Lecture 4: Frameworks/Models

Information Visualization
CPSC 533C, Fall 2007

Tamara Munzner

UBC Computer Science

19 September 2007

Papers Covered

Chapter 1, Readings in Information Visualization: Using Vision to Think. Stuart Card, Jock Mackinlay, and Ben Shneiderman, Morgan Kaufmann 1999.

The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations Ben Shneiderman, Proc. 1996 IEEE Visual Languages, also Maryland HCIL TR 96-13 [citeseer.ist.psu.edu/shneiderman96eyes.html]

Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG 8(1), January 2002. [graphics.stanford.edu/papers/polaris]

The Value of Visualization. Jarke van Wijk. Visualization 2005 [www.win.tue.nl/ vanwijk/vov.pdf]

Low-Level Components of Analytic Activity in Information Visualization. Robert Amar, James Eagan, and John Stasko. Proc. InfoVis 05 [www.cc.gatech.edu/ john.stasko/papers/infovis05.pdf]

Further Readings

The Structure of the Information Visualization Design Space Stuart Card and Jock Mackinlay, Proc. InfoVis 97 [citeseer.ist.psu.edu/card96structure.html]

Automating the Design of Graphical Presentations of Relational Information. Jock Mackinlay, ACM Transaction on Graphics, vol. 5, no. 2, April 1986, pp. 110-141.

Semiology of Graphics. Jacques Bertin, Gauthier-Villars 1967, EHESS 1998

The Grammar of Graphics. Leland Wilkinson, Springer-Verlag 1999

Rethinking Visualization: A High-Level Taxonomy. Melanie Tory and Torsten Möller, Proc. InfoVis 2004, pp. 151-158.

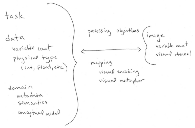
A Function-Based Data Model for Visualization. Lloyd Treinish, Visualization 1999 Late Breaking Hot Topics

Multiscale Visualization Using Data Cubes. Chris Stolte, Diane Tang and Pat Hanrahan, Proc. InfoVis 2002

Frameworks

- Mackinlay/Card/(Bertin)
  - Data Types, Marks, Retinal Attributes (incl Position)
- Shneiderman, Amar/Eagan/Stasko
  - Data, Tasks
- Tory/Moeller, Hanrahan
  - Data/Conceptual Models
- Stolte/Tang/Hanrahan, (Wilkinson)
  - Table Algebra ⇔ Visual Interface
- van Wijk
  - Value

Visualization Big Picture

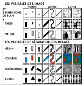


Mapping

- input
  - data semantics
  - use domain knowledge
- output
  - visual encoding
    - visual/graphical/perceptual/retinal
    - channels/attributes/dimensions/variables
  - use human perception
- processing
  - algorithms
  - handle computational constraints

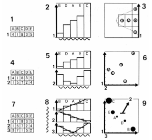
Bertin: Semiology of Graphics

- geometric primitives: marks
  - points, lines, areas, volumes
- attributes: visual/retinal variables
  - parameters control mark appearance
  - separable channels flowing from retina to brain
- x,y
  - position
- z
  - size
  - greyscale
  - color
  - texture
  - orientation
  - shape

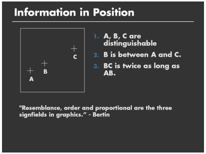


[Bertin, Semiology of Graphics, 1967 Gauthier-Villars, 1998 EHESS]

Design Space = Visual Metaphors



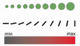
[Bertin, Semiology of Graphics, 1967 Gauthier-Villars, 1998 EHESS]



[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding/walk025.html]

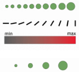
Data Types

- continuous (quantitative)
  - 10 inches, 17 inches, 23 inches



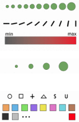
Data Types

- continuous (quantitative)
  - 10 inches, 17 inches, 23 inches
- ordered (ordinal)
  - small, medium, large
  - days: Sun, Mon, Tue, ...



Data Types

- continuous (quantitative)
  - 10 inches, 17 inches, 23 inches
- ordered (ordinal)
  - small, medium, large
  - days: Sun, Mon, Tue, ...
- categorical (nominal)
  - apples, oranges, bananas



[graphics.stanford.edu/papers/polaris]

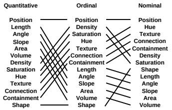
More Data Types: Stevens

- subdivide quantitative further:
- interval: 0 location arbitrary
  - time: seconds, minutes
- ratio: 0 fixed
  - physical measurements: Kelvin temp

[S.S. Stevens, On the theory of scales of measurements, Science 103(2684):677-680, 1946]

Channel Ranking Varies by Data Type

- spatial position best for all types



[Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM TOG 5:2, 1986]

Mackinlay, Card

- data variables
  - 1D, 2D, 3D, 4D, 5D, etc
- data types
  - nominal, ordered, quantitative
- marks
  - point, line, area, surface, volume
  - geometric primitives
- retinal properties
  - size, brightness, color, texture, orientation, shape...
  - parameters that control the appearance of geometric primitives
  - separable channels of information flowing from retina to brain
- closest thing to central dogma we've got

Shneiderman's Data+Tasks Taxonomy

- data
  - 1D, 2D, 3D, temporal, nD, trees, networks
  - text and documents (Hanrahan)
- tasks
  - overview, zoom, filter, details-on-demand,
  - relate, history, extract
- data alone not enough
  - what do you need to do?
- mantra: overview first, zoom and filter, details on demand

[Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations]

Tasks, Amar/Eagan/Stasko Taxonomy

- low-level tasks
  - retrieve value, filter, compute derived value,
  - find extremum, sort, determine range,
  - characterize distribution, find anomalies,
  - cluster, correlate
- standardized set for better comparison between papers
  - bottom-up grouping with affinity diagramming
  - abstraction from domain task down to low-level task

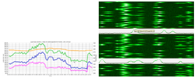
[Amar, Eagan, and John Stasko. Low-Level Components of Analytic Activity in Information Visualization. Proc. InfoVis 05]

Control Room Example

Which location has the highest power surge for the given time period? (extreme y-dimension)

A fault occurred at the beginning of this recording, and resulted in a temporary power surge. Which location is affected the earliest? (extreme xdimension)

Which location has the most number of power surges? (extreme count)



[Overview Use in Multiple Visual Information Resolution Interfaces. Lam, Munzner, and Kincaid. Proc. InfoVis 2007]

Data Models vs. Conceptual Models

- data model: mathematical abstraction
  - set with operations
  - e.g. integers or floats with $*, +$
- conceptual model: mental construction
  - includes semantics, support data
  - e.g. navigating through city using landmarks

[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding/walk005.html]

[Rethinking Visualization: A High-Level Taxonomy. Melanie Tory and Torsten Möller, Proc. InfoVis 2004, pp. 151-158.]

Models Example

- from data model
  - 17, 25, -4, 28.6
  - (floats)

Models Example

- from data model
  - 17, 25, -4, 28.6
  - (floats)
- using conceptual model
  - (temperature)

Models Example

- from data model
  - 17, 25, -4, 28.6
  - (floats)
- using conceptual model
  - (temperature)
- to data type
  - burned vs. not burned (N)
  - hot, warm, cold (O)
  - continuous to 4 sig figures (Q)

Models Example

- from data model
  - 17, 25, -4, 28.6
  - (floats)
- using conceptual model
  - (temperature)
- to data type
  - burned vs. not burned (N)
  - hot, warm, cold (O)
  - continuous to 4 sig figures (Q)
- using task
  - making toast
  - classifying showers
  - finding anamolies in local weather patterns

Time

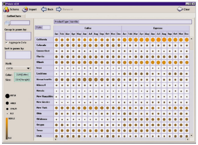
- 2D+T vs. 3D
  - same or different? depends on POV
    - time as input data?
    - time as visual encoding?
- same
  - time just one kind of abstract input dimension
- different
  - input semantics
  - visual encoding: spatial position vs. temporal change
- processing might be different
  - e.g. interpolate differently across timesteps than across spatial position

Polaris

- infovis spreadsheet
- table cell
  - not just numbers: graphical elements
  - wide range of retinal variables and marks
- table algebra ⇔ interactive interface
  - formal language
- influenced by Wilkinson
  - Grammar of Graphics, Springer-Verlag 1999
- commercialized as Tableau

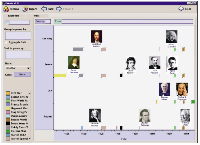
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Circles, State/Product:Month



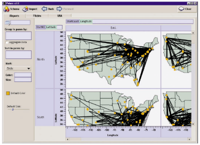
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Gantt Bar, Country/Time



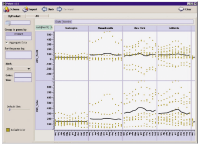
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Circles, Lat/Long



[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Circles, Profit/State:Months



[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Fields Create Tables and Graphs

- Ordinal fields: interpret field as sequence that partitions table into rows and columns:
  - Quarter = (Qtr1),(Qtr2),(Qtr3),(Qtr4) ⇔

Qtr1	Qtr2	Qtr3	Qtr4
95892	101760	105282	98225

- Quantitative fields: treat field as single element sequence and encode as axes:
  - Profit = (Profit) ⇔

[Hanrahan,graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]

Combinatorics of Encodings

- challenge
  - pick the best encoding from exponential number of possibilities $(n+1)^8$
- Principle of Consistency
  - properties of the image should match properties of data
- Principle of Importance Ordering
  - encode most important information in most effective way

[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]

Automatic Design

- Mackinlay, APT
- Roth et al, Sage/Visage
- select visualization automatically given data
  - vs. Polaris: user drag and drop exporation
- limited set of data, encodings
  - scatterplots, bar charts, etc
- holy grail
  - entire parameter space

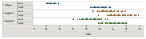
Mackinlay's Criteria

- Expressiveness
  - Set of facts expressible in visual language if sentences (visualizations) in language express all facts in data, and only facts in data.
- consider the failure cases...

[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]

Cannot Express the Facts

- A 1 ⇔ N relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position



[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]

Expresses Facts Not in the Data

- Length interpreted as quantitative value
  - Thus length says something untrue about nominal data

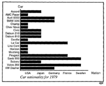


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT]

[Hanrahan,graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]

Mackinlay's Criteria

- Expressiveness
  - Set of facts expressible in visual language if sentences (visualizations) in language express all facts in data, and only facts in data.
- Effectiveness
  - A visualization is more effective than another visualization if information conveyed by one visualization is more readily perceived than information in other.
  - subject of the next lecture

[Hanrahan,graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]

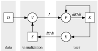
Summary

- formal approach to picture specification
  - declare the picture you want to see
  - compile query, analysis, and rendering commands needed to make the pictures
  - automatically generate presentations by searching over the space of designs
- Bertin's vision still not complete
  - formalize data model
  - formalize the specifications
  - experimentally test perceptual assumptions
- much more research to be done...

[Hanrahan,graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]

Value of Vis

- $I(t) = V(D, S, t)$
  - data D transformed by spec S into time-varying image
- $dK/dt = P(I, K)$
  - perception P of image by user increases knowledge K
- $S(t) = S_0 + \int E(K)$
  - interative exploration E changes spec



[The Value of Visualization. Jarke van Wijk. Proc. Visualization 2005]

Cost Model

- costs
  - $C_i(S_0)$: initial development costs
  - $C_u(S_0)$: initial per-user costs
  - $C_s(S_0)$: initial per-session costs
  - C_e : perception and exploration costs
- benefit
  - $G = nmW(\Delta K)$
- profit
  - $F = G - C$
  - $F = nm(W(\Delta K) - C_s - kC_e) - C_i - nC_u$

"a great visualization method is used by many people, who use it routinely to obtain highly valuable knowledge, without having to spend time and money on hardware, software, and effort. Indeed, quite obvious."

Arguments

- new methods not better by definition
- vis not good by definition
  - must show why automated extraction insufficient
  - e.g. automation not foolproof
- if no clear patterns
  - method limitation?
  - wrong parameters?
  - or truly not there in data?
- inspire new hypotheses vs. verify final truth

Arguments

- "avoid interaction" dictum controversial
  - part of power of computer-based methods
  - but can degenerate into human-powered search
- presentation/exposition vs. exploration
- art vs. science vs. technology

Credits

- Pat Hanrahan
  - graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding
- Torsten Möller, Melanie Tory
  - discussions