

Wiederholung

Überprüfung der erworbenen Fähigkeiten zur Halbzeit des Semesters

Daniela Palleschi

Inhaltsverzeichnis

1	Wiederholung	1
2	Pakete	2
3	Einlesung	2
4	Wrangling	3
5	Coding style	4

1 Wiederholung

Bis jetzt haben wir gelernt:

- wie man einfache mathematische Berechnungen in R durchführt
- wie man R- und Quarto-Skripte erstellt und strukturiert
- wie man Pakete installiert und lädt (z.B. `library()` oder `pacman::p_load()`)
- wie man einen Datensatz inspiziert (z. B. `summary()`)
- wie man einen Datensatz mit Hilfe des `readr`-Pakets lädt
- wie man Daten mit Hilfe des Pakets `ggplot2` visualisiert
- wie man Daten mit Hilfe des `dplyr`-Pakets verarbeitet

-
- von nun an bauen wir auf diesen Themen auf
 - heute werden Sie all diese Fähigkeiten anwenden, indem Sie einen Bericht erstellen

2 Pakete

- installieren: `install.packages("package")`
- laden: `library()`
- *oder*: laden (und installieren, falls erforderlich): `pacman::p_load(package1, package2)`

```
install.packages("tidyverse")
install.packages("here")
library(tidyverse)
library(here)
```

```
pacman::p_load(
  tidyverse,
  here
)
```

3 Einlesung

Um Daten einzulesen, müssen wir:

- das `read_csv()` Paket von `readr` (ein `tidyverse` Paket) benutzen
- `here::here()` benutzen, um R zu sagen, dass es in unserem Projektordner suchen soll
- den Ordner auflisten, in dem sich unsere Daten befinden: `Daten`
- Geben Sie den *exakten Namen* der Datei an, die in `daten` gespeichert ist: z.B. `flights.csv`

```
1 df_flights <-
2   read_csv(
3     here::here(
4       "daten",
5       "flights.csv"
6     )
7   )
```

Rows: 336776 Columns: 19

-- Column specification -----

Delimiter: ","

chr (4): carrier, tailnum, origin, dest

dbl (14): year, month, day, dep_time, sched_dep_time, dep_delay, arr_time, ...

```
dtm (1): time_hour
```

i Use ``spec()`` to retrieve the full column specification for this data.

i Specify the column types or set ``show_col_types = FALSE`` to quiet this message.

4 Wrangling

Warum liefert die folgende Tabelle keine Beobachtungen?

```
df_flights %>%  
  filter(arr_delay > 120,  
         dep_delay < 0,  
         dest=="HOU" |  
         dest=="IAH",  
         carrier=="UA" |  
         carrier=="AA" |  
         carrier=="DL",  
         month== 7 | 8 | 9  
  )
```

```
df_flights %>%  
  filter(arr_delay > 120,  
         dep_delay < 120,  
         dest=="HOU" |  
         dest=="IAH",  
         carrier=="UA" |  
         carrier=="AA" |  
         carrier=="DL",  
         month== 7 | month== 8 | month== 9  
  )
```

A tibble: 8 x 19

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2013	7	10	1629	1520	69	2048	1754
2	2013	7	10	1913	1721	112	2214	2001
3	2013	7	22	2254	2114	100	219	2358

```

4 2013      7    23      824          638      106      1124          911
5 2013      7    29     1229          1030      119      1527         1325
6 2013      8     9     1850          1730       80      2221         2007
7 2013      8     9     2251          2102      109       145         2344
8 2013      8    28     1219          1030      109      1634         1331
# i 11 more variables: arr_delay <dbl>, carrier <chr>, flight <dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dtm>

```

5 Coding style

- dieser Code ist etwas schwer zu lesen.

```
df_flights %>% filter(arr_delay > 120, dep_delay == 0, dest == "IAH" | dest == "HOU", carrier == "UA" | carrier == "AA" | carrier == "DL", month == 07 | month == 08 | month == 09)
```

- dieser Code ist leichter zu lesen, obwohl er mehr Platz einnimmt. Im Allgemeinen sollte der Code so strukturiert sein.

```

df_flights %>%
  filter(
    arr_delay > 120,
    dep_delay == 0,
    dest == "IAH" | dest == "HOU",
    carrier == "UA" |
      carrier == "AA" |
      carrier == "DL",
    month == 07 |
      month == 08 |
      month == 09
  )

```

Eine einfache Möglichkeit, “unordentlichen” Code zu übernehmen, besteht darin, ihn zu markieren und die Tastenkombination **Cmd/Strg+Shift+A** zu verwenden.