

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

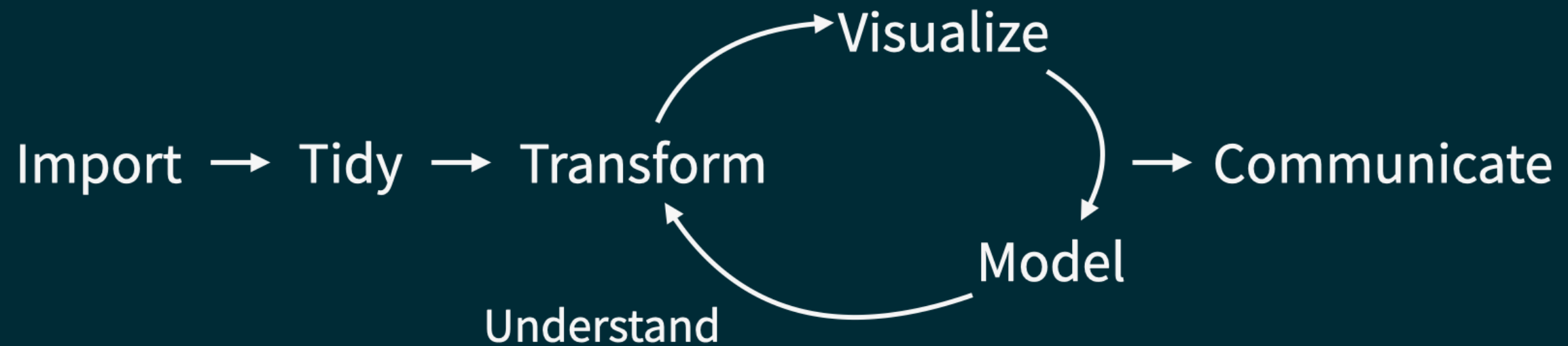
1. die sechs Elemente des *Data Science Lifecycles* aufzählen
2. *aesthetic mappings* für Datenvisualisierung im `{ggplot2}` R Package identifizieren
3. vier Komponenten einer Quarto-Datei identifizieren

Data Science Lifecycle

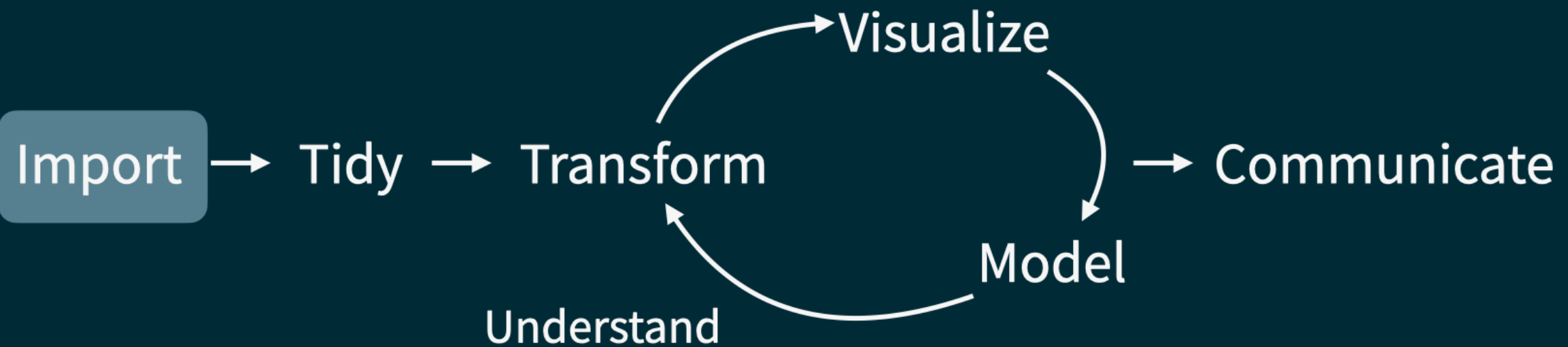
Reproduzierbare Datenanalyse



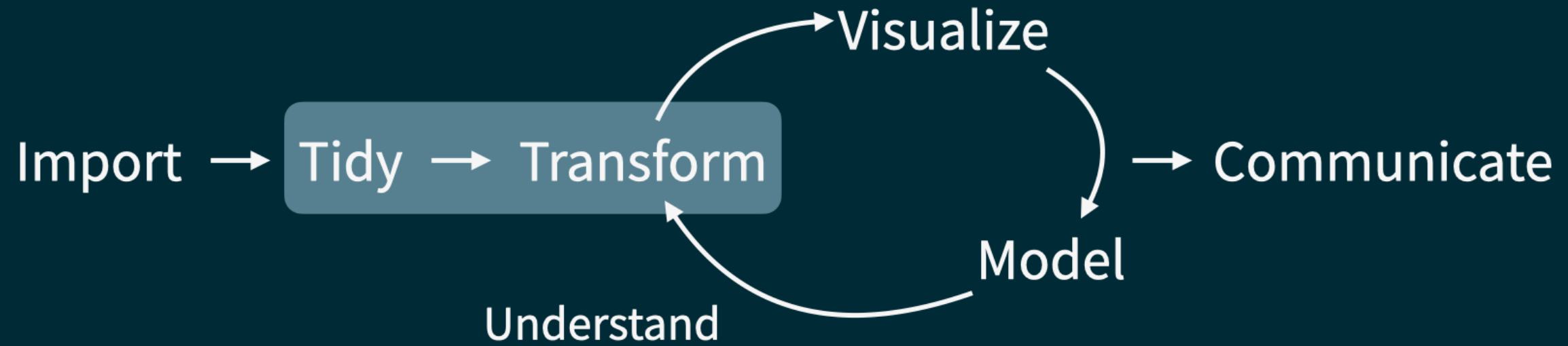
- Tabellen und Grafiken aus Daten und Code reproduzieren?
- Code in anderen Skripten verwenden?
- Stimmt meine Umgebung mit der meiner Kollegen/innen überein?



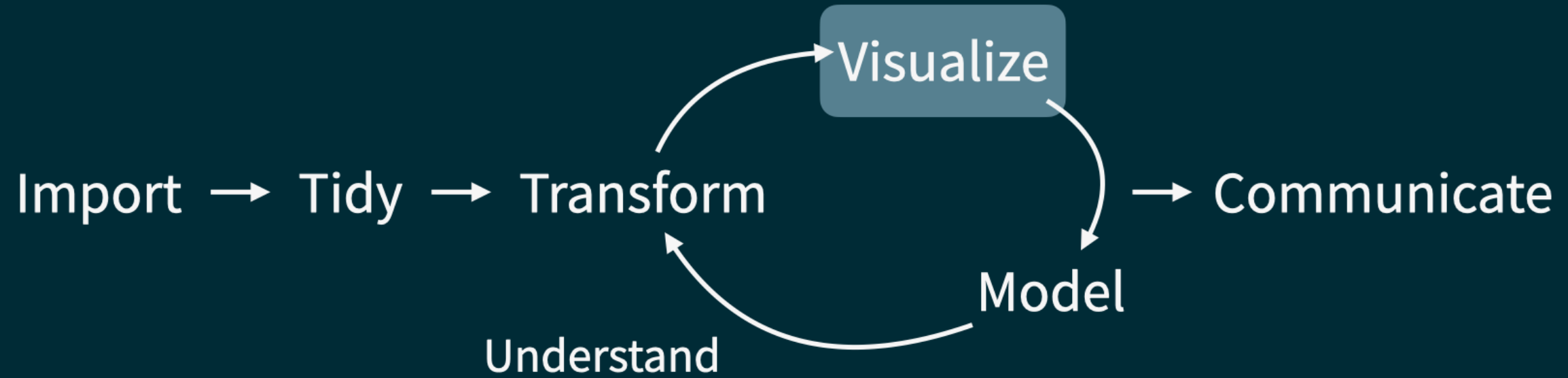
Program



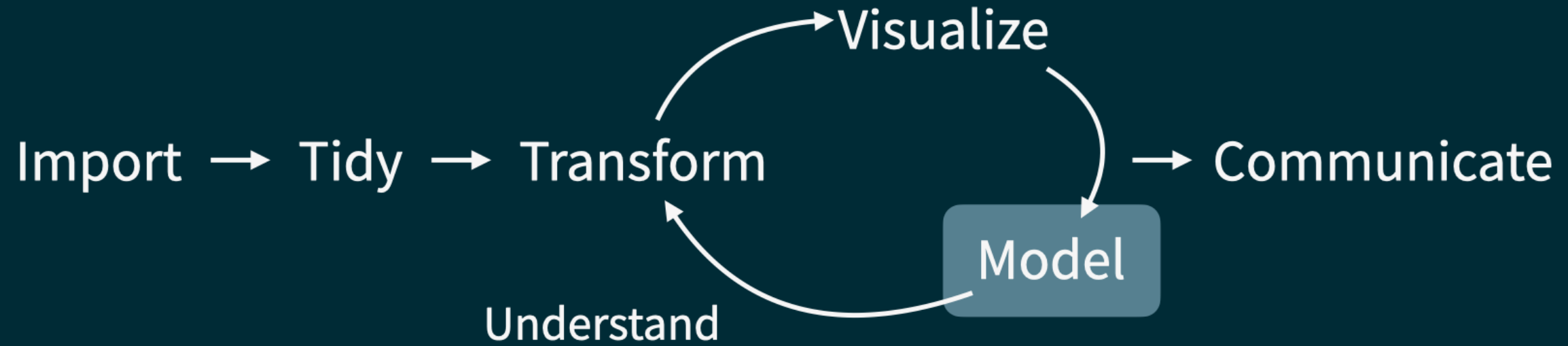
Program



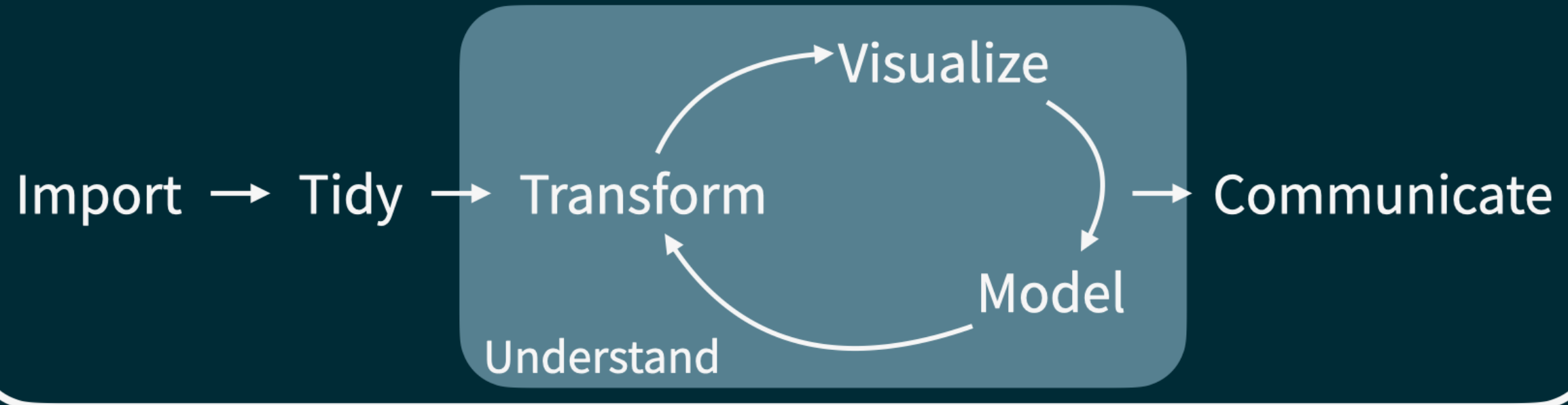
Program



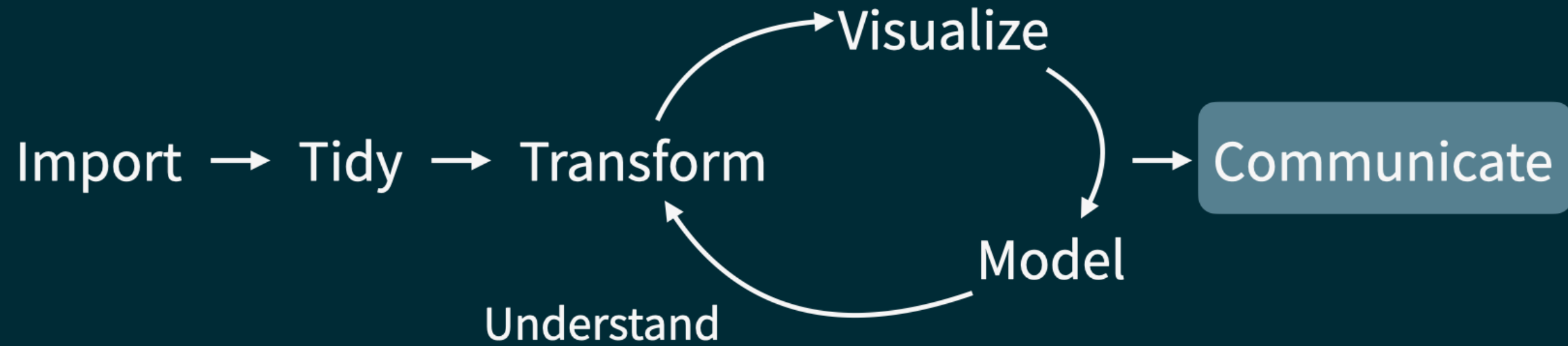
Program



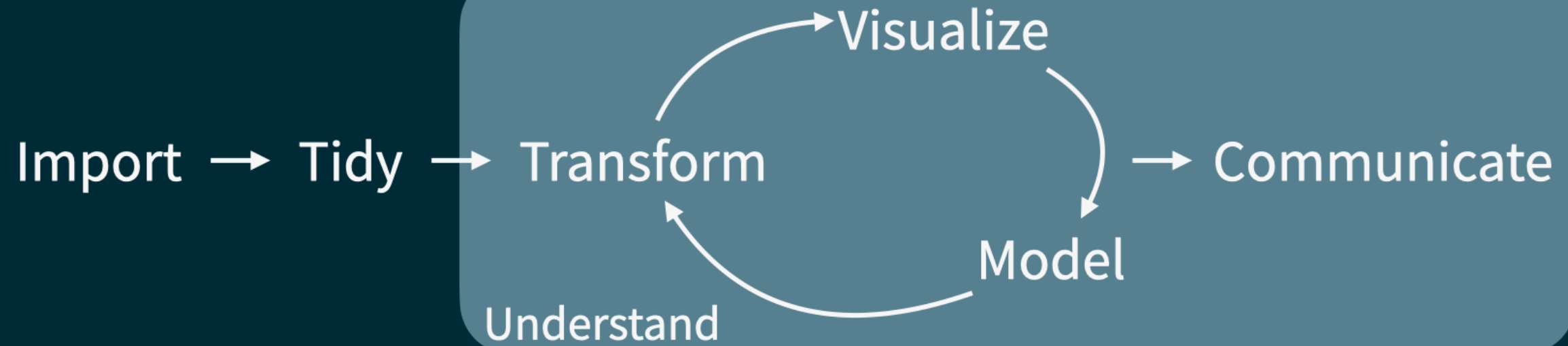
Program



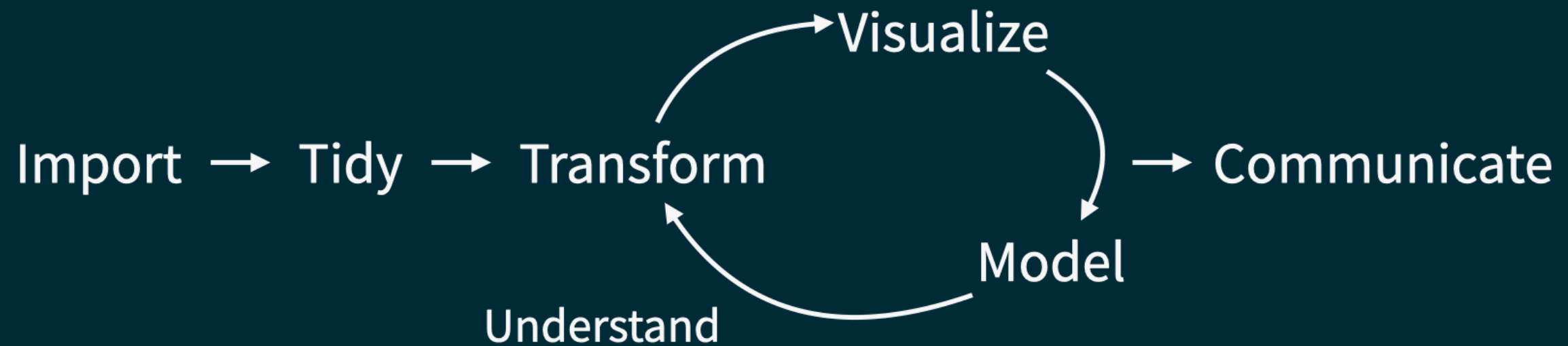
Program



Program



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

File Home Insert Page Layout Formulas Data Review View Developer Tell me what you want to do...

Cut Copy Paste Format Painter Clipboard

Calibri 11 A A Font

Wrap Text Merge & Center Alignment

General Number

Conditional Formatting Table Styles

Normal Bad Good Neutral Calculation Check Cell

Insert

F29 MFH 8

	A	B	C	D	E	F	G	H	I
1	egid	gemeindenummer_bfs	gemeinde	egrid	grundstücksnummer	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

website - main - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

kgwr_2023

Filter Cols: 1 - 50

	egid	gemeindenummer_bfs	gemeindenname	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

R 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: > R_projects > rstatsBL_workshop > exercises

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

RStudio

The screenshot displays the RStudio interface with a Quarto document open. The menu bar at the top is highlighted with an orange box and labeled "Menü Leiste". The document content is as follows:

```

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
8 Frauen zu sehen sind, die über alles andere als einen Mann miteinander reden. Um den Test zu bestehen, muss
9 ein Film drei Grundvoraussetzungen erfüllen:
10
11 - Es gibt mindestens zwei Frauenrollen
12 - Sie sprechen miteinander
13 - Sie unterhalten sich über etwas anderes als einen Mann.
14
15 ![Art by Srravya \[CC0\], via Wikimedia Commons](img/BechdelTest-768x310.jpg)
16
17 In dieser Minianalyse arbeiten wir mit den Daten, die in der *FiveThirtyEight*-Geschichte mit dem Titel
18 ["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/)
19 benutzt wurden.
  
```

The right sidebar shows the Environment pane with the Global Environment and a file explorer at the bottom. The file explorer shows the following files:

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	20.8 KB	Jan 3, 2025, 3:27 PM

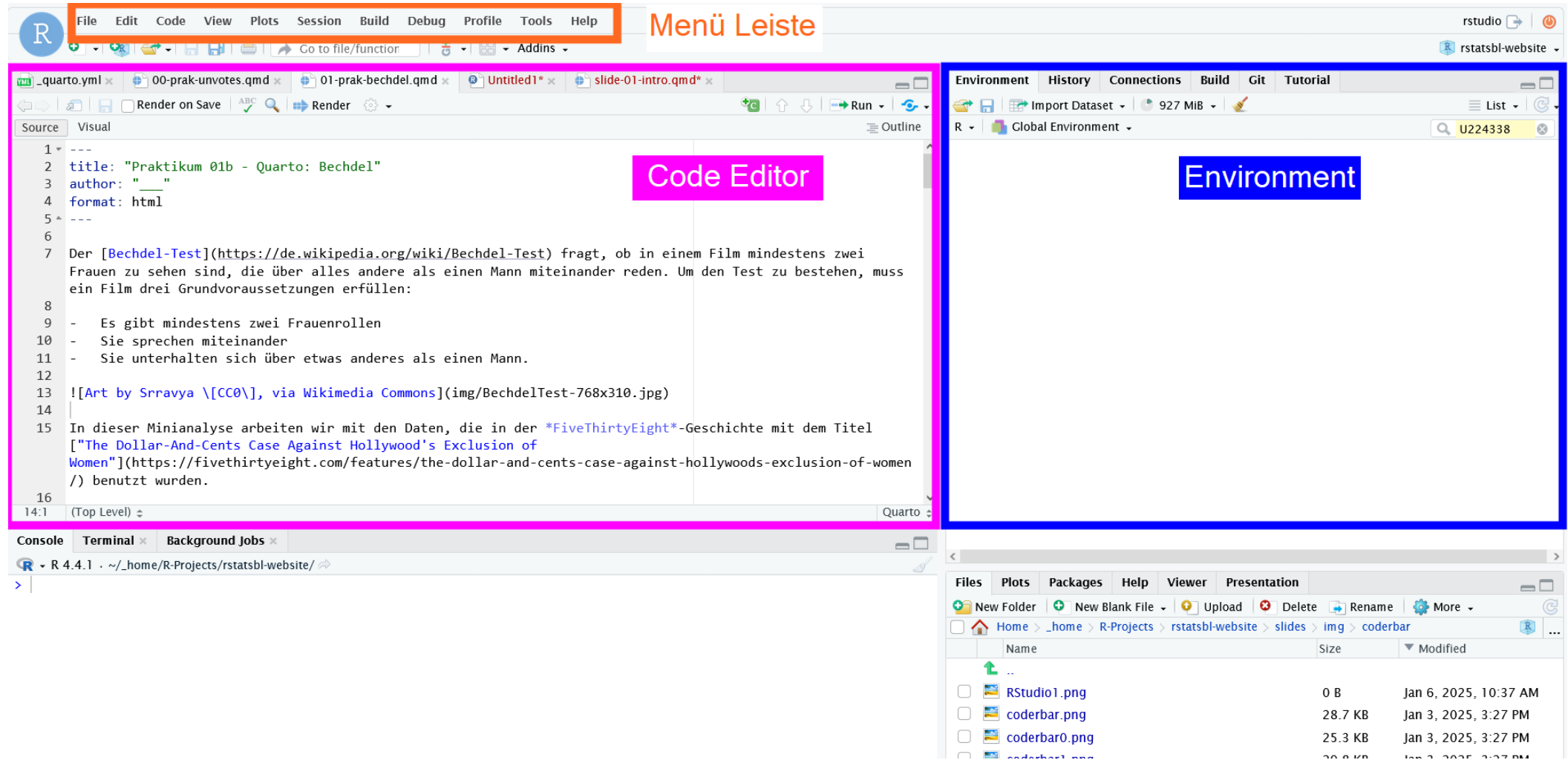
RStudio

The screenshot displays the RStudio environment with the following components and annotations:

- Menü Leiste (Menu Bar):** Highlighted in orange, containing File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help.
- Code Editor:** Highlighted in pink, showing a Quarto document with the following content:


```
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
8 Frauen zu sehen sind, die über alles andere als einen Mann miteinander reden. Um den Test zu bestehen, muss
9 ein Film drei Grundvoraussetzungen erfüllen:
10
11 - Es gibt mindestens zwei Frauenrollen
12 - Sie sprechen miteinander
13 - Sie unterhalten sich über etwas anderes als einen Mann.
14
15 ![Art by Srravya \[CC0\], via Wikimedia Commons](img/BechdelTest-768x310.jpg)
16
17 In dieser Minianalyse arbeiten wir mit den Daten, die in der *FiveThirtyEight*-Geschichte mit dem Titel
18 ["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/)
19 benutzt wurden.
```
- Environment Panel:** Shows the Global Environment with a memory usage of 927 MIB.
- Files Panel:** Displays the file explorer for the project, showing files like RStudio1.png, coderbar.png, and coderbar0.png.

RStudio



RStudio

The screenshot shows the RStudio interface with the following components highlighted:

- Menü Leiste** (Menu Bar): Located at the top, containing File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help.
- Code Editor**: The central area for writing code. It shows a Quarto document with a title "Praktikum 01b - Quarto: Bechdel", author information, and a paragraph about the Bechdel-Test. The code is written in R Markdown format.
- Environment**: The panel on the right side, showing the current environment (Global Environment) and memory usage (927 MIB).
- Console**: The panel at the bottom left, showing the R prompt and the current session path.

RStudio

The screenshot shows the RStudio interface with the following components highlighted and labeled:

- Menü Leiste** (Menu Bar): Highlighted in orange at the top, containing File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help.
- Code Editor**: Highlighted in pink on the left, showing a Quarto document with YAML metadata and R code for a Bechdel Test analysis.
- Environment**: Highlighted in blue on the top right, showing the Global Environment with a memory usage of 927 MIB.
- Console**: Highlighted in green at the bottom left, showing the R prompt and the current directory path.
- Datei Manager (Files)**: Highlighted in cyan at the bottom right, showing a file explorer view of the R-Projects directory.

The Code Editor contains the following content:

```

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
8 Frauen zu sehen sind, die über alles andere als einen Mann miteinander reden. Um den Test zu bestehen, muss
9 ein Film drei Grundvoraussetzungen erfüllen:
10
11 - Es gibt mindestens zwei Frauenrollen
12 - Sie sprechen miteinander
13 - Sie unterhalten sich über etwas anderes als einen Mann.
14
15 ![Art by Srravya \[CC0\], via Wikimedia Commons](img/BechdelTest-768x310.jpg)
16
17 In dieser Minianalyse arbeiten wir mit den Daten, die in der *FiveThirtyEight*-Geschichte mit dem Titel
18 ["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/)
19 benutzt wurden.
  
```

The Console shows the R prompt and the current directory path:

```

> R 4.4.1 ~/_home/R-Projects/rstatsbl-website/
  
```

The Datei Manager (Files) shows a list of files in the R-Projects directory:

Name	Size	Modified
..		
RStudio1.png	28.7 KB	Jan 3, 2025, 3:27 PM
coderbar.png	25.2 KB	Jan 3, 2025, 3:27 PM
coderbar0.png	25.2 KB	Jan 3, 2025, 3:27 PM
coderbar1.png	25.2 KB	Jan 3, 2025, 3:27 PM

RStudio und R-wesentliches

The screenshot displays the RStudio environment with the following components:

- Data/code viewer:** Shows a table of penguin data with columns: species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex, and year. The first 15 rows are visible.
- Environment:** Shows the Global Environment with a variable 'x' assigned the value 2.
- Console:** Contains R code snippets with annotations:
 - `> 2 + 2` followed by `[1] 4`
 - `> x <- 2` with a yellow arrow pointing to it and the text "Object assignment"
 - `> x * 3` followed by `[1] 6`
 - `> library(palmerpenguins)` with a blue arrow pointing to it and the text "Load package"
 - `> view(penguins)`
 - `> mean(penguins$flipper_length_mm)` followed by `[1] NA`
 - `> ?mean`
 - `> mean(penguins$flipper_length_mm, na.rm = TRUE)` with a yellow arrow pointing to it and the text "Function", followed by `[1] 200.9152`
- Help:** Displays the documentation for the `mean` function, including its description, usage, and arguments.

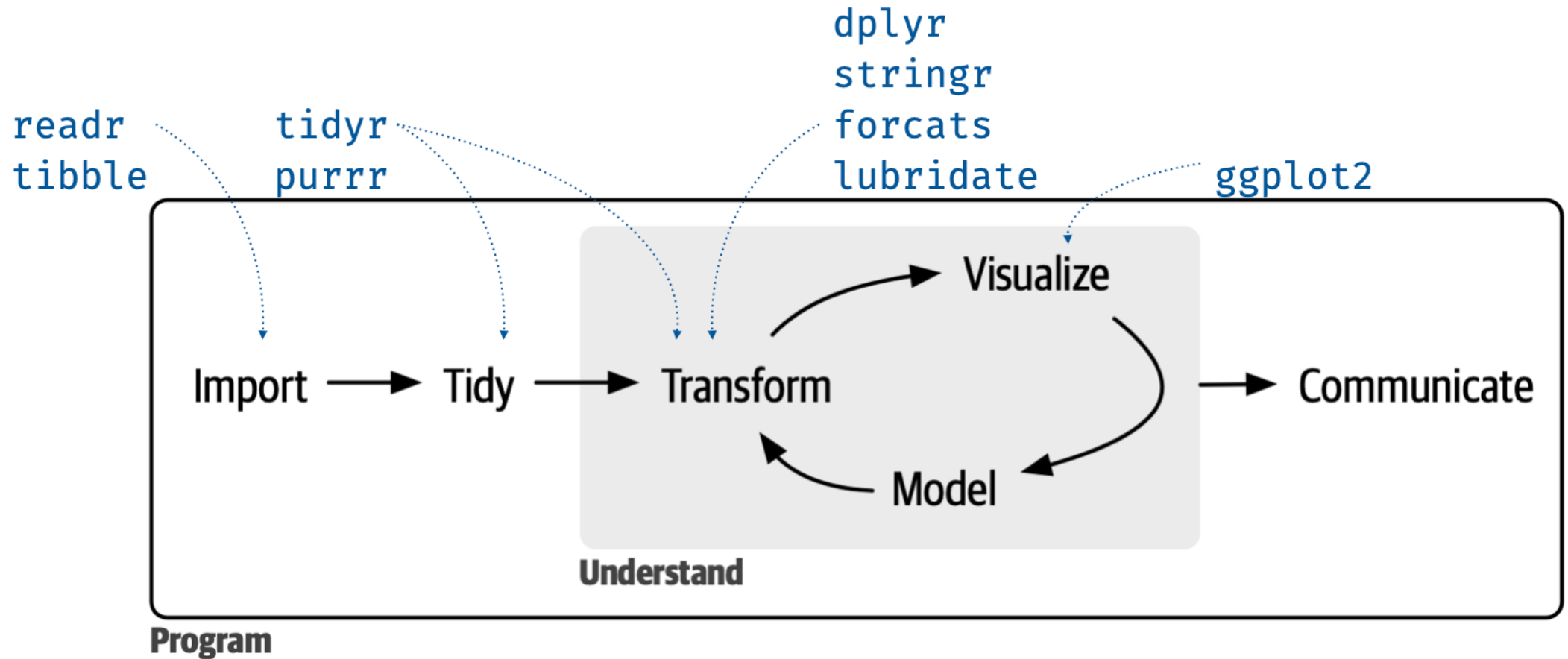
Break   

10:00

Tidyverse



Data Science Lifecycle





The screenshot displays the RStudio interface with a Quarto document open. The document is titled "unvotes-short-de.qmd" and is rendered as a web page titled "UN-Abstimmungen". The rendered page includes the author's name "Souad Guemghar" and the date "August 22, 2023". The document content is in German and discusses UN voting behavior. The rendered output shows the title, author, date, and an introduction section.

Source

```

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
17 hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich
18 ihr Abstimmungsverhalten bei bestimmten Themen?
19
20 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.
21
22 Wir werden die Packages 'tidyverse', 'scales' und 'gt' für die Datenaufbereitung und
23 -visualisierung und das Package 'unvotes' [voeten2013] für die Daten verwenden.
24
25 ```{r}
26 library(tidyverse)
27 library(scales)
28 library(gt)
29 library(unvotes)
30 ```
31
32 ## UN-Wahlverhalten
33
34 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das
35 Abstimmungsverhalten von Algerien, Frankreich und dem Vereinigten Königreich
36 im Laufe der Zeit bei einer Reihe von Themen verändert hat.
37
38 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen,
39 indem wir die Ländernamen in der unten stehenden filter()-Anweisung ersetzen. Die
40 genauen Namen der Länder findest du im Anhang (#appendix).
41
42 ```{r}
43 un_votes |>
44   inner_join(un_roll_calls, by = "rcid") |>
45   inner_join(un_roll_call_issues, by = "rcid") |>
46   filter(country %in% c("Algeria", "France", "United Kingdom")) |>
47   mutate(
48     year = year(date),

```

Table of contents

- Einleitung
- UN-Wahlverhalten
- Anhang

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),

```

The screenshot displays the RStudio interface with a Quarto document open. The document is titled "unvotes-short-de.qmd" and is rendered as a web page titled "UN-Abstimmungen". The rendered page includes the author's name "Souad Guemghar" and the date "August 22, 2023". The document content is in German and discusses UN voting behavior. The introduction section, titled "Einleitung", explains the purpose of the document and the packages used for data analysis: `tidyverse`, `scales`, `gt`, and `unvotes`. The document also includes a section titled "UN-Wahlverhalten" which discusses the data visualization and the countries of interest: **Algerien**, **Frankreich**, and **Vereinigtes Königreich**.

```

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
17 hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich
18 ihr Abstimmungsverhalten bei bestimmten Themen?
19
20 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.
21
22 Wir werden die Packages 'tidyverse', 'scales' und 'gt' für die Datenaufbereitung und
23 -visualisierung und das Package 'unvotes' [voeten2013] für die Daten verwenden.
24
25 ```{r}
26 library(tidyverse)
27 library(scales)
28 library(gt)
29 library(unvotes)
30 ```
31
32 ## UN-Wahlverhalten
33
34 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das
35 Abstimmungsverhalten von Algerien, Frankreich und dem Vereinigten Königreich
36 im Laufe der Zeit bei einer Reihe von Themen verändert hat.
37
38 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen,
39 indem wir die Ländernamen in der unten stehenden filter()-Anweisung ersetzen. Die
40 genauen Namen der Länder findest du im Anhang (#appendix).
41
42 ```{r}
43 un_votes |>
44   inner_join(un_roll_calls, by = "rcid") |>
45   inner_join(un_roll_call_issues, by = "rcid") |>
46   filter(country %in% c("Algeria", "France", "United Kingdom")) |>
47   mutate(
48     year = year(date),

```

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 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
17 hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich
18 ihr Abstimmungsverhalten bei bestimmten Themen?
19
20 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.
21
22 Wir werden die Packages 'tidyverse', 'scales' und 'gt' für die Datenaufbereitung und
23 -visualisierung und das Package 'unvotes' [voeten2013] für die Daten verwenden.
24
25 ```{r}
26 library(tidyverse)
27 library(scales)
28 library(gt)
29 library(unvotes)
30 ```
31
32 ## UN-Wahlverhalten
33
34 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das
35 Abstimmungsverhalten von Algerien, Frankreich und dem Vereinigten Königreich
36 im Laufe der Zeit bei einer Reihe von Themen verändert hat.
37
38 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen,
39 indem wir die Ländernamen in der unten stehenden filter()-Anweisung ersetzen. Die
40 genauen Namen der Länder findest du im Anhang.
41
42 ```{r}
43 un_votes |>
44   inner_join(un_roll_calls, by = "rcid") |>
45   inner_join(un_roll_call_issues, by = "rcid") |>
46   filter(country %in% c("Algeria", "France", "United Kingdom")) |>
47   mutate(
48     year = year(date),

```

Console

Files Plots Packages Help Viewer Presentation

Table of contents

- Einleitung
- UN-Wahlverhalten
- Anhang

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),

```

4:20 UN-Abstimmungen : Quarto

Environment History Connections Build Git Tutorial

The screenshot displays the RStudio interface with a Quarto document open. The left pane shows the source code of 'unvotes-short-de.qmd', and the right pane shows the rendered HTML output.

Source Code (Left Pane):

```

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
17 hat sich ihr Abstimmungsverhalten im Laufe der Zeit entwickelt? Wie unterscheidet sich
18 ihr Abstimmungsverhalten bei bestimmten Themen?
19
20 Wir werden versuchen, diese Fragen durch Visualisierung zu beantworten.
21
22 Wir werden die Packages 'tidyverse', 'scales' und 'gt' für die Datenaufbereitung und
23 -visualisierung und das Package 'unvotes' [voeten2013] für die Daten verwenden.
24
25 ```{r}
26 library(tidyverse)
27 library(scales)
28 library(gt)
29 library(unvotes)
30 ```
31
32 ## UN-Wahlverhalten
33
34 Wir wollen eine Datenvisualisierung erstellen, die zeigt, wie sich das
35 Abstimmungsverhalten von Algerien, Frankreich und dem Vereinigten Königreich
36 im Laufe der Zeit bei einer Reihe von Themen verändert hat.
37
38 Wir können ganz einfach ein Diagramm für jede beliebige Länderkombination erstellen,
39 indem wir die Ländernamen in der unten stehenden filter()-Anweisung ersetzen. Die
40 genauen Namen der Länder findest du im Anhang (#appendix).
41
42 ```{r}
43 un_votes |>
44   inner_join(un_roll_calls, by = "rcid") |>
45   inner_join(un_roll_call_issues, by = "rcid") |>
46   filter(country %in% c("Algeria", "France", "United Kingdom")) |>
47   mutate(
48     year = year(date),

```

Rendered HTML (Right Pane):

Table of contents

- Einleitung
- UN-Wahlverhalten
- Anhang

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),

```

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

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

```
1 [1] 0.4472136
```

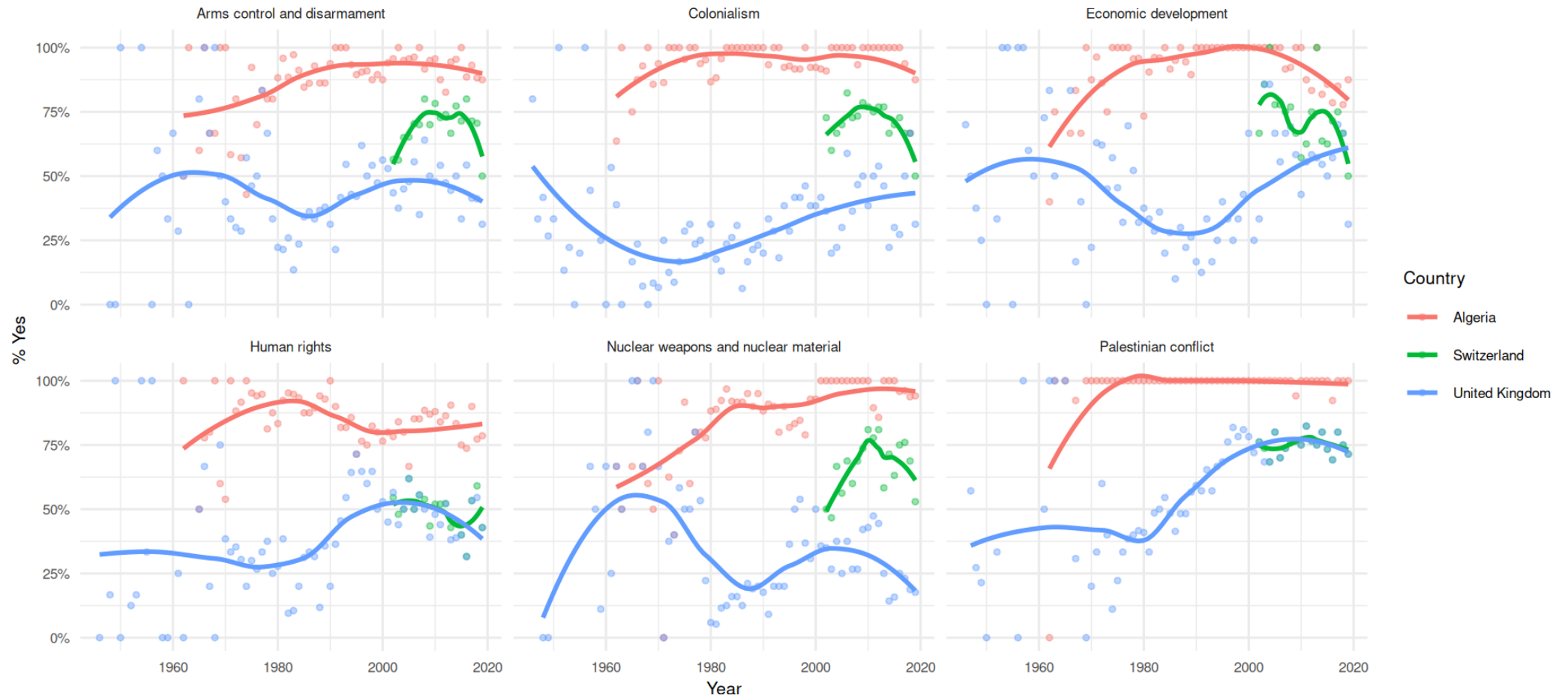
- Code und Text → ***inline code***

``r sqrt(1/5)`` → 0.4472136

Lasst uns eintauchen!



Percentage of 'Yes' votes in the UN General Assembly 1946 to 2019



un_votes x un_roll_calls x un_roll_call_issues x

Filter

rcid session importantvote date unproc amend para short

un_votes x un_roll_calls x un_roll_call_issues x

Filter

rcid short_name issue

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

Praktikum 01b: Quarto – Bechdel Test

prak-01b-bechdel.qmd

20:00

rstatsBL - Data Science mit R

Break   

10:00

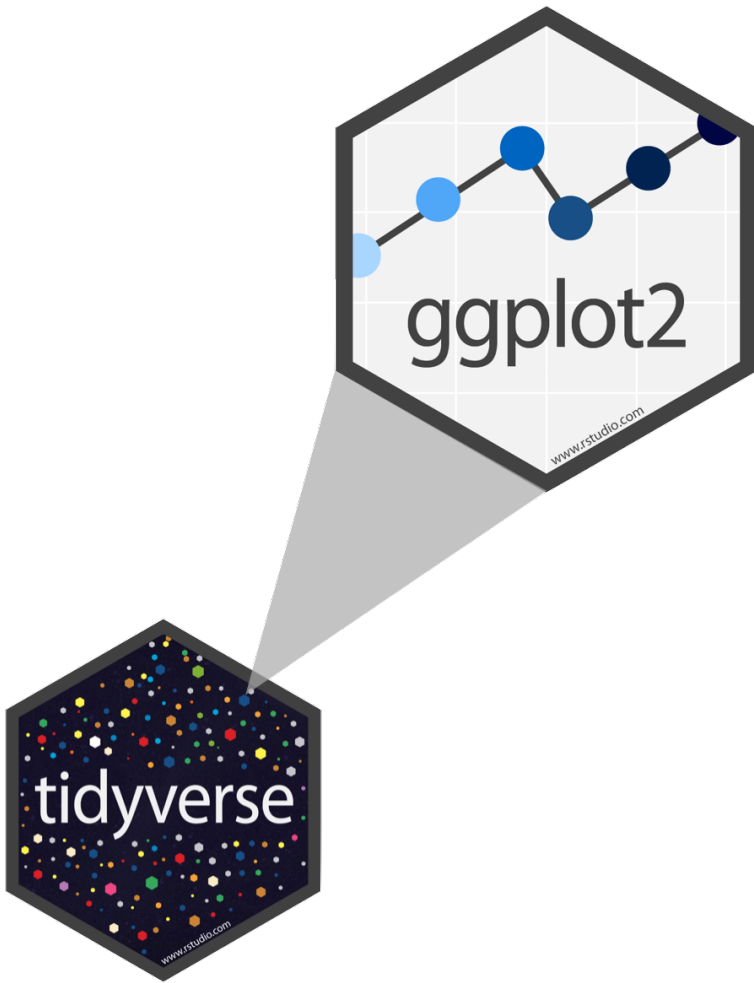
Visualisierung

R Package `ggplot2`

Streudiagramm (*Scatterplot*)

Verteilungen visualisieren

Beziehungen visualisieren



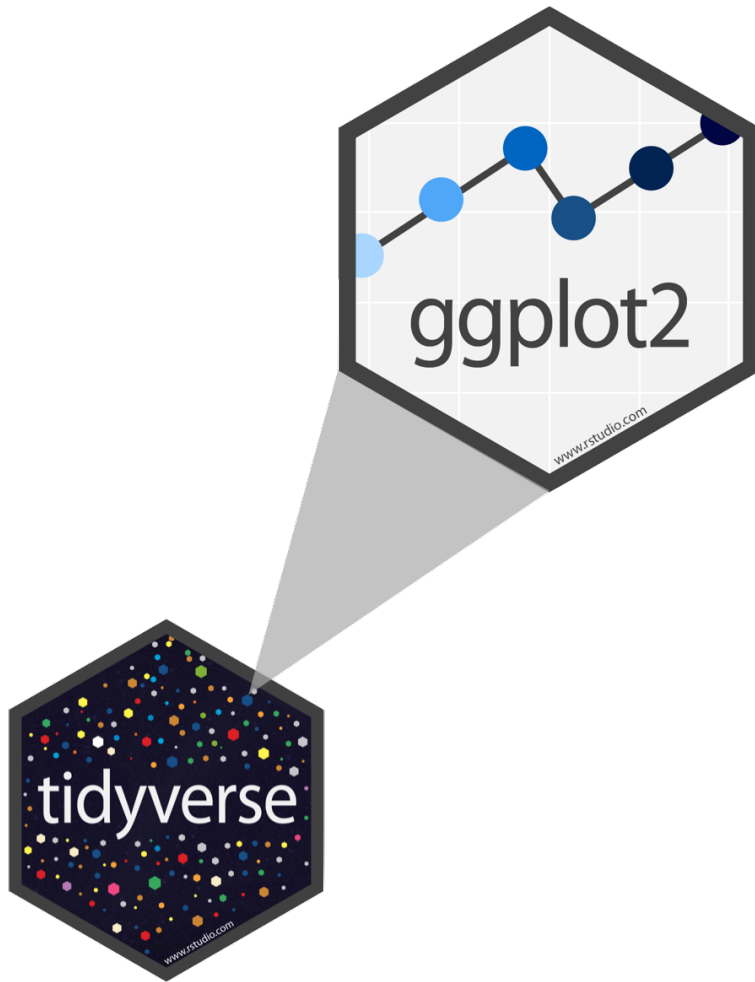
```
1 ggplot(data = [dataset],  
2         mapping = aes(x = [x-variable],  
3                       y = [y-variable])) +  
4   geom_xxx() +  
5   other options
```

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

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_horn

```

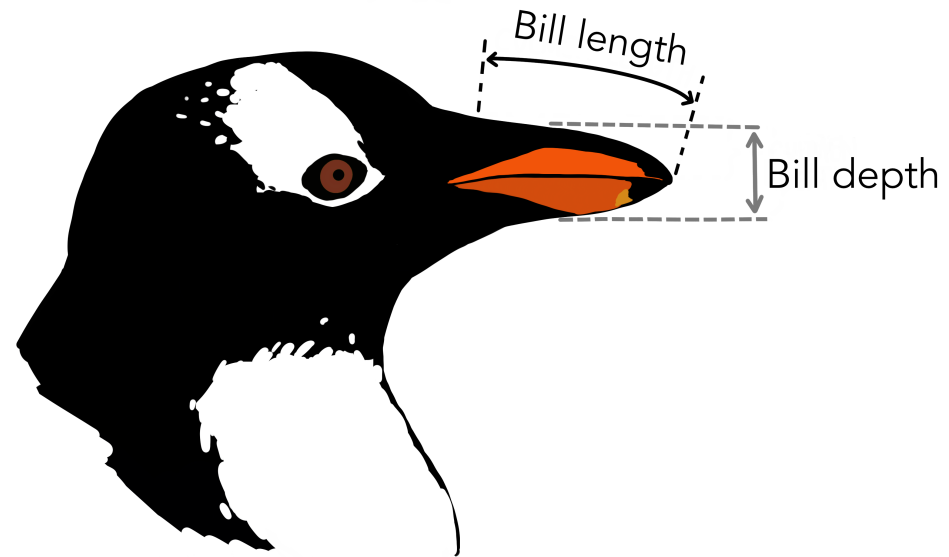
1 ```{r}
2 glimpse(penguins)
3 ```

```

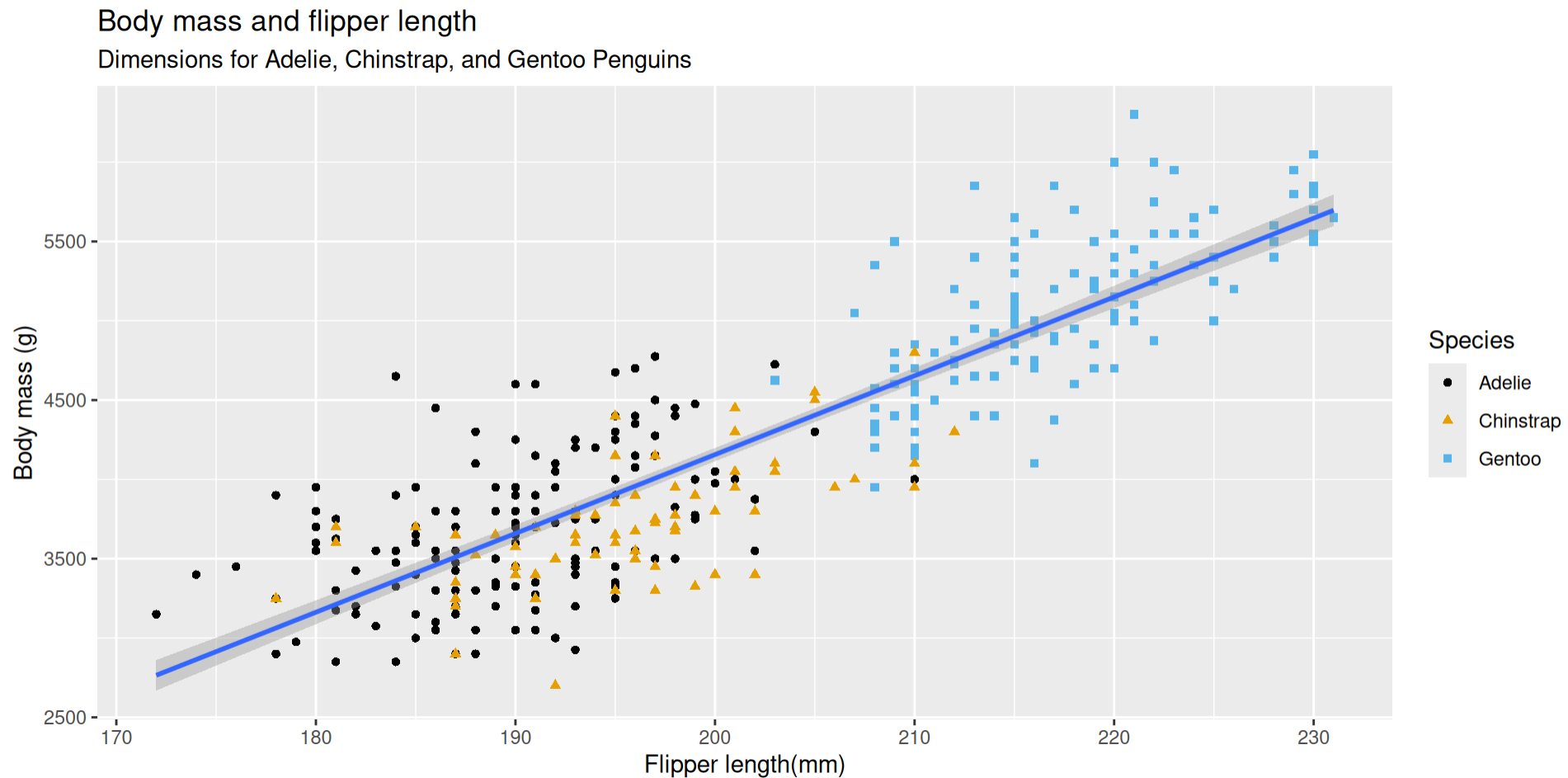
```

1 Rows: 344
2 Columns: 8
3 $ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel...
4 $ island       <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse...
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...

```

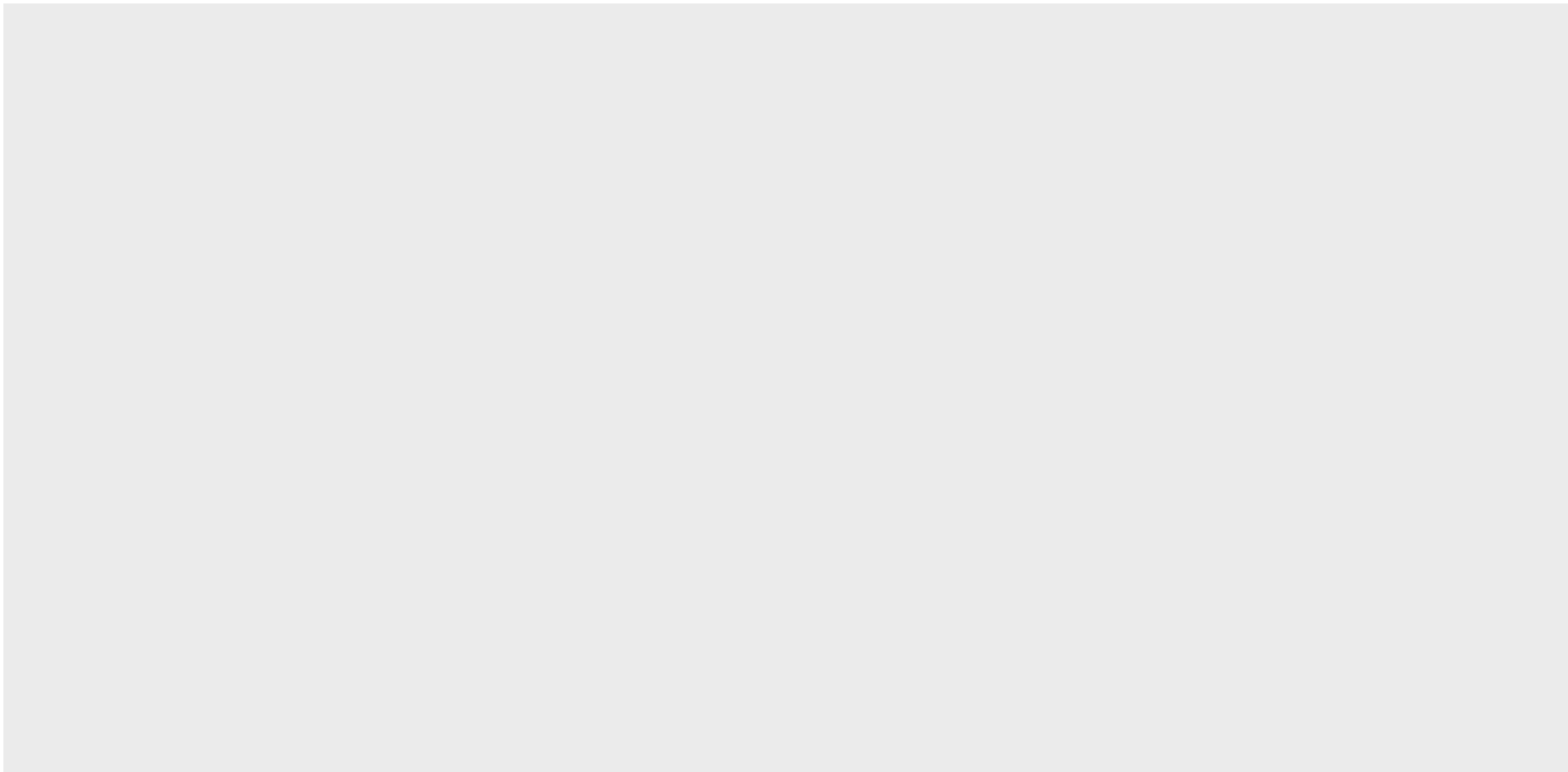


Unser Ziel

[Plot](#)[Code](#)

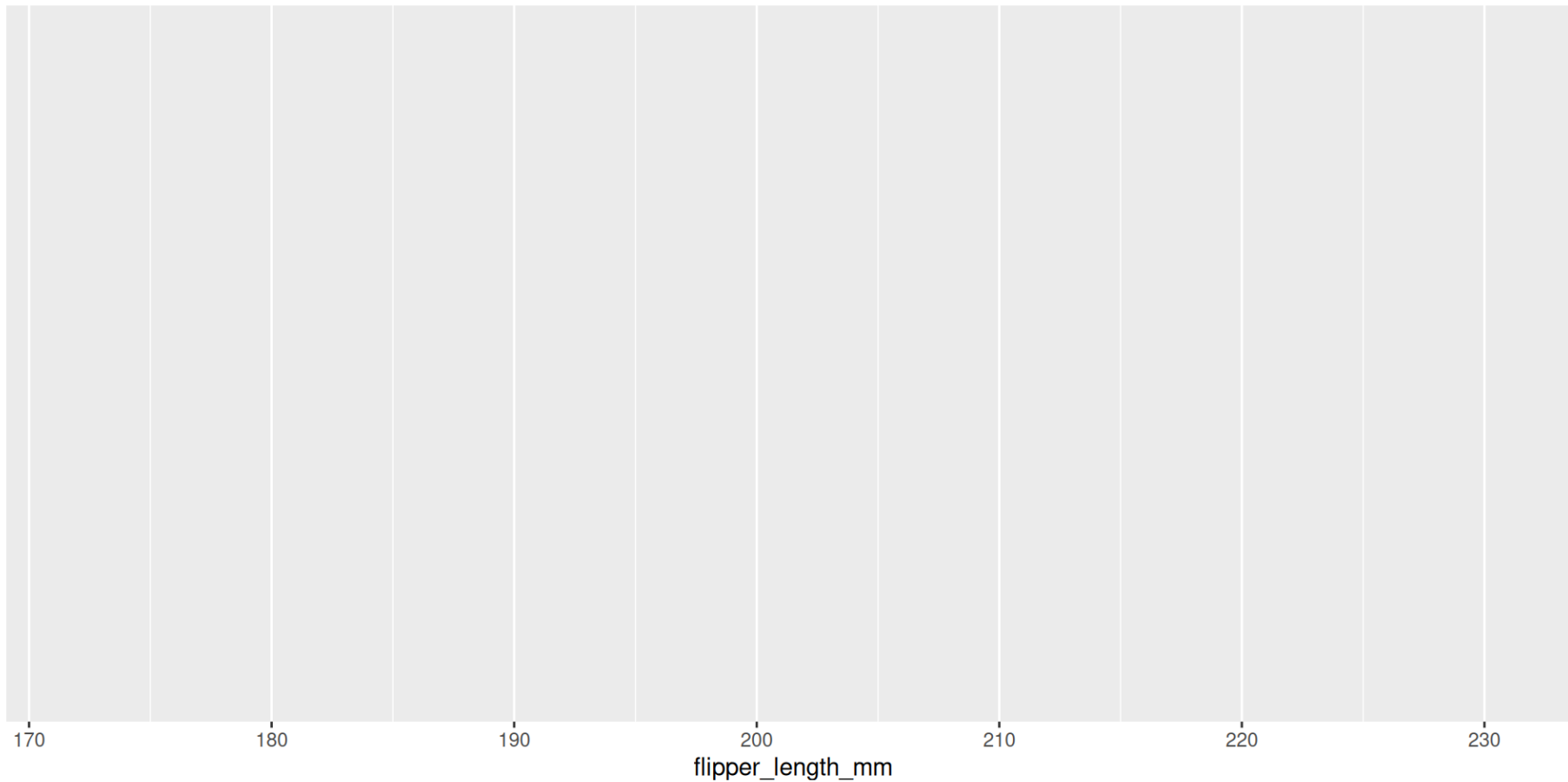
Plot Erstellen

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



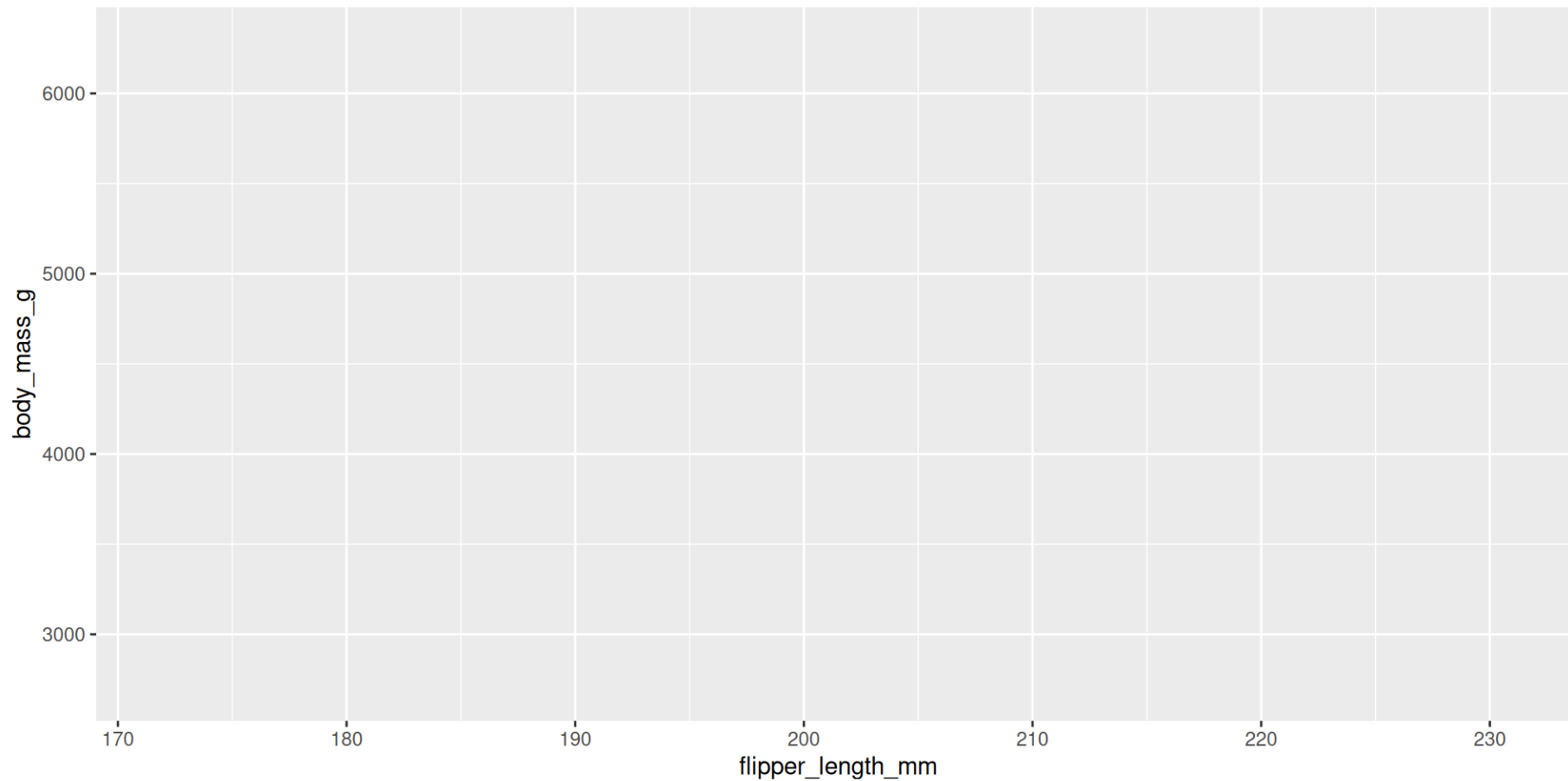
Plot Erstellen

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



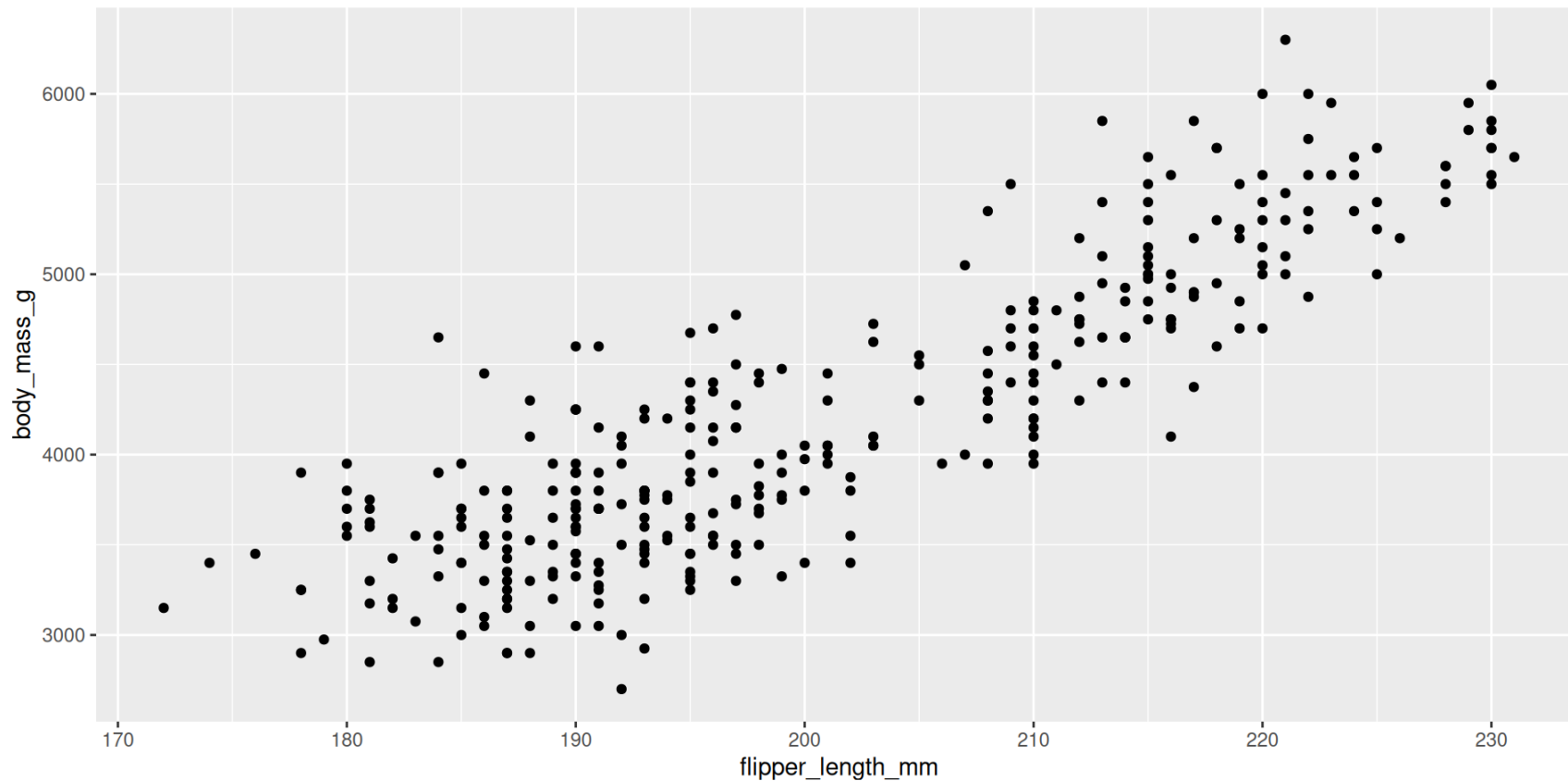
Plot Erstellen

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



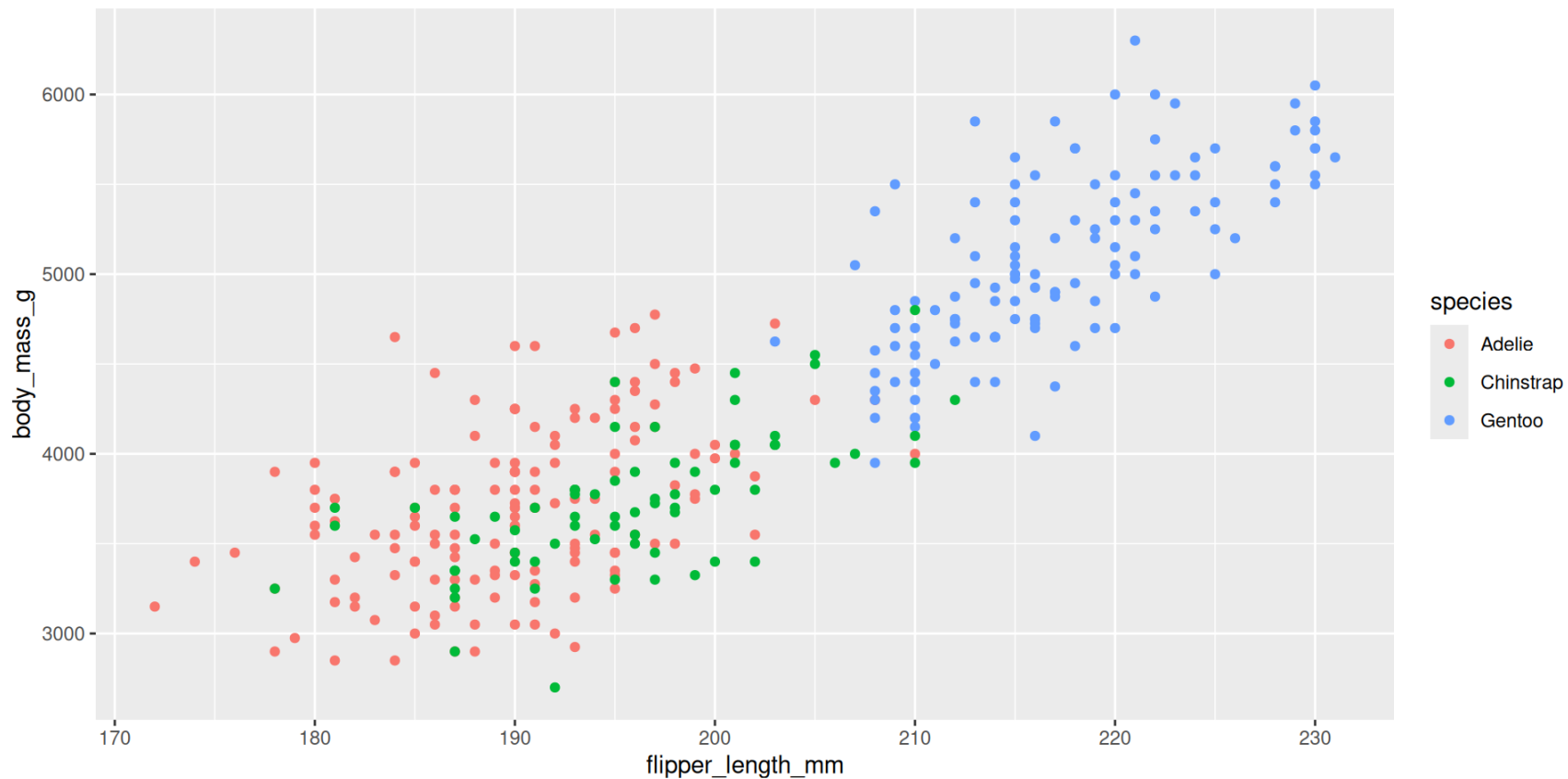
Plot Erstellen

```
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()
```

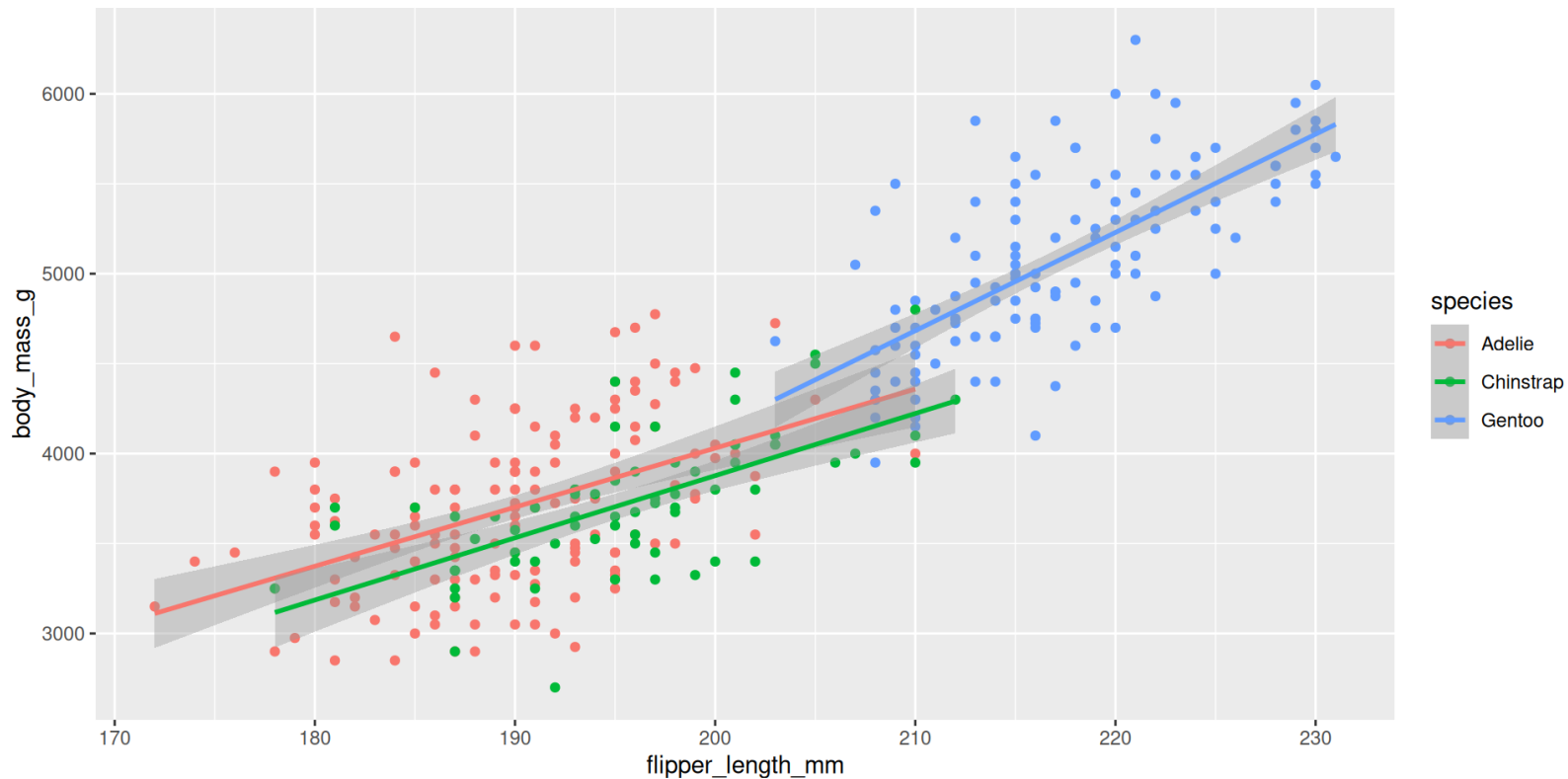


Aesthetics und Schichten

```

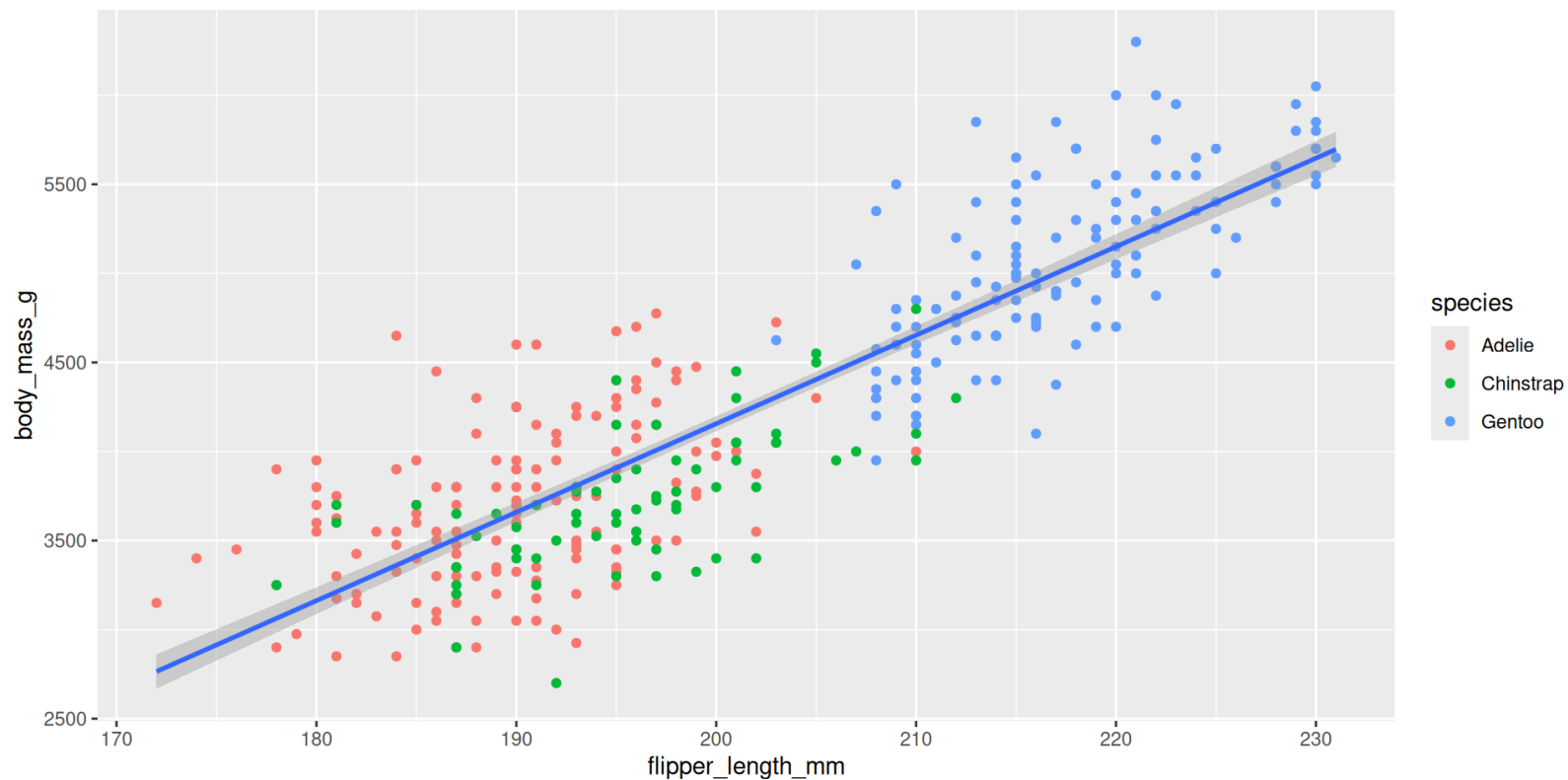
1 ggplot(
2   data = penguins,
3   mapping = aes(x = flipper_length_mm, y = body_mass_g, colour = species)
4 ) +
5   geom_point() +
6   geom_smooth(method = "lm")

```



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")
```

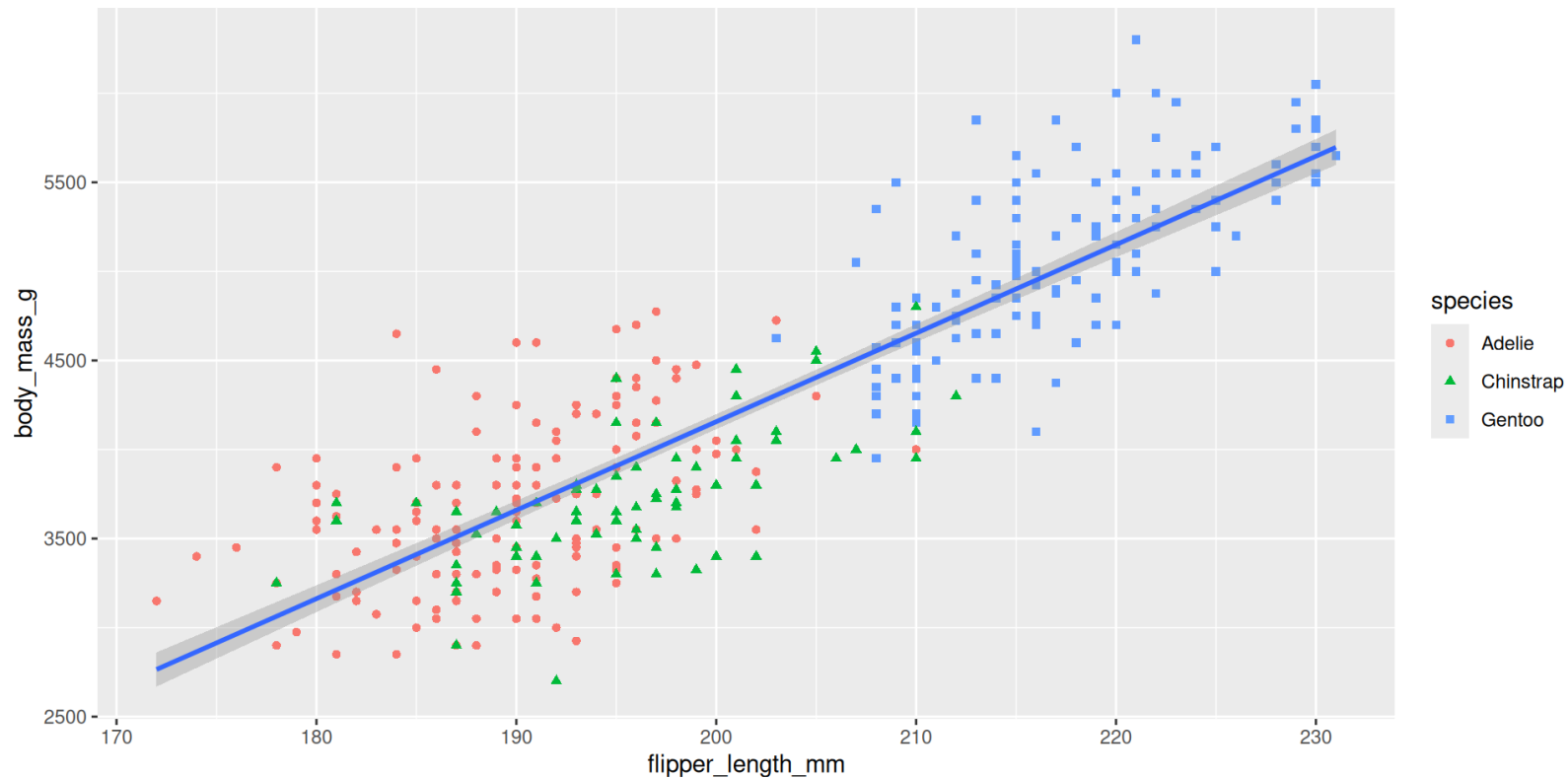


Aesthetics und Schichten

```

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

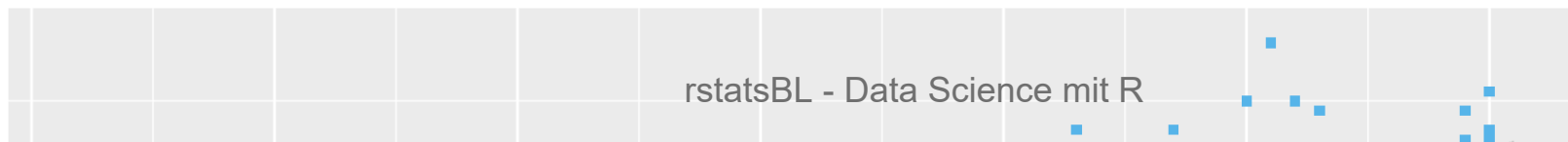
```

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

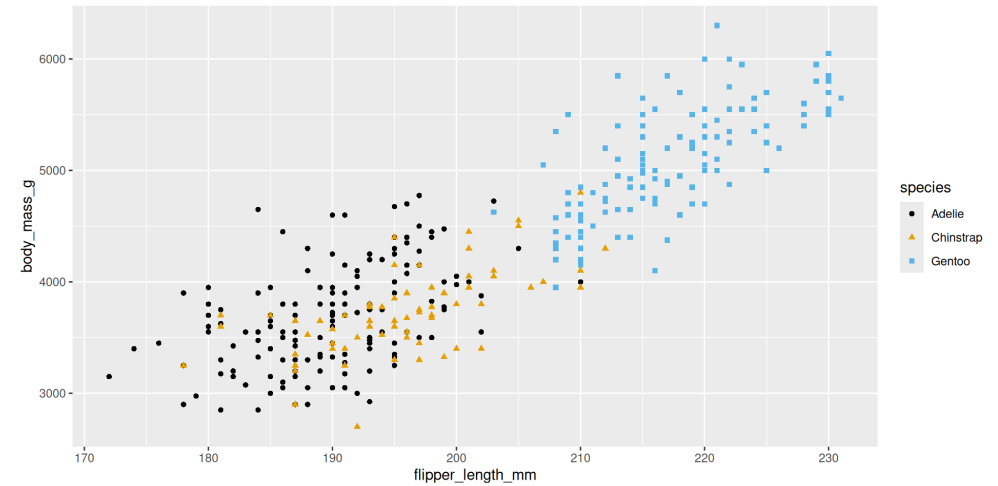
Aesthetics-Optionen

- 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()

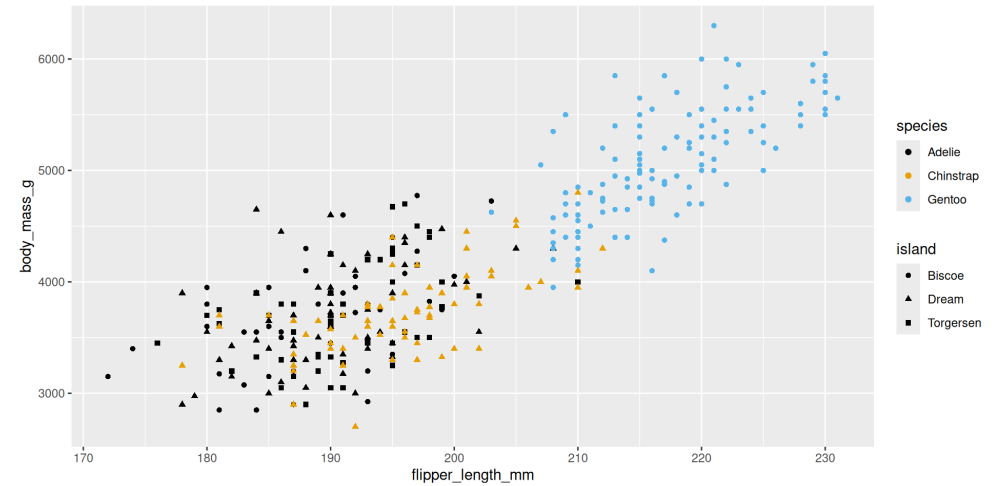
```



```

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()

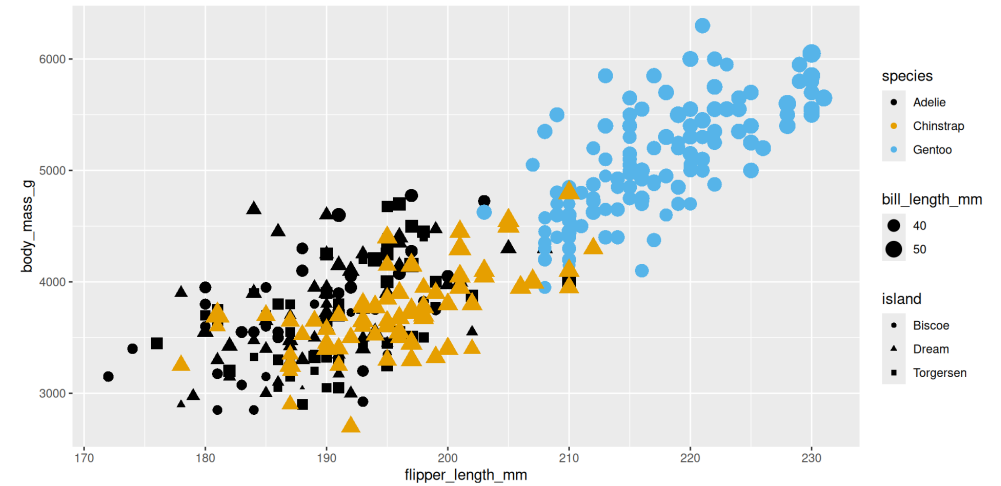
```



```

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()

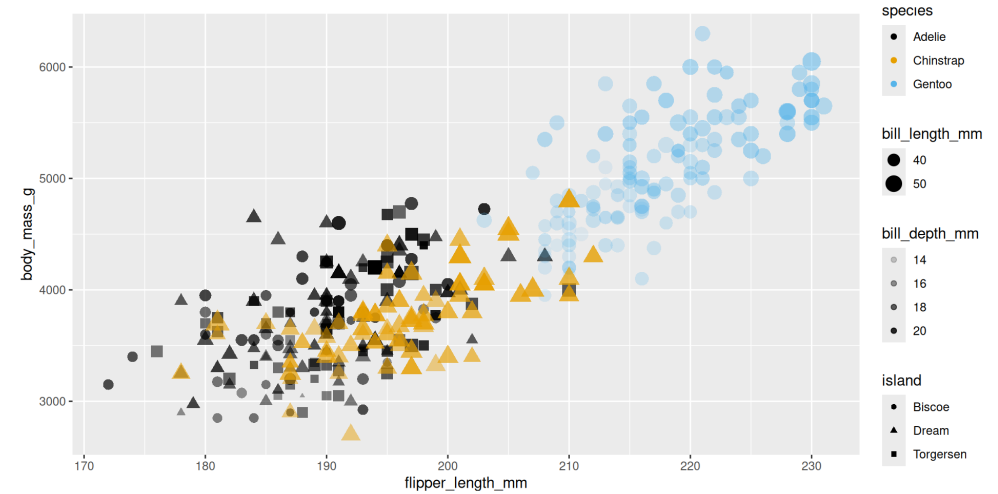
```



```

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()

```

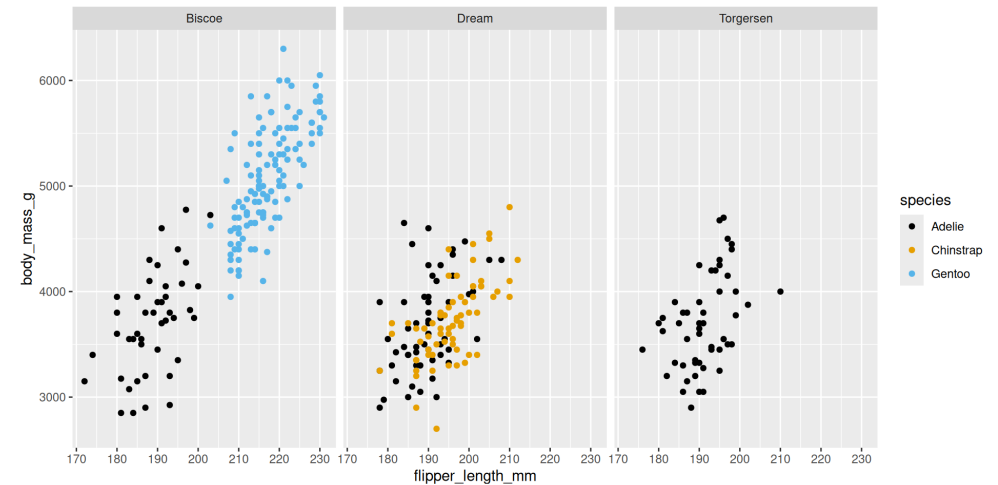


Faceting

```

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()

```

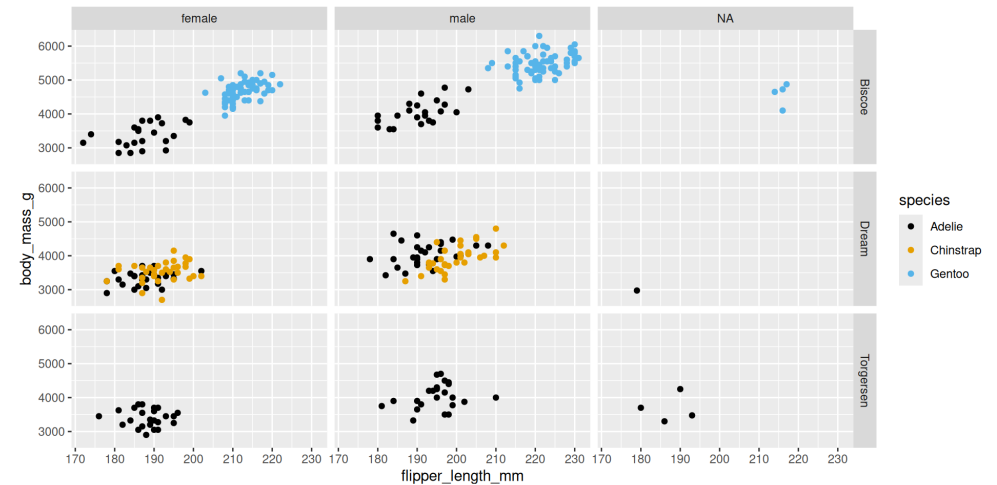


Faceting

```

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)
```

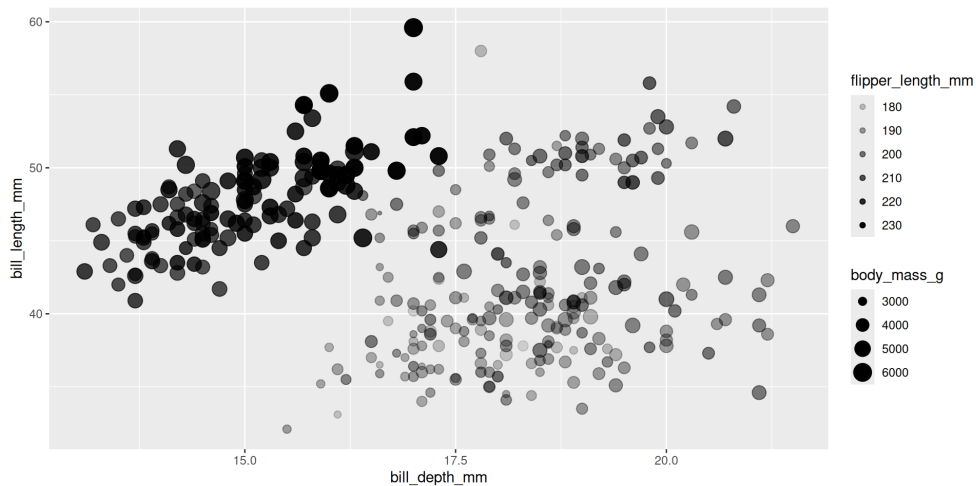
Mapping vs. Setting

```

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()

```

Mapping

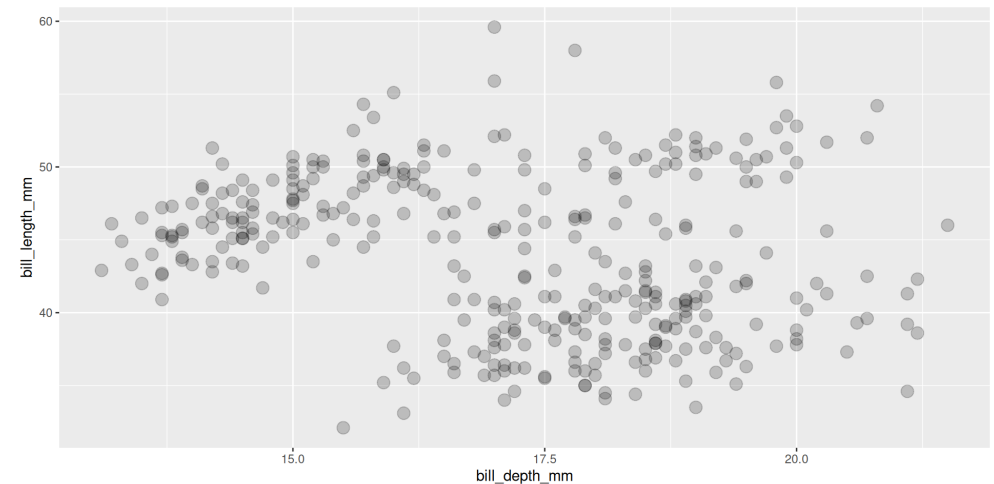


```

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)

```

Setting

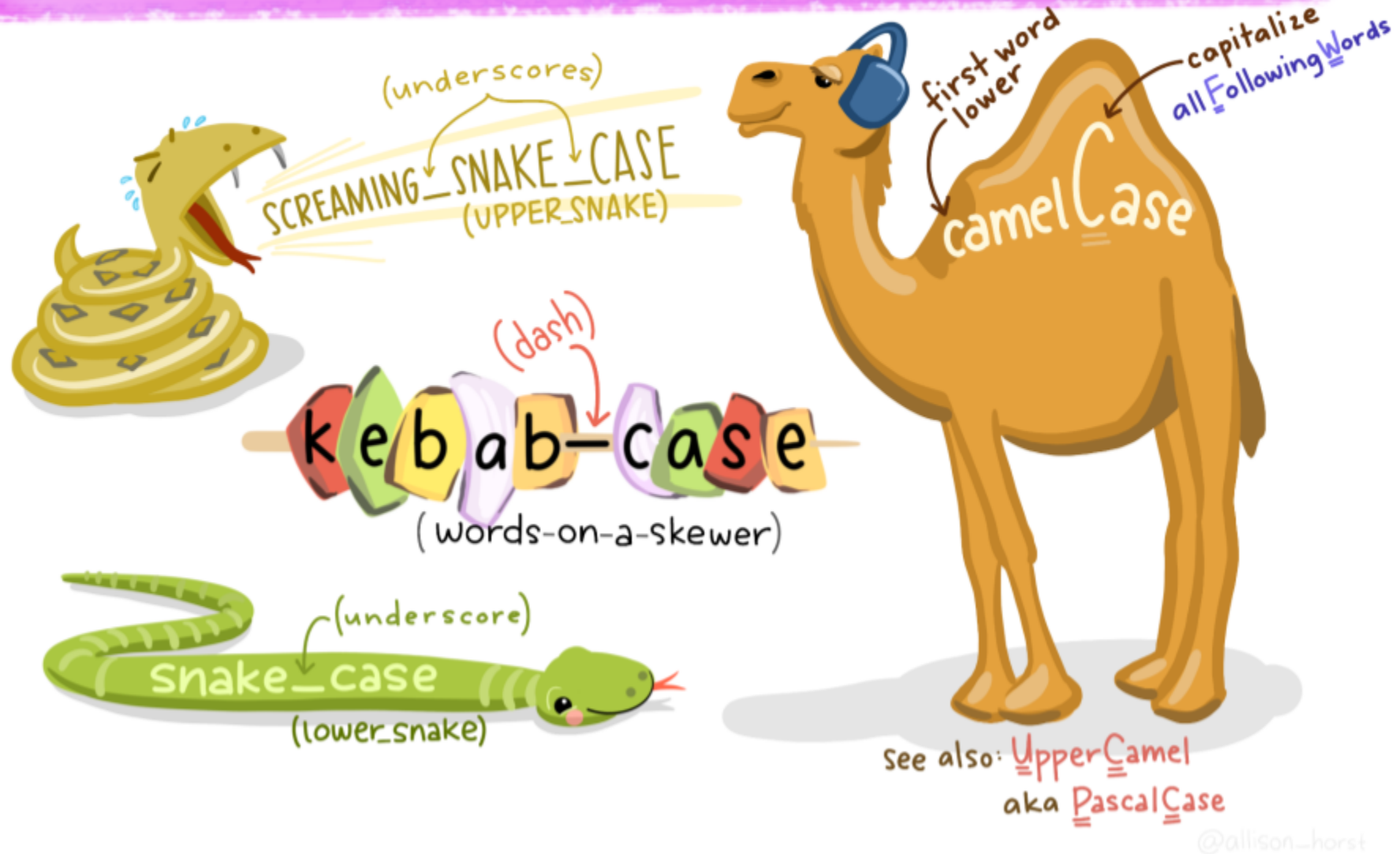


Workflow: Rechtschreibung

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...



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](#).