

Gráfico de boxplot

Transcrição

[0:01] Até o momento, nós já geramos diferentes gráficos, sendo mais exato, 6 gráficos, obtendo diferentes informações para o cursinho.

[0:11] Mas agora o cursinho deseja fazer uma análise mais estatística com as notas da redação, como identificar a variação das notas, como por exemplo as medianas e identificar valores acima do normal.

[0:25] Essas informações devem conter todos os estados.

[0:28] Então, quero fazer uma comparação entre todos os estados com informações mais estatísticas e principalmente onde há filiais do cursinho.

[0:39] Em estatística, há um recurso muito útil que possibilita obter diferentes informações em apenas um único gráfico, que é o chamado boxplot, também conhecido como gráfico de caixa.

[0:50] Esse gráfico possui as informações como quartis, medianas, limites superior e inferior, outlines, que são os valores discrepantes, caso haja nos registros.

[1:05] Então, antes de fazer o gráfico, lembrando que o boxplot, a biblioteca ggplot possui uma função específica.

[1:21] Como eu já falei para você no início, o R é bem intuitivo, então o `geom_boxplot`. Mas antes de gerarmos esse gráfico, vamos fazer alguns filtros nos dados.

[1:28] Lembrando que quem trabalha com análise de dados faz muitos filtros, muitas manipulações nos registros em que está trabalhando.

[1:38] Vamos fazer um filtro aqui eliminando do UF Prova os registros em branco a coluna UF Prova. No início, nós identificamos esses valores, não eliminamos porque eles podem interferir em outras análises distintas que não têm nada a ver com a UF Prova.

[1:56] Então, achamos melhor manter esses registros.

[2:03] E também vamos eliminar `is.na` na nota redação. Ou seja, valores não definidos na coluna da redação.

[2:14] Vamos selecionar apenas, `select`, apenas os campos que nós desejamos, ou seja, UF Prova e nota redação.

[2:33] Vamos salvar tudo isso dentro do objeto chamado `notas redação UF`. Vamos salvar essa seleção, essa manipulação. Pronto. Vamos dar um `view` aqui nesse nosso novo objeto. `View, notas, redação, uf`. Pronto.

[3:02] Temos nossos registros em apenas duas colunas, que é a UF Prova e a nota redação.

[3:11] Agora vamos gerar os `boxplot` utilizando essa nova base de dados.

[3:16] Então vamos aqui `ggplot`, `data`, `notas redação UF`, concatenar e utilizar a função que eu mostrei anteriormente para você, `geom_boxplot`.

[3:33] Maneando os eixos, o eixo x vai receber UF Prova, o eixo y vai receber nota redação.

[3:45] E vamos salvar tudo isso dentro de um objeto chamado `plot_box_uf` underline redação. Vamos executar aqui. Executamos.

[4:08] Demora um pouquinho, pronto.

[4:12] Temos um gráfico com boxplot com as notas da redação para todos os estados. Vamos aumentar o zoom aqui para melhorar a visualização.

[4:24] Antes de fazer uma análise de todas essas informações, esse aqui é o boxplot, com a variação de mediana, os quartis, os limites superior, que é o final dessa linha aqui e o inferior. Para todos os estados.

[4:36] Mas antes de fazer a análise de todas essas informações mais detalhadas, vamos fazer algumas alterações, como destacar os boxplots dos estados que o cursinho tem filial.

[4:45] Porque isso pode facilitar a análise nossa e também a análise dos associados, dos analistas de negócio do cursinho. Então vamos fazer as alterações.