

Taller Análisis Político I. Sesión I

Master de Análisis Político y Electoral

Javier Lorenzo
javier.lorenzo@uc3m.es

Universidad Carlos III de Madrid

November 7, 2023

Presentación

Análisis de
datos en R

1. Presentaciones
2. Instalación de R, Rstudio
3. Breve introducción a R
4. Manipulación de datos

Presentación

Análisis de
datos en R

- Licenciado en Ciencias Políticas UCM
- Doctorado en UCM
- Máster en Sociedad de la Información y Nuevas Tecnologías
- Dr. Europeo Ciencias Políticas y Sociología UC3M
- Prof. Visitante en Universita di Roma III
- Prof. Visitante State University of New York, Binghamton
- ICPSR Advanced Social Research methods
- Investigador visitante en Social Media and Political Participation Lab (NYU)
- EITM Advanced Social Research Methods with computational techniques

Investigador Senior Colegio Oficial de Polítólogos y Sociólogos de Madrid

Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE

Investigación: Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE
- TIOVIVO: Las Redes Sociales en Perspectiva de Genero: Techo Virtual o Ventana de Oportunidad?

Investigación: Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE
- TIOVIVO: Las Redes Sociales en Perspectiva de Genero: Techo Virtual o Ventana de Oportunidad?
- Webcampaigning ESP-IT

Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE
- TIOVIVO: Las Redes Sociales en Perspectiva de Genero: Techo Virtual o Ventana de Oportunidad?
- Webcampaigning ESP-IT
- Desinformación / Fake News

Investigación: Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE
- TIOVIVO: Las Redes Sociales en Perspectiva de Genero: Techo Virtual o Ventana de Oportunidad?
- Webcampaigning ESP-IT
- Desinformación / Fake News
- New media, Political Information, Political discussion, and Participation in Spain

Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE
- TIOVIVO: Las Redes Sociales en Perspectiva de Genero: Techo Virtual o Ventana de Oportunidad?
- Webcampaigning ESP-IT
- Desinformación / Fake News
- New media, Political Information, Political discussion, and Participation in Spain
- Polarización ideológica

Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE
- TIOVIVO: Las Redes Sociales en Perspectiva de Genero: Techo Virtual o Ventana de Oportunidad?
- Webcampaigning ESP-IT
- Desinformación / Fake News
- New media, Political Information, Political discussion, and Participation in Spain
- Polarización ideológica
- Polarización afectiva

Comportamiento político online de las élites

Presentación

Análisis de
datos en R

- Adopción de la tecnología de líderes y partidos UE
- TIOVIVO: Las Redes Sociales en Perspectiva de Genero: Techo Virtual o Ventana de Oportunidad?
- Webcampaigning ESP-IT
- Desinformación / Fake News
- New media, Political Information, Political discussion, and Participation in Spain
- Polarización ideológica
- Polarización afectiva
- Actitudes vs Comportamientos revelados (VAA)

LII:Comportamiento social en escenarios de movilidad: juventud, migración

Presentación

Análisis de
datos en R



The image displays two project websites side-by-side. On the left is the MICADO website, featuring a pink header with the text "MICADO" and "Welcome to MICADO". Below this, it says "Public administration, migration research and IT specialists connect to create a business intelligence solution that supports migrant arrival and participation processes." A "PROJECT GOALS" button is at the bottom. On the right is the MOVE website, which has a dark header with the word "Move" and a globe icon. It features sections for "Researchers" (with the subtext "You are interested in youth research") and "Young people" (with the subtext "You are between 18 and 29 and interested in staying abroad in the EU"). There's also an "Animated Clip" button. At the bottom of the page, there's a sidebar with a teal box labeled "+ estudios" and a red box labeled "+ oficina".

Presentación

Análisis de
datos en R

El Voto electrónico en España

Adopción de perfiles
en redes sociales EPE2014
(*Negative binomial poisson regression models* and Clinton primaries campaign
Tweet and Facebook collecting data out of 2000–2million tweets)
candidates

Opinion Leaders in Social Media: Trump
(*Negative binomial poisson regression models* and Clinton primaries campaign
Tweet and Facebook collecting data out of 2000–2million tweets)
Data mining



Voto residentes ausentes

Webcampaigning ESP-ITA 2004-...
(*Web analysis*
Data Gathering both parties and candidates from SNS)

Seizing the Populist Rhetorical Toolkit: A Comparative Analysis of Trump and Clinton's Discourse on Twitter during the 2016 U.S. Presidential Campaign
Con Paco Seoane.
(*Handcoded tweets, simple OLS*)

Presentación

Análisis de
datos en R

¿Qué es R?

- Es un entorno de programación orientado al cálculo, manipulación de datos, y representación gráfica, publicado como software libre con licencia GNU-GPL.



[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#).

r-project

- Disponible para diferentes plataformas (Unix, MacOS, Windows).
- Muy usado en la investigación científica.

Presentación

Análisis de datos en R

- Una de las grandes riquezas de R es la cantidad de paquetes que amplían sus funcionalidades.
- La lista completa está en <http://cran.es.r-project.org/web/packages/>.
- Las CRAN Task Views agrupan por temáticas: <http://cran.r-project.org/web/views/>

Desventajas

Ventajas

1. Software Libre y Rápido.

Desventajas

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.

Desventajas

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.

Desventajas

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.

Desventajas

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.
5. Te obliga a pensar en tu análisis.

Desventajas

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.
5. Te obliga a pensar en tu análisis.
6. Interfaces con software de almacenamiento de base de datos (SQL)

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.
5. Te obliga a pensar en tu análisis.
6. Interfaces con software de almacenamiento de base de datos (SQL)

Desventajas

1. No amigable para el usuario al inicio: curva de aprendizaje empinada, GUI mínima.

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.
5. Te obliga a pensar en tu análisis.
6. Interfaces con software de almacenamiento de base de datos (SQL)

Desventajas

1. No amigable para el usuario al inicio: curva de aprendizaje empinada, GUI mínima.
2. No hay apoyo comercial; Descubrir los métodos correctos o cómo utilizar una función por tu cuenta puede ser frustrante.

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.
5. Te obliga a pensar en tu análisis.
6. Interfaces con software de almacenamiento de base de datos (SQL)

Desventajas

1. No amigable para el usuario al inicio: curva de aprendizaje empinada, GUI mínima.
2. No hay apoyo comercial; Descubrir los métodos correctos o cómo utilizar una función por tu cuenta puede ser frustrante.
3. Fácil de cometer errores y no saber que lo hiciste.

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.
5. Te obliga a pensar en tu análisis.
6. Interfaces con software de almacenamiento de base de datos (SQL)

Desventajas

1. No amigable para el usuario al inicio: curva de aprendizaje empinada, GUI mínima.
2. No hay apoyo comercial; Descubrir los métodos correctos o cómo utilizar una función por tu cuenta puede ser frustrante.
3. Fácil de cometer errores y no saber que lo hiciste.
4. Trabajar con grandes conjuntos de datos está limitado por RAM.

Ventajas

1. Software Libre y Rápido.
2. State of the Art: Los investigadores estadísticos proporcionan sus métodos como paquetes de R.
3. Comunidad muy activa.
4. Excelente para simulación, programación, análisis intensivo de datos, etc.
5. Te obliga a pensar en tu análisis.
6. Interfaces con software de almacenamiento de base de datos (SQL)

Desventajas

1. No amigable para el usuario al inicio: curva de aprendizaje empinada, GUI mínima.
2. No hay apoyo comercial; Descubrir los métodos correctos o cómo utilizar una función por tu cuenta puede ser frustrante.
3. Fácil de cometer errores y no saber que lo hiciste.
4. Trabajar con grandes conjuntos de datos está limitado por RAM.
5. La preparación y limpieza de los datos puede ser más complicada y más propensa a errores en R vs. SPSS o SAS

Presentación

Análisis de
datos en R

- Es **DIFÍCIL** aprender **todo** de R

Presentación

Análisis de
datos en R

- Es **DIFÍCIL** aprender **todo** de R
- Es **FÁCIL** aprender a resolver problemas cotidianos

Presentación

Análisis de
datos en R

- Es DIFÍCIL aprender **todo** de R
- Es FÁCIL aprender a resolver problemas cotidianos
- Es DIFÍCIL digerirlo todo

Presentación

Análisis de
datos en R

- Es DIFÍCIL aprender todo de R
- Es FÁCIL aprender a resolver problemas cotidianos
- Es DIFÍCIL digerirlo todo
- Es MEJOR ir por partes

Aproximación pedagógica del curso

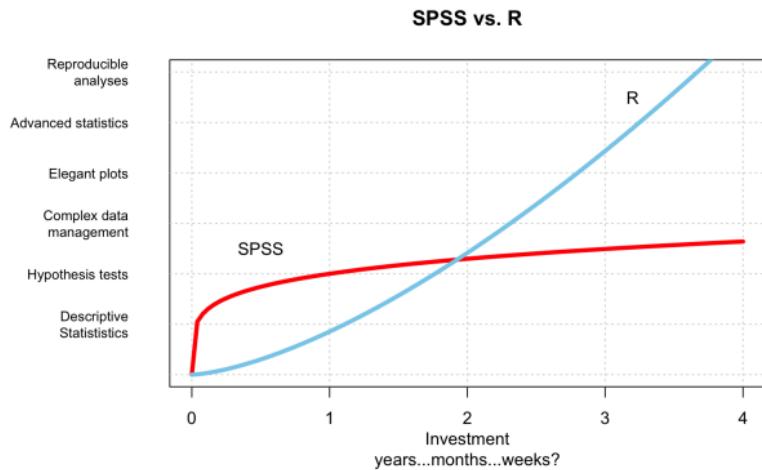
Presentación

Análisis de
datos en R

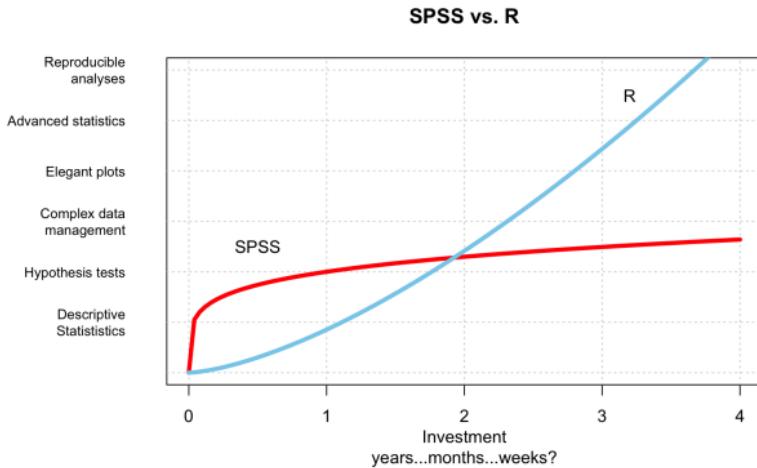


- Es **aprender un idioma**: no necesitamos saberlo todo de una y todo a la vez, si no lo que vamos necesitando.
- Hoy vamos a aprender la estructura, la lógica y algunos de sus términos mas comúnmente usados

- Aprendiendo
haciendo

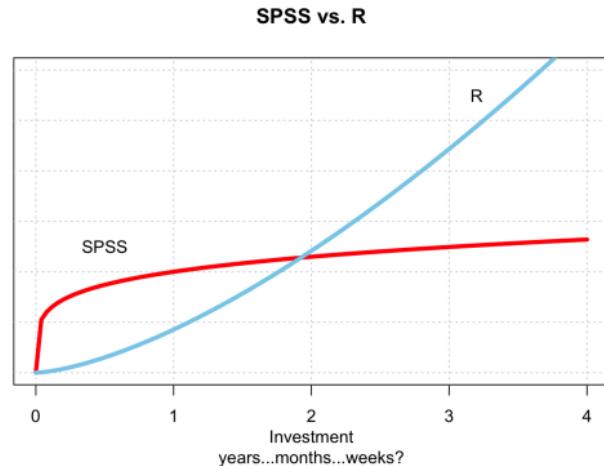


- Aprendiendo haciendo
- Equivocándonos



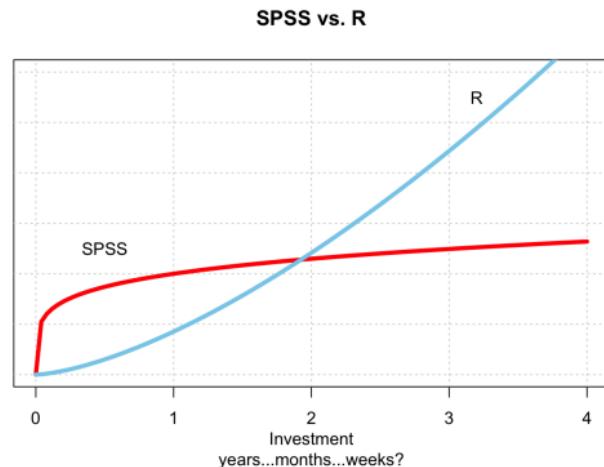
- Aprendiendo haciendo
- Equivocándonos
- Buscando soluciones a lo que nos encontramos cuando estemos a solas

Reproducible analyses
Advanced statistics
Elegant plots
Complex data management
Hypothesis tests
Descriptive Statistics



- Aprendiendo haciendo
- Equivocándonos
- Buscando soluciones a lo que nos encontramos cuando estemos a solas
- Disponer de las herramientas y recursos para resolver problemas que vayamos encontrando

Reproducible analyses
Advanced statistics
Elegant plots
Complex data management
Hypothesis tests
Descriptive Statistics



Empecemos con Rstudio, y un vistazo a POSIT

Análisis de
datos en R

Presentación

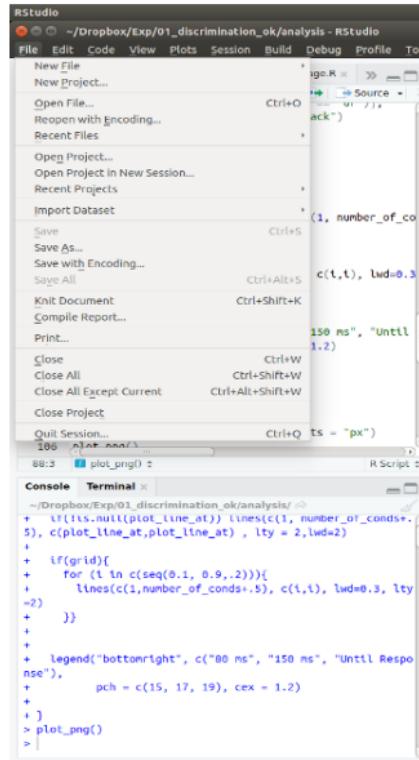
Análisis de
datos en R

- Código: puedes escribir tu código de R y es persistente si lo guardas en un archivo con formato R
- Consola: Ejecuta el código R que le envíes

Presentación

Análisis de datos en R

- La forma más fácil de guardar y almacenar tus tareas. Cada '*Project*' debe ser un proyecto a parte y se guarda en un archivo ".Rproj".
- Abriremos Rstudio y creamos nuestro proyecto
- project (*Rcourse.Rproj*)
- Do!



The screenshot shows the RStudio interface. The title bar says "RStudio ~Dropbox/Exp/01_discrimination_ok/analysis - RStudio". The "File" menu is open, showing options like "New File", "New Project...", "Open File...", "Recent Projects", and "Import Dataset". Below the menu, there's a list of recent files. On the right side, there's a "Console" tab with some R code and output, and a "Terminal" tab with a command-line interface.

```
~Dropbox/Exp/01_discrimination_ok/analysis/ >
+   if(lts.null(plot_line_at)) lINES(c(1, number_of_consts+5), c(plot_line_at,plot_line_at) , lty = 2,lwd=2)
+
+   if(grid){
+     for (l in c(seq(0.1, 0.9,.2))){
+       lINES(c(l,number_of_consts+.5), c(i,t), lwd=0.3, lty=2)
+     }
+   }
+
+   legend("bottomright", c("00 ms", "150 ms", "Until Responce"),
+         pch = c(15, 17, 19), cex = 1.2)
+
+ }
> plot_png()
> |
```

Presentación

Análisis de
datos en R

- Todo lo que existe es un objeto
- Todo lo que ocurre es una función
- Aunque parezca trivial, en el futuro entenderéis su importancia

Presentación

Análisis de
datos en R

- Se utiliza `<-` para asignar
- objeto `<- "valor"`
- Distintos tipos de objetos:
 - integers (1)
 - texto ("hi")
 - listas
 - funciones (a `<- function() print("hi")`,
 - data.frames,
 - lógicos (TRUE) ,

Presentación

Análisis de
datos en R

- R es muy tonto por lo que SOLO seguirá tus instrucciones si las das con cuidado, así que controla tu escritura
- (hello != Hello, hell0 != hellO]), etc.

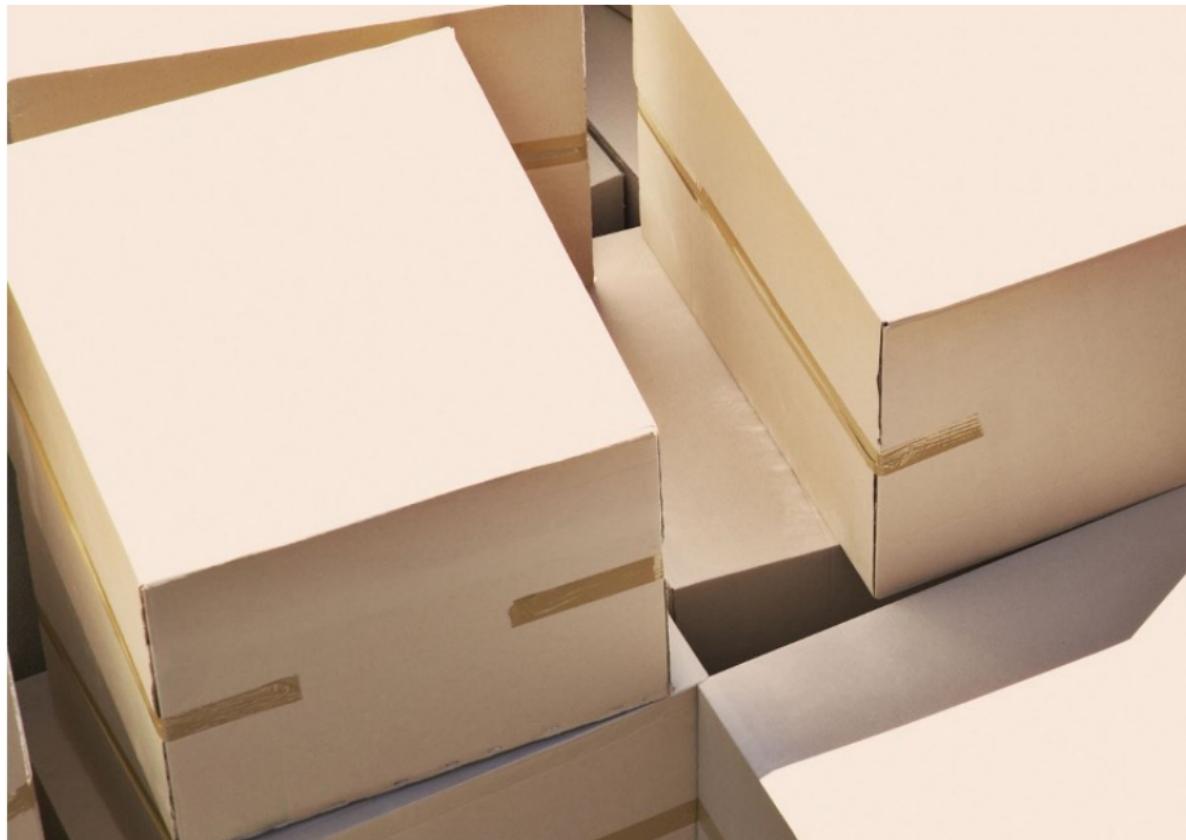
Presentación

Análisis de
datos en R

- Código legible
 - 80 caracteres de longitud por cada linea aconsejable
 - Hazte un favor a tu futuro yo y déjate comentarios
- Se consistente con la forma de escribir
- Mas consejos en el md de reproducibilidad

Presentación

Análisis de
datos en R



Presentación

Análisis de datos en R

- R se mantiene actualizado gracias a que tiene una activa comunidad que comparte código, actualiza paquetes e innova constantemente en sus librerías y funcionalidades.
- Solo en **CRAN** hay cerca de 4000 paquetes, lo que le da a R gran funcionalidad.
- Hay que sumar los paquetes de **github** no verificados pero constantemente actualizados por sus autores
- Una de las ventajas de R es que el código de los paquetes es **abierto**, incluyen **documentación**, y es **reproducible**.

Presentación

Análisis de
datos en R

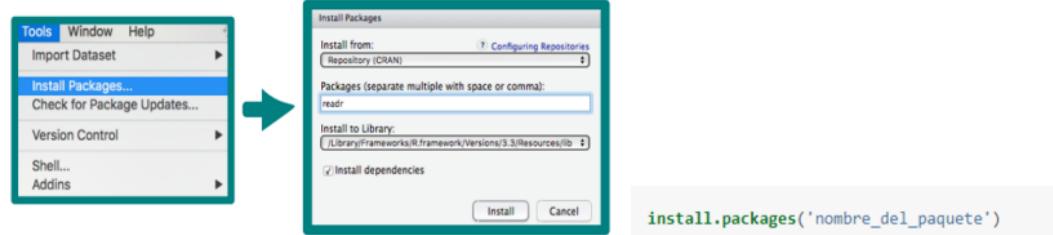
- R tiene siete paquetes predeterminados al iniciar una nueva sesión.
- La función `search()` da la lista de los paquetes cargados en la sesión de R abierta.
- Se puede trabajar con "`base`" y hacer numerosas operaciones

```
## [1] ".GlobalEnv"           "package:stats"      "package:graphics"  
## [4] "package:grDevices"    "package:utils"       "package:datasets"  
## [7] "Autoloads"            "package:base"
```

Presentación

Análisis de
datos en R

- A pesar de los paquetes predeterminados, muchas veces es necesario instalar paquetes de CRAN.
- Existen dos formas de instalar paquetes al espacio de trabajo de R



Presentación

Análisis de
datos en R

- Los paquetes se instalan una vez en el entorno de trabajo local de R.
- No es necesario instalar los paquetes cada sesión nueva que abras.
- Sin embargo, al descargar una nueva versión de R el entorno de trabajo de R local cambia, por lo que deberás instalar de nuevo los paquetes.

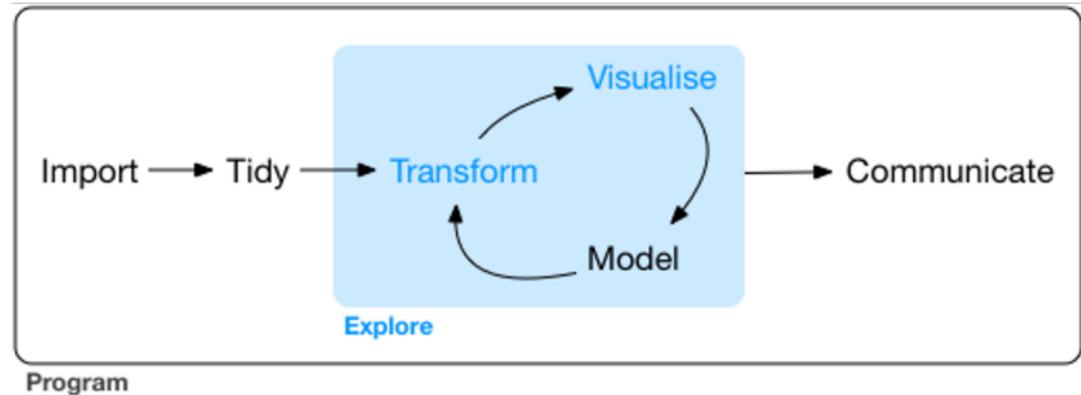


Presentación

Análisis de datos en R

- Un paquete trae al “environment” local, es decir, el lugar donde se almacenan los objetos de R sus propios objetos (funciones, datasets, etc)
- Puede pasar que si un paquete genera una función que tiene el mismo nombre generará un conflicto
- Puedes probar a cargar el paquete dplyr, después del paquete MASS y probar a ver que pasa con la función select(). Como ambos paquetes traen un objeto “select()” R usará el del que se haya cargado último.
- Puedes forzar el uso de una función de un paquete con “::”
- *dplyr :: select()* usará la función select() del paquete dplyr
- Por su parte, *MASS :: select()* usará la del paquete MASS





Exploramos el dataset para:

Presentación

Análisis de
datos en R

- Generar preguntas sobre tus datos.
- Buscar respuestas a estas preguntas.
- Usar lo aprendido para plantearnos nuevos retos sobre este set de datos.



- No existe un guión.
- Objetivo principal: Entender tus datos.
- Hacerse muchas preguntas.

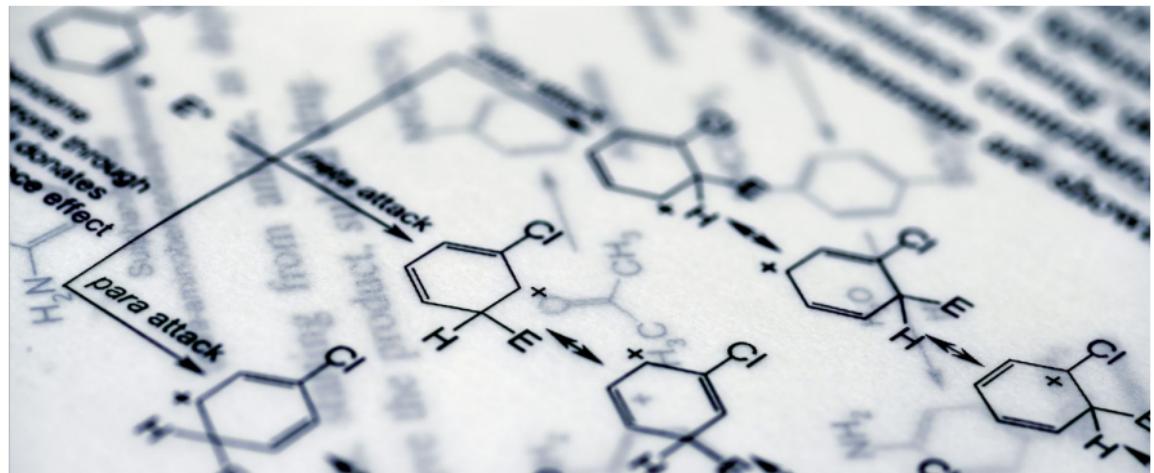
Presentación

Análisis de datos en R

- Variables: Una cantidad, calidad o propiedad que puedes medir. Peso.
- Valor: El estado de las variables cuando las mides. 500 gr.
- Observación: Un conjunto de medidas hechas bajo condiciones similares. Una observación puede contener varios valores, cada uno asociado a una variable. Peso: 500 gr, Altura: 2 m.
- Esto es importante para lo que llamamos tidy data o dato limpio.

Presentación

Análisis de datos en R



Una función es un objeto

```
mi_funcion <- function(x){  
  y <- x+1  
  return(y)  
}
```

argumento

Operaciones sobre los
objetos. Se crea el objeto y

```
mi_funcion(5)  
> 6
```

Devuelve el objeto y. Si no
lo especificas, devuelve el
último objeto creado

Presentación

Análisis de datos en R

```
mi_funcion <- function(x){  
  y <- x+1  
  return(y)  
}
```

```
mi_funcion <- function(x){  
  y <- x+1  
}
```

```
mi_funcion(5)  
> 6
```

Para cualquier función que sea un poco mas complicada que esto, usad el retorno explícito

Las funciones tienen su propio 'environment'

Presentación

Análisis de
datos en R

```
y = 5
mi función <- function(x){
  y <- x+1
  return(y)
}
y
> 5
mi función(5)
> 6
argumentos
y
```

Este objeto solo existe dentro de la función. No puede salir fuera.

Una función puede tener varios argumentos

Presentación

Análisis de
datos en R

```
mi_funcion <- function(x, d){  
  y <- x+d  
}  
  
mi_funcion(5, 1)  
#6
```

Si no indico lo contrario, el primer valor será siempre x y el segundo siempre será d. Si no le doy los valores, o le doy valores de más, lanzará error

Una función puede tener varios argumentos

Presentación

Análisis de
datos en R

```
mi_funcion <- function(x, d = 1){  
  y <- x^d  
}  
  
mi_funcion(5)  
>5
```

Si no especifico el
argumento d, este
será 1.

```
mi_funcion <- function(x, d){  
  y <- x^d  
}  
  
mi_funcion(d = 5, x = 2)  
>32
```

Si cambio el orden
debo especificar a
qué me refiero

Una función puede escribir de múltiples formas

Presentación

Análisis de
datos en R

```
mi_funcion <- function(x){  
  y <- x+1  
}
```

```
mi_funcion(5)  
> 6
```

```
mi_funcion <- function(x) y <- x+1
```

Sé inteligente a la hora de
ser conciso, úsallo con
cuatela

```
mi_funcion(5)  
> 6
```

Presentación

Análisis de
datos en R



Presentación

Análisis de datos en R



- Tidyverse es el grupo de paquetes originalmente creado por Hadley Wickham. Que llevan la filosofía “tidy”. <https://www.tidyverse.org/>
- Estos paquetes incluyen diversos aspectos necesarios para la ciencia de datos. Los dos más importantes son:
 - Gestión de datos: dplyr
 - Gráficos: ggplot2
- Podemos instalarlos de golpe con `install.packages('tidyverse')`

Presentación

Análisis de datos en R



- Conjunto de paquetes que incluyen herramientas para...

- Manejar datos dplyr
- Hacer gráficos ggplot2
- Manejar programación funcional purr
- Trabajar con factores forcats
- Trabajar con fechas lubridate
- Trabajar con cadenas de texto stringr

Install.packages("tidyverse"): pacman::p_load(tidyverse)

- La clave de usar tidyverse en encadenar funciones con la pipa.
- La pipa se invoca con
 $\% > %$
y en Rstudio puedes escribirla con
control + shift + m
- Nos permite ejecutar código de izquierda a derecha
- La idea es facilitar la lectura y poder copiarnos a nosotros mismos más fácilmente

```
## A single command
sqrt(83)

## Base R method of running more than one command
round(sqrt(83), digits = 2)

## Running more than one command with piping
sqrt(83) %>% round(digits = 2)
```

Presentación

Análisis de
datos en R

- El código de la derecha se lee más fácil
- Al vector a, le aplicamos la función suma.
- Al resultado de la suma aplicamos la raíz cuadrada
- “result” es el vector a sumado y al que se obtiene la raíz

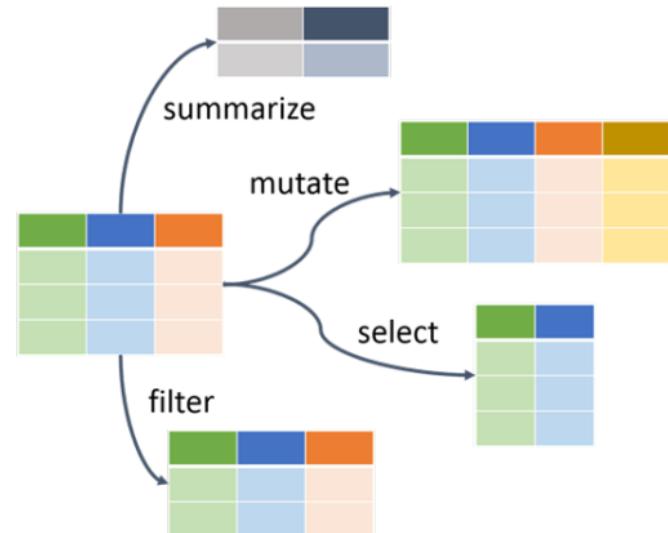
```
library(tidyverse)
a = c(1, 2, 3, 4)
a %>% sum()
a %>% sum() %>% sqrt()
result <- a %>% sum() %>% sqrt()
```

De esta manera yo puedo leer fácil y copiar-pegar mis operaciones fácil.

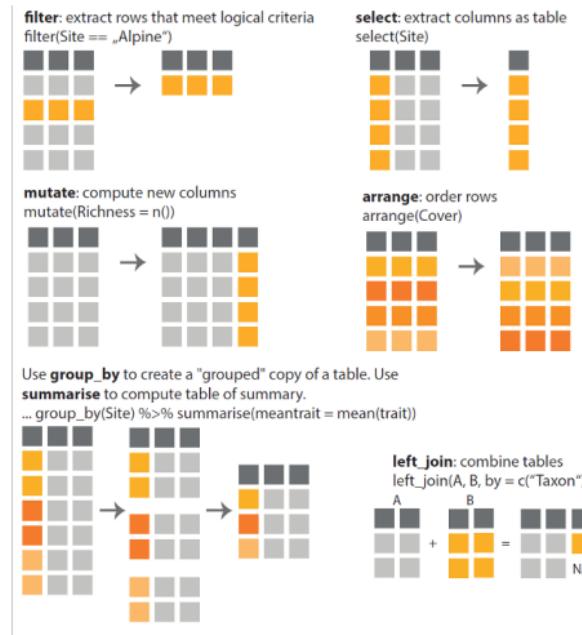
Presentación

Análisis de
datos en R





- `select` = quedarse con columnas
 - `arrange` = ordenar
 - `pivot_longer` = convertir los datos en largos*
 - `pivot_wider` = convertir los datos en anchos*
 - `left_join` = unir de dcha a izqda



Crear sistema de trabajo/ carpetas y subcarpetas

Presentación

Análisis de
datos en R

- Literatura:

- artículos, informes, noticias
- Diario del investigador

Crear sistema de trabajo/ carpetas y subcarpetas

Presentación

Análisis de
datos en R

- Literatura:
 - artículos, informes, noticias
 - Diario del investigador
- Data
 - Identificar la fuente
 - Descarga de codebooks y materiales relacionados
 - Logs, scripts, artículos, todo en misma carpeta o vinculado.
 - Rproject es una buena solución

Crear sistema de trabajo/ carpetas y subcarpetas

Presentación

Análisis de
datos en R

- Literatura:
 - artículos, informes, noticias
 - Diario del investigador
- Data
 - Identificar la fuente
 - Descarga de codebooks y materiales relacionados
 - Logs, scripts, artículos, todo en misma carpeta o vinculado.
 - Rproject es una buena solución
- Flujo de trabajo
 - scripts
 - markdown
 - datasets propios

Crear sistema de trabajo/ carpetas y subcarpetas

- Literatura:
 - artículos, informes, noticias
 - Diario del investigador
- Data
 - Identificar la fuente
 - Descarga de codebooks y materiales relacionados
 - Logs, scripts, artículos, todo en misma carpeta o vinculado.
 - Rproject es una buena solución
- Flujo de trabajo
 - scripts
 - markdown
 - datasets propios
- Output
 - Cuadros y figuras

Crear sistema de trabajo/ carpetas y subcarpetas

Presentación

Análisis de
datos en R

- Literatura:
 - artículos, informes, noticias
 - Diario del investigador
- Data
 - Identificar la fuente
 - Descarga de codebooks y materiales relacionados
 - Logs, scripts, artículos, todo en misma carpeta o vinculado.
 - Rproject es una buena solución
- Flujo de trabajo
 - scripts
 - markdown
 - datasets propios
- Output
 - Cuadros y figuras
- Notas
 - Notas de la literatura
 - ideas, esquemas, mapas mentales

Crear sistema de trabajo/ carpetas y subcarpetas

Presentación

Análisis de datos en R

- Literatura:
 - artículos, informes, noticias
 - Diario del investigador
- Data
 - Identificar la fuente
 - Descarga de codebooks y materiales relacionados
 - Logs, scripts, artículos, todo en misma carpeta o vinculado.
 - Rproject es una buena solución
- Flujo de trabajo
 - scripts
 - markdown
 - datasets propios
- Output
 - Cuadros y figuras
- Notas
 - Notas de la literatura
 - ideas, esquemas, mapas mentales
- Borradores
 - Un documento por sección
 - Todo el documento guardando versiones anteriores
 - Actualización y nombre de archivos
 - Contador de palabras

Crear sistema de trabajo/ carpetas y subcarpetas

- Literatura:
 - artículos, informes, noticias
 - Diario del investigador
- Data
 - Identificar la fuente
 - Descarga de codebooks y materiales relacionados
 - Logs, scripts, artículos, todo en misma carpeta o vinculado.
 - Rproject es una buena solución
- Flujo de trabajo
 - scripts
 - markdown
 - datasets propios
- Output
 - Cuadros y figuras
- Notas
 - Notas de la literatura
 - ideas, esquemas, mapas mentales
- Borradores
 - Un documento por sección
 - Todo el documento guardando versiones anteriores
 - Actualización y nombre de archivos
 - Contador de palabras
- Presentación (ppt, bmr, keynote)
 - Figures, graphs, logos, drafts

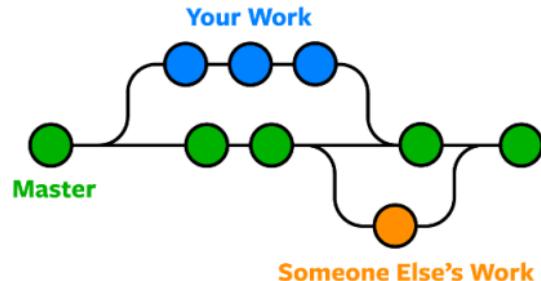
Naming

Renombrar todos los archivos para recordarlos:

- Tipo de documento (a: article, b:book, ds: dataset, fig: figure, tab:table)
- Incluir YYMMDD
- No espacios o caracteres
- Para papers utilizar gestor bibliográfico (Zotero, Mendeley, BibTex) o sistema de keywords en los títulos al menos.
- Crear propio codebook o leyendas

```
dir.create(paste0(getwd(),"/figures"),showWarnings = F)
dir.create(paste0(getwd(),"/processed-
data"),showWarnings = F)
dir.create(paste0(getwd(),"/raw-data"),showWarnings = F)
dir.create(paste0(getwd(),"/scripts"),showWarnings = F)
dir.create(paste0(getwd(),"/manuscript"),showWarnings = F)
```

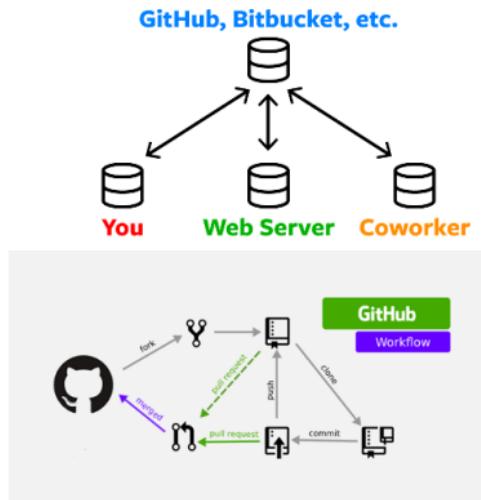
- Una forma de compartir código y proyectos es github
 - Git = protocolo de control de versiones
 - Hub = centro
 - Github es la plataforma para albergar, compartir, descargar y trabajar con proyectos que siguen los protocolos git creados para el trabajo colaborativo en entornos de software libre
 - <https://github.com/skills/introduction-to-github>



Presentación

Análisis de
datos en R

- Diccionario Git
- Repositorio: conjunto de archivos (puedes enviar nuevos archivos o modificaciones de los mismos o descargarlos)
- Branch: rama del trabajo en la que los usuario está trabajando y que puede enviar al “master”
- Master: rama establecida del repositorio
- Commit: cambio propuesto para ser añadido a la rama master
- Fork: desviación de la rama principal
- Pull: enviar tu repositorio al servidor
- Push: descargar un repositorio del servidor



Presentación

Análisis de datos en R

- Veamos un ejemplo con la colección de tutoriales de la asignatura
- <https://github.com/guidocor/R-enlaces-extra>



Presentación

Análisis de
datos en R

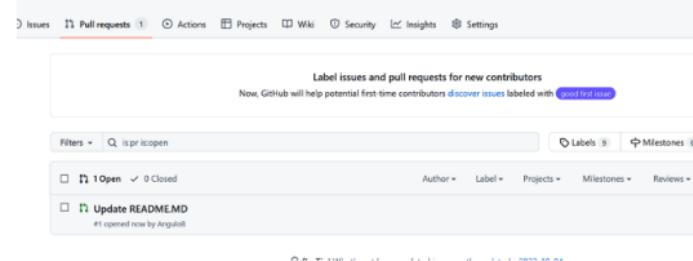
- Podemos clonar el proyecto, descargarlo, modificarlo y...
- O crear nuestro propio “fork” o mandar una sugerencia de mejora
- La mayoría de paquetes de R se desarrollan en GitHub (por ejemplo, tidyverse
<https://github.com/tidyverse>)



Presentación

Análisis de
datos en R

- Recibiendo sugerencias y cambios.
 - La usuaria AnguloB envía una sugerencia de cambio!



Presentación

Análisis de datos en R

- Como administradores de nuestro proyecto tenemos que decidir si el cambio nos parece adecuado

Update README.MD #1

Open AnguloB wants to merge 1 commit into [guidecom/main](#) from [AnguloB/patch-1](#)

Conversation 0 Commits 1 Checks 0 Files changed 1

AnguloB commented 1 minute ago First-time contributor

Añadir un espacio

Update README.MD Verified x790deb6

Add more commits by pushing to the [patch-1](#) branch on [AnguloB/R-enlaces-extra](#).

Require approval from specific reviewers before merging Branch protection rules ensure specific people approve pull requests before they're merged.

This branch has no conflicts with the base branch Merging can be performed automatically.

Merge pull request You also open this in GitHub Desktop or view command line instructions.

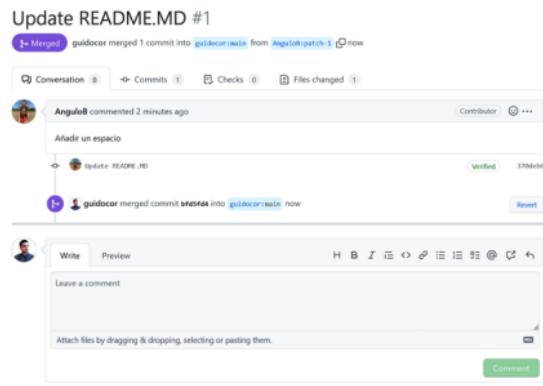
Write Preview

Leave a comment

Presentación

Análisis de datos en R

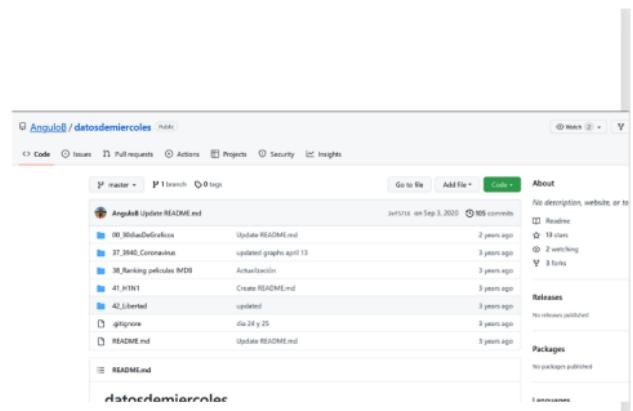
- Una vez aceptado el “pull request” nuestro repositorio ahora tiene una modificación.
- Como veis podemos establecer una conversación sobre los cambios (no todos los cambios te parecerán adecuados, algunos requerirán comentarlos, etc)



Presentación

Análisis de datos en R

- Ahora vamos a sugerir un cambio (pull request) de un repositorio. En este caso, queremos cambiar el archivo README.md de la usuario AnguloB en su repositorio datosdemiercoles



Presentación

Análisis de datos en R

- A editar!



Presentación

Análisis de datos en R

- Cada cambio va con su “commit” es decir la explicación de lo que llevaremos a cabo

You're making changes in a project you don't have write access to. Submitting a change will write it to a new branch in your fork guidcor/datosdemiercoles, so you can send a pull request.

datosdemiercoles / README.md in [Angular](#) · [master](#) · [Cancel changes](#)

[Edit file](#) [Preview](#)

```
1 # datosdemiercoles
2
3 ## [m_dias_de_graficos](https://github.com/cienciasdatos/datos-de-miercoles/blob/master/00_dias-de-graficos-2010.ad)
4
5 La sintaxis puede encontrarse en cada una de las carpetas para los días [m_01](https://github.com/guidcor/datosdemiercoles/tree/master/00_01diasDeGraficos).
6 A continuación un resumen de los gráficos
7 [!summary](https://github.com/guidcor/datosdemiercoles/blob/master/00_01diasDeGraficos/summary.png)
8
9
10 ## Sesión AI 2020-43-27 Libertad de prensa
11 Datos de [The World Bank](https://datacatalog.worldbank.org/indicators/HFB0001?country=88&Indicator=1241&RefYr=11&chartYrs=2001,2015) por datosdemiercoles.
12 [Reportes_sobre_la_liberdad_de_prensa](https://www.rvfe-es.org/).
13
14 La sintaxis puede encontrarse [ aquí](https://github.com/guidcor/datosdemiercoles/blob/master/42_1.libertad_de_prensa.R)
```

49 [!ranking200220](https://github.com/guidcor/datosdemiercoles/blob/master/78_Ranking200películas200208/780_ranking200peliculas200208/780_ranking200.jpg)
50
51
52

Attach files by dragging & dropping, selecting or pasting them.

[Propose changes](#)

Update README.md

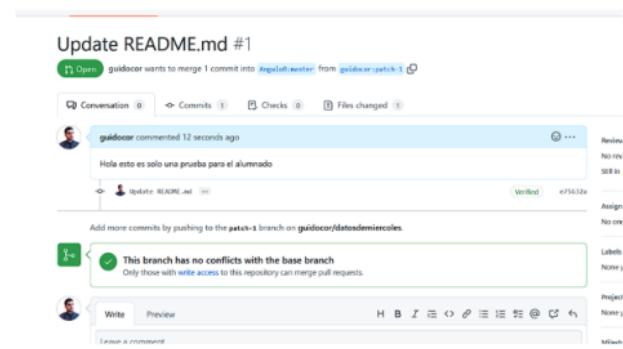
Hola esto es solo una prueba para el alumnado

[Propose changes](#) [Cancel](#)

Presentación

Análisis de datos en R

- Una vez hecho el pull request, tocará esperar que la usuaria acepte el cambio o no.
- En este caso, el cambio no será aceptado porque es una prueba



Presentación

Análisis de
datos en R

A crear vuestro propio perfil de github