SkyTree Visualization Fireside Chat
Is Big Data Visualization Possible?

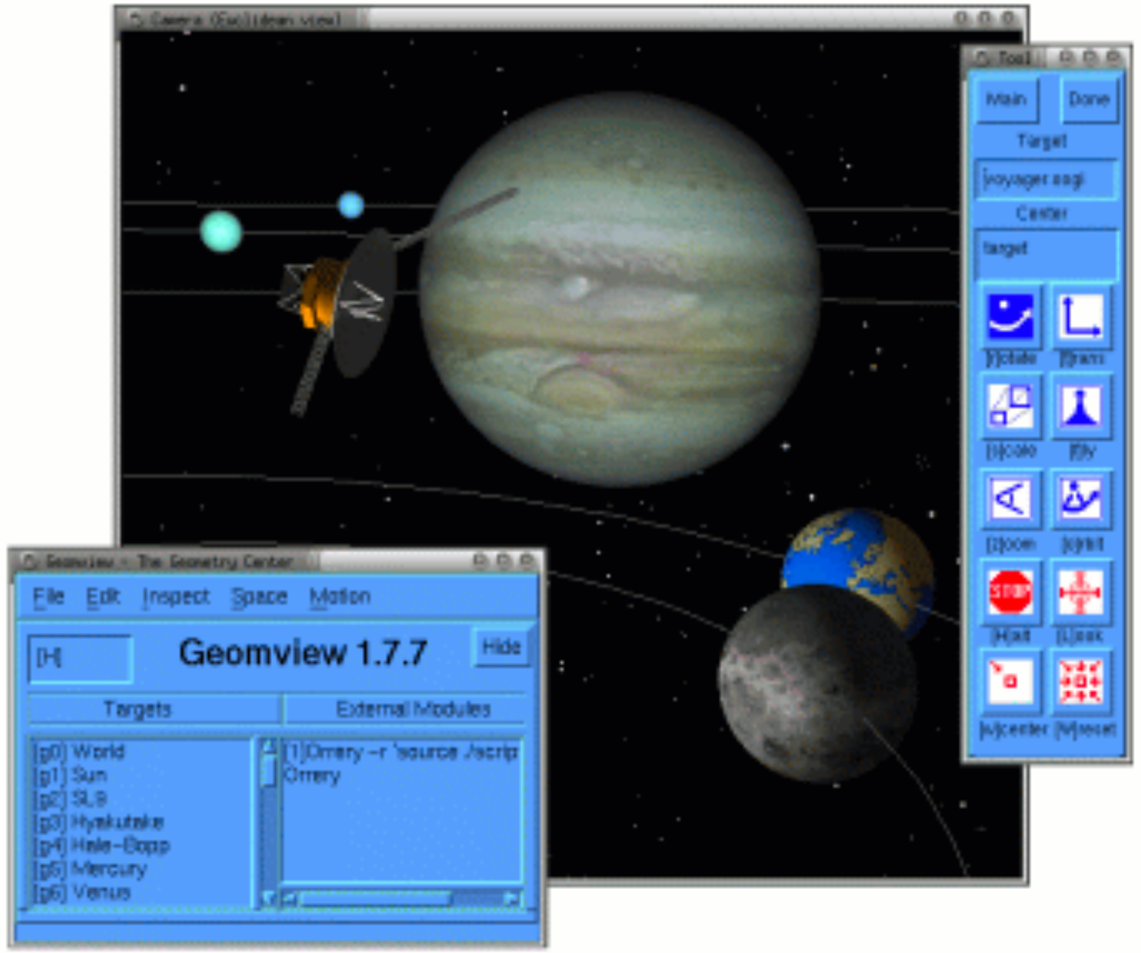
Tamara Munzner
Department of Computer Science
University of British Columbia

Google Hangout on Air
October 1 2014

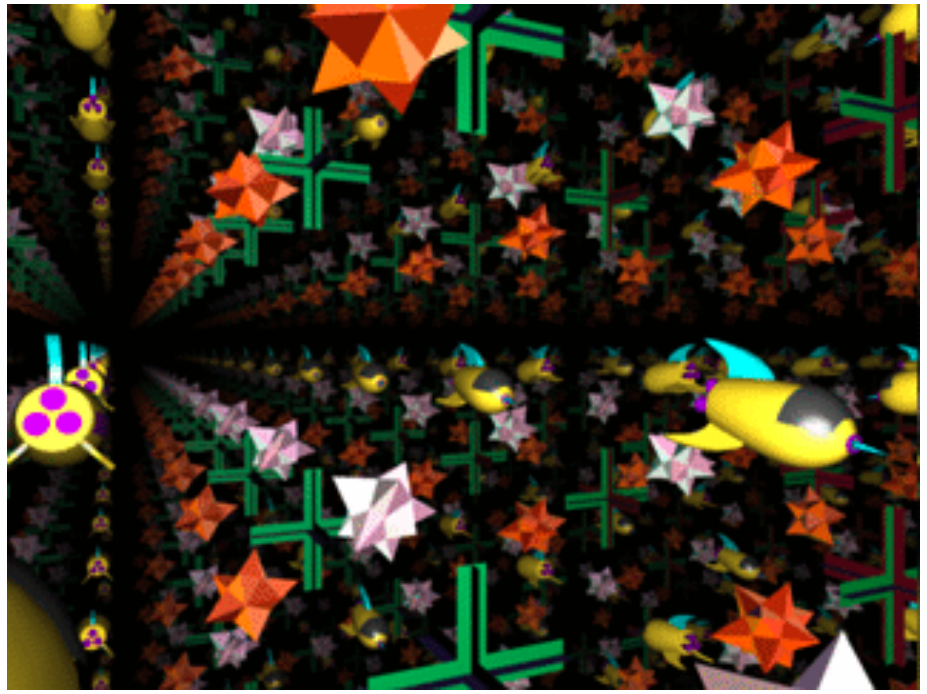
http://www.cs.ubc.ca/~tmm/talks.html#skytree14

About me: Geometry Center 1991-1995

- geometry and topology vis
  - 3D, 4D, non-Euclidean

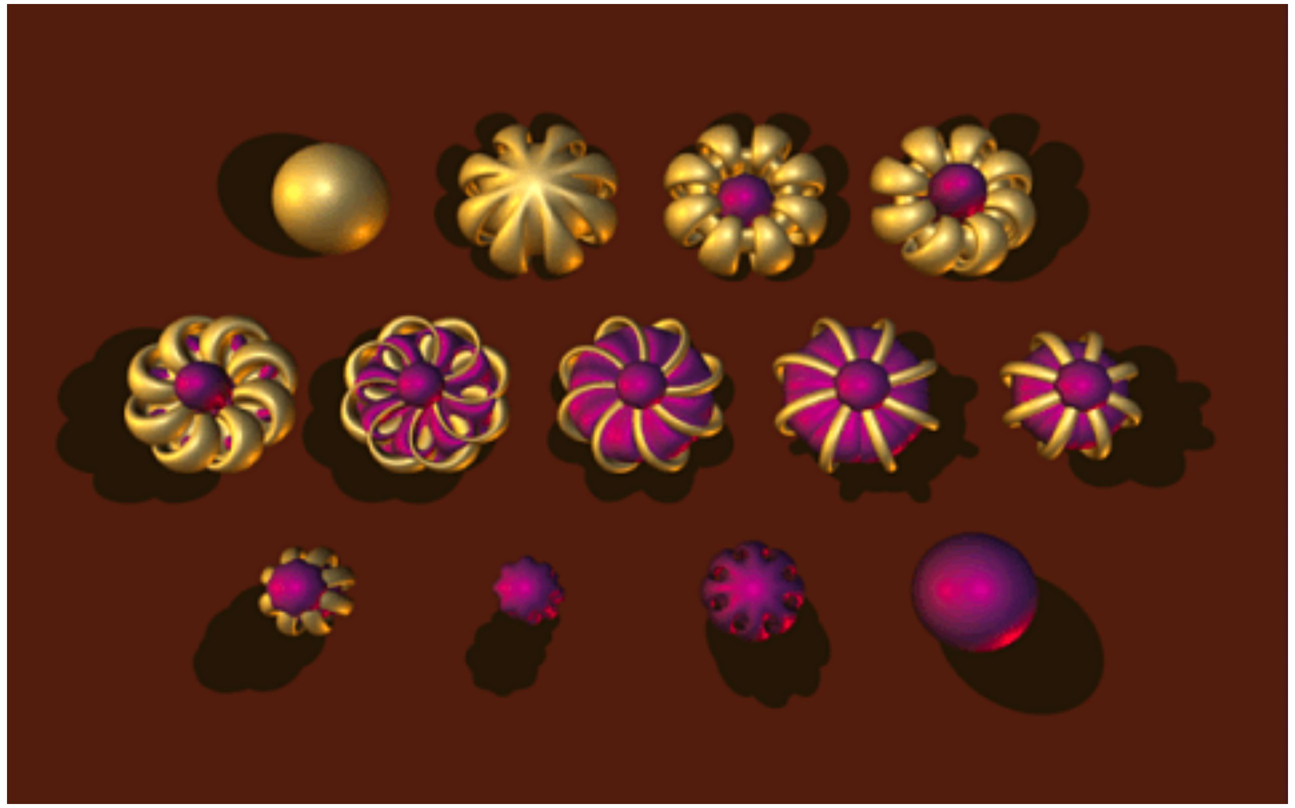


Geomview http://geomview.org/



The Shape of Space

http://youtu.be/-gLNIC_hQ3M



Outside In

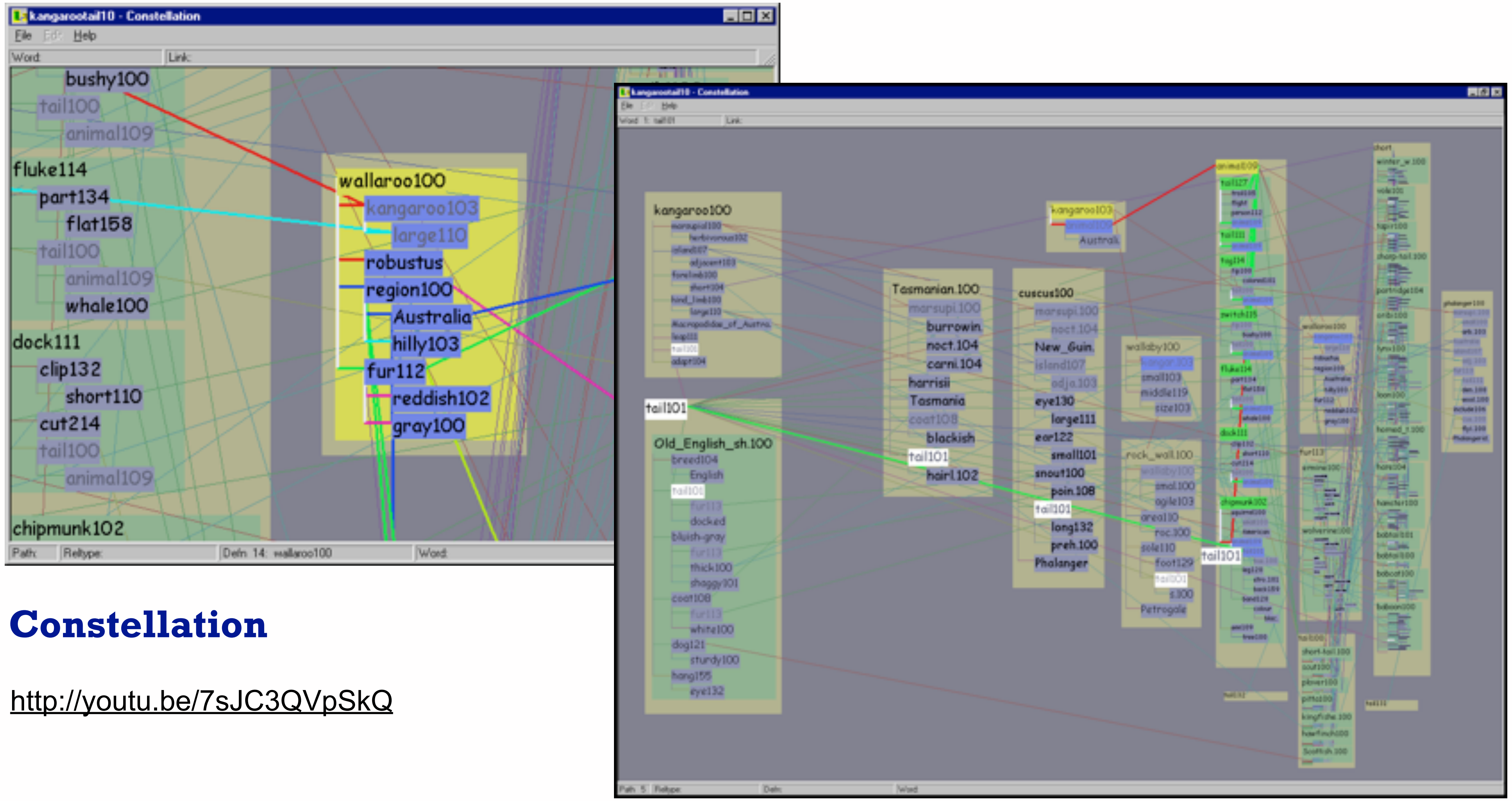
http://youtu.be/sKqt6e7EcCs

http://youtu.be/x7d13SgqUXg

http://youtu.be/6j4T7I49H3Y http://www.crcpress.com/product/isbn/9781568814537

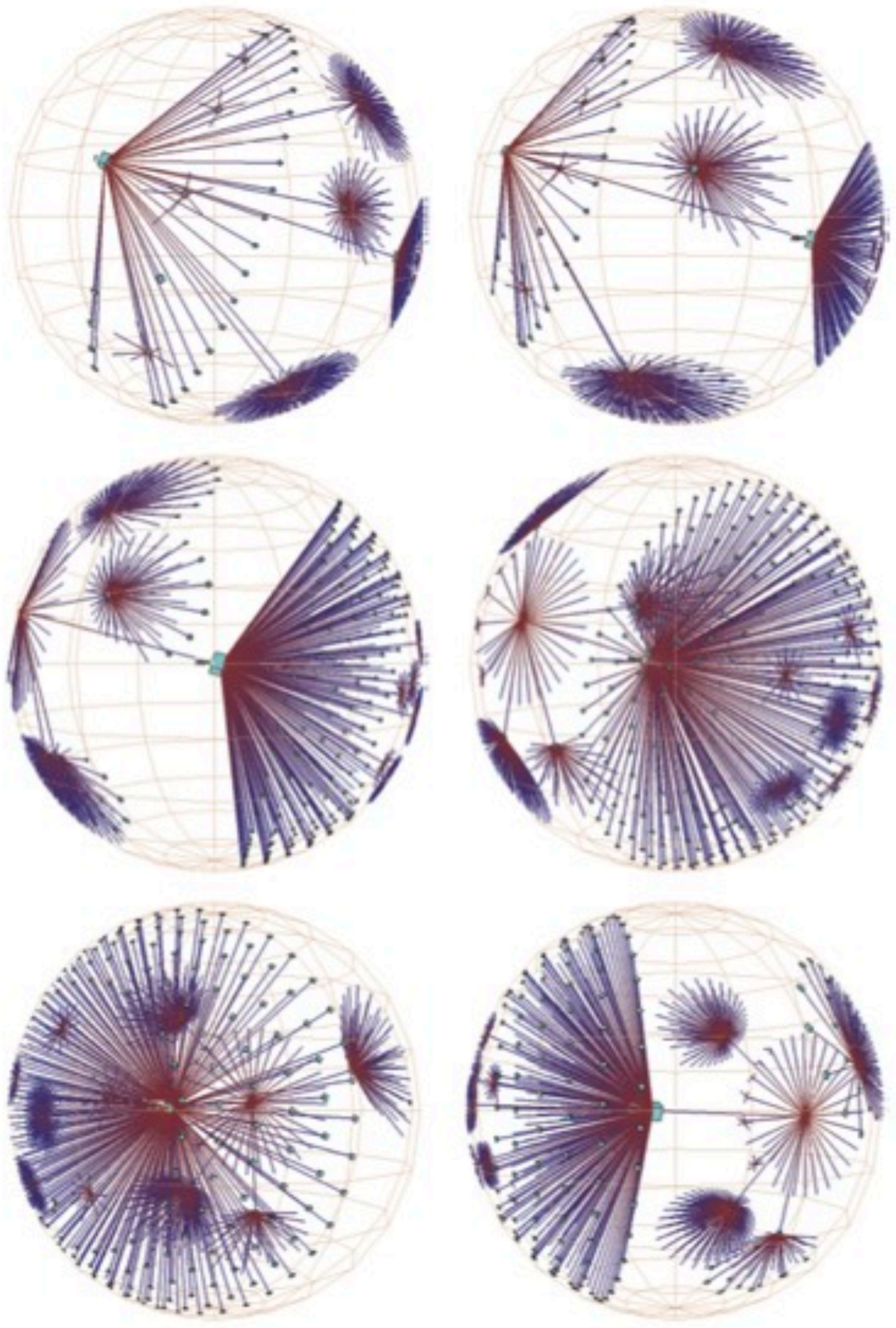
About me: Stanford 1995-2000

- infovis: network vis
  - 3D hyperbolic trees/networks
  - computational linguistics network



Constellation

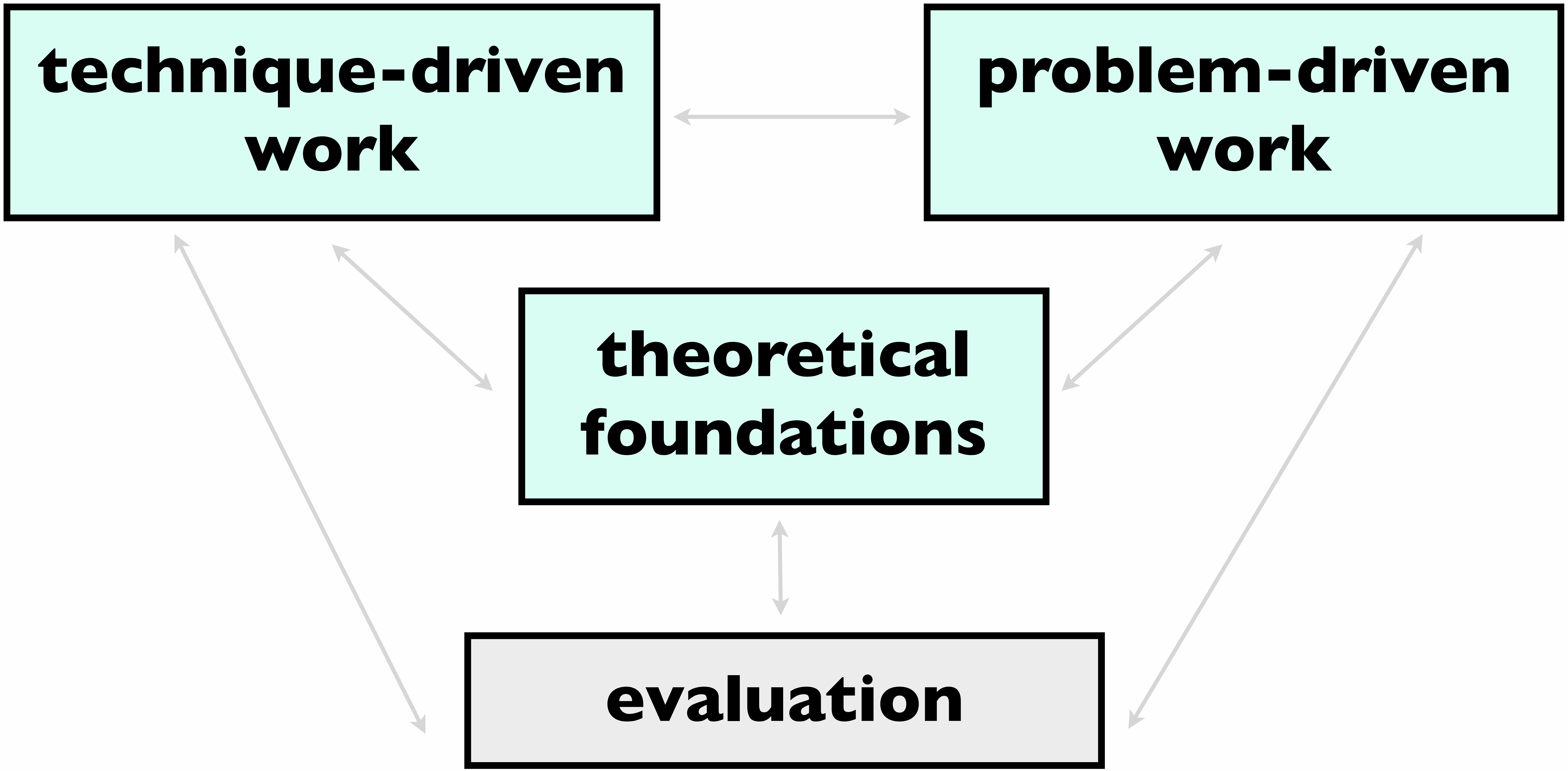
http://youtu.be/7sJC3QVpSkQ



H3

http://youtu.be/fhbQy_NCwWI

About me: UBC 2002-



technique-driven work

problem-driven work

theoretical foundations

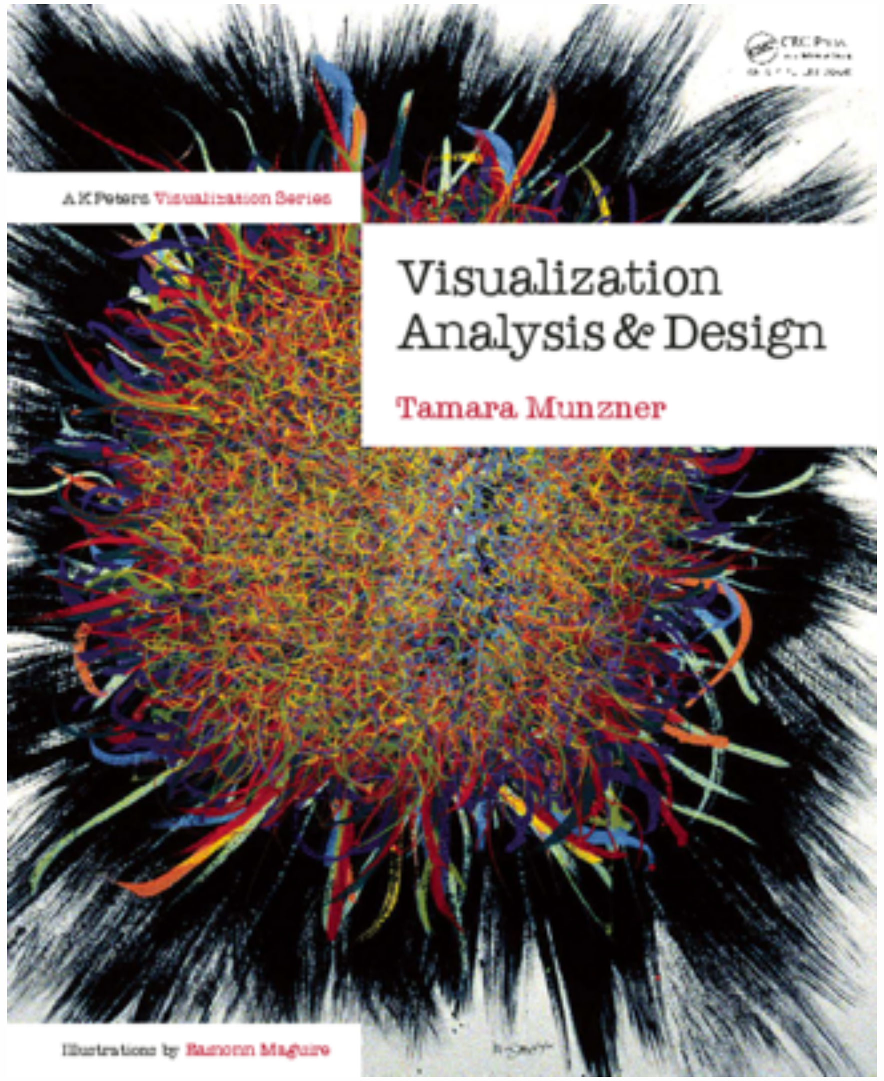
evaluation

When to use visualization

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
  - longterm analysis
  - automation stepping stone, refining, trustbuilding
  - presentation
- external representation: perception vs cognition
- intended task, measurable definitions of effectiveness


A K Peters Visualization Series
Visualization Analysis & Design
Tamara Munzner
Illustrations by Eamonn Maguire

more at:
Visualization Analysis and Design, Chapter 1.
Munzner. AK Peters, 2014, to appear.

Why show data to people?

- summaries lose information
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

Why show data to people?

- summaries lose information
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

Anscombe’s Quartet

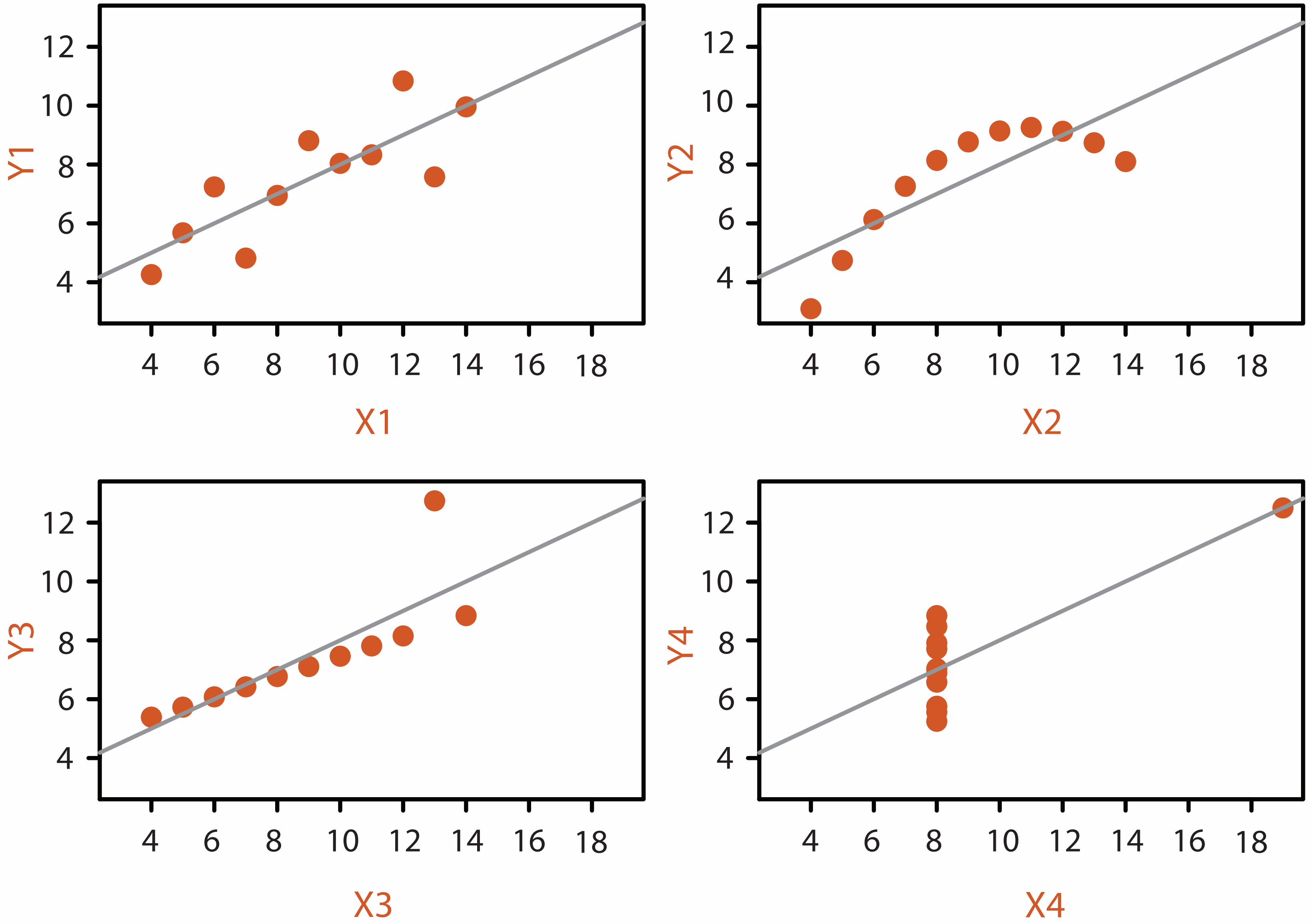
Identical statistics	
x mean	9
x variance	10
y mean	8
y variance	4
x/y correlation	1

Why show data to people?

- summaries lose information
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

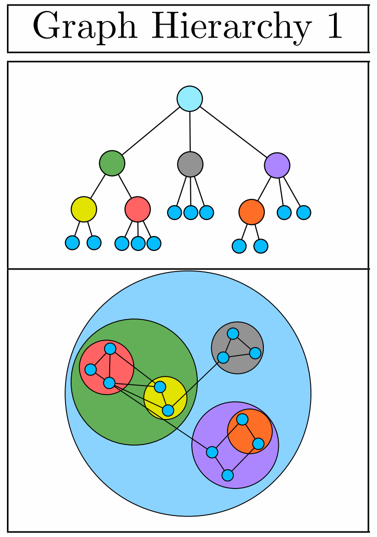
Anscombe's Quartet

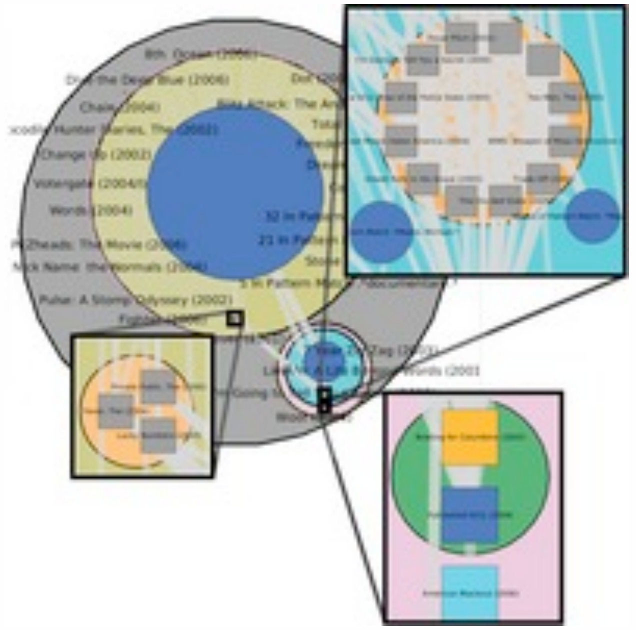
Identical statistics	
x mean	9
x variance	10
y mean	8
y variance	4
x/y correlation	1



Technique-driven work: Networks

- scaling up networks
  - multilevel networks, 10K-100K nodes
    - topologically aware decomposition, layout, browsing
  - trees, millions of nodes
    - guaranteed visibility of semantically meaningful marks

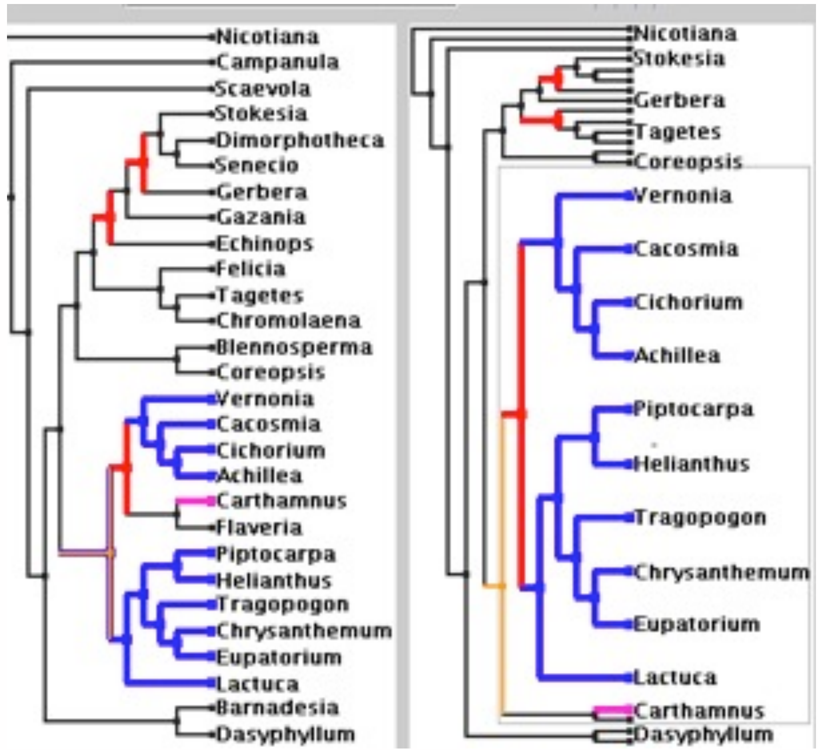


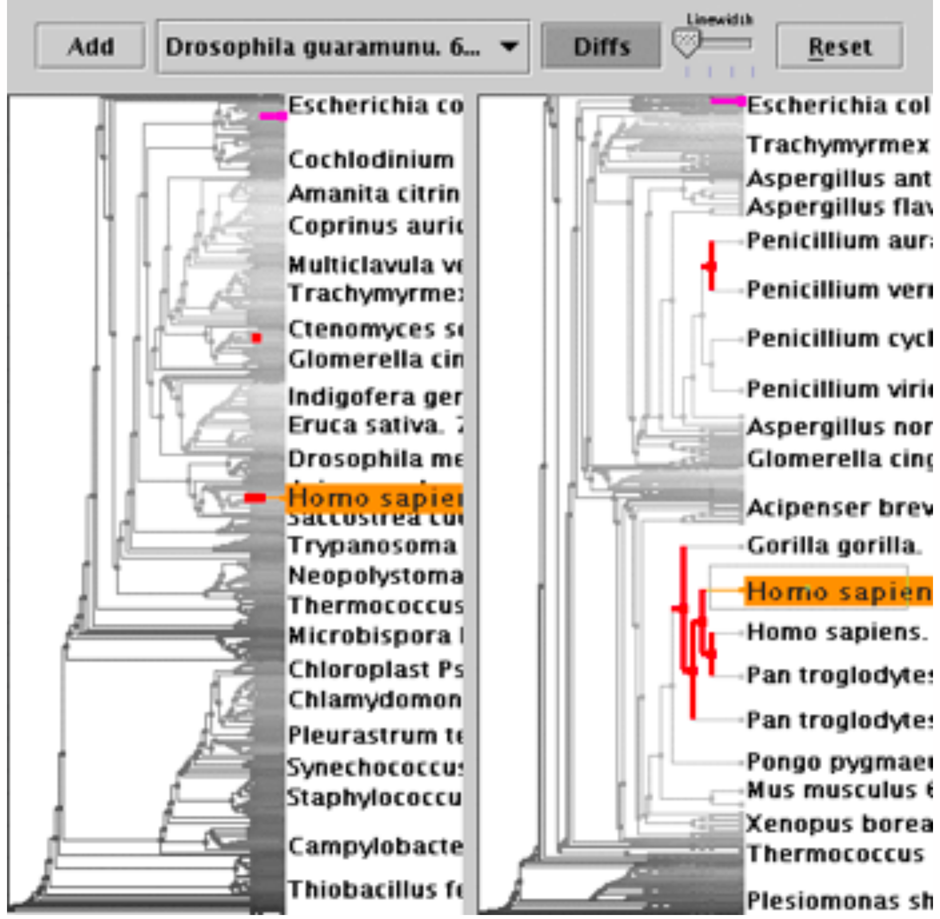


TopoLayout
Smashing Peacocks Further
Grouse
GrouseFlocks
TugGraph

http://youtu.be/t1Xbt6XOWp8

http://youtu.be/AWXAe8zvkt8





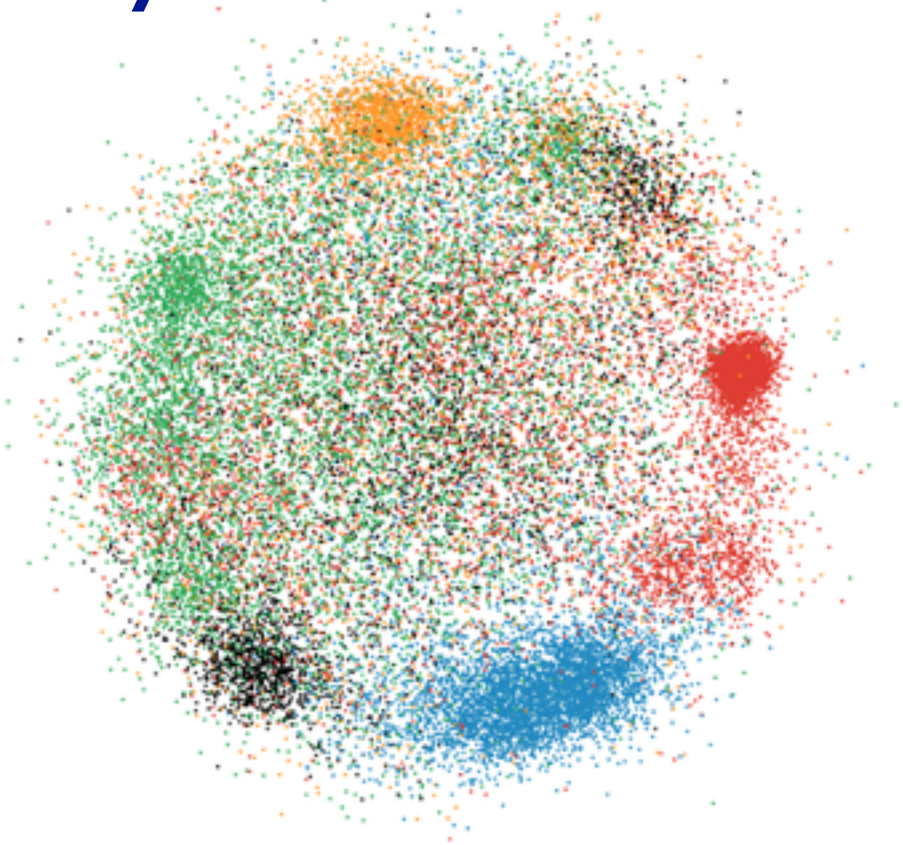
TreeJuxtaposer
PRISAD

http://youtu.be/fq8EIAOutvs

http://youtu.be/GdaPj8a9QEo

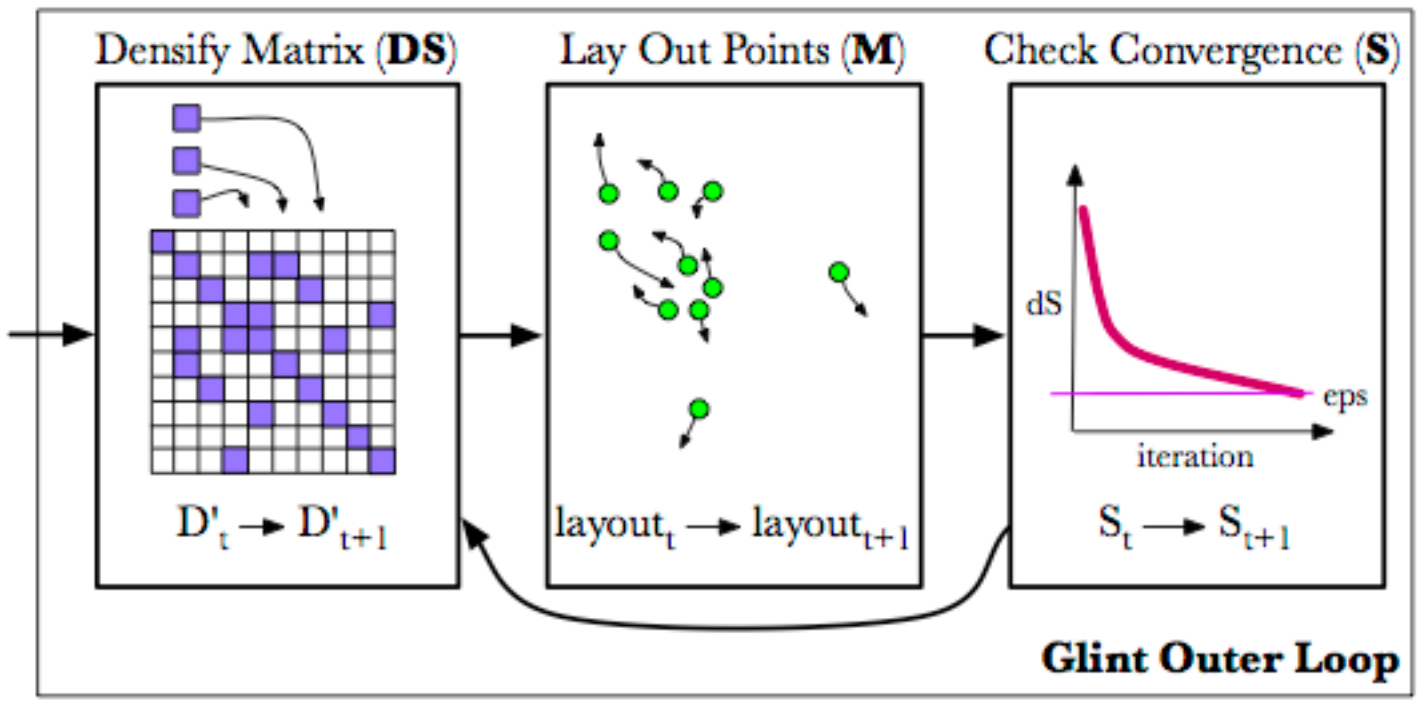
Technique-driven work: Dimensionality reduction

- closest overlap between vis and ML
  - Glimmer: MDS on the GPU
  - Glint: DR for costly distances
  - QSNE: sparse documents
    - high quality for millions of items



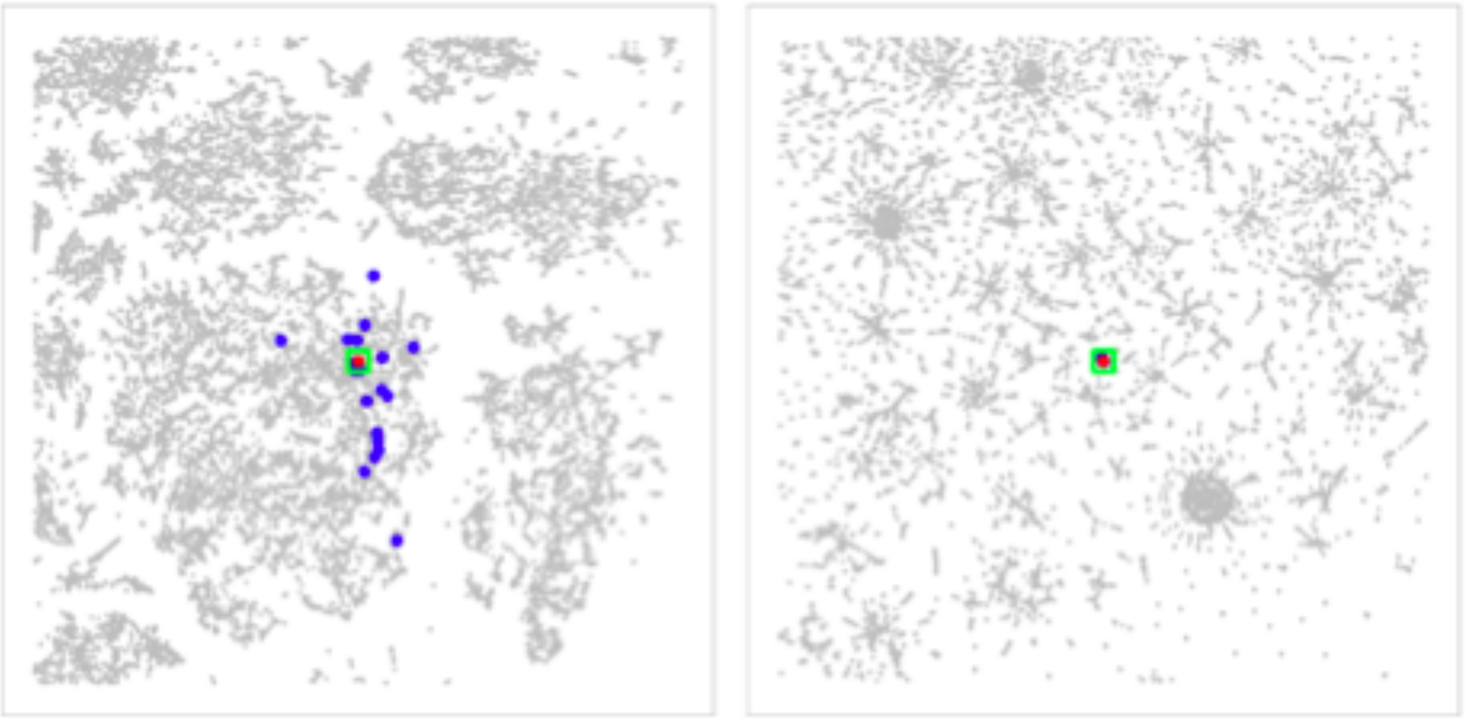
Glimmer

http://youtu.be/PLaBAPM6qLI



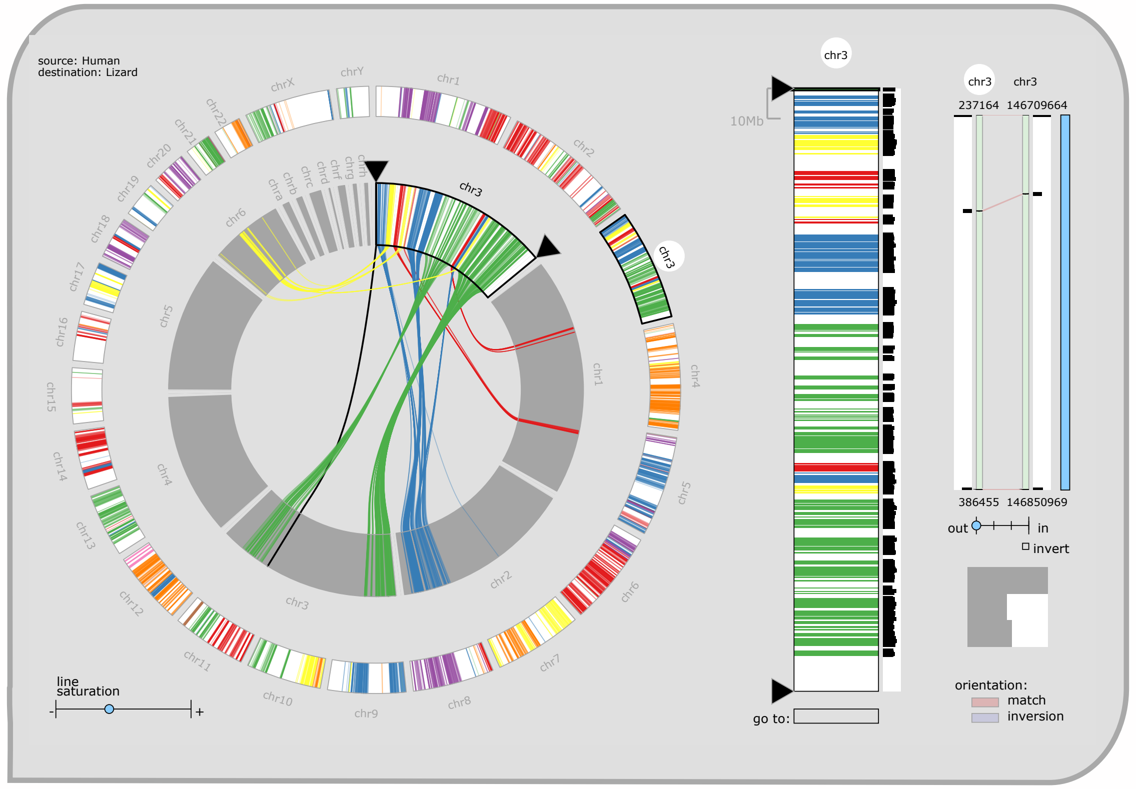
Densify Matrix (**DS**)
Lay Out Points (**M**)
Check Convergence (**S**)
$D'_t \rightarrow D'_{t+1}$
$\text{layout}_t \rightarrow \text{layout}_{t+1}$
dS
eps
iteration
$S_t \rightarrow S_{t+1}$
Glint Outer Loop

Glint

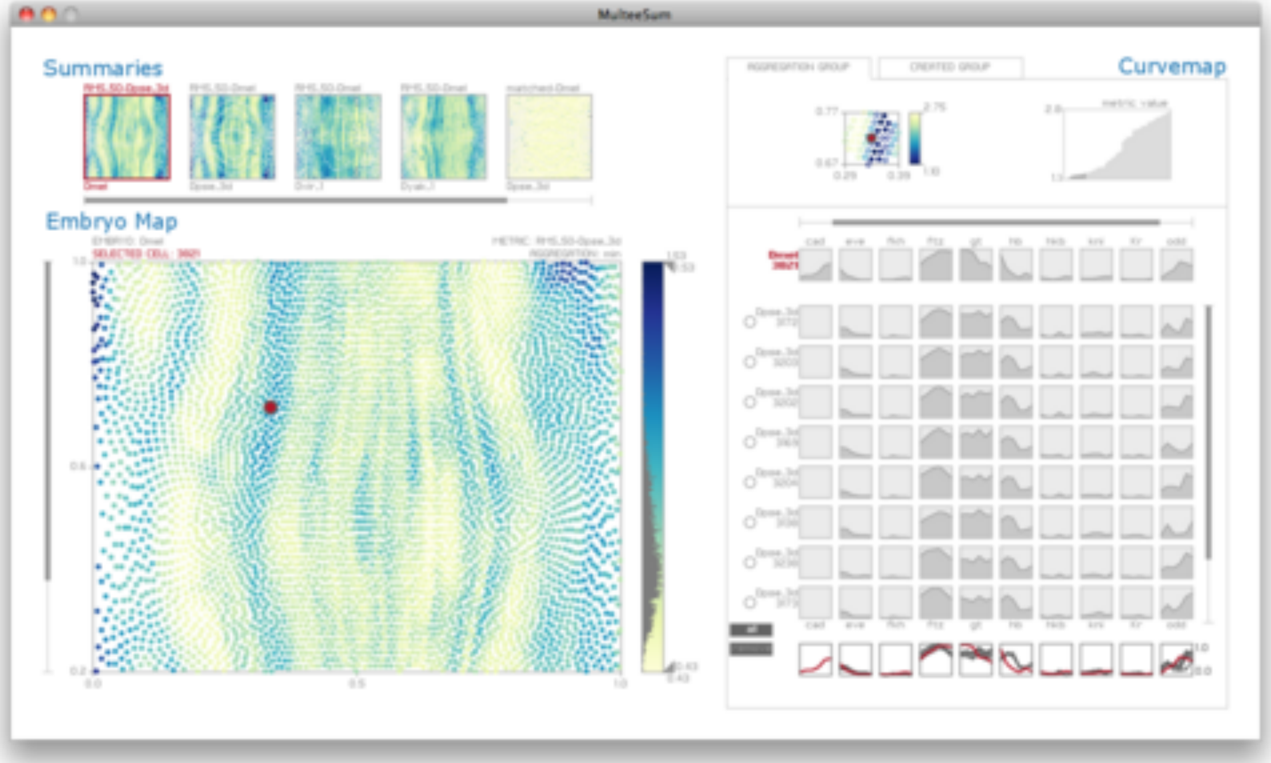


QSNE

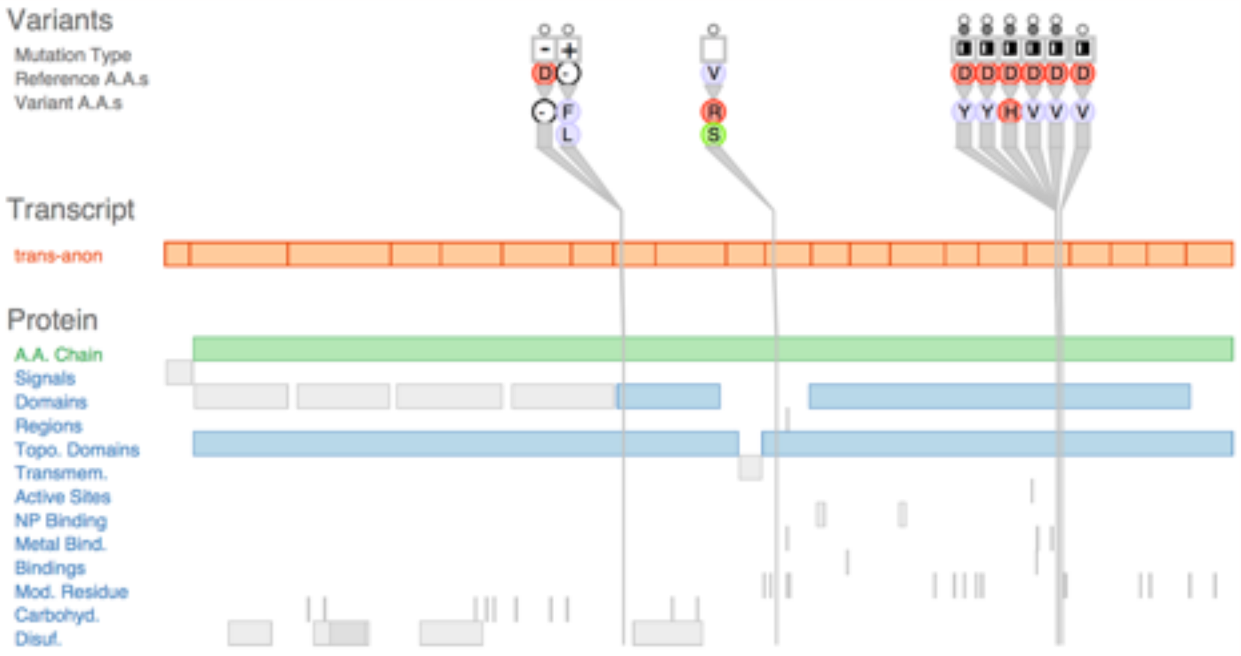
Problem-driven work: Genomics



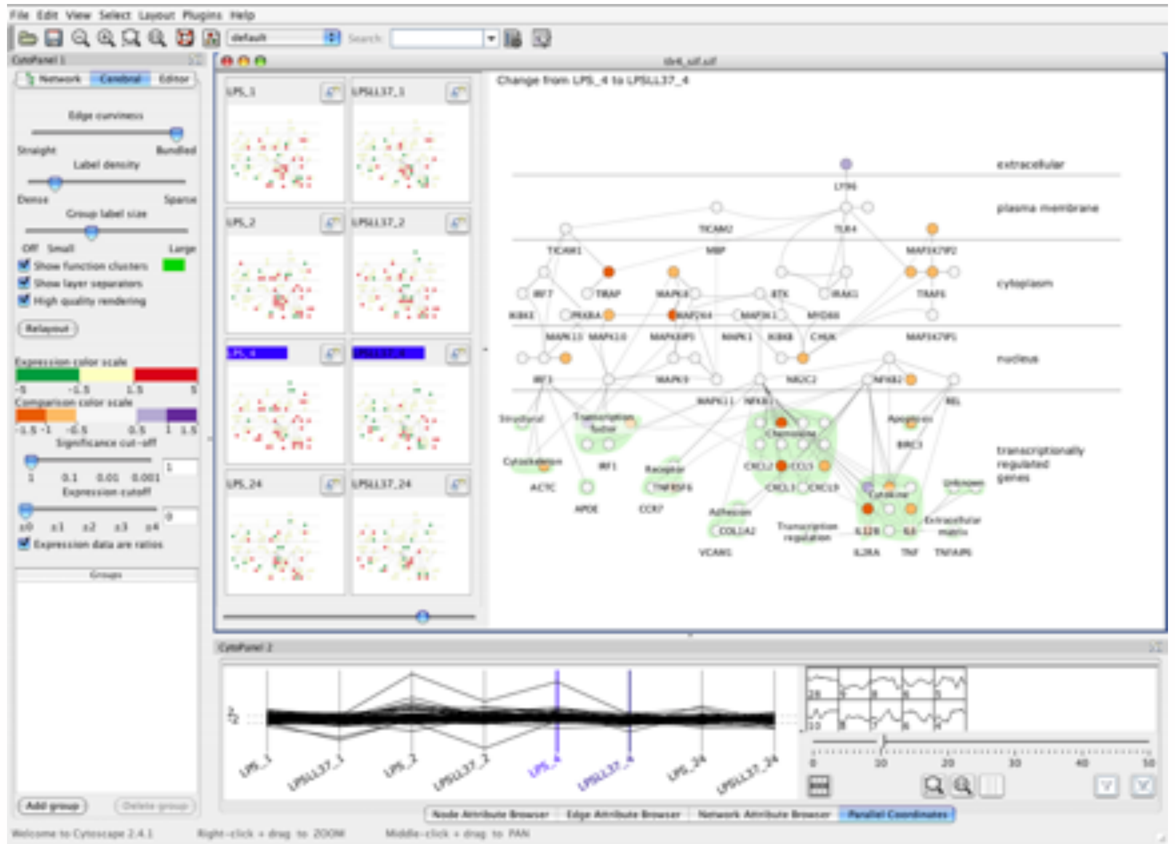
MizBee http://youtu.be/86p7brwuz2g



MulteeSum

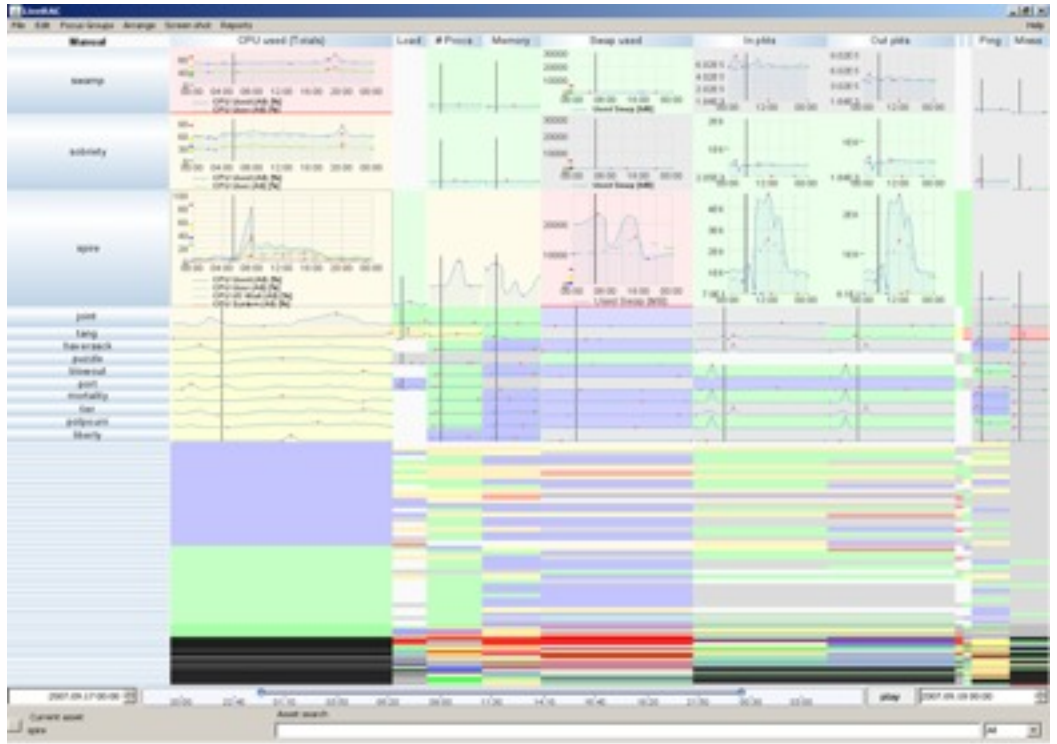


Variant View http://youtu.be/AHDnv_qMXxQ


File Edit View Select Layout Plugins Help
extracellular
plasma membrane
cytoplasm
nucleus
transcriptionally regulated genes

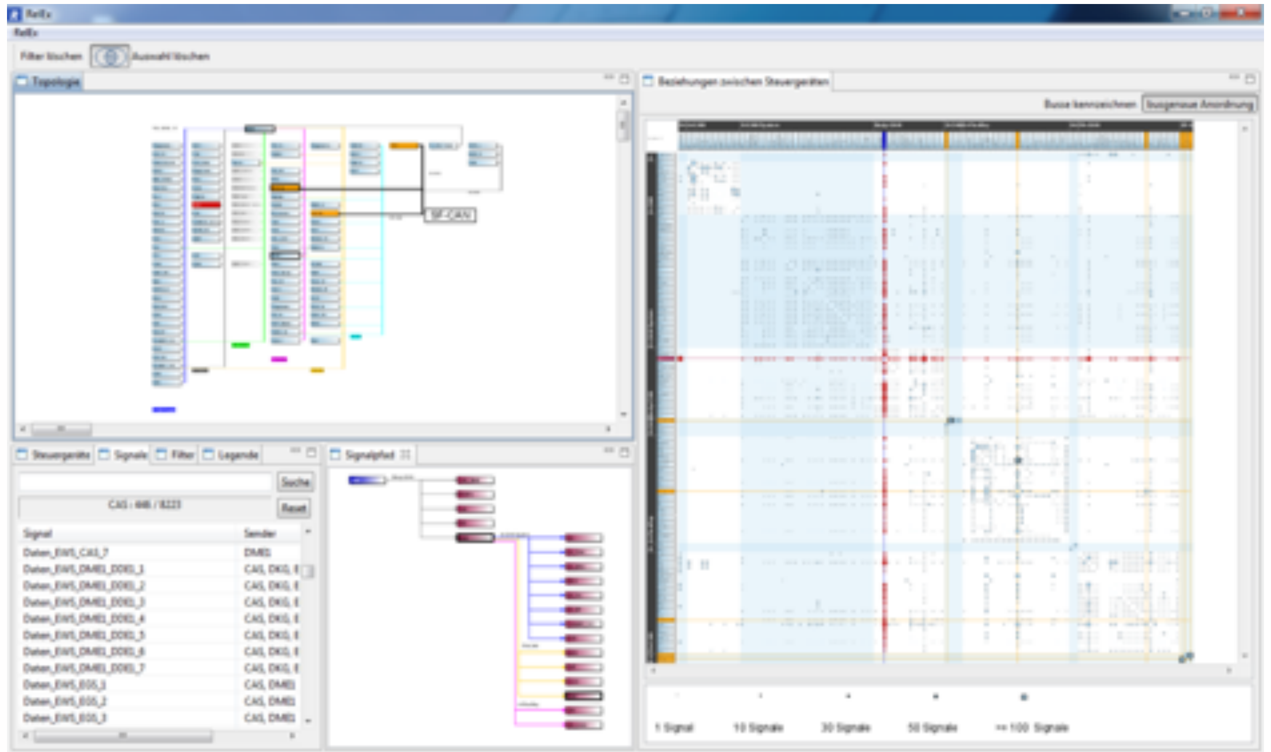
Cerebral http://youtu.be/76HhG1FQngI

Problem-driven work: Many domains

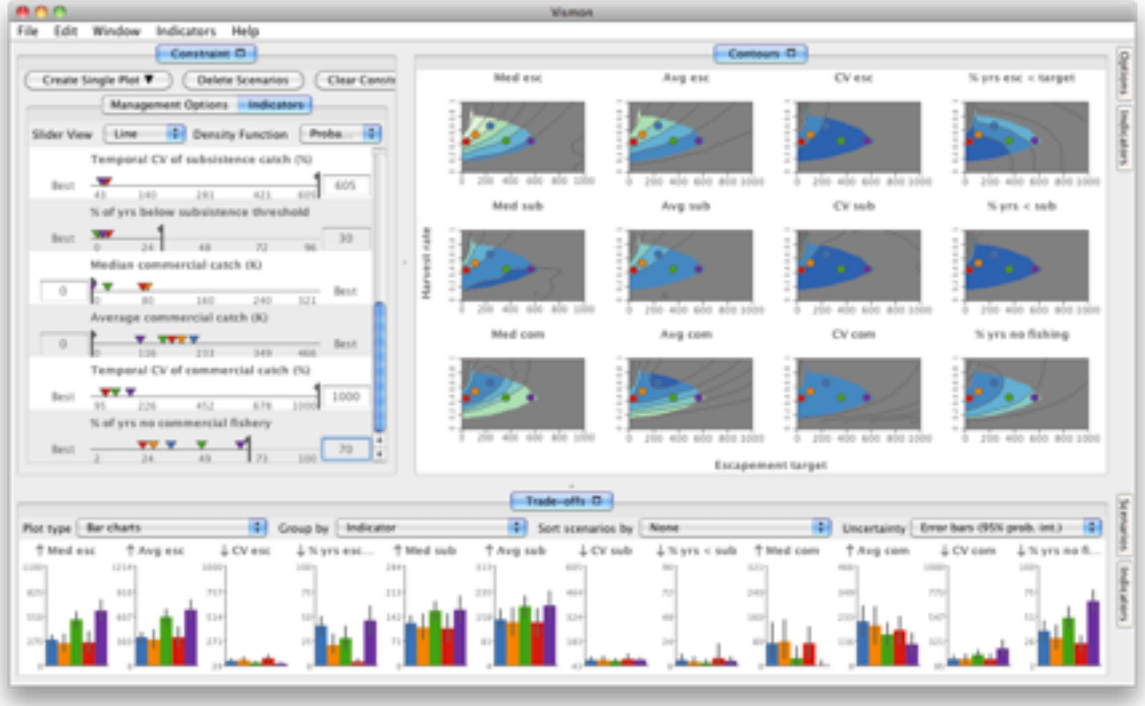


http://youtu.be/ld0c3H0VSkw

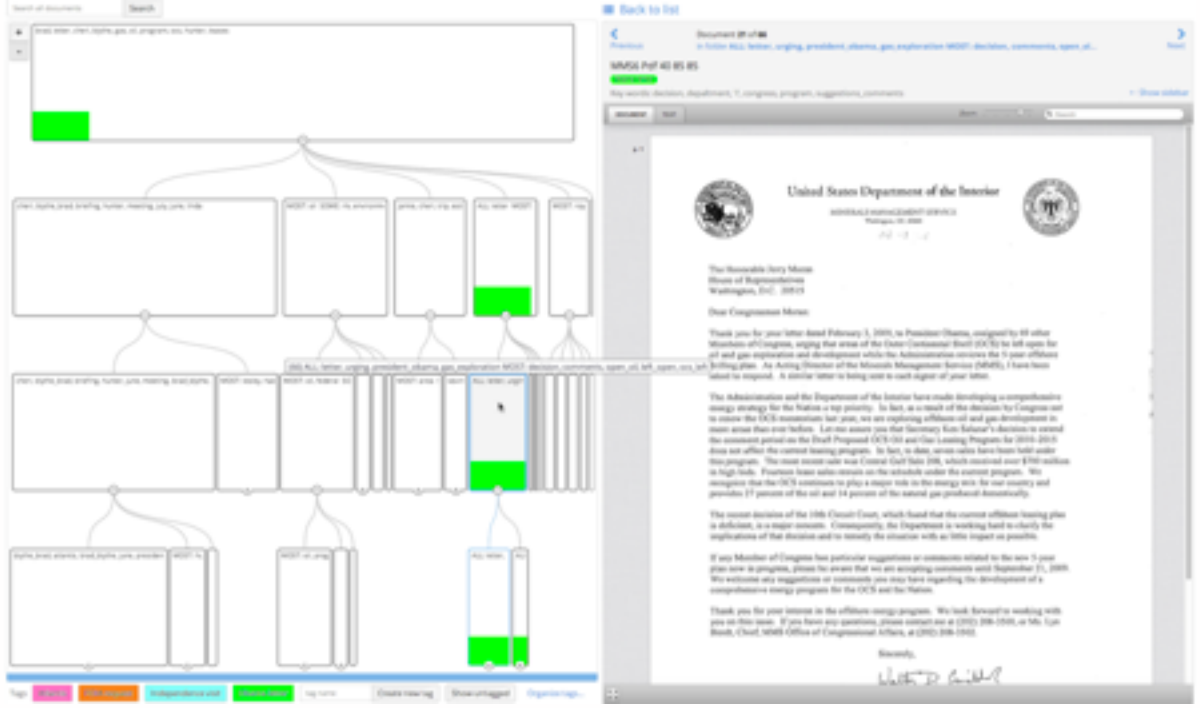
LiveRAC: system management time-series



RelEx: in-car overlay networks http://youtu.be/89lsQXc6Ao4



Vismon: fisheries management http://youtu.be/h0kHoS4VYmk

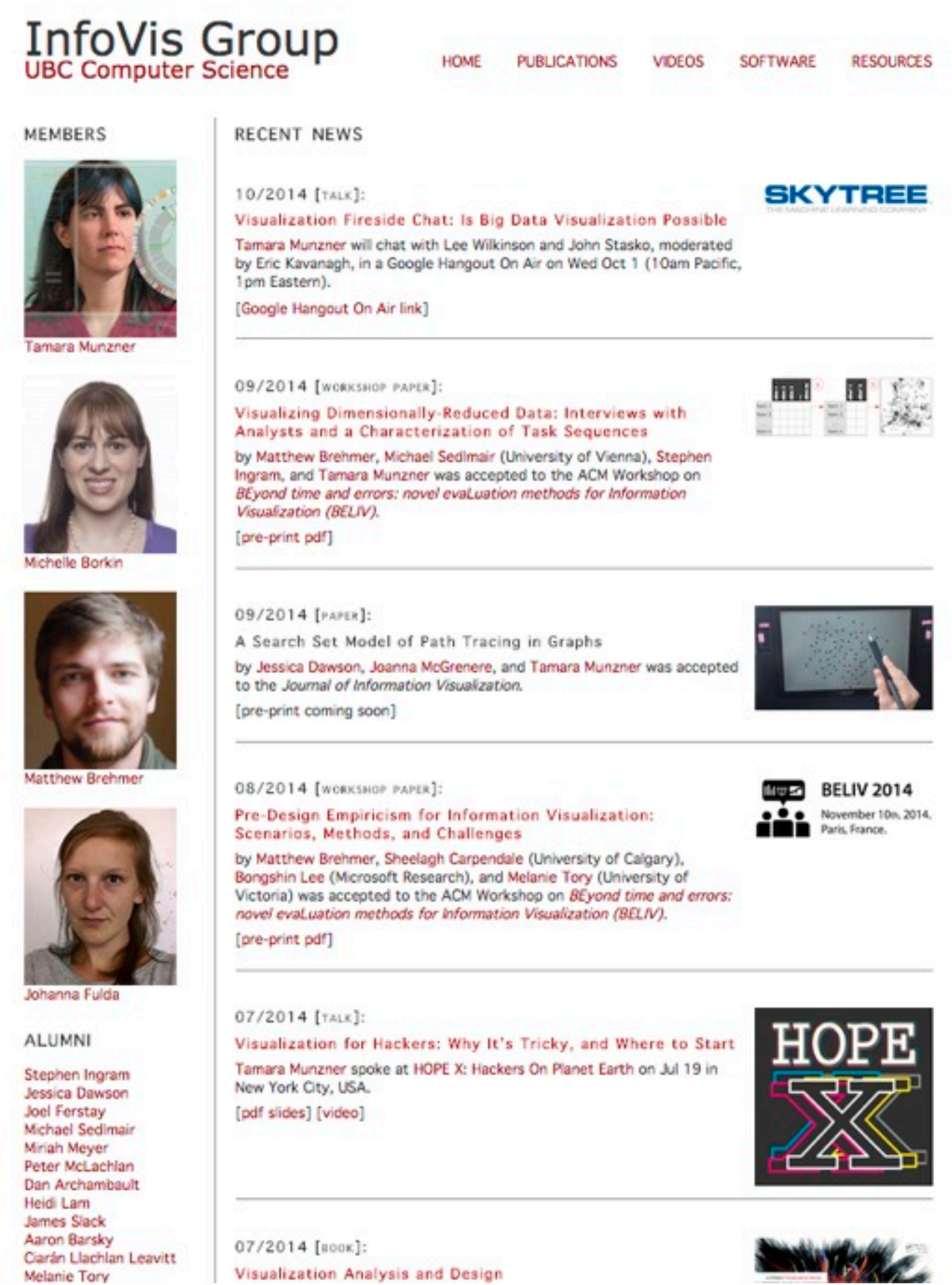


Overview: investigative journalism http://vimeo.com/71483614

More info

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

http://www.cs.ubc.ca/~tmm/talks.html#skytree14



InfoVis Group
UBC Computer Science

HOME PUBLICATIONS VIDEOS SOFTWARE RESOURCES

MEMBERS

Tamara Munzner

Michelle Borkin

Matthew Brehmer

Johanna Fulda

ALUMNI

Stephen Ingram
Jessica Dawson
Joel Ferstay
Michael Sedlmair
Miriah Meyer
Peter McLachlan
Dan Archambault
Heidi Lam
James Slack
Aaron Barsky
Ciarán Llachlan Leavitt
Melanie Tory

RECENT NEWS

10/2014 [TALK]:

Visualization Fireside Chat: Is Big Data Visualization Possible

Tamara Munzner will chat with Lee Wilkinson and John Stasko, moderated by Eric Kavanagh, in a Google Hangout On Air on Wed Oct 1 (10am Pacific, 1pm Eastern).

[Google Hangout On Air link]

09/2014 [WORKSHOP PAPER]:

Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences

by Matthew Brehmer, Michael Sedlmair (University of Vienna), Stephen Ingram, and Tamara Munzner was accepted to the ACM Workshop on *BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV).*

[pre-print pdf]

09/2014 [PAPER]:

A Search Set Model of Path Tracing in Graphs

by Jessica Dawson, Joanna McGrenere, and Tamara Munzner was accepted to the *Journal of Information Visualization.*

[pre-print coming soon]

08/2014 [WORKSHOP PAPER]:

Pre-Design Empiricism for Information Visualization: Scenarios, Methods, and Challenges

by Matthew Brehmer, Sheelagh Carpendale (University of Calgary), Bongshin Lee (Microsoft Research), and Melanie Tory (University of Victoria) was accepted to the ACM Workshop on *BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV).*

[pre-print pdf]

07/2014 [TALK]:

Visualization for Hackers: Why It's Tricky, and Where to Start

Tamara Munzner spoke at HOPE X: Hackers On Planet Earth on Jul 19 in New York City, USA.

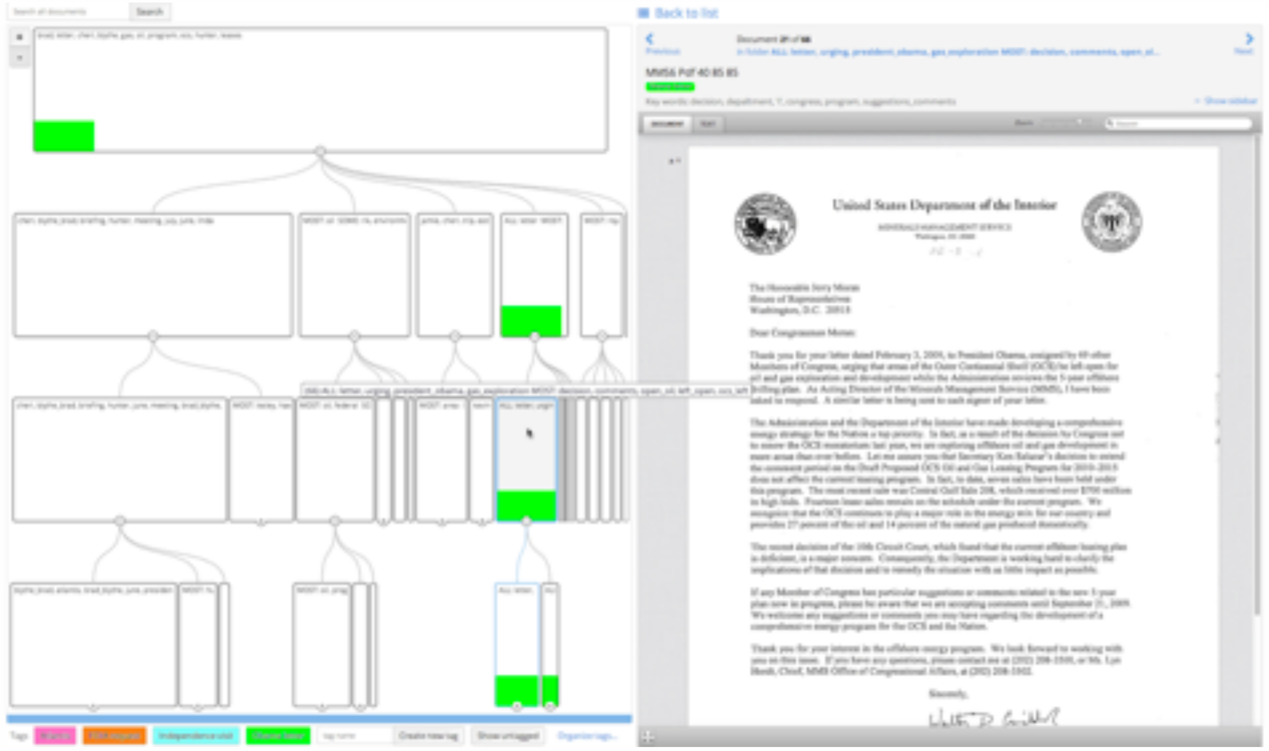
[pdf slides] [video]

07/2014 [BOOK]:

Visualization Analysis and Design

Overview design evolution

v4



MMS6 Pdf 40 85 85

Key words: decision, department, [illegible], congress, program, suggestions, comments

United States Department of the Interior

MINERALS MANAGEMENT SERVICE
Washington, DC 20240

The Honorable Jerry Moran
House of Representatives
Washington, D.C. 20515

Dear Congressman Moran:

Thank you for your letter dated February 3, 2009, to President Obama, cosigned by 69 other Members of Congress, urging that areas of the Outer Continental Shelf (OCS) be left open for oil and gas exploration and development while the Administration reviews the 5-year offshore drilling plan. As Acting Director of the Minerals Management Service (MMS), I have been asked to respond. A similar letter is being sent to each signer of your letter.

The Administration and the Department of the Interior have made developing a comprehensive energy strategy for the Nation a top priority. In fact, as a result of the decision by Congress not to renew the OCS moratorium last year, we are exploring offshore oil and gas development in more areas than ever before. Let me assure you that Secretary Ken Salazar's decision to extend the comment period on the Draft Proposed OCS Oil and Gas Leasing Program for 2010-2015 does not affect the current leasing program. In fact, to date, seven sales have been held under this program. The most recent sale was Central Gulf Sale 208, which received over $700 million in high bids. Fourteen lease sales remain on the schedule under the current program. We recognize that the OCS continues to play a major role in the energy mix for our country and provides 27 percent of the oil and 14 percent of the natural gas produced domestically.

The recent decision of the 10th Circuit Court, which found that the current offshore leasing plan is deficient, is a major concern. Consequently, the Department is working hard to clarify the implications of that decision and to remedy the situation with as little impact as possible.

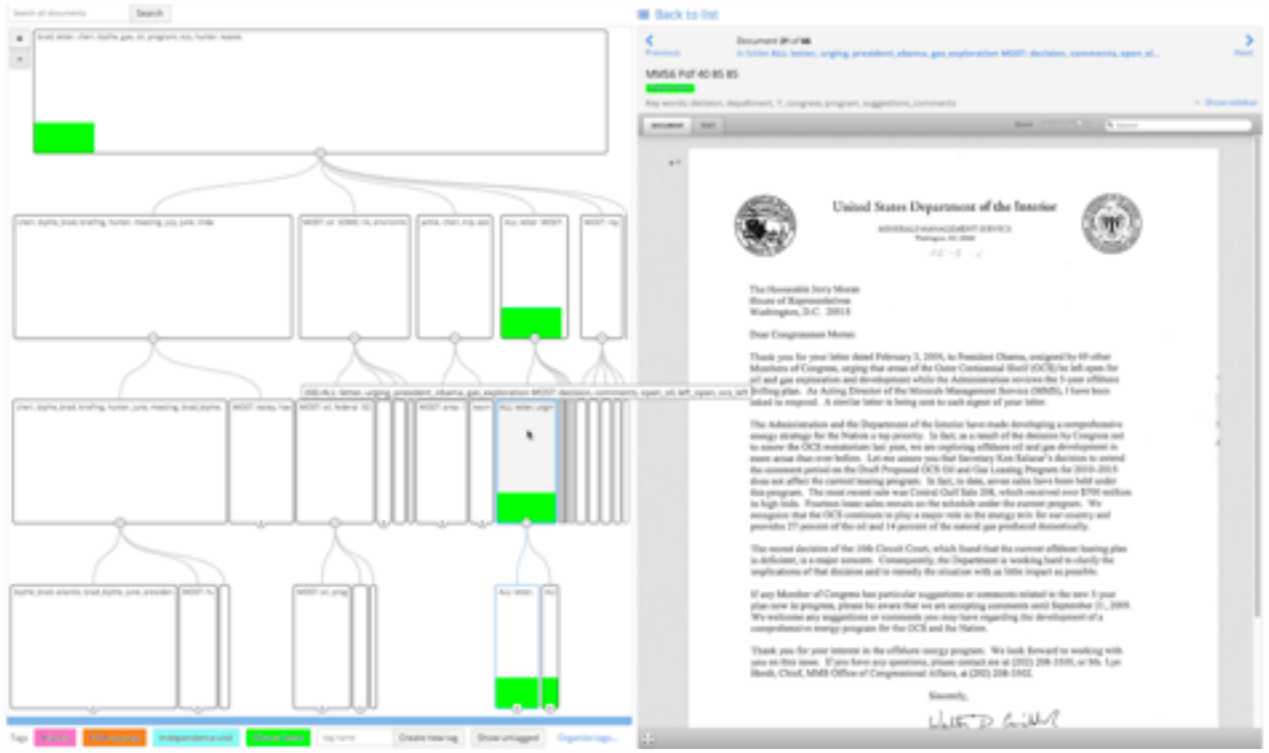
If any Member of Congress has particular suggestions or comments related to the new 5-year plan now in progress, please be aware that we are accepting comments until September 21, 2009. We welcome any suggestions or comments you may have regarding the development of a comprehensive energy program for the OCS and the Nation.

Thank you for your interest in the offshore energy program. We look forward to working with you on this issue. If you have any questions, please contact me at (202) 208-3500, or Ms. Lyn Herdt, Chief, MMS Office of Congressional Affairs, at (202) 208-3502.

Sincerely,

Overview design evolution

v4



United States Department of the Interior

MINERALS MANAGEMENT SERVICE
Washington, DC 20240

The Honorable Jerry Moran
House of Representatives
Washington, D.C. 20515

Dear Congressman Moran:

Thank you for your letter dated February 3, 2009, to President Obama, cosigned by 69 other Members of Congress, urging that areas of the Outer Continental Shelf (OCS) be left open for oil and gas exploration and development while the Administration reviews the 5-year offshore drilling plan. As Acting Director of the Minerals Management Service (MMS), I have been asked to respond. A similar letter is being sent to each signer of your letter.

The Administration and the Department of the Interior have made developing a comprehensive energy strategy for the Nation a top priority. In fact, as a result of the decision by Congress not to renew the OCS moratorium last year, we are exploring offshore oil and gas development in more areas than ever before. Let me assure you that Secretary Ken Salazar's decision to extend the comment period on the Draft Proposed OCS Oil and Gas Leasing Program for 2010-2015 does not affect the current leasing program. In fact, to date, seven sales have been held under this program. The most recent sale was Central Gulf Sale 208, which received over $700 million in high bids. Fourteen lease sales remain on the schedule under the current program. We recognize that the OCS continues to play a major role in the energy mix for our country and provides 27 percent of the oil and 14 percent of the natural gas produced domestically.

The recent decision of the 10th Circuit Court, which found that the current offshore leasing plan is deficient, is a major concern. Consequently, the Department is working hard to clarify the implications of that decision and to remedy the situation with as little impact as possible.

If any Member of Congress has particular suggestions or comments related to the new 5-year plan now in progress, please be aware that we are accepting comments until September 21, 2009. We welcome any suggestions or comments you may have regarding the development of a comprehensive energy program for the OCS and the Nation.

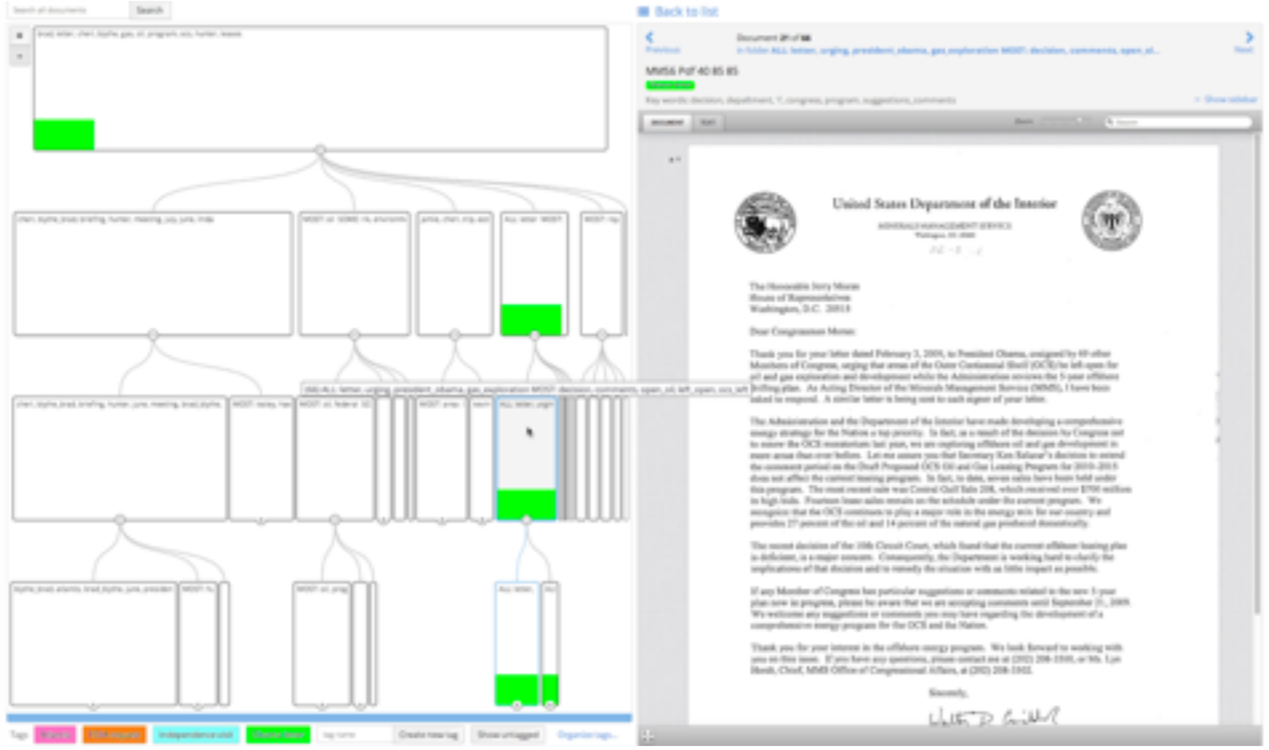
Thank you for your interest in the offshore energy program. We look forward to working with you on this issue. If you have any questions, please contact me at (202) 208-3500, or Ms. Lyn Herdt, Chief, MMS Office of Congressional Affairs, at (202) 208-3502.

Sincerely,

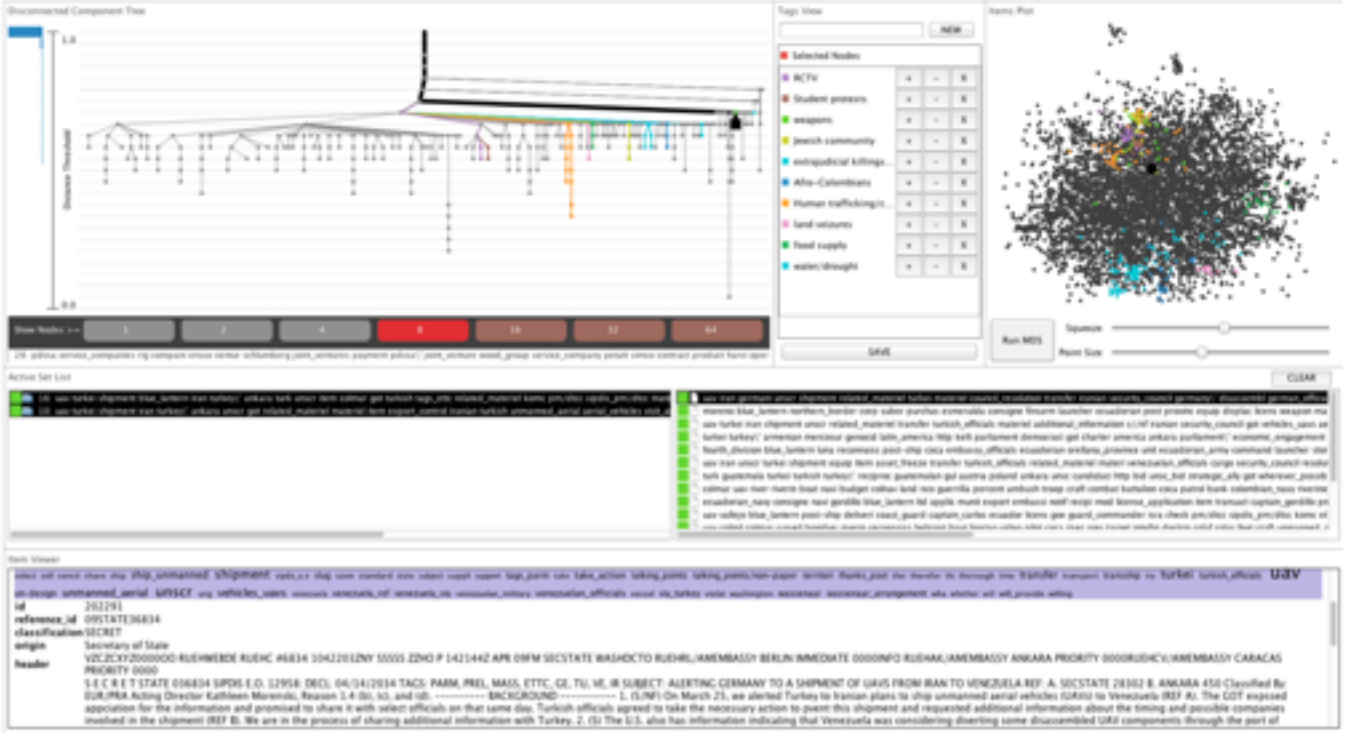
- how to find the needle in the haystack?
- how to convince that the haystack has no needles?

Overview design evolution

v4



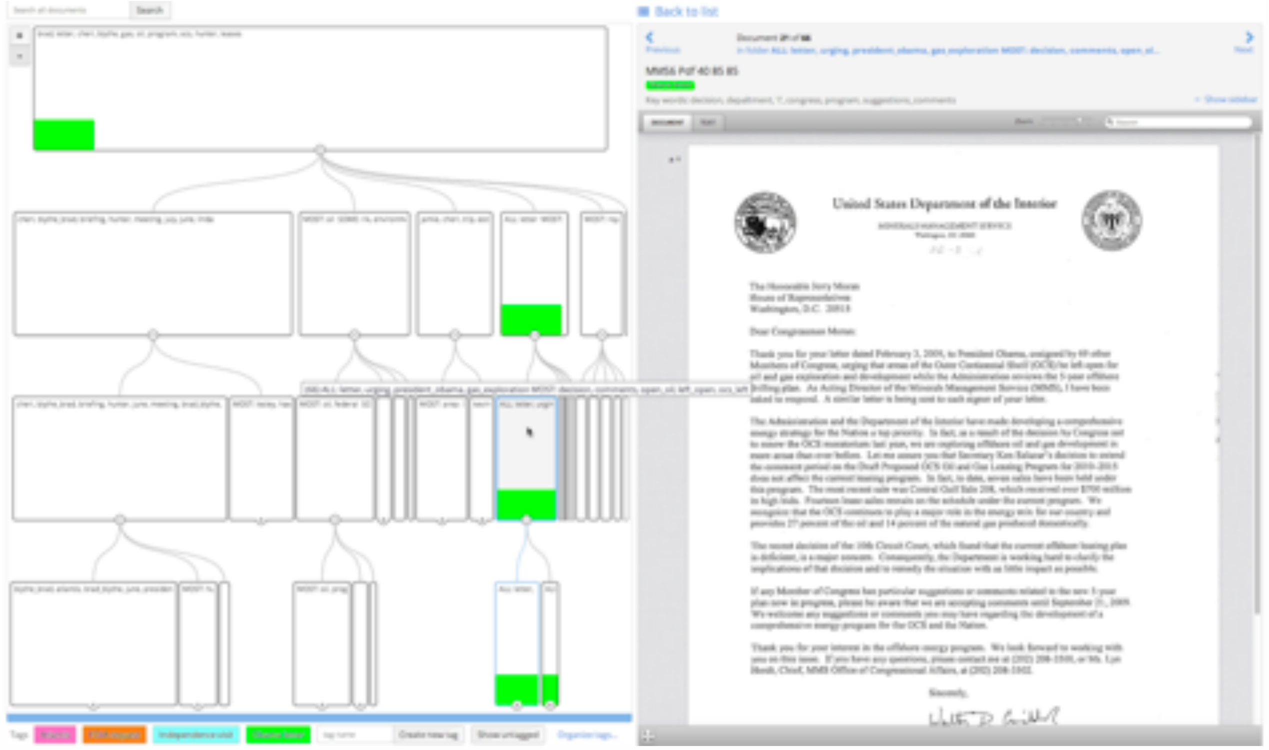
v1



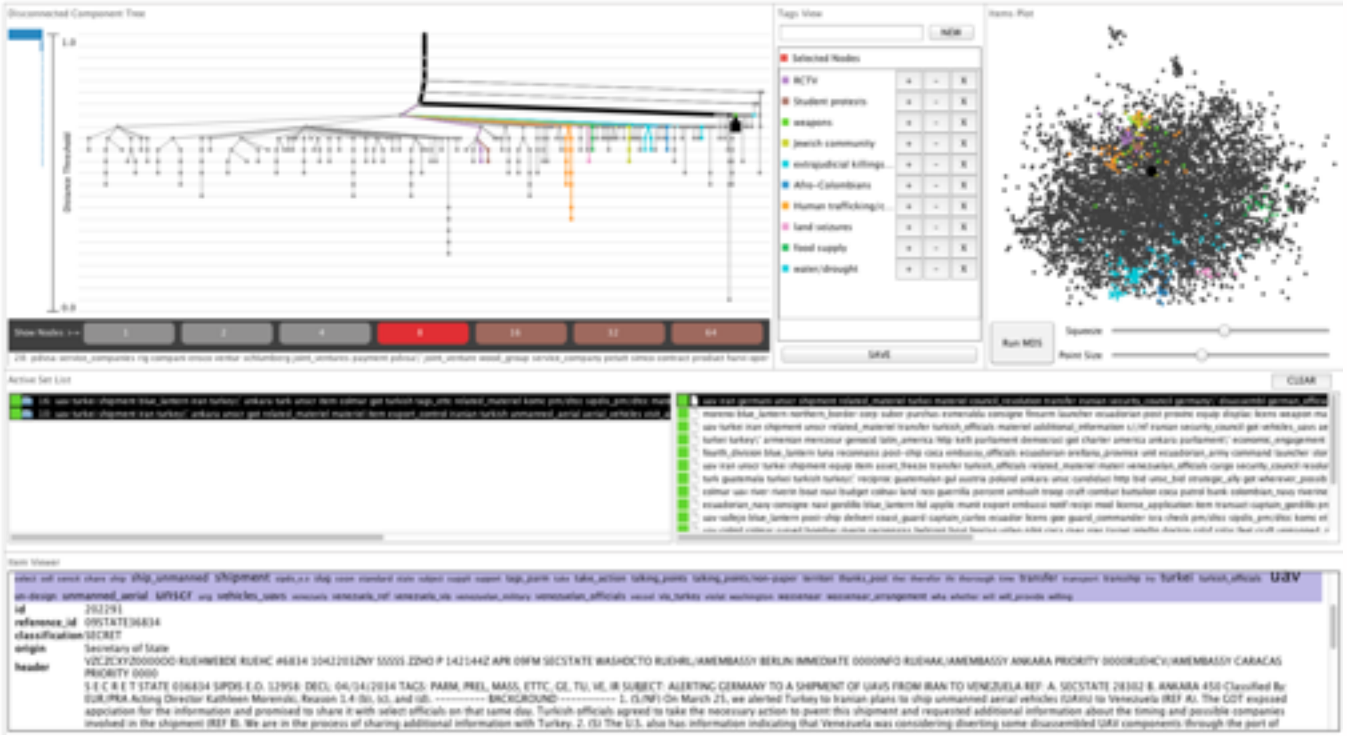
- how to find the needle in the haystack?
- how to convince that the haystack has no needles?

Overview design evolution

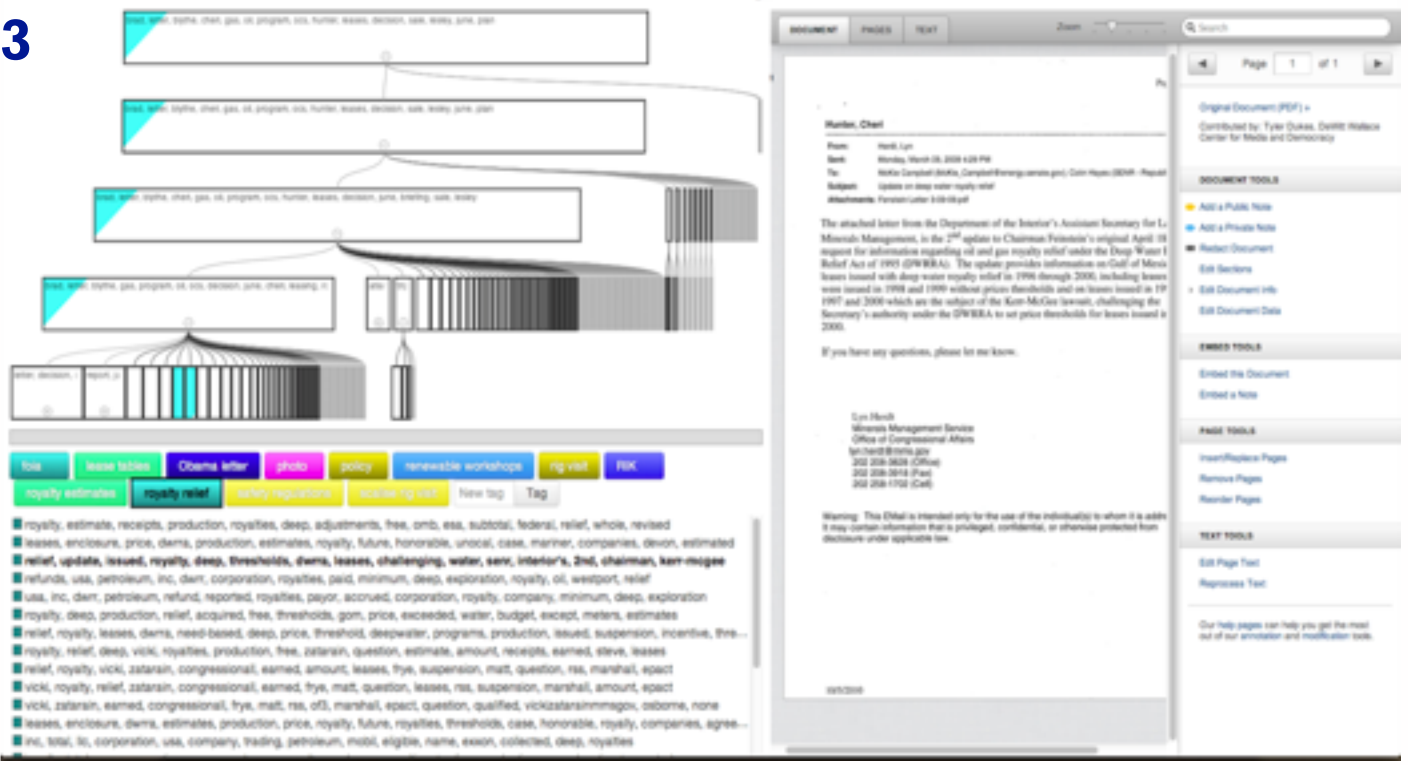
v4



v1



v3

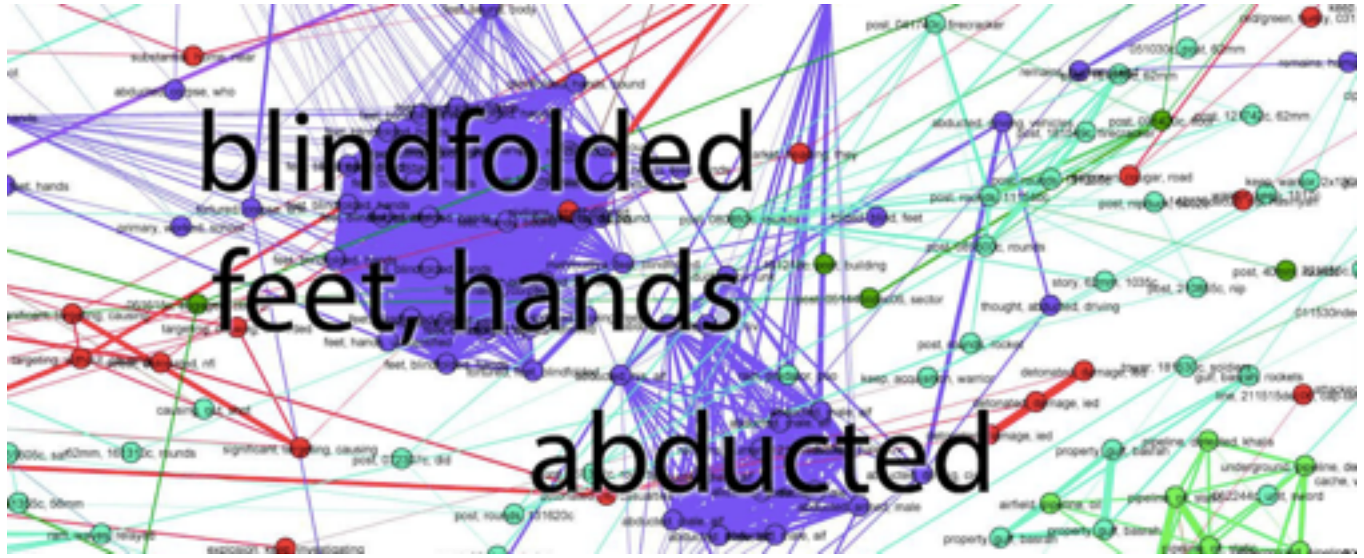


- how to find the needle in the haystack?
- how to convince that the haystack has no needles?

Overview origin story: WikiLeaks meets Glimmer

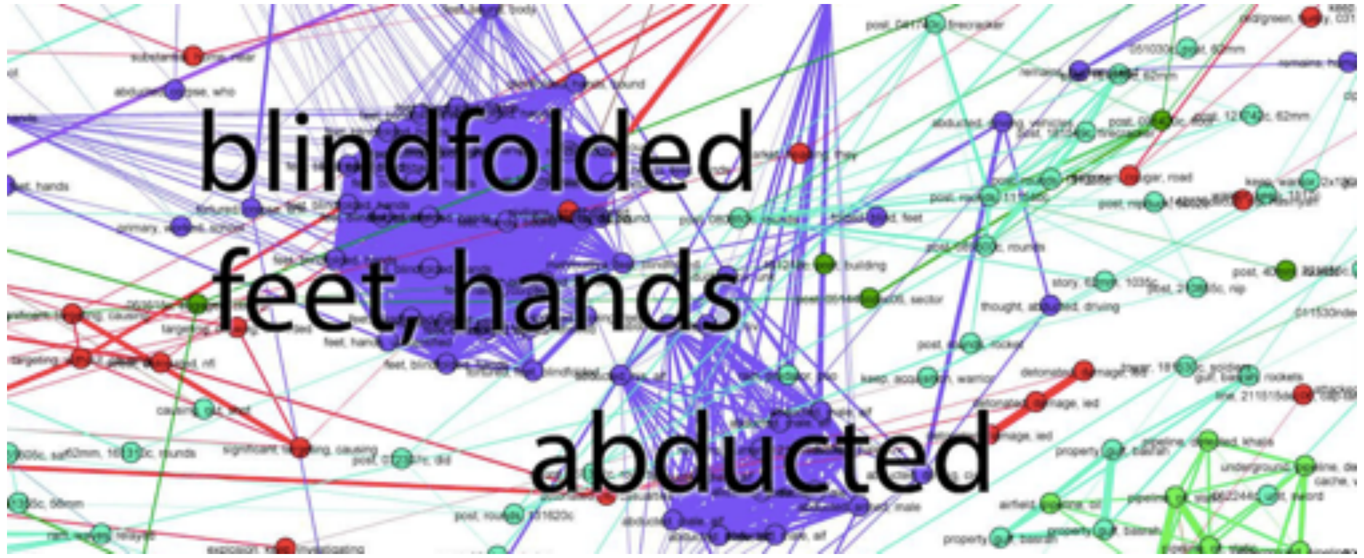
Overview origin story: WikiLeaks meets Glimmer

- WikiLeaks: hacker-journalist Jonathan Stray analyzing Iraq warlogs
  - conjecture that existing label classification falls short of showing all meaningful structure in data
    - friendly action, criminal incident, ...
  - had some NLP, needed better vis tools



Overview origin story: WikiLeaks meets Glimmer

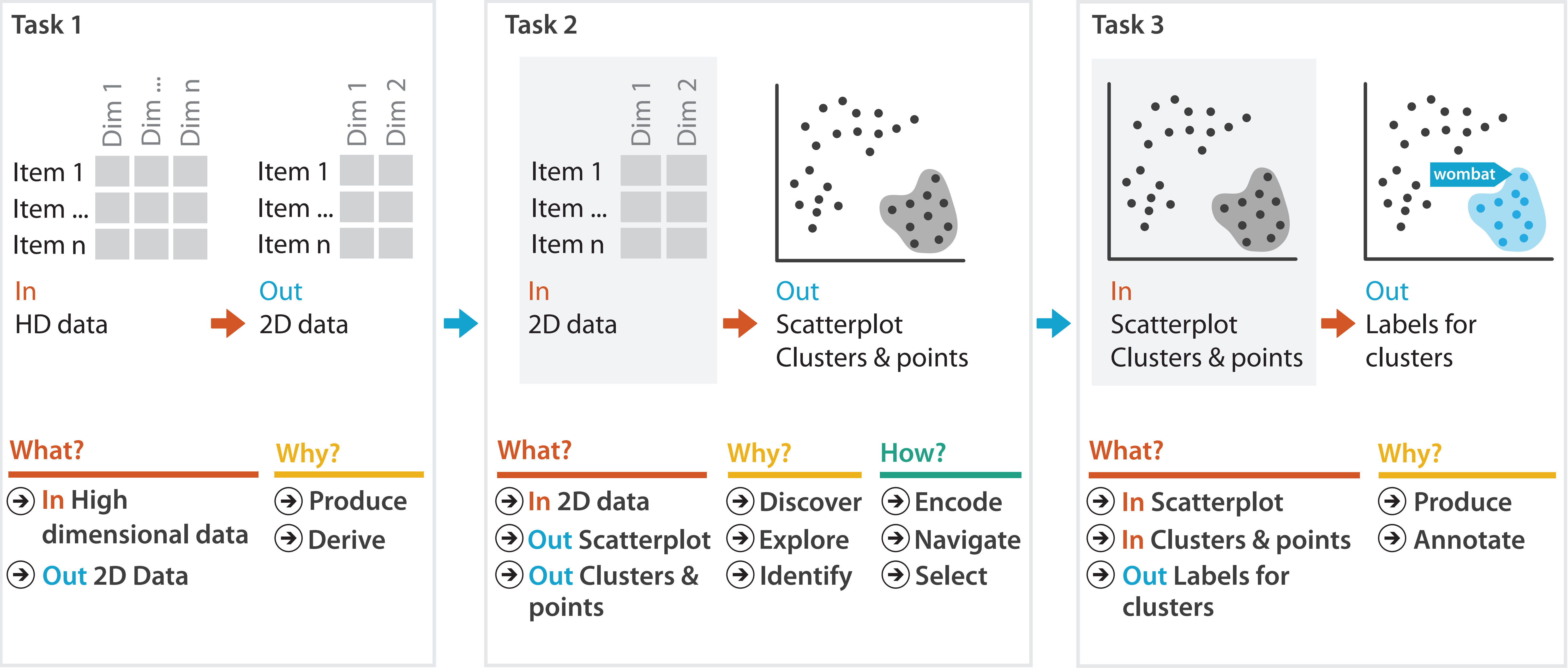
- WikiLeaks: hacker-journalist Jonathan Stray analyzing Iraq warlogs
  - conjecture that existing label classification falls short of showing all meaningful structure in data
    - friendly action, criminal incident, ...
  - had some NLP, needed better vis tools



- 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.]

Visual dimensionality reduction for document datasets



Task 1

In
HD data

Out
2D data

What?

- In High dimensional data
- Out 2D Data

Why?

- Produce
- Derive

Task 2

In
2D data

Out
Scatterplot
Clusters & points

What?

- In 2D data
- Out Scatterplot
- Out Clusters & points

Why?

- Discover
- Explore
- Identify

How?

- Encode
- Navigate
- Select

Task 3

In
Scatterplot
Clusters & points

Out
Labels for clusters

What?

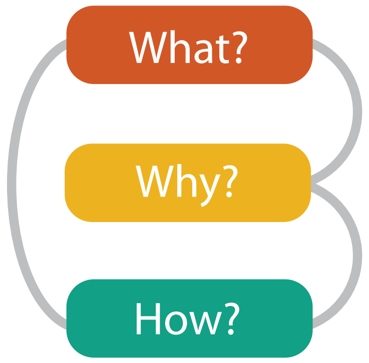
- In Scatterplot
- In Clusters & points
- Out Labels for clusters

Why?

- Produce
- Annotate

- more on visual DR: hour-long talk *Dimensionality Reduction from Several Angles*
  http://www.cs.ubc.ca/~tmm/talks.html#linz14

What/Why/How interplay



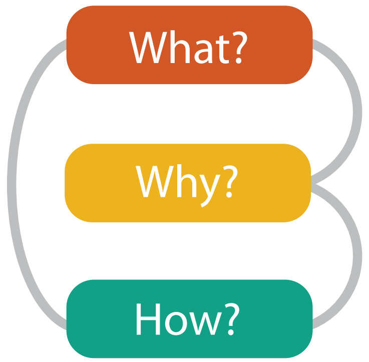
What?

Why?

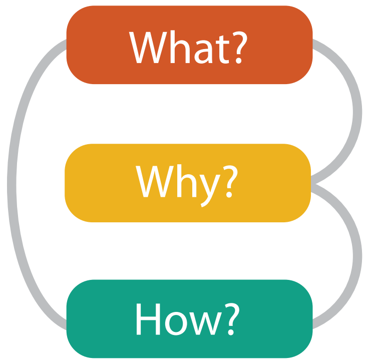
How?

What/Why/How interplay

- why: understand clusters



What/Why/How interplay



What?

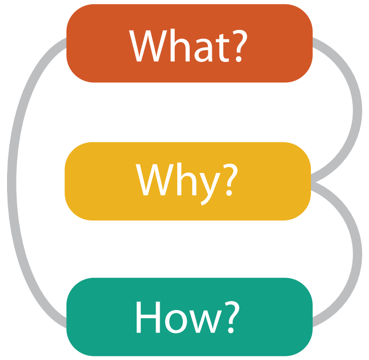
Why?

How?

- why: understand clusters

- what: derive data of full cluster hierarchy

What/Why/How interplay



- why: understand clusters

- what: derive data of full cluster hierarchy
  - explore space of possible clusterings

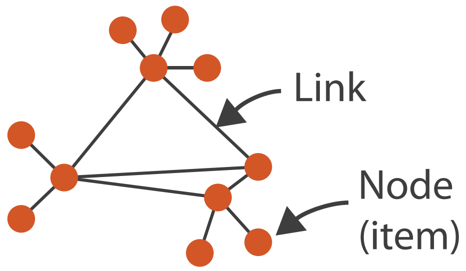
What/Why/How interplay

- why: understand clusters

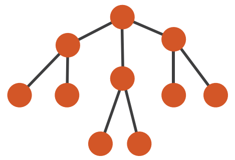
- what: derive data of full cluster hierarchy
  - explore space of possible clusterings

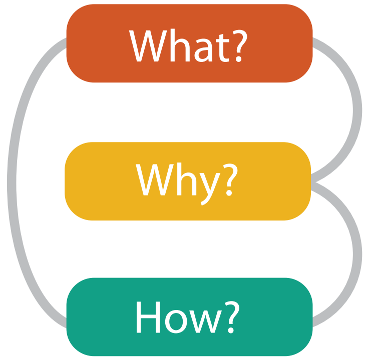
➔ Dataset Types

➔ Networks



➔ *Trees*





What?

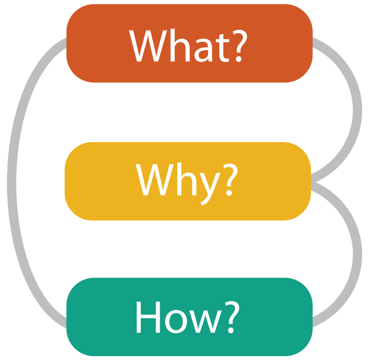
Why?

How?

What/Why/How interplay

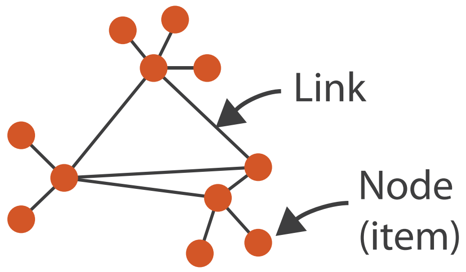
- why: understand clusters

- what: derive data of full cluster hierarchy
  - explore space of possible clusterings

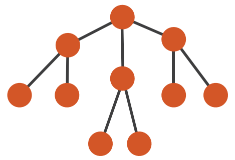


➔ Dataset Types

➔ Networks



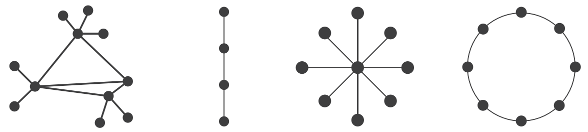
➔ *Trees*



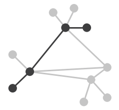
➔ Targets

➔ Network Data

➔ Topology



➔ *Paths*

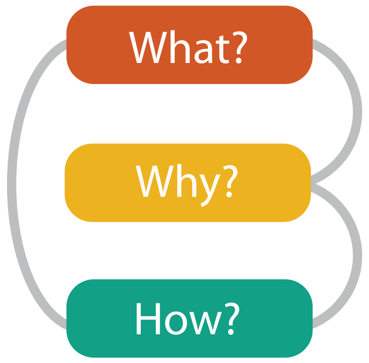


What/Why/How interplay

- why: understand clusters

- what: derive data of full cluster hierarchy
  - explore space of possible clusterings

- how: show cluster hierarchy



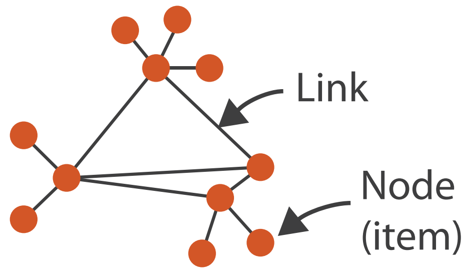
What?

Why?

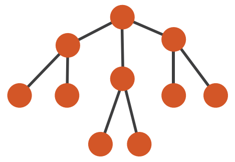
How?

Dataset Types

→ Networks



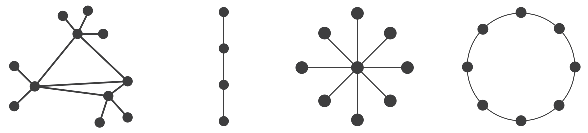
→ *Trees*



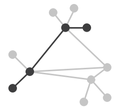
Targets

Network Data

→ Topology



→ *Paths*



What/Why/How interplay

- why: understand clusters

- what: derive data of full cluster hierarchy
  - explore space of possible clusterings

- how: show cluster hierarchy
  - arrange space: node-link

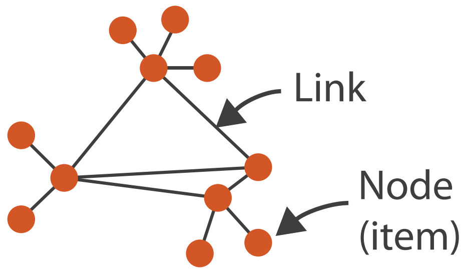
What?

Why?

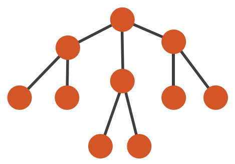
How?

➔ Dataset Types

➔ Networks



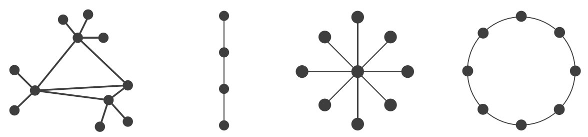
➔ *Trees*



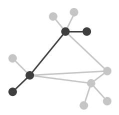
Targets

➔ Network Data

➔ Topology



➔ *Paths*

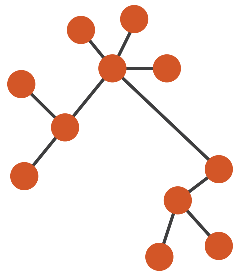


Arrange Networks And Trees

➔ **Node-link Diagrams**

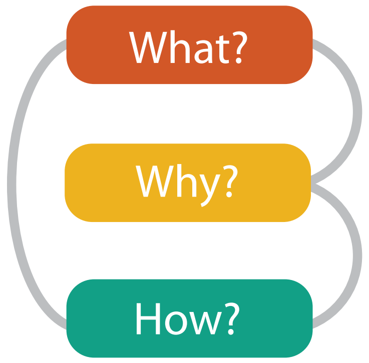
Connections and Marks

✓ NETWORKS ✓ TREES



What/Why/How interplay

- why: understand clusters
- what: derive data of full cluster hierarchy
  - explore space of possible clusterings
- how: show cluster hierarchy
  - arrange space: node-link
- how: support tagging clusters/docs



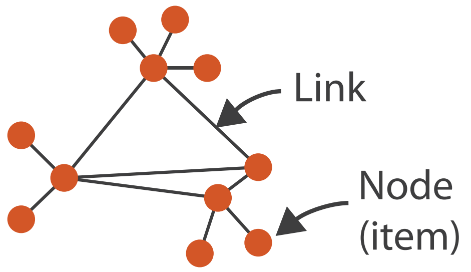
What?

Why?

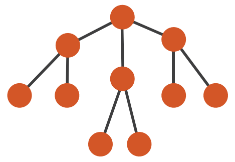
How?

➔ Dataset Types

➔ Networks



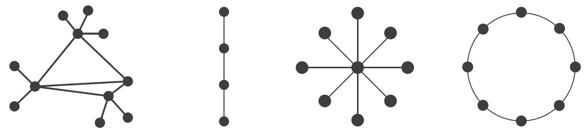
➔ *Trees*



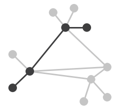
Targets

➔ Network Data

➔ Topology



➔ *Paths*

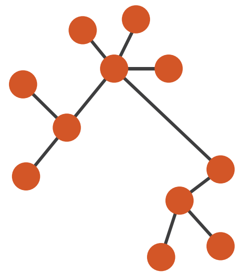


Arrange Networks And Trees

➔ Node-link Diagrams

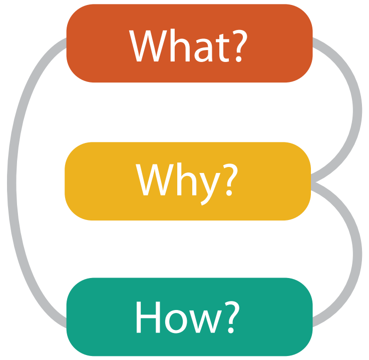
Connections and Marks

✓ NETWORKS ✓ TREES



What/Why/How interplay

- why: understand clusters
- what: derive data of full cluster hierarchy
  - explore space of possible clusterings
- how: show cluster hierarchy
  - arrange space: node-link
- how: support tagging clusters/docs

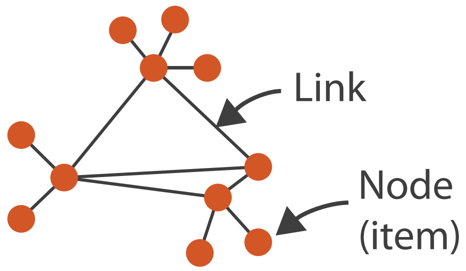

What?

Why?

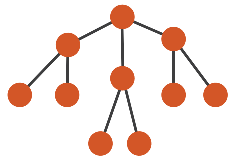
How?

➔ Dataset Types

➔ Networks



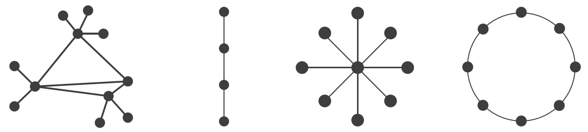
➔ *Trees*



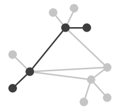
Targets

➔ Network Data

➔ Topology

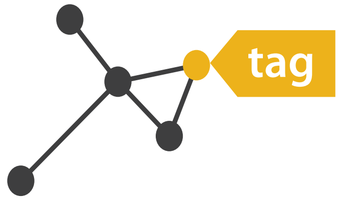


➔ *Paths*



➔ Produce

➔ *Annotate*

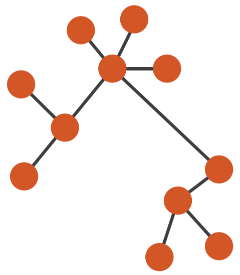


Arrange Networks And Trees

➔ **Node-link Diagrams**

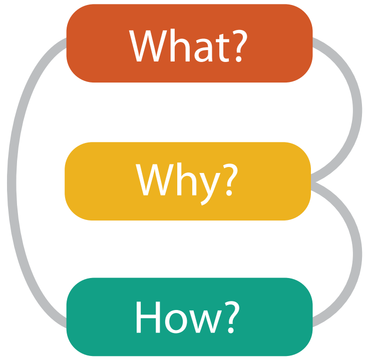
Connections and Marks

✓ NETWORKS ✓ TREES



What/Why/How interplay

- why: understand clusters
- what: derive data of full cluster hierarchy
  - – explore space of possible clusterings
- how: show cluster hierarchy
  - – arrange space: node-link
- how: support tagging clusters/docs
  - – following *or* cross-cutting hierarchy!

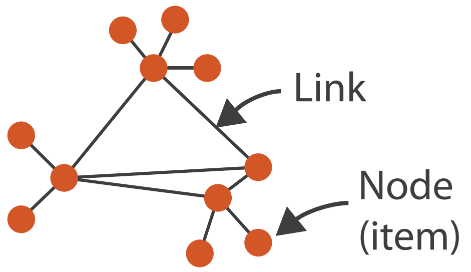

What?

Why?

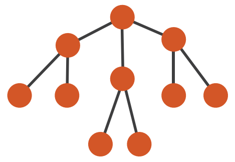
How?

➔ **Dataset Types**

➔ Networks



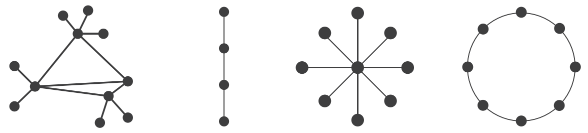
➔ *Trees*



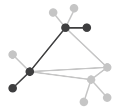
Targets

➔ **Network Data**

➔ Topology

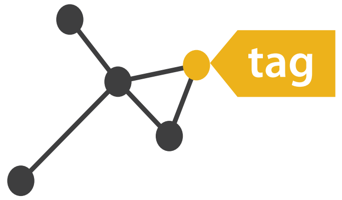


➔ *Paths*



➔ Produce

➔ *Annotate*

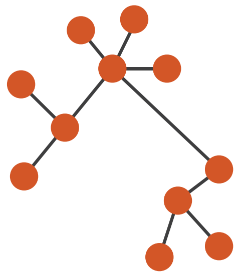


Arrange Networks And Trees

➔ **Node-link Diagrams**

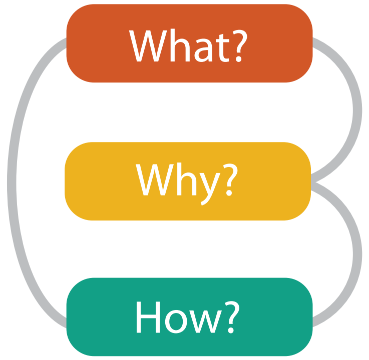
Connections and Marks

✓ NETWORKS ✓ TREES



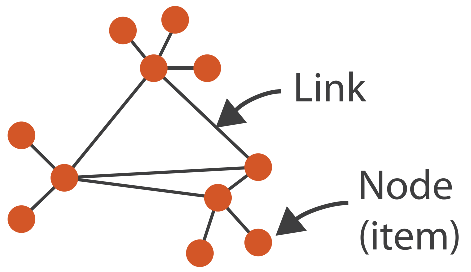
What/Why/How interplay

- why: understand clusters
- what: derive data of full cluster hierarchy
  - explore space of possible clusterings
- how: show cluster hierarchy
  - arrange space: node-link
- how: support tagging clusters/docs
  - following *or* cross-cutting hierarchy!
    - simple annotation

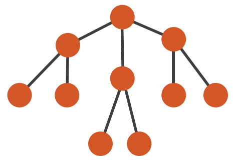


➔ Dataset Types

➔ Networks



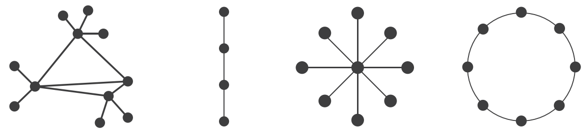
➔ *Trees*



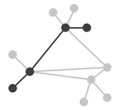
Targets

➔ Network Data

➔ Topology

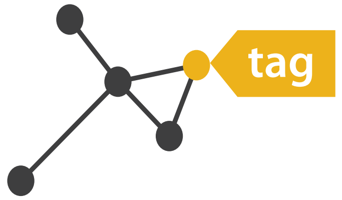


➔ *Paths*



➔ Produce

➔ *Annotate*

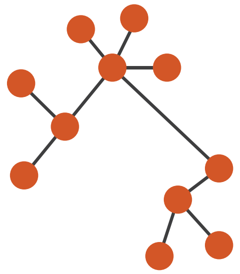


Arrange Networks And Trees

➔ Node-link Diagrams

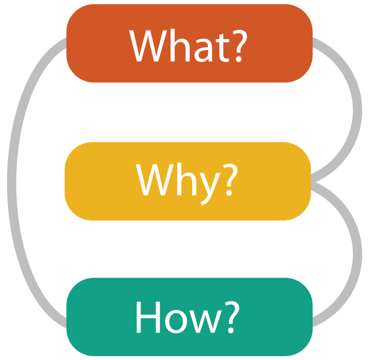
Connections and Marks

✓ NETWORKS ✓ TREES



What/Why/How interplay

- why: understand clusters
- what: derive data of full cluster hierarchy
  - explore space of possible clusterings
- how: show cluster hierarchy
  - arrange space: node-link
- how: support tagging clusters/docs
  - following *or* cross-cutting hierarchy!
    - simple annotation
    - progress tracking

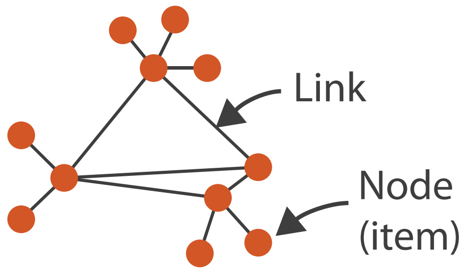

What?

Why?

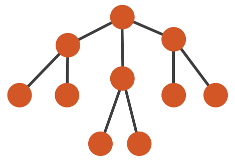
How?

➔ **Dataset Types**

➔ Networks



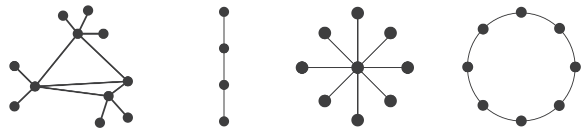
➔ *Trees*



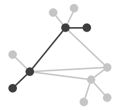
Targets

➔ **Network Data**

➔ Topology

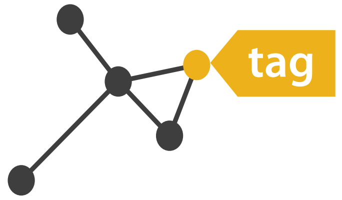


➔ *Paths*



➔ Produce

➔ *Annotate*

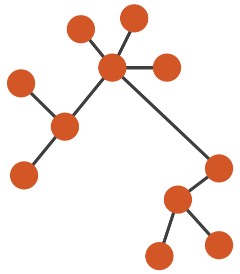


Arrange Networks And Trees

➔ **Node-link Diagrams**

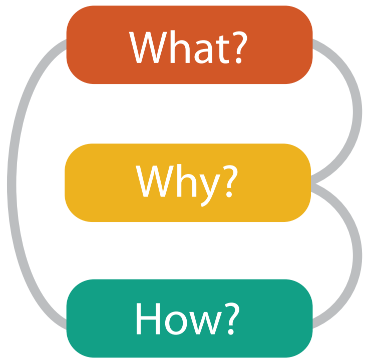
Connections and Marks

✓ NETWORKS ✓ TREES



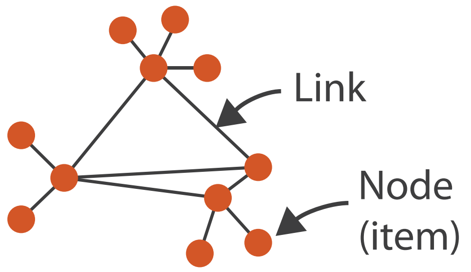
What/Why/How interplay

- why: understand clusters
- what: derive data of full cluster hierarchy
  - explore space of possible clusterings
- how: show cluster hierarchy
  - arrange space: node-link
- how: support tagging clusters/docs
  - following *or* cross-cutting hierarchy!
    - simple annotation
    - progress tracking
    - user-defined semantics

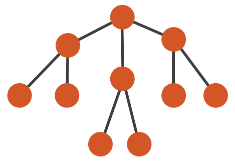


➔ **Dataset Types**

➔ Networks



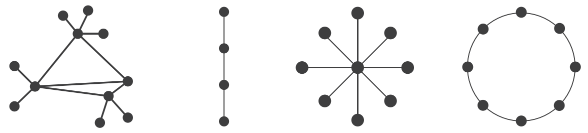
➔ *Trees*



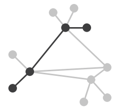
➔ **Targets**

➔ **Network Data**

➔ Topology

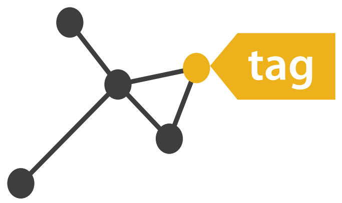


➔ *Paths*



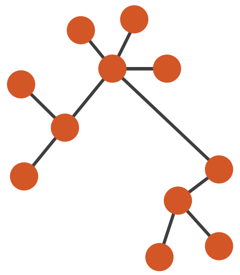
➔ Produce

➔ *Annotate*



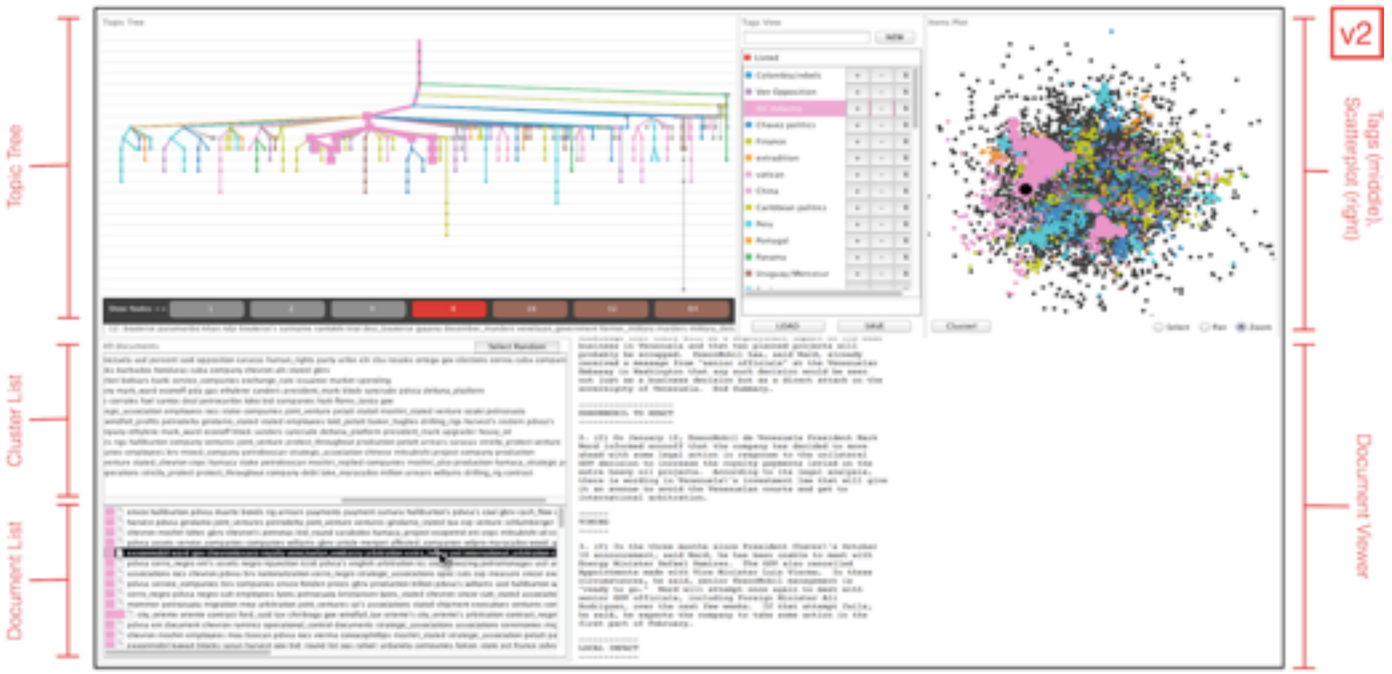
Arrange Networks And Trees

➔ **Node-link Diagrams**
Connections and Marks
✓ NETWORKS ✓ TREES



How: Idiom design decisions

- facet: juxtapose linked views
  - linked color coding
    - cluster hierarchy tree
    - DR scatterplot
    - tags
  - reading text/keywords
    - cluster list
    - doc reader



What?

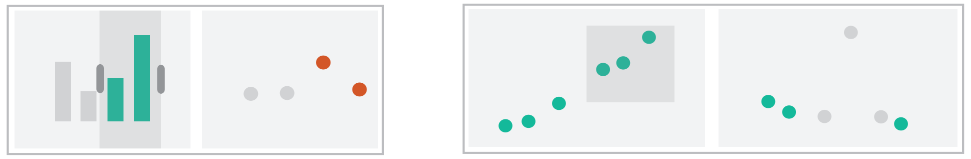
Why?

How?

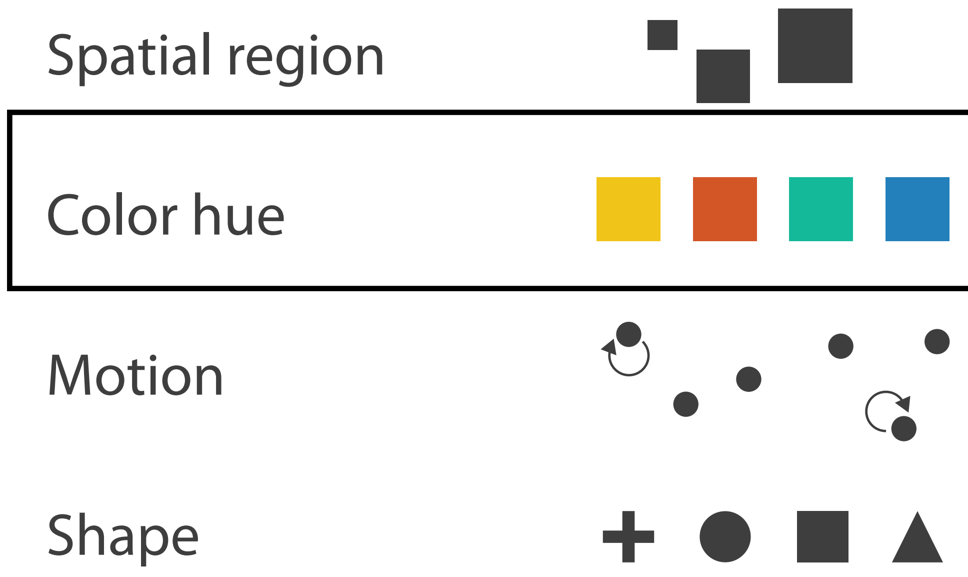
➔ **Juxtapose and Coordinate Views**

➔ Share Encoding: Same/Different

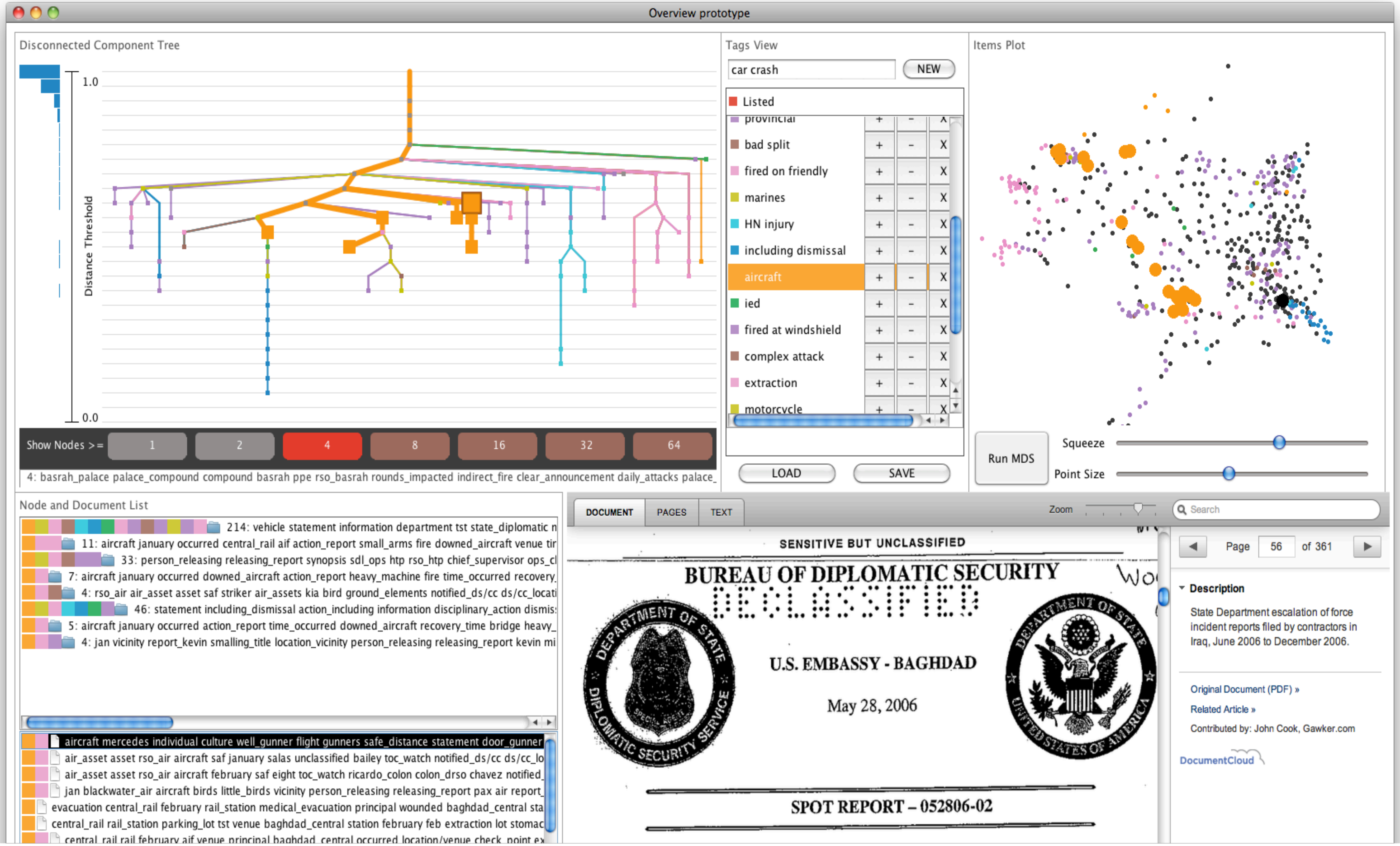
➔ *Linked Highlighting*



➔ **Identity Channels: Categorical Attributes**



Overview video (version 1)



car crash

- provincial
- bad split
- fired on friendly
- marines
- HN injury
- including dismissal
- aircraft
- ied
- fired at windshield
- complex attack
- extraction
- motorcycle

4: basrah_palace palace_compound compound basrah ppe rso_basrah rounds_impacted indirect_fire clear_announcement daily_attacks palace_

214: vehicle statement information department tst state_diplomatic n
11: aircraft january occurred central_rail aif action_report small_arms fire downed_aircraft venue tir
33: person_releasing releasing_report synopsis sdl_ops htp rso_htp chief_supervisor ops_cl
7: aircraft january occurred downed_aircraft action_report heavy_machine fire time_occurred recovery_
4: rso_air air_asset asset saf striker air_assets kia bird ground_elements notified_ds/cc ds/cc_locati
46: statement including_dismissal action_including information disciplinary_action dismis
5: aircraft january occurred action_report time_occurred downed_aircraft recovery_time bridge heavy_
4: jan vicinity report_kevin smalling_title location_vicinity person_releasing releasing_report kevin mi

aircraft mercedes individual culture well_gunner flight gunners safe_distance statement door_gunner
air_asset asset rso_air aircraft saf january salas unclassified bailey toc_watch notified_ds/cc ds/cc_lo
air_asset asset rso_air aircraft february saf eight toc_watch ricardo_colon colon_drso chavez notified_
jan blackwater_air aircraft birds little_birds vicinity person_releasing releasing_report pax air report_
evacuation central_rail february rail_station medical_evacuation principal wounded baghdad_central sta
central_rail rail_station parking_lot tst venue baghdad_central station february feb extraction lot stomac
central_rail rail february aif venue principal baghdad_central occurred location/venue check_point ex

SENSITIVE BUT UNCLASSIFIED

BUREAU OF DIPLOMATIC SECURITY

DECLASSIFIED

U.S. EMBASSY - BAGHDAD

May 28, 2006

SPOT REPORT – 052806-02

State Department escalation of force incident reports filed by contractors in Iraq, June 2006 to December 2006.

Contributed by: John Cook, Gawker.com

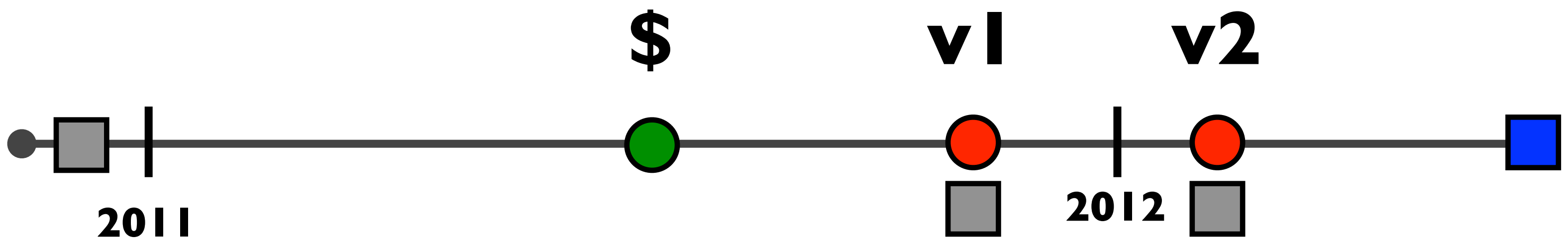
http://www.cs.ubc.ca/labs/imager/tr/2012/modiscotag/

Path to adoption

- version 1
  - fast cluster hierarchy construction for sparse data
  - research prototype by PhD student
  - positive initial assessment from AP Caracas bureau chief
    - barrier to adoption: difficult install/load process

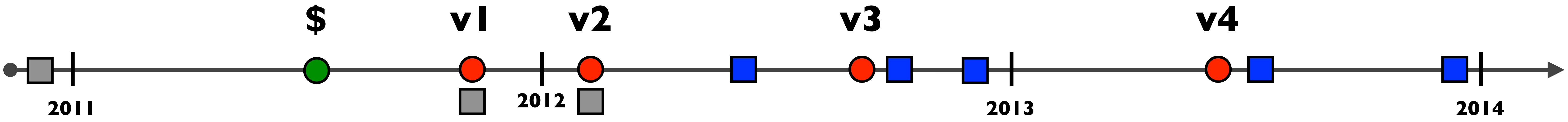
Path to adoption

- version 1
  - fast cluster hierarchy construction for sparse data
  - research prototype by PhD student
  - positive initial assessment from AP Caracas bureau chief
    - barrier to adoption: difficult install/load process
- version 2
  - web deployment, DocumentCloud integration, usability
    - many months of engineering
      - Knight Foundation funding to the rescue!
    - published story by unaffiliated reporter: police corruption in Tulsa

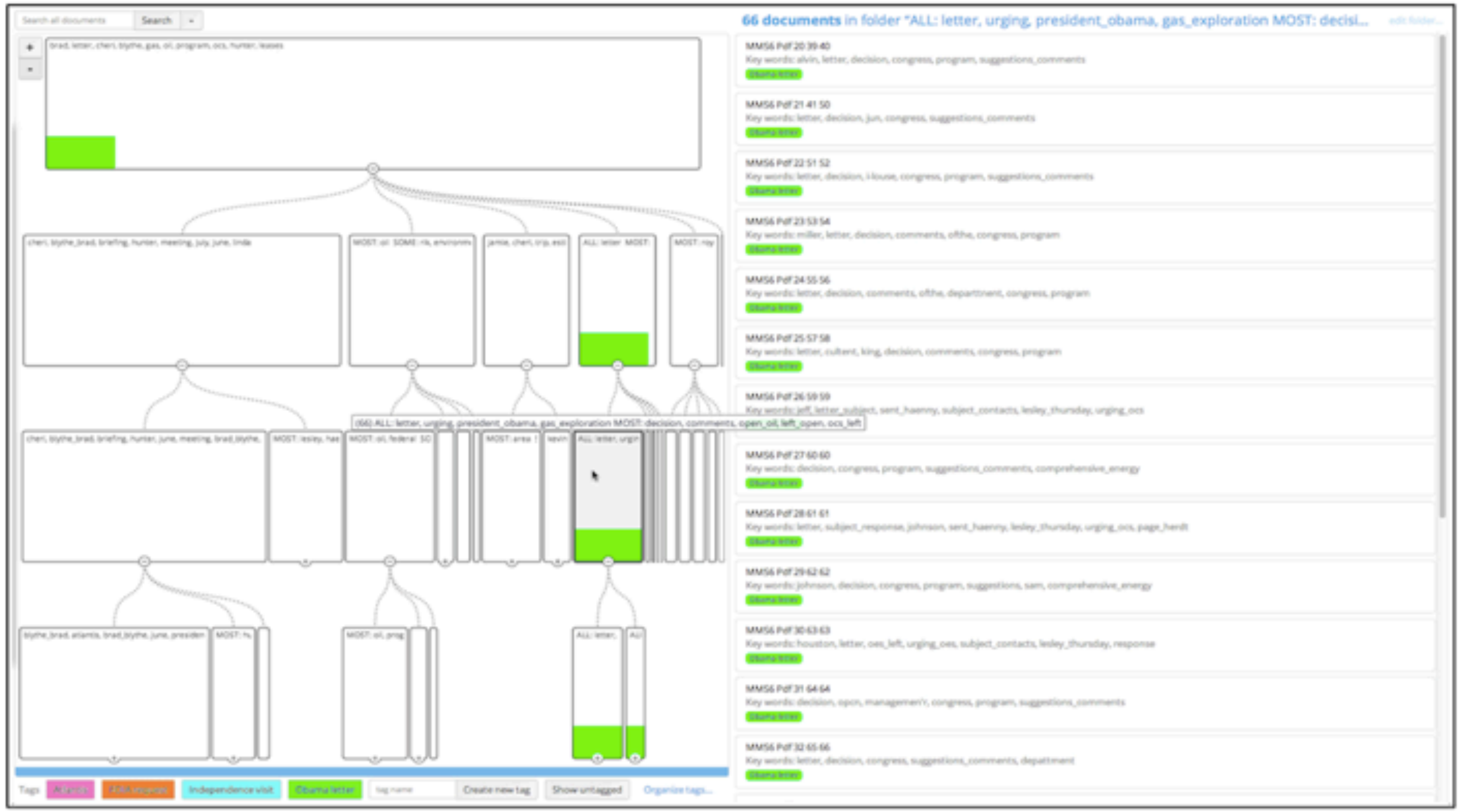


Path to adoption

- even more rounds of what/why/how interplay
  - which views needed? what should they show? how should they show it?
  - usability and utility
- version 3
  - published story: VP candidate Ryan asked for federal help even as championed cuts
  - published story: gun control debate
- version 4
  - followup investigation: government corruption in Texas
  - published story: police misconduct in New York *(Pulitzer prize finalist!)*



Overview video v4


Search all documents
Search
brad, letter, cheri, blythe, gas, oil, program, ocs, hunter, leases
cheri, blythe_brad, briefing, hunter, meeting, july, june, linda
MOST: oil SOME: rik, environm
jamie, cheri, trip, esti
ALL: letter MOST:
MOST: roy
cheri, blythe_brad, briefing, hunter, june, meeting, brad_blythe,
MOST: lesley, hae
MOST: oil, federal SO
MOST: area
kevin
ALL: letter, urgin
(66) ALL: letter, urging, president_obama, gas_exploration MOST: decision, comments, open_oil, left_open, ocs_left
blythe_brad, atlantis, brad_blythe, june, presiden
MOST: ro
MOST: oil, prog
ALL: letter,
ALI
Tags
Atlantis
FOIA request
Independence visit
Obama letter
tag name
Create new tag
Show untagged
Organize tags...

66 documents in folder "ALL: letter, urging, president_obama, gas_exploration MOST: decisi...

MMS6 Pdf 20 39 40
Key words: alvin, letter, decision, congress, program, suggestions_comments
Obama letter

MMS6 Pdf 21 41 50
Key words: letter, decision, jun, congress, suggestions_comments
Obama letter

MMS6 Pdf 22 51 52
Key words: letter, decision, i-louse, congress, program, suggestions_comments
Obama letter

MMS6 Pdf 23 53 54
Key words: miller, letter, decision, comments, ofthe, congress, program
Obama letter

MMS6 Pdf 24 55 56
Key words: letter, decision, comments, ofthe, departtnent, congress, program
Obama letter

MMS6 Pdf 25 57 58
Key words: letter, cultent, king, decision, comments, congress, program
Obama letter

MMS6 Pdf 26 59 59
Key words: jeff, letter_subject, sent_haenny, subject_contacts, lesley_thursday, urging_ocs

MMS6 Pdf 27 60 60
Key words: decision, congress, program, suggestions_comments, comprehensive_energy
Obama letter

MMS6 Pdf 28 61 61
Key words: letter, subject_response, johnson, sent_haenny, lesley_thursday, urging_ocs, page_herdt
Obama letter

MMS6 Pdf 29 62 62
Key words: johnson, decision, congress, program, suggestions, sam, comprehensive_energy
Obama letter

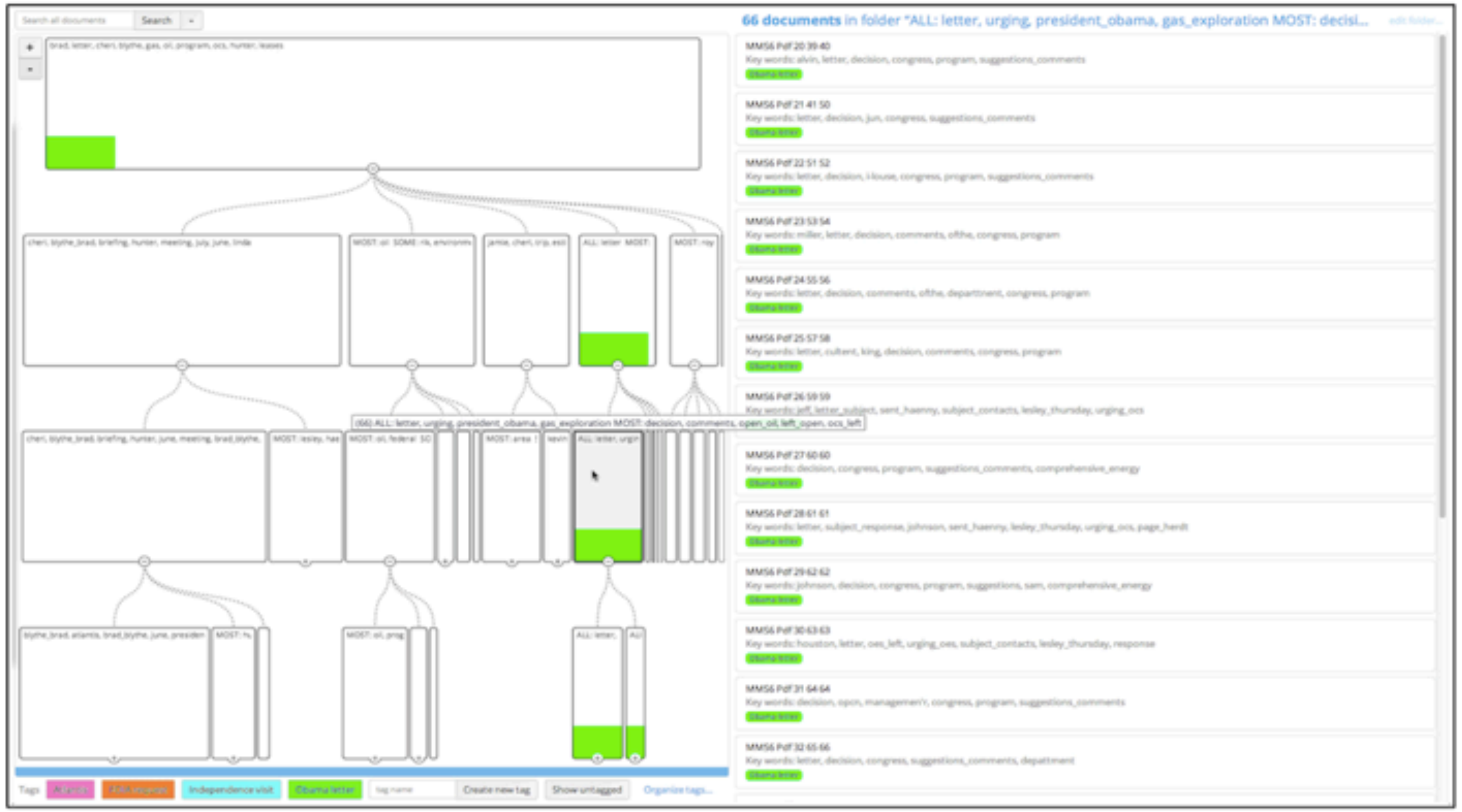
MMS6 Pdf 30 63 63
Key words: houston, letter, oes_left, urging_oes, subject_contacts, lesley_thursday, response
Obama letter

MMS6 Pdf 31 64 64
Key words: decision, opcn, managemen'r, congress, program, suggestions_comments
Obama letter

MMS6 Pdf 32 65 66
Key words: letter, decision, congress, suggestions_comments, depattment
Obama letter

http://vimeo.com/71483614

Overview video v4



Search all documents
Search
+
-
brad, letter, cheri, blythe, gas, oil, program, ocs, hunter, leases
cheri, blythe_brad, briefing, hunter, meeting, july, june, linda
MOST: oil SOME: rls, environm
jamie, cheri, trip, esti
ALL: letter MOST:
MOST: roy
(66) ALL: letter, urging, president_obama, gas_exploration MOST: decision, comments, open_oil, left_open, ocs_left
cheri, blythe_brad, briefing, hunter, june, meeting, brad_blythe,
MOST: lesley, hae
MOST: oil, federal SO
MOST: area
kevin
ALL: letter, urgin
blythe_brad, atlantis, brad_blythe, june, presiden
MOST: hu
MOST: oil, prog
ALL: letter,
ALL
Tags
Atlantis
FOIA request
Independence visit
Obama letter
tag name
Create new tag
Show untagged
Organize tags...
edit folder...

66 documents in folder "ALL: letter, urging, president_obama, gas_exploration MOST: decisi...

MMS6 Pdf 20 39 40
Key words: alvin, letter, decision, congress, program, suggestions_comments
Obama letter

MMS6 Pdf 21 41 50
Key words: letter, decision, jun, congress, suggestions_comments
Obama letter

MMS6 Pdf 22 51 52
Key words: letter, decision, i-louse, congress, program, suggestions_comments
Obama letter

MMS6 Pdf 23 53 54
Key words: miller, letter, decision, comments, ofthe, congress, program
Obama letter

MMS6 Pdf 24 55 56
Key words: letter, decision, comments, ofthe, departtnent, congress, program
Obama letter

MMS6 Pdf 25 57 58
Key words: letter, cultent, king, decision, comments, congress, program
Obama letter

MMS6 Pdf 26 59 59
Key words: jeff, letter_subject, sent_haenny, subject_contacts, lesley_thursday, urging_ocs

MMS6 Pdf 27 60 60
Key words: decision, congress, program, suggestions_comments, comprehensive_energy
Obama letter

MMS6 Pdf 28 61 61
Key words: letter, subject_response, johnson, sent_haenny, lesley_thursday, urging_ocs, page_herdt
Obama letter

MMS6 Pdf 29 62 62
Key words: johnson, decision, congress, program, suggestions, sam, comprehensive_energy
Obama letter

MMS6 Pdf 30 63 63
Key words: houston, letter, oes_left, urging_oes, subject_contacts, lesley_thursday, response
Obama letter

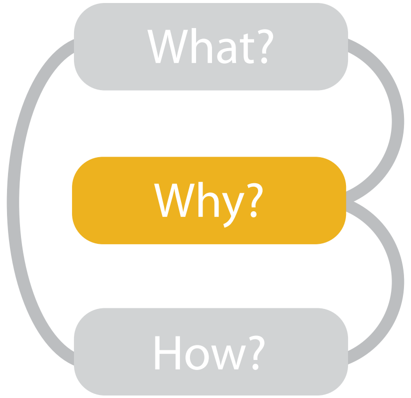
MMS6 Pdf 31 64 64
Key words: decision, opcn, managemen'r, congress, program, suggestions_comments
Obama letter

MMS6 Pdf 32 65 66
Key words: letter, decision, congress, suggestions_comments, depattment
Obama letter

http://vimeo.com/71483614

- versions 3 and 4
  - no DR scatterplot
  - tree arrangement emphasizing nodes not links
  - combined doc/cluster viewer

Why: Task abstractions

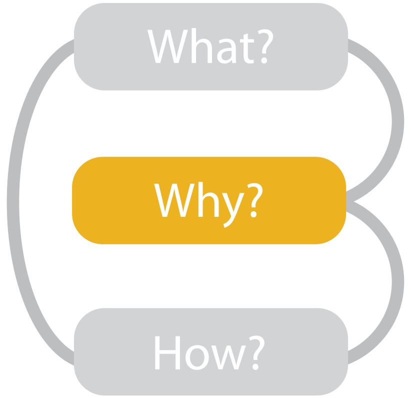


What?

Why?

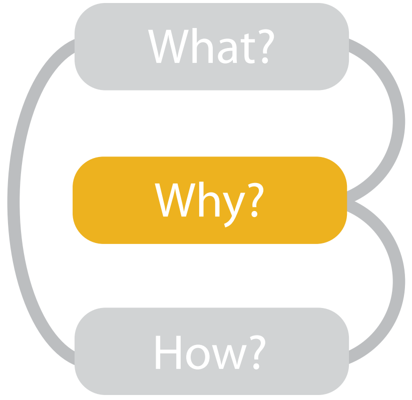
How?

Why: Task abstractions



- what's in this collection? (of leaked docs)

Why: Task abstractions



- what's in this collection? (of leaked docs)
  - *generate* hypothesis

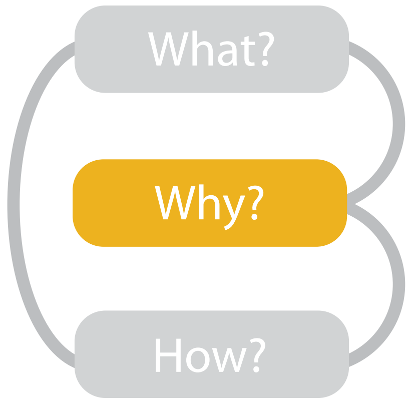
Why: Task abstractions

- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters

Why: Task abstractions

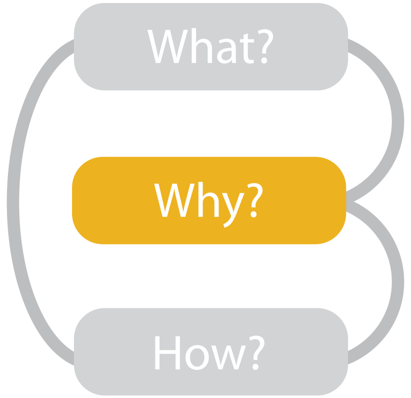
- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters

Why: Task abstractions



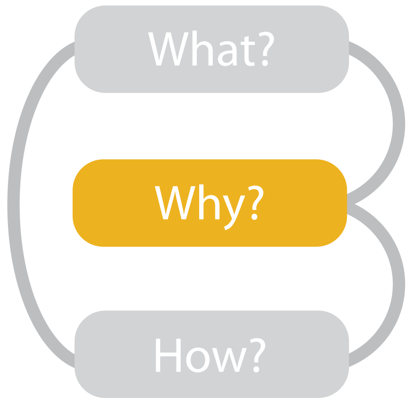
- what’s in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)

Why: Task abstractions



- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis

Why: Task abstractions



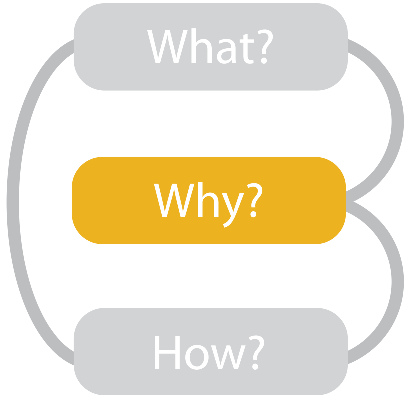
What?

Why?

How?

- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents

Why: Task abstractions

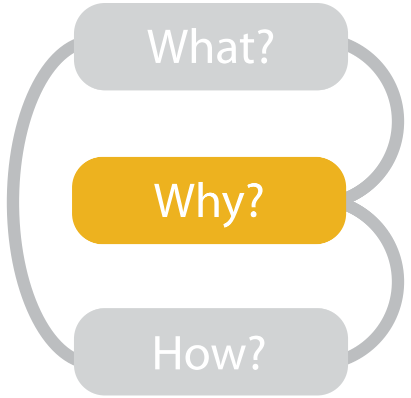

What?

Why?

How?

- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents
  - *locate* clusters/documents

Why: Task abstractions



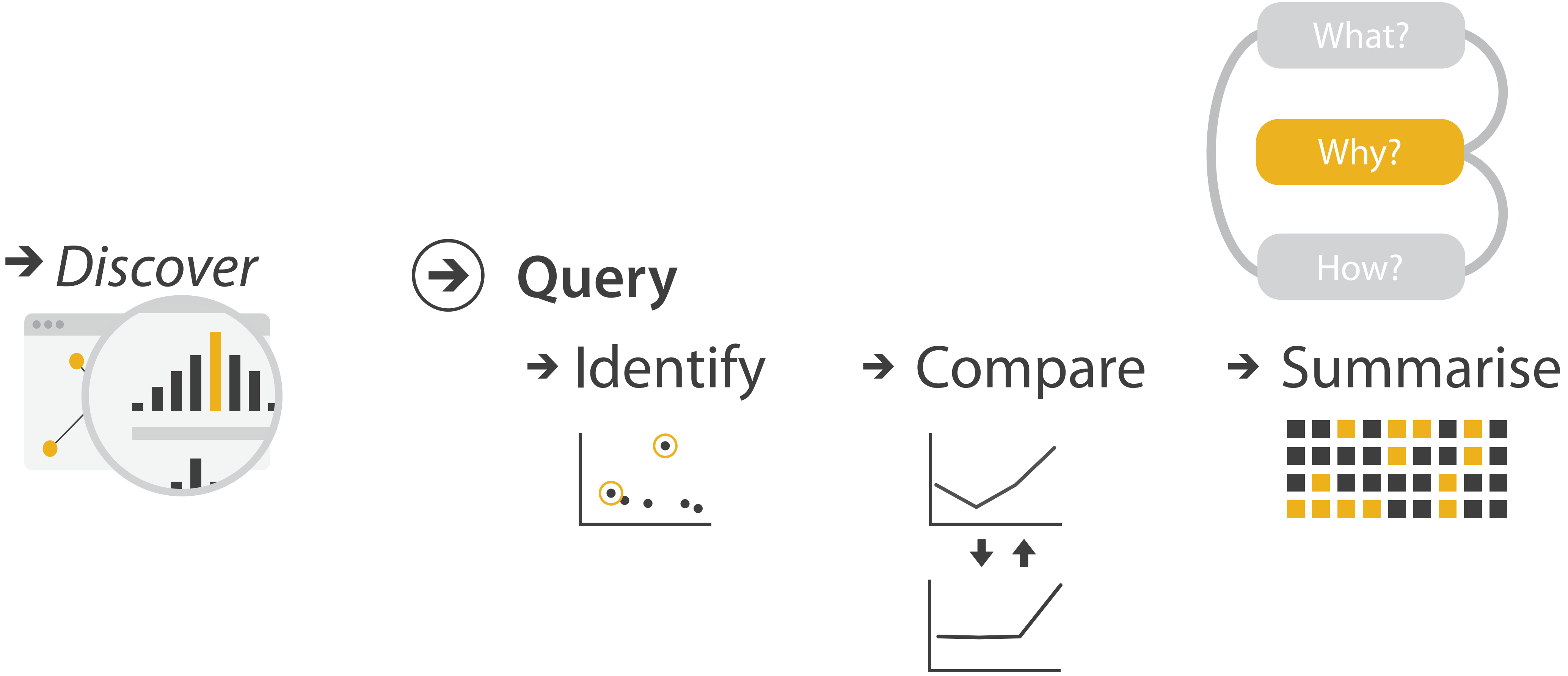
- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters

➔ *Discover*

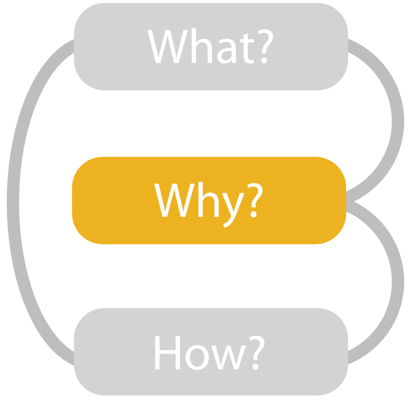
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents
  - *locate* clusters/documents

Why: Task abstractions

- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents
  - *locate* clusters/documents

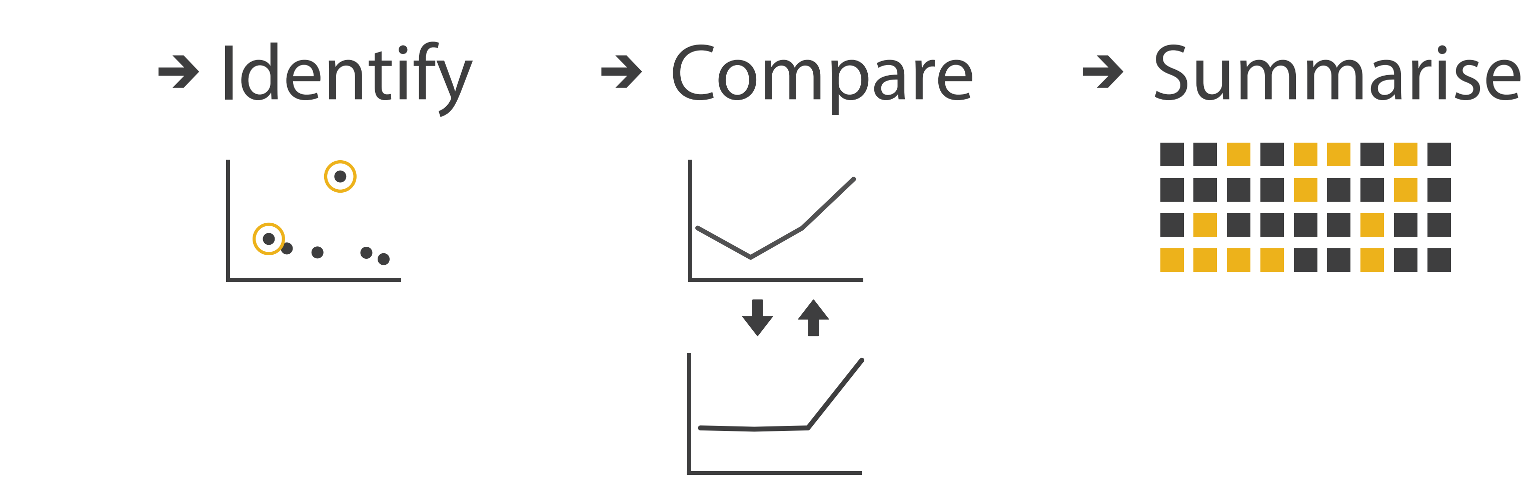


Why: Task abstractions



- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents
  - *locate* clusters/documents

Query

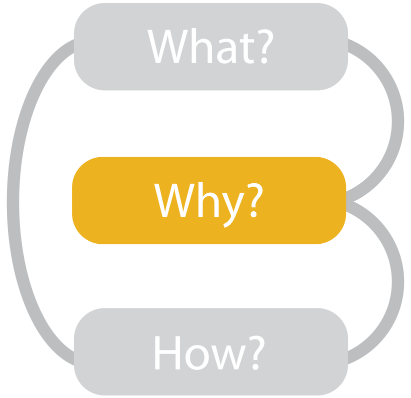


Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

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

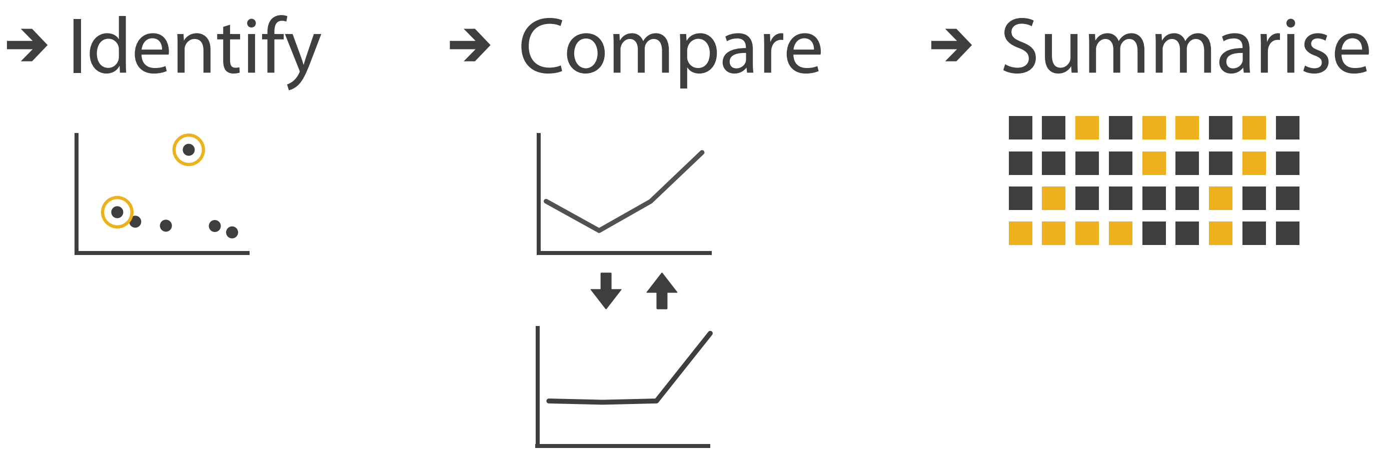
Why: Task abstractions



- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents
  - *locate* clusters/documents
- prove non-existence of evidence

→ *Discover*

→ **Query**

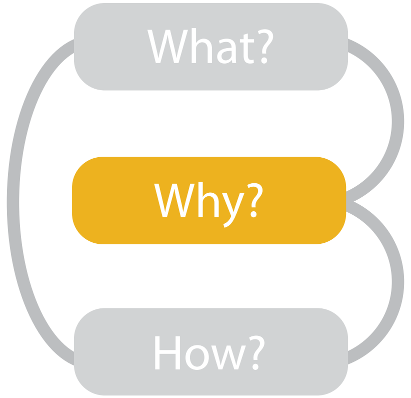


→ **Search**

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

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

Why: Task abstractions

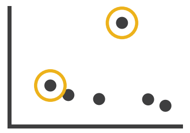


- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents
  - *locate* clusters/documents
- prove non-existence of evidence
  - even harder!

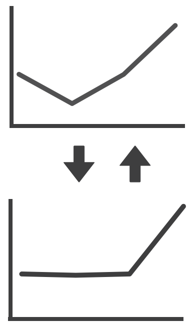
→ *Discover*

➔ **Query**

→ Identify



→ Compare



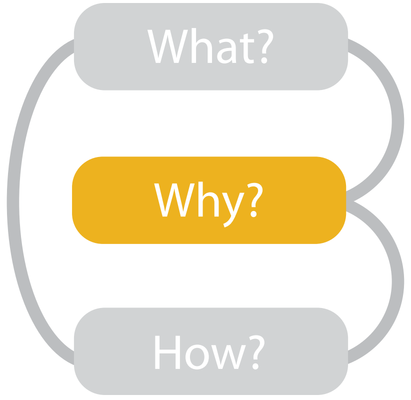
→ Summarise

➔ **Search**

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

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

Why: Task abstractions


What?

Why?

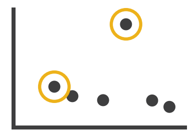
How?

- what's in this collection? (of leaked docs)
  - *generate* hypothesis
  - *summarize* clusters
  - *explore* clusters
- locate evidence (within FOIA dump)
  - *verify* hypothesis
  - *identify* clusters/documents
  - *locate* clusters/documents
- prove non-existence of evidence
  - even harder!
  - exhaustive reading vs filtering out irrelevant

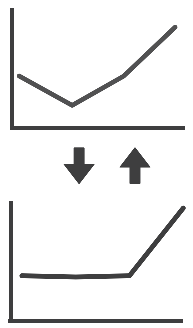
→ *Discover*

→ **Query**

→ Identify



→ Compare



→ Summarise

→ **Search**

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

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

Now what?

Knight Foundation AP ASSOCIATED PRESS

https://www.overviewproject.org/

http://overview.ap.org/

- continuing adoption
  - food stamp distribution delays in North Carolina
  - Surprise! Many credit card agreements allow repossession
  - The brilliance of Louis C.K.'s emails: He writes like a politician
  - Private memo reveals winding tale involving John McCain, the NRA, and... condors
- continuing development
  - Knight Foundation funds v5: named entity recognition, plugin API

- InfoVis14 paper

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists. *Brehmer, Ingram, Stray, and, Munzner.*

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

Algorithm: Spinoff series

- dimensionality reduction for huge text collections
  - great algorithm problem in its own right!
  - QSNE: fast and high-quality DR for millions of documents
    - key feature: handle sparseness appropriately

[Dimensionality Reduction for Documents with Nearest Neighbor Queries. Ingram and Munzner. Neurocomputing (Special Issue on Visual Analytics using Multidimensional Projections), to appear 2014.]

http://www.cs.ubc.ca/labs/imager/tr/2014/QSNE/