

DATA ANALYSIS USING R AND A GRAPHICAL INTERFACE

R. DUTTER

Vienna University of Technology,

Vienna, Austria

e-mail: R.Dutter@tuwien.ac.at

Extended abstract

The computer program system R provides efficient analysis methods for almost any kind of applied problems in statistics and is running on many computer platforms. It is developed on a non-commercial basis and therefore for many practitioners hard to use, nevertheless widely known.

We report on a package called DAS+R under development using a graphical user interface which should ease the application of more or less sophisticated methods. The basis of the graphical user interface comes from the R Commander (see Fox, 2004). It uses Tcl/Tk programming tools (Welch and Jones, 2003). The emphasis is on the analysis of spatially depending uni- or multivariate data, particularly on problems of geochemical data.

Three special properties of DAS+R should be stressed:

- Interactive definition of data subsets (numerically or graphically) together with set operations. Usage of these subsets in almost all graphics and computations.
- Intensive use of possible relations between the geographical information with the values of data in the statistical and graphical analysis.
- The strong requirement of fast reproducibility and repeatability with small variations in the analysis.

For specified subsets many simple graphics can be generated in an easy way by a few mouse clicks (histograms, boxplots, xy-, ternary plots, scatterplot matrices). These nevertheless can become very sophisticated by using the provided advanced options where almost all options of the usual R commands can be specified by clicking graphical icons.

The geographical information is used by generating different kinds of maps. Different symbol sets can be used for representing the values in space. Surface maps may be produced by simple interpolation algorithms or by sophisticated geostatistical methods as kriging.

All these graphical displays may be produced in any specified scale on a user defined worksheet which can be interactively splitted into arbitrary frames which are provided for the different graphics.

Finally many multivariate methods like principal component and factor analysis, cluster and discriminant analysis, are available.

We describe in short the system and illustrate the usability on some geochemical data sets (e.g. Reimann et al., 1998).

Bibliography

- [1] J. Fox (2004). Getting Started with the R Commander: A Basic-statistics Graphical User Interface to R. *'useR 2004' Conference*, May 20-22, 2004, Vienna University of Technology, Austria.
- [2] C. Reimann, M. Äyräs, V. Chekushin, I. Bogatyrev, R. Boyd, P. de Caritat, R. Dutter, T.E. Finne, J.H. Halleraker, Ø. Jæger, G. Kashulina, O. Lehto, H. Niskavaara, V. Pavlov, M.L. Räsänen, T. Strand, and T. Volden. (1998). *Environmental Geochemical Atlas of the Central Barents Region. Geological Survey of Norway (NGU), P.O. Box 3006, Lade, N-7002 Trondheim, Norway.*
- [3] B.B. Welch and K. Jones (2003). *Practical Programming in Tcl and Tk. Prentice Hall PTR, New York.*