

Cómo escribir manuscritos reproducibles

(algunas ideas y herramientas)

Francisco Rodríguez-Sánchez

@frod_san

<https://frodriguezsanchez.net>

¿Qué es un manuscrito reproducible?

texto

datos

código



+



+



DATOS + CÓDIGO

- **trazar** proceso de análisis
- **reproducir** (regenerar) todos los resultados.

La ciencia debe ser reproducible

		Datos	
		Igual	Diferente
Análisis	Igual	Reproducible	Replicable
	Diferente	Robusto	Generalizable

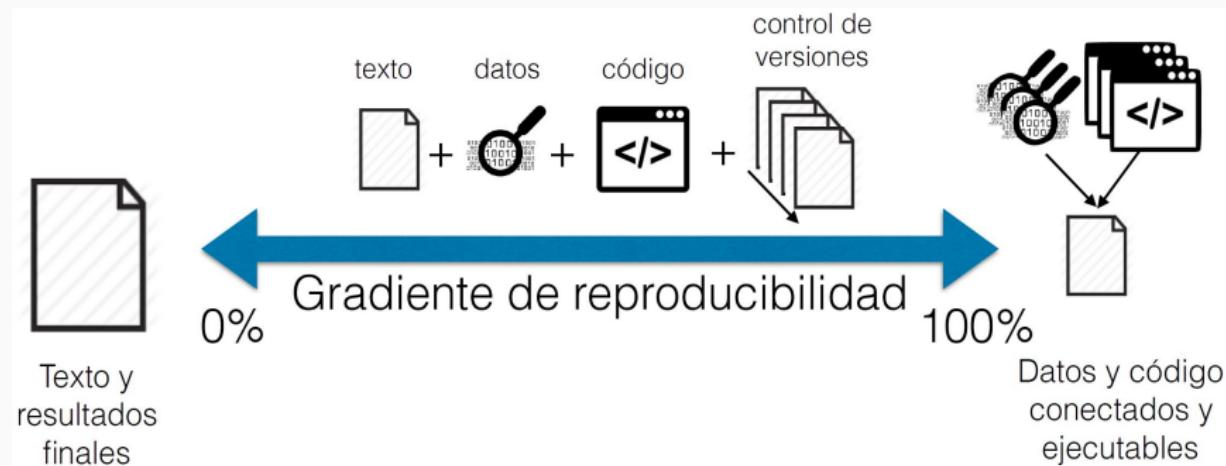
The Turing Way

Reproducibilidad: buena para ti, buena para todos

- Transparencia
- Trazabilidad
- Automatización
- Calidad (- errores)
- Impacto (citas, reconocimiento)
- Reutilización (datos y código)
- Archivo permanente

La inmensa mayoría
de artículos científicos
NO son reproducibles

La reproducibilidad es un gradiente



Rodríguez-Sánchez et al. 2016 (modif. Peng 2011)

Reproducibilidad básica

Reproducibilidad básica

- MANUSCRITO (Texto + Tablas + Figuras)
- DATOS (archivo permanente)
- CÓDIGO (archivo permanente)

Archivo permanente:

- Zenodo, Dryad, OSF, Figshare, Data Paper...
- NO GitHub, web...

Cómo compartir datos

- Formato abierto (csv, txt)

Cómo compartir datos

- Formato abierto (csv, txt)
- Datos brutos + código depuración

Cómo compartir datos

- Formato abierto (csv, txt)
- Datos brutos + código depuración
- README (who, what, when, where, why, how)

Cómo compartir datos

- Formato abierto (csv, txt)
- Datos brutos + código depuración
- README (who, what, when, where, why, how)
- Diccionario (**descripción variables: dataspice, codebook**)

Cómo compartir datos

- Formato abierto (csv, txt)
- Datos brutos + código depuración
- README (who, what, when, where, why, how)
- Diccionario (**descripción variables: dataspice, codebook**)
- Licencia (CC0, CC-BY, ODbL)

Cómo compartir datos

- Formato abierto (csv, txt)
- Datos brutos + código depuración
- README (who, what, when, where, why, how)
- Diccionario (**descripción variables: dataspice, codebook**)
- Licencia (CC0, CC-BY, ODbL)
- Cita (DOI)

Cómo compartir datos

- Formato abierto (csv, txt)
- Datos brutos + código depuración
- README (who, what, when, where, why, how)
- Diccionario (**descripción variables: dataspice, codebook**)
- Licencia (CC0, CC-BY, ODbL)
- Cita (DOI)
- Metadatos estandarizados (JSON, XML)

Cómo compartir código

- Scripts **texto** (.R)

Eglen et al 2016

Cómo compartir código

- Scripts **texto** (.R)
- Archivo permanente (ej. Zenodo) con DOI (citable)

Eglen et al 2016

Cómo compartir código

- Scripts **texto** (.R)
- Archivo permanente (ej. Zenodo) con DOI (citable)
- Licencia

Eglen et al 2016

Cómo compartir código

- Scripts **texto** (`.R`)
- Archivo permanente (ej. Zenodo) con DOI (citable)
- [Licencia](#)
- README (incluir `sessionInfo()`)

Eglen et al 2016

sessionInfo(): registro del entorno computacional

```
sessionInfo()

## R version 4.0.2 (2020-06-22)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.1 LTS
##
## Matrix products: default
## BLAS:    /usr/lib/x86_64-linux-gnublas/libblas.so.3.9.0
## LAPACK:  /usr/lib/x86_64-linux-gnulapack/liblapack.so.3.9.0
##
## locale:
## [1] LC_CTYPE=en_GB.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=es_ES.UTF-8      LC_COLLATE=en_GB.UTF-8
## [5] LC_MONETARY=es_ES.UTF-8   LC_MESSAGES=en_GB.UTF-8
## [7] LC_PAPER=es_ES.UTF-8      LC_NAME=C
## [9] LC_ADDRESS=C              LC_TELEPHONE=C
## [11] LC_MEASUREMENT=es_ES.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics   grDevices  utils      datasets   methods    base
##
## other attached packages:
## [1] knitr_1.30
##
## loaded via a namespace (and not attached):
## [1] compiler_4.0.2  magrittr_1.5    htmltools_0.5.0  tools_4.0.2
## [5] yaml_2.2.1     codetools_0.2-16 stringi_1.5.3   rmarkdown_2.3
## [9] binb_0.0.6      stringr_1.4.0   xfun_0.17      digest_0.6.25
## [13] rlang_0.4.7     evaluate_0.14
```

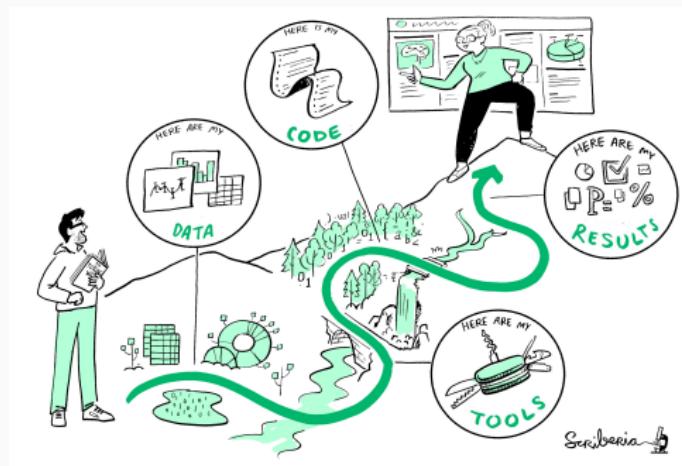
Cómo escribir código más reproducible

BES guide to reproducible code



Cómo escribir código más reproducible

Turing Way



Cómo escribir código más reproducible



PERSPECTIVE

Good enough practices in scientific computing

Greg Wilson^{1*}, Jennifer Bryan², Karen Cranston³, Justin Kitzes⁴, Lex Nederbragt⁵,
Tracy K. Teal⁶

Wilson et al 2017

Cómo escribir código más reproducible

<https://rstats.wtf>

What They Forgot to Teach You About R

Jennifer Bryan

Jim Hester

Cómo escribir código más reproducible

fertile: creating optimal conditions for reproducibility

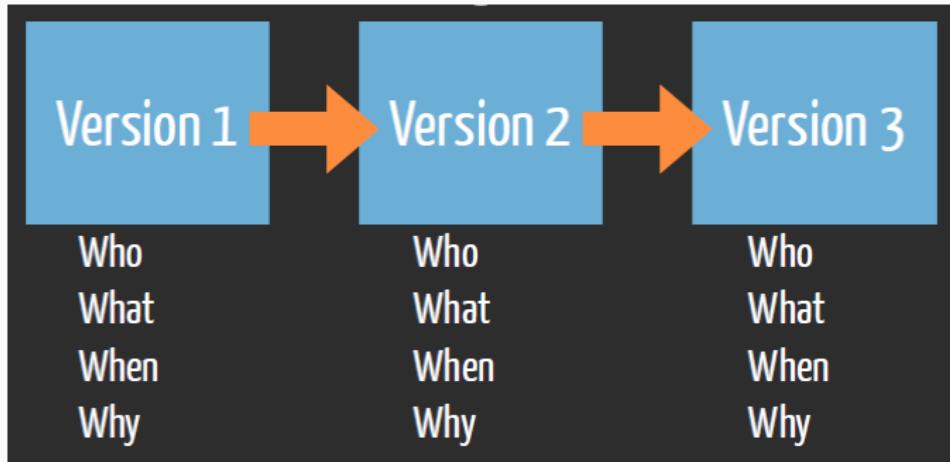
```
library("fertile")
setwd("C:/Users/FRS")
```

Error: setwd() is likely to break reproducibility. Use here::here() instead.

<https://github.com/baumer-lab/fertile>

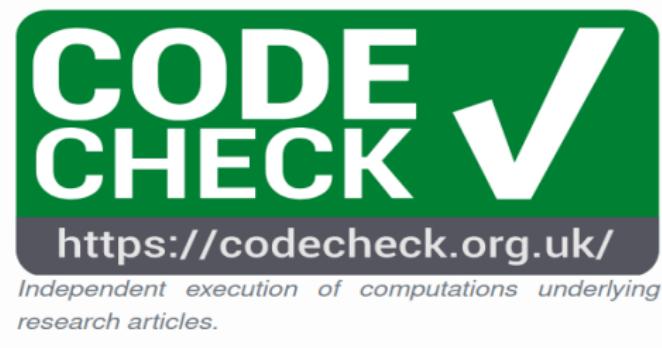
Cómo escribir código más reproducible

Control de versiones (git)



Codecheck: ¿es tu artículo reproducible?

<https://codecheck.org.uk>



Reproducibilidad básica

- **MANUSCRITO** (texto + tablas + figuras)
- **DATOS** (archivo permanente)
- **CÓDIGO** (archivo permanente)

La inmensa mayoría
de artículos científicos
NO son reproducibles

El código existe, pero no se comparte

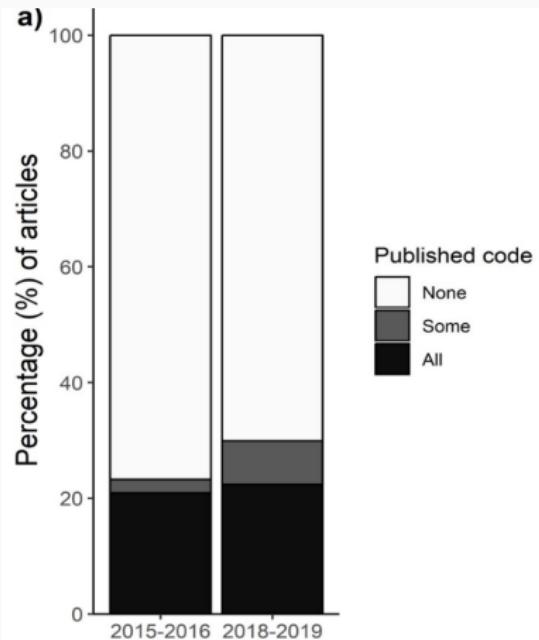


Fig 1. Code-sharing is at its infancy in ecology, where

Las barreras
NO son técnicas

Publish your computer code: it is good enough

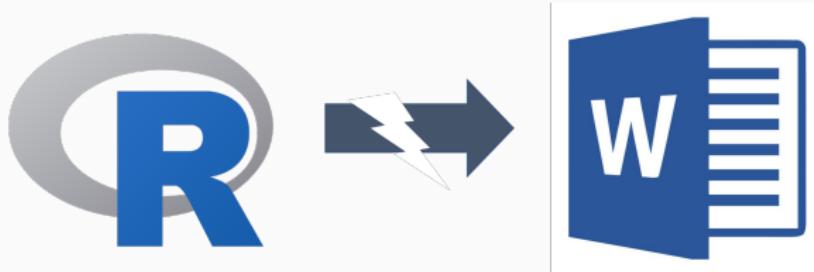
*Freely provided working code – whatever its quality – improves programming and enables others to engage with your research, says **Nick Barnes**.*

Barnes 2010

- Comunidad debe ser constructiva (no *shaming*)
- ‘Ugly code better than no code’

‘Literate programming’ con Rmarkdown

Desconexión de código y manuscrito genera problemas



'Transcribing numbers **from stats software to ms by hand**
was the largest source of errors'

(Eubank 2016)

Había un dato erróneo...

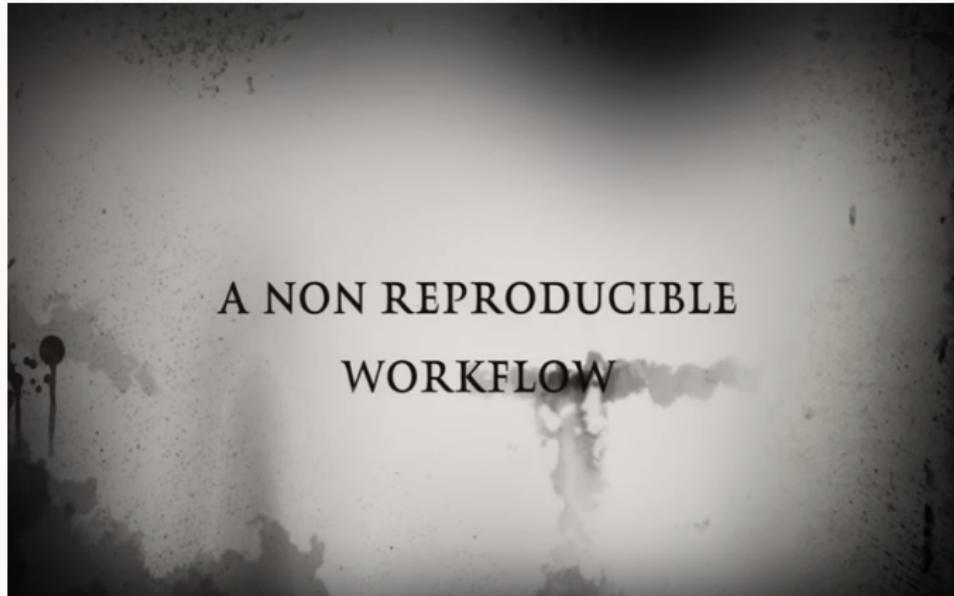
¿Puedes repetir el análisis y actualizar el ms?

Copiar y pegar...

Coefficients:

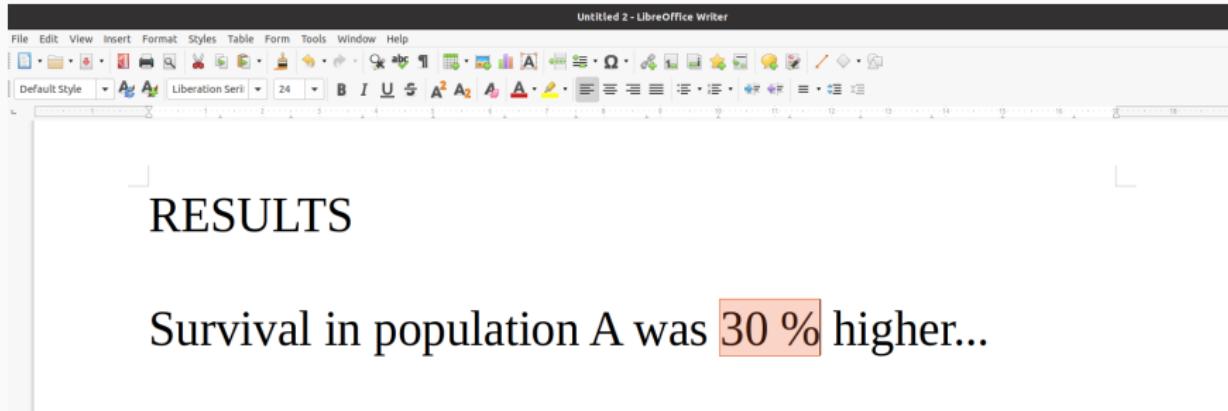
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0651657	0.4264970	-0.153	0.879
sunshine	0.0100228	0.0004232	23.683	<2e-16

Las ventajas de la reproducibilidad en <2 minutos



<https://youtu.be/s3JldKoA0zw>

¿De dónde sale este número?

A screenshot of a LibreOffice Writer document titled "Untitled 2 - LibreOffice Writer". The interface includes a menu bar with File, Edit, View, Insert, Format, Styles, Table, Form, Tools, Window, and Help. Below the menu is a toolbar with various icons for document operations. The main content area contains the word "RESULTS" in a large, bold, black serif font. Below it is a statement in a smaller black serif font: "Survival in population A was 30 % higher...". The percentage "30 %" is highlighted with a light orange rectangular box.

Documentos dinámicos con Rmarkdown

Rmarkdown:

Survival in population A was `r surv.diff` % higher

Output:

Survival in population A was **30** % higher

Documentos dinámicos con Rmarkdown

```
datos <- read.csv("datos.csv")
```

Rmarkdown:

We measured `r nrow(datos)` individuals

Output:

We measured **86** individuals

¡Mucho mejor que copiar y pegar!

Rmarkdown: texto + código

```
---
```

```
title: "Does sunshine make people happy?"  
author: "FRS"  
output: word_document  
---
```

Introduction

It is well known that individual well-being can be influenced by climatic conditions.

Methods

```
```{r echo=FALSE}  
Read data
data <- read.table("data.txt", header = TRUE)

Fit linear model
model <- lm(happiness ~ sunshine, data = data)
```
```

We collected data on `r nrow(data)` individuals and fitted a linear model.

Metadatos
(YAML)

Texto
(Markdown)

Código
(R, Python...)

Regenera Word/PDF/html... con un click

```
---
```

```
title: "Does sunshine make people happy?"
```

```
output: pdf_document
```

```
bibliography: refs.bib
```

```
---
```

```
# Introduction
```

```
climate influences individual well-being [Rehdanz_2005].
```

```
However, ...
```

```
# Methods
```

```
```{r echo=FALSE}
```

```
read data
```

```
data <- read.table("data.txt", header=T)
```

```
data[10,1] <- 11 # correct error
```

```
fit linear model
```

```
model <- lm(happiness ~ sunshine, data=data)
```

```

```

```
we collected data on `r nrow(data)` individuals and fitted a
```

```
linear model.
```

```
Results
```

```
We found that...
```

```
```{r echo=FALSE, results='asis'}
```

```
# make table with model output
```

```
print(xtable::xtable(model), comment = FALSE)
```

```
---
```

```
```{r echo=FALSE, fig.height=3, fig.width=3, fig.align='center'}
```

```
visreg::visreg(model) # plot
```

```

```

```
Discussion
```

```
Our results confirm that happiness is related to
```

```
sunshine (slope = `r coef(model)[2]`).
```

```
References
```

a

## Does sunshine make people happy?

b

### Introduction

Climate influences individual well-being (Rehdanz and Maddison 2005). However, ...

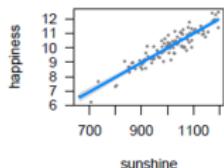
### Methods

We collected data on 100 individuals and fitted a linear model.

### Results

We found that...

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0986	0.4271	-0.23	0.8180
sunshine	0.0101	0.0004	23.75	0.0000



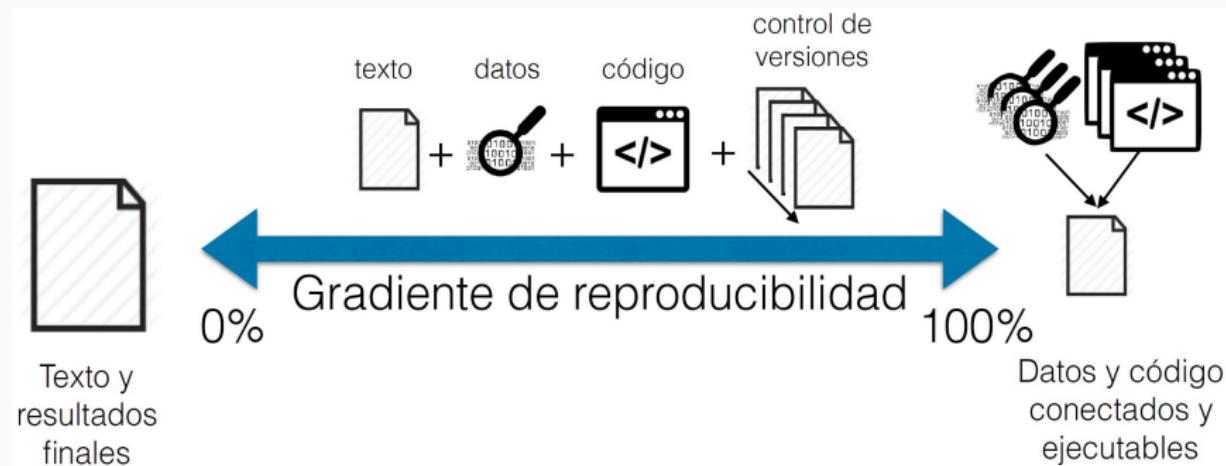
### Discussion

Our results confirm that happiness is related to sunshine (slope = 0.0100652).

### References

Rehdanz, Katrin, and David Maddison. 2005. "Climate and Happiness." *Ecological Economics* 52 (1). Elsevier BV: 111–25. doi:10.1016/j.ecolecon.2004.06.015.

# Rmarkdown: resultados trazables y ejecutables



Rodríguez-Sánchez et al. 2016

## Rmarkdown: generación automática de tablas

```
model <- lm(happiness ~ sunshine, data = datos)
xtable(model)
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0652	0.4265	-0.15	0.8789
sunshine	0.0100	0.0004	23.68	0.0000

## equatiomatic describe la estructura de los modelos

We fitted a linear model:

```
library("equatiomatic")
model <- lm(happiness ~ sunshine, data = datos)
extract_eq(model)
```

$$\text{happiness} = \alpha + \beta_1(\text{sunshine}) + \epsilon$$

# ¡Modelos que se auto-describen!

```
library("report")
model <- lm(happiness ~ sunshine, data = datos)
report(model)
```

We fitted a linear model (estimated using OLS) to predict happiness with sunshine (formula = happiness ~ sunshine). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. Effect sizes were labelled following Cohen's (1988) recommendations.

The model explains a significant and substantial proportion of variance ( $R^2 = 0.85$ ,  $F(1, 98) = 560.90$ ,  $p < .001$ , adj.  $R^2 = 0.85$ ). The model's intercept, corresponding to happiness = 0 and sunshine = 0, is at -0.07 (SE = 0.43, 95% CI [-0.91, 0.78],  $p = 0.879$ ). Within this model:

- The effect of (Intercept) is negative and can be considered as very small and not significant ( $b = -0.07$ , SE = 0.43, 95% CI [-0.91, 0.78], std. beta = 2.28e-16,  $p = 0.879$ ).
- The effect of sunshine is positive and can be considered as large and significant ( $b = 0.01$ , SE = 4.23e-04, 95% CI [9.18e-03, 0.01], std. beta = 0.92,  $p < .001$ ).

# Revisa tu gramática

<https://github.com/ropenscilabs/gramr>

**Ignore**

- Passive Voice
- Duplicate words (the the)
- 'So' at start of sentence
- 'There is/are; at start of sentence
- Avoid weasel words
- Wordiness
- Problematic Adverbs
- Cliches
- Avoid 'Being' words

[Next](#) [Finish](#)

**Text to Check**

So the cat was stolen. This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.

"was stolen" may be passive voice

# rcrossref: cita artículos vía DOI

Add Crossref Citations

Cancel Add Crossref Citations Done

Add a new bibliography entry through Crossref DOI

10.1111/j.2044-8317.2011.02037.x

Philosophy and the practice of Bayesian statistics

Link: [10.1111/j.2044-8317.2011.02037.x](https://doi.org/10.1111/j.2044-8317.2011.02037.x)

Author(s): Andrew Gelman and Cosma Rohilla Shalizi

Journal/Container: British Journal of Mathematical and Statistical Psychology

Year: 2012

Save to references.bib Add to My Citations

Search Metadata Search by DOI

# rcrossref: búsqueda de artículos

Add Crossref Citations

Cancel Add Crossref Citations Done

networks Sort: Relevance ▾ ▾

Author	Journal/Container	Type	Since
Jordano		Journal Article	Any time ▾

Sampling networks of ecological interactions

Link: [10.1111/1365-2435.12763](https://doi.org/10.1111/1365-2435.12763)

Author(s): Pedro Jordano

Journal/Container: Functional Ecology

Issued: 2016-10-05

Save to references.bib Add to My Citations

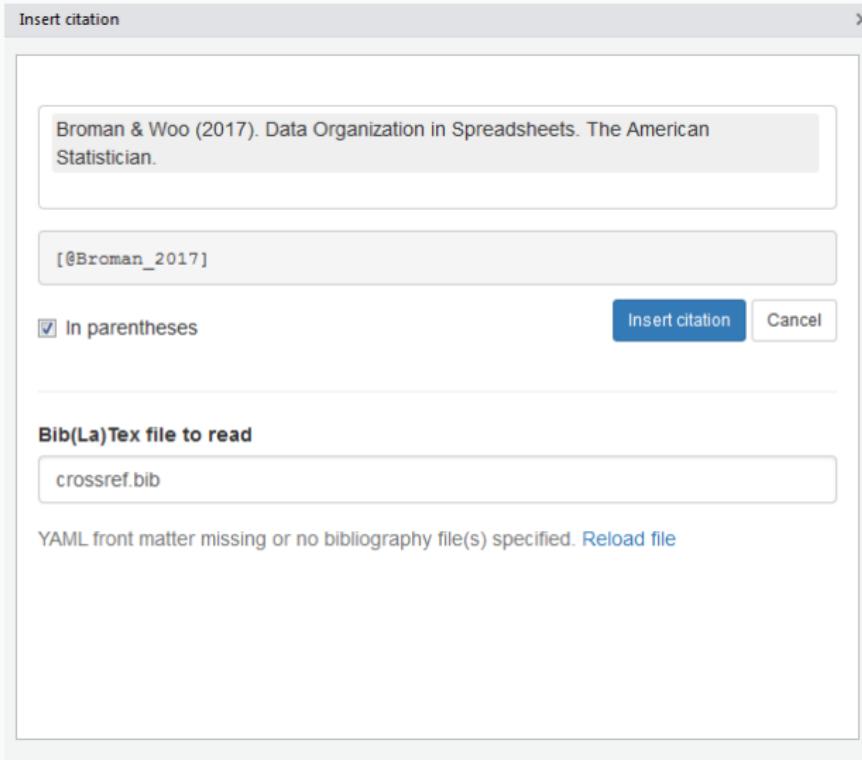
title	issued
Sampling networks of ecological interactions	2016-10-05
Natural history matters: how biological constraints shape diversified interactions in pollination networks	2016-10-19
COEVOLUTION AND THE ARCHITECTURE OF MUTUALISTIC NETWORKS	2012-10-05
Plant-Animal Mutualistic Networks: The Architecture of Biodiversity	2007-12
Geographic Patterns in Plant-Pollinator Mutualistic Networks	2002-09

Showing 1 to 5 of 35 entries Previous 1 2 3 4 5 6 7 Next

 Search Metadata

 Search by DOI

# citr: citar artículo de base de datos (Zotero)



# Rstudio 1.4 facilita cita de múltiples fuentes

The screenshot shows the 'Insert Citation' dialog in RStudio 1.4. On the left, a sidebar titled 'My Sources' lists various citation databases: Bibliography, Zotero, My Library, From DOI, Crossref, DataCite, and PubMed. The main area displays a list of citation entries with icons, keys, titles, authors, and years. A search bar at the top right says 'search for citation'. At the bottom, there's a 'Selected Citation Keys' section, an 'Add to bibliography:' dropdown set to 'book.bib', and 'Insert' and 'Cancel' buttons.

Citation Key	Title	Author(s)	Year
@baghizadehfini2020	What dentists need to know about COVID-19	Baghizadeh Fini, M	2020
@bostanciklioglu2020	Severe Acute Respiratory Syndrome Coronavirus 2 is Penetrating to Dementia Re...	Bostanciklioglu, M	2020
@fran	Functional reactive animation	Elliott, C, and Hudak, P	1997
@guo2020	The origin, transmission and clinical therapies on coronavirus disease 2019 (CO...	Guo, Y, Cao, Q, Hong, Z, Tan, Y, Chen, et al.	2020
@hu2020	The cytokine storm and COVID-19	Hu, B, Huang, S, and Yin, L	2020
@malik2020	Coronavirus Disease Pandemic (COVID-19): Challenges and a Global Perspective	Malik, Y, Kumar, N, Sircar, S et al.	2020
@R-base	R: A language and environment for statistical computing	R Core Team	2017

<https://rstudio.github.io/visual-markdown-editing/#/citations>

¡No olvides citar los paquetes de R!

Rmarkdown:

```
knitr::write_bib()
```

Procesador texto:

```
grateful::cite_packages()
```

# Formatea bibliografía para cualquier revista

```

title: "Does sunshine make people happy?"
author: "FRS"
output: word_document
bibliography: myrefs.bib
csl: ecology-letters.csl

```

Repositorio de 'citation styles' (CSL):

<https://www.zotero.org/styles>

<https://github.com/citation-style-language/styles>

# Plantillas Rmarkdown

- rticles
- papaja
- rrttools
- pinc
- rmdTemplates
- GitHub!

## My cool paper written in Rmarkdown

F. Rodriguez-Sánchez<sup>a,1,2</sup> and And Friendz<sup>b,3</sup>

<sup>a</sup>Now Institute of Technology, Department, Street, City, State, Zip; <sup>b</sup>Another University Department, Street, City, State, Zip

This manuscript was compiled on September 10, 2018.

Please provide an abstract no more than 250 words in a single paragraph. Abstracts should contain no more than one sentence per major component of the manuscript. References in the abstract must be cited in full within the abstract itself and cited in the text.

[one](#) | [two](#) | [optional](#) | [optional](#) | [optional](#)

This PNAS journal template is provided to help you write your work in the correct journal format. Instructions for use are given below.

Note: Please start your introduction without including the word "Introduction" as a section heading (except for math articles in the Physical Sciences section); this heading is implied in the first paragraph.

**Guide to using this template**

Please note that while this template provides a preview of the typeset manuscript for submission, to help in this preparation, it will not necessarily be the final publication layout. For more detailed information please see the [PNAS Information for Authors](#).

**Author Affiliations:** Include department, institution, and complete address, with the ZIP/postal code, for each author. Use lower case letters to match authors with institutions, as shown in the example. Authors with an ORCID ID may supply this information at submission.

**Submitting Manuscripts:** All authors must submit their articles at [PNAScentral](#). If you are using Overleaf to write your article, please use the "Submit to PNAS" option in the top bar of the editor.

**Format:** Many authors find it useful to organize their manuscripts with the following order of sections: Title, Author Affiliation, Keywrods, Abstract, Significance Statement, Results, Discussion, Materials and methods, Acknowledgments, and References. Other orders and headings are permitted.

**Manuscript Length:** PNAS generally uses a two-column format spanning 87 columns, including spaces, per line. The maximum length of a standard research manuscript is 8 pages and a PNAS PLUS research article is 16 pages (including all text, spaces, and the number of characters displayed by figures, tables, and equations). When submitting tables, figures, and/or equations in addition to text, keep the total page count under 20,000 characters (including spaces) for Direct Submissions and 72,000 characters (including spaces) for PNAS PLUS.

**References:** References should be cited in numerical order as they appear in text; this will be done automatically via bibtex, e.g. (1) and (2, 3). All references, including for the SI, should be included in the main manuscript file. References appearing in both section should not be duplicated. SI references

Fig. 1. Placeholder image of a frog with a long sample caption to show justification writing.

included in tables should be included with the main reference section.

**Data Archival:** PNAS must be able to archive the data essential to a published article. Where such archiving is not possible, deposition of data in public databases, such as GenBank, Arxiv, the Protein Data Bank, UniProt, and others outlined in the Information for Authors, is accepted.

**Language Editing Services:** Prior to submission, authors who believe their manuscript would benefit from professional editing are encouraged to use a language-editing service (see list at [www.pnas.org/site/authors/language-editing.shtml](#)). PNAS does not take responsibility for or endorse these services, and their use can have no bearing on acceptance of a manuscript for publication.

**Significance Statement:**

Authors must submit a 120-word maximum statement about the significance of their research paper when at least three students to an undergraduate educated scientist outside their field of specialty. The primary goal of the Significance Statement is to explain the relevance of the work in broad context to a broad readership. The Significance Statement appears in the paper track and is required for all research papers.

[Please provide details of other contributions](#)  
[Please declare any conflict of interest here](#)

[www.pnas.org/pnas/10.1073/pnas.0000000000X](#)



¡No solo R! Python, Julia, C++, SQL, Stan, etc

`knitr` engines:

---

asis	c	fortran	highlight	octave	ruby	sh
asy	cat	fortran95	js	perl	sas	sql
awk	cc	gawk	julia	psql	sass	stan
bash	coffee	go	lein	python	scala	stata
block	css	groovy	mysql	Rcpp	scss	tikz
block2	dot	haskell	node	Rscript	sed	zsh

---

# Documentos dinámicos con Rmarkdown

<https://rmarkdown.rstudio.com/>



# 'Visual Rmarkdown'en Rstudio 1.4 trae múltiples ventajas

¡Como escribir en procesador de texto, pero todas las ventajas de Rmarkdown!

The editor toolbar includes buttons for the most commonly used formatting commands:



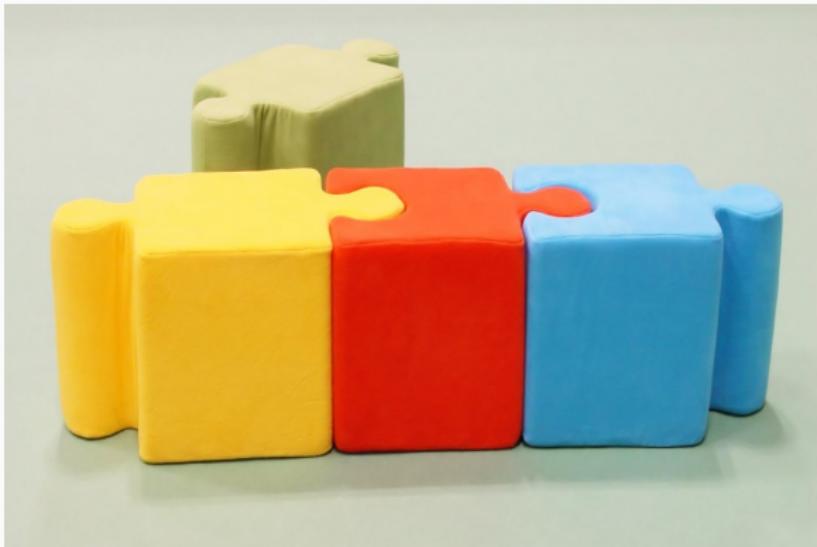
Additional commands are available on the **Format**, **Insert**, and **Table** menus:

A screenshot of the RStudio 1.4 menu bar with three menus open: Format, Insert, and Table. The Format menu contains options for Bold, Italic, Code, Text, Bullets &amp; Numbering, Blockquote, Line Block, Div Block..., Code Block..., Raw, Clear Formatting, and Edit Attributes... with a keyboard shortcut of ⌘\|. The Insert menu contains options for Rmd Chunk, Image..., Link..., Horizontal Rule, Definition, Inline Math, Display Math, Footnote, Citation..., Div Block..., Code Block..., YAML Block, and Comment. The Table menu contains options for Insert Table..., Table Header, Table Caption, Align Column, Insert Row Above, Insert Row Below, Insert Column Left, Insert Column Right, Delete Row, Delete Column, and Delete Table.

## Control de flujos de trabajo

---

En proyectos complejos es necesario **organizar las piezas**



## Script maestro ejecuta scripts en orden

makefile.R

```
source("clean_data.R")

source("fit_model.R")

source("generate_report.R")
```

# Script maestro ejecuta scripts en orden

```
READ AND PREPROCESS DATA
Read site info
read_siteinfo("data-raw/sites_info_raw.csv")

Read and prepare species info
read_sppinfo(sppdata = "data-raw/species_info_raw.csv")

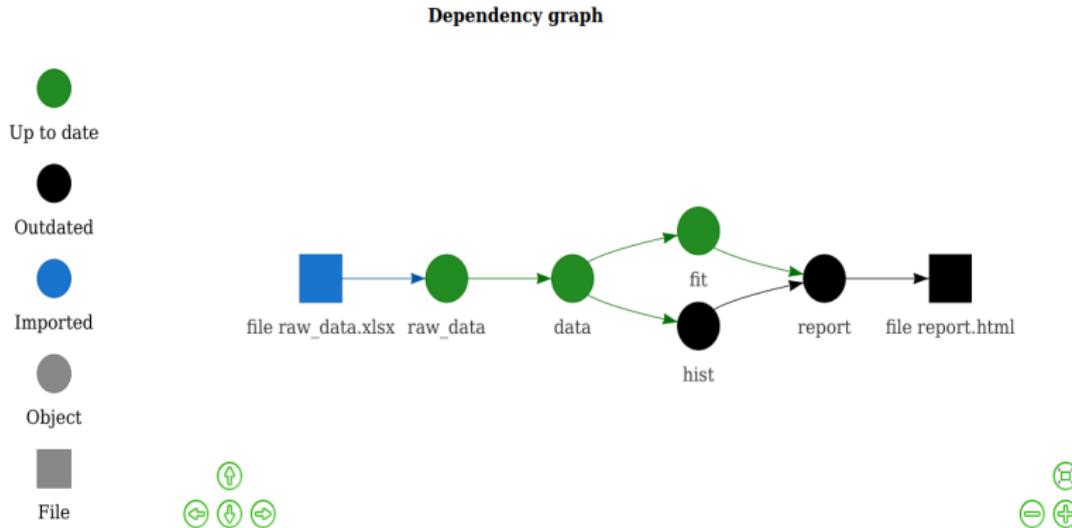
Prepare dataset
make_dataset()

EXPLORATORY ANALYSIS
rmarkdown::render("analyses/EDA.Rmd")

MANUSCRIPT
rmarkdown::render("manuscript/cercados_Almoraima/cercados_Almoraima.Rmd")
```

<https://github.com/Pakillo/exclosures-Almoraima>

# drake/targets: control total del flujo de trabajo



<https://docs.ropensci.org/drake/>

<https://wlandau.github.io/targets/>

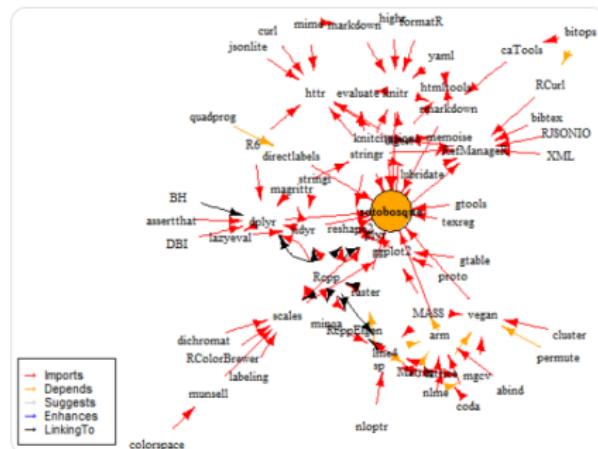
## Control de dependencias

---



F Rodriguez-Sanchez  
@frod\_san

Our last project depends on complex ecosystem of 67 co-evolving #rstats pkgs. Ensuring #reproducibility not trivial



5:39 pm · 27 Jan 2016 · Twitter Web Client

Cambios en paquetes pueden romper/alterar el análisis

¿Cómo reproducir el análisis dentro de un año, o en otro ordenador?

# sessionInfo registra paquetes usados y versiones, OS...

```
sessionInfo()

R version 4.0.2 (2020-06-22)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 20.04.1 LTS
##
Matrix products: default
BLAS: /usr/lib/x86_64-linux-gnublas/libblas.so.3.9.0
LAPACK: /usr/lib/x86_64-linux-gnulapack/liblapack.so.3.9.0
##
locale:
[1] LC_CTYPE=en_GB.UTF-8 LC_NUMERIC=C
[3] LC_TIME=es_ES.UTF-8 LC_COLLATE=en_GB.UTF-8
[5] LC_MONETARY=es_ES.UTF-8 LC_MESSAGES=en_GB.UTF-8
[7] LC_PAPER=es_ES.UTF-8 LC_NAME=C
[9] LC_ADDRESS=C LC_TELEPHONE=C
[11] LC_MEASUREMENT=es_ES.UTF-8 LC_IDENTIFICATION=C
##
attached base packages:
[1] stats graphics grDevices utils datasets methods base
##
other attached packages:
[1] report_0.1.0 equatiomatic_0.1.0 xtable_1.8-4 knitr_1.30
##
loaded via a namespace (and not attached):
[1] magrittr_1.5 insight_0.9.6 tidyselect_1.1.0 performance_0.5.0
[5] R6_2.4.1 rlang_0.4.7 highr_0.8 stringr_1.4.0
[9] dplyr_1.0.2 tools_4.0.2 broom_0.7.0 xfun_0.17
[13] bayestestR_0.7.2 htmltools_0.5.0 ellipsis_0.3.1 yaml_2.2.1
```

checkpoint reconstruye ecosistema de paquetes en fecha determinada

```
library("checkpoint")
checkpoint("2019-10-08")

source("analysis.R")
```

1. Detecta paquetes usados en el proyecto
2. Instala versión correspondiente a esa fecha (solo CRAN)
3. Instalación independiente (no interfiere con paquetes ya instalados)

# automagic detecta e instala paquetes (CRAN + GitHub)

```
automagic::make_deps_file()
```

Crea `deps.yaml` especificando dependencias:

```
- Package: equatiomatic
 Repository: CRAN
 Version: 0.1.0

- Package: report
 GithubUsername: easystats
 GithubRepo: report
 GithubRef: HEAD
 GithubSHA1: c48a4bb0a40df7116bc502aa3ce2cb9d70b7e2
```

Para instalar todas esas dependencias:

```
automagic()
```

renv también controla las dependencias del proyecto

```
renv::init()

renv::snapshot()
Captura dependencias en fichero lockfile

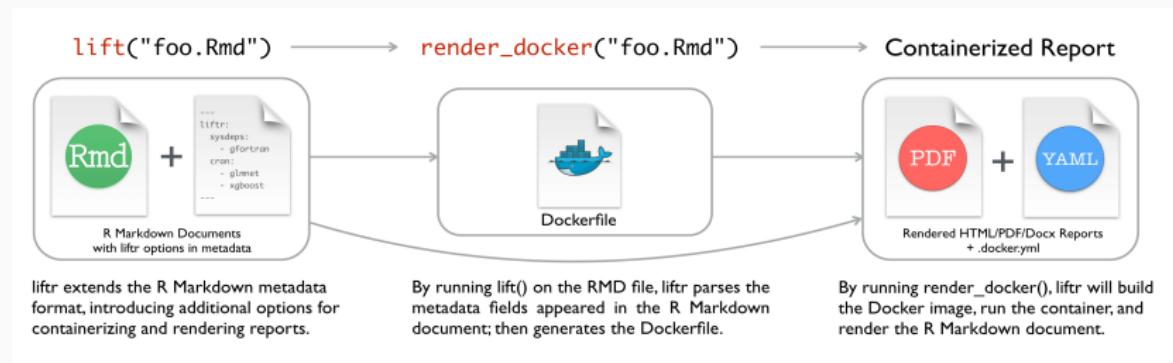
renv::restore()
Regenera dependencias a partir del lockfile
```

<https://environments.rstudio.com/>

Para asegurar la reproducibilidad,  
además de los paquetes de R  
necesitamos controlar el **sistema/entorno computacional**

Docker permite recrear sistemas virtuales  
a partir de un Dockerfile  
que especifica la configuración

# liftr: procesa documento Rmd en un contenedor Docker



<https://liftr.me/>

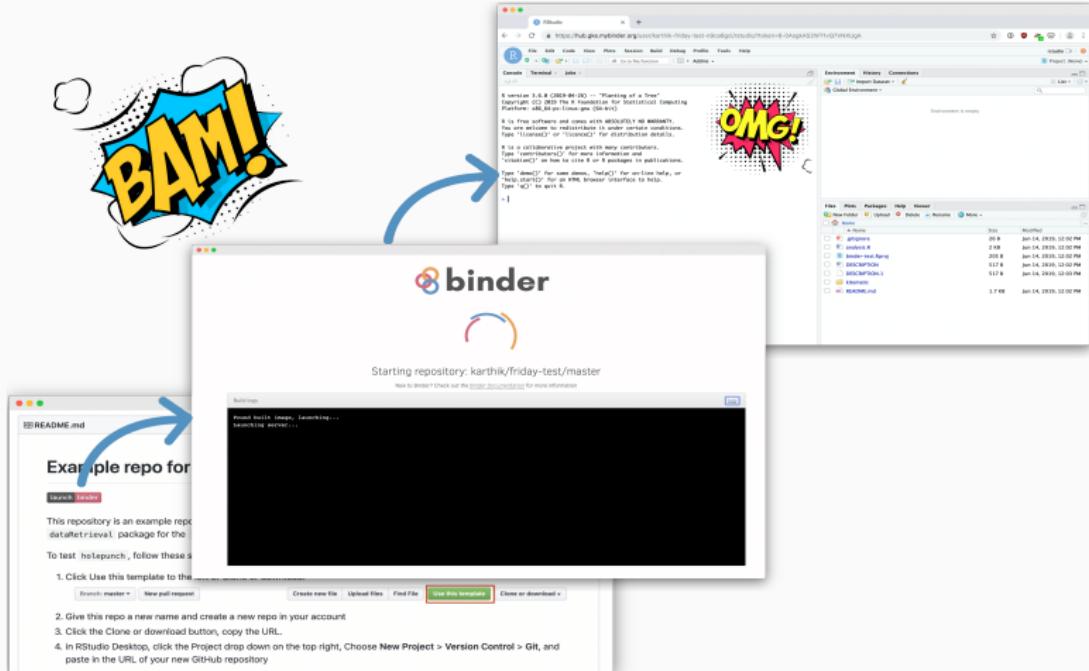
## containerit facilita creación de Dockerfile

```
library("containerit")

dockfile <- dockerfile(from = "mypaper.Rmd")
```

<https://o2r.info/containerit>

# holepunch: reproduce análisis en la nube (Binder)



<https://karthik.github.io/holepunch/>

## Organización del proyecto

---

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros
- README

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros
- README
- Licencia

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros
- README
- Licencia
- Datos brutos separados de limpios (+ código)

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros
- README
- Licencia
- Datos brutos separados de limpios (+ código)
- Funciones independientes del análisis

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros
- README
- Licencia
- Datos brutos separados de limpios (+ código)
- Funciones independientes del análisis
- Funciones documentadas y testadas

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros
- README
- Licencia
- Datos brutos separados de limpios (+ código)
- Funciones independientes del análisis
- Funciones documentadas y testadas
- Makefile

# Principios sobre organización de proyectos

- Una carpeta con todos los ficheros
- README
- Licencia
- Datos brutos separados de limpios (+ código)
- Funciones independientes del análisis
- Funciones documentadas y testadas
- Makefile
- Control de dependencias

## Research compendium:

- Datos + Código + Texto
- Estructura de paquete de R

# Proyectos estructurados como paquetes de R

- Estructura estándar, consistente

Rodríguez-Sánchez et al. 2016, Marwick et al 2018, but see McBain 2020

# Proyectos estructurados como paquetes de R

- Estructura estándar, consistente
- Promueve código modular, bien documentado y testado

Rodríguez-Sánchez et al. 2016, Marwick et al 2018, but see McBain 2020

# Proyectos estructurados como paquetes de R

- Estructura estándar, consistente
- Promueve código modular, bien documentado y testado
- Fácil de compartir y ejecutar

Rodríguez-Sánchez et al. 2016, Marwick et al 2018, but see McBain 2020

# Proyectos estructurados como paquetes de R

- Estructura estándar, consistente
- Promueve código modular, bien documentado y testado
- Fácil de compartir y ejecutar
- Integración continua (Travis, GitHub Actions, etc)

Rodríguez-Sánchez et al. 2016, Marwick et al 2018, but see McBain 2020

# Proyectos estructurados como paquetes de R

- Estructura estándar, consistente
- Promueve código modular, bien documentado y testado
- Fácil de compartir y ejecutar
- Integración continua (Travis, GitHub Actions, etc)
- Revisión automática del código con `goodpractice`

Rodríguez-Sánchez et al. 2016, Marwick et al 2018, but see McBain 2020

# Proyectos estructurados como paquetes de R

- Estructura estándar, consistente
- Promueve código modular, bien documentado y testado
- Fácil de compartir y ejecutar
- Integración continua (Travis, GitHub Actions, etc)
- Revisión automática del código con `goodpractice`
- Muy fácil crear website del proyecto con `pkgdown`

[Rodríguez-Sánchez et al. 2016](#), [Marwick et al 2018](#), but see [McBain 2020](#)

# rrtools facilita la creación de compendios

```
library("rrtools")

use_compendium("~/mynewproject/") # crea proyecto

use_readme_rmd() # crea README

use_analysis() # crea carpeta con manuscrito Rmd

use_dockerfile() # crea Dockerfile

use_travis() # Usa Travis para integración continua

use_testthat() # Usa tests para las funciones
```

# Integración continua: chequeo continuo tras cada ‘commit’

Travis-CI, Circle-CI, GitHub Actions...

Travis CI   About Us   Blog   Status   Help   Sign in with GitHub

Help make Open Source a better place and start building better software today!

## Pakillo / Carex.bipolar

build passing

Current	Branches	Build History	Pull Requests	More options
✓ master	Pakillo	add two more articles to pkgdown	→ #7 passed → 1c006ff ↗ 3 min 22 sec a day ago	
✓ master	Pakillo	added leaflet occurrence maps to appear as a	→ #6 passed → 57f5374 ↗ 5 min 23 sec a day ago	
✓ master	Pakillo	build site with pkgdown	→ #5 passed → 6108a7a ↗ 17 min 35 sec a day ago	
✗ master	Pakillo	still trying to fix error with sf in travis (via rnat)	→ #4 failed → 2c922d4 ↗ 16 min 58 sec 2 days ago	
master	Pakillo	adding more sf dependencies to travis	→ #3 errored → 5a60b49 ↗ 13 min 59 sec 2 days ago	
master	Pakillo	trying to fix error with rgdal on travis	→ #2 errored → 076af29 ↗ 14 min 15 sec 2 days ago	
master	Pakillo	add travis	→ #1 errored → 4bc6e8 ↗ 18 min 54 sec 3 days ago	

# rrtools.addin te guía para crear un nuevo proyecto

The screenshot shows the rrtools Configuration Assistant interface. At the top, there's a search bar labeled "Ask for help or report a bug". Below it, the title "rrtools Configuration Assistant" is displayed. The main content area is titled "Create a research compendium". A red box highlights the "Create an empty, git initialized directory" section, which contains instructions about managing the project with Git and creating a local repository. To the right, a panel for the "use\_compendium" command is shown, with tabs for "use\_compendium" and "R Documentation". The "R Documentation" tab provides a brief description of the command. Below that is a "Description" section and a "Usage" section with a code example. At the bottom, there are navigation icons for Overview, Setup, License, Readme, File structure, Virtualisation, CI, and Tests.

A new project!

Each reproducible research project lives in its own directory on your computer. This directory needs some special files and subfolders to work as an [R package](#). The first step in the creation of a research compendium is to create and configure all of this.

The function `rrtools::use_compendium`

- Creates an R package "directory-name" in a directory
- Writes and Preconfigures R package files
- Opens project in RStudio

Create a research compendium

**Create an empty, git initialized directory**

Usually we want our research compendium to be managed by the version control software [Git](#). Start on Github, Gitlab, or a similar web service, and [create an empty repository](#) called pkgnname (you should use a different name, please follow the rules below) on that service. Then clone that repository to have a local empty directory on your computer, called pkgnname, that is linked to this remote repository.

The name of your project has to be a good representation of its content and also has to fulfil some technical requirements. See [Hadley Wickham's helpful hints](#) for advice. Jim Hesters R package [available](#) makes it easy to check if your name is already in use

use\_compendium

Creates an R package suitable to use as a research compendium, and switches to the working directory of this new package, ready to work

R Documentation

Description

This is useful: `create_package()` with some additional messages to simplify the transition into the new project setting

Usage

```
use_compendium(path = getwd(), field = "compendium")
```

?

Overview

Setup

License

Readme

File structure

Virtualisation

CI

Tests

<https://github.com/nevrome/rrtools.addin>

# Estructura del proyecto

```
- README
- LICENSE
- DESCRIPTION
- travis.yml
- Dockerfile
- analysis/
|
|- paper/
 |- paper.Rmd
 |- references.bib
|
|- figures/
|
|- data/
 |- raw_data/
 |- derived_data/
```

# Estructura del proyecto con funciones

```
- README
- LICENSE
- DESCRIPTION
- travis.yml
- Dockerfile
- analysis/
|
|- paper/
 |- paper.Rmd
 |- references.bib
|
|- figures/
|
|- data/
 |- raw_data/
 |- derived_data/
- R/
 |-functions.R
- man/
```

# workflowr: proyectos reproducibles con website

```
library("workflowr")
wflow_start("~/newproject")
```



## Recapitulación

---

# Recapitulación

- Archivar DATOS + CÓDIGO

# Recapitulación

- Archivar DATOS + CÓDIGO
- Rmarkdown: integra análisis + texto en documento dinámico

# Recapitulación

- Archivar DATOS + CÓDIGO
- Rmarkdown: integra análisis + texto en documento dinámico
- Control de operaciones: `makefile`, `drake`/`targets`

# Recapitulación

- Archivar DATOS + CÓDIGO
- Rmarkdown: integra análisis + texto en documento dinámico
- Control de operaciones: `makefile`, `drake`/`targets`
- Dependencias: `sessionInfo` -> paquetes R -> Docker

# Recapitulación

- Archivar DATOS + CÓDIGO
- Rmarkdown: integra análisis + texto en documento dinámico
- Control de operaciones: `makefile`, `drake/targets`
- Dependencias: `sessionInfo` -> paquetes R -> Docker
- Estructura consistente de proyectos: `rrtools`, `workflowr`...

