Week 5: Manipulate, Facet, Reduce Demo: Text

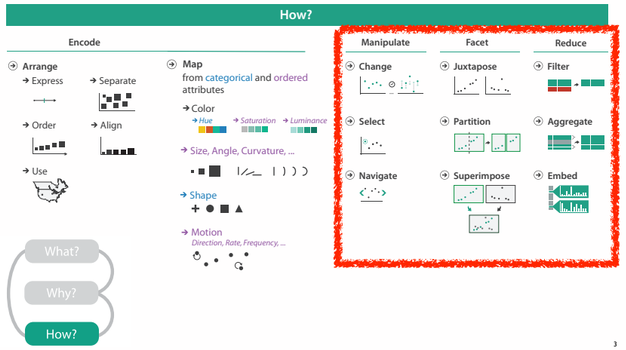
Tamara Munzner
Department of Computer Science
University of British Columbia

JRNL 520M, Special Topics in Contemporary Journalism: Visualization for Journalists
Week 5: 13 October 2015

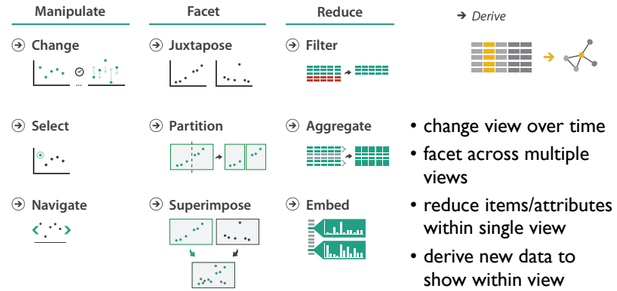
http://www.cs.ubc.ca/~tmm/courses/journ15

Now

- Manipulate
- Facet (not covered last week)
- Reduce
- Demos/Videos
  - LineUp
  - LiveRAC
  - Cerebral
- Demos: Text
  - Overview
  - TimeLineCurator

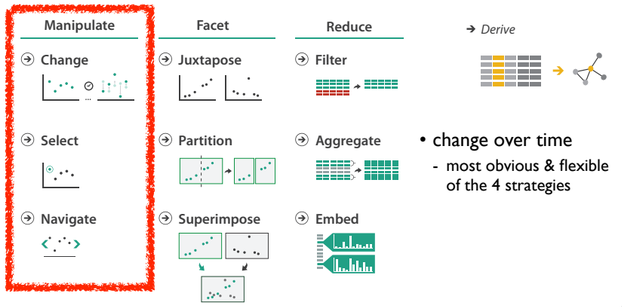


How to handle complexity: 3 more strategies + 1 previous



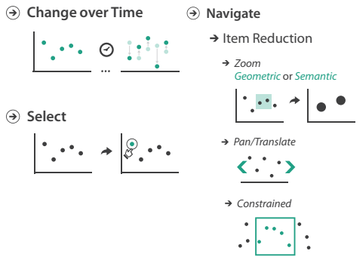
- change view over time
- facet across multiple views
- reduce items/attributes within single view
- derive new data to show within view

How to handle complexity: 3 more strategies + 1 previous



- change over time
  - most obvious & flexible of the 4 strategies

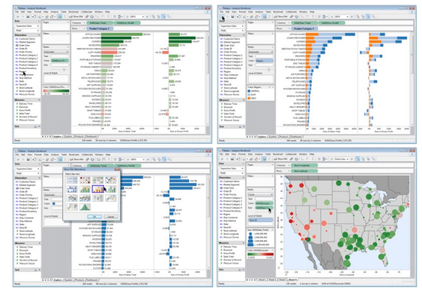
VAD Ch 11: Manipulate



Change over time

- change any of the other choices
  - encoding itself
  - parameters
  - arrange: rearrange, reorder
  - aggregation level, what is filtered...
  - interaction entails change

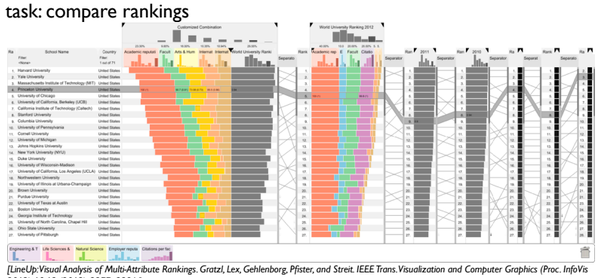
Idiom: **Re-encode** System: **Tableau**



made using Tableau, http://tableausoftware.com

Idiom: **Reorder** System: **LineUp**

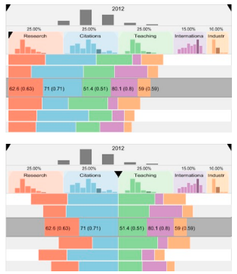
- data: tables with many attributes
- task: compare rankings



[LineUp: Visual Analysis of Multi-Attribute Rankings. Gratzl, Lex, Gehlenborg, Pfister, and Streit. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2013) 19:12 (2013), 2277–2286.]

Idiom: **Realign** System: **LineUp**

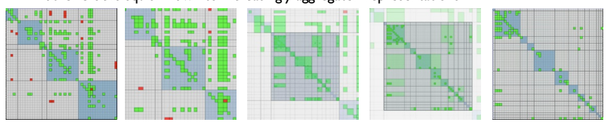
- stacked bars
  - easy to compare
    - first segment
    - total bar
- align to different segment
  - supports flexible comparison



[LineUp: Visual Analysis of Multi-Attribute Rankings. Gratzl, Lex, Gehlenborg, Pfister, and Streit. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2013) 19:12 (2013), 2277–2286.]

Idiom: **Animated transitions**

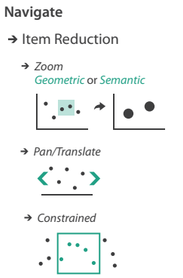
- smooth transition from one state to another
  - alternative to jump cuts
  - support for item tracking when amount of change is limited
- example: multilevel matrix views
  - scope of what is shown narrows down
    - middle block stretches to fill space, additional structure appears within
    - other blocks squish down to increasingly aggregated representations



[Using Multilevel Call Matrices in Large Software Projects. van Ham. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 227–232, 2003.]

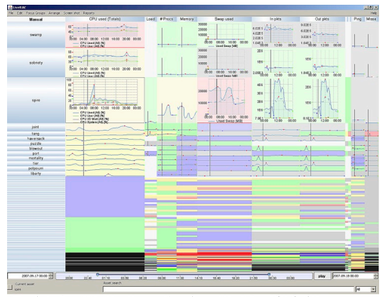
Navigate: Changing item visibility

- change viewpoint
  - changes which items are visible within view
  - camera metaphor
    - zoom
      - geometric zoom: familiar semantics
      - semantic zoom: adapt object representation based on available pixels
        - » dramatic change, or more subtle one
    - pan/translate
    - rotate
      - especially in 3D
  - constrained navigation
    - often with animated transitions
    - often based on selection set



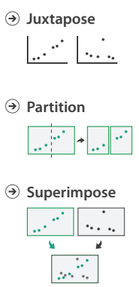
Idiom: **Semantic zooming** System: **LiveRAC**

- visual encoding change
  - colored box
  - sparkline
  - simple line chart
  - full chart: axes and tickmarks

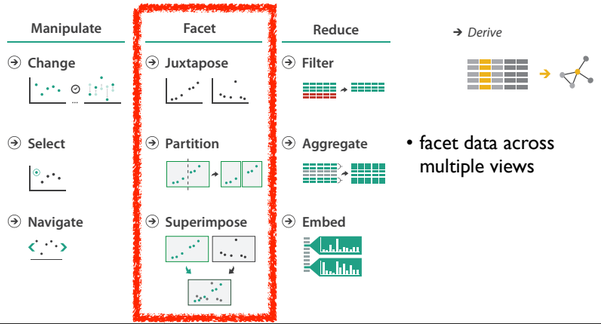


[LiveRAC - Interactive Visual Exploration of System Management Time-Series Data. McLachlan, Munzner, Kautsofios, and North. Proc. ACM Conf. Human Factors in Computing Systems (CHI), pp. 1483–1492, 2008.]

VAD Chap 11: Facet Into Multiple Views

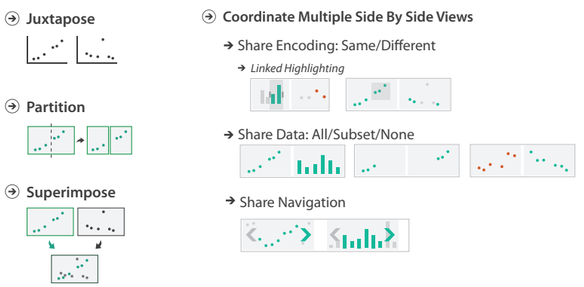


How to handle complexity: 3 more strategies + 1 previous



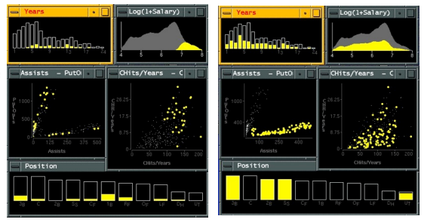
- facet data across multiple views

Facet



Idiom: **Linked highlighting** System: **EDV**

- see how regions contiguous in one view are distributed within another
  - powerful and pervasive interaction idiom
- encoding: different
  - ***multiform***
- data: all shared



[Visual Exploration of Large Structured Datasets. Wills. Proc. New Techniques and Trends in Statistics (NTTS), pp. 237–246. IOS Press, 1995.]

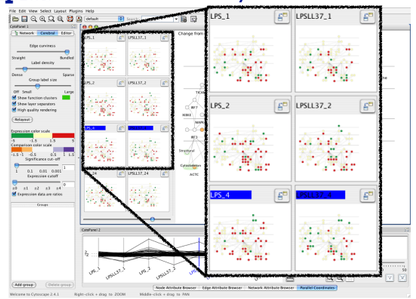
Idiom: **bird's-eye maps** System: **Google Maps**

- encoding: same
- data: subset shared
- navigation: shared
  - bidirectional linking
- differences
  - viewpoint
  - (size)
- ***overview-detail***

[A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.]

Idiom: **Small multiples** System: **Cerebral**

- encoding: same
- data: none shared
  - different attributes for node colors
  - (same network layout)
- navigation: shared



[Cerebral:Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14:6 (2008), 1253–1260.]

Coordinate views: Design choice interaction

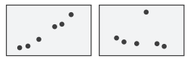
Encoding	Data: All	Data: Subset	Data: None
Same	Redundant	Overview/ Detail	Small Multiples
Different	Multiform	Multiform, Overview/ Detail	No Linkage

- why juxtapose views?
  - benefits: eyes vs memory
    - lower cognitive load to move eyes between 2 views than remembering previous state with single changing view
  - costs: display area, 2 views side by side each have only half the area of one view

Partition into views

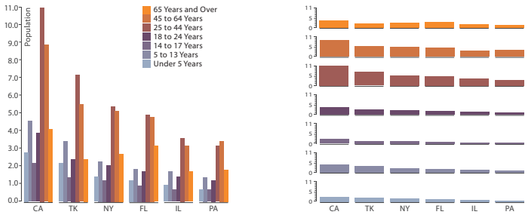
- how to divide data between views
  - encodes association between items using spatial proximity
  - major implications for what patterns are visible
  - split according to attributes
- design choices
  - how many splits
    - all the way down: one mark per region?
    - stop earlier, for more complex structure within region?
  - order in which attribs used to split
  - how many views

Partition into Side-by-Side Views



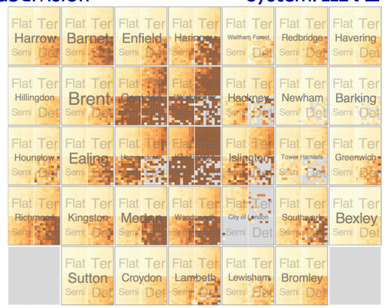
Partitioning: List alignment

- single bar chart with grouped bars
  - split by state into regions
    - complex glyph within each region showing all ages
  - compare: easy within state, hard across ages
- small-multiple bar charts
  - split by age into regions
    - one chart per region
  - compare: easy within age, harder across states



Partitioning: Recursive subdivision System: **HIVE**

- split by neighborhood
- then by type
- then time
  - years as rows
  - months as columns
- color by price
- neighborhood patterns
  - where it's expensive
  - where you pay much more for detached type



[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

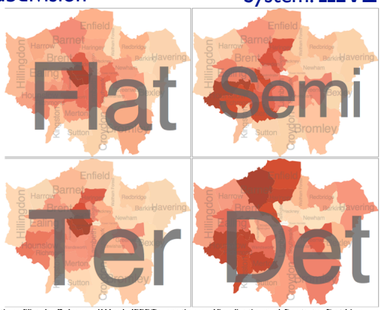
Partitioning: Recursive subdivision System: **HIVE**

- switch order of splits
  - type then neighborhood
- switch color
  - by price variation
- type patterns
  - within specific type, which neighborhoods inconsistent

[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

Partitioning: Recursive subdivision System: **HIVE**

- different encoding for second-level regions
  - choropleth maps



[Configuring Hierarchical Layouts to Address Research Questions. Slingsby, Dykes, and Wood. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009), 977–984.]

Superimpose layers

- ***layer***: set of objects spread out over region
  - each set is visually distinguishable group
  - extent: whole view
- design choices
  - how many layers?
  - how are layers distinguished?
  - small static set or dynamic from many possible?
  - how partitioned?
    - heavyweight with attribs vs lightweight with selection
- distinguishable layers
  - encode with different, nonoverlapping channels
    - two layers achieveable, three with careful design

Superimpose Layers

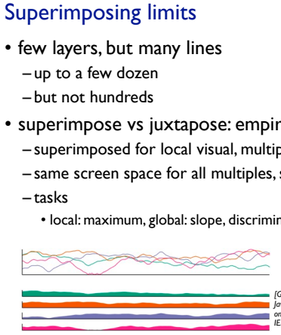
Static visual layering

- foreground layer: roads
  - hue, size distinguishing main from minor
  - high luminance contrast from background
- background layer: regions
  - desaturated colors for water, parks, land areas
- user can selectively focus attention
- "get it right in black and white"
  - check luminance contrast with greyscale view

[Get it right in black and white. Stone. 2010. http://www.stonesc.com/wordpress/2010/03/get-it-right-in-black-and-white]

Superimposing limits

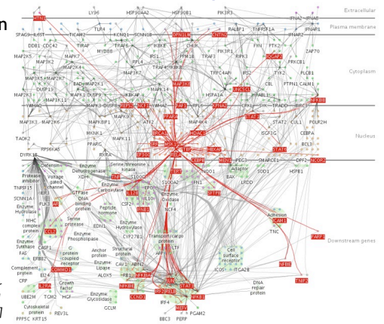
- few layers, but many lines
  - up to a few dozen
  - but not hundreds
- superimpose vs juxtapose: empirical study
  - superimposed for local visual, multiple for global
  - same screen space for all multiples, single superimposed
  - tasks
    - local: maximum, global: slope, discrimination



[Graphical Perception of Multiple Time Series. Javed, McDonnel, and Elmqvist. IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE InfoVis 2010) 16:6 (2010), 927–934.]

Dynamic visual layering System: **Cerebral**

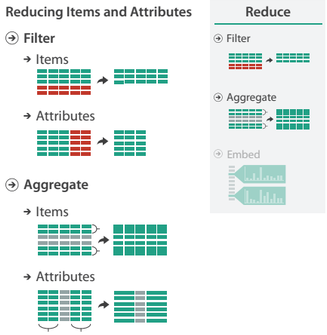
- interactive, from selection
  - lightweight: click
  - very lightweight: hover
- ex: 1-hop neighbors



[Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Barsky, Gardy, Hancock, and Munzner. Bioinformatics 23:8 (2007), 1040–1042.]

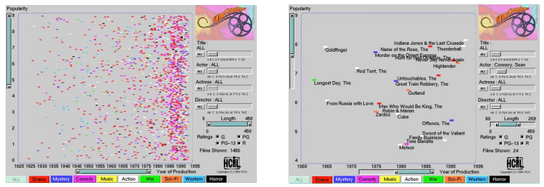
Reduce items and attributes

- reduce/increase: inverses
- filter
  - pro: straightforward and intuitive
    - to understand and compute
  - con: out of sight, out of mind
- aggregation
  - pro: inform about whole set
  - con: difficult to avoid losing signal
- not mutually exclusive
  - combine filter, aggregate
  - combine reduce, change, facet



Idiom: **dynamic filtering** System: **FilmFinder**

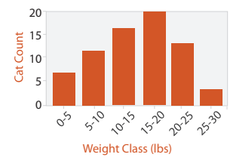
- item filtering
- browse through tightly coupled interaction
  - alternative to queries that might return far too many or too few



[Visual information seeking:Tight coupling of dynamic query filters with starfield displays. Ahlberg and Shneiderman. Proc. ACM Conf. on Human Factors in Computing Systems (CHI), pp. 313–317, 1994.]

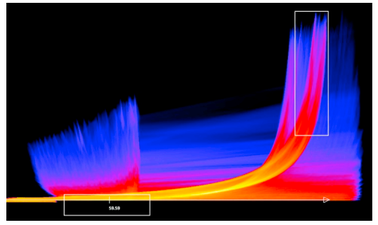
Idiom: **histogram**

- static item aggregation
- task: find distribution
- data: table
- derived data
  - new table: keys are bins, values are counts
- bin size crucial
  - pattern can change dramatically depending on discretization
  - opportunity for interaction: control bin size on the fly



Continuous scatterplot

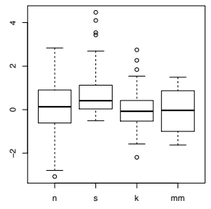
- static item aggregation
- data: table
- derived data: table
  - key attribs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color: sequential categorical hue + ordered luminance colormap



[Continuous Scatterplots. Bachthaler and Weiskopf. IEEE TVCG (Proc. Vis 08) 14:6 (2008), 1428–1435. 2008.]

Idiom: **boxplot**

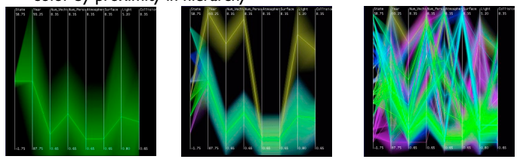
- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown



[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

Idiom: **Hierarchical parallel coordinates**

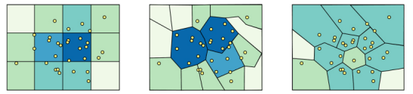
- dynamic item aggregation
- derived data: ***hierarchical clustering***
- encoding:
  - cluster band with variable transparency, line at mean, width by min/max values
  - color by proximity in hierarchy



[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua, Ward, and Rundensteiner. Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

Spatial aggregation

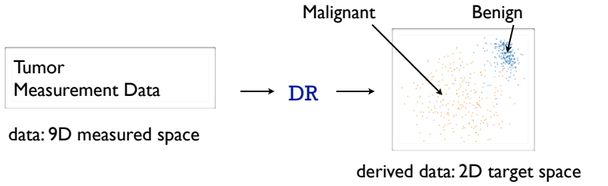
- MAUP: Modifiable Areal Unit Problem
  - gerrymandering (manipulating voting district boundaries) is one example!



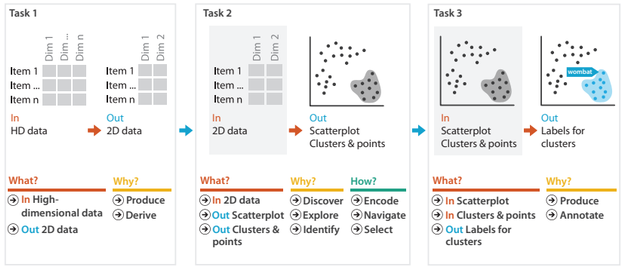
[http://www.e-education.psu.edu/geog486/l4_p7.html, Fig 4.cg.6]

Dimensionality reduction

- attribute aggregation
  - derive low-dimensional target space from high-dimensional measured space
  - use when you can't directly measure what you care about
    - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
    - latent factors, hidden variables



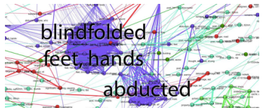
Dimensionality reduction for documents



- bag of words model for text document

Overview origin story: WikiLeaks meets Glimmer

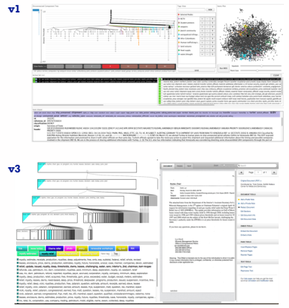
- WikiLeaks: hacker-journalist Jonathan Stray analyzing Iraq warlogs
  - conjecture that existing label classification falls short of showing all meaningful structure in data
    - friendly action, criminal incident, ...
  - had some NLP, needed better vis tools



- 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.]

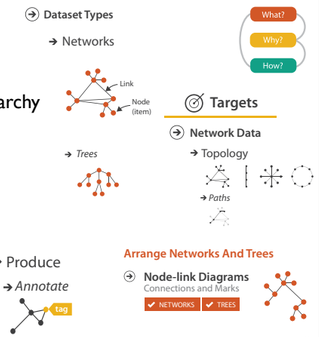
Overview design evolution



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

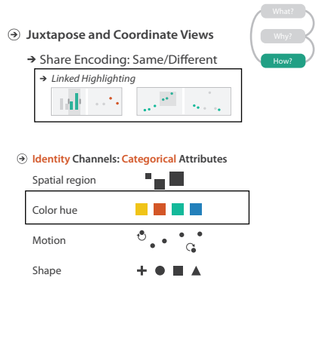
What/Why/How interplay

- 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



How: Idiom design decisions

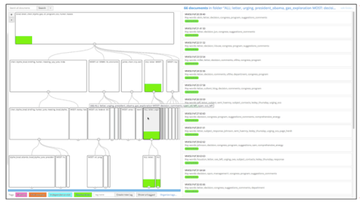
- facet: juxtapose linked views
  - linked color coding
    - cluster hierarchy tree
    - DR scatterplot
    - tags
  - reading text/keywords
    - cluster list
    - doc reader



Overview video (version 1)

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

Overview video v4

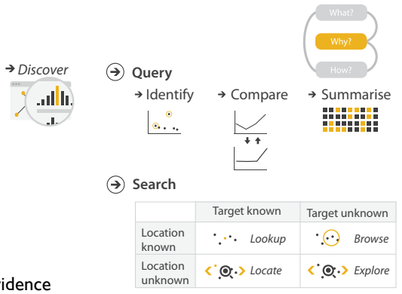


http://vimeo.com/71483614

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

Why: Task abstractions

- what'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



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

Demo

https://www.overviewdocs.com/

http://overview.ap.org/

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

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

Further reading

- Visualization Analysis and Design. Tamara Munzner. CRC Press, 2014.
  - *Chap 11: Manipulate View*
  - *Chap 12: Facet Across Multiple Views*
  - *Chap 13: Reduce Items and Attributes*

Lab/Assignment 5

- Use TimeLineCurator to create visual timelines from free-form text
  - work through BC History example
  - find 1 article where temporal story is worth telling, and curate it for TimelineJS export
    - including media/images is optional
  - find 2 articles that make sense to compare with each other in a mashup
    - curate a combined timeline for TLC export
  - find 1 article where there's nothing interesting to see
    - document that it's uninteresting with screenshot of TLC's initial screen