Lecture 3: Data Principles

Information Visualization
CPSC 533C, Fall 2011

Tamara Munzner

UBC Computer Science

Mon, 19 September 2011

Papers Covered

Chapter 2: Data Principles

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]

Further Readings

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

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

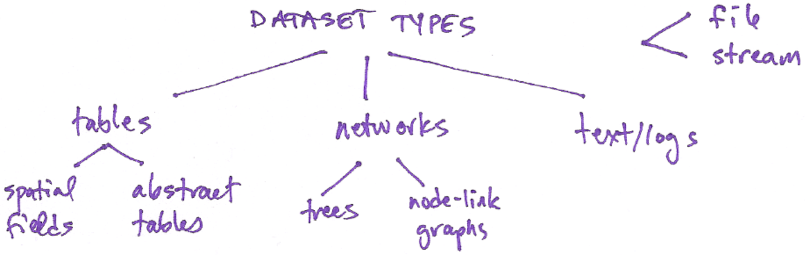
Rethinking Visualization: A High-Level Taxonomy. Melanie Tory and Torsten Möller, Proc. InfoVis 2004, pp. 151-158.

Using Strahler numbers for real time visual exploration of huge graphs. David Auber. Intl Conf. Computer Vision and Graphics, 2002, p 56-69.

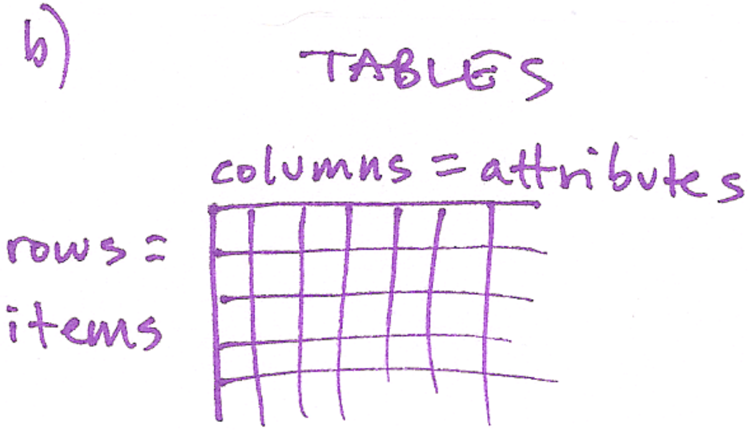
Feature detection in linked derived spaces. Chris Henze. Proc. Visualization (Vis) 1998, p 87-94.

Graph-Theoretic Scagnostics Leland Wilkinson, Anushka Anand, and Robert Grossman. Proc InfoVis 05.

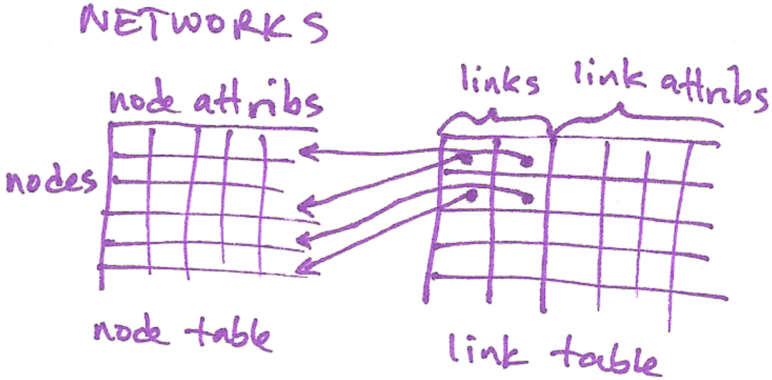
Dataset Types



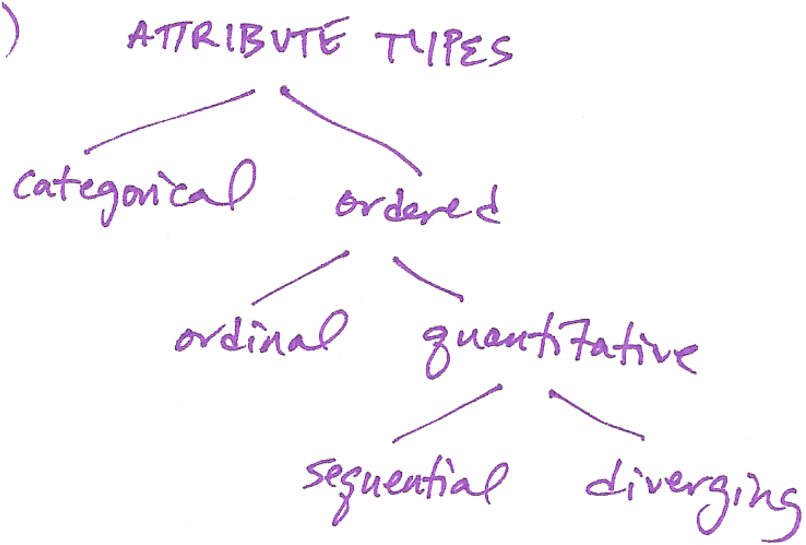
Tables



Networks


NETWORKS
node attribs
nodes
node table
links
link attribs
link table

Attribute Types



- amendment: sequential/diverging crosscuts ordered

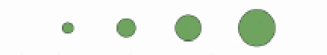
Attribute Types

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

Attribute Types

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

- ordered (ordinal)
  - small, medium, large
  - days: Sun, Mon, Tue, ...

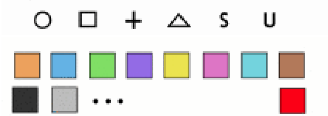


Attribute Types

- 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 Attribute Types: Stevens

- further subdivision of quantitative
  - 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]

Attribute Semantics

- spatial/nonspatial
- temporal/nontemporal
- independent/dependent
- continuous/discrete
- dimensions/measures
    - dimensions: categorical
    - measures: quantitative
    - databases vocab, used in Polaris

Dataset Semantics

- spatial/abstract
  - spatial fields if independent spatial attribs
  - abstract otherwise: must choose spatial layout

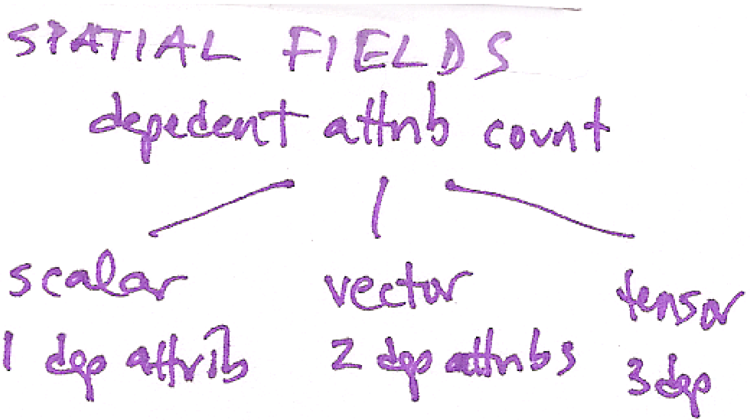
Attributes: Multidimensional Tables

MULTIDIMENSIONAL TABLES

ind: index/independent attrib

dep: dependent/value attrib

Attributes: Spatial Fields



Dataset Semantics

- spatial/abstract
  - spatial fields if independent spatial attribs
  - abstract otherwise: must choose spatial layout
- static/timevarying
  - timevarying if independent temporal dimension
  - tv different than static/dynamic dataset types
  - time series data: simple special case

Other Data Taxonomies

- Shneiderman's data+task taxonomy: data
  - 1D, 2D, 3D, temporal, nD, trees, networks
- Hanrahan's addition:
  - text and documents

[Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proc. 1996 IEEE Visual Languages]

Derived Attributes and Spaces

- derived attribute: compute from originals
  - simple change of type
  - complex transformation using global information
- derived spaces
  - dataset with derived attributes
  - may be only derived attribs, or derived+original attribs
  - dataset transformation as abstraction choice

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
- conceptual model motivates derived data

[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.]

Derived Attributes Example

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

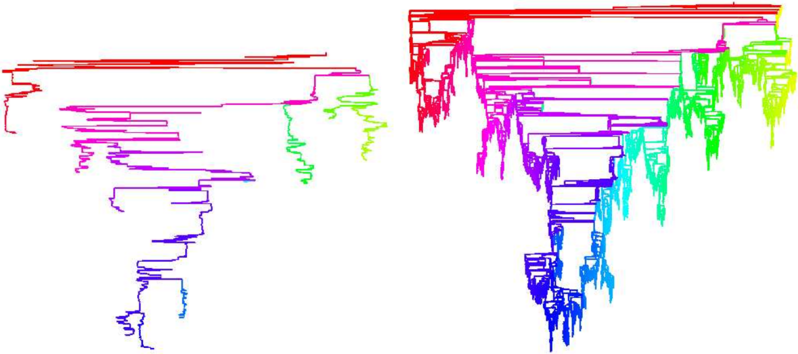
Derived Attributes Example

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

Derived Attributes Example

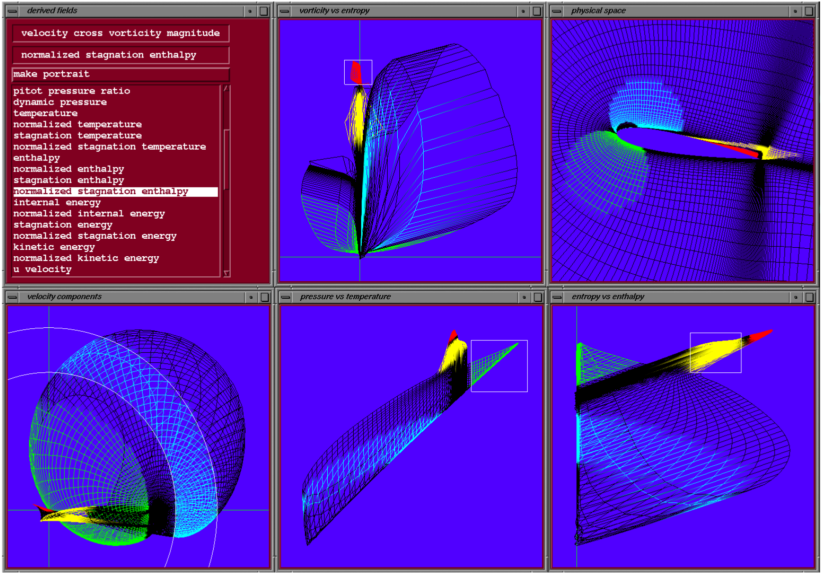
- data model
  - 17, 25, -4, 28.6
  - (floats)
- conceptual model
  - temperature
- depending on task, transform to data type
  - making toast
    - burned vs. not burned (N)
  - classifying showers
    - hot, warm, cold (O)
  - finding anamolies in local weather patterns
    - continuous to 4 sig figures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

Derived Space: Strahler Numbers for Trees



[Using Strahler numbers for real time visual exploration of huge graphs. David Auber. Intl Conf. Computer Vision and Graphics, 2002, p 56-69.]

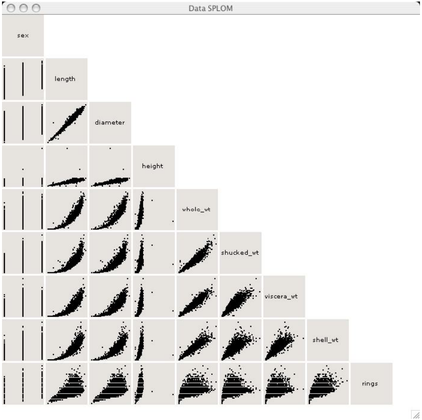
Derived: Feature Detection in Fluids



[Feature detection in linked derived spaces. Chris Henze. Proc. Vis 1998, p 87-94.]

Derived: Graph-Theoretic Scagnostics

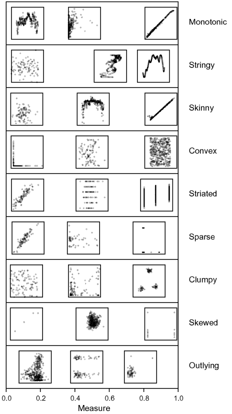
- SPLOM: scatterplot matrix



[Graph-Theoretic Scagnostics Leland Wilkinson, Anushka Anand, and Robert Grossman. Proc InfoVis 05.]

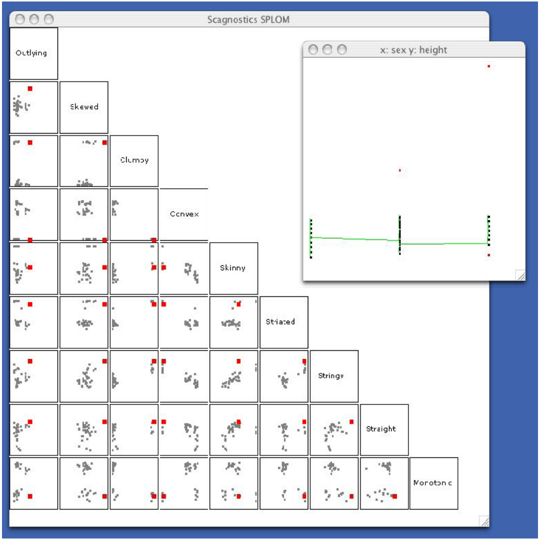
Scagnostics Measures

- scatterplot measures: monotonic, stringy, skinny, convex, striated, sparse, clumpy, skewed, outlying



[Graph-Theoretic Scagnostics Leland Wilkinson, Anushka Anand, and Robert Grossman. Proc InfoVis 05.]

Scagnostics Measures



[Graph-Theoretic Scagnostics Leland Wilkinson, Anushka Anand, and Robert Grossman. Proc InfoVis 05.]

Time

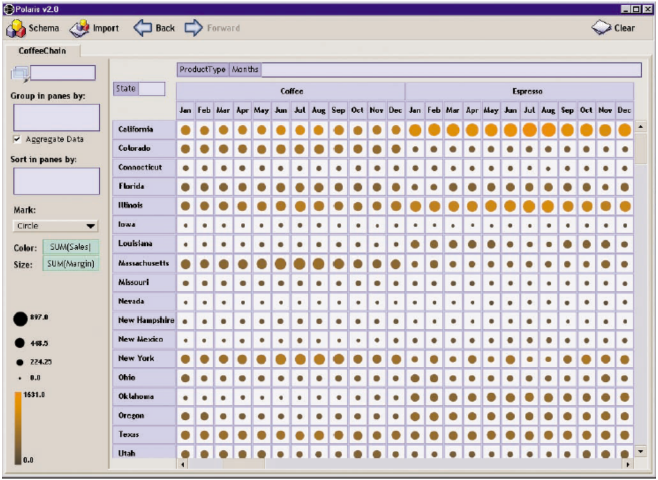
- 2D+T vs. 3D
  - same or different? depends on POV
    - input side vs. output side
- same
  - input: time as just one kind of abstract input dimension
- different
  - input: semantics (time steps of dynamically changing data)
  - output: visual encoding channel of temporal change very different than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

Polaris

- infovis spreadsheet
- table cell
  - not just numbers: graphical elements
  - wide range of retinal variables and marks
- table algebra $\Leftrightarrow$ interactive interface
  - formal language
- influenced by Wilkinson's Grammar of Graphics
  - Grammar of Graphics, Springer-Verlag 1999
- commercialized as Tableau Software
  - good sandbox for projects!

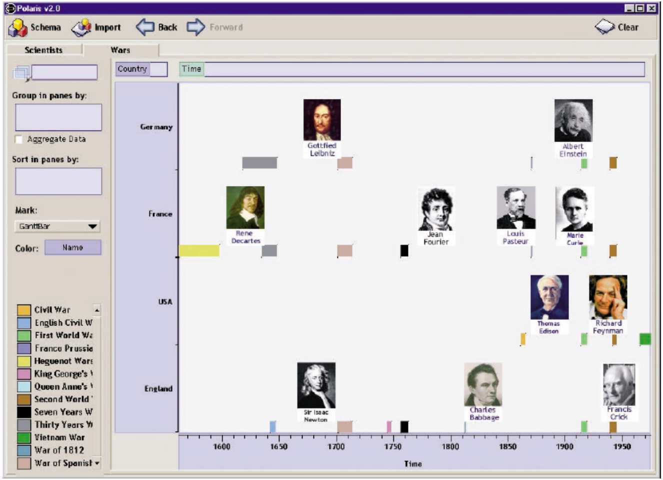
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Circles, State/Product:Month



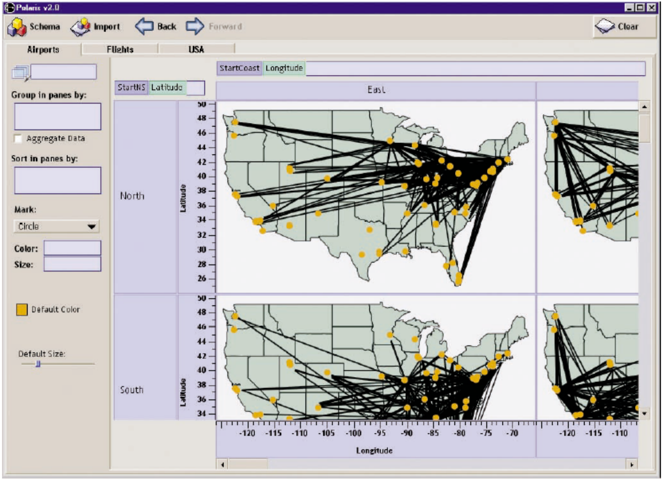
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

Polaris: Gantt Bar, Country/Time



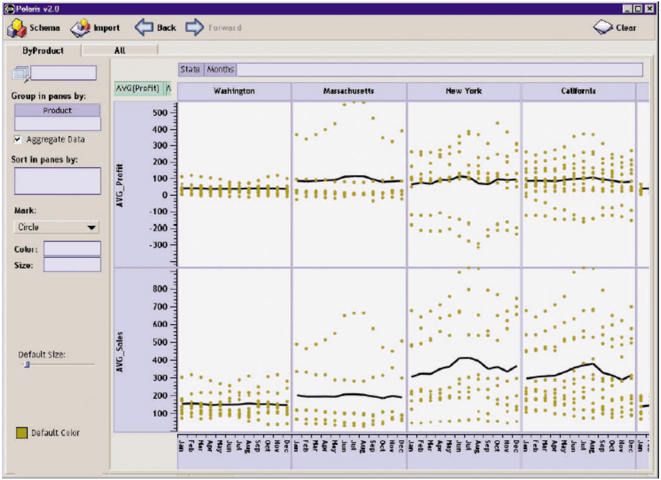
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

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



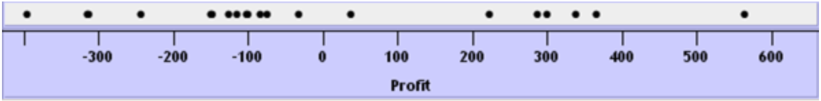
[Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG, 8(1) Jan 2002]

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]