

Statistik für Informatiker, SS 2017

Organisatorisches, ein Startbeispiel und eine sehr knappe R-Einführung

Matthias Birkner

<http://www.staff.uni-mainz.de/birkner/StatInfo17/>

19.4.2017



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

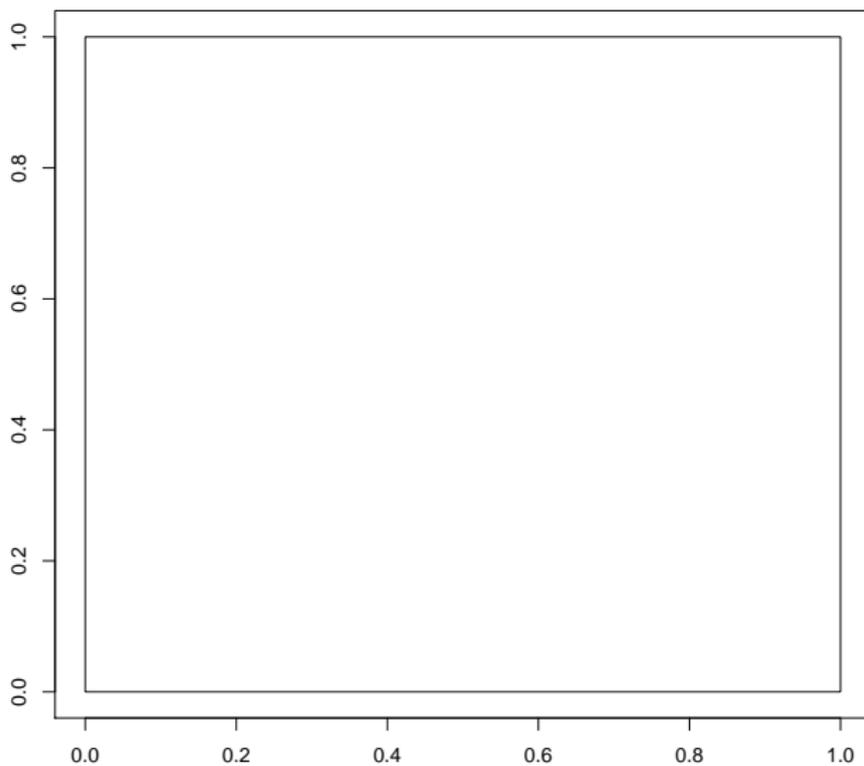
Herzlich willkommen zur Statistik für Informatiker im SS2017!

- 1 Organisatorisches
- 2 Startbeispiel
- 3 Erste Schritte mit R

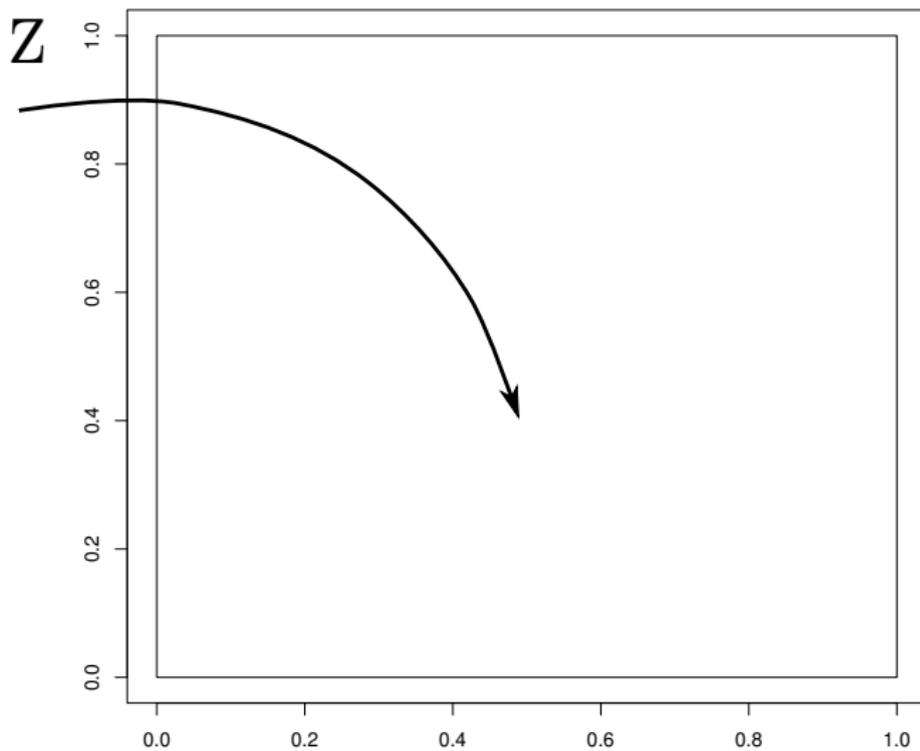
Orte, Zeiten, Zulassungskriterien, ...

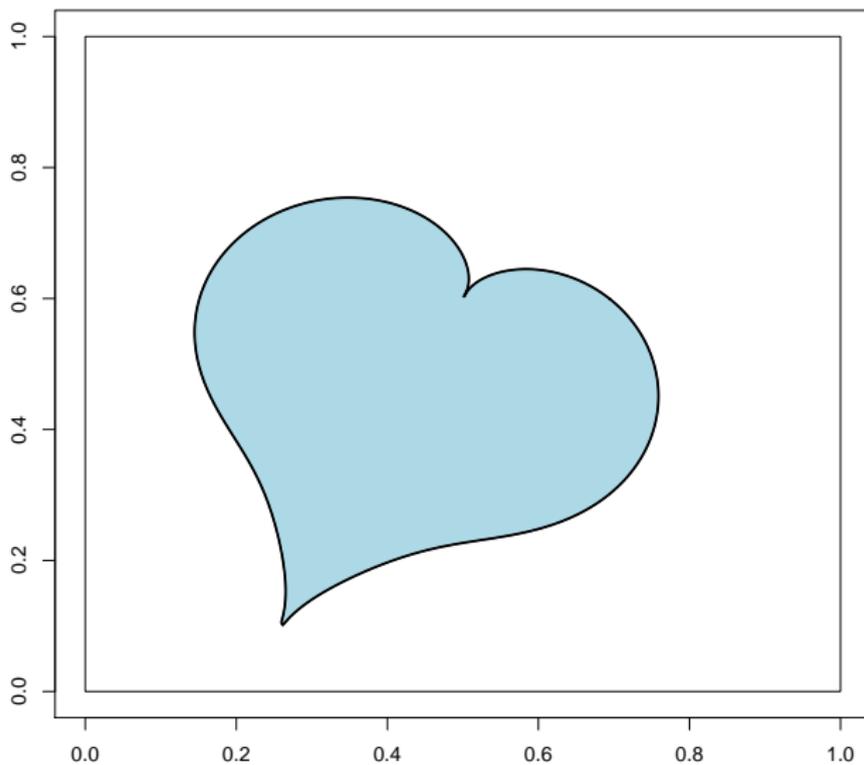
- Mi. 10–12h, Raum 03-428
- Homepage:
<http://www.staff.uni-mainz.de/birkner/StatInfo17/>
(dort werden die Folien, benutzter R-Code/Datensätze, etc. stehen)
- Übungsbetrieb:
Organisiert von Frederik Klement
<https://www.stochastik.mathematik.uni-mainz.de/statistik-fuer-informatik/>
(verwenden [auch] R <http://www.r-project.org/>, dazu später mehr ...)

Einheitsquadrat



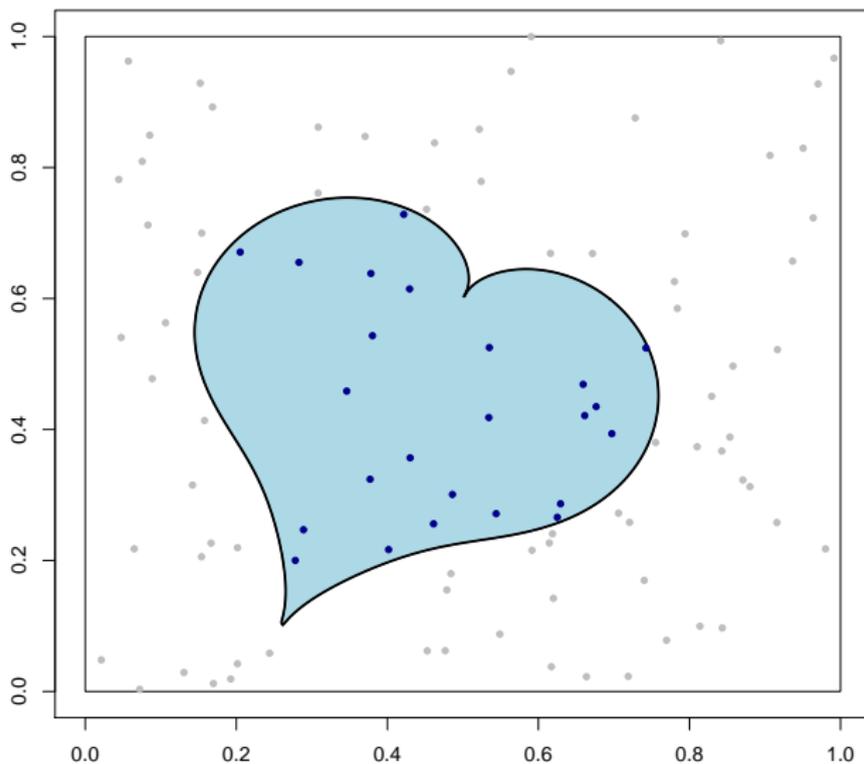
Einheitsquadrat



Einheitsquadrat und Teilmenge B 

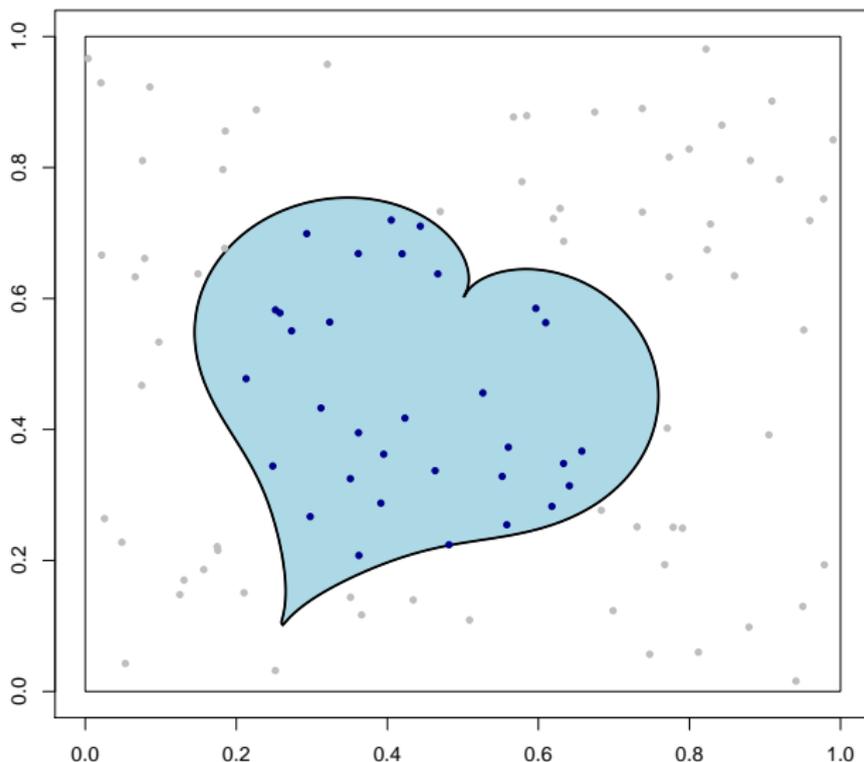
Einheitsquadrat und 100 uniform ausgewählte Punkte

100 zufällige Punkte, Anteil in B : 0.24



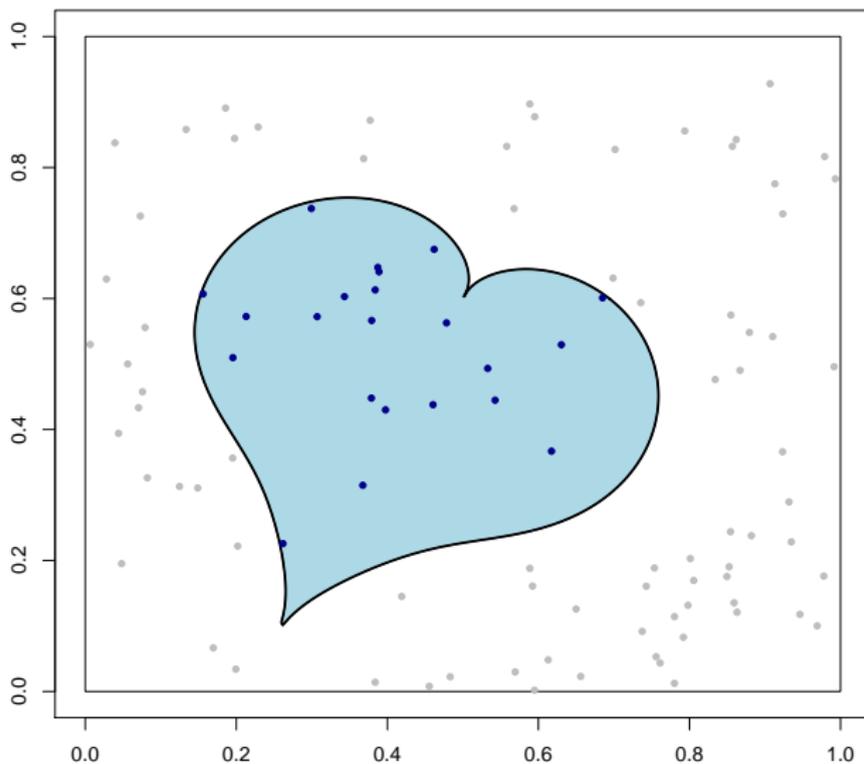
Einheitsquadrat und 100 uniforme ausgewählte Punkte (Wdh.)

100 zufällige Punkte, Anteil in B : 0.32

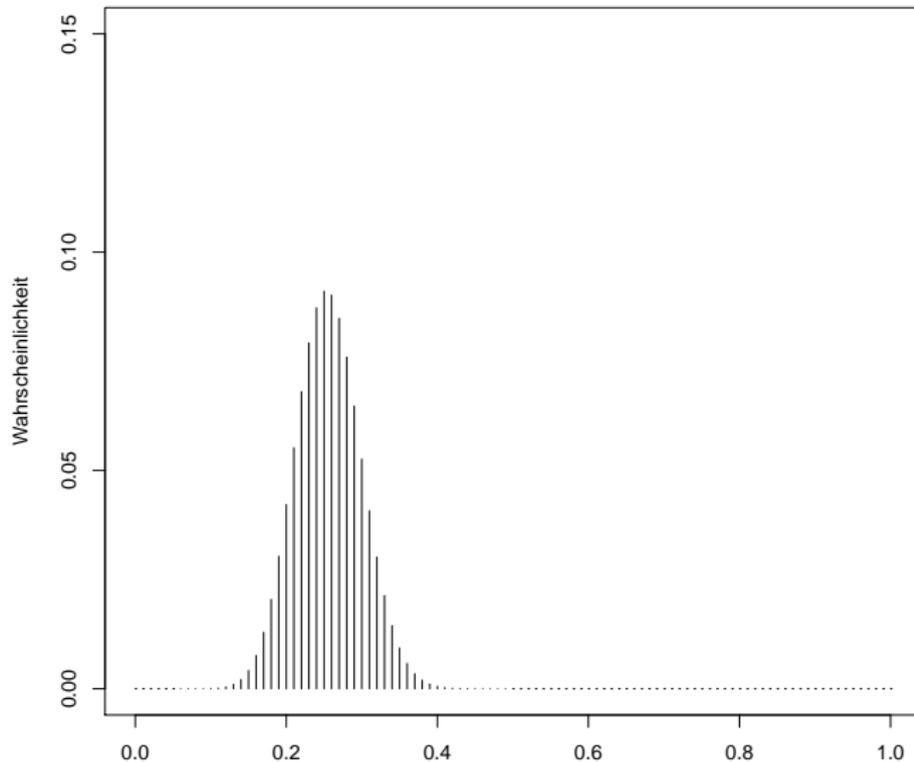


Einheitsquadrat und 100 uniform ausgewählte Punkte (2. Wdh.)

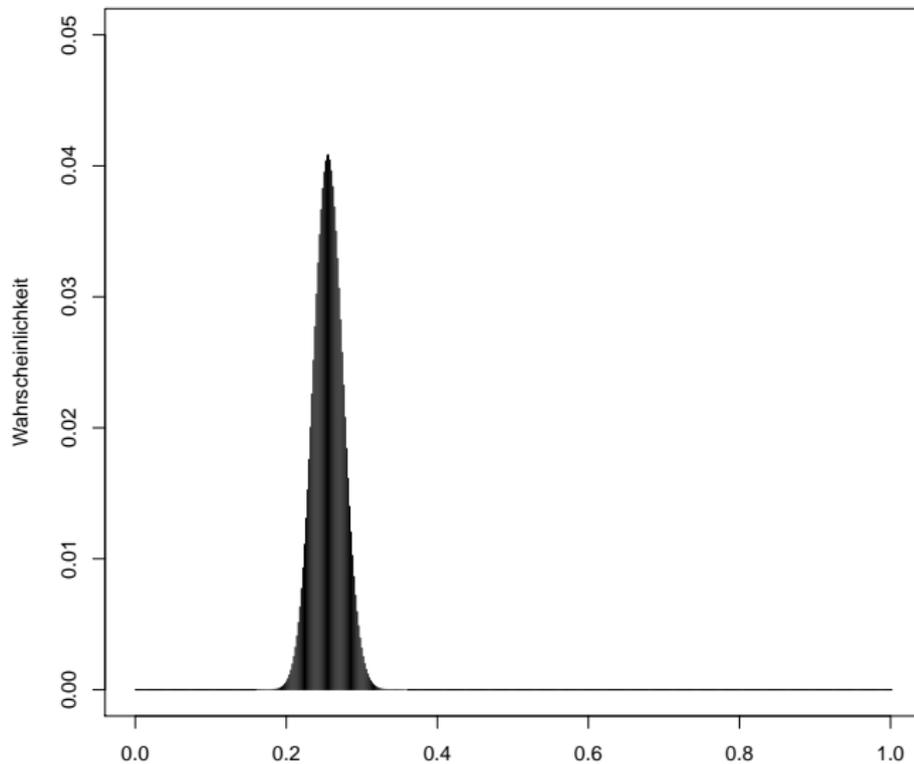
100 zufällige Punkte, Anteil in B : 0.22

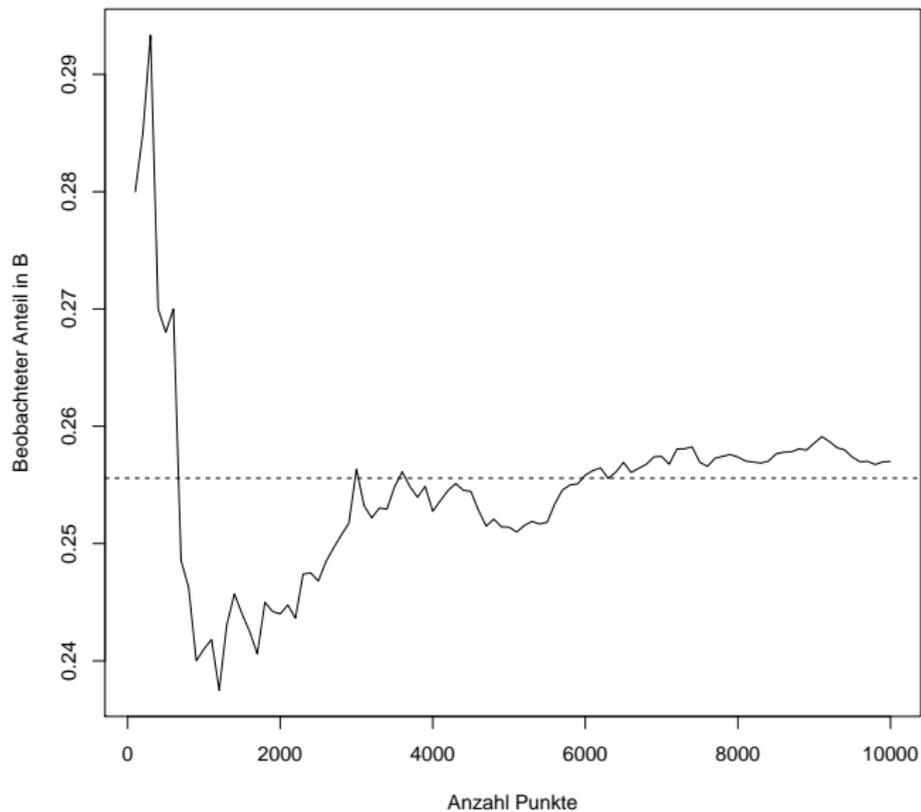


Verteilung der möglichen Werte von M (Mittelung über 100 Punkte)



Verteilung der möglichen Werte von M (Mittelung über 500 Punkte)



Empirischer Anteil in B 

Übrigens:

$$p = \int_0^{2\pi} \frac{1}{98} \left(2 - 2 \sin(w) + \frac{\sin(w) \sqrt{|\cos(w)|}}{\sin(w) + 7/5} \right)^2 dw$$
$$\cong 0.25557221\dots$$

Was ist R?

- R ist eine für die Statistik und für stochastische Simulation entwickelte Programmiersprache, zudem sind viele statistische Standardverfahren bereits in R implementiert oder als Zusatzpaket verfügbar.
- R hat eine sehr aktive Benutzer- und Entwicklergemeinschaft, die nahezu alle Bereiche der Statistik und viele Anwendungsbereiche (z.B. Populationsgenetik, Finanzmathematik) überdeckt.
- R ist frei verfügbar unter der GNU general public license, für (nahezu) alle Rechnerarchitekturen erhältlich:
<http://www.r-project.org/>

Literatur zu R

W.N. Venables et al, *An Introduction to R*,
<http://cran.r-project.org/manuals.html>

Zur ersten Einführung auch die Hinweise zu R, „nullte Schritte“ auf der Vorlesungs-Homepage.

Weiterführend (z.T. wesentlich über den Stoff der Vorlesung hinaus):

- Günther Sawitzki, *Einführung in R*,
<http://sintro.r-forge.r-project.org/>
- William N. Venables, Brian D. Ripley, *Modern applied statistics with S* („Standardwerk“, UB Lehrbuchsammlung)
- Lothar Sachs and Jürgen Hedderich, *Angewandte Statistik – Methodensammlung mit R* (E-Book, UB)
- Christine Duller, *Einführung in die nichtparametrische Statistik mit SAS und R : ein anwendungsorientiertes Lehr- und Arbeitsbuch* (E-Book, UB)
- Helge Toutenburg, Christian Heumann, *Deskriptive Statistik : Eine Einführung in Methoden und Anwendungen mit R und SPSS* (E-Book, UB)
- Uwe Ligges, *Programmieren mit R* (E-Book, UB)

R installieren, starten, anhalten

Installation: Windows, Mac OS: Binaries von

<http://www.r-project.org/> (siehe Link Download, Packages, CRAN dort)

Linux: Für die meisten Distributionen gibt es fertige Pakete (entweder bei CRAN oder im Repository der Distribution)

Bei Fragen oder Problemen: Übungsgruppen

R starten: Windows, Mac OS: Icon (ggf. aus Menu) anklicken,

Linux/Unix: > R auf einer Konsole (oder z.B. mit ESS aus Emacs heraus).

R beenden: q() (fragt, ob Daten gespeichert werden sollen)

laufende Rechnungen unterbrechen: CTRL-C

Hinweis: [RStudio](https://www.rstudio.com/) <https://www.rstudio.com/> bietet eine integrierte Entwicklungsumgebung für R [für Privatanwender frei verfügbar].

Einige einfache (und wichtige) R-Befehle

Mathematische Operatoren und Funktionen:

`+`, `-`, `*`, `/`, `^`, `exp`, `sin`, `log`, etc.

Hilfe aufrufen:

`help(Befehl)` oder `?Befehl`

Online-Hilfe starten:

`help.start()`

Variable einen Wert zuweisen:

`x <- 5`

`y <- "Zeichenkette"`

`z <- TRUE`

`w <- -2+7.5i`

Vektor erzeugen, Elementzugriff:

`v <- c(1,2,3.1415,-17); v[3]`

Rezyklierungs-Regel bei Vektoren:

`v+2` liefert `3,4,5.1415,-15`

Liste erzeugen, Elementzugriff:

`l <- list(1.2,"text",-5,FALSE); v[[2]]`

(Pseudo-)Zufallszahl generieren:

`runif`

Demos starten (und bestaunen):

`demo()`

Einlesen eines Skripts:

`source`

Ausgabe in Datei umlenken:

`sink`

Grafikausgaben:

`plot, ...`

Grafik„geräte“ öffnen:

`x11`, `postscript`, `pdf`,
`dev.copy2eps`, `dev.copy2pdf`, ...

Grafikparameter setzen: `par`

R-Befehle für deskriptive Statistik

Komplexere Datensätze organisiert man in R oft in sogenannten `data.frames`

Einlesen von Daten: `scan`, `read.table`

Nützliche Befehle:

`mean`, `sd`, `median`, `plot`, `hist`, `boxplot`, `barplot`, `pie`,
`rug`, `jitter`

In Grafiken kann man Farben, Titel, Achsenbeschriftungen, Schriftgrößen, etc. individuell anpassen (s.a. Hilfe)