

9. Introducción al análisis reproducible

Donald E. Knuth es el autor de los volúmenes *The Art of Computer Programming* y el creador de TeX

Definition 9.0.1 *Literate programming* (Donald Knuth, 1984)

Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.

No se trata de incluir muchos comentarios en el código de un programa, sino de trabajar como un escritor que explica a los usuarios lo que se está haciendo, introduciendo fragmentos de código que hacen lo que se está explicando

Software que automatice todo el proceso y garantice la reproducibilidad, como el descrito en [Sch+12], evitando tener que copiar y pegar información entre las distintas herramientas que se utilicen.

Implementación de investigación reproducible en R usando knitr [SLP14]

9.1 Exportación de resultados

Se podría crear un guión que automatizase el proceso: R generaScript.r - ejecutaAlg.sh - R procesaResultados.r - latex documento.tex
xtable - exportar a latex y html

9.1.1 xtable

Ejercicio 9.1 Exportar tablas de datos a LaTeX y HTML

```
> if(!is.installed('xtable'))  
+   install.packages('xtable')  
> library('xtable')  
> # Preparar una tabla por cada categoría existente. Cada una  
> # podría escribirse en un archivo distinto e importarse desde  
> # el documento principal
```

```

> tabla <- lapply(split(ebay[,c(-1,-7)], ebay$Category), xtable)

> print(tabla$Books[1:5,], digits=2, include.rownames=FALSE)

% latex table generated in R 3.1.0 by xtable 1.7-3 package
% Wed Aug 20 11:03:23 2014
\begin{table}[ht]
\centering
\begin{tabular}{lrrrlrr}
\hline
currency & sellerRating & Duration & endDay & ClosePrice & Competitive. \\
\hline
US & 352.00 & 7.00 & Sat & 11.50 & 1.00 \\
US & 30.00 & 7.00 & Mon & 0.70 & 0.00 \\
GBP & 149.00 & 7.00 & Thu & 1.77 & 0.00 \\
EUR & 410.00 & 5.00 & Mon & 1.23 & 0.00 \\
EUR & 666.00 & 5.00 & Thu & 1.23 & 0.00 \\
\hline
\end{tabular}
\end{table}

> print(tabla$Books[1:5,], tabular.environment = "longtable",
+       floating = FALSE)

% latex table generated in R 3.1.0 by xtable 1.7-3 package
% Wed Aug 20 11:03:23 2014
\begin{longtable}{rlrrlrr}
\hline
& currency & sellerRating & Duration & endDay & ClosePrice & Competitive. \\
\hline
134 & US & 352.00 & 7.00 & Sat & 11.50 & 1.00 \\
153 & US & 30.00 & 7.00 & Mon & 0.70 & 0.00 \\
177 & GBP & 149.00 & 7.00 & Thu & 1.77 & 0.00 \\
197 & EUR & 410.00 & 5.00 & Mon & 1.23 & 0.00 \\
198 & EUR & 666.00 & 5.00 & Thu & 1.23 & 0.00 \\
\hline
\hline
\end{longtable}

> print(tabla$Books[1:5,], type='HTML')

<!-- html table generated in R 3.1.0 by xtable 1.7-3
package -->
<!-- Wed Aug 20 11:03:23 2014 -->
<TABLE border=1>
<TR> <TH> </TH> <TH> currency </TH> <TH> sellerRating
</TH> <TH> Duration </TH> <TH> endDay </TH> <TH>
ClosePrice </TH> <TH> Competitive. </TH> </TR>
<TR> <TD align="right"> 134 </TD> <TD> US </TD> <TD
align="right"> 352.00 </TD> <TD align="right"> 7.00 </TD>
<TD> Sat </TD> <TD align="right"> 11.50 </TD> <TD
align="right"> 1.00 </TD> </TR>
<TR> <TD align="right"> 153 </TD> <TD> US </TD> <TD
align="right"> 30.00 </TD> <TD align="right"> 7.00 </TD>
<TD> Mon </TD> <TD align="right"> 0.70 </TD> <TD
align="right"> 0.00 </TD> </TR>
<TR> <TD align="right"> 177 </TD> <TD> GBP </TD> <TD
align="right"> 149.00 </TD> <TD align="right"> 7.00 </TD>

```

```
<TD> Thu </TD> <TD align="right"> 1.77 </TD> <TD
align="right"> 0.00 </TD> </TR>
<TR> <TD align="right"> 197 </TD> <TD> EUR </TD> <TD
align="right"> 410.00 </TD> <TD align="right"> 5.00 </TD>
<TD> Mon </TD> <TD align="right"> 1.23 </TD> <TD
align="right"> 0.00 </TD> </TR>
<TR> <TD align="right"> 198 </TD> <TD> EUR </TD> <TD
align="right"> 666.00 </TD> <TD align="right"> 5.00 </TD>
<TD> Thu </TD> <TD align="right"> 1.23 </TD> <TD
align="right"> 0.00 </TD> </TR>
</TABLE>
```

9.1.2 latex - Paquete Hmisc

Ejercicio 9.2 Exportar tablas de datos a LaTeX con latex()

```
> if(!is.installed('Hmisc'))
+   install.packages('Hmisc')
> library('Hmisc')
> tabla <- latex(ebay[ebay$Category=='Books',][1:5,])

> readLines(tabla$file)

%latex.default(ebay[ebay$Category == "Books", ][1:5, ])%
\begin{table}[!tbp]
\begin{center}
\begin{tabular}{llllrrllrrr}
\hline\hline
\multicolumn{1}{l}{ebay} & \multicolumn{1}{c}{Category} &
\multicolumn{1}{c}{currency} &
\multicolumn{1}{c}{sellerRating} &
\multicolumn{1}{c}{Duration} &
\multicolumn{1}{c}{endDay} &
\multicolumn{1}{c}{ClosePrice} &
\multicolumn{1}{c}{OpenPrice} &
\multicolumn{1}{c}{Competitive.}\tabularnewline
\hline
134&Books&US&$352&$7&Sat&$11.500000000000&$0.010000000000&$1$\tabularnewline
153&Books&US&$ 30&$7&Mon&$
0.700000000000&$0.700000000000&$0$\tabularnewline
177&Books&GBP&$149&$7&Thu&$
1.77000000012&$1.77000000012&$0$\tabularnewline
197&Books&EUR&$410&$5&Mon&$
1.230000000000&$1.230000000000&$0$\tabularnewline
198&Books&EUR&$666&$5&Thu&$
1.230000000000&$1.230000000000&$0$\tabularnewline
\hline
\end{tabular}\end{center}

\end{table}
```

9.1.3 Almacenar salidas de tests

sink() - escritura directa de la salida (Para resultados de tests)

mejor alguna opción como `texreg`, `htmlreg`, `screenreg` [Lei13]

Ejercicio 9.3 Exportar resultados de tests y modelos estadísticos

```
> if(!is.installed('texreg'))
+   install.packages('texreg')
> library('texreg')
> modelo1 <- lm(Sepal.Length ~ Sepal.Width, iris)
> modelo2 <- lm(Petal.Width ~ Sepal.Width, iris)
> modelo1      # Esta salida se podría guardar con sink()

Call:
lm(formula = Sepal.Length ~ Sepal.Width, data = iris)

Coefficients:
(Intercept)  Sepal.Width
      6.5262      -0.2234

> str(modelo1)

List of 12
 $ coefficients : Named num [1:2] 6.526 -0.223
 ..- attr(*, "names")= chr [1:2] "(Intercept)" "Sepal.Width"
 $ residuals : Named num [1:150] -0.644 -0.956 ...
 ..- attr(*, "names")= chr [1:150] "1" "2" ...
 $ effects : Named num [1:150] -71.57 -1.19 ...
 ..- attr(*, "names")= chr [1:150] "(Intercept)" "Sepal.Width"
 ...
 $ rank : int 2
 $ fitted.values: Named num [1:150] 5.74 5.86 ...
 ..- attr(*, "names")= chr [1:150] "1" "2" ...
 $ assign : int [1:2] 0 1
 $ qr :List of 5
 ..$ qr : num [1:150, 1:2] -12.2474 0.0816 ...
 .. ..- attr(*, "dimnames")=List of 2
 .. .. ..$ : chr [1:150] "1" "2" ...
 .. .. ..$ : chr [1:2] "(Intercept)" "Sepal.Width"
 .. ..- attr(*, "assign")= int [1:2] 0 1
 ..$ qraux: num [1:2] 1.08 1.02
 ..$ pivot: int [1:2] 1 2
 ..$ tol : num 1e-07
 ..$ rank : int 2
 ..- attr(*, "class")= chr "qr"
 $ df.residual : int 148
 $ xlevels : Named list()
 $ call : language lm(formula = Sepal.Length ~ Sepal.Width, data
 = iris)
 $ terms :Classes 'terms', 'formula' length 3 Sepal.Length ~
 Sepal.Width
 .. ..- attr(*, "variables")= language list(Sepal.Length,
 Sepal.Width)
 .. ..- attr(*, "factors")= int [1:2, 1] 0 1
 .. .. ..- attr(*, "dimnames")=List of 2
 .. .. .. ..$ : chr [1:2] "Sepal.Length" "Sepal.Width"
```

```

.. .. .$. : chr "Sepal.Width"
.. ..- attr(*, "term.labels")= chr "Sepal.Width"
.. ..- attr(*, "order")= int 1
.. ..- attr(*, "intercept")= int 1
.. ..- attr(*, "response")= int 1
.. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
.. ..- attr(*, "predvars")= language list(Sepal.Length,
Sepal.Width)
.. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric"
"numeric"
.. ..- attr(*, "names")= chr [1:2] "Sepal.Length"
"Sepal.Width"
$ model :'data.frame': 150 obs. of 2 variables:
..$ Sepal.Length: num [1:150] 5.1 4.9 4.7 4.6 5 ...
..$ Sepal.Width : num [1:150] 3.5 3 3.2 3.1 3.6 ...
..- attr(*, "terms")=Classes 'terms', 'formula' length 3
Sepal.Length ~ Sepal.Width
.. ..- attr(*, "variables")= language list(Sepal.Length,
Sepal.Width)
.. ..- attr(*, "factors")= int [1:2, 1] 0 1
.. ..- attr(*, "dimnames")=List of 2
.. ..$. : chr [1:2] "Sepal.Length" "Sepal.Width"
.. ..$. : chr "Sepal.Width"
.. ..- attr(*, "term.labels")= chr "Sepal.Width"
.. ..- attr(*, "order")= int 1
.. ..- attr(*, "intercept")= int 1
.. ..- attr(*, "response")= int 1
.. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
.. ..- attr(*, "predvars")= language list(Sepal.Length,
Sepal.Width)
.. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric"
"numeric"
.. ..- attr(*, "names")= chr [1:2] "Sepal.Length"
"Sepal.Width"
- attr(*, "class")= chr "lm"

> screenreg(list(modelo1, modelo2),
+           custom.model.names=c('Sepal length', 'Petal width'))

=====
              Sepal length  Petal width
-----
(Intercept)    6.53 ***      3.16 ***
              (0.48)        (0.41)
Sepal.Width   -0.22          -0.64 ***
              (0.16)        (0.13)
-----
R^2              0.01          0.13
Adj. R^2         0.01          0.13
Num. obs.       150           150
=====
*** p < 0.001, ** p < 0.01, * p < 0.05

> texreg(list(modelo1, modelo2),
+         custom.model.names=c('Sepal length', 'Petal width'))

```

```

\begin{table}
\begin{center}
\begin{tabular}{l c c }
\hline
& Sepal length & Petal width & \\
\hline
(Intercept) &  $6.53^{***}$  &  $3.16^{***}$  & \\
&  $(0.48)$  &  $(0.41)$  & \\
Sepal.Width &  $-0.22$  &  $-0.64^{***}$  & \\
&  $(0.16)$  &  $(0.13)$  & \\
\hline
R2 & 0.01 & 0.13 & \\
Adj. R2 & 0.01 & 0.13 & \\
Num. obs. & 150 & 150 & \\
\hline
\multicolumn{3}{l}{\scriptsize  $p < 0.001$ ,  $p < 0.01$ ,  $p < 0.05$ }}
\end{tabular}
\caption{Statistical models}
\label{table:coefficients}
\end{center}
\end{table}

> htmlreg(list(modelo1, modelo2),
+         custom.model.names=c('Sepal length', 'Petal width'))

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01
Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
<table cellpadding="0" cellspacing="0" align="center" style="border:
none;"> <caption align="bottom"
style="margin-top:0.3em;">Statistical models</caption>
<tr> <th style="text-align: left; border-top: 2px solid
black; border-bottom: 1px solid black; padding-right:
12px;"></th> <th style="text-align: left; border-top: 2px
solid black; border-bottom: 1px solid black;
padding-right: 12px;"><b>Sepal length</b></th> <th
style="text-align: left; border-top: 2px solid black;
border-bottom: 1px solid black; padding-right:
12px;"><b>Petal width</b></th> </tr> <tr> <td
style="padding-right: 12px; border:
none;">(Intercept)</td> <td style="padding-right: 12px;
border: none;">6.53<sup style="vertical-align:
4px;">***</sup></td> <td style="padding-right: 12px;
border: none;">3.16<sup style="vertical-align:
4px;">***</sup></td> </tr> <tr> <td style="padding-right:
12px; border: none;"></td> <td style="padding-right:
12px; border: none;">(0.48)</td> <td
style="padding-right: 12px; border: none;">(0.41)</td>
</tr> <tr> <td style="padding-right: 12px; border:
none;">Sepal.Width</td> <td style="padding-right: 12px;
border: none;">-0.22</td> <td style="padding-right: 12px;
border: none;">-0.64<sup style="vertical-align:
4px;">***</sup></td> </tr> <tr> <td style="padding-right:

```



```

12px; border: none;"></td> <td style="padding-right:
12px; border: none;">(0.16)</td> <td
style="padding-right: 12px; border: none;">(0.13)</td>
</tr> <tr> <td style="border-top: 1px solid black;">R<sup
style="vertical-align: 4px;">2</sup></td> <td
style="border-top: 1px solid black;">0.01</td> <td
style="border-top: 1px solid black;">0.13</td> </tr> <tr>
<td style="padding-right: 12px; border: none;">Adj. R<sup
style="vertical-align: 4px;">2</sup></td> <td
style="padding-right: 12px; border: none;">0.01</td> <td
style="padding-right: 12px; border: none;">0.13</td>
</tr> <tr> <td style="border-bottom: 2px solid
black;">Num. obs.</td> <td style="border-bottom: 2px
solid black;">150</td> <td style="border-bottom: 2px
solid black;">150</td> </tr> <tr> <td
style="padding-right: 12px; border: none;"
colspan="3"><span style="font-size:0.8em"><sup
style="vertical-align: 4px;">***</sup>p &lt; 0.001, <sup
style="vertical-align: 4px;">**</sup>p &lt; 0.01, <sup
style="vertical-align: 4px;">*</sup>p &lt;
0.05</span></td> </tr> </table>

```

	Sepal length	Petal width
(Intercept)	6,53*** (0,48)	3,16*** (0,41)
Sepal.Width	-0,22 (0,16)	-0,64*** (0,13)
R ²	0.01	0.13
Adj. R ²	0.01	0.13
Num. obs.	150	150

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$

Cuadro 9.1: Tabla generada por texreg()

9.1.4 Almacenar gráficas

exportación de gráficos - generar PDF y después importar desde latex

Visto en la presentación previa - producción de múltiples figuras cada una en su archivo

9.2 Introducción a Sweave

Ejercicio 9.4 Ejemplo de documento Sweave combinando LaTeX y un bloque con código R

```
\section{Listas}
```

En R las listas\index{listas} pueden contener datos heterogéneos,

```
incluyendo \textbf{data.frames} y otras listas  
...
```

```
«Listas»=
```

```
lst <- list(3.1415927, 'Hola', TRUE, fdias[4])  
lst  
unlist(lst)  
  
@
```

Al procesar el documento Sweave, el anterior bloque de código R se ejecutará y, aparte de los comandos, en el documento aparecerán también los resultados.

9.3 Introducción a knitr

cheat sheet en <http://shiny.rstudio.com/images/rm-cheatsheet.pdf.zip>

?integración de RStudio con GitHub para facilitar la reproducción de los experimentos