The Data Lifecycle is Social

In many organizations, different groups often do cleaning, integration, visualization, analysis...

Social sharing may improve acquisition, cleaning, integration, hypothesis generation, discussion

The Design Challenge:
Which forms of social interaction and awareness cues will best facilitate successful analysis?
Today:
Collaborative Visual Analysis
Narrative Visualization
Course Wrap-Up
Content Analysis of Comments

Feature prevalence from content analysis (min Cohen’s $K = .74$)
High co-occurrence of Observation, Question, and Hypothesis

16% of sense.us comments and 10% of Many-Eyes comments reference data integrity issues.
Forum-style commenting

limits analysis

Finding important comments is difficult

No support for organizing comments

- No Reordering
- No Grouping

Comments and visualizations lack tight coupling
CommentSpace
Structured Support for Collaborative Visual Analysis

Wesley Willett, et al. CHI 2011
video: http://vimeo.com/19275098

Evaluating the impact of lightweight structure

Studies and Deployments

Study 1 – Controlled lab study to test core analysis subtasks
Live deployments on the web (www.commentspace.net)

Study 2 – Lab study of more complex team analyses

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

No-Tag Condition
Filtering tools
Tags / Links
Study 1 - Tags and Links in Analysis Subtasks

Do tags and links aid identification of existing comments?

How do analysts use tags and links when gathering evidence?

Study 1 - Task

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

1. Identify existing comments that give evidence for and against.
2. Explore the visualization and document new evidence.

Study 1 - Results

Tag participants identified evidence more consistently and more accurately.

(greater in-group agreement) Tag (κ=0.302) No-Tag (κ=0.262)

(greater agreement with experts) Expert & Tag (κ=0.339) Expert & No-Tag (κ=0.303)
Study 1 – Results

Tag participants authored more replies. 
(U=32, p<0.023)  
Tag (Median=7, MAD=3.5)  
No-Tag (Median=2, MAD=1.5)

Tag participants were more likely to reply to tagged comments. 
(χ²(1,308)=27.45, p<0.001)

Live Deployments

but...

Mostly superficial exploratory analysis
Little use of tags or links

Can we manage the analysis process to produce better analytic results?
A between-subjects study (n=16) with 2 conditions (Tag and No-Tag)

Study 2 – Staged Analysis
1. Exploration
2. Organization
3. Synthesis
Results: Exploration & Organization

Tag and No-Tag participants authored a similar number of comments.
- Tag (Median=10, MAD=0)
- No-Tag (Median=10, MAD=0)

Tag participants spent more time than No-Tag participants.
- Tag (Median=23 minutes)
- No-Tag (Median=12 minutes)

Results: Synthesis

Tag participants synthesized longer responses.
- Tag (Median=3082 total characters, MAD=574)
- No-Tag (Median=1480 total characters, MAD=487)

As ranked by three independent evaluators, Tag participants gave better responses.
- (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

Study 2 - Findings

Tag participants spent more time organizing comments and gaining familiarity.

Tag participants’ used link structure to guide their synthesis.

A more formal, staged process appears to facilitate deeper collaborative analyses.

Future Research

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 use these strategies to facilitate collaborative analysis at larger scales?
Final Projects

Final Presentations

Monday June 6, 5-7:30pm, 124 Wallenberg
You can use your own laptop, or send us slides

Prepare a 5 minute presentation:
Motivate the problem. Why does this matter?
Present your solution. Demos recommended.
Discuss remaining challenges or future work

Final Papers – 4 pages, ACM CHI format
Due Wednesday, June 8 by 5pm

Narrative Visualization

“to affect thro’ the Eyes what we fail to convey to the public through their word-proof ears”
Visualization designers are melding the skills of computer science, statistics, artistic design and storytelling.

*The Economist, Feb 2010*

...require[s] skills like those familiar to movie directors, beyond a technical expert's knowledge of computer engineering and science.

*Gershon & Page, 2001*

people have begun to forget how powerful human stories are, exchanging their sense of empathy for a fetishistic fascination with data ...the human stuff is the main stuff, and the data should enrich it.

*Jonathan Harris, 2008*
What is storytelling?

The world’s second oldest profession

*narrative* (n): An account of a series of events, facts, etc., given in order and with the establishing of connections between them.

*Oxford English Dictionary*
Introduction to the people, places, and situations of interest.

A series of events, often with tension, conflict, and challenges.

A resolution or return to form, implications for the future and lessons learned.

**Beginning**  **Middle**  **End**

---

### 5W + H:

I keep six honest serving-men (They taught me all I knew);
Their names are **What** and **Why** and **When** and **How** and **Where** and **Who**.

*Rudyard Kipling, 1902*

---

**Generals Wary of Move to Cut Their Ranks**

By Ginger Thompson and Thom Shumaker

WASHINGTON — Maj. Gen. Paul D. Eaton, a retired Army officer, is familiar with the perks and pitfalls of power, having commanded tens of thousands of troops at Fort Benning, Ga., managed budgets exceeding $2 billion in Iraq, and overseen layers upon layers of staff members who helped manage both his professional duties and his personal life.

He has experienced the full range of lifestyles that come with military leadership, living at one point in an elegant antebellum mansion, and at another, with eight other officers crowded in a marble bathroom behind one of Saddam Hussein’s old palaces.

When he traveled, he was occasionally able to justify the use of military aircraft, but most times, he said, he flew coach. And today he lives on a pension worth 75 percent of his military salary, with health benefits that cover everything except dental and eye care for himself and his wife.

“We are well compensated, and we live very comfortable lives,” General Eaton said, referring to the military’s most senior leaders. “But when you look at all the things going on around a general, the nation is getting a very, very high return on its money.”

Not everyone at the Pentagon agrees. Two weeks ago, Defense Secretary Robert M. Gates announced a sweeping effort to improve efficiency that, among other things, takes aim at the military’s ascendant corps of generals and admirals, ordering his staff to cut at least 50 positions, and making clear that he would be happier if they cut more.
Generals Wary of Move to Cut Their Ranks

By GINGER THOMPSON and THOM SHARIER

WASHINGTON — Maj. Gen. Paul D. Bateman, a retired Army officer, is familiar with the perks and pitfalls of power, having commanded tens of thousands of troops at Fort Benning, Ga., and overseen layers upon layers of staff members who helped manage both his professional duties and his personal life.

He has experienced the full range of benefits that come with military leadership, living at one point in an elegant antebellum mansion, and at another, with eight other officers crowded in a marble bathroom behind one of Saddam Hussein’s old palaces.

When he traveled, he was occasionally able to justify the use of military aircraft, but most times, he said, he flew coach. And today he lives on a pension worth 75 percent of his military salary, with health benefits that cover everything except dental and eye care for himself and his wife.

“We are well compensated, and we live very comfortable lives,” General Bateman said, referring to the military’s most senior leaders. “But when you look at all the things going on around a general, the nation is getting a very, very high return on its money.”

Anecdotal Lead

Not everyone at the Pentagon agrees. Two weeks ago, Defense Secretary Robert M. Gates announced a sweeping effort to improve efficiency that, among other things, takes aim at the military’s senior corps of generals and admirals, ordering his staff to cut at least 50 positions, and making clear that he would be happier if they cut more.

Nut Graph
What **narrative devices** exist for telling stories with data?

**Case Study #1**

Steroids or Not, The Pursuit is On

*The New York Times*
Case Study #2

Human Development Trends

Gapminder
Case Study #3

Budget 2010: Reaction from Around the UK

The Guardian
Consistent Visual Platform

Introductory Text

Drill-Down Story

Anecdotal Stories with Data Stories

Visual Structuring

Establishing Shot / Splash Screen
Consistent Visual Platform
Checklist Structure
Progress Bar / Timeline
Highlighting Direction

Interactivity

Details on Demand / Highlighting
Filtering / Selection / Search
Explicit Instruction
Tacit Tutorial
Single Frame Interactivity

Transition Guidance

Familiar Objects
Viewing Angle
Camera Motion
Continuity Editing
Object Constancy
Animated Transitions
Staged Transitions

Messaging

Captions / Headlines
Annotations
Accompany Article
Multi-Messaging
Repetition
Introductory Text
Summary / Synthesis
Genres for Narrative Visualization?

**Author-Driven**
- Linear Ordering of Images
- Heavy Messaging
- Limited Interactivity

**Reader-Driven**
- No Proscribed Ordering
- No Messaging
- Free Interactivity

**Works well for...**
- Storytelling
- Clear Communication
- Fast Communication
- Data Diagnostics
- Pattern Discovery
- Hypothesis Formation

Most examples fall somewhere in-between...
Future Research

Does the narrative visualization design space suggest new or enhanced techniques?

Which configurations of devices best support narrative communication? How do people “read” a data story? How does this vary by audience and data domain?

How should visualization design tools support authoring of narrative devices?

Course Goals

To explore how a broad class of data analysts can more effectively analyze data using novel interactive tools.

The class will be interdisciplinary in nature, with a goal of identifying and pursuing new research opportunities.

Course Expectations

You should expect to gain:

An overview of the research landscape in interactive data analysis (HCI, Vis, DB).

Engaged discussions with leading researchers and practitioners in the area of “data science.”

Hands-on experience working with scalable data analysis frameworks.

Insights from a quarter-long research project prototyping a new interactive analysis tool.
Course Reflection

What did you find most valuable in the course? What lessons do you expect to take with you?

What are the most pressing issues in need of further research attention?

What emerging technologies and societal trends are likely to shape data analysis in the future?

Thank You!

I used to think correlation implied causation.

Then I took a statistics class. Now I don’t.

Sounds like the class helped.

Well, maybe.

And don’t forget to fill out the course survey...

http://xkcd.com/552/