

Text-Dokument

28

KAPITEL 1. GRUNDLAGEN



Abbildung 1.9: Lego-Prinzip: Zerlege eine komplexe Struktur in einfache Bausteine



Abbildung 1.10: Durchpfeifen: Ein Dataframe wird von Operation zu Operation weitergereicht

1.3.3 Zentrale Bausteine von dplyr

1.3.3.1 Zeilen filtern mit filter

Häufig will man bestimmte Zeilen aus einer Tabelle filtern; `filter`. Zum Beispiel man arbeitet für die Zigarettenindustrie und ist nur an den Rauchern interessiert (die im Übrigen unserer Gesundheitssystem retten [Krämer 2011]), nicht an Nicht-Rauchern; es sollen die nur Umsatzzahlen des letzten Quartals untersucht werden, nicht die vorherigen Quartale; es sollen nur die Daten aus Labor X (nicht Labor Y) ausgewertet werden etc.

Abb. 1.11 zeigt ein Sinnbild für `filter`.

Merk:

Die Funktion `filter` filtert Zeilen aus einem Dataframe.

Schauen wir uns einige Beispiel an; zuerst die Daten laden nicht vergessen. Achtung: "Wohnen" die Daten in einem Paket, muss dieses Paket installiert sein, damit man auf die Daten zugreifen kann.

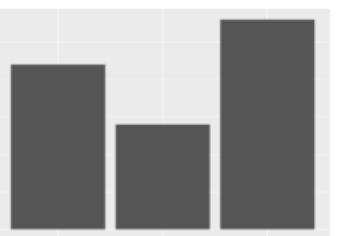
```
data(profiles, package = "okcupidata") # Das Paket muss installiert sein
```

```
df_frauen <- filter(profiles, sex == "f") # nur die Frauen  
df_alt <- filter(profiles, age > 70) # nur die alten Menschen
```

Foliensatz

Slide with R Output

```
library(tidyverse)  
qplot(x = factor(cyl), data = mtcars)
```



Stichwortzettel

