Lecture 13: User Studies
Information Visualization
CPSC 533C, Fall 2007

Ghida Karam
UBC Computer Science
24 October 2007

Readings Covered
Ware, Appendix C: The Perceptual Evaluation of Visualization Techniques and Systems.

Further Readings
Task-Centered User Interface Design, Clayton Lewis and John Rieman, Chapters 0-5.

Ware: Evaluation Appendix
▶ perceptual evaluation of infovis techniques and systems
▶ empirical research methods applied to vis
▶ difficult to isolate evaluation to perception
▶ research method depends on research question and object under study

Psychophysics
▶ method of limits
▶ find limitations of human perceptions
▶ error detection methods
▶ find threshold of performance degradation
▶ staircase procedure to find threshold faster
▶ method of adjustment
▶ find optimal level of stimulus by letting subjects control the level

Cognitive Psychology
▶ repeating simple, but important tasks, and measure reaction time or error
  • Miller’s 7 +/- 2 short-term memory experiments
  • Fitts’ Law (target selection)
  • Hick’s Law (decision making given n choices)
▶ interference between channels
▶ multi-modal studies
  • using haptic feedback for interruption when the participants were visually (and cognitively) busy

Structural Analysis
▶ requirement analysis, task analysis
▶ structured interviews
  • can be used almost anywhere, for open-ended questions and answers
▶ rating/Likert scales
  • commonly used to solicit subjective feedback
  • ex: NASA-TLX (Task Load Index) to assess mental workload
  • “It is frustrating to use the interface”
  • Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree

Comparative User Studies
▶ study design: factors and levels
  ▶ factors
  • independent variables
  • ex: interface, task, participant demographics
  ▶ levels
  • number of variables in each factor
  • limited by length of study and number of participants

Comparative User Studies
▶ result analysis
▶ should know how to analyze the main results/hypotheses BEFORE study
▶ hypothesis testing
  • ex from Snap: Participants will be faster with a coordinated overview+detail display than with a display with the task requires reading details
  • pilots!
▶ should know the main results of the study BEFORE actual study

Evaluation Throughout Design Cycle
▶ user/task centered design cycle
  ▶ initial assessments
  ▶ iterative design process
  ▶ benchmarking
  ▶ deployment
▶ identify problems, go back to previous step

Initial Assessments
▶ what kind of problems are the system aiming to address?
  • analyze a large and complex dataset
  • who are your target users?
  • data analysis
  ▶ what are the tasks? what are the goals?
  • find trends and patterns in the data via exploratory analysis
  ▶ what are their current practices
  • statistical analysis
  ▶ why and how can visualization be useful?
  • visual spotting of trends and patterns
  ▶ talk to the users, and observe what they do
  ▶ task analysis

Iterative Design Process
▶ does your design address the users’ needs?
▶ can they use it?
▶ where are the usability problems?
▶ evaluate without users
  • cognitive walkthrough
  • action analysis
  • heuristics analysis
▶ evaluate with users
  • usability evaluations (think-aloud)
  • bottom-line measurements
  • example: snap paper experiment 1

Benchmarking
▶ how does your system compare to existing ones?
  • snap paper experiment 2
  • empirical, comparative studies
  ▶ ask specific questions
  ▶ compare an aspect of the system with specific tasks
  • Amar/Stasko task taxonomy paper
  ▶ quantitative, but limited
  • The Challenge of Information Visualization Evaluation, Catherine Plaisant, Proc. AVI 2004

Readings Covered
Comparative User Studies
▶ study design: within, or between?
  ▶ within
  • everybody does all the conditions
  • can lead to ordering effects
  • can account for individual differences and reduce noise
  • thus can be more powerful and require fewer participants
  • combinatorial explosion
  ▶ severe limits on number of conditions
  • possible workaround is multiple sessions
  ▶ between
  • divide participants into groups
  • each group does only some conditions

Comparative User Studies
▶ result analysis
▶ should know how to analyze the main results/hypotheses BEFORE study
▶ hypothesis testing analysis (using ANOVA or t-tests) tests how likely observed differences between groups are due to chance alone
  • ex: a p-value of 0.05 means there is a 5% probability the difference occurred by chance
▶ pilots!
▶ should know the main results of the study BEFORE actual study

Comparative User Studies
▶ measurements (dependent variables)
  ▶ performance indicators: task completion time, error rates, mouse movement
  ▶ subjective participant feedback: satisfaction ratings, closed-ended questions, interview
  ▶ observations: behaviors, signs of frustration
  ▶ number of participants
  ▶ depends on effect size and study design: power of experiment
  ▶ possible confounds?
  ▶ learning effect: did everybody use interfaces in a certain order?
  ▶ if so, are people faster because they are more practiced, or because of true interface effect?
**Deployment**
- how is the system used in the wild?
- how are people using it?
- does the system fit into existing work flow? environment?
- contextual studies, field studies

**Comparing Systems vs. Characterizing Usage**
- user/task centered design cycle:
  - initial assessments
  - iterative design process
  - benchmarking: head-to-head comparison
  - deployment
  - identify problems, go back to previous step
  - understanding/characterizing techniques
  - tease apart factors
  - when and how is technique appropriate
- line is blurry: intent

**Snap-Together Visualization: CMV**
- relation :: visualization
- tuple :: item
- primary key :: item ID
- join :: coordination

**Snap Usability Evaluation**
- 6 participants: 3 data analysts, 3 programmers
- census bureau: analysts + 1 programmer (expert?)
- CS students: 2 programmers (novice?)
- 3 tasks
- 2 construct to spec
- 1 open ended, "abstract thinking about coordination"
- measurements
- survey of background knowledge (data, tools)
- success at task
- learning time, time to completion

**Snap Usability Results**
- success, enthusiasm
  - possible confound from please-the-creator effect
- analyst/programmer differences
  - interface building as exploration vs. construction
  - analysts performed better
- snap usability problems
  - explicit overview of coordination setup may help
  - provide attribute lists instead of requiring access queries
  - window rearrangement timesink

**Snap CMV Formalism**
- one-to-one
  - linked selection across views
  - overview select → child load
  - linked scrolling across views
- one-to-many
  - parent select → child load
- architecture
  - independent modules linked via API
  - versus tightly coupled Improvise approach

**Snap User Study**
- hypothesis
  - participants will be faster with a coordinated overview+detail display than with an uncoordinated overview+detail display with the task requires reading details
- factors are and levels
  - interface: 3 levels
  - uncoordinated overview+detail
  - coordinated overview+detail
  - task: 9 levels
  - many browsing tasks, not grouped prior to study
  - closed-ended, with obvious correct answers
  - which task has the highest college degree
  - compare with open-ended usability task: "Please create a user-interface that will support users in efficiently performing the following task: to be able to quickly discover which states have high population and high Per Capita income, and examine their counties with the most employees"

**Snap User Study Design**
- within-subject
  - everybody worked on all interfaces/task combos
- counterbalanced between interfaces
  - 6 permutations to avoid ordering / learning effects
  - 3 groups x 6 permutations = 18 participants
  - need one task set (9) for each interface
  - tasks in each set need to be isomorphic
- 27 tasks per study per participant
  - 3 interfaces x 9 tasks
- time result analysis: hypothesis testing with ANOVA
  - 3 (interface) x 9 (task) within-subjects ANOVA to check for main effects of interface, or task, or interface/task interaction
  - ANOVA
    - (Analysis Of VAriance between groups)
    - commonly used statistics for factorial designs
    - tests difference between means of two or more groups
    - example use: two-way ANOVA to see if there is an effect of interface and task, or interaction between them

**Perceptual Scalability**
- what are perceptual/cognitive limits when screen-space constraints lifted?
  - 2 vs. 32 Mpixel display
  - macro/micro views
  - perceptually scalable
    - no increase in task completion times when normalize to amount of data

**Embedded Visualizations**
- design
  - 2 display sizes, between-subjects
  - (data size also increased proportionally)
  - 3 visualization designs, within
    - small multiples: bars
    - embedded graphs
    - embedded bars
  - 7 tasks, within
  - 42 tasks per participant
    - 3 vis x 7 tasks x 2 trials

**Critique**
- good example of usability vs. comparative study
  - Usability testing
    - improve product, design vs. prototype usable!
    - Discover knowledge (how are methods useful?)
    - How people use them
  - Usability evaluation
    - Do they work? (Are they usable?)
    - How well? (How well do they work?)
    - How long does it take? (How well do they work?)

**Perceptual Scalability**
- design
  - 2 display sizes, between-subjects
  - (data size also increased proportionally)
  - 3 visualization designs, within
    - small multiples: bars
    - embedded graphs
    - embedded bars
  - 7 tasks, within
  - 42 tasks per participant
    - 3 vis x 7 tasks x 2 trials

**Snap CMV Formalism**
- relation :: visualization
- tuple :: item
- primary key :: item ID
- join :: coordination

**Snap Usability Results**
- success, enthusiasm
  - possible confound from please-the-creator effect
- analyst/programmer differences
  - interface building as exploration vs. construction
  - analysts performed better
- snap usability problems
  - explicit overview of coordination setup may help
  - provide attribute lists instead of requiring access queries
  - window rearrangement timesink

**Snap Usability Evaluation**
- 6 participants: 3 data analysts, 3 programmers
- census bureau: analysts + 1 programmer (expert?)
- CS students: 2 programmers (novice?)
- 3 tasks
- 2 construct to spec
- 1 open ended, "abstract thinking about coordination"
- measurements
- survey of background knowledge (data, tools)
- success at task
- learning time, time to completion

**Snap CMV Formalism**
- relation :: visualization
- tuple :: item
- primary key :: item ID
- join :: coordination

**Snap Usability Results**
- success, enthusiasm
  - possible confound from please-the-creator effect
- analyst/programmer differences
  - interface building as exploration vs. construction
  - analysts performed better
- snap usability problems
  - explicit overview of coordination setup may help
  - provide attribute lists instead of requiring access queries
  - window rearrangement timesink

**Snap Usability Evaluation**
- 6 participants: 3 data analysts, 3 programmers
- census bureau: analysts + 1 programmer (expert?)
- CS students: 2 programmers (novice?)
- 3 tasks
- 2 construct to spec
- 1 open ended, "abstract thinking about coordination"
- measurements
- survey of background knowledge (data, tools)
- success at task
- learning time, time to completion

**Snap CMV Formalism**
- relation :: visualization
- tuple :: item
- primary key :: item ID
- join :: coordination

**Snap Usability Results**
- success, enthusiasm
  - possible confound from please-the-creator effect
- analyst/programmer differences
  - interface building as exploration vs. construction
  - analysts performed better
- snap usability problems
  - explicit overview of coordination setup may help
  - provide attribute lists instead of requiring access queries
  - window rearrangement timesink
### Results

- **20x** increase in data, but only **3x** increase in absolute task times

![Image](image1.png)

### Fisheye Multilevel Networks

- **2 interfaces** (fisheye, zoom)
- **2 tasks** (isomorphic)
- stages: find and repair
- within subjects, counterbalanced order
- **20 participants**
- data: **154 nodes, 39 clusters**
- **measurements**
  - completion time
  - number of zooms
  - success

### Lab Experiment

- sig effect of interface: fisheye faster
- but no differences with find subtask
- information visible in both displays
- solution quality differed: fisheye better
- local rerouting difficult in full-zoom

### Critique

- first study of macro/micro effects
- breaking new ground
- many possible followups
- physical navigation vs. virtual navigation

### Field Experiment

- 2 real control room operators
- response times similar
- no statistical analysis, too few subjects
- expressed preference for fisheye over full-zoom
  - (experimenter effect?)
- concerns about fisheye: missing details

### Coding Methodology

- interface
  - which interface used
  - whether picture/chart/graph
- usage (every utterance!)
  - goal
  - extract
    - quant/qual
    - goal-oriented/opportunist
    - integrated/unintegrated
  - brief-writing
    - quant/qual
    - QMM/vis/notes

### Results

- **sig difference** between vis used at CTA stages
  - charts to build QMM
  - images to verify/adjust QMM
  - all kinds during brief-writing
  - many others...

### Critique

- video coding is huge amount of work, but very illuminating
- untying complex story of real tool use
- methodology of CTA construction not discussed here
- often bottomup/topdown mix

### Credits

- Heidi Lam guest lecture