

Webscraping

Dominik Peter

22 12 2016

Webscraping

Laden der Libraries

```
library(data.table)
library(magrittr)
library(stringr)
library(ggplot2)
library(rvest, quietly = TRUE)
```

Extrahieren der Tabelle aus Wikipedia

```
url <- 'https://de.wikipedia.org/wiki/Bern'

df <- url %>%
  read_html(.) %>%
  html_table(fill = TRUE, header = TRUE) %>%
  .[[6]] %>%
  na.omit(.) %>%
  as.data.table(.) %>%
  .[, 1:13, with = FALSE]

setnames(df, 1, "typ")
```

Transformation und Bereinigung der Daten

```
# convert value to numeric and eliminate NA columns
to_numeric <- function(x) {
  x <- str_replace_all(x, ",", "\\.")
  as.numeric(x)
}

months <- colnames(df)[-1]
# apply function
df[, months] <- lapply(df[, months, with = FALSE], to_numeric)

tidy_df <- df[1:3] %>%
  melt(id.vars = "typ", variable.name = "Monat") %>%
  dcast(Monat ~ typ)

col_names <- colnames(tidy_df)
setnames(tidy_df, col_names, c(col_names[1], c("Max", "Min", "Mittelwert")))

pander(tidy_df, style = "rmarkdown")
```

Monat	Max	Min	Mittelwert
Jan	2.8	-3.6	-0.4
Feb	4.7	-3.1	0.7
Mär	9.5	0.2	4.7
Apr	13.4	3	8.1
Mai	18.2	7.4	12.7
Jun	21.6	10.5	16
Jul	24.3	12.5	18.3
Aug	23.7	12.3	17.7
Sep	19.1	8.9	13.7
Okt	13.8	5.4	9.3
Nov	7.3	0.4	3.7
Dez	3.5	-2.3	0.6

Grafische Darstellung

```
mean_temp <- mean(tidy_df$Mittelwert)
min <- scales::comma(tidy_df$Min)
max <- scales::comma(tidy_df$Max)

tidy_df %>%
  ggplot(aes(x = Monat, y = Mittelwert)) +
  geom_point(size = 1.5, color = "#444B54") +
  geom_hline(yintercept = mean_temp, color = "#2980B9", size = 1.5, alpha = 0.2, linetype = 2) +
  geom_errorbar(aes(ymin = Min, ymax = Max), width = .5, color = "#444B54", size = 5) +
  ylab("\nTemperatur (°C)") +
  xlab("\nMonat") +
  geom_text(aes(y=Min), label = min, vjust = 1.9) +
  geom_text(aes(y=Max), label = max, vjust = -1.1) +
  ylim(-8, 30) +
  ggtitle("Monatliche Durchschnittstemperaturen", subtitle = "für Bern 1981 - 2010") +
  theme(panel.background = element_blank(),
        # panel.grid.major = element_line(color = "#F0F1F5", size = 0.2),
        panel.grid.minor = element_blank()
  )
```

FALSE Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x\$label)): Fontmetrik
FALSE für das Zeichen 0x1e unbekannt

FALSE Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x\$label)): Fontmetrik
FALSE für das Zeichen 0x80 unbekannt

FALSE Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x\$label)):
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <e2>

FALSE Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x\$label)):
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <80>

FALSE Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x\$label)):
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <93>

```

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <e2>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <80>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <93>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <e2>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <80>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <93>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <e2>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <80>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <93>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <e2>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <80>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <93>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :

```

```

FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <e2>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <80>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <93>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <e2>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <80>

FALSE Warning in grid.Call(L_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
FALSE Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs': Punkt
FALSE ersetzt <93>

FALSE Warning in grid.Call.graphics(L_text, as.graphicsAnnot(x$label), x$x, x
FALSE $y, : Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs':
FALSE Punkt ersetzt <e2>

FALSE Warning in grid.Call.graphics(L_text, as.graphicsAnnot(x$label), x$x, x
FALSE $y, : Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs':
FALSE Punkt ersetzt <80>

FALSE Warning in grid.Call.graphics(L_text, as.graphicsAnnot(x$label), x$x, x
FALSE $y, : Konvertierungsfehler für 'für Bern 1981 - 2010' in 'mbcsToSbcs':
FALSE Punkt ersetzt <93>

```

Monatliche Durchschnittstemperaturen für Bern 1981 ... 2010

