

# Cheat sheet Práctica 3: Estadística Descriptiva en R

Juan Antonio Villatoro García – Grupo BioestadísticaR



**UNIVERSIDAD  
DE GRANADA**

**Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo BioestadísticaR formado por:**

**Antonio Martín Andrés**

**Juan de Dios Luna del Castillo,**

**Pedro Femia Marzo,**

**Miguel Ángel Montero Alonso,**

**Christian José Acal González,**

**Pedro María Carmona Sáez,**

**Juan Manuel Melchor Rodríguez,**

**José Luis Romero Béjar,**

**Manuela Expósito Ruíz,**

**Juan Antonio Villatoro García.**

**Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de BioestadísticaR.**

**Todos los materiales están protegidos por la Licencia Creative Commons CC BY-NC-ND que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".**

# Práctica 3: : Estadística Descriptiva en R



UNIVERSIDAD DE GRANADA

## Distribuciones de frecuencias

### Obtención paso a paso

ORDEN	FRECUENCIA
<code>table(x)</code>	Frecuencia absoluta. Dentro del paréntesis se coloca la variable que se desea analizar
<code>cumsum(x)</code>	Frecuencia absoluta acumulada. Donde x es un objeto de tipo <code>table</code> .
<code>prop.table(x)</code>	Frecuencia relativa. Donde x es un objetivo del tipo <code>table</code> .
<code>cumsum(prop.table(x))</code>	Frecuencia relativa acumulada. Donde x es un objeto de tipo <code>table</code> .

Al final se uno todo con la función `cbind()`

```
> frec<-table(osteo$grupo_edad)
> frec.acum<-cumsum(frec)
> prop<-prop.table(frec)
> prop.acum<-cumsum(prop)
> tabla.frecuencias<-cbind(frec,
+   frec.acum,prop,prop.acum)
> tabla.frecuencias
```

	frec	frec.acum	prop	prop.acum
< 25	32	32	0.3404255	0.3404255
25 - 33	32	64	0.3404255	0.6808511
> 33	30	94	0.3191489	1.0000000

### Obtención directa

Librería `summarytools` → función `freq()`

```
> library(summarytools)
> freq(osteo$grupo_edad)
```

Frequencies  
osteo\$grupo\_edad  
Type: Factor

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
< 25	32	34.04	34.04	34.04	34.04
25 - 33	32	34.04	68.09	34.04	68.09
> 33	30	31.91	100.00	31.91	100.00
<NA>	0			0.00	100.00
Total	94	100.00	100.00	100.00	100.00

## Síntesis de datos

### Medidas de posición

FUNCIÓN	SIGNIFICADO
<code>mean(x)</code>	Media aritmética
<code>median(x)</code>	Mediana
<code>mfv(x)</code>	Moda (librería <code>modest</code> )
<code>quantile(x, p)</code>	Cuantiles. Donde p es el percentil que se desea calcular

```
> mean(osteo$peso)
[1] 63.83936
> median(osteo$peso)
[1] 62
> library(modeest)
> mfv(osteo$peso)
[1] 62
> quantile(osteo$peso)
 0%  25%  50%  75% 100%
44.60 56.25 62.00 69.75 99.00
> quantile(osteo$peso, probs = c(0.1,0.9))
 10%  90%
49.15 82.38
```

### Medidas de dispersión

ORDEN	SIGNIFICADO
<code>max(x)-min(x)</code>	Rango
<code>var(x)</code>	Varianza muestral
<code>Var(x) * (n-1) / n</code>	Varianza
<code>sd(x)</code>	Desviación típica muestral
<code>sd(x)/var(x)</code>	Coefficiente de variación

```
> max(osteo$peso) - min(osteo$peso)
[1] 54.4
> var(osteo$peso)
[1] 139.3403
> var(osteo$peso)*(length(osteo$peso)-1)/
+   length(osteo$peso)
[1] 137.8579
> sd(osteo$peso)
[1] 11.80425
> sd(osteo$peso)/(mean(osteo$peso))
[1] 0.1849055
```

### Funciones que contienen varias medidas

- Función `summary()`: Devuelve todas las medidas salvo la desviación típica y la varianza
- Función `descr()` de la librería `summarytools`: Devuelve todas las medidas que le indiquemos

```
> summary(osteo$peso)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 44.60  56.25   62.00   63.84  69.75   99.00
```

```
> descr(osteo$peso)
Descriptive Statistics
osteo$peso
N: 94
```

	peso
Mean	63.84
Std.Dev	11.80
Min	44.60
Q1	56.00
Median	62.00
Q3	70.00
Max	99.00
MAD	10.38
IQR	13.50
CV	0.18
Skewness	0.68
SE.Skewness	0.25
Kurtosis	0.14
N.Valid	94.00
Pct.Valid	100.00

## Representaciones gráficas:

### Librería `ggplot2`

Necesita 3 elementos para representar un gráfico:

- Un `dataframe` que contenga los datos que se va a representar
- Una lista de relaciones que permitan interactuar a las variables entre sí
- Los `geoms` que representan los elementos geométricos del gráfico

`ggplot(data.frame, aes(x = variable)) + geom_forma()`

# Práctica 3: : Estadística Descriptiva en R



UNIVERSIDAD DE GRANADA

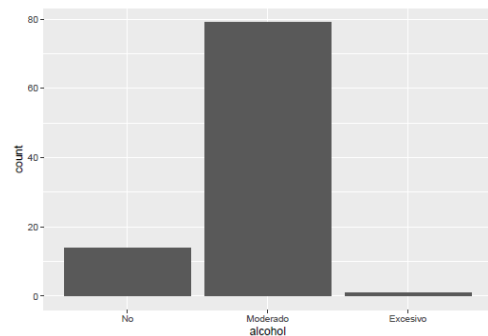
## Representaciones gráficas en la librería ggplot2

### Diagrama de Barras

#### Representación básica

```
ggplot(data.frame, aes(x = variable)) + geom_bar()
```

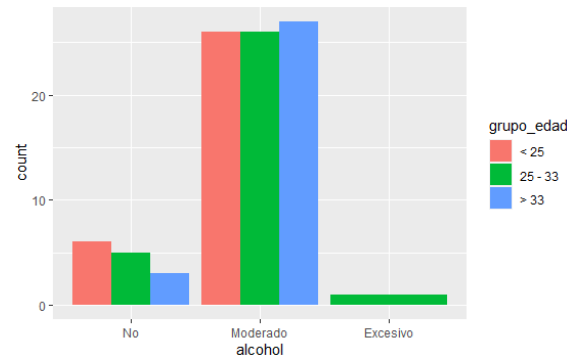
```
> ggplot(osteo, aes(x = alcohol)) + geom_bar()
```



#### Representación en función de las categorías de otra variable

```
ggplot(data.frame, aes(x = variable, fill = variable con las categorías)) + geom_bar(position = "dodge")
```

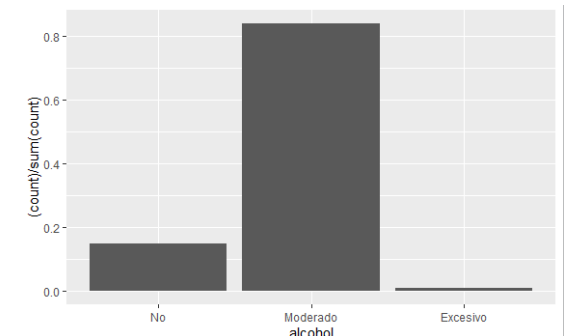
```
> ggplot(osteo, aes(x = alcohol, fill = grupo_edad)) + geom_bar(position="dodge")
```



#### Representación de porcentajes

```
ggplot(data.frame, aes(x = variable, (count)/sum(count))) + geom_bar()
```

```
> ggplot(osteo, aes(x = alcohol, (count)/sum(count))) + geom_bar()
```

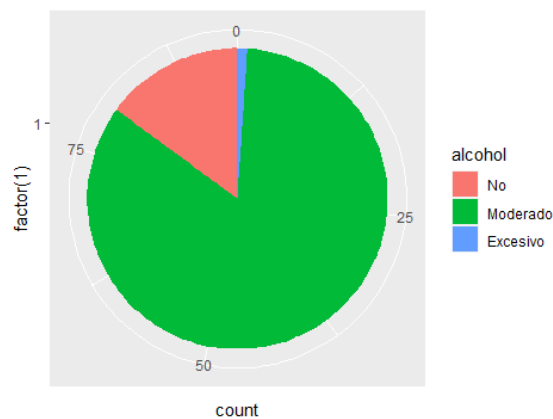


### Diagrama de sectores

#### Representación básica

```
ggplot(data.frame, aes(x=factor(1), fill=variable))+ geom_bar() + coord_polar("y")
```

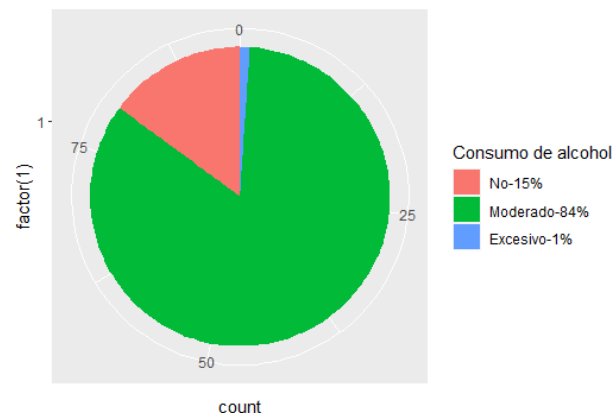
```
> sectores <- ggplot(osteo, aes(x=factor(1), fill=alcohol))+ geom_bar()+coord_polar("y")> sectores
```



#### Modificación de la leyenda

```
ggplot(data.frame, aes(x=factor(1), fill=variable))+ geom_bar() + coord_polar("y") + scale_fill_discrete
```

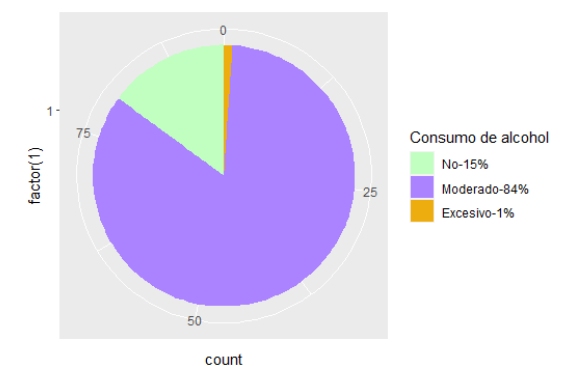
```
> sectores + scale_fill_discrete(name="Consumo de alcohol", labels=c("No-15%", "Moderado-84%", "Excesivo-1%"))
```



#### Modificación de la leyenda y colores

```
ggplot(data.frame, aes(x=factor(1), fill=variable))+ geom_bar() + coord_polar("y") + scale_fill_manual()
```

```
> sectores + scale_fill_manual(values=c("darkseagreen1", "mediumpurple1", "darkgoldenrod2"), name="Consumo de alcohol", labels=c("No-15%", "Moderado-84%", "Excesivo-1%"))
```



# Práctica 3: : Estadística Descriptiva en R



UNIVERSIDAD DE GRANADA

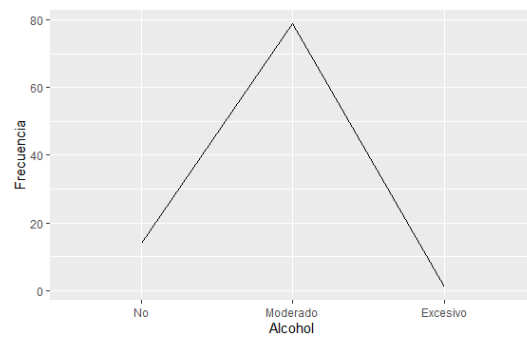
## Representaciones gráficas en la librería ggplot2

### Gráfico de líneas

#### Representación básica

```
ggplot(data.frame(table(data.frame)), aes(x=variable,  
y = Frecuencias absoluta/relativa, group=1)) +  
geom_line()
```

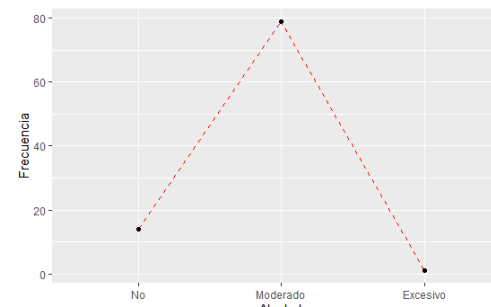
```
> df=table(osteo$alcohol)  
> df=as.data.frame(df)  
> colnames(df)=c("Alcohol", "Frecuencia")  
> ggplot(df, aes(x=Alcohol,y=Frecuencia,  
+ group=1)) + geom_line()
```



#### Cambios de líneas y añadir puntos

```
ggplot(data.frame(table(data.frame)), aes(x=variable,  
y = Frecuencias absoluta/relativa, group=1)) +  
geom_line(linetype = "tipo",col="color")+geom_point()
```

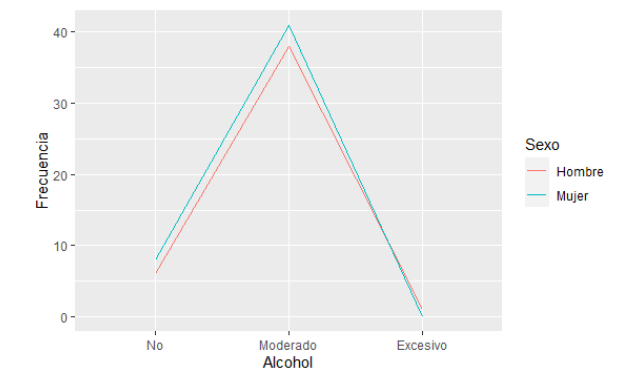
```
> ggplot(df, aes(x=Alcohol,  
+ y=Frecuencia,group=1)) +  
+ geom_line(linetype = "dashed",col="red")+  
+ geom_point()
```



#### Representación en función de las categorías de otra variable

```
ggplot(data.frame(table(data.frame)), aes(x=variable,  
y = Frecuencias absoluta/relativa, group=variable de  
categorias)) + geom_line()
```

```
> df=table(osteo$alcohol,osteo$sexo)  
> df=as.data.frame(df)  
> colnames(df)=c("Alcohol", "Sexo", "Frecuencia")  
> ggplot(df, aes(x=Alcohol,y=Frecuencia,  
+ group=Sexo,colour=Sexo)) + geom_line()
```

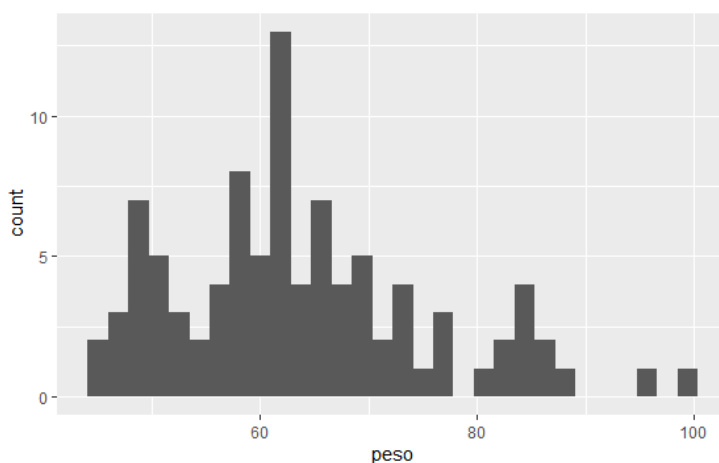


### Histograma

#### Representación básica

```
ggplot(data.frame, aes(x = variable)) +  
geom_histogram()
```

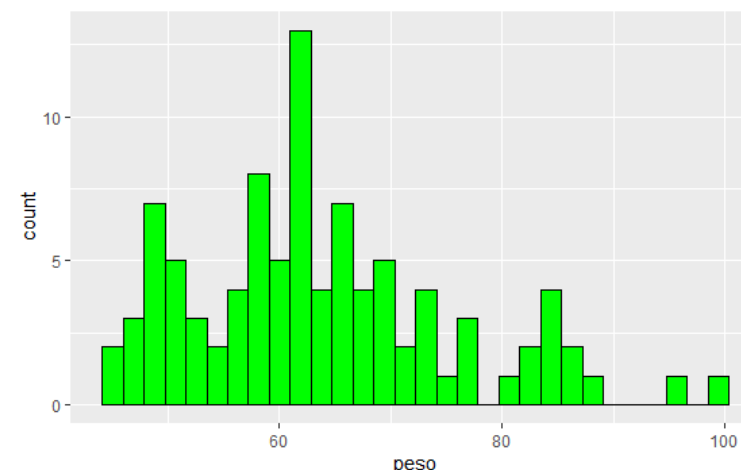
```
> ggplot(osteo, aes(x=peso)) +  
+ geom_histogram()
```



#### Modificación del color y forma

```
ggplot(data.frame, aes(x = variable)) +  
geom_histogram(col = "color", fill = "color")
```

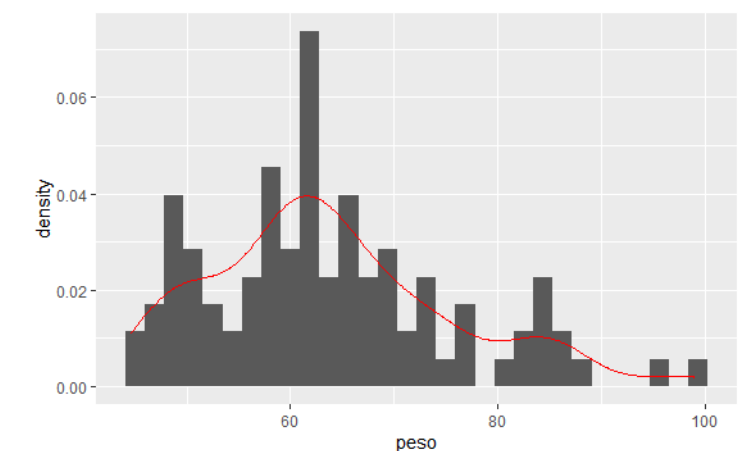
```
> ggplot(osteo, aes(x=peso)) +  
+ geom_histogram(  
+ col='black', fill='green')
```



#### Añadir la densidad

```
ggplot(data.frame, aes(x = variable))  
geom_histogram(aes(y=..density..)) +  
geom_density(col="color")
```

```
> ggplot(osteo, aes(x=peso)) +  
+ geom_histogram(aes(y=..density..)) +  
+ geom_density(col="red")
```



# Práctica 3: : Estadística Descriptiva en R



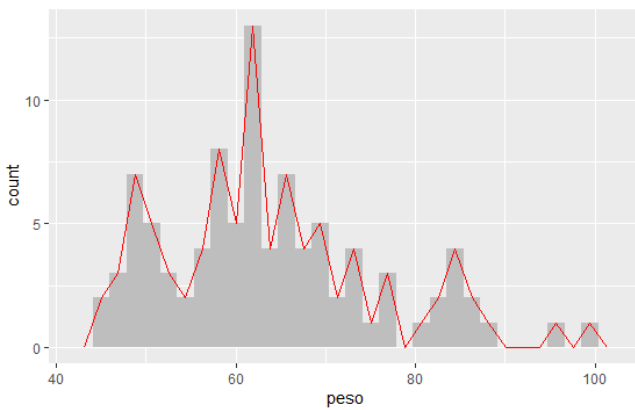
## Representaciones gráficas en la librería ggplot2

### Poligonal de frecuencias

#### Representación básica

```
ggplot(data.frame, aes(x = variable)) +  
geom_histogram() + geom_freqpoly ()
```

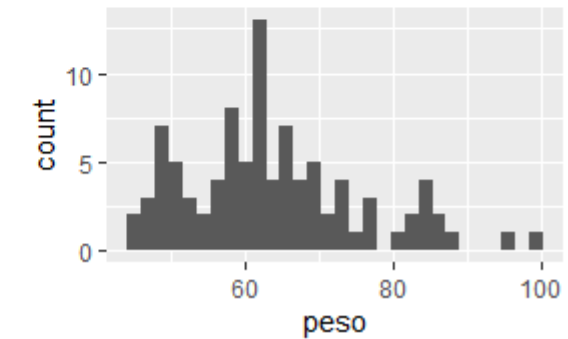
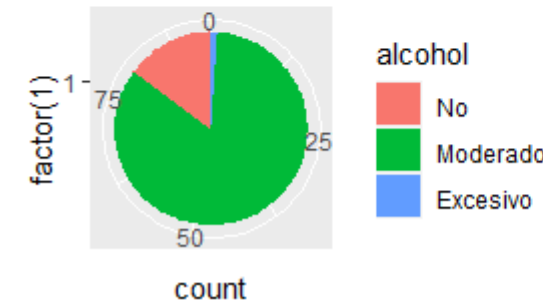
```
> ggplot(osteo, aes(x=peso)) +  
+ geom_histogram(fill="grey") +  
+ geom_freqpoly(col="red")
```



### Representaciones múltiples

Se hace uso de la función `grid.arrange()` de la librería `gridExtra`  
`grid.arrange(gráficos, nrow = número, ncol = número)`

```
> library(gridExtra)  
> histograma = ggplot(osteo, aes(x=peso)) +  
+ geom_histogram()  
> sectores = ggplot(osteo,  
+ aes(x=factor(1), fill=alcohol)) +  
+ geom_bar()+coord_polar("y")  
> barras = ggplot(osteo, aes(x = alcohol)) +  
+ geom_bar()  
> df=table(osteo$alcohol)  
> df=as.data.frame(df)  
> colnames(df)=c("Alcohol", "Frecuencia")  
> lineas= ggplot(df, aes(x=Alcohol,  
+ y=Frecuencia, group=1)) + geom_line()  
>  
> grid.arrange(barras, lineas, sectores,  
+ histograma, nrow = 2, ncol=2)
```

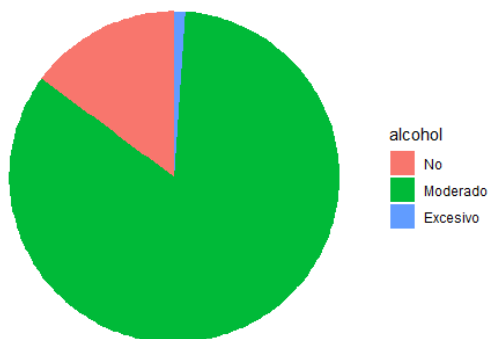


### Órdenes comunes

#### Quitar el fondo

```
ggplot(data.frame, aes(x = variable)) +  
geom_forma() + theme_void()
```

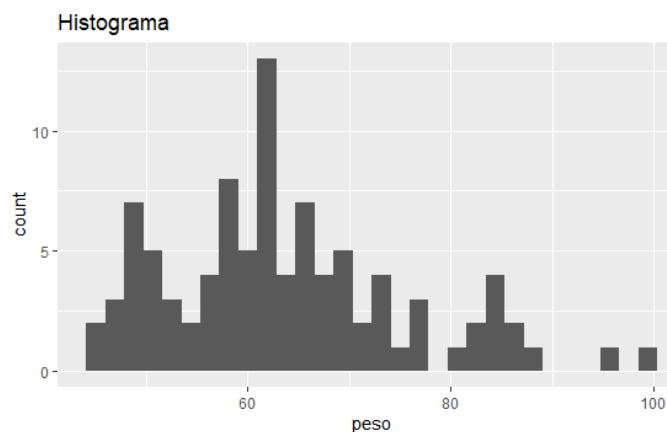
```
> ggplot(osteo, aes(x=factor(1),  
+ fill=alcohol)) +  
+ geom_bar()+coord_polar("y") +  
+ theme_void()
```



#### Poner título al gráfico

```
ggplot(data.frame, aes(x = variable)) +  
geom_forma() + ggtitle("título")
```

```
> ggplot(osteo, aes(x=peso)) +  
+ geom_histogram() +  
+ ggtitle("Histograma")
```



#### Cambiar nombres de los ejes

```
ggplot(data.frame, aes(x = variable)) +  
geom_forma() + xlab("Nombre") + ylab ("Nombre")
```

```
> ggplot(osteo, aes(x = alcohol)) +  
+ geom_bar() +  
+ xlab("Alcohol") +  
+ ylab("Número de casos")
```

