

35 Colour & graphic display aids for correspondence analysis

Irwin Scollar, Irmela Herzog & Michael J. Greenacre

A problem which arises when using Correspondence Analysis is how to incorporate *a-priori* knowledge without disturbing the objectivity of the method. For example, it may be interesting to distinguish between male, female and child graves in a cemetery without using this knowledge in the analysis itself. Alternatively, data obtained from several geographically separated groups to be treated together may be of interest, but one doesn't want to include location in the analysis.

Colour, which may be assigned by the user to units and types, provides a simple solution to coding *a-priori* information. The Bonn Archaeological Statistics Package normally uses a short name, a longer label and an internal sequence number for each type and unit. A two digit colour attribute descriptor for types and units which may be arbitrarily set by the user has now been added to the version 4.5. Any of 16 colours may be assigned via decimal codes. These colours can be shown on EGA/VGA or SuperVGA displays at resolutions up to 1024×768 if available.

The user does not see the internal colour coding. He merely chooses the colours interactively for each type or unit of interest. In addition, each colour may be freely associated with a graphic symbol for output on a plotter or plotter simulator. Colour coding can be added interactively to data previously entered.

The new addition to the Bonn Archaeological Statistics Package has:

- 1) Colour coding of Boolean (and/or) type associations within units.
- 2) Colour flagging of units or types with especially high inertia's using a Tukey (1977:44 ff.) median method for outliers. Separate inertial listings for types and units with flagged outliers are offered.

- 3) Phase marking using user chosen types or units as average or starting markers in the colour display, with the rest of the types or units displayed in the colour of the marker.
- 4) Significant types or units in the graphic display signalled with user chosen colours. Types or units may be displayed separately or both may be shown superimposed. If both are chosen, then a symmetrical display (Greenacre 1984:66 ff.; 1989) is used.
- 5) Short names and/or either round dots or symbols are displayable in colour.
- 6) Mouse supported repeated zooming with scaled quantitative axes or a grid permits viewing and plotting detailed local relationships in the diagram.
- 7) Equal or expanded axes may be chosen to fit the screen or plotter page. All 8 possible rotations and inversions of the image are offered.
- 8) Output of standard HPGL plotter codes with pen colour control is available. Colours and plot symbols may be independently assigned.
- 9) A plot simulator for use with a large number of generally available printers with full screen previewing is available as an option. Conversion of HPGL to Postscript output is included.
- 10) A standardised Common User Architecture (CUA) interface with full mouse support used throughout.

The new tools permit easy detection of multiple "horseshoe" effects without the need to calculate high order components. Simultaneous display of male and female grave sequences is possible without confusion. Larger amounts of data can be analysed more quickly than with any other Correspondence Analysis package on the market known to us.

The colours also discriminate groups which have been separated with Ter Braak (1987)

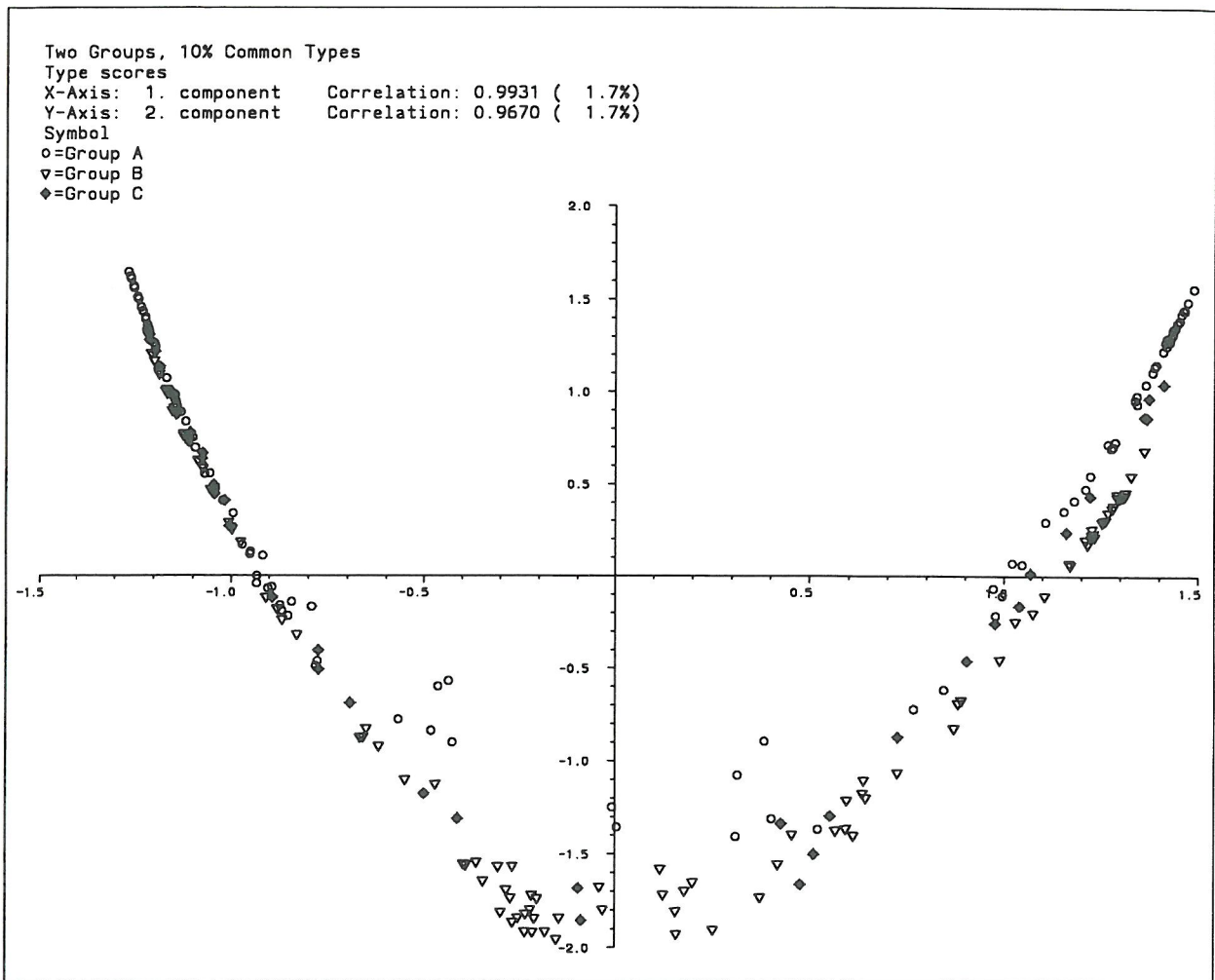


Figure 35.1

detrending, now included in the analysis program. Principal components analysis, with the same colour facilities as in Correspondence Analysis, has been added.

Unfortunately the colours can not be reproduced here, but one may obtain some idea of what is possible with the plotter simulation output of Figures 35.1 and 35.2. A cemetery with two groups of 722 graves containing 10% of 251 types in common was simulated (Scollar & Herzog, 1989). Figure 35.1 shows the resulting classical parabola in the first two components after Correspondence Analysis. This really shows three superimposed parabolas corresponding to the types of the first, second and common group. Figure 35.2 shows the result after using Ter Braak detrending and subsequent stretching of the horizontal axis. The now well-separated parabolas of the three groups are readily visible. Symbols in

both figures have been assigned according to the chosen decimal colour codes as shown in the upper left margin.

The full version 4.5 with handbook and diskettes of the Bonn Archaeological Statistics Package is available from the first author. A public domain version is available on the Göttingen Internet ARCH-L server and can be obtained via Anonymous FTP for ARCH-L subscribers.

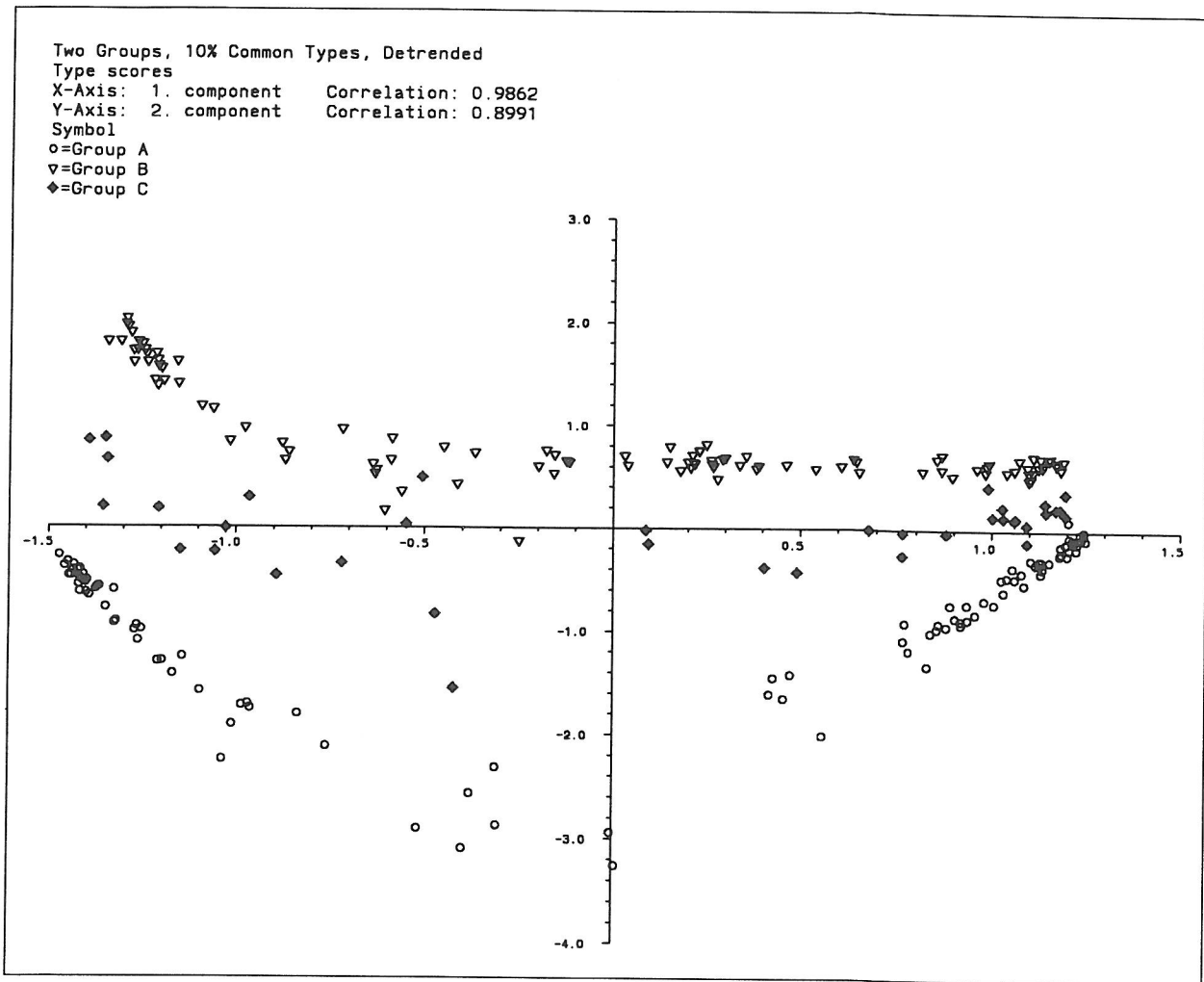


Figure 35.2

References:

- Greenacre, M.J.
 1984 *Theory and Applications of Correspondence Analysis*, Academic Press, London, England.
 1989 The Carroll–Green–Schaffer scaling in correspondence analysis: a theoretical and empirical appraisal. *Journal of Marketing Research* 26:358–365.
- Herzog, I., Scollar, I.
 1988 A mathematical basis for simulation of seriatable data, in *Computer Applications & Quantitative Methods in Archaeology*, 446 2:53–61, BAR, Oxford, England
- Ter Braak, C.J.F.
 1987 *Unimodal Models to Relate Species to Environment*, Agricultural Mathematics Group, Box 100, NL- 6700 AC, Wageningen, The Netherlands
- Tukey, J.W.
 1977 *Exploratory Data Analysis*, Addison–Wesley Publishing, Reading, MA. USA.

Authors' Addresses

Irwin Scollar
 The Unkelbach Valley Software Works
 In der Au 9,
 D-5480 Remagen 4
 E-mail al001@rs1.rrz.uni-koeln.de

Irmela Herzog
 Rheinisches Amt für Bodendenkmalpflege
 D-5300 Bonn 1

Michael J. Greenacre
 Department of Statistics
 The University of South Africa
 PO Box 392
 Pretoria 0001
 South Africa

