Lecture 3: Fundamentals

Information Visualization
CPSC 533C, Fall 2009

Tamara Munzner

UBC Computer Science

Wed, 16 September 2009

Papers Covered

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

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]

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]

A Nested Model for Visualization Design and Validation. Tamara Munzner. IEEE TVCG 15(6) (Proc. InfoVis 2009), to appear. [www.cs.ubc.ca/labs/imager/tr/2009/NestedModel]

MatrixExplorer: a Dual-Representation System to Explore Social Networks. Nathalie Henry and Jean-Daniel Fekete. IEEE Trans. Visualization and Computer Graphics (Proc InfoVis 2006) 12(5), pages 677-684, 2006. [www.aviz.fr/ nhenry/docs/Henry-InfoVis2006.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.

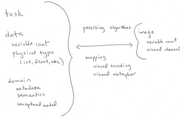
Semiology of Graphics. Jacques Bertin, Gauthier-Villars 1967, EHESS 1998

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.

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]

Visualization Big Picture

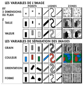


Mapping

- 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

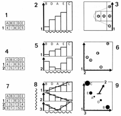
Bertin: Semiology of Graphics

- 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



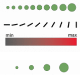
[Bertin, Semiology of Graphics, 1967 Gauthier-Villars, 1998 EHESS]

Data Types

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

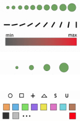
Data Types

- 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]

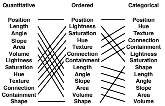
More Data Types: Stevens

- 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, ...
- 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

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]

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

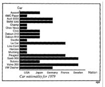


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[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]

Design: Designer vs. Automatic vs. User

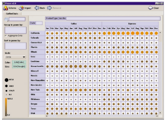
- designer: studies last time
- automatic: select visualization automatically given data
  - Mackinlay, APT
    - limited set of encodings:
    - scatterplots, bar charts...
  - Roth et al, Sage/Visage
  - holy grail: entire space of infovis visual encoding
    - nowhere near goal, esp. with relational/graph data
- human-guided: allow user to change encodings
  - Polaris: user drag and drop exporation

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's Grammar of Graphics
  - Grammar of Graphics, Springer-Verlag 1999
- commercialized as Tableau Software

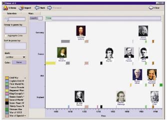
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Circles, State/Product:Month



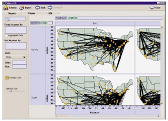
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Gantt Bar, Country/Time



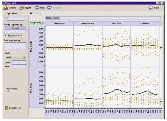
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

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]

Beyond Data Alone

- bigger picture than just visual encoding decisions
- Shneiderman's data+task 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. Proc. 1996 IEEE Visual Languages]

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

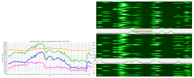
[Amar, Eagan, and John Stasko. Low-Level Components of Analytic Activity in Information Visualization. Proc. InfoVis 05]

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 x-dimension)

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

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

Models Example

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

Models Example

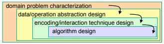
- data model
  - 17, 25, -4, 28.6
  - (floats)
- conceptual model
  - temperature
- depending on task, transform to data type
  - making toast
    - burned vs. not burned (N)
  - classifying showers
    - hot, warm, cold (O)
  - finding anamolies in local weather patterns
    - continuous to 4 sig figures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

Time

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

Nested Model

- separating design into levels
  - not just the visual encoding level!



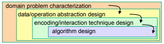
- cascading dependencies: outputs from level above are inputs to level below

[Munzner. A Nested Model for Visualization Design and Validation. IEEE TVCG 15(6) (Proc. InfoVis 2009), to appear. www.cs.ubc.ca/labs/imager/tr/2009/NestedModel]

Nested Levels

- characterizing problems
  - understanding domain concepts, current workflow
  - find gaps where conjecture that vis would help
  - MatrixExplorer case study example
- abstracting into operations on data types
  - Amar/Stasko tasks: abstract operation example
  - MizBee: abstraction on data example
- designing encoding and interaction
  - Bertin, Mackinlay/Card: encoding
  - later in term: interaction design
- creating efficient algorithms
  - classic CS problem: create algorithm given clear specification

Threats To Validity: What Can Go Wrong?

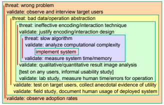


- wrong problem
  - they don't do that
- wrong abstraction
  - you're showing them the wrong thing
- wrong encoding/interaction
  - the way you show it doesn't work
- wrong algorithm
  - your code is too slow

[Munzner. A Nested Model for Visualization Design and Validation. IEEE TVCG 15(6) (Proc. InfoVis 2009), to appear. www.cs.ubc.ca/labs/imager/tr/2009/NestedModel]

Upstream and Downstream Validation

- humans in the loop for outer three levels



[Munzner. A Nested Model for Visualization Design and Validation. IEEE TVCG 15(6) (Proc. InfoVis 2009), to appear. www.cs.ubc.ca/labs/imager/tr/2009/NestedModel]

MatrixExplorer

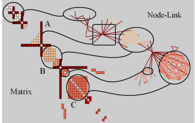
- domain: social network analysis
- validation
  - early: participatory design to generate requirements
  - later: qualitative observations of tool use by target users
- techniques
  - interactively map attributes to visual variables
    - user can change visual encoding on the fly (like Polaris)
  - filtering
  - selection
  - sorting by attribute

Requirements

- use multiple representations
- handle multiple connected components
- provide overviews
- display general dataset info
- use attributes to create multiple views
- display basic and derived attributes
- minimize parameter tuning
- allow manual finetuning of automatic layout
- provide visible reminders of filtered-out data
- support multiple clusterings, including manual
- support outlier discovery
- find where consensus between different clusterings
- aggregate, but provide full detail on demand

Techniques: Dual Views

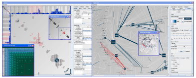
- show both matrix and node-link representations



[Fig 3. Henry and Fekete. MatrixExplorer: a Dual-Representation System to Explore Social Networks. IEEE TVCG 12(5):677-684 (Proc InfoVis 2006)]

MatrixExplorer Views

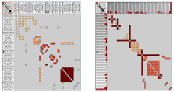
- overviews: matrix, node-link, connected components
- details: matrix, node-link
- controls



[Fig 1. Henry and Fekete. MatrixExplorer: a Dual-Representation System to Explore Social Networks. IEEE TVCG 12(5):677-684 (Proc InfoVis 2006) www.aviz.fr/ nhenry/docs/Henry-InfoVis2006.pdf]

Automatic Clustering/Reordering

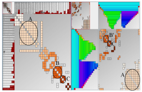
- automatic clustering as good starting point
- then manually refine



[Fig 6. Henry and Fekete. MatrixExplorer: a Dual-Representation System to Explore Social Networks. IEEE TVCG 12(5):677-684 (Proc InfoVis 2006)]

Comparing Clusters

- relayout, check if clusters conserved



- encode clusters with different visual variables
- colorcode common elements between clusters

[Fig 11. Henry and Fekete. MatrixExplorer: a Dual-Representation System to Explore Social Networks. IEEE TVCG 12(5):677-684 (Proc InfoVis 2006)]

Credits

- Pat Hanrahan

graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding

- Torsten Möller, Melanie Tory
  - discussions on conceptual models