Guest & Research Lectures

Tamara Munzner
Department of Computer Science
University of British Columbia

CPSC 547, Information Visualization
7 December 2022

http://www.cs.ubc.ca/~tmm/courses/547-22
Today

• Steve Kasica, UBC
  – qual study: TableScraps, 15 min
  – Q&A, 5 min

• Stephen Kobourov, Univ. Arizona
  – algorithms: Scalable Graph Drawing w/ SGD, 15 min
  – algorithms: MetroSets, 15 min
  – Q&A 5-10 min

• Mara Solen, UBC
  – survey: VisLit, 15 min
  – Q&A, 5 min

• break, 10 min

• me
  – design spaces: Timelines Revisited, GEViT, 30 min
  – design studies: Ocupado, Aggregated Dendrograms, 25 min
  – imperfect models: TimelineCurator, 15 min
  – Q&A, 10 min
Steve Kasica
Stephen Kobourov
Mara Solen
break
design spaces
Design spaces: Continuing theme

The Structure of the Information Visualization Design Space
Stuart K. Card and Jock Mackinlay
Xerox PARC

Exploring the Design Space of Composite Visualization
Waqas Javed* Niklas Elmqvist†

A Design Space of Visualization Tasks
Hans-Jörg Schulz, Thomas Nocke, Magnus Heitzler, and Heidrun Schumann

A Design Space of Vision Science Methods for Visualization Research
Madison A. Elliott, Christine Nothelfer, Cindy Xiong, and Danielle Albers Szafir

Fig. 1. Overview of design space of experimental methods. We present a four component design space to guide researchers in creating visualization studies grounded in vision science research methods.
Design spaces: **What** are they?

- impose **systematic structure** on set of possibilities for specific problem
  - to capture the key variables at play
  - to support **reasoning about design choices**

- delineate
  - **cross-cutting** / independent / orthogonal
  - **axes** / dimensions / categories

- many names
  - design spaces, taxonomies, typologies, classifications, frameworks, models, ...  
  - space within which to express design patterns [*Javed/Elmqvist*]
Design spaces: What are they for?

- describe and analyze portions of design space to understand differences among designs & suggest new possibilities [Card & Mackinlay 1997]

- design spaces provide an actionable structure for systematically reasoning about solutions [Elliott et al 2020]

  - by grouping similar instances together to facilitate reasoning about classes rather than instances
Design spaces: How to **assess**?


  – **descriptive** power: ability to describe significant range of existing examples
  
  – **evaluative** power: ability to help assess multiple design alternatives
  
  – **generative** power: ability to help designers create new designs
Design spaces: How to create?

• **open coding** source material
  – grounded theory / thematic analysis / qualitative analysis

• **literature** review
  – synthesize across existing theories, compare & contextualize

• personal **reflection**
  – reflective synthesis

• complex combinations...
Design spaces: Multiple examples

• datatype: temporal, **timeline** visual encoding

• domain: **genomic epidemiology**, paper figure visual encoding

• domain: **journalism**, data **wrangling** activities

• domain agnostic: **abstract tasks**
Timelines
Timelines Revisited
A Design Space and Considerations for Expressive Storytelling

https://timelinesrevisited.github.io/
https://timelinestoryteller.com

Timelines Revisited: A Design Space and Considerations for Expressive Storytelling
Brehmer, Lee, Bach, Henry Riche, Munzner. IEEE TVCG 23(9):2151-2164

Matt Brehmer
Bongshin Lee
Benjamin Bach
Nathalie Henry Riche
Design space with three axes

- **representation**
  - Linear
  - Radial
  - Grid
  - Spiral
  - Arbitrary

- **scale**
  - Chronological
  - Relative
  - Logarithmic
  - Sequential
  - Sequential + Interim Duration

- **layout**
  - Unified (single timeline)
  - Faceted (multiple timelines)
  - Segmented timeline
  - Faceted + Segmented
Combinations: Characterize narrative, perceptual

- **Narrative point:** present a sequence of events.
- **Perceptual task:** arc position judgments.
- **Comment:** square aspect ratio.

- **Narrative point:** (approximately) compare lengths of sequences between facets.
- **Perceptual task:** arc length comparisons.

- **Narrative point:** compare chronology, duration, periodicity of events over months, weeks, days.
- **Perceptual task:** count and position judgments.
- **Comment:** only supports consecutive events.

- **Narrative point:** present a sequence of events.
- **Perceptual task:** area judgment.
- **Comment:** more compact than radial-sequential-unified timeline.
Viable combinations

- 20 out of 100 criteria
  - purposeful
  - interpretable
  - generalizable
Process

• **create** design space
  – **assemble** source material corpus: 145 timeline visualizations & timeline tools
  – **open code** group timelines together, select example for group, sketch alternatives
  – result: 3-axis design space

• **analyze** design space
  – 24 unique combinations (of 100) found in corpus
  – 20 we deemed viable
Assessment & adoption

• descriptive power
  – validated coverage through checking 118 additional timelines ("test set")
    • all timelines can be described (263 total)
    • 253 characterized as viable

• generative power
  – implemented sandbox authoring software for 20 viable designs
    • & transitions between them
  – created designs for 28 representative datasets
    • 7 full story videos

• adoption
  – open sourced & distributed as Microsoft product
    • free browser version at https://timelinestoryteller.com/
    • free add-on for PowerBI
Genomic Epidemiology
A systematic method for surveying data visualizations and a resulting genomic epidemiology visualization typology: GEViT.

https://amcrisan.github.io/gevit

Propose typology creation method: mixed qual and quant

- Analyzed research articles
- Some analyses are automated (🤖) and others are manual (👤)

Analysis Phase

- Literature Analysis
- Visualization Analysis
- Qualitative Analysis
- Quantitative Analysis
Use method to develop typology in specific domain

- Developed a **Genomic Epidemiology Visualization Typology** (GEViT)

### Literature Analysis

**Topic Clusters**
Sampling Strata

**Article Sampling**
Random stratified sampling

### Visualization Analysis

**Figure Extraction**
Sample articles

**Iterative & Axial Coding**
Development of GEViT

- Chart Type
- Chart Combination
- Chart Enhancement
Domain prevalence design space

A General Method Overview

Analysis Phase

- Literature Analysis
- Visualization Analysis

Qualitative Analysis
Quantitative Analysis

Research Question

- WHY are researchers visualizing data?
- HOW are researchers visualizing data?
- HOW MANY examples are there of specific visualizations?

B Application of our Method to Infectious Disease Genomic Epidemiology

Literature Analysis Steps

- Text mining to identify topics & assign a priori concepts
- Perform a random stratified sampling of articles

Our Objective

Across the many topics of infectious disease gen. epi. research articles, describe and quantify the different kinds of visualizations used

Visualization Analysis Steps

- Generate GEViT to describe elements of a visualization
- Apply GEViT to collection of research figures

- Apply descriptive statistics to GEViT to quantify variability
By the numbers

<table>
<thead>
<tr>
<th>Analysis Step</th>
<th>Number of Articles</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article acquisition and unsupervised topic</td>
<td>17,974</td>
<td>35 topic clusters</td>
</tr>
<tr>
<td>clustering</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15,315</td>
<td>Remove articles that never cluster</td>
</tr>
<tr>
<td>Validation and limitation to human pathogens</td>
<td>9,551</td>
<td>Remove articles with non-human pathogens</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 pathogen topic clusters</td>
</tr>
<tr>
<td>Application of a priori concepts</td>
<td>6,350</td>
<td>Remove pathogens with fewer than 40 articles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23 a priori concepts</td>
</tr>
</tbody>
</table>

**Sampling Round 1**

Reject (No) or Accept (Yes)

- 6074
- 276

**Sampling Round 2**

Reject (No) or Accept (Yes)

- 293
- 97

Finalization of articles

- 204
- 17 manually added

- 221

801 Figures
49 Missed Opportunity Tables
Design space axis: Chart types used in genEpi

Common Statistical Charts
- Bar Chart: Standard, Stacked, Divergent
  - Special Cases: Epidemic Curve, Diversity Chart, LeSe Plot
- Line Chart: Special Cases: Bootstrap, Kaplan-Meier, Skyline Plot
- Scatter Plot: Special Cases: Root-to-tip, Ordination Plot, Q-Q plot

Distribution Plot
- Histogram, PDF, Boxplot, Swarm Plot

Colour Charts
- Pie Chart, Venn Diagram
- Category Stripe, Heatmap, Density Plot

Relational Charts
- Node-link: Special Cases: eBurst, Social network, Molecular network, Minimum Spanning Tree

Flow Diagram
- Chord Diagram, Sankey Diagram

Temporal Charts
- Streamgraph*: Absolute, Relative
- Timeline

Spatial Charts
- Geographic Map, Choropleth Map, Interior Map

Tree Charts
- Phylogenetic Tree: Rooted (Linear & Radial)

Genomic Charts
- Genomic Map: Linear, Radial
- Alignment, Composition Plot
- Sequence Logo Plot

Other Charts
- Table, Image, Gel Image, General Image, Miscellany
Design space axis: Chart combinations of heterogeneous data

- **Spatially Aligned**
  - Horizontal / Vertical Alignment
  - 20%

- **Visually Aligned**
  - Colour / Shape Alignment
  - 14%

- **Small Multiples**
  - Chart Alignment
  - 17%

- **Unaligned**
  - 9%
### Design space axis: Enhancement choices, atop base chart types

<table>
<thead>
<tr>
<th>Size</th>
<th>Shape</th>
<th>Color</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point</td>
<td><img src="image" alt="Point Shapes" /></td>
<td><img src="image" alt="Color Variants" /></td>
<td><img src="image" alt="Texture Variants" /></td>
</tr>
<tr>
<td>Line</td>
<td><img src="image" alt="Line Shapes" /></td>
<td><img src="image" alt="Color Variants" /></td>
<td><img src="image" alt="Texture Variants" /></td>
</tr>
<tr>
<td>Area</td>
<td><img src="image" alt="Area Shapes" /></td>
<td><img src="image" alt="Color Variants" /></td>
<td><img src="image" alt="Texture Variants" /></td>
</tr>
<tr>
<td>Text</td>
<td><img src="image" alt="Text Characters" /></td>
<td><img src="image" alt="Color Variants" /></td>
<td><img src="image" alt="Text Variants" /></td>
</tr>
</tbody>
</table>

**Current Practice**

>80% of all figures have some enhancement
GEViT example
GEViT example

Visualization Breakdown

Literature Analysis (why)

- **Pathogen**: Enterococcus faecium
GEViT example

Visualization Breakdown

Literature Analysis *(why)*
- **Pathogen:** *Enterococcus faecium*

Visualization Analysis *(how)*

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Tree (Rooted Phylogenetic Tree)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Category Stripe</td>
</tr>
<tr>
<td></td>
<td>Heatmap (Variation Profile)</td>
</tr>
<tr>
<td>Chart Combination</td>
<td>Spatially Aligned <em>(horizontal)</em></td>
</tr>
</tbody>
</table>

Tree Chart Type: **Tree**

Heatmap (Variation Profile): **C.S**
GEViT example

Visualization Breakdown

Literature Analysis (why)
- Pathogen: Enterococcus faecium

Visualization Analysis (how)

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Tree (Rooted Phylogenetic Tree)</th>
<th>Category Stripe</th>
<th>Heatmap (Variation Profile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chart Combination</td>
<td>Spatially Aligned (horizontal)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chart Enhancement</td>
<td>Re-encode Marks</td>
<td>Tree – branches</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Add Marks</td>
<td>Tree - Connection Marks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Add Mark (unstructured)</td>
<td>Heatmap – Textboxes</td>
<td></td>
</tr>
</tbody>
</table>
Assessment

• descriptive power
  – provided common language for describing data visualization in genEpi
  – established gap: **unmet tooling needs**
    • no existing tool handled full complexity of what people do manually

• evaluative power
  – **revealed shortfalls** in practices of some genEpi stakeholders
    • eg overuse of text

• generative power
  – validated in followup GEViTRec work
    • **build** automatic recommender system using domain prevalence design space
## Summary: Multiple design spaces

<table>
<thead>
<tr>
<th>Design Space</th>
<th>Open Coding</th>
<th>Sampling Strategy</th>
<th>Reflective Synthesis Timing</th>
<th>Vis Research Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>timeline visual encoding</td>
<td>standalone timelines</td>
<td>assembled corpus</td>
<td>early</td>
<td>some source material</td>
</tr>
<tr>
<td>genEpi visual encoding</td>
<td>figures from papers</td>
<td>stratified random sampling with topic clusters</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>wrangling activities</td>
<td>software from repos</td>
<td>diversity criteria</td>
<td>late</td>
<td>terms: light mapping</td>
</tr>
</tbody>
</table>
## Summary: Multiple design spaces

<table>
<thead>
<tr>
<th>Design Space</th>
<th>Descriptive Power</th>
<th>Generative Power</th>
<th>Descriptive vs Generative</th>
<th>Evaluative Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>timeline visual encoding</td>
<td>validated against test set</td>
<td>software implementation of authoring system, used to create example gallery/videos</td>
<td>analysis to characterize viable subset</td>
<td></td>
</tr>
<tr>
<td>genEpi visual encoding</td>
<td>systematic method yields comprehensive coverage</td>
<td>software implementation of automatic recommender (followup)</td>
<td>same (detailed)</td>
<td></td>
</tr>
<tr>
<td>wrangling activities</td>
<td>high precision, gaps / divergence found for domain</td>
<td>concise framework (followup implementation TBD)</td>
<td>develop entirely new framework</td>
<td></td>
</tr>
</tbody>
</table>
Design spaces: How to assess? Larger context: theory types

• Ben Shneiderman, *Designing the User Interface*: descriptive, explanatory, prescriptive, predictive

• Paul Ralph, *Toward Methodological Guidelines for Process Theories & Taxonomies in Software Engineering, IEEE TSE 2020*

  – theory types
    • theories for **understanding**: organizing what is happening into useful categories (taxonomies)
    • **process** theories: how something happens (often taxonomies++)
    • **variance** theories: why something happens, causal relationships between constructs
      – predictive

  – relevant criteria for taxonomies
    • **yes**: parsimony, transferability, theoretical saturation
    • **sometimes**: utility, originality, resonance/believability, testability
    • **no**: statistical generalizability, construct validity, internal validity, conclusion validity
design studies
Two design studies

- facilities management
- biology
Ocupado

Visualizing Location-Based Counts Over Time Across Buildings


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

Previous measurement required physical counting or installation of additional hardware.
Previous measurement required physical counting or installation of additional hardware.

Previous visualization attempts were limited in space and time.
Design Study
WiFi Connections: Location-Based Counts
WiFi Connections: Location-Based Counts
WiFi Connections: Location-Based Counts
WiFi Connections: Location-Based Counts
WiFi Connections: Location-Based Counts
WiFi Connections: Location-Based Counts

device connections over time
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
Automated HVAC control
Data 

Decision making
WiFi connections as a proxy for occupancy
WiFi connections as a proxy for occupancy
Interviews with potential stakeholders
Focus Domains

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

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

Semi-structured discussions and live demos
Tasks

✅ **Confirm** assumptions or previous observations.

Do students occupy room x in evenings or on weekends?
Tasks

- **Confirm** assumptions or previous observations.
- **Monitor** the current/recent utilization rate.

Which rooms are empty/busy?
Tasks

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

Space usage improved after renovation.
Tasks

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

Sandbox

Data sketches, static data export
Visualization Prototypes

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

Sandbox

Data sketches, static data export
Spatial and Temporal Data Granularities

Regions of interest
Spatial and Temporal Data Granularities

Regions of interest

Periods of interest
Visualization Prototypes

- **Sandbox**
  Data sketches, static data export

- **Campus Explorer**
  Live-data stream, cross-building analysis

- **Building Recent**

- **Building Long-term**

- **Region Compare**

Time
Reusable Visualization Components
### Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
</tr>
</tbody>
</table>

- **Sparkline**: A small chart that shows trends and outliers in a compact form.
- **Juxtaposition**: The side-by-side comparison of visual elements to highlight differences or similarities.
- **Comparisons** (contiguous): Visual comparisons that are contiguous in the layout.
# Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
</tr>
<tr>
<td></td>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (non-contiguous)</td>
</tr>
</tbody>
</table>
# Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(contiguous)</td>
</tr>
<tr>
<td></td>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>(non-contiguous)</em></td>
</tr>
<tr>
<td></td>
<td>Confidence band line chart</td>
<td>Aggregation</td>
<td>Typical utilization profiles</td>
</tr>
</tbody>
</table>
## Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers</td>
</tr>
<tr>
<td></td>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers</td>
</tr>
<tr>
<td></td>
<td>Confidence band</td>
<td>Aggregation</td>
<td>Typical utilization profiles</td>
</tr>
<tr>
<td></td>
<td>Superimposed</td>
<td>Superposition</td>
<td>Within-session patterns, outliers</td>
</tr>
</tbody>
</table>

* (contiguous)
* (non-contiguous)
## Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td></td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
</tr>
<tr>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td></td>
<td>Repeating patterns, trends, outliers (non-contiguous)</td>
</tr>
<tr>
<td>Confidence band</td>
<td>Aggregation</td>
<td></td>
<td>Typical utilization profiles</td>
</tr>
<tr>
<td>line chart</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superimposed</td>
<td>Superposition</td>
<td></td>
<td>Within-session patterns, outliers</td>
</tr>
<tr>
<td>line chart</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
<td></td>
</tr>
<tr>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (non-contiguous)</td>
<td></td>
</tr>
<tr>
<td>Confidence band</td>
<td>Aggregation</td>
<td>Typical utilization profiles</td>
<td></td>
</tr>
<tr>
<td>line chart</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superimposed</td>
<td>Superposition</td>
<td>Within-session patterns, outliers</td>
<td></td>
</tr>
<tr>
<td>line chart</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spatial</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor plan with</td>
<td>Superposition</td>
<td>Within local spatial neighborhood</td>
<td></td>
</tr>
<tr>
<td>symbols</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial heatmap</td>
<td>Containment (nested)</td>
<td>Across distributed regions</td>
<td></td>
</tr>
</tbody>
</table>
Campus Explorer
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
Two design studies

- facilities management
- biology
Aggregated Dendrograms
for Visual Comparison Between Many Phylogenetic Trees


Evolutionary relationships of organisms

Phylogenetic tree

Genetic information

Computational workflow

Human
Chimpanzee
Macaque

ATGACGA
ATGACGA
ACGACGA
Many phylogenetic trees

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

#Trees: how many trees to compare

Pairs

Simplified structure

Full topology

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

TreeJuxtaposer.
Munzner, Guimbretière, Zhang, Zhou.
SIGGRAPH 2003
Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare

- Thousands
- Hundreds
- Dozens
- Pairs

Single point  |  Simplified structure  |  Full topology

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

Tree space.
Hillis, Health, John.
Systematic Biology 2005.
Scalability of Existing Tree Comparison Systems

#Trees: how many trees to compare

- Thousands
  - Many as points

- Hundreds
  - Dozens at multi-scale

- Dozens
  - Few in full

- Pairs
  - Single point
  - Simplified structure
  - Full topology

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

Comparing many phylogenetic trees

#Trees: how many trees to compare

- Thousands: Many as points
- Hundreds: Dozens at multi-scale
- Dozens: Few in full
- Pairs: Single point

Simplified structure

Full topology

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

Hundreds / thousands at multi-scale?
Contributions include idiom & algorithm levels

- Data and task abstractions for comparison of phylogenetic trees
Contributions include idiom & algorithm levels

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

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

- Tree data
- Two crucial tasks
Tree data

Reference tree vs. Tree collection
Two crucial tasks

**Topological** relationships between subtrees / leaf nodes
Two crucial tasks

Topological relationships between subtrees / leaf nodes

Separated

Nested
Two crucial tasks

**Topological** relationships between subtrees / leaf nodes
- Topological distance

**Leaf** node memberships compared to reference tree

Separated      Nested

Separated      Nested

Reference
Two crucial tasks

**Topological** relationships between subtrees / leaf nodes
- Topological distance

**Leaf** node memberships compared to reference tree

![Diagram showing topological relationships and leaf node memberships](image-url)
Two crucial tasks

**Topological** relationships between subtrees / leaf nodes
- Topological distance

**Leaf** node memberships compared to reference tree

Separated      Nested

Reference

Tree1

Tree2
Aggregated Dendrogram (AD)

- Intuition
- Visual design
Use glyphs to compress a tree according to user selections.
Visual design: focus + context
Visual design: focus + context

- Focus
  - Selected subtrees
Visual design: focus + context

- **Focus**
  - Selected subtrees

![Diagram showing focus and context]

- Proportion of matching leaves
- # leaf nodes

(Leaf task)
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
Visual design: focus + context

- Focus
  - Selected subtrees
  - Topological relationships between them
Visual design: focus + context

- **Focus**
  - Selected subtrees
  - Topological relationships between them

- **Context**
  - Neighboring subtrees
Visual design: focus + context

- **Focus**
  - Selected subtrees
  - Topological relationships between them

- **Context**
  - Neighboring subtrees
  - Upstream topology and root
Visual design: focus + context

- **Focus**
  - Selected subtrees
  - Topological relationships between them

- **Context**
  - Neighboring subtrees
  - Upstream topology and root
  - Missing leaf nodes
Visual design: algorithm adapts to space

- Show more info when space permitted
  - Labels
  - #leaf nodes
  - Neighboring blocks
ADView Interface: Multi-level structure across views
Multi-level structure across views

Branch

Individual tree subtree branch and leaf
Interface walkthrough: tree collection main views

Tree collection
Subset of trees

Tree collection
Subset of trees

Individual tree
Subtree
Interface walkthrough: tree collection aux. views
Aggregated Dendrograms for Visual Comparison between Many Phylogenetic Trees

Zipeng Liu, Shing Hei Zhan, Tamara Munzner
Validation with many biologists

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

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

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

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

• two examples
  – different domains
  – different methods
imperfect models
One case study of visualizing imperfect models

• NLP for temporal data
TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text.

Manual creation process

1868 The Typewriter
1885 The Movie
1997 The Sphere
2007 Multi Touch

125
Structured creation process

TimelineJS
timeline.knightlab.com/
## Timeline authoring model

- time required for each task

<table>
<thead>
<tr>
<th></th>
<th>Browse</th>
<th>Extract</th>
<th>Format</th>
<th>Show</th>
<th>Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Drawing</td>
<td><img src="image" alt="snail" /></td>
<td><img src="image" alt="snail" /></td>
<td><img src="image" alt="snail" /></td>
<td><img src="image" alt="snail" /></td>
<td><img src="image" alt="snail" /></td>
</tr>
<tr>
<td>Structured Creation</td>
<td><img src="image" alt="snail" /></td>
<td><img src="image" alt="snail" /></td>
<td><img src="image" alt="gear" /></td>
<td><img src="image" alt="gear" /></td>
<td><img src="image" alt="rabbit" /></td>
</tr>
<tr>
<td><strong>Timeline Curator</strong></td>
<td><img src="image" alt="rabbit" /></td>
<td><img src="image" alt="gear" /></td>
<td><img src="image" alt="gear" /></td>
<td><img src="image" alt="rabbit" /></td>
<td><img src="image" alt="rabbit" /></td>
</tr>
</tbody>
</table>
The general case for curation

- build for human in the loop as continuing need
  - automatic processing to accelerate not replace
  - assume computational results good but not perfect
    - for the indefinite future!
  - visual feedback to accelerate
The importance of being brisk

• cool use case: eureka moment
  – success: enable what was impossible before
  – vis tools for new insights & discoveries

• workhorse use case: workflow speedup
  – success: vis tools accelerate your prior workflow
    • sometimes enables the previously infeasible

• TLC use cases
  – started with speedup use case, for presentation
    • make this doc into a timeline now!
  – two other use cases nudge towards exploration
    • comparison between multiple timelines
    • speculative browsing
TimeLineCurator: Speculative Browsing

https://vimeo.com/jofu/tlc
Q&A
Come talk!

• encourage meeting with me to get advice/feedback before final present
  – chance to get feedback while you can still act on it
  – optional, not mandatory
  – wise to schedule in advance by email
    • can’t meet with all 14 teams in next week office hours, or in last few days!