

Mastering the Information Age Solving Problems with Visual Analytics

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

This work is subject to copyright.

All rights reserved, whether the whole or part of the material is concerned, specifically those of translation, reprinting, re-use of illustrations, broadcasting, reproduction by photocopying machines or similar means, and storage in data banks.

Copyright © 2010 by the authors

Published by the Eurographics Association
–Postfach 8043, 38621 Goslar, Germany–

Printed in Germany, Druckhaus “Thomas Müntzer” GmbH, Bad Langensalza

Cover image: © iStockphoto.com/FotoMak

ISBN 978-3-905673-77-7

The electronic version of this book is available from the Eurographics Digital Library at <http://diglib.eg.org>

In Memoriam of Jim Thomas¹,
a visionary and inspiring person, innovative researcher, enthusiastic leader
and excellent promoter of visual analytics.

¹Jim Thomas passed away on August 6, 2010.

Preface

Today, in many spheres of human activity, massive sets of data are collected and stored. As the volumes of data available to lawmakers, civil servants, business people and scientists increase, their effective use becomes more challenging. Keeping up to date with the flood of data, using standard tools for data management and analysis, is fraught with difficulty. The field of visual analytics seeks to provide people with better and more effective ways to understand and analyse these large datasets, while also enabling them to act upon their findings immediately, in real-time. Visual analytics integrates the analytic capabilities of the computer and the abilities of the human analyst, thus allowing novel discoveries and empowering individuals to take control of the analytical process. Visual analytics sheds light on unexpected and hidden insights, which may lead to beneficial and profitable innovation.

This book is one of the outcomes of a two-year project called VisMaster CA, a coordination action funded by the European Commission from August 2008 to September 2010. The goal of VisMaster was to join European academic and industrial R&D excellence from several individual disciplines, forming a strong visual analytics research community. An array of thematic working groups was set up by the consortium, which focused on advancing the state of the art in visual analytics. These working groups joined research excellence in the fields of data management, data analysis, spatial-temporal data, and human visual perception research with the wider visualisation research community.

This Coordination Action successfully formed and shaped a strong European visual analytics community, defined the research roadmap, exposed public and private stake-holders to visual analytics technology and set the stage for larger follow-up visual analytics research initiatives. While there is still much work ahead to realise the visions described in this book, Europe's most prestigious visual analytics researchers have combined their expertise to determine the next steps.

This research roadmap is the final delivery of VisMaster. It presents a detailed review of all aspects of visual analytics, indicating open areas and strategies for the research in the coming years. The primary sources for this book are the final reports of the working groups, the cross-community reports as well as the resources built up on the Web platform².

The VisMaster consortium is confident that the research agenda presented in this book, and especially the recommendations in the final chapter, will help to support a sustainable visual analytics community well beyond the duration of VisMaster CA, and also serves as the reference for researchers in related

²<http://www.vismaster.eu>

scientific disciplines, which are interested to join and strengthen the community. This research roadmap does not only cover issues that correspond to scientific challenges: it also outlines the connections to sciences, technologies, and industries for which visual analytics can become an 'enabling technology'. Hence, it serves as a reference for research program committees and researchers of related fields in the ICT theme and beyond, to assess the possible implications for their respective field.

Structure

Chapter 1 motivates the topic of visual analytics and presents a brief history of the domain. Chapter 2 deals with the basis of visual analytics including its current application areas, the visual analytics process, its building blocks, and its inherent scientific challenges.

The following Chapters 3 to 8 were written by respective working groups in VisMaster, assisted by additional partners of the consortium and community partners. Each of these chapters introduces the specific community that is linked to visual analytics (e.g., data mining). It then outlines the state of the art and the specific challenges and opportunities that lie ahead for this field with respect to visual analytics research. In particular, Chapter 3 deals with data management for visual analytics, Chapter 4 covers aspects of data mining, Chapter 5 outlines the application of visual analytics to problems with spatial and temporal components, Chapter 6 considers infra-structural issues, Chapter 7 looks at human aspects and Chapter 8 discusses evaluation methodologies for visual analytics.

The final chapter presents a summary of challenges for the visual analytics community and sets out specific recommendations to advance visual analytics research. These recommendations are a collaborative effort of all working groups and specifically address different target groups: the European Commission, the visual analytics research community, the broader research community, industry and governments, together with other potential users of visual analytics technology.

Acknowledgements

We would like to thank all the partners of VisMaster (including community partners) who have contributed to creating this book. Whilst some have produced chapters (authors of each chapter are shown overleaf), others have been involved with the reviewing process and/or coordinating their work groups.

Special thanks goes to Bob Spence and Devina Ramduny-Ellis for their most helpful comments and contributions.

We are appreciative of the excellent technical and creative support given by Florian Stoffel, Juri Buchmüller and Michael Regenscheit. We are truly grateful

once more for the excellent support of Eurographics, and in particular Stefanie Behnke, for publishing this work.

Last but not least, we are indebted to the European Commission, and especially, the project officer of VisMaster CA, Dr. Teresa de Martino, for supporting us throughout; her efforts have contributed appreciably to the success of this project.

This project was funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 225924.

We hope that the content of this book will inspire you to apply current visual analytics technology to solve your real-world data problems, and to engage in the community effort to define and develop visual analytics technologies to meet future challenges.

Daniel Keim (Scientific Coordinator of VisMaster), **Jörn Kohlhammer** (Coordinator of VisMaster), **Geoffrey Ellis**, and **Florian Mansmann**

September 2010

Contents

1	Introduction	1
1.1	Motivation	1
1.2	An Historical Perspective on Visual Analytics	3
1.3	Overview	4
2	Visual Analytics	7
2.1	Application of Visual Analytics	7
2.2	The Visual Analytics Process	10
2.3	Building Blocks of Visual Analytics Research	11
3	Data Management	19
3.1	Motivation	19
3.2	State of the Art	22
3.3	Challenges and Opportunities	32
3.4	Next Steps	36
4	Data Mining	39
4.1	Motivation	39
4.2	State of the Art	44
4.3	Challenges	49
4.4	Opportunities	54
4.5	Next Steps	55
5	Space and Time	57
5.1	Motivation	57
5.2	A Scenario for Spatio-Temporal Visual Analytics	59
5.3	Specifics of Time and Space	62
5.4	State of the Art	68
5.5	Challenges and Opportunities	81
5.6	Next Steps	86
6	Infrastructure	87
6.1	Motivation	87
6.2	State of the Art	91
6.3	Challenges	102
6.4	Opportunities	107
6.5	Next Steps	108
7	Perception and Cognitive Aspects	109
7.1	Motivation	109
7.2	State of the Art	110
7.3	Challenges and Opportunities	123

7.4	Next Steps	130
8	Evaluation	131
8.1	Motivation	131
8.2	State of the Art	134
8.3	Next Steps	141
9	Recommendations	145
9.1	The Challenges	145
9.2	Meeting the Challenges	148
9.3	Future Directions	153
	Bibliography	155
	List of Figures	164
	Glossary of Terms	167

List of Authors

Chapters 1 & 2	Daniel A. Keim Jörn Kohlhammer Florian Mansmann Thorsten May Franz Wanner	University of Konstanz Fraunhofer IGD University of Konstanz Fraunhofer IGD University of Konstanz
Chapter 3	Giuseppe Santucci Helwig Hauser	Sapienza Università di Roma University of Bergen
Chapter 4	Kai Puolamäki Alessio Bertone Roberto Therón Otto Huisman Jimmy Johansson Silvia Miksch Panagiotis Papapetrou Salvo Rinzivillo	Aalto University Danube University Krems Universidad de Salamanca University of Twente Linköping University Danube University Krems Aalto University Consiglio Nazionale delle Ricerche
Chapter 5	Gennady Andrienko Natalia Andrienko Heidrun Schumann Christian Tominski Urska Demsar Doris Dransch Jason Dykes Sara Fabrikant Mikael Jern Menno-Jan Kraak	Fraunhofer IAIS Fraunhofer IAIS University of Rostock University of Rostock National University of Ireland German Research Centre for Geosciences City University London University of Zurich Linköping University University of Twente
Chapter 6	Jean-Daniel Fekete	INRIA
Chapter 7	Alan Dix Margit Pohl Geoffrey Ellis	Lancaster University Vienna University of Technology Lancaster University
Chapter 8	Jarke van Wijk Tobias Isenberg Jos B.T.M. Roerdink Alexandru C. Telea Michel Westenberg	Eindhoven University of Technology University of Groningen University of Groningen University of Groningen Eindhoven University of Technology
Chapter 9	Geoffrey Ellis Daniel A. Keim Jörn Kohlhammer	University of Konstanz University of Konstanz Fraunhofer IGD

1 Introduction

1.1 Motivation

We are living in a world which faces a rapidly increasing amount of data to be dealt with on a daily basis. In the last decade, the steady improvement of data storage devices and means to create and collect data along the way, influenced the manner in which we deal with information. Most of the time, data is stored without filtering and refinement for later use. Virtually every branch of industry or business, and any political or personal activity, nowadays generates vast amounts of data. Making matters worse, the possibilities to collect and store data increase at a faster rate than our ability to use it for making decisions. However, in most applications, raw data has no value in itself; instead, we want to extract the information contained in it.

Raw data has no value in itself, only the extracted information has value

The information overload problem refers to the danger of getting lost in data, which may be:

- irrelevant to the current task at hand,
- processed in an inappropriate way, or
- presented in an inappropriate way.

Due to information overload, time and money are wasted, scientific and industrial opportunities are lost because we still lack the ability to deal with the enormous data volumes properly. People in both their business and private lives, decision-makers, analysts, engineers, emergency response teams alike, are often confronted with large amounts of disparate, conflicting and dynamic information, which are available from multiple heterogeneous sources. There is a need for effective methods to exploit and use the hidden opportunities and knowledge resting in unexplored data resources.

Time and money are wasted and opportunities are lost

In many application areas, success depends on the right information being available at the right time. Nowadays, the acquisition of raw data is no longer the main problem. Instead, it is the ability to identify methods and models, which can turn the data into reliable and comprehensible knowledge. Any technology, that claims to overcome the information overload problem, should answer the following questions:

Success depends on availability of the right information

- Who or what defines the 'relevance of information' for a given task?
- How can inappropriate procedures in a complex decision making process be identified?
- How can the resulting information be presented in a decision-oriented or task-oriented way?

With every new application, processes are put to the test, possibly under circumstances totally different from the ones they have been designed for. The awareness of the problem of how to understand and analyse our data has greatly increased in the last decade. Even though we implement more powerful tools for automated data analysis, we still face the problem of understanding and 'analysing our analyses' in the future – fully automated search, filter and analysis only work reliably for well-defined and well-understood problems. The path from data to decision is typically fairly complex. Fully automated data processing methods may represent the knowledge of their creators, but they lack the ability to communicate their knowledge. This ability is crucial. If decisions that emerge from the results of these methods turn out to be wrong, it is especially important to be able to examine the processes that are responsible.

Visual analytics aims at making data and information processing transparent

The overarching driving vision of visual analytics is to turn the information overload into an opportunity: just as information visualisation has changed our view on databases, the goal of visual analytics is to make our way of processing data and information transparent for an analytic discourse. The visualisation of these processes will provide the means of examining the actual processes instead of just the results. Visual analytics will foster the constructive evaluation, correction and rapid improvement of our processes and models and ultimately the improvement of our knowledge and our decisions.

Visual analytics combines the strengths of humans and computers

On a grand scale, visual analytics provides technology that combines the strengths of human and electronic data processing. Visualisation becomes the medium of a semi-automated analytical process, where humans and machines cooperate using their respective, distinct capabilities for the most effective results. The user has to be the ultimate authority in directing the analysis. In addition, the system has to provide effective means of interaction to focus on their specific task. In many applications, several people may work along the processing path from data to decision. A visual representation will sketch this path and provide a reference for their collaboration across different tasks and at different levels of detail.

The diversity of these tasks cannot be tackled with a single theory. Visual analytics research is highly interdisciplinary and combines various related research areas such as visualisation, data mining, data management, data fusion, statistics and cognition science (among others). One key idea of visual analytics is that integration of all these diverse areas is a scientific discipline in its own right. Application domain experts are becoming increasingly aware that visualisation is useful and valuable, but often ad hoc solutions are used, which rarely match the state of the art in interactive visualisation science, much less the full complexity of the problems, for which visual analytics aims to seek answers. Even if the awareness exists, that scientific analysis and results must be visualised in one way or the other. In fact, all related research areas in the context of visual analytics research conduct rigorous science, each in their vibrant research communities. One main goal of this book is to demonstrate that collaboration can lead to novel, highly effective analysis tools, contributing solutions to the information overload problem in many important domains.

Because visual analytics is an integrating discipline, application specific research areas can contribute existing procedures and models. Emerging from highly application-oriented research, research communities often work on specific solutions using the tools and standards of their specific fields. The requirements of visual analytics introduce new dependencies between these fields.

The integration of the previously mentioned disciplines into visual analytics will result in a set of well-established and agreed upon concepts and theories, allowing any scientific breakthrough in a single discipline to have a potential impact on the whole visual analytics field. In return, combining and upgrading these multiple technologies onto a new general level will have a great impact on a large number of application domains.

1.2 An Historical Perspective on Visual Analytics

Automatic analysis techniques such as statistics and data mining developed independently from visualisation and interaction techniques. However, some key thoughts changed the rather limited scope of the fields into what is today called visual analytics research. One of the most important steps in this direction was the need to move from confirmatory data analysis (using charts and other visual representations to just present results) to exploratory data analysis (interacting with the data/results), which was first stated in the statistics research community by John W. Tukey in his 1977 book, *Exploratory Data Analysis*^[116].

Early visual analytics:
exploratory data analysis

With improvements in graphical user interfaces and interaction devices, a research community devoted their efforts to information visualisation^[25, 27, 104, 122]. At some stage, this community recognised the potential of integrating the user in the knowledge discovery and data mining process through effective and efficient visualisation techniques, interaction capabilities and knowledge transfer. This led to *visual data exploration* and *visual data mining*^[64]. This integration considerably widened the scope of both the information visualisation and the data mining fields, resulting in new techniques and many interesting and important research opportunities.

Visual data exploration
and visual data mining

Two of the early uses of the term *visual analytics* were in 2004^[125] and a year later in the research and development agenda, *Illuminating the Path*^[111]. More recently, the term is used in a wider context, describing a new multidisciplinary field that combines various research areas including visualisation, human-computer interaction, data analysis, data management, geo-spatial and temporal data processing, spatial decision support and statistics^[67, 51].

Since 2004: visual
analytics

Despite the relatively recent use of the term visual analytics, characteristics of visual analytics applications were already apparent in earlier systems, such as the CoCo system created in the early 1990s to achieve improvement in the design of a silicon chip^[32]. In this system, numerical optimisation algorithms alone were acknowledged to have serious disadvantages, and it was found that some of these could be ameliorated if an experienced chip designer continually

Some earlier systems
exhibited the
characteristics of visual
analytics

monitored and guided the algorithm when appropriate. The Cockpit interface supported this activity by showing, dynamically, hierarchically related and meaningful indications of chip performance and sensitivity information, as well as on-the-fly advice by an artificial intelligence system, all of which information could be managed to interactively.

1.3 Overview

This book is the result of a community effort of the partners of the VisMaster Coordinated Action funded by the European Union. The overarching aim of this project was to create a research roadmap that outlines the current state of visual analytics across many disciplines, and to describe the next steps to take in order to form a strong visual analytics community, enabling the development of advanced visual analytic applications. The first two chapters introduce the problem space and define visual analytics. Chapters 3 to 8 present the work of the specialised working groups within the VisMaster consortium. Each of these chapters follow a similar structure – the motivation section gives an outline of the problem and relevant background information; the next section presents an overview of the state of the art in the particular domain, with reference to visual analytics; challenges and opportunities are then identified; and finally in the next steps section, suggestions, pertinent to the subject of the chapter, are put forward for discussion. Higher level recommendations for the direction for future research in visual analytics, put forward by the chapter authors are collectively summarised in the final chapter. We now outline the chapters in more detail.

Daniel A. Keim
Jörn Kohlhammer
Florian Mansmann
Thorsten May
Franz Wanner

Chapter 2 describes some application areas for visual analytics and puts the size of the problem into context, and elaborates on the definition of visual analytics. The interdisciplinary nature of this area is demonstrated by considering the scientific fields that are an integral part of visual analytics.

Giuseppe Santucci
Helwig Hauser

Chapter 3 reviews the field of data management with respect to visual analytics and reviews current database technology. It then summarises the problems that can arise when dealing with large, complex and heterogeneous datasets or data streams. A scenario is given, which illustrates tight integration of data management and visual analytics. The state of the art section also considers techniques for the integration of data and issues relating to data reduction, including visual data reduction techniques and the related topic of visual quality metrics. The challenges section identifies important issues, such as dealing with uncertainties in the data and the integrity of the results, the management of semantics (i.e., data which adds meaning to the data values), the emerging area of data streaming, interactive visualisation of large databases and database issues concerning distributed and collaborative visual analytics.

Chapter 4 considers data mining, which is seen as fundamental to the automated analysis components of visual analytics. Since today's datasets are often extremely large and complex, the combination of human and automatic analysis is key to solving many information gathering tasks. Some case studies are presented which illustrate the use of knowledge discovery and data mining (KDD) in bioinformatics and climate change. The authors then pose the question of whether industry is ready for visual analytics, citing examples of the pharmaceutical, software and marketing industries. The state of the art section gives a comprehensive review of data mining/analysis tools such as statistical and mathematical tools, visual data mining tools, Web tools and packages. Some current data mining/visual analytics approaches are then described with examples from the bioinformatics and graph visualisation fields. Technical challenges specific to data mining are described such as achieving data cleaning, integration, data fusion etc. in real-time and providing the necessary infrastructure to support data mining. The challenge of integrating the human into the data process to go towards a visual analytics approach is discussed together with issues regarding its evaluation. Several opportunities are then identified, such as the need for generic tools and methods, visualisation of models and collaboration between the KDD and visualisation communities.

Kai Puolamäki
Alessio Bertone
Roberto Therón
Otto Huisman
Jimmy Johansson
Silvia Miksch
Panagiotis Papapetrou
Salvo Rinzivillo

Chapter 5 describes the requirements of visual analytics for spatio-temporal applications. Space (as in for example maps) and time (values change over time) are essential components of many data analysis problems; hence there is a strong need for visual analytics tools specifically designed to deal with the particular characteristics of these dimensions. Using a sizeable fictitious scenario, the authors guide the reader towards the specifics of time and space, illustrating the involvement of various people and agencies, and the many dependencies and problems associated with scale and uncertainties in the data. The current state of the art is described with a review of maps, geographic information systems, the representation of time, interactive and collaborative issues, and the implication of dealing with massive datasets. Challenges are then identified, such as dealing with diverse data at multiple scales, and supporting a varied set of users, including non-experts.

Gennady Andrienko
Natalia Andrienko
Heidrun Schumann
Christian Tominski
Urska Demsar
Doris Dransch
Jason Dykes
Sara Fabrikan
Mikael Jern
Menno-Jan Kraak

Chapter 6 highlights the fact that currently most visual analytics application are custom-built stand-alone applications, using for instance, in-memory data storage rather than database management systems. In addition, many other common components of visual analytics applications can be identified and potentially built into a unifying framework to support a range of applications. The author of this chapter reviews architectural models of visualisation, data management, analysis, dissemination and communication components and outlines the inherent challenges. Opportunities and next steps for current research are subsequently identified which encourage a collaborative multidisciplinary effort to provide a much needed flexible infrastructure.

Jean-Daniel Fekete

Chapter 7 discusses visual perception and cognitive issues - human aspects of visual analytics. Following a review of the psychology of perception and cognition, distributed cognition, problem solving, particular interaction issues, the authors suggest that we can learn much from early application

Alan Dix
Margit Pohl
Geoffrey Ellis

examples. Challenges identified, include the provision of appropriate design methodologies and design guidelines, suitable for the expert analyst as well as the naive users; understanding the analysis process, giving the user confidence in the results, dealing with a wide range of devices and how to evaluate new designs.

Jarke van Wijk
Tobias Isenberg
Jos B.T.M. Roerdink
Alexandru C. Telea
Michel Westenberg

Chapter 8 explains the basic concept of evaluation for visual analytics, highlighting the complexities for evaluating systems that involve the close coupling of the user and semi-automatic analytical processes through a highly interactive interface. The exploratory tasks associated with visual analytics are often open ended and hence it is difficult to assess the effectiveness and efficiency of a particular method, let alone make comparisons between methods. The state of the art section outlines empirical evaluation methodologies, shows some examples of evaluation and describes the development of contests in different sub-communities to evaluate visual analytics approaches on common datasets. The authors then argue that a solid evaluation infrastructure for visual analytics is required and put forward some recommendations on how to achieved this.

Geoffrey Ellis
Daniel A. Keim
Jörn Kohlhammer

Chapter 9 summarises the challenges of visual analytics applications as identified by the chapter authors and presents concrete recommendations for funding agencies, the visual analytics community, the broader research community and potential users of visual analytics technology in order to ensure the rapid advancement of the science of visual analytics.