

```
#Importando os pacotes
```

```
if(!require(dplyr)) install.packages("dplyr")  
library(dplyr)
```

```
if(!require(car)) install.packages("car")  
library(car)
```

```
if(!require(RVAideMemoire)) install.packages("RVAideMemoire")  
library(RVAideMemoire)
```

```
if(!require(ggpubr)) install.packages("ggpubr")  
library(ggpubr)
```

```
if(!require(tidyverse)) install.packages("tidyverse")  
library(tidyverse)
```

```
if(!require(viridis)) install.packages("viridis")  
library(viridis)
```

```
#Importando o dataset
```

```
stats = read.csv('Turmas_Machine_Learning.csv')
```

```
glimpse(stats)
```

```
View(stats)
```

```
> glimpse(stats)
```

```
Rows: 7
```

```
Columns: 2
```

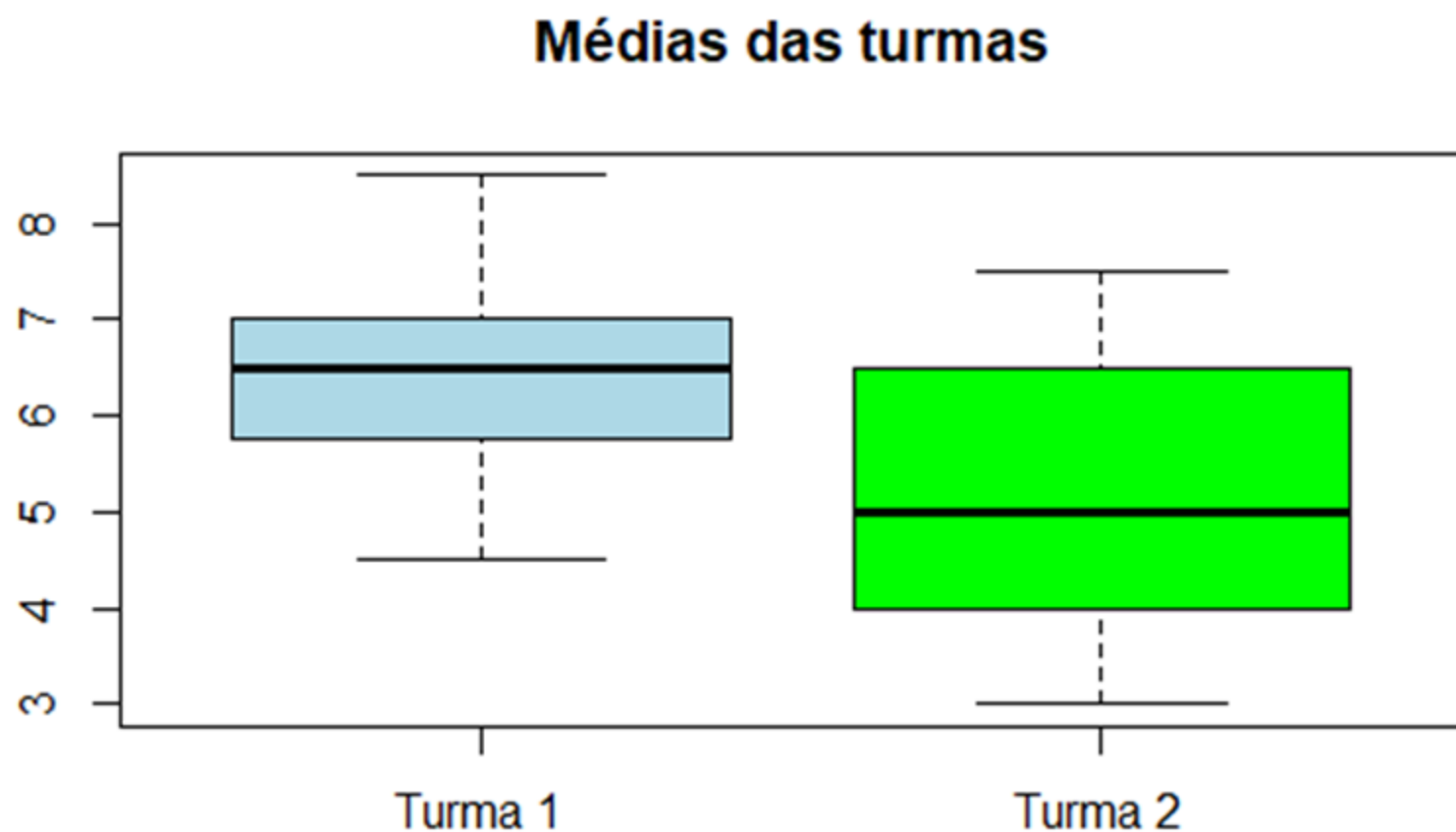
```
$ Turma._ML_01 <dbl> 6.5, 7.0, 5.0, 4.5, 8.5, 7.0, 6.5
```

```
$ Turma._ML_02 <dbl> 6.0, 7.0, 7.5, 3.0, 5.0, 4.0, 4.0
```

	Turma._ML_01	Turma._ML_02
1	6.5	6.0
2	7.0	7.0
3	5.0	7.5
4	4.5	3.0
5	8.5	5.0
6	7.0	4.0
7	6.5	4.0

```
#Visualização dos dados
```

```
boxplot(stats$Turma._ML_01, stats$Turma._ML_02, names=c('Turma 1', 'Turma 2'),  
        main='Médias das turmas', col=c('lightblue', 'green'))
```



```
#testar a normalidade dos dados com o teste shapiro-wilk  
shapiro.test(stats$Turma._ML_01)  
shapiro.test(stats$Turma._ML_02)
```

```
> shapiro.test(stats$Turma._ML_01)
```

Shapiro-Wilk normality test

```
data: stats$Turma._ML_01  
W = 0.93732, p-value = 0.6147
```

```
> shapiro.test(stats$Turma._ML_02)
```

Shapiro-Wilk normality test

```
data: stats$Turma._ML_02  
W = 0.94293, p-value = 0.6652
```

**AS AMOSTRAS TIVERAM VALOR DE P-VALUE = 0.6147 E 0.6652  
RESPECTIVAMENTE,  
COMO OS VALORES SÃO MAIORES QUE O NÍVEL DE SIGNIFICÂNCIA ADOTADO DE  
0.05 ENTÃO AS DISTRIBUIÇÕES SÃO NORMAIS.**

```
#Testar a homogeneidade de variâncias utilizando o teste de Levene
y = c(stats$Turma._ML_01, stats$Turma._ML_02)
grupo = as.factor (c(rep(1, length(stats$Turma._ML_01)),
                      rep(2, length(stats$Turma._ML_02))))
```

```
> leveneTest(y ~ grupo)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	1	0.8504	0.3746
	12		

**AO RODAR O CÓDIGO OBSERVOU-SE QUE O VALOR DE P-VALUE É  
MAIOR QUE 0.05, INDICANDO QUE NÃO HÁ DIFERENÇA NA  
HOMOGENEIDADE DAS VARIÂNCIAS**

#Aplicando o teste com nível de confiança de 95%

```
t.test(stats$Turma._ML_01, stats$Turma._ML_02, conf.level = 0.95)
```

Welch Two Sample t-test

data: stats\$Turma.\_ML\_01 and stats\$Turma.\_ML\_02

t = 1.4968, df = 11.423, p-value = 0.1616

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.5632842 2.9918556

sample estimates:

mean of x mean of y

6.428571 5.214286

**A HIPÓTESE ALTERNATIVA É QUE AS MÉDIAS DAS AMOSTRAS SÃO DIFERENTES. PODEMOS ENTÃO INFERIR QUE A HIPÓTESE NULA DO TESTE É QUE AS MÉDIAS DAS AMOSTRAS SÃO IGUAIS.**

**O TESTE REALIZADO FOI DO TIPO BICAUDAL, QUE É PADRÃO DO R**

## Conclusão final:

Estatisticamente falando, **não** há diferenças significativas entre as notas da turma 1 e da turma 2, apesar das médias serem diferentes, provavelmente se aumentasse o valor de participantes das turmas as médias tenderiam a se igualar.