

```
pergunta1 > code.r
18
19 #ler o ficheiro
20 wine_data <- read.csv("pergunta1/winequality-white-q5.csv")
21
22 #adicionar uma nova coluna com a raíz quadrada do ácido cítrico
23 get_boxplot_outliers <- function(data, x_var, y_var) {
24   outliers_df <- data.frame()
25   unique_x_values <- unique(data[[x_var]])
26   for (val in unique_x_values) {
27     subset_data <- data[data[[x_var]] == val, ]
28     y_values <- subset_data[[y_var]]
29
30     q1 <- quantile(y_values, 0.25, na.rm = TRUE)
31     q3 <- quantile(y_values, 0.75, na.rm = TRUE)
32     iqr <- q3 - q1
33     lower_bound <- q1 - 1.5 * iqr
34     upper_bound <- q3 + 1.5 * iqr
35
36     current_outliers <- subset_data[y_values < lower_bound | y_values > upper_bound, ]
37     if (nrow(current_outliers) > 0) {
38       outliers_df <- rbind(outliers_df, current_outliers)
39     }
40   }
41   return(outliers_df)
42 }
43
44 wine_data$sqrt_citric_acid <- sqrt(wine_data$citric_acid)
45 outliers_data <- get_boxplot_outliers(wine_data, "quality", "sqrt_citric_acid")
46
47
48 ggplot(wine_data, aes(x = factor(quality), y = sqrt(citric_acid))) +
49   geom_boxplot(outlier.shape = NA) +
50   geom_jitter(data = outliers_data,
51             aes(x = factor(quality), y = sqrt_citric_acid, color = "red"),
52             width = 0.2, height = 0, alpha = 0.7, size = 2) + #tirar sobreposição
53   labs(
54     title = "Relationship between Square Root of Citric Acid and Wine Quality",
55     x = "Wine Quality (1 = poor, 5 = excellent)",
56     y = "Square Root of Citric Acid",
57   ) +
58   theme_minimal()
59 ggsave("boxplot_vinho.png", width = 8, height = 5)
```

