



Estadística Empresarial R y R Studio.

Trabajo evaluable





Profesores responsables



Consoli Quintana

consolacion.quintana@uclm.es



Daniel Rojas

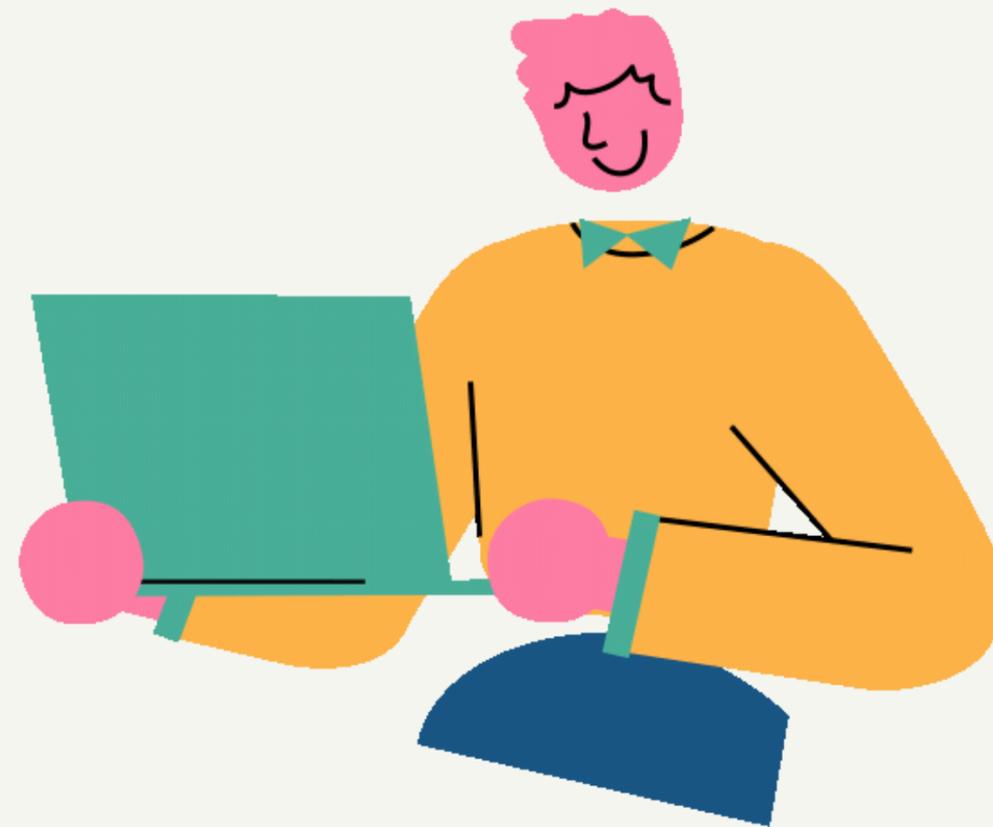
daniel.rojas@uclm.es



En qué consiste el trabajo

Pasos a seguir:

1. Asistir a las clases del día 06 y 07 de abril de 2022.
2. Instalar el software utilizado en clase: R y RStudio, para ello os aconsejamos ver los videos subidos a Campus Virtual.
3. Participar en el "foro de dudas trabajo RStudio" para realizar cualquier pregunta. Recordad que entre todos es más fácil solucionar los problemas.
4. Realizar un informe final reproducible mediante un documento R Markdown, aplicando las enseñanzas en estos dos días y explicándolo a través de los contenidos ya dados con anterioridad en clase.



En qué consiste el trabajo

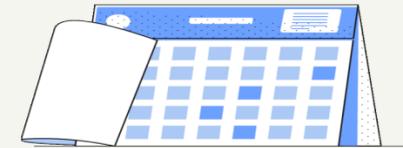
Puntuación:

- El trabajo puntuará el **30% de la nota final del examen**.
- Se requiere que el alumno obtenga **al menos un 4 en la prueba final de evaluación** para hacer media con la calificación obtenida en la actividad propuesta.
- Recordad que la fecha límite de entrega es el **24 de abril a las 23:59 horas**.





Clases 6 y 7 de abril



Conocer R y RStudio

Las bases de R.
Trabajar con proyectos de RStudio.
Informes reproducibles R Markdown.
Importación de datos de un fichero Excel.



Realizar el Informe

Qué es tidyverse.
Realizar tablas de frecuencias.
Gráficos. Histograma.
Mapa de calor.
Realizar una regresión.
Explicación.





Objetivos

Qué queremos lograr

Aplicar los conocimientos teórico-prácticos vistos en la asignatura de Estadística Empresarial.

Familiarizarse con programas informáticos estadísticos que podréis aplicar en vuestro futuro profesional.

Tomar conciencia de la importancia del análisis de los datos.





Comenzamos

Las
bases
de \mathbb{R}





Qué es R

Es un sistema de computación estadística. Un software de análisis de datos y un lenguaje de programación.

Ventajas:

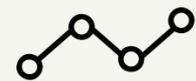
- Es libre y gratuito.
- Multiplataforma.
- Una amplia comunidad de usuarios proporciona recursos que lo alimentan.
- Muy usado en el ámbito de la investigación y profesional (empresas e instituciones).
- Es extensible: pudiendo usarse desde pequeñas funciones, hasta paquetes.
- Es reproducible.





Trabajamos con Scripts

Archivos de texto plano .R con expresiones de R. que al ejecutarlas se obtendrán los resultados para incorporar al informe



Con orden lógico de análisis

1. Cargar los paquetes a utilizar.
2. Importar los datos de un fichero.
3. Transformar, filtrar u organizar los datos.
4. Exploramos los datos.
5. Aplicamos métodos estadísticos.





Flujo de trabajo

Instalación software R e interfaz RStudio

Instalación mediante los videos
colgados en Campus Virtual

Instalación de paquetes

Trabajaremos con los paquetes:
tidyverse
agricolae
formattable

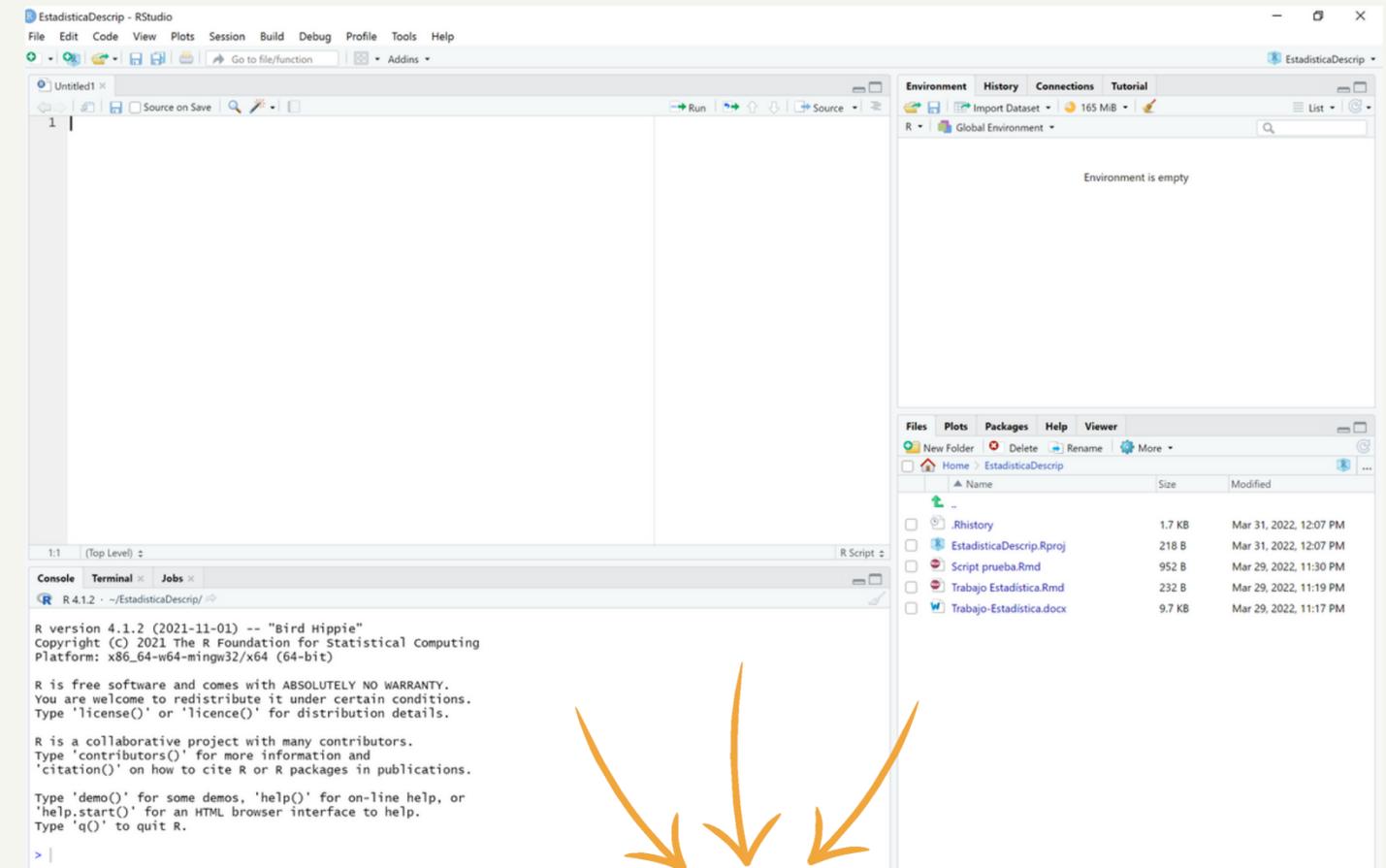
Crear un proyecto

Realizando un informe R
Markdown



Cómo se estructura R

Partes importantes de RStudio



Paneles

Izquierda:

Superior: **editor y visualizador de datos.**

Inferior: **consola.**

Derecha:

Superior: **sesión y herramientas de desarrollo.**

Inferior: **explorador de archivos, salida gráfica de R, paquetes, ayuda, visualizador.**



Llamada a funciones

Nombre de la función

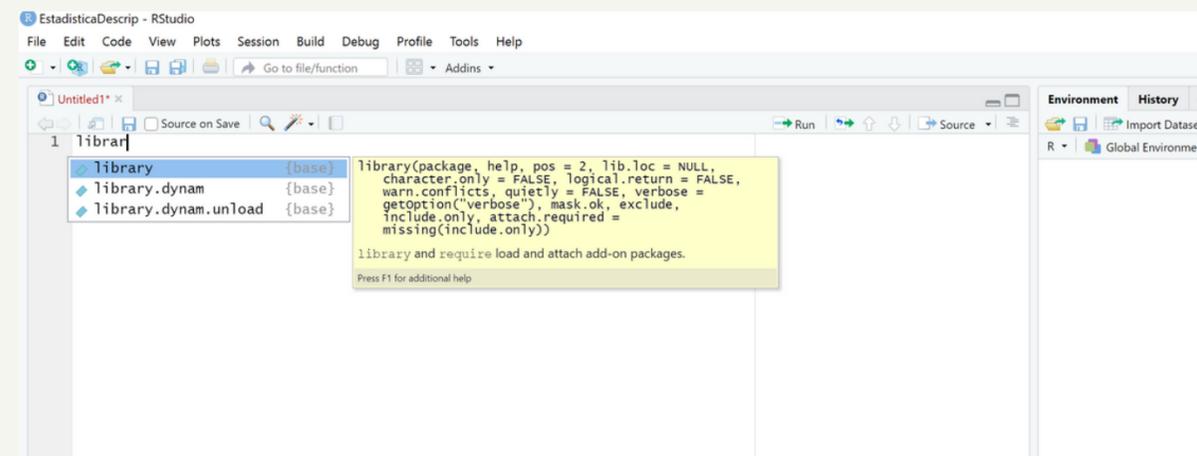
Escribimos el nombre de la función a utilizar, por ejemplo: *library* para cargar los paquetes.

Argumentos

Los argumentos de entrada van entre paréntesis, separados por comas.

Resultado

Puede producir un gráfico, un texto o un resultado "invisible".





Errores frecuentes

Escribir el nombre de la función de forma errónea.

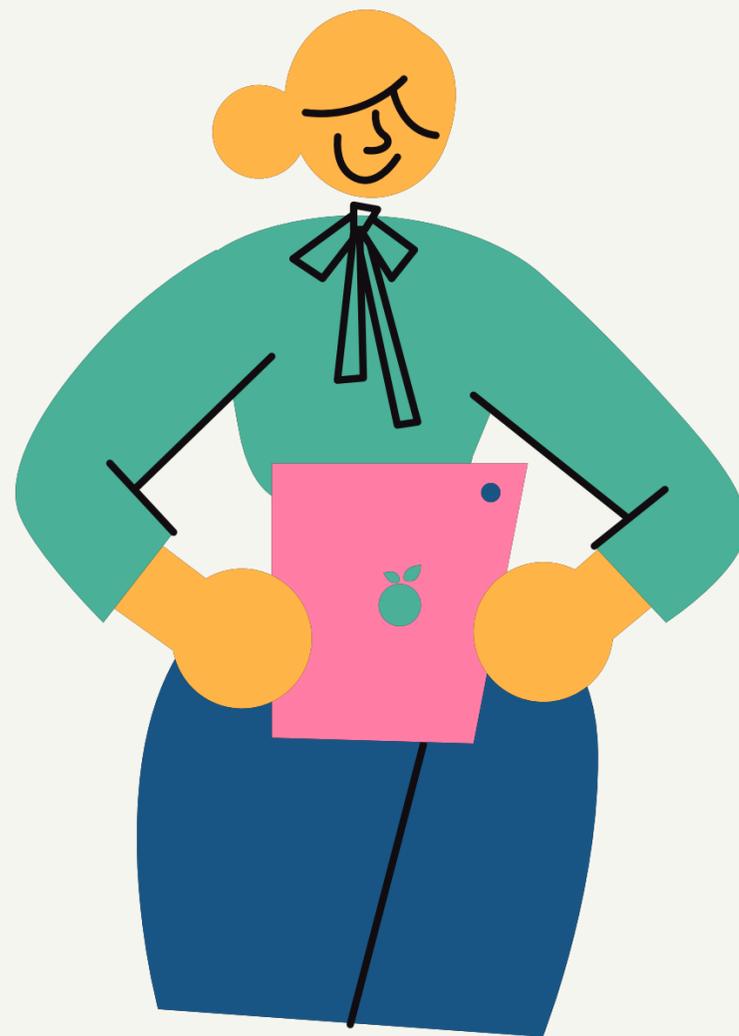
No cargar previamente los paquetes.

Sensible a mayúsculas. Los objetos "datos" y "Datos" por tanto, serán distintos.

13



Trabajamos
con
proyectos





RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

Console Terminal Jobs

```
R 4.1.2 · ~/\n\nR version 4.1.2 (2021-11-01) -- "Bird Hippie"\nCopyright (c) 2021 The R Foundation for Statistical Computing\nPlatform: x86_64-w64-mingw32/x64 (64-bit)\n\nR is free software and comes with ABSOLUTELY NO WARRANTY.\nYou are welcome to redistribute it under certain conditions.\nType 'license()' or 'licence()' for distribution details.\n\nR is a collaborative project with many contributors.\nType 'contributors()' for more information and\n'citation()' on how to cite R or R packages in publications.\n\nType 'demo()' for some demos, 'help()' for on-line help, or\n'help.start()' for an HTML browser interface to help.\nType 'q()' to quit R.\n\n> |
```

Environment History Connections Tutorial

Import Dataset 167 MiB List

R Global Environment

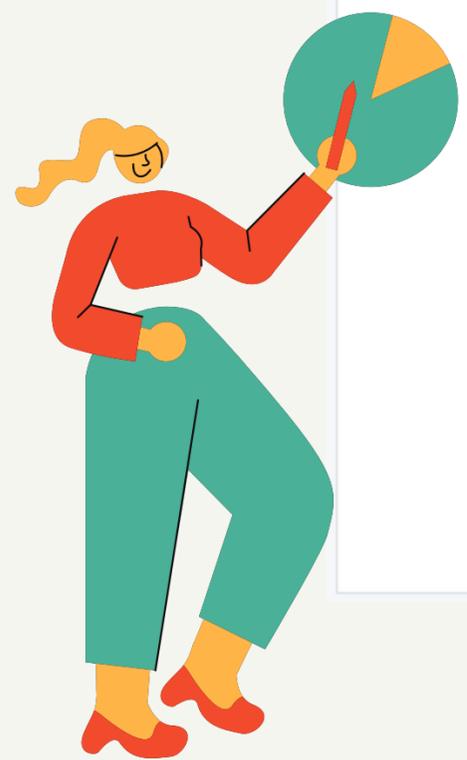
Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home

	Name	Size	Modified
<input type="checkbox"/>	.Rhistory	0 B	Jan 18, 2022, 1:11 PM
<input type="checkbox"/>	Blackmagic Design		
<input type="checkbox"/>	desktop.ini	418 B	Mar 29, 2022, 12:16 PM
<input type="checkbox"/>	EstadisticaDescrip		
<input type="checkbox"/>	Memoria-cientifico-tecnica-individual-2...	316.7 KB	Dec 7, 2021, 6:14 PM
<input type="checkbox"/>	Plantillas personalizadas de Office		
<input type="checkbox"/>	R		



R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Create Project

- New Directory**
Start a project in a brand new working directory
- Existing Directory**
Associate a project with an existing working directory
- Version Control**
Checkout a project from a version control repository

Cancel

Create New Project

Directory name:

Create project as subdirectory of:

Use renv with this project

Open in new session

Name	Size	Modified
.Rhistory	0 B	Jan 18, 2022, 1:11 PM
Blackmagic Design		
desktop.ini	418 B	Mar 29, 2022, 12:16 PM





Flujo de trabajo

Abimos el proyecto de
RStudio

Para realizar el trabajo de clase

Creamos archivos de
código

Script o
R Markdown

Analizamos

Regeneramos los datos,
analizamos y guardamos los
resultados





Informes
reproducibles
R Markdown



Por qué R Markdown

Reproducibilidad



R Markdown:

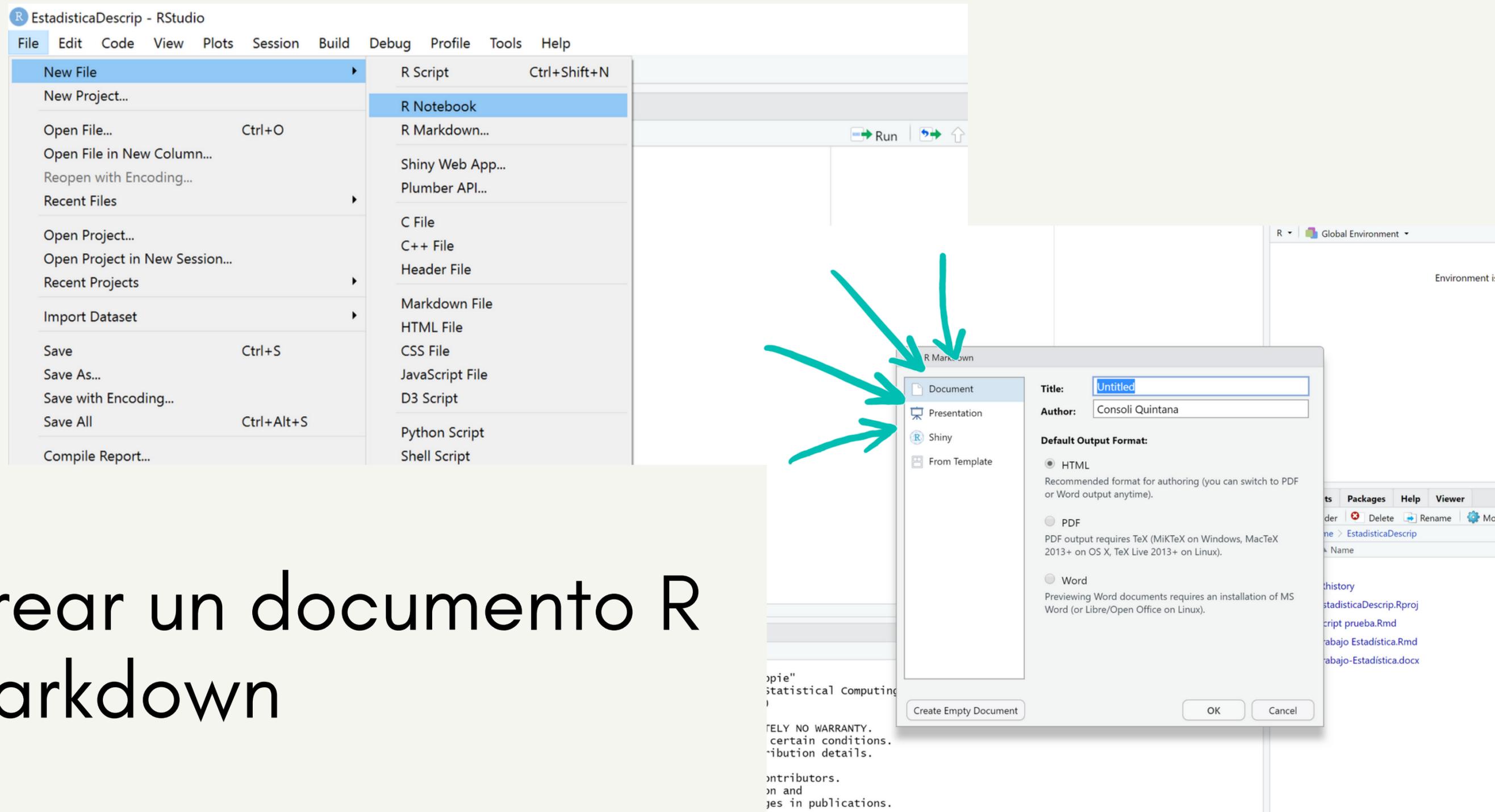
Formato versátil para crear informes reproducibles que permitan trazabilidad de los análisis.

Código:

Sintaxis muy sencilla.

El código se incluirá a través de bloques (*chunks*).





Crear un documento R Markdown





The screenshot illustrates the RStudio interface during the generation of an R Markdown report. The left pane shows the source code with orange arrows pointing to the Knit button and the Run button. The right pane shows the rendered HTML output with teal arrows pointing to the Knit button and the Run button.

Source Code (Left Pane):

```

1 ---
2 |title: "Trabajo R"
3 |author: "Consoli Quintana"
4 |date: "31/3/2022"
5 |output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting sy
15 documents. For more details on using R Markdown see <http://rmarko
16 when you click the **knit** button a document will be generated th
17 output of any embedded R code chunks within the document. You can
18
19 ```{r cars}
20 summary(cars)
21 ```
22
23 ## Including Plots
24 You can also embed plots, for example:
25
26 ```{r pressure, echo=FALSE}
27 plot(pressure)
28 ```

```

Rendered Output (Right Pane):

Trabajo R

Consoli Quintana
31/3/2022

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

##	speed	dist
##	Min. : 4.0	Min. : 2.00
##	1st Qu.:12.0	1st Qu.: 26.00
##	Median :15.0	Median : 36.00
##	Mean :15.4	Mean : 42.98
##	3rd Qu.:19.0	3rd Qu.: 56.00
##	Max. :25.0	Max. :120.00

Including Plots

You can also embed plots, for example:

Generar informe R Markdown



Formateado de texto

Markdown Quick Reference

Sintaxis más usada:

Encabezados con (##).

Negrita con ****palabra****

Formato texto código con backtick

(comillas inversas) `palabra`

Enlaces automáticamente copiados.

Cursiva con *_palabra_*

Enlaces con un texto más link: [texto](url)

Imágenes ![](ruta a la imagen)



Inclusión de código

Chunks

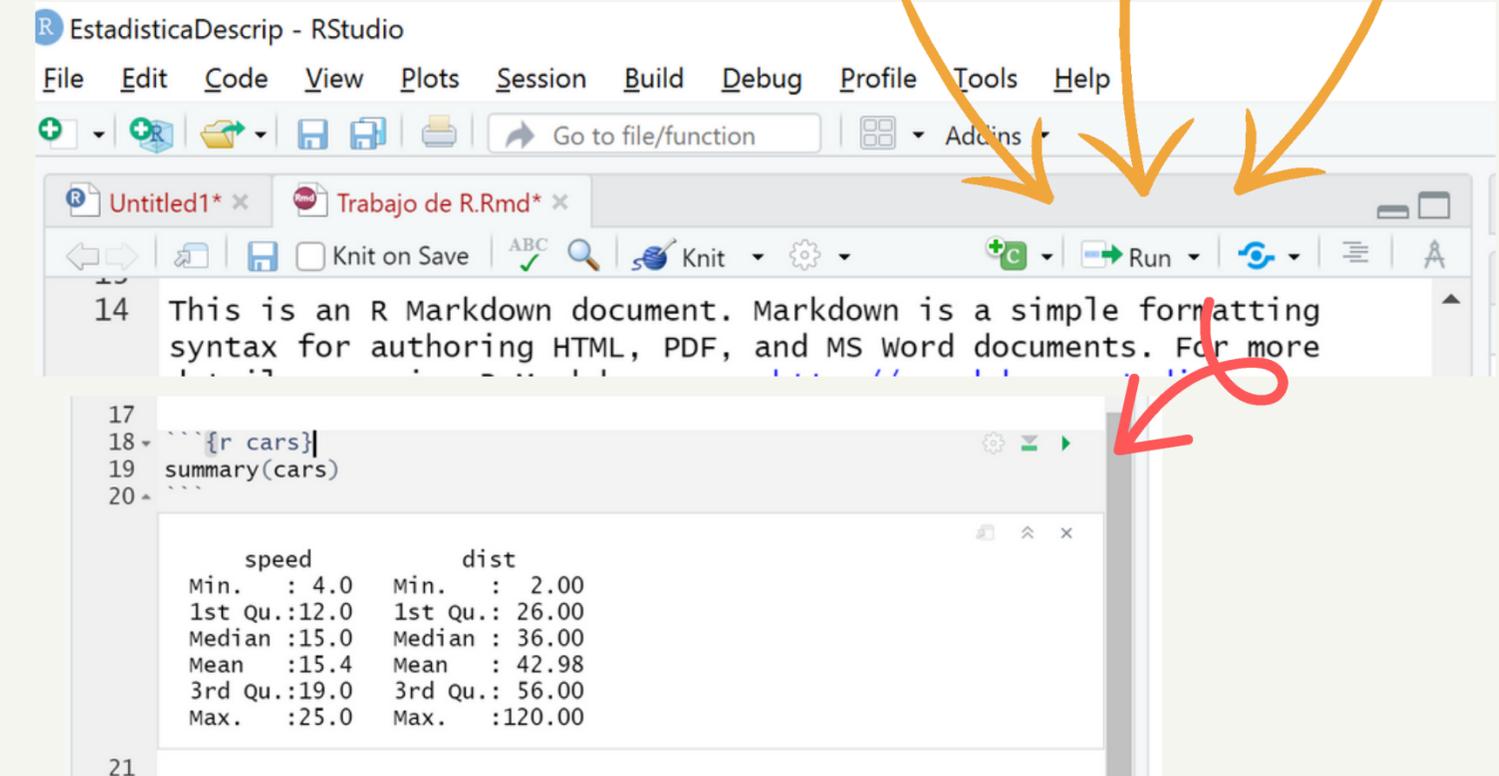
Marcación de bloques con opciones:

```
7  
8 {r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10 {r}  
11
```

Ejecutamos con:

CTRL+MAYUS+ENTER

También menú Run e iconos en chunk



The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu bar is a toolbar with icons for file operations and a search bar. The main editor window shows a code chunk with the following content:

```
14 This is an R Markdown document. Markdown is a simple formatting  
syntax for authoring HTML, PDF, and MS word documents. For more  
17 {r cars}  
18 summary(cars)  
19  
20
```

The output of the chunk is displayed in a separate window below the code editor:

speed	dist
Min. : 4.0	Min. : 2.00
1st Qu.:12.0	1st Qu.: 26.00
Median :15.0	Median : 36.00
Mean :15.4	Mean : 42.98
3rd Qu.:19.0	3rd Qu.: 56.00
Max. :25.0	Max. :120.00

Orange arrows point to the 'Run' button and the chunk options icon in the toolbar. A red arrow points to the 'Run' button in the chunk toolbar.

Importación de datos



Importando datos de Excel

Cuestiones previas



Datos bien organizados:

- No se pueden utilizar celdas combinadas.
- No dejar filas en blanco o con subtotales.
- No usar colores u otros formatos.
- No mezclar tipos de datos: si la columna es numérica, todas las celdas han de ser números o celdas en blanco.

Importar:

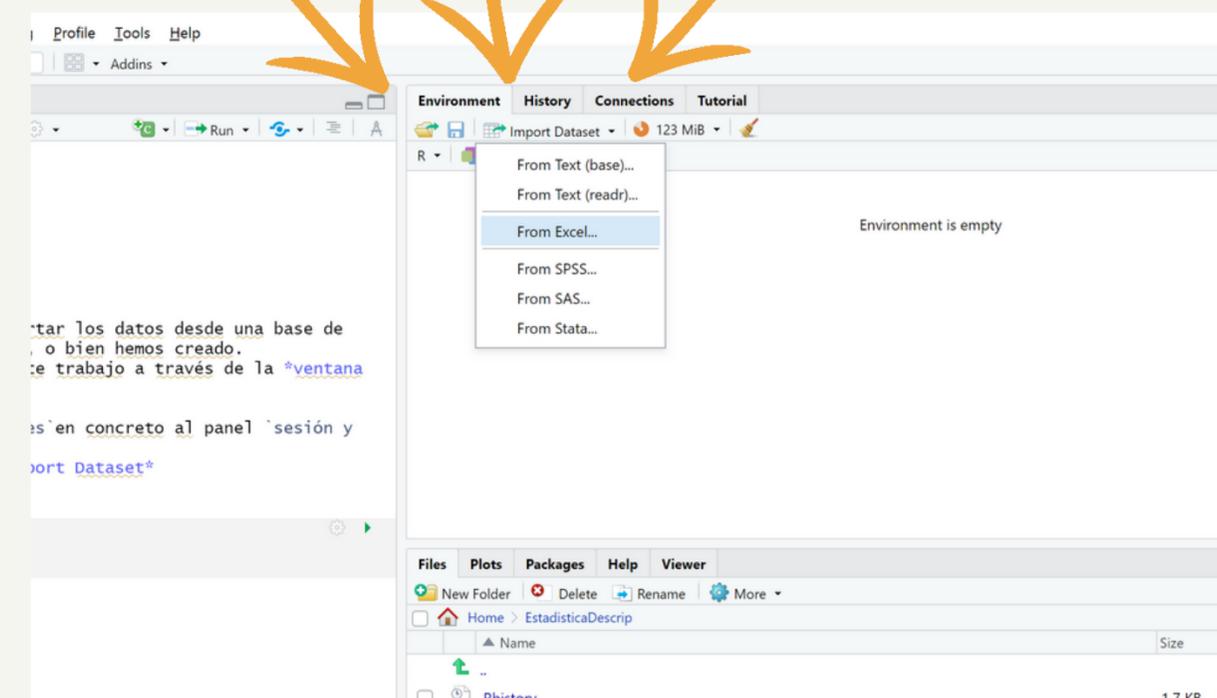
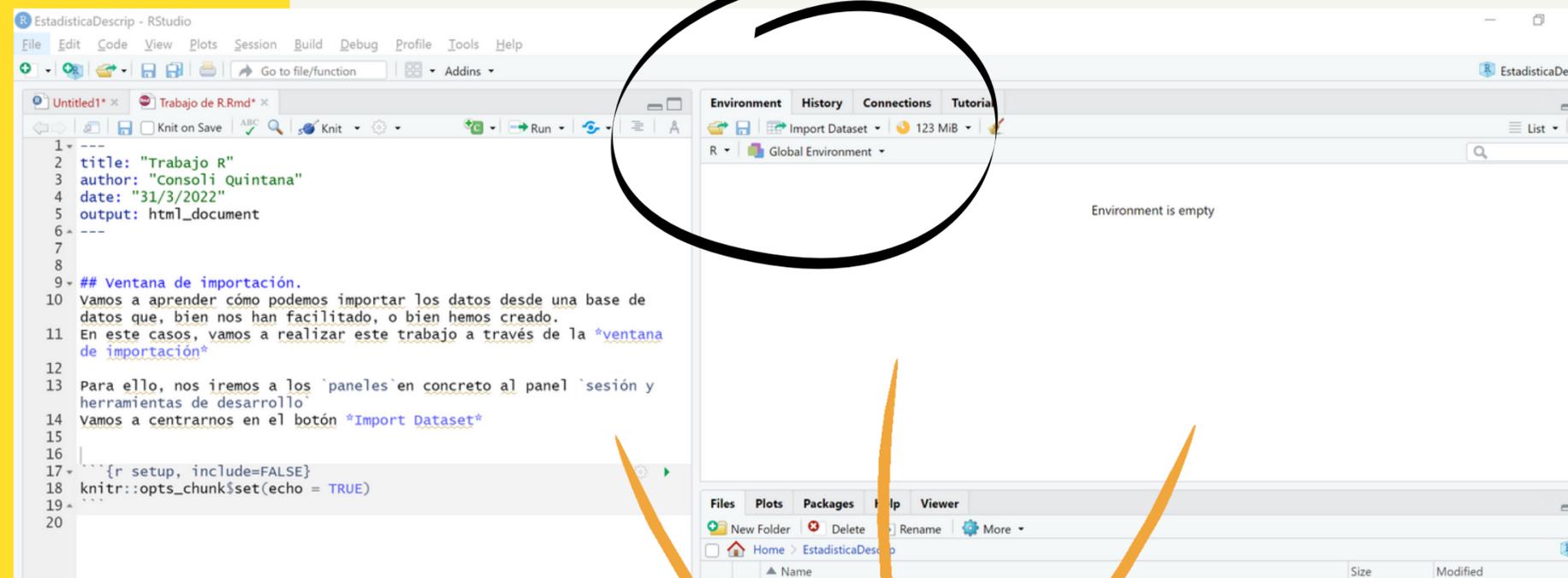
- Dos formas de importar la base de datos:
 1. A través de **download.file()**
 2. A través de la ventana de importación (será la que utilizaremos en este trabajo).





Ventana de importación

Importando datos de Excel





Flujo de trabajo

Panel superior derecho

Ventana de importación

Import Dataset

From Excel

Browse

Cargamos los datos.
Muy importante ser
organizados.



Trabajo de R.Rmd* x Environment History Connections Tutorial

Import Excel Data

File/URL: C:/Users/Consolacion.Quintana/OneDrive - Universidad de Castilla-La Mancha/Curso 2021-22/Estadística Empresarial/Trabajo 21-22/Base de datos.xlsx Browse...

Data Preview:

Periodo (double)	SH (double)	SM (double)	PIBNpc (double)	Deflactor_100 (double)
2009	25001.1	19502.0	23064.17	99.26571
2010	25479.7	19735.2	23030.92	99.41797
2011	25667.9	19767.6	22757.84	99.39834
2012	25682.1	19537.3	22044.72	99.28447
2013	25675.2	19514.6	21886.47	99.68006
2014	25727.2	19744.8	22206.08	99.45740
2015	25992.8	20051.6	23201.50	100.00000
2016	25924.4	20131.4	23961.76	100.32290
2017	26391.8	20607.9	24936.39	101.62665
2018	26738.2	21011.9	25711.90	102.89216
2019	26934.4	21682.0	26401.06	104.23455

Previewing first 50 entries.

Import Options:

Name: Max Rows: First Row as Names

Sheet: Skip: 0 Open Data Viewer

Range: NA:

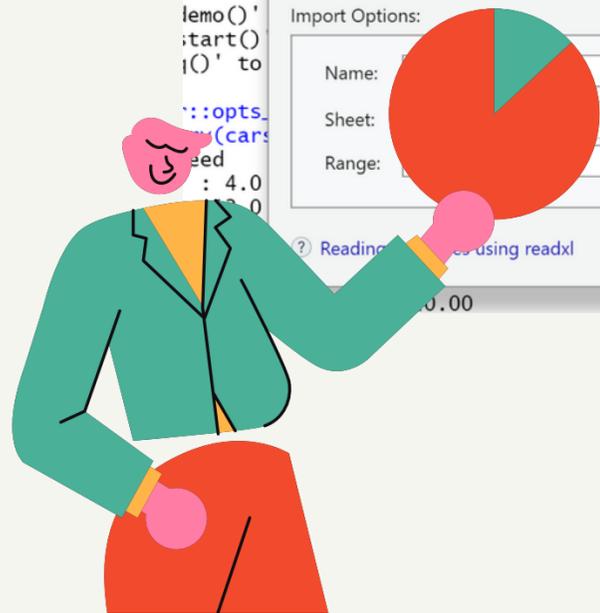
Code Preview:

```
library(readxl)
Base_de_datos <- read_excel("C:/Users/Consolacion.Quintana/OneDrive - Universidad de Castilla-La Mancha/Curso 2021-22/Estadística Empresarial/Trabajo 21-22/Base de datos.xlsx")
View(Base_de_datos)
```

Import Cancel

Tipo de datos:

Podemos modificar el tipo de datos de cada columna.



The screenshot shows the RStudio interface. The top-left pane displays a data table with 11 rows and 5 columns. The top-right pane shows the Environment tab with a data object named 'Base_de_datos' containing 11 observations of 5 variables. The bottom-left pane shows the R console with the following code and output:

```

> knitr::opts_chunk$set(echo = TRUE)
> summary(cars)
      speed      dist
Min.   : 4.0    Min.   : 2.00
1st Qu.:12.0    1st Qu.: 26.00
Median :15.0    Median : 36.00
Mean   :15.4    Mean   : 42.98
3rd Qu.:19.0    3rd Qu.: 56.00
Max.   :25.0    Max.   :120.00
> library(readxl)
> Base_de_datos <- read_excel("C:/Users/Consolacion.Quintana/OneDrive - Universidad de Castilla-La Mancha/Curso 2021-22/Estadística Empresarial/Trabajo 21-22/Base de datos.xlsx",
+   col_types = c("date", "numeric", "numeric",
+   "numeric", "numeric"))
There were 11 warnings (use warnings() to see them)
> View(Base_de_datos)

```

The bottom-right pane shows a file explorer view of the project directory, listing files such as .Rhistory, EstadísticaDescr.Rproj, Script prueba.Rmd, Trabajo de R.Rmd, Trabajo Estadística.Rmd, Trabajo-de-R.html, and Trabajo-Estadística.docx.

Resultado:

En el editor y visualizador de datos, tenemos la tabla con los datos que vamos a usar para trabajar lo visto durante las clases teórico prácticas de la asignatura Estadística empresarial.

En el espacio de trabajo o **Environment** tenemos en **Data** un objeto (tabla de datos) con 11 observaciones y 5 variables. — 29

This close-up screenshot shows the Environment pane in RStudio. Under the 'Data' section, the object 'Base_de_datos' is listed as having 11 observations of 5 variables. The structure of the object is displayed as follows:

```

$ Periodo      : POSIXct[1:11], format: "1905-07-01" ...
$ SH           : num [1:11] 25001 25480 25668 25682 25675 ...
$ SM           : num [1:11] 19502 19735 19768 19537 19515 ...
$ PIBNpc       : num [1:11] 23064 23031 22758 22045 21886 ...
$ Deflactor_100: num [1:11] 99.3 99.4 99.4 99.3 99.7 ...

```

Three orange arrows point from the text above to the object name, the observation count, and the variable list in the Environment pane.

Desde la nube

Importando datos de Excel

```
3- ```{r echo=TRUE}
9 Datos <- read_csv2("https://drive.google.com/uc?id=1oWRGYCcSvQUIVDe
) ^```  
yJuBjiupDx1ca5Ep0&export=download")  
1  
  
i Using ",'" as decimal and "'.'" as grouping mark. Use  
`read_delim()` for more control.  
Rows: 11 Columns: 5  
-- Column specification -----  
Delimiter: ";"  
dbl (5): Periodo, SH, SM, PIBNpc, Deflactor_100  
  
i Use `spec()` to retrieve the full column specification for this  
data.  
i Specify the column types or set `show_col_types = FALSE` to  
quiet this message.
```



Tidyverse



Qué es *tidyverse*

Cuestiones previas



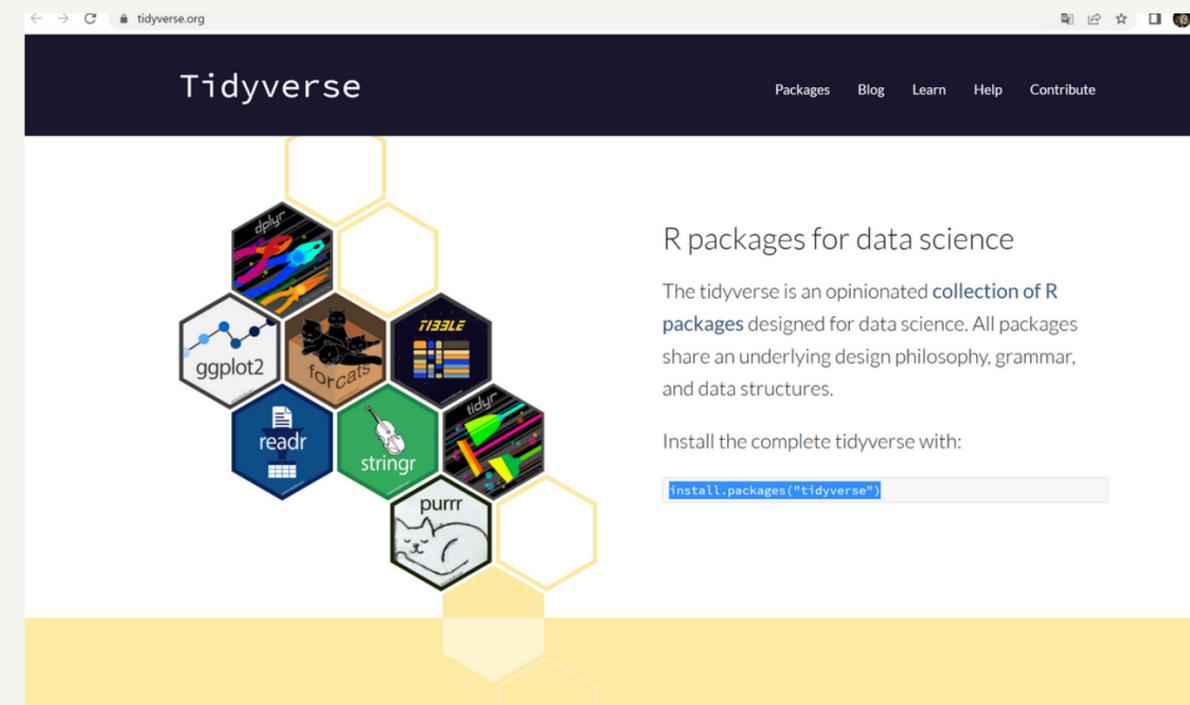
Definición:

- “an opinionated collection of R packages designed for data science”.
- Conjunto de paquetes de R opinables diseñados para ciencia de datos.

Ventajas:

1. Tiene un diseño común.
2. El flujo de trabajo es más intuitivo.
3. Es computacionalmente más eficiente.

En este enlace: <https://www.tidyverse.org/>
tenéis toda la información





Flujo de trabajo



Instalamos tidyverse

```
install.packages("tidyverse")
```

Run

```
library(tidyverse)
```

Pipe

Operador compuesto de 3 caracteres %>%

Atajo teclado: **CTR+MAYUS+M**

Para su uso: hay que cargar el paquete **dplyr**

El operador se pone en medio de las expresiones de R:

objeto1 %>% objeto2



Flujo de datos con "pipes"

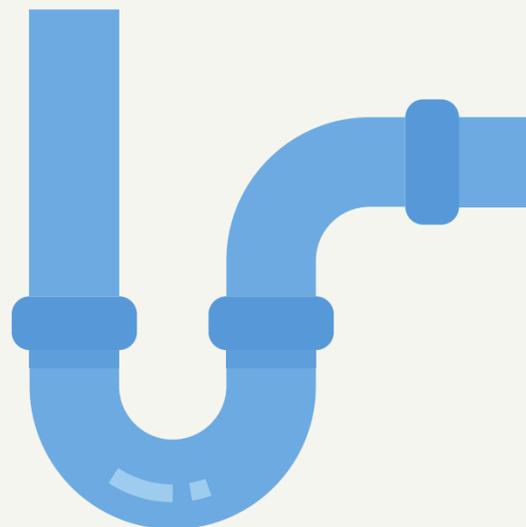
El objeto de la izquierda pasa como primer argumento a la función de la derecha

Los cálculos se hacen al final

Cuando el resultado se reutiliza, guardamos el objeto. Al principio del flujo con <- y si es al final, con ->

— 34

```
objeto1 %>%  
funcion1() %>%  
funcion2() %>%  
objeto2
```



Vamos a utilizar al principio del flujo:

```
nombre_objeto <-  
objeto1 %>%  
objeto2
```



Profesores responsables



Consoli Quintana

consolacion.quintana@uclm.es



Daniel Rojas

daniel.rojas@uclm.es