

# Warenkorbanalyse mit Hilfe der Statistiksoftware R

**Michael Hahsler<sup>(1)</sup>, Kurt Hornik<sup>(2)</sup>  
& Thomas Reutterer<sup>(3)</sup>**

- (1) Institut für Informationswirtschaft, WU-Wien.**
- (2) Institut für Statistik und Mathematik, WU-Wien.**
- (3) Institut für Handel und Marketing, WU-Wien.**



## Inhalt des Vortrags

1. Einführung zu Nachfrageverbund und Warenkorbanalyse
2. R-Erweiterungspaket **arules**
3. Transaktionsdaten
4. Analysen
  - Affinitätsanalyse
  - Verdichtung von Transaktionsdaten
  - Assoziationsregeln
5. Zusammenfassung und Ausblick



## Einführung: Nachfrageverbund

- Analyse des Nachfrageverbunds zwischen Bestandteilen von Einzelhandelssortimenten (Produkte, Warengruppen, etc.)
- Verwertung mittels marketingpolitischer Maßnahmen
  - **Category-Managements des Handels:** Platzierung, Preis- und Sonderangebotspolitik, etc. (vgl. Müller-Hagedorn 2005)
  - **Auf kundenindividuellem oder segment-spezifischem Niveau:** Gestaltung maßgeschneiderter Cross- und Upselling-Aktionen innerhalb von Loyalitätsprogrammen (vgl. Mild/Reutterer 2003; Boztug/Reutterer 2006; Reutterer et al. 2006).



## Einführung: Nachfrageverbund

- **Erklärende Ansätze:** Auswirkungen (Kreuzeffekte) von Marketing-Aktionen in einer Warengruppe auf das Kaufverhalten von Kunden in einer anderen Warengruppe (*Hruschka et al. 1999, Manchanda et al. 1999; Russell/Petersen 2000; Boztug/Hildebrandt 2006*)
  - In der Regel wegen ihrer Komplexität nur auf sehr kleine Ausschnitte des Sortiments beschränkt (*vgl. Boztug/Reutterer 2006*).
- **Explorative Ansätze:** große Datenmengen für die Analyse zu verdichten, Verbundeffekte effizient aufzufinden und für den/die Benutzer zu visualisieren (*Berry/Linoff 1997, Schnedlitz et al. 2001*).



# Explorative Warenkorbanalyse

Ansatz	Ausgewählte Quellen	Methodische Kurzcharakteristik	Aggregationsniveau
<b>Affinitätsanalyse</b>	Böcker (1975, 1978) Merkle (1979, 1981) Dickinson/Harris/Sircar (1992) Julander (1992) Schnedlitz/Kleinberg (1994)	Repräsentation einer Verbundmatrix bestehend aus paarweisen Assoziationsmaßen	Aggregiert
<b>Prototypenbildende Clusterverfahren</b>	Schnedlitz/Reutterer/Joos (2001) Decker/Monien (2003) Decker (2005) Reutterer et al. (2006)	Verdichtung von Verbundbeziehungen in Warenkörben zu prototypischen Warenkorbklassen	Disaggregiert (segment-spezifisch)
<b>Assoziationsregeln</b>	Agrawal/Srikant (1994) Hildermann et al. (1998) Decker/Schimmelpfennig (2002) Brin et al. (1997) Brijs et al. (2004) Hahsler/Hornik/Reutterer (2006)	Generierung von Verbundregeln als Implikationen des Kaufs einer Warengruppe A auf eine (oder mehrere) andere Warengruppe(n) B (C, D, ...)	Aggregiert



## R-Erweiterungspaket arules



- **R** (<http://www.r-project.org>) ist eine freie Programmierumgebung für statistische Anwendungen.
- **Comprehensive R Archive Network** (<http://cran.r-project.org/>)  
Bezug des R Basissystems und von Erweiterungspaketen wie **arules**.
- **arules** stellt die Infrastruktur für explorative Warenkorbanalyse zur Verfügung und ermöglicht unter anderem: **Affinitätsanalyse**, **Prototypenbildende Clusterverfahren** und die Suchen nach **Assoziationsregeln**.



## Transaktionsdaten

- Transaktionsdaten fallen über Scanningsysteme am Kassen-Check-Out („Point-Of-Sale“; POS) in Form von **Warenkörben** an (*Hruschka 1991; Berry/Linoff 1997*)
- Typische Darstellungen als **binäre Kaufinzidenzmatrix**:

		Produkte				
		Milch	Brot	Butter	Bier	...
Transaktionen	1	1	1	0	0	
	2	0	1	1	0	
	3	0	0	0	1	
	4	1	1	1	0	
	5	0	1	1	0	



## Beispiel: Supermarktdaten

```
R> library("arules")
```

```
R> data("Groceries")
```

```
R> summary(Groceries)
```

transactions as itemMatrix in sparse format with **9835 rows**  
(elements/itemsets/transactions) and **169 columns** (items)

most frequent items:

<b>whole milk</b>	other vegetables	rolls/buns	soda	yogurt	(Other)
<b>2513</b>	1903	1809	1715	1372	34055

element (itemset/transaction) length distribution:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	2.000	3.000	4.409	6.000	32.000

includes extended item information - examples:

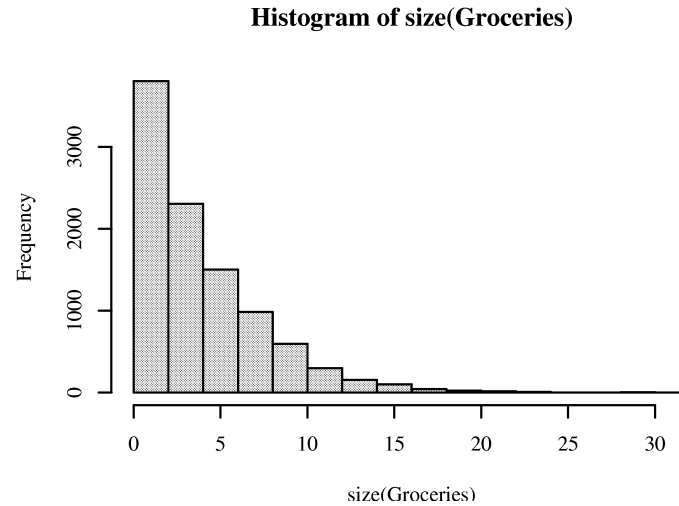
	labels	level2	level1
1	frankfurter	sausage	meet and sausage
2	sausage	sausage	meet and sausage



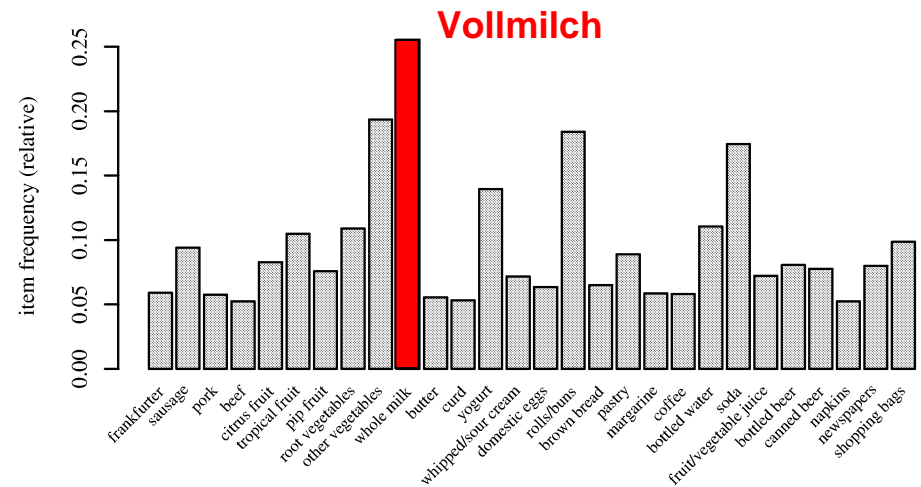


## Beispiel: Supermarktdaten

```
R> which(size(Groceries) > 25)
[1] 1092 1217 2939 2974 4431
    5611 9002
R> hist(size(Groceries))
```



```
R> itemFrequencyPlot(Groceries,
  support = 0.05)
```



## Affinitätsanalyse

- ‚Frequenzmatrix‘ enthält die gemeinsamen Kaufhäufigkeiten für Paare von Warengruppen (vgl. Böcker 1978, Merkle 1981, Hruschka 1991).

```
R> ct <- crossTable(Groceries)
```

```
R> ct[1:5, 1:5]
```

	frankfurter	sausage	liver loaf	ham	meat
frankfurter	580	99	7	25	32
sausage	99	924	10	49	52
liver loaf	7	10	50	3	0
ham	25	49	3	256	9
meat	32	52	0	9	254

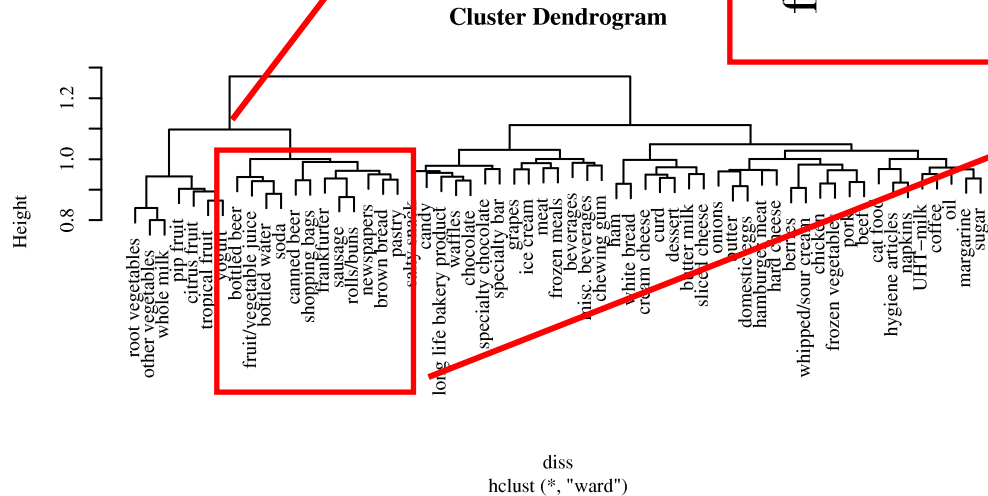
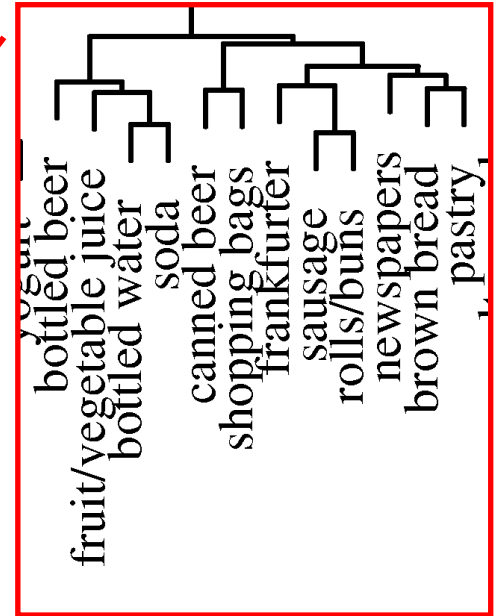
Kann mittels eines geeigneten Assoziationskoeffizienten in eine „Verbundmatrix“ überführt werden.



# Affinitätsanalyse

- Verdichtung durch hierarchische Clusteranalyse (vgl. Merkle 1979; Bordemann 1986; Decker/Schimmelpfennig 2002)

```
R> diss <- dissimilarity(Groceries[,
  itemFrequency(Groceries) > 0.02], method =
  "Jaccard", which = "items")
R> hc <- hclust(diss, method = "ward")
R> plot(hc)
```



diss  
hclust (\*, "ward")



## Affinitätsanalyse

- **Absatzpolitische Relevanz:** Verbundplatzierungen im
  - Verkaufsraum oder für
  - Werbemedien (zum Beispiel Flugblätter, Kataloge).
- **Category-Management:** Nachfragerorientierte Bildung sogenannter '*Categories*' (vgl. Bordemann 1986, Zielke 2002, Müller-Hagedorn 2005).



## Verdichtung von Transaktionsdaten zu Prototypen

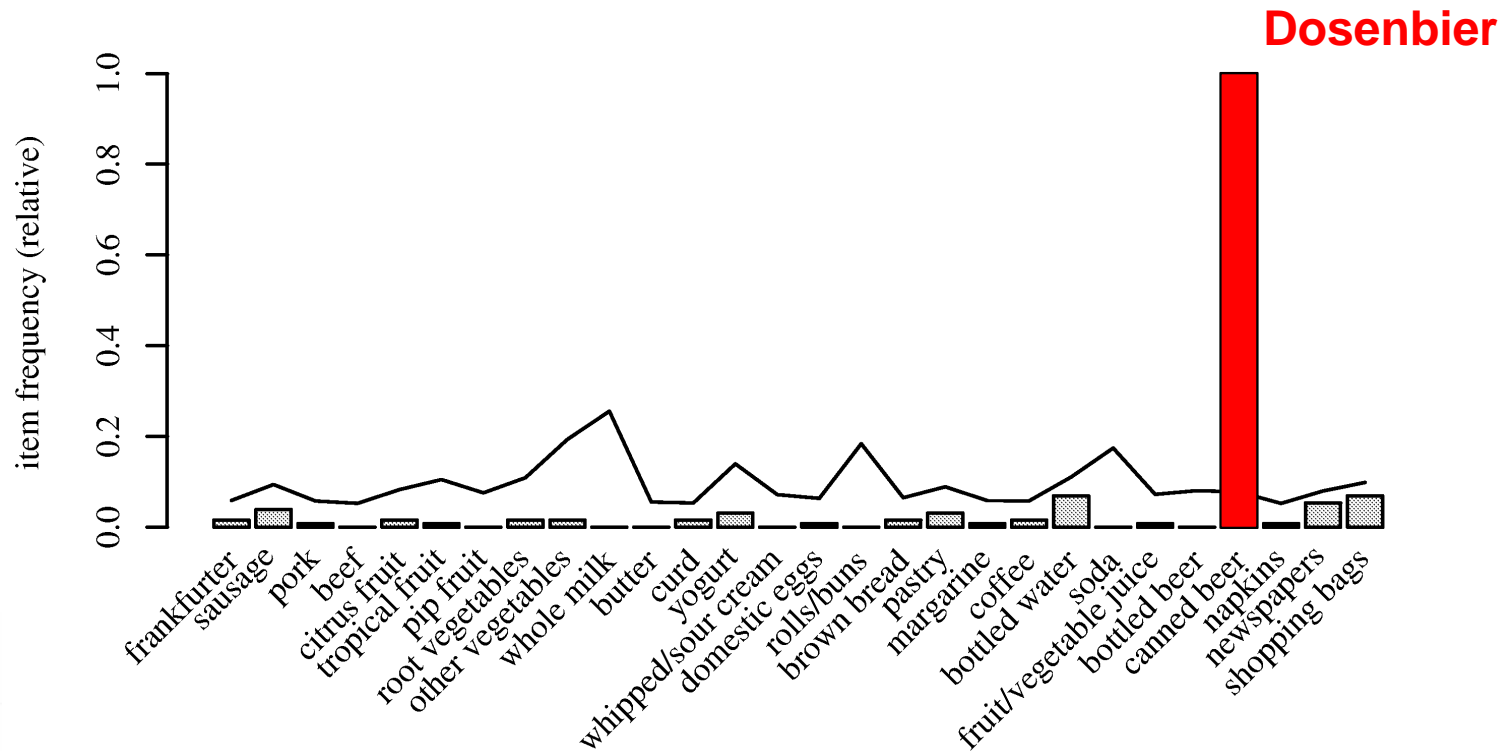
- **Warenkorb-Typologie:** Einzelne Transaktionen zu einer (kleinen) Gruppe von prototypischen Warenkörben zuordnen (*vgl. Schnedlitz et al. 2001*).
- Die Verdichtung zu sog. **Warenkorb-Protoypen** kann beispielsweise durch Clusterbildung basierend auf den binären Transaktionsdaten realisiert werden.

```
R> samp <- sample(groc, 2000)
R> diss <- dissimilarity(samp, method = "Jaccard")
R> library("cluster")
R> clust <- pam(diss, k = 5)
```



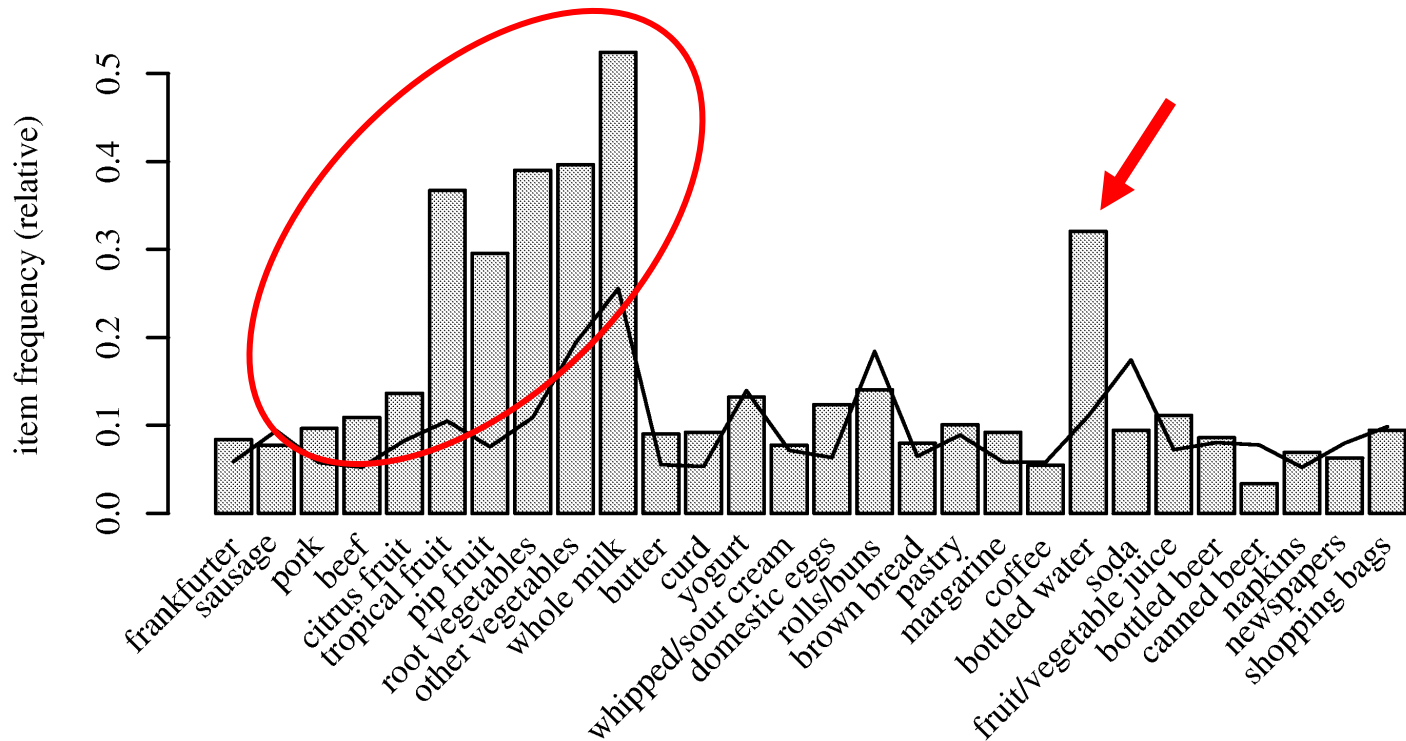
## Verdichtung von Transaktionsdaten zu Prototypen

```
R> itemFrequencyPlot(samp[clust$clustering == 2],
  population = Groceries, support = 0.05)
```



## Verdichtung von Transaktionsdaten zu Prototypen

```
R> itemFrequencyPlot(samp[clust$clustering == 5],  
population = Groceries, support = 0.05)
```



## Verdichtung von Transaktionsdaten zu Prototypen

- **Hauptvorteil:** disaggregierte bzw. segmentspezifische Betrachtungsweise (z.B. bei Loyalitätsprogrammen).
- Erweiterung in Richtung dynamischer Kundensegmentierung *(vgl. Reutterer et al. 2006)*.
- Vorverdichtung und Warenkorbselektion für segmentspezifisch maßgeschneiderte Erklärungsmodelle (Kreuzeffekte zwischen Warengruppen) *(vgl. Boztug und Reutterer 2006)*.





## Generierung von bedeutsamen Assoziationsregeln

- Assoziationsregeln (Association Rules) sind ein Ansatz aus dem **Data-Mining** (*Agrawal/Srikant 1994*).
- Analyse der gemeinsamen Kaufhäufigkeiten von Artikeln/Warengruppen in Transaktionsdaten.
- Assoziationsregeln formalisieren Regeln als Implikationen der Art (*Agrawal/Imielinski/Swami 1993*)

**{Brot, Milch}  $\Rightarrow$  {Butter}**

(Support: 1%, Konfidenz: 50%).



## Generierung von bedeutsamen Assoziationsregeln

```
R> rules <- apriori(Groceries, parameter =
  list(support = 0.001, confidence = 0.2))
```

```
R> summary(rules)
```

set of **21574 rules**

rule length distribution (lhs + rhs):

1	2	3	4	5	
1	620	9337	9824	1792	
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	3.000	4.000	3.593	4.000	5.000

summary of quality measures:

support	<b>confidence</b>	lift
Min. :0.00102	Min. :0.200	Min. : 0.803
1st Qu.:0.00112	1st Qu.:0.263	1st Qu.: 2.115
Median :0.00132	Median :0.355	Median : 2.756
Mean :0.00196	Mean :0.396	Mean : 3.016
3rd Qu.:0.00193	3rd Qu.:0.500	3rd Qu.: 3.612
Max. :0.25552	Max. :1.000	Max. :35.716



# Generierung von bedeutsamen Assoziationsregeln

## Auswahl von Regeln mit Rindfleisch als Konsequenz

```
R> rulesBeef <- subset(rules, rhs %in% "beef")
R> inspect(SORT(rulesBeef, by = "conf")[1:3])
```

lhs	rhs	support	confidence	lift
1 {root vegetables, whole milk, butter, rolls/buns}	=> {beef}	0.001118	<b>0.4783</b>	9.116
2 {sausage, root vegetables, butter}	=> {beef}	0.001017	0.4545	8.664
3 {root vegetables, butter, yogurt}	=> {beef}	0.001525	0.3947	7.524



## Generierung von bedeutsamen Assoziationsregeln

- Entscheidungsunterstützung in den Bereichen **Sortiments-, Platzierungs- und Werbeplanung auf Artikelebene** (vgl. *Brijs et al. 2004, Van den Poel/De Schamphelaere/Wets 2004*).
- **Vorteile:** Effiziente Algorithmen ermöglichen die gleichzeitige Analyse mit sehr vielen Warengruppen oder einzelnen Artikeln.
- **Probleme:** Die Maße Support, Konfidenz und Lift werden systematisch durch die Kaufhäufigkeit der Artikel beeinflusst (vgl. *Hahsler/Hornik/Reutterer 2006*).



## Zusammenfassung und Ausblick

- Das R-Erweiterungspaket **arules** erleichtert die explorative Warenkorbanalyse.
- Eine Vielzahl von aufschlussreichen Analysen können schnell und einfach durchgeführt werden.
- **arules** ermöglicht die unkomplizierte und rasche Erweiterungen durch die Verwendung weiterer R-Pakete. Beispielsweise durch in **cba** enthaltene Algorithmen für:
  - »moderne Kompressionsverfahren
  - »nachbarschaftsbasierte Clusteralgorithmen

