InfoVis Evaluation



CS 7450 - Information Visualization November 22, 2011 John Stasko

Area Focus



- Most of the research in InfoVis that we've learned about this semester has been the introduction of a new visualization technique or tool
 - Fisheyes, cone trees, hyperbolic displays, tilebars, themescapes, sunburst, jazz, ...
 - "Isn't my new visualization cool?..."

Fall 2011 CS 7450 2

1

Reflection



- Creation of new techniques is very important but...
 - It's also important to know that we're getting better
 - So, it's important that we evaluate the visualizations being created

Fall 2011 CS 7450

Evaluation – Why?



• Reasons?

Evaluation – Why?



- Want to learn what aspects of visualizations or systems "works"
- Want to ensure that methods are improving
- Want to insure that technique actually helps people and isn't just "cool"
- NOT: Because I need that section in my paper to get it accepted ... sigh

Fall 2011 CS 7450

Evaluation – Measures?



 How does one judge the quality of work in Information Visualization?

Evaluation - Measures?



- Different possible ways
 - Impact on community as a whole, influential ideas
 - Assistance to people in the tasks they care about

Fall 2011 CS 7450

Strong View



 Unless a new technique or tool helps people in some kind of problem or task, it doesn't have any value



Broaden Thinking



- Sometimes the chain of influence can be long and drawn out
 - System X influences System Y influences
 System Z which is incorporated into a practical tool that is of true value to people
- This is what research is all about (typically)

Fall 2011 CS 7450

Evaluation - How?



- What evaluation techniques should we use?
 - (Channel your HCI knowledge)

Evaluation in HCI



- Takes many different forms
 - Qualitative, quantitative, objective, subjective, controlled experiments, interpretive observations, ...
- So, which ones are best for evaluating InfoVis systems?

Fall 2011 CS 7450 11

Controlled Experiments



- Good for measuring performance or comparing multiple techniques
- What do we measure?
 - Performance, time, errors, ...
- Strengths, weaknesses?

Subjective Assessments



- Find out people's subjective views on tools
 - Was it enjoyable, confusing, fun, difficult, …?
- This kind of personal judgment strongly influence use and adoption, sometimes even overcoming performance deficits

Fall 2011 CS 7450 13

Qualitative, Observational Studies



- Watch systems being used (you can learn a lot)
- Is it being used in the way you expected?
- Ecological validity
- Can suggest new designs and improvements
- (Channel HCI knowledge)

Running Studies



- Beyond our scope here
- You should learn more about this in CS 6750 or 6455

Fall 2011 CS 7450 1

Evaluating UI vs. InfoVis



- Seems comparable but...
- What are some differences?

Usability vs. Utility



- Big difference
- Usability is not the same as utility, which seems to be a key factor for InfoVis
- Can think of visualizations that are very usable but not useful or helpful
- More difficult to measure success of an infovis because more domain knowledge and situated use is required

Fall 2011 CS 7450 17

Evaluating InfoVis in General



- Very difficult in InfoVis to compare "apples to apples"
 - Hard to compare System A to System B
 - Different tools were built to address different user tasks
- UI can heavily influence utility and value of visualization technique

Plaisant Paper



- Discuss
- Challenges identified?
- Possible next steps?

Plaisant AVI '04



Past Review



- Old journal issue whose special topic focus was Empirical Studies of Information Visualizations
 - International Journal of Human-Computer Studies, Nov. 2000, Vol. 53, No. 5
- A bit dated now

Fall 2011 CS 7450 21

Examples



- Let's examine a few example studies, changing focus from
 - Comparative, experimental
 - Hybrid
 - Observational, case studies

Which System is Best?





- Commercial tools evaluation
- Empirical study of 3 InfoVis tools
 - Eureka, Spotfire, InfoZoom
- Methodology
 - 3 data sets
 - 83 students
 - Within subjects, 30 minutes per tool

Kobsa InfoVis '01

Fall 2011 CS 7450 23

More Methodology



- Tasks
 - Very much from the 10 low-level tasks type of questions (specific, not exploratory)
- Measurements
 - Correctness
 - Time

Results



- Time:
 - InfoZoom 80 seconds
 - Spotfire 107 secoonds
 - Eureka 110 seconds

Correctness

Not a lot here

- Spotfire 75%
- Eureka 71%
- InfoZoom 68%

Fall 2011 CS 7450 25

Findings



- Interaction Problems
 - Eureka

Confusion by hidden labels, problems with 3 or more vars., correlation errors

- InfoZoom
 - Correlations errors
- Spotfire

Cognitive set-up costs, scatterplot bias

Findings



- Success depends on
 - Properties of visualization
 - Operations that can be performed on vis
 - Concrete implementation of paradigm
 - Visualization-indept usability problems
- I would have liked even more discussion on how tools assisted with different classes of user tasks

Fall 2011 CS 7450 27

Which Technique is Best?

Start

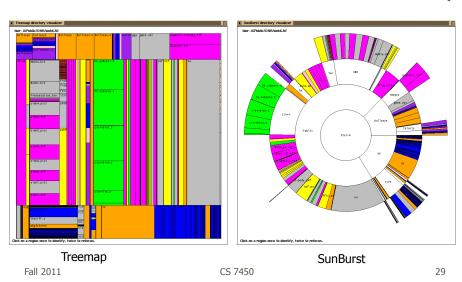


- Space-filling hierarchical views
- Compare Treemap and Sunburst with users performing typical file/directoryrelated tasks
- Evaluate task performance on both correctness and time

Stasko et al IJHCS '00

Tools Compared

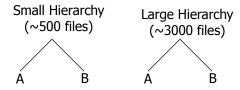




Hierarchies Used



Four in total



 Used sample files and directories from our own systems (better than random)

Methodology



- 60 participants
- Participant only works with a small or large hierarchy in a session
- Training at start to learn tool
- Vary order across participants

SB A, TM B
TM A, SB B
SB B, TM A
TM B, SB A
32 on small hierarchies
28 on large hierarchies

Fall 2011 CS 7450 31

Tasks



- Identification (naming or pointing out) of a file based on size, specifically, the largest and second largest files (Questions 1-2)
- Identification of a directory based on size, specifically, the largest (Q3)
- Location (pointing out) of a file, given the entire path and name (Q4-7)
- Location of a file, given only the file name (Q8-9)
- Identification of the deepest subdirectory (Q10)
- Identification of a directory containing files of a particular type (Q11)
- Identification of a file based on type and size, specifically, the largest file of a particular type (Q12)
- Comparison of two files by size (Q13)
- Location of two duplicated directory structures (Q14)
- Comparison of two directories by size (Q15)
- Comparison of two directories by number of files contained (Q16)

Hypothesis



- Treemap will be better for comparing file sizes
 - Uses more of the area
- Sunburst would be better for searching files and understanding the structure
 - More explicit depiction of structure
- Sunburst would be preferred overall

Fall 2011 CS 7450 33

Small Hierarchy



Hierarch	hy A		Hierarch	ny B	
Tool	Phase	Correct	Tool	Phase	Correct
TM (n = 8)	1	9.88 (3.23)	TM(n = 8)	1	11.50 (2.14)
SB $(n = 8)$	1	12.88 (1.96)	SB(n=8)	1	10.38 (1.69)
TM(n = 8)	2	12.25 (1.75)	TM(n=8)	2	10.75 (2.77)
SB $(n = 8)$	2	12.63 (2.00)	SB(n=8)	2	11.50 (2.00)
TM (collapsed across phase)		11.06 (2.79)	TM (collapsed across phase)		11.13 (2.42)
SB (collapsed across phase)		12.75 (1.91)	SB (collapsed across phase)		10.94 (1.88)

Correct task completions (out of 16 possible)

Large Hierarchy



Hierarch	ny A		Hierarch	ny B	
Tool	Phase	Correct	Tool	Phase	Correct
TM(n=7)	1	8.71 (1.60)	TM(n=7)	1	8.29 (2.14)
SB(n=7)	1	11.43 (1.27)	SB $(n=7)$	1	11.14 (2.67)
TM(n=7)	2	11.57 (1.27)	TM(n=7)	2	10.86 (1.57)
SB(n=7)	2	11.00 (2.16)	SB(n=7)	2	11.00 (2.00)
TM (collapsed across phase)		10.14 (2.03)	TM (collapsed across phase)		9.57 (2.24)
SB (collapsed across phase)		11.21 (1.72)	SB (collapsed across phase)		11.07 (2.27)

Correct task completions (out of 16 possible)

Fall 2011 CS 7450

Performance Results



- Ordering effect for Treemap on large hierarchies
 - Participants did better after seeing SB first
- Performance was relatively mixed, trends favored Sunburst, but not clear-cut
 - Oodles of data!

Subjective Preferences



- Subjective preference:
 SB (51), TM (9), unsure (1)
- People felt that TM was better for size tasks (not borne out by data)
- People felt that SB better for determining which directories inside others
 - Identified it as being better for structure

Fall 2011 CS 7450 37

Strategies



- How a person searched for files etc. mattered
 - Jump out to total view, start looking
 - Go level by level

DQ vs. BH



- Empirical Study
 - Use DataMaps, a geographic (US states) data visualization tool
 - Have participants do different tasks with both methods

How many states have pop between x and y in 1970? Given 3 states, which has the lowest median income? What's the relationship between education and income? List states with pops. 0->x and y->z. What kind of a state is Florida?

We saw this earlier in term

Li & North InfoVis '03

Fall 2011 CS 7450

Findings

Functioned more as its own infovis tool



- Brushing histograms better and more highly rated for more complex discovery tasks
 - Attribute correlation, compare, and trend evaluation
- Dynamic queries better for more simple range specification tasks
 - Single range, multiple ranges, multiple criteria

Functioned more as auxiliary control for other vizs

Animation Really Good?

Start



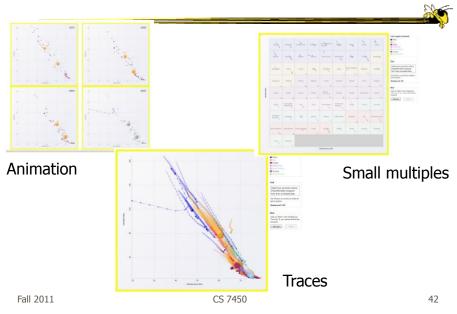
- Examine whether animated bubble charts (a la Rosling and GapMinder) are beneficial for analysis and presentation
- Run an experiment to evaluate the effects of animation



Robertson et al TVCG (InfoVis) '08

Fall 2011 CS 7450 4:

Visualizations Studied



Experiment Design



- 3 (animation types) x 2 (data size: small & large) x 2 (presentation vs. analysis)
- Data
 - UN data about countries
- Tasks
 - 24 tasks, 1-3 requires answers per
 Example: Select 2 countries with significant decreases in energy consumption

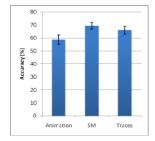
Fall 2011 CS 7450 43

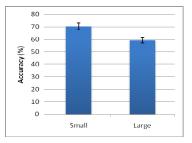
Results



Accuracy

Measured as percentage correct 65% overall (pretty tough)





Significant:

SM better than animation Small data size more accurate than large

Results



- Speed
 - Presentation
 Animation faster than small multiples & traces
 15.8 secs vs. 25.3 secs vs. 27.8 secs.
 - Analysis
 Animation slower than small multiples & traces
 83.1 secs. vs. 45.69 secs. vs. 55.0 secs.

Fall 2011 CS 7450 4

Discussion

More on this study coming up on Animation day



- People rated animation more fun, but small multiples was more effective
- As data grows, accuracy becomes an issue
 - Traces & animation get cluttered
 - Small multiple gets tiny
- Animation:
 - "fun", "exciting", "emotionally touching"
 - Confusing, "the dots flew everywhere"

Useful Junk?

Start

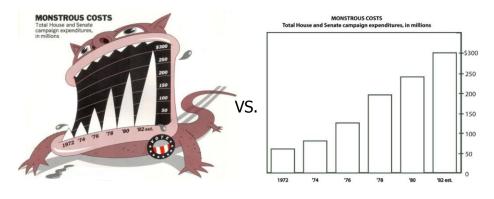


- Tufte claimed that graphs loaded with chartjunk are no good
- Is that really so?
- How could you test this?

Fall 2011 CS 7450 47

Comparing



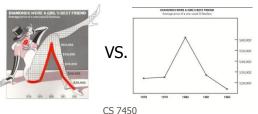


Bateman et al CHI '10

Methodology



- Two versions of each chart
- Participant sees one
 - Asked immediate interpretation accuracy questions
 - Asked similar questions again 5 minutes or 2-3 weeks later



Fall 2011 CS 7450

Results



- No significant difference in immediate interpretation accuracy, or after 5 minute gap
- After 2-3 week gap, recall of chart topic and details was significantly better for chartjunk graphs
- Participants found the chartjunk graphs more attractive, enjoyed them more, and found them easiest and fastest to remember

Caveats



- Small datasets
- "Normal" charts were really plain
- No interaction
- How about other added interpretations from the flowery visuals?
- Be careful reading too much into this

Fall 2011 CS 7450 51

Insight





- Isn't one of the key ideas about InfoVis that it helps generate insights?
- OK, well let's count/measure insights
- What challenges do you see in this?

Saraiya, North, Duca *TVCG* '05

Problem Domain



- Microarray experiments: Gain insight into the extremely complex and dynamic functioning of living cells
- Systems-level exploratory analysis of thousands of variables simultaneously
- Big data sets

Fall 2011 CS 7450 53

Insight



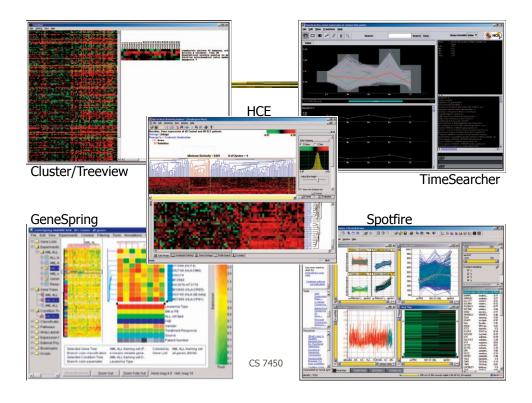
- Insight: An individual observation about the data by the participant, a unit of discovery
- Characteristics
 - Observation
 - Time
 - Domain Value
 - Hypotheses
 - Directed vs Unexpected
 - Category

Experiment Design

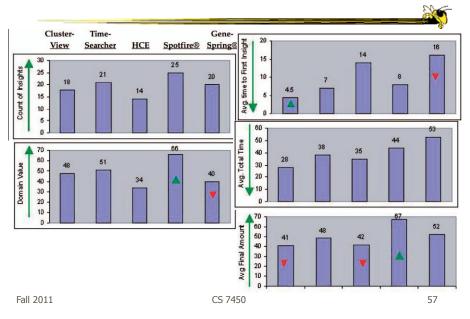


Data: Timeseries, Virus, Lupus

Tool	Visual Representations	Interactions
Cluster/ Treeview	Heat-map, Clustered heat-map	O+D
Time- Searcher	Parallel coordinates, line graph	Brushing, O+D, DQ
HCE	Cluster dendrogram, parallel coordinates, heat-map, scatterplot, histogram	Brushing, Zooming, O+D, DQ
Spotfire® 7.2 Functional Genomics	Parallel coordinates, heat-map, scat- terplots (2D/3D), histogram, bar/pie chart, tree view, spreadsheet view, Clustered parallel coordinates	Brushing, Zooming, O+D, DQ
GeneSpring ® 5.0	Parallel coordinate, heat-map, scatter- plots (2D/3D), histogram, bar chart, block view, physical position view, array layout view, pathway view, spreadsheet view, compare gene to gene, Clusterested parallel coordinates	Brushing, Zooming



Results



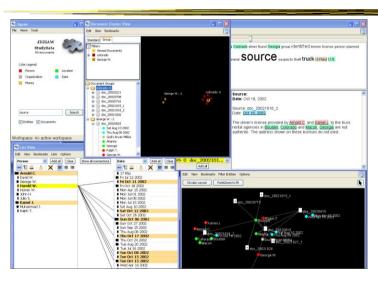
More Complex Task Eval



- Consider investigative analysis tasks involving sensemaking, awareness, and understanding
- Research questions
 - How do people use systems?
 - What characteristics matter?
 - What should we measure/observe?
- Exploring methods for utility evaluation

Kang et al VAST '08 & TVCG'11

System Examined - Jigsaw



Fall 2011 CS 7450 59

Study Design



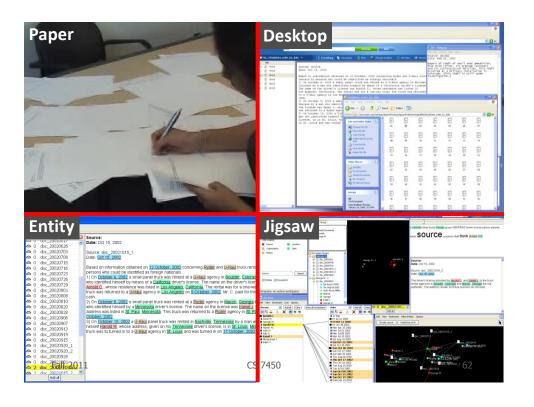
- Task and dataset
 Your HW 8
 - 50 simulated intelligence case reports
 Each a few sentences long
 23 were relevant to plot
 - Identify the threat & describe it in 90 minutes

Source: doc017 **Date**: Oct 22, 2002

Abu H., who was released from custody after the September 11 incidents and whose fingerprints were found in the U-Haul truck rented by Arnold C. [see doc033] holds an Egyptian passport. He is now known to have spent six months in Afghanistan in the summer of 1999.

Study Design - Settings





Performance Measures



- Task sheets (like VAST Contest)
 - Three components (relevant people, events, locations)
 - +1 for correct items, -1 for a misidentified items
- Summary narrative
 - Subjective grading from 1 (low) to 7 (high)
- Two external raters
- Normalized, each part equal, mapped to 100point scale

Fall 2011 CS 7450

Results



		Paper			Des	ktop		Entity				Jigsaw				
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Final Score	22.87	65.00	24.26	87.08	62.08	67.13	42.13	29.41	52.23	15.00	29.26	81.19	95.05	58.07	75.20	90.00
Performance	Fair	Very good	Fair	Excel- lent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excel- lent	Excel- lent	Good	Very good	Excel- lent
Average Score		49	.80			50	.19			44	.42			79	.59	
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23
# of Queries					19	18	48	8	23	61	59	91	44	4	26	8
First Query					40:49	19:55	2:47	12:41	1:31	0:29	0:59	3:12	0:18	5:35	25:37	4:18
Amount of Notes	Many	None	Many	Some	Many	Some	Few	Some	Some	None	None	Few	Some	Few	Few	Few
First Note Taking	0:07		0:05	0:16	1:53	19:57	2:47	8:20	2:37			3:14	0:48	0:32	5:15	78:45
First Task Sheet	43:20	32:53	70:13	3:25	61:35	20:26	7:33	64:11	28:09	0:52	2:55	7:20	48:26	41:48	43:00	5:33

Results

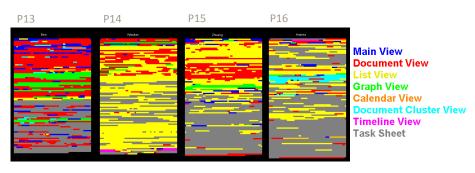


		Pa	per			Des	ktop			En	tity		Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Final Score	22.87	65.00	24.26	87.08	62.08	67.13	42.13	29.41	52.23	15.00	29.26	81.19	95.05	58.07	75.20	90.00
Performance	Fair	Very good	Fair	Excel- lent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excel- lent	Excel- lent	Good	Very good	Excel- lent
Average Score		49	.80			50	.19			44	.42			79	.59	
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23
# of Queries					19	18	48	8	23	61	59	91	44	4	26	8
First Query					40:49	19:55	2:47	12:41	1:31	0:29	0:59	3:12	0:18	5:35	25:37	4:18
Amount of Notes	Many	None	Many	Some	Many	Some	Few	Some	Some	None	None	Few	Some	Few	Few	Few
First Note Taking	0:07		0:05	0:16	1:53	19:57	2:47	8:20	2:37			3:14	0:48	0:32	5:15	78:45
First Task Sheet	43:20	32:53	70:13	3:25	61:35	20:26	7:33	64:11	28:09	0:52	2:55	7:20	48:26	41:48	43:00	5:33

Fall 2011 CS 7450 65

Jigsaw Usage Patterns







Investigative Strategies



- 1. Overview, filter and detail (OFD)
- 2. Build from detail (BFD)
- 3. Hit the keyword (HTK)
- 4. Find a clue, follow the trail (FCFT)

P16: "I like this people-first approach. Once I identify key people, then things that are potentially important come up, too. I'm an impatient person and don't want to read all documents chronologically."

Fall 2011 CS 7450 65

Results by Strategy



		Pa	per			Des	ktop			En	tity			Jig	saw	
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excel- lent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excel- lent	Excel- lent	Good	Very good	Excel- lent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Results by Strategy



		Pa	per			Des	ktop			En	tity			Jigs	saw	
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excel- lent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excel- lent	Excel- lent	Good	Very good	Excel- lent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Fall 2011 CS 7450 69

Results by Strategy



		Pa	per			Des	ktop			En	tity			Jigs	saw	
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excel- lent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excel- lent	Excel- lent	Good	Very good	Excel- lent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Results by Strategy



		Pa	per			Des	ktop			En	tity			Jigs	saw	
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excel- lent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excel- lent	Excel- lent	Good	Very good	Excel- lent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Fall 2011 CS 7450 71

Results by Strategy



		Paper			Desktop				En	tity		Jigsaw				
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT
Performance	Fair	Very good	Fair	Excel- lent	Very good	Very good	Good	Fair	Good	Poor	Fair	Excel- lent	Excel- lent	Good	Very good	Excel- lent
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23

Tool Design Implications



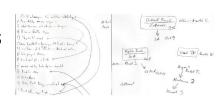
- Support finding starting points/clues
- Guide the analyst to follow the right trail
- Support different strategies of SM process
- Support smooth transition between SM stages
- Provide a workspace
- Allow flexibility in organizing
- Support to find next steps when dead-end
- Facilitate further exploration

Fall 2011 CS 7450 73

Jigsaw's Influence



- Supporting different strategies
- Showing connections between entities
- Helping users find the right clue
- Helping users focus on essential information
- Reviewing hypotheses
- Increasing motivation



Evaluation Recommendations



- Compare system usage to traditional methods
- Collect qualitative data, support with quantitative data
- Consider questions to be answered
- Possible metrics
 - Number of documents viewed
 - When note-taking initiated
 - The quantity of representations created
 - Amount of time and effort in organizing
 - Time spent in reading/processing relevant information

Fall 2011 CS 7450 75

Evaluation Methodology

Start



- Multi-dimensional In-depth Long-term Case Study (MILC)
- M observations, interviews, surveys, logging
- I intense engagement of researchers with domain experts so as to almost become a partner
- L longitudinal use leading to strategy changes
- C Detailed reporting about small number of people working on their own problems in their own domain

Shneiderman & Plaisant BELIV '06

Influences



- Ethnography
 - Preparation
 - Field study
 - Analysis
 - Reporting

Fall 2011 CS 7450 77

Guidelines



- Specify focused research questions & goals
- Identify 3-5 users
- Document current method/tool
- Determine what would constitute professional success
- Establish schedule of observation & interviews
- Instrument tool to record usage

- Provide attractive log book for comments
- Provide training
- Conduct visits & interviews
- Encourage users to continue using best tool for task
- Modify tool as needed
- Document successes and failures

SocialAction

Start

- Evaluation inspired by MILC ideas goals
 - Interview (1 hour)
 - Training (2 hours)
 - Early use (2-4 weeks)
 - Mature use (2-4 weeks)
 - Outcome (1 hour)



Perer & Shneiderman CHI '08

Fall 2011 CS 7450 79

Methodology



- Four case studies
 - Senatorial voting patterns
 - Medical research knowledge discovery
 - Hospital trustee networks
 - Group dynamics in terrorist networks
- Named names
 - I like it!
- Tell what they did with system

My Reflections



- Nice paper
- Stark contrast to comparative, controlled experiments
- We likely need more of this in InfoVis

Fall 2011 CS 7450

How to Evaluate Many Eyes?





- Two main evaluation papers written about system
- Studied use of system, visualizations being created, discussions about system, etc.

Paper 1



- Case study of early use
- System uses
 - Visual analytics
 - Sociability
 - Generating personal and collective mirrors
 - Sending a message

Viégas et al HICSS '08

Fall 2011 CS 7450 83

Use Characteristics



Data Topic/Area	Percentage
Society	14.0
Economics	12.7
Obscured/Anon	12.4
Art & culture	10.8
Web & new media	10.3
Science	10.0
Test data	9.5
Politics	7.4
Technology	6.6

Comment Type	Percentage
Observation	46.3
Question	15.8
Affirmation	13.7
Hypothesis	11.6
Socializing	11.6
System design	11.6
Data integrity	9.5
Testing	4.2
Tips	4.2
To do	4.2

Fall 2011 CS 7450 84

42

Paper 2



- Interview-based study
- Individual phone interviews with 20 users
 - Lots of quotes in paper
- Bloggers vs. regular users
- Also includes stats from usage logs
 - 3069 users
 - 1472 users who uploaded data
 - 5347 datasets
 - 972 users who created visualizations
 - 3449 visualizations
 - 222 users who commented
 - 1268 comments

Danis et al CHI '08

Fall 2011 CS 7450

Findings



- User motivations
 - Analyzing data
 - Broadening the audience, sharing data
- Lots of collaborative discussion
 - Much off the ManyEyes site
- Concerns about data and other eyes

Summary



- Why do evaluation of InfoVis systems?
 - We need to be sure that new techniques are really better than old ones
 - We need to know the strengths and weaknesses of each tool; know when to use which tool

Fall 2011 CS 7450

Challenges



- There are no standard benchmark tests or methodologies to help guide researchers
 - Moreover, there's simply no one correct way to evaluate
- Defining the tasks is crucial
 - Would be nice to have a good task taxonomy
 - Data sets used might influence results
- What about individual differences?
 - Can you measure abilities (cognitive, visual, etc.) of participants?

Challenges



- Insight is important
 - Great idea, but difficult to measure
- Utility is a real key
 - Usability matters, but some powerful systems may be difficult to learn and use
- Exploration
 - InfoVis most useful in exploratory scenarios when you don't know what task or goal is So how to measure that?!

Fall 2011 CS 7450 8

HW 8



- Investigative analysis
- You play the intelligence analyst
- Find the criminal plot embedded across 50 documents
- Paragraph summarizing the threat and a description of what you did
- Due Thursday 1st

Upcoming





- Visual Analytics 1
 - PaperKeim et al '08
- Visual Analytics 2
 - PapersStasko et al '08