Overview of InfoVis

CS 4460 – Intro. to Information Visualization
Aug. 23, 2017
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)

(Will carry over into next class)
Exercise

- Get out pencil and paper

Data
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

Data Transfer

Data

How?

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...
<table>
<thead>
<tr>
<th>Cereal</th>
<th>Manufacturer</th>
<th>Fiber</th>
<th>Potassium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honeycomb</td>
<td></td>
<td>P</td>
<td>0</td>
</tr>
<tr>
<td>Just Right Fruit &amp; Nut</td>
<td></td>
<td>K</td>
<td>2</td>
</tr>
<tr>
<td>Lilt</td>
<td></td>
<td>Q</td>
<td>2</td>
</tr>
<tr>
<td>Lucky Charms</td>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>Maya</td>
<td></td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>Multigrain Raisin, Robots &amp;</td>
<td></td>
<td>R</td>
<td>3</td>
</tr>
<tr>
<td>Nut-Gran Almond-Frutz</td>
<td></td>
<td>K</td>
<td>3</td>
</tr>
<tr>
<td>Nutri-grain Wheat</td>
<td></td>
<td>K</td>
<td>3</td>
</tr>
<tr>
<td>Oatmeal Flakes Crisp</td>
<td></td>
<td>G</td>
<td>1.5</td>
</tr>
<tr>
<td>Post Nix Raisin Bran</td>
<td></td>
<td>P</td>
<td>6</td>
</tr>
<tr>
<td>Product 19</td>
<td></td>
<td>K</td>
<td>1</td>
</tr>
<tr>
<td>Quaker Oatmeal</td>
<td></td>
<td>Q</td>
<td>3.7</td>
</tr>
<tr>
<td>Raisin Bran</td>
<td></td>
<td>K</td>
<td>5</td>
</tr>
<tr>
<td>Raisin Nut Bran</td>
<td></td>
<td>G</td>
<td>2.5</td>
</tr>
<tr>
<td>Rice Krispies</td>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>Shredded Wheat</td>
<td></td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>Shredded Wheat 'n Bran</td>
<td></td>
<td>N</td>
<td>4</td>
</tr>
<tr>
<td>Shredded Wheat spoon</td>
<td></td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>Special K</td>
<td></td>
<td>K</td>
<td>1</td>
</tr>
<tr>
<td>Strawberry Fruit Wheats</td>
<td></td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>Total Corn Flakes</td>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>Total Raisin Bran</td>
<td></td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>Total Whole Grain</td>
<td></td>
<td>G</td>
<td>3</td>
</tr>
<tr>
<td>Tins</td>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>Wheats</td>
<td></td>
<td>G</td>
<td>3</td>
</tr>
<tr>
<td>Wheaties Honey Gold</td>
<td></td>
<td>G</td>
<td>1</td>
</tr>
</tbody>
</table>

Questions:

1. Which cereal has the most/least potassium?
2. Is there a relationship between potassium and fiber?
3. If so, are there any outliers?
4. 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

Anscombe’s quartet
http://astro.swarthmore.edu/astrol21/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, 8.33</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, 5.73</td>
<td>8.0, 6.89</td>
<td></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

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
# Visuals Help Us Think

- Provide a frame of reference, a temporary storage area
- Cognition $\rightarrow$ 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

• Course policies
• Grading
• Overloads
• Surveys
• More...
Administratia

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

http://www.cc.gatech.edu/~stasko/4460

T-Square Site

Should've seen an announcement
Responsibility

• You are responsible for examining and staying up-to-date with information on the class website and t-square website
Course Policies

- Most on class homepage or Assignments page
  - Format
  - Book
  - Attendance
  - Electronics
  - Academic Integrity
  - Grading
  - ...

Books

None required

Will use

Ebook free at http://chimera.labs.oreilly.com/books/1230000000345/
Colored Pencils

Please get some and bring to class when requested

Attendance

- Expected
  - Part of your grade
- Will start promptly at 12:20 and end by 1:10

- Eating (quietly) is OK
Academic Integrity

- Do your own work, unless told otherwise
- Absolutely OK to consult me or Tas
- More to come on programming HWs...

Electronics

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

- Pop quizzes/Attendance 10%
  - Start of class, 5 minutes
- HW Assignments (5) 20%
- Programming Assignments (5) 28%
  - Last one bigger
- Midterm Exam 20%
- Final Exam 22%

At worst 60-70-80-90 for final grades

Schedule

3 types classes
- Lecture/Discuss
- Design
- Lab

Prep material for next class
Link to more info
HW links
**Schedule**

- Slides from class
- Prep material
- Related papers
- Links from class (websites, videos, demos, etc.)

**Class Preparation**

- Paper, video, website to review for the next class
  - Multiple links from Schedule page
  - Potential pop quiz at start of next class
Assignments

Recaps weights

Will have links to HWs

Instructor

Office number out of sequence with others (in corner of building)
About Me

CS PhD, 1989, Brown University

Researcher in information visualization, visual analytics, and HCI

Active in those academic research communities

Teaching Assistants
**CAVEAT**

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

- Graduating seniors: It’s on you now to do the work so no problems later.

- If you are sincerely interested in this topic, I hope you will enjoy the course and learn a lot.

Registration

- Will be able to add some students
  - Room has a little more capacity
  - More if there are drops
  - Cannot go too large

- Not a fan of “seat squatting”

- Please drop the class by Thursday noon.
Survey

• Who wasn’t here on Monday and didn’t fill out a survey?
  – Please fill out even if on waitlist
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

- 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?

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
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

From: Card, Mackinlay, Shneiderman '99
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

http://www.mikewirthart.com/?cat=3

Beer
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 north and west, but compared with China or India, the nation is a vast place.


Atlanta Flight Traffic

Atlanta Journal
April 30, 2000
Country Music

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

Napoleons’s March

Minard graphic

From E. Tufte
The Visual Display of Quantitative Information

size of army
direction
latitude
longitude
temperature
date

Fall 2017 CS 4460 77
Or, for fun...

Percentage of Chart Which Resembles Pac-man

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

http://xkcd.com/197/

Fall 2017  CS 4460  81

Fall 2017  CS 4460  82
A "Bar" Chart
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)
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...
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

Fall 2017 CS 4460

Upcoming

• Design Exercise
  – Bring colored pencils

• Data & Tables