

# Aprendizado de Máquinas (Machine Learning)

Douglas Rodrigues

Universidade Federal Fluminense

- Vamos carregar o banco de dados "faithful". Contém duas variáveis:
  - ① **eruption**: duração da erupção do gêiser.
  - ② **waiting**: tempo de espera para a próxima erupção.
- Vamos tentar prever a variável `waiting` em função da variável `eruption`.

- Vamos criar amostras TREINO e TESTE, com 50% dos dados para cada lado.  
> inTrain <- createDataPartition(faithful\$waiting, p=0.5,list=F)  
> training <- faithful[inTrain,]  
> testing <- faithful[-inTrain,]

- Vamos criar amostras TREINO e TESTE, com 50% dos dados para cada lado.  

```
> inTrain <- createDataPartition(faithful$waiting, p=0.5,list=F)  
> training <- faithful[inTrain,]  
> testing <- faithful[-inTrain,]
```
- Vamos plotar os dados para ter uma ideia do comportamento dos dados.  

```
> plot(waiting ~eruptions, data = training,  
ylab="Waiting", xlab="Eruption")
```

- Observe que há indícios de uma relação linear entre os dados. Vamos treinar o modelo para utilizar MODELOS LINEARES ("lm")

```
> modelFit<- train( waiting ~ eruptions,  
data=training,method="lm")
```

- Observe que há indícios de uma relação linear entre os dados. Vamos treinar o modelo para utilizar MODELOS LINEARES ("lm")

```
> modelFit<- train( waiting ~ eruptions,  
data=training,method="lm")
```

- Vamos observar o resultado.

```
> #Resumo do modelo
```

```
> summary(modelFit)
```

```
> #Observe que o erro é medido pelo RMSE
```

```
> modelFit
```

- Observe que `modelFit` é uma lista com 23 itens. Para chamar apenas o modelo estimado, basta digitar  
`> modelFit$finalModel`

- Observe que `modelFit` é uma lista com 23 itens. Para chamar apenas o modelo estimado, basta digitar  

```
> modelFit$finalModel
```
- Vamos desenhar a reta de regressão estimada.  

```
> abline(modelFit$finalModel, col="red", lwd=3)
```



- Vamos aplicar o modelo na nossa amostra TESTE.  
    `> prediction <- predict(modelFit, testing)`

- Vamos aplicar o modelo na nossa amostra TESTE.  
    `> prediction <- predict(modelFit, testing)`
- Agora, vamos estimar o erro *out of sample*.  
    #Calculando RMSE,  $R^2$  e MAE:  
    `> postResample(prediction, testing$waiting)`

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

