

Fundamentos de ciencia de datos con R - Módulo 4

Clase 3: Gráficos básicos -histogramas

CEPAL - Unidad de Estadísticas Sociales

2025-11-04

Introducción

Los histogramas son una herramienta fundamental para analizar la distribución de variables numéricas continuas. A diferencia de los gráficos de barras —que se usan para variables categóricas—, los histogramas dividen el rango de una variable continua en intervalos (bins) y muestran cuántas observaciones caen dentro de cada uno.

Introducción

Para esta clase utilizaremos la librería tidyverse, que agrupa varias herramientas fundamentales para el análisis de datos en R. Dentro de ella se encuentra ggplot2, el paquete que usaremos para crear nuestros gráficos de barras.

```
library(tidyverse)
```

Introducción

Nota

“Un buen histograma permite ver lo que una tabla de frecuencias oculta.” —
Adaptado de John Tukey

¿Cuándo usar un histograma?

Usa histogramas cuando:

- ▶ Quieres visualizar la forma de la distribución (simetría, sesgo, concentración, colas).
- ▶ Deseas identificar valores atípicos o rangos de mayor frecuencia.
- ▶ Buscas comparar distribuciones de una variable numérica entre grupos (usando color o facetas).

Evita histogramas cuando:

- ▶ Trabajas con variables categóricas (usa barras).
- ▶ Requieres una curva suavizada de densidad (usa `geom_density()`).

Carga base de datos ejemplo

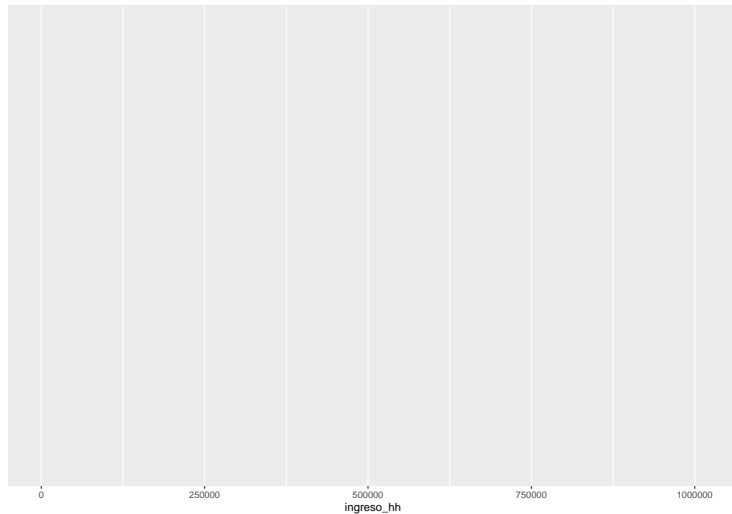
```
datos <- readRDS("../Data/base_personas_gasto.rds")  
head(datos[,2:9], 5)
```

id_pers	upm	estrato	area	fep	pobreza	ingreso_hh	gasto_hh
1	1100100006	11001	1	19	3	10783.05	10783.05
2	1100100006	11001	1	19	3	10783.05	10783.05
1	1100100006	11001	1	16	3	21259.72	21259.72
2	1100100006	11001	1	16	3	21259.72	21259.72
3	1100100006	11001	1	16	3	21259.72	21259.72

Paso a paso: construyendo un histograma (ingreso del hogar)

Paso 1: capa base (solo mapeo estético)

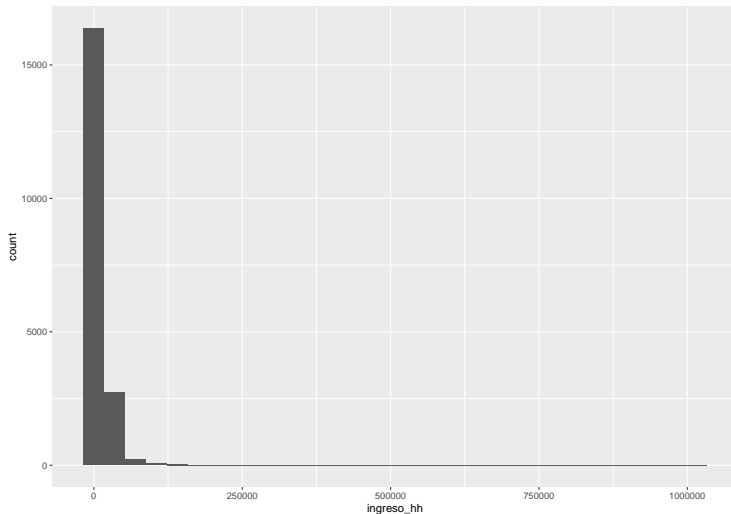
```
ggplot(datos, aes(x = ingreso_hh))
```



Paso a paso: construyendo un histograma (ingreso del hogar)

Paso 2: añadir geometría básica (bins por defecto)

```
ggplot(datos, aes(x = ingreso_hh)) +  
geom_histogram()
```



Paso a paso: construyendo un histograma (ingreso del hogar)

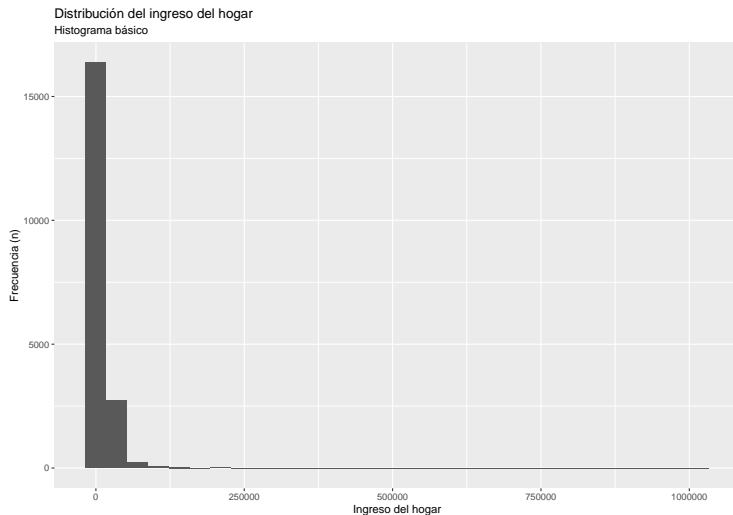
Paso 3: títulos y ejes

```
h_ing_3 <- ggplot(datos, aes(x = ingreso_hh)) +  
  geom_histogram() +  
  labs(  
    title = "Distribución del ingreso del hogar",  
    subtitle = "Histograma básico",  
    x = "Ingreso del hogar", y = "Frecuencia (n)"  
  )
```

Paso a paso: construyendo un histograma (ingreso del hogar)

Paso 3: títulos y ejes

h_ing_3



Paso a paso: construyendo un histograma (ingreso del hogar)

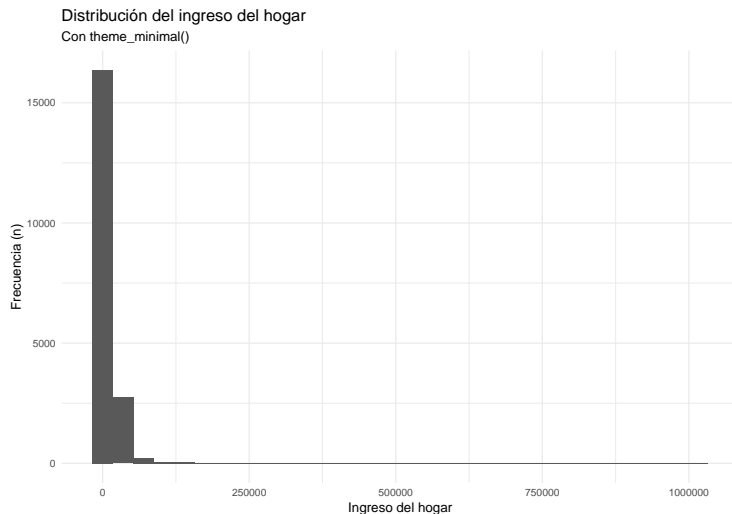
Paso 4: tema y tamaño base

```
h_ing_4 <- ggplot(datos, aes(x = ingreso_hh)) +  
  geom_histogram() +  
  labs(  
    title = "Distribución del ingreso del hogar",  
    subtitle = "Con theme_minimal()",  
    x = "Ingreso del hogar", y = "Frecuencia (n)"  
  ) +  
  theme_minimal(base_size = 13)
```

Paso a paso: construyendo un histograma (ingreso del hogar)

Paso 4: tema y tamaño base

h_ing_4



Paso a paso: construyendo un histograma (ingreso del hogar)

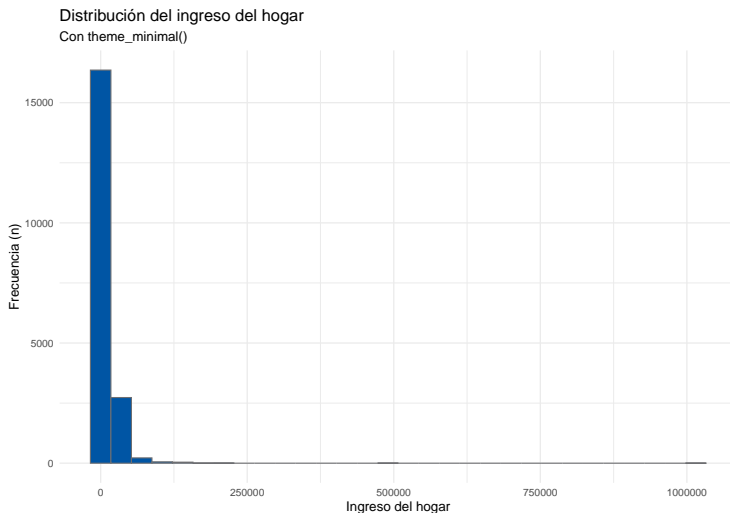
Paso 5: personalizar colores y número de intervalos

```
h_ing_5 <- ggplot(datos, aes(x = ingreso_hh)) +  
  geom_histogram(fill = "#0055A4", color = "#6D6E71") +  
  labs(  
    title = "Distribución del ingreso del hogar",  
    subtitle = "Con theme_minimal()",  
    x = "Ingreso del hogar", y = "Frecuencia (n)"  
  ) +  
  theme_minimal(base_size = 13)
```

Paso a paso: construyendo un histograma (ingreso del hogar)

Paso 5: personalizar colores y número de intervalos

h_ing_5



Ejemplos adicionales de histogramas (ingreso)

Densidad en lugar de conteos + curva de densidad

```
ejem_ing_1 <- ggplot(datos, aes(x = ingreso_hh)) +  
  geom_histogram(aes(y = after_stat(density)), bins = 70,  
    fill = "#A50034", color = "white", alpha = 0.6) +  
  geom_density(color = "#0055A4", linewidth = 1) +  
  labs(  
    title = "Ingreso del hogar: histograma (densidad) + curva",  
    x = "Ingreso del hogar", y = "Densidad"  
  ) +  
  theme_minimal(base_size = 13)
```



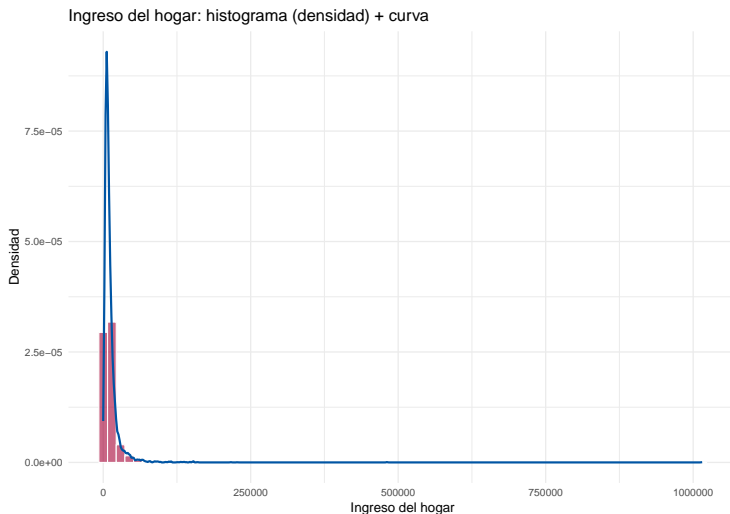
Tip

Ajusta bins para controlar la resolución. Prueba 20–60 bins y quédate con el que mejor muestre la forma sin “ruido”.

Ejemplos adicionales de histogramas (ingreso)

Densidad en lugar de conteos + curva de densidad

`ejem_ing_1`



Ejemplos adicionales de histogramas (ingreso)

Comparar distribuciones por grupo (relleno por pobreza, superpuesto)

```
ejem_ing_2 <- ggplot(datos, aes(x = ingreso_hh, fill = pobreza)) +  
  geom_histogram(position = "identity", alpha = 0.5, bins = 70,  
                 color = "white") +  
  labs(  
    title = "Distribución del ingreso por categoría de pobreza",  
    x = "Ingreso del hogar", y = "Frecuencia (n)", fill = "Pobreza"  
  ) +  
  theme_minimal(base_size = 13)
```

Ejemplos adicionales de histogramas (ingreso)

***Comparar distribuciones por grupo (relleno por pobreza, superpuesto)**

`ejem_ing_2`

