

# Introducción a la ciencia reproducible con R

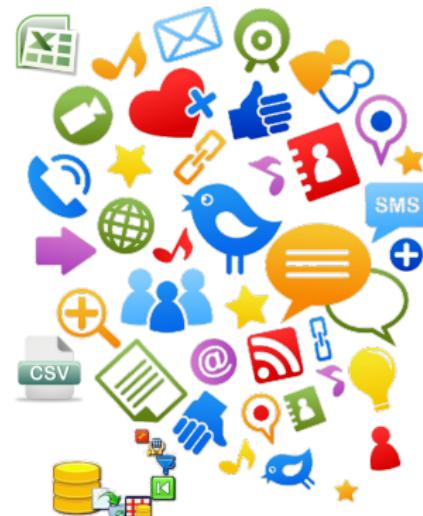
Document Freedom Day - OSLUGR

Francisco Charte (@fcharte, fcharte.com)

25 marzo 2015

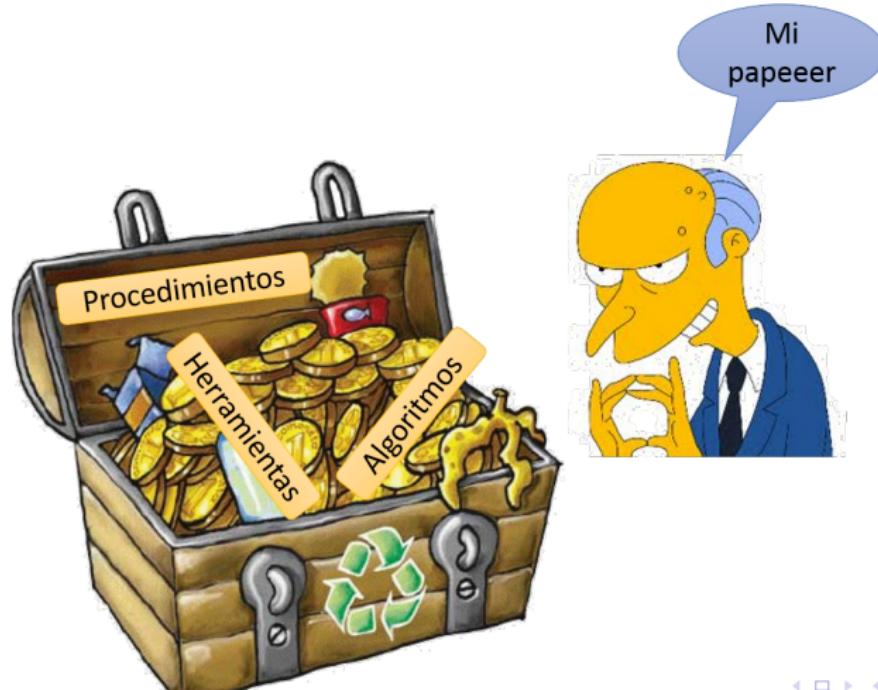
# En qué consiste la ciencia reproducible

## 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 an R Markdown file titled "Introducción a la ciencia reproducible con R". The code includes sections for title, subtitle, author, date, output, theme, and colortheme, followed by two sections for "En qué consiste la ciencia reproducible" and "Procedimientos, algoritmos y herramientas disponibles y accesibles".
- Environment Tab:** Shows the current project is "mldr" and the active file is "R/evaluate.R".
- Plots:** A line plot titled "CRAN downloads for package mldr" showing the number of downloads over time. The y-axis is labeled "Number of downloads" and ranges from 0 to 25. The x-axis is labeled "Date" and shows dates from 2015-01-01 to 2015-12-31. The plot shows a significant peak around December 2015.
- Console:** Displays R code and its output. The code reads a dataset from a local file and plots it as a line graph with black squares and a grey line. The output shows the sum of downloads is 294 and the mean is approximately 8.16.

# 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))
```



# Publicación de datos