

Visualización de datos con `{ggplot2}`

Principales conceptos y funciones para visualizar datos con R

 Estación R

2024-03-20

Bienvenidos y bienvenidas a Estación R

 Linktree

 Web

¿Qué vimos?

- ✓ Conceptos básicos de R
- ✓ Pensar un proyecto de datos con R
- ✓ Procesamiento de datos con `{tidyverse}`

Hoja de Ruta

📌 ¿Por qué visualizar datos?

📌 Gramática de los gráficos y `{ggplot2}`

- Capas y el operador ``+`` (más)

📌 Armandó un gráfico de barras (columnas)

- Función ``geom_col()``

📌 Chapa y pintura de un gráfico (atributos)

Configuración para esta clase

- Armar un proyecto de trabajo nuevo o abrir aquel con el que veníamos trabajando
- Cargar la base del Padrón Único Nacional de Alojamientos (Argentina) y chequear que esté en la carpeta `datos`
- Crear un script de trabajo
- Carga la librería `{tidyverse}`

¿Por qué visualizar datos?

¿Por qué visualizar?

- *“La visualización es el proceso de hacer visibles los contrastes, ritmos y eventos que los datos expresan, que no podemos percibir cuando vienen en forma de áridas listas de números y categorías.”¹*
- Interpretar / decodificar la información de forma visual
- Guiar hacia el hallazgo

ggplot2

Una forma de visualizar



¿Qué es `{ggplot2}`?

- Una implementación del sistema Grammar of graphics (Wilkinson, 2005).
- Un esquema pensado en capas (datos → plano (ejes x e y) → geometrías)
- Un paquete de funciones de aplicación intuitiva.

¿Por qué `{ggplot2}`?

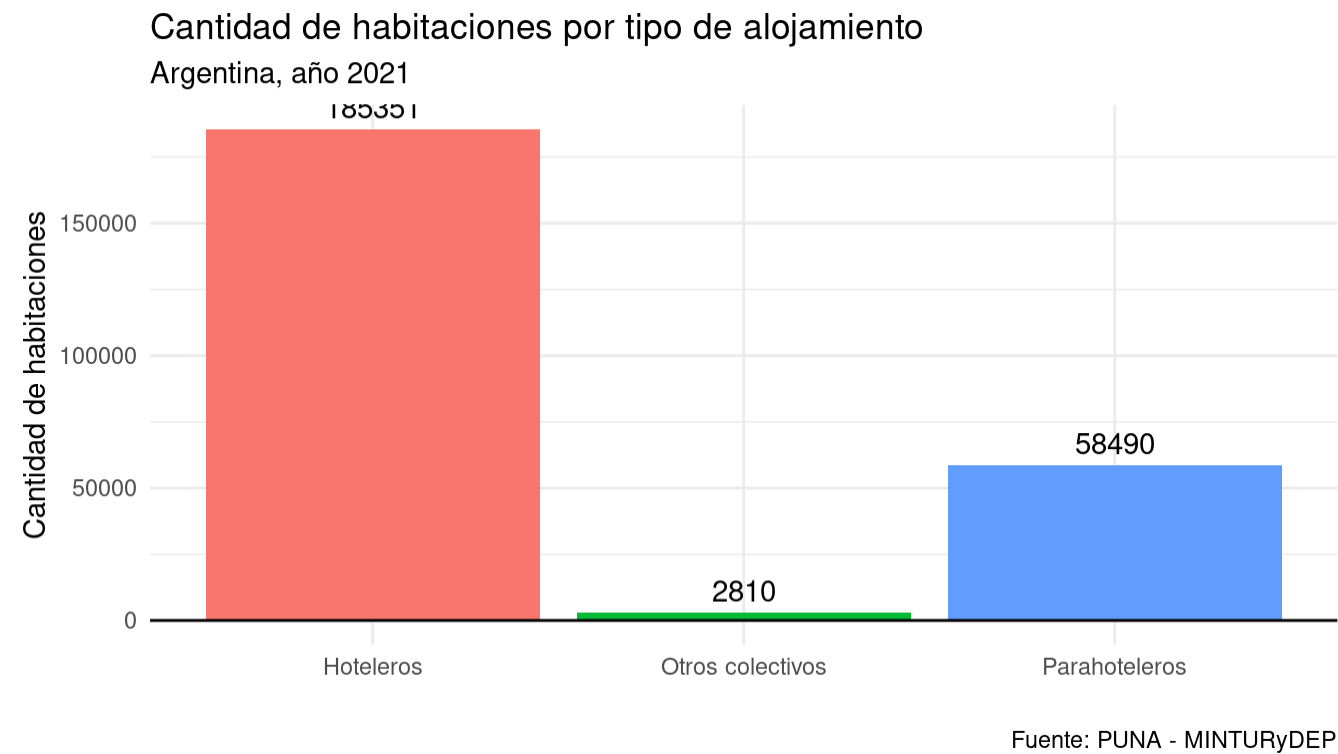
::: incremental

- Tiene un marco de referencia (El grammar of graphics)
- Flexible, con especificaciones a nivel de capas.
- Sistema de `themes`, que permiten *pulir* la apariencia del gráfico
- Decenas de extensiones para ampliar la potencia del paquete
- Comunidad activa y con mucha predisposición a ayudar.

¿A dónde vamos?

Graphic

Code



¿Por dónde empezamos?

Cargamos el paquete

```
library(ggplot2)
```

O...

```
library(tidyverse)
```

Gráfico en clave de capas

3 Capas son las indispensables al pensar nuestro gráfico:

Gráfico en clave de capas

- Los datos (argumento: `data` =):
 - El dataframe que sirve de insumo

Gráfico en clave de capas

- Los datos (argumento: `data` =):
 - El dataframe que sirve de insumo
- Las aesthetics (función `aes()`):
 - Defino el vínculo entre los datos y las propiedades visuales (ejes x e y, por ej.)

Gráfico en clave de capas

- Los datos (argumento: `data` =):
 - El dataframe que sirve de insumo
- Las aesthetics (función `aes()`):
 - Defino el vínculo entre los datos y las propiedades visuales (ejes x e y, por ej.)
- Las geometrías (función `geom_*()`):
 - La geometría con la que se representan los datos

Gráfico en clave de capas

- Pregunta-problema: Quiero representar la cantidad de habitaciones por tipo de alojamiento, para el año 2022

Preparo los datos:

```
library(tidyverse)
df_puna <- read_csv("data/puna_base_agregada.csv")

df_habitaciones_2022 <- df_puna |>
  filter(indice_tiempo == 2022) |>
  group_by(tipo) |>
  summarise(habitaciones_n = sum(habitaciones, na.rm = T))
```

Gráfico en clave de capas

- Así queda nuestra tabla:

```
df_habitaciones_2022
```

```
# A tibble: 3 × 2
```

	tipo	habitaciones_n
	<chr>	<dbl>
1	Hoteleros	185351
2	Otros colectivos	2810
3	Parahoteleros	58490

Gráfico en clave de capas

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n))
```

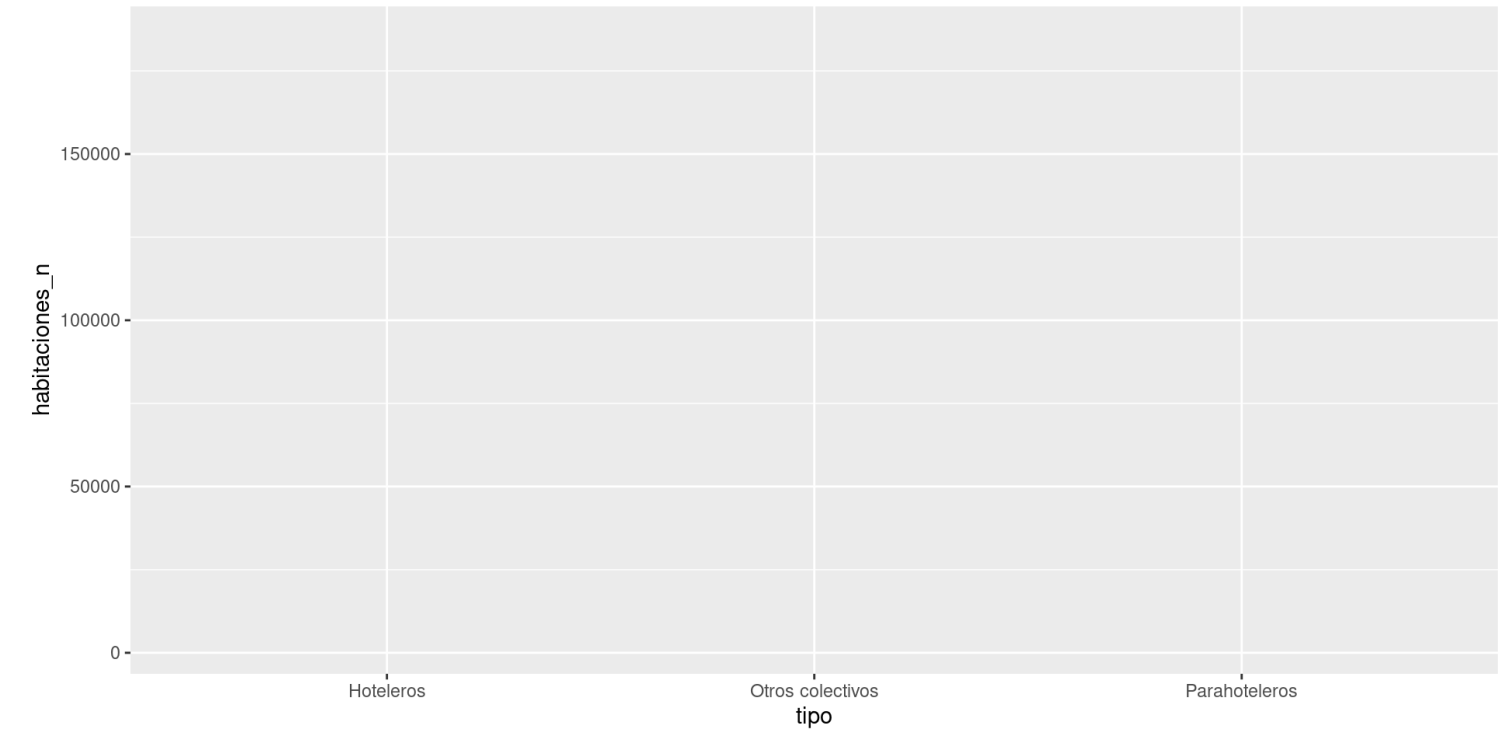
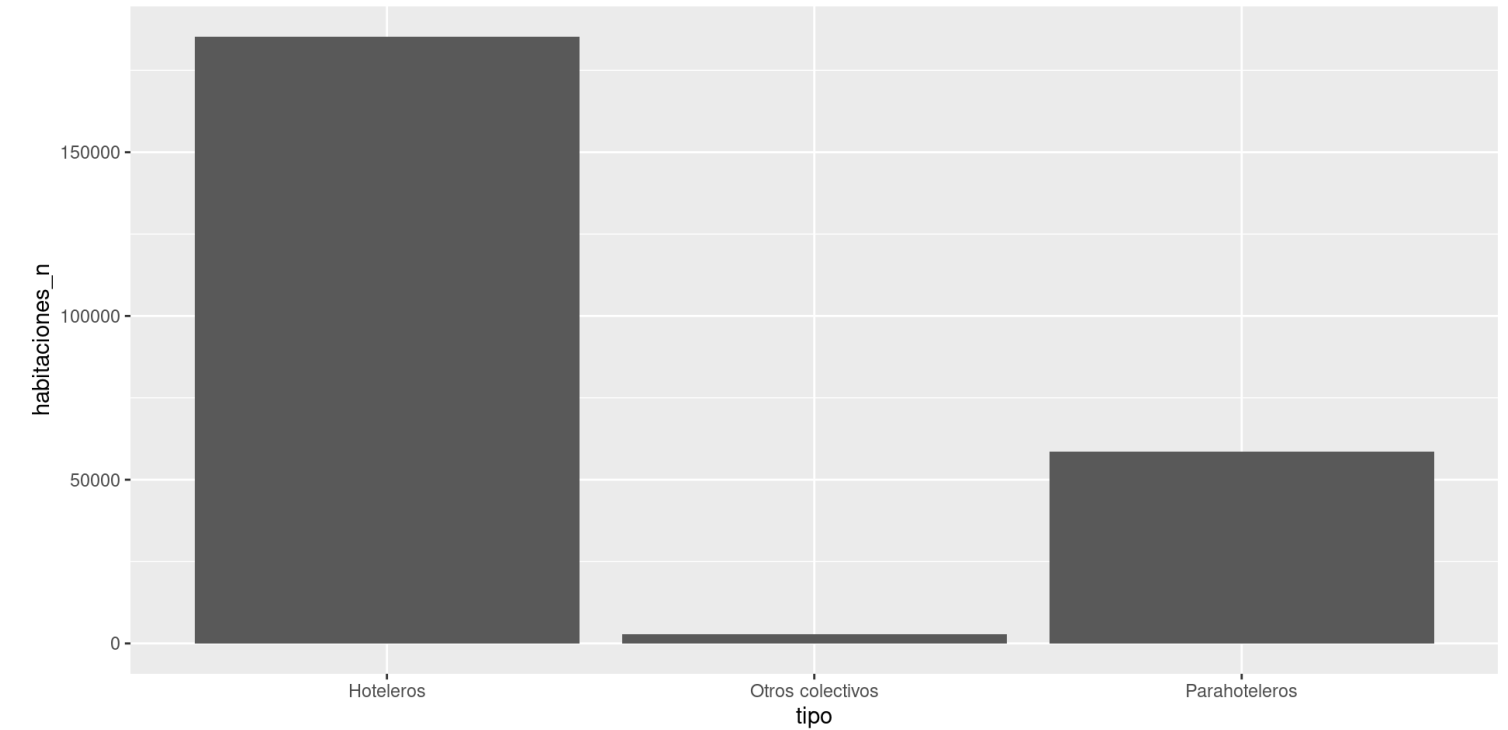


Gráfico en clave de capas

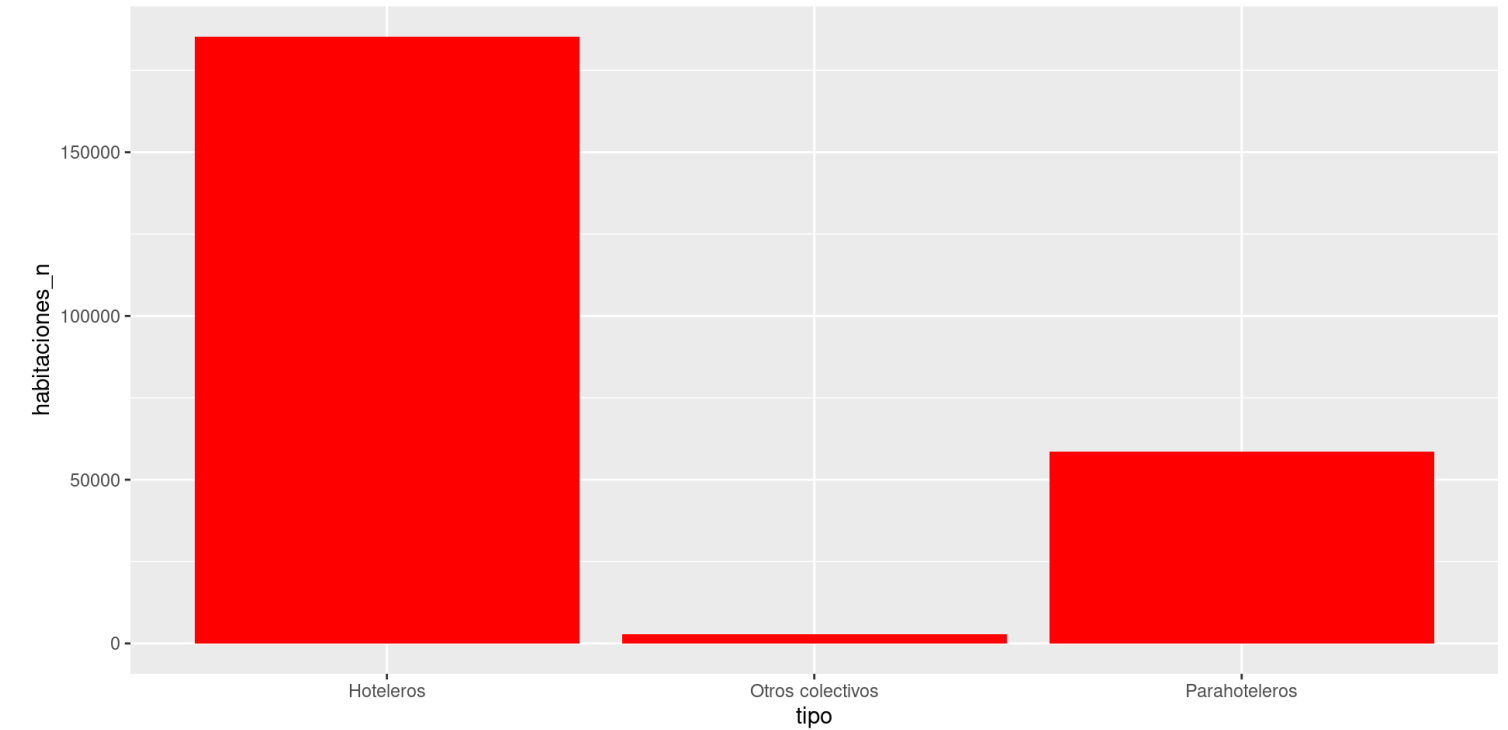
```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col()
```



Chapa y pintura

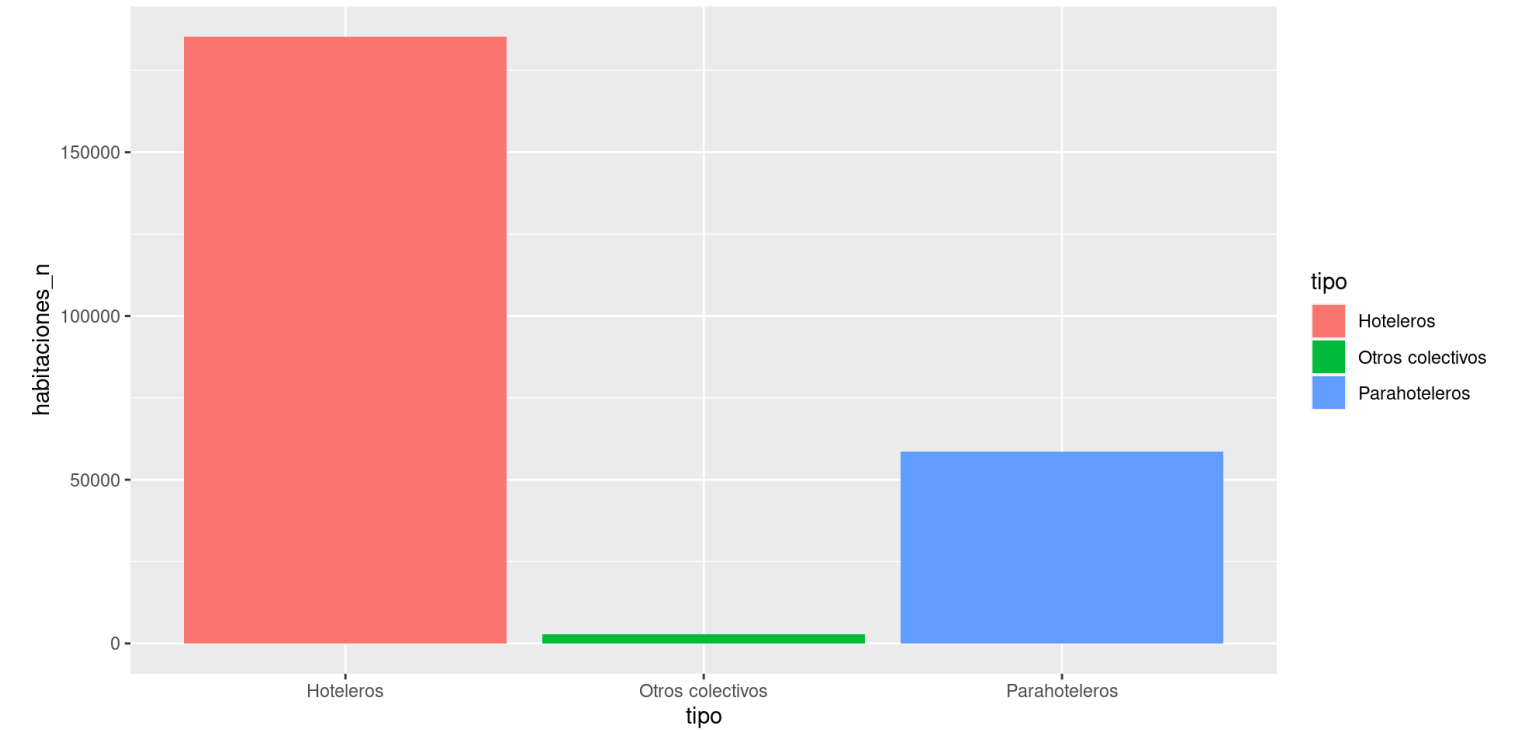
Chapa y pintura - Relleno

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col(fill = "red")
```



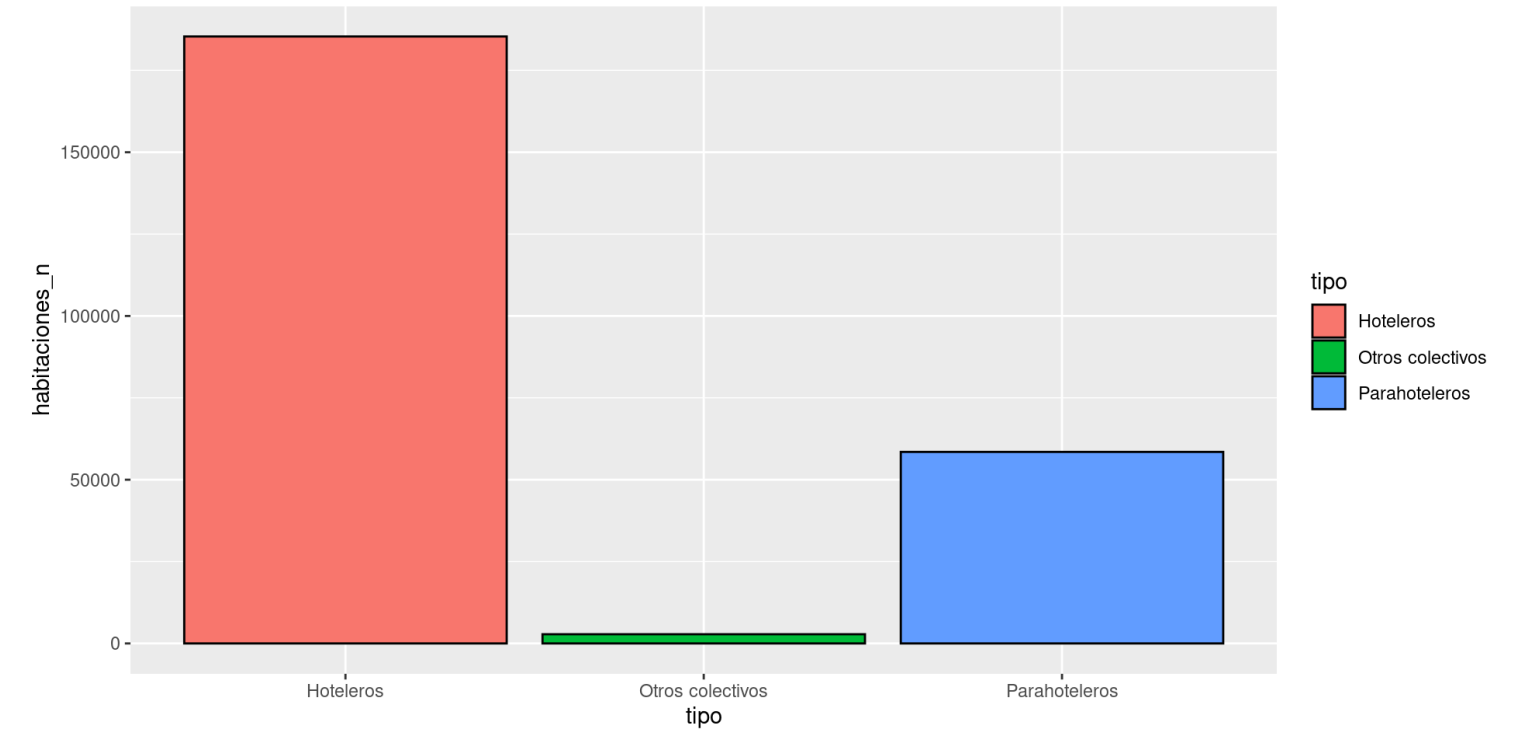
Chapa y pintura - Relleno

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col(aes(fill = tipo))
```



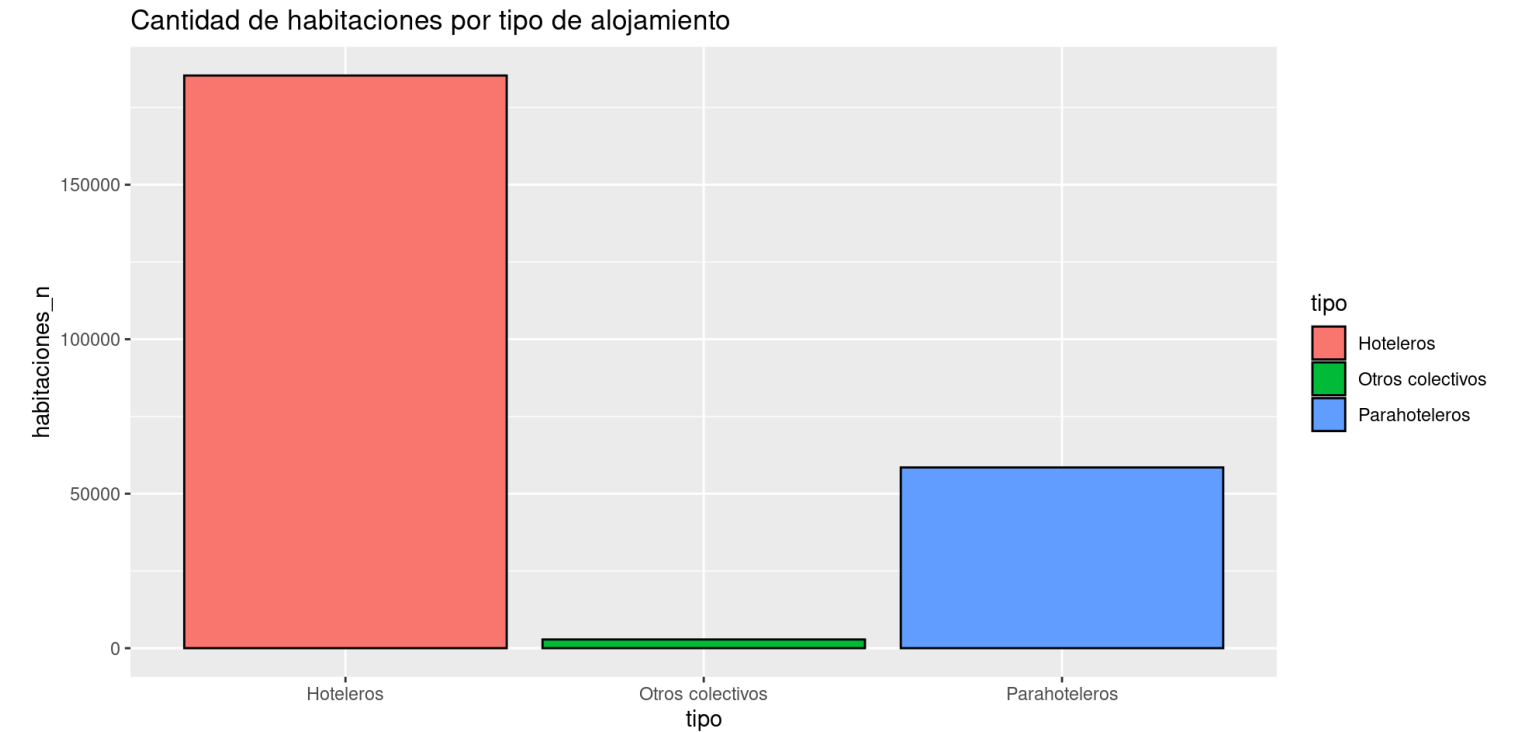
Chapa y pintura - Contorno

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col(aes(fill = tipo),  
4           color = "black")
```



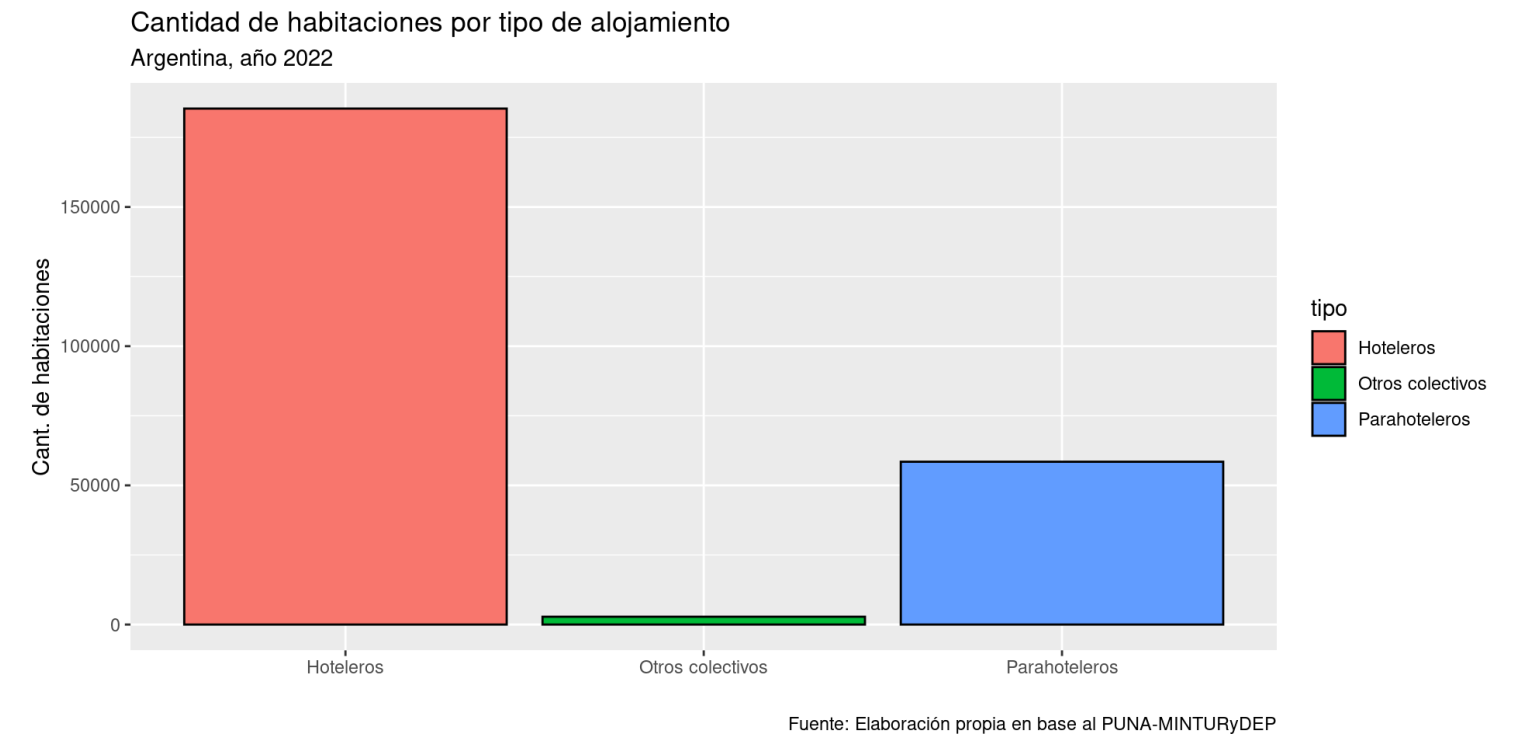
Chapa y pintura - Referencias

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col(aes(fill = tipo),  
4           color = "black") +  
5   labs(title = "Cantidad de habitaciones por tipo de alojam:
```



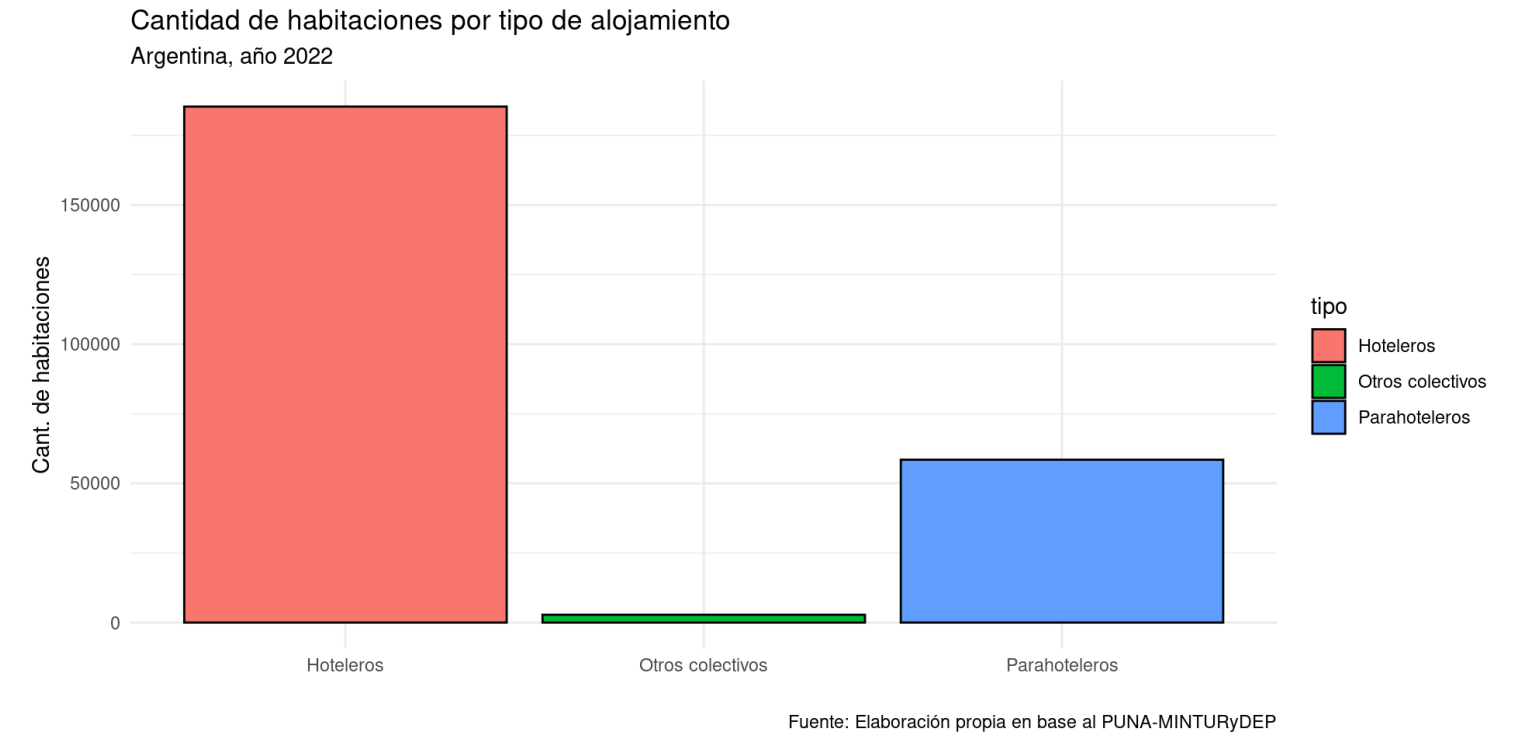
Chapa y pintura - Referencias

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col(aes(fill = tipo),  
4           color = "black") +  
5   labs(title = "Cantidad de habitaciones por tipo de alojam:  
6         subtitle = "Argentina, año 2022",  
7         x = "",  
8         y = "Cant. de habitaciones",  
9         caption = "Fuente: Elaboración propia en base al PUNA-MINTURyDEP")
```



Chapa y pintura - theme

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col(aes(fill = tipo),  
4           color = "black") +  
5   labs(title = "Cantidad de habitaciones por tipo de alojamiento",  
6        subtitle = "Argentina, año 2022",  
7        x = "",  
8        y = "Cant. de habitaciones",  
9        caption = "Fuente: Elaboración propia en base al PUNZ",  
10  theme_minimal())
```



Chapa y pintura - theme

```
1 ggplot(data = df_habitaciones_2022,  
2       aes(x = tipo, y = habitaciones_n)) +  
3   geom_col(aes(fill = tipo),  
4           color = "black") +  
5   labs(title = "Cantidad de habitaciones por tipo de alojamiento",  
6        subtitle = "Argentina, año 2022",  
7        x = "",  
8        y = "Cant. de habitaciones",  
9        caption = "Fuente: Elaboración propia en base al PUNZ",  
10  theme_minimal() +  
11  theme(legend.position = "none")
```

