

XI XORNADA DE USUARIOS DE EN **GALICIA**

The R logo consists of a blue capital letter 'R' with a white outline, positioned inside a grey oval shape that also has a white outline.

**R, Webscraping y Open Science.
Esquivando balas en Matrix**

A decorative light blue wave graphic is located in the bottom right corner of the slide.



DATOS



The Political Economy of Revolving Doors in Spain 1920-2020

Dr. Sebastián Lavezzolo





EL PARLAMENTO ▾

ACTIVIDAD ▾

PARTICIPA ▾

TRANSPARENCIA ▾

Inicio > [El Parlamento](#) > Legislaturas anteriores

Legislaturas anteriores

I Legislatura

II Legislatura

Descargar Listado de Procuradores

(Se descargará un fichero en formato .CSV que podrá ser abierto y manejado con programas de ofimática como hojas de cálculo y gestores de bases de datos)

La sede

Video Institucional

Historia de la institución

X Legislatura

IX Legislatura

VIII Legislatura

VII Legislatura

VI Legislatura

V Legislatura

IV Legislatura

III Legislatura

II Legislatura

I Legislatura

Etapla provisional

Parlamento Cultural

Diputados



Grupo Parlamentario Socialista

- [María Inmaculada Ortega Martínez](#)
- [Francisco Javier Rodríguez Peña](#)
- [Ana María Santos Preciado](#)
- [José Ángel Lacalzada Esquivel](#)
- [Emilia Fernández Núñez](#)
- [Pablo Rubio Medrano](#)
- [Félix Caperos Elosúa](#)
- [Concepción Andreu Rodríguez](#)
- [Jesús María García García](#)
- [Manuela Galdámez Sola](#)
- [María Sonia Iburguren Ruiz](#)



Grupo Parlamentario Popular

- [Pedro Sanz Alonso](#)
- [Carlos Cuevas Villoslada](#)
- [José Ignacio Cenicerros González](#)
- [María Yolanda Preciado Moreno](#)
- [Noemí Manzanos Martínez](#)
- [Emilio del Río Sanz](#)



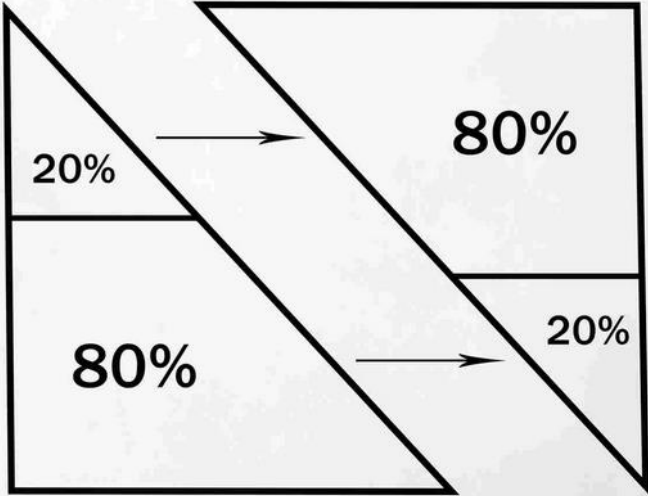


```
library(rvest)
```

```
library(dplyr)
```

```
library(stringr)
```

Pareto Principle




```
scrap_canarias <- function(nombre_variable, link, ano_inicio, ano_fin)
{
  library(rvest)
  library(dplyr)
  read <- read_html(link)

  nombre <- read %>% html_nodes("td:nth-child(2)") %>% html_text()
  n <- length(nombre)
  head(nombre)

  entrada <- rep(ano_inicio, n)
  salida <- rep(ano_fin, n)
  gp <- read %>% html_nodes("td~ td+ td") %>% html_text()

  df <- data.frame(nombre, entrada, salida, gp) #quite cargo
  df$nombre <- normalizar_chr(df$nombre)
  df$gp <- normalizar_chr(df$gp)

  #assign(nombre_variable, df, envir = .GlobalEnv)

  write.csv(df, file=paste0(as.character(nombre_variable), ".csv"), fileEncoding="UTF-8", row.names = F)
}
```



SelectorGadget

Easy, powerful CSS Selector generation.

Detalles

Quitar



Diputados y diputadas por orden alfabético

ABCDEFGHIJKLMNÑOPQRSTUVWXYZ

A	Abrante Brito, Manuel Jesús	GP SC
	Acosta Ármaz, Raúl	GP M
	Alemán Ojeda, Navra	GP SC
B	Barragán Cabrera, José Miguel	GP CCa
	Barreto Hernández, Nieves Lady	GP CCa
	Beato Castellano, Socorro	GP CCa
	Bergaz Villalba, Marcos Antonio	GP SC
	Bermúdez Esparza, José Manuel	GP SC
	Betancort García, Oswaldo	GP SC

Otras legislaturas
XI Legislatura
X Legislatura
IX Legislatura
VIII Legislatura
VII Legislatura
VI Legislatura

tr:nth-child(2) a

Clear (1)

Toggle Position

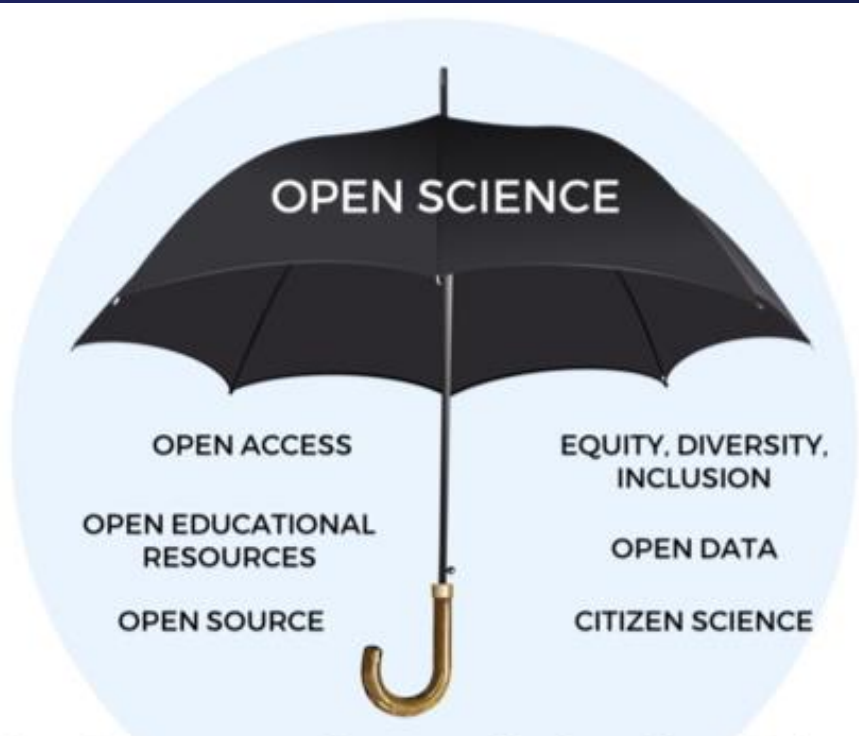
XPath

?

X

Nunca es tan fácil





What do we mean when we talk about Open Science?

Image courtesy of Robin Champieux

Referencias

wickham H (2022). `_rvest: Easily Harvest (Scrape) Web Pages_`. R package version 1.0.3, <<https://CRAN.R-project.org/package=rvest>>.

wickham H (2022). `_stringr: Simple, Consistent Wrappers for Common String Operations_`. R package version 1.5.0, <<https://CRAN.R-project.org/package=stringr>>.

wickham H, François R, Henry L, Müller K, Vaughan D (2023). `_dplyr: A Grammar of Data Manipulation_`. R package version 1.1.2, <<https://CRAN.R-project.org/package=dplyr>>.