Introduction

CS 4460/7450 - Information Visualization
Jan. 8, 2009
John Stasko

Exercise

• Get out pencil and paper
Data Explosion

- Society is more complex
  - There simply is more “stuff”

- Computers, internet and web give people access to an incredible amount of data
  - news, sports, financial, purchases, etc...

How Much Data? (1)

- Estimated info added to digital universe each year will soon approach 1 ZB (zettabyte)*
  - 1000000000000000000000 (10^{21}) bytes

*But only half that goes to my email inbox
How Much Data? (2)

- 6 million FedEx transactions per day
  [http://www.fedex.com/us/about/today/companies/corporation/facts.html]

- Average of 98 million Visa credit-card transactions per day in 2005

- Average of 5.4 petabytes of data crosses AT&T’s network per day
  [http://att.sbc.com/gen/investor-relations?pid=5711]

- Average of 610 to 1110 billion e-mails worldwide per year (based on estimates in 2000)
  [http://www2.sims.berkeley.edu/research/projects/how-much-info/internet.html]

Data Overload

- Confound: How to make use of the data
  - How do we make sense of the data?
  - How do we harness this data in decision-making processes?
  - How do we avoid being overwhelmed?
The Challenge

- Transform the *data* into *information* (understanding, insight) thus making it useful to people

The Problem

- Web, Books, Papers, Game scores, Scientific data, Biotech, Shopping People, Stock/finance, News

How?
- Vision: 100 MB/s
- Ears: <100 b/s
- Telepathy
- Haptic/tactile
- Smell
- Taste

Two slides courtesy of Chris North
Human Vision

- Highest bandwidth sense
- Fast, parallel
- Pattern recognition
- Pre-attentive
- Extends memory and cognitive capacity
  (Multiplication test)
- People think visually

Impressive. Let’s use it!

Some Examples

- Why visualization helps...
Which cereal has the most/least potassium?  
Is there a relationship between potassium and fiber?  
If so, are there any outliers?  
Which manufacturer makes the healthiest cereals?

<table>
<thead>
<tr>
<th>Cereal</th>
<th>Manufacturer</th>
<th>Fiber</th>
<th>Potassium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honeycomb</td>
<td>P</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Just Right Fruit &amp; Nut</td>
<td>K</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>Life</td>
<td>Q</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>Lucky Charms</td>
<td>G</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Marpo</td>
<td>A</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>Muesli Raisins, Cacao, &amp;</td>
<td>R</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>Multi-Grain Cheerios</td>
<td>G</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>Nut Grain Almond Bran</td>
<td>K</td>
<td>3</td>
<td>130</td>
</tr>
<tr>
<td>Nut Grain Wheat</td>
<td>K</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>Oatmeal Bran Crip</td>
<td>G</td>
<td>1.5</td>
<td>120</td>
</tr>
<tr>
<td>Post Nut Raisin Bran</td>
<td>P</td>
<td>6</td>
<td>200</td>
</tr>
<tr>
<td>Product 10</td>
<td>K</td>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>Quaker Oatmeal</td>
<td>G</td>
<td>2.7</td>
<td>110</td>
</tr>
<tr>
<td>Raisin Bran</td>
<td>K</td>
<td>5</td>
<td>240</td>
</tr>
<tr>
<td>Raisin Nut Bran</td>
<td>G</td>
<td>2.5</td>
<td>140</td>
</tr>
<tr>
<td>Rice Krispies</td>
<td>K</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Shredded Wheat</td>
<td>N</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>Shredded Wheat n Bran</td>
<td>N</td>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>Shredded Wheat x Bran</td>
<td>N</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>Smacks</td>
<td>K</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>Special K</td>
<td>K</td>
<td>1</td>
<td>55</td>
</tr>
<tr>
<td>Strawberry Fruit</td>
<td>K</td>
<td>1</td>
<td>55</td>
</tr>
<tr>
<td>Total Corn Flakes</td>
<td>G</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Total Raisin Bran</td>
<td>G</td>
<td>4</td>
<td>200</td>
</tr>
<tr>
<td>Total Whole Grain</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>Tobb</td>
<td>G</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Wheaties</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>Wheaties Honey Gold</td>
<td>G</td>
<td>1</td>
<td>80</td>
</tr>
</tbody>
</table>
Even Tougher?

- What if you could only see one cereal’s data at a time? (e.g. some websites)
- What if I read the data to you?

Another Illustrative Example
Four Data Sets

- Mean of the x values = 9.0
- Mean of the y values = 7.5
- Equation of the least-squared regression line is: $y = 3 + 0.5x$
- Sums of squared errors (about the mean) = 110.0
- Regression sums of squared errors (variance accounted for by x) = 27.5
- Residual sums of squared errors (about the regression line) = 13.75
- Correlation coefficient = 0.82
- Coefficient of determination = 0.67

http://astro.swarthmore.edu/astro121/anscombe.html
### The Values

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.0, 8.04</td>
<td>10.0, 9.14</td>
<td>10.0, 7.46</td>
<td>8.0, 6.58</td>
</tr>
<tr>
<td>2</td>
<td>8.0, 6.95</td>
<td>8.0, 8.14</td>
<td>8.0, 6.77</td>
<td>8.0, 5.76</td>
</tr>
<tr>
<td>3</td>
<td>13.0, 7.58</td>
<td>13.0, 8.74</td>
<td>13.0, 12.74</td>
<td>8.0, 7.71</td>
</tr>
<tr>
<td>4</td>
<td>9.0, 8.81</td>
<td>9.0, 8.77</td>
<td>9.0, 7.11</td>
<td>8.0, 8.84</td>
</tr>
<tr>
<td>5</td>
<td>11.0, 8.33</td>
<td>11.0, 9.26</td>
<td>11.0, 7.81</td>
<td>8.0, 8.47</td>
</tr>
<tr>
<td>6</td>
<td>14.0, 9.96</td>
<td>14.0, 8.10</td>
<td>14.0, 8.84</td>
<td>8.0, 7.04</td>
</tr>
<tr>
<td>7</td>
<td>6.0, 7.24</td>
<td>6.0, 6.13</td>
<td>6.0, 6.08</td>
<td>8.0, 5.25</td>
</tr>
<tr>
<td>8</td>
<td>4.0, 4.26</td>
<td>4.0, 3.10</td>
<td>4.0, 5.39</td>
<td>19.0, 12.50</td>
</tr>
<tr>
<td>9</td>
<td>12.0, 10.84</td>
<td>12.0, 9.13</td>
<td>12.0, 8.15</td>
<td>8.0, 5.56</td>
</tr>
<tr>
<td>10</td>
<td>7.0, 4.82</td>
<td>7.0, 7.26</td>
<td>7.0, 6.42</td>
<td>8.0, 7.91</td>
</tr>
<tr>
<td>11</td>
<td>5.0, 5.68</td>
<td>5.0, 4.74</td>
<td>5.0, 5.73</td>
<td>8.0, 6.89</td>
</tr>
</tbody>
</table>

### Exercise Redux

- Let’s check what you did...

- People work differently
Visualization

• Definition
  – “The use of computer-supported, interactive visual representations of data to amplify cognition.”
    From [Card, Mackinlay Shneiderman '98]

Visualization

• Often thought of as process of making a graphic or an image
• Really is a cognitive process
  – Form a mental image of something
  – Internalize an understanding
• “The purpose of visualization is insight, not pictures”
  – Insight: discovery, decision making, explanation
Main Idea

• Visuals help us think
  – Provide a frame of reference, a temporary storage area
• Cognition → Perception
• Pattern matching
• External cognition aid
  – Role of external world in thinking and reason

Larkin & Simon ’87
Card, Mackinlay, Shneiderman ’98

When to Apply?

• Many other techniques for data analysis
  – Data mining, DB queries, machine learning...

• Visualization most useful in **exploratory data analysis**
  – Don’t know what you’re looking for
  – Don’t have a priori questions
  – Want to know what questions to ask
Part of our Culture

- “I see what you’re saying”
- “Seeing is believing”
- “A picture is worth a thousand words”

Overview

Visualization
“Data visualization”

Scientific visualization

Information visualization
Scientific Visualization

- Primarily relates to and represents something physical or geometric
  - Often 3-D
  - Examples
    Air flow over a wing
    Stresses on a girder
    Torrents inside a tornado
    Organs in the human body
    Molecular bonding

Information Visualization

- What is “information”?
  - Items, entities, things which do not have a direct physical correspondence
  - Notion of abstractness of the entities is important too
  - Examples: baseball statistics, stock trends, connections between criminals, car attributes...
Information Visualization

• What is “visualization”?
  – The use of computer-supported, interactive visual representations of data to amplify cognition.
    From [Card, Mackinlay Shneiderman ’98]

• Components:
  – Taking items without a direct physical correspondence and mapping them to a 2-D or 3-D physical space.
  – Giving information a visual representation that is useful for analysis and decision-making
Two Key Attributes

- **Scale**
  - Challenge often arises when data sets become very large

- **Interactivity**
  - Want to show multiple different perspectives on the data

Example Domains for Info Vis

- Text
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Spring 2009 CS 4460/7450
Examples

- Images
  - Start with static pictures
  - Are they information visualizations?

Atlanta Flight Traffic

Atlanta Journal
April 30, 2000
2000 Election Ballot

Electoral College

The presidential race

Atlanta Journal
November 5, 2000
The Comics

Understanding Comics by Scott McCloud

Hartsfield Airport

Note: Gold icons are live links
Power Costs

Average cost per month to use

Wall Street Journal
August 16, 2001

London Subway

www.thetube.com
SARS Outbreak

Napoleonic's March

From E. Tufte
The Visual Display of Quantitative Information

Minard graphic
size of army
direction
latitude
longitude
temperature
date
NYC Weather

2220 numbers

Tufte, Vol. 1


Or, for fun...

http://www.boingboing.net/2006/11/02/hilarious-piechartvi.html
http://xkcd.com/197/

http://www.flickr.com/photos/91884218@N00/3108768440/in/pool-songchart
But Don’t Do This

Excel

Get rid of those darn 3D bars!
USA Today Graphics

Or worse yet...

Unemployment Rates

Unemployment rates by county, December 2000 - November 2001 averages
(U.S. rate = 4.6 percent)

Source: Bureau of Labor Statistics
Local Area Unemployment Statistics
Examples

- Tools/Systems
  - Now interaction becomes important...

Map of the Market

[Map image]

www.smartmoney.com/marketmap
Philip Glass Music

http://babynamewizard.com/namevoyager/
HomeFinder

HCIL
Univ. Maryland

Tasks for Info Vis?

• Search (OK)
  – Finding a specific piece of information
    How many games did the Braves win in 1995?
    What novels did Ian Fleming author?

• Browsing (Better)
  – Look over or inspect something in a more casual manner, seek interesting information
    Learn about crystallography
    What has Jane been up to lately?
Tasks in Info Vis

- Analysis
  - Comparison-Difference
  - Outliers, Extremes
  - Patterns
- Assimilation
- Monitoring
- Awareness

More to come in a future class...

Good Resources
InfoVis meet MTV
Administratia

- TA
  - Youn-ah Kang

- Office Hours for both of us TBA

Get it all from class website
- Syllabus & Bibliography
- Assignments
- Instructor & TA
- Related Courses
- InfoVis Resources
Administratia

- Assignments
  - Daily readings question
  - HWs
  - Project

HW

- HW1 due next Thursday
  - Data Exploration and Analysis (without infovis)
  - Follow the directions
Upcoming

• Data & Graph/Chart Design
  – Reading:
    S. Few – web article

• Case Studies
  – Reading:
    C. Weaver – Hotel visits