High School Student's Performance in Portugal

Olamide Adu

Introduction

This study is to *investigate* the performance of high school students of two schools in **Portugal**.

Problem Statement

Figure 1: High school student

eustudyassist

Data

```
math_score <-  read_excel("data/student-performance/student-math.xlsx")
```

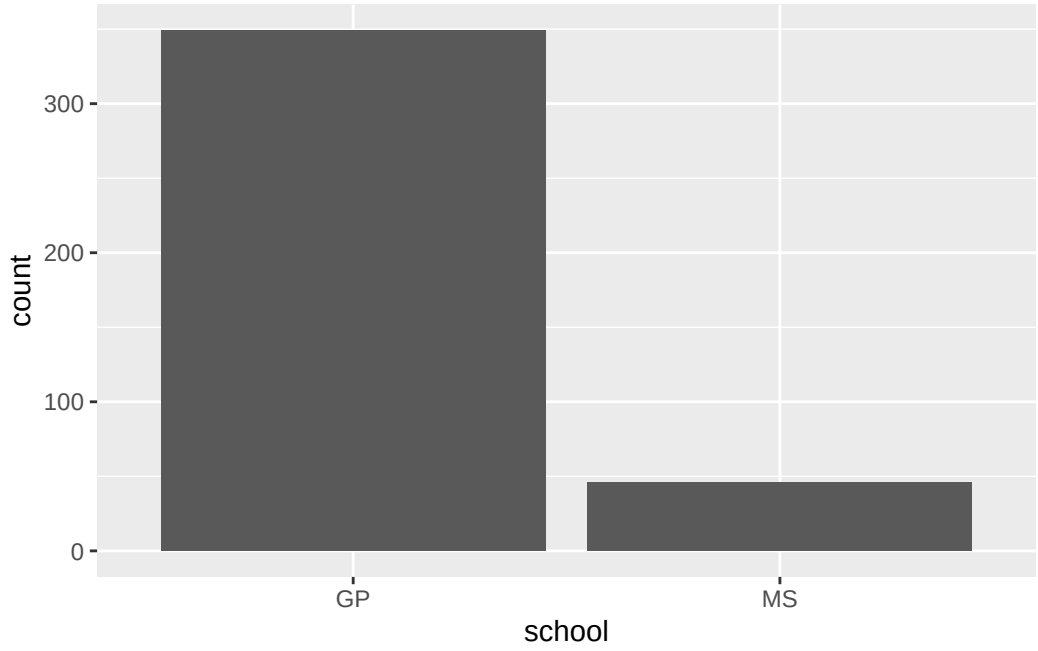
Previewing the data

```
math_score |> select(1:6) |>
  head() |>
  kable()
```

school	sex	age	address	famsize	pstatus
GP	F	18	U	GT3	A
GP	F	17	U	GT3	T
GP	F	15	U	LE3	T
GP	F	15	U	GT3	T
GP	F	16	U	GT3	T
GP	M	16	U	LE3	T

The number of individuals in the school is presented below

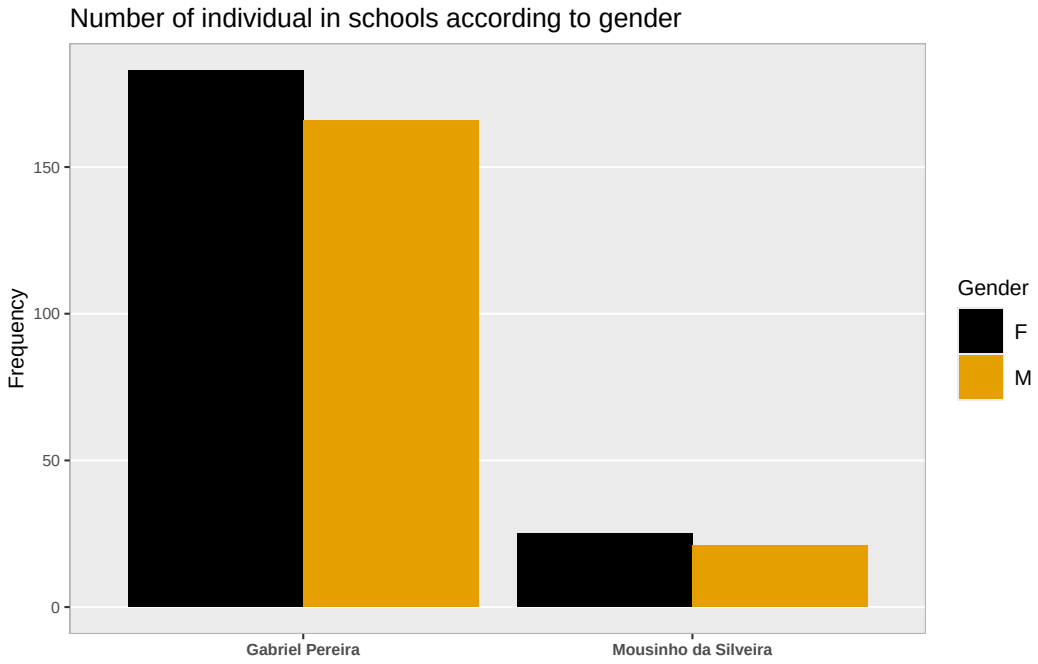
```
math_score |>
  ggplot(aes(school)) +
  geom_bar()
```




```
math_score <- math_score |>
  mutate(
    total_score = g1 + g2 + g3,
    percent_score = total_score/60 * 100,
    internet = ifelse(internet == "no", FALSE, TRUE),
    across(where(is.character), factor)
  )
```

```
math_score |>
  mutate(
    school = factor(
      school,
      levels = c("GP", "MS"),
      labels = c("Gabriel Pereira", "Mousinho da Silveira")
    )
  ) |>
  ggplot(aes(school, fill = sex)) +
  geom_bar(
    position = "dodge"
  ) +
  labs(
    title = "Number of individual in schools according to gender",
    y = "Frequency",
```

```
    fill = "Gender"
  ) +
  scale_fill_colorblind() +
  theme_calc(base_size = 8) +
  theme(
    axis.title.x = element_blank(),
    axis.text.x = element_text(face = "bold")
  )
```




```
math_score_model_1 <-  lm(
  percent_score ~ age,
  data = math_score
)

summary(math_score_model_1)
```

```
Call:
lm(formula = percent_score ~ age, data = math_score)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-48.087 -10.855   0.246  13.297  42.479

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   85.947     12.124   7.089 6.31e-12 ***
age           -1.950      0.724  -2.693  0.00739 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.34 on 393 degrees of freedom
Multiple R-squared:  0.01811,   Adjusted R-squared:  0.01562
F-statistic:  7.25 on 1 and 393 DF,  p-value: 0.007392
```

```
glance(math_score_model_1) |>
  select(1:6) |>
  set_names(c("rsq", "adj_rsq", "sig",
              "stats", "p_value", "df")) |>
  kable()
```

rsq	adj_rsq	sig	stats	p_value	df
0.0181143	0.0156159	18.33904	7.250253	0.0073924	1

```
tidy(math_score_model_1) |>
  kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	85.947115	12.1239029	7.089063	0.0000000
age	-1.949574	0.7240411	-2.692629	0.0073924