

# Teacher student ratio

Oliver Hagger

2023-01-08

## R Markdown

Markdown document for teacher to student ratio data

```
library(tidyverse)
library(ggplot2)
```

```
library(tidyverse)
library(ggplot2)
```

```
df_students <- readr::read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/
                               /master/data/2019/2019-05-07/
                               /student_teacher_ratio.csv")
```

```
head(df_students)
```

```
df_world_tile <- readr::read_csv("https://gist.githubusercontent.com/maartenzam/
                                  787498bbc07ae06b637447dbd430ea0a/raw/
                                  9a9dafafb44d8990f85243a9c7ca349acd3a0d07/
                                  worldtilegrid.csv") %>%
```

```
  mutate(
    ## Namibias two-digit country code is handled as `NA` - let us fix that
    alpha.2 = if_else(name == "Namibia", "NA", alpha.2),
    ## We are going to split "Americas" into "North America" and "Sout America"
    region = if_else(region == "Americas", sub.region, region),
    region = if_else(region %in% c("Northern America", "Central America", "Caribbean"),
                     "North America", region),
    region = if_else(region == "Southern America", "South America", region),
    ## to join both data sets, we need a id column
    country_code = alpha.3
  )
```

```
df_ratios <- df_students %>%
  ## Let's keep only the most recent data per country
  group_by(country, indicator) %>%
  filter(year == max(year)) %>%
  ungroup() %>%
  # Create `NA`s for countries which do not have any data 2012-2018
  complete(indicator, nesting(country, country_code)) %>%
  ## Let's focus on primary education and keep only countries (coded by letters)
  filter(
    indicator == "Primary Education",
```

```

    str_detect(country_code, "[A-Z]")
  ) %>%
  ## merge with world tile map data
  full_join(df_world_tile) %>%
  filter(
    !is.na(region),
    !is.na(indicator)
  ) %>%
  group_by(region) %>%
  mutate(student_ratio_region = median(student_ratio, na.rm = T)) %>%
  ungroup()

df_ratio2 <- df_ratios %>%
  select(c("indicator", "country", "region", "student_ratio", "student_ratio_region"))

ratios_3 <- na.omit(df_ratio2)

ggplot(ratios_3, aes(x = region, y = student_ratio)) +
  geom_boxplot()

df_sorted2 <-
  ratios_3 %>%
  mutate(region = fct_reorder(region, -student_ratio_region))

ggplot(df_sorted2, mapping = aes(x=region, y=student_ratio)) + geom_boxplot()

ggplot(df_sorted2, mapping = aes(x=student_ratio, y=region, color = region)) +
  geom_violin() +
  scale_x_continuous(limits = c(0,90), expand = c(0.02, 0.02))

theme_set(theme_light(base_size = 18, base_family = "Poppins"))

#Sort out the variables of the graphs
g <-
  ggplot(df_sorted2, aes(x = student_ratio, y = region, color = region)) +
  scale_x_continuous(limits = c(0, 90), expand = c(0.02, 0.02)) +

  labs(y = NULL, x = "Student to teacher ratio") +
  theme(
    legend.position = "none",
    axis.title = element_text(size = 16),
    axis.text.x = element_text(family = "Roboto Mono", size = 12),
    panel.grid = element_blank()
  )

#Change to jitter to showcase more points
g + geom_jitter(size = 3, alpha = 0.25, height = 0.25) +
  stat_summary(fun = mean, geom = "point", size = 5)

#Get a world average to compare for the different continents
world_avg <-
  df_sorted2 %>%

```

```

summarize(avg = mean(student_ratio, na.rm = TRUE)) %>%
pull(avg)

#Add text labels to the graph
g_text <- g + geom_jitter(size = 4, alpha = 0.25,height = 0.2)+
  stat_summary(fun = mean, geom = "point", size = 5)+
  geom_vline(aes(xintercept=world_avg, colour="grey"))+
  geom_segment(
    aes(y = region, yend = region,
        x = world_avg, xend = student_ratio_region),
    size = 0.8)+
  annotate(
    "text", x = 35, y = 6.3, family = "Poppins", size = 3, color = "gray20", lineheight = .9,
    label = glue::glue("Worldwide average:\n{round(world_avg, 1)} students per teacher")
  ) +
  annotate(
    "text", x = 10, y = 3.5, family = "Poppins", size = 3, color = "gray20",
    label = "Continental average"
  ) +
  annotate(
    "text", x = 11, y = 1.7, family = "Poppins", size = 3, color = "gray20",
    label = "Countries per continent"
  )

#Set arrows
arrows <-
  tibble(
    y1 = c(6.1, 3.62, 1.8, 1.8),
    y2 = c(5.6, 4, 2.1, 2.76),
    x1 = c(world_avg + 6, 10.5, 9, 9),
    x2 = c(world_avg + 0.1, 18.4, 14.16, 12)
  )

#Plot arrows
g_arrows <-
  g_text +
  geom_curve(
    data = arrows, aes(x = x1, y = y1, xend = x2, yend = y2),
    arrow = arrow(length = unit(0.08, "inch")), size = 0.5,
    color = "gray20", curvature = -0.3
  )

plot(g_arrows)

```