

Gráfico variáveis uf_prova e situacao_conclusao

Transcrição

[0:00] Bom, a próxima solicitação do cursinho é saber sobre a situação de escolaridade das pessoas que estão fazendo a prova do Enem por estado, ou seja, UF.

[0:10] Então, vamos primeiro fazer um plot dos estados, vamos aqui utilizar a função gplot que vocês já conhecem bem, passar um data Enem, o concatenado da função ggplot, que é o mais. Geom_bar, aes, que mapeia o eixo xy. O eixo x vai receber os valores da coluna UF Prova e o stat, como sempre eu usei até agora, o count. Vamos gerar o gráfico.

[0:54] Bem, podemos observar aqui, como eu já tinha alertado você, um valor em branco que a gente identificou no início, porém decidimos também não eliminar esse valor para não interferir em outras análises.

[1:10] Você pode observar também que aqui no eixo y, tem um valor agora bem diferente do que fazemos normalmente. Isso se chama notação científica. Para eliminar essa notação científica, você vai utilizar a função options passando como parâmetro agora scipen 9999.

[1:42] Vamos executar. Vamos executar novamente esse gráfico. Pronto. A notação científica já sumiu. Correto?

[1:54] Agora vamos voltar ao nosso problema aqui que é esse valor em branco aqui da coluna UF Prova. Vamos filtrar aqui só para essa análise. Vamos utilizar a função dplyr, que já é bem conhecida, o filter, que nós utilizamos tanto nesse curso como no anterior, passando UF Prova diferente de vazio, ou seja, duas aspas com vazio. Porém, vamos fazer um novo select, criar um novo conjunto de dados com apenas alguns campos que nós desejamos.

[2:35] Esses campos são: UF Prova, Situação Conclusão. E vamos salvar esses novos registros dentro do novo objeto chamado UF Prova. Vamos alinhar aqui, deixar tudo alinhado. Vamos executar. Pronto, nós temos aqui um novo conjunto de dados UF Prova, eliminando aqui os registros em branco da UF Prova e apenas com 2 variáveis.

[3:17] Agora vamos gerar esse gráfico novamente, porém utilizando esse novo conjunto de dados UF Prova. ggplot passando como data UF Prova, já falei para você, utilizando a função geom_bar, aes x e UF Prova. Olha só: um erro aconteceu.

[3:47] Ao executar esse comando, aconteceu um erro que devemos ficar muito atentos. O geom_bar ao invés de inserir o nome da coluna UF Prova, ele inseriu uf prova em minúsculo.

[4:05] Ou seja, nós inserimos esse novo conjunto de dados que nós criamos e lembrando que o R é case sensitive, ele diferencia as letras maiúsculas das minúsculas. Isso aconteceu porque quando a gente estava apertando UF Prova e ele já auto complementa, se você aperta o enter sem querer, ele vai auto complementar.

[4:24] Então corrigido esse erro, vamos colocar o nome da coluna aqui certinha, que é UF Prova, você olhar aqui ó: UF Prova em letra maiúscula e SITUAÇÃO CONCLUSÃO também em maiúsculas.

[4:41] Vamos aqui continuar, porém vamos já complementar nosso gráfico fazendo a diferenciação da situação da conclusão. Vamos fazer igual fizemos no gráfico anterior, utilizando o parâmetro fill, que esse parâmetro vai criar, vai setar, como você percebeu, cores distintas para cada valor. Situação conclusão. O R define as cores automaticamente para cada valor distinto dessa coluna.

[5:16] Olha só: o gráfico foi gerado. Vamos dar um zoom aqui nesse gráfico para visualizar melhor, vamos aumentar aqui. Como anteriormente, foi gerado um gráfico com barras sobrepostas, ou seja, uma barra em cima da outra, que dependendo

da situação, a visualização fica péssima, como é esse caso aqui. Para melhorar, o recurso é o mesmo, vamos utilizar lá a função parâmetro `position_dodge`, que é uma função, não é apenas uma função de string.

[6:07] É uma função que vai separar as barras lado a lado. Vamos colocar aqui, vamos executar, vamos dar um zoom novamente aqui, vamos aqui, pronto. Agora a visualização está um pouquinho melhor, correto?

[6:31] Porém, há uma forma de melhorar mais ainda esse gráfico, por exemplo, a gente não consegue identificar por exemplo, as quantidades menores, intermediárias, ela está muito ruim de visualização. Conseguimos identificar apenas os valores em vermelho que significa concluído, que estão em mais destaque.

[6:51] Porém, outros valores, da situação de conclusão, por exemplo o roxinho, não concluído e não cursando já fica mais difícil de diferenciar entre os estados. E a ideia dos gráficos é sempre, o objetivo dos gráficos é sempre facilitar a visualização e análise de dados.

[7:12] Então vamos aqui utilizar o novo recurso que é o chamado `facet_grid`, já vou mostrar aqui para você. Vamos acrescentar aqui um mais e vamos chamar a função `facet_grid`. Essa função vai receber como parâmetro a coluna que você deseja fazer a faceta dela, que é a situação conclusão. Depois tem que colocar um tiozinho e um ponto.

[7:41] O que isso significa? Que ela vai fazer a divisão das janelas, você vai ver já na prática, dessa coluna até ela. Não vai ter mais outra coluna. Significa que esse aqui é o final. Então vai pegar somente essa coluna, vai dividir de acordo com o valor dela as barras e a janela completa. Vamos executar aqui, pronto.

[8:11] Assim, essa função `facet_grid` faz essa divisão aqui, você pode ver que cada, temos 4 gráficos em 1 só, divididos por situação da conclusão. Então ele pegou os valores distintos dessa coluna aqui e dividiu em outros gráficos. Vamos dar um zoom para melhorar, vamos aumentar, pronto.

[8:37] Agora é possível visualizar cada situação separadamente, em cada gráfico por estado. Ou seja, eu tenho o eixo x aqui que continua por estado, porém eu tenho uma divisão mais nítida aqui da divisão da situação da conclusão.

[8:55] Podemos observar que o estado com mais concluído, em vermelho, aqui no primeiro gráfico, aqui em cima, é São Paulo. E o estado com não concluído e que não está cursando, que é a cor roxa, o estado com maior, está um valor bem equilibrado entre São Paulo e Rio Grande do Sul. Mas vamos dar uma melhorada nesse gráfico ainda antes de apresentar ao nosso cliente.

[9:28] Pronto, agora que o gráfico está finalizado, já dava para ter feito a visualização, vamos salvar esse gráfico num novo objeto chamado `plot_uf_conclusão`, igual está aqui, vamos copiar para cima, atribuir, vamos alinhar e executar. Pronto. Nós temos o gráfico já salvo nesse objeto e agora vamos fazer as devidas modificações.