

rstatsZH - Data Science mit R

Datenvisualisierung mit ggplot2

Lars Schöbitz

2021-10-31

Rückblick - Woche 1

Viele neue Konzepte

- Git
- GitHub
- RStudio Cloud
- R
- R Markdown
- R Package `{ggplot2}`
- Pinguine 
- Help Files
- Code-chunks
- Argument, Objekt, Funktion, etc. etc.



Lob! - Woche 1

Viele neue Konzepte

- Git
- GitHub
- RStudio Cloud
- R
- R Markdown
- R Package `{ggplot2}`
- Pinguine 🐧
- Help Files
- Code-chunks
- Argument, Objekt, Funktion, etc. etc.

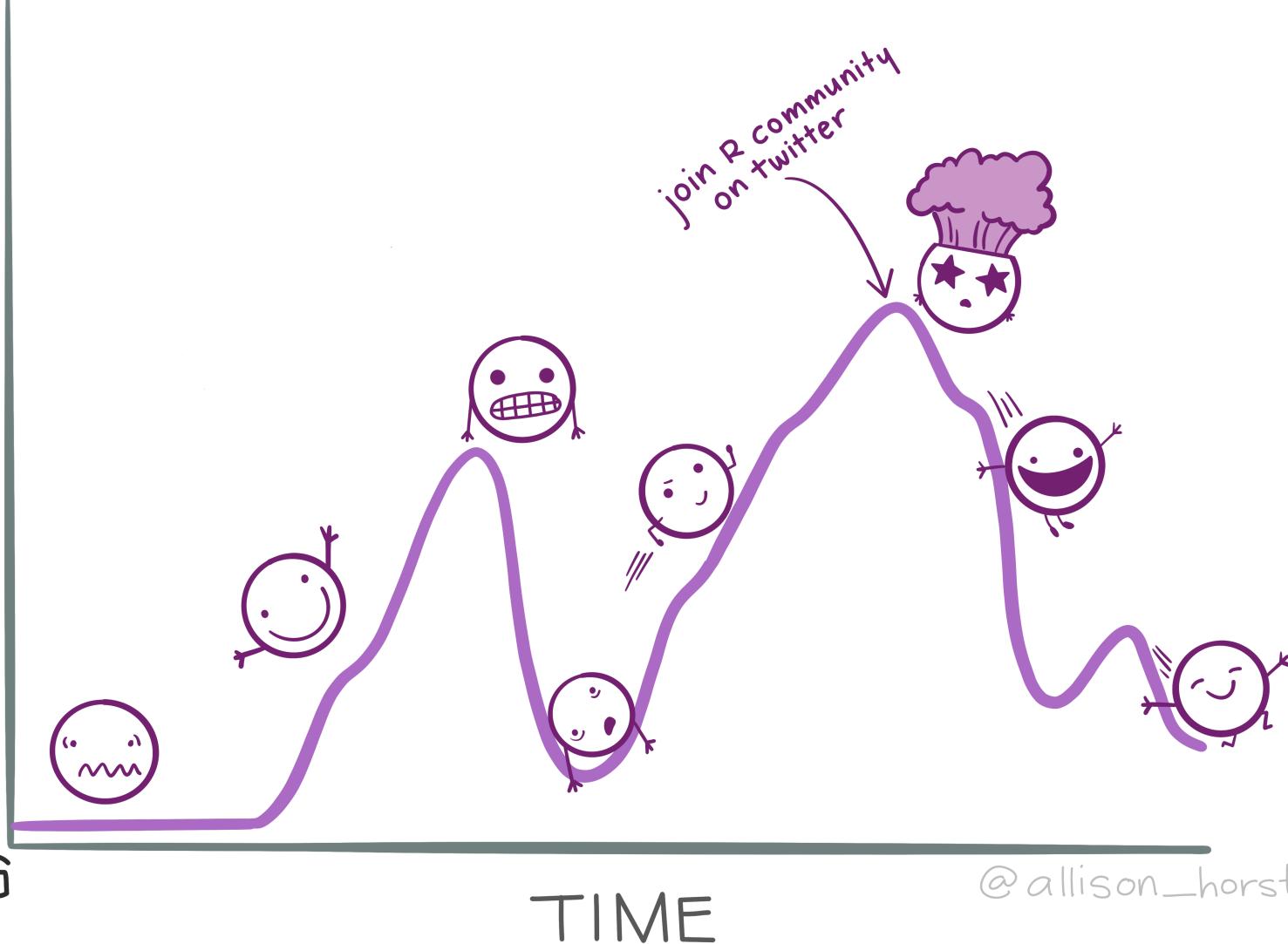


Photo by: [Annie Spratt](#)

HOW
MUCH
I THINK
I KNOW
ABOUT R

I KNOW_
LOTS!

I KNOW_
NOTHING

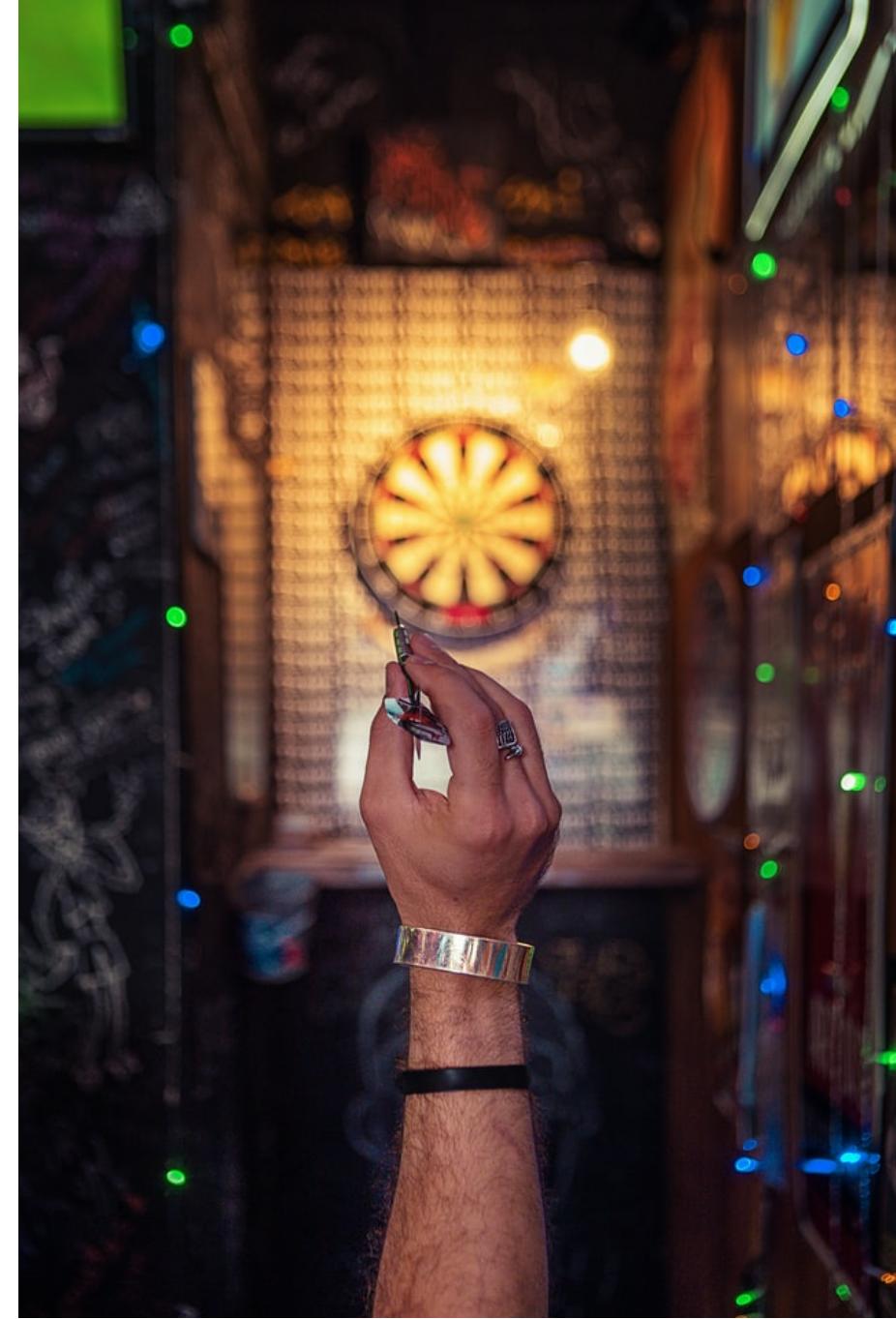


@allison_horst

Ziele für heute

Am Ende dieser Woche könnt ihr:

- den Unterschied zwischen markdown und R Markdown beschreiben
- selbstständig einen Plot mit dem R Package `{ggplot2}` erstellen
- innerhalb von RStudio selbstständig eine neue R Markdown Datei erstellen
- das Aussehen eines HTML Dokuments anpassen



Ein paar Tipps

Lesezeichen

- RStudio Cloud: <https://rstudio.cloud/spaces/179362/projects>
- Kurs Webseite: <https://rstatszh.github.io/website/>
- GitHub: <https://github.com/rstatszh-k005>

GitHub Zugangsdaten

- GitHub Nutzername
- GitHub Passwort
- GitHub Personal Access Token (PAT)

Erste Schritte (replay)

Lars hat das Repo prak02-rainbow-train auf GitHub erstellt prak02-rainbow-train Inbox x



lars@lse.de

to me ▾

GitHub Benutzername

Hallo Larnsce

Ich habe für dich das Repo 'prak02-rainbow-train' auf GitHub erstellt:

<https://github.com/rstatszh-k000/prak02-rainbow-train>



Klick hier

Du kannst den Anweisungen unter 'Erste Schritte' aus Hausaufgabe 01 folgen um das Repo in deine RStudio Cloud zu laden.

<https://rstatszh.github.io/website/ha-01-hallo-git.html>

Bei Fragen kannst du dich jederzeit auf unerem Slack Channel bei mir melden.

Viel Spass!

Lars

RStudio Cloud x rstatszh-k002/ha01-rainbow-train +

github.com/rstatszh-k002/ha01-rainbow-train

Search or jump to... / Pull requests Issues Marketplace Explore

Watch 1 Star 0 Fork 0

rstatszh-k002 / ha01-rainbow-train Private

generated from rstatsZH/ha-01-hallo-git

Code Issues Pull requests Actions Projects Security Insights

GitHub Organisation **Repo(sitory)**

Go to file Add file Code

larnsce Initial commit b8de0dd 4 days ago 1 commit

.gitignore Initial commit 4 days ago

README.md Initial commit 4 days ago

ha-01-hallo-git.Rproj Initial commit 4 days ago

ha-01.Rmd Initial commit 4 days ago

README.md

README

Die Übung in diesem Repository wurde für den Kurs "rstatsZH - Data Science mit R" erstellt. Diese Datei

About
No description, website, or topics provided.

Readme

Releases
No releases published
[Create a new release](#)

Packages
No packages published
[Publish your first package](#)

The screenshot shows a GitHub repository page for 'rstatszh-k002/ha01-rainbow-train'. The repository was generated from 'rstatsZH/ha-01-hallo-git'. The main navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Security, and Insights. Below the navigation, there are buttons for the 'main' branch (1 branch, 0 tags) and a 'Code' dropdown menu. The 'Code' menu is open, showing options for Go to file, Add file, and Clone. The 'Clone' section displays the HTTPS URL 'https://github.com/rstatszh-k002/ha01-rainbow-train'. A pink box highlights the 'Code' dropdown and the copy icon next to the URL. To the right of the repository details, there are three sections: '1. Schritt' (No description, website, or topics provided), '2. Schritt' (Releases, No releases published, Create a new release), and Packages (No packages published, Publish your first package). The repository page also lists files: '.gitignore', 'README.md', 'ha-01-hallo-git.Rproj', and 'ha-01.Rmd', all with 'Initial commit' status.

RStudio Cloud rstatszh-k002/ha01-rainbow-tr

rstudio.cloud/spaces/122222/projects

Projects Members About

All Projects

List All projects Sort By name

ha01-rainbow-train

Rainbow Train

Created Mar 4, 2021 1:41 PM

prak-00-willkommen

Rainbow Train

Created Mar 1, 2021 1:51 PM

Derived from: prak-00-willkommen by Lars Schöbitz

CONTINUE prak-00-willkommen

Lars Schöbitz

Created Mar 1, 2021 11:14 AM

View 1 derived project ...

prak-01-visualisieren

Rainbow Train

Created Mar 1, 2021 2:38 PM

Derived from: prak-01-visualisieren by Lars Schöbitz

New Project

+ New Project

New Project from Git Repository

Delete Move

3. Schritt

Delete Move

Delete Move

RStudio Cloud x rstatszh-k002/ha01-rainbow-tr +

[rstudio.cloud/spaces/122222/projects](#)

All Projects Projects Members About [New Project](#)

4. Schritt

New Project from Git Repository

URL of your Git Repository
`https://github.com/rstatszh-k002/ha01-rainbow-train.git`

Packages
 Add packages from the base project

OK

5. Schritt

CONTINUE prak-00-willkommen

Lars Schöbitz

Created Mar 1, 2021 11:14 AM View 1 derived project ...

prak-01-visualisieren

Rainbow Train

Created Mar 1, 2021 2:38 PM Derived from: prak-01-visualisieren by Lars Schöbitz

The screenshot shows the RStudio Cloud interface with several panels highlighted by colored borders:

- Menu Bar:** The top navigation bar is highlighted with an orange border. It includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help.
- Code Editor:** The left panel, containing an R Markdown file named "prak-01.Rmd", is highlighted with a pink border. The code includes a title and a note about visualizing data. A pink box labeled "Code Editor" is overlaid on this panel.
- Environment:** The top-right panel, showing the R environment and global variables, is highlighted with a blue border. A blue box labeled "Environment" is overlaid on this panel.
- Console:** The bottom-left panel, displaying R startup messages and a command line interface, is highlighted with a teal border. A teal box labeled "Console" is overlaid on this panel.
- File Manager:** The bottom-right panel, showing project files and their modification times, is highlighted with a green border. A green box labeled "Datei Manager (Files)" is overlaid on this panel.

Other visible elements include a RAM usage indicator, a gear icon, and a "Rainbow Train" icon in the top right corner. The R version is listed as 4.0.3 in the top right of the main window area.

Praktikum 2 - R Markdown

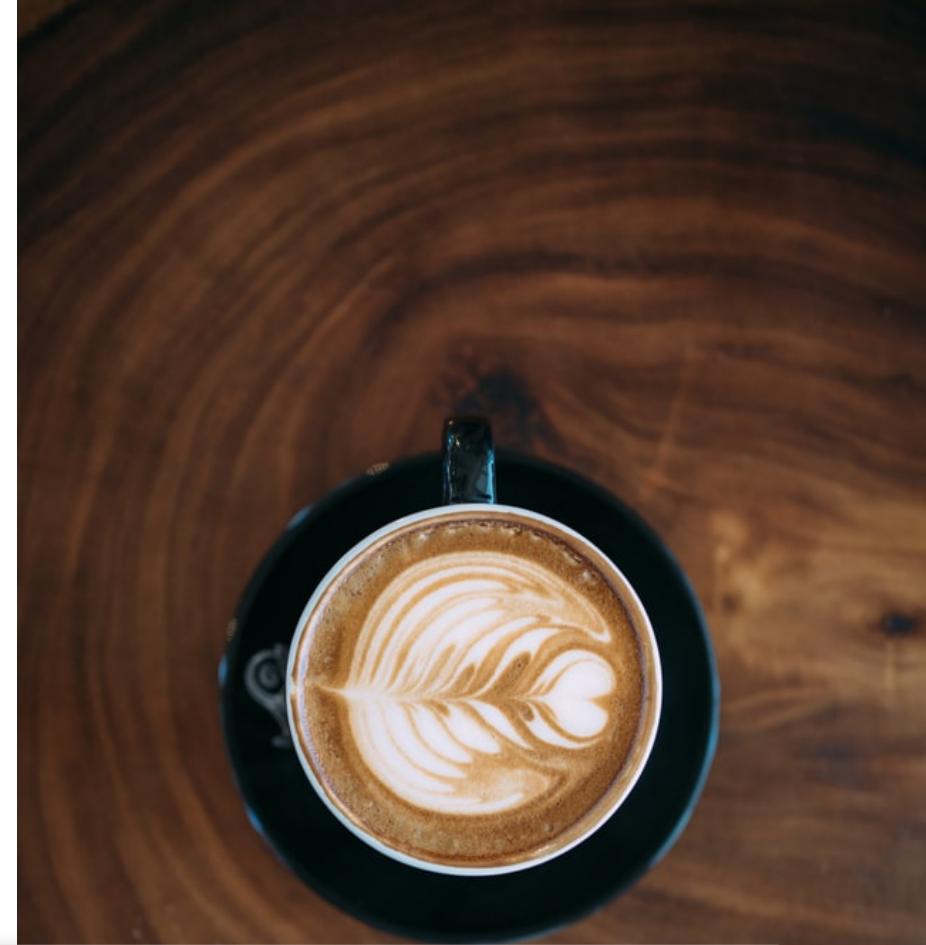
Praktikum 2 - R Markdown

1. **E-Mail:** Öffne deine Email und klicke auf den Link zu deinem persönlichen GitHub repo für **prak-02**
2. **GitHub:** Klicke auf den grünen Button "Code" und kopiere den Link für das Repo in deine Zwischenablage
3. **RStudio Cloud:** Öffne deinen Arbeitsbereich für den Kurs in der RStudio Cloud
4. **RStudio Cloud / Projects:** Klicke auf "New Project from GitHub Repository"
5. **RStudio:** Finde den Datei Manager und Git
6. **Zoom Chat:** Schreibt wenn ihr soweit seid

Pause

10 : 00

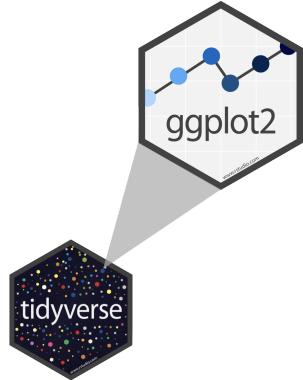
Photo by: [Blake Wisz](#)



ggplot2 ❤️



ggplot2 ∈ tidyverse

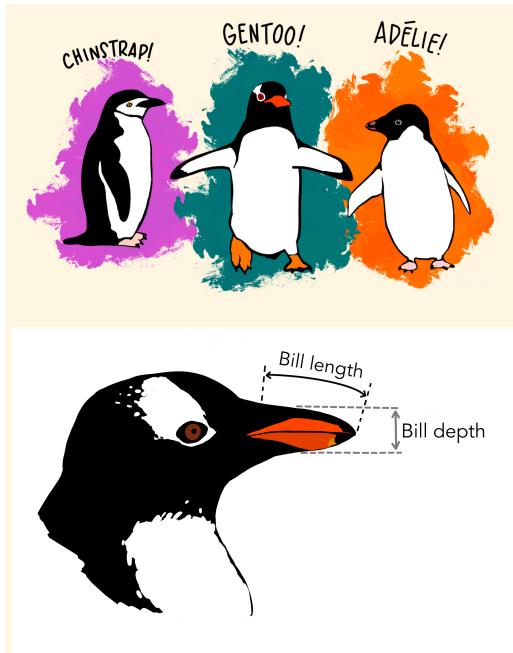


- **ggplot2** ist das R Package für Datenvisualisierung im tidyverse
- Struktur des Codes kann folgerndemassen zusammengefasst werden:

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

Data: Palmer Penguins

Ein Datensatz über Pinguine Spezies auf Inseln im Palmer Archipelago mit Variablen für Grösse (Flossenlänge, Gewicht, Schnabelmasse), und Geschlecht.



Rows: 344

Columns: 8

\$ species

\$ island

\$ hill length mm

\$ bill_dent_mm

\$ bit_depth_min
\$ flipper_length_mm

\$ pepper_tch
\$ body mass g

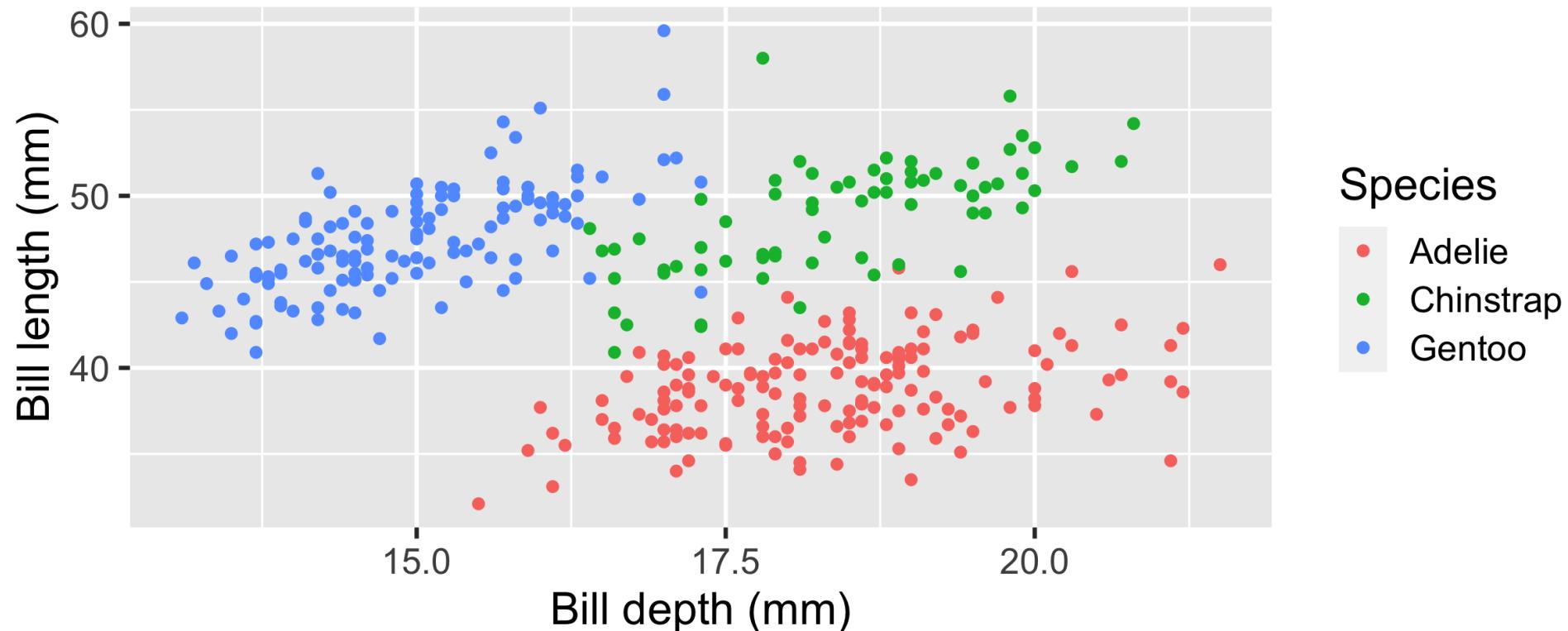
body
copy

\$ 500

```
<fct> Adelie, Adelie, Adelie, Adelie  
<fct> Torgersen, Torgersen, Torgersen  
<dbl> 39.1, 39.5, 40.3, NA, 36.7,  
<dbl> 18.7, 17.4, 18.0, NA, 19.3,  
<int> 181, 186, 195, NA, 193, 190,  
<int> 3750, 3800, 3250, NA, 3450,  
<fct> male, female, female, NA, female  
<int> 2007, 2007, 2007, 2007, 2007
```

Bill depth and length

Dimensions for Adelie, Chinstrap, and Gentoo Penguins



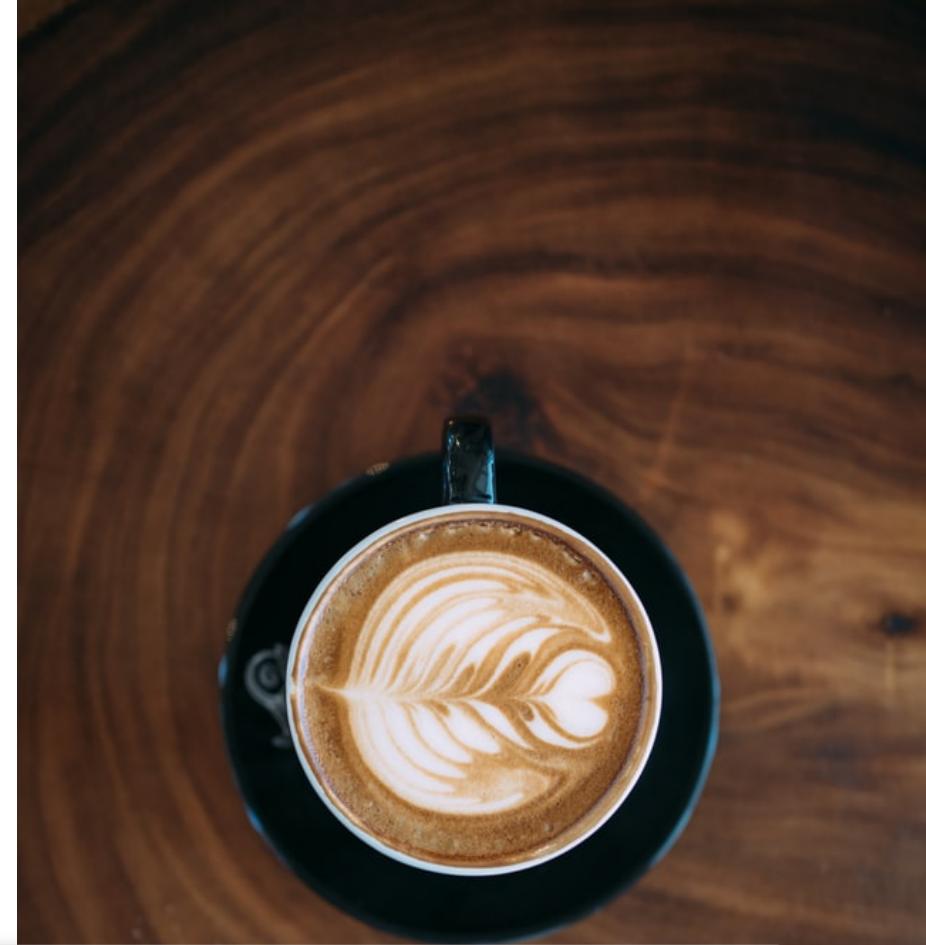
Praktikum 3 - {ggplot2}

1. **E-Mail**: Öffne deine Email und klicke auf den Link zu deinem persönlichen GitHub repo für **prak-03**
2. **GitHub**: Klicke auf den grünen Button "Code" und kopiere den Link für das Repo in deine Zwischenablage
3. **RStudio Cloud**: Öffne deinen Arbeitsbereich für den Kurs in der RStudio Cloud
4. **RStudio Cloud / Projects**: Klicke auf "New Project from GitHub Repository"
5. **RStudio**: Finde den Datei Manager und Git
6. **Zoom Chat**: Schreibt wenn ihr soweit seid

Pause

05 : 00

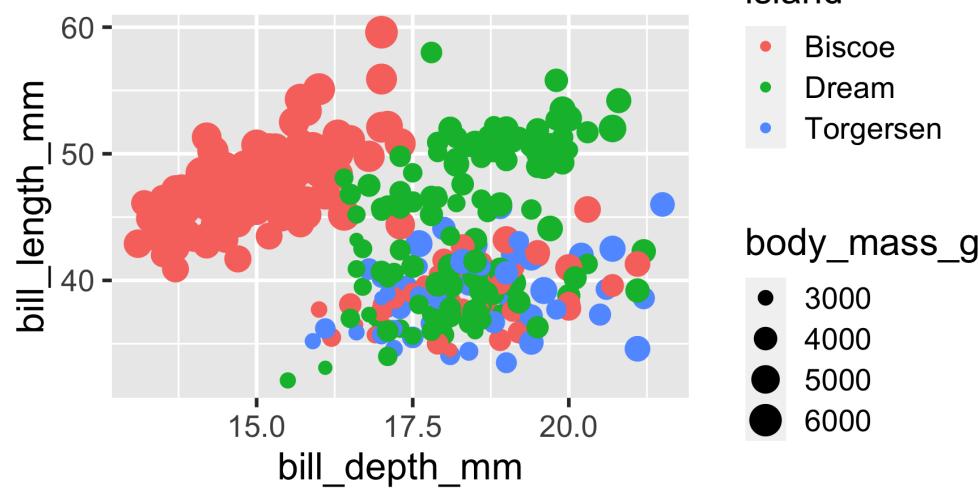
Photo by: [Blake Wisz](#)



Mapping und Faceting

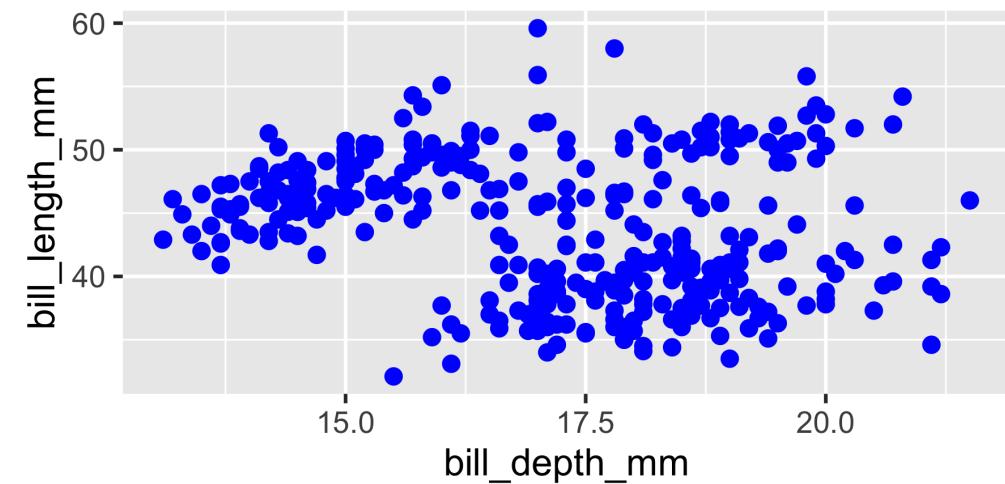
Mapping: Grösse, Alpha, Farbe, etc. der Punkte werden basierend auf den Werten der Variable in den Daten bestimmt.

```
ggplot(data = penguins,  
       mapping = aes(x = bill_depth_mm  
                      , y = bill_length_mm  
                      , size = body_mass_g  
                      , color = island))  
  geom_point()
```



Setting: Grösse, Alpha, Farbe, etc. der Punkte werden **nicht** basierend auf den Werten der Variable in den Daten bestimmt

```
ggplot(data = penguins,  
       mapping = aes(x = bill_depth_mm  
                      , y = bill_length_mm))  
  geom_point(size = 3, color = "blue")
```



Faceting - Zusammenfassung

- `facet_grid()`:
- 2d grid
- `rows ~ cols`
- nutze `.` um nicht zu splitten
- `facet_wrap()`: 1d wird entsprechend der angegebenen Anzahl von Zeilen und Spalten oder der verfügbaren Plotfläche unterbrochen

Praktikum 4 - {ggplot2}

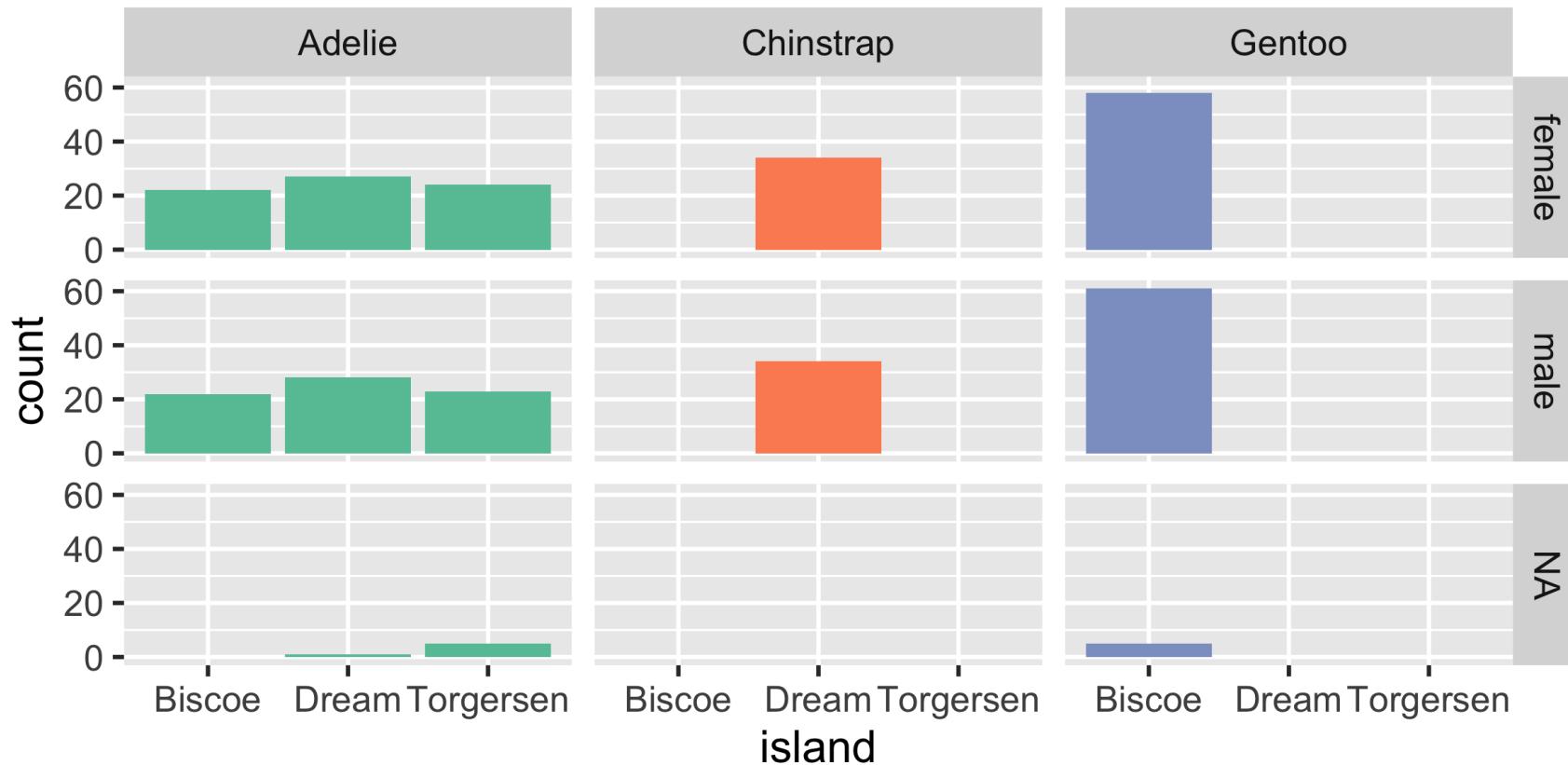
Arbeitet in 2er Teams. Eine Person teilt den Bildschirm. Die andere schreibt mit.

1. **E-Mail**: Öffne deine Email und klicke auf den Link zu deinem persönlichen GitHub repo für **prak-04**
2. **GitHub**: Klicke auf den grünen Button "Code" und kopiere den Link für das Repo in deine Zwischenablage
3. **RStudio Cloud**: Öffne deinen Arbeitsbereich für den Kurs in der RStudio Cloud
4. **RStudio Cloud / Projects**: Klicke auf "New Project from GitHub Repository"

Praktikum 4 - Lösung

Plot

Code



Feedback

Ziele erreicht?

Bitte ausfüllen: kutt.it/rstatszh-eval



Hausaufgabe

Hausaufgabe

- Öffnet jetzt eure E-Mail Inbox
- Bestätigt mir im Zoom Chat, dass ihr eine Email mit Betreff "**rstatsZH - Lars hat das Repo ha02-GitHubName**" erhalten habt
- Die Anweisungen für die Heausaufgabe 02 findet ihr oben rechts auf unserer Kurswebseite
- Kontaktiert mich unter der Woche jederzeit auf Slack



Danke

Für die Aufmerksamkeit!

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

Eine PDF Version der Folien kann hier heruntergeladen werden:

https://github.com/rstatsZH/website/raw/master/slides/e1_d02-data-vis/e1_d02-data-vis.pdf

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