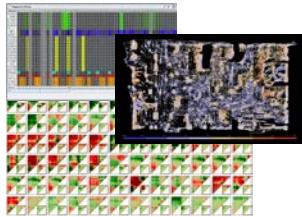


# Exploring Big Data using Visual Analytics



**Daniel A. Keim**  
Data Analysis and Information  
Visualization Group  
University of Konstanz, Germany



**Data Mining for Business Intelligence, Beer Sheva, Israel**  
April 24, 2014

## Automated Analysis

### Good News

**Automated Analysis of big data works!  
(under certain preconditions)**

## Automated Analysis

### Bad News

**The preconditions are rarely met!**

**Preconditions:**

- Data is clearly structured
- Data semantics is well-defined
- Data is complete, correct, and not changing over time

**AND**

- Problem is well-defined

## Automated Analysis

- **Network Security**
  - Data: IP Flows
  - Task: Detection of Novel Viruses
- **Fraud Detection**
  - Data: Credit Card (or Phone Call) Data
  - Task: Detection of Fraud
- **Business Analytics**
  - Data: Customer Records
  - Task: Define Customer Target Groups
- **Molecular Biology**
  - Data: Patient DNA Records
  - Task: Functional Root Cause Analysis for an Illness

## The Role of Visualization

### Visualization is needed in addition to analytics to

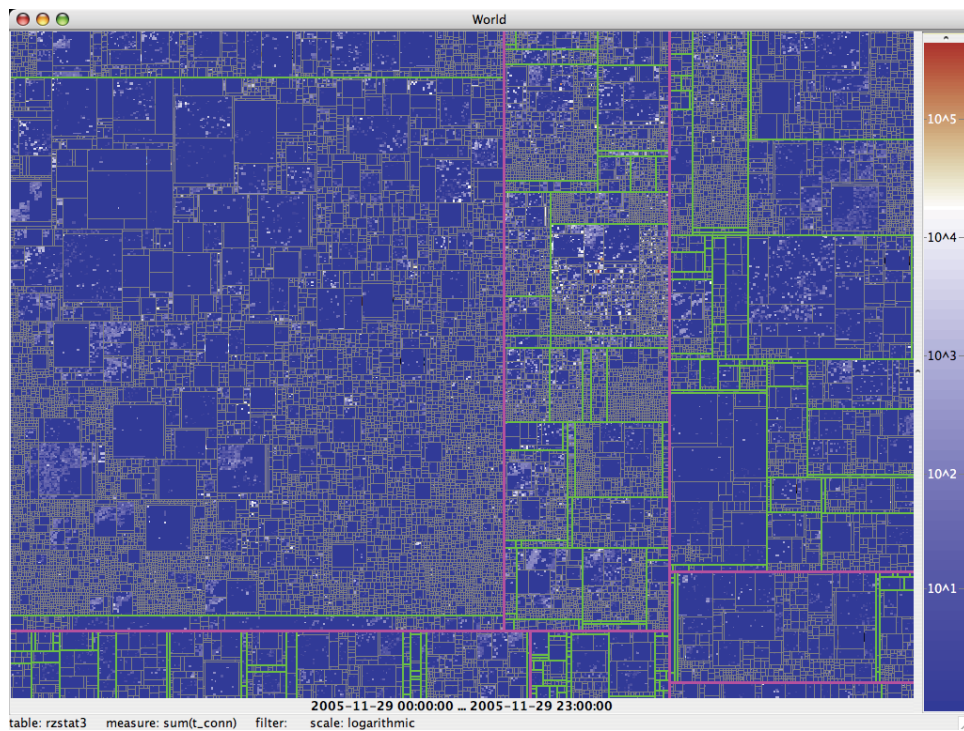
- identify the structure (based on user knowledge)
- bridge the semantic gap (bring in user knowledge)
- help with incomplete or incorrect and changing data
- understand the problem, generate hypotheses and define the problem
- and steer the analysis process in dealing with massive data (local optimization)

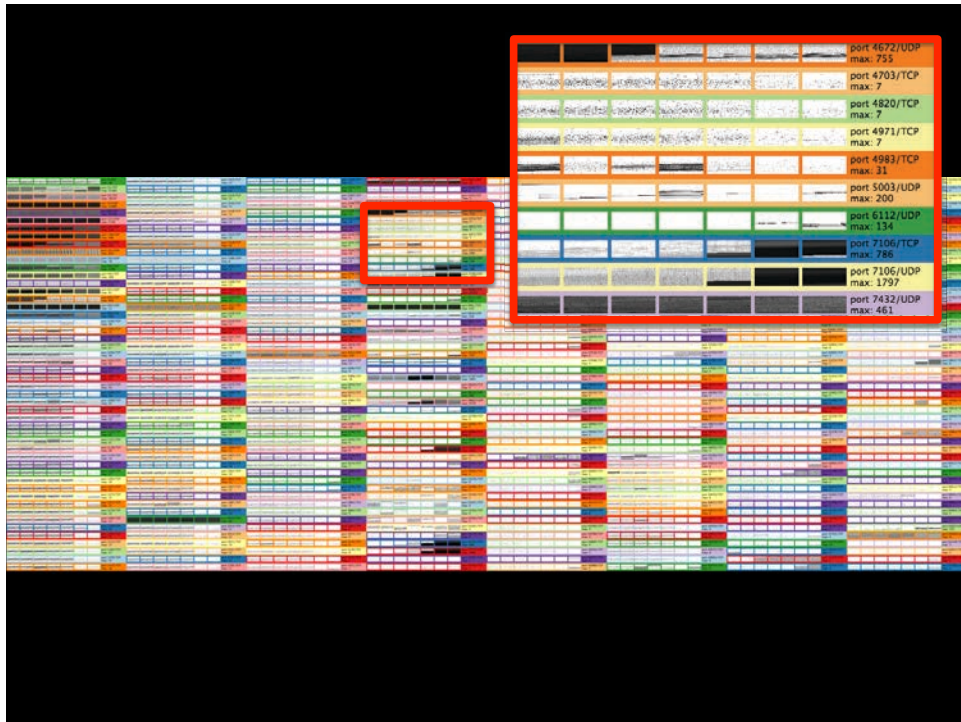
# Visualization

## Bad News

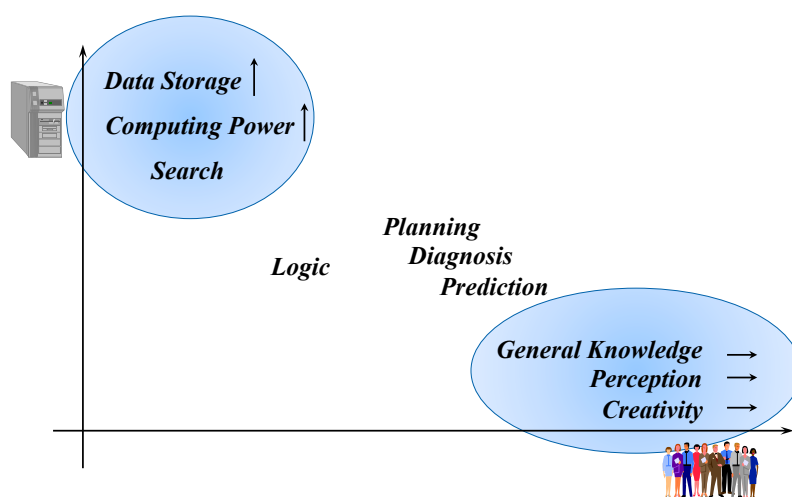
**Visualization of big data  
does not work!**

Universität  
Konstanz





## Visual Analytics



## Visual Analytics

***"Computers are incredibly fast,  
accurate, and stupid; humans are  
incredibly slow, inaccurate, and  
brilliant; together they are powerful  
beyond imagination."***

attributed to Albert Einstein

Roadmap from the  
VisMaster EU Project

[www.visual-analytics.eu](http://www.visual-analytics.eu)

mastering the information age

solving problems with  
**visual analytics**

Keim, Kohlhammer, Ellis, Mansmann (eds.) visual analytics

Edited by Daniel Keim, Jörn Kohlhammer,  
Geoffrey Ellis, and Florian Mansmann

Video

# Outline

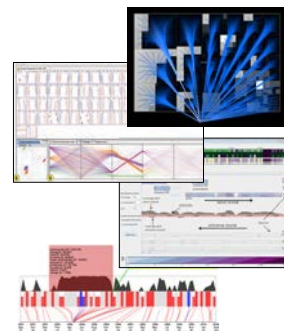
- The Role of Visualization and Analytics in Exploring Big Data

- Why automated analysis does not work
- When visual exploration can help



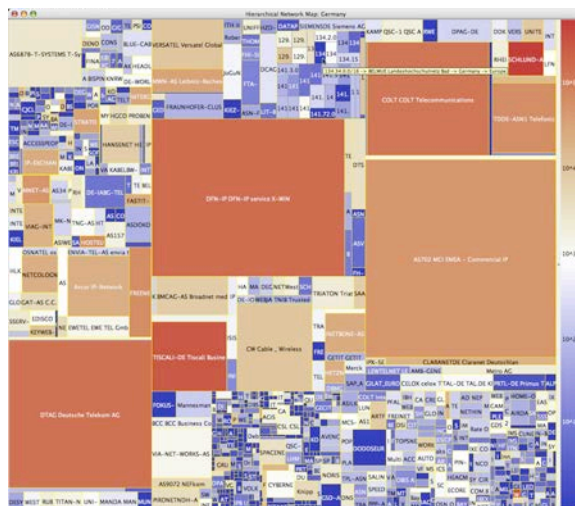
- Visual Analytics Examples

- Network Security
- Document Analysis
- Financial Analysis
- Molecular Biology



- Visual Analytics Perspectives

## Visual Network Analysis

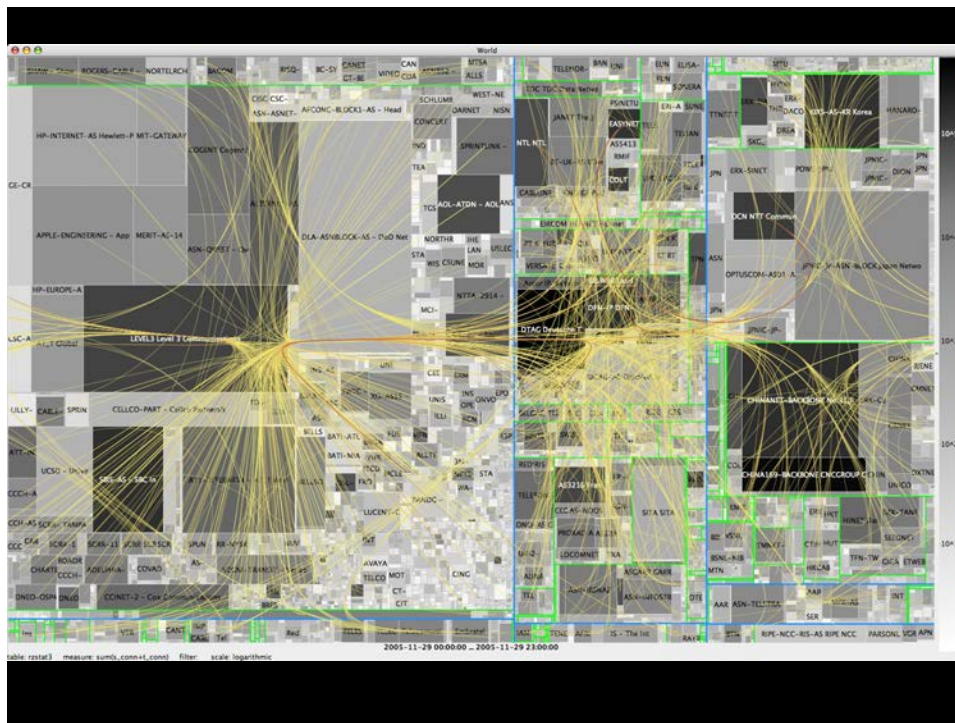


### Hierarchy:

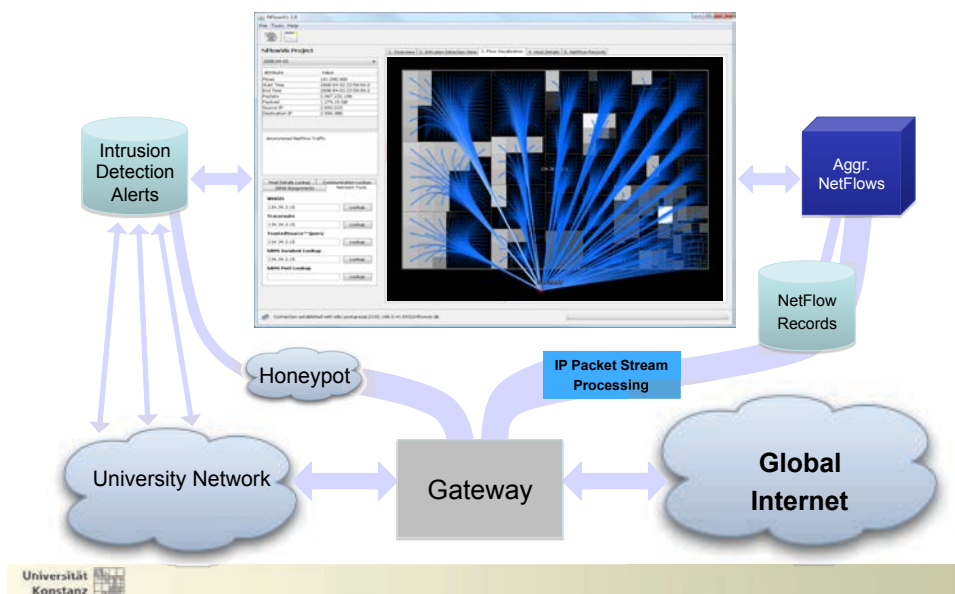
- Continents
- Countries
- Autonomous Systems
- Networks

data: rzstat3 date: 29 Nov 2005 measure: outgoing connections



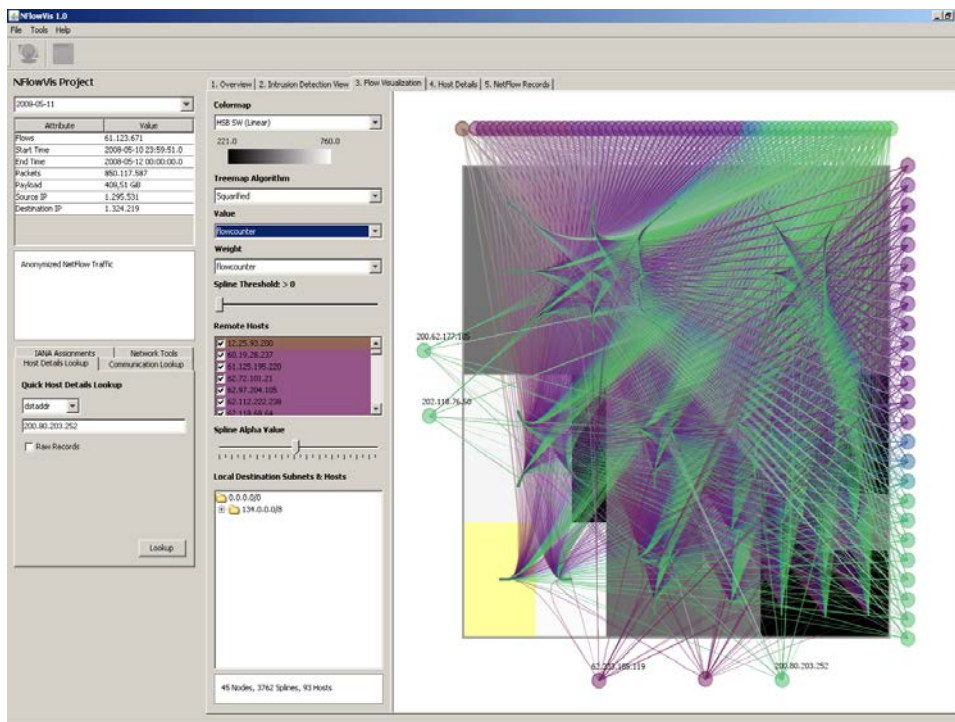
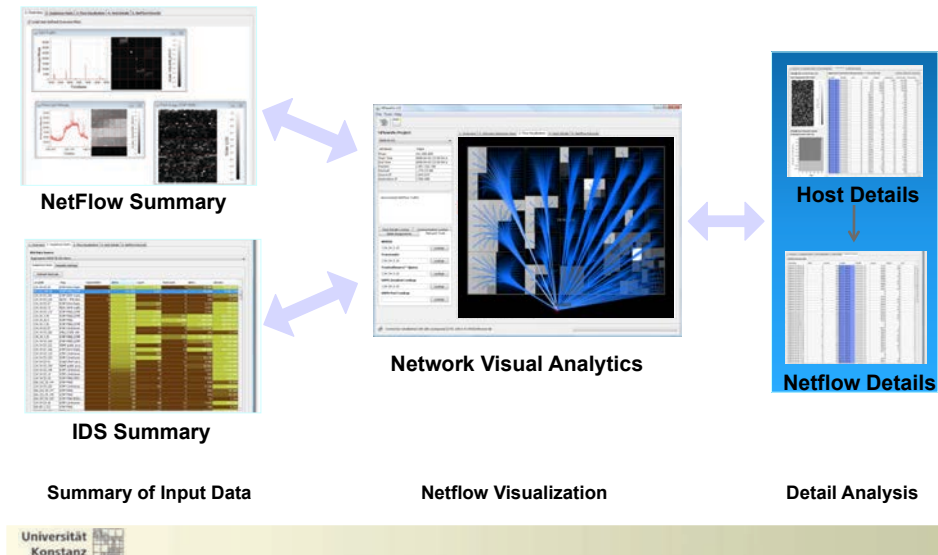


## Visual Analytics of Network Data





# Visual Network Analysis



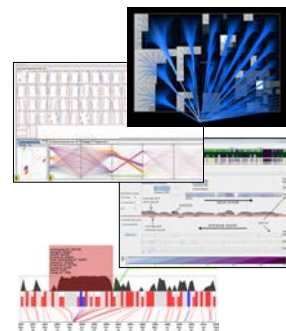


## Uni Konstanz Control Room

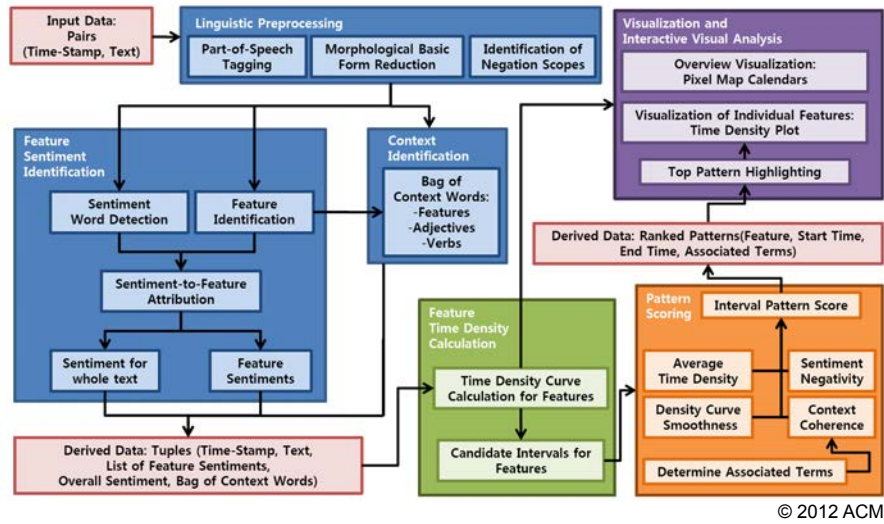


## Outline

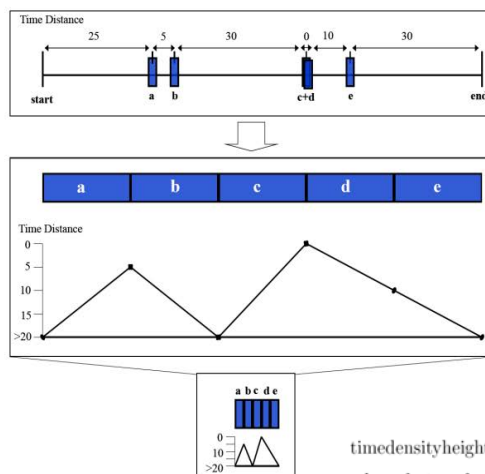
- The Role of Visualization and Analytics in Exploring Big Data
  - Why automated analysis does not work
  - When visual exploration can help
- Visual Analytics Examples
  - Network Security
  - **Document Analysis**
  - Financial Analysis
  - Molecular Biology
- Visual Analytics Perspectives



# Customer Feedback Analysis



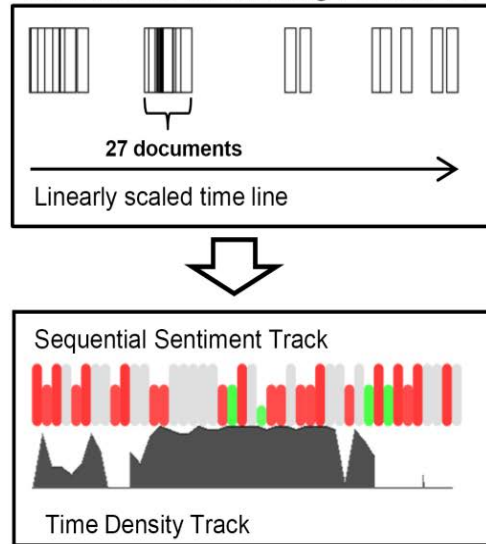
# Customer Feedback Analysis



$$\text{timedensityheight}(f,a,b) = \max \left( 0, \left( 1 - \frac{\text{timedist}(a,b)}{\text{avgtimedist}(f)} \right) \right) (1)$$

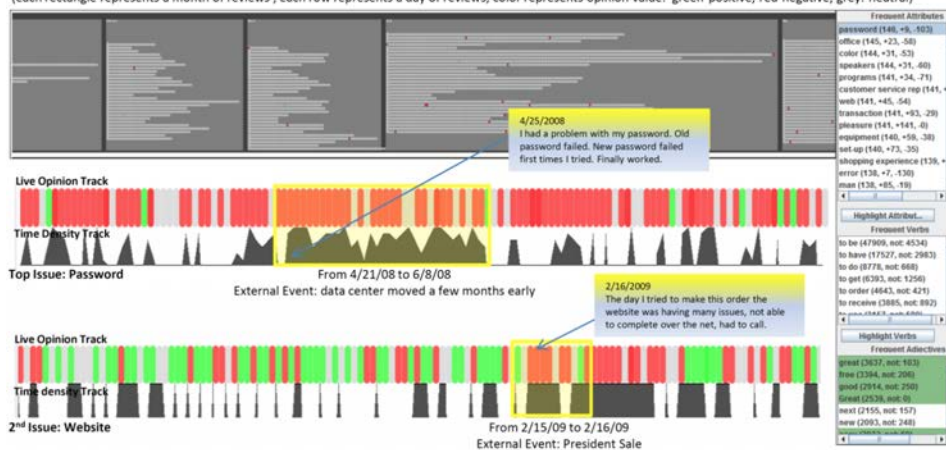
where feature  $f$ , preceding document  $a$ , succeeding document  $b$ .

## Customer Feedback Analysis



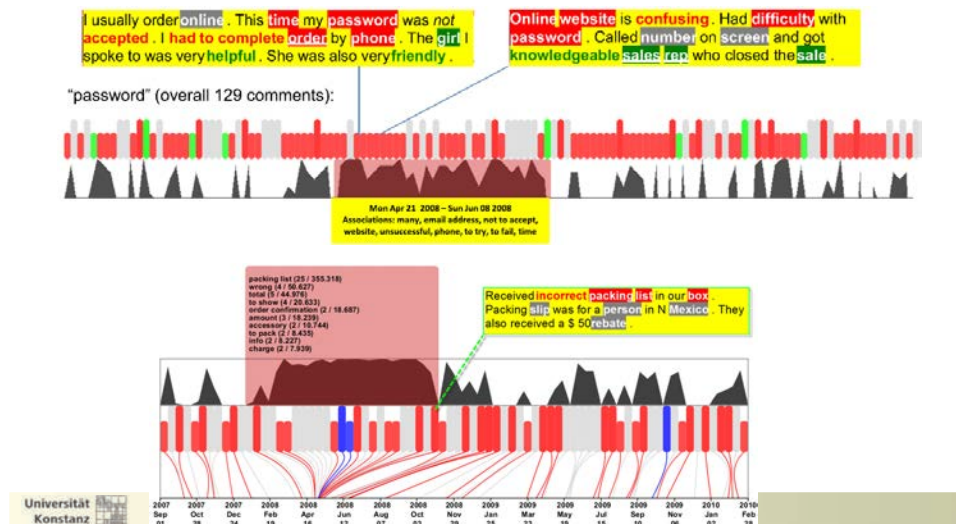
## Customer Feedback Analysis

(each rectangle represents a month of reviews, each row represents a day of reviews; color represents opinion value: green-positive, red-negative, grey-neutral)



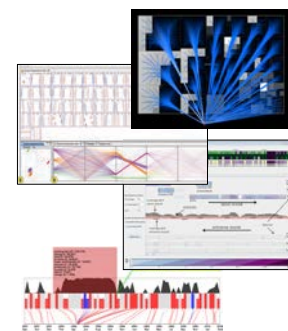


# Customer Feedback Analysis



## Outline

- The Role of Visualization and Analytics in Exploring Big Data
  - Why automated analysis does not work
  - When visual exploration can help
- Visual Analytics Examples
  - Network Security
  - Document Analysis
  - Financial Analysis
  - High-dim. & Subspace Analysis
  - Molecular Biology
- Visual Analytics Perspectives





## Future Visual Analytics Topics

### Technical Challenges

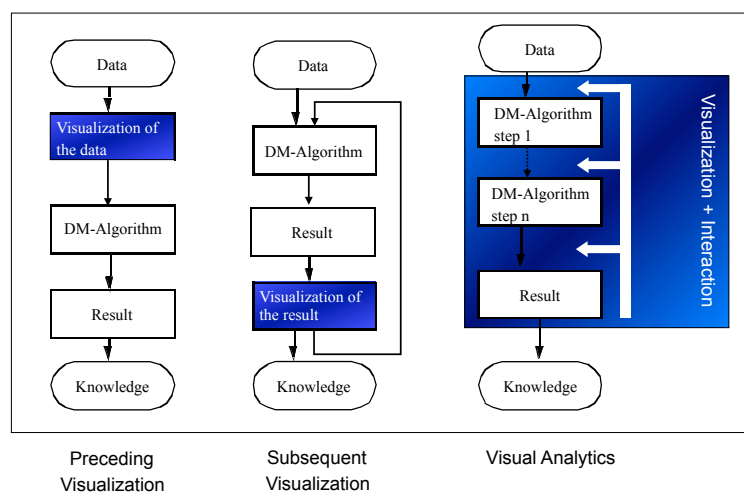
- Streaming Data
- Complex Structure
- Data with Uncertainty
- ...

### Applications

- E-Humanities
- Molecular Biology
- Multimedia
- ...

### Evaluation

## Visual Analytics



## Conclusion

**“All truths are easy to understand once they are discovered; the point is to discover them.”**

Galileo Galilei (1564-1642)

## Questions?

*[infovis.uni-konstanz.de](http://infovis.uni-konstanz.de)*

