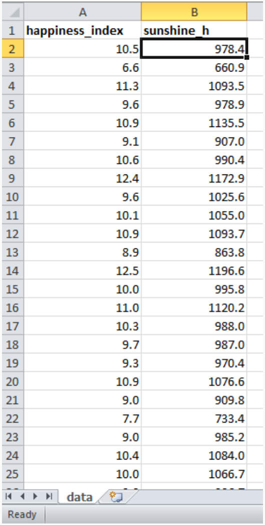
Doing reproducible science: from your hard-won data to a publishable manuscript without going mad

Francisco Rodriguez-Sanchez (@frod_san)

February 2017

A typical research workflow

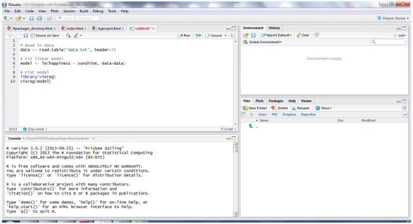
1. Prepare data (**EXCEL**)



	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

A typical research workflow

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

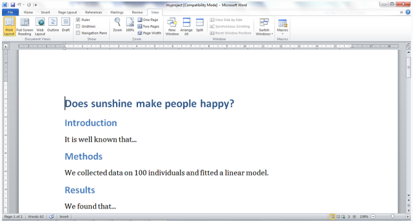


```r

# Read in data
data <- read.table("data.txt", header=T)

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

# Plot model
library(visreg)
visreg(model)
```

```
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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

A typical research workflow

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



Does sunshine make people happy?

Introduction

It is well known that...

Methods

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

Results

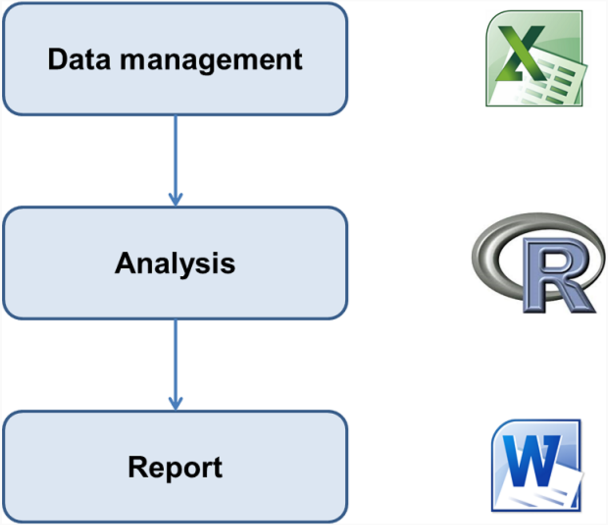
We found that...

A typical research workflow

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

THIS EMAIL CHAIN IS SO FASCINATING
PLEASE MAKE SURE TO KEEP SENDING THEM TO ME
memegenerator.org

This workflow is broken



Problems of a broken workflow

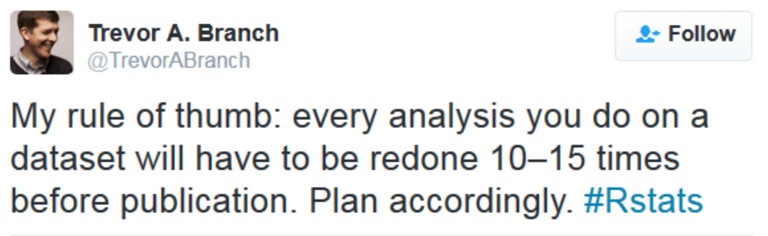
- How did you do this? What analysis is behind this figure? Did you account for . . . ?

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

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

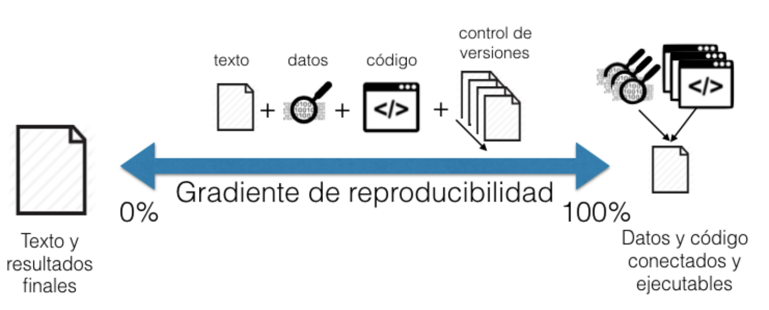
WHAT is Reproducible Science?

Reproducible Science: WHAT

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

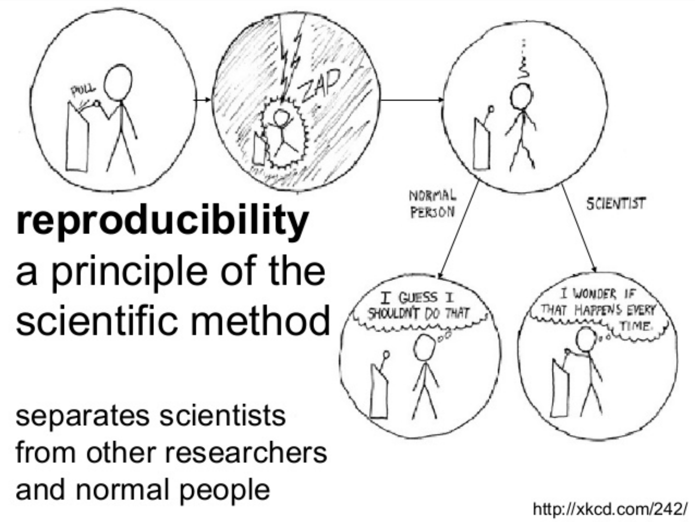

control de versiones
texto
datos
código
+
+
+
Gradiente de reproducibilidad
0%
100%
Texto y resultados finales
Datos y código conectados y ejecutables

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

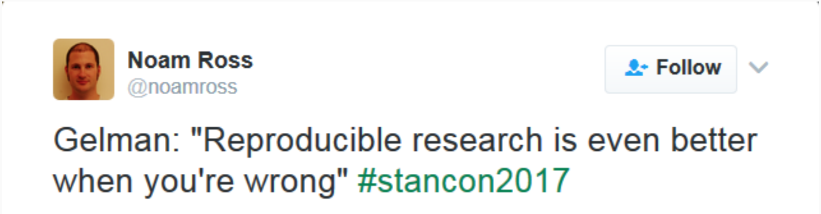
separates scientists from other researchers and normal people

PULL
ZAP
NORMAL PERSON
SCIENTIST
I GUESS I SHOULDN'T DO THAT
I WONDER IF THAT HAPPENS EVERY TIME.

http://xkcd.com/242/

Carole Goble

http://www.slideshare.net/carolegoble/ismb2013-keynotecleangoble



Noam Ross
@noamross

Gelman: "Reproducible research is even better when you're wrong" #stancon2017

Reproducible Science: WHY

- 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?

Reproducible Science: HOW

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

File organisation

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

File organisation

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

File organisation

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

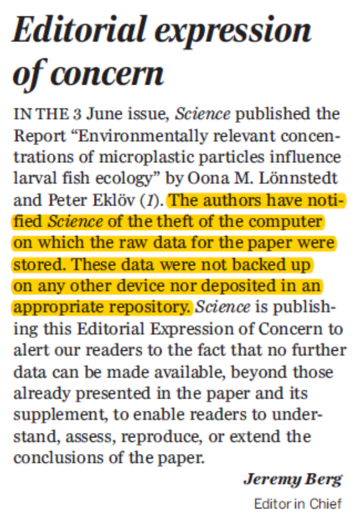
File organisation

- 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
```

Data management



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

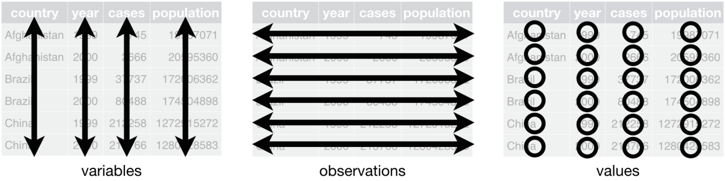
http://science.sciencemag.org/content/354/6317/1242.1

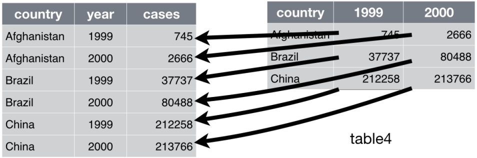
Storage

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

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

Tidy data





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

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

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

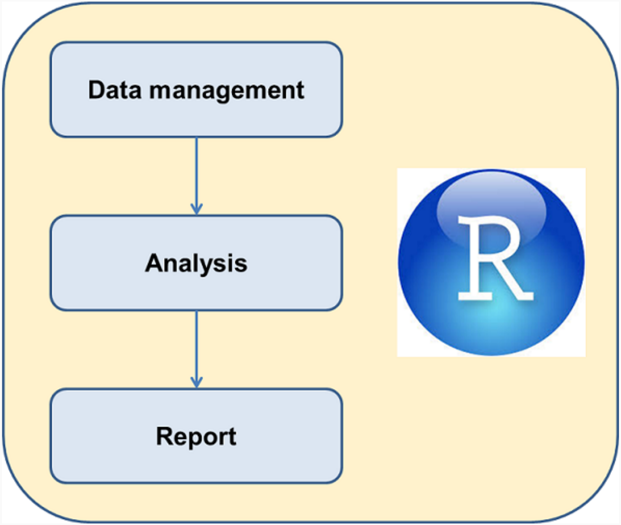
- Put **variables** in **columns** (things you are measuring: height, weight, sex)
- Each **observation** in one **row** (e.g. individuals).
- Avoid spaces, numbers, and special characters in column names.
- Always write zero values, to distinguish from blank/missing data.
- Use blank/empty cells, or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

Data analysis

Always use code

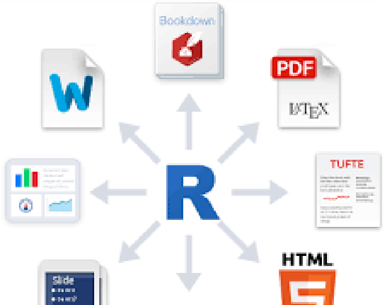
- Reproducible
- Reusable

Dynamic reports



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
    - …

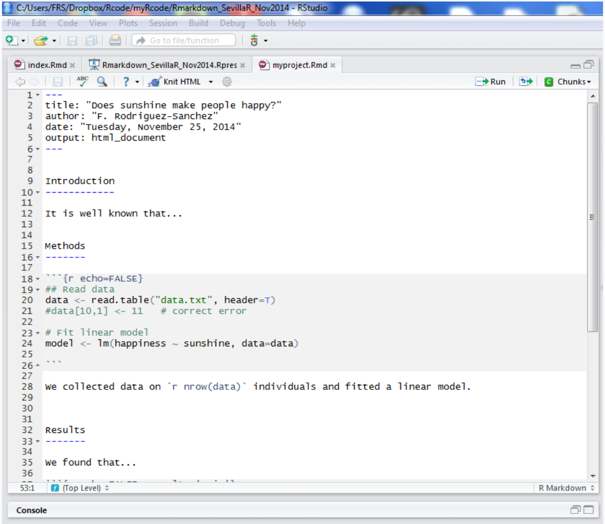


Let's see Rmarkdown in action

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

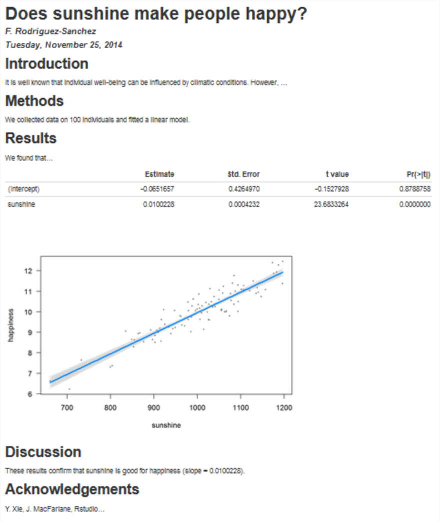
Example: Does sunshine influence happiness?

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



~~~
---
title: "Does sunshine make people happy?"
author: "F. Rodriguez-Sanchez"
date: "Tuesday, November 25, 2014"
output: html_document
---


Introduction
------------

It is well known that...


Methods
-------

```{r echo=FALSE}
## Read data
data <- read.table("data.txt", header=T)
#data[10,1] <- 11   # correct error

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

```

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



Results
-------

We found that...

~~~

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).

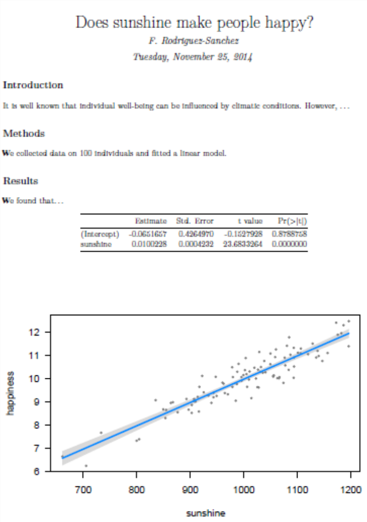
Acknowledgements

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

Spotted error in the data? No problem!

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

Other formats: PDF, Word



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, …

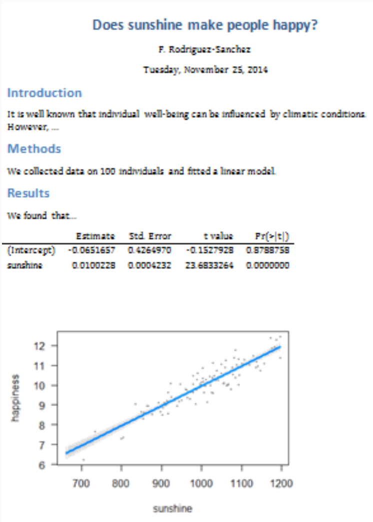
Methods

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.

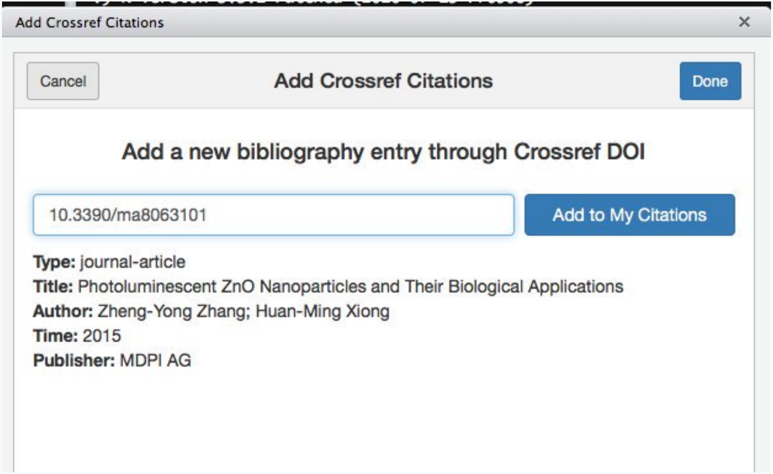
Results

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 by DOI

rcrossref addin



Add a new bibliography entry through Crossref DOI

10.3390/ma8063101

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

citr addin

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

Manuscript templates

- rticles
- rmdTemplates

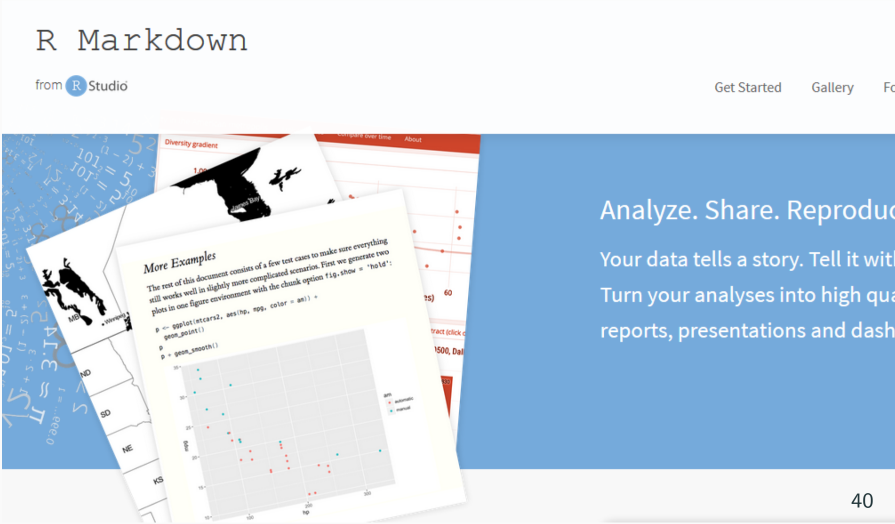
Can write full thesis in Rmarkdown!

See `thesis.Rmd`.

See `thesis.pdf`.

Rmarkdown website

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



R Markdown

Analyze. Share. Reproduc

Your data tells a story. Tell it wit

Turn your analyses into high qua

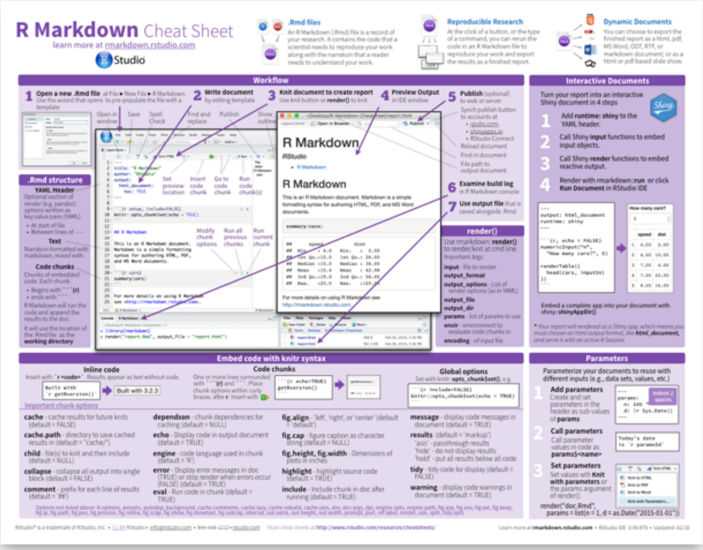
reports, presentations and dash

More Examples

The rest of this document consists of a few test cases to make sure everything still works well in slightly more complicated scenarios. First we generate two plots in one figure environment with the chunk option `fig.show = 'hold'`:

```
p <- ggplot(mtcars2, aes(hp, mpg, color = am)) +
  geom_point()
p
p + geom_smooth()
```

Rmarkdown cheat sheet



R Markdown Cheat Sheet

learn more at rmarkdown.rstudio.com

RStudio

.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()** to knit

4 **Preview Output** in IDE window

5 **Publish** (optional) to web or server

~~~
---
title: "R Markdown"
author: "RStudio"
output:
  html_document:
    toc: TRUE
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
~~~

R Markdown

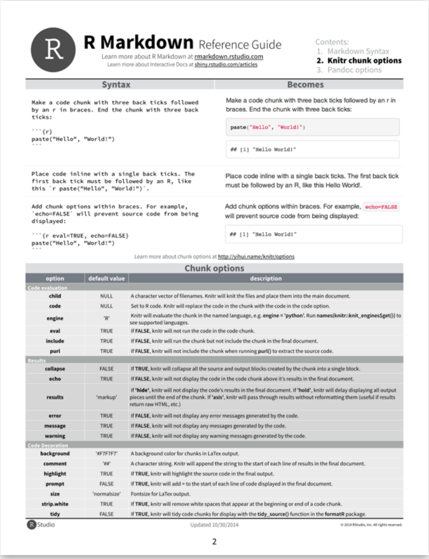
This is an R Markdown document.
Markdown is a simple formatting
syntax for authoring HTML, PDF,
and MS Word documents.

```{r cars}
summary(cars)
```

For more details on using R Markdown
see <http://rmarkdown.rstudio.com>.
```

```
> library(rmarkdown)
> render("report.Rmd", output_file = "report.html")
```

R Markdown

RStudio

- R Markdown

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents.

```
summary(cars)
```

```
##      speed          dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

For more details on using R Markdown see http://rmarkdown.rstudio.com.

6 **Examine build log** in R Markdown console

7 **Use output file** that is saved alongside .Rmd

.Rmd structure

YAML Header
Optional section of render (e.g. pandoc) options written as key:value pairs (YAML).
- At start of file
- Between lines of - - -

Text
Narration formatted with markdown, mixed with:

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**

render()

Use rmarkdown::**render()** to render/knit at cmd line. Important args:

input - file to render
output_format
output_options - List of render options (as in YAML)
output_file
output_dir
params - list of params to use
envir - environment to evaluate code chunks in
encoding - of input file

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 rmarkdown::**run** or click **Run Document** in RStudio IDE

```
---
output: html_document
runtime: shiny
---

```{r, echo = FALSE}
numericInput("n",
  "How many cars?", 5)

renderTable({
  head(cars, input$n)
})
```
```

How many cars?

5

	speed	dist
1	4.00	2.00
2	4.00	10.00
3	7.00	4.00
4	7.00	22.00
5	8.00	16.00

Embed a complete app into your document with shiny::**shinyAppDir()**

* *Your report will rendered as a Shiny app, which means you must choose an html output format, like* ***html_document****, and serve it with an active R Session.*

Embed code with knitr syntax

Inline code

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

```
Built with
`r getRversion()`
```

Built with 3.2.3

Code chunks

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

```
```{r echo=TRUE}
getRversion()
```
```

```
getRversion()
## [1] '3.2.3'
```

Global options

Set with knitr::**opts_chunk$set()**, e.g.

```
```{r include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
```

Important chunk options

cache - cache results for future knits (default = FALSE)
cache.path - directory to save cached results in (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 = '##')

dependson - 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 = 'default')
fig.cap - figure caption as character string (default = NULL)
fig.height, fig.width - Dimensions of plots in inches
highlight - highlight source code (default = TRUE)
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)

Options not listed above: R.options, aniopts, autodep, background, cache.comments, cache.lazy, cache.rebuild, cache.vars, dev, dev.args, dpi, engine.opts, engine.path, fig.asp, fig.env, fig.ext, fig.keep, fig.lp, fig.path, fig.pos, fig.process, fig.retina, fig.scap, fig.show, fig.showtext, fig.subcap, interval, out.extra, out.height, out.width, prompt, purl, ref.label, render, size, split, tidy.opts

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**

```
---
params:
  n: 100
  d: !r Sys.Date()
---
```

Indent 2 spaces

2 **Call parameters**
Call parameter values in code as **params$<name>**

```
Today's date
is `r params$d`
```

3 **Set parameters**
Set values with **Knit with parameters** or the params argument of render():

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

RStudio® is a trademark of RStudio, Inc. • CC BY RStudio • info@rstudio.com • 844-448-1212 • rstudio.com More cheat sheets at http://www.rstudio.com/resources/cheatsheets/ Learn more at **rmarkdown.rstudio.com** • RStudio IDE 0.99.879 • Updated 02/16

https://www.rstudio.com/wp-content/uploads/2016/03/
```

# Rmarkdown reference guide



**R Markdown** Reference Guide

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

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

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

## Chunk options

| option | default value | description |
|---|---|---|
| Code evaluation | | |
| **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. **engine = 'python'**. Run **names(knitr::knit_engines$get())** 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. |
| **purl** | TRUE | If **FALSE**, knitr will not include the chunk when running **purl()** to extract the source code. |
| Results | | |
| **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 **'hide'**, 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** | '#F7F7F7' | 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. |
| **prompt** | FALSE | If **TRUE**, knitr will add > to the start of each line of code displayed in the final document. |
| **size** | 'normalsize' | 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 **tidy_source()** function in the **formatR** package. |

Studio

Updated 10/30/2014

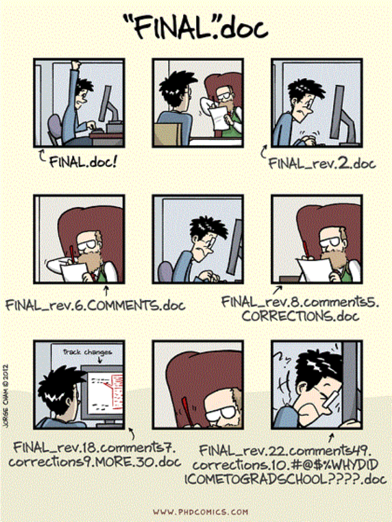
https://www.rstudio.com/wp-content/uploads/2015/03/

# Managing software dependencies

# Managing package dependencies in R

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

# Version control

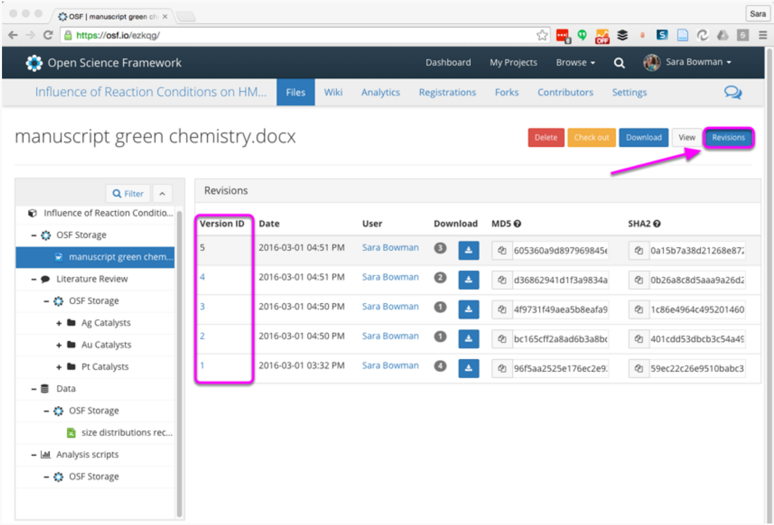

"FINAL".doc
FINAL.doc!
FINAL_rev.2.doc
FINAL_rev.6.COMMENTS.doc
FINAL_rev.8.comments5.
CORRECTIONS.doc
Track changes
FINAL_rev.18.comments7.
corrections9.MORE.30.doc
FINAL_rev.22.comments49.
corrections.10.#@$%WHYDID
ICOMETOGRADSCHOOL????.doc
JORGE CHAM © 2012
WWW.PHDCOMICS.COM

# Dropbox

Dropbox keeps record of deleted/edited files for 30 days

# Open Science Framework

Automatic version control, no time limit.

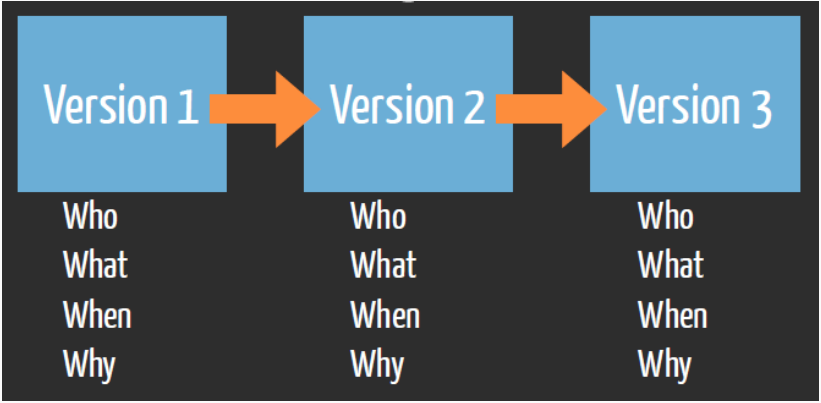


manuscript green chemistry.docx

Revisions

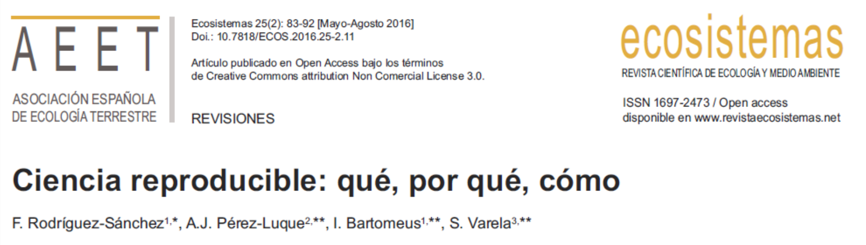
| Version ID | Date | User | Download | MD5 | SHA2 |
|---|---|---|---|---|---|
| 5 | 2016-03-01 04:51 PM | Sara Bowman | 3 | 605360a9d897969845e | 0a15b7a38d21268e87 |
| 4 | 2016-03-01 04:51 PM | Sara Bowman | 2 | d36862941d1f3a9834a | 0b26a8c8d5aaa9a26d |
| 3 | 2016-03-01 04:50 PM | Sara Bowman | 1 | 4f9731f49aea5b8eafa9 | 1c86e4964c495201460 |
| 2 | 2016-03-01 04:50 PM | Sara Bowman | 1 | bc165cff2a8ad6b3a8b | 401cdd53dbcb3c54a4 |
| 1 | 2016-03-01 03:32 PM | Sara Bowman | 4 | 96f5aa2525e176ec2e9 | 59ec22c26e9510babc3 |

# Git & GitHub



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

# To read more



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,**]

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

# Happy writing!

Slides and source code available at

https://github.com/Pakillo/ReproducibleScience