The Semiology of Graphics

Pat Hanrahan
Stanford University

Representations
Number Scrabble [Simon]

Given: The numbers 1 through 9
Goal: Pick three numbers that sum to 15

Example:
   A takes 8
Number Scrabble [Simon]

Given: The numbers 1 through 9
Goal: Pick numbers so that 3 numbers sum to 15
Example:
  A takes 8
  B takes 2

A takes 4

Number Scrabble [Simon]

Given: The numbers 1 through 9
Goal: Pick numbers so that 3 numbers sum to 15
Example:
  A takes 8
  B takes 2
  A takes 4
Number Scrabble [Simon]

Given: The numbers 1 through 9
Goal: Pick numbers so that 3 numbers sum to 15
Example:
  A takes 8
  B takes 2
  A takes 4
  B takes 3
A takes 5
Number Scrabble [Simon]

Given: The numbers 1 through 9
Goal: Pick numbers so that 3 numbers sum to 15
Example:
   A takes 8
   B takes 2
   A takes 4
   B takes 3
   A takes 5
   B takes?

Problem Isomorphs

A takes 8, 4, 5
B takes 2, 3, ?
Brilliant Cognitive Creations

Algebraic relationship:
\[ 1 + 3 + 5 + 7 + 9 = 5^2 \]

Pythagorean theorem:
Chinese proof by dissection

The Representation Effect

The appropriate representation makes solving problems easier

The best representation depends on the task
The Representation Effect

The appropriate representation makes solving problems easier

The best representation depends on the task

Note that this principle is very similar to the use of abstract data structures in computer science

How to Easily Customize Visual Representations?
Common Representations

Maps (Space)

Time

Tables and Charts

Currently

Drawing programs (for professional designers)
Illustrator and photoshop, ...

Graphics libraries (for professional programmers)
OpenGL, Flash, ...

“I was taught assembler, in the second year of school, it’s like construction work, with a toothpick as a tool”
Song about Lisp by Julia Eckler
The properties of the information
The properties of the image
The rules mapping information to images
The analytical tasks

Language Perspective

Sender and receiver use a language with symbols
- Establish code and conventions
- Sender encodes information in these symbols
- Receiver decodes information from these symbols

Semiology – the study of symbol systems
Information in Position

1. A, B, C are distinguishable
   Nominal
2. B is between A and C
   Ordinal
3. BC is twice as long as AB
   Quantitative

"Resemblance, order and proportional are the three signfields in graphics. These signfields are transcribed by visual variables having the same signifying properties" - Bertin

8 Visual Variables

[x,y]
- Position

[z]
- Size
- Value
- Color
- Texture (frequency)
- Orientation
- Shape (pattern)
### Bertins’ “Levels of Organization”

<table>
<thead>
<tr>
<th></th>
<th>N Nominal</th>
<th>O Ordered</th>
<th>Q Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texture</td>
<td><img src="chart.png" alt="Chart" /></td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td><img src="chart.png" alt="Chart" /></td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Q ⊆ O ⊆ N

### Graphical Schemas or Languages

#### STANDARD SCHEMAS

<table>
<thead>
<tr>
<th>1 COMPONENT</th>
<th>2 COMPONENTS</th>
<th>3 COMPONENTS</th>
<th>MORE THAN 3 COMPONENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAGRAMS</td>
<td><img src="chart.png" alt="Diagram" /></td>
<td><img src="chart.png" alt="Diagram" /></td>
<td><img src="chart.png" alt="Diagram" /></td>
</tr>
<tr>
<td>NETWORKS</td>
<td><img src="chart.png" alt="Diagram" /></td>
<td><img src="chart.png" alt="Diagram" /></td>
<td><img src="chart.png" alt="Diagram" /></td>
</tr>
<tr>
<td>MAPS</td>
<td><img src="chart.png" alt="Diagram" /></td>
<td><img src="chart.png" alt="Diagram" /></td>
<td><img src="chart.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

**COMPREHENSIVE IMAGE**
**PROCESSING GRAPHICS** (Comprehensive)
**INVENTORIES** (Comprehensive) **MESSAGES** (Simplified)
Jock Mackinlay’s Thesis

Automatic Presentation Tool

- Rigorous formulation of Bertin’s approach
- Designed a simple set of visual languages
  - Not meant to be complete
- Implemented languages in logic programming language
  - Not necessarily the most practical approach
- Given a relation, enumerated the sentences in the language that encode that relation
- Choose the best one using expressibility and effectiveness criterion (Cleveland)
System for Visual Analysis

- Designed a visual language that allowed for many common visual representations
  - Tables, chart, timelines, maps, ...
- Designed and implemented the language using relational algebra
- Built an easy-to-use interactive system to query, analyze and visualize a relational database

Demonstration
**Visual Query Language (VizQL)**

```sql
SELECT AS SHAPE
  Market * AVG(Sales) ON COLS
  Quarter * AVG(Profit) ON ROWS
  State * Product IN PANES
  ProductType ON COLOR
  Year ON SHAPE
FROM database
```
Bread-and-Butter of Analysis

Selection
Filtering
Sorting
Calculation
Grouping and Aggregation

Basically what SQL and Excel do …
Litmus tests for an analysis system

Visual Queries

Two insights

- Query-By-Example (QBE)
  - Adopted by Microsoft Access, Paradox

- Dimension/Measure model from BI
  - Dimensions are independent, x
  - Measures are dependent, y = f(x)
  - Adds grouping and aggregation to QBE
Query-By-Example [Zloof, 1975]

**Department Relation**

<table>
<thead>
<tr>
<th>SALES</th>
<th>DEPARTMENT</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATIONARY</td>
<td>DISH</td>
<td></td>
</tr>
<tr>
<td>HOUSEHOLD</td>
<td>PEN</td>
<td></td>
</tr>
<tr>
<td>STATIONARY</td>
<td>PENCIL</td>
<td></td>
</tr>
<tr>
<td>COSMETICS</td>
<td>LIPSTICK</td>
<td></td>
</tr>
<tr>
<td>TOY</td>
<td>PEN</td>
<td></td>
</tr>
<tr>
<td>TOY</td>
<td>PENCIL</td>
<td></td>
</tr>
<tr>
<td>TOY</td>
<td>INK</td>
<td></td>
</tr>
<tr>
<td>COSMETICS</td>
<td>PERFUME</td>
<td></td>
</tr>
<tr>
<td>STATIONARY</td>
<td>INK</td>
<td></td>
</tr>
<tr>
<td>HOUSEHOLD</td>
<td>DISH</td>
<td></td>
</tr>
<tr>
<td>STATIONARY</td>
<td>PEN</td>
<td></td>
</tr>
<tr>
<td>HARDWARE</td>
<td>INK</td>
<td></td>
</tr>
</tbody>
</table>

**Supplier Relation**

<table>
<thead>
<tr>
<th>SUPPLY</th>
<th>ITEM</th>
<th>SUPPLIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEN</td>
<td>PEN</td>
<td>PARKER</td>
</tr>
<tr>
<td>PENCIL</td>
<td>PENCIL</td>
<td>BIC</td>
</tr>
<tr>
<td>INK</td>
<td>INK</td>
<td>PARKER</td>
</tr>
<tr>
<td>PERFUME</td>
<td>PERFUME</td>
<td>REVLOH</td>
</tr>
<tr>
<td>INK</td>
<td>INK</td>
<td>BIC</td>
</tr>
<tr>
<td>DISH</td>
<td>DISH</td>
<td>DUPONT</td>
</tr>
<tr>
<td>LIPSTICK</td>
<td>LIPSTICK</td>
<td>REVLOH</td>
</tr>
<tr>
<td>DISH</td>
<td>DISH</td>
<td>BIC</td>
</tr>
<tr>
<td>PEN</td>
<td>PEN</td>
<td>REVLOH</td>
</tr>
<tr>
<td>PENCIL</td>
<td>PENCIL</td>
<td>PARKER</td>
</tr>
</tbody>
</table>

Q2. Find the department(s) that sells an item(s) supplied by the supplier Parker.

Here the user fills in both the SALES and the SUPPLY Tables as follows.

<table>
<thead>
<tr>
<th>SALES</th>
<th>DEPT</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.Toy</td>
<td>PEN</td>
<td></td>
</tr>
</tbody>
</table>

ANS:

<table>
<thead>
<tr>
<th>DEPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOUSEHOLD</td>
</tr>
<tr>
<td>TOY</td>
</tr>
<tr>
<td>STATIONARY</td>
</tr>
<tr>
<td>HARDWARE</td>
</tr>
</tbody>
</table>

Equivalent to the Domain Relational Calculus [Zloof, Ullman]
N.B. the question and answer style of query languages
Generality!

By following the lead of QBE, I can PROVE it is possible for VizQL to generate ANY SQL query

Thus, query by creating a picture that you want to see

Now make it fluid …

Sensemaking Loop [Card, …]

Task/Question

Forage for data

Search for best visual representation

Visual representation shows relationships and patterns

Find answer or discover insight

Decide and act
Visual Statistics?

A simple idea

- Model formula are widely used to specify linear and non linear models (R/S, SAS, …)
- Two examples are linear regression and factor analysis
- Visual specification related to “model formula” in statistics
- Creating a picture can also specify a formula ...

Demonstration
Automatic Graphic Design?

1. Automatic marks
   - Choose a visual mark based on the type of the fields on axes
   - Choose other default visual attributes based on the properties of the field
2. Incrementally adding a field to a shelf
   - Encode using Bertin-like rules
3. Creating a visualization from scratch
   - Read our InfoVis2007 paper

Demonstration
Formalism Enables ...

Formally construct queries using a visual interface
- Map shelves into queries ala QBE
- Enables drag-and-drop visual analysis
Formally construct linear models using visualization
- Model languages are like VizQL
- One-way: Not all models have a visualization
Automatic design of visualizations
- Captures low-level graphic design “rules”
- Picks reasonable defaults

Software Engineering

Declarative (what), not imperative (not how)
- Like database query languages
More efficient software
- Generative versus monolithic components
- Optimized interpreter / scalable
Simplifies useful features
- Undo/redo/bookmarks: save specifications
- Collaborative visualization: share specifications
- History of analysis: log specifications
Future Work

Visualization transformation
- Program transformations create new visualizations from existing visualizations
- Rules for rearranging fields

Learn good visualizations
- Use machine learning to find design rules using examples of good design

Limitations

Currently, rather simple representations
- Bertin did consider networks (node-link)
- Bertin did not consider 3D, animation, ...
- Semantically richer designs such as diagrams
  - Take 3??

Data model is weak
- Unstructured data??
Limitations

Perceptual foundations are shaky
- What are the right visual attributes?
- How can they be combined?

Cognitive models are too simple
- Important additional factors include context, engagement, style, aesthetics, ...
- No easy way to get at task ...

Summary

A journey ...
- Some insights from cognitive science
  - How to choose and create the right representation for a task?
- Combine Bertin’s ideas about the semiology of graphics with relational algebra and databases
- Platform to explore query and analysis, hypothesis testing, and design

Addresses the long tail of analysis and visualization
Thank you