A few points of clarification to add to my lecture, as there were questions afterwards, and I wasn’t sufficiently clear on a few points.

We know from Cleveland and McGill (https://secure.cs.uvic.ca/twiki/pub/Research/Chisel/ComputationalAestheticsProject/cleveland.pdf p536, among others) that we’re better at comparing length (with a common or unshared baseline) than any other visual quantity (area, angle, etc.).

They state their premise very well:
* A graphical form that involves elementary perceptual tasks that lead to more accurate judgements than another graphical form (with the same information) will result in better organization and increase the chances of a correct perception of patterns and behavior.

The practical result is that rectangular bars and stacked bars are easier to accurately compare (because they’re a linear comparison) than graph types that involve areas of varying widths (rectangles with similar widths can be compared linearly), and therefore should be the preferred format when accurate interpretation of the data matters.

Now, the Nightingale roses, and any similar radius-based stacked bar graphs, are ostensibly fine, because the proposition is that the radial distance matters, not area. This format includes the Facebook vs Twitter graph I discussed.

However, we perceive the area, not just height, and give disproportionate weight to the outer portions of the slice. (Examples of the original roses and less distorted variations are here: http://mbostock.github.com/protovis/ex/crimea-rose.html)

Similarly, the hierarchical example shown later attempts to use angular fractions of the whole to show relative volume, but we end up seeing children and grandchildren with larger areas than the parent, because the outer rings comprise so much more area than the inner rings. This too is a huge distortion of the data.

All that said, circular layouts of data are totally valid in situations where there is no requirement to compare relative area at different distances from the center. Examples include formats such as: http://mbostock.github.com/d3/ex/tree.html and http://circos.ca/, both commonly used in the life sciences, as well as representations of cyclical patterns.

I hope that helps to clarify the ambiguity around the circular formats I claimed were inappropriate.

Thanks for the invitation to speak, I had an excellent time.

-Noah
Effective Data Visualizations

Noah Iliinsky • @noahi

UW Genome Sciences Combi • December 7, 2011
Why Visualization?
### Why visualization?

![Anscombe's Quartet](http://en.wikipedia.org/wiki/Anscombe%27s_quartet)

<table>
<thead>
<tr>
<th>Anscombe's Quartet</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>8.04</td>
<td>7.46</td>
<td>6.58</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.14</td>
<td>6.77</td>
<td>5.76</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>8.74</td>
<td>12.74</td>
<td>7.71</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>8.77</td>
<td>7.11</td>
<td>8.84</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>9.26</td>
<td>7.81</td>
<td>8.47</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>8.10</td>
<td>8.84</td>
<td>7.04</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.13</td>
<td>6.08</td>
<td>5.25</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>3.10</td>
<td>5.39</td>
<td>12.50</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>9.13</td>
<td>8.15</td>
<td>5.56</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.26</td>
<td>6.42</td>
<td>7.91</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>4.74</td>
<td>5.73</td>
<td>6.89</td>
</tr>
</tbody>
</table>
Visualization makes data accessible.
Visualization leverages the amazing abilities of our eyes and brains
Visualization gives faster access to actionable insights
Visualization allows access to huge amounts of data

http://www.youtube.com/watch?v=hVimVzgtD6w
Why Stories?
Stories make data *relevant*. 

http://www.youtube.com/watch?v=hVimVzgtD6w

Thursday, December 8, 2011
Part One: Concepts & Definitions
Data Visualization vs Infographics
Data Visualizations are generated by software.
Infographics are manually drawn.
Data visualizations vs Infographics

How much Data or Beauty is feasible (total data points or mH of beauty)

Not much (algorithmic) vs Lots (all manual)

Manual drawing effort required

- Data volume
- Aesthetic treatment

Manual drawing effort required

Thursday, December 8, 2011
Exploration vs Explanation
Visualization for *exploration*, when you don’t (yet) have a story to tell.

http://www.juiceanalytics.com/nfl-visualization/

Thursday, December 8, 2011
Visualization for *explanation*, when you do have a story to tell.
Education vs Persuasion
Visualization for education
Visualization for persuasion (or propaganda)

Source: Joint Economic Committee, Republican Staff
Congressman Kevin Brady, Ranking House Republican Member

http://www.house.gov/apps/list/press/tx08_brady/71509_he_chart.html
Thursday, December 8, 2011
Part Two: How To Do It
Make good choices.
Good Choices are Guided by Three Inputs
Three inputs.

Your visualization

Reader

Data

Designer
You have goals.
You have goals.

Why are you here?

If you can’t concisely articulate your goal, you’re doing it wrong.
Three types of information products.

Data

Informative

Persuasive

Visual art

Designer

Reader
Your Reader has Needs.
Your reader has needs.

Your success is defined by your readers’ success.
Your reader has needs.

Your success is defined by your customers’ success.

If you can’t satisfy their needs, you have failed.
Identify your audience!

Understand their hopes, dreams, and favorite flavors!

Understand their jargon, identity, and contexts of use!
Consider the contexts and needs of:

- a lab mate
- a geneticist
- a scientist
- a member of the general public
Data has Properties
Data has properties.

- Wheel size: numeric (actually categorical)
- Tire width: continuous
- Price: continuous
- Anti-puncture: binary
- Foldable: binary
Now we start designing.
Statement of Goals
Statement of goals.

“Show the results.”

“Show the specific effects of three different drugs across the selected mutations and a control.”
Define Desired Knowledge *Before*
Structure
Knowledge *before* structure.

MOBILE EMAIL IS STEADILY INCREASING, MOSTLY AT THE EXPENSE OF WEBMAIL CLIENTS, WHILE DESKTOP OPEN RATES HAVE STAYED ABOUT THE SAME. OPENS IN MOBILE DEVICES HAVE EXPERIENCED A 150% INCREASE IN THE LAST SIX MONTHS, WITH WEBMAIL SEEING A 20% DECLINE. EMAIL ACCESSED VIA DESKTOP CLIENTS HAS REMAINED FAIRLY STEADY WITH NEARLY A 2% DROP, PERHAPS INDICATING OUR RELIANCE ON EMAIL IN THE WORKPLACE.

http://litmus.com/blog/email-client-market-share-infograph
Thursday, December 8, 2011
Knowledge before structure.

A look at where subscribers open email.

Mobile email is steadily increasing, mostly at the expense of webmail clients, while desktop open rates have stayed about the same. Opens in mobile devices have experienced a 150% increase in the last six months, with webmail seeing a 20% decline. Email access via desktop clients has remained fairly steady with nearly a 2% drop, perhaps indicating our reliance on email in the workplace.
Appropriate Encodings
Data has properties.

- Wheel size: numeric (actually categorical)
- Tire width: continuous
- Price: continuous
- Anti-puncture: binary
- Foldable: binary
Encoding well:
1. Position is everything.
2. Color is difficult.
- @moritz_stefaner
Position is Everything.
Position is everything.
Position is everything.

- absolute & relative departure time (continuous)
- absolute & relative arrival time (continuous)
- absolute & relative length of trip (continuous)
- stopovers (binary)
- absolute & relative stopover duration (continuous)
- absolute & relative stopover start & stop time (continuous)
- sort order (ranked)
Axes give you information for free!

Northerly and Westerly!

1. about targets
2. when searching (think grouping)
Lack of axes gives you spaghetti!

[Image: Spaghetti_alle_vongole.jpg]

http://commons.wikimedia.org/wiki/File:Spaghetti_alle_vongole.jpg
Color is Difficult.
ALTIMETRY.
Source: USGS HYDRO1K
(1 km hydrologically correct Digital Elevation Model).

Meters:

11
27
43
58
72
83
95
108
121
133
144
155
166
179
194
211
231
255
286
322
371
428
489
566
656
758
879
1044
1328
4453

Wrong!
Color is not ordered.
Color is difficult.

(Mostly) Right!
Color is difficult.

Not bad...
Color is meaningful.

Gender
Nationality
Politics
Religion
Morality
Nature
### Appropriate encodings

<table>
<thead>
<tr>
<th>Example</th>
<th>Encoding</th>
<th>Ordered</th>
<th>Useful values</th>
<th>Quantitative</th>
<th>Ordinal</th>
<th>Categorical</th>
<th>Relational</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="position, placement" /></td>
<td>position, placement</td>
<td>yes</td>
<td>infinite</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td><img src="image" alt="1, 2, 3; A, B, C" /></td>
<td>text labels</td>
<td>optional (alphabetical or numbered)</td>
<td>infinite</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td><img src="image" alt="length" /></td>
<td>length</td>
<td>yes</td>
<td>many</td>
<td>Good</td>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="size, area" /></td>
<td>size, area</td>
<td>yes</td>
<td>many</td>
<td>Good</td>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="angle" /></td>
<td>angle</td>
<td>yes</td>
<td>medium</td>
<td>Good</td>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="pattern density" /></td>
<td>pattern density</td>
<td>yes</td>
<td>few</td>
<td>Good</td>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="weight, boldness" /></td>
<td>weight, boldness</td>
<td>yes</td>
<td>few</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="saturation, brightness" /></td>
<td>saturation, brightness</td>
<td>yes</td>
<td>few</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="color" /></td>
<td>color</td>
<td>no</td>
<td>few (&lt; 20)</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="shape, icon" /></td>
<td>shape, icon</td>
<td>no</td>
<td>medium</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="pattern texture" /></td>
<td>pattern texture</td>
<td>no</td>
<td>medium</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="enclosure, connection" /></td>
<td>enclosure, connection</td>
<td>no</td>
<td>infinite</td>
<td>Good</td>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="line pattern" /></td>
<td>line pattern</td>
<td>no</td>
<td>few</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="line endings" /></td>
<td>line endings</td>
<td>no</td>
<td>few</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="line weight" /></td>
<td>line weight</td>
<td>yes</td>
<td>few</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Use defaults.
Unless...


Thursday, December 8, 2011
Unless you’ve got something better.
Design strategies

- Limit the data you include
- Use position for your most important relationship(s)
- Try different axes
- Consider default formats
- Use color for categories, not rank
- Encode other data and relationships with appropriate properties
**Tools**

- **D3.js / protovis**: structured frameworks
- **processing**: flexible, great for data art
- **R + ggplot2**: stats & analytics
- **Tableau**: visual exploration & analytics
Thank you!

@noahi
gmail: iliinsky
ComplexDiagrams.com

How to do it
(this talk)

How they did it
(20 case studies)
Visualization for *persuasion* (or *propaganda*)

http://www.house.gov/apps/list/press/tx08_brady/71509_hc_chart.html

Thursday, December 8, 2011
Wrong!

Thursday, December 8, 2011
Right!

Thursday, December 8, 2011
Wrong!
Apple or Orange Juice Consumption

Orange Juice
74 lbs per capita

Apple Juice
27 lbs per capita

It looks as if orange juice is the clear winner, but

Right...
Wrong!

http://www.smarter.org/research/apples-to-oranges/

Thursday, December 8, 2011
Wrong!


Wrong!

http://www.digitalsurgeons.com/facebook-vs-twitter-infographic/

Thursday, December 8, 2011
Use a format that fits the data

Better.

http://www.digitalsurgeons.com/facebook-vs-twitter-infographic/
Thursday, December 8, 2011
Wrong!


Thursday, December 8, 2011
Right!

http://www.futurechanges.org/2009/11/12/eisenhower-interstate-system-london-underground-style-map/

Thursday, December 8, 2011
Right!

http://sanfrancisco.crimespotting.org/
Thursday, December 8, 2011
# Periodic Table of Beer Styles

A brief description of beer styles with commercial examples:

<table>
<thead>
<tr>
<th>Style Number</th>
<th>Style Name</th>
<th>Key to Yeast Type</th>
<th>Style Family Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Wheat beer</td>
<td>American lager</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Lambic &amp; Sour ale</td>
<td>European lager</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Belgian ale</td>
<td>British lager</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Pale ale</td>
<td>Austrian lager</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>Light beer</td>
<td>German lager</td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>Dark beer</td>
<td>Czech lager</td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>Smoked beer</td>
<td>Austrian lager</td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>Porter</td>
<td>British lager</td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>Brown porter</td>
<td>German lager</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>Oatmeal stout</td>
<td>Scottish lager</td>
<td></td>
</tr>
<tr>
<td>XI</td>
<td>American lager</td>
<td>American lager</td>
<td></td>
</tr>
<tr>
<td>XII</td>
<td>Munich lager</td>
<td>German lager</td>
<td></td>
</tr>
<tr>
<td>XIII</td>
<td>American lager</td>
<td>American lager</td>
<td></td>
</tr>
</tbody>
</table>

Key to SRM:
- SRM number
- Color
- Clear
- Light straw
- Pale straw
- Dark straw
- Light amber
- Pale amber
- Dark amber
- Very dark amber
- Black
- 0
- 5.5
- 5.5
- 5.5
- 5.5
- 10.0
- 10.0
- 26.0
- 40.0
- 12.0
- 15.0
- 20.0
- 30.0
- 40.0
- 50.0
- 60.0
- 80.0
- 100.0


Thursday, December 8, 2011
Wrong!

http://code.google.com/more/table/
Thursday, December 8, 2011
# A Periodic Table of Visualization Methods

<table>
<thead>
<tr>
<th>Continuum</th>
<th>Data Visualization</th>
<th>Strategy Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tb</td>
<td>Information Visualization</td>
<td>The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.</td>
</tr>
<tr>
<td>Pi</td>
<td>Concept Visualization</td>
<td>The use of interactive visual representations of data to simplify cognition. This means that the data is transformed into an image, it is mapped to a screen space. The image can be changed by users as they proceed working with it.</td>
</tr>
<tr>
<td>B</td>
<td>Compound Visualization</td>
<td>The complementary use of different graphic representation formats in one single schema or frame.</td>
</tr>
</tbody>
</table>

## Process Visualization

Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

© Ralph Lengler & Martin J. Eppler, www.visual-literacy.org

## Structure Visualization

<table>
<thead>
<tr>
<th>Overview</th>
<th>Detail</th>
<th>Detail AND Overview</th>
<th>Divergent thinking</th>
<th>Convergent thinking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Su</td>
<td>Pc</td>
<td>Ho</td>
<td>Ft</td>
<td>Mg</td>
</tr>
<tr>
<td>Pe</td>
<td>St</td>
<td>Fd</td>
<td>Mq</td>
<td>Ld</td>
</tr>
<tr>
<td>Oc</td>
<td>Hc</td>
<td>Ft</td>
<td>Po</td>
<td>Sr</td>
</tr>
<tr>
<td>Ft</td>
<td>Mq</td>
<td>Ld</td>
<td>Po</td>
<td>Ta</td>
</tr>
<tr>
<td>Ft</td>
<td>Mq</td>
<td>Ld</td>
<td>Po</td>
<td>Ta</td>
</tr>
</tbody>
</table>

## Visualization Tools


Wrong!