Visual Analytics - Introduction

Eduard Gröller

Institute of Computer Graphics and Algorithms

Vienna University of Technology

Goals of VA [VisMaster, 2010]

Creation of tools and techniques to enable people 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 these assessment effectively for action

What is Visual Analytics?

"Visual Analytics is the science of analytical reasoning supported by a highly interactive visual interface." [Wong and Thomas 2004]

"Visual Analytics combines **automated analysis** techniques with **interactive visualisations** for an effective **understanding, reasoning and decision making** on the basis of **very large** and **complex datasets**" [Keim 2010]

Detect the expected and discover the unexpected

Visual Analytics Process

First step: preprocess and transform data

- Data cleaning, normalization, grouping, data fusion

Automated methods

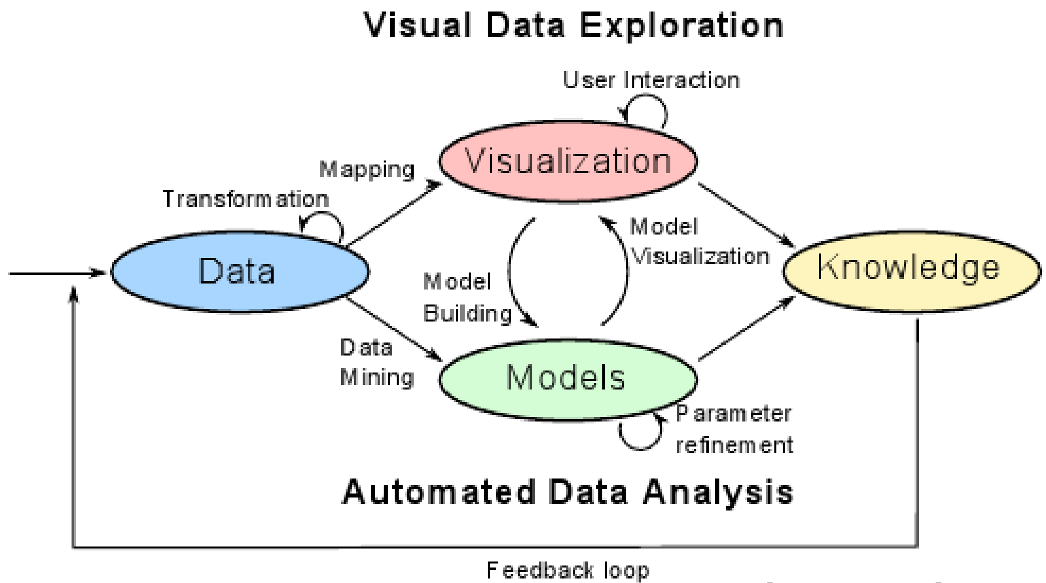
- + Scale well
- - Get stuck in local optima
- - Run in a black box fashion

Visualization

- + Interactive data analysis
- - Scalability

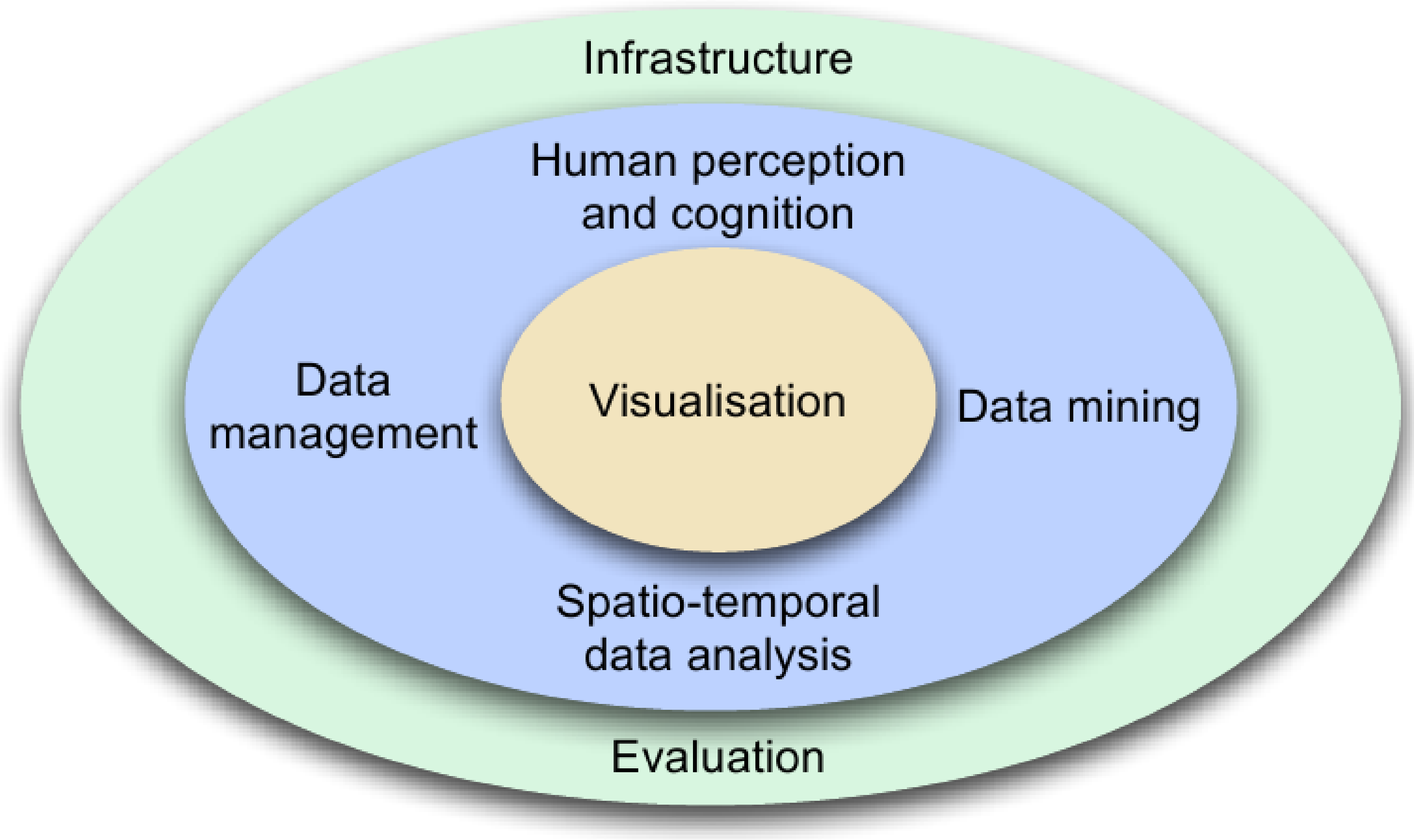
Visual Analytics integrates both

- Tied together by the user
- Alternating between visual and automatic methods



[Keim 2006]

Interdisciplinary!



Challenges

Data

Dealing with very large, diverse, variable quality datasets

Users

Meeting the needs of the users

Design

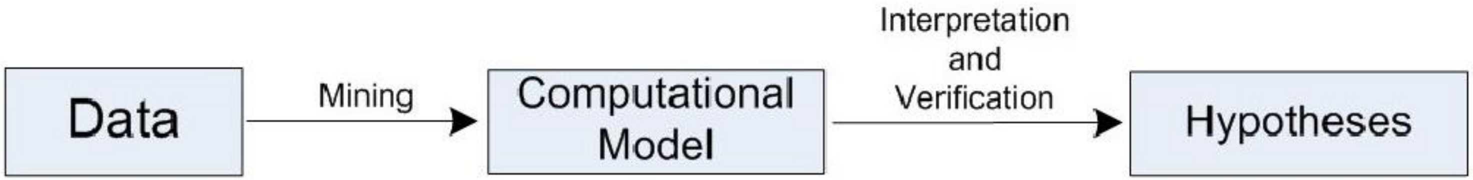
Assisting designers of visual analytic systems

Technology

Providing the necessary infrastructure

Data Mining Definition

Automatic algorithmic extraction of valuable information from raw data



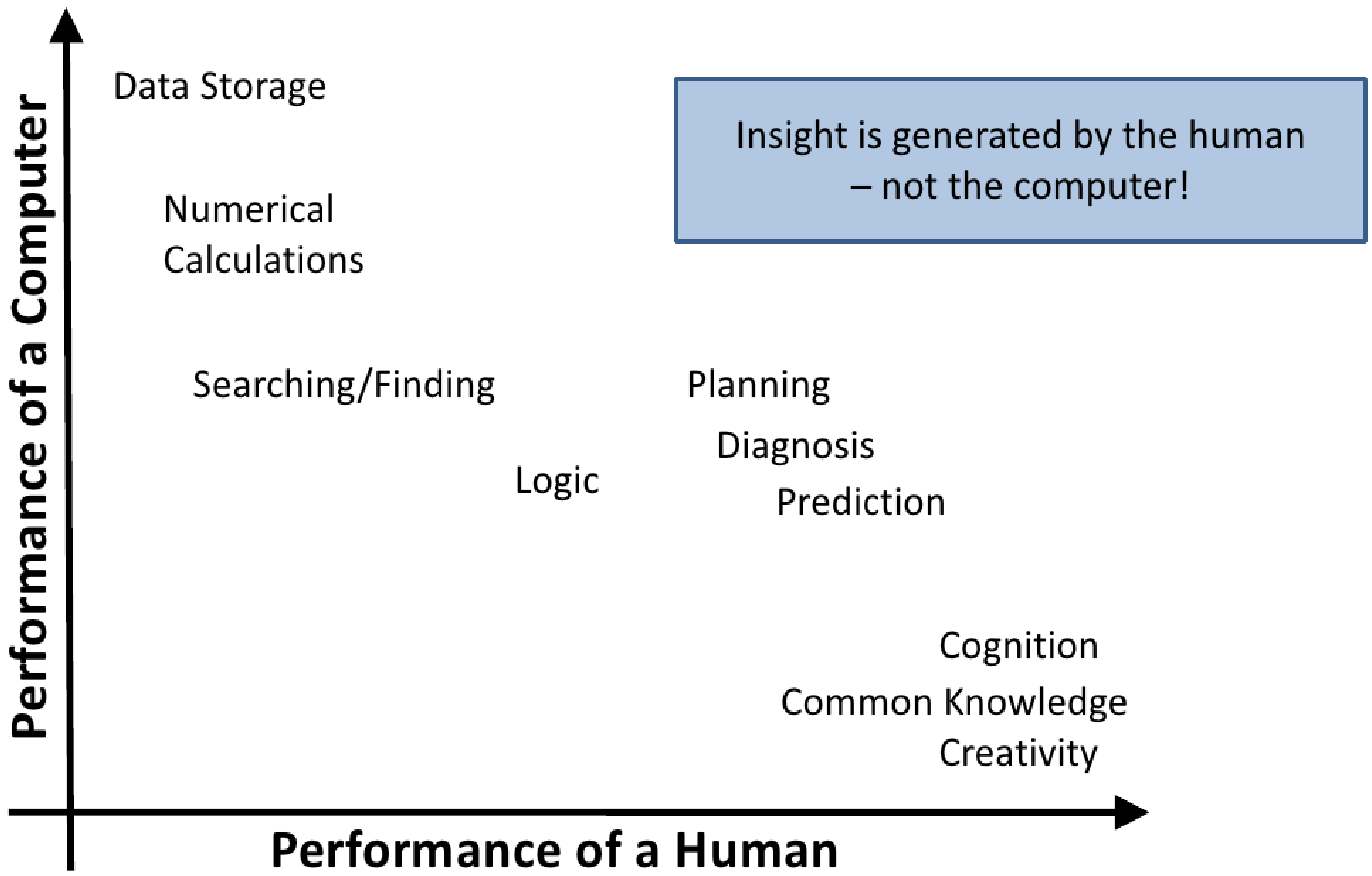
Knowledge Discovery and Data Mining (KDD)

- Semi or fully automated analysis of massive data sets

- Contributions are more about general methodologies

- Black-box methods in the hands of end users
  - Users need to understand the algorithms for using them
  - What attributes to use? What similarity measure? etc.
  - Often trial and error

The Ability Matrix



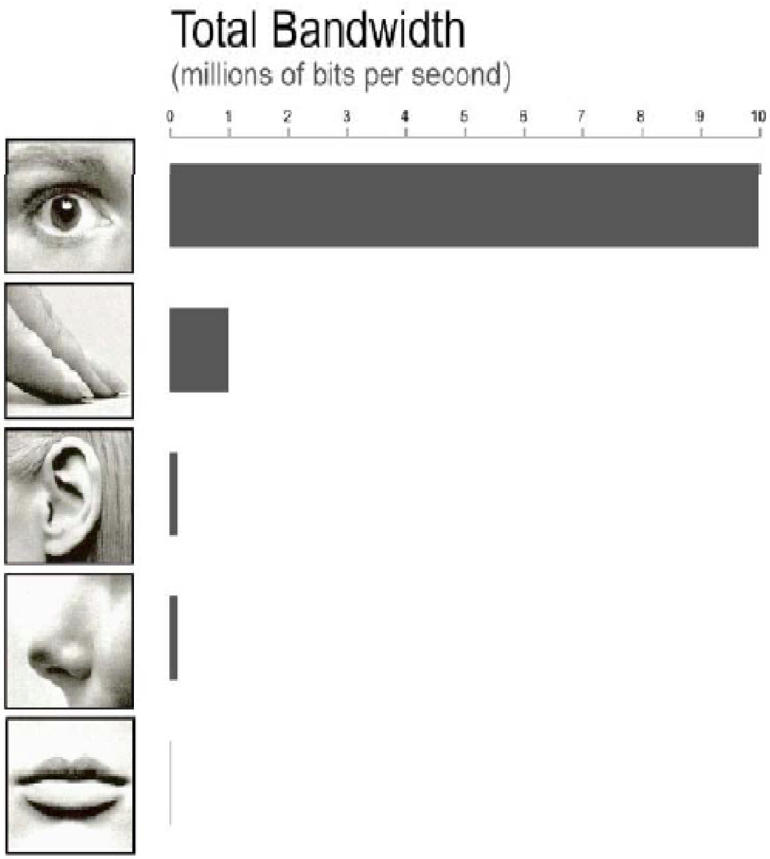
adapted from Daniel Keim, Uni. Konstanz

Why Graphics?

Figures are richer; provide more information with less clutter and in less space.

Figures provide the 'gestalt' effect: they give an overview; make structure more visible.

Figures are more accessible, easier to understand, faster to grasp, more comprehensible, more memorable, more fun, and less formal.



list adapted from: [Stasko et al. 1998]

Statistics vs. Visualization: Anscombe's Quartet

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

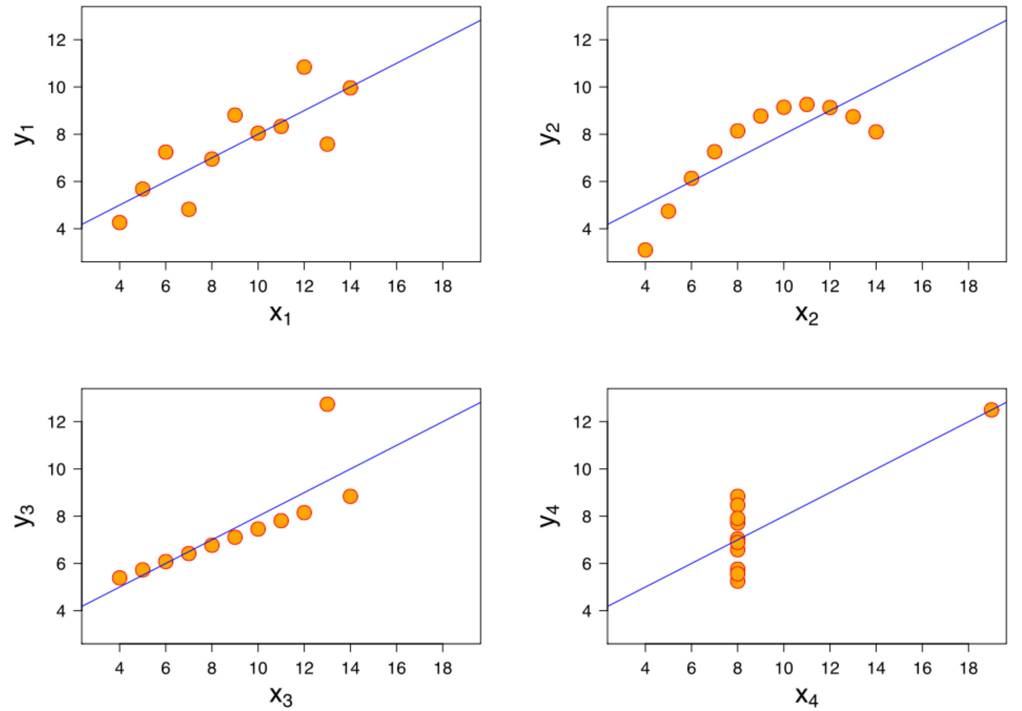
Statistics vs. Visualization: Anscombe's Quartet

Statistics profile is the same for all!

Property	Value
Mean of x in each case	9 (exact)
Variance of x in each case	11 (exact)
Mean of y in each case	7.50 (to 2 decimal places)
Variance of y in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

Anscombe’s Quartet

Four datasets that have identical simple statistical properties, yet appear very different when graphed.

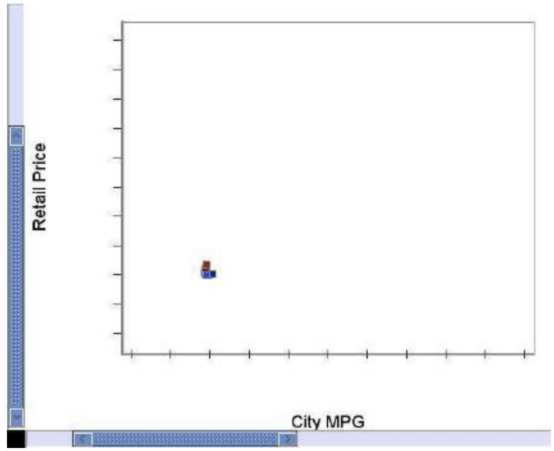


Wikimedia Commons

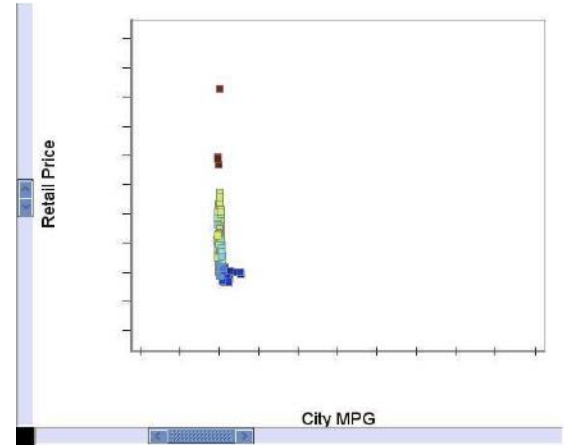
Property	Value
Mean of *x* in each case	9 (exact)
Variance of *x* in each case	11 (exact)
Mean of *y* in each case	7.50 (to 2 decimal places)
Variance of *y* in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between *x* and *y* in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

Visualization Can Be Biased

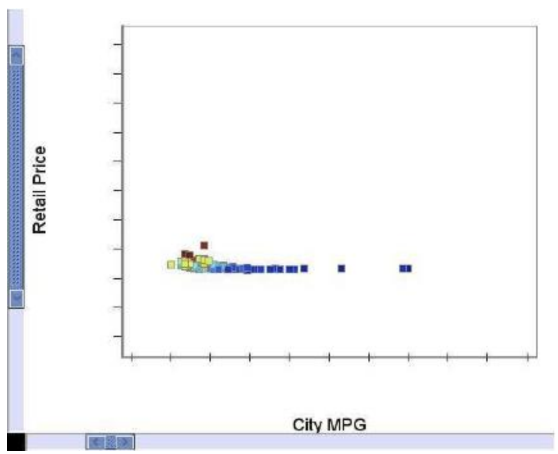
The same data plotted with different scales is perceived dramatically differently.



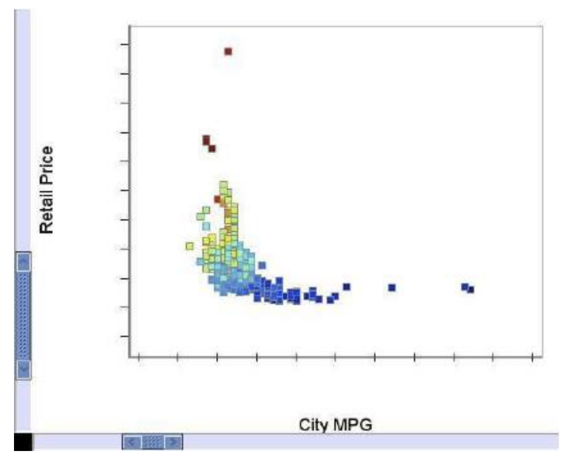
(a) Equally (uniformly) large scale in both *x* and *y*



(b) Large scale in *y*



(c) Large scale in *y*



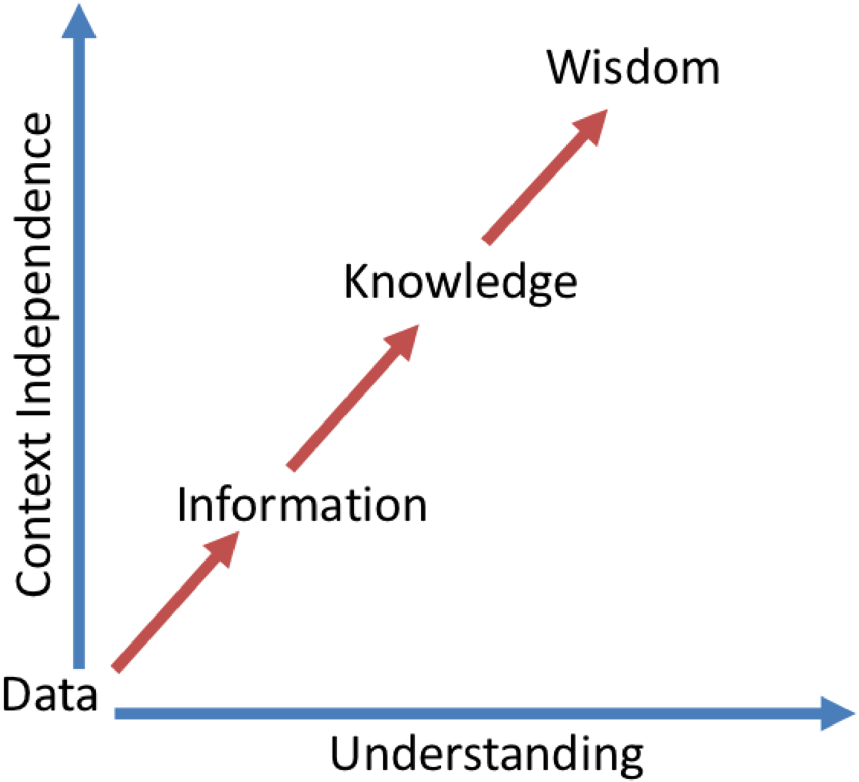
(d) Scale determined by range of *x*- and *y*-values.

[Ward, Grinstein, Keim 2011]

Diagram vs. Visualization

A diagram represents **information**.

A visualization represents **data**.



The DIKW-Hierarchy
according to [Bellinger 2004]

From Epistemology:
- Data = ground truth
- Information = phenomena
- Knowledge = causes
- Wisdom = possible (inter-)actions

"Above all else, show the data"
- Edward R. Tufte

Mantras

Guide to visually explore data - Shneiderman‘s Mantra:

Overview first, zoom/filter, details on demand

[Shneiderman, 1996]

Describes how data should be presented on screen

For massive datasets it is difficult to create overview without loosing interesting patterns

Extended Mantra for VA:

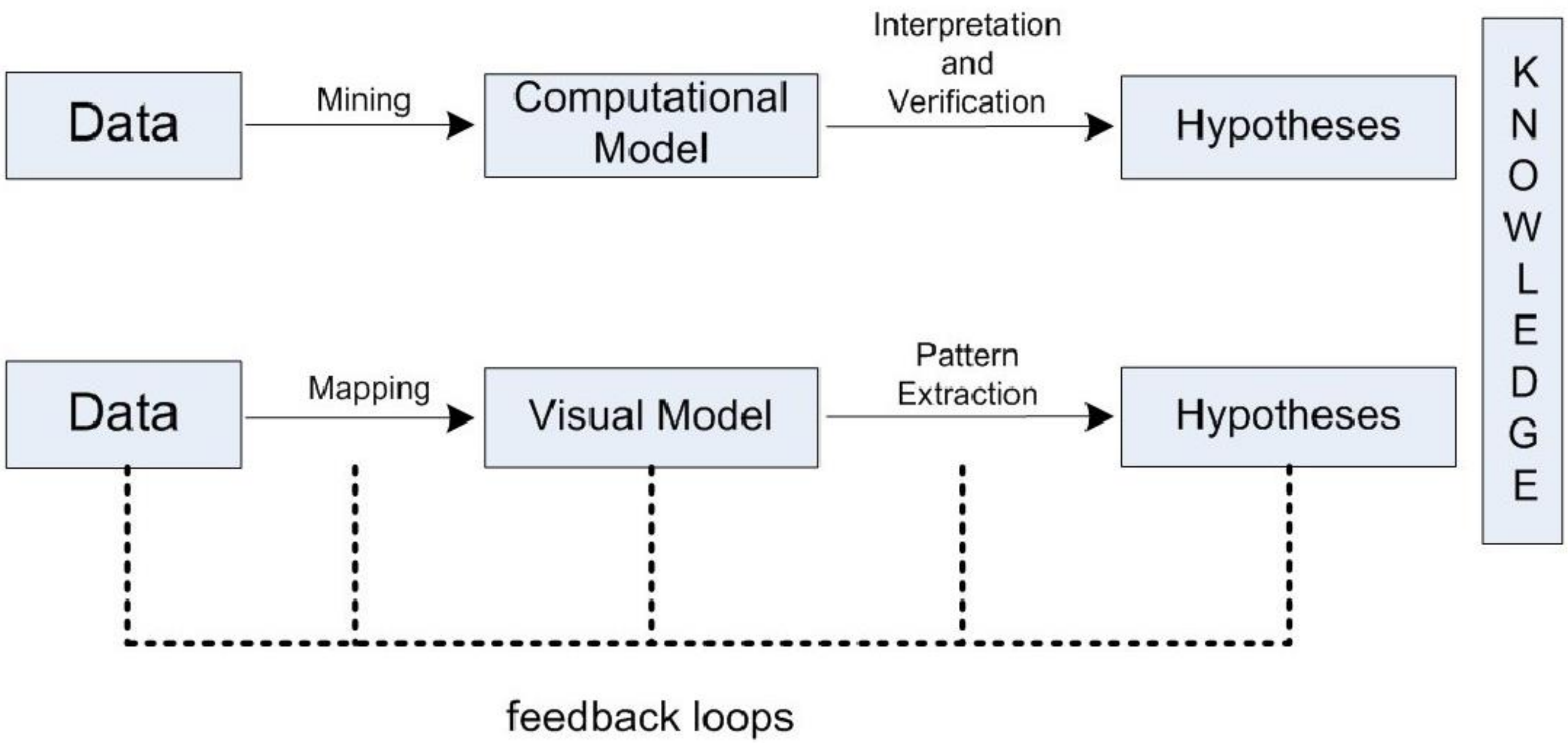
Analyse first, show the important, zoom/filter, analyse further, details on demand

[Keim, 2006]

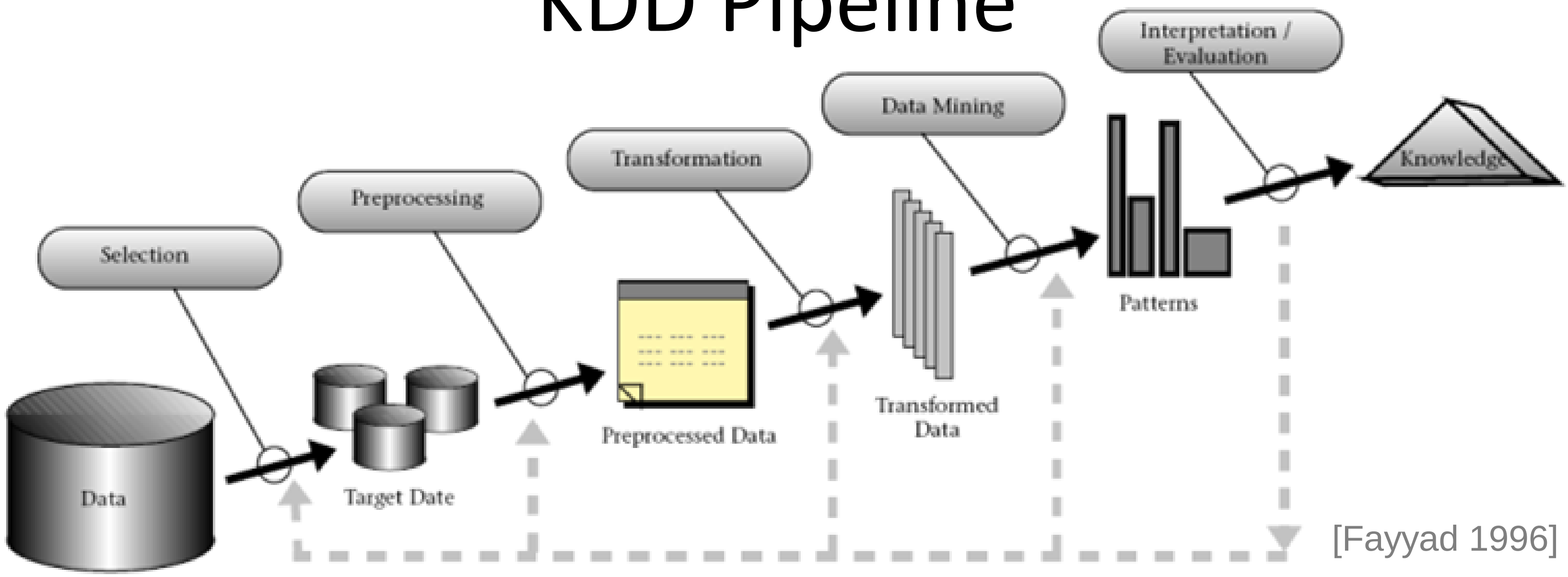
Institute of Computer Graphics

JKU JOHANNES KEPLER UNIVERSITY LINZ

Traditional Data Mining vs. Visual Analysis Processes

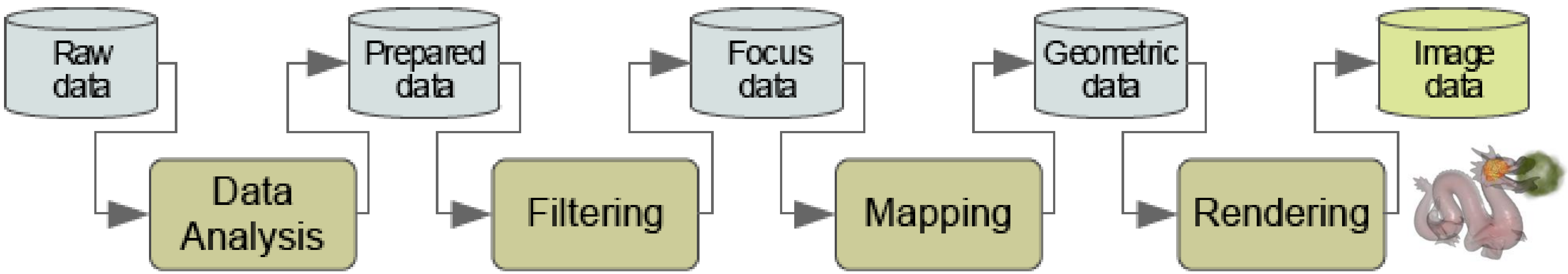


KDD Pipeline



[Fayyad 1996]

Visualization Pipeline



[dos Santos and Brodlie 2004]

Uncertainty

- What is not surrounded by uncertainty cannot be the truth [Richard Feynman]
- True genius resides in the capacity for evaluation of uncertain, hazardous, and conflicting information [Winston Churchill]
- Doubt is not a pleasant condition, but certainty is absurd [Voltaire]

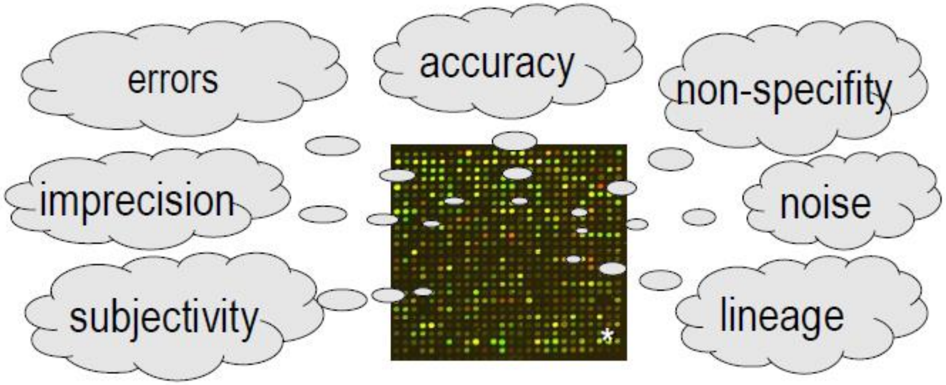
Uncertainty

Definition

"Degree to which the lack of knowledge about the amount of error is responsible for hesitancy in accepting results and observations with caution" [Hunter 1993]

Measurement data

e.g., DNA microarray expression data



[Holzhüter 2010]

Can be handled in data or view space

Data Management Challenges

“ Big Data“

Uncertainty

Semantics Management

Data Streaming

Distributed and Collaborative VA

VA for the Masses

What is " Big Data"?

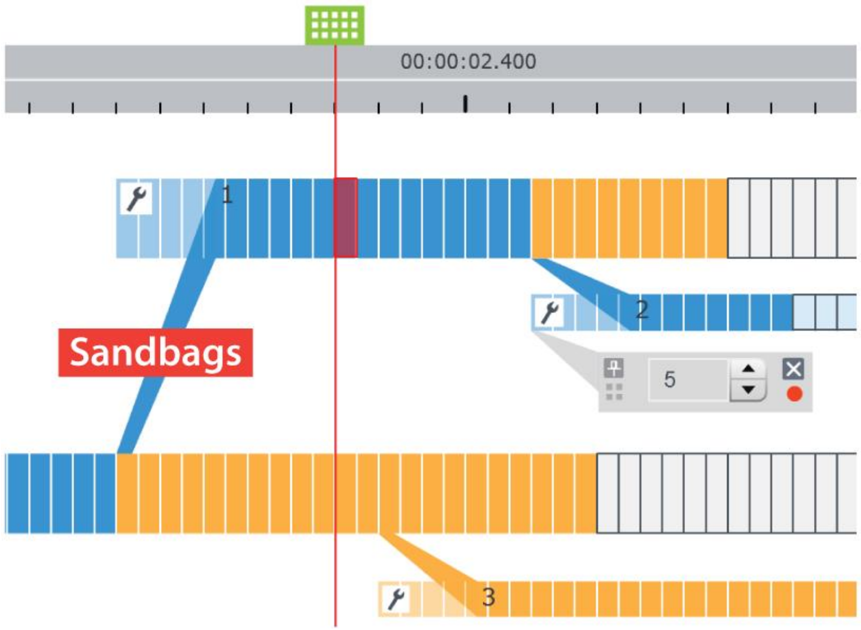
→ Moving target

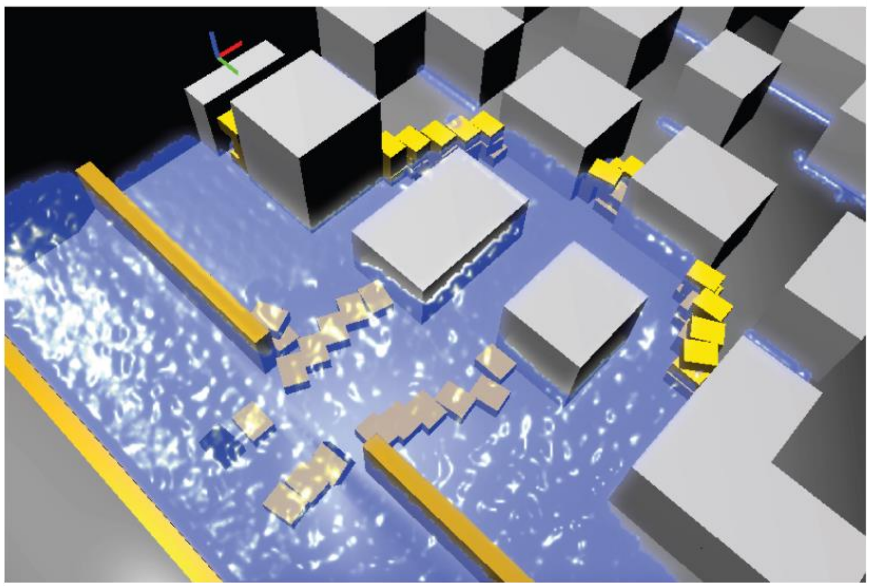
Fields dealing with this kind of data:

- Meteorology
- Genomics
- Connectomics
- Complex physics simulations
- Biological and environmental research
- Business intelligence

Multiples of bytes					V · T · E
SI decimal prefixes		Binary usage	IEC binary prefixes		
Name (Symbol)	Value		Name (Symbol)	Value	
kilobyte (kB)	10^3	2^{10}	kibibyte (KiB)	2^{10}	
megabyte (MB)	10^6	2^{20}	mebibyte (MiB)	2^{20}	
gigabyte (GB)	10^9	2^{30}	gibibyte (GiB)	2^{30}	
terabyte (TB)	10^{12}	2^{40}	tebibyte (TiB)	2^{40}	
petabyte (PB)	10^{15}	2^{50}	pebibyte (PiB)	2^{50}	
exabyte (EB)	10^{18}	2^{60}	exbibyte (EiB)	2^{60}	
zettabyte (ZB)	10^{21}	2^{70}	zebibyte (ZiB)	2^{70}	
yottabyte (YB)	10^{24}	2^{80}	yobibyte (YiB)	2^{80}	
See also: Multiples of bits · Orders of magnitude of data					

http://en.wikipedia.org/wiki/Template:Quantities_of_bits





Visual Steering to Support Decision Making in Visdom

Jürgen Waser

http://www.cg.tuwien.ac.at/research/publications/2011/waser_2011_VSD/

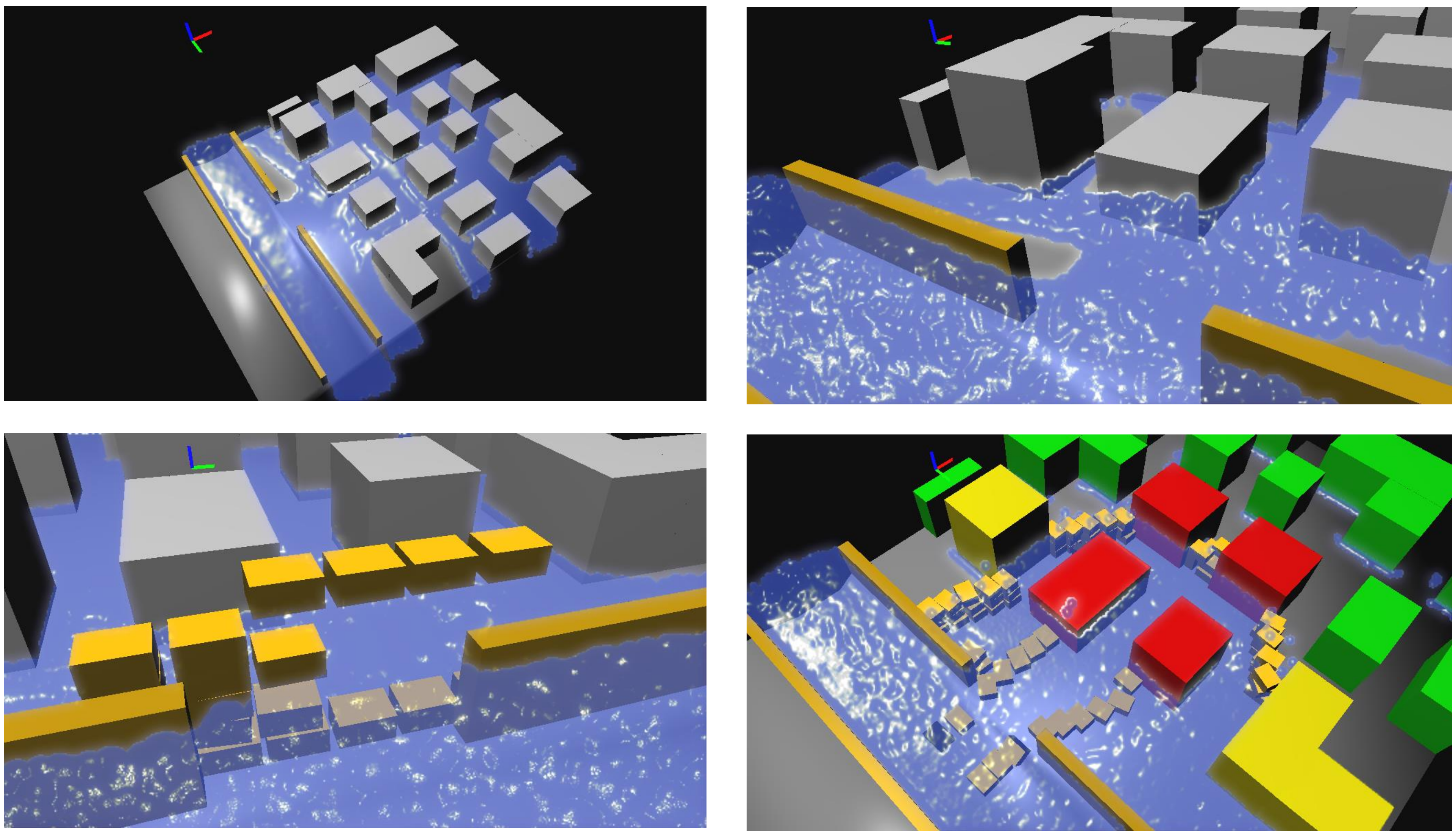
Flood emergency assistance

- **New Orleans 2005: 17th canal levee breach**

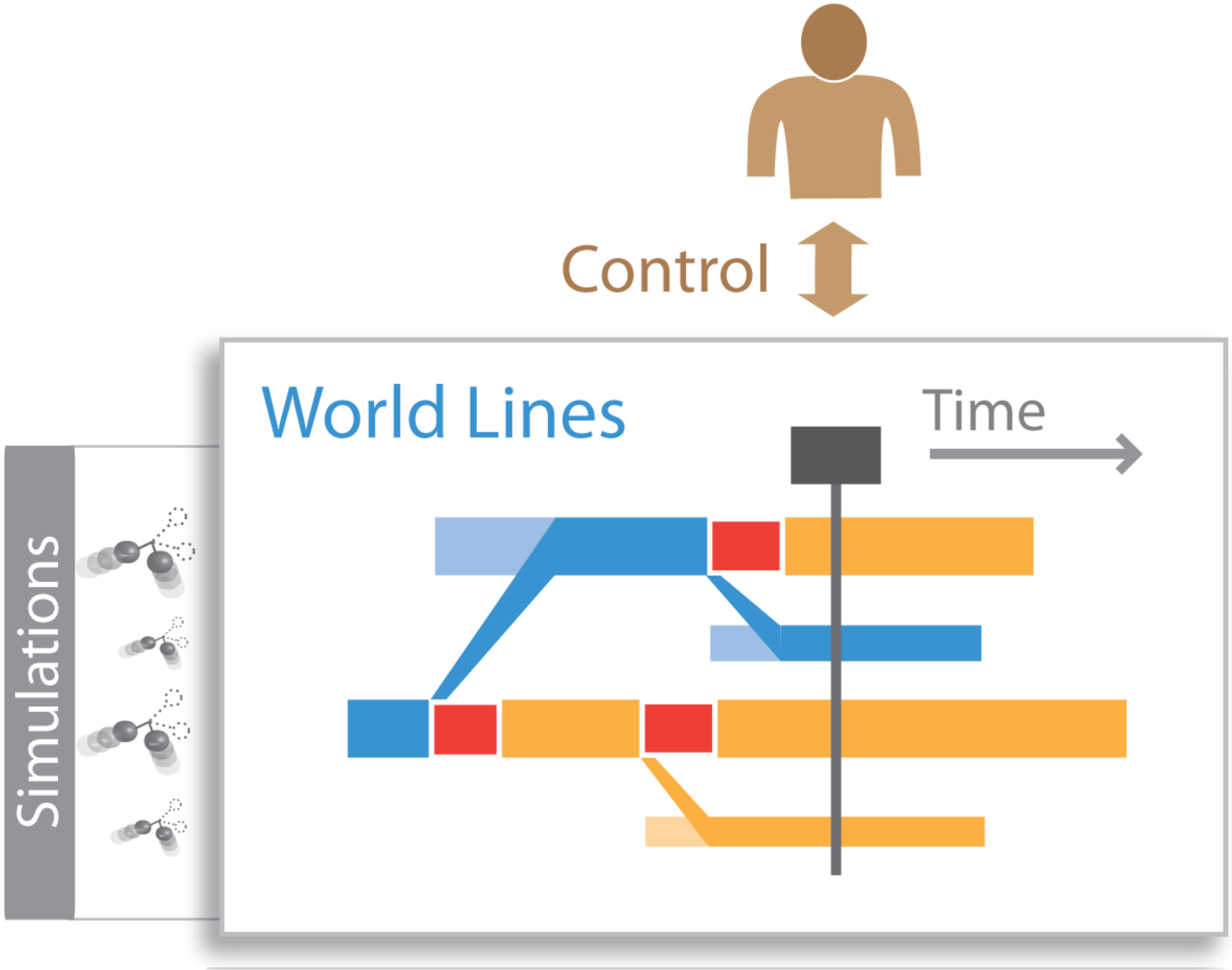
Image courtesy of USACE, US Army Corps of Engineers

Flood emergency assistance

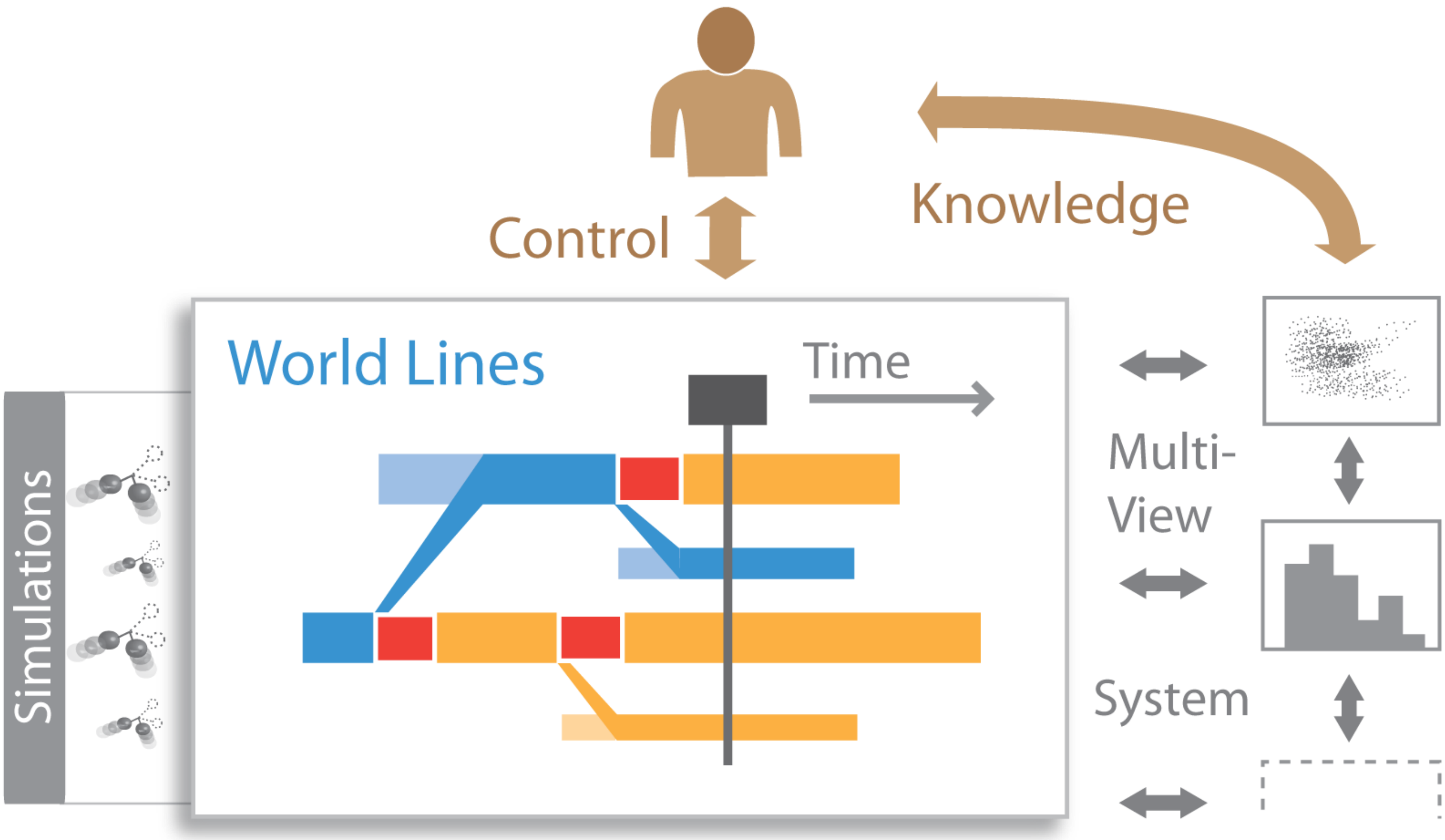
- **Testing sandbag configurations in a virtual environment**



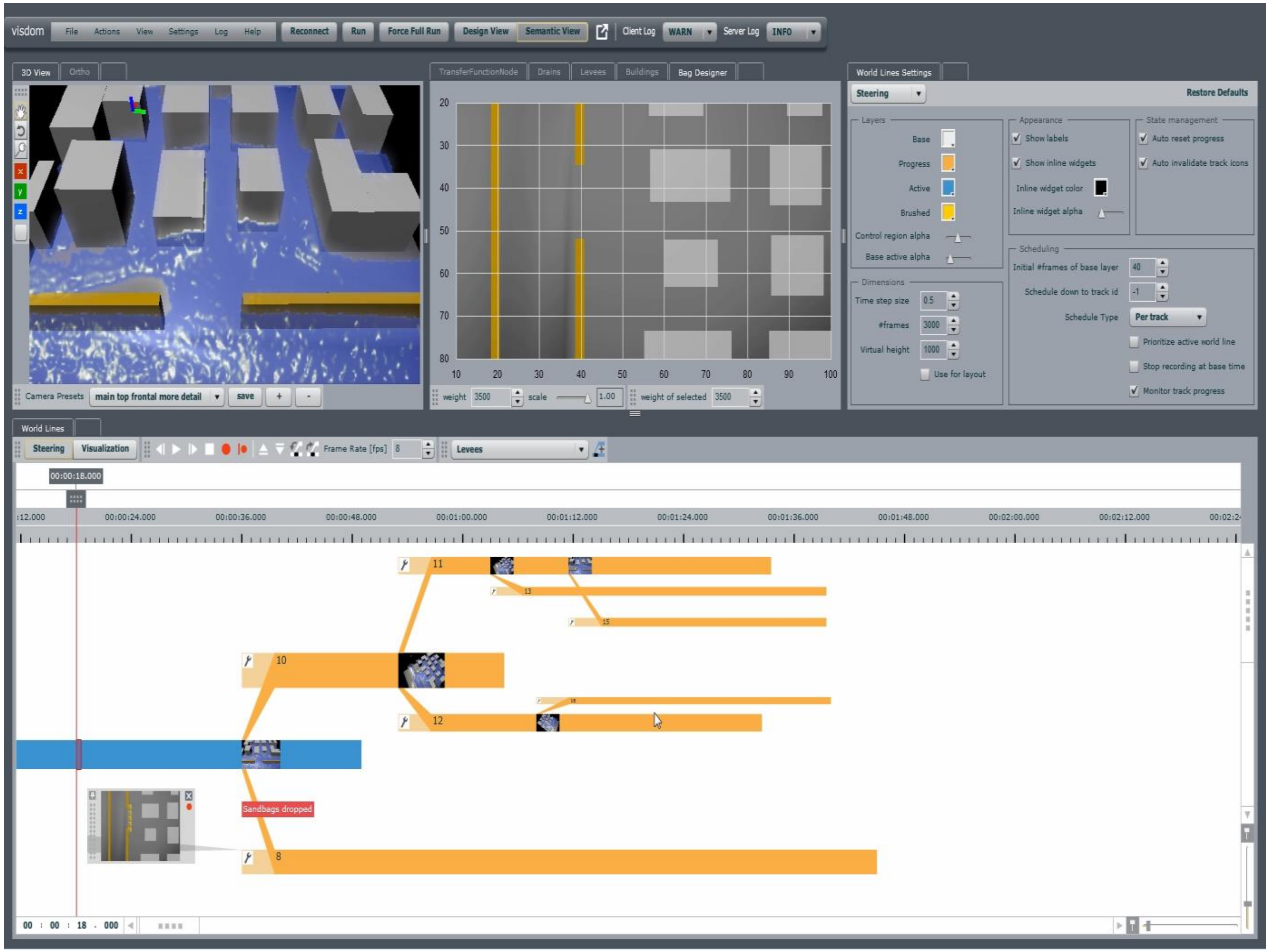
Solution: World Lines



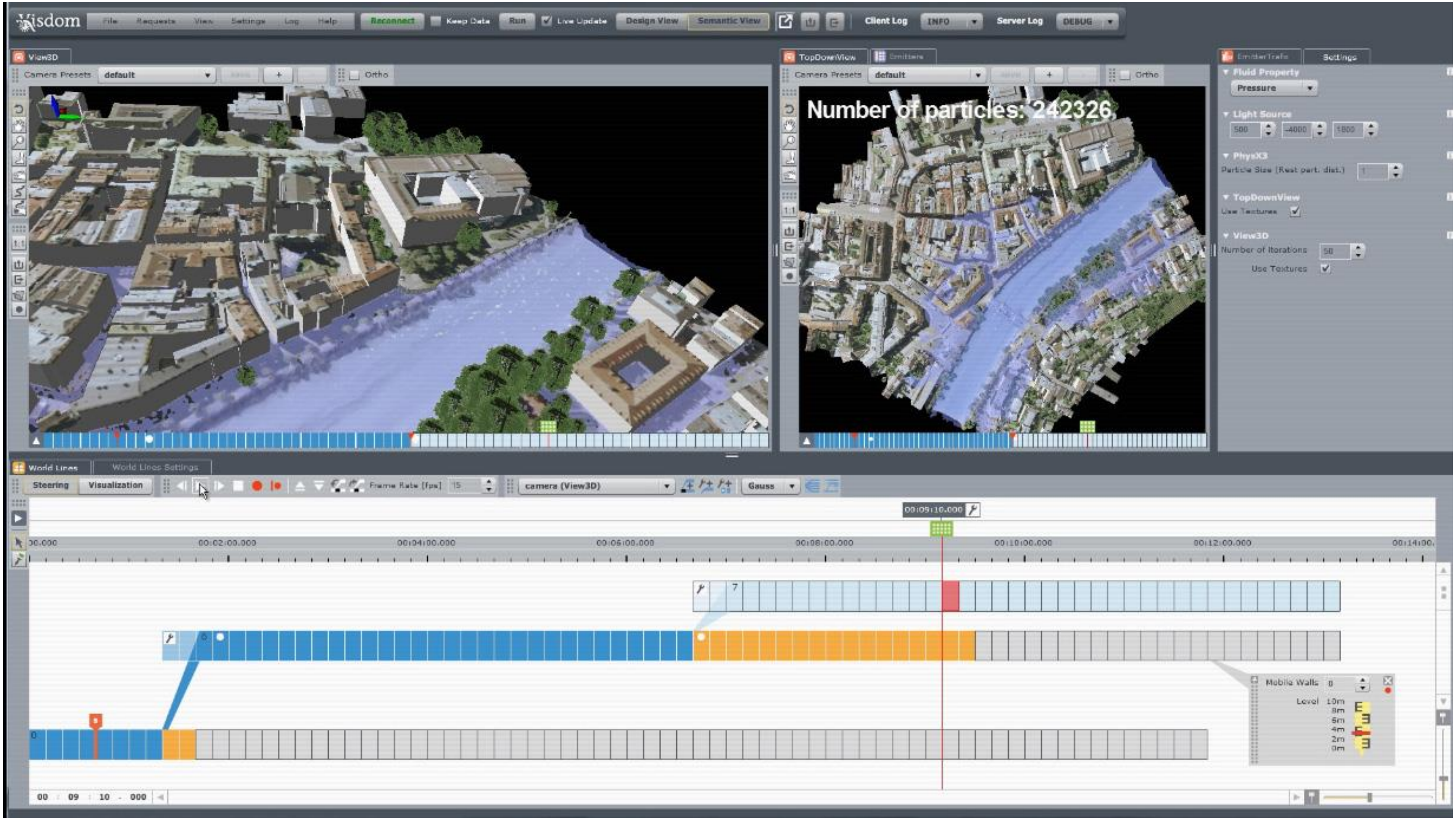
Solution: World Lines



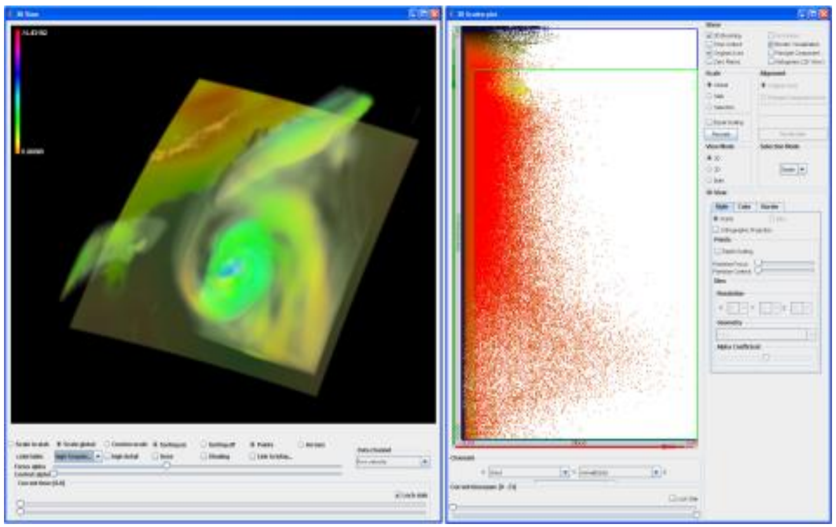
Video

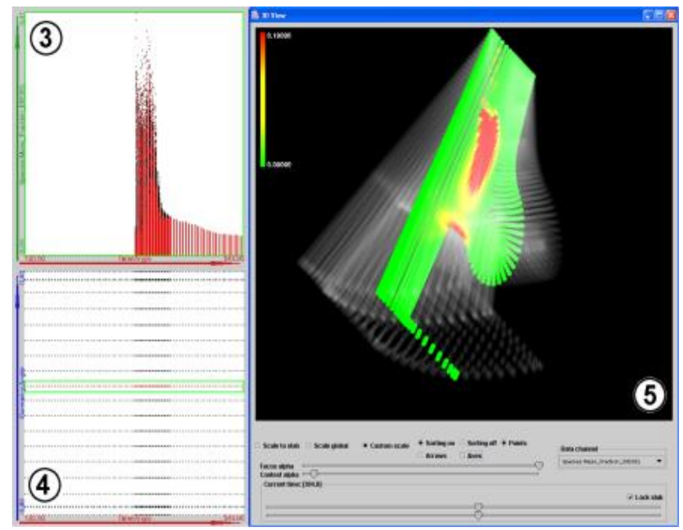

visdom
File
Actions
View
Settings
Log
Help
Reconnect
Run
Force Full Run
Design View
Semantic View
Client Log
WARN
Server Log
INFO
3D View
Ortho
TransferFunctionNode
Drains
Levees
Buildings
Bag Designer
World Lines Settings
Steering
Restore Defaults
Layers
Base
Progress
Active
Brushed
Control region alpha
Base active alpha
Appearance
Show labels
Show inline widgets
Inline widget color
Inline widget alpha
State management
Auto reset progress
Auto invalidate track icons
Scheduling
Initial #frames of base layer
Schedule down to track id
Schedule Type
Per track
Prioritize active world line
Stop recording at base time
Monitor track progress
Dimensions
Time step size
#frames
Virtual height
Use for layout
Camera Presets
main top frontal more detail
save
weight
scale
weight of selected
World Lines
Visualization
Frame Rate [fps]
Levees
00:00:18.000
Sandbags dropped

Worldlines – Multiple Linked Views



SimVis: Interactive Visual Analysis of Large & Complex Simulation Data



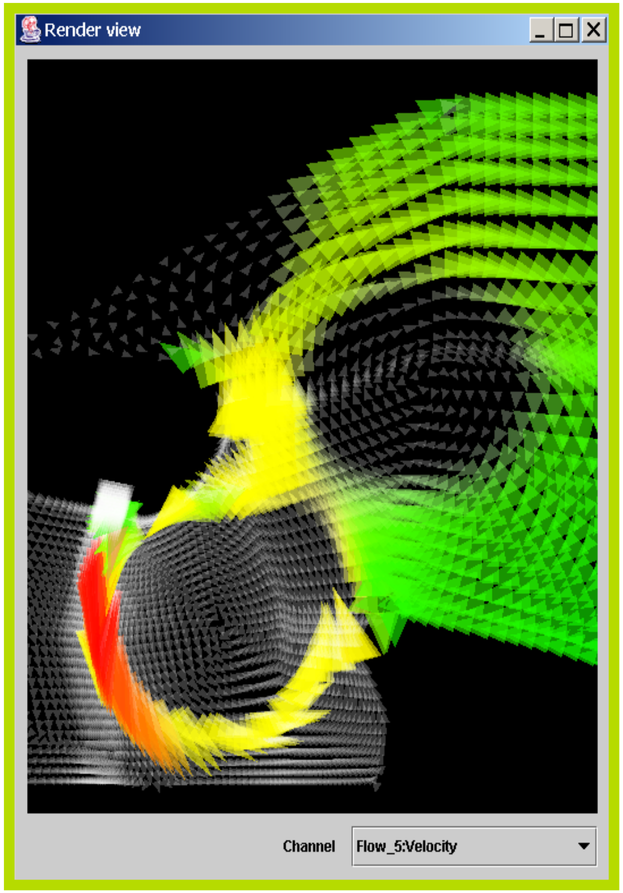


Dr. Helmut Doleisch

VRVis Research Center

Motivation

- large data sets from simulation
- **goal**: support **exploration** and **analysis** of results
  - analyze n-dim. data interactively
  - use 3D visualization
  - overview, zoom and filter, detail on demand (Shneidermans' information seeking mantra)
- challenge:
  - occlusion
  - interactive data handling



Interactive Data Handling

- sample data set size:
  - 540 million data items
  - currently working to expand to billions

cells	timesteps	attributes	cells * timesteps	cells * timesteps * attributes
704.900	20	16	14.098.000	225.568.000
150.124	600	6	90.074.400	540.446.400
7.680.000	288	15	2.211.840.000	33.177.600.000

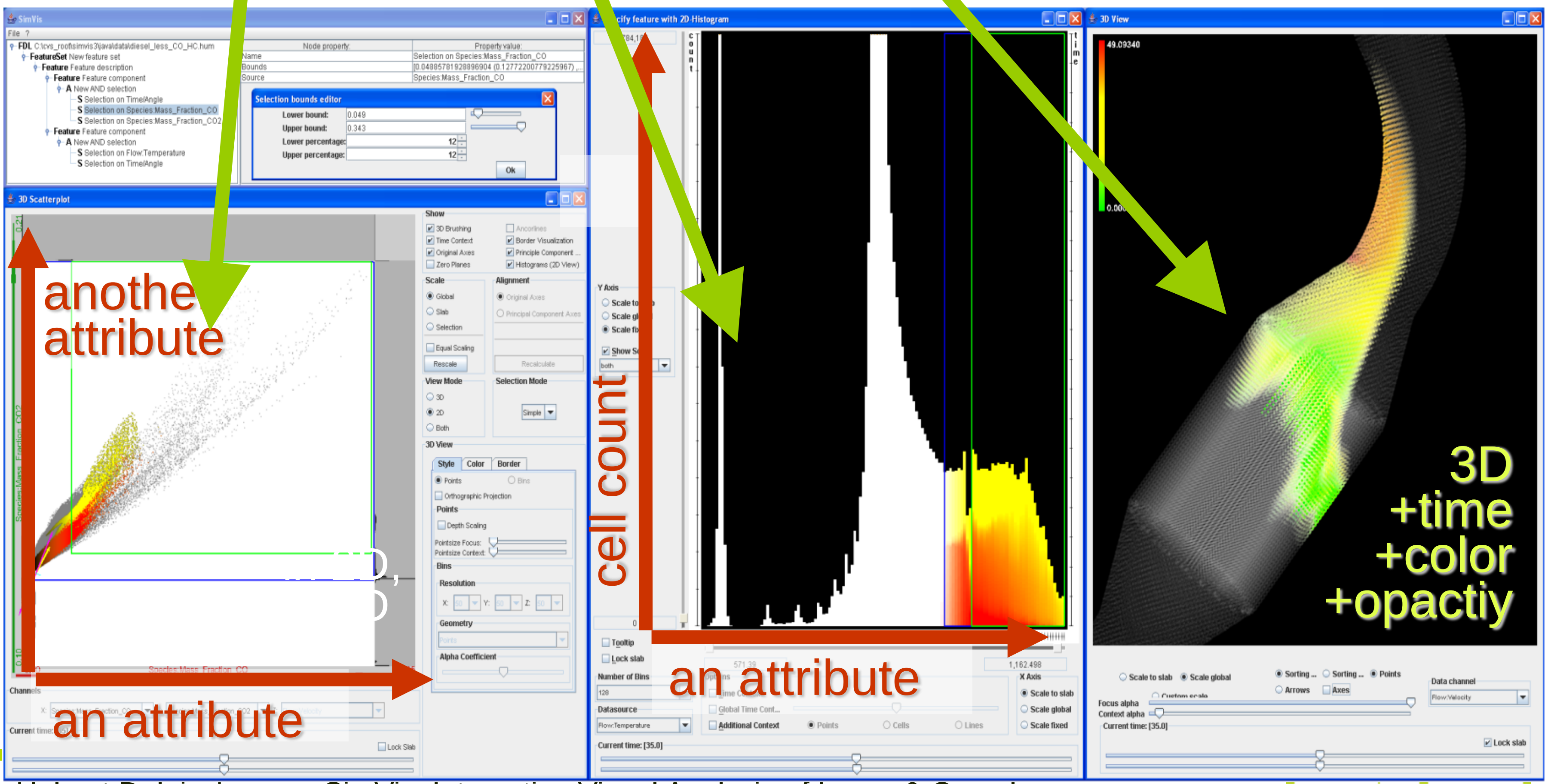
SimVis

- VRVis´ solution for these challenges
- Feature-based visualization framework

- SimVis key features:
  - Multiple, linked views
  - Interactive feature specification
  - Focus+Context visualization
  - Smooth feature boundaries
  - Explicit feature representation
  - On-the-fly attribute derivation

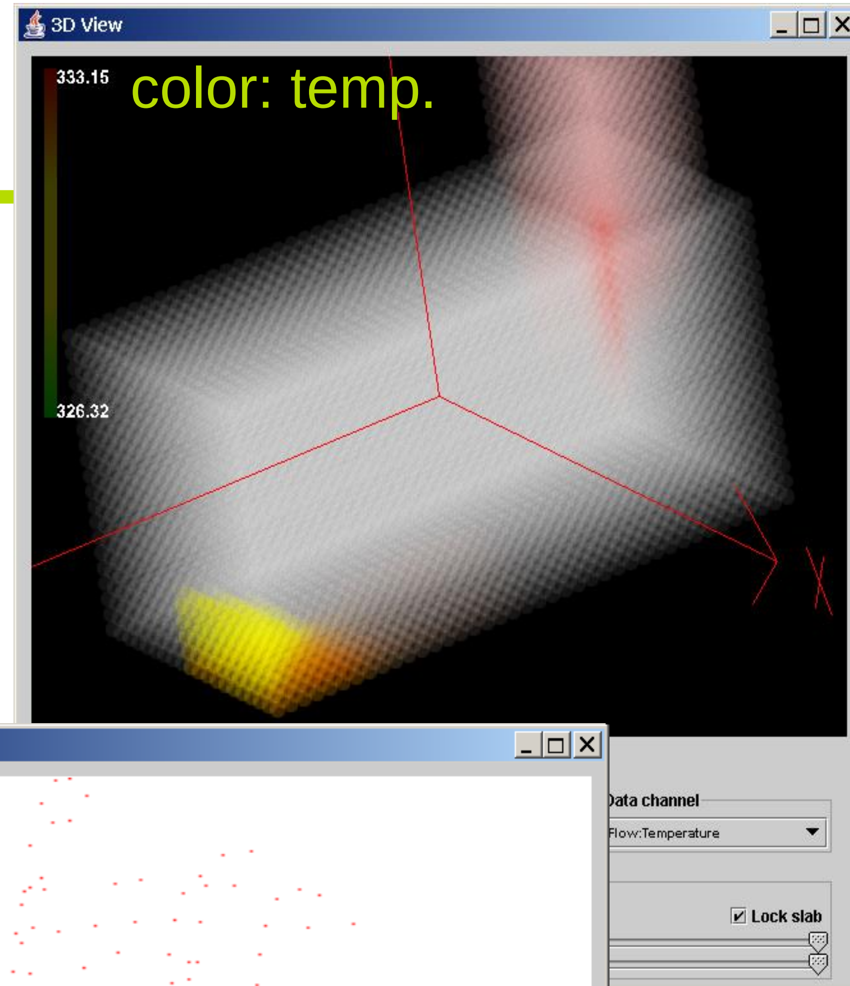
SimVis: Multiple Views

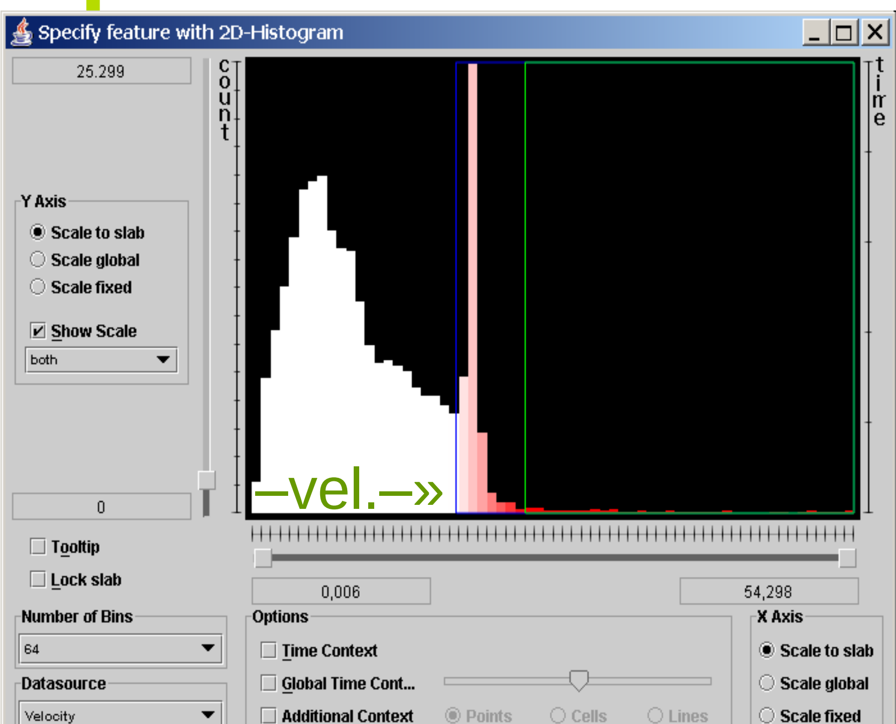
- Scatterplots, histogram, 3D(4D) view, etc.

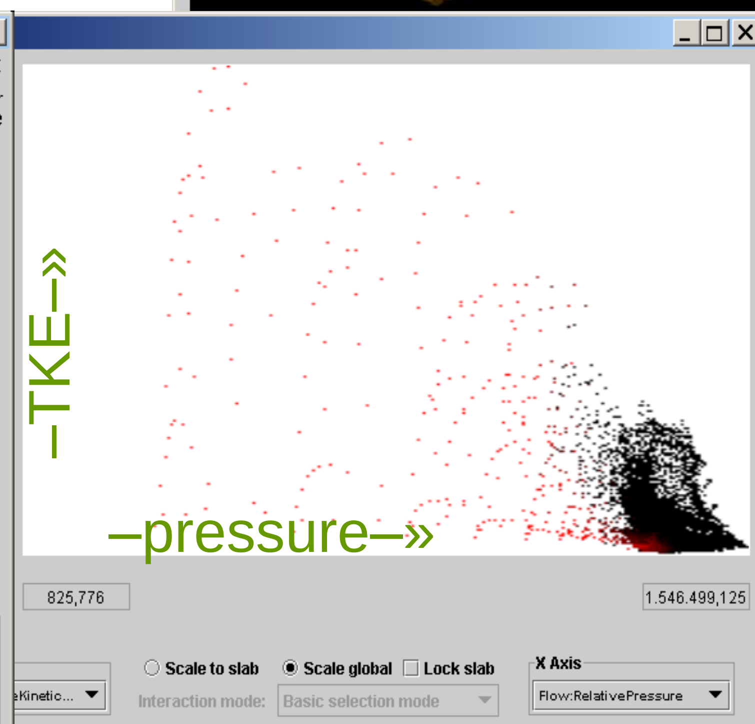


Brushing

- Move/alter/extend brush interactively
- Update linked F+C views in real-time





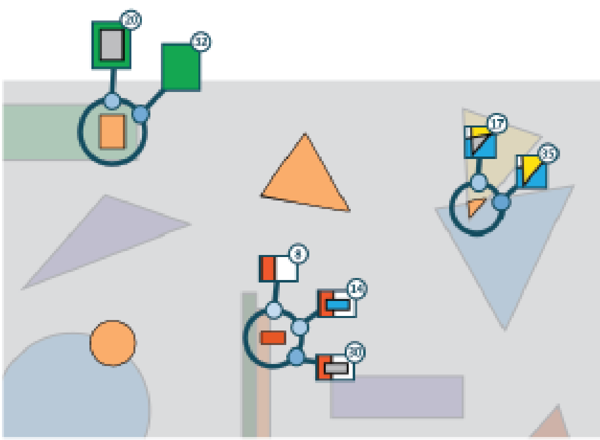


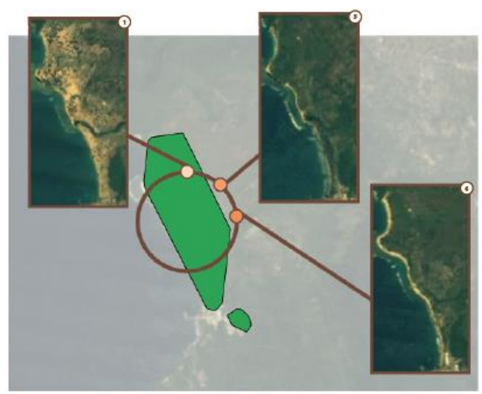
VAICo: Visual Analysis for Image Comparison

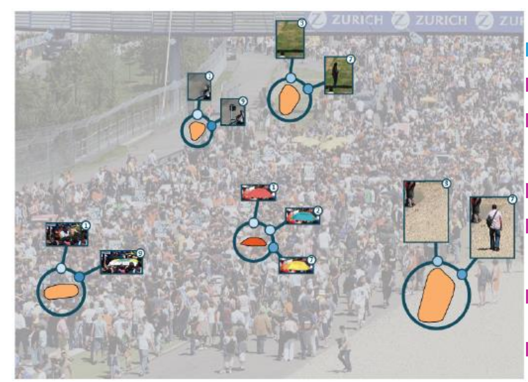
Johanna Schmidt[1], M. Eduard Gröller[1], Stefan Bruckner[2]

[1]Vienna University of Technology, Austria

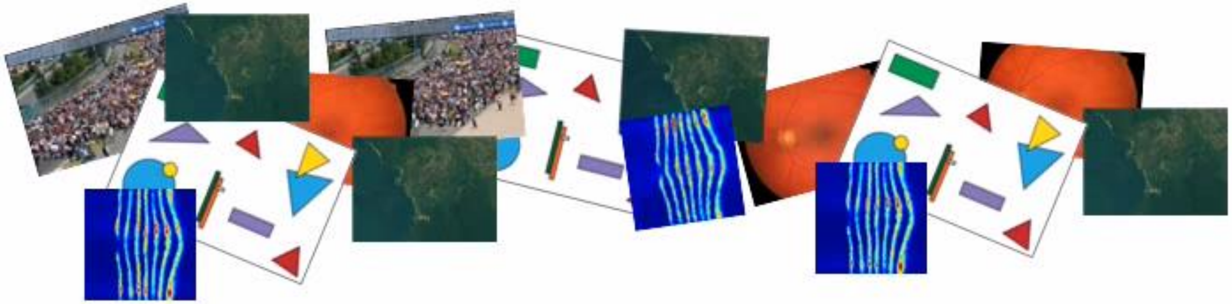
[2]University of Bergen, Norway







VAICo – Example Video



VAICo

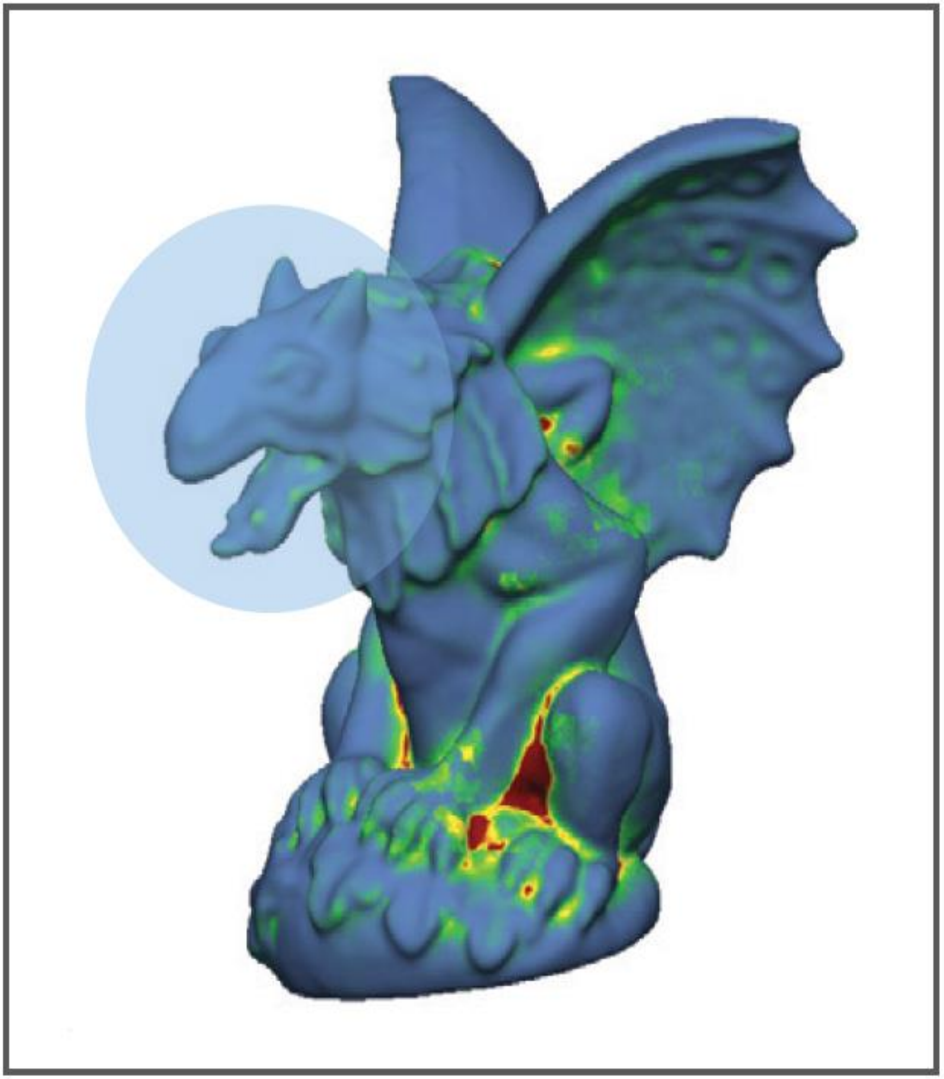
Visual Analysis for Image Comparison

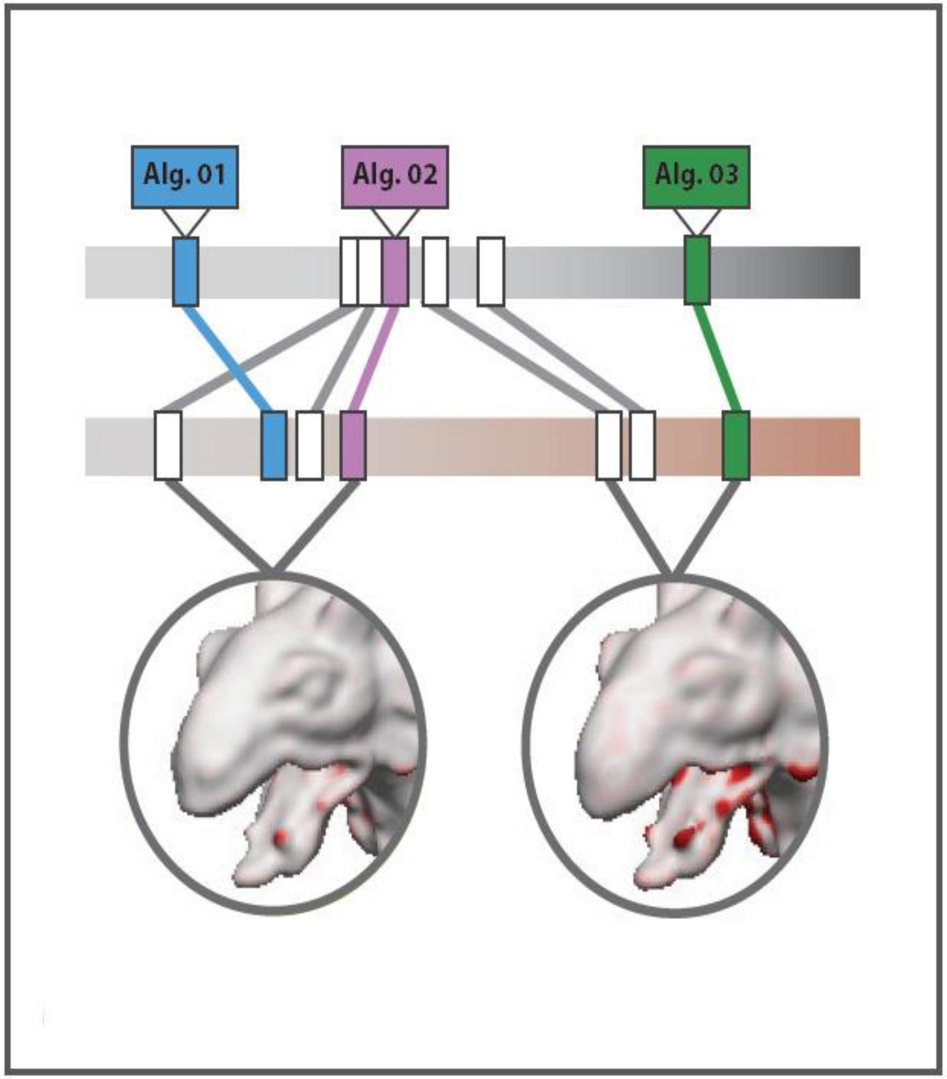
Johanna Schmidt
Vienna University of Technology, Austria

M. Eduard Gröller
Vienna University of Technology, Austria

Stefan Bruckner
University of Bergen Norway

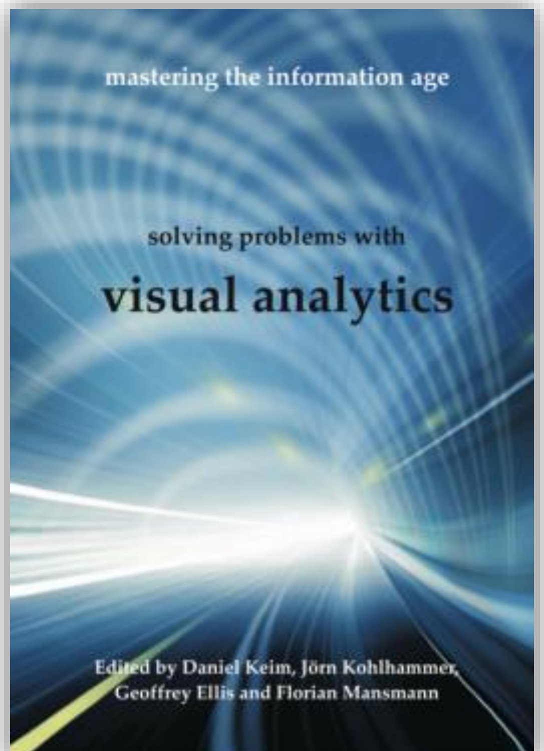
YMCA - Your Mesh Comparison Application

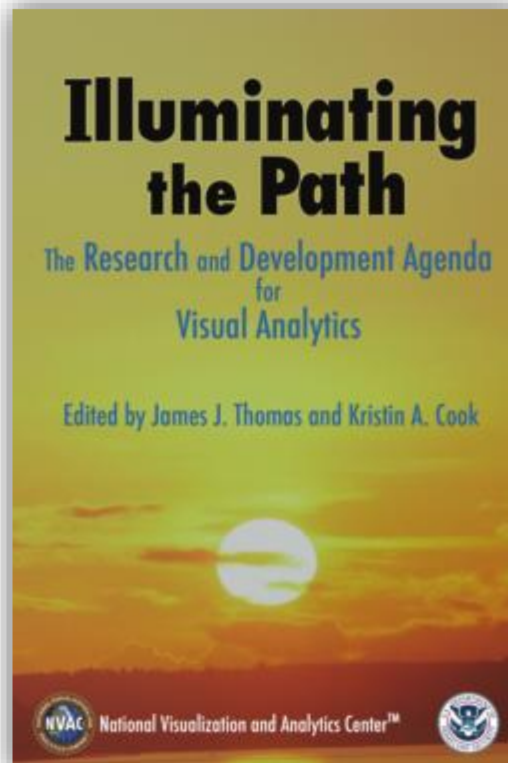




[Johanna Schmidt et al.]

Literature on Visual Analytics

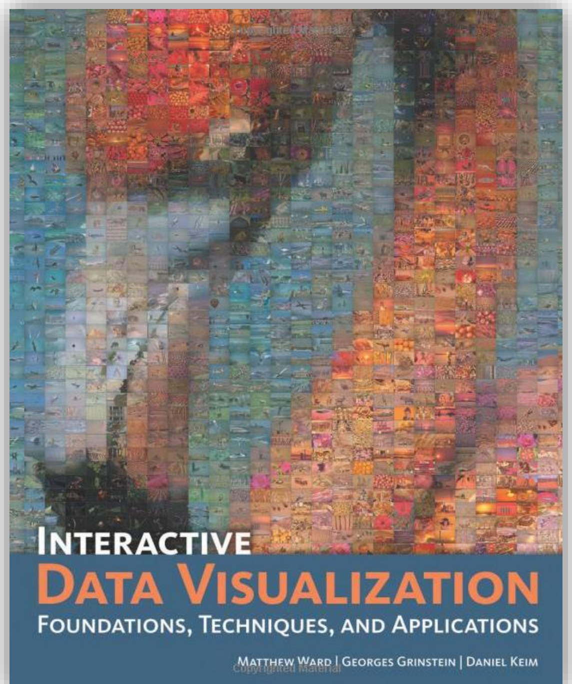




Daniel A. Keim, Jörn Kohlhammer, Geoffrey Ellis and Florian Mansmann: *Mastering the Information Age - Solving Problems with Visual Analytics,* Eurographics Association, 2010.
ISBN: 978-3905673777.
Free download: http://www.vismaster.eu/book/

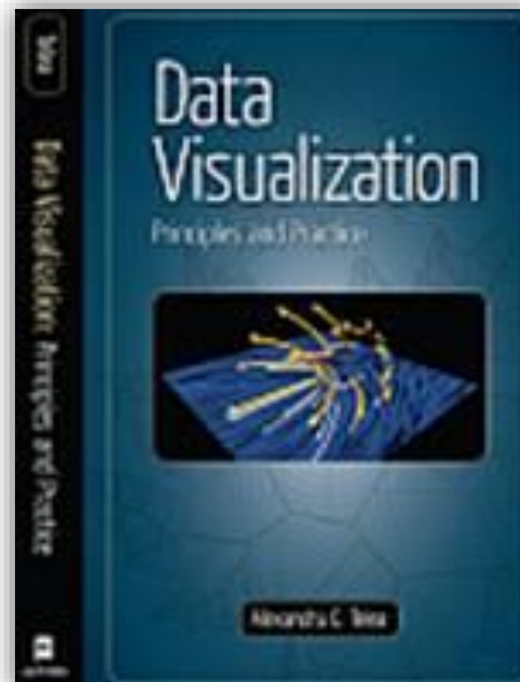
James J. Thomas and Kristin A. Cook : *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, National Visualization and Analytics Ctr, 2005.
ISBN: 978-0769523231
Free download: http://vis.pnnl.gov/

Literature on Visualization

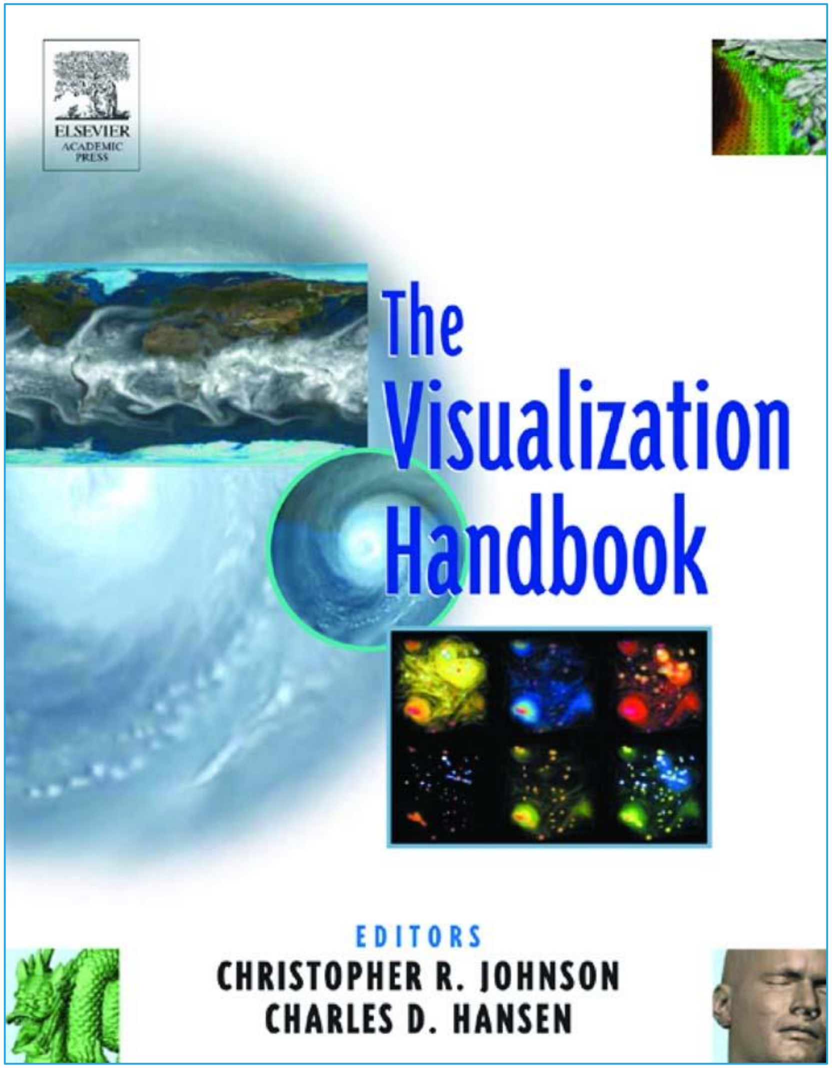


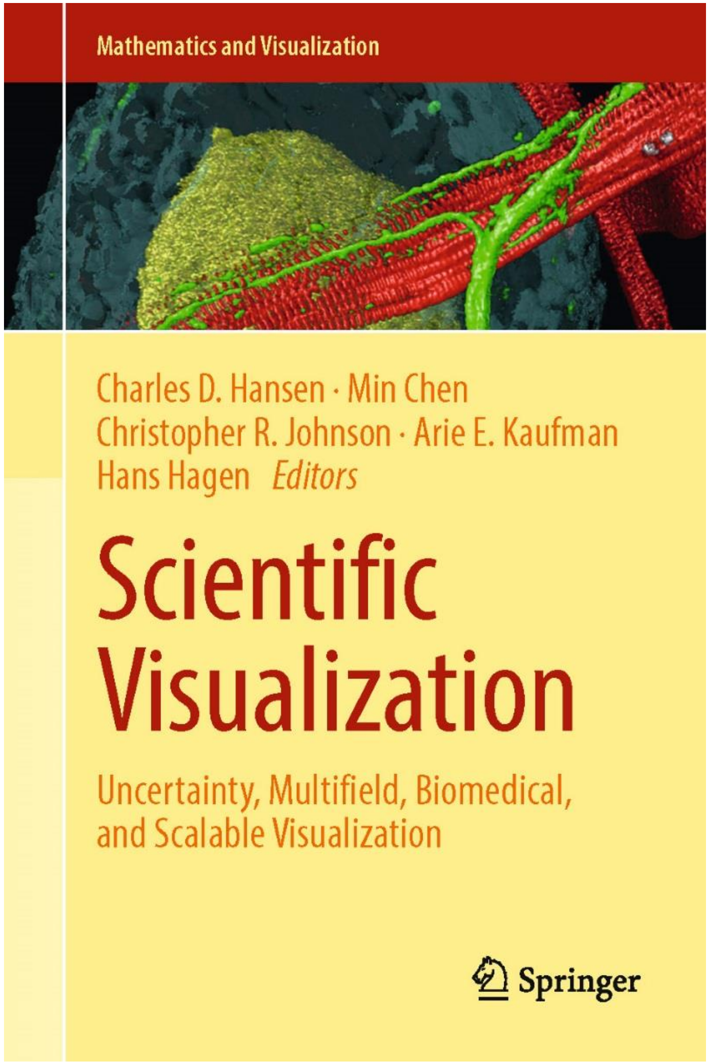
Matthew Ward, George Grinstein, Daniel Keim: *Interactive Data Visualization: Foundations, Techniques, and Applications,*
AK Peters Verlag, 2010.
ISBN: 1568814739

Heidrun Schumann, Wolfgang Müller: *Visualisierung - Grundlagen und allgemeine Methoden,*
Springer Verlag, 2000.
ISBN: 3540649441

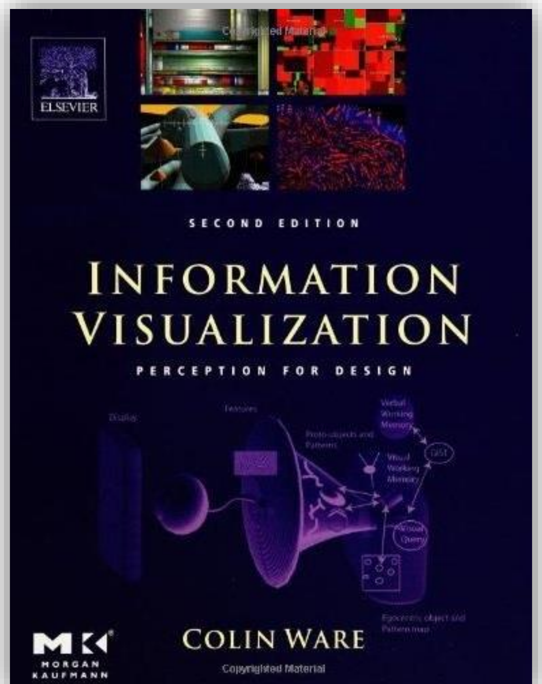


Alexandru Telea: Data *Visualizaton – Principles and Practice,*
AK Peters Verlag, 2008.
ISBN: 9781568813066

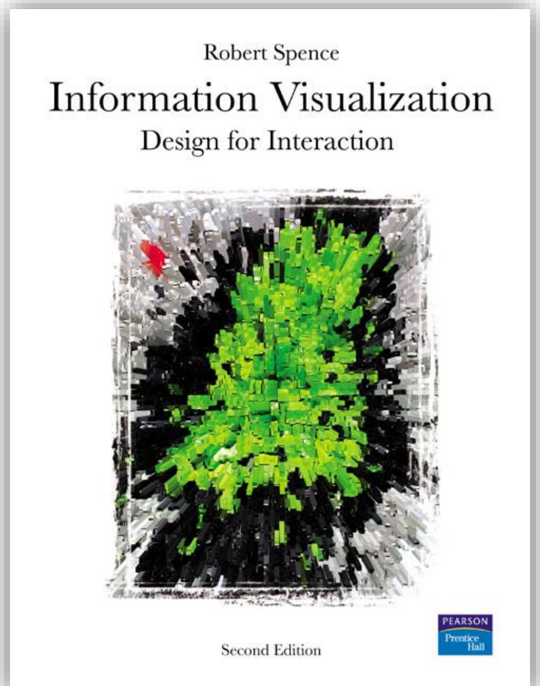

ELSEVIER
ACADEMIC PRESS
The Visualization Handbook
EDITORS
CHRISTOPHER R. JOHNSON
CHARLES D. HANSEN


Mathematics and Visualization
Charles D. Hansen · Min Chen
Christopher R. Johnson · Arie E. Kaufman
Hans Hagen Editors
Scientific Visualization
Uncertainty, Multifield, Biomedical, and Scalable Visualization
Springer

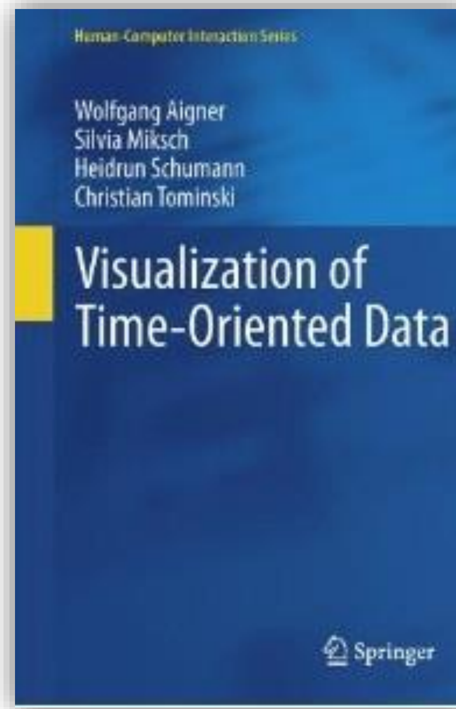
Literature on Information Visualization



Colin Ware: *Information Visualization, Second Edition: Perception for Design*, Morgan Kaufmann, 2nd edition, 2004.
ISBN: 1558608192



Robert Spence: *Information Visualization - Design for Interaction*,
Pearson Verlag, 2001.
ISBN13: 9780132065504



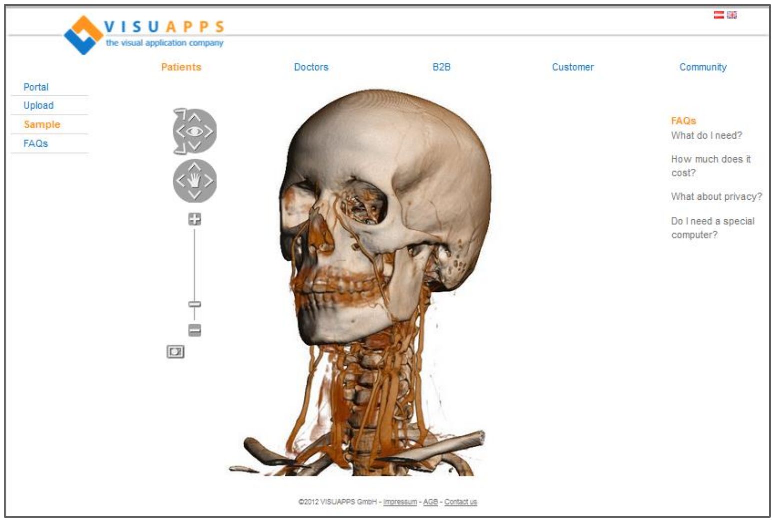
Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, Christian Tominski: *Visualization of Time-Oriented Data*,
Springer Verlag, 2011.
ISBN13: 978-0857290786

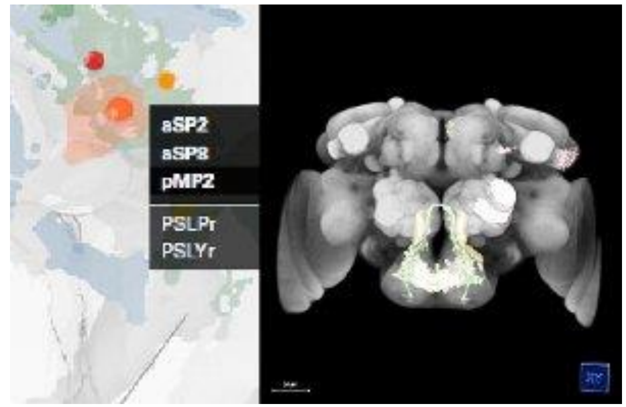
Acknowledgements

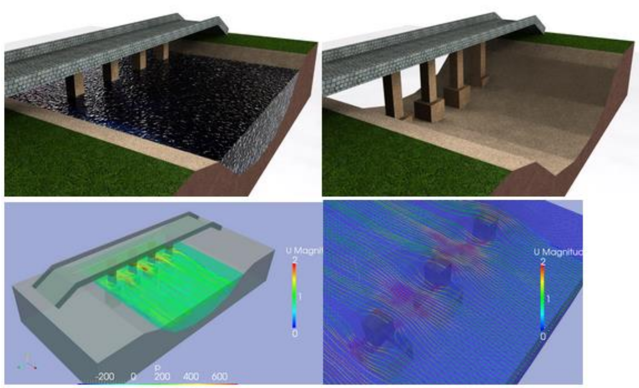
- For material for this lecture unit
  - Marc Streit, Johannes Kepler University Linz

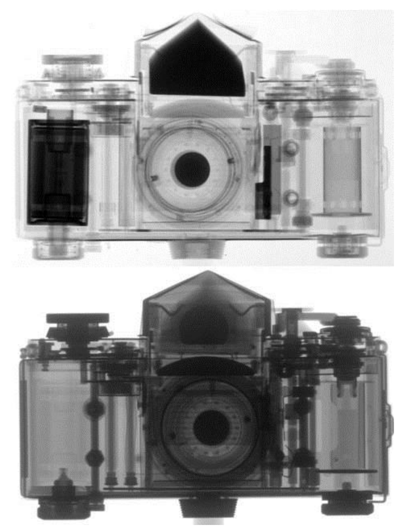
Praktika, Bachelorarbeiten, Diplomarbeiten

- http://www.cg.tuwien.ac.at/courses/projekte/









Christmas Tree Case Study