

Data Science Lifecycle und Visualisierung

Unit 1

Willkommen! 🙌

Themen

- *Data Science Lifecycle*
- Daten visualisieren
- *Tidy data*, Daten aufräumen und umwandeln
- mit Quarto kommunizieren

Kennenlernen

- Dienststelle / Aufgabe
- was willst du lernen?
- was machst du gerne ausserhalb der Arbeit?

Ziele für heute

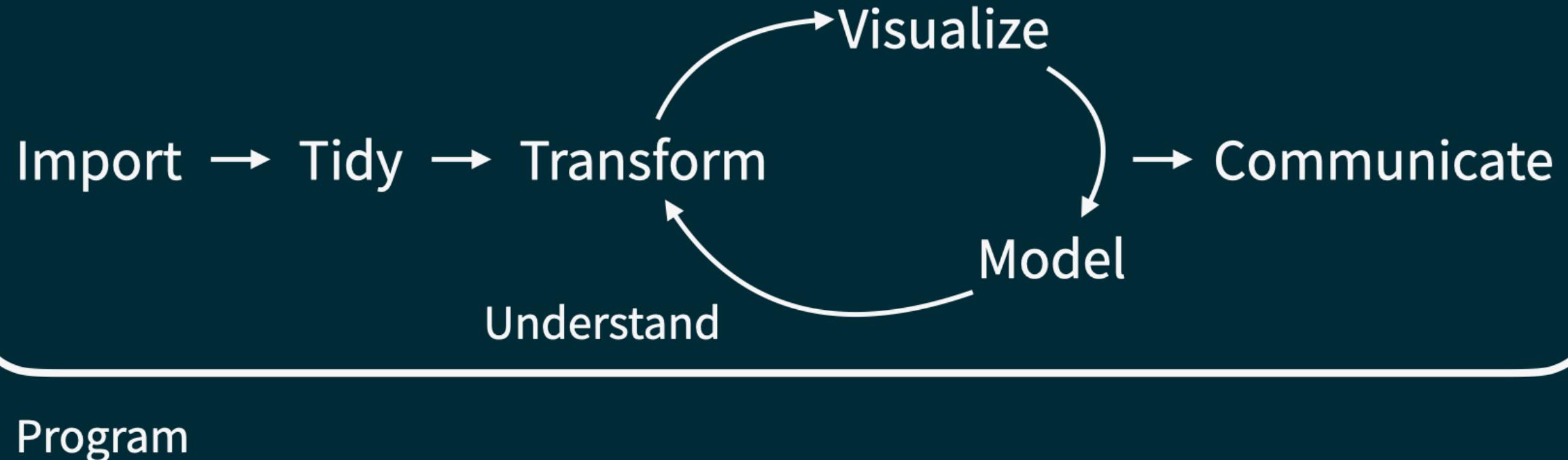
1. die sechs Elemente des *Data Science Lifecycles* aufzählen
2. *aesthetic mappings* für Datenvisualisierung im `{ggplot2}` R Package identifizieren
3. vier Komponenten einer Quarto-Datei identifizieren

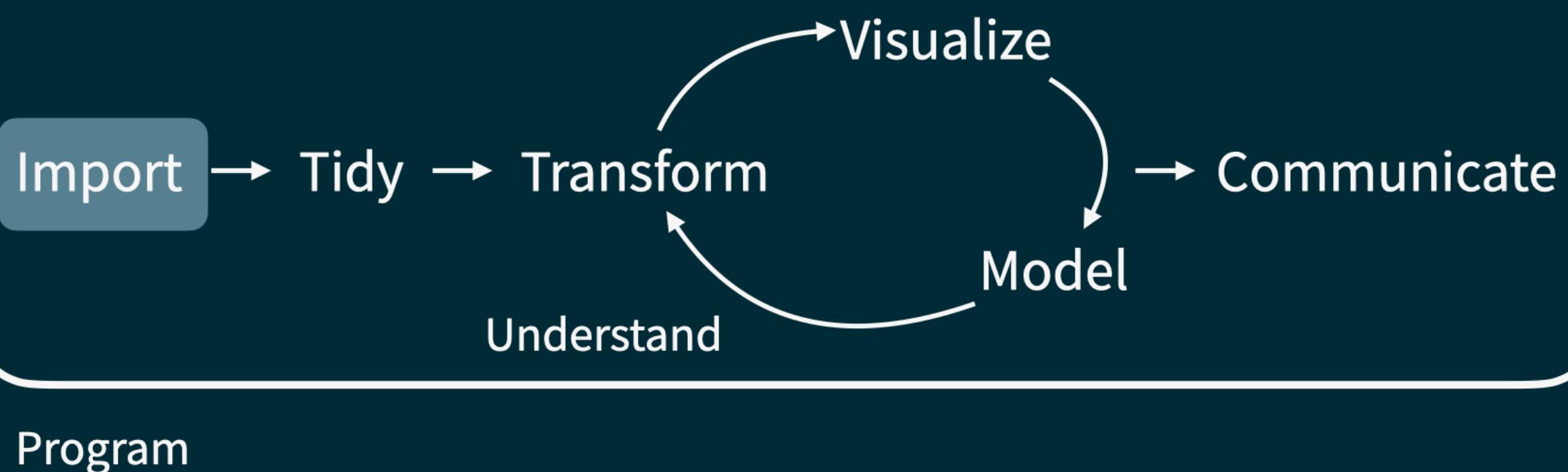
Data Science Lifecycle

Reproduzierbare Datenanalyse

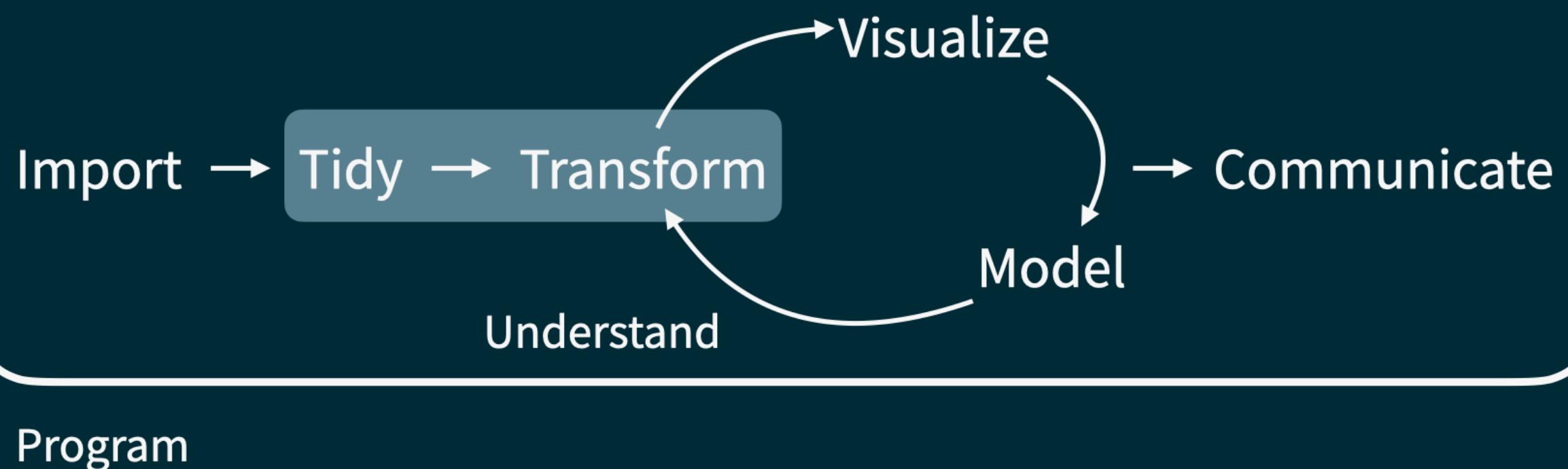


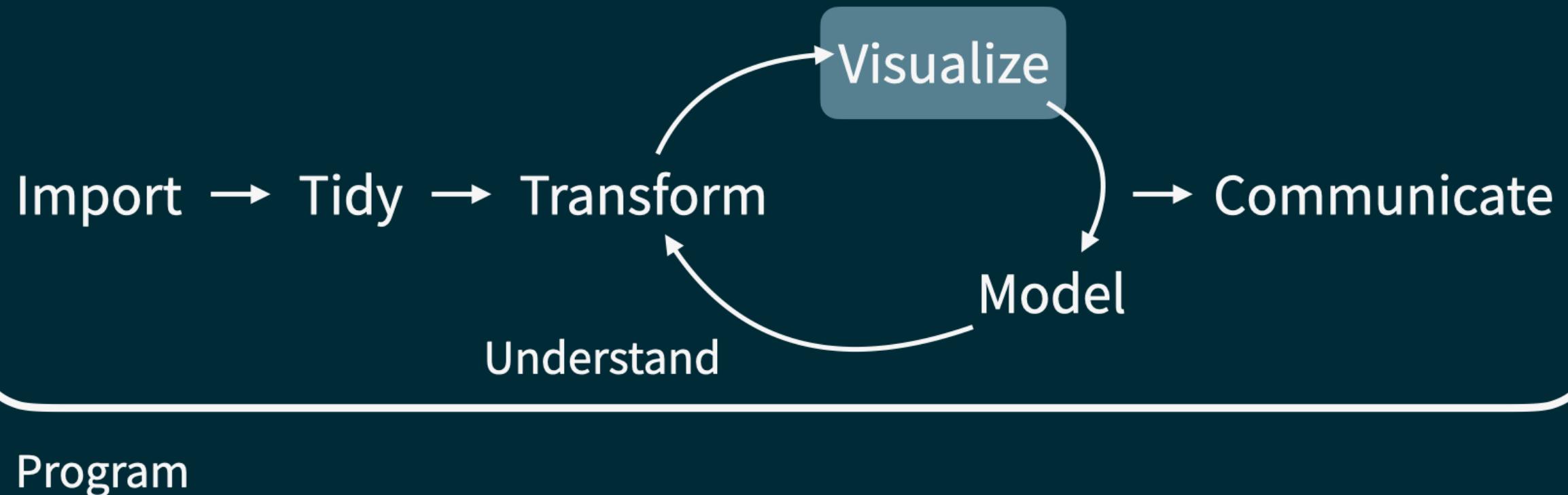
- Tabellen und Grafiken aus Daten und Code reproduzieren?
- Code in anderen Skripten verwenden?
- Stimmt meine Umgebung mit der meiner Kollegen/innen überein?

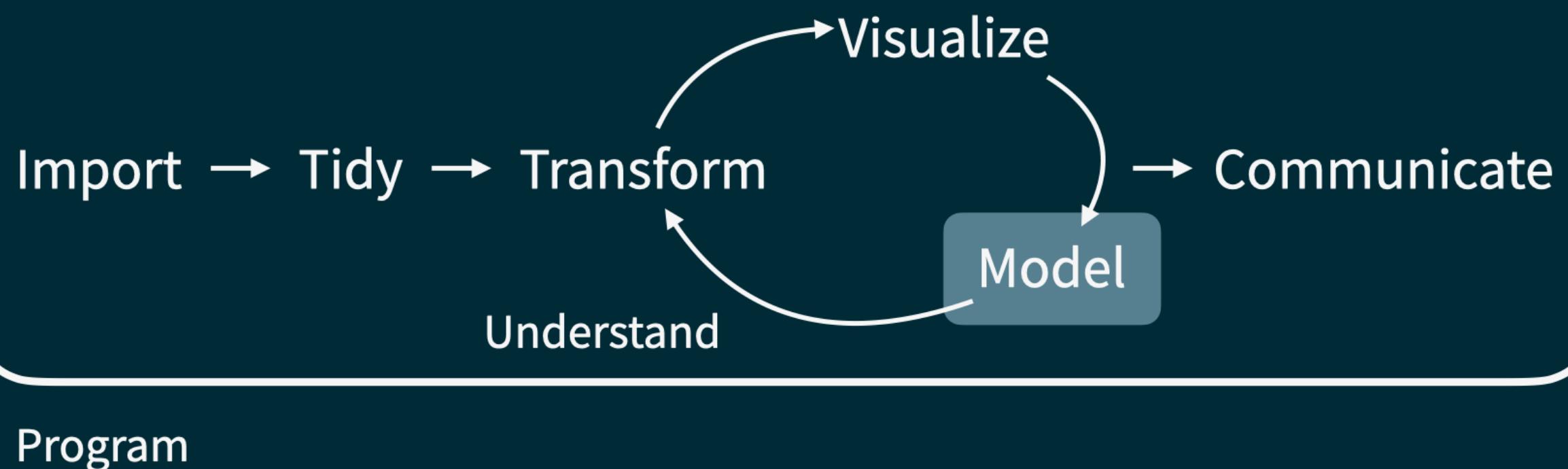




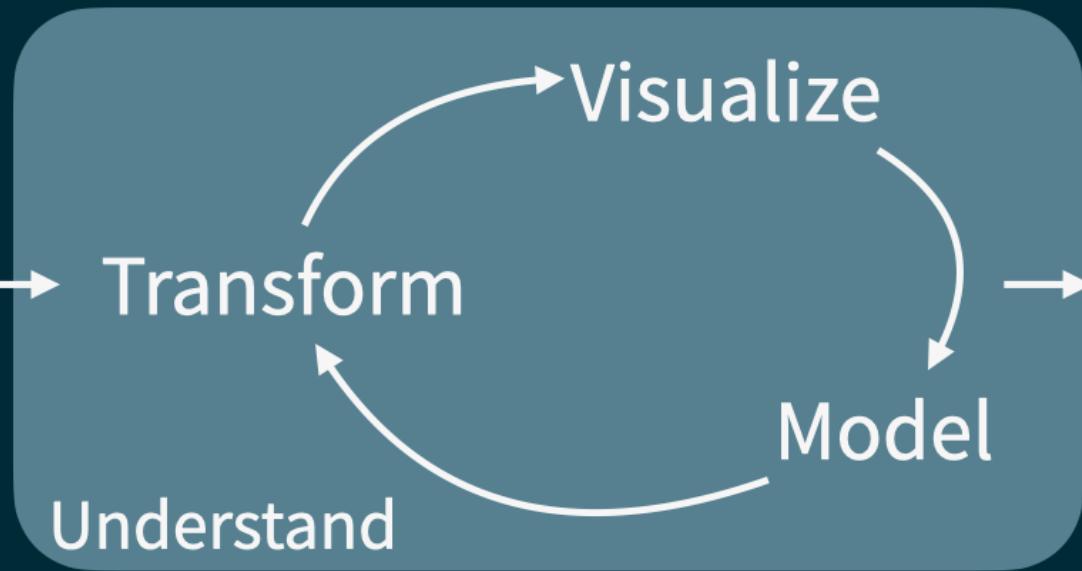
Program



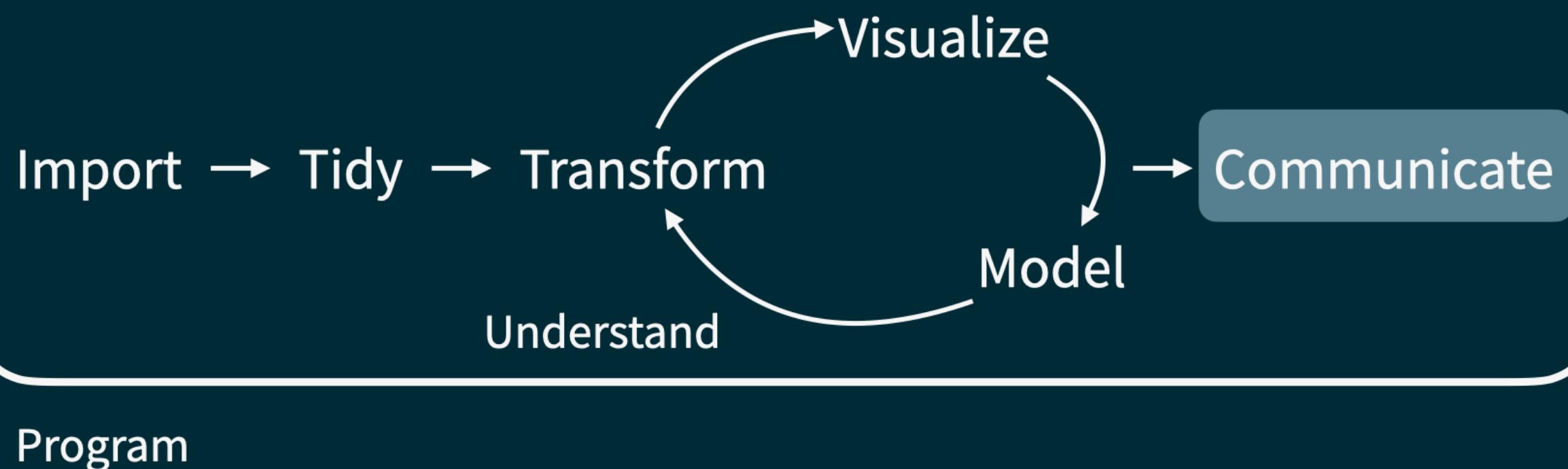




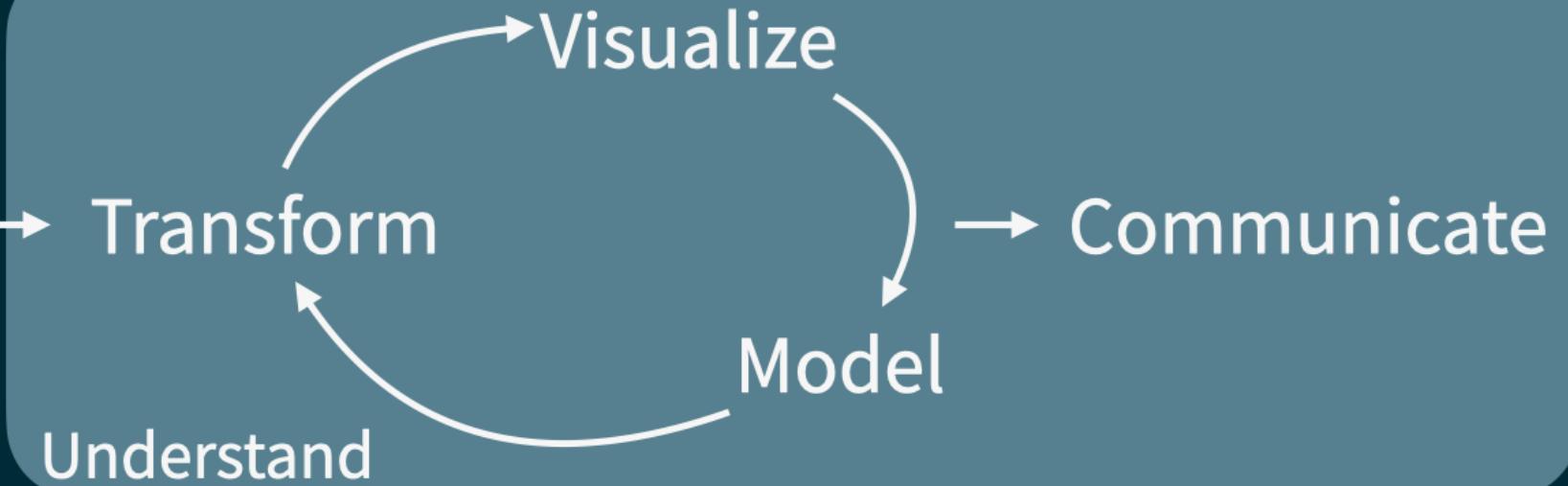
Import → Tidy → Transform → Communicate



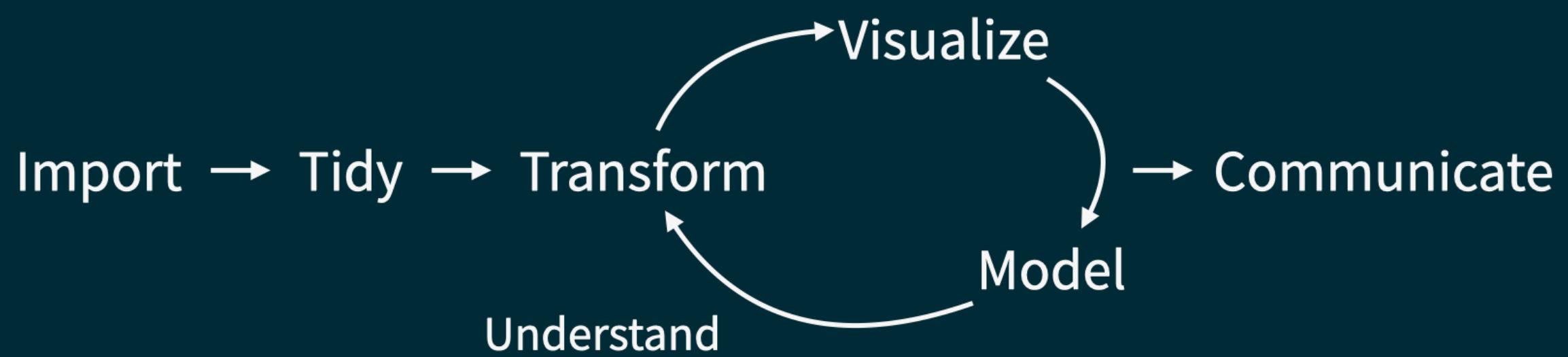
Program



Import → Tidy → Transform



Program



Program

Kurs-Werkzeuge

Programmierung:

- R
- RStudio
- tidyverse
- Quarto

Versionskontrolle & Kollaboration:

-  git
- 
- 

Reproduzierbare Datenanalyse

Hallo R! 🙌

Du brauchst die Sprache



und das IDE



R-Packages

Du benutzt R durch *packages*

```
1 install.packages("package")
2 library(package)
```

...die *functions* enthalten

...die häufig nur Befehle sind

```
1 do_this(to_this)
2 do_that(to_this, to_that, with_those)
```

Comprehensive R Archive Network (CRAN)

2023_kgwr - Excel

	A	B	C	D	E	F	G	H	I
1	egid	gemeindenummer_bfs	gemeindenname	egrid	grundstuecksnummer	name_des_gebaeudes	e_gebaeudekoordinate	n_gebaeudekoordinate	koordinatenherkunft_
2	399274	2765	Binningen	CH430779396888	4425	Wohnhaus	2609253.2435524277	1265322.5141663	901
3	399264	2765	Binningen	CH843907797612	881	Wohnhaus	2609106.991	1265210.808	901
4	2349133	2765	Binningen	CH627740073945	193	Wohnhaus	2609241.831	1265966.129	901
5	245071585	2867	Wittinsburg	CH367761088075	1149	Wohnwagen	2630482.114	1252480.956	901
6	398808	2765	Binningen	CH413923077972	1768	EFH mit Garagen	2610488.611106762	1265609.558332078	901
7	245070458	2821	Arisdorf	CH427902175960	5232	Doppelgarage	2624584.9785663174	1261333.6971606945	901
8	2353830	2789	Nenzlingen	CH141283798493	623	Lager- Und Buerogebaeude 3 Stock	2608404.9355578646	1254835.3211961393	901
9	245070516	2765	Binningen	CH487607397711	1309	MFH	2609882.555269362	1264290.6638168446	901
10	245071338	2792	Wahlen	CH410490377735	2241	EFH mit Carport	2605866.204107465	1250318.511324841	901
11	1368538	2789	Nenzlingen	CH367783127861	518	Wohnhaus	2609617.2212955626	1255197.1511377348	901
12	245043651	2768	Ettingen	CH327112073976	3819	Wohnhaus mit Carport	2608106.835	1258675.896	901
13	245073835	2768	Ettingen	CH957107783943	2853	DEFH A	2608760.002553694	1259293.164164539	901
14	245043603	2765	Binningen	CH567007397809	2222	4 EFH mit AEH	2609997.136	1265457.024	901
15	245043601	2765	Binningen	CH567007397809	2222	4 EFH mit AEH	2609986.113	1265462.835	901
16	245043567	2765	Binningen	CH513939077988	6687	EFH	2609703.92	1264356.392	901
17	245043570	2765	Binningen	CH790779473987	6638	DEFH	2609023.335	1265581.5	901
18	245043596	2765	Binningen	CH930771399362	6909	EFH	2609799.887	1265322.793	901
19	245043597	2765	Binningen	CH943907719382	6910	EFH	2609802.892	1265328.246	901
20	245043387	2841	Anwil	CH980161087916	1543	EFH mit Carport	2637988.5159741053	1255818.6108785928	901
21	245073806	2861	Sissach	CH570878336128	1245	Autoeinstellhalle	2628043.8160701953	1257782.3801204867	901
22	245043563	2765	Binningen	CH407139350706	7062	EFH	2609125.176	1264229.117	901
23	245043518	2865	Wenslingen	CH607908296162	1061	EFH	2635387.413	1254443.17	901
24	245043562	2765	Binningen	CH713914077159	6892	EFH	2609133.624	1264238.866	901
25	433835	2887	Lampenberg	CH840865770615	262	Wohnhaus	2624048.902354877	1252342.0210463002	901

R website - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Addins Go to file/function Addins

kgwr_2023

Filter Cols: << 1 - 50 >>

egid	gemeindenummer_bfs	gemeindename	egrid	grundstücksnummer	name_des_gebaeudes
1	399274	2765 Binningen	CH430779396888	4425	Wohnhaus
2	399264	2765 Binningen	CH843907797612	881	Wohnhaus
3	2349133	2765 Binningen	CH627740073945	193	Wohnhaus
4	245071585	2867 Wittinsburg	CH367761088075	1149	Wohnwagen
5	398808	2765 Binningen	CH413923077972	1768	EFH mit Garagen
6	245070458	2821 Arisdorf	CH427902175960	5232	Doppelgarage
7	2353830	2789 Nenzlingen	CH141283798493	623	Lager- Und Buerogebaeude 3 Stock
8	245070516	2765 Binningen	CH487607397711	1309	MFH
9	245071338	2792 Wahlen	CH410490377735	2241	EFH mit Carport
10	1368538	2789 Nenzlingen	CH367783127861	518	Wohnhaus
11	245043651	2768 Ettingen	CH327112073976	3819	Wohnhaus mit Carport
12	245073835	2768 Ettingen	CH957107783943	2853	DEFH A
13	245043603	2765 Binningen	CH567007397809	2222	4 EFH mit AEH
14	245043601	2765 Binningen	CH567007397809	2222	4 EFH mit AEH
15	245043567	2765 Binningen	CH513939077988	6687	EFH
16	245043570	2765 Binningen	CH790779473987	6638	DEFH
17	245043596	2765 Binningen	CH930771399362	6909	EFH
18	245043597	2765 Binningen	CH943907719382	6910	EFH
19	245043387	2841 Anwil	CH980161087916	1543	EFH mit Carport
20	245073806	2861 Sissach	CH570878336128	1215	Autoeinstellhalle
21	245043563	2765 Binningen	CH407139350706	7062	EFH

Showing 1 to 21 of 127,150 entries, 61 total columns

Console Terminal Background Jobs

R 4.3.0 · C:/R_projects/rstatsBL_workshop/website/

Environment History Connections Build Git Tutorial

Import Dataset 353 MiB

Global Environment

Data kgwr_2023 127150 obs. of 61 variables

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Delete Rename More

C > C:/R_projects/rstatsBL_workshop/exercises

Name	Size	Modified
..	14.6 KB	Aug 22, 2023, 11:41 AM
.Rhistory	218 B	Aug 22, 2023, 11:37 AM
exercises.Rproj	257 B	Aug 22, 2023, 10:19 AM
_quarto.yml	57 B	Aug 22, 2023, 10:02 AM
.gitignore	628 B	Aug 21, 2023, 5:29 PM
references.bib	519.2 KB	Aug 10, 2023, 5:37 PM
week7_starwars.csv	1.1 KB	Jun 27, 2023, 1:15 PM
draft_05.R	6.2 KB	Jun 19, 2023, 10:20 AM
README.md		
archive		
data		
img		
week01		
week02		
week03		
week04		

RStudio

Menü Leiste

The screenshot shows the RStudio IDE interface. The menu bar at the top is highlighted with an orange box and contains the following items: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help. Below the menu bar is a toolbar with various icons for file operations like Open, Save, Print, and a search bar labeled 'Go to file/function'. The main workspace on the left is a code editor showing R Markdown code. The code includes a YAML front matter section and a Bechdel-Test analysis. The right side of the interface features a 'Global Environment' pane and a 'Files' browser pane. The 'Files' browser pane shows a directory structure under 'R-Projects/rstatsbl-website/slides/img/coderbar' with files: RStudio1.png, coderbar.png, coderbar0.png, and coderbar1.png. The 'Global Environment' pane shows memory usage of 927 MB.

```

1 ---  

2 title: "Praktikum 01b - Quarto: Bechdel"  

3 author: "__"  

4 format: html  

5 ---  

6  

7 Der [Bechdel-Test](https://de.wikipedia.org/wiki/Bechdel-Test) fragt, ob in einem Film mindestens zwei Frauen zu sehen sind, die über alles andere als einen Mann miteinander reden. Um den Test zu bestehen, muss ein Film drei Grundvoraussetzungen erfüllen:  

8  

9 - Es gibt mindestens zwei Frauenrollen  

10 - Sie sprechen miteinander  

11 - Sie unterhalten sich über etwas anderes als einen Mann.  

12  

13 ![Art by Srravya \[CC0\], via Wikimedia Commons](img/BechdelTest-768x310.jpg)  

14  

15 In dieser Minianalyse arbeiten wir mit den Daten, die in der *FiveThirtyEight*-Geschichte mit dem Titel ["The Dollar-And-Cents Case Against Hollywood's Exclusion of Women"](https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/) benutzt wurden.  

16
  
```

Console Terminal x Background Jobs x

R 4.4.1 . ~/home/R-Projects/rstatsbl-website/

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Upload Delete Rename More

Home > _home > R-Projects > rstatsbl-website > slides > img > coderbar

Name	Size	Modified
..		
RStudio1.png	0 B	Jan 6, 2025, 10:37 AM
coderbar.png	28.7 KB	Jan 3, 2025, 3:27 PM
coderbar0.png	25.3 KB	Jan 3, 2025, 3:27 PM
coderbar1.png	26.9 KB	Jan 3, 2025, 3:27 PM

RStudio

Menü Leiste

The screenshot shows the RStudio desktop application. The menu bar at the top includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. A red box highlights the 'File' through 'Help' items. Below the menu bar is a toolbar with various icons for file operations like Open, Save, Print, and a search bar labeled 'Go to file/function...'. The main workspace is divided into several panes:

- Code Editor**: The leftmost pane contains R code. A pink box highlights the code area. The code is related to a Bechdel Test analysis.
- Environment**: The top right pane shows the R environment with tabs for Environment, History, Connections, Build, Git, and Tutorial. It displays the global environment and memory usage (927 MB).
- Files**: The bottom right pane is a file browser with tabs for Files, Plots, Packages, Help, Viewer, and Presentation. It shows a directory structure under Home > _home > R-Projects > rstatsbl-website > slides > img > coderbar. A table below lists files with their names, sizes, and modification dates.
- Console**: The bottom left pane shows the R console output for session R 4.4.1.

```

1 ---  

2 title: "Praktikum 01b - Quarto: Bechdel"  

3 author: "__"  

4 format: html  

5 ---  

6  

7 Der [Bechdel-Test](https://de.wikipedia.org/wiki/Bechdel-Test) fragt, ob in einem Film mindestens zwei Frauen zu sehen sind, die über alles andere als einen Mann miteinander reden. Um den Test zu bestehen, muss ein Film drei Grundvoraussetzungen erfüllen:  

8  

9 - Es gibt mindestens zwei Frauenrollen  

10 - Sie sprechen miteinander  

11 - Sie unterhalten sich über etwas anderes als einen Mann.  

12  

13 ![Art by Srravya \[CC0\], via Wikimedia Commons](img/BechdelTest-768x310.jpg)  

14  

15 In dieser Minianalyse arbeiten wir mit den Daten, die in der *FiveThirtyEight*-Geschichte mit dem Titel ["The Dollar-And-Cents Case Against Hollywood's Exclusion of Women"](https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/) benutzt wurden.  

16
  
```

Name	Size	Modified
..	0 B	Jan 6, 2025, 10:37 AM
RStudio1.png	28.7 KB	Jan 3, 2025, 3:27 PM
coderbar.png	25.3 KB	Jan 3, 2025, 3:27 PM
coderbar0.png	26.9 KB	Jan 3, 2025, 3:27 PM

RStudio

The screenshot displays the RStudio interface with several highlighted sections:

- Menü Leiste:** The top navigation bar is highlighted with an orange box.
- Code Editor:** The main editor pane on the left is highlighted with a pink box. It shows R code for a Quarto document, including a Bechdel Test section and a note about the "FiveThirtyEight" case.
- Environment:** The global environment pane on the right is highlighted with a blue box.
- Files:** The file browser pane at the bottom right is highlighted with a grey box. It lists files in the directory: RStudio1.png, coderbar.png, coderbar0.png, and coderbar1.png.

```

1 ---  

2 title: "Praktikum 01b - Quarto: Bechdel"  

3 author: "__"  

4 format: html  

5 ---  

6  

7 Der [Bechdel-Test](https://de.wikipedia.org/wiki/Bechdel-Test) fragt, ob in einem Film mindestens zwei Frauen zu sehen sind, die über alles andere als einen Mann miteinander reden. Um den Test zu bestehen, muss ein Film drei Grundvoraussetzungen erfüllen:  

8  

9 - Es gibt mindestens zwei Frauenrollen  

10 - Sie sprechen miteinander  

11 - Sie unterhalten sich über etwas anderes als einen Mann.  

12  

13 ![Art by Srravya \[CC0\], via Wikimedia Commons](img/BechdelTest-768x310.jpg)  

14  

15 In dieser Minianalyse arbeiten wir mit den Daten, die in der *FiveThirtyEight*-Geschichte mit dem Titel ["The Dollar-And-Cents Case Against Hollywood's Exclusion of Women"](https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/) benutzt wurden.  

16
  
```

RStudio

The screenshot displays the RStudio interface with several labeled components:

- Menü Leiste**: The top menu bar is highlighted with an orange box.
- Code Editor**: The main workspace where R code and Quarto documents are edited. A pink box highlights the editor area.
- Environment**: The right-hand pane showing the R environment, including the Global Environment and a search bar.
- Console**: The bottom-left pane showing the R console output.
- Files**: The bottom-right pane showing the file system structure and a list of files.

Code Editor Content (Quarto Document):

```

1 ---  

2 title: "Praktikum 01b - Quarto: Bechdel"  

3 author: "__"  

4 format: html  

5 ---  

6  

7 Der [Bechdel-Test](https://de.wikipedia.org/wiki/Bechdel-Test) fragt, ob in einem Film mindestens zwei Frauen zu sehen sind, die über alles andere als einen Mann miteinander reden. Um den Test zu bestehen, muss ein Film drei Grundvoraussetzungen erfüllen:  

8  

9 - Es gibt mindestens zwei Frauenrollen  

10 - Sie sprechen miteinander  

11 - Sie unterhalten sich über etwas anderes als einen Mann.  

12  

13 ![Art by Srravya \[CC0\], via Wikimedia Commons](img/BechdelTest-768x310.jpg)  

14  

15 In dieser Minianalyse arbeiten wir mit den Daten, die in der *FiveThirtyEight*-Geschichte mit dem Titel ["The Dollar-And-Cents Case Against Hollywood's Exclusion of Women"](https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/) benutzt wurden.  

16
  
```

Files Content:

Name	Size	Modified
..	0 B	Jan 6, 2025, 10:37 AM
RStudio1.png	28.7 KB	Jan 3, 2025, 3:27 PM
coderbar.png	25.3 KB	Jan 3, 2025, 3:27 PM
coderbar0.png	26.9 KB	Jan 3, 2025, 3:27 PM

RStudio

The screenshot displays the RStudio interface with four main panels:

- Menü Leiste:** The top navigation bar containing File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins.
- Code Editor:** The left panel showing an R Markdown file (Untitled1.qmd) with code and text. The code includes a YAML header and a Bechdel-Test analysis. A pink box highlights the title "Code Editor".
- Environment:** The right panel showing the Global Environment tab with a list of objects. A blue box highlights the title "Environment".
- Console:** The bottom panel showing the R console output. A green box highlights the title "Console".

RStudio und R-wesentliches

The screenshot shows the RStudio interface with several key components:

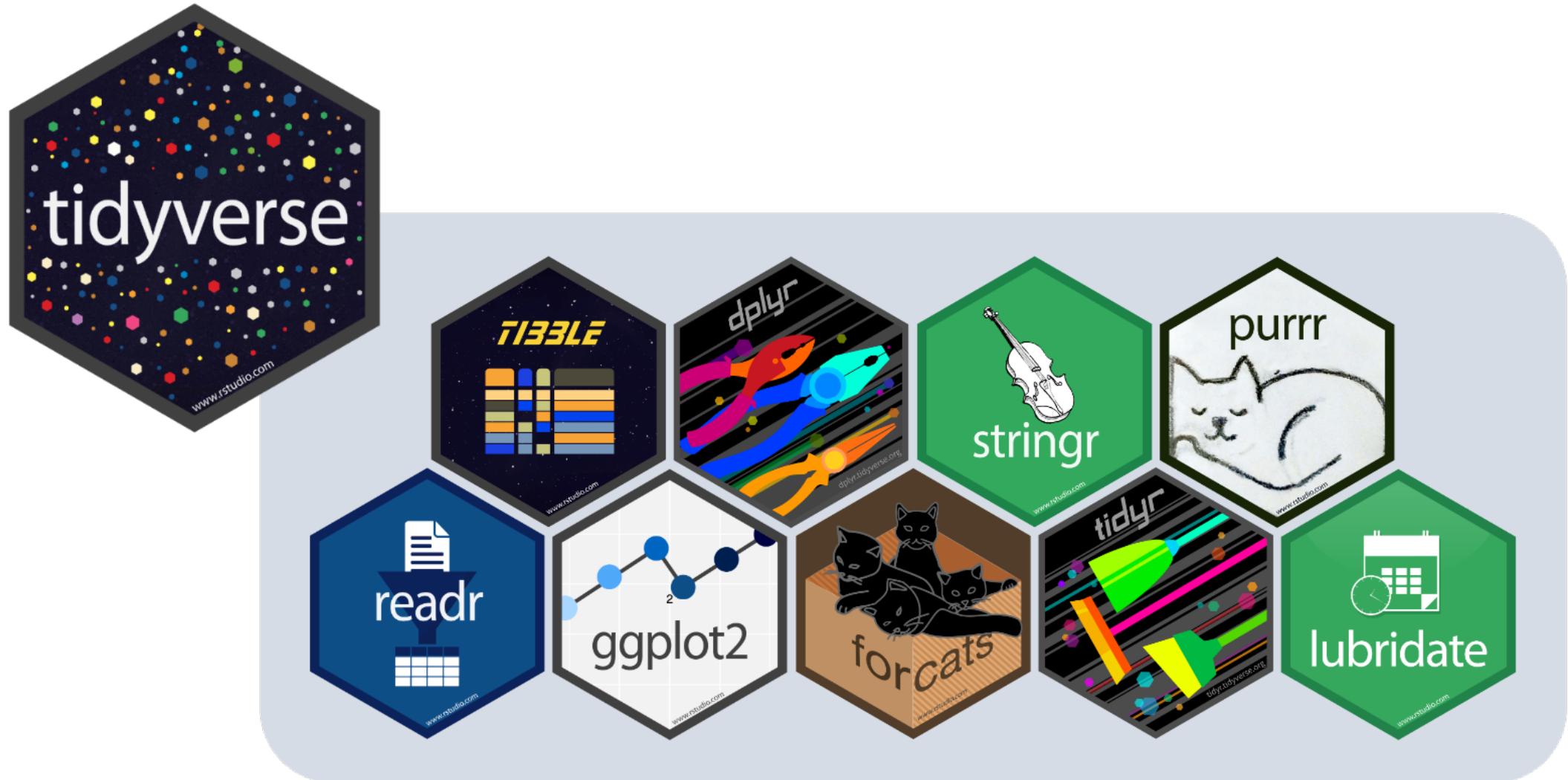
- Data/code viewer:** Displays the "penguins" dataset as a table with columns: species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex, and year. The table shows 15 rows of data for Adelie penguins from Torgersen.
- Environment:** Shows the variable `x` assigned the value 2.
- Help:** Provides documentation for the `mean` function, including its usage, arguments, and description.
- Console:** Shows R session history with annotations:
 - `> 2 + 2` [1] 4 → **Object assignment**
 - `> library(palmerpenguins)` ← **Load package**
 - `> mean(penguins$flipper_length_mm)` [1] NA
 - `> ?mean`
 - `> mean(penguins$flipper_length_mm, na.rm = TRUE)` ← **Function**
 - `[1] 200.9152`

Break

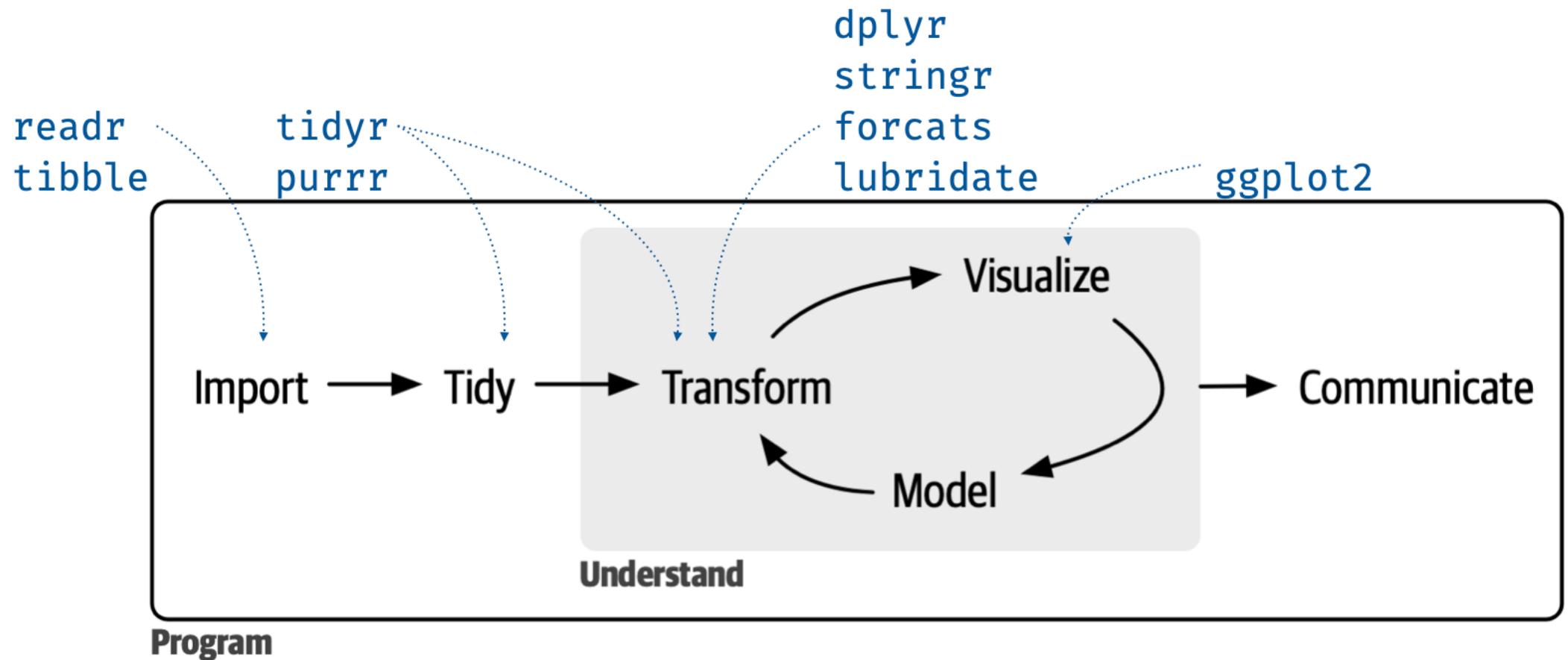


10:00

Tidyverse



Data Science Lifecycle





R exercises - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

unvotes-short-de.qmd

Source Visual Outline

```

1 ---  

2 title: "UN-Abstimmungen"  

3 author: "Souad Guemgar"  

4 date: last-modified  

5 format:  

6 html:  

7 fontsize: 16px  

8 toc: true  

9 toc-location: left  

10 nocite: |  

11   @robinson2021, @mincr2023  

12 ---  

13  

14 ## Einleitung  

15  

16 Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?  

17  

18 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.  

19  

20 Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` [Voeten 2013] für die Daten verwenden.  

21  

22 ``{r}  

23 library(tidyverse)  

24 library(scales)  

25 library(gt)  

26 library(unvotes)  

27 ``  

28  

29 ## UN-Wahlverhalten  

30  

31 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von **Algerien**, **Frankreich** und dem **Vereinigten Königreich** im Laufe der Zeit bei einer Reihe von Themen verändert hat.  

32  

33 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden `filter()`-Anweisung ersetzen. Die genauen Namen der Länder findest du im [Anhang] (#appendix).  

34  

35 ``{r}  

36 un_votes |>  

37   inner_join(un_roll_calls, by = "rcid") |>  

38   inner_join(un_roll_call_issues, by = "rcid") |>  

39   filter(country %in% c("Algeria", "France", "United Kingdom")) |>  

4:20 UN-Abstimmungen Quarto

```

Console

Files Plots Packages Help Viewer Presentation

Table of contents

Einleitung
UN-Wahlverhalten
Anhang

UN-Abstimmungen

Souad Guemgar August 22, 2023

Einleitung

Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?

Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.

Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` (Voeten 2013) für die Daten verwenden.

```

library(tidyverse)
library(scales)
library(gt)
library(unvotes)

```

UN-Wahlverhalten

Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von **Algerien**, **Frankreich** und dem **Vereinigten Königreich** im Laufe der Zeit bei einer Reihe von Themen verändert hat.

Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden `filter()`-Anweisung ersetzen. Die genauen Namen der Länder findest du im [Anhang](#).

```

un_votes |>
  inner_join(un_roll_calls, by = "rcid") |>
  inner_join(un_roll_call_issues, by = "rcid") |>
  filter(country %in% c("Algeria", "France", "United Kingdom")) |>
  mutate(
    year = year(date),

```

The screenshot shows the RStudio interface with a Quarto document titled "unvotes-short-de.qmd". The left pane displays the R code and comments, while the right pane shows the rendered HTML output.

Code (Left Panel):

```

1 ---  
2 title: "UN-Abstimmungen"  
3 author: "Souad Guemgar"  
4 date: last-modified  
5 format:  
6   html:  
7     fontsize: 16px  
8 toc: true  
9 toc-location: left  
10 nocite: |  
11   @robinson2021, @mincr2023  
12 ---  
13  
14 ## Einleitung  
15  
16 Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?  
17  
18 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.  
19  
20 Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` [voeten2013] für die Daten verwenden.  
21 ~~~{r}  
22 library(tidyverse)  
23 library(scales)  
24 library(gt)  
25 library(unvotes)  
26 ~~~  
27  
28  
29 ## UN-Wahlverhalten  
30  
31 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von **Algerien**, **Frankreich** und dem **Vereinigten Königreich** im Laufe der Zeit bei einer Reihe von Themen verändert hat.  
32  
33 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden `filter()`-Anweisung ersetzen. Die genauen Namen der Länder findest du im [Anhang] (#appendix).  
34  
35 ~~~{r}  
36 un_votes |>  
37   inner_join(un_roll_calls, by = "rcid") |>  
38   inner_join(un_roll_call_issues, by = "rcid") |>  
39   filter(country %in% c("Algeria", "France", "United Kingdom")) |>  
4:20 UN-Abstimmungen Quarto

```

Rendered Output (Right Panel):

The rendered output is a Quarto page titled "UN-Abstimmungen" by Souad Guemgar, dated August 22, 2023.

UN-Abstimmungen

Souad Guemgar August 22, 2023

Einleitung

Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?

Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.

Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` (Voeten 2013) für die Daten verwenden.

```

library(tidyverse)
library(scales)
library(gt)
library(unvotes)

```

UN-Wahlverhalten

Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von **Algerien**, **Frankreich** und dem **Vereinigten Königreich** im Laufe der Zeit bei einer Reihe von Themen verändert hat.

Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden `filter()`-Anweisung ersetzen. Die genauen Namen der Länder findest du im [Anhang](#).

```

un_votes |>
  inner_join(un_roll_calls, by = "rcid") |>
  inner_join(un_roll_call_issues, by = "rcid") |>
  filter(country %in% c("Algeria", "France", "United Kingdom")) |>
  mutate(
    year = year(date),

```

The screenshot shows the RStudio interface with a Quarto document titled "unvotes-short-de.qmd" on the left and its rendered output on the right.

Quarto Document (Left):

```

1 ---  

2 title: "UN-Abstimmungen"  

3 author: "Souad Guemghar"  

4 date: last-modified  

5 format:  

6 html:  

7 fontsize: 16px  

8 toc: true  

9 toc-location: left  

10 nocite: |  

11   @robinson2021, @mincr2023  

12 ---  

13  

14 ## Einleitung  

15  

16 Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?  

17  

18 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.  

19  

20 Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` [Voeten 2013] für die Daten verwenden.  

21  

22 ````{r}  

23 library(tidyverse)  

24 library(scales)  

25 library(gt)  

26 library(unvotes)  

27 ````  

28  

29 ## UN-Wahlverhalten  

30  

31 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von Algerien, Frankreich und dem Vereinigten Königreich im Laufe der Zeit bei einer Reihe von Themen verändert hat.  

32  

33 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden filter()-Anweisung ersetzen. Die genauen Namen der Länder findest du im \[Anhang\] (#appendix).  

34  

35 ````{r}  

36 un_votes |>  

37   inner_join(un_roll_calls, by = "rcid") |>  

38   inner_join(un_roll_call_issues, by = "rcid") |>  

39   filter(country %in% c("Algeria", "France", "United Kingdom")) |>  

4:20 UN-Abstimmungen
  
```

RStudio Environment: Environment, History, Connections, Build, Git, Tutorial

RStudio Addins: Go to file/function, Render on Save, ABC, Render, Run, Outline, Publish

Console: Files, Plots, Packages, Help, Viewer, Presentation

Table of contents:

- Einleitung
- UN-Wahlverhalten
- Anhang

Rendered Output (Right):

UN-Abstimmungen

Souad Guemghar August 22, 2023

Einleitung

Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?

Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.

Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` (Voeten 2013) für die Daten verwenden.

```
library(tidyverse)
library(scales)
library(gt)
library(unvotes)
```

UN-Wahlverhalten

Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von **Algerien**, **Frankreich** und dem **Vereinigten Königreich** im Laufe der Zeit bei einer Reihe von Themen verändert hat.

Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden `filter()`-Anweisung ersetzen. Die genauen Namen der Länder findest du im [\[Anhang\]](#).

```
un_votes |>
  inner_join(un_roll_calls, by = "rcid") |>
  inner_join(un_roll_call_issues, by = "rcid") |>
  filter(country %in% c("Algeria", "France", "United Kingdom")) |>
  mutate(
    year = year(date),
```

R exercises - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

unvotes-short-de.qmd

Source Visual

```

1 ---  

2   title: "UN-Abstimmungen"  

3   author: "Souad Guemgar"  

4   date: last-modified  

5   format:  

6     html:  

7       fontsize: 16px  

8   toc: true  

9   toc-location: left  

10  nocite: |  

11    @robinson2021, @mincr2023  

12 ---  

13  

14 ## Einleitung  

15  

16 Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?  

17  

18 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.  

19  

20 Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` [Voeten 2013] für die Daten verwenden.  

21  

22 ``{r}  

23 library(tidyverse)  

24 library(scales)  

25 library(gt)  

26 library(unvotes)  

27 ``  

28  

29 ## UN-Wahlverhalten  

30  

31 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von **Algerien**, **Frankreich** und dem **Vereinigten Königreich** im Laufe der Zeit bei einer Reihe von Themen verändert hat.  

32  

33 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden `filter()`-Anweisung ersetzen. Die genauen Namen der Länder findest du im [Anhang] (#appendix).  

34  

35 ``{r}  

36 un_votes |>  

37   inner_join(un_roll_calls, by = "rcid") |>  

38   inner_join(un_roll_call_issues, by = "rcid") |>  

39   filter(country %in% c("Algeria", "France", "United Kingdom")) |>  

4:20 UN-Abstimmungen Quarto

```

Console

Files Plots Packages Help Viewer Presentation

Table of contents

Einleitung

UN-Wahlverhalten

Anhang

Souad Guemgar August 22, 2023

UN-Abstimmungen

Einleitung

Wie stimmen die verschiedenen Länder in der Generalversammlung der United Nations ab? Wie hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich ihr Abstimmungsverhalten bei bestimmten Themen?

Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.

Wir werden die Packages `tidyverse`, `scales` und `gt` für die Datenaufbereitung und -visualisierung und das Package `unvotes` (Voeten 2013) für die Daten verwenden.

```

library(tidyverse)
library(scales)
library(gt)
library(unvotes)

```

UN-Wahlverhalten

Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das Abstimmungsverhalten von **Algerien**, **Frankreich** und dem **Vereinigten Königreich** im Laufe der Zeit bei einer Reihe von Themen verändert hat.

Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen, indem wir die Ländernamen in der unten stehenden `filter()`-Anweisung ersetzen. Die genauen Namen der Länder findest du im [Anhang](#).

```

un_votes |>
  inner_join(un_roll_calls, by = "rcid") |>
  inner_join(un_roll_call_issues, by = "rcid") |>
  filter(country %in% c("Algeria", "France", "United Kingdom")) |>
  mutate(
    year = year(date),

```

- Text → **Markdown**
- Code → **code chunk**

```
1  ```{r}
2  sqrt(1/5)
3  ```

1 [1] 0.4472136
```

- Code und Text → ***inline code***

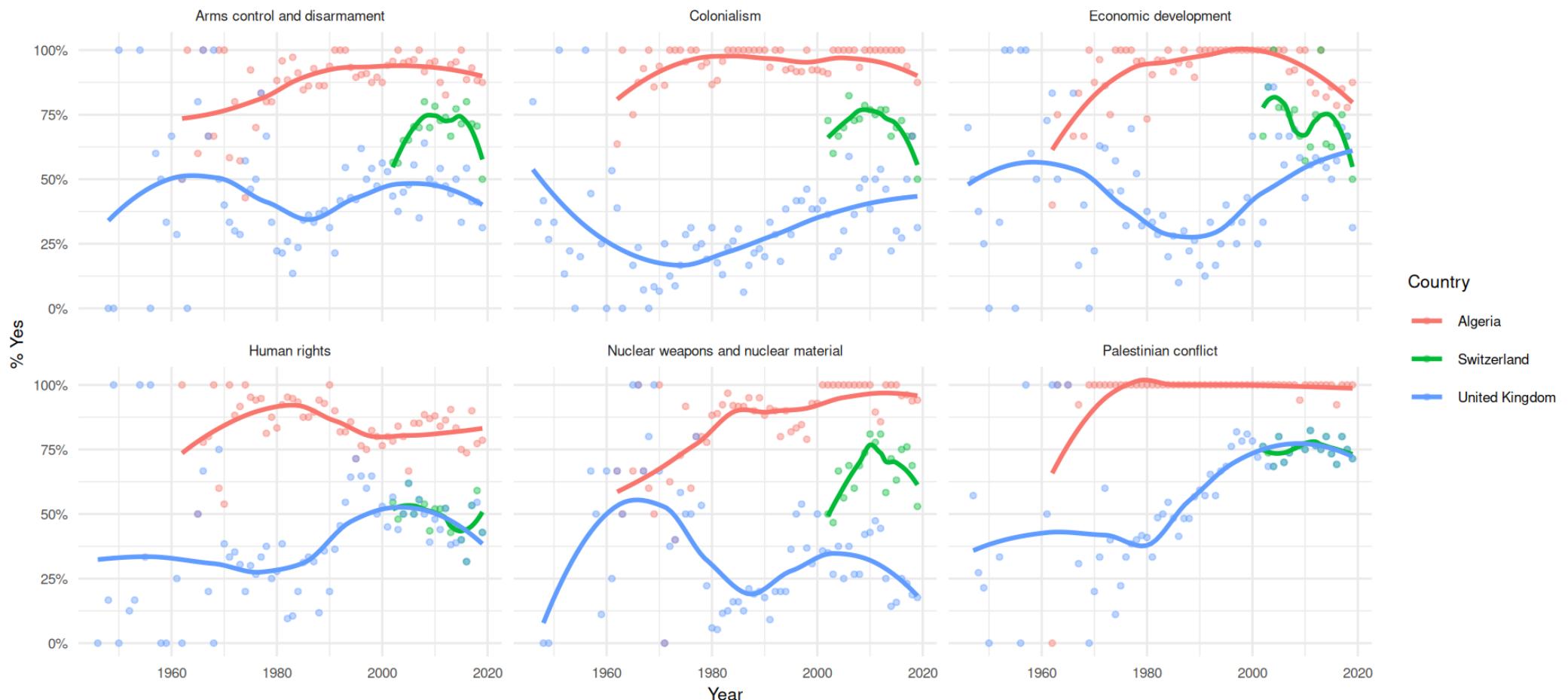
`r sqrt(1/5)` → 0.4472136

Lasst uns eintauchen!



Percentage of 'Yes' votes in the UN General Assembly

1946 to 2019



The screenshot shows three RStudio data frames stacked vertically. The top frame is titled 'un_votes' and has columns 'rcid', 'country', 'country_code', and 'vote'. The middle frame is titled 'un_roll_calls' and has columns 'rcid', 'session', 'importance', 'date', 'unres', 'amend', 'para', and 'short'. The bottom frame is titled 'un_roll_call_issues' and has columns 'rcid', 'short_name', and 'issue'. All three frames show data related to the 'Palestinian conflict'.

	rcid	country	country_code	vote
1	1	un_votes	un_roll_calls	un_roll_call_issues
2	2			
3	3			
4	4			
5	5			
6	6			
7	7			
8	8			
9	9			
10	10			
11	11			
12	12			
13	13			
14	14			
15	15			
16	16			
17	17			
18	18			
19	19			
20	20			
21	21			
22	22			
23	23			
24	24			
25	25			
26	26			

	rcid	session	importance	date	unres	amend	para	short
1	1	un_votes	un_roll_calls	un_roll_call_issues				
2	2							
3	3							
4	4							
5	5							
6	6							
7	7							
8	8							
9	9							
10	10							
11	11							
12	12							
13	13							
14	14							
15	15							
16	16							
17	17							
18	18							
19	19							
20	20							
21	21							
22	22							
23	23							
24	24							
25	25							
26	26							

	rcid	short_name	issue
1	3372	me	Palestinian conflict
2	3658	me	Palestinian conflict
3	3692	me	Palestinian conflict
4	2901	me	Palestinian conflict
5	3020	me	Palestinian conflict
6	3217	me	Palestinian conflict
7	3298	me	Palestinian conflict
8	3429	me	Palestinian conflict
9	3558	me	Palestinian conflict
10	3625	me	Palestinian conflict
11	3714	me	Palestinian conflict
12	3368	me	Palestinian conflict
13	3410	me	Palestinian conflict
14	3539	me	Palestinian conflict
15	3634	me	Palestinian conflict
16	4880	me	Palestinian conflict
17	4126	me	Palestinian conflict
18	4078	me	Palestinian conflict
19	3016	me	Palestinian conflict
20	4290	me	Palestinian conflict
21	4717	me	Palestinian conflict
22	4790	me	Palestinian conflict
23	4483	me	Palestinian conflict
24	4555	me	Palestinian conflict
25	4646	me	Palestinian conflict
26	5020	me	Palestinian conflict

Showing 1 to 26 of 5,281 entries, 3 total columns

Code Plot

```

1  ```{r}
2 #| code-line-numbers: "|5-7|8|16|22"
3 #| eval: false
4
5 un_votes |>
6   inner_join(un_roll_calls, by = "rcid") |>
7   inner_join(un_roll_call_issues, by = "rcid") |>
8   filter(country %in% c("Algeria", "Switzerland", "United Kingdom")) |>
9   mutate(
10     year = year(date),
11     issue = fct_relevel(issue, "Arms control and disarmament"),
12     issue = fct_relevel(issue, "Palestinian conflict", after = Inf)
13   ) |>
14   group_by(country, year, issue) |>
15   summarise(percent_yes = mean(vote == "yes")) |>
16   ggplot(mapping = aes(x = year, y = percent_yes, colour = country)) +
17   geom_point(alpha = 0.4, size = 1) +
18   geom_smooth(method = "loess", se = FALSE) +

```

Praktikum 01a: UN Votes

prak-01a-unvotes.qmd

- Render oder **Ctrl + Shift + K**,
- andere Länder auswählen.

20:00
rstatsBL - Data Science mit R

BASEL
LANDSCHAFT

FINANZ- UND KIRCHENDIREKTION
AMT FÜR DATEN UND STATISTIK

Praktikum 01b: Quarto – Bechdel Test

prak-01b-bechdel.qmd

20:00
rstatsBL - Data Science mit R

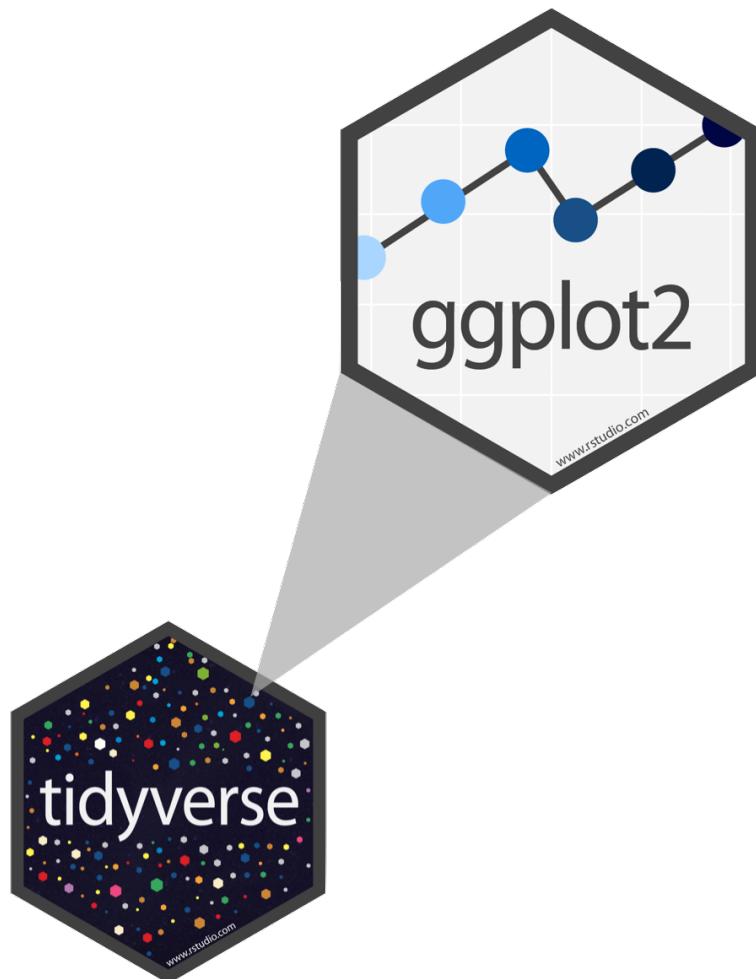
Break



10:00

Visualisierung

R Package `ggplot2`



Streudiagramm (*Scatterplot*)

Verteilungen visualisieren

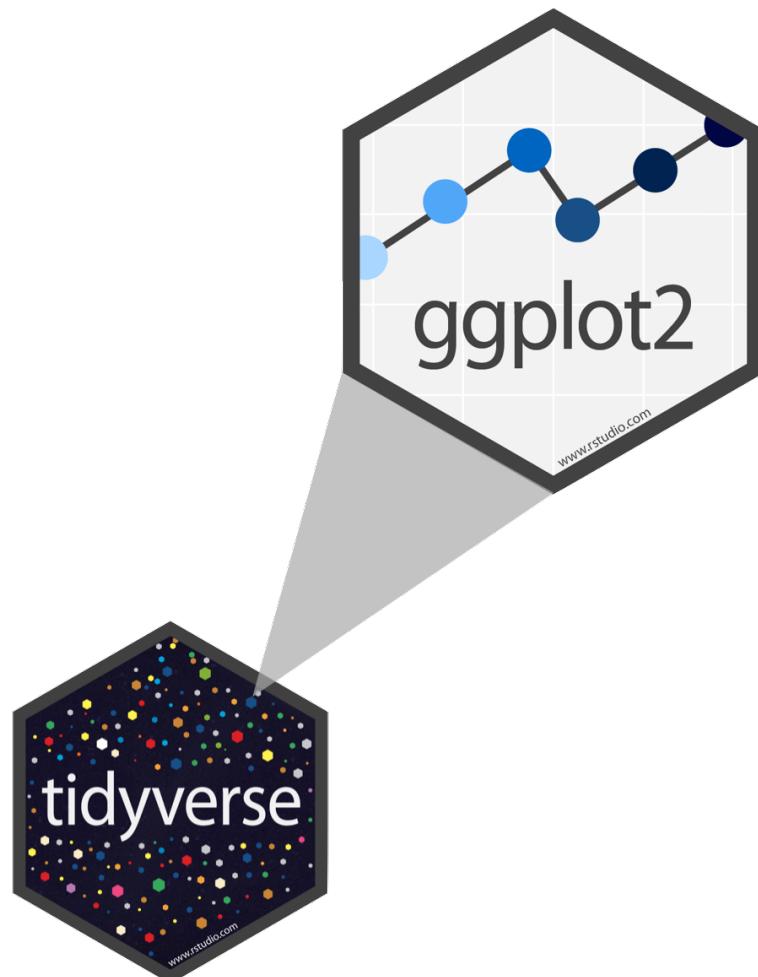
Beziehungen visualisieren

```
1 ggplot(data = [dataset],  
2         mapping = aes(x = [x-variable],  
3                             y = [y-variable])) +  
4         geom_xxx() +  
5         other options
```

Grammar of Graphics

rstatsBL - Data Science mit R

R Package `ggplot2`



Streudiagramm (*Scatterplot*)

Verteilungen visualisieren

Beziehungen visualisieren

```
1 ggplot(data = [dataset],  
2         mapping = aes(x = [x-variable],  
3                             y = [y-variable])) +  
4         geom_xxx() +  
5         other options
```

Grammar of Graphics

rstatsBL - Data Science mit R

Erste Schritte: Streudiagramm

Voraussetzung

```
1  ```{r}
2 # install.packages("tidyverse")
3 library(tidyverse)
4 ```
```

```
1  ```{r}
2 library(palmerpenguins) # data
3 library(ggthemes) # colourblind safe colour palette
4 ```
```

CHINSTRAP!

GENTOO!

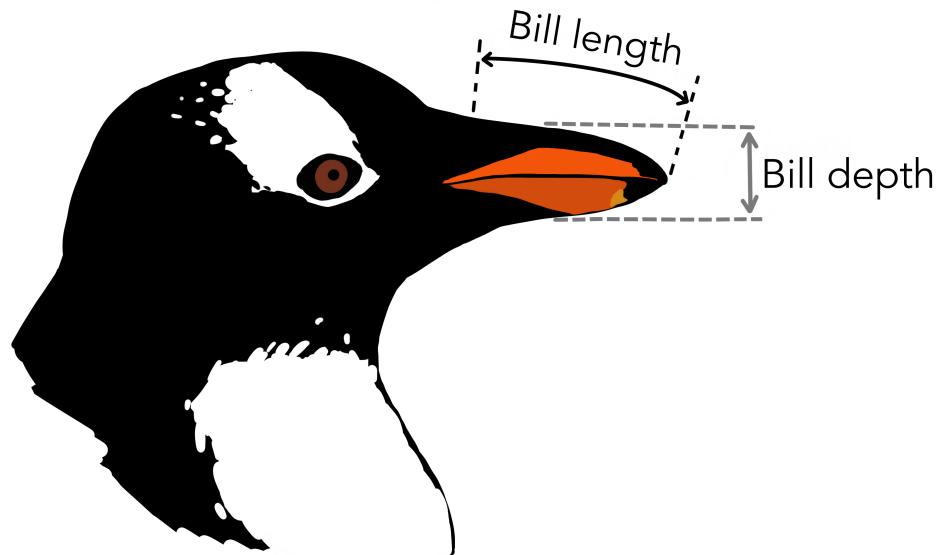
ADÉLIE!



@allison_kern

```
1 ` ``{r}
2 glimpse(penguins)
3 ```

1 Rows: 344
2 Columns: 8
3 $ species <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel...
4 $ island <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgers...
5 $ bill_length_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, ...
6 $ bill_depth_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, ...
7 $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186...
8 $ body_mass_g <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, ...
9 $ sex <fct> male, female, female, NA, female, male, female, male...
10 $ year <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007...
```



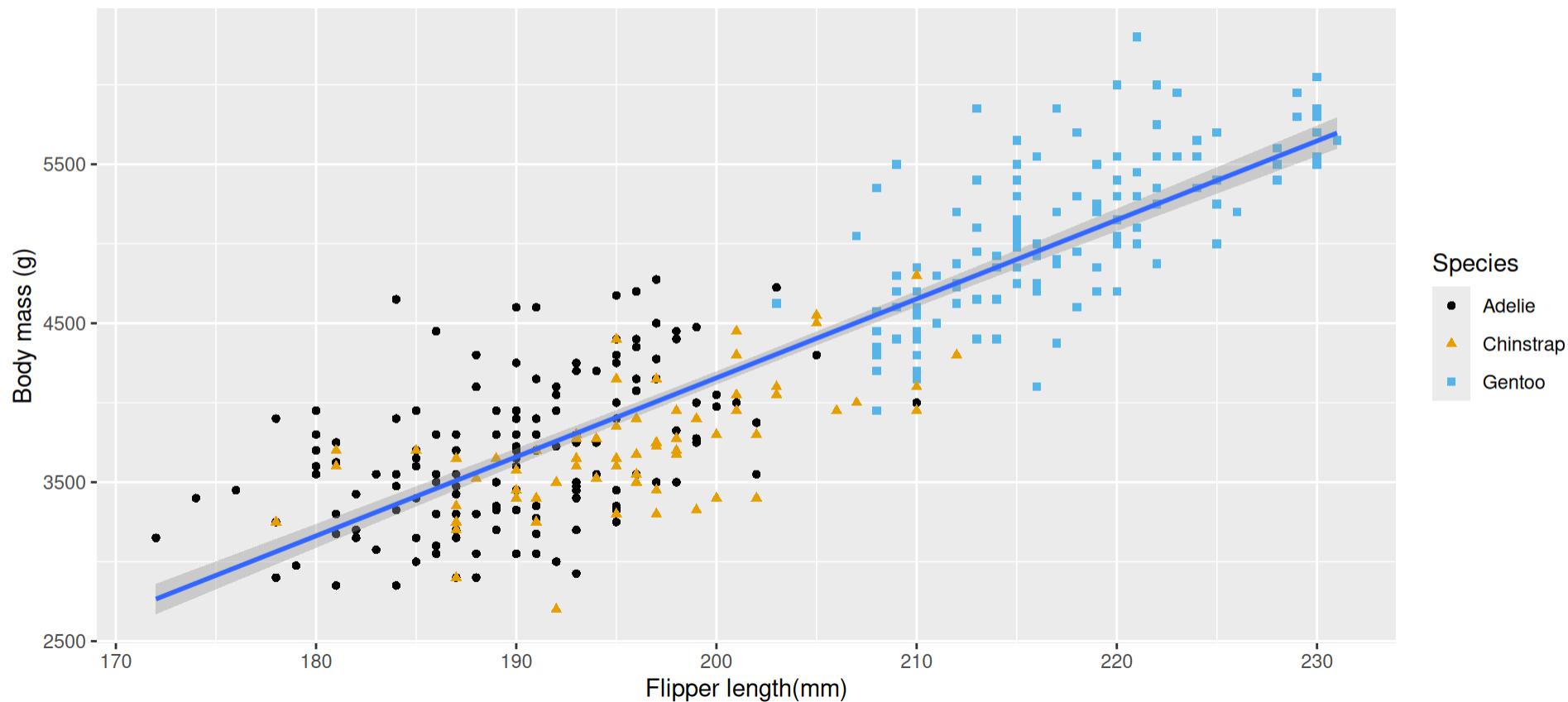
Unser Ziel

Plot Code

1

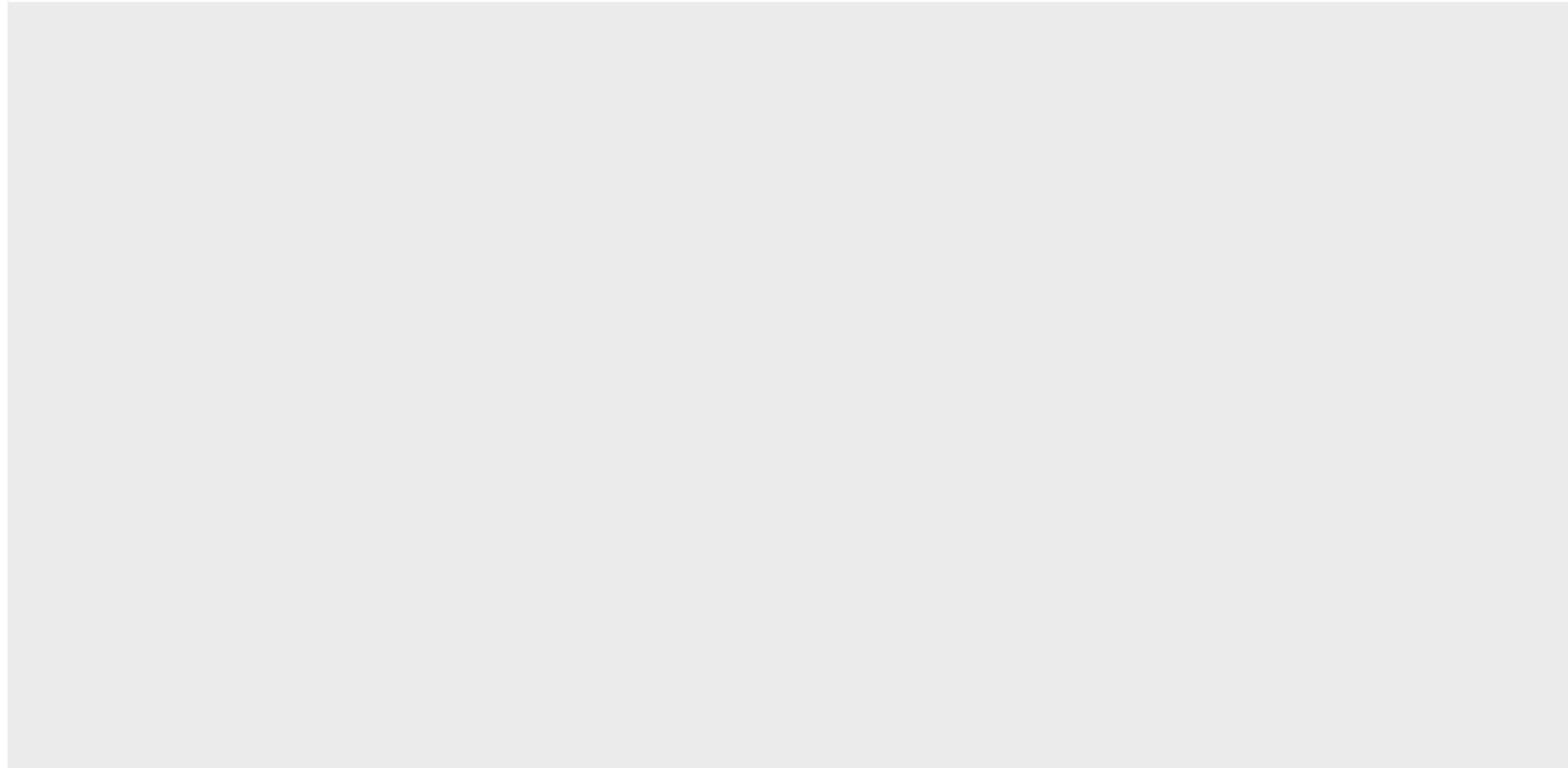
Body mass and flipper length

Dimensions for Adelie, Chinstrap, and Gentoo Penguins



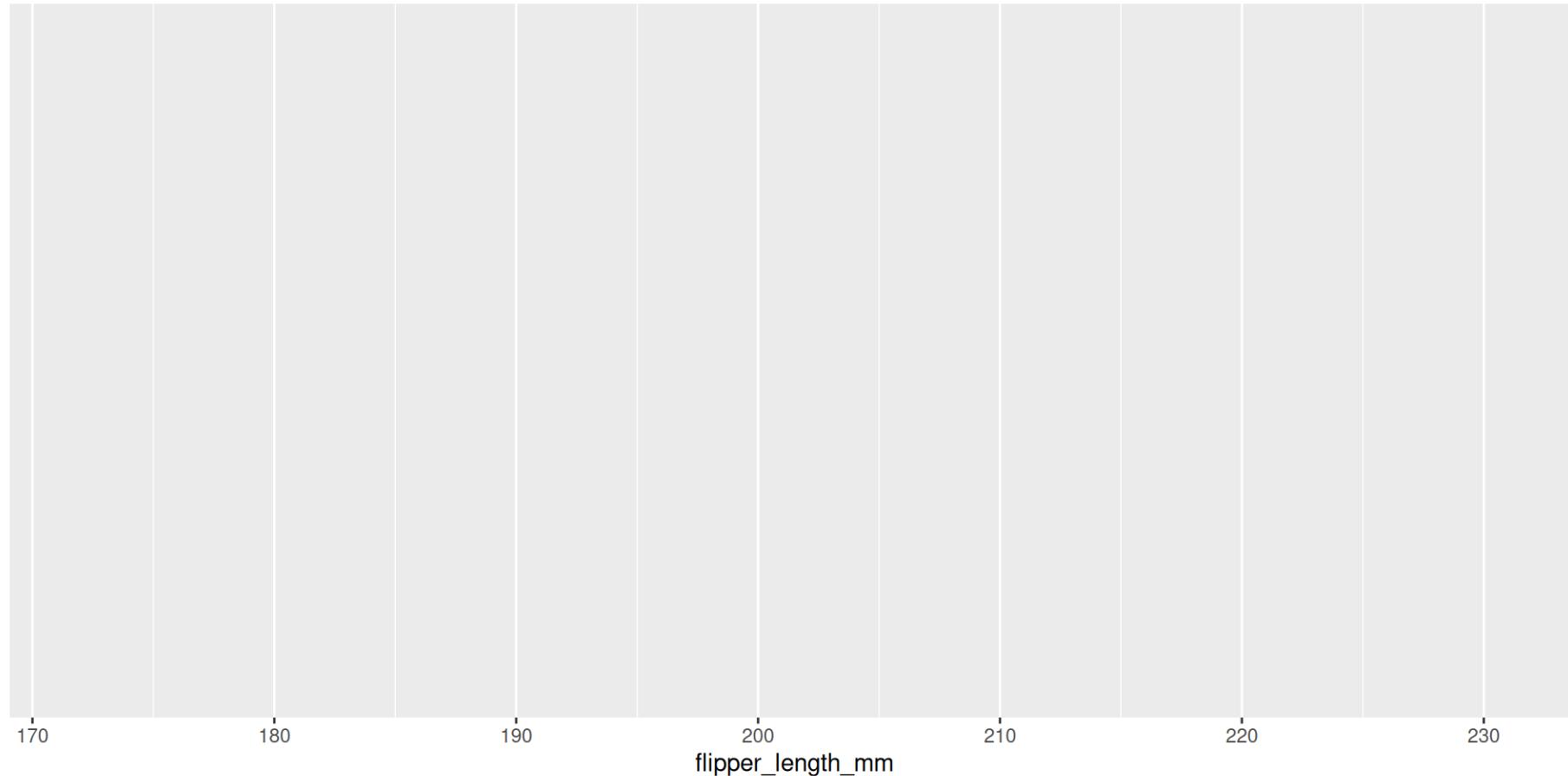
Plot Erstellen

```
1 ggplot(  
2   data = penguins  
3 )
```



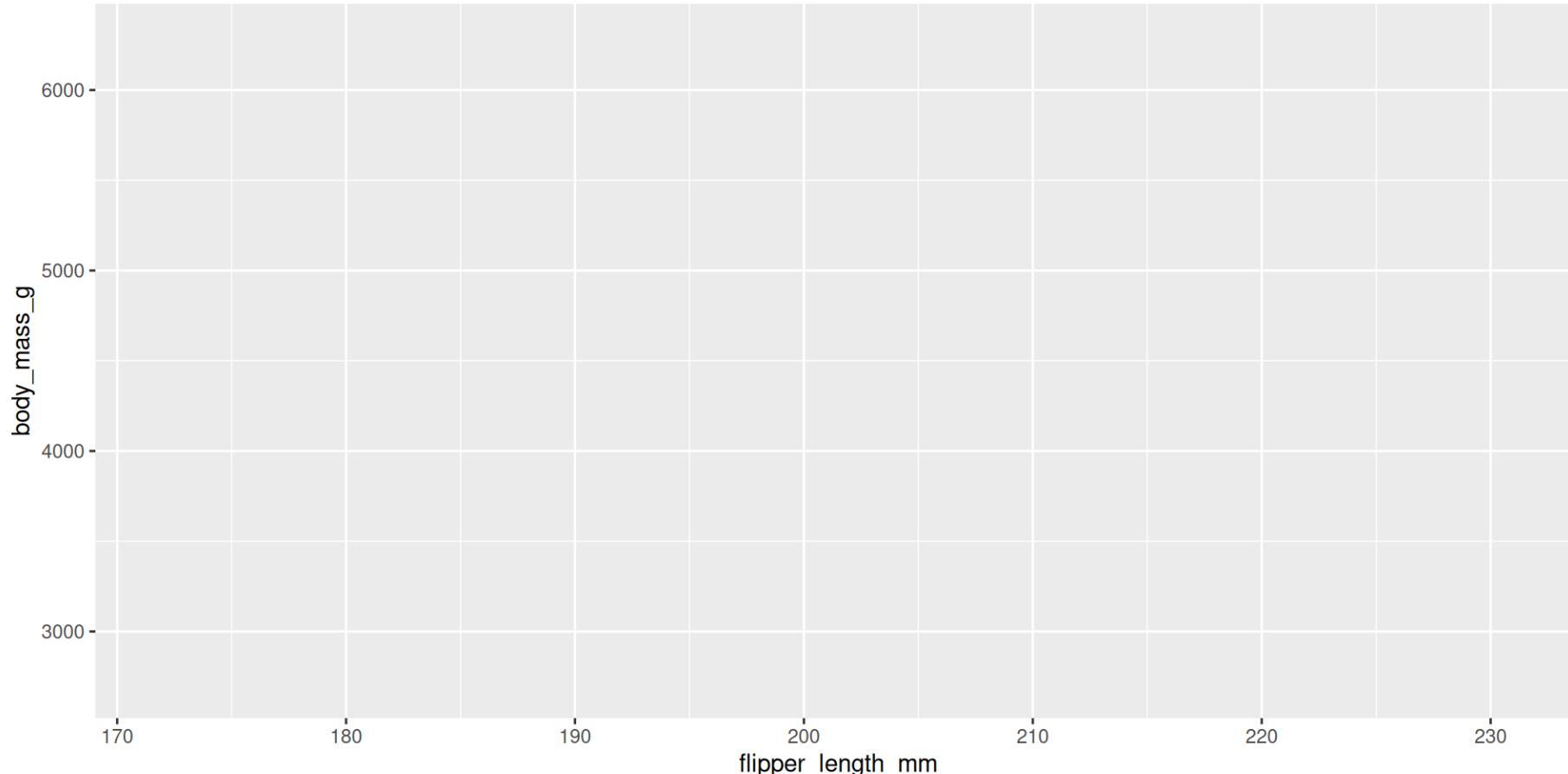
Plot Erstellen

```
1 ggplot(  
2   data = penguins,  
3   mapping = aes(x = flipper_length_mm))  
4 )
```



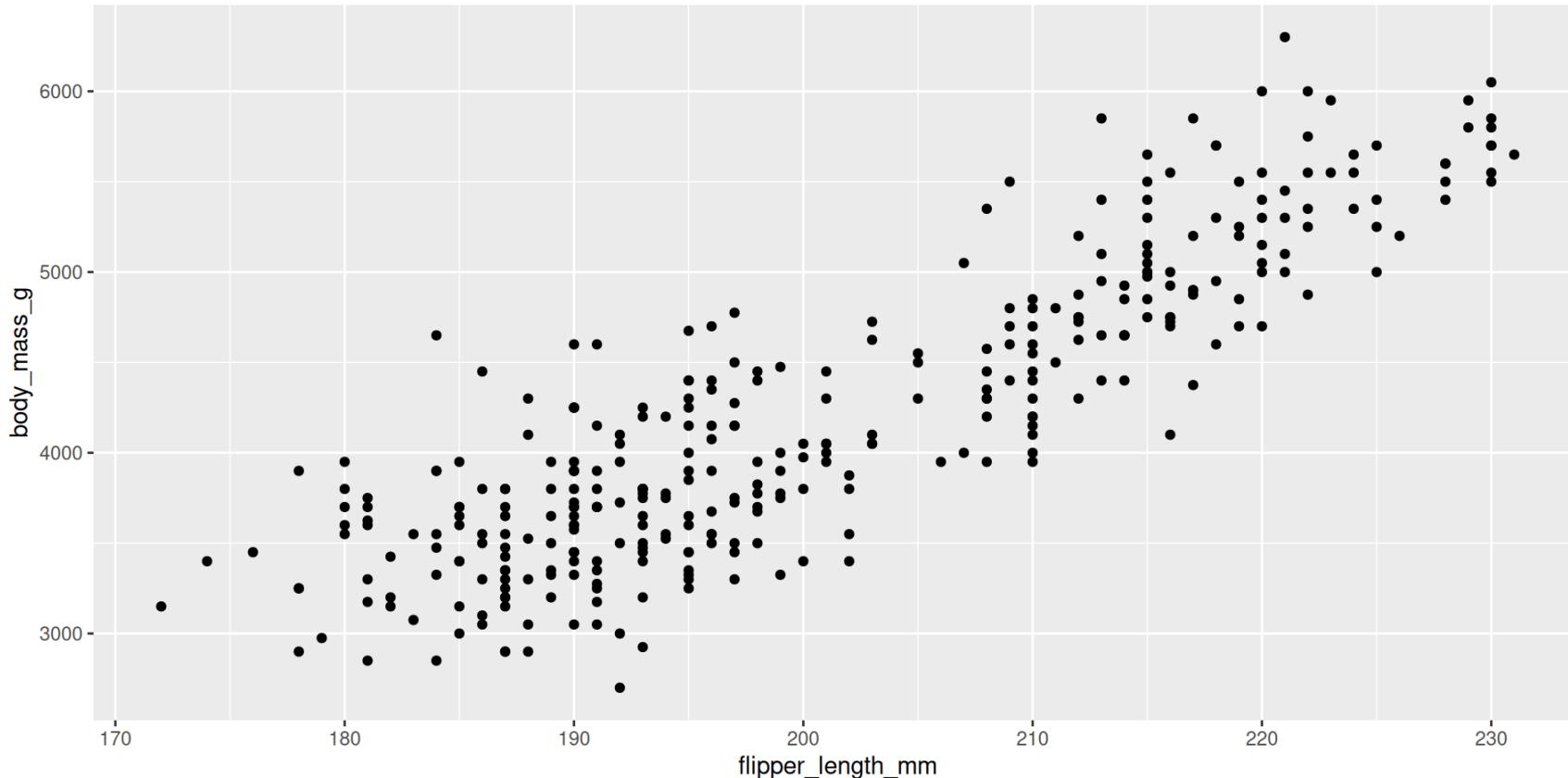
Plot Erstellen

```
1 ggplot(  
2   data = penguins,  
3   mapping = aes(x = flipper_length_mm, y = body_mass_g)  
4 )
```



Plot Erstellen

```
1 ggplot(  
2   data = penguins,  
3   mapping = aes(x = flipper_length_mm, y = body_mass_g)  
4 ) +  
5   geom_point()
```

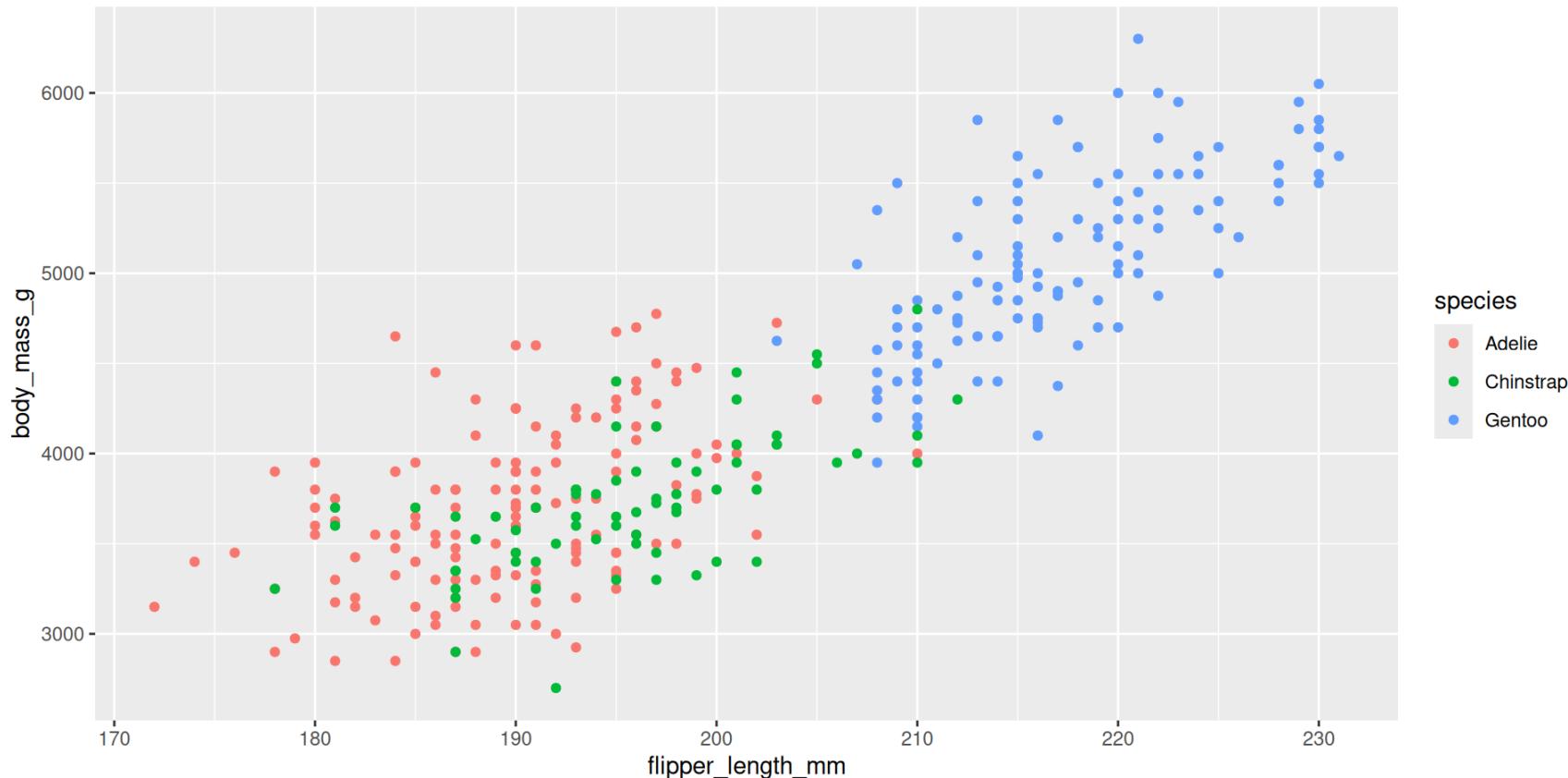


Aesthetics und Schichten

```

1 ggplot(
2   data = penguins,
3   mapping = aes(x = flipper_length_mm, y = body_mass_g, colour = species)
4 ) +
5   geom_point()

```

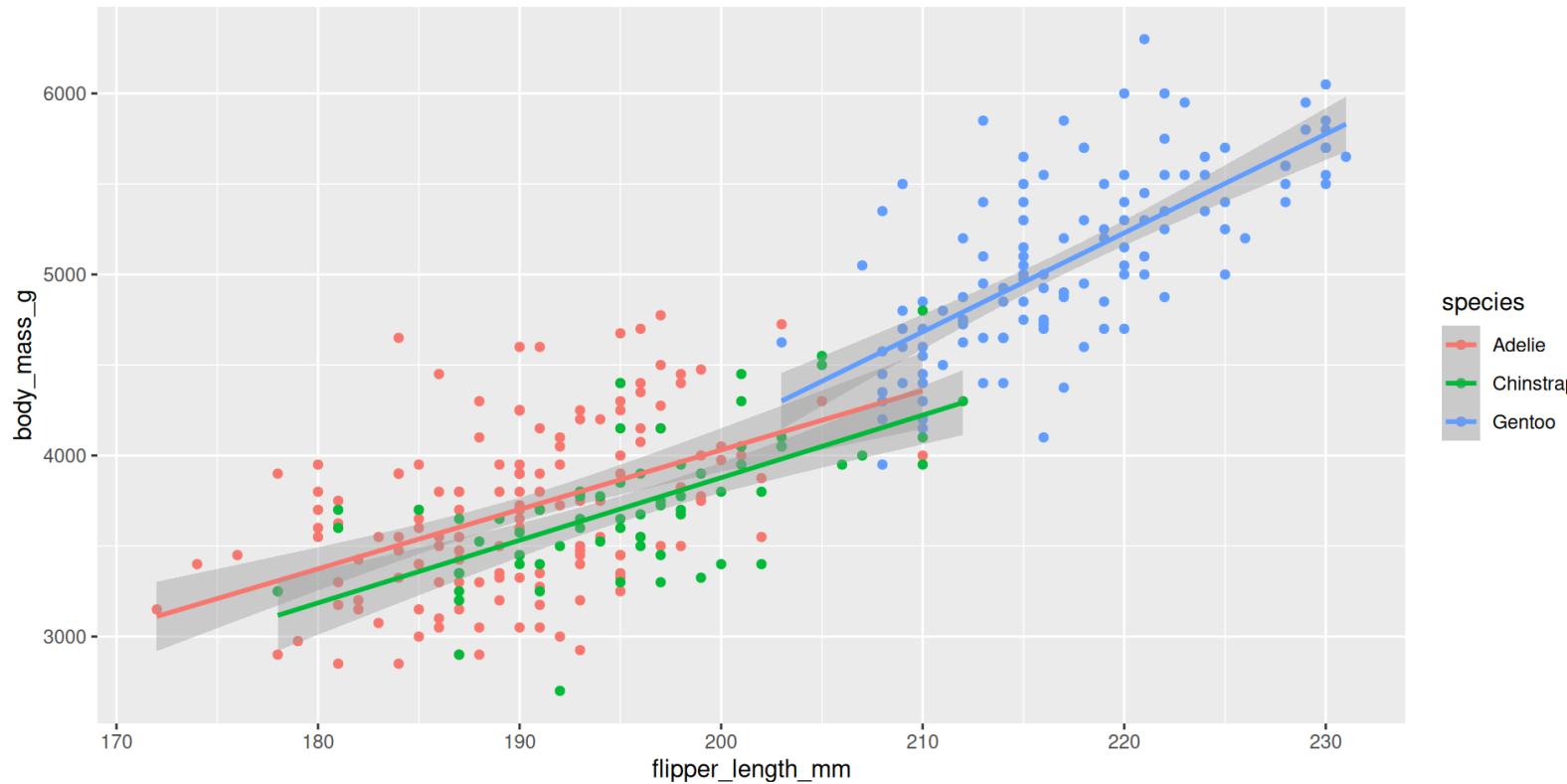


Aesthetics und Schichten

```

1 ggplot(
2   data = penguins,
3   mapping = aes(x = flipper_length_mm, y = body_mass_g, colour = species)
4 ) +
5   geom_point() +
6   geom_smooth(method = "lm")

```

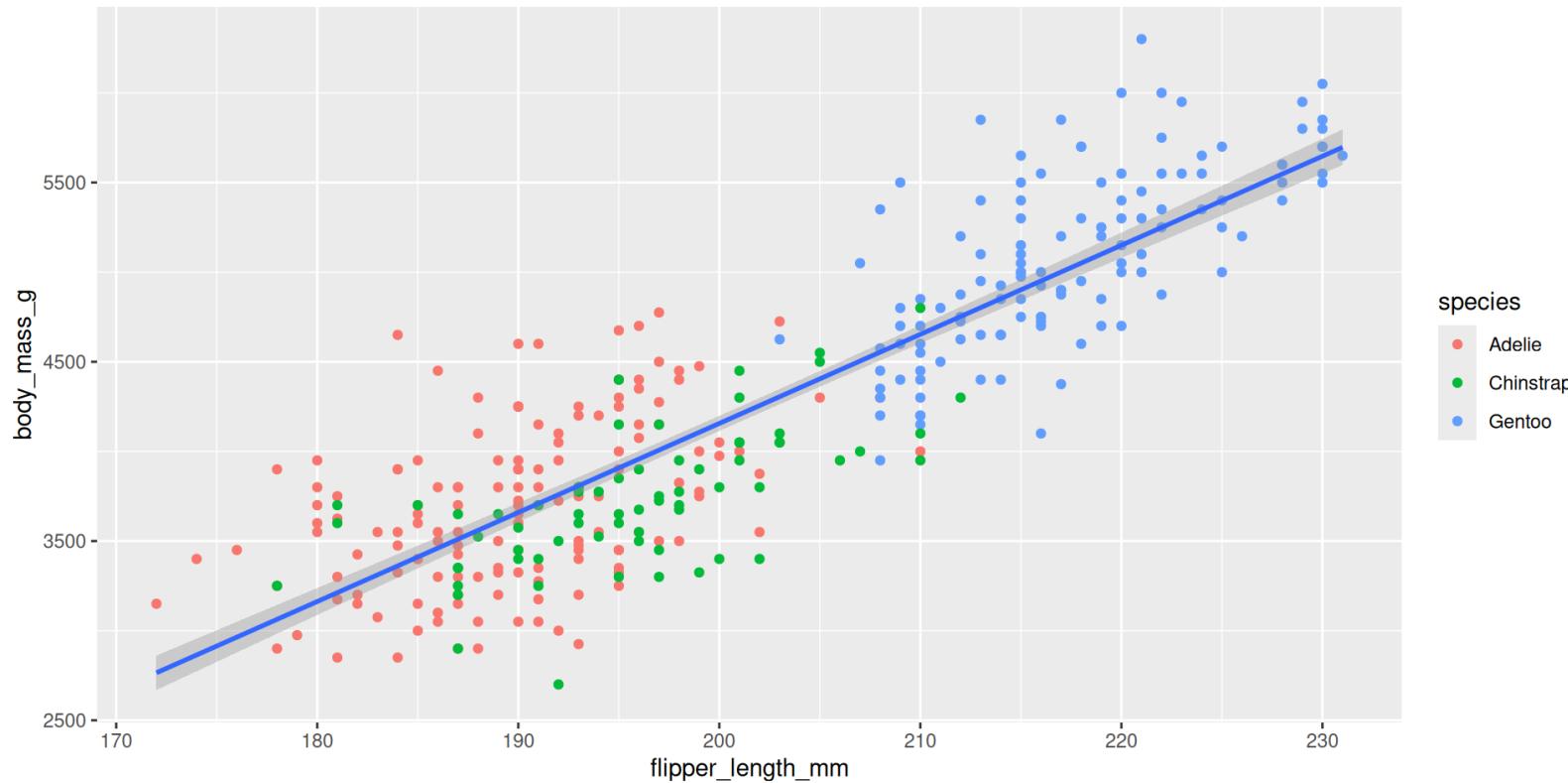


Aesthetics und Schichten

```

1 ggplot(
2   data = penguins,
3   mapping = aes(x = flipper_length_mm, y = body_mass_g)
4 ) +
5   geom_point(mapping = aes(colour = species)) +
6   geom_smooth(method = "lm")

```

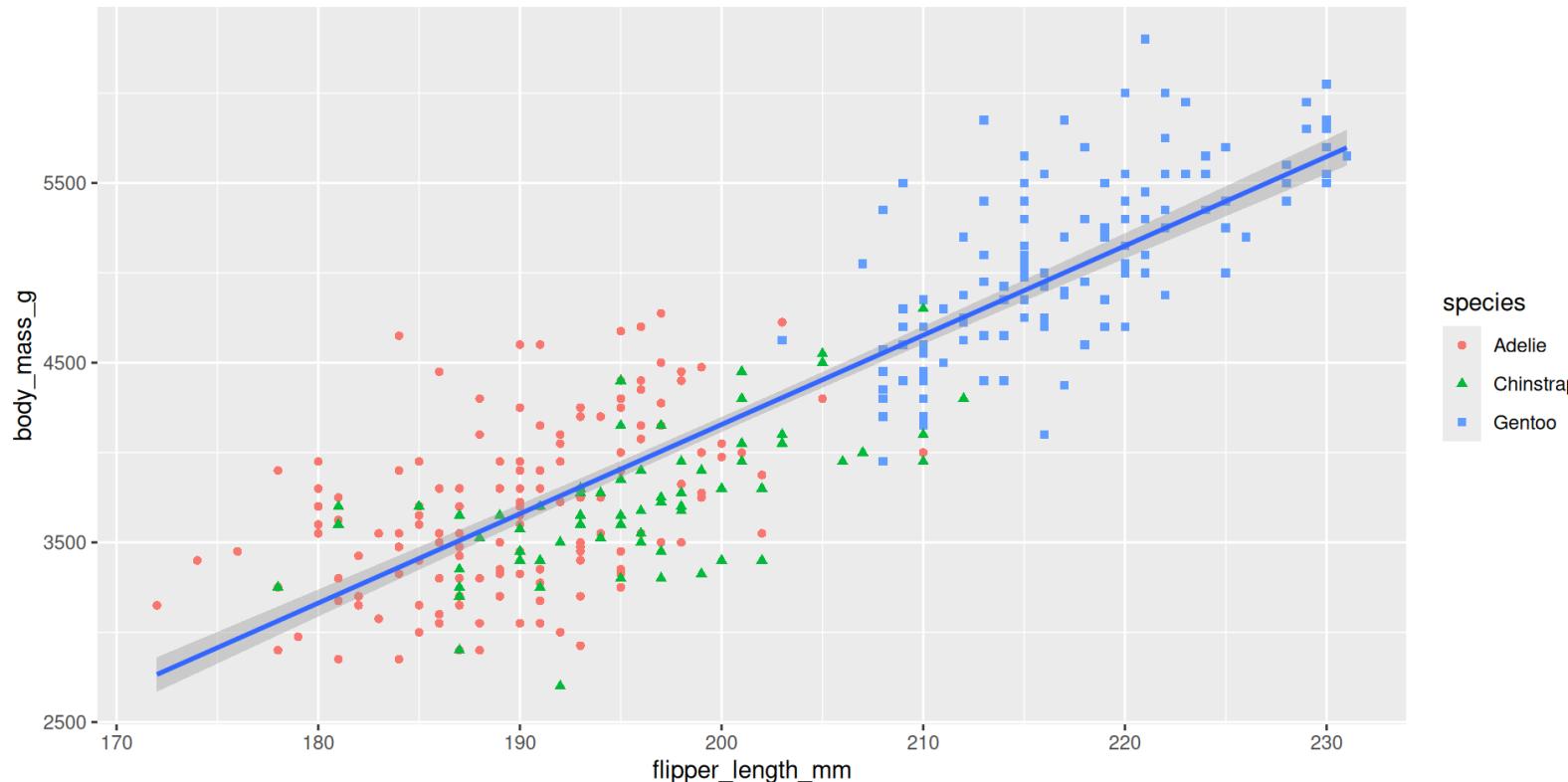


Aesthetics und Schichten

```

1 ggplot(
2   data = penguins,
3   mapping = aes(x = flipper_length_mm, y = body_mass_g)
4 ) +
5   geom_point(mapping = aes(colour = species, shape = species)) +
6   geom_smooth(method = "lm")

```



Aesthetics und Schichten

```
1 ggplot(  
2   data = penguins,  
3   mapping = aes(x = flipper_length_mm, y = body_mass_g)  
4 ) +  
5   geom_point(aes(colour = species, shape = species)) +  
6   geom_smooth(method = "lm") +  
7   labs(  
8     title = "Body mass and flipper length",  
9     subtitle = "Dimensions for Adelie, Chinstrap, and Gentoo Penguins",  
10    x = "Flipper length (mm)",  
11    y = "Body mass (g)",  
12    color = "Species",  
13    shape = "Species"  
14  ) +  
15  scale_colour_colorblind()
```

Aesthetics und Schichten

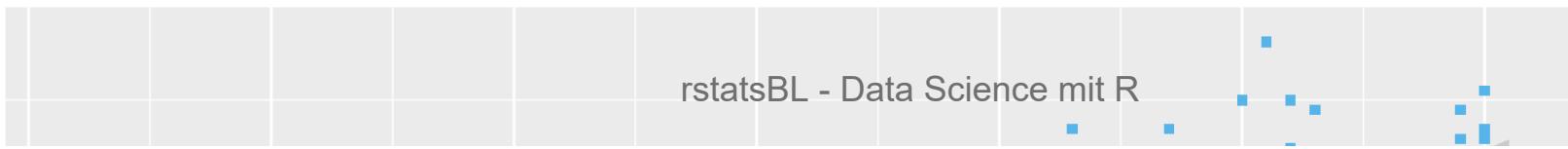
► Code

```

1 ggplot(
2   data = penguins,
3   mapping = aes(x = flipper_length_mm, y = body_mass_g)
4 ) +
5   geom_point(aes(colour = species, shape = species)) +
6   geom_smooth(method = "lm") +
7   labs(
8     title = "Body mass and flipper length",
9     subtitle = "Dimensions for Adelie, Chinstrap, and Gentoo Penguins",
10    x = "Flipper length (mm)",
11    y = "Body mass (g)",
12    color = "Species",
13    shape = "Species"
14  ) +
15  scale_colour_colorblind()

```

Body mass and flipper length
Dimensions for Adelie, Chinstrap, and Gentoo Penguins



Praktikum 01c

prak-01c-ggplot-scatter.qmd

20:00
rstatsBL - Data Science mit R

Break

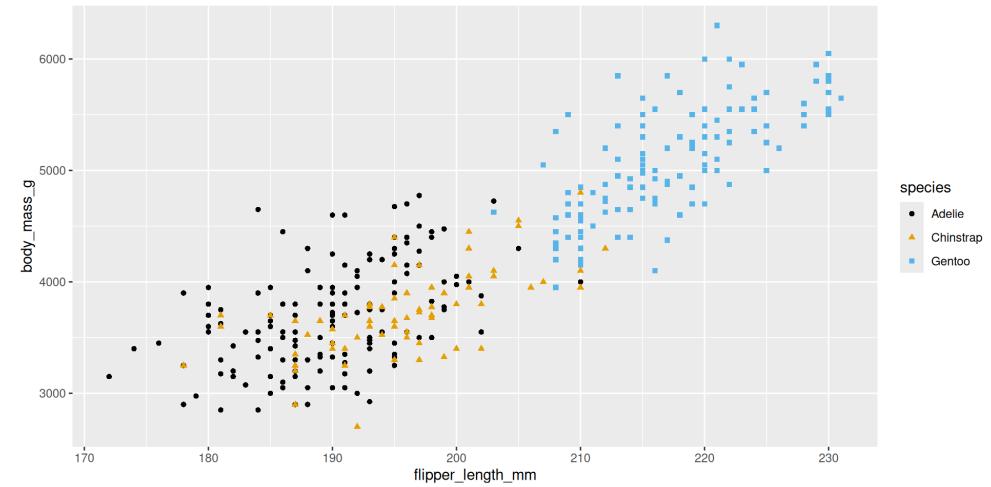


10:00

Aesthetics-Optionen

- colour 
- shape 
- size
- alpha (transparency)

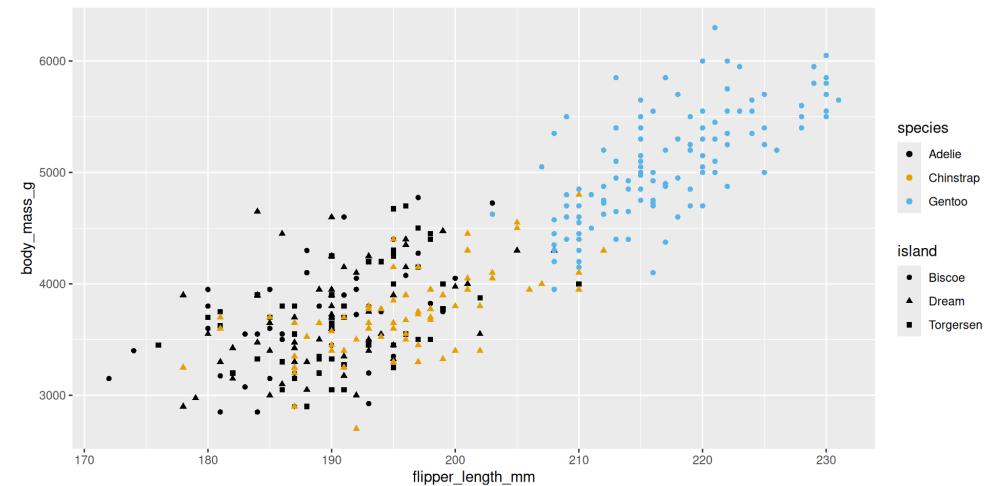
```
1 ggplot(  
2   penguins,  
3   aes(  
4     x = flipper_length_mm,  
5     y = body_mass_g,  
6     colour = species,  
7     shape = species  
8   )  
9 ) +  
10 geom_point() +  
11 scale_colour_colorblind()
```



```

1 ggplot(
2   penguins,
3   aes(
4     x = flipper_length_mm,
5     y = body_mass_g,
6     colour = species,
7     shape = island
8   )
9 ) +
10 geom_point() +
11 scale_colour_colorblind()

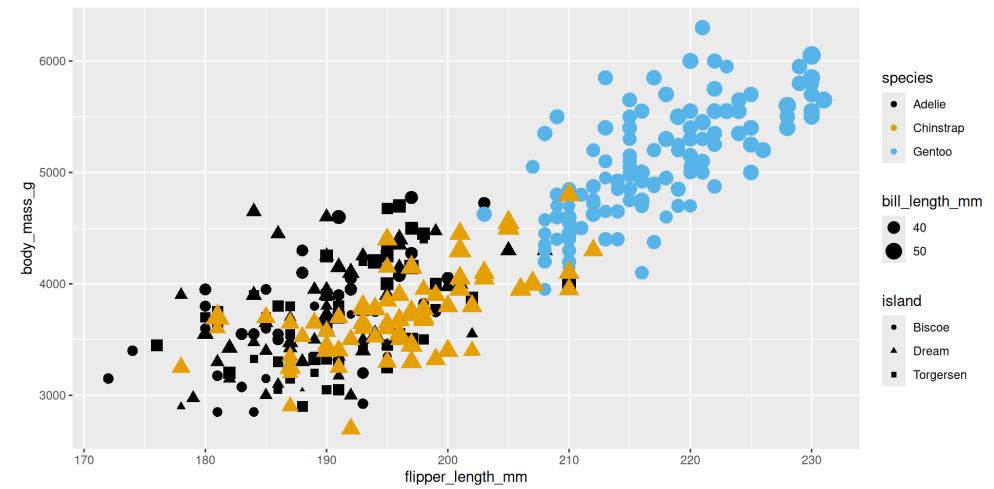
```



```

1 ggplot(
2   penguins,
3   aes(
4     x = flipper_length_mm,
5     y = body_mass_g,
6     colour = species,
7     shape = island,
8     size = bill_length_mm
9   )
10 ) +
11   geom_point() +
12   scale_colour_colorblind()

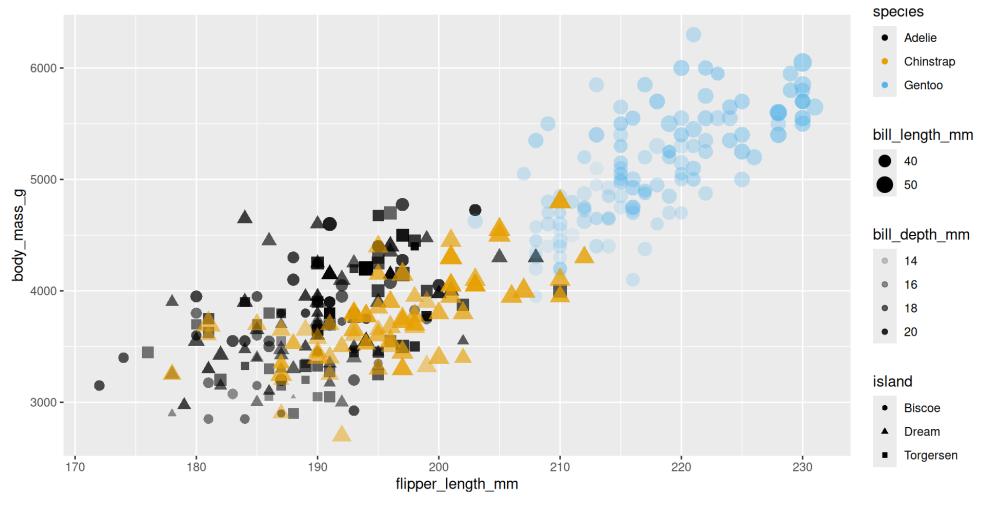
```



```

1 ggplot(
2   penguins,
3   aes(
4     x = flipper_length_mm,
5     y = body_mass_g,
6     colour = species,
7     shape = island,
8     size = bill_length_mm,
9     alpha = bill_depth_mm
10   )
11 ) +
12   geom_point() +
13   scale_colour_colorblind()

```

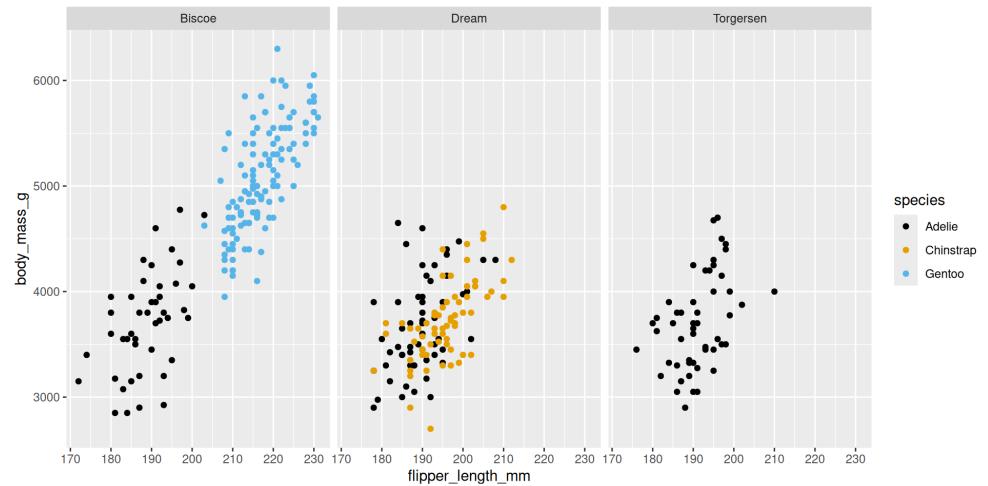


Faceting

```

1 ggplot(
2   penguins,
3   aes(
4     x = flipper_length_mm,
5     y = body_mass_g,
6     colour = species
7   )
8 ) +
9   geom_point() +
10  facet_wrap(~island) +
11  scale_colour_colorblind()

```

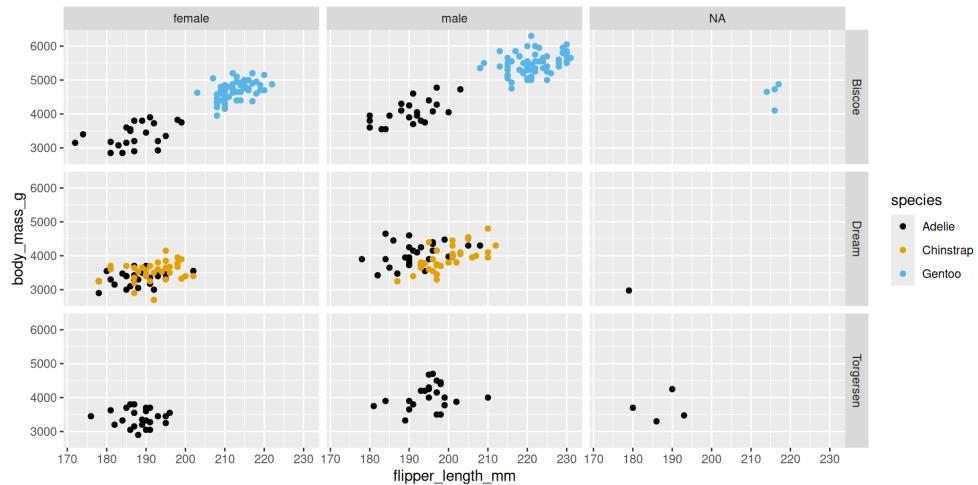


Faceting

```

1 ggplot(
2   penguins,
3   aes(
4     x = flipper_length_mm,
5     y = body_mass_g,
6     colour = species
7   )
8 ) +
9   geom_point() +
10  facet_grid(island ~ sex) +
11  scale_colour_colorblind()

```



Mapping vs. Setting ?

```

1 ggplot(
2   penguins,
3   aes(
4     x = flipper_length_mm,
5     y = body_mass_g,
6     size = bill_depth_mm,
7     alpha = bill_length_mm
8   )
9 ) +
10 geom_point()

```

```

1 ggplot(
2   penguins,
3   aes(
4     x = flipper_length_mm,
5     y = body_mass_g
6   )
7 ) +
8 geom_point(size = 4, alpha = 0.2)

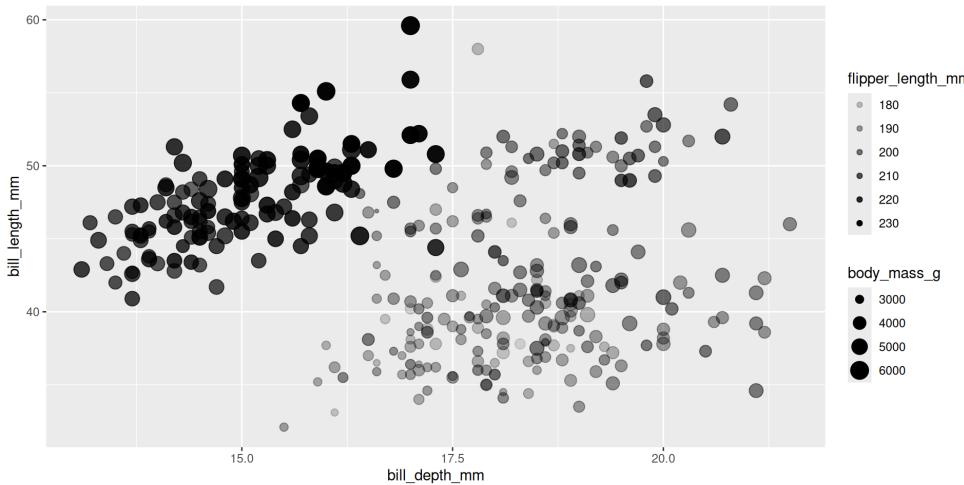
```

Mapping vs. Setting

```

1 ggplot(
2   penguins,
3   aes(
4     x = bill_depth_mm,
5     y = bill_length_mm,
6     size = body_mass_g,
7     alpha = flipper_length_mm
8   )
9 ) +
10 geom_point()

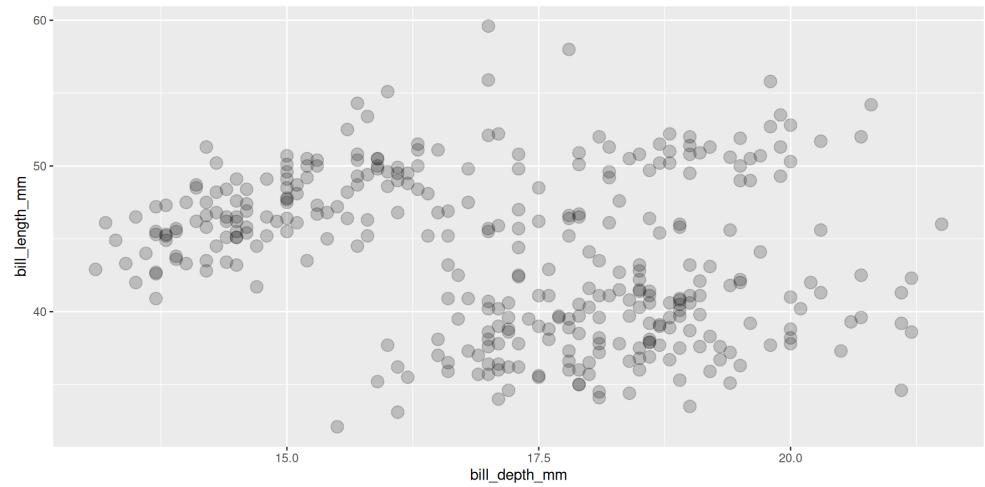
```



```

1 ggplot(
2   penguins,
3   aes(
4     x = bill_depth_mm,
5     y = bill_length_mm
6   )
7 ) +
8 geom_point(size = 4, alpha = 0.2)

```

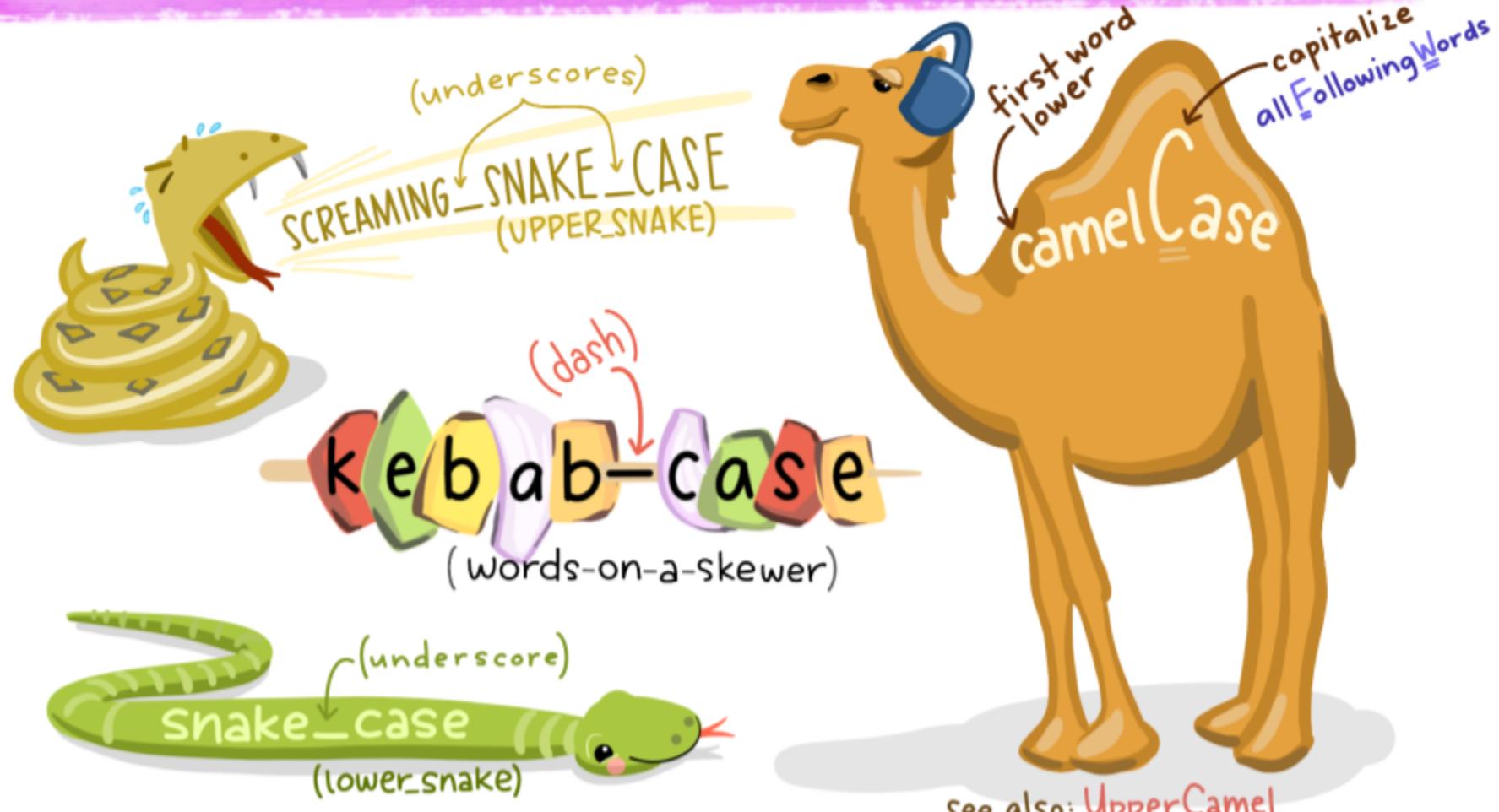


Workflow: Rechtschreibung

Namen

```
1 # Good
2 i_use_snake_case
3
4 # acceptable
5 otherPeopleUseCamelCase
6 some.people.use.periods
7
8 # Bad
9 And_aFew.People_Are.FREEspirits
```

in that case...



see also: UpperCamel
aka PascalCase

@allison_horst

in that case...



⚠ Case matters...

```
1 welcome_to_r <- "Welcome to R"  
1 welcome_to_R  
2 #> Error: object 'welcome_to_R' not found
```

... and so does punctuation!

Hausaufgabe

prak-01d-style.qmd

R for Data Science

- Das Buch für den Kurs
- Kostenfrei [Online](#)
- Tidyverse-Philosophie

Danke! 

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

Access slides as [PDF](#).

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