



DataMeadow

A Visual Canvas for Analysis of
Large-Scale Multivariate Data



IEEE Symposium on Visual Analytics Science & Technology 2007

Niklas Elmqvist (elm@lri.fr) – John Stasko (stasko@cc.gatech.edu) – Philippas Tsigas (tsigas@chalmers.se)



CHALMERS

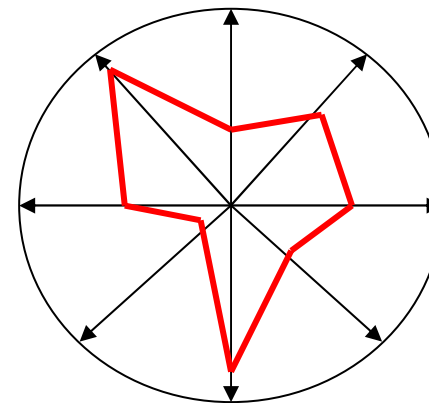
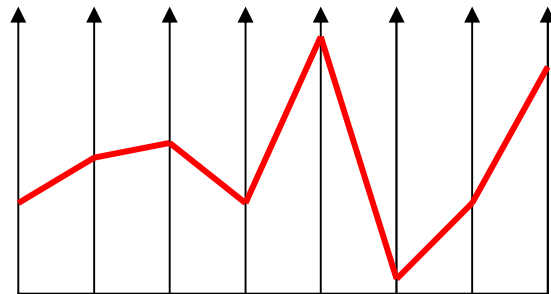


In media res...

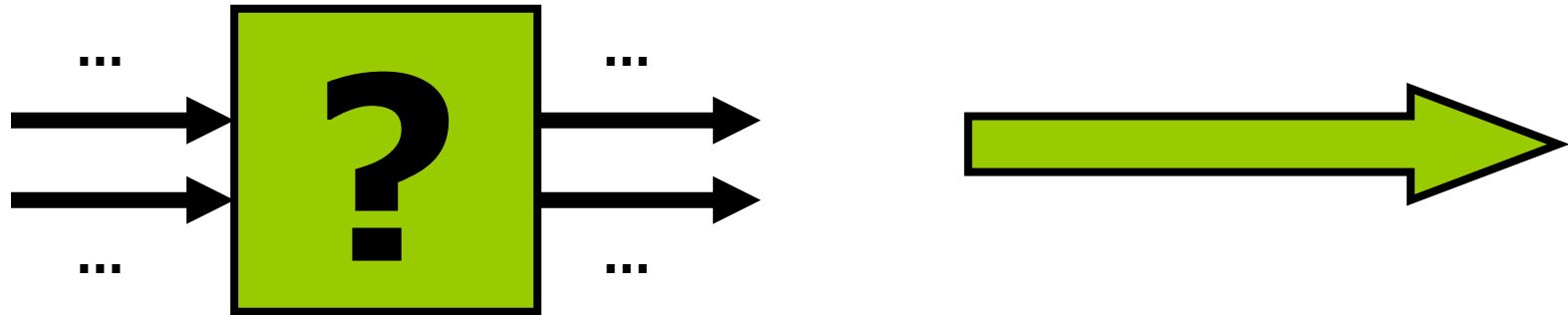
- Let's start with a demo!

Parallel Coordinates

- First proposed by Alfred Inselberg in *The Visual Computer* in 1985
- **Basic idea:** stack dimension axes in parallel, points become polylines
- **Advantage:** easy to add new dimensions
- As we add dimensions, the parallel coordinate diagram grows horizontally
- Another solution is to transform to **polar coordinates** and make radial axes
 - **Starplot** diagram



Elements and Dependencies



- Visual Elements

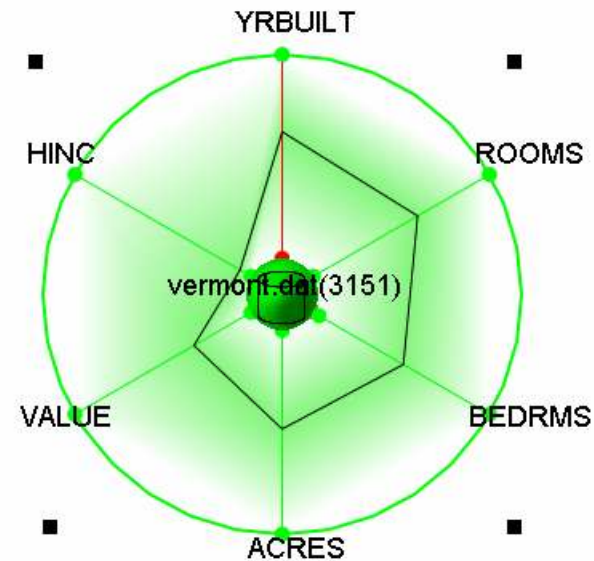
- Conforms to system-wide data format
- Visual appearance depending on input
- Input and output ports
- **Three types:** Sources, sinks, transformers

- Dependencies

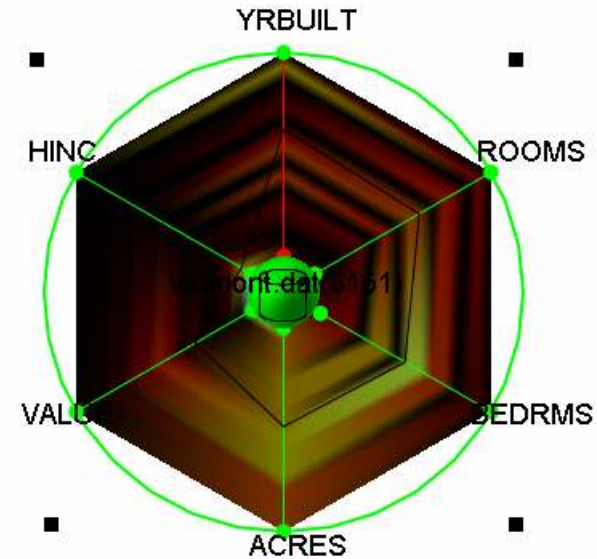
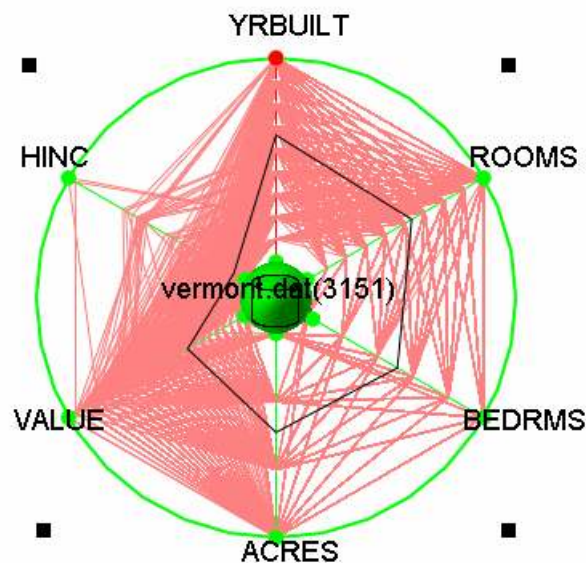
- Conforms to system-wide data format
- Directed link between two elements
- Propagates data from source to destination
- Interactive updates

The DataRose

- 2D starplot display
- **Transformer** visual element
- Average as black polyline
- Shows data distribution using different representations
 - Opacity bands [Fua et al. 1999]
 - Color histogram bands
 - Parallel coordinates

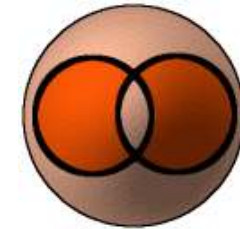
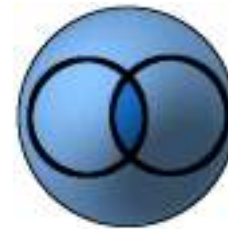
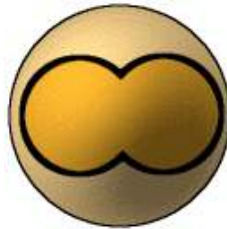


DataRose Representations



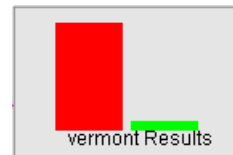
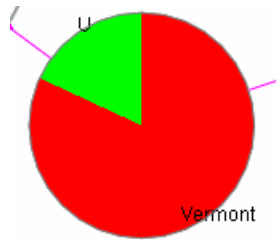
- Parallel coordinate mode
 - See all details
 - Cannot see distribution
- Color histogram mode
 - LOCS color scale
 - Brightness = high value

DataRose Types



- DataRoses can be of several different types
 - Each type represents a specific multi-set operation
- Four types:
 - **Source**: external database loaded from a file
 - **Union**: **all** input cases combined
 - **Intersection**: input cases that exist in **all** input sets
 - **Uniqueness**: input cases that exist in **one** input set

Viewers and Annotations



Vermont query

New York query

- Viewers are sink elements: accept input - no output
- Shows quantitative information for the inputs
 - Barchart
 - Piecharts
 - Histogram
- Annotations support communication of analyses
 - Labels
 - Notes
 - Images
 - Reports



Evaluation

- **Evaluation:** expert review (think-aloud protocol)
- **Participants:** two visualization researchers
- **Dataset:** US Census 2000
 - Three types of open-ended questions:
 - Direct facts: “What is the average house value in Georgia?”
 - Comprehension: “Which state has the highest ratio of small and expensive houses?”
 - Extrapolation: “Is there a relation between fuel type and building size in Alaska?”
- **Results:** positive, has lead to new design iterations
 - Participants were able to solve questions
 - Interaction quoted as main benefit



Contributions

- A highly interactive visual canvas (**DataMeadow**) for multivariate data analysis using multiple small visualization components
- A visual representation (**DataRose**) based on axis-filtered parallel coordinate starplots that can be linked together into interactive visual queries
- Results from a qualitative user study showing the use of our system for multivariate data analysis



Future Work

- Additional visual components for the DataMeadow
 - More complex visual queries
 - Additional annotation and communication support
- Non-standard input devices
 - Pen-based interfaces
- Non-standard output devices
 - Large displays
 - Collaborative visual analytics

Questions?



*Pictures courtesy of Helene Gregerström
Taken at the Atlanta Botanical Gardens*

- Contact information:
Niklas Elmqvist

E-mail: elm@lri.fr

WWW: <http://www.lri.fr/~elm/>

Phone: +33 1 69 15 61 97

Fax: +33 1 69 15 65 86



Aviz

Visual Analytics Project

<http://www.aviz.fr/>