Collaborative Visual Analysis

Jeffrey Heer  Stanford University
CS448B Data Visualization  30 Nov 2010

A Tale of Two Visualizations

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.

Analysis of data coupled with social interpretation and deliberation.

How can user interfaces catalyze and support collaborative visual analysis?
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

Don’t disrupt individual exploration

User Study Design

30 participant laboratory study
25 minute, unstructured sessions with job voyager
3-week live deployment on IBM intranet
Employees logged in using intranet accounts

Data analyzed
12.5 hours of qualitative observation
258 comments (41 pilot, 85 ibm, 60 ucb, 72 live)
Usage logs of user sessions

Voyagers and Voyeurs

Complementary faces of analysis

Voyager – focus on visualized data
Active engagement with the data
Serendipitous comment discovery

Voyeur – focus on comment listings
Investigate others’ explorations
Find people and topics of interest
Catalyze new explorations
Social Data Analysis

Spotfire Decision Site Posters
Many-Eyes

Collaborative Sensemaking

Data Jokes
Content Analysis of Comments

<table>
<thead>
<tr>
<th>Observation</th>
<th>Question</th>
<th>Hypothesis</th>
<th>Sense.us</th>
<th>Service</th>
<th>Many-Eyes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Feature prevalence from content analysis (min Cohen’s $\kappa = .74$)
High co-occurrence of Observations, Questions, and Hypotheses

Sharing in External Media
Data Integrity

No cooks in 1910? ... There may have been cooks then. But maybe not.

The great postmaster scourge of 1910? Or just a bug in the data?
Content Analysis of Comments

16% of sense.us comments and 10% of Many-Eyes comments reference data quality or integrity.

Building Off of Others

Integrating Data in Context
What factors enable viable collaborations?

How might we design systems to facilitate social data analysis?

Administrivia

Final Project Milestones

Poster Presentations
Held *Tue Dec 7 4-6pm*, Packard Hall (EE) lobby
Poster + Laptop/device for demos
Arrive *30min early* to setup
*In CS224W?* Come to Packard at *4:30pm*

Final Project Reports
Due *Fri Dec 10*, by *5pm*
6-8 pages in ACM Conf or IEEE TVCG format
Paper, documentation posted to course wiki

Poster Presentation

Poster guidelines on the Final Project wiki page
Remember to include: Problem, Motivation, Approach, Results, and Future Work
Prepare 5-min presentation with poster + demo

Need a 30” x 20” poster (PDF)
*Sign up for printing with Vadim!*
Design Considerations

Modules of Contribution

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

Raw Data
- Data Tables
- Visual Structures
- Views

Data Transformations
- Visual Mappings
- View Transformations

Sensemaking

Visualization
- Select Data Sources
- Apply Visual Encoding
- Author Software

Author Software
- Select Data Sources
- Apply Visual Encoding
- Author Software

Foraging Loop
- 3. Search for Information
- 6. Search for Evidence
- 9. Search for Evidence
- 12. Search for Support
- 15. Reevaluate

Sensemaking Loop
- 1. External Data Sources
- 2. Read & Extract
- 5. Read & Extract
- 8. Schematize
- 11. Build Case
- 14. Tell Story

Foraging Loop
- 3. Search for Information
- 6. Search for Evidence
- 9. Search for Evidence
- 12. Search for Support
- 15. Reevaluate

Sensemaking Loop
- 1. External Data Sources
- 2. Read & Extract
- 5. Read & Extract
- 8. Schematize
- 11. Build Case
- 14. Tell Story
Design Considerations [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

Social Data Analysis

How can users’ activity traces be used to improve awareness in collaborative analysis?

Social Navigation

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

---

**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.

---

“Technology is costing jobs by making occupations obsolete.”

---

**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?
Do you see what I see?
http://sense.us/birthplace#region=Middle+East

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]
“Look at the spike in the middle.”

Use of Annotations

<table>
<thead>
<tr>
<th>Shape</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrows</td>
<td>25.1%</td>
</tr>
<tr>
<td>Text</td>
<td>24.6%</td>
</tr>
<tr>
<td>Ovals</td>
<td>17.9%</td>
</tr>
<tr>
<td>Pencil</td>
<td>16.2%</td>
</tr>
<tr>
<td>Lines</td>
<td>14.5%</td>
</tr>
<tr>
<td>Rectangles</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

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

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)
Visual Queries

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

Applicable to **dynamic, time-varying data**

Retarget selection **across visual encodings**

Support **social navigation and data mining**

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 **communication of analytic findings**?
Graphical Analysis Histories

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 can user contributions be better integrated?

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**

Merging Analysis Structures (Brennan et al ‘06)

**Fusion of Private Views**

Integration: Evidence Matrices (Billman et al ‘06)

**Alpha’s CACHE workspace**

α’s matrix  β’s matrix  γ’s matrix

Chat tool

Sharing & coordination

Individual analysis

Design Considerations [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
Remaining Challenges

Weave visualizations into the web: data access, visualization creation, address data quality issues, view sharing and pointing.

Support discussion, discovery, and integration of contributions.

Improve processes and technologies for communication and dissemination.