Introducción a la ciencia reproducible con R

Document Freedom Day - OSLUGR

Francisco Charte (@fcharte, fcharte.com)

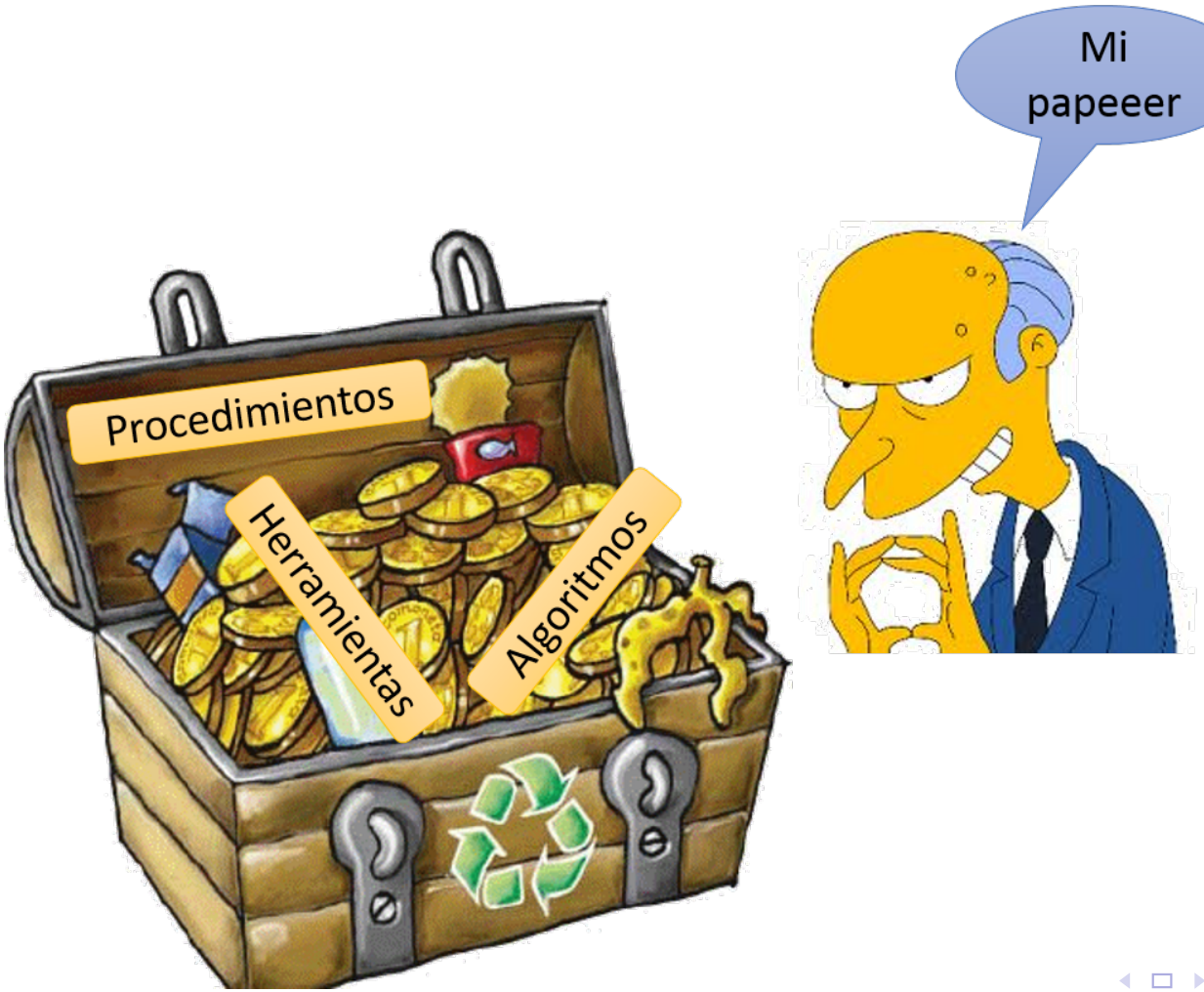
25 marzo 2015

En qué consiste la ciencia reproducible

Datos disponibles y accesibles

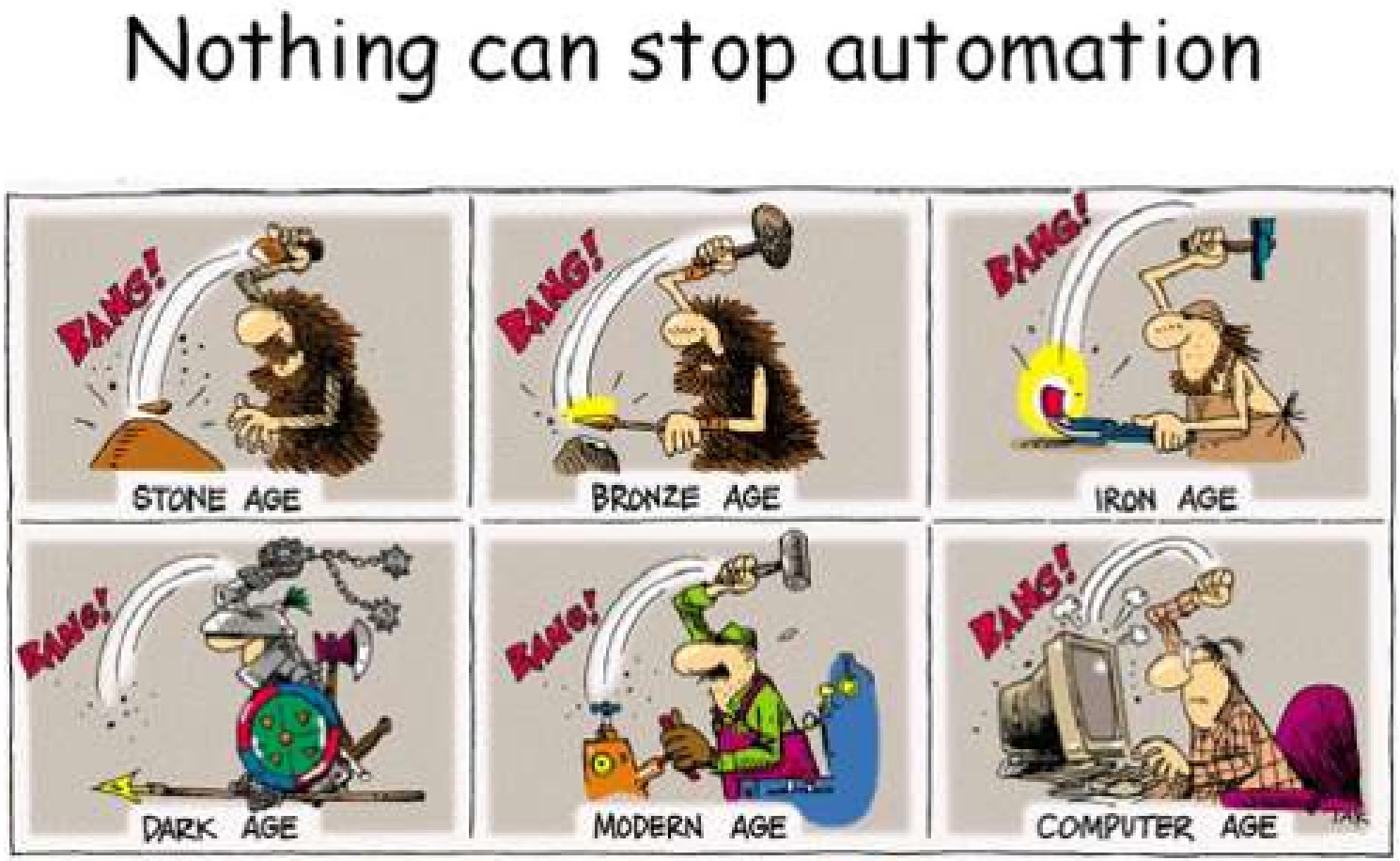
En qué consiste la ciencia reproducible

Procedimientos, algoritmos y herramientas disponibles y accesibles

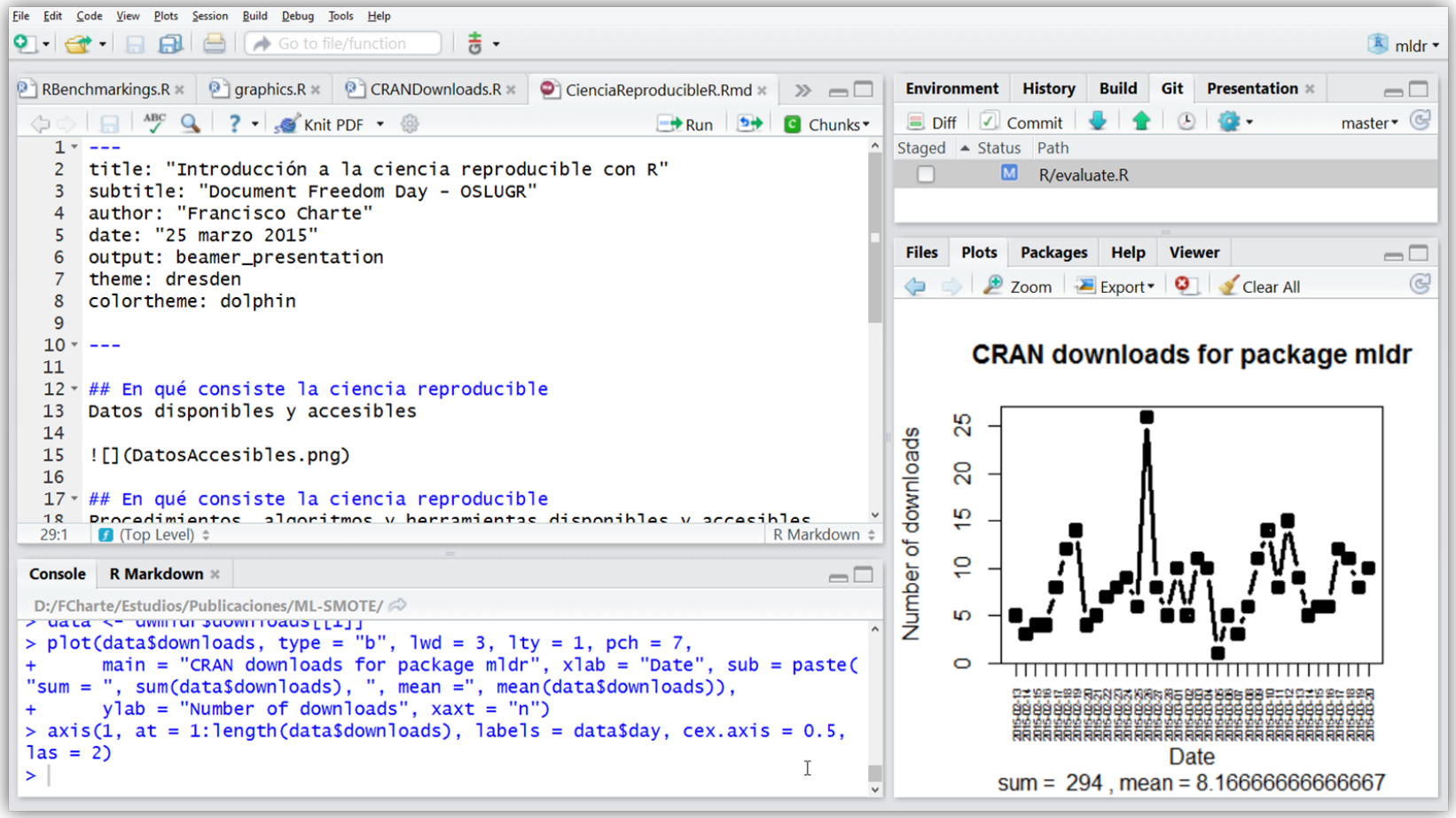


En qué consiste la ciencia reproducible

Las acciones manuales no son fácilmente reproducibles



Introducción a R



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`

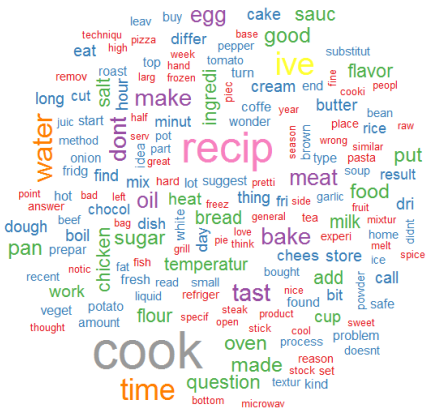
Importación de datos - Otras fuentes

- 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

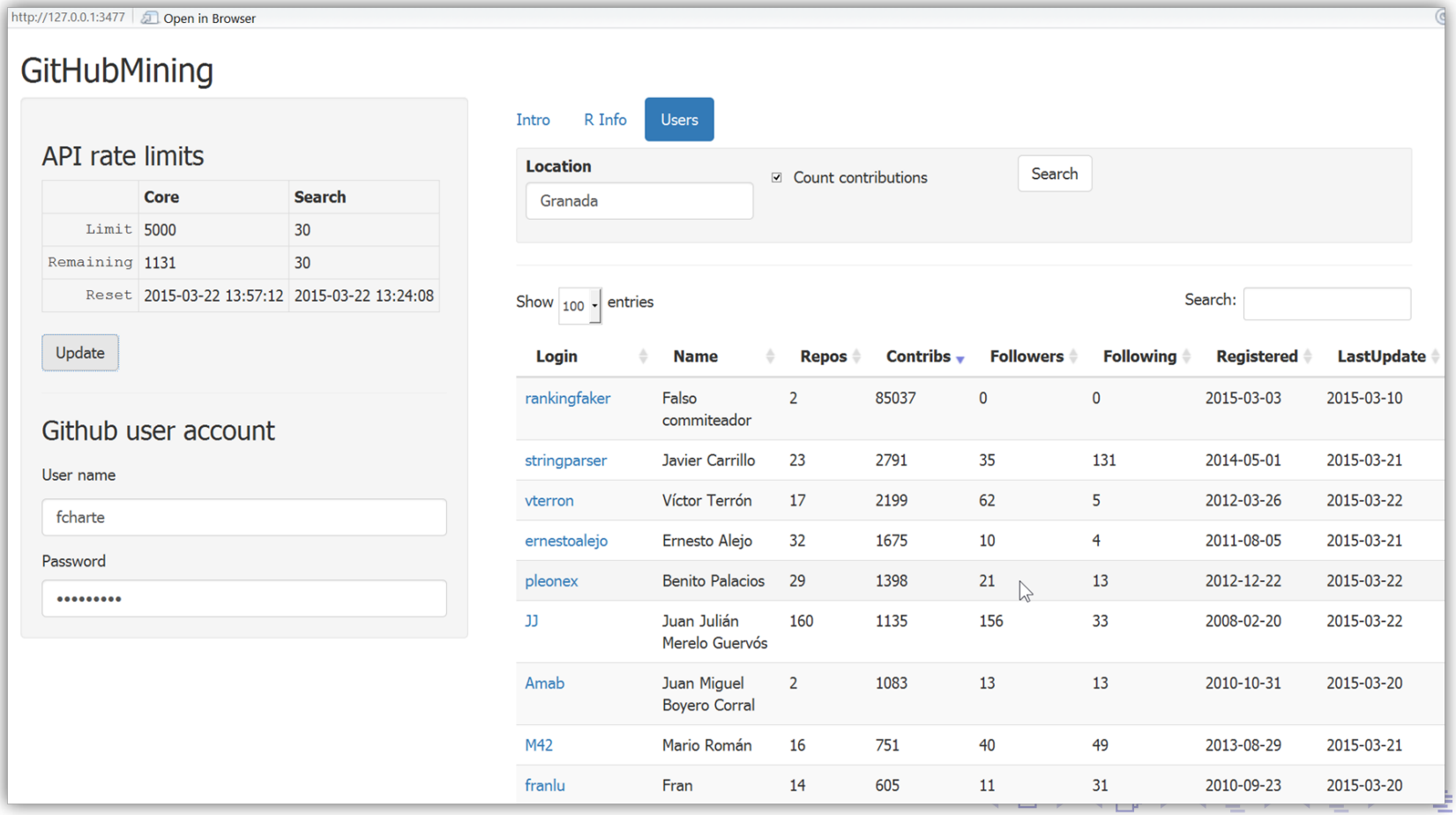
```
library(XML)
library(tm)
library(wordcloud)
content<-xmlTreeParse(filename)
...
docs<-Corpus(VectorSource(content))
docs<-tm_map(docs,
content_transformer(tolower))
...
wordcloud(DocumentTermMatrix(docs))
```



Importación de datos - Ejemplo

Datos obtenidos de la API de GitHub en formato JSON

`runGitHub('GitHubMining', 'fcharte')`



Publicación de datos