

# Testes de Wilcoxon e Mann-Whitney

Erika Fialho

28 outubro, 2021

## Contents

<b>Pacotes</b>	<b>1</b>
<b>Teste Wilcoxon</b>	<b>1</b>
Exemplo 1: amostra pequena . . . . .	1
Exemplo 2: amostras grandes . . . . .	3
<b>Teste de Mann-Whitney</b>	<b>6</b>
Exemplo 1: amostra pequena . . . . .	6
Exemplo 2: amostra grande . . . . .	8

## Pacotes

```
library(rio)
library(tidyverse)
library(gridExtra)
library(ggplot2)
```

## Teste Wilcoxon

### Exemplo 1: amostra pequena

Os dados para este exemplo são referentes a colheitas de espigas de milho (em libras por acre) de dois diferentes tipos de sementes (normais e secas no forno) que foram usados em lotes adjacentes. O objetivo do estudo é testar se existe diferença entre a colheita utilizando ambas as sementes.

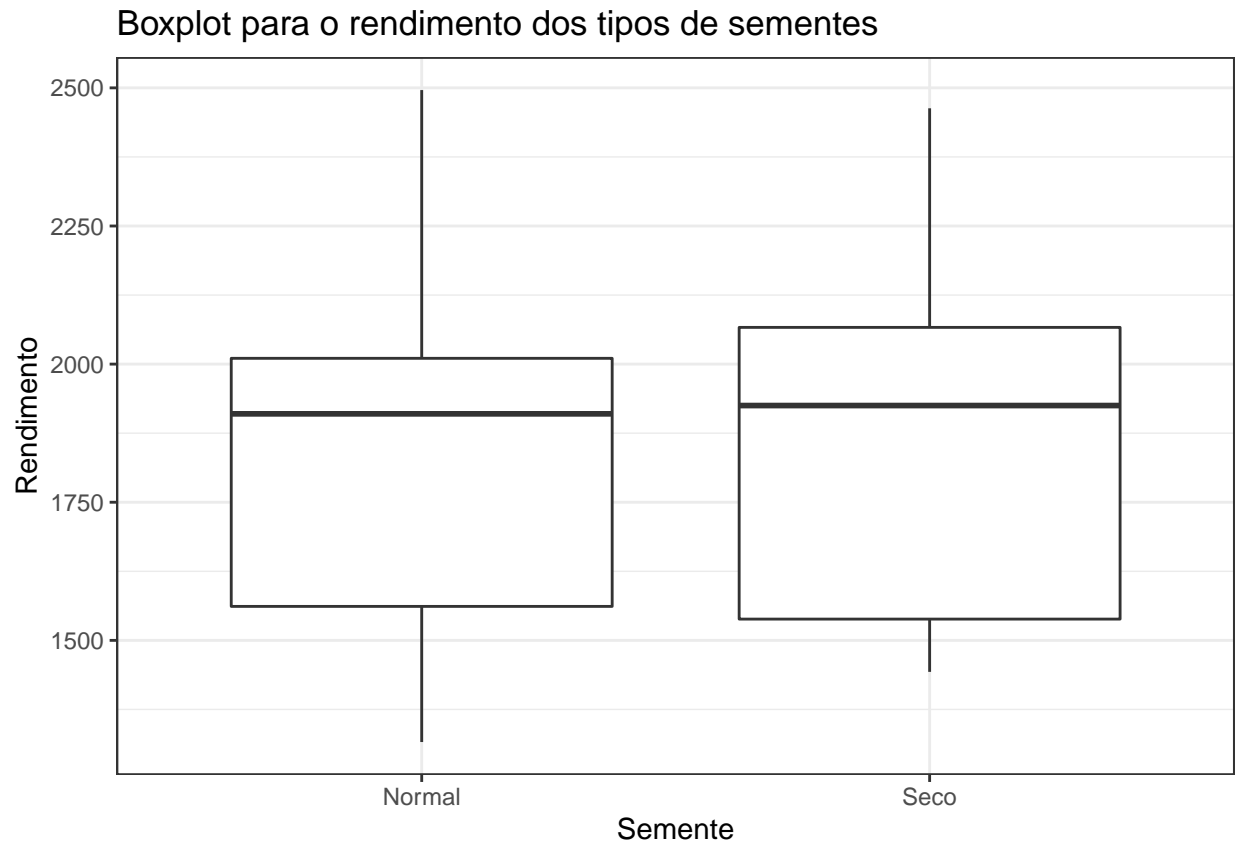
```
seed <- import(file = "http://lea.estadistica.ccet.ufrn.br/tutoriais/dados/seed.xls")
dados_seed <- dados <- data.frame(Semente = c(rep("Normal",11), rep("Seco",11)),
                                   Rendimento = c(seed$'Regular seed',seed$'Kiln-dried seed'))

Seca <- dados_seed[dados_seed$Semente == "Seco",]
Normal <- dados_seed[dados_seed$Semente == "Normal",]
completo <- data.frame(Normal,Seca)
completo
```

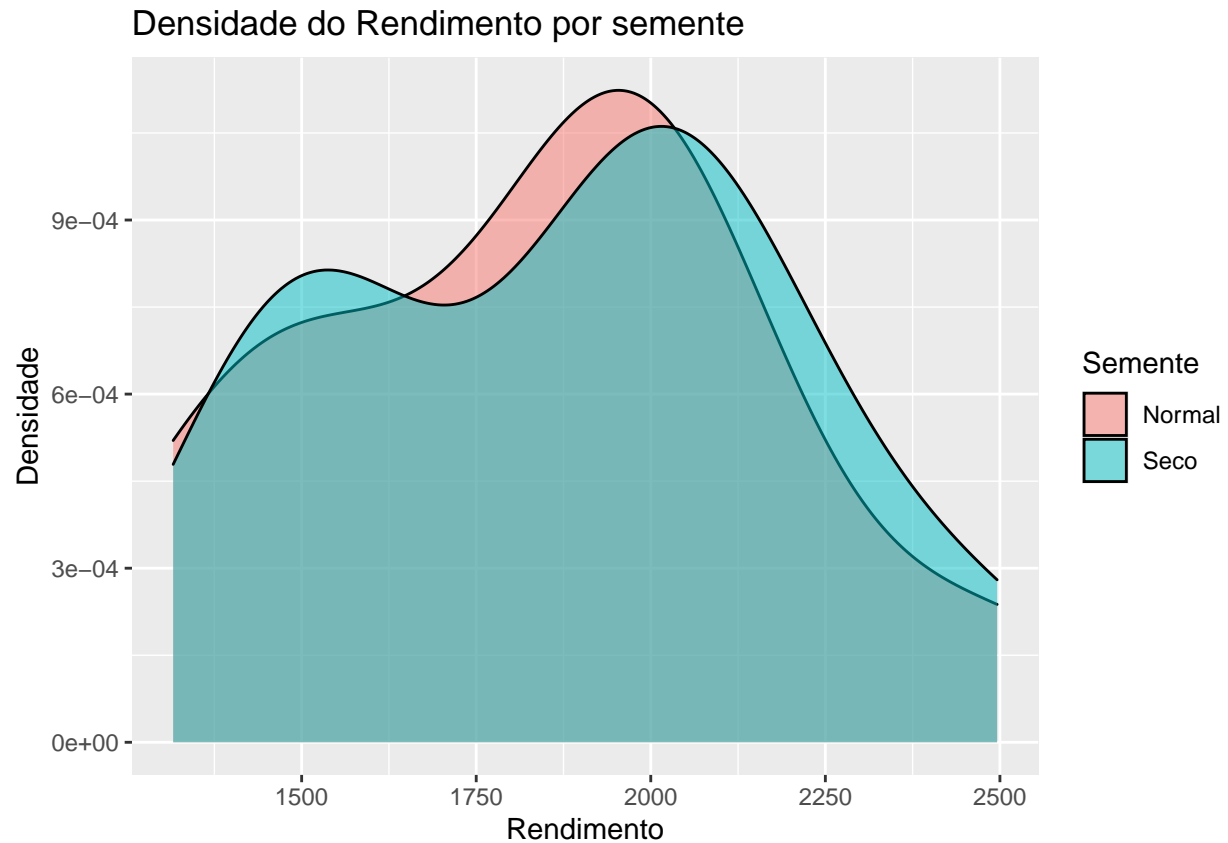
##	Semente	Rendimento	Semente.1	Rendimento.1
## 1	Normal	1903	Seco	2009
## 2	Normal	1935	Seco	1915
## 3	Normal	1910	Seco	2011
## 4	Normal	2496	Seco	2463
## 5	Normal	2108	Seco	2180
## 6	Normal	1961	Seco	1925
## 7	Normal	2060	Seco	2122
## 8	Normal	1444	Seco	1482
## 9	Normal	1612	Seco	1542
## 10	Normal	1316	Seco	1443
## 11	Normal	1511	Seco	1535

Investigação inicial

```
ggplot(dados_seed, aes(x = Semente, y = Rendimento)) + geom_boxplot() +
  labs(x = "Semente", y = "Rendimento", title =
    "Boxplot para o rendimento dos tipos de sementes") + theme_bw()
```



```
ggplot(dados_seed, aes(Rendimento, fill= Semente)) + geom_density(alpha=.5) +
  labs(title="Densidade do Rendimento por semente", x="Rendimento", y="Densidade")
```



Teste bilateral de Wilcoxon

```
wilcox.test(Normal$Rendimento, Seca$Rendimento, paired = T, alternative =
            "two.sided", conf.level = 0.95)
```

```
##
## Wilcoxon signed rank exact test
##
## data: Normal$Rendimento and Seca$Rendimento
## V = 15, p-value = 0.123
## alternative hypothesis: true location shift is not equal to 0
```

Portanto, não há evidências contra a hipótese  $H_0$ , não parece haver diferença estatisticamente significativa entre os tipos de sementes.

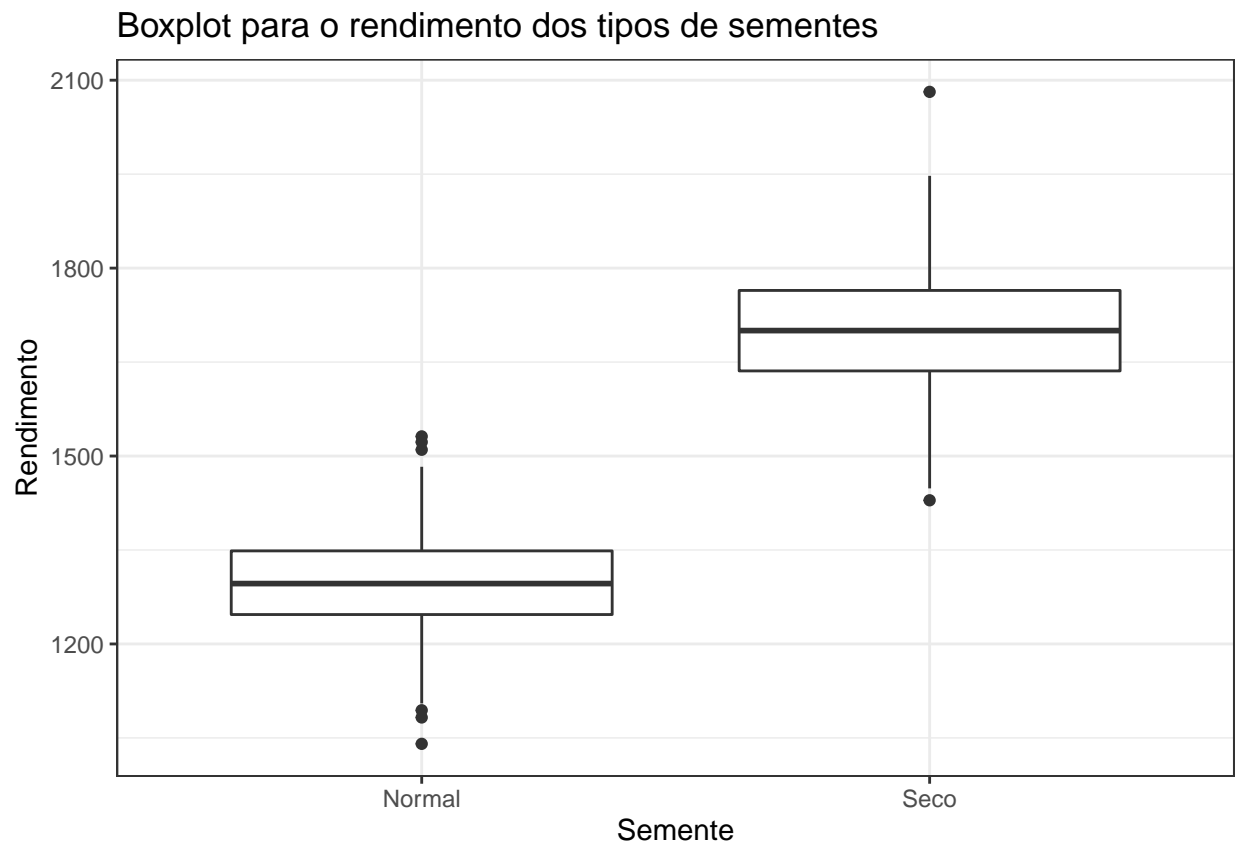
## Exemplo 2: amostras grandes

Vamos considerar o mesmo exemplo, agora com um banco de dados maior.

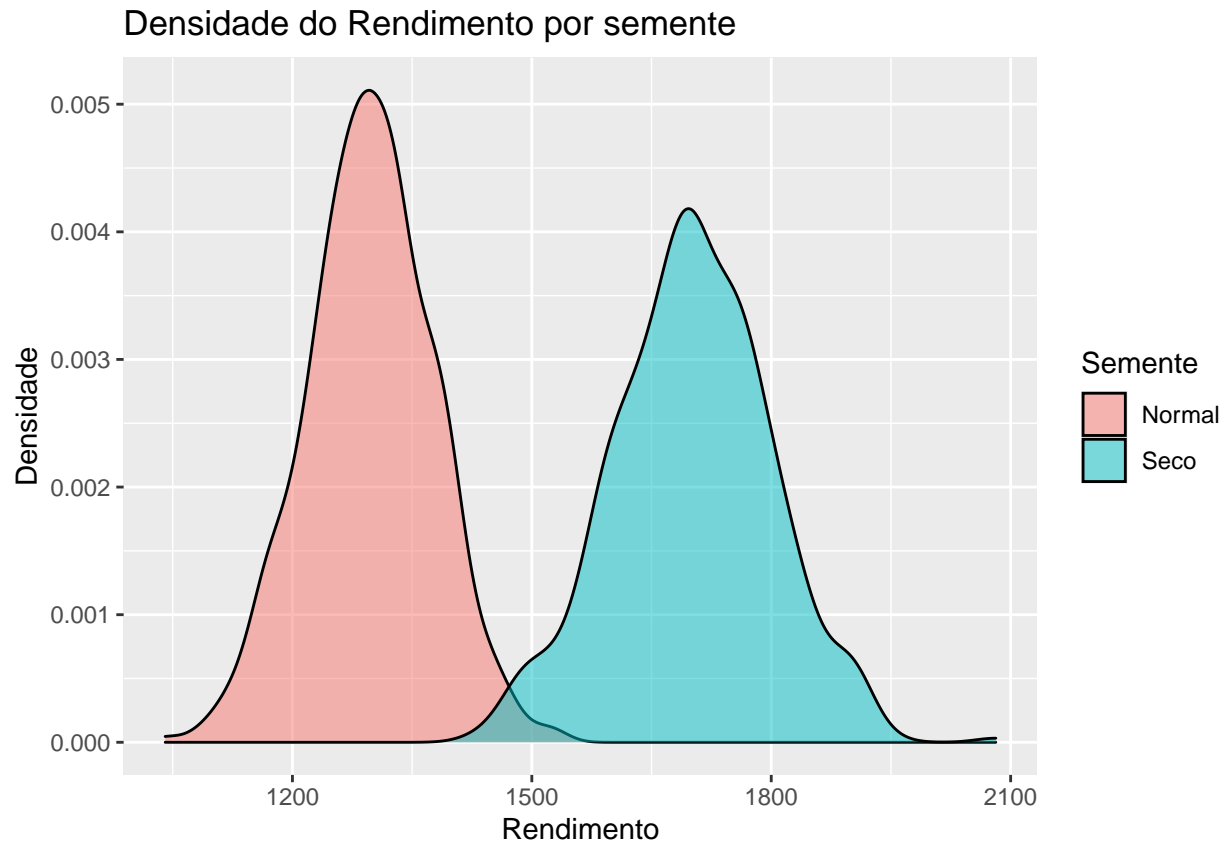
```
Rendimento1 <- rnorm(500, 1300, 80)
Rendimento2 <- rnorm(500, 1700, 100)
Seed <- data.frame(Semente = c(rep("Normal", 500), rep("Seco", 500)),
                  Rendimento = c(Rendimento1, Rendimento2))
```

Investigação inicial

```
ggplot(Seed, aes(x = Semente, y = Rendimento)) + geom_boxplot() +
  labs(x = "Semente", y = "Rendimento", title =
    "Boxplot para o rendimento dos tipos de sementes") + theme_bw()
```



```
ggplot(Seed, aes(Rendimento, fill= Semente)) + geom_density(alpha=.5) +
  labs(title="Densidade do Rendimento por semente", x="Rendimento", y="Densidade")
```



Teste unilateral de Wilcoxon

```
wilcox.test(Rendimento1, Rendimento2, paired = T, alternative =
            "two.sided", conf.level = 0.95)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: Rendimento1 and Rendimento2
## V = 0, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(Rendimento1, Rendimento2, paired = T, alternative =
            "greater", conf.level = 0.95)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: Rendimento1 and Rendimento2
## V = 0, p-value = 1
## alternative hypothesis: true location shift is greater than 0
```

Portanto, há evidências contra a hipótese  $H_0$ , pode-se dizer que parece haver diferença estatisticamente significativa entre os tipos de sementes. Além disso, com o teste unilateral, pode-se ainda dizer que o Rendimento2 (sementes secas) parece ser maior do que o Rendimento1 (sementes normais).

# Teste de Mann-Whitney

## Exemplo 1: amostra pequena

Uma pesquisa de mercado foi realizada em uma cidade com a finalidade de comparar duas marcas de refrigerante, A e B. Cada participante provou o refrigerante sem que soubesse qual das duas marcas estava provando e atribuiu uma nota entre 0 e 10.

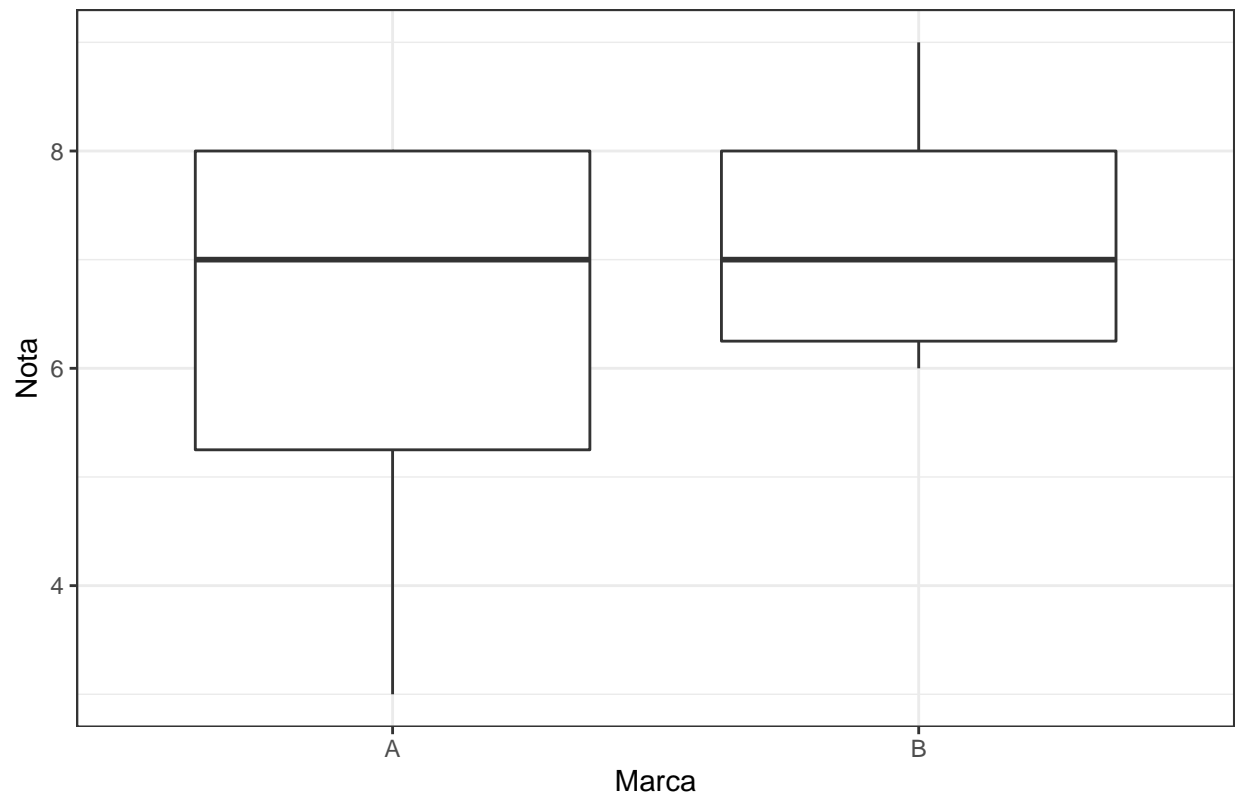
```
A <- c(5,7,8,8,4,8,6,7,3,8)
B <- c(9,6,8,7,6,9,7,7,8,6)
Refrigerantes <- data.frame(Marca = c(rep("A",10), rep("B",10)),
                             Nota = c(A, B))
Refrigerantes
```

##	Marca	Nota
## 1	A	5
## 2	A	7
## 3	A	8
## 4	A	8
## 5	A	4
## 6	A	8
## 7	A	6
## 8	A	7
## 9	A	3
## 10	A	8
## 11	B	9
## 12	B	6
## 13	B	8
## 14	B	7
## 15	B	6
## 16	B	9
## 17	B	7
## 18	B	7
## 19	B	8
## 20	B	6

Investigação inicial

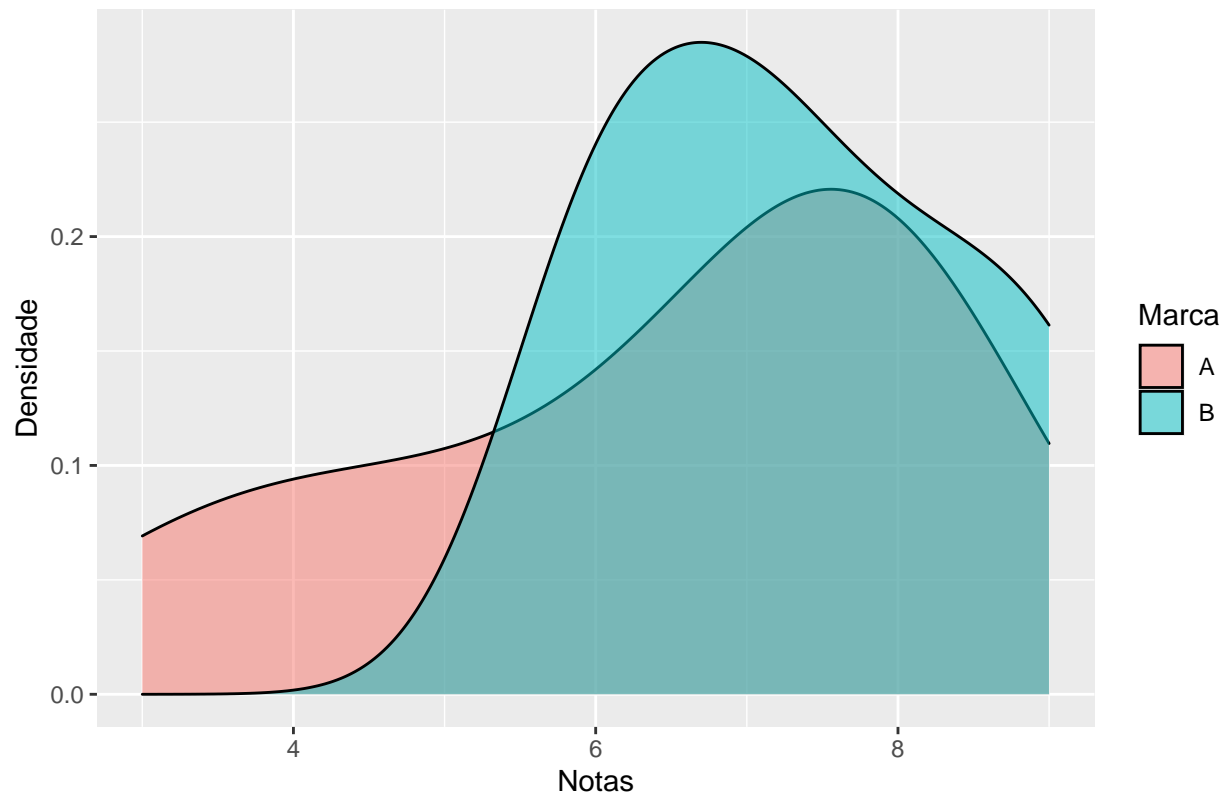
```
ggplot(Refrigerantes, aes(x = Marca, y = Nota)) + geom_boxplot() +
  labs(x = "Marca", y = "Nota", title =
    "Boxplot para as notas de cada refrigerante") + theme_bw()
```

Boxplot para as notas de cada refrigerante



```
ggplot(Refrigerantes, aes(Nota, fill= Marca)) + geom_density(alpha=.5) +  
  labs(title="Densidade das notas por Refrigerante", x="Notas", y="Densidade")
```

## Densidade das notas por Refrigerante



Teste bilateral de Wilcoxon

```
wilcox.test(A,B, paired = FALSE, alternative = "two.sided", conf.level = 0.95)
```

```
## Warning in wilcox.test.default(A, B, paired = FALSE, alternative =  
## "two.sided", : cannot compute exact p-value with ties
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: A and B  
## W = 38.5, p-value = 0.3937  
## alternative hypothesis: true location shift is not equal to 0
```

Portanto, não há evidências contra a hipótese  $H_0$ , não parece haver diferença estatisticamente significativa entre os tipos de refrigerantes.

## Exemplo 2: amostra grande

Novamente, vamos ampliar o exemplo e supor um conjunto de dados maior.

```
A2 <- rnorm(2000, 8, 0.5)  
B2 <- rnorm(2000, 7.5, 1)  
Refrigerantes2 <- data.frame(Marca = c(rep("A",2000), rep("B",2000)),
```

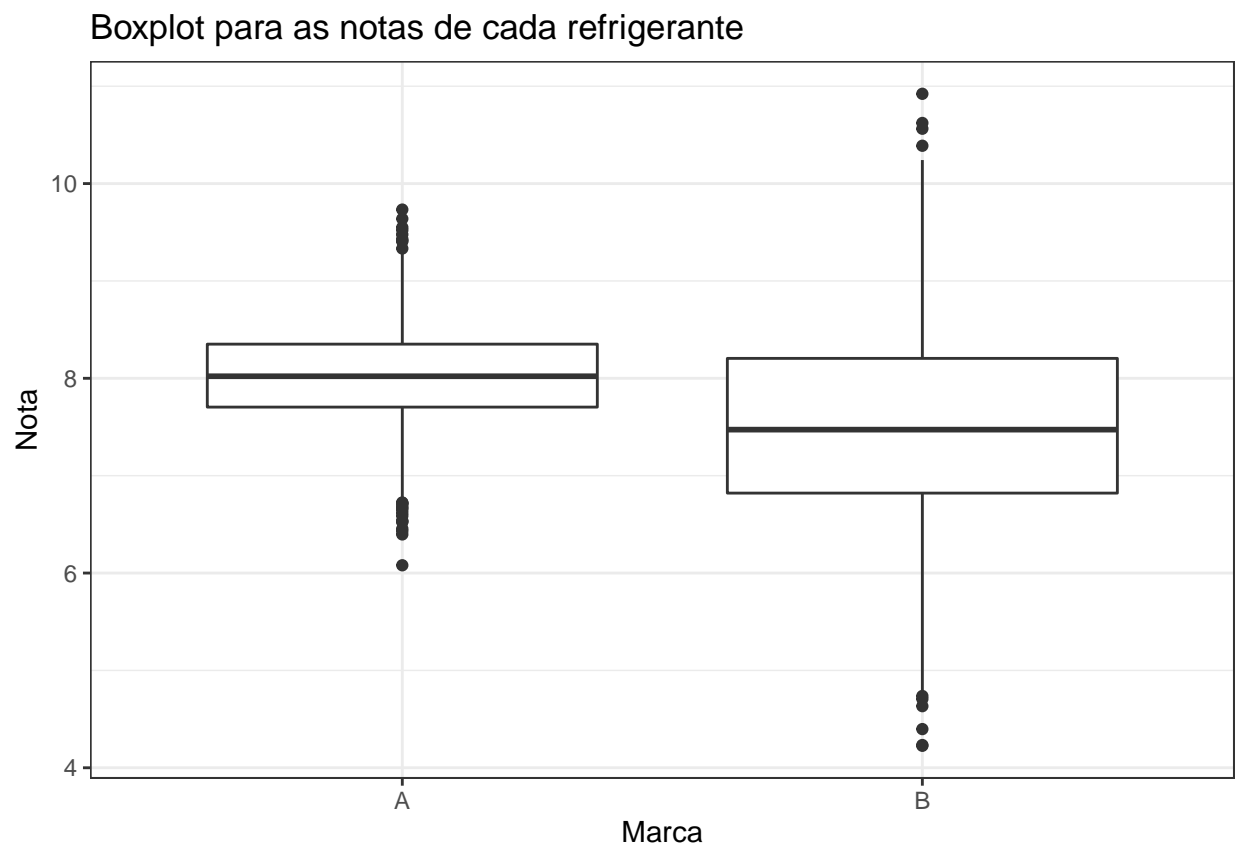


```
Nota = c(A2, B2))
Refrigerantes2[1:10,]
```

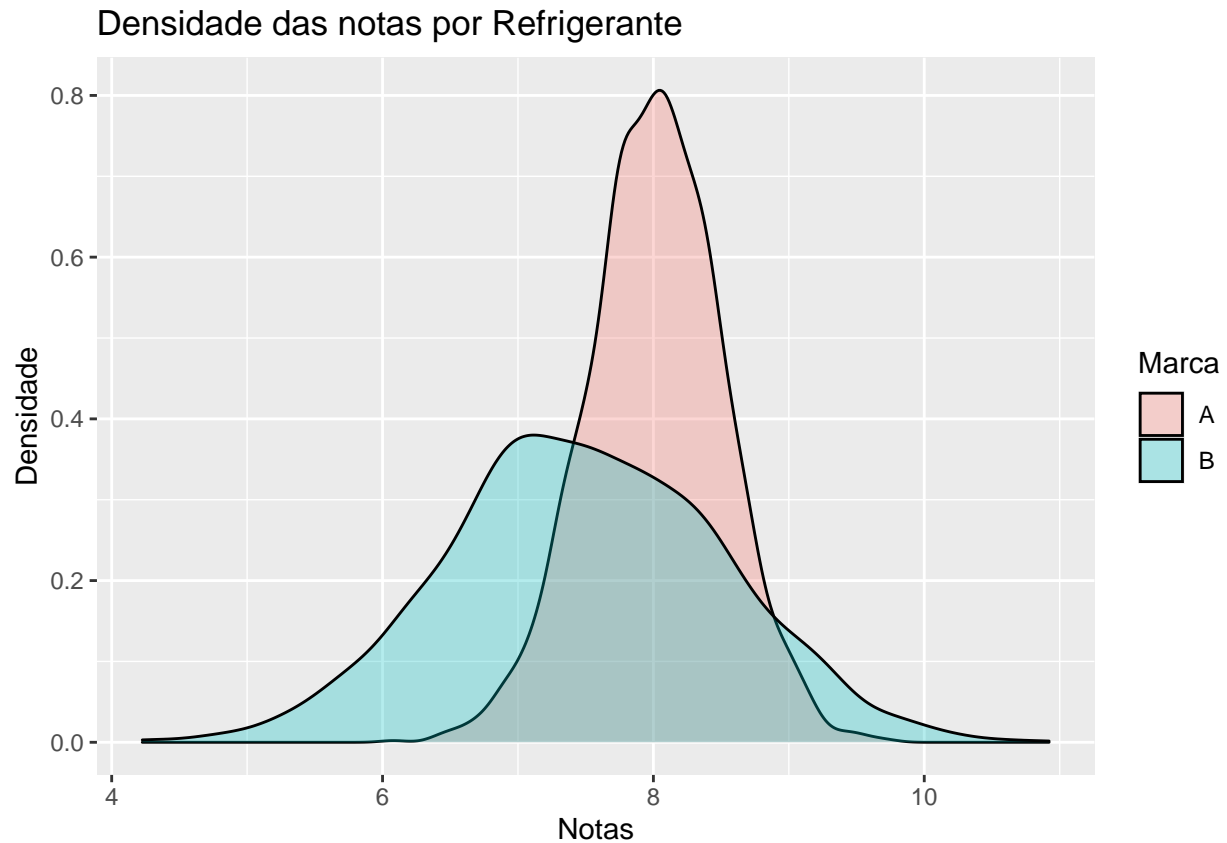
```
##   Marca   Nota
## 1     A 7.816249
## 2     A 8.458241
## 3     A 7.968606
## 4     A 8.681250
## 5     A 8.001267
## 6     A 7.087911
## 7     A 7.986208
## 8     A 8.248447
## 9     A 6.539190
## 10    A 8.375015
```

Investigação inicial

```
ggplot(Refrigerantes2, aes(x = Marca, y = Nota)) + geom_boxplot() +
  labs(x = "Marca", y = "Nota", title =
    "Boxplot para as notas de cada refrigerante") + theme_bw()
```



```
ggplot(Refrigerantes2, aes(Nota, fill= Marca)) + geom_density(alpha=.3) +
  labs(title="Densidade das notas por Refrigerante", x="Notas", y="Densidade")
```



Teste unilateral de Wilcoxon

```
wilcox.test(A2,B2, paired = FALSE, alternative = "two.sided", conf.level = 0.95)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: A2 and B2
## W = 2701630, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(A2, B2, paired = FALSE, alternative = "greater", conf.level = 0.95)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: A2 and B2
## W = 2701630, p-value < 2.2e-16
## alternative hypothesis: true location shift is greater than 0
```

Portanto, há evidências contra a hipótese  $H_0$ , pode-se dizer que parece haver diferença estatisticamente significativa entre os tipos de refrigerante. Além disso, com o teste unilateral, pode-se ainda dizer que a mediana das notas do Refrigerante A parece ser maior do que do Refrigerante B.