Information Visualization and the Study of Visual Perception

Ronald A. Rensink

Departments of Psychology and Computer Science
University of British Columbia
Vancouver, Canada
Information visualization (InfoVis): “Using vision to think”

- map problems into graphical form
- use visual intelligence to find the solution
E.g., find the outliers in a dataset

<table>
<thead>
<tr>
<th>Ship</th>
<th>Reporting Date (1A)</th>
<th>UW BBLs</th>
<th>NUW BBLs</th>
<th>UW Hrs</th>
<th>NUW Hrs</th>
<th>CI Hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOWDITCH</td>
<td>10/31/2002</td>
<td>2054</td>
<td>88</td>
<td>574</td>
<td>77</td>
<td>93</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>11/30/2002</td>
<td>1940</td>
<td>107</td>
<td>616</td>
<td>104</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>12/31/2002</td>
<td>1610</td>
<td>67</td>
<td>484</td>
<td>33</td>
<td>227</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>1/31/2003</td>
<td>856</td>
<td>141</td>
<td>192</td>
<td>51</td>
<td>501</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>2/28/2003</td>
<td>1797</td>
<td>114</td>
<td>567</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>3/31/2003</td>
<td>2173</td>
<td>150</td>
<td>633</td>
<td>21</td>
<td>90</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>4/30/2003</td>
<td>1919</td>
<td>170</td>
<td>614</td>
<td>106</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>5/31/2003</td>
<td>2178</td>
<td>200</td>
<td>640</td>
<td>104</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>6/30/2003</td>
<td>1450</td>
<td>202</td>
<td>383</td>
<td>56</td>
<td>281</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>7/31/2003</td>
<td>1671</td>
<td>133</td>
<td>522</td>
<td>183</td>
<td>39</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>8/31/2003</td>
<td>1693</td>
<td>352</td>
<td>538</td>
<td>206</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>9/30/2003</td>
<td>1270</td>
<td>434</td>
<td>419</td>
<td>229</td>
<td>72</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>10/31/2003</td>
<td>2096</td>
<td>294</td>
<td>525</td>
<td>219</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>11/25/2003</td>
<td>1813</td>
<td>58</td>
<td>561</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>12/25/2003</td>
<td>1510</td>
<td>472</td>
<td>431</td>
<td>289</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>1/25/2004</td>
<td>1588</td>
<td>333</td>
<td>548</td>
<td>196</td>
<td>0</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>2/25/2004</td>
<td>2007</td>
<td>122</td>
<td>619</td>
<td>68</td>
<td>57</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>3/25/2004</td>
<td>566</td>
<td>107</td>
<td>143</td>
<td>96</td>
<td>457</td>
</tr>
<tr>
<td>BOWDITCH</td>
<td>4/25/2004</td>
<td>2055</td>
<td>180</td>
<td>627</td>
<td>117</td>
<td>0</td>
</tr>
</tbody>
</table>
E.g., find the outliers in a dataset
Information visualization and vision science can interact in at least three ways...
1. Knowledge of basic *mechanisms* of human vision can help with *design*

- color (Szafir, VIS 2017; Schloss, VSS 2017)
- texture (Hagh-Shenas, Interrante, & Park, 2006)
- motion (Bartram & Ware, 2002)
- ensembles (Szafir, Haroz, Gleicher, & Franconeri, 2016)
- etc., etc…
Knowledge of vision science *methodologies* can help with *evaluation*

- just noticeable differences (JNDs)
- multidimensional scaling
- timing methods (e.g., use of masking)
- dual-task techniques
- etc., etc…
3. Knowledge of vision science *experimental approaches* can help with *understanding* (i.e., why a given visualization works)

In terms of information visualization, this might

- simplify / speed up parts of its evaluation
- inspire new, more effective designs

In terms of vision science, we might discover

- interesting new laws of behavior
- new mechanisms of visual perception / visual intelligence
Potential Objection:

Visualizations aren’t natural; can be arbitrary. Also, they’re recent; we didn’t evolve for them.
However:

1. We’re not looking at arbitrary constructions, just the successful ones—those in common use.

2. Although humans didn’t evolve for these, successful visualizations evolved for us.
**Diagram based on Table I.**
(all female heights are multiplied by 1.08)

<table>
<thead>
<tr>
<th>Mid-parents</th>
<th>Adult Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heights in inches</td>
<td>Deviates in inches</td>
</tr>
<tr>
<td>72</td>
<td>+3</td>
</tr>
<tr>
<td>71</td>
<td>+2</td>
</tr>
<tr>
<td>70</td>
<td>+1</td>
</tr>
<tr>
<td>69</td>
<td>0</td>
</tr>
<tr>
<td>68</td>
<td>-1</td>
</tr>
<tr>
<td>67</td>
<td>-2</td>
</tr>
<tr>
<td>66</td>
<td>-3</td>
</tr>
<tr>
<td>65</td>
<td>-4</td>
</tr>
</tbody>
</table>

- Column headers for heights are 64, 65, 66, 67, 68, 69, 70, 71, 72, 73 inches.
- Column headers for deviations are -4, -3, -2, -1, 0, +1, +2, +3, +4.

Legend:
- Minor axis
- Major axis
- Locus of vertical
- Locus of horizontal
- Tangential points

Graph indicates the relationship between mid-parental heights and adult-child heights, with deviations from 68.5 inches.
An Experimental Approach to Visualization:

*Investigate minimal systems that actually exist, and are easy to manipulate.*

**Drosophila** (fruit fly)
Example: **Scatterplots—correlation r**
(Rensink & Baldridge, 2010; Rensink, 2014, 2017)
1. **Accuracy**  
(Rensink & Baldridge, 2010; Rensink, 2017)

Relate *physical quantity* $r$ (Pearson correlation) to *psychological quantity* $g(r)$ (perceived correlation)

Stimuli with
- gaussian distributions
- equal variance in both dimensions
- $n = 100$ points
Perceived correlation via bisection

*Bisection* *(relative quantities)* rather than *absolute numbers*

Adjust test plot to be midway between reference plots

*(Rensink & Baldridge, 2010)*

*(cf. Plateau, 1872)*
Results (n=20)

(Rensink & Baldridge, 2010)
Results (n=20)

Letting $u = 1 - b_{est}r$

$g(u) \propto \ln(u)$

Fechner’s law

bias $b_{est}$

(accuracy)

$g(r) = \frac{\ln(1 - b_{est}r)}{\ln(1 - b_{est})}$

($b_{est} = 0.9$)

$\text{Pearson correlation } r$
Note: Does not follow Stevens’ Law
- not a sensory quantity?
- cf. perception of numerosity
2. Precision  (Rensink & Baldridge, 2010; Rensink, 2017)

Question:

How much precision exists in correlation estimates?

i.e. - Given a scatterplot with correlation $r$, how much scatter exists in the estimates made?
Forced-choice: Which one has the higher correlation?

Just noticeable difference (JND) – separation for accuracy of 75%
Results (n=20)

\[ \Delta r = k \left( \frac{1}{b_{\text{disc}}} - r \right) \]

\( k: \text{ variability} \)
\( (= 0.22) \)

\( b_{\text{disc}}: \text{ bias} \)
\( (= 0.91) \)

Letting \( u = 1 - b_{\text{disc}} \)
\[ \Delta u = ku \]
\[ \frac{\Delta u}{u} = k \]

Weber’s Law

\[ \Delta r = \text{JND (75% correct)} \]

\[ \text{bias } b_{\text{disc}} \]
\( \text{(1/intercept)} \)

\( Rensink & Baldridge, 2010 \)

Precision
\[ \text{JND}(r) = k \left( \frac{1}{b_{\text{disc}}} - r \right) \]

Accuracy
\[ g(r) = \frac{\ln(1 - b_{\text{est}} r)}{\ln(1 - b_{\text{est}})} \]

\[ b_{\text{disc}} = b_{\text{est}}? \]
For standard condition (gaussian, 100 points)

- magnitude estimation, \( b_{\text{est}} = 0.91 \)
- discrimination, \( b_{\text{disc}} = 0.90 \)

Is this just a coincidence? No. (Rensink, 2017)

Fechner assumption:

\[
\text{Each } \Delta g = \text{constant difference in } g(r) \\
\Rightarrow b_{\text{est}} = b_{\text{disc}} = b
\]
Performance described by just two parameters \((k, b)\)

\[
\text{JND}(r) = k \left( \frac{1}{b} \right) - r
\]

\[
g(r) = \frac{\ln(1 - br)}{\ln(1 - b)}
\]

Evaluation via just two JND measurements(!)

Four conditions:

1. Basic gaussian (equal variances; 100 dots)
2. Low density (25 dots)
3. High aspect ratio (horizontal x 1/2)
4. Uniform distribution

⇒ Much the same performance for all.
5. Nonspatial mappings (Rensink, 2012; 2014)

Here, data dimension #1 is mapped to *horizontal* location, data dimension #2 is mapped to *vertical* location.

Here, data dimension #1 is mapped to *horizontal* location, data dimension #2 is mapped to *blue-yellow value*.

Augmented stripplot
Results — Blue-Yellow (n = 20)

\[ R^2 = 0.85 \]
\[ \text{RMSE} = 0.02 \]

\[ b_{\text{disc}} = 0.67 [0.58, 0.74]; \quad b_{\text{est}} = 0.65 [0.52, 0.74]; \quad p = 0.75 \]

Fechner assumption holds
Carrier:
Any visual property that conveys quantitative information.
Several visual properties seem to be carriers:

- spatial position (*horizontal, vertical*)
- color (*blue-yellow, red-green*)
- orientation (-45° to +45°)
- size (area)
- luminance

All show logarithmic-linear behavior

⇒ new, general laws for information / data visualization

⇒ *a new way to characterize visual features? (= carriers?)*

⇒ connections to perception of numerosity?
6. Mechanisms

What underlies all this?
Proposal (*Rensink*, 2017):

1. The visual system infers a *probability distribution* in an *abstract 2-D space*, likely via *ensemble coding*.

2. Logarithm of the *width of this distribution* $\approx$ *entropy*.
Information entropy (Shannon information)

\[ H = - \sum p_i \log p_i \; ; \; p_i = \text{probability of selection at location } i \]

Often used as a measure of low-level disorder in images (e.g., Chang, Du, Wang, Guo, & Thouin, 2006)

For a bivariate gaussian: \( H(r) \approx \hat{H}(r) \propto \log (1 - br) \)
Proposal (Rensink, 2017):

1. The visual system infers a *probability distribution* in an *abstract 2-D space*, likely via *ensemble coding*.

2. Logarithm of the *width of this distribution* $\approx$ *entropy*.

3. Entropy $\hat{H}$ is used as a *proxy* for Pearson correlation $r$. 
Summary — Scatterplots

The “fruit fly” approach can help us understand the visualization of correlation in scatterplots.

1. Simple, general laws for accuracy and precision
2. Possible new way to explore visual features (carriers)
3. Possible new insights into ensemble coding, numerosity

So: A useful way to explore visual perception?
Objection: This only works for correlation in scatterplots.
But: Dozens of issues can be investigated this way...

- scatterplots — averages
- scatterplots — correlation
- bar charts — averages
- bar charts — correlation
- line graphs — averages
- line graphs — correlation
- pie charts — proportions

- one population
- two populations
- $n$ populations
- learning
- individual diffs
- cultural diffs

- outliers, clustering, perceived similarity, memorability, etc.
Lollipop Chart

- A
- B
- C
- D
- E
- F

http://datavizproject.com/
And these are just some of the *simple* issues…

Can also explore issues concerning, e.g.,

- animation (incl. flickering points)
- data-induced clutter (chartjunk)
- visual metaphor

Can eventually move up from “fruit flies” to more complex & interactive visualizations…

- includes connections to higher-level cognition
Conclusions
The study of correlation perception in scatterplots is a useful way to investigate human vision.

Example: Perception of correlation in scatterplots

- obeys laws that are *simple* (Weber, Fechner)
- based on a mechanism that is *sophisticated* (entropy)
- obeys laws that are *general* (invariance to carrier)
  - Weber laws for stripplots, line graphs, bar charts, etc.
    
    *(Harrison, Yang, Franconeri, & Chang, 2014)*
Many other kinds of visualization issues can also be investigated this way.

Examples: Averages, correlations, clustering, etc. in

- line graphs
- bar charts
- pie charts
- parallel co-ordinate plots
- many, many others…

(And even more in the following talks…)
Visualizations are a useful class of stimuli for exploring our visual intelligence.
A Huge Amount of Thanks to…

• Past & present members of the UBC Visual Cognition Lab:
  - Akbar Alikhan
  - Gideon Baldridge
  - Jacky Chung
  - Mario Cimet
  - Kim De Rosa
  - Madison Elliott
  - Jessica Ip
  - Adelena Leon
  - Natália Lopes
  - Praveena Manogaran
  - Kyle Melnick
  - Yana Pertels
  - Theo Rosenfeld
  - Benjamin Shear
  - Ramyar Sigarchy
  - Sai Venkatasubramanian
  - Kristen Waterman
  - Spencer Williams

• The Boeing Company for their support