Evaluation, When and How

Tamara Munzner
University of British Columbia

Evaluation: How Much Evaluation Is Enough?
Panel, VIS13
Wrestling with the question

• in private: ponder on case-by-case basis
  - as an author
  - as a reviewer

• in public: series of 4 meta-papers

  “She generally gave herself very good advice, (though she very seldom followed it).”
  - Lewis Carroll

  - process and pitfalls Dagstuhl07
  - nested model of design and validation InfoVis09
    • NM revisited: blocks & guidelines BELIV12, Inf Vis Journal 13
  - design study methodology InfoVis12
Evaluation: broadly interpreted


problem domain:
observe target users using existing tools

data/task abstraction:

encoding/interaction technique:
justify design wrt alternatives

algorithm:
measure system time
analyze computational complexity

analyze results qualitatively
measure human time with lab experiment (“user study”)

observe target users post-deployment (“field study”)

measure adoption

• mismatch: cannot show technique good with system timings
• mismatch: cannot show abstraction good with lab study
When to evaluate?

“Begin at the beginning and go on till you come to the end; then stop.”
- Lewis Carroll

• at the end, to show you got it right: summative
• as you go, to make it better: formative
• iterate!

Victories and challenges: I

- evolving sophistication: the user study pendulum swings
- we’ve come a long way!
  - no user studies at all
  - a few dubious ones, lacking rigor
  - some good ones appear
  - rigorous studies are common
- but pushes to change culture often overshoot...
  - some reviewers expect all papers to have user studies
  - some authors do user studies without understanding why or how
User studies: tricky to do right

“If you don't know where you are going, any road will get you there.”
- Lewis Carroll

• many ways to go wrong that are not obvious from reading finished papers
  - ideal: collaborate with experts... several times!

• good resources
  - How to Design and Report Experiments, Andy Field & Graham Hole, SAGE
  - Experimental Human-Computer Interaction, Helen Purchase, Cambridge Univ Press
Victories and challenges: II

• significance testing with controlled experiments
  - we’ve moved beyond “my friends liked it”
  - new frontier: multiple regression for comparison
  - new frontier: thinking beyond time and error
    • qualitative vs quantitative
      - different axis from lab/field
    • BELIV workshops
      - 06 AVI, 08 CHI, 10 CHI, 12 VisWeek
Victories and challenges: III

- post-deployment studies with target users
  - we’ve moved beyond “I’m the only one who’s used it”
  - new frontier: post-adoption studies

- Seven Scenarios: only 5 out of 800!


- what happens after you get that first paper out?...
Of course...

• ...you should evaluate your work
  - use appropriate methods!

• ...you should **not** have a user study in every paper
  - avoid litmus test and cargo cult thinking