# EVALUATING VISUALIZATIONS

Petra Isenberg

informatics mathematics

# RECAP

SO FAR:

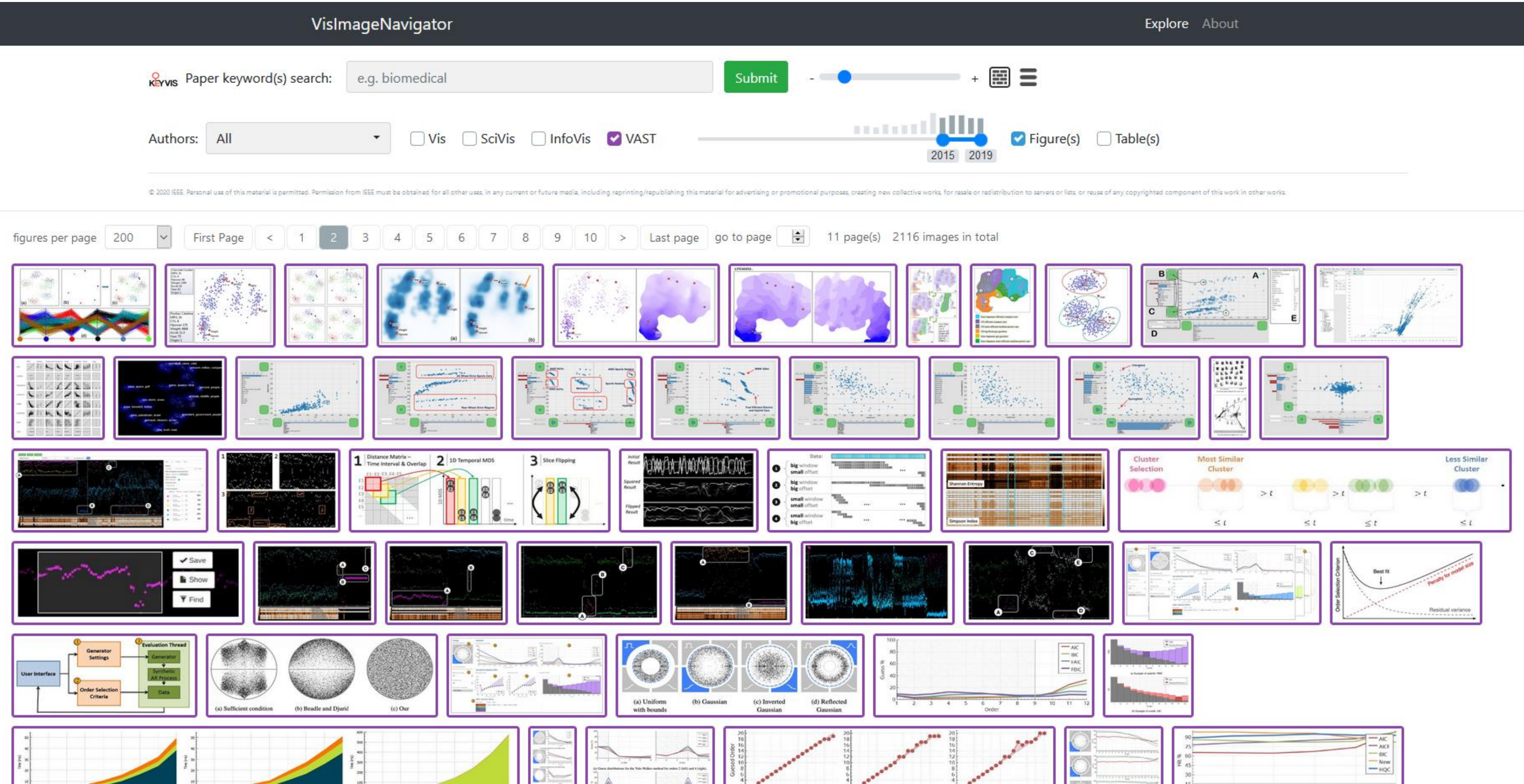
- How to find & prepare data
- How to create visualizations

TODAY:

- How do we know a visualization has value?

MANY VISUALIZATIONS / VISUAL ANALYTICS SYSTEMS, TOOLS, TECHNIQUES EXIST

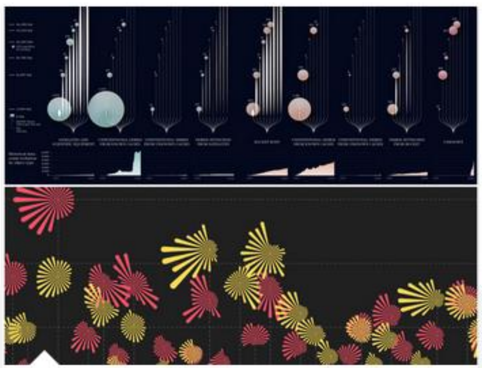
# FROM RESEARCH...



# PRACTITIONERS…

KANTAR
Information is Beautiful
Awards 2019

About News Awards Entry Showcase Sponsor



10 MONTHS AGO

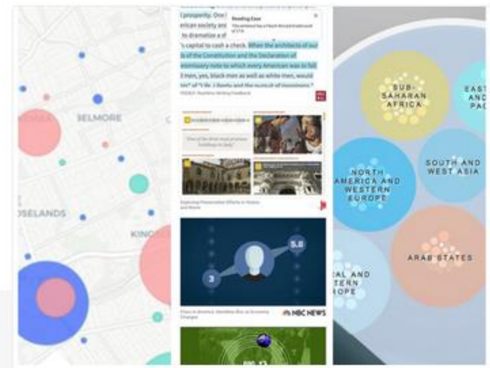
### Best of the Web: January 2020

There's only one cure for the depths of January... bright, gorgeous, dazzling data. Scroll down to see which vizzes got all the likes, shares, slow claps, and nods of approval in the past... →


11 MONTHS AGO

### Interview with 2019's Best Non-English-Language Winner

Our Best Non-English-Language category presents us with an opportunity to celebrate vizzes featuring topics and commentary outside of the English-speaking world - i.e., most of the global... →



11 MONTHS AGO

### 6 Years of Outstanding Outfits

The Information is Beautiful Awards have been celebrating extraordinary outlets and astounding studios since 2014... that's 6 years of data visual goodness from some stunningly creative &... →

11 MONTHS AGO

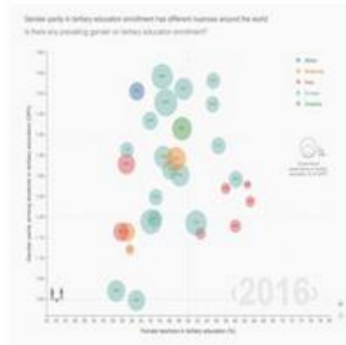
11 MONTHS AGO

# STUDENTS…

## Finished Project Gallery

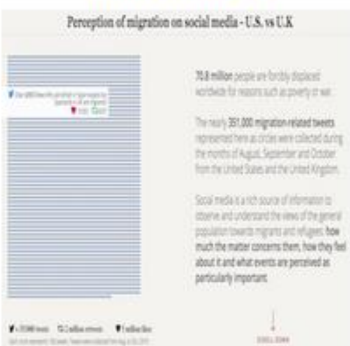

Twitter Influencers


Depression Among Students




Gendergap in Education

Avian Flu Outbreak




Public Opinion on Migration

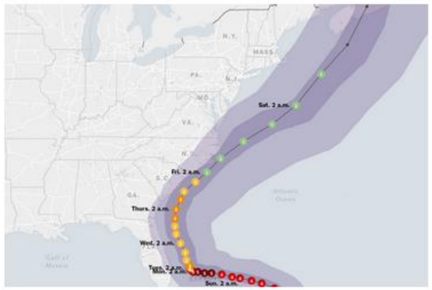
Refugee Integration & Resettlement

Illicit drug Use

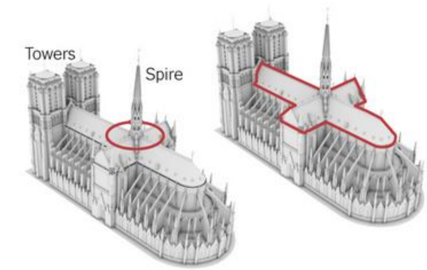
Income Inequality

# ORGANIZATIONS….

The New York Times — 2019: The Year in Visual Stories and Graphics

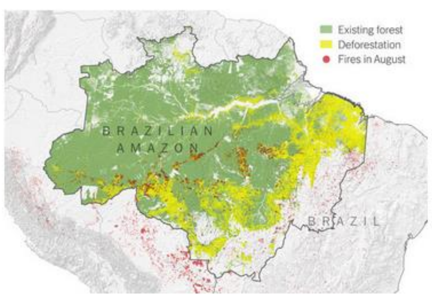


Live Maps: Tracking Hurricane Dorian's Path SEPT. 6, 2019

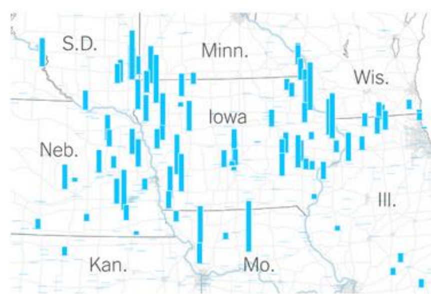



How the Notre-Dame Cathedral Fire Spread APRIL 15, 2019


California Wildfires: How PG&E Ignored Risks in Favor of Profits MARCH 18, 2019




What Satellite Imagery Tells Us About the Amazon Rain Forest Fires AUG. 24, 2019




Rising Waters: See How Quickly the Midwest Flooded MARCH 19, 2019


Where Brazilians Live in High-Risk Areas Downhill From Mining Dams FEB. 14, 2019


# DO THESE VISUALIZATIONS HAVE VALUE?

# EVALUATION

an assessment of a visualization or its context of use, including:

- a multitude of methods / methodologies
- at any point in the development cycle

→ not just with participants

# EVALUATING VISUALIZATIONS IS DIFFICULT

# EVALUATION

…is difficult because:

- Visualizations are often part of a creative activity
  - Make hypotheses, look for patterns, refine hypotheses
- Insight/discovery can happen any time – even after tool use
- Data analysis is often collaborative
- Insight/discovery are not predictable and often rare

# YOU SHOULD ASK YOURSELF

WHY DO YOU WANT TO EVALUATE SOMETHING?

WHEN SHOULD YOU EVALUATE SOMETHING?

WHAT DO YOU WANT TO KNOW?

HOW CAN YOU EVALUATE THAT?

# QUESTION 1

WHY DO YOU WANT TO EVALUATE?

# DEPENDS ON WHO YOU ARE

| Researcher | Developer | End User |
| --- | --- | --- |
| • Did I have a good idea / hypothesis?<br>• What makes people use my idea/system in a certain way?<br>• What are its limits? | • What should we build?<br>• Which option should we focus on?<br>• Do users like our product?<br>• How well does the product work? | • What should I use?<br>• When should I use something?<br>• What product should I buy? |

# SEVEN SCENARIOS...

# ConTour: Data-Driven Exploration of Multi-Relational Datasets for Drug Discovery

Christian Partl, Alexander Lex, Marc Streit, Hendrik Strobelt,
Anne-Mai Wassermann, Hanspeter Pfister and Dieter Schmalstieg

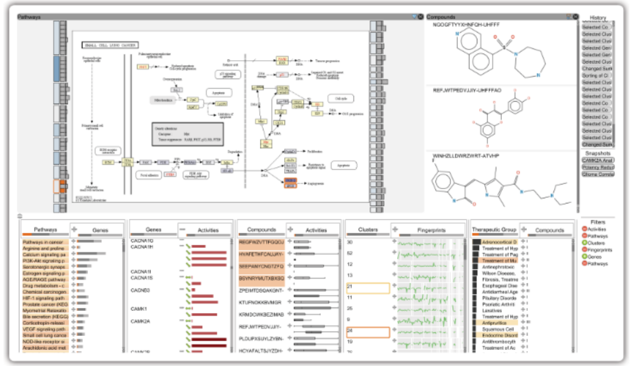


Fig. 1. ConTour shows a multitude of heterogeneous data items in several columns in the relationship view (bottom). The detail views display a selected pathway and selected chemical structures of compounds (top).

**Abstract**—Large scale data analysis is nowadays a crucial part of drug discovery. Biologists and chemists need to quickly explore and evaluate potentially effective yet safe compounds based on many datasets that are in relationship with each other. However, there is a lack of tools that support them in these processes. To remedy this, we developed ConTour, an interactive visual analytics technique that enables the exploration of these complex, multi-relational datasets. At its core ConTour lists all items of each dataset in a column. Relationships between the columns are revealed through interaction: selecting one or multiple items in one column highlights and re-sorts the items in other columns. Filters based on relationships enable drilling down into the large data space. To identify interesting items in the first place, ConTour employs advanced sorting strategies, including strategies based on connectivity strength and uniqueness, as well as sorting based on item attributes. ConTour also introduces interactive nesting of columns, a powerful method to show the related items of a child column for each item in the parent column. Within the columns, ConTour shows rich attribute data about the items as well as information about the connection strengths to other datasets. Finally, ConTour provides a number of detail views, which can show items from multiple datasets and their associated data at the same time. We demonstrate the utility of our system in case studies conducted with a team of chemical biologists, who investigate the effects of chemical compounds on cells and need to understand the underlying mechanisms.

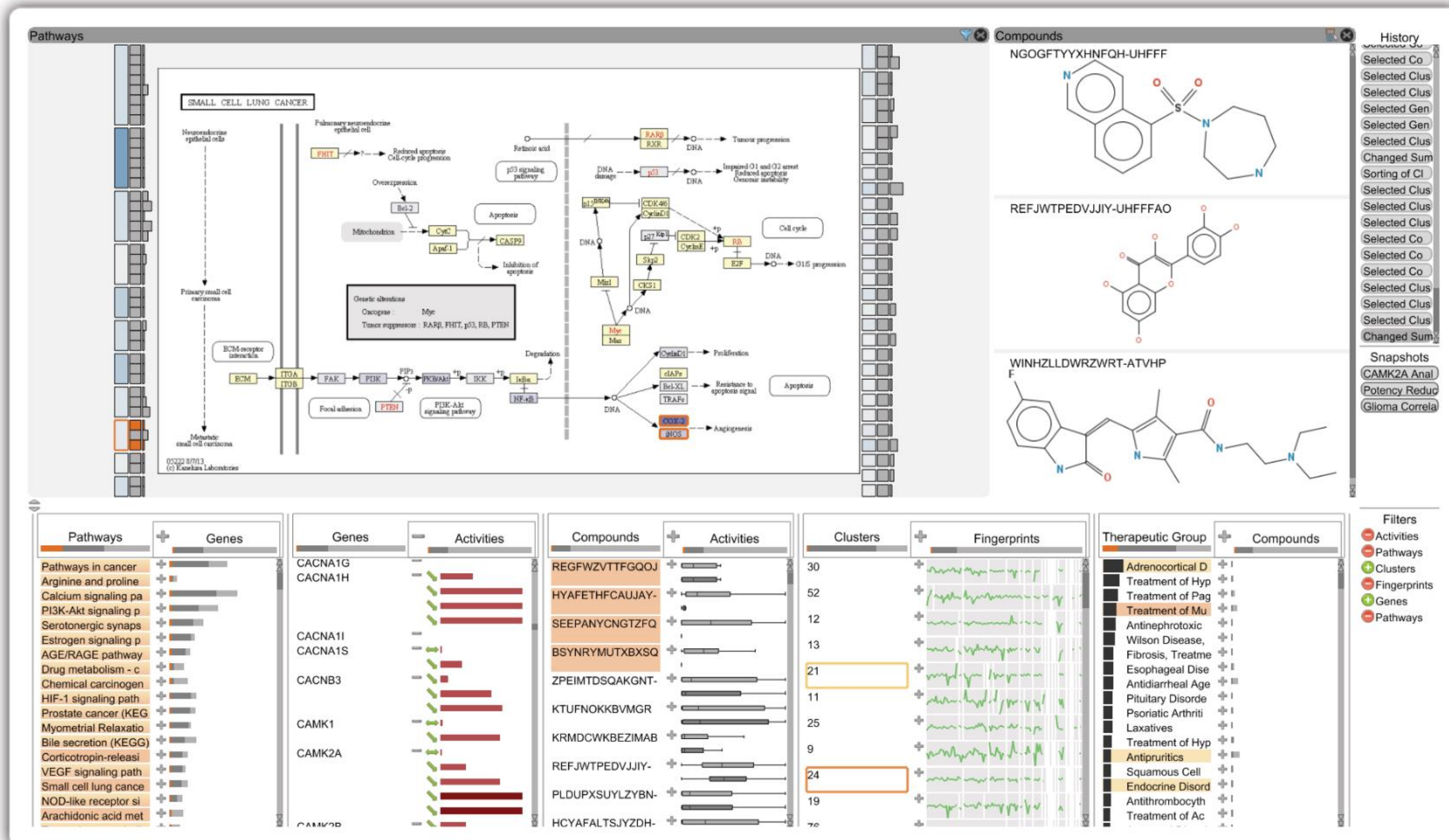
**Index Terms**—Multi-relational data, visual data analysis, drug discovery

## 1 Introduction

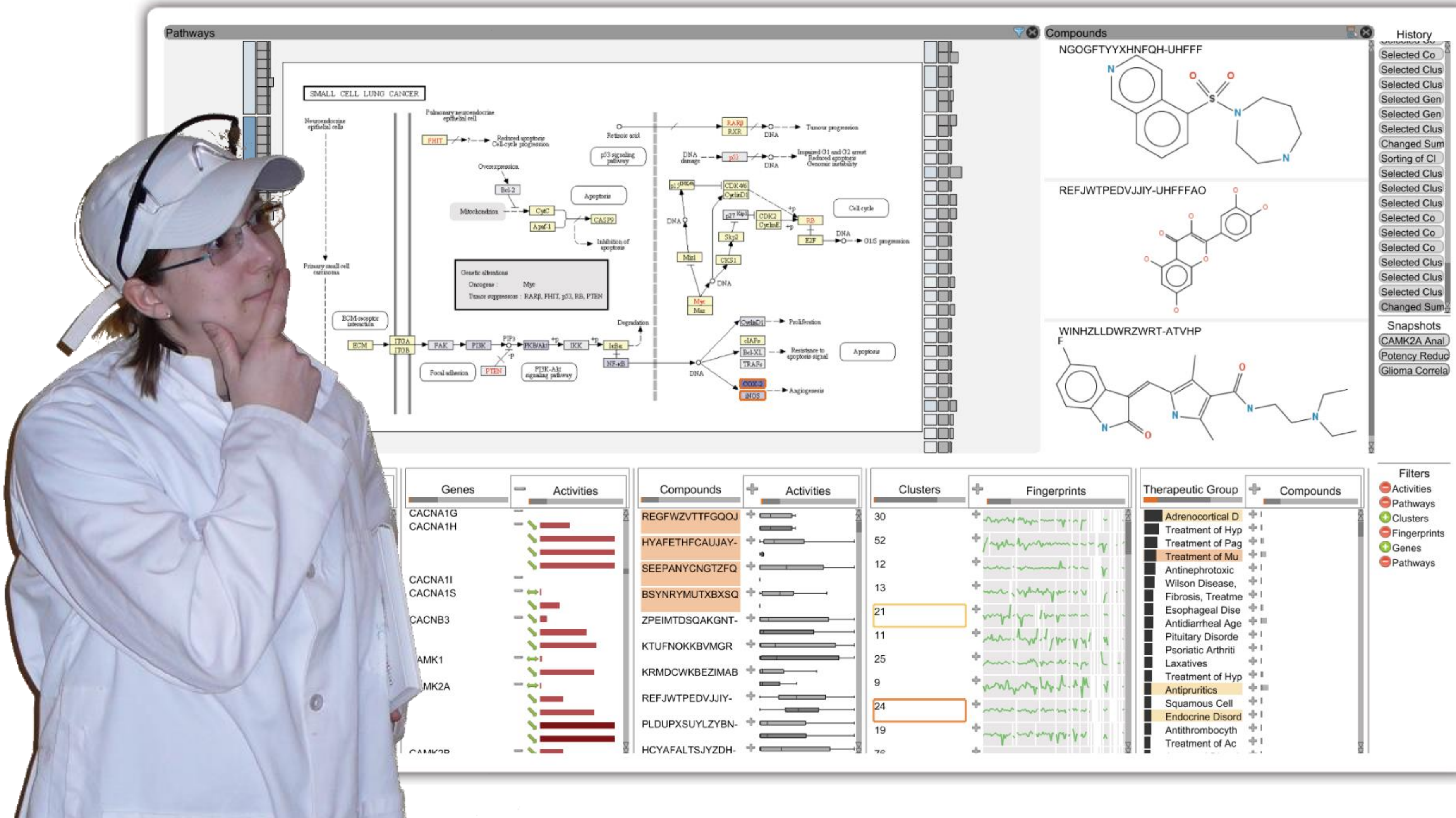
The need to explore multi-relational data is common in many domains. Answering questions such as whether relationships of particular entities across datasets exist or how strong or specific a relationship is, is important for a variety of applications. This is also true in drug discovery. Researchers want to learn whether there are chemical compounds, i.e., drugs or drug candidates, that modulate a specific biological process without influencing others, or want to see which drugs induce a characteristic change in a cell's phenotype. However, due to the complexity of the experimental data, the manifold interactions between compounds and cellular components, and the rich associated data, the


- *Christian Partl and Dieter Schmalstieg are with Graz University of Technology. E-mail: {partl, schmalstieg}@icg.tugraz.at.*
- *Alexander Lex, Hendrik Strobelt and Hanspeter Pfister are with Harvard University. E-mail: {alex, strobelt, pfister}@seas.harvard.edu.*
- *Marc Streit is with Johannes Kepler University Linz. E-mail: marc.streit@jku.at.*
- *Anne-Mai Wassermann is with Novartis Institutes for BioMedical Research. E-mail: anne.mai.wassermann@novartis.com.*


*Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014. Date of publication 11 Aug. 2014; date of current version 9 Nov. 2014. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.*
*Digital Object Identifier 10.1109/TVCG.2014.2346752*



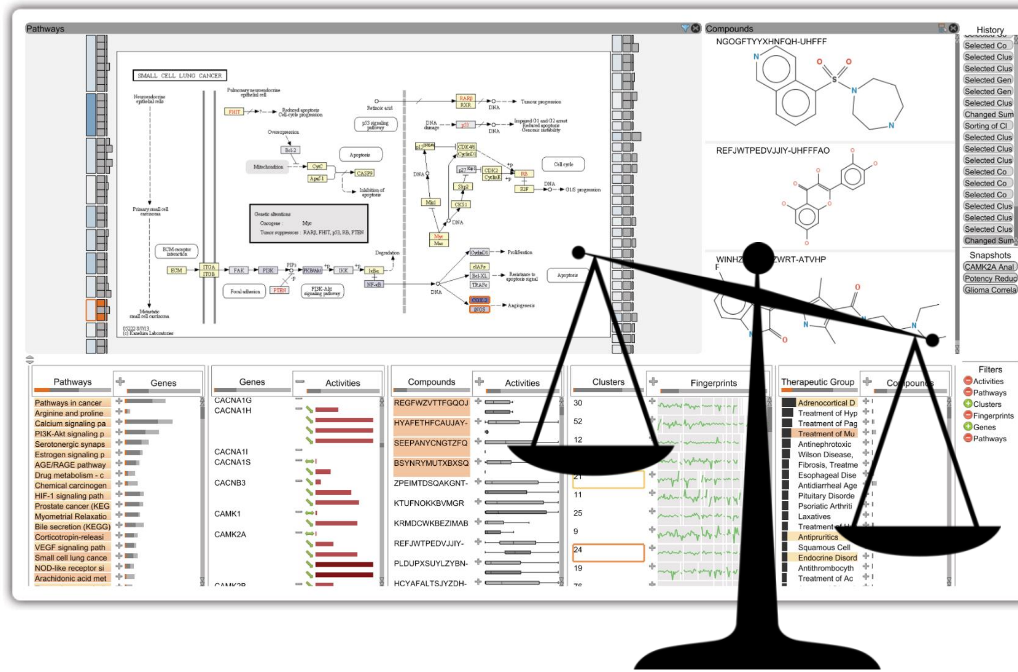
Pathways
Compounds
History
Snapshots
CAMK2A Anal
Potency Reduc
Glioma Correla
Filters
Activities
Pathways
Clusters
Fingerprints
Genes
Pathways
SMALL CELL LUNG CANCER
NGOGFTYYXHNFQH-UHFFF
REFJWTPEDVJJIY-UHFFFAO
WINHZLLDWRZWRT-ATVHP
Genes
Activities
Clusters
Therapeutic Group


Pathways
SMALL CELL LUNG CANCER
Compounds
NGOGFTYYXHNFQH-UHFFF
REFJWTPEDVJJIY-UHFFFAO
WINHZLLDWRZWRT-ATVHP
History
Snapshots
CAMK2A Anal
Potency Reduc
Glioma Correla
Genes
Activities
Compounds
Activities
Clusters
Fingerprints
Therapeutic Group
Compounds
Filters
Activities
Pathways
Clusters
Fingerprints
Genes
Pathways

# USER PERFORMANCE

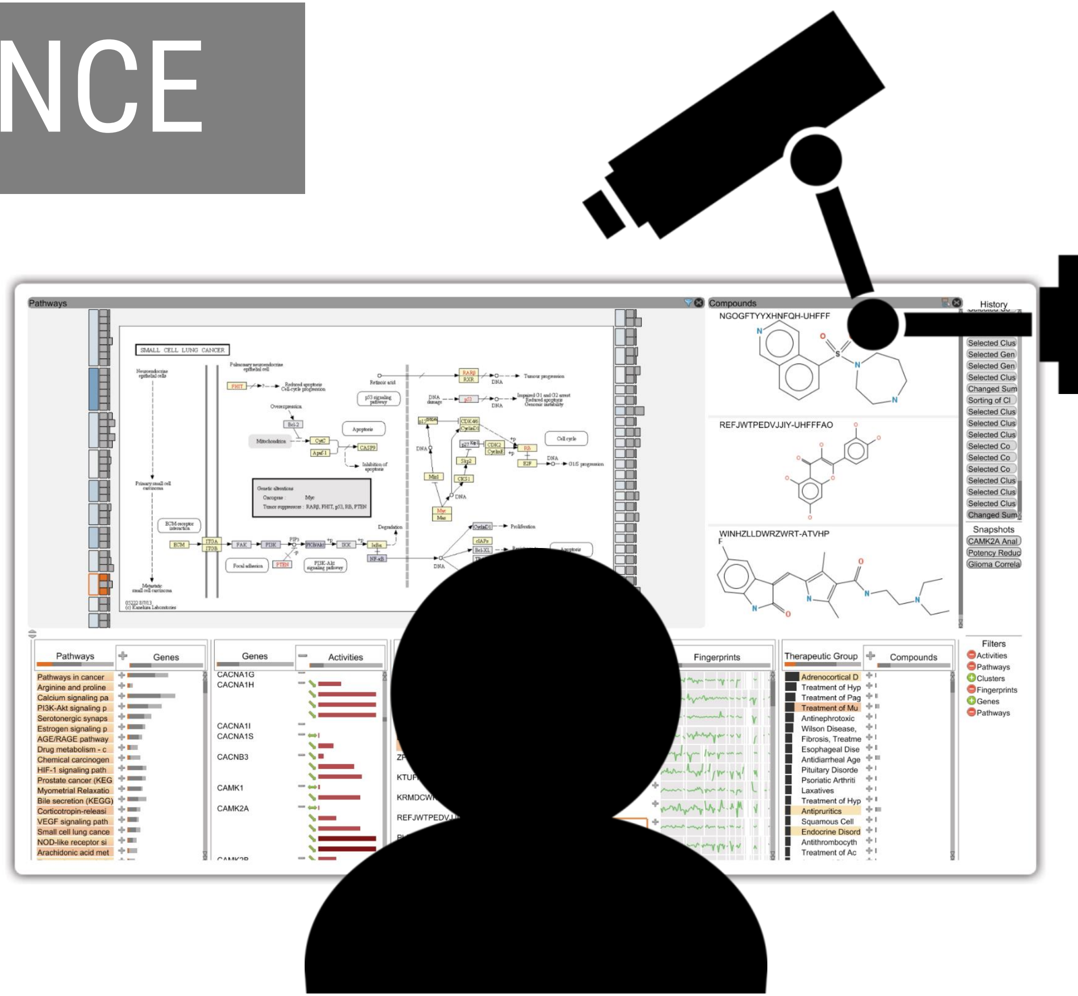
How fast or effectively can people perform a task?

*Given an item of type A, how quickly / correctly can a user find all items of type B that are directly or indirectly related?*




Typically you get a larger set of participants performing a specific task

# USER EXPERIENCE




How do people react to this tool?

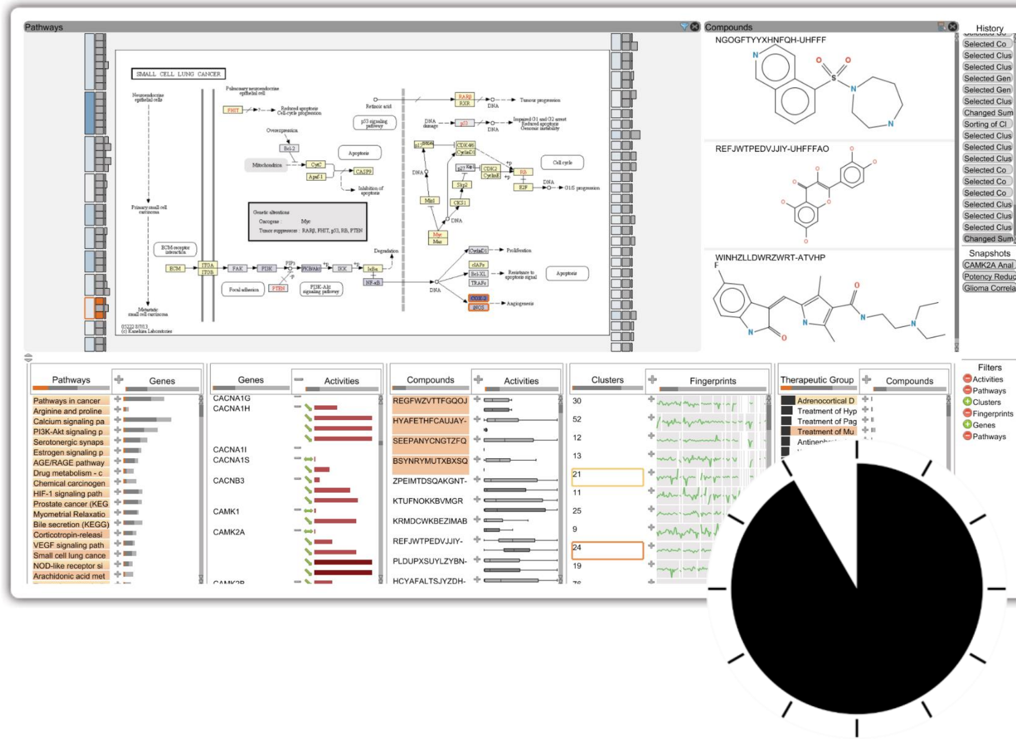
*Which features do users find most useful?*

Typically you get a smaller set of participants performing an in-depth analysis task

# ALGORITHMIC PERFORMANCE

How performant is the tool?

*How quickly can it render datasets of 100MB, 1GB, 1TB, …?*

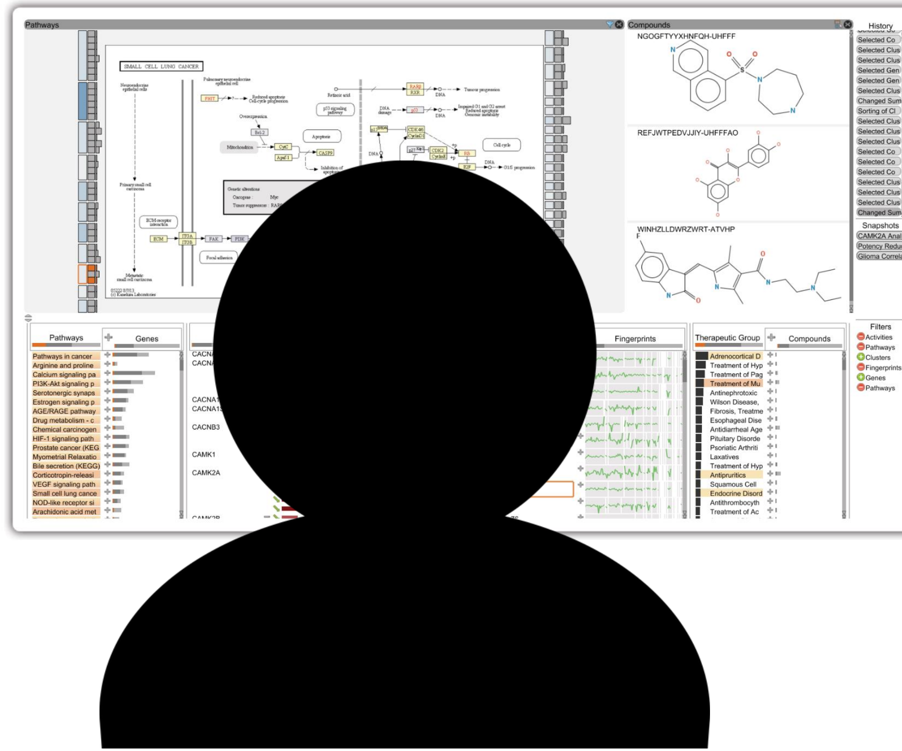



Typically you don't get any human participants

# WORK PRACTICES

How is / should the tool (be) used in practice?

- *Does it support users' questions?*
- *Does it support their tasks/goals?*
- *What visualizations are currently in use?*
- *Is it being adopted?*



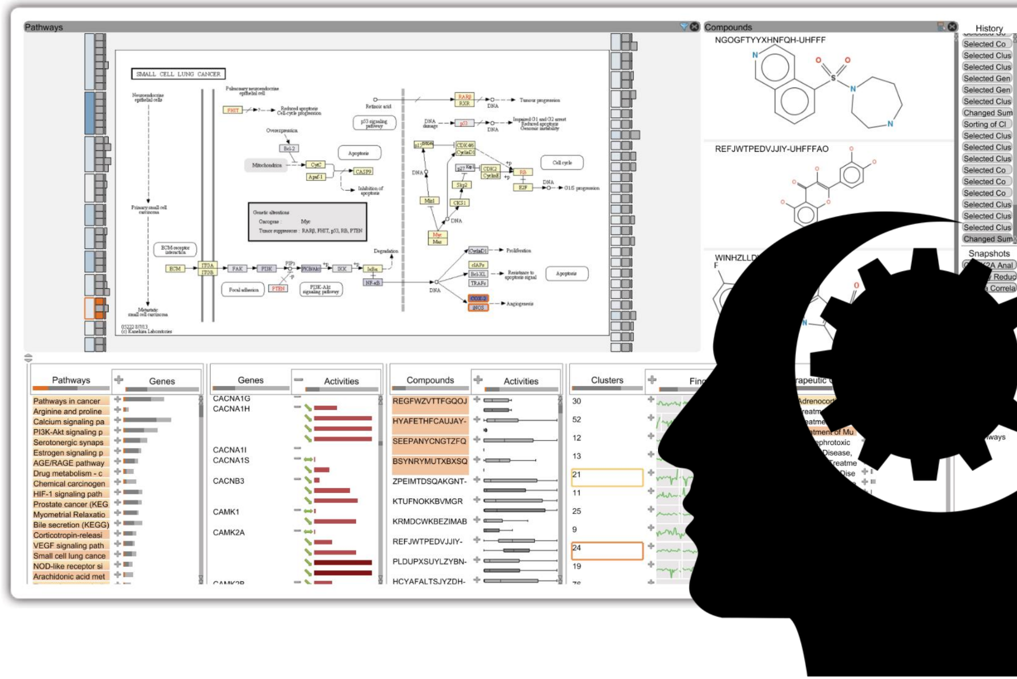

Typically you observe experts or end-users

# DATA ANALYSIS / REASONING

Does the tool lead to new discoveries or insights?

*Does the tool support*

- *the generation of hypotheses?*
- *the extraction of information?*
- *Decision making?*



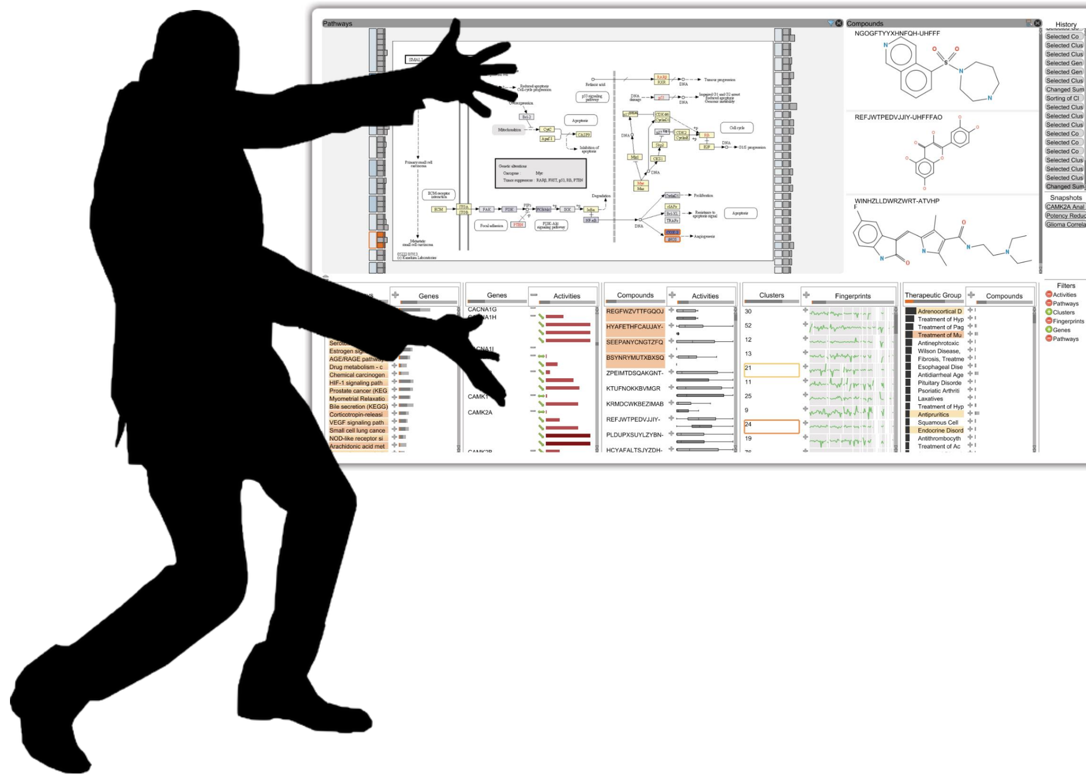

Typically you get a smaller set of participants performing an in-depth analysis task

# COMMUNICATION

Does the tool support communicating discoveries, insights, results?

*Can people learn with the visualizations?*

*Can the tool be used to explain a finding to third parties?*



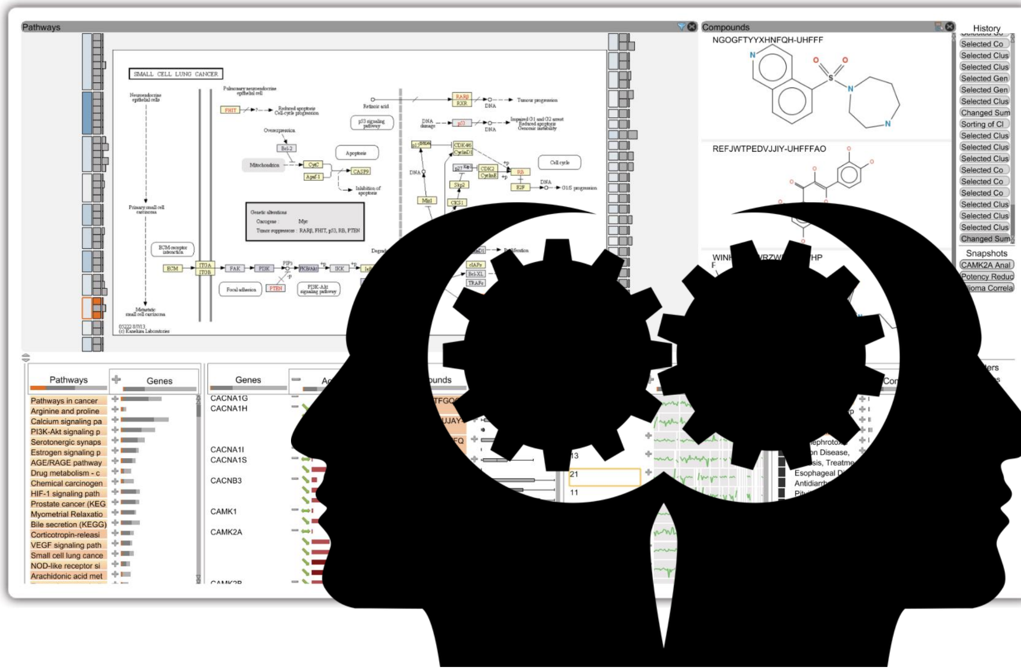

Typically you get a larger set of participants who answer questions

# COLLABORATION

Does the tool support collaborative analysis or decision making?

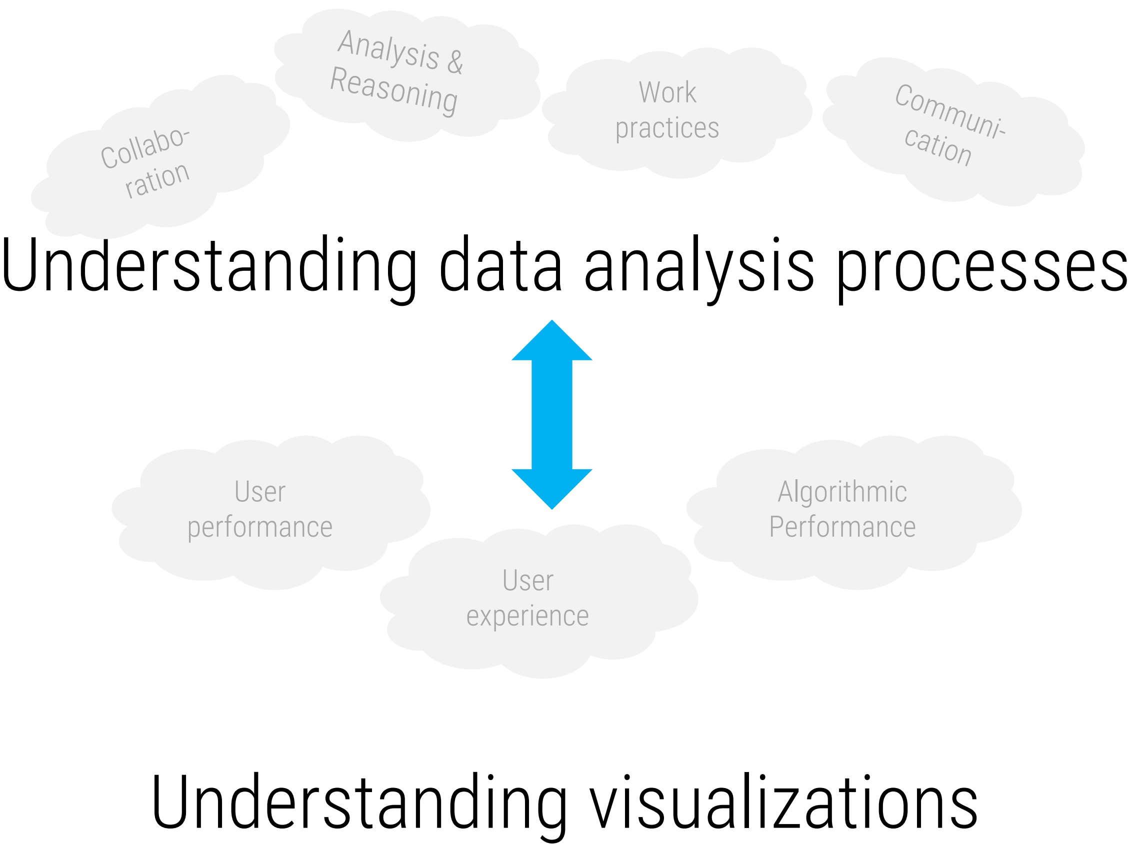
*Can analysis effectively work together?*

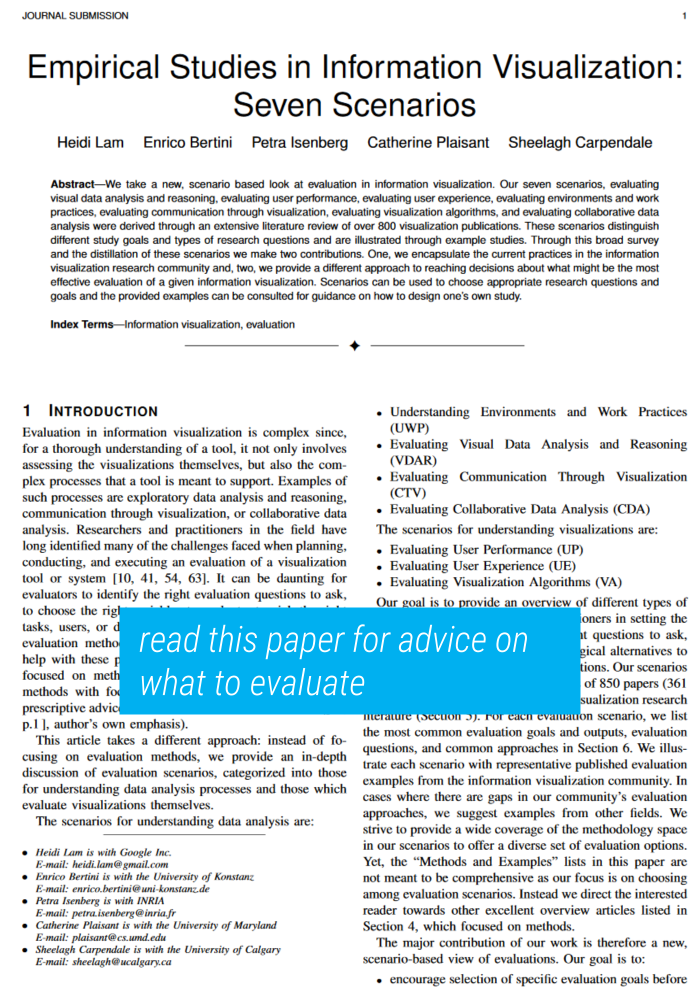
*Can users easily share and communicate about findings?*



Typically you get a multiple of participants performing an in-depth analysis task

# SEVEN SCENARIOS




# Empirical Studies in Information Visualization: Seven Scenarios

Heidi Lam Enrico Bertini Petra Isenberg Catherine Plaisant Sheelagh Carpendale

**Abstract**—We take a new, scenario based look at evaluation in information visualization. Our seven scenarios, evaluating visual data analysis and reasoning, evaluating user performance, evaluating user experience, evaluating environments and work practices, evaluating communication through visualization, evaluating visualization algorithms, and evaluating collaborative data analysis were derived through an extensive literature review of over 800 visualization publications. These scenarios distinguish different study goals and types of research questions and are illustrated through example studies. Through this broad survey and the distillation of these scenarios we make two contributions. One, we encapsulate the current practices in the information visualization research community and, two, we provide a different approach to reaching decisions about what might be the most effective evaluation of a given information visualization. Scenarios can be used to choose appropriate research questions and goals and the provided examples can be consulted for guidance on how to design one's own study.

**Index Terms**—Information visualization, evaluation

## 1 Introduction

Evaluation in information visualization is complex since, for a thorough understanding of a tool, it not only involves assessing the visualizations themselves, but also the complex processes that a tool is meant to support. Examples of such processes are exploratory data analysis and reasoning, communication through visualization, or collaborative data analysis. Researchers and practitioners in the field have long identified many of the challenges faced when planning, conducting, and executing an evaluation of a visualization tool or system [10, 41, 54, 63]. It can be daunting for evaluators to identify the right evaluation questions to ask, to choose the rig
tasks, users, or d
evaluation metho
help with these p
focused on meth
methods with fo
prescriptive advic
p.1 ], author's own emphasis).

This article takes a different approach: instead of focusing on evaluation methods, we provide an in-depth discussion of evaluation scenarios, categorized into those for understanding data analysis processes and those which evaluate visualizations themselves.

The scenarios for understanding data analysis are:

- Understanding Environments and Work Practices (UWP)
- Evaluating Visual Data Analysis and Reasoning (VDAR)
- Evaluating Communication Through Visualization (CTV)
- Evaluating Collaborative Data Analysis (CDA)

The scenarios for understanding visualizations are:

- Evaluating User Performance (UP)
- Evaluating User Experience (UE)
- Evaluating Visualization Algorithms (VA)

Our goal is to provide an overview of different types of
ioners in setting the
nt questions to ask,
gical alternatives to
tions. Our scenarios
of 850 papers (361
sualization research
literature (Section 5). For each evaluation scenario, we list the most common evaluation goals and outputs, evaluation questions, and common approaches in Section 6. We illustrate each scenario with representative published evaluation examples from the information visualization community. In cases where there are gaps in our community's evaluation approaches, we suggest examples from other fields. We strive to provide a wide coverage of the methodology space in our scenarios to offer a diverse set of evaluation options. Yet, the "Methods and Examples" lists in this paper are not meant to be comprehensive as our focus is on choosing among evaluation scenarios. Instead we direct the interested reader towards other excellent overview articles listed in Section 4, which focused on methods.

The major contribution of our work is therefore a new, scenario-based view of evaluations. Our goal is to:

- encourage selection of specific evaluation goals before


- *Heidi Lam is with Google Inc.*
  *E-mail: heidi.lam@gmail.com*
- *Enrico Bertini is with the University of Konstanz*
  *E-mail: enrico.bertini@uni-konstanz.de*
- *Petra Isenberg is with INRIA*
  *E-mail: petra.isenberg@inria.fr*
- *Catherine Plaisant is with the University of Maryland*
  *E-mail: plaisant@cs.umd.edu*
- *Sheelagh Carpendale is with the University of Calgary*
  *E-mail: sheelagh@ucalgary.ca*


*read this paper for advice on what to evaluate*

Updated information for another visualization sub-field

# A Systematic Review on the Practice of Evaluating Visualization

Tobias Isenberg, *Senior Member, IEEE*, Petra Isenberg, Jian Chen, *Member, IEEE*, Michael Sedlmair, *Member, IEEE*, and Torsten Möller, *Senior Member, IEEE*

**Abstract**—We present an assessment of the state and historic development of evaluation practices as reported in papers published at the IEEE Visualization conference. Our goal is to reflect on a meta-level about evaluation in our community through a systematic understanding of the characteristics and goals of presented evaluations. For this purpose we conducted a systematic review of ten years of evaluations in the published papers using and extending a coding scheme previously established by Lam et al. [2012]. The results of our review include an overview of the most common evaluation goals in the community, how they evolved over time, and how they contrast or align to those of the IEEE Information Visualization conference. In particular, we found that evaluations specific to assessing resulting images and algorithm performance are the most prevalent (with consistently 80–90% of all papers since 1997). However, especially over the last six years there is a steady increase in evaluation methods that include participants, either by evaluating their performances and subjective feedback or by evaluating their work practices and their improved analysis and reasoning capabilities using visual tools. Up to 2010, this trend in the IEEE Visualization conference was much more pronounced than in the IEEE Information Visualization conference which only showed an increasing percentage of evaluation through user performance and experience testing. Since 2011, however, also papers in IEEE Information Visualization show such an increase of evaluations of work practices and analysis as well as reasoning using visual tools. Further, we found that generally the studies reporting requirements analyses and domain-specific work practices are too informally reported which hinders cross-comparison and lowers external validity.

**Index Terms**—Evaluation, validation, systematic review, visualization, scientific visualization, information visualization

## 1 Motivation

In this paper, we report a systematic review of 581 papers from ten years of IEEE Visualization conference publications with respect to their use of evaluation. We provide a quantitative and objective report of the types of evaluations encountered in the literature. At the same time, we also qualitatively assess our observations from coding these 581 papers. Specifically, we put evaluation practices into historic perspective and assess and compare them in context to those of the larger visualization community. Our goal in pursuing this work is to get an understanding of the practices of evaluation in visualization research as a whole.

The importance of evaluation to the field of visualization has become well recognized—demonstrated by the growing body of work on how to conduct visualization evaluation and by the growing amount of research papers that incorporate some form of formal or informal evaluation. In this article we contribute to the body of work by providing a systematic assessment and u… published peer-re… to such a systema…

Our work is b… in which they id… research articles. … reflect on the ent… known as the 'inf… all other visualiz… identifying evalu… We aim to comple… nity by answering the question: What are evaluation practices in the 'scientific visualization' part of our community? What are similarities

*read this paper for more advice on what to evaluate & what mistakes have been made in the past*

and differences between these sub-communities? To do so, we use and extend Lam et al.'s scenarios to systematically analyze the literature that appeared at the IEEE Visualization conference. We believe that our extended work is fundamental to understanding all subcultures in visualization and to properly sample all aspects of visualization work, not only those labeled as 'information visualization.'

By looking at the historic record, we were hoping to uncover some trends by examining how the field of visualization has been changing over the last 15 years. We were wondering whether some of the self-reflection by some of the field's leaders in the early 2000's has left its mark on our community and whether it led to more rigor in our evaluations. Likewise, our work is an opportunity to compare the IEEE Information Visualization and IEEE Visualization conferences to better understand their differences and commonalities. Our analysis of evalu… …oth weaknesses …learn for future …on practices but …to improve their …emplary papers …l.

…efold. First, we …he visualization …n the works in …e work done by …w of the use of evaluation in the visualization community as reported in the IEEE Information Visualization and IEEE Visualization conferences and put evaluation practices into perspective. This is a qualitative assessment and provides a historical perspective by comparing current and past evaluation practices. And, third, we provide information for researchers conducting evaluation by assisting them to identify, justify, and refine evaluation approaches as well as helping them to recognize and avoid pitfalls that can be learned from previous research.

## 2 Fundamentals and Related Work

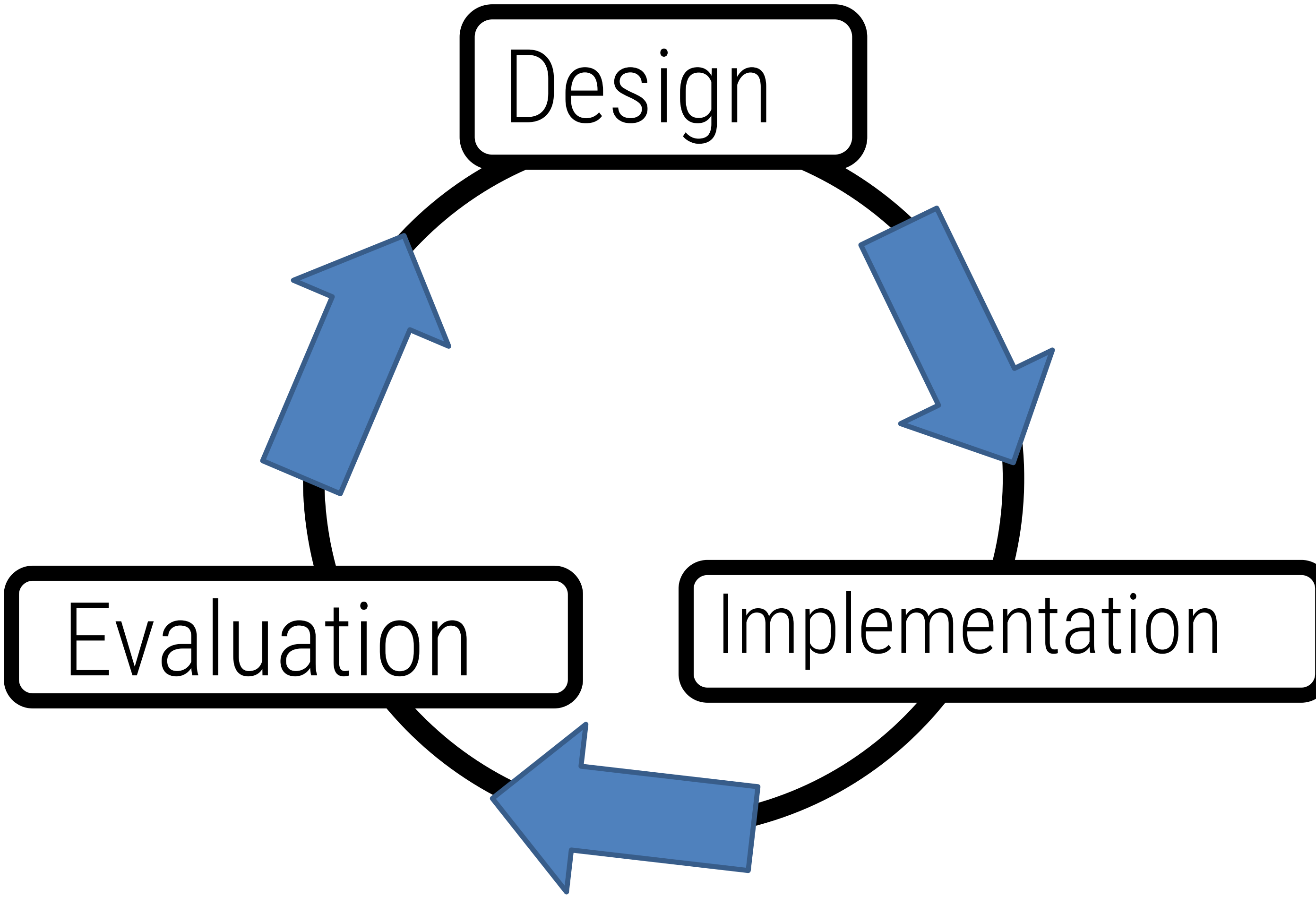
There are two traditions of evaluation that the visualization community draws from—evaluation in the sciences (both social and natural) and evaluation in design. On the one hand, scientists try to understand the world and seek a representative model, often a mathematical model (e.g., Newton's law or Fitts' law), while designers and engineers introduce a tool and henceforth seek to alter the world in which they live and


- *Tobias Isenberg is with INRIA, France. E-mail: tobias.isenberg@inria.fr.*
- *Petra Isenberg is with INRIA, France. E-mail: petra.isenberg@inria.fr.*
- *Jian Chen is with the University of Maryland, Baltimore County, USA. E-mail: jichen@umbc.edu.*
- *Michael Sedlmair is with the University of Vienna, Austria. E-mail: michael.sedlmair@univie.ac.at.*
- *Torsten Möller is with the University of Vienna, Austria. E-mail: torsten.moeller@univie.ac.at.*


*Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.*
*For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.*

# QUESTION 2

WHEN SHOULD YOU EVALUATE SOMETHING?




Design Life Cycle

# EVALUATION

IN VISUALIZATION

| | |
|---|---|
| pre-design | context of solution |
| design | cognition and perception for encodings |
| prototype | design goals, comparisons |
| deployment | understand adoption |
| re-design | usability problems |

# QUESTION 2+3

WHEN & WHAT SHOULD YOU EVALUATE?

# THE NESTED MODEL

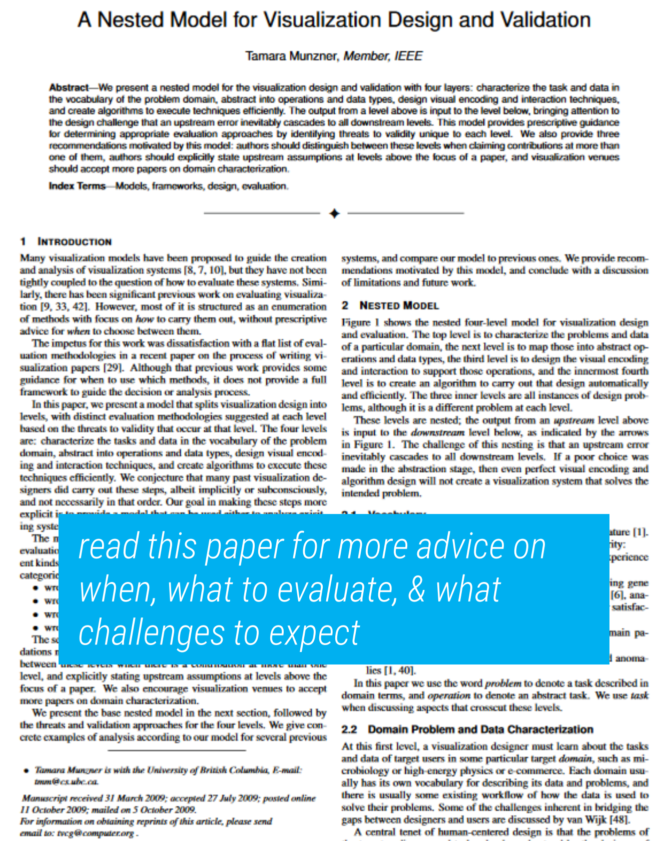
Domain problem characterization

Data/task abstraction design

Encoding/interaction technique design

Algorithm design

A model for visualization creation



# A Nested Model for Visualization Design and Validation

Tamara Munzner, *Member, IEEE*

**Abstract**—We present a nested model for the visualization design and validation with four layers: characterize the task and data in the vocabulary of the problem domain, abstract into operations and data types, design visual encoding and interaction techniques, and create algorithms to execute techniques efficiently. The output from a level above is input to the level below, bringing attention to the design challenge that an upstream error inevitably cascades to all downstream levels. This model provides prescriptive guidance for determining appropriate evaluation approaches by identifying threats to validity unique to each level. We also provide three recommendations motivated by this model: authors should distinguish between these levels when claiming contributions at more than one of them, authors should explicitly state upstream assumptions at levels above the focus of a paper, and visualization venues should accept more papers on domain characterization.

**Index Terms**—Models, frameworks, design, evaluation.

## 1 Introduction

Many visualization models have been proposed to guide the creation and analysis of visualization systems [8, 7, 10], but they have not been tightly coupled to the question of how to evaluate these systems. Similarly, there has been significant previous work on evaluating visualization [9, 33, 42]. However, most of it is structured as an enumeration of methods with focus on *how* to carry them out, without prescriptive advice for *when* to choose between them.

The impetus for this work was dissatisfaction with a flat list of evaluation methodologies in a recent paper on the process of writing visualization papers [29]. Although that previous work provides some guidance for when to use which methods, it does not provide a full framework to guide the decision or analysis process.

In this paper, we present a model that splits visualization design into levels, with distinct evaluation methodologies suggested at each level based on the threats to validity that occur at that level. The four levels are: characterize the tasks and data in the vocabulary of the problem domain, abstract into operations and data types, design visual encoding and interaction techniques, and create algorithms to execute these techniques efficiently. We conjecture that many past visualization designers did carry out these steps, albeit implicitly or subconsciously, and not necessarily in that order. Our goal in making these steps more explicit i[illegible] ing syste[illegible]

The m[illegible] evaluatio[illegible] ent kinds[illegible] categori[illegible]

- wr[illegible]
- wr[illegible]
- wr[illegible]
- wr[illegible]

The s[illegible] dations [illegible] between [illegible] level, and explicitly stating upstream assumptions at levels above the focus of a paper. We also encourage visualization venues to accept more papers on domain characterization.

We present the base nested model in the next section, followed by the threats and validation approaches for the four levels. We give concrete examples of analysis according to our model for several previous systems, and compare our model to previous ones. We provide recommendations motivated by this model, and conclude with a discussion of limitations and future work.

## 2 Nested Model

Figure 1 shows the nested four-level model for visualization design and evaluation. The top level is to characterize the problems and data of a particular domain, the next level is to map those into abstract operations and data types, the third level is to design the visual encoding and interaction to support those operations, and the innermost fourth level is to create an algorithm to carry out that design automatically and efficiently. The three inner levels are all instances of design problems, although it is a different problem at each level.

These levels are nested; the output from an *upstream* level above is input to the *downstream* level below, as indicated by the arrows in Figure 1. The challenge of this nesting is that an upstream error inevitably cascades to all downstream levels. If a poor choice was made in the abstraction stage, then even perfect visual encoding and algorithm design will not create a visualization system that solves the intended problem.

### 2.1 [illegible]

[illegible] ature [1]. [illegible] ity: [illegible] perience [illegible] ing gene [illegible] [6], ana- [illegible] satisfac- [illegible] main pa- [illegible] anoma- lies [1, 40].

In this paper we use the word *problem* to denote a task described in domain terms, and *operation* to denote an abstract task. We use *task* when discussing aspects that crosscut these levels.

### 2.2 Domain Problem and Data Characterization

At this first level, a visualization designer must learn about the tasks and data of target users in some particular target *domain*, such as microbiology or high-energy physics or e-commerce. Each domain usually has its own vocabulary for describing its data and problems, and there is usually some existing workflow of how the data is used to solve their problems. Some of the challenges inherent in bridging the gaps between designers and users are discussed by van Wijk [48].

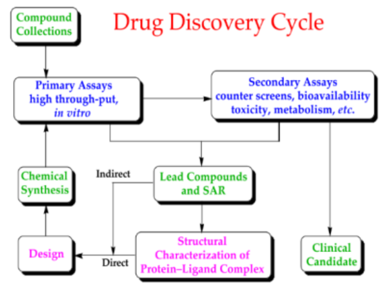
A central tenet of human-centered design is that the problems of the target audience need to be clearly understood by the designer of

• *Tamara Munzner is with the University of British Columbia, E-mail: tmm@cs.ubc.ca.*

*Manuscript received 31 March 2009; accepted 27 July 2009; posted online 11 October 2009; mailed on 5 October 2009.*
*For information on obtaining reprints of this article, please send email to: tvcg@computer.org .*

*read this paper for more advice on when, what to evaluate, & what challenges to expect*

# THE NESTED MODEL




Domain problem characterization

understand as much as possibly about users, their tasks, and context of use in order to produce a stable and detailed set of requirements

“discover new candidate medications” vs.
“identify clusters of compounds where all compounds interact with only one specific pathway”

# THE NESTED MODEL

Domain problem characterization

Data/task abstraction design

Translate the domain problems into generic visualization tasks & usable data types for visualization

"identify clusters of compounds where all compounds interact with only one specific pathway"
→ "analyze network connectivity"

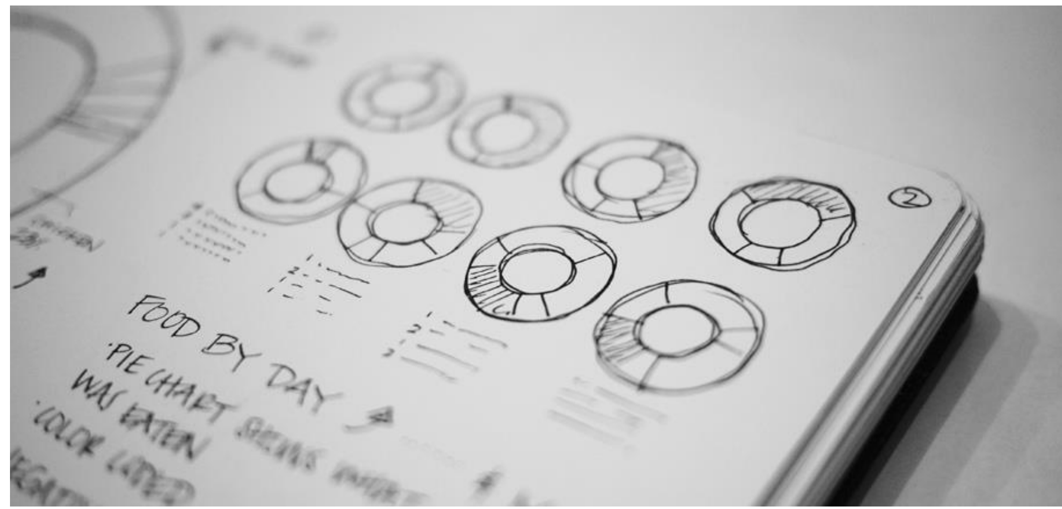
# THE NESTED MODEL

Domain problem characterization

Data/task abstraction design

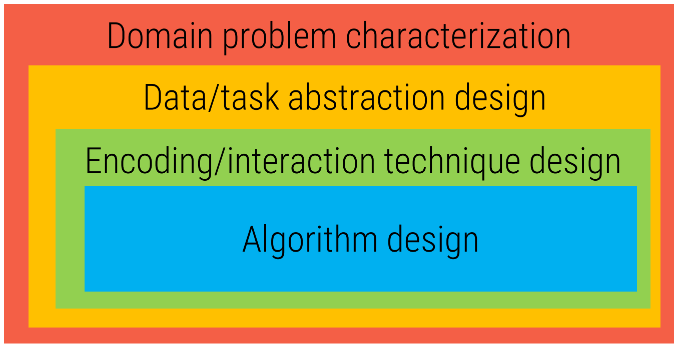
Encoding/interaction technique design

## Design visual encoding and interaction

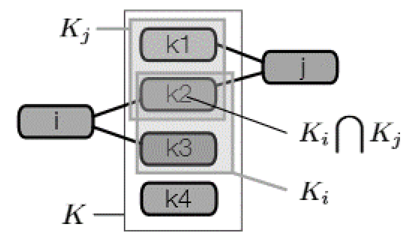



by Lauren Manning Sketchbook

# THE NESTED MODEL

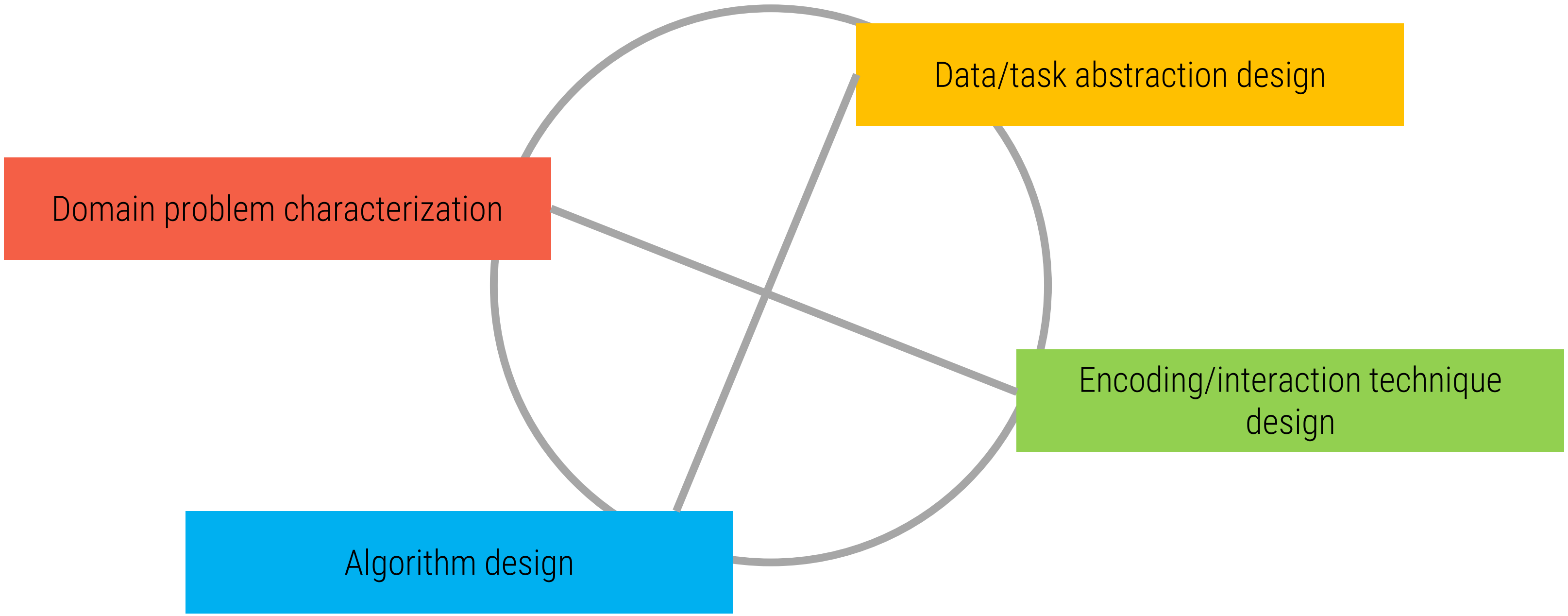



Create algorithm to carry out the visual encoding and interaction design automatically

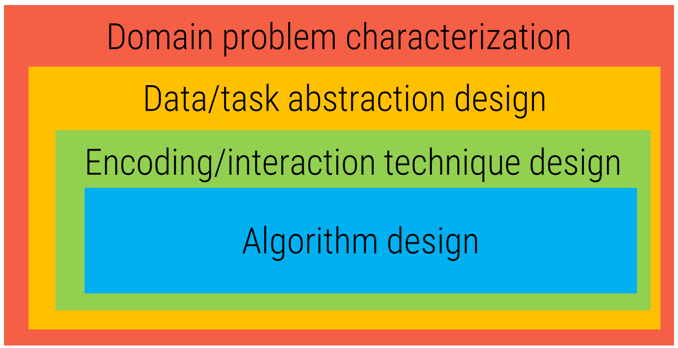



Enrichment for $i$ in $j$ via $K$: $s_{i,j}(K) = (1/2)/(2/4) = 1$

# THE NESTED MODEL



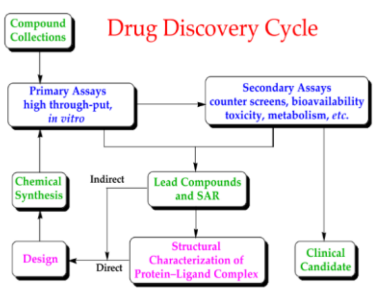
# EVALUATION QUESTIONS




- What can go wrong at each step?
- How can you avoid doing a step incorrectly?
- How can you validate that you completed each step correctly?

# EVALUATION QUESTIONS

Domain problem characterization




THREAT:

- You chose the wrong problem
- You did not understand the domain language/problem

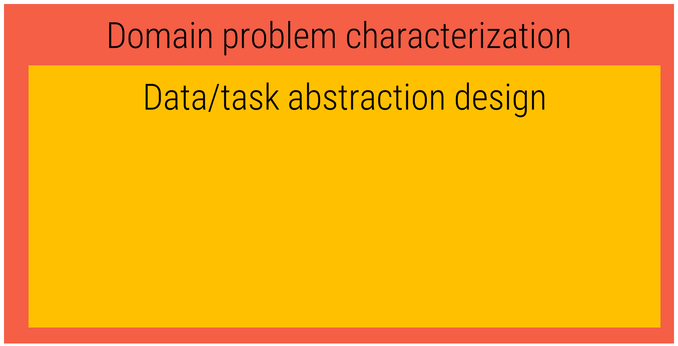
WHAT CAN BE DONE:

- Observe and interview domain users

# QUESTIONS TO ASK:

- who are the users?
- what are their needs?
- what are their tasks?
- how do they currently work?

# EVALUATION QUESTIONS




THREAT:

- General tasks / data types do not solve domain problem

WHAT CAN BE DONE:

- Compare against existing approaches
- Analyze your previously collected data carefully with specific methods

# TASK TAXONOMIES IN VISUALIZATION

Many task taxonomies exist in visualization

- Some are very general
- Some domain specific

Recent call for publishing more domain specific analyses and design studies

BMC Bioinformatics. 2017; 18(Suppl 2): 21.
Published online 2017 Feb 15. doi: 10.1186/s12859-016-1443-5
PMCID: PMC5333192
PMID: 28251869

## A taxonomy of visualization tasks for the analysis of biological pathway data

Paul Murray,[1] Fintan McGee,[2] and Angus G. Forbes[1]



### Abstract

Go to:

#### Background

Go to:

Understanding complicated networks of interactions and chemical components is essential to solving contemporary problems in modern biology, especially in domains such as cancer and systems research. In these domains, biological pathway data is used to represent chains of interactions that occur within a given biological process. Visual representations can help researchers understand, interact with, and reason about these complex pathways in a number of ways. At the same time, these datasets offer unique challenges for visualization, due to their complexity and heterogeneity.

#### Results

Go to:

Here, we present taxonomy of tasks that are regularly performed by researchers who work with biological pathway data. The generation of these tasks was done in conjunction with interviews with several domain experts in biology. These tasks require further classification than is provided by existing taxonomies. We also examine existing visualization techniques that support each task, and we discuss gaps in the existing visualization space revealed by our taxonomy.

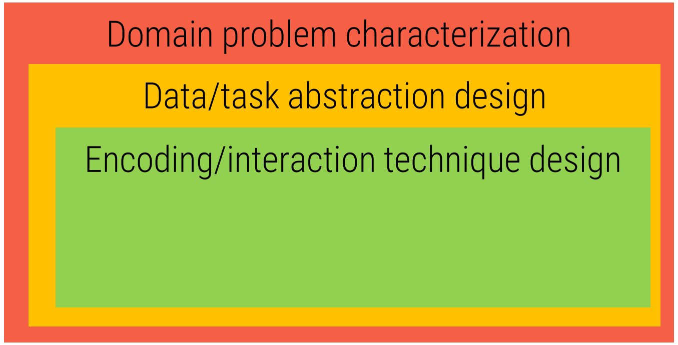
#### Conclusions

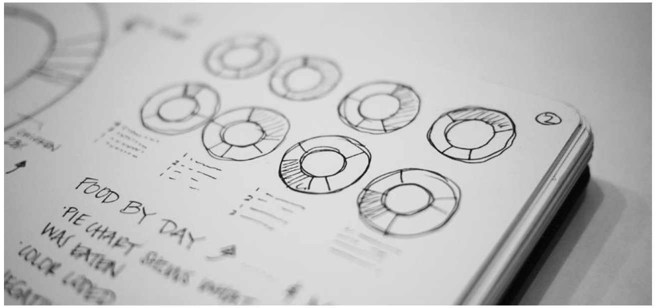
Go to:

Our taxonomy is designed to support the development and design of future biological pathway visualization applications. We conclude by suggesting future research directions based on our taxonomy and motivated by the comments received by our domain experts.

**Keywords:** Biological pathways, Pathway visualization, Task taxonomy

# EVALUATION QUESTIONS



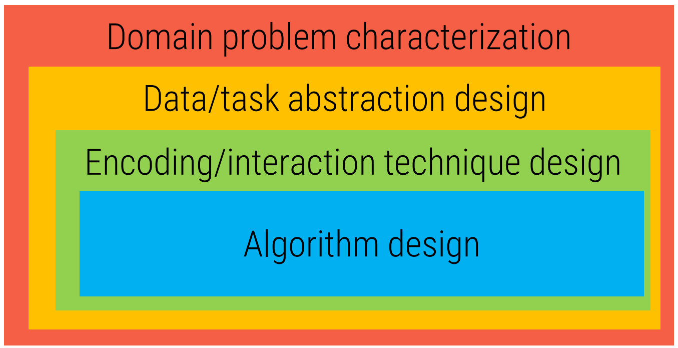

THREAT:

- Ineffective encoding/interaction technique

WHAT CAN BE DONE:

- Justify the design with known guidelines

# EVALUATION QUESTIONS




THREAT:

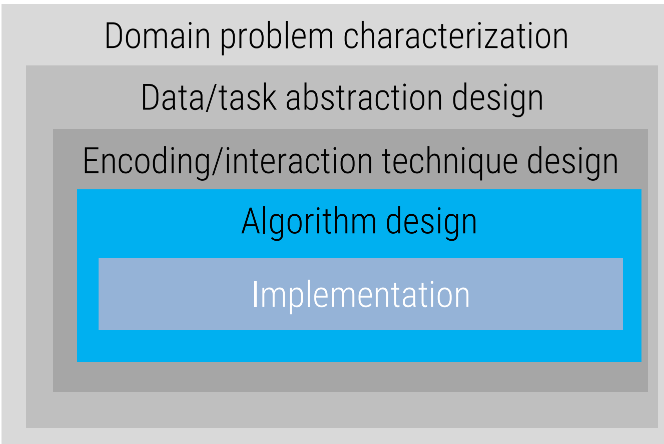
- Suboptimal algorithms in terms of speed and memory
- Incorrect algorithm

WHAT CAN BE DONE:

- Analyze computational complexity


IMPLEMENT

# EVALUATION QUESTIONS




THREAT:

- Suboptimal algorithms in terms of speed and memory
- Incorrect algorithm

WHAT CAN BE DONE:

- Benchmark testing

# Visual Analytics Benchmark Repository

*A service of the SEMVAST Project*
*Managed by HCIL, University of Maryland*



Benchmarks contain **datasets and tasks**, as well as materials describing the **uses** of those benchmarks (the results of analysis, contest entries, controlled experiment materials etc.) Most benchmarks contain ground truth described in a solution provided with the benchmark, allowing accuracy metrics to be computed.

## List of Benchmarks

### VAST Challenge 2017

- Grand Challenge
- Mini-Challenge 1
- Mini-Challenge 2
- Mini-Challenge 3

## Contact information

Kris Cook, Pacific Northwest National Laboratory

Georges Grinstein, University of Massachusetts Amherst

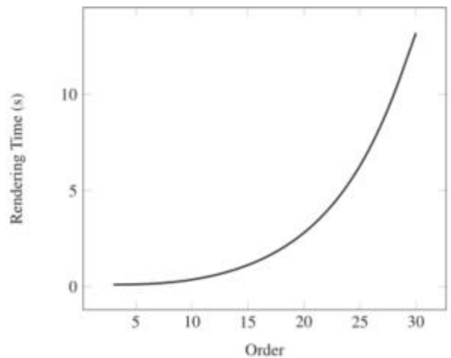
Mark Whiting, Pacific Northwest National Laboratory

Kristen Liggett, Air Force Research Laboratory

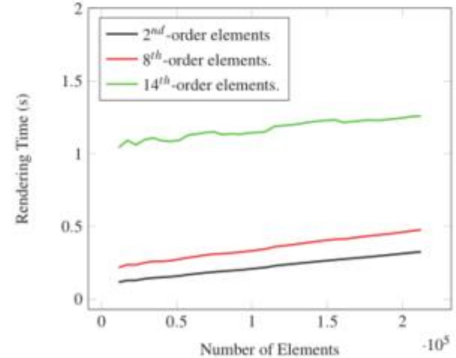
Diane Staheli, MIT Lincoln Laboratory

Jordan Crouser, Smith College

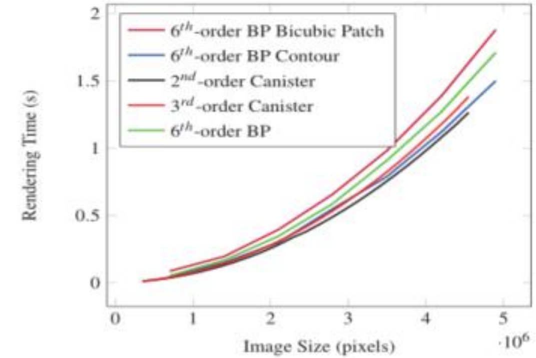
John Fallon, University of Massachusetts Amherst



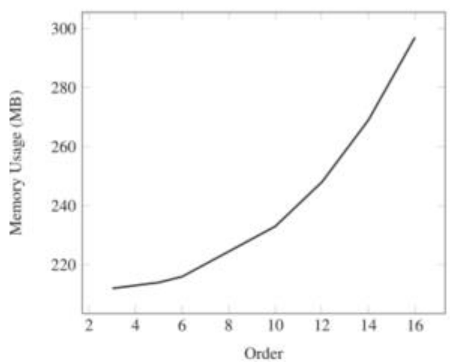

(a) Time in seconds to render volumes with increasing polynomial order.



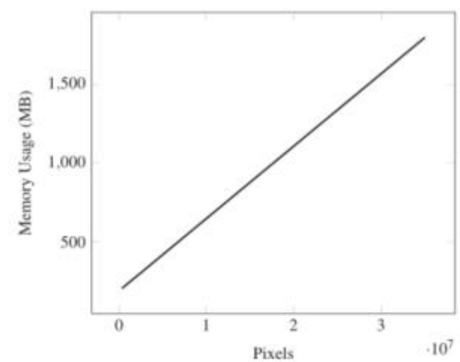

(b) Time in seconds to render volumes with a varying number of elements.



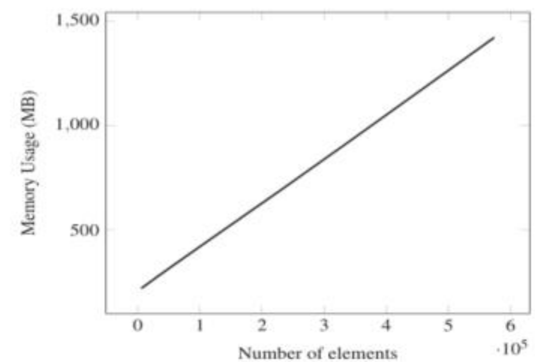

(c) Time in seconds to render high-order data sets for a variety of image sizes and cut-surfaces, measured in pixels. BP refers to the block/plate configuration.




(d) Varying order with $50,000$ elements and a $1000 \times 1000$ image.



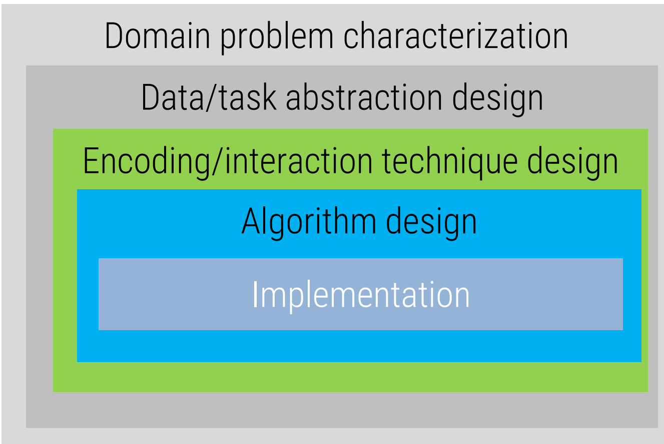

(e) Varying image size with $50,000$ $8^{th}$ order elements.




(f) Varying the number of elements with $8^{th}$ order solutions and $1000 \times 1000$ image size.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6064943

# EVALUATION QUESTIONS



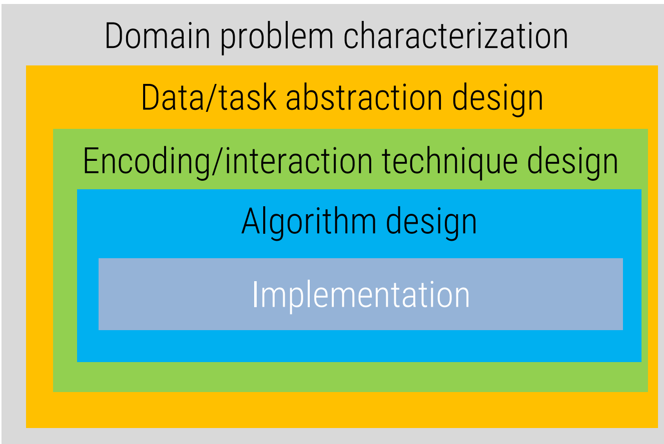

THREAT:

- Ineffective encoding/interaction technique

WHAT CAN BE DONE:

- Formal user study
- Present and discuss the implemented system with experts, end users
- Quantitatively assess result images

# EVALUATION QUESTIONS



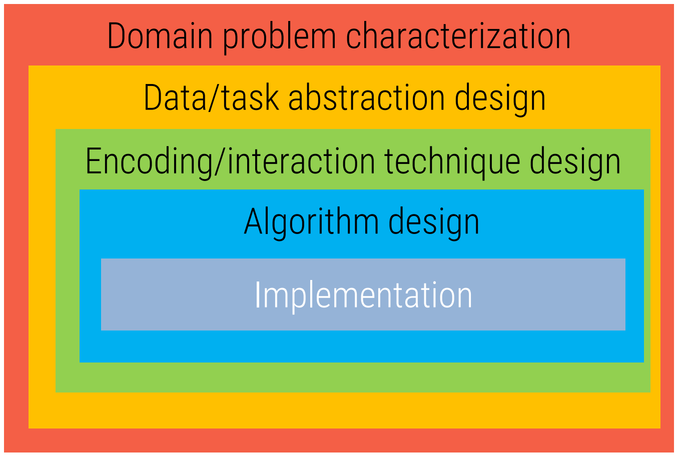

THREAT:

- General tasks / data types do not solve domain problem

WHAT CAN BE DONE:

- Test tool on target users, check if tasks can be completed. E.g. can hypotheses be found?
- Long term field studies

# EVALUATION QUESTIONS



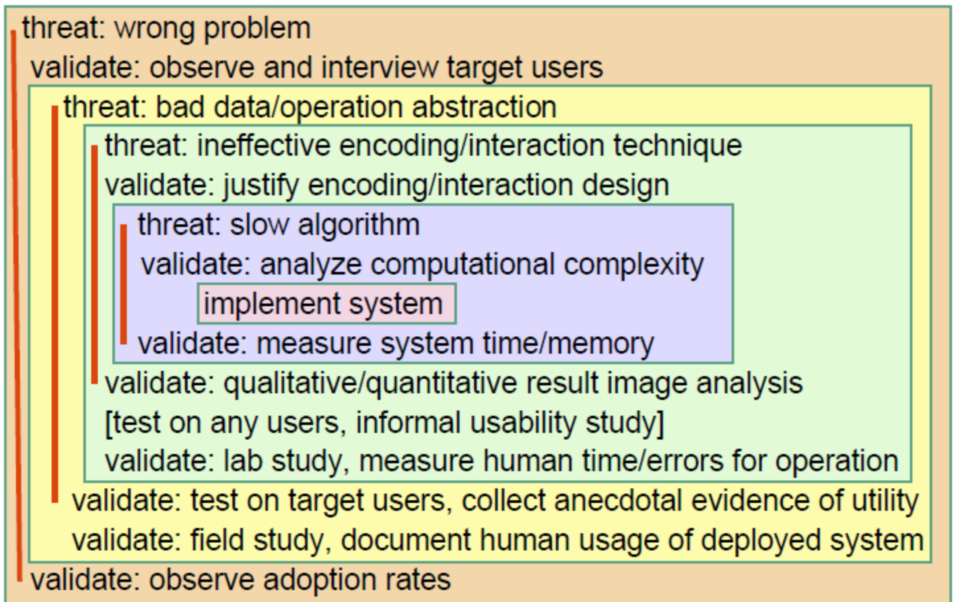

THREAT:

- You chose the wrong problem
- You did not understand the domain language/problem

WHAT CAN BE DONE:

- Check adoption rates

# IN SUMMARY – THE NESTED MODEL

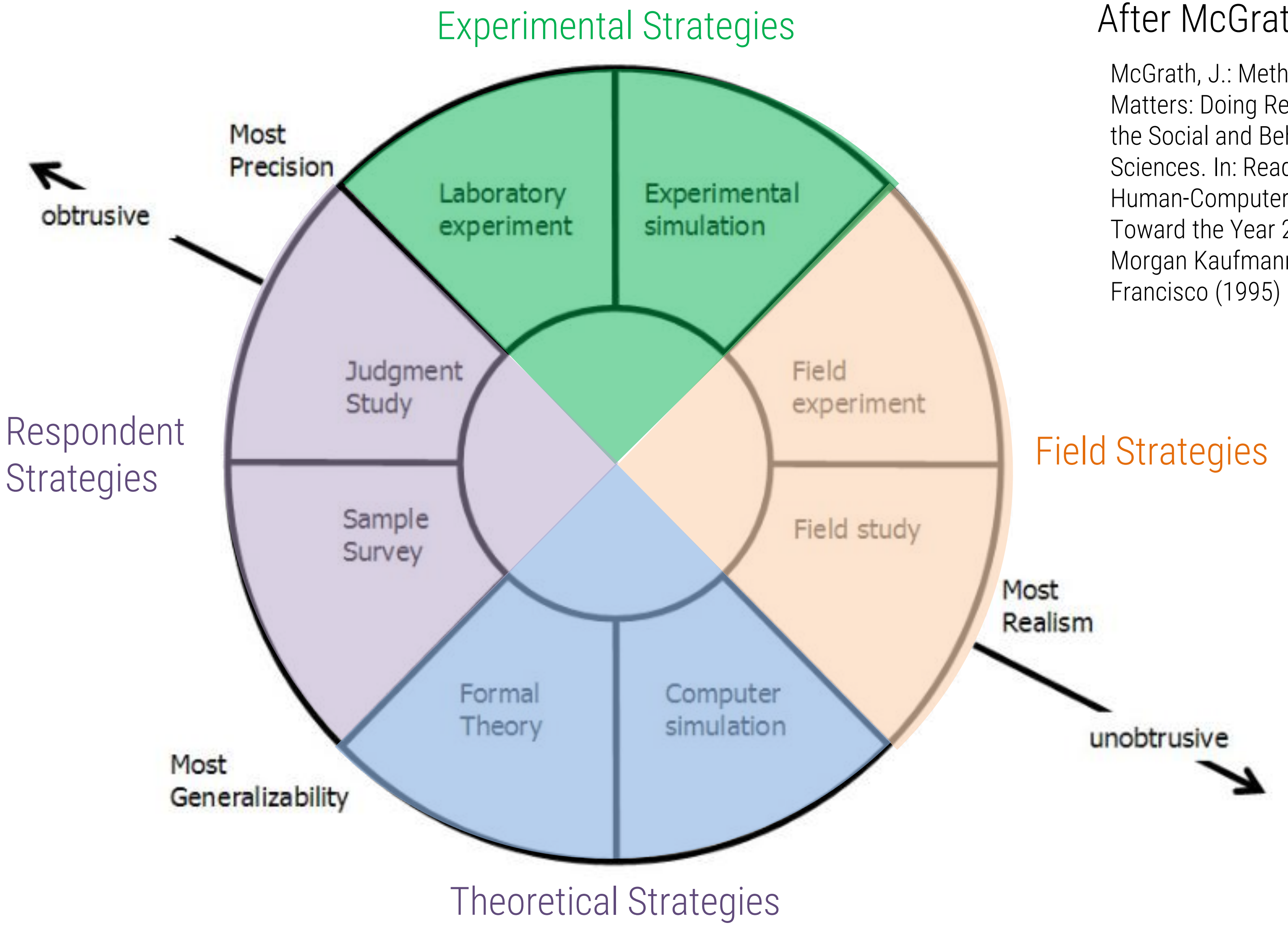



read this paper for advice on evaluation in the development cycle

Munzner, InfoVis 2009

# QUESTION 4

HOW TO EVALUATE?



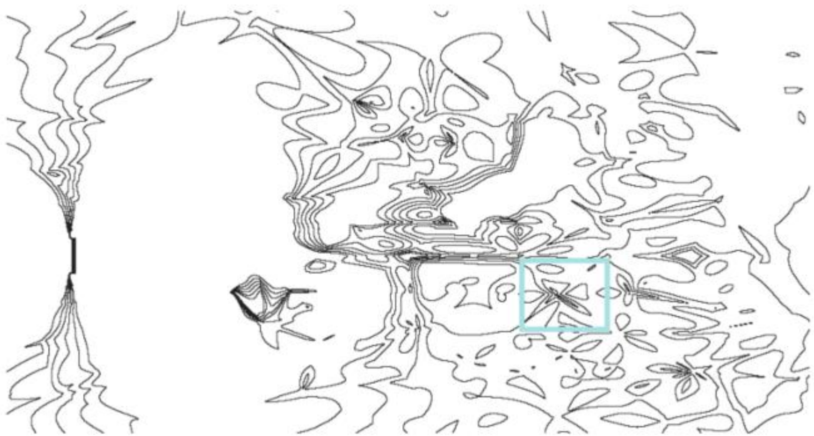

After McGrath 1995

McGrath, J.: Methodology Matters: Doing Research in the Social and Behavioural Sciences. In: Readings in Human-Computer Interaction: Toward the Year 2000, Morgan Kaufmann, San Francisco (1995)

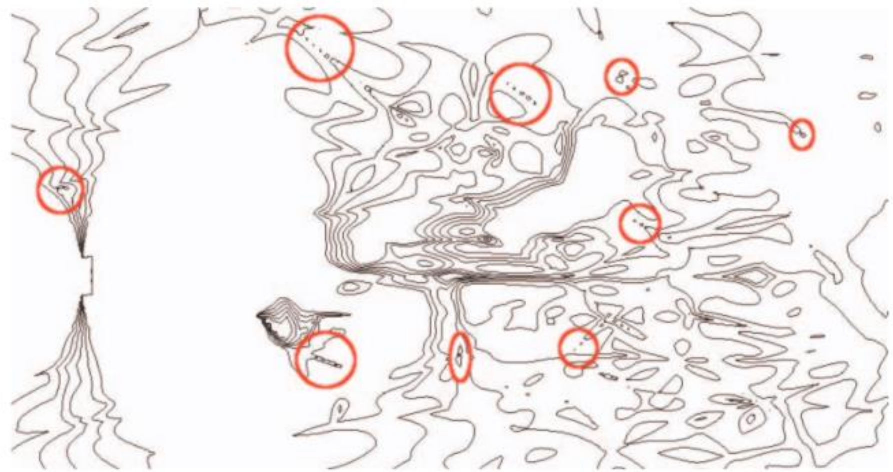
# LOW-COST METHODS

OFTEN EARLY ON PROTOTYPES

# QUALITATIVE RESULT INSPECTION



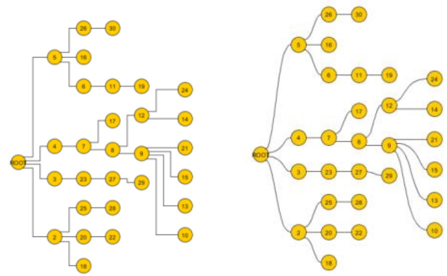
(c) High-Order Rendering. Rendering time for 2000×2000 image = 0.3 seconds.



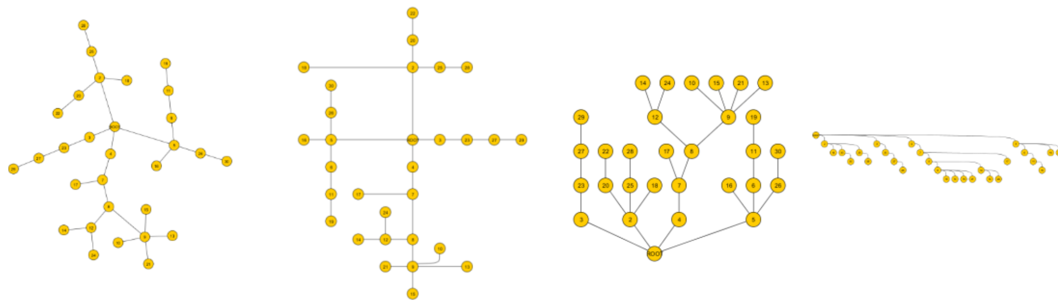
(a) 524, 288 Triangles. VTK Contour Generation Time = 0.265 seconds.

# AESTHETICS

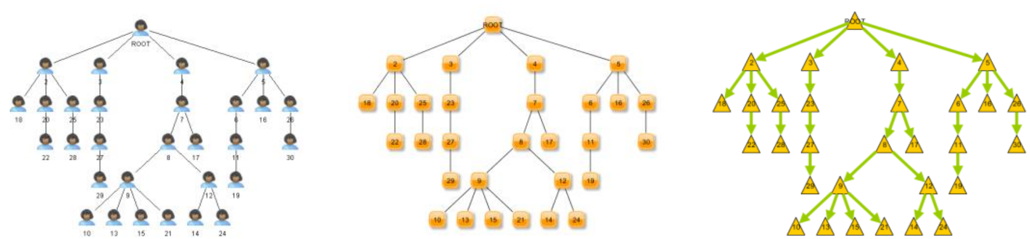
vary edges



node placement



marks



Images created with yEd: http://www.yworks.com

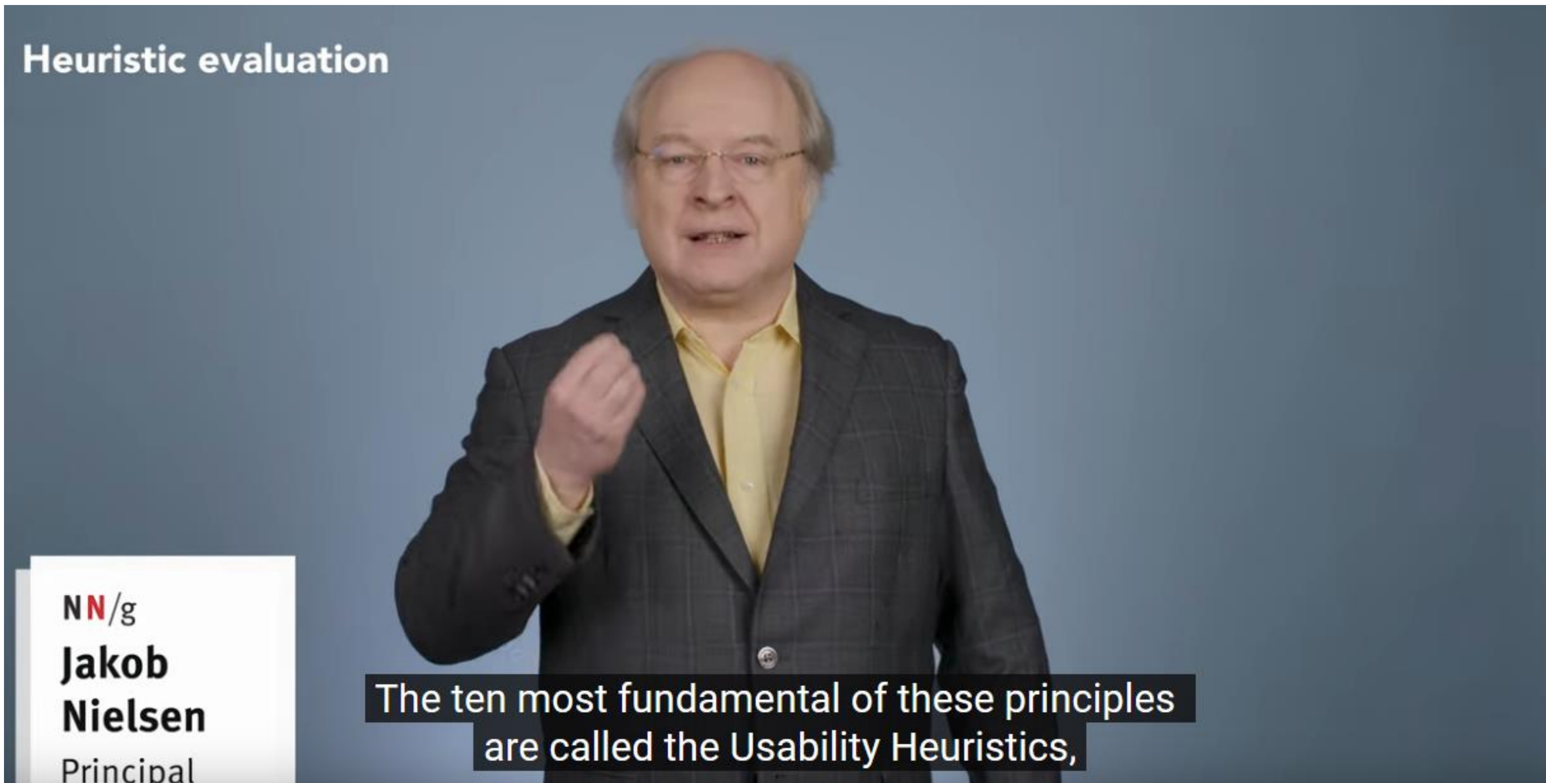
# AESTHETICS

aesthetics of node-link tree algorithms describe properties that improve the perception of the data that is being layed out

- **area**: match area of your layout to the size of the display and data
- **aspect ratio**: usually optimal if close to 1
- **subtree separation**: try not to overlap subtrees
- **root-leaf distance**: minimize distance from root to leaves
- **edge lengths**: minimize total, average, maximum, edge lengths & try to make edge lengths uniform
- **angular resolution**: increase angles formed by edges
- **symmetry**: symmetric layouts usually considered pleasing

# HEURISTIC EVALUATION

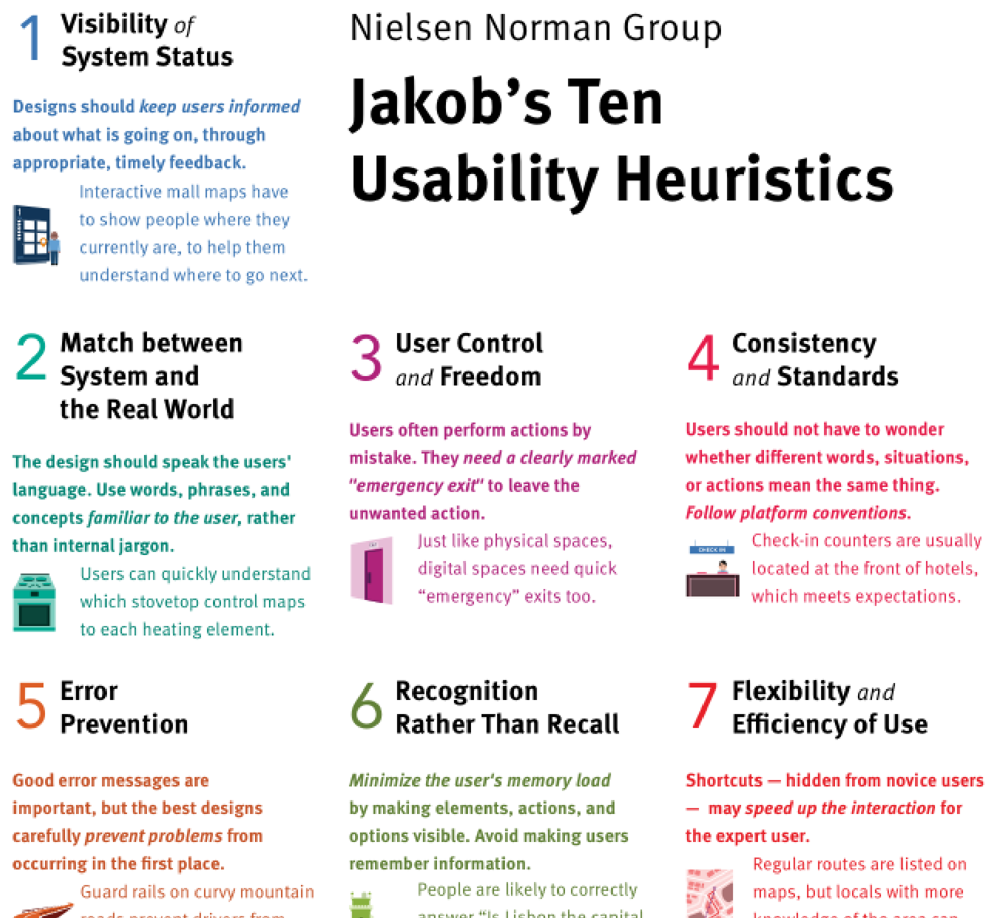
by Jakob Nielsen (1994) and others




https://www.youtube.com/watch?v=6Bw0n6Jvwxk

# HEURISTIC EVALUATION

use of design principles/heuristics to inspect an interface for usability problems




https://www.nngroup.com/articles/ten-usability-heuristics/

# HEURISTIC EVALUATION

- choose the guidelines
- define a rating systems
- documents

Also choose: number of evaluators

  - single inspector
  - multiple inspectors

HEURISTIC EVALUATION = recruit USABILITY EXPERTS

USER TESTING = recruit PARTICIPANTS

# HEURISTIC EVALUATION IN VISUALIZATION

## Lots of recent research interest

| Set | Heuristics |
|---|---|
| Zuk and Carpendale's *Selection of perceptual and cognitive heuristics [26]* | Ensure visual variable has sufficient length [3][25][26]<br>Don't expect a reading order from color [3][25][26]<br>Color perception varies with size of colored item [25][3][26]<br>Local contrast affects color & gray perception [25][26]<br>Consider people with color blindness [25][26][22]<br>Preattentive benefits increase with field of view [3][25][26][11]<br>Quantitative assessment requires position or size variation [3][26]<br>Preserve data to graphic dimensionality [24][3][26]<br>Put the most data in the least space [24][26]<br>Remove the extraneous (ink) [24][26]<br>Consider Gestalt Laws [25][26]<br>Provide multiple levels of detail [24][25][26]<br>Integrate text wherever relevant [24][25][26] |
| Shneiderman's *"Visual Information-Seeking Mantra"[20]* | Overview first [20]<br>Zoom and filter [20]<br>Details on demand [20]<br>Relate [20]<br>Extract [20]<br>History [20] |
| Amar and Stasko's *Knowledge and task-based framework [1]* | Expose uncertainty [1]<br>Concretize relationships [1]<br>Determination of Domain Parameters [1]<br>Multivariate Explanation [1]<br>Formulate cause & effect [1]<br>Confirm Hypotheses [1] |

Many more…

# THE ICE-T MODEL

- http://visvalue.org/
- A form of heuristic evaluation

# THE FOUR CHARACTERISTICS

INSIGHT + TIME + ESSENCE + CONFIDENCE

# INSIGHT

A visualization's ability to spur and discover **insights** and/or **insightful questions** about the data

# CONFIDENCE

A visualization's ability to generate **confidence** , knowledge, and trust about the data, its domain and context

# ESSENCE

A visualization's ability to convey an overall **essence** or take-away sense of the data

# TIME

A visualization's ability to minimize the total **time** needed to answer a wide variety of questions about the data

# LAB STUDIES

OFTEN WITH HIGH-FIDELITY PROTOTYPES

# LAB-BASED TESTING: ESSENTIALLY…

- bring in real users
- have them complete tasks with your design, while you watch <u>with your entire team</u>
- use a think-aloud protocol, so you can "hear what they are thinking"
- measure
  - task completion, task time
  - satisfaction, problem points, etc.
- identify problems (major ones | minor ones)
- provide design suggestions to design/engineering team
- iterate on the design, repeat

# TESTING ENVIRONMENTS…

# IN-THE-WILD STUDIES

OFTEN WITH HIGH-FIDELITY PROTOTYPES

# RESEARCH METHODS - IDEAL

- observing and/or interviewing the real end users
  - find out what current methods users use for doing their tasks
    (paper, competing systems, antiquated systems, …)
  - abstract users → real people with real needs

*example:*
*if you are interested in customers who do data analysis for drug discovery, observe and talk to them in their current work environment*

# RESEARCH METHODS – SECOND BEST

interviewing the end-user representative

- if you absolutely cannot get hold of end-users
- carefully select and interview end-user representatives
- MUST be people with direct contact with end users and intimate knowledge and experience of their needs and what they do
- people who work with end users are the best

*Example:*
*talk to managers/team leaders if you cannot get hold of actual analysts.*
*Better: interview/observe how the representatives analyze data*

# RESEARCH METHODS – IF ALL ELSE FAILS

make your beliefs about the end users and the task space explicit

- if you cannot get in touch with real end users or their representatives
- use your team to articulate their assumptions about end users and their tasks
- risk: resulting user and task descriptions do not resemble reality → only use as last resort

# WHEN LOOKING IS NOT ENOUGH…

- LOOKing gives you great insight into the state of the world

- But it doesn't tell you <u>why</u> people are acting the way they do, or what their goals, needs, or feelings are

# ASK

- Surveys
- Interviews
- Focus Groups
- Diary Studies
- Experience Sampling
- …

# LESSON TO LEARN ABOUT INTERVIEWS

- what people say they want and what they want is not always the same
  - through observation you can uncover the latter
- what people say they do is not always what they actually do
  - through observation you can see what they do

# WHAT CAN HAPPEN WHEN TALKING TO PEOPLE

# IDEALLY, COMBINE INTERVIEWS WITH OBSERVATIONS

- watch people in their own environment
- watch people do everyday tasks

To look up further:
Contextual Inquiry

- opportunities for new designs arise from:
  - workarounds
  - breakdowns
  - unexpected uses of existing tools

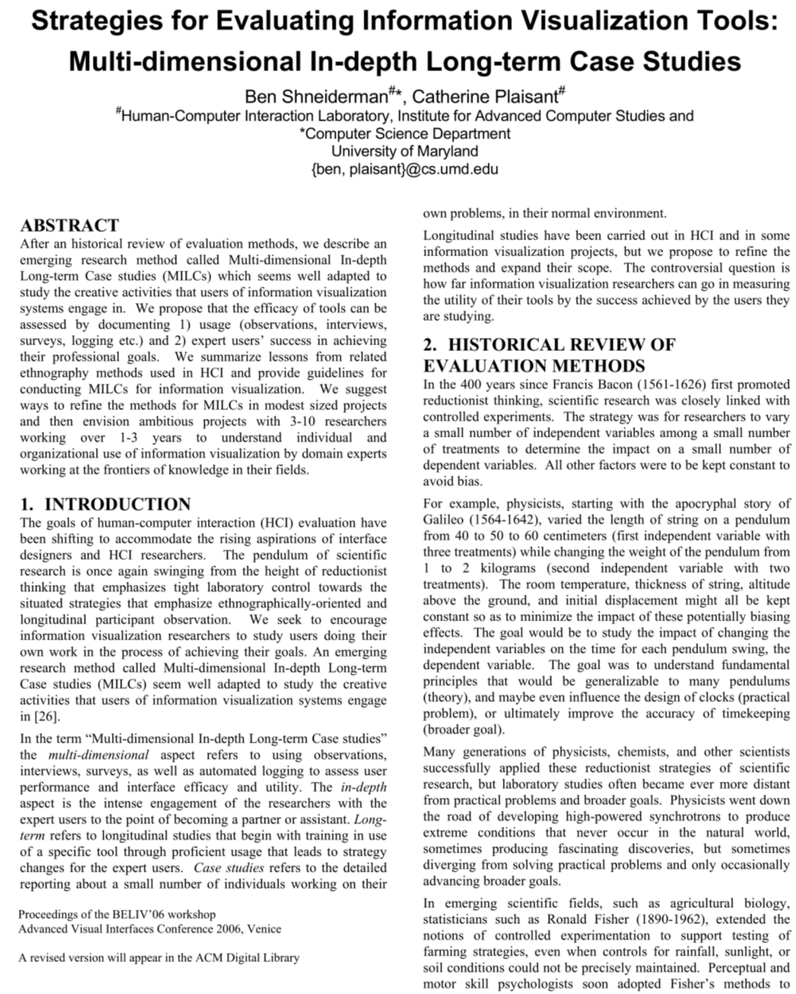
# MILCS

## Uses

- Ethnographical observatior
- Interviews
- Automated logging

*Especially useful in situations where replicability is not attainable.*

*The outcome may be specific suggestions for tool improvements and a better understanding of design principles.*



**Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies**

Ben Shneiderman[#*], Catherine Plaisant[#]
[#]Human-Computer Interaction Laboratory, Institute for Advanced Computer Studies and
*Computer Science Department
University of Maryland
{ben, plaisant}@cs.umd.edu

**ABSTRACT**
After an historical review of evaluation methods, we describe an emerging research method called Multi-dimensional In-depth Long-term Case studies (MILCs) which seems well adapted to study the creative activities that users of information visualization systems engage in. We propose that the efficacy of tools can be assessed by documenting 1) usage (observations, interviews, surveys, logging etc.) and 2) expert users' success in achieving their professional goals. We summarize lessons from related ethnography methods used in HCI and provide guidelines for conducting MILCs for information visualization. We suggest ways to refine the methods for MILCs in modest sized projects and then envision ambitious projects with 3-10 researchers working over 1-3 years to understand individual and organizational use of information visualization by domain experts working at the frontiers of knowledge in their fields.

**1. INTRODUCTION**
The goals of human-computer interaction (HCI) evaluation have been shifting to accommodate the rising aspirations of interface designers and HCI researchers. The pendulum of scientific research is once again swinging from the height of reductionist thinking that emphasizes tight laboratory control towards the situated strategies that emphasize ethnographically-oriented and longitudinal participant observation. We seek to encourage information visualization researchers to study users doing their own work in the process of achieving their goals. An emerging research method called Multi-dimensional In-depth Long-term Case studies (MILCs) seem well adapted to study the creative activities that users of information visualization systems engage in [26].

In the term "Multi-dimensional In-depth Long-term Case studies" the *multi-dimensional* aspect refers to using observations, interviews, surveys, as well as automated logging to assess user performance and interface efficacy and utility. The *in-depth* aspect is the intense engagement of the researchers with the expert users to the point of becoming a partner or assistant. *Long-term* refers to longitudinal studies that begin with training in use of a specific tool through proficient usage that leads to strategy changes for the expert users. *Case studies* refers to the detailed reporting about a small number of individuals working on their own problems, in their normal environment.

Proceedings of the BELIV'06 workshop
Advanced Visual Interfaces Conference 2006, Venice

A revised version will appear in the ACM Digital Library

Longitudinal studies have been carried out in HCI and in some information visualization projects, but we propose to refine the methods and expand their scope. The controversial question is how far information visualization researchers can go in measuring the utility of their tools by the success achieved by the users they are studying.

**2. HISTORICAL REVIEW OF EVALUATION METHODS**
In the 400 years since Francis Bacon (1561-1626) first promoted reductionist thinking, scientific research was closely linked with controlled experiments. The strategy was for researchers to vary a small number of independent variables among a small number of treatments to determine the impact on a small number of dependent variables. All other factors were to be kept constant to avoid bias.

For example, physicists, starting with the apocryphal story of Galileo (1564-1642), varied the length of string on a pendulum from 40 to 50 to 60 centimeters (first independent variable with three treatments) while changing the weight of the pendulum from 1 to 2 kilograms (second independent variable with two treatments). The room temperature, thickness of string, altitude above the ground, and initial displacement might all be kept constant so as to minimize the impact of these potentially biasing effects. The goal would be to study the impact of changing the independent variables on the time for each pendulum swing, the dependent variable. The goal was to understand fundamental principles that would be generalizable to many pendulums (theory), and maybe even influence the design of clocks (practical problem), or ultimately improve the accuracy of timekeeping (broader goal).

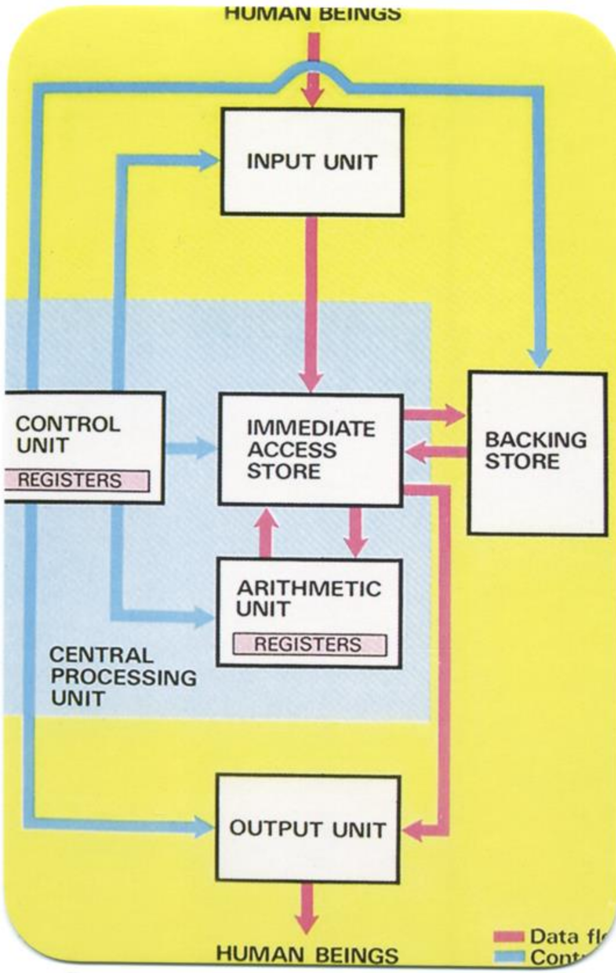
Many generations of physicists, chemists, and other scientists successfully applied these reductionist strategies of scientific research, but laboratory studies often became ever more distant from practical problems and broader goals. Physicists went down the road of developing high-powered synchrotrons to produce extreme conditions that never occur in the natural world, sometimes producing fascinating discoveries, but sometimes diverging from solving practical problems and only occasionally advancing broader goals.

In emerging scientific fields, such as agricultural biology, statisticians such as Ronald Fisher (1890-1962), extended the notions of controlled experimentation to support testing of farming strategies, even when controls for rainfall, sunlight, or soil conditions could not be precisely maintained. Perceptual and motor skill psychologists soon adopted Fisher's methods to

# ANALYZING YOUR DATA

# LEARN FROM YOUR DATA

- Now that you have a huge stack of notes and ideas from all of your LOOKing and ASKing, it's time to make some sense of the data
- Methods are intended to help you organize your thinking, and express it to help make it concrete and real




Learn Look Ask Try

# Flow Analysis

HOW: Represent the flow of information or activity through all phases of a system or process.

WHY: This is useful for identifying bottlenecks and opportunities for functional alternatives.

Designing an online advice website, flow analysis helped the IDEO team to design a more seamless experience navigating the site.

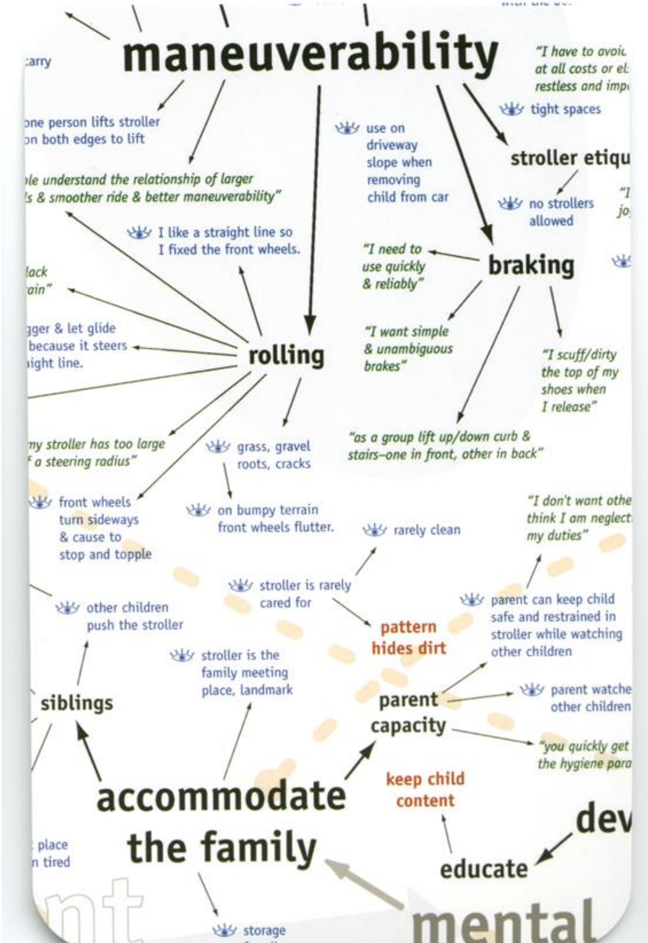



# Cognitive Task Analysis

HOW: List and summarize all of a user's sensory inputs, decision points, and actions.

WHY: This is good for understanding users' perceptual, attentional, and informational needs and to identify bottlenecks where errors may occur.

Cognitive task analysis helped the IDEO team understand the proximity and disorientation problems that remote-vehicle operators suffered due to the design of their controls.






# Affinity Diagrams

HOW: Cluster design elements according to intuitive relationships such as similarity, dependence, proximity, etc.

WHY: This method is a useful way to identify connections between issues and reveal innovation opportunities.

Clustering the elements related to transporting the family helped the IDEO team to discover some significant opportunities for stroller design.