



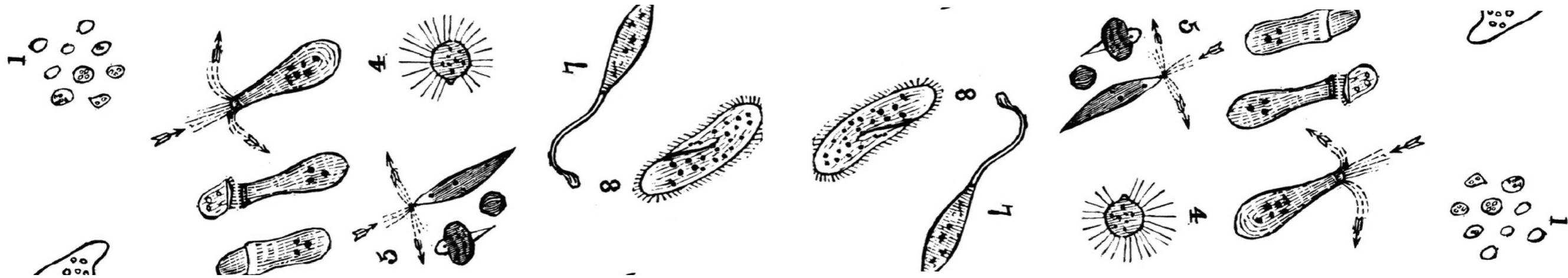
Diagramas de cajas (boxplots), de violín (Violin plots) y sinaplots

www.castrolab.org

www.cbib.cl

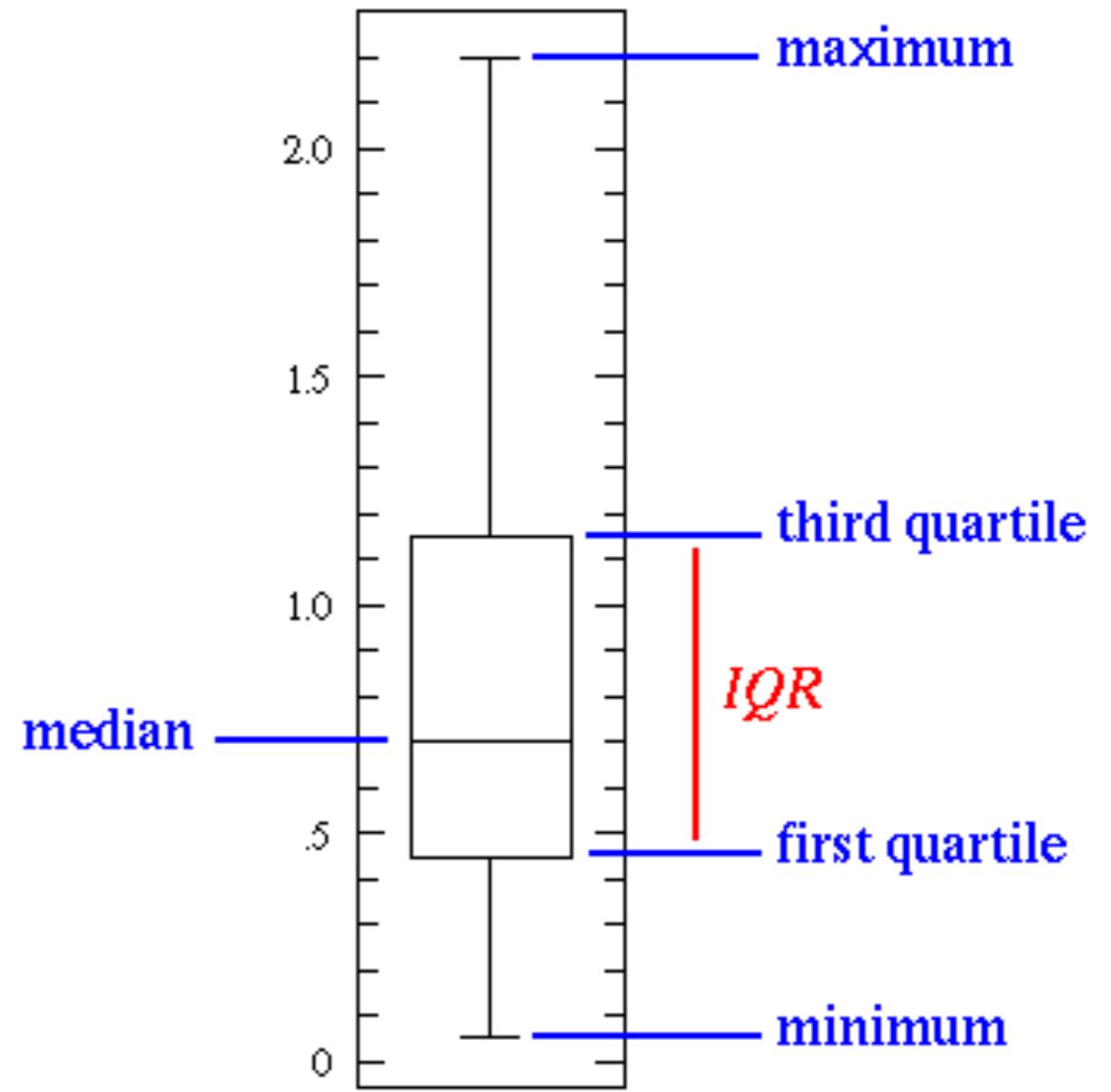
www.ucdavischile.org

Eduardo Castro, PhD
Universidad Andrés Bello
26 de marzo de 2018



Boxplots o diagramas de caja

- Otra forma de visualizar una distribución
- Percentiles, quintiles, cuartiles, etc.



¿Qué es un percentil?

- El valor por debajo del cual un porcentaje de observaciones cae
- Percentil 20, es el valor donde el 20% de las observaciones están ubicadas

Table 1. Test Scores.

Number	Rank
3	1
5	2
7	3
8	4
9	5
11	6
13	7
15	8

¿Qué es un percentil?

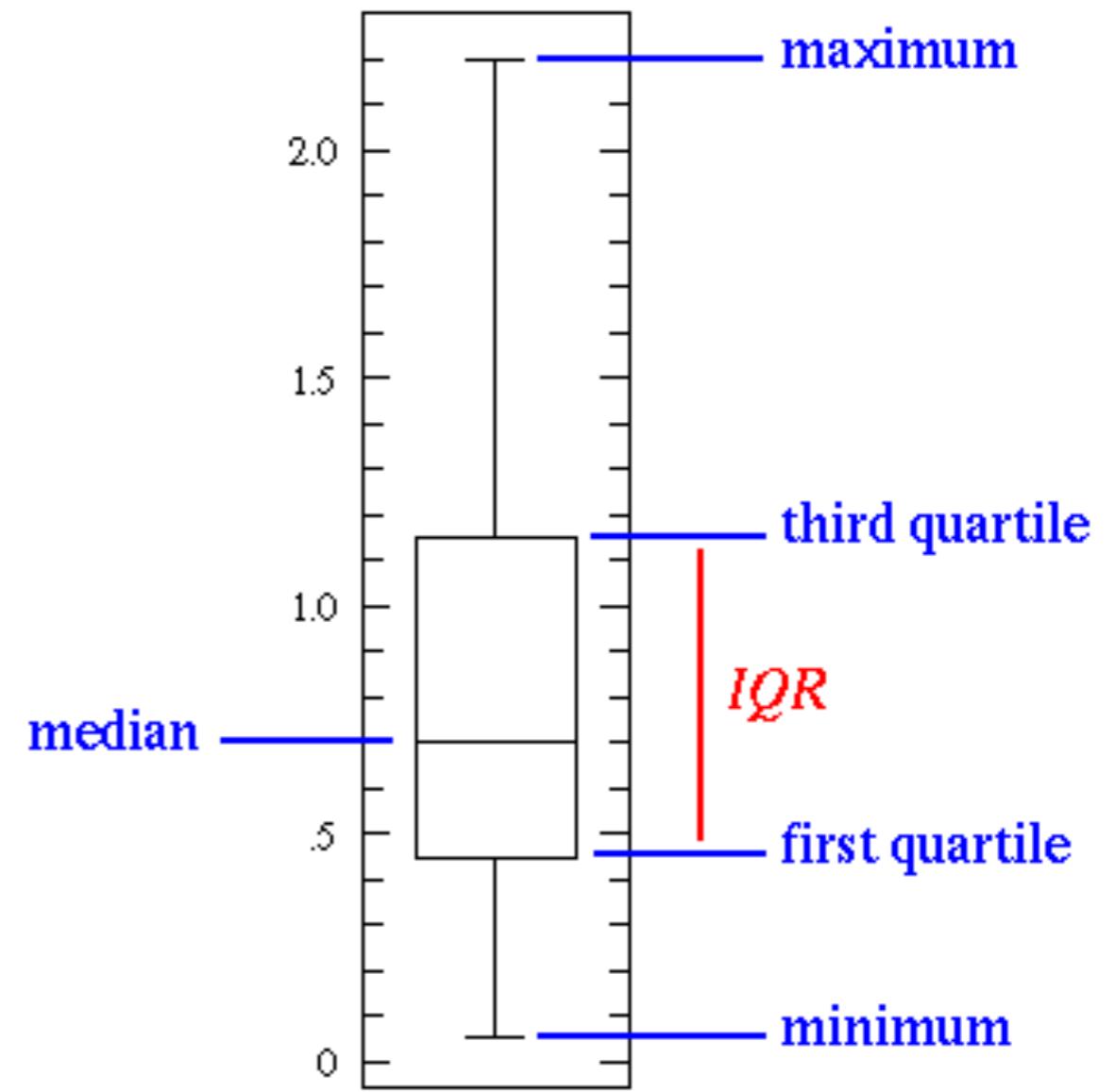
- Ejemplo con percentil 25
- $R = P/100 \times (N + 1)$
- $R = 25/100 \times (8 + 1) = 9/4 = 2.25$
- Entre 5 y 7
- $0.25 * (7 - 5) + 5 = 5.5$

Table 1. Test Scores.

Number	Rank
3	1
5	2
7	3
8	4
9	5
11	6
13	7
15	8

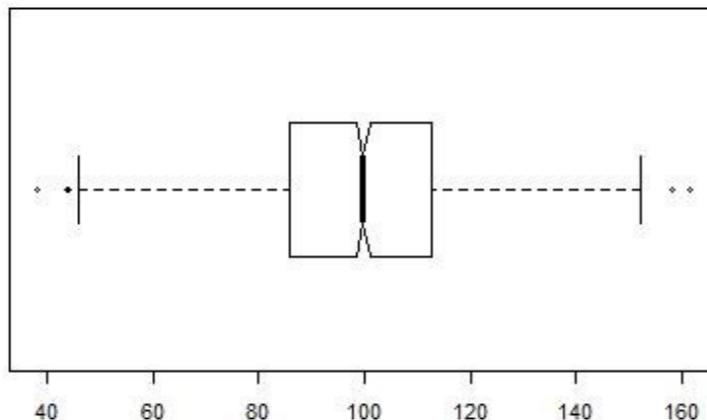
Boxplots o diagramas de caja

- Cuartil = 25avo de la distribución
- Quintil = un quinto de la distribución

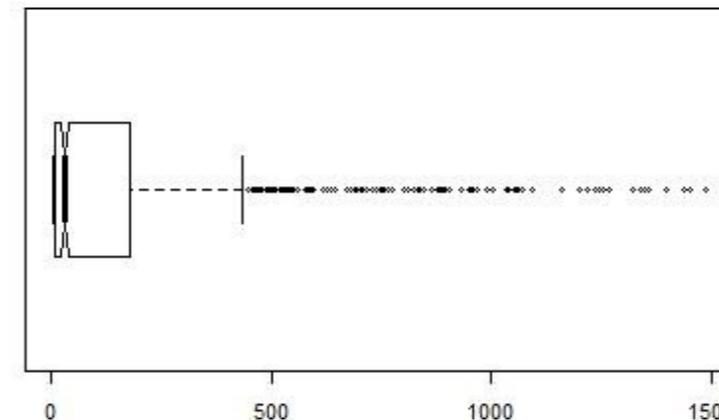


Boxplots - parientes del histograma y del gráfico de densidad

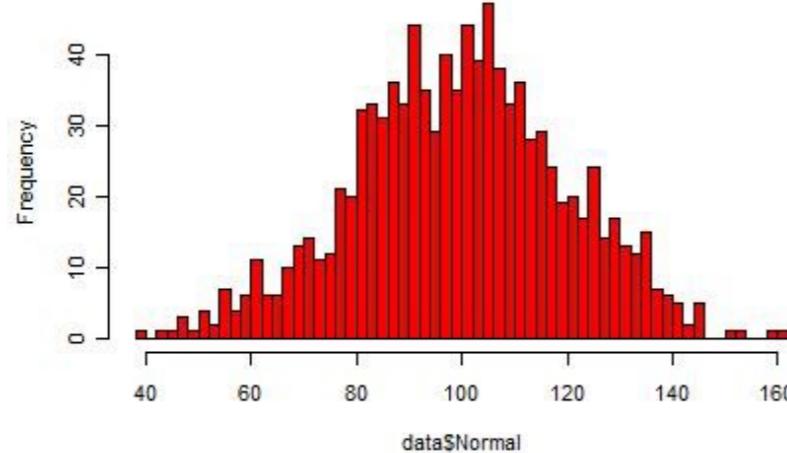
Notched Box Plot Normal Data



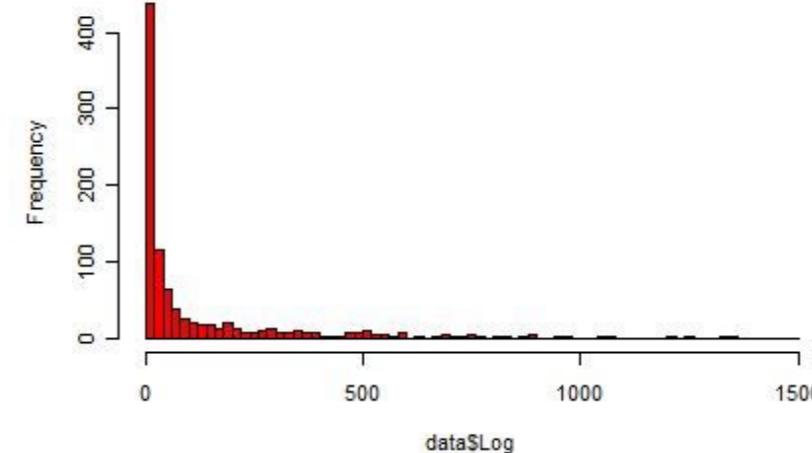
Notched Box Plot of Skewed Data



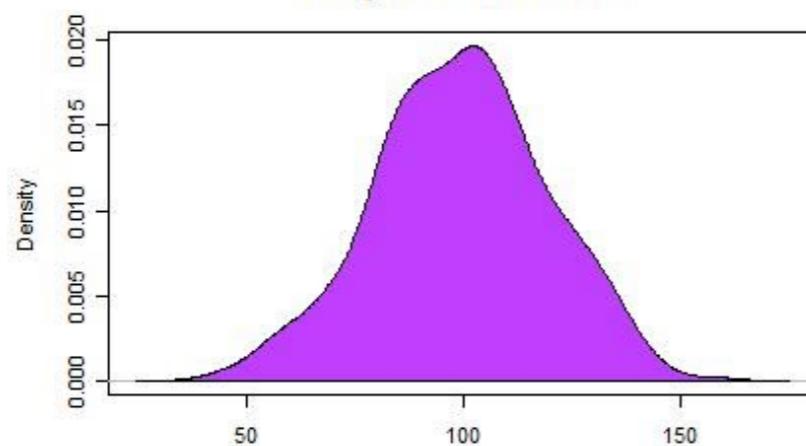
Histogram of Normal Data



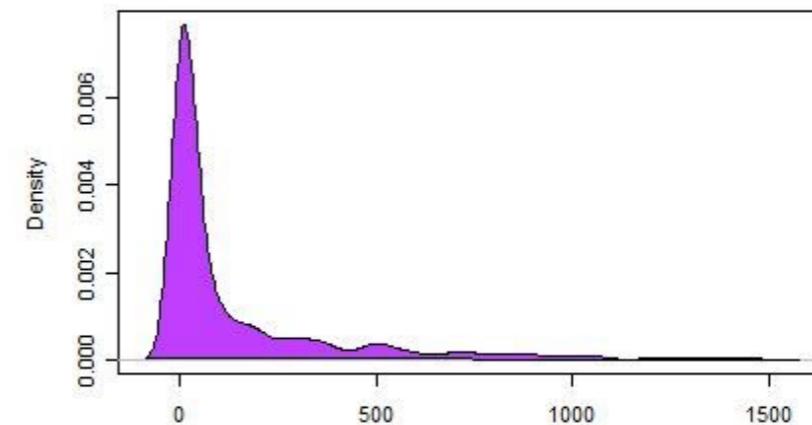
Histogram of Skewed Data



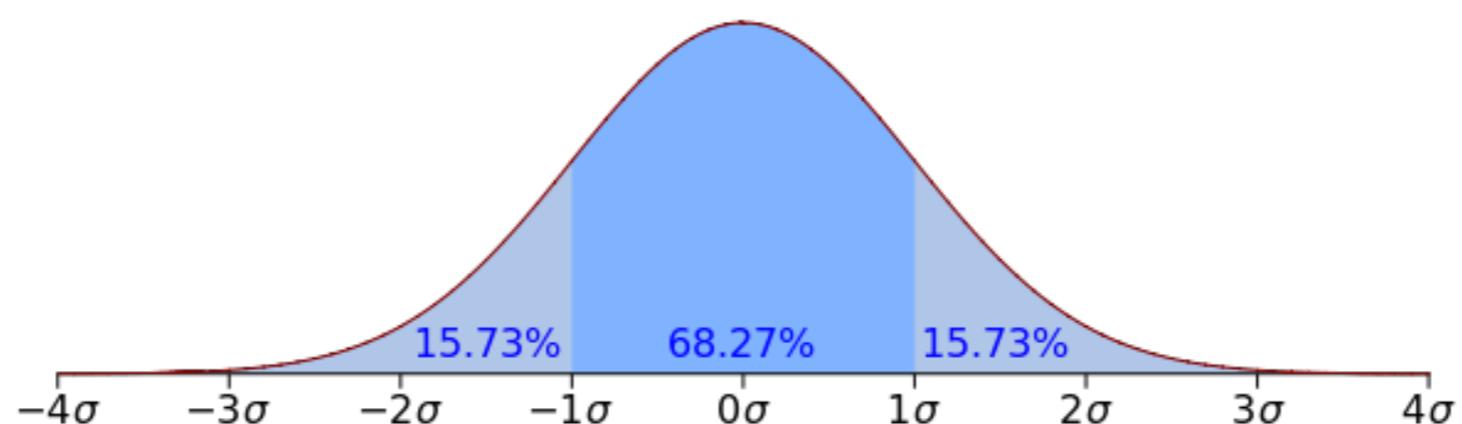
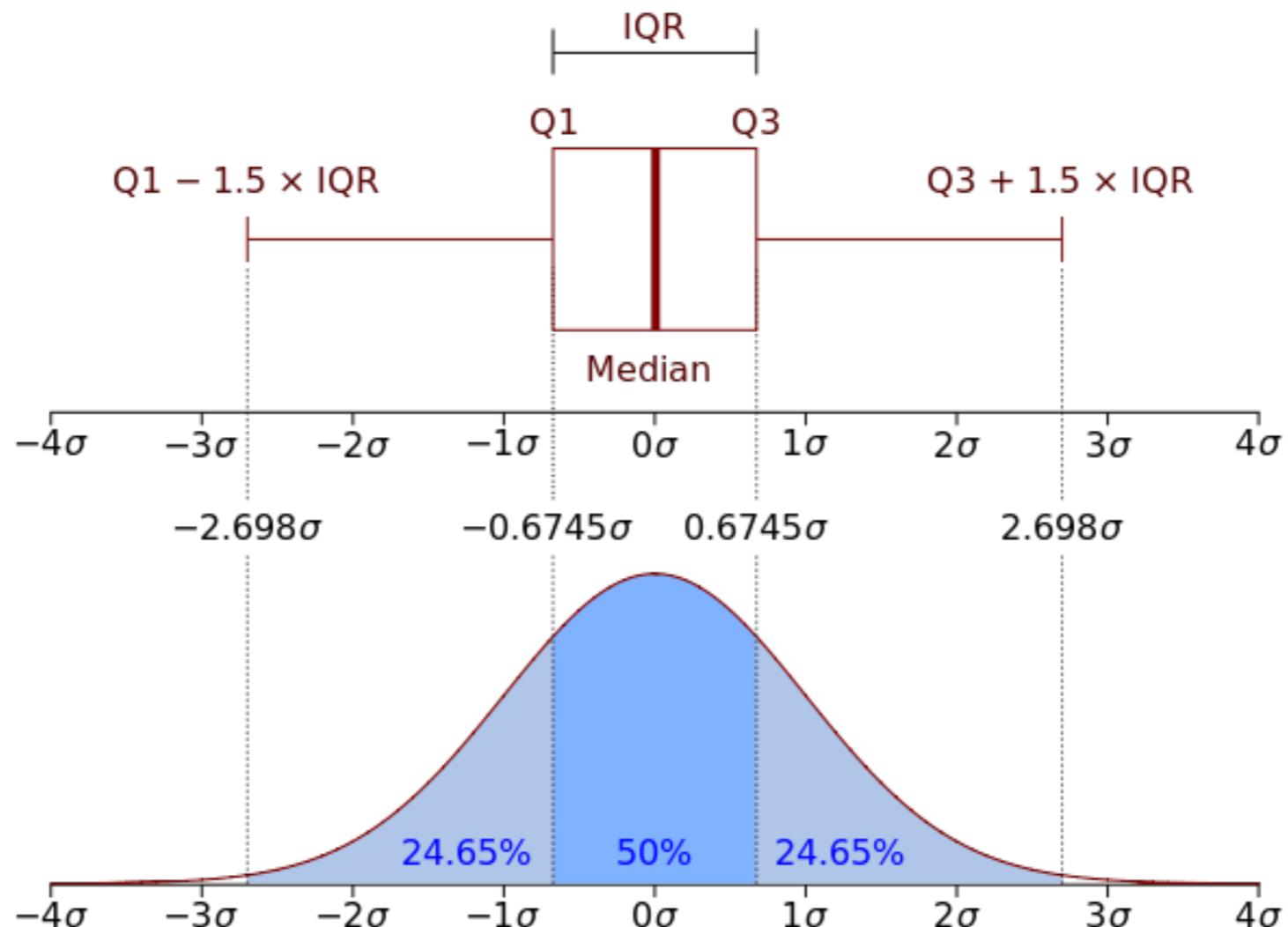
Density Plot of Normal Data



Density Plot of Skewed Data



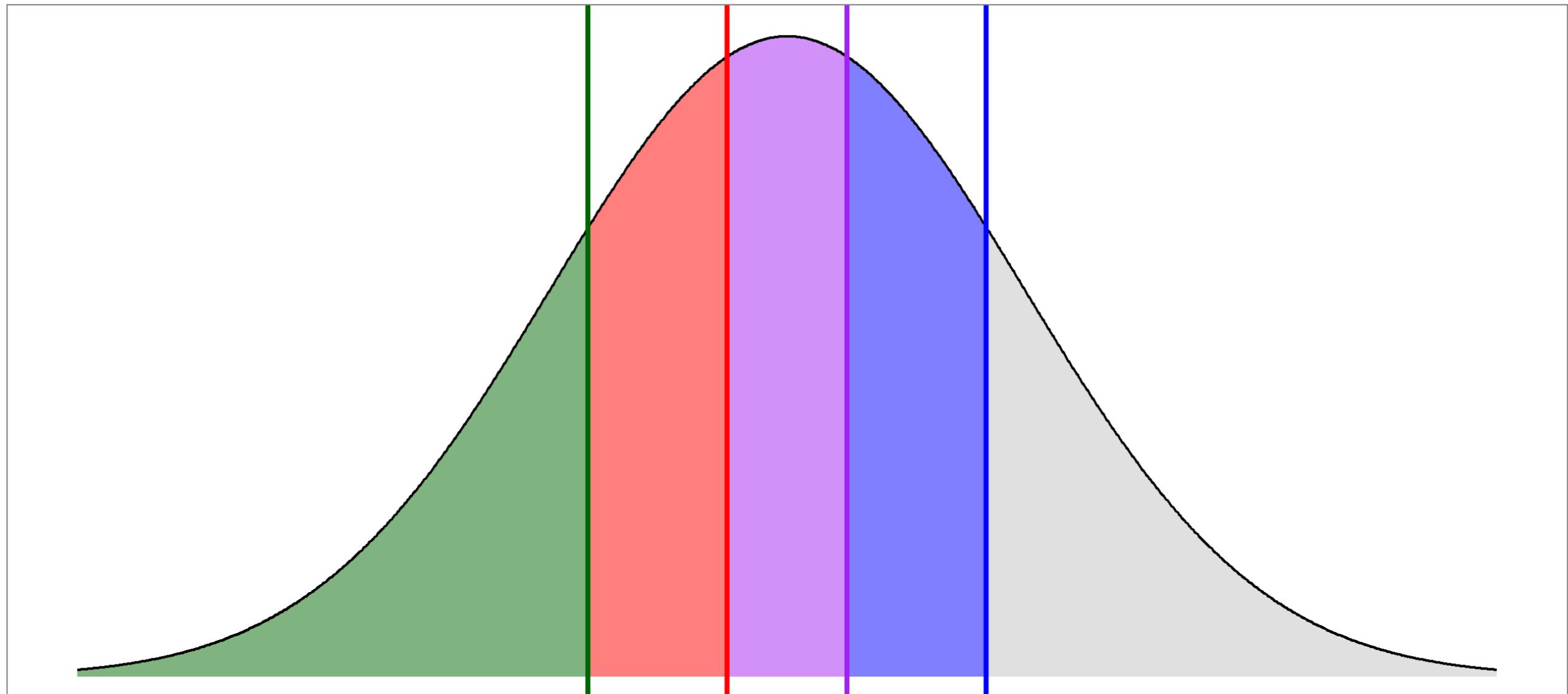
Boxplots



Gratuidad

- Los requisitos para poder acceder a este beneficio es ser estudiante que ingresa a primer año de Educación Superior y pertenecer al primer, segundo o tercer **quintil** de ingreso per cápita

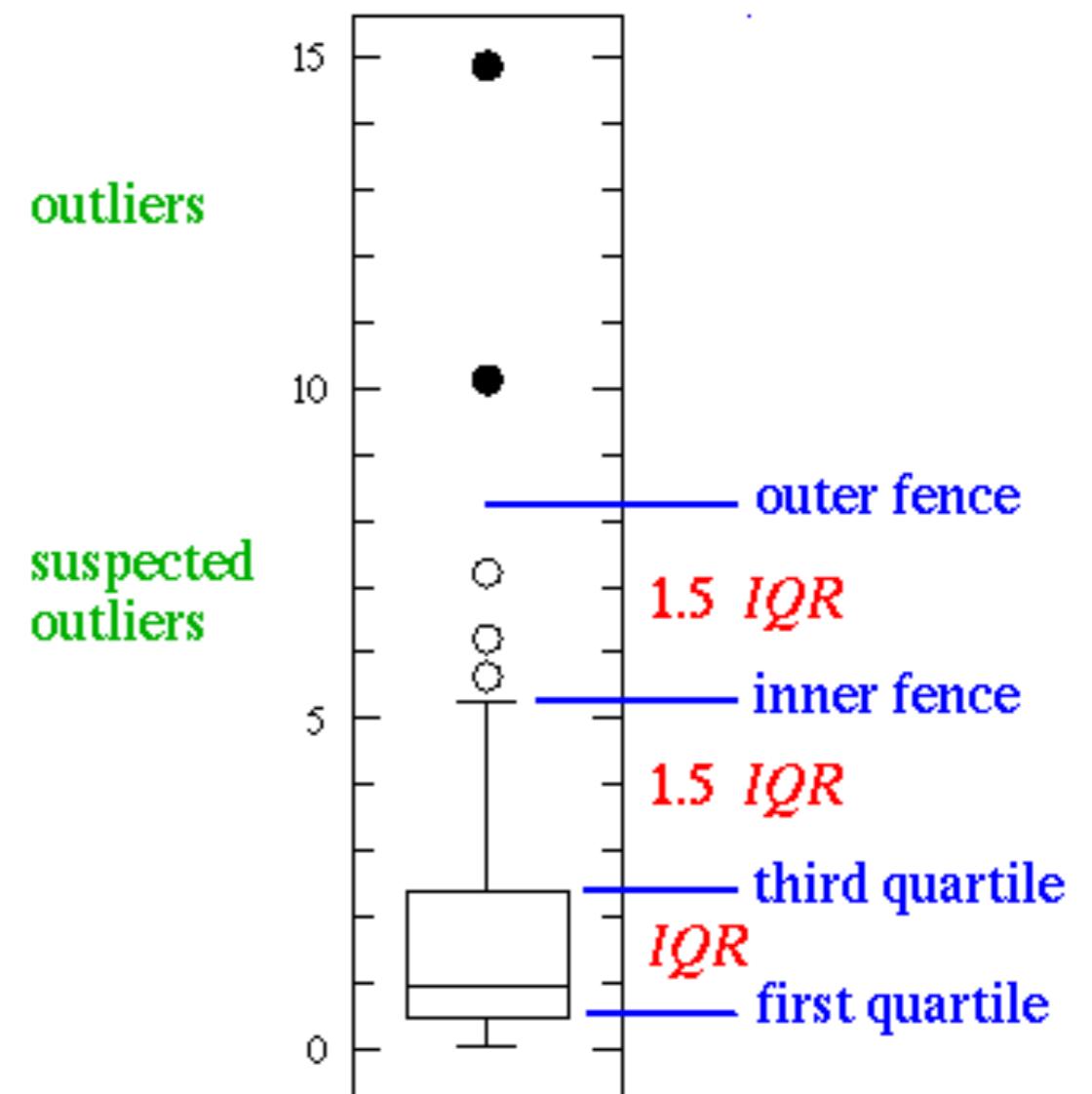
¿Qué es un quintil?



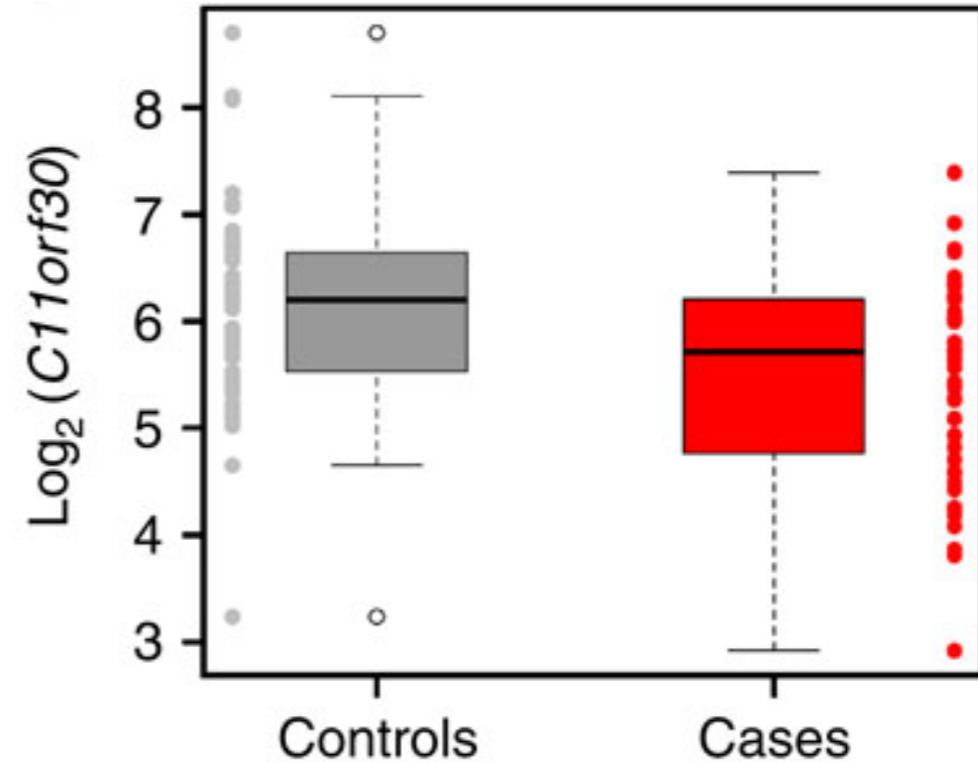
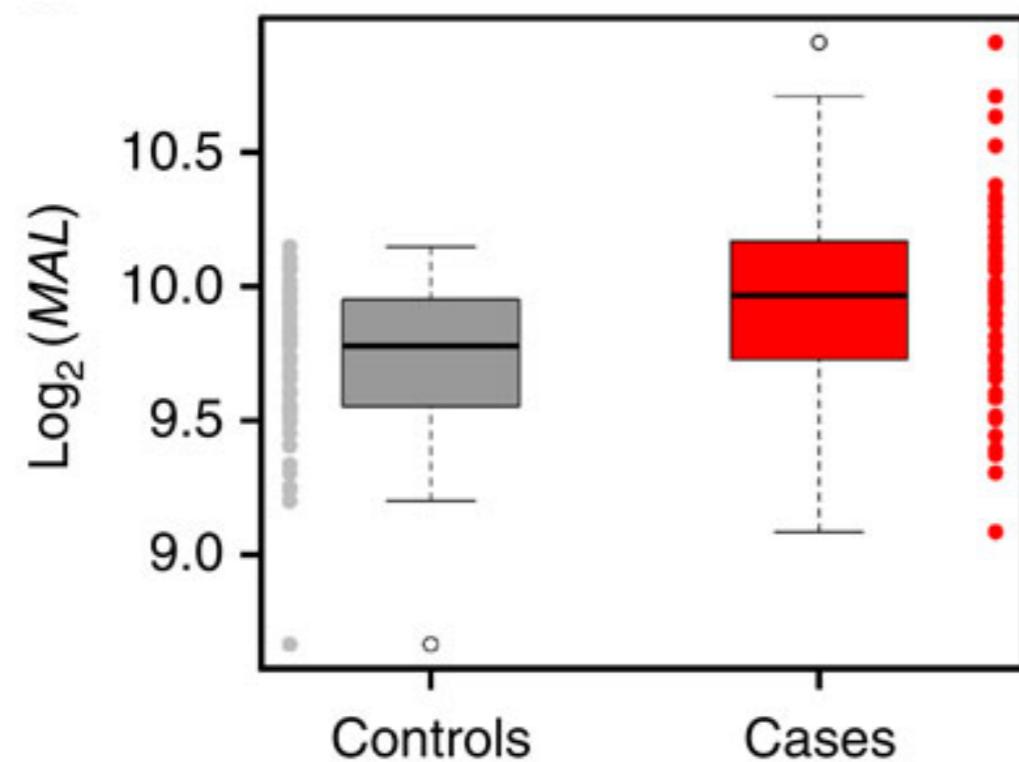
Quintile Bottom Second Middle Fourth Top

Utilidad de los boxplots

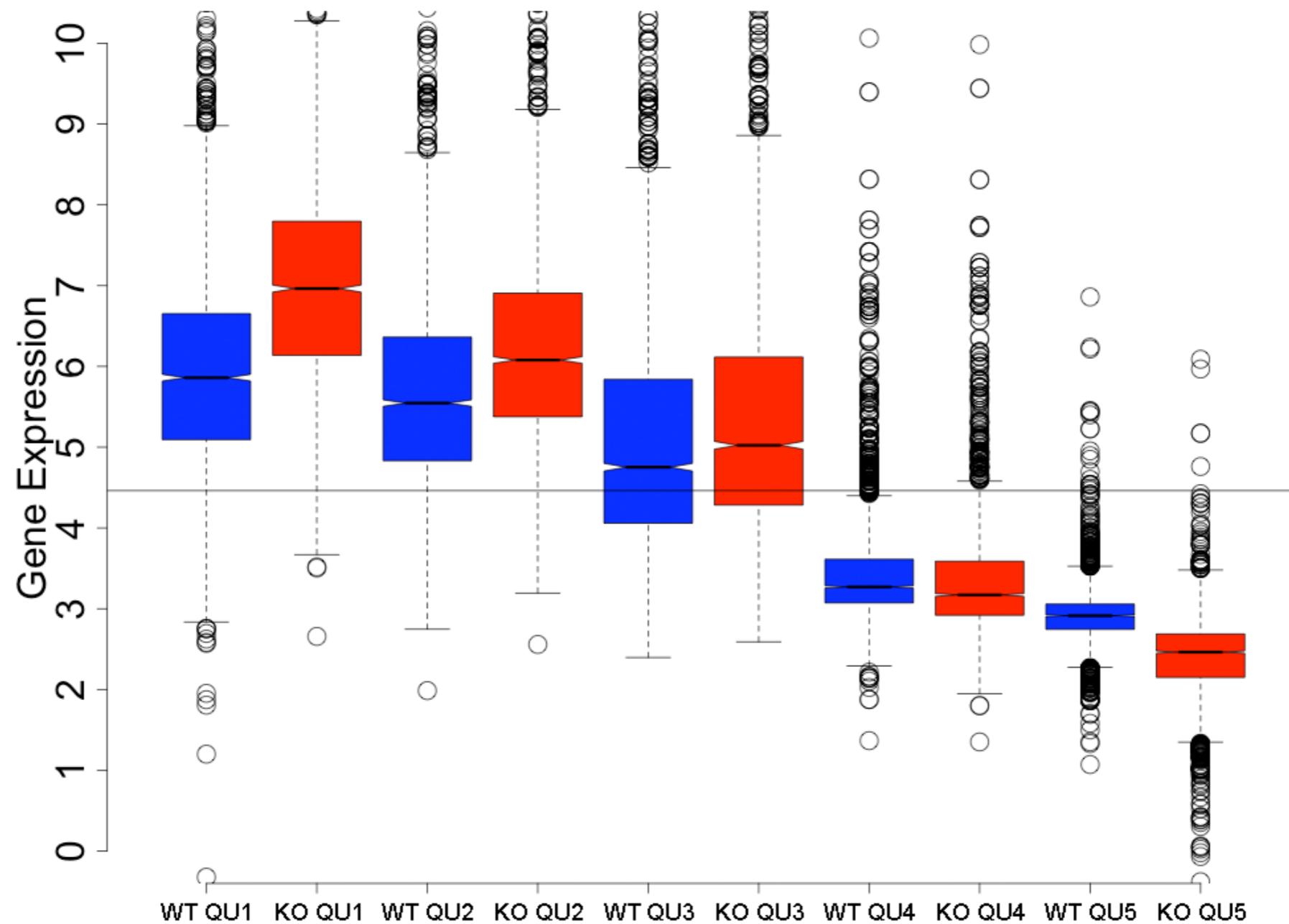
- Valores atípicos probables - 1.5 IQR
- Valores atípicos - 3 IQR



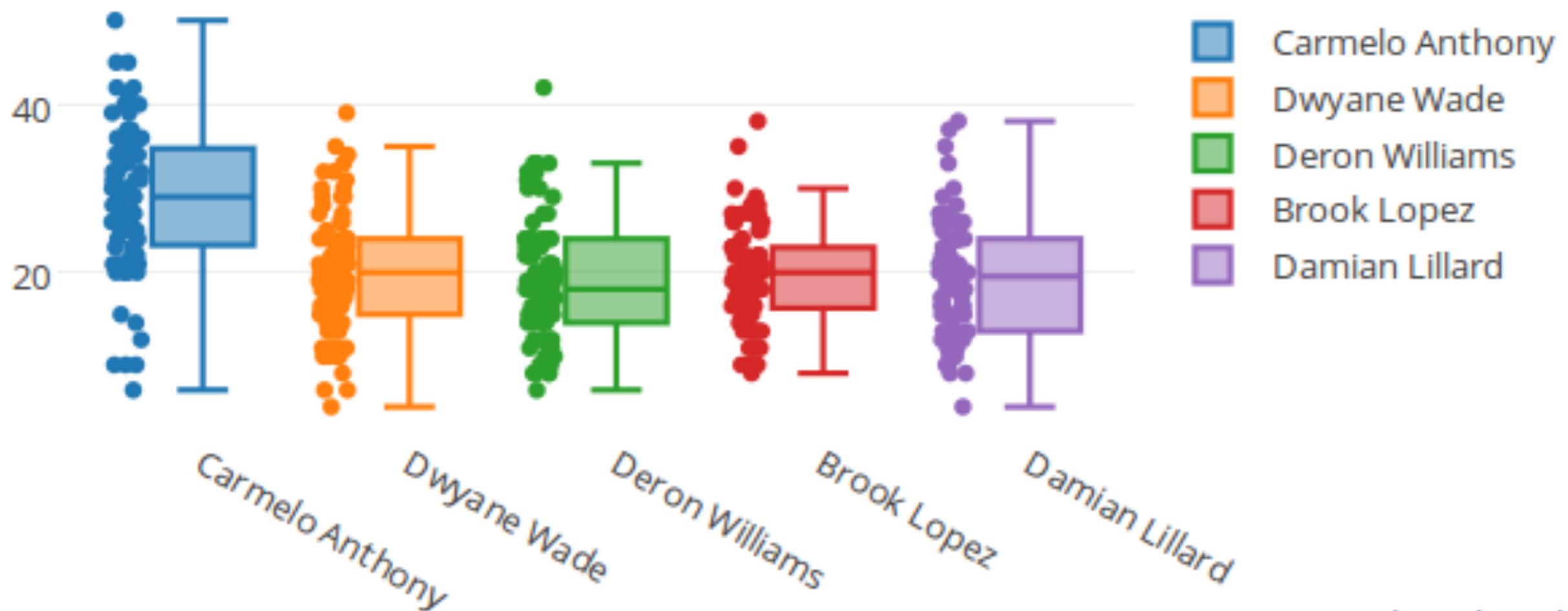
Boxplots sirven para comparar distribuciones y detectar valores atípicos



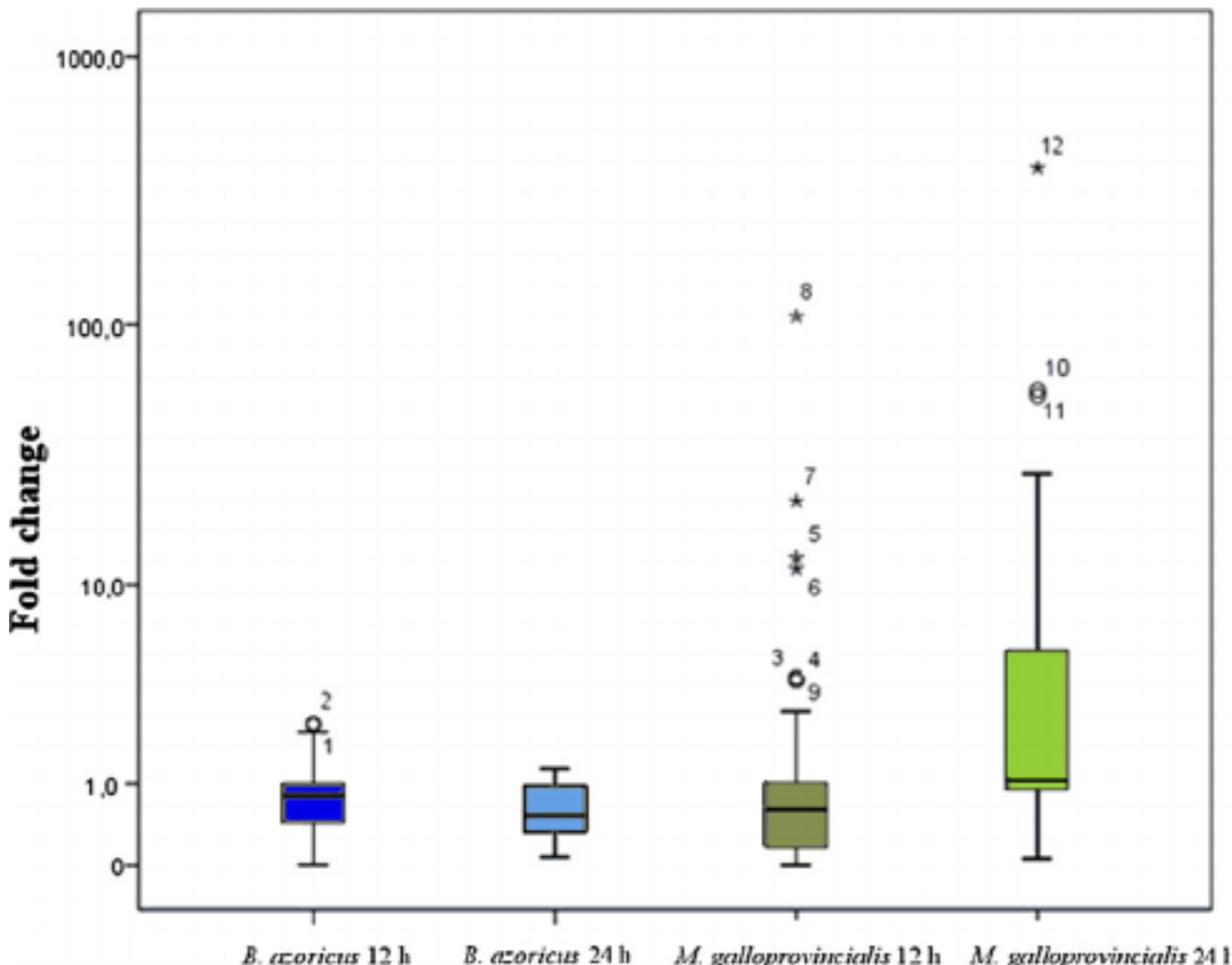
Boxplots sirven para comparar distribuciones y detectar valores atípicos



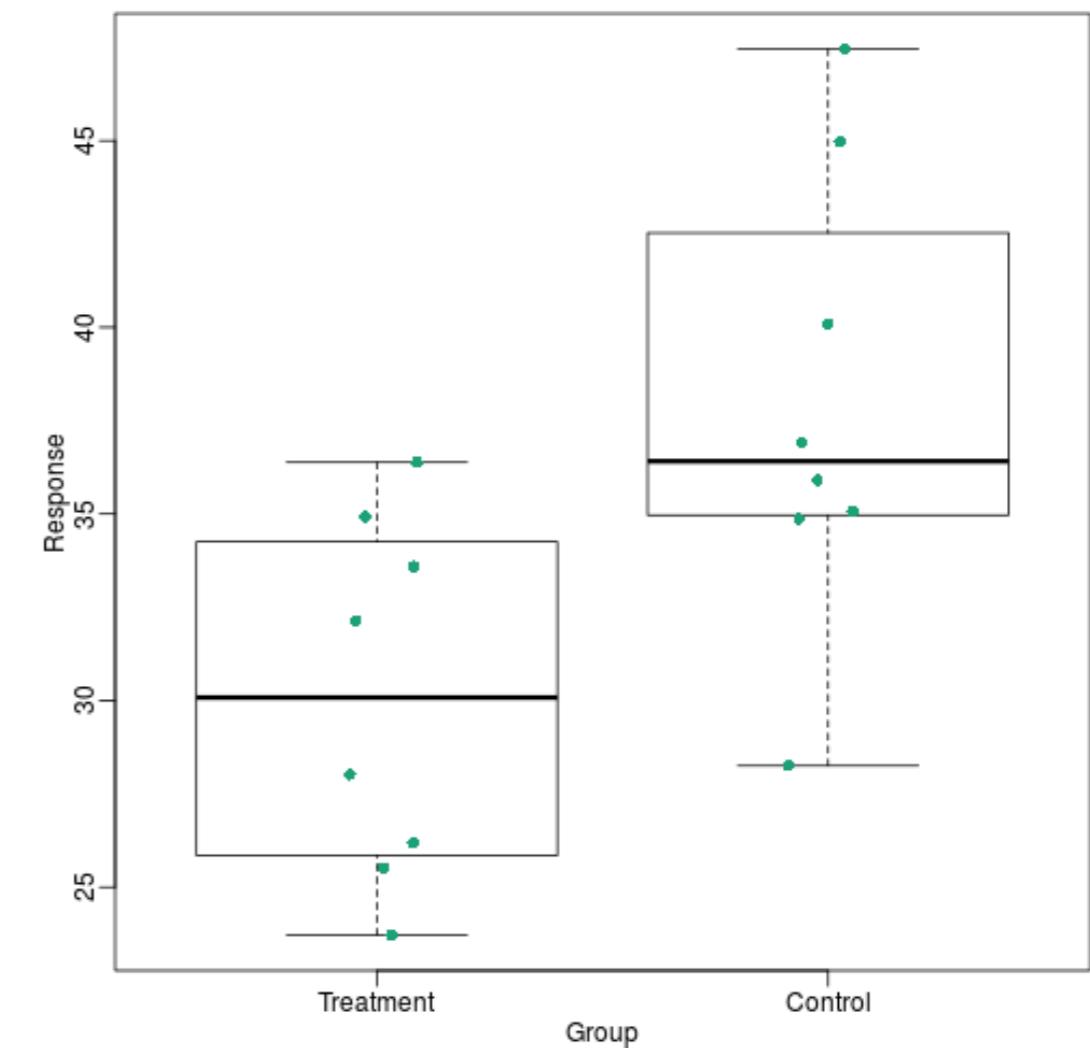
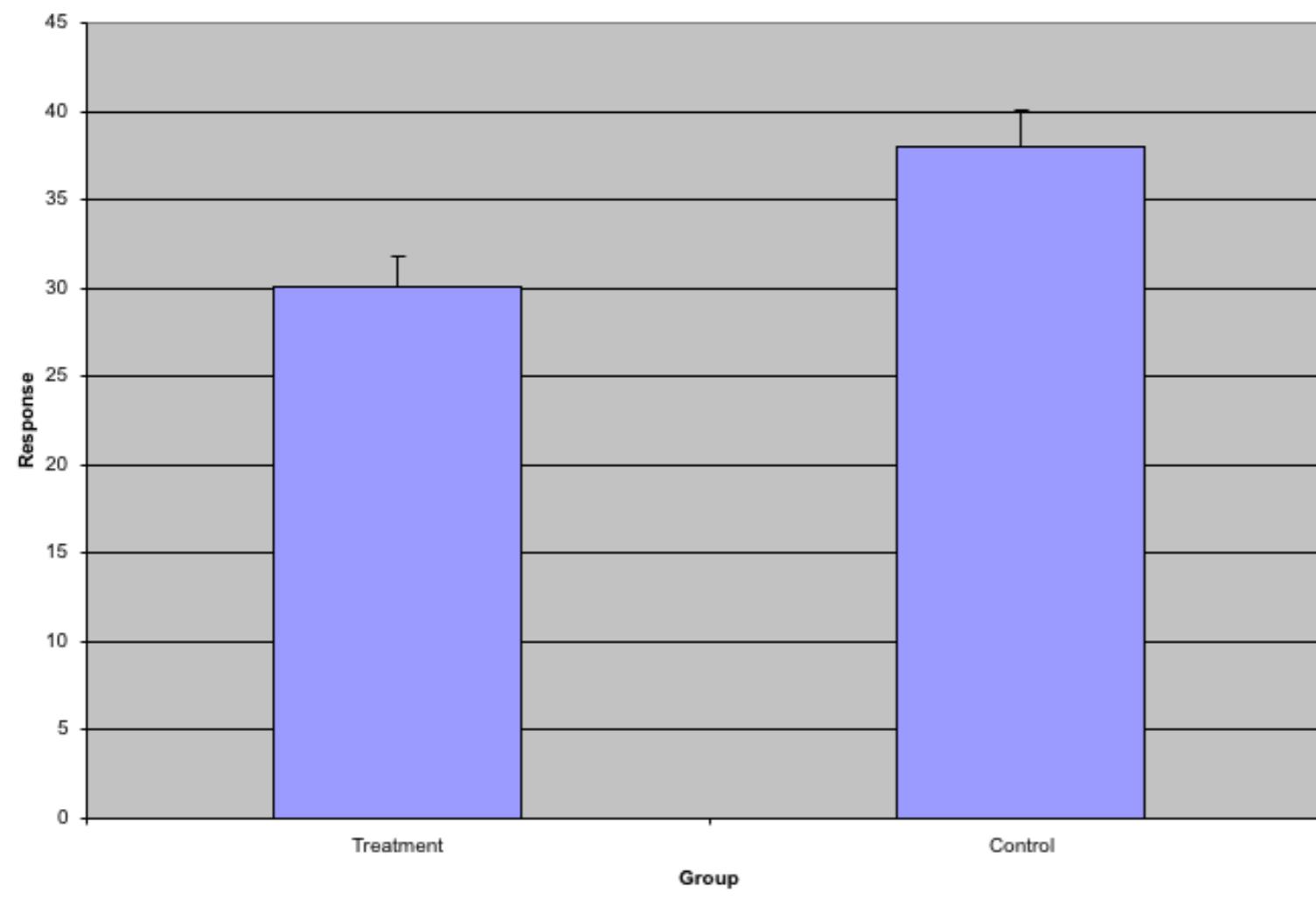
Buenas prácticas - mostrar los puntos del boxplot



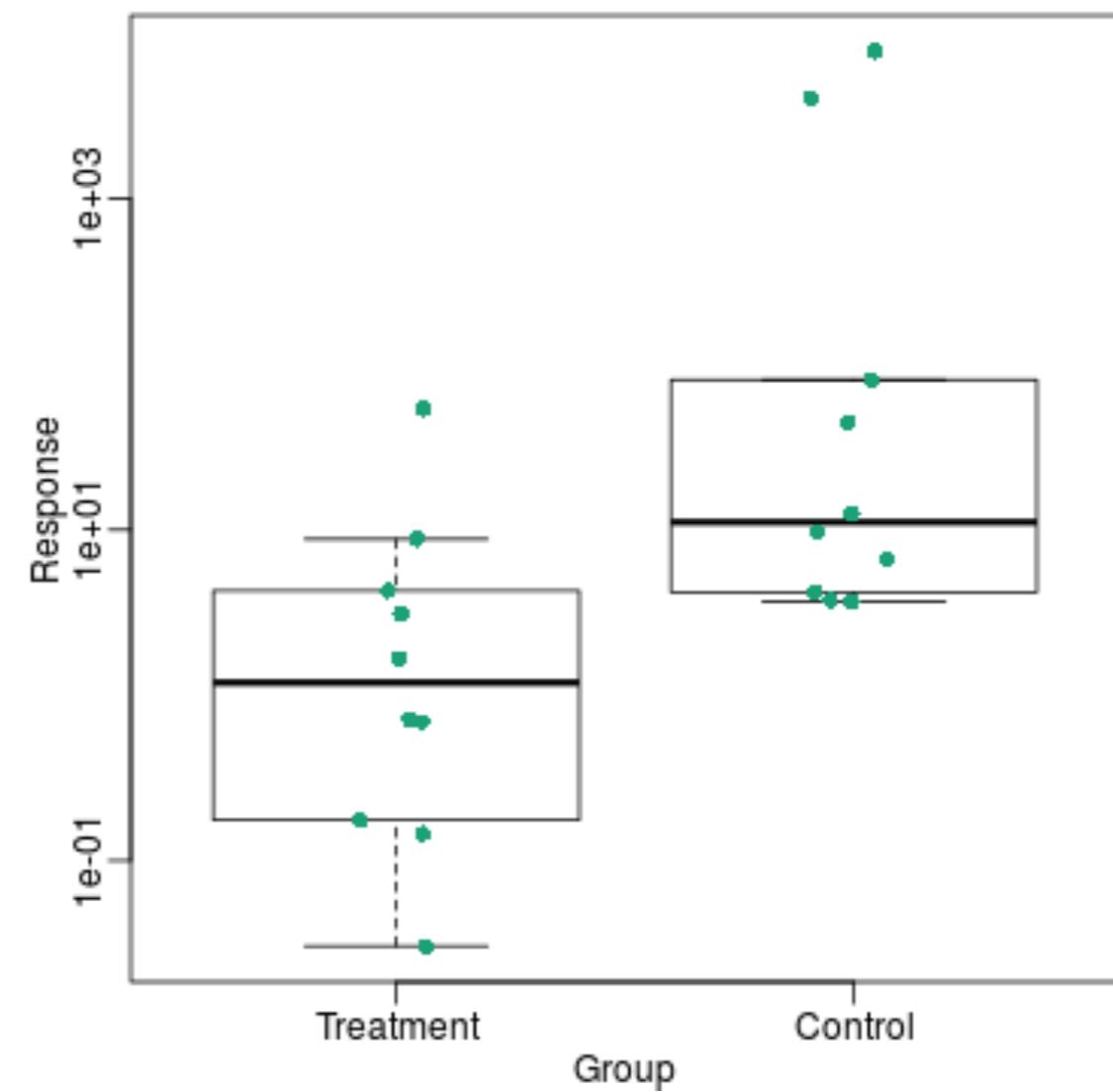
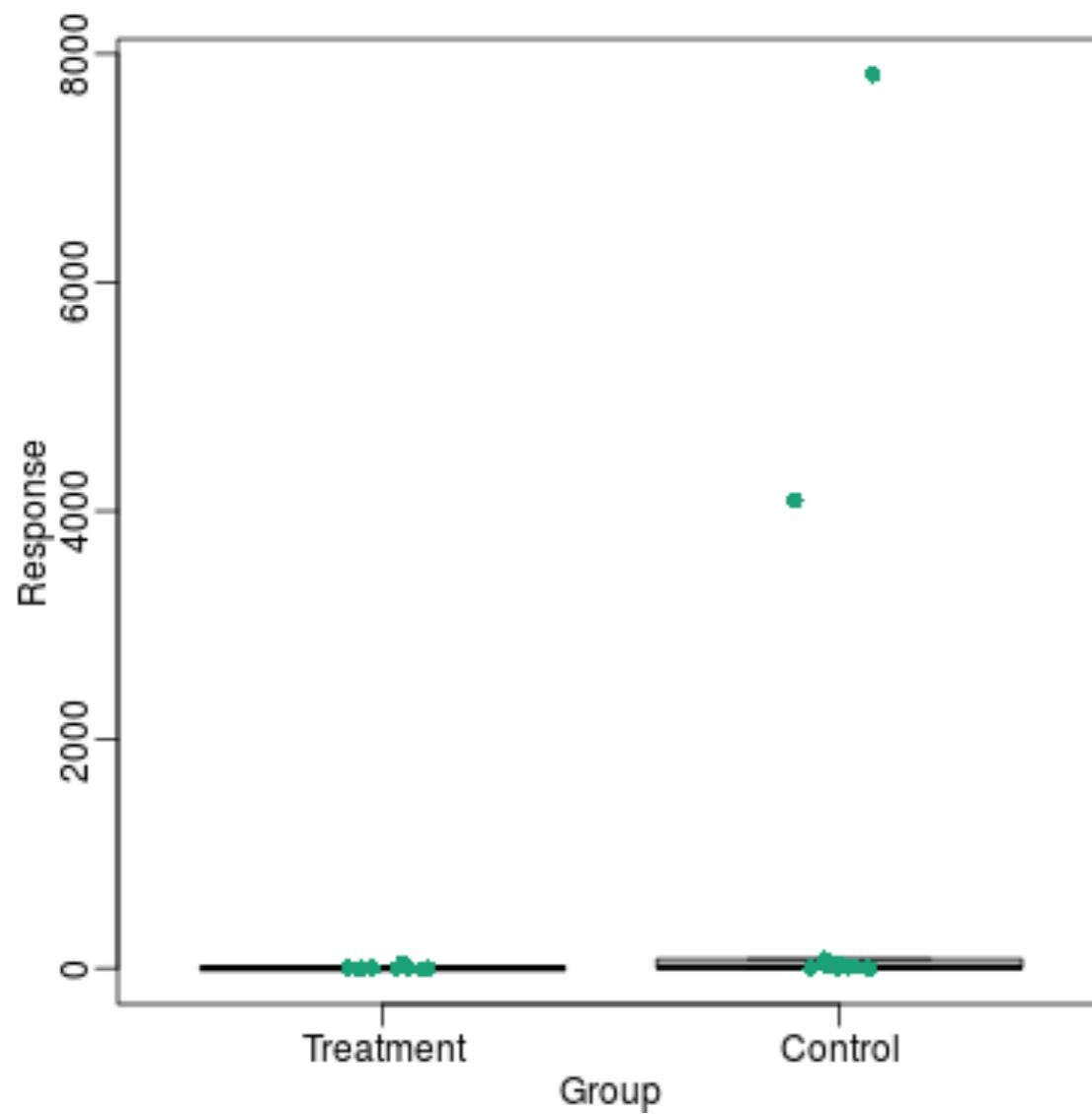
Buenas prácticas - identificar outliers



Malas prácticas - ocultar la distribución de datos con un gráfico de barras

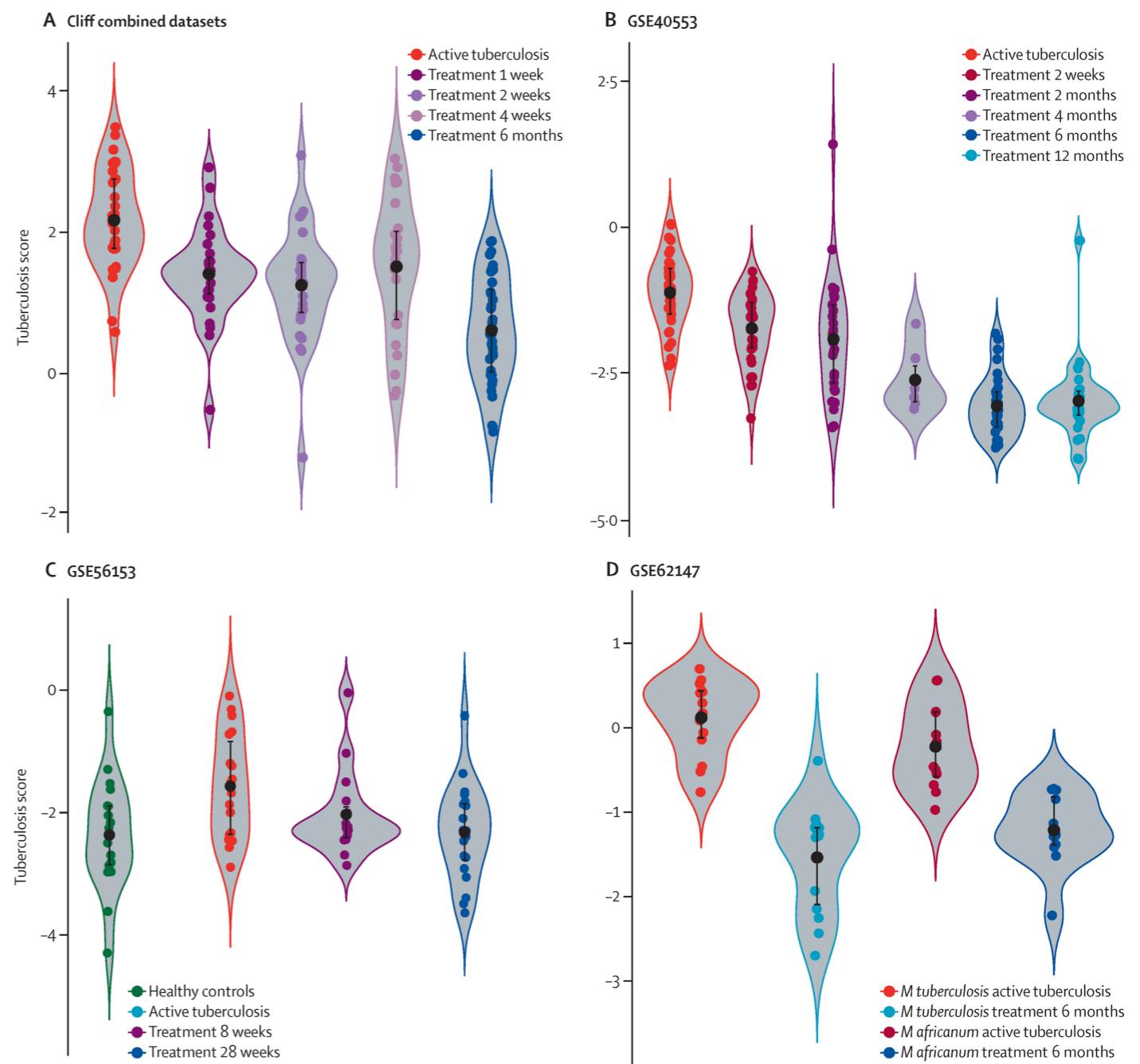


Malas prácticas - ocultar la distribución de datos al usar una escala inadecuada



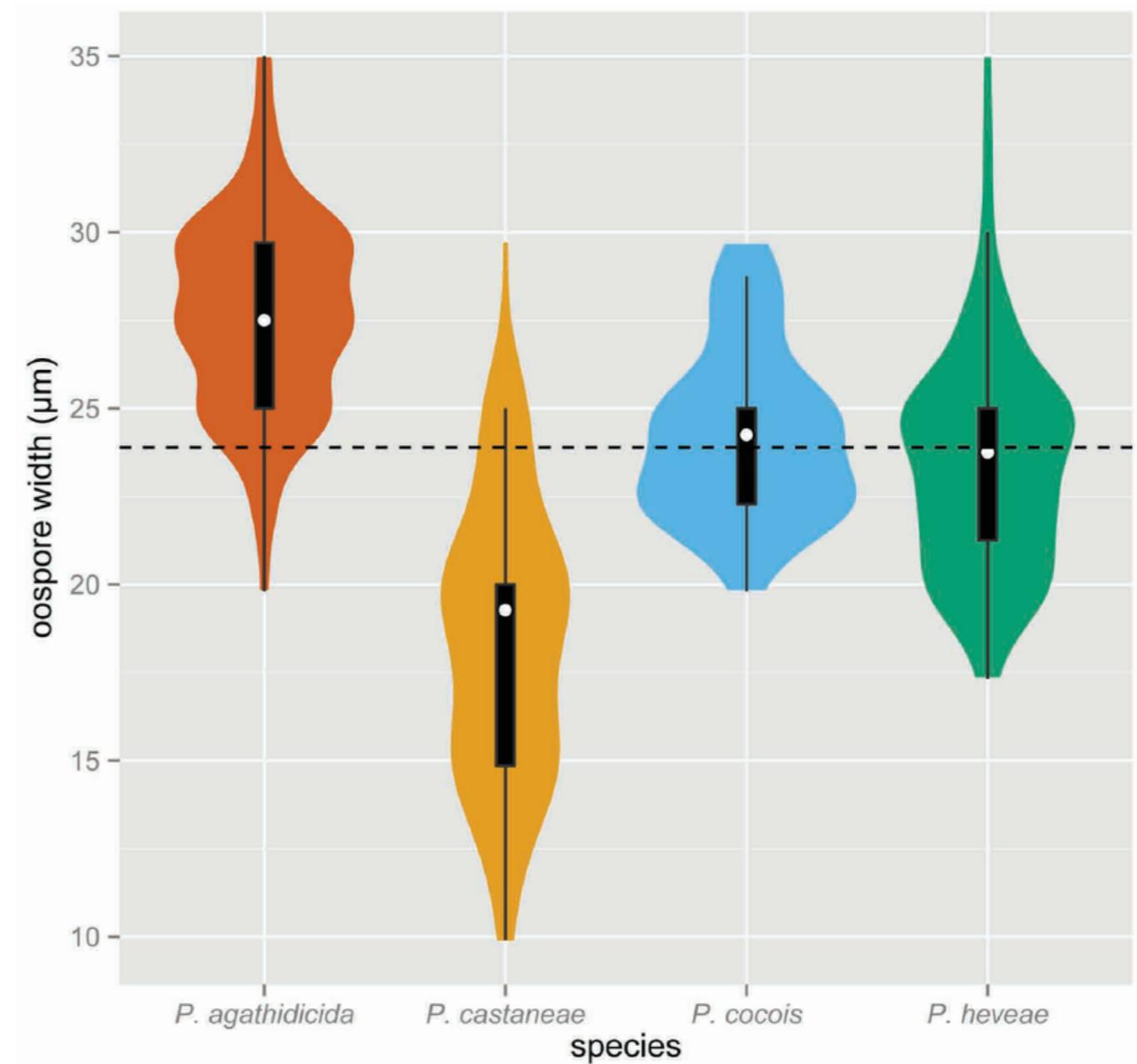
Violin plot

- Análogo a histogramas y gráficos de densidad
- Es un boxplot con un gráfico de densidad a cada lado



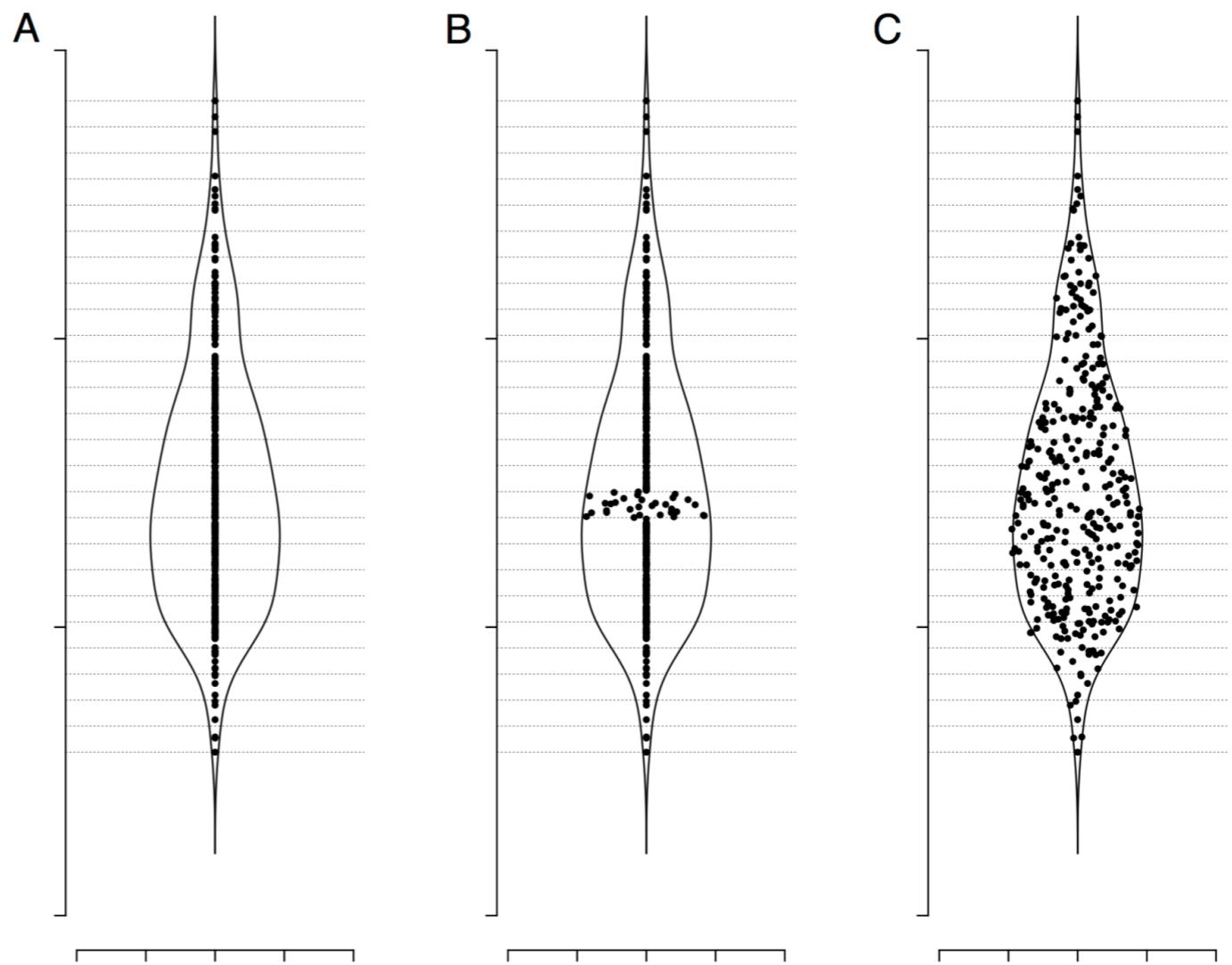
Violin plot

- También puede mostrar el IQR y la mediana

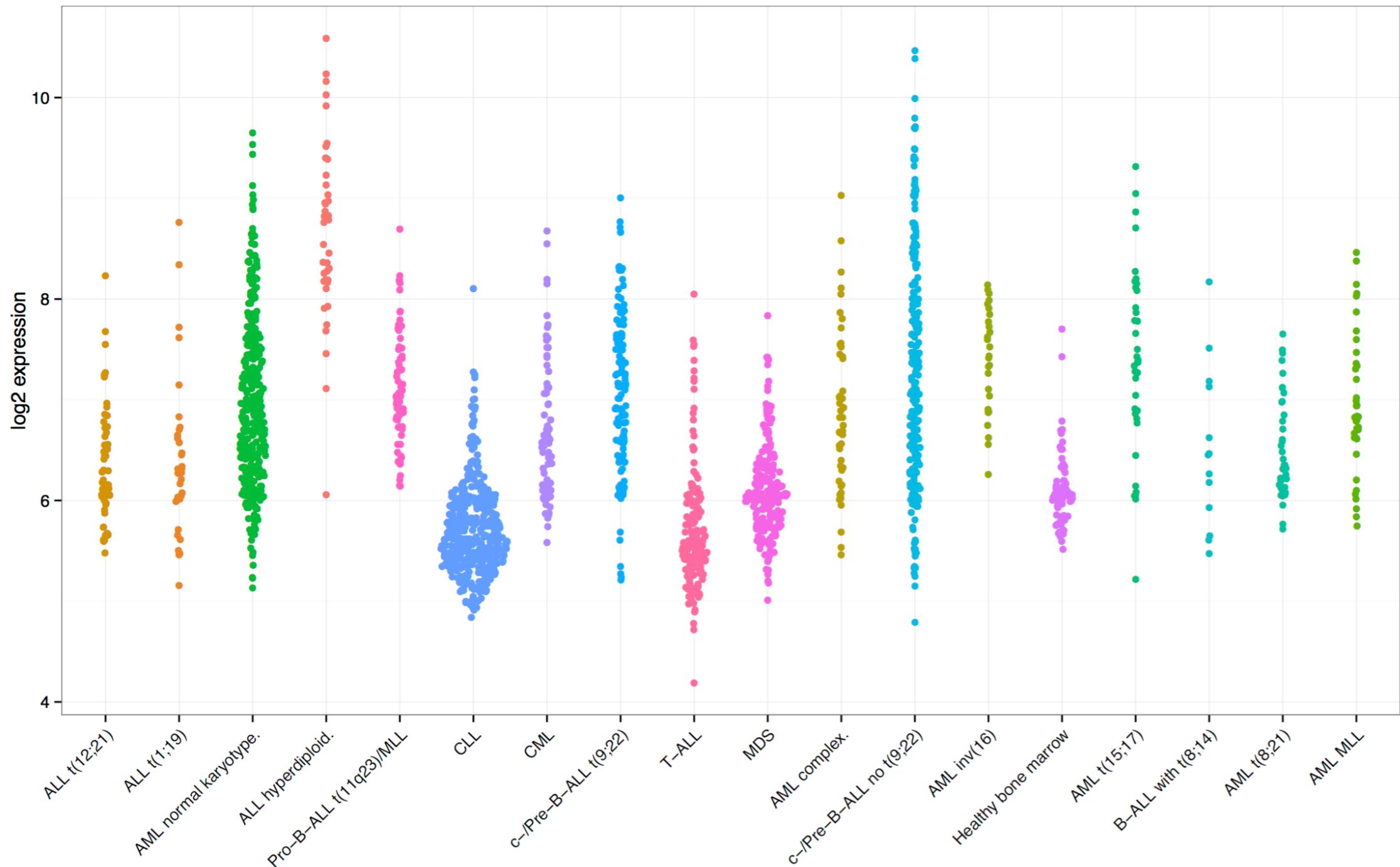


Sina plot

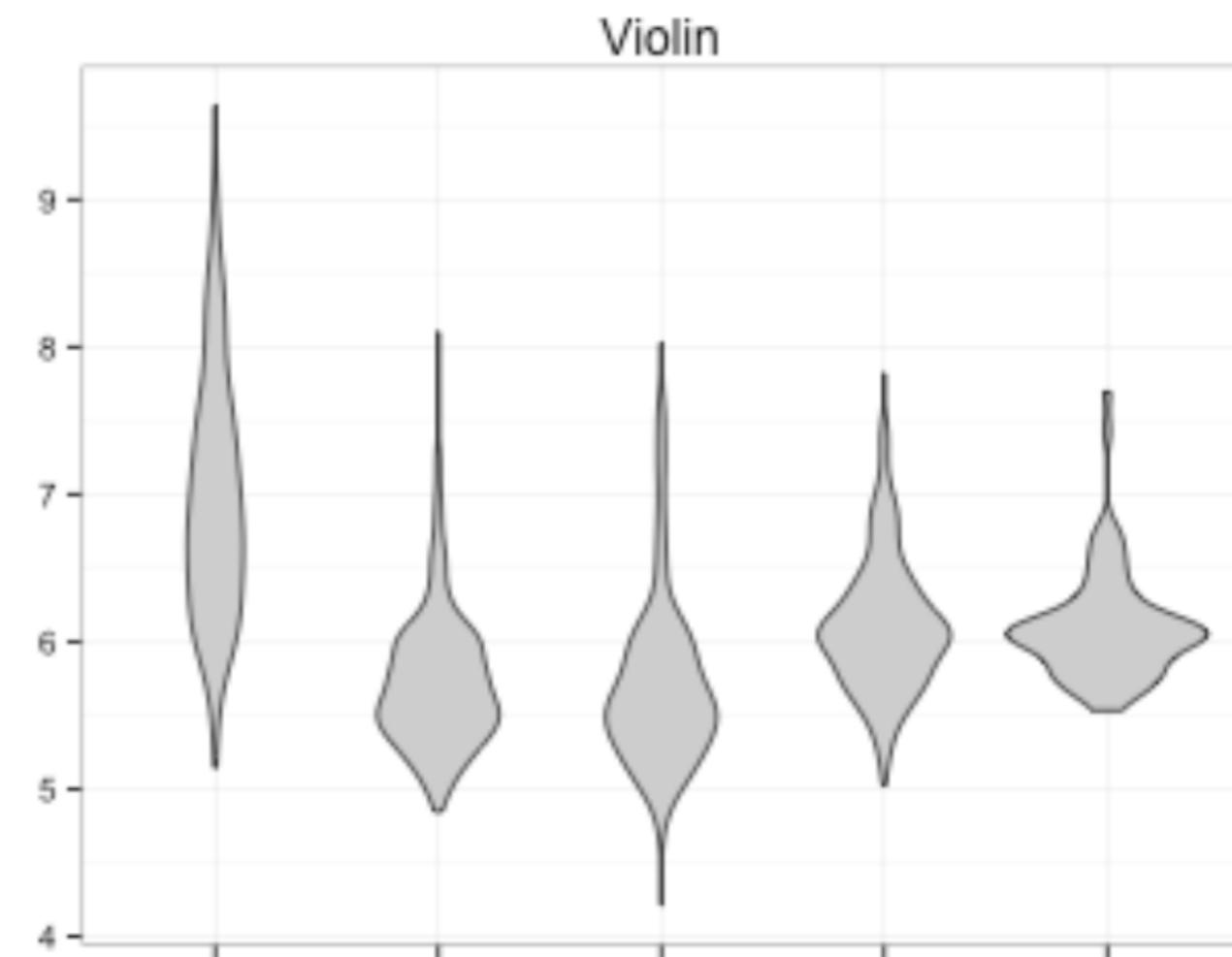
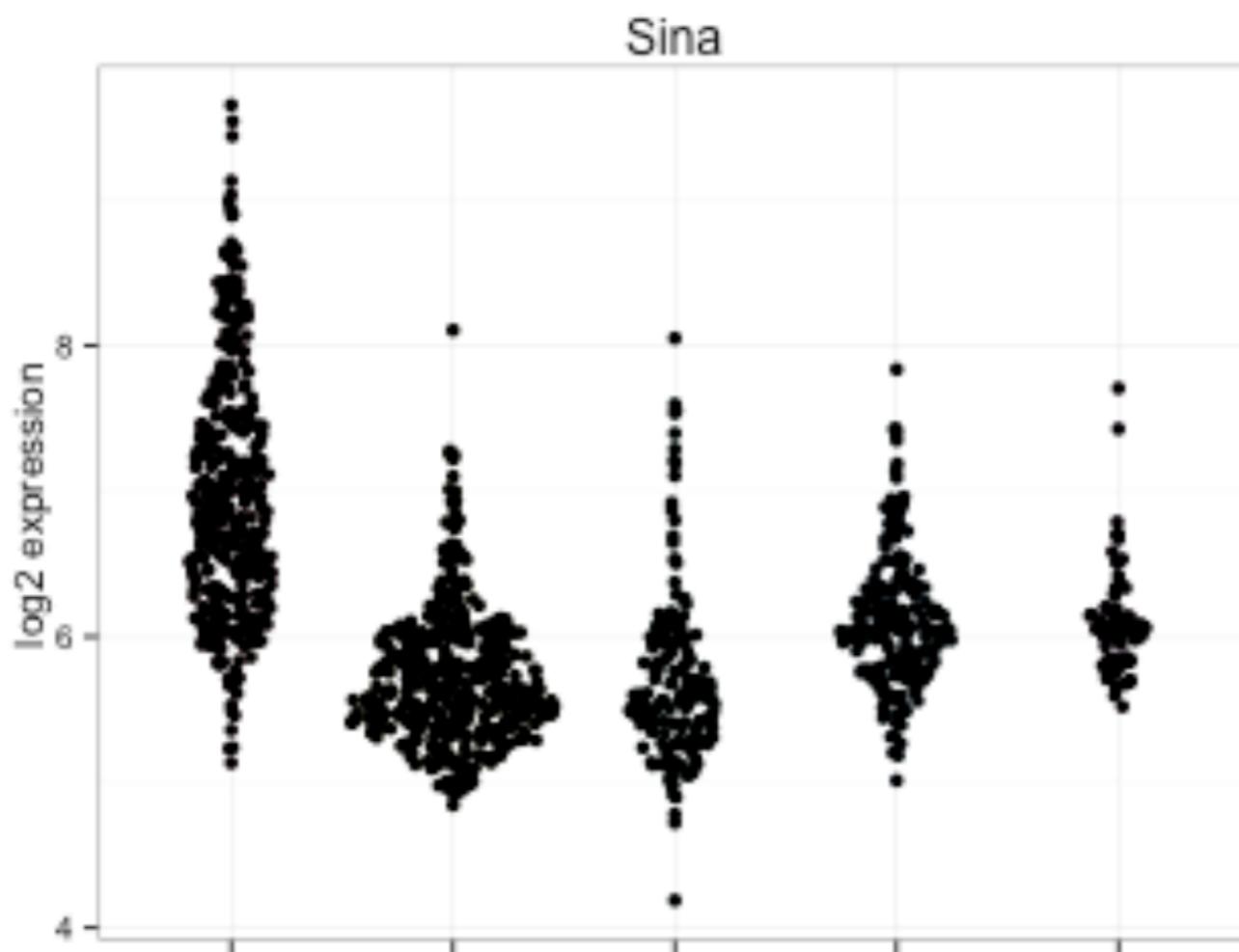
- Muestra el número de puntos, la distribución, valores atípicos, y la dispersión de los datos



Sina plot

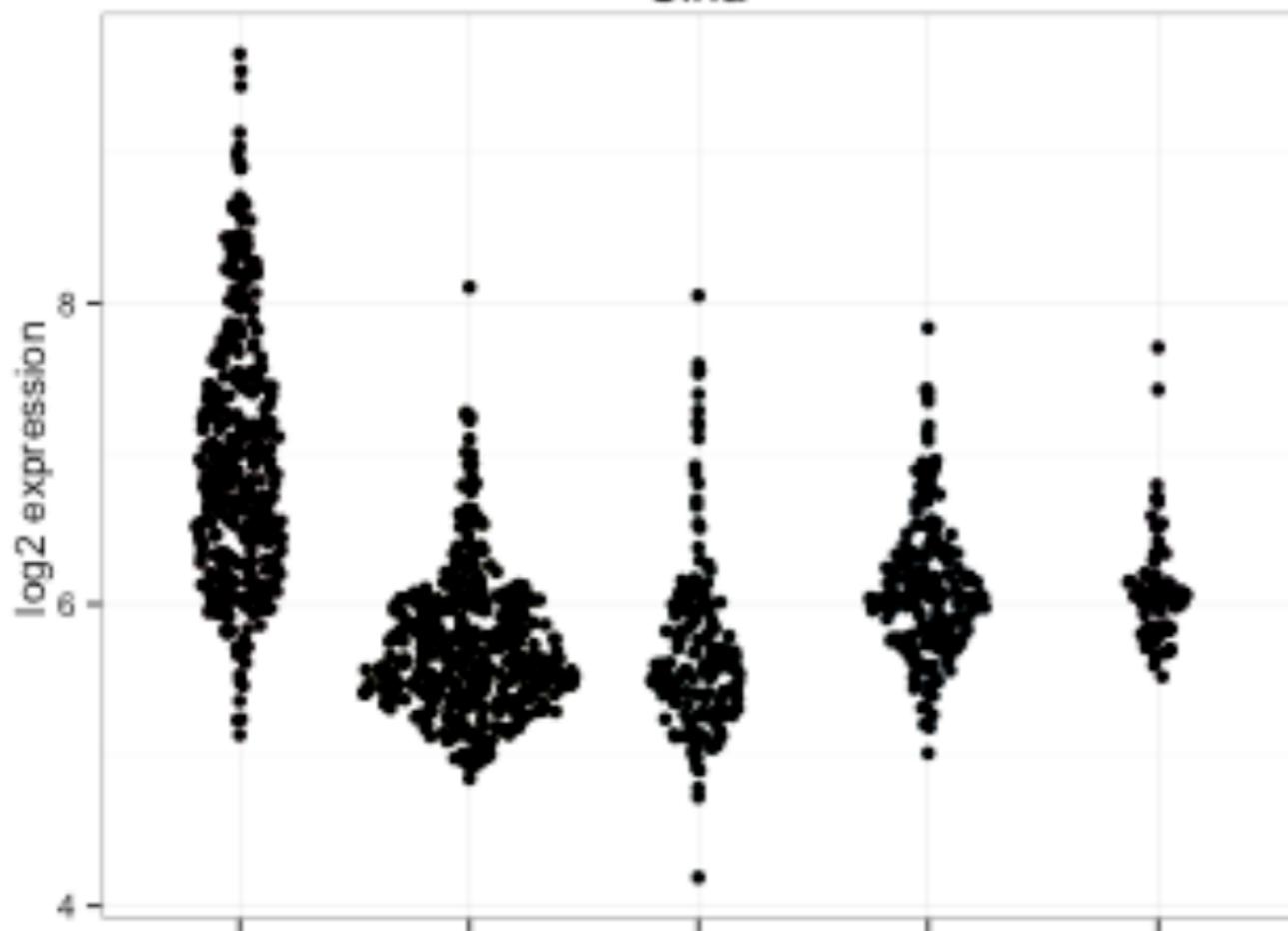


Sina plot

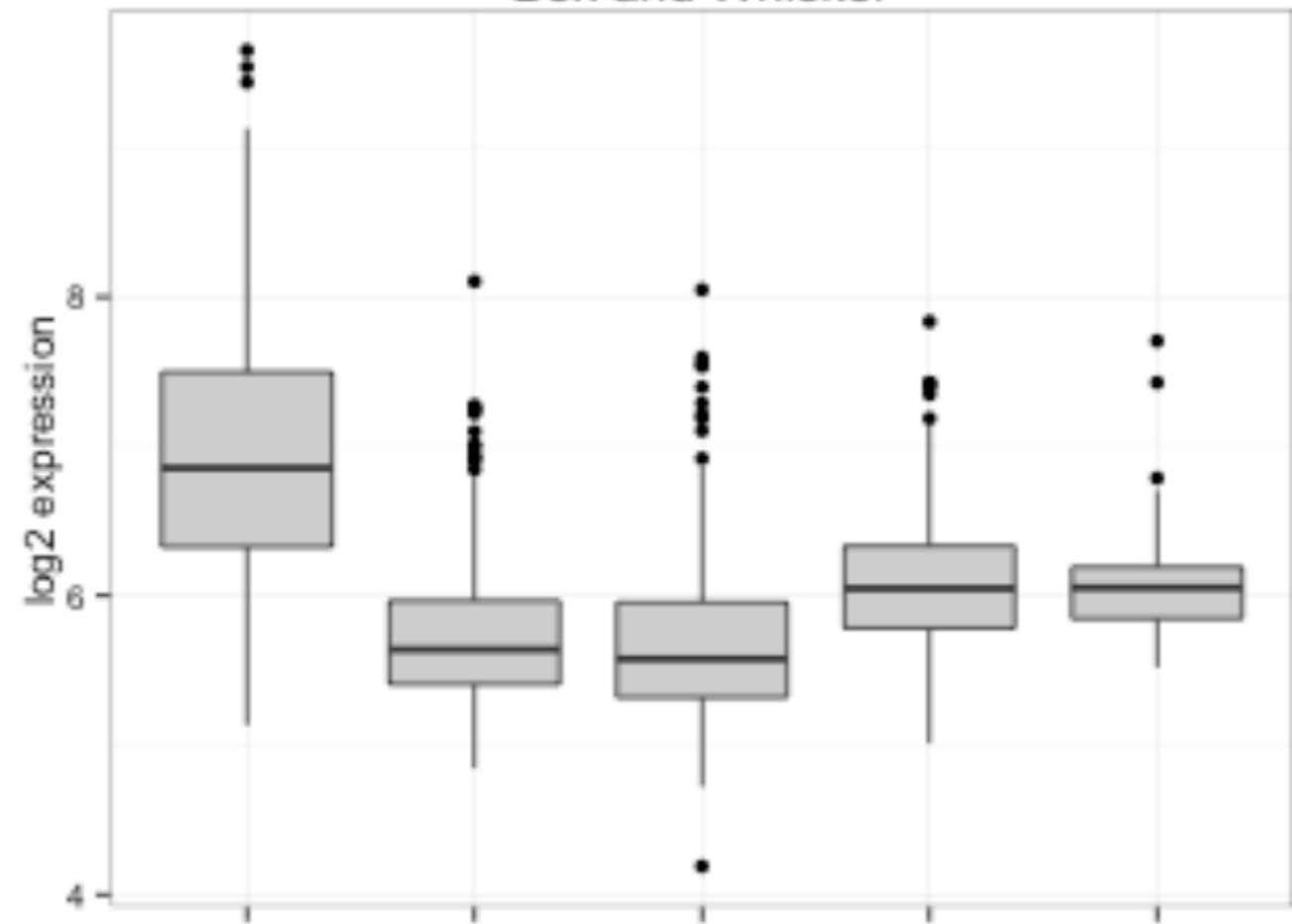


Sina plot

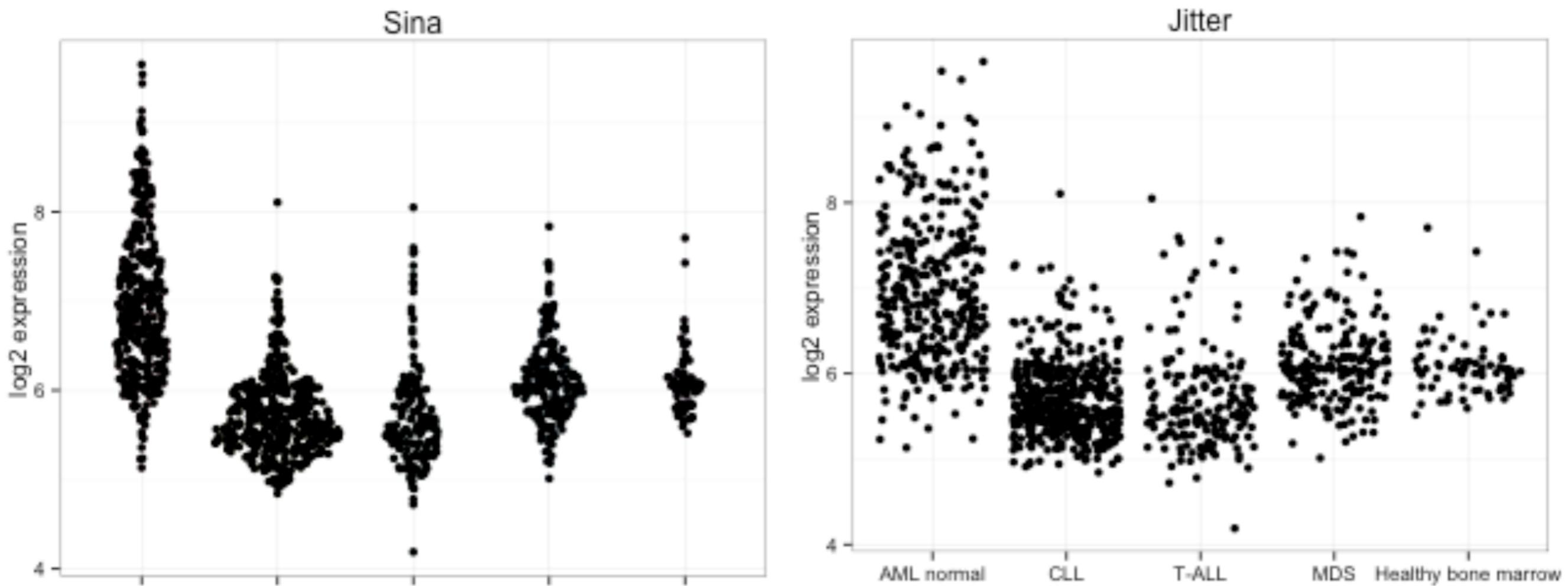
Sina



Box and Whisker

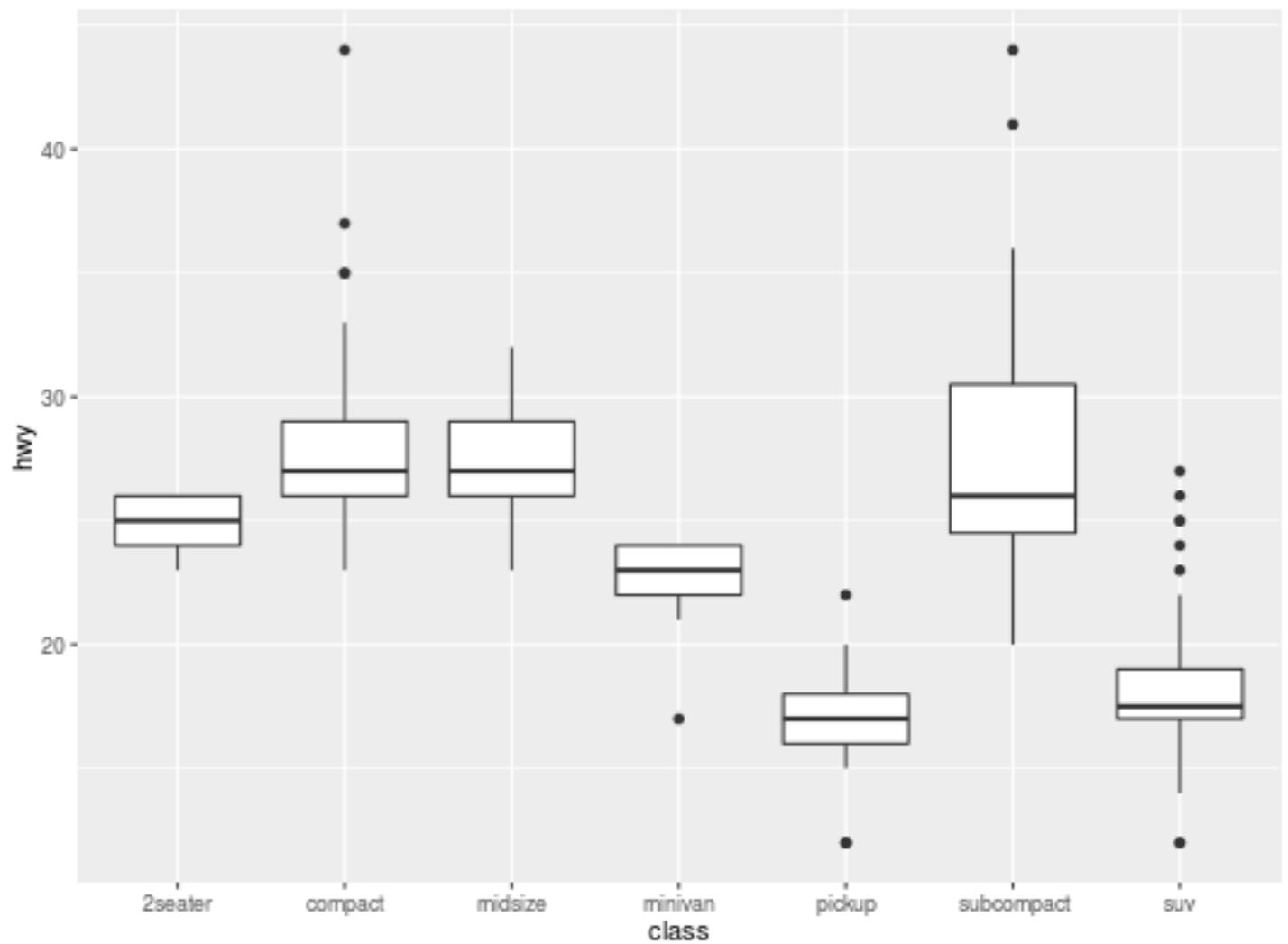


Sina plot



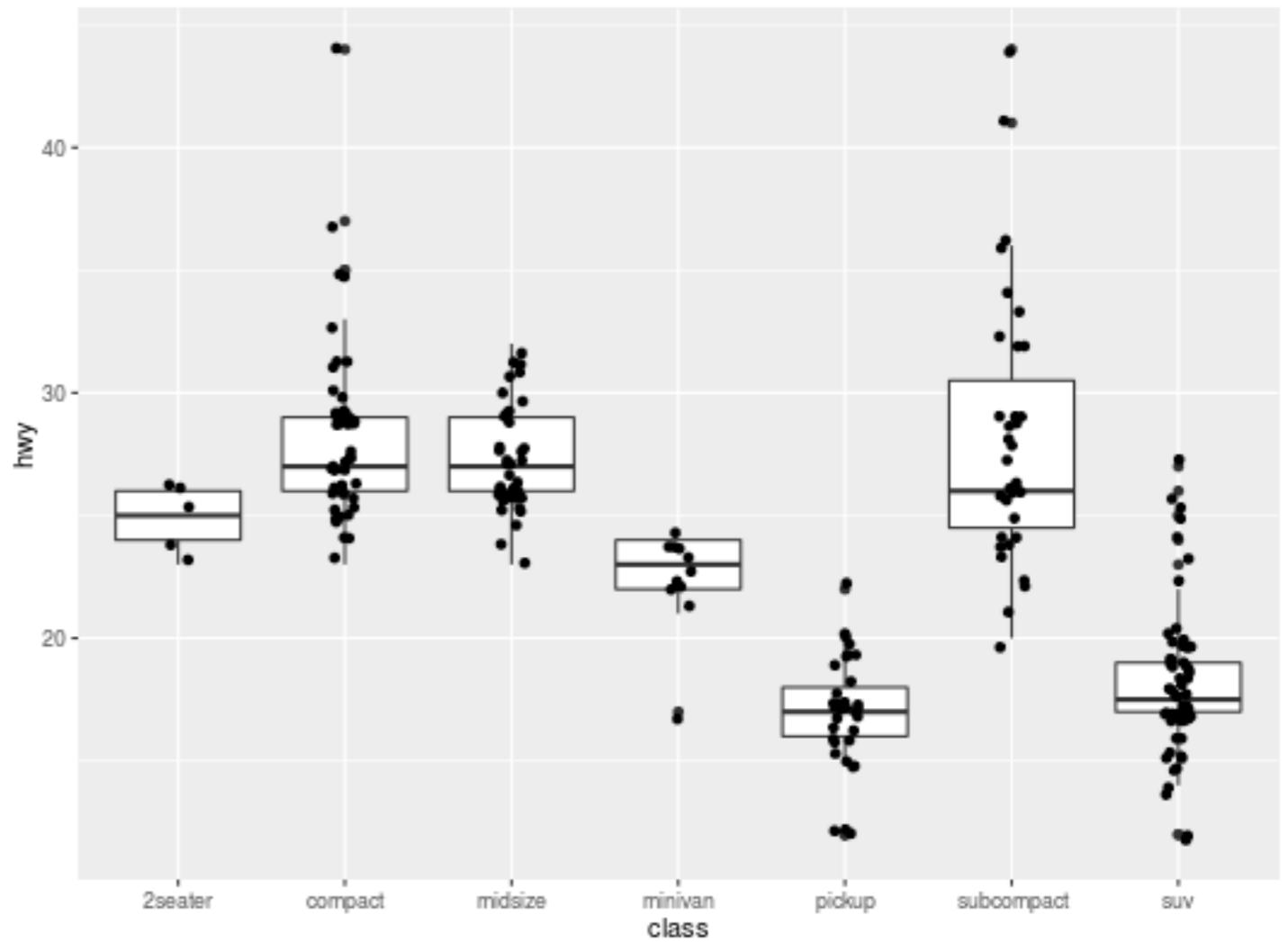
Un poco de ggplot2

- probemos
geom_boxplot()
- p <- ggplot(mpg,
aes(class, hwy))
- p + geom_boxplot()



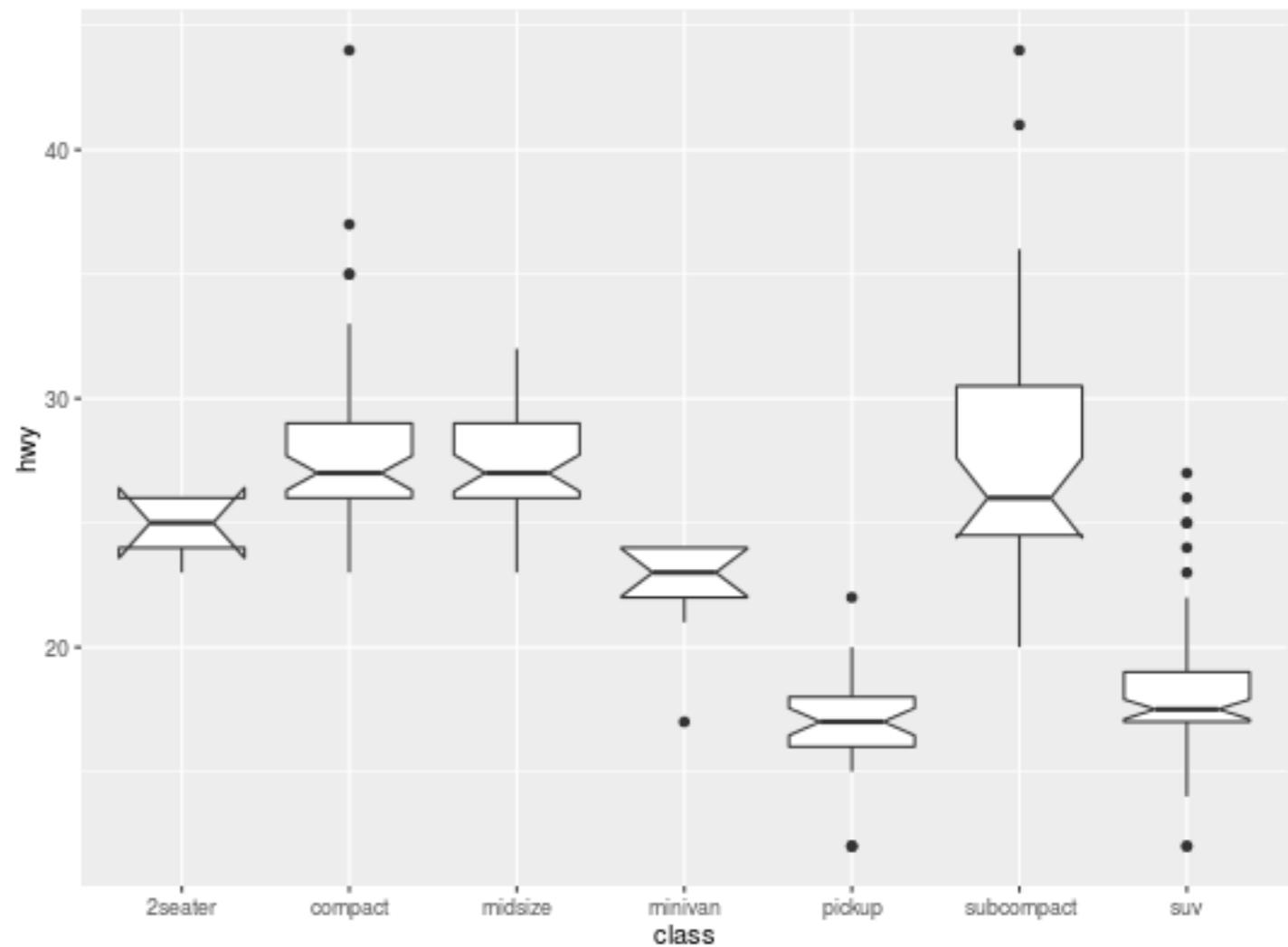
Un poco de ggplot2

- Podemos decirle a ggplot que nos muestre los puntos
- `p + geom_boxplot() + geom_jitter(width = 0.2)`



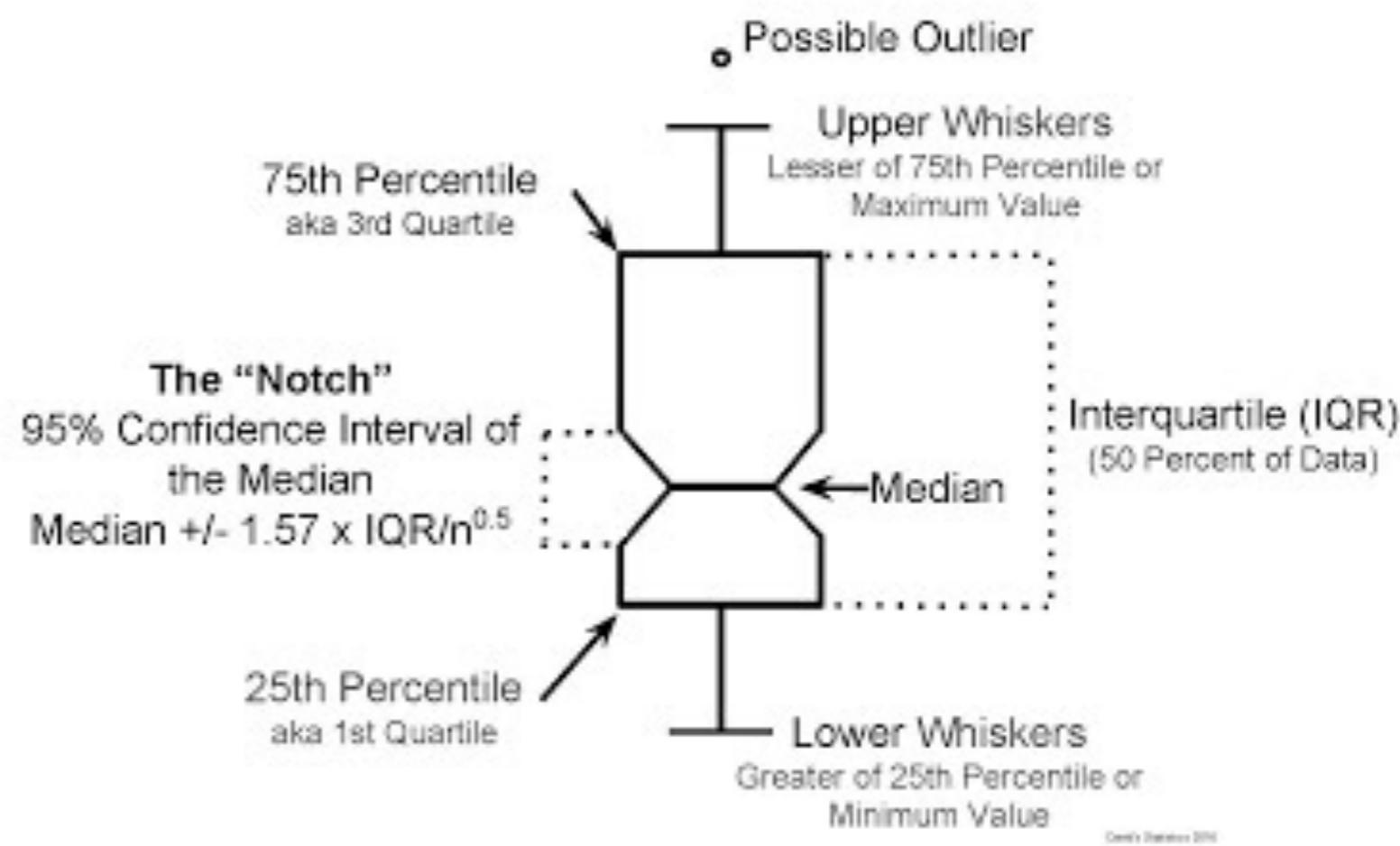
Un poco de ggplot2

- Podemos decirle a ggplot que queremos ver intervalos de confianza en nuestro boxplot
- ```
p +
 geom_boxplot(notch =
 TRUE)
```



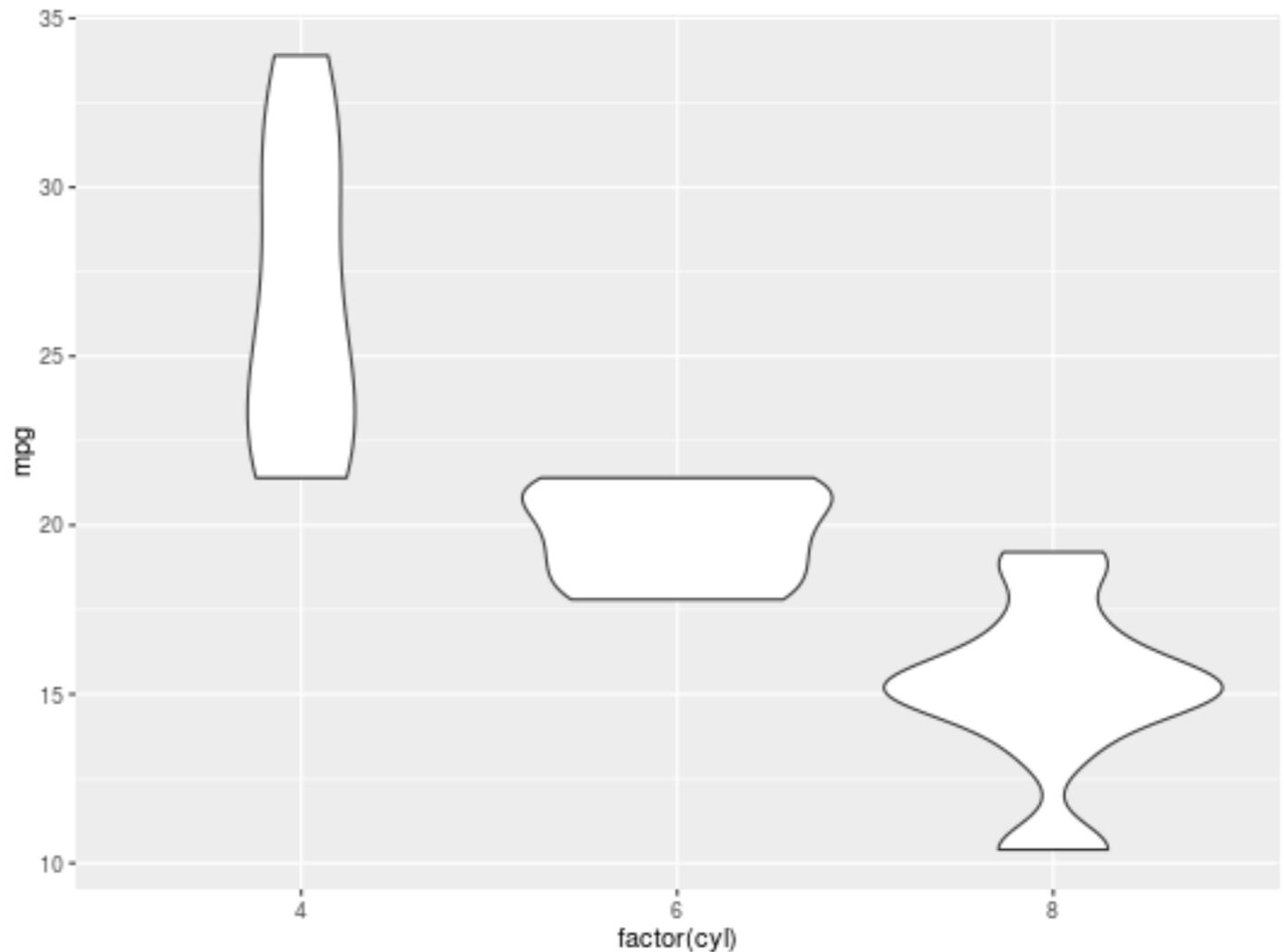
# Agregar un “notch” o muesca

- Si las muescas no se superponen, hay evidencia fuerte de que las medianas son significativamente diferentes



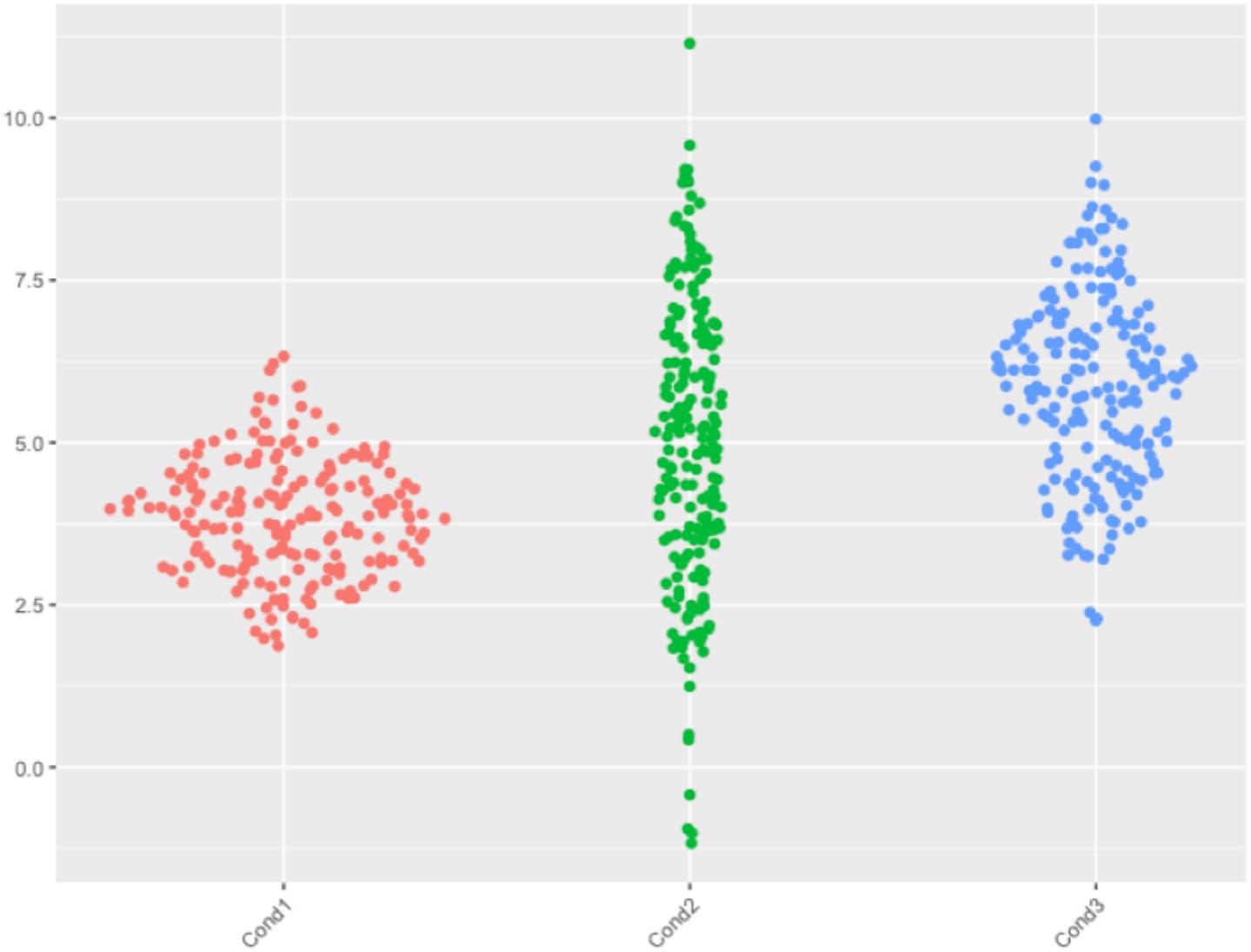
# Un poco de ggplot2

- De manera similar podemos construir un violin plot
- ```
p <- ggplot(mtcars, aes(factor(cyl), mpg))
```
- ```
p + geom_violin()
```



# Dejemos instalado el paquete sinaplot

- `install.packages("sinaplot")`
- `library("sinaplot")`
- `x <- c(rnorm(200, 4, 1),  
rnorm(200, 5, 2),  
rnorm(200, 6, 1.5))`
- `groups <- c(rep("Cond1",  
200), rep("Cond2", 200),  
rep("Cond3", 200))`
- `sinaplot(x, groups)`



“They are different, but not different enough to matter [...]”

– Roald Hoffmann, Premio Nobel de Química 1981