Knowledge Visualisation
Introduction

Week 1 Lecture 1
IAT 814
Lyn Bartram
Introductions

Instructor

• Lyn Bartram
• lyn@sfu.ca

Research

• Information visualization
• Perception, attention and visual representations
• Visualization for sustainable living/Ambient visualization
• Motion and Animation visual techniques
• And generally interactive visual systems
What is visualization?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

-- Tamara Munzner, 2014
What’s the difference here?
Illustration or visualization?

Illustration

- a visual representation (a picture or diagram) that is used to make some subject more pleasing or easier to understand
Illustration or visualization?

Visualization **communicates data.**

- from something that is abstract or at least not immediately visible (like the inside of the human body)....
- Visualization transforms from the invisible to the visible.
  ---- Robert Kosara

- Visualization is about MAPPING.
What is visualisation?

- [**geek**] branches of computer graphics and UI design concerned with presenting data to users by means of images
  - A computational tool or method for generating images

- [**psychologist**] The formation of mental visual images
  - The act or process of interpreting in visual terms

- [**designer**] The process of putting into visual form
  - The craft of assigning representational “codes” and techniques to data attributes and conveying meaning
  - The practice of assembling images
Some of the things we will cover in this course

• How we see and “think visually”
• Contrast and colour
• PreAttentive: things that pop out
• Motion and Animation
• Visual Awareness and Attention
• Space perception
• Interaction and Navigation
• Images, words and gesture

• Visualization in practice: scientific, information and affective
• 2D Visualization techniques
• Single and multiple views
• Interaction
• Transforming data and visualizing multidimensional data
Administrivia

- **Course textbook:**
  - Colin Ware. Visual Thinking for Design

- **Course web site**
  - [www.sfu.ca/siatclass/IAT814/Fall2014](http://www.sfu.ca/siatclass/IAT814/Fall2014)

- **Course management site**
  - canvas.sfu.ca
Course Evaluation

• 3 components:

1. 2 Visualisation critiques (20%, 10% each)
   • Part 1. Find an example of practice, critique and present
   • Part 2. Discuss from perceptual perspective!
   • We will revisit these.

2. Research critique (20%)
   • Critical review of recent research, presentation and report

3. Final project (60%)
   • Research paper
   • Design proposal of technique or method to address a visualization challenge
   • Study of existing techniques applied to visualization/analytics problem
Policies and expectations

- Participation
  - Graduate classes are most effective when grad students engage.

- Academic honesty
  - SFU Code of Conduct
  - Plagiarism will result in automatic 0

- Publication. Students are often interested in publishing the results of their projects and I thoroughly encourage this effort
  - the work is reviewed with me and deemed appropriate and a venue selected;
  - the senior supervisor is consulted and gives explicit permission for the extra effort required in preparing work for publication;
  - the authorship includes the senior supervisor and myself unless either of us waives it.
Policies and expectations

• Project work across multiple courses.
  • In principle I have no problem with the same research project serving two different courses, but this means that the actual work is much bigger than required in either course, and that the applicability of the visualization component of the work suits the nature of this course. This must be cleared with me and the instructor of the other course before the project is approved. Failure to do so will result in the project being marked out of 50%.
Why do we care?
Why?

• We are in a new era of human history:
  • Since 1994 we have witnessed an information explosion
    • Everyone can get all of the data that’s out there
      • News, sports, financial, purchases, etc...

• What do we do with it?
Kegs of data

- Between 1 and 2 exabytes of unique data produced per year
  - $1000000000000000000$ (10^{18}) bytes
  - 800 meg for every person (2003)
  - Printed documents only .003% of total
    Lyman and Varian, 2000
    Cal-Berkeley, Info Mgmt & Systems
    www.sims.berkeley.edu/how-much-info
- 90 trillion emails sent on the Internet in 2009
More data generated in next 4 years than in the history of the world

INFORMATION OVERLOAD

People are connected up to 12 hours a day:
- 2 million bits per second
- IT Social Media users generating data at an exponential rate
- Amount of Information That Is Unique: 25%
- Human Absorption Rate: 126 Bits per second
- Rate humans can listen: 40 Bits per second
- 29 hours
- A lifetime of learning can be transferred over the Internet.
- Internet Audience: unique visitors
  - Asia: 400 million
  - Worldwide: >1 billion

AMOUNT OF DATA IN 2010
- 1.2 zettabytes

AMOUNT OF DATA IN 2020
- 35 zettabytes

LIMITED INFORMATION INTAKE

Time constraints:
- 24 hours a day = 1,440 minutes a day
- Average minutes spent shopping a day in the U.S.:
  - 1965: 34
  - Today: 28

Americans spend < 3% of waking hours actually buying.

Time spent shopping has not changed in 4 decades
• 6 million FedEx transactions per day

• Average of 98 million Visa credit-card transactions per day in 2005
  http://www.corporate.visa.com/md/nr/press278.jsp

• 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
### SAP prediction

#### The data explosion - unwound

<table>
<thead>
<tr>
<th>Time frame</th>
<th>Data volume growth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In 2010</strong></td>
<td>- 1200 exabytes of data</td>
</tr>
<tr>
<td><strong>In 2011</strong></td>
<td>- 1.8 zettabytes of data</td>
</tr>
<tr>
<td><strong>In 2020</strong></td>
<td>- 35 zettabytes will exist</td>
</tr>
</tbody>
</table>

1 million terabytes = 1 exabyte  
1000 exabytes = 1 zettabyte

Data from *The 2011 Digital Universe Study: Extracting Value from Chaos*, by IDC.
Unlike Before…

• Used to be (15 years ago), you had to go to a library
  • read the info, put it on some sort of storage device, take notes, run a specialized program
  • On a computer 1000 times slower than today

• Now:
  • 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 only reasonable solution

- Computing + Human Vision
  - Highest bandwidth sense
  - Fast, parallel
  - Pattern recognition
  - Extends memory and cognitive capacity
  - Many People think visually
When do we need to visualize?

• 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 did Toyota outpace Chrysler in the auto market?
• How did their business decisions differ?
• Have their cars changed in design? If so, in what major ways?
EDA 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?
Visualizations make data into information

Causes of Mortality in the Army in the East
April, 1854 to March 1855

Non-Battle
Battle

From: Florence Nightingale, "Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army", 1858
At first glance: the Cain 9-9-9 tax proposal
Not the whole story
More is not necessarily better

**A closer look at Cain’s tax plan**

A new independent analysis of Republican presidential candidate Herman Cain’s 9-9-9 tax plan shows how it would impact different income groups:

<table>
<thead>
<tr>
<th>Income group</th>
<th>Average federal tax change under the 9-9-9 proposal, in dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than $10,000</td>
<td>$1,122</td>
</tr>
<tr>
<td>10,000-20,000</td>
<td>2,705</td>
</tr>
<tr>
<td>20,000-30,000</td>
<td>3,833</td>
</tr>
<tr>
<td>30,000-40,000</td>
<td>4,196</td>
</tr>
<tr>
<td>40,000-50,000</td>
<td>4,399</td>
</tr>
<tr>
<td>50,000-75,000</td>
<td>4,326</td>
</tr>
<tr>
<td>75,000-100,000</td>
<td>4,368</td>
</tr>
<tr>
<td>100,000-200,000</td>
<td>2,105</td>
</tr>
<tr>
<td>200,000-500,000</td>
<td>-11,155</td>
</tr>
<tr>
<td>500,000-1 million</td>
<td>-59,489</td>
</tr>
<tr>
<td>More than 1 million</td>
<td>-455,247</td>
</tr>
</tbody>
</table>

Source: Tax Policy Center  
Graphic: Judy Treible, Robert Dorrell  
© 2011 MCT  

Those who would pay less
Key benefits

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

- depict and present constantly growing large data volumes of increasing dimensionality
- Support the revealing and discovery of patterns
- Display and reveal complex relationships and inter-relationships
- Convey an impression, evoke an emotion -- affect
Visualization goals

- **✓** depict and present constantly growing large data volumes of increasing dimensionality
- **✓** Support the revealing and discovery of patterns
- **✓** Display and reveal complex relationships and inter-relationships
- **✓** Convey an impression, evoke an emotion -- affect
Using vision to think

• Comprehend huge amounts of data

• Emergent properties and relations

• Detect problems and inconsistencies in data

• Facilitates understanding of large- and small-scale features of the data

• Facilitates hypothesis formation: the forming of new questions and insights
Visualization goals: not quite there
So what is visualization again? (take 2)

Visual thinking involves:

• Constructing visual queries on displays
• Visual search strategies through eye movements and attention to relevant patterns
• Visual notification and attention “redirection” to new patterns and events

• Well structured balance of elements and tasks
Essential issues

• What **mental models** most effectively carry various kinds of information?
• Which definable and recognizable **visual attributes** of these models are most useful for conveying specific information either independently or in conjunction with other attributes?
• How can we most effectively induce chosen mental models in the mind of an observer?
• How can we provide guidance on choosing appropriate models and their attributes to a human or automated display designer?
  • ------- G. Robertson
• Progress in [scientific] visualization can be accelerated if workers could more readily find visualization techniques relevant to a given problem
  • ---Wehrend and Lewis
Essential issues

- What mental models most effectively carry various kinds of information?
- Which definable and recognizable visual attributes of these models are most useful for conveying specific information either independently or in conjunction with other attributes?
- How can we most effectively induce chosen mental models in the mind of an observer?
- How can we provide guidance on choosing appropriate models and their attributes to a human or automated display designer?

------ G. Robertson

- Progress in [scientific] visualization can be accelerated if workers could more readily find visualization techniques relevant to a given problem

--- Wehrend and Lewis
What makes/is a good visualisation?

• a **model** that captures the essence of a system
  • A model is an abstraction with the important things in and the unimportant out

• Different visualizations provide different levels of detail, show and hide different things; so support **different abstractions**

• useful to **aid understanding**, not just realistic representations (what color is a carbon atom?)

• map the important part of the **tasks** onto techniques that show the relevant **characteristics** best
Key challenges

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

- **Diversity**
  - Data of data types, forms, sizes

- **Evaluation**
  - How to measure and prove?
  - All those benefits are not easily quantifiable and measured
  - perhaps primary open research challenge for visualization
Scientific visualization

- the visual display of spatial data associated with scientific processes such as the bonding of molecules in computational chemistry

- deals with data that has a natural geometric structure (e.g. chemical data or wind flows)
Information visualization

- visual metaphors for non-inherently spatial data such as the exploration of text-based document databases.

- More abstract

- Assign structure and position to information that has none
Visual analytics

• analytical reasoning supported by the interactive visual interface

• Intersection of visualization with data analysis

• Biology
• Business
• National security
Affective (experience) visualisation
What is (Information) Visualization?

“Transformation of the symbolic into the geometric”

(McCormick et al., 1987)

“... finding the artificial memory that best supports our natural means of perception.”

(Bertin, 1983)
Visual Aids for Thinking

- We build tools to amplify cognition.
- Example: multiplication (Card, Moran, & Shneiderman.)
  - In your head, multiply 35 x 95
  - Now do it on paper
  - People are 5 times faster with the visual aid

![Bar chart showing time to multiply mentally versus using paper and pencil.](image)

**FIGURE 1.1**

Use of external aids amplifies ability to do multiplication.
We’ve been doing this for a long time
We look at the pattern(s) first

Focus on individual numbers … >2K data points!
Nightingale’s Coxcomb

Causes of Mortality in the Army in the East
April, 1854 to March 1855

Non-Battle
Battle

From: F. Nightingale, "Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army", 1858
The Power of Info. Visualization
Two key aspects

- Representation
- Interaction

“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
The Power of Visualization

1. Start out going Southwest on ELLSWORTH AVE Towards BROADWAY by turning right.
2. Turn RIGHT onto BROADWAY.
3. Turn RIGHT onto QUINCY ST.
4. Turn LEFT onto CAMBRIDGE ST.
5. Turn SLIGHT RIGHT onto MASSACHUSETTS AVE.
6. Turn RIGHT onto RUSSELL ST.
The Power of Visualization

The estimated travel time is 5 minutes for 2.16 miles of travel, total of 6 steps.

<table>
<thead>
<tr>
<th>Directions</th>
<th>Elapsed Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Begin at <strong>17 Ellsworth Ave</strong> on Ellsworth Ave and go Southwest for 0.1 miles</td>
<td>0.1</td>
</tr>
<tr>
<td>2 Turn right on <strong>Broadway</strong> and go Northwest for 0.4 miles</td>
<td>0.5</td>
</tr>
<tr>
<td>3 Turn right on <strong>Quincy St</strong> and go North for 0.5 miles</td>
<td>0.5</td>
</tr>
<tr>
<td>4 Turn left on <strong>Cambridge St</strong> and go West for 0.8 miles</td>
<td>0.8</td>
</tr>
<tr>
<td>5 Bear right on <strong>Massachusetts Ave</strong>, Mass Ave, RT-2A and go North for 2.0 miles</td>
<td>2.0</td>
</tr>
<tr>
<td>6 Turn right on <strong>Russell St</strong> and go Northeast for 1.2 miles to <strong>77 Russell St</strong></td>
<td>2.2</td>
</tr>
</tbody>
</table>

Line drawing tool by Maneesh Agrawala http://graphics.stanford.edu/~maneesh/
Visualization for Problem Solving

Mystery: what is causing a cholera epidemic in London in 1854?
Visualization for Problem Solving

Illustration of John Snow’s deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Illustration of John Snow’s deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Horizontal lines indicate location of deaths.

Visualization for Eliciting Knowledge from Data

- Which state has highest Income?
- Relationship between Income and Education?
- Outliers?
Visualization for Eliciting Knowledge from Data

![Graph showing the relationship between College Degree % and Per Capita Income for different states.](image-url)
What are the voids made of?
Visualization for Clarification

- London Subway Map Example
- Abstract away details for easier understanding
- Harry Beck’s schematic
London Underground Map 1927
Visualization for Clarification

- Horizontal, vertical and 45° segments
- Key insight: topology and relative location of stations
Visualization for telling a story
Visualization: Two Primary Goals

Analyze, Explore, Discover

Explain, Illustrate, Communicate

Express
A Good Use of TreeMaps and Interactivity
www.smartmoney.com/marketmap
Analysis vs. Communication

- MarketMap’s use of TreeMaps allows for sophisticated analysis
- Peets’ use of TreeMaps is more for presentation and communication
- This is a key contrast
Visualizations: Fuel Economy Treemap

Can't see the visualization? Download the latest Java plugin here. On Macs: best viewed in Safari.

Created by: Martin Wattenberg  Created on: Wednesday January 10, 10:33 AM

[Image of a treemap visualization showing different categories of fuel economy data.]

Explore
visualizations
data sets
comments
topic hubs

Participate
register
create visualization
upload data set
create topic hub

Learn more
quick start
visualization types
data format & style
about Many Eyes
FAQ
blog

Contact
contact us
report a bug

Legal
terms of use
Truck Sales Slip, Tripping Up Chrysler

Over the past few years, Chrysler executives said they were following the lead of Toyota and Honda, focusing on vehicles that met the needs of their customers. But as American consumers turned away from large trucks and S.U.V.s in 2006, Chrysler continued to churn out big vehicles, which are now sitting unsold at dealerships across the country.

The Chevrolet Impala, with or without flashing lights, did well in 2006, when a redesign came out.

The New York Times
lead of Toyota continued to the country.

**TRUCKS, VANS, S.U.V.'S | CARS**

<table>
<thead>
<tr>
<th>Toyota Tacoma</th>
<th>Toyota Sienna</th>
<th>Toyota Camry</th>
<th>Toyota Corolla</th>
</tr>
</thead>
<tbody>
<tr>
<td>178,361</td>
<td>163,269</td>
<td>448,445</td>
<td>387,288</td>
</tr>
<tr>
<td>Toyota RAV4</td>
<td>Toyota Highlander</td>
<td></td>
<td></td>
</tr>
<tr>
<td>152,047</td>
<td>129,794</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toyota Tundra</td>
<td>Toyota 4Runner</td>
<td>Toyota Prius</td>
<td>Scion xB</td>
</tr>
<tr>
<td>124,508</td>
<td>103,086</td>
<td>166,971</td>
<td>61,006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scion tC</td>
<td>Toyota Yaris</td>
</tr>
<tr>
<td></td>
<td></td>
<td>79,125</td>
<td>70,308</td>
</tr>
<tr>
<td>Lexus RX</td>
<td>Toyota, FJ Cruiser</td>
<td>Lexus ES</td>
<td>Scion xA</td>
</tr>
<tr>
<td>108,348</td>
<td>56,225</td>
<td>75,987</td>
<td>32,603</td>
</tr>
<tr>
<td>Toyota Sequoia</td>
<td>Toyota Avalon</td>
<td>Lexus IS</td>
<td>Lexus GS</td>
</tr>
<tr>
<td>34,315</td>
<td>88,938</td>
<td>54,287</td>
<td>27,290</td>
</tr>
</tbody>
</table>

**Change in sales from 2005 to 2006**

<table>
<thead>
<tr>
<th>Below</th>
<th>Above industry average</th>
<th>No 2005 sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10%</td>
<td>-2.6</td>
<td>0</td>
</tr>
<tr>
<td>+10</td>
<td>+100</td>
<td></td>
</tr>
</tbody>
</table>

*Many of these vehicles were introduced in 2005.*

**Toyota**

+12.5%

**Trucks/vans/S.U.V.'s** 1.1 million

**Cars** 1.5 million

Toyota rolled out a new version of the Camry, and once again it was the country's best-selling car.

Corolla sales also jumped, along with gas prices. Toyota could not escape the decline in sales of supersized S.U.V.'s like its Sequoia.
Other Taxonomies of Goals

- **Others:**
  - Analysis
  - Monitoring
  - Planning
  - Communication

- **Tufte:**
  - Description
  - Exploration
  - Tabulation
  - Decoration

- **Others:**
  - Aid to thinking
  - Problem solving/Decision making
  - Insight
  - Clarifying
  - Entertainment / Art

Ideas from this slide from Stone & Zellweger
New twists

• Implicit vs explicit

• Pragmatic vs artistic

• Ambient vs focused

• Casual vs expert …
Another way to think about it

- Answer this question: Do you know the answer?
  - If yes,
    - Presentation, communication, education
  - If no,
    - Exploration, analysis
    - Problem solving, planning,
    - Aid to thinking, reasoning

- Answer this question: Are you the creator or the viewer of the information?
  - Often there is a loop between analysis and presentation
Open Issues

• Does visualization help?
  • The jury is still out
  • Still supplemental at best for text collections
    • A correlation with spatial ability
    • Learning effects: with practice ability on visual display begins to equal that of text

• Does visualization sell?
  • Value of vis tools has exploded
  • Data become more ubiquitous

• Evaluation of GOOD visualizations critical
Utility

• Have we made advances in communication with computer assisted visualization?
  • Two of my favourite examples
Organizational Overview

-87  Now

New Nations
Presentation in the extreme

• Joh Cleese uses infovis to explain an issue of language
Assignment 1: Vis critiques

• Find two examples of visualization used in practice
  • One good, one bad
  • Example sources
    • Journal (Journal of Applied Optics, ...)
    • Science magazine (Nature, Science, Scientific American, ...)
    • News Magazine or newspaper (Newsweek, The Economist, NY Times, USA Today, ...)
    • Not from a vis paper, design textbook, design blog etc.
• Make a Canvas submission with critique
• 5-8 min presentation