

Combinando gráficos barras e limite erro

Transcrição

[0:01] Ainda analisando as notas das redações o cursinho deseja analisar as médias das redações por UF, comparando com a média nacional, inclusive dos estados onde o cursinho tem filiais.

[0:12] Ele quer um destaque maior para esses estados.

[0:15] Essa informação é útil para o cursinho ter conhecimento de como está a diferença da média estadual em relação à nacional.

[0:25] E essa informação poderá complementar a análise feita anteriormente com o boxpot, que também foi feita com base na nota de redação.

[0:35] Então, auxiliando os gestores do cursinho, eles podem traçar melhores metodologias de estudo para os alunos melhorarem o desempenho na matéria de redação.

[0:44] Para isso, vamos gerar esse gráfico com informações: a média local, ou estadual, por UF e a média nacional, que possibilite visualizar e comparar facilmente essas duas informações. Vamos lá então.

[0:58] Primeiro, vamos selecionar os dados que nós queremos e calcular a média nacional, a geral e depois uma média por estado.

[1:11] Como é que vamos fazer isso? Vamos chamar aqui o `enem`, filter UF Prova, esse comando você já conhece bem, né, estou eliminando os registros em branco da UF Prova, colocar um `and is.na` da nota redação, vamos concatenar outra função aqui agora, vamos utilizar um `mutate`, ou seja, inserir uma nova coluna, com a coluna média nacional.

[1:43] Estou criando uma coluna chamada média nacional, minha, nota redação.

[1:53] Estou calculando a média nacional para a redação.

[1:56] Vamos fazer um `group_by`, ou seja, nós vamos calcular agora a média por UF, então UF Prova e temos que colocar aqui a média nacional para manter essa coluna também, então vamos fazer um agrupamento UF Prova e a média nacional.

[2:14] Vamos agrupar com duas colunas, `summarise`, vamos criar um novo conjunto de dados, média UF.

[2:26] E vai ser a coluna média UF, `mean` nota redação.

[2:33] Vamos salvar esse procedimento em média redação, vamos executar, vamos dar um `view`. Está aqui.

[2:48] Temos a coluna UF Prova, média nacional que vai ser igual para todos os registros, mas nós precisamos dessa informação aqui, que é um dos métodos de fazer e a média para a respectiva UF.

[2:55] Então, ao gerar esse novo conjunto de dados, uma coluna chamada média nacional, que ela tem aqui, e média UF, agora uma das opções para atender essa demanda, ou seja, gerar um gráfico para atender essas duas informações, média nacional e média UF, é gerar um gráfico de barras com linhas de erro, que é chamada linhas de erro neste caso.

[3:20] Vamos criar o gráfico de barras primeiro. Vamos aqui `ggplot`, `data`, média, opa, média redação, mapear o eixo x vai receber, vamos ordenar, `reorder` UF Prova, esse procedimento você já conhece bem, que fizemos isso no primeiro curso e

também já fizemos aqui nesse curso, média UF, o y que vai receber média UF e vamos inserir aqui `geom_bar`, `stat`.

[4:14] Vamos executar. Pronto. Geramos aqui o gráfico de barras, vamos inserir a função `coord_flip` para rotacionar e virar um gráfico de barras horizontais.

[4:28] Pronto, ordenadas porque a gente ordenou aqui com o `reorder`.

[4:35] E você pode observar uma coisa diferente que eu fiz aqui, eu mapeei o eixo x e y dentro do `ggplot`.

[4:42] Essa é uma das formas de fazer o gráfico, não há mudança. Porém para algumas situações, há a necessidade de mapear aqui dentro, quando vamos acessar os dados, fazer algumas alterações mais bruscas, então eu estou só mostrando para você aqui mais uma forma de mostrar para você como é que faz.

[5:00] Agora vamos fazer, vamos inserir a barra, a linha de erro, que é chamar `geom_erro_bar`, vamos mapear, o mesmo esquema, `aes`, só que ele recebe o parâmetro `ymin`, que vai ser a média nacional e o `ymax`, que vai receber a média nacional também.

[5:32] Porque esse é um recurso de erro de valor máximo e mínimo.

[5:40] E não é o nosso caso, no nosso caso queremos só indicar de forma mais fácil a média nacional. `Size 1`. Vamos executar. Pronto, o gráfico foi gerado.

[5:54] BOM, você pode observar aqui no canto direito que foi inserido uma linha junto ao gráfico de barras, porém estamos utilizando o mesmo valor de `ymin` e `ymax`.

[6:06] Essa linha representa o valor da média nacional e cada barra representa a média por UF da prova.

[6:15] Então vamos organizar essa linha aqui, porque ela está um pouco estranha, não está visualmente agradável de se ver, e esse problema aconteceu lembrando que colocamos um mesmo valor `ymin` e `ymax`.

[6:37] Vamos simplesmente dividir o `ymin` por 2, você pode fazer aqui, vamos executar esse código, pronto. Vamos dar um zoom aqui.

[6:46] Agora a linha está um pouco maior.

[6:47] Então você pode observar aqui que tem uma margem de diferença entre a média local, que são as barras comparando com a média nacional que são as linhas. Fica mais fácil de visualizar essas duas médias, certo?

[7:04] Porém, há um problema ainda nesse gráfico: um problema de design, visualização.

[7:07] Essa linha que indica a média nacional, você pode ver aqui, ela ficaria melhor por baixo da barra, então ela ficaria visível somente essa parte aqui da diferença.

[7:20] Para isso, vamos fazer uma alteração aqui no nosso código, você pode copiar esse código novamente, basta você inverter a ordem de execução do `geom_bar` com o `geom_erro_bar`.

[7:35] Você vai inverter. Aqui no código de cima, primeiro era executado o `geom_bar` e depois o `geom_erro_bar`. Então nós vamos inverter aqui e executar primeiro o `geom_erro_bar` e depois o `geom_bar`.

[7:51] E vamos salvar nosso gráfico dentro do `plo bar erro`. Vamos executar.

[8:10] O código foi executado. Vamos dar um zoom aqui.

[8:18] Agora sim temos um gráfico mais bonito de se olhar.

[8:24] Agora temos a média local em comparação com a média nacional. É possível observar que aproximadamente metade dos estados tem a média acima da média nacional aqui em cima ó e poucos estados estão com a média muito abaixo da média nacional.

[8:47] Mas não é fácil identificar quais estados que mais interessam o cursinho, ou seja, onde o cursinho tem filiais.

[8:57] Vamos fazer mais algumas alterações para facilitar e melhorar a nossa análise e deixar nosso gráfico mais profissional, mais bonito e principalmente, mais fácil de ser interpretado.