

Willkommen zum Kurs & Werkzeugkiste

rstatsZH - Data Science mit R

Lars Schöbitz

Feb 25, 2025

Willkommen!



Triff die Kursleitung

Lars Schöbitz



- Umweltingenieur 
- Forscher im Ruhestand 
- RStudio zertifizierte Lehrkraft
- Data Steward an der ETH Zurich

Lernziele des Kurses

Die folgenden Übergeordneten Lernziele sind für diesen Kurs definiert:

1. Teilnehmer:innen werden in der Lage sein, eine Reihe gängiger datenwissenschaftlicher Werkzeuge (R, RStudio IDE, Git, GitHub, R tidyverse packages, Quarto) zu verwenden, um die Ergebnisse von Datenanalyseprojekten zu veranschaulichen und öffentlich zu kommunizieren.
2. Teilnehmer:innen können das Quarto-Dateiformat und den visuellen Bearbeitungsmodus der RStudio IDE zu verwenden, um technisch-wissenschaftliche Dokumente mit Zitaten, Fußnoten, Querverweisen, Abbildungen und Tabellen zu erstellen.

3. Teilnehmer:innen können Git Versionskontrolle verwenden um:

(1) Änderungen an Dateien zu GitHub zu übertragen und von GitHub zu übernehmen, (2) Branches zu nutzen und Pull Requests auf GitHub zu öffnen, (3) mittels des GitHub Issue Tracker innerhalb eines Projekts zu kommunizieren und an Aufgaben zusammenzuarbeiten.

Du bist dran: Über Dich

- Wähle einen Gegenstand aus und mache dir eine Minute lang Notizen.
- Was hat der Gegenstand, den du ausgewählt hast, mit dem Grund zu tun, warum du hier bist?

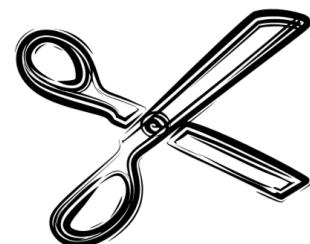
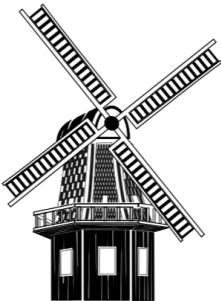


Images from: <https://openclipart.org/>

In Paaren

Nehmt euch jeweils 2 Minuten Zeit, um euch auszutauschen:

- Was hat der Gegenstand, den ihr ausgewählt habt, mit dem Grund eurer Anwesenheit zu tun?



Images from: <https://openclipart.org/>

Kurskalender

modul	datum	titel
Modul 01	25 Februar 2025	<u>Willkommen zum Kurs & Werkzeugkiste & Kollaboratives Arbeiten mit GitHub I</u>
Modul 02	04 März 2025	<u>Explorative Datenanalyse durch Visualisierung & Digitaler Arbeitsplatz (DAP) & R-Community</u>
Modul 03	11 März 2025	<u>Daten Transformation mit dplyr</u>
Modul 04	18 März 2025	<u>Daten Import & Daten Management & Kollaboratives Arbeiten mit GitHub II</u>
Modul 05	25 März 2025	<u>Bedingte Anweisungen & Datumswerte & OGD Ressourcen</u>
Modul 06	01 April 2025	<u>Daten Typen & Vektoren & Iteration I (for loops)</u>
Modul 07	08 April 2025	<u>Datenformen lang und breit & Daten zusammenfügen</u>
Modul 08	15 April 2025	<u>Mit Text Daten arbeiten & KTZH Corporate Design mit statR</u>
Modul 09	22 April 2025	<u>Erstellung von Berichten und Folien mit Quarto</u>
Modul 10	29 April 2025	<u>Funktionen schreiben & Parameterisierung von Berichten</u>
Abschlussprojekt	06 Mai 2025	Arbeitszeit für Abschlussprojekt
Abschlussprojekt	13 Mai 2025	Abgabe Abschlussprojekt
Abschlussveranstaltung	20 Mai 2025	Abschlussveranstaltung

Kursstruktur

- Ich bin dran: Vortragssegmente + Live-Coding
- Wir sind dran: Live-Coding + Mitmachen
- Ihr seid dran: Übungen in Paaren

Ich bin dran: Vortragssegmente + Live-Coding

- Der Kursleiter schreibt und spricht den Code laut vor
- Der Kursleiter erklärt Konzepte und Prinzipien, die relevant sind
- Code wird auf dem Bildschirm angezeigt

Wir sind dran: Live-Coding + Mitmachen

- Der Kursleiter schreibt und spricht den Code laut vor
- Der Kursleiter erklärt relevante Konzepte und Prinzipien
- Der Code wird auf dem Bildschirm angezeigt
- Die Lernenden machen mit, indem sie denselben Code schreiben und ausführen

Ihr seid dran: Übungen in Paaren

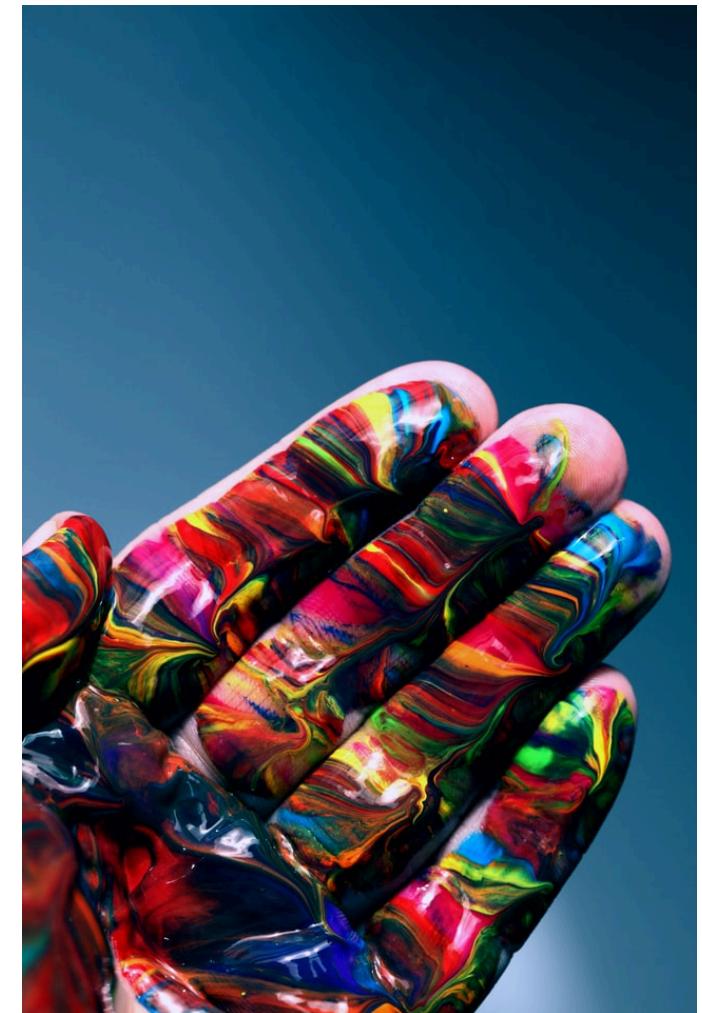
- Zwei Lernende arbeiten in Paaren in Break-out Räumen zusammen
- Eine Person tippt und teilt die Bildschirmanzeige
- Eine Person bietet Kommentare und Vorschläge an, und tippt selbst auch mit

Unterstützung bekommen / Fragen stellen

- Während Live-Coding: hebt die Hand mittels Zoom und unterbrecht mich. Ich werde versuchen, mich um dein Anliegen zu kümmern, wenn es die Zeit erlaubt. Offene Fragen werden dokumentiert.
- Während Übungen in Break-out Räumen: Ihr könnt mich zu euch rufen. Bei einem Grossteil der Übungen werde ich versuchen in jedem Raum einmal vorbeizuschauen.

Verhaltenskodex

- Diversität und Inklusion
- Namen und Pronomen
- Lernprozess
- rstatszh-k010.github.io/website/verhaltenskodex.html



Website zum Kurs

rstatszh-k010.github.io/website/

Wir sind dran: Öffne und speichere die Kurswebsite als Lesezeichen

1. Öffne einen Webbrowser auf deinem Arbeitsgerät.
2. Gehe zur Website des Kurses: rstatszh-k010.github.io/website/
3. Erstelle einen Lesezeichenordner mit dem Namen rstatszh-k010 in deiner Lesezeichenleiste.
4. Füge dem Ordner rstatszh-k010 das Lesezeichen zur Kurs-Webseite hinzu.

Lernziele (für diese Woche)

1. Die Lernenden können ein GitHub Issue öffnen und den Kursleiter markieren.
2. Die Lernenden können ein Repository von GitHub klonen und das GitHub PAT verwenden, um einen Commit von ihrem lokalen Repository auf GitHub zu übertragen.
3. Die Lernenden können vier Komponenten einer Quarto-Datei identifizieren (YAML, Code Chunk, R-Code, Markdown).
4. Die Lernenden können die vier Fenster der RStudio IDE identifizieren.



Quarto

Wir holen euch dort ab, wo ihr seid

Posit Cloud & RStudio IDE

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

rstudio.cloud/spaces/122222/project/2262471

RstatsZH-K002 / prak-01-visualisieren

RAM Settings More RT 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 Knit Run

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
: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

RStudio Cloud

x [rstatszh-k002/ha01-rainbow-tr](#) x | +

→ C [rstudio.cloud/spaces/122222/project/2262471](#)

RAM ⚙️ RT Rainbow Train

RstatsZH-K002 / prak-01-visualisieren

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

prak-01.Rmd x

ABC Knit R Markdown

```
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  
:22 # Daten visualisieren
```

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.

Menü Leiste

File Plots Packages Help Viewer

New Folder Upload Delete Rename More

	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

RStudio Cloud

x [rstatszh-k002/ha01-rainbow-tr](#) | +

→ C [rstudio.cloud/spaces/122222/project/2262471](#)

RAM ⚙️ RT Rainbow Train

Menü Leiste

File Edit Code View Plots Session Build Debug Profile Tools Help

Addins Go to file/function Run

Code Editor

```
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  
:22 # Daten visualisieren R Markdown
```

In dieser ersten Übung arbeiten wir mit einem Datensatz über

Environment History Connections Git Tutorial

Import Dataset Global Environment

Environment is empty

Console Terminal Jobs

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

RStudio Cloud

x [rstatszh-k002/ha01-rainbow-tr](#) x | +

→ C [rstudio.cloud/spaces/122222/project/2262471](#)

RAM Settings More Rainbow Train

RstatsZH-K002 / prak-01-visualisieren

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Code Editor

```
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  
:22 # Daten visualisieren
```

In this first exercise, we will work with a dataset about

Environment Git

Environment History Connections Git Tutorial

Import Dataset Global Environment

Files

New Folder Upload Delete Rename More

Cloud > project

	Name	Size	Modified
	..		
<input type="checkbox"/>	.gitignore	731 B	Mar 3, 2021, 2:46 PM
<input type="checkbox"/>	.Rhistory	0 B	Mar 3, 2021, 2:46 PM
<input type="checkbox"/>	prak-01-visualisieren.Rproj	205 B	Mar 5, 2021, 2:27 PM
<input type="checkbox"/>	prak-01.Rmd	3.7 KB	Mar 3, 2021, 2:46 PM
<input type="checkbox"/>	README.md	1.1 KB	Mar 3, 2021, 2:46 PM

File Plots Packages Help Viewer

rstudio.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.

@ [rstatszh-k010.github.io/website/](#)

27

RStudio Cloud

x  rstatszh-k002/ha01-rainbow-tr x | +

rstudio.cloud/spaces/122222/project/2262471

RAM  Settings  More  RT Rainbow Train

Menü Leiste

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Code Editor

```
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  
:22 # Daten visualisieren
```

R Markdown

Environment Git

Environment History Connections Git Tutorial

Import Dataset Global Environment

Datei Manager (Files)

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

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

2:46 PM
2:46 PM
2:27 PM
2:46 PM
2:46 PM

28

@ rstatszh-k010.github.io/website/

RStudio Cloud

x rstatszh-k002/ha01-rainbow-tr | +

rstudio.cloud/spaces/122222/project/2262471

RAM

RT Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

prak-01.Rmd x

ABC Knit Run

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
11 # Daten visualisieren R Markdown

Code Editor

Environment History Connections Git Tutorial

Import Dataset Global Environment

Environment Git

Console Terminal Jobs

cloud/project/

This free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

This is a collaborative project. Type 'contributors()' for more information on how to contribute.
Type 'citation()' for citation information, and 'publication()' for 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

Datei Manager (Files)

Plots Help File Viewer

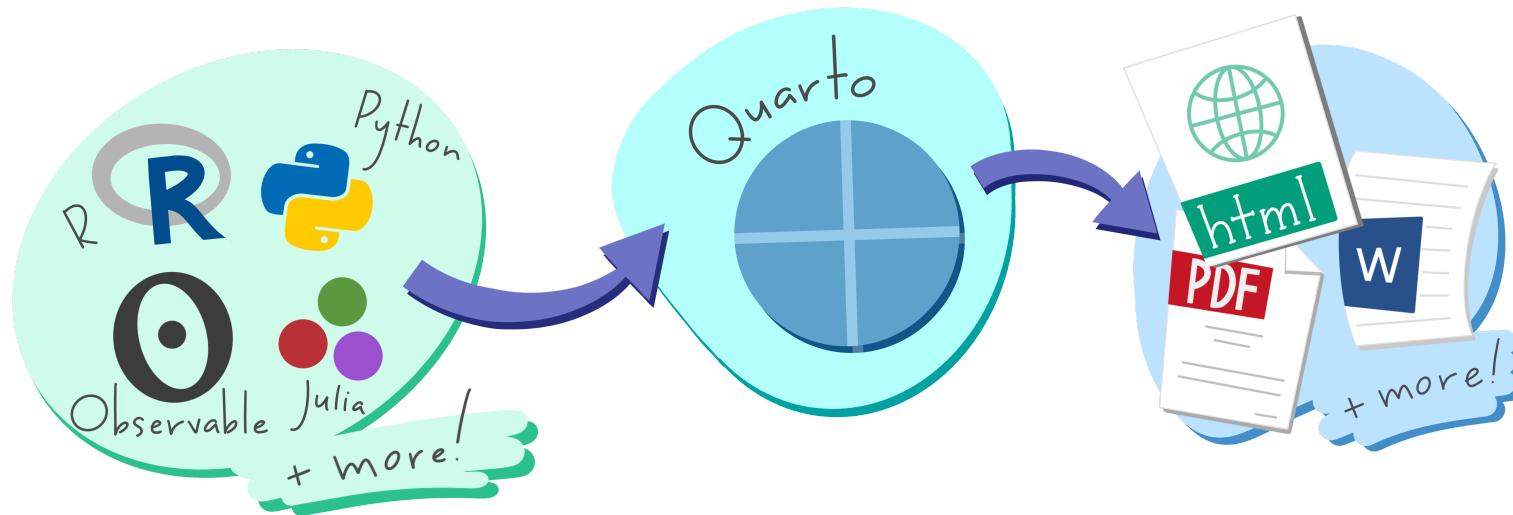
2:46 PM 2:46 PM 2:27 PM 2:46 PM 2:46 PM

29

Was ist Quarto?

Quarto ...

- ist ein neues, open-source, wissenschaftliches und technisches Publikationssystem
- zielt darauf ab, den Prozess der Erstellung von Berichten und der Zusammenarbeit drastisch zu verbessern



Artwork from “Hello, Quarto” keynote by Julia Lowndes and Mine Çetinkaya-Rundel, presented at RStudio Conference 2022. Illustrated by Allison Horst.

Ich bin dran: Ein Rundgang durch RStudio IDE und Quarto

zurücklehnen und
genießen!

Pause machen

Bitte steh auf und beweg dich. Lasst eure E-Mails in Frieden ruhen.



Ihr seid dran: md-01-uebungen

1. Öffne [posit.cloud](#) in deinem Browser (verwende dein Lesezeichen).
2. Öffne den rstatszh-k010 Arbeitsbereich (Workspace) für den Kurs.
3. Klicke auf Start neben md-01-uebungen.
4. Suche im Dateimanager im Fenster unten rechts die Datei “hallo-quarto.qmd” und klicke darauf, um sie im Fenster oben links zu öffnen.
5. Render das Dokument.
6. Füge deinen Namen zum Schlüssel `author:` im YAML-Header hinzu
7. Render das Dokument.
8. Untersuche die Komponenten des Dokuments und führe eine weitere Aktualisierung und ein erneutes Rendern durch.
9. Besprich die Notizen zu den von dir vorgenommenen Aktualisierungen. Notiere alle Aspekte des Dokuments, die nach der Tour und deiner ersten Interaktion damit unklar sind.

Von deiner eigenen Arbeitsumgebung aus

The screenshot shows a Jupyter Notebook interface with two open files: 'python.qmd' and 'Interactive-1'. The 'python.qmd' file contains a Quarto document and Python code for generating a polar plot. The 'Interactive-1' file is an IPython shell where the plot is displayed.

python.qmd Content:

```
1 ---  
2 title: "matplotlib demo"  
3 format:  
4   html:  
5     code-fold: true  
6 jupyter: python3  
7 ---  
8  
9 For a demonstration of a line plot on a  
polar axis, see @fig-polar.  
10  
11 > Run Cell  
12 ````{python}  
13 #| label: fig-polar  
14 #| fig-cap: "A line plot on a polar axis"  
15  
16 import numpy as np  
17 import matplotlib.pyplot as plt  
18  
19 r = np.arange(0, 2, 0.01)  
20 theta = 2 * np.pi * r  
21 fig, ax = plt.subplots(  
22   subplot_kw = {'projection': 'polar'}  
23 )  
24 ax.plot(theta, r)  
25 ax.set_rticks([0.5, 1, 1.5, 2])  
26 ax.grid(True)  
27 plt.show()  
```
```

**Interactive-1 Shell Output:**

```
Python 3.9.5 (v3.9.5:0a7dcdb13, May 3 2021,
13:17:02)
Type 'copyright', 'credits' or 'license' for more
information
IPython 7.25.0 -- An enhanced Interactive
Python. Type '?' for help.
```

**Polar Plot Result:**



# Quarto formate

# Mit einer Installation ist alles enthalten

- RMarkdown entwickelte sich zu einem großen Ökosystem mit unterschiedlicher Syntax.
- Quarto kommt mit allem was es braucht um verschiedene Formate zu erstellen.
  - HTML Berichte und Webseiten
  - PDF Berichte
  - MS Office (Word, Powerpoint)
  - Präsentationen (Powerpoint, Beamer, `revealjs`)
  - Bücher

# Viele Quarto Formate

| Feature         | R Markdown                                                                                     | Quarto                                                              |
|-----------------|------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| Basic Formats   | <a href="#">html_document</a><br><a href="#">pdf_document</a><br><a href="#">word_document</a> | <a href="#">html</a><br><a href="#">pdf</a><br><a href="#">docx</a> |
| Beamer          | <a href="#">beamer_presentation</a>                                                            | <a href="#">beamer</a>                                              |
| PowerPoint      | <a href="#">powerpoint_presentation</a>                                                        | <a href="#">pptx</a>                                                |
| HTML Slides     | <a href="#">xaringan</a><br><a href="#">ioslides</a><br><a href="#">revealjs</a>               | <a href="#">revealjs</a>                                            |
| Advanced Layout | <a href="#">tufte</a><br><a href="#">distill</a>                                               | <a href="#">Quarto Article Layout</a>                               |

# Viele Quarto Formate

| Feature          | R Markdown                                                                                        | Quarto                                                          |
|------------------|---------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Cross References | <a href="#">html_document2</a><br><a href="#">pdf_document2</a><br><a href="#">word_document2</a> | <a href="#">Quarto Crossrefs</a>                                |
| Websites & Blogs | <a href="#">blogdown</a><br><a href="#">distill</a>                                               | <a href="#">Quarto Websites</a><br><a href="#">Quarto Blogs</a> |
| Books            | <a href="#">bookdown</a>                                                                          | <a href="#">Quarto Books</a>                                    |
| Interactivity    | <a href="#">Shiny Documents</a>                                                                   | <a href="#">Quarto Interactive Documents</a>                    |
| Journal Articles | <a href="#">rticles</a>                                                                           | <a href="#">Journal Articles</a>                                |
| Dashboards       | <a href="#">flexdashboard</a>                                                                     | <a href="#">Quarto Dashboards</a>                               |

# Pause machen

Bitte steh auf und beweg dich. Lasst eure E-Mails in Frieden ruhen.



# Versionskontrolle

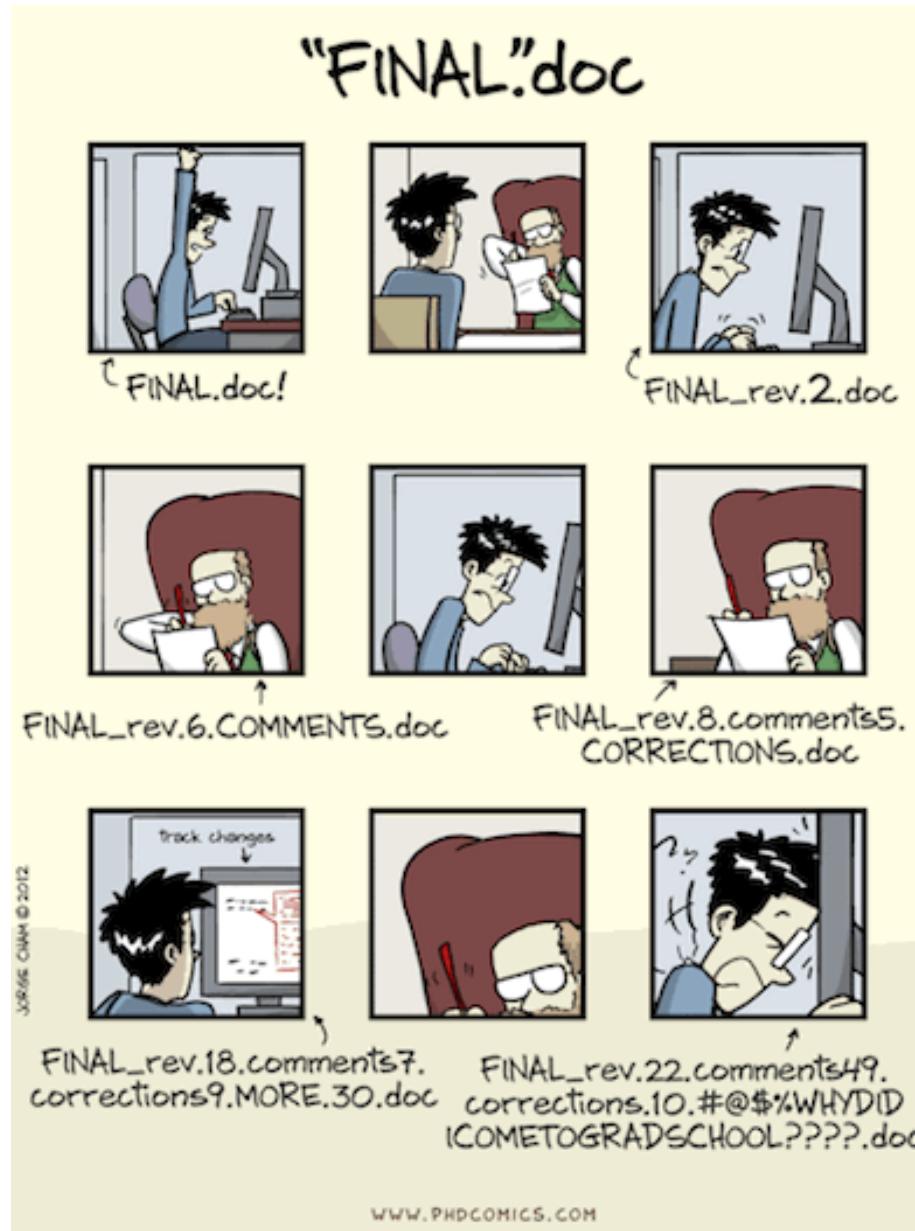
# Versionskontrolle mit Git und GitHub

Eine Möglichkeit, Dateien mit anderen zu teilen, damit sie:

- Dateien herunterladen,
- wiederverwenden,
- Etwas beitragen können.

Du kannst den Verlauf von Dateien anzeigen und zu jedem beliebigen Zeitpunkt zu einer früheren Version zurückgehen.

# Warum ist es nützlich?



# Git und GitHub



- Git ist eine Software zur Versionskontrolle
- Erstellt im Jahr 2005
- Beliebt bei Programmierern, die gemeinsam Code entwickeln
- Verfolgt Änderungen in einer Gruppe von Dateien (Verzeichnis/Ordner/Repository)
- GitHub is a hosting platform for version control using Git
- Launched in 2008, acquired by Microsoft in 2018, Microsoft for US\$ 7.5 billion
- 100 million Users (20.5 in 2022 alone) ([October, 2023](#))
- Social media for software developers

# Ich bin dran: Ein Rundgang durch GitHub

# zurücklehnen und genießen!

# Ihr seid dran: Klonen eines Repositorys von GitHub

1. Öffne einen Webbrowser auf deinem Laptop.
2. Gehe zur Website des Kurses: [rstatszh-k010.github.io/website/](https://rstatszh-k010.github.io/website/)
3. Falls noch nicht geschehen, füge die Website des Kurses zu deinen Lesezeichen hinzu.
4. Klicke im Menü auf der linken Seite auf **Modul 1**
5. In der Modulübersicht, scrolle nach unten zur Sektion **“Zusatzaufgaben”** und öffne **“Zusatzaufgabe 1: Git configuration”**
6. Folge den Anweisungen auf der Website, um das Repository von GitHub zu klonen.

# Zusatzaufgaben Modul 1

# Modul 1 Dokumentation

[rstatszh-k010.github.io/website/module/md-01.html](https://rstatszh-k010.github.io/website/module/md-01.html)

# Danke!

Danke! 🌻

Folien erstellt mit revealjs und Quarto:

<https://quarto.org/docs/presentations/revealjs/>

Access slides als [PDF auf GitHub](#)

Alle Materialien sind lizenziert unter [Creative Commons Attribution Share Alike 4.0 International.](#)