

 SAN JUAN BAUTISTA



UNIVERSIDAD PRIVADA  
**SAN JUAN BAUTISTA**

SIS Y MÉTODOS ESTADÍSTICOS

# Práctica calificada 1-1

Docente:

- Castro Lopez, Segundo Vicente

Estudiantes:

- Cardenas Sanchez, Alai Alexis
- Cholán Sánchez, Ariana Maricielo
- Perez Flores, Andrea Ximena
- Ortiz Alvarez, María Fernanda
- Pinto Coaquira, Ines

## 1 Descargamos el dataset asignado "AUTISMO"

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	Puntuacion_P1	Puntuacion_P2	Puntuacion_P3	Puntuacion_P4	Puntuacion_P5	Puntuacion_P6	Puntuacion_P7	Puntuacion_P8	Puntuacion_P9		
2	1,1,1,0,0,1,1,0,1,0,0,6	Masculino	Otros	No	No	Jordan	No	5	4-11 years	Padre/Madre	No	
3	2,1,1,0,0,1,1,0,1,0,0,6	Masculino	Oriente Medio	No	No	Jordan	No	5	4-11 years	Padre/Madre	No	
4	1,1,1,0,0,0,1,1,0,0,6	Masculino	Desconocido	No	No	Jordan	Si	5	4-11 years	Desconocido	No	
5	4,0,1,0,0,1,1,0,0,0,1,5	Femenino	Desconocido	Si	No	Jordan	No	4	4-11 years	Desconocido	No	
6	5,1,1,1,1,1,1,1,1,1,5	Masculino	Otros	Si	No	Estados Unidos	No	10	4-11 years	Padre/Madre	Si	
7	6,0,0,1,0,1,1,0,1,0,1,4	Masculino	Desconocido	No	Si	Egypt	No	5	4-11 years	Desconocido	No	
8	7,1,0,1,1,1,1,0,1,0,1,5	Masculino	Blanco-Europeo	No	No	Reino Unido	No	7	4-11 years	Padre/Madre	Si	
9	8,1,1,1,1,1,1,1,1,0,0,5	Femenino	Oriente Medio	No	No	Baréin	No	8	4-11 years	Padre/Madre	Si	
10	9,1,1,1,1,1,1,1,0,0,0,11	Femenino	Oriente Medio	No	No	Baréin	No	7	4-11 years	Padre/Madre	Si	
11	10,0,0,1,1,1,0,1,1,0,0,11	Femenino	Desconocido	No	Si	Austria	No	5	4-11 years	Desconocido	No	
12	11,1,0,0,0,1,1,1,1,1,1,10	Masculino	Blanco-Europeo	Si	No	Reino Unido	No	7	4-11 years	Autocompletado	Si	
13	12,0,1,0,0,1,0,0,0,0,1,5	Femenino	Desconocido	No	No	Kuwait	No	3	4-11 years	Desconocido	No	
14	13,0,1,1,1,1,1,1,1,1,1,4	Masculino	Blanco-Europeo	Si	No	Estados Unidos	No	9	4-11 years	Padre/Madre	Si	
15	14,1,0,0,0,0,0,1,0,0,0,4	Femenino	Negro	No	No	Emiratos Árabes Unidos	No	2	4-11 years	Padre/Madre	No	
16	15,1,1,1,1,1,1,1,1,1,1,6	Masculino	Blanco-Europeo	No	No	Europe	No	10	4-11 years	Padre/Madre	Si	
17	16,1,1,1,1,1,1,1,1,1,1,8	Masculino	Blanco-Europeo	No	No	Malta	No	10	4-11 years	Padre/Madre	Si	
18	17,1,1,1,1,1,1,0,1,1,1,4	Masculino	Asiático del Sur	No	No	Bulgaria	No	9	4-11 years	Padre/Madre	Si	
19	18,0,0,0,0,0,0,1,0,0,0,7	Masculino	Otros	No	No	Estados Unidos	No	1	4-11 years	Padre/Madre	No	
20	19,1,0,1,1,1,0,1,1,1,1,11	Masculino	Blanco-Europeo	No	Si	Estados Unidos	No	8	4-11 years	Padre/Madre	Si	
21	20,1,1,1,1,1,0,1,0,1,5	Masculino	Desconocido	No	No	Egypt	No	8	4-11 years	Desconocido	Si	
22	21,1,1,1,1,1,1,0,1,0,5	Masculino	Blanco-Europeo	Si	No	Sudáfrica	No	8	4-11 years	Padre/Madre	Si	

## 2 Instalamos los paquetes

Damos click

```
{r}
install.packages("tidyverse")
install.packages("rio")
install.packages("here")
install.packages("janitor")
install.packages("skimmer")
install.packages("visdat")
install.packages("dplyr")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>  
Installing package into 'C:/Users/&Usuario HP/AppData/Local/R/win-library/4.4'

Chunk 1

Console

R 4.4.3 · C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB\_PRACTICA\_RSTUDIO/semana\_03/

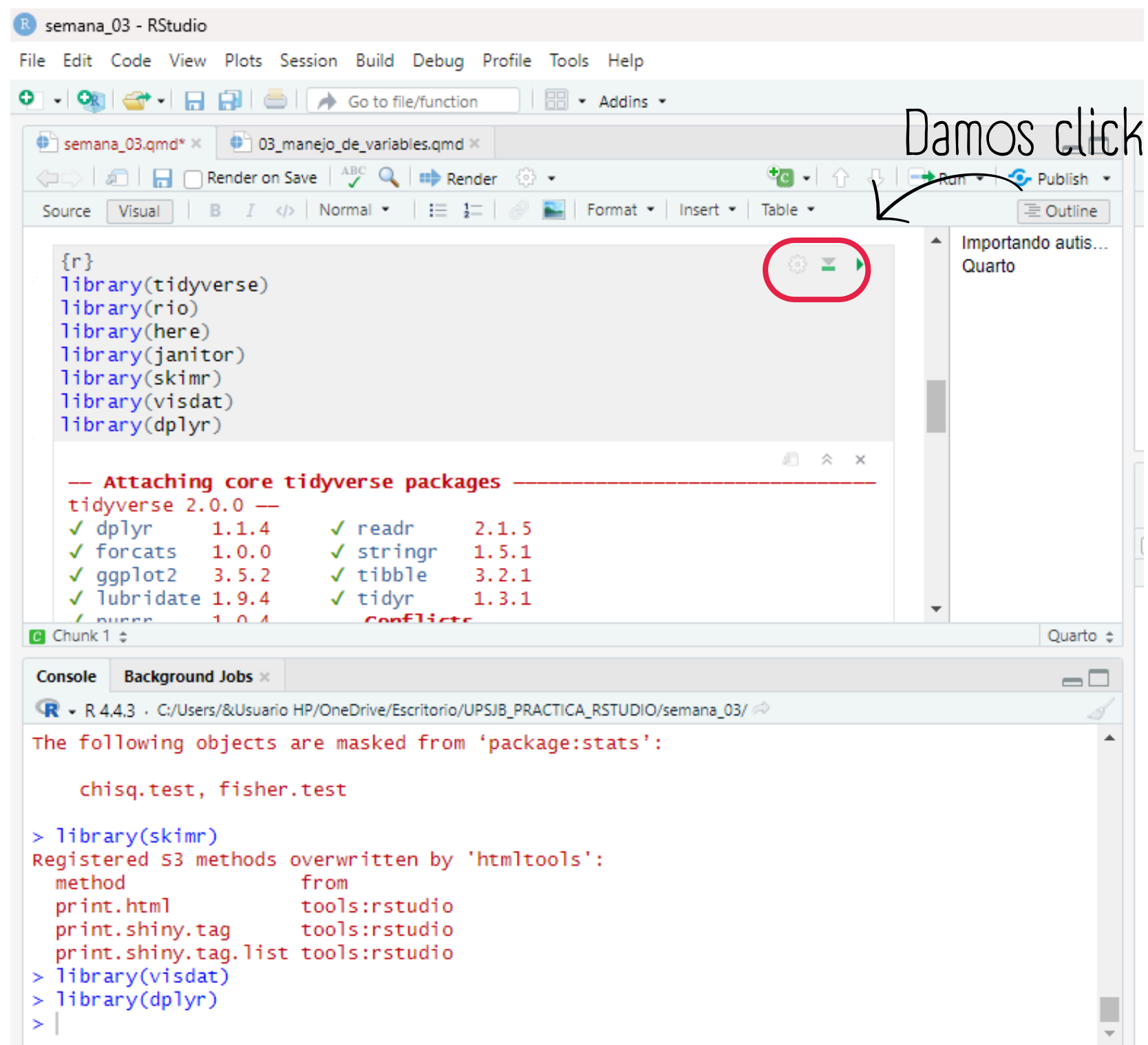
<https://cran.rstudio.com/bin/windows/Rtools/>  
Installing package into 'C:/Users/&Usuario HP/AppData/Local/R/win-library/4.4'  
(as 'lib' is unspecified)  
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/dplyr\_1.1.4.zip'  
Content type 'application/zip' length 1590760 bytes (1.5 MB)  
downloaded 1.5 MB

package 'dplyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in  
C:/Users/&Usuario HP/AppData/Local/Temp/RtmpYZWZdJ/downloaded\_packages

3

## CARGAR LOS PAQUETES



semana\_03 - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

semana\_03.qmd\* 03\_manejo\_de\_variables.qmd

Render on Save Render Run Publish

Source Visual B I </> Normal Format Insert Table Outline

```
{r}
library(tidyverse)
library(rio)
library(here)
library(janitor)
library(skimr)
library(visdat)
library(dplyr)
```

Importando autis...  
Quarto

— Attaching core tidyverse packages —

tidyverse 2.0.0			
✓ dplyr	1.1.4	✓ readr	2.1.5
✓ forcats	1.0.0	✓ stringr	1.5.1
✓ ggplot2	3.5.2	✓ tibble	3.2.1
✓ lubridate	1.9.4	✓ tidyr	1.3.1

Chunk 1

Console Background Jobs

R 4.4.3 C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB\_PRACTICA\_RSTUDIO/semana\_03/

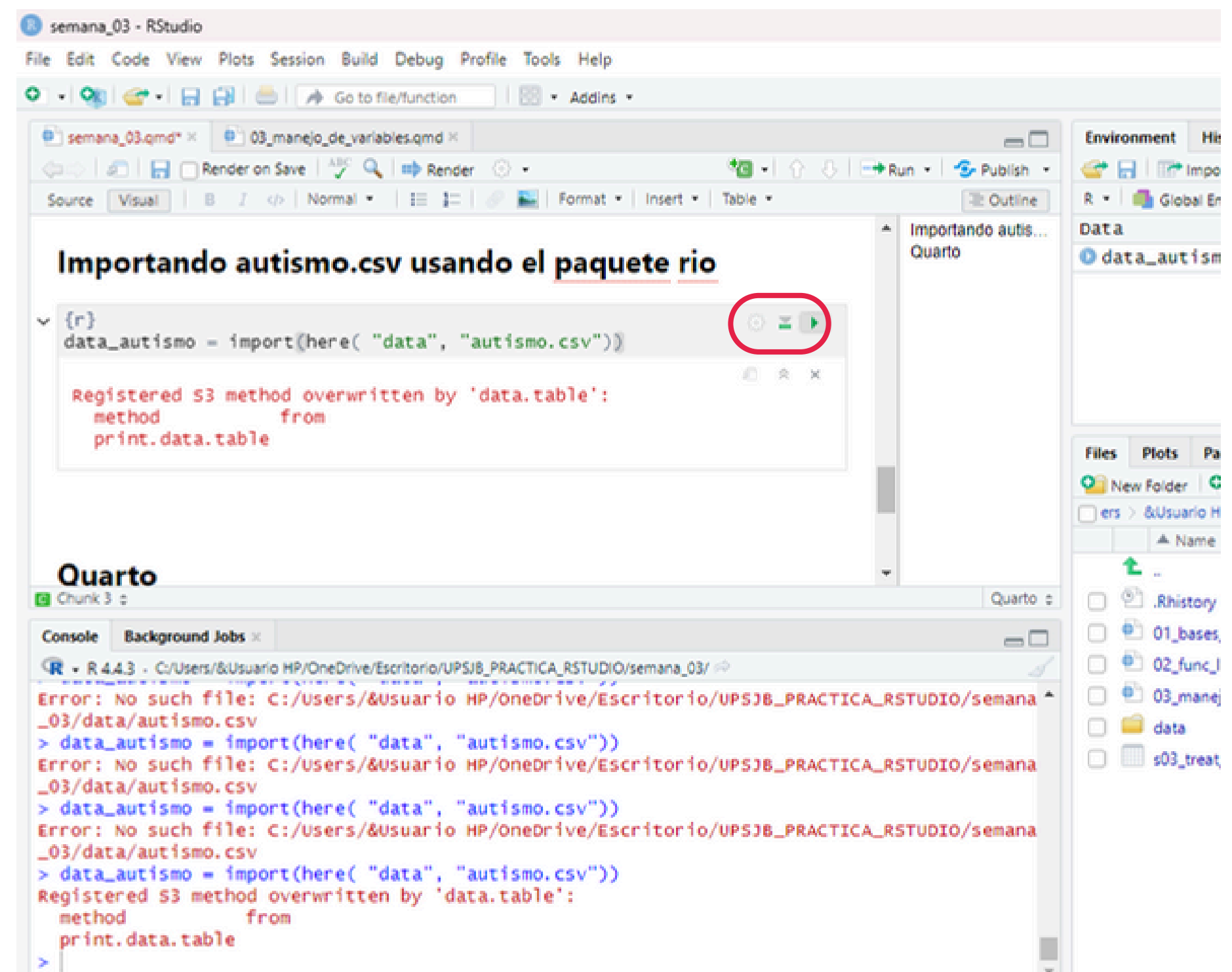
The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
> library(skimr)
Registered S3 methods overwritten by 'htmltools':
  method      from
print.html    tools:rstudio
print.shiny.tag tools:rstudio
print.shiny.tag.list tools:rstudio
> library(visdat)
> library(dplyr)
>
```

Se cargan los paquetes para que esten activos, si no estan cargados no funcionan.

## IMPORTANCION DEL DATASET



semana\_03 - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

semana\_03.qmd\* 03\_manejo\_de\_variables.qmd

Render on Save Render Run Publish

Source Visual B I </> Normal Format Insert Table Outline

Importando autis...  
Quarto

### Importando autismo.csv usando el paquete rio

```
{r}
data_autismo = import(here("data", "autismo.csv"))
```

Registered S3 method overwritten by 'data.table':

method	from
print.data.table	

Quarto

Chunk 3

Console Background Jobs

R 4.4.3 C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB\_PRACTICA\_RSTUDIO/semana\_03/

```
Error: No such file: C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB_PRACTICA_RSTUDIO/semana_03/data/autismo.csv
> data_autismo = import(here("data", "autismo.csv"))
Error: No such file: C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB_PRACTICA_RSTUDIO/semana_03/data/autismo.csv
> data_autismo = import(here("data", "autismo.csv"))
Error: No such file: C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB_PRACTICA_RSTUDIO/semana_03/data/autismo.csv
> data_autismo = import(here("data", "autismo.csv"))
Registered S3 method overwritten by 'data.table':
  method      from
print.data.table
>
```

Environment

R Global Environment

Data

data\_autismo

Files Plots Pa

New Folder

ers &Usuario H

Name

..

.Rhistory

01\_bases

02\_func

03\_manej

data

s03\_treat

Si no esta instalado, ni cargado el paquete "rio", sale error



## TIPOS DE VARIABLES Y CUANTOS HAY

Funcion "dim": para obtener o establecer las dimensiones de una matriz, matriz o marco de datos

The screenshot shows the RStudio interface. In the source editor, the code `dim(data_autismo)` is entered. The console at the bottom shows the output: `[1] 292 22`. A red circle highlights the run button in the source editor. The environment pane on the right shows the variable `data_aut` as a data frame.

```
{r}
dim(data_autismo)

[1] 292 22
```

```
> data_autismo = import(here("data", "autismo.csv"))
Error: No such file: C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB_PRACTICA_RSTUDIO/semana_03/data/autismo.csv
> data_autismo = import(here("data", "autismo.csv"))
Error: No such file: C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB_PRACTICA_RSTUDIO/semana_03/data/autismo.csv
> data_autismo = import(here("data", "autismo.csv"))
Registered S3 method overwritten by 'data.table':
  method      from
print.data.table
> dim(data_autismo)
[1] 292 22
```

TOTAL DE DATOS

VISTASO DEL CONTENIDO

Funcion "str" Y Funcion "glimpse"

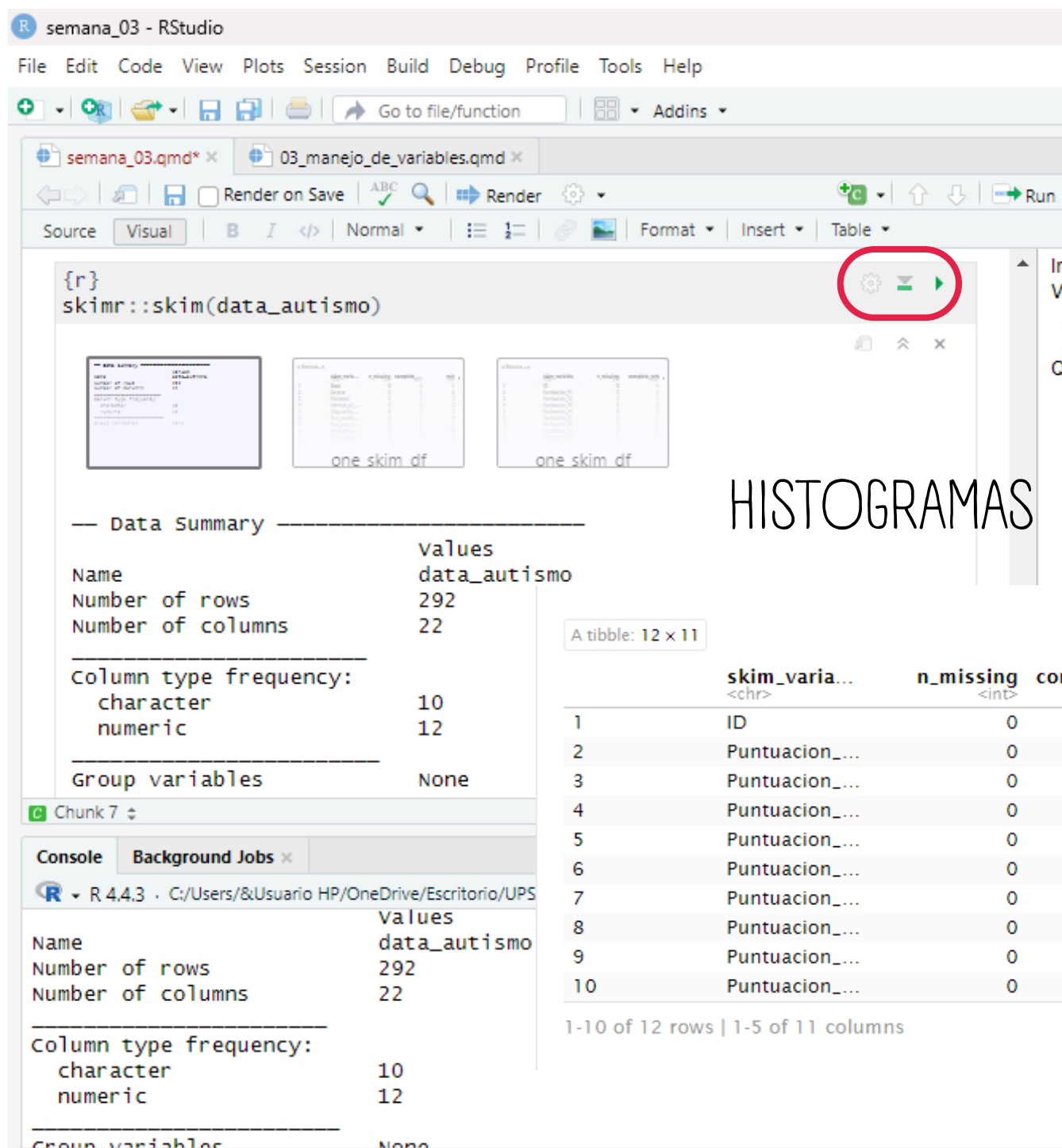
The screenshot shows the RStudio interface. In the source editor, the code `str(data_autismo)` is entered. The console at the bottom shows the output of the `str()` function, displaying the structure of the data frame with 292 observations and 22 variables. A red circle highlights the run button in the source editor.

```
{r}
str(data_autismo)

'data.frame': 292 obs. of 22 variables:
 $ ID          : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Puntuacion_P1 : int  1 1 1 0 1 0 1 1 1 0 ...
 $ Puntuacion_P2 : int  1 1 1 1 1 0 0 1 1 0 ...
 $ Puntuacion_P3 : int  0 0 0 0 1 1 1 1 1 1 ...
 $ Puntuacion_P4 : int  0 0 0 0 1 0 1 1 1 1 ...
 $ Puntuacion_P5 : int  1 1 0 1 1 1 1 1 1 1 ...
 $ Puntuacion_P6 : int  1 1 1 1 1 1 1 1 1 0 ...
 $ Puntuacion_P7 : int  0 0 1 0 1 0 0 1 1 1 ...
 $ Puntuacion_P8 : int  1 1 1 0 1 1 1 1 0 1 ...
 $ Puntuacion_P10 : int  0 0 0 1 1 1 1 0 0 0 ...
 $ Edad          : chr  "6" "6" "6" "5" ...
 $ Genero        : chr  "Masculino" "Masculino" "Masculino" "Femenino" ...
 $ Etnicidad     : chr  "Otros" "Oriente Medio" "Desconocido" "Desconocido" ...
 $ Ictericia_al_nacer : chr  "No" "No" "No" "Sí" ...
 $ Diagnostico_autismo : chr  "No" "No" "No" "No" ...
 $ Pais_residencia : chr  "Jordan" "Jordan" "Jordan" "Jordan" ...
 $ Uso_prev_aplicacion : chr  "No" "No" "Sí" "No" ...
 $ Resultado     : int  5 5 5 4 10 5 7 8 7 5 ...
 $ Descripcion_edad : chr  "4-11 years" "4-11 years" "4-11 years" "4-11 years" ...
 $ Quien_completa_prueba : chr  "Padre/Madre" "Padre/Madre" "Desconocido" "Desconocido" ...
 $ Diagnostico_ASD : chr  "No" "No" "No" "No" ...
```

# Estadísticos descriptivos y otros parámetros para exploración de datos

Funcion "skim": para obtener una vision mas general de dataframe, genere un pequeño histograma para mostrar la distribucion de los datos (numericos/enteros).



Resumen por variable

Funcion "summary"

The screenshot shows the RStudio interface with the `summary(data_autismo)` command in the console. The output is displayed in the Environment pane, showing a summary of the data for each variable. The summary includes minimum, first quartile, median, mean, third quartile, and maximum values for each variable. The word "HISTOGRAMAS" is written in large letters over the table.

ID	Puntuacion_P1	Puntuacion_P2	Puntuacion_P3
Puntuacion_P4	Min. : 1.00	Min. :0.0000	Min. :0.0000
Min. :0.0000	1st Qu.: 73.75	1st Qu.:0.0000	1st Qu.:0.0000
1st Qu.:0.0000	Median :146.50	Median :1.0000	Median :1.0000
Median :1.0000	Mean :146.50	Mean :0.6336	Mean :0.7432
Mean :0.5514	3rd Qu.:219.25	3rd Qu.:1.0000	3rd Qu.:1.0000
3rd Qu.:1.0000	Max. :292.00	Max. :1.0000	Max. :1.0000
Max. :1.0000	Puntuacion_P5	Puntuacion_P6	Puntuacion_P7
Puntuacion_P5	Min. :0.0000	Min. :0.0000	Min. :0.0000
Min. :0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
1st Qu.:0.0000	Median :1.0000	Median :1.0000	Median :0.0000
Median :0.0000	Mean :0.7432	Mean :0.7123	Mean :0.6062
Mean :0.4932	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
3rd Qu.:1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000
Max. :1.0000			



# Observamos los datos y variables del dataset asignado "AUTISMO"

Funcion: VIEW

```
{r}  
view(data_autismo)
```



semana\_03 - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

semana\_03.qmd data\_autismo 03\_manejo\_de\_variables.qmd

Filter

	ID	Puntuacion_P1	Puntuacion_P2	Puntuacion_P3	Puntuacion_P4	Puntuacion_P5	Puntuacion_P6	Puntuacion_P7	Puntuacion_P8	Puntuacion_P9
1	1	1	1	0	0	1	1	0	1	1
2	2	1	1	0	0	1	1	0	1	1
3	3	1	1	0	0	0	1	1	1	1
4	4	0	1	0	0	1	1	0	0	0
5	5	1	1	1	1	1	1	1	1	1
6	6	0	0	1	0	1	1	0	1	1
7	7	1	0	1	1	1	1	0	1	1
8	8	1	1	1	1	1	1	1	1	1
9	9	1	1	1	1	1	1	1	1	0
10	10	0	0	1	1	1	0	1	1	1
11	11	1	0	0	0	1	1	1	1	1
12	12	0	1	0	0	1	0	0	0	0
13	13	0	1	1	1	1	1	1	1	1
14	14	1	0	0	0	0	0	1	0	0
15	15	1	1	1	1	1	1	1	1	1
16	16	1	1	1	1	1	1	1	1	1

Showing 1 to 16 of 292 entries, 22 total columns

Console Background Jobs

R 4.4.3 · C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB\_PRACTICA\_RSTUDIO/semana\_03/

Resultado	Descripcion_edad	quien_completa_prueba	diagnostico ASD
Min. : 0.00	Length:292	Length:292	Length:292
1st Qu.: 5.00	Class :character	Class :character	Class :character
Median : 6.00	Mode :character	Mode :character	Mode :character
Mean : 6.24			

semana\_03 - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

semana\_03.qmd data\_autismo 03\_manejo\_de\_variables.qmd

Filter

Edad	Genero	Etnicidad	Ictericia_al_nacer	Diagnostico_autismo	Pais_residencia	Uso_prev_aplicacion	Resultado	Descripcion_edad
6	Masculino	Otros	No	No	Jordan	No	5	4-11 years
6	Masculino	Oriente Medio	No	No	Jordan	No	5	4-11 years
6	Masculino	Desconocido	No	No	Jordan	Sí	5	4-11 years
5	Femenino	Desconocido	Sí	No	Jordan	No	4	4-11 years
5	Masculino	Otros	Sí	No	Estados Unidos	No	10	4-11 years
4	Masculino	Desconocido	No	Sí	Egypt	No	5	4-11 years
5	Masculino	Blanco-Europeo	No	No	Reino Unido	No	7	4-11 years
5	Femenino	Oriente Medio	No	No	Baréin	No	8	4-11 years
11	Femenino	Oriente Medio	No	No	Baréin	No	7	4-11 years
11	Femenino	Desconocido	No	Sí	Austria	No	5	4-11 years
10	Masculino	Blanco-Europeo	Sí	No	Reino Unido	No	7	4-11 years
5	Femenino	Desconocido	No	No	Kuwait	No	3	4-11 years
4	Masculino	Blanco-Europeo	Sí	No	Estados Unidos	No	9	4-11 years
4	Femenino	Negro	No	No	Emiratos Árabes Unidos	No	2	4-11 years
6	Masculino	Blanco-Europeo	No	No	Europe	No	10	4-11 years
8	Masculino	Blanco-Europeo	No	No	Malta	No	10	4-11 years

Showing 1 to 16 of 292 entries, 22 total columns

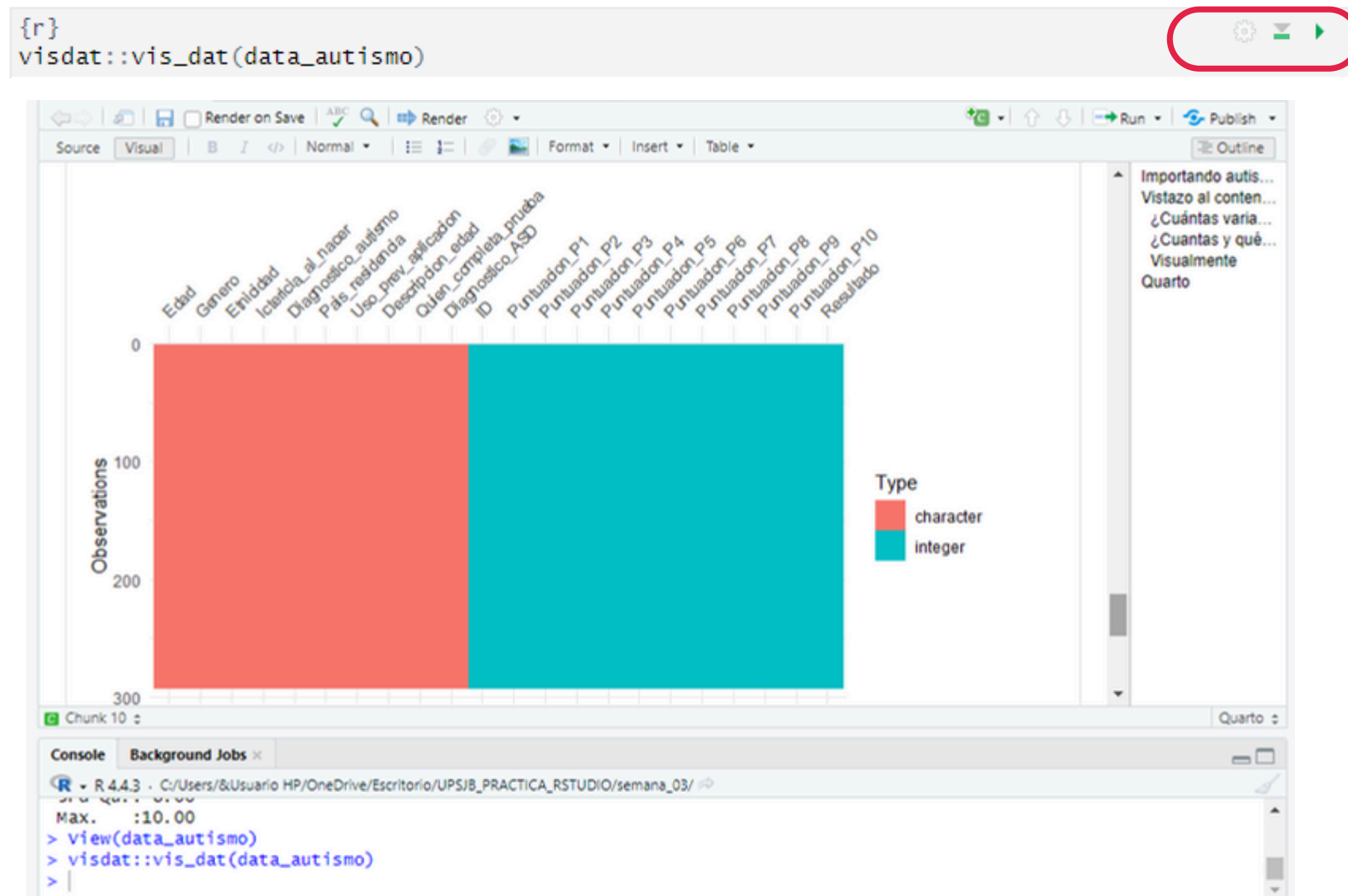
Console Background Jobs

R 4.4.3 · C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB\_PRACTICA\_RSTUDIO/semana\_03/

Resultado	Descripcion_edad	quien_completa_prueba	diagnostico ASD
Min. : 0.00	Length:292	Length:292	Length:292
1st Qu.: 5.00	Class :character	Class :character	Class :character
Median : 6.00	Mode :character	Mode :character	Mode :character
Mean : 6.24			

# LIMPIEZA DE DATOS

Funcion: "vis\_dat": Proporciona una visualizacion rapida de un marco de datos completo, util para identificar y entender los valores faltantes.



semana\_03 - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

semana\_03.qmd \* data\_autismo \* 03\_manejo\_de\_variables.qmd \*

Source Visual B I </> Normal Format Insert Table

## Limpieza de datos

### Paso uno: corregir los nombres de variables.

Clean names es una función del paquete janitor, importante para estandarizar y limpiar los nombres de las columnas de un data frame (nombres unico, en minusculas con guiones bajos)

Generamos un nuevo data

```
{r}
data_autismo_1 = clean_names(data_autismo)
```

Generamos un nueva data

Environment History

R Global Environ

Data

- data\_autismo
- data\_autismo\_1

Files Plots Packag

- New Folder
- C: > Users > &Usui
- Name
- ..
- .Rhistory
- 01\_bases\_de\_
- 02\_func\_lib\_p
- 03\_manejo\_de
- data
- s03\_treat\_artr

Console Background Jobs

```
R 4.4.3 - C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB_PRACTICA_RSTUDIO/semana_03/
> view(data_autismo)
> visdat::vis_dat(data_autismo)
> visdat::vis_dat(data_autismo)
> data_autismo_1 = clean_names(data_autismo)
>
```

data\_autismo\_1



Nota el contraste (la función "names ()" imprime los nombres de columnas de un dataset).

```
{r}  
names(data_autismo_1)  
  
[1] "id" "puntuacion_p1" "puntuacion_p2"  
"puntuacion_p3"  
[5] "puntuacion_p4" "puntuacion_p5" "puntuacion_p6"  
"puntuacion_p7"  
[9] "puntuacion_p8" "puntuacion_p9" "puntuacion_p10"  
"edad"  
[13] "genero" "etnicidad" "ictericia_al_nacer"  
"diagnostico_autismo"  
[17] "pais_residencia" "uso_prev_aplicacion" "resultado"  
"descripcion_edad"  
[21] "quien_completa_prueba" "diagnostico_asd"
```

data\_autismo\_1

Convertir celdas vacías a NA y Eliminar columnas o filas vacías.

Función "mutate\_if":  
Permite aplicar una transformación a un grupo de columnas en un marco de datos basado en una condición lógica.

data\_autismo\_2

### Paso dos: convertir celdas vacías a NA

```
{r}  
data_autismo_2 = mutate_if(data_autismo_1,  
is.character, list(~na_if(., "")))
```

### Paso tres: eliminar columnas o filas vacías.

```
{r}  
data_autismo_3 = remove_empty  
(data_autismo_2, which = c("rows", "cols"))
```

Funcion:  
"remove\_empty"  
disponible en el paquete  
"Janitor" sirve para  
eliminar filas y columnas  
vacías en un data frame o  
matriz

data\_autismo\_3

# Convertir celdas vacías a NA y Eliminar columnas o filas vacías.

Funcion: SELECT

**Para eliminar columnas con los mismos resultados**

```
{r}  
data_autismo_4 <- select(data_autismo_3,-diagnostico_autismo)
```

- La función "select", se utiliza para seleccionar un subconjunto de variables o columnas de un conjunto de datos.

data\_autismo\_4

Funcion: CLEAN NAMES

```
{r}  
data_autismo_5 = data_autismo |>  
  clean_names() |>  
  mutate_if(is.character, list(~ na_if(., ""))) |>  
  remove_empty(which = c("rows", "cols"))
```

data\_autismo\_5

## Corregir errores ortográficos o valores inválidos

Corregir errores ortográficos usando `mutate()` y `case_when()`

```
{r}  
data_autismo_7 = data_autismo_6 |>  
  mutate(edad = case_when(  
    edad == "?" ~ "Na",  
    TRUE ~ edad))
```

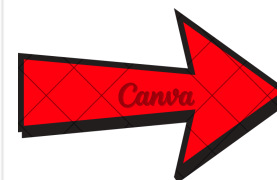
```
{r}  
data_autismo_7 |> count(edad) # Cambia de variable categórica
```

```
{r}  
data_autismo_5 |> count(edad) # Cambia de variable categórica
```

Description: df [9 x 2]

edad <chr>	n <int>
10	18
11	26
4	92
5	45
6	39
7	27
8	21
9	20
?	4

INSPECCION TABULAR



A tibble: 9 x 2

edad <chr>	n <int>
10	18
11	26
4	92
5	45
6	39
7	27
8	21
9	20
Na	4

9 rows

CORREGIDO


# INSPECCION TABULAR

```
{r}
data_autismo_5 |> count(genero) # Cambia de variable categórica
```

Description: df [2 x 2]

genero <chr>	n <int>
Femenino	84
Masculino	208

2 rows




```
{r}
data_autismo_5 |> count(ictericia_al_nacer) # Cambia de variable categórica
```

Description: df [2 x 2]

ictericia_al_n... <chr>	n <int>
No	212
Sí	80


2 rows



```
{r}
data_autismo_5 |> count(etnicidad) # Cambia de variable categórica
```

Description: df [11 x 2]

etnicidad <chr>	n <int>
Asiático	46
Asiático del Sur	21
Blanco-Europeo	108
Desconocido	43
Hispano	7
Latino	8
Negro	14
Oriente Medio	27
Otros	14
Pacífico	2




```
{r}
data_autismo_5 |> count(diagnostico_asd) # Cambia de variable categórica
```

Description: df [2 x 2]

diagnostico_... <chr>	n <int>
No	151
Sí	141

2 rows




```
{r}
data_autismo_5 |> count(quien_completa_pru...
categorica
```

Description: df [5 x 2]

quien_compl... <chr>	n <int>
Autocompletado	5
Desconocido	43
Familiar	17
Padre/Madre	214
Profesional de...	13

5 rows



```
{r}
data_autismo_5 |> count(pais_residencia) # Cambia de variable categórica
```

Description: df [52 x 2]

pais_residencia <chr>	n <int>
Afghanistan	2
Alemania	1
Arabia Saudita	4
Argentina	1
Armenia	3
Australia	23
Austria	2
Bangladés	6
Baréin	2
Brazil	2

1-10 of 52 rows

Previous 1 2 3 4 5 6 Next





## Transformando de data.frame a as tibble

data\_autismo\_6

```
{r}  
data_autismo_6 = as_tibble(data_autismo_5)
```

## Corregir errores ortográficos usando mutate() y case\_when()

- Funcion "mutate": define las nuevas variables o modificaciones.
- Funcion: "case\_when": proporciona una forma de definir valores condicionalmente.

```
{r}  
data_autismo_7 = data_autismo_6 |>  
  mutate(edad = case_when(  
    edad == "?" ~ "Na",  
    TRUE ~ edad))
```

data\_autismo\_7

Para observar la correccion de datos

```
{r}  
data_autismo_7 |> count(edad) # Cambia de variable categórica
```

# Colapsar una variable categórica en menos niveles

Ver la variable de interés

```
{r}  
data_autismo_7 |> count((puntuacion_p1),(puntuacion_p2), (puntuacion_p3),  
(puntuacion_p4),(puntuacion_p5), (puntuacion_p6),(puntuacion_p7  
,(puntuacion_p8), (puntuacion_p9),(puntuacion_p10))
```

A tibble: 186 × 11

(puntuacion_p1) <int>	(puntuacion_p2) <int>	(puntuacion_p3) <int>
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0

1-10 of 186 rows | 1-3 of 11 columns Previous 1 2 3 4 5 6 ... 19 Next

```
{r}  
data_autismo_8 = data_autismo_7 |>  
  mutate(resultado_2 = case_when(  
    resultado %in% c("1", "2", "3", "4") ~ "no autismo", # tiene menos rasgos  
    autista por la puntuacion de 1 al 4  
    resultado %in% c("5", "6", "7", "8", "9", "10") ~ "si autismo ", # tiene  
    mas rasgos autistas por la puntuacion de las columnas puntuacion_p1 a  
    puntuacion_p10 ))  
|
```

Comprobando el cambio

```
{r}  
data_autismo_4 |> count(resultado_2)
```

data\_autismo\_8

Creando una nueva columna con las variables

iso_prev_aplicacion	resultado	descripcion_edad	quien_completa_prueba	diagnostico_asd	resultado_2
lo	5	4-11 years	Padre/Madre	No	si autismo
lo	5	4-11 years	Padre/Madre	No	si autismo
if	5	4-11 years	Desconocido	No	si autismo
lo	4	4-11 years	Desconocido	No	no autismo
lo	10	4-11 years	Padre/Madre	Sí	si autismo
lo	5	4-11 years	Desconocido	No	si autismo
lo	7	4-11 years	Padre/Madre	Sí	si autismo
lo	8	4-11 years	Padre/Madre	Sí	si autismo
lo	7	4-11 years	Padre/Madre	Sí	si autismo
lo	5	4-11 years	Desconocido	No	si autismo
lo	7	4-11 years	Autocompletado	Sí	si autismo
lo	3	4-11 years	Desconocido	No	no autismo
lo	9	4-11 years	Padre/Madre	Sí	si autismo
lo	2	4-11 years	Padre/Madre	No	no autismo
lo	10	4-11 years	Padre/Madre	Sí	si autismo
lo	10	4-11 years	Padre/Madre	Sí	si autismo

Comprobando el cambio

```
{r}  
data_autismo_8 |> count(resultado_2)
```

A tibble: 3 × 2

resultado_2 <chr>	n <int>
no autismo	69
si autismo	222

data\_autismo\_8

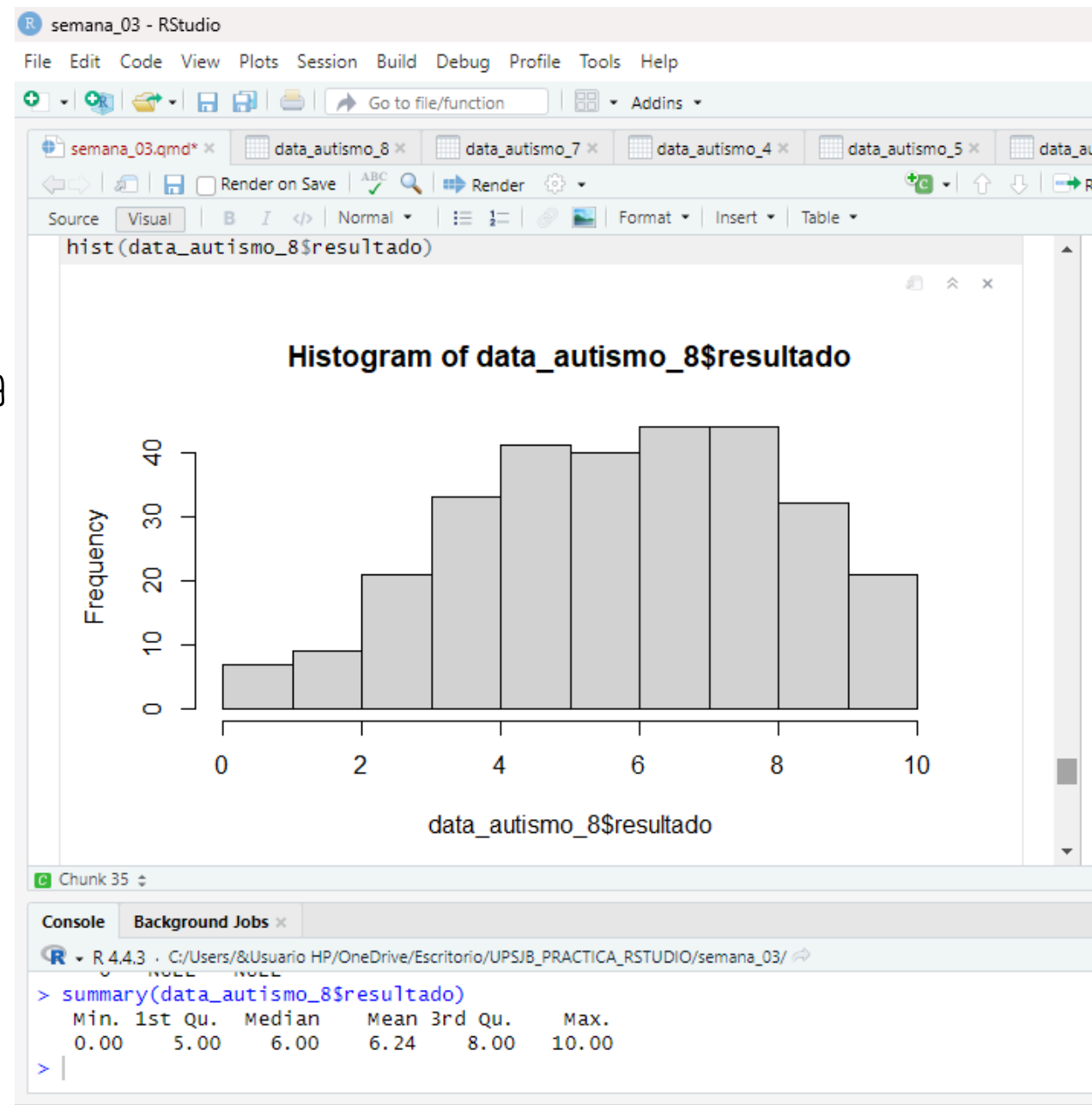
# Transformar una variable

Transformación a logaritmo

```
{r}  
summary(data_autismo_8$resultado)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	5.00	6.00	6.24	8.00	10.00

Funcion  
"hist": crea  
un histograma  
de la variable.



Transformando valores a valores perdidos usando la función `na_if()`

```
{r}  
data_autismo_8= data_autismo_7 |>  
  mutate(resultado_2 = na_if(resultado_2, NA))
```

Para eliminar columnas repetidas luego de la revision

```
{r}  
data_autismo_9 <- select(data_autismo_8, -puntuacion_p1, -puntuacion_p2,  
  -puntuacion_p3, -puntuacion_p4, -puntuacion_p5, -puntuacion_p6,  
  -puntuacion_p7, -puntuacion_p8, -puntuacion_p9, -puntuacion_p10,  
  -diagnostico_autismo)
```

data\_autismo\_9



## Transformar una variable

Imprimir los nombres para ver que cambios necesita

```
{r}
names(data_autismo_9)
```

[1]	"id"	"edad"	"genero"
[4]	"etnicidad"	"ictericia_al_nacer"	"pais_residencia"
[7]	"uso_prev_aplicacion"	"resultado"	"descripcion_edad"
[10]	"quien_completa_prueba"	"diagnostico_asd"	"resultado_2"

## Cambiando un nombre de variables o varios a la vez

Funcion "rename": permite renombrar una variable en un data frame.

```
{r}
data_autismo_10 <- data_autismo_9 |>
  rename(pais_de_residencia = pais_residencia)
```

Varios a la vez

data\_autismo\_10

```
{r}
data_autismo_10 <- data_autismo_9 |>
  rename(pais_de_residencia = pais_residencia,
        niño_diagnosticado_con_asd = diagnostico_asd,
        uso_previo_aplicacion = uso_prev_aplicacion,
        resultado_de_prueba = resultado,
        determinacion_segun_la_puntuacion_1_a_10 = resultado_2,
        Persona_responsable_que_completo_la_prueba = quien_completa_prueba)
```

## COMPROBANDO EL CAMBIO

Comprobando

```
{r}
names(data_autismo_10)
```

[1]	"id"	"edad"
[3]	"genero"	"etnicidad"
[5]	"ictericia_al_nacer"	"pais_de_residencia"
[7]	"uso_previo_aplicacion"	"resultado_de_prueba"
[9]	"descripcion_edad"	
	"Persona_responsable_que_completo_la_prueba"	
[11]	"niño_diagnosticado_con_asd"	
	"determinacion_segun_la_puntuacion_1_a_10"	

# Dataset asignado "AUTISMO"

data\_autismo\_10

semana\_03 - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

semana\_03.qmd\* x data\_autismo\_10 x data\_autismo\_9 x data\_autismo\_7 x

Filter

	id	edad	genero	etnicidad	ictericia_al_nacer	pais_de_residencia	uso_previo_aplicacion	resultado_de_prueba
1	1	6	Masculino	Otros	No	Jordan	No	
2	2	6	Masculino	Oriente Medio	No	Jordan	No	
3	3	6	Masculino	Desconocido	No	Jordan	Sí	
4	4	5	Femenino	Desconocido	Sí	Jordan	No	
5	5	5	Masculino	Otros	Sí	Estados Unidos	No	10
6	6	4	Masculino	Desconocido	No	Egypt	No	
7	7	5	Masculino	Blanco-Europeo	No	Reino Unido	No	
8	8	5	Femenino	Oriente Medio	No	Baréin	No	
9	9	11	Femenino	Oriente Medio	No	Baréin	No	
10	10	11	Femenino	Desconocido	No	Austria	No	
11	11	10	Masculino	Blanco-Europeo	Sí	Reino Unido	No	
12	12	5	Femenino	Desconocido	No	Kuwait	No	
13	13	4	Masculino	Blanco-Europeo	Sí	Estados Unidos	No	
14	14	4	Femenino	Negro	No	Emiratos Árabes Unidos	No	
15	15	6	Masculino	Blanco-Europeo	No	Europe	No	10
16	16	8	Masculino	Blanco-Europeo	No	Malta	No	10

Showing 1 to 16 of 292 entries, 12 total columns

Console Background Jobs x

R 4.4.3 · C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB\_PRACTICA\_RSTUDIO/semana\_03/

```
+      Persona_responsable_que_completo_la_prueba = quien_completa_prueba)
> view(data_autismo_10)
> view(data_autismo_10)
>
```

semana\_03 - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

semana\_03.qmd\* x data\_autismo\_10 x data\_autismo\_9 x data\_autismo\_7 x

Filter

	resultado_de_prueba	descripcion_edad	Persona_responsable_que_completo_la_prueba	niño_diagnosticado_con_asd	determinacion_segun_la...
	5	4-11 years	Padre/Madre	No	si autismo
	5	4-11 years	Padre/Madre	No	si autismo
	5	4-11 years	Desconocido	No	si autismo
	4	4-11 years	Desconocido	No	no autismo
	10	4-11 years	Padre/Madre	Sí	si autismo
	5	4-11 years	Desconocido	No	si autismo
	7	4-11 years	Padre/Madre	Sí	si autismo
	8	4-11 years	Padre/Madre	Sí	si autismo
	7	4-11 years	Padre/Madre	Sí	si autismo
	5	4-11 years	Desconocido	No	si autismo
	7	4-11 years	Autocompletado	Sí	si autismo
	3	4-11 years	Desconocido	No	no autismo
	9	4-11 years	Padre/Madre	Sí	si autismo
	2	4-11 years	Padre/Madre	No	no autismo
	10	4-11 years	Padre/Madre	Sí	si autismo
	10	4-11 years	Padre/Madre	Sí	si autismo

Showing 1 to 16 of 292 entries, 12 total columns

Console Background Jobs x

R 4.4.3 · C:/Users/&Usuario HP/OneDrive/Escritorio/UPSJB\_PRACTICA\_RSTUDIO/semana\_03/

```
+      Persona_responsable_que_completo_la_prueba = quien_completa_prueba)
> view(data_autismo_10)
> view(data_autismo_10)
>
```



**UNIVERSIDAD PRIVADA  
SAN JUAN BAUTISTA**