

Reproduzierbare Forschung mit R und Quarto – Praktischer Teil

Workshop für die GSE Wuppertal

Björn S. Siepe

June 27, 2025

Quarto/Markdown Grundlagen

YAML Header

- Enthält Metadaten für das Dokument
- Beispiel:

```
1 title: "Using features of dynamic networks to guide treatment selection and outcome prediction"
2 subtitle: "Simulation Code"
```

```
1 author:
2   - name: Björn S. Siepe
3     orcid: 0000-0002-9558-4648
4     affiliations: University of Marburg
5   date: "2025-06-24"
```

```
1 format:
2   html:
3     toc: true
4     number-sections: true
5     theme: cosmo
6     code-fold: true
7     code-tools: true
8     code-summary: "Show the code"
9     fig-width: 7
10    fig-height: 4.5
11    fig-align: "center"
12    embed-resources: true
```

```
1 execute:
2   message: false
3   warning: false
4   eval: false
```

Text

Basierend auf markdown:

```
1 ## Formatierung
2 *italic* **bold** ~~strikeout~~ `code`
3 superscript^2^ subscript~2~
4 [underline]{.underline} [small caps]{.smallcaps}
5
6 ## Überschriften
7 # Erste Ebene
8 ## Zweite Ebene
9 ### Dritte Ebene
10
11 ## Listen
12 - Punkt 1 einer Liste
13 - Punkt 2
14   - Punkt 2a
15   - Punkt 2b
16
17 1. Nummerierte Liste Punkt 1
18 2. Punkt 2.
19   Die Zahlen der Liste werden automatisch angepasst.
20
21 ## Quotes
22 > Gutes Zitat hier.
```

Wickham, Çetinkaya-Rundel, and Grolemund (2023)

Code Chunks - Basics

Quarto Input

```
1 ```{r}
2 mean(c(1, 2, 3, 4, 5))
3 ```
```

```
1 ```{r}
2 #| label: my-plot
3 #| echo: false
4 plot(cars)
5 ```
```

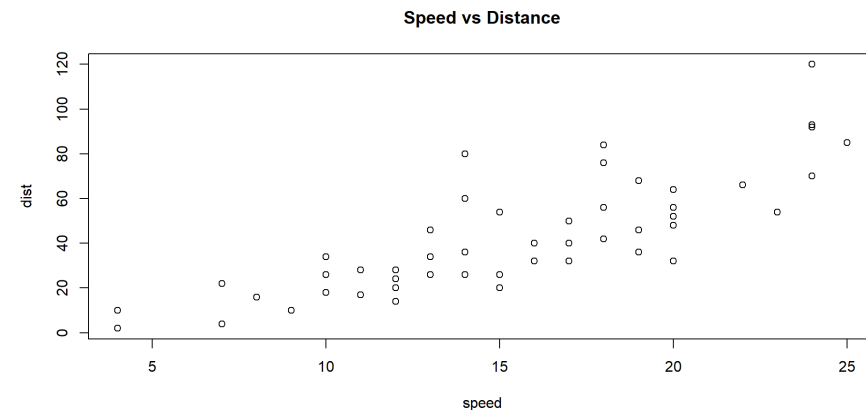
```
1 ```{python}
2 import pandas as pd
3 df = pd.DataFrame({'x': [1, 2, 3]})
4 print(df)
5 ```
```

Wickham, Çetinkaya-Rundel, and Grolemund (2023)

Rendered Output

[1] 3

Plot (with echo: false):



Python output:

```
x
0 1
1 2
2 3
```

Code Chunks: Anpassung

- Code Chunks können mit verschiedenen Optionen angepasst werden
 - Namen
 - Ausführung
 - Abbildungsoptionen

```
1 ```{r}
2 #| label: Beispielchunk
3 #| echo: true
4 #| fig-width: 6
5 ```
```

Code Chunks: Optionen

Verschiedene Optionen können in den Code Chunks verwendet werden:

Option	Run code	Show code	Output	Plots	Messages	Warnings
<code>eval: false</code>	X		X	X	X	X
<code>include: false</code>		X	X	X	X	X
<code>echo: false</code>		X				
<code>results: hide</code>			X			
<code>fig-show: hide</code>				X		
<code>message: false</code>					X	
<code>warning: false</code>						X

Wickham, Çetinkaya-Rundel, and Grolemund ([2023](#))

Inline Code

- Inline Code wird mit `r` oder `python` in Backticks geschrieben:

```
1 Der Mittelwert der ganzen Zahlen von 1 bis 5 ist 3.
```

Ergibt:

Der Mittelwert der ganzen Zahlen von 1 bis 5 ist 3.

Visual Editor

- Quarto bietet einen Visual Editor, der die Bearbeitung von Markdown-Dateien erleichtert
- Er kann über das Menü “View” → “Visual Editor” aktiviert werden
- Er ermöglicht eine WYSIWYG-Erfahrung

Übung 1

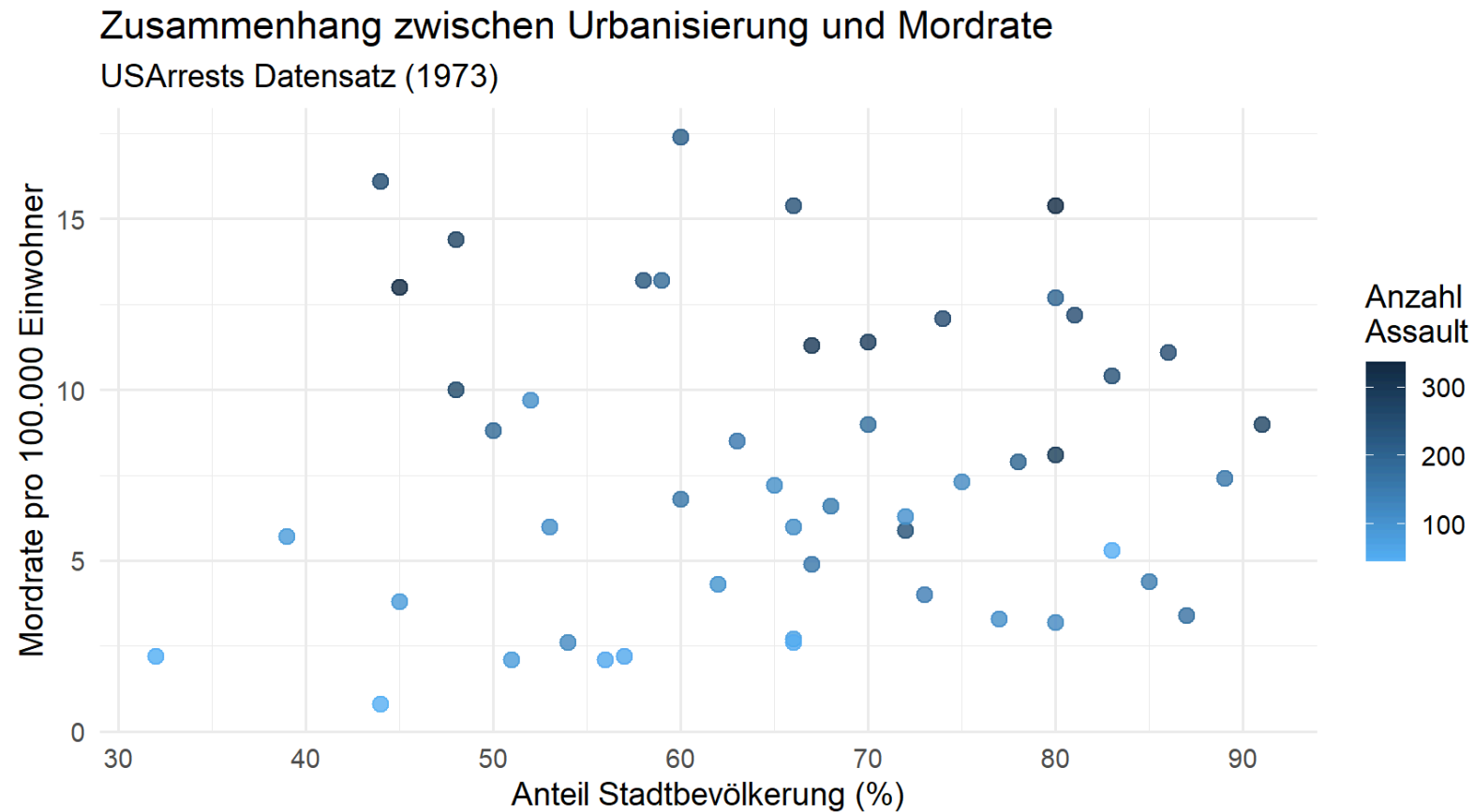
Zeit zum Üben!

- Laden Sie entweder `aufgabe_01.qmd` herunter oder erstellen Sie eine neue Quarto-Datei.
- Personalisieren Sie den YAML-Header.
- Laden Sie den “USArrests” Datensatz in R. Dies geht mit `data("USArrests")`.
- Schreiben Sie einen kurzen Einleitungstext für eine Beispielanalyse. Verwenden Sie unterschiedliche Formatierungen.
- Berichten Sie über das Minimum und Maximum einer der Variablen in einem Satz.

Abbildungen

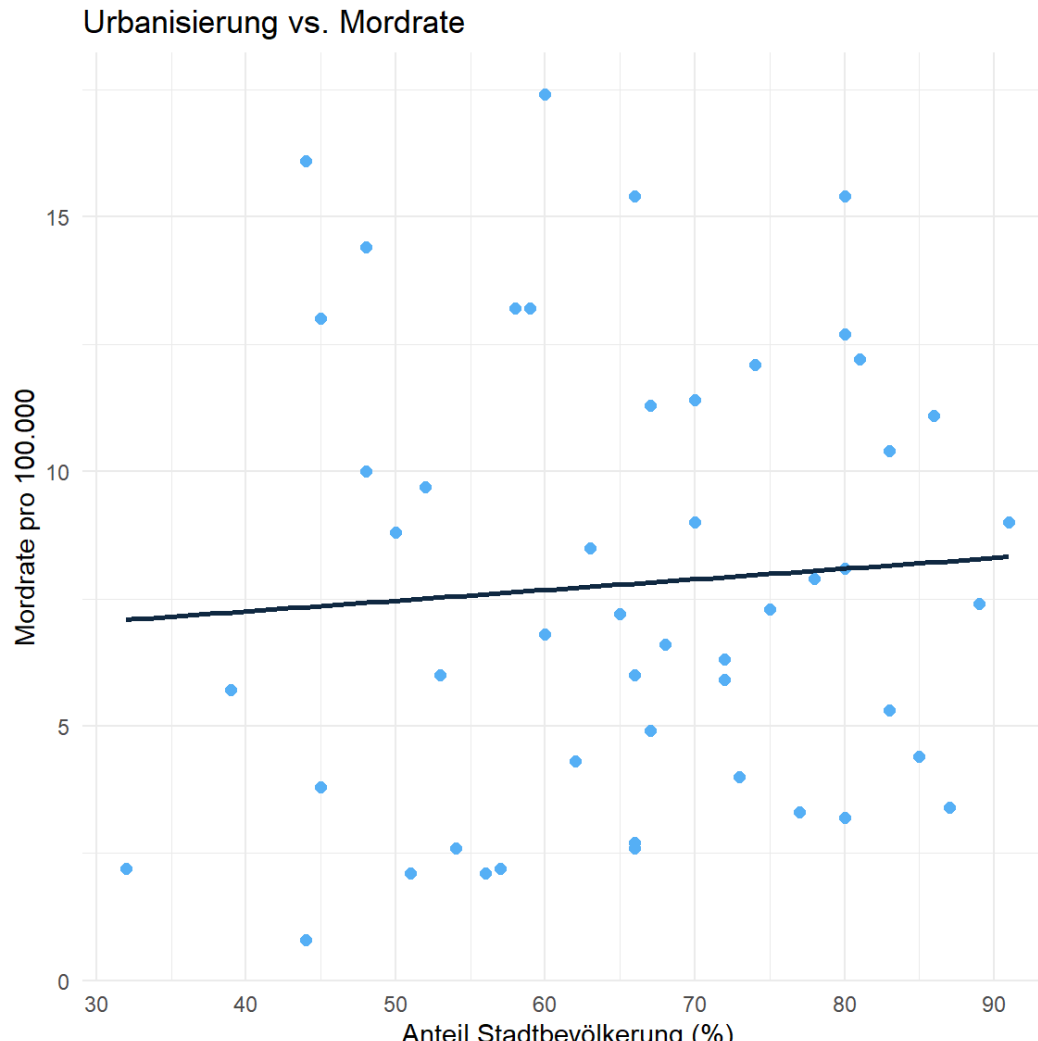
- Beispiel:

► Code



Mehrere Abbildungen

- Wir können auch mehrere Abbildungen zusammen einfügen:



Externe Grafiken

- Externe Bilder können mit Markdown-Syntax oder `knitr::include_graphics()` eingefügt werden:

Markdown-Syntax:

```
1 ![Bildbeschreibung](pfad/zum/bild.png){width=50%}
```

R Code Chunk:

```
1 knitr::include_graphics("images/my_plot.png")
```

Online-Bilder:

```
1 ![R Logo](https://www.r-project.org/logo/Rlogo.png){width=200px}
```

Tabellen

- Tabellen können mit vielen Paketen erstellt werden, z.B. `kable`, `gt`, `flextable` oder `pander`.
- Beispiel mit `kable`:

```
1 data("USArrests")
2 library(knitr)
3 kable(head(USArrests), caption = "Erste 6 Zeilen des USArrests Datensatzes")
```

Erste 6 Zeilen des USArrests Datensatzes

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7

Tabellen: Formatierung

Tabellen-Export für Manuskripte

- Tabellen können in verschiedenen Formaten exportiert werden, z.B. LaTeX, HTML oder Word.
- Beispiel mit `flextable` für eine APA-konforme Tabelle:

```
1 library(flextable)
2 tabelle <- flextable(head(USArrests))
3
4 tabelle |>
5   save_as_docx("USArrests_table.docx", path = here::here("tables"))
```


Übung 2

- Verwenden Sie entweder Ihr existierendes Quarto-Dokument oder laden Sie [aufgabe_02.qmd](#) herunter.
- Fügen Sie eine Abbildung hinzu, die den Zusammenhang zwischen “UrbanPop” und “Assault” im USArrests Datensatz zeigt.
 - Ändern Sie die Größe der Abbildung.
 - Optional: Speichern Sie die Abbildung und laden Sie als externes Bild ein.
- Fügen Sie eine Tabelle hinzu.
- Versuchen Sie, die Tabelle schöner zu formatieren.

Referenzen

- Referenzen können mit BibTeX-Dateien eingebunden werden.
- Als Beispiel verwenden wir `referenzen.bib`. Dies müssen wir im YAML-Header angeben:

```
1 bibliography: referenzen.bib
2 csl: apa.csl # Optional: Style
```

- Ein Zitat kann dann so aussehen:
 - `[@wickham2016]` → (Wickham, 2016)
 - `@wickham2016` → Wickham (2016)
 - `[@wickham2016; @r2023]` → (Wickham, 2016; R Core Team, 2023)

BibTeX Beispiel

Eine typische `referenzen.bib` Datei sieht so aus:

```
1 @book{wickham2016,  
2   title = {R for Data Science},  
3   author = {Hadley Wickham and Garrett Golemund},  
4   year = {2016},  
5   publisher = {O'Reilly Media},  
6   isbn = {978-1491910399}  
7 }  
8  
9 @manual{r2023,  
10  title = {R: A Language and Environment for Statistical Computing},  
11  author = {{R Core Team}},  
12  organization = {R Foundation for Statistical Computing},  
13  address = {Vienna, Austria},  
14  year = {2023},  
15  url = {https://www.R-project.org/}  
16 }
```

Verschiedene Zitierweisen

Syntax	Ausgabe	Verwendung
<code>[@autor2023]</code>	(Autor, 2023)	Normale Klammerzitate
<code>@autor2023</code>	Autor (2023)	Autor als Teil des Satzes
<code>[@autor2023, S. 15]</code>	(Autor, 2023, S. 15)	Mit Seitenzahl
<code>[-@autor2023]</code>	(2023)	Nur Jahr, ohne Autor
<code>[@autor2023; @autor2024]</code>	(Autor, 2023, 2024)	Mehrere Quellen

Referenzen hinzufügen

Manuell in der BibTeX-Datei: - Direkt in [referenzen.bib](#) editieren - Google Scholar: “Zitieren” → “BibTeX” kopieren (und DOI hinzufügen)

Zotero Integration: - Better BibTeX Plugin installieren - Auto-Export zu [.bib](#) Datei - Sync zwischen Zotero und Quarto

RStudio Visual Editor: - [Insert](#) → [Citation](#) - Automatisches Hinzufügen zur [.bib](#) Datei

Literaturverzeichnis

- Wird automatisch am Ende des Dokuments eingefügt
- Nur zitierte Quellen werden angezeigt
- Anpassung der Position:

```
1 # Literatur
2 ::: {#refs}
3 :::
4
5 # Anhang
6 Weiterer Inhalt nach dem Literaturverzeichnis...
```

Rechenintensive Chunks

- Teilweise möchten wir Code nicht ausführen, wenn das Dokument gerendert wird
 - z.B. bei rechenintensiven Operationen
 - Ausführung des Codes auf einem Server
 - oder bei Zufallsimulationen
 - oder wenn wir nur den Code zeigen wollen
- Beispiel:

```
1 data("USArrests")
2
3 # Bootstrap durchführen, um Konfidenzintervall einer Regression zu schätzen
4 n_boot <- 1000
5 bootstrap_results <- numeric(n_boot)
6
7 # Warnung: Dieser Code dauert ca. 30 Sekunden!
8 system.time({
9   for(i in 1:n_boot) {
10     # Zufälliges Sampling mit Zurücklegen
11     boot_sample <- sample(nrow(USArrests), replace = TRUE)
12     boot_data <- USArrests[boot_sample, ]
13
14
15     boot_model <- lm(Murder ~ UrbanPop, data = boot_data)
16     # Extrahieren der Slope
17     bootstrap_results[i] <- coef(boot_model)[2]
18   }
```

Übung 3

- Arbeiten sie entweder in Ihrer existierenden Quarto- Datei weiter oder verwenden Sie [aufgabe_03.qmd](#).
- Fügen Sie eine Referenz zu einem Buch oder Artikel hinzu, den Sie interessant finden.
- Lassen Sie das Literaturverzeichnis am Ende des Dokuments erscheinen.
- Führen Sie eine Regression auf den USArrests Datensatz durch und speichern Sie das Ergebnis. Sorgen Sie dafür, dass der Code nicht ausgeführt wird, wenn das Dokument gerendert wird.
- Optional: Versuchen Sie, das Dokument so zu formatieren, dass es möglichst nah an einen echten Artikel herankommt.

Kleine Zusatztips

- `.qmd` zu `.r` umwandeln mit `knitr::`

Session Info

R version 4.5.0 (2025-04-11 ucrt)

Platform: x86_64-w64-mingw32/x64

locale: *LC_COLLATE=German_Germany.utf8, LC_CTYPE=German_Germany.utf8, LC_MONETARY=German_Germany.utf8, LC_NUMERIC=C and LC_TIME=German_Germany.utf8*

attached base packages: *stats, graphics, grDevices, utils, datasets, methods and base*

other attached packages: *knitr(v.1.50) and ggplot2(v.3.5.2.9001)*

loaded via a namespace (and not attached): *Matrix(v.1.7-3), gtable(v.0.3.6), jsonlite(v.2.0.0), dplyr(v.1.1.4), compiler(v.4.5.0), tidyselect(v.1.2.1), Rcpp(v.1.0.14), dichromat(v.2.0-0.1), splines(v.4.5.0), scales(v.1.4.0), yaml(v.2.3.10), fastmap(v.1.2.0), lattice(v.0.22-7), R6(v.2.6.1), labeling(v.0.4.3), generics(v.0.1.3), tibble(v.3.2.1), pander(v.0.6.6), pillar(v.1.10.2), RColorBrewer(v.1.1-3), rlang(v.1.1.6), xfun(v.0.52), S7(v.0.2.0), cli(v.3.6.5), withr(v.3.0.2), magrittr(v.2.0.3), mgcv(v.1.9-3), digest(v.0.6.37), grid(v.4.5.0), lifecycle(v.1.0.4), nlme(v.3.1-168), vctrs(v.0.6.5), evaluate(v.1.0.3), glue(v.1.8.0), farver(v.2.1.2), rmarkdown(v.2.29), tools(v.4.5.0), pkgconfig(v.2.0.3) and htmltools(v.0.5.8.1)*

Literatur

Wickham, H., M. Çetinkaya-Rundel, and G. Grolemund. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.