loops and functions in R

Berry Boessenkool, berry-b@gmx.de Jannes Breier, jbreier@gfz-potsdam.de

gitext URL

These slides and tasks are a subset of Berry's teaching material at github.com/brry/course

These slides are licenced under so you can use the material freely as long as you cite us.

R installation instructions: github.com/brry/course#install

PDF created on 2019-11-18, 11:07

TOC 1/27 4 🗇 github.com/brry/course

Outline

For loops Functions Debugging

 $\bullet \ \mathsf{Berry} \ \mathsf{Boessenkool} \to \mathsf{Geoecology} \ \mathsf{@} \ \mathsf{Potsdam} \ \mathsf{University}$

- $\bullet \ \mathsf{Berry} \ \mathsf{Boessenkool} \to \mathsf{Geoecology} \ \mathsf{@} \ \mathsf{Potsdam} \ \mathsf{University}$
- R Fan

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic

github.com/brry/course TOC 3/27 🖃 🕨

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic since 2010

github.com/brry/course TOC 3/27 4 🗇 🕨

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic since 2010
- Developer of rdwd,

github.com/brry/course TOC 3/27 4 🗇 🕨

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic since 2010
- Developer of rdwd, Freelance trainer & consultant

github.com/brry/course TOC 3/27 4 🗇 🕨

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic since 2010
- Developer of rdwd, Freelance trainer & consultant

ullet Jannes Breier o Geoecology @ Potsdam University

github.com/brry/course TOC 3/ 27 🖽 🗇 🕨

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic since 2010
- Developer of rdwd, Freelance trainer & consultant
- Jannes Breier → Geoecology @ Potsdam University
- Berry taught me R in 2013 ©

github.com/brry/course TOC 3/27 🖅

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic since 2010
- Developer of rdwd, Freelance trainer & consultant
- Jannes Breier → Geoecology @ Potsdam University
- Berry taught me R in 2013 ©
- in Research Software Engineer at GFZ, Sec.4.4: Hydrology

github.com/brry/course TOC 3/27 🐇

- ullet Berry Boessenkool o Geoecology @ Potsdam University
- R Fanatic since 2010
- Developer of rdwd, Freelance trainer & consultant
- Jannes Breier → Geoecology @ Potsdam University
- Berry taught me R in 2013 ©
- in Research Software Engineer at GFZ, Sec.4.4: Hydrology

• If we're proceeding too fast, please interrupt!

github.com/brry/course TOC 3/27 🖅

Outline

For loops

Functions Debugging

Execute a block of code several times, with different input values.

Syntax: for(aRunningVariable in aSequence){ doSomething }

github.com/brry/course TOC 5/ 27 🖽 🗇 🕨

Execute a block of code several times, with different input values.

Syntax: for(aRunningVariable in aSequence){ doSomething }

Often, i (for index) is used, thus for(i in 1:n) doThis(i)

github.com/brry/course TOC 5/ 27 🖽 🤊 🕨

```
Execute a block of code several times, with different input values.
Syntax: for(aRunningVariable in aSequence) { doSomething }
Often, i (for index) is used, thus for(i in 1:n) doThis(i)
```

```
help("for") # needs quotation marks!
```

github.com/brry/course TOC 5/ 27 🖽 🗇 🕨

```
Execute a block of code several times, with different input values.

Syntax: for(aRunningVariable in aSequence){ doSomething }

Often, i (for index) is used, thus for(i in 1:n) doThis(i)
```

```
help("for") # needs quotation marks!
```

```
print(1:2)
print(1:5)
print(1:9)
```

github.com/brry/course TOC 5/ 27 🖽 🗇 🕨

```
Execute a block of code several times, with different input values.

Syntax: for(aRunningVariable in aSequence) { doSomething }

Often, i (for index) is used, thus for(i in 1:n) doThis(i)
```

```
help("for") # needs quotation marks!
```

```
print(1:2)
print(1:5)
print(1:9)
```

This is easier and less prone to human errors with:

Execute a block of code several times, with different input values.

Syntax: for(aRunningVariable in aSequence) { doSomething }

Often, i (for index) is used, thus for(i in 1:n) doThis(i)

```
help("for") # needs quotation marks!
```

```
print(1:2)
print(1:5)
print(1:9)
```

This is easier and less prone to human errors with:

```
for(i in c(2,5,9) ) { print(1:i) }

## [1] 1 2
## [1] 1 2 3 4 5
## [1] 1 2 3 4 5 6 7 8 9

github.com/brry/course
TOC 5/27 4 5 9
```

```
v <- vector(mode="numeric", length=20)
v
## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</pre>
```

for(i in 3:17) { v[i] <- (i+2)^2 }</pre>

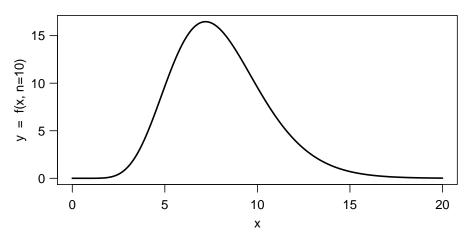
```
v <- vector(mode="numeric", length=20)</pre>
V
   for(i in 3:17) { v[i] \leftarrow (i+2)^2 }
v # this code was executed once for each i
##
   [1] 0 0 25 36 49 64 81 100 121
  [10] 144 169 196 225 256 289 324 361
##
  [19]
        0
```

```
v <- vector(mode="numeric", length=20)</pre>
V
   ##
for(i in 3:17) { v[i] \leftarrow (i+2)^2 }
v # this code was executed once for each i
   [1] 0 0 25 36 49 64 81 100 121
  [10] 144 169 196 225 256 289 324 361
  [19]
        0
```

In R, for loops are slow. Always try to vectorize (the best option, not always possible) or use lapply (saves you the initiation of the empty vector, easier to parallize).

For loops: execute code multiple times I

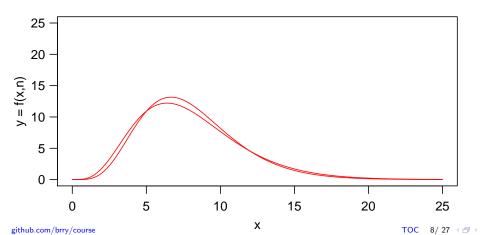
$$y = f(x, n) = \frac{12.5 * n}{(n-1)!} * (\frac{nx}{8})^{(n-1)} * e^{-\frac{nx}{8}}$$



github.com/brry/course TOC 7/27 ← 🗇 ト

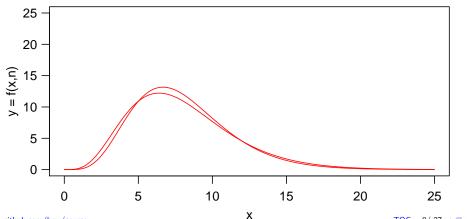
For loops: execute code multiple times II

```
x <- seq(0,25,0.1)
plot(x,x, type="n", ylab="y = f(x,n)")
lines(x, 12.5*5/factorial(5-1)*(x/8*5)^(5-1)*exp(-x/8*5), col=2)
lines(x, 12.5*6/factorial(6-1)*(x/8*6)^(6-1)*exp(-x/8*6), col=2)</pre>
```



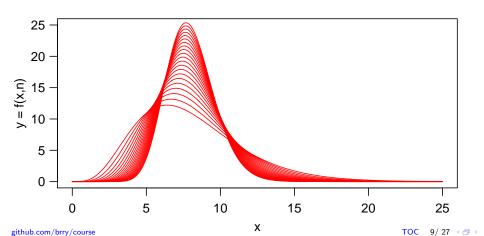
For loops: execute code multiple times II

```
x <- seq(0,25,0.1)
plot(x,x, type="n", ylab="y = f(x,n)")
lines(x, 12.5*5/factorial(5-1)*(x/8*5)^(5-1)*exp(-x/8*5), col=2)
lines(x, 12.5*6/factorial(6-1)*(x/8*6)^(6-1)*exp(-x/8*6), col=2)</pre>
```



For loops: execute code multiple times III

```
x <- seq(0,25,0.1)
plot(x,x, type="n", ylab="y = f(x,n)")
for (n in 5:25)
lines(x, 12.5*n/factorial(n-1)*(x/8*n)^(n-1)*exp(-x/8*n), col=2)</pre>
```

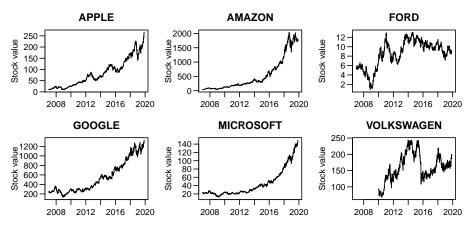


For loops exercise

Exercise 1: for loop

- Read stocks.txt (rightclick Raw, save as), so that there are no factors in the data.frame
- Ohange the first column type from char to date with ?as.Date
- What do you get with plot(stocks[,c(1,2)])? Make it a line graph by setting the argument type.
- With a for loop, plot each stock time series, i.e. plot the ith column over the first column.
- BONUS 1: Use good annotations (main, ylab, xlab)
- BONUS 2: Turn y axis labels upright (las)
- O BONUS 3: With par(mfrow..., set up a two by three panel plot

For loops exercise solution



lapply: apply a function to elements of a list (or vector)

list: R object containing other objects

lapply: apply a function to elements of a list (or vector) list: R object containing other objects

```
files <- dir("../rawdata", pattern="*.csv", full=TRUE)
```

```
lapply: apply a function to elements of a list (or vector)
list: R object containing other objects
files <- dir("../rawdata", pattern="*.csv", full=TRUE)

# bad and slow way:
ldfs <- list() # initiate empty list
for(i in 1:length(files))
   ldfs[[i]] <- read.csv(files[i], as.is=TRUE)</pre>
```

github.com/brry/course TOC 12/ 27 🖽 🗈

```
lapply: apply a function to elements of a list (or vector)
list: R object containing other objects
files <- dir("../rawdata", pattern="*.csv", full=TRUE)

# bad and slow way:
ldfs <- list() # initiate empty list
for(i in 1:length(files))
   ldfs[[i]] <- read.csv(files[i], as.is=TRUE)</pre>
```

```
# much better way: apply function to each file
ldfs <- lapply(X=files, FUN=read.csv, as.is=TRUE)</pre>
```

github.com/brry/course TOC 12/27 🖅

```
# bad and slow way:
ldfs <- list() # initiate empty list</pre>
for(i in 1:length(files))
   ldfs[[i]] <- read.csv(files[i], as.is=TRUE)</pre>
# much better way: apply function to each file
ldfs <- lapply(X=files, FUN=read.csv, as.is=TRUE)</pre>
# progress bar with remaining time (+ parallelized!)
library("pbapply")
ldfs <- pblapply(X=files, FUN=read.csv, as.is=TRUE)</pre>
ldfs <- pblapply(X=files, FUN=read.csv, as.is=TRUE, cl=8)</pre>
                                                         TOC 12/27 4 🗇 🕨
github.com/brry/course
```

lapply: apply a function to elements of a list (or vector)

files <- dir("../rawdata", pattern="*.csv", full=TRUE)

list: R object containing other objects

Outline

For loops

Functions

Debugging

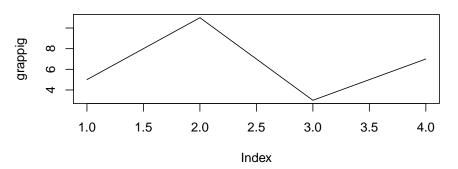
Functions I

http://r4ds.had.co.nz/functions.html

http://r4ds.had.co.nz/functions.html
?"function"

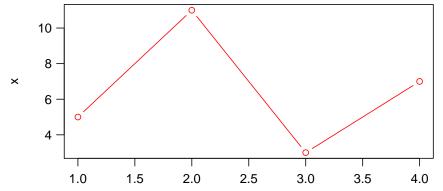
After return()ing, the execution of the function is terminated, so it should only be positioned at the end. It can also be left away, the last instruction ("expression") will then be returned.

```
myfunct( c(5,11,3,7) )
## [1] 35 77 21 49
```



Functions with more arguments + default values

```
myfunct <- function(x, type="b", ...) plot(x, type=type, ...)
# type="o" is now the default, thus used unless specified
# The ellipsis (...) passes arguments to other functions
myfunct( c(5,11,3,7) , col="red", las=1)</pre>
```



github.com/brry/course

16/27 < 🗗 ト

TOC

Syntax: if(this_is_true) {do_something}

```
Syntax: if(this_is_true) {do_something}
if(this_is_true) {do_something} else {do_other_thing}
```

```
Syntax: if(this_is_true) {do_something}
if(this_is_true) {do_something} else {do_other_thing}

If condition == TRUE, then expression1 is evaluated,
if condition == FALSE, then expression2 is evaluated.
```

```
Syntax: if(this_is_true) {do_something}
if(this_is_true) {do_something} else {do_other_thing}

If condition == TRUE, then expression1 is evaluated,
if condition == FALSE, then expression2 is evaluated.

7-3 > 2
class(7-3 > 2)
if(7-3 > 2) 18
if(7-3 > 5) 18
else 17
```

github.com/brry/course TOC 17/27 🖽 🗇 🕨

```
Syntax: if(this_is_true) {do_something}
if(this_is_true) {do_something} else {do_other_thing}
If condition == TRUE, then expression 1 is evaluated,
if condition == FALSE, then expression2 is evaluated.
7-3 > 2
                            TRUE
class(7-3 > 2)
                            logical = truth value, boolean
if(7-3 > 2) 18
                            Condition is TRUE, so 18 is returned
if(7-3 > 5) 18
                            Condition is FALSE, so nothing happens
if(7-3 > 5) 18 else 17 Condition FALSE, so 17 is returned.
if(length(input)>1)
  stop("length must be 1, not ", length(input))
```

warning: continues but gives warning

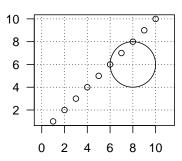
message: to inform instead of worry the user

```
Syntax: if(this_is_true) {do_something}
if(this_is_true) {do_something} else {do_other_thing}
If condition == TRUE, then expression 1 is evaluated,
if condition == FALSE, then expression2 is evaluated.
7-3 > 2
                            TRUE
class(7-3 > 2)
                            logical = truth value, boolean
if(7-3 > 2) 18
                            Condition is TRUE, so 18 is returned
if(7-3 > 5) 18
                            Condition is FALSE, so nothing happens
if(7-3 > 5) 18 else 17 Condition FALSE, so 17 is returned.
if(length(input)>1)
  stop("length must be 1, not ", length(input))
stop: Interrupts function execution and gives error
```

github.com/brry/course TOC 17/27 🗗 🕨

Exercise: add circles with given radius

```
plot(1:10, asp=1) # aspect ratio y/x of graph range
grid(col=1)
# the next part sould go into a function:
x <- 8; y <- 6; r <-2
p <- seq(0, 2*pi, len=50)
cx <- x+r*cos(p); cy <- y+r*sin(p)
polygon(cx, cy)</pre>
```



Time to practice programming

Exercise 2: Writing functions

Write a function that

- draws a circle with a certain radius at user-specified locations of an existing plot (see last slide).
- ② uses ellipsis (...) to allow the user to customize the appearance
- checks all the arguments and gives useful warnings if the wrong type of input is provided
- has useful explanations for each argument (documentation)
- a has readable indentation, spacing and comments explaining the code
- Now let your neighbor use it without explaining how it is to be used (this should be inferred from the code and comments!)
- Use your neighbor's function with a vector to draw several circles at once. (unintended use?) What happens?
- **3** BONUS: Learn writing packages at packdev.R (rightclick Raw, save as)

github.com/brry/course TOC 19/ 27 < 🗇

```
# Small helper function drawing circles into existing graphics
circle <- function(
    x, # x-coordinate of points, numeric value of length 1
    y, # ditto for y
    r, # radius of the circle, in the graphic's units
    locnum=100, # number of points on circle (more means smoother but slower)
    ...) # Further Arguments passed to polygon, like col, border, lwd
{
#
```

```
# Small helper function drawing circles into existing graphics
circle <- function(
 x, # x-coordinate of points, numeric value of length 1
 v. # ditto for u
 r, # radius of the circle, in the graphic's units
 locnum=100, # number of points on circle (more means smoother but slower)
  ...) # Further Arguments passed to polygon, like col, border, lwd
 # input checking - only one circle can be drawn:
 if(length(x) >1 | length(y) >1 | length(r) >1 | length(locnum) >1)
   warning("Only the first element of the vectors is used.")
   x \leftarrow x[1]; y \leftarrow y[1]; r \leftarrow r[1]; locnum \leftarrow locnum[1]
```

#

```
# Small helper function drawing circles into existing graphics
circle <- function(
 x, # x-coordinate of points, numeric value of length 1
 v. # ditto for u
 r, # radius of the circle, in the graphic's units
 locnum=100, # number of points on circle (more means smoother but slower)
  ...) # Further Arguments passed to polygon, like col, border, lwd
 # input checking - only one circle can be drawn:
 if(length(x) >1 | length(y) >1 | length(r) >1 | length(locnum) >1)
   warning("Only the first element of the vectors is used.")
   x \leftarrow x[1]; y \leftarrow y[1]; r \leftarrow r[1]; locnum \leftarrow locnum[1]
if(!is.numeric(x)) stop("x must be numeric, not ", class(x))
if(!is.numeric(y)) stop("y must be numeric, not ", class(y))
if(!is.numeric(r)) stop("r must be numeric, not ", class(r))
 #
```

```
# Small helper function drawing circles into existing graphics
circle <- function(
 x, # x-coordinate of points, numeric value of length 1
 v. # ditto for u
 r, # radius of the circle, in the graphic's units
 locnum=100, # number of points on circle (more means smoother but slower)
  ...) # Further Arguments passed to polygon, like col, border, lwd
 # input checking - only one circle can be drawn:
 if(length(x) >1 | length(y) >1 | length(r) >1 | length(locnum) >1)
   warning("Only the first element of the vectors is used.")
   x \leftarrow x[1]; y \leftarrow y[1]; r \leftarrow r[1]; locnum \leftarrow locnum[1]
if(!is.numeric(x)) stop("x must be numeric, not ", class(x))
if(!is.numeric(y)) stop("y must be numeric, not ", class(y))
if(!is.numeric(r)) stop("r must be numeric, not ", class(r))
 # prepare circle line coordinates:
cx \leftarrow x+r*cos(seq(0,2*pi,len=locnum))
cy <- y+r*sin( seq(0,2*pi,len=locnum) )</pre>
polygon(cx, cy, ...) # actually draw it
# Note: if circles look like ellipsis, use plot(... asp=1)
```

Solution for exercise 2 II: functions

github.com/brry/course

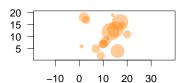
```
plot(1:20, type="n", asp=1, cex=2)
circle(5,5, r=3)
circle(15,10, r=4, locnum=12, col=2, border=4, lwd=3)
  20
  10
        -10
                                               20
                                                            30
                                                           21/27 4 🗇 🕨
```

Solution for exercise 2 III: functions

```
# can not be vectorized:
x <- sample(1:20, 15); y <- sample(1:20, 15); r <- runif(20)*4
circle(x,y,r, col=rgb(1,0.5,0,alpha=0.4), border=NA)

## Warning in circle(x, y, r, col = rgb(1, 0.5, 0, alpha = 0.4), border = NA):
Only the first element of the vectors is used.

for(i in 1:15) circle(x[i],y[i],r[i], col=rgb(1,0.5,0,alpha=0.4), border=NA)</pre>
```



github.com/brry/course TOC 22/27 🖅

Outline

For loops Functions

Debugging

Debugging

Your code throws an error. You didn't call the mentioned function.
 Obviously, your code calls some function calling some function calling some function calling [you get the idea] which in the end creates an error. To trace back this path, you can use traceback().

github.com/brry/course TOC 24/27 🖽

Debugging

- Your code throws an error. You didn't call the mentioned function.
 Obviously, your code calls some function calling some function calling some function calling [you get the idea] which in the end creates an error. To trace back this path, you can use traceback().
- Now that you know where the error originates from, you set options(error=recover). You run your code again, but this time R waits at the level creating an error. You examine the environment within the function, play around with the objects and internal function code, until the bug has been fixed. You have just debugged a function.

github.com/brry/course TOC 24/27 🐇

Debugging

- Your code throws an error. You didn't call the mentioned function.
 Obviously, your code calls some function calling some function calling some function calling [you get the idea] which in the end creates an error. To trace back this path, you can use traceback().
- Now that you know where the error originates from, you set options(error=recover). You run your code again, but this time R waits at the level creating an error. You examine the environment within the function, play around with the objects and internal function code, until the bug has been fixed. You have just debugged a function.
- You want to step into the function you are developing at a specific point. You add browser() at that point of the code.

github.com/brry/course TOC 24/27 🖽

source("projectFuns.R") execute complete file

```
source("projectFuns.R") execute complete file find error source in sequence of function calls
```

source("projectFuns.R")
traceback()
options(warn=2)

execute complete file find error source in sequence of function calls warnings to error. default $\boldsymbol{0}$

```
source("projectFuns.R")
traceback()
options(warn=2)
browser()
```

execute complete file find error source in sequence of function calls warnings to error. default 0 go into function environment: n, s, f, c, Q

```
source("projectFuns.R")
traceback()
options(warn=2)
browser()
options(error=recover)
```

execute complete file find error source in sequence of function calls warnings to error. default 0 go into function environment: n, s, f, c, Q open interactive session where error occurred

```
source("projectFuns.R")
traceback()
options(warn=2)
browser()
options(error=recover)
debug(funct)
```

execute complete file find error source in sequence of function calls warnings to error. default 0 go into function environment: n, s, f, c, Q open interactive session where error occurred toggle linewise function execution

```
source("projectFuns.R")
traceback()
options(warn=2)
browser()
options(error=recover)
debug(funct)
undebug(funct)
```

execute complete file find error source in sequence of function calls warnings to error. default 0 go into function environment: n, s, f, c, Q open interactive session where error occurred toggle linewise function execution after calling and fixing funct

```
source("projectFuns.R")
traceback()
options(warn=2)
browser()
options(error=recover)
debug(funct)
undebug(funct)
```

execute complete file find error source in sequence of function calls warnings to error. default 0 go into function environment: n, s, f, c, Q open interactive session where error occurred toggle linewise function execution after calling and fixing funct

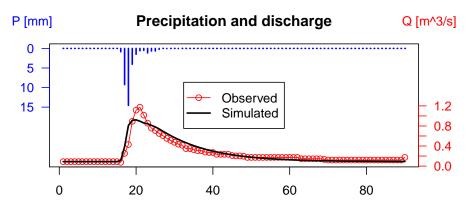
```
R. Peng (2002): Interactive Debugging Tools in R
D. Murdoch (2010): Debugging in R
H. Wickham (2015): Advanced R: debugging
Example: Pete Werner Blog Post (2013)
```

github.com/brry/course TOC 25/27 4 🗇 🕨

Practice debugging, Isc_functions.R (rightclick Raw, save as)

Exercise 3: Debugging

- Load your package and the datasets. Correct the functions until lsc(calib\$P, calib\$Q, area=1.6) returns the result below.
- BONUS: commit each change to git.



github.com/brry/course

TOC 26/27 4 🗇 🕨

Solution for exercise 3: Debugging

- stupid error you can easily remove traceback find location of error lsc#73 just comment it out
- harder to find but still stupid traceback nse#11 ditto
- Error in plot: need finite 'ylim' value browser/options(error=recover) - lsc#105 - NAs in Q - range(Q, na.rm=TRUE) - also in other applicable locations
- There were 50 or more warnings come from rmse being called in optimization - add argument quietNA (or similar) to lsc that is passed to rmse in lsc#79

github.com/brry/course TOC 27/27 🐇 TOC 27/27