

# 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.

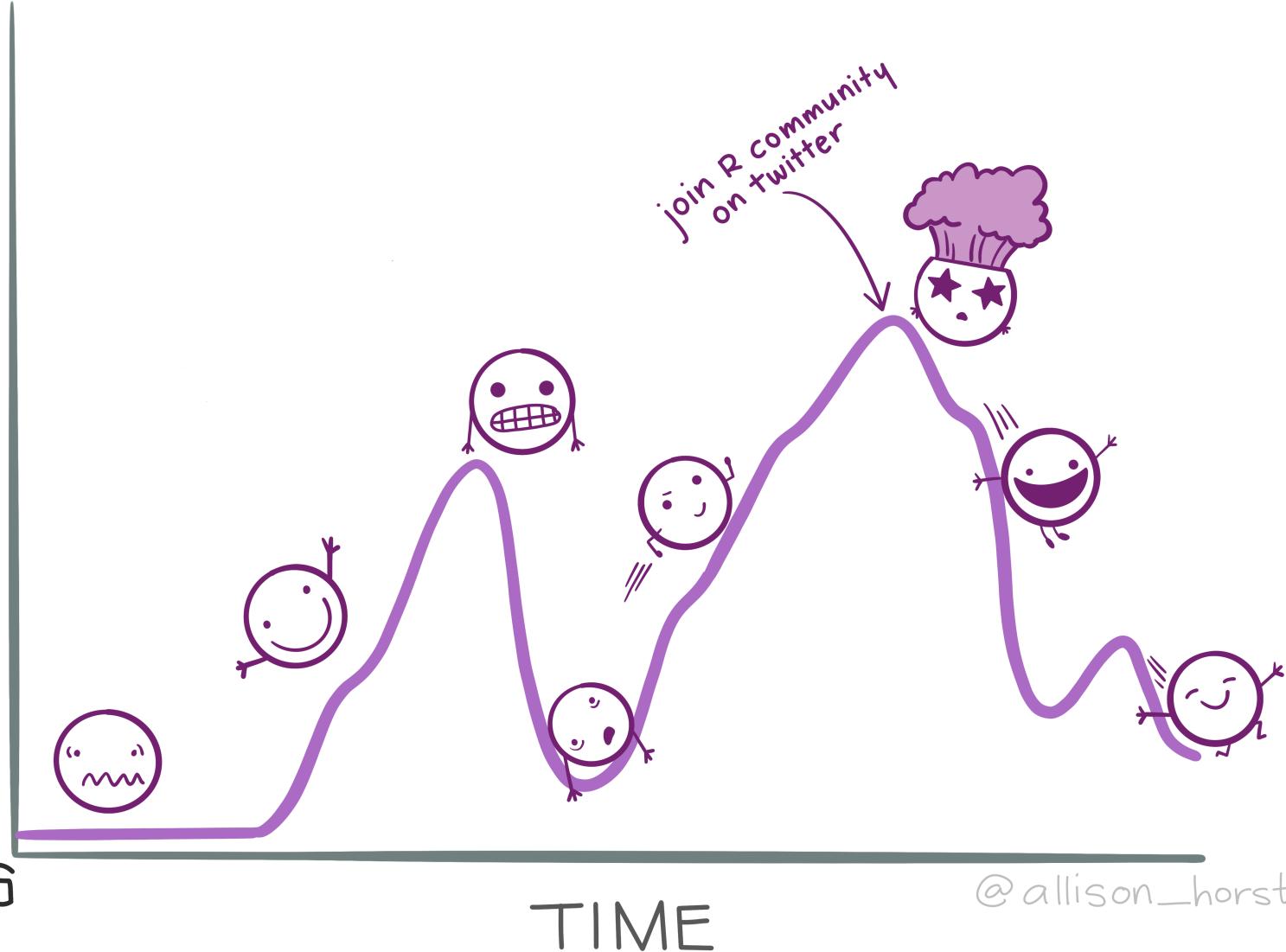


Photo by: [Annie Spratt](#)

HOW  
MUCH  
I THINK  
I KNOW  
ABOUT R

I KNOW -  
NOTHING

I KNOW -  
LOTS!

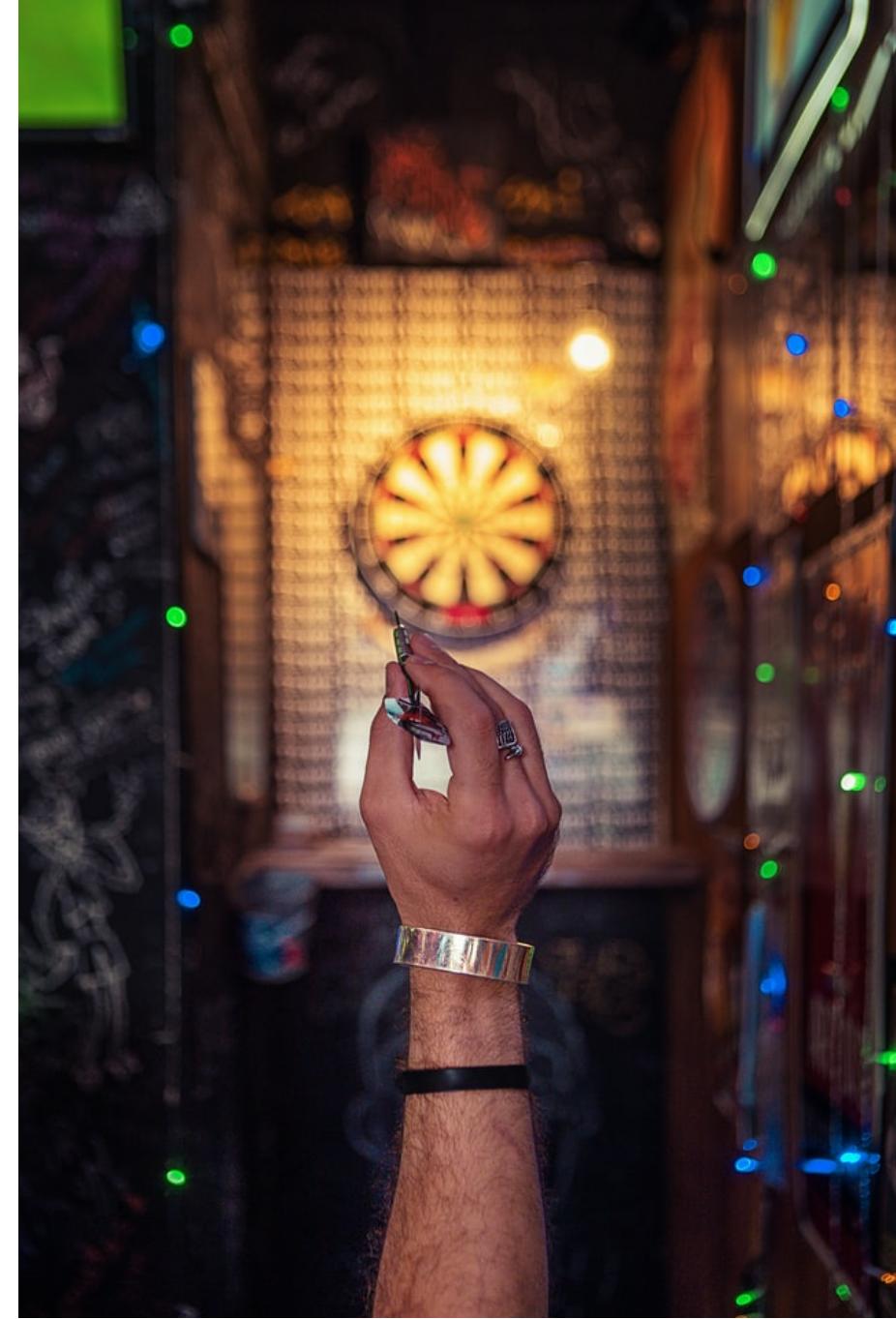


@allison\_horst

# 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/122221/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**

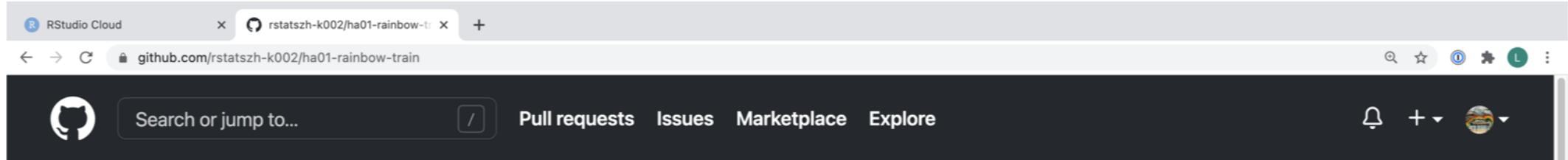
Du kannst den Anweisungen unter 'Erste Schritte' aus Hausaufgabe 01 folgen um das Repo in deine RStudio Cloud zu laden.

<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



**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
	.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. At the top, there's a navigation bar with links for Pull requests, Issues, Marketplace, and Explore. Below the header, the repository name 'rstatszh-k002 / ha01-rainbow-train' is displayed, along with a 'Private' badge, a 'Watch' button (with 1 watch), a 'Star' button (0 stars), and a 'Fork' button (0 forks). The main content area shows a commit from 'larnsce' with the message 'Initial commit'. Below the commit, there are files listed: '.gitignore', 'README.md', 'ha-01-hallo-git.Rproj', and 'ha-01.Rmd', each with an 'Initial commit' note. To the right, there's a 'Code' dropdown menu with options: 'Go to file', 'Add file', 'Code' (which is highlighted with a pink box), 'Clone' (with a URL field containing 'https://github.com/rstatszh-k002/ha01-rainbow-train'), 'Open with GitHub Desktop', and 'Download ZIP'. A pink box also highlights the URL field in the 'Clone' section. On the far right, there are sections for '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). At the bottom, there's a large 'README' section with the text: '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    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

RstatsZH-K002 / prak-01-visualisieren



Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

# Menü Leiste

R 4.0.3

prak-01.Rmd x

```
1 ---  
2 title: "Daten visualisieren"  
3 name: ""  
4 date: ""  
5 output: html  
6 ---  
7  
8 # Daten  
9  
10 In dieser ersten Übung arbeiten wir mit einem Datensatz über  
5:22 # Daten visualisieren R Markdown
```

# Code Editor

Environment History Connections Git Tutorial

Import Dataset

R Global Environment

# Environment Git

Console Terminal Jobs

```
/cloud/project/ ↵  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
R is a collaborative project...tors.  
Type 'contributors()' for ...itors.  
'citation()' on how to cite R or its publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

# Console

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

- ..
- .git
- .Rhistory
- prak-01.Rmd
- prak-01.Rproj
- README.md

# Datei Manager (Files) Plots Help File Viewer

	2:46 PM
..	2:46 PM
.git	2:46 PM
.Rhistory	2:27 PM
prak-01.Rmd	2:46 PM
prak-01.Rproj	2:46 PM
README.md	2:46 PM

# 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

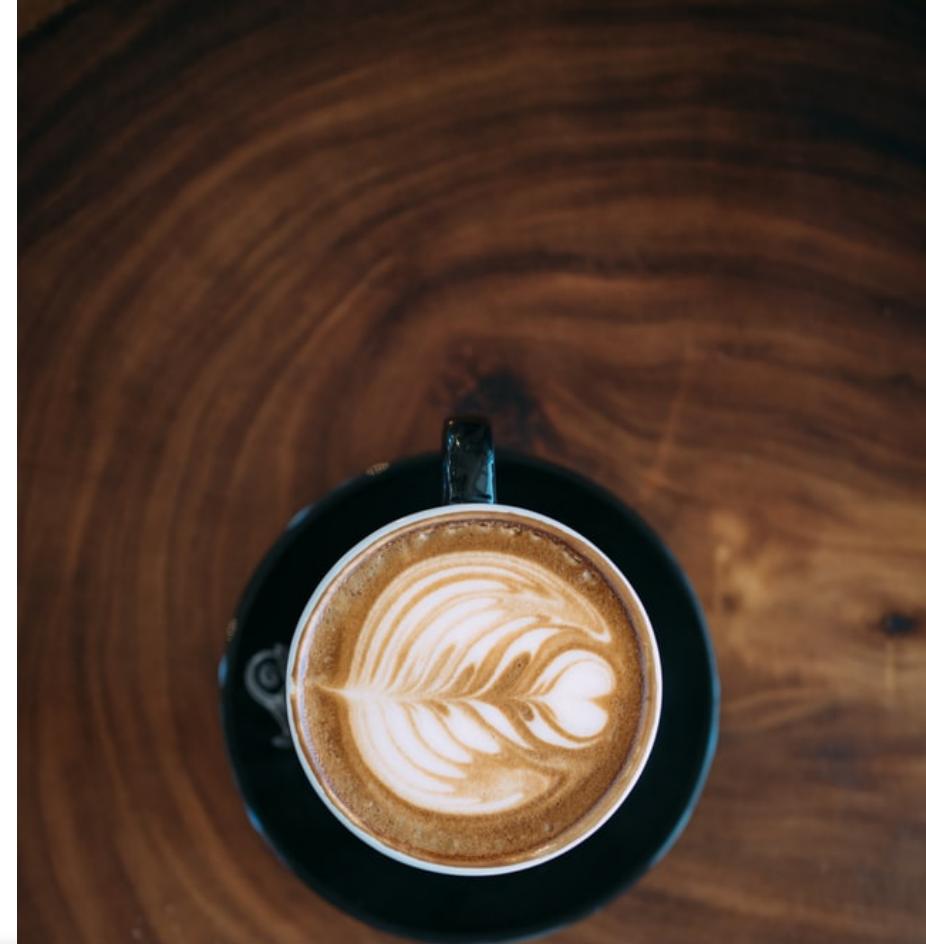
# Praktikum 2 - R Markdown

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

# Pause

05 : 00

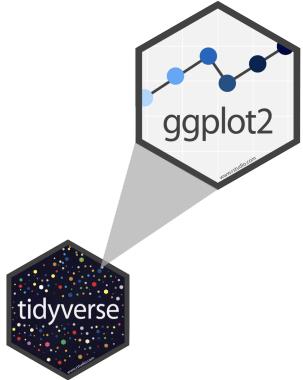
Photo by: [Blake Wisz](#)



ggplot2 ❤️



# ggplot2 ∈ tidyverse

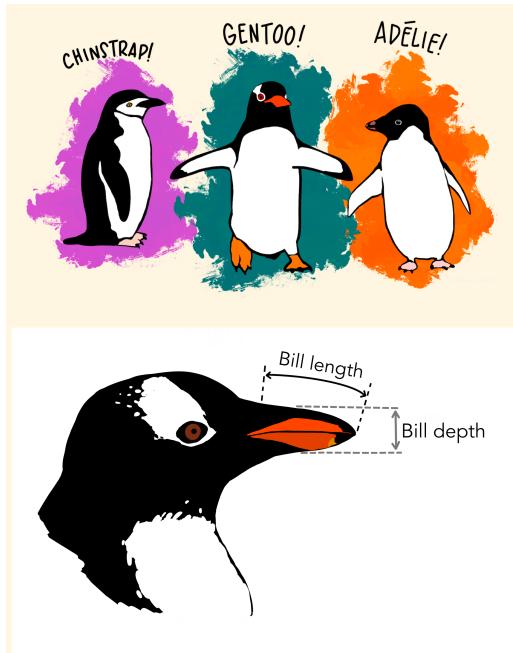


- **ggplot2** ist das R Package für Datenvisualisierung im tidyverse
- Struktur des Codes kann folgerndermassen zusammengefasst werden:

```
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

\$\text{hill length mm}

\$ bill\_denth\_mm

+ slice\_depth\_mm  
\$ flipper\_length\_mm

\$ flipper\_tch  
\$ body mass &

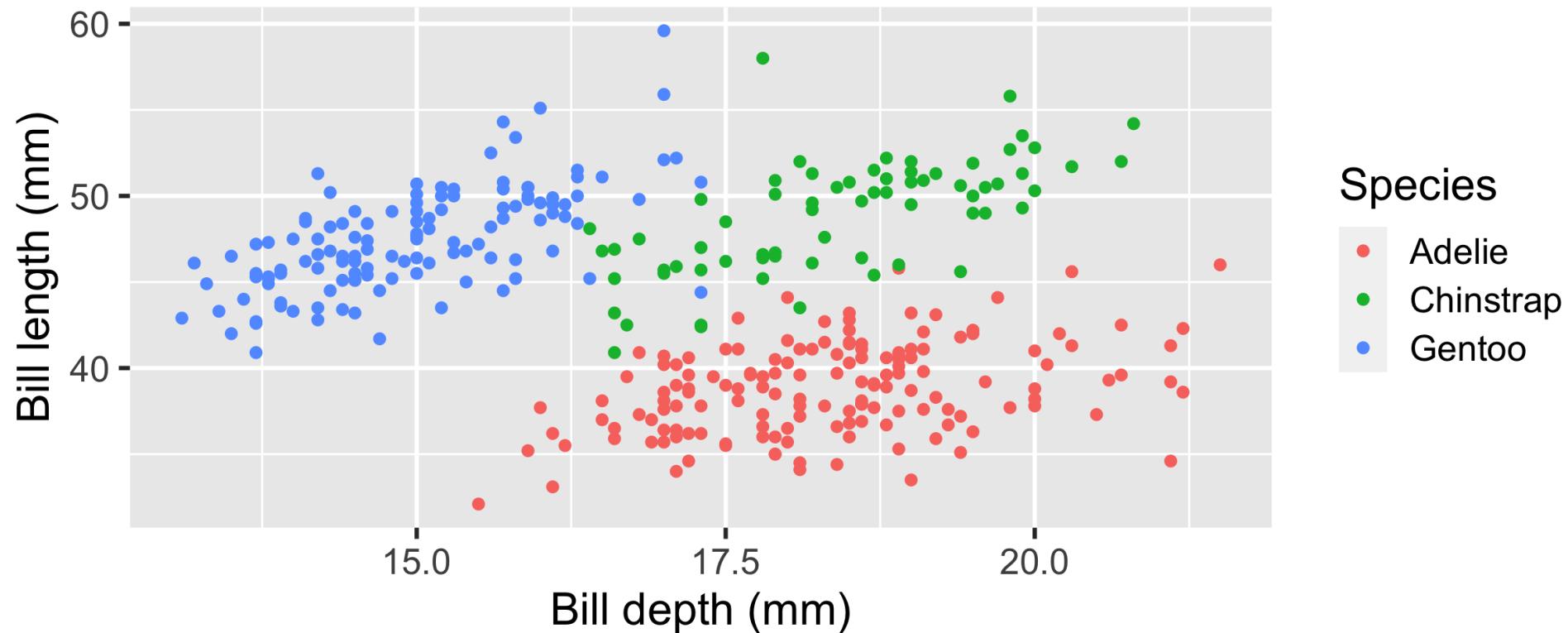
\$ body  
\$ son

58

```
<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



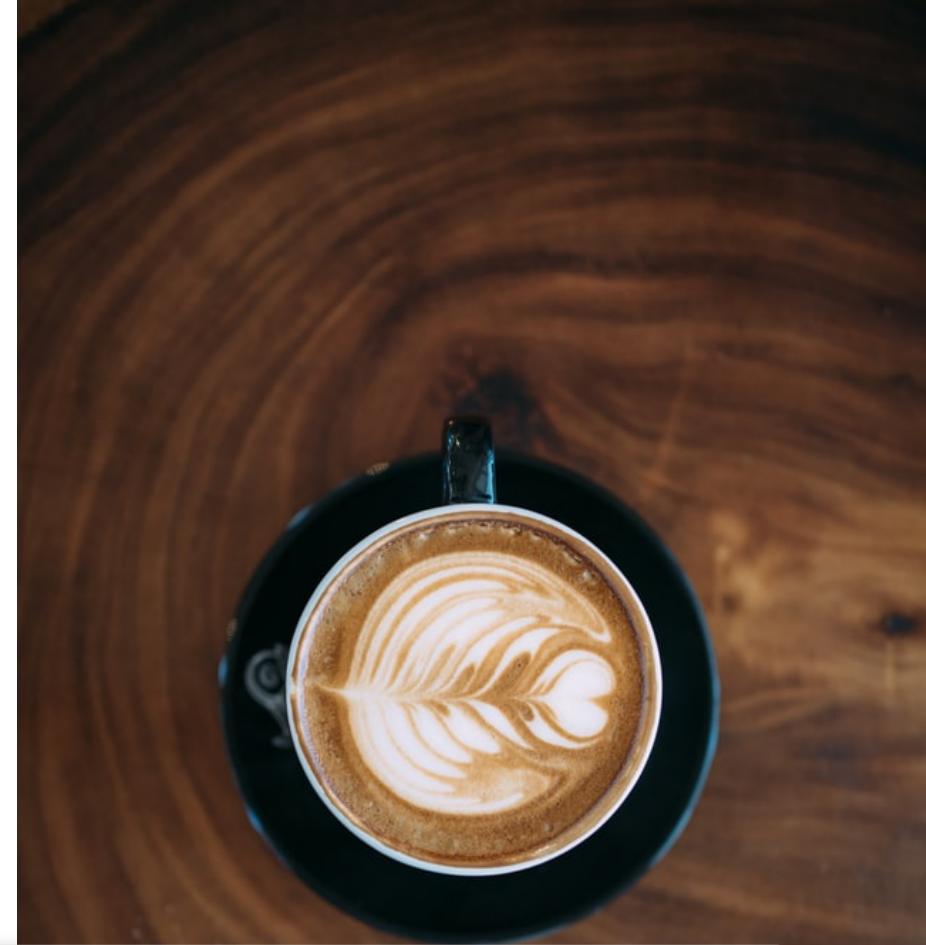
# Praktikum 3 - {ggplot2}

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

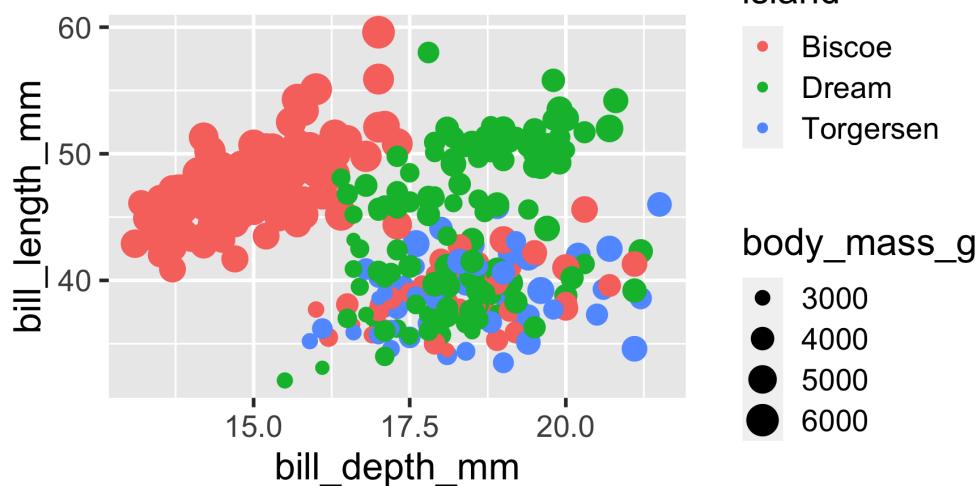
Photo by: [Blake Wisz](#)



# Mapping und Faceting

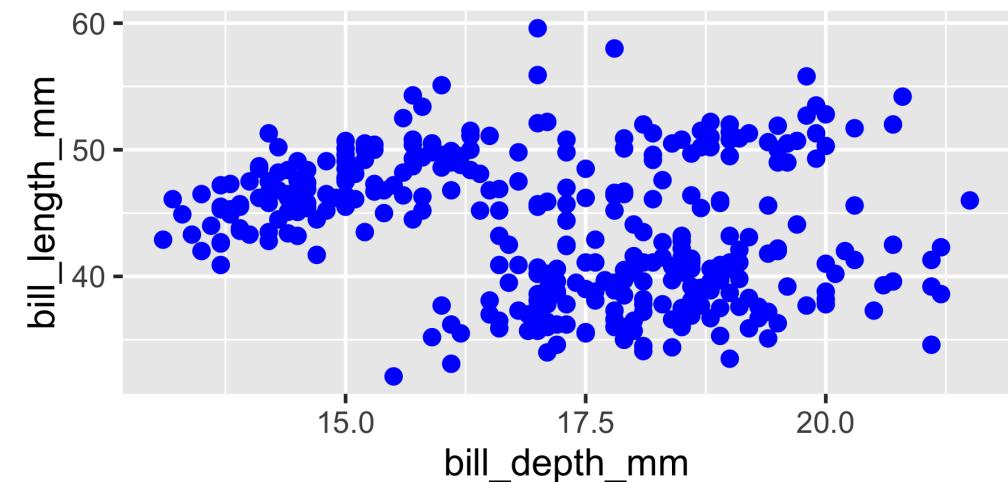
**Mapping:** Grösse, Alpha, Farbe, etc. der Punkte werden basierend auf den Werten der Variable in den Daten bestimmt.

```
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 3 - {ggplot2}

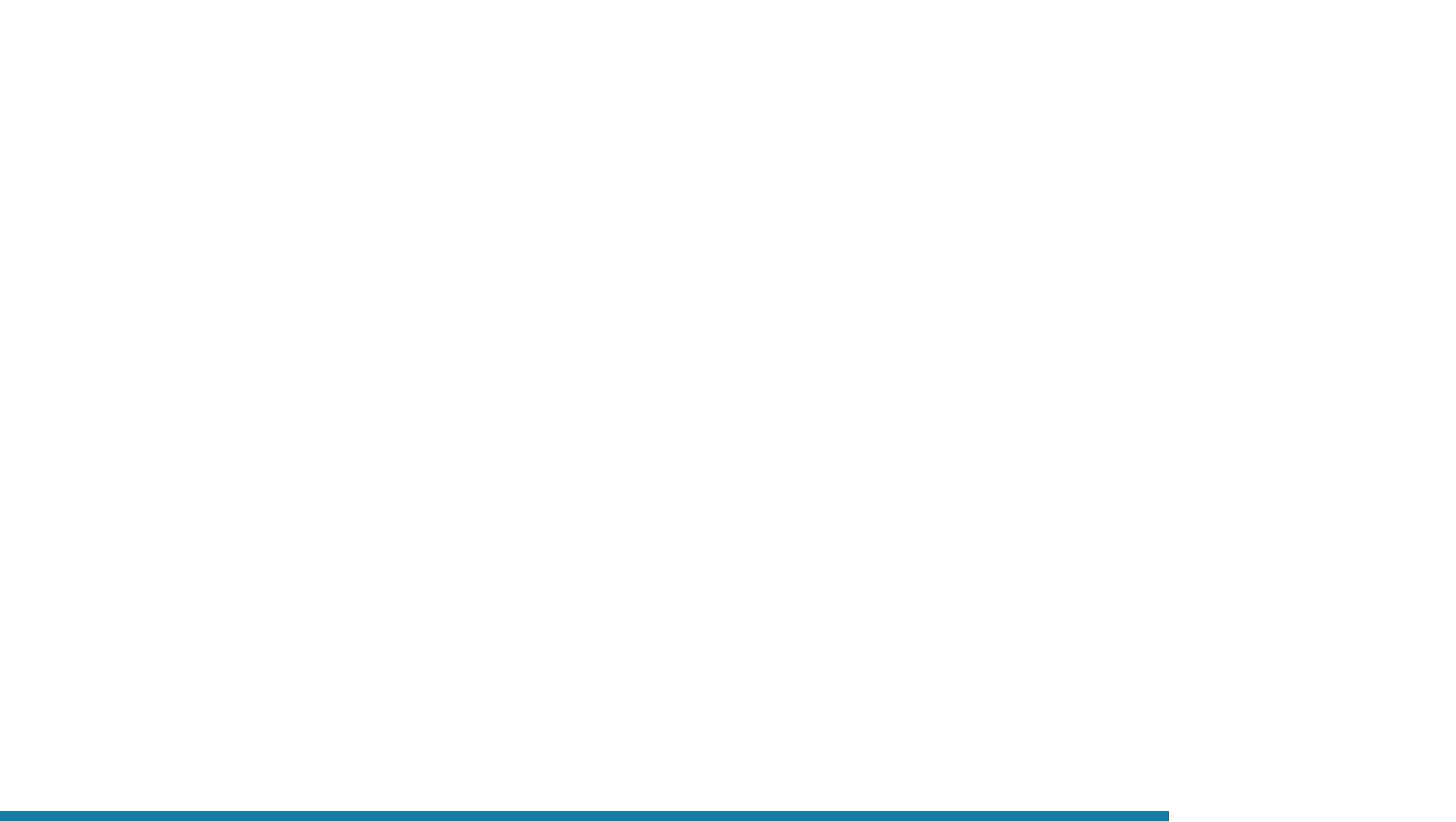
1. **RStudio Cloud**: Öffne deinen Arbeitsbereich für den Kurs in der RStudio Cloud
2. **RStudio Cloud / Projects**: Öffne erneut das Praktikum 03
3. Folgt wieder auf dem Bildschirm

# Eine Skala ersetzen - Farbskalen (color)

```
ggplot(penguins, aes(x = bill_length_mm, y = bill_depth_mm, colour = species)) +  
  geom_point() +  
  scale_color_viridis_d()
```

# Eine Skala ersetzen - Farbskalen (fill)

```
ggplot(penguins, aes(x = species, fill = island)) +  
  geom_bar() +  
  scale_fill_viridis_d()
```



# Praktikum 4 - {ggplot2}

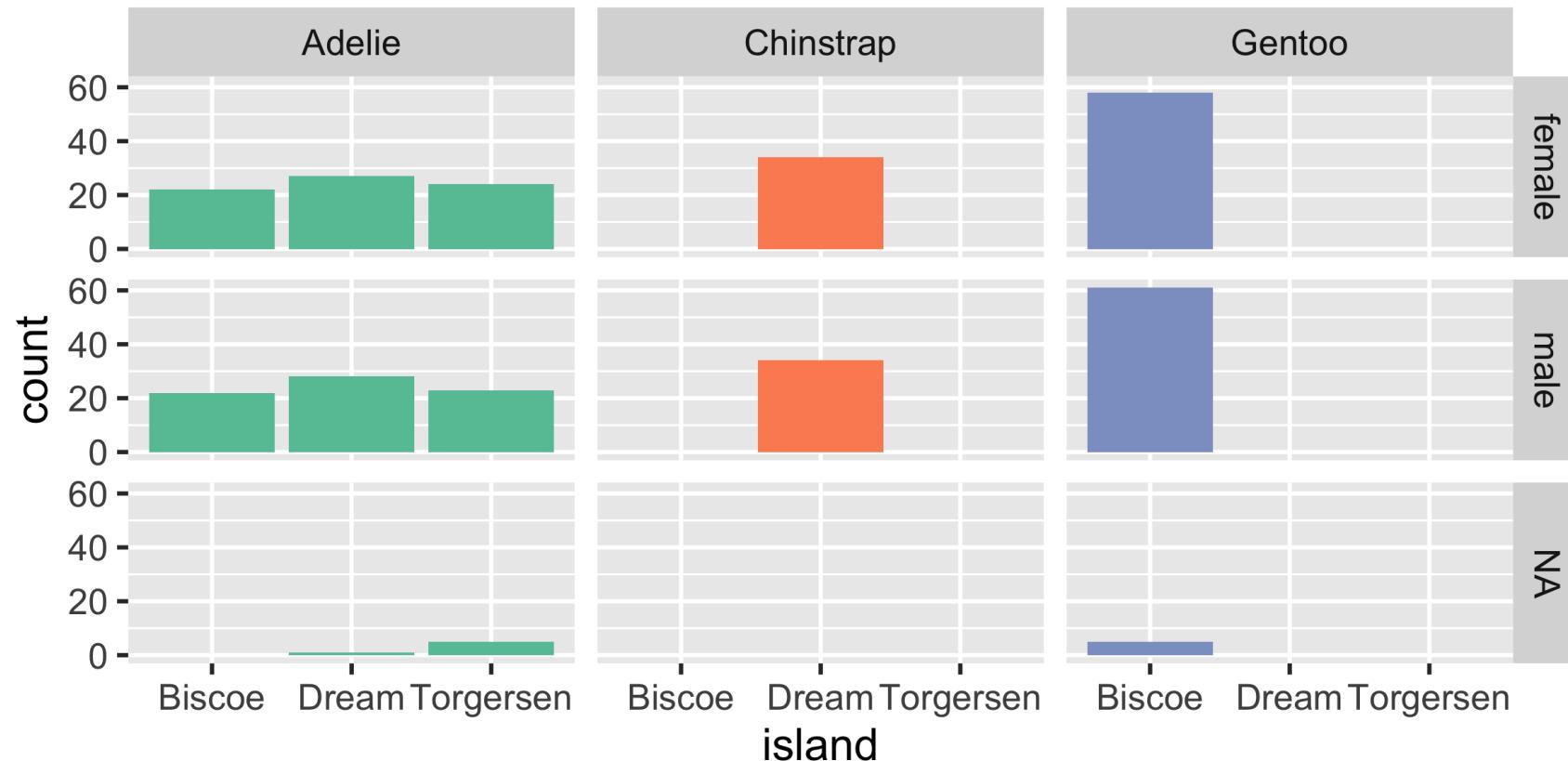
Arbeitet in 2er Teams. Eine Person teilt den Bildschirm. Die andere schreibt mit.

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](https://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](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.