

# 05

## Para saber mais: histograma e boxplot

Falamos bastante sobre histograma e boxplot durante as nossas aulas e construímos estes visuais utilizando as bibliotecas do Python. Mas afinal o que são essas representações visuais?

### Histograma

Um histograma é uma representação visual que mostra a distribuição dos dados em um dado intervalo por meio de colunas. Os dados são divididos em intervalos, chamados de “bins”, que apresentam quantas observações estão presentes naquela faixa de dados. Ele é amplamente utilizado para visualizar a distribuição de frequências ou contagens em um conjunto de dados.

Este tipo de gráfico é eficaz para mostrar a distribuição de variáveis discretas ou categorias em que a altura de cada barra indica a contagem ou proporção de observações naquela categoria. Isso facilita a identificação de tendências, como a concentração de observações em determinados intervalos, e ajuda a compreender também a distribuição geral dos dados.

Como vimos na aula, é importante termos atenção no equilíbrio entre o número apropriado de “bins” e/ou a largura dos “bins” para representação dos dados, pois bins estreitos podem gerar ruídos por poucas observações em determinados intervalos e bins largos podem descaracterizar a curva e reduzir a sensibilidade na variação das faixas.

### Boxplot

O Boxplot, também conhecido como “diagrama de caixa”, é uma representação visual que apresenta a distribuição de um conjunto de dados por meio de uma série de estatísticas descritivas: quartis, mediana, mínimo e máximo.

Este tipo de gráfico auxilia na visualização da dispersão e centralidade dos dados, identificando assimetrias, outliers e variações nos dados.

---

Você pode aprender um pouco mais sobre estas visualizações de dados no curso [Estatística com Python: frequências e medidas](#).