(Installation and first steps)

INTRODUCTION TO R

Created: 2021-07-08 Do 20:38

- Why are we using R?
- Getting in/out of R
- Installing R on Windows and Mac
- R Packages and libraries

WHY WE ARE USING R

Programming Language	2021	2016	2011	2006	2001	1996	1991	1986
C	1	2	2	2	1	1	1	1
Java	2	1	1	1	3	28	-	-
Python	3	5	6	7	23	16	-	-
C++	4	3	3	3	2	2	2	8
C#	5	4	5	6	9	-	-	-
JavaScript	6	7	9	9	6	30	-	-
PHP	7	6	4	4	20	-	-	-
R	8	14	35	-	-	-	-	-
SQL	9	-	-	-	-	-	-	-
Go	10	56	15	-	-	-	-	-
Perl	14	8	7	5	4	3	-	-
Lisp	32	23	12	13	16	7	3	2
Ada	34	22	20	15	15	5	9	3

Figure 1: Source: **tiobe.com**

Check some of these out!

OBTAINING AND INSTALLING R FROM CRAN

URL: **https://cran.r-project.org/mirrors.html**

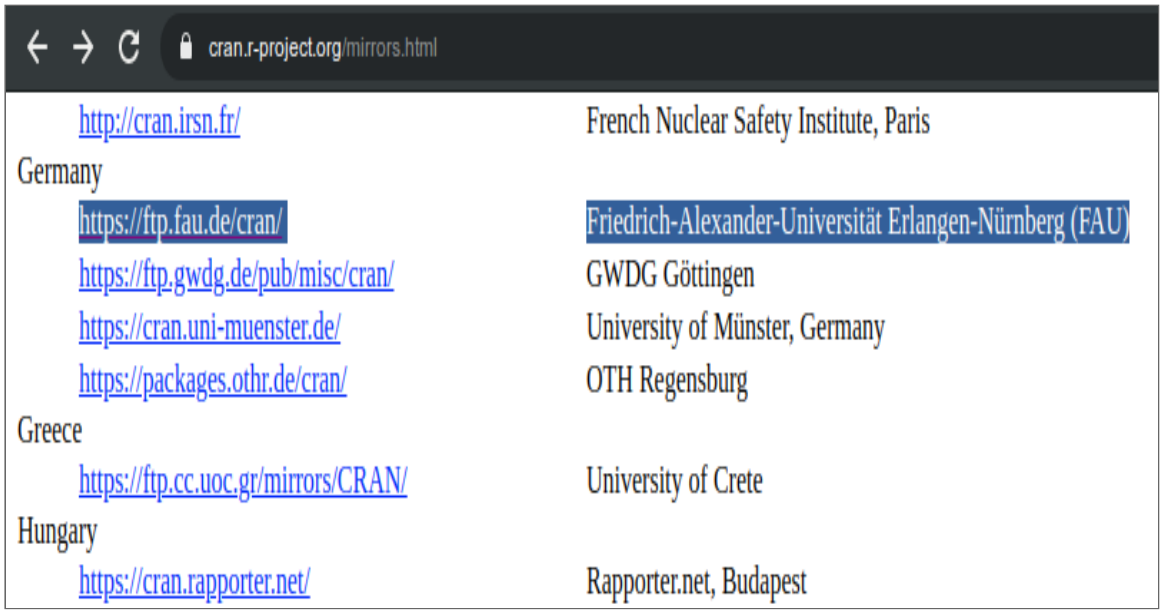
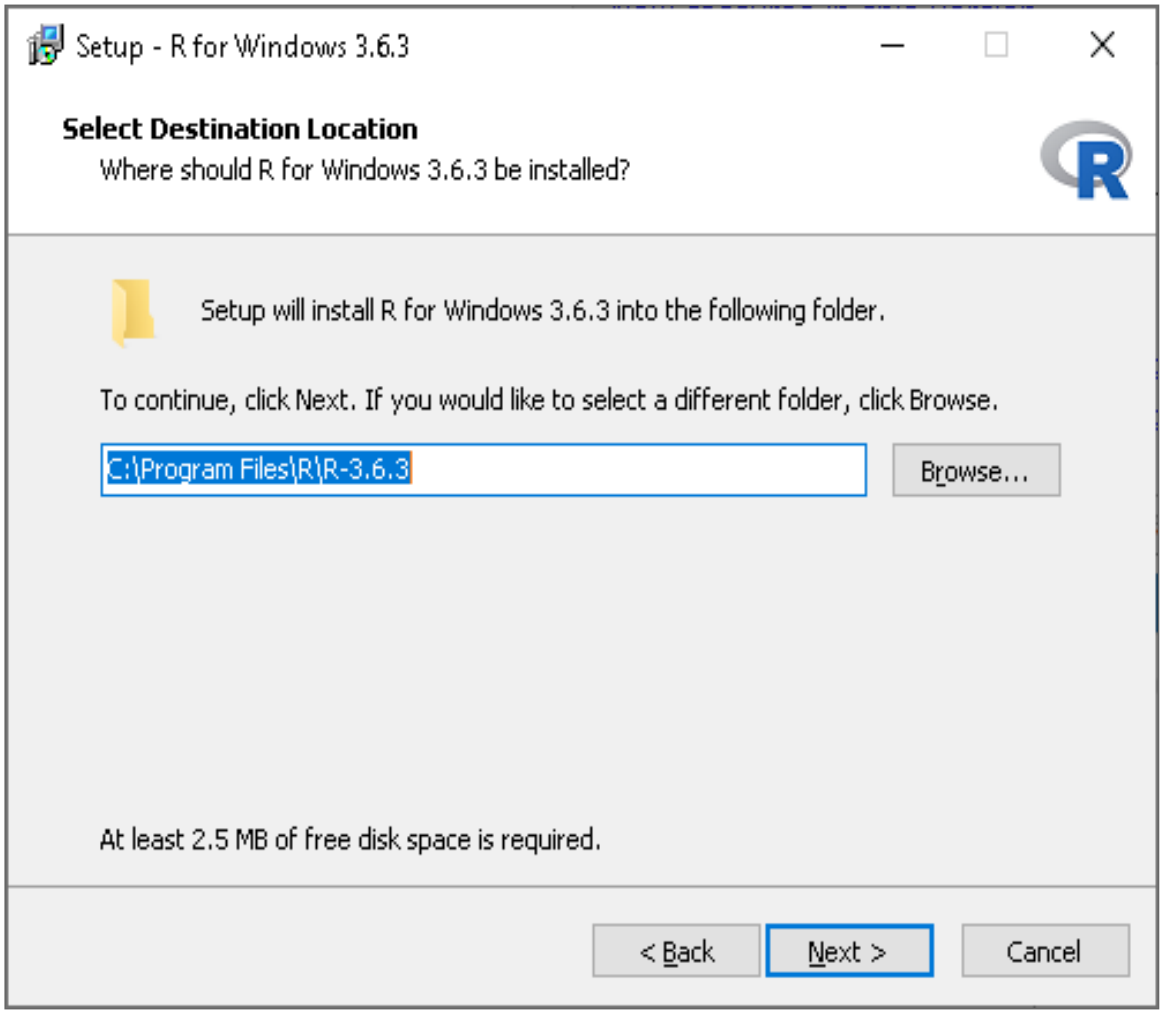
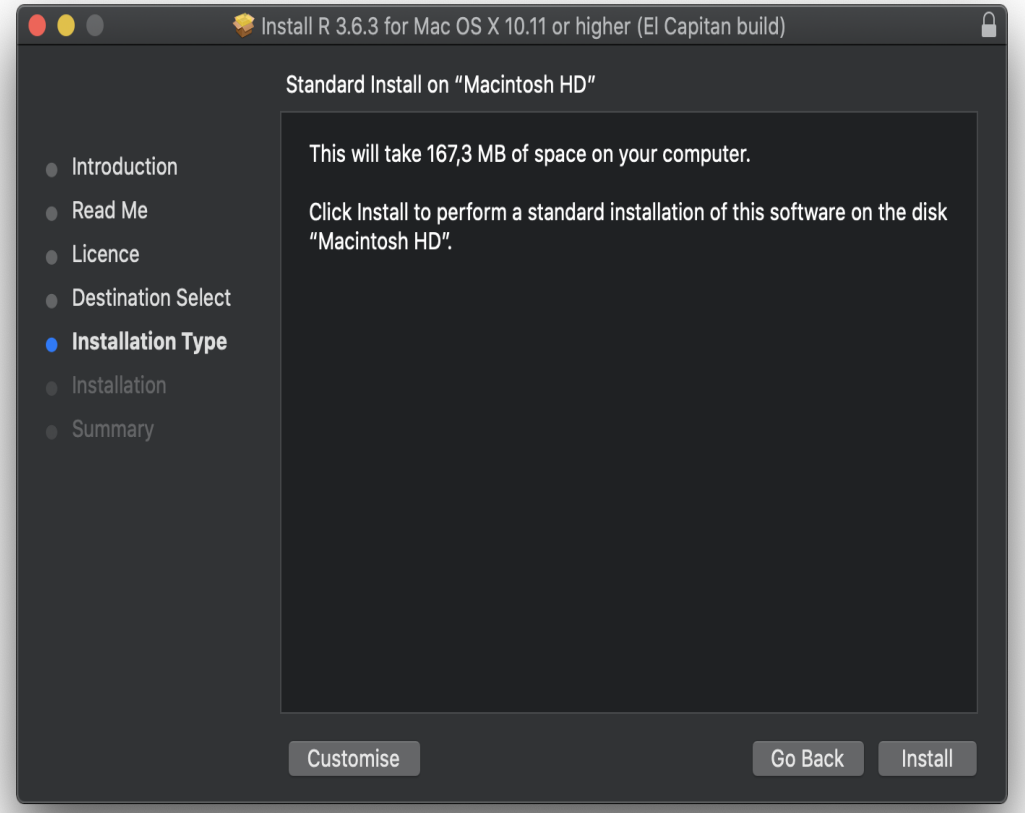


Figure 2: List of mirrors at CRAN
CRAN = Comprehensive R Archive Network

HOW THIS LOOKS IN WINDOWS



HOW THIS LOOKS ON A MAC



INSTALL R NOW

- Breakout room 1: Windows
- Breakout room 2: MacOS
- Max. 30 min - return when done

OPENING R FOR THE FIRST TIME

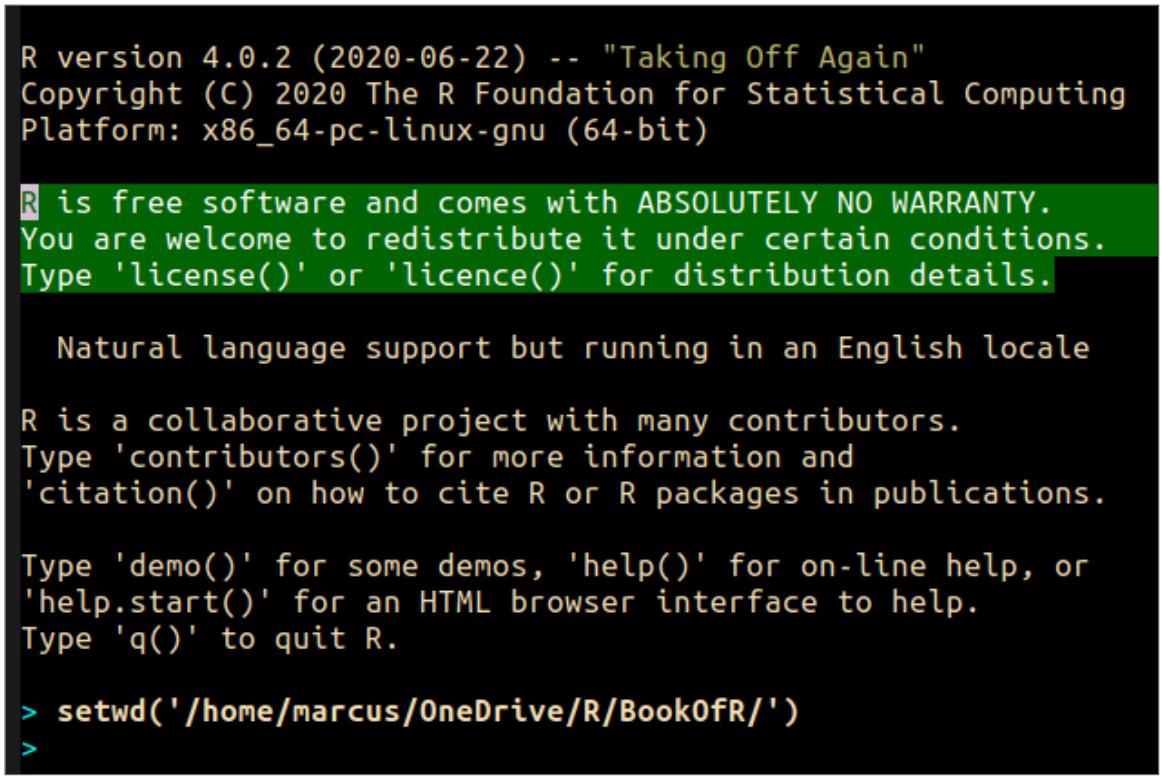
- Go to **https://www.r-project.org**
- Check FAQ and "related projects" pages

VERSION AND PLATFORM

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> setwd('/home/marcus/OneDrive/R/BookOfR/')
>
```

What type of bit-architecture do you have?

DISTRIBUTION LICENSE



Type `license()`. What is "GNU"?

THE R PROJECT

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> setwd('/home/marcus/OneDrive/R/BookOfR/')
>
```

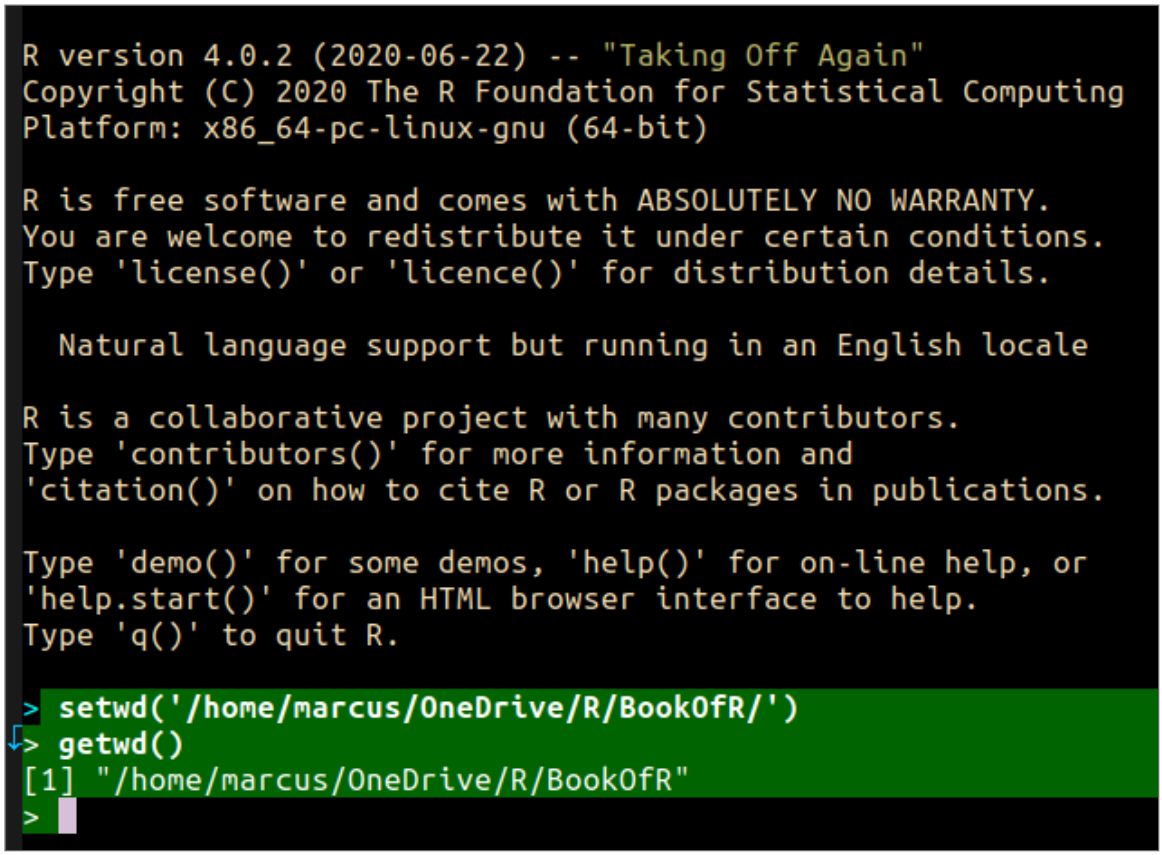
- Enter `citation()`. Why cite software?
- Enter `contributors()`. Who can contribute?

DEMO AND HELP

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> setwd('/home/marcus/OneDrive/R/BookOfR/')
>
```

- Enter `demo(graphics)` and marvel.
- Enter `help.start()` - where is this page?

WORKING DIRECTORY



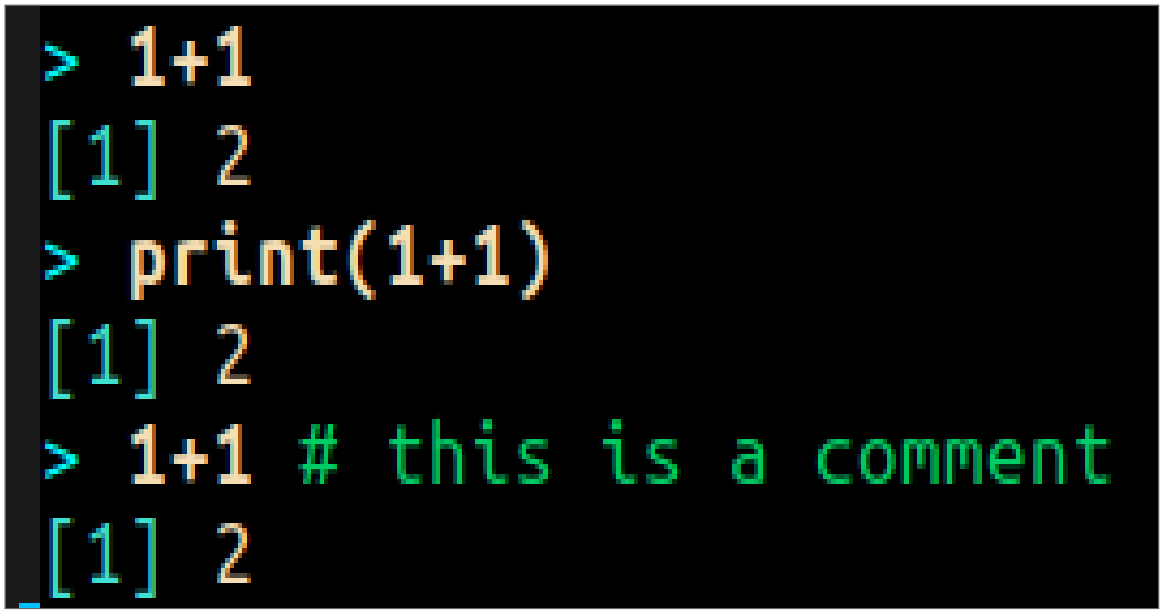
- Enter `getwd()` ("get working dir")
- Use `setwd()` to change directory

R "PROMPT"

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> setwd('/home/marcus/OneDrive/2020_Winter/DS101/2_R_intro/')
> getwd()
[1] "/home/marcus/OneDrive/2020_Winter/DS101/2_R_intro"
> options(prompt="R> ")
R>
```

- Change your prompt to your name
- Change it back to "> "

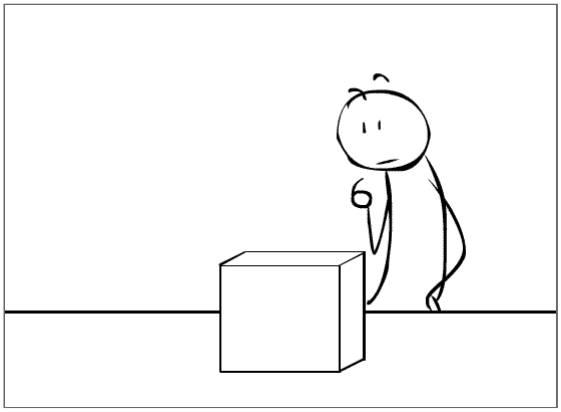
COMPUTING



- Compute "2×2" and print it
- Do it again with a comment (#)

R PACKAGES

- Contain functions and data sets
- Must be installed and loaded for use



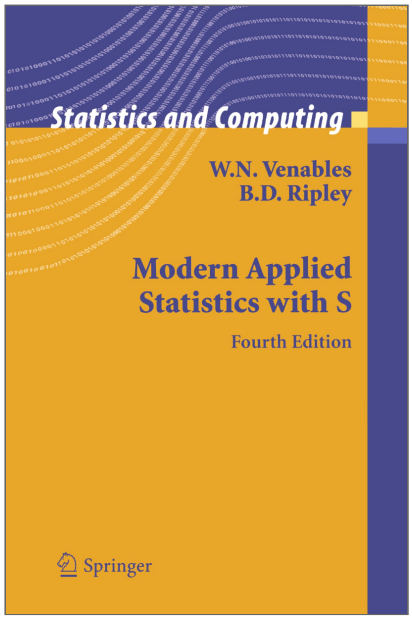
- Can be created with relative ease
- Default data sets: `?datasets`

INSTALL PACKAGES

```
> install.packages("MASS")
Installing package into ‘/home/marcus/R/x86_64-pc-linux-gnu-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'https://ftp.fau.de/cran/src/contrib/MASS_7.3-54.tar.gz'
Content type 'application/x-gzip' length 506246 bytes (494 KB)
==================================================
downloaded 494 KB
```

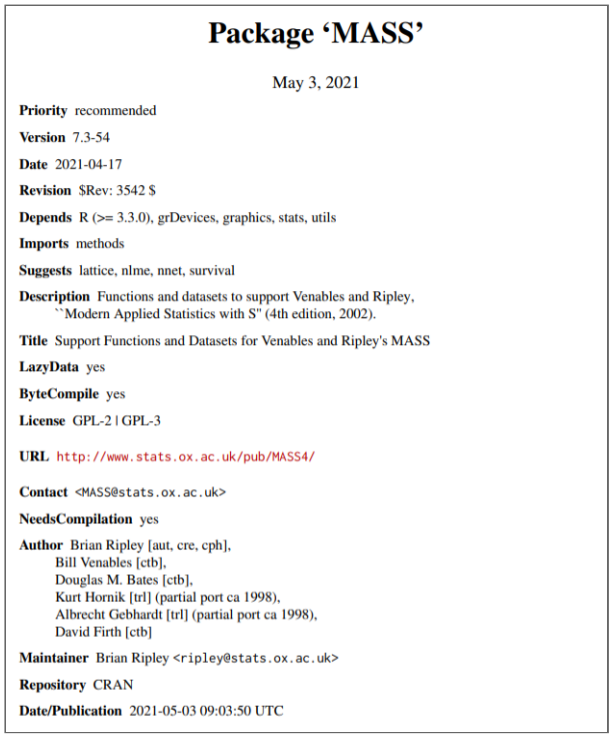
- install package "MASS": enter `install.packages("MASS")`

CHECK DATASETS



- Which datasets are in `MASS`?
- Enter `data(package="MASS")`

LOAD PACKAGE



Package 'MASS'

May 3, 2021

Priority recommended

Version 7.3-54

Date 2021-04-17

Revision $Rev: 3542 $

Depends R (>= 3.3.0), grDevices, graphics, stats, utils

Imports methods

Suggests lattice, nlme, nnet, survival

Description Functions and datasets to support Venables and Ripley, ``Modern Applied Statistics with S'' (4th edition, 2002).

Title Support Functions and Datasets for Venables and Ripley's MASS

LazyData yes

ByteCompile yes

License GPL-2 | GPL-3

URL http://www.stats.ox.ac.uk/pub/MASS4/

Contact <MASS@stats.ox.ac.uk>

NeedsCompilation yes

Author Brian Ripley [aut, cre, cph], Bill Venables [ctb], Douglas M. Bates [ctb], Kurt Hornik [trl] (partial port ca 1998), Albrecht Gebhardt [trl] (partial port ca 1998), David Firth [ctb]

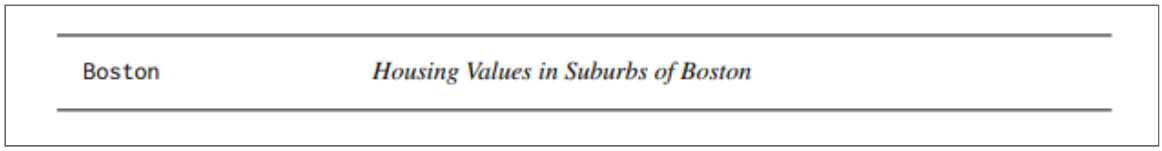
Maintainer Brian Ripley <ripley@stats.ox.ac.uk>

Repository CRAN

Date/Publication 2021-05-03 09:03:50 UTC

- Load MASS in current R session
- Enter `library(MASS)`
- See **documentation** @CRAN

LOAD DATASET



- Load the data set "`Boston`"
- What is in `MASS::Boston`?
- There are **different ways** to find out!

EXPLORE DATASET

```
> str(Boston)
'data.frame':   506 obs. of  14 variables:
 $ crim   : num  0.00632 0.02731 0.02729 0.03237 0.06905 ...
 $ zn     : num  18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...
 $ indus  : num  2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...
 $ chas   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nox    : num  0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...
 $ rm     : num  6.58 6.42 7.18 7 7.15 ...
 $ age    : num  65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...
 $ dis    : num  4.09 4.97 4.97 6.06 6.06 ...
 $ rad    : int  1 2 2 3 3 3 5 5 5 5 ...
 $ tax    : num  296 242 242 222 222 222 311 311 311 311 ...
 $ ptratio: num  15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...
 $ black  : num  397 397 393 395 397 ...
 $ lstat  : num  4.98 9.14 4.03 2.94 5.33 ...
 $ medv   : num  24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...
```

- Print first/last lines: `head()` / `tail()`
- Show structure: `str()`

LEAVING R

```
> q()
Save workspace image? [y/n/c]: y

Process R finished at Thu Jul  8 20:25:09 2021
```

- Leave R with q ( )
- Save your workspace with y
- Check which files were created!

HOUSEKEEPING

```
/home/marcus/OneDrive/2021_Fall/ds101:
total used in directory 13596 available 240.9 GiB
drwxrwxr-x  6 marcus marcus     4096 Jul  8 20:25 .
-rw-rw-r--  1 marcus marcus     2564 Jul  8 20:25 .Rhistory
-rw-rw-r--  1 marcus marcus 12891303 Jul  8 20:25 .RData
```

- Saved R commands: `.Rhistory`
- Saved R variables: `.RData`
- R profile settings: `.Rprofile`
- **Sample profiles**

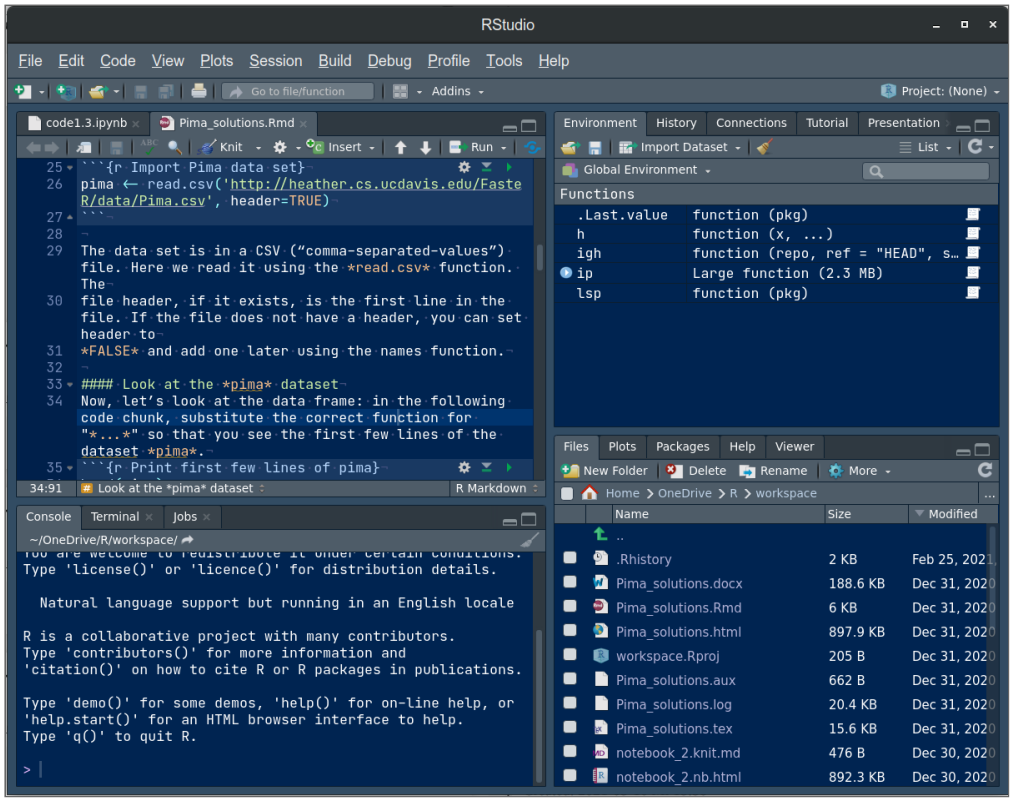
CUSTOMIZE STARTUP

- Create a file `.Rprofile`:

```
options(
repos = c(CRAN = "https://ftp.fau.de/cran/")
)
```

- To check: restart R, re-install MASS

THE RSTUDIO IDE



- Use it at your own **peril**!
- Give **Emacs** + ESS a chance!
- Learn "stick shift" first (=CLI)

CONCEPT SUMMARY

- R is an easy to **learn** language to quickly and interactively analyse datasets. R is especially strong on visualization.
- R can be downloaded from `r-project.org` and installed on your computer.
- There is plenty of **help** on R available from within the program, or on the Internet using the wider community of practitioners.
- When you open R, you establish a working **environment**, which includes packages, functions and variables.

CODE SUMMARY I

`license()`, `licence()`	License info
`help()`, `?help`, `??cars`	get help
`demo()`	R demos
`getwd()`, `setwd()`	get/set working dir
`options(prompt=)`	set prompt
`print(1+1)`	result of 1+1
`quit()`, `q()`	leave R
`# ...`	comment

CODE SUMMARY II

`library("MASS")`	load
`install.packages("MASS")`	install
`installed.packages()`	list all packages
`update.packages()`	update
`packageDescription("MASS")`	describe
`help(package="MASS")`	show
`data()`	built-in datasets

WHAT NEXT?

WHAT NOW? READ!

- Read frequently and widely
- Go both deep and stay shallow

WHAT NOW? PLAY!

Data Scientists Should Learn Through Play

WHAT'S NEXT?

Arithmetic with R:	Just numbers!
Vectors in R:	The workhorse

THANK YOU! QUESTIONS?

REFERENCES

- Adolfo Alvarez (25 Mar 2019). R Packages: A Beginner's Guide. Online: **datacamp.com**.
- Robert Becker (2004). A Brief History of S. Online: **sas.waterloo.ca**.
- Tilman M. Davies (2016). **The Book of R. No Starch Press.**
- Kyle Gallatin (1 Nov 2018). Some Important Data Science Tools that aren't Python, R, SQL or Math. Online: **towardsdatascience.com**.
- Michael Grogan (23 Jul 2020). How R Still Excels Compared To Python. Online: **towardsdatascience.com.**
- Knuth D (1992). **Literate Programming**. Stanford, Center for the Study of Language and Information Lecture Notes 27.
- Norman Matloff (2019). TidyverseSceptic. Online: **github.com**.
- Keith McNulty (23 Jun 2020). Data Scientists Should Learn Through Play. Online: **drkeithmcnulty.com**.
- Robert A. Muenchen (2017). Why R is Hard to Learn. Online: **r4stats.com**.
- Brien Posey (5 Feb 2018). How To Navigate the File System in Windows 10's Bash Shell. Online: **redmondmag.com**.

- Dario Radecic (10 Sept 2020). Trying R for the First Time. Online: **towardsdatascience.com**.
- Gordon Shotwell (30 Dec 2019). Why I use R. Online: **blog.shotwell.ca**.
- Sagar Uprety (23 Jul 2020). Data Cleaning and Exploratory Analysis in Python and R. Online: **towardsdatascience.com**.
- Yuleng Zeng (28 Aug 2018). An Introduction to R and LaTeX. Online: **bookdown.org**.

SOLUTIONS TO THE CHALLENGES

DOWNLOAD FROM CRAN

Mirror sites are called that way because they are actual identical copies of the original site. The quality of the cloned page is monitored. **The result looks interesting** (to me). You can see how well maintained a particular mirror site is.

OPENING R FOR THE FIRST TIME

The projects listed here (by no means a complete list!) are divided in applications and infrastructure projects. **Applications** of R include bioinformatics (e.g. in the medical sciences or in genomics), geospatial statistics (anything related to maps), and finance (R is strong with this one!). **Infrastructure** includes incorporation of R in Wikis (like Wikipedia) - for example to generate plots on the fly - and ESS ("Emacs Speaks Statistics"), which is the interface to the extensible text editor that I'm using (e.g. to create all documentation for this course - essentially from one text file). An alternative to ESS is the highly popular IDE (Integrated Development Environment) RStudio. We will not be using it in this course but I encourage you to check it out, try it and see if you like it, especially if my teaching tempo is too slow for you!

VERSION AND PLATFORM

See here to find out details of your CPU and computer architecture for **Windows** or **MacOS**.

DISTRIBUTION LICENSE

Go to **GNU Software** to see a list of all programs distributed under the GPL. These programs constitute the GNU system of free software. Looking through the list, I noticed the following programs that I have used: Chess (chess game implementation), Emacs (extensible text editor that I am using in this very moment), Gimp (image manipulation), Gnome (desktop for my operating system, Ubuntu Linux), and so on…425 programs are listed here alone (29 Aug 2020).

THE R PROJECT

There is no special connection between LaTeX and R, except that both are free software programs, one for formatting (especially when mathematical formulas need to be presented), the other one for statistical calculations and visualisation. However, to communicate data analysis results and to make the analysis process itself reproducible, a combination between these two goals (formatting/programming) is desirable. This is exactly what "literate programming" (**Knuth 1984**) does. There is also a program called "R Markdown" to create documents that enables you e.g. to created HTML, PDF, ePUB and Kindle books with only one source. You can find examples at **bookdown.org**. See also **Zeng (2018)** for a brief introduction to both R and LateX - sufficient to get started - written apparently as a minimal example for bookdown. For LaTeX there are also cloud editors like **overleaf.com**.

R PACKAGES

You can directly search for this dataset - I usually take the search string "`r doc [name]`, in this case `r doc MASS boston`, which gets me straight **to this page**. At the top, you can read that "The `Boston` data frame has 506 rows and 14 columns". There's also an R Notebook, which shows various aspects of this dataset.
Another way to find the answer is by using the command `str()` that you already know: `str(Boston` contains the answer in the first line - as long as `MASS` has been loaded. (Check out what happens if not by closing the R session with `q()` (don't save the workspace) and reopening it again. The simplest way is to type `help(Boston)` (again, only after loading the `MASS` package).