Exploratory Analysis of Spatial and Temporal Data

Natalia Andrienko · Gennady Andrienko

Exploratory Analysis of Spatial and Temporal Data

A Systematic Approach

With 245 Figures and 34 Tables

Springer

Authors

Natalia Andrienko
Gennady Andrienko

Fraunhofer Institute AIS
Schloss Birlinghoven
53754 Sankt Augustin, Germany
gennady.andrienko@ais.fraunhofer.de
http://www.ais.fraunhofer.de/and

Library of Congress Control Number: 2005936053

ACM Computing Classification (1998): J.2, H.3

ISBN-10 3-540-25994-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-25994-7 Springer Berlin Heidelberg New York

Typeset by the authors
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: KünkelLopka Werbeagentur, Heidelberg

Printed on acid-free paper 45/3142/YL - 5 4 3 2 1 0

Preface

This book is based upon the extensive practical experience of the authors in designing and developing software tools for visualisation of spatially referenced data and applying them in various problem domains. These tools include methods for cartographic visualisation; non-spatial graphs; devices for querying, search, and classification; and computer-enhanced visual techniques. A common feature of all the tools is their high user interactivity, which is essential for exploratory data analysis. The tools can be used conveniently in various combinations; their cooperative functioning is enabled by manifold coordination mechanisms.

Typically, our ideas for new tools or extensions of existing ones have arisen from contemplating particular datasets from various domains. Understanding the properties of the data and the relationships between the components of the data triggered a vision of the appropriate ways of visualising and exploring the data. This resulted in many original techniques, which were, however, designed and implemented so as to be applicable not only to the particular dataset that had incited their development but also to other datasets with similar characteristics. For this purpose, we strove to think about the given data in terms of the generic characteristics of some broad class that the data belonged to rather than stick to their specifics.

From many practical cases of moving from data to visualisation, we gained a certain understanding of what characteristics of data are relevant for choosing proper visualisation techniques. We learned also that an essential stage on the way from data to the selection or design of proper exploratory tools is to envision the questions an analyst might seek to answer in exploring this kind of data, or, in other words, the data analysis tasks. Knowing the questions (or, rather, types of questions), one may look at familiar techniques from the perspective of whether they could help one to find answers to those questions. It may happen in some cases that there is a subset of existing tools that covers all potential question types. It may also happen that for some tasks there are no appropriate tools. In that case, the nature of the tasks gives a clue as to what kind of tool would be helpful. This is an important initial step in designing a new tool.

Having passed along the way from data through tasks to tools many times, we found it appropriate to share the knowledge that we gained from

this process with other people. We would like to describe what components may exist in spatially referenced data, how these components may relate to each other, and what effect various properties of these components and relationships between them may have on tool selection. We would also like to show how to translate the characteristics of data and structures into potential analysis tasks, and enumerate the widely accepted principles and our own heuristics that usually help us in proceeding from the tasks to the appropriate approaches to accomplishing them, and to the tools that could support this. In other words, we propose a methodological framework for the design, selection, and application of visualisation techniques and tools for exploratory analysis of spatially referenced data. Particular attention is paid to spatio-temporal data, i.e. data having both spatial and temporal components.

We expect this book to be useful to several groups of readers. People practising analysis of spatially referenced data should be interested in becoming familiar with the proposed illustrated catalogue of the state-of-the-art exploratory tools. The framework for selecting appropriate analysis tools might also be useful to them. Students (undergraduate and postgraduate) in various geography-related disciplines could gain valuable information about the possible types of spatial data, their components, and the relationships between them, as well as the impact of the characteristics of the data on the selection of appropriate visualisation methods. Students could also learn about various methods of data exploration using visual, highly interactive tools, and acknowledge the value of a conscious, systematic approach to exploratory data analysis. The book may be interesting to researchers in computer cartography, especially those imbued with the ideas of cartographic visualisation, in particular, the ideas widely disseminated by the special Commission on Visualisation of the International Cartographic Association. Our tools are in full accord with these ideas, and our data- and task-analytic approach to tool design offers a way of putting these ideas into practice. It can also be expected that the book will be interesting to researchers and practitioners dealing with any kind of visualisation, not necessarily the visualisation of spatial data. Many of the ideas and approaches presented are not restricted to only spatially referenced data, but have a more general applicability.

The topic of the book is much more general than the consideration of any particular software: we investigate the relations between the characteristics of data, exploratory tasks (questions), and data exploration techniques. We do this first on a theoretical level and then using practical examples. In the examples, we may use particular implementations of the techniques, either our own implementations or freely available demonstrators. However, the main purpose is not to instruct readers in how to use

this or that particular tool but to allow them to better understand the ideas of exploratory data analysis.

The book is intended for a broad reader community and does not require a solid background in mathematics, statistics, geography, or informatics, but only a general familiarity with these subjects. However, we hope that the book will be interesting and useful also to those who do have a solid background in any or all of these disciplines.

Acknowledgements

This book is a result of a theoretical generalisation of our research over more than 15 years. During this period, many people helped us to establish ourselves and grow as scientists. We would like to express our gratitude to our scientific "parents" Nadezhda Chemeris, Yuri Pechersky, and Sergey Soloview, without whom our research careers would not have started. We are also grateful to our colleagues and partners who significantly influenced and encouraged our work from its early stages, namely Leonid Mikulich, Alexander Komarov, Valeri Gitis, Maria Palenova, and Hans Voss.

Since 1997 we have been working at GMD, the German National Research Centre for Information Technology, which was later transformed into the AIS (Autonomous Intelligent Systems) Fraunhofer Institute. Institute directors Thomas Christaller and Stefan Wrobel and department heads Hans Voss and Michael May always supported and approved our work. All our colleagues were always cooperative and helpful. We are especially grateful to Dietrich Wettschereck, Alexandr Savinov, Peter Gatalsky, Ivan Denisovich, Mark Ostrovsky, Simon Scheider, Vera Hernandez, Andrey Martynkin, and Willi Kloesgen for fruitful discussions and cooperation.

Our research was developed in the framework of numerous international projects. We acknowledge funding from the European Commission and the friendly support of all our partners. We owe much to Robert Peckham, Jackie Carter, Jim Petch, Oleg Chertov, Andreas Schuck, Risto Paivinen, Frits Mohren, Mauro Salvemini, and Matteo Villa. Our work was also greatly inspired by a fruitful (although informal) cooperation with Piotr Jankowski and Alexander Lotov.

Our participation in the ICA commissions on Visualisation and Virtual Environments, Maps and the Internet, and Theoretical Cartography had a strong influence on the formation and refinement of our ideas. Among all the members of these commissions, we are especially grateful to Alan MacEachren, Menno-Jan Kraak, Sara Fabrikant, Jason Dykes, David Fairbain, Terry Slocum, Mark Gahegan, Jürgen Döllner, Monica Wachowicz,

Corne van Elzakker, Michael Peterson, Georg Gartner, Alexander Volodtschenko, and Hans Schlichtmann.

Discussions with Ben Shneiderman, Antony Unwin, Robert Haining, Werner Kuhn, Jonathan Roberts, and Alfred Inselberg were a rich source of inspiration and provided apt occasions to verify our ideas. Special thanks are due to the scientists whose books were formative for our research, namely John Tukey, Jacques Bertin, George Klir, and Rudolf Arnheim.

The authors gratefully acknowledge the encouraging comments of the reviewers, the painstaking work of the copyeditor, and the friendly cooperation of Ralf Gerstner and other people of Springer-Verlag.

We thank our family for the patience during the time that we used for discussing and writing the book in the evenings, weekends, and during vacations.

Almost all of the illustrations in the book were produced using the CommonGIS system and some other research prototypes developed in our institute. Online demonstrators of these systems are available on our Web site http://www.ais.fraunhofer.de/and and on the web site of our institute department http://www.ais.fraunhofer.de/SPADE. People interested in using the software should visit the site of CommonGIS, http://www.CommonGIS.com.

The datasets used in the book were provided by our partners in various projects.

1. **Portuguese census.** The data set was provided by CNIG (Portuguese National Centre for Geographic Information) within the EU-funded project CommonGIS (Esprit project 28983). The data were prepared by Joana Abreu, Fatima Bernardo, and Joana Hipolito.
2. **Forests in Europe.** The dataset was created within the project "Combining Geographically Referenced Earth Observation Data and Forest Statistics for Deriving a Forest Map for Europe" (15237-1999-08 F1ED ISP FI). The data were provided to us by EFI (the European Forest Institute within the project EFIS (European Forest Information System), contract number: 17186-2000-12 F1ED ISP FI.
3. **Earthquakes in Turkey.** The dataset was provided within the project SPIN! (Spatial Mining for Data of Public Interest) (IST Programme, project IST-1999-10536) by Valery Gitis and his colleagues.
4. **Migration of white storks.** The data were provided by the German Research Centre for Ornithology of the Max Planck Society within a German school project called "Naturdetektive". The data were prepared by Peter Gatalsky.

5. **Weather in Germany.** The dataset was published by Deutscher Wetterdienst at the URL http://www.dwd.de/de/FundE/Klima/KLIS/daten/online/nat/index_monatswerte.htm. Simon Scheider prepared the data for application of the tools.
6. **Crime in the USA.** The dataset was published by the US Department of Justice, URL http://bjsdata.ojp.usdoj.gov/dataonline/. The data were prepared by Mohammed Islam.
7. **Forest management scenarios.** The dataset was created in the project SILVICS (Silvicultural Systems for Sustainable Forest Resources Management) (INTAS EU-funded project). The data were prepared for analysis by Alexey Mikhaylov and Peter Gatalsky.
8. **Forest fires in Umbria.** The dataset was provided within the NEFIS (Network for a European Forest Information Service) project, an accompanying measure in the Quality of Life and Management of Living Resources Programme of the European Commission (contract number QLK5-CT-2002-30638). The data were collected by Regione dell'Umbria, Servizio programmazione forestale, Perugia, Italy; the survey was performed by Corpo Forestale dello Stato, Italy
9. **Health care in Idaho.** The dataset was provided by Piotr Jankowski within an informal cooperation project between GMD and the University of Idaho, Moscow, ID.

August 2005

Sankt Augustin, Germany

Natalia Andrienko
Gennady Andrienko

Contents