Voyagers and Voyeurs
Visualization and Social Data Analysis

Jeffrey Heer
Computer Science Dept
Stanford University

Stanford Computer Forum
15 April 2009
<table>
<thead>
<tr>
<th>Set A</th>
<th>Set B</th>
<th>Set C</th>
<th>Set D</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>8.04</td>
<td>10</td>
<td>9.14</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8</td>
<td>8.14</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>13</td>
<td>8.74</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>9</td>
<td>8.77</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>11</td>
<td>9.26</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>14</td>
<td>8.1</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6</td>
<td>6.13</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>4</td>
<td>3.1</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>12</td>
<td>9.11</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7</td>
<td>7.26</td>
</tr>
<tr>
<td>5</td>
<td>5.68</td>
<td>5</td>
<td>4.74</td>
</tr>
</tbody>
</table>

**Summary Statistics**

- $\mu_X = 9.0$, $\sigma_X = 3.317$
- $\mu_Y = 7.5$, $\sigma_Y = 2.03$

**Linear Regression**

- $Y = 3 + 0.5 \times X$
- $R^2 = 0.67$

[Anscombe 73]
Improve the analysis and communication of information using interactive visualization
1. Visualization Techniques
2. Visualization Tools
3. Social Data Analysis
Visualization Techniques

Improve visual analysis and communication via novel algorithms, encodings, and interactions
Visualization Tools

Simplify creation and customization by crafting toolkits for interactive visualization

Prefuse  CHI 2005, InfoVis 2006

Flare  2008
Social Data Analysis

Leverage the insights of multiple analysts with interfaces for collaborative visual analysis

Where have all the dentists gone?
Visualization Techniques
Visualizing Large Hierarchies

Indented Layout

Reingold-Tilford Layout
Space-constrained, multi-focal tree layout

Degree-of-Interest Trees

[AV104]
Aspect Ratio Optimization

Maximize perceptual discriminability of line segment angles

\[ \sum_{i} \sum_{j} | \theta_i(\alpha) - \theta_j(\alpha) |^2 \]

Identify trends using spectral analysis; find ratios for trend lines
Enron E-Mail Corpus
Get your message. I'm testifying at the Congressional hearing and Dave is covering. I think Jeff's comments were taken out of context. He said policymakers need to take care of small customers whose bills are tripling. Frankly, we'd get slaughtered if we said anything else. But he also said there is a right way and a wrong way to do it. Union and others had provided a market-based answer by offering a fixed price deal to SCE (which would have enabled them to cap rates to those who had not switched). California elected instead to cap rates and deficit spend (ie create a deferral account). I don't think we can stand for anything that doesn't protect the small customers, but we can continue to emphasize the market based solutions. One of the messages in my testimony will be: customers should be encouraged to choose. Those who did are doing fine.
Enron 'Mastermind' Pleads Guilty

SAN FRANCISCO, Oct. 17, 2002

(AP) A former top energy trader, considered the mastermind of Enron Corp.'s scheme to drive up California's energy prices, pleaded guilty Thursday to a federal conspiracy charge.

Deputy Attorney General Larry Thompson, center, head of the Justice Department's Corporate Fraud Task Force, comments Thursday on the guilty plea by Timothy N. Belden, Enron's chief energy trader. (Photo: CBS/AP)

Timothy Belden, the former head of trading in Enron's Portland, Ore., office, admitted to one count of conspiracy to commit wire fraud and promised to cooperate with state and federal prosecutors as well as any non-criminal effort to investigate the energy industry.

"I did it because I was trying to maximize profit for Enron," Belden told U.S. District Judge Martin Jenkins.
Visualization Tools
Setbacks to Adoption

Visualizations are hard to create
Layout algorithms, dynamic graphics

Pre-built “widgets” are not enough
Good designs tailored to an application domain
The Prefuse and Flare Toolkits

Integrate data handling, graphics, and interaction
Extensible operator language for visual encodings
Exploration via dynamic queries, search, zooming
Usage and Uptake

Open Source projects  95,000+ downloads
Used by students, researchers, corporations

Design Patterns describing architecture [InfoVis 06]
Phan et al, *Flow Map Layout*, InfoVis 05
Collins, *DocuBurst*, Featured in the Toronto Star
Perer, *SocialAction*, InfoVis 06, IUI 07, CHI 08
Collins et al, *VisLink*, InfoVis 07
Social Data Analysis
A Tale of Two Visualizations
Observations

Groups spent more time in front of the visualization than individuals.

Friends encouraged each other to unearth relationships, probe community boundaries, and challenge reported information.

Social play resulted in informal analysis, often driven by story-telling of group histories.
The Baby Name Voyager

name >> je|
Social Data Analysis

Visual sensemaking can be social as well as cognitive.

How can user interfaces support and encourage collaborative analysis?

Understand impact by building and deploying real systems.
sense.us
A Web Application for Collaborative Visualization of Demographic Data
Dogear Tags
Dogear Tag Usage, May 2005 to August 2006
source: IBM Dogear
10 comments

Dogear People
Dogear Bookmarking by Person, May 2005 to August 2006
source: IBM Dogear
1 comment

Job Voyager
Reported Occupations of U.S. Labor Force, 1850-2000
source: http://ipums.org
139 comments

Birthplace Voyager
Reported Birthplace of U.S. Residents, 1850-2000
source: http://ipums.org
10 comments

U.S. Census State Map
State Map of 2000-2005 Census Data
source: U.S. Census Bureau
16 comments

Population Pyramid
U.S. Population Demographics, 1850-2000
source: http://ipums.org
7 comments

sense.us - social data visualization

sense.us is a prototype system for collaborative visualization.
- See the data. See what people have to say about it.
- Dive into the data and share your explorations.

The site requires Java 1.5+ and either Firefox or Internet Explorer.
Use your IBM W3/BluePages e-mail and password to login.
Use at least 1024x768 resolution for the best experience.

Check out the user's guide and privacy policy before getting started.

Having problems using Firefox with Java 1.4? Some users using Firefox and Java 1.4 have found that comments aren't loading properly. If you run into this problem, consider upgrading to Java 1.5 (Windows, Linux) or using Internet Explorer (on Windows systems). Sorry for any inconvenience!
Building Off of Others

![Graph showing percentage of work force in military from 1850 to 2000, with notable increases during Civil War and World War I. The graph indicates a significant rise in military personnel, especially for men, with a peak around 1940.](image-url)
Data Jokes

Great depression "killed" a lot of brokers
Collaborative Sensemaking
Content Analysis of Comments

Feature prevalence from content analysis (min Cohen’s $\kappa = .74$)  
High co-occurrence of Observations, Questions, and Hypotheses
Voyagers and Voyeurs

Complementary faces of analysis

**Voyager** – focus on visualized data
Active engagement with the data
Serendipitous comment discovery

**Voyeur** – focus on comment listings
Investigate others’ explorations
Find people and topics of interest
Catalyze new explorations
Spotfire Decision Site Posters
### Analytic Dashboard

#### Discounting by Product

<table>
<thead>
<tr>
<th>Labels</th>
<th>Feb 1</th>
<th>Mar 1</th>
<th>Apr 1</th>
<th>May 1</th>
<th>Jun 1</th>
<th>Jul 1</th>
<th>Aug 1</th>
<th>Sep 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pens &amp; Art Supplies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Top Products by Profitability

<table>
<thead>
<tr>
<th>Product</th>
<th>Sum of Gross Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canon PC840 Copier</td>
<td>57,911</td>
</tr>
<tr>
<td>Polycom Voicestation 100</td>
<td>37,264</td>
</tr>
<tr>
<td>Hon Olson Stack Chair</td>
<td>30,369</td>
</tr>
<tr>
<td>Phone 916</td>
<td>25,421</td>
</tr>
<tr>
<td>Ericsson 8360</td>
<td>23,349</td>
</tr>
<tr>
<td>Siemens 5185</td>
<td>14,903</td>
</tr>
</tbody>
</table>

#### Summary by Year and Product Type

<table>
<thead>
<tr>
<th>Product Type</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furniture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Discount</td>
<td>13</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Total Profit</td>
<td>$480.4K</td>
<td>$57.4K</td>
<td>$5.6K</td>
<td>$5.2K</td>
</tr>
<tr>
<td>Office Supplies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Discount</td>
<td>37</td>
<td>57</td>
<td>50</td>
<td>58</td>
</tr>
<tr>
<td>Total Profit</td>
<td>$70.8K</td>
<td>$51.3K</td>
<td>$50.1K</td>
<td>$41.9K</td>
</tr>
</tbody>
</table>
Conclusion
Visualization Techniques

Algorithms for optimized data displays

Visual analysis tools for evolving hierarchies, social networks, and e-mail archives
Visualization Tools

**Toolkits** which simplify visualization creation and support fine-grained reuse of techniques.
Social Data Analysis

Systems and techniques for social data analysis
Interfaces for social navigation and annotation
New tools for finding and visualizing web data
Improve the analysis and communication of information using interactive visualization
Voyagers and Voyeurs
Visualization and Social Data Analysis

Jeffrey Heer
Computer Science
Stanford University

jheer@cs.stanford.edu