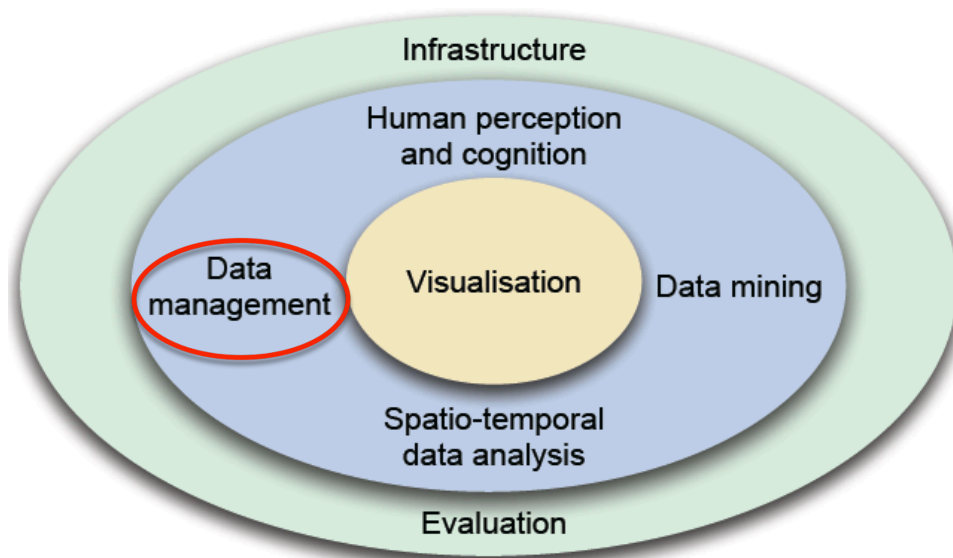


## Data Management

Dr. Michele C. Weigle

<http://www.cs.odu.edu/~mweigle/CS796-S11/>

### Building Blocks of Visual Analytics Research



# Data Management in Visual Analytics

## *Obstacles*

---

- ▶ Heterogeneity of data sources
  - ▶ integrate and query data from diverse sources
- ▶ Different data types
  - ▶ numeric, non-numeric, images, video, ...
- ▶ Streams of data
  - ▶ data source frequently produces new information

## Data Management

---

- ▶ Well-understood field
- ▶ Methods for effectively dealing with large datasets
  - ▶ ensure data consistency
  - ▶ avoid duplication
  - ▶ handle data transactions in a formal way
- ▶ Uses relational databases, SQL queries

# Data Management

## *State of the Art*

---

### ▶ Data Integration

- ▶ providing unified and transparent access to a set of heterogeneous sources

### ▶ Data Warehousing, OLAP, Data Mining

- ▶ data warehouse – integrated repository of data, optimized for complex decision-support queries (vs. relational transactions)
- ▶ OLAP (on-line analytical processing) – apps for interactive exploration of large multidimensional datasets
- ▶ data mining – discovering knowledge or patterns from massive amounts of data through ad hoc algorithms

# Data Management

## *State of the Art*

---

### ▶ Data Reduction and Abstraction

- ▶ used to obtain summary statistics
- ▶ improve interactivity of OLAP apps
- ▶ purpose to save computational or disk access costs

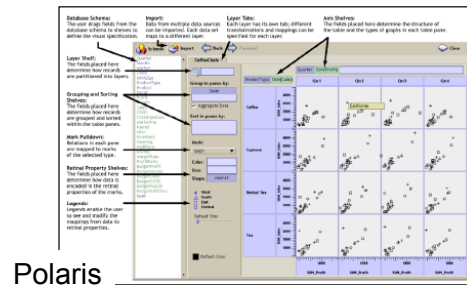
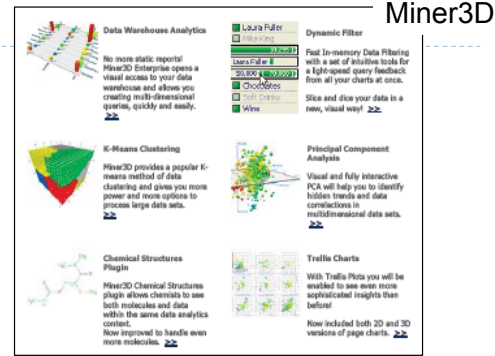
### ▶ Data Quality

- ▶ data conflicts – *duplicate records*
- ▶ missing data
- ▶ incorrect or inaccurate data

# Data Management and Info Vis

## State of the Art

- ▶ Visual Data Mining
- ▶ Visual OLAP
- ▶ Visual Data Reduction
  - ▶ visualizing large datasets often produces cluttered images - *clutter reduction techniques*
- ▶ Visualization for the Masses
  - ▶ e.g., GapMinder



▶ 7

CS 796/896 - Spring 2011 - Weigle

From Mastering the Information Age: Solving Problems with Visual Analytics, Keim et al.

# Data Management in Visual Analytics

## Challenges and Opportunities

- ▶ Uncertainty
- ▶ Data Integration
- ▶ Semantics Management
- ▶ Data Provenance and Integrity of Results
- ▶ Data Streaming
- ▶ Interactive Vis of Large Databases
- ▶ Distributed and Collaborative Visual Analytics
- ▶ Visual Analytics for the Masses

▶ 8

CS 796/896 - Spring 2011 - Weigle

From Mastering the Information Age: Solving Problems with Visual Analytics, Keim et al.

# Data Management in Visual Analytics

*Papers – Feb 8, Feb 15*

---

- ▶ P5 - Duplicate Record Detection: A Survey
- ▶ P6 - Give Chance a Chance: Modeling Density to Enhance Scatter Plot Quality Through Random Data Sampling
- ▶ P7 - A Taxonomy of Clutter Reduction for Information Visualisation
- ▶ P8 - Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering