

# sesion6\_formato\_tablas

Fabiana Hidalgo Fallas

2023-06-27

## Uso del paquete gt para estilo de cuadros.

Cuando comunicamos nuestros resultados, es importante que tengamos en cuenta el formato de los cuadros que estamos generando. Existen varios paquetes de R que nos pueden ayudar a estilizar nuestras tablas, por ejemplo:

- gt
- DT
- kableExtra
- flextable

En esta clase aprenderemos cómo trabajar con gt.

Primero, vamos a crear un dataframe a partir del dataset islands de la librería datasets.

```
df_islas <- tibble(  
  nombre = names(datasets::islands),  
  tamaño = islands  
)
```

De este dataframe, vamos a extraer las 10 masas terrestres más grandes del mundo y crear una nueva columna que indique si la masa es un Continente, un país o una subregión:

```
# Cargar datos  
datos_islas <- df_islas |>  
  arrange(desc(tamaño)) |>  
  head(10) |>  
  mutate(tipo = c("Continente", "Continente", "Continente", "Continente", "Continente",  
                  "Continente", "País", "País", "Subregión", "Subregión"))
```

Así se ve este dataframe sin hacerle ningún formato:

```
datos_islas
```

```
## # A tibble: 10 x 3  
##   nombre      tamaño tipo  
##   <chr>      <dbl> <chr>  
## 1 Asia      16988 Continente  
## 2 Africa    11506 Continente  
## 3 North America  9390 Continente
```

```
## 4 South America 6795 Continente
## 5 Antarctica 5500 Continente
## 6 Europe 3745 Continente
## 7 Australia 2968 País
## 8 Greenland 840 País
## 9 New Guinea 306 Subregión
## 10 Borneo 280 Subregión
```

Para mejorar el formato de este dataframe con la librería gt, primero convertimos el df a un objeto gt como se muestra a continuación:

```
islas_gt <- datos_islas |>
  gt()
```

Observe que hemos transformado el dataframe en un objeto que parece una lista, como aparece en nuestro entorno. Visualicemos cómo queda hasta este momento:

```
islas_gt
```

nombre	tamano	tipo
Asia	16988	Continente
Africa	11506	Continente
North America	9390	Continente
South America	6795	Continente
Antarctica	5500	Continente
Europe	3745	Continente
Australia	2968	País
Greenland	840	País
New Guinea	306	Subregión
Borneo	280	Subregión

No está mal pero queda muy básico. Con gt podemos añadirle varias partes a nuestra tabla, como títulos, subtítulos, notas al pie, referencias y más.

## Agregar títulos y subtítulos

La forma en que agregamos partes como el encabezado de la tabla y las notas al pie en el pie de tabla es utilizando la familia de funciones `tab_*`(`.`). Es fácil agregar un encabezado de tabla, así que veamos cómo se ve la tabla anterior con un título y un subtítulo. Podemos agregar esta parte usando la función `tab_header()`.

```
islas_gt <- datos_islas |>
  gt() |>
  tab_header(
    title = "Grandes masas terrestres del mundo",
    subtitle = "Se presentan las 10 más grandes" # es opcional
  )
```

Veamos cómo se ve la tabla con el título y subtítulo:

```
islas_gt
```

## Grandes masas terrestres del mundo

Se presentan las 10 más grandes

nombre	tamano	tipo
Asia	16988	Continente
Africa	11506	Continente
North America	9390	Continente
South America	6795	Continente
Antarctica	5500	Continente
Europe	3745	Continente
Australia	2968	País
Greenland	840	País
New Guinea	306	Subregión
Borneo	280	Subregión

## Mover posición de columnas y ajustar sus nombres

Suponga que deseamos reacomodar las columnas de nuestra tabla. Podemos hacerlo antes de convertir el dataframe en un objeto gt (usando la función `relocate` de dplyr), o podemos hacerlo directamente con gt:

```
islas_gt <- islas_gt |>
  cols_move_to_start(
    columns = c(nombre, tipo, tamano)
  )

islas_gt
```

## Grandes masas terrestres del mundo

Se presentan las 10 más grandes

nombre	tipo	tamano
Asia	Continente	16988
Africa	Continente	11506
North America	Continente	9390
South America	Continente	6795
Antarctica	Continente	5500
Europe	Continente	3745
Australia	País	2968
Greenland	País	840
New Guinea	Subregión	306
Borneo	Subregión	280

## Añadir `tab_spanners`

Los tabs spanner sirven para agrupar varias columnas en grupos. Esto no siempre es necesario pero en algunos casos puede ayudarnos a mejorar la presentación de la información. Se crean con la función `tab_spanner`:

```
islas_gt <- islas_gt |>
  tab_spanner(
    label = "Información política",
    columns = c(nombre, tipo)
  ) |>
  tab_spanner(
    label = "Información geográfica",
    columns = c(tamano)
  )
```

## Añadir Nota de Origen y pie de tabla (source\_note)

Se puede agregar una nota de origen al pie de página de la tabla mediante el uso de la función `tab_source_note()`. La idea es que esto nos permite indicar la fuente de nuestros datos. Funciona de la misma manera que `tab_header()` (también permite entradas de Markdown), excepto que se puede llamar varias veces: cada invocación da como resultado la adición de una nota de origen.

```
islas_gt <- islas_gt |>
  tab_source_note(
    source_note = "Fuente: The World Almanac and Book of Facts, 1975, page 406."
  ) |>
  tab_source_note(
    source_note = "McNeil, D. R. (1977) Interactive Data Analysis. Wiley."
  )

# Para ver el resultado:
islas_gt
```

Grandes masas terrestres del mundo  
Se presentan las 10 más grandes

Información política		Información geográfica
nombre	tipo	tamano
Asia	Continente	16988
Africa	Continente	11506
North America	Continente	9390
South America	Continente	6795
Antarctica	Continente	5500
Europe	Continente	3745
Australia	País	2968
Greenland	País	840
New Guinea	Subregión	306
Borneo	Subregión	280

Fuente: The World Almanac and Book of Facts, 1975, page 406.  
McNeil, D. R. (1977) Interactive Data Analysis. Wiley.

Las notas al pie se encuentran dentro de la parte del Pie de tabla y sus marcas de referencia están adjuntas a los datos de las celdas. Las notas al pie se agregan con la función `tab_footnote()`.

Dependiendo de dónde queremos añadir el pie de página, deberemos usar una función distinta para indicar el location de ese pie.

Aquí tenemos un ejemplo sencillo de cómo se pueden agregar notas al pie a una celda de tabla. Añadamos una nota al pie que haga referencia a las celdas de América del Norte y América del Sur en la columna de nombres:

```
islas_gt <- islas_gt |>
  tab_footnote(
    footnote = "Las Américas", # es lo que va a decir el pie de tabla
    locations = cells_body(columns = nombre, rows = 3:4) # acá indicamos la referencia al pie de tabla
  )
```

Veamos cómo se ve esto:

```
islas_gt
```

## Grandes masas terrestres del mundo

Se presentan las 10 más grandes

Información política		Información geográfica
nombre	tipo	tamano
Asia	Continente	16988
Africa	Continente	11506
North America <sup>1</sup>	Continente	9390
South America <sup>1</sup>	Continente	6795
Antarctica	Continente	5500
Europe	Continente	3745
Australia	País	2968
Greenland	País	840
New Guinea	Subregión	306
Borneo	Subregión	280

<sup>1</sup>Las Américas

Fuente: The World Almanac and Book of Facts, 1975, page 406.

McNeil, D. R. (1977) Interactive Data Analysis. Wiley.

Si queremos añadir una nota al pie directamente en el nombre de una columna, en locations usamos la función

```
islas_gt <- islas_gt |>
  tab_footnote(
    footnote = "Tamaño en miles de millas cuadradas", #lo que va a decir al pie de la tabla
    locations = cells_column_labels(columns = tamano), #la referencia al pie
    placement = "right"
  )
```

```
islas_gt
```

## Grandes masas terrestres del mundo

Se presentan las 10 más grandes

Información política	Información geográfica
----------------------	------------------------

nombre	tipo	tamano <sup>1</sup>
Asia	Continente	16988
Africa	Continente	11506
North America <sup>2</sup>	Continente	9390
South America <sup>2</sup>	Continente	6795
Antarctica	Continente	5500
Europe	Continente	3745
Australia	País	2968
Greenland	País	840
New Guinea	Subregión	306
Borneo	Subregión	280

<sup>1</sup>Tamaño en miles de millas cuadradas

<sup>2</sup>Las Américas

Fuente: The World Almanac and Book of Facts, 1975, page 406.

McNeil, D. R. (1977) Interactive Data Analysis. Wiley.

Ejercicios para la observacion 1688 que diga “El tamaño más grande”

```
islas_gt <- islas_gt |>
tab_footnote(
  footnote = "El tamaño más grande",
  locations = cells_body(columns = tamano, rows = 1 ))

islas_gt
```

## Grandes masas terrestres del mundo

Se presentan las 10 más grandes

Información política		Información geográfica
nombre	tipo	tamano <sup>1</sup>
Asia	Continente	<sup>2</sup> 16988
Africa	Continente	11506
North America <sup>3</sup>	Continente	9390
South America <sup>3</sup>	Continente	6795
Antarctica	Continente	5500
Europe	Continente	3745
Australia	País	2968
Greenland	País	840
New Guinea	Subregión	306
Borneo	Subregión	280

<sup>1</sup>Tamaño en miles de millas cuadradas

<sup>2</sup>El tamaño más grande

<sup>3</sup>Las Américas

Fuente: The World Almanac and Book of Facts, 1975, page 406.

McNeil, D. R. (1977) Interactive Data Analysis. Wiley.

En resumen, en location usamos:

- cells\_body si la nota va dentro del cuerpo de la tabla
- cells\_column\_labels si la nota va en el nombre de una columna

Vea la ayuda de tab\_footnote para más detalles.

## Añadir colores

Podemos hacer más cosas interesantes con `gt`. Por ejemplo, digamos que queremos añadir colores en la columna tamaño. Esto lo podemos hacer usando la función `data_color`. OJO: los colores pueden verse cuando se renderiza en docx o html, pero no en pdf.

```
islas_gt <- islas_gt |>
  data_color(
    columns = tamaño,
    palette = c("yellow", "orange", "navyblue")
  )

islas_gt
```

Grandes masas terrestres del mundo  
Se presentan las 10 más grandes

Información política		Información geográfica
nombre	tipo	tamaño <sup>1</sup>
Asia	Continente	<sup>2</sup> 16988
Africa	Continente	11506
North America <sup>3</sup>	Continente	9390
South America <sup>3</sup>	Continente	6795
Antarctica	Continente	5500
Europe	Continente	3745
Australia	País	2968
Greenland	País	840
New Guinea	Subregión	306
Borneo	Subregión	280

<sup>1</sup>Tamaño en miles de millas cuadradas

<sup>2</sup>El tamaño más grande

<sup>3</sup>Las Américas

Fuente: The World Almanac and Book of Facts, 1975, page 406.

McNeil, D. R. (1977) Interactive Data Analysis. Wiley.

## Ajustar los nombres de las columnas:

Podemos hacer esto con la función `cols_label`:

```
islas_gt <- islas_gt |>
  cols_label(
    nombre = "Nombre",
    tipo = "Tipo",
    tamaño = "Area"
  )

islas_gt
```

## Grandes masas terrestres del mundo

Se presentan las 10 más grandes

Información política		Información geográfica
Nombre	Tipo	Area <sup>1</sup>
Asia	Continente	<sup>2</sup> 16988
Africa	Continente	11506
North America <sup>3</sup>	Continente	9390
South America <sup>3</sup>	Continente	6795
Antarctica	Continente	5500
Europe	Continente	3745
Australia	País	2968
Greenland	País	840
New Guinea	Subregión	306
Borneo	Subregión	280

<sup>1</sup>Tamaño en miles de millas cuadradas

<sup>2</sup>El tamaño más grande

<sup>3</sup>Las Américas

Fuente: The World Almanac and Book of Facts, 1975, page 406.

McNeil, D. R. (1977) Interactive Data Analysis. Wiley.

Práctica en grupos

Para los datos de airquality, realice lo siguiente:

- Cargue los datos

```
datos_calidad_aire <- datasets::airquality
```

- Obtenga un cuadro con los 10 registros con las temperaturas reportadas más altas

```
Datos_aire <- datos_calidad_aire |>
  arrange(desc(Temp)) |>
  head(10)
```

- Agregue títulos, subtítulos, fuente y pie de tabla al cuadro obtenido. Recuerde usar ?airquality para obtener información sobre el origen de los datos.

```
Datos_aire <- Datos_aire |>
  gt() |>
  tab_header(
    title = "New York Air Quality Measurements",
    subtitle = "Daily air quality measurements in New York, May to September 1973"
  )

Datos_aire
```



## New York Air Quality Measurements

Daily air quality measurements in New York, May to September 1973

Ozone	Solar.R	Wind	Temp	Month	Day
76	203	9.7	97	8	28
84	237	6.3	96	8	30
118	225	2.3	94	8	29
85	188	6.3	94	8	31
NA	259	10.9	93	6	11
73	183	2.8	93	9	3
91	189	4.6	93	9	4
NA	250	9.2	92	6	12
97	267	6.3	92	7	8
97	272	5.7	92	7	9

```
Datos_aire <- Datos_aire |>
  tab_source_note(
    source_note = "Fuente: The data were obtained from the New York State Department of Conservation (o
  ) |>
  tab_source_note(
    source_note = "Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Met
  )

Datos_aire
```

## New York Air Quality Measurements

Daily air quality measurements in New York, May to September 1973

Ozone	Solar.R	Wind	Temp	Month	Day
76	203	9.7	97	8	28
84	237	6.3	96	8	30
118	225	2.3	94	8	29
85	188	6.3	94	8	31
NA	259	10.9	93	6	11
73	183	2.8	93	9	3
91	189	4.6	93	9	4
NA	250	9.2	92	6	12
97	267	6.3	92	7	8
97	272	5.7	92	7	9

Fuente: The data were obtained from the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data)

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Belmont, CA: Wadsworth

```
Datos_aire <- Datos_aire |>
  tab_footnote(
    footnote = "Temperatura más alta", # es lo que va a decir el pie de tabla
    locations = cells_body(columns = Temp, rows = 1) # acá indicamos la referencia al pie de tabla
  )
Datos_aire
```

## New York Air Quality Measurements

Daily air quality measurements in New York, May to September 1973

Ozone	Solar.R	Wind	Temp	Month	Day
76	203	9.7	<sup>1</sup> 97	8	28
84	237	6.3	96	8	30
118	225	2.3	94	8	29
85	188	6.3	94	8	31
NA	259	10.9	93	6	11
73	183	2.8	93	9	3
91	189	4.6	93	9	4
NA	250	9.2	92	6	12
97	267	6.3	92	7	8
97	272	5.7	92	7	9

<sup>1</sup>Temperatura más alta

Fuente: The data were obtained from the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data)

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Belmont, CA: Wadsworth

- d. Además, agrúpelas columnas Ozone, Solar.R y Wind bajo un `tab_spanner` llamada “Mediciones”, y las columnas Temp, Month y Day bajo un `tab_spanner` llamado “Tiempo”.

```
Datos_aire <- Datos_aire |>
  tab_spanner(
    label = "Mediciones",
    columns = c(Ozone, Solar.R, Wind)
  ) |>
  tab_spanner(
    label = "Tiempo",
    columns = c(Temp, Month, Day)
  )
Datos_aire
```

## New York Air Quality Measurements

Daily air quality measurements in New York, May to September 1973

Mediciones			Tiempo		
Ozone	Solar.R	Wind	Temp	Month	Day
76	203	9.7	<sup>1</sup> 97	8	28
84	237	6.3	96	8	30
118	225	2.3	94	8	29
85	188	6.3	94	8	31
NA	259	10.9	93	6	11
73	183	2.8	93	9	3
91	189	4.6	93	9	4
NA	250	9.2	92	6	12
97	267	6.3	92	7	8
97	272	5.7	92	7	9

<sup>1</sup>Temperatura más alta

Fuente: The data were obtained from the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data)

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Belmont, CA: Wadsworth

- e. Renombre las columnas para que estén en español y coloque colores a su gusto a la columna Temperatura.

```
Datos_aire <- Datos_aire |>
  cols_label(
    Ozone = "Ozono",
    Solar.R = "Radiacion Solar",
    Wind = "Viento",
    Temp = "Temperatura",
    Month = "Mes",
    Day = "Dia"
  )
```

Datos\_aire

## New York Air Quality Measurements

Daily air quality measurements in New York, May to September 1973

Mediciones			Tiempo		
Ozono	Radiacion Solar	Viento	Temperatura	Mes	Dia
76	203	9.7	<sup>1</sup> 97	8	28
84	237	6.3	96	8	30
118	225	2.3	94	8	29
85	188	6.3	94	8	31
NA	259	10.9	93	6	11
73	183	2.8	93	9	3
91	189	4.6	93	9	4
NA	250	9.2	92	6	12
97	267	6.3	92	7	8
97	272	5.7	92	7	9

<sup>1</sup>Temperatura más alta

Fuente: The data were obtained from the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data)

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Belmont, CA: Wadsworth

Paleta de colores

```
Datos_aire <- Datos_aire |>
  data_color(
    columns = Temp,
    palette = c("red", "pink", "purple")
  )
```

Datos\_aire

# New York Air Quality Measurements

Daily air quality measurements in New York, May to September 1973

Mediciones			Tiempo		
Ozono	Radiacion Solar	Viento	Temperatura	Mes	Dia
76	203	9.7	<sup>1</sup> 97	8	28
84	237	6.3	96	8	30
118	225	2.3	94	8	29
85	188	6.3	94	8	31
NA	259	10.9	93	6	11
73	183	2.8	93	9	3
91	189	4.6	93	9	4
NA	250	9.2	92	6	12
97	267	6.3	92	7	8
97	272	5.7	92	7	9

<sup>1</sup>Temperatura más alta

Fuente: The data were obtained from the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data)

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Belmont, CA: Wadsworth