More complex RMardown example

Joseph Casillas

2019-02-07

# RMarkdown

## ¿Qué?

RMarkdown es un tipo de lenguaje que sirve para generar distintos tipos de documentos que son comunes en el mundo académico (.docx, .pdf, .html). Lo más probable es que hayas usado un programa como Microsoft Word para generar documentos para tus clases. Este tipo de programa se clasifica WYSIWYG (what you see is what you get), lo cual significa que el usuario puede ver exactamente cómo va a quedar el documento a la vez que trabaja en él. En cambio, si usamos un lenguaje para generar el documento—en este caso el lenguaje es RMarkdown—tenemos que darle al programa instrucciones específicas acerca de cómo queremos formatear distintos parámetros (i.e., el tamaño de la fuente, negrita, cursiva, etc.). A primera vista usar un lenguaje como RMarkdown parece introducir una complicación innecesaria, pues teclear directamente en Word es bastante fácil. Pero, como veremos abajo, un lenguaje como RMarkdown nos puede facilitar muchas tareas comunes que tenemos que hacer al llevar a cabo una investigación académica. Por lo tanto, podemos considerar que RMarkdown es una herramienta para la investigación.

## ¿Cómo?

RMarkdown tiene su propia sintaxis. A saber…

### Texto

Al teclear de manera normal (es decir, sin añadir nada), producimos texto normal. Podemos terminar una línea y comenzar un párrafo nuevo en el documento añadiendo dos espacios al final.

Usamos el asterisco (\*) para *escribir en cursiva*. Usamos dos asteriscos (\*\*) para **escribir en negrita**. También podemos cambiar el texto de otras maneras, por ejemplo con unsuperscript o ~~tachando las palabras~~. Incluso podemos incluir [enlaces](https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf) (pincha ese enlace para aprender más acerca de la sintaxis).

### Listas

Podemos producir:

* unordered lists
* usando el guion (-)

y también…

1. ordered lists
2. usando números (1. , 2. , etc.)

### Tablas

También es posible crear tablas…

|  |  |
| --- | --- |
| Table Header | Second Header |
| Table Cell | Cell 2 |
| Cell 3 | Cell 4 |

## ¿Por qué?

Hemos visto un poco acerca de cómo funciona RMarkdown. Ahora vamos a ver algunas de las razones por las que es útil para la investigación.

### Funcionalidad

RMardown se escribe en un archivo .Rmd, que es un simple archivo de texto. Escribir en un archivo de texto aporta varias ventajas.

* Un documento de texto no pesa
* No necesitamos software caro (lo podemos abrir y editar con cualquier editor de texto)
* Funciona en cualquier plataforma

### Output variado

El archivo .Rmd puede convertirse en un documento word, un pdf, o una página web. Sólo tenemos que cambiar output: "html\_document" a “pdf\_document” o “word\_document”.

### La investigación reproducible

Normalmente nuestros análisis estadísticos y los documentos que usamos para presentarlos se mantienen separados. Por ejemplo, hacemos un ANOVA en SPSS y luego reportamos los resultados en nuestro paper (que luego se manda a una revista para que se publique). Esta separación es el enemigo. Es contraproducente si la meta es hacer investigaciones reproducibles. Mediante RMarkdown podemos incluir nuestros análisis (scripts de R) en el mismo documento.

# R

## Comments

#############################  
# Always comment your code! #  
#############################  
  
# This is a comment  
2 + 2

## [1] 4

## Matemáticas básicas

#####################  
# R as a calculator #  
#####################  
  
2 + 2

## [1] 4

4^2

## [1] 16

(12 \* 15) / 2

## [1] 90

## Asignación de variables

# R uses objects  
  
# Store the value '2' to the object 'x'  
x <- 2  
print(x)

## [1] 2

# Perform an operation on 'x'  
x + 4

## [1] 6

## Instalar paquetes

# install a package  
install.packages("tidyverse")

## Analizar datos

library(tidyverse)  
glimpse(cars)

## Observations: 50  
## Variables: 2  
## $ speed <dbl> 4, 4, 7, 7, 8, 9, 10, 10, 10, 11, 11, 12, 12, 12, 12, 13, …  
## $ dist <dbl> 2, 10, 4, 22, 16, 10, 18, 26, 34, 17, 28, 14, 20, 24, 28, …

my\_analysis <- lm(speed ~ dist, data = cars)  
  
summary(my\_analysis)

##   
## Call:  
## lm(formula = speed ~ dist, data = cars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.5293 -2.1550 0.3615 2.4377 6.4179   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.28391 0.87438 9.474 1.44e-12 \*\*\*  
## dist 0.16557 0.01749 9.464 1.49e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.156 on 48 degrees of freedom  
## Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438   
## F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

## Make a plot

ggplot(cars, aes(x = dist, y = speed)) +   
 geom\_point() +   
 geom\_smooth(method = 'lm')

