

Daten transformieren mit `{dplyr}`

Unit 2

Ziele für heute

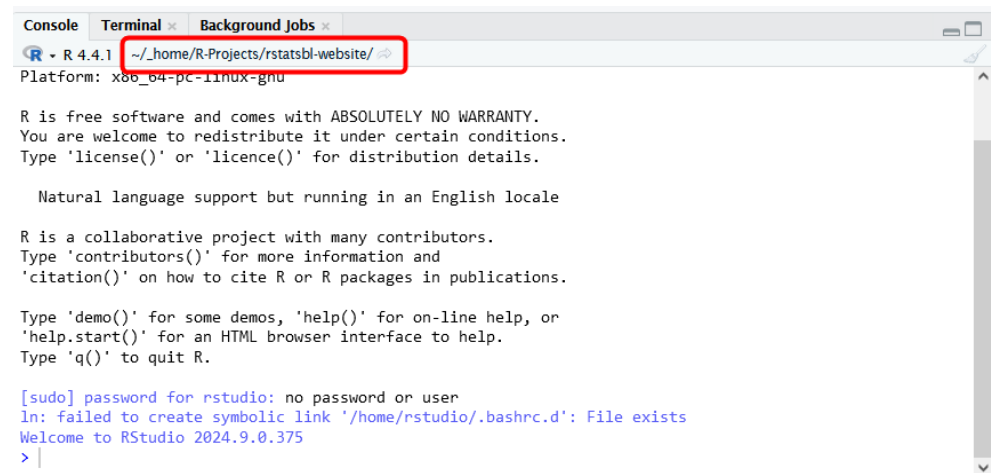
1. ein Projekt in RStudio beschreiben
2. Regeln der `{dplyr}`-Befehle auflisten
3. Befehle zur Transformation von Zeilen identifizieren und anwenden

Workflow: R-Projects

Working Directory

`setwd()` ↔ `getwd()`

RStudio Projekte 



```
Console Terminal Background Jobs
~/_home/R-Projects/rstatsbl-website/
Platform: x86_64-pc-linux-gnu

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

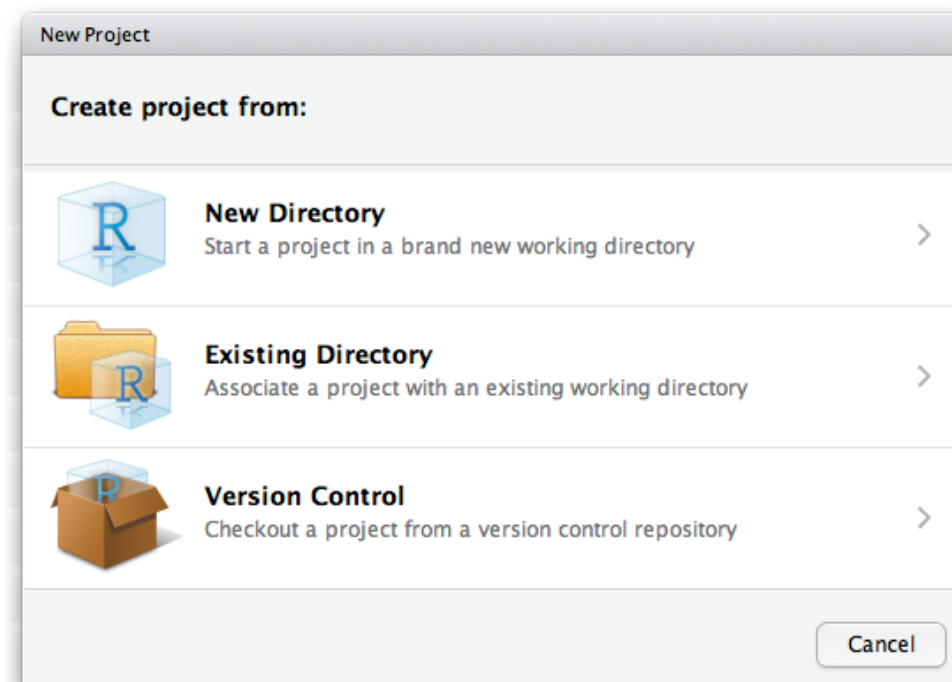
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[sudo] password for rstudio: no password or user
ln: failed to create symbolic link '/home/rstudio/.bashrc.d': File exists
Welcome to RStudio 2024.9.0.375
> |
```

RStudio-Projekte

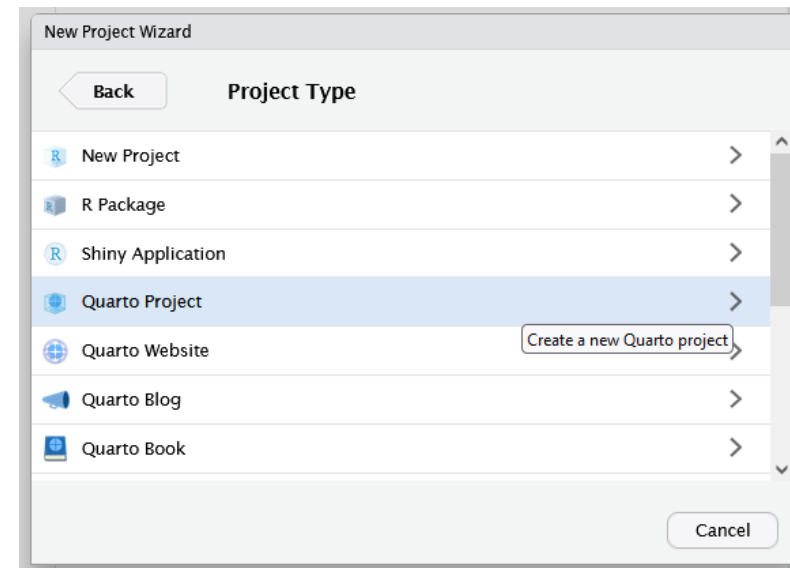
Einer der stärksten Aspekte von RStudio:

- Relative Pfade (`setwd()` nicht erforderlich)
- Dateien und Ordner zusammen, die zusammengehören



Praktikum: RStudio-Projekte

- Erstellt ein neues Quarto-Projekt namens `test`



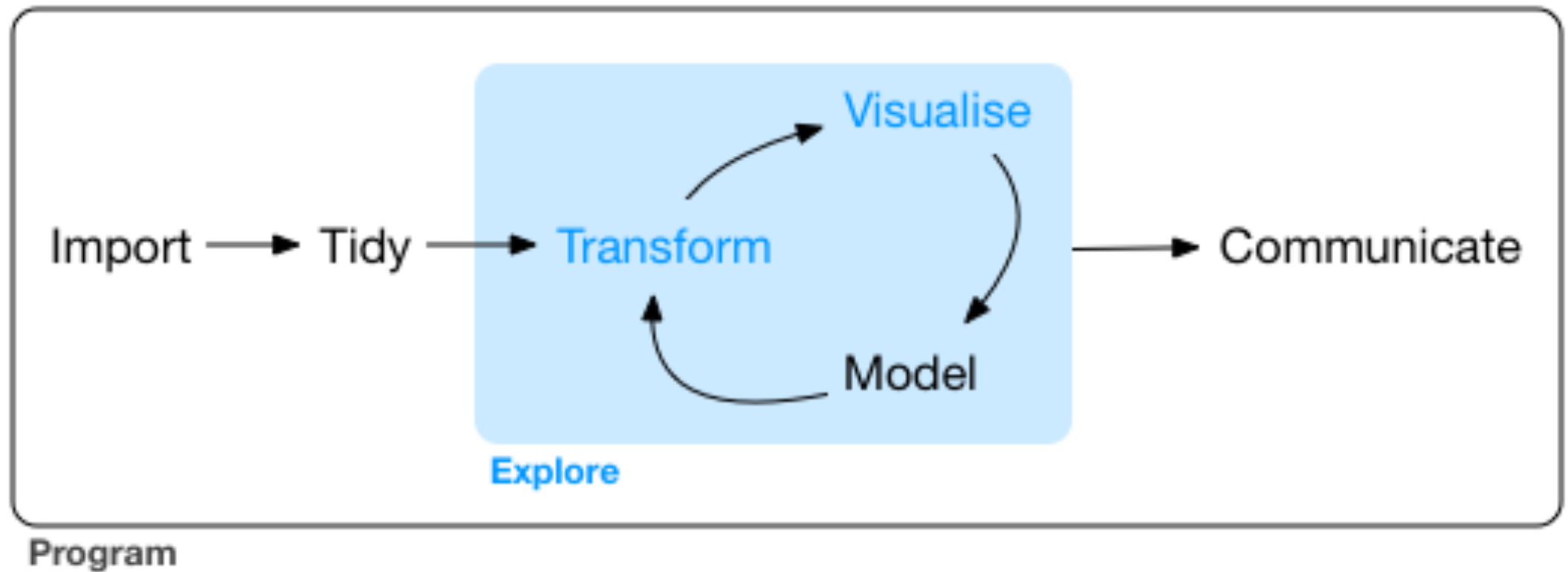
- Erstellt ein RStudio-Project mit einem existierenden Ordner

20:00

rstatsBL - Data Science mit R

Break   

10:00



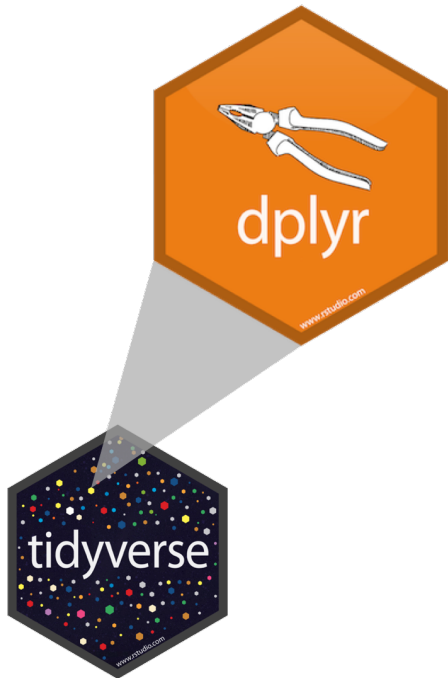
dplyr : go wrangling



Illustration von @allison_horst

rstatsBL - Data Science mit R

Daten-Transformation mit `{dplyr}`



- **Zeilen:** filtern, anordnen
- **Spalten:** auswählen, anordnen, umbenennen, erstellen
- **Gruppen:** zusammenfassen, zählen
- **Tabellen:** zusammenfügen

Regeln der `{dplyr}`-Functions

`BEFEHL(DATEN_ZUM_UMWANDELN, WAS_MACHT_ES)`

1. das erste Argument ist immer ein *data frame*
2. die nachfolgenden Argumente geben an was mit dem *data frame* gemacht werden soll
3. Geben immer ein *data frame* zurück
4. ändern nichts an Ort und Stelle

Zeilen transformieren

`filter()`, `arrange()`, `distinct()`, `slice()`

Daten

← → ↺

https://data.bl.ch/explore/dataset/11990/table/?disjunctive.domain0&disjunctive.name&disjunctive.type&disjunctive.answer&sort=date

🔍 ⭐

📄 ⬇️ 📧

Getting Started R R AFDs BFS R Workshop Quarto ggplot2 Workgroups Train the Trainer

BASEL
LANDSCHAFT

OGD

Daten

Dashboards

API

Dokumentation

Kontakt

🔍 Suchen Ctrl+K

31'734 Dateneinträge

Abstimmungsarchiv nach Vorlage, Gemeinde und Datum (seit 2003)

✕ f in

Keine aktiven Filter

Information Tabelle Exporte API Kommentare (0)

Filter

Suche nach Einträgen ... 🔍

date

2003 2'150

2004 1'548

2005 1'462

2006 860

2007 602

2008 1'548

> Mehr

domain

federation 15'996

canton 15'738

name

Aesch (BL) 369

Allschwil 369

Anwil 369

Arboldswil 369

Arisdorf 369

Arlesheim 369

> Mehr

type

counter-proposal 1'204

proposal 29'326

tie-breaker 1'204

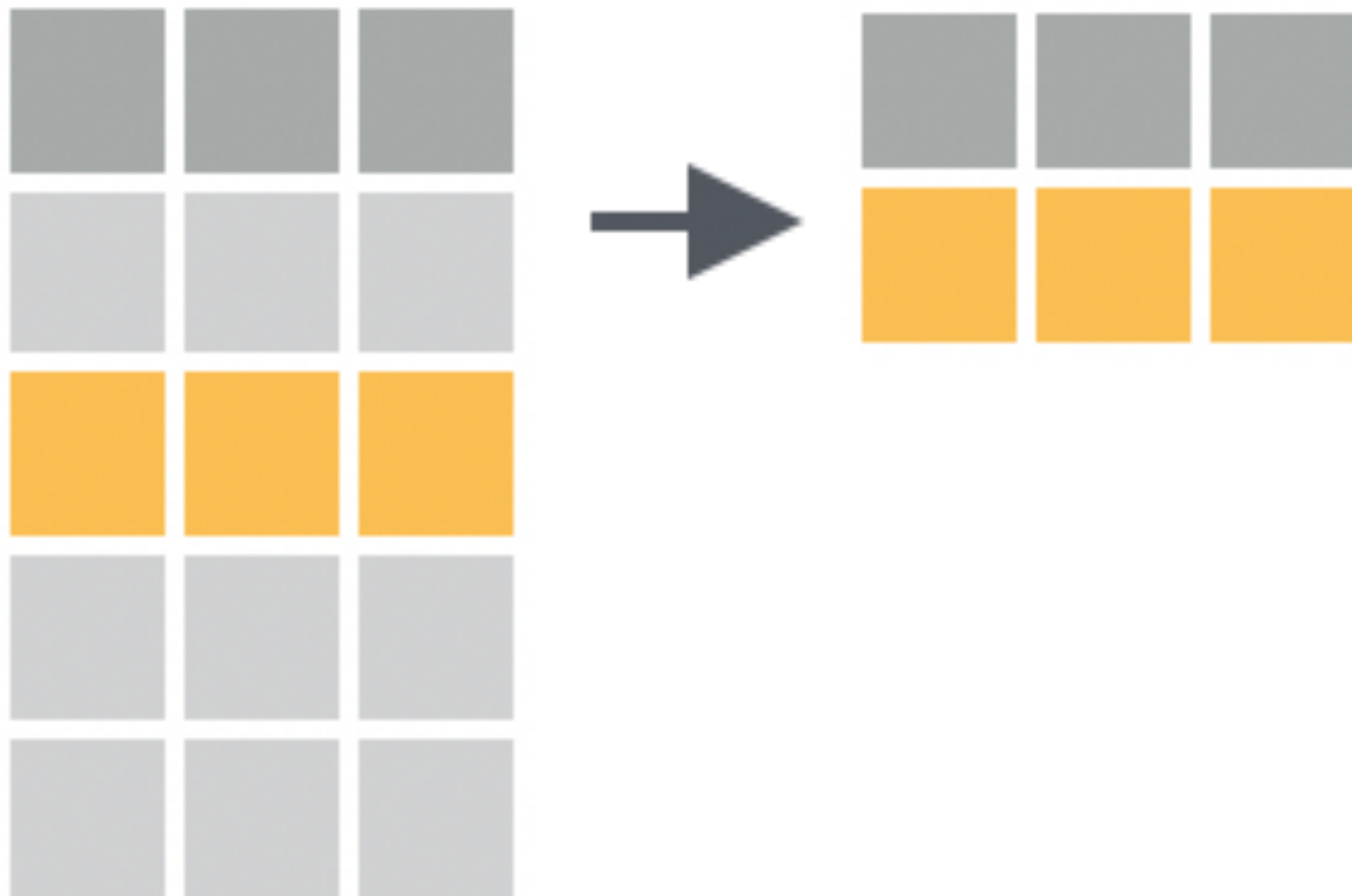
	date	entity_id	name	district	vote_id	domain	type	title_de_CH	counted	answer	percent_years	percent_nays	percent
1	24. November 2024	2761	Aesch (BL)	Arlesheim	20241124_E1	federation	proposal	Bundesbeschluss vom 29. Septe...	True	accepted	57.7 %	42.3 %	39.7 %
2	24. November 2024	2761	Aesch (BL)	Arlesheim	20241124_E2	federation	proposal	Änderung vom 29. September 202...	True	accepted	52.1 %	47.9 %	38.4 %
3	24. November 2024	2761	Aesch (BL)	Arlesheim	20241124_E3	federation	proposal	Änderung vom 29. September 202...	True	rejected	47.6 %	52.4 %	39.0 %
4	24. November 2024	2761	Aesch (BL)	Arlesheim	20241124_E4	federation	proposal	Änderung vom 22.Dezember 2023...	True	accepted	55.8 %	44.2 %	39.6 %
5	24. November 2024	2762	Allschwil	Arlesheim	20241124_E1	federation	proposal	Bundesbeschluss vom 29. Septe...	True	accepted	52.4 %	47.6 %	43.6 %
6	24. November 2024	2762	Allschwil	Arlesheim	20241124_E2	federation	proposal	Änderung vom 29. September 202...	True	rejected	44.3 %	55.7 %	42.9 %
7	24. November 2024	2762	Allschwil	Arlesheim	20241124_E3	federation	proposal	Änderung vom 29. September 202...	True	rejected	40.7 %	59.3 %	43.6 %
8	24. November 2024	2762	Allschwil	Arlesheim	20241124_E4	federation	proposal	Änderung vom 22.Dezember 2023...	True	accepted	52.4 %	47.6 %	43.2 %
9	24. November 2024	2763	Arlesheim	Arlesheim	20241124_E1	federation	proposal	Bundesbeschluss vom 29. Septe...	True	rejected	48.3 %	51.7 %	50.5 %
10	24. November 2024	2763	Arlesheim	Arlesheim	20241124_E2	federation	proposal	Änderung vom 29. September 202...	True	rejected	47.8 %	52.2 %	50.3 %
11	24. November 2024	2763	Arlesheim	Arlesheim	20241124_E3	federation	proposal	Änderung vom 29. September 202...	True	rejected	44.8 %	55.2 %	50.2 %
12	24. November 2024	2763	Arlesheim	Arlesheim	20241124_E4	federation	proposal	Änderung vom 22.Dezember 2023...	True	accepted	57.7 %	42.3 %	50.2 %
13	24. November 2024	2764	Biel-Benken	Arlesheim	20241124_E1	federation	proposal	Bundesbeschluss vom 29. Septe...	True	accepted	59.6 %	40.4 %	51.2 %
14	24. November 2024	2764	Biel-Benken	Arlesheim	20241124_E2	federation	proposal	Änderung vom 29. September 202...	True	accepted	63.1 %	36.9 %	50.6 %
15	24. November 2024	2764	Biel-Benken	Arlesheim	20241124_E3	federation	proposal	Änderung vom 29. September 202...	True	accepted	60.1 %	39.9 %	50.7 %
16	24. November 2024	2764	Biel-Benken	Arlesheim	20241124_E4	federation	proposal	Änderung vom 22.Dezember 2023...	True	accepted	67.8 %	32.2 %	50.9 %
17	24. November 2024	2765	Binningen	Arlesheim	20241124_E1	federation	proposal	Bundesbeschluss vom 29. Septe...	True	accepted	52.0 %	48.0 %	51.6 %
18	24. November 2024	2765	Binningen	Arlesheim	20241124_E2	federation	proposal	Änderung vom 29. September 202...	True	accepted	51.0 %	49.0 %	51.5 %
19	24. November 2024	2765	Binningen	Arlesheim	20241124_E3	federation	proposal	Änderung vom 29. September 202...	True	rejected	48.4 %	51.6 %	51.4 %
20	24. November 2024	2765	Binningen	Arlesheim	20241124_E4	federation	proposal	Änderung vom 22.Dezember 2023...	True	accepted	60.5 %	39.5 %	51.5 %
21	24. November 2024	2766	Birsfelden	Arlesheim	20241124_E1	federation	proposal	Bundesbeschluss vom 29. Septe...	True	rejected	35.7 %	64.3 %	44.7 %
22	24. November 2024	2766	Birsfelden	Arlesheim	20241124_E2	federation	proposal	Änderung vom 29. September 202...	True	rejected	34.6 %	65.4 %	44.2 %
23	24. November 2024	2766	Birsfelden	Arlesheim	20241124_E3	federation	proposal	Änderung vom 29. September 202...	True	rejected	27.1 %	72.9 %	44.2 %
24	24. November 2024	2766	Birsfelden	Arlesheim	20241124_E4	federation	proposal	Änderung vom 22.Dezember 2023...	True	rejected	46.4 %	53.6 %	43.8 %
25	24. November 2024	2767	Bottmingen	Arlesheim	20241124_E1	federation	proposal	Bundesbeschluss vom 29. Septe...	True	accepted	58.1 %	41.9 %	51.4 %
26	24. November 2024	2767	Bottmingen	Arlesheim	20241124_E2	federation	proposal	Änderung vom 29. September 202...	True	accepted	60.1 %	39.9 %	49.6 %

Daten

```
1 library(tidyverse)
2 abstimmung <- read_delim("data/ogd_abstimmungsarchiv_11990.csv", delim = ";")
3 glimpse(abstimmung)
```

```
1 Rows: 31,734
2 Columns: 21
3 $ date                <date> 2017-05-21, 2017-05-21, 2017-05-21, 2017-05-21...
4 $ entity_id           <dbl> 2765, 2766, 2766, 2766, 2766, 2766, 2767, 2767,...
5 $ name                <chr> "Binningen", "Birsfelden", "Birsfelden", "Birsf...
6 $ district            <chr> "Arlesheim", "Arlesheim", "Arlesheim", "Arleshe...
7 $ vote_id             <chr> "20170521_K5", "20170521_E1", "20170521_K2", "2...
8 $ domain              <chr> "canton", "federation", "canton", "canton", "ca...
9 $ type                <chr> "proposal", "proposal", "proposal", "proposal",...
10 $ title_de_CH         <chr> "Änderung der Verfassung des Kantons Basel-Land...
11 $ counted             <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,...
12 $ answer              <chr> "accepted", "accepted", "rejected", "rejected",...
13 $ percent_yeas        <dbl> 87.93388, 60.14460, 29.46679, 26.06433, 35.0667...
14 $ percent_nays        <dbl> 12.06612, 39.85540, 70.53321, 73.93567, 64.9332...
15 $ percent_turnout     <dbl> 46.24283, 34.89440, 34.20569, 33.97612, 33.8689...
16 $ eligible_voters     <dbl> 9941, 6534, 6534, 6534, 6534, 6534, 4192, 4192,...
17 $ expats              <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
18 $ emntv               <dbl> 321. 22. 54. 62. 71. 127. 16. 24. 46. 57. 123. ...
```

Zeilen filtern



```
filter(Datensatz, Zeilenbedingung)
```

Zeilen filtern

```
1 filter(  
2   abstimmung,  
3   name == "Oberwil (BL)"  
4 )
```

- {dplyr}-Befehl: `filter()`

Zeilen filtern

```
1 filter(  
2   abstimmung,  
3   name == "Oberwil (BL)"  
4 )
```

- {dplyr}-Befehl: `filter()`
- Erstes Argument: *data frame*

Zeilen filtern

```
1 filter(  
2   abstimmung,  
3   name == "Oberwil (BL)"  
4 )
```

- {dplyr}-Befehl: `filter()`
- Erstes Argument: *data frame*
- Zweites Argument:
Zeilenbedingung

Zeilen filtern

```
1 filter(
2   abstimmung,
3   name == "Oberwil (BL)"
4 )
```

- {dplyr}-Befehl: `filter()`
- Erstes Argument: *data frame*
- Zweites Argument:
Zeilenbedingung
- Resultat: *data frame*

```
1 # A tibble: 369 × 21
2   date      entity_id name    district vo
3   <date>      <dbl> <chr>    <chr>    <cl
4 1 2017-05-21    2771 Oberw... Arleshe... 20
5 2 2017-05-21    2771 Oberw... Arleshe... 20
6 3 2017-05-21    2771 Oberw... Arleshe... 20
7 4 2017-05-21    2771 Oberw... Arleshe... 20
8 5 2017-05-21    2771 Oberw... Arleshe... 20
9 6 2017-02-12    2771 Oberw... Arleshe... 20
10 7 2017-02-12    2771 Oberw... Arleshe... 20
11 8 2017-02-12    2771 Oberw... Arleshe... 20
12 9 2016-11-27    2771 Oberw... Arleshe... 20
13 10 2016-11-27    2771 Oberw... Arleshe... 20
14 # i 359 more rows
15 # i 12 more variables: answer <chr>, perce
16 #   percent_turnout <dbl>, eligible_voters
17 #   invalid <dbl>, yeas <dbl>, nays <dbl>,
18 #   url_web <chr>
```

Zeilen filtern

```
1 filter(  
2   abstimmung,  
3   name == "Oberwil (BL)" & percent_turnout > 60  
4 )
```

Zeilen filtern

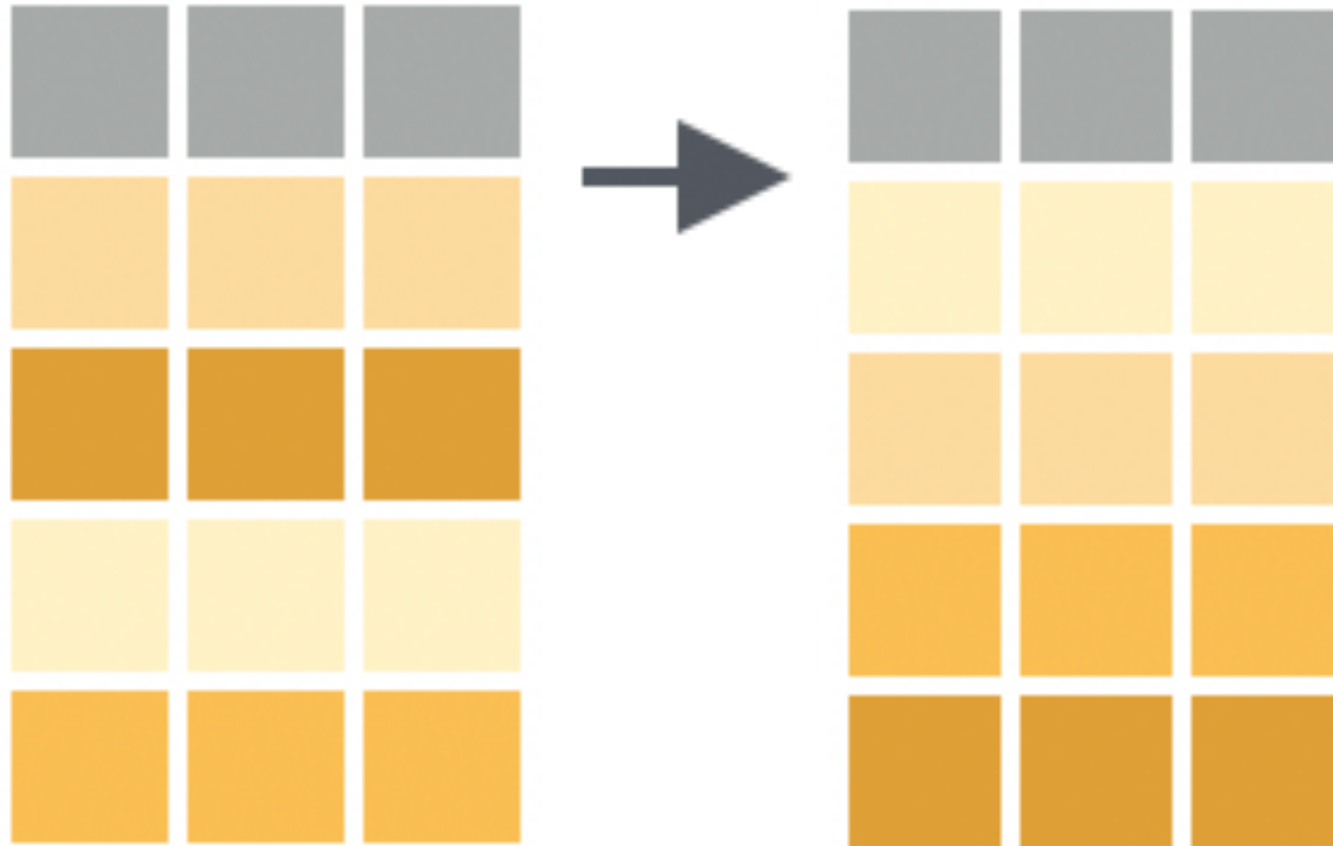
```
1 filter(
2   abstimmung,
3   name == "Oberwil (BL)" & percent_turnout > 60
4 )
```

```
1 # A tibble: 12 × 21
2   date      entity_id name    district vote_id domain type  title_de_CH counted
3   <date>      <dbl> <chr>   <chr>      <chr>   <chr> <chr> <chr>      <lgl>
4 1 2016-02-28    2771 Oberw... Arleshe... 201602... feder... prop... die Volksi... TRUE
5 2 2016-02-28    2771 Oberw... Arleshe... 201602... feder... prop... die Volksi... TRUE
6 3 2016-02-28    2771 Oberw... Arleshe... 201602... feder... prop... die Volksi... TRUE
7 4 2016-02-28    2771 Oberw... Arleshe... 201602... feder... prop... die Änderu... TRUE
8 5 2021-11-28    2771 Oberw... Arleshe... 202111... feder... prop... Volksiniti... TRUE
9 6 2021-11-28    2771 Oberw... Arleshe... 202111... feder... prop... Volksiniti... TRUE
10 7 2021-11-28    2771 Oberw... Arleshe... 202111... feder... prop... Änderung v... TRUE
11 8 2020-09-27    2771 Oberw... Arleshe... 202009... feder... prop... Volksiniti... TRUE
12 9 2020-09-27    2771 Oberw... Arleshe... 202009... feder... prop... Änderung v... TRUE
13 10 2020-09-27    2771 Oberw... Arleshe... 202009... feder... prop... Bundesbesc... TRUE
14 11 2005-06-05    2771 Oberw... Arleshe... 200506... feder... prop... Bundesbesc... TRUE
15 12 2005-06-05    2771 Oberw... Arleshe... 200506... feder... prop... Bundesgese... TRUE
16 # i 12 more variables: answer <chr>, percent_yeas <dbl>, percent_nays <dbl>,
17 #   percent_turnout <dbl>, eligible_voters <dbl>, expats <dbl>, empty <dbl>,
18 #   invalid <dbl>, yeas <dbl>, nays <dbl>, link_to_canton_results <chr>.
```

Logische Operatoren

<	kleiner	$x \mid y$	x Oder y
<=	kleiner gleich	<code>is.na(x)</code>	x ist NA
>	grösser	<code>!is.na(x)</code>	x ist nicht NA
>=	grösser gleich	$x \%in\% y$	x ist in y
==	genau gleich	$!(x \%in\% y)$	x ist nicht in y
!=	ungleich	$!x$	nicht x
$x \& y$	x UND y		

Reihenfolge der Zeilen ändern



```
arrange(Datensatz, Spalte)
```

Reihenfolge der Zeilen ändern

```

1 arrange(abstimmung, date)

1 # A tibble: 31,734 × 21
2   date      entity_id name    district vote_id domain type  title_de_CH counted
3   <date>      <dbl> <chr>   <chr>      <chr>   <chr> <chr> <chr>      <lgl>
4 1 2003-02-09      2761 Aesch... Arleshe... 200302... feder... prop... Volksrechte TRUE
5 2 2003-02-09      2761 Aesch... Arleshe... 200302... feder... prop... Kantonale ... TRUE
6 3 2003-02-09      2762 Allsc... Arleshe... 200302... feder... prop... Volksrechte TRUE
7 4 2003-02-09      2762 Allsc... Arleshe... 200302... feder... prop... Kantonale ... TRUE
8 5 2003-02-09      2763 Arles... Arleshe... 200302... feder... prop... Volksrechte TRUE
9 6 2003-02-09      2763 Arles... Arleshe... 200302... feder... prop... Kantonale ... TRUE
10 7 2003-02-09      2764 Biel-... Arleshe... 200302... feder... prop... Volksrechte TRUE
11 8 2003-02-09      2764 Biel-... Arleshe... 200302... feder... prop... Kantonale ... TRUE
12 9 2003-02-09      2765 Binni... Arleshe... 200302... feder... prop... Volksrechte TRUE
13 10 2003-02-09      2765 Binni... Arleshe... 200302... feder... prop... Kantonale ... TRUE
14 # i 31,724 more rows
15 # i 12 more variables: answer <chr>, percent_yeas <dbl>, percent_nays <dbl>,
16 #   percent_turnout <dbl>, eligible_voters <dbl>, expats <dbl>, empty <dbl>,
17 #   invalid <dbl>, yeas <dbl>, nays <dbl>, link_to_canton_results <chr>,
18 #   url_web <chr>

```


Reihenfolge der Zeilen ändern

```

1 arrange(abstimmung, desc(date))

1 # A tibble: 31,734 × 21
2   date      entity_id name    district vote_id domain type  title_de_CH counted
3   <date>      <dbl> <chr>    <chr>      <chr>    <chr> <chr> <chr>      <lgl>
4 1 2024-11-24      2761 Aesch... Arleshe... 202411... feder... prop... Bundesbesc... TRUE
5 2 2024-11-24      2761 Aesch... Arleshe... 202411... feder... prop... Änderung v... TRUE
6 3 2024-11-24      2761 Aesch... Arleshe... 202411... feder... prop... Änderung v... TRUE
7 4 2024-11-24      2761 Aesch... Arleshe... 202411... feder... prop... Änderung v... TRUE
8 5 2024-11-24      2762 Allsc... Arleshe... 202411... feder... prop... Bundesbesc... TRUE
9 6 2024-11-24      2762 Allsc... Arleshe... 202411... feder... prop... Änderung v... TRUE
10 7 2024-11-24      2762 Allsc... Arleshe... 202411... feder... prop... Änderung v... TRUE
11 8 2024-11-24      2762 Allsc... Arleshe... 202411... feder... prop... Änderung v... TRUE
12 9 2024-11-24      2763 Arles... Arleshe... 202411... feder... prop... Bundesbesc... TRUE
13 10 2024-11-24      2763 Arles... Arleshe... 202411... feder... prop... Änderung v... TRUE
14 # i 31,724 more rows
15 # i 12 more variables: answer <chr>, percent_yeas <dbl>, percent_nays <dbl>,
16 #   percent_turnout <dbl>, eligible_voters <dbl>, expats <dbl>, empty <dbl>,
17 #   invalid <dbl>, yeas <dbl>, nays <dbl>, link_to_canton_results <chr>,
18 #   url_web <chr>

```

Eindeutige Zeilen finden

```
1 distinct(abstimmung, district)
```

```
1 # A tibble: 5 × 1
2   district
3   <chr>
4 1 Arlesheim
5 2 Laufen
6 3 Liestal
7 4 Sissach
8 5 Waldenburg
```

```
1 count(abstimmung, district)
```

```
1 # A tibble: 5 × 2
2   district      n
3   <chr>    <int>
4 1 Arlesheim  5535
5 2 Laufen    4797
6 3 Liestal   5166
7 4 Sissach   10701
8 5 Waldenburg 5535
```

Zeilen nach *Indices* auswählen

slice()-Familie

```
1 slice(abstimmung, 21:23)
```

```
1 # A tibble: 3 × 21
2   date          entity_id name      distric
3   <date>          <dbl> <chr>    <chr>
4 1 2017-05-21      2769 Münche... Arleshe
5 2 2017-05-21      2770 Muttentz Arleshe
6 3 2017-05-21      2770 Muttentz Arleshe
7 # i 12 more variables: answer <chr>, p
8 #   percent_turnout <dbl>, eligible_vo
9 #   invalid <dbl>, yeas <dbl>, nays <d
10 #   url_web <chr>
```

```
1 slice_sample(abstimmung, n = 3)
```

```
1 # A tibble: 3 × 21
2   date          entity_id name      distric
3   <date>          <dbl> <chr>    <chr>
4 1 2016-06-05      2785 Duggin... Laufen
5 2 2018-06-10      2884 Diegten Waldenb
6 3 2010-11-28      2890 Lieder... Waldenb
7 # i 12 more variables: answer <chr>, p
8 #   percent_turnout <dbl>, eligible_vo
9 #   invalid <dbl>, yeas <dbl>, nays <d
10 #   url_web <chr>
```

Praktikum - Zeilen transformieren mit {dplyr}

prak-02c-dplyr-abstimmung.qmd

20:00

rstatsBL - Data Science mit R

Danke! 🍷

Slides created via [revealjs](#) and [Quarto](#).

Access slides as [PDF](#).

All material is licensed under [Creative Commons Attribution Share Alike 4.0 International](#).