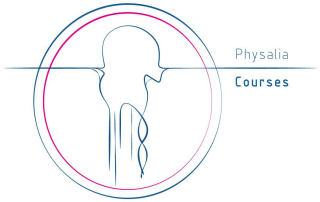
Getting started with **R**

Alexandre Courtiol

Leibniz Institute of Zoo and Wildlife Research

June 2019



Getting started with R

R in brief

R is a programming language and software environment for statistical computing & graphics.

Key points about **R**:

- free for all
- open source (explore: `https://github.com/wch/r-source`)

R in brief

R is a programming language and software environment for statistical computing & graphics.

Key points about **R**:

- free for all
- open source (explore: `https://github.com/wch/r-source`)
- polyvalent:
  - from laptop to most advanced supercomputers
  - local or remote
  - Windows, MacOS, linux or many other Unix-based systems

R in brief

R is a programming language and software environment for statistical computing & graphics.

Key points about **R**:

- free for all
- open source (explore: `https://github.com/wch/r-source`)
- polyvalent:
  - from laptop to most advanced supercomputers
  - local or remote
  - Windows, MacOS, linux or many other Unix-based systems
- rich (tons of **R** packages out there)
- cutting edge (check updates for today: `http://dirk.eddelbuettel.com/cranberries/cran/updated`)

R in brief

R is a programming language and software environment for statistical computing & graphics.

Key points about **R**:

- free for all
- open source (explore: `https://github.com/wch/r-source`)
- polyvalent:
  - from laptop to most advanced supercomputers
  - local or remote
  - Windows, MacOS, linux or many other Unix-based systems
- rich (tons of **R** packages out there)
- cutting edge (check updates for today: `http://dirk.eddelbuettel.com/cranberries/cran/updated`)
- used by millions

R in brief

R is a programming language and software environment for statistical computing & graphics.

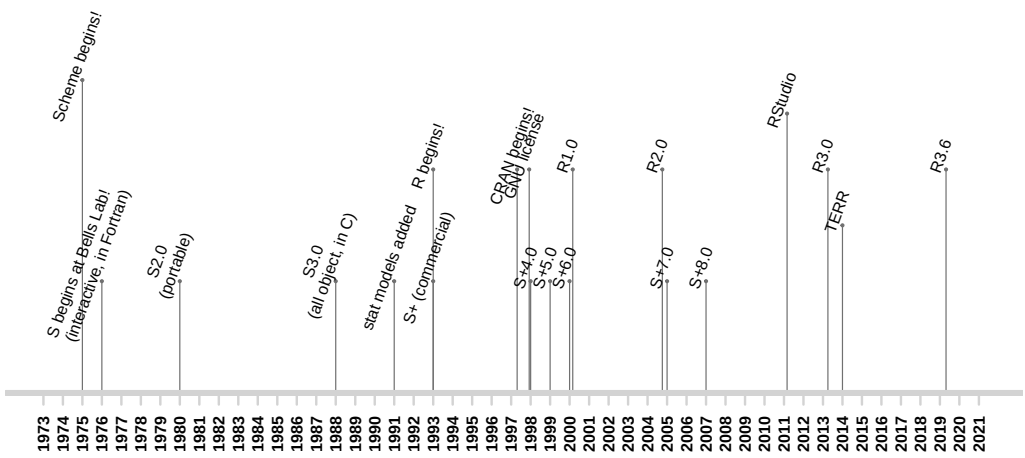
Key points about **R**:

- free for all
- open source (explore: `https://github.com/wch/r-source`)
- polyvalent:
  - from laptop to most advanced supercomputers
  - local or remote
  - Windows, MacOS, linux or many other Unix-based systems
- rich (tons of **R** packages out there)
- cutting edge (check updates for today: `http://dirk.eddelbuettel.com/cranberries/cran/updated`)
- used by millions

CONCLUSION: **R** is the best software environment for statistical computing (but it is far from perfect!)

Getting started with R

A short history of S/R



R tooks its syntax and tools from the **S**, and its semantic from `Scheme`

R's Who's Who

John Chambers
creator of **S**

Ross Ihaka
creator of **R**

Robert Gentleman
creator of **R** (& *Bioconductor*)

Hadley Wickham
creator of the `tidyverse`

Note: many other persons have been (and continue to be) strongly influencing **R** and its community!

Getting started with R

Is **R** good for you?

Good for:

- small, medium and big data
- data manipulation
- data analyses
- plotting (including GIS, web . . .)
- programming (even for biologists)

Is **R** good for you?

Good for:

- small, medium and big data
- data manipulation
- data analyses
- plotting (including GIS, web . . .)
- programming (even for biologists)

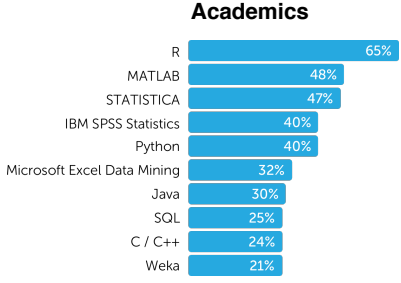
Not optimal for:

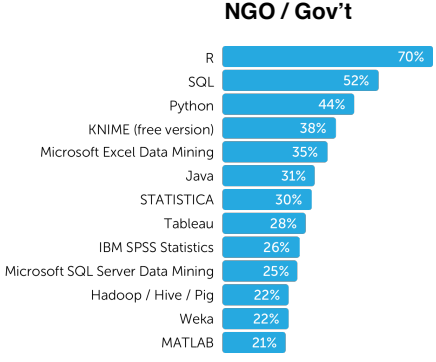
- beginners
- data entry
- formal algebra
- general purpose programming

Getting started with R

Most Data Scientists use Multiple Tools

What data science / analytic tools, technologies, and languages did you use in the past year?





All tools used by more than 20% of a group are shown

Rich companies rely on **R** too!

Some examples:
(`http://blog.revolutionanalytics.com/2014/05/companies-using-r-in-2014.html`)
(`https://blog.revolutionanalytics.com/2014/06/more-companies-using-r.html`)
(`https://blog.revolutionanalytics.com/2017/07/more-companies-using-r.html`)

- Facebook (data analysis, big-data visualization, user behaviour analysis)
- Google (advertising effectiveness, economic forecasting, and big-data statistical modeling)
- Lloyds of London (risk analysis and catastrophe modeling)
- Merk (analysis of factors that affect the cold chain)
- Microsoft (Xbox matchmaking + plus much more these days!!)
- Monsanto (statistical analysis in plant breeding, fertility mapping and yield forecasting)
- The New York Times (interactive features such as the Dialect Quiz and the Election Forecast)
- Twitter (data visualization and semantic clustering)
- Uber (various analyses)

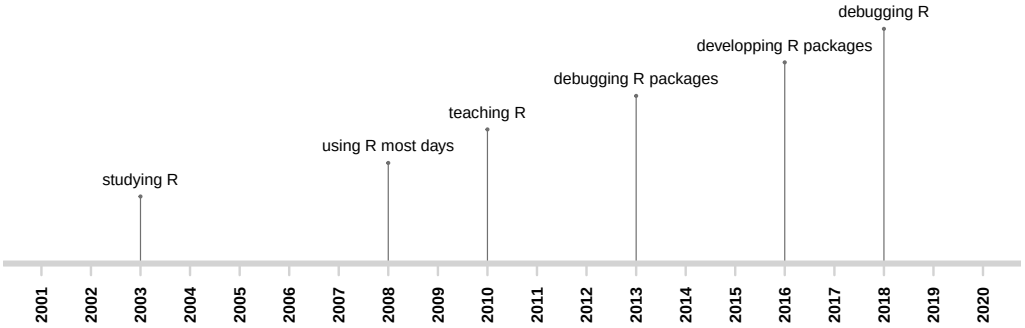
Getting started with **R**

Who am I?

- evolutionary biologist / statistician
- studies in France (Montpellier), postdoc in the UK (Sheffield)
- senior researcher at Leibniz IZW / lecturer at Freie University (Berlin)

Who am I?

- evolutionary biologist / statistician
- studies in France (Montpellier), postdoc in the UK (Sheffield)
- senior researcher at Leibniz IZW / lecturer at Freie University (Berlin)
- experience with **R**:



All of these activities are still ongoing (**R** is rich and evolving)

Getting started with **R**

What will we do?

I will:

- give you all the slides (so write only what is not being displayed)
- explain you the main principles of the **R** language and some useful functions
- provide you with short exercises to put concepts into practice
- provide you with suggestions on how to learn more about **R** on your own

What will we do?

I will:

- give you all the slides (so write only what is not being displayed)
- explain you the main principles of the **R** language and some useful functions
- provide you with short exercises to put concepts into practice
- provide you with suggestions on how to learn more about **R** on your own

You will not:

- be fluent in the **R** language after one week
- be able to remember how to implement many of the things we will see

What will we do?

I will:

- give you all the slides (so write only what is not being displayed)
- explain you the main principles of the **R** language and some useful functions
- provide you with short exercises to put concepts into practice
- provide you with suggestions on how to learn more about **R** on your own

You will not:

- be fluent in the **R** language after one week
- be able to remember how to implement many of the things we will see

You should:

- "*accept that you are gonna be frustrated, accept that you are gonna be bad at it for a while, but you have to keep going anyway...*" (Wickham 2018, `https://www.youtube.com/watch?v=Uxdf8evD6pQ`)
- ask any silly question that pops up in your creative minds
- let me know immediately when you stop following

Structure of the course

Day 1	Getting ready to work with **R**
Day 2-3	Getting to use data in **R**
Day 4	Getting to plot data in **R**
Day 5	Getting to analyses data in **R**