

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

Francisco Rodriguez-Sanchez (@frod_san)

November 2016

A typical research workflow

1. Prepare data (**EXCEL**)

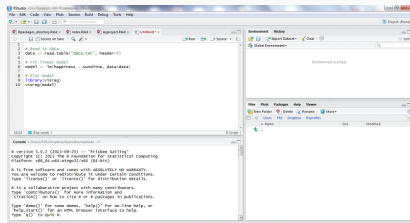
	A	B
1	happiness_index	sunshine_h
2	10.5	978.4
3	6.6	660.9
4	11.3	1093.5
5	9.6	978.9
6	10.9	1135.5
7	9.1	907.0
8	10.6	990.4
9	12.4	1172.9
10	9.6	1025.6
11	10.1	1055.0
12	10.9	1093.7
13	8.9	863.8
14	12.5	1196.6
15	10.0	995.8
16	11.0	1120.2
17	10.3	988.0
18	9.7	987.0
19	9.3	970.4
20	10.9	1076.6
21	9.0	909.8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

data

Ready

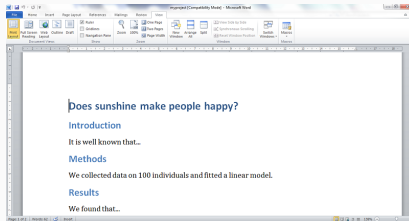
A typical research workflow

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



A typical research workflow

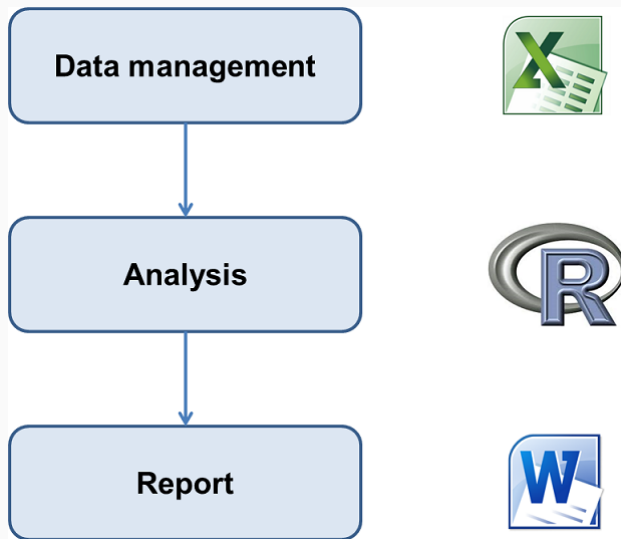
1. Prepare data (**EXCEL**)
2. Analyse data (**R**)
3. Write report/paper (**WORD**)



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



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

- 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?

- 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?

- Fundamental pillar of scientific method

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

- Fundamental pillar of scientific method
- Much less prone to errors
- Automatically regenerate results

- Fundamental pillar of scientific method
- Much less prone to errors
- Automatically regenerate results
- Code reuse & sharing accelerates scientific progress

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

- Fundamental pillar of scientific method
- Much less prone to errors
- Automatically regenerate results
- 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

myproject

```
| - 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

1. Planification (e.g. [DMPTool](#))
2. Collection
3. Metadata description (EML, [Morpho](#))
4. Quality control
5. Storage

Use the cloud: safe, persistent, easy to share

- Dropbox
- OSF
- Figshare, etc
- See all data repositories in www.re3data.org

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 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 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 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 with R 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 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 with R 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 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 with R 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 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 with R code.
- Export data as plain text (txt, csv)
- <http://www.datacarpentry.org/spreadsheet-ecology-lesson/>
- <http://kbroman.org/dataorg/>

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		lake site Jun 12 2012						12-Jun		lake site Jun 19 2012						19-Jun		lake site Jun 26 2012						26-Jun			
3	plot bug1 bug2						avr	SEM	plot bug1 bug2						avr	SEM	plot bug1 bug2						avr	SEM	plot bug1 bug2						avr	SEM		
4	1	T1	1	1	2	T1	2.6	0.51	1	T1	6	85	91	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	T2	0.2	0.2	0.2	
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			0	0	
7	4	T1	1	0	1				4	T1	0	6	6				4	T1	0	14	14	control 0.4		0.4	4	T1	0	39	39	control 0		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		Barn site Jun 12 2012						12-Jun		Barn site Jun 19 2012						19-Jun		Barn Site Jun 26 2012						26-Jun			
22	plot bug1 bug2 general						avr	SEM	plot bug1 bug2 general						avr	SEM	plot bug1 bug2 general						avr	SEM	plot bug1 bug2 general						avr	SEM		
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	5				2	T1	36	74	110				2	T1	65	502	567				2	T1	44	2057	2101					
25	3	T1	0	0	0	T1	2.4	1.288	3	T1	13	0	13	T1	30.6	20.10124	3	T1	10	7	17	T1	19.4	11.92882	3	T1	12	20	32	T2	0.4	0.4	0.4	
26	4	T1	0	0	0	T2	0.4	0.245	4	T1	7	0	7	T2	1	0.774597	4	T1	0	6	6	T2	5	2.1908902	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		0.969556	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																																		

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

Using formatting, comments, etc to convey information

Plot: 2			
Date collected	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			

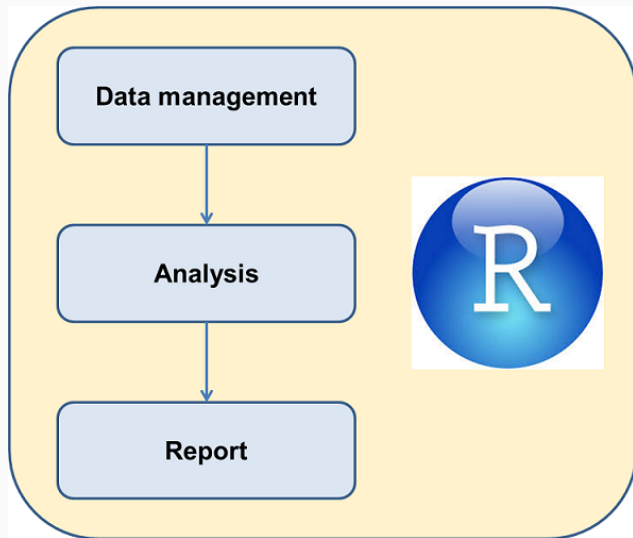
Plot: 2				
Date collected	Species	Sex	Weight	Calibrated
1/8/14	NA			
1/8/14	DM	M	44	Y
1/8/14	DM	M	38	Y
1/8/14	OL			
1/8/14	PE	M	22	Y
1/8/14	DM	M	38	Y
1/8/14	DM	M	48	Y
1/8/14	DM	M	43	Y
1/8/14	DM	F	35	Y
1/8/14	DM	M	43	Y
1/8/14	DM	F	37	Y
1/8/14	PF	F	7	Y
1/8/14	DM	M	45	Y
1/8/14	OT			
1/8/14	DS	M	157	N
1/8/14	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

- Reproducible
- Reusable



Rmarkdown documents

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

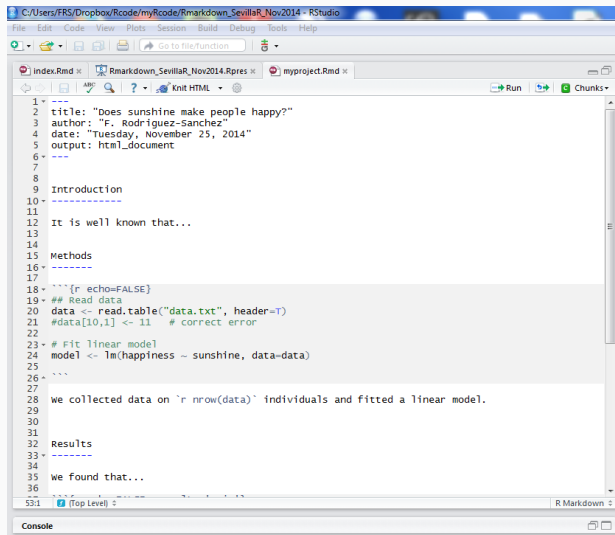


Let's see Rmarkdown in action

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

Example: Does sunshine influence happiness?

See [myproject.Rmd](#)



```
1 ---
2 title: "Does sunshine make people happy?"
3 author: "F. Rodriguez-Sanchez"
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) >
```


Does sunshine make people happy?

F. Rodríguez-Sánchez

Tuesday, November 25, 2014

Introduction

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

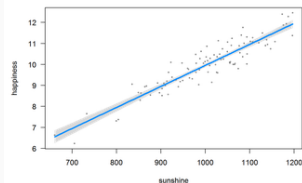
Methods

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	-0.1527926	0.8786756
sunshine	0.0100228	0.0004232	23.6533264	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!

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, ...

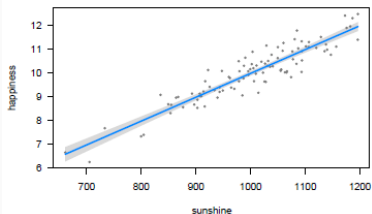
Methods

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



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, ...

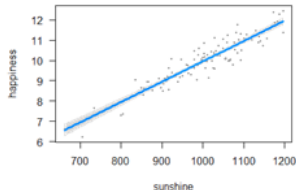
Methods

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

```
| bibliography: references.bib
```

```
---{r citations_setup, include=FALSE}
```

```
library(knitr)
cite_options(citation_format = "pandoc")
```

```
Introduction
```

It is well known that individual well-being can be influenced by climatic conditions ``r citep("10.1016/j.ecolecon.2004.06.015")``. However, ...

See [output](#)

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](https://doi.org/10.1016/j.ecolecon.2004.06.015).

Useful: rcrossref addin

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

Can write full thesis in Rmarkdown!

See `thesis.Rmd`.

See `thesis.pdf`.

Rmarkdown cheat sheet

R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com



.Rmd files

An R Markdown (.Rmd) file is a record of your research. It contains the code that a scientist needs to reproduce your work along with the narration that a reader needs to understand your work.

Reproducible Research

At the click of a button, or the type of a command, you can rerun the code in an R Markdown file to reproduce your work and export the results as a finished report.

Dynamic Documents

You can choose to export the finished report as a html, pdf, MS Word, ODT, RTF, or markdown document, or as a html or pdf based slide show.

Workflow

- 1 Open a new .Rmd file** at File > New File > R Markdown. Use the wizard that opens to pre-populate the file with a template.
- 2 Write document** by editing template.
- 3 Knit document to create report** (Use knit button or `render()` in knitr).
- 4 Preview Output** in IDE window.
- 5 Publish (optional)** to web or server. Sync publish button to accounts at:
 - GitHub
 - Amazon S3
 - RStudio ConnectRebuild document. Find in document. File path to output document.
- 6 Examine build log** in R Markdown console.
- 7 Use output file** that is saved alongside .Rmd.

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.

.Rmd structure

YAML Header
Optional section of header in .Rmd documents written as key-value pairs (YAML).
• At start of file
• Between lines of `---`

Text
Narration formatted with markdown, mixed with code chunks.

Code chunks
Chunks of embedded code. Each chunk:
• Begins with ````{r}`
• Ends with `````
• R Markdown will run the code and append the results to the doc.
• It will use the location of the .Rmd file as the working directory.

Embed code with knitr syntax

Inline code
Insert with `<code>`. Results appear as text without code.

Code chunks
One or more lines surrounded with ````{r}` and `````. Place chunk options within curly braces, after the ````{r}`.

Global options
Set with knitr: `opts_chunk$set()`, e.g. `opts_chunk$set(echo = TRUE)`.

Important chunk options

cache - cache results for future knits (default = FALSE)
cache.path - directory to save cached results (default = "cache")
child - file(s) to knit and then include (default = NULL)
collapse - collapse all output into single block (default = FALSE)
comment - prefix for each line of results (default = "#")

dependencies - chunk dependencies for caching (default = NULL)
echo - Display code in output document (default = TRUE)
engine - code language used in chunk (default = "R")
error - Display error messages in doc (TRUE) or stop render when errors occur (FALSE) (default = TRUE)
eval - Run code in chunk (default = TRUE)

fig.align - "left", "right", or "center" (default = "center")
fig.cap - figure caption as character string (default = NULL)
fig.height, **fig.width** - Dimensions of plots in inches
highlight - highlight source code (default = FALSE)
include - include chunk in doc after running (default = TRUE)
message - display code messages in document (default = TRUE)
results (default = "markup")
"asis" - passthrough results
"hide" - do not display results
"hold" - put all results below all code
tidy - tidy code for display (default = FALSE)
warning - display code warnings in document (default = TRUE)

render("doc.Rmd")
`params = list(n = 1, d = as.Date("2015-01-01"))`

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 of `params`.
- 2 Call parameters**
Call parameter values in code as `params$name`.
- 3 Set parameters**
Set values with `knitr::params` or the parameters argument of `render()`.

RStudio® is a trademark of RStudio, Inc. | © 2016 RStudio | <http://www.rstudio.com> | +1 617 549 3212 | <http://cheatsheets.rstudio.com>

Learn more at rmarkdown.rstudio.com • RStudio IDE 0.98.870 • Updated: 02/16



R Markdown Reference Guide

Learn more about R Markdown at rmarkdown.rstudio.com
Learn more about Interactive Docs at shiny.rstudio.com/articles

Contents:

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

Syntax	Becomes
<p>Make a code chunk with three back ticks followed by an <code>r</code> in braces. End the chunk with three back ticks:</p> <pre>```{r} paste("Hello", "World!") ```</pre>	<p>Make a code chunk with three back ticks followed by an <code>r</code> in braces. End the chunk with three back ticks:</p> <pre>paste("Hello", "World!") ## [1] "Hello World!"</pre>
<p>Place code inline with a single back ticks. The first back tick must be followed by an <code>R</code>, like this <code>r paste("Hello", "World!")</code>.</p>	<p>Place code inline with a single back ticks. The first back tick must be followed by an <code>R</code>, like this <code>Hello World!</code>.</p>
<p>Add chunk options within braces. For example, <code>echo=FALSE</code> will prevent source code from being displayed:</p> <pre>```{r eval=TRUE, echo=FALSE} paste("Hello", "World!") ```</pre>	<p>Add chunk options within braces. For example, <code>echo=FALSE</code> will prevent source code from being displayed:</p> <pre>## [1] "Hello World!"</pre>

Learn more about chunk options at <http://yihui.name/knitr/options>

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

RStudio

Updated 10/30/2014

© 2014 RStudio, Inc. All rights reserved.

Managing software dependencies

Managing package dependencies in R

- sessionInfo (or session_info)
- switchr
- rctrack
- checkpoint
- packrat
- docker

Version control

"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



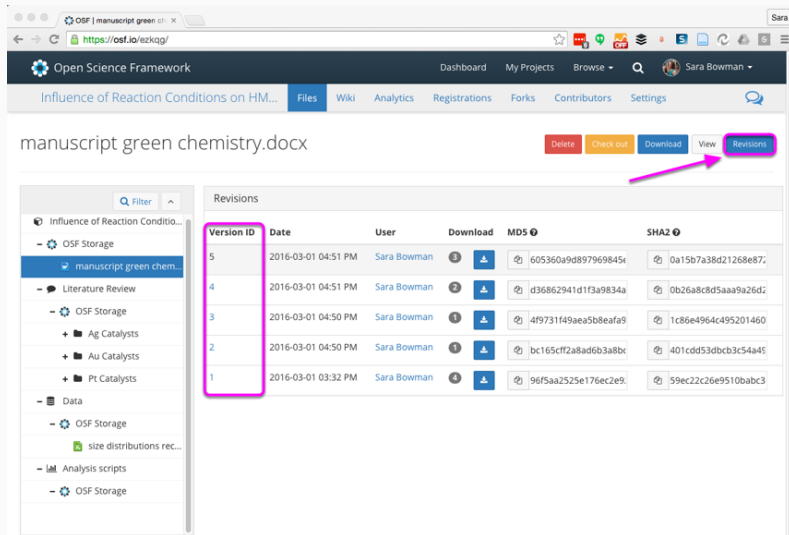
FINAL_rev.18.comments7.
corrections9.MORE.30.doc



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

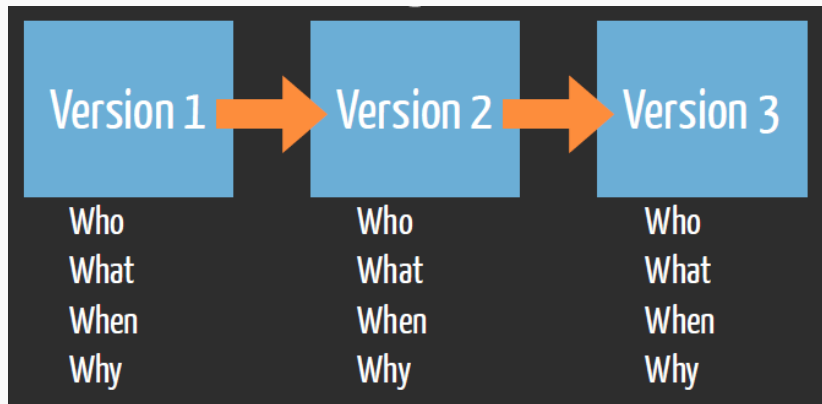
Dropbox keeps record of deleted/edited files for 30 days

Automatic version control, no time limit.



The screenshot displays the Open Science Framework (OSF) web interface. The browser address bar shows the URL <https://osf.io/ezkqg/>. The page title is "manuscript green chemistry.docx". The navigation bar includes links for Dashboard, My Projects, Browse, Search, and a user profile for Sara Bowman. The main content area features a sidebar on the left with a search filter and a list of project components, including "Influence of Reaction Conditions on HM...", "OSF Storage", "Literature Review", "Ag Catalysts", "Au Catalysts", "Pt Catalysts", "Data", "size distributions rec...", and "Analysis scripts". The main content area displays a table of revisions for the document. The "Revisions" button is highlighted with a pink box, and a pink arrow points to it. The "Revisions" table has columns for Version ID, Date, User, Download, MD5, and SHA2. The "Version ID" column is highlighted by a pink box.

Version ID	Date	User	Download	MD5	SHA2
5	2016-03-01 04:51 PM	Sara Bowman		605360a9d897969845f	0a15b7a38d21268e87
4	2016-03-01 04:51 PM	Sara Bowman		d36862941d1f3a9834a	0b26a8c8d5aaa9a26d
3	2016-03-01 04:50 PM	Sara Bowman		4f9731f49aea5b8eafa9	1c86e4964c495201460
2	2016-03-01 04:50 PM	Sara Bowman		bc165cff2a8ad6b3a8bc	401cdd53dbcb3c54a4f
1	2016-03-01 03:32 PM	Sara Bowman		96f5aa2525e176ec2e9	59ec22c26e9510abc3



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

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

Happy writing!