From Visual Data Mining
towards Visual Analytics

Dr. Gennady Andrienko
Dr. Natalia Andrienko

http://geoanalytics.net
http://visual-analytics.info

October 2009
The Value of Visualisation

- Visualise: “to make perceptible to the mind or imagination”
  - Random House Webster’s College Dictionary

- “Visualisation is the process of representing abstract business or scientific data as images that can aid in understanding the meaning of the data.”

- “Visualisation offers a method for seeing the unseen.”

- “An estimated 50 percent of the brain's neurons are associated with vision. Visualization <…> aims to put that neurological machinery to work.”
  - Ibid.
Visual Data Mining pipeline

Source:
S.J. Simoff et al. (Eds.): Visual Data Mining, LNCS 4404, pp. 1–12, 2008
Examples of Visual Data Mining: Decision Trees

- Interactive tuning of decision trees by manipulating and visualizing
  - size of the node (number of training records corresponding to the node)
  - quality of the split (purity of the resulting partitions)
  - class distribution (frequency and location of the training instances of all classes)

Source: Ankerst & Ester & Kriegel, ACM KDD 2000
Examples of Visual Data Mining: Association Rules

Figure 3: (Labelled) Double Decker Plot of the mosaic

Figure 4: Example of a “good” association rule: the bin heineken & coke & chicken is filled almost entirely with highlighting, while none of the other bins is filled.

- Visual Inspection and interactive modification of association rules on mosaic plots

Source: Hofmann & Siebes & Wilhelm, ACM KDD 2000
Examples of Visual Data Mining: Subgroups

- Interpretation of subgroups in attribute and geographic spaces

Examples of Visual Data Mining: Clusters

- Visualisation of the attribute values statistics for the clusters in comparison to the whole dataset

Source: IBM DB2 Intelligent Miner; http://www-3.ibm.com/software/data/iminer/fordata/
Examples of Visual Data Mining: Hierarchical Clusters

- Visually-driven hierarchical clustering

Source: Seo and Shneiderman, InfoVis, 2004
Visual Analytics: Similar Techniques, Different Focus

- **Data Mining** is **computer-centred**:  
  - Computer performs data analysis, human somehow uses the results  
  - Visualisation may be involved for  
    - (mainly) helping the user to understand the results;  
    - (sometimes) enabling the user to select and prepare input data;  
    - (sometimes) enabling the user to direct the work of the algorithm

- **Visual Analytics** is **human-centred**:  
  - Human solves a complex problem, computer helps the human  
  - Visualisation is needed for activating the perceptual and cognitive capabilities of the human:  
    - perception of patterns;  
    - identification and association;  
    - abstraction and generalisation;  
    - reasoning and insight.
Computer and Human Can Work Synergistically

Computers
- can store and process great amounts of information
- are very fast in searching information
- are very fast in processing data
- can extend their capacities by linking with other computers
- can efficiently render high quality graphics, both static and dynamic

Humans
- are flexible and inventive, can deal with new situations and problems
- can solve problems that are hard to formalise
- can reasonably act in cases of incomplete and/or inconsistent information
- can simply see things that are hard to compute
- can employ their knowledge and experience
The Goal of Visual Analytics

- Visual analytics must develop solutions
  - enabling analysts to focus their **full perceptual and cognitive capabilities** on their analytical processes
  - while allowing them to apply **advanced computational capabilities** to augment their discovery process
Visual Analytics Integrates Scientific Disciplines to Improve the Division of Labour between Human and Machine

Machine
- Statistical Analysis
- Data Mining
- Data Management
- Compression & Filtering

Human
- Human Cognition
- Perception
- Visual Intelligence
- Decision Making Theory

Semantics-based approaches

Information Design

Information Visualization

Graphics and Rendering

"The best of both sides"
Definition of Visual Analytics

Visual Analytics is the science of analytical reasoning facilitated by interactive visual interfaces.

People use visual analytics tools and techniques to

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, and understandable assessments
- Communicate assessment effectively for action

Components of Visual Analytics

- **Analytical reasoning**
  - How to maximise human capacity to perceive, understand, and reason about complex and dynamic data and situations?

- **Visual representations and interaction techniques**
  - How to augment cognitive reasoning with perceptual reasoning through visual representations and interaction?

- **Data representations and transformations**
  - How to transform data into a representation that is appropriate to the analytical task and effectively conveys the important content?

- **Production, presentation, and dissemination**
  - How to convey analytical results in meaningful ways to various audiences?
Emergence of Visual Analytics

Initially driven by the USA Homeland Security…

…but now has a much broader scope and impact
Conferences, symposia, workshops

Visualization, Analytics & Spatial Decision Support

Call for papers for the Workshop on Visualization, Analytics & Spatial Decision Support at the GIScience conference (September 20, 2006, Münster) and for a special issue of the International Journal of Geographical Information Science

University courses and seminars

Visualization and MultiMedia Lab

Department of Informatics, University of Zürich

VisMaster: Visual Analytics - Mastering the Information Age

What is VisMaster?

Visual Analytics - Mastering the Information Age

VisMaster is a European Coordination Action Project focused on the research discipline of Visual Analytics. One of the most important challenges of the emerging Information Age is to
Visual Analytics Aims at Supporting the Whole Analytical Process

- Plan the process
- Gather relevant information and become familiar with it
- Incorporate the relevant information with the existing knowledge
- Generate candidate explanations (hypotheses)
- Evaluate the hypotheses in light of evidence and assumptions
- Develop a judgement about the most likely explanations or outcomes
- Try to find other possible explanations that were not previously considered
- Draw conclusions
- Create a report or presentation of the results; explain why
- Collaboratively review the results and the arguments (with colleagues and/or external experts)
- Share the results with customers or other audience
Analytical Discourse

The issue to be addressed
- understand
- refine

Prior knowledge

Evolving knowledge

Judgements

Assumptions

Hypotheses

Relevant information

Supporting technology

Raw data

Transformed data

Supporting technology

- store
- organize
- visualize

- transform
- visualize

- link
- evaluate

- perceive
- understand
- select

- store
- organize
- visualize

- understand
- refine
Supporting Technology

- Data pre-processing and computer-adapted representation
  - e.g. extraction of structured data from images, video, texts
- Data transformations
  - e.g. aggregation; clustering; dimensionality reduction; interpolation; smoothing
- Automatic extraction of potentially interesting features and patterns (relations, regularities, anomalies, trends)
- Techniques for hypotheses testing (statistics)
- Annotation support
- Support for workspaces and workflows
- Support for collaborative analyses

Visualization of data (original and derived)

Visualization of derived knowledge, argumentation, and analysis process
Example of Data Pre-processing (Text Processing)

Text Documents

FEMA* Situation Updates

Statistical Text Extraction/Processing Module
- Entity Extraction
- Relation Extraction
- Topic-based Segmentor
- Topic Classifier
- Event Detector

Disambiguation Module
- Named Entity Disambiguation
- Geographical Disambiguation
- Temporal Disambiguation

Chi-Chun Pan, Prasenjit Mitra
Pennsylvania State University

* Federal Emergency Management Agency
Example of Data Transformation: Aggregation, Smoothing

Events (traffic accidents)  Densities of events

Darya Filippova, Joonghoon Lee, Andreea Olea, Michael VanDaniker, Krist Wongsuphasawat
University of Maryland, College Park
Example of Data Transformation: Clustering, Classification

Diansheng Guo, Jin Chen, Alan M. MacEachren, Ke Liao
University of South Carolina; Pennsylvania State University
Example of Data Transformation: Clustering, Dimensionality Reduction

Time series of 2 variables clustered using SOM (Self-Organizing Map)

Tobias Schreck, Tatiana Tekušová, Jörn Kohlhammer, Dieter Fellner
Technische Universität Darmstadt
Fraunhofer IGD Darmstadt
Example of Feature Extraction

Machon Gregory, Anthony Don, Elena Zheleva, Sureyya Tarkan, Catherine Plaisant, Ben Shneiderman
University of Maryland, College Park
Example of Annotation Support

- Choose Annotation Style
- Set Annotation Color
- Delete Annotations

- New Text Section
- Text Section Color
- Hide Sections by Color
- Take Snapshot
- Report Title
- Section Title
- Clickable Link to snapshot
- Snapshot Preview Panel

Ryan Eccles, Thomas Kapler, Robert Harper, William Wright
Oculus Info Inc.
Example of Workspace Support

Relevant Information, Hypotheses, Evidence, Arguments, ...

Pascale Proulx, Sumeet Tandon, Adam Bodnar, David Schroh, Robert Harper, William Wright
Oculus Info Inc.
Example of Workflow and Workspace Support

Yedendra B. Shrinivasan, Jarke J. van Wijk
Technische Universiteit Eindhoven

From Visual Data Mining towards Visual Analytics
October 2009
Examples of Support for Collaborative Analyses
(synchronous, co-located collaboration)

Alan M. MacEachren, Isaac Brewer
Pennsylvania State University
Example of Support for Collaborative Analyses
(asynchronous, distributed collaboration)

Jeffrey Heer, Fernanda B. Viégas, Martin Wattenberg
University of California, Berkeley
Conclusion

- Visual Analytics science and technology is meant to help people to make sense from complex data and solve complex problems
- Complexities: massive amounts, high dimensionality, heterogeneity, multiple facets, time variance, incompleteness, uncertainty, inconsistency
- Visual Analytics combines interactive visual interfaces with algorithmic methods for data pre-processing, transformation, and feature/pattern extraction
- Visual Analytics also includes interactive visual tools supporting reasoning, knowledge synthesis, and knowledge management
- Interactive visual interfaces help analysts to utilize their perceptual and cognitive capabilities fully and effectively
- Computer technologies compensate for the natural limitations in human skills and abilities and augment the discovery process
- The ultimate goal is to enable a synergistic collaboration of human and computer where each side can utilize its intrinsic capabilities in the best possible way