

## Mudando parâmetros do gráfico

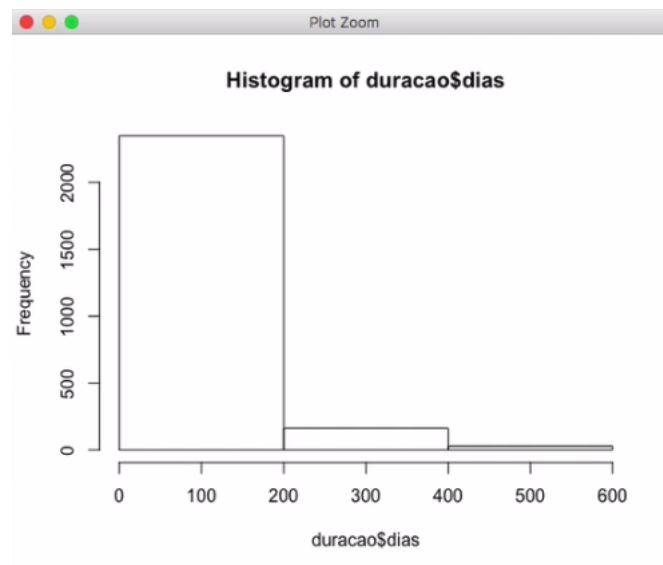
### Transcrição

Produzimos um histograma proveniente da configuração padrão do RStudio. O número *default* de suas quebras pode ou não ser apropriado para as informações trabalhadas no gráfico. O programa analisa os dados disponíveis e escolhe um número de quebras. Muitas vezes ele é bom, mas às vezes precisamos alterar o parâmetro para criarmos um gráfico mais adequado.

Para informarmos ao RStudio que queremos alterar a quantidade de barras, modificaremos o código `hist`, utilizando o parâmetro `breaks`. Nele, estabelecemos o número de quebra que queremos, no caso `3`.

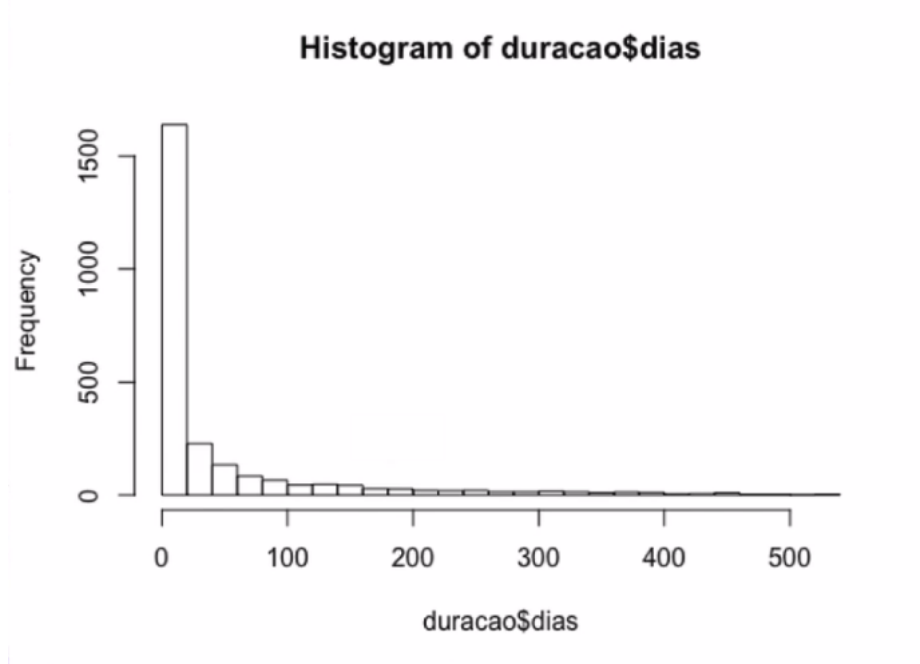
```
hist(duracao$dias, breaks = 3)
```

Assim, definimos que queremos um histograma com três barras. Vamos executar para ver o resultado. Aplicado o zoom, teremos a seguinte imagem:



De acordo com o que gostaríamos, observamos uma agregação maior de dados, em um número menor de possibilidades. De 0 a 200 dias, temos a conclusão da maioria dos cursos, mais de 2000. Há uma fatia de cursos que duram de 200 a 400 dias e uma minoria — ou *outlier*, termo utilizado para casos discrepantes — que dura de 400 a 600 dias.

Essa é uma maneira de informarmos ao cliente quanto tempo os alunos levam para concluir os cursos. Porém, podemos extrair mais informações a partir dos mesmos dados. No mesmo parâmetro que adicionamos anteriormente, alteraremos as 3 quebras para 20. Após executarmos o comando, aplicaremos zoom e expandiremos a janela, e o histograma ficará da seguinte forma:



O gráfico ficou bem diferente em relação ao anterior. O histograma com 20 barras é mais detalhado em relação ao número de dias que os alunos levam até a conclusão dos cursos. A frequência, referente à quantidade de vezes que o número de dias aparece no banco de dados, se mantém no eixo vertical.

Percebemos que a agregação de 0 a 200 no eixo horizontal poderia ocultar muitas informações, pois há uma concentração dos cursos concluídos entre 0 e 20 dias, tornando a primeira barra dominante em relação às outras. À medida em que `duracao$dias` aumenta, a frequência diminui.

De 200 a 500 dias a frequência se torna ínfima no histograma. Poderíamos agregá-la em uma barra de 200 dias para cima. Com isso, poderemos considerar casos discrepantes. As categorias de durações relevantes seriam de 0 a 20, de 20 a 100, de 100 a 200 e, por fim, mais de 200 dias.

Essa seria outra forma de informarmos a empresa sobre a duração dos cursos. Eventualmente, na apresentação, poderemos mostrar os dois histogramas, sendo que o primeiro seria apresentado para demonstrar ao cliente os cursos que duram menos que 200 dias, e aqueles que duram mais de 200 dias. Ou poderíamos mostrar o segundo histograma, com interpretação em 3 partes, demonstrando os cursos que duram menos de 20 dias, os que duram de 20 a 200, e os que duram mais de 200 dias.

Ou seja, depende de como queremos apresentar os dados ao cliente; em um ou dois histogramas, o importante é passarmos de forma eficiente o máximo de informação a partir dos gráficos. Uma boa maneira de fazermos isso é por meio da organização das divisões nos parâmetros dos gráficos.