Information Visualization: Principles, Promise, and Pragmatics
Marti Hearst

CHI 2003 Tutorial
Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up
Introduction

- Goals of Information Visualization
- Case Study: The Journey of the TreeMap
- Key Questions
What is Information Visualization?

Visualize: to form a mental image or vision of ...

Visualize: to imagine or remember as if actually seeing.

American Heritage dictionary, Concise Oxford dictionary
What is Information Visualization?

“Transformation of the symbolic into the geometric”
(McCormick et al., 1987)

“... finding the artificial memory that best supports our natural means of perception.”
(Bertin, 1983)

The depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system.
Information Visualization

• **Problem:**
  – HUGE Datasets: How to understand them?

• **Solution**
  – Take better advantage of human perceptual system
  – Convert information into a graphical representation.

• **Issues**
  – How to convert abstract information into graphical form?
  – Do visualizations do a better job than other methods?
Visualization
Success Stories

Images from yahoo.com
The Power of Visualization

1. Start out going Southwest on ELLSWORTH AVE Towards BROADWAY by turning right.
2. Turn RIGHT onto BROADWAY.
3. Turn RIGHT onto QUINCY ST.
4. Turn LEFT onto CAMBRIDGE ST.
5. Turn SLIGHT RIGHT onto MASSACHUSETTS AVE.
6. Turn RIGHT onto RUSSELL ST.
The Power of Visualization

The estimated travel time is 5 minutes for 2.16 miles of travel, total of 6 steps.

Directions | Elapsed Distance
---|---
1. Begin at **17 Ellsworth Ave** on **Ellsworth Ave** and go Southwest for 500 feet | 0.1
2. Turn right on **Broadway** and go North for 0.4 miles | 0.5
3. Turn right on **Quincy St** and go North for 200 feet | 0.5
4. Turn left on **Cambridge St** and go West for 0.3 miles | 0.8
5. Bear right on **Massachusetts Ave, Mass Ave, RT-2A** and go North for 1.2 miles | 2.0
6. Turn right on **Russell St** and go Northeast for 1000 feet to **77 Russell St** | 2.2
Visualization Success Story

Mystery: what is causing a cholera epidemic in London in 1854?
Visualization Success Story

Illustration of John Snow’s deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Horizontal lines indicate location of deaths.

Illustration of John Snow’s deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Horizontal lines indicate location of deaths.
Purposes of Information Visualization

To help:
- Explore
- Calculate
- Communicate
- Decorate
Two Different Primary Goals: Two Different Types of Viz

**Explore/Calculate**
- Analyze
- Reason about Information

**Communicate**
- Explain
- Make Decisions
- Reason about Information
Goals of Information Visualization

More specifically, visualization should:

- Make large datasets coherent
  (Present huge amounts of information compactly)
- Present information from various viewpoints
- Present information at several levels of detail
  (from overviews to fine structure)
- Support visual comparisons
- Tell stories about the data
Why Visualization?

Use the eye for pattern recognition; people are good at
  scanning
  recognizing
  remembering images

Graphical elements facilitate comparisons via
  length
  shape
  orientation
  texture
Animation shows changes across time
Color helps make distinctions
Aesthetics make the process appealing
The Need for Critical Analysis

- We see many creative ideas, but they often fail in practice

- The hard part: how to apply it judiciously
  - Inventors usually do not accurately predict how their invention will be used

- This tutorial will emphasize
  - Getting past the coolness factor
  - Examining usability studies
Case Study: The Journey of the TreeMap

- The TreeMap (Johnson & Shneiderman ‘91)
- Idea:
  - Show a hierarchy as a 2D layout
  - Fill up the space with rectangles representing objects
  - Size on screen indicates relative size of underlying objects.
Early Treemap Applied to File System
Treemap Problems

• Too disorderly
  – What does adjacency mean?
  – Aspect ratios uncontrolled leads to lots of skinny boxes that clutter

• Color not used appropriately
  – In fact, is meaningless here

• Wrong application
  – Don’t need all this to just see the largest files in the OS
Successful Application of Treemaps

• Think more about the use
  – Break into meaningful groups
  – Fix these into a useful aspect ratio

• Use visual properties properly
  – Use color to distinguish meaningfully
    • Use only two colors:
      – Can then distinguish one thing from another
    • When exact numbers aren’t very important

• Provide excellent interactivity
  – Access to the real data
  – Makes it into a useful tool
TreeMaps in Action

http://www.smartmoney.com/maps

A Good Use of TreeMaps and Interactivity

www.smartmoney.com/marketmap
Treemaps in Peets site
Analysis vs. Communication

- MarketMap’s use of TreeMaps allows for sophisticated analysis
- Peets’ use of TreeMaps is more for presentation and communication
- This is a key contrast
Open Issues

• **Does visualization help?**
  - The jury is still out
  - Still supplemental at best for text collections
    • A correlation with spatial ability
    • Learning effects: with practice ability on visual display begins to equal that of text

• **Does visualization sell?**
  - Jury is still out on this one too!

• This is a *hot* area!  More ideas will appear!
Key Questions to Ask about a Viz

1. What does it teach/show/elucidate?
2. What is the key contribution?
3. What are some compelling, useful examples?
4. Could it have been done more simply?
5. Have there been usability studies done? What do they show?
What we are *not* covering

- Scientific visualization
- Statistics
- Cartography (maps)
- Education
- Games
- Computer graphics in general
- Computational geometry
Agenda

• Introduction
• Visual Principles
• What Works?
• Visualization in Analysis & Problem Solving
• Visualizing Documents & Search
• Comparing Visualization Techniques
• Design Exercise
• Wrap-Up
Visual Principles
Visual Principles

- Types of Graphs
- Pre-attentive Properties
- Relative Expressiveness of Visual Cues
- Visual Illusions
- Tufte’s notions
  - Graphical Excellence
  - Data-Ink Ratio Maximization
  - How to Lie with Visualization
References for Visual Principles

- Kosslyn: Types of Visual Representations
- Lohse et al: How do people perceive common graphic displays
- Bertin, MacKinlay: Perceptual properties and visual features
- Tufte/Wainer: How to mislead with graphs
A Graph is: (Kosslyn)

- A visual display that illustrates one or more relationships among entities
- A shorthand way to present information
- Allows a trend, pattern, or comparison to be easily apprehended
Types of Symbolic Displays (Kosslyn 89)

- Graphs
- Charts
- Maps
- Diagrams
Types of Symbolic Displays

- **Graphs**
  - at least two scales required
  - values associated by a symmetric “paired with” relation
  - Examples: scatter-plot, bar-chart, layer-graph
Types of Symbolic Displays

Charts
- discrete relations among discrete entities
- structure relates entities to one another
- lines and relative position serve as links

Examples:
family tree
flow chart
network diagram
Types of Symbolic Displays

- **Maps**
  - internal relations determined (in part) by the spatial relations of what is pictured
  - labels paired with locations

Examples:
map of census data
topographic maps

From www.thehighsierra.com
Types of Symbolic Displays

Diagrams
- schematic pictures of objects or entities
- parts are symbolic (unlike photographs)
  - how-to illustrations
  - figures in a manual

Anatomy of a Graph  (Kosslyn 89)

- **Framework**
  - sets the stage
  - kinds of measurements, scale, ...

- **Content**
  - marks
  - point symbols, lines, areas, bars, ...

- **Labels**
  - title, axes, tic marks, ...
Basic Types of Data

- **Nominal (qualitative)**
  - (no inherent order)
  - city names, types of diseases, ...

- **Ordinal (qualitative)**
  - (ordered, but not at measurable intervals)
  - first, second, third, ...
  - cold, warm, hot

- **Interval (quantitative)**
  - list of integers or reals
Common Graph Types

- URL vs. # of accesses
- Length of access vs. # of accesses
- Length of page vs. # of accesses
- Length of page vs. days
- # of accesses vs. url 1 to url 7
Combining Data Types in Graphs

Examples?

<table>
<thead>
<tr>
<th>Nominal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Ordinal</td>
</tr>
<tr>
<td>Nominal</td>
<td>Interval</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Ordinal</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Interval</td>
</tr>
<tr>
<td>Interval</td>
<td>Interval</td>
</tr>
</tbody>
</table>
Scatter Plots

- **Qualitatively determine if variables**
  - are highly correlated
    - linear mapping between horizontal & vertical axes
  - have low correlation
    - spherical, rectangular, or irregular distributions
  - have a nonlinear relationship
    - a curvature in the pattern of plotted points

- **Place points of interest in context**
  - color representing special entities
When to use which type?

• Line graph
  – x-axis requires quantitative variable
  – Variables have contiguous values
  – familiar/conventional ordering among ordinals
• Bar graph
  – comparison of relative point values
• Scatter plot
  – convey overall impression of relationship between two variables
• Pie Chart?
  – Emphasizing differences in proportion among a few numbers
Classifying Visual Representations

Lohse, G L; Biolsi, K; Walker, N and H H Rueter,
A Classification of Visual Representations
CACM, Vol. 37, No. 12, pp 36-49, 1994

Participants sorted 60 items into categories

Other participants assigned labels from Likert scales

Experimenters clustered the results various ways.
Subset of Example Visual Representations
From Lohse et al. 94

1. soil triangle
2. missile crisis
3. USA tornadoes
4. microscope
5. circular tree diagram
6. auto repair records
Likert Scales
(and percentage of variance explained)

16.0  emphasizes whole – parts
11.3  spatial – nonspatial
10.6  static structure – dynamic structure
10.5  continuous – discrete
10.3  attractive – unattractive
10.1  nontemporal – temporal
   9.9  concrete – abstract
9.6   hard to understand – easy
9.5   nonnumeric – numeric
2.2   conveys a lot of info – conveys little
Experimentally Motivated Classification (Lohse et al. 94)

- Graphs
- Tables (numerical)
- Tables (graphical)
- Charts (time)
- Charts (network)
- Diagrams (structure)
- Diagrams (network)
- Maps
- Cartograms
- Icons
- Pictures
Interesting Findings
Lohse et al. 94

• Photorealistic images were least informative
  – Echos results in icon studies – better to use less complex, more schematic images
• Graphs and tables are the most self-similar categories
  – Results in the literature comparing these are inconclusive
• Cartograms were hard to understand
  – Echos other results – better to put points into a framed rectangle to aid spatial perception
• Temporal data more difficult to show than cyclic data
  – Recommend using animation for temporal data
Visual Properties

- Preattentive Processing
- Accuracy of Interpretation of Visual Properties
- Illusions and the Relation to Graphical Integrity

All Preattentive Processing figures from Healey 97
http://www.csc.ncsu.edu/faculty/healey/PP/PP.html
Preattentive Processing

• A limited set of visual properties are processed preattentively
  – (without need for focusing attention).
• This is important for design of visualizations
  – what can be perceived immediately
  – what properties are good discriminators
  – what can mislead viewers
Example: Color Selection

Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in color.
Example: Shape Selection

Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in form (curvature)
Pre-attentive Processing

• < 200 - 250ms qualifies as pre-attentive
  – eye movements take at least 200ms
  – yet certain processing can be done very quickly,
    implying low-level processing in parallel

• If a decision takes a fixed amount of time
  regardless of the number of distractors, it is
  considered to be preattentive.
Example: Conjunction of Features

Viewer *cannot* rapidly and accurately determine whether the target (red circle) is present or absent when target has two or more features, each of which are present in the distractors. Viewer must search sequentially.

All Preattentive Processing figures from Healey 97
http://www.csc.ncsu.edu/faculty/healey/PP/PP.html
Example: Emergent Features

Target has a unique feature with respect to distractors (open sides) and so the group can be detected preattentively.
Example: Emergent Features

Target does not have a unique feature with respect to distractors and so the group cannot be detected preattentively.
Asymmetric and Graded Preattentive Properties

- Some properties are asymmetric
  - a sloped line among vertical lines is preattentive
  - a vertical line among sloped ones is not
- Some properties have a gradation
  - some more easily discriminated among than others
Use Grouping of Well-Chosen Shapes for Displaying Multivariate Data
# Preattentive Visual Properties

*(Healey 97)*

<table>
<thead>
<tr>
<th>Property</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>Triesman &amp; Gormican [1988]</td>
</tr>
<tr>
<td>width</td>
<td>Julesz [1985]</td>
</tr>
<tr>
<td>size</td>
<td>Triesman &amp; Gelade [1980]</td>
</tr>
<tr>
<td>curvature</td>
<td>Triesman &amp; Gormican [1988]</td>
</tr>
<tr>
<td>number</td>
<td>Julesz [1985]; Trick &amp; Pylyshyn [1994]</td>
</tr>
<tr>
<td>terminators</td>
<td>Julesz &amp; Bergen [1983]</td>
</tr>
<tr>
<td>intersection</td>
<td>Julesz &amp; Bergen [1983]</td>
</tr>
<tr>
<td>closure</td>
<td>Enns [1986]; Triesman &amp; Souther [1985]</td>
</tr>
<tr>
<td></td>
<td>Kawai et al. [1995]; Bauer et al. [1996]</td>
</tr>
<tr>
<td>intensity</td>
<td>Beck et al. [1983]; Triesman &amp; Gormican [1988]</td>
</tr>
<tr>
<td>flicker</td>
<td>Julesz [1971]</td>
</tr>
<tr>
<td>direction of motion</td>
<td>Nakayama &amp; Silverman [1986]; Driver &amp; McLeod [1992]</td>
</tr>
<tr>
<td>binocular lustre</td>
<td>Wolfe &amp; Franzel [1988]</td>
</tr>
<tr>
<td>stereoscopic depth</td>
<td>Nakayama &amp; Silverman [1986]</td>
</tr>
<tr>
<td>3-D depth cues</td>
<td>Enns [1990]</td>
</tr>
<tr>
<td>lighting direction</td>
<td>Enns [1990]</td>
</tr>
</tbody>
</table>
Gestalt Properties

- **Gestalt**: form or configuration
- Idea: forms or patterns transcend the stimuli used to create them.
  - Why do patterns emerge?
  - Under what circumstances?

Why perceive pairs vs. triplets?
Gestalt Laws of Perceptual Organization (Kaufman 74)

- **Figure and Ground**
  - Escher illustrations are good examples
  - Vase/Face contrast
- **Subjective Contour**
More Gestalt Laws

- **Law of Proximity**
  - Stimulus elements that are close together will be perceived as a group

- **Law of Similarity**
  - like the preattentive processing examples

- **Law of Common Fate**
  - like preattentive motion property
    - move a subset of objects among similar ones and they will be perceived as a group
Which Properties are Appropriate for Which Information Types?
Accuracy Ranking of Quantitative Perceptual Tasks
Estimated; only pairwise comparisons have been validated
(Mackinlay 88 from Cleveland & McGill)
Interpretations of Visual Properties

Some properties can be discriminated more accurately but don’t have intrinsic meaning
(Senay & Ingatious 97, Kosslyn, others)

- Density (Greyscale)
  Darker -> More

- Size / Length / Area
  Larger -> More

- Position
  Leftmost -> first, Topmost -> first

- Hue
  ??? no intrinsic meaning

- Slope
  ??? no intrinsic meaning
### Ranking of Applicability of Properties for Different Data Types

(Mackinlay 88, Not Empirically Verified)

<table>
<thead>
<tr>
<th>QUANTITATIVE</th>
<th>ORDINAL</th>
<th>NOMINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Saturation</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density</td>
</tr>
<tr>
<td>Density</td>
<td>Containment</td>
<td>Color Saturation</td>
</tr>
<tr>
<td>Color Saturation</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Color Hue</td>
<td>Angle</td>
<td>Length</td>
</tr>
</tbody>
</table>
Color Schemes

Order these (low -> hi)

- Purple
- Green
- Orange
- Yellow
- Red
Color Schemes

- Gray scale
- Full spectral scale
- Single sequence part spectral scale
- Single sequence single hue scale
- Double-ended multiple hue scale
Color Purposes

- Call attention to specific items
- Distinguish between classes of items
  - Increases the number of dimensions for encoding
- Increase the appeal of the visualization
Using Color

- **Proceed with caution**
  - Less is more
  - Representing magnitude is tricky
- **Examples**
  - Red-orange-yellow-white
    - Works for costs
    - Maybe because people are very experienced at reasoning shrewdly according to cost
  - Green-light green-light brown-dark brown-grey-white
    - Works for atlases
  - Grayscale is unambiguous but has limited range
Visual Illusions

- People don’t perceive length, area, angle, brightness they way they “should”.
- Some illusions have been reclassified as systematic perceptual errors
  - e.g., brightness contrasts (grey square on white background vs. on black background)
  - partly due to increase in our understanding of the relevant parts of the visual system
- Nevertheless, the visual system does some really unexpected things.
Illusions of Linear Extent

- Mueller-Lyon (off by 25-30%)
- Horizontal-Vertical
Illusions of Area

- Delboeuf Illusion

- Height of 4-story building overestimated by approximately 25%
What are good guidelines for Infoviz?

• **Use graphics appropriately**
  – Don’t use images gratuitously
  – Don’t lie with graphics!
    • Link to original data
  – Don’t conflate area with other information
    • E.g., use area in map to imply amount

• **Make it interactive (feedback)**
  – Brushing and linking
  – Multiple views
  – Overview + details

• **Match mental models**
Tufte

- Principles of Graphical Excellence
  - Graphical excellence is
    - the well-designed presentation of interesting data – a matter of substance, of statistics, and of design
    - consists of complex ideas communicated with clarity, precision and efficiency
    - is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
    - requires telling the truth about the data.
Tufte’s Notion of Data Ink Maximization

• What is the main idea?
  – draw viewers attention to the substance of the graphic
  – the role of redundancy
  – principles of editing and redesign

• What’s wrong with this? What is he really getting at?
Tufte Principle

Maximize the data-ink ratio:

\[
\text{Data-ink ratio} = \frac{\text{data ink}}{\text{total ink used in graphic}}
\]

Avoid “chart junk”
Tufte Principles

• Use multifunctioning graphical elements
• Use small multiples
• Show mechanism, process, dynamics, and causality
• High data density
  – Number of items/area of graphic
  – This is controversial
    • White space thought to contribute to good visual design
    • Tufte’s book itself has lots of white space
Tufte’s Graphical Integrity

- Some lapses intentional, some not
- Lie Factor = \( \frac{\text{size of effect in graph}}{\text{size of effect in data}} \)
- Misleading uses of area
- Misleading uses of perspective
- Leaving out important context
- Lack of taste and aesthetics
A common example of a high lie factor occurs when both dimensions of a two-dimensional figure are made proportional to the same data, so that the size of the figure is proportional to the square of the data, for instance.

<table>
<thead>
<tr>
<th>Year</th>
<th>Books circulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>100</td>
</tr>
<tr>
<td>2002</td>
<td>141</td>
</tr>
<tr>
<td>2003</td>
<td>200</td>
</tr>
</tbody>
</table>

An example of a low lie factor can be seen in the "Cones" custom chart format in Microsoft Excel.

The heights of the (truncated) cones are proportional to the data, but their areas on the screen and their apparent volumes make the larger data values seem relatively small.
How to Exaggerate with Graphs from Tufte ‘83

"Lie factor" = 2.8
How to Exaggerate with Graphs
from Tufte ‘83
Howard Wainer
How to Display Data Badly (Video)

http://www.dartmouth.edu/~chance/ChanceLecture/AudioVideo.html
Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up
Promising Techniques
Promising Techniques & Approaches

- **Perceptual Techniques**
  - Animation
  - Grouping / Gestalt principles
  - Using size to indicate quantity
  - **Color** for Accent, Distinction, Selection
    - NOT FOR QUANTITY!!!!

- **General Approaches**
  - Standard Techniques
    - Graphs, bar charts, tables
  - Brushing and Linking
  - Providing Multiple Views and Models
  - Aesthetics!
Standard Techniques

- **It’s often hard to beat:**
  - Line graphs, bar charts
  - Scatterplots (or Scatterplot Matrix)
  - Tables

- **A Darwinian view of visualizations:**
  - Only the fittest survive
  - We are in a period of great experimentation; eventually it will be clear what works and what dies out.

- **A bright spot:**
  - Enhancing the old techniques with interactivity
  - Example: Spotfire
    - Adds interactivity, color highlighting, zooming to scatterplots
  - Example: TableLens / Eureka
    - Adds interactivity and length cues to tables
Spotfire: Integrating Interaction with Scatterplots

Ahlberg & Shneiderman, Color plate 1. The FilmFinder.
Spotfire/IVEE: Integrating Interaction with Scatterplots

Ahlberg & Shneiderman, Color plate 2. Categories have been selected, the displayed is zoomed into 1960-85 and popularity 4-9, and Sean Connery has been selected.
Brushing and Linking

- Interactive technique
  - Highlighting
  - Brushing and Linking
- At least two things must be linked together to allow for brushing
  - select a subset of points
  - see the role played by this subset of points in one or more other views
- Example systems
  - Graham Will’s EDV system
  - Ahlberg & Sheiderman’s IVEE (Spotfire)
Linking types of assist behavior to position played (from Eick & Wills 95)
Baseball data: Scatterplots and histograms and bars (from Eick & Wills 95)

- How long in majors
- Select high salaries
- Avg assists vs avg putouts (fielding ability)
- Avg career HRs vs avg career hits (batting ability)
- Distribution of positions played
What was learned from interaction with this baseball data?

- Seems impossible to earn a high salary in the first three years
- High salaried players have a bimodal distribution (peaking around 7 & 13 yrs)
- Hits/Year a better indicator of salary than HR/Year
- High paid outlier with low HR and medium hits/year. Reason: person is player-coach
- There seem to be two differentiated groups in the put-outs/assists category (but not correlated with salary) Why?
Animation

- "The quality or condition of being alive, active, spirited, or vigorous" (dictionary.com)
- "A dynamic visual statement that evolves through movement or change in the display"
- "... creating the illusion of change by rapidly displaying a series of single frames" (Roncarelli 1988).
We Use Animation to...

- Tell stories / scenarios: cartoons
- Illustrate dynamic process / simulation
- Create a character / an agent
- Navigate through virtual spaces
- Draw attention
- Delight
Cartoon Animation Principles

- Chang & Unger ‘93
- Solidity (squash and stretch)
  - Solid drawing
  - Motion blur
  - Dissolves
- Exaggeration
  - Anticipation
  - Follow through
- Reinforcement
  - Slow in and slow out
  - Arcs
  - Follow through
Why Cartoon-Style Animation?

- Cartoons’ theatricality is powerful in communicating to the user.
- Cartoons can make UI engage the user into its world.
- The medium of cartoon animation is like that of graphic computers.
Application using Animation: Gnutellavision

- **Visualization of Peer-to-Peer Network**
  - Hosts (with color for status and size for number of files)
  - Nodes with closer network distance from focus on inner rings
  - Queries shown; can trace queries

- **Gnutellavision as exploratory tool**
  - Very few hosts share many files
  - Uneven propagation of queries
  - Qualitative assessment of queries (simple)
Layout - Illustration
Animation in Gnutellavision

Goal of animation is to help maintain context of nodes and general orientation of user during refocus

- **Transition Paths**
  - Linear interpolation of polar coordinates
  - Node moves in arc not straight line
  - Moves along circle if not changing levels (like great circles on earth)
  - Spirals in or out to next ring
Animation (continued)

- **Transition constraints**
  - Orientation of transition to minimize rotational travel
  - (Move former parent away from new focus in same orientation)
  - Avoid cross-over of edges
  - (to allow users to keep track of which is which)

- **Animation timing**
  - Slow in Slow out timing (allows users to better track movement)
Transition Constraint - Orientation
Transition Constraint - Order
Usability Testing

- In general, users appreciated the subtleties added to the general method when the number of nodes increased.

- Perhaps the most interesting result is that most people preferred rectangular movement for the small graph and polar coordinate movement for the large one.

<table>
<thead>
<tr>
<th>Overall Preference of Users</th>
<th>No Features</th>
<th>All Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Graph</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Large Graph</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>
Hyperbolic Tree

- Also uses animation
- Tree-based layout; leaves stretch to infinity
- Only a few labels can be seen at a time
Issues

- **Displaying text**
  - The size of the text
    - Works good for small things like directories
    - Not so good for URLs
- **Only a portion of the data can be seen in the focus at one time**
- **Only works for certain types of data - Hierarchical**
- **Not clear if it is actually useful for anything.**
Animating Algorithms

- Kehoe, Stasko, and Taylor, “Rethinking Evaluation of Algorithm Animations as Learning Aids”

- Why previous studies present no benefits:
  - No or limited benefits from particular animations
  - Benefits are not captured in measurements
  - Design of experiments hides the benefits

- Methods for this study:
  - Combination of qualitative & quantitative
  - More flexible setting
  - Metrics: score for each type of questions, time used, usage of materials, qualitative data from observations & interviews
Findings

- Value of animation is more apparent in interactive situations
- Most useful to learn procedural operations
- Makes subject more accessible & less intimidating → increase motivation
What Isn’t Working?

The existing studies indicate that we don’t yet know how to make the following work well for every-day tasks:

- Pan-and-Zoom
- 3D Navigation
- Node-and-link representations of concept spaces
Zoom, Overview + Detail

- An exception, possibly:
Overview + Detail

Overview + Detail


- A study on integrating Overview + Detail on a Map search task
  - Incorporating panning & zooming as well.
  - They note that panning & zooming does not do well in most studies.

- Results seem to be
  - Subjectively, users prefer to have a linked overview
  - But they aren’t necessarily faster or more effective using it
  - Well-constructed representation of the underlying data may be more important.

- More research needed as each study seems to turn up different results, sensitive to underlying test set.
Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up
Problem Solving
Problem Solving

- **A Detective Tool for Multidimensional Data**
  - Inselberg on using Parallel Coordinates

- **Analyzing Web Clickstream Data**
  - Brainerd & Becker, Waterson et al.

- **Information Visualization for Pattern Detection**
  - Carlis & Konstan on Periodic Data

- **Visualization vs. Analysis**
  - Comments by Wesley Johnson of Chevron
Multidimensional Detective


Figure 1: The full dataset consisting of 473 batches
A Detective Story


Inselberg’s Principles for analysis using visualizations:
1. Do not let the picture scare you
2. Understand your objectives
   – Use them to obtain visual cues
3. Carefully scrutinize the picture
4. Test your assumptions, especially the “I am really sure of’s”
5. You can’t be unlucky all the time!
A Detective Story


• The Dataset:
  – Production data for 473 batches of a VLSI chip
  – 16 process parameters
  – The yield: % of produced chips that are useful
    • X1
  – The quality of the produced chips (speed)
    • X2
  – 10 types of defects (zero defects shown at top)
    • X3 ... X12
  – 4 physical parameters
    • X13 ... X16

• The Objective:
  – Raise the yield (X1) and maintain high quality (X2)
Multidimensional Detective


Do Not Let the Picture Scare You!!

Figure 1: The full dataset consisting of 473 batches
Multidimensional Detective

- Each line represents the values for one batch of chips
- This figure shows what happens when only those batches with both high $X_1$ and high $X_2$ are chosen
- Notice the separation in values at $X_{15}$
- Also, some batches with few $X_3$ defects are not in this high-yield/high-quality group.

Figure 2: The batches high in Yield, $X_1$, and Quality, $X_2$. 
Multidimensional Detective

- Now look for batches which have *nearly* zero defects.
  - For 9 out of 10 defect categories
- Most of these have low yields
- Surprising because we know from first diagram that some defects are ok.
- Go back to first diagram, looking at defect categories
- Notice that X6 behaves differently than the rest
- Allow two defects, where one defect in X6
- This results in the very best batch appearing
Multidimensional Detective

- Fig 5 and 6 show that high yield batches don’t have non-zero values for defects of type X3 and X6
  - Don’t believe your assumptions …
- Looking now at X15 we see the separation is important
  - Lower values of this property end up in the better yield batches
Automated Analysis

A. Inselberg, Automated Knowledge Discovery using Parallel Coordinates, INFOVIS ‘99
Case Study: E-Commerce Clickstream Visualization

- Brainerd & Becker, IEEE Infovis 2001
- Aggregate nodes using an icon (e.g. all the checkout pages)
- Edges represent transitions
  - Wider means more transitions
Customer Segments

• **Collect**
  – Clickstream
  – Purchase history
  – Demographic data

• **Associates customer data with their clickstream**

• **Different color for each customer segment**
Layout

- Aggregation based on file system path
Initial Findings

- Gender shopping differences
Initial Findings (cont)

- Checkout process analysis
- Newsletter hurting sales
WebQuilt

Interactive, zoomable directed graph
- Nodes = web pages
- Edges = aggregate traffic between pages

Directed graph

- **Nodes:** visited pages
  - Color marks entry and exit nodes
- **Arrows:** traversed links
  - Thicker: more heavily traversed
  - Color
    - **Red/yellow:** Time spend before clicking
    - **Blue:** optimal path chosen by designer
http://pda.edmunds.com

Where Smart Car
Buyers Start

Edmunds2Go!

- Vehicle Prices & Reviews
- Dealer Locator
- Auto Tools

About Us | Help

Pilot Usability Study

- Edmunds.com PDA web site
- Visor Handspring equipped with a OmniSky wireless modem
- 10 users asked to find...
  - Anti-lock brake information on the latest Nissan Sentra model
  - The Nissan dealer closest to them.
In the Lab vs. Out in the Wild

Comparing in-lab usability testing with WebQuilt remote usability testing

• 5 users were tested in the lab
• 5 were given the device and asked to perform the task at their convenience
• All task directions, demographic data, and follow up questionnaire data was presented and collected in web forms as part of the WebQuilt testing framework.
### Findings

<table>
<thead>
<tr>
<th>Browser</th>
<th>Device</th>
</tr>
</thead>
</table>
| - Interact before load (3)  
- No forward button (2) | - Difficulty with input in questionnaire (3)  
- **Difficulty scrolling** (2)  
- Device errors unrelated to testing (1)  
- Tried writing on screen (0) |

<table>
<thead>
<tr>
<th>Site Design</th>
<th>Test Design</th>
</tr>
</thead>
</table>
| - Falsely completed task (4)  
- Long download times (4)  
- Ping-pong behavior (3)  
- Interact before load (3)  
- **Too much scrolling** (2)  
- Save address functionality not clear (1)  
- **Back button navigation** (0)  
- Would like more features (0)  
- Finds site useful (0) | - Falsely completed task (4)  
- **Difficulty remembering task description** (3)  
- Difficulty with input in questionnaire (3)  
- Questionnaire wording problems (3)  
- Forgot how to end task (1)  
- Confusing task description (1) |
Findings

• WebQuilt methodology is promising for uncovering site design related issues.
• 1/3 of the issues were device or browser related.
  • Browser and device issues can not be captured automatically with WebQuilt unless they cause an interaction with the server
  • Can be revealed via the questionnaire data.
Visualization for Analysis

- Carlis & Konstan, UIST 1998

- **Problem: data that is both periodic and serial**
  - Time students spend on different activities
  - Tree growth patterns
    - Time: which year
    - Period: yearly
  - Multi-day races such as the Tour de France
  - Calendars arbitrarily wrap around at end of month
  - Octaves in music

- **How to find patterns along both dimensions?**
Analyzing Complex Periodic Data

Figure 1. A Spiral of Archimedes. (For color figures see the electronic proceedings or www.cs.unm.edu/~carls)

Figure 2. An indented spiral, with spokes, showing monthly consumption percentages for Baphia Capparidifolia during the period 1980 – 1988.

Carlis & Konstan, UIST 1998.
Analyzing Complex Periodic Data

• Consumption values for each month appear as spikes
• Each food has its own color
• Boundary line (in black) shows when season begins/ends

Figure 4. A spiral display of year-month consumption percentages for 12 highly consumed foods during the period 1980 – 1988. Rotated and zoomed in to show one season and boundary lines.

Carlis & Konstan, UIST 1998.
Carlis & Konstan, UIST 1998.
Visualization vs. Analysis?

- Applications to data mining and data discovery.
- Wesley Johnson ’02:
  - Visualization tools are helpful for exploring hunches and presenting results
    - Examples: scatterplots
  - They are the WRONG primary tool when the goal is to find a good classifier model in a complex situation.
  - Need:
    - Solid insight into the domain and problem
    - Tools that visualize several alternative models.
    - Emphasize “model visualization” rather than “data visualization”
 Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up
Visualizing Documents and Search
Documents and Search

• Why Visualize Text?
• Why Text is Tough
• Visualizing Concept Spaces
  – Clusters
  – Category Hierarchies
• Visualizing Retrieval Results
• Usability Study Meta-Analysis
Why Visualize Text?

• **To help with Information Retrieval**
  - give an overview of a collection
  - show user what aspects of their interests are present in a collection
  - help user understand why documents retrieved as a result of a query

• **Text Data Mining**
  - Mainly clustering & nodes-and-links

• **Software Engineering**
  - not really text, but has some similar properties
Why Text is Tough

- **Text is not pre-attentive**
- **Text consists of abstract concepts**
  - which are difficult to visualize
- **Text represents similar concepts in many different ways**
  - space ship, flying saucer, UFO, figment of imagination
- **Text has very high dimensionality**
  - Tens or hundreds of thousands of features
  - Many subsets can be combined together
Why Text is Tough

As the man walks the cavorting dog, thoughts arrive unbidden of the previous spring, so unlike this one, in which walking was marching and dogs were baleful sentinals outside unjust halls.

How do we visualize this?
Why Text is Tough

- Abstract concepts are difficult to visualize
- Combinations of abstract concepts are even more difficult to visualize
  - time
  - shades of meaning
  - social and psychological concepts
  - causal relationships
Why Text is Tough

- Language only hints at meaning
- Most meaning of text lies within our minds and common understanding
  - “How much is that doggy in the window?”
    - how much: social system of barter and trade (not the size of the dog)
    - “doggy” implies childlike, plaintive, probably cannot do the purchasing on their own
    - “in the window” implies behind a store window, not really inside a window, requires notion of window shopping
Why Text is Easy

- **Text is highly redundant**
  - When you have lots of it
  - Pretty much any simple technique can pull out phrases that seem to characterize a document

- **Instant summary:**
  - Extract the most frequent words from a text
  - Remove the most common English words

- **People are very good at attributing meaning to lists of otherwise unrelated words**
Guess the Text:

10 PEOPLE
10 ALL
9 STATES
9 LAWS
8 NEW
7 RIGHT
7 GEORGE
6 WILLIAM
6 THOMAS
6 JOHN
6 GOVERNMENT
5 TIME
5 POWERS
5 COLONIES
4 LARGE
4 INDEPENDENT
4 FREE
4 DECLARATION
4 ASSENT
3 WORLD
3 WAR
3 USURPATIONS
3 UNITED
3 SEAS
3 RIGHTS
Visualization of Text Collections

• How to summarize the contents of hundreds, thousands, tens of thousands of texts?
• Many have proposed clustering the words and showing points of light in a 2D or 3D space.
• Examples
  – Showing docs/collections as a word space
  – Showing retrieval results as points in word space
Alice's Adventures in Wonderland

TextArc.org (Bradford Paley)
Winter storm dumps more snow on weary region
severe weather, trouble, general weather, winter

NEW YORK (UPI) — Winter - weary metropolitan area residents backed down the hatches Tuesday and readied their rock salt to another snowstorm threatened to dump more than a foot of snow over the region.

The storm swept over northern New Jersey and southeastern New York including New York City and Long Island. Moderate to heavy snow spread across the entire area, with snow falling at a rate of two inches to three inches an hour in northern suburbs, the city and western Long Island.

By Wednesday the region, already blanketed with snow, turned to bitter cold making travel difficult.

Weather Satellite Photo, Feb. 8, 1994

NEW YORK (AP) — The weather satellite photo taken at 5:30 a.m. EST, Tuesday, February 8, 1994 shows cloud cover over much of the country.

Galaxy of News
Rennison 95
Example: Themescapes (Wise et al. 95)
Scatter Plot of Clusters (Chen et al. 97)
Clustering for Collection Overviews

- **Two main steps**
  - cluster the documents according to the words they have in common
  - map the cluster representation onto a (interactive) 2D or 3D representation

- **Since text has tens of thousands of features**
  - the mapping to 2D loses a tremendous amount of information
  - only very coarse themes are detected
<table>
<thead>
<tr>
<th>Cluster 1 Size: 8</th>
<th>key army war francis spangle banner air song scott word poem british</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Star–Spangled Banner, The</td>
</tr>
<tr>
<td></td>
<td>Key, Francis Scott</td>
</tr>
<tr>
<td></td>
<td>Fort McHenry</td>
</tr>
<tr>
<td></td>
<td>Arnold, Henry Harley</td>
</tr>
<tr>
<td></td>
<td>Milbank, Author</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 2 Size: 68</th>
<th>film play career win television role record award york popular stage play</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Burstyn, Ellen</td>
</tr>
<tr>
<td></td>
<td>Stanwyck, Barbara</td>
</tr>
<tr>
<td></td>
<td>Berle, Milton</td>
</tr>
<tr>
<td></td>
<td>Zukor, Adolph</td>
</tr>
<tr>
<td></td>
<td>Paulette, Tallulah</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 3 Size: 97</th>
<th>bright magnitude cluster constellation line type contain period spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>star</td>
</tr>
<tr>
<td></td>
<td>Galaxy, The</td>
</tr>
<tr>
<td></td>
<td>extragalactic systems</td>
</tr>
<tr>
<td></td>
<td>interstellar matter</td>
</tr>
<tr>
<td></td>
<td>cluster star</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 4 Size: 67</th>
<th>astronomer observatory astronomy position measure celestial telescope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>astronomy and astrophysics</td>
</tr>
<tr>
<td></td>
<td>astrometry</td>
</tr>
<tr>
<td></td>
<td>Agena</td>
</tr>
<tr>
<td></td>
<td>astronomical catalogs and atlases</td>
</tr>
<tr>
<td></td>
<td>Homak-Garcia, William</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster 5 Size: 10</th>
<th>family specie flower animal arm plant shape leaf brittle tube foot horn</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>blazing star</td>
</tr>
<tr>
<td></td>
<td>brittle star</td>
</tr>
<tr>
<td></td>
<td>bishop’s–cap</td>
</tr>
<tr>
<td></td>
<td>feather star</td>
</tr>
</tbody>
</table>
How Useful is Collection Cluster Visualization for Search?

Three studies find negative results
Study 1


• This study compared
  – a system with 2D graphical clusters
  – a system with 3D graphical clusters
  – a system that shows textual clusters

• Novice users

• Only textual clusters were helpful (and they were difficult to use well)
Study 2: Kohonen Feature Maps

H. Chen, A. Houston, R. Sewell, and B. Schatz, *JASIS 49(7)*

- **Comparison: Kohonen Map and Yahoo**
- **Task:**
  - “Window shop” for interesting home page
  - Repeat with other interface
- **Results:**
  - Starting with map could repeat in Yahoo (8/11)
  - Starting with Yahoo unable to repeat in map (2/14)
Study 2 (cont.)

• Participants liked:
  – Correspondence of region size to # documents
  – Overview (but also wanted zoom)
  – Ease of jumping from one topic to another
  – Multiple routes to topics
  – Use of category and subcategory labels
Study 2 (cont.)

- **Participants wanted:**
  - hierarchical organization
  - other ordering of concepts (alphabetical)
  - integration of browsing and search
  - correspondence of color to meaning
  - more meaningful labels
  - labels at same level of abstraction
  - fit more labels in the given space
  - combined keyword and category search
  - multiple category assignment (sports+entertain)
Study 3: NIRVE

NIRVE Interface by Cugini et al. 96. Each rectangle is a cluster. Larger clusters closer to the “pole”. Similar clusters near one another. Opening a cluster causes a projection that shows the titles.
Study 3


- This study compared:
  - 3D graphical clusters
  - 2D graphical clusters
  - textual clusters
- 15 participants, between-subject design
- Tasks
  - Locate a particular document
  - Locate and mark a particular document
  - Locate a previously marked document
  - Locate all clusters that discuss some topic
  - List more frequently represented topics
Study 3

- **Results (time to locate targets)**
  - Text clusters fastest
  - 2D next
  - 3D last
  - With practice (6 sessions) 2D neared text results; 3D still slower
  - Computer experts were just as fast with 3D
- **Certain tasks equally fast with 2D & text**
  - Find particular cluster
  - Find an already-marked document
- **But anything involving text (e.g., find title) much faster with text.**
  - Spatial location rotated, so users lost context
- **Helpful viz features**
  - Color coding (helped text too)
  - Relative vertical locations
Summary: Visualizing Clusters

- Huge 2D maps may be inappropriate focus for information retrieval
  - cannot see what the documents are about
  - space is difficult to browse for IR purposes
  - (tough to visualize abstract concepts)
- Perhaps more suited for pattern discovery and gist-like overviews
IR Infovis Meta-Analysis

(Empirical studies of information visualization: a meta-analysis, Chen & Yu IJHCS 53(5), 2000)

• Goal
  – Find invariant underlying relations suggested collectively by empirical findings from many different studies

• Procedure
  – Examine the literature of empirical infoviz studies
    • 35 studies between 1991 and 2000
    • 27 focused on information retrieval tasks
    • But due to wide differences in the conduct of the studies and the reporting of statistics, could use only 6 studies
IR Infovis Meta-Analysis

(Empirical studies of information visualization: a meta-analysis, Chen & Yu IJHCS 53(5),2000)

• Conclusions:
  - IR Infoviz studies not reported in a standard format
  - Individual cognitive differences had the largest effect
    • Especially on accuracy
    • Somewhat on efficiency
  - Holding cognitive abilities constant, users did better with simpler visual-spatial interfaces
  - The combined effect of visualization is not statistically significant
So What Works?


- Color highlighting of query terms in results listings
- Sorting of search results according to important criteria (date, author)
- Grouping of results according to well-organized category labels.
  - Cha-cha
  - Flamenco
- Only if highly accurate:
  - Spelling correction/suggestions
  - Simple relevance feedback (more-like-this)
  - Certain types of term expansion
- Note: most don’t benefit from visualization!
Cha-Cha

Teoma: appears to combine categories and clusters
(this version before it was bought by askjeeves)
Teoma: Now in prime time
Cat-a-Cone

Marti Hearst and Chandu Karadi, *Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy*.

*Proceedings of the 20th Annual International ACM/SIGIR Conference*

Philadelphia, PA, July 1997
Better to reduce the viz

- Flamenco – allows users to steer through the category space
- Uses
  - Dynamically-generated hypertext
  - Color for distinguishing and grouping
  - Careful layout and font choices
- Focused first on the users’ needs
Flamenco Image Search

Refine your search further within these categories:

Media
- aquatint (3), drawing (18), drypoint (2), etching (29), lithograph (12), mezzotint (14), painting (1)
- photograph (1), print (10), wood engraving (4), more...

Date: 19th century
- 1800 - 1809 (3), 1810 - 1819 (8), 1820 - 1829 (7), 1830 - 1839 (6), 1840 - 1849 (2), 1850 - 1859 (6), 1860 - 1869 (8), 1870 - 1879 (7), 1880 - 1889 (9), 1890 - 1899 (10), more...

Nature: heavens > cloud
- all items, within current results

89 items (grouped by media)

aquatint
- Distant View of...
  - Daniell
  - 1817

Verlassen (Aband...
- Klinger
- 1884

Village au Bord ...
- Henier
- 19th century

drawing
- A Lady at a Shrine
  - Anonymous
- A View of the No...
  - Eustache
- Alsation Landscape
  - Anonymous
- Boar Avatar of Y...
  - Anonymous

view ungrouped items

start a new search
Using Thumbnails to Search the Web


Design Goals
- Enhance features that help the user decide whether document is relevant to their query
  - Emphasize text that is relevant to query
    - Text callouts
  - Enlarge (make readable) text that might be helpful in assessing page
    - Enlarge headers
Text and Image Summaries

- **Text summaries**
  - Lots of abstract, semantic information

- **Image summaries (plain thumbnails)**
  - Layout, genre information
  - Gist extraction faster than with text

- **Benefits are complementary**

- **Create textually-enhanced thumbnails that leverage the advantages of both text summaries and plain thumbnails**
Putting Callouts in a Separate Visual Layer

- Transparency
- Occlusion

Junctions indicate the occurrence of these events.
Design Issues:

- **Color Management**
  - Problems: Callouts need to be both readable and draw attention
  - Solution: Desaturate the background image, and use a visual search model to choose appropriate colors
  - Colors look like those in highlighter pens

- **Resizing of Text**
  - Problem: We want to make certain text elements readable, but not necessarily draw attention to them
  - Solution: Modify the HTML before rendering the thumbnail
Examples
Tasks

- **Criteria: tasks that**...
  - Are representative of common queries
  - Have result sets with different characteristics
  - Vary in the number of correct answers

- **4 types of tasks**
  - **Picture:** “Find a picture of a giraffe in the wild.”
  - **Homepage:** “Find Kern Holoman’s homepage.”
  - **Side-effects:** “Find at least three side effects of halcion.”
  - **E-commerce:** “Find an e-commerce site where you can buy a DVD player. Identify the price in dollars.”
Conditions

- **Text summary**
  - Page title
  - Extracted text with query terms in bold
  - URL
- **Plain thumbnail**
- **Enhanced thumbnail**
  - Readable H1, H2 tags
  - Highlighted callouts of query terms
  - Reduced contrast level in thumbnail

*The Lycaem -- GHB*

.....seems to be without serious *side effects." His almost off-hand...
...recovery with no long-term *side effects is universal." They...

www.algonet.se/~spot/arch/texts/ghbfqa.html
Collections of Summaries

- 100 results in random order

Approximately same number of each summary type on a page
Method

18 questions, with 100 query results each
Entire process took about 75 minutes

• Procedure
  - 6 practice tasks
  - 3 questions for each of the 4 task types
    • e.g., each participant would do one E-commerce question using text, one E-commerce question using plain thumbnails, and one E-commerce question using enhanced thumbnails
  - Questions blocked by type of summary
  - WebLogger recorded user actions during browsing
  - Semi-structured interview

• Participants
  - 12 members of the PARC community
Results

• **Average total search times, by task:**
  - Picture: 61 secs
  - Homepage: 80 secs
  - E-commerce: 64 secs
  - Side effects: 128 secs

• **Results pooled across all tasks:**
  - Subjects searched 20 seconds faster with enhanced thumbnails than with plain
  - Subjects searched 30 seconds faster with enhanced thumbnails than with text summaries
  - Mean search time overall was 83 seconds
Results

![Normalized total search time (s)](image)

- **Picture**
- **Homepage**
- **E-commerce**
- **Side-effects**

Legend:
- **text**
- **plain**
- **enhanced**
Results: User Responses

- Participants preferred enhanced thumbnails
  - 7/12 preferred overall
  - 5/12 preferred for certain task types

- Enhanced thumbnails are intuitive and less work than text or plain thumbnails
  - One subject said searching for information with text summaries did not seem hard until he used the enhanced thumbnails.

- Many participants reported using genre information, cues from the callouts, the relationship between search terms, etc.
Agenda

• Introduction
• Visual Principles
• What Works?
• Visualization in Analysis & Problem Solving
• Visualizing Documents & Search
• Comparing Visualization Techniques
• Design Exercise
• Wrap-Up
Comparing Approaches
Comparing 3 Commercial Systems

Alfred Kobsa, An Empirical Comparison of Three Commercial Information Visualization Systems, INFOVIS'01.

Figure 1a: A screenshot from Gizmo, which shows how a user might solve the question "Did males cheat more on their girlfriends than females on their boyfriends?" (an actual question used in the experiment). After grouping the attribute "Gender" and sorting the column "Did you cheat?", a user can compare the number of "Yes" entries and then find that more females than males indicated having cheated. (One male and one female gave no answers.)

Figure 1b: This screenshot shows one possible way for solving the same problem in InfoZoom, specifically in its overview mode. After clicking at, and thereby zooming into, the "Yes" entries in the attribute "Did you cheat?", users can see from the length of the bars in the Gender category that females indicated more frequently having cheated than males.

Figure 2: Spotlens' geographical representation of heavy metal concentrations through a scatterplot diagram.
Comparing 3 Commercial Systems

Eureka (InXight)

Figure 1a. A screenshot from Eureka that shows how a user might solve the question “Did males cheat more on their girlfriends than females on their boyfriends?” (an actual question used in the experiment). After grouping the attribute “Gender” and sorting the column “Did you cheat?,” a user can compare the number of “Yes” entries and thus find that more females than males indicated having cheated. (One male and one female gave no answers.)
Comparing 3 Commercial Systems

InfoZoom (HumanIT)

Figure 1b: This screenshot shows one possible way for solving the same problem in InfoZoom, specifically in its overview mode. After clicking at, and thereby zooming into, the “Yes” entries in the attribute “Did you cheat?”, users can see from the length of the bars in the Gender category that females indicated more frequently having cheated than males.
Comparing 3 Commercial Systems
Infozoom Overview

- Presents data in three different views.
  - Wide view shows data set in a table format.
  - Compressed view packs the data set horizontally to fit the window width.
  - Overview mode has all attributes in ascending or descending order and independent of each other.
**InfoZoom Overview View**

![InfoZoom Overview View](image)

The InfoZoom Overview View is a tool for visualizing and analyzing large datasets. In this example, the view is focused on film attributes, including the title, year, length, subject, actor, actress, director, popularity, and awards. The view is interactive, allowing users to explore different attributes and filter data based on various criteria.

### Attributes Overview

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>A, B, C, D</td>
</tr>
<tr>
<td>Year</td>
<td>1986, 1987,</td>
</tr>
<tr>
<td>Length</td>
<td>96, 91, 94, 95, 96, 97</td>
</tr>
<tr>
<td>Subject</td>
<td>Action, Comedy, Drama</td>
</tr>
<tr>
<td>Popularity</td>
<td>2, 3, 4, 5, 6, 7, 8</td>
</tr>
<tr>
<td>Awards</td>
<td>No, Yes</td>
</tr>
</tbody>
</table>
InfoZoom Overview View
### InfoZoom Compressed Table View

<table>
<thead>
<tr>
<th>Column</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key</td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>Company Short Name</td>
<td></td>
</tr>
<tr>
<td>Company Name</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>Structure of Corporation</td>
<td></td>
</tr>
<tr>
<td>CEO</td>
<td></td>
</tr>
<tr>
<td>CEO (Photo)</td>
<td></td>
</tr>
<tr>
<td>Upper Management</td>
<td></td>
</tr>
<tr>
<td>Director of Staff</td>
<td></td>
</tr>
<tr>
<td>Director of Staff (Photo)</td>
<td></td>
</tr>
<tr>
<td>Division Head</td>
<td></td>
</tr>
<tr>
<td>Type of Company Relation</td>
<td></td>
</tr>
<tr>
<td>Markets</td>
<td></td>
</tr>
<tr>
<td>Lines of Business</td>
<td></td>
</tr>
<tr>
<td>Parent Company</td>
<td></td>
</tr>
</tbody>
</table>
### InfoZoom Wide Table View

**Table Contents:**
- Title
- Year
- Length
- Subject
- Actor
- Actress
- Director
- Popularity
- Awards

<table>
<thead>
<tr>
<th>Title</th>
<th>Year</th>
<th>Length</th>
<th>Subject</th>
<th>Actor</th>
<th>Actress</th>
<th>Director</th>
<th>Popularity</th>
<th>Awards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild at Heart</td>
<td>1990</td>
<td>125</td>
<td>Drama</td>
<td>Cage, Nicolas</td>
<td>De Palma</td>
<td>Lynch, David</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Goodbye Again</td>
<td>1991</td>
<td>120</td>
<td>Drama</td>
<td>Perkins, Anthony</td>
<td>Bergman</td>
<td>Litvak, Anatole</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Hunt for Red October, The</td>
<td>1990</td>
<td>135</td>
<td>Drama</td>
<td>Connelly, Sean</td>
<td>Kristin</td>
<td>McTiernan, J.</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Terminator 2</td>
<td>1991</td>
<td>136</td>
<td>Action</td>
<td>Schwarzenegger A.</td>
<td>Hamilton</td>
<td>Cameron, J.</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>John Clee Cee on How to Irritate People</td>
<td>1993</td>
<td>65</td>
<td>Comedy</td>
<td>Clee Cee, John</td>
<td>Manesse</td>
<td>Racette, Francine</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Au Revou Ies Enfan</td>
<td>1987</td>
<td>103</td>
<td>Drama</td>
<td>De Palma</td>
<td>Gaspard</td>
<td>Racette, Francine</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>The Ballad of Narayama</td>
<td>1983</td>
<td>128</td>
<td>Drama</td>
<td>De Palma</td>
<td>Gaspard</td>
<td>Racette, Francine</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Cyrano De Bergerac</td>
<td>1999</td>
<td>131</td>
<td>Drama</td>
<td>De Palma</td>
<td>Gaspard</td>
<td>Racette, Francine</td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>
Datasets

- Multidimensional data: three databases were used
  - Anonymized data from a web based dating service (60 records, 27 variables)
  - Technical data of cars sold in 1970 – 82 (406 records, 10 variables)
  - Data on the concentration of heavy metals in Sweden (2298 records, 14 variables)
Sample Questions

- Do more women than men want their partners to have a higher education?
- What proportion of the men live in California?
- Do all people who think the bar is a good place to meet a mate also believe in love at first site?
- Do heavier cars have more horsepower?
- Which manufacturer produced the most cars in 1980?
- Is there a relationship between the displacement and acceleration of a vehicle?
Experiment Design

- The experimenters generated 26 tasks from all three data sets.
- 83 participants. Between-subjects design.
- Each was given one visualization system and all three data sets.
- Type of visualization system was the independent variable between them.
- 30 mins were given to solve the tasks of each data set i.e 26 tasks in 90 mins.
Overall Results

• Mean task completion times:
  • Infozoom users: 80 secs
  • Spotfire users: 107 secs
  • Eureka users: 110 secs

• Answer correctness:
  • Infozoom users: 68%
  • Spotfire users: 75%
  • Eureka users: 71%

• Not a time-error tradeoff
  • Spotfire more accurate only 6 questions
Eureka - problems

• Hidden labels: Labels are vertically aligned, max 20 dimensions

• 3+ Attributes: Problems with queries involving three or more attributes

• Correlation problems: Some participants had trouble answering questions correctly that involved correlations between two attributes.
Spotfire - problems

• Cognitive setup costs: Takes participants considerable time to decide on the right representation and to correctly set the coordinates and parameters.

• Biased by scatterplot default: Though powerful, many problems cannot be solved (well) with it.
Infozoom - problems

- Erroneous Correlations
  - Overview mode has all attributes sorted independent of each other

- Narrow row height in compressed view

- Participants did not use row expansion and scatterplot charting function which shows correlations more accurately
Geographic Questions

• Spotfire should have done better on these
  • Which part of the country has the most copper
  • Is there a relationship between the concentration of vanadin and that of zinc?
  • Is there a low-level chrome area that is high in vanadin
  • Spotfire was only better only for the last question (out of 6 geographic ones)
Discussion

• Many studies of this kind use relatively simple tasks that mirror the strengths of the system
  • Find the one object with the maximum value for a property
  • Count how many of certain attributes there are
• This study looked at more complex, realistic, and varied questions.
Discussion

• Success of a visualization system depends on many factors:
  • Properties supplied
    • Spotfire doesn’t visualize as many dimensions simultaneously
  • Operations
    • Zooming easy in InfoZoom; allows for drill-down as well
    • Zooming in Eureka causes context to be lost
    • Column view in Eureka makes labels hard to see
Information Exploration “Shootout”

- http://ivpr.cs.uml.edu/shootout/about.html
- Data Mining Applications
- One component focuses on visualization
Comparing Tree Views

- T. Barlow and P. Neville, Comparison of 2D Visualizations of Hierarchies, INFOVIS’01.

- **Problem**
  - Organization Chart is de facto standard for visualizing decision trees. Is there a better compact view of the tree for the overview window?

- **Solution**
  - Two usability studies to determine which tree works best.
Goal: Compact View of Tools

Figure 1. Example of compact view in data mining

T. Barlow and P. Neville, Comparison of 2D Visualizations of Hierarchies, INFOVIS’01.
Decision Trees

- Each split constitutes a rule or variable in predictive model
- Begin Splitting into nodes
- Often hundreds of leaves
Decision Trees – What makes a good visualization

- **Uses**
  - For novice-helps them understand models
  - Experts-initial evaluation of decisions tree without looking at models

- **Criteria for usability in study**
  - Ease of Interpretation of Topology (Parent Child Sibling relations)
  - Comparison of Node Size
  - User preference
Different views examined in study

Org Chart  Tree Ring Icicle Plot  TreeMap

Figure 2. Different views of the same tree
Usability Test 1:

- **Users:**
  - 15 colleagues familiar with org chart but not others

- **Tasks**
  - Is the tree binary or n-ary?
  - Is the tree balanced or unbalanced?
  - Find deepest common ancestor of two nodes
  - Number of levels?
  - Find three largest leaves (excluding org chart)

- **Data:** Created 8 trees for analysis

- **Study Design**
  - Randomized order of tasks
  - 4X5 design (almost)
  - Timed task from appearance on screen until spacebar tap
Results

• **Response Time**
  - TreeMap slowest; no statistical difference between others

• **Response Accuracy**
  - No significant difference

• **User Preference**
  - Prefer icicle map and org chart (faster)
  - Dislike tree map
Discussion

- Org chart served as benchmark
- Icicle plot favored amongst others
  - Hypothesis: Same left to right / top to bottom structure
- TreeRing did well
- TreeMap suffered from poor accuracy
  - Offset of rectangles required because of off (which is needed for selection)
Usability Test II: Tree implementation

- **Three views:**
  - TreeMap eliminated from this round

- **Tasks**
  - Node Description
    - Four versions – select those nodes or leaves that meet certain criteria
  - Node Analysis:
    - Memorize a highlighted node – find again after tree redrawn in different position
Results

- Tree rings slower for description but fast and accurate for memory tasks
- Perhaps due to unique geometric forms / spatial clues
Conclusions

- TreeMap not useful for this type of task
- Org Chart/Icicle seem to be best overall
- TreeRing has merits for certain tasks

- Icicle chosen for implementation
  - Best design considering Org Chart could not be used for node size tasks
- However:
  - Didn’t seem to actually do tests on trees as large as the ones they describe as typical of datamining
Visualizing Conversations
Text-Based Chat

Katesmiles1 enters
You tell Horse_99 me too
Horse_99 says Real.
Speci_Man_98 says Where you from Kim?
Soapbox_7 leaves, heading for the Gen-X Love #19
Horse_99 says On here!
Horse_99 says Lets go private and find out.
Muta4 leaves heading for another room
Muta4 leaves
Horse_99 says Sure.
Speci_Man_98 says Ever been to new York?

Katesmiles1 enters
You tell Horse_99 me too
Horse_99 says Real.
Speci_Man_98 says Where you from Kim?
Soapbox_7 leaves, heading for the Gen-X Love #19
Soapbox_7 says
godess_oflove leaves, heading for the Gen-X Love #19
Kim_24_98 says connecticut
Horse_99 says How old are you Mika?
Sycam leaves
Kim_24_98 says where are you from
Muta4 says 16
You tell Horse_99 are you talking to me when you say Mike?
Speci_Man_98 says Paducah, ky
Horse_99 says 24m/a
Horse_99 says Yes
You say 26
Kim_24_98 says cool, I'm 25!
Horse_99 says On here!
Horse_99 says Lets go private and find out.
Muta4 leaves heading for another room
Muta4 leaves
Horse_99 says Sure.
Speci_Man_98 says Ever been to new York?
Chat Circles

Fernanda Viegas and Judith Donath, Chat Circles, Proceedings of CHI'99.
Chat Circles

• “Chat Circles is a graphical interface for synchronous communication that uses abstract shapes to convey identity and activity.”
• Each participant appears as a colored circle, which is accompanied by the user name
• Location of circles will also identify participants (important for many users having similar colors associated)
• Participants’ circles become larger when posting occurs (circle adapts to text length)
• Circle appears bright when posting occurs
• Circles of inactive users fade in the background
Chat Circles – Conversational Groupings

• There is only ONE room in Chat Circles
• Groupings are achieved by moving closer to other participants
• At any time, a participant can view all other participants
• A participant can also detect interesting conversations in different areas of the room by looking at how many circles are gathered and how often circles become larger
• Overview panel in Chat Circles II nice example of focus + context
Chat Circles History
History Log Patterns

+ Easy to see “lurkers”
+ Sequence and size of messages quickly visible
- Not very scalable
History Log Patterns

+/- User-centric: only 1 point of view represented

- Impossible to see all the text at once – requires individual mouse rollovers
- Easy to see “out of range” conversations – but why would you want to?
Agenda

• Introduction
• Visual Principles
• What Works?
• Visualization in Analysis & Problem Solving
• Visualizing Documents & Search
• Comparing Visualization Techniques
• Design Exercise
• Wrap-Up
Design Exercise
Design Exercise

- **BreakingStory**
  (Reffel, Fitzpatrick, Ayedelott SIMS final project, at CHI 2003)
  - Create an application that supplies a visualization for trends over time in web-based news. The primary purpose is to provide an overview, but it should also be possible to view text from individual news sources on specific days. Its goal is to inform, inspire, and enlighten, and also to make people want to look more deeply at the news.
What is BreakingStory?

BreakingStory is a tool to help you explore online news. Curious about when a phrase was first mentioned? Wondering if references to a current event vary by geographic region? Here you can find out!

How do you use it?

You can use this site to search for words or phrases that have appeared in online news. The results will be displayed in one or more charts that show the history of references.

**Single Chart View** allows you to examine references from one geographic area or news site.

**Multiple Chart View** shows a number of small charts next to each other, allowing you to compare references from different geographic areas or news sites.

You can also view the **full text** of the front pages that matched your search.

Don't know where to begin? Try looking at an example: World Aids Day

Want More?

- [Help](#) - suggestions, advice for using BreakingStory
- [Site List](#) - the news sites we collect and their locations
- [FAQ](#) - the who what when where how of BreakingStory
- [Project Page](#) - detailed information on the project
- [E-mail us](#) - we want to hear your comments, really!
Front page hits for:

oil  peace

Selected date range:

05 March 2003 to 04 April 2003 (31 days total)
Front page hits:

Average references per page (total references)
- oil: 0.26 (1794 total references in 1305 pages)
- peace: 0.22 (1494 total references in 1163 pages)

Selected date range:

17 August 2002 to 04 April 2003 (231 days total)

World > North America
Another Approach: ThemeRiver

Wrap-up: Guidelines for Success
Key Questions to Ask about a Viz

1. Is it for analysis or presentation?
2. What does it teach/show/elucidate?
3. What is the key contribution?
4. What are some compelling, *useful* examples?
5. Could it have been done more simply?
6. Have there been usability studies done?
   What do they show?
Holistic Design Goals for Information Visualization

- Tailor to the application and the domain
- Create highly interactive and integrated systems
- Embed the visualization within a larger application
- Provide alternative views
Visualization with a Light Touch: Orbitz.com
Visualization with a Light Touch: Orbitz.com
Visualization with a Light Touch: Orbitz.com

one moment please...

we're searching to find you the
low fares and options you want...
Visualization with a Light Touch: Orbitz.com

We checked for better fares at nearby airports, and your search found our lowest prices. Lowest fare beats this trip's 30-day average by $81.
Visualization with a Light Touch: Orbitz.com
For more information

- My course:  
  http://www.sims.berkeley.edu/courses/is247/s02/Lectures.html
- Atlas of Cyberspaces:  
  http://www.geog.ucl.ac.uk/casa/martin/atlas/atlas.html
- Gallery of Data Visualization; The Best and Worst of Statistical Graphics  
  http://www.math.yorku.ca/SCS/Gallery/
- Tamara Munzner’s collection:  
  http://graphics.stanford.edu/courses/cs348c-96-fall/resources.html
Thank you!