Communicating With Tangible Abstractions
A quick intro to data visualization
Skye Bender-deMoll
skyebend@skyeome.net

Impact 2.0 - New mechanisms for linking research and policy
Kingston Jamaica, February 22-23, 2011
Why should we visualize data?

- To communicate concepts and conclusions more effectively
- To explore data that would otherwise be difficult to grasp
Why should we visualize data?

- To communicate concepts and conclusions more effectively
- To explore data that would otherwise be difficult to grasp
- Because everyone else is doing it
- It makes reports look fancier
- I don’t know but my boss asked me to do it
Making abstractions tangible

Many of us work with concepts, processes and programs that are very abstract, they are too big, too small, too fast, or too slow to be visible on a human scale.

A visualization should be a concrete example that conveys your abstraction.

- Something both you and the person you are communicating with can point to
- Should ideally show how you imagine the data or process, not just an arbitrary choice
At what level are you trying to reach people?
What is data anyway?

Semi-structured information

<table>
<thead>
<tr>
<th>Date</th>
<th>Name</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994 (probably)</td>
<td>Mr. John A. Smith</td>
<td>10,000</td>
</tr>
<tr>
<td>2001</td>
<td>Smith, John</td>
<td>250-500</td>
</tr>
<tr>
<td>3/5/2002</td>
<td>Johnson, Ann</td>
<td>8045.26</td>
</tr>
<tr>
<td>3.5.2002</td>
<td>Laura Rossi</td>
<td>2.3M</td>
</tr>
<tr>
<td></td>
<td>Perez, Jose</td>
<td>250,000</td>
</tr>
</tbody>
</table>

Structured data

<table>
<thead>
<tr>
<th>Year</th>
<th>Year Certainty</th>
<th>Name</th>
<th>US dollar amout</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>0.7</td>
<td>Smith, John</td>
<td>10000</td>
</tr>
<tr>
<td>2001</td>
<td>0.9</td>
<td>Smith, John</td>
<td>375</td>
</tr>
<tr>
<td>2002</td>
<td>0.9</td>
<td>Johnson, Ann</td>
<td>8045</td>
</tr>
<tr>
<td>2002</td>
<td>0.9</td>
<td>Rossi, Laura</td>
<td>23000000</td>
</tr>
<tr>
<td>-</td>
<td>0.0</td>
<td>Pérez, José</td>
<td>250000</td>
</tr>
</tbody>
</table>
The world is messy, so is real data

- Real world data is messy
- Need to know context, how it was collected
- Measurement issues and abstraction error
- Some tools can help. Google refine, etc.
Basic kinds of data and measurement

Categorical: cat, dog, goldfish

Ordinal: small, medium, large

Numeric:
  ▶ Integer: 0, 6, 1589
  ▶ Real: 0.0, 25.35, 322.01
Kinds of data and measurement: compound

**Time series:** pairs of ordinal and Numeric values
(0, 1.5), (1, 2.3), (2, 0.6)

**Matrix / array / raster:** multi dimensional associations of numeric values

\[
\begin{pmatrix}
0.3 & 1.2 & 7.0 \\
6.1 & 2.0 & 12.3 \\
0.6 & 5.0 & 4.5
\end{pmatrix}
\]

**Geographic / Geometric:** points (groups of real coordinates)

**Textual:** ‘Once upon a time, in a faraway land’

**Relational:** associations between categorical values:

- cat → mammal
- dog → mammal
- mammal → vertebrate
Building a visualization from data

- Process of constructing “mappings” from elements of the dataset to visual properties, color (hue, intensity), shape, position, etc.
- Sometimes converting data, standardizing or merging in other sources (country codes to latitude and longitude)
- Making sure there are “hooks” to highlight and engage viewers. (humor, surprise)
- Usually its a process of throwing away data or compressing it to show what is relevant
Many types of data have useful and culturally specific standard views.

- Pairs of real values + science audience $\Rightarrow$ x-y plots
- Percentages + business audience $\Rightarrow$ pie charts\(^1\)
- Postal codes + web audience $\Rightarrow$ pins on a map

Look at what people have done, and adapt it

\(^1\)Please don’t tell anyone I said you should use pie charts.
Eric Fisher’s Map of Racial and Ethnic Divides

http://www.flickr.com/photos/walkingsf/4981425631/in/set-72157624812674967/
What task are you trying to support?

Some examples

- Illustrating dynamics of a process
- Comparing outcomes
  - Typical (what is most likely / most common?)
  - Extremal (what are the best / worst or edge cases?)
- Facilitating exploration (familiarization)
- Narrating a historical explanation
Visual narrative genres

Jeff Heer’s core techniques for guiding a viewer through a data presentation
Things are complicated on the Wild Wild Web

- Diagrams can be interactive
- Can link out to lots of resources and actions
- Can be viewed (almost) anywhere in the world
- Easy to share and spread

But...

- Lots of different technologies
- Lots of different platforms and configurations
- Various bandwidth and processor speeds
- Viewed in many contexts
Emile Daigle’s ManyEyes USAID

http://www-958.ibm.com/v/16843
A few web tools:

- **ManyEyes**: Good data loading, fair number of good options, good social elements, hard to customize
  http://www-958.ibm.com/software/data/cognos/manyeyes/visualizations

- **TableauPublic**: Excellent tool, free version has some limits, windows only
  http://www.tableausoftware.com/public/gallery

- **Google Charts / Visualization API**: May take a little hacking, but some good options, robust.

- **Swivel**: dead

- **Mapequation** Network data and clustering
  http://mapequation.org

- **JavaScript libraries** (Raphael JS, Protoviz): there are lots of them, probably need to have a programmer on your team

Example comparisons for geo charts:
http://skyeome.net/projects/geo_chart_comparisons.html
An Average Consumer’s Spending

Each shape below represents how much the average American spends in different categories. Larger shapes make up a larger part of spending.

Food and beverages 15%
The high price of oil is a factor that has made food prices rise quickly.

Miscellaneous 3%

Recreation 6%

Education/Communication 6%
Cellphones were added to the index in 1997. Because the Consumer Price Index can be slow to add new goods, which are often cheaper, it may overstate parts of inflation.

Housing 42%
In the C.P.I., home ownership costs track rent prices more closely than housing prices. This means inflation may have been understated when home prices were rising faster than rents.

Transportation 18%
Gas is 5.2 percent of spending nationwide, but only 3.8 percent in the New York area.

Health care 6%
As a group, the elderly spend about twice as much of their budget on medical care.

Apparel 4%
The ratio of spending on women’s clothes to that on men’s clothes is about 2 to 1.
Good concepts for deeper interaction (via Jeff Heer)

**Details on demand**  Show an overview, let viewer interact for specific details

**Tacit tutorial**  Introduce the visualization with examples that teach how to use it

**Stimulating default views**  Start the viewer with something interesting, let them explore

Other ways

**Linked views**  Multiple views of data with synchronized highlighting

**Make it a game**  Provide scoring and interactive feedback
Mark Newman’s World Population Cartogram

http://www-personal.umich.edu/~mejn/cartograms/
Things to watch out for

- Try not to lie
- Compress and distort confusing non-relevant info, not contradictory facts
- Don’t try to fit everything into one image or visualization.
XKCD’s Social Media map

http://xkcd.com/802/
David McCandless’ Billion Dollar-O-Gram

http://www.informationisbeautiful.net/visualizations/the-billion-dollar-o-gram-2009/

The Billion Dollar-o-Gram

2003 figures
Gapminder’s Dollar Street

http://www.gapminder.org/downloads/dollar-street/
Eigefactor.org’s 2004 Map of science

http://www.eigefactor.org/map/maps.htm
VoteSmart’s VoteEasy Tool

http://www.votesmart.org/voteeasy/
Oil Change International’s Proposition 23 Funding

http://prop23.dirtyenergymoney.com
In my opinion

Visualizations should:

- Be a form of principled storytelling
- Support alternate readings and interpretations
- Get viewers to ask questions
- Convey main point immediately, reward inspection with refined understanding
- Use appropriate conventions, but violate them with good reason
- Communicate the concept, not the technology you built with
More tools and resources

Visualizing Information for Advocacy pamphlet / pdf,
Tactical Technology Collective
http://archive.tacticaltech.org/infodesign

Maps for Advocacy pamphlet/pdf
http://www.tacticaltech.org/mapsforadvocacy

Protoviz examples
http://vis.stanford.edu/protovis/ex/

VisualComplexity blog / catalog
(Manual Lima)
http://www.visualcomplexity.com/

Information Aesthetics blog
http://infosthetics.com/