

rstatsZH - Data Science mit R

Willkommen

Lars Schöbitz

2020-02-08

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:**
 - Team Arbeit mit Daten eurer Wahl
 - Bericht (1 Seite)
 - 5 Minuten Präsentation
- **Zertifikat:** Bestätigung über Teilnahme und Aufwand

Werkzeugkiste



Werkzeugkiste

Kursarbeit

- Zoom
- Slack
- (Google Docs)
- <https://rstatszh.github.io/website/>

Data Science

Programmieren

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

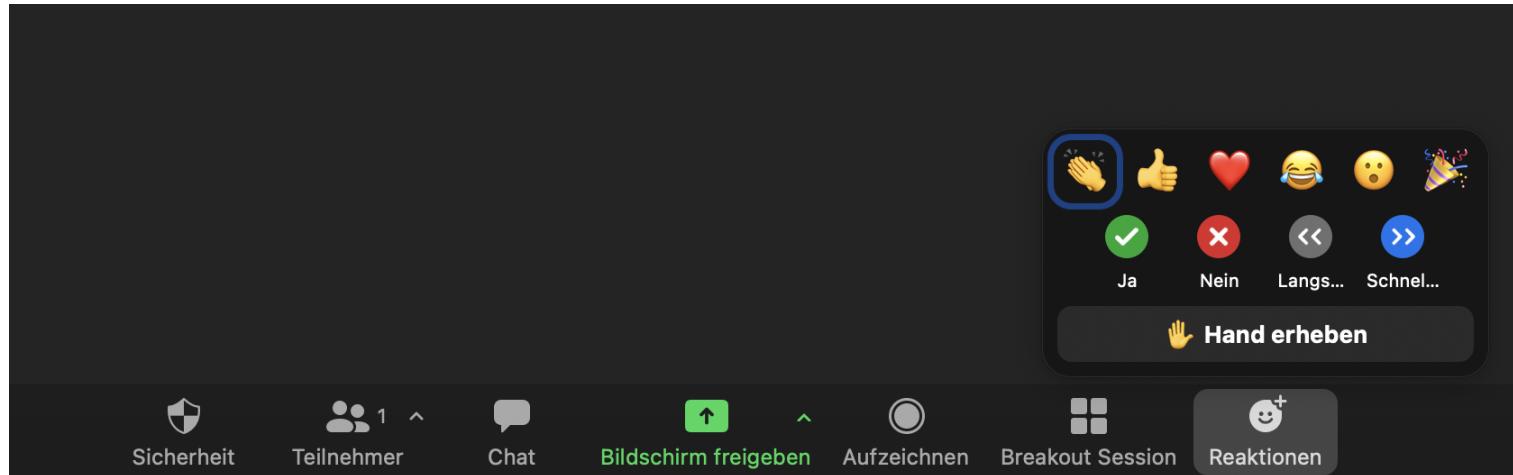
Versionskontrolle und Kollaboration

- 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



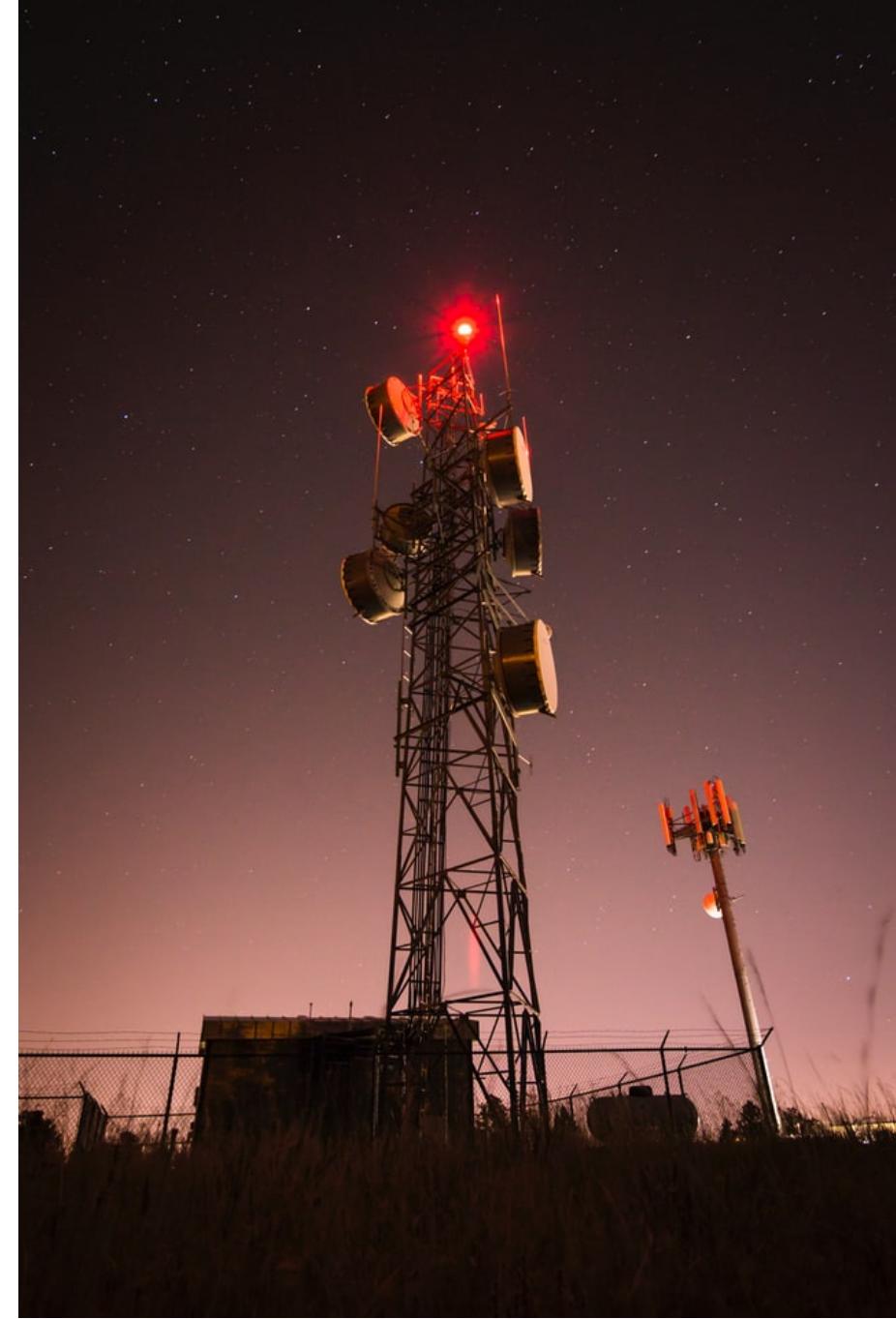
Slack

Unser Channel für die Kommunikation

Was gehört hinein?

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

Photo by: Steve Halama



Google Docs

Unser geteiltes Notizbuch (ein Experiment)

Wozu nutzen wir es?

- Für bestimmte Übungen und Aufgaben
- Zum Teilen von Bildern

KeyCastr

- MacOS
- Schweizer Keyboard Layout
- cmd ⌘ (Mac) = Ctrl (Windows)
- ⌘ (Mac) = Alt (Windows)



Wer von euch nutzt Windows? Antwortet mit "ja" als Reaktion in Zoom.

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 (einfache lineare Modellierung machen hier einen kleinen Teil aus)
- 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 - moderner NutzeR

Ein **moderner NutzeR** zu werden 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



Artwork by [@allison_horst](#)

R Community

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



Verhaltenskodex

- Diversität und Inklusion
- Namen und Pronomen
- Lernprozess

Photo by: [Sharon McCutcheon](#)



Tour: R, RStudio, R Markdown

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

R Markdown Hilfe

R Markdown Cheat Sheet Help -> Cheatsheets

R Markdown :: CHEAT SHEET

What is R Markdown?

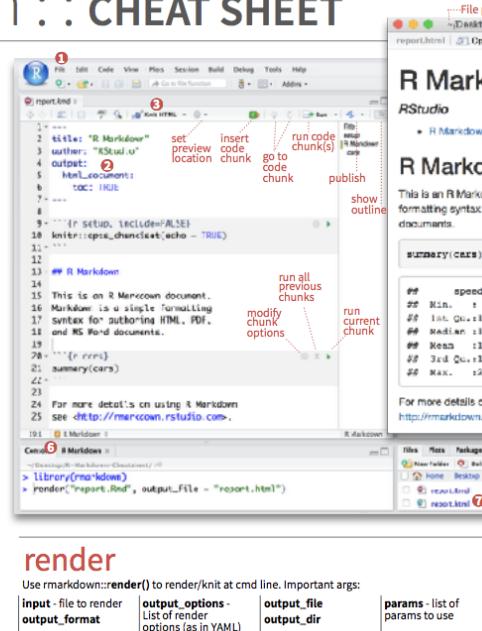
.Rmd files • An R Markdown (.Rmd) file is a record of your research. It contains the code that a scientist needs to reproduce your work along with the narration that a reader needs to understand your work.

Reproducible Research • At the click of a button, or the type of a command, you can run the code in an R Markdown file to reproduce your work and export the results as a finished report.

Dynamic Documents • You can choose to publish your work in a variety of formats, including html, pdf, MS Word, or RTF documents; html or pdf based slides, Notebooks, and more.

Workflow

- ① Open a new .Rmd file at File ► New File ► R Markdown. Use the wizard that opens to pre-populate the file with a template
- ② Write document by editing template
- ③ Knit document to create report; use knit button or render() to knit
- ④ Preview Output in IDE window
- ⑤ Publish (optional) to web server
- ⑥ Examine build log in R Markdown console
- ⑦ Use output file that is saved along side .Rmd



render

Use rmarkdown::render() to render/knit at cmd line. Important args:

input - file to render	output_options - List of render options (as in YAML)	output_file	params - list of params to use
------------------------	--	-------------	--------------------------------

Embed code with knitr syntax

INLINE CODE
Insert with `r <code>`. Results appear as text without code.
Built with `r getRVersion()`

CODE CHUNKS
One or more lines surrounded with `{{r}}` and `{{ }}`. Place chunk options within curly braces, after r. Insert with
`{{r echo=TRUE}}`
`getRVersion()`
`{{ }}`

GLOBAL OPTS
Set with knitr::opts_chunk
`{{r include=F}}`
`knitr::opts_chunk
`{{ }}`

Markdown Quick Reference Help -> Markdown Quick Reference

The screenshot shows the R Markdown Quick Reference page within the RStudio interface. The top navigation bar includes Files, Plots, Packages, Help, and Viewer. The main content area is titled "Markdown Quick Reference" and contains sections for Emphasis, Headers, Lists, and Manual Line Breaks, each with examples of the syntax.

Emphasis

italic **bold**
italic bold

Headers

Header 1
Header 2
Header 3

Lists

Unordered List

- * Item 1
- * Item 2
 - + Item 2a
 - + Item 2b

Ordered List

1. Item 1
2. Item 2
3. Item 3
 - + Item 3a
 - + Item 3b

Manual Line Breaks

End a line with two or more spaces:
Roses are red,
Violets are blue.

Links

Use a plain http address or add a link to a phrase:
<http://www.rstudio.com>

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

05 : 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)
```

R - Wesentliches

- Auf Spalten (Variablen) in Dataframes wird mit `$` zugegriffen:

```
# Beispiel  
dataframe$var_name  
  
# Praxis  
penguins$species
```

- Auf die Dokumentation von Objekten, Packages, Funktionen, etc. kann mit `?` zugegriffen werden

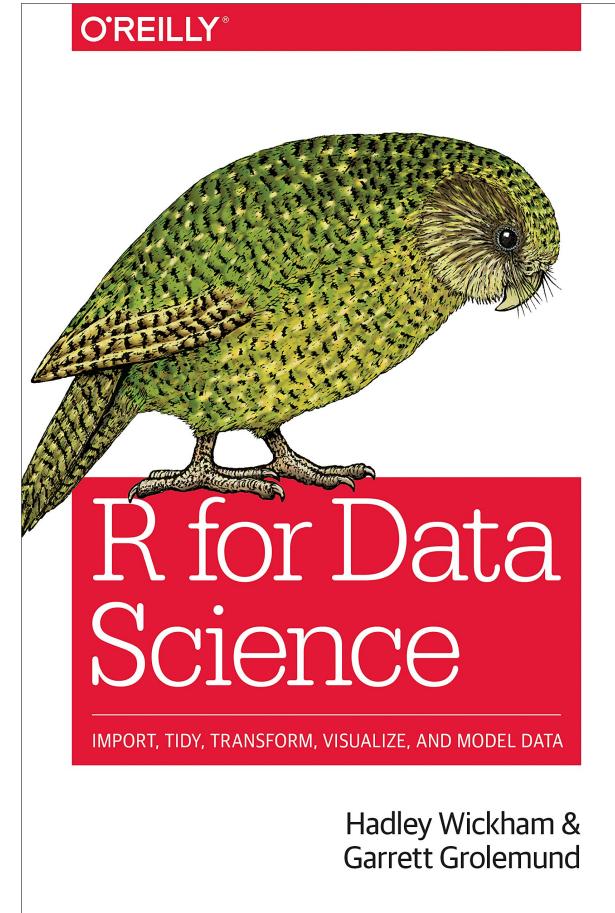
tidyverse.org

- 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

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

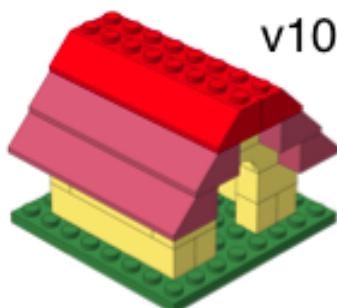
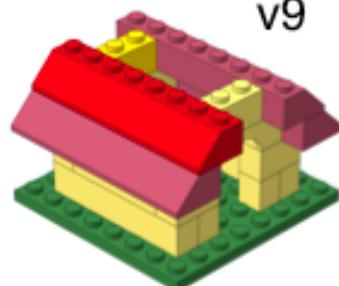
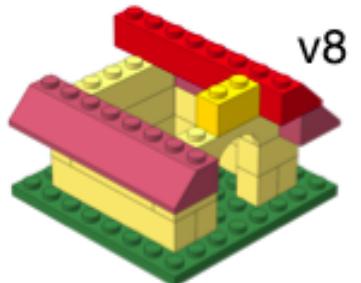
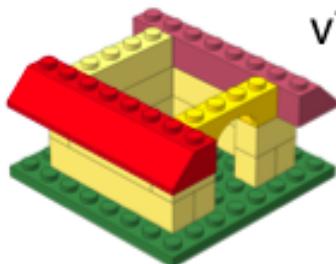
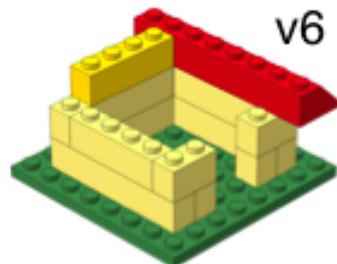
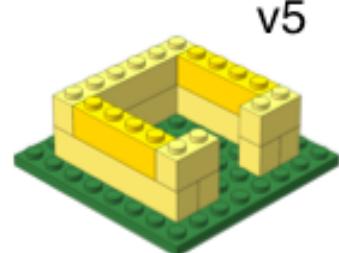
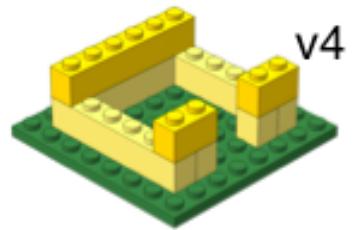
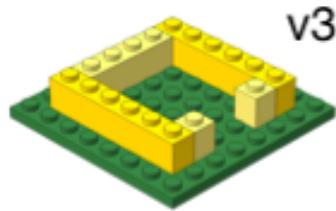
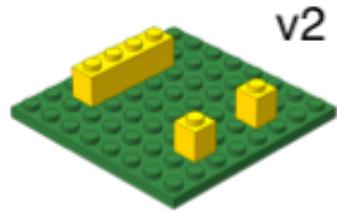
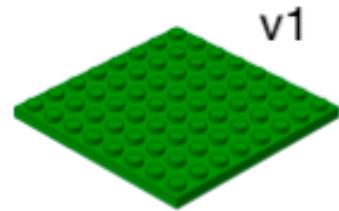


- Git ist ein System zur Versionsverwaltung - ähnlich wie "Track Changes" in Microsoft Word
- Es ist nicht das einzige System, aber ein sehr beliebtes



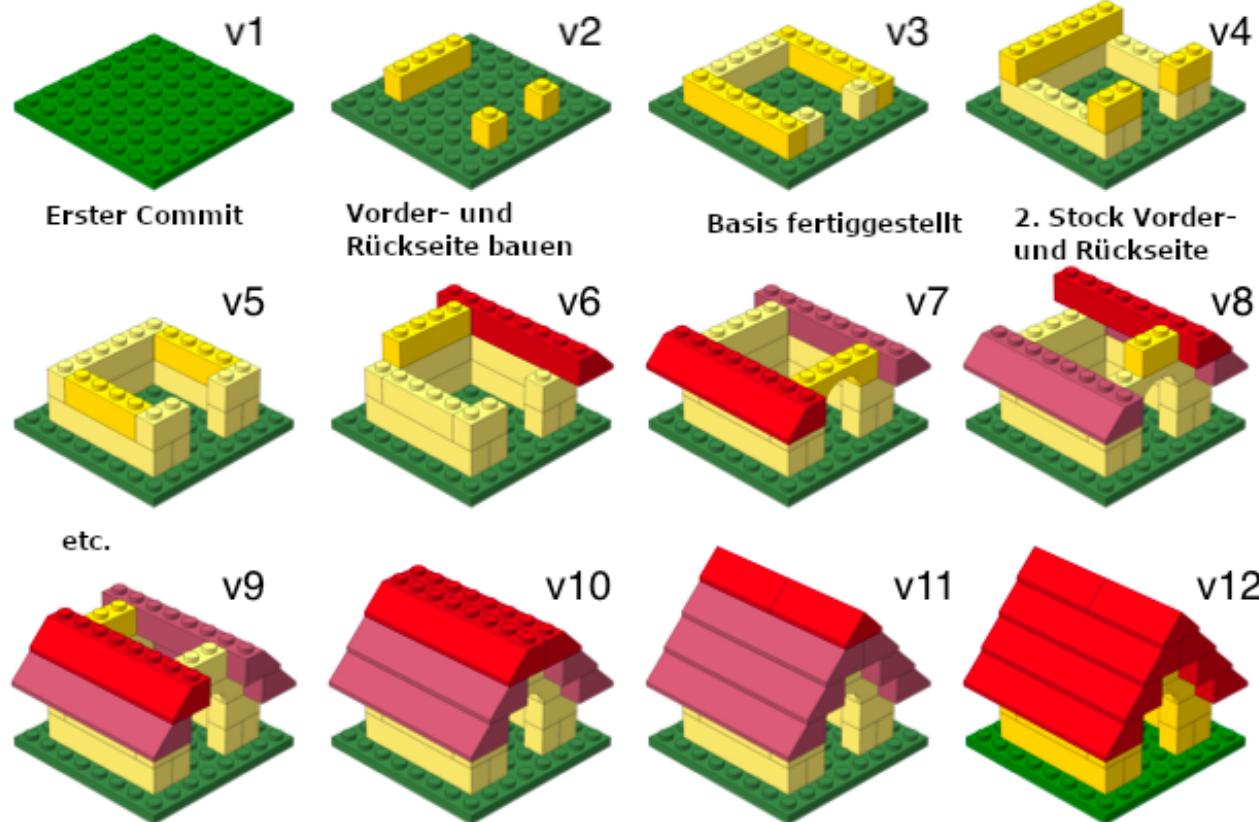
- GitHub ist das Zuhause für deine Git-basierten Projekte im Internet - ähnlich wie Dropbox
- Wir werden GitHub als Plattform für das Webhosting und die Zusammenarbeit nutzen (und als unser Kursmanagement System!)

Versionierung

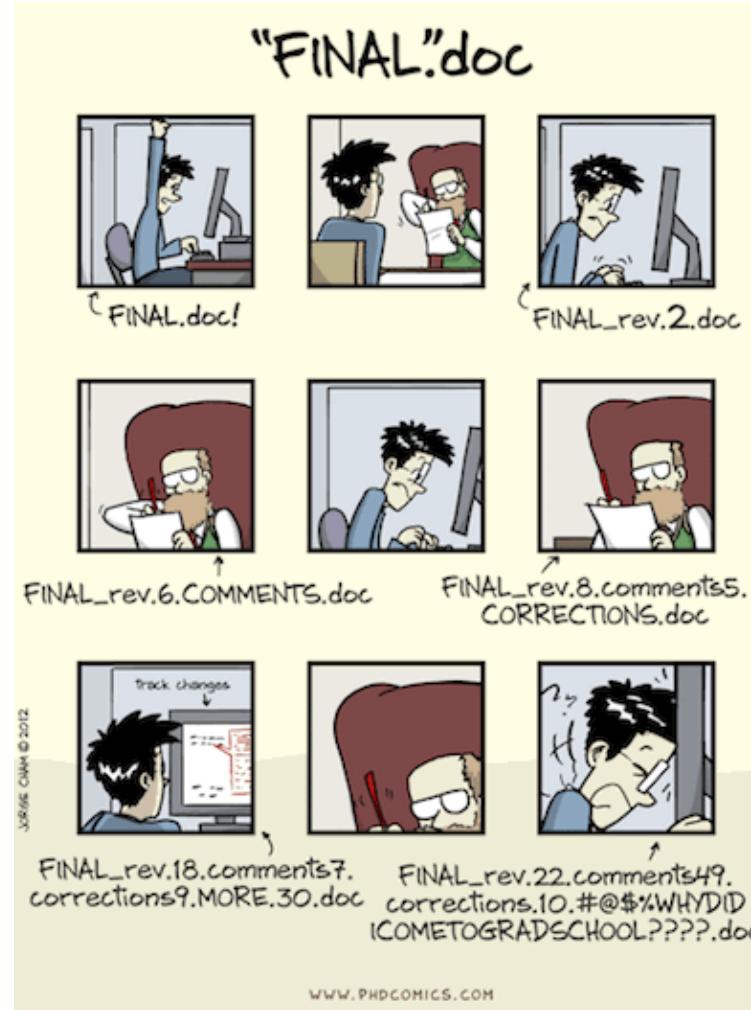


Versionierung

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 Heimarbeit 1 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.

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.