DATAVIZ Task 3

Lucas Cardoso - 179737

29 September, 2025

Contents

```r
# Carregando arquivo
small_file <- read_delim("small_file.txt", delim = "\t")
```

```
## Rows: 40 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (2): Sample, Category
## dbl (1): Length
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
message("Dados de small_file.txt carregados com sucesso!")
```

```
## Dados de small_file.txt carregados com sucesso!
```

```r
# Inspeção
head(small_file)
```

```
## # A tibble: 6 x 3
##   Sample Length Category
##   <chr>   <dbl> <chr>
## 1 x_1        45 A
## 2 x_2        82 B
## 3 x_3        81 C
## 4 x_4        56 D
## 5 x_5        96 A
## 6 x_6        85 B
```

```r
glimpse(small_file)
```

```
## Rows: 40
## Columns: 3
## $ Sample   <chr> "x_1", "x_2", "x_3", "x_4", "x_5", "x_6", "x_7", "x_8", "x_9"~
## $ Length   <dbl> 45, 82, 81, 56, 96, 85, 65, 96, 60, 62, 80, 63, 50, 64, 43, 9~
## $ Category <chr> "A", "B", "C", "D", "A", "B", "C", "D", "A", "B", "C", "D", "~
```

```
# Filtrar Categoria D e ordenar
category_d <- small_file %>% filter(Category == "D") %>% arrange(Length)
category_d
```

```
## # A tibble: 10 x 3
##    Sample Length Category
##    <chr>   <dbl> <chr>
##  1 x_4        56 D
##  2 x_12       63 D
##  3 y_11       65 D
##  4 z_6        72 D
##  5 y_7        79 D
##  6 z_2        83 D
##  7 z_10       84 D
##  8 z_14       93 D
##  9 x_8        96 D
## 10 y_3        98 D
```

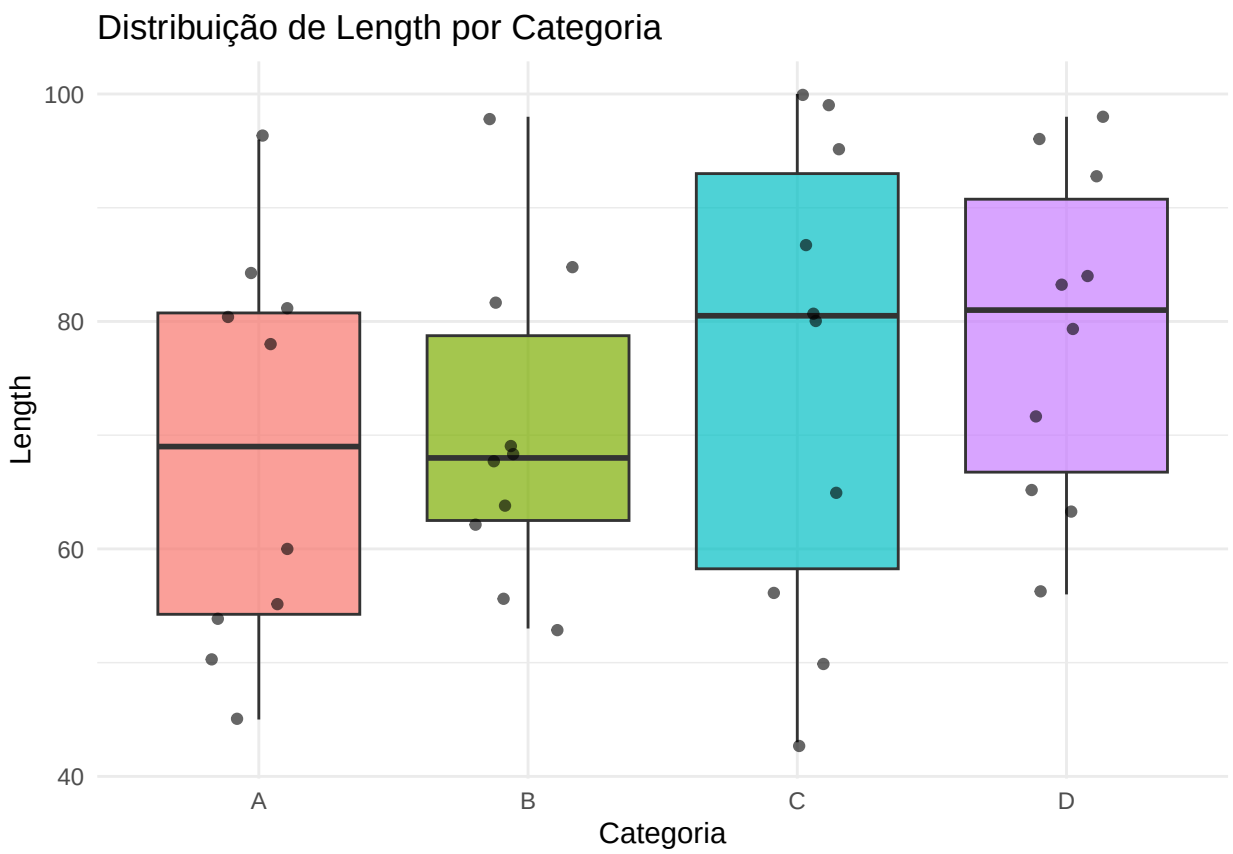
```
# Médias
mean_length_d <- mean(category_d$Length)
mean_length_a <- mean(small_file %>% filter(Category == "A") %>% pull(Length))
message("Média do Length para categoria D: ", round(mean_length_d, 2))
```

```
## Média do Length para categoria D: 78.9
```

```
message("Média do Length para categoria A: ", round(mean_length_a, 2))
```

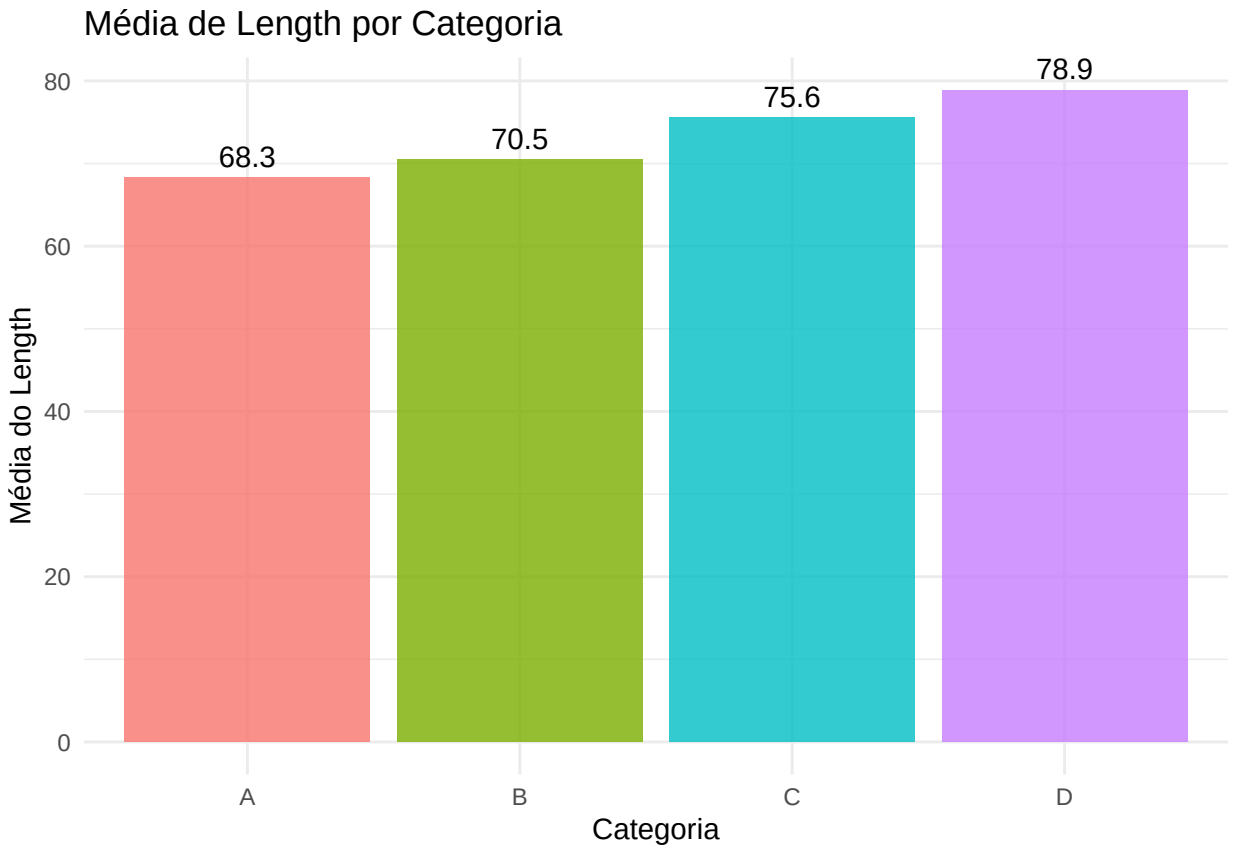
```
## Média do Length para categoria A: 68.3
```

```
ggplot(small_file, aes(x = Category, y = Length, fill = Category)) +
  geom_boxplot(alpha = 0.7) +
  geom_jitter(width = 0.2, alpha = 0.6) +
  labs(title = "Distribuição de Length por Categoria",
       x = "Categoria", y = "Length") +
  theme_minimal() +
  theme(legend.position = "none")
```




```
means_data <- small_file %>%
  group_by(Category) %>%
  summarise(mean_length = mean(Length), .groups = "drop")

ggplot(means_data, aes(x = Category, y = mean_length, fill = Category)) +
  geom_col(alpha = 0.8) +
  geom_text(aes(label = round(mean_length, 1)), vjust = -0.5) +
  labs(title = "Média de Length por Categoria",
       x = "Categoria", y = "Média do Length") +
  theme_minimal() +
  theme(legend.position = "none")
```




```
student_data <- read_csv("student_grade.csv")
```

```
## Rows: 43 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): Class, Student
## dbl (12): Year, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
message("Dados de student_grade.csv carregados com sucesso!")
```

```
## Dados de student_grade.csv carregados com sucesso!
```

```
head(student_data)
```

```
## # A tibble: 6 x 14
##    Year Class   Student      Q1    Q2    Q3    Q4    Q5    Q6    Q7    Q8    Q9
##   <dbl> <chr>   <chr>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2022 Student Lucca       7.5   6.23  6.5   7.15 NA     5.43  8.58  8.19  7.96
```

```
## 2  2022 Student Salles    10    10    10    10    10    NA    10    10    10
## 3  2022 Student Bueno      9.5   9     9     9.25  9.25  8     9.75  9.75  7.5
## 4  2022 Student Simas      9.5   9     9     9.25  9.25  8     9.75  9.75  7.5
## 5  2022 Student Goncalves  1.67  3.17  4.67  1.67  4     1.67  4.83  0.83  0.83
## 6  2022 Student Dornelles  9.1   8.75  9.83  9     9.75  9     9.5   9.25  9
## # i 2 more variables: Q10 <dbl>, Q11 <dbl>
```

```
dim(student_data)
```

```
## [1] 43 14
```

```
colnames(student_data)
```

```
##  [1] "Year"    "Class"   "Student" "Q1"      "Q2"      "Q3"      "Q4"
##  [8] "Q5"      "Q6"      "Q7"      "Q8"      "Q9"      "Q10"     "Q11"
```

```
student_tidy <- student_data %>%
  pivot_longer(cols = Q1:Q11, names_to = "Question", values_to = "Grade") %>%
  filter(!is.na(Grade)) %>%
  select(Year, Class, Student, Question, Grade)
head(student_tidy, 10)
```

```
## # A tibble: 10 x 5
##     Year Class   Student Question Grade
##    <dbl> <chr>   <chr>   <chr>    <dbl>
##  1  2022 Student Lucca   Q1        7.5
##  2  2022 Student Lucca   Q2        6.23
##  3  2022 Student Lucca   Q3        6.5
##  4  2022 Student Lucca   Q4        7.15
##  5  2022 Student Lucca   Q6        5.43
##  6  2022 Student Lucca   Q7        8.58
##  7  2022 Student Lucca   Q8        8.19
##  8  2022 Student Lucca   Q9        7.96
##  9  2022 Student Lucca   Q10       7.92
## 10  2022 Student Lucca   Q11       6.48
```

```
# Q1 e Q2
q1_q2_stats <- student_tidy %>%
  filter(Question %in% c("Q1", "Q2")) %>%
  group_by(Question) %>%
  summarise(mean_grade = mean(Grade), sd_grade = sd(Grade), n_students = n(), .groups="drop")
q1_q2_stats
```

```
## # A tibble: 2 x 4
##   Question mean_grade sd_grade n_students
```

```
##   <chr>          <dbl>    <dbl>      <int>
## 1 Q1              8.50     1.61         43
## 2 Q2              7.95     1.62         43
```

```
# Performance geral
student_performance <- student_tidy %>%
  group_by(Student) %>%
  summarise(mean_grade = mean(Grade), n_questions = n(), .groups="drop") %>%
  arrange(desc(mean_grade))
head(student_performance, 5)
```

```
## # A tibble: 5 x 3
##   Student   mean_grade n_questions
##   <chr>          <dbl>       <int>
## 1 Junior         10             10
## 2 Salles         10             10
## 3 Pedro           9.81          11
## 4 Gabriel         9.70          11
## 5 Francisca       9.64          11
```

```
tail(student_performance, 5)
```

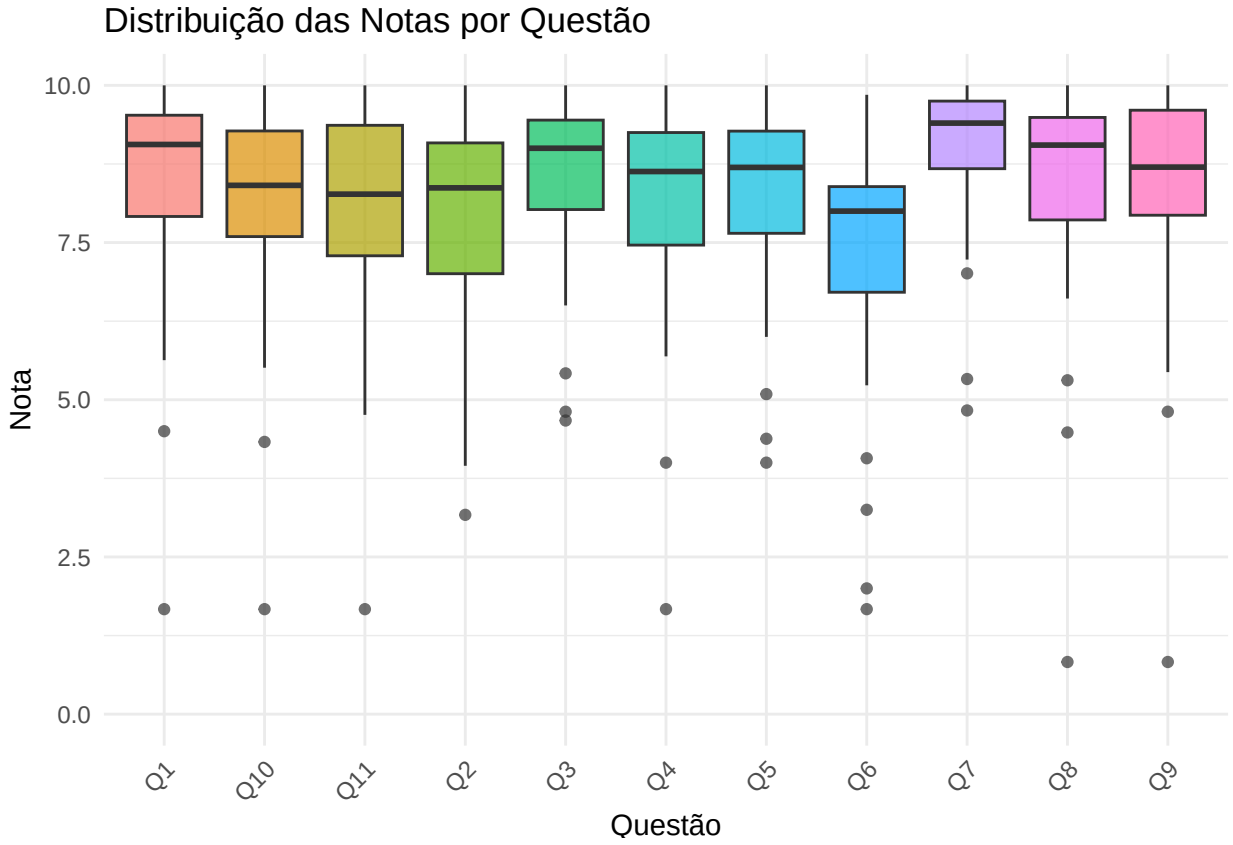
```
## # A tibble: 5 x 3
##   Student   mean_grade n_questions
##   <chr>          <dbl>       <int>
## 1 Rafaela         6.92          11
## 2 Ramos           6.13          11
## 3 Samara          5.48          11
## 4 Gleiser         4.42          11
## 5 Goncalves       2.43          11
```

```
# Dificuldade das questões
question_difficulty <- student_tidy %>%
  group_by(Question) %>%
  summarise(mean_grade = mean(Grade), sd_grade = sd(Grade), n_students = n(), .groups="drop") %
  arrange(mean_grade)
question_difficulty
```

```
## # A tibble: 11 x 4
##    Question mean_grade sd_grade n_students
##    <chr>          <dbl>    <dbl>      <int>
##  1 Q6              7.32     1.93         41
##  2 Q2              7.95     1.62         43
##  3 Q11             8.02     1.72         43
##  4 Q4              8.18     1.63         43
```

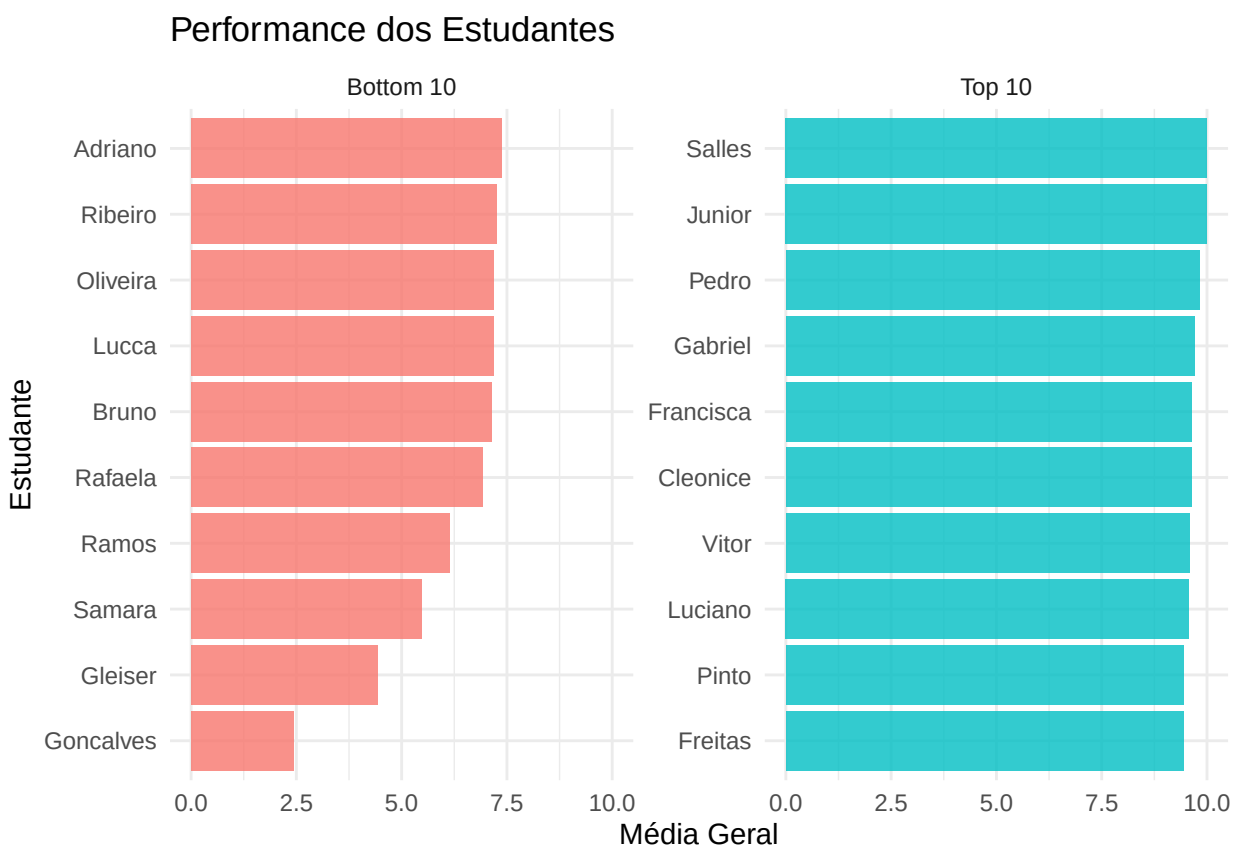
```
##  5 Q10            8.18     1.61         43
##  6 Q5             8.30     1.48         42
##  7 Q9             8.37     1.67         43
##  8 Q8             8.39     1.73         43
##  9 Q1             8.50     1.61         43
## 10 Q3             8.57     1.38         42
## 11 Q7             9.04     1.16         43
```

```
ggplot(student_tidy, aes(x = Question, y = Grade, fill = Question)) +
  geom_boxplot(alpha = 0.7) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position="none") +
  labs(title="Distribuição das Notas por Questão", x="Questão", y="Nota") +
  scale_y_continuous(limits=c(0,10))
```

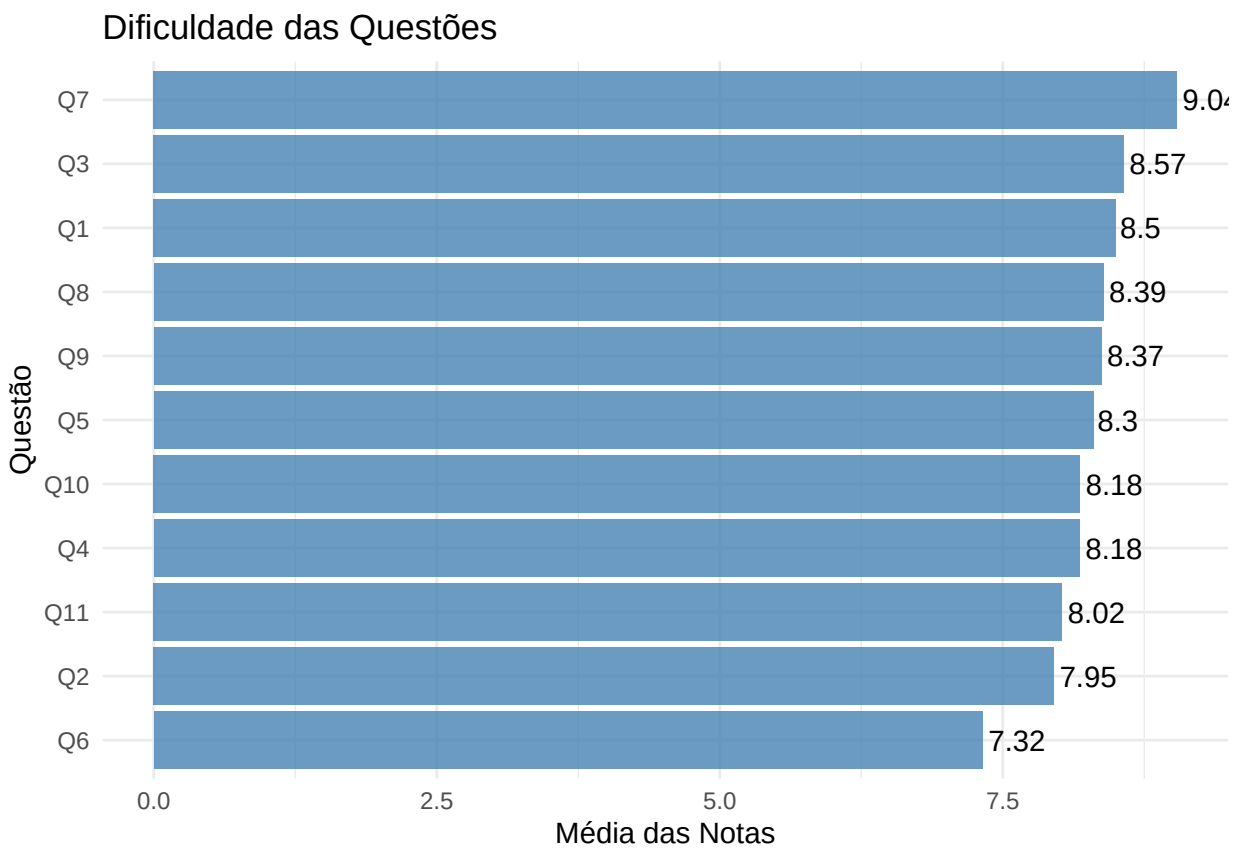



```
top_bottom_students <- rbind(
  head(student_performance, 10) %>% mutate(group="Top 10"),
  tail(student_performance, 10) %>% mutate(group="Bottom 10")
)
ggplot(top_bottom_students, aes(x=reorder(Student, mean_grade), y=mean_grade, fill=group)) +
  geom_col(alpha=0.8) +
  coord_flip() +
  facet_wrap(~group, scales="free_y") +
```

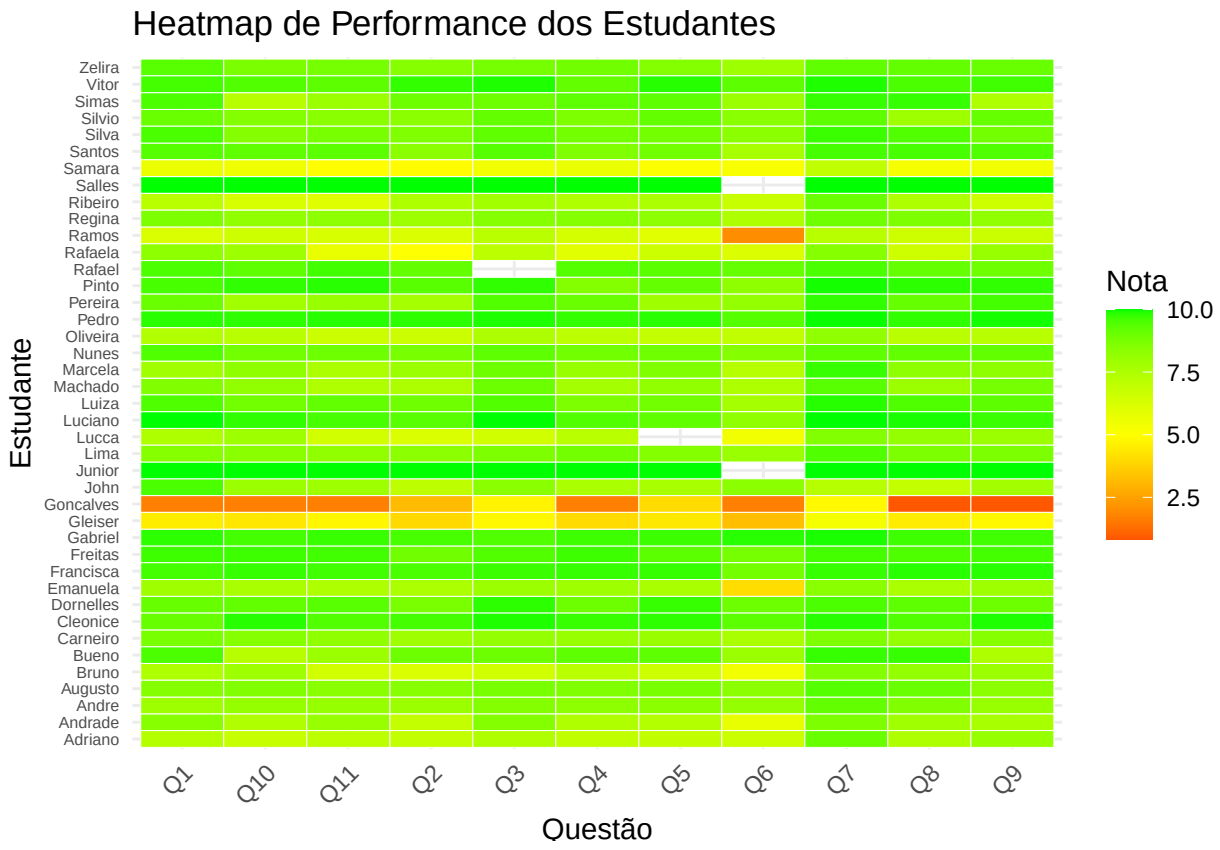
```
  labs(title="Performance dos Estudantes", x="Estudante", y="Média Geral") +
  theme_minimal() +
  theme(legend.position="none")
```



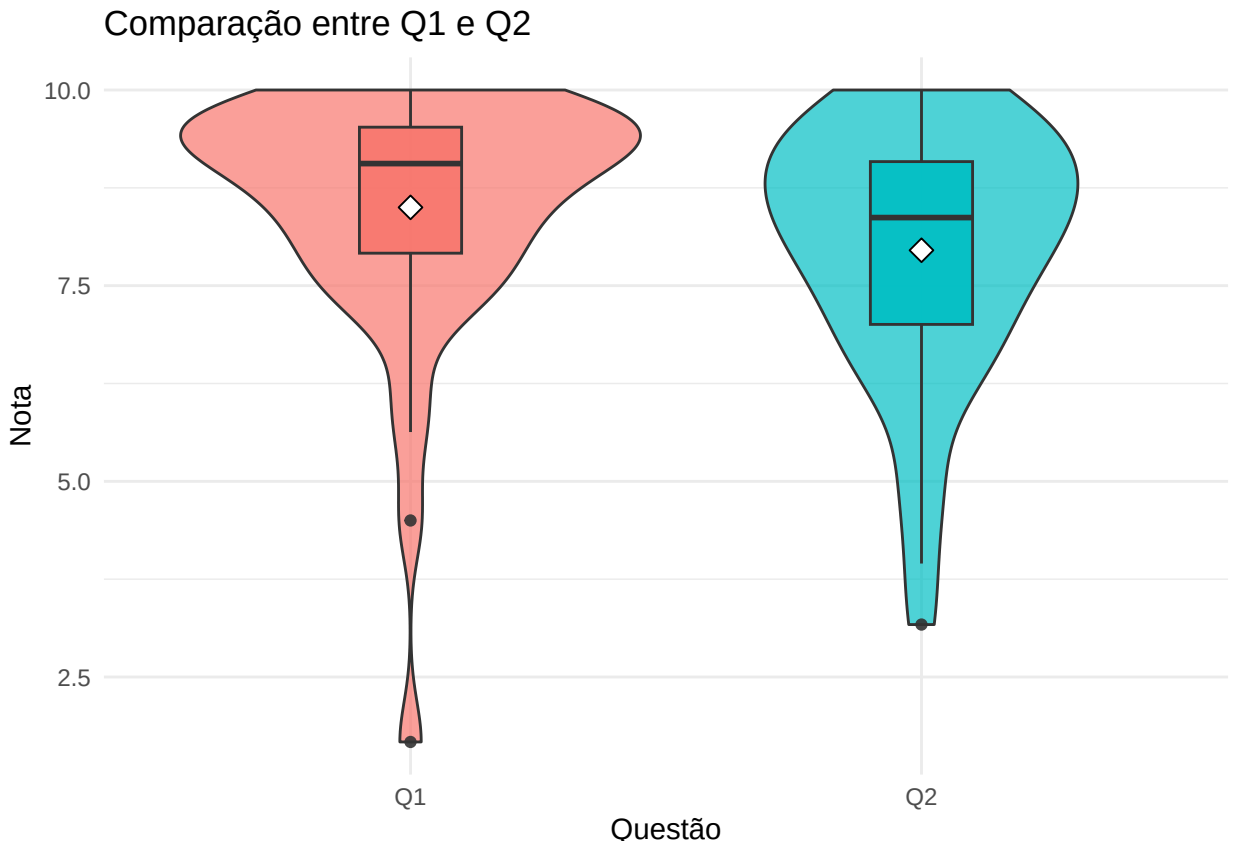

```
ggplot(question_difficulty, aes(x=reorder(Question, mean_grade), y=mean_grade)) +
  geom_col(fill="steelblue", alpha=0.8) +
  geom_text(aes(label=round(mean_grade,2)), hjust=-0.1) +
  coord_flip() +
  labs(title="Dificuldade das Questões", x="Questão", y="Média das Notas") +
  theme_minimal()
```




```
ggplot(student_tidy, aes(x=Question, y=Student, fill=Grade)) +
  geom_tile(color="white") +
  scale_fill_gradient2(low="red", mid="yellow", high="green", midpoint=5, name="Nota") +
  labs(title="Heatmap de Performance dos Estudantes", x="Questão", y="Estudante") +
  theme_minimal() +
  theme(axis.text.y=element_text(size=6), axis.text.x=element_text(angle=45,hjust=1))
```




```
q1_q2_data <- student_tidy %>% filter(Question %in% c("Q1","Q2"))
ggplot(q1_q2_data, aes(x=Question, y=Grade, fill=Question)) +
  geom_violin(alpha=0.7) +
  geom_boxplot(width=0.2, alpha=0.9) +
  stat_summary(fun=mean, geom="point", shape=23, size=3, fill="white") +
  labs(title="Comparação entre Q1 e Q2", x="Questão", y="Nota") +
  theme_minimal() +
  theme(legend.position="none")
```


Comparação entre Q1 e Q2
10.0
7.5
5.0
2.5
Nota
Q1
Q2
Questão