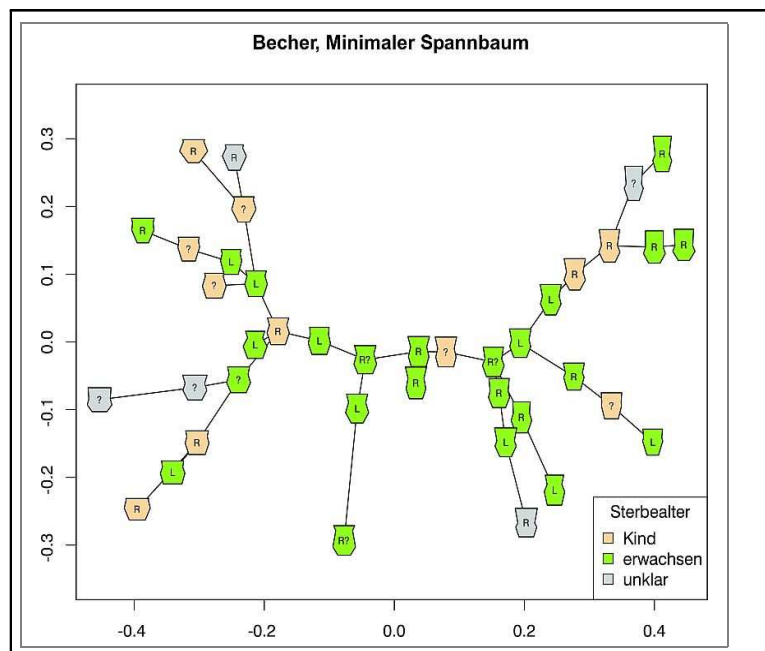


Programm für den R-Kurs beim CAA-AG-Workshop am Sa. 21.01.2012 (Kursleiter Georg Roth/Köln und Jörg Wicke/Halle)

R ist eine Open-Source-Anwendung zur Bewältigung statistischer Fragestellungen (R Development Core Team 2011). Für das Basisprogramm sind tausende online installierbare Erweiterungen (26.10.11: 3415) erhältlich, die modernste statistische Methoden zur Verfügung stellen. Durch seine befehlsorientierte Struktur ist die genaue Dokumentation angewandeter Verfahren möglich. Seine Flexibilität gestattet die Umsetzung eigener Algorithmen. Eine weitere Stärke ist die individuelle Generierung publikationsreifer Grafiken. R ist ein von international führenden Spezialisten betreutes Open-Source-Projekt. Es wird weltweit von zehntausenden Wissenschaftlern verwendet. R ist kostenlos und stets aktuell. Erschwerend für Anfänger ist die Kommandozeilenstruktur und die technische Logik des Programms, welche nur teilweise durch visuell orientierte Editoren (GUI) abgedeckt wird. Der Überwindung dieses Einstiegshindernisses und dem zielorientierten Verfahrenseinsatz ist der Kurs gewidmet.



Schnurkeramische Becher, anhand eines Ähnlichkeitswertes (Prokrustes-Abstand) geordnet und in einem Minimalen Spannbaum dargestellt.

Die für die Abstandsberechnung verwendeten Maße wurden außerdem zur Rekonstruktion der jeweiligen Gefäßumrisse im Diagramm verwendet.

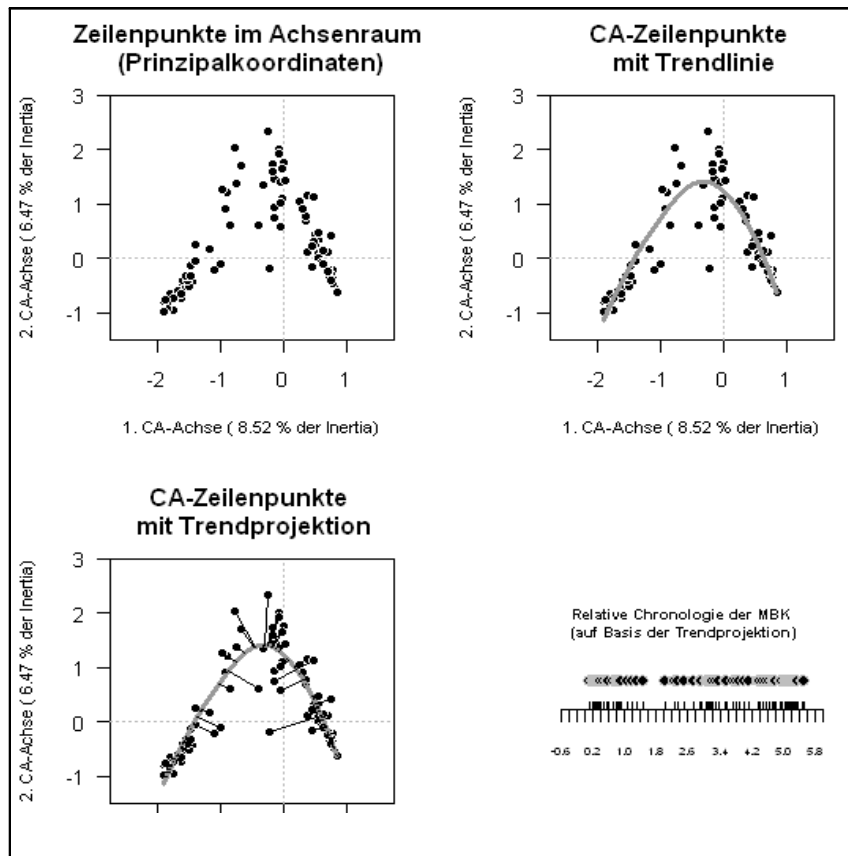
Für Berechnung und Darstellung wurde ausschließlich R verwendet; der Programmieraufwand umfasst ca. 25 Zeilen. (Jörg Wicke/Halle).

Kursteil 1 – Basiskenntnisse (90 min, Kursleiter Jörg Wicke, Tutor Georg Roth):

Im ersten Teil des Kurses führen wir die Teilnehmer an das Programm R heran. Programmierkenntnisse sind von Vorteil aber keine Voraussetzung. Erläutert werden folgende Grundlagen:

- Arbeitsumgebung, Bedienung und Installation von Zusatzpaketen,
- Aufbau der Grundfunktionen (Befehlsstruktur, Operatoren),
- Kontrollstrukturen (Schleifen, Abfrage von Variablenwerten),
- Datenein- und -ausgabe (Dateneinlesen, Zwischenspeichernutzung, Speicherung),
- Datenstruktur in R und Datenaufbereitung (Auswahl, Sortieren),
- einfache deskriptive statistische Funktionen (Mittelwerte, Häufigkeitsverteilung etc.),
- statistische Tests (Chi-Quadrat, Rangsummentests etc.),
- graphische Ausgabe (Bildschirmgrafiken und Grafikenexport),
- R-Hilfe und Internet-Hilfesysteme für R.

In diesem Kursteil lernen die Teilnehmer Grundkenntnisse, um eigene Fragestellungen in R umzusetzen, Programmanpassungen vorzunehmen, Daten einzulesen, zu verarbeiten, auszugeben und graphisch darzustellen. Als Vorbereitung wird empfohlen, sich mit (deutschsprachigen) Einführungen dem Programm anzunähern. Eine Ressourcensammlung steht auf dem R-Blog der Kursleiter, www.rchaeology.tk, bereit.



Korrespondenzanalyse von Fundkomplexen der Michelsberger Kultur (Daten: Höhn 2002) mit dem R-Paket *ca* (Greenacre/Nenadic 2010). In Anlehnung an den Vortrag zur CA-Trend-Projektion (Roth) wird der Verfahrensablauf von einem Prinzipalkoordinaten-Biplot, über die Einsetzung der Trendlinie, die Projektion und das Projektionsergebnis illustriert.

Für Berechnung und Darstellung wurde ausschließlich R verwendet; der Programmieraufwand zur Berechnung umfasst ca. 25 Zeilen, der für die Grafiken ebenfalls (Georg Roth/Köln).

Teil 2 – Überblick zu multivariaten Verfahren (90 min, Kursleiter G. Roth, Tutor J. Wicke): Im zweiten Kursteil wird mit bereitgestellten, archäologischen Datensätzen gearbeitet. Inhalt ist zunächst die Anwendung multivariater Verfahren in R vor allem mit den Paketen *ca* und *vegan* (Greenacre/Nenadic 2010; Oksanen et al. 2011). Grundkenntnisse multivariater Statistik sind von Vorteil, da die Theorie im Kurs nicht erklärt werden kann. Anschließend werden zwei Verfahren für räumliche Daten vorgestellt, Punktfeldstatistik mit *spatstat* (Baddeley/Turner 2005) und 3D-Kerndichteschätzung mit *ks* ():

- Korrespondenzanalyse in *ca* und *vegan* (Ergebnisaufbau und Grafikausgabe),
- Ergebnisevaluierung und -export in *ca* (Qualitätskennwerte),
- kanonische Korrespondenzanalyse in *vegan* (Ergebnisaufbau und Grafikausgabe),
- Punktfeldstatistik in *spatstat* (Datenstruktur und Basisfunktionen),
- 2/3D-Kerndichteschätzung mit *ks* (Bandbreitenschätzung, Visualisierung),

Zur Vorbereitung sei für die PF-Statistik auf ein Skript von Adrian Baddeley (www.csiro.au/files/files/p10ib.pdf) und für die multivariaten Verfahren auf Online-Material von Jari Oksanen (cc.oulu.fi/~jarioksa/opetus/metodi/index.html) verwiesen. Ein fertiges Skript zur Korrespondenzanalyse in *ca* steht unter www.rchaeology.tk bereit.

Der zweite Kursteil bietet einen Zugang, wie wichtige moderne Verfahren in R zielorientiert auf archäologische Daten angewendet werden können. Die bereitgestellten Datensätze sind eine Abundanztafel von MBK-Fundkomplexen (Höhn 2002) und die Verteilung von Ha-D "Fürstensitzen" und Wagengräbern.

- A. Baddeley/R. Turner, *Spatstat: an R package for analyzing spatial point patterns*. *Journal of Statistical Software* 12/6, 2005, 1-42.
- T. Duong, *ks: Kernel smoothing*. R package version 1.8.2. (2011). cran.r-project.org/package=ks
- M. Greenacre/O. Nenadic, *ca: Simple, Multiple and Joint Correspondence Analysis*. R package version 0.33 (2010). cran.r-project.org/package=ca
- B. Höhn, *Michelsberger Kultur in der Wetterau*. UPA 87 (Bonn 2002).
- J. Oksanen et al., *vegan: Community Ecology Package*. R package version 1.17-11 (2011). cran.r-project.org/package=vegan
- R Development Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing (Vienna/Austria 2011). ISBN 3-900051-07-0. www.r-project.org