

# Hands-On: Visualization

**Tamara Munzner**

Department of Computer Science

**University of British Columbia**

*IRS Cross-Border Reporting Workshop*

*May 2017, Vancouver BC*

<http://www.cs.ubc.ca/~tmm/talks.html#cb17>

[@tamaramunzner](#)

# Visualization (vis) defined & motivated

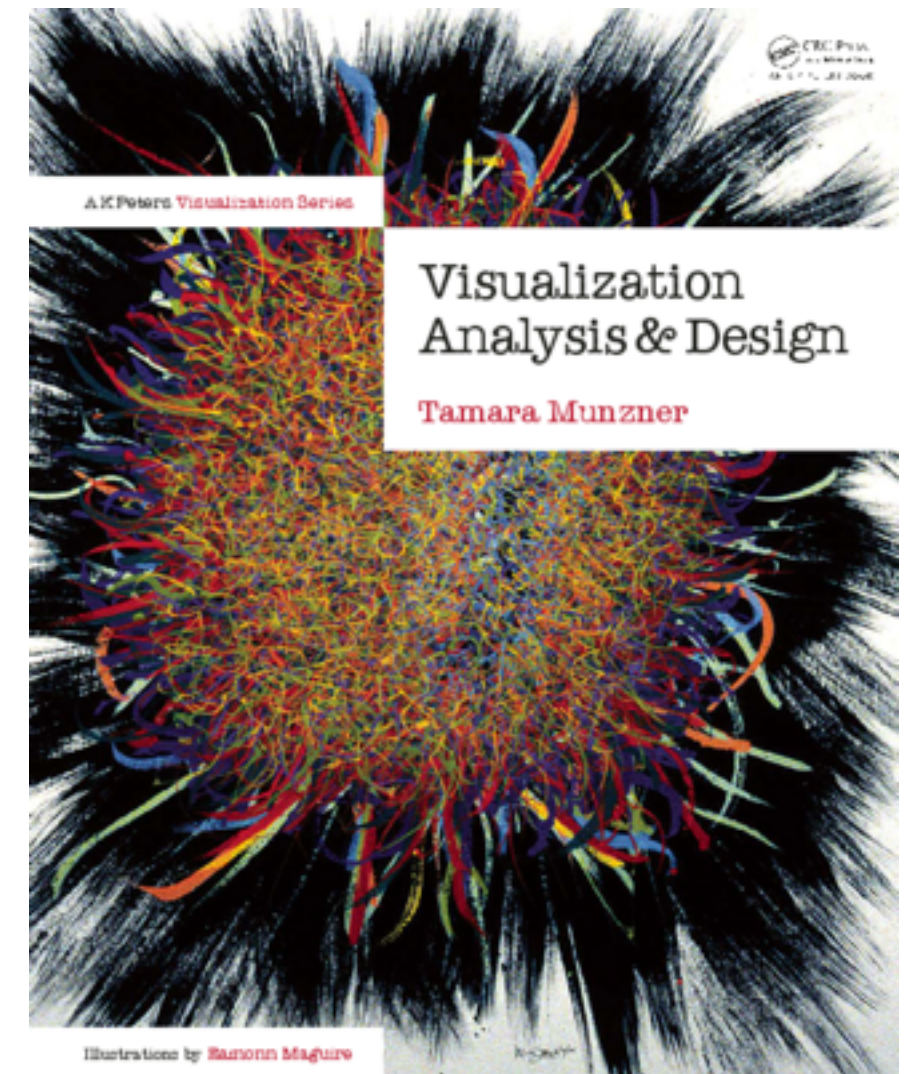
**Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.**

**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 the details
  - doesn't know exactly what questions to ask in advance
  - exploration: longterm exploratory data analysis (EDA)
  - presentation: known results
  - stepping stone towards automation: refining, trustbuilding
- 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.*



# Two vignettes

- a tale of two tools created for journalistic use
  - shared frameworks of interdisciplinary methods from my research group
    - thinking about collaboration
      - roles & rewards, for computer scientists & journalists
    - reasoning about visualization design
      - beyond pretty pictures
  - divergent goals & audiences
    - TimeLineCurator: presentation / explanatory
    - Overview: investigation / exploratory

# Collaboration incentives

- why do CS/vis people need to understand journalism's problems?
  - we work with you to understand your driving problems
  - we build tools intended to help
    - only works out if we understood the problems deeply enough
  - we observe how you use them
    - if they're good enough
      - CS win: research success stories
      - journalist win: access to better tools
  - we develop guidelines on how to build better tools in general
    - CS win: research progress in visualization

# Vignette 1: Vis Tool for Journalistic Presentation



Johanna Fulda  
@jofu\_



Matthew Brehmer  
@mattbrehmer



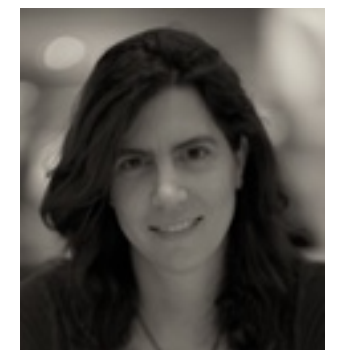
# TimeLineCurator

*Interactive Authoring of Visual Timelines from Unstructured Text*

<http://about.timelinecurator.org>

<http://timelinecurator.org>

Tamara Munzner  
@tamaramunzner



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.

# Origin story: Tedium in the newsroom

- Johanna Fulda: interactive infographics developer, Sueddeutsche Zeitung
  - then Munich CS master's student, visiting UBC
- what pain point could we address with interactive visualization?
  - plus some natural language processing (NLP)

# TimeLineCurator

visual & browser-based

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



# Manual creation process



**Mighty Mouse**  
In 1980, Apple Computer asked a group of guys fresh from Stanford's product design program to take a \$400 device and make it mass-producible, reliable and cheap.  
Their work transformed personal computing.  
By Alex Seojung Kim-Pang  
Dean Hovey was hungry. His young industrial design firm, Hovey-Kelly Design, had been working on projects for Apple Computer for a couple of years but wanted to develop entire products, not just casings and keyboards. Hovey had come to pitch Apple co-founder Steven Jobs some ideas. But before he could get started, the legendary high-tech pioneer interrupted him. "Stop, Steve," Hovey recalls Jobs saying. "What you guys need to do, what we need to do together, is build a mouse."  
Hovey was dumbfounded. A what? Jobs told him about an amazing computer, code-named Alto, he had just seen at Xerox's Palo Alto Research Center (PARC). In early 1980, most computers (including Apple's) required users to memorize text commands to perform tasks. The Alto had a graphical user interface—a symbolic world with little pictures of folders, documents and other icons—that users navigated with a handheld input device called a mouse. Jobs explained that Apple was working on two computers, named Lisa and Macintosh, that would bring that technology to market. The mouse would help revolutionize computers, making them more accessible to ordinary people. "When I walked out that door," recalls Hovey, "I, 35 '85, "I was ready to change the world."  
Just one problem: a commercial mouse based on the Xerox technology cost \$400, malfunctioned regularly and was nearly impossible to clean. That device—a descendant of the original computer mouse invented by Douglas Engelbart at the Stanford Research Institute in the early 1960s—was a masterpiece of high-concept technology, but a hopeless product. Jobs wanted a mouse that could be mass-produced for \$12 to \$25, survive everyday use and work on his terms. "We

**The History of user interfaces**  
Independently, Douglas Engelbart at the Stanford Research Institute (SRI) and Bill English at the Xerox International (Xerox) invented their first mouse prototypes. They christened the device attached to the rear part of the device looking like a mouse. Engelbart never received the patent, which ran out before it became a common mouse. Engelbart's invention of the mouse was just a small part of augmenting human intellect via the Augmented Reality (AR) system. The mouse was one of the first pointing devices. The mouse was used in the Xerox's off-Line System (OLS) for document editing. The mouse was used in the Xerox's group had been using the image of that mouse at the time. The mouse was used in the Xerox's group had been using the image of that mouse at the time. The mouse was used in the Xerox's group had been using the image of that mouse at the time.

**1868 The Typewriter**  
Invented by Christopher Sholes, typewriters quickly became indispensable tools for practically all writing other than personal correspondence. They were widely used by professional writers, in offices, and for business correspondence in private homes.

**188 Douglas C. Engelbart Mouse + One Button Computer**

**1897 The Mouse**  
Some additional information here

**1977 The Stylus**  
a small pen-shaped instrument that is used to input commands to a computer screen, mobile device or graphics tablet.

**2007 Multi Touch**  
With the start of iPhones Multi-touch became a thing

**2012 Speech Recognition**

*we only have 2 columns for this piece*  
*Why is there such a big gap that mostly disappears and we only have 2 columns*  
*can we not have as much whitespace here?*  
*was there anything else happening in the time between Mouse and Stylus?*  
*has to be mentioned here, since the new bit 3D Touch*

# Structured creation process

















Start Date	End Date	Headline	Text	Media	Media Credit
10/28/2003		Facemash	Zuckerberg wrote a program called Facemash on October 28, 2003 while attending Harvard as a sophomore	<a href="http://dubindigital.in/wp-content/uploads/2011/03/facemash-700x429.jpg">http://dubindigital.in/wp-content/uploads/2011/03/facemash-700x429.jpg</a>	Dublin Digital
1/1/2004		new website	The following semester, Zuckerberg began writing code for a new website in January 2004		
2/4/2004		thefacebook.com	On February 4, 2004, Zuckerberg launched "thefacebook", originally located at thefacebook.com		
2/10/2004		Harvard Connection accusations	Six days after the site launched, three Harvard seniors (Cameron Winklevoss, Tyler Winklevoss, and Divya Narendra) accused Zuckerberg of intentionally misleading them into believing he would help them build a social network called HarvardConnection.com	<a href="http://www.capiberg.com/wp-content/uploads/2011/05/3Harvard-wider.jpg">http://www.capiberg.com/wp-content/uploads/2011/05/3Harvard-wider.jpg</a>	Capital Berg
1/1/2008	12/31/2008	Connection	They later filed a lawsuit against Zuckerberg, subsequently settling in 2008[17] for 1.2 million shares (worth \$300 million at Facebook's IPO)	<a href="http://4.bp.blogspot.com/_KMKs4_C0jys/YY5336_YFsi/AAAAAAAAAT8/btIse-cG01A/s16005.jpg">http://4.bp.blogspot.com/_KMKs4_C0jys/YY5336_YFsi/AAAAAAAAAT8/btIse-cG01A/s16005.jpg</a>	Estymene
2/4/2004	3/4/2004	thefacebook at Harvard only	Membership was initially restricted to students of Harvard College; within the first month, more than half the undergraduates at Harvard were registered on the service		
3/1/2004	3/31/2004	thefacebook expands to other universities	In March 2004, Facebook expanded to the universities of Columbia, Stanford, and Yale [20] In mid-2004, entrepreneur Sean Parker (an informal advisor to		

**TimelineJS**  
[timeline.knightlab.com/](http://timeline.knightlab.com/)

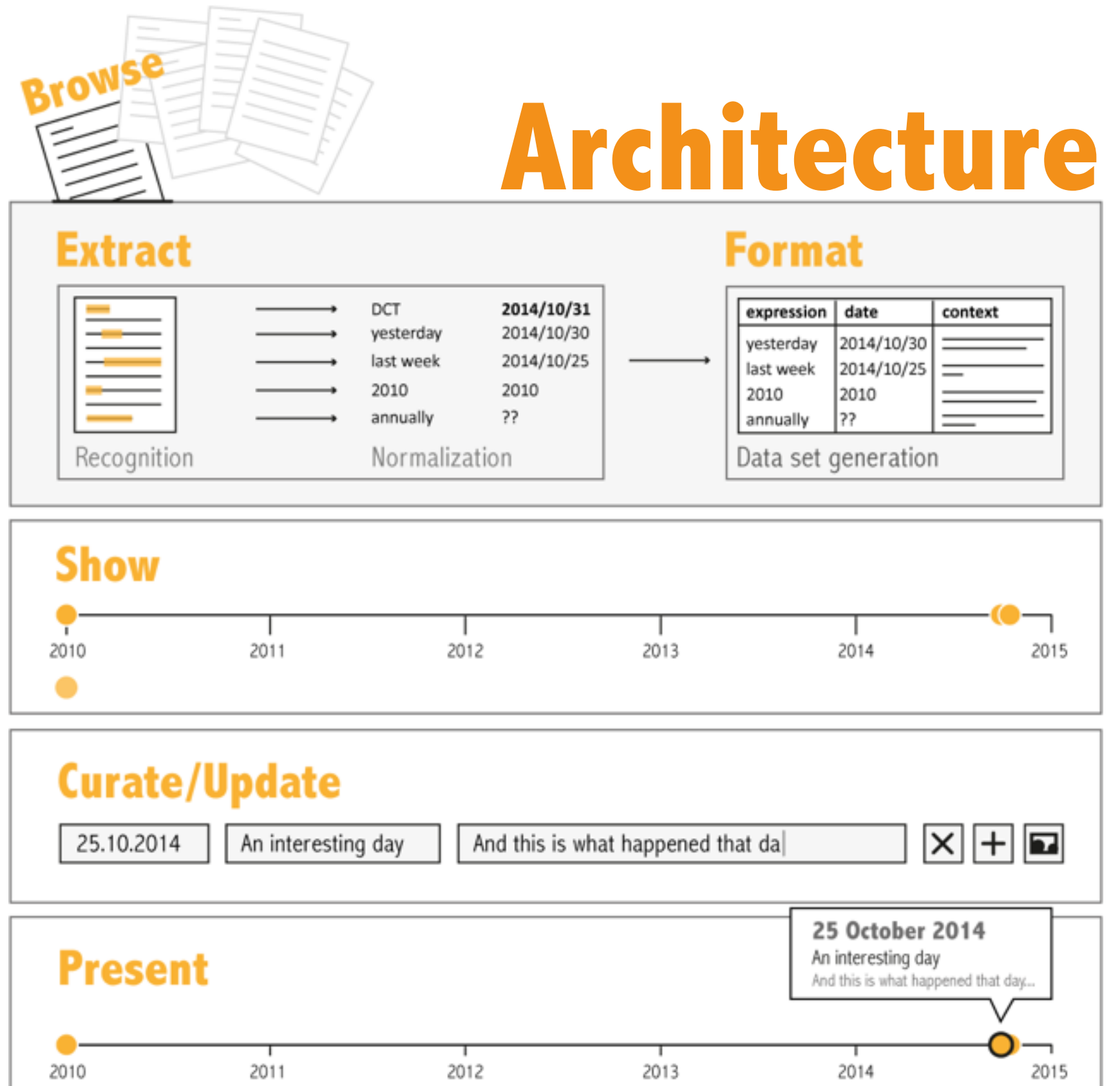
# Timeline authoring model

- time required for each task

	<b>Browse</b>	<b>Extract</b>	<b>Format</b>	<b>Show</b>	<b>Update</b>
Manual Drawing	 slow	 slow		 slow	 slow
Structured Creation	 slow	 slow	 slow	 automated	 fast
<b>TimeLine Curator</b>	 fast	 automated	 automated	 fast	 fast

# 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
  - enable what was impossible before
  - vis tools for new insights & discoveries
- workhorse use case: workflow speedup
  - vis tools to accelerate what you're already doing
    - 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

s p e c u l a t i v e   b r o w s i n g

Try it out!

<http://timelinecurator.org>

# Vignette 2: Vis Tool for Investigative Reporting





Matthew Brehmer  
@mattbrehmer



Stephen Ingram  
@FroweFace



# Overview

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

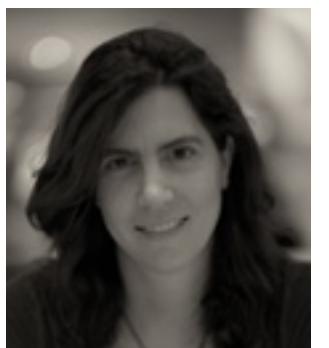
<http://www.cs.ubc.ca/labs/imager/tr/2014/Overview/>

<https://www.overviewdocs.com>

Jonathan Stray  
@jonathanstray

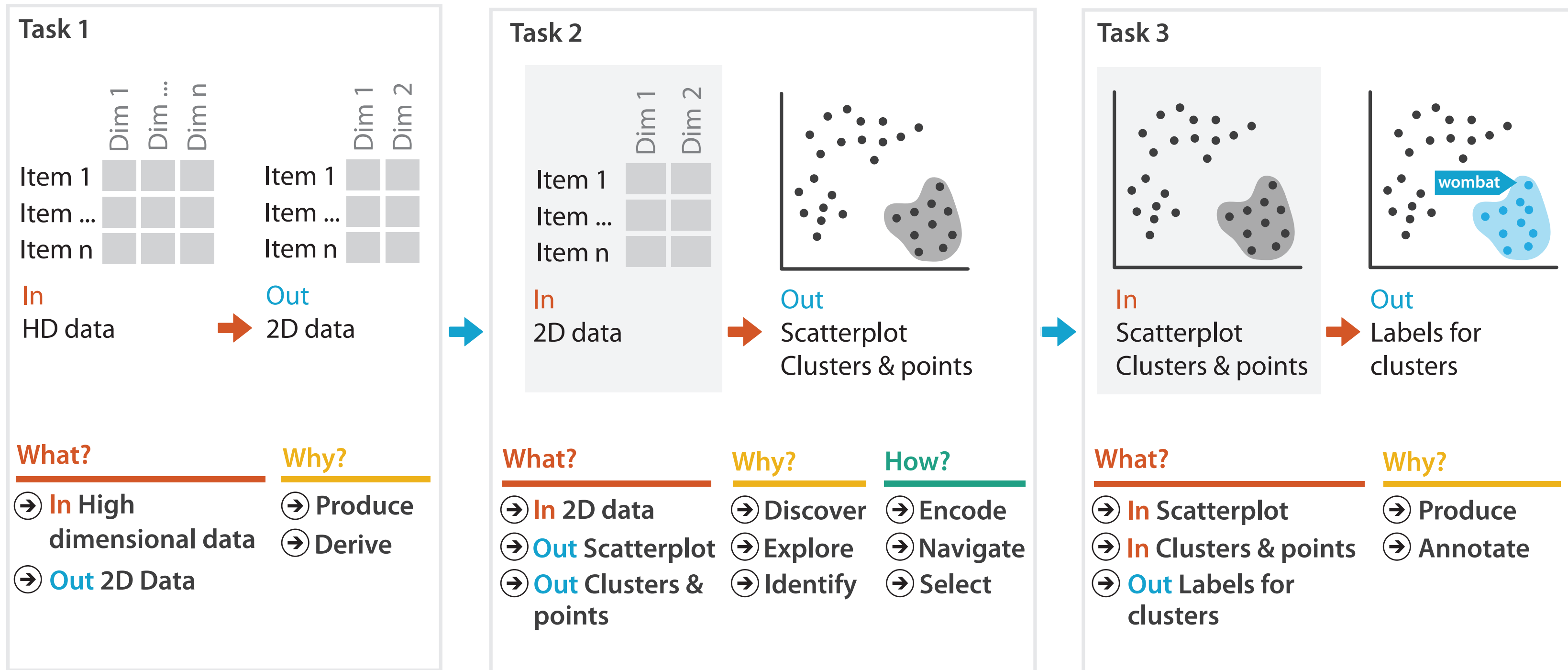


Tamara Munzner  
@tamaramunzner



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.

# Starting point: Dimensionality reduction for document datasets



- more on DR: hour-long talk *Dimensionality Reduction from Several Angles*

<http://www.cs.ubc.ca/~tmm/talks.html#kelowna16>

# Origin story: WikiLeaks meets Glimmer

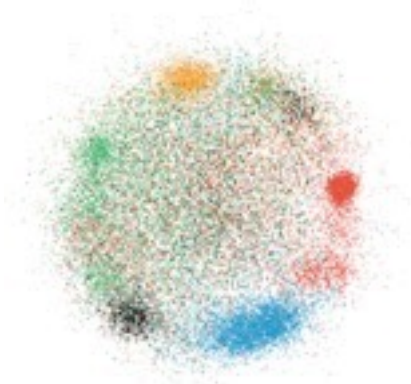
- WikiLeaks: hacker-journalist Jonathan Stray analyzing Iraq warlogs
  - one instance of general problem: Too Many Documents
  - conjectured that existing label classification falls short of showing all meaningful structure in data
    - friendly action, criminal incident, ...
  - he had some NLP, needed better vis tools



- Glimmer: multilevel dimensionality reduction algorithm
  - scalability to 30K documents and terms

*[Glimmer: Multilevel MDS on the GPU.*

*Ingram, Munzner, Olano. IEEE TVCG 15(2):249-261, 2009.]*



# Overview: Early version

The screenshot displays the MoDiscoTag software interface, which is divided into several functional panels:

- Disconnected Component Tree:** A tree diagram showing hierarchical relationships between nodes. A vertical axis on the left is labeled "Distance Threshold" with a value of 1.0. Below the tree, a "Show Nodes >=" control has buttons for 1, 2, 4, 8, 16, 32, and 64. The number 64 is highlighted in red. Below these buttons, the text "107: surinam dutch bouters guyana paramaribo jagdeo khan" is visible.
- Tags View:** A panel for managing tags. It includes a "NEW" button and a table of "Selected Nodes":

Selected Nodes	+	-	X
Oil industry	+	-	X
Finance	+	-	X
extradition	+	-	X
Banks	+	-	X
Airlines	+	-	X

A "SAVE" button is located below the table.
- Items Plot:** A scatter plot showing a dense cluster of points. Below the plot are two sliders: "Squeeze" and "Point Size", each with a "Run..." button to its left.
- Active Set List:** A list of active sets, each with a folder icon and a list of terms. A "CLEAR" button is positioned to the right of the list.
  - 16: uav turkei shipment blue\_lantern iran turkey\` ankara turk unscr item colr
  - 10: uav turkei shipment iran turkey\` ankara unscr got related\_materiel mater
  - 6: shipment uav turkei iran unscr ankara materiel related\_materiel export\_cor
  - 4: blue\_lantern ecuadorian\_navy pm/dtcc sipdis\_pm/dtcc komc coordinator\_e
  - 4: uav shipment turkei iran unscr materiel related\_materiel turkish\_officials gc

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

# Overview: current version

# OVERVIEW

[Blog](#) [Help](#) [Contact us](#) admin@overview-project.org Admin ▾ Your document sets Log out

+

-

service\_offshore

investigation, independence, inc, atlantis\_platform, production, whether, let

ALL: investigation, atlantis\_pl

MOST: hub, facility,

ALL: leases, royalty

ALL: investiga

ALL: c

ALL

ALL: indej

ALL

Tags atlantis form letter rig visit

[organize tags...](#)

[Back to list](#)

Document 1 of 11

Previous [in folder MOST: hub, facility, canyon SOME: independence...](#) Next


MMS2 Pdf 68 168 168

rig visit

Key words: arrive, port\_fourchon, aviation\_service, depart, shell [← Show sidebar](#)

DOCUMENT PAGES TEXT Zoom

p. 1



## Offshore Schedule

Friday, September 25, 2009

11:30 a.m.	Arrive Atlantic Aviation Service, 749 Lockheed Dr, Kenner, LA Gen. Mgr. Keith Myer, (Phone 504-453-8207); Receive helicopter safety briefing, personal protective equipment
12:00 noon	Depart Atlantic Aviation Service
1:15 p.m.	Arrive at Shell Offshore Inc.'s (Shell) Mars or Ursa Tension Leg Platform (TLP) located at Mississippi Canyon Block 807 or 809. Personnel introductions and personnel receive safety briefing/presentation by Shell.
2:00 p.m.	Tour either the Mars or Ursa TLP
4:00 p.m.	Depart TLP and flyover Louisiana Offshore Oil Port (LOOP) located at Grand Isle Block 59 enroute to Port Fourchon
5:00 p.m.	Arrive at Port Fourchon, Louisiana

# Deploy in the real world

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	<i>What did security contractors do during Iraq war?</i>	<i>Were municipal police funds mismanaged?</i>	<i>Were Paul Ryan's campaign statements hypocritical?</i>	<i>What is the gun ownership debate about?</i>	<i>Was gov't response to emergency incident effective?</i>	<i>Did gov't fail to pass bills addressing police misconduct?</i>

# Deploy in the real world, understand user goals

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	<i>What did security contractors do during Iraq war?</i>	<i>Were municipal police funds mismanaged?</i>	<i>Were Paul Ryan's campaign statements hypocritical?</i>	<i>What is the gun ownership debate about?</i>	<i>Was gov't response to emergency incident effective?</i>	<i>Did gov't fail to pass bills addressing police misconduct?</i>
	the trend story: getting the big picture	find the needle in the haystack / smoking gun	wheat vs chaff: filtering out irrelevant material	categorize and count: turning docs into data	wheat vs chaff: filtering out irrelevant material	prove haystack contains no needles!

<https://blog.overviewdocs.com/2014/01/09/what-do-journalists-do-with-documents-the-different-kinds-of-document-driven-stories/>

# Evolution across levels

- evolution of task abstraction

- task 1: **generate hypotheses** → **explore** → **summarize**

- *obviously you can't read everything; speed up with tool for categorizing and counting*

- task 2: **verify hypotheses** → **locate** → **identify**

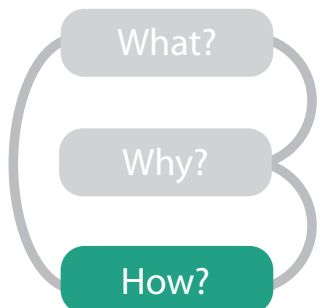
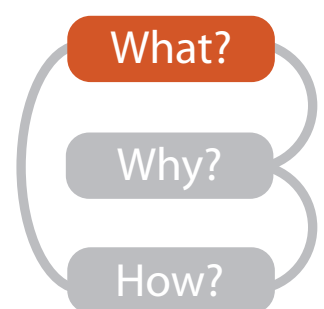
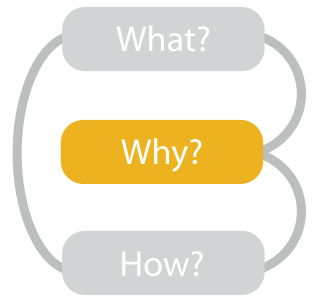
- *you really do read each doc; speed up with tool to keep track of findings*

- evolution of data abstraction & idioms

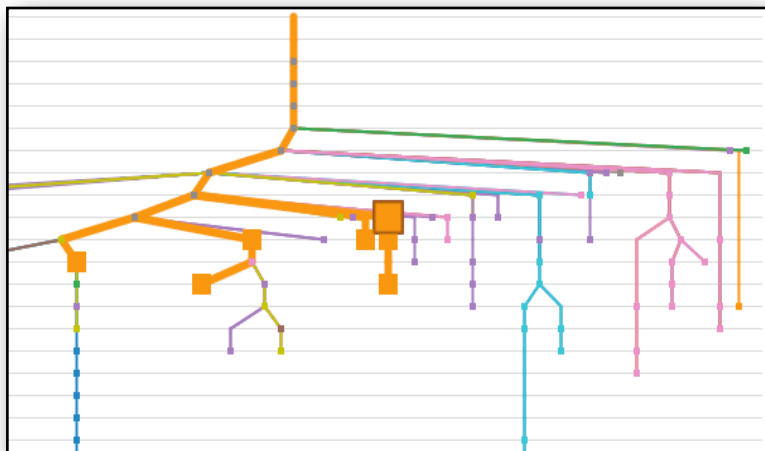
- arrange cluster tree to emphasize nodes vs links

- new vis insight: DR scatterplot less effective than cluster tree vis + tagging

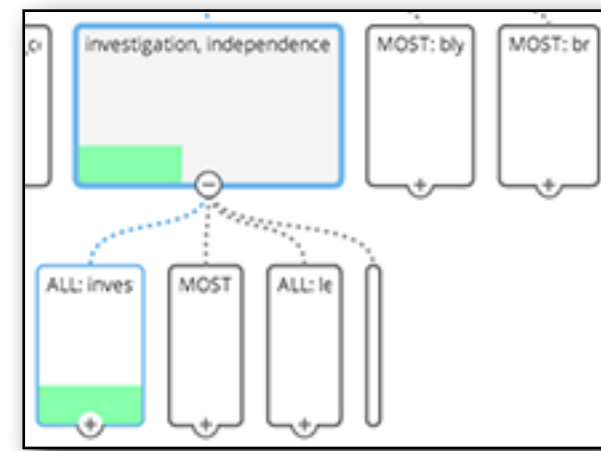
- better affordance for systematic traversal of document cluster hierarchy



early



current





Try it out!

<https://www.overviewdocs.com/>

<https://blog.overviewdocs.com/completed-stories/>

# Discussion

- how many of you have jumped into data journalism?
  - what kind of tools have you used?
- have any of you grappled with timelines?
  - what kind of tools did you use?
- have any of you grappled with large document collections?
  - what kind of story did you have in mind?

# More Information

[@tamaramunzner](https://twitter.com/tamaramunzner)

- this talk

[www.cs.ubc.ca/~tmm/talks.html#cb17](http://www.cs.ubc.ca/~tmm/talks.html#cb17)

- book

<http://www.cs.ubc.ca/~tmm/vadbook>

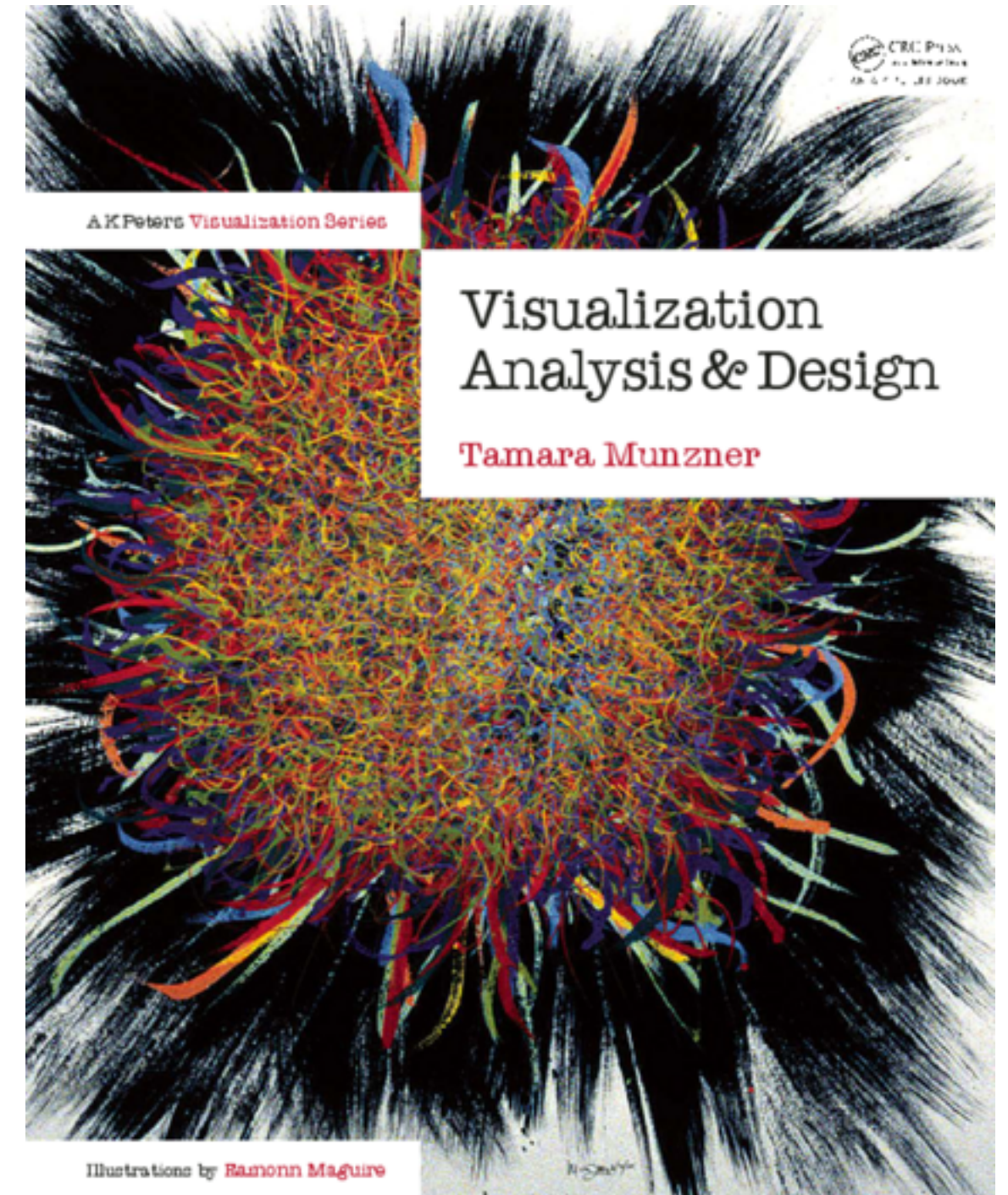
– 20% off promo code, book+ebook combo: HVN17

– <http://www.crcpress.com/product/isbn/9781466508910>

- papers, videos, software, talks, courses

<http://www.cs.ubc.ca/group/infovis>

<http://www.cs.ubc.ca/~tmm>



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