

# Gráficos en R

Eduardo Fernández

Extraído de <https://www.r-graph-gallery.com/> y de <https://www.data-to-viz.com/>

# Importancia de los Gráficos

1. **Resumen Visual:** Proporcionan una representación fácil y rápida de patrones, tendencias y distribuciones.
2. **Descubrimiento de Patrones:** Ayudan a identificar patrones y tendencias que no son evidentes en tablas numéricas.
3. **Identificación de Valores Atípicos:** Permiten detectar valores atípicos que pueden indicar errores o áreas de investigación.
4. **Comparación de Datos:** Facilitan la comparación de diferentes conjuntos de datos y su interrelación.
5. **Comunicación Efectiva:** Son herramientas efectivas para comunicar hallazgos, especialmente a audiencias no técnicas.

# Importancia de los Gráficos

- 6. Análisis Exploratorio:** Son fundamentales para entender las características de los datos antes de análisis más detallados.
- 7. Verificación de Modelos:** Útiles para comprobar la adecuación de los modelos estadísticos a los datos.
- 8. Dirección de Análisis Futuros:** Indican áreas de interés o preocupación que podrían requerir análisis o recolección de datos adicionales.
- 9. Retención y Compromiso:** Son más atractivos y memorables que las tablas numéricas, mejorando la comprensión y recordación de la información.

**En resumen, los gráficos estadísticos son herramientas esenciales para resumir, explorar, interpretar y comunicar datos.**

# Datos y Gráficos

## **Tipos de Datos**

- 1. Cuantitativos (Intervalo, Ratio)**
- 2. Categóricos (Ordinal, Nominal)**
- 3. Numéricos & Categóricos**
- 4. Mapas**
- 5. Redes**
- 6. Series Temporales**

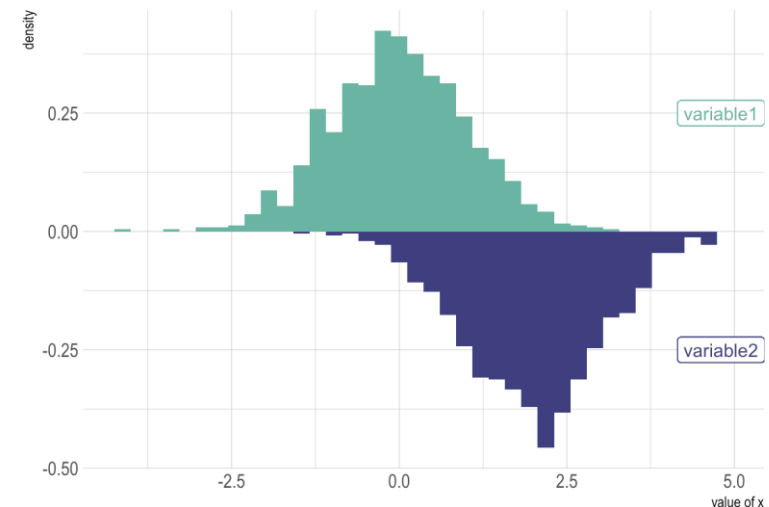
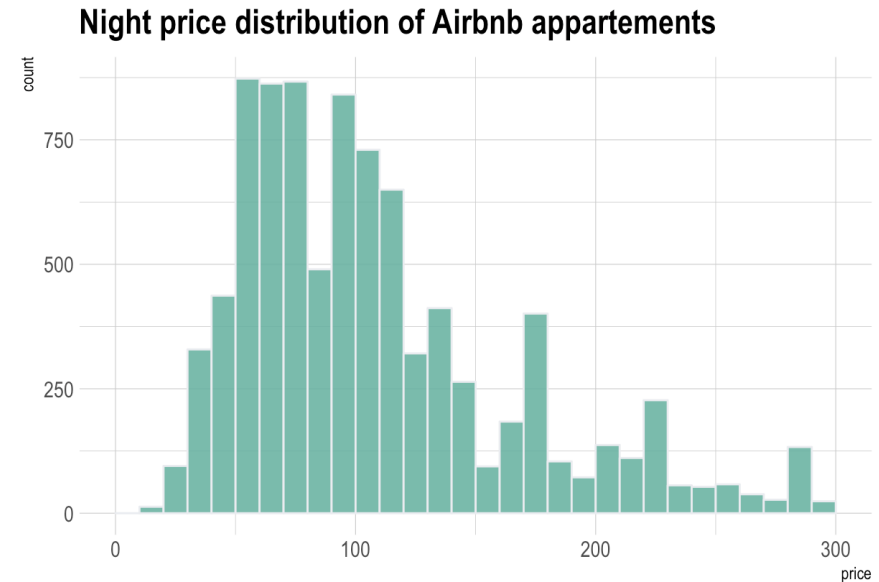
## **Tipos de gráficos**

- 1. Distribución**
- 2. Correlación**
- 3. Ranking**
- 4. La parte en el todo**
- 5. Evolución**
- 6. Mapa**
- 7. Flujo**

# Variable simple cuantitativa: Histograma (Histogram)

- **Distribución de una variable continua, representando la frecuencia de diferentes intervalos de valores.**
- Puede dar una idea visual de la mediana, la moda y la dispersión de los datos.
- La variable es cortada en secciones, y el número de observaciones por sección es representada en la altura de la barra.

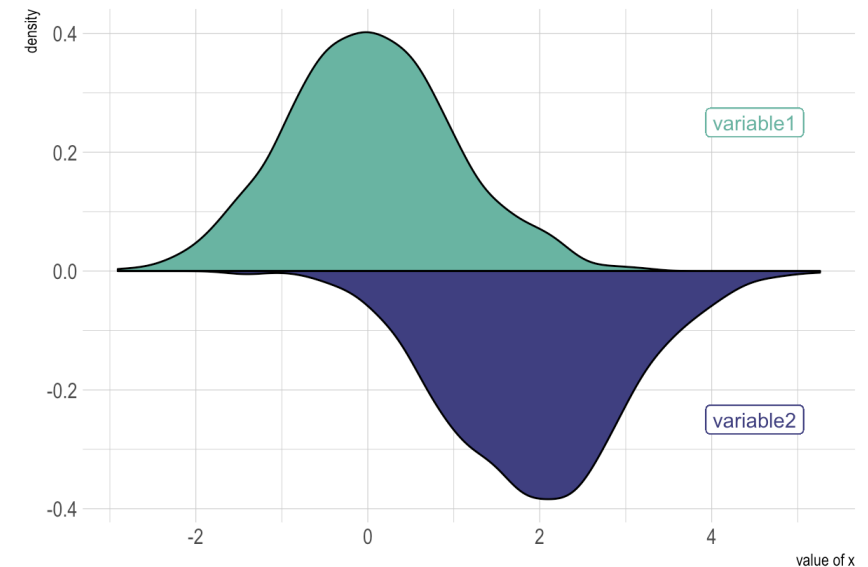
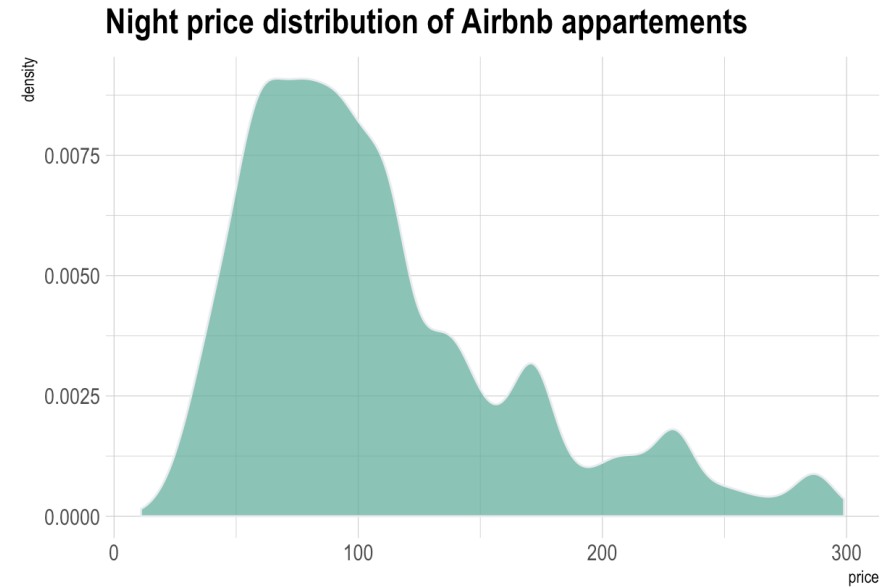
<https://www.data-to-viz.com/graph/histogram.html>



# Variable simple cuantitativa: Gráfico de densidad (Density plot)

- Muestra la función de densidad de probabilidad de una variable continua.
- Similar a un histograma, pero continuo.

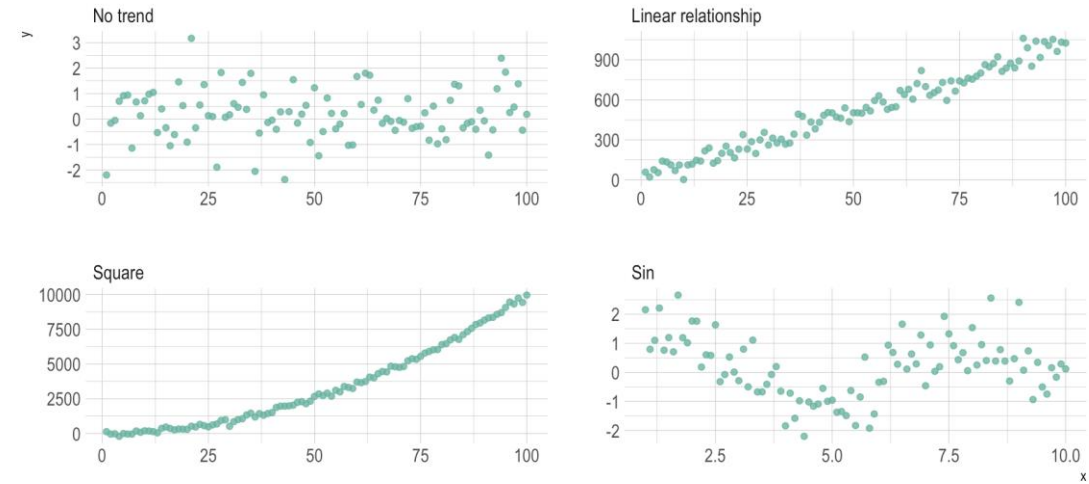
<https://www.data-to-viz.com/graph/density.html>



# Dos variables cuantitativas: Gráfico de dispersión (Scatter plot)

- Muestra la relación entre dos variables cuantitativas.
- Cada eje codifica los valores de una variable cuantitativa diferente y los datos individuales se representan como puntos (o puntos) en el gráfico.
- A menudo se utiliza con una línea de regresión (gráfico de línea) para mostrar la tendencia en los datos.

<https://www.data-to-viz.com/graph/scatter.html>



Diferentes tipos de relaciones entre variables.

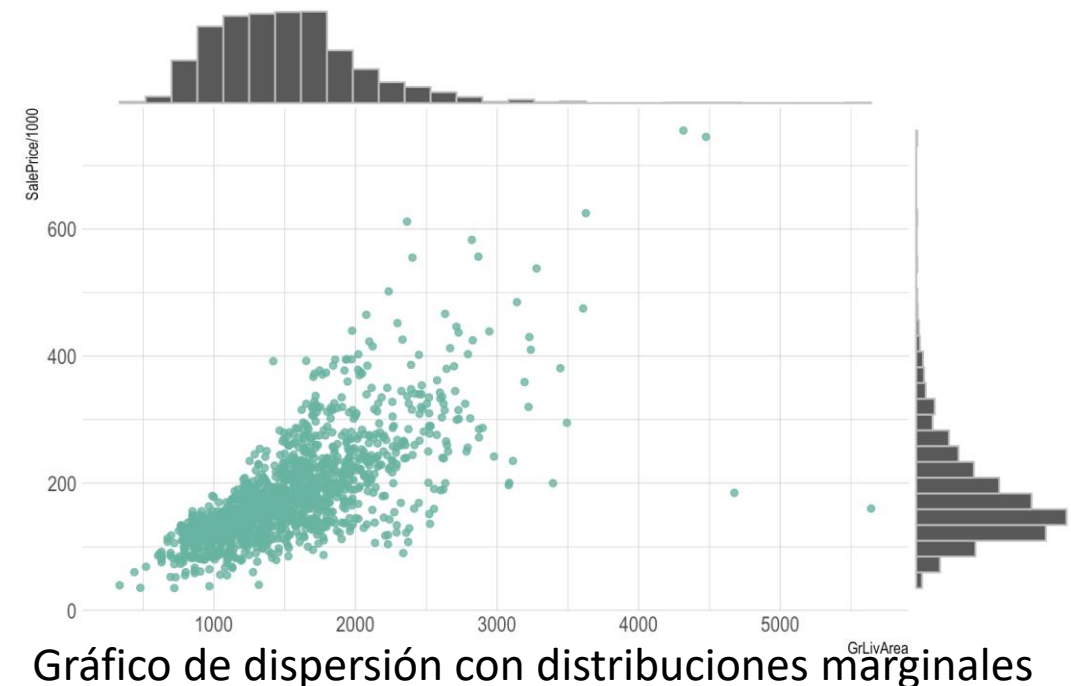
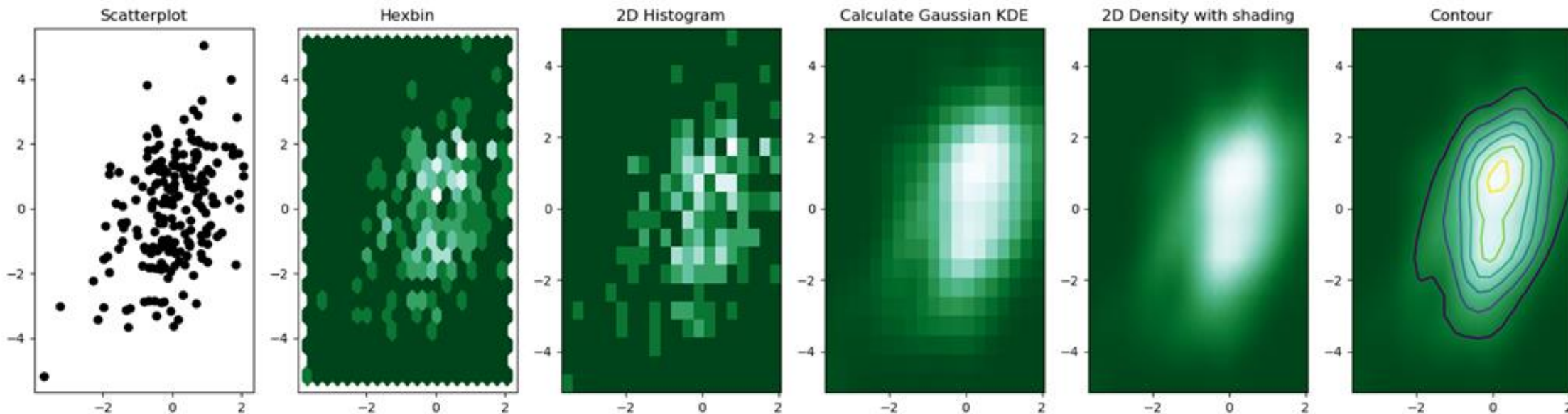
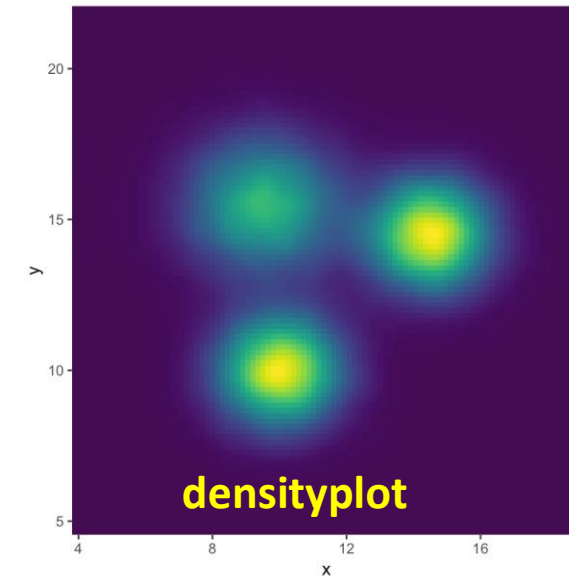
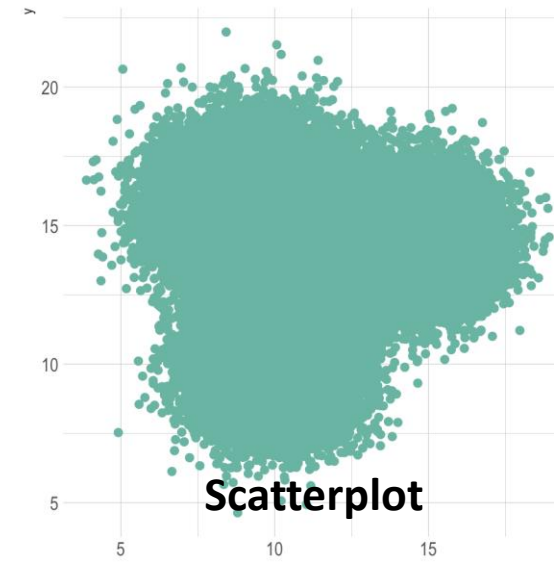


Gráfico de dispersión con distribuciones marginales

# Dos variables cuantitativas: Gráfico de densidad 2D (2D Density plot)

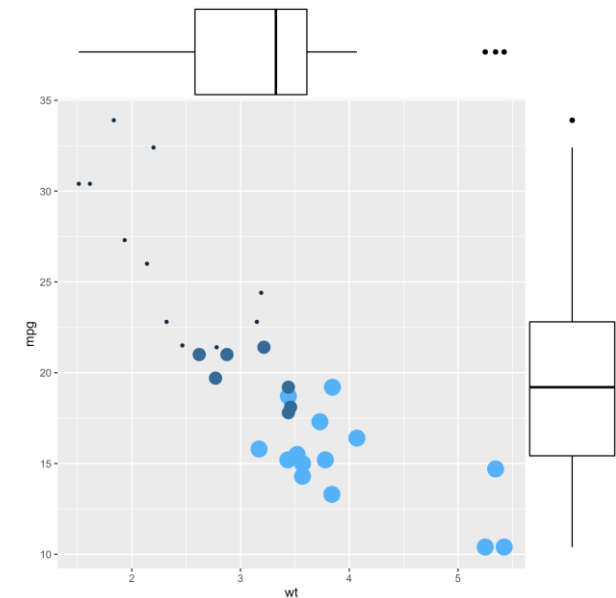
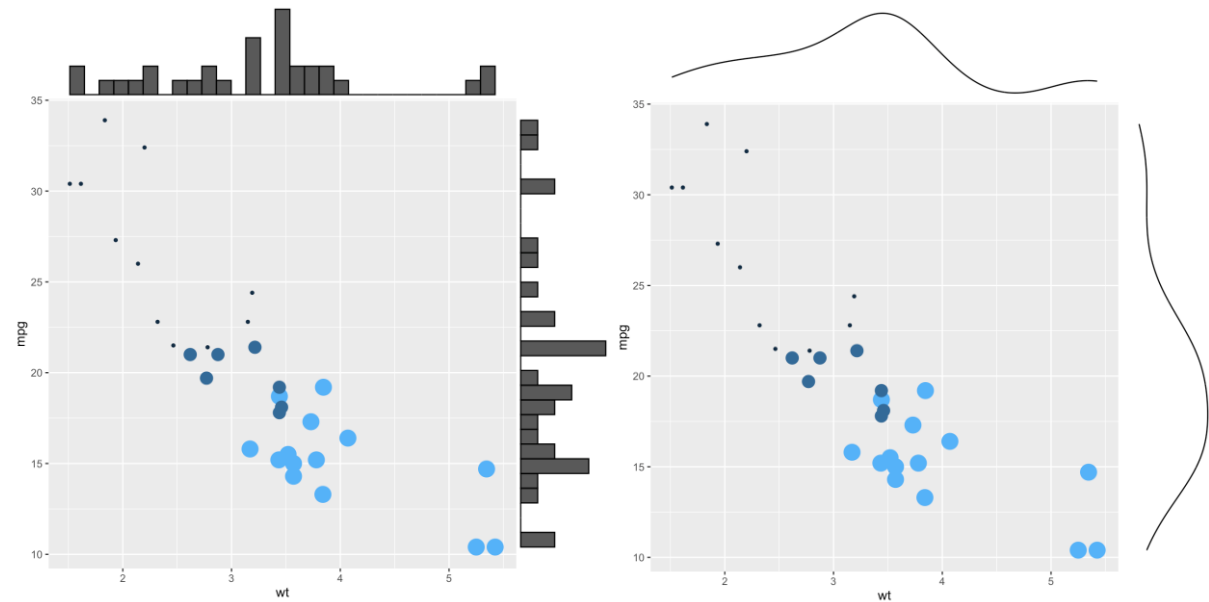
- Muestra la distribución de valores en un conjunto de datos en el rango de dos variables cuantitativas.
- Cuenta el número de observaciones dentro de un área particular del espacio 2D y lo representa utilizando una estimación de densidad cuadrada, hexagonal o kernel.
- <https://www.data-to-viz.com/graph/density2d.html>





## Dos variables cuantitativas: Con distribución marginal

- Muestra la distribución marginal agregada a los ejes x e y de un plot de dos variables continuas.
- Se puede usar histograma, densidad, boxplot



<https://www.r-graph-gallery.com/277-marginal-histogram-for-ggplot2.html>

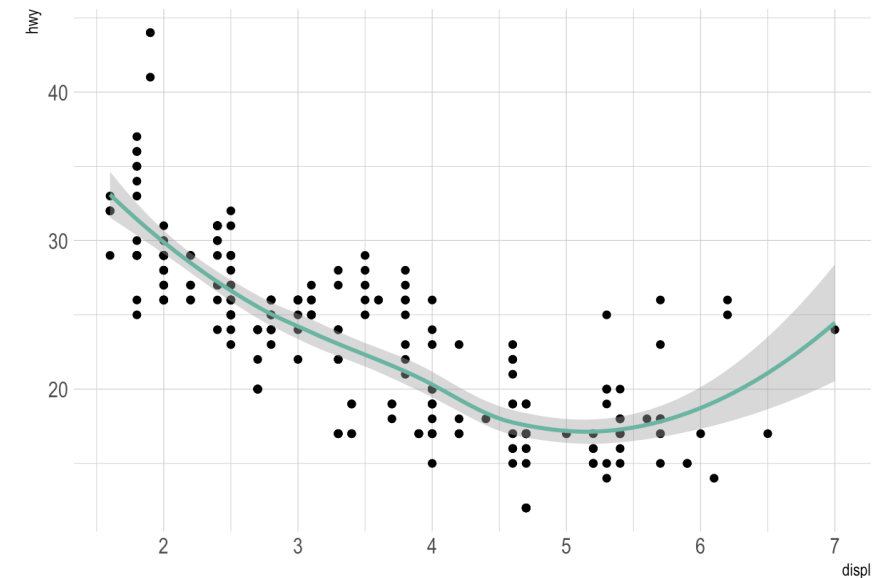
## Dos variables cuantitativas: Gráfico de línea (Line plot)

- Muestra la relación entre dos variables cuantitativas.
- Muestra la tendencia a lo largo del tiempo (una serie temporal)
- Los puntos de datos están conectados por líneas rectas. Son comunes en las series de tiempo.

<https://www.data-to-viz.com/graph/line.html>



Evolución del bitcoin.



Estimación de relación (line plot)  
Junto con los datos en sí (gráfico de dispersión).

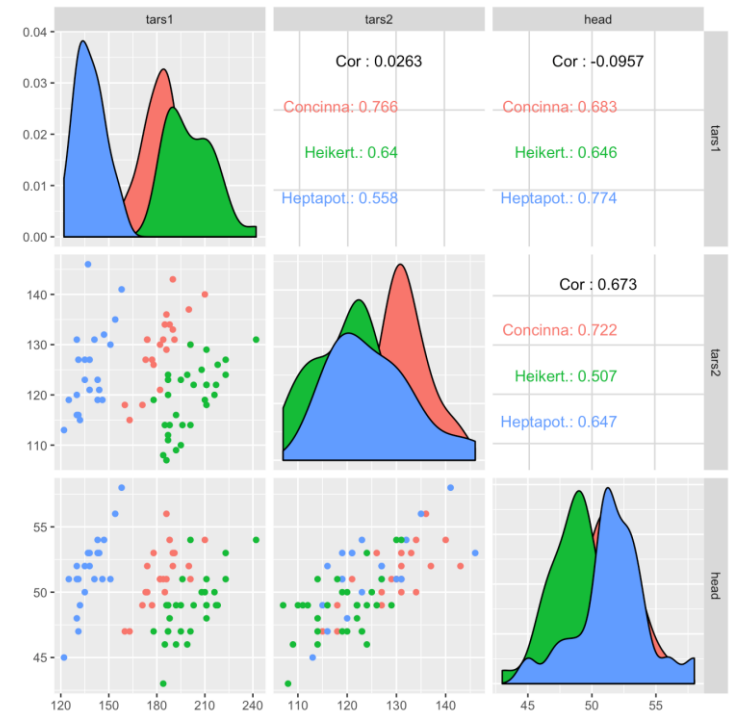
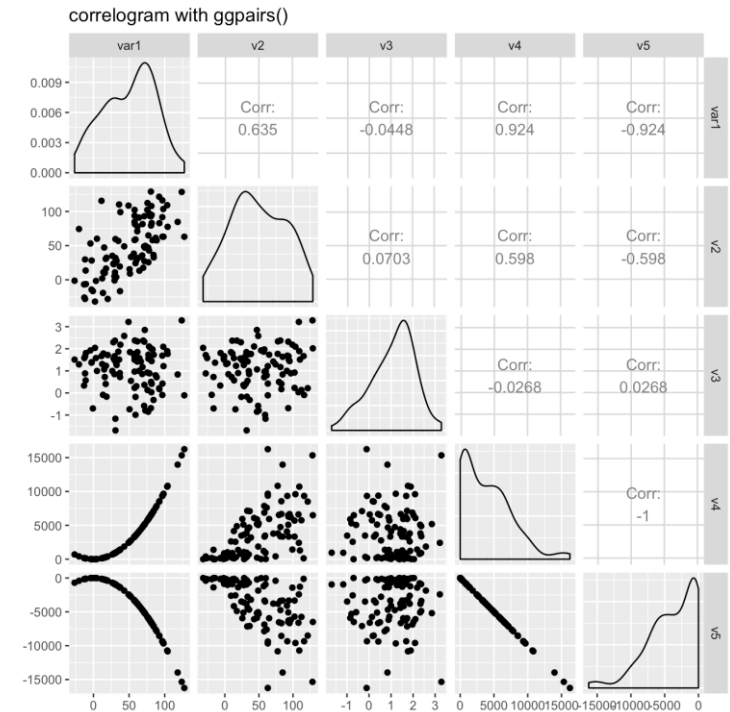
# Muchas variables cuantitativas

## Correlograma (Correlogram)

- Matriz cuadrada de gráficos que ayuda a visualizar la relación entre varios pares de variables numéricas.
- Existen diversas formas de visualizar.
- Matriz simétrica

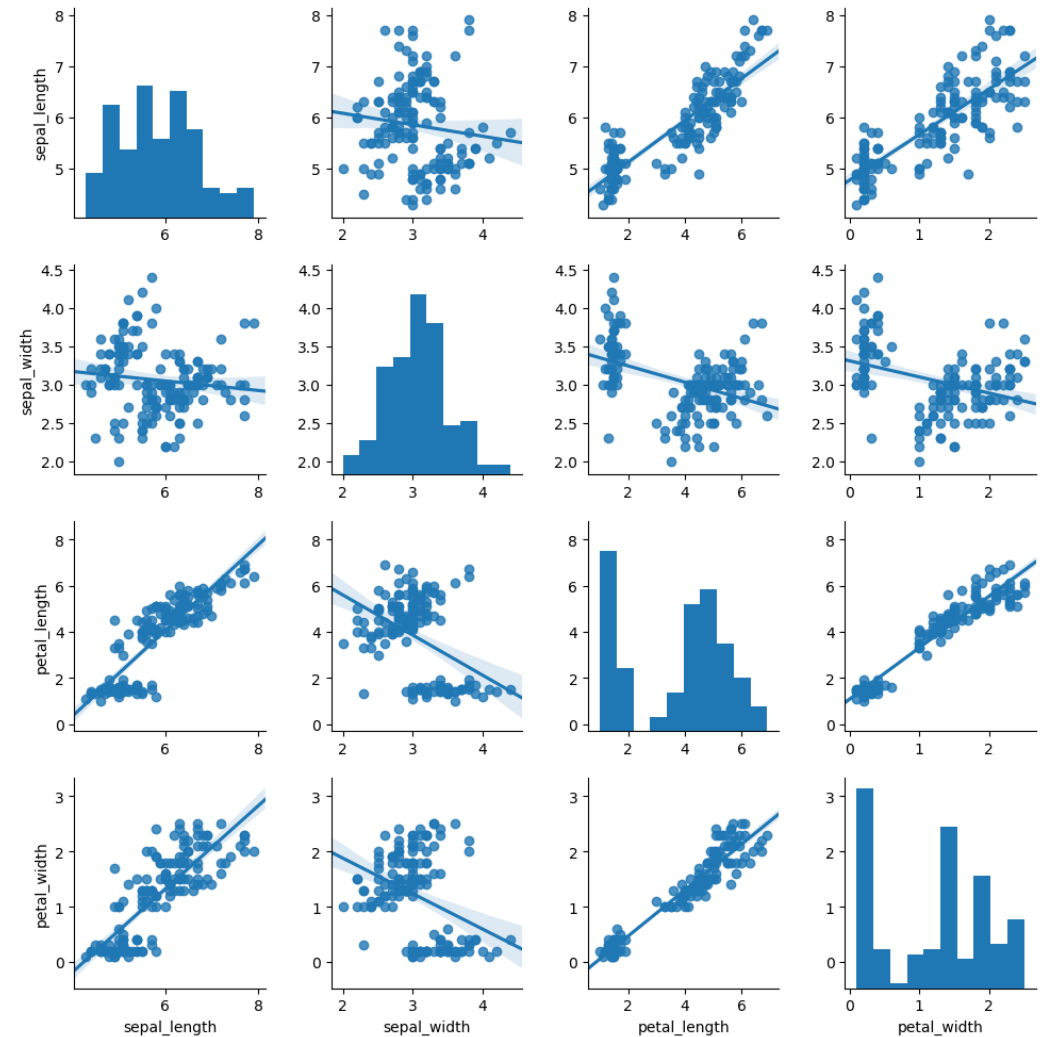
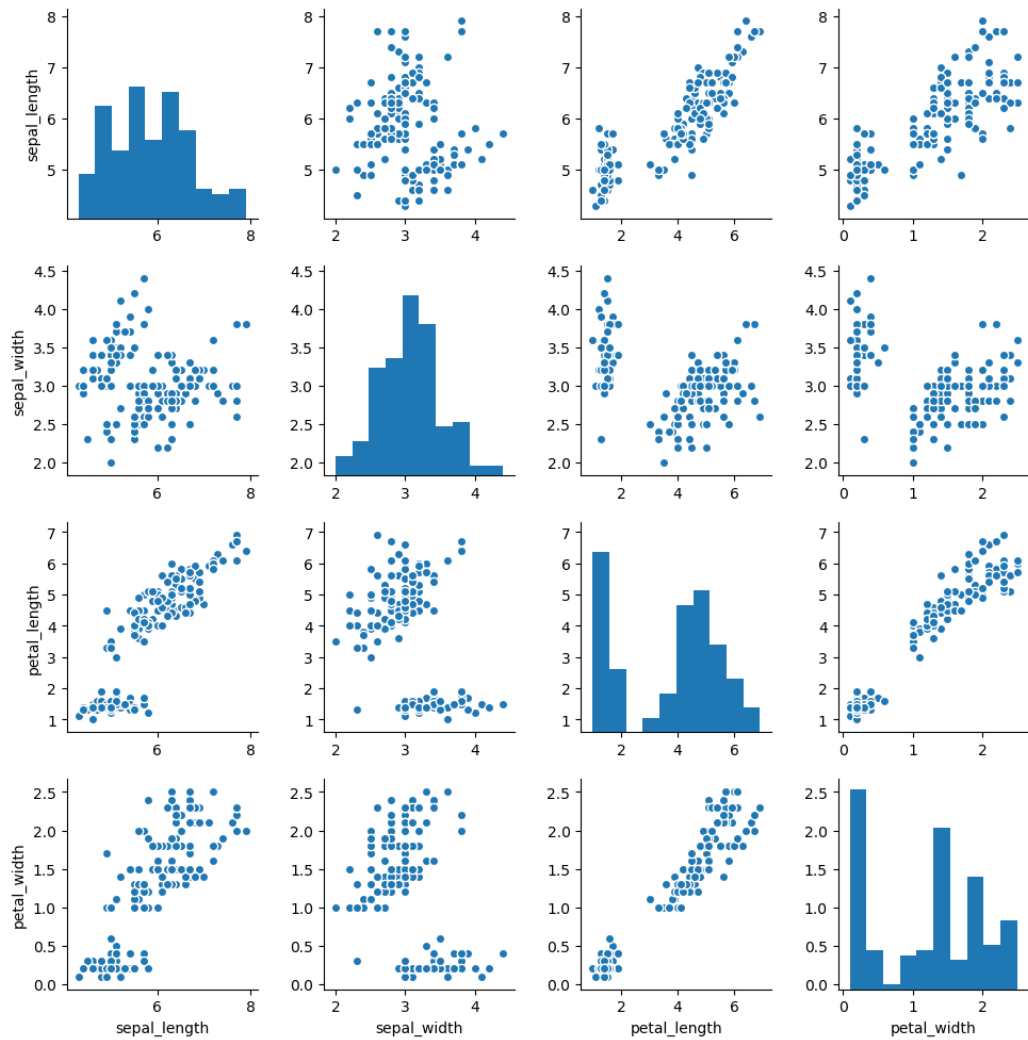
<https://www.r-graph-gallery.com/199-correlation-matrix-with-ggally.html>

<https://www.data-to-viz.com/graph/correlogram.html>



# Muchas variables cuantitativas

## Correlograma (Correlogram)

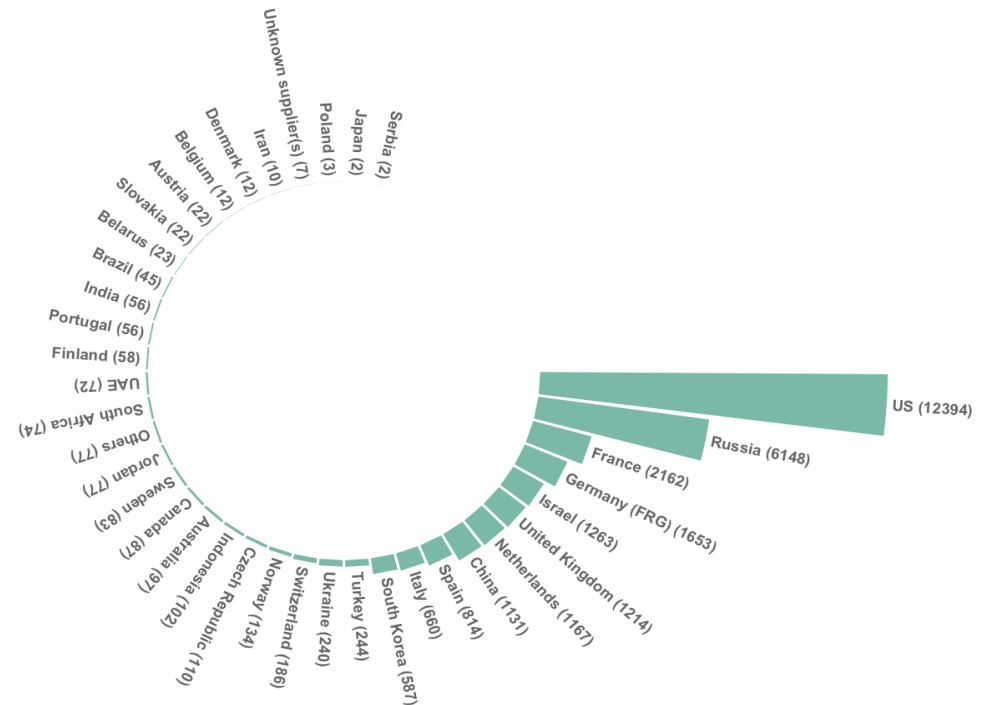
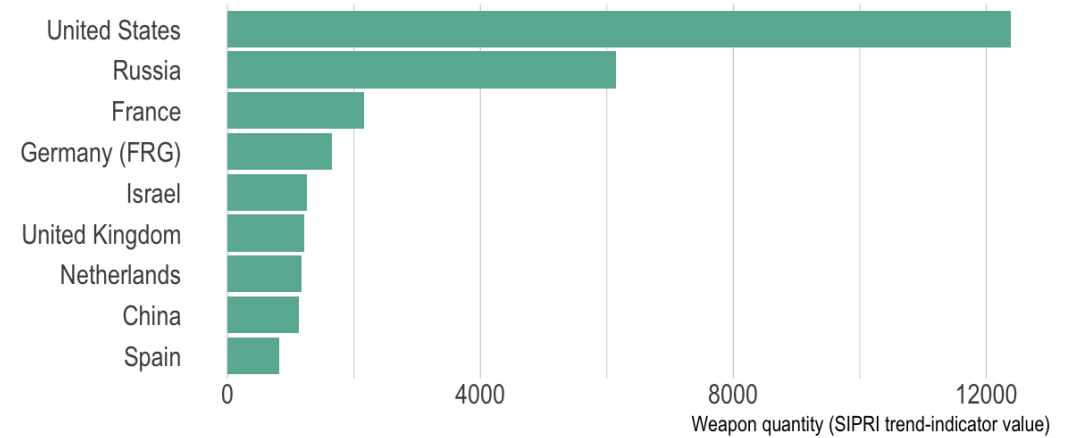


# var cuantitativa + var. categórica

## Gráfico de barras (Bar chart)

- Establece relaciones entre variables numéricas y categóricas.
- Compara varias categorías de datos. Puede ser útil para mostrar datos de frecuencia o proporciones.
- Las barras representan la cantidad de datos en diferentes categorías de una variable. Un eje codifica las frecuencias de los datos cuantitativos y el otro eje las categorías de los datos cualitativos.

<https://www.data-to-viz.com/graph/barplot.html>

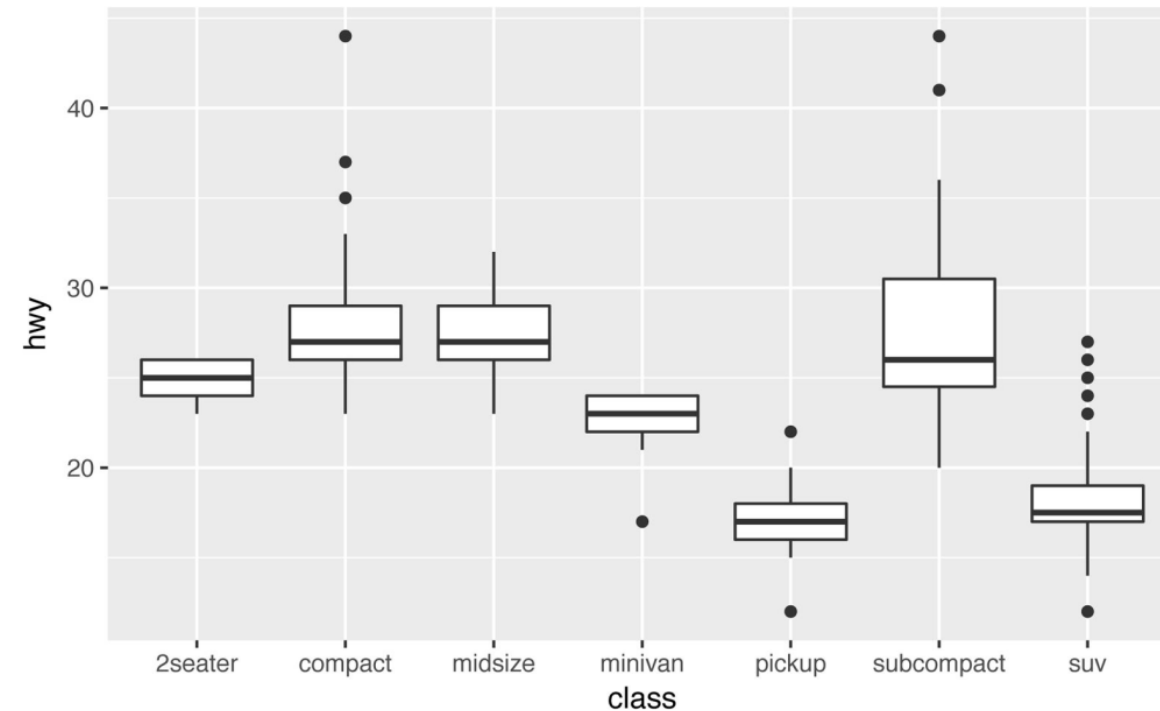
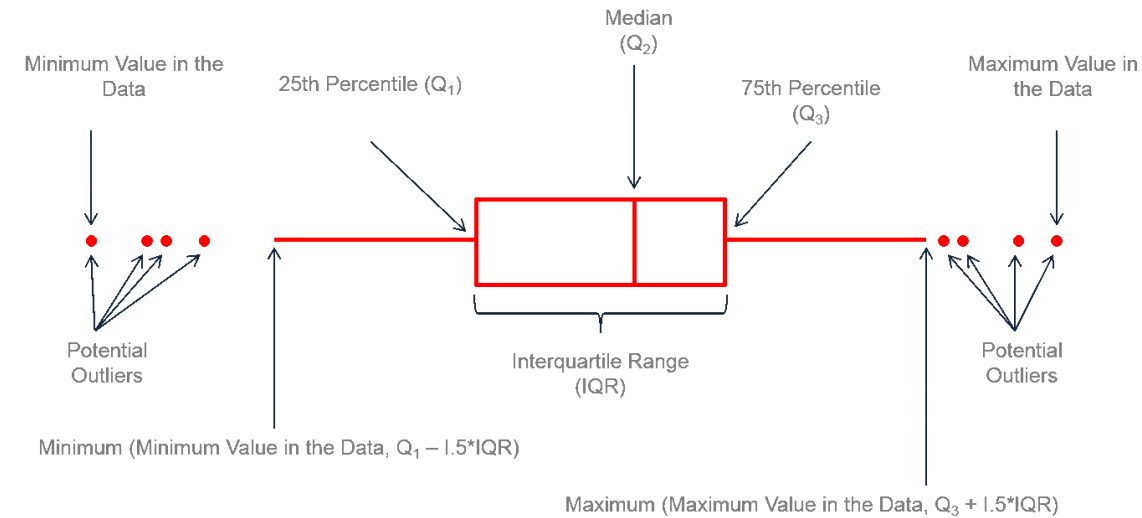


Who sells more weapons?

# var cuantitativa + var. categórica

## Gráfico de caja (Box plot)

- **Mediana, cuartiles, el rango intercuartílico y los posibles valores atípicos en los datos.**
- Se utilizan los cuartiles, desplegando:
  - 1er cuartil ( $Q_1 = 25\%$  de los datos).
  - 2do cuartil ( $Q_2 = 50\%$  de los datos).
  - 3er cuartil ( $Q_3 = 75\%$  de los datos).
  - “Bigotes” que se extienden hasta los valores más alejados mientras no midan más que  $1.5 \cdot (Q_3 - Q_1)$
  - Los valores por fuera (atípicos) se marcan individualmente.
- <https://www.data-to-viz.com/caveat/boxplot.html>

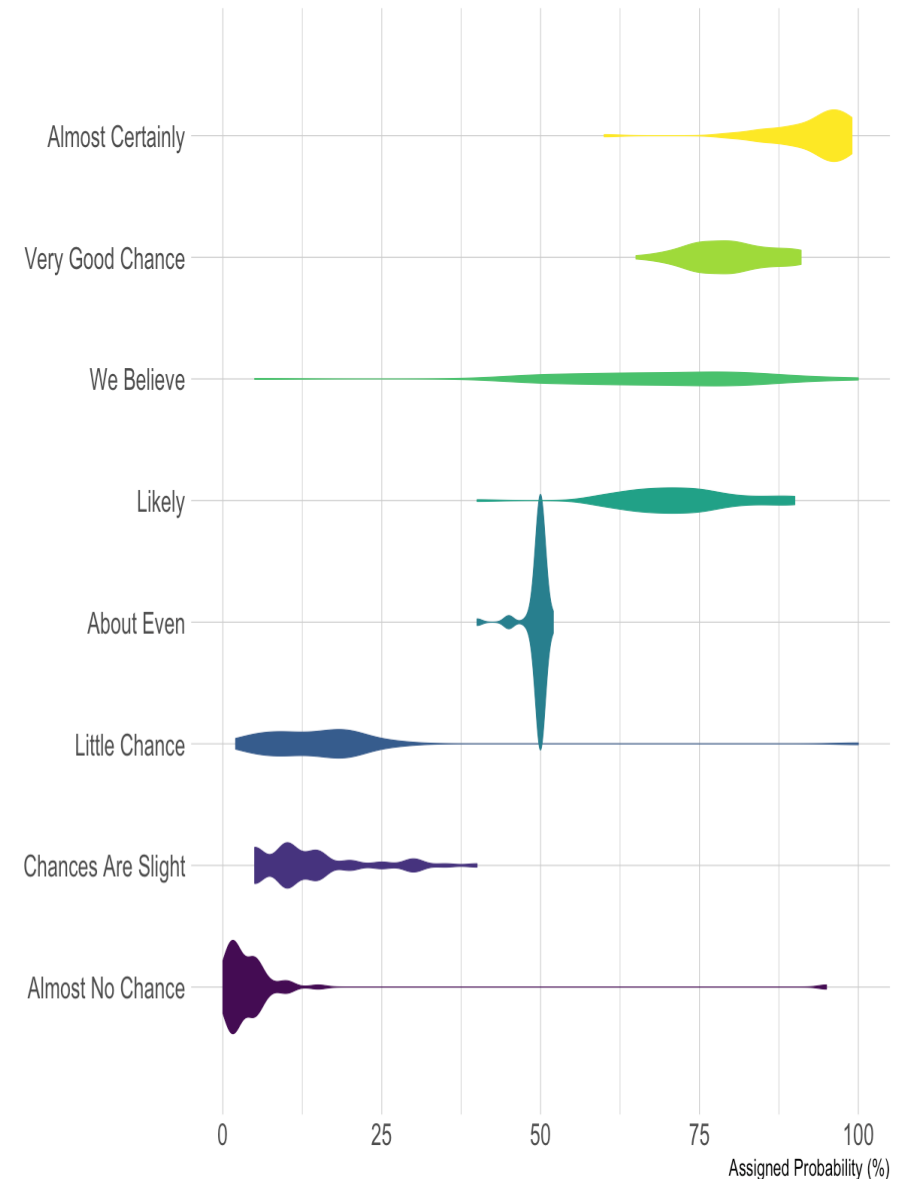


# var cuantitativa + var. categórica

## Gráfico de violín (violin plot)

- **Muestra la distribución de una variable para uno o varios grupos.**
- en esencia, un boxplot que también muestra la distribución/densidad de la variable cuantitativa.

<https://www.data-to-viz.com/graph/violin.html>

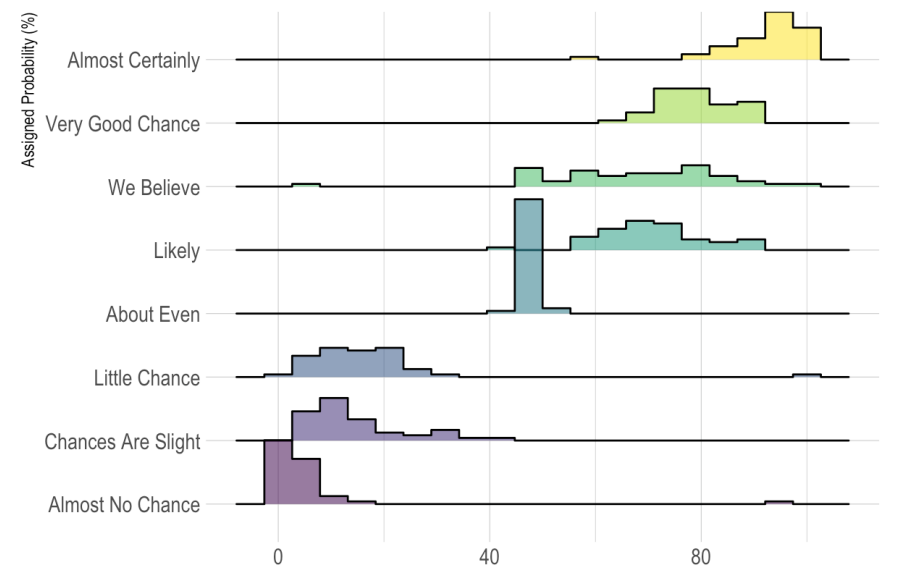
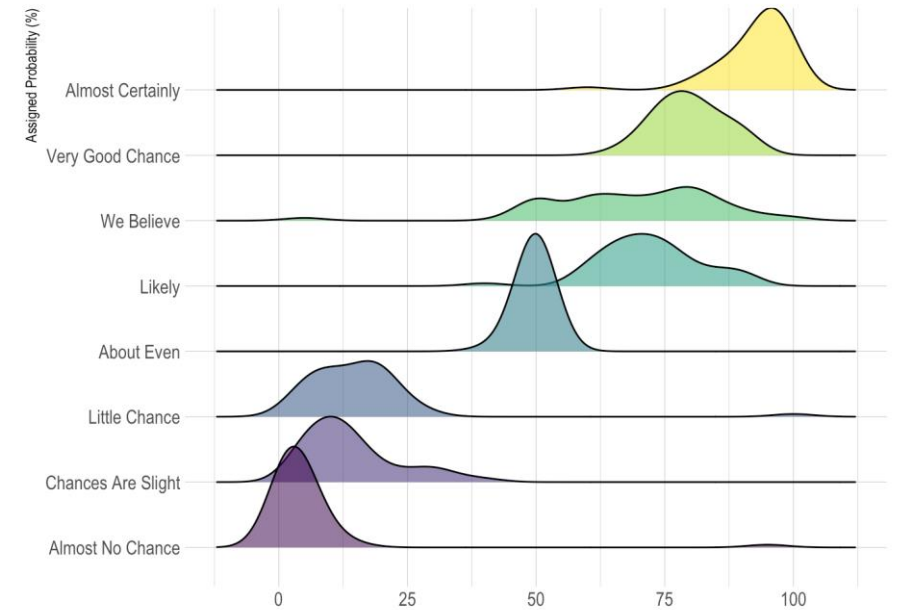


# var cuantitativa + var. categórica

## Gráfico de Cresta (Ridgeline plot)

- **Muestra la distribución de un valor numérico para varios grupos.**
- La distribución se puede representar mediante histogramas o diagramas de densidad, todos alineados en la misma escala horizontal y presentados con una ligera superposición.
- Es especialmente útil cuando el número de grupos a representar es alto.

<https://www.data-to-viz.com/graph/ridgeline.html>





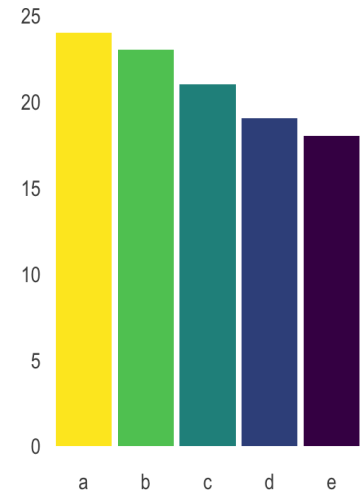
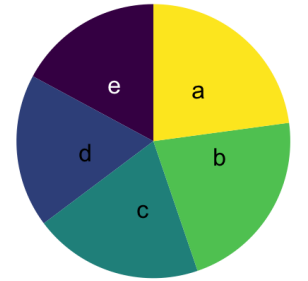
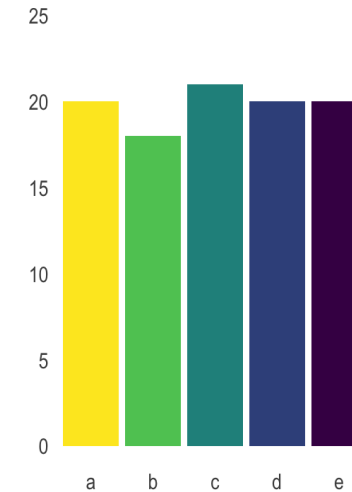
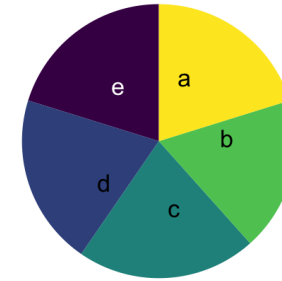
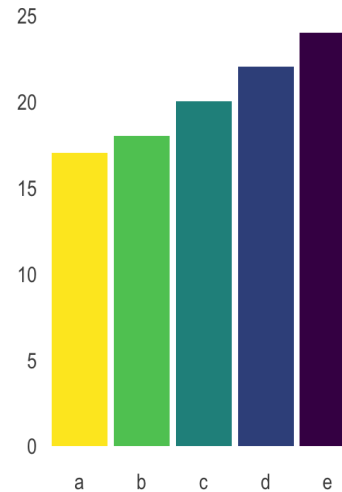
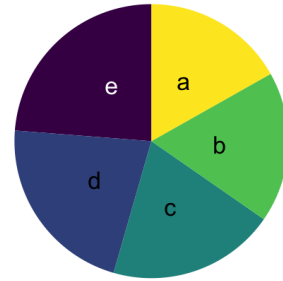


# var cuantitativa + var. categórica

## Gráfico de Torta (Pie chart)

- Usa tamaños relativos para mostrar qué tan grande es cada categoría en relación con el todo.
- **¡resista la tentación de usarlos!**
- Los humanos somos muy malos comparando ángulos.

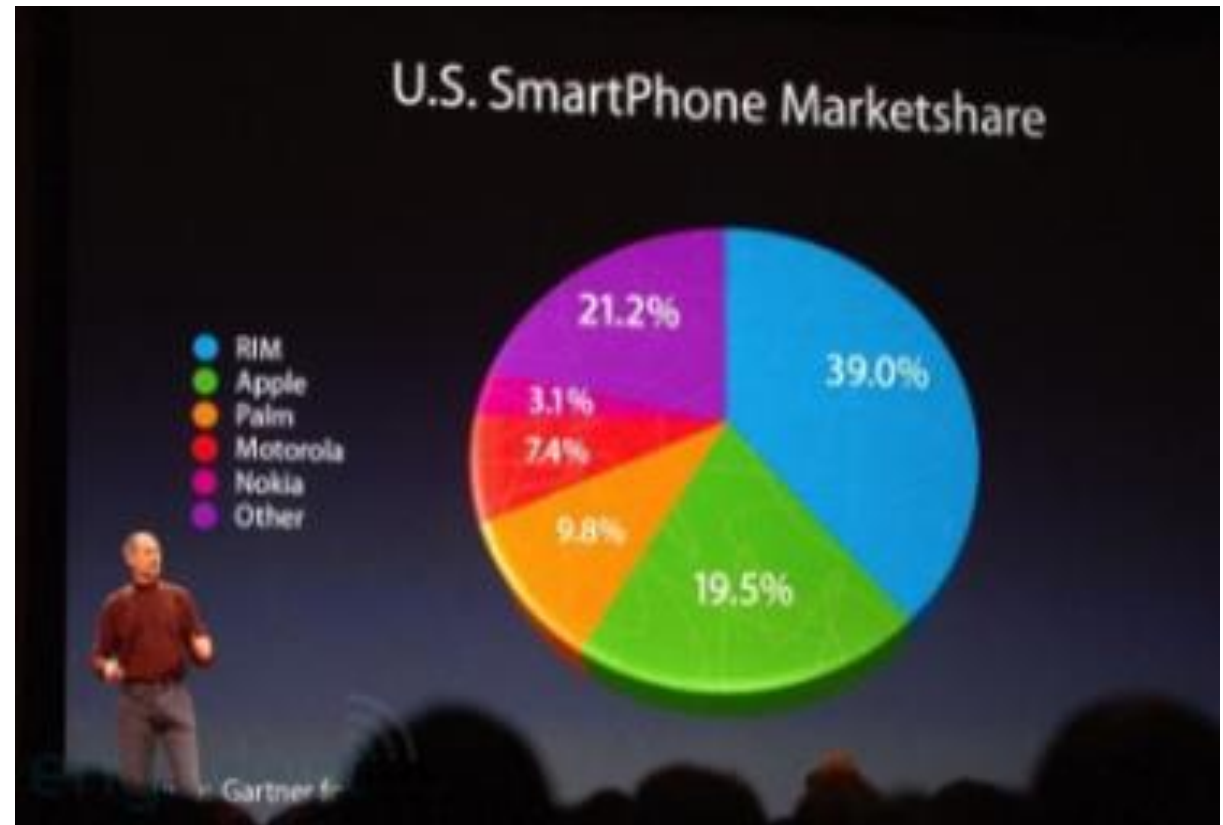
<https://www.data-to-viz.com/caveat/pie.html>



## var cuantitativa + var. categórica Gráfico de Torta (Pie chart)

- Usa tamaños relativos para mostrar qué tan grande es cada categoría en relación con el todo.
- **¡resista la tentación de usarlos!**
- Los humanos somos muy malos comparando ángulos.

<https://www.data-to-viz.com/caveat/pie.html>

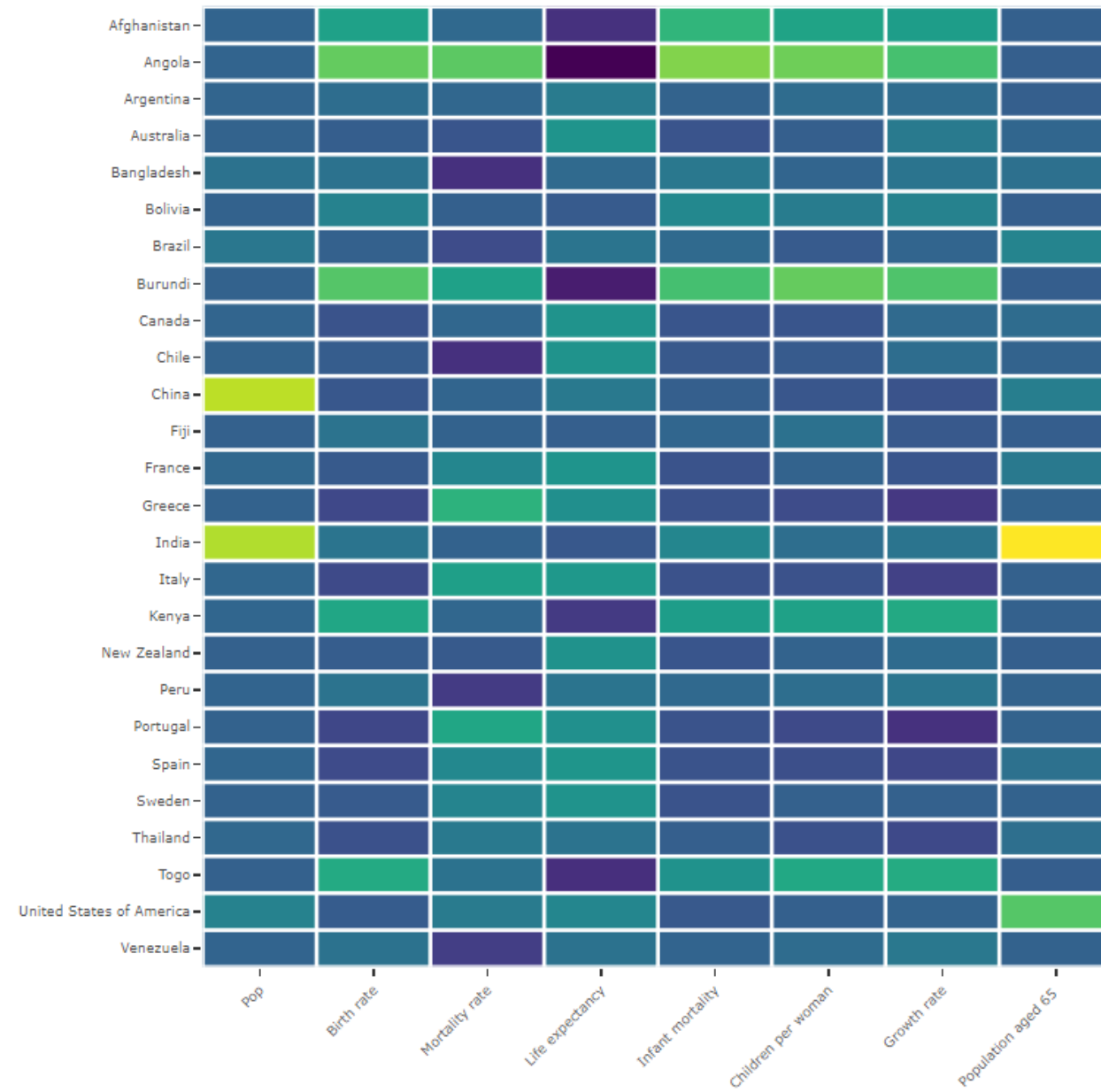


# Multiples variables cuantitativas + variable categórica

## Mapa de calor (Heat map)

- **Representación gráfica de datos donde los valores individuales contenidos en una matriz se representan como colores.**
- Este ejemplo muestra 8 características generales como la población o la esperanza de vida de unos 30 países en 2015.
- Los datos se deben normalizar (se llevan a media 0 y desviación estándar 1) para unificar colores.
- [Datos del Instituto Nacional de Estudios Demográficos de Francia.](https://www.data-to-viz.com/graph/heatmap.html)

<https://www.data-to-viz.com/graph/heatmap.html>

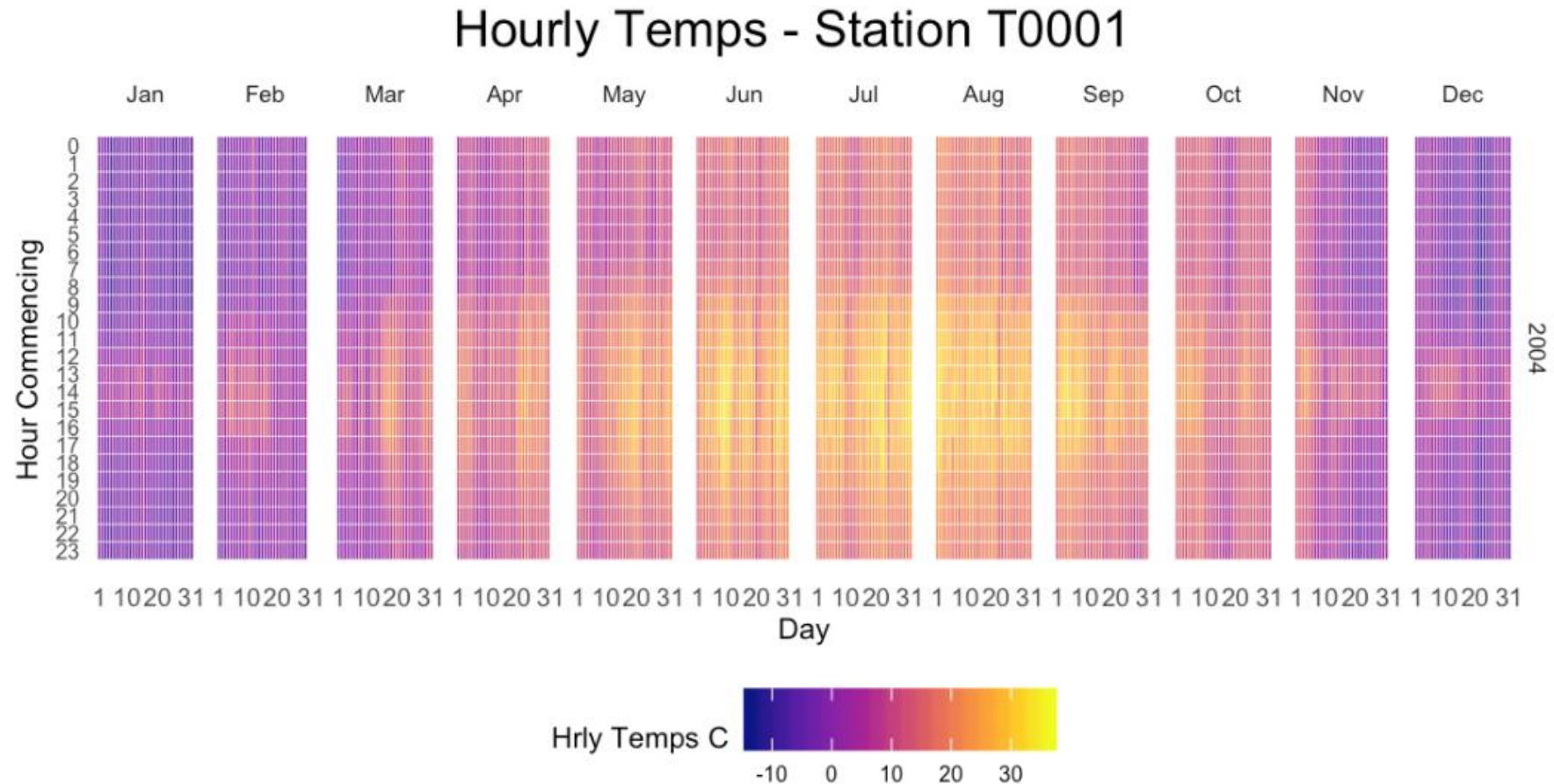


# Series de tiempo

## Mapa de calor horario (Hourly Heat map)

Los mapas de calor pueden ser una muy buena alternativa para visualizar series de tiempo, especialmente cuando el marco de tiempo que se estudia se repite, como semanas.

- <https://r-graph-gallery.com/283-the-hourly-heatmap.html>

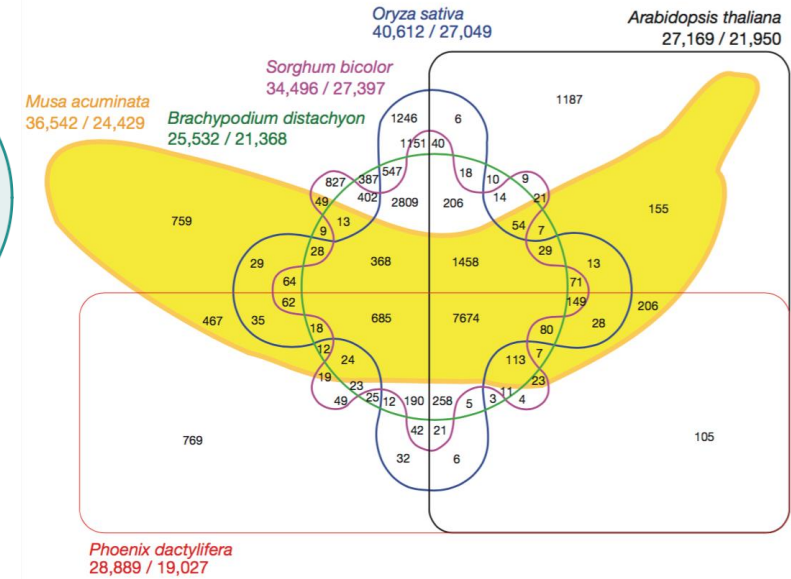
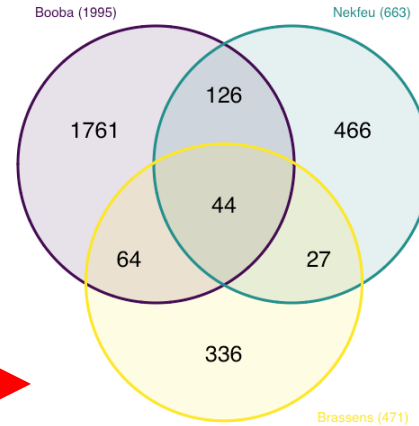


# variable categórica: Venn Diagram – Upset plot

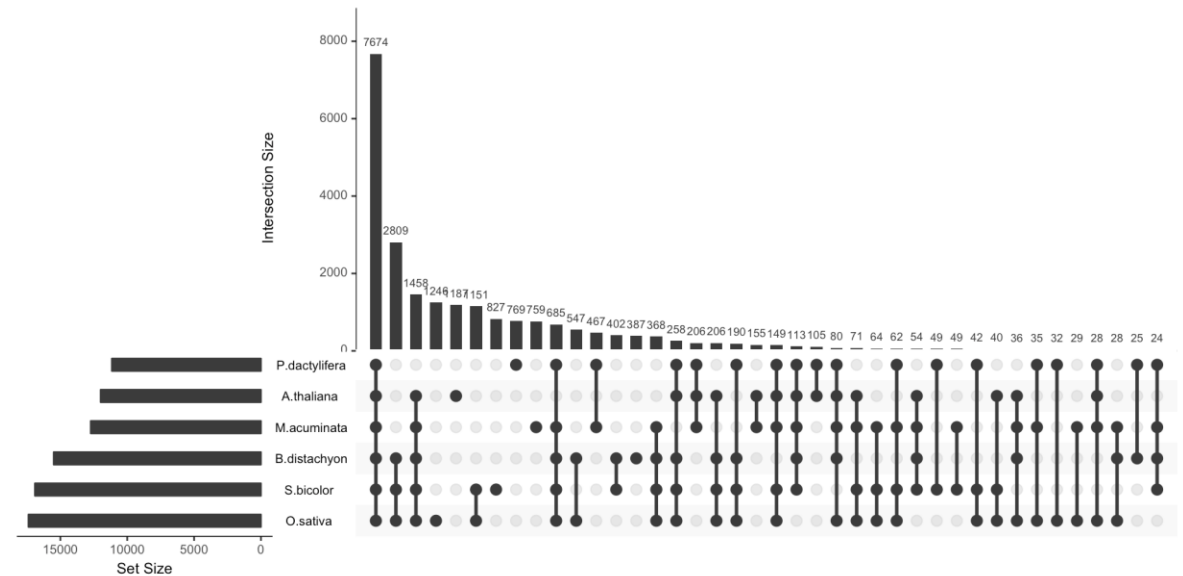
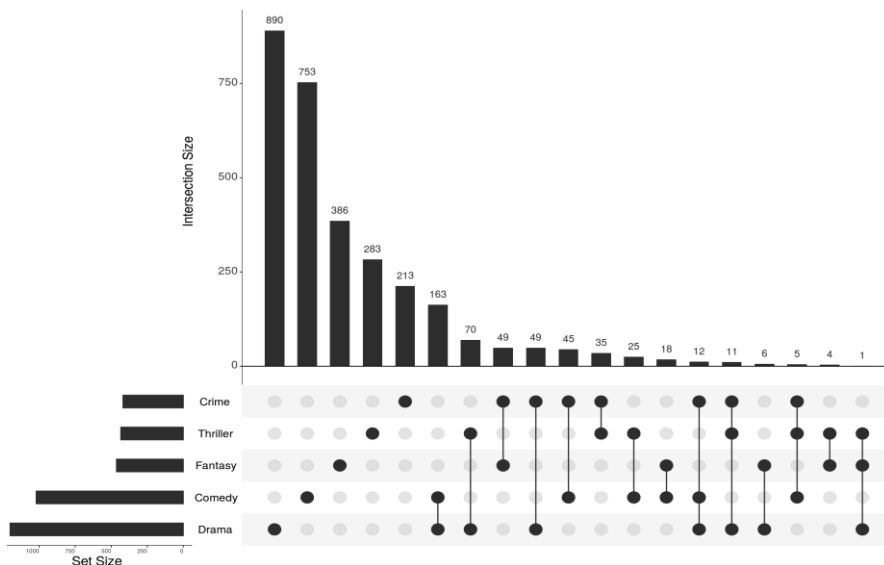
- El diagrama de Venn sólo funciona bien para estudiar la intersección de dos o tres conjuntos.
- El upset plot es una alternativa para visualizar la intersección de más de 3 conjuntos.

<https://www.data-to-viz.com/graph/venn.html>

<https://github.com/hms-dbmi/UpSetR>



Diags. de Venn de 3 y 6 variables



Upset plots para visualizar 5 y 6 variables.

## variable categóricas (anidadas): Mapas de árboles (Treemap)

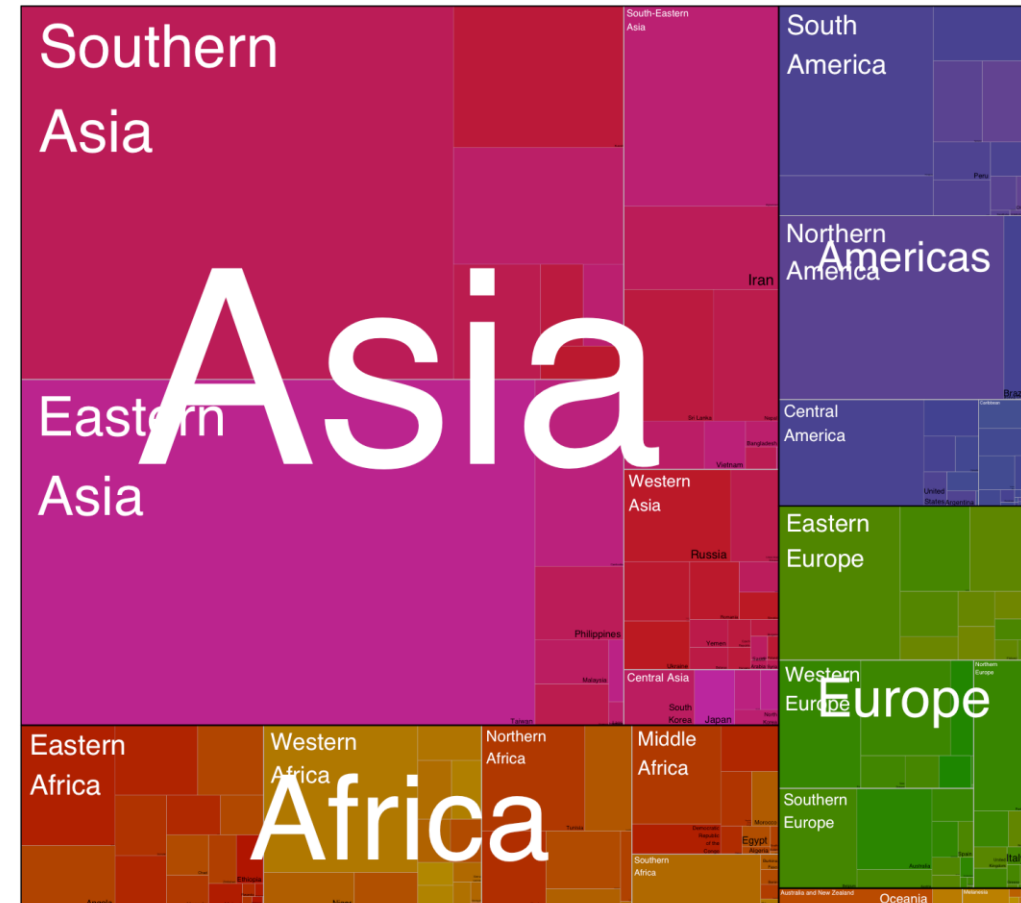
- **Un Treemap muestra datos jerárquicos como un conjunto de rectángulos anidados.**
- Cada grupo está representado por un rectángulo. Usando esquemas de color y/o interactividad, es posible representar varias dimensiones: grupos, subgrupos, etc.
- También se puede combinar con var. cuantitativas (p.ej. aquí el área representa la población).

• <https://www.data-to-viz.com/graph/treemap.html>

**Ver las variaciones interactivas.**

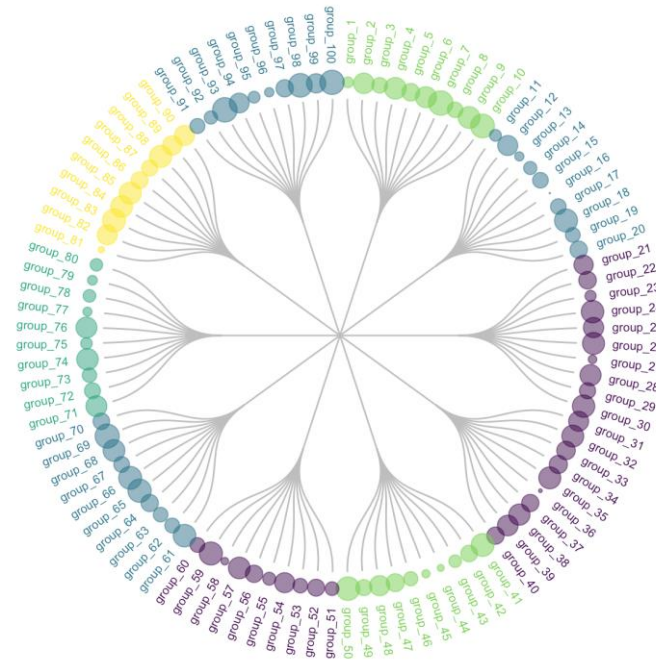
Ver también circular packing <https://www.data-to-viz.com/graph/circularpacking.html>

Aquí hay un ejemplo que describe la población mundial de 250 países. El mundo se divide en continentes (grupo), los continentes se dividen en regiones (subgrupos) y las regiones se dividen en países. En esta estructura de árbol, los países se consideran como hojas: están al final de las ramas.

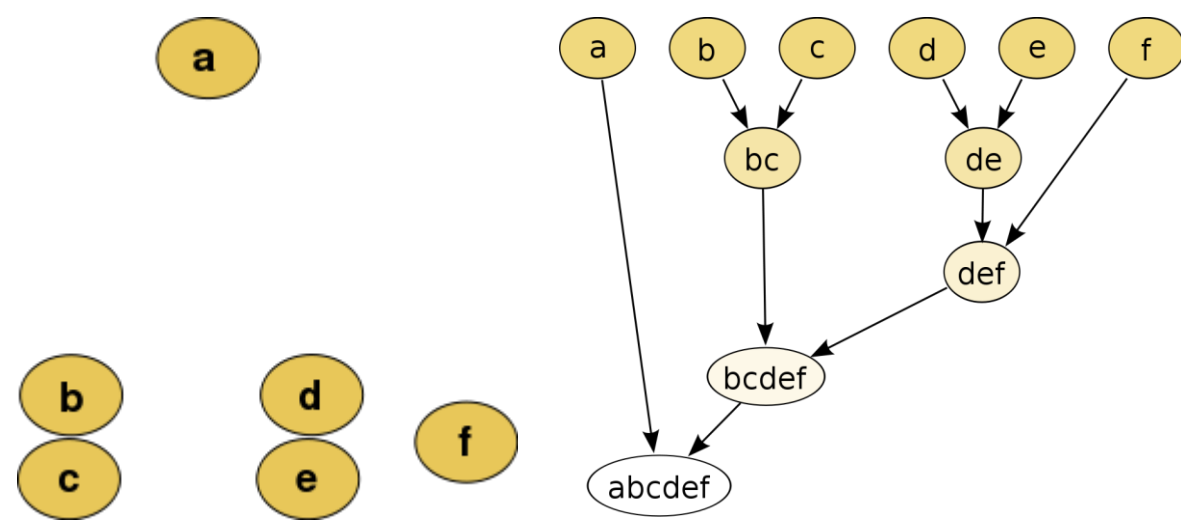


# Dendogramas

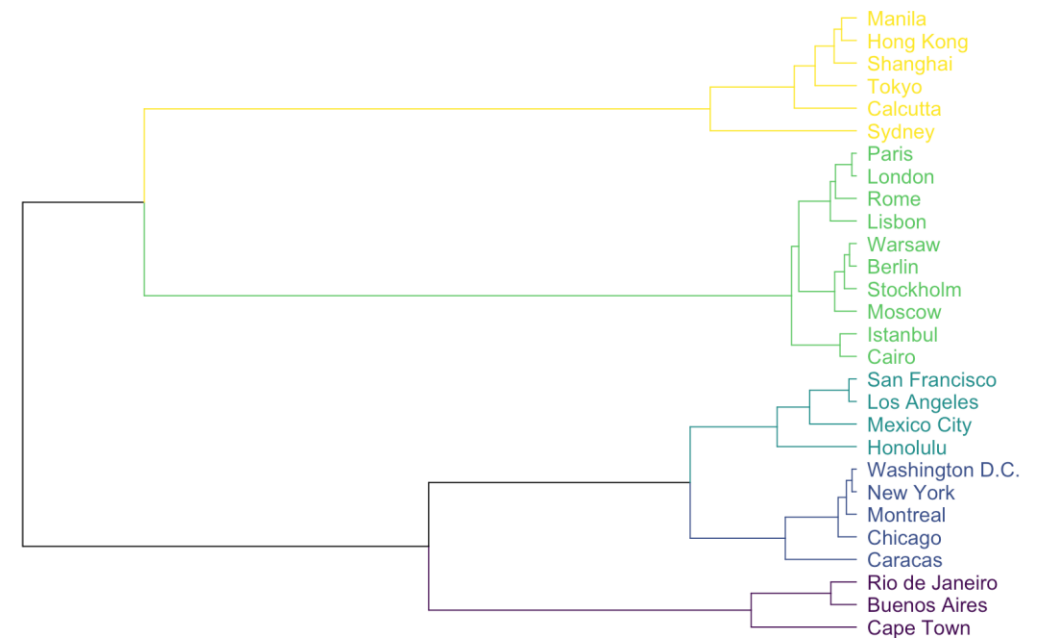
- Se utiliza cuando se quiere agrupar datos que son similares.
- Se mide la distancia euclídea entre los objetos y se van agrupando.



<https://www.data-to-viz.com/graph/dendrogram.html>



Agrupar de forma jerárquica los objetos según su distancia.

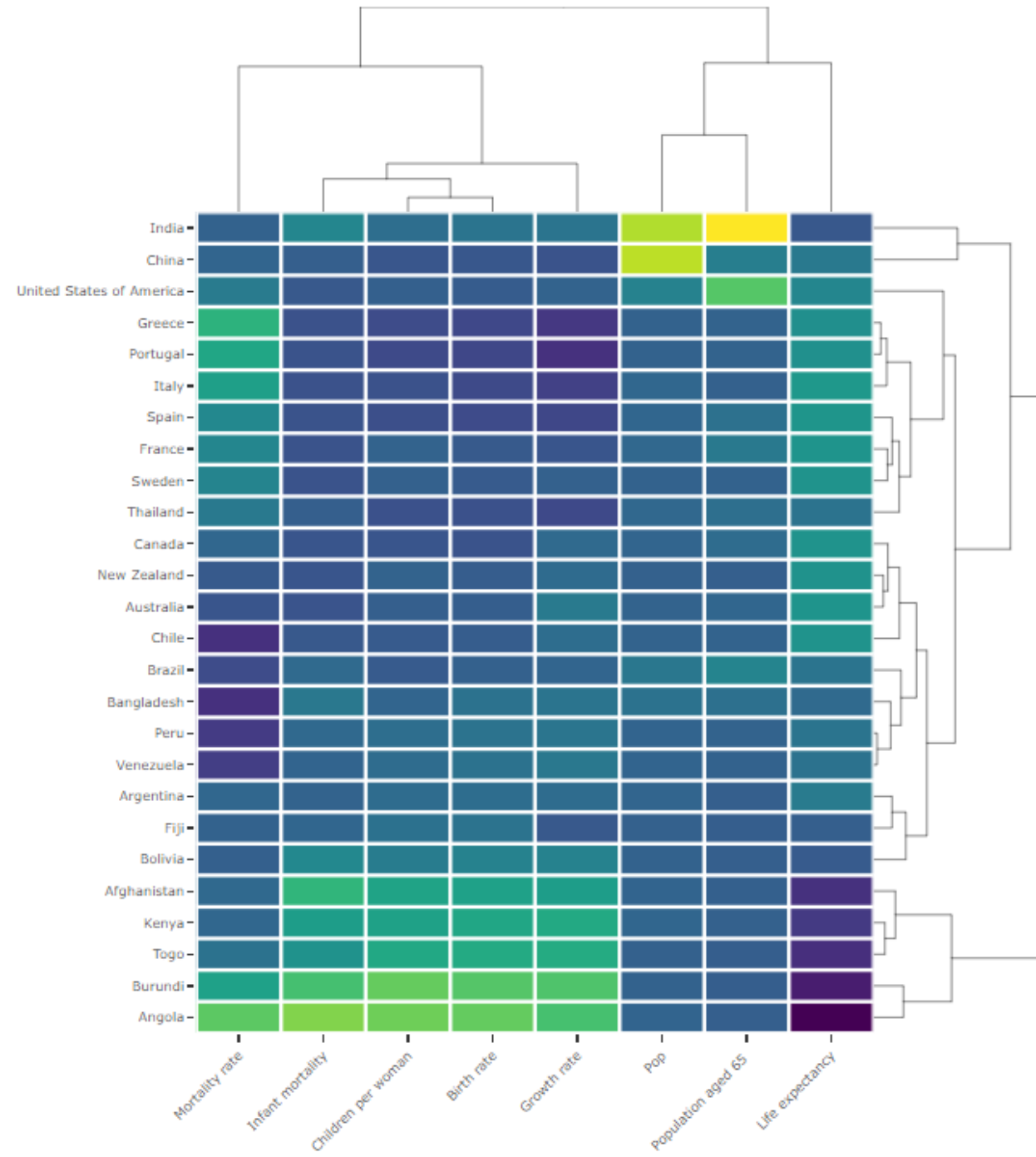




# Dendogramas

- Se utiliza cuando se quiere agrupar datos que son similares.
- Se mide la distancia euclídea entre los objetos y se va agrupando.

• <https://www.data-to-viz.com/graph/dendrogram.html>

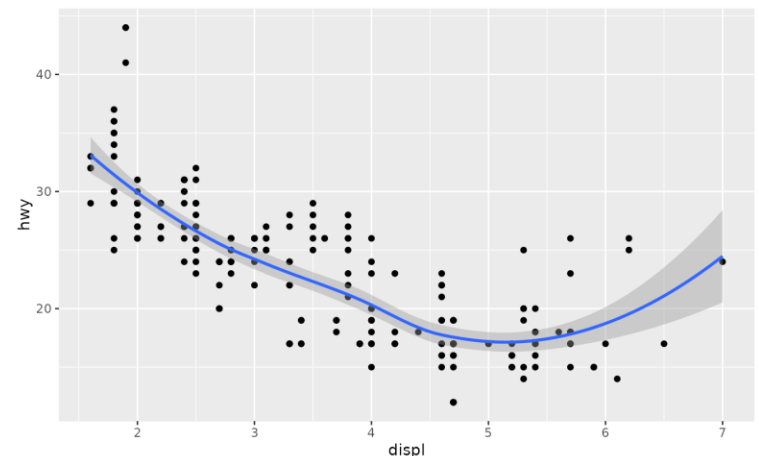
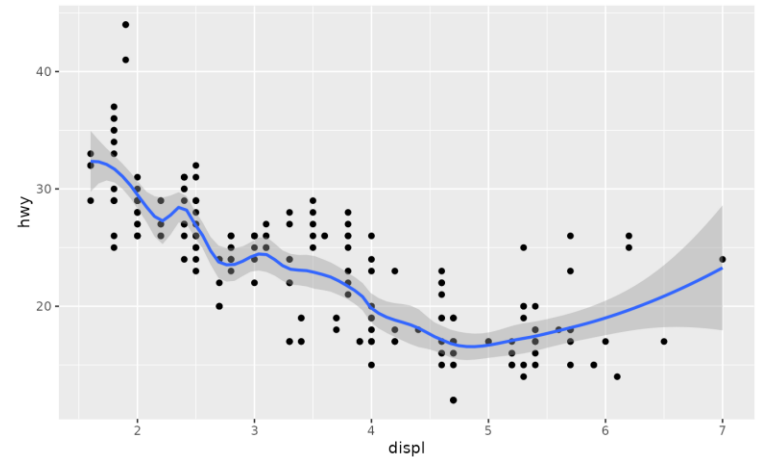
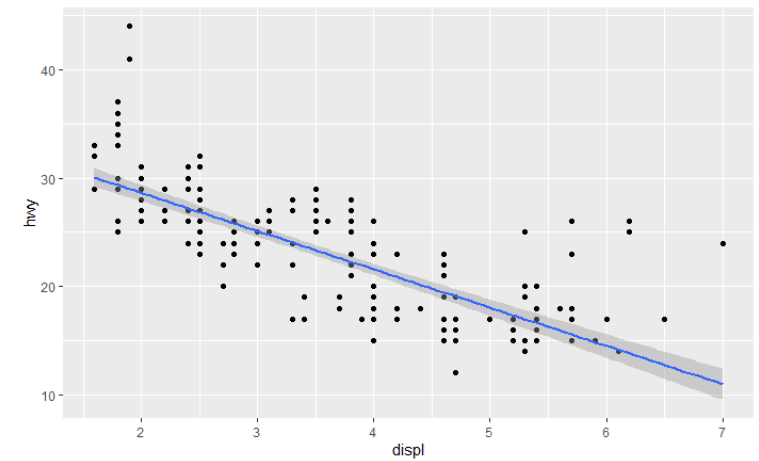


# Modelado: medias condicionales suavizadas

- Ayuda a ver patrones entre variables continuas.
- Se calcula la curva media y el intervalo de confianza.
- Se puede elegir el método de cálculo y el tipo de curva usada para modelar la correlación de variables.

En el help de `geom_smooth` dice: Smoothing method (function) to use, accepts either NULL or a character vector, e.g. "lm", "glm", "gam", "loess" or a function, e.g. MASS::rlm or mgcv::gam, stats::lm, or stats::loess. "auto" is also accepted for backwards compatibility.

[https://ggplot2.tidyverse.org/reference/geom\\_smooth.html](https://ggplot2.tidyverse.org/reference/geom_smooth.html)



# Más gráficos

## WHAT DO YOU WANT TO SHOW ?

- |   |                 |   |           |
|---|-----------------|---|-----------|
|  | Distribution    |  | Evolution |
|  | Correlation     |  | Maps      |
|  | Ranking         |  | Flow      |
|  | Part of a whole |   |           |

<https://www.data-to-viz.com/>

