Einführung in R(Studio)

Explorative Datenanalyse mit R
The R Bootcamp @ CSS

in









Dezember 2019

R ist eine Programmiersprache

Eine Programmiersprache ist eine **formale Sprache** die eine Reihe Instruktionen für alle möglichen Ziele spezifiziert. Programmiersprachen bestehen aus instruktionen für einen Computer und werden genutzt um Algorithmen zu implementieren.

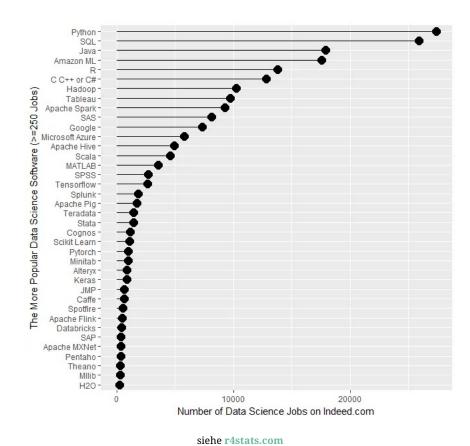
Algorithmus in Worten

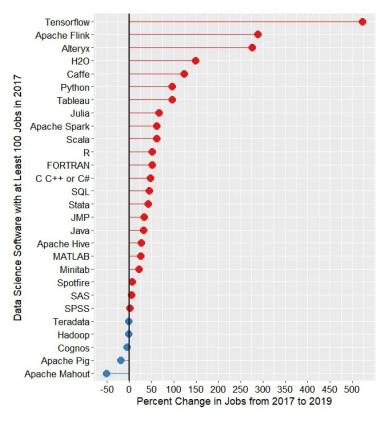
- 1. Lade Daten
- 2. Extrahiere Variablen
- 3. Führe Analyse durch
- 4. Zeige Resultate

Implementation in R

```
data <- read.table(link)</pre>
variables <- data[,c('gruppe','variable')]</pre>
analysis <- lm(variable ~ gruppe, data = variables)</pre>
summary(analysis)
```

Warum R?





siehe r4stats.com

(cc) BY-SA www.therbootcamp.com

R wird relevant bleiben

R ist de facto die lingua franca für Statistik und Datenanalyse.

Pro

- 1. **Open-source** und umsonst.
- 2. Community (e.g., stackoverflow)
- 3. Erweiterbarkeit (CRAN)
- 4. Tidyverse
- 5. RStudio
- 6. Produktivität: Latex, Markdown, GitHub

Ehemals Contra

- 1. **Unschön**e Sprache wird überarbeitet (Tidyverse)
- 2. Langsame Elemente werden ersetzt (Rcpp)
- 3. Brücken zu externen Tools/Sprachen (rPython, tensorflow)

Komponenten von R





RStudio



adapted from rstudio.com

R Packages



adapted from towardsdatascience.com

Das mächtige tidyverse

Das tidyverse ist im Kern eine Sammlung hoch-performanter, nutzerfreundlicher Pakete, die speziell für eine effizientere Datenanalyse entwickelt wurden.

- 1. ggplot2 für Grafiken.
- 2. dplyr für Datenverarbeitung.
- 3. tidyr für Datenverarbeitung.
- 4. readr für Daten I/O.
- 5. purrr für funktionales Programmieren.
- 6. tibble für moderne data. frames.













Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

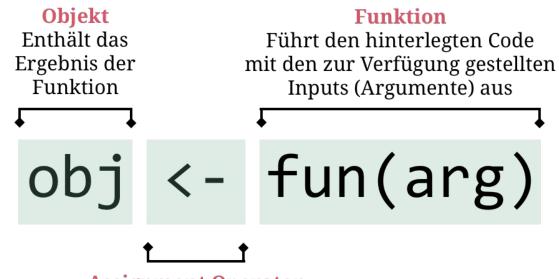
- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

Syntax

- **Der Assignment Operator <-**
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio



Assignment Operator

Kreiert/verändert das Objekt rechts durch das Ergebnis der Funktion rechts

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Ein Objekt namens eins_zwei_drei
eins_zwei_drei <- c(1, 2, 3)
# Printe das Objekt
print(eins_zwei_drei)
## [1] 1 2 3
```

```
# Printe das Objekt
eins_zwei_drei
```

[1] 1 2 3

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Ein Objekt namens eins_zwei_drei
eins_zwei_drei <- c(1, 2, 3)
# Berechne den Mittelwert
mean(eins_zwei_drei)
```

[1] 2

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Ein Objekt namens eins_zwei_drei
eins_zwei_drei <- c(1, 2, 3)
# Berechne den Mittelwert
ergebnis <- mean(eins_zwei_drei)</pre>
# Benutze Ergebnis
ergebnis * 10
```

12 / 40

[1] 20

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Ein Objekt namens eins_zwei_drei
eins_zwei_drei <- c(1, 2, 3)
# Printe das Objekt
eins_zwei_drei
## [1] 1 2 3
# Addiere 100
eins_zwei_drei + 100
## [1] 101 102 103
# Printe das Objekt
eins_zwei_drei
```

[1] 1 2 3

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Ein Objekt namens eins_zwei_drei
eins_zwei_drei <- c(1, 2, 3)
# Printe das Objekt
eins_zwei_drei
## [1] 1 2 3
```

```
# Ändere das Objekt
eins_zwei_drei <- eins_zwei_drei + 100
# Printe das Objekt
eins_zwei_drei
```

[1] 101 102 103

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

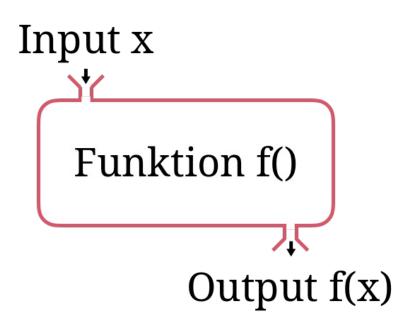
```
# Ein Objekt namens eins_zwei_drei
eins_zwei_drei <- c(1, 2, 3)
# Ergänze um weitere Zahl
und_vier <- c(eins_zwei_drei, 4)</pre>
# Printe das Objekt
und_vier
```

[1] 1 2 3 4

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio



Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Funktion c()
eins_zwei_drei <- c(1, 2, 3)
# Funktion `+`()
eins_zwei_drei + 100
## [1] 101 102 103
# Funktion print()
eins_zwei_drei
## [1] 1 2 3
# Funktion mean()
mean(x = eins_zwei_drei)
## [1] 2
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Kein Argument
mean()
## Error in mean.default(): argument "x" is missing, with no
# Ein (notwendiges) Argument
mean(c(1, 2, 3))
## [1] 2
# Hinzufügen eines fehlenden Werts (NA)
mean(c(1, 2, 3, NA))
## [1] NA
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Kein Argument
mean()
## Error in mean.default(): argument "x" is missing, with no
# Ein (notwendiges) Argument
mean(c(1, 2, 3))
## [1] 2
# Ändere den Default zur Entfernung des NAs
mean(c(1, 2, 3, NA), na.rm = TRUE)
## [1] 2
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Berechne Mittelwert von c(1, 2, 3)
mean(c(1, 2, 3))
## [1] 2
# Berechne Mittelwert von c('1', '2', '3')
mean(c("1", "2", "3"))
## Warning in mean.default(c("1", "2", "3")): argument is no
## [1] NA
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Klasse von c(1, 2, 3)
class(c(1, 2, 3))
## [1] "numeric"
# Klasse von c('1', '2', '3')
class(c("1", "2", "3"))
## [1] "character"
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

?mean

mean {base} Arithmetic Mean Description Generic function for the (trimmed) arithmetic mean Usage mean(x, ...) mean(x, trim = 0, na.rm = FALSE, ...) An R object. Currently there are methods for numeric/logical vectors and date, date-time and time interval objects. Complex vectors are allowed for trim the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the na.rm a logical value indicating whether NA values should be stripped before the computation proceeds ... further arguments passed to or from other methods. Value If trim is zero (the default), the arithmetic mean of the values in x is computed, as a numeric or complex vector of length one. If x is not logical (coerced to numeric), numeric (including integer) or complex, $NA_real_$ is returned, with a warning. If trim is non-zero, a symmetrically trimmed mean is computed with a fraction of trim observations deleted from each end before the mean is computed Becker, R. A., Chambers, I. M. and Wilks, A. R. (1988) The New S Language, Wadsworth & Brooks/Cole See Also weighted.mean, mean.POSIXct, colMeans for row and column means. x < -c(0:10, 50)c(xm, mean(x, trim = 0.10))

[Package base version 3.6.0 Index]

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

?mean

Usage

```
mean(x, ...)
## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)
```

Arguments

- An R object. Currently there are methods for numeric/logical vectors and date, date-time and time interval objects. Complex vectors are allowed for trim = 0, only.
- trim the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
- na.rm a logical value indicating whether NA values should be stripped before the computation proceeds.
- further arguments passed to or from other methods.

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Mittelwert von c(1, 2, 3)
mean(c(1, 2, 3))
## [1] 2
# Mittelwert von c(TRUE, FALSE, TRUE)
mean(c(TRUE, FALSE, TRUE))
## [1] 0.6667
# Mittelwert von c("1", "2", "3")
mean(c("1", "2", "3"))
## Warning in mean.default(c("1", "2", "3")): argument is no
## [1] NA
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

?mean

Usage

```
mean(x, ...)
## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)
```

Arguments

- An R object. Currently there are methods for numeric/logical vectors and date, date-time and time interval objects. Complex vectors are allowed for trim = 0, only.
- trim the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
- na.rm a logical value indicating whether NA values should be stripped before the computation proceeds.
- further arguments passed to or from other methods.

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Mittelwert von c(1, 2, 3, NA)
mean(c(1, 2, 3, NA), na.rm = TRUE)
## [1] 2
# Mittelwert von c(1, 2, 3, NA)
mean(c(1, 2, 3, NA), na.rm = "TRUE")
## [1] 2
# Mittelwert von c(1, 2, 3, NA)
mean(c(1, 2, 3, NA), na.rm = "ABCD")
## Error in if (na.rm) x <- x[!is.na(x)]: argument is not in
```

Syntax

- Der Assignment Operator <--</p>
- <- kreiert/verändert Objekte</p>
- 3 Alles passiert durch Funktionen
- 4 Funktionen haben (Default) Argumente
- **5** Argumente erwarten Klassen
- **6** Findet Hilfe mit?

- **7** Funktionen leben in Paketen
- 8 Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- 10 Benutzt Projekte in Rstudio

```
# Klasse von TRUE
class(TRUE)

## [1] "logical"

# Klasse von "TRUE"
class("TRUE")

## [1] "character"

# Klasse von "ABCD"
class("ABCD")

## [1] "character"
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Wandle "TRUE" in logical um
as.logical("TRUE")
## [1] TRUE
# Wandle "ABCD" in logical um
as.logical("ABCD")
## [1] NA
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

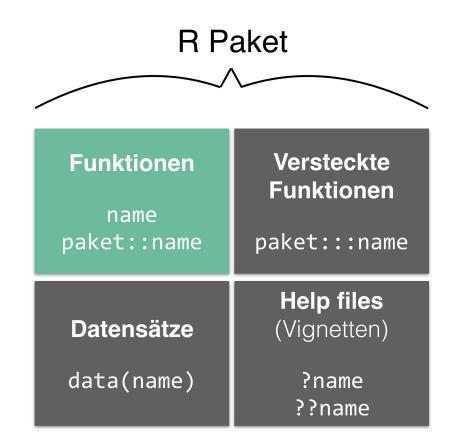
- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

Kurze Pause?

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio



Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

R RDocumentation

Search for packages, functions, etc

Leaderboard

indexed packages

indexed functions

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

Installiere Pakete einmal mit install.packages()

```
install.packages("tidyverse")
```

Lade existierende Pakete jedes mal mit library()

```
library(tidyverse)
```

```
    Attaching packages

tidyverse 1.2.1 —

✓ ggplot2 3.1.0

                  ✓ purrr 0.2.5

✓ tibble 2.0.1

✓ dplyr 0.7.6

✓ tidyr 0.8.1 ✓ stringr 1.3.1

✓ readr 1.1.1

✓ forcats 0.3.0

— Conflicts —
se_conflicts() —
# dplyr::filter() masks stats::filter()
# dplyr::lag() masks stats::lag()
Warning message:
package 'tibble' was built under R version 3.5.2
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

```
# Message: Interessant
basel <- type_convert(baselers)</pre>
## Parsed with column specification:
## cols(
     sex = col_character()
## )
# Warning: Wichtig
result <- mean('NA')</pre>
## Warning in mean.default("NA"): argument is not numeric or
# Error: Beheben
lenth(x = 1)
## Error in lenth(x = 1): could not find function "lenth"
```

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

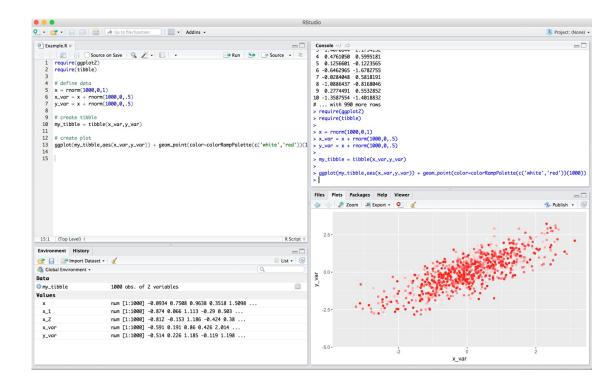
- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio

Fehler	Beschreibung
<pre>'could not find function'</pre>	Typo oder Paket nicht geladen
'error in eval'	Ein Objekt wird verwendet, dass nicht existiert.
'cannot open()'	Typo oder fehlende Ordner im Pfad.
'no applicable method'	Funktion erwartet andere Klassen.
package errors	Fehler im installieren, kompilieren, oder laden von Paketen.

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- **Benutzt R in RStudio**
- Benutzt Projekte in Rstudio

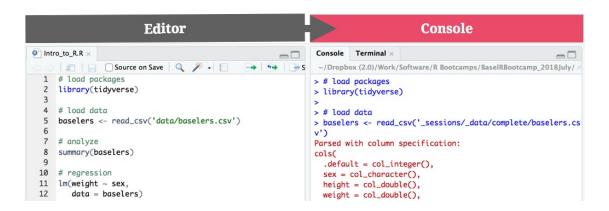


Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio



Shortcut für schicke Code zur Console:

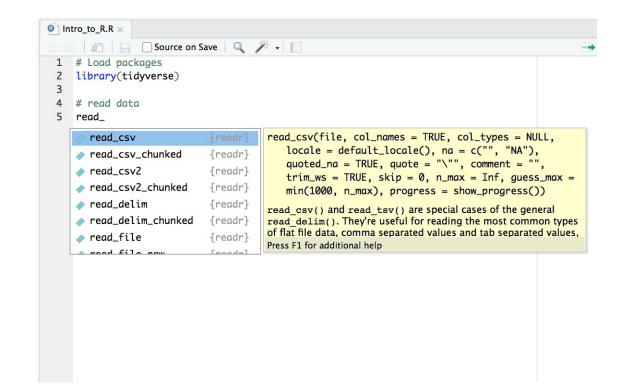
Shortcut für führe Chunk wiederholt aus:

$$\mathcal{H}/\text{ctrl} + \mathcal{O} + p$$

Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- Benutzt Projekte in Rstudio



Syntax

- Der Assignment Operator <-
- <- kreiert/verändert Objekte
- Alles passiert durch Funktionen
- Funktionen haben (Default) Argumente
- Argumente erwarten Klassen
- Findet Hilfe mit?

Produktivität

- Funktionen leben in Paketen
- Lernt von Fehlern und Warnungen
- Benutzt R in RStudio
- **Benutzt Projekte in Rstudio**

Projekte helfen bei...

workspace und history speichern • projektspezifische Optionen setzen • Dateien finden • Versionskontrolle • etc.

- 0_Materials
- 1_Data
- 2_Code
- 3_Figures
- 9_Graveyard
- projectXY.Rproj

Downloads



Interactive

