Overview of InfoVis

CS 7450 - Information Visualization
Aug. 24, 2016
John Stasko

Learning Objectives

- Articulate definition and purpose of visualization
- Describe two main uses or applications of visualization
- List two primary components of visualizations
- Describe the different areas of academic visualization research
- Explain the infovis “pipeline” (process)
Exercise

• Get out pencil and paper

Electronics

• This is now a NO laptops/cellphones class
• Exceptions will be noted (exercises, etc)
• Note-takers, see me

Reminder
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

Data

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

How?

Data Transfer

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

Two slides courtesy of Chris North

Human Vision

• Highest bandwidth sense
• Fast, parallel
• Pattern recognition
• Pre-attentive
• Extends memory and cognitive capacity
• People think visually

Impressive. Let’s use it!
An Example

- Why visualization helps...

Questions:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cereal A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>2</td>
<td>100% Bran</td>
<td>9</td>
<td>10</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>100% Natural Bran</td>
<td>2</td>
<td>2</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>All-Bran</td>
<td>K</td>
<td>9</td>
<td>320</td>
</tr>
<tr>
<td>5</td>
<td>All-Bran with Extra Fiber</td>
<td>K</td>
<td>14</td>
<td>330</td>
</tr>
<tr>
<td>6</td>
<td>Almond Delight</td>
<td>R</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Apple Cinnamon Cheerios</td>
<td>G</td>
<td>1.5</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>Bran Chex</td>
<td>R</td>
<td>4</td>
<td>125</td>
</tr>
<tr>
<td>9</td>
<td>Bran Flakes</td>
<td>P</td>
<td>6</td>
<td>190</td>
</tr>
<tr>
<td>10</td>
<td>Cap'n Crunch</td>
<td>G</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>Cheerios</td>
<td>G</td>
<td>2</td>
<td>105</td>
</tr>
<tr>
<td>12</td>
<td>Cocoa Puffs</td>
<td>G</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>13</td>
<td>Corn Chex</td>
<td>R</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>14</td>
<td>Corn Flakes</td>
<td>K</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>15</td>
<td>Count Chocula</td>
<td>G</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>16</td>
<td>Crispix</td>
<td>K</td>
<td>4</td>
<td>160</td>
</tr>
<tr>
<td>17</td>
<td>Cream of Wheat (Quick)</td>
<td>K</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>Crispy Wheat &amp; Raisins</td>
<td>G</td>
<td>2</td>
<td>120</td>
</tr>
<tr>
<td>19</td>
<td>Double Chex</td>
<td>R</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>20</td>
<td>Froot Loops</td>
<td>K</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>21</td>
<td>Frosted Flakes</td>
<td>K</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>22</td>
<td>Fruit &amp; Fiber Cereal, Wal</td>
<td>P</td>
<td>5</td>
<td>200</td>
</tr>
<tr>
<td>23</td>
<td>Fruity Bran</td>
<td>K</td>
<td>5</td>
<td>190</td>
</tr>
<tr>
<td>24</td>
<td>Fruity Pebbles</td>
<td>P</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>Golden Grahams</td>
<td>G</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>26</td>
<td>Grape Nut Flakes</td>
<td>P</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>27</td>
<td>Honey Nut Cheerios</td>
<td>G</td>
<td>1.5</td>
<td>70</td>
</tr>
</tbody>
</table>

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?
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 Data Sets

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>
More on this Topic

- “Value of visualization” lecture later in term

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

Visualizations

• 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
**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

---

**Expressed Well**

“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland  
*The Elements of Graphing Data*
Part of our Culture

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

Admin Intermission

- Overloads
- Surveys
- More...
Administratia

- Get it all from class website
  - Schedule
  - Assignments
  - Instructor & TA
  - Related Courses
  - InfoVis Resources

http://www.cc.gatech.edu/~stasko/7450

T-Square Site

Get lecture slides from last time there
Books

Required
- Design for Information

Recommended
- Envisioning Information

Schedule

Overview
- Aug 22: Introduction
- Readings
  - Tutorial: Introduction to Information Visualization
  - Tutorial: Data Visualization
  - Workshop: Introduction to Information Visualization

Detail
- Aug 22: Introduction to Information Visualization
  - Lecture: Introduction to Information Visualization
  - Reading: Tutorial: Introduction to Information Visualization
  - Workshop: Introduction to Information Visualization

"Fall 2016 CS 7450"
**Flipped Readings**

- No reading assignment previewing next class’ material
- Instead, reading of papers and viewing of videos/websites from current class for next time
  - Potential pop quiz at start of next class

Look for R on Schedule webpage

---

**Grading**

- Participation
- Pop Quizzes
- Assignments
  - HWs (about 7)
- Project
- Final Exam

- (Details still being finalized)
Survey

• Who wasn’t here on Monday and didn’t fill out a survey?

• If you want to change your “I’m on WL and not sure I want in” to “I’m on WL and I definitely want in” see me after class

***CAVEAT***

• This course is a lot of work. If you’re just looking for some easy grade, I would advise you to drop now.

• If you are sincerely interested in this topic, I hope you will enjoy the course and learn a lot
Waitlist/Overload Update

- Status
  - Room capacity
  - Let in 15 yesterday
  - 29 of original 60 said “might drop”
  - Will let in more tomorrow
  - Please drop by Thursday noon
  - On Friday it becomes the “Wild West”

Electronics

- This is now a NO laptops/cellphones class
- Exceptions will be noted (exercises, etc)
- Note-takers, see me

Reminder
Purpose

• Two main uses of infovis
  – Analysis – Understand your data better and act upon that understanding
  – Communication – Communicate and inform others more effectively
1. Analysis

- Given all the data, then
  - understand, compare, decide, judge,
    evaluate, assess, determine, ...

- Ultimately, about solving problems

When to Apply?

- Many other techniques for data analysis
  - Statistics, DB, data mining, 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

"A graphic display has many purposes but it achieves its highest value when it forces us to see what we were not expecting."

H. Wainer
EDA Example 1

- Business
  - Why has Hyundai made such great strides in the US market?
  - How influential was their “Lose your job, we’ll buy the car back” campaign?
  - Have their cars improved in quality? If so, in what major ways?
  - Is the Genesis as good of a car as the Lexus ES?
EDA Example 2

- Airlines
  - What are the key factors causing flight delays in the US?
  - Are delays worse in the summer or winter?
  - Is the seasonal effect influenced by geographic location?
  - How does competition at an airport affect flight delays?

More on EDA

“Information visualization is ideal for exploratory data analysis. Our eyes are naturally drawn to trends, patterns, and exceptions that would be difficult or impossible to find using more traditional approaches, such as tables or text, including pivot tables. When exploring data, even the best statisticians often set their calculations aside for a while and let their eyes take the lead.”

S. Few

Now you see it
2. Communication

- Use visualization to communicate ideas, present, influence, explain, persuade

- Visuals can serve as evidence or support

When to Apply?

- Visuals can frequently take the place of many words
- Visuals can summarize, aggregate, unite, explain, ...

- Sometimes words are needed, however
Key Benefits of Visualization

- Facilitating awareness and understanding
- Helping to raise new questions and supply answers
- Generating insights
- Telling a story and making a point

Key Challenge

- How to measure and prove?
  - All those benefits are not easily quantifiable and measured

- Evaluation is perhaps primary open research challenge for visualization

More to come later in term
Data Analysis is cool

Academic Areas

- Where does InfoVis fit in the academic world?
Overview

Visualization
“Data visualization”

Scientific visualization
Information visualization

Scientific Visualization

- Primarily relates to and represents something spatial, 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

Not the focus of this class
Information Visualization

1. What is “information”?
   - Non-spatial data: 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...

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

- Characteristics:
  - Taking things without a direct physical correspondence (non-spatial) and mapping them to a 2-D or 3-D physical space
  - Giving information a visual representation that is useful for analysis and presentation
  - "A key challenge in information visualization is designing a cognitively useful spatial mapping of a dataset that is not inherently spatial and accompanying the mapping by interaction techniques that allow people to intuitively explore the dataset. Information visualization draws on the intellectual history of several traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, statistical graphics, cartography, and art."

http://conferences.computer.org/infovis/

Constituents

- Two key aspects of infovis
  - Representation
  - Interaction (too often overlooked)

"The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means."

S. Few, Now you see it
Two Key Challenges

- Scale
  - Challenge often arises when data sets become large

- Diversity
  - Data of data types, forms, sizes

Example Domains for Info Vis

- Text
- Statistics
- Financial/business data
- Internet information
- Software
- ...
InfoVis Process Model

- Raw data → Data tables → Visual Structures → Views
- Data transformations → Visual mappings → View transformations

From: Card, Mackinlay, Shneiderman '99

New Area Emerging: Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces

Available at http://nvac.pnl.gov/ in PDF form

More to come later in term
Back to InfoVis (Examples)

- Start with static pictures (InfoGraphics)
  - Very popular on the web
  - But are they information visualizations?

NYC Weather

2220 numbers

Tufte, Vol. 1

Data Values

365  High temp for each day
365  Low temp for each day
365  Avg high temp for each day
365  Avg low temp for each day
365  Precipitation for each day
365  Humidity for each day
12  Precipitation for each month
12  Avg precipitation for each month
1  Precipitation for the year
1  Avg precipitation per year
1  Highest temp (& day) for the year
1  Lowest temp (& day) for the year
1  Avg daily temp for the year
1  Avg daily temp per year

Updated Version

http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00014g
Income and Religion

http://awesome.good.is/transparency/web/1002/almighty-dollar/transparency.jpg
Population

Where We Live...

Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie.


Atlanta Flight Traffic

Atlanta Journal
April 30, 2000
Figure 14. States Mentioned in Country-Music Lyrics
Note: The size of each state is proportional to the number of times it is mentioned.

From E. Tufte
The Visual Display of Quantitative Information
Or, for fun...

Percentage of Chart Which Resembles Pac-man

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

http://xkcd.com/197/
A "Bar" Chart

Drunkness

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...
Baby Name Wizard

http://babynamewizard.com/namevoyager/

NY Times

- Has been a wonderful source of interactive data visualizations

- Some examples...

Naming Names
Names used by major presidential candidates in the series of Democratic and Republican debates leading up to the Iowa caucuses.

- Each thin line represents one candidate speaking the last name of another candidate.
- Every line ends at an arrow, which points to the name that was spoken.
- The length of each circle segment represents the total number of words spoken by the candidate during the debates. Each tick mark represents 1,000 words.

Source: Election Transcripts
Fall 2016 CS 7450

A Peek Into Netflix Queues

- Examine Netflix rental patterns, neighbor by neighbor, in a dozen cities. Some titles with distinct patterns are Med Men, Obsessed and Lost Chance Harvey.

Source: Netflix
Fall 2016 CS 7450
Good Resources

• Some places to look for more information

InfoVis Wiki

http://www.infovis-wiki.net
Scoop.It!

http://www.scoop.it/t/data-visualization-by-guilhes-damian

Graphic Detail - Economist

http://www.economist.com/blogs/graphicdetail
A Compendium

Learning Objectives

- Articulate definition and purpose of visualization
- Describe two main uses or applications of visualization
- List two primary components of visualizations
- Describe the different areas of academic visualization research
- Explain the infovis “pipeline” (process)
HW

- HW1 due next Monday
  - Data Exploration and Analysis
  - Bring 2 hardcopies

Reading

- Card, Mackinlay, Shneiderman
  Chapter 1 of their book
- Check out some of the websites on the Schedule page
Upcoming

• Multivariate data & tables

• Graphs & Charts