

**Doing reproducible science:
from your hard-won data
to a publishable manuscript
without going mad**

Francisco Rodriguez-Sanchez
@frod_san

A typical research workflow

1. Prepare data (**EXCEL**)

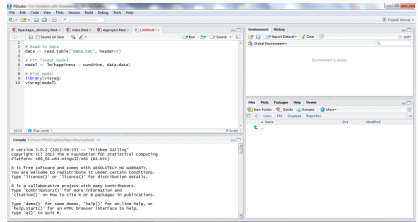
	A	B
1	happiness_index	sunshine_h
2	10.5	978.4
3	6.6	660.9
4	11.3	1093.5
5	9.6	978.9
6	10.9	1135.5
7	9.1	907.0
8	10.6	990.4
9	12.4	1172.9
10	9.6	1025.6
11	10.1	1055.0
12	10.9	1093.7
13	8.9	863.8
14	12.5	1196.6
15	10.0	995.8
16	11.0	1120.2
17	10.3	988.0
18	9.7	987.0
19	9.3	970.4
20	10.9	1076.6
21	9.0	909.8
22	7.7	733.4
23	9.0	985.2
24	10.4	1084.0
25	10.0	1066.7

data

Ready

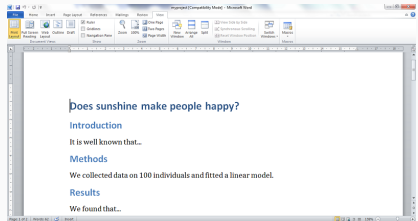
A typical research workflow

1. Prepare data (**EXCEL**)
2. Analyse data (**R**)



A typical research workflow

1. Prepare data (**EXCEL**)
2. Analyse data (**R**)
3. Write report/paper
(**WORD**)

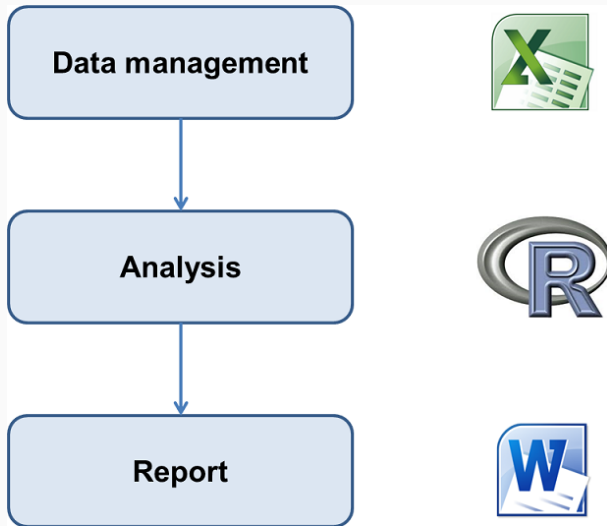


A typical research workflow

1. Prepare data (**EXCEL**)
2. Analyse data (**R**)
3. Write report/paper
(**WORD**)
4. Start the email attachments
nightmare...



This workflow is broken



Problems of a broken workflow

- How did you do this? What analysis is behind this figure? Did you account for ...?

Problems of a broken workflow

- How did you do this? What analysis is behind this figure? Did you account for ...?
- What dataset was used? Which individuals were left out? Where is the clean dataset?

Problems of a broken workflow

- How did you do this? What analysis is behind this figure? Did you account for ...?
- What dataset was used? Which individuals were left out? Where is the clean dataset?
- Oops, there is an error in the data. Can you repeat the analysis? And update figures/tables in Word!



Trevor A. Branch

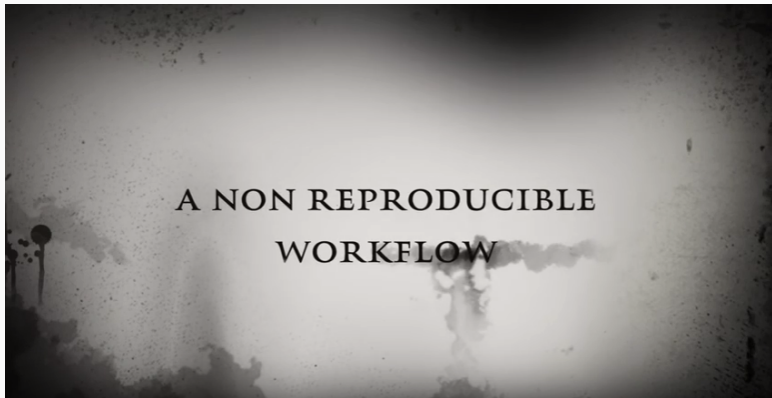
@TrevorABranch



Follow

My rule of thumb: every analysis you do on a dataset will have to be redone 10–15 times before publication. Plan accordingly. [#Rstats](#)

Our everyday scary movie



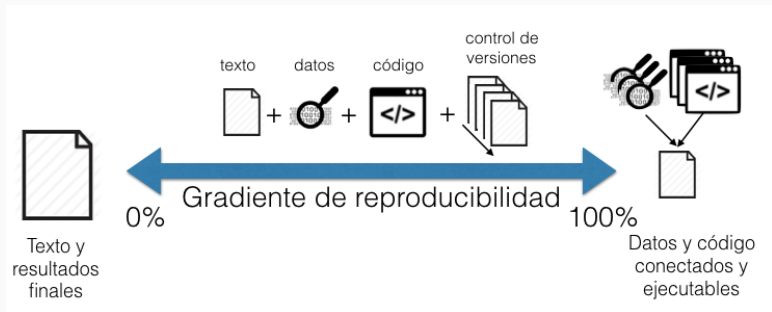
<https://youtu.be/s3JldKoA0zw>

WHAT is Reproducible Science?

A scientific article is **reproducible** if there is computer **code** that can **regenerate** all results and figures from the original data.

- Transparent
- Traceable
- Comprehensive
- Useful

Most science is not reproducible

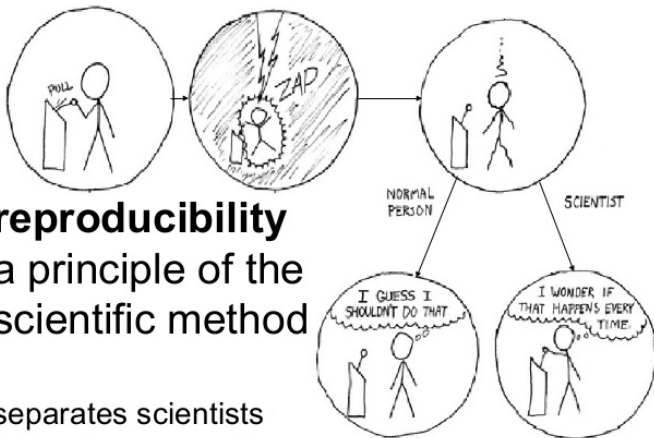


Even **you** will struggle to reproduce **your own results** from a few weeks/months ago.

*You can't reproduce if you don't understand where a number came from.
You can't reproduce what you don't remember. And trust me: you won't.
You can't reproduce what you've lost. What if you need access to a file as
it existed 1, 10, 100, or 1000 days ago?*

Ben Bond-Lamberty

WHY Reproducible Science?



reproducibility
a principle of the
scientific method

separates scientists
from other researchers
and normal people

<http://xkcd.com/242/>

Carole Goble <http://www.slideshare.net/carolegoble/ismb2013-keynotecleangoble>



Noam Ross

@noamross



Follow



Gelman: "Reproducible research is even better when you're wrong" [#stancon2017](#)

- Fundamental pillar of **scientific method**

Reproducible Science: WHY

- Fundamental pillar of **scientific method**
- Much less prone to **errors**

Reproducible Science: WHY

- Fundamental pillar of **scientific method**
- Much less prone to **errors**
- Regenerate results **automatically**

Reproducible Science: WHY

- Fundamental pillar of **scientific method**
- Much less prone to **errors**
- Regenerate results **automatically**
- **Code reuse** & sharing accelerates scientific progress

Reproducible Science: WHY

- Fundamental pillar of **scientific method**
- Much less prone to **errors**
- Regenerate results **automatically**
- **Code reuse** & sharing accelerates scientific progress
- Increasingly required by **journals**

Reproducible Science: WHY

- Fundamental pillar of **scientific method**
- Much less prone to **errors**
- Regenerate results **automatically**
- **Code reuse** & sharing accelerates scientific progress
- Increasingly required by **journals**
- Higher publication **impact** (citations, future collaborations, etc)

HOW TO DO Reproducible Science?

1. File **organisation**.
2. **Data management**. Spreadsheet good practices.
3. **Code-based** data analysis. **Rmarkdown**
4. Software **dependencies**.
5. **Version control** & collaborative writing.

- All files in **same directory** (Rstudio project).

- All files in **same directory** (Rstudio project).
- **Raw data untouched** in independent folder.

- All files in **same directory** (Rstudio project).
- **Raw data untouched** in independent folder.
- Derived, **clean data** in another folder.

- All files in **same directory** (Rstudio project).
- **Raw data untouched** in independent folder.
- Derived, **clean data** in another folder.
- Figures, code, etc also have their own folder.

File organisation example

```
| - README          # general info about the project

| - analysis.R      # master script that executes everything

| - data-raw/       # original raw data

| - data/           # clean data (produced w/ script)

| - R/              # functions definitions

| - doc/            # manuscript files

| - figs/           # final figures

| - output/         # other code output
```

Data management

See <https://www.dataone.org/best-practices>

.

1. Planification (e.g. DMPTool)
2. Collection
3. Metadata description (EML, Morpho, Data Packages, DataPackageR)
4. Quality control (e.g. assertr, validate, pointblank)
5. Storage

Editorial expression of concern

IN THE 3 June issue, *Science* published the Report “Environmentally relevant concentrations of microplastic particles influence larval fish ecology” by Oona M. Lönnstedt and Peter Eklöv (1). The authors have notified *Science* of the theft of the computer on which the raw data for the paper were stored. These data were not backed up on any other device nor deposited in an appropriate repository. *Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available, beyond those already presented in the paper and its supplement, to enable readers to understand, assess, reproduce, or extend the conclusions of the paper.

Jeremy Berg

Editor in Chief

Use the **cloud**: safe, persistent, easy to share

- [Open Science Framework](#)
- GitHub
- Dropbox
- Figshare, Zenodo, etc
- See all data repositories in www.re3data.org

Tidy data

country	year	cases	population
Afghanistan	1999	37737	172206362
Afghanistan	2000	80488	174204898
Brazil	1999	212258	1272015272
Brazil	2000	213766	128201583

variables

country	year	cases	population
Afghanistan	1999	37737	172206362
Afghanistan	2000	80488	174204898
Brazil	1999	212258	1272015272
Brazil	2000	213766	128201583

observations

country	year	cases	population
Afghanistan	1999	37737	172206362
Afghanistan	2000	80488	174204898
Brazil	1999	212258	1272015272
Brazil	2000	213766	128201583

values

country	year	cases
Afghanistan	1999	745
Afghanistan	2000	2666
Brazil	1999	37737
Brazil	2000	80488
China	1999	212258
China	2000	213766

country	1999	2000
Afghanistan	745	2666
Brazil	37737	80488
China	212258	213766

table4

<http://r4ds.had.co.nz/tidy.html>

Very careful with data entry and management in Excel!

COMMENT

Open Access



Gene name errors are widespread in the scientific literature

Mark Ziemann¹, Yotam Eren^{1,2} and Assam El-Osta^{1,3*}

Abstract

The spreadsheet software Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.

frequently reused. Our aim here is to raise awareness of the problem.

We downloaded and screened supplementary files from 18 journals published between 2005 and 2015 using a suite of shell scripts. Excel files (.xls and .xlsx suffixes) were converted to tabular separated files (tsv) with ssconvert (v1.12.9). Each sheet within the Excel file was converted to a separate tsv file. Each column of data in the tsv file was screened for the presence of gene sym-

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.
- Use **Data validation** in Excel (or GForms) to constrain data entry to accepted values.

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.
- Use **Data validation** in Excel (or GForms) to constrain data entry to accepted values.
- Don't combine multiple pieces of information in one cell.

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.
- Use **Data validation** in Excel (or GForms) to constrain data entry to accepted values.
- Don't combine multiple pieces of information in one cell.
- **Don't touch raw data**. Do all data manipulation through code.

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.
- Use **Data validation** in Excel (or GForms) to constrain data entry to accepted values.
- Don't combine multiple pieces of information in one cell.
- **Don't touch raw data**. Do all data manipulation through code.
- Export data as plain text (txt, csv).

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.
- Use **Data validation** in Excel (or GForms) to constrain data entry to accepted values.
- Don't combine multiple pieces of information in one cell.
- **Don't touch raw data**. Do all data manipulation through code.
- Export data as plain text (txt, csv).
- <http://www.datacarpentry.org/spreadsheet-ecology-lesson/>

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.
- Use **Data validation** in Excel (or GForms) to constrain data entry to accepted values.
- Don't combine multiple pieces of information in one cell.
- **Don't touch raw data**. Do all data manipulation through code.
- Export data as plain text (txt, csv).
- <http://www.datacarpentry.org/spreadsheet-ecology-lesson/>
- <http://kbroman.org/dataorg/>

Spreadsheet good practices

- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- **Avoid** spaces, numbers, and **special characters** in column names.
- Always **write zero values**, to distinguish from blank/missing data.
- Use blank/empty cells, or NA, for missing data.
- Input dates as **year, month, day** in separate columns. Or **YYYY-MM-DD** as text.
- Use **Data validation** in Excel (or GForms) to constrain data entry to accepted values.
- Don't combine multiple pieces of information in one cell.
- **Don't touch raw data**. Do all data manipulation through code.
- Export data as plain text (txt, csv).
- <http://www.datacarpentry.org/spreadsheet-ecology-lesson/>
- <http://kbroman.org/dataorg/>
- Broman & Woo: [Data organization in spreadsheets](#)

Common spreadsheet errors

More than one variable per column

Date collected	Plot	Species-Sex	Weight
1/9/78	1	DM-M	40
1/9/78	1	DM-F	36
1/9/78	1	DS-F	135
1/20/78	1	DM-F	39
1/20/78	2	DM-M	43
1/20/78	2	DS-F	144
3/13/78	2	DM-F	51
3/13/78	2	DM-F	44
3/13/78	2	DS-F	146

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Source: Data Carpentry

Multiple tables

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1																																	
2	lake site May 29 2012						29-May avr	SEM	lake site Jun 12 2012						12-Jun avr	SEM	lake site Jun 19 2012						19-Jun	Lake site Jun 26 2012						26-Jun	SEM		
3			bug1	bug2					plot	bug1	bug2	general		plot			bug1	bug2	general		plot	bug1		bug2	general								plot
4	1	T1	1	1	2	T1	2.6	0.51	1	T1	6	85	95	T1	30.4	15.47126	1	T1	17	80	97		avr	SEM	1	T1	52	191	243		avr	SEM	
5	2	T1	1	2	3	T2	0.2	0.2	2	T1	8	13	21	T2	0.2	0.2	2	T1	44	136	180	T1	77.8	30.384865	2	T1	50	270	320	T1	141.6	60.313	
6	3	T1	1	3	4	control	0.2	0.2	3	T1	11	0	11	control	0.6	0.6	3	T1	18	0	18	T2	1.8	1.5620499	3	T1	6	0	6	T2	0.2	0.2	
7	4	T1	1	0	1				4	T1	0	6	6				4	T1	0	14	14	control	0.4	0.244949	4	T1	0	39	39	control	0	0	
8	5	T1	0	3	3				5	T1	3	20	23				5	T1	10	70	80				5	T1	4	96	100				
9	6	T2	1	0	1				6	T2	0	0	0				6	T2	1	7	8				6	T2	0	1	1				
10	7	T2	0	0	0				7	T2	0	0	0				7	T2	0	1	1				7	T2	0	0	0				
11	8	T2	0	0	0				8	T2	1	0	1				8	T2	0	0	0				8	T2	0	0	0				
12	9	T2	0	0	0				9	T2	0	0	0				9	T2	0	0	0				9	T2	0	0	0				
13	10	T2	0	0	0				10	T2	0	0	0				10	T2	0	0	0				10	T2	0	0	0				
14	11	control	0	0	0				11	control	0	0	0				11	control	0	0	0				11	control	0	0	0				
15	12	control	0	0	0				12	control	0	0	0				12	control	0	0	0				12	control	0	0	0				
16	13	control	0	0	0				13	control	0	0	0				13	control	0	0	0				13	control	0	0	0				
17	14	control	0	0	0				14	control	0	0	0				14	control	0	1	1				14	control	0	0	0				
18	15	control	1	0	1				15	control	3	0	3				15	control	0	1	1				15	control	0	0	0				
19																																	
20																																	
21	Barn site May 29 2012						29-May	SEM	Barn site Jun 12 2012						12-Jun	SEM	Barn site Jun 19 2012						19-Jun	Barn Site Jun 26 2012						26-Jun	SEM		
22		plot	bug1	bug2	general				plot	bug1	bug2	general		plot			bug1	bug2	general		plot	bug1		bug2	general								plot
23	1	T1	3	3	6				1	T1	21	0	21				1	T1	5	0	5				1	T1	0	0	0				
24	2	T1	1	4	6		avr	SEM	2	T1	36	74	110		avr	SEM	2	T1	65	502	567		avr	SEM	2	T1	44	2057	2101	T1	431.8	417.33	
25	3	T1	0	0	0	T1	2.4	1.288	3	T1	13	0	13	T1	30.6	10.10124	3	T1	10	7	17	T1	119.4	11.92882	3	T1	12	20	32	T2	0.4	0.4	
26	4	T1	0	0	0	T2	0.4	0.245	4	T1	7	0	7	T2	5	0.774597	4	T1	0	16	6	T2	5	1.1908902	4	T1	0	16	16	control	1.2	0.5831	
27	5	T1	0	1	1	control	1	0.316	5	T1	2	0	2	control	2.2	1.714643	5	T1	0	2	2	control	2.8	0.969536	5	T1	0	10	10				
28	6	T2	0	0	0				6	T2	1	0	1				6	T2	0	8	8				6	T2	0	0	0				
29	7	T2	0	0	0				7	T2	0	4	4				7	T2	0	12	12				7	T2	0	0	0				
30	8	T2	0	1	1				8	T2	0	0	0				8	T2	0	0	0				8	T2	0	0	0				
31	9	T2	0	1	1				9	T2	0	0	0				9	T2	3	0	3				9	T2	0	0	0				
32	10	T2	0	0	0				10	T2	0	0	0				10	T2	2	0	2				10	T2	0	2	2				
33	11	control	0	0	0				11	control	1	0	1				11	control	0	5	5				11	control	0	2	2				
34	12	control	0	1	1				12	control	0	0	0				12	control	1	1	2				12	control	1	0	1				
35	13	control	0	1	1				13	control	0	0	0				13	control	0	0	0				13	control	0	0	0				
36	14	control	0	1	1				14	control	8	1	9				14	control	0	5	5				14	control	0	3	3				
37	15	control	0	2	2				15	control	0	1	1				15	control	0	2	2				15	control	1	0	0				
38																																	
39																																	

Multiple tabs

Could you avoid new tab by adding a column to original spreadsheet?

Using formatting, comments, etc to convey information

Plot: 2					
Date collect	Species	Sex	Weight		
1/8/14	NA				
1/8/14	DM	M	44		
1/8/14	DM	M	38		
1/8/14	OL				
1/8/14	PE	M	22		
1/8/14	DM	M	38		
1/8/14	DM	M	48		
1/8/14	DM	M	43		
1/8/14	DM	F	35		
1/8/14	DM	M	43		
1/8/14	DM	F	37		
1/8/14	PF	F	7		
1/8/14	DM	M	45		
1/8/14	OT				
1/8/14	DS	M	157		
1/8/14	OX				
2/18/14	NA	M	218		
2/18/14	PF	F	7		
2/18/14	DM	M	52		
	measurement device not calibrated				

Date collect	Species	Sex	Weight	Calibrated
1/8/14	NA			
1/8/14	DM	M	44	Y
1/8/14	DM	M	38	Y
1/8/14	OL			
1/8/14	PE	M	22	Y
1/8/14	DM	M	38	Y
1/8/14	DM	M	48	Y
1/8/14	DM	M	43	Y
1/8/14	DM	F	35	Y
1/8/14	DM	M	43	Y
1/8/14	DM	F	37	Y
1/8/14	PF	F	7	Y
1/8/14	DM	M	45	Y
1/8/14	OT			
1/8/14	DS	M	157	N
1/8/14	OX			
2/18/14	NA	M	218	N
2/18/14	PF	F	7	Y
2/18/14	DM	M	52	Y

Your turn: tidy up this messy dataset

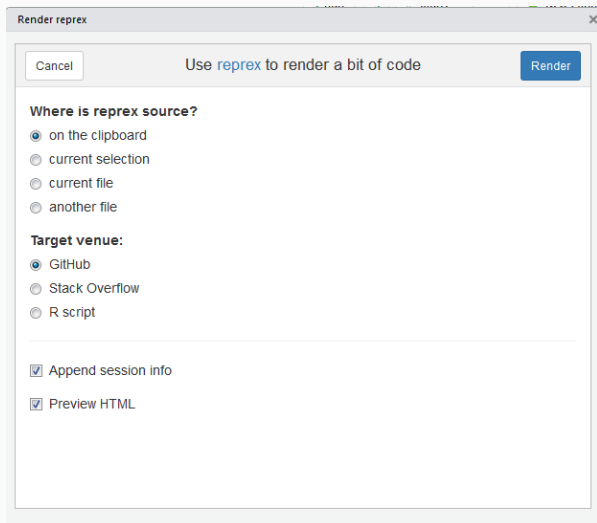
<https://ndownloader.figshare.com/files/2252083>

Data analysis

Always use code

- Reproducible
- Reusable

Stuck with error / Found bug? Use reprex



The screenshot shows a dialog box titled "Render reprex" with a close button (X) in the top right corner. The dialog has a light gray header bar containing a "Cancel" button on the left, the text "Use **reprex** to render a bit of code" in the center, and a "Render" button on the right. The main content area is white and contains the following sections:

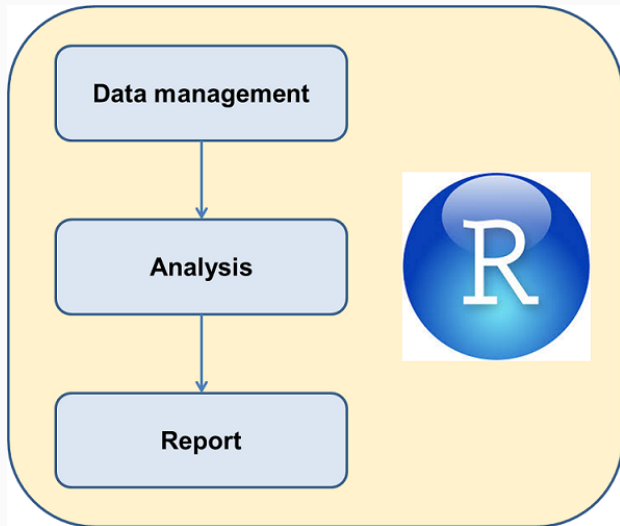
- Where is reprex source?**
 - ☒ on the clipboard
 - ☐ current selection
 - ☐ current file
 - ☐ another file
- Target venue:**
 - ☒ GitHub
 - ☐ Stack Overflow
 - ☐ R script

Below the radio buttons are two checked checkboxes:

- ☒ Append session info
- ☒ Preview HTML

<https://reprex.tidyverse.org/>

Dynamic reports



Rmarkdown documents

- Fully reproducible (trace all results inc. tables and plots)
- Dynamic (regenerate with 1 click)
- Suitable for
 - documents (HTML, Word, PDF)
 - presentations (HTML, PDF, PowerPoint)
 - books
 - websites
 - ...

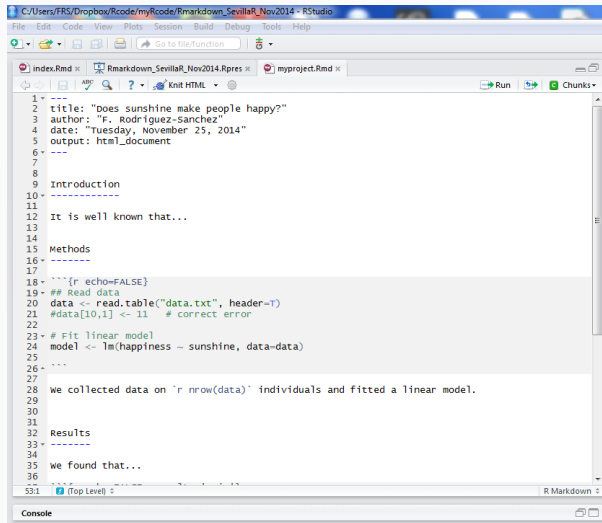


Let's see Rmarkdown in action

In Rstudio, create new Rmarkdown document and click on `Knit HTML`.

Example: Does sunshine influence happiness?

See `myproject.Rmd` (<http://bit.ly/rmdsun>)



```
1 ---
2 title: "Does sunshine make people happy?"
3 author: "F. Rodríguez-Sánchez"
4 date: "Tuesday, November 25, 2014"
5 output: html_document
6 ---
7
8
9 Introduction
10 ---
11
12 It is well known that...
13
14
15 Methods
16 ---
17
18 ```{r echo=FALSE}
19 ## Read data
20 data <- read.table("data.txt", header=T)
21 #data[10,1] <- 11 # correct error
22
23 # Fit linear model
24 model <- lm(happiness ~ sunshine, data=data)
25
26 ```
27
28 we collected data on `r nrow(data)` individuals and fitted a linear model.
29
30
31
32 Results
33 ---
34
35 we found that...
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53:1 (Top Level) <
R Markdown
```

HTML output includes text, plot and formatted table

Does sunshine make people happy?

F. Rodriguez-Sanchez

Tuesday, November 25, 2014

Introduction

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

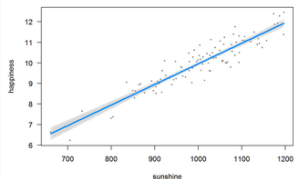
Methods

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

Results

We found that...

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0651657	0.4264970	-0.1527928	0.8788758
sunshine	0.0100228	0.0004232	23.6833264	0.0000000



Discussion

These results confirm that sunshine is good for happiness (slope = 0.0100228).

Acknowledgements

Y. Xie, J. MacFarlane, Rstudio...

Spotted error in the data? No problem!

Make changes in Rmarkdown document, click `knit` and report will **update automatically!**

Why Rmarkdown?



Convince me to start using R Markdown

R Markdown

rmarkdown



Darren_Dahly

1. Start using R Markdown to generate reports of your data analyses.
2. If the data changes, rerun the report with a click of the mouse.
3. Take 3 days off of work.
4. On the 4th day, tell your collaborators that the re-analysis is complete.
5. Be hailed as a hero.

<https://community.rstudio.com/t/convince-me-to-start-using-r-markdown/1636/12>

Other formats: PDF, Word, PowerPoint...

Does sunshine make people happy?

F. Rodriguez-Sanchez
Tuesday, November 25, 2014

Introduction

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

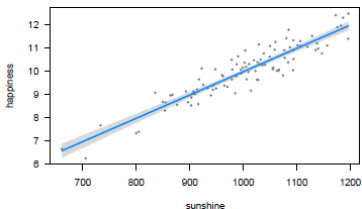
Methods

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

Results

We found that...

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0651657	0.4964970	-0.1327928	0.8788758
sunshine	0.0100228	0.0004232	23.6833264	0.0000000



Does sunshine make people happy?

F. Rodriguez-Sanchez
Tuesday, November 25, 2014

Introduction

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

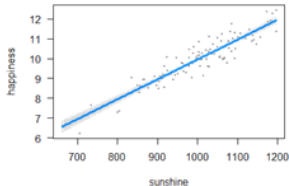
Methods

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

Results

We found that...

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0651657	0.4264970	-0.1527928	0.8788758
sunshine	0.0100228	0.0004232	23.6833264	0.0000000



Adding citations by DOI: rcrossref

Add Crossref Citations

Cancel

Add Crossref Citations

Done

Add a new bibliography entry through Crossref DOI

10.3390/ma8063101

Add to My Citations

Type: journal-article

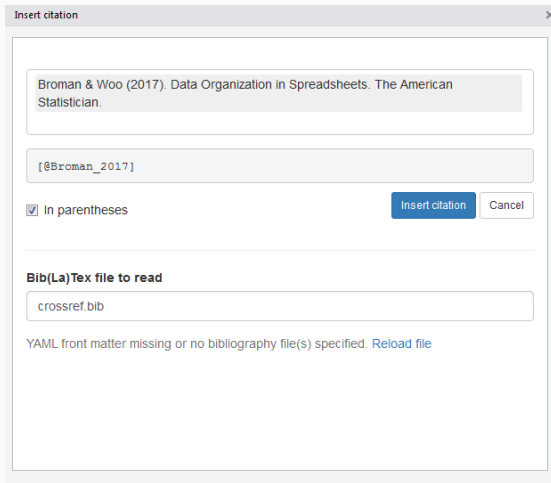
Title: Photoluminescent ZnO Nanoparticles and Their Biological Applications

Author: Zheng-Yong Zhang; Huan-Ming Xiong

Time: 2015

Publisher: MDPI AG

Adding citations from BibTeX file (or Zotero): citr



The screenshot shows a window titled "Insert citation" with a close button (X) in the top right corner. Inside the window, there are several input fields and buttons:

- A text box containing the citation: "Broman & Woo (2017). Data Organization in Spreadsheets. The American Statistician."
- A text box containing the BibTeX key: "[@Broman_2017]"
- A checkbox labeled "In parentheses" which is checked.
- Two buttons: "Insert citation" (blue) and "Cancel" (white with a grey border).
- A section titled "Bib(La)Tex file to read" with a text box containing "crossref.bib".
- A message at the bottom: "YAML front matter missing or no bibliography file(s) specified. [Reload file](#)"

<https://github.com/crsh/citr/>

Insert equations

- Using LaTeX (\$) or (\$\$)
- bookdown::mathquill
- mathpix: <https://github.com/jonocarroll/mathpix>

Revise writing style: gramr

gramr

build **passing** coverage **unknown** last change **2017-05-19** r>= **3.4.0** license **MIT**

The goal of gramr is to help R programmers who can't write good and and wanna learn to do other stuff good too by checking a RMarkdown document for grammatical errors.

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

Spell and grammar checking

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

Find synonyms

gramr

build passing coverage unknown last change 2017-05-19 r>= 3.4.0 license MIT

The goal of gramr is to help R programmers who can't write good and and wanna learn to do other stuff good too by checking a RMarkdown document for grammatical errors.

<https://github.com/gadenbuie/synamyn>

<https://github.com/benmarwick/wordcountaddin>

- [rticles](#): Elsevier, Springer, PeerJ, PNAS, PLoS...
- [rmdTemplates](#)
- Check out GitHub for Rmarkdown templates...

Write your next PNAS/PLoS/PeerJ/whatever in Rmarkdown!

My cool paper written in Rmarkdown

F. Rodriguez-Sanchez^{1,2} and And Friends^{3,4}

¹Stem Institute of Technology, Department, Street, City, State, Zip; ²Stem University, Department, Street, City, State, Zip

The manuscript was compiled on September 10, 2018.

Please provide an abstract of no more than 300 words in a single paragraph. Abstracts should explain to the general reader the major contributions of the article. References in the abstract must be cited in full within the abstract itself and cited in the text.

one | two | options | options | options

This PNAS journal template is provided to help you write your work in the current journal format. Instructions for use are provided 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 whilst 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 as well.

Submitting Manuscripts. All authors must submit their article as [PNAS Central](#). If you are using Overleaf to write your article, you can use the "Submit to PNAS" option in the top bar of the editor window.

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

Manuscript Length. PNAS generally uses a two-column format averaging 67 characters, including spaces, per line. The maximum length of a Direct Submission research article is six pages and a PNAS PLUS research article is ten pages including all text, space, and the number of characters displayed by figures, tables, and equations. When submitting tables, figures, and/or equations in addition to text, keep the text for your manuscript under 20,000 characters (including spaces) for Direct Submissions and 70,000 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 sections should not be duplicated. SI references



Fig. 1. Placeholder image of a frog with a large sample caption to show publication styling.

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

Data Archiving. 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, ArrayExpress, Protein Data Bank, UniProt, and others outlined in the Information for Authors, is acceptable.

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/editservices/language-editing-services](#)). PNAS does not take responsibility for or endorse these services, and their use has 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 written at a level understandable 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 file and is required for all research papers.

The page number of author contributions is:

The use of this template is not required.

Impactos de la herbivoría por ungulados en las comunidades de plantas leñosas del Parque Natural Los Alcornocales

Francisco Rodríguez-Sánchez^{1,4}, Virginia Luque-Oliva^{1,2}, Vicente Jurado³

1. Departamento de Ecología Integrativa, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Avda. Américo Vespucio 26, 41092 Sevilla, España.
1. Universidad Pablo de Olavide, Ctra. de Utrera s/n, Dos Hermanas, Sevilla, España.
1. Facultad de Geografía e Historia, C/ Doña María de Padilla s/n, 41004 Sevilla, España.
1. Autor para correspondencia [frrodriguez.work@gmail.com]

Resumen

Las poblaciones de ciervos y otros ungulados se han incrementado enormemente en muchas áreas mediterráneas en las últimas décadas. Los impactos del incremento de la presión de herbivoría en las comunidades de plantas y la eficacia de distintas medidas de conservación están aún poco claras. En este trabajo se evalúan los efectos de la herbivoría en comunidades arbustivas del Campo de Gibraltar y Parque Natural de Los Alcornocales. Se ha comparado la diversidad de plantas, la cobertura, la altura y la intensidad de daños dentro y fuera de un total de 20 cercados de exclusión de herbívoros establecidos hace 10-15 años para favorecer la regeneración del bosque. Aunque la composición de la comunidad de plantas (riqueza de especies y diversidad) fue similar en las zonas cercadas y no cercadas, la altura de los arbustos y su cobertura fueron mayores dentro de los cercados de protección. En contraste, el daño de herbivoría fue

Can write full thesis in Rmarkdown!

See [thesis.Rmd](#).

See [thesis.pdf](#).

BOOKDOWN

Write HTML, PDF, ePub, and Kindle books with R Markdown

The **bookdown** package is an [open-source R package](https://bookdown.org) that facilitates writing books and long-form articles/reports with R Markdown. Features include:

- Generate printer-ready books and ebooks from R Markdown documents.
- A markup language easier to learn than LaTeX, and to write elements such as section headers, lists, quotes, figures, tables, and citations.
- Multiple choices of output formats: PDF, LaTeX, HTML, EPUB, and Word.
- Possibility of including dynamic graphics and interactive applications (HTML widgets and Shiny apps).
- Support a wide range of languages: R, C/C++, Python, Fortran, Julia, Shell scripts, and SQL, etc.
- LaTeX equations, theorems, and proofs work for all output formats.
- Can be published to GitHub, bookdown.org, and any web servers.
- Integrated with the RStudio IDE.
- One-click publishing to <https://bookdown.org>.



<https://bookdown.org/>

<https://github.com/yihui/xaringan>

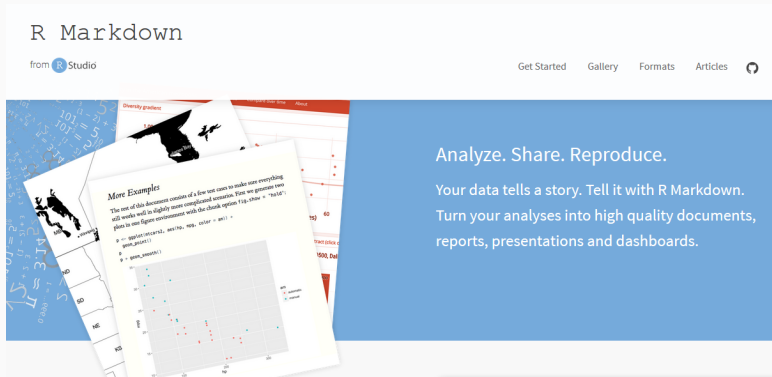
[https:](https://bookdown.org/yihui/rmarkdown/parameterized-reports.html)

[//bookdown.org/yihui/rmarkdown/parameterized-reports.html](https://bookdown.org/yihui/rmarkdown/parameterized-reports.html)

Resources

Rmarkdown website

<http://rmarkdown.rstudio.com/index.html>



The screenshot shows the R Markdown website homepage. At the top, it says "R Markdown" in a large, serif font, followed by "from R Studio" in a smaller font. To the right, there are navigation links: "Get Started", "Gallery", "Formats", "Articles", and a circular icon. Below the header, there is a large blue banner with the text "Analyze. Share. Reproduce." and "Your data tells a story. Tell it with R Markdown. Turn your analyses into high quality documents, reports, presentations and dashboards." On the left side of the banner, there is a collage of images including a map of Europe, a scatter plot, and a document page with the title "More Examples" and some R code.

R Markdown

from R Studio

Get Started Gallery Formats Articles

Analyze. Share. Reproduce.

Your data tells a story. Tell it with R Markdown.

Turn your analyses into high quality documents, reports, presentations and dashboards.

Rmarkdown cheat sheet

R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com

Studio

Workflow

1. Open a new .Rmd file at File > New File > R Markdown. Use the wizard that opens. RStudio also has the file with a template.
2. Write document. Edit and preview. Publish. Share output.
3. Knit document to create report. You knit (build or render) to knit.
4. Preview Output. You knit (build or render) to knit.
5. Publish (optional). To web or server. Launch publish button to accounts at: GitHub, RStudio Connect, Netlify, Heroku, AWS S3, etc. Find in document. File path to output document.
6. Examine building. In R Markdown console.
7. Use output file that is knit (build or render).

Interactive Documents

Turn your report into an interactive Shiny document in 4 steps:

1. Add runtime: shiny to the YAML header.
2. Call Shiny input functions to embed input objects.
3. Call Shiny render functions to embed reactive output.
4. Render with markdown: run or click Run Document in RStudio IDE.

Embed code with knitr syntax

Insert code: `<code>` (Built with 3.2.3)

Global options: `knitr::opts_chunk$set(echo = TRUE)`

Cache: `cache` (cache results for future knits (default = FALSE))

Cache path: `cache.path` (directory to save cached results in (default = "cache"))

Child: `child` (link to file and then include (default = NULL))

Collapsible: `collapsible` (collapse all output into single blocks (default = FALSE))

Comment: `comment` (pre for each line of results (default = ""))

Depends: `depends` (chunk dependencies for cache (default = NULL))

Engine: `engine` (display code in output document (default = TRUE))

Eval: `eval` (code language used in chunk (default = "R"))

Error: `error` (display error messages in doc (TRUE) or stop-render when errors occur (FALSE) (default = TRUE))

Fig: `fig` (figure caption as character string (default = NULL))

Fig height: `fig.height` (Dimensions of figure in inches)

Fig width: `fig.width` (Dimensions of figure in inches)

Highlight: `highlight` (highlight source code (default = TRUE))

Include: `include` (include chunk in doc after knitting (default = TRUE))

Message: `message` (display code messages in document (default = TRUE))

Results: `results` (default = "markup")

Self: `self` (pass-through "true" or "false" (default = "true"))

Tidy: `tidy` (code style for display (default = FALSE))

Warning: `warning` (display code warnings in document (default = TRUE))

Parameters

Parameterize your documents to reuse with different inputs (e.g., data sets, values, etc.)

1. Add parameters. Create and set parameters in the header as sub-values.
2. Call parameters. Call parameter values in code as `params$parameter`.
3. Set parameters. Set values with `knitr::set_parameters()` or the parameter argument of `render()`.

Knitr is a trademark of RStudio, Inc. • © 2022 RStudio • RStudio - R Markdown Cheat Sheet

Learn more at rmarkdown.rstudio.com • RStudio IDE v2022.02.10

https://www.rstudio.org/links/r_markdown_cheat_sheet

Rmarkdown reference guide



R Markdown Reference Guide

Learn more about R Markdown or cran.rstudio.com/doc
Learn more about Interactive Docs or rpubs.com/articles

Contents

1. Markdown Syntax
2. Knitr chunk options
3. Pandoc options

Syntax

Make a code chunk with three back ticks followed by an `r` in braces. End the chunk with three back ticks:

```
```r```  
paste("Hello", "World")
```

Place code inline with a single back tick. The first back tick must be followed by an `R`, like this: `r paste("Hello", "World")`.

Add chunk options within braces. For example, `echo=FALSE` will prevent source code from being displayed:

```
```r eval=TRUE, echo=FALSE```  
paste("Hello", "World")
```

Becomes

Make a code chunk with three back ticks followed by an `r` in braces. End the chunk with three back ticks:

```
paste("Hello", "World")
```

```
## [1] "Hello World"
```

Place code inline with a single back tick. The first back tick must be followed by an `R`, like this: `Hello World`.

Add chunk options within braces. For example, `echo=FALSE` will prevent source code from being displayed:

```
## [1] "Hello World"
```

Learn more about chunk options or <http://yihui.name/en/option>

option	default value	description
Code chunk options		
<code>child</code>	<code>NULL</code>	A character vector of filenames. Knitr will knit the files and place them into the main document.
<code>code</code>	<code>NULL</code>	Set to R code. Knitr will replace the code in the chunk with the code in the code option.
<code>engine</code>	<code>"R"</code>	Knitr will evaluate the chunk in the named language, e.g. <code>engine = "python"</code> . For named <code>echo=knitr::knit_engine_get()</code> to see supported languages.
<code>eval</code>	<code>TRUE</code>	If <code>FALSE</code> , knitr will not run the code in the code chunk.
<code>include</code>	<code>TRUE</code>	If <code>FALSE</code> , knitr will not knit the chunk but not include the chunk in the final document.
<code>port</code>	<code>TRUE</code>	If <code>FALSE</code> , knitr will not include the chunk when running <code>port()</code> to extract the source code.
Knitr options		
<code>collapse</code>	<code>FALSE</code>	If <code>TRUE</code> , knitr will collapse all the source and output blocks created by the chunk into a single block.
<code>echo</code>	<code>TRUE</code>	If <code>FALSE</code> , knitr will not display the code in the code chunk above it's results in the final document.
<code>results</code>	<code>"markup"</code>	If <code>"hide"</code> , knitr will not display the code's results in the final document. If <code>"hide"</code> , knitr will display displaying all output given until the end of the chunk. If <code>"ask"</code> , knitr will pass through results without reformating them (useful if results return raw HTML, etc.).
<code>error</code>	<code>TRUE</code>	If <code>FALSE</code> , knitr will not display any error messages generated by the code.
<code>message</code>	<code>TRUE</code>	If <code>FALSE</code> , knitr will not display any messages generated by the code.
<code>warning</code>	<code>TRUE</code>	If <code>FALSE</code> , knitr will not display any warning messages generated by the code.
Code formatting		
<code>background</code>	<code>"#FFFFFF"</code>	A background color for chunks in LaTeX output.
<code>comment</code>	<code>"#"</code>	A character string. Knitr will append the string to the start of each line of results in the final document.
<code>highlight</code>	<code>TRUE</code>	If <code>TRUE</code> , knitr will highlight the source code in the final output.
<code>prompt</code>	<code>FALSE</code>	If <code>TRUE</code> , knitr will add to the start of each line of code displayed in the final document.
<code>size</code>	<code>"normalfont"</code>	Fontsize for LaTeX output.
<code>strip.white</code>	<code>TRUE</code>	If <code>TRUE</code> , knitr will remove white spaces that appear at the beginning or end of a code chunk.
<code>tidy</code>	<code>FALSE</code>	If <code>TRUE</code> , knitr will tidy code chunks for display with the <code>tidy_source()</code> function in the <code>formatt</code> package.

 (CC) BY-SA

Updated 10/30/2014

© 2014 Yihui Xie. All rights reserved.

https://www.rstudio.org/links/r_markdown_reference_guide

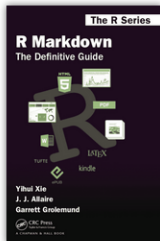
R Markdown: The Definitive Guide

by Yihui Xie, J. J. Allaire, Garrett Grolemond

2018-09-11



239



The first official book authored by the core R Markdown developers that provides a comprehensive and accurate reference to the R Markdown ecosystem. With R Markdown, you can easily create reproducible data analysis reports, presentations, dashboards, interactive applications, books, dissertations, websites, and journal articles, while enjoying the fun and the great power of R and other

<https://bookdown.org/yihui/rmarkdown/>

Managing software dependencies

Managing package dependencies in R

- **sessionInfo** (or `session_info`)
- **checkpoint**
- packrat
- switchr
- docker

Version control

"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



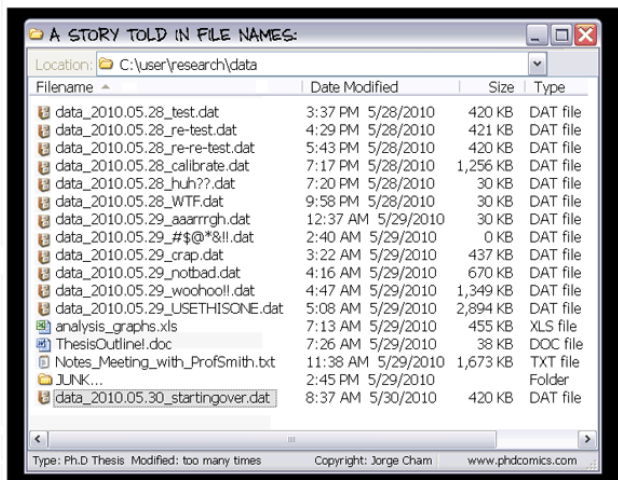
FINAL_rev.8.comments5.
CORRECTIONS.doc


















































FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10. #@\$%WHYDID
ICOMETOGRADSCHOOL?????.doc



 Coordinates_rawdata.csv	18/05/2015 19:07
 coords_2015-09-09.csv	23/09/2015 17:18
 coords_2015-09-09_modif.csv	05/11/2015 15:20
 coords_2015-10-11_modif_YC.csv	17/11/2015 13:49
 coords_2015-10-18_modif_YC.csv	18/11/2015 17:26
 coords_2015-12-26_modif_YC.csv	28/12/2015 13:33
 coords_2015-12-26_modif_YC_years.csv	30/03/2016 19:38
 Pulido et al_SMI_Data.csv	20/10/2015 11:55
 Pulido et al_SMI_Data_modif_YC_2015-12-26.csv	28/12/2015 13:30
 qualitative_data.csv	04/07/2016 15:50
 cleandata.xlsx	25/06/2015 01:14
 cleandata_YC.xlsx	30/06/2015 16:22
 COORDENADAS PACO_20-05-2016 CON REVIEWS.xlsx	20/05/2016 16:23
 COORDENADAS PACO_20-05-2016 CON REVIEWS_FRS.xlsx	27/05/2016 19:41
 COORDENADAS_paper195(Girella_elevata).xlsx	08/06/2016 13:09
 coordenadas_raw_2016-06-08.xlsx	09/06/2016 15:53
 coordenadas_raw_2016-06-08_old.xlsx	08/06/2016 16:00
 coordenadas_raw_2016-06-21.xlsx	21/06/2016 16:12
 coords_2015-09-09_modif.xlsx	05/11/2015 15:23
 coords_2015-10-11_modif_YC.xlsx	17/11/2015 13:37
 coords_2015-10-11_modif_YC_PACO.xlsx	17/11/2015 17:06
 coords_2015-10-18_modif_YC.xlsx	18/11/2015 17:24
 coords_2015-12-26_modif_YC.xlsx	30/03/2016 19:38
 coords_2016-04-02.xlsx	06/04/2016 17:46
 coords_2016-04-02_YC.xlsx	06/04/2016 18:03
 coords_2016-04-08_YC.xlsx	11/04/2016 13:51
 dataset_y_coords_09_15.xlsx	23/09/2015 17:18
 Datos metaanalysis_18-04-2016.xlsx	19/04/2016 16:24
 FINAL METAANALYSIS_14-6-2016_WITH REVIEWS.xlsx	21/06/2016 16:15
 FINAL METAANALYSIS_16-6-2016_WITH REVIEWS.xlsx	21/06/2016 16:13
 FINAL METAANALYSIS_2016-04-27_WITH REVIEWS.xlsx	25/05/2016 18:05
 FINAL METAANALYSIS_2016-04-27_WITH REVIEWS_FRS.xlsx	27/05/2016 18:44
 FINAL METAANALYSIS_2016-04-29_EXCLUDING REVIEWS.xlsx	08/06/2016 13:06
 FINAL VOTECOUNTING_1-7-2016.xlsx	04/07/2016 15:46
 fitnessdata_2016-06-22.xlsx	22/06/2016 21:00
 Ifs for Bacton_19-3-2016_YC.xlsx	28/03/2016 19:26
 Metaanalysis final_01-05-2015 with coordinates.xlsx	18/05/2015 19:20
 Metaanalysis final_22-05-2015 coords.xlsx	24/06/2015 15:50
 Metaanalysis final_25-06-2015.xlsx	30/06/2015 16:55
 Metaanalysis y coords revisadas_06-08-2015_AH_E.xlsx	23/09/2015 12:57
 Pulido et al_SMI_Data_2016-05-27.xlsx	27/05/2016 18:48
 Pulido et al_SMI_Data_2016-06-21_PACO.xlsx	20/06/2016 16:23

 exclosure_damage_raw.csv	04/07/2016 21:21
 exclosures_cover_raw.csv	04/07/2016 20:49
 sitenames.csv	04/07/2016 20:42
 sites_info_raw.csv	30/06/2016 20:03
 species_info_raw.csv	05/07/2016 15:53

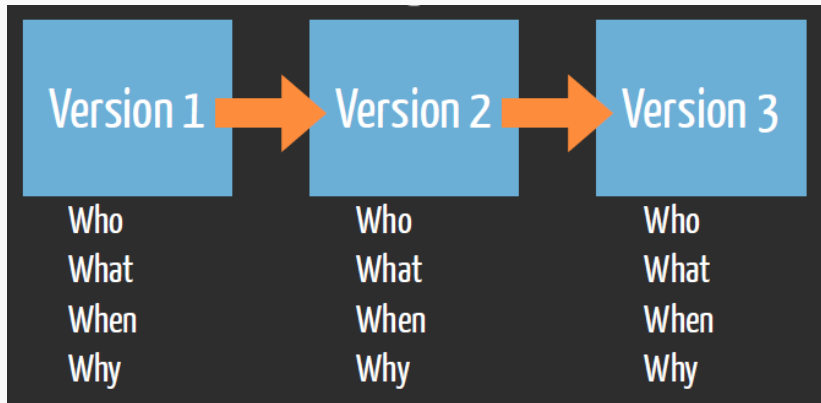
Dropbox keeps record of deleted/edited files for 30 days

Open Science Framework

Automatic version control, no time limit.

The screenshot displays the Open Science Framework (OSF) interface. The top navigation bar includes 'Dashboard', 'My Projects', 'Browse', and a search icon. The user 'Sara Bowman' is logged in. The main content area shows a file named 'manuscript green chemistry.docx'. Below the file name are buttons for 'Delete', 'Check out', 'Download', 'View', and 'Revisions'. A pink arrow points to the 'Revisions' button. The 'Revisions' table lists five versions of the file, with the 'Version ID' column highlighted by a pink box.

Version ID	Date	User	Download	MD5	SHA2
5	2016-03-01 04:51 PM	Sara Bowman	3	605360a9d897969845e	0a15b7a38d21268e87;
4	2016-03-01 04:51 PM	Sara Bowman	2	d36862941d1f3a9834a	0b26a8c8d5aaa9a26d2
3	2016-03-01 04:50 PM	Sara Bowman	1	4f9731f49aea5b8eafa9	1c86e4964c495201460
2	2016-03-01 04:50 PM	Sara Bowman	1	bc165cff2a8ad6b3a8bc	401cdd53dbcb3c54a45
1	2016-03-01 03:32 PM	Sara Bowman	4	96f5aa2525e176ec2e9;	59ec22c26e9510bac3



R. Fitzjohn (<https://github.com/richfitz/reproducibility-2014>)

- Sign up for GitHub
- Install Git
- Introduce yourself
- Create repo on GitHub
- Clone repo in Rstudio
- Make changes, push, pull
- Collaboration

Collaborative writing

Many alternatives

- Rmarkdown + GitHub
- Word + Dropbox
- Google Docs
- Overleaf
- Authorea
- ...



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

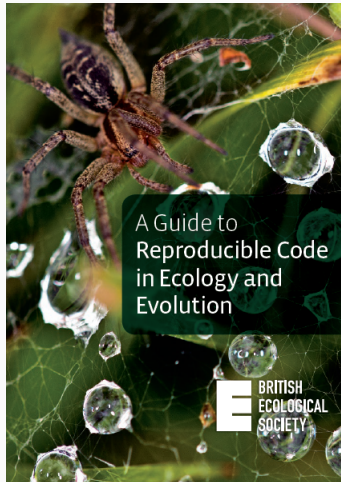
ISSN 1697-2473 / Open access
disponible en www.revistaecosistemas.net

Ciencia reproducible: qué, por qué, cómo

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

<http://www.revistaecosistemas.net/index.php/ecosistemas/article/viewFile/1178/973>

To read more



<https://www.britishecologicalsociety.org/publications/guides-to/>

Happy writing!



Slides and source code available at

<https://github.com/Pakillo/ReproducibleScience>