuition
- Use glyphs to compress a tree according to user selections
Interface walkthrough: tree collection aux. views

1. Visual design: focus + context
   - Focus
     - Selected subtrees
   - Context
     - Neighboring subtrees
   - Topological relationships between them

2. Visual design: focus + context
   - Focus
     - Selected subtrees
   - Context
     - Neighboring subtrees
          - Upstream topology and root
          - Missing leaf nodes
   - Topological relationships between them

3. Visual design: algorithm adapts to space
   - Show more info when space permitted
   - Labels
   - Leaf nodes
   - Neighboring blocks

4. ADView Interface: Multi-level structure across views

5. Multi-level structure across views

6. Interface walkthrough: tree collection main views

7. Interface walkthrough: tree collection aux. views

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

9. 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
   - Validity of data and task abstractions
   - Utility of ADView

10. Validation with many biologists
    - Work closely with a biology PhD student (second author)
    - Demos, interviews and discussions
      - Using their own datasets
    - 10 biologists at different times throughout project
    - User study sessions
      - 5 biologists
      - Using their own datasets
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
All Passion/Visualization/Information
CRC Press, 2014.

http://www.cs.ubc.ca/~tmm/talks.html#chinavis20
@tamaramunzner

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

- this talk
http://www.cs.ubc.ca/~tmm/talks.html/chinavis18