

08

Faça o que eu fiz na aula

Separar os dados para o teste e treino é de extrema importância. Embora o recomendado seja separar 70% dos dados para o treino e o restante para o teste, é possível modificar essa porcentagem de acordo com sua necessidade. Vamos dividir 50% e 50% (.5).

Lembre-se que para isso devemos remover os valores "na" do banco e utilizar o comando "createDataPartition" do pacote **caret**:

```
dengueSemNa <- na.omit(dengue) dataindex <- createDataPartition(dengueSemNa$NoYes, p= .5, list=FALSE)
```

Uma vez particionado, vamos criar os bancos separados:

```
dengueTreino <- dengueSemNa[dataindex,] dengueTeste <- dengueSemNa[-dataindex,]
```