Cargando los datos

```
{r}
autismo_limpio <- import(here("data", "data_autismo_12"))
</pre>
```

Antes de <u>iniciar</u> el <u>análisis</u> descriptivo de <u>tus</u> datos, <u>estos</u> deben <u>ser inspeccionados</u> para la <u>eliminación</u> de <u>errores</u> en la <u>codificación</u> o <u>transformación</u> de <u>valores perdidos</u> o anómalos.

La "estructura" de los datos

1. Resumen de variables categóricas

Para resumir variables categóricas usamos frecuencias simples y frecuencias porcentuales. Es importante resumir, también, los datos perdidos. Resumen con la función table()

```
{r}
table(autismo_limpio$genero, useNA = "ifany")

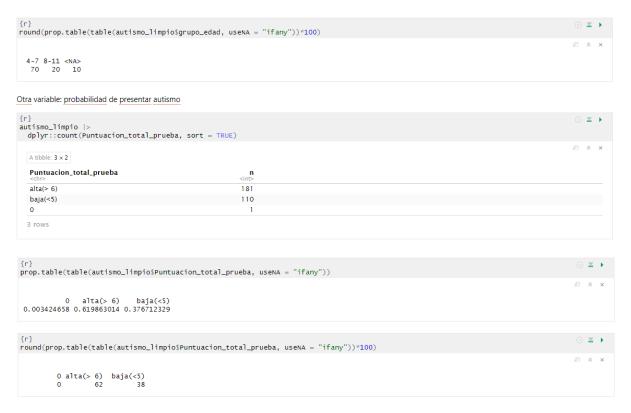
Femenino Masculino
84 208

{r}
table(autismo_limpio$grupo_edad, useNA = "ifany")

4-7 8-11 <NA>
203 59 30
```

Con tidyverse

Valores perdidos en proporciones



2. Resumen de variables continuas

Para resumir variables numéricas hay funciones en la base de R que permiten calcular los principales estadísticos. Por ejemplo, la función mean() y sd(), calcula el promedio y desviación estandard de un conjunto de datos.



Ojo, en este ejemplo sólo será posible calcular el promedio y desviación estandard si el valor para el argumento na.rm es TRUE. Esto elimina los valores perdidos en el cálculo.

Otras funciones que puedes encontrar útil son min(), max(), median() y IQR()

Ojo, en este ejemplo sólo será posible calcular el promedio y desviación estandard si el valor para el argumento na m es TRUE. Esto elimina los valores perdidos en el cálculo.

Otras funciones que puedes encontrar útil son min(), max(), median() y IQR()

```
{r}
min(autismo_limpio_1$puntuacion_total, na.rm = TRUE) # Proporciona el valor mínimo

IQR(autismo_limpio_1$puntuacion_total, na.rm = TRUE) # Calcula el rango intercuartilico

median(autismo_limpio_1$puntuacion_total, na.rm = TRUE) # Calcula la mediana

[1] 0
[1] 3
[1] 6
```

La función summary() para calcular todas las medidas, a la vez.

summary() es una función nativa de R que permite calcular todas las medidas estadísticas a la vez. Abajo, un ejemplo con la variable puntuacion total

¿Mediana o media?

Una de las cosas que queremos saber al realizar un análisis descriptivo es la distribución de los datos. La media y mediana puede informarnos al respecto. Abajo, un ejemplo para la variable puntuación total



La media y mediana de la variable resultado son similares. Estos sugiere que los datos tienen una distribución simétrica.

Debajo otro ejemplo con la variable puntuacion_p1

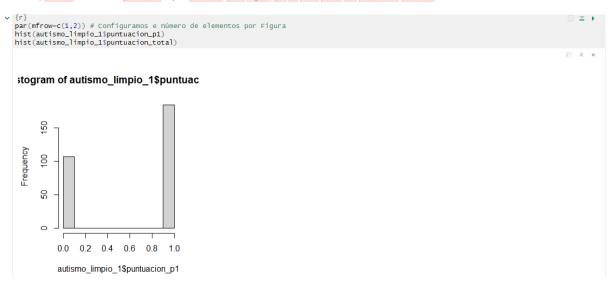
```
{r}
mean(autismo_limpio_1$puntuacion_p1, na.rm = T)

[1] 0.6335616

median(autismo_limpio_1$puntuacion_p1, na.rm = T)

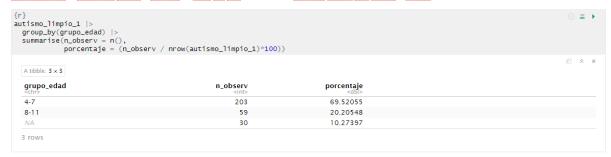
[1] 1
```

La media y mediana de la variable puntuacion_p1 son diferentes. Estos sugiere que los datos tienen una distribución asimétrica

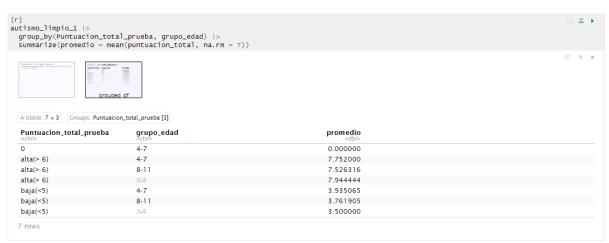


[resumen] Por otra variable

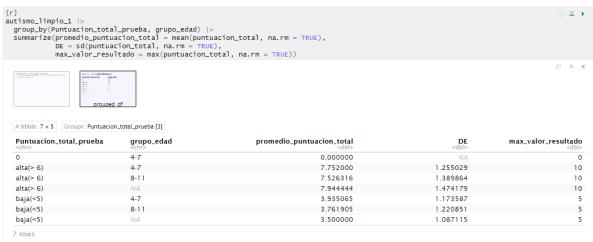
Frecuentemente es importante realizar el resumen de datos por otra variable. En R, podemos calcular esto usando la función <code>group_by()</code> y <code>summarize()</code>







Otros estadísticos

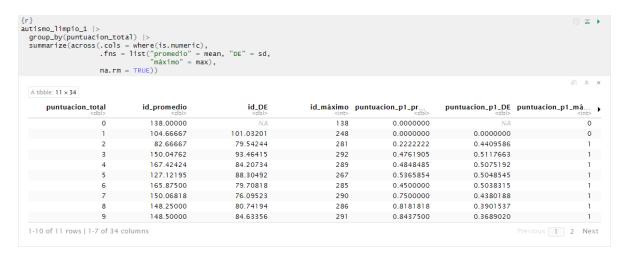


Percentiles del nivel de puntuacion total en participantes con autismos usando la función filter ()

```
∅ x •
autismo_limpio_1 |>
    dplyr::filter(puntuacion_total == "si") |> # Filtra los registros donde el resultado es "si"
    group_by(grupo_edad) |> # Agrupa por grupo de edad
       summarize
           ummarrze(
pp25_puntuacion_p1 = quantile(puntuacion_p1, probs = 0.25, na.rm = TRUE),
p50_puntuacion_p1 = quantile(puntuacion_p1, probs = 0.50, na.rm = TRUE),
p75_puntuacion_p1 = quantile(puntuacion_p1, probs = 0.75, na.rm = TRUE),
           p25_puntuacion_p2 = quantile(puntuacion_p2, probs = 0.25, na.rm = TRUE), p50_puntuacion_p2 = quantile(puntuacion_p2, probs = 0.50, na.rm = TRUE), p75_puntuacion_p2 = quantile(puntuacion_p2, probs = 0.75, na.rm = TRUE),
           p25_puntuacion_p3 = quantile(puntuacion_p3, probs = 0.25, na.rm = TRUE), p50_puntuacion_p3 = quantile(puntuacion_p3, probs = 0.50, na.rm = TRUE), p75_puntuacion_p3 = quantile(puntuacion_p3, probs = 0.75, na.rm = TRUE),
           \label{eq:p25_puntuacion_p4} \begin{array}{lll} \texttt{p25\_puntuacion\_p4} & \texttt{quantile(puntuacion\_p4}, \texttt{probs} = 0.25, \texttt{na.rm} = \texttt{TRUE}), \\ \texttt{p50\_puntuacion\_p4} & \texttt{quantile(puntuacion\_p4}, \texttt{probs} = 0.50, \texttt{na.rm} = \texttt{TRUE}), \\ \texttt{p75\_puntuacion\_p4} & \texttt{quantile(puntuacion\_p4}, \texttt{probs} = 0.75, \texttt{na.rm} = \texttt{TRUE}), \\ \end{array}
           p25_puntuacion_p5 = quantile(puntuacion_p5, probs = 0.25, na.rm = TRUE), p50_puntuacion_p5 = quantile(puntuacion_p5, probs = 0.50, na.rm = TRUE), p75_puntuacion_p5 = quantile(puntuacion_p5, probs = 0.75, na.rm = TRUE),
           p25_puntuacion_p6 = quantile(puntuacion_p6, probs = 0.25, na.rm = TRUE), p50_puntuacion_p6 = quantile(puntuacion_p6, probs = 0.50, na.rm = TRUE), p75_puntuacion_p6 = quantile(puntuacion_p6, probs = 0.75, na.rm = TRUE),
              p25_puntuacion_p4 = quantile(puntuacion_p4, probs = 0.25, na.rm = TRUE), p50_puntuacion_p4 = quantile(puntuacion_p4, probs = 0.50, na.rm = TRUE), p75_puntuacion_p4 = quantile(puntuacion_p4, probs = 0.75, na.rm = TRUE),
              p25_puntuacion_p5 = quantile(puntuacion_p5, probs = 0.25, na.rm = TRUE), p50_puntuacion_p5 = quantile(puntuacion_p5, probs = 0.50, na.rm = TRUE), p75_puntuacion_p5 = quantile(puntuacion_p5, probs = 0.75, na.rm = TRUE),
              p25_puntuacion_p6 = quantile(puntuacion_p6, probs = 0.25, na.rm = TRUE), p50_puntuacion_p6 = quantile(puntuacion_p6, probs = 0.50, na.rm = TRUE), p75_puntuacion_p6 = quantile(puntuacion_p6, probs = 0.75, na.rm = TRUE),
              p25_puntuacion_p7 = quantile(puntuacion_p7, probs = 0.25, na.rm = TRUE), p50_puntuacion_p7 = quantile(puntuacion_p7, probs = 0.50, na.rm = TRUE), p75_puntuacion_p7 = quantile(puntuacion_p7, probs = 0.75, na.rm = TRUE),
              p25_puntuacion_p8 = quantile(puntuacion_p8, probs = 0.25, na.rm = TRUE), p50_puntuacion_p8 = quantile(puntuacion_p8, probs = 0.50, na.rm = TRUE), p75_puntuacion_p8 = quantile(puntuacion_p8, probs = 0.75, na.rm = TRUE),
              p25_puntuacion_p9 = quantile(puntuacion_p9, probs = 0.25, na.rm = TRUE), p50_puntuacion_p9 = quantile(puntuacion_p9, probs = 0.50, na.rm = TRUE), p75_puntuacion_p9 = quantile(puntuacion_p9, probs = 0.75, na.rm = TRUE),
              p25_puntuacion_p10 = quantile(puntuacion_p10, probs = 0.25, na.rm = TRUE), p50_puntuacion_p10 = quantile(puntuacion_p10, probs = 0.50, na.rm = TRUE), p75_puntuacion_p10 = quantile(puntuacion_p10, probs = 0.75, na.rm = TRUE)
```

Usando across () para multiples estadísticos descriptivos para multiples variables

```
⊕ ≖ ▶
        group_by(uso_previo_aplicacion)
group_by(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p)(co_p
                                                                                               na.rm = TRUE))
         A tibble: 2 × 34
           puntuacion_p5_promedio
                                                                                                                                                                                                                                      puntuacion_p5_DE
                                                                                                                                                                                                                                                                                                                                                                         puntuacion_p5_máximo
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         puntuacion_p6_promedio
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       puntuacion_p6_DE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   0.7046263
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             0.4570245
                                                                                                                                                                                                                                                                        0.4670994
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   0.9090909
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             0.3015113
                                                                                                                 0.7272727
        2 rows | 17-21 of 34 columns
```



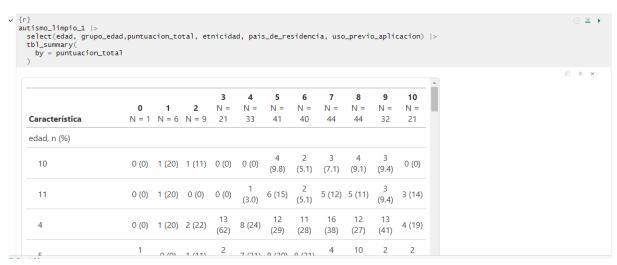
La "Tabla 1" con gtsummary



Característica	N = 292
edad, n (%)	
10	18 (6.3)
11	26 (9.0)
4	92 (32)
5	45 (16)
6	39 (14)
7	27 (9.4)
8	21 (7.3)
9	20 (6.9)
Desconocido	4

	grupo_edad, n (%)	
	4-7	203 (77
	8-11	59 (23)
	Desconocido	30
	puntuacion_total, Mediana (IQR)	6.00 (5.00 –
	etnicidad, n (%)	
	Asiático	46 (16)
	Asiático del Sur	21 (7.2)
	Blanco-Europeo	108 (37
	Desconocido	43 (15)
	Hispano	7 (2.4)
pais	_de_residencia, n (%)	
Α	fghanistan	2 (0.7)
Α	lemania	1 (0.3)
Α	rabia Saudita	4 (1.4)
Α	rgentina	1 (0.3)
Α	rmenia	3 (1.0)
Α	ustralia	23 (7.9)
Α	ustria	2 (0.7)
В	angladés	6 (2.1)
В	aréin	2 (0.7)
В	razil	2 (0.7)

uso_previo_aplicacion, n ((%)
No	281 (96)
Sí	11 (3.8)



Configurando el estilo e idioma



Característica	No N = 151	Sí N = 141
puntuacion_total, Mediana (IQR)	5.00 (4.00 – 6.00)	8.00 (7.00 – 9.00)
grupo_edad, n (%)		
4-7	102 (75)	101 (80)
8-11	34 (25)	25 (20)
Desconocido	15	15
pais_de_residencia, n (%)		
Afghanistan	2 (1.3)	0 (0)
Alemania	0 (0)	1 (0.7)
Arabia Saudita	3 (2.0)	1 (0.7)
Argentina	1 (0.7)	0 (0)
Armenia	0 (0)	3 (2.1)

Puntuacion_total_prueba	, n (%)	
0	1 (0.7)	0 (0)
alta(> 6)	40 (26)	141 (100)
baja(<5)	110 (73)	0 (0)

Recodificación de variables y adición de nombres correctos a variables

Tabla 1. Variables consideradas en el estudio

		Resultados autismo	
Característica	Overall N = 292	No N = 151	Sí N = 141
Grupo de Edad, n (%)			
4-7	203 (77)	102 (75)	101 (80)
8-11	59 (23)	34 (25)	25 (20)
Desconocido	30	15	15
genero, n (%)			
Femenino	84 (29)	46 (30)	38 (27)
Masculino	208 (71)	105 (70)	103 (73)
Puntaje Total, n (%)			
0	1 (0.3)	1 (0.7)	0 (0)
alta(> 6)	181 (62)	40 (26)	141 (100)
baja(<5)	110 (38)	110 (73)	0 (0)
ictericia_al_nacer, n (%)			
No	212 (73)	108 (72)	104 (74)
Sí	80 (27)	43 (28)	37 (26)
uso_previo_aplicacion, n (%)			
No	281 (96)	144 (95)	137 (97)
Sí	11 (3.8)	7 (4.6)	4 (2.8)
Persona responsable, n (%)			
Autocompletado	5 (1.7)	3 (2.0)	2 (1.4)
Desconocido	43 (15)	28 (19)	15 (11)
Familiar	17 (5.8)	12 (7.9)	5 (3.5)
Padre/Madre	214 (73)	102 (68)	112 (79)
Profesional de salud	13 (4.5)	6 (4.0)	7 (5.0)

Esta tabla presenta las variables consideradas en un estudio sobre autismo, comparando características entre participantes que **no presentan autismo** (N=151) y los que **sí presentan autismo** (N=141), dentro de un total de **292** participantes. Aquí están los principales resultados:

• Grupo de Edad:

- Entre 4-7 años: 77% del total (75% en no autismo, 80% en autismo).
- o Entre 8-11 años: 23% del total (25% en no autismo, 20% en autismo).
- o Datos de edad desconocida: 30 casos (15 en cada grupo).

Género:

- Femenino: 29% del total (30% en no autismo, 27% en autismo).
- o Masculino: 71% del total (70% en no autismo, 73% en autismo).
- Puntaje Total (en una escala que parece medir riesgo o severidad):
 - Puntaje 0: 0.3% del total (solo en no autismo).
 - Puntaje alto (>6): 62% del total, **100% en grupo con autismo**.
 - Puntaje bajo (<5): 38% del total, sólo en grupo sin autismo.

Ictericia al nacer:

- No: 73% del total.
- Sí: 27% (distribuido casi igual en ambos grupos).

• Uso previo de la aplicación:

- No: 96% del total.
- Sí: 3.8% (ligeramente más en grupo sin autismo).

• Persona responsable de responder la encuesta:

- Mayoritariamente respondido por padres o madres (73%).
- Otras respuestas incluyen profesionales de salud, familiares, autocompletado y algunos casos desconocidos.