

03

Personalizando gráfico boxplot

Transcrição

[0:00] Para que seja possível fazermos as alterações, destacar os boxplots das regiões das filiais do cursinho, vamos fazer uma nova manipulação dos dados, porém não vamos utilizar os dados originais, ou seja esse Enem aqui e nem o nota redação.

[0:21] Há um recurso muito interessante no ggplot, que é esse aqui: nós podemos acessar os dados diretamente os dados do objeto onde está o gráfico.

[0:33] Você pode ver que aqui é o objeto onde está armazenado esse gráfico aqui.

[0:38] Colocando o cifrão, nós acessamos diferentes parâmetros dentro do objeto, dentre eles o data frame chamado data.

[0:50] Vamos salvar dentro de um objeto temporário chamado dados, vamos dar um view aqui dados. Executar, pronto.

[0:58] Você pode ver que temos todos os registros com apenas 2 colunas, que foi o que a gente fez anteriormente.

[1:05] Então, em alguns casos isso é muito útil se nós não temos mais acesso a esses dados aqui, nós podemos acessar diretamente do objeto, mas isso daqui é só para você ver uma nova forma de acessar os dados de onde foi gerado os gráficos.

[1:22] Agora nós temos todos os dados necessários para gerar esse gráfico, salvamos os dados dentro de um objeto chamado dados, contendo todas as colunas que nós precisamos.

[1:33] Agora nós vamos criar uma nova coluna chamada filial para indicar se determinado estado tem ou não uma filial do cursinho.

[1:44] E vamos selecionar apenas essas duas colunas.

[1:46] Então vamos selecionar aqui dados, vamos utilizar a função mutate, summarise, o summarise cria uma nova base de dados e o mutate cria uma nova coluna.

[2:02] Então vamos chamar aqui de filial, vamos utilizar a função ifelse, por que nós queremos verificar cada linha de UF Prova, se esse registro está dentro das filiais desse cursinho, que é o Ceará, DF, Minas Gerais e Rio Grande do Sul, que é o RS.

[2:31] Essa função verifica se essa condição está aqui, se for verdade, ela vai retornar um valor, vamos chamar de True, T. Se for falso, nesse caso aqui, F.

[2:46] Vamos salvar tudo isso dentro do objeto chamado dados.

[2:52] Vamos executar. Finalizou, vamos abrir esse objeto, Pronto. Nós temos aqui o mesmo objeto chamado dados com uma nova coluna que indica se é true ou false, indica se aquele registro faz parte de uma filial do cursinho.

[3:20] Então no caso, Minas Gerais faz parte, São Paulo não faz parte, Rio Grande do Sul faz parte, São Paulo não, não, não, e assim por diante.

[3:28] Então, nós temos um novo conjunto de dados com um novo atributo para trabalharmos. Agora, vamos gerar o gráfico novamente, mas utilizando esse novo conjunto de dados chamado dados.

[3:41] Então aqui ggplot, data dados, chamar a função idêntica anteriormente, geom_boxplot, aes x vai receber UF Prova, y vai receber nota redação e agora vamos atribuir um novo parâmetro. Fill, que você já conhece bem, já utilizamos em diferentes ocasiões aqui, que ela faz o preenchimento dos elementos do gráfico, vamos utilizar aqui filial.

[4:22] Vamos salvar tudo isso, vamos executar aqui. Pode demorar um pouquinho porque era bastante dados. Pronto, vamos dar o zoom.

[4:44] Temos o mesmo gráfico gerado anteriormente, só que agora com os estados do Ceará, Distrito Federal, Minas Gerais e Rio Grande do Sul com cores diferentes, porque nesses estados o cursinho tem filial.

[4:57] Agora vamos fazer algumas alterações nos outlayers, por exemplo aqui em Santa Catarina, mais à direita, temos dois pontinhos que indicam valores discrepantes, ou seja, esse estado tem notas de redação acima do normal. Correto?

[5:16] Vamos defini-lo como vermelho e aumentar um pouco o tamanho dele para ficar mais visível.

[5:24] Isso é possível utilizando alguns parâmetros que eu vou mostrar agora para você.

[5:25] Vamos copiar aqui, fazer essas alterações. O parâmetro que nós vamos utilizar é o outlier.color, vamos utilizar red para ficar bem nítido e vamos aumentar o tamanho, que tem um parâmetro aqui chamado outlier.size, vamos utilizar o tamanho 3.5 e vamos salvar essas alterações na nossa conhecida variável p.

[5:57] Então vamos salvar tudo isso na variável p, que é uma variável temporária que utilizamos para fazer alterações nos gráficos.

[6:10] Vamos executar novamente. Demora um pouquinho, executou, vamos dar o zoom.

[6:21] Os outliers do estado de Santa Catarina mudaram de cor, agora estão vermelhos e ficaram maiores, ficando mais nítidos, mais perceptíveis.

[6:31] Agora vamos fazer as alterações que já fizemos repetidas vezes até aqui, que é alterar o título, o rótulo e o layout do gráfico.

[6:43] Então vamos aqui inserir xlab UF Prova, ylab vamos colocar nota redação e, por fim, vamos alterar o theme, themebw, e vamos salvar aqui dentro do objeto p mesmo. Vamos executar.

[7:19] Alteramos o layout e a alteração.

[7:23] Porém, as cores não estão muito profissionais.

[7:26] Essas cores do R, automáticas, elas facilitam bastante, mas às vezes não são agradáveis de serem visualizadas.

[7:33] Então vamos alterá-las deixando a visualização mais suave e mais profissional.

[7:42] Alterando os rótulos da legenda, indicando se tem filial ou não, porque está false e true, né. Então esses rótulos da legenda também não estão muito bons.

[7:52] Vamos fazer essas alterações.

[7:52] Vamos alterar primeiro a cor, scale_fill_manual, ou seja, vou definir a cor de preenchimento manualmente, vamos deixar o título da legenda filial em branco, values é a lista de cores, vamos inserir aqui chocolate3 e outra cor chamada chartreuse3.

[8:32] E agora vamos alterar o rótulo da legenda, vamos trocar esse false e true por nomes mais intuitivos.

[8:44] Vamos utilizar o parâmetro labels, vamos passar uma lista, um vetor sem filial, que é o do false, que vai trocar para sem filial e o true que vai trocar para com filial.

[9:00] Vamos salvar tudo isso em p, novamente, vamos executar. Pronto, vamos dar um zoom aqui.

[9:19] O nosso gráfico está ponto, com um layout mais profissional, os estados que tem filial destacados em verde, você pode ver aqui que está bem mais fácil de visualizar, e os outliers em vermelho, também aqui mais perceptíveis e a legenda está mais intuitiva, ou seja, o marrom é sem filial e o verde é com filial.

[9:38] Agora vamos realizar a nossa análise com base nesse gráfico.