

LERNEN EINFACH GEMACHT



3. Auflage

R

für
dummies[®]



Daten analysieren und
verarbeiten

Hochwertige Diagramme und
Grafiken erstellen

Mithilfe von Modellen
Vorhersagen machen

Andrie de Vries
Joris Meys

R für Dummies

Schummelseiten

R ist mehr als nur eine Statistiksprache. R ist ein mächtiges Werkzeug, mit dem sich Daten auf vielerlei Weise verarbeiten und manipulieren lassen. R wird von einer großen Gemeinschaft von Programmieren und Anwendern in der Wissenschaft und der Praxis verwendet. Um jedoch alles aus R rauszuholen, müssen Sie manchmal auf die R-Hilfe oder andere Quellen zurückgreifen. Um Daten mit R schön darzustellen, müssen Sie sie manchmal kurz und bündig rausfiltern. Und schließlich müssen Sie natürlich erst mal diese Daten einlesen – mit einer der vielen Funktionen, die R zu bieten hat.

HILFE ERHALTEN

Auch mit dem allerbesten Einführungsbuch zu R werden Sie ab und an bei der Lösung eines Problems stecken bleiben. R verfügt glücklicherweise über exzellente Hilfeseiten, die Ihnen alle Funktionen und deren Verwendung detailliert erklären. In den meisten Fällen werden die Funktionen an anschaulichen Beispielen demonstriert.

- ✓ `?:` zeigt die Hilfe für eine konkrete Funktion an. Beispielsweise gibt `?data.frame` Hilfe zur Funktion `data.frame()`.
- ✓ `??:` sucht allgemein nach einem Begriff in den Hilfeseiten. `??list` zeigt Ihnen beispielsweise alle

Funktionen an, die *list* entweder im Funktionsnamen oder in ihrer Beschreibung enthalten.

- ✓ `RSiteSearch()`: Diese Funktion nimmt Ihren Suchbegriff und gibt ihn direkt an die Suchmaschine auf RSiteSearch (<http://search.r-project.org/nmz.html>). Die Treffer können Sie sich dann in Ihrem Webbrowser ansehen.

`RSiteSearch("linear models")` durchsucht die Website zum Beispiel nach dem Ausdruck *linear models*.

Natürlich sind Sie nicht auf die R-Hilfe beschränkt, wenn Sie nach der Lösung eines Problems suchen. Das Paket `sos`, das Sie vom CRAN (<http://cran.r-project.org/web/packages/sos/index.html>) herunterladen und installieren können, bietet beispielsweise nette Zusatzfunktionen für `RSiteSearch` und ist auch browserbasiert. Um das Paket `sos` zu verwenden, installieren Sie es mit dem Befehl `install.packages("sos")` über Ihre Konsole und laden es danach mit `library("sos")`.

Sie können anschließend die `findFn()`-Funktion für Ihre Suche nutzen. Wenn Sie zum Beispiel `findFn("regression")` über Ihre R-Konsole ausführen lassen, führt Sie das zu einer Website mit den Namen, Beschreibungen und weiterführenden Links zu vielen Hundert Funktionen, die das Wort *regression* in ihrem Funktionsnamen oder ihrer Beschreibung enthalten.

DATEN IMPORTIEREN

Sie haben viele Möglichkeiten, Daten in R zu importieren. In der folgenden Tabelle finden Sie einige davon:

Funktion	Wofür	Beispiel
<code>read.table()</code>	importiert sämtliche tabellarischen Daten, unabhängig davon, welche Trennzeichen verwendet werden	<code>read.table(file="ihredatei", sep="t", header=TRUE)</code>
<code>read.csv()</code>	eine vereinfachte Version von <code>read.table</code> , die darauf ausgelegt ist, CSV-Dateien wie beispielsweise Excel-Dateien einzulesen	<code>read.csv(file="ihredatei")</code>
<code>read.csv2()</code>	eine vereinfachte Version von <code>read.csv</code> , die darauf ausgelegt ist, CSV-Dateien mit Kommata als Dezimalpunkt und Semikola als Trennzeichen einzulesen	<code>read.csv2(file="ihredatei", header=TRUE)</code>
<code>read.delim()</code>	nützlich für Daten mit Tabulatorvorschüben als Trennzeichen	<code>read.delim(file="ihredatei", header=TRUE)</code>
<code>scan()</code>	gewährt mehr Kontrolle über den Einlesevorgang für nicht tabulare Daten	<code>scan("ihredatei", skip = 1, nmax=100)</code>
<code>readLines()</code>	liest Text zeilenweise aus einer Datei	<code>readLines("ihredatei")</code>
<code>read.fwf()</code>	liest Daten aus einer Datei mit festgelegter Spaltenbreite. Anders ausgedrückt: Jede Spalte hat eine feste Anzahl an Zeichen.	<code>read.fwf("ihredatei", widths=c(1,2,3))</code>

Tabelle 1: Datenimport leicht gemacht

Zusätzlich haben Sie über das Paket `foreign` die Möglichkeit, andere Datenformate einzulesen, zum Beispiel aus SPSS. Um diese Funktion zu nutzen, müssen Sie zunächst das bereits vorinstallierte `foreign`-Paket laden:

```
> library("foreign")
```

Die folgende Tabelle führt einige Funktionen auf, mit denen Sie Daten aus SPSS, Stata oder SAS importieren können:

Funktion Wofür		Beispiel
read.spss	liest eine SPSS-Datei	read.spss("")
read.dta	liest eine Stata-Datei	read.dta("")
read.xport	liest eine SAS-Datei	read.xport("")

Tabelle 2: Noch mehr Funktionen zum Datenimport



Andrie de Vries

R

**für
dummies®**

3. Auflage

Übersetzung aus dem Amerikanischen von
Rainer G. Haselier, Robert Leidenfrost und
Prof. Dr. Karl-Kuno Kunze

WILEY
WILEY-VCH GmbH

R für Dummies

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

3. Auflage 2021

© 2021 Wiley-VCH GmbH, Weinheim

Original English language edition R for Dummies © 2015 by Wiley Publishing, Inc.

All rights reserved including the right of reproduction in whole or in part in any form. This translation published by arrangement with John Wiley and Sons, Inc.

Copyright der englischsprachigen Originalausgabe R for Dummies © 2015 by Wiley Publishing, Inc

Alle Rechte vorbehalten inklusive des Rechtes auf Reproduktion im Ganzen oder in Teilen und in jeglicher Form. Diese Übersetzung wird mit Genehmigung von John Wiley and Sons, Inc. publiziert.

Wiley, the Wiley logo, Für Dummies, the Dummies Man logo, and related trademarks and trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries. Used by permission.

Wiley, die Bezeichnung »Für Dummies«, das Dummies-Mann-Logo und darauf bezogene Gestaltungen sind Marken oder eingetragene Marken von John Wiley & Sons, Inc., USA, Deutschland und in anderen Ländern.

Das vorliegende Werk wurde sorgfältig erarbeitet. Dennoch übernehmen Autoren und Verlag für die

Richtigkeit von Angaben, Hinweisen und Ratschlägen
sowie eventuelle Druckfehler keine Haftung.

Coverfoto: fotomek - stock.adobe.com

Korrektur: Claudia Lötschert

Print ISBN: 9783527719068

ePub ISBN: 978-3-527-83609-3

Über den Autoren

Andrie de Vries verwendete R im Jahr 2009 erstmalig, um Umfragedaten zu analysieren, und ist seitdem regelmäßig überrascht über die Fähigkeit einer Open-Source-Gemeinde, hervorragende und innovative Software herzustellen. Ebenfalls 2009 gründete er PentaLibra Limited, ein kleines, aber feines Marktforschungsunternehmen, das sich auf Umfragen und statistische Analysen spezialisiert hat. Nachdem er sich immer stärker innerhalb der R-Community engagierte, wurde er Mitarbeiter bei Revolution Analytics, um R bei Großkunden zu etablieren und ihnen und anderen bei den Herausforderungen von Data Science und Big Data zu helfen. Zur Erhaltung seines seelischen Gleichgewichts lernt und praktiziert Andrie de Vries Yoga.

Joris Meys, MSc, ist Berater in Sachen Statistik sowie R-Programmierer und Dozent an der Universität Gent (Belgien). Nachdem er seinen Master in Biologie erwarb, arbeitete er zunächst sechs Jahre in Umweltforschung und -management, bis er fortgeschrittene Studien für einen Masterabschluss in statistischer Datenanalyse begann. Joris Meys schreibt Pakete sowohl für spezielle Projekte als auch generelle Methoden, die in seiner Abteilung entwickelt wurden. Darüber hinaus unterhält er mehrere Pakete bei R-Forge. Er ist Koautor mehrerer wissenschaftlicher Arbeiten, bei denen er den statistischen Part übernahm. Um Wissenschaft und Kultur in Einklang zu bringen, spielt er im überwiegenden Teil seiner Freizeit Saxofon in einigen lokalen Bands.

Widmung

Dieses Buch ist für meine Frau Annemarie, für ihre Ermutigung, Unterstützung und Geduld. Es ist auch für meine Nichte Tanya, die wirklich gut in Mathe ist und nicht aufhörte, mich an den Abgabetermin des Manuskripts zu erinnern. Schlussendlich ist es für meine Eltern, für ihren lebenslangen Zuspruch.

– Andrie de Vries

Ich widme dieses Buch den wichtigsten Frauen in meinem Leben. Für meine Mutter, denn sie machte den Menschen aus mir, der ich jetzt bin. Für Eva, denn sie liebt den Mann, der ich bin. Für Amelie, weil ihr kleines Lächeln jedes Mal mein Herz zum Schmelzen bringt. Und für meine Großmutter – sie ist einfach wirklich super!

– Joris Meys

Danksagungen

Dieses Buch wurde nur möglich durch die enorme Unterstützung, die wir vom Wiley-Verlag erfahren haben. Insbesondere danken wir Elizabeth Kuball (erste Auflage) und Katie Mohr (zweite Auflage).

Danke an unseren technischen Lektor, Gavin Simpson, für sein aufmerksames Lesen und viele hilfreiche Kommentare.

Wir danken Patrick Burns, dem Autor von R Inferno, für seine enthusiastische Unterstützung und die sehr detaillierte Rezension der ersten Auflage.

Danke an das *R Core Team* für die Entwicklung von R, für die Pflege des CRAN und die Hingabe an die R-Gemeinde in Form von Mailinglisten, Dokumentationen und Seminaren. Danke auch an die R-Gemeinde für Tausende hilfreicher Pakete, Blogs und beantwortete Fragen.

In diesem Buch verwenden wir einige Pakete von Hadley Wickham, dessen Paket `ggplot2` und andere, wie zum Beispiel `plyr`, eine nicht versiegende Quelle der Inspiration für uns sind.

Beim Schreiben dieses Buchs profitierten wir von sehr hilfreicher Unterstützung vieler Teilnehmer der R-Gruppe von Stack Overflow. Vielen Dank an James (JD) Long, David Winsemius, Ben Bolker, Joshua Ulrich, Barry Rowlingson, Roman Luštrik, Joran Elias, Dirk Eddelbuettel, Richie Cotton, Colin Gillespie, Simon Urbanek, Gabor Grotendieck und die vielen weiteren, die Stack Overflow zu einer großen Informationsquelle für die ganze R-Gemeinde machen.

Von Andrie: Es ist keine Übertreibung, zu behaupten, dass dieses Buch zumindest teilweise dafür verantwortlich ist, dass sich mein Leben verändert hat. R zu erlernen, einen Beitrag zu seiner Open-Source-Community zu leisten und dieses Buch zu schreiben, hat alles dazu beigetragen, dass ich heute bei Revolution Analytics angestellt bin. Ich möchte all meinen Kollegen danken, insbesondere Derek McCrae Norton, David Smith und Joseph Rickert.

Joris Meys: Danke an die Professoren und Kollegen an der Fakultät für Mathematische Modellierung, Statistik und Bioinformatik der Universität Gent für erhellende Diskussionen und die Unterstützung, die ich während des Schreibens dieses Buchs erhalten habe.

Über die Übersetzer/Bearbeiter

Karl-Kuno Kunze (Übersetzer der ersten Auflage), geboren 1970, ist Professor für Wirtschaftsmathematik und -statistik an der Ostfalia Hochschule Braunschweig/Wolfenbüttel und arbeitet mit R seit 2004. Er promovierte zunächst in theoretischer Physik, dann in Wirtschaftswissenschaft. **Robert Leidenfrost** (Übersetzer der zweiten Auflagen), geboren 1986, ist Molekularmediziner und Bioinformatiker. Er arbeitet, forscht und lehrt an der Hochschule Mittweida, University of Applied Sciences. Die Sprache R begleitet ihn schon seit 2011. Die bei seinen wissenschaftlichen Arbeiten gewonnenen Daten prozessiert und visualisiert er am liebsten mit R. **Rainer G. Haselier** (Bearbeiter der 3. Auflage), geboren 1958, ist Diplom-Sozialpädagoge sowie Groß- und Außenhandelskaufmann und hat in den vergangenen 30 Jahren mehrere Hundert Bücher zu den Themenbereichen Programmierung, Betriebssysteme sowie Microsoft Office sowohl verfasst als auch übersetzt.

Inhaltsverzeichnis

Cover

Titelblatt

Impressum

Über den Autoren

Widmung

Danksagungen

Über die Übersetzer/Bearbeiter

Einleitung

Über dieses Buch

Änderungen der zweiten Auflage

Änderungen der dritten Auflage

Konventionen in diesem Buch

Was Sie nicht lesen müssen

Törichte Annahmen über den Leser

Wie dieses Buch aufgebaut ist

Symbole, die in diesem Buch verwendet werden

Wie es weitergeht

Teil I: Sind Sie beReit?

Kapitel 1: R im Überblick

Die Vorteile der Anwendung von R erkennen

Einige bemerkenswerte Eigenschaften von R

Kapitel 2: R erkunden

Mit einem Code-Editor arbeiten

Ihre erste R-Sitzung starten

Ein Skript einlesen

Ihr Programm mit ausgeben

Sich im Arbeitsbereich zurechtfinden

Kapitel 3: Die Grundlagen von R

[Die ganze Power von Funktionen ausschöpfen](#)

[Lesbaren Code verfassen](#)

[Von der R-Basis abheben](#)

Teil II: Arbeiten mit R

Kapitel 4: Erste Schritte mit Arithmetik

[Mit Zahlen, Unendlichkeit und fehlenden Werten arbeiten](#)

[Daten in Vektoren organisieren](#)

[Werte in Vektoren hinein- und aus ihnen herausbekommen](#)

[Mit logischen Vektoren arbeiten](#)

[Turbomathematik mit Vektorfunktionen](#)

Kapitel 5: Erste Schritte im Lesen und Schreiben

[Zeichenvektoren für Text verwenden](#)

[Text bearbeiten](#)

[Mit Faktoren arbeiten](#)

Kapitel 6: Ihr erstes Date mit R

[Mit Datumsangaben arbeiten](#)

[Datumsangaben in verschiedenen Formaten ausgeben](#)

[Datumsangaben um die Uhrzeit erweitern](#)

[Datumsangaben und Uhrzeiten formatieren](#)

[Verschiedene Operationen mit Datumsangaben und Uhrzeiten durchführen](#)

Kapitel 7: Arbeiten in höheren Dimensionen

[Eine zweite Dimension hinzufügen](#)

[Indizes verwenden](#)

[Matrixzeilen und -spalten benennen](#)

[Mit Matrizen rechnen](#)

[Weitere Dimensionen hinzufügen](#)

[Verschiedene Datentypen in Datensätzen vereinen](#)

[Werte in Datensätzen verändern](#)

[Verschiedene Objekte in Listen vereinen](#)

Teil III: Programmieren in R

Kapitel 8: Mehr Fun mit Funktionen

[Von Skripten zu Funktionen gelangen](#)

[Argumente geschickt einsetzen](#)

[Geltungsbereiche verstehen](#)

[Aufgaben an Methoden delegieren](#)

Kapitel 9: Die Ablauflogik kontrollieren

[Mit »if« Verzweigungen einbauen](#)

[Mit »if ... else« eine Alternative vorgeben](#)

[Verzweigungen vektorisieren](#)

[Mehrere Alternativen vorgeben](#)

[Schleifen mit unterschiedlichen Werten durchlaufen](#)

[Schleifen ohne Schleifen bauen: Die »apply«-Familie](#)

Kapitel 10: Fehlersuche

[Wissen, wonach man suchen soll](#)

[Fehler- und Warnmeldungen entschlüsseln](#)

[Auf Fehlerjagd gehen](#)

[Ihre eigenen Meldungen verfassen](#)

[Fehler erkennen, die Sie sicher machen werden](#)

Kapitel 11: Hilfe erhalten

[Informationen in den Hilfeseiten finden](#)

[Das Internet nach Hilfe zu R durchsuchen](#)

[Der R-Onlinegemeinde beitreten](#)

[Ein reproduzierbares Minimalbeispiel erstellen](#)

Teil IV: Daten zum Reden bringen

Kapitel 12: Daten lesen und schreiben

[Daten in R einlesen](#)

[Daten aus R herausbekommen](#)

[Mit Dateien und Verzeichnissen arbeiten](#)

Kapitel 13: Mit Daten arbeiten

[Die passende Datenstruktur finden](#)

[Teilmengen von Daten bilden](#)
[Berechnete Felder hinzufügen](#)
[Daten verbinden und zusammenführen](#)
[Daten sortieren und ordnen](#)
[Daten mit den »apply«-Funktionen durchlaufen](#)
[Die Formelschnittstelle kennenlernen](#)
[Daten in Form bringen](#)

Kapitel 14: Daten verdichten

[Mit den richtigen Daten beginnen](#)
[Kontinuierliche Variablen beschreiben](#)
[Kategoriale Daten beschreiben](#)
[Verteilungen beschreiben](#)
[Mehrere Variablen beschreiben](#)
[Mit Tabellen arbeiten](#)

Kapitel 15: Differenzen und Relationen untersuchen

[Verteilungen genauer untersuchen](#)
[Zwei Stichproben vergleichen](#)
[Häufigkeiten und Verhältnisse testen](#)
[Mit Modellen arbeiten](#)

Teil V: Mit Grafiken arbeiten

Kapitel 16: Mit den Basisfunktionen für Grafik arbeiten

[Unterschiedliche Arten von Diagrammen erzeugen](#)
[Optionen und Argumente von »plot\(\)« nutzen](#)
[Grafiken in Bilddateien speichern](#)

Kapitel 17: Rastergrafiken mit »lattice«

[Eine Rastergrafik mit »lattice« erzeugen](#)
[Grafikoptionen verändern](#)
[Verschiedene Diagrammtypen erzeugen](#)
[Daten in Gruppen darstellen](#)
[Eine »lattice«-Grafik ausgeben und speichern](#)

Kapitel 18: Grammatik für Grafik: »ggplot2«

[Das Paket »ggplot2« installieren und laden](#)

[Ebenen verstehen](#)

[»geom«-und »stat«-Argumente verwenden](#)

[Raster, Skalen und Optionen](#)

[Mehr Informationen erhalten](#)

Teil VI: Der Top-Ten-Teil

Kapitel 19: Zehnmal R statt Excel

[Zeilen- und Spaltensummen hinzufügen](#)

[Zahlen formatieren](#)

[Daten sortieren](#)

[Mit »if« Bedingungen einbauen](#)

[Teilsummen berechnen](#)

[Spalten oder Zeilen transponieren](#)

[Eindeutige Werte und Duplikate finden](#)

[Mit Wertetabellen arbeiten](#)

[Mit Pivot-Tabellen arbeiten](#)

[Zielwertsuche einsetzen](#)

Kapitel 20: Zehn Tipps zum Arbeiten mit Packages

[In den Ecken und Ritzen des CRAN herumschnüffeln](#)

[Interessante Pakete finden](#)

[Pakete installieren](#)

[Pakete laden](#)

[Das Handbuch und die Vignette zum Paket lesen](#)

[Pakete aktualisieren](#)

[R-Forge erforschen](#)

[R-Pakete von GitHub laden](#)

[Pakete aus dem Bioladen: BioConductor nutzen](#)

[Das R-Handbuch lesen](#)

Anhang A: R und RStudio installieren

[R installieren und konfigurieren](#)

[RStudio installieren und konfigurieren](#)

Anhang B: Das »rfordummies«-Paket

[Verwenden Sie »rfordummies«](#)

Stichwortverzeichnis

Advertisement Page

End User License Agreement

Tabellenverzeichnis

Schummelseiten

[Tabelle 1: Datenimport leicht gemacht](#)

[Tabelle 2: Noch mehr Funktionen zum Datenimport](#)

Kapitel 4

[Tabelle 4.1: Arithmetische Basisoperationen](#)

[Tabelle 4.2: Nützliche mathematische Funktionen in R](#)

[Tabelle 4.3: Ergebnisse für unendliche, NaN- und fehlende Werte](#)

[Tabelle 4.4: Werte vergleichen in R](#)

[Tabelle 4.5: Vektoroperationen](#)

Kapitel 6

[Tabelle 6.1: Gebräuchliche Funktionen für Datumsangaben](#)

[Tabelle 6.2: Einige Formatkürzel für die Datumsumwandlung
\(geeignet für `as.Date`,](#)

[Tabelle 6.3: Formatkürzel für den Zeitanteil von `POSIXct`- und
`POSIXlt`-Objekten](#)

Kapitel 9

[Tabelle 9.1: `apply`, `sapply` und `lapply` verwenden](#)

Kapitel 12

[Tabelle 12.1: Funktionen zum Import von kommerziellen
Statistiksystemen im Paket](#)

[Tabelle 12.2: Nützliche Funktionen für Dateien](#)

Kapitel 13

[Tabelle 13.1: Hilfreiche Objekte für die Datenanalyse](#)

[Tabelle 13.2: Mögliche Teilmengen festlegen](#)

[Tabelle 13.3: Einige Formeloperatoren und ihre Bedeutung](#)

Kapitel 15

[Tabelle 15.1: Auf Modellobjekte zugreifen](#)

Kapitel 18

[Tabelle 18.1: Eine Auswahl von geom-Elementen und zugehörigen stat-Elementen](#)

[Tabelle 18.2: Einige nützliche stat-Elemente und zugehörige geom-Elemente](#)

Illustrationsverzeichnis

Kapitel 2

[Abbildung 2.1: Die Verknüpfungen zu RGui, 32- und 64-Bit, Version 4.0.4](#)

[Abbildung 2.2: Eine frisch geöffnete R-Sitzung in der RGui](#)

[Abbildung 2.3: R bittet Sie um Auskunft.](#)

[Abbildung 2.4: RStudio hat vier Arbeitsfenster.](#)

[Abbildung 2.5: Ein Skript vom Editor an die Konsole schicken](#)

[Abbildung 2.6: Die Verarbeitung Ihres Skripts mit und ohne ECHO](#)

Kapitel 5

[Abbildung 5.1: Jedes Element eines Zeichenvektors ist ein Stück Text, häufig als ...](#)

[Abbildung 5.2: Numerische Vektoren und Faktoren im Vergleich](#)

Kapitel 7

[Abbildung 7.1: Vektor, Matrix und Datenfeld](#)

Kapitel 8

[Abbildung 8.1: Wie R die Umgebungen sieht.](#)

Kapitel 10

[Abbildung 10.1: Mit RStudio zum Debuggen durch eine Funktion wandern](#)

Kapitel 12

[Abbildung 12.1: Das Arbeitsblatt mit Elementen des Periodensystems dient uns in d...](#)

[Abbildung 12.2: Daten bearbeiten im interaktiven Texteditor](#)

[Abbildung 12.3: Ein Excel-Arbeitsblatt nach der Verwendung von writeClipboard\(.\) U...](#)

[Abbildung 12.4: Die ersten sechs Zeilen von iris nach dem Einfügen in eine Excel-...](#)

Kapitel 13

[Abbildung 13.1: Verschiedene Möglichkeiten, Daten zu verbinden](#)

[Abbildung 13.2: Verschiedene Arten, merge\(.\) zu verwenden, und deren Entsprechung ...](#)

[Abbildung 13.3: Verwenden Sie apply bei Matrizen; verwenden Sie lapply oder sapply...](#)

[Abbildung 13.4: Daten im schmalen \(eingeschmolzenen\) Format sind für die Arbeit m...](#)

Kapitel 14

[Abbildung 14.1: Ein Histogramm für Ihre Daten erzeugen](#)

[Abbildung 14.2: Die Dichtefunktion darstellen und mit einem Histogramm überlagern](#)

[Abbildung 14.3: Verwenden Sie die Funktion boxplot\(\), um diese Grafik zu erhalten...](#)

[Abbildung 14.4: Die Relationen zwischen Variablen eines Datensatzes](#)

Kapitel 15

[Abbildung 15.1: Histogramme für verschiedene Gruppen erzeugen](#)

[Abbildung 15.2: Ein Q-Q-Diagramm mit zwei unterschiedlichen Ensembles erzeugen](#)

[Abbildung 15.3: Merkmale mithilfe von Q-Q-Diagrammen mit der Normalverteilung ver...](#)

[Abbildung 15.4: Die Ergebnisse des Tukey'schen HSD-Tests als Diagramm](#)

[Abbildung 15.5: Eine Trendlinie durch die Daten legen](#)

Kapitel 16

[Abbildung 16.1: Ein Diagramm mit \(Achsen-\)Beschriftungen, \(Haupt-\)Überschrift und...](#)

[Abbildung 16.2: Ein Streudiagramm erstellen](#)

[Abbildung 16.3: **Abbildung 16.3** Verschiedenfarbige Punkte zum Diagramm hinzufügen](#)

[Abbildung 16.4: Einem Diagramm Linien hinzufügen](#)

[Abbildung 16.5: Den Diagrammtyp festlegen](#)

[Abbildung 16.6: Hinzufügen der \(Haupt-\)Überschrift sowie der Beschriftungen von x...](#)

[Abbildung 16.7: Die Ausrichtung der Beschriftung ändern](#)

[Abbildung 16.8: Die Rahmeneinstellungen ändern](#)

[Abbildung 16.9: Ausrichtung der Beschriftung, Rahmen, Farbe und Symbol ändern](#)

[Abbildung 16.10: Ändern der Schriftgröße der »Etiketten« sowie der Überschrift un...](#)

[Abbildung 16.11: Zwei Diagramme nebeneinander](#)

Kapitel 17

[Abbildung 17.1: Rastergrafiken wie diese ermöglichen eine synoptische Ansicht ver...](#)

[Abbildung 17.2: Ein lattice-Streudiagramm der Daten in `mtcars`](#)

[Abbildung 17.3: `xyplot\(\)` aus `lattice` mit Regressionslinien](#)

[Abbildung 17.4: Ein `lattice`-Diagramm mit Beschriftung](#)

[Abbildung 17.5: Die Schriftgröße in `lattice`-Grafiken ändern](#)

[Abbildung 17.6: Themen verwenden, um die Farbe von Datenpunkten und Linien zu änd...](#)

[Abbildung 17.7: Ein Balkendiagramm mit `barchart\(\)` aus dem Paket `lattice`](#)

[Abbildung 17.8: Einen Boxplot in `lattice` erstellen](#)

[Abbildung 17.9: Daten im schmalen Format verwenden, um unterschiedliche Variablen...](#)

[Abbildung 17.10: Ein `lattice`-Balkendiagramm \(`bar chart`\) mit Gruppen und Legende](#)

Kapitel 18

[Abbildung 18.1: Ein `ggplot2`-Diagramm des `faithful`-Datensatzes](#)

[Abbildung 18.2: Ein mit `geom:histogram\(\)` beziehungsweise `geom:bar\(\)` erstelltes Hi...](#)

[Abbildung 18.3: Verdichtete Daten in einem Balkendiagramm](#)

[Abbildung 18.4: Ein Streudiagramm mit `geom:point\(\)` erstellen](#)

[Abbildung 18.5: Ein Liniendiagramm mit `geom:line\(\)` erzeugen](#)

[Abbildung 18.6: Regressionslinien hinzufügen mit `stat_smooth\(\)`](#)

[Abbildung 18.7: Raster hinzufügen mit `facet_grid\(\)`](#)

[Abbildung 18.8: Die Einstellungen von `ggplot2` verändern](#)

Kapitel 19

[Abbildung 19.1: Ein Modell erwarteter Mengen und Umsätze abhängig vom Preis](#)

Anhang A

[Abbildung A.1: Das Optionenfenster von RStudio. Je nach Version und Installation ...](#)

Einleitung

Willkommen bei *R für Dummies*, dem Buch, das die Lernkurve für die Statistik- und Programmiersprache R zu einem aufregenden Erlebnis ähnlich der Nordkurve macht.

Wir garantieren keineswegs, dass Sie nach der Lektüre dieses Buchs ein R-Guru sind, die folgenden Dinge werden Sie allerdings auf jeden Fall beherrschen:

- ✓ Sie führen Datenanalysen mit verschiedenen leistungsstarken Werkzeugen durch.
- ✓ Sie benutzen R für statistische Analysen sowie andere Daten verarbeitende Aufgaben.
- ✓ Sie lernen die Schönheit vektorbasierter Operationen im Vergleich zur Verwendung von Schleifen schätzen.
- ✓ Sie verstehen die Bedeutung folgender Codezeile:
`knowledge <- apply(theory, 1, sum)`
- ✓ Sie wissen, wie Sie Code von Mitgliedern der R-Gemeinde und anderen Entwicklern finden, herunterladen und einsetzen können.
- ✓ Sie wissen, wo Sie zusätzliche Hilfe und Ressourcen erhalten, um Ihre Fähigkeiten noch weiter auszubauen.
- ✓ Sie erzeugen wundervolle Grafiken und Visualisierungen Ihrer Daten und Ergebnisse.

Über dieses Buch

R für Dummies ist eine Einführung in die statistische Programmiersprache R. Wir beginnen zunächst mit der Benutzeroberfläche und arbeiten uns dann von ziemlich einfachen Konzepten der Sprache vor bis zu schon recht

anspruchsvollen Themen der Datenverarbeitung und -analyse.

Jeder Schritt wird von einfach durchzuführenden Beispielen begleitet. Dieses Buch enthält zahlreiche Codeausschnitte, einige Baukastenfunktionen, die Sie später weiterverwenden können, sowie komplette Analyseskripte. All das hat im Wesentlichen einen Zweck: Sie sollen es selbst ausprobieren.

Wir versuchen erst gar nicht, eine technische Beschreibung zu unternehmen, wie R selbst programmiert wurde. Davon abgesehen sollen sich unsere Ausführungen, wie etwas funktioniert, in etwa die Waage halten mit den Gründen, warum das so funktioniert. R ist keine durchschnittliche Skriptsprache und hat einige Eigenschaften, die auf den ersten Blick überraschen mögen. Anstatt Ihnen einfach nur zu erzählen, wie Sie mit R sprechen sollen, glauben wir schon, dass es wichtig ist, Ihnen zu verraten, wie R Ihre Eingaben liest und interpretiert. Nachdem Sie dieses Buch gelesen haben, sollten Sie in der Lage sein, Daten in der von Ihnen gewünschten Form zu verarbeiten und Funktionen zu verwenden, die wir in diesem Buch nicht vorgestellt haben (die wir vorgestellt haben, möglichst auch).

Dieses Buch ist als Referenz gedacht. Sie müssen es nicht von Anfang bis Ende lesen. Stattdessen können Sie einfach das Inhaltsverzeichnis und das Stichwortverzeichnis nutzen, um die Informationen, die Sie brauchen, zu finden. In jedem Kapitel verweisen wir auf andere Kapitel, in denen Sie weitere Informationen finden.

Änderungen der zweiten Auflage

Seit der Publikation der ersten Auflage hat sich R kontinuierlich weiterentwickelt und verbessert. Um die Korrektheit des Buchs zu gewährleisten, haben wir den Code entsprechend der letzten Version von R (Version 3.4.1) angepasst. Basierend auf dem Feedback von Lesern, Studenten und Kollegen konnten wir einige Abschnitte überarbeiten und so Fragen klären und Ungenauigkeiten beheben. Beispielsweise haben wir den Code dahin gehend geändert, dass wir nun hochgestellte Gänsefüße statt Hochkommata innerhalb von Text-Strings verwenden. Auch bezeichnen wir jetzt die Basiseinheiten von Listen als Komponenten statt Elementen.

Änderungen der dritten Auflage

Die dritte Auflage des Buchs basiert auf der R-Version 4.0.4, die im Februar 2021 veröffentlicht wurde. Falls sich geänderte und neue Features auf die Themen und Beispiele auswirken, die im Buch vorgestellt werden, dann wurden Text und Code entsprechend angepasst.

Das aktuelle `rfordummies`-Paket enthält die Codebeispiele aus dem Buch. Sie können alles Weitere über das Paket in Anhang B erfahren.

R und RStudio

R für Dummies kann mit jedem Betriebssystem verwendet werden, auf dem R läuft. Ob Sie macOS, Linux oder Windows nutzen, mit diesem Buch bringen

wir Sie auf den Weg mit R.

R ist mehr eine Programmiersprache als ein Anwendungsprogramm. Wenn Sie R herunterladen, laden Sie auch eine Konsolenanwendung herunter, die zu Ihrem Betriebssystem passt. Erwarten Sie jedoch keine Wunder: Die Funktionalität ist sehr eingeschränkt und variiert zwischen den einzelnen Betriebssystemen.

RStudio ist eine Plattform-übergreifende Anwendung, auch benannt als Integrated Development Environment (IDE), mit einigen sehr netten Eigenschaften in Bezug auf R. Dieses Buch setzt zwar keine spezielle Konsolenanwendung voraus. Da RStudio jedoch eine über alle Betriebssysteme einheitliche Anwendung ist, glauben wir, dass Sie es schnell zum Laufen bekommen. Daher nutzen wir lieber RStudio als den uneinheitlichen Editor, um die Konzepte im Buch vorzustellen.

Konventionen in diesem Buch

Codeausschnitte erscheinen wie in dem folgenden Beispiel, indem wir eine Million Würfe zweier sechsseitiger Würfel simulieren:

```
> set.seed(42)
> throws <- 1e6
> dice <- sapply(1:2,
+ function(x) sample(1:6, throws, replace = TRUE)
+ )
> table(rowSums(dice))
```

2	3	4	5	6	7	8
28007	55443	83382	110359	138801	167130	138808
9	10	11	12			
110920	83389	55816	27945			

Jede Zeile des R-Codes in diesem Beispiel beginnt mit einem der folgenden Symbole:

- ✓ **>:** Das Anweisungssymbol `>`. Es ist nicht Teil des Codes. Geben Sie es nicht ein, wenn Sie den Code selbst ausprobieren.
- ✓ **+:** Das Fortsetzungssymbol `+`. Es zeigt an, dass diese Zeile eigentlich noch zur vorhergehenden gehört. Genau genommen müssen Sie gar keine Zeilenumbrüche in Ihren Code einbauen. Wir tun dies jedoch häufig, um die Lesbarkeit des Codes zu verbessern. Darüber hinaus ist es hilfreich, damit er auf die Buchseiten passt.

Die Zeilen, die weder mit dem einen noch mit dem anderen Symbol beginnen, sind Ausgaben von R. In unserem Beispiel erhalten Sie die Gesamtzahl an Würfeln, in denen die Summe der Augen 2, 3, ..., 12 betrug. Zum Beispiel war die Summe der Augen in 28.007 von einer Million Würfeln gleich zwei.




Sie können diese Codeausschnitte kopieren und in R ausführen. Achten Sie darauf, sie genau abzuschreiben. Es gibt nur drei Ausnahmen:

- ✓ Geben Sie nicht das Kommandosymbol `>` ein.
- ✓ Geben Sie nicht das Fortsetzungssymbol `+` ein.
- ✓ Tabulatoren oder Leerzeichen können Sie beliebig im Code verteilen, solange es nicht innerhalb von Schlüsselwörtern ist. Mit dem Zeilenvorschub sollten Sie etwas vorsichtiger umgehen.

Wenn R eine Eingabe von Ihnen erwartet, zeigt es das mit dem Symbol `>` ganz links in der Zeile, etwa so:

```
> print("Hallo Welt!")
```

Wenn Sie diese Anweisung in die Konsole eingeben und  drücken, antwortet R mit:

```
[1] "Hallo Welt"
```

Aus Bequemlichkeit werden sowohl die Eingabe als auch die Ausgabe in einem Block angezeigt:

```
> print("Hallo Welt!")
```

```
[1] "Hallo Welt!"
```

Schließlich sei noch auf die Schriftart von R-Code im Buch hingewiesen. Die meisten Wörter in R sind von englischen Wörtern abgeleitet. Zwar ist dies bei einem deutschen Text nicht so verwirrend wie bei einem englischen. Dennoch setzen wir R-Funktionen, Argumente und Schlüsselwörter in Monofont. Funktionen werden immer zusammen mit nachgestellten geschlossenen Klammern dargestellt – zum Beispiel `plot()`. Funktionsargumente geben wir grundsätzlich nicht an und weichen davon nur in wichtigen Fällen ab.

Manchmal geht es um Menübefehle, zum Beispiel DATEI | SICHERN (FILE | SAVE). Dies bedeutet lediglich, dass Sie gebeten werden, das Menü DATEI (FILE) zu öffnen und anschließend die Option SICHERN (SAVE) zu wählen.

Was Sie nicht lesen müssen

Sie können dieses Buch so lesen, wie es für Sie am besten passt. Wenn Sie jedoch unter Zeitdruck stehen

oder weniger an Details interessiert sind, können Sie problemlos alle rein technischen Informationen auslassen (zum betreffenden Symbol siehe weiter hinten in dieser Einleitung). Darüber hinaus können Sie auch alle grau hinterlegten Kästen auslassen. Zwar enthalten sie interessante Informationen, jedoch nichts, was für das Verständnis des Themas benötigt wird. Wenn Sie aber doch alles lesen, seien Sie gnädig wegen so mancher (absichtlicher) Wiederholung.

Törichte Annahmen über den Leser

Folgende Annahmen treffen wir über Sie als Leser sowie Ihren Computer:

- ✓ **Sie kennen sich mit Ihrem Computer bestens aus.** Sie wissen, wie man Software herunterlädt und installiert. Sie haben Zugang zum Internet und wissen, wie man dort Informationen findet.
- ✓ **Sie sind kein Programmierer.** Wenn Sie doch Programmierer sind und weitere oder andere Sprachen gewohnt sind, schauen Sie sich gern die rein technischen Informationen an (zum betreffenden Symbol siehe weiter hinten in dieser Einleitung). Dort erfahren Sie mehr dazu, wo R genauso oder anders tickt als andere Sprachen.
- ✓ **Sie sind kein Statistiker, aber Sie verstehen die Grundlagen der Statistik.** *R für Dummies* ist kein Buch über Statistik, obgleich wir Ihnen zeigen, wie man mit R einfache statistische Analysen durchführen kann. Wenn Sie mehr über Statistik erfahren wollen, empfehlen wir Ihnen *Statistik für Dummies* oder

Statistik mit R für Dummies (beide erschienen im Wiley-VCH Verlag).

- ✓ **Sie wollen neue Dinge entdecken.** Sie mögen es, Probleme zu knacken, und haben keine Angst, mal etwas in der R-Konsole auszuprobieren.

Wie dieses Buch aufgebaut ist

R für Dummies gliedert sich in sechs Teile. Folgende Themen erwarten Sie in den einzelnen Teilen:

Teil I: Sind Sie beReit?

In diesem Teil lernen Sie R kennen und schreiben Ihr erstes Skript. Sie machen sich mit dem sehr nützlichen Vektorkonzept in R vertraut und führen Berechnungen simultan auf vielen Variablen aus. Sie lernen den R-Arbeitsbereich (englisch *workspace*) kennen, das heißt, wie Sie Variablen erzeugen, verändern und entfernen. Sie finden heraus, wie Sie Ihre Arbeit speichern und wie Sie Skriptdateien laden und verändern, die Sie in vorangegangenen Sitzungen erstellt haben. Darüber hinaus zeigen wir Ihnen ein paar Grundlagen in R, zum Beispiel wie Sie den Funktionsumfang erweitern, indem Sie Packages installieren.

Teil II: Arbeiten mit R

In diesem Teil füllen wir Sie ab mit den drei R: Reading (lesen), wRiting (schreiben) und aRithmetic (rechnen) – mit anderen Worten, wie Sie mit Text und Zahlen arbeiten, und nicht zu vergessen: mit Datumswerten. Hier lernen Sie auch die für das Leben mit R unerlässlichen Konzepte der Listen und Datensätze (*data frame*) kennen.