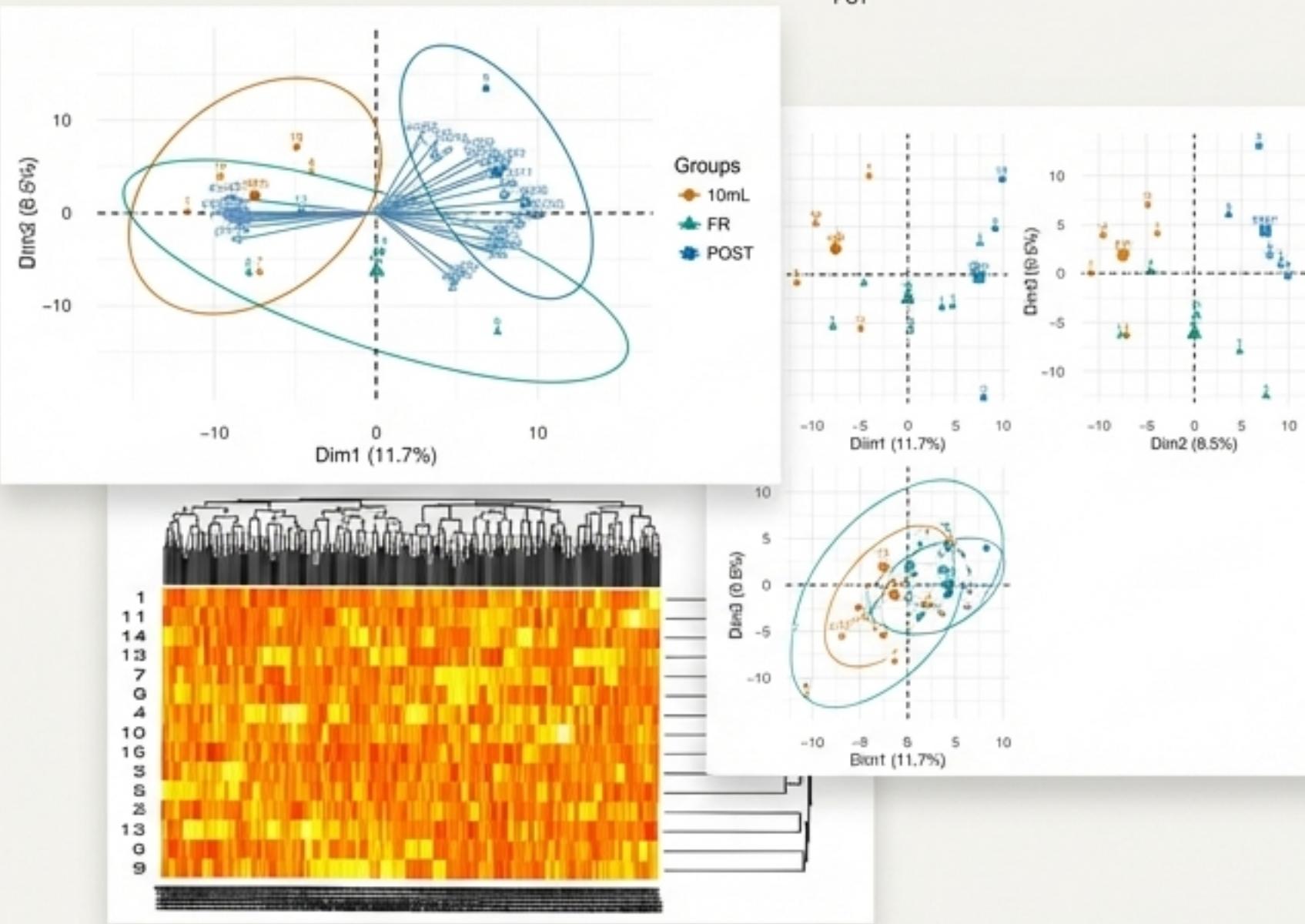


# R



# La Travesía del Análisis de Datos con R

Una guía visual basada en Análisis de datos y métodos estadísticos con R y Análisis espacial con R.

# R es el lenguaje franco que conecta las ciencias con el análisis de datos.



## Es Libre:

Se distribuye bajo licencia GNU. No solo se puede utilizar, sino también mejorar y distribuir libremente.



## Es Multiplataforma:

Con versiones para Linux, Windows, y Mac, funciona en cualquier entorno de trabajo.



## Es Potente:

Su capacidad de análisis estadístico y manejo de datos complejos es de nivel profesional, ideal para la investigación reproducible.



## Capacidad Gráfica Insuperable:

Sus herramientas de visualización, como se verá en este viaje, difícilmente son superadas por otro software estadístico.

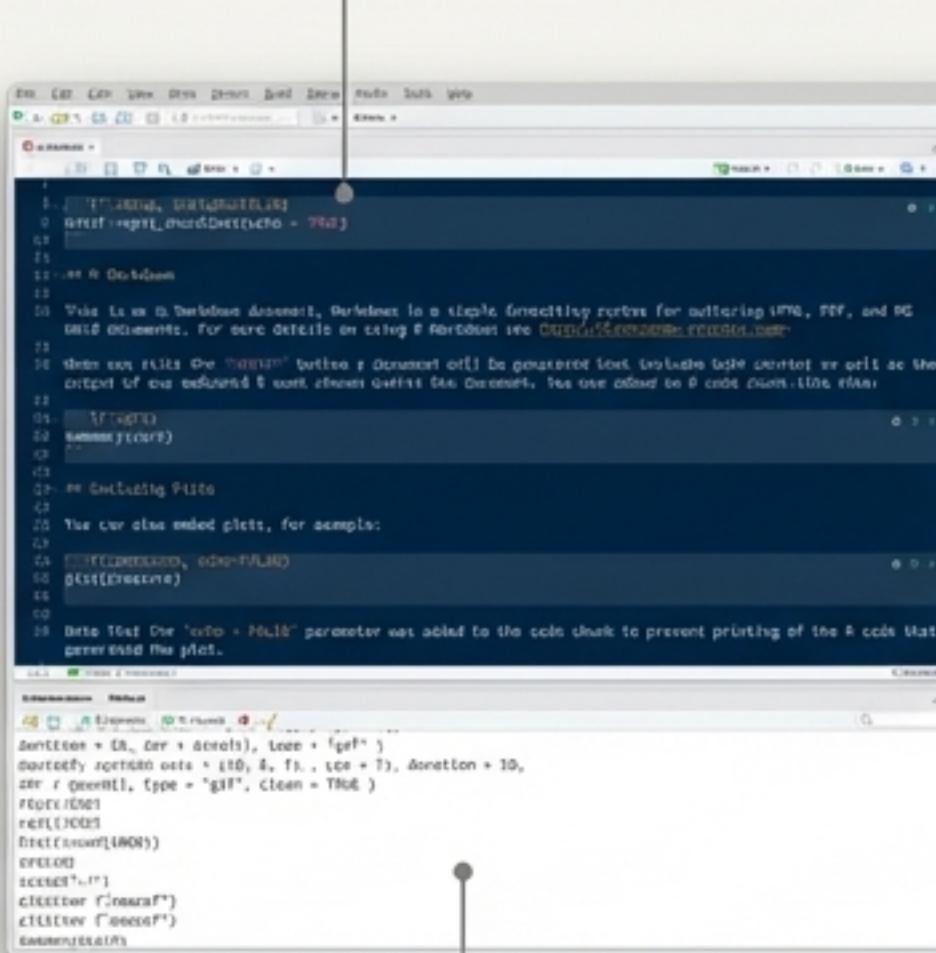


## Una Comunidad Activa:

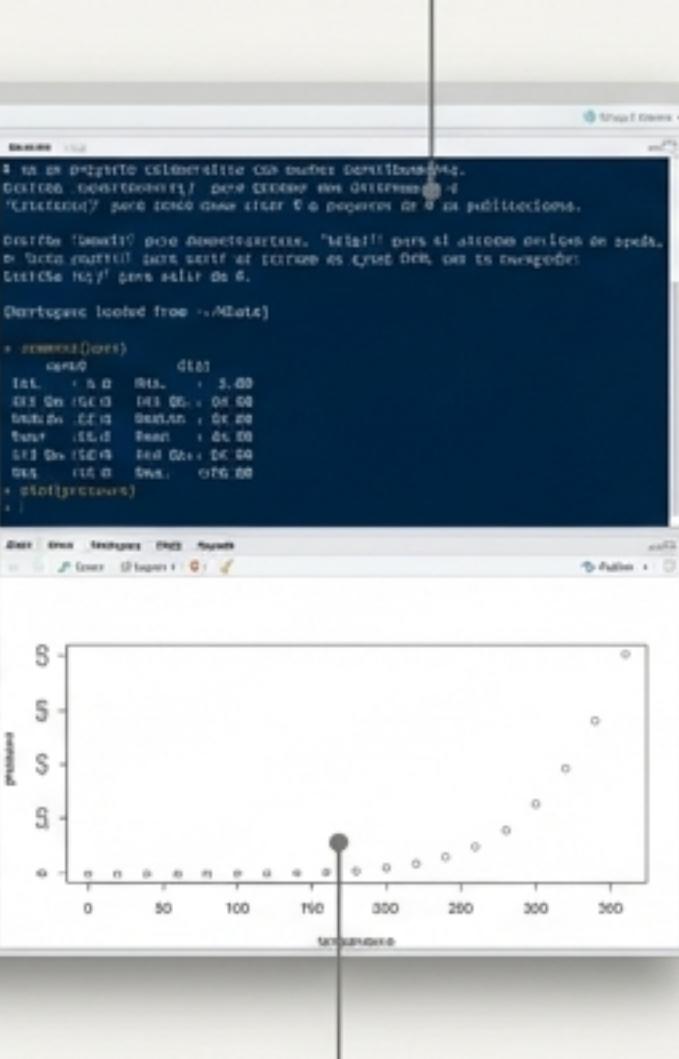
El texto ‘AdymeR’ nace de la comunidad ‘00Rteam’ y el curso ‘FEIR’, demostrando un ecosistema colaborativo que enriquece la herramienta constantemente.

# RStudio es su centro de comando para la exploración de datos.

Editor de Scripts



Entorno / Historial



Consola

Archivos / Gráficos / Ayuda

- Permite trabajar con múltiples scripts simultáneamente.
- Ejecución de código por fragmentos (chunks).
- Visualización integrada del workspace, historial y gráficos.
- Gestión de paquetes y ayuda interactiva.
- Integración con R Markdown para crear reportes dinámicos y reproducibles.

# El `dataframe` es la estructura fundamental para organizar y manipular datos en R.

**¿Qué es un `dataset`?**: Un conjunto de datos ordenado por filas (observaciones) y columnas (variables).

**Anatomía de un `dataframe`**: Una generalización de una matriz donde cada columna puede tener un tipo de dato diferente (numérico, de caracteres, lógico, factor), lo que lo hace ideal para datos del mundo real.

Variables					
Observaciones	manager	date	country	gender	age
1	1	10/11/08	US	♂ M	NA
2	2	10/12/08	US	♀ F	45
3	3	10/13/08	US	♀ M	25
4	4	10/14/08	UK	♂ F	39
5	5	10/11/08	UK	NA	NA

```
# Creación de un dataframe
manager <- c(1:5)
date <- c("10/11/08", "10/12/08", "10/13/08",
         "10/14/08", "10/15/08")
country <- c("US", "US", "UK", "UK", "UK")
gender <- c("M", "F", "F", NA, "F")
age <- c(NA, 45, 25, 39, 99)
df <- data.frame(manager, date, country,
                  gender, age)
```

# R se conecta con prácticamente cualquier fuente de datos.

```
df <- read.table("datos.csv",  
                 header=T, sep=";")
```



```
library(xlsx);  
read.xlsx("fichero.xlsx", 1)
```

```
read.csv("https://example.  
com/data.csv")
```



```
library(foreign);  
read.spss("datos.sav",  
to.data.frame=TRUE)
```



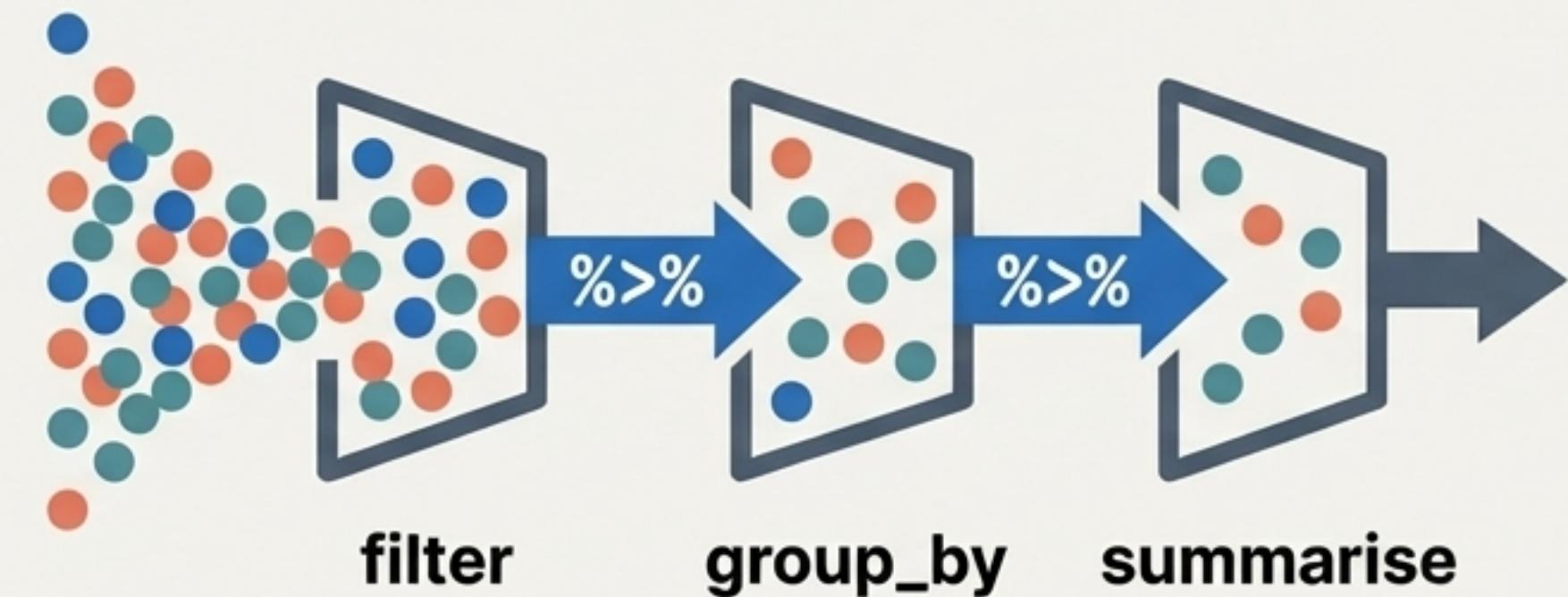
```
library(haven);  
read_dta("datos.dta")
```

# El Tidyverse ofrece una gramática consistente para la manipulación de datos.



Introducción a los cōres del verbos **dplyr**, operator ad el operador del **pipe**:

- **select()**: para elegir columnas.
- **filter()**: para elegir filas.
- **arrange()**: para ordenar filas.
- **mutate()**: para crear nuevas columnas.
- **group\_by()** y **summarise()**: para agrupar y resumir.



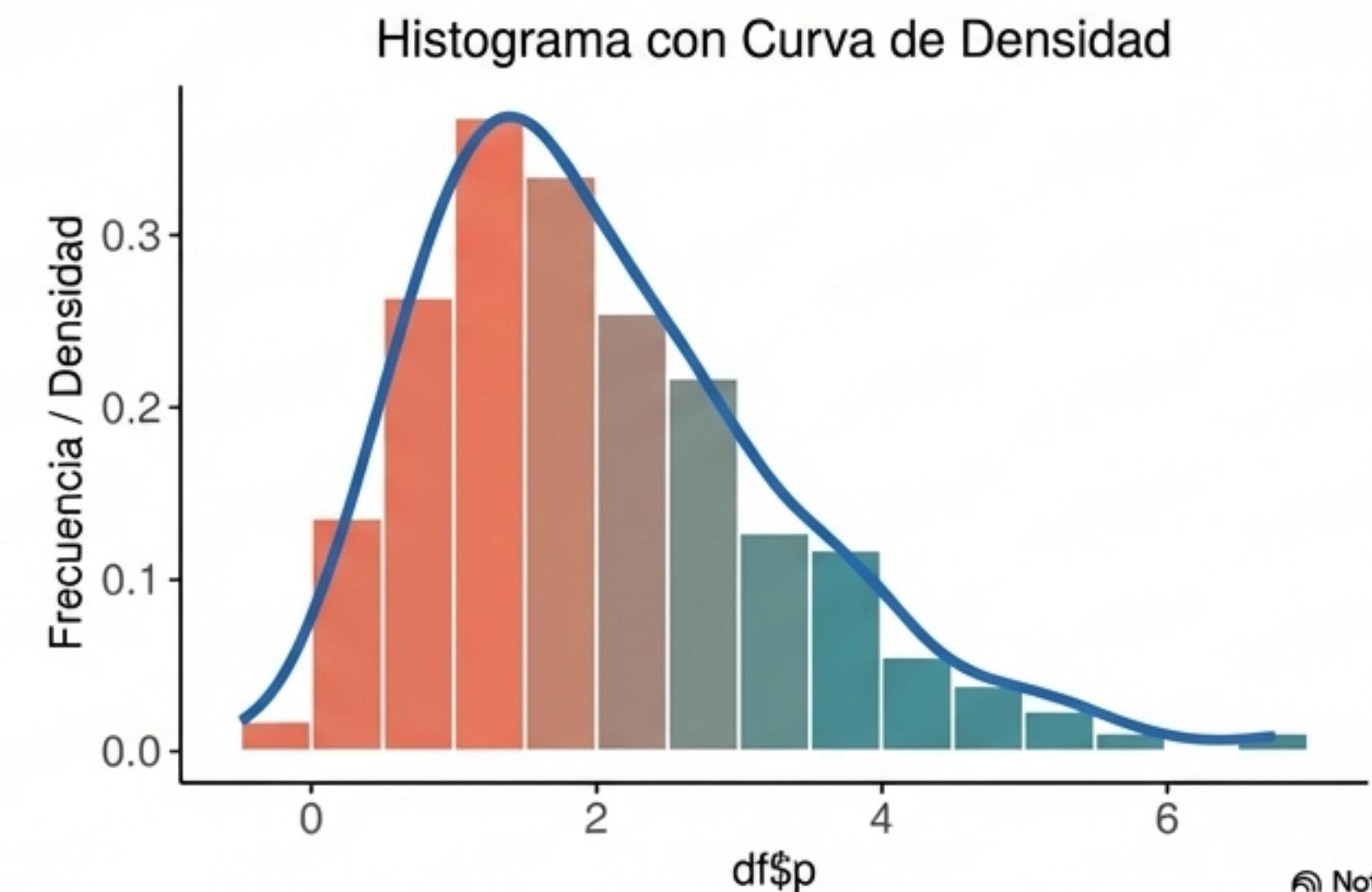
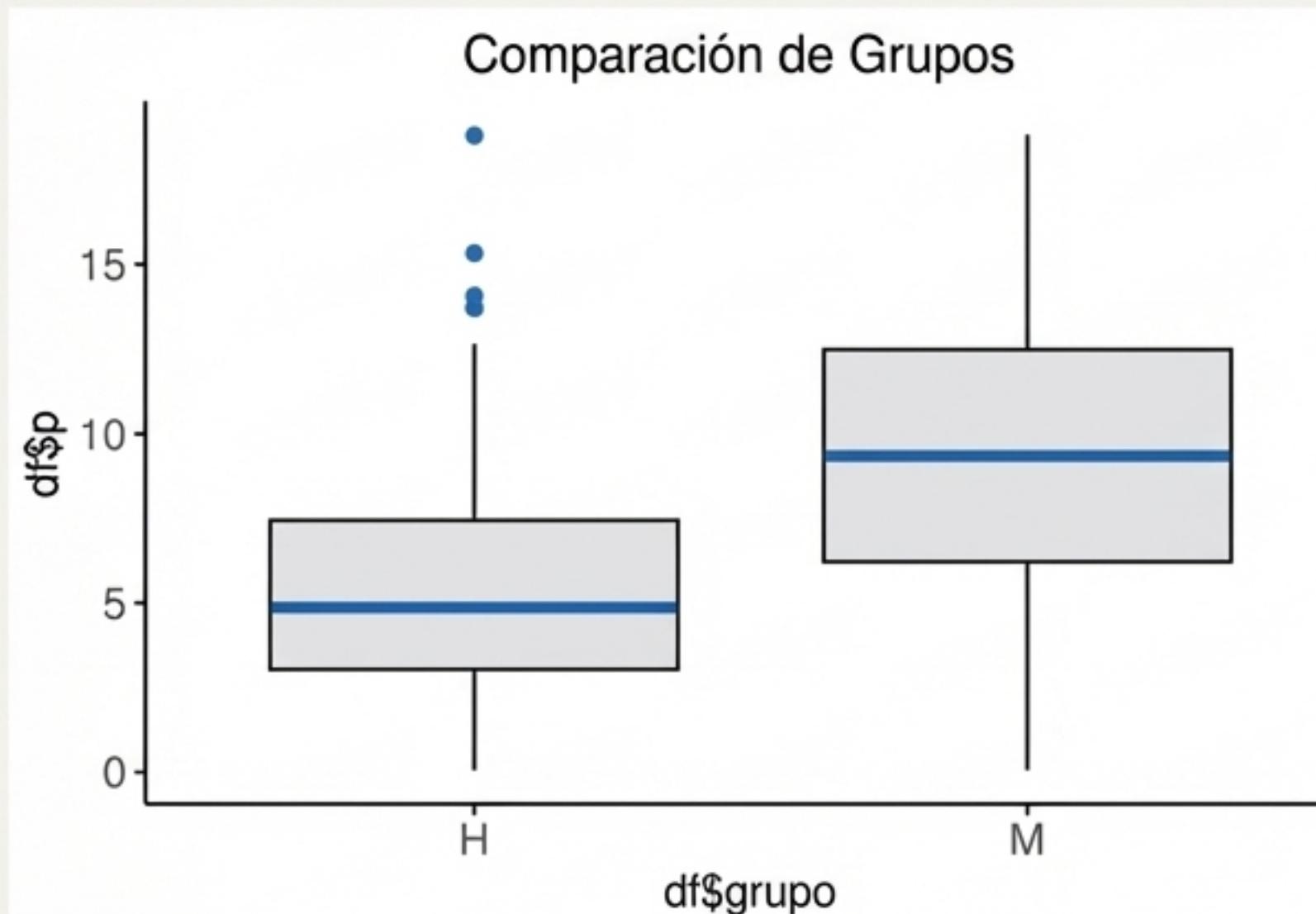
```
# Sintaxis con Tidyverse y pipes
mtcars %>%
  filter(cyl > 4) %>%
  group_by(gear) %>%
  summarise(media_mpg = mean(mpg))
```

# La exploración inicial revela la forma y estructura de sus datos.

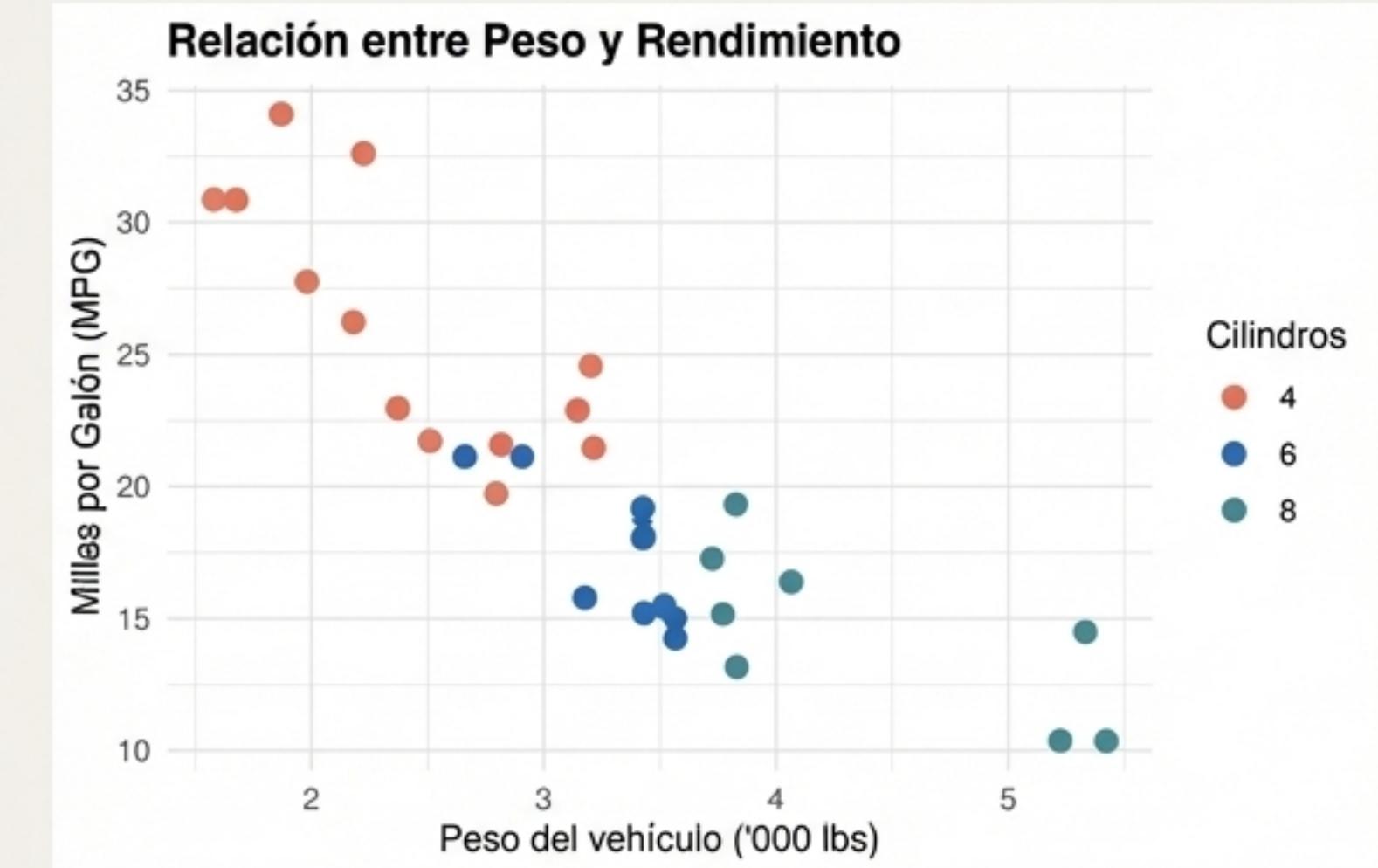
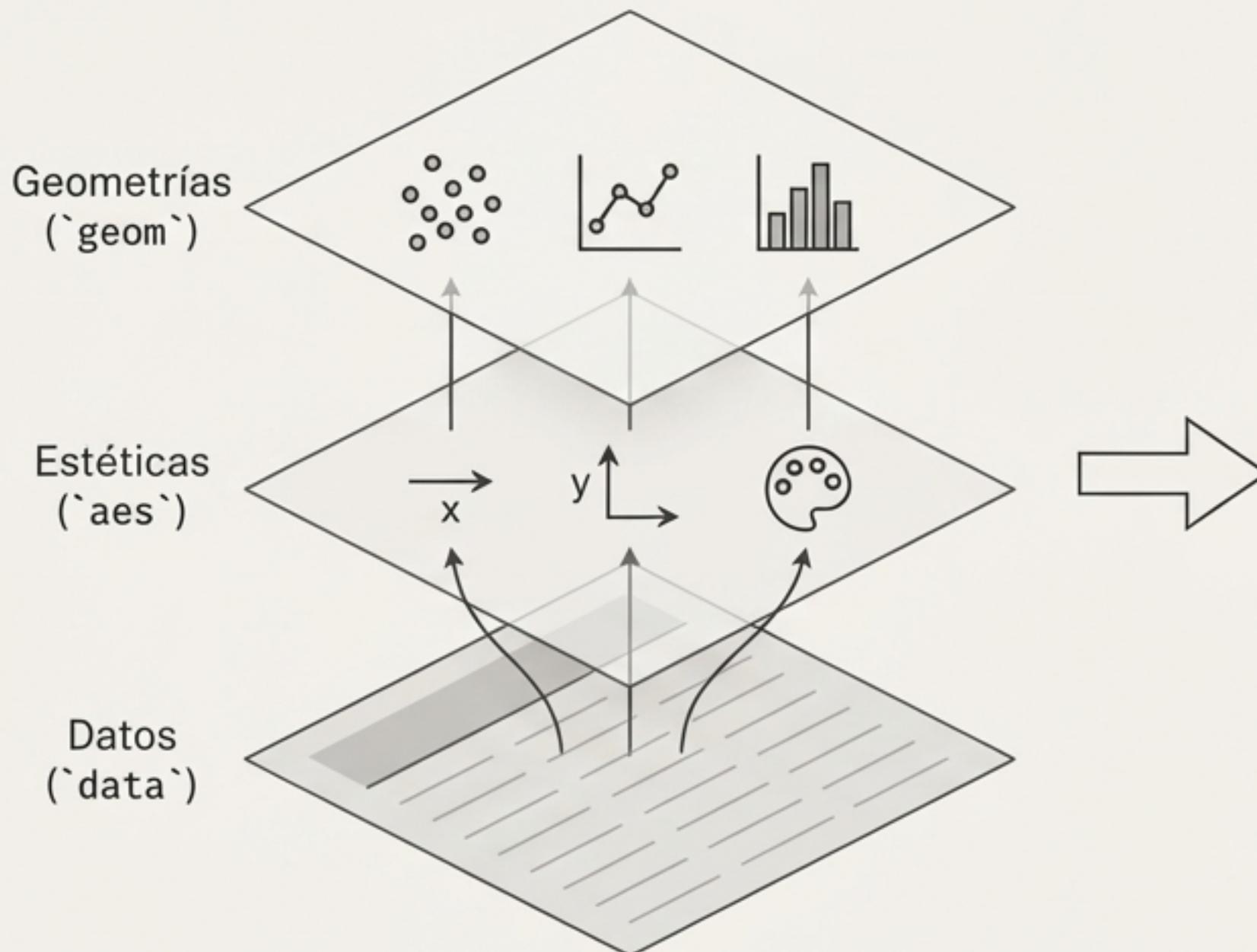
A brief text block introduces the core concepts of Exploratory Data Analysis (EDA).

**Resúmenes numéricos:** La función `summary()` proporciona un resumen estadístico rápido de cada variable.

**Visualización de distribuciones:** Histogramas (`hist()`): Para ver la frecuencia de valores de una variable continua. Diagramas de caja (`boxplot()`): Ideales para comparar distribuciones entre grupos.



# Con `ggplot2`, usted construye gráficos capa por capa, desde los datos hasta la estética.



```
ggplot(data = mtcars, aes(x = wt, y = mpg, color = factor(cyl))) +  
  geom_point(size = 3) +  
  labs(title = "Relación entre Peso y Rendimiento",  
       x = "Peso del vehículo ('000 lbs)",  
       y = "Millas por Galón (MPG)",  
       color = "Cilindros") +  
  theme_minimal()
```

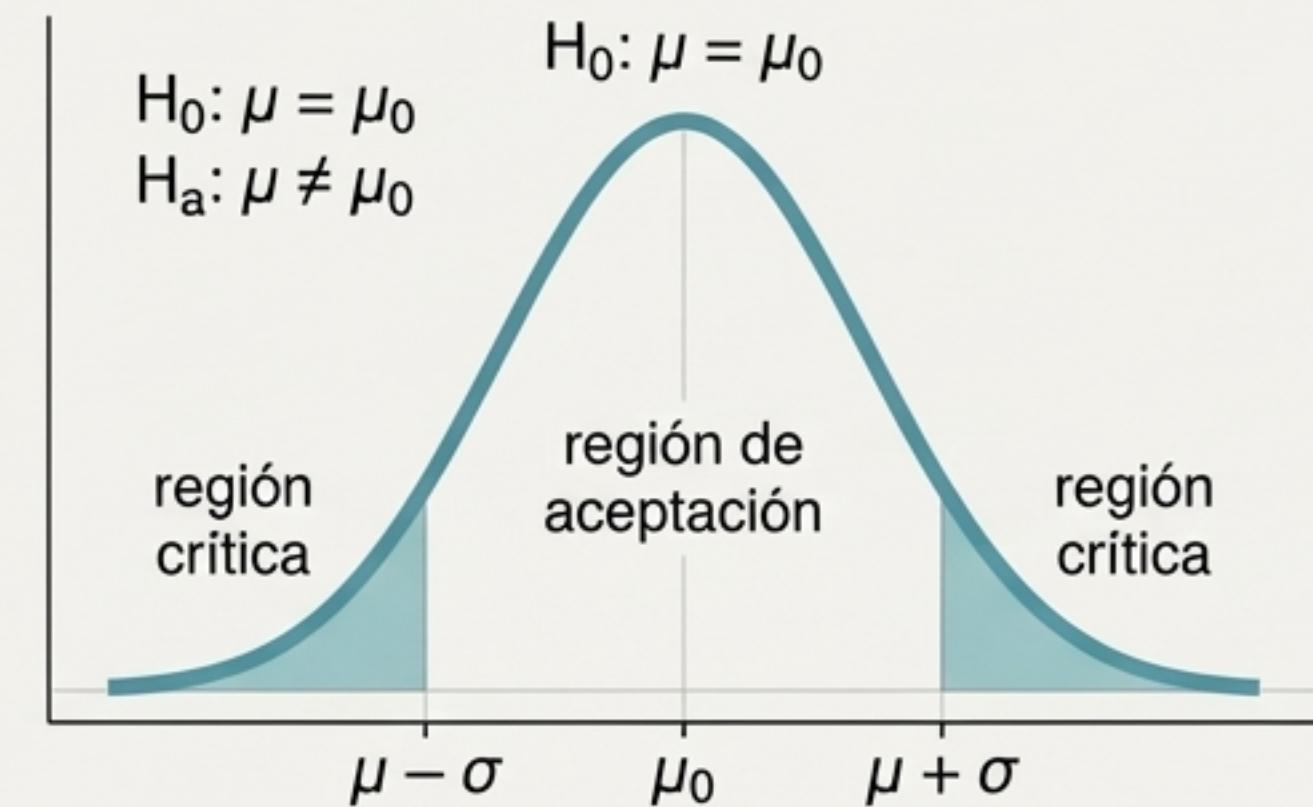
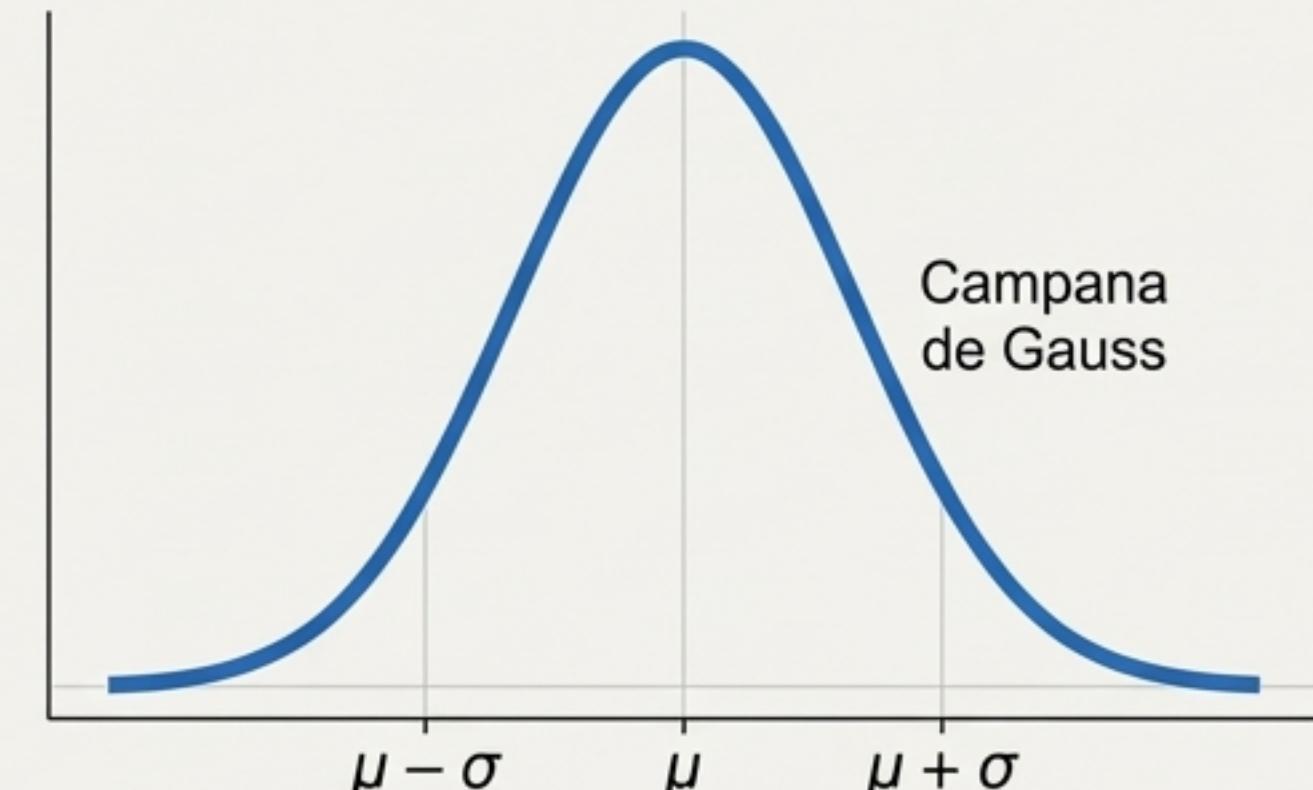
# La inferencia estadística nos permite sacar conclusiones sobre una población a partir de una muestra.

**Población vs. Muestra:** Extrapolar desde los datos recogidos para llegar a conclusiones generales.

**Hipótesis Nula ( $H_0$ ):** La hipótesis de ‘no hay efecto’ o ‘no hay diferencia’.

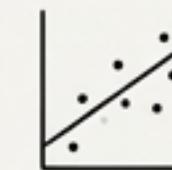
**El p-valor:** La probabilidad de observar nuestros datos (o más extremos) si la  $H_0$  fuera cierta. Un p-valor bajo ( $< 0.05$ ) es evidencia para rechazar la  $H_0$ .

**Intervalos de Confianza:** Un rango de valores plausible para el parámetro de la población.



# La regresión lineal permite cuantificar y predecir la relación entre variables continuas.

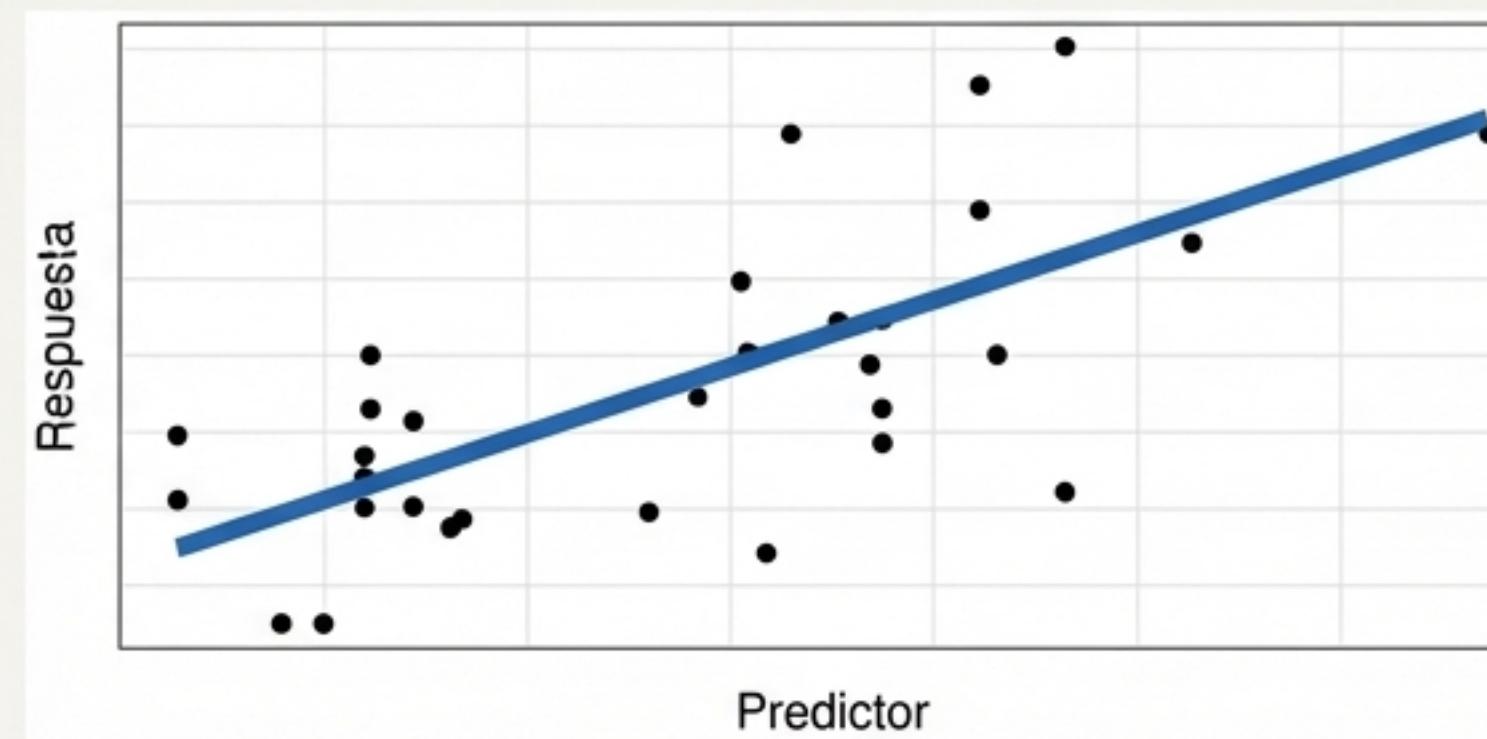
**El Modelo:**  $Y = \beta_0 + \beta_1 X + \varepsilon$



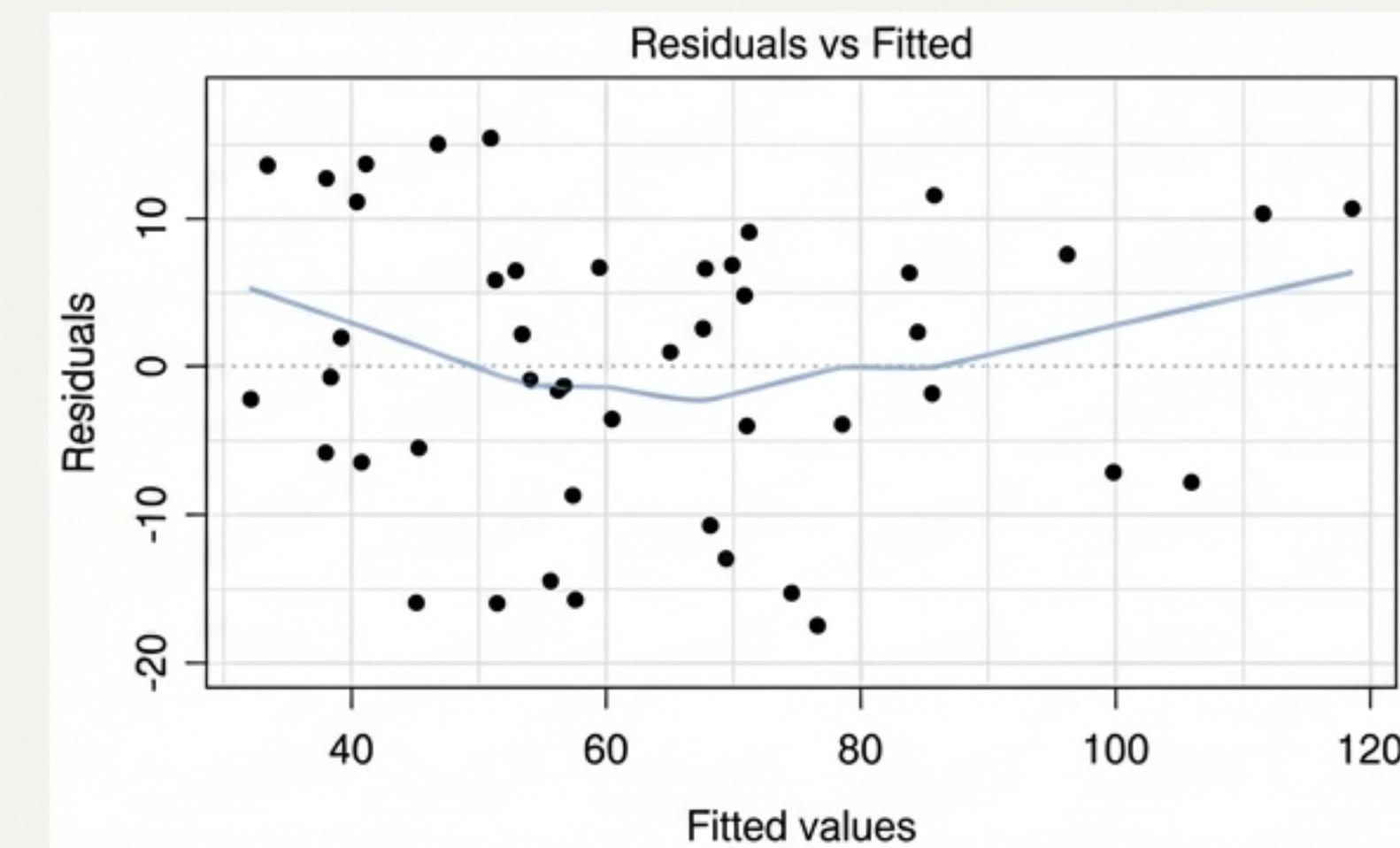
**Implementación en R:**

```
modelo <- lm(respuesta ~ predictor, data = datos)
```

**Interpretación de summary(modelo):** Se explican los Coefficients y el R-squared.



**Diagnóstico del Modelo:** Se enfatiza la importancia de verificar supuestos analizando los residuos. 



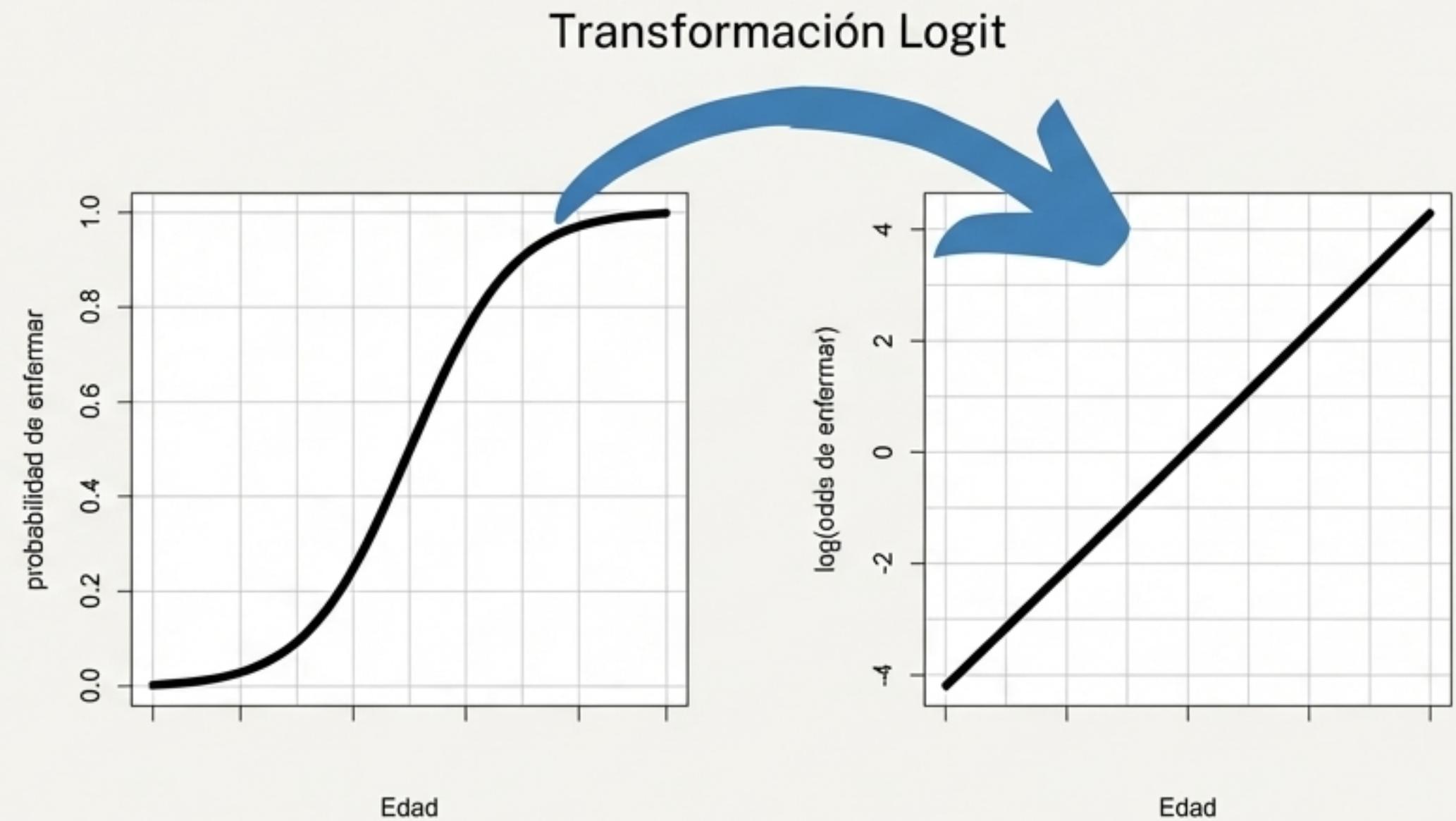
# La regresión logística modela la probabilidad de un resultado categórico.

**El Objetivo:** Predecir la probabilidad de que ocurra un evento (Éxito/Fracaso).

**La Transformación Logit:**  
Explica cómo se convierte la probabilidad (0-1) en una escala lineal.

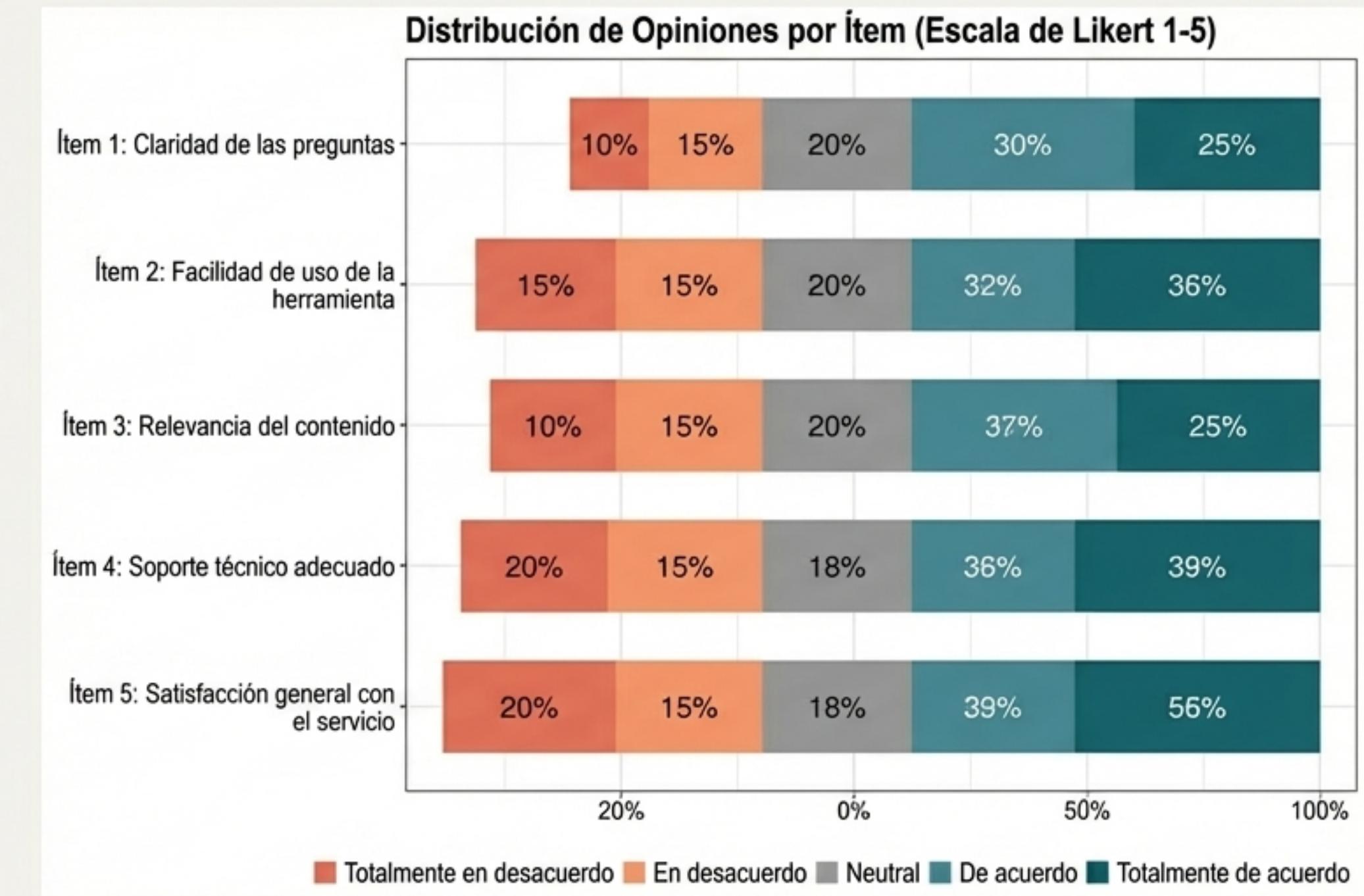
**Implementación en R:**  
`glm()` con `family = binomial()`

**Interpretación:** Se introduce el concepto de Odds Ratios (OR).



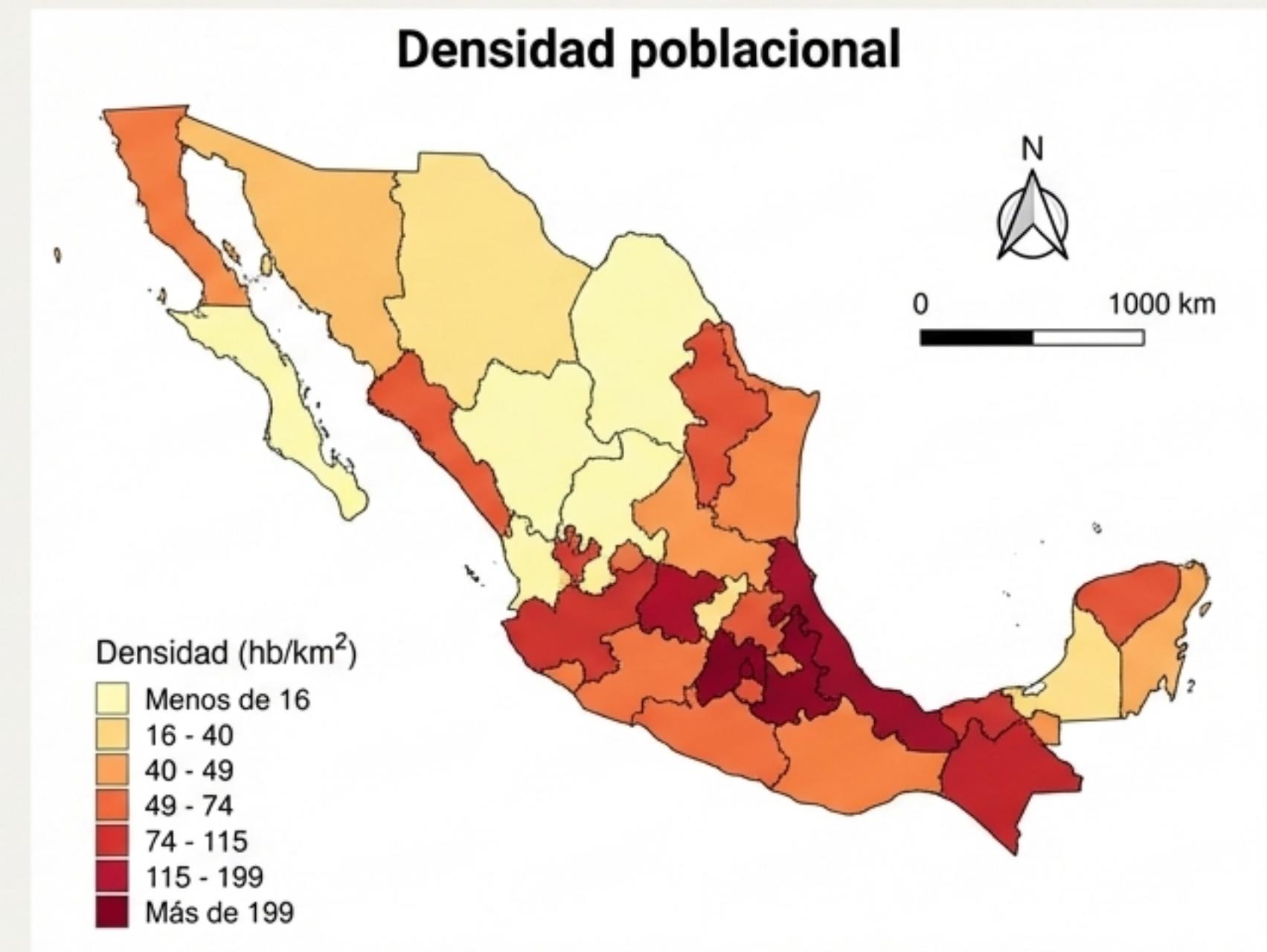
# Expedición Especializada: Análisis de cuestionarios con escalas tipo Likert.

- **Validación del Instrumento:** Se introduce el **Alfa de Cronbach** para medir la consistencia interna de la escala.
- **Visualización:** Se destaca el uso de paquetes como **likert** para crear gráficos de barras divergentes, permitiendo una representación clara de las respuestas.
- **Análisis Inferencial:** Se mencionan pruebas no paramétricas como la **U de Mann-Whitney** y **Kruskal-Wallis** para comparar grupos.

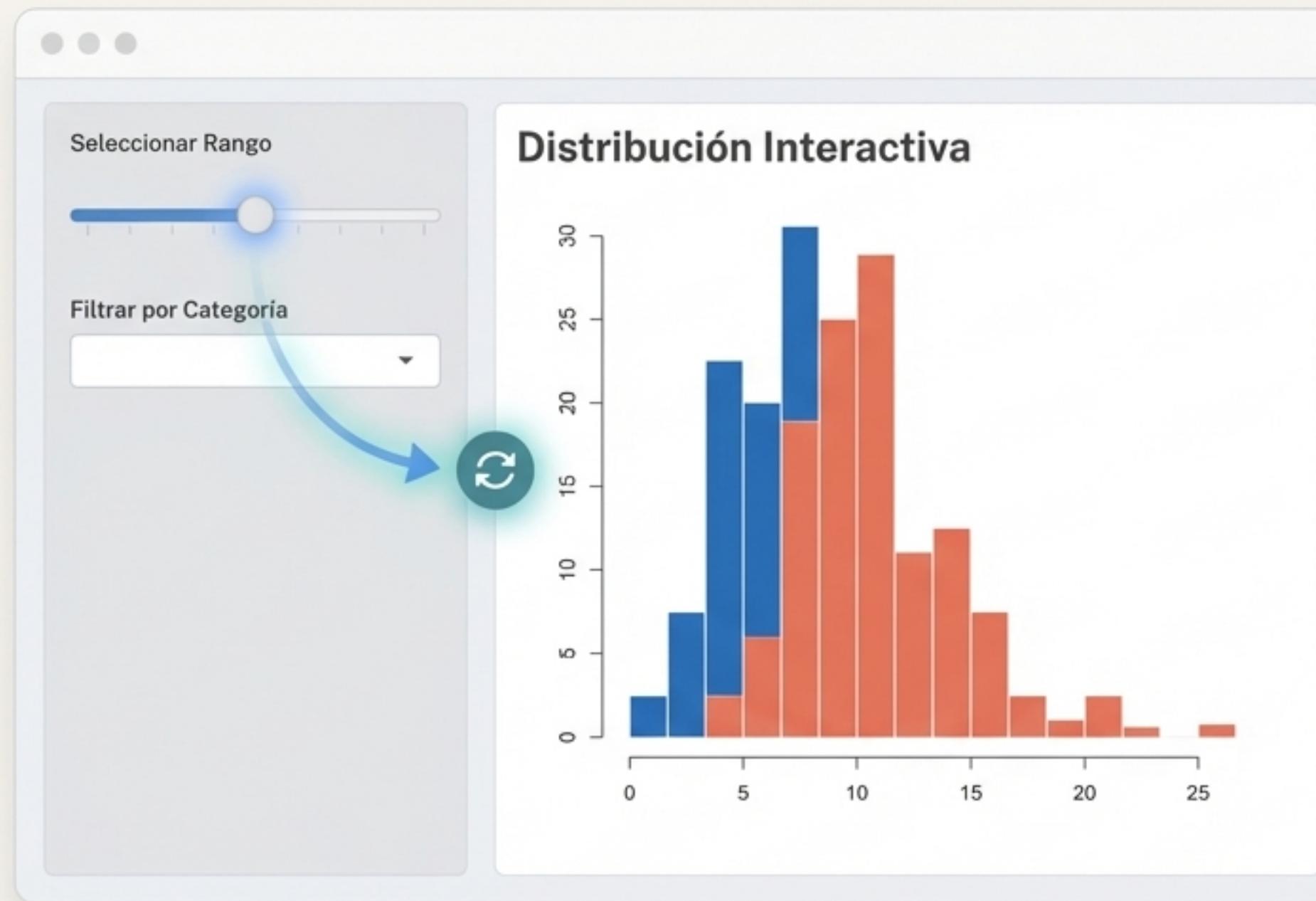


# Expedición Especializada: R como un potente Sistema de Información Geográfica (SIG).

- **El Ecosistema Geoespacial de R:** Se presentan los paquetes fundamentales:
  - `sf` (simple features): El estándar moderno para datos vectoriales.
  - `raster`: Para el manejo de datos en formato de malla.
- **Flujo de Trabajo:**
  1. **Importación:** `st_read()`
  2. **Manipulación:** `st_intersects`, `st_join`
  3. **Visualización:** Creación de mapas temáticos.



# Con Shiny, sus análisis cobran vida como aplicaciones web interactivas.



## ¿Qué es Shiny?

Un paquete para construir aplicaciones web interactivas directamente desde R.

## Estructura

- **ui (Interfaz de Usuario)**  
Define la apariencia y los controles.
- **server (Servidor)**  
Contiene la lógica reactiva que genera las salidas.

## El Poder de la Reactividad

El servidor “escucha” los cambios y actualiza las salidas automáticamente.

# R es un ecosistema integrado para todo el ciclo del análisis de datos.

R offers a unique, potent, and reproducible environment for the entire data science workflow, eliminating the need to switch between different programs.

