Algebraic Visualization Design for Pedagogy

Gordon Kindlmann
Computer Science & Computation Institute
University of Chicago

Carlos Scheidegger
Computer Science
University of Arizona
Our Basic Goal

Empower students with future-proof critical thinking about reading and making visualizations

(not: training on popular toolkits)

U of Chicago T-shirts: “That’s all very well in practice, but how does it work in theory?”
Our Strategy

Structure presentation of vis methods and guidelines around a common underlying theory:

**Algebraic Visualization Design**


So first, a quick primer on Category Theory, Functors, and Fiber Products

A joke! actually ...
Don’t need math to ask focused questions.

Two questions to ask of any vis:
1) If world was different in some interesting way, how would the data be different (α)?
2) How then would the visualization look different (ω)? Does it make visual sense?

Connections: 1) world–data, 2) vis–user

Knowing Taxonomies ≠ Answering questions
Include zero or not on vertical axis?

History of men’s 100m dash world records

What are the $\alpha$?

(axis includes zero)

Cotgreave: “... doesn’t really expose the change of the record over time ... or highlight impact Usain Bolt had on record”

Better because $\omega$ are easier to see!

(axis does not include zero)

“Once you learn the guidelines, you’ll be able to fine tune your charts by bending or breaking them according to your use case and objective.” (not actionable)
Virginia (US) 2014 Senate Election results, per precinct

What, exactly, does this colormap not show about the data?
Virginia (US) 2014 Senate Election results, per precinct

So which one is better, for what purpose?
Teach using $\alpha$s to probe vis properties

$\alpha$: move towards tie

This $\alpha$ explicitly captures source of ambiguity: confuser
... now with a different visualization

New colormap removes this confuser

(\alpha \rightarrow \text{move towards tie})
Where was vote evenly split?

\( \alpha : \text{flip parties} \)
Where was vote evenly split??

\[ \alpha : \text{flip parties} \]

Yikes - so much purple, not informative: \textbf{jumbler}
... now with a different visualization

So here is where vote was evenly split.

Diverging colormap fixes jumbler $\alpha$. 

$\alpha$ : flip parties
What is bad about this visualization?

29 Sept 2015 US Congressional hearing on Planned Parenthood

Visualization shown by Rep. Jason Chaffetz, (R Utah)

http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-chaffetz/chartShown-planned-parenthood-hearing-misleading-/
It is the visual symmetry that’s misleading

(Strong) visual symmetry $\omega$ suggests meaningful data change $\alpha$, i.e. Preserving strong negative correlation, but flipping temporal trend

$\omega$: swap arrows
Actual data change not as interesting

Dual axes create a misleading: an apparent visual symmetry $\omega$ not grounded in interesting data change $\alpha$.

$\omega$: swap arrows
Representation Invariance is an old idea

For seven years a committee of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A Classification of Scales of Measurement

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads
Representation Invariance is an old idea

<table>
<thead>
<tr>
<th>Scale</th>
<th>Basic Empirical Operations</th>
<th>Mathematical Group Structure</th>
<th>Permissible Statistics (invariantive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Determination of equality</td>
<td>$\text{Permutation group } x' = f(x)$</td>
<td>Number of cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$f(x)$ means any one-to-one substitution</td>
<td>Mode</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Determination of greater or less</td>
<td>$\text{Isotonic group } x' = f(x)$</td>
<td>Contingency correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$f(x)$ means any monotonic increasing function</td>
<td></td>
</tr>
<tr>
<td>Interval</td>
<td>Determination of equality of intervals or differences</td>
<td>$\text{General linear group } x' = ax + b$</td>
<td>Median</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Percentiles</td>
</tr>
<tr>
<td>Ratio</td>
<td>Determination of equality of ratios</td>
<td>$\text{Similarity group } x' = ax$</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Standard deviation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rank-order correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Product-moment correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coefficient of variation</td>
</tr>
</tbody>
</table>

Possible hallucinators: e.g. taking median commutes with applying a monotonic function, but taking the mean does not
Like others: pedagogy via critique

Algebraic vis design provides operational vocabulary for **describing a visualization** (confuser, jumbler/misleader, hallucinator)

Also: tensor glyphs and tensor field features

“Tools, not Rules” (George Gopen, The Sense of Structure)

In progress: library of αs that implements Brehmer & Munzner’s task typology  “A Multi-Level Typology of Abstract Visualization Tasks” TVCG 9(12): 2376-2385 (2013)