Developing Design Spaces for Visualization

Tamara Munzner
Department of Computer Science
University of British Columbia

Stanford HCI Seminar
4 March 2022, virtual

http://www.cs.ubc.ca/~tmm/talks.html#stanf22
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 and Niklas Elmqvist

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 19, NO. 12, DECEMBER 2013

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 Notherfer, 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]

• taxonomies increase cognitive efficiency & support inferences [Ralph, Toward Methodological Guidelines for Process Theories & Taxonomies in Software Engineering. IEEE. TSE 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
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
- 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/](https://timelinestoryteller.com/)
    • free add-on for PowerBI
A systematic method for surveying data visualizations and a resulting genomic epidemiology visualization typology:

GEViT

https://amcrisan.github.io/gevit

A systematic method for surveying data visualizations and a resulting genomic epidemiology visualization typology: GEViT.

Propose typology creation method: mixed qual and quant

- Analyzed research articles
- Some analyses are automated (🤖) and others are manual (👤)
Use method to develop typology in specific domain

- Developed a Genomic Epidemiology Visualization Typology (GEViT)
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>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>Application of <em>a priori</em> concepts</td>
<td>6,350</td>
<td>Remove pathogens with fewer than 40 articles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 pathogen topic clusters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23 <em>a priori</em> concepts</td>
</tr>
<tr>
<td>Sampling Round 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Reject (No) or Accept (Yes)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6074</td>
<td></td>
</tr>
<tr>
<td></td>
<td>276</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampling Round 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Reject (No) or Accept (Yes)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>293</td>
<td></td>
</tr>
<tr>
<td></td>
<td>179</td>
<td></td>
</tr>
<tr>
<td></td>
<td>97</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finalization of articles</td>
<td>204</td>
<td>801 Figures</td>
</tr>
<tr>
<td></td>
<td>+ 17 manually added</td>
<td>49 Missed Opportunity Tables</td>
</tr>
<tr>
<td></td>
<td>221</td>
<td></td>
</tr>
</tbody>
</table>
Design space axis: Chart types used in genEpi

Common Statistical Charts
- Bar Chart
  - Standard
  - Stacked
  - Divergent
- Line Chart
  - Special Cases
    - Epidemic Curve
    - Diversity Chart
    - LeSe Plot
- Scatter Plot
  - Special Cases
    - Bootstrap
    - Kaplan-Meier
    - Skyline Plot

Distribution Plot
- Histogram
- PDF
- Boxplot
- Swarm Plot

Relational Charts
- Node-link
  - Special Cases
    - eBurst
    - Social network
    - Molecular network
    - Minimum Spanning Tree
- Flow Diagram
  - Chord Diagram
  - Sankey Diagram

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

Temporal Charts
- Streamgraph*
  - Absolute
  - Relative
- Timeline

Spatial Charts
- Geographic Map
- Choropleth Map
- Interior Map

Tree Charts
- Phylogenetic Tree
  - Rooted (Linear & Radial)
- Unrooted (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></th>
<th>Size</th>
<th>Shape</th>
<th>Color</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point</strong></td>
<td><img src="image" alt="point" /></td>
<td><img src="image" alt="square" /></td>
<td><img src="image" alt="circle" /></td>
<td><img src="image" alt="patterns" /></td>
</tr>
<tr>
<td><strong>Line</strong></td>
<td><img src="image" alt="lines" /></td>
<td><img src="image" alt="curved lines" /></td>
<td><img src="image" alt="dashes" /></td>
<td><img src="image" alt="dotted lines" /></td>
</tr>
<tr>
<td><strong>Area</strong></td>
<td><img src="image" alt="areas" /></td>
<td><img src="image" alt="area patterns" /></td>
<td><img src="image" alt="area colors" /></td>
<td><img src="image" alt="area textures" /></td>
</tr>
<tr>
<td><strong>Text</strong></td>
<td><img src="image" alt="text" /></td>
<td><img src="image" alt="text patterns" /></td>
<td><img src="image" alt="text colors" /></td>
<td><img src="image" alt="text faces" /></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>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Chart Combination</th>
<th>Spatially Aligned <em>(horizontal)</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Chart Enhancement</th>
<th>Re-encode Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Marks</td>
<td>Tree - Connection Marks</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Add Mark <em>(unstructured)</em></th>
<th>Heatmap - Textboxes</th>
</tr>
</thead>
</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
GEViTRec: Data Reconnaissance Through Recommendation Using a Domain-Specific Visualization Prevalence Design Space

https://github.com/amcrisan/GEVitRec

Data Wrangling
An Actionable Framework for Multi-Table Data Wrangling

From an Artifact Study of Computational Journalism


An Actionable Framework for Multi-Table Data Wrangling From an Artifact Study of Computational Journalism.
Journalists are data wranglers...
...who show their work publicly

- lots of wrangling behind the scenes
- enter the “nerd box”
  - article sidebars or snippet
  - provide / link
    - methods, analysis materials
- publish code/data to public repos
  - hundreds on GitHub & Observable
- editorial transparency
  - public can scrutinize
  - colleague can reproduce

Note: All the refugee data in this post comes from the Department of State’s Refugee Processing Center, and cover through November 18, 2015. The raw data and supporting data analysis can be found here.
What are the wrangling practices of journalists with programming skills?

Technical observation

Repo selection  Qualitative coding
Process overview

What are the wrangling practices of journalists with programming skills?

Technical observation

Repo selection  Qualitative coding

Data-flow sketches
Process overview

**What are the wrangling practices of journalists with programming skills?**

**Technical observation**
- Repo selection
- Qualitative coding

**Taxonomies of data wrangling in computational journalism - initial**
Process overview

What are the wrangling practices of journalists with programming skills?

- Technical observation
  - Repo selection
  - Qualitative coding

Which practices align with or diverge from existing characterizations?

- Literature search
  - Term harmonization
  - Gap discovery

Taxonomies of data wrangling in computational journalism - initial
What are the wrangling practices of journalists with programming skills?

- Technical observation
  - Repo selection
  - Qualitative coding

Which practices align with or diverge from existing characterizations?

- Literature search
  - Term harmonization
  - Gap discovery

Taxonomies of data wrangling in computational journalism - finalized
Process overview

What are the wrangling practices of journalists with programming skills?

Technical observation
- Repo selection
- Qualitative coding

Which practices align with or diverge from existing characterizations?

Literature search
- Term harmonization
- Gap discovery

How to re-characterize wrangling to match the observed practices?

Reflective Synthesis

Taxonomies of data wrangling in computational journalism
Reflective Synthesis

How to re-characterize wrangling to match the observed practices?

Which practices align with or diverge from existing characterizations?

What are the wrangling practices of journalists with programming skills?

Technical observation
- Repo selection
- Qualitative coding

Literature search
- Term harmonization
- Gap discovery

Taxonomies of data wrangling in computational journalism

Multi-table framework of data wrangling
By the numbers

Phase 1: Technical observation study

1.1: Repo selection
- GitHub Repos
- Observable Notebooks
- Inclusion criteria
- 1,301 journalists’ repos
- Exclusion criteria
- Select w.r.t. diversity criteria
- 225 curated repos

Primary contributions
In supplemental materials

1.2: Qualitative coding
- Selected repo
- Codeset
- Update
- Annotate
- Add to coded pool
- Group by axial codes
- Taxonomy
- Re-annotate: after 25, 50
- 50 annotated repos

Phase 2: Literature search
- Computer Science
- Computational Journalism
- Judge for relevance and treatment of multiple tables
- Curated set of related work papers
- Align names, assess novelty
- Final taxonomies of data wrangling by journalists: 165 codes

Phase 3: Reflective synthesis
- Multi-table wrangling framework: 21 operations
- Coverage check
- Taxonomy-framework cross check
Two taxonomies of data wrangling in journalism

- **Actions** taken by journalists
- **Process** interpreted by researchers

- **descriptive** power: excellent
  - total codes: 165
  - max depth: 5 levels

- **generative** power: limited

**Actions**
- Import
- Clean
- Merge
- Profile
- Drive
- Transform
- Export

**Process**
- Source
- Workflow
- Cause
- Themes
- Analysis
- Management
- Pain Points

- descriptive power: excellent
  - total codes: 165
  - max depth: 5 levels

- generative power: limited
Key finding: journalists use many, many tables
Key finding: journalists use many, many tables

- workflow complexity varies greatly
Key finding: journalists use many, many tables

- workflow complexity varies greatly
- current interactive wrangling applications do not scale well
Key finding: journalists use many, many tables

- workflow complexity varies greatly
- current interactive wrangling applications do not scale well
- re-characterize wrangling design space to match these observed practices
Two axes of multi-table wrangling design space
# Two axes of multi-table wrangling design space

<table>
<thead>
<tr>
<th>Object type</th>
<th>Table</th>
<th>Row</th>
<th>Column</th>
</tr>
</thead>
</table>

Two axes of multi-table wrangling design space

<table>
<thead>
<tr>
<th>Object type</th>
<th>Table</th>
<th>Row</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create</td>
<td></td>
<td>0 → 1</td>
<td></td>
</tr>
<tr>
<td>Delete</td>
<td></td>
<td>1 → 0</td>
<td></td>
</tr>
<tr>
<td>Transform</td>
<td></td>
<td>1 → 1</td>
<td></td>
</tr>
<tr>
<td>Separate</td>
<td></td>
<td>1 → n</td>
<td></td>
</tr>
<tr>
<td>Combine</td>
<td></td>
<td>n → 1</td>
<td></td>
</tr>
</tbody>
</table>
Multi-table data wrangling design space

• concise and actionable
  – **generative** power achieved
  – suitable framework for building tool

<table>
<thead>
<tr>
<th>I/O Cardinality</th>
<th>Table</th>
<th>Row</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create</td>
<td>$0 \rightarrow 1$</td>
<td><img src="Image" alt="Diagram" /></td>
<td><img src="Image" alt="Diagram" /></td>
</tr>
<tr>
<td>Delete</td>
<td>$1 \rightarrow 0$</td>
<td><img src="Image" alt="Diagram" /></td>
<td><img src="Image" alt="Diagram" /></td>
</tr>
<tr>
<td>Transform</td>
<td>$1 \rightarrow 1$</td>
<td><img src="Image" alt="Diagram" /></td>
<td><img src="Image" alt="Diagram" /></td>
</tr>
<tr>
<td>Separate</td>
<td>$1 \rightarrow n$</td>
<td><img src="Image" alt="Diagram" /></td>
<td><img src="Image" alt="Diagram" /></td>
</tr>
<tr>
<td>Combine</td>
<td>$n \rightarrow 1$</td>
<td><img src="Image" alt="Diagram" /></td>
<td><img src="Image" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Assessment: Cross-check

- cross-check coverage of multi-table framework vs actions taxonomy
  - verify descriptive power
Abstract Tasks
A Multi-Level Typology of Abstract Visualization Tasks.


https://www.cs.ubc.ca/labs/imager/tr/2013/MultiLevelTaskTypology/
Task abstraction: Gap

Previous Work

Classifying Tasks, Goals, Intentions, Objectives, Activities, Interactions

A mid-level gap?

Meyer, Sedlmair, & Munzner (BELIV 2012)

low level of abstraction
e.g. “retrieve value”

high level of abstraction
e.g. “integration of insight”

Heer & Shneiderman (2012)
Mullins & Treu (1993)
Pike, Stasko, et al. (2009)
Springmeyer et al. (1992)
RE Roth (2012)

Card, Mackinlay, Shneiderman (1999)
Klein, Moon, & Hoffman (2006)
Liu & Stasko (2010)
Pirolli & Card (2005)
Spence (2007)

Amar, Eagan, & Stasko (2005)
Andrienko & Andrienko (2006)
Buja et al. (1996)
Casner (1991)
Chi & Riedl (1998)
Chuah & Roth (1996)
Dix & Ellis (1998)
Gotz & Zhou (2008)
Keim (2002)
Lee et al. (2006)
Raskin (1990)
Roth & Mattis (1990)
Shneiderman (1996)
Tweedie (1997)
Valiati et al. (2006)
Wehrend & Lewis (1990)
Yi, Stasko, et al. (2007)
Process

• reflective synthesis
• open coding

1. read and think
2. code: arrange and abstract
3. simplify and repeat...

open coding of literature rather than empirical study with human subjects
Process

Iteration

27 terms
Final design space: three axes

- why, what, how

why?
- present
- discover
- enjoy

search
- target known
  - lookup
  - browse
- target unknown
  - locate
  - explore

why?
- consume

produce
- encode
- select
- navigate
- arrange
- derive
- record
- manipulate
- introduce

what?
- [input]
- [output]
- manipulate
- introduce
- aggregate
- filter
- change
- record
- derive
- import
- navigate
- select
- inject
- introduce

query
- identify
- compare
- summarize
- lookup
- locate
- browse
- consume

[location unknown]
Mapping terms

compare [5, 31, 42, 50, 57, 66, 72, 73, 83]* [40], compare (within a relation vs. across / between relations) [59, 78]*, relation seeking [5]*, read comparison [11]*, making comparisons [10]*, [76], discriminate [42]*, associate [57]*

Our 27 terms (left column)

navigate [23, 64, 75]* [40, 44, 52, 76, 80], focus [10, 15]* [13], details-on-demand [11, 61]*, [13], flip through [13] zoom [10, 11, 15, 19, 29, 42, 50, 57, 61, 82]* [13, 44, 80], pan [10, 19, 42, 50, 57, 82]* [80], elaborate [50, 82]*, abstract [50, 82]*, change (range) [19]*, drill down [15]*, maneuver / navigate [66]*, rotate [13, 80] revisit [19, 37]*

Terms from 30 extant classification systems (right column)

Mapping our Vocabulary to Previous Work

Table 1: lookup table of task vocabulary

Our 27 terms

<table>
<thead>
<tr>
<th>Terms from extant classification systems</th>
<th>Additional references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our 27 terms</td>
<td></td>
</tr>
</tbody>
</table>
Constructing a Typology

Stages of Action 1 +
Gulf of Goal Formation 2,
Distributed Cognition 3, Sensemaking 4,
Play Theory 5, Nested Model 6

1 Norman (1988)
2 Lam (TVCG 2008)
3 e.g. Hollan et al. (2000)
4 e.g. Pirolli and Card (2005)
6 Munzner (TVCG 2009)
Assessment & adoption

• **descriptive** power
  – analyze & compare task sequences, clarify means and ends

• **generative** power
  – early stages of problem-driven work: abstracting & requirements gathering

• **evaluative** power
  – codeset for field studies, task set for lab studies

• adoption
  – hundreds of papers
# How?

## Encode

- **Arrange**
  - Express
  - Separate
- **Order**
  - Align
- **Use**

## Manipulate

- **Map**
  - from categorical and ordered attributes
  - Color
    - Hue
    - Saturation
    - Luminance
  - Size, Angle, Curvature, ...
- **Shape**
  - + ● ■ △
- **Motion**
  - Direction, Rate, Frequency, ...

## Facet

- **Change**
- **Select**
- **Partition**
- **Navigate**
- **Superimpose**

## Reduce

- **Filter**
- **Aggregate**
- **Embed**

---

Visualization Analysis and Design. Munzner.

*CRC/Routledge, AK Peters Visualization Series, 2014.*
### Analyze
- **Consume**
  - Discover
  - Present
  - Enjoy
- **Produce**
  - Annotate
  - Record
  - Derive

### Search
<table>
<thead>
<tr>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td>&quot;•••&quot; Lookup</td>
</tr>
<tr>
<td>Location unknown</td>
<td>&quot;••&quot; Locate</td>
</tr>
</tbody>
</table>

### Query
- **Identify**
- **Compare**
- **Summarize**

### Why?
- **All Data**
  - Trends
  - Outliers
  - Features
- **Attributes**
  - One
    - Distribution
  - Many
    - Dependency
    - Correlation
    - Similarity
- **Network Data**
  - Topology
    - Paths
- **Spatial Data**
  - Shape
Bridging From Goals to Tasks with Design Study Analysis Reports

Heidi Lam

Melanie Tory
@vizstudylady

Bridging From Goals to Tasks With Design Study Analysis Reports.

http://www.cs.ubc.ca/labs/imager/tr/2017/GoalsToTasks/

design space: analysis goals
source material: analysis reports extracted from design study papers
# Summary: Multiple design spaces

<table>
<thead>
<tr>
<th>Design Space</th>
<th>Open Coding Source Material</th>
<th>Sampling Strategy</th>
<th>Reflective Synthesis Timing</th>
<th>Vis Research Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>timeline</strong></td>
<td>visual encoding</td>
<td>standalone timelines</td>
<td>assembled corpus</td>
<td>early</td>
</tr>
<tr>
<td><strong>genEpi</strong></td>
<td>visual encoding</td>
<td>figures from papers</td>
<td>stratified random sampling with topic clusters</td>
<td>-</td>
</tr>
<tr>
<td><strong>wrangling</strong></td>
<td>activities</td>
<td>software from repos</td>
<td>diversity criteria</td>
<td>late</td>
</tr>
<tr>
<td><strong>abstract tasks</strong></td>
<td>tasks from papers</td>
<td>comprehensive</td>
<td>early</td>
<td>terms: thorough mapping</td>
</tr>
</tbody>
</table>

- **timeline**: visual encoding
  - Source Material: standalone timelines
  - Sampling Strategy: assembled corpus
  - Reflective Synthesis Timing: early
  - Vis Research Literature: some source material

- **genEpi**: visual encoding
  - Source Material: figures from papers
  - Sampling Strategy: stratified random sampling with topic clusters
  - Reflective Synthesis Timing: -
  - Vis Research Literature: -

- **wrangling**: activities
  - Source Material: software from repos
  - Sampling Strategy: diversity criteria
  - Reflective Synthesis Timing: late
  - Vis Research Literature: terms: light mapping

- **abstract tasks**: tasks from papers
  - Sampling Strategy: comprehensive
  - Reflective Synthesis Timing: early
  - Vis Research Literature: terms: thorough mapping
## 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>
<tr>
<td>abstract tasks</td>
<td>widespread adoption</td>
<td>widespread adoption</td>
<td>same (concise)</td>
<td>widespread adoption</td>
</tr>
</tbody>
</table>
Design spaces: How to assess? Larger context: theory types

- Ben Shneiderman, *Designing the User Interface*: descriptive, explanatory, prescriptive, predictive
  - 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
More information

• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#stanf22

• book
  http://www.cs.ubc.ca/~tmm/vadbook

• full courses, papers, videos, software, talks
  http://www.cs.ubc.ca/group/infovis
  http://www.cs.ubc.ca/~tmm

@tamaramunzner