

## Analyse und Dokumentation

BSc Psychologie SoSe 2024

#### Belinda Fleischmann

Datum	Einheit	Thema	Lehrperson
10.04.24	Seminar	(1) Ethik und Ethische Formalitäten	BF
17.04.24	Seminar	(2) Wissenschaftliche Berichte	BF
24.04.24	Seminar	(3) Offenheit und Transparenz	BF
01.05.24	Tag der Arbeit		
08.05.24	Seminar	(4) R, Quarto und Zotero	BF
15.05.24	Praxisseminar	Offene Übung	BF
22.05.24	Präsentationen	Einfache Lineare Regression	JS
29.05.24	Präsentationen	Korrelation	JS
05.06.24	Präsentationen	Einstichproben-T-Test	JS
12.06.24	Präsentationen	Zweistichproben-T-Test	JS
19.06.24	Präsentationen	Einfaktorielle Varianzanalyse	BF
26.06.24	Präsentationen	Zweifaktorielle Varianznalyse	BF
03.07.24	Präsentationen	Multipe Regression	BF
10.07.24	Präsentationen	Kovarianzanalyse	BF
26.07.24	Klausurtermin		
Feb 2025	Klausurwiederholungstermin		

(4) R, Quarto und Zotero

R Tools

Quarto

Zotero

## **R** Tools

Quarto

Zotero

#### Hilfreiche Quellen

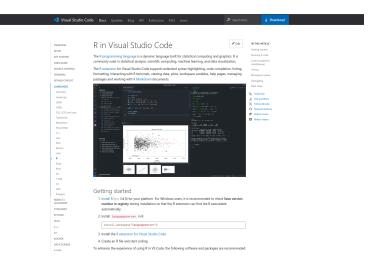
Visual Studio Code (VS Code) Website

VS Code-R Wiki

R for Data Science (2e)

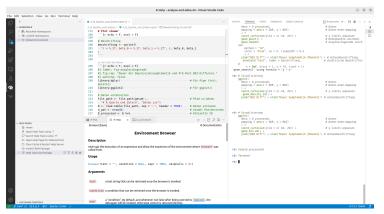
ggplot2: Elegant Graphics for Data Analysis (3e)

## Wiederholung: R und VS Code



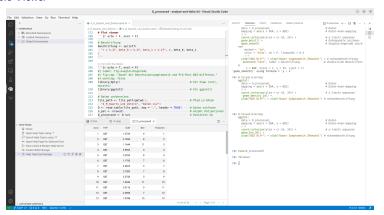
VS Code Website

#### Help Viewer



Mit dem Befehl ?funktionsname oder über HELP PAGES öffner

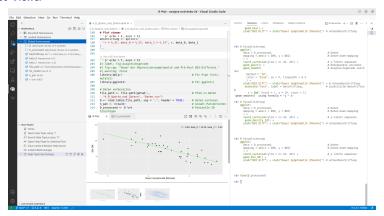
#### Table Viewer



#### List Viewer



#### Plot Viewer



Mit dem R Paket httpgd erleichtert die Ansicht erstellter Grafiken.

Nach Installation über r.plot.useHttpgd in VS Code Einstellungen freischalten.

### Debugging mit browser()

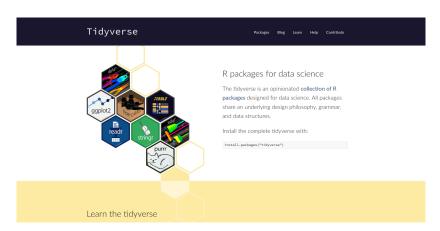
Die R base Funktion browser() erlaubt das Pausieren der Exekution eines Skripts und Inspektion der aktuellen *environment*.

#### Beispiel

Über das Argument expr kann auch eine Bedingung als boolesche Operation spezifiziert werden.

Mit Enter wird die Exekution fortgeführt.

Mit Q wird der browser beendet.



Tidyverse

Cheat Sheets

#### 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....) slice\_sample(.data, ..., n, prop, weight\_by = NULL\_replace = FALSE) Randomly select rows. Compute table of summaries 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) everythi 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 rename\_with() to rename with a function. 💥 posit 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

dplyr Cheat Sheet

D <- read.table("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

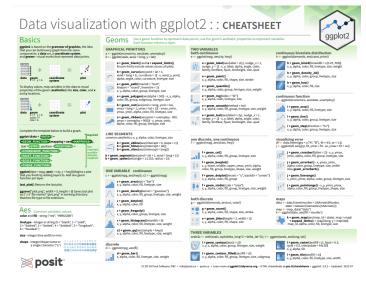
Der Pipe operater %>% oder |> ermöglicht es, Funktionen in einer Reihe nacheinander auszuführen.

Mit der R-Funktion mutate() können wir neue Spalten erzeugen (auch 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

 ${\sf Mit\; der\; R-Funktion\; filter()\; k\"onnen\; wir\; Zeilen\; gem\"{a}{\it B}\; bestimmten\; Bedingungen\; ausw\"{a}hlen}.$ 

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

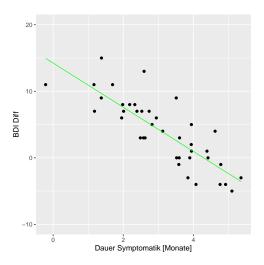
### Plotten mit ggplot

#### 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
```



 $\mathsf{R}\ \mathsf{Tools}$ 

Quarto

Zotero

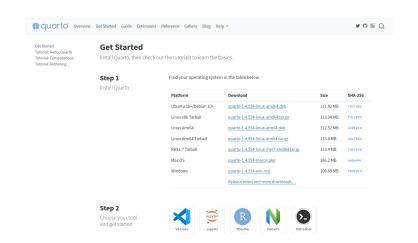


Quarto Website

#### Was ist Quarto?

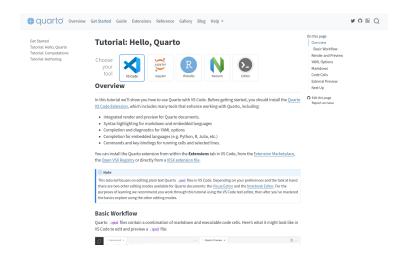
- Ein seit 2022 verfügbares freies wissenschaftlich-technisches Publikationssystem
- Eine Weiterentwicklung von RMarkdown und RBookdown durch Posit
- RMarkdown/RBookdown sind RStudio Adaptationen von Markdown und Jupyter Notebooks
- Allgemeines Ziel ist hier die einfache Integration von ausführbarem 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.

#### Installation von Quarto



Quarto Website Installation

## Quarto und VS Code



Quarto Website Tutorial: VS Code

#### 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 chi>Sub-heading chi>Alternative heading chi>Alternative sub-heading	Heading Sub-heading Alternative heading Alternative sub-heading Panagagains are separated by a blank line. Two spaces at the end of a line produce a line break.
Text attributes _italic_, **bold**, amonspace .  Horizontal rule:	<pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre>	Text attributes ifalio, <b>bold</b> , monospace . Horizontal rule:

- Ein Softwarepaket zur Vereinfachung von TeX
- TeX ist ein von Donald E. 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_interporemiploatenverteilung des Allgemeinen Linearen Modells]
\Justifying
\Justifying
\Justifying
\Justifying
\Lambda_interporemiploaten

Es sei
\Lambda_geaution}
\Lambda_geaution
\Lambd
```



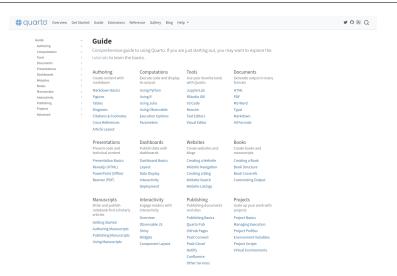
# Theorem (Datenverteilung des Allgemeinen Linearen Modells) Es sei $\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{7}$ das ALM. Dann gilt $\upsilon \sim N(\mu, \sigma^2 I_n) \text{ mit } \mu := X\beta \in \mathbb{R}^n. \tag{8}$

#### Pandoc

- Pandoc ist eine kostenlose Document Conversion Software von John MacFarlane
- · Pandoc ist in die Quarto Engine integriert
- Einige von Pandocs Konvertierungsoptionen sind im Folgenden aufgeführt



## Quarto Guide



Quarto Website Guide

#### Quarto Beispiel

```
title: "Quarto Demonstration"
author: "Belinda Fleischmann"
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
$\mu := X\beta \in \mathbb{R}^n$ aufschreiben.
```

#### **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
- 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.

## Berichte und Präsentationen mit Quarto

Beispielbericht

Beispielpräsentation

 $\mathsf{R}\;\mathsf{Tools}$ 

Quarto

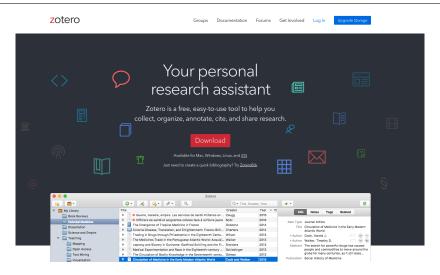
Zotero

## Reference Manager

#### 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.

#### 7otero



Zotero Website

Zotero Documentation