Information Visualization
CPSC 683

Sheelagh Carpendale

Overview for today

• Project
• Project data sets

• Introductions
• Discussion - information visualization
CPSC 683  
Information Visualization: Theory & Practice

- Monday 15:00-17:50
- MS 680J 220-6055,
- sheelagh@cpsc.ucalgary.ca,
- Office Hours: TR 10:45 – 11:45

Marks
- In Class Participation 20%
- Visual Journal 20%
- Project 60%

In Class Participation
(worth 20%)

- You will be expected to take part in discussions
- Most weeks will be conducted as seminars
- You will have assigned readings and part and/or questions in regards to the readings which you will discuss in class
- Sometimes we will also do data sketching in class – these will be followed by a discussion in which you will be expected to take part.
Keeping a Visual Journal
Highly recommended!! (worth 20%)

- Sketch: ideas, data, concepts
- Collect
  - images that others have created for information visualization and/or visual communication in general
  - chosen because you liked/disliked them or because you can not figure out some else’s reaction
- React
  - why and which parts you like or dislike
  - annotate your images, draw on them, write on them, look at them in different scales
  - which parts might you use yourself, as is, with changes, or never!
- Generate
  - keeping track of your developing ideas
  - including, scribbles, sketches, math, and words

In class data-sketch exercises
Structure:
- Part 1 (innovation part):
  - creation in class of a visual representation of small data set
  - encourage creation, innovation, risk taking – ‘just do it’
  - sketches will be handed in at the end of class and photographed to create a record
  - like a quiz in that you must be present and take part
  - not like a quiz in that you do not have to memorize facts
In class exercises

Structure:
• Part 2 (review skills, the ‘crit’):
  – You will be able to discuss your representations and to discuss your response to your classmates representations
  – sometimes using specific criteria that have been taught in class
  – for all you will discuss both the positive and negative aspects of the sketch and make suggestions for improvements

Course Project

*Projects must be approved*. Two possible types of project

1) Visualize 5 aspects of your own data. Data that affects your life or that you generate. Synthesize into a single application

Marks break down

<table>
<thead>
<tr>
<th>Part</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project proposal</td>
<td>part of project synthesis</td>
</tr>
<tr>
<td>Part #1 through part #5</td>
<td>15% each 75%</td>
</tr>
<tr>
<td>Project synthesis</td>
<td>25%</td>
</tr>
</tbody>
</table>

2) A normal information visualization project will consist of creating a visual representation for your chosen information/data, organizing the presentation of this representation on the display space and implementing this with interaction methods.

Marks break down

<table>
<thead>
<tr>
<th>Part</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project proposal</td>
<td>part of project report</td>
</tr>
<tr>
<td>The project itself</td>
<td>80%</td>
</tr>
<tr>
<td>Project report</td>
<td>20%</td>
</tr>
</tbody>
</table>
Information Visualization

Possible data sets for your projects

Movie Data (infovis contest 2007)

- all US movies released from 2000 to 2006
- for each movie
  - director(s)
  - cinematographer(s)
  - first ten billed actors and actresses
  - genre(s)
  - US release date
  - IMDB ratings
- Format: XML

Note: NetFlix data are missing
http://eagereyes.org/InfoVisContest2007Data.html
MovieLens data sets

- 100,000 movie ratings
- 943 users (each user rated at least 20 movies)
- 1683 movies
- Format: tab separated list

global peace index (GPI)

- GPI rankings for 140+ countries (2007)
- 24 peace indicators, e.g.
  - # of external and internal conflicts fought
  - level of disrespect for human rights
  - military expenditure ...
- 32 related indicators
  - Freedom of the press
  - GDP per capita
  - Life expectancy
- Format: XML
  http://www.visionofhumanity.org/gpi-data/#/2009/scor
book data base

• 2000 books from freebase library
  – author
  – title
  – cover image
  – genre
  – date of first publication
  – abstract
• Format: Json

Desktop input

• interaction input logs - keyboard and mouse
• data can include:
  – all keyboard events (e.g. letters, numbers, punctuation, corrections, backspaces, etc.), mouse events (e.g. clicks, location, etc.)
  – related temporality of these events (e.g. rhythm, frequency, velocity, acceleration, etc.)
• possible purposes include:
  – identity comparisons,
  – augmenting with personality,
  – development of digital signatures.
other data sources

• freebase.com
  – huge variety of open data
  – JSON format

• NY Times API http://developer.nytimes.com/docs
  – for querying the NY Times data base

• Twitter API http://apiwiki.twitter.com/
  – for querying the twitter data base

• UN data http://data.un.org/

• infochimps.org – datasets uploaded by users

Music Listening Dataset

• 150,000 people profiles including
  – artists they listen to
  – how much each artist was played

• Format: tab separated list

audioscrobbler data
Language: Processing

- Introduction during tutorial
- open source programming language
- Web site: http://processing.org/

Why do Information Visualization?
• 2008 It is estimated that 800 exabyte ($800 \times 10^{19}$) of digital information will be generated this year
• 2010 – 1 zettabyte = 1 trillion gigabytes
• 2011 to date 1.8 zettabytes

"The ability to take data to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it that's going to be a hugely important skill in the next decades."

Hal Varian, chief economist at Google
Example: Anscombe’s Quartet

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>7.46</td>
</tr>
<tr>
<td>y</td>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>6.77</td>
</tr>
<tr>
<td></td>
<td>10.0</td>
<td>15.0</td>
<td>15.0</td>
<td>12.74</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>7.11</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>7.81</td>
</tr>
<tr>
<td></td>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.84</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>7.74</td>
<td>6.0</td>
<td>6.03</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>8.15</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>6.42</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>5.73</td>
</tr>
</tbody>
</table>

For all 4 columns the stats are identical.

Statistical Analysis

Mean of x: 9.0
Variance of x: 11.0
Mean of y: 7.5
Variance of y: 4.12
Correlation between x and y: 0.816
Linear regression line: \( y = 3 + 0.5x \)
Visual representation reveals a different story

Why visual data representations?

- Vision is our most dominant sense
- We are very good at recognizing visual patterns
- We need to see and understand in order to explain, reason, and make decisions

common examples:

- graphs / hierarchies
- charts
- maps

all examples from: [http://vis.stanford.edu/protovis/]
benefits

- expand human working memory
  - offload cognitive resources to visual system
- improve search
  - large amount of data in small space
- enhance patterns recognition
  - making patterns visual explicit
- aids monitoring of incoming events
- manipulable medium supports exploration

"One hundred rumors are not comparable to one look."
An Old Chinese Inscription

Via Brinton, Graphic Presentation, 1939
Information visualization

- Create visual representation
- Concentrates on abstract data
- Includes interaction

Common definition:
The use of computer supported, interactive, visual representations of abstract data to support cognition. [Card et al. 1999]

Functions of visualization

- recording information
  - tables, blueprints, satellite images
- processing information
  - needs feedback and interaction
- presenting information
  - share, collaborate, revise
  - for oneself, one’s peers, and to teach
- seeing the unseen
Information visualization research field

- ~25 years old (computer infovis)
- from scientific visualization, computer graphics, HCI
- Major conferences
  - IEEE VisWeek (Vis, InfoVis, VAST)
  - EG EuroVis
  - ACM CHI (general HCI)

**Information vs. Scientific Visualization**

**InfoVis**
- focus on abstract data
- position of data in space can be chosen freely
- typical techniques:
  - (we will see later)

**SciVis**
- focus on scientific data
- position of data in space is typically fixed
- typical techniques
  - flow visualization
  - volume visualization (CT, MRT)

Different spatial graph layouts of abstract data

Spatial representation of gravity waves & a PET scan

There exists a gray region between both fields (e.g. maps)
Visual Analytics

- Related to both information visualization and scientific visualization
- Focus on whole data analysis process

Visualization of abstract data has been practiced for hundreds of years...

HISTORICAL EXAMPLES
The Broadway Street Pump

- In 1854 cholera broke out in London
  - 127 people near Broad Street died within 3 days
  - 616 people died within 30 days
- People thought it spread by “miasma in the atmosphere”
- Dr. John Snow was the first to link contaminated water to the outbreak of cholera
- How did he do it?
  - he talked to local residents
  - identified a water pump as a likely source
  - used maps to illustrate his theory
  - convinced authorities to disable the pump

More info here: http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak
Napoleon’s March on Moscow  Charles Minard, 1869

Named the best statistical graphic ever drawn (by Edward Tufte)
- Includes: spatial layout linked with stats on: army size, temperature, time
- Tells a story in one overview

C. J. Minard, French engineer, 1851
Combined statistical diagrams and maps

More info: The Visual Display of Quantitative Information (Tufte)
Influencing Hospital Management

Florence Nightingale
English nurse, 1858

Diagrams of motion

- Using white tape and black velvet, Marey created time series images.
- E. J. Marey, (1830 – 1904)

Comparative Scheduling

- E. J. Marey. 1885. Train schedules from Paris to Lyon
- Stations spaced according to distances, time from left to right

1981 – new express train – trip now 3 hours instead of 9
MORE RECENTLY

http://www.zeit.de/datenschutz/malte-spitz-data-retention
TrashTrack

- Winner of the NSF International Science and Engineering Visualization Challenge.
- http://senseable.mit.edu/trashtrack/

http://vimeo.com/23554190
Artificial Intelligence

http://www.turbulence.org/spotlight/thinking/chess.html

Open Data

- Movement making government data freely available
- Encourage participation by everyone

OECD Better Life Index: http://www.oecdbetterlifeindex.org/
Family Trees

http://www.aviz.fr/geneaquilts/

Weather

http://weatherspark.com/#!dashboard;a=Canada/AB/Calgary
It is difficult to create

CREATE VISUALIZATIONS

Mis-interpretation is common
AND ALL WERE IN THE WRONG!

Good Presentation Should Be Susceptible to Only One Interpretation

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant—
(Though all of them were blind—)
That each by observation
Might satisfy his mind.

The First (side) "Is very like a wall!"
The Second (tusk) "Is very like a spear!"
The Third (trunk) "Is very like a snake!"
The Fourth (trunk) "Is very like a snake!"
The Fifth (ear) "Is mighty like a fan!"
The Sixth (tail) "Is very like a rope!"

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong.
Though each was partly in the right,
And all were in the wrong!