Módulo 5. Informes con RStudio y R Markdown

Curso Herramientas para Data Science

Prof. Jose Jacobo Zubcoff Vallejo, PhD Universidad de Alicante

2017 @ Licencia Creative Common BY

Módulo 5. Informes con RStudio y R Markdown

Una de las grandes ventajas de usar RStudio es que permite generr informes a partir de R. En realidad utiliza un lenguaje llamado RMarkdown para enriquecer el informe final. Este permite incluir texto, comandos de R, imágenes y gráficos a un documento. Pero lo más importante es que permite que se reproduzca el análisis realizado y si se incorporan nuevos datos, los resultados se actualizarán.

Realmente es una herramienta muy potente. Vamos a conocer un poco la metodología para generar informes con RStudio y RMarkdown.

Estos informes están compuestos por dos partes:

- el fichero que genera el informe (en formato RMarkdown cuya extensión es .RMD)
- el informe generado (en Word, PDF o HTML)

El fichero generador del informe se debe escribir siguiendo las recomendaciones del lenguaje RMarkdown (http://rmarkdown.rstudio.com). Nuestra recomendación es crear el informe desde RStudio, éstegenera un fichero de muestra básico que sirve para editar su contenido para adaptarlo a nuestras necesidades.

Lección 1. Informes con RStudio y RMarkdown

Para crear un informe en RStudio solo hay que elegir la opción "File"+"New File"+"R Markdown".

Podemos elegir el destino del informe: una página web que se podrá ver con un navegador, un documento Word o un PDF.

Esto abrirá un fichero que incluye ya contenido de texto y de lenguaje RMarkdown.

Primer informe con RMarkdown

Una vez seleccionado el formato de salida del informe, aparecerá en el panel de guiones o scripts, el fichero que acabamos de crear. Este contiene una cabecera, con el título, autor, fecha y formato de salida del informe. En este caso he seleccionado como formato de salida una página Web por lo que pone "html document".

Es recomendable utilizar ese informe como plantilla, y editar el fichero cambiando el contenido respetando la estructura del fichero, en cuanto a que hay partes de texto y hay partes de comandos.

Los comandos se ejecutarán cada vez que se quiera generar la salida del informe (PDF, Word o HTML), para lo que se usarán las opciones de menú "Code"+"Run" (o el botón dedicado que hay en el panel de scripts).

Los comandos irán entre unas líneas de cabecera y pie, como muestra la figura 3.

Tipos de salida del Informe

Podremos generar informes en Word, PDF o HTML (formato página web visible con cualquier navegador). En cada formato cambia el aspecto del mismo.

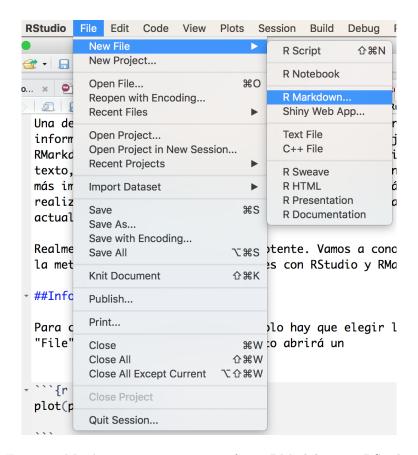


Figure 1: Menú para crear un nuevo informe RMarkdown en RStudio

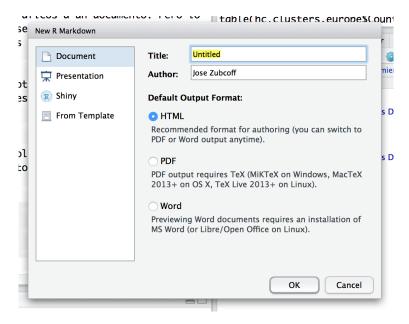


Figure 2: Opciones para crear un nuevo informe RMarkdown en RStudio

```
RStudio File Edit Code View Plots Session Build Debug Profile
                                                                                                                                                                                                                                  Tools Win
                                                                                                                                                                                                                                      RStudio
Contenido de curso Herramientas ... 🗶 🎱 Modulo 5 - Informes con R Markdo...* 🗴 👰 Untitled1 🗴 \gg 👝 🗔

☐ In the second of the se
                                                                                                                                          2 title: "Primer Informe con RMarkdown'
                  author: "Jose Zubcoff"
                 date: "9/28/2017'
                output: html_document
         6
         8 - ```{r setup, include=FALSE}
         9
                  knitr::opts_chunk$set(echo = TRUE)
     10
     11
     12 - ## R Markdown
     13
                This is an R Markdown document. Markdown is a simple formatting
     14
                   syntax for authoring HTML, PDF, and MS Word documents. For more
                    details on using R Markdown see <a href="http://rmarkdown.rstudio.com">http://rmarkdown.rstudio.com</a>>.
                  When you click the **Knit** button a document will be generated
                  that includes both content as well as the output of any embedded R
                   code chunks within the document. You can embed an R code chunk
                    like this:
     17
     2:1
                    Primer Informe con RMarkdown $
                                                                                                                                                                                                                           R Markdown #
```

Figure 3: Nuevo informe RMarkdown en RStudio

Ejemplo de informe con salida en Word

Cuando se genera el informe en fichero Word, el documento es completamente editable. Es un documento de Word normal. Si has incluído gráficos verás esos gráficos como imágenes en el documento de Word. El título del documento se genera en tipo de letra Calibrí, tamaño 18 puntos, en un color azul marino.

El cuerpo del documento usa la tipografía Cambria con tamaño 12. Esto puede cambiar según el sistema operativo, o incluso puede cambiar con actualizaciones siguientes. Por supuesto estas características se pueden cambiar editando las cabeceras YAML y las hojas de estilo CSS y valores por defecto de RStudio, pero se sale fuera del contenido de este curso. Sin embargo, es muy fácil cambiar estas características directamente en Word.

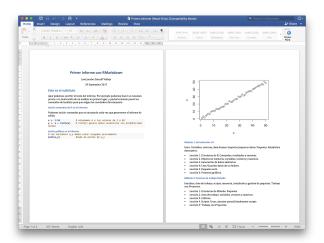


Figure 4: Documento en Word del primer informe en RMarkdown

El mismo informe con salida en PDF

Como se puede observar, cuando se genera un documento PDF la tipografía es distinta a cuando generamos un Word (también cuando generamos HTML). En general, los informes que generan un PDF usan tipografías tipo Times, con tamaño estándar para el cuerpo de 13 y 16 para títulos. Para editar ficheros PDF se necesitan herramientas especiales como Adobe Acrobat u otras alternativas. Pero lo importante es que podemos cambiar o editar el propio generador del PDF para obtener un documento final, que no necesite ninguna edición.

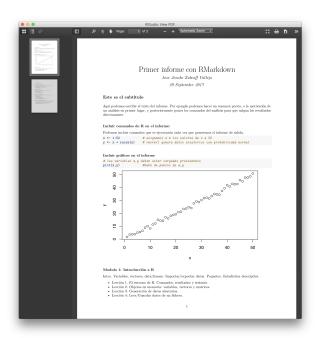


Figure 5: Documento en PDF del primer informe en RMarkdown

El mismo informe con salida en HMTL

También se puede generar informes HTML, que resultan muy útiles para mantener actualizadas páginas web con análisis, o centros de mandos con gráficos, etc. La ventaja de una salida en HTML es que es visible en cualquier navegador o explorador. También se puede editar con cualquier editor de páginas webs (o incluso con cualquier editor de textos). Pero la idea es no tener que editar el documento final sino editar el generador de informes y obtener el informe final.

Lección 2. Comandos básicos. Comentarios. Inclusión de resultados en el informe.

Cualquier texto que se incluya en el fichero generador del informe aparacerá tal cual. Si queremos incluir subt'ítulos, o caracteres en negrita, o ecuaciones, ..., tendremos que introducir una sintaxis específica.

Por ejemplo, para introducir títulos, subtítulos, y cabeceras podemos usar el caracter #.

Texto en el generador del informe:

"## Este es el título"

Este es el título

Texto en el generador del informe:



Figure 6: Documento en HTML del primer informe en RMarkdown

"### Este es un subtítulo"

Este es un subtítulo

Texto en el generador del informe:

"#### Este es un subsubtítulo"

Este es un subsubtítulo

Para resaltar una palabra o una frase debemos ponerla entre doble asteriscos.

Este es el texto en el generador del informe:

Dentro de una frase se pueden resaltar **varias palabras**.

El resultado es:

Dentro de una frase se pueden resaltar varias palabras.

Para poner una palabra (o una frase) en cursiva debemos ponerla entre asteriscos simples.

Este es el texto en el generador del informe:

Dentro de una frase se pueden resaltar *varias palabras*.

El resultado es:

Dentro de una frase se pueden resaltar varias palabras.

Para poner una palabra (o una frase) en cursiva debemos ponerla entre asteriscos simples.

Este es el texto en el generador del informe:

Dentro de una frase se pueden resaltar _varias palabras_.

El resultado es:

Dentro de una frase se pueden resaltar varias palabras.

Lección 3. Mezclar título, cabeceras y texto con resultados de R en el informe.

Una vez que controlamos como poner títulos, subtítulos y cabeceras y texto, podemos incluir análisis y resultados de R en el informe. Para ello debemos escribir el código de R entre una cabecera y pie que permite interpretar la sintaxis de R.

Cabecera y pie que indica el código de R ejecutable en el generador de informes:

Figure 7: Notación de la cabecera y pie de código con RMarkdown

```
#aquí el código de R

"'{r datos_aleatorios}
```

Por ejemplo, podemos calcular la media de unos valores generados con los siguientes comandos (que deberán ir entre cabecera y pie de comandos del generador de informes):

Figure 8: Cabecera y pie de código ejecutable en el generador de informes con RMarkdown

Y el resultado es:

```
set.seed(1234)  #semilla de aleatorización para reproducibilidad
x <- 1:50  # asignamos a x los valores de 1 a 50
y <- x + rnorm(x)  # rnorm() genera datos aleatorios con probabilidad normal
mean(y)  #calcula la media de y</pre>
```

[1] 25.04695

Incluir imágenes en el informe

Podemos incluir imágenes, o mapas o cualquier captura de pantalla que tenga formato JPG o PNG. Además, podemos controlar el ancho de la imagen, por ejemplo ajustando al 90% del tamaño de la misma (la altura se ajustará automáticamente).

Por ejemplo aquí el logo del curso:



Figure 9: Aquí el pie de figura



Figure 10: Comandos para insertar una imagen

Lección 4. Incluir gráficos en el informe.

Para incluir gráficos en el informe debemos introducir los comandos de gráficos que hemos visto en los Módulos 3 y 4 de este curso, o cualquier otro, entre la cabecera y pie de código de R.

Por ejemplo, generamos datos aleatorios y dibujamos una nube de puntos. Cada vez que se generen aleatoriamente datos la nube de puntos será distinta a menos que usemos el comando "set.seed()" para que sea una generación de datos aleatoria pero reproducible, como vimos en el Módulo 1 de este curso.

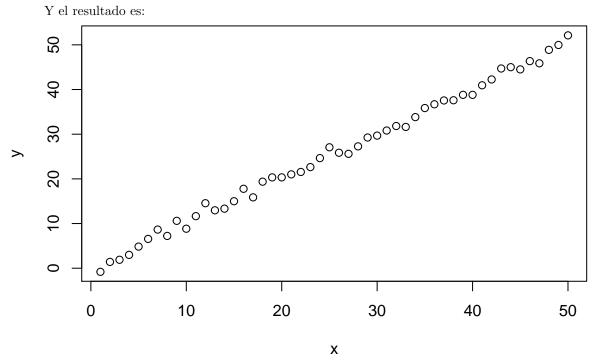
Este es el texto que iría en el genrador de informes:

```
"'{r datos aleatorios}
x <- 1:50 # asignamos a x los valores de 1 a 50
y <- x + rnorm(x) # rnorm() genera datos aleatorios con probabilidad normal
plot(x,y) #nube de puntos de x,y
```

```
`{r datos_aleatorios, echo=FALSE}
    y \leftarrow x + rnorm(x)
                             rnorm() genera datos aleatorios con
     probabilidad normal
224 plot(x,y)
                           #nube de puntos de x.v
225
226
```

Figure 11: Código ejecutable del Plot con RMarkdown

Y el resultado es:



Lección 5. Control total del informe: qué debe salir en el informe final.

Hemos visto como generar un informe automáticamente a través de RMarkdown y RStudio. Por supuesto, podemos controlar lo que aparece en el informe así como controlar que tipo de documento final vamos a

generar. Podemos por ejemplo generar un artículo cientifíco o una tesis de Máster o Doctorado totalmente con RMarkdown porque se puede incluir la gestión de citas y referencias (aunque excede los objetivos de este curso). También podremos generar todo tipo de informes únicos, o repetitivos (semanales, mensuales, anuales, etc.) que son frecuentes en estudios de seguimiento en experimentos científicos o en proyecto a lo largo del tiempo.

El control del informe final lo tendremos desde el generador del informe, podremos insertar código para generar gráficos (tan avanzados como los que hemos visto en el Módulo 3 y 4 de este curso), o hacer análisis, o ambos. Además podremos insertar imágenes. Por lo que tenemos control total del informe final. Este es el final de este curso, si has aprendido el contenido de cada módulo entonces habrás alcanzado el primer nivel de un científico de datos: conocer las herramientas para hacer Data Science. Buena suerte!