Instituto Nacional de Salud

R aplicado a epidemiología

Jorge Mario Estrada
Juan Camilo Rojas
Paola Cortés Molano

14-17 mayo 2024

Salud

Agenda

- Pretest
- Unidad 1. Iniciando con R base: conociendo la sintaxis
- Unidad 2. Administración de archivos, importación
- Unidad 3. Limpieza de datos I
- Unidad 4 Limpieza de datos II
- Unidad 5. Agrupación y tabla estadísticas
- Unidad 6. Transformaciones uniones
- Unidad 7. Visualización de datos General
- Unidad 8. Visualización de datos en epidemiología
- Presentación final
- Postest

Generalidades

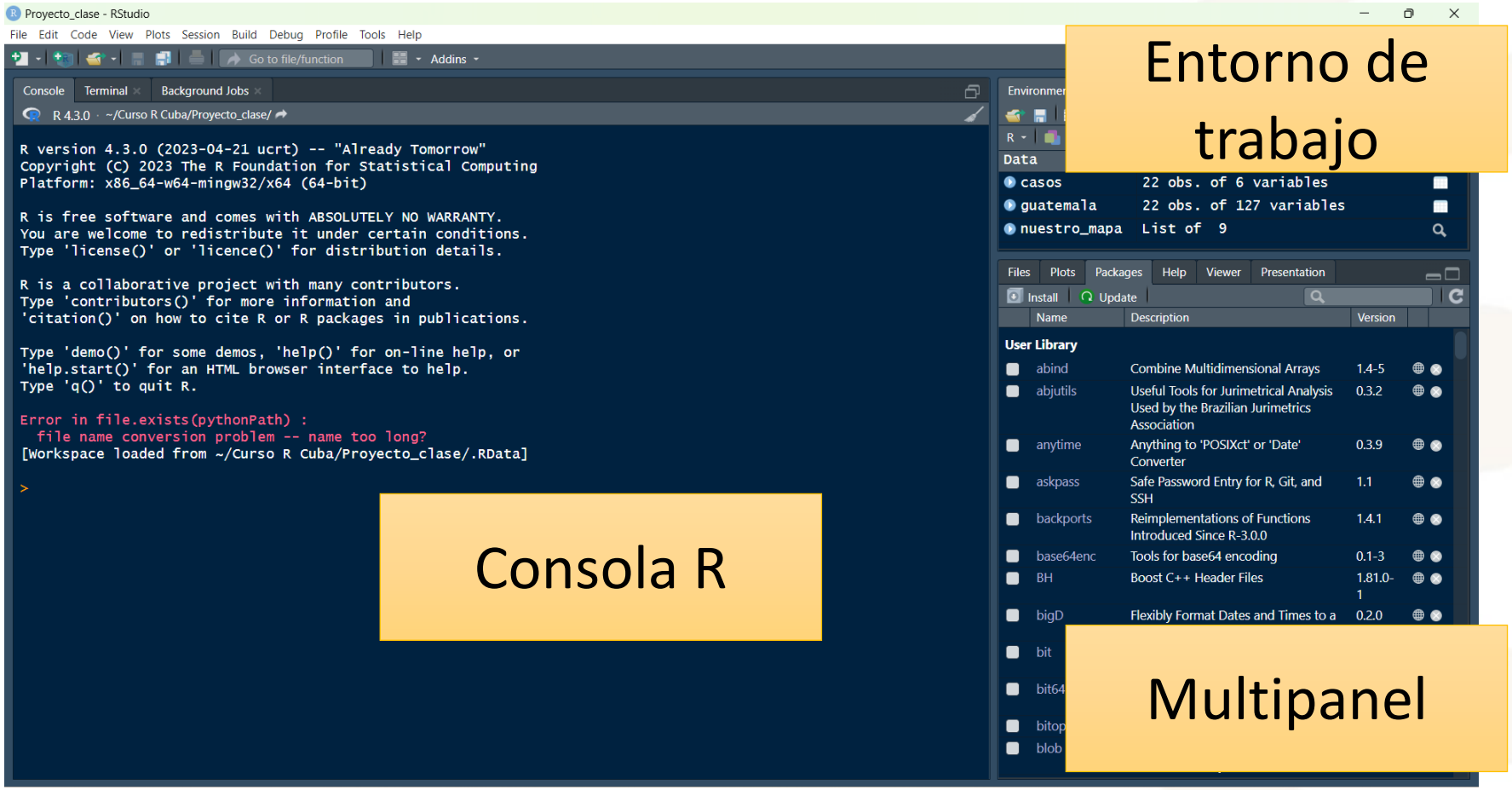
R

- Lenguaje de programación
- Acceso libre
- Orientado a la estadística
- Paquetes que permiten elaboración de visualizaciones y documentos técnicos

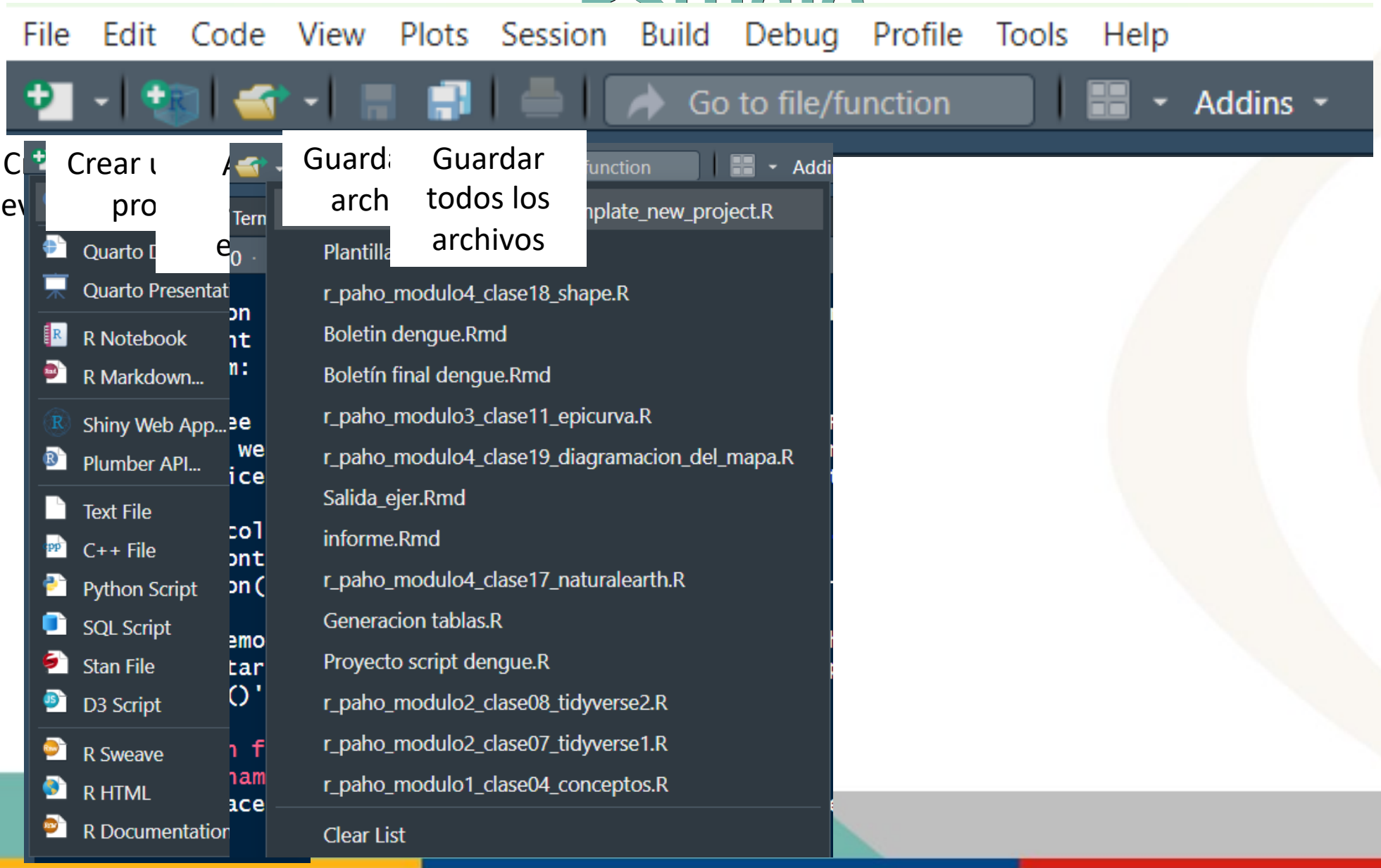
RStudio

"Entorno de Desarrollo Integrado"

Conjunto de herramientas creadas para facilitar el uso del lenguaje R



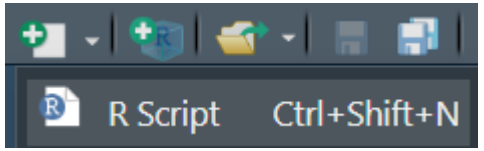
RStudio


File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Crear u
pro
Guarda
arch
Guardar todos los archivos
Quarto D
Quarto Presentat
R Notebook
R Markdown...
Shiny Web App...
Plumber API...
Text File
C++ File
Python Script
SQL Script
Stan File
D3 Script
R Sweave
R HTML
R Documentation
Plantilla
r_paho_modulo4_clase18_shape.R
Boletin dengue.Rmd
Boletín final dengue.Rmd
r_paho_modulo3_clase11_epicurva.R
r_paho_modulo4_clase19_diagramacion_del_mapa.R
Salida_ejer.Rmd
informe.Rmd
r_paho_modulo4_clase17_naturalearth.R
Generacion tablas.R
Proyecto script dengue.R
r_paho_modulo2_clase08_tidyverse2.R
r_paho_modulo2_clase07_tidyverse1.R
r_paho_modulo1_clase04_conceptos.R
Clear List

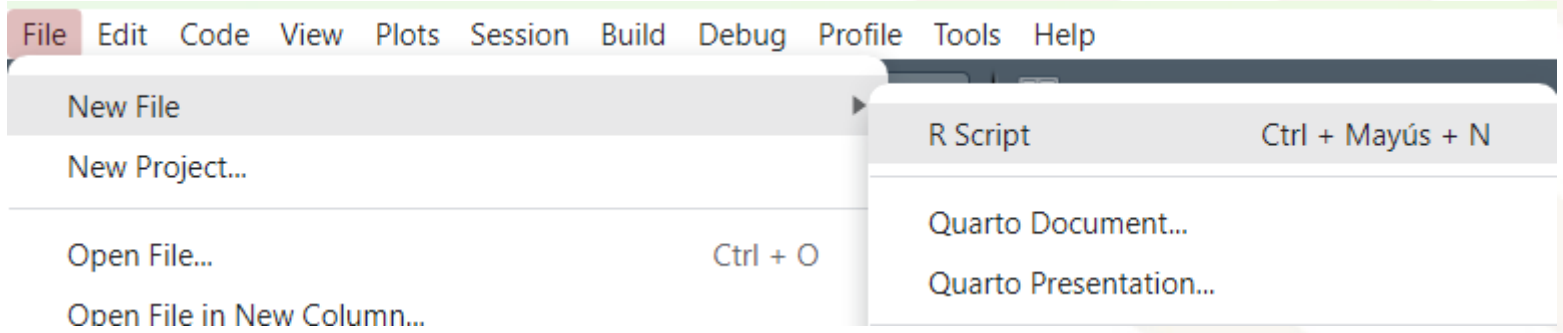
Scripts

- Documentos de texto donde se registran los códigos o instrucciones que necesitamos que R realice
- Estos documentos tienen extensión de archivo .R
- Para crear un script nuevo, se tienen dos formas:

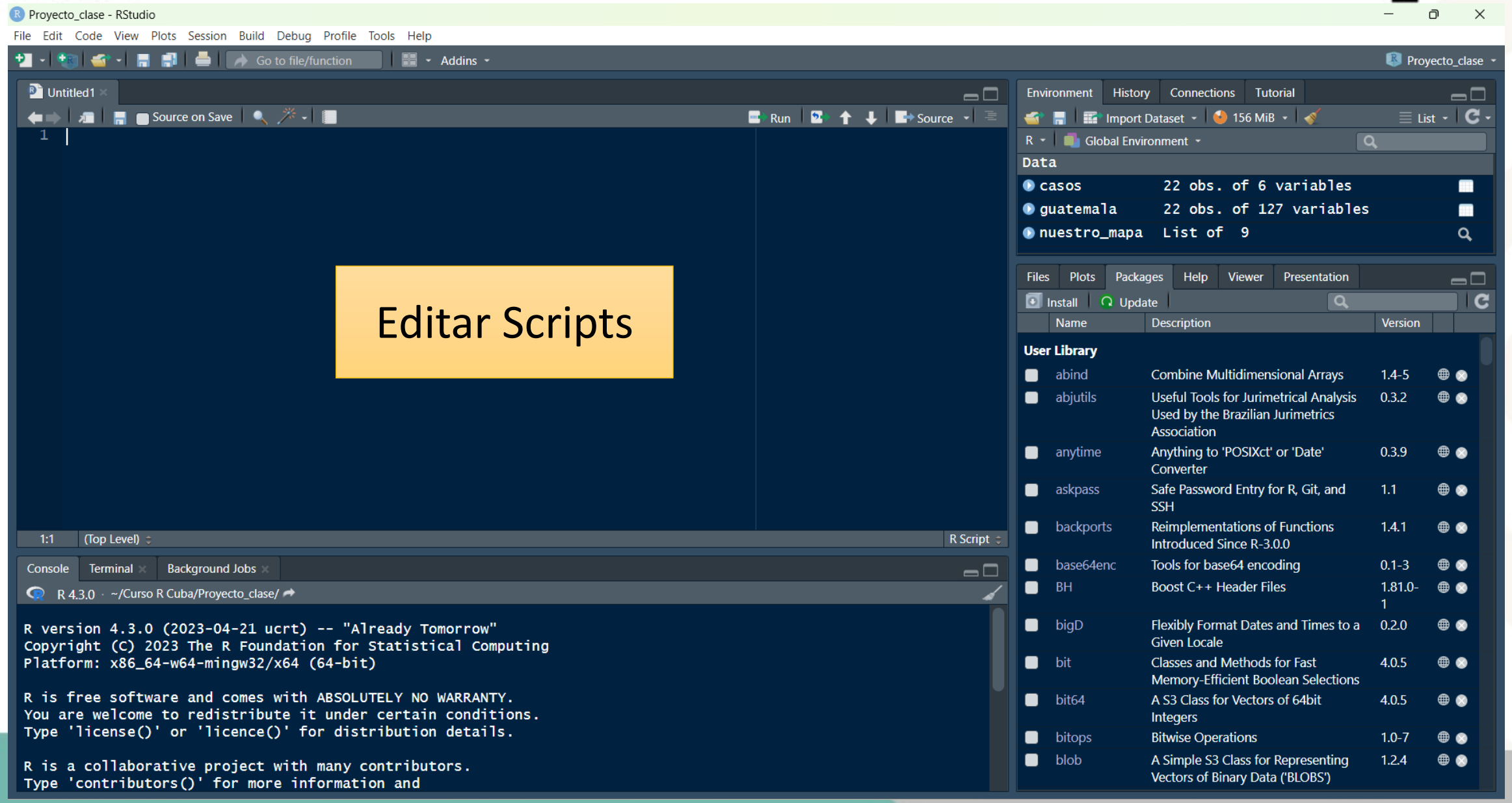
1



2

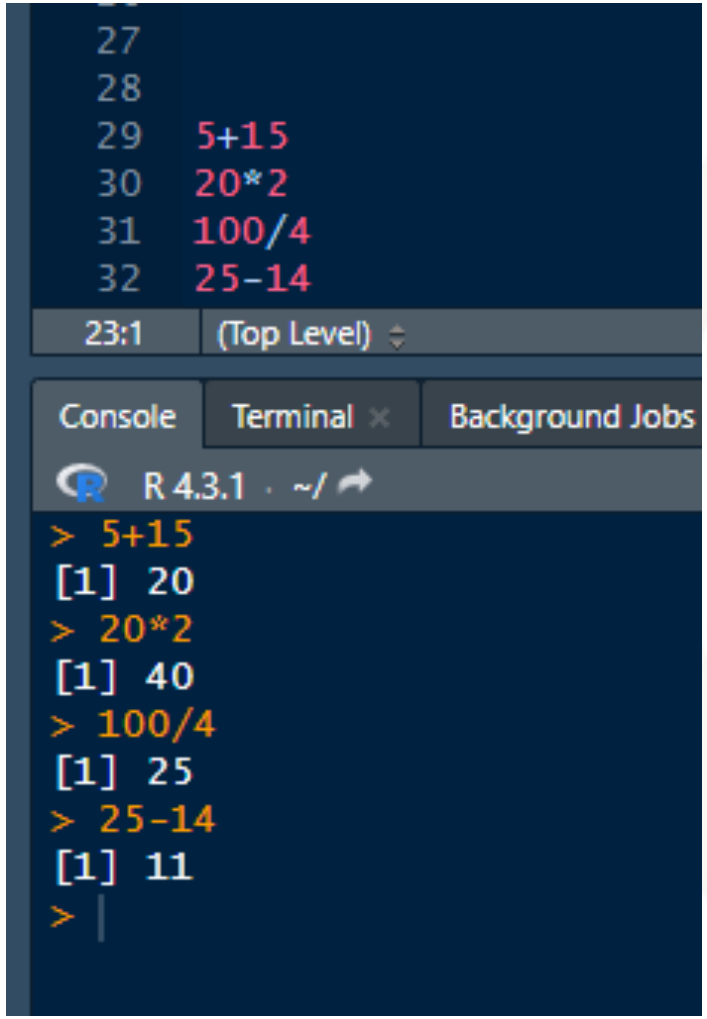


Scripts


Proyecto_clase - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Proyecto_clase
Untitled1
Source on Save
Run
Source
Editar Scripts
1:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.0 · ~/Curso R Cuba/Proyecto_clase/
R version 4.3.0 (2023-04-21 ucrt) -- "Already Tomorrow"
Copyright (C) 2023 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
Environment
History
Connections
Tutorial
Import Dataset
156 MiB
List
R
Global Environment
Data
casos 22 obs. of 6 variables
guatemala 22 obs. of 127 variables
nuestro_mapa List of 9
Files
Plots
Packages
Help
Viewer
Presentation
Install
Update
Name
Description
Version
User Library
abind Combine Multidimensional Arrays 1.4-5
abjutils Useful Tools for Jurimetrical Analysis Used by the Brazilian Jurimetrics Association 0.3.2
anytime Anything to 'POSIXct' or 'Date' Converter 0.3.9
askpass Safe Password Entry for R, Git, and SSH 1.1
backports Reimplementations of Functions Introduced Since R-3.0.0 1.4.1
base64enc Tools for base64 encoding 0.1-3
BH Boost C++ Header Files 1.81.0-1
bigD Flexibly Format Dates and Times to a Given Locale 0.2.0
bit Classes and Methods for Fast Memory-Efficient Boolean Selections 4.0.5
bit64 A S3 Class for Vectors of 64bit Integers 4.0.5
bitops Bitwise Operations 1.0-7
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.4

Ejercicio básico

- 5+15
- 20*2
- 100/4
- 25-14



Recomendaciones

- R diferencia entre minúsculas y mayúsculas

  grupo_edad es diferente a Grupo_edad

- Evite espacios en los nombres de objetos. Puede usar guiones o puntos.

  grupo_edad o grupo.edad

- Los paréntesis o corchetes siempre deben cerrarse
- Al usar “#” al inicio, se entiende como texto
- Explicar que se quiere obtener del código

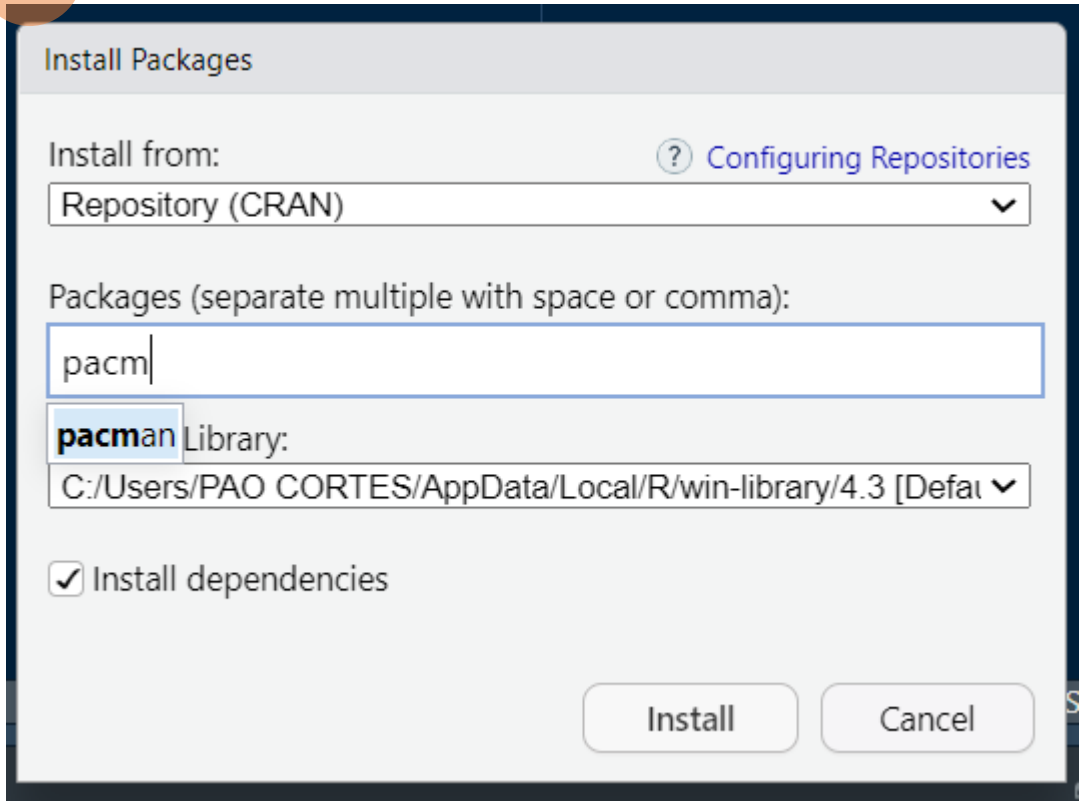
Paquetes en R

- Los paquetes son colecciones de códigos para realizar tareas específicas.
- Vienen acompañados de una descripción detallada de cómo usar los paquetes
- Muchos de ellos requieren ser instalados.

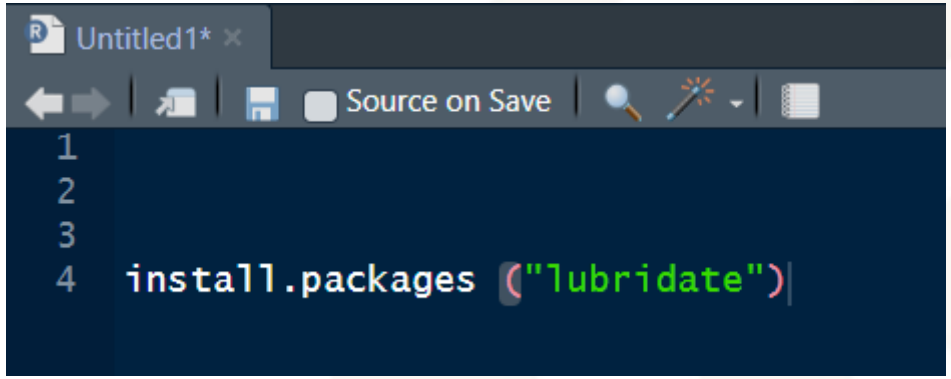
Los paquetes se instalan solo una vez!!!

Instalar paquetes

1



2



Usar un paquete instalado

1

Untitled1*

Source on Save

```
1
2
3
4 library('lubridate')
5 library("lubridate")
6
7
```

2

Files | Plots | Packages | Help | Viewer | Presentation

Install | Update | lu

	Name	Description	Version
☑	lubridate	Make Dealing with Dates a Little Easier	1.9.2
☐	cellranger	Translate Spreadsheet Cell Ranges to Rows and Columns	1.1.0
☐	conflicted	An Alternative Conflict Resolution Strategy	1.2.0
☐	evaluate	Parsing and Evaluation Tools that Provide More Details than the Default	0.21
☐	glue	Interpreted String Literals	1.6.2
☐	pillar	Coloured Formatting for Columns	1.9.0
☐	remotes	R Package Installation from Remote Repositories, Including 'GitHub'	2.4.2
☐	rnaturaleart...	High Resolution World Vector Map Data from Natural Earth used in rnaturalearth	0.2.1
☐	cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.1.4

Los paquetes se deben activar en cada sesión de trabajo!!!

Paquetes mas usados

- Pacman — Instalación y activación de paquetes
- Rio — Importación y exportación
- Tidyverse — Manipulación y tratamiento de datos
- Dplyr — Manipular conjuntos de datos
- Lubridate — Manipulación de fechas
- Maptools — Transformación de datos
- Ggplot2 — Visualización de datos
- Ggspatial — Manipulación de mapas
- Sf — Manipulación e importación de datos espaciales

Ayuda en R

- Información de una función y paquetes instalados
  help()
  ?

- Buscar términos/conceptos/ buscará las funciones y paquetes general
  ??

help("lubridate-package")

R: Dates and times made easy with lubridate Find in Topic

lubridate-package {lubridate} R Documentation

Dates and times made easy with lubridate

Description

Lubridate provides tools that make it easier to parse and manipulate dates. These tools are grouped below by common purpose. More information about each function can be found in its help documentation.

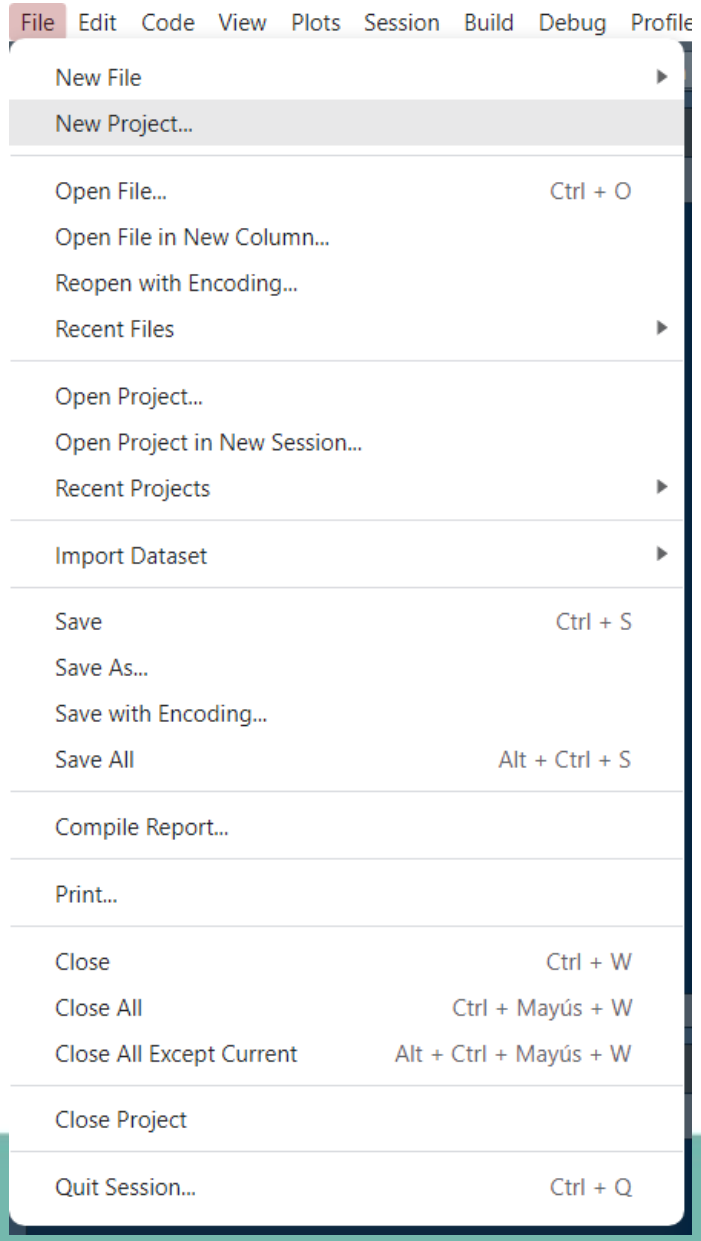
Parsing dates

Lubridate's parsing functions read strings into R as POSIXct date-time objects. Users should choose the function whose name models the order in which the year ('y'), month ('m') and day ('d') elements appear the string to be parsed: `dmy()`, `myd()`, `ymd()`, `ydm()`, `dym()`, `mdy()`, `ymd_hms()`). A very flexible and user friendly parser is provided by `parse_date_time()`.

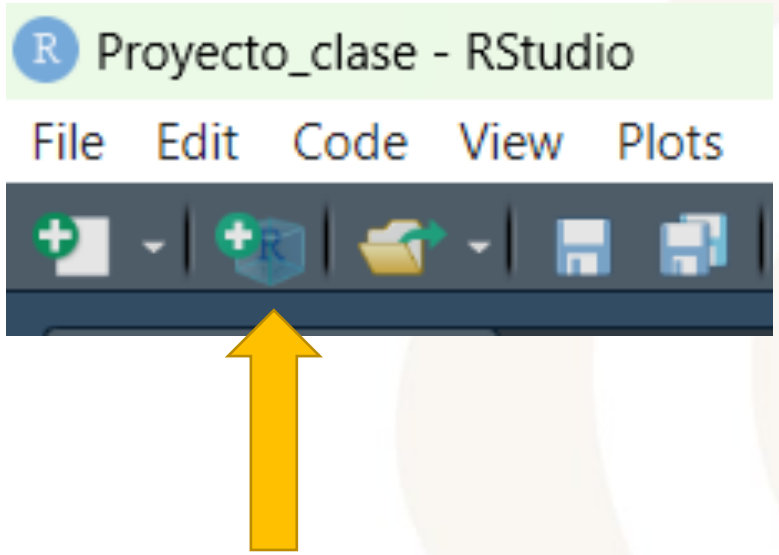
Proyectos en R

- Un proyecto en RStudio es una colección de documentos (datos, scripts, resultados, gráficos) organizados en una carpeta de trabajo.
- Para cada proyecto de análisis de datos, generación de reportes y/o visualización, es conveniente de crear un proyecto de Rstudio.
- El proyecto encapsula todo lo requerido para un proceso o análisis en particular, lo que facilita el poder transportarlo o compartirlo como un todo.

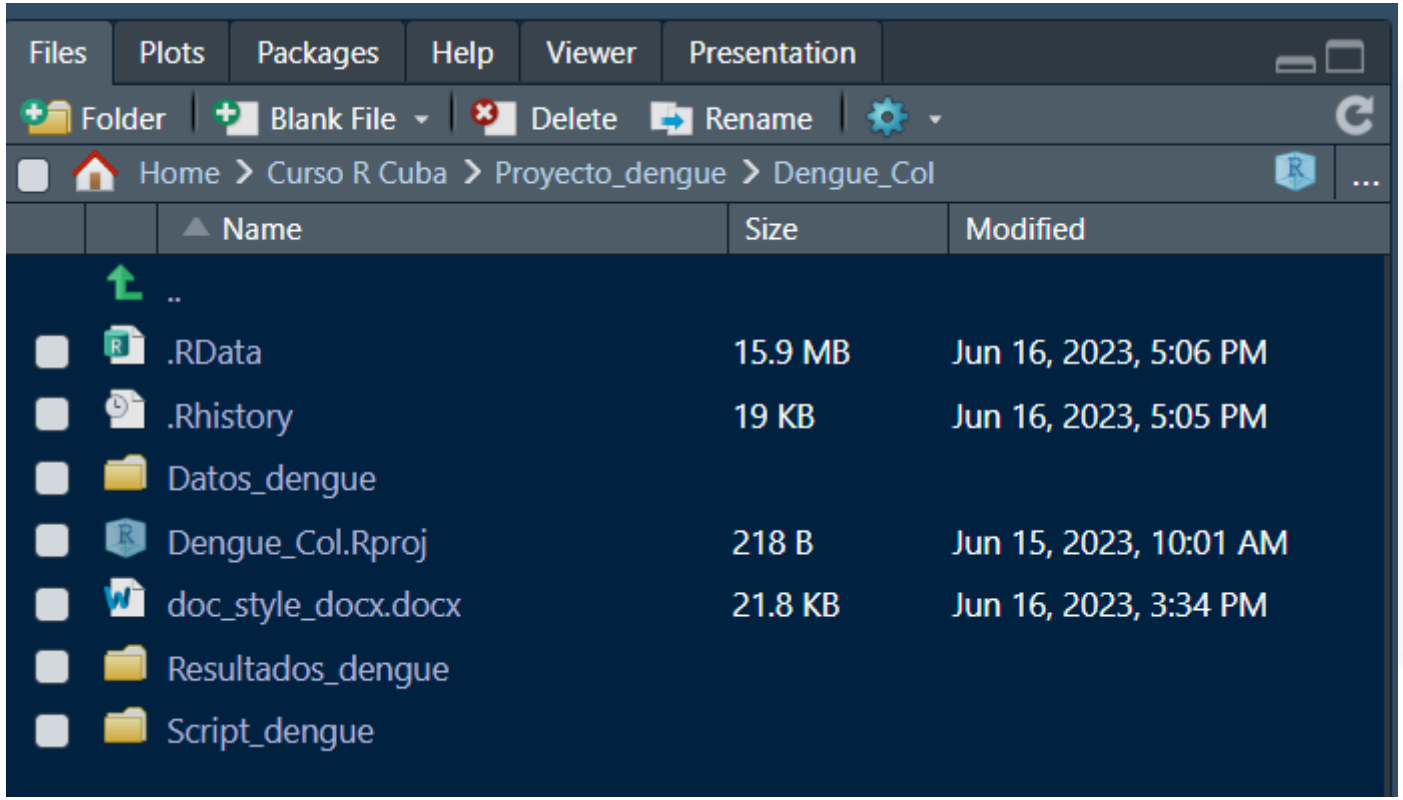
1


File Edit Code View Plots Session Build Debug Profile
New File
New Project...
Open File... Ctrl + O
Open File in New Column...
Reopen with Encoding...
Recent Files
Open Project...
Open Project in New Session...
Recent Projects
Import Dataset
Save Ctrl + S
Save As...
Save with Encoding...
Save All Alt + Ctrl + S
Compile Report...
Print...
Close Ctrl + W
Close All Ctrl + Mayús + W
Close All Except Current Alt + Ctrl + Mayús + W
Close Project
Quit Session... Ctrl + Q

2


Proyecto_clase - RStudio
File Edit Code View Plots

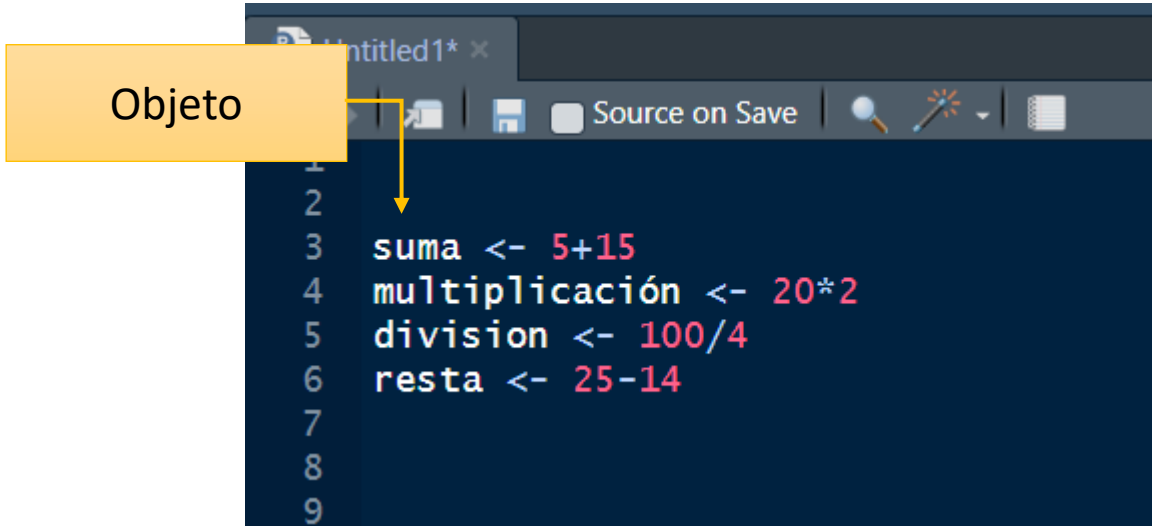
Organización de carpetas dentro de un proyecto



Operaciones

Objetos

- Todos los datos y estructura de datos en R son objetos, además estos deben de tener un nombre.
- Para crear objetos en R debe de usar el operador de asignación: <–



Tipo de datos

Tipo	En inglés	Ejemplo
Entero	integer	25 64 92
Decimal	numeric	2.5 6.4 9.2
Texto	character	"bienvenidos FETP al curso R"
Factor	factor	"femenino" "masculino"
Lógico	logical	TRUE FALSE
Fecha	date	"2024-05-14"

Función para conocer de qué tipo es un dato en R

- class()
- typeof()

Operadores

- Son símbolos que le indican a R que debe de realizar operaciones matemáticas o lógicas específicas. Estas operaciones son realizadas sobre objetos y valores.

- Operadores básicos son los siguientes:
  - Operadores aritméticos
  - Operadores relacionales
  - Operadores lógicos
  - Operadores de asignación

Operadores aritméticos

Operador	Operación	Ejemplo	Resultado
+	Suma	15 + 5	20
-	Resta	15 - 5	10
*	Multiplica	15 * 5	75
/	Divide	15 / 5	3
^ **	Potencia	15^5 15**5	759375

Operadores relacionales

Operador	Comparación	Ejemplo	Resultado
<	Menor que	15 < 5	FALSE
<=	Menor o igual que	15 <= 5	FALSE
>	Mayor que	15 > 5	TRUE
>=	Mayor o igual que	15 >= 5	TRUE
==	Exactamente igual que	15 == 5	FALSE
!=	Diferente que	15 != 5	TRUE

Operadores lógicos

Operador	Lógica	Ejemplo	Resultado
&	Y (AND)	x > 0 & y < 0	FALSE
|	O (OR)	x == 0 | y > 0	TRUE
!	NO	!(x >= 6 & x <= 20)	TRUE
isTRUE()	Es verdad	isTRUE(y==5)	TRUE
isFALSE()	Es falso	isFALSE(x==10)	FALSE

X= 15
Y= 5

Operadores de asignación

- Permite designar valores a objetos
- Luego de efectuar la operación de asignación, podemos emplear la variable para hacer operaciones con ella.

Operador	Operación
<-	Asigna un valor a un objeto

Funciones

- Funciones matemáticas
- Funciones estadísticas
- Funciones para textos
- Funciones que convierten un tipo de dato a otro
- Funciones del paquete lubridate que extraen componentes de una fecha

Funciones matemáticas

Función	Que realiza		Función	Que realiza
round()	redondeo		sqrt()	raíz cuadrada
ceiling()	redondeo arriba		exp()	exponencial
floor()	redondeo abajo		log()	logaritmo natural
trunc()	trunca la parte entera		log10()	log base 10
abs()	valor absoluto		log2()	log base 2

Funciones estadísticas

Función	Que realiza		Función	Que realiza
mean()	media		sum()	suma
median()	mediana		min()	valor mínimo
sd()	desviación estándar		max()	valor máximo
quantile()	cuantiles		range()	rango
IQR()	rango intercuartil		summary()	resumen descriptivo

Funciones para textos

Función	Que realiza
nchar()	Cuenta el número de caracteres
toupper()	Convierte a mayúsculas
tolower()	Convierte a minúsculas
substring()	Extrae una cadena de un texto
strsplit()	Separa un texto en función de un carácter
gsub()	Reemplaza una cadena de texto por otra
paste()	Une dos o más cadena de texto

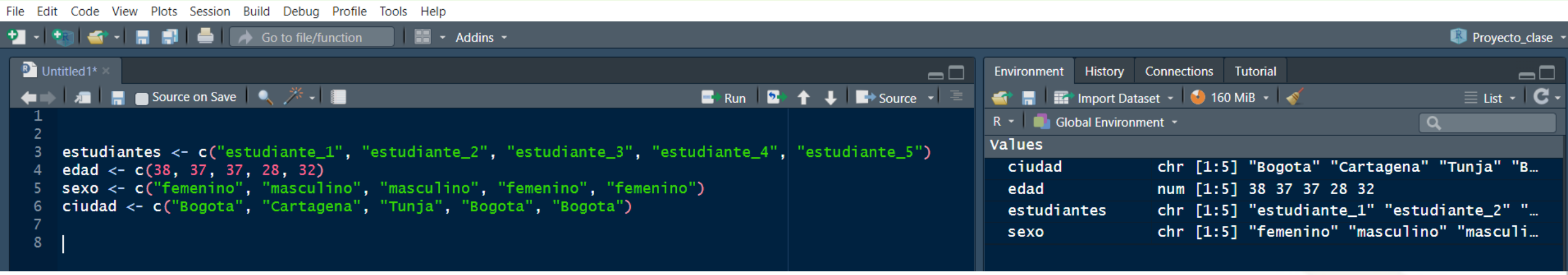
Funciones que convierten un tipo de dato a otro

Función	Tipo de dato al que se quiere convertir
as.integer()	Entero
as.numeric()	Numérico
as.character()	Texto
as.factor()	Factor
as.logical()	Lógico
as.Date()	Fecha

Vectores y Data frames

Creación de vectores

- Es una colección de uno o más datos del **mismo tipo**.
- función c()



Operación a vector

- edad * 2

```
8:1   (Top Level)                                                    R Script
Console   Terminal   Background Jobs
R 4.3.0 · ~/Curso R Cuba/Proyecto_clase/
> estudiantes <- c("estudiante_1", "estudiante_2", "estudiante_3", "estudiante_4", "estudiante_5")
> edad <- c(38, 37, 37, 28, 32)
> sexo <- c("femenino", "masculino", "masculino", "femenino", "femenino")
> ciudad <- c("Bogota", "Cartagena", "Tunja", "Bogota", "Bogota")
> edad * 2
[1] 76 74 74 56 64
>
```

Data frames

Los data frames se utilizan para almacenar una colección de vectores (columnas) de la misma longitud, pero posiblemente de diferentes tipos

```
Values
  ciudad        chr [1:5] "Bogota" "Cartagena" "Tunja" "Bogota" …
  edad          num [1:5] 38 37 37 28 32
  estudiantes   chr [1:5] "estudiante_1" "estudiante_2" "estudia…
  sexo          chr [1:5] "Femenino" "masculino" "masculino" "fe…
```

Si los vectores que usamos para construir el data frame no son del mismo largo, se generará un error.

Data frames

data.frame()

Untitled1*

```r
mi_df <- data.frame(

"estudiantes" = c("estudiante_1", "estudiante_2", "estudiante_3", "estudiante_4", "estudiante_5"),
"edad" = c(38, 37, 37, 28, 32),
"sexo" = c("femenino", "masculino", "masculino", "femenino", "femenino"),
"ciudad" = c("Bogota", "Cartagena", "Tunja", "Bogota", "Bogota")

)
```

Environment

```
Data
mi_df          5 obs. of 4 variables
Values
  ciudad       chr [1:5] "Bogota" "Cartagena" "Tunja" "B…
  edad         num [1:5] 38 37 37 28 32
  estudiantes  chr [1:5] "estudiante_1" "estudiante_2" "…
  sexo         chr [1:5] "femenino" "masculino" "masculi…
```

Data frames

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1* mi_df

Filter

	estudiantes	edad	sexo	ciudad
1	estudiante_1	38	femenino	Bogota
2	estudiante_2	37	masculino	Cartagena
3	estudiante_3	37	masculino	Tunja
4	estudiante_4	28	femenino	Bogota
5	estudiante_5	32	femenino	Bogota

COLOMBIA
POTENCIA DE LA
VIDA

¡Gracias!

#OrgullosamenteINS

@INSColombia

@insaludcolombia

Instituto Nacional de Salud de Colombia

Avenida Calle 26 # 51 - 20 / Bogotá, Colombia • PBX: (601) 220 77 00 exts. 1101 - 1214

INSTITUTO
NACIONAL DE
SALUD