



Journal of Statistical Software

July 2004, Volume 11, Book Review 7.

<http://www.jstatsoft.org/>

Reviewer: Frederic Udina
Pompeu Fabra University

Visualizing Categorical Data

Michael Friendly
SAS Institute Inc., Carey, NC, 2000.
ISBN 1-58025-660-0. 456 pp. \$69.95 (P).
<http://www.math.yorku.ca/SCS/vcd/>

Visualization of Categorical Data is an ever growing field of research and Michael Friendly is one of its main contributors. Many techniques he and others have developed in recent years are not included in standard statistical packages. This is one of the main values of this book: it may be seen as a complete review of techniques for visualization of categorical data.

Michael Friendly wrote *The SAS System for Statistical Graphics* in 1991. Its final chapter, titled *Displaying Categorical Data*, included this comment (page 499):

While graphical display techniques are common adjuncts to analysis of variance and regression, methods for plotting contingency table data are not as widely used

The author claims that this is the starting point for the present book.

Content

The book includes eight chapters (see below) with plenty of real life data examples. In each case, careful guidance in the technical aspects of the solution is provided. Detailed directions and code listings on how to produce the graphical solution through SAS procedures and macros are also included. The data sets, as well as the SAS macro listings, are included in the appendices and are also available in electronic form through the book's home page

<http://www.math.yorku.ca/SCS/vcd/>

In fact, Friendly's home page is a good place to learn and play with categorical data visualization involved.

As already mentioned, the appendices include the description of the SAS programs and macros (some of them are listed too) and the datasets used (obviously in the DATA step format).

Comment

The book is really what its title announces: it covers in full the currently available techniques

for analysis and visualization of discrete, frequency, ordinal, and all kinds of categorical data. Each technique is briefly introduced in its mathematical and statistical background, but the emphasis is placed on how to use it for specific problems. It also has a good sample of real life data sets (the example index makes it easy to locate the different treatment given to a single data set). All this means that not only SAS users will be interested in the book; any data analyst willing to learn the right methods for categorical data analysis and presentation should also take a look at it.

It is a fact that graphical culture in statistical work still needs to be encouraged: for example, the worst of the data smoothers, the histogram, is still the one used in the vast majority of published scientific papers. This book can constitute a major step forward towards promoting visualization techniques that are still not very popular. But it is also true that some of the graphics in the book look too clumsy and difficult to interpret for the average reader.

The effort of making accessible some techniques that are not available in most statistical packages is one of the stronger points of this work. On his home page the author provides an even more accessible way to create mosaic displays, but it would be a blessing to have a version of this book providing access to these tools in some other popular, not proprietary, statistical environments.

Reviewer:

Frederic Udina
Pompeu Fabra University
Department of Economics and Business
Barcelona, Spain 08005
E-mail: udina@upf.es