CPSC 583
Fall 2010
Information Visualization

Sheelagh Carpendale
Overview for today

• Course structure
• Project
• Project data sets

• Introductions
• Discussion - information visualization
CPSC 583: Course Structure

• Class
  when:  Monday & Wednesday 10:00 to 11:15
  place:  MS156

• Tutorial (Uta Hinrichs)
  when:  Monday & Wednesday 11:30 to 12:45
  place:  MS156

• Office hours
  when:  Monday & Wednesday 12:45 to 13:45
  place:  MS680J

• Evaluation
  In class exercises  25%
  Project  75%

• Contact information
  sheelagh@ucalgary.ca
  220-6055
  office MS680J
In class exercises

Purpose:
• develop representation mapping skills
• create and recognize useful information visualizations

Evaluation:
• 5 of these exercises worth 5% each for a total of 25% of your course mark

• Each exercise will have 2 parts.
  – Part 1 (innovation part): worth 2% each (total 10% for all 5 of these in the term)
  – Part 2 (review skills, the ‘crit’): worth 3% each (total 15% for all 5 of these in the term)
In class exercises

Structure:
• Part 1 (innovation part):
  – creation in class of a visual representation of small data set
  – full marks will be awarded for doing this exercise
  – 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’):
  – short 1 to max 2 page review a sketch that is not your own
  – hand in at the beginning of the next class
  – 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

• One project - 75% of grade
• Parts marked at regular intervals throughout the semester

Break down:

• **Project proposal**  5%  Wed. Sept. 22\textsuperscript{nd}
• Representation sketches  10%  initial Sept. 29\textsuperscript{th}, full Mon. Oct. 4\textsuperscript{th}
• Presentation sketches  10%  initial Oct. 20\textsuperscript{th}, full Mon. Oct. 25\textsuperscript{th}
• Interaction sketches  10%  initial Nov. 3\textsuperscript{rd}, full Mon. Nov. 8\textsuperscript{th}
• Implementation/demo  30%  Wed. Nov. 30\textsuperscript{th}, 23:59 pm
• Progress portfolio  10%  Wed. Dec. 8\textsuperscript{th}
• Project report  25%  Wed. Dec. 8\textsuperscript{th} (all parts)
  – Written report  10%
  – Video  10%
  – Presentation  5%
Project proposal

- Project proposal  5%  Wed. Sept. 22\textsuperscript{nd}
- *Projects must be approved*.

- 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.

- 1 page (2 max)
- data/information source
- at least one (two or three is better) scenarios describing how someone would use a visualization of your chosen data.
Information Visualization

CPSC 583
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

book data - freebase.com
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
  - for querying the NY Times database
- **Twitter API** [http://apiwiki.twitter.com/](http://apiwiki.twitter.com/)
  - for querying the twitter database
- **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/
• Book: http://www.oreilly.com/catalog/9780596514556/toc.html
Course outline:

• Week 1
  • Monday: Introduction
  • Wednesday: Ex #1

• Week 2
  • Monday: Visual representation #1 (visual variables)
  • Monday: Tutorial: Ex #1 discussion
  • Wednesday: Visual representation #2: representing trees and graphs
  • Wednesday: Tutorial: Introduction to Processing

• Week 3
  • Monday: Visual representation #3 (excellence & disasters)
  • Monday: Tutorial: Introduction to basic interaction
  • Wednesday: Visual representation #4: (2D rep)
  • Wednesday: Tutorial: representation sketches discussion
Course outline:

• Week 4
  • Monday : Ex #2
  • Monday: Tutorial: colour and transparency
  • Wednesday: Presentation #1 screen space
  • Wednesday: Tutorial: Ex #2 discussion

• Week 5
  • Monday : Thanksgiving
  • Wednesday: Presentation #2: (lenses)
  • Wednesday: Tutorial: simple animation

• Week 6
  • Monday : Presentation #3: (layering)
  • Monday: Tutorial: lenses
  • Wednesday: Presentation #2: (zooming)
  • Wednesday: Tutorial: presentation sketches discussion
Course outline:

- Week 7
  - Monday: Ex #3
  - Monday: Tutorial: discrete lenses
  - Wednesday: Interaction #1 dynamic queries
  - Wednesday: Tutorial: Ex #3 discussion

- Week 8
  - Monday: Interaction #2: brushing and linking
  - Monday: Tutorial: interactive filtering
  - Wednesday: Interaction #3: (drill down)
  - Wednesday: Tutorial: Interaction sketches discussion

- Week 9
  - Monday: Ex #4:
  - Wednesday: Interaction #4: (juxtaposition)
  - Wednesday: Tutorial: Ex #4 discussion
Course outline:

• Week 10
  • Monday: cognitive science and interaction
  • Wednesday: evaluating infovis

• Week 11
  • Monday: Ex #5
  • Wednesday: applications #1: (bio vis)
  • Wednesday: Tutorial: Ex #4 discussion

• Week 12
  • Monday: applications #1: (bio vis)
  • Wednesday: use of 3D
  • Wednesday: Tutorial: project demos

• Week 13
  • Monday: Infovis – open questions
  • Monday: Tutorial: video
  • Wednesday: project presentations
Keeping a Visual Journal

Highly recommended!!

• 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 cannot 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
• Can go in your progress report
Why do Information Visualization?
Mis-interpretation is common
Anscombe’s Quartet

N: 11.0
mean X’s : 9.0
mean Y’s : 7.5
standard error of slope estimate: 0.1
sum of squares: 110.0
regression sum of squares: 27.5
residual sum of squares of Y: 13.8
correlation coefficient: 0.8
r squared: 0.7
regression line: Y=3+0.5X
John Snow
1854

- Discovered cause of a cholera epidemic through visualizing the data in conjunction with a map.

Influencing Hospital Management

Florence Nightingale
English nurse, 1858

Diagram by Florence Nightingale, corrected by Hugh Small

The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex. The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes. The black lines across the red triangles in Sept & Nov 1854 mark the boundaries of the deaths from all other causes during those months. In October 1854, April 1855, & November 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black. The entire areas may be compared by following the blue, the red & the black lines enclosing them.
C. J. Minard, French engineer, 1851
Combined statistical diagrams and maps
Infosthetics

Visual Complexity

Data Vis Blog

- http://datavisualization.ch/