

Explorative Daten Analyse (EDA)

- **Visuelle Exploration**

Da das Ziel der EDA die Extraktion von qualitativer Information ist, kommen primär Graphiken zum Einsatz.

- **Extensive Benutzer-Interaktion**

Die explorative Arbeitsweise bietet im Allgemeinen keine beste, ausschließliche Methode an. Daher muss der Anwender flexible Werkzeuge zur Modifikation von Graphiken und Daten besitzen.

- **Re-expressing**

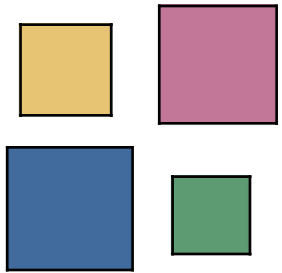
Dynamische Transformationen, wie z.B.

- Box-Cox-Transformation
- Ausreißerelimination

sollen möglichst schnell viele verschiedene Ansichten der Daten liefern.

- **Unterstützung von Meta Daten**

Da das Verständnis der Daten im Vordergrund steht, reicht es meist nicht aus, nur "die" Datenmatrix zu speichern. Daten aus anderen Relationen, Ergebnisse etc. sollen zur Analyse hinzugefügt werden.

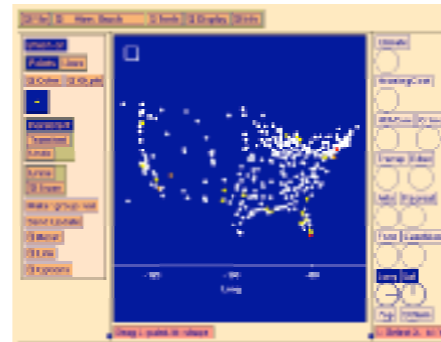


ISG: Historie

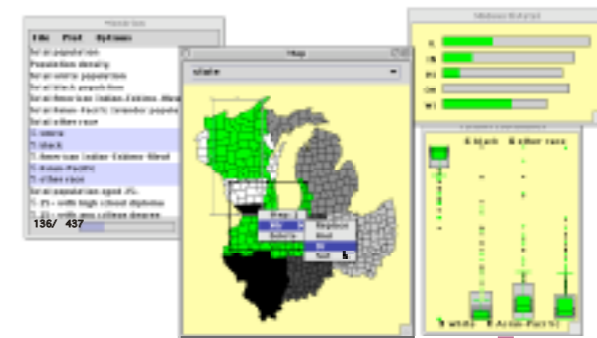
1983
SPLOM
Becker et al.



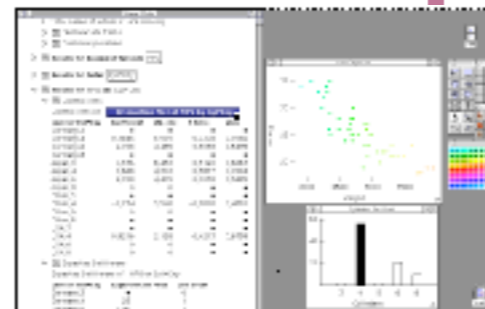
1991
XGobi
Buja et al.



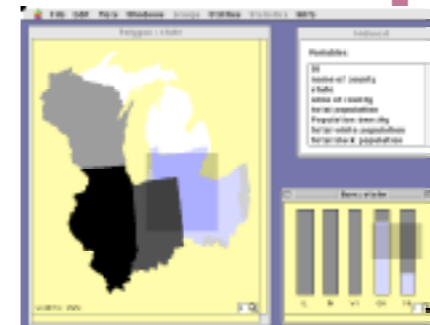
1997
Mondrian
Theus



1973
PRIM-9
Tukey et al.

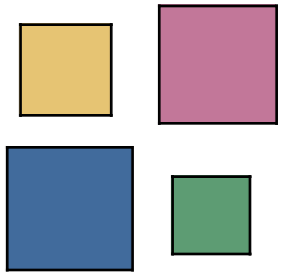


1985
DataDesk
Velleman



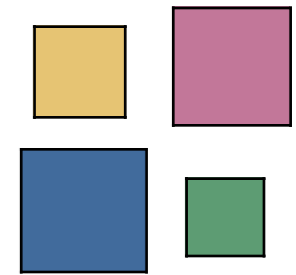
1993
MANET
Unwin et al.





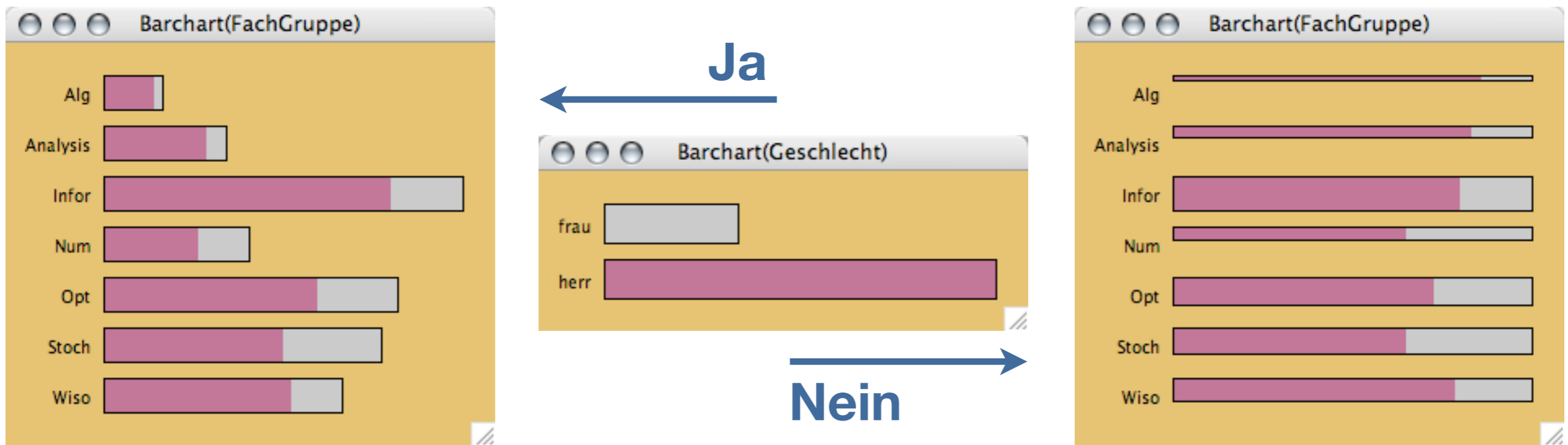
Interaktive Statistische Graphik

- **Interaktionen**
 - Selektion mit linked Highlighting
 - Ändern von Parametern der Plots
- **ISG \neq Dynamische Graphik**
- **Vier Grundelemente**
 - **Highlighting**
Hervorgehobene Darstellung einer Untermenge der Daten in allen Graphiken
 - **Selektion**
Auswahl einer Untermenge von Daten in einer Graphik
 - **Abfragen**
Weitere, differenzierende Information wird auf Wunsch durch einfache Mausklicks zur Verfügung gestellt
 - **Warnungen**
Nicht offensichtliche Informationen werden vom System geliefert
 - **Interaktive, dynamische Variation**
Sofortige Änderung der Parametrisierung der Graphik

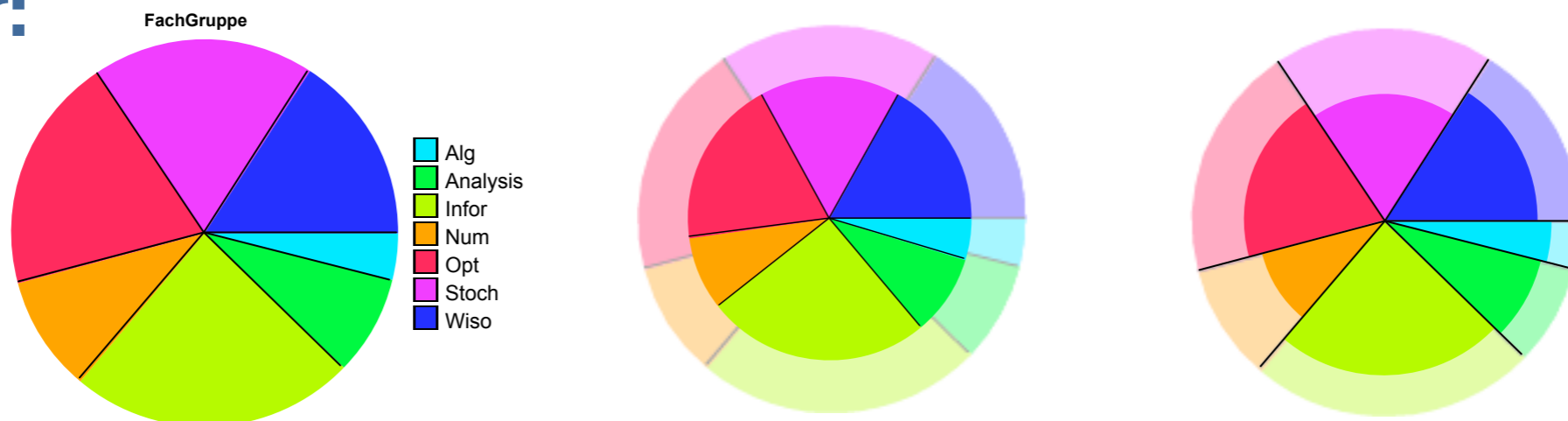


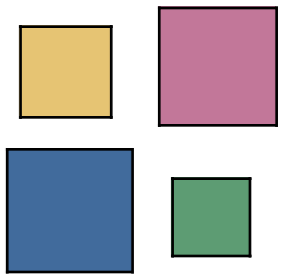
ISG: Highlighting

- Muss das Highlighting vom gleichen Typ sein, wie die Graphik?



- Aber:





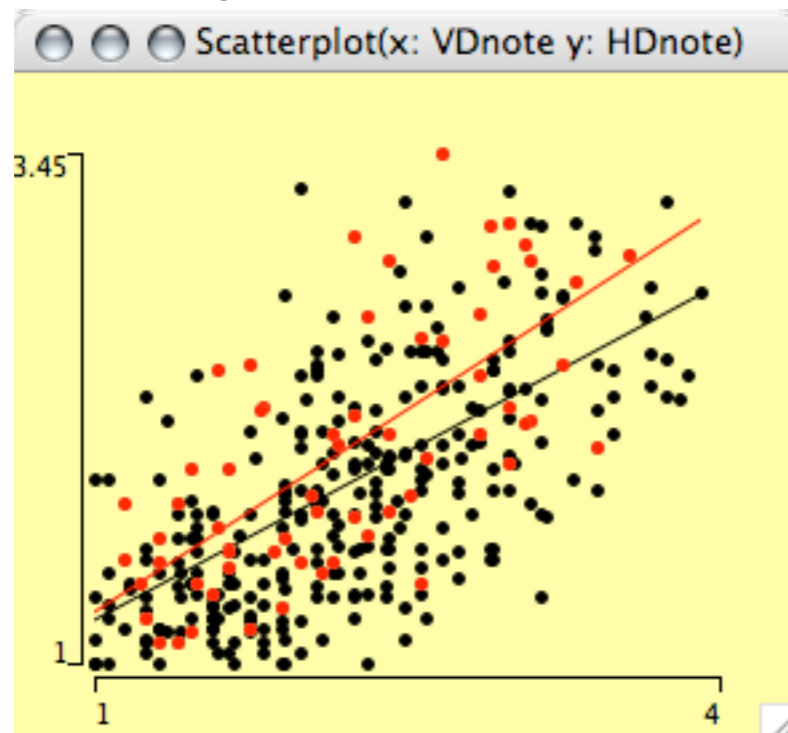
ISG: Highlighting

- Highlighting vs. non-Highlighting

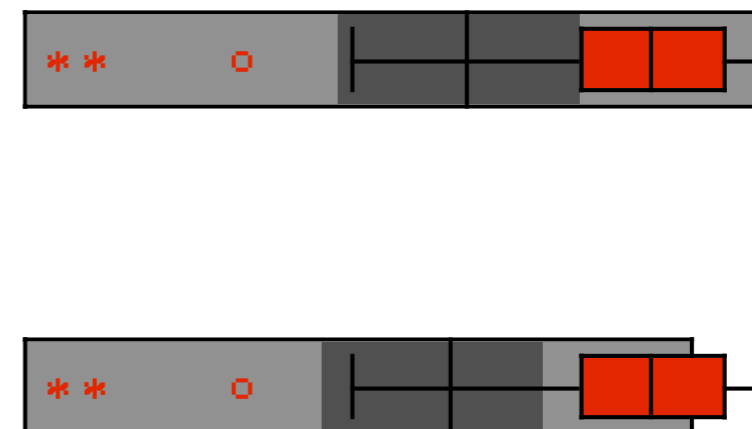
Soll die ausgewählte Menge mit ihrem Komplement oder mit der Gesamtheit der Daten verglichen werden?

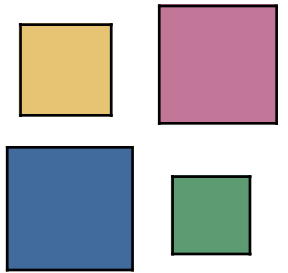
- bei vielen Plots macht es keinen Unterschied: Scatterplots, Parallel Coordinates
- bei manchen Plots schon, z.B. Box-Plots

Scatterplot



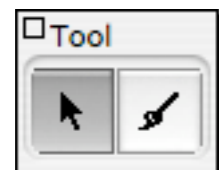
Box-Plot





ISG: Selektions Werkzeuge

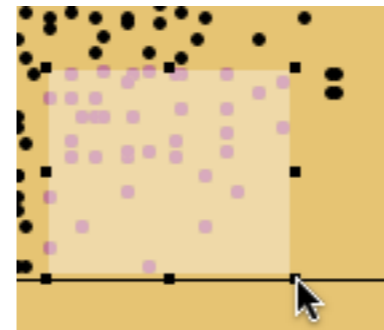
- Für die Selektion gibt es i.A. verschiedene **Werkzeuge**:
 - **Lasso**:
Mit dem Lasso lassen sich beliebige Formen, Untermengen einfangen. Start- und Endpunkt werden immer verbunden.
 - **Drag-Box**:
Die Drag-Box kann Rechtecke in Graphiken selektieren.
 - **Brush**:
Mit der Brush (Bürste) kann man rechteckige Bereiche dynamisch highlighten.
 - **Pointer**:
Der Pointer selektiert einzelne Punkte.
 - **Slicer**:
Mit dem Slicer (Messer) können achsenparallele Intervalle selektiert werden.



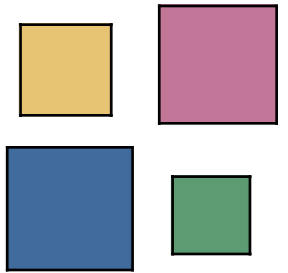
MANET



DataDesk

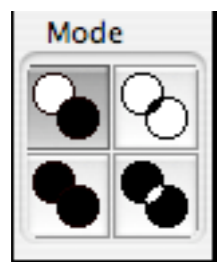


Mondrian



ISG: Selektions Modi

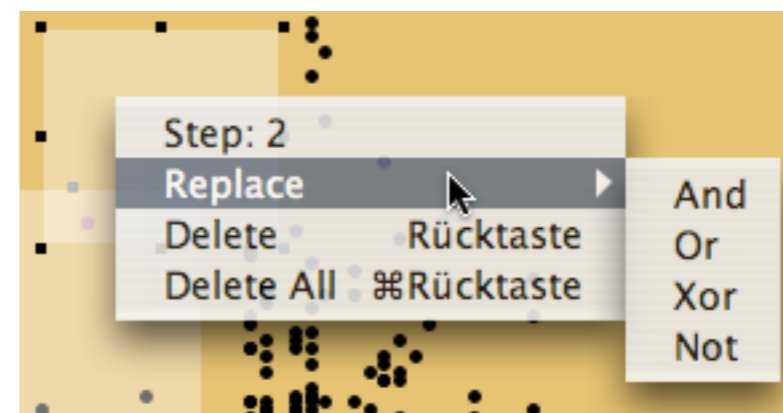
- Einfach / Standard / Default
 - Es wird genau der gerade ausgewählte Bereich selektiert.
- Schnitt / AND / \cap
 - Es bleiben nur die Objekte selektiert, die bereits selektiert waren, und auch in der neuen Selektion enthalten sind.
- Vereinigung / OR / \cup
 - Zu den schon selektierten Daten wird die neue Selektion hinzugefügt.
- Toggle / XOR / \oplus
 - Schon selektierte Objekte werden deselektiert, alle anderen selektiert.
- Nicht / NOT / \neg
 - Die ausgewählten Objekte werden aus der Selektion herausgenommen.



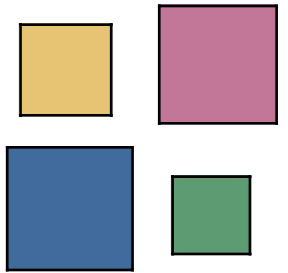
MANET



DataDesk



Mondrian



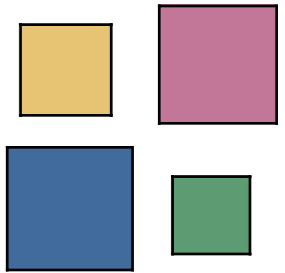
ISG: Selektionssequenzen

- **Probleme bei der Selektion**

- Selektion existiert nur als Menge der selektierten Daten
- Komplexe Selektionen sind schwer zu erstellen
- Vorhandene Selektionsmenge kann nicht beschrieben werden
- Fehler bei der Selektion sind fatal
- Änderungen an der aktuellen Selektion können nicht vorgenommen werden

- **Lösung: Selektionssequenzen**

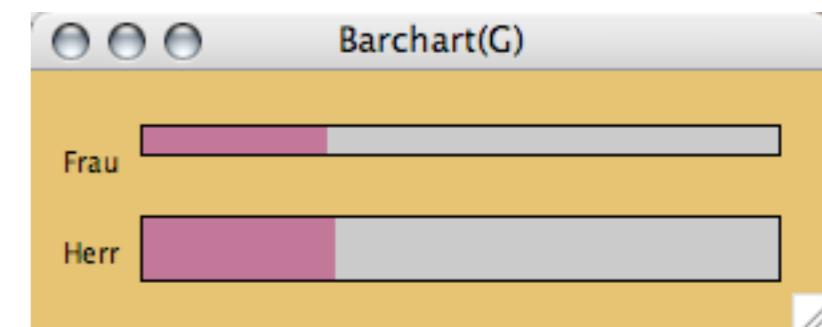
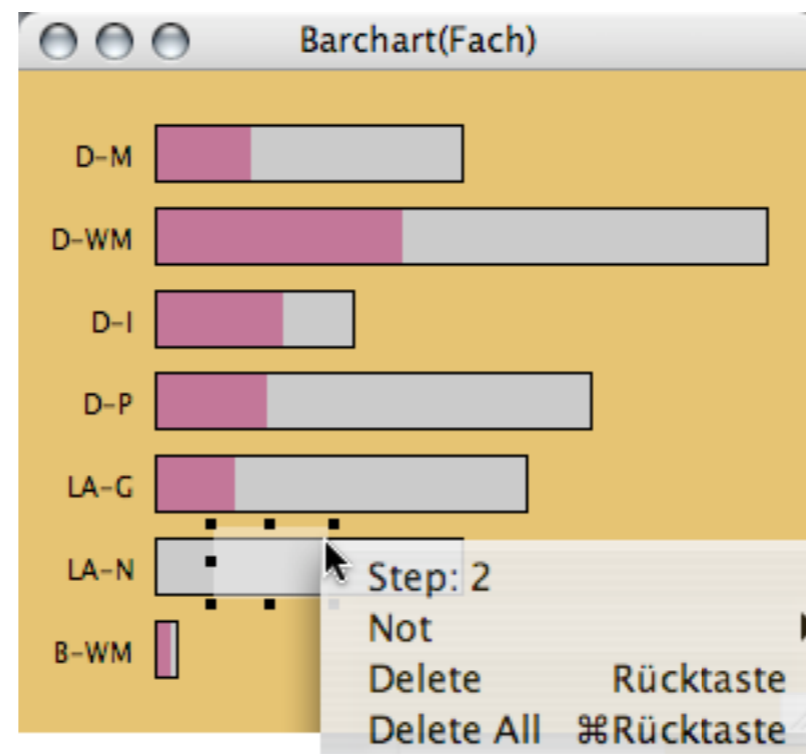
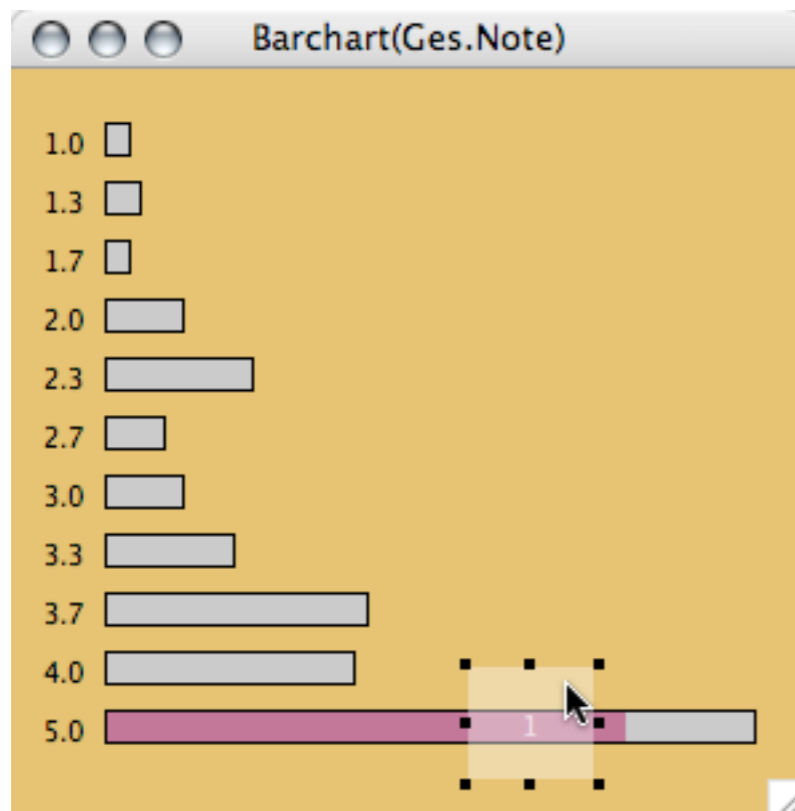
- Speicherung der Selektionen in parametrischer Form, d.h. für jede Selektion speichere:
 - Plot
 - Koordinaten der Selektion
 - Nummer in Sequenz
 - Selektions Modus
- Jeder Selektionsschritt ist erkennbar
- Änderungen können an vorhandenen Selektionsobjekten vorgenommen werden.

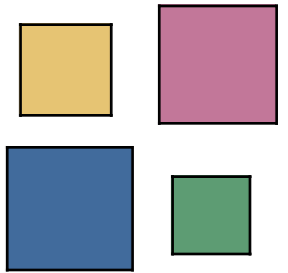


ISG: Selektionssequenzen

- Implementierungen in
 - MANET
 - Mondrian
 - (Spotfire)
- Beispiel: Daten aus der LA1 Vorlesung

“Unterscheidet sich die Durchfallquote bei Männern und Frauen, wenn man “LA nicht vertieft” nicht betrachtet?”





ISG: Selektionssequenzen

- Selektionssequenzen sind gerichtet, d.h. für drei Selektionsmengen A, B, C gilt:

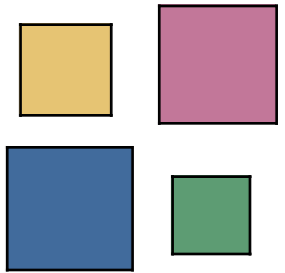
$$A \text{ OR } B \text{ AND } C = A \text{ OR } (B \text{ AND } C) \neq ((A \text{ OR } B) \text{ AND } C)$$

bzw.

$$A \text{ U } B \cap C = A \text{ U } (B \cap C) \neq ((A \text{ U } B) \cap C)$$

d.h. explizite Linksklammerung!

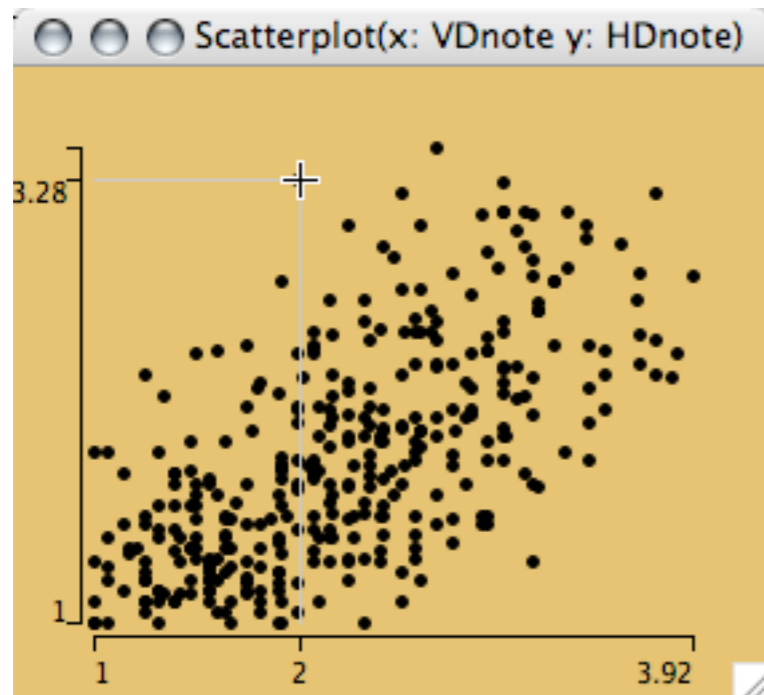
- Im Allgemeinen entspricht dies jedoch der Sichtweise des Benutzers
- Selektionssequenzen lassen sich ohne weiteres in SQL übersetzen. Auch hier ist die Reihenfolge zu beachten



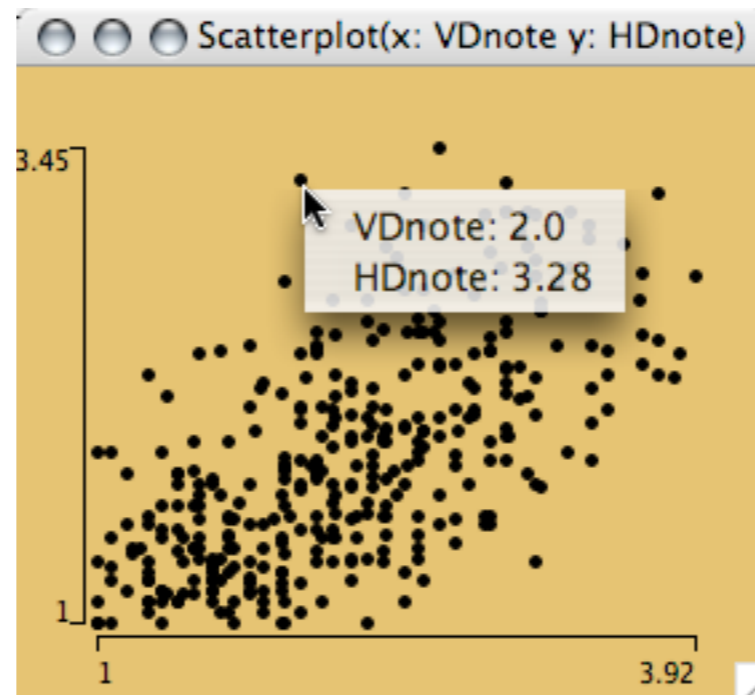
ISG: Abfragen

- Da Grafiken meist nur qualitative Information liefern, sollen genaue Quantitäten über Abfragen bereitgestellt werden.
- Die Information sollte dabei in verschiedenen Ebenen (Granularität) bereitgestellt werden.
- Beispiel: Scatterplot

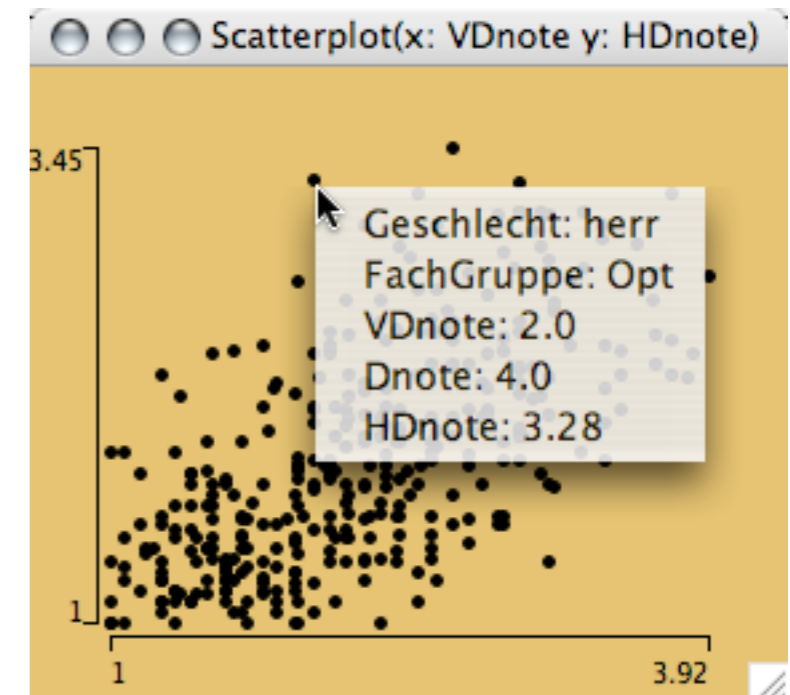
Orientierung

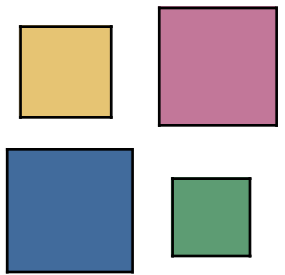


Standard



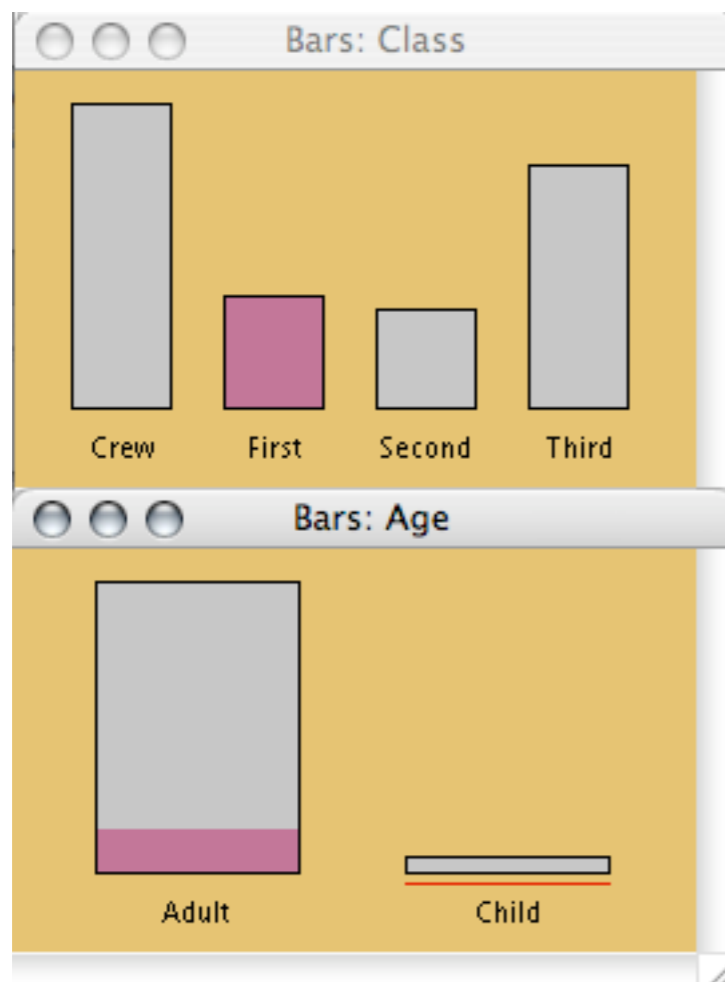
Erweitert





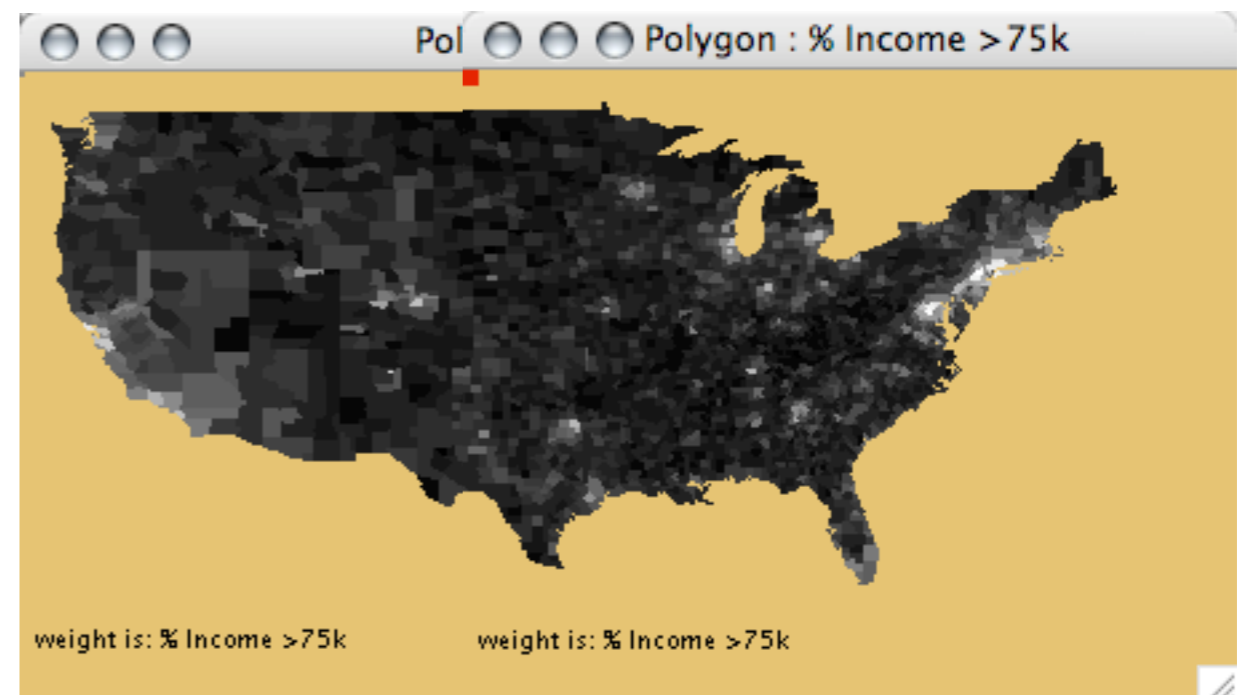
ISG: Warnungen

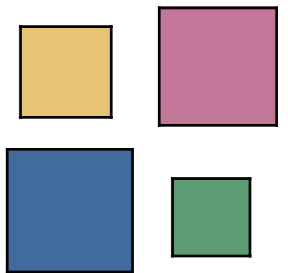
- Auf nicht offensichtliche Informationen soll das System durch Warnungen hinweisen, oder
“What you see is not what you think you see!”
- Beispiele



unsichtbare
Anteile

Daten außerhalb des Viewports





ISG: Typische Interaktionen

- Selektion und Highlighting ✓
- Umordnen von Elementen, z.B.
 - x-, y-Achse im Scatterplot vertauschen
 - Achsen im Parallel Coordinate Plot umsortieren
 - Reihenfolge der Kategorien in Barcharts ändern
 - Reihenfolge von Variablen in Mosaic Plots
- Dito mit automatischem Sortieren, z.B. nach
 - absolutem oder relativem Highlighting
 - Name
 - Parametern
 - ...
- Änderung der Skalierung
 - Zooming
 - Startpunkt und Binbreite in Histogrammen
 - Farben