Spalten trennen

Sophie C. Schmidt

2 Mai 2019

## Workflow

Hier lge ich meinen workflow dar, wie ich Spalten, die mehrere Informationen beinhalten, 1. trennen 2. in ein langes Datenformat überführen 3. in ein breites Datenformat überführen kann

library(tidyr)

### Beispieldatensatz

## x  
## 1 <NA>  
## 2 a,b  
## 3 a,b,d  
## 4 b,c  
## 5 d,a,f,e

### Spalten trennen -> die Komma-seperierten mehrfachen Angaben in einzelne Spalten unterteilen

## Warning: Expected 4 pieces. Missing pieces filled with `NA` in 3 rows [2,  
## 3, 4].

## A B C D id  
## 1 <NA> <NA> <NA> <NA> 1  
## 2 a b <NA> <NA> 2  
## 3 a b d <NA> 3  
## 4 b c <NA> <NA> 4  
## 5 d a f e 5

WICHTIG \* man benötigt so viele Spaltennamen (c(“A”, “B”…), wie max unterschiedliche Angaben in der Spalte zu finden sind, die getrennt wird \* es entsteht einiges an NA. Nicht schlimm.

### Die einzelnen Spalten in einzelne Zeilen umwandeln (von breitem zu langem Format überführen)

## id fignr fig  
## 2 2 1 a  
## 3 3 1 a  
## 4 4 1 b  
## 5 5 1 d  
## 7 2 1 b  
## 8 3 1 b  
## 9 4 1 c  
## 10 5 1 a  
## 13 3 1 d  
## 15 5 1 f  
## 20 5 1 e

WICHTIG: \* key bezeichnet einen NEUEN Spaltennamen, in dem die NAMEN der Spalten auftauchen, die von breit zu lang überführt werden \* value bezeichnet den NEUEN Spaltennamen, in dem die WERTE der Spalten auftauchen, die überführt werden \* na.rm entfernt alle NAs aus der Tabelle, die durch das seperate entstanden sind

### In 0 und 1 Werte umrechnen

## id a b c d e f  
## 1 2 1 1 NA NA NA NA  
## 2 3 1 1 NA 1 NA NA  
## 3 4 NA 1 1 NA NA NA  
## 4 5 1 NA NA 1 1 1

WICHTIG: \* key ist der Spaltenname, dessen WERTE zu den neuen Spaltennamen werden

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.