

# Data Exploration

*Vincent Schmalor*

*6 12 2019*

## Einleitung

Dieses Dokument wendet einige Operationen auf den Datensets an, um sie näher kennen zu lernen und zu verstehen

## Funktionen

## Data-set vorbereiten

## Rohdaten

Kurzer Überblick über die vorhandenen Daten in ihrer Rohform

```
##   X7.1 X7.2 X7.3 X7.4 X7.5 X7.6 X7.7 X9.1 X9.2 X9.3 X9.4 X9.5 X9.6 X9.7
## 1    5    5    5    1    5    3    2    1    1    1    0    0    0    0
## 2    3    1    1    1    1    3    1    0    1    1    0    0    1    0
## 3    1    7    7    1    1    1    1    1    0    0    0    0    1    0
## 4    5    5    5    1    1    5    2    0    1    1    0    1    0    0
## 5    6    7    7    3    2    4    5    1    1    0    0    0    0    0
## 6    5    1    5    4    6    6    4    1    1    1    0    0    1    0
##   X9.8 D1 D2 X10.1
## 1    0 27  1   300
## 2    0 18  2   150
## 3    0 25  1    50
## 4    0 36  2    50
## 5    0 30  1   100
## 6    0 51  1  1000

##  min  Q1 median  Q3  max   mean      sd   n missing
##    1 200    400 600 3500 486.69 427.1325 400      0
```

## Richtige Datentypen und Bezeichnungen

Kurzer Überblick über die Daten mit richtigen Spaltenbeschriftungen, um sie interpretieren zu können

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()` :
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

	Beratung	Angebote	Bequemlichkeit	Einkaufsatmosphaere	Marken
## 1	5	5	5	1	5
## 2	3	1	1	1	1
## 3	1	7	7	1	1
## 4	5	5	5	1	1
## 5	6	7	7	3	2
## 6	5	1	5	4	6

	Guete der Produkte und Marke	Naehe Partner	Eltern	Verwandte	Kommilitonen	
## 1	3	2	1	1	1	0
## 2	3	1	0	1	1	0
## 3	1	1	1	0	0	0
## 4	5	2	0	1	1	0
## 5	4	5	1	1	0	0
## 6	6	4	1	1	1	0

	Kinder	Freunde	Arbeitskollegen	Soziale Institutionen	Alter	Geschlecht
## 1	0	0	0	0	27	1
## 2	0	1	0	0	18	2
## 3	0	1	0	0	25	1
## 4	1	0	0	0	36	2
## 5	0	0	0	0	30	1
## 6	0	1	0	0	51	1

	EKWert
## 1	300

```
## 2    150
## 3     50
## 4     50
## 5    100
## 6   1000
```

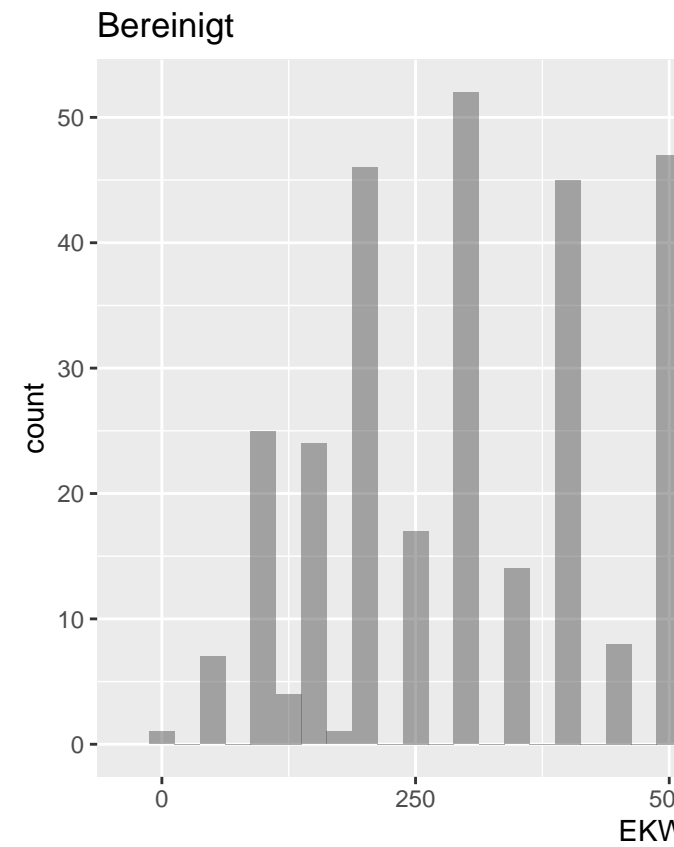
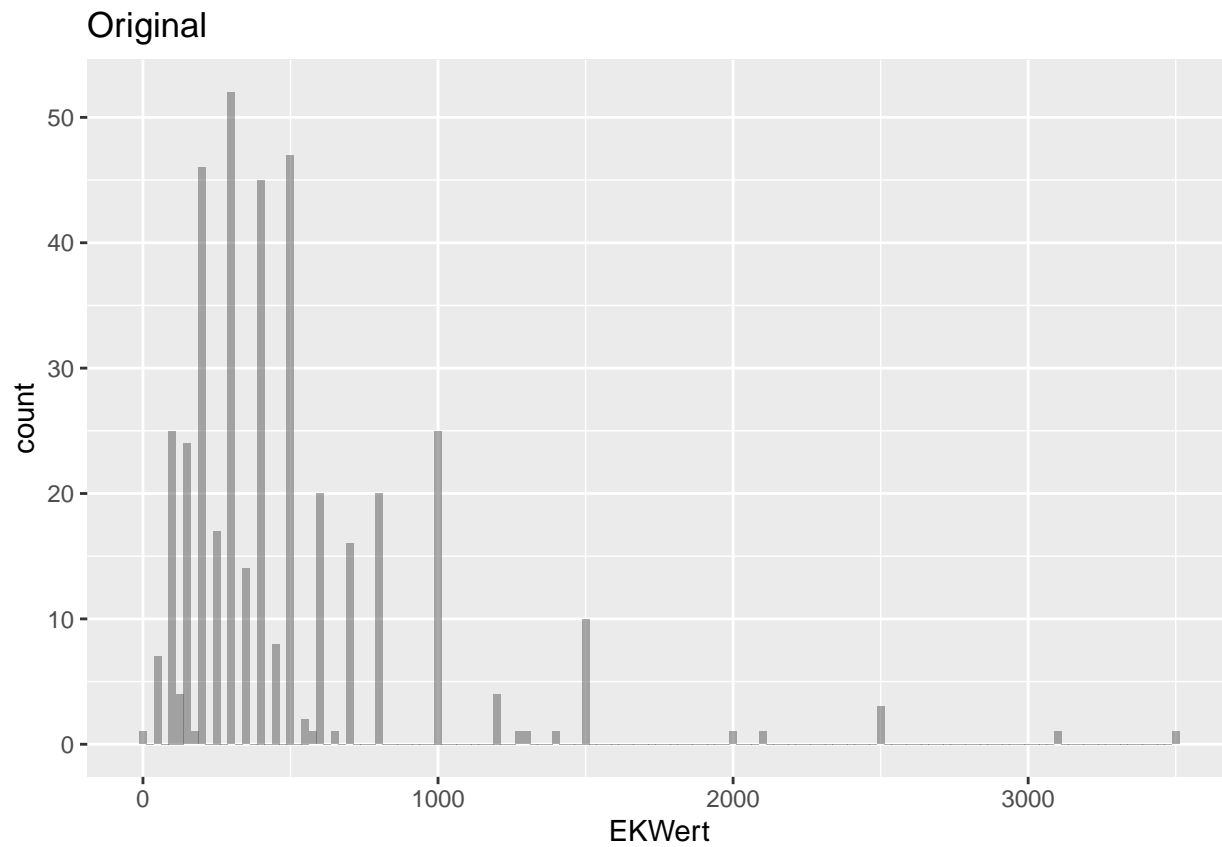
## Ausreißer bereinigen

Ausreißer identifizieren und bereinigtes Dataset in “cl\_test” speichern.

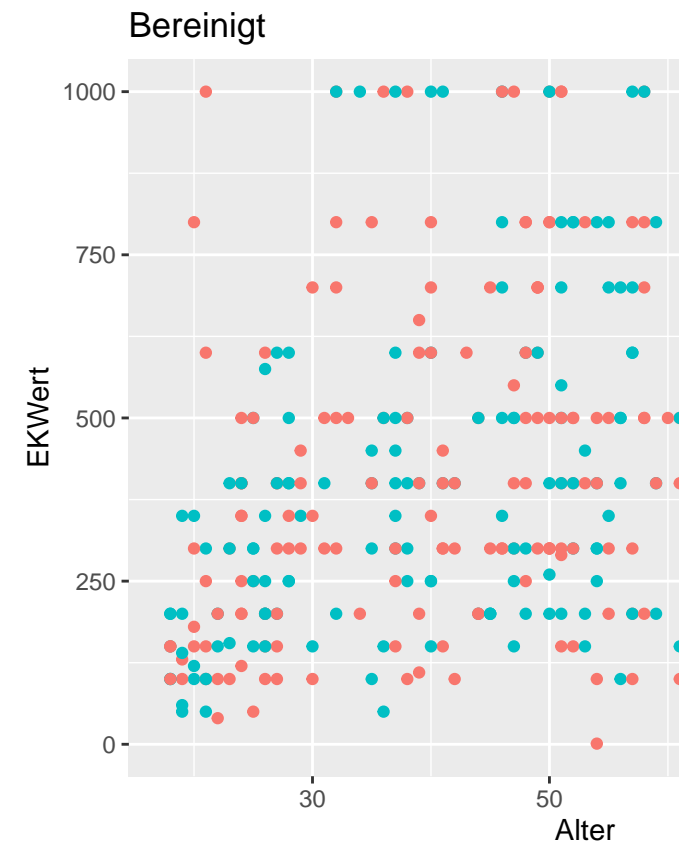
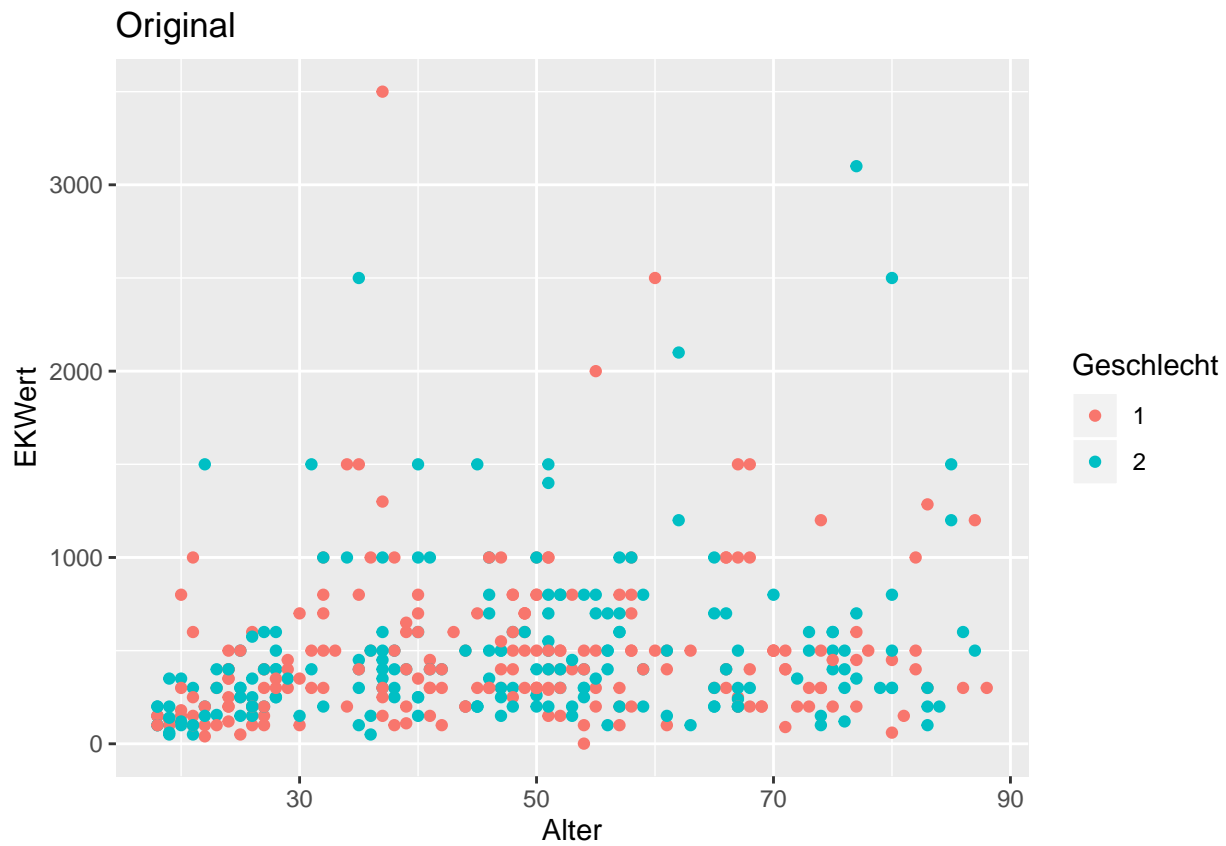
```
## Beratung Angebote Bequemlichkeit Einkaufsatmosphaere Marken
## 1      5         5              5              1      5
## 2      3         1              1              1      1
## 3      1         7              7              1      1
## 4      5         5              5              1      1
## 5      6         7              7              3      2
## 6      5         1              5              4      6
## Guete der Produkte und Marke Naeh Partner Eltern Verwandte Kommilitonen
## 1              3      2      1      1      1      0
## 2              3      1      0      1      1      0
## 3              1      1      1      0      0      0
## 4              5      2      0      1      1      0
## 5              4      5      1      1      0      0
## 6              6      4      1      1      1      0
## Kinder Freunde Arbeitskollegen Soziale Institutionen Alter Geschlecht
## 1      0      0              0              0      27      1
## 2      0      1              0              0      18      2
## 3      0      1              0              0      25      1
## 4      1      0              0              0      36      2
## 5      0      0              0              0      30      1
## 6      0      1              0              0      51      1
## EKWert
## 1    300
## 2    150
## 3     50
## 4     50
## 5    100
## 6   1000
```

## Data-set anschauen

### Histogramm

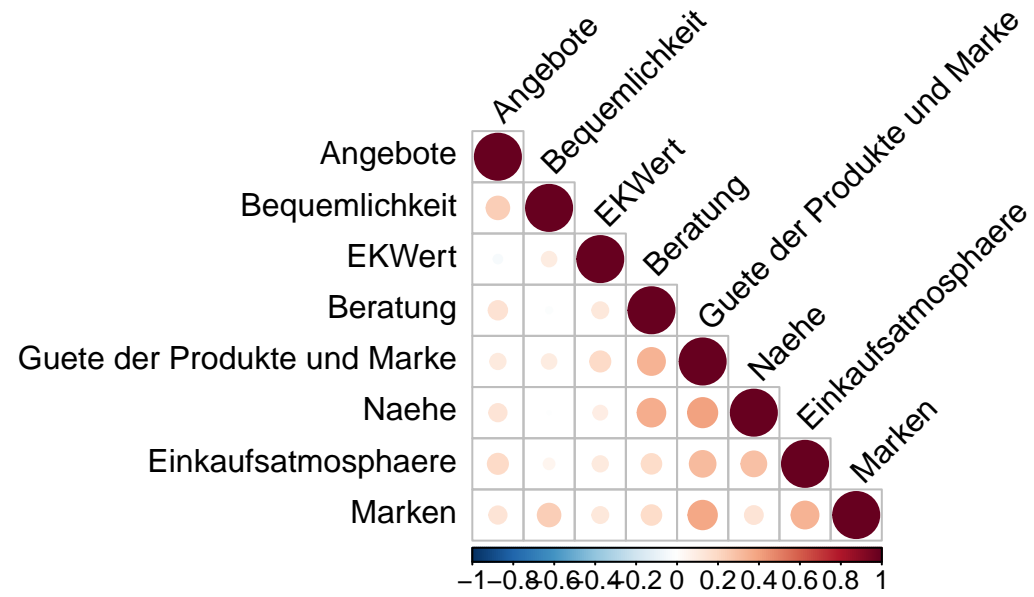


## Punktwolken



## Analysen anstellen

Korrelationsmatrix erstellen, um Abhängigkeiten zu identifizieren.



Hier ist zu erkennen, dass der EKWert kaum Abhängigkeiten aufweist.