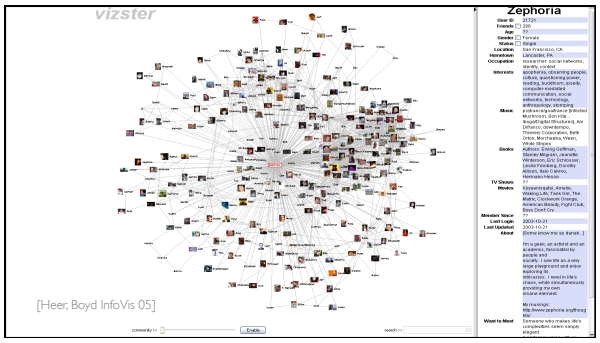
Social Data Analysis

Wesley Willett
CS294-10 Visualization

[with some slides from Jeffrey Heer Stanford University]

A Tale of Two Visualizations



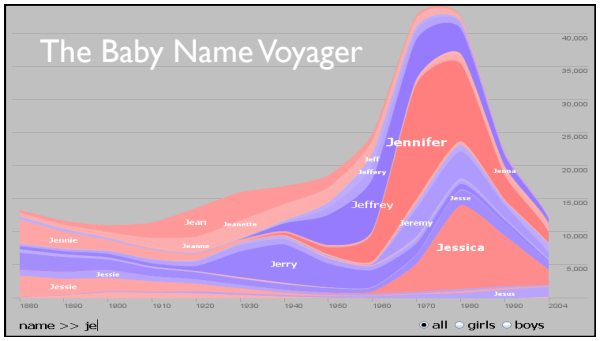
[Heer, Boyd InfoVis 05]

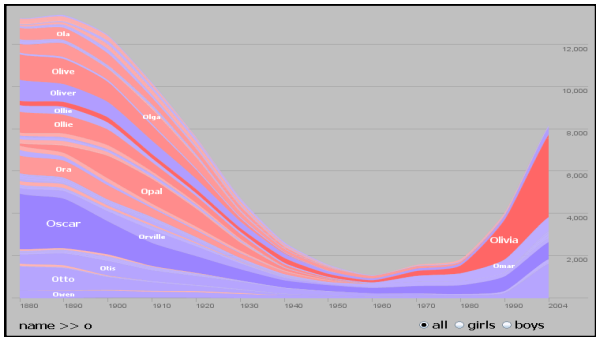
Observations

Groups spent more time in front of the visualization than individuals.

Friends encouraged each other to unearth relationships, probe community boundaries, and challenge reported information.

Social play resulted in informal analysis, often driven by story-telling of group histories.





Social Data Analysis

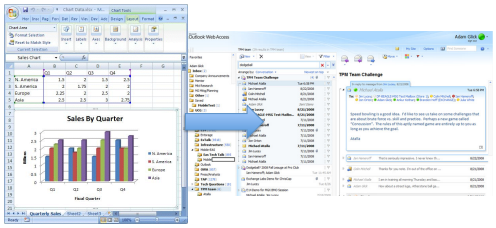
Visual sensemaking can be **social** as well as cognitive.

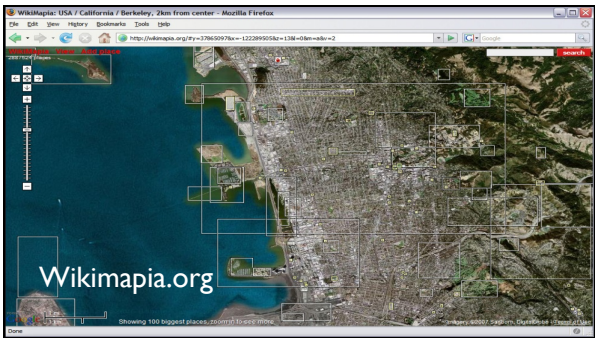
How can user interfaces catalyze and support **collaborative visual analysis**?

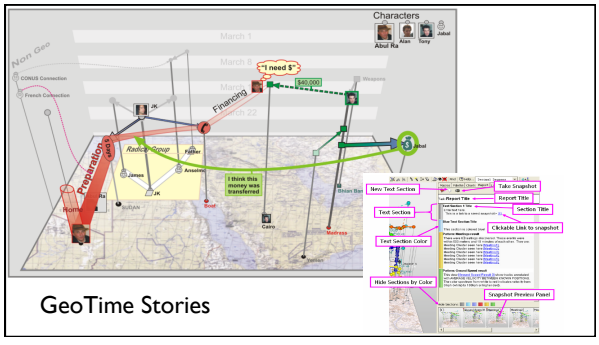
Inspired to design and evaluate both **real systems** and **targeted techniques**.

Social Data Analysis

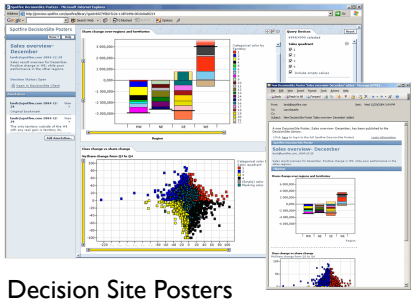
Excel + Email



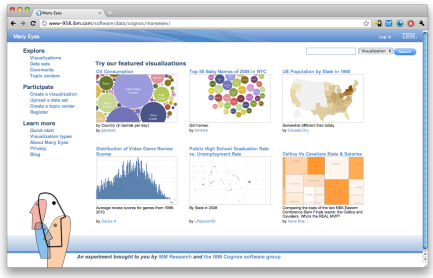




GeoTime Stories

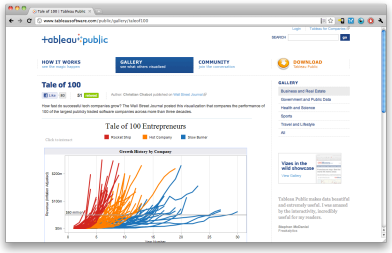


Spotfire Decision Site Posters



Many-Eyes

Tableau Public

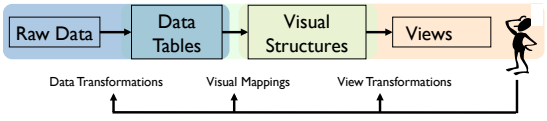


Where can Collaborators Contribute?

Data Management
Contribute Data
Clean Data
Categorize Data
Moderate Data
Create Metadata

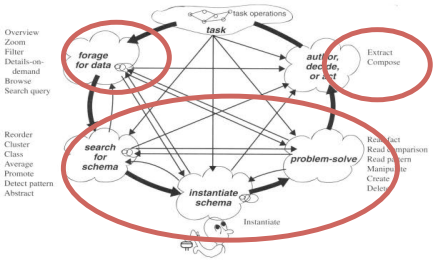
Visualization
Select Data Sources
Apply Visual Encoding
Author Software

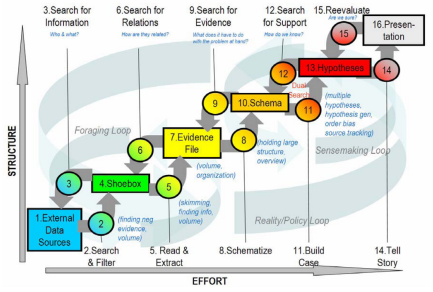
Visual Analytics
Observations
Hypotheses
Evidence (+/-)
Summarize
Report / Presentation



Sensemaking

[Card, Mackinlay, & Schneiderman]



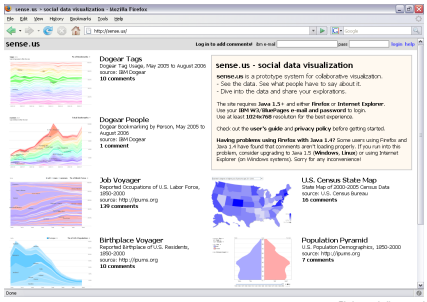


[Pirolli & Card, 1999]

Design Considerations [Heer & Agrawala VAST 07, IVS 08]

Division, allocation, and integration of work
Common ground and awareness
Reference and deixis (pointing)
Identity, trust, and reputation
Group formation and management
Incentives and engagement
Presentation and decision-making

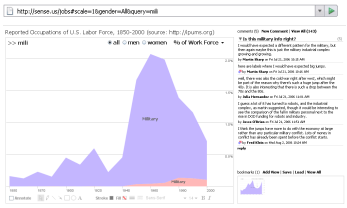
Sense.us: Collaborative Visualization of Demographic Data



[Heer, Viégas, Wattenberg CHI 07]

Exploratory Design Rationale

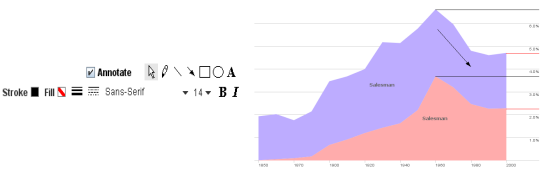
Sharing within visualization and across the web



Exploratory Design Rationale

Sharing within visualization and across the web

Pointing at interesting trends, outliers



Exploratory Design Rationale

Sharing within visualization and across the web

Pointing at interesting trends, outliers

Collecting and linking related views

Exploratory Design Rationale

Sharing within visualization and across the web

Pointing at interesting trends, outliers

Collecting and linking related views

Awareness of social activity

Exploratory Design Rationale

Sharing within visualization and across the web

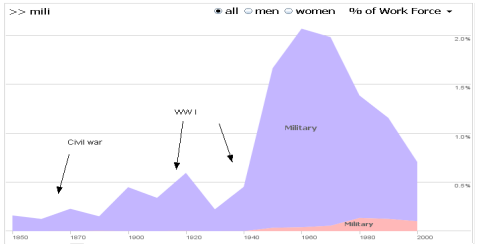
Pointing at interesting trends, outliers

Collecting and linking related views

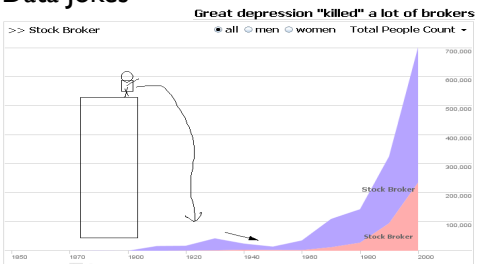
Awareness of social activity

Don't disrupt individual exploration

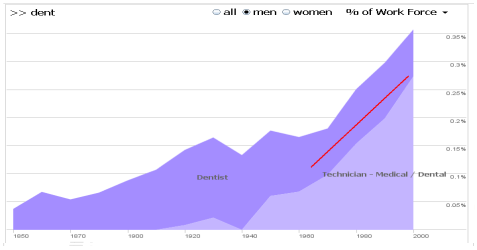
Building Off of Others



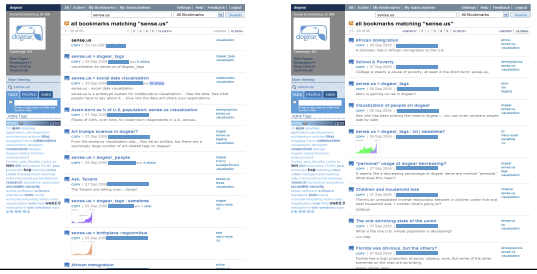
Data Jokes



Collaborative Sensemaking



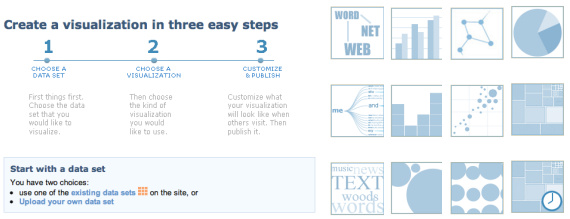
Sharing in External Media



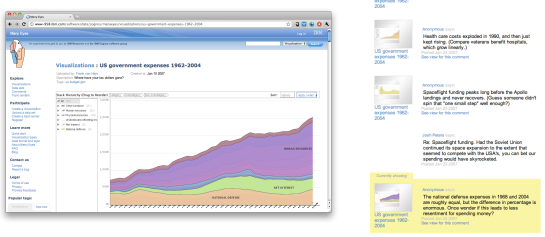
Many-Eyes: Social Data Analysis at an Internet Scale

[Viégas, Wattenberg, et al. CHI 07]

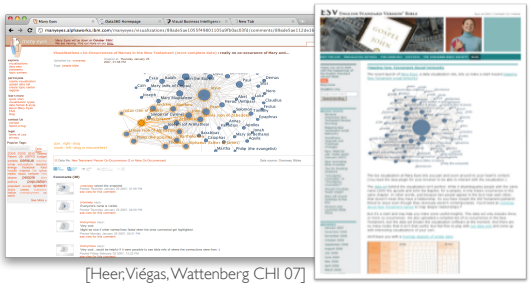
Templates for Visualizing Data



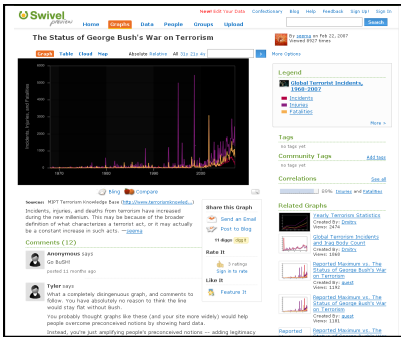
Visualizations & Commenting



Outside Communities



[Heer, Viégas, Wattenberg CHI 07]



Swivel.com, and others…

A graveyard of "YouTubes for Data"

verifiable.com

swivel

Track-n-Graph™

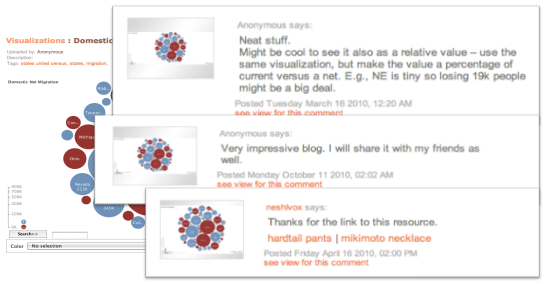
icharts

Chartle.net
simple & interactive charts online

Web 2.0 ready!

EARLY BETA

Emergent analysis and discussion isn't very good



Many Eyes – circa 10/2010

144,480 Data Sets

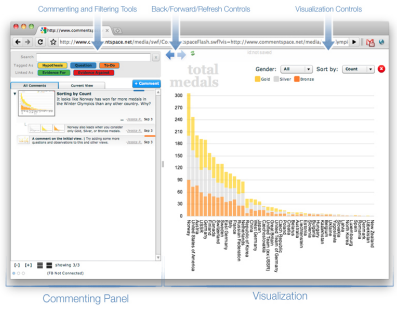
71,454 Visualizations

11,818 Comments

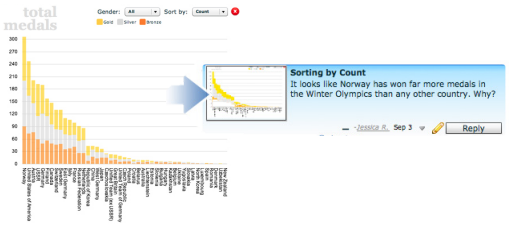
~**90%** spam and insubstantial comments

CommentSpace: Structured Support for Social Data Analysis

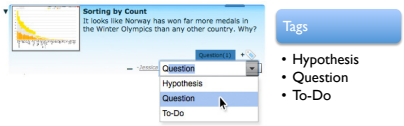
Can we augment social data analysis to support deeper **analysis** and **synthesis**?



Tight coupling of comments and views



Organizing comments using tags and links



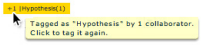
Tags

- Hypothesis
- Question
- To-Do

Links

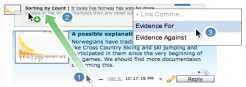
- Evidence For
- Evidence Against
- (Related)

Hypothesis generation and evidence gathering

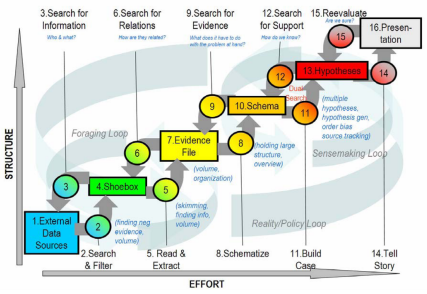


Contributors highlight important items with tags.

Tags help late-joiners identify and build on important comments.

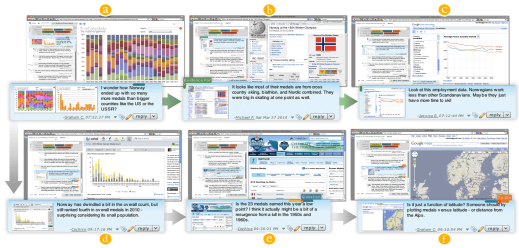


Contributors use links to organize contributions and build narratives.

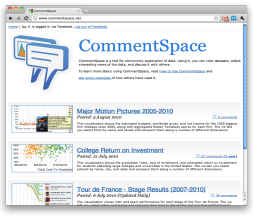


[Pirolli & Card, 1999]

Using tags and links to connect observations



Studies and Deployments



1. Controlled lab studies to test core analysis subtasks
2. Live deployments on the web (www.commentspace.net)

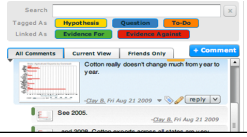
Study 1 - Tags and Links in Support of Analysis

Hypothesis: Tags and links can provide common ground and encourage continued discussion.

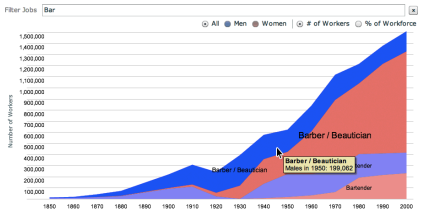
A between subjects study (n=24) with 2 conditions.

"No-Tag" Condition

"Tag" Condition



Study 1 – Prompt



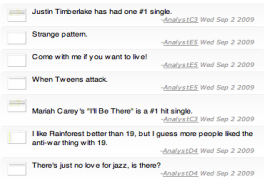
Hypothesis: Stereotypically male jobs have remained almost entirely male even as women have joined the work force.

Study 1 – Results

Participants who used tags and links classified comments more **consistently** and **accurately** than those who didn't.
(greater in-group agreement) (greater agreement with experts)

Participants using tags and links generated significantly more replies to existing comments.
Tag (Median=7)
No-Tag (Median=2)

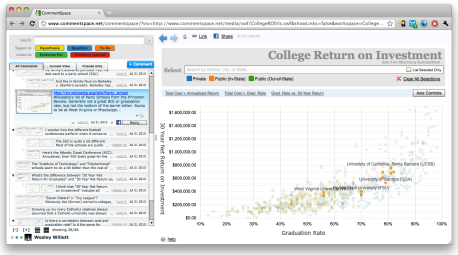
But…



In open-ended tasks, participants still engaged mostly in superficial, exploratory analysis.

We saw very little use of tags or links.

More carefully **managing** the analysis process to produce better analytic results.



3-stage experiment with two analysis teams.

A between subjects study (n=16) with 2 conditions (Tag and No-Tag).

Study 2 - We Broke Analysis into Three Phases:

1. Exploration
2. Organization
   (for participants using tags and links)
3. Synthesis

Tags and links make this staged integration possible.

Exploration

Participants are asked to generate 10 comments in response to two general prompts:

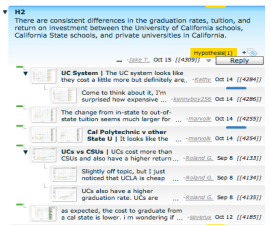
1. The relationship between graduation rate, total cost of attendance, and return on investment. Is there a clear relationship? Are there good examples of this? Are there schools that buck the trend?

2. The distribution of schools from each of the university systems in California (UC system, Cal State system, etc). Are there any trends?

Organization

Tag participants use tags and links to organize their findings around more specific prompts.

No-Tag participants review comments, but do no organizing.



Synthesis

Participants in both conditions use their prior comments to complete a decision-making task.

1. **A.** Is there a correlation between graduation rate, the total cost of attendance, and return on investment? Explain.

   B. You are advising a group of students who are applying to universities. Based on the possible relationship between graduation rate, the total cost of attendance, and return on investment, list the top 5 universities you would recommend and explain why you would recommend them.

Study 2 – Results

Organization – Tag participants spent longer on the task than **No-Tag** participants.

Tag (Median=23 minutes)
No-Tag (Median=12 minutes)

Study 2 – Results

Tag participants synthesized longer responses than those in the no-tag condition.

Tag (Median=3082 total characters, MAD=574)
No-Tag (Median=1480 total characters, MAD=487)

As ranked by three independent evaluators, **participants using tags and links produced better results** than those who did not.

($U=5.5$, $p<0.0013$).
Tag (Median rank*=3.83, MAD=0.5)
No-Tag (Median rank*=6.17, MAD=1)
*Lower is better

Results

Tag participants spent more time with the comments in the organization task, gaining familiarity.

Tag participants cited comments in ways that suggests they used link structure to help guide synthesis.

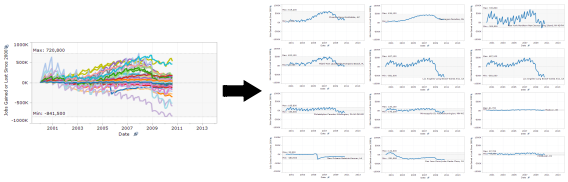
Crowdsourcing Visual Analysis

Can we **decouple** tasks completely?
(e.g. one group explores, another synthesizes)

Are **different staging models** necessary for other types of analysis tasks? *(beyond Explore – Organize – Synthesize)*

Can we identify **design patterns** for distributing common visual analysis tasks?

Can we **break down**, **distribute**, and **recompose** visual analysis tasks?

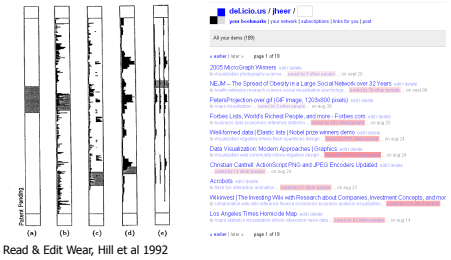


Designing for Social Data Analysis

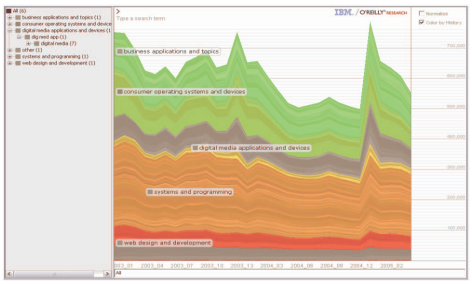
Designing for Social Data Analysis

How can users' **activity traces** be used to improve collaborative analysis?

Social Navigation



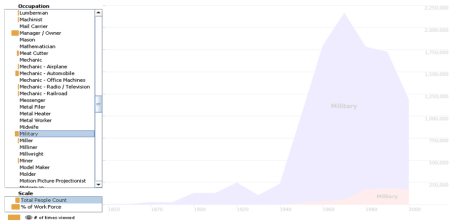
Read & Edit Wear, Hill et al 1992



Wattenberg & Kriss – Color by history: grayed out regions have already been visited

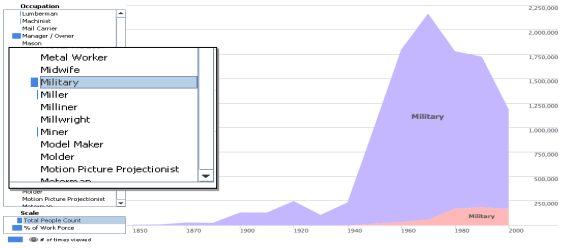
Scented Widgets

[InfoVis 07]

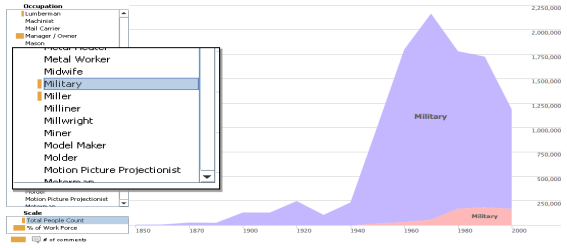


Visual navigation cues embedded in interface widgets

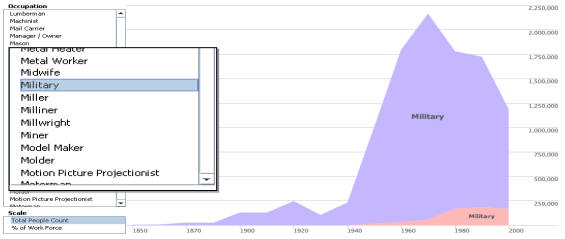
Visitation counts



Comment counts



No scent (baseline)

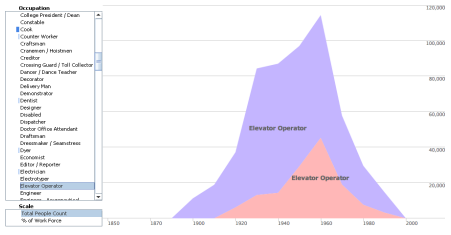


Do social activity cues affect usage?

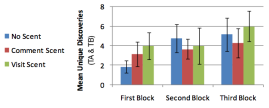
Hypotheses: With activity cues, subjects will
1. Exhibit more revisitation of popular views
2. Make more unique observations

Controlled experiment with 28 subjects
Collect evidence for and against an assertion
Varied scent cues (3) and foraging task (3)
Activity metrics collected from sense.us study

"Technology is costing jobs by making occupations obsolete."



Results



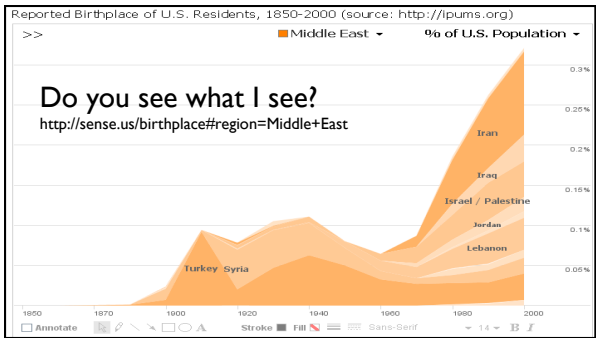
Unique Discoveries
Visit scent had sig. higher rate of discoveries in first block. Less reliance on scent when subjects were familiar with data and visualization.

Revisitation
Visit and comment scent conditions correlate more highly with seed usage than no scent.

Designing for Social Data Analysis

How can users' **activity traces** be used to improve collaborative analysis?

How should **annotation techniques** be designed to provide nuanced pointing behaviors?



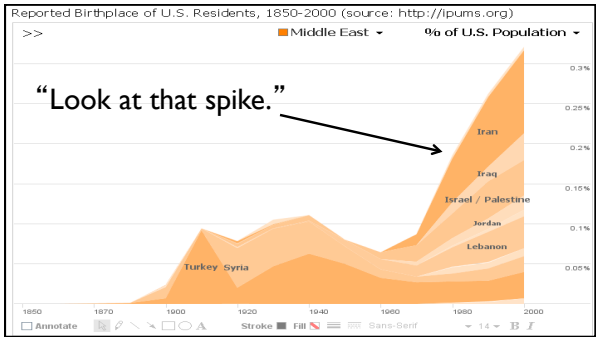
Common Ground

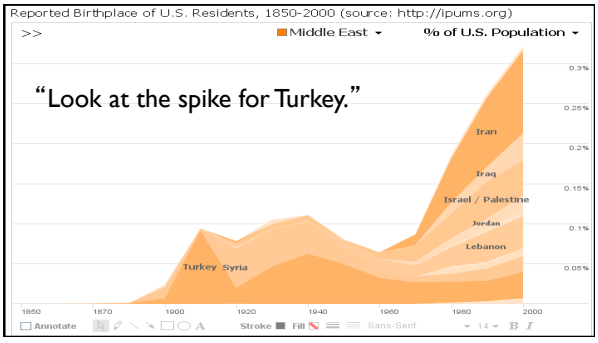
Common Ground: the shared understanding enabling conversation and collaborative action [Clark & Brennan '91]

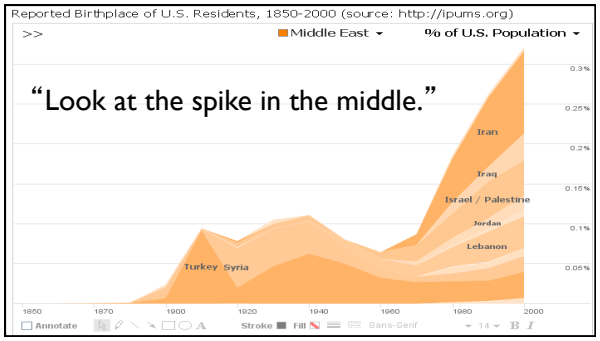
Do you see what I see? → View sharing (URLs)

How do collaboration models affect grounding? Linked discussions vs. embedded comments vs. …

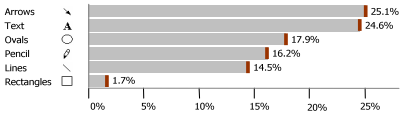
Principle of Least Collaborative Effort: participants will exert *just enough* effort to successfully communicate.
•[Clark & Wilkes-Gibbs '86]





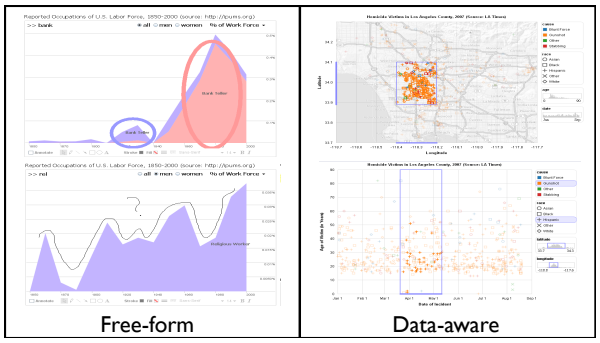


Use of Annotations in Sense.us



39.0% of comments included annotations
Pointing to specific points, trends, or regions (88.6%)
Drawing to socialize or tell jokes (11.4%)

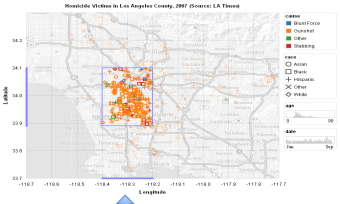
Variety of subject responses
'Not always necessary', but 'surprisingly satisfying'
Some concern about professional look



Free-form | Data-aware

Visual Queries

Model selections as **declarative queries** over interface elements or underlying data



(-118.371 ≤ lon AND lon ≤ -118.164) AND (33.915 ≤ lat AND lat ≤ 34.089)

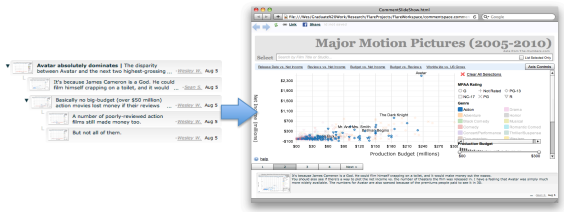
Designing for Social Data Analysis

How can users' **activity traces** be used to improve collaborative analysis?

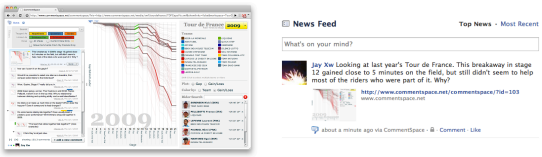
How should **annotation techniques** be designed to provide nuanced pointing behaviors?

How can interface design better support **presentation of analytic findings**?

Building and Presenting Stories



Analysis tools that utilize existing social media



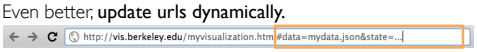
Tools to support and bring in existing communities.

Bootstrap initial analyses by mining social media and engaging with domain experts and enthusiasts.

(An Aside) Support Sharing!

If you want to support social use of your vis provide **unique, bookmarkable URLs.**

Even better, **update urls dynamically.**



Shortcuts for posting to social media (**Email, IM,** Twitter, etc.) are nice, but less important.

Designing for Social Data Analysis

How can users' **activity traces** be used to improve collaborative analysis?

How should **annotation techniques** be designed to provide nuanced pointing behaviors?

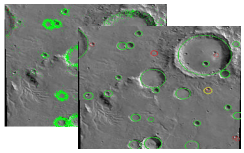
How can interface design better support **presentation of analytic findings**?

How to better **integrate analytic findings?**

Structured Conversation

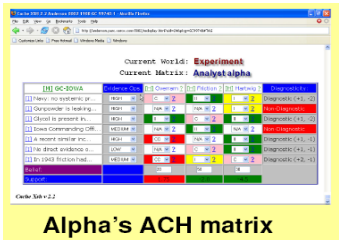
Reduce the cost of synthesizing contributions

Wikipedia: Shared Revisions



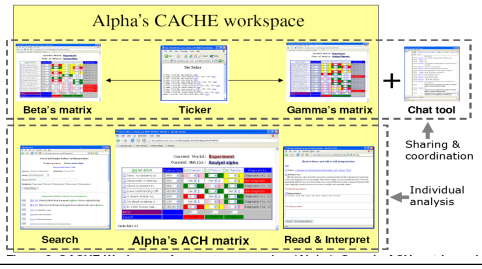
NASA ClickWorkers: Statistics

Integration: Evidence Matrices [Billman et al '06]

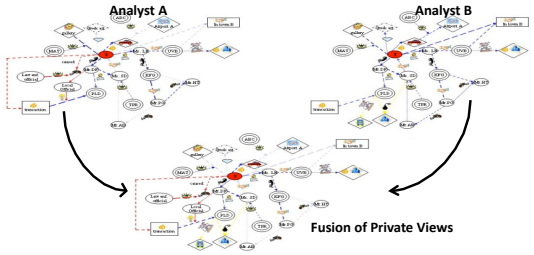


Alpha's ACH matrix

Integration: Evidence Matrices [Billman et al '06]



Merging Analysis Structures [Brennan et al '06]



Summary

As visualization becomes a citizen of the web, opportunities for collaborative analysis abound

Challenges

Weave data visualizations into the web: data access, visualization creation, view sharing and pointing

Support discussion and discovery, but also the integration of contributions to leverage the collective

We need improved processes and technologies for communication and dissemination