

```
pergunta1 > code.r
1 # INFO
2 ## 4898 amostras de vinho verde branco
3 ## variáveis 1 a 11 -> características físico-químicas dos vinhos
4 ## variáveis 12 -> avaliação feita por especialistas (1-má a 5-excelente)
5
6 # EXERCÍCIO
7 ## pacote ggplot2
8 ## produzir gráfico com box plots paralelos
9 ## como é que a raiz quadrada da variável citric.acid (variável 3) se relaciona com quality (variável 12)
10 ## realçar os outliers e tentar contrariar a sua sobreposição
11
12 # SUBMISSÃO
13 ## PDF com uma página A4 que inclua o código e o gráfico produzido
14
15
16 #pacote ggplot2
17 library(ggplot2)
18
19 #ler o ficheiro
20 dados <- read.csv("pergunta1/winequality-white-q5.csv", header = TRUE, sep = ",")
21
22 #adicionar uma nova coluna com a raiz quadrada do ácido cítrico
23 dados$sqrt_citric_acid <- sqrt(dados$citric.acid)
24
25 #criar o boxplot com pontos sobrepostos deslocados (jitter)
26 ggplot(dados, aes(x = factor(quality), y = sqrt_citric_acid)) +
27   geom_boxplot(outlier.shape = NA) +
28   geom_jitter(
29     color = "red",
30     width = 0.2,
31     alpha = 0.5,
32     size = 1.5,
33     data = subset(dados, sqrt_citric_acid %in% boxplot.stats(dados$sqrt_citric_acid)$out)
34   ) +
35   labs(
36     title = "Relation between the square root of citric.acid and the wine quality",
37     x = "Wine's quality",
38     y = "Square root of citric.acid"
39   ) +
40   theme_minimal()
41 ggsave("boxplot_vinho.png", width = 8, height = 5)
```

