Lecture 7: Statistical Graphics

Information Visualization
CPSC 533C, Fall 2007

Tamara Munzner

UBC Computer Science

1 October 2007

Readings Covered

Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Chris Ahlberg and Ben Shneiderman, Proc SIGCHI '94, pages 313-317

Metric-Based Network Exploration and Multiscale Scatterplot. Yves Chiricota, Fabien Jourdan, Guy Melancon. Proc. InfoVis 04, pages 135-142.

Graph-Theoretic Scagnostics. Leland Wilkinson, Anushka Anand, and Robert Grossman. Proc InfoVis 05

The Visual Design and Control of Trellis Display. R. A. Becker, W. S. Cleveland, and M. J. Shyu Journal of Computational and Statistical Graphics, 5:123-155. (1996).

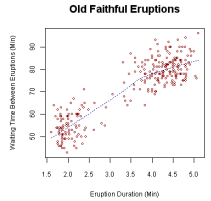
Multi-Scale Banking to 45 Degrees. Jeffrey Heer, Maneesh Agrawala. IEEE TVCG 12(5) (Proc. InfoVis 2006), Sep/Oct 2006, pages 701-708.

Statistical Graphics

- long history for paper-based views of data
  - springboard for infovis
- interacting with scatterplots
  - interactive dynamic queries
  - multiscale structure
  - matrix of scatterplots, level of indirection
  - linked views (more on this next time)
- ordering dot plots
- improving line charts

Scatterplots

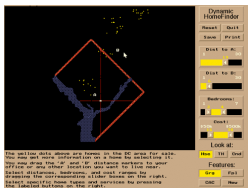
- encode two input variables with spatial position
- show positive/negative/no correllation between variables



[http://upload.wikimedia.org/wikipedia/commons/0/0f/Oldfaithful3.png]

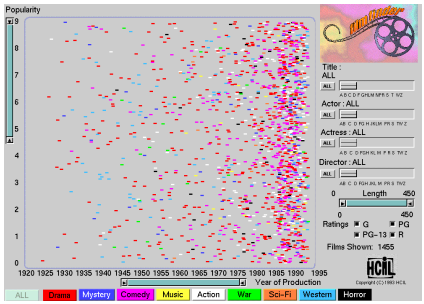
Dynamic Queries on Scatterplots

- tight coupling: immediate feedback after action
- starfield = interactive scatterplot
- dynamic queries as lightweight visual exploration
  - vs. composing SQL query



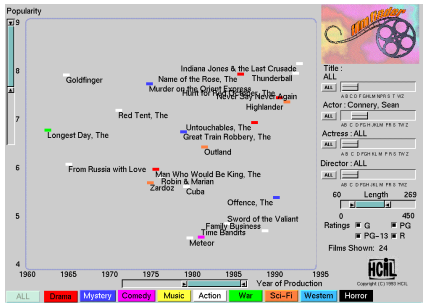
[Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Chris Ahlberg and Ben Shneiderman, Proc SIGCHI '94, p 313-317] [http://www.cs.umd.edu/hcil/pubs/screenshots/FilmFinder/]

FilmFinder



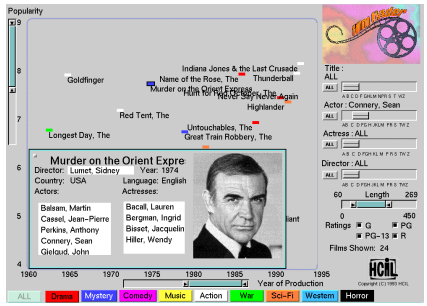
[Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Chris Ahlberg and Ben Shneiderman, Proc SIGCHI '94, p 313-317] [http://www.cs.umd.edu/hcil/pubs/screenshots/FilmFinder/]

FilmFinder



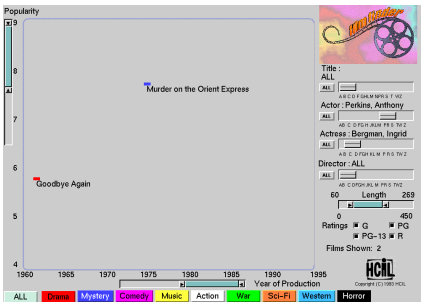
[Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Chris Ahlberg and Ben Shneiderman, Proc SIGCHI '94, p 313-317] [http://www.cs.umd.edu/hcil/pubs/screenshots/FilmFinder/]

FilmFinder



[Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Chris Ahlberg and Ben Shneiderman, Proc SIGCHI '94, p 313-317] [http://www.cs.umd.edu/hcil/pubs/screenshots/FilmFinder/]

FilmFinder



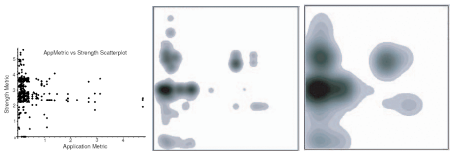
[Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Chris Ahlberg and Ben Shneiderman, Proc SIGCHI '94, p 313-317] [http://www.cs.umd.edu/hcil/pubs/screenshots/FilmFinder/]

Critique

- clear successes
  - fast, lightweight visual queries
  - details on demand
  - easy to use for novices
- more arguable: alphasliders
  - other techniques: data vis sliders, fisheye menus, speed-dependent automatic zooming

Multiscale Scatterplots

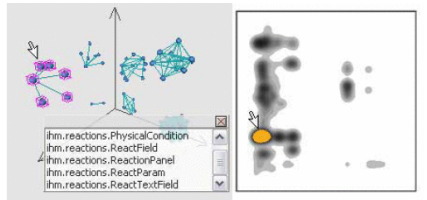
- blur shows structure at multiple scales
  - convolve with Gaussian
  - slider to control scale parameter interactively
- easily selectable regions in quantized image



[Metric-Based Network Exploration and Multiscale Scatterplot. Yves Chiricota, Fabien Jourdan, Guy Melancon. Proc. InfoVis 04]

Metric-Based Exploration: Software Engr

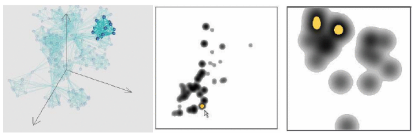
- linked views for metric-based exploration
  - graph view
  - axis 1: strength metric (topological graph structure)
  - axis 2: software engr metric (public methods)



[Metric-Based Network Exploration and Multiscale Scatterplot. Chiricota, Jourdan, and Melancon. Proc. InfoVis 04]

Metric-Based Exploration: IMDB

- axis 1: centrality, for locating cliques
  - axis 2: node degree, for size of clique
  - axis 2: clustering index



[Metric-Based Network Exploration and Multiscale Scatterplot. Chiricota, Jourdan, and Melancon. Proc. InfoVis 04]

Critique

- interesting followup to Wattenberg paper
  - exploiting perceptual mechanisms
- suitable for intermediate/expert analysis
  - abstraction might be difficult for novice use

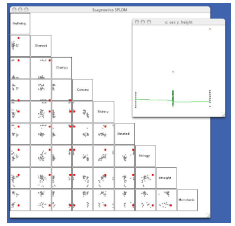
SPLOM: Scatterplot Matrix

- show all pairwise variable combos side by side
  - matrix size grows quadratically with variable count

[Graph-Theoretic Scagnostics. Wilkinson, Anand, and Grossman. Proc InfoVis 05.]

Graph-Theoretic Scagnostics

- reduce problem to constant size
  - overview matrix of 9 geometric metrics
- meta-SPLOM: each point represents scatterplot
  - detail on demand to see individual scatterplots



Graph-Theoretic Scagnostics. Leland Wilkinson, Anushka Anand, and Robert Grossman. Proc InfoVis 05.

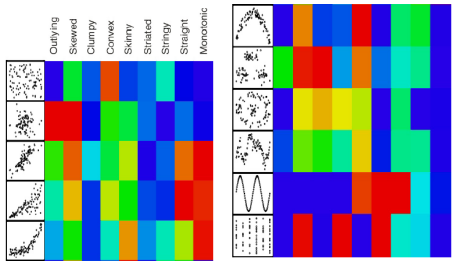
Measuring Scatterplots

- aspects and measures
  - outliers: outlying
  - shape: convex, skinny, stringy, straight
    - computed with convex hull, alpha hull, min span tree

  - trend: monotonic
  - density: skewed, clumpy
  - coherence: striated

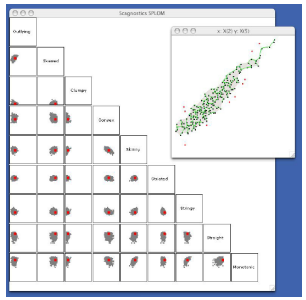
[Graph-Theoretic Scagnostics. Wilkinson, Anand, and Grossman. Proc InfoVis 05.]

Measuring Scatterplots



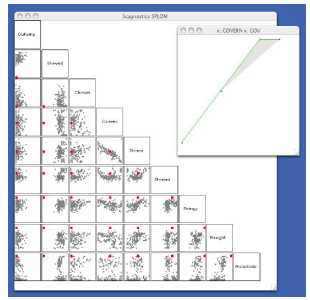
[Graph-Theoretic Scagnostics. Wilkinson, Anand, and Grossman. Proc InfoVis 05.]

Results



[Graph-Theoretic Scagnostics. Wilkinson, Anand, and Grossman. Proc InfoVis 05.]

Results



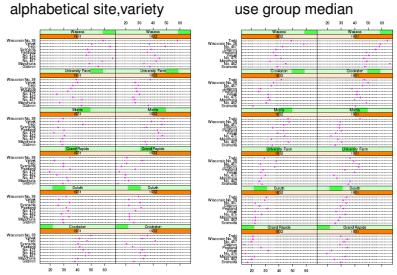
[Graph-Theoretic Scagnostics. Wilkinson, Anand, and Grossman. Proc InfoVis 05.]

Critique

- very powerful and elegant method
  - curse of dimensionality is hard problem
- abstraction level clearly appropriate for experts

Automatic Dotplot Ordering: Trellis

alphabetical site,variety | use group median



[The Visual Design and Control of Trellis Display. Becker, Cleveland, and Shyu. JCSG 5:123-155 1996]

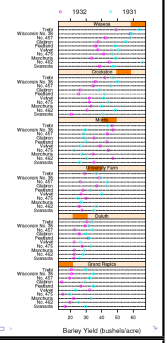
Trellis Structure

- conditioning/trellising: choose structure
  - pick how to subdivide into panels
  - pick x/y axes for indiv panels
  - explore space with different choices
    - multiple conditioning
- ordering
  - large-scale: between panels
  - small-scale: within panels
  - main-effects: sort by group median
    - derived space, from categorical to ordered

Confirming Hypothesis

- dataset error with Morris switched?
- old trellis: yield against variety given year/site
- new trellis: yield against site and year given variety
  - exploration suggested by previous main-effects ordering

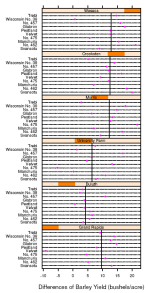
[The Visual Design and Control of Trellis Display. Becker, Cleveland, and Shyu. JCSG 5:123-155 1996]



Partial Residuals

- fixed dataset, Morris data switched
- explicitly show differences
  - take means into account
  - line is 10% trimmed mean (toss outliers)

[The Visual Design and Control of Trellis Display. Becker, Cleveland, and Shyu. JCSG 5:123-155 1996]

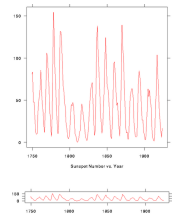


Critique

- careful attention to statistics and perception
- finding signals in noisy data
  - trends, outliers
- exploratory data analysis (EDA)
  - Tukey work fundamental, Cleveland continues

Banking to 45 Degrees

- mentioned but not explained in Trellis paper
  - previous work by Cleveland
- perceptual principle: most accurate angle judgement at 45 degrees
- pick aspect ratio (height/width) accordingly



[www.research.att.com/~rab/trellis/sunspot.html]

Multiscale Banking to 45

- frequency domain analysis
- find interesting regions at multiple scales

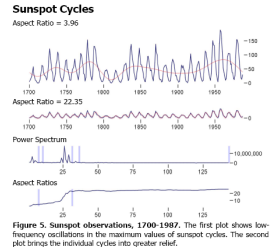
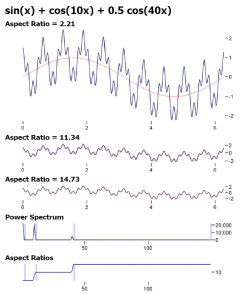


Figure 5. Sunspot observations, 1700-1987. The first plot shows low-frequency oscillations in the maximum values of sunspot cycles. The second plot brings the individual cycles into greater relief.

[Multi-Scale Banking to 45 Degrees. Heer and Agrawala, Proc InfoVis 2006 vis.berkeley.edu/papers/banking]

Choosing Aspect Ratios

- FFT the data, smooth by convolve with Gaussian
- find interesting spikes/ranges in power spectrum
- cull nearby regions if too similar, ensure overview shown
- create trend curves for each aspect ratio



Multiscale Banking to 45

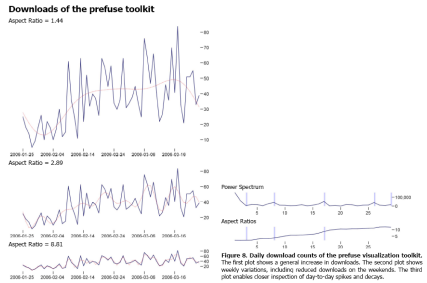


Figure 8. Daily download counts of the prefuse visualization toolkit. The first plot shows a general increase in downloads. The second plot shows weekly variations, including reduced downloads on the weekends. The third plot enables closer inspection of day-to-day spikes and decays.

[Multi-Scale Banking to 45 Degrees. Heer and Agrawala, Proc InfoVis 2006 vis.berkeley.edu/papers/banking]

Presentation Topics Due Oct 19

- pick three topics that you want
- optional: veto one of the three days
  - Nov 5 or Nov 7 or Nov 19
- send me email by Oct 19
  - Subject: 533 submit topics

Topics

- application domains
  - software viz
  - computer networks viz
  - db/datamine viz
  - cartographic viz
  - social networks viz
  - ...
- data domains
  - time-series data viz
  - text/document collection viz
  - tree/hierarchy visualization
  - graph drawing
  - high dimensional data
  - ...
- techniques/approaches
  - interaction
  - focus+context
  - navigation/zooming
  - glyphs
  - animation
  - brushing/linking
  - statistical graphics
  - ...
- other
  - frameworks/taxonomies
  - perception
  - evaluation
  - ....