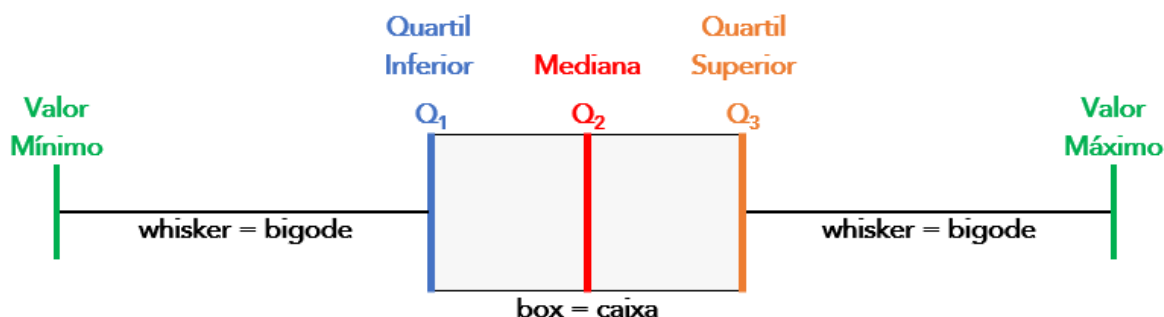


BOX PLOT

Um **boxplot** (também chamado de **box-and-whisker plot**) é uma ferramenta gráfica frequentemente utilizada na análise exploratória de dados que permite visualizar a distribuição dos dados e os valores discrepantes (outliers), assim como a distância dos valores extremos em relação à maioria dos dados. Essa ferramenta **resume cinco medidas descritivas** de um conjunto de dados, incluindo: **o valor mínimo, o primeiro quartil, a mediana, o terceiro quartil e o valor máximo**.

Para construir um gráfico de **boxplot**, usamos uma haste horizontal ou vertical e uma caixa retangular (**box**). O local em que a haste começa (da esquerda para a direita) indica o **valor mínimo** e o ponto em que a haste termina indica o **valor máximo**.

A caixa retangular, localizada no meio da haste, em geral, possui três linhas. A primeira linha, na extremidade esquerda da caixa, indica o **primeiro quartil**. A terceira linha, na extremidade direita, indica o **terceiro quartil**. A linha do meio, no interior da caixa, indica o **segundo quartil ou a mediana**. O segundo quartil pode estar entre o primeiro e o terceiro quartis, ou pode coincidir com um, ou outro, ou ambos.



Além disso, há dois traços, chamados de **whiskers** (ou bigodes), ligando o valor mínimo à extremidade esquerda da caixa e o valor máximo à extremidade direita da caixa. Cada um desses traços comporta, aproximadamente, 25% dos dados. O restante, cerca de 50%, está distribuído no interior da caixa.

Também podemos encontrar gráficos de **box plot** com pontos ou asteriscos marcando valores **discrepantes (outliers)**. Nesses casos, os **whiskers** não se estendem aos valores mínimo e máximo do conjunto de dados, mas ficam limitados a um comprimento máximo de $1,5 \times DIQ$, em que **DIQ** é a **distância interquartílica**.

A **distância interquartílica** (ou amplitude interquartílica, ou intervalo interquartílico) é calculada pela fórmula:

