

Quiénes somos

María Hurtado de Mendoza Romo

- FPI Plan Estatal, ecología (INMAR y EBD).
- Github: <https://github.com/MariaHdMR>

Sergio Picó Jorda

- FPI UCA, ecología (INMAR).
- Github: <https://github.com/spicjor>

Repositorio curso

https://github.com/spicjor/reproducible_uca23

En este repositorio podéis encontrar todos los materiales que vamos a usar durante el curso.

Objetivos

1. Entender los motivos de la crisis de reproducibilidad en ciencia.
2. Conocer los fundamentos de la ciencia reproducible.
3. Aprender a usar los elementos básicos de un flujo de trabajo reproducible.
4. Aprender el uso básico del control de versión con Git y GitHub.

IS THERE A REPRODUCIBILITY CRISIS?

A *Nature* survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help.

BY MONYA BAKER

1,576

RESEARCHERS SURVEYED

52%
Yes, a significant
crisis

38%
Yes, a slight
crisis

7%
Don't know

3%
No, there is no crisis

1. La crisis de reproducibilidad

¿Qué es la crisis de reproducibilidad?

De una encuesta a 1576 investigadores en Nature (2016):

“ More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. ”

Baker, M., Penny, D. (2016). Is there a reproducibility crisis? *Nature*, 533, 452-454.

Business as usual

Lo tradicional: Solo el artículo y los resultados que los autores eligen se publican.

ON THE RELATIVE ABUNDANCE OF SPECIES

ROBERT MACARTHUR

Department of Zoology, University of Pennsylvania,
Philadelphia, Pennsylvania

This paper will contain a discussion of the ecological consequences which can be deduced from data on the comparative abundances of species found together.

Let $N_i(t)$ be the abundance of the i -th species at time t . Then if $r_i(t)$ is defined by

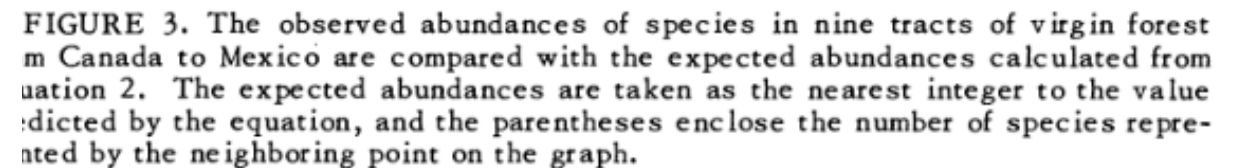
$$r_i(t) = \frac{1}{N_i(t)} \frac{dN_i(t)}{dt},$$

integrating, we obtain

$$(1) \quad \log N_i(t) = \log N_i(0) + \int_0^t r_i(t) dt.$$

Notice that r is permitted to vary. There are two opposed schools in ecology, one maintaining that the integral in equation (1) is important compared with $\log N_i(0)$ and the other maintaining that it is unimportant. These two views may lead to different ideas of the relative sizes of the $N_i(t)$ and so they will be discussed separately.

No datos brutos y métodos poco detallados = irreproducible.



Efectos no solo académicos

Errores que se propagan a la política económica mundial (5000 citas en Google Scholar).

Growth in a Time of Debt

Carmen M. Reinhart & Kenneth S. Rogoff

WID
R 15639

DOI 10.3386/w15639

ISSUE DATE January 2010

REVISION DATE December 2011

...y economic growth and inflation at different levels of government and external debt. Our
...is based on new data on forty-four countries spanning about two hundred years. The
...incorporates over 3,700 annual observations covering a wide range of political systems,
...ns, exchange rate arrangements, and historic circumstances. Our main findings are: First,
...ionship between government debt and real GDP growth is weak for debt/GDP ratios below
...old of 90 percent of GDP. Above 90 percent, median growth rates fall by one percent, and

Efectos no solo académicos

Un error en una hoja de Excel con tremendas consecuencias.

Si se hubiera seguido una metodología reproducible, probablemente se habría evitado.

Does High Public Debt Consistently
Stifle Economic Growth?
A Critique of Reinhart and Rogoff

Thomas Herndon, Michael Ash and Robert Pollin



Muchos más ejemplos

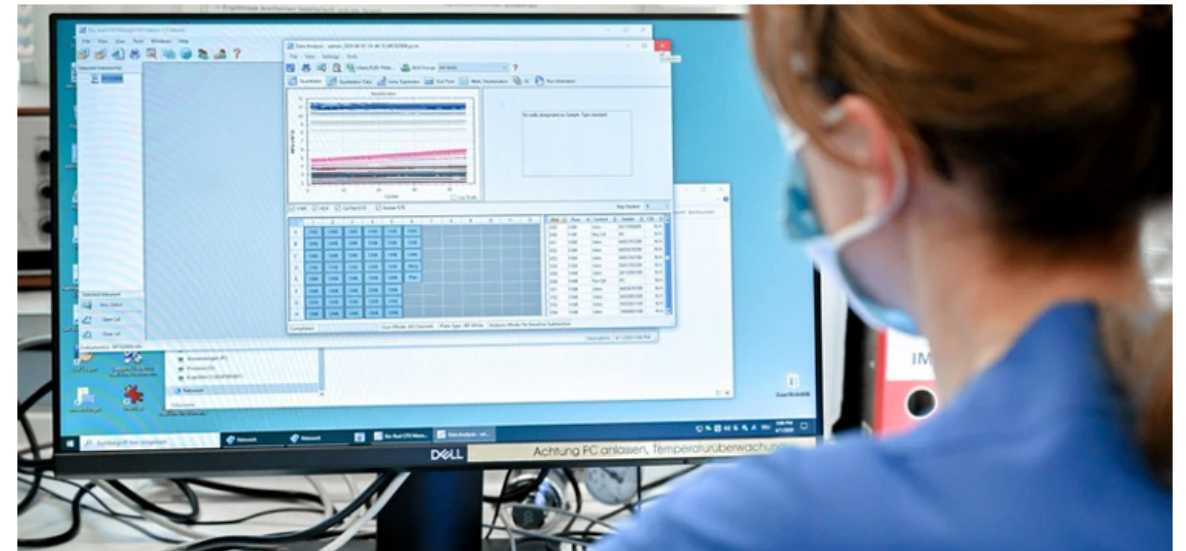
Excel corrigiendo nombres de genes en miles de artículos.

NEWS | 13 August 2021 | Correction [25 August 2021](#)

Autocorrect errors in Excel still creating genomics headache

Despite geneticists being warned about spreadsheet problems, 30% of published papers contain mangled gene names in supplementary data.

[Dyani Lewis](#)



Muchos más ejemplos

Fabricación de datos detectada gracias al requisito de publicar los datos brutos.

NEWS | 07 February 2020 | Clarification [13 February 2020](#)

'Avalanche' of spider-paper retractions shakes behavioural-ecology community

Allegations of fabricated data have prompted a university investigation and some soul-searching.

[Giuliana Viglione](#)



2. Ciencia reproducible

2.1 Qué es la ciencia reproducible

Reproducible (reproducibility, computational reproducibility):

"Un estudio es reproducible si el texto del artículo viene acompañado de código... ...que permite recrear exactamente a partir de los datos originales todos los resultados y figuras incluidos en el artículo."

vs Repetible (replicability):

"...replicar el mismo estudio (con nuevos datos) a partir de la información proporcionada en el artículo."

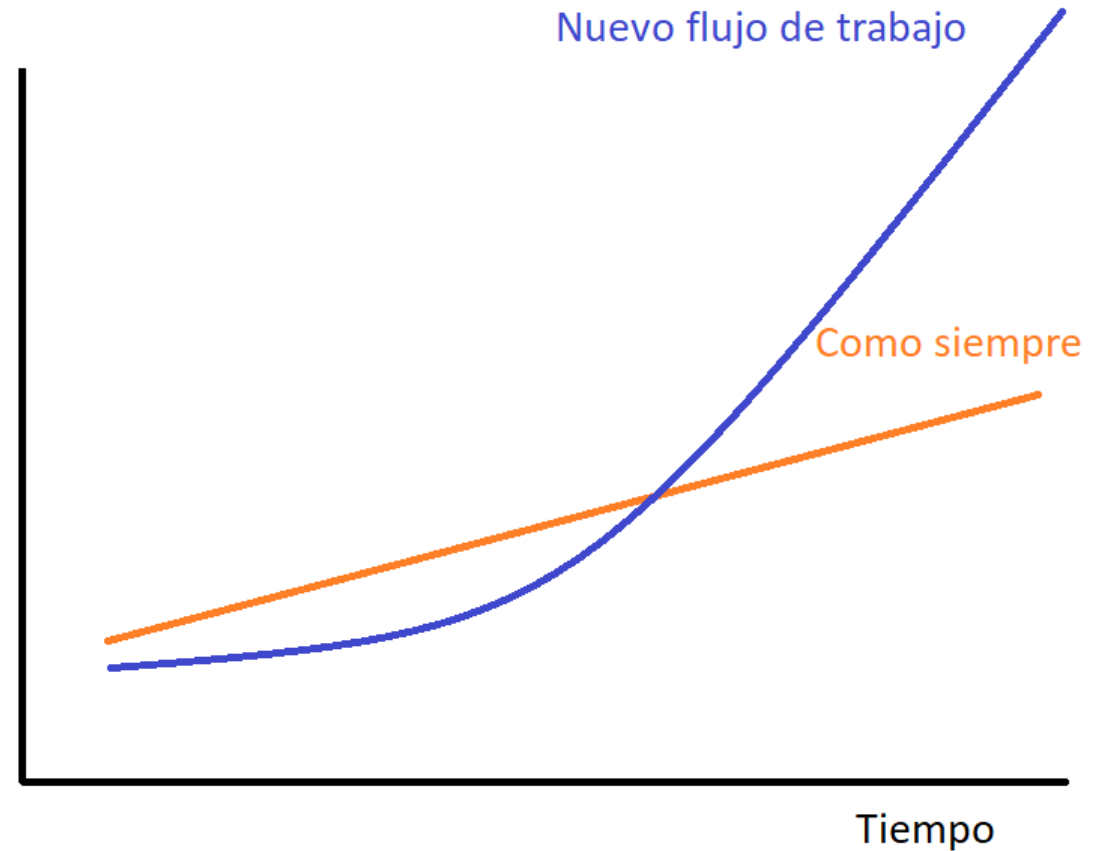
Rodríguez-Sánchez, F., Pérez-Luque, A.J. Bartomeus, I., Varela, S.
2016. Ciencia reproducible: qué, por qué, cómo. Ecosistemas 25(2):
83-92. Doi.: 10.7818/ECOS.2016.25-2.11

2.2 Por qué es necesaria

1. Garantía de transparencia y calidad. Menos errores y estos son detectables y corregibles.
2. Poder reutilizar código ayuda a acelerar el progreso científico.
3. Datos y código requeridos por revistas y entidades financiadoras.
4. Beneficios personales: ahorro de tiempo y esfuerzo, facilita la colaboración, signo de calidad, más probabilidad de citas, etc.

2.3 Cómo implementarla?

Necesitamos cambiar la manera en que trabajamos por un **flujo de trabajo reproducible**.



3. El flujo de trabajo reproducible

3.1. Buenas prácticas con datos y código

El principio: los datos

Planificar: Cómo se van a obtener, recoger, almacenar y compartir los datos.

Recolectar: Intentar conservar **siempre** los datos brutos en el formato original.

Metadatos: Descripción detallada de lo que representa cada variable, cómo se tomó, en qué unidades está, las personas encargadas, etc.

Control de calidad: Plantillas o formularios ayudan a evitar errores al introducir los datos. Ser consistente en los códigos y la estructura.

Preservación: Asegurar que nuestros datos seguirán disponibles a largo plazo. Archivo de formato público (txt o csv por ejemplo) en un repositorio con DOI (DRYAD, Zenodo o figshare por ejemplo). Estos datos muchas veces se pueden vincular con el artículo donde se han estudiado estos datos.

Abundance	Biomass	Family	Genus	Species	SampleDescr	Plot	Latitude	Longitude	DepthElevati	Day	Month	Year	StudyID
4.07	NA	Aegithalidae	Aegithalos	caudatus	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
1.84	NA	Certhiidae	Certhia	brachydactylus	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.52	NA	Corvidae	Garrulus	glandarius	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.39	NA	Emberizidae	Emberiza	caesia	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.26	NA	Fringillidae	Chloris	chloris	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
1.57	NA	Fringillidae	Fringilla	coelebs	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.26	NA	Fringillidae	Pyrrhula	pyrrhula	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
8.66	NA	Muscicapidae	Erithacus	rubecula	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
8.14	NA	Paridae	Cyanistes	caeruleus	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
2.62	NA	Paridae	Lophophanes	cristatus	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.92	NA	Paridae	Parus	major	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.66	NA	Paridae	Periparus	ater	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.66	NA	Phylloscopid	Phylloscopus	collybita	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.13	NA	Picidae	Dendrocopos	major	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.26	NA	Prunellidae	Prunella	modularis	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
3.28	NA	Regulidae	Regulus	ignicapilla	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.13	NA	Sittidae	Sitta	europaea	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
3.28	NA	Sylviidae	Sylvia	atricapilla	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.92	NA	Sylviidae	Sylvia	melanocephala	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.13	NA	Sylviidae	Sylvia	undata	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
0.26	NA	Troglodytidae	Troglodytes	troglodytes	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
1.97	NA	Turdidae	Turdus	merula	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA
4.37	NA	Aegithalidae	Aegithalos	caudatus	Cazorla_birds	NA	37.94	-2.87	NA	NA	NA	1978	NA

Buenas prácticas escribiendo código

- Antes de empezar:
 - Estructura de carpetas
 - Añadir Readme
 - Añadir licencia
- Durante:
 - COMENTAR, COMENTAR, COMENTAR Y COMENTAR.
 - Citar software y paquetes (con versión!).

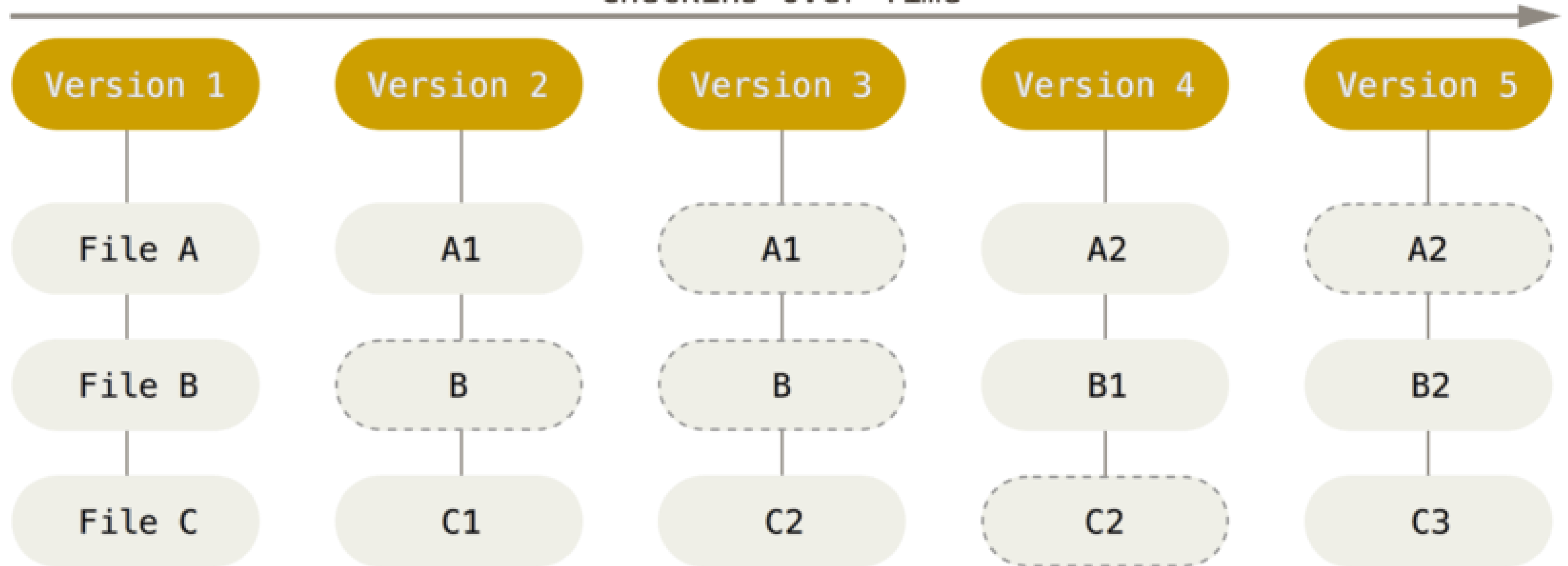
3.2 Control de versión con Git





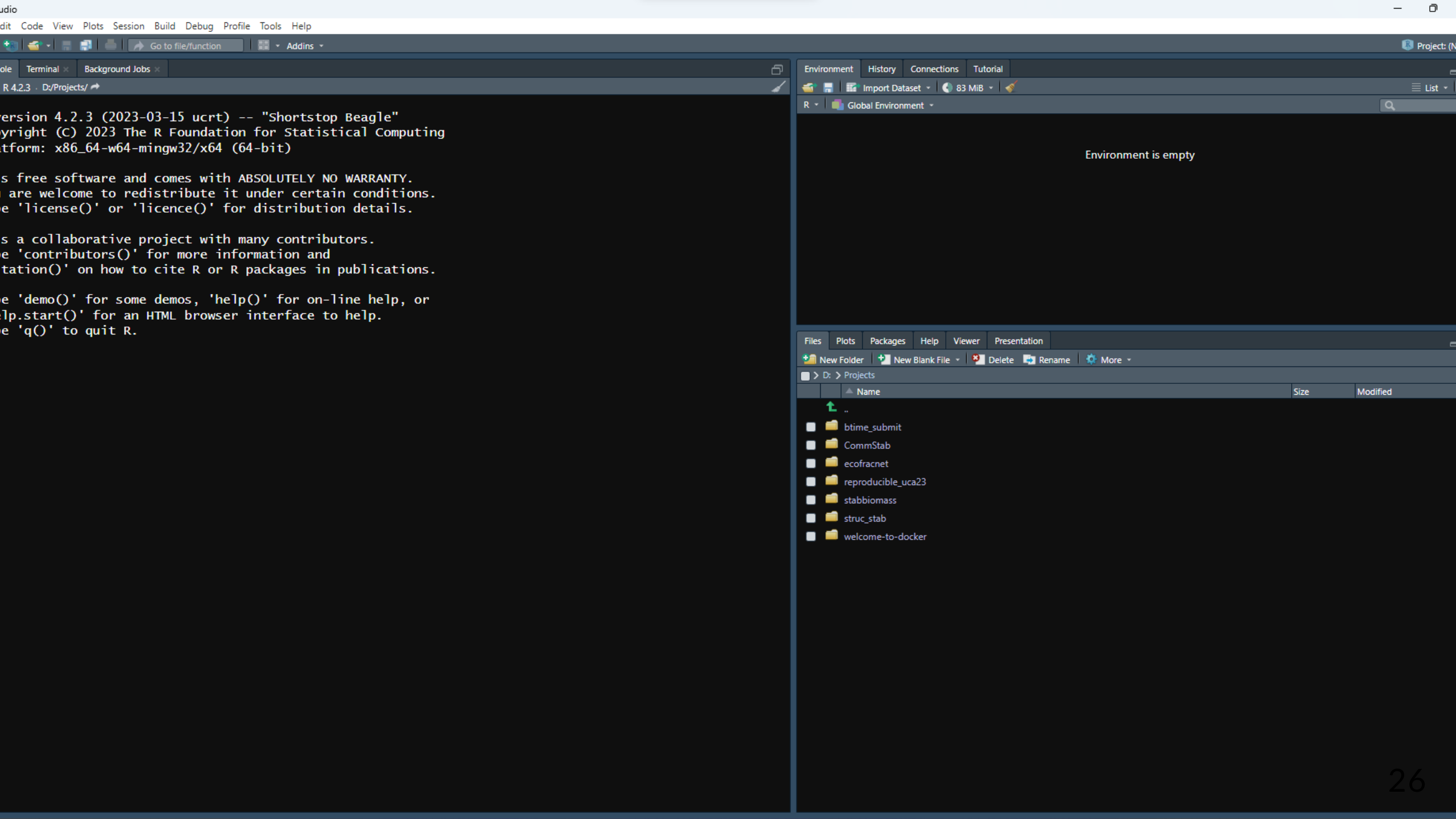
git

Checkins Over Time



3.3 Integración con RStudio y Github





```
R version 4.2.3 (2023-03-15 ucrt) -- "Shortstop Beagle"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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

R is a collaborative project with many contributors.
For more information and contribution guidelines see the 'contributors()'
function, or 'help(contributing)' for more information and
documentation on how to cite R or R packages in publications.

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

EnvironmentHistoryConnectionsTutorial

Import Dataset83 MiB

RGlobal Environment

Environment is empty

FilesPlotsPackagesHelpViewerPresentation

New FolderNew Blank FileDeleteRenameMore

D: > Projects

Name	Size	Modified
..		
btime_submit		
CommStab		
ecofracnet		
reproducible_uca23		
stabbiomass		
struc_stab		
welcome-to-docker		

GitHub



spicjor Advancing ppt

8991ab9 2 days ago 26 commits



ppt

Advancing ppt

2 days ago



.gitignore

Updated gitignore

last month



LICENSE

Initial commit

3 months ago



README.md

Updated readme

2 days ago

[README.md](#)

Curso ciencia reproducible - Introducción a Git y GitHub. UCA 2023

Responsables:

Sergio Picó Jordá - INMAR (UCA)

María Hurtado de Mendoza Romo - INMAR (UCA)

Justificación

La ciencia lleva tiempo inmersa en una crisis de reproducibilidad. La presión por publicar favorece la proliferación de malas prácticas que hacen que se multipliquen los estudios en los que es muy difícil o imposible reproducir el proceso de obtención de los resultados, entorpeciendo el avance científico. Como respuesta, surgió el movimiento

Curso ciencia reproducible - Introducción a Git y GitHub. UCA 2023

[Readme](#)[MIT license](#)[Activity](#)[0 stars](#)[1 watching](#)[0 forks](#)

Releases

No releases published

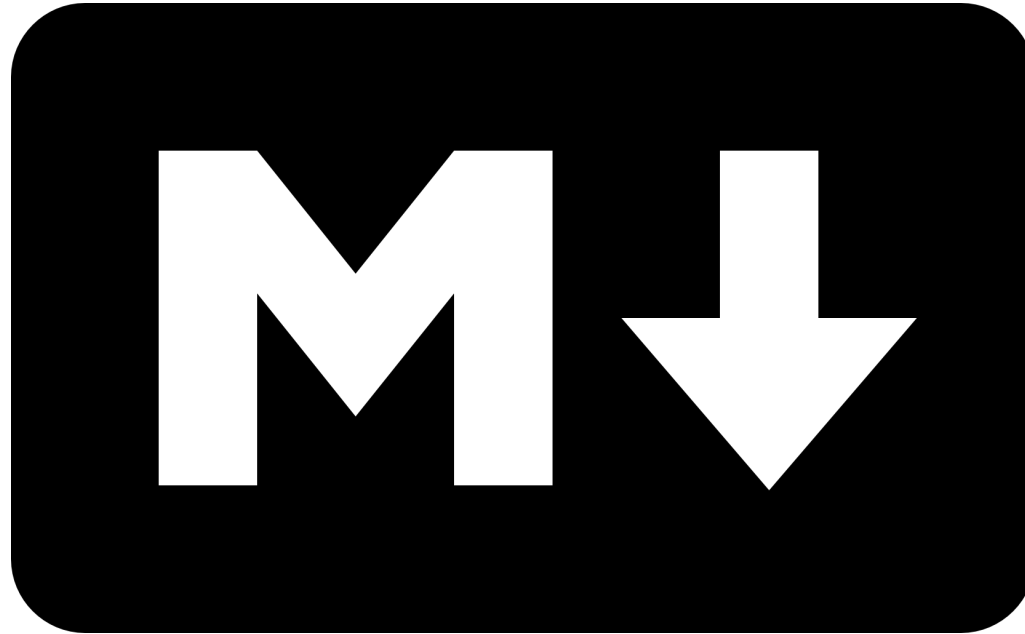
[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

3.4 Redacción de artículos reproducibles



ejemplo_proyecto > Informe_reptiles.Rmd > ...

```
1 ---
2 title: "Seguimiento a largo plazo de reptiles en
3         Doñana 2005-2021"
4 author: Sergio Picó Jordá^[Universidad de Cádiz,
5         sergio.pico@uca.es],
6         María Hurtado de Mendoza Romo^[Universidad de
7         Cádiz, maria.hurtadodemendoza@uca.es]
8 output: html_document
9 ---
10 ```{r setup, include = FALSE}
11
12 knitr::opts_chunk$set(echo = FALSE, warning = FALSE,
13                       message = FALSE)
14
15 # Cargar los paquetes necesarios
16 library(readr)
17 library(tidyr)
18 library(dplyr)
19 library(terra)
20 library(stringr)
```