Introduction to an Algebraic Process for Visualization Design

Gordon Kindlmann
Computer Science & Computation Institute
University of Chicago
glk@uchicago.edu
@glk1

Carlos Scheidegger
Computer Science
University of Arizona
cscheid@cscheid.net
@scheidegger

http://algebraicvis.net
The basic mapping of visualization

Data → Visual

1) How to use 2 planar dimensions? (layout, arrange)
2) What to draw at each location? (encode)

How will these be perceived by the viewer?

example:
(a particular graph on 4 vertices)
Vis methods use **computational** representation

**“Data”** \[\rightarrow\] **Representation** \[\rightarrow\] **Visual**

Underlying thing of interest

How we can measure or store it on computer

"Show data variation, not design variation" [Tufte 1983]

V=(A,B,C,D);

not equal: bug?

V=(D,A,B,C);
Basic ideas of Algebraic Vis Design

[Kindlmann & Scheidegger 2014]

Are important data changes well-matched with obvious visual changes?

Not a taxonomy of tasks, data types, etc.

Mathematical vocabulary for describing how a visualization does or doesn’t work.
Underlying commutative diagram

\[
\begin{array}{ccc}
D & \xrightarrow{r_1} & R & \xrightarrow{v} & V \\
\downarrow{\alpha} & & \downarrow{h} & & \downarrow{\omega} \\
D & \xrightarrow{r_2} & R & \xrightarrow{v} & V
\end{array}
\]

Data change \{ \alpha \} \quad \text{representation change} \{ h \} \quad \text{visual change} \{ \omega \}
Design goal: Task $\rightarrow \alpha, \omega \rightarrow$ affordance

Low-level abstract tasks
[Munzner 2009]
[Meyer et al. 2012]

Perception,
Affordances
[Cleveland & McGill 1984]
[Gibson 1986] [Ware 2012]

$D \xrightarrow{\alpha} \ r_1 \ D \xrightarrow{r_2} \ R \xrightarrow{\nu} V \xrightarrow{\omega} V$

Data change
{well-defined, domain-specific}

Visual change
Three Algebraic Design Principles

All derived from one diagram

Tools, not Rules

Does $\omega$ make sense, given $\alpha$?

→ 1. Principle of Visual-Data Correspondence

For all important $\alpha$, is $\omega$ obvious?

→ 2. Principle of Unambiguous Data Depiction

Can obvious $\omega$ arise without data change ($\alpha=1$)?

→ 3. Principle of Representation Invariance
3. Principle of Representation Invariance

Visualization is invariant w.r.t changes in data representation.

If $\alpha=1$, then $\omega=1$.

- Underlying data $D \neq$ representation $R$ of data
  - e.g. sets as lists, eigenvectors as vectors
- Invariantive: Scale of measurement (nominal, ordinal, interval, ratio) limits permissible statistics [Stevens 1946]
- If change $h$ in representation is visible ($\omega \neq 1$), $h$ is the “hallucinator”
Representation Invariance is an old idea.

---

**SCIENCE**

Vol. 103, No. 2684

Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

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 science—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. Possible hallucinators:

- e.g. taking median commutes with applying a monotonic function; taking the mean does not

<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>Permutation group ( x' = f(x) ) ( f(x) ) means any one-to-one substitution</td>
<td>Number of cases \ Mode \ Contingency correlation</td>
</tr>
<tr>
<td>ORDINAL</td>
<td>Determination of greater or less</td>
<td>Isotonic group ( x' = f(x) ) ( f(x) ) means any monotonic increasing function</td>
<td>Median \ Percentiles</td>
</tr>
<tr>
<td>INTERVAL</td>
<td>Determination of equality of intervals or differences</td>
<td>General linear group ( x' = ax + b )</td>
<td>Mean \ Standard deviation \ Rank-order correlation \ Product-moment correlation</td>
</tr>
<tr>
<td>RATIO</td>
<td>Determination of equality of ratios</td>
<td>Similarity group ( x' = ax )</td>
<td>Coefficient of variation</td>
</tr>
</tbody>
</table>
Invariance example: Graph layout

Data: a graph on 4 vertices $D$ ($\alpha=1$)

### Representation:
- **lists** of verts, edges

### Example:
- $R$: permute vert list
- $\omega \neq 1$: layout depends on vertex ordering

- $\{A, B, C, D\}$
- $\{D, A, B, C\}$
- $\{A, B, C, D\}$
- $\{D, A, B, C\}$

- \( v \): layout
Invariance example: Graph layout

Data: a graph on 4 vertices

$D$ (\(\alpha=1\))

Representation: \textbf{lists}

of verts, edges

$R$

$\omega=1$: with order-invariant layout

$A$

$B$

$C$

$D$

\(h\): permute vert list

$R$

$\nu$

$\nu$
Invariance example: alpha-blended marks

Data: set of locations of taxi pickups & drop-offs

Representation: list of locations

\[ D \quad (\alpha=1) \]

\[ R \quad v \]

\[ h: \text{permute list} \]

\[ \omega \neq 1: \text{“over” operator does not commute: permutation } h \text{ is a hallucinator} \]
Invariance example: alpha-blended marks

Data: set of locations of taxi pickups & drop-offs

Representation: list of locations

$h$: permute list

$\omega = 1$ with order-invariant (commutative) compositing
Three Algebraic Design Principles

Tools, not Rules

Does $\omega$ make sense, given $\alpha$?

→ 1. Principle of Visual-Data Correspondence

For all important $\alpha$, is $\omega$ obvious?

→ 2. Principle of Unambiguous Data Depiction

Can obvious $\omega$ arise without data change ($\alpha=1$)?

→ 3. Principle of Representation Invariance
Important $\alpha$ map to obvious $\omega$.

If $\omega=1$, then $\alpha=1$.

- **Expressiveness**: visualization shows all facts about data (and nothing more) [Mackinlay 1986]

- **Injectivity**: visualization preserves distinctness so viewer can invert it (read it) [Ziemkiewicz & Kosara 2009]

- If not $\nu$ injective, $\alpha$ explicitly indicates the ambiguity; $\alpha$ is the “confuser”
Unambiguity example: treemaps

$\omega = 1$: $\alpha$ is a “confuser” for this treemap
Unambiguity example: treemaps

\[ \alpha \]

\[ \omega \neq 1: \]

Cushion treemaps removes confuser

[van Wijk & H. van de Wetering 1999]
Unambiguity example: parallel coordinates
Unambiguity example: parallel coordinates
Three Algebraic Design Principles

All derived from one diagram

Tools, not Rules

Does \( \omega \) make sense, given \( \alpha \) ?

→ 1. Principle of Visual-Data Correspondence

For all important \( \alpha \), is \( \omega \) obvious?

→ 2. Principle of Unambiguous Data Depiction

Can obvious \( \omega \) arise without data change (\( \alpha = 1 \))?  

→ 3. Principle of Representation Invariance
1. Principle of Visual-Data Correspondence

Important $\alpha$ produce obvious and meaningful $\omega$

- $\alpha$ and $\omega$ well-matched, "$\alpha \cong \omega$"
- $\omega$ makes sense, given $\alpha$

- **Congruence**: visual (external) structure $\cong$ viewer’s mental (internal) structure [Tversky et al. 2002]
- **Effectiveness**: important data attributes mapped to readily perceived visual attributes [Mackinlay 1986]
- **Visual embedding**: visualization preserves distance (in spaces of data, perception) [Demiralp et al. 2014]
Correspondence example: elevation colormap

Data: signed elevation relative to sea level

\[ D \overset{v}{\longrightarrow} \]

\[ \alpha(e) = -e \]

\[ D \overset{v}{\longrightarrow} \]

\[ \omega = \text{negate hue} \]

meaningful \( \alpha \) not matched with perception: “jumbler”
Correspondence example: elevation colormap

Data: signed elevation relative to sea level

\( \alpha(e) = -e \)

\( D \xrightarrow{v} -v(e) \approx v(-e) \) colormapping commutes with negation
Correspondence example: scatterplots

Data: % men vs women employed as senior managers in various countries

$\alpha$: decrease gender gap for one country: EST

$\omega$? Not clear how big that change was

http://economix.blogs.nytimes.com/2013/04/02/comparing-the-worlds-glass-ceilings/?_r=0
Correspondence example: scatterplots

Data: % men vs women employed as senior managers in various countries

\[\alpha: \text{decrease gender gap for one country: EST}\]

\[\omega: \text{change in position along a common scale [Cleveland & McGill 1984]}\]

add diagonal line (%men = %women) and support lines

http://economix.blogs.nytimes.com/2013/04/02/comparing-the-worlds-glassceilings/?_r=0
Correspondence example: simple plots

29 Sept 2015 US Congressional hearing on Planned Parenthood
Visualization shown by Rep. Jason Chaffetz, (Republican-Utah)
Note two distinct vertical scalings!
Correspondence example: simple plots

29 Sept 2015 US Congressional hearing on Planned Parenthood

http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading/
Correspondence example: simple plots

29 Sept 2015 US Congressional hearing on Planned Parenthood

http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading-/
So what is misleading, exactly?

Original data values:

<table>
<thead>
<tr>
<th></th>
<th>2006</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abortions</td>
<td>0.29M</td>
<td>0.33M</td>
</tr>
<tr>
<td>Cancer Scrs &amp; PSs</td>
<td>2.0M</td>
<td>0.94M</td>
</tr>
</tbody>
</table>

$\omega$: exchange arrows (~reflect vertically)
So what is misleading, exactly?

Reading off values (of swapped lines) implied by two distinct vertical scales:

<table>
<thead>
<tr>
<th></th>
<th>2006</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abortions</td>
<td>0.34M</td>
<td>0.29M</td>
</tr>
<tr>
<td>Cancer Scrns &amp; PSs</td>
<td>1.0M</td>
<td>1.7M</td>
</tr>
</tbody>
</table>
Correspondence example: simple plots

The different vertical scales mean that a clear and obvious $\omega$ corresponds to an $\alpha$ that is not especially important $\Rightarrow \omega$ is a misleader.
Correspondence example: simple plots

With single vertical scale: same $\omega$ would correspond to meaningful $\alpha$: swapping values, or reflecting across $x=y$ (preserving the implied negative correlation)
Colormaps and color ordering

Categorical data: no ordering
Color ordering: (primarily) luminance
Can trust “L” of LAB or HCL colorspace, or can experimentally compare luminances
  Students more empowered
Face-based luminance matching [Kindlmann et al. 2002]
  web demo by Kai Li
All colors: L=62 in HCL space

Permuting categories will be a jumbler
Visualizing Principle Comp. Analysis (PCA)

Students tasked with creating colormaps to visualize principle components:

- Maximize correspondence
- Minimize hallucinators

(web demo)
Summary of 3 Principles

Visual-Data Correspondence
or else a **jumbled** $\alpha$, or **misleading** $\omega$

Unambiguous Data Depiction
or else a **confuser** $\alpha$

Representation Invariance
or else a **hallucinator** $h$
Questions to ask of a visualization

• If the data were different, would the vis be different (Unambiguous), and different in an informative way? (Correspondence)
• If ambiguous: what are the data changes am I blind to? (Confuser) Is that a problem?
• If not informative: is there another way to lay out or encode the data to create a better correspondence? (removing Jumblers)
• Are there apparent properties in the vis that are not actually in the data (Misleader)
• Could the vis have ended up appearing differently, in a way that is not determined by the data? (Invariance)
• What are changes in the computational/numerical representation, or the execution of algorithm, that should be inconsequential, but are not? (Hallucinator)
References

Acknowledgements

Planned Parenthood plots: http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading/

Conversations with: Tamara Munzner, Stephen Ingram, Hadley Wickham, Çağatay Demiralp, Xavier Tricoche, and Thomas Schultz

2009 Dagstuhl Scientific Visualization Seminar 09251

Questions?