

# Einstieg in R Markdown

*Benedict Witzenberger, SZ*

*29. Oktober 2019*

Markdown ist eine einfache Auszeichnungssprache - mit der wir schnell und einfach Dokumente erstellen können, die wir in verschiedensten Vormaten ausspielen können.

RMarkdown ist die R-Variante davon. Sie kann R-Code ausführen und die Ergebnisse in das Dokument einbinden.

Das tolle: Wir können unsere Analyse durchführen - und die Ergebnisse hinterher ganz einfach teilen:

- als HTML-Seite
- als PDF
- als Word-Datei
- als Präsentation
- als Buch
- ...

## Ein Beispiel

Wir laden die Daten des Oktoberfest-Datensatzes in R:

```
df_oktoberfest <- read_csv("https://www.opengov-muenchen.de/dataset/8d6c8251-7956-4f92-8c96-f79106aab822")
```

```
## Parsed with column specification:
## cols(
##   jahr = col_double(),
##   dauer = col_double(),
##   besucher_gesamt = col_double(),
##   besucher_tag = col_double(),
##   bier_preis = col_double(),
##   bier_konsum = col_double(),
##   hendl_preis = col_double(),
##   hendl_konsum = col_double()
## )
```

Der Datensatz enthält 34 Jahre. Von 1985 bis 2018.

Diese Variablen enthält der Datensatz:

```
df_oktoberfest %>%
  str()
```

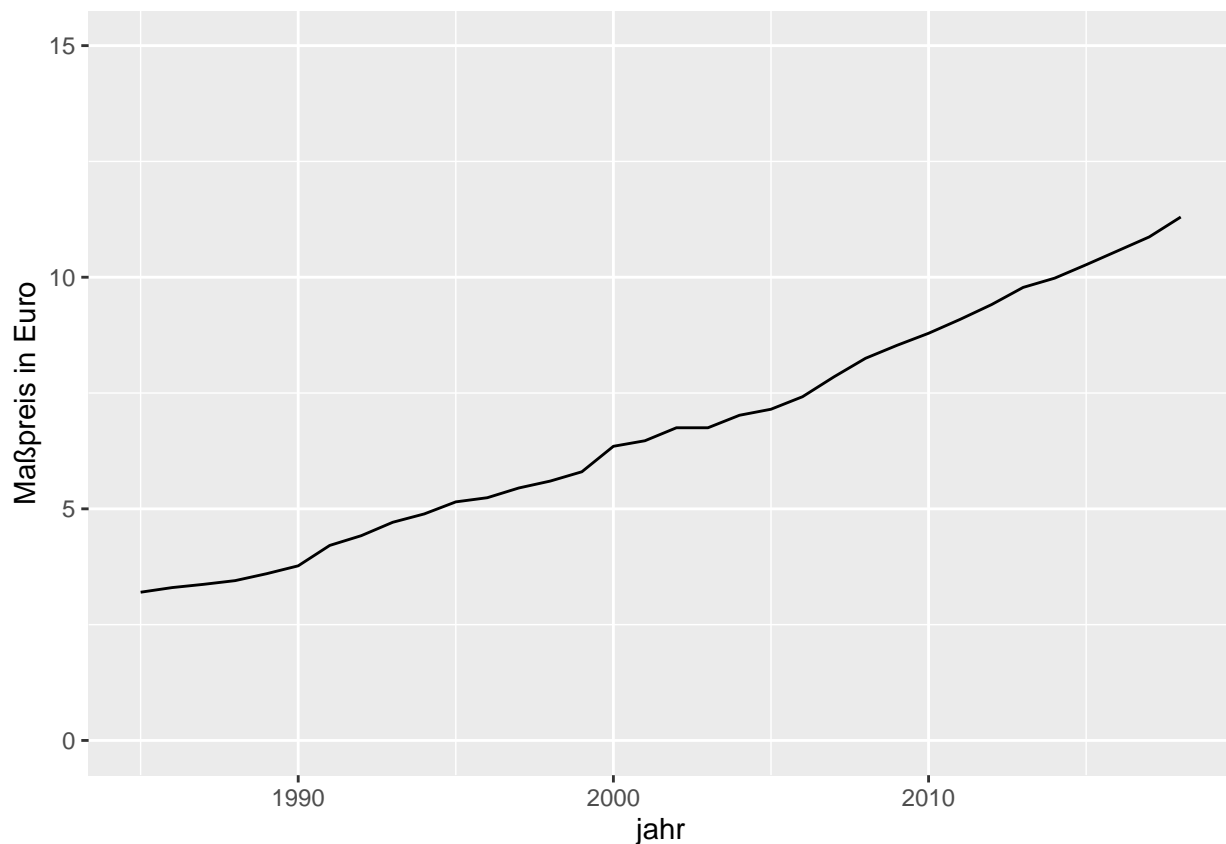
```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 34 obs. of 8 variables:
## $ jahr          : num  1985 1986 1987 1988 1989 ...
## $ dauer          : num  16 16 16 16 16 16 16 16 16 16 ...
## $ besucher_gesamt: num  7.1 6.7 6.5 5.7 6.2 6.7 6.4 5.9 6.5 6.6 ...
## $ besucher_tag   : num  444 419 406 356 388 419 400 369 406 413 ...
## $ bier_preis      : num  3.2 3.3 3.37 3.45 3.6 3.77 4.21 4.42 4.71 4.89 ...
## $ bier_konsum     : num  54541 53807 51842 50951 51241 ...
## $ hendl_preis     : num  4.77 3.92 3.98 4.19 4.22 4.47 4.81 5.11 5.25 5.39 ...
## $ hendl_konsum    : num  629520 698137 732859 720139 775674 ...
## - attr(*, "spec")=
```

```
## .. cols(
## ..   jahr = col_double(),
## ..   dauer = col_double(),
## ..   besucher_gesamt = col_double(),
## ..   besucher_tag = col_double(),
## ..   bier_preis = col_double(),
## ..   bier_konsum = col_double(),
## ..   hendl_preis = col_double(),
## ..   hendl_konsum = col_double()
## .. )
```

Im Schnitt dauerte das Fest 16 Tage.

So hat sich der Bierpreis verändert:

```
ggplot(df_oktoberfest, aes(jahr, bier_preis)) +
  geom_line() +
  scale_y_continuous(limits = c(0, 15)) +
  ylab("Maßpreis in Euro")
```



Was waren die besuchterstärksten Jahre?

- Im Jahr **1985** kamen insgesamt 7.1 Mio. Menschen in 16 Tagen auf das Fest. Im Schnitt waren das etwa 444.000 pro Tag.
- Im Jahr **1996** kamen insgesamt 6.9 Mio. Menschen in 16 Tagen auf das Fest. Im Schnitt waren das etwa 431.000 pro Tag.
- Im Jahr **2000** kamen insgesamt 6.9 Mio. Menschen in 18 Tagen auf das Fest. Im Schnitt waren das etwa 383.000 pro Tag.
- Im Jahr **2011** kamen insgesamt 6.9 Mio. Menschen in 17 Tagen auf das Fest. Im Schnitt waren das

etwa 406.000 pro Tag.