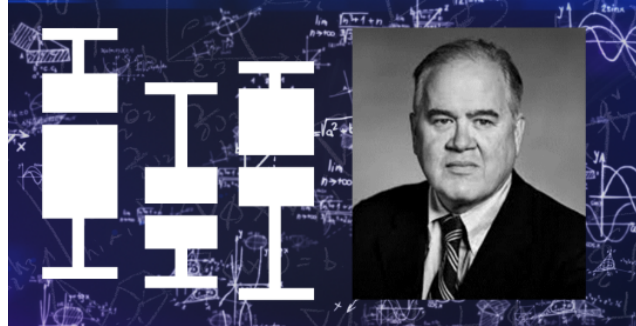
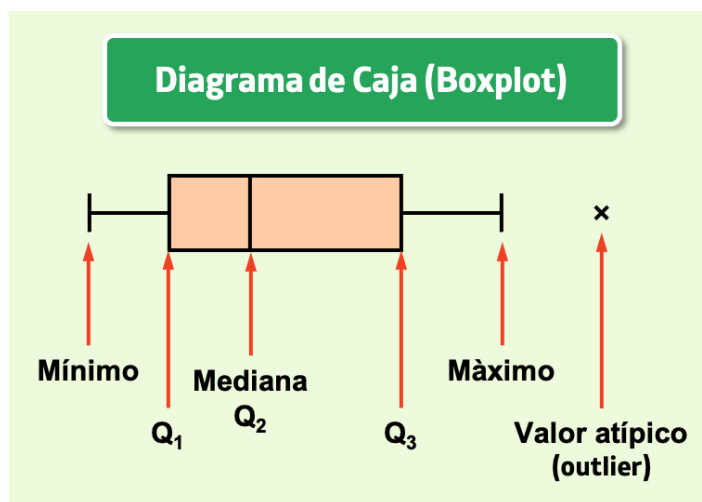


DIAGRAMA DE CAJA



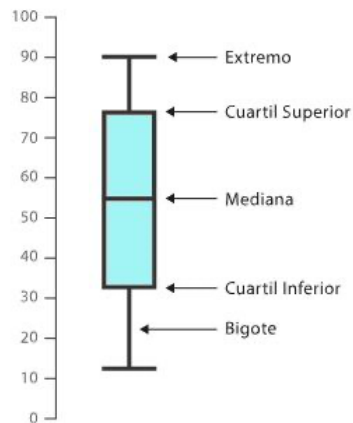
El diagrama de caja fue inventado por un genio de la estadística llamado John Tukey en los años 70. Como parte de su enfoque de análisis exploratorio de datos, Tukey propuso este tipo de gráfico porque permite ver mucha información estadística en una sola imagen, incluyendo la mediana, la dispersión y posibles valores atípicos.

El **diagrama de caja** o **boxplot** es una representación gráfica que muestra la distribución de un conjunto de datos a través de sus medidas de posición: mínimo, primer cuartil (Q1), mediana (Q2), tercer cuartil (Q3) y máximo. Además, permite identificar valores atípicos (outliers).



<https://share.google/images/PZrRKCxNBuFKJKhDI>

Componentes principales del diagrama de caja:



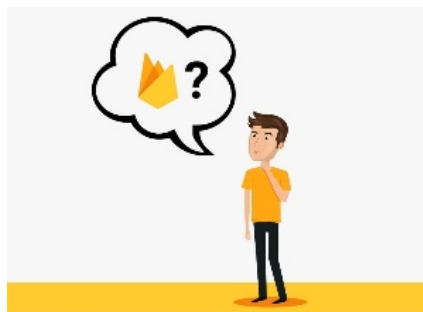
<https://share.google/images/G4cUNzoMfy8SS54cB>

Caja (box): Va desde el primer cuartil (Q1) hasta el tercer cuartil (Q3). Representa el rango intercuartílico (RIC), es decir, el 50% central de los datos.

Línea dentro de la caja: Indica la mediana (Q2), que divide a la mitad la muestra.

Bigotes (whiskers): Líneas que se extienden desde los extremos de la caja hasta el valor mínimo y máximo dentro de un rango aceptable (normalmente hasta $1.5 \times \text{RIC}$ desde Q1 y Q3).

Valores atípicos (outliers): Puntos fuera de los bigotes, que pueden ser errores o valores inusualmente altos o bajos.



¿Para qué sirve?

- Visualizar la distribución y dispersión de los datos.
- Comparar varias muestras o grupos rápidamente.
- Identificar asimetrías o sesgos en los datos.
- Detectar valores atípicos.

.Referencia:

Tukey, J. W. (1977). Exploratory Data Analysis. Estados Unidos. Addison-Wesley.

Triola, M. F. (2019). Estadística (12.ª ed.). México. Pearson Educación.

Montgomery, D. C., & Runger, G. C. (2014). Estadística Aplicada y Probabilidad para Ingenieros (6.ª ed.).

México. Wiley.