Information Visualization for Knowledge Discovery

Ben Shneiderman  ben@cs.umd.edu

Founding Director (1983-2000), Human-Computer Interaction Lab
Professor, Department of Computer Science
Member, Institute for Advanced Computer Studies
Interdisciplinary research community
- Computer Science & Info Studies
- Psych, Socio, Poli Sci & MITH

(www.cs.umd.edu/hcil)
Scientific Approach (beyond user friendly)

- Specify users and tasks
- Predict and measure
  - time to learn
  - speed of performance
  - rate of human errors
  - human retention over time
- Assess subjective satisfaction
  (Questionnaire for User Interface Satisfaction)
- Accommodate individual differences
- Consider social, organizational & cultural context
Design Issues

• Input devices & strategies
  • Keyboards, pointing devices, voice
  • Direct manipulation
  • Menus, forms, commands

• Output devices & formats
  • Screens, windows, color, sound
  • Text, tables, graphics
  • Instructions, messages, help

• Collaboration & communities

• Manuals, tutorials, training

www.awl.com/DTUI
• Scholars, Journalists, Citizens
• Teachers, Students
Visible Human Explorer (NLM)

- Doctors
- Surgeons
- Researchers
- Students
NASA Environmental Data

• Scientists
• Farmers
• Land planners
• Students
Bureau of the Census

- Economists, Policy makers, Journalists
- Teachers, Students
NSF Digital Government Initiative

- Find what you need
- Understand what you Find

Census, NCHS, BLS, EIA, NASS, SSA

www.ils.unc.edu/govstat/

FedStats
International Children’s Digital Library

Our Mission
We are a library that provides free access to children’s books from around the world. By ensuring access to books from many cultures and in diverse languages, we foster a love of reading, a readiness to learn, and a response to the challenges of world illiteracy.

We invite you to explore our growing library and learn how you can contribute to our efforts.

FEATURED BOOKS
Colorful illustrations with Jeanette Marin
May 8, 2006 Selected by Jeanette Marin, a student at the University of Maryland. Jeanette has three daughters.
Piccolo: Toolkit for 2D zoomable objects

Structured canvas of graphical objects in a hierarchical scenegraph
- Zooming animation
- Cameras, layers

Open, Extensible & Efficient
Java, C#, PocketPC versions

www.cs.umd.edu/hcil/piccolo
The eye…
the window of the soul,
is the principal means
by which the central sense
can most completely and
abundantly appreciate
the infinite works of nature.

*Leonardo da Vinci*

(1452 - 1519)
Using Vision to Think

- **Visual bandwidth is enormous**
  - Human perceptual skills are remarkable
    - Trend, cluster, gap, outlier...
    - Color, size, shape, proximity...
  - Human image storage is fast and vast

- **Opportunities**
  - Spatial layouts & coordination
  - Information visualization
  - Scientific visualization & simulation
  - Telepresence & augmented reality
  - Virtual environments
Spotfire: Retinol’s role in embryos & vision
Spotfire: DC natality data
Information Visualization: Mantra

- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
### Information Visualization: Data Types

<table>
<thead>
<tr>
<th>SciViz</th>
<th>InfoViz</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1-D Linear</strong></td>
<td>Document Lens, SeeSoft, Info Mural, Value Bars</td>
</tr>
<tr>
<td><strong>2-D Map</strong></td>
<td>GIS, ArcView, PageMaker, Medical imagery</td>
</tr>
<tr>
<td><strong>3-D World</strong></td>
<td>CAD, Medical, Molecules, Architecture</td>
</tr>
<tr>
<td><strong>Multi-Var</strong></td>
<td>Parallel Coordinates, Spotfire, XGobi, Visage, Influence Explorer, TableLens, DEVise</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
<td>Perspective Wall, LifeLines, Lifestreams, Project Managers, DataSpiral</td>
</tr>
<tr>
<td><strong>Tree</strong></td>
<td>Cone/Cam/Hyperbolic, TreeBrowser, Treemap</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td>Netmap, netViz, SeeNet, Butterfly, Multi-trees</td>
</tr>
</tbody>
</table>

(Online Library of Information Visualization Environments) [otal.umd.edu/Olive](otal.umd.edu/Olive)
ManyEyes: A web sharing platform

Welcome to the alpha version of Many Eyes!
View your data, ask questions, and share your discoveries.
Harness the collective intelligence of the net for insight and analysis.

http://services.alphaworks.ibm.com/manyeyes/app
Treemap: view large trees with node values

+ Space filling
+ Space limited
+ Color coding
+ Size coding
- Requires learning

TreeViz (Mac, Johnson, 1992)
NBA-Tree(Sun, Turo, 1993)
Winsurfer (Teittinen, 1996)
Diskmapper (Windows, Micrologic)
SequoiaView, Panopticon,
HiveGroup, Solvern
Treemap4 (UMd, 2004)

Treemap: Stock market, clustered by industry
Market falls steeply Feb 27, 2007, with one exception

Microsoft -4.09%
MSFT: last sale 27.88, change -1.19
(click for more detail)
Treemap: Newsmap

1. Russian violence leaves 98 dead
2. Iranian frees British troops seized on border with Iraq
3. Dozens die as fighting intensifies ahead of Iraqi transfer of power
4. Saudi offers militants one-month amnesty
5. Quartet delegations meet over Israeli withdrawal plan
6. Iraq attacks leave 83 dead
7. Peacekeeping without the US
8. S. Korean president orders probe of foreign ministry
9. He's suddenly the center of attention as media try to snare him for TV shows
10. Oklahoma Wal-Mart plaintiff says battle already won
11. Gates Muller sues CEO of PeopleSoft
12. No cakewalk
13. SpaceShipOne makes history barely
14. Devil Rays-Blue Jays Preview
15. Doping-Montgomery summoned to Doping Hearing
16. NBA Draft to be held in New York, Thursday
17. Yankee pitcher set to return to NYC with his hand
18. Czech reserves being hammered by Germany crashing
19. Tiger loses confidence
20. Yankee on Fame, Non-Fiction Films, and Fahrenheit 9/11
21. Secrets of muscle growth revealed
22. Billion with Cervical Stiff get pop smear
23. Human breast milk can effectively treat HIV
24. Bob Dylan awarded honorary doctorate of music degree
25. www.hivegroup.com
Treemap: Gene Ontology

http://www.cs.umd.edu/hcil/treemap/
Treemap: Product catalogs
LifeLines: Patient Histories
LifeLines: Customer Histories

Mrs Linda Simpson  
Maryland  41  
Allergic to polyester

<table>
<thead>
<tr>
<th>LifeLine</th>
<th>90/95</th>
<th>96</th>
<th>97</th>
<th>98</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalogs</td>
<td>Fall95</td>
<td>Fall96</td>
<td>Fall97</td>
<td>Spring97</td>
<td>Spring98</td>
</tr>
<tr>
<td>Store</td>
<td>Web</td>
<td>Store</td>
<td>Web</td>
<td>Web</td>
<td>Web</td>
</tr>
<tr>
<td>WWW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PURCHASES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clothes</td>
<td>$101</td>
<td>$194</td>
<td>$127</td>
<td>$150 back</td>
<td></td>
</tr>
<tr>
<td>Week-end</td>
<td></td>
<td>$160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accessories</td>
<td></td>
<td>$152</td>
<td></td>
<td>$27</td>
<td></td>
</tr>
<tr>
<td>House</td>
<td></td>
<td>$152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Return</td>
<td></td>
<td></td>
<td>too-small</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credit</td>
<td></td>
<td>Credit Refused</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birthday</td>
<td>Linda BD</td>
<td>Linda BD</td>
<td>Linda BD</td>
<td>Linda BD</td>
<td></td>
</tr>
</tbody>
</table>

Linda Simpson ORDER#614345 (MOST RECENT)  
11-1-1997, $150, Visa card, order by phone

- D45F224: Boots - 4 - Black
  ----> BACKORDER - estimated shipping
  12-15
- D45F236: Gloves - Black - M
- A78156: Corduroy pants - Tan - L
- 157826: Hat - Grey - M

Kids Backcountry hiking boots  
Leather boots with super grip soles.
Temporal Data: TimeSearcher 1.3

- Time series
  - Stocks
  - Weather
  - Genes
- User-specified patterns
- Rapid search
Temporal Data: TimeSearcher 2.0

- Long Time series (>10,000 time points)
- Multiple variables
- Controlled precision in match (Linear, offset, noise, amplitude)
Goal: Find Features in Multi-Var Data

• Clear vision of what the data is
• Clear goal of what you are looking for

• Systematic strategy for examining all views
• Ranking of views to guide discovery
• Tools to record progress & annotate findings
“HCE enabled us to find important clusters that we didn’t know about.”
- a user
Do you see anything interesting?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Element</td>
<td>+P1</td>
<td>+P2</td>
<td>Atomic Num</td>
<td>Atomic Mass</td>
<td>Atomic Radi</td>
<td>Ionic Radi</td>
<td>lonization E</td>
<td>Electroneg</td>
<td>+C1</td>
</tr>
<tr>
<td>2</td>
<td>Ag</td>
<td>110</td>
<td>0</td>
<td>55</td>
<td>221</td>
<td>200</td>
<td>126</td>
<td>51</td>
<td>11</td>
<td>62</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>Hg</td>
<td>120</td>
<td>0</td>
<td>63</td>
<td>27</td>
<td>144</td>
<td>129</td>
<td>75</td>
<td>18</td>
<td>124</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>Al</td>
<td>130</td>
<td>0</td>
<td>13</td>
<td>27</td>
<td>143</td>
<td>87</td>
<td>80</td>
<td>16</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Ba</td>
<td>140</td>
<td>0</td>
<td>18</td>
<td>39</td>
<td>98</td>
<td>154</td>
<td>158</td>
<td>32</td>
<td>176</td>
<td>51</td>
</tr>
<tr>
<td>6</td>
<td>As</td>
<td>150</td>
<td>0</td>
<td>23</td>
<td>75</td>
<td>120</td>
<td>72</td>
<td>98</td>
<td>22</td>
<td>115</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>Ac</td>
<td>160</td>
<td>0</td>
<td>28</td>
<td>85</td>
<td>210</td>
<td>76</td>
<td>95</td>
<td>22</td>
<td>119</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>Au</td>
<td>170</td>
<td>0</td>
<td>33</td>
<td>93</td>
<td>197</td>
<td>89</td>
<td>91</td>
<td>25</td>
<td>151</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>Br</td>
<td>180</td>
<td>0</td>
<td>38</td>
<td>10</td>
<td>85</td>
<td>83</td>
<td>41</td>
<td>83</td>
<td>101</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>Bi</td>
<td>190</td>
<td>0</td>
<td>43</td>
<td>15</td>
<td>137</td>
<td>149</td>
<td>51</td>
<td>46</td>
<td>56</td>
<td>46</td>
</tr>
<tr>
<td>11</td>
<td>Be</td>
<td>200</td>
<td>0</td>
<td>48</td>
<td>9</td>
<td>112</td>
<td>59</td>
<td>93</td>
<td>15</td>
<td>82</td>
<td>15</td>
</tr>
<tr>
<td>12</td>
<td>Br</td>
<td>210</td>
<td>0</td>
<td>53</td>
<td>15</td>
<td>209</td>
<td>150</td>
<td>73</td>
<td>20</td>
<td>140</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>Cr</td>
<td>220</td>
<td>0</td>
<td>58</td>
<td>10</td>
<td>114</td>
<td>182</td>
<td>118</td>
<td>30</td>
<td>161</td>
<td>44</td>
</tr>
<tr>
<td>14</td>
<td>Cu</td>
<td>230</td>
<td>0</td>
<td>63</td>
<td>15</td>
<td>12</td>
<td>77</td>
<td>113</td>
<td>25</td>
<td>82</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>Co</td>
<td>240</td>
<td>0</td>
<td>68</td>
<td>20</td>
<td>40</td>
<td>197</td>
<td>114</td>
<td>80</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>16</td>
<td>Cd</td>
<td>250</td>
<td>0</td>
<td>73</td>
<td>25</td>
<td>112</td>
<td>151</td>
<td>109</td>
<td>90</td>
<td>17</td>
<td>113</td>
</tr>
<tr>
<td>17</td>
<td>Cl</td>
<td>260</td>
<td>0</td>
<td>78</td>
<td>30</td>
<td>35</td>
<td>100</td>
<td>167</td>
<td>32</td>
<td>173</td>
<td>47</td>
</tr>
<tr>
<td>18</td>
<td>Al</td>
<td>270</td>
<td>0</td>
<td>83</td>
<td>35</td>
<td>79</td>
<td>114</td>
<td>182</td>
<td>118</td>
<td>30</td>
<td>161</td>
</tr>
<tr>
<td>19</td>
<td>Cr</td>
<td>280</td>
<td>0</td>
<td>88</td>
<td>40</td>
<td>12</td>
<td>77</td>
<td>113</td>
<td>25</td>
<td>82</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>Cu</td>
<td>290</td>
<td>0</td>
<td>93</td>
<td>45</td>
<td>209</td>
<td>150</td>
<td>73</td>
<td>20</td>
<td>140</td>
<td>27</td>
</tr>
<tr>
<td>21</td>
<td>Fe</td>
<td>300</td>
<td>0</td>
<td>98</td>
<td>50</td>
<td>114</td>
<td>182</td>
<td>118</td>
<td>30</td>
<td>161</td>
<td>44</td>
</tr>
<tr>
<td>22</td>
<td>Fe</td>
<td>310</td>
<td>0</td>
<td>103</td>
<td>55</td>
<td>12</td>
<td>77</td>
<td>113</td>
<td>25</td>
<td>82</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>Ga</td>
<td>320</td>
<td>0</td>
<td>108</td>
<td>60</td>
<td>132</td>
<td>265</td>
<td>181</td>
<td>39</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>24</td>
<td>Ge</td>
<td>330</td>
<td>0</td>
<td>113</td>
<td>66</td>
<td>137</td>
<td>135</td>
<td>76</td>
<td>60</td>
<td>18</td>
<td>89</td>
</tr>
<tr>
<td>25</td>
<td>Hf</td>
<td>340</td>
<td>0</td>
<td>118</td>
<td>71</td>
<td>227</td>
<td>152</td>
<td>43</td>
<td>8</td>
<td>37</td>
<td>56</td>
</tr>
<tr>
<td>26</td>
<td>Hg</td>
<td>350</td>
<td>0</td>
<td>123</td>
<td>76</td>
<td>132</td>
<td>128</td>
<td>87</td>
<td>76</td>
<td>19</td>
<td>118</td>
</tr>
</tbody>
</table>
What features stand out?
Correlation...What else?

![Scatter Plot](image)

Scatter Plot

- Ionization Energy
- Electron Energy

Axes:
- Ionization Energy
- Electron Energy

Values:
- 50, 75, 100, 125, 150, 175, 200, 225, 250
- 0, 10, 20, 30, 40

Linear trend line.
... and Outliers

Scatter Plot

- Ionization Energy
- Electron Energy

He
Rn
Demonstration

• US counties census data
  • 3138 counties
  • 14 dimensions: population density, poverty level, unemployment, etc.
Rank-by-Feature Framework: 1D

Ranking Criterion

Rank-by-Feature Prism

Score List

Manual Projection Browser
A Ranking Example

3138 U.S. counties with 17 attributes

Ranking Criterion: Uniformity (entropy) (6.7, 6.1, 4.5, 1.5)

Ranking Criterion: Pearson correlation (0.996, 0.31, 0.01, -0.69)
HCE Status

- In collaboration and sponsored by Eric Hoffman: Children’s National Medical Center
- Phd work of Jinwook Seo
- 72K lines of C++ codes
- 4,000+ downloads since April 2002

www.cs.umd.edu/hcil/hce
Evaluation Methods

Ethnographic Observational Situated

• Multi-Dimensional
• In-depth
• Long-term
• Case studies
Evaluation Methods

Ethnographic Observational Situated

- Multi-Dimensional
- In-depth
- Long-term
- Case studies

Domain Experts Doing Their Own Work for Weeks & Months
Evaluation Methods

Ethnographic Observational Situated
• Multi-Dimensional
• In-depth
• Long-term
• Case studies

MILCs

Shneiderman & Plaisant, BeLIV workshop, 2006
MILC example

- Evaluate Hierarchical Clustering Explorer
- Focused on rank-by-feature framework
- 3 case studies, 4-8 weeks (molecular biologist, statistician, meteorologist)
- 57 email surveys
- Identified problems early, gave strong positive feedback about benefits of rank-by-feature

Seo & Shneiderman, *IEEE TVCG* 12,3, 2006
MILC example

• Evaluate SocialAction

• Focused on integrating statistics & visualization
• 4 case studies, 4-8 weeks (journalist, bibliometrician, terrorist analyst, organizational analyst)
• Identified desired features, gave strong positive feedback about benefits of integration

Perer & Shneiderman, 2007
Case Study Methodology

1) Interview (1 hr)
2) Training (2 hr)
3) Early Use (2-4 weeks)
4) Mature Use (2-4 weeks)
5) Outcome (1 hr)
Take Away Message

Rank-by-Feature Framework

• Decomposition of complex problems into multiple simpler problems wins
• Ranking guides discovery
• Systematic strategies

www.cs.umd.edu/hcil/hce