

Statistik und Graphiken mit R

[Termine 1 & 2]

Miguel Alvarez

28. Februar & 02. März 2023

▶ Termin 1 & 2

- ▶ Grundlagen
- ▶ Datentypen

▶ Termin 3 & 4

- ▶ Objekten
- ▶ Lesen und Schreiben

▶ Termin 5 & 6

- ▶ Statistiken
- ▶ Graphiken (1)

▶ Termin 7 & 8

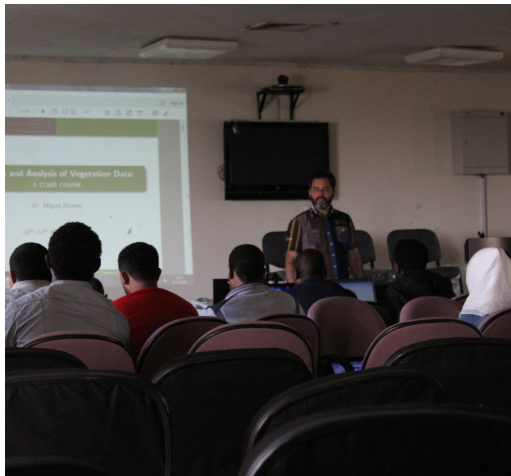
- ▶ Graphiken (2)
- ▶ Fortgeschrittenes Programmieren
- ▶ Abschluss

18:00 – 20:30

Pause 19:15 – 19:30

- ▶ Methoden
 - ▶ Vortrag
 - ▶ Life-Codierung
 - ▶ Übungen

<https://kamapu.github.io/GrundkursR/>



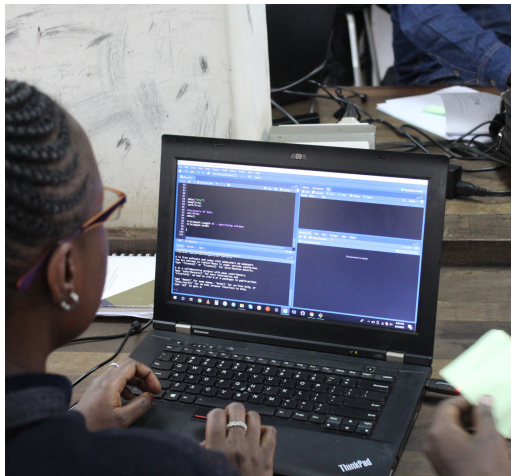
Miguel Alvarez



- ▶ Vegetationsökologist
- ▶ Data Scientist
- ▶ R-Programmierer
 - ▶ Umwelt und Biodiversität
 - ▶ Geographische Informationssysteme
 - ▶ Datenbanken
 - ▶ Reproduzierbarkeit

Die Teilnehmer?

- ▶ Warum R?
- ▶ Erwartung von dem Kurs



Geschichte

- ▶ S (1975)
 - ▶ S-PLUS
 - ▶ TIBCO Spotfire S+
- ▶ R (1992)

R ist die kostenlose Alternative zu **S**

Ross Ihaka

Robert Gentleman



Was ist R?

- ▶ Programmiersprache
 - ▶ Open Source (offene Quelle)
 - ▶ Freeware (kostenlos)
- ▶ Statistische Umgebung
 - ▶ Interface
 - ▶ Terminal
 - ▶ Editoren

<https://www.r-project.org/>


[\[Home\]](#)

Download
CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Conferences
Search
Get Involved: Mailing Lists
Get Involved: Contributing
Developer Pages
R Blog

R Foundation

Foundation
Board
Members
Donors
Donate

Help With R

Getting Help

Documentation

Manuals
FAQs
The R Journal
Books
Certification

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 4.2.2 (Innocent and Trusting) prerelease versions** will appear starting Friday 2022-10-21. Final release is scheduled for Monday 2022-10-31.
- **R version 4.2.1 (Funny-Looking Kid)** has been released on 2022-06-23.
- **R version 4.1.3 (One Push-Up)** was released on 2022-03-10.
- Thanks to the organisers of useR! 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the [R Consortium YouTube channel](#).
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

News via Twitter

 The R Foundation Retweeted

 **R Contributo...** @R_Contributo... · Oct 10
R Contribution Working Group meeting Oct 18, 18:30 - 19:30 UTC

We'll discuss progress on current issues (github.com/r-devel/rcontr...), including

- Translation hackathons (LatinR, AsiaR)
- Office hours
- Code of conduct

All welcome. Zoom registration:
us02web.zoom.us/joining/regist...

Was ist R?

- ▶ Schwerpunkt
 - ▶ Mathematik
 - ▶ Grafiken
- ▶ Kommandozeilen
- ▶ REPL (Read-Eval-Print-Schleife)

```
(5 + 100) * 25
```

```
## [1] 2625
```

```
50/10
```

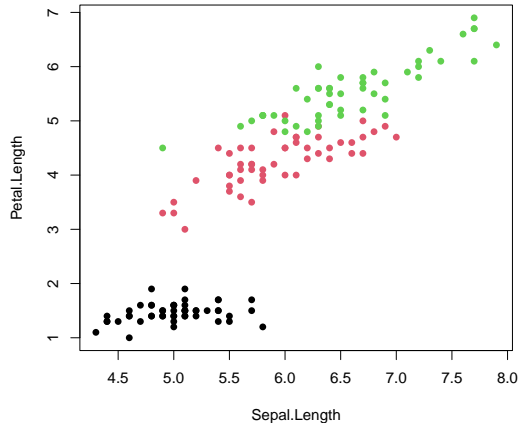
```
## [1] 5
```

```
60 > 15
```

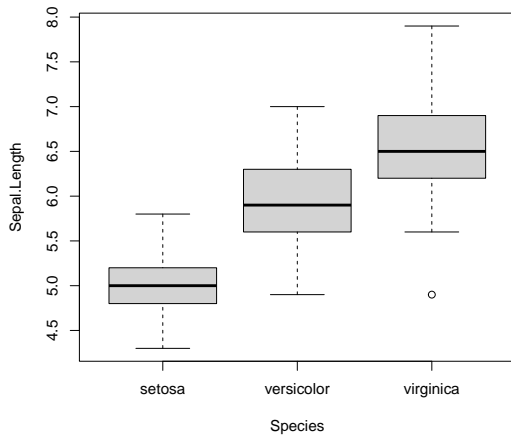
```
## [1] TRUE
```



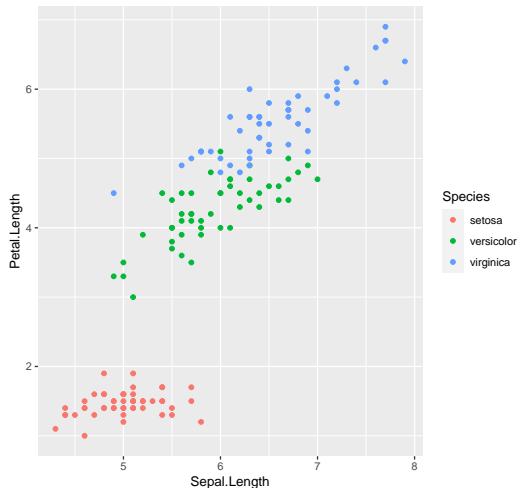
```
plot(iris[ , c("Sepal.Length",
               "Petal.Length")],
     pch = 16, col = iris$Species)
```



```
boxplot(Sepal.Length ~ Species,  
        data = iris)
```



```
library(ggplot2)
ggplot(iris,
       aes(x = Sepal.Length,
           y = Petal.Length,
           color = Species)) +
  geom_point()
```



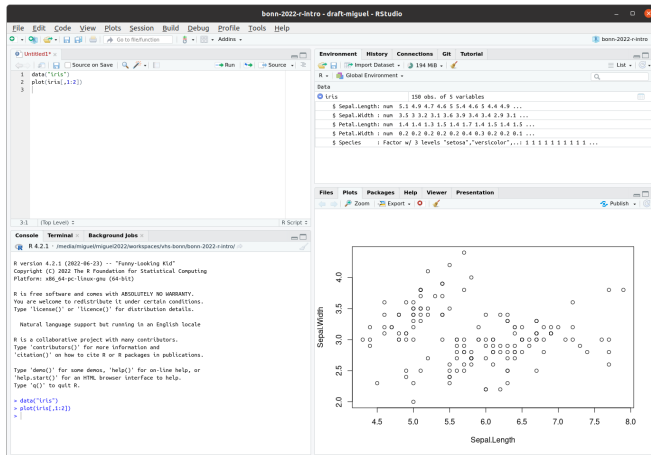
Warum R?

- ▶ Kostet nichts
- ▶ Steuerung von Analysen
- ▶ Skripten
 - ▶ Protokolle
 - ▶ Reproduzierbarkeit
- ▶ Vielseitig

Jargon

- ▶ Konsole
- ▶ Sitzung
- ▶ Workspace (Environment)
- ▶ Arbeitsverzeichnis (working directory)
- ▶ Skript
- ▶ Paket (Package)

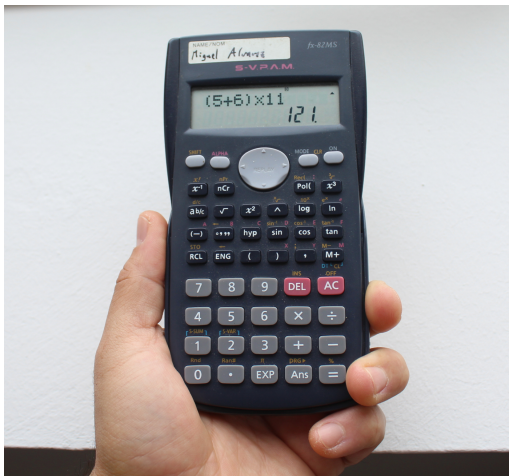
- ▶ Text Editor
- ▶ IDE (integrated development environment)



eclipse (StatET)

The screenshot shows the Eclipse IDE with the following components:

- Project Explorer:** Displays a file tree for the project 'vhs-bonn - bonn-2022-r-intro'. The file 'FolienRKurs-12.Rmd' is selected.
- Editor:** Shows the R code for 'FolienRKurs-12.Rmd'. The code includes comments for RStudio and R as a calculator, and a plot command: `plot(liris[, 1:2])`.
- Console:** Displays the R startup message, including the R version (4.4.0) and the location of the R installation.
- Scatter Plot:** A plot of Sepal.Length vs Petal.Length for the 'liris' dataset. The x-axis is labeled 'Sepal.Length' and ranges from 4.5 to 8.0. The y-axis is labeled 'Petal.Length' and ranges from 2.0 to 4.0. The plot shows a clear separation between the three species of Iris.



- ▶ Mathematische Operatoren
- ▶ Klammern

$5+6*11$

[1] 71

$(5+6)*11$

[1] 121

$(5-6)/11$

[1] -0.09090909

Zuweisung

`<-`

Zuweisung von einem Wert zu einem Objekt.

Konvention

- ▶ Nutze `<-` anstatt `->`
- ▶ Reserviere `=` für Argumente in Funktionen
- ▶ Nutze Leerzeichen für eine bessere Übersicht in Quellcode
- ▶ Nutze Einrückungen für gebrochene Kommandozeilen

Logische Operatoren

==

!=

>

>=

<

<=

&

|

%in%

!

any()

all()

```
10 > 15
```

```
## [1] FALSE
```

```
10 < 15
```

```
## [1] TRUE
```

Vektoren

Der Vektor ist die grundlegende Datenstruktur in **R**

- ▶ Länge `length()`
- ▶ Typ `class()`
- ▶ Evtl. Namen `names()`

```
c(1:10)
## [1] 1 2 3 4 5 6 7 8 9 10
rep(5, times = 10)
## [1] 5 5 5 5 5 5 5 5 5 5
LETTERS[1:5]
## [1] "A" "B" "C" "D" "E"
```

Indexieren

- ▶ Eckige Klammern
- ▶ Index
 - ▶ integer
 - ▶ logical (Bedingung)
 - ▶ character (Namen)

```
# Mit integer
```

```
letters[15]
```

```
## [1] "o"
```

```
# Mit logischen Werten
```

```
letters[!letters %in% c("a", "b", "c")]
```

```
## [1] "d" "e" "f" "g" "h" "i" "j" "k" "l" "m"
```

```
## [20] "w" "x" "y" "z"
```

```
# Mit Namen
```

```
names(letters) <- letters
```

```
letters["m"]
```

```
## m
```

```
## "m"
```

Datentypen

- ▶ integer
- ▶ numeric
- ▶ logical
- ▶ factor
- ▶ character

```
A <- c(1:10)
is.numeric(A)
## [1] TRUE
```

```
B <- as.character(A)
B
## [1] "1" "2" "3" "4" "5" "6" "7" "8"
is.numeric(B)
## [1] FALSE
```

Sonderklassen

- ▶ NA
- ▶ NaN
- ▶ NULL
- ▶ Inf
- ▶ -Inf

5/0

```
## [1] Inf
```

```
log(0)
```

```
## [1] -Inf
```

```
sqrt(-1)
```

```
## Warning in sqrt(-1): NaNs produced
```

```
## [1] NaN
```

Rückmeldungen

- ▶ Konsolenoutput
- ▶ Nachricht (message)
- ▶ Warnung (warning)
- ▶ Fehlermeldung (error)

Eine Warnung ist nicht zwingend ein Fehler.

```
sqrt(-1)
```

```
## Warning in sqrt(-1): NaNs produced
```

```
## [1] NaN
```

```
5 + "B"
```

```
## Error in 5 + "B": non-numeric argument to bin
```

Fehlerbehebung

- ▶ Dokumentation
- ▶ Foren
- ▶ Programfehler melden

```
# help()
?mean
# help.search()
??"multinomial"
RSiteSearch("Import")
```


Regeln für Hilfeanfrage

- ▶ Recherechiere vorher
- ▶ KESS (*keep it simple stupid*)
- ▶ Zeige ein **minimales Beispiel**
- ▶ Bleibe höflich
- ▶ Melde die Endlösung
- ▶ Lerne **markdown**

<https://stackoverflow.com/questions>
<https://www.markdowntutorial.com/>

stackoverflow Products Search...

Home

PUBLIC

Questions

Tags

Users

Companies

COLLECTIVES

Explore Collectives

TEAMS

Stack Overflow for Teams – Start collaborating and sharing organizational knowledge.

Create a free Team

Why Teams?

P-values from aov in R

Asked 7 years, 1 month ago Modified 6 years, 10 months ago Viewed 2k times

Take this month's Pulse Survey

I'm conducting a simulation study in R. Basically, I generate fake data sets and then run an ANOVA on the data using the aov function. But I'm having difficulty extracting p-values. Previous questions do not help ([Extract p-value from aov](#)) – I am running a mixed ANOVA.

First I have an ANOVA:

```
results <- summary(aov(dv={A*B*C*D*E}*Error(subj)/(A*B*C*D)), data = mdata)) # con
```

which generate this output:

```
Error: subj
      Df Sum Sq Mean Sq F value Pr(>F)
E       1 1039157 1039157    0.95  0.334
Residuals 58 63428016 1093586

Error: subj:A
      Df Sum Sq Mean Sq F value Pr(>F)
A       1 1996 1996    0.229  0.641
A:E     1 2294 2294    0.253  0.617
Residuals 58 526389 9076

...
```

I'm truncating the output for space. What I want list of p-values with the effect name (A or A:E). I have halfway succeeded, but it's messy. I can extract the p-values using this `get_p` function that I made.

```
#Function
get_p = function(results,headIf
```

Weitere Dokumente

- ▶ Demos
 - ▶ `demo()`
- ▶ Beispielssitzungen
 - ▶ `browsevignette()`
 - ▶ `vignette()`
- ▶ Zahlreiche Tutorials

```
demo("graphics")  
vignette("ggplot2-specs")
```

Vielen Dank!

```
library(fortunes)  
fortune(283)
```

```
##  
## The good way to do it is to include the following comment at the beginning:  
## # This is a holy Script, please edit it not  
## -- Kenn Konstabel (on "... how to protect R Script files from inadvertent  
## editing by users.")  
## R-help (April 2011)
```