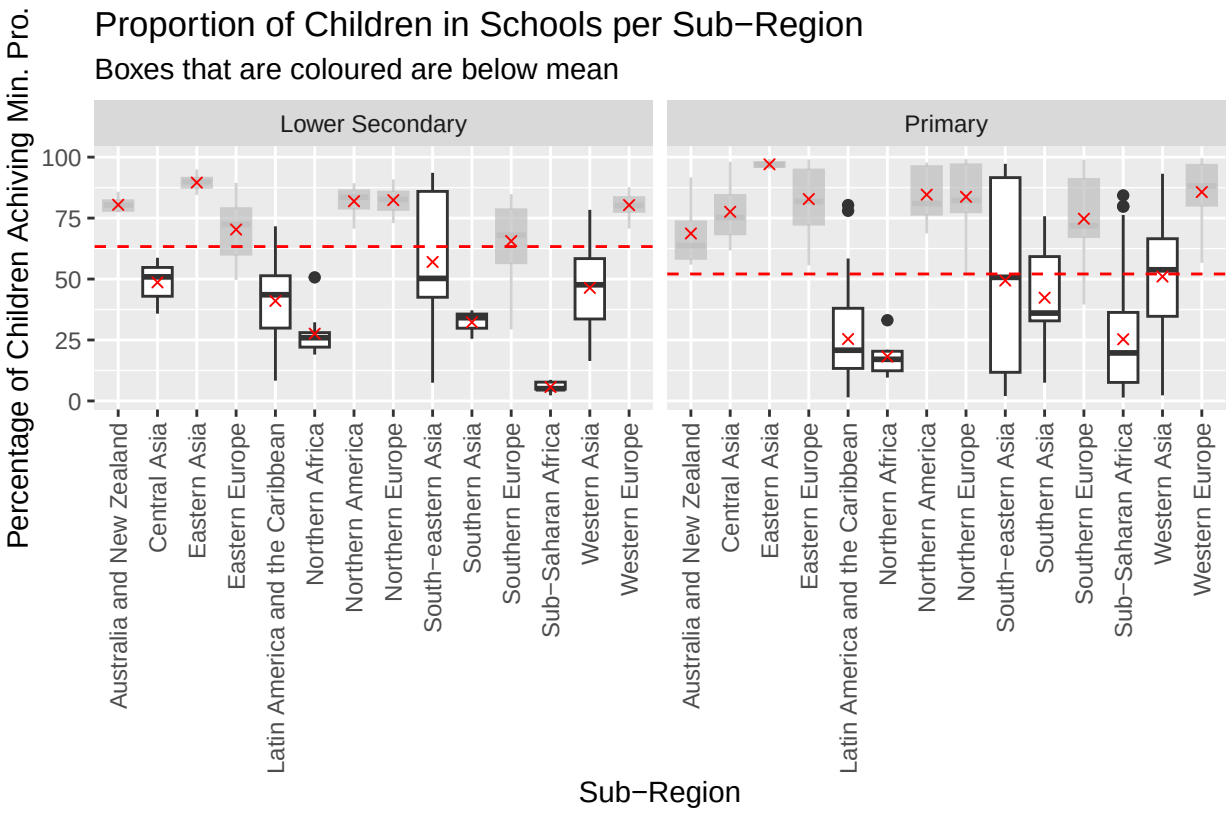
Question One

Andrew Dawson, Lewis Miller

2025-12-01

Comparing Sub-Regions



Within the box plots, the highlighted box plots show each Sub Region's mean being less than the mean for the overall Education Level – for clarity the red crosses are the means for each Sub Region, whilst the black points are the outliers. Within the Lower Secondary section, only one Sub-Region displays an outlier, being Northern Africa. Interestingly, this outlier is still below the grand mean of the section, displaying that Northern Africa is an area of concern. Additionally, South-east Asia is the only Sub-Region where the mean falls below the grand mean, but the third quartile is higher. This shows much more spread across the Sub-Region. Within the Primary section, outlier are shown in Latin America and the Caribbean, Northern Africa, and Sub-Saharan Africa. Additonally, South-east Asia shows an incredible spread of data, with the IQR being NA.

Whilst the Lower Secondary grand mean, being 63.32, is higher than the Primary's grand mean, being 52.07, it is important to consider the two means are not independent of another. In other words, in order to get into secondary school one must pass primary, thus the grand mean of the Primary schools would affect the

mean of the Lower Secondary schools.

South-east Asia also demonstrates the most spread across the sub-regions. Contextually, this aligns with the general developmental growth of the region. For instance, the education within Singapore is usually shown as the world's best standard for education, whilst a nation like Laos or Thailand, lacks behind, which all three countries are within the same sub-region.

Recomendation

For Sub-Regions which have it's maximum underneath the grand mean, more allocation of resources is needed there, whilst for Sub-Regions where its maximum does go above the grand mean but the Sub-Regions' mean is below the grand mean, further analysis per Sub-Region is needed to ensure proper allocation of resources, i.e. the Singaporean example within South-east Asia.

Added

This analysis examines global variation in educational access using several SDG 4 indicators, with a primary focus on Indicator 4.1.2 (completion rate across primary, lower secondary, and upper secondary levels). By extracting the most recent available data for each country, we identify those lagging furthest behind and assess how completion rates differ across continents. The findings highlight persistent regional inequalities and reveal where progress has been slow or stagnant.

Table 1: Worst 10 countries for Indicator 4.1.2 (Completion Rate)

GeoAreaName	Continent	SubRegion	TimePeriod	Value
Côte d'Ivoire	Africa	Sub-Saharan Africa	2016	0.0
Nepal	Asia	Southern Asia	2019	0.0
Qatar	Asia	Western Asia	2012	0.0
Turks and Caicos Islands	Americas	Latin America and the Caribbean	2019	0.0
Angola	Africa	Sub-Saharan Africa	2015	1.5
Chad	Africa	Sub-Saharan Africa	2019	2.0
Malawi	Africa	Sub-Saharan Africa	2015	2.4
Madagascar	Africa	Sub-Saharan Africa	2018	2.5
Mali	Africa	Sub-Saharan Africa	2018	2.7
Rwanda	Africa	Sub-Saharan Africa	2015	3.6

Using the latest available values for each country, we extracted the ten lowest performers. All ten countries fall within low-income or lower-middle-income regions, and eight of the ten are located in Sub-Saharan Africa, including Côte d'Ivoire, Angola, Malawi, Madagascar, Mali, and Rwanda. This strongly suggests that educational access continues to be highly constrained in African regions, where systemic barriers such as poverty, conflict, and limited infrastructure are well-documented.

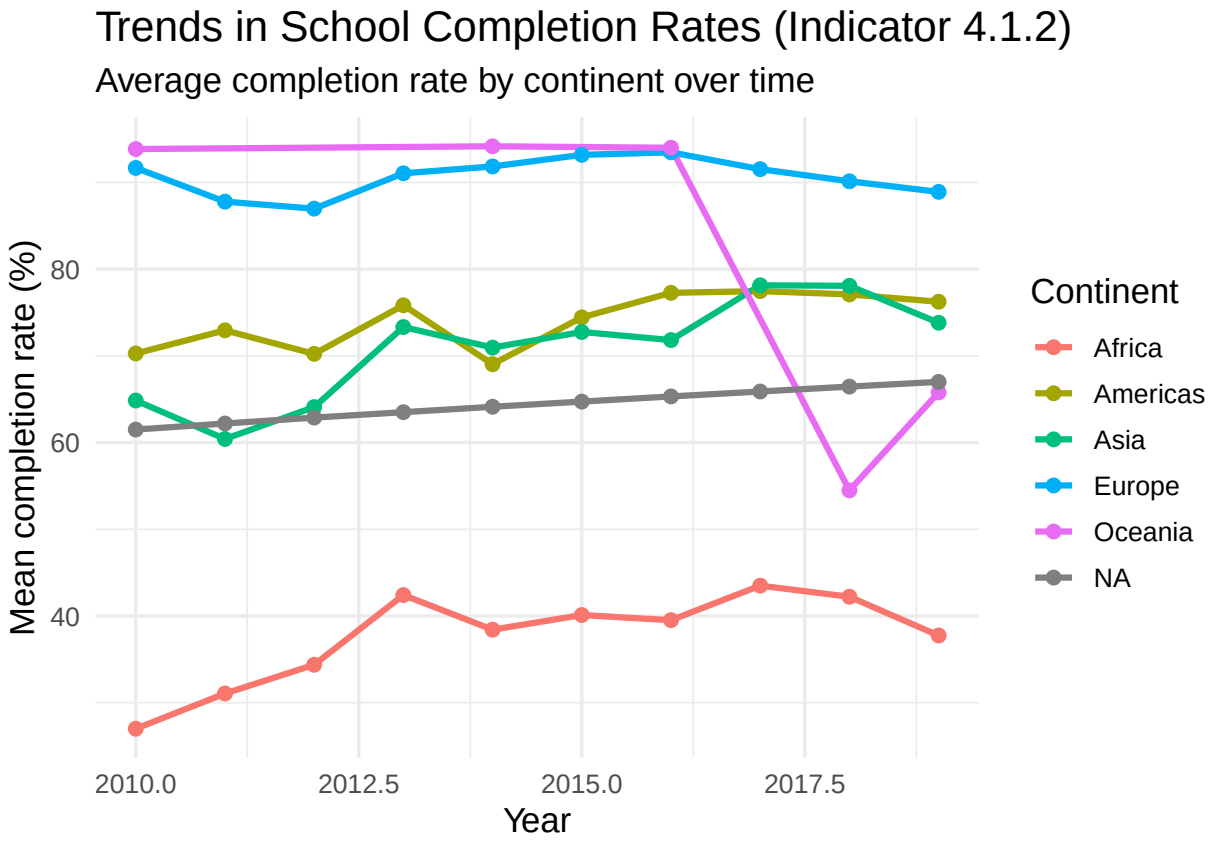
A notable methodological limitation arises from the fact that countries have uneven reporting years. For example, Qatar's most recent observation is from 2012, whereas others have data from 2018–2019. As a result, some "low-performing" countries may have improved since their last measurement. This highlights a broader challenge in using SDG datasets for cross-country comparisons: data recency varies significantly.

Now, looking at each continents development from their earliest datapoint to their most recent:

Continental trends in school completion

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
```

```
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



To reduce noise arising from irregular reporting and to understand broader patterns, we computed average completion rates over time for each continent. The resulting plot (Figure 1) reveals several key insights:

Clearly, as shown previously, we notice African countries struggling with school completion rates. Not only are their school completion rates currently significantly lower than other continents, but their trend over the last 5 years does not look promising. Asia and the Americas show higher completion rates with stable increasing trends throughout the timeline. Europe reports the highest overall completion rates, with minimal variation over time and near-universal completion for primary and lower secondary education.

Oceania displays a sharp and irregular decline in completion rates, which appears misleading at first glance. This pattern could well be driven by the dominance of data from Small Island Developing States (SIDS), such as Tuvalu and Samoa, whose extremely small populations make their indicators statistically volatile. For example, despite having a population over 2,000 times smaller than Australia, Tuvalu contributes nearly half as many data points (380 vs 899). As a result, small fluctuations in enrolment or reporting can disproportionately shift regional averages.

This highlights a key limitation: regional aggregates for Oceania should be interpreted cautiously, as they are highly sensitive to small-island reporting and do not reliably reflect the trend for the region's largest country (Australia). Analysts should consider weighting by population or examining Australia separately if a stable trend is required.

References

Duncalfe, Luke. 2024. "ISO-3166 Country and Dependent Territories Lists with UN Regional Codes." github.com/lukes/ISO-3166-Countries-with-Regional-Codes?tab=readme-ov-file.
United Nations. 2023. "SGD Indicators Database." https://unstats.un.org/sdgs/dataportal/database.

Appendix

```
#for future plots
tiltXText <- theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#filter data
DataPro <- Data %>% filter(Indicator == "4.1.1", Units == "PERCENT", Sex == "BOTHSEX", SubRegion != "NA

#set up data for plot
lines <- DataPro %>% group_by(`Education level`) %>% summarise(grandMean = mean(Value))
points <- DataPro %>% group_by(SubRegion, `Education level`) %>% summarise(mean = mean(Value))

#merge for highlighted plot
DataPro <- left_join(DataPro, lines, by = "Education level")

#plot filtered data
DataPro %>% ggplot(aes(SubRegion, Value))+
  geom_boxplot()+
  facet_wrap(~`Education level`, labeller = labeller(`Education level` = c("LOWSEC" = "Lower Secondary"
  gghighlight(mean(Value) < mean(grandMean), calculate_per_facet = T)+
  tiltXText+
  geom_hline(aes(yintercept = grandMean), lty = 2, colour = "red", data=lines)+
  geom_point(data = points, aes(y=mean), pch = 4, colour="red")+
  labs(x = "Sub-Region", y = "Percentage of Children Achiving Min. Pro.", title = "Proportion of Childr
latest_values <- Data %>%
  filter(Sex == "BOTHSEX",
         Indicator %in% c("4.1.1", "4.1.2", "4.6.1", "4.a.1", "4.c.1")) %>%
  arrange(GeoAreaName, TimePeriod) %>%
  group_by(GeoAreaName, Indicator) %>%
  slice_tail(n = 1) %>%
  ungroup()

worst10 <- function(ind_code) {
  latest_values %>%
    filter(Indicator == ind_code) %>%
    arrange(Value) %>%
    slice_head(n = 10) %>%
    select(GeoAreaName, Continent, SubRegion, TimePeriod, Value)
}

worst_412 <- worst10("4.1.2")

kable(worst_412,
      caption = "Worst 10 countries for Indicator 4.1.2 (Completion Rate)",
      digits = 1) %>%
  kable_styling(full_width = FALSE)
```

```r
continent_trends <- Data %>%
  filter(Indicator == "4.1.2",
         Sex == "BOTHSEX") %>%
  group_by(Continent, TimePeriod) %>%
  summarise(
    mean_completion = mean(Value, na.rm = TRUE),
    n_countries      = n_distinct(GeoAreaName),
    .groups = "drop"
  )


ggplot(continent_trends,
       aes(x = TimePeriod, y = mean_completion, colour = Continent)) +
  geom_line(size = 1.1) +
  geom_point(size = 2) +
  labs(
    title = "Trends in School Completion Rates (Indicator 4.1.2)",
    subtitle = "Average completion rate by continent over time",
    x = "Year",
    y = "Mean completion rate (%)",
    colour = "Continent"
  ) +
  theme_minimal(base_size = 13)
```