

Fundamentos de ciencia de datos con R - Módulo 4

Clase 5: Gráficos básicos - líneas

CEPAL - Unidad de Estadísticas Sociales

2025-11-05

Introducción

Los gráficos de líneas son una de las formas más comunes de visualizar la evolución de una variable numérica a lo largo del tiempo o de un orden lógico. Permiten identificar tendencias, ciclos y variaciones entre categorías o grupos.

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

“Las líneas conectan puntos, pero también conectan historias: nos permiten ver el cambio en el tiempo.”

¿Cuándo usar un gráfico de líneas?

Usa líneas cuando:

- ▶ Quieres mostrar la evolución de una variable en el tiempo (años, meses, periodos, etc.).
- ▶ Deseas comparar tendencias de varios grupos o categorías.
- ▶ La variable del eje X tiene orden lógico o cronológico.

Evítalo cuando:

- ▶ Los datos son categóricos sin orden (usa barras).
- ▶ Hay pocos puntos o intervalos irregulares (considera puntos o áreas).
- ▶ El eje X no tiene sentido secuencial (podría inducir interpretaciones erróneas).

¿Cuándo usar un gráfico de líneas?



Advertencia

- ▶ Usa líneas **solo** cuando el eje X tenga **orden temporal o lógico**.
- ▶ Evita conectar categorías sin secuencia (¡podría parecer una tendencia falsa!).
- ▶ Si hay **NA** en los datos, la línea se cortará.

Carga base de datos ejemplo

```
datos <- readRDS("../Data/base_personas_gasto.rds")  
head(datos[,1:8], 5)
```

id_hogar	id_pers	upm	estrato	area	fep	pobreza	ingreso_hh
262	1	1100100006	11001	1	19	3	10783.05
262	2	1100100006	11001	1	19	3	10783.05
265	1	1100100006	11001	1	16	3	21259.72
265	2	1100100006	11001	1	16	3	21259.72
265	3	1100100006	11001	1	16	3	21259.72

Carga base de datos ejemplo

Para este análisis, y dado que necesitamos una variable temporal para graficar líneas, se crea la variable anio de forma simulada (2018–2024). Esta variable es solo para fines didácticos.

```
set.seed(123)
datos <- datos %>%
  dplyr::mutate(anio = sample(2018:2024, dplyr::n(), replace = TRUE))

head(datos[, c("anio", "sexo", "ingreso_hh", "gasto_hh")], 5)
```

anio	sexo	ingreso_hh	gasto_hh
2024	Hombre	10783.05	10783.05
2024	Mujer	10783.05	10783.05
2020	Mujer	21259.72	21259.72
2023	Hombre	21259.72	21259.72
2020	Hombre	21259.72	21259.72

Paso a paso: líneas de ingreso promedio por año

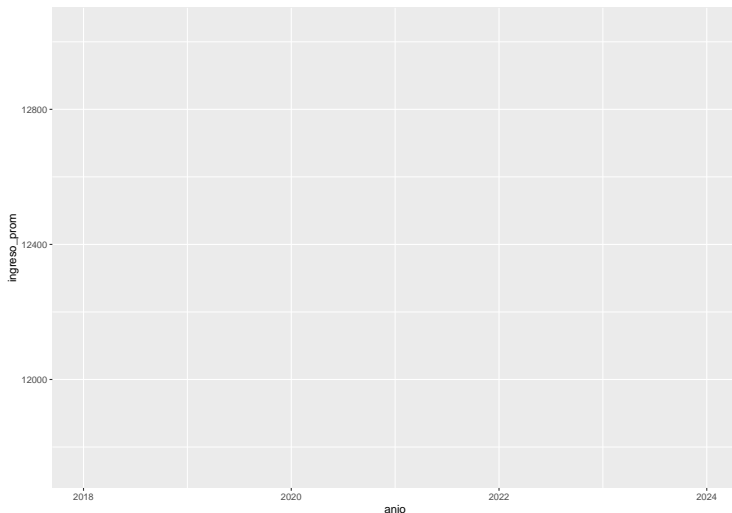
```
serie_ingreso <- datos %>%  
group_by(anio) %>%  
summarise(ingreso_prom = mean(ingreso_hh, na.rm = TRUE), .groups = "drop")  
arrange(anio)  
head(serie_ingreso)
```

anio	ingreso_prom
2018	11835.59
2019	11743.39
2020	11946.41
2021	12143.31
2022	11889.75
2023	12228.08

Paso a paso: líneas de ingreso promedio por año

Paso 1: capa base

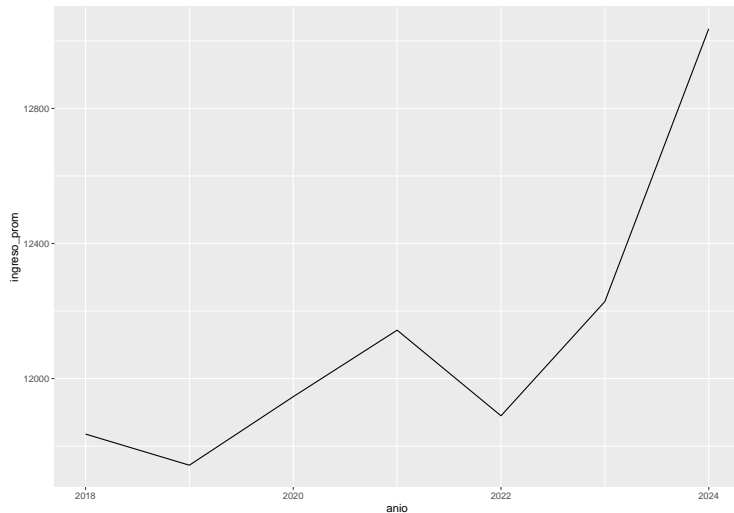
```
ggplot(serie_ingreso, aes(x = anio, y = ingreso_prom))
```



Paso a paso: líneas de ingreso promedio por año

Paso 2: añadir línea

```
ggplot(serie_ingreso, aes(x = anio, y = ingreso_prom)) +  
geom_line()
```



Paso a paso: líneas de ingreso promedio por año

Paso 3: títulos + Color + tema

```
linea_ing_3 <- ggplot(serie_ingreso, aes(x = anio, y = ingreso_prom)) +  
  geom_line(color = "blue") +  
  geom_point(size = 2) +  
  labs(title = "Ingreso del hogar promedio por año",  
        subtitle = "Serie anual (promedio simple)",  
        x = "Año", y = "Ingreso promedio del hogar") +  
  theme_minimal(base_size = 13)
```

Paso a paso: líneas de ingreso promedio por año

Paso 3: títulos + Color + tema

```
linea_ing_3
```

