Foundations of Data Visualization

Min Chen · Helwig Hauser · Penny Rheingans ·
Gerik Scheuermann
Editors

Foundations of Data Visualization

Springer

Editors
Min Chen
University of Oxford
Oxford, UK

Penny Rheingans
University of Maryland, Baltimore County
Baltimore, MD, USA

Helwig Hauser
Department of Informatics
University of Bergen
Bergen, Norway

Gerik Scheuermann
Institut für Informatik
Universität Leipzig
Leipzig, Germany

ISBN 978-3-030-34443-6 ISBN 978-3-030-34444-3 (eBook)
https://doi.org/10.1007/978-3-030-34444-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Foundations of Data Visualization should become required reading for visualization researchers. In my Visualization Viewpoints article on Top Scientific Visualization Problems,[1] I encouraged visualization researchers to engage in what Bill Hibbard called "foundational problems" in visualization.[2] Researchers responded with exciting new papers describing theoretic frameworks, new conceptual models, ontologies, and taxonomies.[3] The editors of *Foundations of Visualization*, Min Chen, Helwig Hauser, Penny Rheingans, and Gerik Scheuermann, have assembled a who's who among international visualization researchers contributing to foundational visualization problems.

Foundations of Data Visualization begins by introducing important basic concepts related to visual abstractions, what measurement means in visual spaces, and knowledge-assisted models, including an information-theoretic perspective. The book then moves on to explore the fundamental mathematical and computer science underpinnings of visualization, illuminating many essential links between computer science and mathematical theory and visualization theory.

An observation variously attributed to Yogi Berra, Albert Einstein, Richard Feynman, and Jan L. A. van de Snepscheut goes, "In theory, there is no difference between theory and practice. But, in practice, there is." In a nod to this paradoxical truth, the book concludes with several chapters on empirical studies in visualization and multiple "real-life, feet on the ground" examples of visualization collaborations with university and industrial domain, scientific, engineering, and medical researchers, reminding us how visualization in its practice reaches into and influences virtually all disciplines and aspects of our lives.

[1]C. R. Johnson: Top Scientific Visualization Research Problems; IEEE Computer Graphics and Applications 24(4): Visualization Viewpoints, pp. 13–17, July/August 2004.

[2]B. Hibbard: Top Ten Visualization Problems; Proc. ACM SIGGRAPH 33(2), ACM Press, 1999, pp. 21–22; https://doi.org/10.1145/326460.326485.

[3]M. Chen, J. Kennedy: References for Theoretic Researches in Visualization; https://sites.google.com/site/drminchen/themes/theory-refs, 2017.

Likewise, as da Vinci understood the need for practitioners to study their own practices, whether the art of science or the science of art, so too did he comprehend the need to theorize those practices in order to understand them and hence to strengthen them. I cannot help but agree with him when he says, "He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may cast."

This book reminds us that, as practice evolves, theories of visualization emerge and themselves evolve. The book *Foundations of Data Visualization* proposes many "stakes in the ground" for future discussion and debate. I encourage all visualization researchers to read this volume and to engage and further the discussion—even to propose new theoretical ideas—on foundational visualization research.

August 2019

Chris R. Johnson
University of Utah
Salt Lake City, USA

Preface

Data visualization is the user- and task-oriented transformation of data from measurement or simulation, as well as from models (empirically crafted or machine-learned) into interactive images for exploration, analysis, and presentation. It has become an indispensable central 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 dealing with a wide variety of data. One of the powerful strengths of data visualization is the effective and efficient utilization of the human sensory and cognitive system to enable and support instructive exploration, complex analysis, and critical decision making, through the recognition of relevant patterns, the observation of unseen relations, and the identification of new connections with other data and complementing facts, concepts, theories, goals, and opinions, which are known to the users. Since vision dominates our sensory system, a significant amount of effort has been made to bring meaningful abstractions or other useful information to our eyes through interactive computer graphics. The foundations of data visualization should therefore address the fundamentals of visualization techniques, the intrinsics of visualization processes, the conceptualization of visualization users, their mind and tasks, and the principles of developing visualization applications. The interplay of these multidisciplinary foundations of data visualization and currently emerging, new research challenges in visualization constitute the broader basis of this book.

As the title indicates, this book focuses on the foundations of visualization as seen by about fifty experts from all areas of visualization, including scientific visualization, information visualization, and visual analytics, providing an in-depth discourse on a wide range of foundational topics, based on their broad expertise. The rapid advances in data visualization have resulted in a large collection of visual designs, algorithms, software tools, and development kits. We also commonly refer to a substantial body of work on mathematical methods in data visualization such as topological methods, feature extraction techniques, and information-theoretic solutions. However, we are still lacking a widely accepted and unified description of theoretical, perceptual, and cognitive aspects of visualization that would

allow visualization practitioners to derive even better solutions—facilitated by a sound theoretical basis. With this book, we identify promising, related ideas and contribute to their further discussion, evaluation, validation, or falsification. Currently, many visualization researchers and developers employ empirical studies to decide if a visual design is more effective. They could benefit from a comprehensive theory that answers why one visual design is more effective than another and how a visual design can be optimized. Fortunately, the visualization community has accumulated a substantial amount of knowledge about the role of existing visualization techniques in specific analytic processes, but the generalization of such knowledge for explaining many phenomena in practice or guiding the development of new applications has been challenging, especially in terms of using mathematical theories and quantitative measures. Accordingly, progress in such principle research on data visualization would also contribute quantitative measures of visualization quality. In addition, the community seeks a better understanding of the merits and demerits of conducting different forms of empirical studies involving domain experts and may benefit from the development of theory-guided methodologies for evaluating visualization techniques and systems.

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 theoretical foundations 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 representations, to the perception and cognition by the human visual system and the interpretation by the human mind. While we have good empirical methods for evaluating visualization techniques and systems in applications, more effort will be necessary for using empirical studies to inform theory formation and to validate or falsify proposed theories. Such theories, once adequately validated, would in return provide the basis for more effective and efficient methods of evaluation. Modern visualization includes advanced numerical and algorithmic data processing, so the correctness of such processing requires a critical look at its assumptions, considering the application at hand. Only then, visualization can establish strong correlations between visualization algorithms and questions in the application domains. Further, 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 be related to uncertainty in the data, errors due to numerical processing, errors due to visual depiction, and, finally, uncertainty in human perception and cognition.

This book does not provide the absolute or final foundations of data visualization, and indeed no book could ever do. Nevertheless, to date, it is the most extensive collection of this discourse, by the visualization experts representing different areas of visualization, on four important foundational aspects of visualization:

Theoretical Underpinnings of Data Visualization: As the research field of data visualization evolves, and lots of individual contributions are made to a quickly growing corpus of the visualization literature, theoretic considerations become increasingly important. Explaining visualization and how it works become a pressing question as well as sorting out essential theoretic underpinnings of the visualization process. In Part I of this book, six chapters contribute a rich in-depth discussion of central theoretic questions in visualization research. A fabric of visualization (Chap. 1) is described, demonstrating the complex interaction of different aspects in data visualization, before the central topic of abstraction is addressed (Chap. 2). In Chap. 3, the question of what can we measure in visualization (and how) is discussed, before the focus is set on the role of prior knowledge in visualization (Chap. 4). Part I of this book then closes with two extensive chapters on important mathematical foundations of visualization (Chap. 5) and on essential concepts for mappings and transformations in data visualization (Chap. 6).

Empirical Studies in Visualization: While empirical studies provide useful means for evaluating visualization techniques and systems, it has become more common to use empirical studies to gain new insight about various fundamental questions about human perception and cognition in visualization. Meanwhile, there has been concern about whether empirical studies can serve as an effective and efficient means of evaluation in applications involving domain experts and aspiration for finding more cost-effective evaluation methods. The six chapters in this part address the topics on empirical studies from several perspectives that are rare in the visualization literature, including a survey of variables used in controlled and semi-controlled experiments (Chap. 7), a rational discourse on evaluation involving domain experts (Chap. 8), an in-depth discourse on evaluation in the form of long-term case studies (Chap. 9), an inspiring argument for using visualization as an analytical tool for handling data resulting from empirical studies (Chap. 10), a philosophical examination of different schools of thought that represent some most consequential hypotheses in visualization (Chap. 11), and a summary of the challenges and opportunities in empirical visualization research (Chap. 12).

Collaboration with Domain Experts: Many visualizations address questions and needs from researchers, engineers, or analysts. These users know the data, the underlying model, and the tasks well, usually even better than the visualization experts involved. This part discusses successful examples and draws conclusions from them in collaboration with such domain experts. The reader can also find some advice on how to find and start good collaborations. The three chapters in this part address the topic of collaboration with domain experts from practical and theoretical points of view. Seven successful case studies are described including learned lessons (Chap. 13), the view of industry in collaboration with universities is given special attention (Chap. 14), and more theoretical considerations about the collaboration between domain experts and visualization researchers are presented (Chap. 15).

Visualization for Broad Audiences: Besides domain experts, there is strong need for visualizations for broad audiences like the general public. A substantial part of science communication and public debate relies on effective visualizations allowing to understand and to draw conclusions for a lay audience on data. This part concerns the foundations of this specific challenge (Chap. 16), as well as remaining challenges (Chap. 21). It also includes descriptions of the goals, characteristics, and examples of visualization for broad audiences in four distinct settings: a research institute (Chap. 17), a large government agency (Chap. 18), a science center or museum (Chap. 19), and three different perspectives on educational settings (Chap. 20).

This book follows an inspiring, engaging, and energetic Dagstuhl Seminar in January 2018 on the topic. The editors and all the authors are very grateful to Schloss Dagstuhl, its staff, and its funding organizations for the unique opportunity to hold the seminar there. Without this support and great atmosphere, this book would not have been possible. The section of Acknowledgment details our gratitude to all seminar participants, authors, reviewers, the individuals, and organizations that helped to produce this book.

Oxford, UK — Min Chen
Bergen, Norway — Helwig Hauser
Orono, USA — Penny Rheingans
Leipzig, Germany — Gerik Scheuermann
February 2020

Acknowledgements

Fig. 1 Attendees of the Dagstuhl Seminar 18041 on *Foundations of Data Visualization*, which was held on January 21–26, 2018. *Source* Schloss Dagstuhl, Leibniz-Zentrum für Informatik GmbH, Oktavie-Allee, 66687 Wadern, Deutschland; Postprocessing: H. Hauser

The preparation of this book started mainly during the Dagstuhl Seminar 18041 on *Foundations of Data Visualization*, which was held on January 21–26, 2018. We are grateful to all attendees of this event (Fig. 1), showing an unprecedented amount of enthusiasm and contributing large collaborative efforts for collecting, analyzing, and consolidating the foundations of the subject which is so dear to every attendee—*visualization*. We acknowledge especially the contributions made by those attendees

who took part in discussions, shared their views—some even started their writing—but eventually are not listed as the contributors of the chapters in this book. They include: Hamish Carr, Eduard Gröller, Hans-Christian Hege, Nathalie Henry-Riche, Gordon Kindlmann, Laura A. McNamara, and Jarke van Wijk.

We are very, very grateful to all the authors of the twenty chapters in this book and appreciate their collaborative spirit and their enormous endurance in completing their ambitious writing plans. We value tremendously the contributions made by all co-authors recruited by the attendees of the Dagstuhl Seminar to provide their crucial knowledge. We are certain that the observation, analysis, critique, insight, and vision offered by the authors will collectively have a profound impact on the development of visualization as a scientific and technological subject.

We are also in debt to the hardworking reviewers of all chapters in the book. They include: Alfie Abdul-Rahman, Jim Ahrens, Fabian Bolte, Michael Böttinger, Roxana Bujack, Soumya Dutta, Christian Heine, Ingrid Hotz, Tobias Isenberg, Barbora Kozlikova, Helene-Nicole Kostis, Alark Joshi, Heidi Lam, Liz Marai, Krešimir Matković, Laura A. McNamara, Silvia Miksch, Margit Pohl, Ivan Viola, Thomas Wischgoll, and Ross Whitaker. Their comments, critiques, and suggestions have been indispensable in improving the quality of the book, and their time, effort, knowledge, and wisdom are deeply appreciated. In addition, the editors of this book have also acted as anonymous reviewers for some chapters.

The book was prepared primarily on a ShareLaTeX host at Leipzig University, Germany. Christian Blecha has provided vital technical support to the editors and to the authors throughout the writing period of this book. We cannot thank him enough for his great effort and marvelous skill. We also thank Michael Böttinger for his advice and assistance in dealing with some difficult LaTeX issues.

We also record here our enormous gratitude to Helen Desmond (Springer) for her advice and patience throughout this book project. We appreciate the support offered by the Springer teams for managing the Springer LaTeX style and for typesetting, cover design, Web access, and so on.

Last but not least, on behalf of all attendees of the Dagstuhl Seminar 18041 on *Foundations of Data Visualization*, we would like to take our hats off and thank once again the brilliant Dagstuhl staff, who (as always!) made the activities surrounding the highly challenging research topic very enjoyable and productive. Especially, we thank Director Prof. Dr. Raimund Seidel, Susanne Bach-Bernhard, Annette Beyer, Jutka Gasiorowski, Michael Gerke, Thomas Schillo, and Dr. Michael Wagner.

Oxford, UK — Min Chen
Bergen, Norway — Helwig Hauser
Orono, USA — Penny Rheingans
Leipzig, Germany — Gerik Scheuermann
February 2020

Contents

Contributors

Alfie Abdul-Rahman — King's College London, UK

Assaf Anyamba — Universities Space Research Association and NASA Goddard Space Flight Center, Greenbelt, MD, USA

Jordan Bell — University of Alabama in Huntsville, AL, USA

Johanna Beyer — Harvard University, Cambridge, MA, USA

Christian Blecha — Leipzig University, Germany

Alexander Bock — Linköping University, Norrköping, Sweden

Ryan Boller — NASA Goddard Space Flight Center, Greenbelt, MD, USA

Fabian Bolte — University of Bergen, Norway

Michael Böttinger — DKRZ, Hamburg, Germany

Stefan Bruckner — University of Bergen, Norway

Roxana Bujack — Los Alamos National Laboratory, Santa Fe, NM, USA

Min Chen — University of Oxford, UK

Darren J. Edwards — Swansea University, UK

Christoph Garth — Technische Universität Kaiserslautern, Germany

Hans Hagen — Technische Universität Kaiserslautern, Germany

Charles Hansen — University of Utah, Salt Lake City, UT, USA

Helwig Hauser — University of Bergen, Norway

Mario Hlawitschka — Leipzig University of Applied Sciences, Germany

Ingrid Hotz — Linköping University, Norrköping, Sweden

Tobias Isenberg — Inria Saclay, Palaiseau, France

Alark Joshi — University of San Francisco, CA, USA

Virginia Kalb — NASA Goddard Space Flight Center, Greenbelt, MD, USA

Helen-Nicole Kostis — Universities Space Research Association and NASA Goddard Space Flight Center, Greenbelt, MD, USA

Barbora Kozlikova — Masaryk University, Brno, Czech Republic

David H. Laidlaw — Brown University, Providence, RI, USA

Heike Leitte — TU Kaiserslautern, Germany

Patric Ljung — Linköping University, Norrköping, Sweden

G. Elisabeta Marai — University of Illinois at Chicago, IL, USA

Krešimir Matković — VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH, Vienna, Austria

Silvia Miksch — TU Wien, Vienna, Austria

Torsten Möller — University of Vienna, Austria

Andrew Molthan — NASA Marshall Space Flight Center, Huntsville, AL, USA

Daniela Oelke — SIEMENS AG, Munich, Germany

Paulo A. Oemig — New Mexico State University, Las Cruces, NM, USA

Bernhard Preim — Otto-von-Guericke-Universität Magdeburg, Germany

Penny Rheingans — University of Maine, Orono, ME, USA

Geraldine B. Robbins — ASRC/AFSS NASA Goddard Space Flight Center, Greenbelt, MD, USA

Miguel O. Román — Universities Space Research Association (USRA), Columbia, MD, USA

Gerik Scheuermann — Leipzig University, Germany

Lori Schultz — University of Alabama in Huntsville, AL, USA

Ranjay M. Shrestha — Science Systems and Applications, Inc., Greenbelt, MD, USA

Eleanor C. Stokes — University of Maryland, College Park, MD, USA

Markus Stommel — TU Dortmund, Germany

Marc Streit — Johannes Kepler University Linz, Austria

Qingsong Sun — Science Systems and Applications, Inc., Greenbelt, MD, USA

Ariane Sutor — SIEMENS AG, Munich, Germany

Amitabh Varshney — University of Maryland, College Park, MD, USA

Maria Velez-Rojas — CA Technologies, Santa Clara, CA, USA

Ivan Viola — King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Yong Wan — University of Utah, Salt Lake City, UT, USA

Bei Wang — University of Utah, Salt Lake City, UT, USA

Zhuosen Wang — University of Maryland, College Park, MD, USA

Johannes Waschke — University of Applied Sciences Leipzig, Germany

Daniel Weiskopf — University of Stuttgart, Germany

Ross Whitaker — University of Utah, Salt Lake City, UT, USA

Thomas Wischgoll — Wright State University, Dayton, OH, USA

Anders Ynnerman — Linköping University, Norrköping, Sweden

Caroline Ziemkiewicz — Forrester Research, Inc, Cambridge, MA, USA