

# Installation und erste Schritte mit R

Dr. Matthias Kohl

Lehrstuhl für Stochastik

<http://www.stoch.uni-bayreuth.de>



UNIVERSITÄT  
BAYREUTH

Oktober 2008

## Inhaltsverzeichnis

1	Was ist R?	2
2	Der Aufbau von R	4
3	Installation von R	4
4	Installation von Erweiterungspaketen	5
5	Wo bekomme ich Hilfe?	7
6	Erste Schritte	8
7	Arbeiten mit R	10
8	Daten Import und Export	11

## 1 Was ist R?

R [R Development Core Team, 2008] ist eine freie, nicht-kommerzielle Implementierung der bei den AT&T Bell Laboratories von Rick Becker, John Chambers und Mitarbeitern entwickelten Statistik-Programmiersprache S. Die Entwicklung von S begann im Jahr 1976 – genauer am 5. Mai 1976 (vgl. S. 476 in [Chambers, 2008]) – mit Version 1. Im Jahre 1980 folgte Version 2, die im Jahr 1988 von Version 3 abgelöst wurde. Seit 1998 ist nunmehr die Version 4 auf dem Markt; vgl. <http://cm.bell-labs.com/cm/ms/departments/sia/S/history.html>. Version 4 von S ist ausführlich in [Chambers, 1998] beschrieben.

Die kommerzielle Version “S-Plus” wurde dann nicht mehr von den Bell Laboratories, sondern von der Insightful Corporation weiterentwickelt und vertrieben. Im Sommer 2008 wurde Insightful Corporation von TIBCO Software Inc. übernommen und gehört nun zur TIBCO Spotfire enterprise analytics Produktgruppe. Eine kostenlose S-Plus Version für Studenten mit eingeschränkter Funktionalität kann unter <http://elms03.e-academy.com/splus/> bezogen werden.

Das R Projekt wurde im Jahr 1991 von Ross Ihaka und Robert Gentleman initiiert. Die ersten Dateien von R erschienen 1993 auf Statlib (<http://lib.stat.cmu.edu/>). Die Quelldateien wurden erstmals im Jahr 1995 unter GPL veröffentlicht. Die erste offizielle Version R 0.60 wurde am 5. Dezember 1997 herausgegeben. Die Version R 1.0.0, welche am 29. Februar 2000 veröffentlicht wurde, stellte eine Implementation von S Version 3 (S3) dar. Am 4. Oktober 2004 wurde R 2.0.0 herausgegeben, welches eine Implementation von S Version 4 (S4) darstellt, aber auch bereits Elemente enthält, die über S4 hinausgehen. Die aktuelle Version von R, die eine Weiterentwicklung von S4 darstellt ist in [Chambers, 2008] sehr gut beschrieben.

Für die Entwicklung des Grundsystems von R zeichnet sich das sog. R Core verantwortlich, welches aktuell aus 19 Mitgliedern besteht. Im Jahr 2002 wurde außerdem die R Foundation gegründet (vgl. <http://www.r-project.org/foundation/main.html>), an der die R Core Mitglieder als ordentliche Mitglieder beteiligt sind. Die Ziele dieser Foundation umfassen das Fortführen der Entwicklung von R, die Erforschung neuer Methoden, die Lehre und das Training im Bereich von Statistical Computing sowie die Organisation von Zusammenkünften und Konferenzen mit einer Ausrichtung auf Statistical Computing.

Zusammenfassend läßt sich R als eine Entwicklungsumgebung und Programmiersprache für Statistik und Graphik beschreiben, welche unter GNU GPL-2 entwickelt wird und somit jederzeit ohne Einschränkung auf beliebig vielen Rechnern installiert werden kann, wobei es Versionen für Windows, Linux und MacOS X gibt.

```
R > license()
```

```
This software is distributed under the terms of the GNU General  
Public License Version 2, June 1991. The terms of this license  
are in a file called COPYING which you should have received with
```

```
this software and which can be displayed by RShowDoc("COPYING").
```

If you have not received a copy of this file, you can obtain one at <http://www.R-project.org/licenses/>.

A small number of files (the API header files listed in R\_DOC\_DIR/COPYRIGHTS) are distributed under the Lesser GNU General Public License version 2.1. This can be displayed by `RShowDoc("COPYING.LIB")`, or obtained at the URI given.

'Share and Enjoy.'

Der freie Charakter sowie der modulare Aufbau von R, haben dazu geführt, dass es mittlerweile eine nahezu exponentiell wachsende Anzahl (aktuell ca. 1500) von Erweiterungen für R gibt. Diese und andere interessante Fakten zur Entwicklung von R wurden von John Fox

<http://www.statistik.uni-dortmund.de/useR-2008/abstracts/Fox.pdf>

und Kurt Hornik

<http://www.statistik.uni-dortmund.de/useR-2008/abstracts/Hornik.pdf>

auf der useR! – The R User Conference 2008 vorgestellt.

Konkret ist R durch diese Erweiterungen z.B. in der Lage, Dateien von anderen gängigen Statistik-Paketen wie etwa SPSS, EpiInfo, Stata, Minitab, SAS und verschiedener Versionen von S-Plus zu laden und weiterzuverarbeiten. Zudem bietet R die Möglichkeit, qualitativ hochwertige Graphiken in nahezu allen gängigen Graphikformaten zu erzeugen und man kann außerdem Berichte in HTML-, XML-, LaTeX- (und damit in dvi-, ps- bzw. pdf-Format) oder Open Document (und damit in MS-Word-Format) Format generieren. Die Seite <http://addictedtor.free.fr/graphiques/> gibt einen Einblick in die Möglichkeiten, die R bezüglich der Erzeugung von Graphiken bietet und stellt zusätzlich für alle dort vorgestellten Bilder jeweils den R Code zur Verfügung, mit dem diese erzeugt wurden.

In den letzten Jahren sind weitere Projekte entstanden, die auf R basieren; siehe <http://www.r-project.org/other-projects.html>. Hervorzuheben sind das Projekt Bioconductor (vgl. [www.bioconductor.org](http://www.bioconductor.org)), welches R Pakete für die Analyse von biologischen Daten zu Verfügung stellt, das Projekt Omega (vgl. [www.omegahat.org](http://www.omegahat.org)), welches insbesondere R Pakete mit Schnittstellen zu anderen Sprachen und Systemen bereitstellt, sowie das Projekt R GUI (vgl. [http://www.sciviews.org/\\_rgui/](http://www.sciviews.org/_rgui/)), welches verschiedene Projekte im Zusammenhang mit der Bereitstellung einer benutzerfreundlichen GUI für R zusammenfasst.

## 2 Der Aufbau von R

Wie bereits erwähnt ist der Aufbau von R modular, genauer spricht man in R von Paketen (packages), die in eine Bibliothek (library) installiert werden. Das Grundsystem von R, welches vom sog. R Core betreut wird, besteht aus den Paketen "base", "datasets", "graphics", "grDevices", "grid", "methods", "splines", "stats", "stats4", "tcltk", "tools" und "utils".

Die nächste Hierarchiestufe bilden die sog. "recommended packages", die aktuell aus den Paketen "boot", "cluster", "codetools", "foreign", "KernSmooth", "lattice", "mgcv", "nlme", "rpart" und "survival" sowie dem R bundle (Zusammenfassung von mehreren Paketen) "VR" besteht. Das bundle "VR" beinhaltet die Pakete "class", "MASS", "nnet" und "spatial". Es ist geplant, dass beim Release der Version 2.9.0 (ca. April 2009) außerdem das Paket "Matrix" den recommended Status erhält.

Zu den base und recommended Paketen kommen dann eine riesige Anzahl (aktuell ca. 1500) von Erweiterungspaketen hinzu; vgl. <http://cran.at.r-project.org/web/packages/index.html>, aber auch <http://www.bioconductor.org/download> und <http://www.omegahat.org/>.

Einen Überblick über die Pakete geben sog. "task views" (vgl. <http://cran.r-project.org/web/views/>), bei denen ein Teil der Erweiterungspakete in Kategorien wie z.B. Econometrics, Finance, Pharmacokinetics oder SocialSciences zusammengefasst sind.

Der Einsatz von R – genauer der base und recommended Pakete, aber nicht der anderen Erweiterungspakete – in klinischen Studien wird in "R: Regulatory Compliance and Validation Issues. A Guidance Document for the Use of R in Regulated Clinical Trial Environments" behandelt; vgl. <http://www.r-project.org/certification.html>.

## 3 Installation von R

Die Files, die benötigt werden, um R unter Windows, Linux oder MacOS X zu installieren, können von CRAN (Comprehensive R Archive Network; vgl. <http://cran.r-project.org/>) oder einem der Mirrors (vgl. <http://cran.r-project.org/mirrors.html>) heruntergeladen werden.

**Windows** Der Windows Installer findet sich unter <http://cran.r-project.org/bin/windows/base/>. Weitere Informationen zu Installation, Update und Deinstallation unter Windows finden sich in den FAQs für Windows unter <http://cran.r-project.org/bin/windows/base/rw-FAQ.html>.

**MacOS X** Die notwendigen Dateien für MacOS X sowie eine kurze Anleitung finden sich unter <http://cran.r-project.org/bin/macosx/>. Ähnlich wie im Fall von Windows gibt es auch für MacOS X eine FAQ-Seite (<http://cran.r-project.org/bin/macosx/RMacOSX-FAQ.html>), auf der zusätzliche Informationen zu finden sind.

**Linux** Es stehen Debian (<http://cran.r-project.org/bin/linux/debian/>) und Ubuntu (<http://cran.r-project.org/bin/linux/ubuntu/>) Pakete sowie Dateien für Redhat (<http://cran.r-project.org/bin/linux/redhat/>) und OpenSUSE (<http://cran.r-project.org/bin/linux/suse/>) zur Verfügung. Auf den Seiten von Debian, Ubuntu und OpenSUSE finden sich auch kurze Anleitungen zur Installation.

Die offizielle und umfassende Dokumentation zur Installation von R findet sich in dem Manual “R Installation and Administration” [R Development Core Team, 2008], welches unter <http://cran.r-project.org/manuals.html> als HTML- und pdf-Version zur Verfügung gestellt wird. Darin finden sich u.a. Informationen dazu, wie R unter der Verwendung der Quelldateien auf verschiedenen Systemen installiert werden kann.

## 4 Installation von Erweiterungspaketen

Nach dem Start von R kann die Installation von Erweiterungspaketen in Windows über die Menüleiste der GUI vorgenommen werden. Genauer ist dies über den Menüpunkt “Pakete” und dort den Unterpunkt “Installiere Paket(e) ...” möglich.

Neben dieser Möglichkeit gibt es generell die Möglichkeit, die R Funktionen `available.packages`, `install.packages`, `update.packages`, `download.packages`, `old.packages`, `new.packages`, `installed.packages` und `remove.packages` zu verwenden. Mit diesen ist es möglich, festzustellen, welche Pakete es gibt, diese zu installieren, upzudaten oder auch deinstallieren; zum Beispiel

```
R > ## Installation des Paketes "distr"
R > install.packages("distr", repos = "http://cran.at.r-project.org")
R > ## Entfernen des Paketes "distr"
R > remove.packages("distr")
```

Es ist auch möglich, Pakete mittels Kommandozeilenbefehlen von außerhalb von R aus zu installieren. Weitere Einzelheiten dazu finden sich im Manual “R Installation and Administration”; vgl. <http://cran.at.r-project.org/manuals.html>.

Die in einer Bibliothek (`library`) installierten Pakete lassen sich mit Hilfe der Funktion `library` feststellen und dann auch laden.

```
R > ## installierte Pakete
R > library()
R > ## Lade Paket "distr"
R > library(distr)
```

Ein Überblick über ein Paket läßt sich ebenfalls mittels der Funktion `library` gewinnen; z.B.

```
R > ## Überblick über Paket "distr"
R > library(help = distr)
```

Die grundlegenden Informationen eines Paketes sind in der sog. DESCRIPTION-Datei enthalten, die es für jedes Paket gibt. Deren Inhalt läßt sich auch von R aus mittels der Funktion `packageDescription` inspizieren; z.B.

```
R > ## Das "base" Paket
R > packageDescription("base")

Package: base
Version: 2.8.0
Priority: base
Title: The R Base Package
Author: R Development Core Team and contributors worldwide
Maintainer: R Core Team <R-core@r-project.org>
Description: Base R functions
License: GPL (>= 2)
Built: R 2.8.0; ; Mi 15. Okt 15:04:23 CEST 2008; unix

-- File: /home/btm722/RTOP/R-2-8-branch/library/base/Meta/package.rds
```

```
R > ## Das "distr" Paket
R > packageDescription("distr")

Package: distr
Version: 2.0.1
Date: 2008-10-10
Title: Object orientated implementation of distributions
Description: Object orientated implementation of distributions
Author: Florian Camphausen, Matthias Kohl, Peter Ruckdeschel, Thomas
       Stabla
Maintainer: Peter Ruckdeschel <Peter.Ruckdeschel@itwm.fraunhofer.de>
Depends: R(>= 2.2.0), methods, graphics, startupmsg, sfsmisc
Imports: stats
LazyLoad: yes
License: GPL (version 2 or later)
URL: http://distr.r-forge.r-project.org/
Packaged: Sat Oct 11 09:38:27 2008; btm722
Built: R 2.8.0; ; 2008-10-16 10:35:45; unix

-- File: /home/btm722/RTOP/R-2-8-branch/library/distr/Meta/package.rds
```

## 5 Wo bekomme ich Hilfe?

In R gibt es verschiedene Formen von Hilfe. Als erstes zu nennen sind die offiziellen Manuals “An Introduction to R”, “The R language definition”, “Writing R Extensions”, “R Data Import/Export”, “R Installation and Administration”, “R Internals” und “The R Reference Index” [R Development Core Team, 2008]; vgl. <http://cran.r-project.org/manuals.html>. Darüber hinaus gibt es eine Vielzahl weiterer Manuals, die unter <http://www.r-project.org/other-docs.html> näher beschrieben werden. Es handelt sich dabei um Bücher <http://www.r-project.org/doc/bib/R-books.html>, aber auch um viele kostenlose Manuskripte <http://cran.r-project.org/other-docs.html>. Interessantes Material zu einführenden und weiterführenden Themen findet man auch auf den Seiten des Bioconductor Projekts; vgl. <http://www.bioconductor.org/workshops>.

Nach der erfolgreichen Installation von R steht eine weitere Hilfsfunktionalität in Form der Funktionen `?`, `help`, `help.search` sowie `help.start` zur Verfügung.

```
R > ?help
R > help()
R > ?help.search
R > ?help.start
```

Zum Beispiel

```
R > ?exp
R > ??exp
```

Die Funktionen `?` und `help` ermöglichen es, zu einem R Objekt (z.B. einer R Funktion) oder einem Datensatz die entsprechende Hilfeseite zu öffnen. Mit `??` und `help.search` ist es möglich, alle Hilfeseiten zu einem gewissen Thema zu suchen, wobei sich die Suche über alle installierten R Pakete erstreckt. Die Funktion `help.start` schließlich startet die HTML-Version der R Hilfe.

Seit etwa Mitte 2006 bietet R eine weitere Art der Hilfe in Form einer Wiki, die unter <http://wiki.r-project.org/> zu finden ist. Darüber hinaus gibt es auf der R Homepage unter <http://www.r-project.org/search.html> Links zu verschiedenen Suchmaschinen. Sowie die bereits im vorangegangenen Abschnitt angesprochenen, umfangreichen FAQs-Seiten; vgl. <http://cran.r-project.org/faqs.html>.

Liefern alle diese vorgestellten Hilfemöglichkeiten keine Lösung zu einem vorgegebenen Problem, so gibt es eine weitere, “ultimative” Möglichkeit der Hilfe in Form verschiedener Mailinglisten; vgl. <http://www.r-project.org/mail.html>.

Dies ist eine sehr effektive und schnelle Möglichkeit, an Hilfe zu gelangen. Es ist jedoch dringend angeraten, den sogenannten “Posting Guide” zu lesen (siehe <http://www.r-project.org/posting-guide.html>), bevor man diesen Weg beschreitet. Außerdem sollte man vorher kontrollieren, dass die Frage bzw. eine ähnliche Frage nicht bereits (mehrfach) gestellt wurde. Dies ist möglich über <http://finzi.psych.upenn.edu/search.html> bzw. indem man innerhalb von R die Funktion `RSiteSearch` verwendet; z.B.

```
R > RSiteSearch("exponential")
```

## 6 Erste Schritte

Im einfachsten Fall kann man R auch als einen überdimensionierten Taschenrechner verwenden. Es lassen sich die verschiedensten Funktionen mit R auswerten; z.B.

```
R > ## Exponentialfunktion  
R > exp(2)
```

```
[1] 7.389056
```

```
R > ## natürlicher Logarithmus  
R > log10(10)
```

```
[1] 1
```

```
R > ## Gamma Funktion  
R > gamma(3)
```

```
[1] 2
```

Genauso sind die bekannten arithmetischen Operationen implementiert.

```
R > ## Addition  
R > 1+1
```

```
[1] 2
```

```
R > ## Subtraktion  
R > 2-1
```

```
[1] 1
```

```
R > ## Multiplikation  
R > 2*2
```

```
[1] 4
```

```
R > ## Division  
R > 3/2
```

```
[1] 1.5
```

```
R > ## Potenzieren  
R > 2^3
```

```
[1] 8
```

```
R > ## Division mit Rest ("modulo")  
R > 4 %% 2
```

```
[1] 0
```

```
R > 4.5 %% 2
```

```
[1] 0.5
```

```
R > ## Integer-Division  
R > 4 %/% 2
```

```
[1] 2
```

```
R > 4.5 %/% 2
```

```
[1] 2
```

Um Hilfe für arithmetische oder andere (etwa logische) Operatoren zu bekommen, ist es nötig, diese in Anführungszeichen zu setzen, also z.B. `'+'` oder auch `help('+')`.

Natürlich möchte man die Ergebnisse von Berechnungen auch gerne zwischenspeichern; d.h., das Ergebnis einer Variablen zu weisen. Hierfür kann in R `<-` oder `=` verwendet werden (oder auch: `<<-` oder allgemeiner `assign`). Es empfiehlt sich jedoch `<-` zu verwenden. Einige gute Gründe hierfür liefert Bill Venables unter <http://finzi.psych.upenn.edu/R/Rhelp02a/archive/126700.html>.

```
R > x <- 2 + 3
R > x
```

```
[1] 5
```

```
R > y <- 5-2
R > y
```

```
[1] 3
```

```
R > x+y
```

```
[1] 8
```

Durch die Eingabe des Namens einer Variable, gefolgt von dem Drücken der Enter/Return-Taste wird die Ausgabe des Inhalts der Variable erreicht. Genauer wird dadurch die Funktion `print` aufgerufen, welche dann für die Ausgabe sorgt.

## 7 Arbeiten mit R

Für das Arbeiten mit R genügt ein einfacher Texteditor, mit dem man die einzelnen R Befehle in einer Textdatei abspeichern kann. Üblicherweise verwendet man für diese Textdateien den Dateityp “.r” oder “.R”. Um sich die Arbeit zu erleichtern, empfiehlt es sich jedoch einen Editor zu verwenden, der zusätzliche Funktionalität wie etwa das Hervorheben von Schlüsselwörtern oder evtl. sogar das Übermitteln von R Code nach R übernimmt. Je nach Betriebssystem gibt es eine Vielzahl von Möglichkeiten; vgl. [http://www.sciviews.org/\\_rgui/projects/Editors.html](http://www.sciviews.org/_rgui/projects/Editors.html).

**Windows** Hervorzuheben sind Tinn-R (vgl. <http://www.sciviews.org/Tinn-R/>), WinEdt (vgl. <http://www.winedt.com/>) in Kombination mit dem R Paket "RWinEdt" von Uwe Ligges sowie XEmacs (vgl. <http://www.xemacs.org/>) zusammen mit ESS (Emacs Speaks Statistics, vgl. <http://ess.r-project.org/> bzw. <http://socserv.mcmaster.ca/jfox/Books/Companion/ESS/>).

**MacOS X** Examples are bluefish (vgl. <http://bluefish.openoffice.nl/index.html>), SubEthaEdit (vgl. <http://www.codingmonkeys.de/subethaedit/>) oder Alpha (vgl. <http://alphancl.sourceforge.net/wikit/>).

**Linux** Es sind Emacs (vgl. <http://www.gnu.org/software/emacs/>) bzw. XEmacs (vgl. <http://www.xemacs.org/>), aber auch viele andere Editoren wie etwa Vim (<http://www.vim.org/>) oder Kate (vgl. <http://kate.kde.org/>), die Funktionalität für R bereitstellen.

Eine andere Möglichkeit ist die Verwendung spezieller GUIs (graphical user interfaces) (siehe <http://www.r-project.org/GUI>). Hier sind aus meiner Sicht besonders der R Commander (vgl. <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/>), JGR (vgl. <http://stats.math.uni-augsburg.de/JGR/>) und das Sciviews Projekt (vgl. <http://www.sciviews.org/SciViews-R>) zu nennen. Einen Überblick gibt Philippe Grosjean unter [http://www.sciviews.org/\\_rgui/](http://www.sciviews.org/_rgui/).

## 8 Daten Import und Export

Für den Import und Export von Daten steht in R eine Vielzahl von Möglichkeiten zur Verfügung. Wir werden hier nur die grundlegenden Funktionen vorstellen. Die Basisfunktion ist `scan`. Mit dieser Funktion können Daten von der Konsole oder einer Datei in einen Vektor oder eine Liste eingelesen werden. Die Funktion `read.table` kann dazu verwendet werden, Daten in Tabellenformat einzulesen. Abgeleitet von `read.table` sind die Funktionen `read.csv`, `read.csv2`, `read.delim` und `read.delim2`. Diese Funktionen rufen ihrerseits `read.table` auf, wobei Sie gewisse Defaulteinstellungen verwenden. So eignet sich etwa `read.csv2` dazu, `.csv`-Dateien einzulesen, bei denen die Spalten mit “;” getrennt sind und als Dezimalzeichen “.” verwendet wird. Sollten Sie mit einer deutschen MS Windows Version und MS Office bzw. OpenOffice arbeiten, so können Sie Dateien mit dieser Formatierung erzeugen, indem Sie vorhandene Tabellen einfach unter Verwendung dieses Dateityps abspeichern.

Eine weitere Form des Datenimports stellt die R Funktion `load` dar. Diese eignet sich dazu, sogenannte `.RData`-Files einzuladen; d.h., Dateien, die man z.B. mit den R Funktionen `save` oder `save.image` erzeugen kann. Was uns direkt zum Datenexport führt. Es ist also möglich, mit Hilfe der R Funktionen `save` oder `save.image`, den Inhalt einzelner Variablen oder den gesamten Speicherinhalt einer R Sitzung in einer `.Rdata`-Datei abzuspeichern, wobei man zusätzlich wählen kann, ob die Datei komprimiert werden soll oder nicht. Das Gegenstück zu den `read.*` Funktionen stellen die Funktionen `write.table`, `write.csv` und `write.csv2` dar. Im Fall, dass Sie mit einer deutschen MS Windows Version und MS Office bzw. OpenOffice arbeiten, können Sie mit Hilfe von `write.csv2` eine `.csv`-Datei erzeugen, die sich dann direkt mit MS Excel bzw. dem Tabellenkalkulationsprogramm von OpenOffice öffnen läßt.

Einen Überblick über die Vielzahl von weiteren Möglichkeiten, die es zum Datenimport und Export gibt – etwa auch die Verbindung zu relationalen Datenbanken –, findet sich im Manual “R Data Import/Export” [R Development Core Team, 2008]; siehe <http://cran.r-project.org/doc/manuals/R-data.html>.

## Literatur

[Chambers, 1998] Chambers, J. (1998). *Programming with Data. A Guide to the S Language*. Springer. 2

- 
- [Chambers, 2008] Chambers, J. (2008). *Software for Data Analysis: Programming with R*. Springer. [2](#)
- [R Development Core Team, 2008] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [2](#), [5](#), [7](#), [11](#)