Developing Design Spaces for Visualization

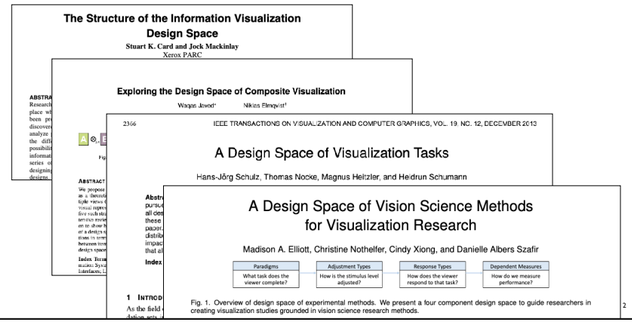
Tamara Munzner
Department of Computer Science
University of British Columbia

Stanford HCI Seminar
4 March 2022, virtual

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

@tamaramunzner

Design spaces: Continuing theme



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**?

- Michel Beaudoin-Lafon, *Designing Interaction, not Interfaces. AVI 2004.*
  - **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

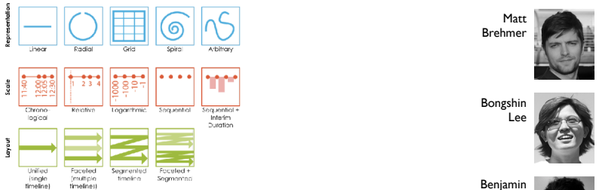
Design spaces: How to **create**?

- **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

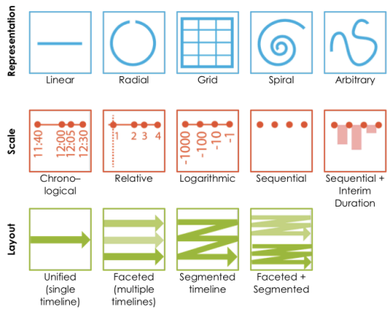
Matt Brehmer
Bongshin Lee
Benjamin Bach
Nathalie Henry Riche

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

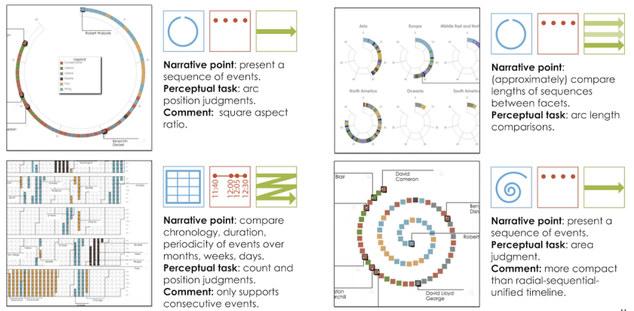
Timelines Revisited: A Design Space and Considerations for Expressive Storytelling
Brehmer, Lee, Bach, Henry Riche, Munzner. IEEE TVCG 23(9):2151-2164

Design space with three axes

- representation
- scale
- layout



Combinations: Characterize narrative, perceptual



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

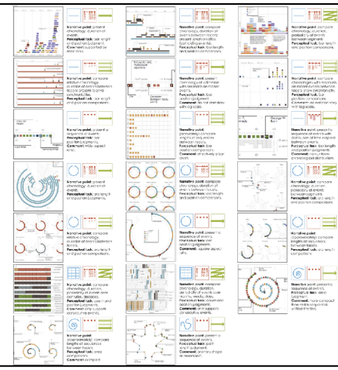
Narrative point: (approximately) compare lengths of sequences between facets.
Perceptual task: arc length comparisons.

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

Anamaria Crisan
@amcrisan

Jenn Gardy
@jennifergardy

https://amcrisan.github.io/gevit

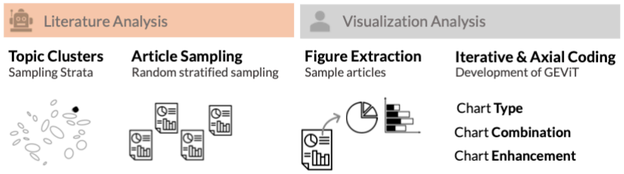
A systematic method for surveying data visualizations and a resulting genomic epidemiology visualization typology: GEViT.
Crisan, Gardy, Munzner. Oxford Bioinformatics 35(10):1668-1676, 2018.

Propose typology creation method: mixed qual and quant

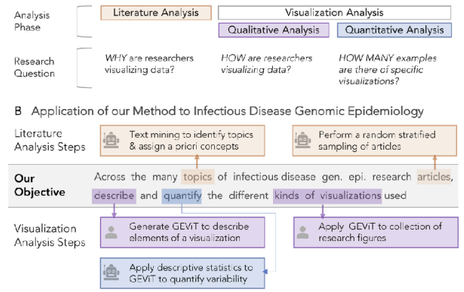
- Analyzed research articles
- Some analyses are automated ( ) and others are manual ( )

Use method to develop typology in specific domain

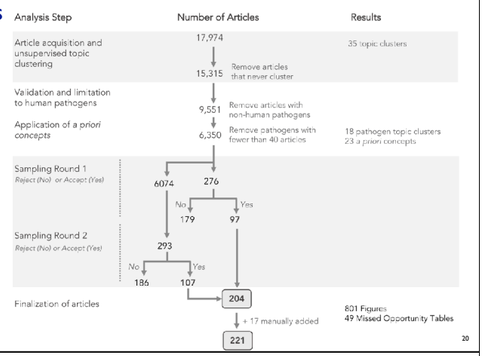
- Developed a **G**enomic **E**pidemiology **Vis**ualization **T**ypology (GEViT)



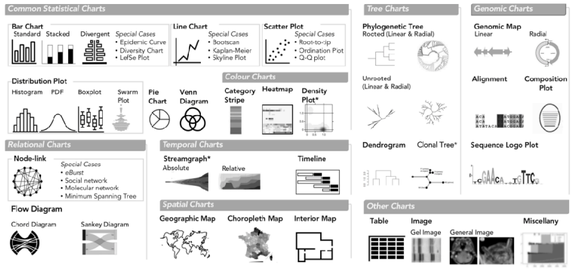
Domain prevalence design space



By the numbers



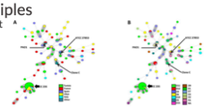
Design space axis: Chart types used in genEpi



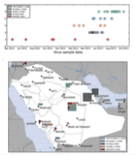
Design space axis: Chart combinations of heterogeneous data

Spatially Aligned
Horizontal / Vertical Alignment
20%

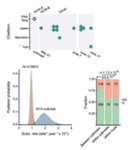
Small Multiples
Chart Alignment
17%



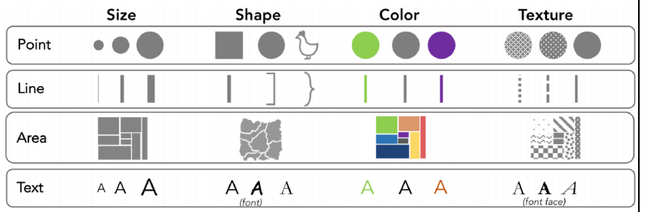
Visually Aligned
Colour / Shape Alignment
14%



Unaligned
9%

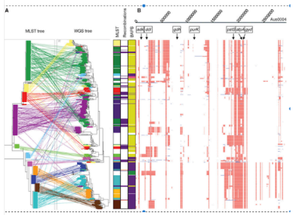


Design space axis: Enhancement choices, atop base chart types

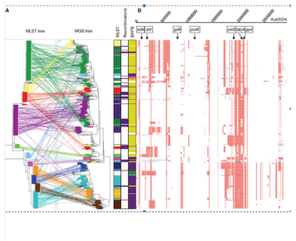


Current Practice >80% of all figures have some enhancement

GEViT example



GEViT example

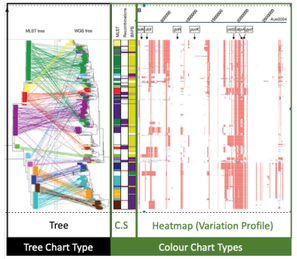


Visualization Breakdown

Literature Analysis (why)

- ***Pathogen**: Enterococcus faecium*

GEViT example



Visualization Breakdown

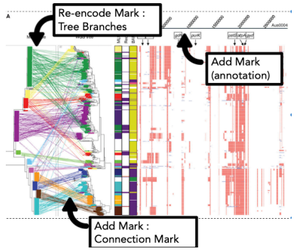
Literature Analysis (why)

- ***Pathogen**: Enterococcus faecium*

Visualization Analysis (how)

Chart Type	Tree (Rooted Phylogenetic Tree) Category Stripe Heatmap (Variation Profile)
Chart Combination	Spatially Aligned (*horizontal*)

GEViT example



Visualization Breakdown

Literature Analysis (why)

- ***Pathogen**: Enterococcus faecium*

Visualization Analysis (how)

Chart Type	Tree (Rooted Phylogenetic Tree) Category Stripe Heatmap (Variation Profile)	
Chart Combination	Spatially Aligned (*horizontal*)	
Chart Enhancement	Re-encode Marks	Tree – *branches*
	Add Marks	Tree - *Connection Marks*
	Add Mark (*unstructured*)	Heatmap – *Textboxes*

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

Anamaria Crisan
@amcrisan

Shannah Fisher

Jenn Gardy
@jennifergardy

https://github.com/amcrisan/GEViTRec

GEViTRec: Data Reconnaissance Through Recommendation Using a Domain-Specific Visualization Prevalence Design Space.
Crisan, Fisher, Gardy, Munzner. IEEE TVCG to appear, 2022.

Data Wrangling

An Actionable Framework for Multi-Table Data Wrangling

From an Artifact Study of Computational Journalism

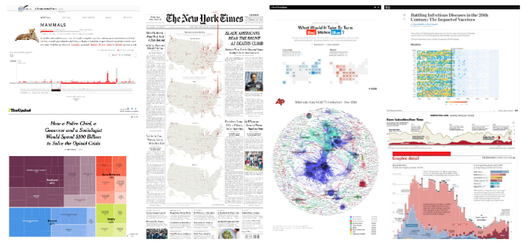
Steve
Kasica
@stevekasica

Charles
Berret
@cberret

http://www.cs.ubc.ca/group/infovis/pubs/2020/table-scraps/

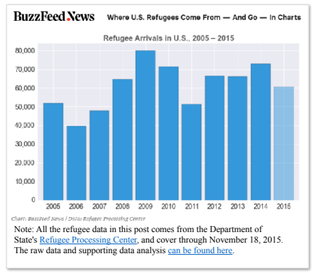
An Actionable Framework for Multi-Table Data Wrangling From an Artifact Study of Computational Journalism.
Kasica, Berret, Munzner. IEEE TVCG 27(2):957-966 2021. (Proc. InfoVis 2020).

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.

Process overview

What are the wrangling practices of journalists with programming skills?

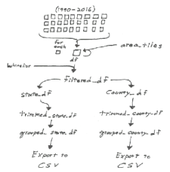
Technical observation: Repo selection | Qualitative coding

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

Taxonomies of data wrangling in computational journalism - initial

Which practices align with or diverge from existing characterizations?

Literature search: Term harmonization | Gap discovery

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 - 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

Taxonomies of data wrangling in computational journalism

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

Reflective Synthesis

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

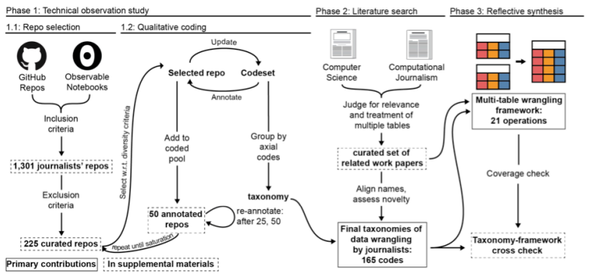
Taxonomies of data wrangling in computational journalism

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

Reflective Synthesis

Multi-table framework of data wrangling

By the numbers



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

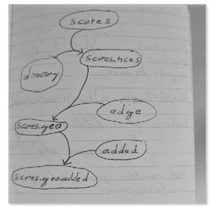
Actions
- Import
- Clean
- Merge
- Profile
- Drive
- Transform
- Export

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

Key finding: journalists use many, many tables

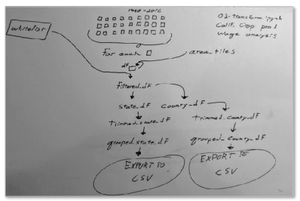
Key finding: journalists use many, many tables

- workflow complexity varies greatly



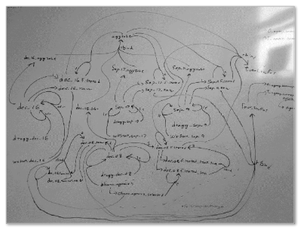
Key finding: journalists use many, many tables

- 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



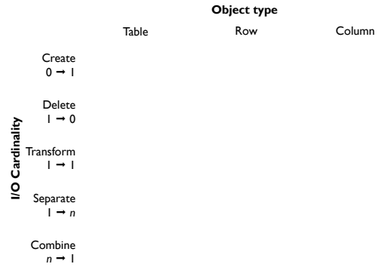
Two axes of multi-table wrangling design space

Two axes of multi-table wrangling design space

Object type

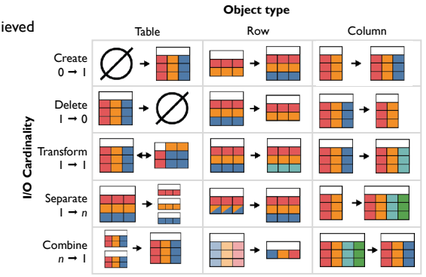
Table | Row | Column

Two axes of multi-table wrangling design space



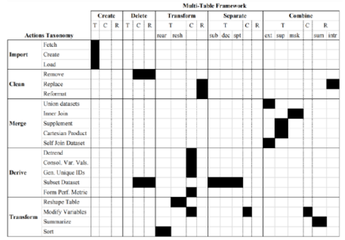
Multi-table data wrangling design space

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



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

Matt Brehmer
@mattbrehmer

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

A Multi-Level Typology of Abstract Visualization Tasks.
Brehmer, Munzner. IEEE TVCG 19(12):2376–2385 (Proc. InfoVis 2013).

Task abstraction: Gap

low level of abstraction
e.g. "retrieve value"

high level of abstraction
e.g. "integration of insight"

Previous Work

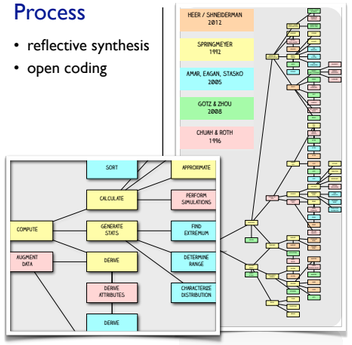
Classifying Tasks, Goals, Intentions, Objectives, Activities, Interactions

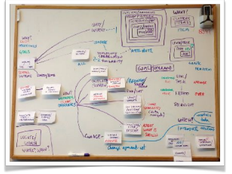
Amar, Eagan, & Stasko (2005)
Andrienko & Andrienko (2006)
Buja et al. (1996)
Casner (1991)
Chi & Riedl (1998)
Chuah & Roth (1996)
Dix & Ellis (1998)
Gotz & Zhou (2008)
Keim (2002)
Lee et al. (2006)
Raskin (1990)
Roth & Mattis (1990)
Shneiderman (1996)
Tweedie (1997)
Valiati et al. (2006)
Ward & Yang (2004)
Wehrend & Lewis (1990)
Yi, Stasko, et al. (2007)
Zhou & Feiner (1998)
Heer & Shneiderman (2012)
Mullins & Treu (1993)
Pike, Stasko, et al. (2009)
Springmeyer et al. (1992)
RE Roth (2012)
Amar & Stasko (2004)
Card, Mackinlay, Shneiderman (1999)
Klein, Moon, & Hoffman (2006)
Liu & Stasko (2010)
Pirolli & Card (2005)
Spence (2007)

A mid-level gap?
Meyer, Sedlmair, & Munzner (BELIV 2012)

Process

- reflective synthesis
- open coding

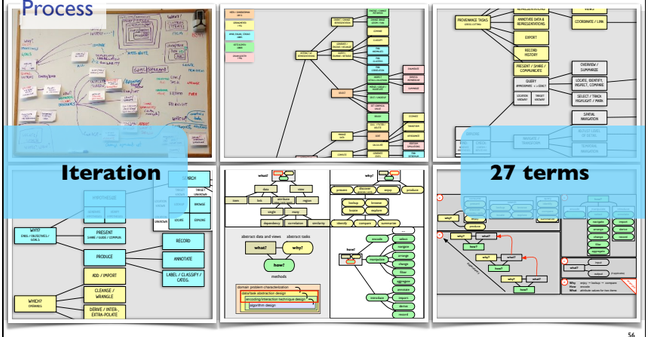




1. read and think
2. *code*: arrange and abstract
3. simplify and repeat…

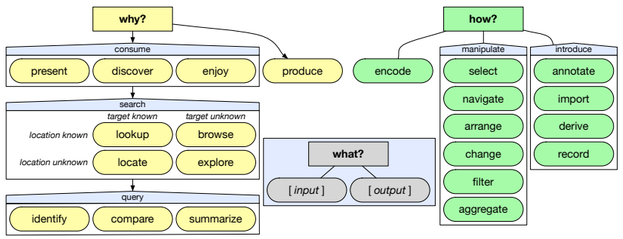
open coding of literature rather than empirical study with human subjects

Process



Final design space: three axes

- why, what, how



Mapping terms

→ compare — *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]*

Mapping our Vocabulary to Previous Work

→ navigate — *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]*

← **Table 1**: lookup table of

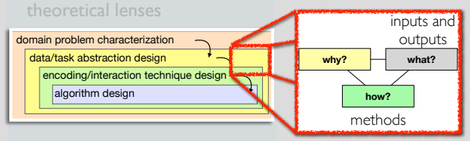
Our **27** terms *(left column)*

Terms from **30** extant classification systems *(right column)*

Directionality

Bottom-Up
previous classification systems
Top-Down
theoretical lenses

Constructing a Typology



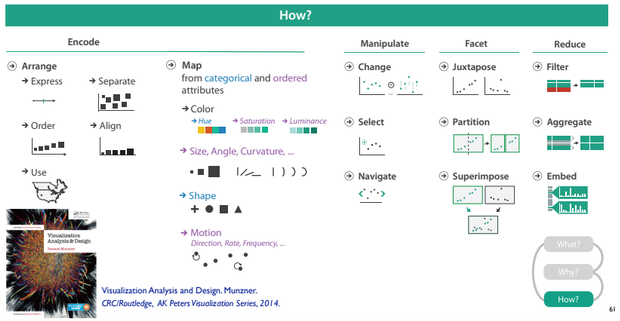
[1] Norman (1988)
[2] Lam (TVCG 2008)
[3] e.g. Hollan et al. (2000)
[4] e.g. Pirolli and Card (2005)
[5] Stephenson (1967) ;Toms (2000)
[6] Munzner (TVCG 2009)

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

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

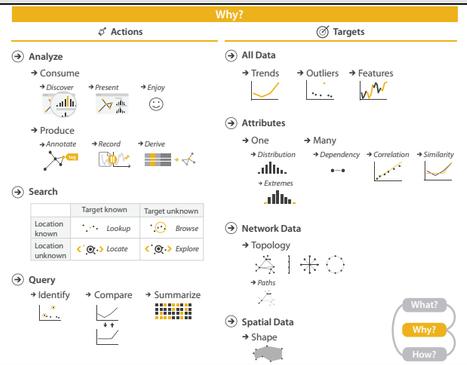
VAD Book: Visualization Analysis and Design



Visualization Analysis and Design. Munzner. *CRC/Routledge, AK Peters Visualization Series, 2014.*

VAD Book

Visualization Analysis and Design. Munzner. *CRC/Routledge, AK Peters Visualization Series, 2014.*



Specificity / # Populations	Explore	Describe	Explain	Confirm
Single	Discover Observation	Describe Observation	Identify Main Cause	Collect Evidence
Multiple		Compare Entities	Explain Differences	Evaluate Hypothesis

Bridging From Goals to Tasks

with Design Study Analysis Reports

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

Heidi Lam

Melanie Tory
@vizstudylady

design space: analysis goals

source material: analysis reports extracted from design study papers

Bridging From Goals to Tasks With Design Study Analysis Reports.
Lam,Tory, Munzner. IEEE TVCG 24(1):435-445 (Proc. InfoVis 2017).

Summary: Multiple design spaces

Design Space	Open Coding Source Material	Sampling Strategy	Reflective Synthesis Timing	Vis Research Literature
timeline visual encoding	standalone timelines	assembled corpus	early	some source material
genEpi visual encoding	figures from papers	stratified random sampling with topic clusters	-	-
wrangling activities	software from repos	diversity criteria	late	terms: light mapping
abstract tasks	tasks from papers	comprehensive	early	terms: thorough mapping

Summary: Multiple design spaces

Design Space	Descriptive Power	Generative Power	Descriptive vs Generative	Evaluative Power
timeline visual encoding	validated against test set	software implementation of authoring system, used to create example gallery/videos	analysis to characterize viable subset	
genEpi visual encoding	systematic method yields comprehensive coverage	software implementation of automatic recommender (followup)	*same (detailed)*	
wrangling activities	high precision, gaps / divergence found for domain	concise framework (followup implementation TBD)	develop entirely new framework	
abstract **tasks**	widespread adoption	widespread adoption	*same (concise)*	widespread adoption

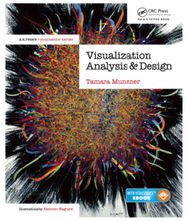
Design spaces: How to assess? Larger context: theory types

- 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

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



Visualization Analysis and Design. Munzner.
CRC Press, AK Peters Visualization Series, 2014.