

Analyse und Dokumentation

BSc Psychologie SoSe 2025

Belinda Fleischmann und Dirk Ostwald

	Gruppe $1/2$	Gruppe 3	Format	Thema
1	Do, 10.04	Fr, 11.04	Seminar	(1) Quarto, Zotero, Tidyverse
2	Do, 17.04	Fr, 25.04	Seminar	(2) Ethik und Ethische Formalitäten
3	Do, 24.04	Fr, 02.05	Seminar	(3) Wissenschaftliche Berichte
4	Mi, 30.04	Fr, 09.05	Seminar	(4) Offenheit und Transparenz
5	Do, 08.05	Fr, 16.05	Praxisseminar	Offene Übung
6	Do, 15.05	Fr, 23.05	Präsentationen	Einfache Lineare Regression
7	Mi, 21.05	Fr, 30.05	Präsentationen	Korrelation
8	Do, 05.06	Fr, 06.06	Präsentationen	Einstichproben-T-Test
9	Do, 12.06	Fr, 13.06	Präsentationen	Zweistichproben-T-Test
10	Do, 19.06	Fr, 20.06	Präsentationen	Einfaktorielle Varianzanalyse
11	Do, 26.06	Fr, 27.06	Präsentationen	Zweifaktorielle Varianzanalyse
12	Do, 03.07	Fr, 04.07	Präsentationen	Multipe Regression
13	Do, 10.07	Fr, 11.07	Präsentationen	Kovarianzanalyse
	Juli		Klausurtermin	

(1) Quarto, Zotero, Tidyverse

Zotero

Tidyverse

Zotero

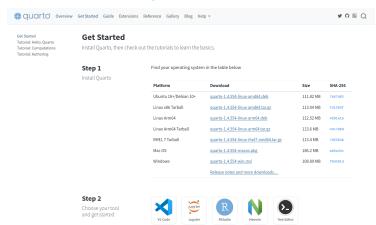
Tidyverse



Was ist Quarto?

- Ein seit 2022 verfügbares freies wissenschaftlich-technisches Publikationssystem
- Eine Weiterentwicklung von RMarkdown und RBookdown durch Posit
- RMarkdown/RBookdown sind RStudio Adaptationen von Markdown und Jupyter Notebooks
- Allgemeines Ziel ist hier die einfache Integration von ausführbarem Programmiercode in ein ansprechendes Text-. Tabellen- und Abbildungslavout für Web- und Printdokumente.
- Quarto nutzt Markdown und Latex für Layoutprozesse.
- Quarto nutzt Pandoc für multiple Outputformate (.html, .docx, .pdf, etc.)
- Quarto läuft smoother und schneller als RMarkdown und RBookdown.

Quarto Installation



Quarto VSCode Tutorial



Was ist Markdown?

- Eine Markup Language (Auszeichnungssprache) zur Erzeugung formatierten Texts
- Eine HTML Alternative zur Erstellung von Webseiten etc. mithilfe einfacher Texteditoren
- Von John Gruber und Aaron Swartz 2004 mit dem Ziel hoher Lesbarkeit entwickelt

Text using Markdown syntax	Corresponding HTML produced by a Markdown processor	Text viewed in a browser
Heading # Alternative heading # Alternative sub-heading Paragraphs are separated by a blank line. Two spaces at the end of a line produce a line break.	chi>Heading <pre>chi>Sub-heading chi>Alternative heading chi>Alternative beading chi>Alternative sub-heading chi>Alternative sub-heading chi>Day alternative sub-heading</pre> chi>Day alternative sub-heading chi>Day alternative sub	Heading Sub-heading Alternative heading Alternative sub-heading Panagraphs are separated by a blank line. Two spaces at the end of a line produce a line break.
Text attributes _{tolic_, **bold**, monospace'. Horizontal rule:	<pre>rext attributes <embitalic< em="">, bold, <code>monospace</code>.</embitalic<></pre>	Text attributes ifailo, bold , monospace . Horizontal rule:

Was ist Latex?

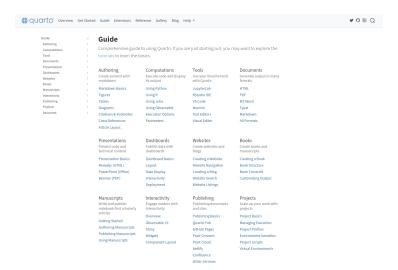
- Ein Softwarepaket zur Vereinfachung von TeX
- TeX ist ein von Donald Knuth ab 1977 entwickeltes Textsatzsystem mit Makrosprache
- LaTeX wurde von Leslie Lamport Anfang 1984 entwickelt
- LaTeX ist insbesondere für mathematische Berichte und Präsentationen (Beamer) nützlich

```
\footnotesize\
\lambda_intercera[Datenverteilung des Allgemeinen Linearen Modells]
\fustifying
\lambda_iving
\lamb
```



Theorem (Datenverteilung des Allgemeinen Linearen Modells) Es sei $v = X\beta + \varepsilon \ \text{mit} \ \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{7}$ das ALM. Dann gilt $v \sim N(\mu, \sigma^2 I_n) \ \text{mit} \ \mu := X\beta \in \mathbb{R}^n. \tag{8}$

Quarto Guide



\$\mu := X\beta \in \mathbb{R}^n\$ aufschreiben.

```
title: "Quarto Demonstration"
author: "Toni Demo"
date: today
format: pdf
# Überschrift zu Kapitel 1.
Hier steht der Text für Kapitel 1. Darin könnte auch eine Abbildung enthalten sein.
![](../Abbildungen/otto.png){width="10%"}
## Überschrift zum Unterkapitel 1.1
Hier steht der Text für Unterkapitel 1.1. Manche Worte möchte ich **fett** und manche Worte _kursiv_, und Befehle
in `monospace` schreiben. Mögliche Farben möchte ich mit Stichpunkten auflisten.
* \textcolor{blue}{blau}
* \textcolor{green}{grün}
* \textcolor{red}{rot}
* \textcolor{gray}{grau}
Wenn wir mathematische Ausdrücke mit Dollarzeichen umrahmen, werden sie mithilfe von \LaTeX formatiert.
So können wir z.B. die Verteilung eines Zufallsvektors formal mit $\upsilon \sim N(\mu, \sigma^2 I_n)$ mit
```

Quarto Demonstration

Belinda Fleischmann 2024-05-02

Überschrift zu Kapitel 1.

Hier steht der Text für Kapitel 1. Darin könnte auch eine Abbildung enthalten sein.



Überschrift zum Unterkapitel 1.1

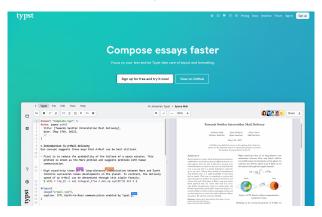
Hier steht der Text für Unterkapitel 1.1. Manche Worte möchte ich fett und manche Worte kursiv. und Befehle in monospace schreiben. Mögliche Farben möchte ich mit Stichpunkten auflisten.

- blau
- grün
- rot • grau

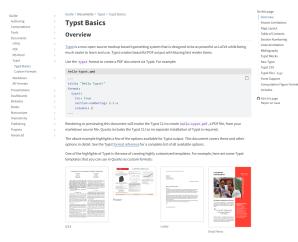
Wenn wir mathematische Ausdrücke mit Dollarzeichen umrahmen, werden sie mithilfe von LaTeX formatiert. So können wir z.B. die Verteilung eines Zufallsvektors formal mit $v \sim N(\mu, \sigma^2 I_n)$ mit $\mu := X\beta \in \mathbb{R}^n$ aufschreiben. Beispielbericht

Beispielpräsentation

Typst



Quarto Typst Integration



Learn more about how to use them, and how to create your own in Custom Formats.

Zotero

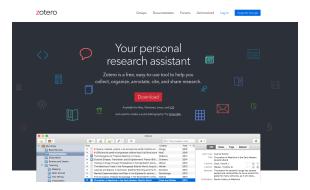
Tidyverse

Was ist ein Reference Manager?

- Reference Manager sind Literaturverwaltungsprogramme
- Reference Manager unterstützen Zitationen und das Erstellen von Literaturverzeichnissen
- Zitierstile können automatisch auf bestimmte Spezifikationen (z.B. APA) eingestellt werden
- Reference Manager dienen auch als digitale Bibliotheken
- Kommerzielle Reference Manager sind z.B. EndNote, Citavi, Mendeley und Papers
- Kostenlose/Freemium Reference Manager sind z.B. JabRef und Zotero
- Eine Integration in Quarto erlaubt z.B. der Export der eigenen Library in das BibTex Format.

Zotero Website

Zotero Documentation



Quarto

Zotero

Tidyverse

Tidyverse



Cheat Sheets

💥 posit

Data transformation with dplyr:: **CHEATSHEET** dplyr functions work with pipes and expect tidy data. In tidy data EXTRACT CASES EXTRACT VARIABLES Row functions return a subset of rows as a new table. Column functions return a set of columns as a new vector or table. Each variable is in Each observation, or case, is in its own row becomes f(x, y) its own column filter(.data, ..., .preserve = FALSE) Extract rows pull(.data, var = -1, name = NULL, ...) Extract that meet logical criteria. column values as a vector, by name or index. distinct(.data, ..., keep_all = FALSE) Remove select(.data, ...) Extract columns as a table. Apply summary functions to columns to create a new table of rows with duplicate values. summary statistics. Summary functions take vectors as input and return one value (see hark) relocate(.data, ..., .before = NULL, .after = NULL) slice(.data, preserve = FALSE) Select rows Move columns to new position. summary function mtcars |= relocate(mpg, cyl, after = last_col()) summarize(.data....) Compute table of summaries. slice_sample(.data, ..., n, prop, weight_by = NULL_replace = FALSE) Randomly select rows. Use these helpers with select() and across() Use n to select a number of rows and prop to count(.data, ..., wt = NULL, sort = FALSE, name = select a fraction of rows. contains(match) num_range(prefix, range) i, e.g., mpg:cyl NULL) Count number of rows in each group defined → by the variables in ... Also tally(), add_count(), ends_with(match) all_of(x)/any_of(x, ..., vars) i.e.g., is starts_with(match) matches(match) everyth slice_min(.data, order_by, ..., n, prop, with ties = TRUE) and slice max() Select rows MANIPULATE MULTIPLE VARIABLES AT ONCE with the lowest and highest values. slice_head(.data, ..., n, prop) and slice_tail() Use group_by(.data, ..., .add = FALSE, .drop = TRUE) to create a "grouped" copy of a table grouped by columns in ... dolyr across(.cols, .funs, ..., .names = NULL) Summarize or mutate multiple columns in the same way. Select the first or last rows. functions will manipulate each "group" separately and combine Logical and boolean operators to use with filter() c_across(.cols) Compute across columns in is.na() %in% lis.na() ! See ?base::Logic and ?Comparison for help MAKE NEW VARIABLES ARRANGE CASES Apply vectorized functions to columns. Vectorized functions take Use rowwise(.data, ...) to group data into individual rows. dplyr arrange(.data, ..., .by_group = FALSE) Order vectors as input and return vectors of the same length as output functions will compute results for each row. Also apply functions rows by values of a column or columns flow to (see back). to list-columns. See tidyr cheat sheet for list-column workflow. vectorized function high), use with desc() to order from high to low. mutate(.data, keep = "all", .before = NULL, .after = NULL) Compute new column(s). Also add_column(). ADD CASES add_row(,data, ..., before = NULL, after = NULL) ungroup(x,...) Returns ungrouped copy of table. Add one or more rows to a table. rename(,data, ...) Rename columns. Use

dplyr Cheat Sheet

CC BY SA Posit Software, PBC - info@posit.co - posit.co - Learn more at dplyr.tidyverse.org - HTML cheatsheets at posit;cheatsheets - dplyr 1.1.2 - Updated: 2023-07

rename_with() to rename with a function.

Tidyverse dplyr

D <- read.table("./Daten/Daten_1.csv", sep = ",", header = TRUE)

Daten einlesen

Variable_1	Variable_2	Variable_3
34.87	34.61	33.56
32.16	22.89	15.75
33.95	31.82	28.83
28.78	25.91	20.04
30.13	26.83	22.00
30.50	26.50	24.42
32.48	26.92	22.96
31.66	31.84	28.83
32.76	33.00	33.28
31.60	26.77	21.21
32.44	28.55	28.63
29.48	25.33	24.19
31.24	28.97	25.18
34.33	31.31	28.22
31.56	27.11	22.92
31.87	30.95	30.30
27.07	21.94	17.60
29.36	25.41	19.32
36.07	33.56	33.41
33.03	28.81	26.58
33.12	32.20	29.44

Tidyverse dplyr

Der Pipe operater %>% oder |> ermöglicht es, Funktionen in einer Reihe nacheinander auszuführen.

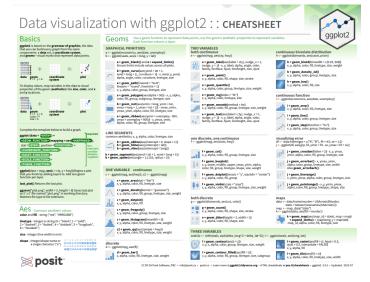
mutate() erlaubt das Erzeugen neuer Spalten als Funktionen bestehender Spalten

Variable_1	Variable_2	Variable_3	ID	Summe
34.87	34.61	33.56	1	103.04
32.16	22.89	15.75	2	70.79
33.95	31.82	28.83	3	94.60
28.78	25.91	20.04	4	74.74
30.13	26.83	22.00	5	78.96
30.50	26.50	24.42	6	81.42
32.48	26.92	22.96	7	82.37
31.66	31.84	28.83	8	92.34
32.76	33.00	33.28	9	99.05
31.60	26.77	21.21	10	79.58
32.44	28.55	28.63	11	89.62
29.48	25.33	24.19	12	79.00
31.24	28.97	25.18	13	85.40
34.33	31.31	28.22	14	93.86
31.56	27.11	22.92	15	81.59
31.87	30.95	30.30	16	93.12
27.07	21.94	17.60	17	66.61
29.36	25.41	19.32	18	74.09
36.07	33.56	33.41	19	103.04
33.03	28.81	26.58	20	88.42
33.12	32.20	29.44	21	94.75

Tidyverse dplyr

filter() erlaubt es, Zeilen gemäß bestimmten Bedingungen auswählen

Variable_1	Variable_2	Variable_3	ID	Summe
34.87	34.61	33.56	1	103.04
33.95	31.82	28.83	3	94.60
31.66	31.84	28.83	8	92.34
32.76	33.00	33.28	9	99.05



ggplot2 Cheat Sheet

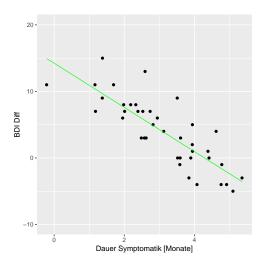
Tidyverse ggplot2

Beispieldatensatz

Die ersten 12 Zeilen des Dataframes:

DUR	BDI	PatientIn
1.37	9	1
2.18	8	2
1.16	11	3
3.60	0	4
2.33	8	5
1.18	7	6
2.49	3	7
2.74	7	8
2.58	3	9
1.69	11	10
3.51	9	11
2.39	7	12

```
library(ggplot2)
                                                         # Für ggplot()
# Visualisierung
ggplot(
 data = D_processed,
                                                         # Daten
  mapping = aes(x = DUR, y = BDI)
                                                         # Daten-Axen-mapping
  ) +
  coord_cartesian(ylim = c(-10, 20)) +
                                                         # y-limits anpassen
  geom_point() +
                                                         # Datenpunkte zeichnen
  geom_smooth(
                                                         # Ausgleichsgerade zeichnen
    method = "lm".
    color = "green", se = F, linewidth = 0.4
   ) +
 ylab("BDI Diff") + xlab("Dauer Symptomatik [Monate]") # Achsenbeschriftung
graphics.off()
                                                         # Schließt browser
ggsave(
                                                         # Abbildung speichern
 filename = "ggplot_beispiel.pdf",
 height = 5, width = 5
```



Quellen

VS Code Website

VS Code-R Wiki

R for Data Science (2e)

ggplot2: Elegant Graphics for Data Analysis (3e)