

rstatsZH - Data Science mit R

Vektoren in R und mehr Git

Lars Schöbitz

2021-09-28

Rückblick - Woche 4

dplyr Funktionen

- `select()` Hilfe Funktionen
 - `starts_with()`
 - `ends_with()`
 - `contains()`
- `relocate()`
- `rename()`
- `mutate()`
- `group_by()`
- `summarise()`
- `count()`

Hausaufgabe 4 - Übung 1

Was glaubst du wird mit der Zeile:

```
locale = locale(encoding = "ISO-8859-1")
```

erreicht?

```
# Der Link zu den Daten wird hier als Objekt gespeichert und taucht oben rechts i
link <- "https://www.web.statistik.zh.ch/ogd/data/bista/ZH_Uebersicht_alle_Lernen

# Hier wird nun das Objekt "link" genutzt um die CSV zu lesen
lernende <- read_csv(file = link,
                     locale = locale(encoding = "ISO-8859-1"))
```

Hausaufgabe 4 - Übung 1

Wenn Daten im UTF-8 Standard kodiert (gespeichert) wurden wir das `locale =` Argument nicht benötigt, da die Funktion `read_csv` standardmässig UTF-8 für die Dekodierung nutzt.

```
# Der Link zu den Daten wird hier als Objekt gespeichert und taucht oben rechts i
link <- "https://www.web.statistik.zh.ch/ogd/data/bista/ZH_Uebersicht_alle_Lernen

# Hier wird nun das Objekt "link" genutzt um die CSV zu lesen
lernende <- read_csv(file = link)
```

Tipp: UTF-8 als Standard festlegen um Daten aus R zu speichern:
<https://github.com/rstatsZH/kochbuch/tree/main/06-file-encoding>

Hausaufgabe 4 - Übung 3

Plot

Code

Hausaufgabe 4 - Übung 5

Was gibt die Spalte `n` an?

```
# Häufigkeitstabelle 1 - Zusammenfassen  
lernende2019 %>%  
  count(Schultyp)
```

```
# A tibble: 6 × 2  
  Schultyp      n  
  <chr>    <int>  
1 Berufsfachschule      8  
2 Brückenangebot      8  
3 Heim- und Sonderschule 22  
4 Höhere Berufsbildung   8  
5 Mittelschule         16  
6 Volksschule          24
```

```
# Häufigkeitstabelle 2 - Zusammenfassen  
lernende2019 %>%  
  group_by(Schultyp) %>%  
  summarise(  
    n = n()  
  )
```

```
# A tibble: 6 × 2  
  Schultyp      n  
  <chr>    <int>  
1 Berufsfachschule      8  
2 Brückenangebot      8  
3 Heim- und Sonderschule 22  
4 Höhere Berufsbildung   8  
5 Mittelschule         16  
6 Volksschule          24
```

Hausaufgabe 4 - Übung 5

Wie kommen wir auf die Anzahl Schüler_innen.

```
# Häufigkeitstabelle 1 - Zusammenfassen  
lernende2019 %>%  
  count(Schultyp, wt = Anzahl)
```

```
# A tibble: 6 × 2  
  Schultyp      n  
  <chr>      <dbl>  
1 Berufsfachschule 45508  
2 Brückenangebot  2073  
3 Heim- und Sonderschule 3021  
4 Höhere Berufsbildung 19757  
5 Mittelschule 20591  
6 Volksschule 161073
```

```
# Häufigkeitstabelle 2 - Zusammenfassen  
lernende2019 %>%  
  group_by(Schultyp) %>%  
  summarise(  
    n = sum(Anzahl)  
  )
```

```
# A tibble: 6 × 2  
  Schultyp      n  
  <chr>      <dbl>  
1 Berufsfachschule 45508  
2 Brückenangebot  2073  
3 Heim- und Sonderschule 3021  
4 Höhere Berufsbildung 19757  
5 Mittelschule 20591  
6 Volksschule 161073
```

Hausaufgabe 4 - Übung 6

```
lernende %>%
  drop_na %>%
  group_by(Jahr) %>%
  summarise(
    Total = sum(Anzahl)
  ) %>%
  mutate(
    Differenz = Total - lag(Total)
  ) %>%
  mutate(
    Veraenderung = case_when(
      Differenz > 0 ~ "Zunahme",
      Differenz < 0 ~ "Abnahme"
    )
  )
```

```
# A tibble: 21 × 4
  Jahr   Total Differenz Veraenderung
  <dbl>   <dbl>   <dbl>   <chr>
1  2000  217135      NA   <NA>
2  2001  220048    2913  Zunahme
3  2002  226706    6658  Zunahme
4  2003  229676    2970  Zunahme
5  2004  228468   -1208  Abnahme
6  2005  230183    1715  Zunahme
# ... with 15 more rows
```

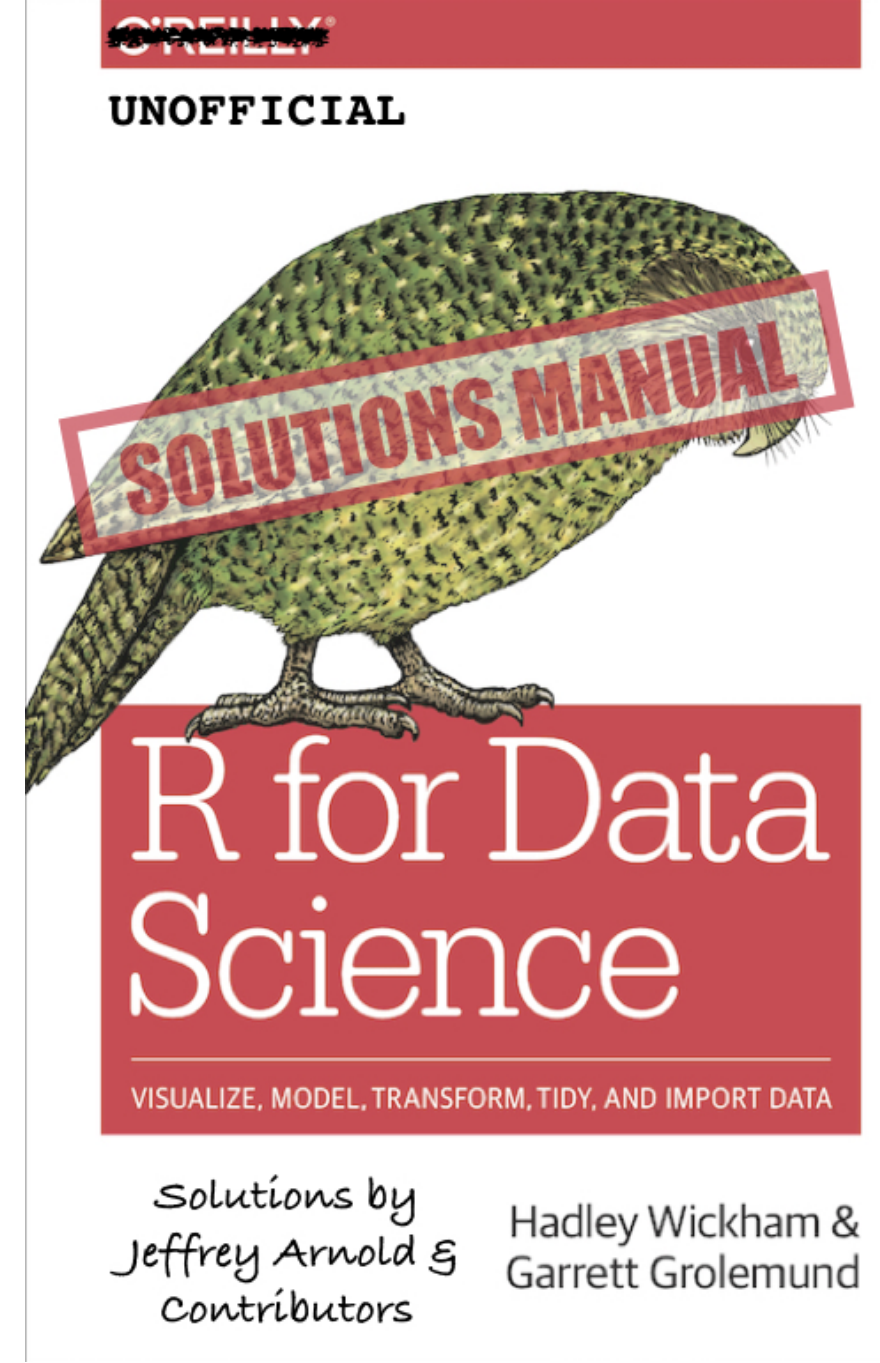

Hausaufgabe 4 - Lösungen

- **GitHub Organisation:** rstatsZH
 - <https://github.com/rstatsZH/>
- **Repo:** ha-04-mehr-dplyr
 - <https://github.com/rstatsZH/ha-04-mehr-dplyr>
- **R Markdown Datei:** ha-04-solutions.Rmd
 - <https://github.com/rstatsZH/ha-04-mehr-dplyr/blob/main/ha-04-solutions.Rmd>

R for Data Science - Lösungsbuch

Lösungen zu den Übungen aus dem Buch mit nützlichen Erklärungen gibt es hier:

<https://jrnold.github.io/r4ds-exercise-solutions/>



Praktikum 7 - dplyr

Live Code

1. **RStudio Cloud**: Öffne deinen Arbeitsbereich für den Kurs in der RStudio Cloud
2. **RStudio Cloud / Projects**: Öffne erneut das Praktikum 07
3. Folgt wieder auf dem Bildschirm

Ziele für diese Woche

Am Ende dieser Woche könnt ihr:

- Einer Anleitung folgen um kollaborativ mit Git und GitHub an einem Projekt zu arbeiten
- Die Begriffe `git add`, `git commit`, `git push`, und `git pull` unterscheiden
- Den Unterschied zwischen einem Vektor vom Typ "character" und "factor" erläutern

Demonstration 1 - Ehedaten

1. Schaut mir bei dem Erkunden eines Datensatz zu
2. Schreibt selbst **keinen** Code
3. Macht euch Notizen und stellt Fragen

Pause

10 : 00

Photo by: Blake Wisz



Praktikum 8 - Git / GitHub

Ihr könnt bereits:

- `git add` (Das Häkchen bei Staged setzen)
- `git commit`
- `git push`

Neu lernt ihr:

- `git pull`

Praktikum 8 - Zwei Rollen

Person 1 - Push

- Öffnet die R Markdown Datei für die Übung
- Teilt den Bildschirm
- Schreibt den Code
- Führt die Git Befehle "add", "commit" und "push" aus

Person 2 - Pull

- Öffnet die R Markdown Datei für die Übung
- Schreibt selbst **keinen** Code
- Unterstützt Person 1 bei den Übungen
- Führt den Git Befehl "pull" aus

Praktikum 8 - Git / GitHub - Ehedaten

2er Teams

Lest die Anleitungen aufmerksam und entscheidet welche Person zu Beginn den Bildschirm teilt

1. **E-Mail:** Öffne deine Email und klicke auf den Link zu deinem GitHub **Team** repo für **prak-08-team0X**
2. **GitHub:** Klicke auf den grünen Button "Code" und kopiere den Link für das Repo in deine Zwischenablage
3. **RStudio Cloud:** Öffne deinen Arbeitsbereich für den Kurs in der RStudio Cloud
4. **RStudio Cloud / Projects:** Klicke auf "New Project from GitHub Repository"

Pause

10 : 00

Photo by: Blake Wisz



Demonstration 2 - Vektoren

1. Schaut mir nochmals beim Programmieren zu
2. Macht euch Notizen und stellt Fragen

Hausaufgabe

Hausaufgabe 5 - Erstes Projekt mit Git und GitHub

- Keine Vorlage von meiner Seite
- Alle Schritte sind auf der Webseite beschrieben:

<https://rstatszh.github.io/website/ha-05-erstes-projekt.html>

Feedback

Ziele erreicht?

Bitte ausfüllen: kutt.it/rstatszh-eval

Photo by: Virgil Cayasa





Für die Aufmerksamkeit!

Für die R packages `{xaringan}` und `{xaringanthemer}` mit welchen die Folien geschrieben wurden.

Eine PDF Version der Folien kann hier heruntergeladen werden:

https://github.com/rstatsZH/website/raw/master/slides/e1_d05-vektoren/e1_d05-vektoren.pdf

Für [Data Science in a Box](#) und [Remaster the Tidyverse](#), von welchen ich Materialien für diesen Kurs nutze und welche genau wie diese Folien mit [Creative Commons Attribution Share Alike 4.0 International](#) lizenziert sind.