

BIOINFORMATICA

STATISTIEK IN R_

Week 2

ONDERWERPEN

Werken met Rstudio

Datastructuren

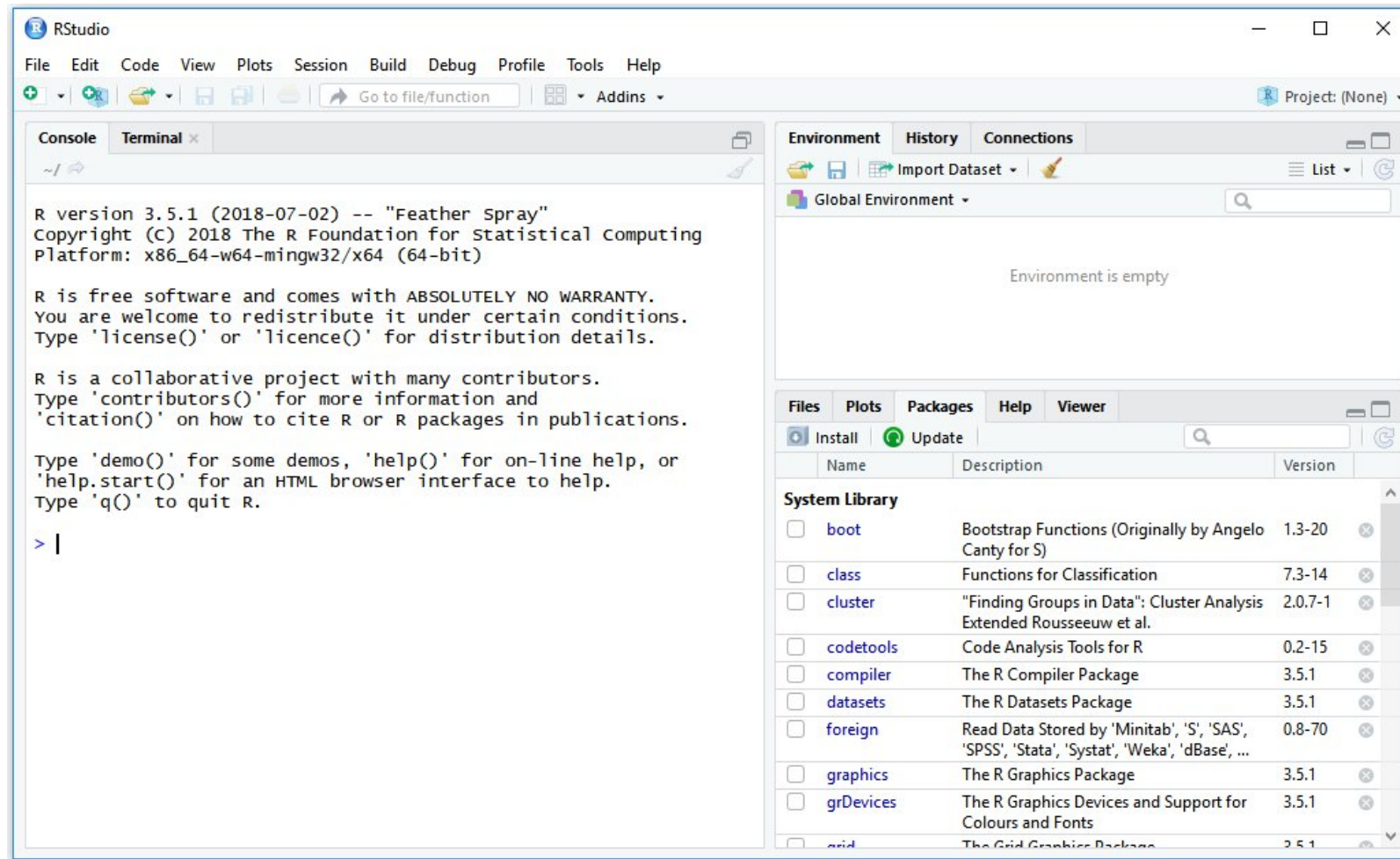
Packages

Ingebouwde documentatie

Beschrijvende statistiek

NA waarden

WERKEN MET RSTUDIO



DATASTRUCTUREN

Vector	<- lijst van 1 datatype
Matrix	<- 2d-lijst van 1 datatype
Factor	<- categorische lijst (nominaal/ordinaal)
Data Frame	<- 2d-lijst met verschillende datatypen
List	<- lijst met verschillende datatypen

Let op: R begint met tellen bij 1!

PACKAGES

R heeft veel ingebouwde packages

- Datasets
- Base
- Graphs
- Stats
- Etc.

Packages zijn standaard aanroepbaar

Bevatten onder andere functies of gevulde datastructuren

INGEBOUWDE DOCUMENTATIE

?datasets

datasets-package {datasets}

R Documentation

The R Datasets Package

Description

Base R datasets

Details

This package contains a variety of datasets. For a complete list, use `library(help = "datasets")`.

Author(s)

R Core Team and contributors worldwide

Maintainer: R Core Team R-core@r-project.org

[Package *datasets* version 3.5.1 [Index](#)]

INGEBOUWDE DOCUMENTATIE

library(help = "datasets")

```
Information on package 'datasets'

Description:

Package:      datasets
Version:      3.5.1
Priority:      base
Title:        The R Datasets Package
Author:       R Core Team and contributors
              worldwide
Maintainer:   R Core Team <R-core@r-project.org>
Description:  Base R datasets.
License:      Part of R 3.5.1
Built:        R 3.5.1; ; 2018-07-02 09:11:07 UTC;
              windows

Index:

AirPassengers      Monthly Airline Passenger Numbers 1949-1960
BJsales            Sales Data with Leading Indicator
BOD                Biochemical Oxygen Demand
CO2                Carbon Dioxide Uptake in Grass Plants
ChickWeight        Weight versus age of chicks on different diets
DNase              Elisa assay of DNase
EuStockMarkets     Daily Closing Prices of Major European Stock
                  Indices, 1991-1998
Formaldehyde       Determination of Formaldehyde
HairEyeColor       Hair and Eye Color of Statistics Students
Harman23.cor        Harman Example 2.3
Harman74.cor        Harman Example 7.4
Indometh            Pharmacokinetics of Indomethacin
InsectSprays       Effectiveness of Insect Sprays
JohnsonJohnson    Quarterly Earnings per Johnson & Johnson Share
```

INGEBOUWDE DOCUMENTATIE

?stats

stats-package {stats}

R Documentation

The R Stats Package

Description

R statistical functions

Details

This package contains functions for statistical calculations and random number generation.

For a complete list of functions, use `library(help = "stats")`.

Author(s)

R Core Team and contributors worldwide

Maintainer: R Core Team R-core@r-project.org

[Package *stats* version 3.5.1 [Index](#)]

INGEBOUWDE DOCUMENTATIE

library(help = "stats")

Information on package 'stats'

Description:

```
Package:      stats
Version:      3.5.1
Priority:      base
Title:        The R Stats Package
Author:        R Core Team and contributors worldwide
Maintainer:    R Core Team <R-core@r-project.org>
Description:   R statistical functions.
License:       Part of R 3.5.1
Imports:       utils, grDevices, graphics
Suggests:      MASS, Matrix, SuppDists, methods, stats4
NeedsCompilation: yes
Built:         R 3.5.1; x86_64-w64-mingw32; 2018-07-02 09:10:19 UTC;
               windows
```

Index:

.checkMFClasses	Functions to Check the Type of Variables passed to Model Frames
AIC	Akaike's An Information Criterion
ARMAacf	Compute Theoretical ACF for an ARMA Process
ARMAtoMA	Convert ARMA Process to Infinite MA Process
Beta	The Beta Distribution
Binomial	The Binomial Distribution
Box.test	Box-Pierce and Ljung-Box Tests
C	Sets Contrasts for a Factor
Cauchy	The Cauchy Distribution
Chisquare	The (non-central) Chi-Squared Distribution
Distributions	Distributions in the stats package
Exponential	The Exponential Distribution
FDist	The F Distribution
GammaDist	The Gamma Distribution
Geometric	The Geometric Distribution

INGEBOUWDE DOCUMENTATIE

??zoekterm

?term

Library(help="term")

Library(term)

BESCHRIJVENDE STATISTIEK

Zoek dataset ChickWeight op.

Wat voor informatie staat erin?

?ChickWeight

summary(ChickWeight)

gewicht <- ChickWeight\$weight

mean(gewicht)

median(gewicht)

sd(gewicht)

NA WAARDES

Soms heb je geen waarden bij bepaalde metingen.

Deze wil je niet zomaar weggooien, en R gebruikt daar NA (not available) voor.

Voor meer informatie ?NA

Je kunt NA waarden laten verwijderen wanneer je berekeningen wilt doen.

Probeer maar uit:

```
test <- c(1,2,3,4,NA)
```

```
test +1
```

```
is.na(test)
```

```
sum(test)
```

```
sum(na.omit(test))
```

OEFENINGEN OP AIRQUALITY

Wat is het resultaat? Omschrijf wat je krijgt bij de volgende scripts:

Begin met het bestuderen van de package airquality, voordat je de vragen probeert te beantwoorden.

- a) `subset(airquality, Day==6)`
- b) `head(airquality)`
- c) `tail(airquality)`
- d) `subset(airquality, Temp >80, select = c(Ozone, Temp))`
- e) `summary(airquality)`
- f) `tempcelsius <- 5/9*(airquality$Temp-32)`
- g) `is.na(airquality$Ozone)`
- h) `na.omit(airquality$Ozone)`

OEFENINGEN OP AIRQUALITY

Begin met het bestuderen van de package airquality, voordat je de vragen probeert te beantwoorden.

- a) Selecteer de subset, waarvoor de temperatuur onder de 70°F en laat alleen rijnummer, Ozone en Temp zien.
- b) Selecteer de subset, waarvoor Wind>12.0 en laat alleen rijnummer, Ozone en Wind zien.
- c) Selecteer de subset, waarvoor Day = 1 en laat alleen rijnummer, Ozone en Temp en Day zien.
- d) Selecteer de subset, waarvoor Day = 1 en laat de kolom Temp weg.
- e) Selecteer de kolommen Ozone tot en met Wind.
- f) Selecteer regel 8 tot en met 14 (week 2) uit het dataframe.
- g) Selecteer alle dagen uit het dataframe waarop Ozone de waarde NA heeft.
- h) Selecteer de zondagen uit het dataframe.

VERANTWOORDING

In deze uitgave is géén auteursrechtelijk beschermd werk opgenomen

Alle teksten © Gonny Henkes tenzij expliciet externe bronnen zijn
aangegeven

Screenshots op basis van vernoemde websites

Eventuele images zijn opgenomen met vermelding van bron