# Part 10

## Evaluation & Usability

## Content :: Evaluation & Usability

Information Visualization Evaluation

Evaluation in Practice
- in2vis
- DisCō
- Stardinates

Crucial InfoVis Challenges



## Content :: Evaluation & Usability

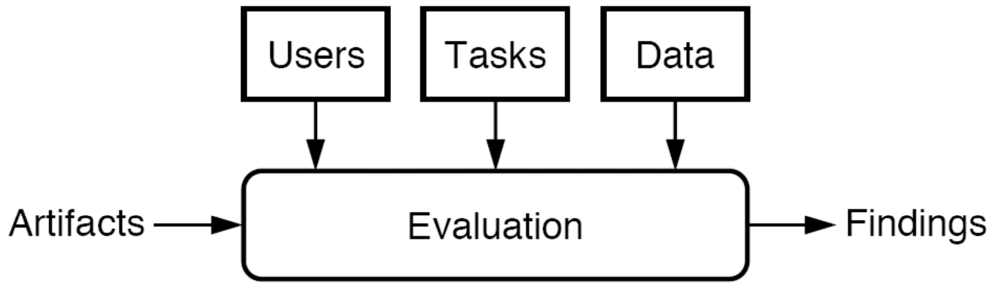
Information Visualization Evaluation

Evaluation in Practice
- in2vis
- DisCō
- Stardinates

Crucial InfoVis Challenges



## The Main Ingredients of Evaluation [Keim, et al. 2010 - RoadMap]




For Example,
- ***Artifact*** :: scatterplots
- ***Task*** :: helpful to find clusters
- ***Data*** :: a limited number of real valued attributes
- ***Users*** :: training in the proper interpretation

## Users

[Keim, et al. 2010 - RoadMap]

Can be professional well trained or lay persons
Can be proficient with computers or not
Can be young or old
…

***Difficult issues***

Expert are well trained and know the tasks but their time is precious and they are scarce resources

Students as found in our labs will not exhibit the same kinds of performance as experts for real tasks



## Tasks

[Keim, et al. 2010 - RoadMap]

**Several levels**

Low level: important but not "ecologically valid" and not sufficient

Can be done in clean lab settings



## Artifacts

[Keim, et al. 2010 - RoadMap]

**Several levels**

**Low Level Encodings**
e.g., grey value vs. size

**Component Level**
e.g., visualization/interaction technique

**System Level**
e.g., system X vs. system Y

**Environment Level**
e.g., integration of system X in environment Z



## Data

[Keim, et al. 2010 - RoadMap]

**Several levels**

**Low level** are homogeneous

**Mid level** are heterogeneous/multiple

**High level** are dynamic, varying, under specified and noisy

## Evaluation Areas

[Plaisant 2004]

**Controlled experiments comparing design elements**

- to compare specific widgets (e.g., alphaslider designs) or mappings of information to graphical display

**Usability evaluation** of a tool

- to provide feedback on the problems users encountered with a tool
- to show how designers can refine the design

**Controlled experiments comparing two or more tools**

- common type of study
- to compare a novel technique with the state of the art

**Case studies** of tools **in realistic settings**

- least common type of studies
- *advantage*
  - report on users in their natural environment doing real tasks
  - demonstrating feasibility and in-context usefulness
- *disadvantage*
  - time consuming to conduct,
  - and results may not be replicable and generalizable



## Approaches

GOMS: [Card, et al. 1983]

**Time to completion**

**Error rates**

**GOMS** - **Modeling and describing human task performance**

GOMS = Goals, Operators, Methods, and Selection Rules

**Goals** represent the goals that a user is trying to accomplish, usually specified in a hierarchical manner. **Operators** are the set of atomic-level operations with which a user composes a solution to a goal. **Methods** represent sequences of operators, grouped together to accomplish a single goal. **Selection Rules** are used to decide which method to use for solving a goal when several are applicable.

**Benchmarks Repositories**

**Infovis Contest**

http://www.cs.umd.edu/hcil/InfovisRepository/

**Visual Analytics Benchmark Repository**

http://hcil.cs.umd.edu/localphp/hcil/vast/archive/

**Insights**

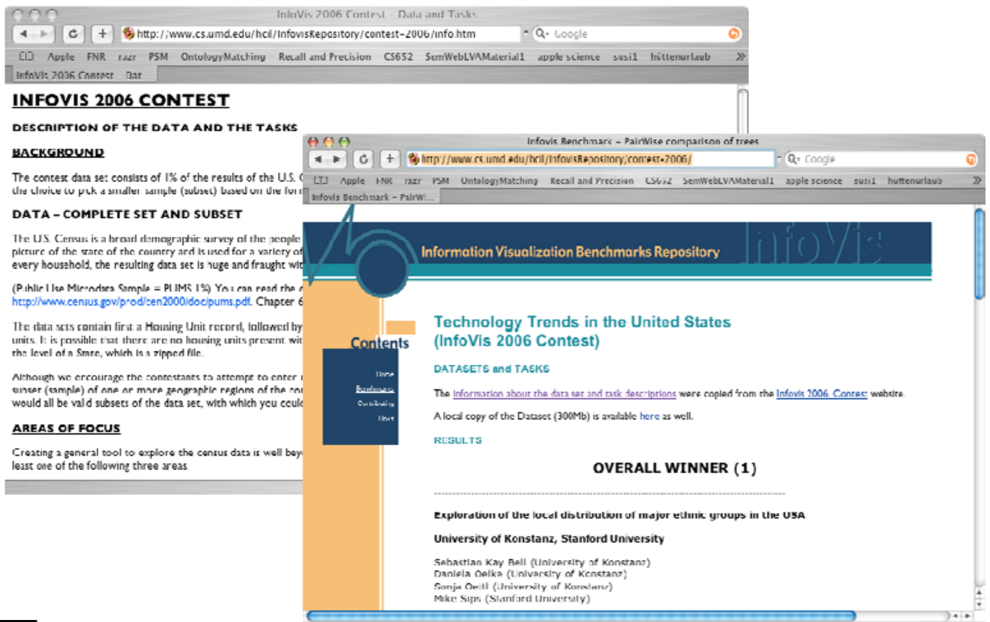
High level cognitive processes: reasoning, causality, explanation, ...

**BELIV'06**
**BE**yond time and errors: novel eva**L**uation methods for **I**nformation **V**isualization



## InfoVis Contest 2006





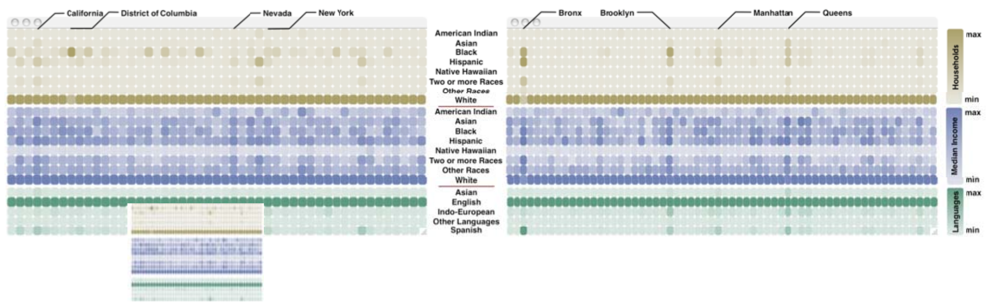



## InfoVis Contest 2006 Winners

[Belle, et al. 2006]



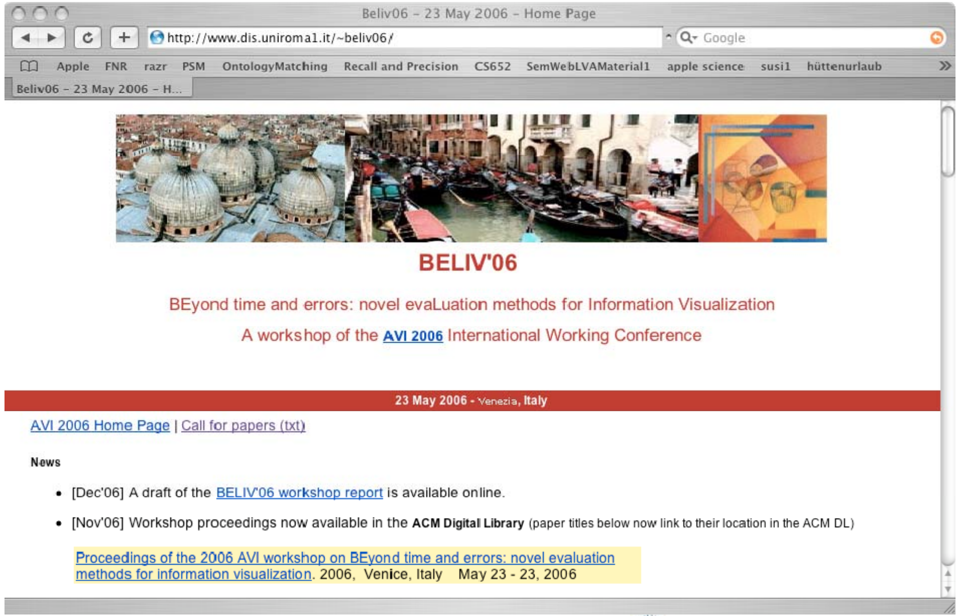
Exploration of the Local Distribution of Major Ethnic Groups in the USA




Visualization of the local distribution of major ethnic groups, their income and the regionally spoken languages. Geographical units are represented by columns, the data for the categories such as household, income, and language data by rows. Left: state level, middle: county level for state New York, bottom left: again state level, but with an iPod-resolution of 220x176 pixel (in comparison to the other screenshots having a resolution of approx. 800x400 pixel).
(Column-by-column normalization strategy)

## BELIV'06 Workshop








## BELIV'06 Workshop Description



'Controlled experiments remain the workhorse of evaluation but there is a growing sense that information visualization systems need new methods of evaluation, from longitudinal field studies, insight based evaluation and other metrics adapted to the perceptual aspects of visualization as well as the exploratory nature of discovery.'
[...]
'e.g. new ways of conducting user studies, definition and assessment of infovis effectiveness through the formal characterization of perceptual and cognitive tasks and insights, definition of quality criteria and metrics. Case study and survey papers are also welcomed when clearly presenting general guidelines, practical advices, and lessons learned.'



## Evaluation - Specification of Goals



What to investigate? What are the research questions?
How to investigate in order to get answers?

Domain knowledge helps to identify relevant research questions

Example: E-learning system

- **Question** 1: Did the participants learn the content?
  - **Method**: Exam
- **Question** 2: Did the participants like to use the system?
  - **Method**: Interviews
- **Question** 3: Is the system easy to use?
  - **Methods**: Observation, Software logs



## Evaluation - Implementation of a Study



Select and find participants for the study (subjects)

Laboratory setting

+ clear conditions allow for good identification of causality
– simulated and restricted setting could yield irrelevant statements

Field study

+ lifelike and informative
– identification of valid statements is difficult because of the complexity (high number of variables)

## Types of Evaluation (1)

[Robert Stakes] 

### Formative evaluation

evaluation and development are done in parallel
(iterative development process)
feedback about usability and utility
results cause improvement of the tool

### Summative evaluation

development of the tool is finished
assessment of efficacy and features (e.g., comparative evaluation)
results may support buyers' decisions

'When the cook tastes the soup, that’s formative;
when the guests taste the soup, that’s summative.'



## Types of Evaluation (2)

[Robert Stakes] 

### Quick-and-dirty

informal and non-systematic
small number (2 to 10) subjects use the product and
tell what they think about it
usually conducted during product development
low cost

### Scientific evaluation

elaborated process
definition and validation of scientific hypotheses
minimum of 20 subjects for quantitative studies
standardized evaluation methods: quantitative or qualitative
conducted to investigate core questions of a product or research topic, e.g., command-line interaction versus direct manipulation of objects



## Evaluation Methods

Interviews / focus groups
Questionnaire
Observation
Software logs
Thinking Aloud



## Interviews / Focus Groups



### Interviews

can give a differentiated idea of the usability and efficacy of a tool
subjects cannot always report their behavior,
since some cognitive processes are automatic and unconscious
subjects' intentions can provide reasons
for measurements and objective data
allows for in-depth analysis
based on guidelines

### Focus groups

discussions with groups
sometimes a problem to ensure equal participation
group situation could influence topics
based on guidelines for discussion and moderation

## Questionnaire

In contrast to interviews questionnaires allow for studying large groups of people (quantitative evaluation)

Can yield representative data

Should avoid bias

Difficult to prevent misunderstandings because of different interpretations

Simple questions

Closed questions: given answer categories

Open questions: free answers, etc.



## Observation

Collection of information does not depend on subjects' reports (sometimes subjects can give no information about their activities)

Subjective falsifications are impossible

Problem to understand why persons set certain actions.

No guarantee that the observed person behaves naturally (Hawthorne effect)

Observations can take place in laboratories or in real-world situations

Yields an abundance of data

Difficult to select relevant data

Based on guidelines (what to observe)



## Software logs

Monitoring tool collects data about computer and user activities, e.g., about number and location of clicks or type of keyboard input

Observes only a limited number of activities

Delivers high amount of data

Procedure is not visible for user

Does not intervene user's activities

Activity sequences yield more information than single step

Analysis of activity sequences is difficult

Software logs do not register the intentions or goals of the users



## Thinking Aloud

Mixes observation and questioning

Subjects are asked to describe their thoughts while using the product

Gives more details than interviews, because information filtering is reduced

Thinking aloud could impede the interaction processes

It is difficult to express the thoughts if interaction with the tool requires attention

Sometimes crucial situations are not reported

Provides with highly relevant and interesting data

# Usability Evaluation

**Guidelines checklist**
Broad principles, empirically-derived results, established conventions

**Cognitive walkthrough**
Based on specific tasks: 'simulation' of a user (model)
How difficult is it for the user to identify and operate the interface element most relevant to their current subgoal?

**Pluralistic walkthrough**
Users + developers + HCI experts: Identify primary tasks, step through those tasks
Different Stakeholders adopt different goals / perspectives
=> more usability problems are identified

**Consistency inspection**
Quality control technique: consistency in: design, graphics, text, interaction

**User testing**
4-10++ "users", series of tasks, observation, thinking aloud, log files, ...

**Performance measurement**
Efficiency of use, task completion times; useful for comparative studies



# Usability Evaluation: Relevant Links

| | |
|---|---|
| www.useit.com | Jakob Nielsen |
| www.jnd.org | Don Norman |
| www.nngroup.com | Nielsen Norman Group |
| www.asktog.com | Bruce Tognazzini |
| www.usabilityfirst.com | Diamond Bullet Design |



[Nielsen 1994]

# Heuristic Evaluation (1)

A small number of trained evaluators (typically 3 to 5) separately inspect a user interface by applying a set of **'heuristics'**, **broad guidelines** that are generally relevant

Use more evaluators if usability is critical or evaluators aren't domain experts

Go through interface **at least twice**:
1. Get a feeling for the flow of the interaction
2. Focus on specific interface elements

**Write reports**
Reference rules, describe problem, one report for each problem.

**Don't communicate** before all evaluations are completed!

Observer assists evaluators

Use additional usability principles

Provide **typical usage scenario** for domain-dependent systems

Conduct a debriefing session (provides design advice)

Phases:
**pre-evaluation training / evaluation / debriefing** / severity rating



[Nielsen 1994]

# Heuristic Evaluation (2): Rules

**Visibility of system status**
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

**Match between system and the real world**
The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

**User control and freedom**
Users often chose system functions by mistake and will need a clearly marked „emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

**Consistency and standards**
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

**Error prevention**
Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

[Nielsen 1994]

## Heuristic Evaluation (3): Rules

### Recognition rather than recall

Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

### Flexibility and efficiency of use

Accelerators — unseen by the novice user — may often speed up the interaction for the expert user to such an extent that the system can carter to both inexperienced and experienced users. Allow users to tailor frequent actions.

### Aesthetic and minimalist design

Dialogues should not contain information which is irrelevant od rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

### Help users recognize, diagnose, and recover from errors

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

### Help and documentation

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.



## Heuristic Usability Evaluation (1) Forsell & Johansson, 2010

A new set of 10 heuristics out of 63 heuristics (from 6 earlier published heuristic sets)

Especially tailored to the evaluation of common and important usability problems in *Information Visualization techniques*



## Heuristic Usability Evaluation (1) Forsell & Johansson, 2010

1. **B5. Information coding.** Perception of information is directly dependent on the mapping of data elements to visual objects. This should be enhanced by using realistic characteristics/techniques or the use of additional symbols.

2. **E7. Minimal actions.** Concerns workload with respect to the number of actions necessary to accomplish a goal or a task.

3. **E11: Flexibility.** Flexibility is reflected in the number of possible ways of achieving a given goal. It refers to the means available to customization in order to take into account working strategies, habits and task requirements.

4. **B7: Orientation and help.** Functions like support to control levels of details, redo/undo of actions and representing additional information.

5. **B3: Spatial organization.** Concerns users' orientation in the information space, the distribution of elements in the layout, precision and legibility, efficiency in space usage and distortion of visual elements.



## Heuristic Usability Evaluation (1) Forsell & Johansson, 2010

6. **E16: Consistency.** Refers to the way design choices are maintained in similar contexts, and are different when applied to different contexts.

7. **C6: Recognition rather than recall.** The user should not have to memorize a lot of information to carry out tasks.

8. **E1: Prompting.** Refers to all means that help to know all alternatives when several actions are possible depending on the contexts

9. **D10: Remove the extraneous.** Concerns whether any extra information can be a distraction and take the eye away from seeing the data or making comparisons.

10. **B9: Data set reduction.** Concerns provided features for reducing a data set, their efficiency and ease of use

## Jean-Daniel Fekete

*Slide Dagstuhl 2010*

# Newer Methodologies

Recently, Ronald Rensink advertised using "Vision Science" methods to evaluate visualizations
– Ronald A. Rensink, and Gideon Baldridge, The Perception of Correlation in Scatterplots. Computer Graphics Forum, 29: 1203-1210. 2010.

Instead of counting insights, look at decisions on sample datasets (decision theory)
– Expressing insight is a high-level complex process
– Decision is much more direct, does not need verbal expression

Use Log/Trace analysis for longitudinal studies
– Instrument programs (at the right levels) and analyze the logs (use visualization to explore)
– Nathalie Henry, Niklas Elmqvist and Jean-Daniel Fekete. **A Methodological Note on Setting Up Logging and Replay Mechanisms in InfoVis Systems**. *In BELIV'08, a workshop at the ACM CHI 2008 conference*, April 2008.

Use MRI or BCI to study brain response to VA systems
– Detect insight?
– Measure cognitive load and fatigue

Use Eye Tracking to study attention and cognitive load
– Chris Weaver. "Look Before You Link: Eye Tracking in Multiple Coordinated View Visualization". *BELIV '10*, Atlanta, GA, April 2010.



## Content :: Evaluation & Usability

Information Visualization Evaluation

Evaluation in Practice
- in2vis
- DisCō
- Stardinates

Crucial InfoVis Challenges



## in2vis Project: Visualization



Domain: therapy of anorectic young women

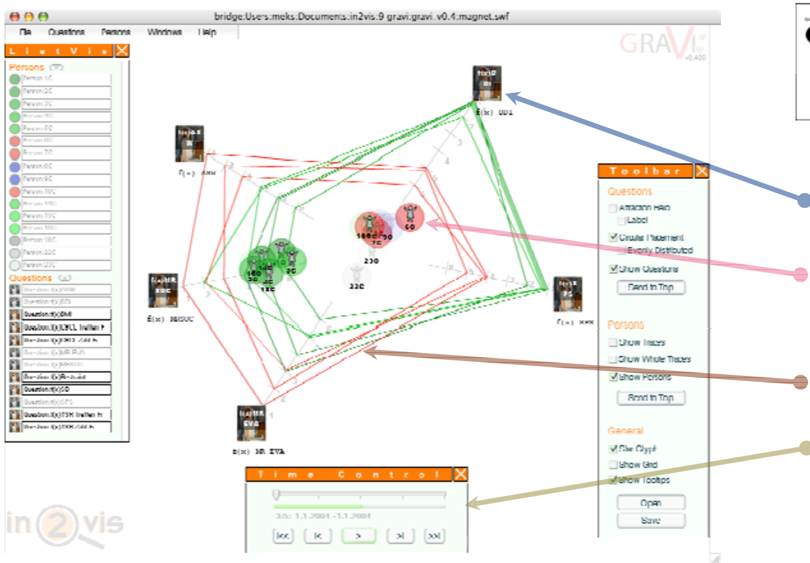
Support psychotherapists
- during therapy a large amount of highly complex data is collected
- patients and parents have to fill in numerous questionnaires
- (before, during, and after the therapy)

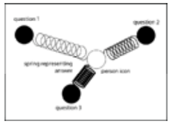
Statistical methods are insufficient
- small sample size (~27 patients in three years)
- high number of variables (~40 different questionnaires with ~40 items. some of them every week, others every 3 months)
- time-oriented data

### Aims of the therapists
- **predict success or failure of the therapy** for the individual patients
- analyze the **factors influencing anorexia nervosa**
- reduce the number of questionnaires the patients have to fill out

## in2vis Project: Visualization

[in2vis]





spring-based

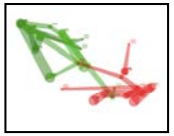
questions/ questionnaires

patients

attraction field

star glyph

time steps





## in2vis Project: Evaluation

[Rester, et al. 2006]

| Stage | Method | Subjects | Aim | Collected Material |
|---|---|---|---|---|
| Usability | usability inspection | 1 usability expert | spot most obvious glitches | 31 usability problems |
| | heuristic evaluation | 27 semi-experts in usability | in depth testing | 447 reports documenting 576 problems (221 different) |
| | focus groups | | additional usability assessment | no new problems BUT different perspective |
| Insight Study (Gravi++, EDA, Machine Learning) | insight reports | 33 domain novices | patterns of insight & cognitive strategies | 876 reports documenting 2166 insights |
| | log files | | used vis. options & exploration strategies | 56055 log file entries |
| | focus groups | | relativize findings & aids correct interpretation | transcription of 3x 100min |
| Case Study | interviews | 2 real users | feasibility & usefulness in real life | transcription of 1x 60min |
| | thinking aloud | | | notes on 1x 180min |
| Transferability | interviews | 14 experts of other domains | usefulness in other domains | transcription of 14x 60min |



## in2vis Project: Usability Evaluation Setting



[Rester, et al. 2006]

### Motivation

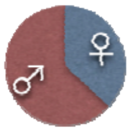
improve visualization application

preclude mix-up of usability problems with weaknesses of visualization method as such

### Sample

27 students of informatics-related studies

semi-experts




### Methods

informal usability inspection / guideline review

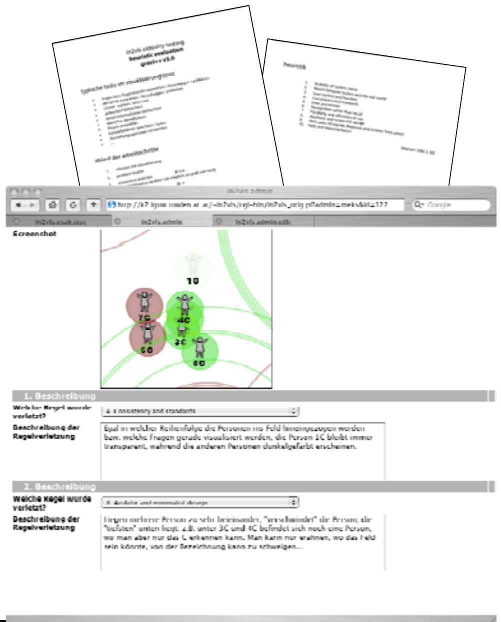
heuristic evaluation

focus groups



## in2vis Project: Usability Evaluation Setting

[Rester, et al. 2006]



### Handouts

typical tasks

detailed procedure

heuristics (outline)

### Report system

screenshot upload

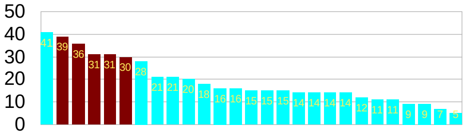
violated rule(s)

description(s)

# in2vis Project: Usability Evaluation Results



**3h** sessions
**27** subjects
**447** reports
**513** violations




max 41
avg 19
min 5

| Rule | Mentions | Percentage |
|---|---|---|
| 1. Visibility of system status | 63 | 12.28 |
| 2. Match between system and the real world | 40 | 7.80 |
| 3. User control and freedom | 59 | 11.50 |
| 4. Consistency and standards | 105 | 20.47 |
| 5. Error prevention | 23 | 4.48 |
| 6. Recognition rather than recall | 19 | 3.70 |
| 7. Flexibility and efficiency of use | 32 | 6.24 |
| 8. Aesthetic and minimalist design | 52 | 10.14 |
| 9. Help users recognize, diagnose, and recover from errors | 12 | 2.34 |
| 10. Help and documentation | 33 | 6.43 |
| 11. Other Rule | 75 | 14.62 |
| | 513 | 100.00 |

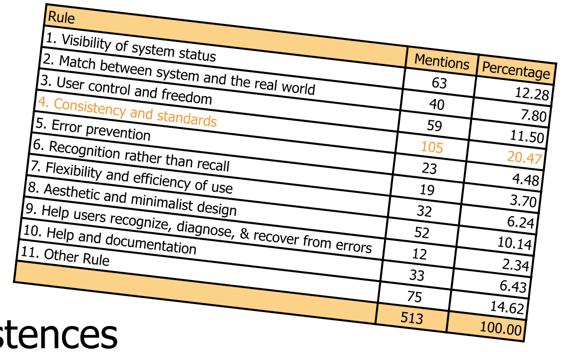




# in2vis Project: Usability Evaluation Results



Frequency of assigned principles is affected amongst others by:



| Rule | Mentions | Percentage |
|---|---|---|
| 1. Visibility of system status | 63 | 12.28 |
| 2. Match between system and the real world | 40 | 7.80 |
| 3. User control and freedom | 59 | 11.50 |
| 4. Consistency and standards | 105 | 20.47 |
| 5. Error prevention | 23 | 4.48 |
| 6. Recognition rather than recall | 19 | 3.70 |
| 7. Flexibility and efficiency of use | 32 | 6.24 |
| 8. Aesthetic and minimalist design | 52 | 10.14 |
| 9. Help users recognize, diagnose, & recover from errors | 12 | 2.34 |
| 10. Help and documentation | 33 | 6.43 |
| 11. Other Rule | 75 | 14.62 |
| | 513 | 100.00 |

- quantity of true existences
- comprehension of the principles by subjects
- difficulty of tracking down violations of the different principles
- domain knowledge needed to find problems of different categories

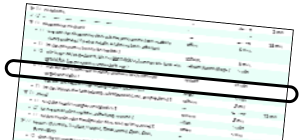




# in2vis Project: **Usability Evaluation Results (Focus Groups)**



3 groups : 90 minutes : 15 questions
biggest usability problem(s)
27 different problems in 46 statements



| Biggest Usability Problem (Total Mentions >1) | FG1 | FG2 | FG3 | Total Mentions | |
|---|---|---|---|---|---|
| Undo/Redo is missing. | 3 | 4 | | 7 | 10 |
| Attraction Field: which circle & person do correspond. | 3 | | | 3 | |
| Performance problem. | | 2 | 1 | 3 | heur. |
| Time control feedback is confusing. | 3 | | | 3 | |
| Traces: many bugs (size, disappear, remain, numbers remain) | 1 | 2 | | 3 | eval. |
| Everything should be controllable via menu. | 2 | | | 2 | |
| Help is missing. | 2 | | | 2 | 18 |
| Reset Window Position is missing. | 2 | | | 2 | |
| Bug: load / save. | 1 | 1 | | 2 | |
| No project-files but saved states. | | 2 | | 2 | |
| | | | | 29 | |

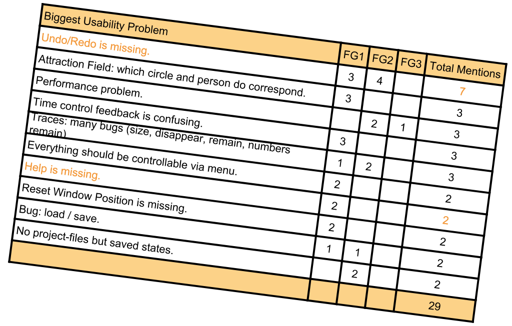




# in2vis Project: **Usability Evaluation Results (Focus Groups)**



A problem's importance may be assessed among others by:

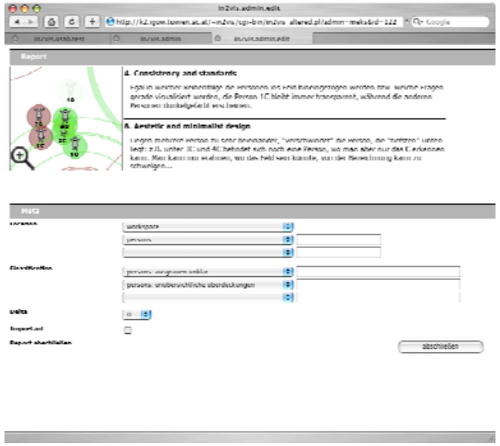


| Biggest Usability Problem | FG1 | FG2 | FG3 | Total Mentions |
|---|---|---|---|---|
| Undo/Redo is missing. | 3 | 4 | | 7 |
| Attraction Field: which circle and person do correspond. | 3 | | | 3 |
| Performance problem. | | 2 | 1 | 3 |
| Time control feedback is confusing. | 3 | | | 3 |
| Traces: many bugs (size, disappear, remain, numbers remain) | 1 | 2 | | 3 |
| Everything should be controllable via menu. | 2 | | | 2 |
| Help is missing. | 2 | | | 2 |
| Reset Window Position is missing. | 2 | | | 2 |
| Bug: load / save. | 1 | 1 | | 2 |
| No project-files but saved states. | | 2 | | 2 |
| | | | | 29 |

- total number of mentions within all groups
- number of groups in which it is stated
- distribution of the total number across groups

## in2vis Project: Usability Evaluation Results





3-tier location
Unique classification(s)

Some results

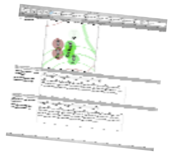
- 221 unique problems
- 576 documentations (513)
- top-evaluator(s): 47 (41)
- easy to spot problems
- many bugs (20%)
- feature requests (15%)
- person-icons (9%)
- inconsistencies (6%)
- question-icons (5%)
- menu (5%)





## in2vis Project: Usability Evaluation - Summary

Informal usability inspection identifies obvious weaknesses
- increases quality of heuristic evaluation



Heuristic evaluation proper method
- general framework is useful for training
- screenshots help comprehending, reproducing, interpreting

Focus groups reveal overall view of evaluators
- efficiently identify dramatic problems

3 methods give a different perspective on usability issues
- complement each other to a broader view





## in2vis Project: Insight Study



Tools used by subjects

| | |
|---|---|
| gravi++ | interactive infovis |
| eda | explorative data analysis |
| ml | machine learning |

| | | | |
|---|---|---|---|
| intro domain | | | 60 min |
| intro eda | | | 30 min |
| intro ml | | | 30 min |
| intro gravi | | | 30 min |
| 9 subj. | 12 subj. | 12 subj. | |
| ml | gravi | eda | 60 min |
| eda | ml | gravi | 60 min |
| gravi | eda | ml | 60 min |

Comparative study

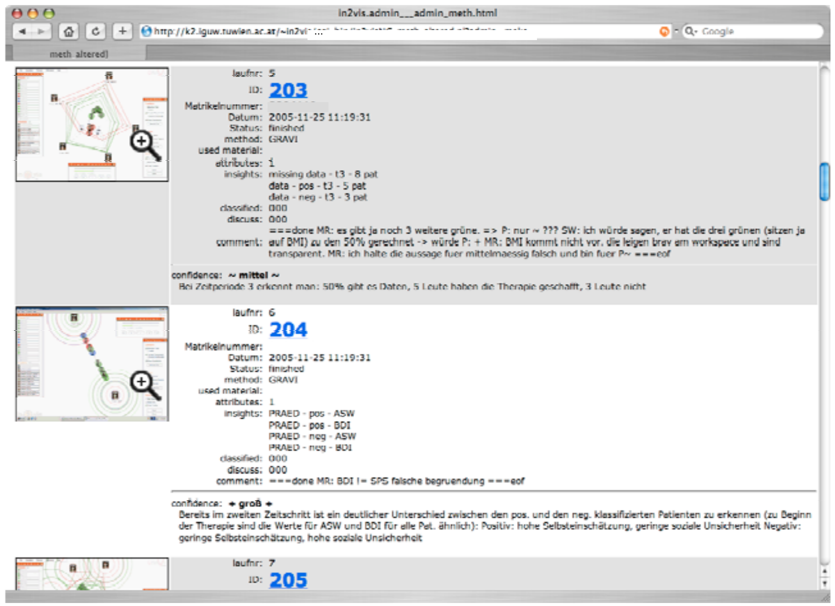
- **scenarios** (data subset): undirected exploration
- **concrete tasks** (data subset + question): still argument required

Goals

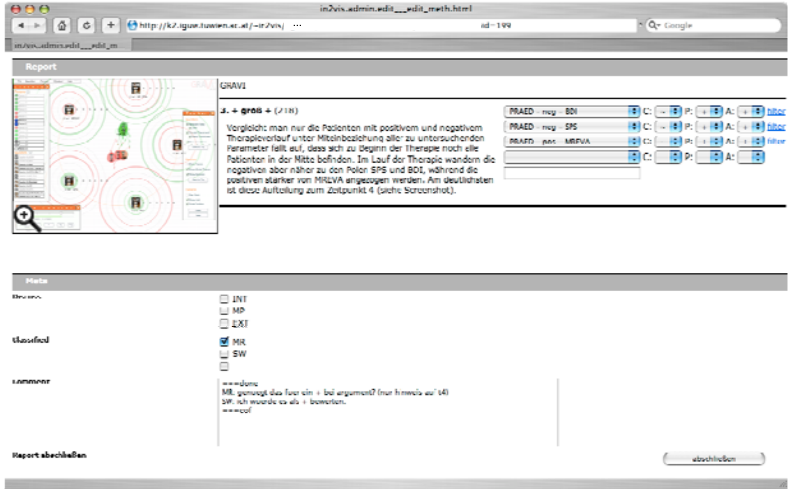
- types of insight gained with different tools
- different insights by varying orders of used tools?
- patterns of insight & cognitive strategies



## in2vis Project: Report System (1)

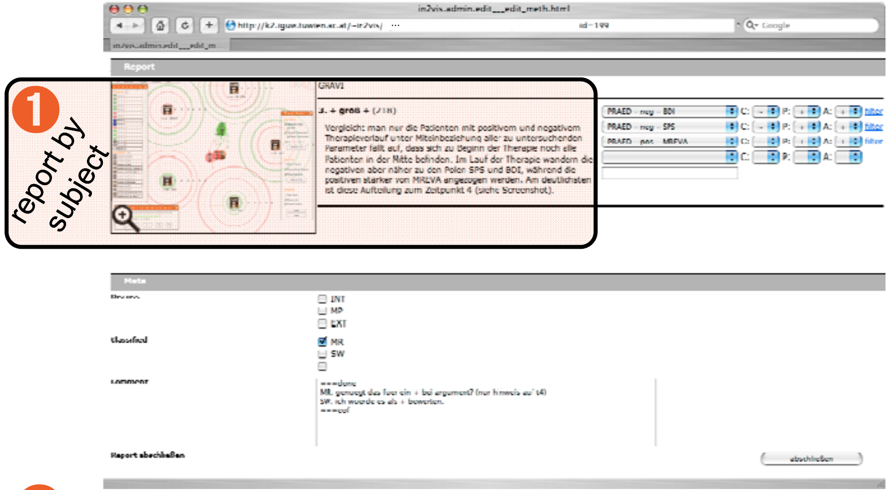


## in2vis Project: Report System (2)





## in2vis Project: Report System (3)

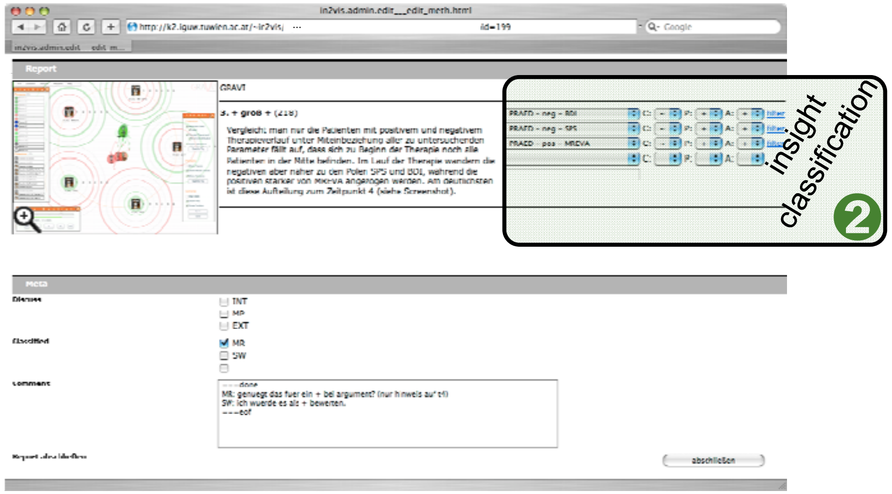



1 report generated by subject including

uploaded screenshot
confidence rating ( high | mid | low )
insight description



## in2vis Project: Report System (4)

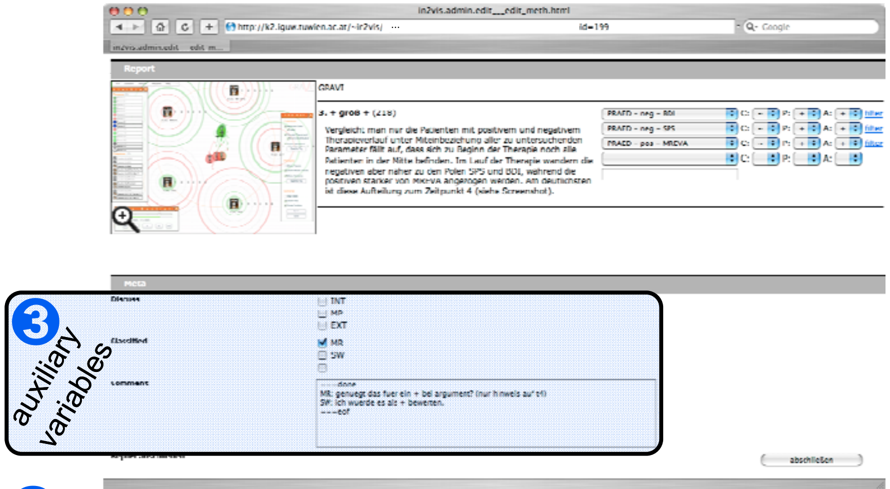



2 insight classification including

insight identifier
complexity ( complex | regular | trivial )
plausability ( high | mid | low )
argument ( correct | missing | wrong )



## in2vis Project: Report System (5)




3 auxiliary variables including

various to-discuss flags (e.g., between investigators, with domain experts)
classification status (proofread by a 'second set of eyes')
comment/discussion field for investigators

[Rester, et al. 2006]

## in2vis Project: Evaluation Issues



### Insight reports

Should long reports be split in basic insights or are they a unique occurrence of a complex insight?

Are they simply a cumulative documentation from a subject who did not adhere to the test procedure of reporting insights immediately after having them?

➔ for comparability splitting is necessary.

### Log files

How should one account for the learning curve?
Log file chunks between later insights will probably not reflect the explorative interactions leading to an insight.

➔ analyze log files as whole and identify different subjects and compare their insights without time-dependency.



## Content :: Evaluation & Usability

Information Visualization Evaluation

Evaluation in Practice

- in2vis
- DisCō
- Stardinates

Crucial InfoVis Challenges



## Project DisCō (lat. ich lerne)

**visual DIScovery and COmmunication of complex time patterns in non regularly gathered multigranular and multivariate data**

FIT-IT [ Visual Computing

Silvia Miksch, Wolfgang Aigner, Alessio Bertone, Tim Lammarsch, Thomas Turic

Johannes Gärtner, Dieter Punzengruber, Sabine Wahl

Hanna Risku, Eva Mayr, Michael Smuc



## DisCō Project (lat. I learn)

**Data**
time-oriented,
irregularly sampled
multivariate, multigranular

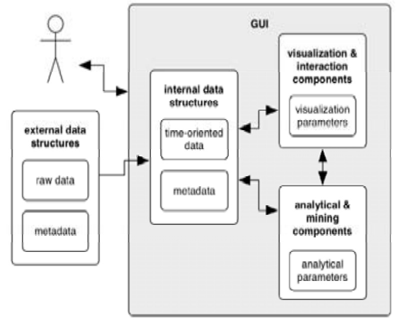
**Task**
Discovery of complex patterns and relationships

**Goals**

Interactive visualization of data and results with visual parameterization

Analytical methods for analyzing time-oriented data

Ensuring usability and utility of developed methods via User-Centered Design



[Smuc et al., 2008]

## Research and Development Process

(1) Task & user analysis

In-depth interviews: users tasks, needs & goals

(2) Iterative process & user-driven design

Iterative design, Usability-inspection, focusgroups

(3) Usability testing & data analysis

Usability-evaluation



DisCō

Visual **DIS**covery and **CO**mmunication of complex time patterns

funded by FIT-IT

## Visualizations at first sight.
## Do insights require training?

Michael Smuc, Eva Mayr, Tim Lammarsch, Alessio Bertone, Wolfgang Aigner, Hanna Risku and Silvia Miksch

http://www.donau-uni.ac.at/disco

[Smuc et al., 2008]

## DisCō: Insight Study

**Insights**

**Insight Study: Visualizations at First Sight.**

Material: Cycleplot & Multiscale

Method

Insight Counters

Insight Visualization

**Discussion: Do Insights Require Training?**



[Smuc et al., 2008]

## Insights

= the generation of new knowledge by individuals out of visualization for data analysis.
(Low granularity – single observations)

# Insight Study: **Visualizations at First Sight**

## Research Questions:

Can users generate insights without prior knowledge about the visualization?

Can users generate insights without domain knowledge?





# Method

Mockup-interviews

Think-aloud technique

Instruction:

„Take a look at this visualization and think aloud while exploring it"

Analysis:

- Transcription of interviews
- Segmentation
- Coding of insights





# Insight Categories

| | | |
|---|---|---|
| **Integration of Prior Knowledge** | | "It decreases until 6 in the morning, to a minimum. I assume this is due to […], to my knowledge, change of shift." |
| **Visualization Insights** | **How-insight** | "The more green the less assignments, the more blue the more assignments." |
| | **Meta-insight** | "Okay, first I'm looking at the days, if I can detect any patterns." |
| | **Improvement-insight** | "It would be good to be able to filter out one day." |
| **Data Insights** | **Cycleplot: Cycle** | "Starting in the morning it rises to a peak around 10, 11 am. Then it calms down at noon with a second peak around 4, 5 pm. Then it falls down again." |
| | **Cycleplot: Trend** | "The first Monday is high, descending on the second, and rising again on the third and forth." |
| | **Multiscale: Overview** | "Sundays are rather low, on average." |
| | **Multiscale: Detail** | "Especially at noon it's higher than before or after noon. It's always darkest then." |



# Innovations: Highlight 1

**Goal**

Development of **methods and measures** for the Usability of visualizations and visualization tools

**Problem**

Benefits of classical Usability measures like **completion time** and **errors** are limited, esp. for design of Visual Analytics tools

**State of the Art**

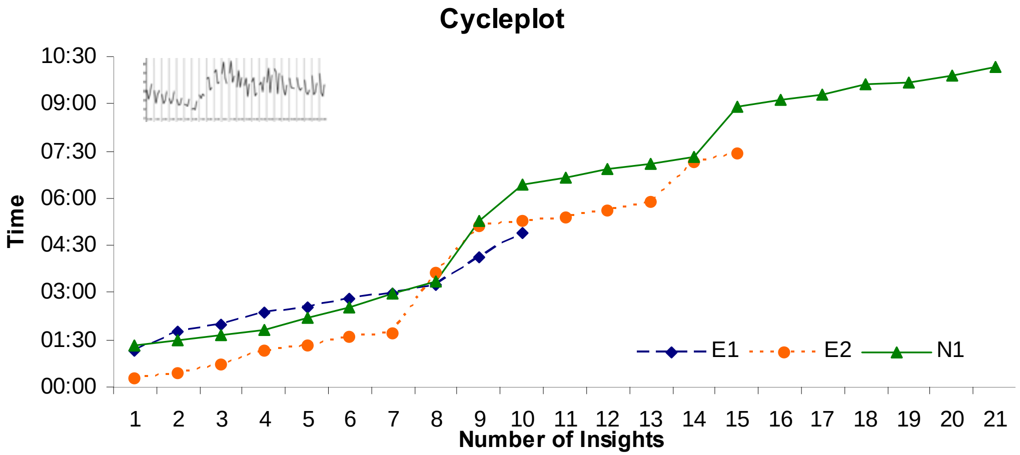
Productivity measures like counting the number of **insights** [North, 06]

**Our Solution**

Development of the **Relational Insight Organizer** (RIO) optimized for iterative design [Smuc et al., 2008]

[Smuc et al., 2008]

# Insight Counters

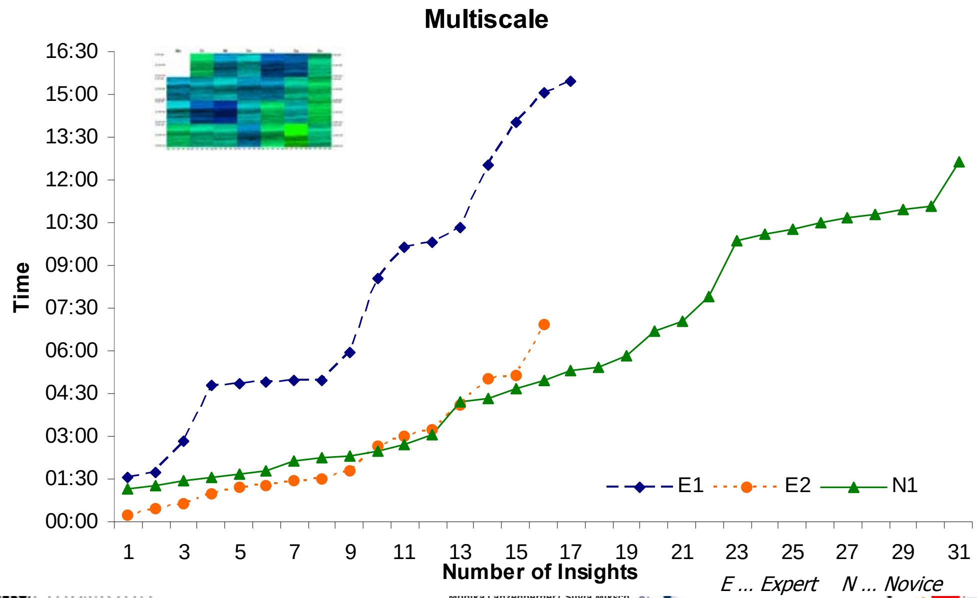



*E … Expert N … Novice*



[Smuc et al., 2008]

# Insight Counters

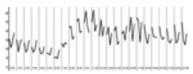


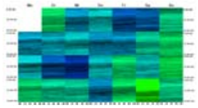

*E … Expert N … Novice*

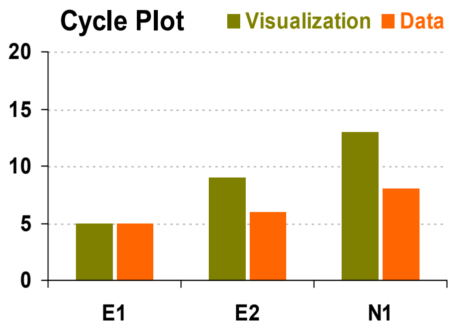
informations-visualisierung Monika Lanzenberger/ Silvia Miksch DEPARTMENT FÜR INFORMATION UND KNOWLEDGE ENGINEERING ieg TU VIENNA

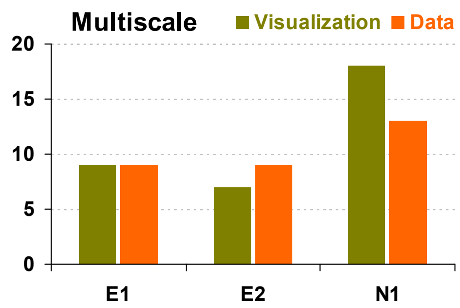
[Smuc et al., 2008]

# Insight Counters







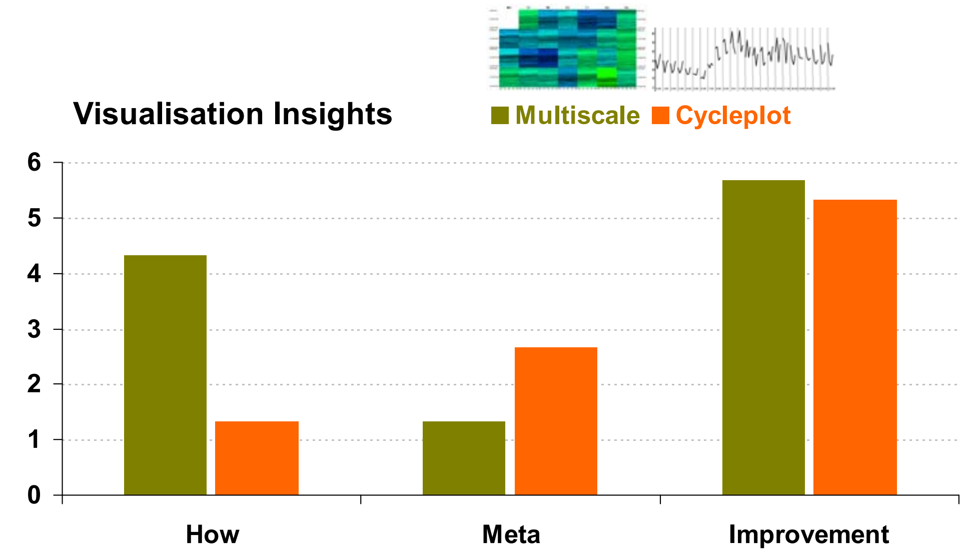




*E … Expert N … Novice*

informations-visualisierung Monika Lanzenberger/ Silvia Miksch DEPARTMENT FÜR INFORMATION UND KNOWLEDGE ENGINEERING ieg TU VIENNA

[Smuc et al., 2008]

# Insight Counters

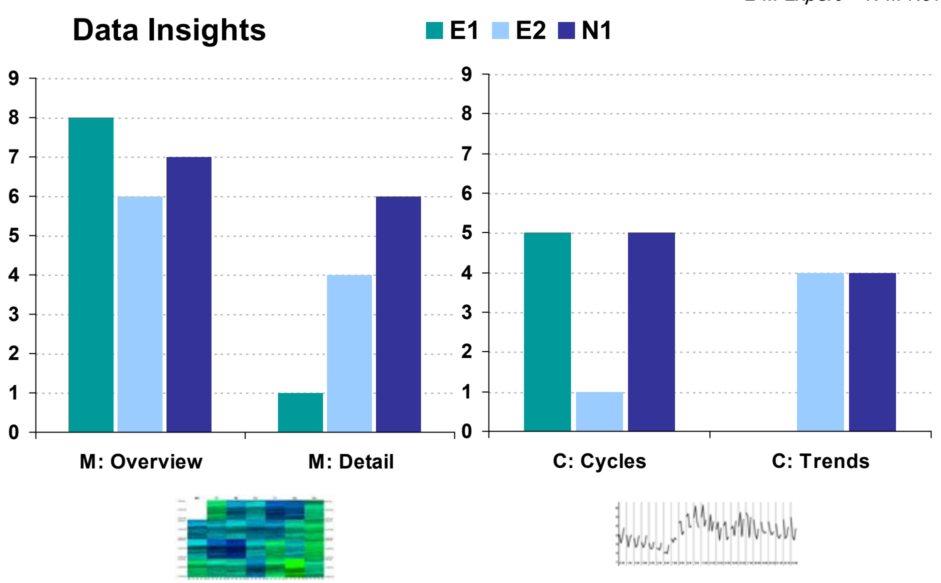



informations-visualisierung Monika Lanzenberger/ Silvia Miksch DEPARTMENT FÜR INFORMATION UND KNOWLEDGE ENGINEERING ieg TU VIENNA

[Smuc et al., 2008]

# Insight Counters

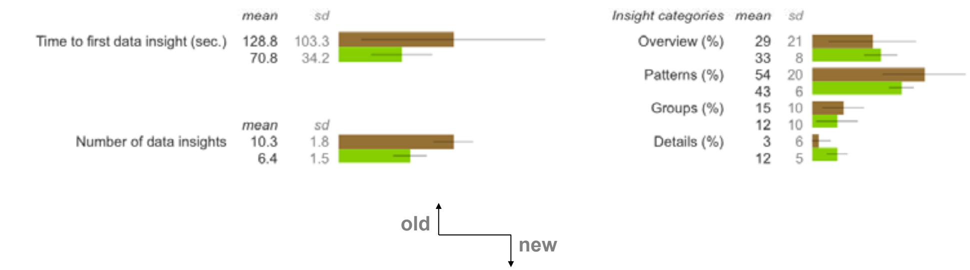
*E ... Expert   N ... Novice*



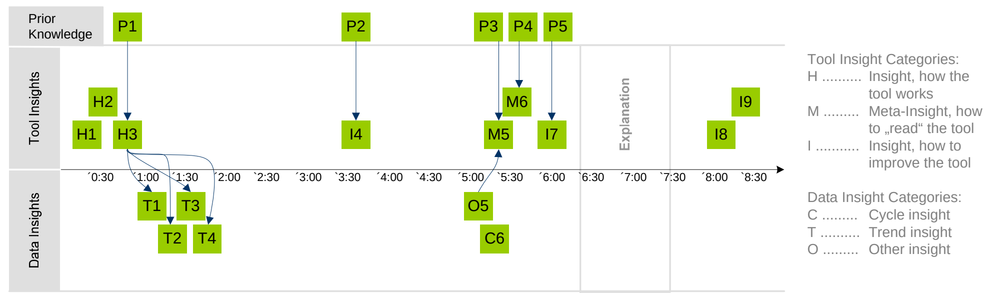



[Smuc et al., 2008]

**Multi Scale Plot vs. Cycle Plot**




**RIO of user 3 for Cycle Plot**

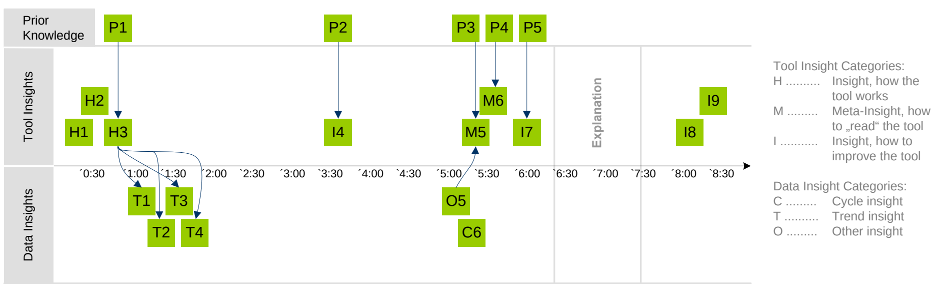




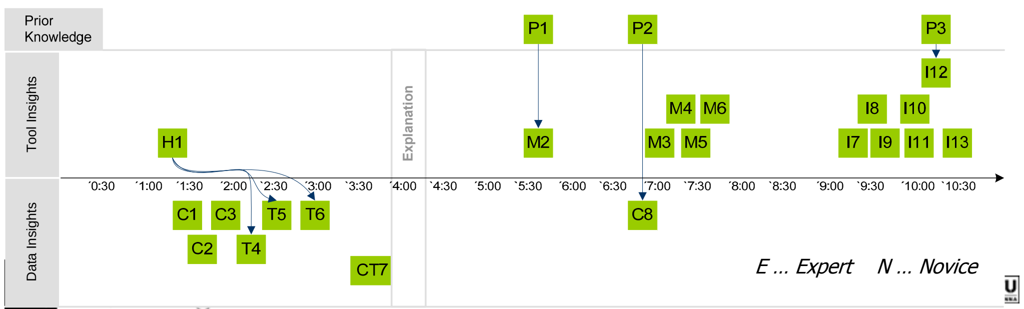

[Smuc et al., 2008]

# Insight Visualization

E2: Uptake Graph for Cycle Plot




N1: Uptake Graph for Cycle Plot




*E ... Expert   N ... Novice*

[Smuc et al., 2008]

# Discussion

## Do insights require training?

- Participants were able to generate insights from the start
- Domain knowledge was not necessary for insights
- Insights into the visualization were needed prior to data insights, but no full understanding

## Is expert knowledge beneficial?

- Not necessarily
- Prior knowledge was used to interpret data
- Experts‘ existing cognitive scripts maybe hindered more flexible analysis

[Smuc et al., 2008]

## Discussion of Methodology

Similar insights by expert and novice users

Mockup tests did generate complex data insights

Insight counters provide limited findings for iterative design, rather qualitative analysis of insights is needed

Small sample can provide useful ideas for improvements

Limitations
- Open task
- Sample size



## Content :: Evaluation & Usability

Information Visualization Evaluation

Evaluation in Practice
- in2vis
- DisCō
- Stardinates

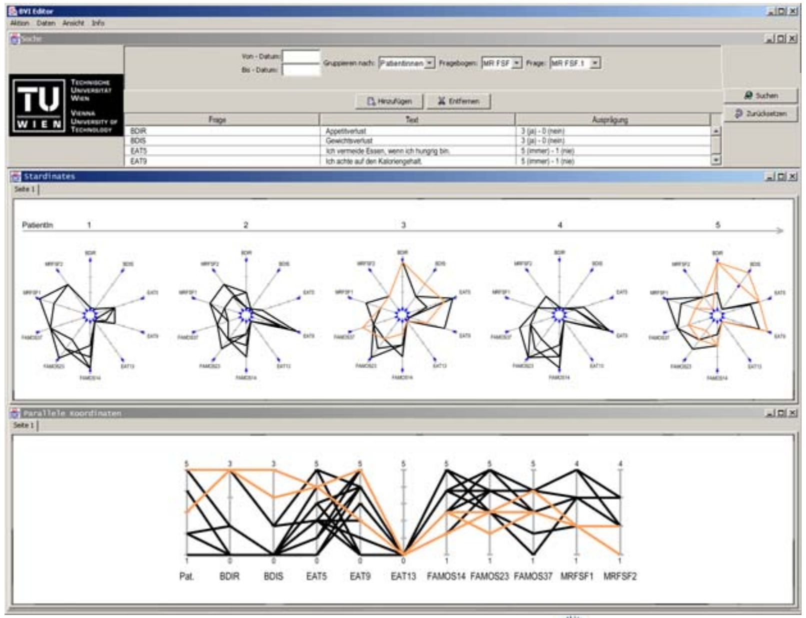
Crucial InfoVis Challenges



[Lanzenberger 2003]

## LinkStar







[Gärtner, et al. 2002, Lanzenberger 2003]

## Evaluation of the Interactive Stardinates



ViCo - Metric to measure the complexity of visualization
- Analyze the tasks of the users
- Define basic operations (e.g., Read, Compare, Highlight)
- Develop an algorithm

Compare Parallel Coordinates and the Stardinates by calculating the complexity of their algorithms

[Lanzenberger 2003]

## Complexity of Interactive Stardinates

| Task | Code |
|---|---|
| Task 1:<br>The Algorithm | Read Shape & Decide(Op1)<br>/*One may be able to recognize the relation if it is strong enough and the changes over time occur in a homogeneous way */<br>IF no clear Relation THEN<br>  FOR MANY Data Bundles (a*#P)<br>    Read Shape of Data Bundle & Decide(Op2)<br>IF still no clear Relation THEN<br>  FOR MANY Lines (a*#T)<br>    Highlight(Op4)<br>    Read Area Shape (Op5)<br>    Compare Area (Op7)<br>IF still no clear Relation THEN<br>  FOR EACH Data Point (#P*#T*#B)<br>    Read Data Point (Op6)<br>    Compare Data Point (Op8) |
| Complexity for Task 1 | Best Case:<br>Op1<br>Middle Cases:<br>Op1 + a*#P*(Op2)<br>or: Op1 + a*#P*(Op2)<br>  + a*#T(Op4+Op5+Op7)<br>Worst Case:<br>Op1+ a*#P*(Op2) + a*#T(Op4+Op5+Op7)<br>  +#P*#T*#B (Op6+Op8) |





## Complexity of Parallel Coordinates

| Task | Code |
|---|---|
| Task 1:<br>The Algorithm | FOR each Patient (#P)<br>  Highlight (Op4) /*Select every patient*/<br>  /* Read Shape of one patient's data */<br>  Read Shape & Decide(Op1)<br>  IF no clear Relation THEN<br>    FOR MANY Data Bundles (a*#P)<br>      Highlight (Op4) /*Select every patient*/<br>      Read Line (b*Op5)<br>      Compare Line (b*Op7)<br>  IF still no clear Relation THEN<br>    FOR MANY Lines (a*#T*#P)<br>      Highlight (Op4)<br>      Read Line Shape (b*Op5)<br>      Compare Line (b*Op7)<br>  IF still no clear Relation THEN<br>    FOR EACH Data Point (#P*#T*#B)<br>      Highlight (Op4)<br>      Read Data Point (Op6)<br>      Compare Data Point (Op8) |
| Complexity for Task 1 | Best Case:<br>#P*(Op4+Op1)<br>Middle Cases:<br>#P*(Op4+Op1) + a*#P*(Op4+b*Op5+b*Op7)<br>or: #P*(Op4+Op1) + a*#P*(Op4+b*Op5+b*Op7)<br>  + a*#T*#P*(Op4+b*Op5+b*Op7)<br>Worst Case:<br>#P*(Op4+Op1) + a*#P*(Op4+b*Op5+b*Op7)<br>  + a*#T*#P*(Op4+b*Op5+b*Op7)<br>  + #P*#T*#B*(Op4+Op6+Op8) |



[Lanzenberger 2003]

## Concept Testing

Comparative study (Controlled experiment) with 22 participants (35 participants for each visualization method), 2 examples

| Age | # of Sub. |
|---|---|
| - 20 | 1 |
| 21 - 25 | 6 |
| 26 - 30 | 6 |
| 31 - 35 | 4 |
| 36 - 40 | 1 |
| 41 - 45 | 3 |
| 45 - | 1 |
| Total | 22 |

Research questions:

- Are the users able to find information at the first glance?
- Are the users able to find the crucial information?
- Which visualization supports the creation of hypotheses?

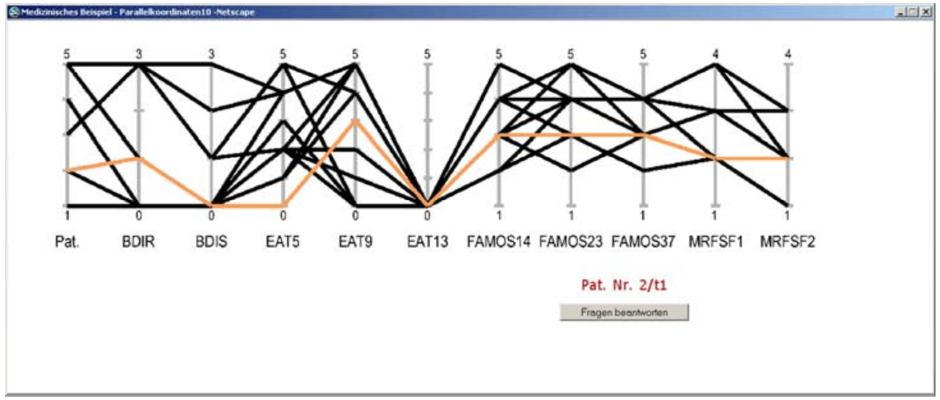
Evaluation:

- Time measurements, questionnaires
- Classification of strategies (categories)
- Expert defined 'Key Statements'



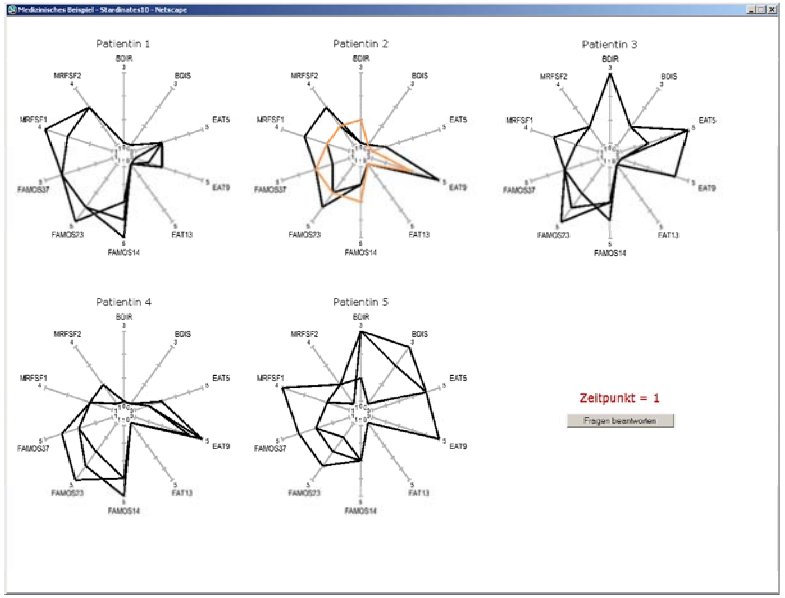
[Lanzenberger 2003]

## Visualization Method: Parallel Coordinates



[Lanzenberger 2003]

# Visualization Method: Stardinates

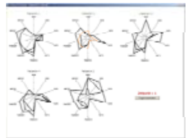


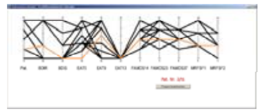



[Lanzenberger, et al. 2005]

# Evaluation Results: Time Measurement







| Task | Stardinates Average | Stardinates Minimum | Stardinates Maximum | Parallel Coordinates Average | Parallel Coordinates Minimum | Parallel Coordinates Maximum |
|---|---|---|---|---|---|---|
| Viewing 1.Example | 2:28 | 0:10 | 7:56 | 4:17 | 0:11 | 11:51 |
| 1. Answering | 9:29 | 2:14 | 18:32 | 9:09 | 2:43 | 22:16 |
| Viewing 1.Example (Correct answer) | 2:31 | 0:10 | 7:56 | 3:37 | 0:11 | 9:58 |
| 1. Answering (Correct answer) | 8:57 | 2:14 | 18:25 | 11:32 | 4:24 | 22:16 |
| Viewing 1.Example (Incorrect answer) | 2:09 | 0:38 | 5:01 | 5:28 | 0:24 | 11:51 |
| 1. Answering (Incorrect answer) | 12:50 | 8:13 | 18:32 | 4:49 | 2:53 | 8:30 |
| Total time 1. Example | 11:57 | 5:57 | 20:31 | 13:26 | 3:24 | 26:40 |
| Viewing 2.Example | 4:58 | 0:14 | 16:56 | 3:57 | 0:08 | 16:43 |
| 2. Answering | 20:41 | 6:47 | 46:02 | 16:08 | 3:03 | 41:34 |
| Total time 2. Example | 25:39 | 7:46 | 53:50 | 20:05 | 4:43 | 48:33 |
| Total time (1. + 2. Example) | 37:36 | 16:37 | 1:14:21 | 33:31 | 12:42 | 1:04:11 |



[Lanzenberger, et al. 2005]

# **Evaluation** Example 1 **- Aircraft Collision**

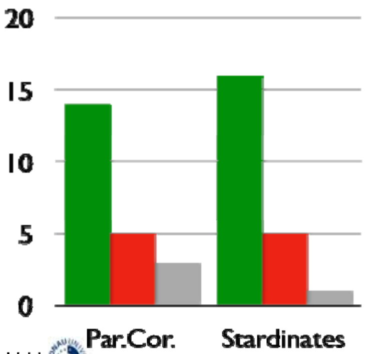


Questions:

- Ist deiner Meinung nach eine Kollision aufgetreten?
  - Wenn ja, welche Flugzeuge waren beteiligt?
- Bei welcher Graphik (welchen Graphiken) konntest du etwas ablesen? Wenn ja, was hast du dort abgelesen?
- Welche Probleme / Schwierigkeiten hattest du bei der Interpretation?

Results:

- Parallel Coordinates:
  - 63.6% (14 subjects) **correct** answer,
  - 22.7% (5 subjects) **incorrect** answer,
  - 13.6% (3 subject) no answer.
- Stardinates:
  - 72.7% (16 subjects) **correct** answer,
  - 22.7% (5 subjects) **incorrect** answer,
  - 4.5% (1 subject) no answer.
- Two strategies with the Stardinates:
  - Compare triangles (shapes)
  - Read exact values






[Lanzenberger, et al. 2005]

# **Evaluation** Example 2 **– Psychotherapeut. Data**

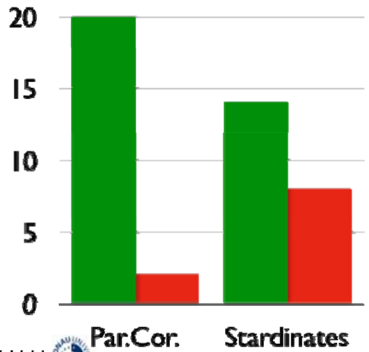


Questions:

- Gibt es Aussagen, die auf den ersten Blick auffallen?
- Bei welcher Graphik (welchen Graphiken) konntest du etwas ablesen? Wenn ja, was hast du dort abgelesen?
- Welche Probleme / Schwierigkeiten hattest du bei der Interpretation?

Results - 1. Question:

- Parallel Coordinates:
  - 90.9 % (20 subjects) found **information at the first glance**
- Stardinates:
  - 63.6 % (14 subjects) found **information at the first glance**



[Lanzenberger, et al. 2005]

# Evaluation Results: Key Statements

| Key Statement | Stardinates # of Sub. | Stardinates % of Sub. | Parallel Coordinates # of Sub. | Parallel Coordinates % of Sub. |
|---|---|---|---|---|
| Patients do not feel sick after eating. | 12 | 54.55% | 16 | 72.72% |
| Pat. 1: good starting basis. | 15 | 68.18% | 5 | 22.73% |
| Pat. 2: unstable. | 6 | 27.27% | 2 | 9.09% |
| Pat. 3: contradicting answers. | 3 | 13.64% | 3 | 13.64% |
| Pat. 4: positive progress in therapy. Cares more about herself. | 3 | 13.64% | 0 | 0% |
| Pat. 5: significantly positive progress in therapy between second and third time point. | 12 | 54.55% | 3 | 13.64% |



[Lanzenberger, et al. 2005]

# Evaluation Results: Key Statements

## Are the users able to find the crucial information?

Although unfamiliar with psychotherapeutic data, users were able to find crucial insights.

## Statistical analysis:

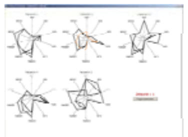
Stardinates were significantly better for finding crucial information (represented by the key statements).
Mean number of key statements:
- 2.32 with the Stardinates,
- 1.32 with the Parallel Coordinates.
- ($t$=2.687, $df$=21, level of significance: 5%).

Parallel Coordinates showed a high result in the first category, which is based on one dimension (EAT13) only,
but did not perform significantly better



[Lanzenberger, et al. 2005]

# Evaluation Results: Classification of Strategies



| Category | Stardinates # of Sub. | Stardinates % of Sub. | Parallel Coordinates # of Sub. | Parallel Coordinates % of Sub. |
|---|---|---|---|---|
| 1: Comparing Patients | 15 | 68.18% | 10 | 45.45% |
| 2: Overview | 15 | 68.18% | 5 | 22.73% |
| 3: Changes over Time | 15 | 68.18% | 10 | 45.45% |
| 4. Examining Single Axes | 14 | 63.64% | 19 | 86.36% |
| 5: General Conclusions | 5 | 22.73% | 1 | 4.55% |
| 6: Causal dependency | 8 | 36.37% | 1 | 4.55% |
| 7: None | 0 | 0% | 1 | 4.55% |



[Lanzenberger, et al. 2005]

# Evaluation Results: Classification of Strategies

## Which visualization supports the creation of hypotheses?

Subjects produced significantly more statements with the Stardinates than with the Parallel Coordinates.
They did not need more time when using the Stardinates.

## Statistical Analysis:

Mean number of statements
3.27 with the Stardinates and
2.14 with the Parallel Coordinates
($t$=3.504, $df$=21, level of significance: 5%)

## Content :: Evaluation & Usability

Information Visualization Evaluation

Evaluation in Practice
- in2vis
- DisCō
- Stardinates

Crucial InfoVis Challenges



[Chen 2005]

## Crucial InfoVis Challenges: Top 10 Problems

| Problems | Category |
|---|---|
| Usability<br>Understanding perceptual-cognitive tasks<br>Prior knowledge<br>Education and training | user-centered perspective |
| Quality measures<br>Scalability<br>Aesthetics | technical challenges |
| Paradigm shift from structures to dynamics<br>Causality, visual inference, and predictions<br>Knowledge domain visualization | disciplinary challenges |



## Top 10 Problems: Usability

[Chen 2005]

user-centered perspective

Relevant for researchers and developers
- compare
  Spotfire (http://www.spotfire.com) and
  Inspire (http://in-spire.pnl.gov)

InfoVis is growing much faster than its usability research

Lack of low-cost or open source InfoVis tools

Usability studies need to address critical details specific to InfoVis
- e.g., recognition of the intended patterns or
  interaction with possible cognitive paths in a network visualization



## Top 10 Problems: Perceptual-cognitive tasks

[Chen 2005]

user-centered perspective

Evaluation of the usefulness of InfoVis components is done:
- Identifying & decoding visualized objects, preattentive perception

But evaluation of high-level user tasks is needed:
- Browsing, searching, recognition of clusters, identification of trends, discovery of previously unknown connections, insightful discovery

## Top 10 Problems: Prior knowledge



user-centered perspective

Two types of prior knowledge:

- the **knowledge of how to operate the device**, such as a telescope, a microscope, or, in our case, an InfoVis system, and
- the **domain knowledge** of how to interpret the content

Good usability and utility can reduce the dependence on the first type of prior knowledge

Distinguish perception, cognition and learning



## Top 10 Problems: Education and training

[Chen 2005]

user-centered perspective

Learn and share various principles and skills of visual communication and semiotics

Language of InfoVis must become comprehensible

Potential beneficiaries outside the immediate field of InfoVis to see the value and how it might contribute in practice



## Top 10 Problems: Quality measures

[Chen 2005]

technical challenges

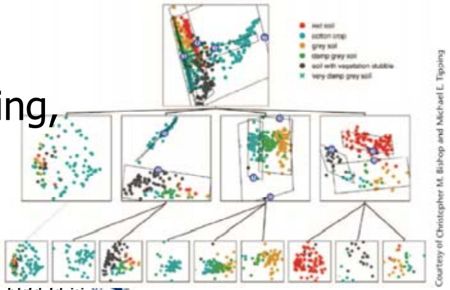
Quantifiable measures of quality, benchmarks are missing

Simplifies development and evaluation of algorithms

Answer key questions such as:

- To what extent does an InfoVis design represent the underlying data faithfully and efficiently?
- To what extent does it preserve intrinsic properties of the underlying phenomenon?

Integrating machine learning for topic detection, trend tracking, adaptive information filtering, and detecting concept drifts in streaming data





## Top 10 Problems: Scalability (1)

[Chen 2005]

technical challenges

Long-lasting challenge for InfoVis

Unlike to scientific visualization, supercomputers have not been the primary source of data suppliers

Parallel computing and other high-performance computing techniques are not used

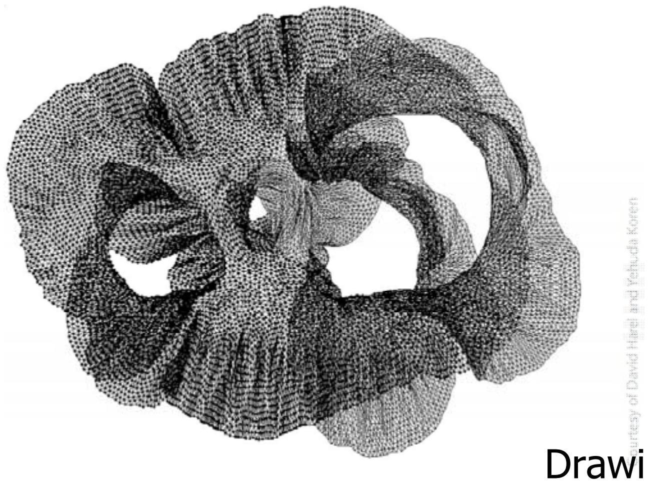
Visualization of data streams and the urgency to understand its contents

# Top 10 Problems: Scalability (2)

[Chen 2005]



technical challenges




Drawing a 15,606-vertex and 45,878-edge graph within a matter of seconds

Interaction?



# Top 10 Problems: Aesthetics

[Chen 2005], Graph: [Rester and Pohl 2005]

technical challenges

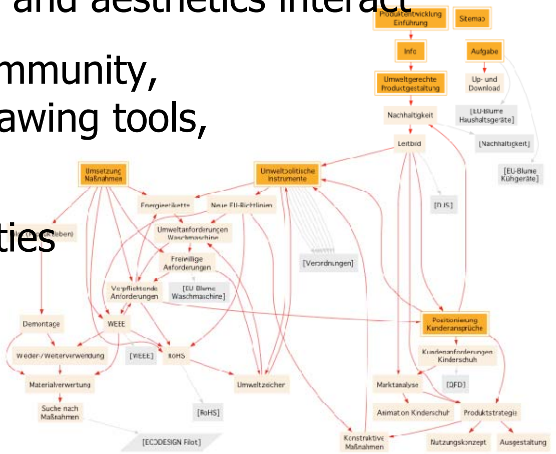
Insights, not just pretty pictures

Goal is to enhance utility

Understand how insights and aesthetics interact

Active graph-drawing community, e.g., automatic graph-drawing tools,

But often focuses on graph-theoretical properties and rarely involves the semantics associated with the data





# Top 10 Problems: Paradigm shift

[Chen 2005]

disciplinary challenges

In 1990s most InfoVis tools dealt with structures such as cone tree, treemap, and hyperbolic views

Paradigm shifts to dynamic visualization

Changes over time and thematic trends

Draw users' attention to changes and trends: built-in trend detection mechanisms

Collaboration with data mining and artificial intelligence communities



# Top 10 Problems: *Causality, visual inference, & predictions*

[Chen 2005]

disciplinary challenges

Visual thinking, reasoning, and analytics:
InfoVis powerful medium for finding causality, forming hypotheses, and assessing available evidence

Tufte's re-visualization of
the data from the challenger space shuttle disaster and Snow's map of cholera deaths

Challenge is to resolve conflicting evidence and suppress background noises

Freely interact with raw data as well as with its visualizations to find causality

Potential areas: evidence-based medicine, technology forecasting, collaborative recommendation, intelligence analysis, and patent examination

## Top 10 Problems: Knowledge domain visualization

[Chen 2005]

disciplinary challenges

Difference between knowledge and information can be seen in terms of the role of social construction

Knowledge involves interpretations and decisions

Interacting with InfoVis can be more than retrieving individual items of information

Entire body of domain knowledge is subject to the rendering

The KDViz problem is rich in detail, large in scale, extensive in duration, and widespread in scope



## References



**[Belle, et al.]** Sebastian Kay Belle, Daniela Oelke, Sonja Oettl, and Mike Sips: Exploration of the local distribution of major ethnic groups in the USA, 2006. http://www.cs.umd.edu/hcil/InfovisRepository/contest-2006/files/Belle.pdf (checked online 6. Jan. 2009).

**[Card, et al. 1983]** Stuart Card, Thomas P. Moran, and Allen Newell:The Psychology of Human-Computer Interaction. Hillsdale, NJ: Erlbaum, 1983.

**[Chen 2005 ]** Chaomei Chen, Top 10 unsolved information visualization problems. IEEE Computer Graphics and Applications, 25 (4), pp.12-16, 2005. http://www.pages.drexel.edu/~cc345/papers/cga2005.pdf (checked online 6. Jan. 2009).

**[Gärtner, et al. 2002]** Johannes Gärtner, Silvia Miksch, Stefan Carl-McGrath: ViCo: A Metric for the Complexity of Information Visualizations, in Second International Conference on Theory and Application of Diagrams (Diagrams 2002), Springer, Berlin, pp. 249-263, 2002.

**[in2vis]** Silvia Miksch, Klaus Hinum, Margit Pohl, Markus Rester, Sylvia Wiltner, Christian Popow, and Susanne Ohmann: Interactive Information Visualization (in2vis), Exploring and Supporting Human Reasoning Processes, http://ieg.ifs.tuwien.ac.at/projects/in2vis/ (checked online 6. Jan. 2009).

**[Lanzenberger 2003]** Monika Lanzenberger: The Interactive Stardinates - An Information Visualization Technique Applied in a Multiple View System , Vienna University of Technology, 2003.

[**Lanzenberger, et al. 2005**] Monika Lanzenberger, Silvia Miksch, and Margit Pohl: Exploring Highly Structured Data - A Comparative Study of Stardinates and Parallel Coordinates. In Proceedings of the IV05, 9th International Conference on Information Visualisation, July 6-8, 2005, London, UK, IEEE Computer Science Society, pp. 312-320, 2005.



## References



**[Nielsen 1994]** Jakob Nielsen: Heuristic Evaluation, chapter 2, pp. 25–62. John Wiley & Sons, Inc., New York, 1994.

**[Plaisant 2004]** Catherine Plaisant: The challenge of information visualization evaluation. In M. F. Costabile (ed) Proceedings of the working conference on Advanced visual interfaces, pp. 109–116. ACM Press, 2004.

**[Rester and Pohl 2005]** Markus Rester and Margit Pohl: Ecodesign - an Online University Course for Sustainable Product Design, World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA), Montreal, Canada, In Proceedings of of ED-MEDIA, pp. 316 - 323, 2005.

**[Smuc et al., 2008]** Smuc, M.; Mayr, E.; Lammarsch, T.; Bertone, A.; Aigner, W.; Risku, H. & Miksch, S.: Visualizations at First Sight: Do Insights Require Traininig?, *Proceedings of 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society (USAB 2008)*, p. 261-280, Springer, November, 2008.

**[Smuc et al., 2009]** Smuc, M.; Mayr, E.; Lammarsch, T.; Aigner, W.; Miksch, S. & Gärtner, J.: To Score or Not to Score? Tripling Insights for Participatory Design, *IEEE Computer Graphics and Applications*, Vol. 29, No. 3, p. 29-38, IEEE Computer Society Press, May, 2009.

**[Rester, et al. 2006]** Markus Rester, Margit Pohl, Klaus Hinum, Silvia Miksch, Christian Popow, Susanne Ohmann, and Slaven Banovic: Methods for the Evaluation of an Interactive InfoVis Tool Supporting Exploratory Reasoning Processes, BELIV '06: Proceedings of the 2006 conference on BEyond time and errors: novel evaLuation methods for Information Visualization, Venice, Italy, ACM Press, New York, NY, USA, pp. 32-37, 2006.



## Thanks to ...

Monika Lanzenberger, TUW
Markus Rester, TUW
Margit Pohl, TUW

Alessio Bertone, DUK
Tim Lammarsch, DUK
Eva Mayr, DUK
Michael Smuc, DUK

for making nice slides of previous classes available.