

Entry Points to Visualization: Different Methods for Different Problems



Tamara Munzner
Department of Computer Science
University of British Columbia

Viz@UBC Kickoff: Lunchtime Lecture Series
12 March 2019

www.cs.ubc.ca/~tmm/talks.html#vizatubc19-entry

@tamaramunzner

Visualization (vis) defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- human in the loop needs details about data
 - entry point: exploratory data analysis
 - don't know exactly what questions to ask in advance
 - entry point: presentation of known results
 - refining model, trustbuilding/monitoring, mixed-initiative
- external representation: perception vs cognition
- intended task, measurable definitions of effectiveness

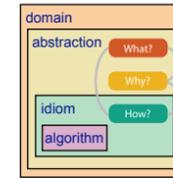


more at:
Visualization Analysis and Design, Chapter 1.
Munzner. AK Peters Visualization Series, CRC Press, 2014.

Analysis framework: Four levels, three questions

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

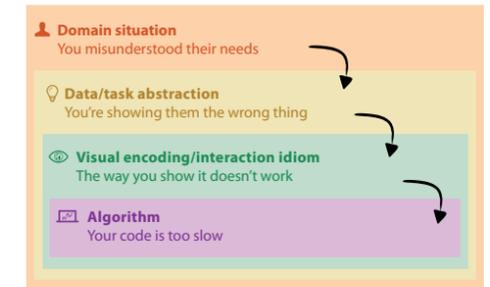
- domain situation
 - who are the target users?
- abstraction
 - translate from specifics of domain to vocabulary of visualization
 - what is shown? data abstraction
 - why is the user looking at it? task abstraction
- idiom
 - how is it shown?
 - visual encoding idiom: how to draw
 - interaction idiom: how to manipulate
- algorithm
 - efficient computation



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

Different threats to validity at each level

- cascading effects downstream



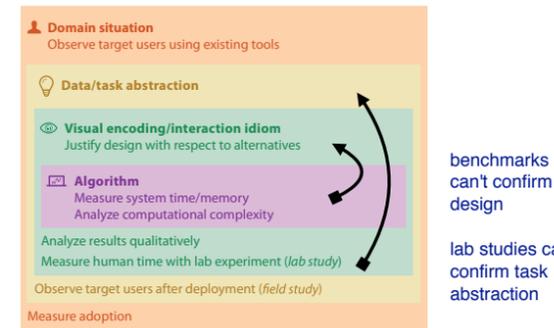
Different methods for different problems, from different fields

- interdisciplinary, mix of qual and quant approaches (typically)

anthropology/ethnography	Domain situation Observe target users using existing tools	qual	problem-driven work
	Data/task abstraction		
design	Visual encoding/interaction idiom Justify design with respect to alternatives	qual	
computer science	Algorithm Measure system time/memory Analyze computational complexity	quant	
psychology	Analyze results qualitatively Measure human time with lab experiment (lab study)	qual quant	
anthropology/ethnography	Observe target users after deployment (field study) Measure adoption	qual quant	technique-driven work

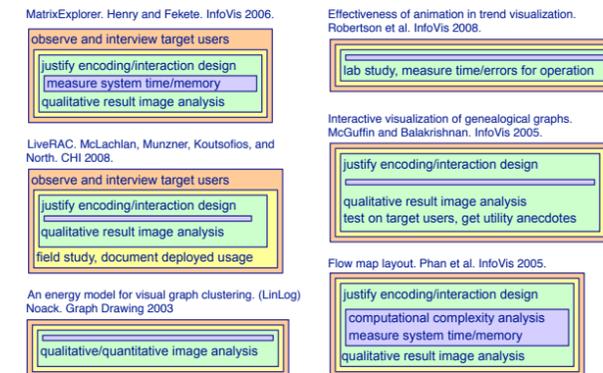
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Method mismatches: Common problem

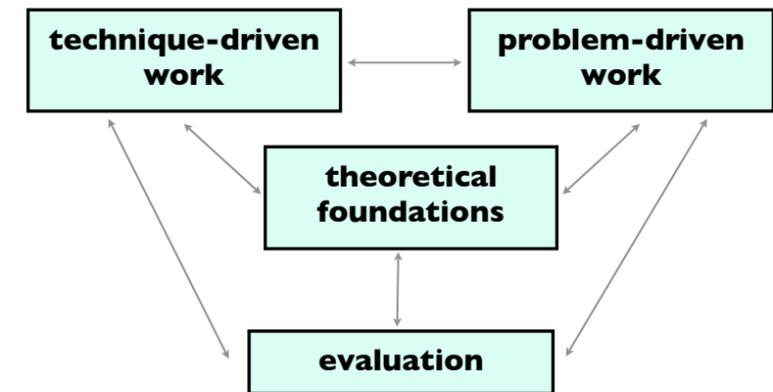


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis examples: Single paper includes only subset of methods



Different angles of attack for different problems



Technique-driven work

- scalable algorithms & systems
 - typical evaluation: computational benchmarks
- new layout & interaction techniques
 - typical evaluation: usage scenarios
 - typical evaluation/characterization: controlled experiments on human subjects

Technique-driven: Graph/network drawing

Daniel Archambault **David Auber (Bordeaux)** <https://youtu.be/AWXAe8zvtk8>

Benjamin Renoust **Guy Melançon (Bordeaux)** <https://youtu.be/QOinHSuUV6k>

TopoLayout
SPF
Grouse
GrouseFlocks
TugGraph

Detangler

Technique-driven: Tree drawing

Zipeng Liu **Shing Hei Zhan**

Treejuxtaposer <https://youtu.be/GdaPi8a9QEo>

Aggregated Dendrograms <https://youtu.be/2SLcz7KNLJw>

Evaluation experiments: Graph/tree drawing

Dmitry Nekrasovski **Adam Bodnar** **Joanna McGrenere**

Jessica Dawson **Joanna McGrenere**

lab study led to "focus+context" idiom disenchantment

Stretch and squish navigation

Search set model of path tracing

1 qualitative study: coding observational video
2 create & implement behavioral model
3 multiple regression to untangle factor relationships

Technique-driven: Dimensionality reduction

Stephen Ingram

Glimmer

DimStiller

QSNE

Dimensionality reduction for documents

- derive low-dimensional target space from high-dimensional measured space

Task 1: In HD data → Out 2D data

Task 2: In 2D data → Out Scatterplot Clusters & points

Task 3: In Scatterplot Clusters & points → Out Labels for clusters

What? Why? How?

⊕ In High-dimensional data → Produce ⊕ Discover ⊕ Encode
⊕ Out 2D data → Derive ⊕ Explore ⊕ Navigate
⊕ Out Clusters & points → Identify ⊕ Select

Evaluation experiments: Dimensionality reduction

Melanie Tory **Michael Sedlmair** **Melanie Tory**

traditional user study: many people for short time, few datasets

data studies: many datasets, few people for long time (experts qual+quant coding)

Points vs landscapes for dimensionally reduced data

Taxonomy of cluster separation factors

Guidance on DR & scatterplot choices

Evaluation in the field: Dimensionality reduction

DR in the Wild

interview study & qualitative coding led to task abstractions: specific to data type, agnostic to domain

Matt Brehmer **Michael Sedlmair** **Melanie Tory** **Stephen Ingram**

Problem-driven work

- "design studies"
 - in collaboration with target users
 - real data, real tasks
 - intensive requirements analysis
 - iterative refinement
 - deploy tools/systems
 - typical evaluation: field studies
- my strategy: opportunistic collaboration
 - many domains
 - both industrial and academic partners

Problem-driven: Genomics

Aaron Barsky Jenn Gardy Robert Kincaid (Microbio) (Agilent)

Cerebral <https://youtu.be/76HhG1FQngI>

Miriam Meyer Hanspeter Pfister (Harvard)

MizBee <https://youtu.be/86p7brwuz2g> MulteeSum, Pathline

Problem-driven: Genomics, fisheries

Joel Ferstay Cydney Nielsen (BC Cancer)

Variant View https://youtu.be/AHDnv_qMXxQ

Maryam Booshehrian

Torsten Moeller (SFU)

Vision <https://youtu.be/h0kHoS4VYmk>

Problem-driven: Tech industry

Heidi Lam Diane Tang (Google)

Session Viewer: web log analysis <https://youtu.be/T4MaTZd56G4>

Peter McLachlan Stephen North (AT&T Research)

LiveRAC: systems time-series <https://youtu.be/ld0c3H0VSkw>

methods reflection: staged model of access to target users

Problem-driven: Building energy mgmt, journalism

Matt Brehmer Kevin Tate (Pulse/EnerNOC)

Energy Manager

redesign success: industrial swdev resources committed

Matt Brehmer Stephen Ingram Jonathan Stray (Assoc Press)

Overview <https://vimeo.com/71483614>

Matthew Brehmer @mattbrehmer

Stephen Ingram @FroweFace

Jonathan Stray @jonathanstray

Tamara Munzner @tamaramunzner

Overview
The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists

<http://www.cs.ubc.ca/labs/imager/tr/2014/Overview/>

<https://www.overviewdocs.com>

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists. Brehmer, Ingram, Stray, and, Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2014), 20(12):2271-2280, 2014.

From design

Case Study	#1
Document Collection	4,500 pages from FOIA
Question	What did security contractors do during Iraq war?

From design, to deploy, ...

Case Study	#1	#2	#3	#4	#5	#6
Document Collection	4,500 pages from FOIA	5,996 emails from FOIA	8,680 pages from FOIA	1,278 survey comments	4,653 emails from FOIA	1,680 bills
Question	What did security contractors do during Iraq war?	Were municipal police funds mismanaged?	Were Paul Ryan's campaign statements hypocritical?	What is the gun ownership debate about?	Was gov't response to emergency incident effective?	Did gov't fail to pass bills addressing police misconduct?

... to redesign, to reflect on task abstractions...

Case Study	#1	#2	#3	#4	#5	#6
Document Collection	4,500 pages from FOIA	5,996 emails from FOIA	8,680 pages from FOIA	1,278 survey comments	4,653 emails from FOIA	1,680 bills
Question	What did security contractors do during Iraq war?	Were municipal police funds mismanaged?	Were Paul Ryan's campaign statements hypocritical?	What is the gun ownership debate about?	Was gov't response to emergency incident effective?	Did gov't fail to pass bills addressing police misconduct?

find the needle in the haystack

prove haystack contains no needles!

THOUSANDS OF DOCUMENTS

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists
IEEE InfoVis '14

M. Brehmer, S. Ingram, J. Stray, & T. Munzner
University of British Columbia (brehmer@img.ubc.ca)
Columbia Journalism School & Associated Press
jonathanstray@gmail.com
img.ubc.ca

Problem-driven: In-car networks, e-commerce

Michael Sedlmair

RelEx (BMW) <https://youtu.be/89lsQXc6Ao4>

Kim Dextras-Romagnino

latest work: Segmentifier (Mobify): e-commerce clickstreams

sneak preview video

Theoretical foundations: Methodology

domain
abstraction
idiom
algorithm

Nested Model

- Visual Encoding Pitfalls
 - Unjustified Visual Encoding
 - Hammer In Search Of Nail
 - 2D Good, 3D Better
 - Color Cacophony
 - Rainbows Just Like In The Sky
- Strategy Pitfalls
 - What I Did Over My Summer
 - Least Publishable Unit
 - Dense As Plutonium
 - Bad Slice and Dice

Papers Process & Pitfalls

Anamaria Crisan

Design Study Methodology

Michael Sedlmair Miriah Meyer

Regulatory & Organizational Constraints

Michael Sedlmair

Miriam Meyer

Tamara Munzner @tamaramunzner

Design Study Methodology

Reflections from the Trenches and from the Stacks

<http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/>

Design Study Methodology: Reflections from the Trenches and from the Stacks. Sedlmair, Meyer, Munzner. IEEE Trans. Visualization and Computer Graphics 18(12):2431-2440, 2012 (Proc. InfoVis 2012).

Methodology for problem-driven work

- definitions
- 9-stage framework
- 32 pitfalls & how to avoid them
- comparison to related methodologies

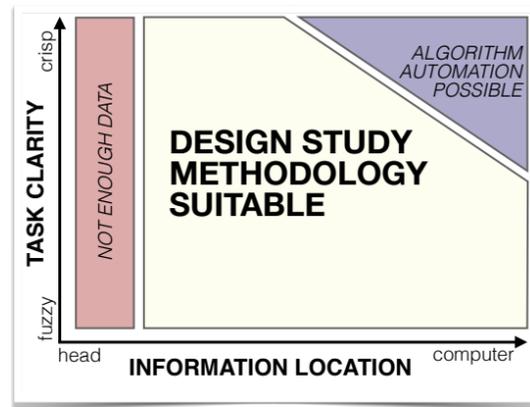
Lessons learned from the trenches: 21 between us

Cerebral genomics MizBee genomics Pathline genomics MulteeSum genomics Vision fisheries management QuestVis sustainability WiKeVis in-car networks

MostVis in-car networks Car-X-Ray in-car networks ProgSpy2010 in-car networks RelEx in-car networks Cardigram in-car networks AutobahnVis in-car networks VisTra in-car networks

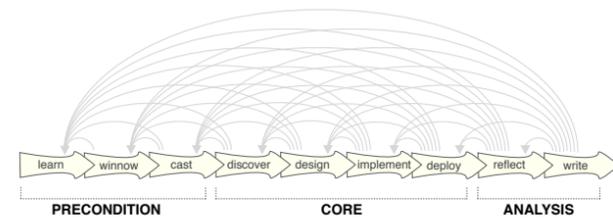
Constellation linguistics LibVis cultural heritage Gaidants multicast SessionViewer web log analysis LiveRAC server hosting PowerSetViewer data mining LastHistory music listening

Design study methodology: definitions



33

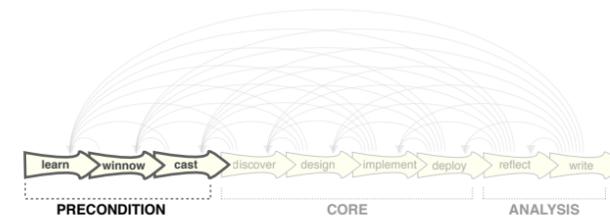
9 stage framework



34

9-stage framework

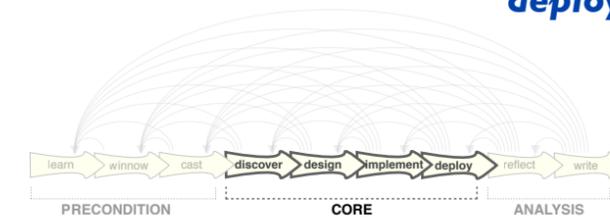
**learn
winnow
cast**



35

9-stage framework

**discover
design
implement
deploy**

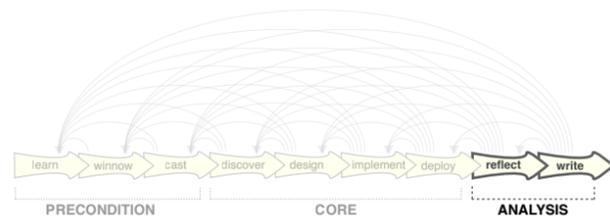


36

9-stage framework

- guidelines: confirm, refine, reject, propose

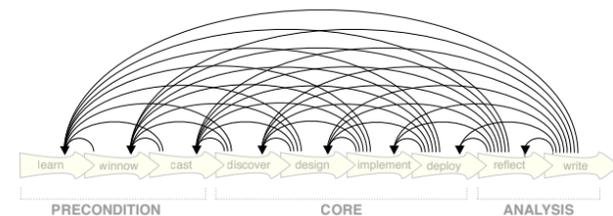
**reflect
write**



37

9-stage framework

iterative



38

Design study methodology: 32 pitfalls

- and how to avoid them

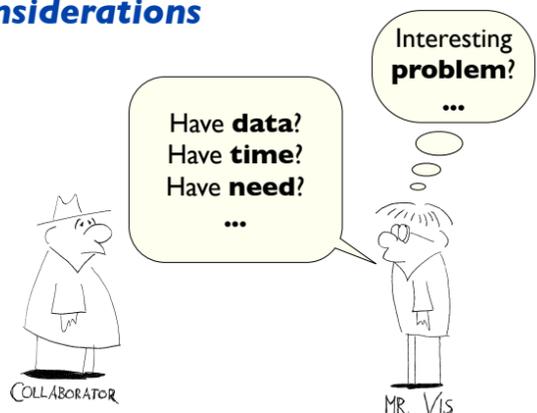
PF-1	premature advance: jumping forward over stages	general
PF-2	premature start: insufficient knowledge of vis literature	learn
PF-3	premature commitment: collaboration with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow

39



40

considerations



41

Design study methodology: 32 pitfalls

PF-21	mistaking technique-driven for problem-driven work	design
PF-22	nonrapid prototyping	implement
PF-23	usability: too little / too much	implement
PF-24	premature end: insufficient deploy time built into schedule	deploy
PF-25	usage study not case study: non-real task/data/user	deploy
PF-26	liking necessary but not sufficient for validation	deploy
PF-27	failing to improve guidelines: confirm, refine, reject, propose	reflect
PF-28	insufficient writing time built into schedule	write
PF-29	no technique contribution ≠ good design study	write
PF-30	too much domain background in paper	write
PF-31	story told chronologically vs. focus on final results	write
PF-32	premature end: win race vs. practice music for debut	write

42

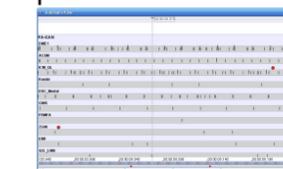
METAPHOR Horse Race vs. Music Debut



43

EXAMPLE FROM THE TRENCHES Don't step on your own toes!

First design round published



AutobahnVis 1.0 [Sedlmair et al., Smart Graphics, 2009]

Subsequent work not stand-alone paper

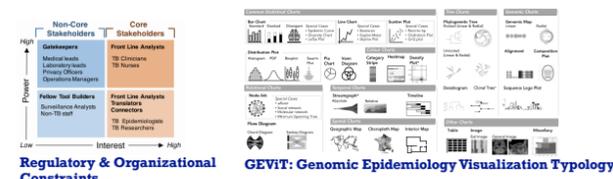
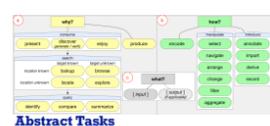


AutobahnVis 2.0 [Sedlmair et al., Information Visualization 10(3), 2011]

44

Theoretical foundations: Typologies

T F P E



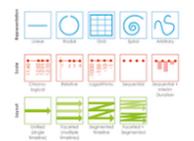
45

Curation & Presentation: Timelines

T F P E



TimeLineCurator <https://vimeo.com/123246662>



Timelines Revisited [timelinesrevisited.github.io/](https://github.com/timelinesrevisited)



46

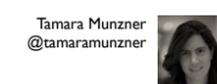
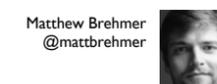


TimeLineCurator Interactive Authoring of Visual Timelines from Unstructured Text

<http://about.timelinecurator.org>

<http://timelinecurator.org>

TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text. Fulda, Brehmer, Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. IEEE VAST 2015) 22(1):300-309, 2015.

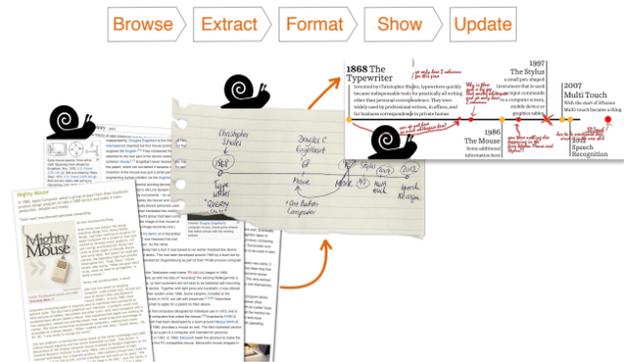


47

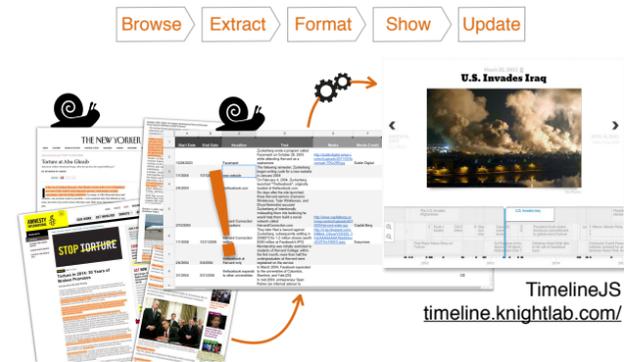
TimeLineCurator visual & browser-based

<https://vimeo.com/jofu/tlc>

Manual creation process



Structured creation process



Timeline authoring model

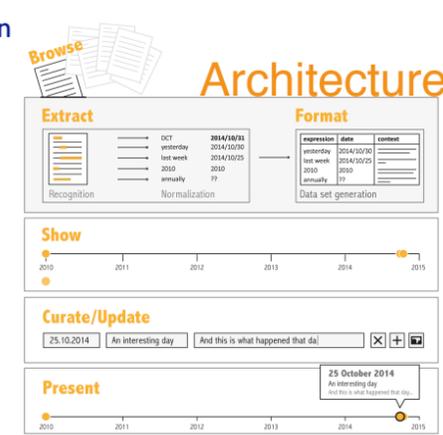
- time required for each task

	Browse	Extract	Format	Show	Update
Manual Drawing	slow	slow		slow	slow
Structured Creation	slow	slow	slow	automated	fast
TimeLine Curator	fast	automated	automated	fast	fast

51

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

- sexy 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

speculative browsing

<https://vimeo.com/jofu/tlc>

Curation & Presentation: Timelines

TimeLineCurator
<https://vimeo.com/123246662>

Timelines Revisited
[timelinesrevisited.github.io/](https://github.com/timelinesrevisited)

55

Presentation: Geometry Center math vis videos

Outside In

The Shape of Space

Stuart Levy, Mark Phillips, Delle Maxwell

56



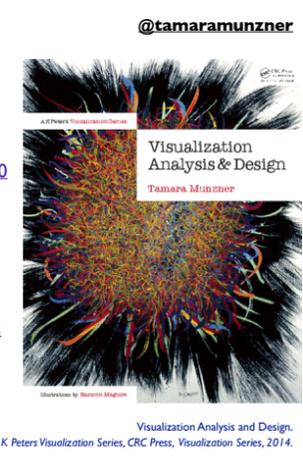
Visualization entry points

- goals
 - exploratory data analysis
 - presentation
 - curation / authoring
 - methods
 - algorithm development
 - system building & software development
 - lab studies with human subjects
 - field studies with human subjects
- 58

Teaching

Visualization Analysis and Design

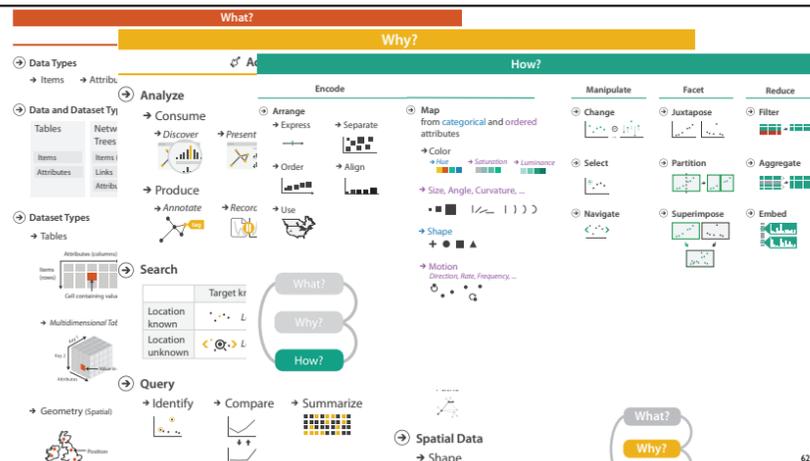
- book page
<http://www.cs.ubc.ca/~tmm/vadbook>
 - 20% promo code for book+ebook combo: HVN17
 - <http://www.crcpress.com/product/isbn/9781466508910>
 - free to read online within UBC
<http://resolve.library.ubc.ca/cgi-bin/catsearch?bid=7678980>
- slide decks at many talk lengths (1, 2, 3, 6, 8+ hrs), some w/ videos
<http://www.cs.ubc.ca/~tmm/talks.html#vadallslides>



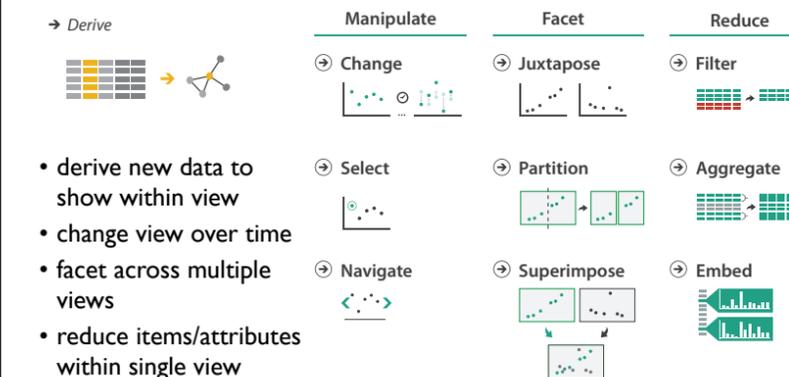
Visualization Analysis & Design

IEEE VIS 2014 Tutorial
Video Preview

Tamara Munzner
Department of Computer Science
University of British Columbia



How to handle complexity: 4 families of strategies



Visualization Teaching at UBC: Me

- Computer Science grad
 - open to all students, no CS prereqs, non-programming project options available
 - tooling not taught (most use D3 or R)
 - Computer Science ugrad
 - coming in January 2020, 4th year CS majors
 - tooling: D3
 - Data Science
 - tooling: R
 - Journalism
 - tooling: Tableau
 - Ed Psych, Forestry, Geography, iSchool, Psychology
<https://dfp.ubc.ca/initiatives/viz-ubc/visualization-courses>
- 64

Visualization Teaching Across UBC

- many other visualization-focused courses
 - Ed Psych
 - Forestry
 - Geography
 - iSchool
 - Psychology
- initial list compiled <https://dfp.ubc.ca/initiatives/viz-ubc/visualization-courses>
 - please contact vizatubc-info@cs.ubc.ca with additions/corrections!
- still todo: compile list of courses with significant visualization content

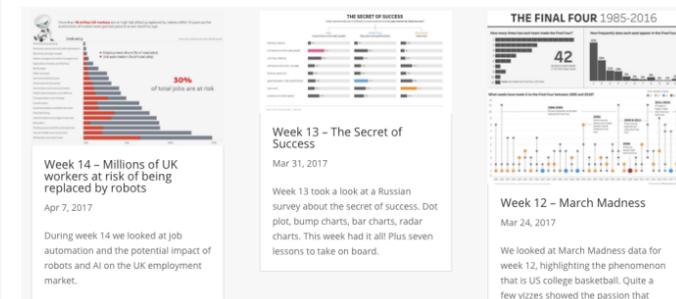
Engaging with visualization teaching

- teach/take a visualization-focused course
- teach/take domain-oriented course where visualization plays a role
 - presentation
 - exploratory data analysis
- offer your domain problem as project topic
 - research or administrative data

Engaging: Possible Next Steps

Redesign En Masse: Makeover Mondays

- easy entry point, Tableau focus



<http://www.makeovermonday.co.uk/blog/>

Visual Design Process In Depth: Dear Data

- inspiring celebration of data humanism

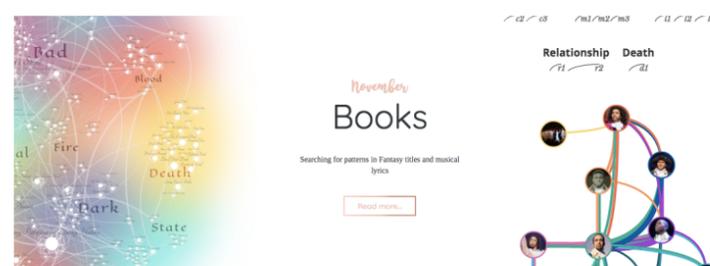


<http://www.dear-data.com/by-week/>

Giorgia Lupi and Stefanie Posavec

Visual Design Process In Depth: Data Sketches

- detailed process notes, from sketching through coding



<http://www.datasketch.es/>

Shirley Wu and Nadieh Brehmer

Pathways to participate

- join Viz@UBC
 - <https://dfp.ubc.ca/initiatives/viz-ubc>
 - get on visatubc-announce email list (send mail to vizatubc-info@cs.ubc.ca)
 - upcoming kickoff events: 2 more talks + 1 mixer
 - join as core, so you're findable in people index
 - join as organizer, help us decide what to do next
- join Data Visualization Society
 - <https://www.datavisualizationsociety.com/>
 - brand new! resources, jobs board,...

Pathways to participate

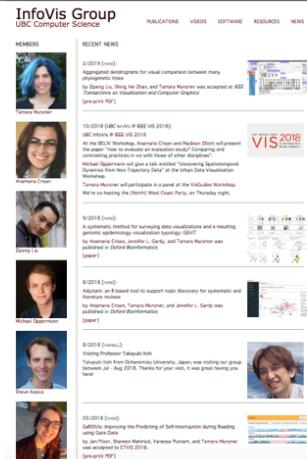
- participate in IEEE VIS 2019 in Vancouver, Oct 20-25
 - <http://ieevis.org>
 - big 3 research tracks: VA, InfoVis, SciVis
 - many associated events
 - Vis in Data Science
 - Vis In Practice
 - Large Data Analysis & Visualization
 - Application Spotlights
 - many workshops including bio, security, ...
 - job fair (asynchronous)



More Information

- this talk <https://www.cs.ubc.ca/~tmm/talks.html#vizatubc19-entry>
- papers, videos, software, talks, courses <http://www.cs.ubc.ca/group/infovis> <http://www.cs.ubc.ca/~tmm>

@tamaramunzner



Q&A References

- entry points for practitioners?
 - D3 resources for advanced programmers: <https://github.com/d3/d3/wiki/Gallery> <https://bl.ocks.org/>
 - R resources for range of programming experience: <https://www.tidyverse.org/> <https://ggplot2.tidyverse.org/>
 - Tableau resources, for non-programmers: <https://www.tableau.com/>
 - Andy Kirk's continuously updated resources list <http://www.visualisingdata.com/resources/>
 - many of these do not require programming!