

Introducción a la ciencia reproducible con R

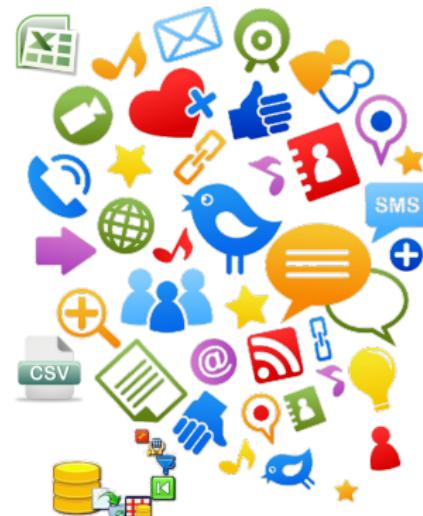
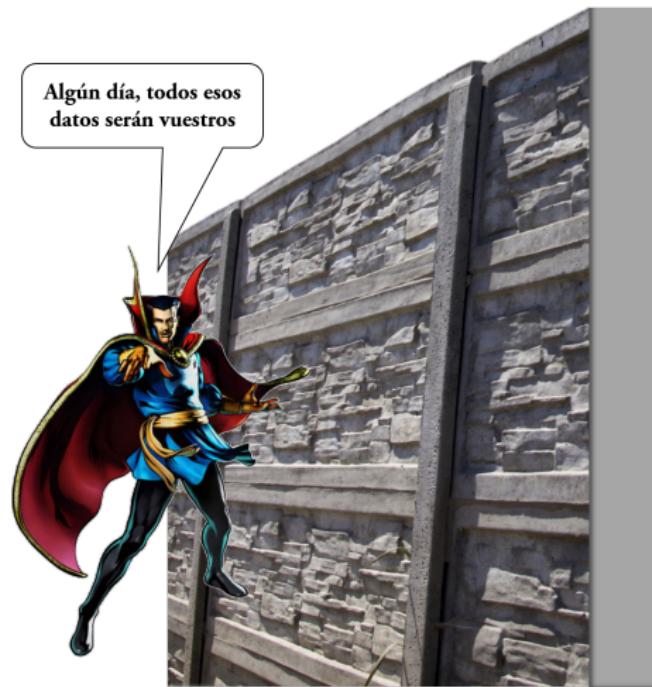
Document Freedom Day - OSLUGR

Francisco Charte (@fcharte, fcharte.com)

25 marzo 2015

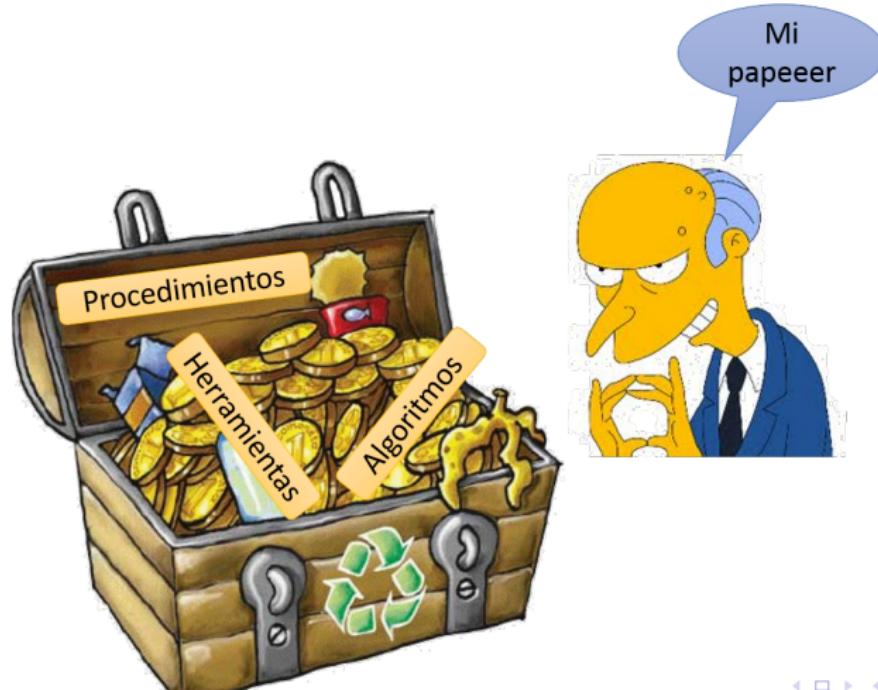
En qué consiste la ciencia reproducible

Datos disponibles y accesibles



En qué consiste la ciencia reproducible

Procedimientos, algoritmos y herramientas disponibles y accesibles



En qué consiste la ciencia reproducible

Las acciones manuales no son fácilmente reproducibles

Nothing can stop automation



Introducción a R

The screenshot shows the RStudio interface with the following components:

- Code Editor:** Displays R code in the `RBenchmarkings.R` file. The code includes sections for a presentation, a reproducibility check, and a plot generation.
- File Browser:** Shows the project structure with files like `graphics.R`, `CRANDownloads.R`, and `ClenciaReproducible.Rmd`.
- Plot Viewer:** Displays a line graph titled "CRAN downloads for package mldr". The y-axis is "Number of downloads" (0 to 25) and the x-axis is "Date" (from 2013-01-01 to 2015-01-01). The plot shows several peaks, with a major peak around January 2015 reaching approximately 25 downloads.

Importación de datos - Desde archivos

- ▶ CSV: `read.table()`
- ▶ Paquete `foreign`
 - ▶ Stata: `read.dta()`
 - ▶ SPSS: `read.spss()`
 - ▶ SAS: `read.ssd()`
 - ▶ dBase: `read.dbf()`
 - ▶ ARFF: `read.arff()`
 - ▶ Octave: `read.octave()`
- ▶ Excel: `loadWorkbook()` - Paquete `XLConnect`
- ▶ OpenOffice/LibreOffice: `read.ods()` - Paquete `ROpenOffice`
- ▶ XML: `xmlParse()` - Paquete `XML`

Importación de datos - Desde bases de datos

- ▶ SQLite: `RSQLite`
- ▶ Oracle: `ROracle`
- ▶ MySQL: `RMySQL`
- ▶ PostgreSQL: `RPostgreSQL`
- ▶ ODBC/JDBC: `RODBC/RJDBC`
- ▶ MongoDB: `rmongodb`
- ▶ CouchDB: `R4CouchDB`
- ▶ Cassandra: `RCassandra`

Importación de datos - Otras fuentes

- ▶ Hadoop: RHadoop
- ▶ Spark: SparkR
- ▶ JSON: jsonlite
- ▶ Web: parseHTML() y readHTMLTable() - Paquete XML
- ▶ Web: GET() - Paquete httr
- ▶ Web: getURL() y getForm() - Paquete RCurl

Importación de datos - Ejemplo

Tags de los posts del foro de cocina de Stack Exchange

```
library(XML)
library(tm)
library(wordcloud)
content<-xmlTreeParse(filename)
...
docs<-Corpus(VectorSource(content))
docs<-tm_map(docs,
content_transformer(tolower))
...
wordcloud(DocumentTermMatrix(docs))
```



Importación de datos - Ejemplo

Datos obtenidos de la API de GitHub en formato JSON
runGitHub('GitHubMining', 'fcharte')

The screenshot shows the GitHubMining R package interface. It includes:

- API rate limits:** Shows Core Limit (5000), Search Limit (30), Remaining (1131), and Reset (2015-03-22 13:57:12).
- Github user account:** Fields for User name (fcharte) and Password.
- Users:** A table listing GitHub users with columns: Login, Name, Repos, Contribs, Followers, Following, Registered, and LastUpdate.

Login	Name	Repos	Contribs	Followers	Following	Registered	LastUpdate
rankingfaker	Falso commiteador	2	85037	0	0	2015-03-03	2015-03-10
stringparser	Javier Carrillo	23	2791	35	131	2014-05-01	2015-03-21
vterrón	Víctor Terrón	17	2199	62	5	2012-03-26	2015-03-22
ernestoalejo	Ernesto Alejo	32	1675	10	4	2011-08-05	2015-03-21
pleonex	Benito Palacios	29	1398	21	13	2012-12-22	2015-03-22
JJ	Juan Julián Merelo Guervós	160	1135	156	33	2008-02-20	2015-03-22
Amab	Juan Miguel Boyero Corral	2	1083	13	13	2010-10-31	2015-03-20
M42	Mario Román	16	751	40	49	2013-08-29	2015-03-21
frenlu	Fran	14	605	11	31	2010-09-23	2015-03-20

Publicación de datos