

Lista 2.1

Juan Belieni

Maio/2021

Importações e configurações

```
library("ggplot2")
library("readr")
library("dplyr")
library("socviz")
library('ggrepel')

dados <- read_csv("./StudentsPerformance.csv")
head(dados)
```

A tibble: 6 x 8

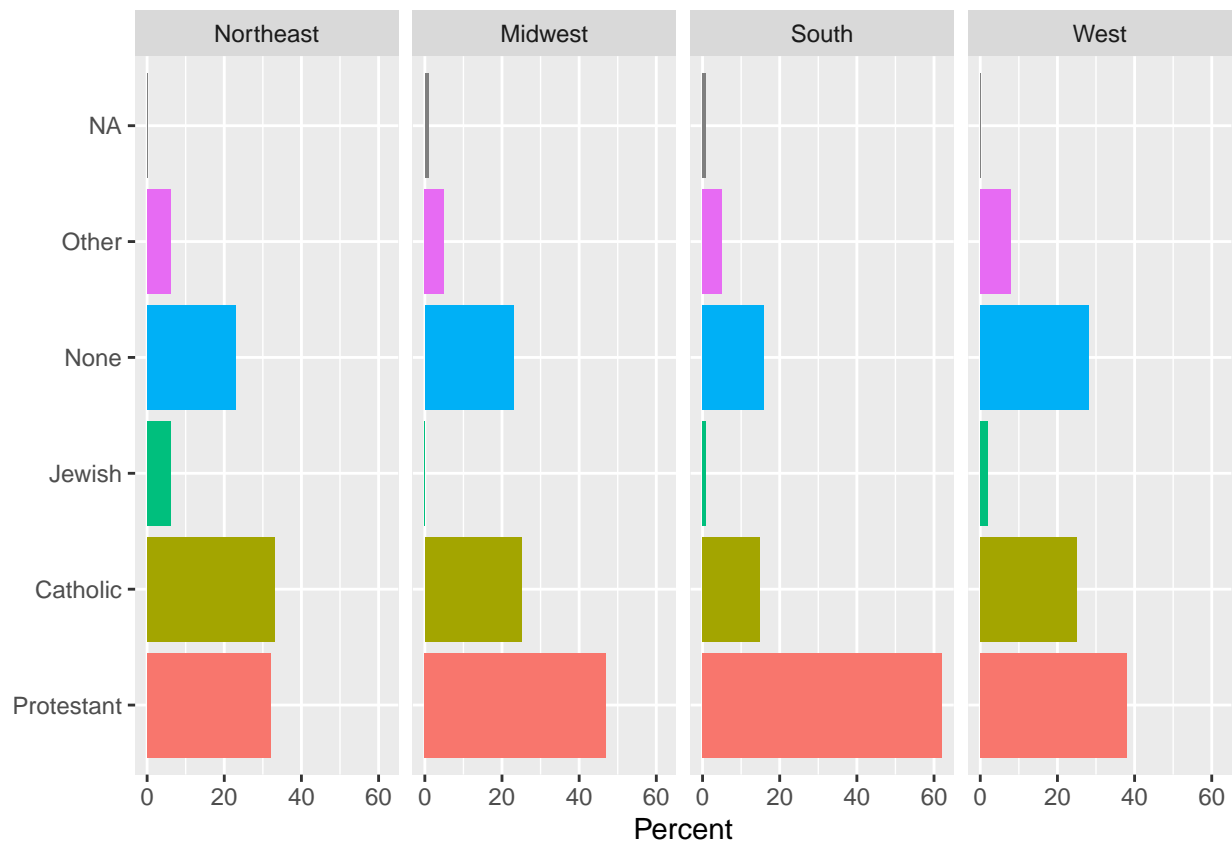
##	gender	race_ethnicity	parental_level_of_	lunch	test_preparation~	math_score
##	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>
## 1	female	group B	bachelor's degree	standa~	none	72
## 2	female	group C	some college	standa~	completed	69
## 3	female	group B	master's degree	standa~	none	90
## 4	male	group A	associate's degree	free/r~	none	47
## 5	male	group C	some college	standa~	none	76
## 6	female	group B	associate's degree	standa~	none	71

... with 2 more variables: reading_score <dbl>, writing_score <dbl>

Gráficos do capítulo 5

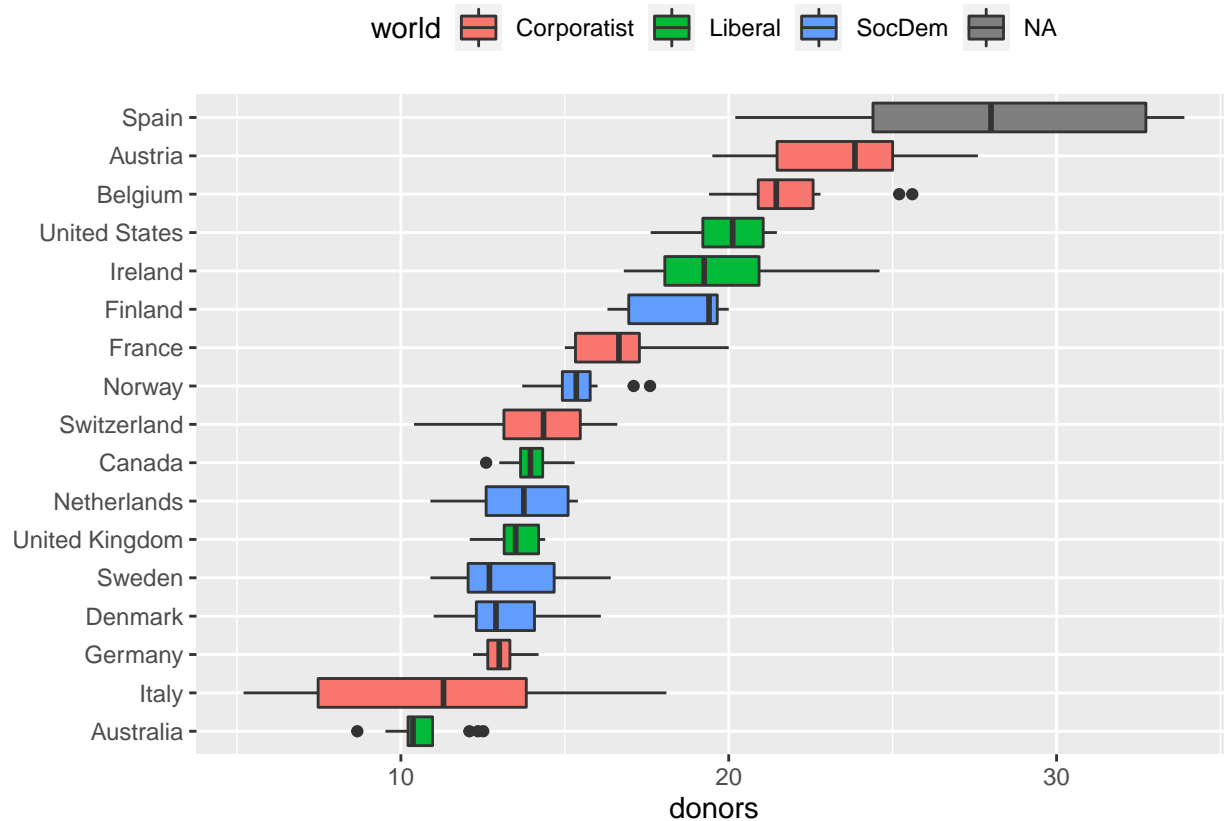
```
rel_by_region <- gss_sm %>%
  group_by(bigregion, religion) %>%
  summarize(N = n()) %>%
  mutate(freq = N / sum(N), pct = round((freq * 100), 0))

ggplot(rel_by_region, aes(x = religion, y = pct, fill = religion)) +
  geom_col(position = "dodge2") +
  labs(x = NULL, y = "Percent", fill = "Religion") +
  guides(fill = FALSE) +
  coord_flip() +
  facet_grid(~bigregion)
```



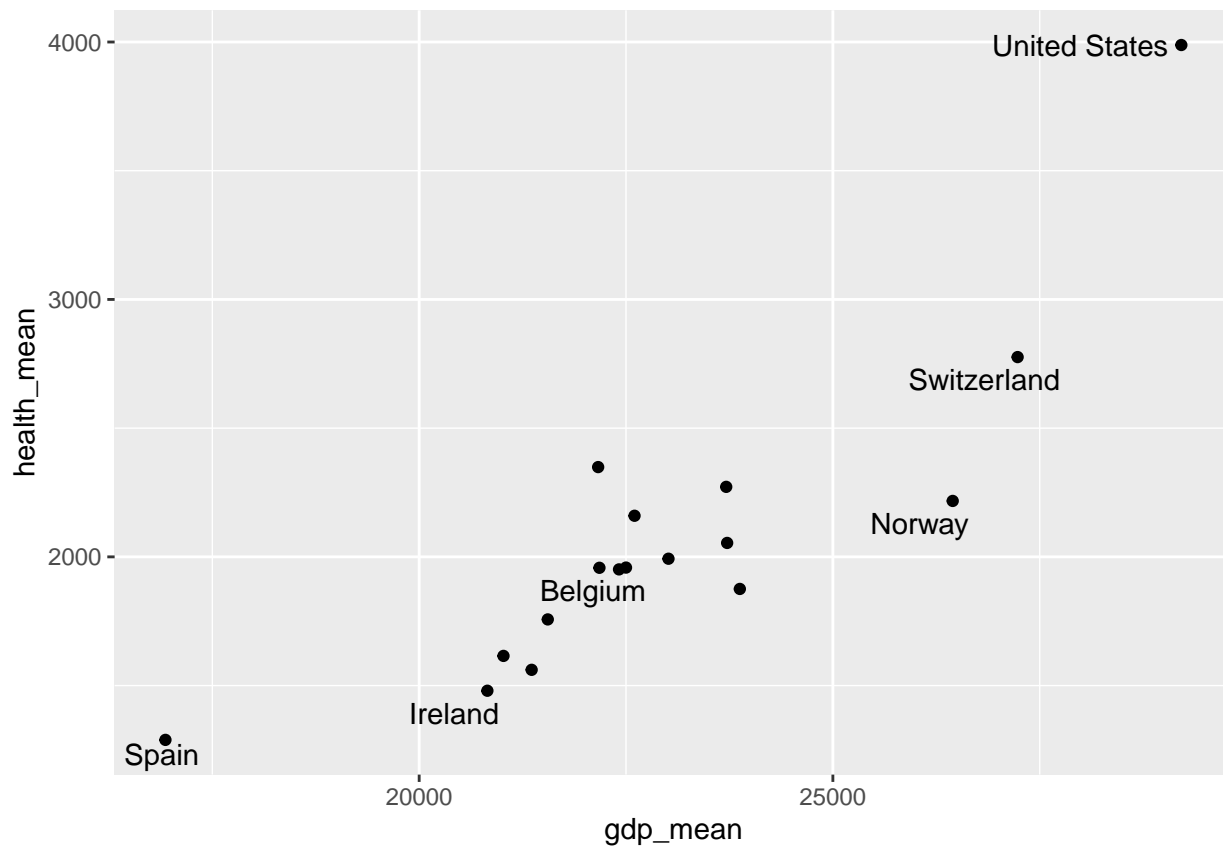
```
ggplot(
  data = organdata,
  mapping = aes(
    x = reorder(country, donors, na.rm = TRUE),
    y = donors,
    fill = world
  )
) +
  geom_boxplot() +
  labs(x = NULL) +
  coord_flip() +
  theme(legend.position = "top")
```

Warning: Removed 34 rows containing non-finite values (stat_boxplot).



```
by_country <- organdata %>%
  group_by(consent_law, country) %>%
  summarise_if(is.numeric, list(mean = mean, sd = sd), na.rm = TRUE) %>%
  ungroup()

by_country %>% ggplot(mapping = aes(x = gdp_mean, y = health_mean)) +
  geom_point() +
  geom_text_repel(
    data = subset(by_country, gdp_mean > 25000 |
      health_mean < 1500 |
      country %in% "Belgium"),
    mapping = aes(label = country)
  )
```



```
organdata %>% ggplot(aes(roads, donors)) +
  geom_point() +
  annotate(geom = "rect", xmin = 125, xmax = 155, ymin = 30, ymax = 35, fill = "red", alpha = 0.2) +
  annotate(geom = "text", x = 157, y = 33, label = "A surprinsigly high \n recovery rate.", hjust = 0)
```

```
## Warning: Removed 34 rows containing missing values (geom_point).
```

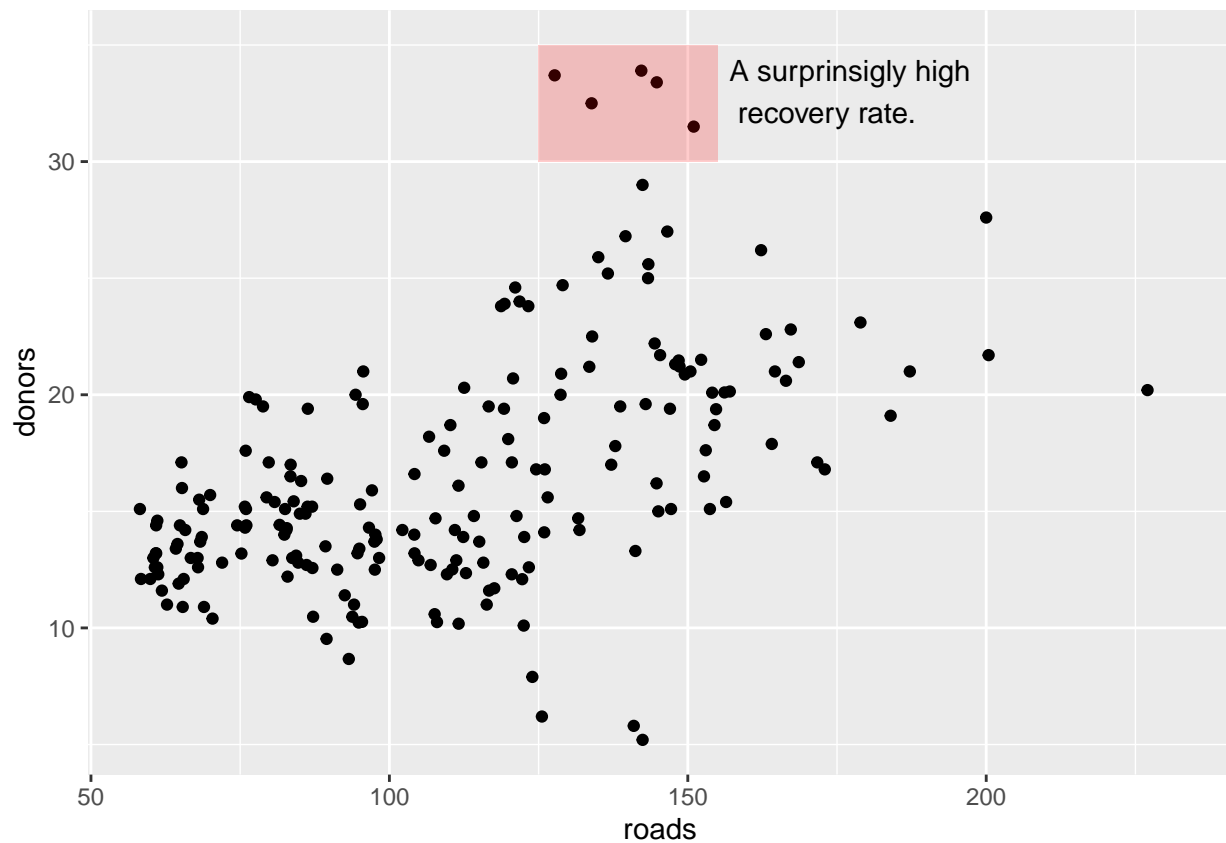


Gráfico 1

Gráfico original

```
dados %>% ggplot(aes(parental_level_of_education, fill = race_ethnicity)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  xlab("Nível de educação dos pais") +
  ylab("Porcentagem") +
  labs(fill = "Raça/etnia")
```

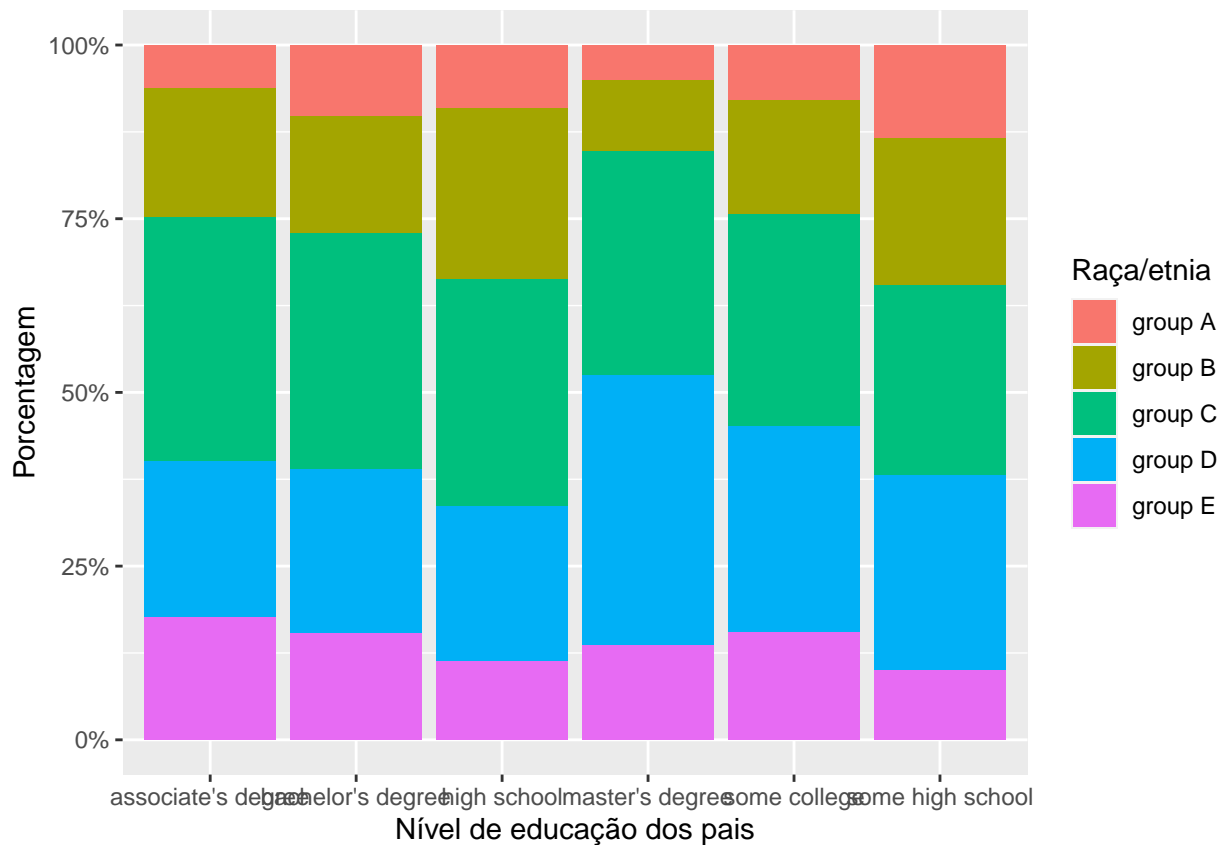


Gráfico com modificações

```
dados %>%
  group_by(parental_level_of_education, race_ethnicity) %>%
  summarise(n = n()) %>%
  ggplot(aes(n)) +
  geom_col(aes(y = reorder(race_ethnicity, n), fill = race_ethnicity)) +
  facet_wrap(~parental_level_of_education) +
  guides(fill = FALSE) +
  labs(y = NULL, x = "Quantidade")
```

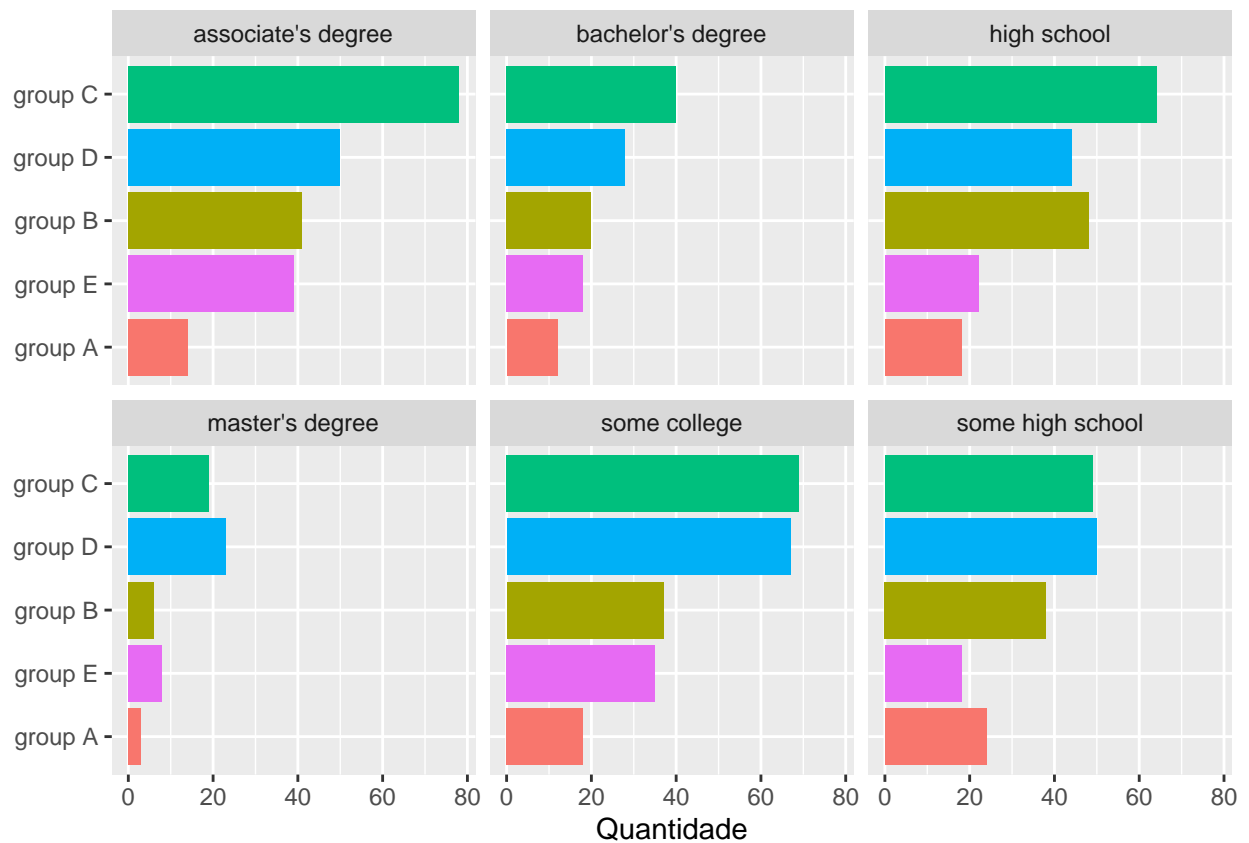


Gráfico 2

Gráfico original

```
dados %>% ggplot(aes(reading_score, writing_score, color = test_preparation_course)) +
  geom_jitter() +
  facet_wrap(~test_preparation_course) +
  xlab("Nota em leitura") +
  ylab("Nota em escrita") +
  labs(color = "Curso de preparação")
```

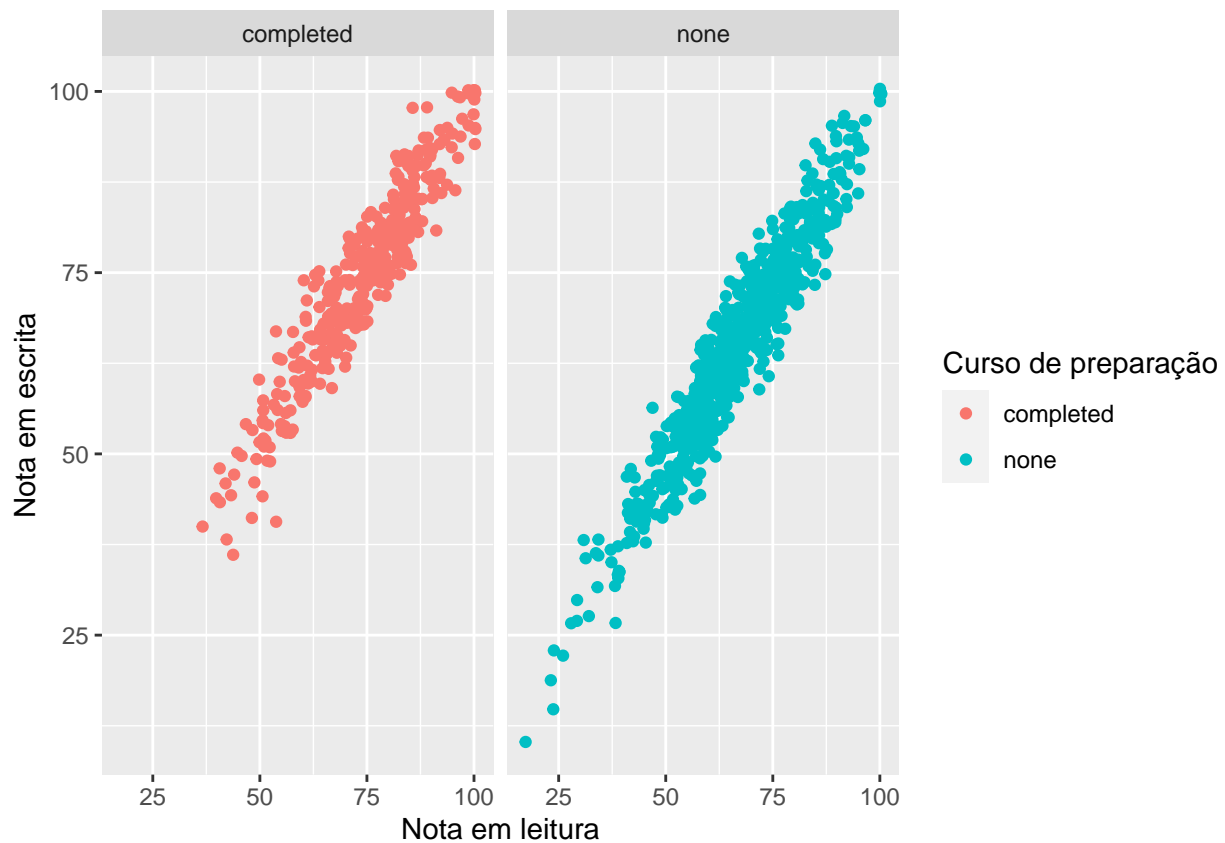


Gráfico com modificações

```
dados %>%
  ggplot(aes(reading_score, writing_score, color = test_preparation_course)) +
  geom_jitter(alpha = 0.5) +
  scale_color_manual(breaks = c("completed", "none"), values = c("red", "grey")) +
  annotate(geom = "rect",
    xmin = 15,
    xmax = 40,
    ymin = 8,
    ymax = 39,
    fill = "red",
    alpha = 0.1) +
  annotate(geom = "text",
    x = 42,
    y = 24,
    label = "Alunos que tiraram notas\nbaixas nas duas matérias",
    hjust = 0) +
  labs(x = "Nota em leitura", y = "Nota em escrita", color = "Curso de preparação") +
  theme(legend.position = "top")
```

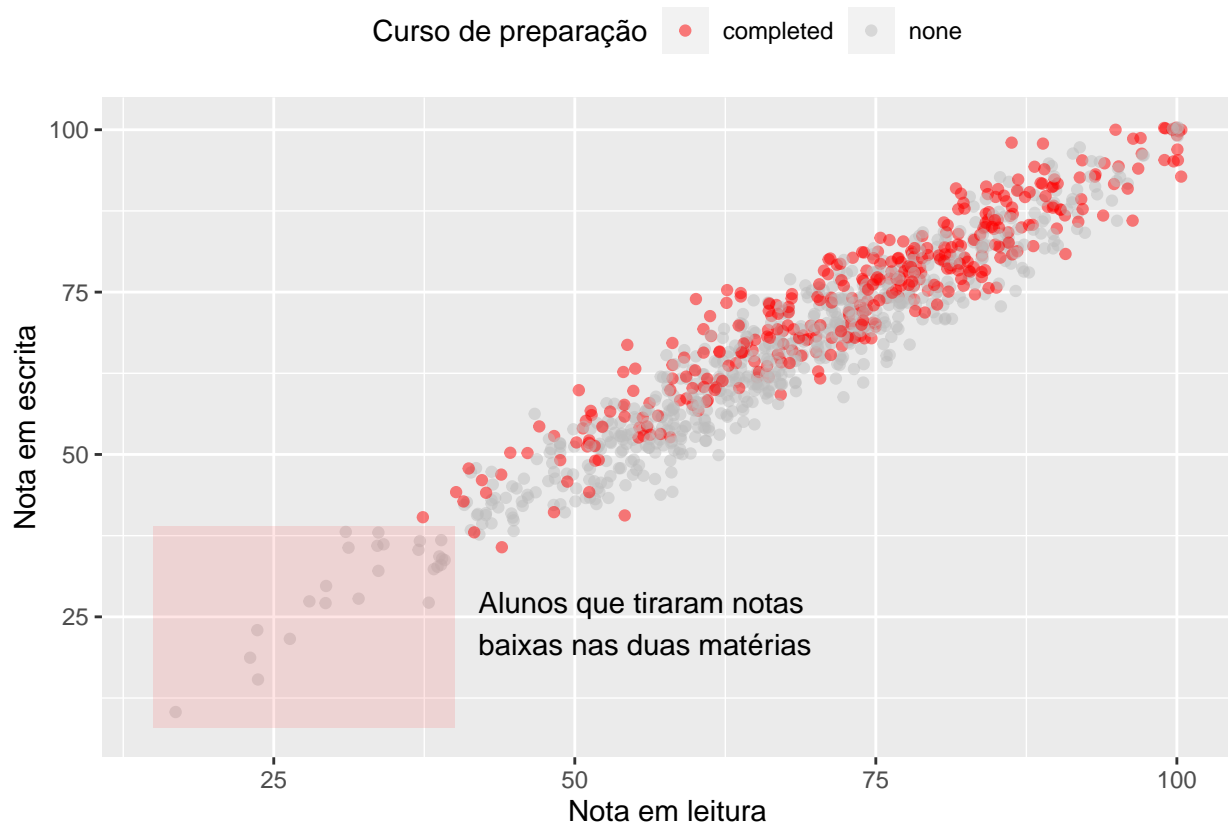



Gráfico 3

Gráfico original

```
dados %>% ggplot(aes(
  parental_level_of_education,
  math_score, fill = test_preparation_course
)) +
  geom_boxplot() +
  xlab("Nível de educação dos pais") +
  ylab("Nota em matemática") +
  labs(fill = "Curso de preparação")
```

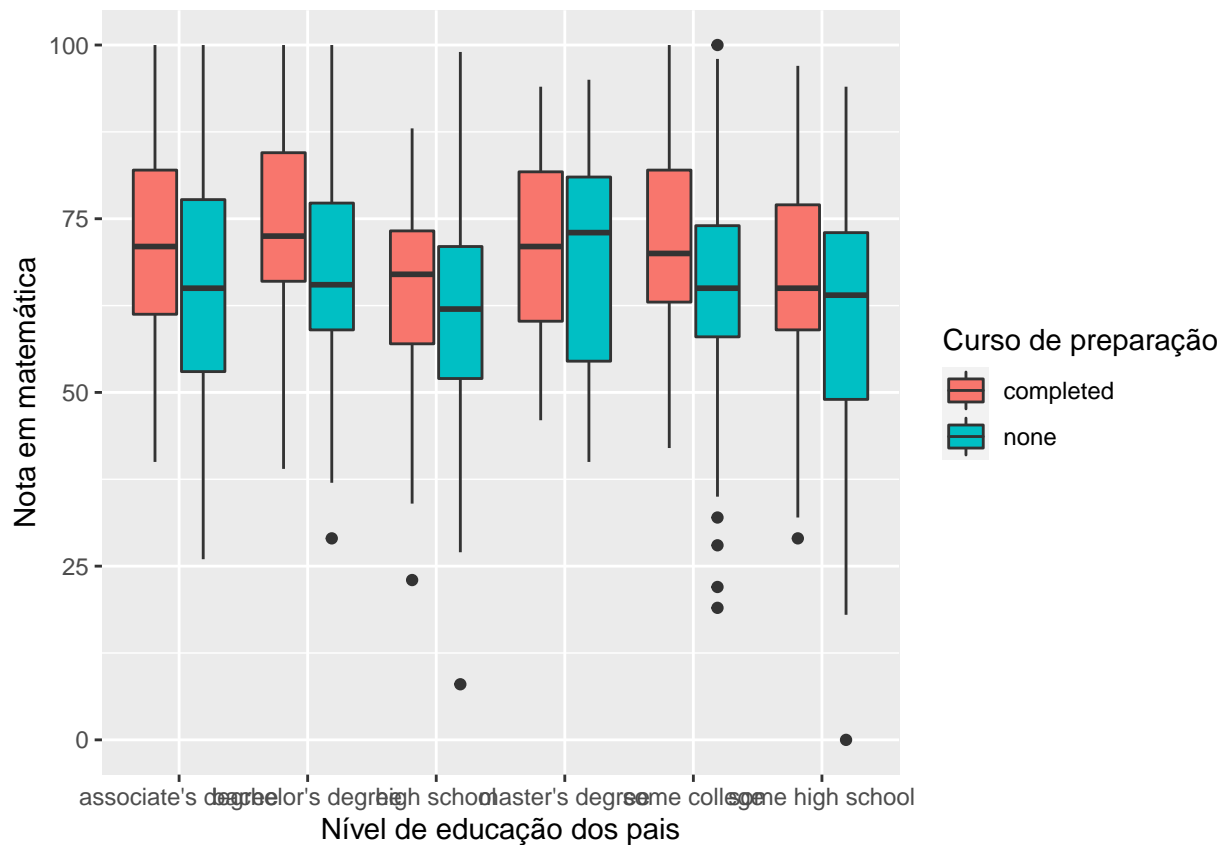


Gráfico com modificações

```
outliers <- dados %>%
  group_by(test_preparation_course, parental_level_of_education) %>%
  summarise(minimo = quantile(math_score, 0.25) - 1.5 * IQR(math_score))

dados_com_minimo <- dados %>%
  left_join(outliers)

dados_com_minimo %>%
  ggplot(aes(math_score, test_preparation_course, fill = test_preparation_course, label = math_score)) +
  geom_boxplot() +
  geom_label_repel(data = dados_com_minimo %>% filter(math_score < minimo)) +
  facet_wrap(~parental_level_of_education) +
  labs(x = "Nota em matemática", y = "Curso de preparação") +
  guides(fill = FALSE)
```

