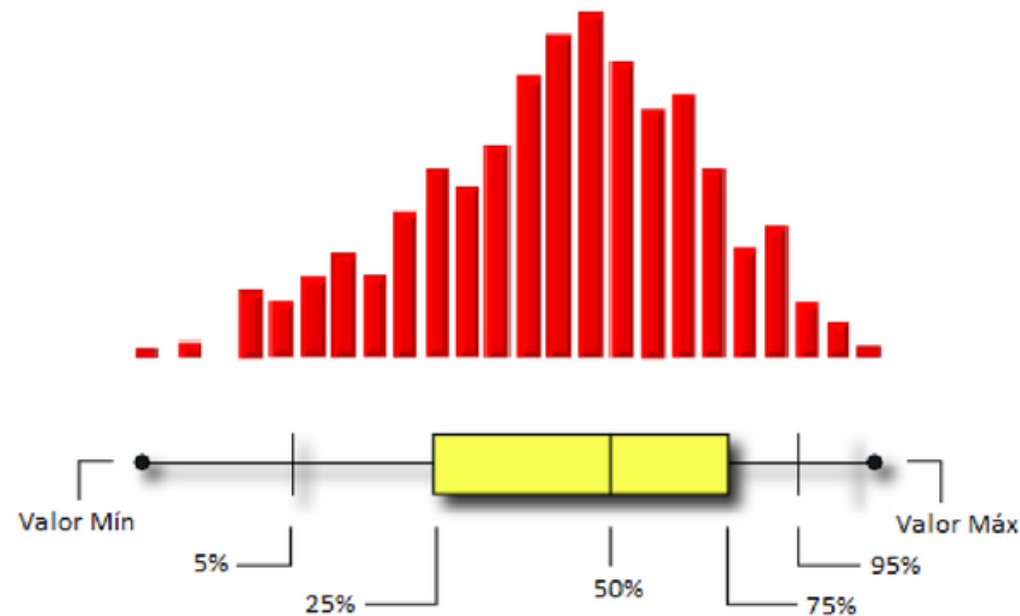


# Estadística descriptiva



## Definición

La estadística descriptiva es la encargada del análisis de los datos recopilados mediante el uso de determinados parámetros relevantes y gráficos que describan la situación de los datos.



Nombre	Definición
Datos <i>ordenados</i>	$x = (x_1, \dots, x_n)$
Media	$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_i$
Varianza	$\hat{s}^2 = \frac{1}{n-1} \sum_{k=1}^n (x_i - \bar{x})^2$
Desviación típica	$\hat{s} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_i - \bar{x})^2}$
Percentil $k \cdot 100\%$	$x_{i_0}$ con $i_0 = \lceil k \cdot n \rceil$
Q1	$Q_1 = \text{Percentil } 25\%$
Mediana	$Q_2 = \text{Percentil } 50\%$
Q3	$Q_3 = \text{Percentil } 75\%$
Moda	$x_{i_0}$ valor más repetido.

Gráficas principales
Histograma
Diagrama de barras
Diagrama de Pareto
Gráfico circular
Boxplot
Diagrama de tallo y hoja
Diagrama de dispersión
Serie temporal
Gráfico de correlaciones



### Formato ideal de datos

Variables de estudio	$X_1$	$X_2$	$\dots$	$X_n$
Observación 1	$x_{11}$	$x_{12}$	$\dots$	$x_{1n}$
Observación 2	$x_{21}$	$x_{22}$	$\dots$	$x_{2n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Observación m	$x_{m1}$	$x_{m2}$	$\dots$	$x_{mn}$

```
# Librería
library("tidyverse")

# Datos de vehículos
df = mpg
```

Variable	Descripción
manufacturer	empresa constructora del vehículo
model	modelo del vehículo
displ	tamaño del motor en litros
year	año de construcción
cyl	número de cilindros
trans	típo de transmisión
drv	El tipo de tracción
cty	consumo en ciudad (millas por galón)
hwy	consumo en carretera (millas por galón)
fl	tipo de combustible
class	clase del vehículo

```
# Librería  
library("tidyverse")
```

```
# Datos de vehículos  
df = mpg
```

Variable	Descripción
manufacturer	empresa constructora del vehículo
model	modelo del vehículo
displ	tamaño del motor en litros
year	año de construcción
cyl	número de cilindros
trans	típo de transmisión
drv	El tipo de tracción
cty	consumo en ciudad (millas por galón)
hwy	consumo en carretera (millas por galón)
fl	tipo de combustible
class	clase del vehículo

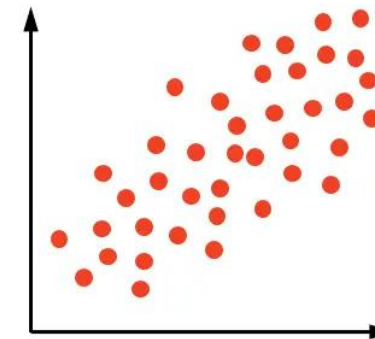
## Ejemplo

¿Existe alguna relación entre el tamaño del motor y el consumo en carretera?

### Ventajas

- Correlaciones
- Estudio previo
- Fácil

### Diagrama de dispersión



### Desventajas

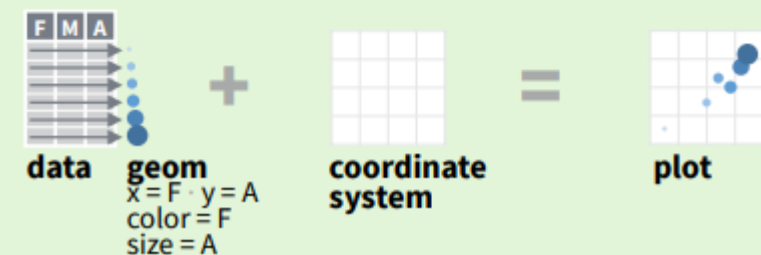
- NO cualitativas
- NO causa-efecto
- Solo 2 variables

## Basics

**ggplot2** is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and **geoms**—visual marks that represent data points.



To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.



Complete the template below to build a graph.

```
ggplot (data = <DATA>) +  
  <GEOM_FUNCTION> (mapping = aes(<MAPPINGS>),  
    stat = <STAT>, position = <POSITION>) +  
  <COORDINATE_FUNCTION> +  
  <FACET_FUNCTION> +  
  <SCALE_FUNCTION> +  
  <THEME_FUNCTION>
```

required

Not required, sensible defaults supplied

**ggplot(data = mpg, aes(x = cty, y = hwy))** Begins a plot that you finish by adding layers to. Add one geom function per layer.

**last\_plot()** Returns the last plot.

**ggsave("plot.png", width = 5, height = 5)** Saves last plot as 5' x 5' file named "plot.png" in working directory. Matches file type to file extension.

## Aes Common aesthetic values.

**color** and **fill** - string ("red", "#RRGGBB")

**linetype** - integer or string (0 = "blank", 1 = "solid", 2 = "dashed", 3 = "dotted", 4 = "dotdash", 5 = "longdash", 6 = "twodash")

**lineend** - string ("round", "butt", or "square")

**linejoin** - string ("round", "mitre", or "bevel")

**size** - integer (line width in mm)

**shape** - integer/shape name or a single character ("a")



## Ejemplo

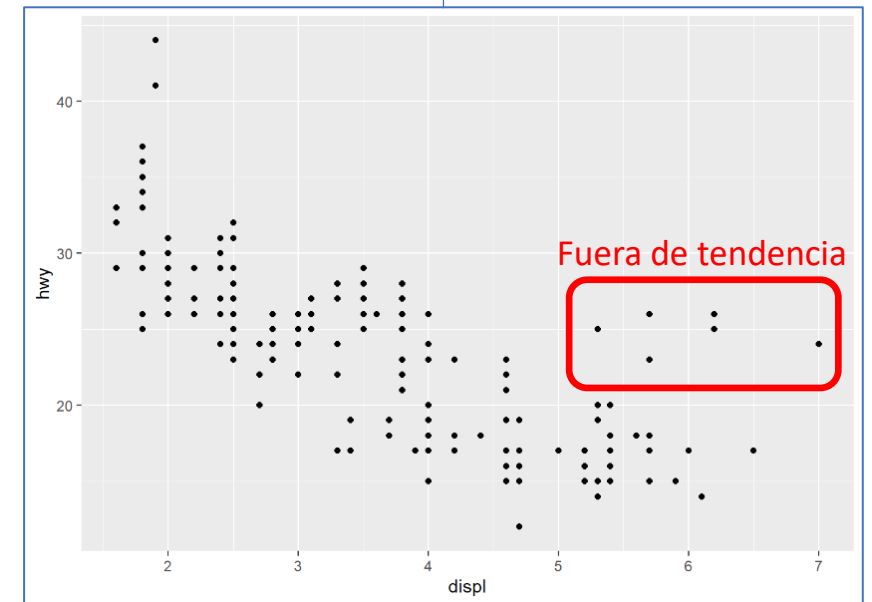
¿Existe alguna relación entre el tamaño del motor y el consumo en carretera?

Observamos una tendencia general de correlación entre displ y hwy.

```
# Librería
library("tidyverse")

# Datos de vehículos
df = mpg

# geom_point para diagrama de dispersión.
ggplot( data = df ) +
  geom_point( mapping = aes(x = displ, y = hwy))
```



## Ejemplo

¿Existe alguna relación entre el número de cilindros y el consumo en carretera?

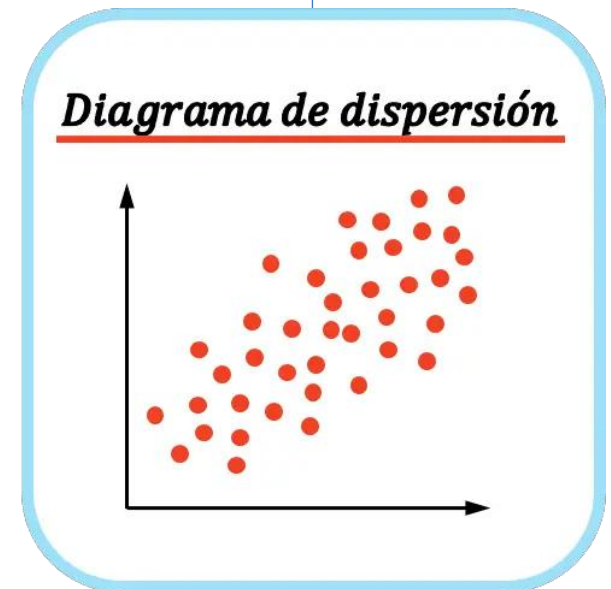


```
# Librería
library("tidyverse")

# Datos de vehículos
df = mpg

# geom_point para diagrama de dispersión.
ggplot( data = df ) +
  geom_point( mapping = aes(???????????) )
```

????????







## Ejemplo

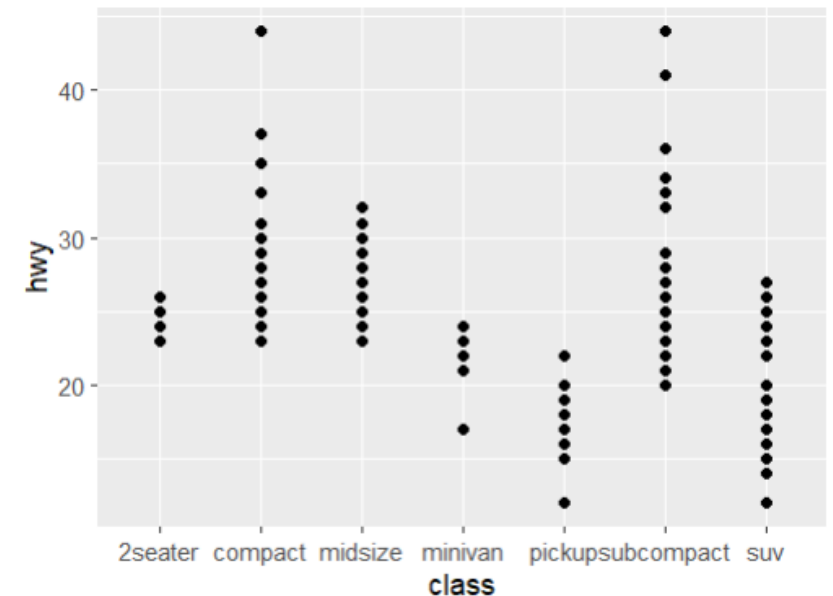
¿Dispersión de discreta y continua? Por ejemplo  
clase de vehículo y consumo en carretera.

Mejor boxplot.

```
# Librería
library("tidyverse")

# Datos de vehículos
df = mpg

# geom_point para diagrama de dispersión.
ggplot(data = df) +
  geom_point(mapping = aes(x = class, y = hwy))
```



## Ejemplo

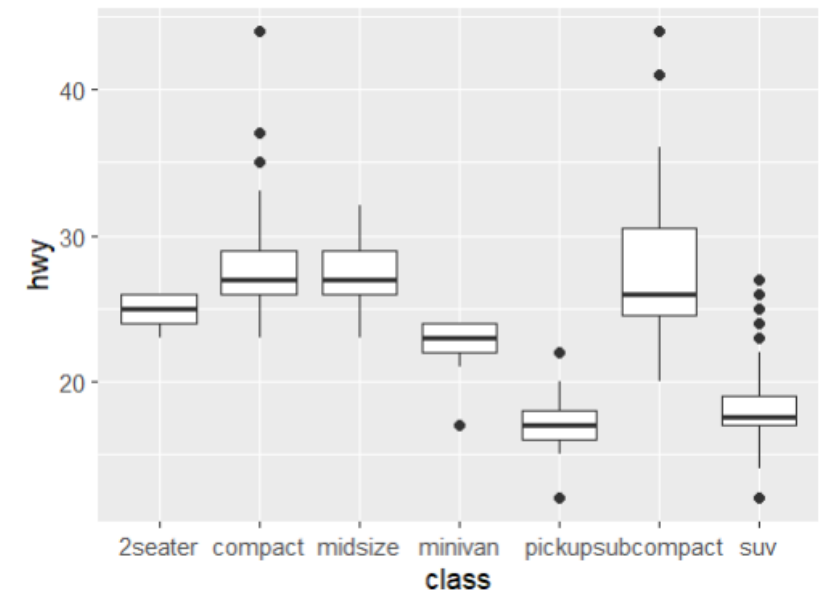
¿Dispersión de discreta y continua? Por ejemplo clase de vehículo y consumo en carretera.

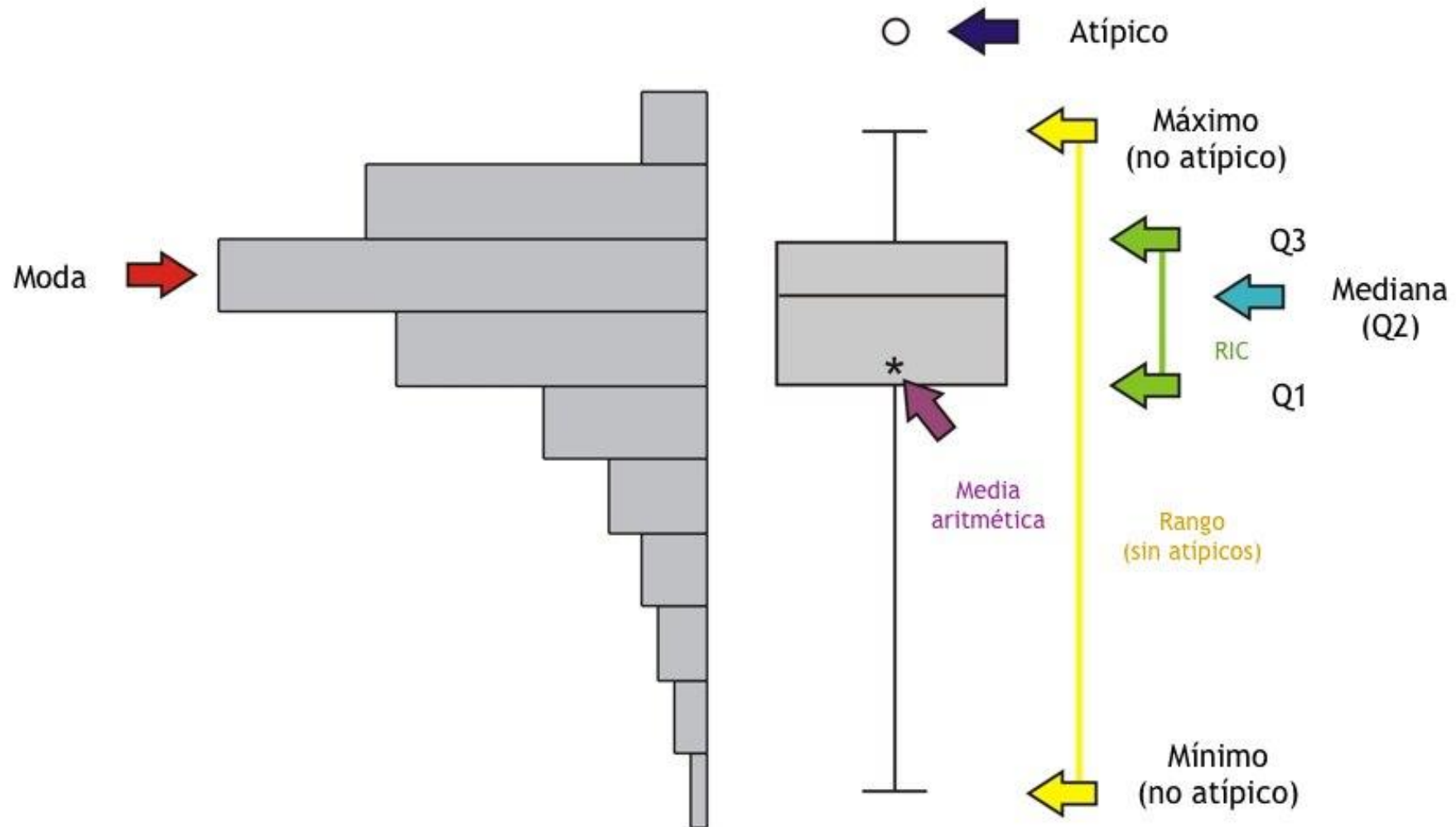
Mejor boxplot.

```
# Librería
library("tidyverse")

# Datos de vehículos
df = mpg

# Boxplot.
ggplot(data = df) +
  geom_boxplot(mapping = aes(x = class, y = hwy))
```





```
# Librería
library("tidyverse")

# Datos de diamantes
df = diamonds
```

Variable	Descripción
carat	Peso del diamante
cut	Calidad del corte
color	Color del diamante
clarity	Claridad del diamante
x	Longitud (mm)
y	Ancho (mm)
z	Profundidad (mm)
depth	Porcentaje de profundidad total
table	Anchura parte superior
price	Precio del diamante

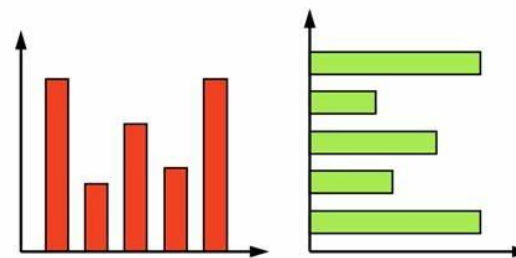
## Ejemplo

¿Qué tipo de distribución siguen los datos?

### Barras

Para variables  
discretas.

### Diagrama de barras



### Histograma

Para variables  
continuas.

## Ejemplo

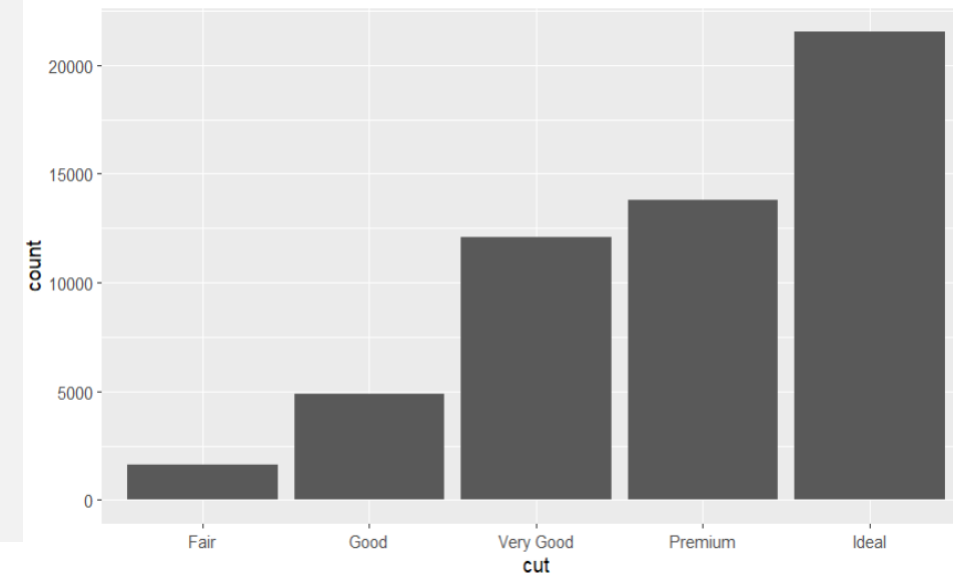
¿Cuántos diamantes hay para cada tipo de corte?

Podemos observar que la moda está en el corte ideal, habiendo más de 20.000 cortes ideales.

```
# Librería
library("tidyverse")

# Datos de diamantes
df = diamonds

# Diagrama de barras.
ggplot( data = df ) +
  geom_bar( mapping = aes(x = cut))
```



## Ejemplo

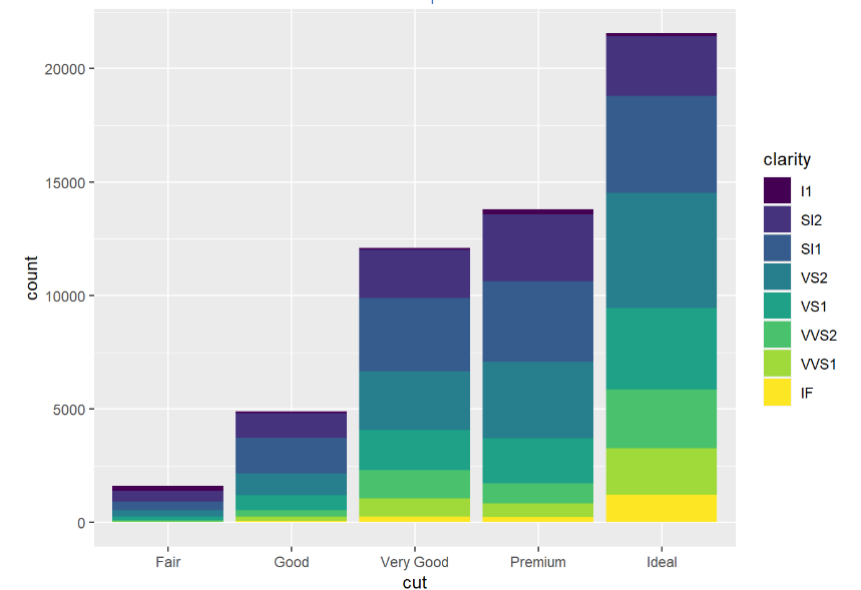
¿Cómo distinguir en cada barra según claridad (discreta)?

```
# Librería
library("tidyverse")

# Datos de diamantes
df = diamonds

# Diagrama de barras.
ggplot( data = df ) +
  geom_bar(mapping = aes(x = cut, fill = clarity))
```

En cada categoría distinguimos según claridad.





## Ejemplo

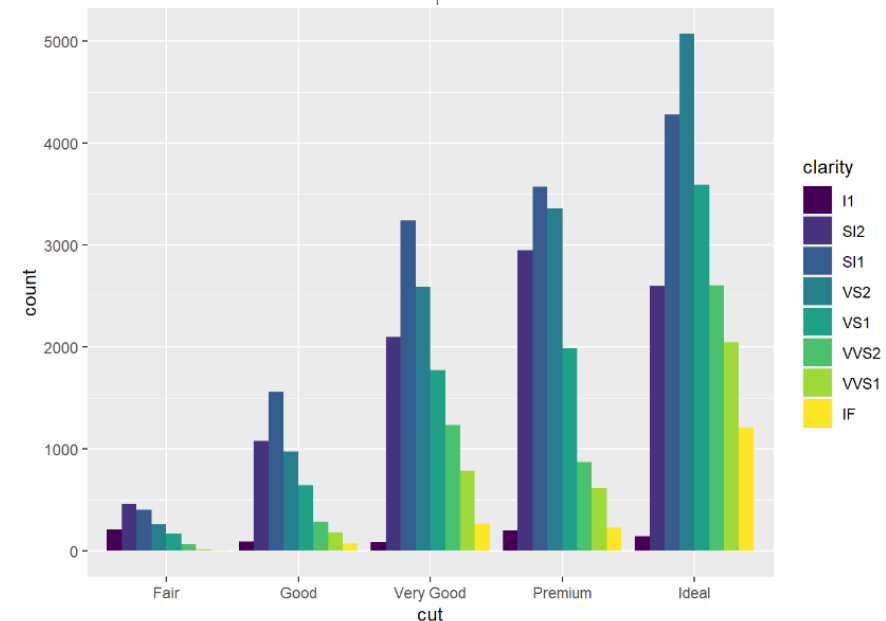
¿Cómo dividir cada barra según claridad (discreta)?

En cada categoría distinguimos según claridad.

```
# Librería
library("tidyverse")

# Datos de diamantes
df = diamonds

# Diagrama de barras.
ggplot( data = df ) +
  geom_bar(mapping = aes(x = cut, fill = clarity),
           position = "dodge")
```



## Ejemplo

¿Cuántos diamantes hay para cada precio?

Con binwidth elegimos el tamaño de las barras.

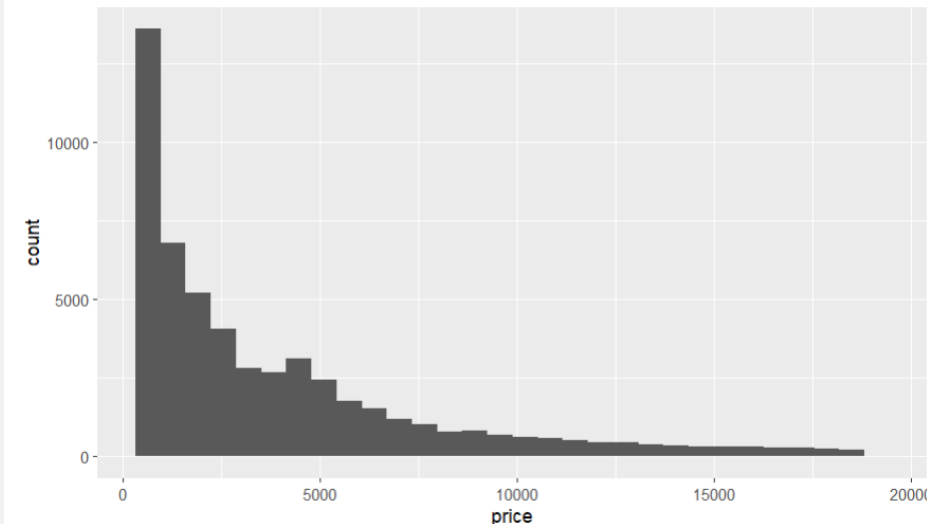
```
# Librería  
library("tidyverse")
```

```
# Datos de diamantes  
df = diamonds
```

```
# Histograma.  
ggplot( data = df ) +  
  geom_histogram( mapping = aes(x = price))
```

Prueba bar

binwidth = controla el ancho de las barras





## Ejemplo

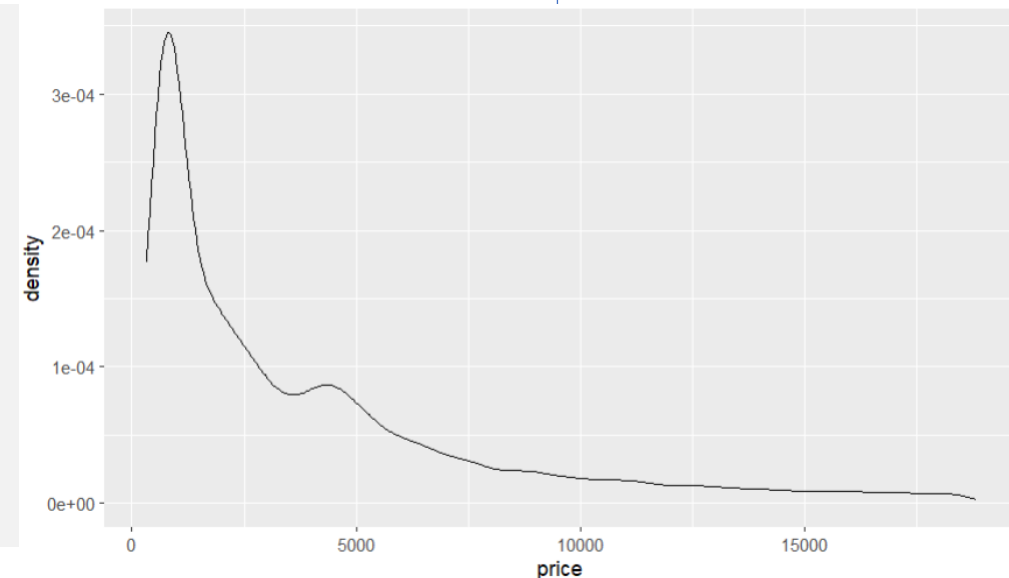
¿Cuántos diamantes hay para cada precio?

En forma de función de densidad.

```
# Librería
library("tidyverse")

# Datos de diamantes
df = diamonds

# geom_bar para diagrama de dispersión.
ggplot( data = df ) +
  geom_density( mapping = aes(x = price))
```





# R CHARTS

por R CODER

**¡Bienvenido a R CHARTS! En este sitio encontrarás ejemplos de gráficos de R con su código hechos con el paquete graphics de R base, ggplot2 y otras librerías. Contribuye, sugiere nuevas visualizaciones o soluciona cualquier bug a través de GitHub**

[Enlace](#)