

# Question 3 EDA

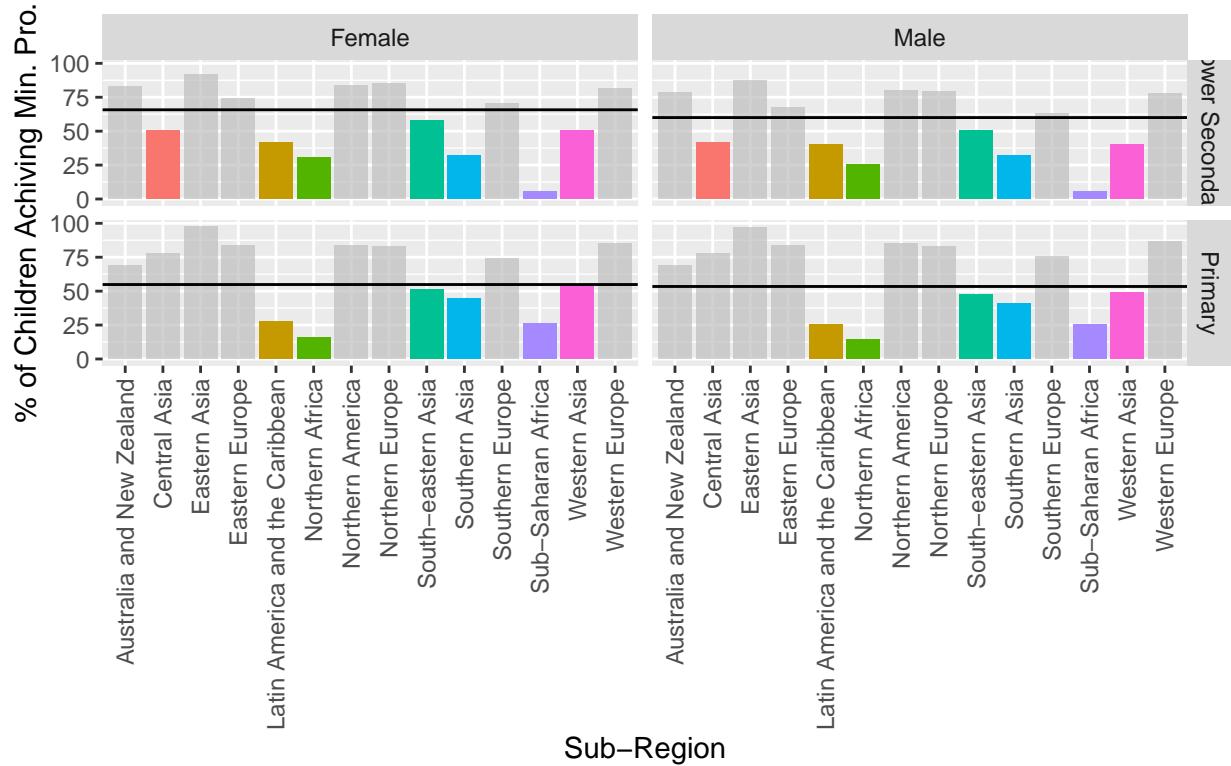
Andrew Dawson

## Comparing Disparities in Educational Access

### Sex

Comparing this with the two sexes also shows different patterns.

### Sex and Education Level Across the Different Sub–Regions



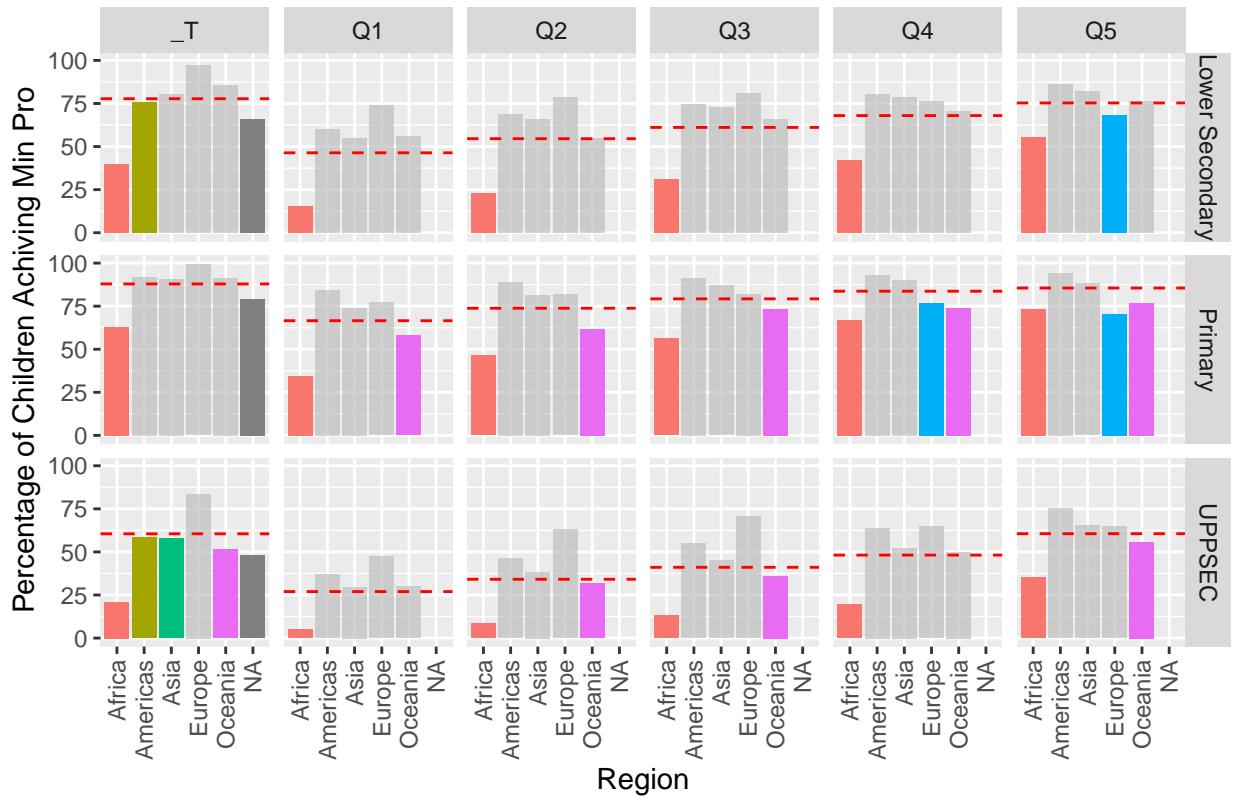
Using the GDP Indicator 5.1.1

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. Interestingly, both sexes have an equal number of Sub-Regions whereby the Sub-Regions mean is less than the grand mean of the section. In addition, the primary section has less sub-regions where the mean is less than the grand mean; in other words, Central Asia has a greater mean than the grand mean in Primary Education whilst it has a smaller mean in Secondary for both sexes.

The IQR for the sexes differ. The IQR for males seems to be less than females.

## Wage

### Comparing Wealth with Completion Rate

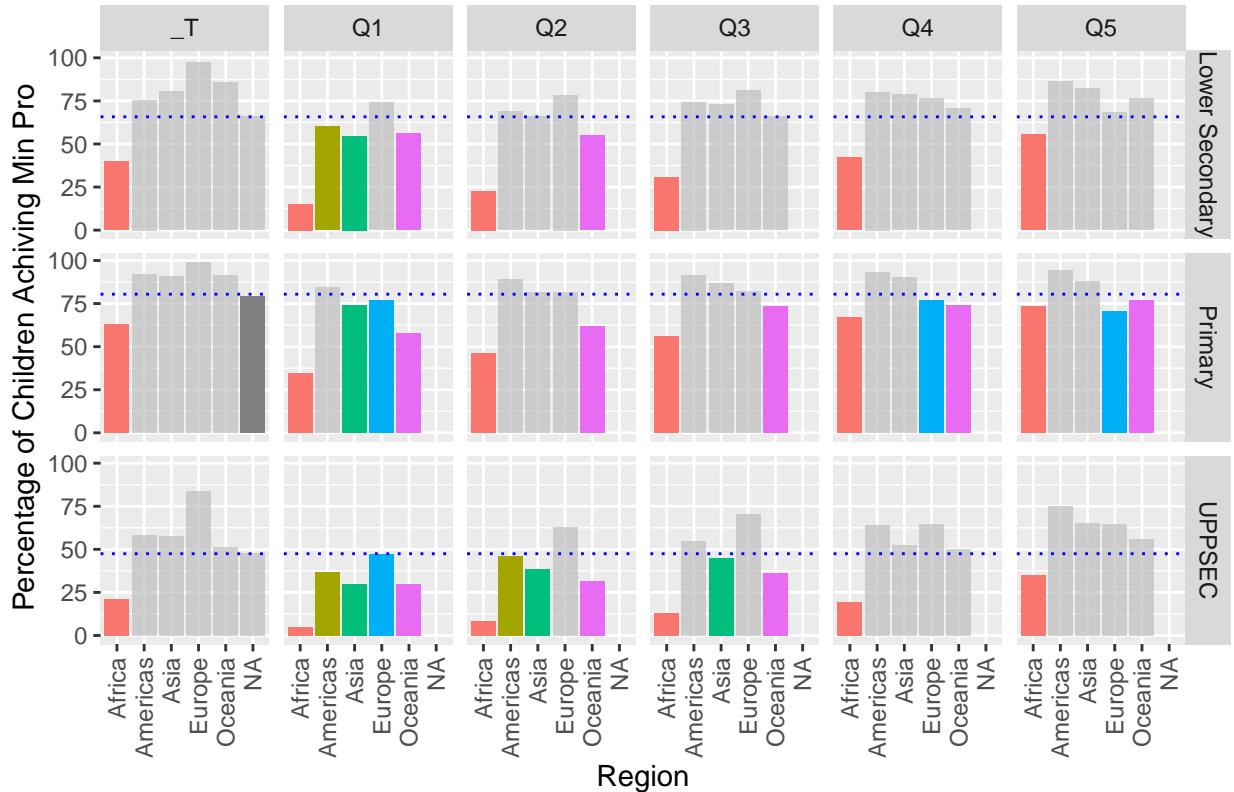


Within this plot, the general regions are compared with wealth quantiles, i.e. Q1 indicates the poorest of the population, and Q5 the wealthiest.

The mean completion rate per region seems to almost have more wealthy people lack behind. I don't think this is a fair comparison so adjustment of the graph is needed.

Take this plot with a grain of salt, for example look at Europe for Q5. IQR is the entire range, and the median is 100 yet the mean is less than the grandmean.

## Comparing Wealth with Completion Rate



This plot better highlights the wealth disparity, as it highlights the poorer countries to have more issues in educational completion rate, which is known to be historically proven

## Recommendation

Sexes may not exactly demonstrate a need to differentiate, however there may be bias within the data.

Wealth shows that less fortunate families may need further support.

## 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](https://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 dataset
DataProSex <- Data %>%
  filter(Indicator == "4.1.1", Units == "PERCENT", Sex != "BOTHSEX", `sub-region` != "NA")
linesSex <- DataProSex %>%
  group_by(`Education level`, Sex) %>%
```

```

    summarise(sexMean = mean(Value))
pointsSex <- DataProSex %>%
  group_by(`sub-region`, `Education level`, Sex) %>%
  summarise(regionalMean = mean(Value))

#join together
pointsSex <- left_join(pointsSex, linesSex, by = c("Education level" = "Education level", "Sex" = "Sex"))

#plot
pointsSex %>% ggplot(aes(`sub-region`, regionalMean, fill = `sub-region`))+
  geom_col()+
  facet_grid(`Education level` ~ Sex, labeller = labeller(`Education level` = c("LOWSEC" = "Lower Secondary", "HIGHSEC" = "Higher Secondary", "Tertiary" = "Tertiary", "Post-Secondary" = "Post-Secondary", "Other" = "Other"), Sex = c("Male" = "Male", "Female" = "Female")))
  gghighlight(regionalMean < sexMean, calculate_per_facet = T)+
  geom_hline(aes(yintercept = sexMean))+
  theme(legend.position = "hide")+
  tiltXText+
  labs(x = "Sub-Region", y = "% of Children Achieving Min. Pro.", title = "Sex and Education Level Across Sub-Regions")
## lets make a bar plot
#make the lines and means
DataComp <- Data %>% filter(Indicator == "4.1.2", Units == "PERCENT")
linesWage <- DataComp %>%
  group_by(`Education level`, Quantile) %>%
  summarise(quantileMean = mean(Value))
linesWage2 <- DataComp %>%
  group_by(`Education level`) %>%
  summarise(eduMean = mean(Value))
pointsWage <- DataComp %>%
  group_by(`Education level`, Quantile, region) %>%
  summarise(regionalMean = mean(Value)) %>%
  left_join(linesWage, by = c("Education level" = "Education level", "Quantile" = "Quantile")) %>%
  left_join(linesWage2, by = c("Education level" = "Education level"))

#plot the data
pointsWage %>% ggplot(aes(region, regionalMean, fill = region))+
  geom_col()+
  facet_grid(`Education level` ~ Quantile, labeller = labeller("Education level" = c("LOWSEC" = "Lower Secondary", "HIGHSEC" = "Higher Secondary", "Tertiary" = "Tertiary", "Post-Secondary" = "Post-Secondary", "Other" = "Other"), Quantile = c("1st" = "1st", "2nd" = "2nd", "3rd" = "3rd", "4th" = "4th", "5th" = "5th")))
  gghighlight(regionalMean < quantileMean, calculate_per_facet = T)+
  #geom_hline(aes(yintercept = eduMean), lty = 3, colour = "blue")+
  geom_hline(aes(yintercept = quantileMean), lty = 2, colour = "red")+
  theme(legend.position = "none")+
  tiltXText+
  labs(x = "Region", y = "Percentage of Children Achieving Min Pro", title = "Comparing Wealth with Completing Education Across Regions")
#plot the data
pointsWage %>% ggplot(aes(region, regionalMean, fill = region))+
  geom_col()+
  facet_grid(`Education level` ~ Quantile, labeller = labeller("Education level" = c("LOWSEC" = "Lower Secondary", "HIGHSEC" = "Higher Secondary", "Tertiary" = "Tertiary", "Post-Secondary" = "Post-Secondary", "Other" = "Other"), Quantile = c("1st" = "1st", "2nd" = "2nd", "3rd" = "3rd", "4th" = "4th", "5th" = "5th")))
  gghighlight(regionalMean < eduMean, calculate_per_facet = T)+
  geom_hline(aes(yintercept = eduMean), lty = 3, colour = "blue")+
  #geom_hline(aes(yintercept = quantileMean), lty = 2, colour = "red")+
  theme(legend.position = "none")+
  tiltXText+
  labs(x = "Region", y = "Percentage of Children Achieving Min Pro", title = "Comparing Wealth with Completing Education Across Regions")

```