

Ch 2: Data Abstraction

Tamara Munzner

Department of Computer Science
University of British Columbia

CPSC 547, Information Visualization

Day 3: 17 September 2015

<http://www.cs.ubc.ca/~tmm/courses/547-15>

News

- Waitlist update: 32 registered and waitlist cleared
- Signup sheet - add yourself if you weren't here before
 - probably just new auditors?

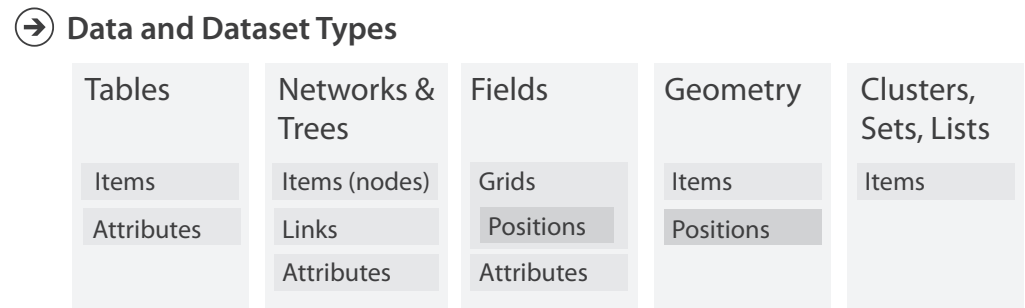
VAD Ch 2: Data Abstraction

What?

Datasets Attributes

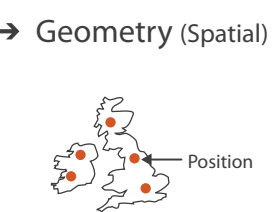
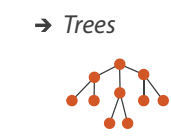
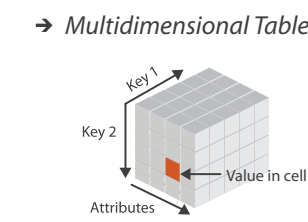
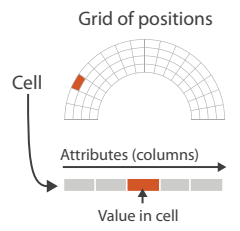
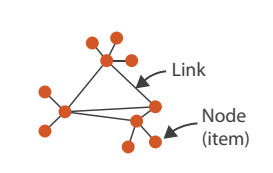
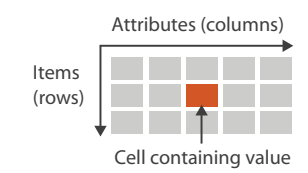
- Data Types
 → Items → Attributes → Links → Positions → Grids

- Attribute Types
 → Categorical
 + ● ■ ▲

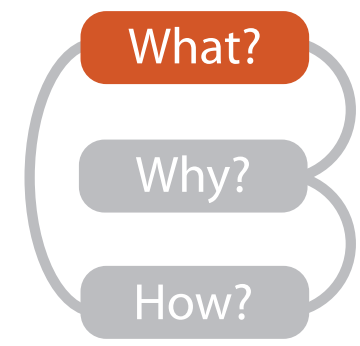


- Ordered
 → Ordinal
 ↗ ↘ ↕
- Quantitative
 | | |

- Dataset Types
 → Tables → Networks → Fields (Continuous)



- Ordering Direction
 → Sequential
 → Diverging
 → Cyclic

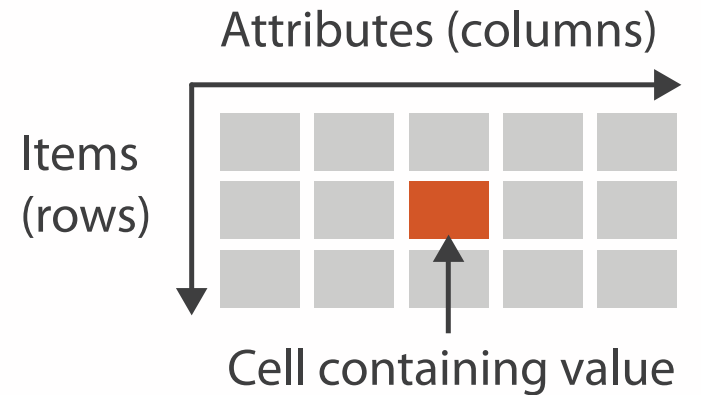


[VAD Fig 2.1]

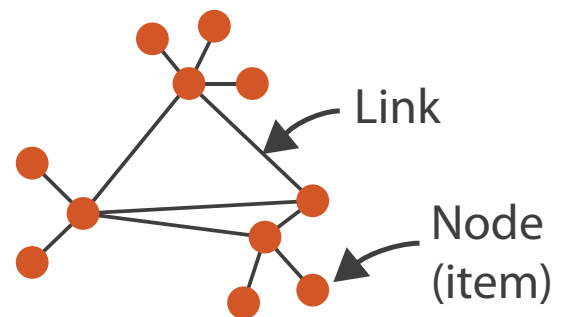
Dataset types

➔ Dataset Types

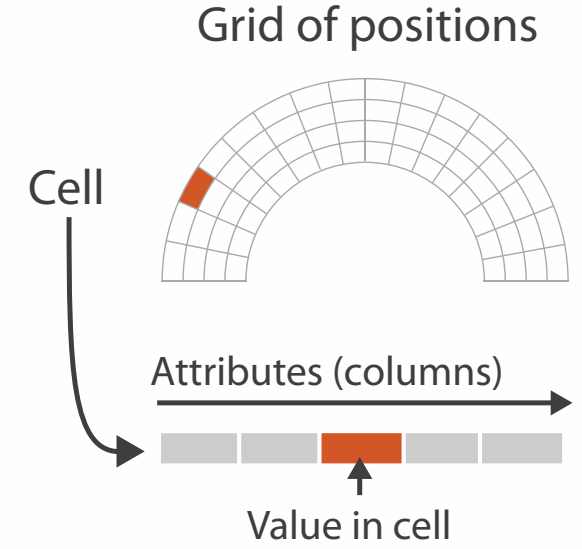
➔ Tables



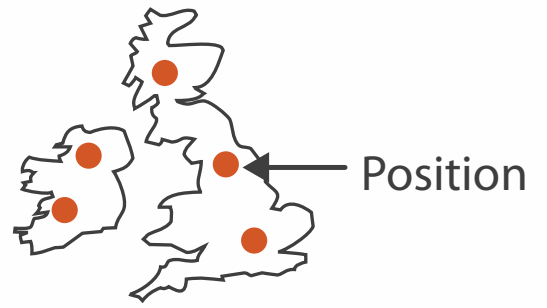
➔ Networks



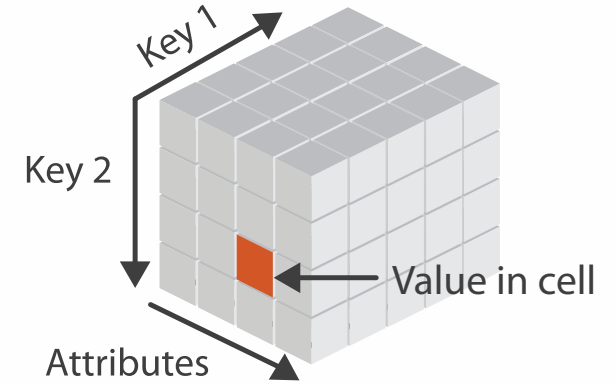
➔ Fields (Continuous)



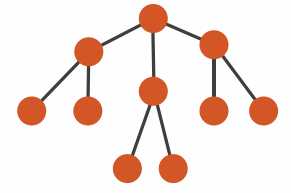
➔ Geometry (Spatial)



➔ *Multidimensional Table*

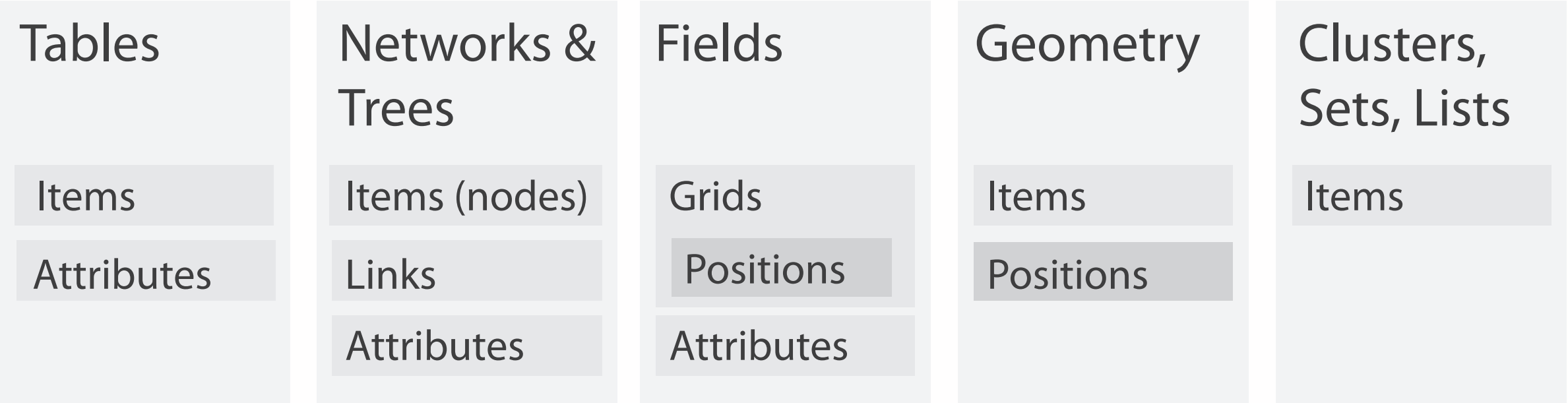


➔ Trees



Dataset and data types

→ Data and Dataset Types



→ Data Types

- Items
- Attributes
- Links
- Positions
- Grids

→ Dataset Availability

- Static
- Dynamic



Attribute types

➔ Attribute Types

➔ Categorical

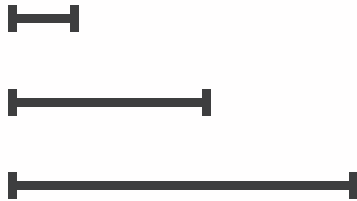


➔ Ordered

➔ *Ordinal*



➔ *Quantitative*



➔ Ordering Direction

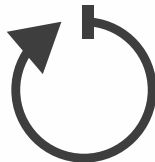
➔ Sequential



➔ Diverging



➔ Cyclic



Further reading:Articles

- Mathematics and the Internet:A Source of Enormous Confusion and Great Potential. Walter Willinger, David Alderson, and John C. Doyle. Notices of the AMS 56(5):586-599, 2009.
- Rethinking Visualization:A High-Level Taxonomy. InfoVis 2004, p 151-158, 2004.
- The Eyes Have It:A Task by Data Type Taxonomy for Information Visualizations Ben Shneiderman, Proc. 1996 IEEE Visual Languages
- The Structure of the Information Visualization Design Space. Stuart Card and Jock Mackinlay, Proc. InfoVis 97.
- Polaris:A System for Query,Analysis and Visualization of Multi-dimensional Relational Databases. Chris Stolte, Diane Tang and Pat Hanrahan, IEEE TVCG 8(1): 52-65 2002.

Further reading: Books

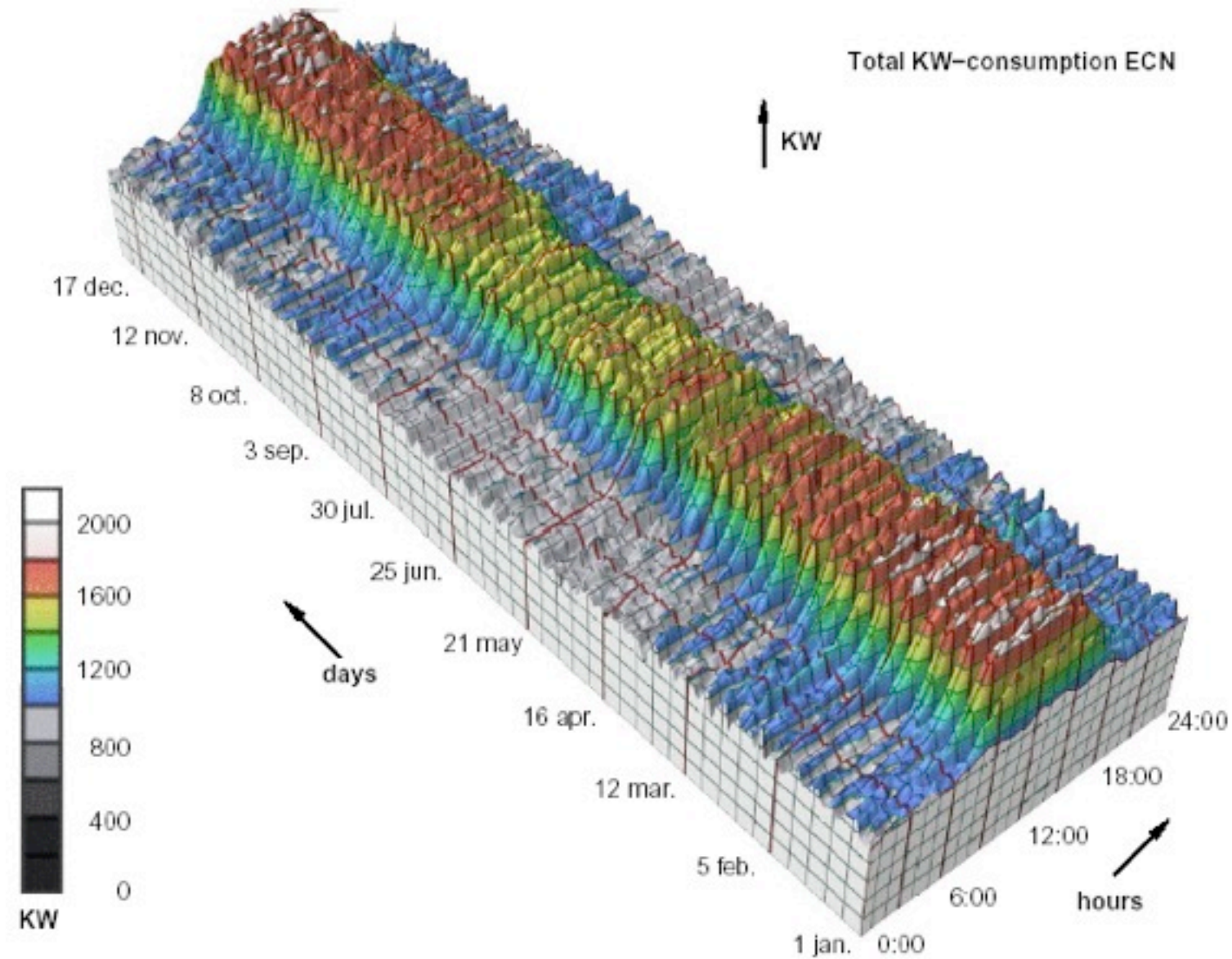
- Visualization Analysis and Design. Munzner. CRC Press, 2014.
 - Chap 2: Data Abstraction
- Information Visualization: Using Vision to Think. Stuart Card, Jock Mackinlay, and Ben Shneiderman.
 - Chap 1
- Data Visualization: Principles and Practice, 2nd ed. Alexandru Telea, CRC Press, 2014.
- Interactive Data Visualization: Foundations, Techniques, and Applications, 2nd ed. Matthew O. Ward, Georges Grinstein, Daniel Keim. CRC Press, 2015.
- The Visualization Handbook. Charles Hansen and Chris Johnson, eds. Academic Press, 2004.
- Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, 4th ed. Will Schroeder, Ken Martin, and Bill Lorensen. Kitware 2006.
- Visualization of Time-Oriented Data. Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, Chris Tominski. Springer 2011.

Now: In-class Design Exercise

- Five time-series data scenarios
 - A: every 5 min, duration 1 year, 1 thing: building occupancy rates
 - B: every 5 min, 1 year, 2 things: currency exchange rates
 - C: several years and several things: 5 years, 10 currencies
 - D: 1 year, many things: 1000 machines (CPU load)
 - E: 1 year, several parameters, many things: 1 year, 10 params, 1000 machines
- Group exercise: 15-20 min
 - one group per table (4 max), 10 groups total
 - discuss/sketch possible visual encodings appropriate for data assigned to your group
- Reportback: 20-30 min
 - 2-3 min from each group
- Design space: 15-20 min

Time-series data: Case A naive

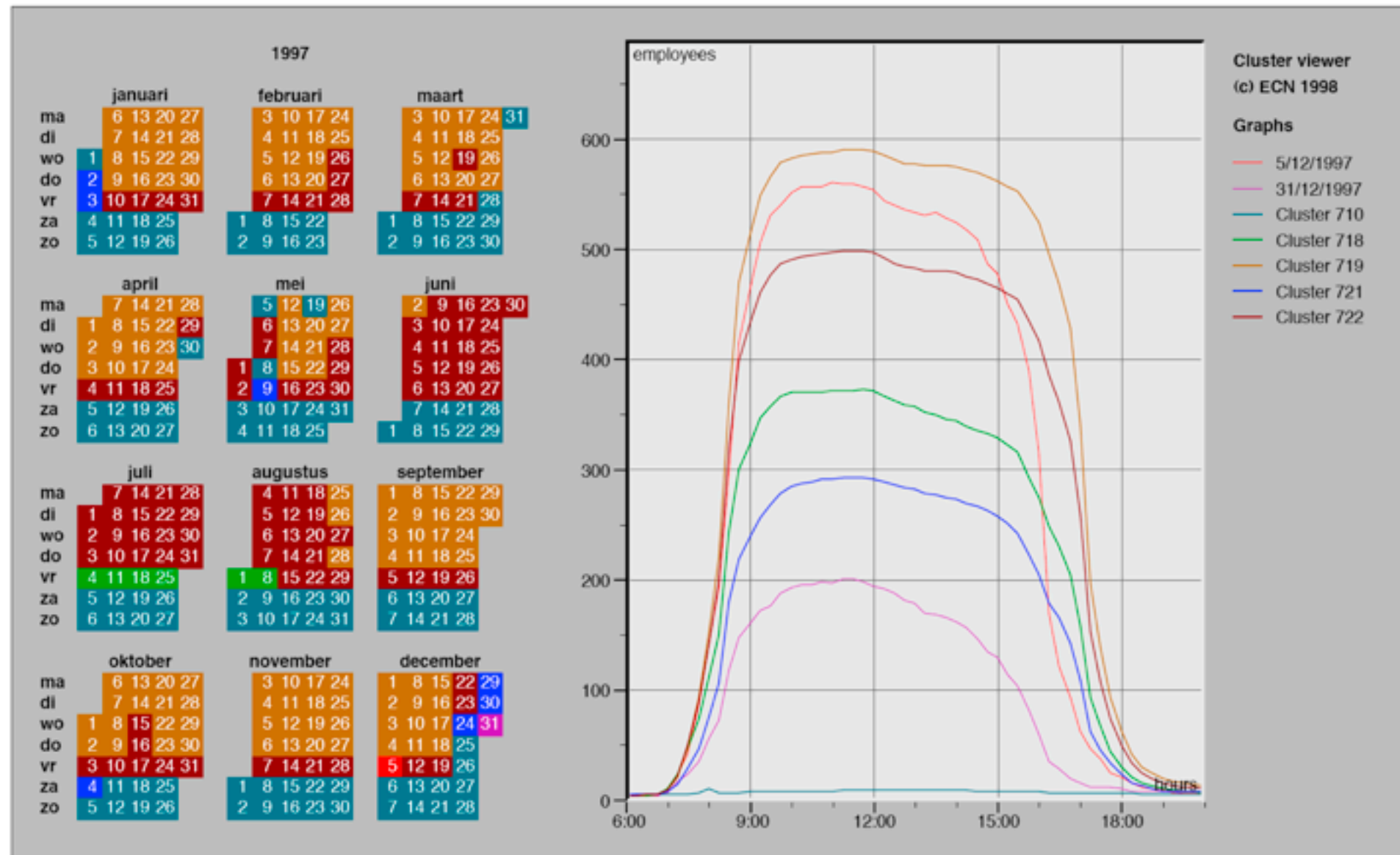
- extruded curves: detailed comparisons impossible



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Case A: Better Cluster-Calendar Solution

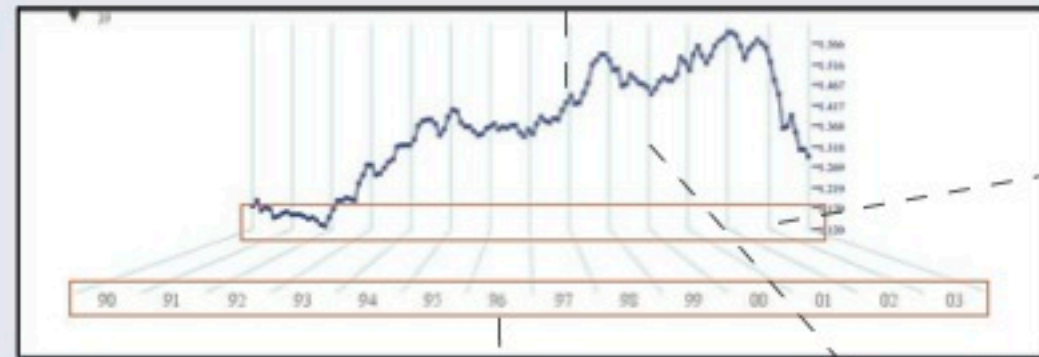
- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves



[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

Case A: BinX

Line Graph at a controllable aggregation level

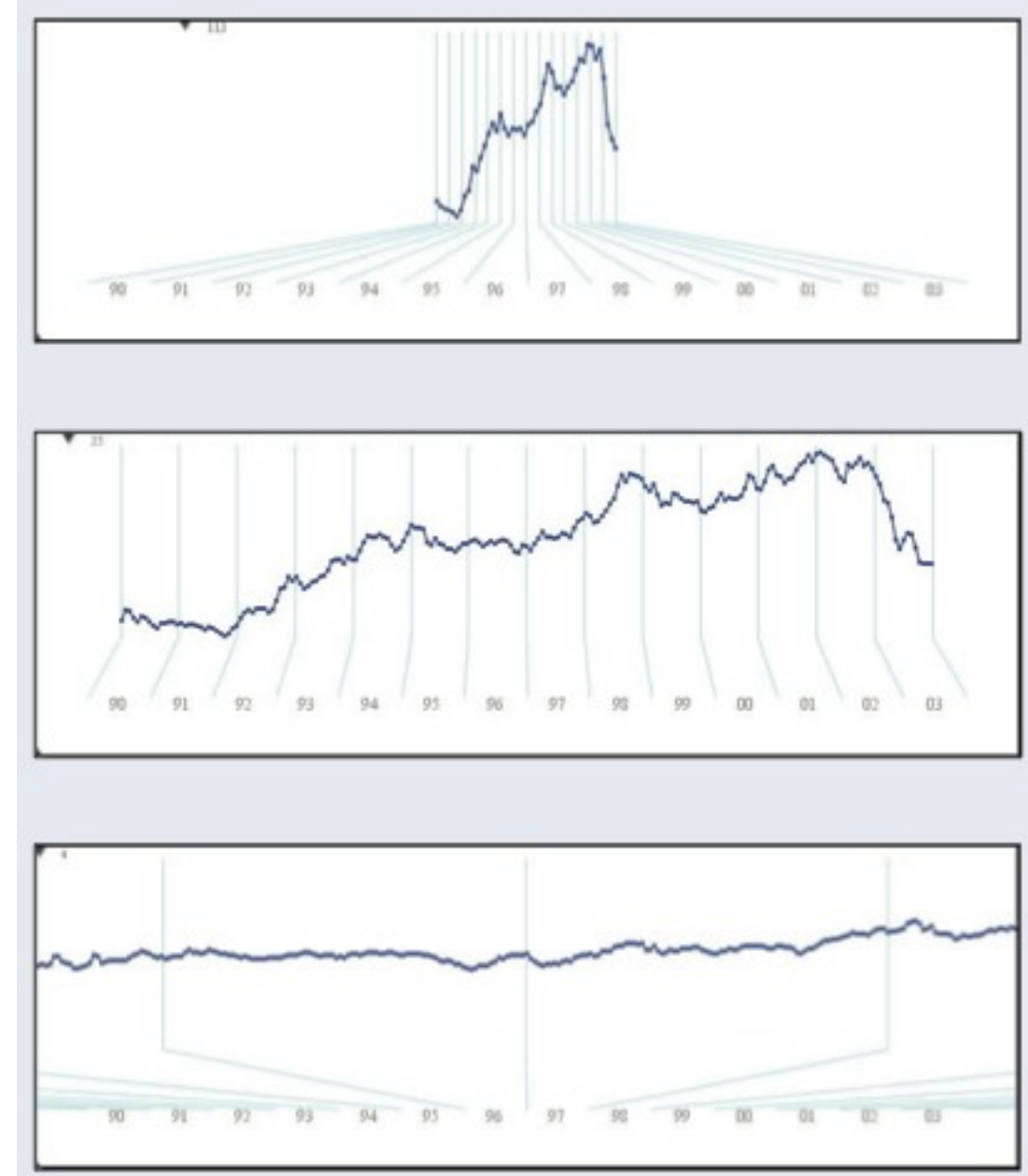


Vertical and slanted timelines are tick marks demarcating time intervals on the binned data

Fixed time coordinate axis shows full range of dataset

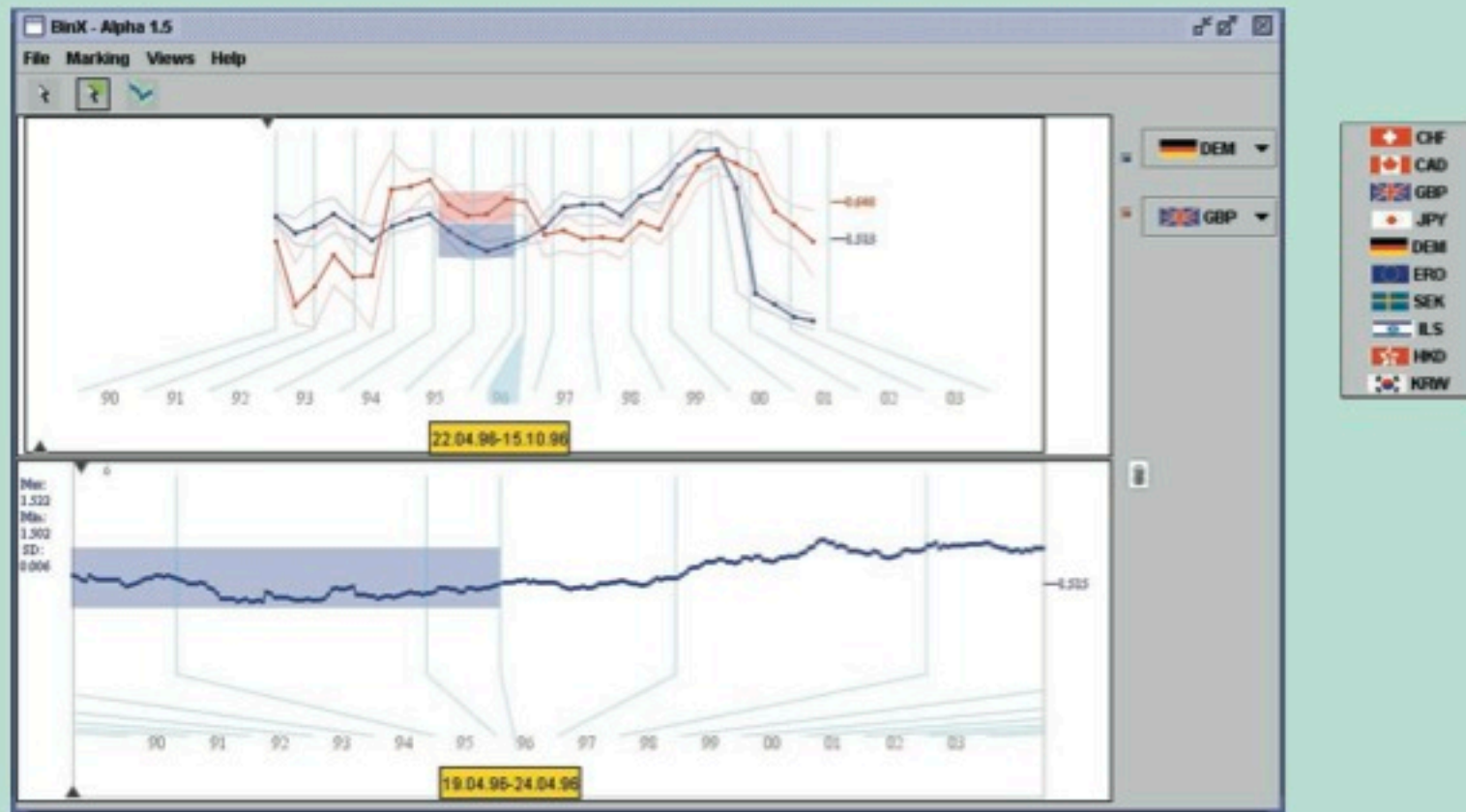


Optional Information on binned data



[BinX: Dynamic Exploration of Time Series Datasets Across Aggregation Levels. Lior Berry and Tamara Munzner.]
InfoVis 2004 Posters Compendium, pp 5-6.

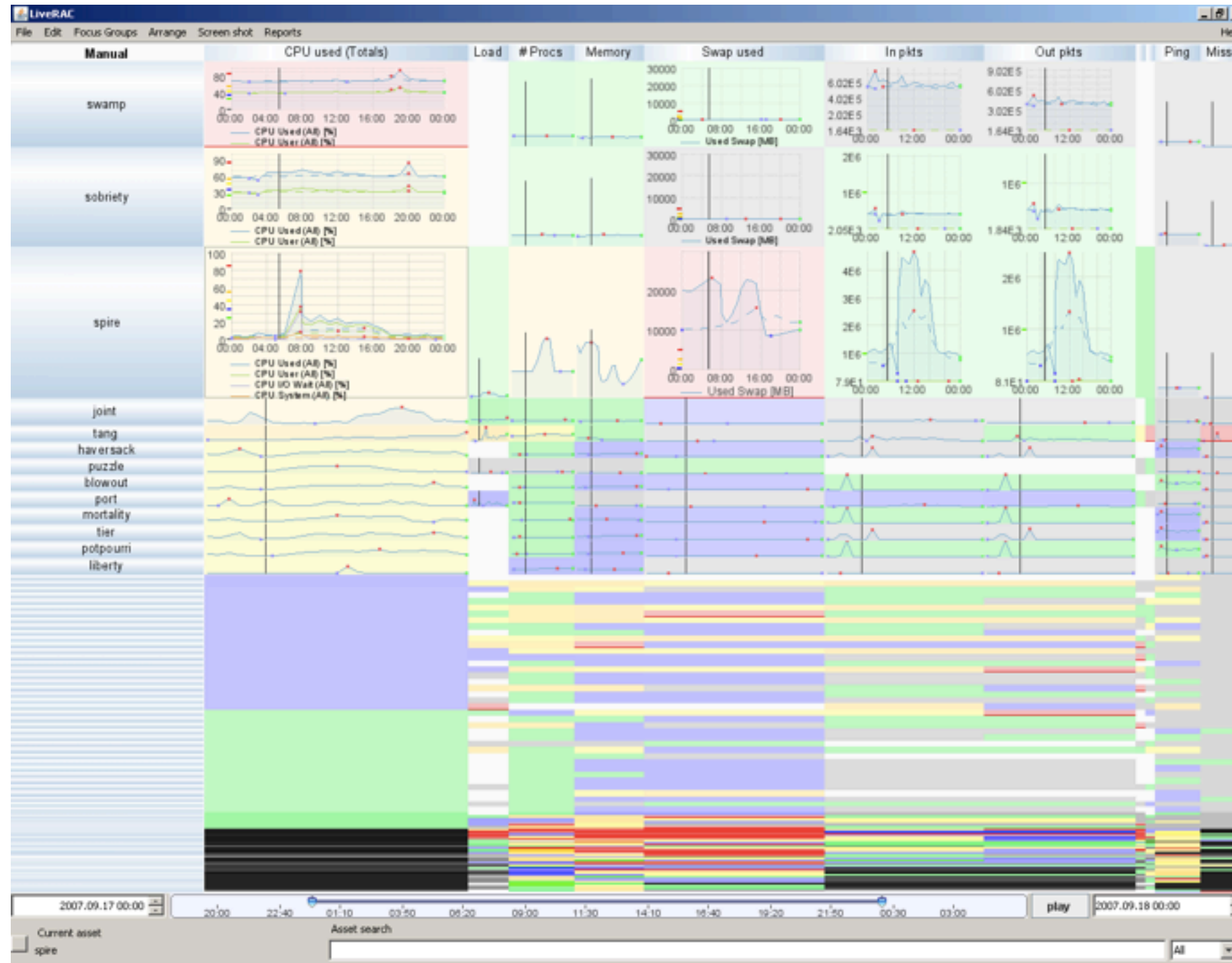
Case B:



- 5000 samples (daily rate over 15 years)
- Using two DTVC components
- Linked or separate navigation and marking
- Detail + Overview: two distinct aggregation levels

[BinX: Dynamic Exploration of Time Series Datasets Across Aggregation Levels. Lior Berry and Tamara Munzner.]
InfoVis 2004 Posters Compendium, pp 5-6.

Case E: LiveRAC video



<http://youtu.be/Id0c3H0VSkw>

[LiveRAC - Interactive Visual Exploration of System Management Time-Series Data. McLachlan, Munzner, Koutsofios, North. Proc. Conf. on Human Factors in Computing Systems (CHI) 2008, pp 1483-1492.]

Case E: LiveRAC data abstraction

- multidimensional table: time series data

- key attributes

- time

- 50,000: 5-minute intervals over 6 months
- multiscale levels of interest

- devices

- 4000

- parameters

- 20
- ex: CPU usage, memory load, network traffic, alarms, ...

- value attributes

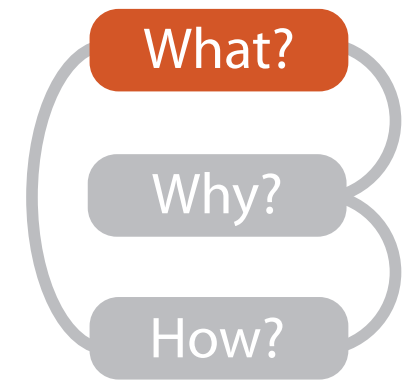
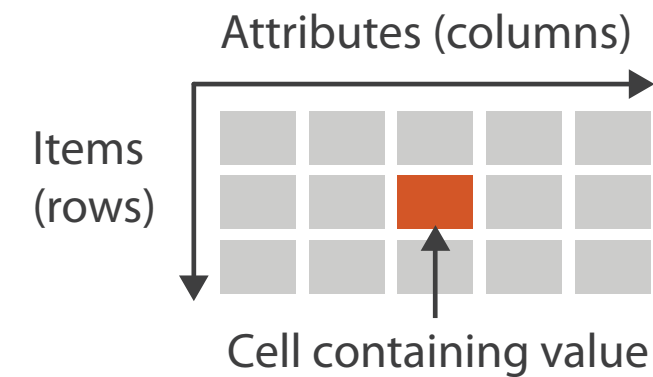
- parameter value for device at time point

- quantitative

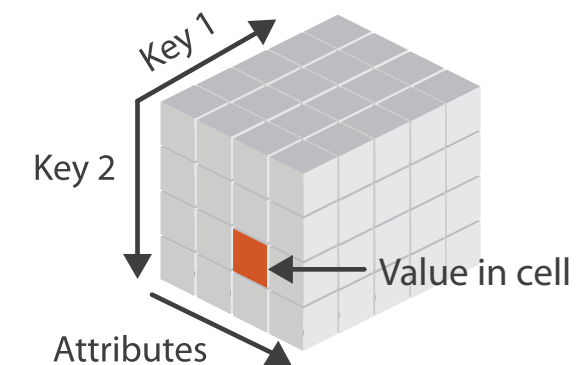
- device groups

- categorical

→ Tables



→ Multidimensional Table



⊕ Attribute Types

→ Categorical



→ Ordered

→ Quantitative



Next Time

- to read
 - VAD Ch. 3: Task Abstraction
 - Design Study Methodology: Reflections from the Trenches and the Stacks. Michael Sedlmair, Miriah Meyer, and Tamara Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2012), 18(12):2431-2440, 2012.
 - paper type: model