

Analyse alt- und mittelpaläolithischer Steinartefaktinventare mittels Parallelkoordinatenplots

von

Irmela Herzog

Einleitung

Im Rahmen der Herbsttagung der AG DANK (Datenanalyse und Numerische Klassifikation) stellt häufig ein Teilnehmer im Vorfeld einen Datensatz zur Verfügung, der dann von einem Teil der anderen Tagungsteilnehmern analysiert wird. Im Jahr 2012 fand die Herbsttagung am 5. und 6. Oktober zusammen mit der AG CAA (Computeranwendungen und quantitative Methoden in der Archäologie) in Bonn statt. Der Archäologe Thomas Weber vom Landesamt für Denkmalpflege und Archäologie in Sachsen-Anhalt bot einen Datensatz an, der 81 alt- und mittelpaläolithische Steinartefaktinventare umfasst (siehe auch: Weber 2009). In dem vorliegenden Beitrag werden die Daten zu diesen Steinartefaktinventare mithilfe von Parallelkoordinatenplots ausgewertet.

In dem von Herrn Weber bereit gestellten Datensatz ist jedes Inventar mit einem eindeutigen Namenskürzel versehen. Außerdem liegt für jedes Inventar die Anzahl der untersuchten Artefakte (N), eine Datierung (DAT) in Zahlen (-80 entspricht 80000 v.Chr.) und acht numerische Merkmale vor. Die von Herrn Weber erarbeitete Klassifikation unterscheidet vier Gruppen (group1, group2, group3 und group4). Es liegen noch weitere textlich erfasste Merkmale vor, z.B. ein Kürzel für die Lage der Fundstelle, doch diese Merkmale bleiben in der hier vorgestellten Analyse bzw. Visualisierung der Daten unberücksichtigt.

Nicht für alle Inventare konnten die numerischen Merkmale vollständig erhoben werden, es gibt also auch fehlende Werte. Bei zwei Inventaren war der Anteil der fehlenden Daten sehr hoch, deshalb wurden sie aus dem Datensatz entfernt. In allen anderen Fällen wurden fehlende Daten durch den Mittelwert ersetzt – dies ist sicherlich die einfachste, aber nicht die beste Methode, um fehlende Werte zu schätzen.

Anwendung von Parallelkoordinatenplots

Parallelkoordinatenplots (Brunsdon et al. 1997; Cook & Swayne 2007; Inselberg 1985) eignen sich gut zur Visualisierung von Datensätzen dieser Größenordnung. Doch haben Parallelkoordinatenplots in der Archäologie bisher kaum Verwendung gefunden. Dies mag zum Teil daran liegen, dass die frei verfügbare Software zu diesem Zweck verschiedene Nachteile aufweist. Deshalb wurde von der Autorin dieses Beitrags das Programm MultiPCP entwickelt, das eine interaktive Analyse mithilfe von Parallelkoordinatenplots erlaubt (Herzog & Siegmund im Druck). Bei Benutzerführung und Funktionen des Programms dienten zum Teil geographische Informationssysteme (GIS) als Vorbild.

MultiPCP kann gleichzeitig mehrere Parallelkoordinatenplots erstellen, wobei jeweils eine bestimmte Gruppe von Daten in den Vordergrund gerückt wird (siehe Abb. 1). Dabei erhält der Nutzer die Möglichkeit ein Merkmal auszuwählen, das die Farbgebung steuert (im Beispiel: DAT2). Wie in einem GIS zeigt eine Legende die Farben an, die der Anwender den einzelnen Ausprägungen des Merkmals zugewiesen hat.

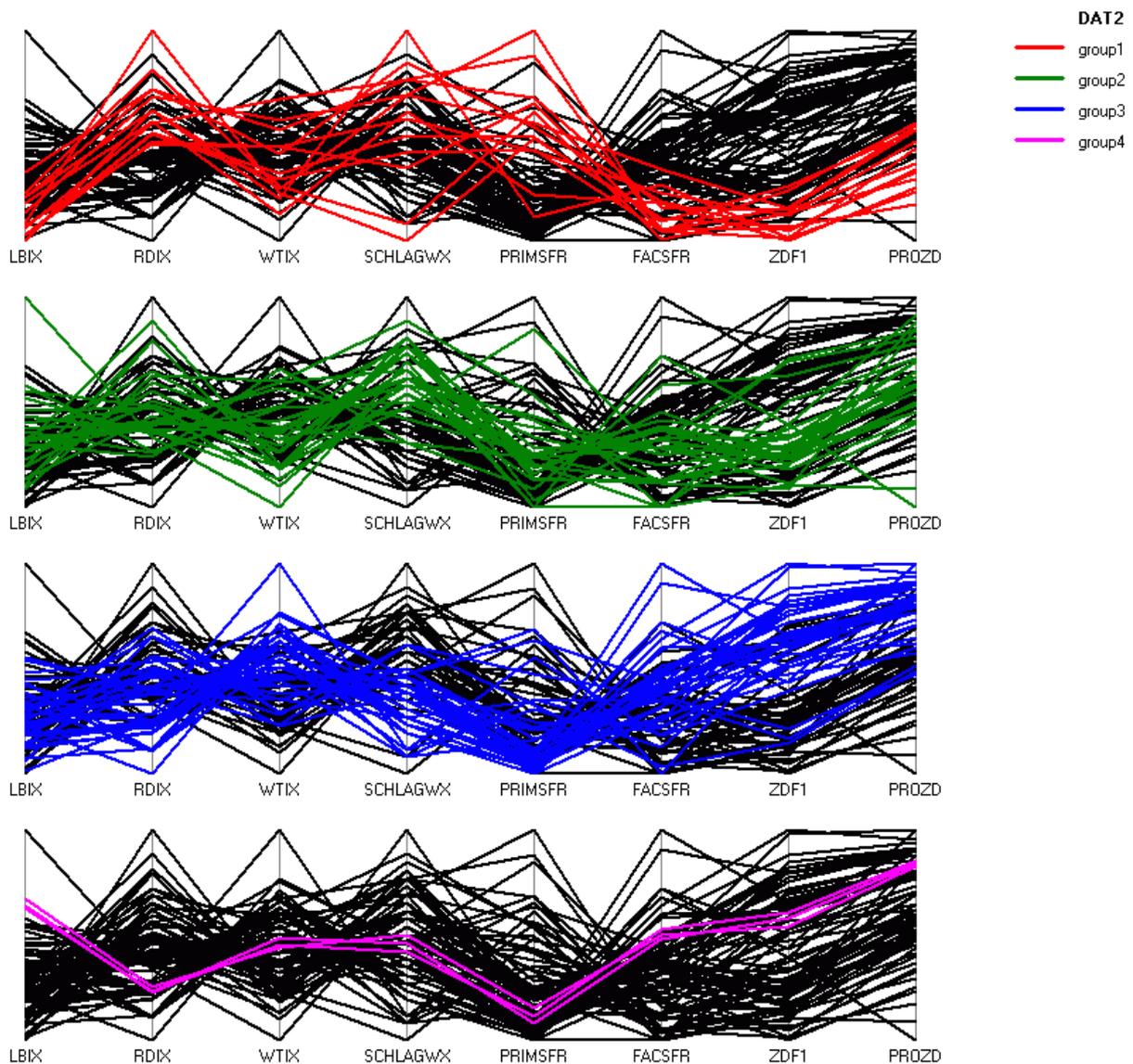


Abb. 1: Vier Parallelkoordinatenplots der acht numerischen Merkmale, bei denen jeweils eine Gruppe in den Vordergrund gerückt und farblich hervorgehoben ist.

In Abb. 1 sieht man, dass „group4“ nur aus wenigen Inventaren besteht und alle numerischen Messwerte dieser Gruppe sehr nah beieinander liegen. Bei „group1“ gibt es Merkmale wie ZDF1, bei denen alle Messwerte im unteren Wertebereich liegen; daneben treten jedoch auch andere Merkmale wie SCHLAGWX auf, die annähernd den gesamten Bereich abdecken.

Abb. 1 zeigt die Reihenfolge der Merkmale, wie in der Datentabelle von Herrn Weber geliefert. In Abb. 2 sind die Merkmale so umgeordnet, dass diejenigen nebeneinander stehen, die für „group1“ nur beschränkte Wertebereiche im Vergleich zu den anderen Inventaren aufweisen.

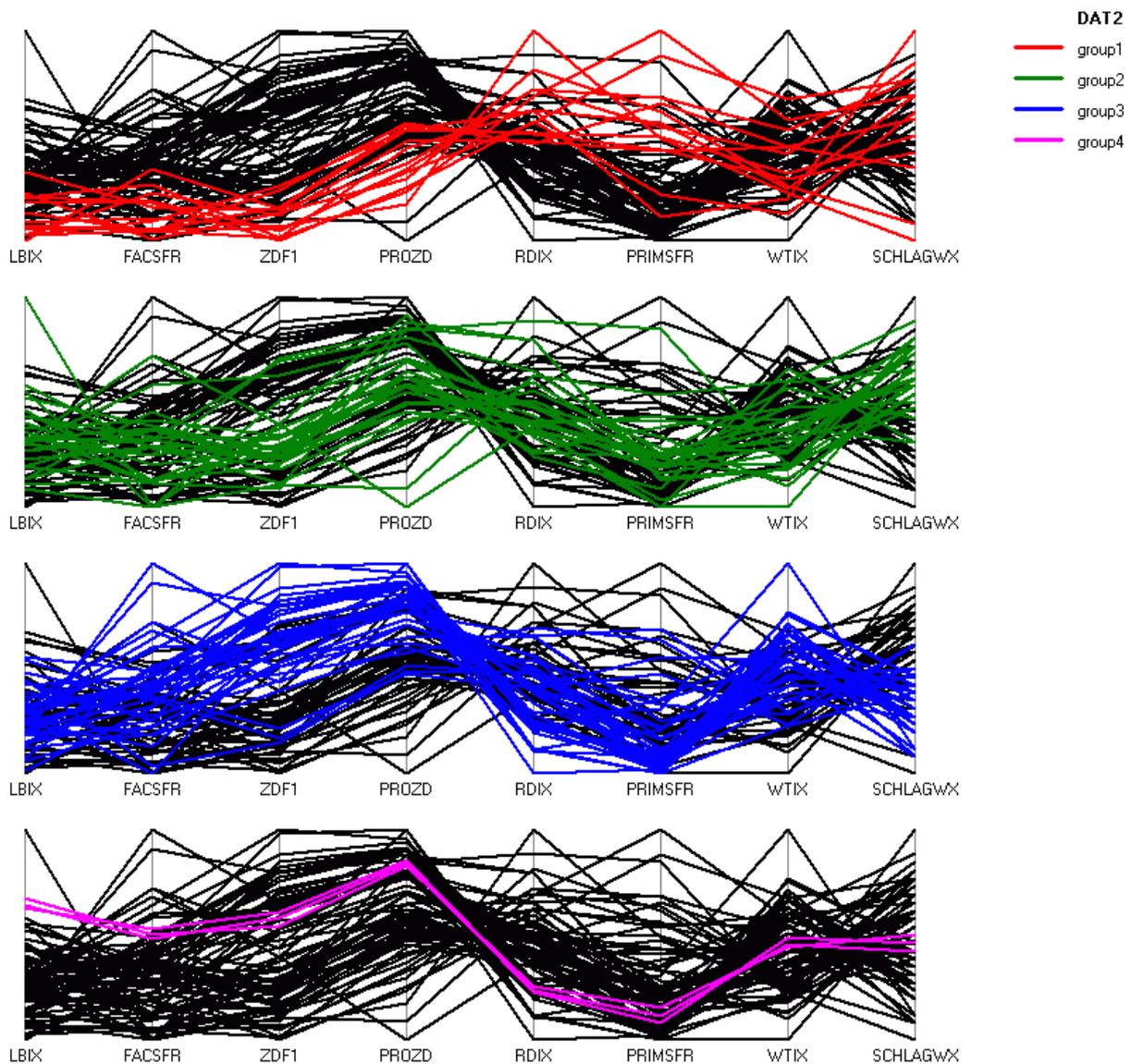


Abb. 2: Umsortierung der Achsen (Reihenfolge der numerischen Merkmale) von Abb. 1.

Die Anzahl N der Steinartefakte pro Inventar variiert zwischen 5 und 4111. Es ist anzunehmen, dass die Messwerte (Durchschnittswerte) umso zuverlässiger sind, je mehr Einzelmessungen Berücksichtigung finden. Das Programm MultiPCP erlaubt es, die Strichstärke entsprechend anzupassen: Abb. 3 verwendet vier Strichstärken, die dünnsten Linien erscheinen für $N \leq 56$, also für N im 1. Quartil. Die Legende in Abb. 3 ist entsprechend ergänzt, so dass auch der Median (122) ablesbar ist. Die Parallelkoordinatenplots zeigen, dass beispielsweise der höchste Wert des Merkmals LBIX von einem Inventar mit vergleichsweise wenigen Steinartefakten stammt.

Möchte man nun genau wissen, um welches Inventar es sich hierbei handelt, so erlaubt das i-Werkzeug im Programm MultiPCP wie in einem GIS, die Daten zu diesem Inventar abzurufen – auch die nicht im Diagramm gezeigten. In vorliegendem Fall handelt es sich um ein Inventar mit dem Namenskürzel va und $N = 5$.

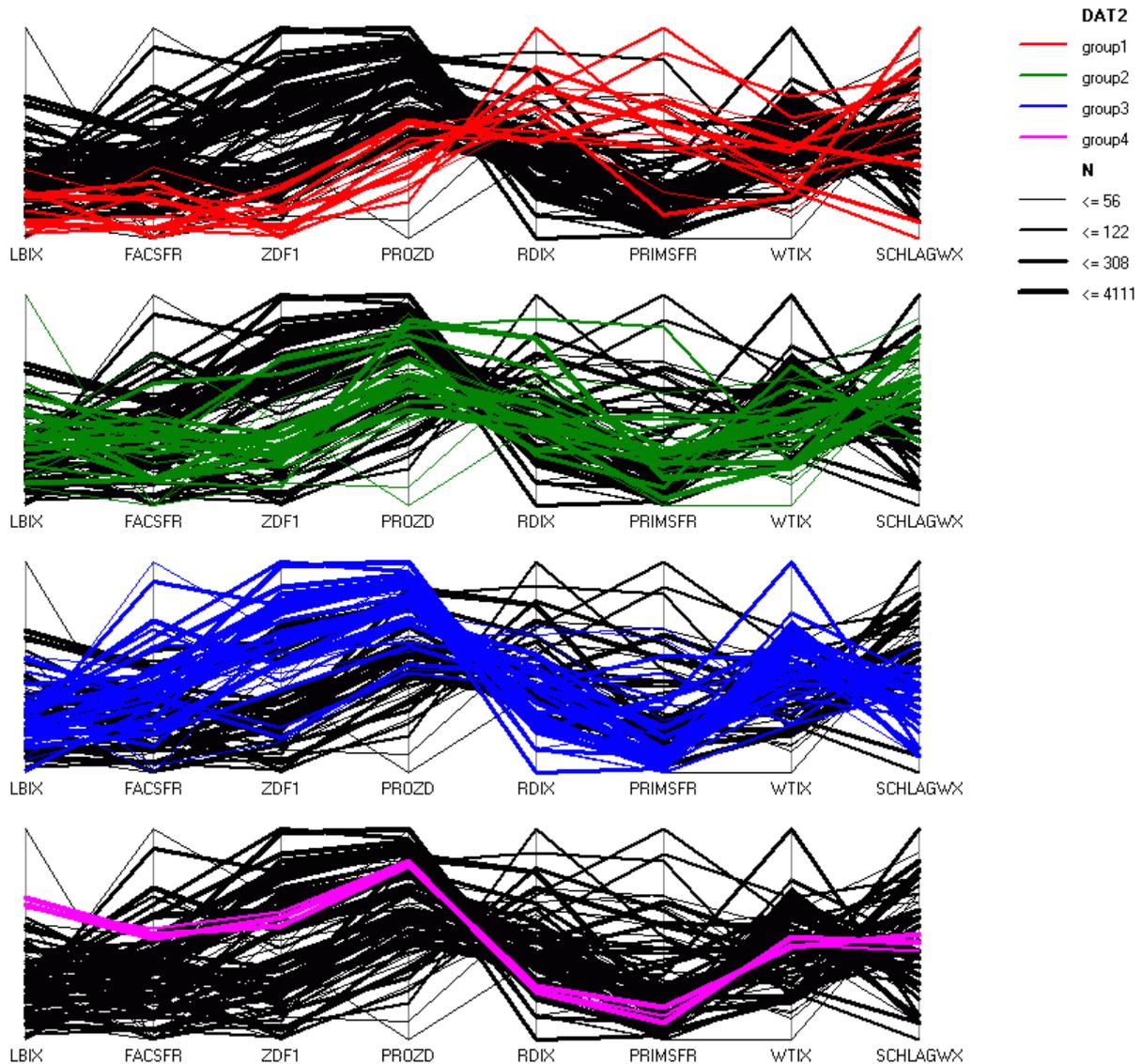


Abb. 3: Berücksichtigung der Anzahl der Artefakte pro Inventar.

Es besteht außerdem die Möglichkeit, die Werte des Inventars mit dem Namens Kürzel *va* durch Punkte besonders zu markieren (siehe Abb. 4). So wird deutlich, dass dieses Inventar auch für das Merkmal WTIX einen extremen Wert – nämlich den niedrigsten aller Inventare – aufweist.

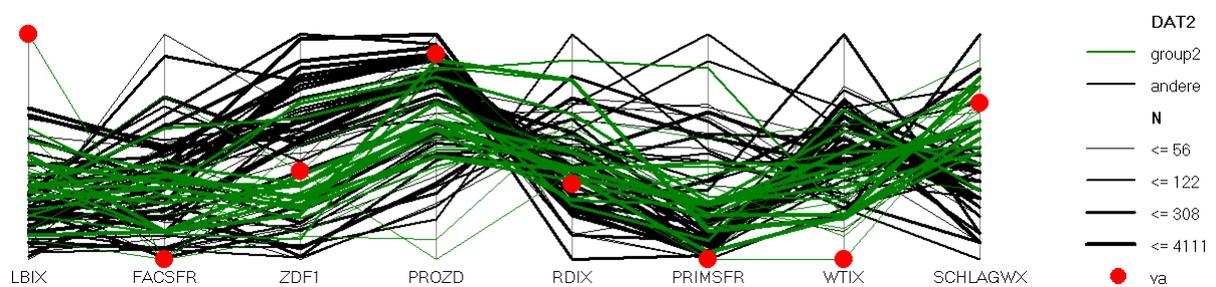


Abb. 4: Das Inventar mit dem Namens Kürzel *va* ist besonders hervorgehoben.

Dagegen gehört der mit Abstand höchste Wert für das Merkmal WTIX zu einem Inventar mit $N=291$, der hohe Wert von WTIX ist also nicht aufgrund der geringen Zahl der betrachteten Steinartefakte zu erklären (siehe Abb. 5). Bei anderen Datenbeständen hat die Analyse mit

Parallelkoordinatenplots gezeigt, dass solche besonders hohen bzw. niedrigen Werte des öfteren auf Eingabefehler zurückzuführen sind.

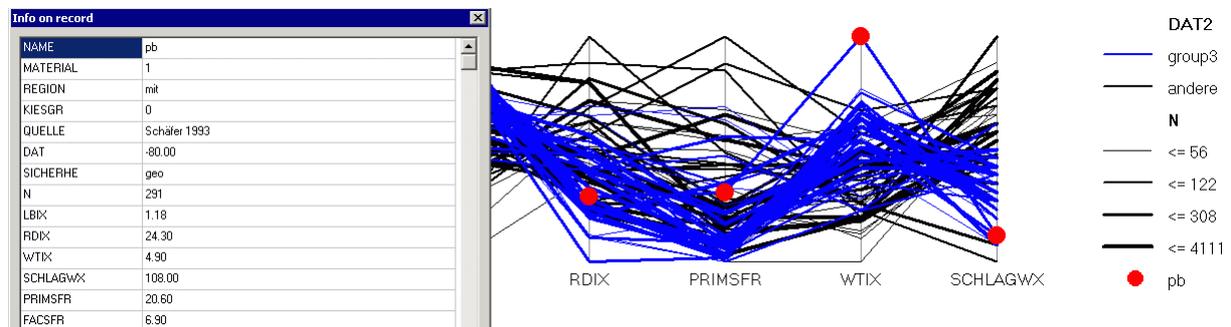


Abb. 5: Ausschnitt aus der Benutzeroberfläche von MultiPCP: Das Inventar mit dem höchsten WTIX-Wert ist hervorgehoben; es handelt sich um das Inventar mit Namenskürzel pb; das Fenster im Vordergrund zeigt die Daten zu diesem Inventar.

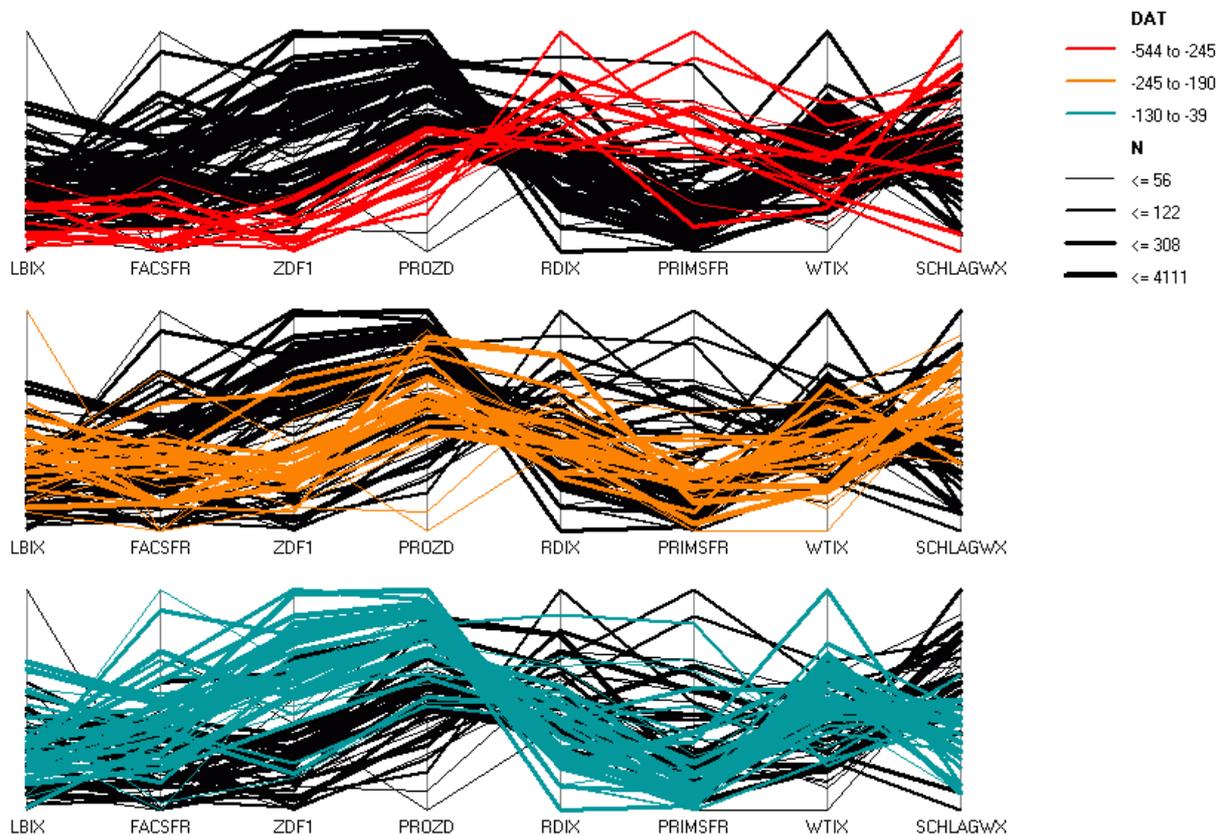


Abb. 6: Auswertung der Datierung mit Jahreszahlen: Die Zeitintervalle sind durch die von Herrn Weber angegebenen Warmzeiten bestimmt.

In Abb. 6 sind die Farben in den drei Parallelkoordinatenplots durch Jahreszahlen-Intervalle festgelegt. Dabei schließt jedes Intervall eine Warmzeit ein. Anzunehmen ist, dass die Inventare einer Warmzeit untereinander ähnlich sind. Die Inventare der ältesten betrachteten Warmzeit sind unverkennbar mit „group1“ identisch.

Wie Abb. 6 zeigt, trennt das Merkmal PROZD besonders gut die Inventare der frühesten betrachteten Warmzeit von allen späteren Inventaren. Eine Ausnahme bilden zwei Inventare mit kleinen Artefaktanzahlen, nämlich die Inventare mit Namenskürzel arn bzw. ms mit N=5 bzw. N=17.

Mithilfe des i-Werkzeugs lässt sich ermitteln, in welchen Bereichen das Merkmal PROZD innerhalb jeder Warmzeit variiert: Für die früheste Warmzeit ist der höchste Wert 69, die PROZD-Merkmale der mittleren Warmzeit (bei Inventaren ab N=20) bewegen sich zwischen 58 und 90, während für die Inventare der letzten betrachteten Warmzeit nur Werte ab 63 auftreten.

Wie die Beispiele zeigen, sind Parallelkoordinatenplots für Datenbestände dieser Größenordnung gut einsetzbar, nicht nur um vorgegebene Gruppierungen zu visualisieren, sondern auch um problematische oder fehlerhafte Daten zu finden. Parallelkoordinatenplots können helfen, den Zusammenhang zwischen einzelnen Merkmalen zu untersuchen und schließlich die Struktur eines Datenbestandes zu verstehen.

Literaturhinweise:

Brunsdon, C., Fotheringham, A. S. & Charlton M. E. (1997). An Investigation of Methods for Visualising Highly Multivariate Datasets.

<http://www.agocg.ac.uk/reports/visual/casestud/brunsdon/brunsdon.pdf>

Cook, D. & Swayne, D.F. (2007). Interactive and Dynamic Graphics for Data Analysis: With R and GGobi (Use R) (S. 24–34). New York: Springer.

Herzog, I. & Siegmund, F., Parallel Coordinate Plots in Archaeology. Akzeptiert für: Proceedings volume GfKI/IFCS 2011.

Inselberg, A. (1985). The plane with parallel coordinates. *The Visual Computer*, 1, 69–91.

Weber, Th. (2009). The Lower/Middle Palaeolithic transition - is there a Lower/Middle Palaeolithic transition? *Preistoria Alpina* 44: 1–6

Irmela Herzog
LVR-Amt für Bodendenkmalpflege im Rheinland
Endenicher Str. 133
53115 Bonn
mail: i.herzog at LVR.de