

Graficación Bivariada y Multivariada

Juan C. Correa

Material de uso exclusivo para
INGENIO PANTALEON, S.A.
Diagonal 6, 10-31, Zona 10

Ciudad de Guatemala



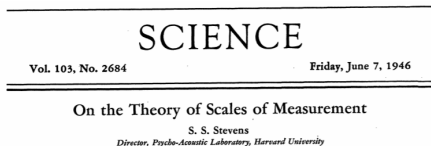
- 1 Preliminares de Graficación Bivariada
- 2 Graficación Bivariada: Ejemplos Particulares
- 3 Graficación Multivariada: Ejemplos Particulares
- 4 Asignación



Preliminares de Graficación Bivariada

La **graficación bivariada** es nuestra primera técnica para realizar informes o reportes ejecutivos de alta calidad. Por bivariado se entiende que lo que se grafica es una relación entre dos variables.

Pero como hay varias clases de variables, no siempre resulta evidente cuál es el gráfico que se debe usar. Por eso, es fundamental recordar los tipos o clases de variables.



Preliminares de Graficación Bivariada

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
NOMINAL	Determination of equality	<i>Permutation group</i> $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	<i>Isotonic group</i> $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	<i>General linear group</i> $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	<i>Similarity group</i> $x' = ax$	Coefficient of variation



Preliminares de Graficación Bivariada

Variables Nominales (Categorías) Son aquellas cuyos casos solo pueden analizarse en términos de identidad. Ejemplo, los números en la camiseta que identifican a un jugador de futbol.



En estas variables (aunque estén en números) no puede hacerse ningún tipo de análisis más allá de conteos, moda, y correlación por tablas de contingencia.



Preliminares de Graficación Bivariada

Variables Ordinales Son aquellas cuyos casos pueden ordenarse en términos de mayor a menor, o viceversa.



En estas variables (aunque estén en números) no puede hacerse ningún tipo de análisis más allá de su conteo por agrupación. Por ejemplo, análisis de campeones (medallistas de oro) versus medallistas (medallistas de plata y bronce) versus competidores (no medallistas). Percentiles y Mediana, también pueden calcularse.



Preliminares de Graficación Bivariada

Variables Intervalo/Razón son variables propiamente cuantitativas (e.g., distancia, peso, temperatura).



Con este tipo de variables todas los análisis estadísticos son aplicables.



En R, las variables desde un punto de vista computacional suelen diferenciarse en:

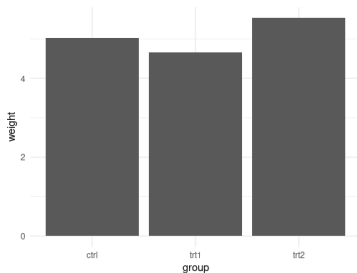
- Characters (Nominales)
- Factor (Nominales)
- Integer (Ordinales)
- Double (Ordinales, intervalo/Razón)
- Logical (Cierto/Falso)
- Date (Fechas)

La graficación bivariada puede aplicarse a cualquier combinación de estas.



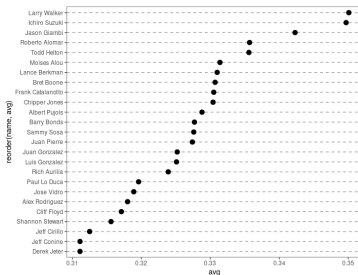
Ejemplos Particulares: Nominal-Intervalo

```
library(ggplot2)
library(gcookbook)
ggplot(pg_mean, aes(x=group, y=weight)) + geom_bar(stat="identity")
```



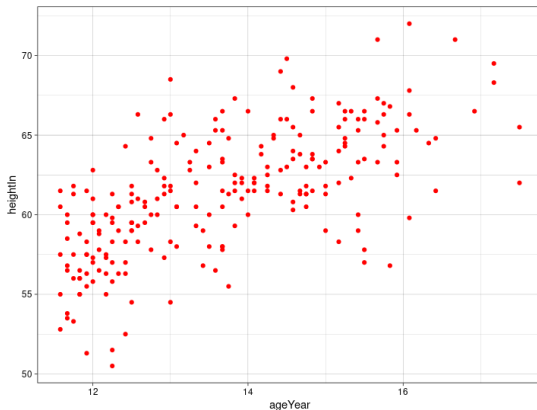
Ejemplos Particulares: Nominal-Intervalo

```
library(gcookbook)
tophit <- tophitters2001[1:25, ]
ggplot(tophit, aes(x=avg, y=reorder(name, avg))) +
  geom_point(size=3) +
  # Use a larger dot
  theme_bw() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(colour="grey60", linetype="dashed"))
```



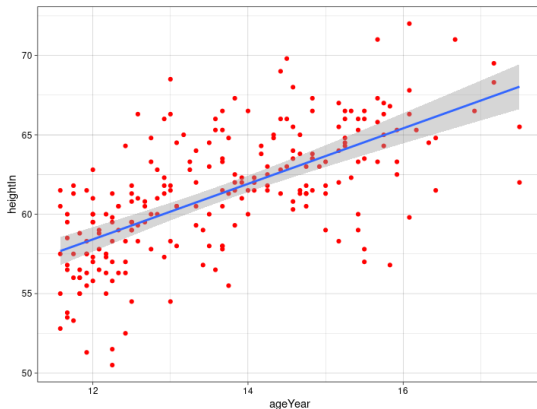
Ejemplos Particulares: Intervalo-Intervalo

```
library(gcookbook)
heightweight[, c("ageYear", "heightIn")]
ggplot(heightweight, aes(x=ageYear, y=heightIn)) + geom_point(colour="red")
+ theme_linedraw()
```



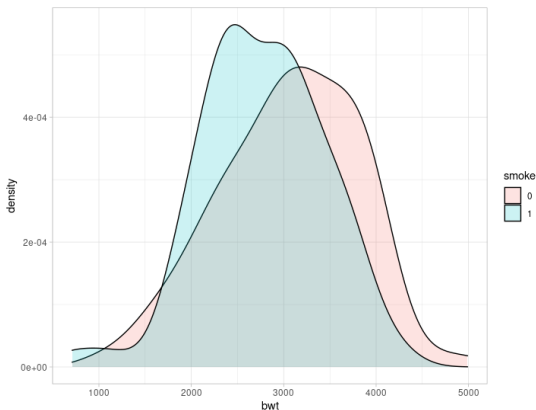
Ejemplos Particulares: Intervalo-Intervalo

```
library(gcookbook)
heightweight[, c("ageYear", "heightIn")]
ggplot(heightweight, aes(x=ageYear, y=heightIn)) + geom_point(colour="red")
+ theme_linedraw() + stat_smooth(method=lm, level=0.99)
```



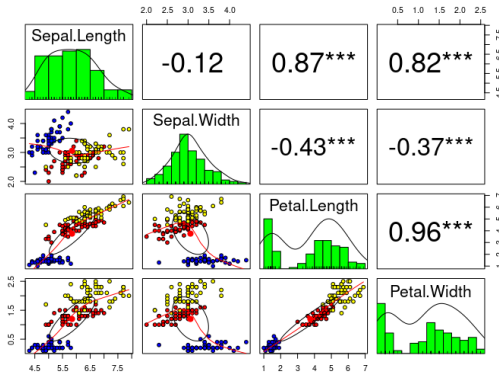
Ejemplos Particulares: Dicotómica-Intervalo

```
library(MASS)
birthwt1 <- birthwt
birthwt1$smoke <- factor(birthwt1$smoke)
ggplot(birthwt1, aes(x=bwt, fill=smoke)) + geom_density(alpha = 0.2) + theme_light()
```



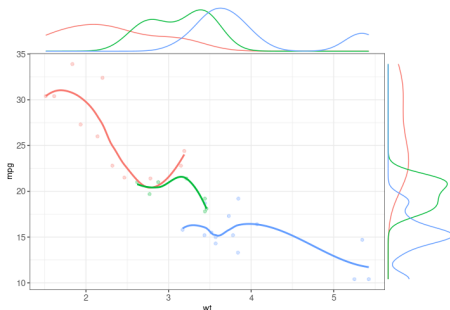
Ejemplos Particulares: Correlation Chart

```
library(psych)
pairs.panels(iris[-5],
             bg=c("blue","red","yellow")[iris$Species],
             hist.col = "green", pch=21, stars = TRUE)
```



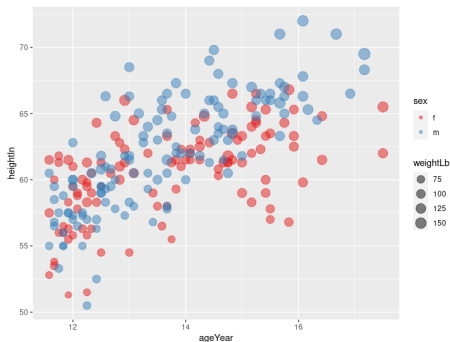
Ejemplos Particulares: Scatterplot trivariado

```
p <- ggplot(mtcars, aes(x=wt, y=mpg, color=as.factor(cyl))) +  
  theme_bw() +  
  geom_point(alpha = 0.3) +  
  geom_smooth(formula = y ~ x, method = loess, se=FALSE) +  
  theme(legend.position = "bottom",  
        legend.title = element_blank(),  
        axis.text.x = element_text(size = 10),  
        axis.text.y = element_text(size = 10),  
        axis.title.x = element_text(size = 10),  
        axis.title.y = element_text(size = 10),  
        legend.text = element_text(size = 10)) +  
  theme(legend.background=element_blank()) +  
  theme(legend.key=element_blank())  
ggMarginal(p, type="density", groupColour = TRUE)
```



Ejemplos Particulares: Scatterplot trivariado

```
ggplot(heightweight, aes(x=ageYear, y=heightIn, size=weightLb, colour=sex)) +  
  geom_point(alpha=.5) +  
  scale_size_area() +  
  scale_colour_brewer(palette="Set1")
```



Usando la base de datos PantaleonData que se encuentra en el siguiente repositorio

<https://github.com/jcorrean/Pantaleon/blob/main/PantaleonData.xlsx>

Desarrolle un par de gráficos bivariados y un gráfico multivariado dentro de un documento RMarkdown.

