Visualization

CS 347

Parastoo Abtahi & Danaë Metaxa

Thanks to Michael Bernstein (as usual), Jeff Heer, and Diana Maclean
Announcements

Shorter lecture today because of Project Fair (round two)

Reminders

- Loop through your presentation
- Visit other teams and give feedback
- Report is due tomorrow at noon (only one page)
## Anscombe’s Quartet

*Anscombe, 1973*

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>8.04</td>
<td>10</td>
<td>9.14</td>
<td>10</td>
<td>7.46</td>
<td>8</td>
<td>6.58</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8</td>
<td>8.14</td>
<td>8</td>
<td>6.77</td>
<td>8</td>
<td>5.76</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>13</td>
<td>8.74</td>
<td>13</td>
<td>12.74</td>
<td>8</td>
<td>7.71</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>9</td>
<td>8.77</td>
<td>9</td>
<td>7.11</td>
<td>8</td>
<td>8.84</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>11</td>
<td>9.26</td>
<td>11</td>
<td>7.81</td>
<td>8</td>
<td>8.47</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>14</td>
<td>8.1</td>
<td>14</td>
<td>8.84</td>
<td>8</td>
<td>7.04</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6</td>
<td>6.13</td>
<td>6</td>
<td>6.08</td>
<td>8</td>
<td>5.25</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>4</td>
<td>3.1</td>
<td>4</td>
<td>5.39</td>
<td>19</td>
<td>12.5</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>12</td>
<td>9.11</td>
<td>12</td>
<td>8.15</td>
<td>8</td>
<td>5.56</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7</td>
<td>7.26</td>
<td>7</td>
<td>6.42</td>
<td>8</td>
<td>7.91</td>
</tr>
<tr>
<td>5</td>
<td>5.68</td>
<td>5</td>
<td>4.74</td>
<td>5</td>
<td>5.73</td>
<td>8</td>
<td>6.89</td>
</tr>
</tbody>
</table>

### Summary Statistics
- $u_X = 9.0$  
  $\sigma_X = 3.317$
- $u_Y = 7.5$  
  $\sigma_Y = 2.03$

### Linear Regression
- $Y = 3 + 0.5 \times X$
- $R^2 = 0.67$
Anscombe's Quartet [Anscombe, 1973]
What is 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 1967]

“The use of computer-generated, interactive, visual representations of data to amplify cognition.”  
[Card, Mackinlay, and Shneiderman 1999]
Exports and Imports to and from Denmark & Norway from 1700 to 1780

The Bottom line is divided into Years, the Right hand line into £10,000 each.

Published at the Act direct, 1st May 1786, by Wm. Playfair
Nash sculpt, 352 Strand, London.

William Playfair, 1786
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 areas for area the deaths from Preventible orMitigable 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 line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 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.

Florence Nightingale, 1857
Charles Minard, 1869
Today

How visualizations convey information

Principles and techniques for creating effective visualizations
Visualization and perception

“The visual decoding of information encoded on graphs”
[Cleveland and McGill, ‘84]
Which best encodes quantities?

Position
Length
Area
Value (brightness)
Color hue
Orientation (angle)
Shape

What do you think? [1 min]
Mackinlay’s ranking [1986]

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Ordinal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density</td>
<td>Color hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color saturation</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density</td>
</tr>
<tr>
<td>Density</td>
<td>Containment</td>
<td>Color saturation</td>
</tr>
<tr>
<td>Color saturation</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Color hue</td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Connection</td>
<td>Area</td>
<td>Slope</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Area</td>
</tr>
</tbody>
</table>

But why this ranking?
Sémiologie Graphique
[Bertin 1967]

Foundations of infovis theory. Terms such as:

**marks**: represent something other than themselves

**visual variables**: properties of marks that can vary, e.g., position, size, shape, and color
Pre-attentive processing

[Healey, TVCG 2012]
Pre-attentive processing

[Healey, TVCG 2012]
Implication: visualizations should render variation along the characteristics where our perceptual systems support pre-attentive processing.
We are poor at perceiving (some) differences [Cleveland & McGill, 1984]

Classic result: experimental tests of the relationship between encodings and the accuracy of the differences that we perceive from those encodings.
Intentionally difficult?

[Hullman and Adar, Infovis 2011]

Generally, visualization (and HCI more broadly) argue optimizing for clear and correct interpretation.

Yet difficult visualizations may support better comprehension and recall.

Why? It induces active processing.

Difficult, with chartjunk

[Li and Moacdieh, HFES 2014]
Designing effective visualization
Touring the visualization zoo

[Heer, Bostock, and Ogievetsky 2010]

Visualization researchers develop many new techniques for visualizing complex data. Here are a few popular ones...

Stacked graph: good for aggregate patterns

Small multiples: makes comparison easier
Touring the visualization zoo

[Heer, Bostock, and Ogievetsky 2010]

Comparing statistical distributions...

Scatter plot matrix: compare all pairs

Parallel coordinates: more compact
Touring the visualization zoo
[Heer, Bostock, and Ogievetsky 2010]

Visualizing hierarchies...

Reingold-Tilford tree layout algorithm

Sunburst diagram

Treemap
[Shneiderman 1991]
Touring the visualization zoo

[Heer, Bostock, and Ogievetsky 2010]

Visualizing networks…

- Force directed layout
- Arc diagram
- Adjacency matrix
Conveying uncertainty

[Kay et al., CHI 2016]

We over-rely on point estimates. People simplify distributions and attend to point estimates on the right (10 min until the bus comes)
Conveying uncertainty

[Kay et al., CHI 2016]

Suggestion: quantile dot plots

Probability density of Normal distribution

To generate a discrete plot of this distribution, we could try taking random draws from it. However, this approach is noisy: it may be very different from one instance to the next.

Instead, we use the quantile function (inverse CDF) of the distribution to generate “draws” from evenly-spaced quantiles.

We plot the quantile “draws” using a Wilkinsonian dotplot, yielding what we call a quantile dotplot: a consistent discrete representation of a probability distribution.

By using quantiles we facilitate interval estimation from frequencies: e.g., knowing there are 50 dots here, if we are willing to miss our bus 3/50 times, we can count 3 dots from the left to get a one-sided 94% (1 - 3/50) prediction interval corresponding to that risk tolerance.
Brushing and Linking

[Buja et al., IEEE Vis 1991; Ahlberg and Shneiderman CHI 1994]

Example by Jeff Heer
Exploratory analysis

[Wongsuphasawat et al., TVCG 2015]

User inputs variables of interest, and recommender automatically generates visualizations of relevant other variables.
See you next Wednesday :) 

No lecture on Monday 

Programming commentaries are due one day later 
   Tuesday at 11:30 AM 

Critiques of HCI commentaries are due as usual 
   Wednesday at 11:30 AM 

Studio next week will be focused on programming 
   With guest lecturer, Will Crichton, leading the first half of studio