Report from Dagstuhl Seminar 18041

# Foundations of Data Visualization

**Edited by**

**Helwig Hauser[1], Penny Rheingans[2], and Gerik Scheuermann[3]**

1 **University of Bergen, NO,** helwig.hauser@uib.no
2 **University of Maryland, Baltimore County, US,** rheingan@umbc.edu
3 **Universität Leipzig, DE,** scheuermann@informatik.uni-leipzig.de

**Abstract**

This report documents the program and the outcomes of Dagstuhl Seminar 18041 "Foundations of Data Visualization". It includes a discussion of the motivation and overall organization, an abstract from each of the participants, and a report about each of the working groups.

**Seminar** January 21–26, 2018 – https://www.dagstuhl.de/18041

**2012 ACM Subject Classification** Human-centered computing → Empirical studies in visualization, Human-centered computing → Visualization design and evaluation methods, Human-centered computing → Visualization theory, concepts and paradigms

**Keywords and phrases** Foundations, Interdisciplinary Cooperation, Theory, Visualization

**Digital Object Identifier** 10.4230/DagRep.8.1.100

## 1 Executive Summary

*Penny Rheingans (University of Maryland, Baltimore County, US)*

**License** Creative Commons BY 3.0 Unported license
© Penny Rheingans

Data Visualization is the transformation of data, derived from observation or simulation, and models into interactive images. It has become an indispensable part of the knowledge discovery process in many fields of contemporary endeavor. Since its inception about three decades ago, the techniques of data visualization have aided scientists, engineers, medical practitioners, analysts, and others in the study of a wide variety of data, including numerical simulation based on high-performance computing, measured data from modern scanners (CT, MR, seismic imaging, satellite imaging), and survey and sampled data, and metadata about data confidence or provenance. One of the powerful strengths of data visualization is the effective and efficient utilization of the broad bandwidth of the human sensory system in interpreting and steering complex processes involving spatiotemporal data across a diverse set of application disciplines. Since vision dominates our sensory input, strong efforts have been made to bring the mathematical abstraction and modeling to our eyes through the mediation of computer graphics. The interplay between these multidisciplinary foundations of visualization and currently emerging, new research challenges in data visualization constitute the basis of this seminar.

The rapid advances in data visualization have resulted in a large collection of visual designs, algorithms, software tools, and development kits. There is also a substantial body of work on mathematical approaches in visualizations such as topological methods, feature extraction approaches, and information theoretical considerations. However, a unified description of theoretical and perceptual aspects of visualization would allow visualization practitioners to derive even better solutions using a sound theoretical basis. There are promising ideas

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

but they need further discussion. Currently, we employ user studies to decide if a visual design is more effective, but a comprehensive theory would allow visualization researchers to answer why one visual design is more effective than another and how the visual design can be optimized. Furthermore, we usually have an understanding of the role of a specific visualization in a specific analytic workflow, but we would like to formalize the general role of visualization in the analytic workflow. This would also allow for more quantitative measures of visualization quality. In addition, the community needs a deeper, general understanding of the most informative way to conduct perceptual and usability studies involving domain experts.

For this seminar, we chose to take a focused consideration of the foundations of visualization in order to establish an integrated discussion on the fundamental understanding and generic methodologies of data visualization, including theories, models and workflows of data visualization, evaluation metrics, and perceptual and usability studies. We included experts from all areas of visualization such as scientific visualization, information visualization, and visual analytics to allow for an in-depth discussion of our shared research foundations based on a broad expertise.

With the experience of delivering technical advances over the past three decades, it is timely for the visualization community to address these fundamental questions with a concerted effort. Such an effort will be critical to the long-term development of the subject, especially in building a theoretical foundation for the subject. The community needs to develop suitable models for the whole visualization process from cleaning and filtering the data, analysis processing, mapping to graphical scenes, to the interpretation by the human visual system. While there are some methods of evaluation based on user studies and findings in applications, a complete theoretical foundation for evaluations is missing. Modern visualization includes advanced numerical and combinatorial data processing, so the correctness of this processing including a critical look at its assumptions with respect to the application at hand is needed. Only then, visualization can establish strong correlations between visualization algorithms and questions in the application domains. In addition, uncertainty has received attention from the visualization community in recent years, but a full analysis of uncertainty at all stages of the established visualization pipeline is still not available. Theoretical foundations of uncertainty in visualization need to look at uncertainty in the data, errors due to numerical processing, errors due to visual depiction and, finally, uncertainty in the results based on human misinterpretation of interactive visual depictions.

This workshop addressed five important topics:

Theory of overall visualization process. A theory of the whole visualization process needs to cover all parts of the visualization pipeline and should be applicable to broad classes of application domains. Of course, it is the ultimate foundation, but there are a few formulation attempts and the seminar discussed them. Such a theory should allow to find optimal visualizations and to quantify the value of visualizations. In addition, it is strongly believed by most experts that such a theory needs to cover the challenge of uncertainty in the data, the processing including visual mapping and potential misinterpretation by human observers.

Foundations of evaluation. Evaluation allows designers and analysts to select visualization approaches from among different options for a specific problem. One evaluation method is a user study, usually with a larger group of subjects. Here, it is often a challenge that there is only a very small set of experts available that understand the scientific questions behind the data. Guidelines for user study design in these situations are necessary. In addition, evaluation needs to look at limits of the human visual system. In advanced analytic applications, it is also very important to study the relation between user interest and visualization. There are many open questions in this area that will be discussed in the seminar.

Collaboration with domain experts. Many visualizations address questions and needs from expert researchers, engineers, analysts, or decision makers. Therefore, visualization nearly always involves people outside the visualization community. The seminar included some representatives from large applied research centers so that the discussion about relations between visual data analysis and application semantics was not carried out without domain experts. These participants also commented on methodologies for defining domain requirements and realistic roles of application researchers in evaluation.

Visualization for broad audiences. Visualizations developed for broad audiences involve context and constraints different from those developed for expert domain collaborators. Such visualizations include those for personal information, school use, science centers and other public settings, and communication with a broad general public. Issues with developing visualizations for broad audiences include a higher need for intuitive metaphors and conventions, a larger imperative for drawing participants into interaction, and more requirements for robust interfaces and systems.

Mathematical foundations of visual data analysis. There is a rich tradition of mathematical/computational methods used in visualization, such as topological approaches, mathematical descriptions of feature extraction, numerical sampling and reconstruction methods, integration, differential operators, filtering, dimension reduction, and applications of information theory. In addition, we have seen promising attempts to incorporate uncertainty in these mathematical approaches. While all these methods have a solid mathematical foundation, a careful look at the relation between theories in applications and these mathematical approaches in visual data analysis was taken in this seminar.

The format of the seminar incorporated several elements: overview talks on each topic, clusters of short talks on a single topic followed by a joint panel discussion, and breakout groups on each of the five topics. Unlike the typical arrangement, all presentations in each session were given in sequence without a short Q&A session at the end of each talk. Instead, all speakers of a session were invited to sit on the stage after the presentations, and answer questions in a manner similar to panel discussions. This format successfully brought senior and junior researchers onto the same platform, and enabled researchers to seek a generic and deep understanding through their questions and answers. It also stimulated very long, intense, and fruitful discussions that were embraced by all participants. The breakout groups focused on the general themes and are reported in a later section.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Towards a Theory for Massive, Multidimensional Data Analysis and Visualization

*James Ahrens (Los Alamos National Lab., US)*

**License** Creative Commons BY 3.0 Unported license
© James Ahrens

Sensors and simulations are producing massive amounts of multidimensional data on the order of 10Î2-10Î8 bytes that need to be visualized and understood. The human visual system can process data on the scale of the order of 106̂. Therefore some type of data reduction or sampling is required to produce a visualization.

In this talk, I focus on in situ visualization, visualizing data while it is being generated by a simulation on a supercomputer. Three in situ approaches that use sampling are presented. The first approach, which we refer to as Cinema, conceptually visualizes all results needed while simulation data is in memory for later exploration. Results are generated via rendering a complete Cartesian project of all interesting operators, parameters and camera positions. Results are selected via a set of sliders for the parameters. The second approach extends Cinema to sparse experimental data using parallel coordinates to identify and select the sparse entries. The third approach, proposes the use of a sample-based data representation as a common representation for all data. Each visualization operator inputs and outputs samples. A pipeline-based composition of these operators reduces data to the target size.

## 3.2 Who Are We (in a Collaboration)?

*Johanna Beyer (Harvard University - Cambridge, US)*

**License** Creative Commons BY 3.0 Unported license
© Johanna Beyer

Visualization researchers and practitioners face many challenges when collaborating with domain experts. In particular, the different roles of visualization researchers as compared to visualization engineers or practitioners have a huge influence on the goals and measures for success for a collaboration. While visualization researchers should focus on novel algorithms, tools, and ultimately publications in visualization-related fields, the main focus of visualization engineers generally lies in creating usable software tools that are used beyond the initial prototype stage. Therefore, these different roles should be explicitly addressed at the beginning of a collaboration, to avoid common pitfalls and differing expectations between collaborating visualization and domain experts.

## 3.3 A Model of Spatial Directness in Interactive Visualization

*Stefan Bruckner (University of Bergen, NO)*

**License** Creative Commons BY 3.0 Unported license
© Stefan Bruckner

The ability to interactively explore a visual representation is a core aspect of all visualization systems. The term “directness”, as in “direct manipulation”, is commonly used to discuss properties of interaction techniques in the context of visualization. Unfortunately, the terms referring to the directness of spatial interaction are largely used by intuition and without a clear definition. In this talk, I introduce a model of directness in interactive visualization that characterizes it as an emerging property of the involved mapping processes, from the data space to the perception and cognition of the user. Based on such a formulation, we can further proceed to quantify the different dimensions of directness, leading us to an approach that forms the basis of formulating testable predictions for visualizations that may ultimately allow us to perform in-silico user studies and even allow the synthesis of novel visualization methods based on different objective functions.

## 3.4 Color, Math, and Visualization

*Roxana Bujack (Los Alamos National Laboratory, US)*

**License** Creative Commons BY 3.0 Unported license
© Roxana Bujack

Perceptual scientists' experiments indicate that human color perception is non-Euclidean, which induces new challenges on colormap design. How can we generalize methods for the evaluation, optimization, generalization and interpolation of colormaps?

## 3.5 Visualization of Climate Projections for Communication to the Public

*Michael Böttinger (DKRZ Hamburg, DE)*

**License** Creative Commons BY 3.0 Unported license
© Michael Böttinger

Increasing public attention to climate and climate change triggers a demand by the media, policy makers and the general public for meaningful visualizations showing key outcomes of future climate simulations. At the German Climate Computing Center (DKRZ), visualizations of IPCC simulations have regularly been produced for more than 20 years. One of the keys for successful visualization is simplicity. However, to help recipients of these visualizations in identifying the main outcomes, accompanying annotation in the form of text or narration proved to be useful.

In this talk, several examples of successful visualizations are discussed which had been adopted by various media. The latest example refers to one of the key visualizations of the IPCC AR5 summary for policymakers that shows the temperature change of the CMIP5 multi model ensemble with two levels of robustness overlaid by stippling and hatching. We present an alternative, simplified and animated version for the same data set that draws the attention of the viewer to robust areas by dehighlighting non-robust areas. In this way, the viewer's focus is guided to the trustworty part of the data.

## 3.6 My Math Keeps Breaking!

*Hamish Carr (University of Leeds, GB)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Hamish Carr

Visualization relies heavily on mathematics, both because the input data is often defined mathematically, and because the mathematics is the tool that we use to describe the stages of data processing in our data pipelines. However, the complexity of our pipelines and the nature of the mathematical computations we perform causes increasing problems in our mathematics. One way this occurs is that different stages in the computation are often handled by different people, with different mathematical assumptions. In the worst case, their mathematical assumptions are irreconcilable, but even when they are formally reconcilable, their cumulative effect is to make the overall computation unreliable. Moreover, much of the mathematics we use has formal assumptions that are computationally difficult or impossible to guarantee, leading to the need for new mathematics.

## 3.7 Empirical Studies in Visualization

*Min Chen (University of Oxford, GB)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Min Chen

In the field of visualization, empirical studies are typically conducted under the major scope of "Evaluation". The emphases have typically been placed on "testing" some visual designs or visualization systems as part of a software engineering workflow. While empirical studies can and should support "evaluation" in visualization, there have not been enough emphases given to the more ambitious goal of empirical studies, that is, to make new discoveries about how and why visualization works in some conditions and not in others, and to inform and verify proposed theories advances.

Most of us agree that in some circumstances, visualization is more effective and/or efficient than viewing data in numerical, textual, or tabular forms, and than being simply informed by a computer about the decision. When visualization works in these circumstances, there must be some merits in perception and cognition. Hence any causal factors that make visualization work may potentially be the causal factors that make perception and cognition work. Therefore, visualization researchers are in the right place at the right time to look for these causal factors. For example, can the cost-benefit metric proposed by Chen and Golan also be the fitness function for the development or evaluation of some perceptual and cognitive capabilities (e.g., visual search, selective attention, gestalt grouping, heuristics, and memory)?

## 3.8 Topological data analysis and topology-based visualization

*Leila De Floriani (University of Maryland - College Park, US)*

**License** Creative Commons BY 3.0 Unported license
© Leila De Floriani

The talk deals with common topological and algorithmic tools to two Topological Data Analysis and topology-based visualization. Specifically, it focuses on an algebraic topology tool, Discrete Morse Theory (DMT), applied in both disciplines, and its relation with persistent homology. New developments in dealing with multivariate data which led to multi-parameter persistent homology in TDA are discussed as well a new approach for computing multi-parameter persistent homology and its possible applications to critical feature extraction in multifield data analysis and visualization.

## 3.9 Fundamental Mathematics in Visualization

*Christoph Garth (TU Kaiserslautern, DE)*

**License** Creative Commons BY 3.0 Unported license
© Christoph Garth

Mathematical concepts, methods, and tools have played a key role in the development of a large variety of visualization techniques. This raises the question, which mathematical techniques should visualization researchers be familiar with. In my talk, I will report on an informal survey of the mathematical underpinnings of the past decade of visualization research, and examine its implications for the education of students.

## 3.10 Adjust, Just Adjust

*Eduard Gröller (TU Wien, AT)*

**License** Creative Commons BY 3.0 Unported license
© Eduard Gröller

Developing visualizations for broad audiences requires glanceable graphics and graspable interactions. This talk will concentrate on interaction facilitation through automatic adjustments. The first example illustrates an automatic color scale adjustment in a biomolecular setting to accommodate contradicting and overlapping color schemes across scales. The second example discusses output-sensitive interaction to make changes in the input proportional to changes in the output, or to visually indicate the sensitivity of input changes with respect to output changes. The third example deals with visualization of 4D ultrasound data, which is targeted to a broad audience in prenatal imaging and diagnosis. Lessons learned during this project are presented. The talk makes a case for automatically reducing interaction complexity in visualizations for broad audiences.

## 3.11 Visualize Insight??

*Hans Hagen (TU Kaiserslautern, DE)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Hans Hagen

One purpose of data visualization is to help the viewer to obtain insight. But how does insight emerge from data? Is insight part of the visualization? Can we somehow characterize the insights to be found in a visualization?

## 3.12 Effective Collaboration with Domain Experts: FluoRender

*Charles D. Hansen (University of Utah - Salt Lake City, US)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Charles D. Hansen

Effective collaboration with domain experts requires knowledge, joint interest, and coordination to achieve joint scientific goals. FluoRender is an example of close collaborations with biologists and visualization experts that resulted in a widely used visualization tool that has contributed results to the visualization community and enabled scientific results in the biology field. There are several lessons that can be learned from this collaboration. First, communication is key and a common language and vocabulary is fundamental. Both parties, visualization and domain experts, should accomplish scientific contributions in their respected fields. Close collaboration requires detailed application knowledge by the visualization research and visualization knowledge by the domain expert. It is important not to ask the domain expert what problems need solving or which features are required. It is better to understand the domain scientist's workflow by spending time in their research laboratory, work closely with them and do not limit interaction to meetings and discussions. Lastly, it is important to be creative, have fun, and collaborate.

## 3.13 About the scales and limits of visualization

*Helwig Hauser (University of Bergen, NO)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Helwig Hauser

Discussing what works in visualization, and what not, can be done in terms of several principal aspects of influence. At the side of the user, perceptual and cognitive aspects of the non-uniform human visual system are important, enabling (and also limiting) visualization. Then, of course, the extent of the data has a major influence and there is a certain range of extent that lends itself to visualization solutions. Similarily, the richness of the data, for example, in terms of multivariate data is criticial. Thirdly, in this respect, the dimensionality of the data leads to major differences – a few dozens of dimensions are very different from hundreds! Last, but not least, a more technical perspective is impoortant: good hardward and good software. All in all, these aspects possibly form a space for visualization solutions.

## 3.14 Benefits of and Questions to a Theory of Visualization

*Hans-Christian Hege (Zuse Institute Berlin, DE)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Hans-Christian Hege

In this talk, "theory" does mean no single universal theory (which might not be achievable) but a bundle of theories. What would the benefits of such a theory building be and which fundamental questions should it help to answer?

The practical benefits are obvious: Qualitative, conceptual models could help us to describe, understand and reason about visualization processes and provide hints, which visual representations, analysis actions, and work flows are more efficient than others. Quantitative models would help us, to make predictions about quantitative dependencies in visualization processes and help us to optimize mathematically components of visualization processes.

Beside that there are strategic benefits; in particular, a common core theory would be an effective countermeasure to the danger of fragmentation of data visualization. It would also increase its survival capability in the landscape of competing disciplines.

The list of fundamental questions to be answered, is long. Here are some, commented in the talk: What is visualization and what is it for? What is information? What is the solution of the data–information–knowledge conundrum? What are the elementary knowledge units? How can prior knowledge be captured in detail? What are the elementary acts of reasoning, given new external information? What are the elementary acts in which humans increase their knowledge using external (nonvisual/visual) information? What leads to the emergent phenomenon of a eureka moment? How are external and mental images related? What is the role of mental images in reasoning? What are the limits of visualization? If we have (better) answers to such questions: How can the theories be made operational and how can they be practically utilized?

Almost all these fundamental questions can be answered only in collaboration with other sciences, especially cognitive sciences. That will not happen, if it is not initiated by us.

## 3.15 Theory of Visualization and Domain Experts

*Mario Hlawitschka (Hochschule Leipzig, DE)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Mario Hlawitschka

Working with "domain scientists" may be challenging, especially when they come from different domains and a scientific language barrier must be lowered. On the first day of setting up a project, the group should be aware of what they can expect from their partners. Even in the field of visualization, the definition of visualization is rather vague and the first step of a fruitful collaboration is to explain the potentials of visualization. A "theory of visualization" could aid in finding good definitions of the process of visualization and its potentials. Guidelines should be derived from that, which should be used by domain experts as well as visualization designers and researchers. An example is "information theory" where a profound basis has been set in a very specific topic, which is now used in a much broader way. Such building blocks, both algorithmically but also as parts of a theory (or theories) of visualization lay a foundation and correctly applied, may lower that entry barrier. Ultimately, this may lead to one or many "theories of visualization" that may be a foundation to impact many other fields of research.

## 3.16 What I am thinking about when I am biking to work: Spaces – mappings – projections

*Ingrid Hotz (Linköping University, SE)*

**License** Creative Commons BY 3.0 Unported license
© Ingrid Hotz

This talk reasons about scientific knowledge discovery process as a sequence of mappings from and to various spaces. Spaces involved in such a pipeline could be defined by models, data, application areas, or humans exhibiting a certain experience. The mapping between these spaces can be of different nature preserving the dimension or reducing the dimensions describing projections. A carful design of these spaces, their parametrization and the mappings between is essential for the success of the process. Within the visualization pipeline one can exemplarily consider the data space representing the data according to some model space, the space spanned by relevant questions in one application, and a space used by the visualization.

## 3.17 Pathways for Theoretical Advances in Visualization

*Christopher R. Johnson (University of Utah - Salt Lake City, US)*

**License** Creative Commons BY 3.0 Unported license
© Christopher R. Johnson

In my 2004 article, Top Scientific Visualization Research Problems, I proposed creating a Theory of Visualization as a top research problem. Since 2004, there has been some progress in theoretical aspects of visualization, but much more needs to be done in this area. In 2017, Min Chen lead a co-author team of M. Chen, G. Grinstein, myself, J. Kennedy, and M. Tory who proposed Pathways for Theoretical Advances in Visualization [1]. We hope that many visualization researchers will contribute to this foundational area within visualization.

**References**

1 Chen, M., Grinstein, G., Johnson, C. R., Kennedy, J., and Tory, M. Pathways for theoretical advances in visualization. IEEE Computer Graphics and Applications, 37(4):103–112, 2017.

## 3.18 Empirical Studies with Domain Experts

*Alark Joshi (University of San Francisco, US)*

**License** Creative Commons BY 3.0 Unported license
© Alark Joshi

Tool adoption by domain users is a strong measure of success when working with domain experts. Working with domain experts requires deep, longer conversations that go from learning each others language to working closely on prototypes to help solve their problem. When working with atmospheric physicists, we developed a system to predict hurricane dissipation and even though we conducted formative and summative evaluations, it was eventually not adopted for regular use. In our collaboration with neurosurgeons, we developed a tool that works with an image-guided navigation system. We conducted various empirical studies to evaluate the use of a novel interaction technique for multimodal visualization, applying existing visualization techniques for vascular visualization, and so on. Empirical studies can truly help you learn about specific aspects in your system/technique. I believe that empirical studies should not be an afterthought and working with human factors experts can help us design better studies to learn from them.

## 3.19 Making sense of Math in Vis

*Gordon Kindlmann (University of Chicago, US)*

**License** Creative Commons BY 3.0 Unported license
© Gordon Kindlmann

Visualization research sometimes has a complicated relationship to mathematics. Many accounts of data visualization do not include a presentation or discussion of the underlying mathematics employed. When there is math, it can come myriad forms. The types of mathematics used for one type of research may or may not be similar to those for other research: the linear algebra for tensor visualization is distinct from the statistics used to measure the results of user studies. This talk attempts to locate the places *in* visualization where math arises, as well as outlining some recent work on the math *of* visualization. The necessity of math in visualization will likely remain an ongoing topic of consideration and debate.

## 3.20 Data-driven Storytelling at NASA

*Helen-Nicole Kostis (USRA/GESTAR SVS NASA/GSFC, US)*

**License** Creative Commons BY 3.0 Unported license
© Helen-Nicole Kostis

In this talk, I will provide an overview of the storytelling efforts at NASA Goddard Space Flight Center. The goal of the Scientific Visualization Studio (SVS) is to promote greater understanding of NASA science programs through visualization. The products of the visualization efforts are data-driven high quality computer graphics animations that are developed and produced in collaboration between producers, science writers, visualization experts and scientists. NASA's heartbeat are the scientific results and engineering accomplishments. Through the years, data-driven visualizations from the Scientific Visualization Studio have clearly become a critical component on leading outreach, education and science communication efforts.

## 3.21 Collaboration with the Domain Experts - molecular visualization

*Barbora Kozlíková (Masaryk University - Brno, CZ)*

**License** Creative Commons BY 3.0 Unported license
© Barbora Kozlíková

Understanding the structure and behavior of protein molecules is crucial in many biological and biochemical, such as drug design and protein engineering. This process requires studying the proteins from many aspects, including their constitution, physico-chemical properties, temporal behavior, or interactions with other molecules. Observing that by traditional approaches, i.e., animation of the 3D structural model, is not feasible anymore, due to the amount of data to be processed. Therefore, specialized visualization techniques have to be involved into the exploration process. The talk covers short introduction to the domain problem and then focuses mainly on the experience in collaboration with the experts and lessons learned.

## 3.22 Accidental Broad Audiences in Virtual Reality Visualization

*David H. Laidlaw (Brown University - Providence, US)*

**License** Creative Commons BY 3.0 Unported license
© David H. Laidlaw

Over five minutes I will share some of the lessons learned showing several thousand audience members our large-scale virtual reality display and, within it, several scientific and academic applications we have developed. In particular, the short, pithy messages that are appropratie for broad audiences contrast with the more exploratory or formative activities that occur with our scientific research tool development. This has implications on the design of tools and how they are presented and used.

## 3.23 Foundations of visualization - Where we stand and where to go

*Heike Leitte (TU Kaiserslautern, DE)*

**License** Creative Commons BY 3.0 Unported license
© Heike Leitte

A theory is a set of scientifically founded statements used to describe a part of the world and make predictions about it. In the visualization field a number of such theories have been published over the last decades that help understand different aspects of the visualization process. Their validity, interrelationships, and impact have been discussed in a number of panels, but summarizing papers giving an overview over theories in and of visualization are scarce. Hence, it is time to join forces and structure the presented ideas, identify shortcomings, and think about future directions.

## 3.24 Empirical Studies on Human-in-the-Loop

*Ross Maciejewski (Arizona State University - Tempe, US)*

**License** Creative Commons BY 3.0 Unported license
© Ross Maciejewski

Currently, a large variety of empirical studies in information visualization have provided insights into how people perceive information, what the just noticeable differences are, response times, etc. However, less work has focused on understanding the use of knowledge being generated. This talk discusses issues in knowledge generation, open challenges, and the notion of algorithmic aversion and its potential relationship to visualization.

## 3.25 Activity-Centered Domain Characterization

*Georgeta Elisabeta Marai (University of Illinois - Chicago, US)*

**License** Creative Commons BY 3.0 Unported license
© Georgeta Elisabeta Marai

Domain characterization is the first stage of the visualization process. Activity Theory helps lay an activity-centered foundation for this stage. In a departure from existing visualization models, this approach assigns value to a visualization based on user activities; ranks user tasks above user data; partitions requirements into activity-related capabilities and nonfunctional characteristics and constraints; and explicitly incorporates user workflows into the requirements process. A quantitative evaluation supports the merits of the activity-centered model and leads to several questions regarding the sparsity of the vis theoretical landscape, and about the evaluation models we use for theories.

## 3.26 Empirical Studies in Visualization

*Kresimir Matkovic (VRVis - Wien, AT)*

**License** Creative Commons BY 3.0 Unported license
© Kresimir Matkovic

Empirical studies represents a well-established research field. Visualization researchers are often required to evaluate their research results. Empirical studies represent a possibility of evaluation. However, they are particularly suitable for well-defined tasks which can be easily quantizable (how long does it take to do something, what is in front what is behind, etc.). Such low level tasks are useful in visualization, but typical tasks are mostly more complex. How much knowledge is gained, what insights are gained, etc. Providing quantitative measures for such questions is not easy. This is why some of the visualization researches are not so enthusiastic with evaluation. A possible solution is to identify basic tasks and to test it by means of an empirical study. Another requirement which is often posed is to evaluate a specific technique developed for a specific domain in a user study and or to generalize it. Both requirements are not easy to fulfill. The experts are rare, so we cannot find enough of them for a proper user study. If we generalize it, we need a lot of users again. This might require additional resources (time) which are not always available. Finally, we often base evaluation on tasks abstraction. The data and user abstraction is usually neglected. Further, the tasks are rarely compared with similar tasks from peers' research. We argue, it is necessary to base the evaluation on abstraction of tasks, data, and users. Having a list of tasks, data and users with corresponding solution would be a valuable contribution to the visualization community.

## 3.27 Theory of Visualization Process: Survey? Overview? Challenges and Opportunities?

*Silvia Miksch (TU Wien, AT)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Silvia Miksch

Is there any (unified) theory of the visualization process, which is "somehow" accepted by the visualization communities? In this talk I present some definitions of visualization, with particular focus on the process characteristics, and various models in visualization to identify possible challenges and opportunities. Definitions and models on various levels of abstraction, functionality, and complexity exits, but no real unified ones. I propose a conceptual model of knowledge-assisted visual analytics incorporating the role of explicit knowledge as well as characterizing guidance in visual analytics.

**References**

1 Federico, P., M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics. Proceedings of the IEEE Conference on Visual Analytics Science and Technology (IEEE VAST 2017). (http://www.cvast.tuwien.ac.at/node/785)

2 Ceneda, D., T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, and C. Tominski. Characterizing Guidance in Visual Analytics. IEEE Transactions on Visualization and Computer Graphics 23(1):111–120, 2017. (http://www.cvast.tuwien.ac.at/node/765)

## 3.28 Bridging the gap between domain experts and data analysts

*Daniela Oelke (Siemens AG - München, DE)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Daniela Oelke

An analyst needs an analysis question to work with. This question is provided by the domain expert, but has to be translated into a (more general) data analysis question before a data analyst can work with it. This process requires the data analyst and the domain expert to work closely together and can be challenging.

In my talk I presented experiences from industry of what has proven useful to bridge this gap including educating the domain expert, interviewing domain experts in the right way, and using visual analytics to facilitate the communication.

Furthermore, I pointed out that in order to have an impact, additional stakeholders or domain experts have to be included in the process in a company such as customers, sales representatives, management, etc. I reported on an experimental project in which all these stakesholders were working together for a week combining methods of visual data analytics and business innovation to come up with ideas for novel business opportunities.

## 3.29 I work with Experts

*Kristi Potter (NREL - Golden, US)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Kristi Potter

This proposal aims to close the loop on the traditional flow of knowledge by relating simulation and analysis results to a conceptual model. This new framework will relate relevant pieces of scientific workflow, including analysis results, uncertainty information, and background knowledge, to help to solve the intractable data problem faced by exascale computing by explicitly conveying relationships between complex computational systems, large-scale data, and theoretical scientific concepts.

## 3.30 A critical analysis of evaluation in medical visualisation

*Bernhard Preim (Universität Magdeburg, DE)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Bernhard Preim

Medical visualisation is typically evaluated anecdotical only. In addition there are some perception-based studies related to shape and depth perception. These are serious quantitative evaluation but they are limited to understanding low level perceptual issues and not the high level cognitive activities like the decision-making and problem solving in diagnosis, treatment planning and medical education. Eye-tracking studies, think about, interaction protocols and long term case studies are needed to better understand what works in medical visualisation.

## 3.31 Mathematical Foundations in my Work

*Gerik Scheuermann (Universität Leipzig, DE)*

**License** Ⓒ Creative Commons BY 3.0 Unported license
© Gerik Scheuermann

In a short position statement for the mathematical foundations panel, I name the areas of mathematics that I have used in the past. From the insight that pretty much all areas of mathematics have been used to some extent for visualization purposes in the literature, I raise the question which areas should be part of a curriculum and which areas are just optional depending on the specific material covered in class. Besides, I also pointed at the problem that the meaning of mathematical concepts behind visualization algorithms does not fit the applications in some cases, leading to unsatisfying results.

## 3.32 Collaborating with Domain Experts

*Marc Streit (Johannes Kepler Universität Linz, AT)*

**License** Creative Commons BY 3.0 Unported license
© Marc Streit

In the first part of my talk, I summarize what advice researchers and practitioners can get from a theory of visualization. We – as a community – currently provide advice by publishing models and theories, by collecting techniques and methods, and by describing best practices. While this is very useful, it is often not actionable. A less explored possibility is to provide cheat sheets in the form of decision trees that can help practitioners to create effective visualizations. These decision trees could be created as a community effort, underpinned with our models, and carefully annotated. In the second part, I talk about why generalizing design studies is hard, why data and task abstraction is key to create impact in visualization through collaboration with domain experts, and what lessons I've learned in previous collaborations.

## 3.33 Mathematical Foundations of Visualization – Different Kinds

*Holger Theisel (Universität Magdeburg, DE)*

**License** Creative Commons BY 3.0 Unported license
© Holger Theisel

Which mathematical foundations should we expect from Visualization experts? There are two kinds: foundations for all Visualization experts, and foundations that only a few visualization experts work with. This is fine: it is an established way of Data Vis development to constantly discover (not invent) new mathematical theories for Visualization, and to make them useful and applicable for visualize concrete data. Further, visualization experts can and should contribute in developing the foundations We should not wait for the results and "only" visualize them then! This is, however, an approach limited to only a few problems in visualization (perhaps the ones visualization is most matured)

## 3.34 Bringing your research to broad audiences

*Jarke J. van Wijk (TU Eindhoven, NL)*

**License** Creative Commons BY 3.0 Unported license
© Jarke J. van Wijk

Having your research published at one of our venues is not an endpoint. It can be highly useful and rewarding to bring it to broader audiences. I describe one of my experiences in this. After having developed the cushion treemap technique (1999), I had a student integrate that in a tool just for hard disk visualization, Sequoiaview. That attracted much attention, and led to generalization of the method and a start-up company, MagnaView. One of their successes was a tool for visualizing high school data, which is used on a large scale. To be succesful in this, dedication to the needs and wants of the audience and careful tuning of presentation and interaction is crucial.

## 3.35 Domain Expert Collaboration: when it went well

*Anna Vilanova (TU Delft, NL)*

**License** Creative Commons BY 3.0 Unported license
© Anna Vilanova

I present an example in which the collaboration was a success according to my definition of success.

Key factors:

- New data that the domain experts could not analyze without visualization aid.
- A problem that suits the visualization field and has challenges that are unsolved in the vis community.
- Two vis people: Have a person in the project just focused on the development of general Vis techniques that are inspired but not directly application dependent. Have another person between domains that transforms advances to an adapted framework that can be used by the domain experts.
- Engineering factor needed was limited
- Funding was available, with quite some freedom on how to use it.
- Great respect, and effort to understand each others field.
- Talented, communicative and enthusiastic people involved.

## 3.36 On Visual Abstraction

*Ivan Viola (TU Wien, AT)*

**License** Creative Commons BY 3.0 Unported license
© Ivan Viola

Visual abstraction is a fundamental concept in visual arts and data visualization. While we have an intuitive understanding what the term "visual abstraction" stands for, there is no consensus. Abstract, originating from Latin abstrahere, means drawn away, and is often used in terms like abstract data, abstract class, abstract art, where it represents aspects that are derived from a concrete corresponding object. Abstraction is a process and also an outcome of that process. Visual abstraction is therefore a process of abstraction, where information is transformed into visual representations. We can recognize multiple fundamentally different directions of visual abstractions: geometric abstraction, photometric abstraction, and temporal abstraction.

## 3.37 Vis4Vis: Visualization in Empirical Visualization Research

*Daniel Weiskopf (Universität Stuttgart, DE)*

**License** Creative Commons BY 3.0 Unported license
© Daniel Weiskopf

Appropriate evaluation in visualization research is a longstanding, relevant, and often-discussed issue. My talk focuses on empirical studies with user involvement. I argue that one of the underlying difficulties is the varying role of visualization research: it has facets of engineering and (natural) science, depending on the research objective at hand. We may adopt study methodology from other fields, such as psychology or HCI, but have to be careful to adapt them to the specific needs of visualization research. One promising direction is the use of data-rich observations that we can acquire during studies in order to obtain more reliable interpretations of empirical studies. For example, we have been witnessing an increased availability and use of physiological sensor information from eye tracking, EEG, and other modalities as well as user logging. Such data-rich empirical studies promise to be especially useful for studies "in the wild" and similar scenarios. However, with the growing availability of large, complex, time-dependent, heterogeneous, and unstructured observational data, we are facing the new challenge of how we can analyze such data. I argue that we need Vis4Vis: visualization as a means of data analysis of empirical study data to advance visualization research.

## 3.38 Data transformations, embeddings, summaries

*Ross Whitaker (University of Utah - Salt Lake City, US)*

**License** Creative Commons BY 3.0 Unported license
© Ross Whitaker

Fundamentally, data visualization is the process of placing dabs of ink or color on a 2D plane. However, the complexity of data is increasing so that we see large numbers of instances, dimensions, paramaters, etc. Such data surpases what can readily shown on a 2D or 3D display. One solution to this challenge is the development of better or more complex interfaces, that include, for instance, linked views, large displays, dynamic visualizations, and sophisticated user interactions. The alternative and complementary approach is to develop sets of mathematical and statistical tools to transform, map, or summarize data and reduce its complexity so that visualization and understanding of large, complex becomes more feasible. The role of visualization research, in this case, is to identify common use cases and develop methods and tools that can readily be adapted to particular applications.

## 3.39 Trust in Visualization (and what it has to do with Theory)

*Thomas Wischgoll (Wright State University - Dayton, US)*

**License** Creative Commons BY 3.0 Unported license
© Thomas Wischgoll

There are different issues with trust involved when working with domain experts to visualize their data. There may be limitations with the data that require special precautions, such as sensitivity or security limitations. It may have taken a lot of effort to collect or create the data so that a certain level of trust is required for the domain expert to share the data. At the same time, the domain expert needs to be able to trust in the final visualization results. This presentation discusses these issues with trust and what requirements for a theoretical foundation this results in. Furthermore, additional requirements are discussed for user interfaces and other elements within the visualization.

## 3.40 Exploranation

*Anders Ynnerman (Linköping University, SE)*

**License** Creative Commons BY 3.0 Unported license
© Anders Ynnerman

This presentation discusses visualization approaches to reach broad audiences. The area is wide and includes aspects of infographics, science communication, interfaces for human in the loop applications, and indeed specific visualization for large groups of domain experts. This presentation introduces the confluence of exploratory and explanatory visualization denoted "Exploranation" as a means to reach large user groups with engaging visualization. Examples are give from the field of science communication at public venues and presents derived design principles for interactive installations in museums as well as requirements and challenges for mediated visual science communication. The presentation is concluded with reflections on the need for visualization in human in the loop applications such as autonomous systems and presents a visions for visual cognitive companions.

## 3.41 Using Empirical Results in Practice

*Caroline Ziemkiewicz (Forrester Research Inc., US)*

**License** Creative Commons BY 3.0 Unported license
© Caroline Ziemkiewicz

There is a growing and welcome tendency in the visualization community to reflect on how and why to perform empirical studies, particularly user evaluations. For this process of reflection to be productive, it is necessary to consider the various audiences of empirical results and what they need. One important such audience includes visualization practitioners and designers. Practitioners use empirical research results to support decisions about what techniques to use for an application and how to tell whether a design is effective. Generalizing and making use of results in this way requires a full understanding of the context in which the study was performed: task abstractions, user models, assumptions, and tested requirements. Many common methods of designing and reporting empirical studies in visualization lack this context, particularly in system evaluation and technique comparisons. New approaches and methods are needed to make this context concrete and produce results that are specific enough to be generalized.

# 4 Working groups

There were five working groups for the five central topics, i.e.

- Theory of overall visualization process
- Foundations of evaluation
- Collaboration with domain experts
- Visualization for broad audiences
- Mathematical foundations of visual data analysis

One key target product of the workshop was an edited volume on Foundations of Data Visualization. The five working groups explored their topics and organization for sections of that planned book. Seminar participants were surveyed before the seminar about their level of interest in each working group topic and assigned to working groups based on those interests. Each working group developed a plan for the creation of their book section over the months following the seminar. By the end of the seminar, working groups had identified chapters and authors for the section, as well as a schedule for authoring and review. Work on the chapters themselves continues after the conclusion of the seminar.

## Participants

- James Ahrens
Los Alamos National Lab., US
- Johanna Beyer
Harvard University – Cambridge, US
- Michael Böttinger
DKRZ Hamburg, DE
- Stefan Bruckner
University of Bergen, NO
- Roxana Bujack
Los Alamos National Laboratory, US
- Hamish Carr
University of Leeds, GB
- Min Chen
University of Oxford, GB
- Leila De Floriani
University of Maryland – College Park, US
- Christoph Garth
TU Kaiserslautern, DE
- Eduard Gröller
TU Wien, AT
- Hans Hagen
TU Kaiserslautern, DE
- Charles D. Hansen
University of Utah – Salt Lake City, US
- Helwig Hauser
University of Bergen, NO
- Hans-Christian Hege
Zuse Institute Berlin, DE
- Nathalie Henry Riche
Microsoft Research – Redmond, US
- Mario Hlawitschka
Hochschule Leipzig, DE
- Ingrid Hotz
Linköping University, SE
- Christopher R. Johnson
University of Utah – Salt Lake City, US
- Alark Joshi
University of San Francisco, US
- Gordon Kindlmann
University of Chicago, US
- Helen-Nicole Kostis
USRA/GESTAR SVS NASA/GSFC, US
- Barbora Kozlíková
Masaryk University – Brno, CZ
- David H. Laidlaw
Brown University – Providence, US
- Heike Leitte
TU Kaiserslautern, DE
- Ross Maciejewski
Arizona State University – Tempe, US
- Georgeta Elisabeta Marai
University of Illinois – Chicago, US
- Kresimir Matkovic
VRVis – Wien, AT
- Laura A. McNamara
Sandia National Labs – Albuquerque, US
- Silvia Miksch
TU Wien, AT
- Torsten Möller
Universität Wien, AT
- Daniela Oelke
Siemens AG – München, DE
- Kristi Potter
NREL – Golden, US
- Bernhard Preim
Universität Magdeburg, DE
- Penny Rheingans
University of Maryland, Baltimore County, US
- Gerik Scheuermann
Universität Leipzig, DE
- Marc Streit
Johannes Kepler Universität Linz, AT
- Holger Theisel
Universität Magdeburg, DE
- Jarke J. van Wijk
TU Eindhoven, NL
- Amitabh Varshney
University of Maryland – College Park, US
- Maria Velez-Rojas
CA Technologies – Santa Clara, US
- Anna Vilanova
TU Delft, NL
- Ivan Viola
TU Wien, AT
- Daniel Weiskopf
Universität Stuttgart, DE
- Ross Whitaker
University of Utah – Salt Lake City, US
- Thomas Wischgoll
Wright State University – Dayton, US
- Anders Ynnerman
Linköping University, SE
- Caroline Ziemkiewicz
Forrester Research Inc., US