

---

# Cómo hacer un reporte estadístico con Sweave: L<sup>A</sup>T<sub>E</sub>X y R

How to make statistical reports with Sweave. L<sup>A</sup>T<sub>E</sub>X y R

Carlos Lesmes<sup>a</sup>  
lesmes@member.ams.org

---

## Resumen

Este documento muestra un ejemplo del uso de **Sweave** para manejar la entrada y salida de la consola de R en un documento L<sup>A</sup>T<sub>E</sub>X. Este es un formato útil para hacer tutoriales de R, análisis estadísticos, proyectos con R y reproducir investigaciones.

**Palabras clave:** ambiente, integrado, tejido, R, código, trozo.

## Abstract

This document provides an example of using **Sweave** to handle input and output in R console in a L<sup>A</sup>T<sub>E</sub>X document. This is a useful format for R tutorials, statistical analyses, projects with R and reproducing researches.

**Key words:** environment, integrated tissue, R, code, chunk.

## 1. Introducción

Este artículo es un ejemplo de un documento creado usando **Sweave**. **Sweave** es una herramienta para combinar L<sup>A</sup>T<sub>E</sub>X y código R en el mismo archivo. Se asume que el lector tiene cierta familiaridad con L<sup>A</sup>T<sub>E</sub>X, R y RStudio.

Se ha procesado por el sistema **Sweave** en R, que corre el código R para generar una salida de texto y gráficos, y también crea un documento L<sup>A</sup>T<sub>E</sub>X. En realidad **Sweave** es una función de R que permite integrar código R en un documento L<sup>A</sup>T<sub>E</sub>X con el propósito de crear documentos dinámicos, que se puedan actualizar automáticamente si hay algún cambio en los datos. Para más detalles de **Sweave** ver Lenth (2012) y Leisch (2002). Lo anterior se puede realizar en el ambiente RStudio. Para esto se debe descargar e instalar RStudio de <http://www.rstudio.com> y tener previamente instaladas versiones recientes de R (Team 2012) y L<sup>A</sup>T<sub>E</sub>X. RStudio

---

<sup>a</sup>Docente tiempo completo, Facultad de Estadística, Universidad Santo Tomás.

es un *Integrated Development Environment (IDE)* : un ambiente integrado para desarrolladores. Este es un ambiente para programar, integrado al programa R, el cual le provee una interfaz gráfica, un editor de texto o código, un interpretador y un depurador. Este ambiente está desarrollado para usar y programar con R, entre otras aplicaciones. Para más información acerca de Rstudio consultar Verzani (2011).

## 2. RStudio

Los principales componentes de RStudio son cuatro paneles que incluyen la consola interactiva de R, un editor de R que permite administrar los scripts generados, un navegador de variables e historia de trabajo y un cuarto panel con las ayudas y acceso al manejo de paquetes de R, un manejador de gráficas y archivos. El editor de R soporta los comandos de R, archivos de texto, documentos **Sweave**, documentos .tex y documentación de R.

## 3. ¿Cómo se hace?

```

\documentclass{minimal}
\usepackage[...]{...}
\title{...}
\author{...}
\begin{document}
\maketitle
.
.
(segmento latex)
.
.
<<(opciones...)>>=
(código R)
@ .
.
\end{document}

```

Figura 1: Estructura mínima para un documento Sweave. Fuente: elaboración propia.

Escriba un archivo  $\text{\LaTeX}$  con extensión .Rnw en vez de la extensión .tex. El archivo debe tener segmentos de código R alternado por segmentos de  $\text{\LaTeX}$ . Deberá tener la forma que se ve en la figura 1. Un trozo de código R se llama *chunk*. Un *chunk* es código R entre los delimitadores <<>>= y @ dentro de un documento  $\text{\LaTeX}$ .

Simplemente se escoge la opción **Sweave Document** en la esquina inferior derecha del editor de R y se escribe el documento. El documento quedará con la extensión `.Rnw` requerida. Por último se obtiene el archivo  $\text{\LaTeX}$  y el PDF correspondiente con **Compile PDF** en la parte superior del editor (vea Figura 2). El archivo  $\text{\LaTeX}$ , (con extensión `.tex`) que se crea no se edita y en realidad solo lo utiliza RStudio para crear el PDF. Los *chunks* admiten diferentes opciones, de acuerdo con las necesidades del autor. A continuación se presentan algunas de ellas.

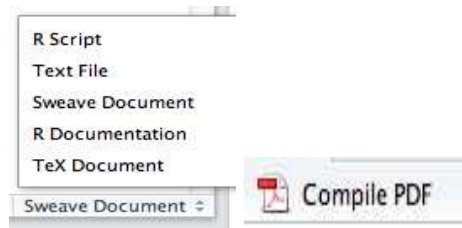


Figura 2: Opciones del editor de RStudio.

## 4. Algunos ejemplos de *chunks*

El *chunk* a continuación, con nombre uno.

```
<<uno>>=
x <- 2
y <- 3
x+y
@
```

produce

```
> x<-2
> y<-3
> x+y
[1] 5
```

Si agregamos el *chunk* dos

```
<<label=dos>>=
x*y
@
```

se obtiene

```
> x*y
[1] 6
```

Los chunks con nombre se pueden reutilizar, como en el siguiente script:

```
<<tres>>=
<<uno>>
<<dos>>
@
```

y se tiene:

```
> x<-2
> y<-3
> x+y
[1] 5
> x*y
[1] 6
```

En el siguiente *chunk* nótese que la única opción es `label=cuatro`, que es el nombre del chunk. La palabra `label` puede ser omitida, el nombre debe ir en la primera posición, de lo contrario se produce un error. Un *chunk* puede no tener nombre.

```
<<label=cuatro>>=
datos=rexp(100)
summary(datos)
@
```

La salida correspondiente es:

```
> datos=rexp(100)
> summary(datos)

    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.03336 0.31170 0.71440 1.03500 1.50700 4.55900
```

El mismo *chunk* con la opción `echo=FALSE`: *no* muestra los comandos de la consola R, únicamente los resultados. En el *chunk* `cuatro` el defecto es `echo=TRUE`.

```
<<cinco, echo=F>>=
datos=rexp(100)
summary(datos)
@
```

con este *chunk* se obtiene:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.006007	0.348000	0.820300	1.016000	1.559000	5.352000

En el siguiente *chunk* los resultados de R salen con formato L<sup>A</sup>T<sub>E</sub>X:

```
<<label=seis, echo=F,results=teX>>=
datos=rexp(10)
datos
@

> datos=rexp(10)
> datos
[1] 3.5420565 0.9458364 0.3135035 0.9355242 1.4517435 1.9733489 0.4764576
[8] 0.2863853 0.8825239 1.9308057
```

Ahora, un *chunk* con `echo=F` y `results=hide` no muestra el código y tampoco los resultados.

```
<<label=siete, echo=F,results=hide>>=
datos=rexp(100)
summary(datos)
@
```

## 5. Resultados en el texto con Sexpr.

Se pueden insertar resultados de los comandos de R en el texto con el comando `\ Sexpr{}`. La media de `datos` se escribe en L<sup>A</sup>T<sub>E</sub>X:

```
\ Sexpr{mean(datos)}
```

y por lo tanto se puede escribir:

La media de `datos` es 1.2246 y la varianza 1.6595.

## 6. Mejorando las opciones de un *chunk*

Un *chunk* para ver la salida de R en verbatim:

```
<<label=ocho, echo=T,results=verbatim>>=
datos[1:10]
@
```

y la salida es:

```
> datos[1:10]
[1] 2.2713786 0.4022673 0.8288273 0.3506155 0.2636374 1.6728966 2.6034021
[8] 1.5087466 0.3309880 0.4270607
```

Se pueden modificar las opciones de R usando el comando `options()` dentro del *chunk*, en este caso con `width` y `prompt`, como en el siguiente *chunk*:

```
<<label=nueve, echo=F,results=verbatim>>=
options(width=60,prompt="SwR> ",continue="...")
datos[1:10]
@
```

obteniendo:

```
> options(width=60,prompt="SwR> ",continue="...")
SwR> datos[1:10]

[1] 2.2713786 0.4022673 0.8288273 0.3506155 0.2636374
[6] 1.6728966 2.6034021 1.5087466 0.3309880 0.4270607
```

## 7. Poniendo color

Hay comandos para ponerle color al código y a los resultados en `verbatim`, no olvide utilizar el paquete `fancyvrb`:

```
\DefineVerbatimEnvironment{Sinput}{Verbatim}{formatcom=%
{\color{rgb}{0.8,0,0}},xleftmargin=2em}
\DefineVerbatimEnvironment{Soutput}{Verbatim}{formatcom=%
{\color{rgb}{0,0,0.8}},xleftmargin=2em}
\DefineVerbatimEnvironment{Scode}{Verbatim}{formatcom=%
{\color{rgb}{0,0.8,0}},xleftmargin=2em}
```

Para el siguiente *chunk*:

```
<<label=diez, echo=F,results=verbatim>>=
summary(datos)
@
```

Se tiene:

```
SwR> summary(datos)
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.01452 0.41630 0.82620 1.22500 1.53300 7.88500
```

## 8. Código para figuras

También se pueden insertar gráficas de R integradas con `caption` y `label` para hacer referencia (vea Figura 3). Note que debe estar la opción `fig=TRUE` y que se puede determinar el tamaño de la gráfica con `width` y `height`. Se pueden personalizar las gráficas con el comando

```
\SweaveOpts{width=6,height=4}
```

en el preámbulo. Las gráficas resultantes son del tamaño de 4 por 6 pulgadas. Las opciones de este comando aplican para todos los *chunks* del documento.

El siguiente código realiza la figura 3.

```
\begin{figure}
<<once, echo=F,fig=T,width=10,height=4>>=
par(mfrow=1,2)
hist(rexp(100),main="Hist 1",ylab="Frecuencia")
hist(rexp(1000),main="Hist 2",ylab="Frecuencia")
@
\caption{Histogramas}
\label{F1}
\end{figure}
```

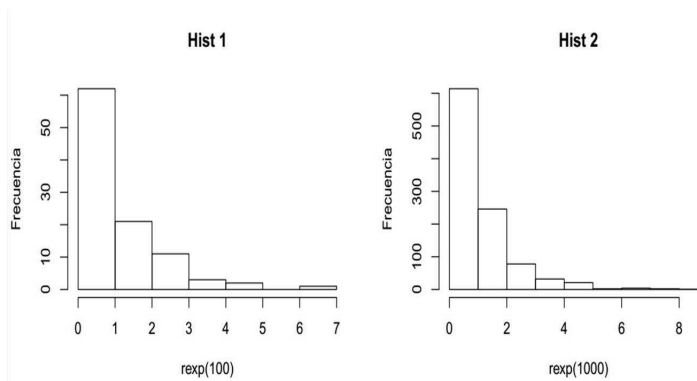


Figura 3: Histogramas. Fuente: elaboración propia.

Recuerde que todos los *chunks* hacen parte de la misma sesión de R. Un código para hacer un histograma con `datos` (Figura 3), generado anteriormente, es el siguiente:

```
\begin{figure}
<<doce, echo=F,fig=T>>=
hist(datos,main="Otro",ylab="Frecuencia",col='gray90')
@
\caption{Datos}
\label{F2}
\end{figure}
```

El mismo *chunk* cambiando el tamaño de la figura (vea Figura 5) usando el comando `\setkeys{Gin}{width=0.5\textwidth}`.

## 9. Código para tablas

Un *chunk* para generar una tabla (vea tabla 1). Se utiliza el comando `xtable` y la opción `results=tex`.

Más opciones para los *chunks* se pueden obtener solicitando ayuda para el comando `RweaveLatex`: `??RweaveLatex` en la consola de R o también se puede consultar en la página [http://yihui.name/knitr/options#chunk\\_options](http://yihui.name/knitr/options#chunk_options).

No olvide hacer las citas necesarias:



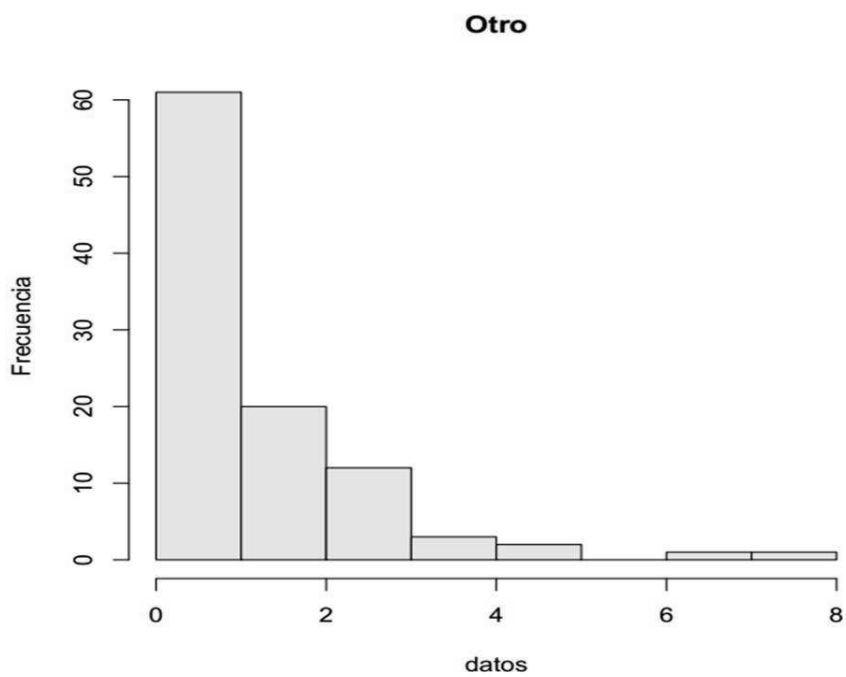


Figura 4: Datos. Fuente: elaboración propia.

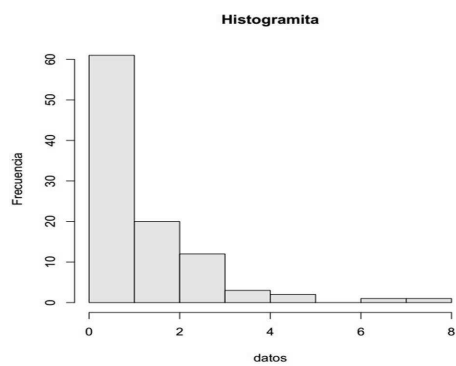


Figura 5: Datos. Fuente: elaboración propia.

```
<< echo=F,results=te>=>
library(xtable)
tabla=as.matrix(summary(datos))
xtable(tabla,caption="Resumen",label="F3")
@
```

	x
Min.	0.01
1st Qu.	0.42
Median	0.83
Mean	1.23
3rd Qu.	1.53
Max.	7.88

Tabla 1: Resumen. Fuente: elaboración propia.

```
<< echo=F>>=
citation()
citation(...)
@
```

## 10. Discusión

Hay que anotar que esta no es la única forma de hacer documentos con **Sweave**, pero sí una forma cómoda y rápida. Como parte de la llamada *programación literaria*, el *tejido* de programas fuente en los documentos se hace cada vez con más frecuencia. *Literate programming* o programación literaria se define como la combinación de documentación y código de tal manera que pueda ser entendida por seres humanos, algo así como un ensayo con macros. Para más información consulte la página [http://en.wikipedia.org/wiki/Literate\\_programming](http://en.wikipedia.org/wiki/Literate_programming) y <http://www.ntg.nl/maps/24/15.pdf>.

Además, al enseñar el uso de R, **Sweave** hace más eficiente la presentación de documentos para las clases, también se puede guardar los programas de R hechos en clase para luego ser comentados en documentos.

**Recibido: 24 de septiembre de 2012**

**Aceptado: 26 de octubre de 2012**

## Referencias

Leisch, F. (2002), 'Sweave: Dynamic generation of statistical reports using literate data analysis', *Proceedings in Computational Statistics*, Springer, pp. 575–580.

Lenth, R. V. (2012), *StatWeave User's Manual*, University of Iowa.

\*<http://www.stat.uiowa.edu/~rlenth/StatWeave/StatWeave-manual.pdf>

Team, R. C. (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

\*<http://www.R-project.org>

Verzani, J. (2011), *Getting Started with RStudio*, O'Reilly Media, USA.