

Ejemplo de uso de RMarkdown para construir informes en formato PDF

Curso Introducción a R para Ciencias Sociales - Abril 2020

Felipe Ruiz Bruzzzone

22 julio, 2020

Contents

1	Presentación	2
2	Detalle de los elementos de RMarkdown	3
3	Construcción de resultados	4
3.1	Tablas de información	4
3.2	Tablas de frecuencias	5
3.3	Estadísticos descriptivos	6
3.4	Regresión lineal simple	8
3.5	Gráfico	9
4	¿Cómo incorporar bibliografía?	10
5	Enlaces útiles	11
	Referencias bibliográficas	12

1 Presentación

Este documento es un ejemplo de exportación de resultados utilizando RMarkdown, para el curso introductorio al uso de RStudio para Ciencias Sociales efectuado durante **diciembre de 2019** en Estación Lastarria.¹

El objetivo del documento es lograr incluir en este ejemplo ocho elementos, con formato adecuado a reportes académicos y/o profesionales:

1. Encabezado con información relevante junto con un índice al inicio.
2. Texto con diversas opciones de formato.
3. Tabla con información construida manualmente.
4. Tabla de frecuencias simples.
5. Tablas de estadísticos descriptivos construidas manualmente y mediante función ad hoc.
6. Resultados principales de un modelo de regresión lineal simple.
7. Un gráfico de dispersión.
8. Bibliografía.

Esta plantilla de RMarkdown permite articular todos los elementos anteriores para producir un reporte. Recuerden que *nunca es necesario aprenderse de memoria los comandos*. Basta construir una planilla de este estilo e ir modificando lo que se desee, según los análisis requeridos.

Para ejecutarlo, necesitarán tener instaladas algunas cuestiones adicionales. Existen diversas fuentes en línea en las cuales pueden encontrarse orientaciones generales respecto al funcionamiento y uso de RMarkdown que incluimos a continuación como referencias (Grolemund, 2014; Miller, 2016, 2018; Workshop, 2016), podrán encontrar el final de este documento acceso a ellas mediante enlaces url.

Para contar con orientaciones detalladas respecto a la instalación, funcionamiento y uso de RMarkdown en español pueden revisar el [tutorial en línea](#) de Giorgio Boccardo y Felipe Ruiz (2019) (en español) disponible para descarga [en el siguiente enlace](#).

También es altamente recomendable un libro (en inglés) publicado recientemente por los desarrolladores de las diferentes herramientas que componen RMarkdown titulado *RMarkdown: The Definitive Guide* (Yihui, Allaire, & Grolemund, 2018). En el enlace asociado a la referencia se encuentra una versión gratuita en línea de tal material.

¹Si lo ejecutan tal como está, pueden exportarlo a Html de manera directa. Si quieren exportarlo a PDF deberán editar el comando *stargazer* del apartado de regresión.

2 Detalle de los elementos de RMarkdown

Una sintaxis de RMarkdown se estructura en torno a tres elementos.

En primer lugar tenemos el encabezado. Aquí se configuran las *propiedades* del reporte final. Se pueden indicar distintos elementos: título del documento, subtítulo, autor, fecha, bibliografía y su formato, si se numerarán los títulos, si se integrará un índice, así como el estilo de las notas y enlaces a la web.

En segundo lugar tenemos el editor de texto, que es el formato base de esta sintaxis. Aquí lo escrito por defecto no es un código de R. Debemos pensar que éste es nuestro procesador de texto (**y no Word!**). Desde aquí no sólo configuraremos las “salidas de resultados”, sino que también escribiremos el contenido.

En tercer lugar, tenemos los *trozos de código de R* o *code chunks*. Se trata de bloques de código delimitados por la siguiente estructura. Como puede observarse en la sintaxis de RMarkdown, el inicio de un código está delimitado por tres apóstrofes seguidos por un `r` entre corchetes curvos `{r}`, y su cierre por otros tres apóstrofes. Eso delimita lo que se ejecutará como código de computación, diferenciándolo respecto al texto simple.

Como se observa a continuación, también puede aplicarse una sangría mayor para un párrafo en específico.

¿Cuáles son, entonces, las características de RMarkdown que lo diferencian de esta forma de trabajo? En términos sencillos RMarkdown es un procesador de texto que ofrece además la posibilidad de incluir trozos de código desde R (u otros formatos). El principal beneficio de esta herramienta es que permite trabajar en un sólo documento tanto la redacción del contenido narrativo de reportes de investigación, como también la construcción y presentación formal de resultados de análisis estadísticos. (Boccardo & Ruiz Bruzzzone, 2019)

El primer trozo de código se muestra en este reporte y se ejecuta al mismo tiempo. Refiere a la carga de datos y algunos paquetes. Similar a una sintaxis simple de R, cada trozo de código contempla lo que se hizo antes en la sesión de RMarkdown. En cualquier caso, es útil cargar datos y paquetes a usar en cada trozo de código, pues al ejecutarse desde el principio todos juntos, a veces se producen errores.²

```
# Se configura el código para que se imprima pero sin mostrar ningún mensaje.  
library(knitr)  
library(summarytools)  
library(stargazer)  
load("datos.RData")
```

² Así se incorporan notas al pie.

3 Construcción de resultados

3.1 Tablas de información

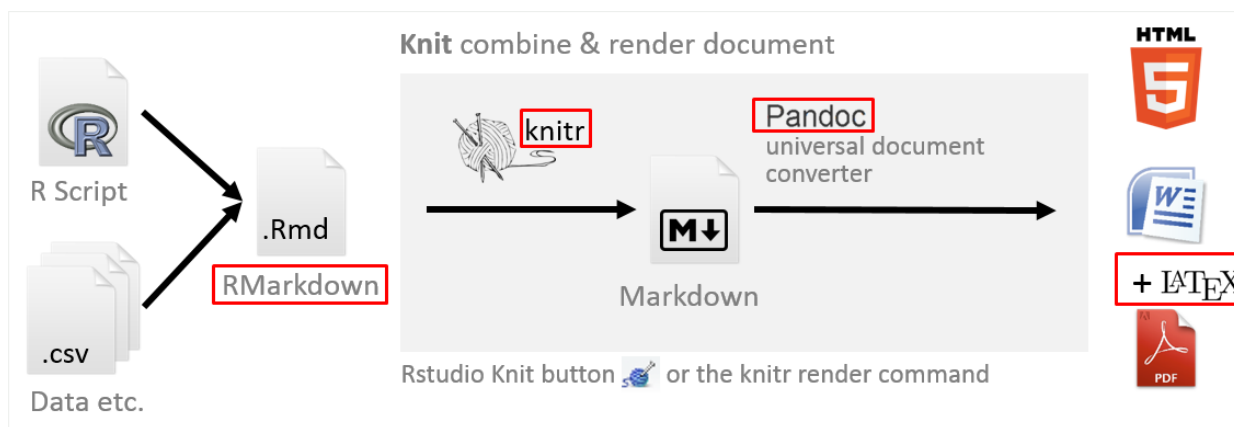
Una de las cuestiones básicas a considerar en el uso de RMarkdown es comprender qué elementos se ponen en juego en su funcionamiento. En la tabla 1 se muestran los elementos necesarios para el funcionamiento de RMarkdown y en la imagen 1, se muestra un diagrama que indica su articulación en el flujo de trabajo.

Al mismo tiempo, se está mostrando cómo insertar tablas con contenido construidas manualmente y también como insertar imágenes.

Table 1: **Elementos necesarios para ejecución de RMarkdown: procedimientos de instalación**

Elemento	Procedimiento de instalación	Código para instalación
RMarkdown	No es necesario ningún procedimiento adicional a la instalación de R y RStudio pues viene instalado con este último.	<i>No aplica</i>
Knitr	Debe instalarse como cualquier otro paquete de R, asegurando su disponibilidad para ser utilizado por RMarkdown al compilar documentos.	<code>install.packages("knitr")</code>
Pandoc	No es necesario ningún procedimiento adicional a la instalación de R y RStudio pues viene instalado con este último.	<i>No aplica</i>
LaTeX	Debe descargarse como un paquete de R. Para asegurar su disponibilidad para ser utilizado por RMarkdown al compilar documentos, debe instalarse con un comando adicional.	<code>install.packages("tinytex")</code> <code>tinytex::install_tinytex()</code>

Imagen 1: Elementos necesarios para construir reportes con RMarkdown



3.2 Tablas de frecuencias

Para compilar tablas de frecuencias, se utiliza la función **freq** del paquete *summarytools*. Esta función permite indicar qué tipo de compilación se busca (*style*), la alineación del texto en tabla (*justify*), si se desea o no compilar los encabezados del resultado (*headings = ...*) y si se desea incluir en la tabla el conteo diferenciado de casos considerando valores perdidos (*report.nas = ...*).

```
library(summarytools)
freq(datos$sexo, style = "simple", justify = "center", headings = TRUE,
      report.nas = FALSE)
```

```
## Frequencies
## datos$sexo
## Type: Factor
##
##           Freq      %      % Cum.
## -----
##   Hombre    553   38.83   38.83
##   Mujer     871   61.17  100.00
##   Total    1424  100.00  100.00
```

Si se cambia el argumento *style* a la opción *'rmarkdown'*, y se indica la opción *results = 'asis'* para el código de R, se obtendrá una tabla compilada en un formato procesable por RMarkdown.

```
library(summarytools)
freq(datos$sexo, style = "rmarkdown", justify = "center", headings = FALSE,
      report.nas = FALSE)
```

	Freq	%	% Cum.
Hombre	553	38.83	38.83
Mujer	871	61.17	100.00
Total	1424	100.00	100.00

3.3 Estadísticos descriptivos

3.3.1 Tabla manual

En este ejemplo se observa la construcción de una tabla de resultados elaborada de forma manual. Se calcula cada valor y se compila como vector (para dos variables). Luego se combinan tales vectores para construir una matriz de datos con dos filas. Luego de indicar que cada valor numerico se redondee a dos valores, se crea una variable con dos valores, que se unirá como columna, para indicar una etiqueta para cada variable.

```
n1<-length(datos$edad)
min1<- min(datos$edad)
med1<- median(datos$edad)
prom1<- mean(datos$edad)
de1 <- sd(datos$edad)
max1 <- max(datos$edad)
edad<- cbind(n1,min1,med1,prom1,de1,max1)

n2 <- length(datos$satisfaccion_vida)
min2<-min(datos$satisfaccion_vida, na.rm = T)
med2 <- median(datos$satisfaccion_vida, na.rm = T)
prom2 <- mean(datos$satisfaccion_vida,na.rm = T)
de2 <- sd(datos$satisfaccion_vida, na.rm = T)
max2 <- max(datos$satisfaccion_vida, na.rm = T)
satisfaccion<- cbind(n2,min2,med2,prom2,de2,max2)

descriptivos <- rbind(edad, satisfaccion)

descriptivos <- round(descriptivos, digits = 1)

variables <- c("Edad", "Satisfacción con la vida")

tabla_1 <- cbind(variables,descriptivos)
```

De esto resulta una matriz de datos que se observa como sigue.

```
##      variables                n1      min1 med1 prom1 de1      max1
## [1,] "Edad"                  "1424" "18" "50" "49.9" "17.8" "97"
## [2,] "Satisfacción con la vida" "1424" "1"  "7"  "7.3"  "2.1"  "10"
```

Posteriormente, utilizando la función *kable* del paquete **knitr** se imprimirá esta matriz de datos en formato *tabla*, indicando un título, la alineación de los valores y los nombres de las columnas.

```
library(knitr)
kable(tabla_1, caption = "Estadísticos descriptivos de dos variables", align = 'c', digits = 2,
      col.names = c("Variables", "N", "Mínimo", "Mediana", "Media", "Desviación Estándar", "Máximo"))
```

Table 3: Estadísticos descriptivos de dos variables

Variables	N	Mínimo	Mediana	Media	Desviación Estándar	Máximo
Edad	1424	18	50	49.9	17.8	97
Satisfacción con la vida	1424	1	7	7.3	2.1	10

3.3.2 Tabla usando función

Seguiremos usando una función del paquete *summarytools*, en este caso la función **descr**. Esta función compila inicialmente una tabla vertical, para ajustarla a formato horizontal se usa especifica el argumento *transpose* como **TRUE**. Mediante el argumento *stats* se indican los estadísticos a calcular. El argumento *style* permite indicar el formato en que se busca compilar el resultado, en este caso, se indica el formato por defecto para la consola de R. Al igual que la función *freq* ya vista para tablas de frecuencias, los siguientes argumentos permiten indicar la alineación del texto en tabla (*justify*) y si se desea o no compilar los encabezados del resultado (*headings* = ...).

```
descr(datos$edad, transpose = TRUE,
      stats = c("N.Valid", "min", "q1", "med", "mean", "sd", "q3", "max", "iqr"),
      style = "simple", justify = "c", headings = TRUE)

## Warning: 'funs()' is deprecated as of dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with 'tibble::lst()':
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_warnings()' to see where this warning was generated.
```

Descriptive Statistics
datos\$edad
N: 1424

	N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
edad	1424.00	18.00	36.00	50.00	49.87	17.79	64.00	97.00	28.00

Ahora, si se indica el formato *rmarkdown* y se utiliza la función *kable* del paquete **knitr** es factible imprimir la tabla adecuada a formato RMarkdown, pero editando el título de la tabla y los encabezados de las columnas.

En el archivo *.Rmd* se observa que se indica la el argumento **results** = 'asis' y **echo** = **FALSE** como especificaciones del trozo de código; esto permite que ni **knitr** ni **pandoc** compilen el formato del resultado, respetando la configuración establecida con *kable*, a la vez que se indica ejecutar el código y compilar su resultado, pero no mostrarlo en el reporte final.

Table 5: Estadísticos descriptivos usando función

	N válidos	Mín.	Q1	Mediana	Media	DE	Q3	Máx.	RIC.
edad	1424	18	36	50	49.87	17.79	64	97	28

3.4 Regresión lineal simple

A continuación se muestra como exportar el resultado de un modelo de regresión especificado mediante la siguiente sintaxis: `lm(ozone ~ temp, data = medias)`. El resultado se guarda en un objeto llamado *regresion*

Luego, si sobre tal objeto ejecutamos el comando `stargazer` del paquete del mismo nombre, se obtiene un resumen en formato tabla, de los estadísticos descriptivos principales del modelo.

Cabe observar que para obtener archivos compilados desde RMarkdown, es vital el argumento `type` de esta función. Para documentos en formato Html, sirve mantener la configuración `html`; sin embargo, si buscamos compilar un archivo PDF, deberemos cambiar el comando `type` a la opción `type = "latex"`. Si buscamos compilar un archivo word, deberemos indicar la opción `type = "text"`

```
library(stargazer)
stargazer(regresion, header = FALSE, type = "latex")
```

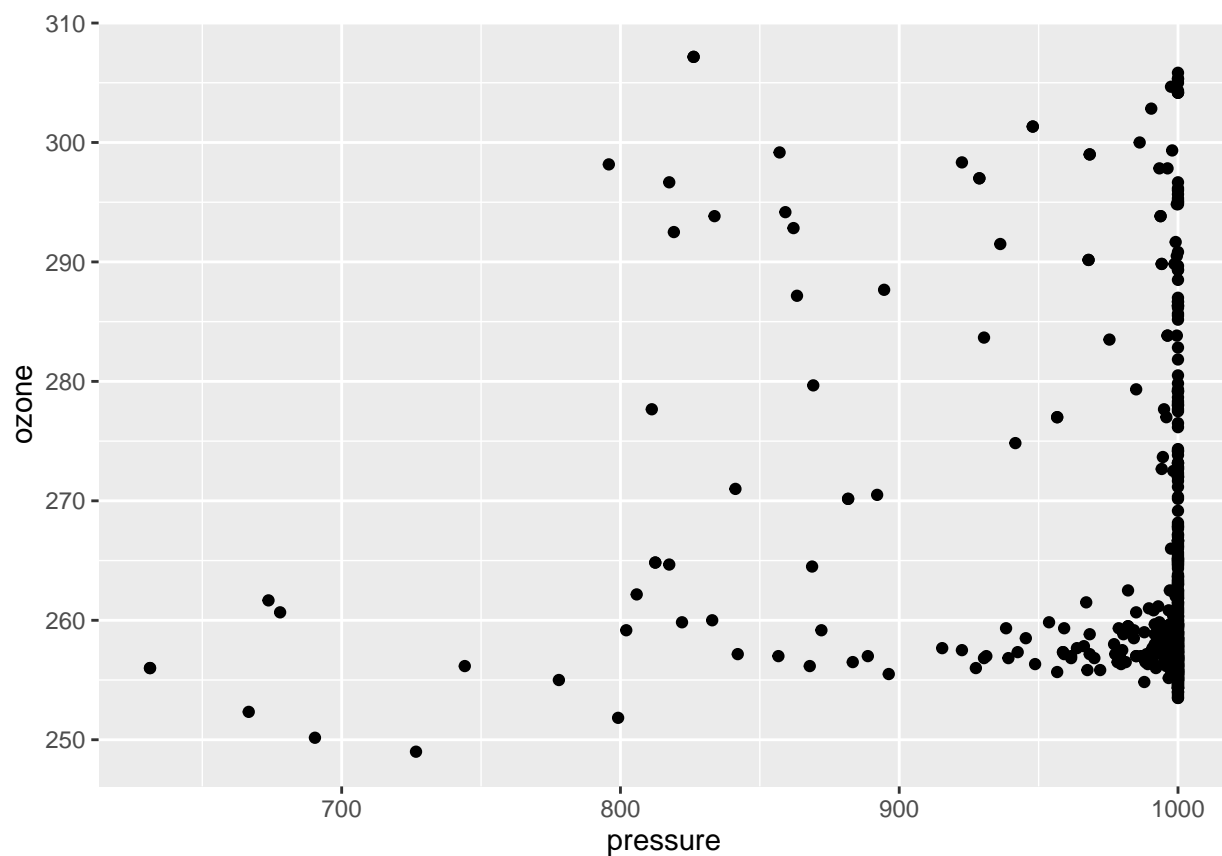
Table 6:

	<i>Dependent variable:</i>
	satisfaccion_vida
edad	−0.012*** (0.003)
Constant	7.887*** (0.166)
Observations	1,414
R ²	0.009
Adjusted R ²	0.009
Residual Std. Error	2.102 (df = 1412)
F Statistic	13.504*** (df = 1; 1412)
Note:	*p<0.1; **p<0.05; ***p<0.01

3.5 Gráfico

La misma lógica se sigue para los gráficos. A continuación se muestra el código que genera un gráfico. Es importante recordar que **ggplot** es la función, mientras que **aes** permite especificar la relación entre los datos y los ejes o dimensiones del gráfico. Luego del signo '+' la siguiente línea especifica la geometría, en este caso **geom_point** indica que se trata de un diagrama de dispersión (la geometría indicada son puntos).

```
ggplot(medias, aes(x = pressure, y = ozone)) + geom_point()
```



4 ¿Cómo incorporar bibliografía?

El último elemento a revisar es cómo incorporar una bibliografía formal al reporte.

Para hacerlo es preciso manejar dos elementos.

1. *Zotero y Zotero connector*. Mediante este software se puede crear un listado bibliográfico compatible con LaTeX (archivo *.bib*) que - cargado a RMarkdown - permitirá insertar referencias y que luego se construya un listado bibliográfico. Ambos pueden descargarse desde [este enlace](#).
2. Contar con un archivo *.csl* para indicarle el formato de cita al documento (*Citation Language Style*). Se puede descargar el adecuado para formato APA 6a Edición desde [esta página](#).

El listado bibliográfico final se contruirá siempre en base a las citas realizadas en el cuerpo del texto y no en base a todas las referencias presentes en el archivo de referencias exportado desde zotero. La bibliografía siempre se compila al final del documento. Eso significa que estará al final del último elemento escrito. Por eso, si se deja un título al final del documento, permite anteponer un encabezado al apartado de bibliografía.

5 Enlaces útiles

Links para descargar Zotero y Zotero connector

- <https://www.zotero.org/download/>

Link para descargar diferentes formatos para referencias bibliográficas.

- <https://www.zotero.org/styles?q=apa>

Referencias bibliográficas

- Boccardo, G., & Ruiz Bruzzone, F. (2019). *RStudio para Estadística Descriptiva en Ciencias Sociales. Manual de apoyo docente para la asignatura Estadística Descriptiva* (Segunda edición). Facultad de Ciencias Sociales: Departamento de Sociología, Universidad de Chile. Retrieved from <https://bookdown.org/gboccardo/manual-ED-UCH/>
- Grolemund, G. (2014). Introduction to R Markdown. Retrieved from https://rmarkdown.rstudio.com/articles_intro.html
- Miller, S. V. (2016). An R Markdown Template for Academic Manuscripts. *Steven V. Miller*. Retrieved from <http://svmiller.com/blog/2016/02/svm-r-markdown-manuscript/>
- Miller, S. V. (2018, June). A Pandoc Markdown Article Starter and Template. Retrieved from <https://github.com/svmiller/svm-r-markdown-templates>
- Workshop, R. R. (2016). Writing publications with R. *Writing publications with R*. Retrieved from http://www.geo.uzh.ch/microsite/reproducible_research/post/rr-r-publication/
- Yihui, X., Allaire, J., & Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman & Hall/CRC. Retrieved from <https://bookdown.org/yihui/rmarkdown/>