



5. Auflage

Luhmann

# R für Einsteiger

Einführung in die Statistiksoftware  
für die Sozialwissenschaften



Online-Material

**BELTZ**

Luhmann

**R für Einsteiger**

Maike Luhmann

# R für Einsteiger

Einführung in die Statistik-Software für die Sozialwissenschaften

Mit Online-Material

5., überarbeitete Auflage

**BELTZ**

Prof. Dr. Maike Luhmann  
Ruhr-Universität Bochum  
Fakultät für Psychologie  
Psychologische Methodenlehre  
Universitätsstr. 150  
Raum IB E4/185 - Postfach 17  
44780 Bochum  
E-Mail: maike.luhmann@rub.de

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlags unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronische Systeme.



Dieses Buch ist erhältlich als:  
ISBN 978-3-621-28790-6 Print  
ISBN 978-3-621-28791-3 E-Book (PDF)

5., überarbeitete Auflage 2020

© 2020 Programm PVU Psychologie Verlags Union  
in der Verlagsgruppe Beltz · Weinheim Basel  
Werderstraße 10, 69469 Weinheim  
Alle Rechte vorbehalten

Lektorat: Salome Fabianek  
Coverbild: IconicBestiary/iStock/Getty Images Plus; R-Logo © 2016 The R Foundation, unter CC-BY-SA 4.0 modifiziert  
Herstellung: Lelia Rehm  
Satz: Reproduktionsfähige Vorlagen der Autorin  
Gesamtherstellung: Beltz Grafische Betriebe, Bad Langensalza  
Printed in Germany

Weitere Informationen zu unseren Autor\_innen und Titeln finden Sie unter: [www.beltz.de](http://www.beltz.de)

# Inhaltsübersicht

<b>Vorwort zur fünften Auflage</b>	<b>12</b>
<b>1 Einleitung</b>	<b>13</b>
<b>2 Installation</b>	<b>17</b>
<b>3 Ein erster Überblick</b>	<b>19</b>
<b>4 Einführung in die Programmiersprache</b>	<b>36</b>
<b>5 Objekte</b>	<b>45</b>
<b>6 Daten importieren</b>	<b>63</b>
<b>7 Datenaufbereitung</b>	<b>71</b>
<b>8 Univariate deskriptive Statistik</b>	<b>113</b>
<b>9 Bivariate deskriptive Statistik</b>	<b>126</b>
<b>10 Datenvisualisierung</b>	<b>137</b>
<b>11 Wahrscheinlichkeitsverteilungen</b>	<b>183</b>
<b>12 Mittelwertsvergleiche mit <math>t</math>-Tests</b>	<b>189</b>
<b>13 Varianzanalyse ohne Messwiederholung</b>	<b>204</b>
<b>14 Varianzanalyse mit Messwiederholung</b>	<b>221</b>
<b>15 Grundlagen der Regressionsanalyse</b>	<b>238</b>
<b>16 Spezielle Regressionsmodelle</b>	<b>256</b>
<b>17 Nonparametrische Verfahren</b>	<b>285</b>
<b>18 Verfahren für die Testkonstruktion</b>	<b>294</b>
<b>19 Lineare Strukturgleichungsmodelle</b>	<b>309</b>

<b>20 Mehrebenenanalyse</b>	<b>327</b>
<b>21 Poweranalysen und Stichprobenumfangsplanung</b>	<b>343</b>
<b>22 Ausgaben exportieren</b>	<b>346</b>
<b>23 Tipps für den Workflow</b>	<b>353</b>
<b>24 Troubleshooting</b>	<b>355</b>
<b>Anhang A: Datensätze</b>	<b>358</b>
<b>Anhang B: Pakete</b>	<b>360</b>
<b>Hinweise zum Online-Material</b>	<b>361</b>
<b>Literatur</b>	<b>362</b>
<b>Sachwortverzeichnis</b>	<b>367</b>

# Inhalt

<b>Vorwort zur fünften Auflage</b>	<b>12</b>
<b>1 Einleitung</b>	<b>13</b>
1.1 Warum R?	14
1.2 Für wen ist dieses Buch?	14
1.3 Wie benutzt man dieses Buch?	15
1.4 Weiterentwicklungen und Aktualität des Buchs	15
1.5 Verwendete Schriftarten	16
<b>2 Installation</b>	<b>17</b>
2.1 Download und Installation von R	17
2.2 Download und Installation von RStudio	18
<b>3 Ein erster Überblick</b>	<b>19</b>
3.1 RStudio im Überblick	19
3.2 Zusätzliche Pakete	27
3.3 Hilfe zu R	31
3.4 Die wichtigsten Funktionen im Überblick	35
<b>4 Einführung in die Programmiersprache</b>	<b>36</b>
4.1 R als Taschenrechner	36
4.2 Logische Abfragen	37
4.3 Funktionen	38
4.4 Kommentare	43
4.5 Ein paar Stilregeln	43
4.6 Übungen	44
<b>5 Objekte</b>	<b>45</b>
5.1 Neue Objekte anlegen	45
5.2 Objekttypen	48
5.3 Elemente aus Objekten auswählen	53
5.4 Der Workspace	56
5.5 Objekte speichern und öffnen	58
5.6 Die wichtigsten Funktionen im Überblick	61
5.7 Übungen	62

<b>6</b>	<b>Daten importieren</b>	<b>63</b>
6.1	Daten aus Textdateien einlesen	63
6.2	Andere Datenformate einlesen	68
6.3	Daten betrachten	68
6.4	Daten speichern	69
6.5	Die wichtigsten Funktionen im Überblick	70
6.6	Übungen	70
<b>7</b>	<b>Datenaufbereitung</b>	<b>71</b>
7.1	Klassische und moderne Datenaufbereitung	71
7.2	Variablen erstellen und bearbeiten	74
7.3	Data Frames reduzieren	89
7.4	Data Frames sortieren	95
7.5	Data Frames zusammenfügen	97
7.6	Data Frames umstrukturieren	101
7.7	Fortgeschrittene Datenaufbereitung	107
7.8	Funktionen aus den R-Basispaketen im Überblick	109
7.9	tidyverse-Funktionen im Überblick	110
7.10	Übungen	111
<b>8</b>	<b>Univariate deskriptive Statistik</b>	<b>113</b>
8.1	Häufigkeitstabellen	113
8.2	Deskriptive Kennwerte	117
8.3	Gruppenvergleiche	122
8.4	Die wichtigsten Funktionen im Überblick	124
8.5	Übungen	125
<b>9</b>	<b>Bivariate deskriptive Statistik</b>	<b>126</b>
9.1	Kontingenztabellen	126
9.2	Zusammenhangsmaße für metrische Variablen	129
9.3	Zusammenhangsmaße für nicht-metrische Variablen	134
9.4	Die wichtigsten Funktionen im Überblick	135
9.5	Übungen	136
<b>10</b>	<b>Datenvisualisierung</b>	<b>137</b>
10.1	Einführung in die klassischen Grafik-Funktionen	138
10.2	Einführung in ggplot2	150
10.3	Ausgewählte Diagramme	159

10.4	Die wichtigsten klassischen Grafik-Funktionen im Überblick	180
10.5	Zusätzliche Argumente für klassische Grafik-Funktionen	180
10.6	Die wichtigsten ggplot2-Funktionen im Überblick	181
10.7	Zusätzliche Argumente für ggplot2-Funktionen	181
10.8	Übungen	182
<b>11</b>	<b>Wahrscheinlichkeitsverteilungen</b>	<b>183</b>
11.1	Grafische Darstellung der Wahrscheinlichkeitsverteilung	184
11.2	Berechnung von Quantilen	186
11.3	Berechnung von Flächenanteilen bzw. $p$ -Werten	186
11.4	Tests für die Prüfung der Normalverteilungsannahme	187
11.5	Die wichtigsten Funktionen im Überblick	188
11.6	Übungen	188
<b>12</b>	<b>Mittelwertvergleiche mit <math>t</math>-Tests</b>	<b>189</b>
12.1	Einstichproben- $t$ -Test	189
12.2	$t$ -Test für unabhängige Stichproben	194
12.3	$t$ -Test für abhängige Stichproben	199
12.4	Die wichtigsten Funktionen im Überblick	202
12.5	Übungen	203
<b>13</b>	<b>Varianzanalyse ohne Messwiederholung</b>	<b>204</b>
13.1	Einfaktorielle Varianzanalyse ohne Messwiederholung	204
13.2	Mehrfaktorielle Varianzanalyse ohne Messwiederholung	211
13.3	Die wichtigsten Funktionen im Überblick	219
13.4	Übungen	220
<b>14</b>	<b>Varianzanalyse mit Messwiederholung</b>	<b>221</b>
14.1	Datenstruktur	221
14.2	Einfaktorielle Varianzanalyse mit Messwiederholung	222
14.3	Mehrfaktorielle gemischte Varianzanalyse	228
14.4	Die wichtigsten Funktionen im Überblick	237
14.5	Übungen	237
<b>15</b>	<b>Grundlagen der Regressionsanalyse</b>	<b>238</b>
15.1	Einfache lineare Regression	238
15.2	Multiple Regression	243
15.3	Hierarchische Regression	245
15.4	Modellannahmen prüfen	249

15.5	Die wichtigsten Funktionen im Überblick	254
15.6	Übungen	255
<b>16</b>	<b>Spezielle Regressionsmodelle</b>	<b>256</b>
16.1	Kategoriale Prädiktoren	256
16.2	Kovarianzanalyse	260
16.3	Moderierte Regression	263
16.4	Nicht-lineare Regression	275
16.5	Logistische Regression	278
16.6	Die wichtigsten Funktionen im Überblick	283
16.7	Übungen	284
<b>17</b>	<b>Nonparametrische Verfahren</b>	<b>285</b>
17.1	Der $\chi^2$ -Test	285
17.2	Der Wilcoxon-Test	288
17.3	Der Kruskal-Wallis-Test	292
17.4	Die wichtigsten Funktionen im Überblick	293
17.5	Übungen	293
<b>18</b>	<b>Verfahren für die Testkonstruktion</b>	<b>294</b>
18.1	Exploratorische Faktorenanalyse	294
18.2	Itemanalyse und interne Konsistenz	304
18.3	Die wichtigsten Funktionen im Überblick	308
18.4	Übungen	308
<b>19</b>	<b>Lineare Strukturgleichungsmodelle</b>	<b>309</b>
19.1	Multiple Regression mit lavaan	309
19.2	Pfadmodell mit Mediatorvariable	311
19.3	Konfirmatorische Faktorenanalyse	317
19.4	Kombination von Mess- und Strukturmodell	322
19.5	Erstellen eines Pfaddiagramms	324
19.6	Weitere Funktionen und ergänzende Pakete	325
19.7	Die wichtigsten Funktionen im Überblick	325
19.8	Übungen	326
<b>20</b>	<b>Mehrebenenanalyse</b>	<b>327</b>
20.1	Das Nullmodell	327
20.2	Das Random-Intercept-Modell	330
20.3	Das Random-Slopes-Modell	335

20.4	Modelle mit Ebene-2-Prädiktoren	338
20.5	Die wichtigsten Funktionen im Überblick	341
20.6	Übungen	342
<b>21</b>	<b>Poweranalysen und Stichprobenumfangsplanung</b>	<b>343</b>
21.1	Poweranalysen mit dem pwr-Paket	343
21.2	Simulationsbasierte Poweranalysen	344
21.3	Übungen	345
<b>22</b>	<b>Ausgaben exportieren</b>	<b>346</b>
22.1	Daten exportieren	346
22.2	Tabellen exportieren	347
22.3	Diagramme exportieren	348
22.4	Kommentierte Ausgaben mit R Markdown erstellen	349
22.5	Die wichtigsten Funktionen im Überblick	352
<b>23</b>	<b>Tipps für den Workflow</b>	<b>353</b>
23.1	Ordner- und Dateienstruktur	353
23.2	RStudio-Projekte und relative Pfade	353
23.3	Aufbau eines R-Skripts	354
<b>24</b>	<b>Troubleshooting</b>	<b>355</b>
24.1	Typische Tippfehler	355
24.2	Häufige Fehlermeldungen	356
24.3	Vorgehen bei der Fehlersuche	356
24.4	Probleme mit der Darstellung von R-Skripten	357
<b>Anhang A: Datensätze</b>		<b>358</b>
<b>Anhang B: Pakete</b>		<b>360</b>
<b>Hinweise zum Online-Material</b>		<b>361</b>
<b>Literatur</b>		<b>362</b>
<b>Sachwortverzeichnis</b>		<b>367</b>

## Vorwort zur fünften Auflage

In den fünf Jahren seit Erscheinen der letzten Auflage dieses Buchs ist R zu einem der beliebtesten Statistik-Programme in Lehre und Forschung und auch außerhalb der Hochschulen geworden. Auch die Ressourcen für R haben stark zugenommen. Habe ich in der letzten Auflage noch auf wenige einschlägige Webseiten zu R verweisen können, ist das Angebot an kostenlos verfügbaren Video-Tutorials, eBooks, Webinaren und Kurzanleitungen mittlerweile fast unüberschaubar.

Wozu dann noch ein klassisches Lehrbuch zu R? Ich gebe mittlerweile seit vielen Jahren Kurse zur Einführung in R. In diesen Kursen wird immer wieder deutlich, dass das große Angebot an Ressourcen für Einsteigerinnen und Einsteiger häufig überfordernd ist. Ziel dieses Buchs ist es daher, den Einstieg in R zu strukturieren und damit den Leserinnen und Lesern ein Grundverständnis von R zu vermitteln, mit dem sie anschließend die zahlreichen großartigen Online-Angebote effizient nutzen können. In diesem Buch werden die statistischen Verfahren behandelt, die in den Sozialwissenschaften am häufigsten gelehrt und angewandt werden. Für die fünfte Auflage wurden alle Kapitel aktualisiert und teilweise komplett neu geschrieben (z. B. die Kapitel zu Datenaufbereitung und Datenvisualisierung). Zudem sind einige kleinere Kapitel dazugekommen, z. B. zu Workflow und Troubleshooting.

Eine solche umfassende Überarbeitung des Buchs war nur möglich durch tatkräftige Unterstützung. Ich möchte mich besonders bei Peter Hähner, Flavio Schröder, Marie von Rogal und Merle Widlok bedanken, die intensiv die Texte und Befehle der vorherigen und der neuen Auflage geprüft und korrigiert haben. Ein großer Dank geht auch an die R Community auf Twitter, über die ich immer wieder neue Tipps und Updates zu R erhalte und die meine Fragen zu R schnell und hilfreich beantwortet haben. Zudem bedanke ich mich herzlich bei meinem Team, meinen Studierenden, Workshop-Teilnehmerinnen und -Teilnehmern, Kolleginnen und Kollegen und natürlich bei allen Leserinnen und Lesern, deren Fehlermeldungen, konstruktive Verbesserungsvorschläge und positive Rückmeldungen diese neue Auflage möglich gemacht haben.

Für alle verbleibenden Unklarheiten und Fehler bin ich natürlich allein verantwortlich. Bitte kontaktieren Sie mich, wenn Sie Fehler entdecken sollten oder Vorschläge zur Verbesserung des Buchs machen möchten.

# 1 Einleitung

In den letzten Jahren hat die Statistik-Software R (R Core Team, 2019b) deutlich an Beliebtheit gewonnen. Nicht nur an den meisten Universitäten wird R mittlerweile in Lehre und Forschung eingesetzt, sondern auch in vielen Unternehmen (Bhalla, 2016). Die Vorteile von R sind schon lange bekannt – warum also jetzt dieser rasante Aufstieg? Dafür gibt es mindestens zwei Gründe.

Zum einen leben wir in einem Daten-Zeitalter, d. h., Daten werden kontinuierlich und von in großen Mengen gesammelt. Die Nutzung dieser Daten nimmt sowohl in der Forschung als auch in der freien Wirtschaft einen immer größeren Stellenwert ein, und unter dem Stichwort »Data Science« hat sich mittlerweile eine ganze Berufsgruppe etabliert, deren Spezialität das Aufbereiten und Auswerten großer Datenmengen ist. Eines der in diesem Feld am häufigsten eingesetzten Statistik-Programme ist R (Dataflair Team, 2019).

Zum anderen findet in der sozialwissenschaftlichen Forschung, besonders in der Psychologie, derzeit ein Umbruch statt. Immer mehr Forscherinnen und Forscher, wissenschaftliche Institutionen und Fachzeitschriften verschreiben sich den Open-Science-Prinzipien (z. B. Schönbrodt et al., 2017). Zu Open Science gehört unter anderem, dass man Daten und Auswertungsskripte zur Verfügung stellt, damit andere die veröffentlichten Ergebnisse reproduzieren können. Damit wirklich alle Menschen, egal woher sie kommen oder wie viel Geld sie haben, diese Dateien nutzen können, müssen diese in nicht-proprietärer Software erstellt worden sein – also in einer Software, die nicht kommerziell vertrieben wird. Ehemals beliebte kommerzielle Statistik-Programme wie SPSS oder Stata erfüllen dieses Kriterium nicht. R schon.

Es gibt also eine Menge Gründe, R zu lernen. Allerdings ist R nicht die intuitivste Software. Das Ziel dieses Buchs ist es, Ihnen den Einstieg in R zu erleichtern. Sie werden sehen: Wenn Sie die ersten Kapitel gemeistert haben, wird es sehr viel einfacher. Ich wünsche Ihnen beim Einstieg in R viel Erfolg und hoffentlich auch etwas Spaß!

## 1.1 Warum R?

Es gibt viele Gründe, die für R sprechen. Hier sind die wichtigsten:

- ▶ **R kann mehr.** Die Software enthält viele Funktionen, die man bei kommerziellen Statistik-Programmen vergeblich sucht.
- ▶ **R ist aktuell.** Das Programm wird ständig weiterentwickelt, sodass viele neue statistische Methoden zuerst in R und erst Jahre später in anderen Programmen implementiert sind.
- ▶ **R ist ansprechbar.** Die Programmierer der einzelnen Pakete sind über E-Mail erreichbar und reagieren meiner Erfahrung nach innerhalb weniger Tage. So können Fragen schnell geklärt und eventuelle Programmierfehler behoben werden.
- ▶ **R schläft nicht.** R-Nutzer gibt es auf der ganzen Welt, und viele von ihnen sind in den einschlägigen Foren und in den sozialen Medien aktiv. So kann man seine Fragen rund um die Uhr klären.
- ▶ **R ist nachgefragt.** Wenn Sie R-Kenntnisse nachweisen können, stehen Ihnen viele Jobmöglichkeiten offen – in und außerhalb der Wissenschaft.
- ▶ **R ist kostenlos!**

## 1.2 Für wen ist dieses Buch?

Dieses Buch richtet sich sowohl an Einsteiger, die zum ersten Mal mit einer Statistik-Software arbeiten, als auch an Umsteiger, die R als eine Alternative zu anderen Statistik-Programmen ausprobieren möchten. Besondere Computer- oder gar Programmierkenntnisse brauchen Sie nicht. Da dieses Buch aber keine Statistik-Einführung ist, sollten Sie die Grundlagen der hier behandelten statistischen Tests kennen. Dafür empfehle ich das Lehrbuch von Eid et al. (2017), an dem auch die Gliederung dieses Buchs orientiert ist.

Ich bin selbst Psychologin und habe daher Datenbeispiele aus psychologischen Studien gewählt. Darüber hinaus behandle ich einige Funktionen und statistische Verfahren, die besonders für Psychologinnen und Psychologen von Interesse sind, z. B. die Berechnung von Skalenwerten (Kap. 7) oder Verfahren für die Testkonstruktion wie Itemanalysen und Faktorenanalysen (Kap. 18). Grundsätzlich richtet sich das Buch jedoch an Interessierte aus allen Fachrichtungen der Sozialwissenschaften.

In einem Einführungsbuch ist natürlich kein Platz, alle speziellen statistischen Verfahren detailliert zu besprechen. Sie werden jedoch bald in der Lage sein, sich in die Funktionen und Pakete, die Sie für solche Verfahren brauchen, selbst einzuarbeiten. Weitere Hinweise dazu finden Sie im Anhang B: Pakete.

### 1.3 Wie benutzt man dieses Buch?

Den Umgang mit einer Statistik-Software lernt man am besten durch Ausprobieren. Daher empfehle ich, bei der Lektüre dieses Buchs immer einen Computer mit R dabei zu haben. So können Sie die beschriebenen Funktionen direkt ausprobieren. Die Datensätze dafür können Sie von der Webseite zu diesem Buch herunterladen: <http://www.beltz.de/r-fuer-einsteiger>. Dort finden Sie auch die Skripte für jedes Kapitel, die oftmals noch weitere Beispiele enthalten, sowie zahlreiche Zusatzmaterialien.

#### Hinweise für Einsteiger

Ich habe mich bemüht, dieses Buch so zu schreiben, dass auch Einsteiger ohne jegliche Erfahrung mit Statistik-Software mit R zurecht kommen können. In diesem Buch stelle ich die statistischen Verfahren in der Reihenfolge vor, in der sie typischerweise gelehrt werden. Wenn Sie dieses Buch also in Begleitung zu einer Statistik-Vorlesung in Psychologie oder anderen sozialwissenschaftlichen Fächern lesen, können Sie die Kapitel einfach von vorne nach hinten durcharbeiten.

#### Hinweise für Umsteiger

Als Umsteiger haben Sie bereits Erfahrung mit anderen Statistik-Programmen und interessieren sich möglicherweise nur für bestimmte Funktionen in R. In diesem Fall brauchen Sie natürlich nicht das Buch von vorne bis hinten durchzuarbeiten, sondern können sich auf die Funktionen konzentrieren, die Sie benötigen. Beginnen Sie aber auf jeden Fall mit den Kapiteln 2 bis 5, in denen die Grundlagen von R vermittelt werden.

### 1.4 Weiterentwicklungen und Aktualität des Buchs

R wird ständig weiterentwickelt. Sowohl die Basisversion als auch die zusätzlichen Pakete werden regelmäßig aktualisiert, und fast wöchentlich werden neue Pakete zur Verfügung gestellt. Wenn Sie regelmäßig mit R arbeiten, werden Sie vermutlich immer wieder neue Pakete entdecken, die Ihnen die statistischen Analysen noch leichter machen. Für die Erstellung dieses Buchs habe ich mich bemüht, immer die aktuellsten Versionen zu verwenden (Stand April 2020). Dieses Buch verwendet die R Version 3.6.2. Die hier vorgestellten Funktionen und Pakete sollten jedoch auch mit neueren Versionen funktionieren. Wenn Sie trotzdem veraltete Funktionen oder gar Fehler finden, lassen Sie mich dies bitte wissen. Aktualisierungen und Fehlerkorrekturen werden regelmäßig unter <http://www.beltz.de/r-fuer-einsteiger> und unter

dem Twitter-Account [@RfuerEinsteiger](#) veröffentlicht. Und noch ein Tipp: Die aktuellsten Hinweise zu den Funktionen und den entsprechenden Argumenten finden Sie in der dazugehörigen Hilfe-Datei (s. Abschn. 4.3.3).

## 1.5 Verwendete Schriftarten

Um die Übersichtlichkeit des Texts zu erhöhen, werden in diesem Buch verschiedene Schriftarten verwendet (s. Tab. 1.1).

**Tabelle 1.1** Verwendung von Schriftarten im Buch

Text	Beispiel
Menübefehle	DATEI → SPEICHERN
Namen von Paketen	psych-Paket
Namen von Funktionen	mean-Funktion
Befehle und Ausgaben	> c(1, 2, 3)

## 2 Installation

Um gut mit R arbeiten zu können, installieren wir zwei Programme: R (R Core Team, 2019b) und RStudio (RStudio Team, 2019). R ist die eigentliche Software, die zwar schon alles kann, was uns interessiert, aber leider optisch wenig ansprechend ist. Daher verwenden wir zusätzlich die Benutzeroberfläche RStudio. Wir installieren zunächst R und anschließend RStudio. Beide Programme sind kostenlos.

### 2.1 Download und Installation von R

Die Internetseite <http://cran.r-project.org> ist der Ausgangspunkt für die Arbeit mit R. Die Abkürzung CRAN steht für Comprehensive R Archive Network. Hier werden sowohl die Software als auch verschiedene Pakete (s. Abschn. 3.2) und Dokumente zur Verfügung gestellt. Direkt auf der Startseite befindet sich die Option *Download and Install R*. Hier kann man R für die Betriebssysteme Linux, MacOS X und Windows herunterladen.

#### Windows

Zum Download der Windows-Version gelangt man über den Link *Download R for Windows*. Auf der neuen Seite wählen wir den Link *base*. Auf der folgenden Seite klicken wir auf den Link *Download R X.X.X for Windows*. Damit laden wir die Set-up-Datei herunter. (Das X steht für die aktuelle Versionsnummer. Auf der CRAN-Seite wird immer nur die neueste R-Version angeboten.)

Wir starten die Installation von R, indem wir die heruntergeladene Set-up-Datei öffnen. Als erstes müssen wir die Sprache für das Set-up auswählen. Danach öffnet sich der Set-up-Assistent. Folgende Schritte werden bei der Installation durchlaufen:

- (1) Der Lizenztext wird angezeigt und sollte durchgelesen werden.
- (2) ZIELORDNER AUSWÄHLEN. Wir wählen den Ordner, in dem das Programm installiert wird. Standardmäßig wird der Ordner `C:\Programme\R\R-X.X.X` vorgeschlagen.
- (3) KOMPONENTEN AUSWÄHLEN. Hier wählen wir die Dateien aus, die installiert werden sollen. Für die meisten Leser wird die Option `BENUTZERINSTALLATION` geeignet sein. Dabei werden alle Dateien installiert.
- (4) STARTOPTIONEN ANPASSEN. Wählen wir die Option `JA`, können wir im Folgenden bestimmte Einstellungen von R bei der Installation festlegen. Dazu zählen der

Anzeigemodus (Darstellung der einzelnen Fenster), der Hilfestil (Hilfe-Dateien als PDF oder im Windows-Format) sowie der Internetzugang. Wählen wir **NEIN**, werden die einzelnen Schritte übersprungen und die Standardeinstellungen für R übernommen. Wir wählen die letztere Option.

- (5) **STARTMENÜORDNER AUSWÄHLEN**. Hier kann man entscheiden, ob ein Ordner im Startmenü erstellt werden soll und den Namen im Startmenü ändern.
- (6) **ZUSÄTZLICHE AUFGABEN AUSWÄHLEN**. Wir entscheiden nun, ob zusätzliche Symbole auf dem Desktop oder in der Schnellstartleiste angelegt werden sollen. Außerdem legen wir hier die Art der Verknüpfung von R mit Windows fest (Speicherung der Versionsnummer, R mit .RData-Dateien verknüpfen). Für die meisten Leser sind die Standardeinstellungen ausreichend.
- (7) Fertig!

### MacOS

Zum Download der Mac-Version gelangt man über den Link *Download R for MacOS X*. Wir gelangen nun auf die Seite *R for MacOS X*, wo wir die Datei *R-X.X.X.pkg* auswählen und herunterladen. Wenn wir diese Datei öffnen, startet der Set-up-Assistent. Dieser Assistent führt uns Schritt für Schritt durch die Installation. Bei der Installation werden die folgenden Schritte durchlaufen:

- (1) Wichtige Informationen lesen und zustimmen.
- (2) Software-Lizenzvereinbarungen lesen und akzeptieren.
- (3) Zielvolumen wählen: Soll die Software für alle Benutzer oder nur auf einem bestimmten Laufwerk installiert werden?
- (4) Installationstyp: Hier kann der Ort der Installation verändert werden.
- (5) Durchführung der Installation.
- (6) Fertig!

An dieser Stelle noch ein allgemeiner Hinweis: Dieses Lehrbuch wurde mit Windows erstellt. Die Benutzeroberfläche sieht auf dem Mac manchmal etwas anders aus. Grundsätzlich lassen sich aber alle Funktionen an denselben Stellen wiederfinden. Alle Pakete lassen sich auch auf MacOS verwenden.

## 2.2 Download und Installation von RStudio

Nachdem die Installation von R erfolgreich abgeschlossen ist, laden wir die Benutzeroberfläche RStudio von der Seite <http://www.rstudio.com/ide/> herunter und installieren sie. Nach der Installation steht RStudio als eigenes Programm zur Verfügung. R muss zwar installiert sein, aber nicht gestartet werden, um RStudio zu nutzen. Ab jetzt arbeiten wir immer direkt in RStudio!

# 3 Ein erster Überblick

Wenn wir R und RStudio erfolgreich installiert haben, können wir loslegen. Wir öffnen immer direkt RStudio. R muss und sollte nicht zusätzlich gestartet werden, sondern läuft automatisch im Hintergrund mit. In diesem Kapitel verschaffen wir uns einen ersten Überblick über das Programm (Abschn. 3.1) und besprechen, wie man zusätzliche Pakete installieren und laden kann (Abschn. 3.2). Wenn man mal nicht weiter weiß, kann man sich an verschiedenen Stellen Hilfe zu R holen (Abschn. 3.3).

## 3.1 RStudio im Überblick

Die Benutzeroberfläche von RStudio setzt sich üblicherweise aus vier nebeneinander angeordneten Fenstern zusammen (Abb. 3.1): das Skript-Fenster oben links, die Konsole unten links, das Environment-Fenster oben rechts, und ein Fenster mit einer ganzen Reihe von Reitern unten rechts. Bei der ersten Nutzung von RStudio wird das Skript-Fenster nicht immer sofort angezeigt. Wie man es öffnet, wird in Abschn. 3.1.2 beschrieben. In den folgenden Abschnitten gehen wir die Fenster einzeln durch.

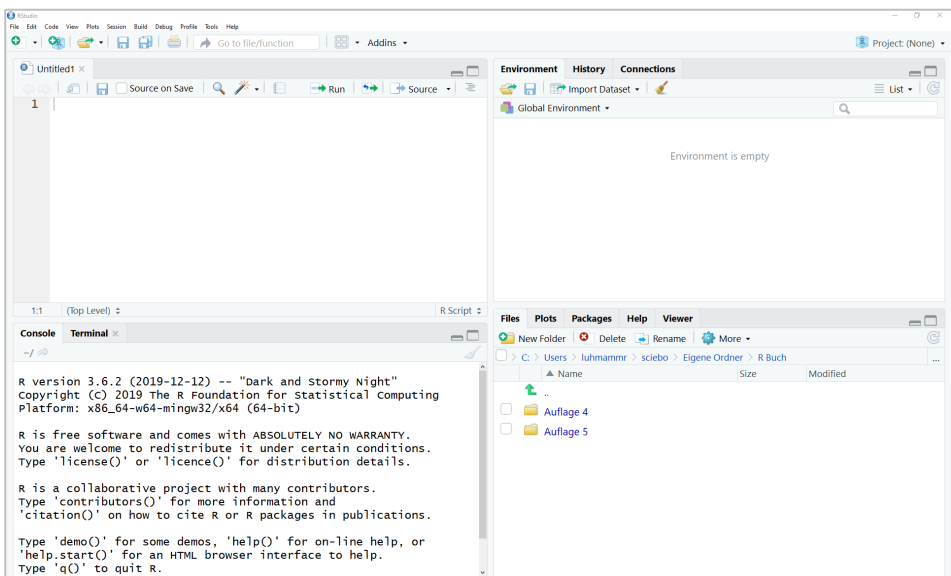


Abbildung 3.1 RStudio

### 3.1.1 Die Konsole

Die Konsole bietet eigentlich alles, was wir für die Arbeit mit R brauchen. Hier können wir sowohl Befehle eingeben als auch Ausgaben betrachten. Wenn wir RStudio öffnen, steht in der Konsole ein kurzer Begrüßungstext (s. Abb. 3.1), der u. a. angibt, welche R Version gerade geladen ist (hier: Version 3.6.2, veröffentlicht am 12.12.2019 unter dem Spitznamen »Dark and Stormy Night«). Unter dem Begrüßungstext erscheint das folgende Symbol:

```
>
```

Bei diesem Symbol handelt es sich um ein Eingabesymbol, d. h., dahinter können wir einen Befehl eingeben. In der R-Sprache wird dieses Symbol als »Prompt« (deutsch: Eingabeaufforderung) bezeichnet. Um einen Befehl in der Konsole auszuführen, drücken wir die Enter-Taste. Die Ausgabe erscheint dann direkt unter dem Befehl in einer anderen Farbe. Wenn alles gut gegangen ist, erwartet uns darunter schon wieder das nächste Eingabesymbol. Wir probieren das mal mit dem folgenden Befehl aus:

```
> 1+2
[1] 3
>
```

In diesem Beispiel haben wir eine kleine Rechenaufgabe lösen lassen. R kann man nämlich auch als Taschenrechner verwenden (mehr dazu in Abschn. 4.1). Das Ergebnis erscheint in der Zeile unter dem Befehl. Diesen Teil bezeichnet man auch als Ausgabe oder Output.

Die Ausgabe beginnt mit [1]. Das eigentliche Ergebnis steht dahinter. Das Zeichen [1] hat nichts mit dem Ergebnis der Rechenaufgabe zu tun, sondern wird verwendet, um bei langen Ausgaben, die über mehrere Zeilen gehen, die Elemente zu nummerieren. Das klingt jetzt noch sehr abstrakt, wird aber klarer, wenn wir später Beispiele für solche Ausgaben betrachten (s. Abschn. 5.4).

Unter der Ausgabe erscheint erneut das Eingabesymbol. Das ist ein gutes Zeichen, denn es bedeutet, dass der Befehl vollständig ausgeführt wurde und die Konsole jetzt bereit für den nächsten Befehl ist. Wurde der Befehl dagegen nicht vollständig ausgeführt (z. B. weil eine Klammer nicht geschlossen wurde), fügt die Konsole automatisch eine zweite Zeile ein, die mit einem + beginnt. Das + zeigt an, dass diese Zeile mit der vorhergehenden Zeile verknüpft ist.

Im Folgenden sind zwei Befehle dargestellt. Der erste Befehl geht über eine Zeile, daher folgt die Ausgabe direkt in der zweiten Zeile. In den zweiten Befehl wurde dagegen ein Zeilenumbruch eingefügt, sodass dieser Befehl über zwei Zeilen geht. Das

+ am Anfang der zweiten Zeile ist hier kein mathematisches Symbol, sondern es verknüpft lediglich die beiden Befehlszeilen miteinander. Die Ausgabe folgt in der dritten Zeile:

```
> (3+7+2) / 3  
[1] 4
```

```
> (3+7+2) /  
+ 3  
[1] 4
```

### Tipp

Bevor man einen neuen Befehl ausführt, sollte man immer sicherstellen, dass die unterste Zeile in der Konsole mit `>` beginnt. Beginnt sie stattdessen mit `+`, so ist dies ein Zeichen, dass der vorangegangene Befehl unvollständig eingegeben und daher noch nicht beendet wurde. Das passiert z. B., wenn man vergisst, Klammern oder Anführungszeichen zu schließen. Diese und weitere typische Fehlerquellen werden in Kapitel 24 behandelt.

### Konsole aufräumen

Nach einer längeren R-Sitzung ist die Konsole meistens ziemlich voll und unübersichtlich. Wenn uns das stört, können wir die Konsole mit der Tastenkombination `STRG + L` (Windows) bzw. `CMD + L` (Mac) aufräumen. Damit werden alle Inhalte der Konsole gelöscht.

### Inhalt der Konsole speichern

Die Konsole enthält alle Befehle und Ausgaben, die in einer Sitzung angefordert wurden. Und damit ist wirklich alles gemeint – also auch alle Befehle mit Tippfehlern, mehrfach ausgeführte Befehle und Fehlermeldungen. Daher wird der Inhalt der Konsole üblicherweise nicht gespeichert. Stattdessen kann man bestimmte Ausgaben (z. B. Tabellen oder Diagramme) gezielt in externe Dateien zu exportieren oder eine sog. R Markdown-Datei anlegen (s. Kap. 22). Wenn man trotzdem den Inhalt der Konsole speichern möchte, kann man diesen markieren, kopieren und in ein einfaches Textdokument einfügen.

### 3.1.2 Das Skript-Fenster

Zwar könnte man theoretisch alle Befehle direkt in die Konsole eingeben, wie im vorherigen Abschnitt gezeigt, aber wir lassen das lieber bleiben. Der Nachteil an der Konsole ist nämlich, dass die Befehle immer wieder neu eingegeben werden müssen und nicht gut gespeichert werden können. Deshalb geben wir Befehle ab jetzt immer in das Skript-Fenster ein. Falls dieses Fenster noch nicht zu sehen ist, können wir es über `FILE → NEW FILE → R SCRIPT` öffnen.

#### **Befehle im Skript-Fenster eingeben und ausführen**

Im Skript-Fenster erscheint kein Eingabesymbol, sondern der Befehl wird direkt eingegeben. Ein Befehl kann über eine oder mehrere Zeilen gehen. Bei langen Befehlen mit vielen Argumenten (s. Abschn. 4.3) ist es übersichtlicher, wenn man den Befehl durch Zeilenumbrüche auf verschiedene Zeilen aufteilt. Dazu setzen wir den Cursor an die Stelle, wo der Zeilenumbruch eingefügt werden soll, und drücken die `ENTER`-Taste. RStudio rückt alle folgenden Zeilen etwas ein, sodass man schnell erkennen kann, wo ein Befehl anfängt und aufhört. So wird die Rechenaufgabe aus dem vorherigen Abschnitt im Skript-Fenster dargestellt, wenn wieder ein Zeilenumbruch eingefügt wird:

```
(3+7+2) /  
3
```

Wie führt man nun diesen Befehl aus? Anders als in der Konsole reicht hier die `ENTER`-Taste nicht aus, denn damit würde nur ein Zeilenumbruch eingefügt werden. Wir können stattdessen auf ein Symbol klicken oder eine spezielle Tastenkombination verwenden.

Zunächst müssen wir aber festlegen, welche Befehle genau ausgeführt werden sollen. Dazu gibt es mehrere Möglichkeiten: (1) Wir setzen den Cursor irgendwo in den Befehl, der ausgeführt werden soll. Es wird dann genau dieser Befehl von Anfang bis Ende ausgeführt. (2) Wir markieren den Befehl mit der Maus. Dann wird alles ausgeführt, was markiert ist. Wir können damit mehrere Befehle gleichzeitig ausführen. Aber Vorsicht: Hier wird wirklich nur der markierte Teil ausgeführt – haben wir z. B. vergessen, eine Klammer mitzumarkieren, dann wird der Befehl unvollständig ausgeführt. Diese Eigenschaft klingt zunächst etwas unpraktisch, hat aber einen wichtigen Vorteil: Wir können dadurch gezielt Teile von Befehlen ausführen, was besonders nützlich ist, wenn wir es später mit langen, geschachtelten Befehlen zu tun haben. Zudem kann man diese Eigenschaft für die Fehlersuche nutzen (s. Kap. 24).

Um den Befehl auszuführen, klicken wir entweder auf das Symbol RUN auf der rechten Seite der oberen Leiste des Skript-Fensters oder verwenden die Tastenkombination STRG + ENTER (Windows) bzw. CMD + ENTER (Mac). Dadurch erscheint die Ausgabe zusammen mit dem Befehl in der Konsole. Der Text im Skript-Fenster bleibt unverändert. Diese Trennung zwischen Eingaben und Ausgaben hat u. a. den Vorteil, dass man alte Befehle jederzeit wieder aufrufen und erneut ausführen kann, sogar mit anderen Datensätzen.

### Darstellung von Befehlen in RStudio

Eben haben wir schon gesehen, dass RStudio nachfolgende Zeilen bei mehrzeiligen Befehlen einrückt. Auch bei anderen Dingen ist RStudio sehr bemüht, die Befehle so übersichtlich wie möglich darzustellen.

Ein Beispiel dafür ist die Verwendung von verschiedenen Farben. So werden Funktionen (mehr dazu in Abschn. 4.3), Texteingaben (also alles, was von Anführungszeichen umgeben ist) und sonstige Bestandteile von Befehlen in unterschiedlichen Farben dargestellt. Welche das sind, kann man selbst beeinflussen: Unter TOOLS → GLOBAL OPTIONS → APPEARANCE kann man u. a. die farbliche Darstellung der Befehle und die Hintergrundfarbe verändern.

### Darstellung von Befehlen in diesem Buch

In diesem Lehrbuch stellen wir Befehle immer so dar, wie sie in der Konsole erscheinen. Das heißt, die Befehle beginnen immer mit dem Eingabesymbol, um sie besser von Ausgaben zu unterscheiden. Wenn man diese Befehle im Skript-Fenster eingibt, muss das Eingabesymbol weggelassen werden (s. Tab. 3.1).

**Tabelle 3.1** Darstellung von Befehlen im Lehrbuch und Eingabe dieser Befehle im Skript-Fenster

Darstellung im Lehrbuch	Eingabe im Skript-Fenster
> 3+4	3+4
> mean(x)	mean(x)
> vektor <- c(1, 2, 3)	vektor <- c(1, 2, 3)
> plot(y ~ x, pch = 2, las = 1)	plot(y ~ x, pch = 2, las = 1)

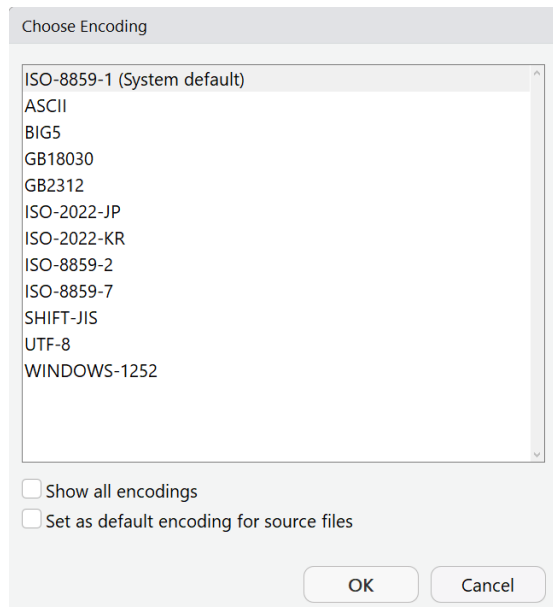
### Skript speichern

Die Befehle, die wir in das Skript-Fenster schreiben, können wir als Datei speichern. Diese Datei wird Skript genannt und erhält die Dateierdung .R. Um das Skript zu

speichern, wählen wir im Menü die Option FILE → SAVE AS oder verwenden die Tastenkombination STRG + S (Windows) bzw. CMD + S (Mac).

Beim erstmaligen Speichern kann es sein, dass sich ein Fenster öffnet, in dem wir die Encodierung der Skript-Datei festlegen sollen (s. Abb. 3.2). Hier wird festgelegt, wie Sonderzeichen wie ä, ö, ü oder ß gespeichert werden. Als erste Option wird immer die Encodierung angezeigt, die auf dem aktuellen Computer als Standard eingestellt ist. Wenn man diese Encodierung wählt, kann man das Skript an allen Computern, die diese Encodierung als Standard eingestellt haben, problemlos öffnen. Öffnet man das Skript jedoch an einem Computer mit einer anderen Encodierung, werden Sonderzeichen kryptisch dargestellt, z. B. wird aus einem hübschen ä ein weniger hübsches Åª. Das passiert z. B. dann, wenn man eine auf Mac erstellte Skript-Datei auf einem Windows-PC öffnet, da Mac und Windows unterschiedliche Encodierungen verwenden. Wie man dieses Problem vermeidet und wie man mit einer anders encodierten Skript-Datei umgeht, wird in Kapitel 24 vertieft.

Warum sollte man überhaupt ein Skript anlegen und speichern? Der wichtigste Vorteil ist, dass man Befehle einfach modifizieren und wiederholt ausführen kann. Ein gutes Skript kann man vollständig markieren und in einem Rutsch ausführen, ohne dass man zwischendurch noch etwas per Hand machen muss. Wie ein gutes Skript aufgebaut und gestaltet ist, besprechen wir in Kapitel 23.



**Abbildung 3.2** Auswahlfenster zur Wahl der Encodierung des Skripts

### Skript öffnen

Von RStudio aus öffnen wir ein Skript über `FILE` → `OPEN FILE`. Das Skript wird dann in einem neuen Skript-Fenster angezeigt. Man kann aber eine Skript-Datei auch durch Anklicken im Explorer (Windows) bzw. im Finder (Mac) öffnen.

### 3.1.3 Das Environment-Fenster

Oben rechts befindet sich das Environment-Fenster. Hier werden später alle Objekte angezeigt, die wir in einer R-Sitzung erstellt oder geladen haben (s. Kap. 5). Außerdem erscheinen hier standardmäßig noch zwei weitere Reiter: `HISTORY` und `CONNECTIONS`. Im History-Fenster werden alle Befehle aufgelistet, die wir ausführen – genau in der Reihenfolge, wie wir sie aufgeführt haben, ohne Kommentare, ohne Ergebnisse und vor allem inklusive sämtlicher Tippfehler. Diese Auflistung ist selten interessant und noch seltener hilfreich, daher ignorieren wir sie in diesem Buch. Im Connections-Fenster können wir Verbindungen zu Datenquellen an anderen Orten (z. B. im Internet) aufbauen. Auch diese Funktion benötigen wir in diesem Buch nicht.

### 3.1.4 Files, Plots, Packages, Help, Viewer

Im Bereich unten rechts finden wir mehrere Fenster, die wir über verschiedene Reiter ansteuern können. Unter `FILES` befindet sich ein Bereich, in dem wir durch die Verzeichnisse auf unserem Computer navigieren und Dateien öffnen können. Wenn wir später Diagramme erstellen (s. Kap. 10), erscheinen diese im Fenster `PLOTS`. Der Bereich `PACKAGES` enthält eine Liste aller Pakete, die wir installiert haben (s. Abschn. 3.2). Wenn wir Hilfe-Dateien zu bestimmten R-Paketen oder Funktionen aufrufen (s. Abschn. 4.3.3), erscheinen diese im Fenster `HELP`. Der `VIEWER` ist vor allem relevant, wenn wir Dokumente in R Markdown erstellen (s. Abschn. 22.4).

### 3.1.5 RStudio anpassen

Unter `TOOLS` → `GLOBAL OPTIONS` kann man das Aussehen und einige Funktionalitäten von RStudio anpassen (s. Abb. 3.3). So können wir z. B. unter `GENERAL` ein bestimmtes Verzeichnis als Standard-Arbeitsverzeichnis (`DEFAULT WORKING DIRECTORY`) festlegen (s. Abschn. 5.5.1) oder unter `CODE` Einstellungen zur Darstellung und Ausführung von R-Befehlen verändern. Wir lassen erstmal alle Einstellungen so, wie sie sind. Wenn Sie aber bei der fortgeschrittenen Arbeit mit R merken, dass manche Standardeinstellungen nicht zu Ihrem Arbeitsstil passen, finden Sie in den `GLOBAL OPTIONS` häufig eine Lösung.

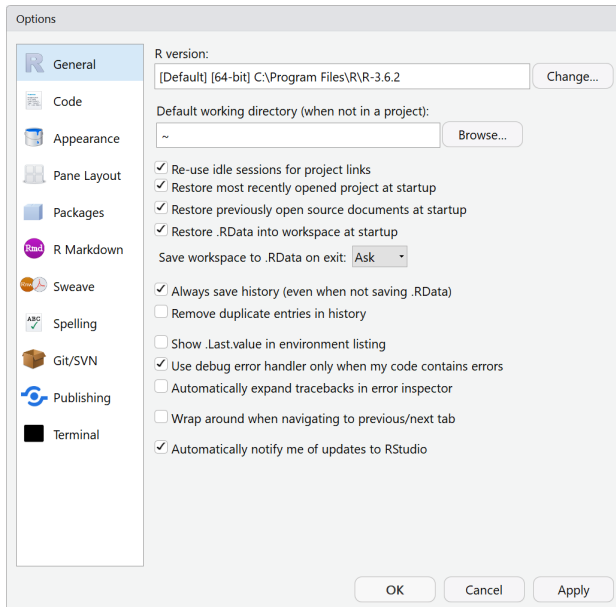


Abbildung 3.3 Global Options

### 3.1.6 RStudio schließen

Bevor wir RStudio schließen, sollten wir prüfen, ob alle Skripte als Skript-Dateien gespeichert sind (s. Abschn. 3.1.2). Wir schließen das Programm dann wie gewohnt, indem wir auf das Kreuz oben rechts klicken, im Menü die Option `FILE` → `QUIT SESSION` wählen, oder die Tastenkombination `STRG + Q` (Windows) bzw. `CMD + Q` (Mac) verwenden.

Dann passiert allerdings noch etwas, was für Einsteiger erstmal verwirrend ist: Wir werden gefragt, ob der Workspace bzw. das Global Environment gesichert werden soll. Den Workspace lernen wir in Kapitel 5 kennen. Er enthält sog. Objekte, z. B. Daten, mit denen wir in der R-Sitzung gearbeitet haben. Wenn wir auf `JA` klicken, wird der Inhalt des Workspaces in einer namenlosen Datei im Arbeitsverzeichnis gespeichert. Diese Datei wird dann jedesmal automatisch geöffnet, wenn wir R erneut starten. Dies ist fast nie wünschenswert, daher sollten wir hier immer die Option `NEIN` wählen. Bei den `GLOBAL OPTIONS` kann man außerdem einstellen, dass diese Meldung gar nicht mehr erscheint. Unter `TOOLS` → `GLOBAL OPTIONS` → `GENERAL` ändern wir die Einstellung bei `SAVE WORKSPACE TO .RDATA ON EXIT` zu `NEVER`.

## 3.2 Zusätzliche Pakete

R ist eine Open-Source-Software. Allgemein bedeutet das, dass der Quellcode für das Programm offen ist und von den Nutzerinnen und Nutzern geändert werden kann. Bei R ist das eingeschränkt der Fall: Die Basisversion, die wir mit der Installation von R erhalten, kann nicht von Ihnen oder von mir verändert werden. Das ist sicher sinnvoll, denn die Wahrscheinlichkeit, dass wir dabei etwas kaputtmachen, ist hoch (zumindest bei mir). Was tun also Menschen, die R verbessern oder erweitern möchten? Sie programmieren eigene Funktionen und stellen diese in sog. Paketen der Community zur Verfügung. Durch dieses Prinzip sind neu entwickelte statistische Methoden meistens als erstes in R verfügbar, Jahre bevor sie in kommerziellen Statistik-Programmen implementiert werden. Und die Anzahl der Pakete wächst exponentiell: Zum Zeitpunkt der vorherigen Auflage dieses Buchs (März 2015) standen ca. 6000 Pakete zur Verfügung, mittlerweile (Mai 2020) sind es bereits über 15000! Auf der Webseite <https://cran.r-project.org> sind alle verfügbaren Pakete aufgelistet und beschrieben.

Um Pakete zu nutzen, müssen zwei Schritte durchgeführt werden: Die Pakete müssen einmalig installiert werden (Abschn. 3.2.1) und die Pakete müssen in jeder Sitzung neu geladen werden (Abschn. 3.2.2).

### 3.2.1 Pakete installieren

In RStudio installieren wir Pakete, indem wir uns entweder im Menü durchklicken oder direkt einen Befehl eingeben. Wir beginnen mit der Klick-Variante, da diese für die meisten Einsteiger intuitiver ist. Fortgeschrittene bevorzugen aber meistens die Eingabe eines Befehls, da dies schneller geht.

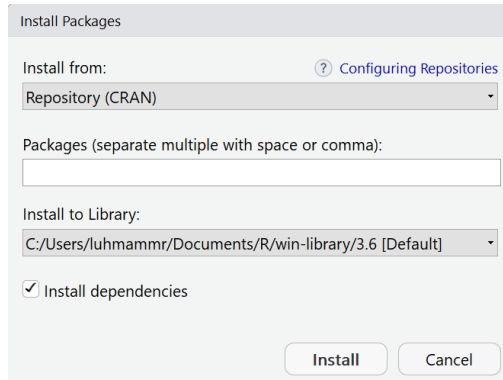
Zur Veranschaulichung installieren wir jetzt das Paket `psych` (Revelle, 2019). Dieses Paket wurde vom Persönlichkeitspsychologen William Revelle programmiert und enthält eine Reihe nützlicher Funktionen, die wir häufig anwenden werden.

#### Menübefehl

In RStudio finden wir die Menüoption für das Installieren von Paketen unter **TOOLS** → **INSTALL PACKAGES**. Alternativ können wir auch im Fenster unten rechts den Reiter **PACKAGES** auswählen und auf **INSTALL** klicken.

In dem sich öffnenden Fenster (Abb. 3.4) geben wir die Namen der Pakete ein, die wir installieren möchten, bei uns also `psych`. Wenn wir mehrere Pakete auf einmal installieren möchten, können wir die Namen der Pakete einfach hintereinander eingeben, getrennt mit einem Leerzeichen oder einem Komma. Die anderen Einstellun-

gen in diesem Fenster beziehen sich darauf, von wo die Pakete heruntergeladen werden und wo sie auf dem Rechner installiert werden sollen. Ich empfehle, an diesen Einstellungen nichts zu ändern. Wir achten nun noch darauf, dass bei der Option `INSTALL DEPENDENCIES` ein Häkchen gesetzt ist. Was ist das? Viele Pakete nutzen Funktionen aus anderen Paketen und funktionieren daher nur dann reibungslos, wenn diese anderen Pakete ebenfalls direkt installiert werden. Wenn das Häkchen gesetzt ist, werden also alle Pakete heruntergeladen und installiert, von denen das uns eigentlich interessierende Paket abhängt.



**Abbildung 3.4** Eingabefenster für die Installation von Paketen

Wenn wir nun auf `INSTALL` klicken, wird das Paket heruntergeladen und installiert. In der Konsole erscheint dann die folgende (hier verkürzt dargestellte) Ausgabe:

```
> install.packages("psych", dependencies = TRUE)
Installing package into 'C:/Users/.../R/win-library/3.6'
(as 'lib' is unspecified)
also installing the dependencies 'abind', 'coda', 'arm',
'pbivnorm', 'numDeriv', 'matrixcalc', 'mi', 'minqa', 'nloptr',
'Rcpp', 'RcppEigen', 'psychTools', 'GPArotation', 'lavaan',
'sem', 'lme4', 'Rcscdp'

trying URL 'https://cran.rstudio.com/bin/windows/
contrib/3.6/psych_1.9.12.31.zip'
Content type 'application/zip' length 3802421 bytes (3.6 MB)
downloaded 3.6 MB
package 'psych' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\...\downloaded_packages
```

Hier wurde also nicht nur das `psych`-Paket, sondern auch diverse andere Pakete installiert, von denen das `psych`-Paket abhängt. Woran erkenne ich, ob die Installation geklappt hat? Achten Sie auf die Meldung `package 'psych' successfully unpacked and MD5 sums checked`. Wenn die kommt, ist alles gut.

### Funktion

Pakete können auch über die folgende Funktion installiert werden:

```
> install.packages("Name.des.Pakets", dependencies = TRUE)
```

Das Argument `dependencies` bezieht sich wiederum darauf, ob auch andere Pakete installiert werden sollen. Steht dahinter `TRUE`, werden alle Pakete, von denen unser Paket abhängt, direkt mitinstalliert. Der entsprechende Befehl für die Installation des `psych`-Pakets lautet:

```
> install.packages("psych", dependencies = TRUE)
```

#### Tipp

R unterscheidet zwischen Groß- und Kleinschreibung. Folgende Pakete gibt es nicht: `PSYCH`, `Psych`, `psYch` usw.

### 3.2.2 Pakete laden

Die zusätzlich installierten Pakete werden beim Öffnen von Rstudio nicht automatisch geladen. Das heißt, wir müssen ein Paket jedes Mal laden, wenn wir es in einer neuen Sitzung verwenden möchten. Es ist allerdings nicht notwendig, das Paket jedes Mal neu zu installieren. Woher wissen wir, ob ein Paket bereits installiert ist? In RStudio können wir im Fenster unten rechts unter dem Reiter `PACKAGES` eine Liste aller bereits installierten Pakete einsehen. Diese Liste erhalten wir auch mit dem folgenden Befehl:

```
> library()
```

Auch das Laden von Paketen kann wieder über das Menü oder über einen Befehl erfolgen. Wenn wir bald richtig mit R arbeiten, werden wir immer den Befehl nutzen (s. Kap. 23). Für den Einstieg schauen wir uns aber auch an, wie es über das Menü funktioniert.

**Menübefehl**

Wir öffnen wieder im Fenster unten rechts den Reiter `PACKAGES`. Wir können nun das Paket aus der Liste suchen und mit einem Häkchen versehen – schon wird es geladen. Dieser Weg wird allerdings etwas umständlich, wenn wir sehr viele Pakete installiert haben. Daher verwenden wir ab jetzt lieber die Funktion.

**Funktion**

Pakete laden wir mit der folgenden Funktion:

```
> library(Name.des.Pakets)
```

Warum heißt diese Funktion so? Pakete werden bei der Installation in der sog. Library gespeichert. Wir können uns die Library als Bücherregal vorstellen, in dem die Pakete gelagert werden. Tatsächlich handelt es sich dabei um einen Ordner auf Ihrem Rechner, mit Unterordnern für jedes Paket. Wichtig: Rühren Sie diesen Ordner auf keinen Fall an! Die Installation von Paketen sollte immer über RStudio erfolgen. Wenn wir nun ein Paket verwenden möchten, müssen wir es aus dem Bücherregal herausholen. Daher heißt die Funktion dazu `library`. Zur Veranschaulichung laden wir nun das Paket `psych`:

```
> library(psych)
```

Normalerweise wird das Paket ohne Probleme geladen, was man daran erkennen kann, dass nur der Befehl, jedoch keine zusätzlichen Meldungen in der Konsole erscheinen. Unter Umständen gibt es jedoch zusätzliche Meldungen. Eine häufige Meldung ist die folgende:

```
Warning message:
Paket 'psych' wurde unter R Version 3.6.3 erstellt
```

Eine solche Meldung erscheint, wenn wir eine ältere R-Version verwenden. In diesem Beispiel wurde das Paket in R 3.6.2 geladen. Das Paket selbst wurde jedoch für die Version R 3.6.3 optimiert. In den meisten Fällen stellt dies kein Problem dar, da sich die Pakete von einer Version zur nächsten meistens nur wenig ändern. Um möglichen Kompatibilitätsproblemen vorzubeugen, empfiehlt es sich jedoch, stets die aktuellste R-Version installiert zu haben.

Wir werden im Laufe des Buchs eine Reihe von Paketen kennenlernen. Die Pakete werden jeweils dann installiert und geladen, wenn sie gebraucht werden. Mit dem folgenden Befehl erhalten wir eine Liste aller bereits geladenen Pakete:

```
> search()
```

### Tipp

Manche Pakete funktionieren auf Mac nur, wenn bestimmte zusätzliche Dateien installiert sind. Dabei handelt es sich zum einen um XQuartz, eine Erweiterung des Mac Betriebssystems. XQuartz kann unter <http://xquartz.macosforge.org> heruntergeladen werden. Zum anderen sollte die Tcl/Tk-Erweiterung installiert sein. Die Tcl/Tk-Erweiterung steht unter <https://cran.r-project.org/bin/macosx/> zum Download zur Verfügung. Möglicherweise erhalten Sie bei der Installation die Fehlermeldung, dass diese App von nicht verifizierten Entwicklern stammt. Klicken Sie in diesem Fall auf das Fragezeichen und folgen Sie den Anweisungen in der Hilfe zu dieser Fehlermeldung.

### 3.2.3 Pakete aktualisieren

Viele Pakete werden immer weiter verbessert und erweitert. Es ist daher sinnvoll, die bereits installierten Pakete von Zeit zu Zeit zu aktualisieren. Dazu wählt man in RStudio das Menü **TOOLS** → **CHECK FOR PACKAGE UPDATES**. Man erhält dann eine Übersicht, welche Versionen der Pakete man installiert hat und welche verfügbar sind. Unterscheiden sich die Versionsnummern, kann man das Paket anklicken und anschließend mit **INSTALL UPDATES** die Aktualisierung starten. Mit dem folgenden Befehl werden alle installierten Pakete auf einmal aktualisiert:

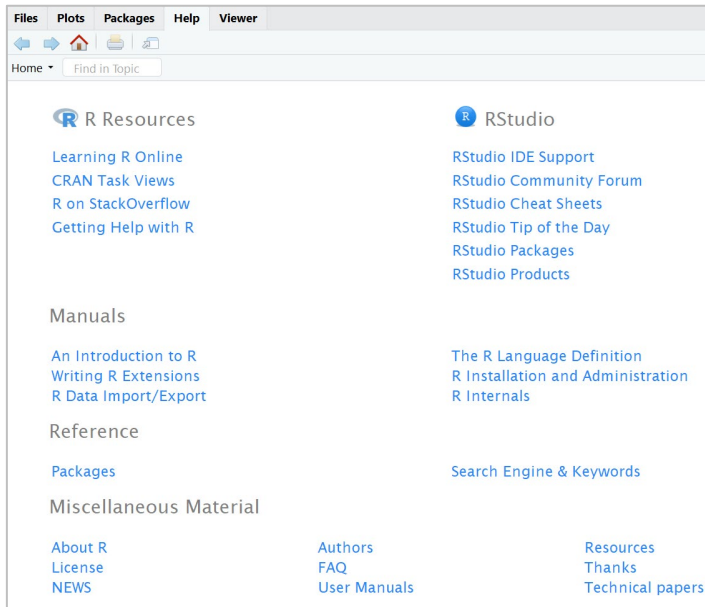
```
> update.packages(ask = FALSE)
```

## 3.3 Hilfe zu R

R entwickelt sich rasant, und mittlerweile gibt es wohl niemanden mehr, der alle Pakete und Funktionen kennt. Ein wichtiges Ziel beim Einstieg in R ist deshalb, dass man lernt, sich neue Pakete und Funktionen selbständig anzueignen. Selbständig heißt aber nicht mutterseelenallein. Es gibt unzählige Stellen, wo man Hilfe finden kann: in RStudio selbst, in einschlägigen Fachbüchern und vor allem im Internet.

### 3.3.1 Hilfe in RStudio

Die erste Anlaufstation für Hilfe in RStudio ist das Menü **HELP** → **R HELP**. Damit öffnet sich im Fenster unten rechts eine Hilfe-Seite mit vielen nützlichen Links zu Ressourcen und Materialien zu R und RStudio (s. Abb. 3.5). Viele dieser Links führen zu Internetseiten und können daher nur verwendet werden, wenn eine Internetverbindung besteht. Die im Abschnitt **MANUALS** verlinkten Handbücher können aber auch offline abgerufen werden.



**Abbildung 3.5** Hilfe-Seite in RStudio, zu erreichen über **HELP** → **R HELP**

### 3.3.2 Hilfe zu bestimmten Paketen und Funktionen

Alle Pakete enthalten Hilfe-Dateien, die direkt mitinstalliert werden. Um nähere Informationen zu einem bestimmten Paket zu bekommen, muss das Paket zunächst geladen werden:

```
> library(NameDesPakets)
```

Mit der `help`-Funktion kann man die Hilfe-Dateien aufrufen. Hilfe zum gesamten Paket erhält man, indem man den Namen des Pakets eingibt:

```
> help(NameDesPakets)
```

Alternativ kann man auch einfach ein Fragezeichen vor den Namen des Pakets setzen:

```
> ?NameDesPakets
```

Hilfe zu bestimmten Funktionen erhält man, indem man den Namen der Funktion in den Befehl eingibt (s. auch Abschn. 4.3.3):

```
> help(NameDerFunktion) oder ?NameDerFunktion
```

Mit der `help.search`-Funktion erhält man eine Liste aller Pakete, die den gesuchten Begriff in einer Funktion enthalten:

```
> help.search("Suchbegriff")  
> ??Suchbegriff
```

### 3.3.3 Hilfe im Internet

Das Internet bietet eine Unmenge an Informationen zu R. Wir gehen hier die wichtigsten Ressourcen durch.

#### Homepages von R und RStudio

Sowohl auf der offiziellen R-Homepage (<http://www.r-project.org>) als auch auf der Support-Webseite von RStudio (<https://support.rstudio.com>) findet man viele nützliche Ressourcen, z. B. Handbücher zu R. Informationen zu den einzelnen Paketen findet man unter <http://cran.r-project.org/web/packages>. Hier ist jedes Paket mit einer eigenen Seite verlinkt, von der man auch das Manual des jeweiligen Pakets als pdf-Dokument herunterladen kann.

#### Einführende Webseiten

Die Webseite Quick-R (<http://www.statmethods.net>) fasst die am häufigsten benötigten Funktionen kurz und prägnant zusammen. Für blutige Anfänger sind die Funktionen manchmal etwas zu knapp kommentiert, aber sobald man die grundlegende Logik von R verstanden hat, kann man hier sehr gut und schnell einzelne Funktionen nachschlagen. Die Homepage des Persönlichkeitspsychologen William Revelle (<http://www.personality-project.org/R/R.guide.html>) gibt eine Einführung in R und geht besonders auf statistische Methoden ein, die in der psychologischen Forschung häufig zum Einsatz kommen, z. B. die Faktorenanalyse. Revelle ist auch der Autor des `psych`-Pakets, das wir noch viel verwenden werden. Eine hilfreiche

deutschsprachige Seite ist R-Lernplattform des Lehrstuhls für Psychologische Methodenlehre an der Humboldt-Universität Berlin: <http://methods-berlin.com/de/r-lernplattform/>. Hier findet man anschauliche Kurzanleitungen, Übungsaufgaben und FAQs zu R.

### Suchmaschinen

Natürlich lassen sich alle Funktionen auch über reguläre Suchmaschinen wie Ecosia oder Google finden. Der Name der Software erschwert die Suche jedoch ungemein. Unter dem Stichwort »R« findet man nicht nur Statistik-Seiten. Daher sollte man lieber eine auf R spezialisierte Suchmaschine verwenden. Gut bewährt hat sich die Suchmaschine R-Seek (<http://www.rseek.org>). Der Vorteil dieser Suchmaschine ist, dass die Ergebnisse nach verschiedenen Kategorien sortiert werden können. Sucht man z. B. eine Funktion für eine bestimmte Art von Analyse, so kann man unter dem Stichwort *Function* nachschauen. Auch Blogs, Forumdiskussionen und Wikis sind hier verlinkt. Die Suchmaschine greift auch auf die Seite <https://www.rdocumentation.org> zu, auf der man Dokumentationen für alle Pakete findet.

### Internetforen und soziale Medien

Bei spezifischen Fragen oder Problemen findet man häufig in Internetforen eine Antwort. Die populärsten Foren für R sind derzeit [stackoverflow](https://stackoverflow.com) (<https://stackoverflow.com>) und das RStudio-Forum (<https://community.rstudio.com>). Beide Foren werden auch von der Suchmaschine R-Seek abgesucht. Bei der Nutzung von Foren gehört es zur Etikette, dass man immer erstmal prüft, ob jemand anderes schon etwas zu dieser Frage gepostet hat. Ist das nicht der Fall, kann man selbst einen Eintrag erstellen. Dabei sollte man die Frage oder das Problem so präzise wie möglich beschreiben. Wenn man z. B. aus einer Fehlermeldung nicht schlau wird, sollte man nicht nur den Text dieser Fehlermeldung posten, sondern den gesamten R-Befehl – am besten in einer Form, dass andere ihn direkt kopieren und ausführen können, auch wenn sie keinen Zugriff auf die Daten haben.

Auch in den sozialen Medien kann man viel über R lernen. Auf Twitter gibt es diverse Accounts, die regelmäßig zu R twittern, z. B. [@rstudiotips](#), [@inside\\_R](#), [@hadleywickham](#) oder [@RLangTip](#). Anlässlich der 5. Auflage dieses Buchs gibt es neuerdings auch den Account [@RfuerEinsteiger](#), auf dem regelmäßig Updates und Tipps zu R gepostet werden. Tweets zu R findet man außerdem unter dem Hashtag [#rstats](#). Auf Facebook gibt es viele öffentliche Gruppen, z. B. [R Users Psychology](https://www.facebook.com/groups/346735129027556/) (<https://www.facebook.com/groups/346735129027556/>).

### Cheatsheets bzw. Reference Cards

Cheatsheets (auch als Reference Cards bekannt) sind übersichtliche Zusammenstellungen der wichtigsten R-Funktionen, die man ausdrucken und sich als eine Art

Spickzettel neben den Computer legen kann. Viele Spickzettel sind direkt von RStudio aus zu finden unter `HELP` → `CHEATSHEETS`. Eine noch größere Sammlung findet man auf der Webseite von RStudio unter <https://rstudio.com/resources/cheatsheets/>. Auch für dieses Buch gibt es eine solche Reference Card; sie kann auf der begleitenden Webseite <http://www.beltz.de/r-fuer-einsteiger> heruntergeladen werden.

### 3.3.4 Bücher

Es gibt mittlerweile sehr viele gute Bücher zu R. Besonders empfehlen möchte ich die folgenden. Das Buch »Moderne Datenanalyse mit R« von Sauer (2019) gibt einen unterhaltsamen Einstieg in R. Der Fokus liegt dabei auf modernen R-Funktionen wie die aus den `tidyverse`-Paketen (s. Kap. 7) sowie auf speziellen Analyseverfahren, die weit über die im typischen Studium gelehrteten Verfahren hinausgehen, z. B. prädiktive Modellierung oder Textmining. Das Buch »R for Data Science« von Wickham und Grolemund (2017) ist eines der am meisten empfohlenen englischsprachigen Bücher zu R. Inhaltlich hat es einen ähnlichen Schwerpunkt wie das Buch von Sauer (2019). Für die Erstellung von Diagrammen gibt das »R Graphics Cookbook« von Chang (2013) einen guten Überblick. Darüber hinaus wird R auch zunehmend in Statistik-Lehrbüchern verwendet (z. B. Field et al., 2012).

## 3.4 Die wichtigsten Funktionen im Überblick

Funktion	Beschreibung
<code>install.packages()</code>	Lädt ein Paket herunter und installiert es.
<code>library()</code>	Lädt ein Paket in R. Dieser Vorgang muss in jeder Sitzung neu durchgeführt werden. Wird kein Paketname in der Funktion angegeben, wird eine Liste mit allen installierten Paketen angezeigt.
<code>search()</code>	Zeigt eine Liste mit allen aktuell geladenen Paketen an.
<code>update.packages()</code>	Aktualisiert alle installierten Pakete.
<code>help()</code> bzw. <code>?</code>	Öffnet die Hilfe-Datei für das angegebene Paket bzw. die angegebene Funktion.
<code>help.search()</code> bzw. <code>??</code>	Nennt alle Funktionen, in denen der eingegebene Suchbegriff verwendet wird.

# 4 Einführung in die Programmiersprache

Das Erlernen von R hat gewisse Parallelen zum Erlernen einer Fremdsprache. Wir müssen uns mit einer fremden Grammatik vertraut machen und neue Vokabeln lernen. Darüber hinaus gibt es bestimmte Stilregeln und Konventionen, die wir uns aneignen müssen. Und es gilt: Nur durch regelmäßiges Üben werden wir besser und vergessen weniger. Einen wichtigen Unterschied gibt es jedoch auch: Anders als bei einer gesprochenen Sprache erwartet niemand von uns, dass wir die Programmiersprache fließend sprechen und alle Vokabeln im Kopf haben.

In diesem Kapitel beschäftigen wir uns mit der grundlegenden Grammatik von R und lernen ein paar erste Vokabeln und Stilregeln kennen. Wir behandeln R als Taschenrechner (Abschn. 4.1), geben logische Abfragen ein (Abschn. 4.2), lernen erste Funktionen kennen (Abschn. 4.3), kommentieren unser bisheriges Werk (Abschn. 4.4) und besprechen einige Konventionen, mit denen wir unsere Skripte übersichtlich gestalten können (Abschn. 4.5).

## 4.1 R als Taschenrechner

In R kann man nicht nur komplexe Datensätze analysieren, sondern auch ganz einfache Berechnungen durchführen. R ersetzt also nicht nur gängige Statistik-Programme, sondern auch den simplen Taschenrechner (hoffentlich aber nicht den Kopf). Tabelle 4.1 enthält eine Übersicht der wichtigsten mathematischen Operatoren und Funktionen in R.

**Tabelle 4.1** Mathematische Operatoren und Funktionen in R

Zeichen	Bedeutung	Zeichen	Bedeutung
+	Addition	-	Subtraktion
*	Multiplikation	/	Division
^	Potenz	sqrt(x)	Quadratwurzel
exp(x)	Exponentialfunktion $e^x$	log(x)	Natürlicher Logarithmus
sum(x)	Summe	abs(x)	Betrag (absoluter Wert)

Im folgenden Beispiel berechnen wir die Summe von 1, 5 und 7. Unter dem eingegebenen Befehl erscheint direkt das Ergebnis, nämlich 13:

```
> 1+5+7
[1] 13
```

## 4.2 Logische Abfragen

In R können wir auch logische Abfragen durchführen. Logische Abfragen benötigt man z. B., wenn man einzelne Analysen nur dann ausführen möchte, wenn bestimmte Bedingungen erfüllt sind. Diese Bedingungen werden dann als logische Abfrage formuliert.

Bei einer logischen Abfrage wird ausgegeben, ob unsere Eingabe richtig (`TRUE`) oder falsch (`FALSE`) ist. In einer logischen Abfrage werden verschiedene logische Zeichen verwendet (s. Tab. 4.2). Hier sind ein paar sehr einfache Beispiele für logische Abfragen:

```
> 3 > 2
[1] TRUE
```

```
> 3 == 2
[1] FALSE
```

**Tabelle 4.2** Zeichen für logische Abfragen in R

Zeichen	Bedeutung	Zeichen	Bedeutung
<code>==</code>	Gleich	<code>!=</code>	Ungleich
<code>&gt;</code>	Größer	<code>&lt;</code>	Kleiner
<code>&gt;=</code>	Größer gleich	<code>&lt;=</code>	Kleiner gleich
<code>&amp;</code>	Logisches Und	<code> </code>	Logisches Oder

Es ist auch möglich, mehrere logische Abfragen miteinander zu verknüpfen. Dabei muss festgelegt werden, ob alle Bedingungen erfüllt sein müssen, um ein `TRUE` zu erhalten (»und«-Verknüpfung, Symbol `&`) oder ob es dafür reicht, wenn eine einzige Bedingung erfüllt ist (»oder«-Verknüpfung). Im folgenden Beispiel werden zwei Bedingungen mit dem Symbol `&` verknüpft:

```
> 3 < 2 & 2 < 3
[1] FALSE
```

Da nur eine, jedoch nicht beide Bedingungen zutreffen, ist das Ergebnis `FALSE`. Wenn wir jedoch stattdessen eine »oder«-Verknüpfung verwenden, erhalten wir das Ergebnis `TRUE`, denn eine der beiden Bedingungen ist ja in der Tat richtig. Hier werden die beiden Bedingungen mit dem Symbol `|` verknüpft:

```
> 3 < 2 | 2 < 3
[1] TRUE
```

Die bisherigen logischen Abfragen waren Beispiele mit Zahlen. Man kann aber auch Texte eingeben. Dabei werden Anführungszeichen verwendet, um die Texteingabe von Objektnamen (s. Kap. 5) zu unterscheiden:

```
> "abcd" == "abcd"
[1] TRUE
```

Nun kommt leider eine schlechte Nachricht: R weiß nicht alles. Die folgende Abfrage führt bspw. zu einem eher unromantischen Ergebnis:

```
> "typ aus statistikkurs" == "grosse liebe"
[1] FALSE
```

## 4.3 Funktionen

Funktionen haben den folgenden Aufbau:

```
> funktion(argument1 = wert1, argument2 = wert2, ...)
```

Eine Funktion beginnt immer mit einem Namen (hier allgemein: `funktion`), gefolgt von Klammern. In den Argumenten innerhalb der Klammer werden die Details angegeben, z. B. auf welche Daten die Funktion angewandt werden soll. Ein Argument ist folgendermaßen aufgebaut:

- ▶ Der Name des Arguments (hier verallgemeinernd `argument1` und `argument2`) steht vor dem einfachen Gleichheitszeichen.
- ▶ Die Einstellung bzw. der Wert des Arguments (hier verallgemeinernd `wert1` und `wert2`) steht hinter dem einfachen Gleichheitszeichen.

Vereinfacht kann man sich Argumente als verschiedene Schalter vorstellen, mit denen wir die Funktion steuern. Für jeden Schalter gibt es eine Standardeinstellung, die genommen wird, wenn das Argument nicht in der Funktion aufgeführt wird. Um die