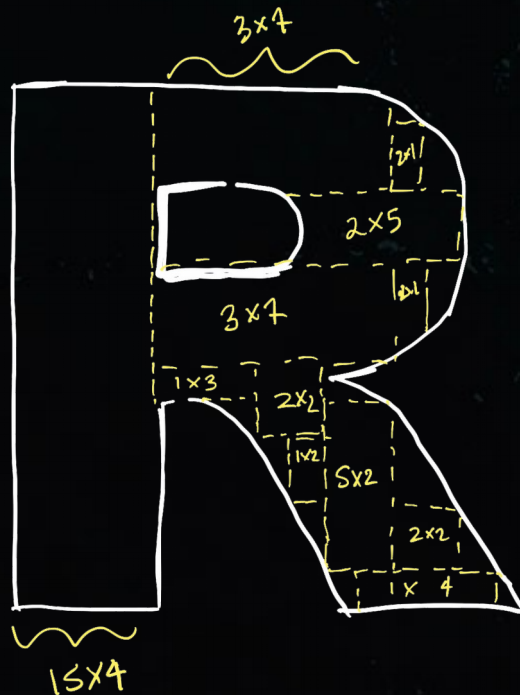
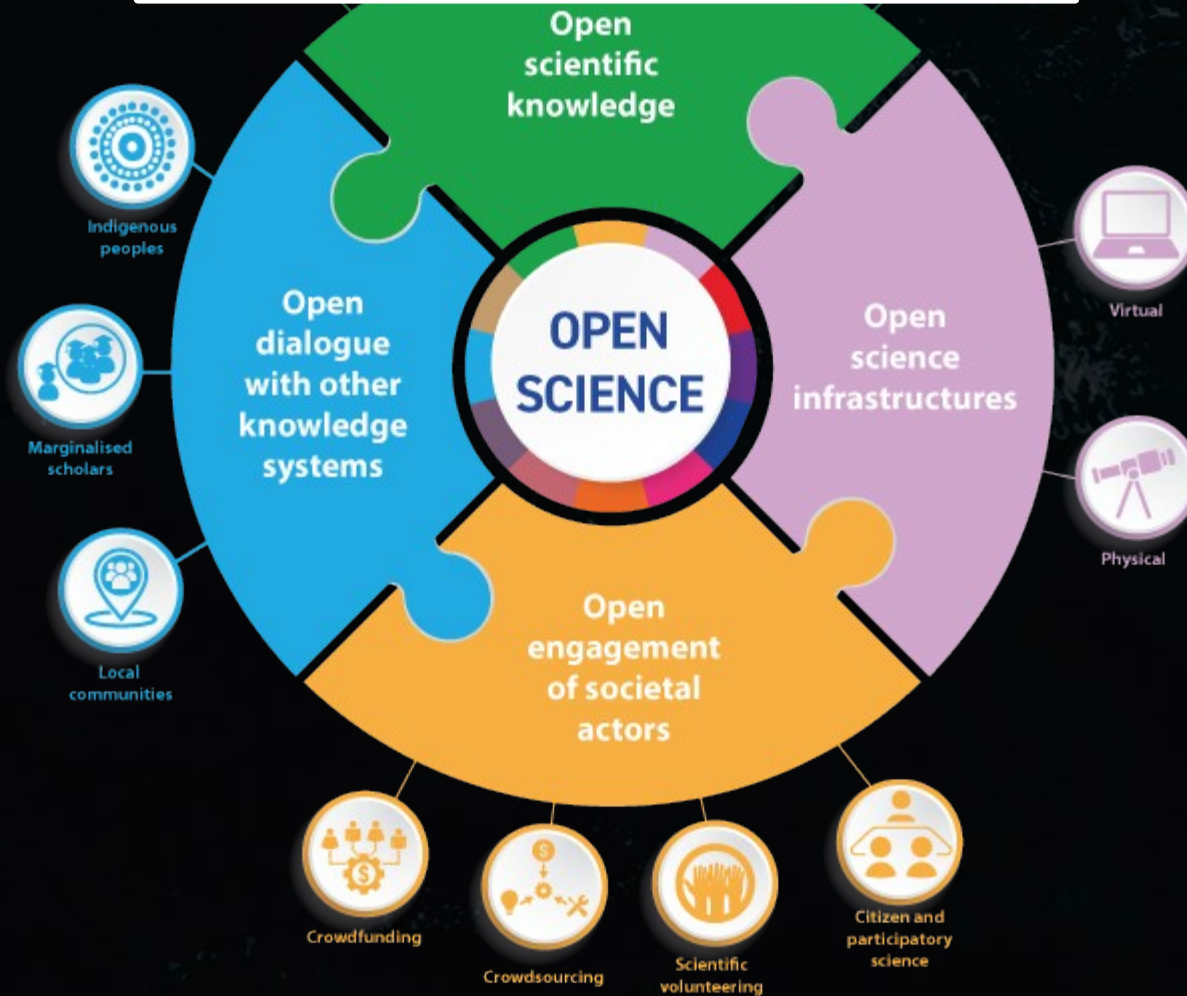
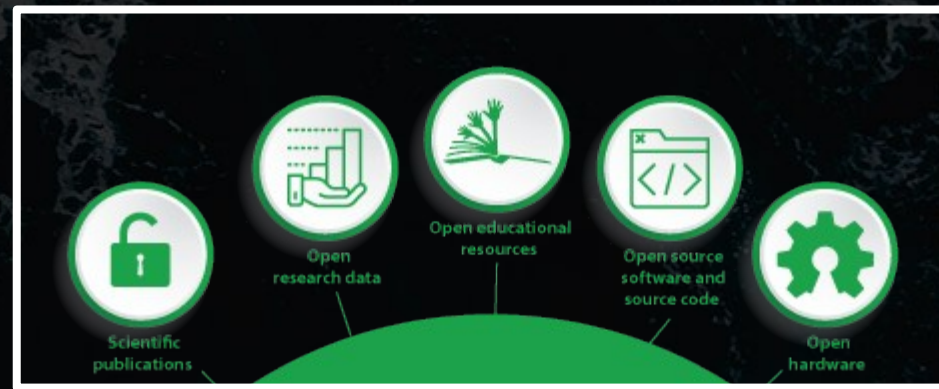


Datos y Proyectos

Martín Amodeo
DBBF, UNS
IADO CONICET-UNS
CCT Bahía Blanca





- Set of principles and practices that aim to make scientific research from all fields accessible to everyone for the benefit of scientists and society as a whole.
- ... not only that scientific knowledge is accessible but also that the production of that knowledge itself is inclusive, equitable and sustainable.
- Opening science involves opening up the processes of creating, evaluating, sharing, exploring and storing scientific knowledge, practices and perspectives.
- make multilingual scientific knowledge openly available, accessible and reusable for everyone
- increase scientific collaborations and sharing of information for the benefits of science and society, and to
- open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community.





Ciencia Reproducible

A E E T

ASOCIACIÓN ESPAÑOLA
DE ECOLOGÍA TERRESTRE

Ecosistemas 25(2): 83-92 [Mayo-Agosto 2016]

Doi.: 10.7818/ECOS.2016.25-2.11

Artículo publicado en Open Access bajo los términos
de Creative Commons attribution Non Comercial License 3.0.

REVISIONES

ecosistemas

REVISTA CIENTÍFICA DE ECOLOGÍA Y MEDIO AMBIENTE

ISSN 1697-2473 / Open access
disponible en www.revistaecosistemas.net

Ciencia reproducible: qué, por qué, cómo

F. Rodríguez-Sánchez^{1,*}, A.J. Pérez-Luque^{2,**}, I. Bartomeus^{1,**}, S. Varela^{3,**}

(1) Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD-CSIC), Consejo Superior de Investigaciones Científicas, Avda. Américo Vespucio s/n, E-41092 Sevilla, España.

(2) Laboratorio de Ecología (iEcolab), Instituto Interuniversitario Sistema Tierra (CEAMA), Universidad de Granada, Avda. del Mediterráneo s/n, Granada 18006, España.

(3) Departamento de Ciencias de la Vida, Facultad de Biología, Ciencias Ambientales y Química, Universidad de Alcalá, Campus Universitario. Ctra. Madrid-Barcelona, Km. 33,600, 28805 Alcalá de Henares, Madrid, España.

<https://doi.org/10.7818/ECOS.2016.25-2.11>

- Los estudios científicos reproducibles incluyen código informático capaz de recrear todos los resultados a partir de los datos originales.
- Acelera el progreso científico
- Ahorro de tiempo y esfuerzo de los/as investigadores/as
- Incremento de la calidad e impacto de sus publicaciones

A E E T
ASOCIACIÓN ESPAÑOLA
DE ECOLOGÍA TERRESTRE

Ecosistemas 25(2): 83-92 [Mayo-Agosto 2016]
Doi: 10.7818/ECOS.2016.25-2.11

Artículo publicado en Open Access bajo los términos
de Creative Commons attribution Non Commercial License 3.0.

REVISIONES

ecosistemas
REVISTA CIENTÍFICA DE ECOLOGÍA Y MEDIO AMBIENTE

ISSN 1697-2473 / Open access
disponible en www.revistaecosistemas.net

Ciencia reproducible: qué, por qué, cómo

F. Rodríguez-Sánchez^{1,*}, A.J. Pérez-Luque^{2,**}, I. Bartomeus^{1,**}, S. Varela^{3,**}

(1) Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD-CSIC), Consejo Superior de Investigaciones Científicas, Avda. Américo Vespucio s/n, E-41092 Sevilla, España.
(2) Laboratorio de Ecología (IEcolab), Instituto Interuniversitario Sistema Tierra (CEAMA), Universidad de Granada, Avda. del Mediterráneo s/n, Granada 18006, España.
(3) Departamento de Ciencias de la Vida, Facultad de Biología, Ciencias Ambientales y Química, Universidad de Alcalá, Campus Universitario. Ctra. Madrid-Barcelona, Km. 33,600, 28805 Alcalá de Henares, Madrid, España.



Repositorios de datos y código



v_1	v_2	v_3	v_4
\vdots \vdots \vdots $-$ $-$ $-$	$-$ $-$ $-$ $-$ \vdots $-$	$-$ $-$ $-$ $-$ \vdots \vdots $-$	$-$ $-$ $-$ $-$ \vdots \vdots $-$

	A	B	C	D	E	F
1	fecha	n.obs	vel.max	vel.media	dir	
2	2013-06-30 10:50:00	612	45.6	27.7	NW	
3	2013-07-07 10:50:00	608	33.3	19.1	NW	
4	2013-08-08 10:50:00	514	NA	30.5	NW	
5	2013-08-17 10:50:00	508	NA	33.7	NW	
6	2013-09-25 10:50:00	567	50.6	3.3	NW	
7	2013-10-20 10:50:00	539	66.2	21.9	NW	
8	2013-11-21 10:50:00	430	93.1	38.4	NA	
9	2013-12-07 10:50:00	584	52.4	24.7	NA	
10	2013-12-23 10:50:00	588	53.5	18.1	NW	
11	2014-04-21 10:50:00	539	26.8	14.4	NW	
12	2014-06-17 10:50:00	454	NA	NA	NA	
13	2014-06-24 10:50:00	394	NA	NA	NA	
14	2014-07-10 10:50:00	654	NA	NA	NA	
15	2014-07-26 10:50:00	568	NA	NA	NA	
16	2014-08-04 10:50:00	641	42.4	32.9	NW	
17	2014-08-20 10:50:00	125	54.4	16.6	NW	
18	2014-08-27 10:50:00	109	16.5	7.4	NW	
19	2014-09-12 10:50:00	72	49.5	17.5	NW	
20	2014-11-15 10:50:00	554	41.3	7.4	NW	
21	2014-11-24 10:50:00	206	45.4	NA	NA	
22	2014-12-01 10:50:00	622	40.5	16.6	NW	
23						
24						
25						
26						
27						

Planilla interpretable

Simple, plana, completa, con NA

- Respetar Reglas Básicas para Nombres
- Unificar rótulos
("cuatr", "cuatr.", "cuatrerros", "cuatrero", "Cuatrerros")
- Chequear errores de tipeo y mayúsculas/minúsculas
- Sin columnas ni filas vacías
- Sin medidas de resumen, tablitas, grafiquitos....
- SOLO DATOS CRUDOS
- Definir los datos faltantes como NA

Planilla interpretable

NA

Dato faltante

- Se perdió la muestra,
- Se perdió el número
- No se hizo ese punto
- El equipo tiró error

Cero

- Hay dato y el valor es CERO

Funciones *read*

```
read.table(file, header = FALSE, sep = "", quote = "\"'",
           dec = ".", numerals = c("allow.loss", "warn.loss", "no.loss"),
           row.names, col.names, as.is = !stringsAsFactors,
           na.strings = "NA", colClasses = NA, nrows = -1,
           skip = 0, check.names = TRUE, fill = !blank.lines.skip,
           strip.white = FALSE, blank.lines.skip = TRUE,
           comment.char = "#",
           allowEscapes = FALSE, flush = FALSE,
           stringsAsFactors = FALSE,
           fileEncoding = "", encoding = "unknown", text, skipNul = FALSE)

read.csv(file, header = TRUE, sep = ",", quote = "\"",
         dec = ".", fill = TRUE, comment.char = "", ...)

read.csv2(file, header = TRUE, sep = ";", quote = "\"",
          dec = ",", fill = TRUE, comment.char = "", ...)

read.delim(file, header = TRUE, sep = "\t", quote = "\"",
           dec = ".", fill = TRUE, comment.char = "", ...)

read.delim2(file, header = TRUE, sep = "\t", quote = "\"",
            dec = ",", fill = TRUE, comment.char = "", ...)
```

Ayuda de *read.table()*

Formatos y funciones *read*

- .txt (genérico)
- .csv (genérico)
- .xlsx (Office)
- .ods (Libreoffice)
- read.delim()
- read.csv()
- read_xlsx() {readxl}
- read_ods() {readODS}

Y más... readr, read_open, read_sheet... y más... y más... y más...

Funcion *read.csv()*

base

```
read.csv(file, header = TRUE, sep = ",", quote = "\"",  
         dec = ".", fill = TRUE, comment.char = "", ...)
```

`read.csv()`

Argumentos más usados:

- `file` = "micarpeta/miarchivo.csv"
- `dec` = "." ó ","
- `sep` = separado por ",", " " ó "[tab]" ó por ";"?
- `na.strings` = "NA"

Ojo! Excel graba con ; aunque dice "separado por comas"
Libreoffice es más serio... Bill... media pila...

Funcion *read_xlsx()*

library(readxl)

```
read_excel(path, sheet = NULL, range = NULL, col_names = TRUE,  
  col_types = NULL, na = "", trim_ws = TRUE, skip = 0,  
  n_max = Inf, guess_max = min(1000, n_max),  
  progress = readxl_progress(), .name_repair = "unique")  
  
read_xls(path, sheet = NULL, range = NULL, col_names = TRUE,  
  col_types = NULL, na = "", trim_ws = TRUE, skip = 0,  
  n_max = Inf, guess_max = min(1000, n_max),  
  progress = readxl_progress(), .name_repair = "unique")  
  
read_xlsx(path, sheet = NULL, range = NULL, col_names = TRUE,  
  col_types = NULL, na = "", trim_ws = TRUE, skip = 0,  
  n_max = Inf, guess_max = min(1000, n_max),  
  progress = readxl_progress(), .name_repair = "unique")
```



Argumentos más usados:

- path = "micarpeta/miarchivo.csv"
- na = ""
- sheet = la hoja donde están mis datos



Funciones *readr*

`library(readr)`



`read_csv(), read_txt()...`

Funciones *write*

```
write.table(x, file = "", append = FALSE, quote = TRUE, sep = " ",  
           eol = "\n", na = "NA", dec = ".", row.names = TRUE,  
           col.names = TRUE, qmethod = c("escape", "double"),  
           fileEncoding = "")
```

```
write.csv(...)  
write.csv2(...)
```

```
write.csv(x= midataframe, file= "planilla_exportada.csv")
```

La extensión es parte del nombre del
archivo

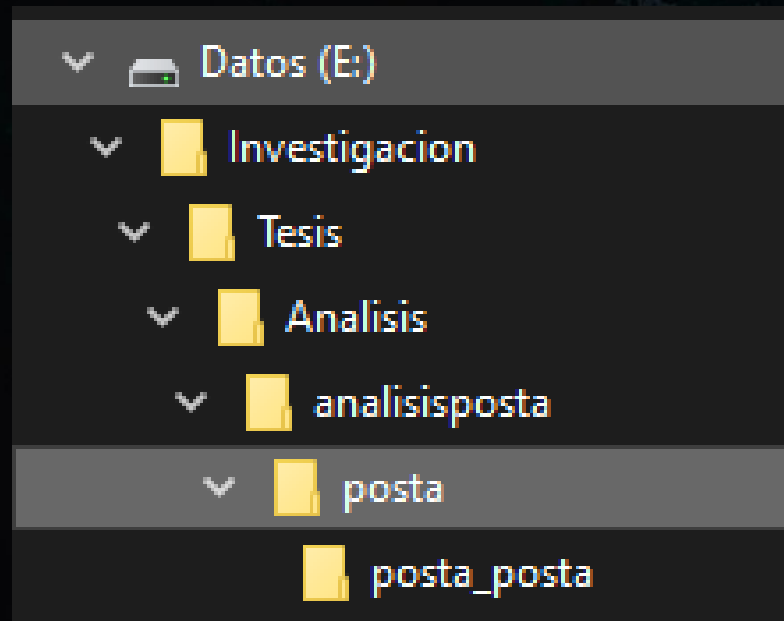
nombre.csv

nombre.xlsx

nombre.txt

Rutas

¿Dónde estamos?



Rutas

/

Rutas

Rutas Absolutas

E:/Investigacion/Tesis/
Analisis/analisisposta/datos

Se expresa desde la raíz del disco

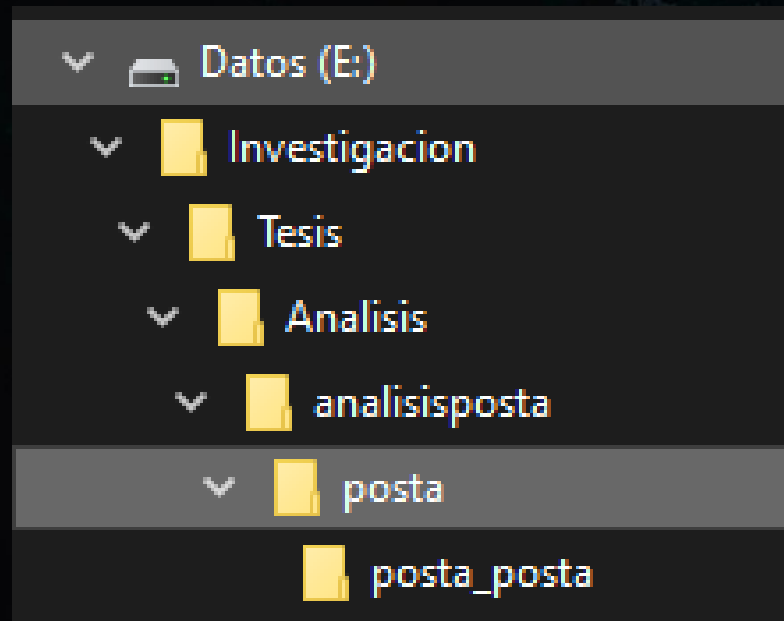
Rutas relativas

./datos

Se expresa desde una ubicación
definida previamente
(nuestro working directory)

Working directory

¿Dónde estamos?



Las rutas absolutas no son reproducibles

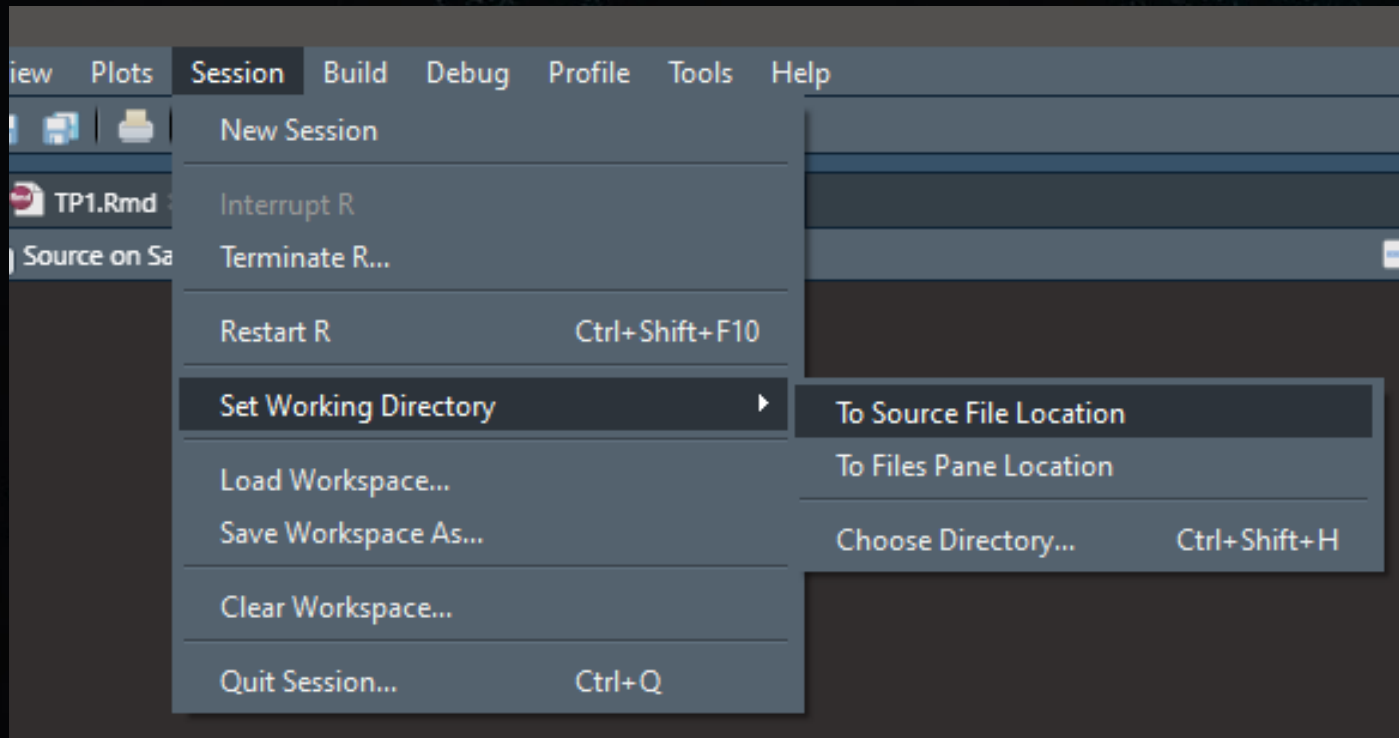
Evitemos esto...

E:/PC_mama_te_quiero/BillCAPO/Tesis/Analisis/
estafa_maestra/microsoft

Las rutas relativas son reproducibles

- `./Analisis/GLM/metales`
- El `"."` quiere decir "Donde está el *working directory*"
- ¿Cómo defino el *working directory*?
 - 1) A mano cuando abro Rstudio
 - 2) Usando `setwd(PROYECTOS!!!)`

1) A mano en Rstudio

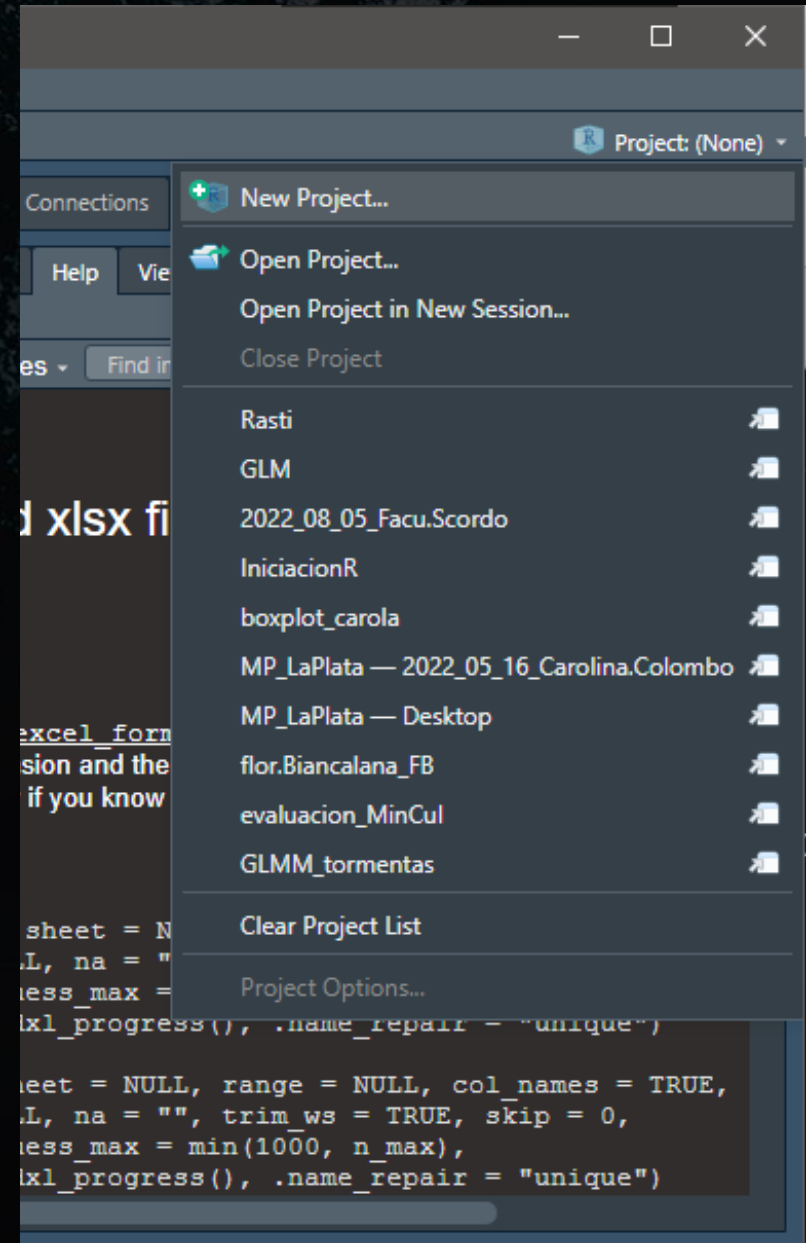


donde está el
← archivo de script .R
ó
← elegimos...

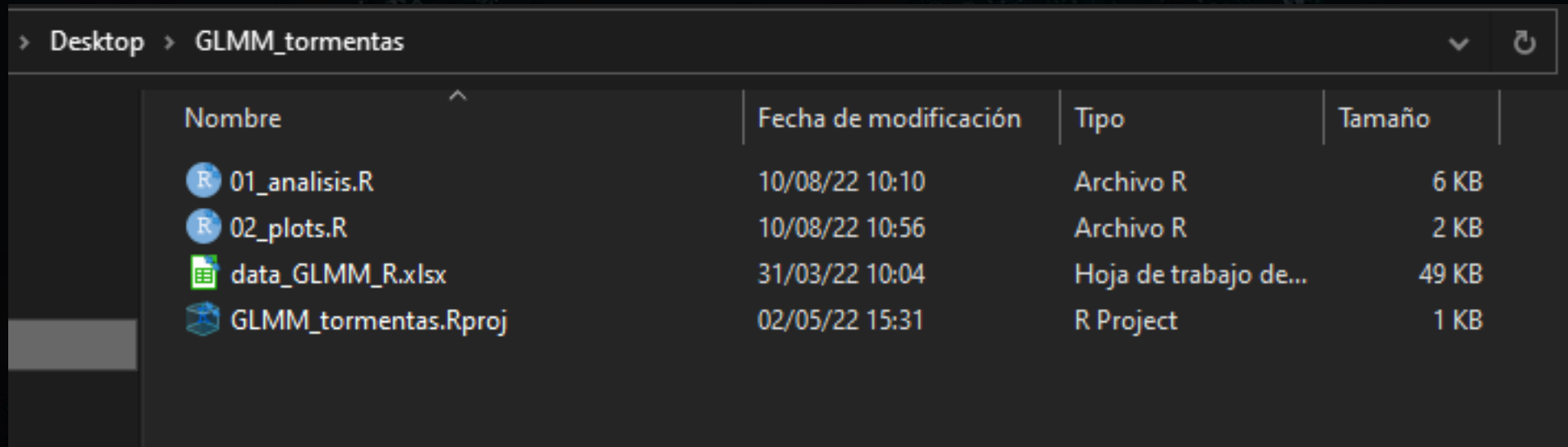
- Es práctico, rápido
- No es reproducible
- Mejor que usar rutas absolutas
- MAS ADELANTE: línea de comando para quien la quiera





2) Proyectos

- Crea un archivo .Rproj que señala el working directory
- Para trabajar ordenadamente, discontinuamente, compartir
- Reproducible y ordenado



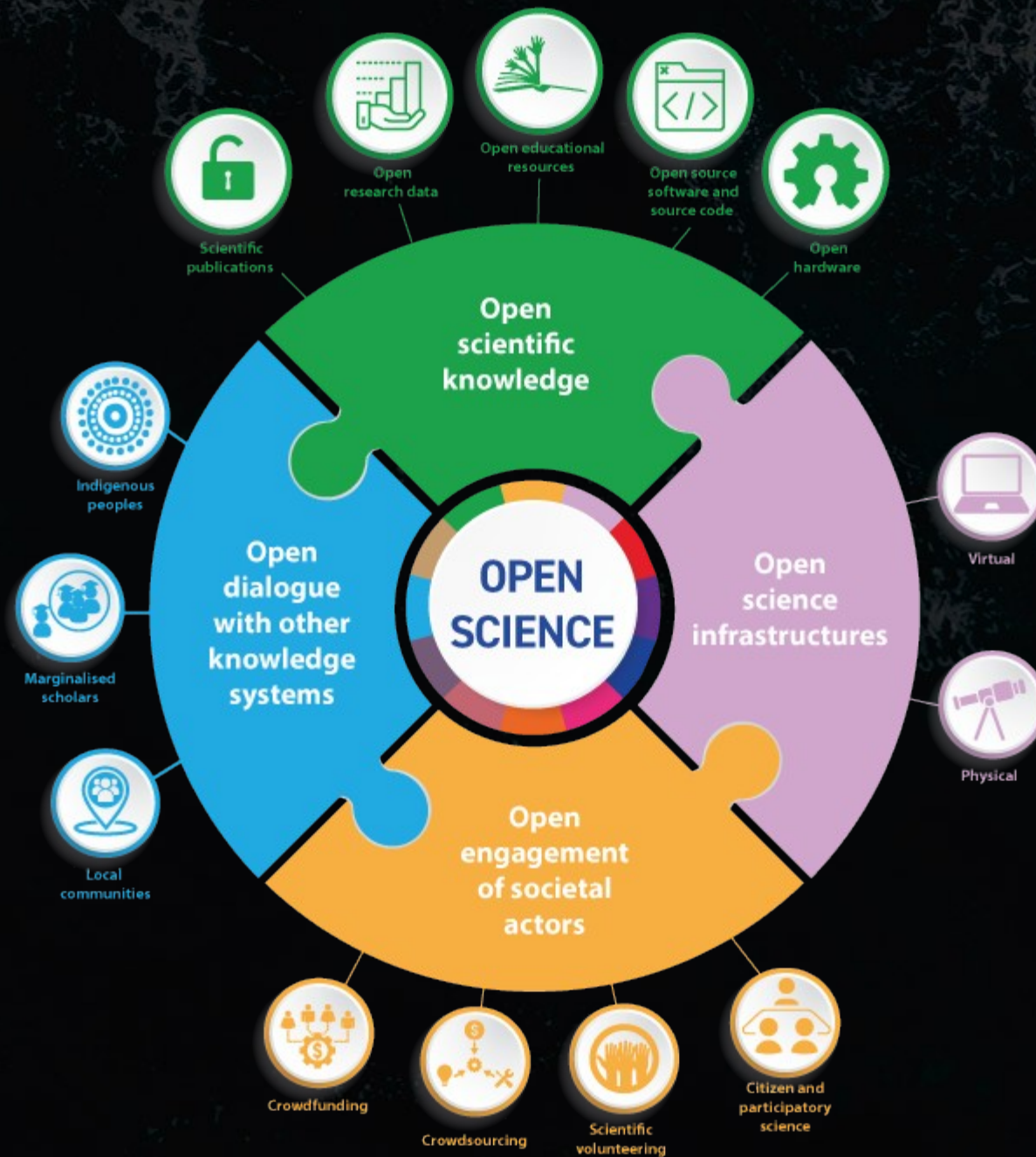
Proyectos .Rproj



Desktop > GLMM_tormentas				
Nombre	Fecha de modificación	Tipo	Tamaño	
 01_analisis.R	10/08/22 10:10	Archivo R	6 KB	
 02_plots.R	10/08/22 10:56	Archivo R	2 KB	
 data_GLMM_R.xlsx	31/03/22 10:04	Hoja de trabajo de...	49 KB	
 GLMM_tormentas.Rproj	02/05/22 15:31	R Project	1 KB	

- Archivo (.Rproj) ← Lo ejecuto para iniciar Rstudio (ubicado)
- Mis datos (.csv, .xlsx, .txt, ...)
- Mis scripts (.R)
- Puedo organizarlo usando subdirectorios

./data ./figs ./tabs ./tmp ./R ./analysis



Ciencia Reproducible

A E E T

ASOCIACIÓN ESPAÑOLA
DE ECOLOGÍA TERRESTRE

Ecosistemas 25(2): 83-92 [Mayo-Agosto 2016]

Doi.: 10.7818/ECOS.2016.25-2.11

Artículo publicado en Open Access bajo los términos
de Creative Commons attribution Non Comercial License 3.0.

REVISIONES

ecosistemas

REVISTA CIENTÍFICA DE ECOLOGÍA Y MEDIO AMBIENTE

ISSN 1697-2473 / Open access
disponible en www.revistaecosistemas.net

Ciencia reproducible: qué, por qué, cómo

F. Rodríguez-Sánchez^{1,*}, A.J. Pérez-Luque^{2,**}, I. Bartomeus^{1,**}, S. Varela^{3,**}

(1) Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD-CSIC), Consejo Superior de Investigaciones Científicas, Avda. Américo Vespucio s/n, E-41092 Sevilla, España.

(2) Laboratorio de Ecología (iEcolab), Instituto Interuniversitario Sistema Tierra (CEAMA), Universidad de Granada, Avda. del Mediterráneo s/n, Granada 18006, España.

(3) Departamento de Ciencias de la Vida, Facultad de Biología, Ciencias Ambientales y Química, Universidad de Alcalá, Campus Universitario. Ctra. Madrid-Barcelona, Km. 33,600, 28805 Alcalá de Henares, Madrid, España.

Paquetes de R como compendios de investigación

```
miproyecto
|- README           # información general
|- DESCRIPTION      # metadatos y dependencias del paquete
|- NAMESPACE       # generado automáticamente
|- makefile         # script maestro que ejecuta el análisis
|- data/            # datos depurados
|- data-raw/        # datos brutos
|- R/               # definición de funciones
|- man/             # documentación de las funciones
|- tests/           # tests de las funciones
|- analysis/        # ficheros Rmarkdown, figuras, etc.
```

zenodo



OSFHOME



PANGAEA.

Data Publisher for Earth & Environmental Science



figshare



DRYAD

Paquetes de R como compendios de investigación



E:/PC_mama_te_quiero/BillCAPO/Tesis/Analisis/
estafa_maestra/microsoft