Overview for today

• Course structure
• Project

• Introductions
• Discussion - information visualization
Grading break down

• In class exercises 10%
• Class participation 10%
• Project #1 20%
• Project #2 20%
• Project #3 20%
• Project synthesis 20%

Sketching exercises

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

exercises

Structure:
• Part 2 (review skills, the ‘crit’):
  – 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
Language: Processing

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

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 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
Why do Information Visualization?

- 2008
- 2010 – 1 zettabyte = 1 trillion gigabytes
- 2011 to date 1.8 zettabytes

It is estimated that 800 exabyte (800x10^19) of digital information will be generated this year.
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>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
</tr>
<tr>
<td>14.0</td>
<td>7.58</td>
<td>14.0</td>
<td>8.74</td>
<td>14.0</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
</tr>
<tr>
<td>14.0</td>
<td>8.46</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
</tr>
<tr>
<td>6.0</td>
<td>7.74</td>
<td>6.0</td>
<td>6.19</td>
<td>6.0</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
</tr>
<tr>
<td>12.0</td>
<td>10.04</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Example: Anscombe’s Quartet

Source: Anscombe’s quartet, Wikipedia

It’s not easy to get a handle on jobs in data science. However, data from O’Reilly Research shows a steady year-over-year increase in Cassandra and Cassandra job ratings, which are good proxies for the “data science” market as a whole. The graph shows the increase in Cassandra jobs, and the companies hiring Cassandra positions, over time.

“...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
Statistical Analysis

For all 4 columns the stats are identical

<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>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Y</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>X</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Y</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>X</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td>Y</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>X</td>
<td>14.0</td>
<td>14.0</td>
<td>14.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Y</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
</tr>
</tbody>
</table>

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

**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

(all examples from: http://vis.stanford.edu/protovis/)
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](image1.png)

![Spatial representation of gravity waves & a PET scan](image2.png)

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

---

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
  - 656 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

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

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

E. J. Marey, “La Méthode Graphique,” (Paris 1885), p.20. This method is attributed to the French engineer, Irby (Tufte, 1983, p.31)

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