

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**

Visualização de dados

Tipos de gráficos e visualizações

Aula 2

Código da aula: [DADOS]ANO1C3B3S22A2

Visualização de
dados

Mapa da Unidade 3 Componente 4

Você está aqui!
Tipos de gráficos e
visualizações

semana

22

Tipos de gráficos e
visualizações

semana

24

semana

17

Introdução à
visualização de
dados

semana

20

Tipos de gráficos e
visualizações

semana

26

Cores e elementos
visuais

Visualização de
dados

Mapa da
Unidade 3
Componente 4

Você está aqui!

Tipos de gráficos e
visualizações

Aula 2

Código da aula: [DADOS]ANO1C3B3S22A2

22



Objetivo da aula

- Conhecer os diferentes tipos de gráficos e visualizações de dados, ensinando como escolher e criar cada um deles para representar informações de forma clara e eficaz.



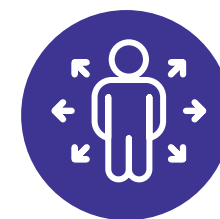
Recursos didáticos

- Recurso audiovisual para exibição de vídeos e imagens;
- Acesso ao laboratório de informática e/ou à internet.



Duração da aula

50 minutos



Competências técnicas

- Compreender e dominar técnicas de manipulação de dados;
- Extrair, transformar e carregar conjuntos de dados de diferentes fontes, garantindo a qualidade e a integridade dos dados;
- Criar e compreender visualizações gráficas.



Competências socioemocionais

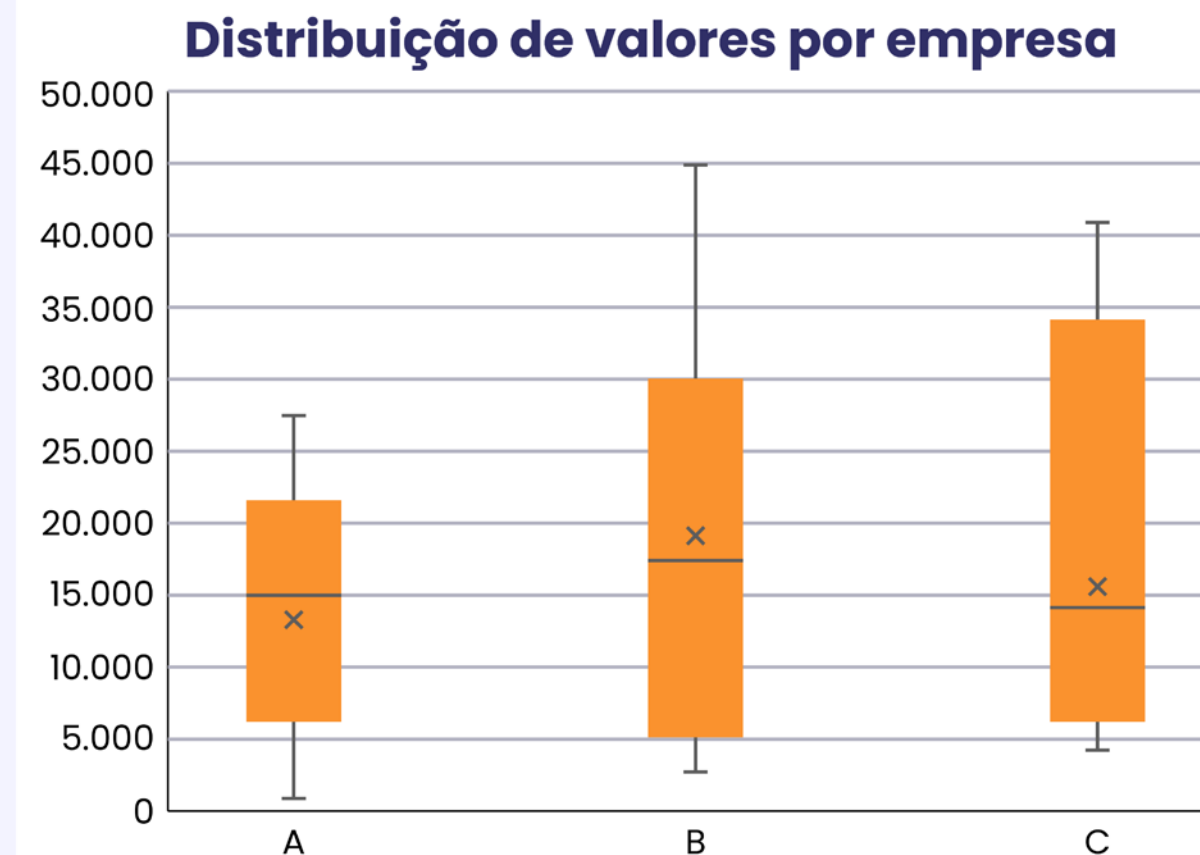
- Colaborar efetivamente com outros profissionais, como cientistas de dados e engenheiros de dados;
- Trabalhar em equipes multifuncionais, colaborando com colegas, gestores e clientes.

Construindo o conceito

Revisando: gráfico de caixa

O gráfico de caixa é útil para comparar a distribuição de dados entre diferentes grupos ou para identificar valores discrepantes em um conjunto de dados.

Na primeira semana de aula sobre tipos de gráficos e visualizações, vimos rapidamente como plotar no Excel um gráfico de caixa, porém não entramos muito em detalhes. Este tipo de gráfico auxilia o cientista de dados a entender a distribuição dos dados na fase de análise exploratória.



Elaborado especialmente para o curso.

Construindo
o **conceito**

Revisando: gráfico de caixa

Entender este tipo de gráfico exige conhecer de antemão alguns conceitos sobre distribuição de dados em Estatística.

Por isso, nesta aula, antes de praticar um pouco mais o gráfico de caixa (*Boxplot*), vamos discorrer sobre os seguintes conceitos estatísticos: média, mediana, quartis e *outliers*.

Construindo
o **conceito**

***Outlier* (Valor discrepante)**

Em Estatística, um *outlier* é um valor ou uma observação que se destaca significativamente dos outros dados em um conjunto.

Imagine que você está medindo a altura de uma turma de estudantes, e a maioria tem entre 1,50 e 1,80 metros, mas, de repente, você encontra alguém com 2,10 metros. Esse valor é muito diferente dos demais, e é isso que chamamos de *outlier*.

Os *outliers* podem ser causados por vários fatores, como erros na coleta de dados, variações naturais ou até mesmo algo incomum que realmente aconteceu.

Situação fictícia elaborada especialmente para o curso.

Construindo
o **conceito**

***Outlier* (Valor discrepante)**

Saber identificar e entender os *outliers* é importante, porque eles podem afetar a análise dos dados.

Por exemplo, eles podem distorcer a média ou a mediana de um conjunto de dados, levando a conclusões erradas. Por isso, os estatísticos precisam decidir como lidar com os *outliers*, o que pode envolver investigá-los mais a fundo, removê-los ou ajustar os métodos de análise para levar em conta essas variações.

Em resumo, os *outliers* são pontos de dados que se destacam, e lidar com eles de forma adequada é essencial para uma análise estatística precisa.

Construindo o **conceito**

Média

Em Estatística, a média é uma medida central que nos ajuda a entender a tendência geral de um conjunto de dados.

Imaginem que temos as notas de uma prova de Matemática de uma turma e queremos saber, de forma geral, como a turma se saiu. Para encontrar a média, somamos todas as notas e dividimos pelo número total de alunos.

Por exemplo, se temos cinco alunos, com as notas 6, 8, 9, 7 e 10, somamos essas notas, obtendo 40, e, em seguida, dividimos por 5, resultando em uma média de 8. Isso significa que, em média, os alunos tiraram nota 8 na prova.

Situação fictícia elaborada especialmente para o curso.

Construindo o **conceito**

Média

A média é útil, porque nos dá uma ideia da distribuição dos dados em torno de um ponto central. No entanto, é importante lembrar que a média pode ser afetada por valores muito altos ou muito baixos, conhecidos como *outliers*.

Por isso, em algumas situações, podemos usar outras medidas, como a mediana, para ter uma visão mais equilibrada do conjunto de dados. Mas, de forma geral, a média é uma ferramenta valiosa para resumir informações e nos ajudar a entender as características principais de um grupo de dados.

Construindo o **conceito**

Mediana

Em Estatística, a mediana é uma medida de tendência central que nos indica o valor central de um conjunto de dados quando esses dados estão ordenados. Para encontrar a mediana, primeiro organizamos os dados em ordem crescente ou decrescente.

Se o número de observações é ímpar, a mediana é o valor que está exatamente no meio da lista. Por exemplo, se temos as idades de cinco estudantes: 13, 15, 16, 17 e 20, ao organizá-las em ordem crescente (13, 15, 16, 17, 20), vemos que a idade do meio é 16, então essa é a mediana. Se o número de observações é par, a mediana é a média dos dois valores centrais. Por exemplo, se temos seis estudantes com idades de 13, 14, 15, 17, 18 e 20, os valores do meio são 15 e 17, então a mediana seria a média deles, ou seja, 16.

Construindo o **conceito**

Mediana

A mediana é particularmente útil quando temos *outliers* ou valores extremos em nossos dados, pois, diferentemente da média, ela não é influenciada por esses valores extremos.

Por exemplo, se em um grupo de salários, a maioria ganha em torno de R\$ 2.000,00, mas há uma pessoa que ganha R\$ 50.000,00, a média seria puxada para cima por esse valor alto, mas a mediana ainda refletiria a realidade da maioria dos salários. Portanto, a mediana nos dá uma ideia mais precisa do que é "típico" ou "comum" em um conjunto de dados, especialmente quando a distribuição é assimétrica.

Construindo
o **conceito**

Quartis

Em estatística, os quartis são valores que dividem um conjunto de dados ordenados em quatro partes iguais, cada uma contendo aproximadamente 25% dos dados. Eles são uma forma de resumir a distribuição dos dados, ajudando a entender como eles estão espalhados e a identificar possíveis outliers ou valores extremos. Existem três quartis principais: o primeiro quartil (Q1), o segundo quartil (Q2) e o terceiro quartil (Q3).

Construindo o **conceito**

Quartis

Ao analisar os quartis, podemos ter uma visão mais clara da dispersão dos dados e das tendências gerais, o que é fundamental para a tomada de decisões baseadas em dados.

O primeiro quartil, Q_1 , é o valor que divide os 25% menores dados do restante do conjunto. Para encontrar Q_1 , organizamos os dados em ordem crescente e identificamos o valor que separa os 25% inferiores dos outros 75%. De forma similar, o terceiro quartil, Q_3 , é o valor que divide os 75% menores dados dos 25% maiores. É o ponto médio dos dados que estão acima da mediana geral. Assim, Q_1 e Q_3 nos dão uma ideia da variação dos dados nos extremos inferior e superior, respectivamente.

Construindo o **conceito**

Quartis

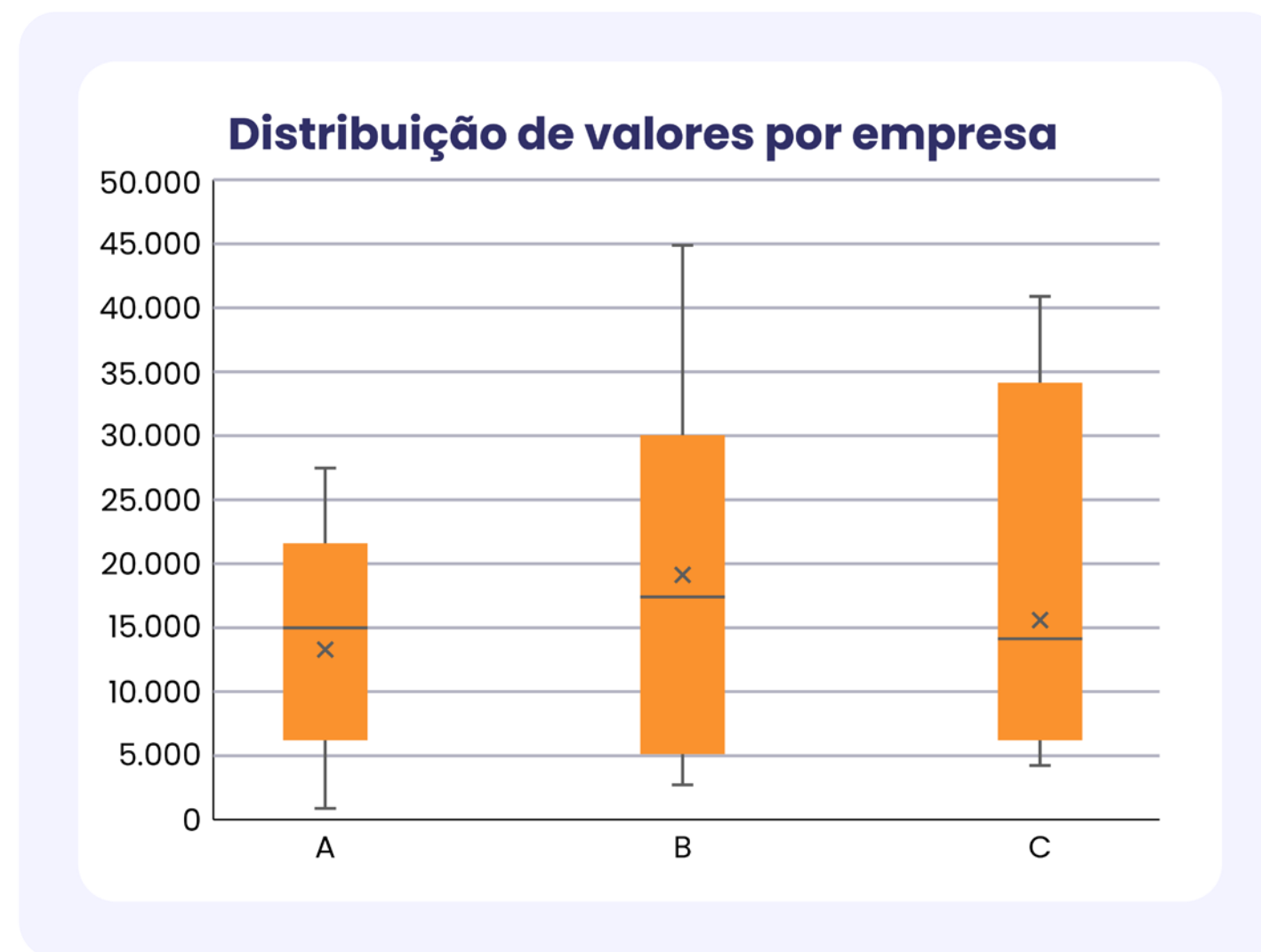
O segundo quartil, Q_2 , é simplesmente a mediana do conjunto de dados, dividindo-o ao meio. Metade dos dados está abaixo de Q_2 , e a outra metade está acima.

Os quartis são particularmente úteis para criar gráficos de caixa, ou box plots, que fornecem uma representação visual da distribuição dos dados, incluindo a mediana, os quartis e os possíveis outliers. Ao analisar os quartis, podemos ter uma visão mais clara da dispersão dos dados e das tendências gerais, o que é fundamental para a tomada de decisões baseadas em dados.

Construindo
o **conceito**

Gráfico de caixa

O *boxplot* (ou diagrama de caixa, numa tradução livre) mostra a distribuição quantitativa dos dados de um jeito que facilita a comparação entre as variáveis ou através dos níveis categóricos das variáveis.



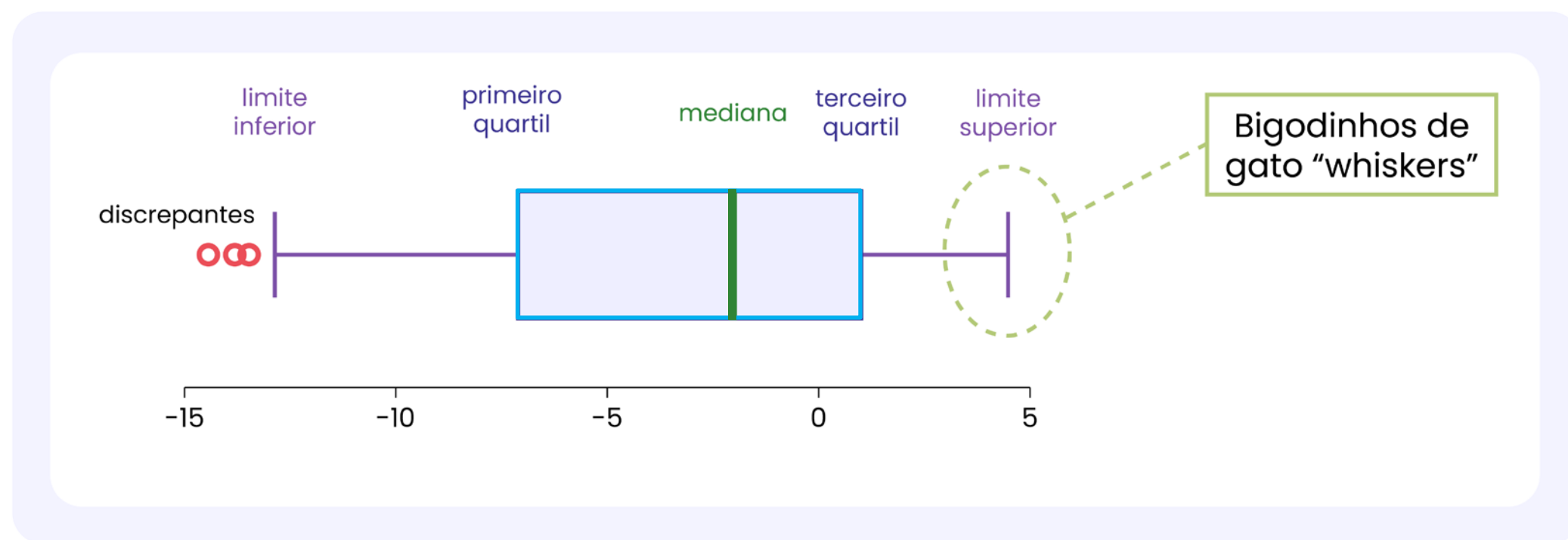
Elaborado especialmente para o curso.

Construindo o **conceito**

Gráfico de caixa

Esta caixa (*box*) mostra os quartis do *dataset*, enquanto os *whiskers* mostram o resto da distribuição, exceto os pontos que são chamados de *outliers*.

A linha no centro da figura em verde, no nosso caso, representa a mediana. A linha azul que está na aresta esquerda da figura representa 25% da distribuição, e a linha que está na aresta direita representa 75% da distribuição dos dados.



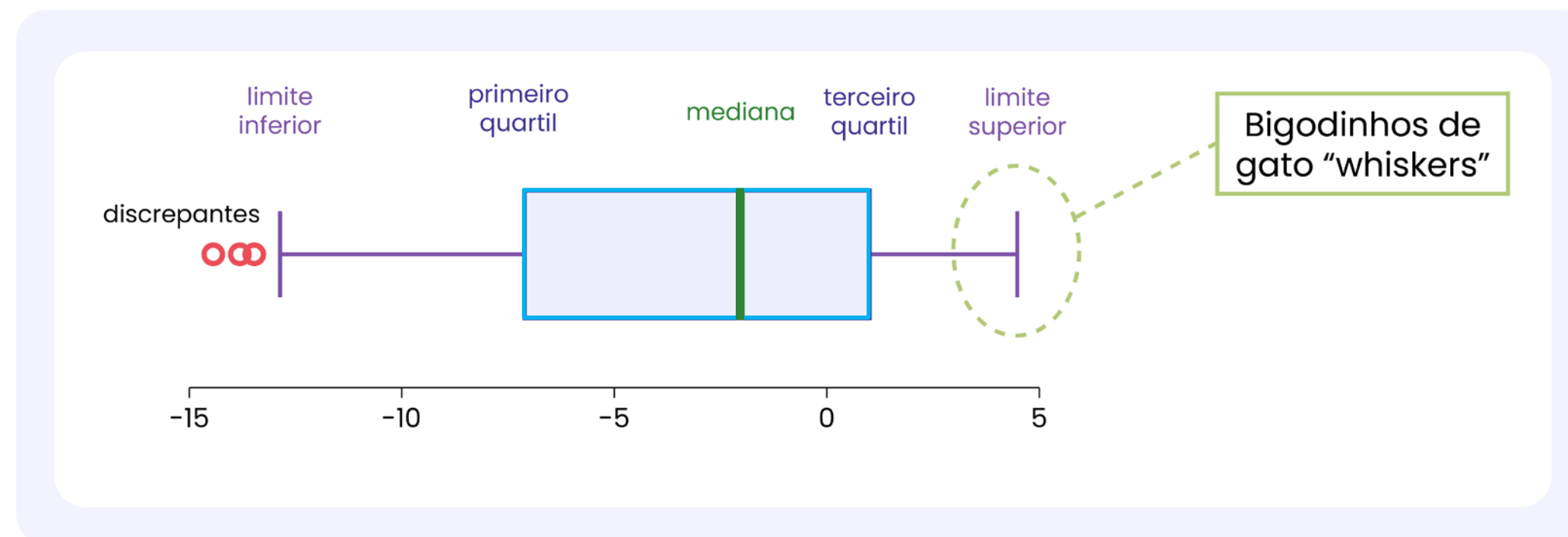
Elaborado especialmente para o curso.

Construindo o conceito

Gráfico de caixa

As linhas roxas, que estão nas extremidades de ambos lados, saem da figura central e mostram uma delimitação. Temos o início e o fim da área dos quartis.

Agora, repare que há alguns pontos após essa delimitação no lado esquerdo. Esses pontos são chamados de *outliers*, ou seja, são pontos “isolados” na nossa distribuição.



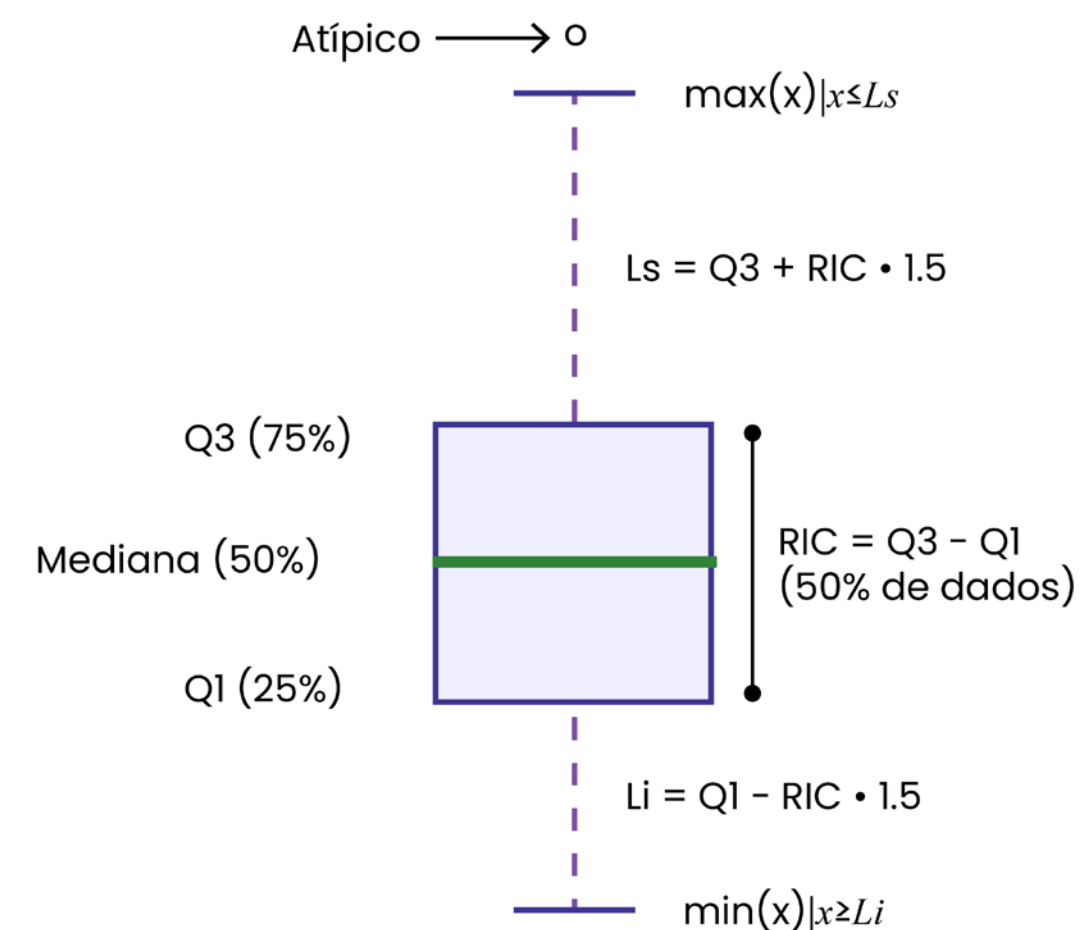
Elaborado especialmente para o curso.

Construindo o conceito

Gráfico de caixa

Para sabermos até onde vão os *whiskers*, que são esses “bigodinhos de gato” que delimitam os quartis e marcam onde se iniciam os *outliers*, vamos fazer um cálculo.

Uma vez que encontramos a mediana, o ponto central da distribuição, sabemos que temos 50% da distribuição à esquerda e 50% à direita. Após isso, vamos dividir novamente os lados na metade, para obtermos duas partes de 25% à direita e 25% à esquerda.



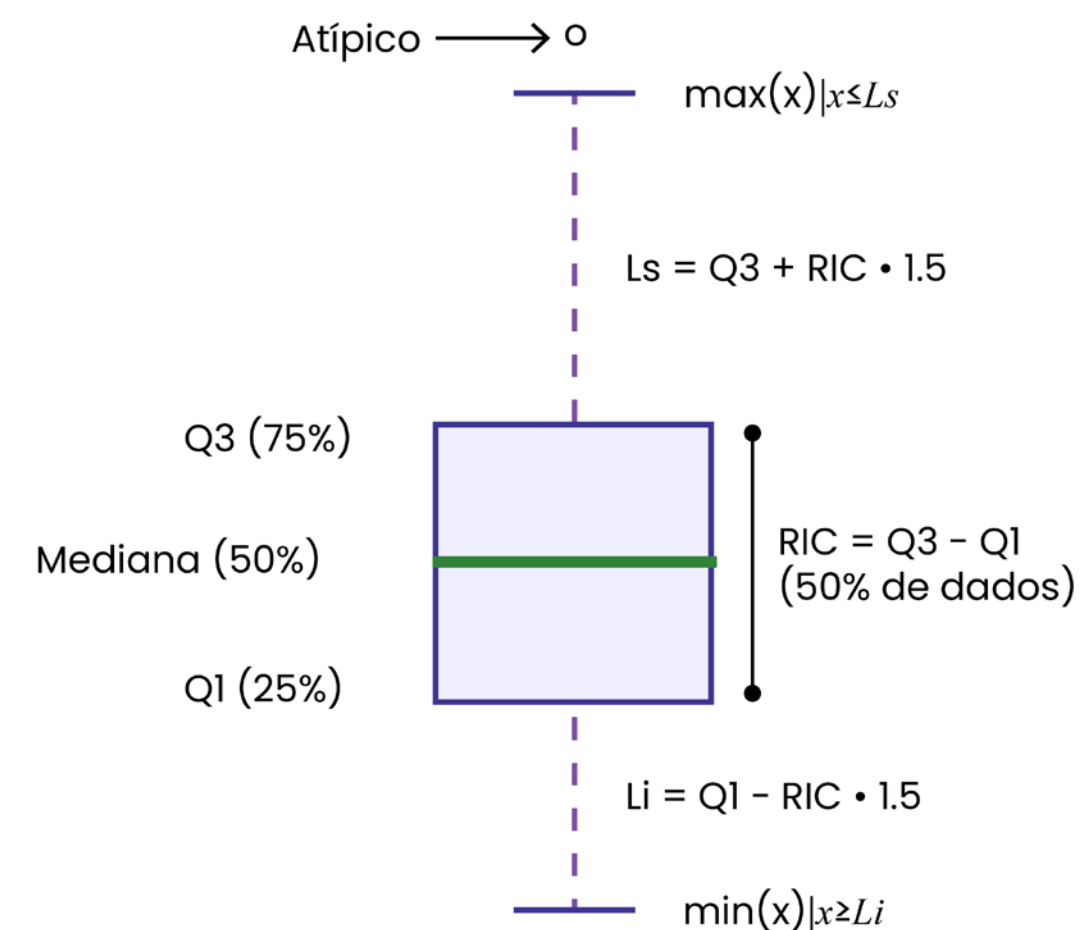
Elaborado especialmente para o curso.

Construindo o conceito

Gráfico de caixa

Conforme é possível visualizar na imagem, agora que temos os quatro quartis, cada um com 25%, podemos desenhar a caixa (*box*), e delimitamos com o 2º e o 3º quartis.

Para encontrarmos, então, a delimitação dos *whiskers*, vamos multiplicar a distância do 2º quartil até a mediana por 1,5. Fazemos o mesmo procedimento com a distância do 3º quartil até a mediana: vamos multiplicá-la por 1,5.



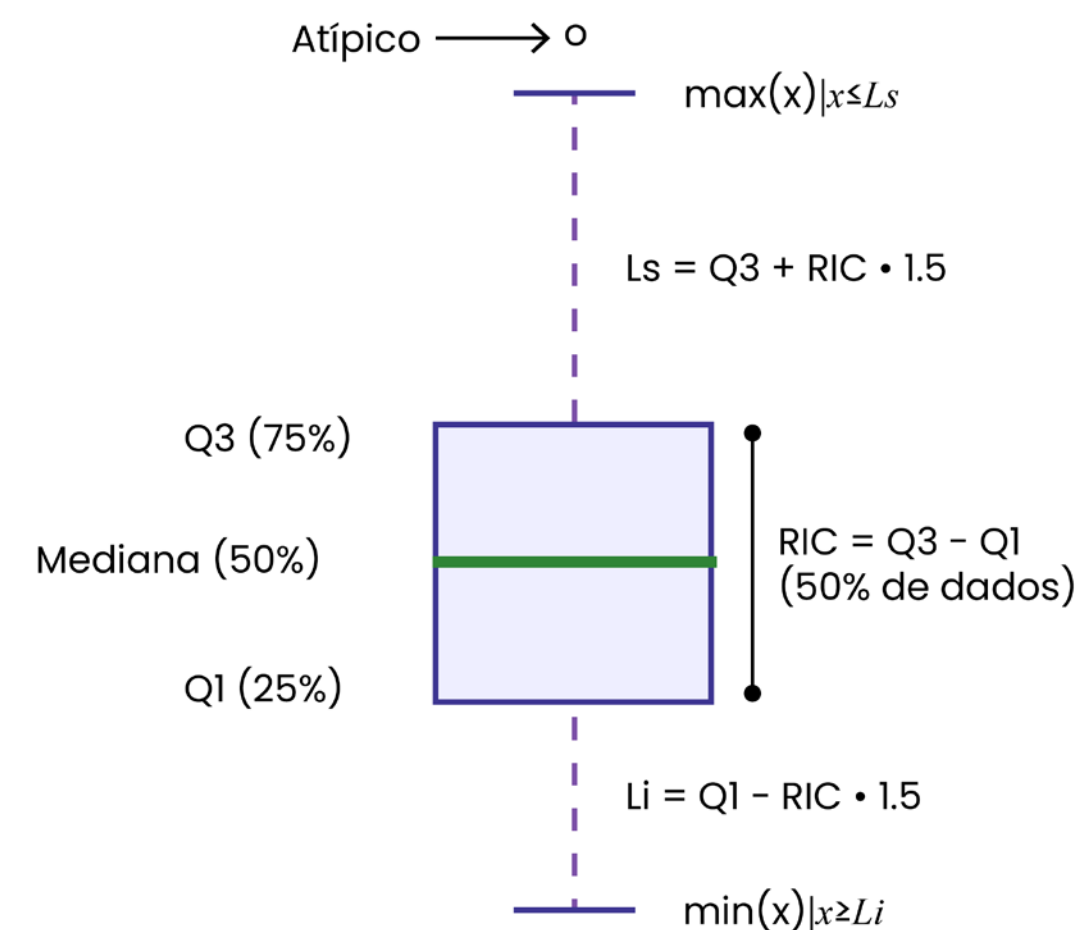
Elaborado especialmente para o curso.

Construindo o conceito

Gráfico de caixa

Em suma, o *boxplot* nos auxilia a visualizar a distribuição de dados divididos nos quartis.

Ademais, ele mostra onde os dados estão mais concentrados e se há *outliers* fora dos nossos quartis.



Elaborado especialmente para o curso.

Colocando
em **prática**

Estudo de caso: área da saúde

Análise do tempo de resposta em um teste de reação

O laboratório de neurociência Mente Ágil realizou um teste de reação com um grupo de voluntários. Neste teste, os participantes tinham que pressionar um botão o mais rápido possível após verem um estímulo visual.

Situação fictícia elaborada especialmente para o curso.

No material de apoio desta aula, encontra-se um gráfico *boxplot* que representa o conjunto de dados com os tempos de resposta (em milissegundos) de 30 voluntários para o teste.



15 minutos



Individual

Colocando
em **prática**

Estudo de caso: área da saúde

Instruções:

A partir do gráfico *boxplot* fornecido, que representa os tempos de resposta dos voluntários:

- observe o gráfico e descreva com suas próprias palavras o que você vê, especialmente em relação aos *outliers*;
- tire conclusões, com base no gráfico, sobre a distribuição dos tempos de resposta e o desempenho dos voluntários no teste de reação;
- produza o texto-síntese.



15 minutos



individual



© Getty Images

O que nós
**aprendemos
hoje?**

Então ficamos assim...

- 1** Aprendemos que o *boxplot*, também conhecido como diagrama de caixa, é uma ferramenta gráfica útil para visualizar a distribuição de um conjunto de dados;
- 2** Compreendemos que os *outliers* são pontos de dados que se destacam por serem muito diferentes dos demais valores do conjunto;
- 3** Vimos que a mediana e os quartis são medidas robustas em relação aos *outliers*, pois não são tão facilmente afetados por valores extremos quanto a média.

Saiba mais

Quer aprender mais sobre como melhorar uma análise em *boxplot*?

SIQUEIRA, D. Melhorando a análise com o boxplot. *Alura*, 3 maio 2021. Disponível em:

<https://www.alura.com.br/artigos/melhorando-a-analise-com-o-boxplot>. Acesso em: 20 jun. 2024.

Referências da aula

KNAFLIC, C. N. *Storytelling com dados*: um guia sobre visualização de dados para profissionais de negócios. Rio de Janeiro: Alta Books, 2019.

SIQUEIRA, D. Melhorando a análise com o boxplot. *Alura*, 3 maio 2021. Disponível em: <https://www.alura.com.br/artigos/melhorando-a-analise-com-o-boxplot>. Acesso em: 20 jun. 2024.

Identidade visual: imagens © Getty Images.

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**