

Gráficos avanzados con R

CNE/ISCIII

Informes automatizados

Rmarkdown

- El paquete **rmarkdown** es una extensión de R que permite crear documentos automatizados que combinan texto con código, tablas y gráficos generados directamente por R.
- Rmarkdown permite generar documentos dinámicos al mezclar texto formateado y resultados generados por R. Los documentos generados pueden estar en HTML, PDF, Word y muchos otros formatos.

Las ventajas de esta herramienta son numerosas:

- El código y sus resultados no están separados de los comentarios asociados a ellos
- El documento final es reproducible
- El documento se puede actualizar fácilmente, por ejemplo, si los datos de origen se han modificado.

Por lo tanto, es una herramienta muy práctica para exportar, comunicar y difundir resultados estadísticos.

Estructura

- Markdown, un *formato* para escribir documentos en modo texto, fácil de leer y procesable para volcarlos a otros formatos
- La integración entre R y markdown

Aquí, un documento de R Markdown básico:

```
---  
title: "Ejemplo de R Markdown"  
output:  
  html_document:  
    df_print: paged  
---
```

R Markdown permite mezclar :

- texto libre puesto en formato
- bloques de código de R

Los bloques de código se pueden ejecutar para incluir sus resultados en el documento, así por ejemplo :

```
```${r}  
mean(airquality$Ozone, na.rm=TRUE)
```
```

Compilación

Al “compilar” el documento, el texto se formatea, los bloques de código se ejecutan, sus resultados se agregan al documento y todo se transforma en uno de los formatos disponibles (html, pdf, word).

Aquí, la representación del documento anterior en formato HTML

Ejemplo de R Markdown

R Markdown permite mezclar :

- texto libre puesto en formato
- bloques de código de R

Los bloques de código se pueden ejecutar para incluir sus resultados en el documento, así por ejemplo :

```
mean(airquality$Ozone, na.rm=TRUE)
```

```
## [1] 42.12931
```

Rmarkdown en 15 mn

Para aprender Markdown, se recomiendan los dos siguientes ejercicios :

Crear un fichero `.Rmd` usando `File > New File > R Markdown`.

Al crear un nuevo fichero de tipo **R Markdown**, RStudio proporciona, en lugar de uno vacío, una plantilla que muestra algunas de las opciones disponibles en este formato. Eso facilita el siguiente ejercicio:

Modificar el fichero de ejemplo creado en el ejercicio anterior añadiéndole títulos de varios niveles, párrafos de texto, cursivas, negritas, enlaces, listas (numeradas y sin numerar), etc. usando como guía el [Cheat Sheet](#) del paquete.

Compile el documento (p.e., pulsando el botón con la etiqueta *Knit HTML* situado encima del panel de edición de RStudio) para inspeccionar el resultado final.

Se puede también generar documentos en formato Word y PDF. Para estos formatos, es necesario tener instalados los programas : MS Word, LibreOffice o similar para el primero y LaTeX para el segundo.

Elementos de un documento Rmarkdown

Encabezado (préambulo)

La primera parte del documento es su *encabezado*. Se encuentra al principio del documento y está delimitado por tres guiones (---) antes y después:

```
---  
title: "Titulo"  
author: "Nombre Apellido"  
date: "2 de mayo de 2018"  
output: html_document  
---
```

Este encabezado contiene los metadatos del documento, como su título, autor, fecha, más una serie de opciones posibles que permiten configurar o personalizar todo el documento y su representación. Aquí, por ejemplo, la línea **output: html_document** indica que el documento generado tendrá un formato HTML.

Texto del documento

Listas

El cuerpo del documento consiste en texto con la sintaxis de Markdown: un marcado ligero que permite establecer niveles de títulos o formatear texto. Por ejemplo, el siguiente texto:

Este es un texto **en cursiva** y ****en negrita****.

Se puede definir una lista así:

- primer elemento
- segundo elemento

Que dará la siguiente salida

Este es un texto *en cursiva* y **en negrita**.

Se puede definir una lista así:

- primer elemento
- segundo elemento

Titulos y vinculos

Los títulos de diferentes niveles se pueden definir comenzando una línea con uno o más caracteres #:

Título de nivel 1

Título de nivel 2

Título de nivel 3

Cuando se han definido los títulos, al hacer clic en el icono *Outline* en el extremo derecho de la barra de herramientas asociada al archivo, se muestra una tabla dinámica de contenidos que permite navegar fácilmente en el documento.

La sintaxis de Markdown permite también insertar enlaces o imágenes. Por ejemplo, la siguiente sintaxis:

```
[ISCI II](http://www.isciii.es/)
```

Dará el siguiente vinculo:

[ISCI II](http://www.isciii.es/)

Bloques de código

Además del texto libre en formato Markdown, un documento R Markdown contiene, como su nombre indica, código R. Este código se incluye en fragmentos definidos por la siguiente sintaxis:

Como esta cadena de caracteres no es muy fácil de escribir, se puede usar *R* en el menú Insertar de RStudio, o teclear el atajo **Ctrl+Alt+i**.

Se puede dar un nombre al bloque y se indica directamente después de `r`:

```
{r nombre_del_bloque}
```

No es obligatorio, pero puede ser útil en caso de error de compilación, para identificar el bloque que causó el problema. Atención, no podemos tener dos bloques con el mismo nombre.

Opción de salida del código echo

Además de un nombre, se puede pasar a un bloque una serie de opciones para modificar su comportamiento.

```
```{r echo = FALSE, warning = FALSE}  
x <- 1:5
```
```

Una de las opciones más útiles es la opción **echo**. Por defecto **echo=TRUE**, y el bloque de código R se inserta en el documento generado:

```
x <- 1:5  
print(x)
```

```
## [1] 1 2 3 4 5
```

Pero, si la opción **echo=FALSE**, entonces el código R ya no se inserta en el documento, y solo se el resultado será visible:

```
## [1] 1 2 3 4 5
```

Otras opciones

Aquí, una lista de algunas de las opciones más comunes:

| opción | valores | descripción |
|---------|--------------|---|
| echo | TRUE / FALSE | Mostrar o no el código R en el documento |
| eval | TRUE / FALSE | Ejecutar o no el código R en tiempo de compilación |
| warning | TRUE / FALSE | Mostrar o no las advertencias generadas por el bloque |
| message | TRUE / FALSE | Mostrar o no los mensajes generados por el bloque |

Hay muchas otras opciones descritas en la [Guía de referencia de R Markdown](#).

Tablas con Rmarkdown

Tablas cruzadas

Las tablas generadas por la función `table` se muestran tal y como aparecen en la consola de R:

```
titanic<-apply(Titanic,c(1,4),sum) #Supervivencia al Titanic según clase
titanic
```

```
##           Survived
## Class    No Yes
##  1st    122 203
##  2nd    167 118
##  3rd    528 178
##  Crew    673 212
```

Tablas cruzadas con kable

Su presentación se puede mejorar utilizando la función `kbl` de la extensión `kableExtra`:

```
require(kableExtra)
kbl(titanic,caption="Supervivencia a la catastrofe del Titanic según la clase") %>%
  kable_classic("striped")
```

Supervivencia a la catastrofe del Titanic según la clase

| | No | Yes |
|------|-----|-----|
| 1st | 122 | 203 |
| 2nd | 167 | 118 |
| 3rd | 528 | 178 |
| Crew | 673 | 212 |

Base de datos

La representación HTML de bases de datos (**data.frame**) por defecto es el contenido que aparece en consola. Este formato puede ser poco adecuado si la tabla excede una cierta dimensión. Una alternativa es usar la función **datatable** del paquete **DT**, que ofrece además interactividad:

```
require(DT)
datatable(gapminder, options=list(pageLength=6))
```

Show entries

Search:

| | country | continent | year | lifeExp | pop | gdpPercap |
|---|-------------|-----------|------|---------|----------|-------------|
| 1 | Afghanistan | Asia | 1952 | 28.801 | 8425333 | 779.4453145 |
| 2 | Afghanistan | Asia | 1957 | 30.332 | 9240934 | 820.8530296 |
| 3 | Afghanistan | Asia | 1962 | 31.997 | 10267083 | 853.10071 |
| 4 | Afghanistan | Asia | 1967 | 34.02 | 11537966 | 836.1971382 |
| 5 | Afghanistan | Asia | 1972 | 36.088 | 13079460 | 739.9811058 |
| 6 | Afghanistan | Asia | 1977 | 38.438 | 14880372 | 786.11336 |

Showing 1 to 6 of 1,704 entries

Previous 2 3 4 5 ... 284 Next

Resumen de base de datos

El paquete **gtsummary** permite generar automáticamente resúmenes de base de datos:

```
require(gtsummary)
temp <- subset(nhs,select=c(edad:raza,imc:pas,dth))
tbl_summary(temp, by=sexo)
```

| Characteristic | Hombre, N = 4,349 ¹ | Mujer, N = 4,901 ¹ |
|----------------------------------|--------------------------------|-------------------------------|
| edad | 58 (42, 65) | 58 (42, 66) |
| raza | | |
| Blanco | 3,815 (88%) | 4,274 (87%) |
| Negro | 449 (10%) | 549 (11%) |
| Otro | 85 (2.0%) | 78 (1.6%) |
| imc | 25.6 (23.2, 28.2) | 25.2 (22.2, 29.1) |
| pulso | 78 (68, 84) | 80 (72, 88) |
| Unknown | 16 | 28 |
| pas | 132 (121, 147) | 130 (116, 149) |
| Unknown | 6 | 15 |
| dth | 1,258 (29%) | 887 (18%) |
| ¹ Median (IQR); n (%) | | |

Tabla de resultados de análisis de regresión

Este mismo paquete permite también elaborar de manera eficiente tablas de resultados de regresión:

```
fit = lm(lifeExp ~ continent, data=gapminder)
T1 = tbl_regression(fit)
T1
```

| Characteristic | Beta | 95% CI ¹ | p-value |
|---------------------------------------|------|---------------------|---------|
| continent | | | |
| Africa | — | — | |
| Americas | 16 | 15, 17 | <0.001 |
| Asia | 11 | 10, 12 | <0.001 |
| Europe | 23 | 22, 24 | <0.001 |
| Oceania | 25 | 22, 29 | <0.001 |
| ¹ CI = Confidence Interval | | | |

Tabla de resultados de análisis de regresión

Este mismo paquete permite también elaborar de manera eficiente tablas de resultados de regresión:

```
fit2 = lm(lifeExp ~ log(gdpPercap,2) + continent, data=gapminder)
T2 = tbl_regression(fit2)
tbl_merge(list(T1,T2),c("Efectos crudos","Efectos ajustados"))
```

| Characteristic | Efectos crudos | | | Efectos ajustados | | |
|---------------------------------------|----------------|---------------------|---------|-------------------|---------------------|---------|
| | Beta | 95% CI ¹ | p-value | Beta | 95% CI ¹ | p-value |
| continent | | | | | | |
| Africa | — | — | | — | — | |
| Americas | 16 | 15, 17 | <0.001 | 7.0 | 5.9, 8.1 | <0.001 |
| Asia | 11 | 10, 12 | <0.001 | 5.9 | 5.0, 6.8 | <0.001 |
| Europe | 23 | 22, 24 | <0.001 | 9.6 | 8.4, 11 | <0.001 |
| Oceania | 25 | 22, 29 | <0.001 | 9.2 | 6.2, 12 | <0.001 |
| log(gdpPercap, 2) | | | | 4.5 | 4.2, 4.7 | <0.001 |
| ¹ CI = Confidence Interval | | | | | | |

Referencias

Este curso está basado en las siguientes referencias:

- *R para profesionales de los datos*, Carlos Bellosta, 2017, https://datanalytics.com/libro_r
- El libro [R for data science](#) disponible en línea, que contiene un capítulo dedicado a R Markdown .

Y también muy inspirado de estos otros libros:

- *Data Visualisation with R*, Thomas Rahlf, 2014, <http://www.datavisualisation-r.com/>
- *R Graph Cookbook* , Hrish Mittal, 2011