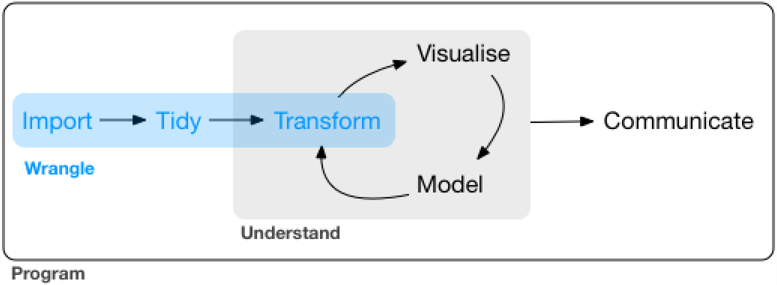
Intro to R for Datascience

DPV

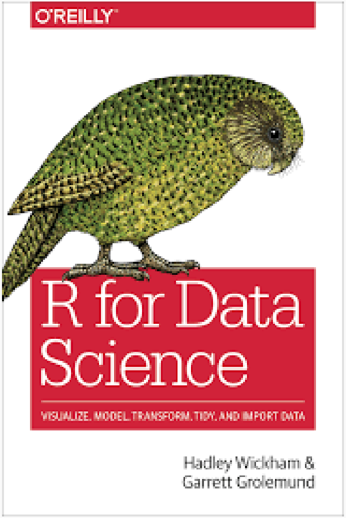
Karin Groothuis-Oudshoorn

sep 5, 2019

R for Data Science



R for Data Science boek



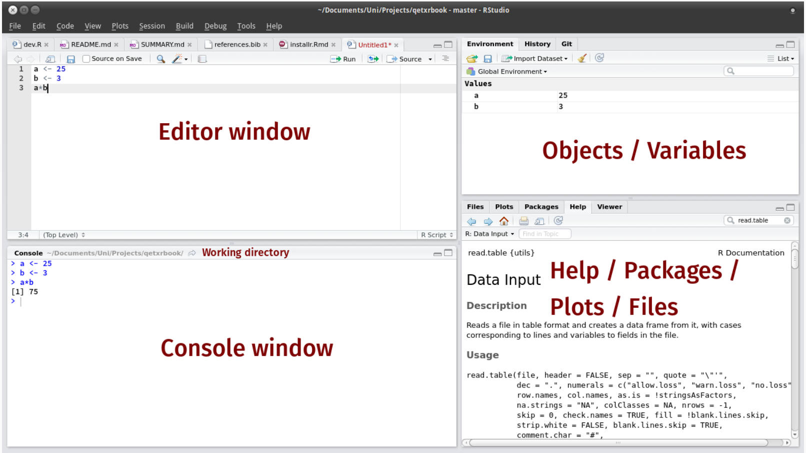
book online: http://r4ds.had.co.nz/

R en RStudio

R: the engine

RStudio: Dashboard

RStudio



```
a <- 25
b <- 3
a*b
```

Editor window

Values

a	25
b	3

Objects / Variables

Working directory

```
> a <- 25
> b <- 3
> a*b
[1] 75
>
```

Console window

read.table {utils}   R Documentation

Data Input

Description

Reads a file in table format and creates a data frame from it, with cases corresponding to lines and variables to fields in the file.

Usage

```
read.table(file, header = FALSE, sep = "", quote = "\"'",
           dec = ".", numerals = c("allow.loss", "warn.loss", "no.loss"
           row.names, col.names, as.is = !stringsAsFactors,
           na.strings = "NA", colClasses = NA, nrows = -1,
           skip = 0, check.names = TRUE, fill = !blank.lines.skip,
           strip.white = FALSE, blank.lines.skip = TRUE,
           comment.char = "#",
```

Help / Packages / Plots / Files

R and R packages

R: a new phone

R packages: Apps that you can download

Installing and using packages

1. Install package: only once
2. Lading package: every session
3. Reinstall package: if you update R
4. List with default packages:

```r
## which packages are default loaded?
search()
```

```
## [1] ".GlobalEnv"        "package:stats"     "package:graphics"
## [4] "package:grDevices" "package:utils"     "package:datasets"
## [7] "package:methods"   "Autoloads"         "package:base"
```

Installing packages

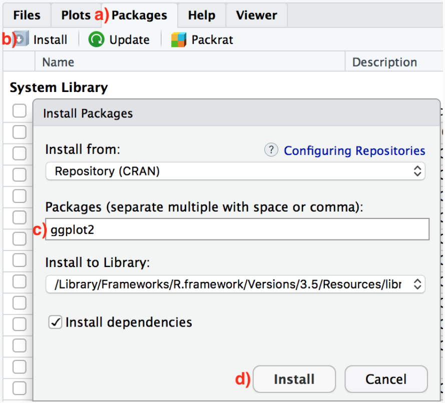
The easy way: in the lower right panel of RStudio:

a) Click on the 'Packages' tab

b) Click on 'Install'

c) Type the name of the package under 'Packages'

d) Click on 'Install'



Install from: Repository (CRAN)

Packages (separate multiple with space or comma): ggplot2

Install to Library: /Library/Frameworks/R.framework/Versions/3.5/Resources/libr

☑ Install dependencies

Tidyverse

- Developer: Hadley Wickham (van RStudio)
- Collection of packages: `dplyr`, `ggplot2`, `tibble`, `readr`, `tidyr`, `purrr`, `stringr`, `forcats`
- More consistent than standard R
- A good starting point to learn R
- Is not standard R!
- Webpage: http://www.tidyverse.org

```
install.packages(tidyverse)
library(tidyverse)
```

R is a program

- Created by Ross Ihaka and Robert Gentleman;
- Inspired by S, statistical programming languague developed by John Chambers, Rick Becker and Allan Wilks from Bell Laboratories.
- First stable beta version in 2000;
- Originally developed for statistics and data science.

R as calculator

```
3 + 4

sqrt(8)

sequence <- c(1,3,5,7)
another_sequence <- seq(from = 1, to = 10, by = 0.5)

random_sequence <- runif(n = 10)
```

Example R-script

```
library(tidyverse) # package
library(ggplot2) # package

# create a table in R and name it "auto"
car <- tibble(
    speed = c(33.0, 33.0, 49.1, 65.2, 78.5, 93.0),
    stop_distance = c(4.69, 4.05, 10.3, 22.3, 34.4, 43.5))

# plot the data and draw a line
ggplot(data = car, aes(x = speed, y = stop_distance)) +
    geom_point() +
    geom_smooth(method = "lm", se = FALSE)

model <- lm(stop_distance ~ speed, data = car)
model
```

Characteristics R statements

Functions

- You work with functions in statements
- Function has a name, arguments and value(s)
- The arguments are between brackets: ‘(’and’)’.

Objects

- *object* contains result of a statement
- *object* can be used in new statements
- *object* is made by assignment operator <− (short cut “Alt -”).

Name object

- Name *object* starts with a letter;
- Name *object* contains only letters, numbers, _ and .;
- Captitals and small letters are **NOT** the same (case sensitive)!;
- Look out for type errors!

Projects

Data-analysis

- Data-analysis is about data and scripts;

RStudio project

- Directory on your hard disk;
- Place scripts and data in project directory;
- Projectmap is *working directory* for R;
- RStudio saves default the files in this project directory.

Advice

- Distinguish between R-scripts, data files and output files, graphics by using subdirectories;
- Use logical names for files, objects and variables;
- Scripts are more important than the output (reproducibility);
- Result = Data files + Scripts.
- Do not save your working directory at the end of a session!

Information on the web

- https://www.r-project.org.org
- https://rweekly.org
- https://www.r-bloggers.com
- https://stackoverflow.com
- cheatsheets