

rstatsZH - Data Science mit R

Datenvisualisierung mit ggplot2

Lars Schöbitz

2021-05-25

Rückblick - Woche 1

Viele neue Konzepte

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

Photo by: [Annie Spratt](#)



Lob! - Woche 1

Viele neue Konzepte

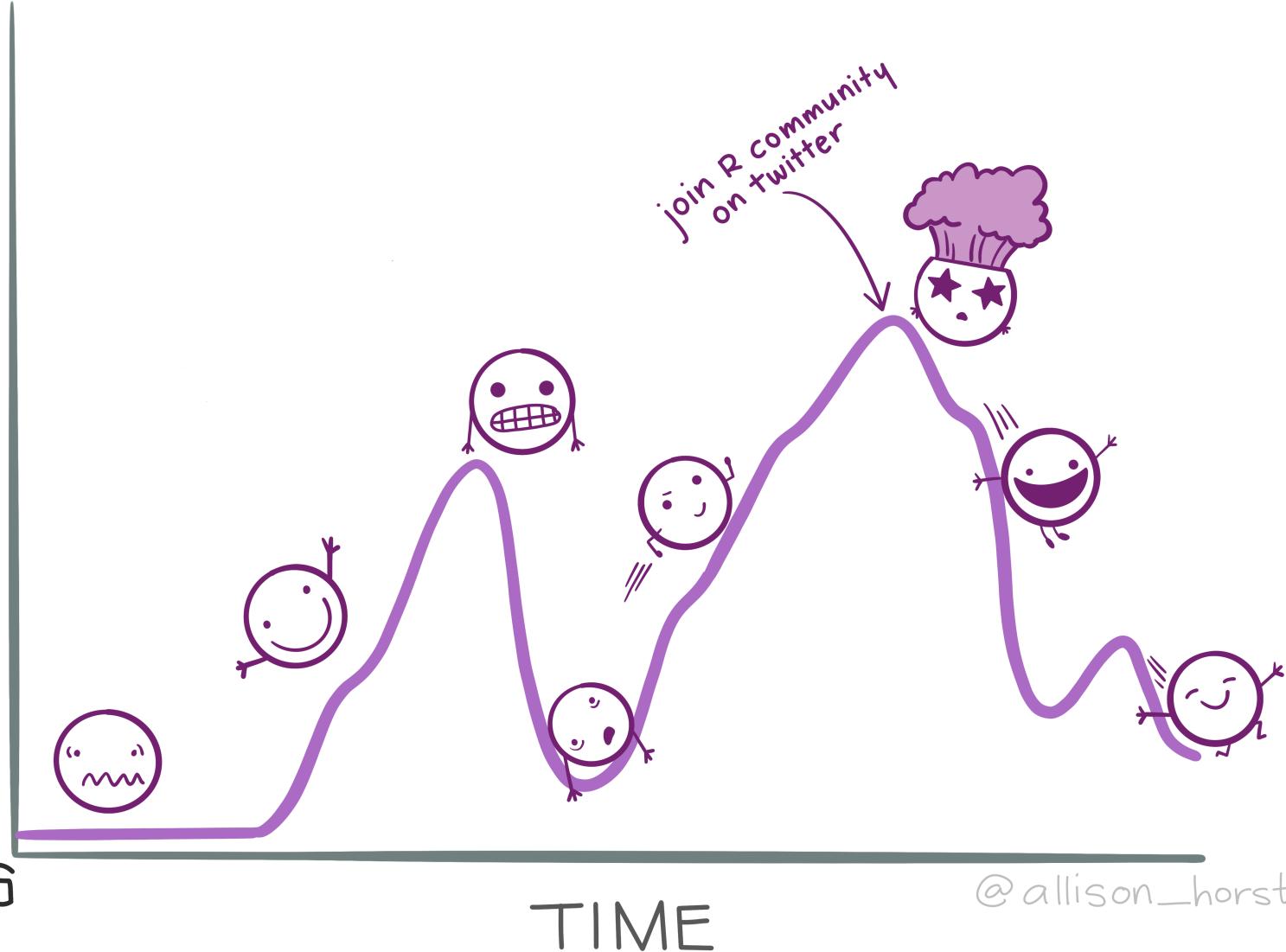
- Git
- GitHub
- RStudio Cloud
- R
- R Markdown
- R Package `{ggplot2}`
- Pinguine 🐧
- Help Files
- Code-chunks
- Cheatheets
- Argument, Objekte, Funktionen, etc. etc.



HOW
MUCH
I THINK
I KNOW
ABOUT R

I KNOW -
NOTHING

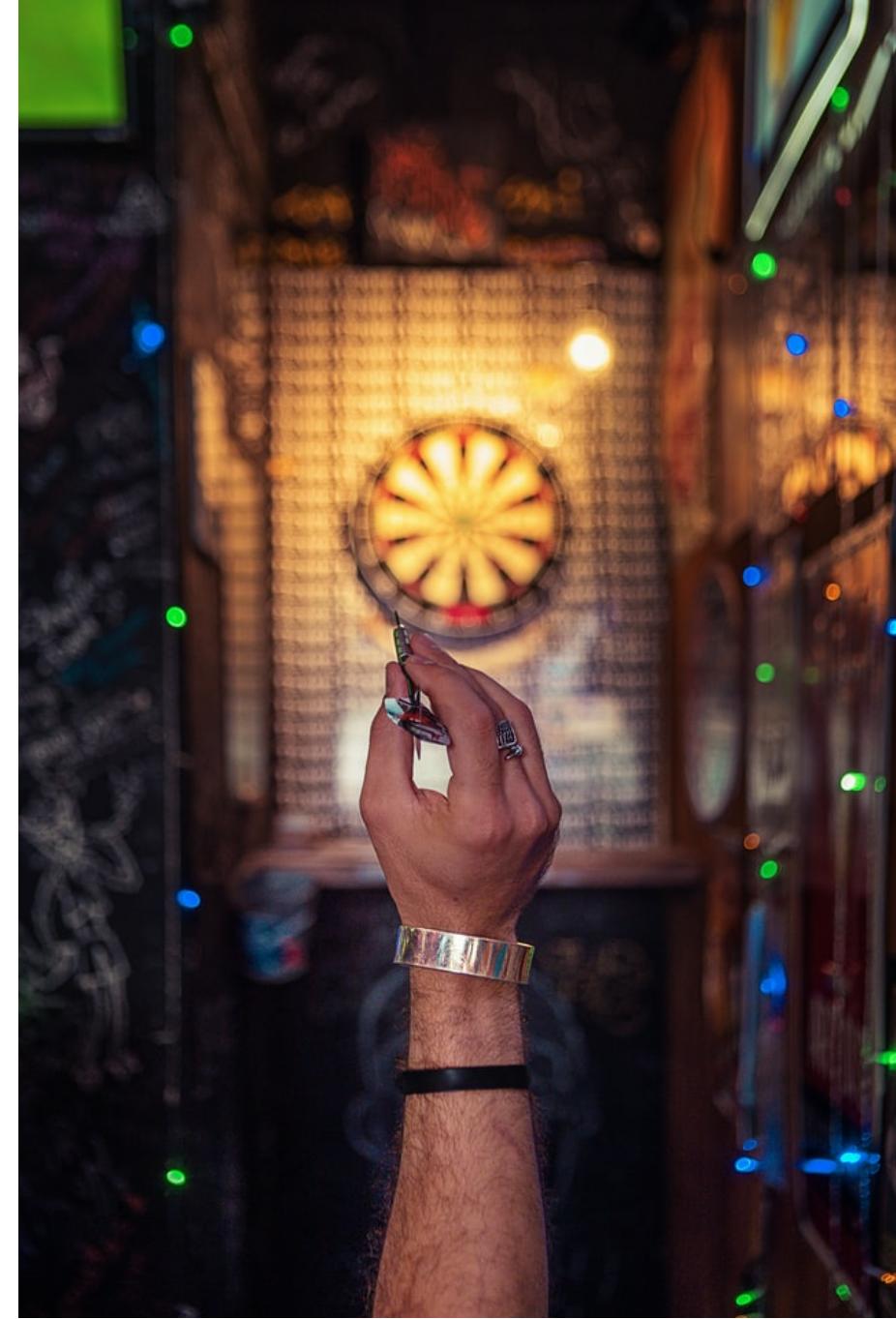
I KNOW -
LOTS!



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/142876/projects>
- Kurs Webseite: <https://rstatszh.github.io/website/>
- GitHub: <https://github.com/rstatszh-k003>

GitHub Credentials

- GitHub Nutzername
- GitHub Passwort

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](#)

RStudio Cloud

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

main 1 branch 0 tags

Go to file Add file Code

Clone

HTTPS SSH GitHub CLI

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

Use Git or checkout with SVN using the web URL

Open with GitHub Desktop

Download ZIP

1. Schritt
No description, website, or topics provided.

2. Schritt

Releases
No releases published
Create a new release

Packages
No packages published
Publish your first package

README

README.md

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

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 rstatszh-k002/ha01-rainbow-tr

rstudio.cloud/spaces/122222/projects

All Projects

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

4. Schritt

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

Delete Move

The screenshot shows the RStudio Cloud interface with several panels:

- Menu Bar:** Located at the top, it includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. A yellow box highlights the "File" menu.
- Code Editor:** Shows an R Markdown file named "prak-01.Rmd". The code includes a title, date, and output type, followed by a note about working with a dataset. A pink box highlights the code area. The word "Code Editor" is overlaid in white text on a pink background.
- Environment:** Displays the Global Environment pane. A blue box highlights the "Global Environment" tab. The word "Environment" is overlaid in white text on a blue background.
- Console:** Shows the R console output. It starts with the standard R welcome message, followed by information about the project's license and contributors. A green box highlights the "Console" tab. The word "Console" is overlaid in white text on a green background.
- File Manager:** Shows the project's directory structure. A green box highlights the "New Folder" button. The words "Datei Manager (Files)", "Plots", "Help File", and "Viewer" are overlaid in white text on a green background.

Additional interface elements include a RAM usage indicator, a gear icon, a three-dot menu, and a "Rainbow Train" icon in the top right corner. The R version is shown as R 4.0.3.

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

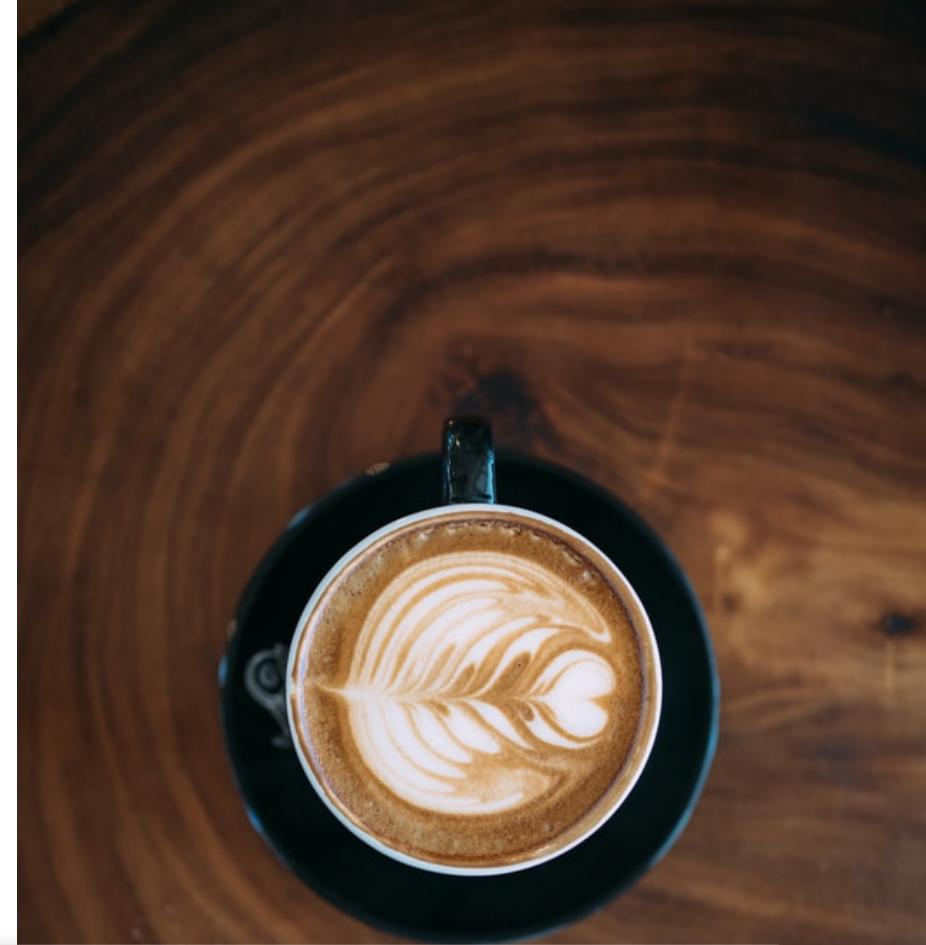
Praktikum 2 - R Markdown

1. **Bildschirm:** Konfiuriere deinen Bildschirm so, dass du mir gleichzeitig folgen kannst und selbst Code schreiben kannst

Pause

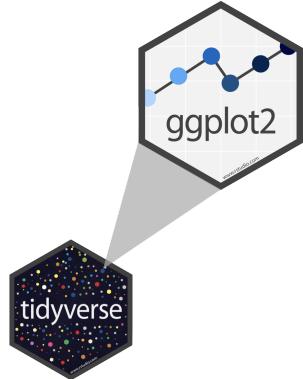
10 : 00

Photo by: [Blake Wisz](#)



ggplot2 ❤️ 🐧

ggplot2 ∈ tidyverse

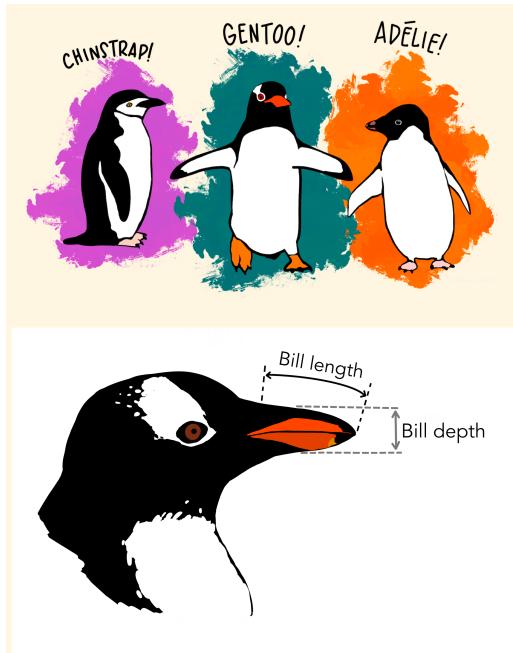


- **ggplot2** ist das R Package für Datenvisualisierung im tidyverse
- Struktur des Codes kann folgerndermassen zusammengefasst werden:

```
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_denth_mm

flipper_length_mm

Pepper_zen
body mass g

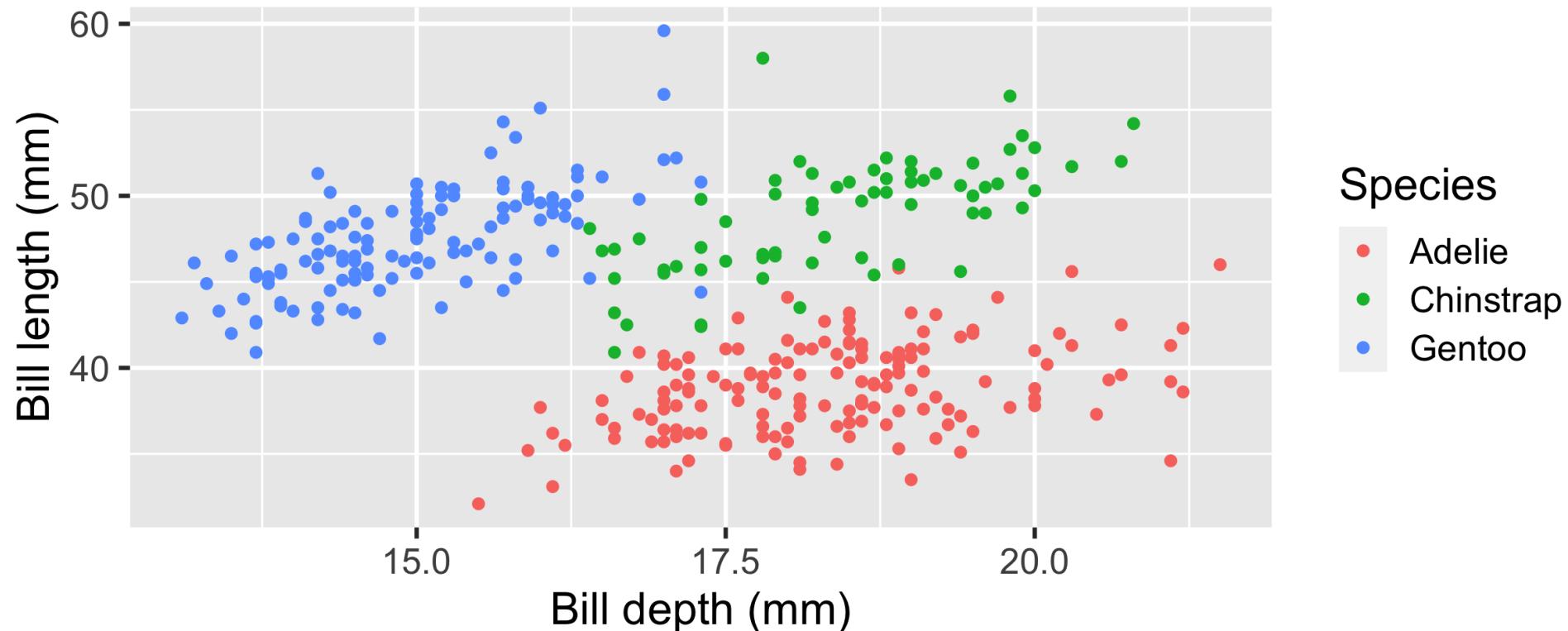
9
t

\$ sex
\$ mean

```
<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

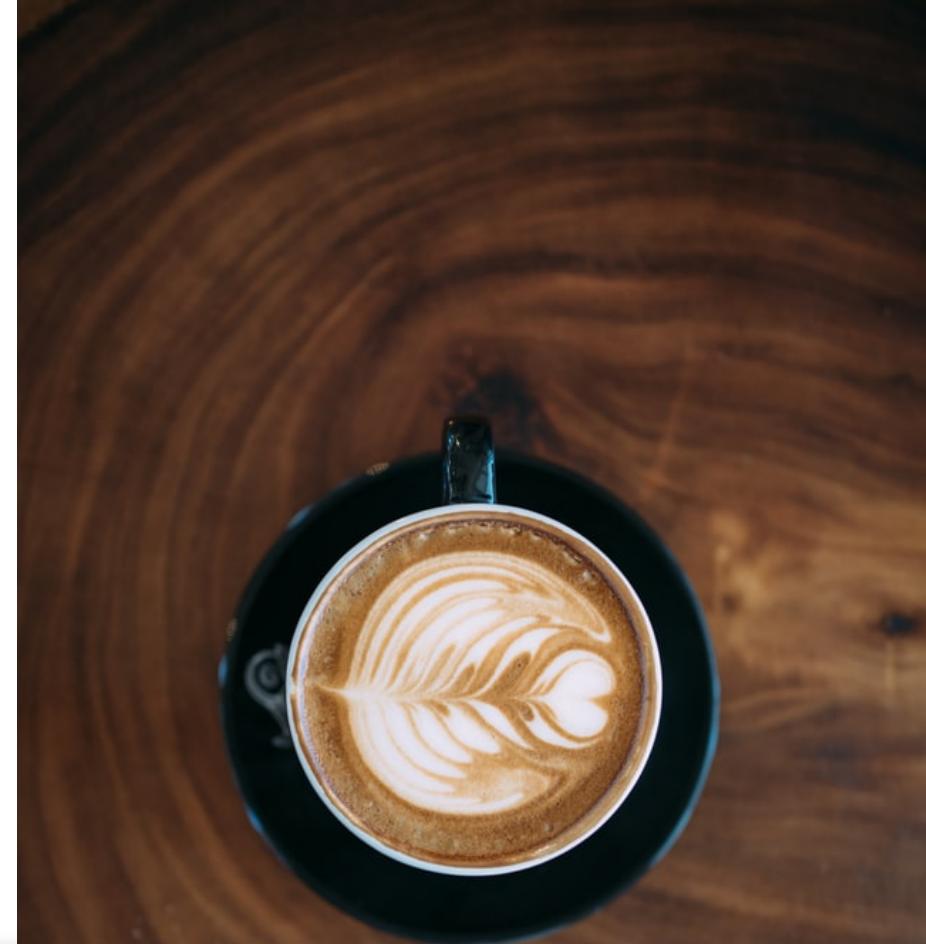
Dimensions for Adelie, Chinstrap, and Gentoo Penguins



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

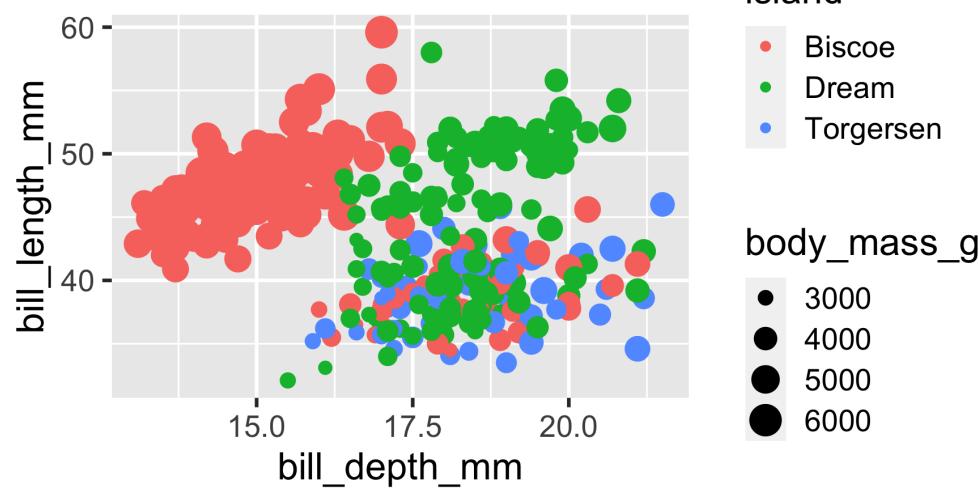


Photo by: [Blake Wisz](#)

Mapping und Faceting

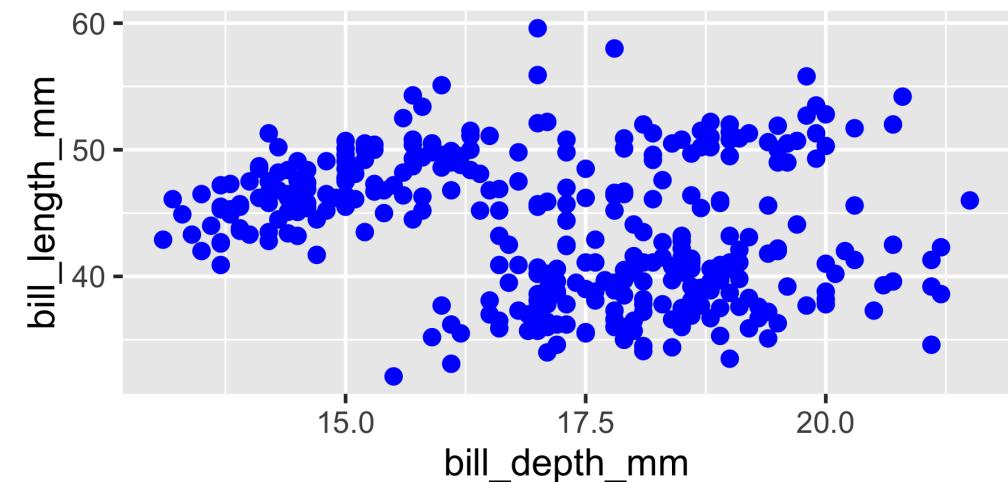
Mapping: Grösse, Alpha, Farbe, etc. der Punkte werden basierend auf den Werten der Variable in den Daten bestimmt.

```
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}

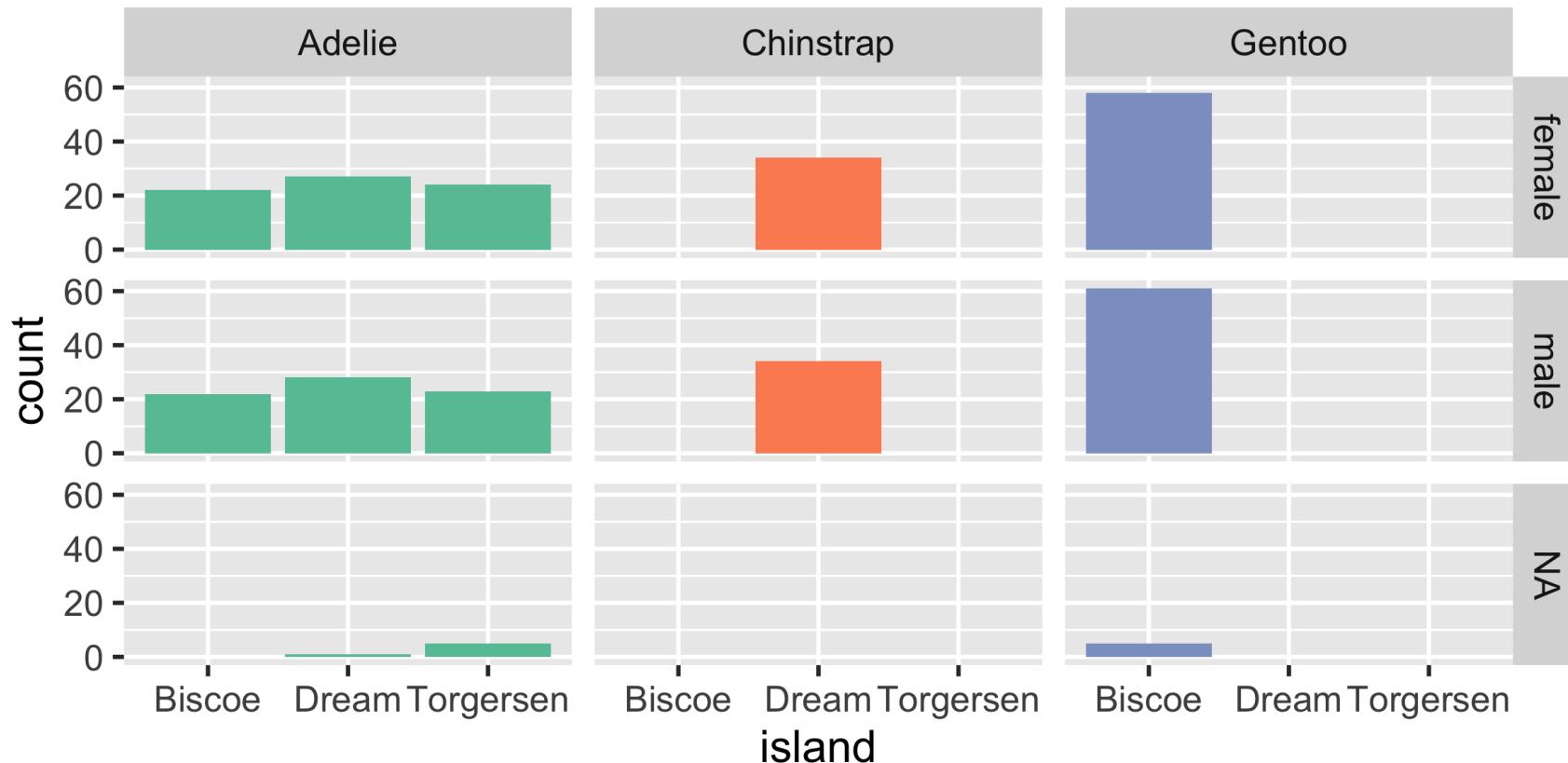
Arbeitet in 2er Teams. Eine Person teilt den Bildschirm. Die andere schreibt mit.

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.