

rstatsZH - Data Science mit R

Willkommen

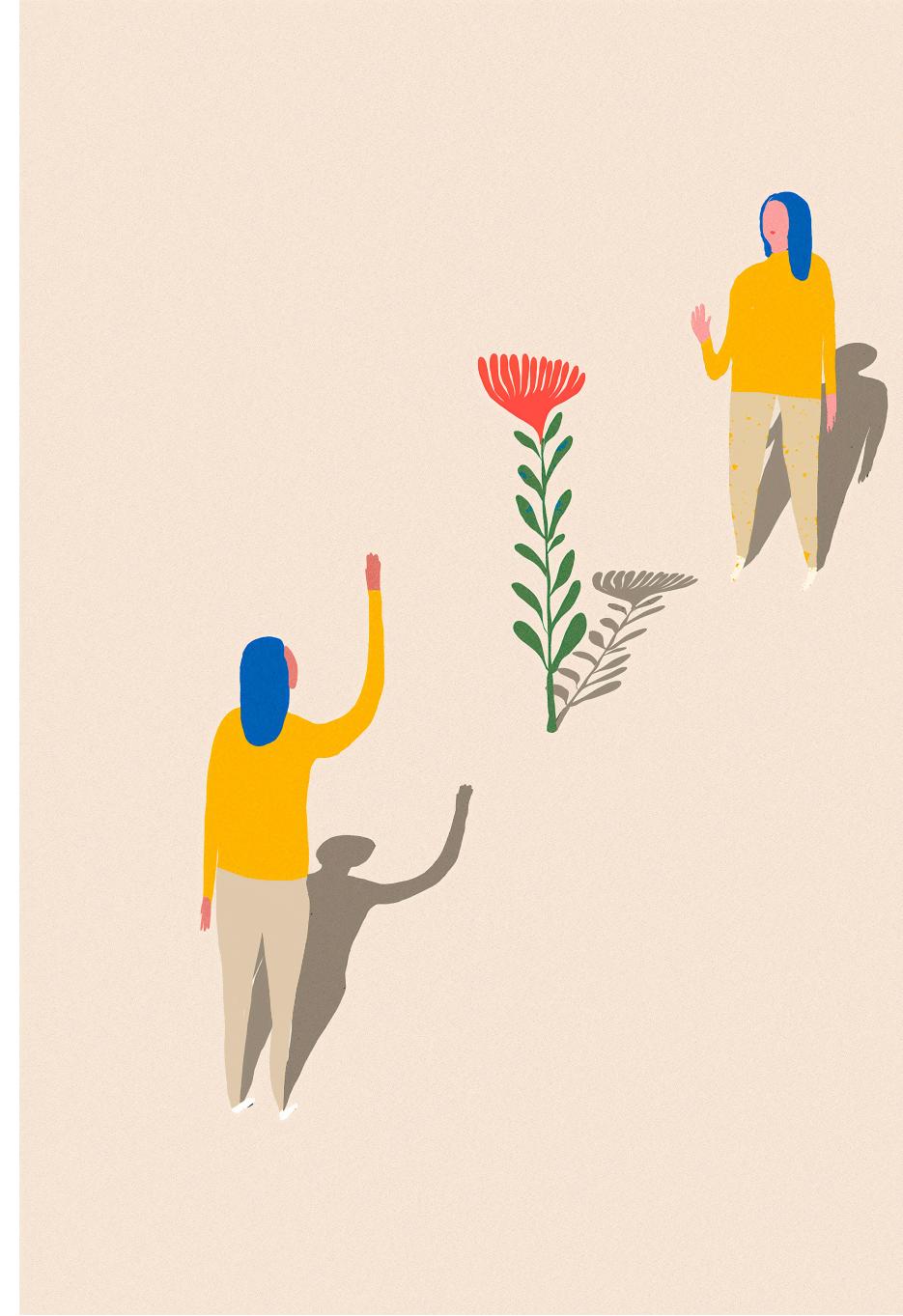
Lars Schöbitz

2021-10-25

Hallo R! 

Hallo Ihr!

Image created by Danielo Baretto



Ziele für heute

Am Ende des Tages:

- haben wir uns kennengelernt
- habt ihr die Werkzeugkiste kennengelernt
- könnt ihr beschreiben was die Ziele des Kurses sind
- könnt ihr einen ersten Plot erstellen



Kennenlernen

Stellt euch jeweils gegenseitig in 2er Gruppen vor (pro Person 2 Minuten)

- Mögliche Fragen
 - In welcher Verwaltungseinheit bist du?
 - Was für Erfahrungen hast du mit R?
 - Welche Ziele hast du für den Kurs?
 - Was machst du gern ausserhalb der Arbeit?

05 : 00

Kennenlernen

In der Gruppe: Stellt eure Arbeitspartner_in vor

Wer bin ich?

- Lars Schöbitz
 - Umweltingenieur
 - Toiletten-Forscher
 - Datenwissenschaftler
- R Nutzer seit 2016
- R Trainer seit 2020
- Website: lse.de
- Twitter: [@larnsce](https://twitter.com/larnsce)



Kursinfo

Kursinfo

- **Unterricht:** 3 Stunden Sessions (Pause jede Stunde)
- **Übungen:** Praktika während der Session
- **Hausaufgaben:** Einzureichen bis zum nächsten Kurstag
- **Persönliches Tutoring:** Sprechstunden ausserhalb der Kurstage
- **Abschluss Projekt (freiwillig):**
 - Arbeit mit Daten eurer Wahl
 - Bericht
 - 5 Minuten Präsentation
- **Zertifikat:** Bestätigung über Teilnahme

Werkzeugkiste



Werkzeugkiste

Kursarbeit

- Zoom
- Slack
- <https://rstatszh.github.io/website/>

Data Science

Programmieren

- R
- RStudio (Cloud)
- tidyverse
- R Markdown

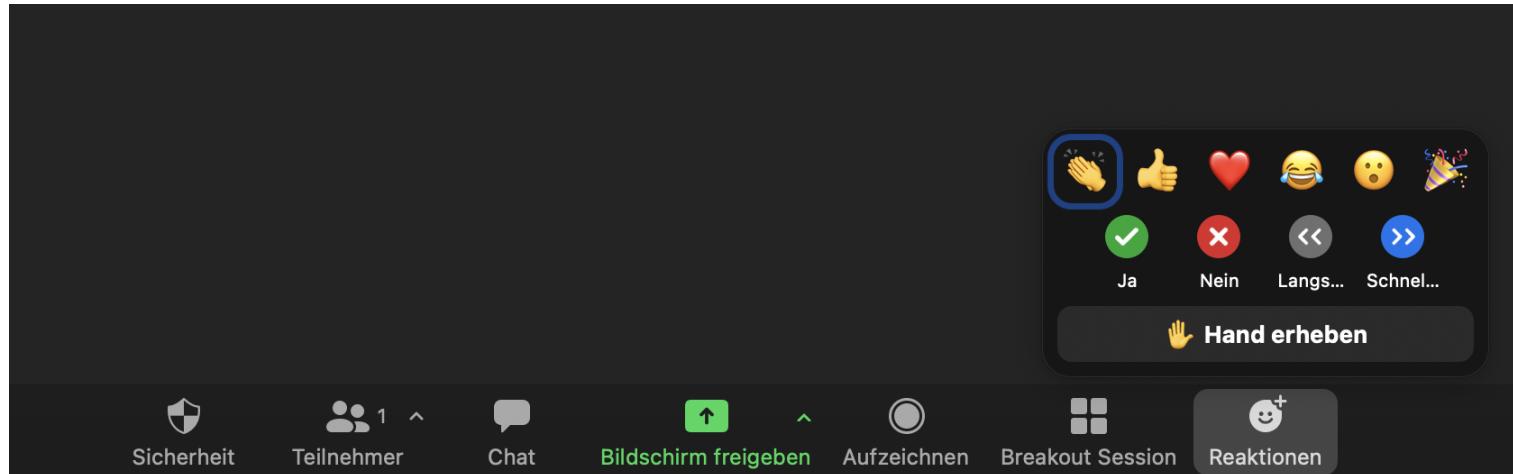
Versionskontrolle und Kollaboration

- Git
- GitHub

Zoom - Unsere Konferenzplatform

Tipps für die Nutzung

- Das Mikrofon auf stumm schalten
- Wenn ihr mögt, das Video gern einschalten
- Zoom Chat für Fragen
- Bei Fragen, bitte "Hand heben"
- Nutzt die Reaktionen für Feedback



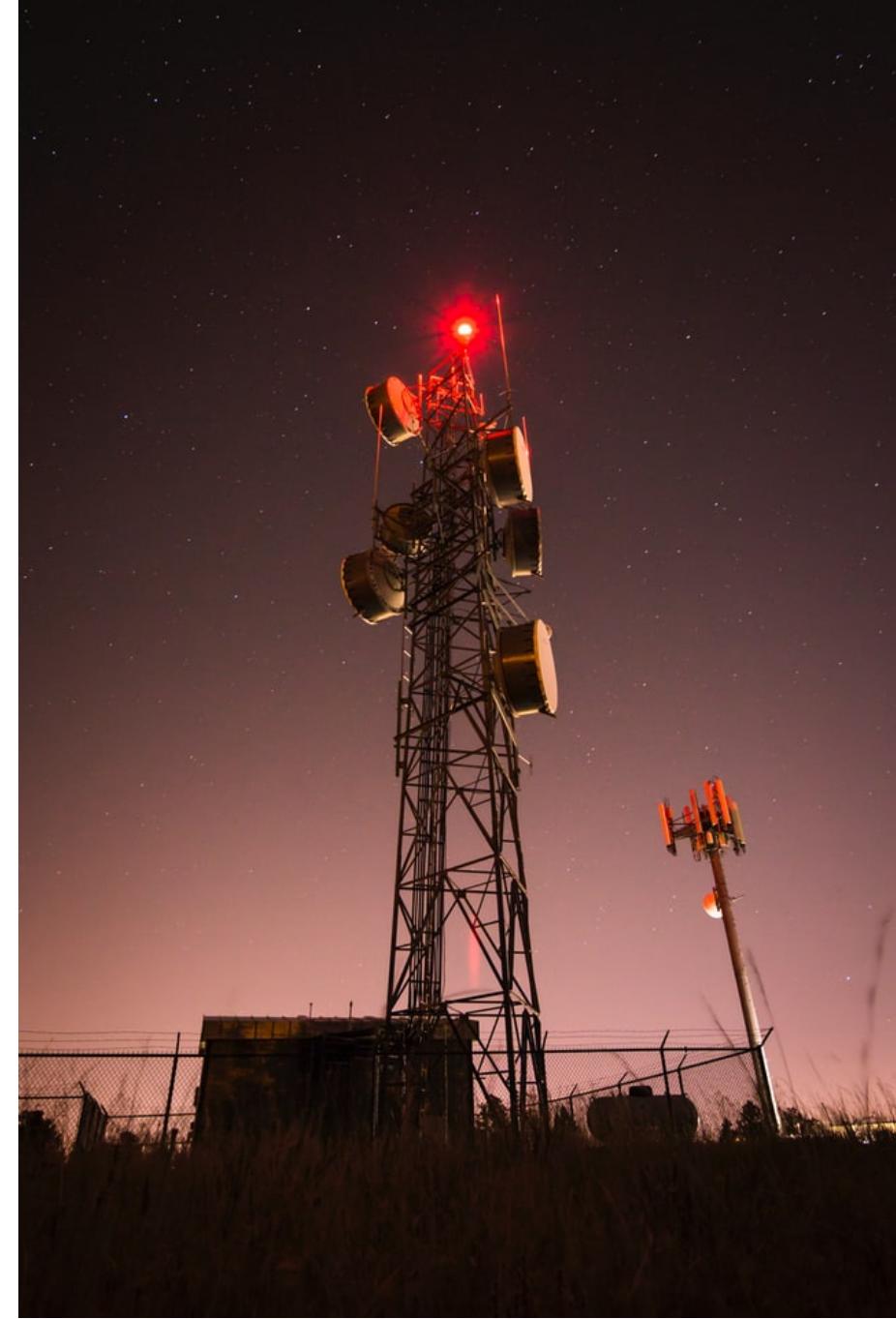
Slack

Unser Channel für die Kommunikation

Was gehört hinein?

- Fragen
- Unterstützung beim Programmieren
- Technische Probleme
- Ankündigungen
- Unterhaltungen

Photo by: Steve Halama



Data Science

Data Science - Lernziele

Am Ende dieses Kurses kannst du:

- (Daten)projekte reproduzierbar publizieren
- kollaborativ mit git und GitHub arbeiten
- das Konzept von tidy data verstehen und anwenden
- explorative Datenanalysen durchführen
- dir selbst und Anderen helfen mehr über die grosse Welt der Möglichkeiten für Data Science mit R zu lernen

Reproduzierbare Datenanalyse

- Denkt über folgende Frage für 2 Minuten nach
- Danach, teilt eure Ansichten in 2er Gruppen für 5 Minuten
- Wichtig: Wir haben alle ein anderes mentales Modell

Frage: Was macht eine Datenanalyse reproduzierbar?

02 : 00

- Denkt über folgende Frage für 2 Minuten nach
- Danach, teilt eure Ansichten in 2er Gruppen für 5 Minuten
- Wichtig: Wir haben alle ein anderes mentales Modell

Frage: Was macht eine Datenanalyse reproduzierbar?

Kurzfristige Ziele

- Können Tabellen und Grafiken aus den Daten und dem Code neu erstellt (reproduziert) werden?
- Kann Code aus einem Skript in einem Anderen wieder verwendet werden?
- Stimmt meine Arbeitsumgebung mit der meiner Kolleg_innen überein?

rstatsZH - moderne Nutzer_innen

Moderne Nutzer_innen zu sein bedeutet, Fähigkeiten, Gewohnheiten und Denkweisen für eine reproduzierbare und kollaborative Arbeitsweise zu entwickeln.

Wir lernen:

1. Coding mit bewährten Verfahren und Skripts (R/RStudio/tidyverse)
2. Reproduzierbare Dokumente schreiben und veröffentlichen (R Markdown)
3. Kollaboratives Arbeiten und Versionsverwaltung (Git/GitHub)

R and RStudio

R und RStudio



- R ist eine freie (open-source) Programmiersprache, erschienen in 1993
- R ist eine Arbeitsumgebung für statistische Analysen und Grafiken
- R kann mit *packages* erweitert werden



- RStudio ist eine grafische Benutzeroberfläche, welche auch integrierte Entwicklungsumgebung (i.e. **IDE** - integrated development environment) genannt wird
- RStudio IDE wird entwickelt von **RStudio, PBC** (Public Benefit Corporation), eine US-amerikanische Firma

R packages

- **Packages** sind die grundlegenden Einheiten von reproduzierbarem R-Code
- Packages enthalten wiederverwendbare R Funktionen, die Dokumentation welche beschreibt wie sie verwendet werden, und Beispieldaten
- Mit Stand Februar 2021 gibt es mehr als 17'000 R-Pakete auf **CRAN** (dem Comprehensive R Archive Network)¹
- Wir werden mit einem kleinen (aber wichtigen) Teil

¹ CRAN contributed packages.

R Community



R Community

- Twitter: #rstats
- Online learning community @R4DScommunity
- Minority R Users @miR_community
- We are @R-Ladies
- Zurich R User Group @ZurichRUsers
- RStudio Community



Verhaltenskodex¹

- Diversität und Inklusion
- Namen und Pronomen
- Lernprozess

Photo by: [Sharon McCutcheon](#)

¹ https://rstatszh.github.io/website/code_of_conduct.html.



Tour: R, RStudio, R Markdown

RstatsZH-K002 / prak-01-visualisieren



Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

R 4.0.3

prak-01.Rmd x

ABC



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

5:22 # Daten visualisieren

R Markdown

Console Terminal x Jobs x

/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 with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in 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.

>

Environment History Connections Git Tutorial

Import Dataset

List

R Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

Name	Size	Modified
..		
.gitignore	731 B	Mar 3, 2021, 2:46 PM
.Rhistory	0 B	Mar 3, 2021, 2:46 PM
prak-01-visualisieren.Rproj	205 B	Mar 5, 2021, 2:27 PM
prak-01.Rmd	3.7 KB	Mar 3, 2021, 2:46 PM
README.md	1.1 KB	Mar 3, 2021, 2:46 PM

RstatsZH-K002 / prak-01-visualisieren



Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Menü Leiste

R 4.0.3

prak-01.Rmd x

ABC 🔎 Knit ⚙️ Run ⚙️ ↕ Global Environment

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

Environment History Connections Git Tutorial

Import Dataset

R Global Environment

Environment is empty

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 with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in 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.

> |

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

Name	Size	Modified
..		
.gitignore	731 B	Mar 3, 2021, 2:46 PM
.Rhistory	0 B	Mar 3, 2021, 2:46 PM
prak-01-visualisieren.Rproj	205 B	Mar 5, 2021, 2:27 PM
prak-01.Rmd	3.7 KB	Mar 3, 2021, 2:46 PM
README.md	1.1 KB	Mar 3, 2021, 2:46 PM

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

ABC 🔎 Knit ⚙️ Run ⚙️ Global Environment

```
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 is empty

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 with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in 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.

> |

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

Name	Size	Modified
..		
.gitignore	731 B	Mar 3, 2021, 2:46 PM
.Rhistory	0 B	Mar 3, 2021, 2:46 PM
prak-01-visualisieren.Rproj	205 B	Mar 5, 2021, 2:27 PM
prak-01.Rmd	3.7 KB	Mar 3, 2021, 2:46 PM
README.md	1.1 KB	Mar 3, 2021, 2:46 PM

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 with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in 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.

> |

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

Name	Size	Modified
..		
.gitignore	731 B	Mar 3, 2021, 2:46 PM
.Rhistory	0 B	Mar 3, 2021, 2:46 PM
prak-01-visualisieren.Rproj	205 B	Mar 5, 2021, 2:27 PM
prak-01.Rmd	3.7 KB	Mar 3, 2021, 2:46 PM
README.md	1.1 KB	Mar 3, 2021, 2:46 PM

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

- Menu Bar (Orange border):** Contains the standard RStudio menu items: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help.
- Code Editor (Pink border):** Displays 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 labeled "Code Editor" is overlaid on the editor area.
- Environment (Blue border):** Shows the R environment with tabs for Environment, History, Connections, Git, and Tutorial. A blue box labeled "Environment Git" is overlaid on the Git tab.
- Console (Grey border):** Displays the R startup message and information about the R project.
- File Manager (Green border):** Shows the project structure with files like ".git", ".Rhistory", "prak-01.Rmd", "prak-01.html", and "README.md". A green box labeled "Datei Manager (Files)" is overlaid on the left side of the file list.

Overlaid text labels identify the sections:

- Menü Leiste (Orange box)
- Code Editor (Pink box)
- Environment Git (Blue box)
- Datei Manager (Files) (Green box)

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 contains the following items: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help.
- Code Editor:** The left panel, highlighted with a pink border, displays an R Markdown file named "prak-01.Rmd". The code includes a title, date, and output settings, followed by a note about working with a dataset. A large pink box labeled "Code Editor" covers the bottom half of this panel.
- Environment:** The second panel from the top, highlighted with a blue border, shows the R environment with tabs for Environment, History, Connections, Git, and Tutorial. It includes sections for Import Dataset and Global Environment. A large blue box labeled "Environment Git" covers the bottom half of this panel.
- Console:** The third panel from the top, highlighted with a teal border, shows the R console output. It includes the standard R welcome message, information about the license, and details about the current project. A large teal box labeled "Console" covers the bottom half of this panel.
- File Manager:** The bottom panel, highlighted with a green border, shows the file structure of the project. It includes New Folder, Upload, Delete, Rename, and More options. A large green box labeled "Datei Manager (Files)" covers the bottom half of this panel, with additional labels "Plots", "Help File", and "Viewer" appearing below it.

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 R 4.0.3 in the top right of the environment panel.

RStudio Cloud

rstudio.cloud/spaces/122222/project/2262471

RstatsZH-K002 / prak-01-visualisieren

Fenster 1: RStudio Cloud

File Edit
prak-01.Rmd
85
86 - # Üb
87 1. F
88 2. E
89 schau nach wie du einen Boxplot erstellt hast. Was könnte hier fehlen?
90 3. Führe den Code aus
91 4. Warum wurde das Argument 'color' durch 'fill' ersetzt?
(*Tipp:** Probiere aus das Argument 'fill' wieder durch 'color' zu ersetzen)
92
93 **Lösch diesen Text und beschreibe hier was du gemacht hast**
94
95 ~`{r penguin-vis-
96 ggplot(data = peng
97 mapping = c(bill_l
98
99
100 ~`
101 ~`
102
103 - # Übung 6
104
105 - Füge deinen Namen und das Datum zum YAML Header hinzu (**Tipp:** Zwischen die drei gestrichelten Linien)
106 - Klicke auf 'Knit' und betrachte dein fertiges Dokument
107
108 Super gemacht! Wir sind fertig hier.

Hier schreibst du Code

Übung 5 R Markdown

Console Terminal Jobs

/cloud/project/

You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in 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.



Praktikum - Willkommen

- Öffne rstudio.cloud und login (Link ist im Zoom Chat)
- Unter "Projects", klicke auf **Start** neben **prak-00-willkommen**
- Folge meinen Anweisungen im Zoom Video

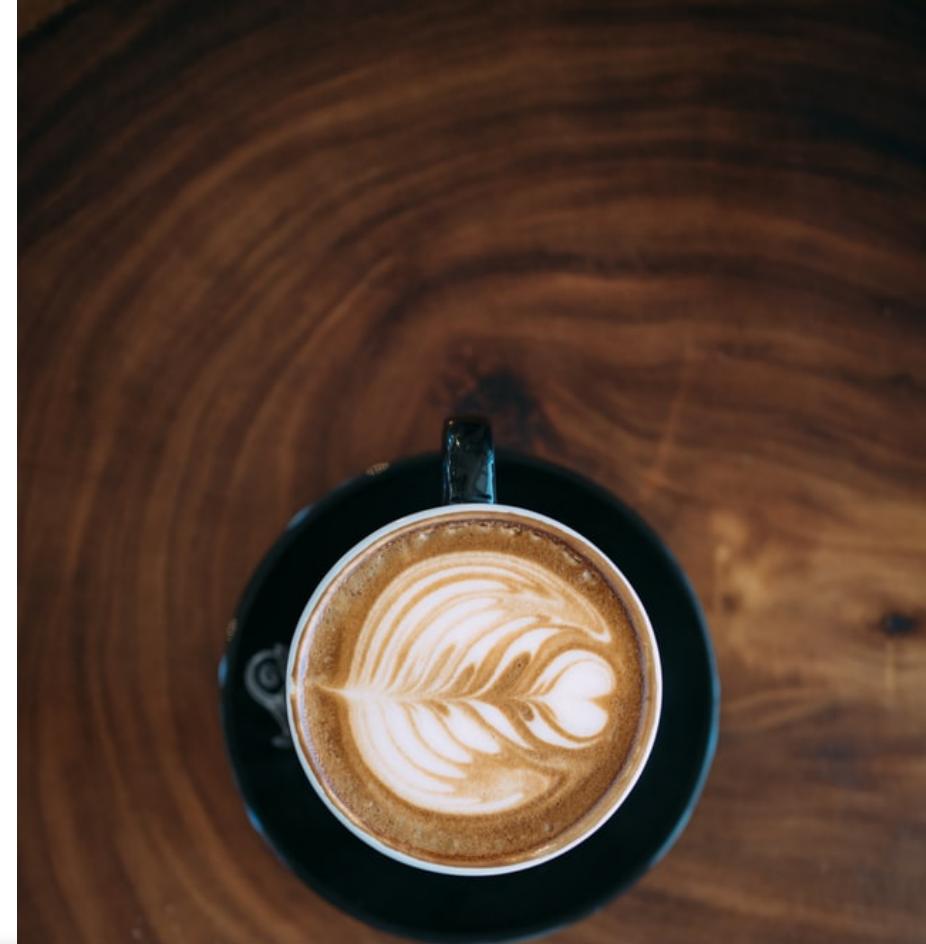
Wie nutzen wir R Markdown?

- Jede Aufgabe / Bericht / Projekt / etc. ist ein R Markdown Dokument
- Du wirst immer eine Vorlage für ein R Markdown Dokument haben, mit der du beginnen kannst.
- Die Menge an vorgebenen Hilfen in der Vorlage wird im Laufe des Kurses abnehmen

Pause

10 : 00

Photo by: [Blake Wisz](#)



Frage: Wobei handelt es sich um eine Zahl?

Wobei handelt es sich um eine Zahl:

1. 10
2. ten
3. "ten"
4. "10"

Schickt die Antwort in den Zoom Chat.

Frage: Wobei handelt es sich um eine Zahl?

1. 10 
2. ten --- **Objekt**
3. "ten" --- **Wort (string)**
4. "10" --- **Wort (string)**

R - Wesentliches

R - Wesentliches

Funktionen

- Funktionen sind (häufig) Verben, gefolgt vom dem worauf sie angewendet werden

```
mach_dies(mit_dem)
mach_dies(mit_dem, auf_das, mit_diesen)
```

Packages

- Packages werden einmalig der Funktion `install.packages` installiert und einmal pro Sitzung mit der `library` Funktion geladen:

```
install.packages("package_name")
library(package_name)
```

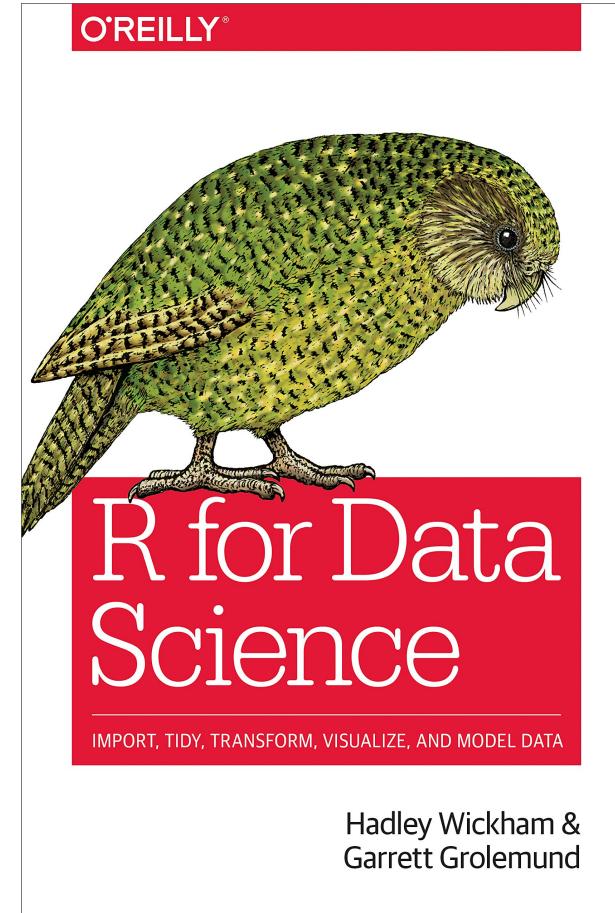
tidyverse.org

- Ein logisch zusammenhängendes System von R Packages zur
 - Daten Aufbereitung
 - Daten Exploration
 - Daten Visualisierung
 - Daten Modellierung
- Eine Designphilosophie, Grammatik und Daten Struktur
- Entwickelt um Produktivität zu unterstützen



Buch - R for Data Science

- Das Buch für den Kurs
- Kostenfrei auf
<https://r4ds.had.co.nz/>
- Geschrieben von Hadley Wickham
- Nutzt die Philosophie des Tidyverse



Praktikum 1

Praktikum - Daten Visualisierung

Breakout rooms (2er Gruppen)

- Eine Person teilt den Bildschirm
- Öffne "Projects" in RStudio Cloud Workspace
- Klicke auf **Start** neben **prak-01-visualisieren**

30 : 00

Pause



07:30



Praktikum - Daten Visualisierung

Bericht

- Hat die Zeit gereicht?
- Was hat Schwierigkeiten bereitet?
- Was habt ihr gelernt?

Werkzeugkiste



Werkzeugkiste

Kursarbeit

- Zoom
- Slack
- Google Docs
- <https://statistikzh.github.io/rstatsZH/>

Data Science

Programmieren

- R
- RStudio (Cloud)
- tidyverse
- R Markdown

Versionskontrolle und Kollaboration

- Git
- GitHub

Git und GitHub

Git und GitHub

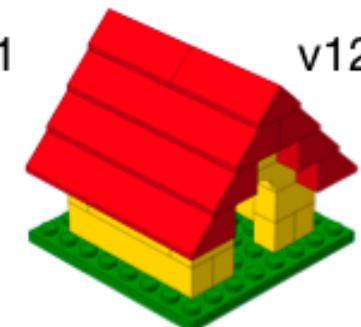
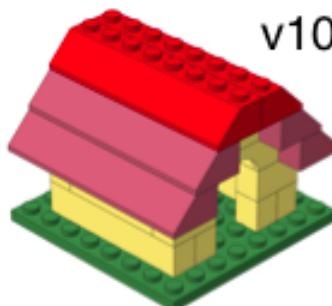
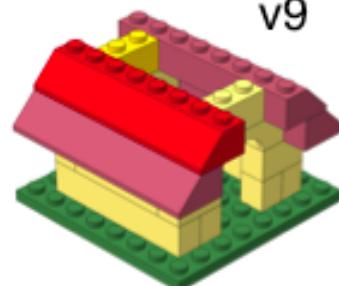
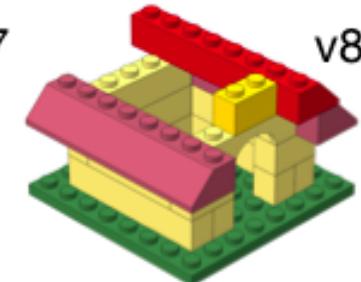
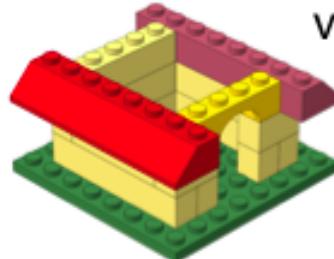
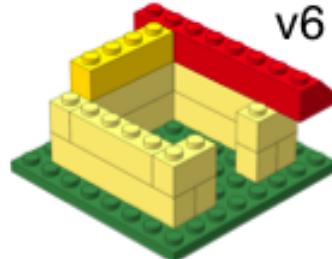
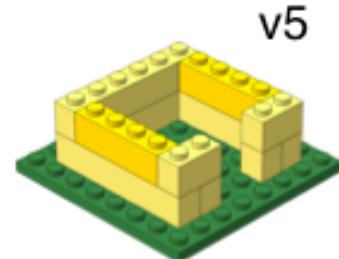
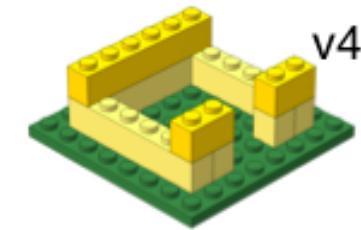
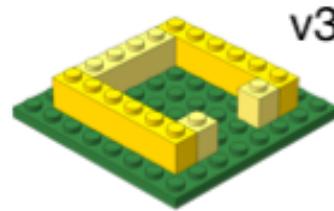
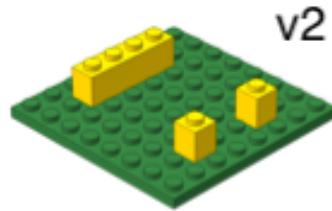
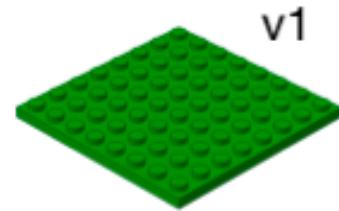


- Git ist ein System zur Versionsverwaltung - ähnlich wie "Track Changes" in Microsoft Word
- Es ist nicht das einzige System, aber ein sehr beliebtes



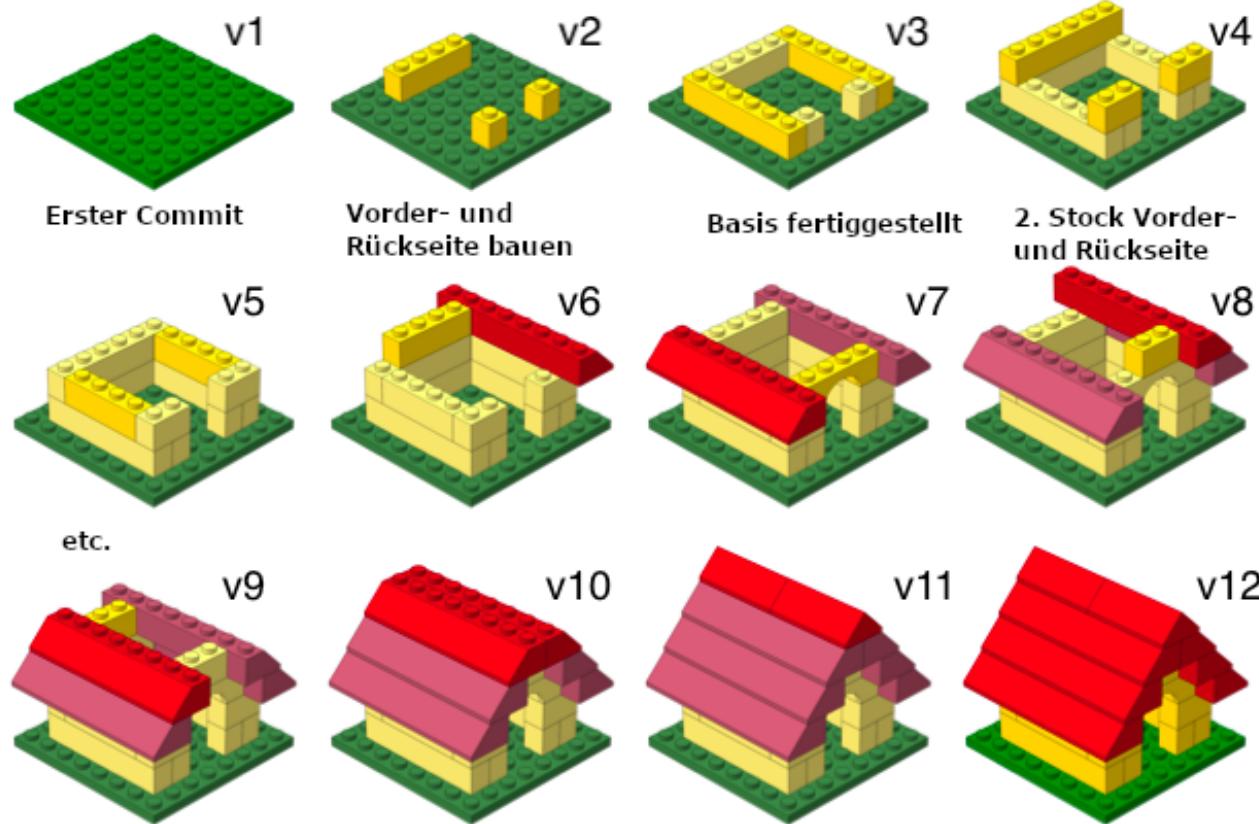
- GitHub ist das Zuhause für deine Git-basierten Projekte im Internet - ähnlich wie Dropbox
- Wir werden GitHub als Plattform für das Webhosting und die Zusammenarbeit nutzen (und als unser Kursmanagement System!)

Versionierung

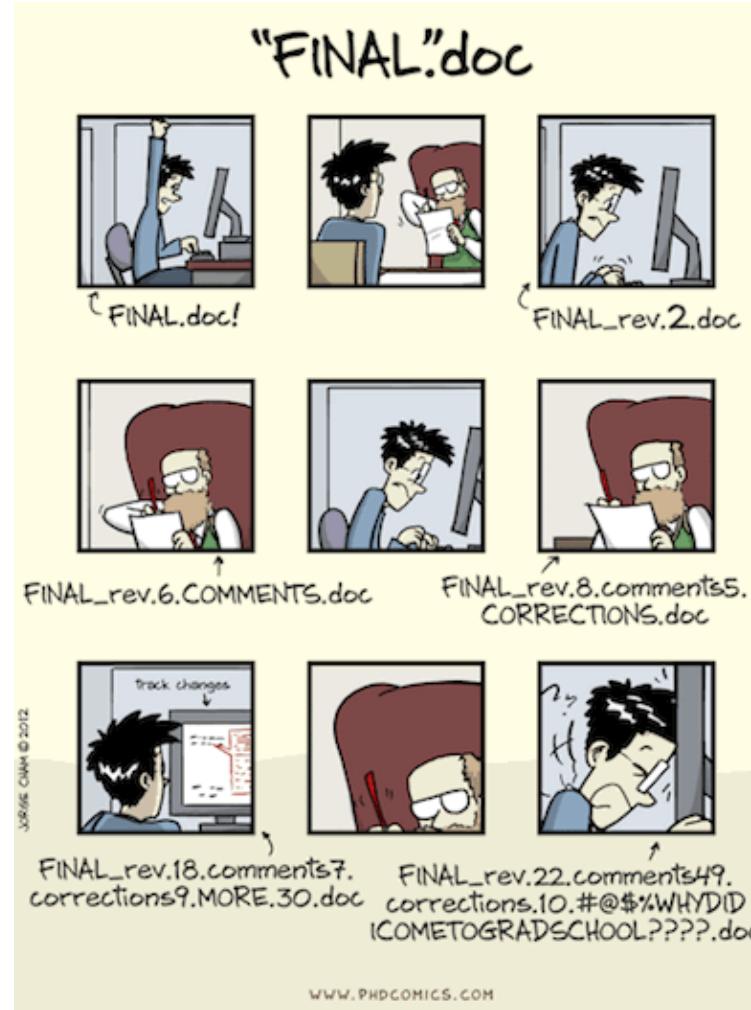


Versionierung

mit (von Menschen) lesbaren Nachrichten



Warum ist Versionsverwaltung nützlich?



Tour: GitHub

Wie nutzen wir Git und GitHub?

Hausaufgabe

Hausaufgabe

- Öffnet jetzt eure E-Mail Inbox
- Bestätigt mir im Zoom Chat, dass ihr eine Email mit Betreff "**rstatsZH - Lars hat das Repo ha01-GitHubName**" erhalten habt
- Die Anweisungen für die Heausaufgabe 01 findet ihr oben rechts auf unserer Kurswebseite
- Kontaktiert mich unter der Woche jederzeit auf Slack

Feedback

Ziele erreicht?

Bitte ausfüllen: kutt.it/rstatszh-eval





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_d01-willkommen/e1_d01-willkommen.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.