

K 3.6 Multivariate Graphiken

3.6.1 Parallelkoordinatenplots (für stetige Variablen)

Um viele Variablen gleichzeitig darzustellen. Matrizen von Streudiagrammen sind nur für ein Paar Variablen brauchbar. Parallele Koordinaten Plots sind von AI Inselberg in den achziger Jahren vorgeschlagen worden.

- Jede Variable hat ihre eigene vertikale Achse. Als Default wird der Wertebereich auf $[0, 1]$ transformiert.
- Jeder Fall wird durch einen Polygonzug mit Knoten auf den Achsen bei den entsprechenden Werten dargestellt.
- Die Skalierung und die Reihenfolge der Achsen üben einen wesentlichen Einfluss auf den Plot aus.

Beispiel von Parallelkoordinatenplots: Olympische Zehnkampf (Sydney 2000 und Athens 2004)

- Für alle 10 Disziplinen gibt es die tatsächliche Leistung (in Sekunden oder Metern) und die dafür vergebene Punktzahl.
- Es gab 24 und 28 Teilnehmer, die in 2000, bzw. 2004, für alle Disziplinen Punkte bekommen haben. In beiden Olympischen Spielen gab es insgesamt 39 Teilnehmer.
- Mögliche graphische Darstellungen:
 - Histogramme
 - Boxplots
 - Streudiagramme
 - Parallelkoordinatenplots

Interaktive Optionen für Parallelkoordinatenplots

- Abfrage
- Selektion — Punkte (Linien, Winkel)
- Skalierung
 - Invertierung
 - gemeinsame Skalierung
 - Ausrichtung nach Statistiken oder nach Fällen
- Neuordnen und Sortierung von Variablen
 - per Hand
 - nach der Reihenfolge der Variablen
 - nach Statistiken der Variablen (für ausgewählte Daten)
- Darstellungsmöglichkeiten
 - alphablending
 - Hotselektion
 - Stutzen
 - Boxplots

3.6.2 Mosaic Plots (für kategoriale Variablen)

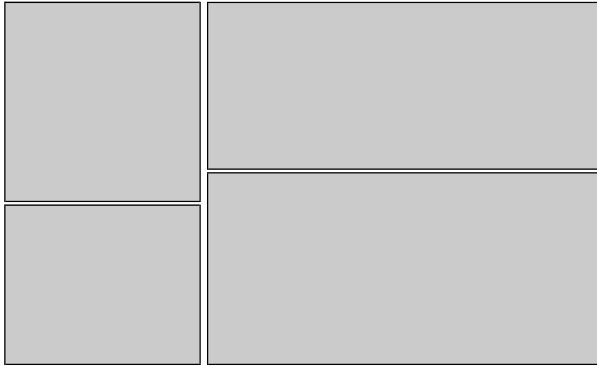
Mosaicplots stellen multivariate kategoriale Kombinationen in Rechtecken dar. Deren Größe zeigt (normalerweise) die Anzahl und deren Lage spiegelt die Kombination wider.

1. Zuerst wird die horizontale Achse nach der ersten Variable aufgeteilt (so daß ein eindimensionales Mosaic Plot einem Spine Plot gleicht).



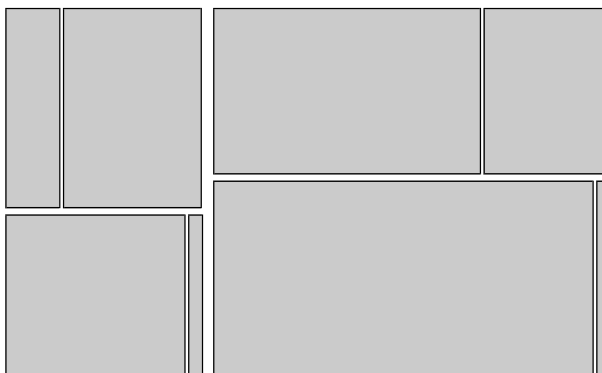
Frau arbeitet/nicht.

2. Dann wird jede Spalte vertikal nach der zweiten Variable aufgeteilt. Daraus ergeben sich den zweidimensionalen Kombinationshäufigkeiten entsprechende Rechtecke.



Frau arbeitet/nicht, älter als 38/nicht.

3. Diese Rechtecke können im Prinzip immer weiter horizontal und vertikal aufgeteilt werden. Ein Mosaic Plot für acht binäre Variablen hat höchstens $256 (= 2^8)$ Rechtecke. Der Übersicht wegen werden leere Zellen (im Datensatz nicht vorhandenen Kombinationen) mit 0 markiert.



Frau arbeitet/nicht, älter als 38/nicht, Kind unter 4/nein

Beispiel von Mosaicplots:

Der Rochdale Datensatz

665 Haushalte aus Rochdale, England, haben an einer Umfrage teilgenommen. Unter anderem wird die Information in 8 binären Variablen zusammengefasst:

wEcon Frau arbeitet/nicht

husUnemp Mann arbeitslos/nicht

c4 Kind ≤ 4

Wife38 Frau älter als 38/nicht

WifeEdx Frau hat eine Ausbildung/nicht

HusEdx Mann hat eine Ausbildung/nicht

Asi Familie aus Asien

OHwork Andere im Haushalt arbeiten/nicht

Mosaic Plot Variationen

- Anzahl
- Gleiche Bingröße
- Multiple Säulendiagramme
- Fluctuationsdiagramme
- Gewichtete Mosaicplots
- Doubledeckerplots

Interaktive Optionen für Mosaicplots

- Abfrage
- Selektion
- Anzahl der Variablen
- Reihenfolge der Variablen
- Reihenfolge der Kategorien innerhalb der Variablen
- Variationen (einschließlich Größe und Seitenverhältnis)