Visualization for Hackers: Why It’s Tricky, and Where to Start

Tamara Munzner
Department of Computer Science
University of British Columbia

Outline
- Introduction
  - What’s visual anyway?
- LiveRAC
  - Server log managed web hosting (with AT&T)
- Overview
  - Text: visual document mining for journalists (with Associated Press)
  - Big picture and wrapup

Defining visualization (vis)

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Why?

- Many analysis problems ill-specified, not clear what questions to ask in advance
  - Don’t need via when fully automatic solution exists and is trusted

Why have a human in the loop?

- Computer-based information systems provide visual representations to help people carry out tasks more effectively.

Visualization is reliable when there is a need to augment human capabilities rather than replace them with computational decision-making methods.

Why use an external representation?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- External representation: replace cognition with perception

Analysis framework: Four levels, three questions

- Domain situation
  - Who are the targets?
- Abstraction
  - Translate from specifics of domain to vocabulary of vis
  - What is shown? data abstraction
- Why is the user looking at it? Task abstraction
- Idiom
  - How is it shown?
  - Visual encoding: combinatorial explosion of choices
  - Add interaction: even bigger
- Data abstraction transformation: truly enormous
- Most possibilities ineffective for particular task/data combination
  - Implication: avoid random walk; be guided by principles
  - Analysis framework: scaffold to think systematically about design space
  - Ensure that consideration space encompasses full scope of possibilities
  - Improve chances that selected solution is good not mediocre
  - Today’s focus: abstractions and idioms, what why-how

LiveRAC

Interactive Visual Exploration of System Management Time-Series Data

LiiveRAC video

What: Data abstraction

- Multidimensional table: time series data
  - Key attributes
  - Time
  - 50,000 5-minute intervals over 6 months
  - Multiple levels of interest
  - Devices
  - Parameters
  - 20
  - CPU Usage: memory load, network traffic, alarms... Value attributes
  - Parameter value for device at time point
  - Qualitative
  - Device groups
  - Categorical

Trends

- All data
- Outliers
- Features
- Targets

What: Facet

- Facet: partition data into multiple views
- Juxtapose views side by side
- Same encoding, different data; small multiples

How: Facet

- Juxtapose linked views
- Inkblot highlighting
- Markers: line tracks across views

How: Juxtapose

- Juxtapose and coordinate views
- Share encoding: same/different

Why: Task abstraction

- Browse and correlate across combinations of parameter, device, time
- Correlate alarm attributes with other parameter attributes
- Find trends across groups of devices
- Summarize over different time intervals
- Identify devices at or beyond parameter thresholds
- Identify critical parameter values

Why: Tasks in domain language

- Interpret network environment status
- Report generation
- Capacity planning
- Event investigation/forensics
- Coordination between customers, engineering, ops

Outline

- Introduction
- What’s visual anyway?
- LiveRAC
  - Server log managed web hosting (with AT&T)
- Overview
  - Text: visual document mining for journalists (with Associated Press)
  - Big picture and wrapup
More Information

• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#hope14

• papers, videos, software, talks, courses
  http://www.cs.ubc.ca/~tmm/publications
  http://www.cs.ubc.ca/~tmm/research

  http://www.cs.ubc.ca/~tmm/vadbook

• acknowledgements
  – funding: AT&T, Knight Foundation, NSERC
  – talk feedback: Matt Brehmer

Visualization Analysis and Design