Comput Stat (2011) 26:561–565
DOI 10.1007/s00180-011-0278-9

EDITORIAL

Data Viz VI

Adalbert F. X. Wilhelm · Lars Linsen

Published online: 11 October 2011

Data Viz VI, an international workshop on Data and Information Visualization[1] was held at Jacobs University Bremen from June 24th to 28th, 2008 and brought together more than fifty researchers from Asia, Europe, and North America. This workshop followed the tradition of similar meetings that took place in the decade before. That we labeled it as the sixth workshop in this series is a rather subjective estimate, based on personally attending a series of precursors, namely at

- University of Augsburg, 1996
- Drew University, 1998,
- George Mason University, 2001
- University of Augsburg, 2002, and
- Humboldt University Berlin, 2006.

[1] The organizers gratefully acknowledge the financial support of this workshop by the Commerzbank Foundation, Frankfurt am Main, and Jacobs University Bremen.

A. F. X. Wilhelm (✉) · L. Linsen
Jacobs University Bremen, Bremen, Germany
e-mail: a.wilhelm@jacobs-university.de

L. Linsen
e-mail: l.linsen@jacobs-university.de

Present Address:
A. F. X. Wilhelm
George Mason University, Fairfax, VA, USA

As you base your count on more researchers in this field, some might leave out the first one in 1996 (Unwin and Theus 2004) or refer to earlier gatherings in Heidelberg.[2] It is worthwhile noting that in the 1995 meeting at Heidelberg two researchers –Ross Ihaka and Robert Gentleman—came all the way from New Zealand, to present "an S-like environment that can work on small Macintoshes" that they called R emphasizing that "the current version is used for teaching introductory courses on Macs with less than 4MB of memory".[3]

While the Heidelberg meeting in 1995 marked the early days of R, the meeting in Bremen showed a certain consolidation in the statistical graphics research by finally implementing in R (http://www.r-project.org) ideas that have been around for quite some while. This trend of consolidation has continued since then, making now many graphical tools available to a broad user community. It might be fair to state that this process fulfills one of the major requests and concerns of the research community on statistical graphics (Unwin and Theus 2004, p. 8). In many of the early meetings, one of the most-widely heard comments always has been: "These are great ideas, but does your software also run on other computers than Macs?" Now, the question more likely is: "Is there an R package available on CRAN?"

The specific theme of the workshop held in Bremen was *Statistical Graphics: Data and information visualization in today's multimedia society*. The graphical representation of data is common place in modern media and the increasing number of online publications allows for ubiquitous use of color and interactive techniques. Not always comes the increase in technical possibilities with a corresponding increase in quality. One of the major goals of this workshop was to take up the challenge to investigate how well prepared both producers and recipients of statistical graphics are to present, decipher, and interpret information from complex statistical graphics. A second major aspect was the question of adaptations of graphical tools to the current demands of non-standard data sets, in particular social and biological networks, movement data or other time and space referenced data.

The participants of the workshop not only traversed national borders—coming from Canada, China, Germany, Ireland, Israel, Italy, Japan, Korea, the Netherlands, Taiwan, and the United States of America—but also transgressed disciplinary boundaries bringing together statisticians, computer scientists, geographers, climatologists, and other data analysts. Having been able to schedule all 36 talks without any parallel session offered all participants full access to the information provided. Moreover, having accommodated all participants on the university campus made it possible to continue discussions and software presentations outside the official schedule. This also fostered the communication and cooperation between young researchers and more experienced folks among participants.

The presentations covered a broad spectrum of topics and the selected papers that appear in this special issue highlight some of it. While exploratory data analysis and

[2] Günter Sawitzki reminded us about at least three workshops of this kind that took place in the late 1980s, early 1990s as satellite meetings for the SoftStat Conferences. The programme of the last of these workshops, which took place in 1995, is still accessible on the world wide web, http://www.statlab.uni-heidelberg.de/projects/workshop/index.html.

[3] see http://www.statlab.uni-heidelberg.de/projects/workshop/session1.html.

visualization has been often seen as a preliminary method of data exploration done prior to numerical analysis, the application papers in this issue focus on the aspect of using data and information visualization to interpret and enhance the results of numerical methods. The order of the papers starts out with the more general aspect of quality in the visualization process and then moves on to specific visualization approaches for particular data settings. The final paper presents a link between numerical analysis techniques and interactive visualization.

The first paper by Ward et al. (2011) addresses the general visualization and data analysis process. The authors propose and evaluate some strategies for quality assessment of the whole data visualization process by splitting the data analytics process into four stages—data collection, transformation, graphical mapping, and display. They discuss theoretical aspects of quality for graphics and report results from a user study on some specific multivariate displays. For the quality aspect their focus is on data quality, abstraction quality, and visual quality. This paper presents a powerful framework for the inclusion of quality aspects into the multi-variate data analysis process and addresses important issues for further research.

Graph theory and data visualization have a dual linkage: on the one hand, network graphs pose a specific data problem for visualization purposes, on the other hand, graphs can be used to represent the hierarchy of displays or the movements from one representation to another display. It is the latter aspect that Hurley and Oldford (2011a,b) focus on. In Hurley and Oldford (2011a) a general navigational framework is developed that is applied to scatterplots and scatterplot matrices. Visual comparisons are heavily dependent on the ordering of the displays to be compared. Hurley and Oldford (2011b) tackle this issue by transforming this seriation problem into the problem of constructing edge-traversals of graphs. They describe the theoretical background of various edge traversals using Eulerian tours and Hamiltonian decompositions and their implementation in an R package. The potential and flexibility of the approach is illustrated by two applications, the assessment of rater agreement as it is commonly a task in psychology and model comparison in regression.

Telea and Voinea (2011) tackle one of the top ten challenges of visual analytics, namely the adaptation of visual analytics to software engineering and maintenance. In their paper, they give an excellent introduction and overview on the interface between visual analytics and software maintenance. They successfully present the emerging field of Software Visual Analytics and describe a concrete application scenario. Their application addresses the issue of build optimization of large-scale computer codes by adapting interactive data mining and information visualization techniques.

One of the major limitations of all visualization techniques is the dimensionality restriction of our three-dimensional world. Long and Linsen (2011) present an implementation of a 3D star coordinate representation for high density clusters in large multi-dimensional data. They separate the two core issues of the problem: first of all, how to reduce the number of observations to a manageable amount, and secondly, how to represent the resulting multi- dimensional structure within the limits of a two-dimensional computer screen. For the first aspect, they suggest a clustering approach based on density using a parameter-free density estimation procedure to create a cluster hierarchy. For the second aspect, two visualization concepts are proposed: one that is based on 2D and 3D star coordinates, the other uses a 2D radial layout of the cluster

hierarchy. For the first approach cluster envelopes are constructed and the layout of the 2D and 3D star coordinates is optimized such that the nesting structure of the clusters gets visible with a minimum overlap of cluster representations. The multidimensional data space is projected into the two- or three-dimensional visual space by linear mappings to preserve the cluster hierarchies. The second approach uses a linked view of 2D radial layouts to represents the cluster tree with a parallel coordinates layout of the raw data.

Tree visualization is a widely addressed topic in the information visualization community. Linsen and Behrendt (2011) contribute to the research in this field by incorporating tree maps and node-link diagrams into a 3-d layout. The visualization tool described enhances interpretation of the tree structure by supporting various navigation techniques such as zooming, and transparent rendering of subtrees. A case study approach is used to document the efficacy and efficiency of the presented visual exploration system.

In their paper on benchmarking experiments, Eugster and Leisch (2011) bridge the aspect of specific data application and the R software environment. The use of interactive graphical tools for visualizing benchmark experiment data allows for a more flexible and rapid exploration of this kind of data. At the same time, this paper provides an interesting example of extending the **iPlots** interactive graphics package in R.

The paper by Conversano (2011) presents a detailed description of bringing together visualization techniques and data mining tools. In this paper interactive tree visualization tools are used to navigate the cluster hierarchy in a multiclass learning situation with the aim of analytically interpreting the results of the cluster analysis. The approach is based on a combination of the Sequential Automatic Search of Subset of Classifiers Algorithm (SASSC) and the interactive visualization of classification trees as implemented in KLIMT. The approach is illustrated using the vowel data set, a publicly available benchmark data set for recognizing vowel sounds from multiple speakers. Unfortunately, the static version printed in this journal can only unfold a small portion of the method's potential.

The last two decades have seen a tremendous amount of innovation in data and information visualization, leading to a variety of different software approaches and application examples. The workshop has shown that—at least within the statistics community—there is a kind of consolidation by adding more and more interactive features to the R graphics environment and by implementing novel ideas within R. Powerful graphical tools that allow sophisticated and artful arrangement of data are hence more easily and more widely available than ever. Given a reasonable amount of education in exploratory data analysis, statistical computing, and statistical visualization these tools can be applied effectively by a broad user community. Further improvement and sophistication will come from different disciplines but equally important is a broad dissemination of the ideas and concepts of visual analysis and data exploration.

References

Conversano C (2011) Interactive visualization in multiclass learning: integrating the SASSC algorithm with KLIMT. Comput Stat 26. doi:10.1007/s00180-011-0255-3

Eugster M, Leisch F (2011) Exploratory analysis of Benchmark experiments: an interactive approach. Comput Stat 26. doi:10.1007/s00180-010-0227-z
Hurley CB, Oldford RW (2011a) Graphs as navigational infrastructure for high dimensional data spaces. Comput Stat 26. doi:10.1007/s00180-011-0228-6
Hurley CB, Oldford RW (2011b) Eulerian tour algorithms for data visualization and the pairviz package. Comput Stat 26. doi:10.1007/s00180-011-0229-5
Linsen L, Behrendt S (2011) Linked treemap: a 3D treemap-nodelink layout for visualizing hierarchical structures. Comput Stat 26. doi:10.1007/s00180-011-0272-2
Long TV, Linsen L (2011) Visualizing high density clusters in multidimensional data using optimized star coordinates. Comput Stat 26. doi:10.1007/s00180-011-0271-3
Telea A, Voinea L (2011) Visual software analytics for the build optimization of large-scale software systems. Comput Stat 26. doi:10.1007/s00180-011-0248-2
Unwin A, Theus M (2004) Papers on data visualisation—what can we see in them? Comput Stat 19: 5–8. doi:10.1007/BF02915273
Ward M, Xie Z, Yang D, Rundensteiner E (2011) Quality-aware visual data analysis. Comput Stat 26. doi:10.1007/s00180-010-0226-0