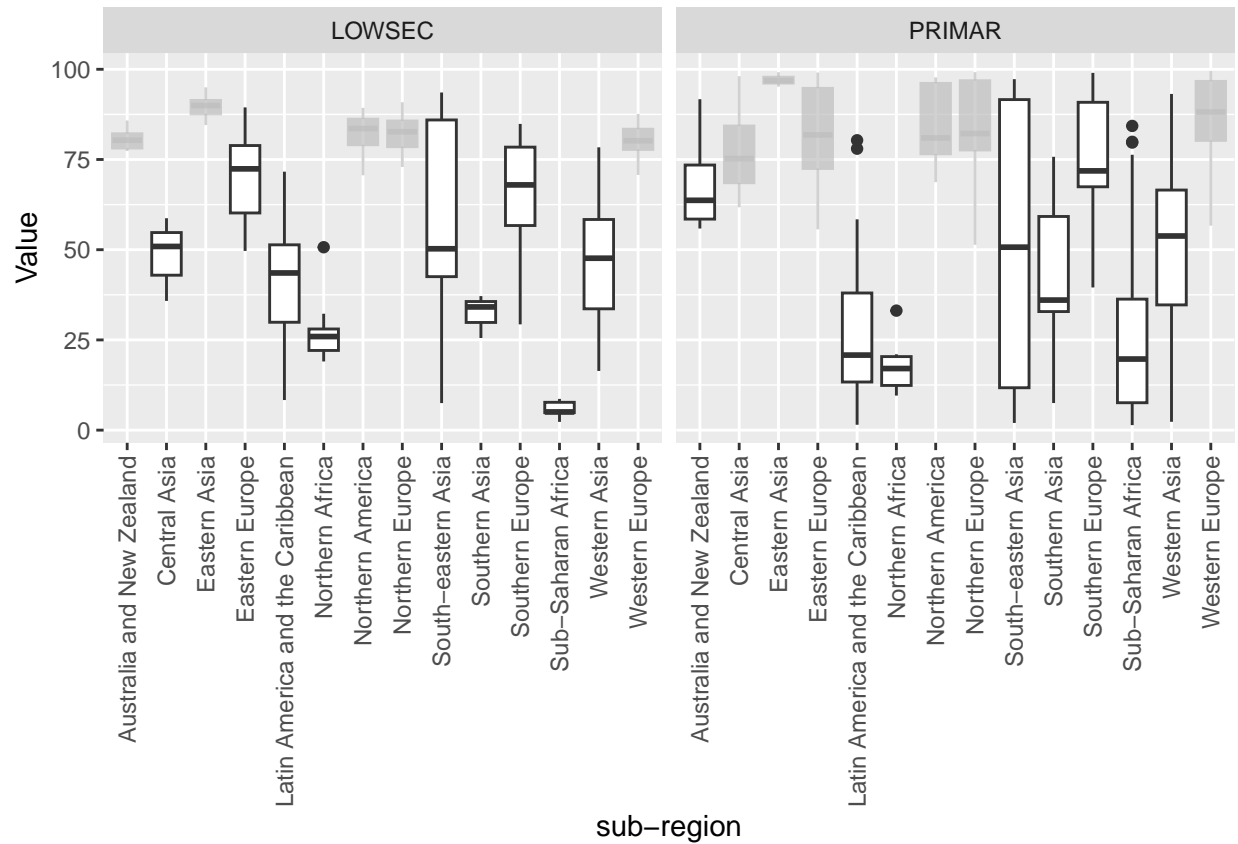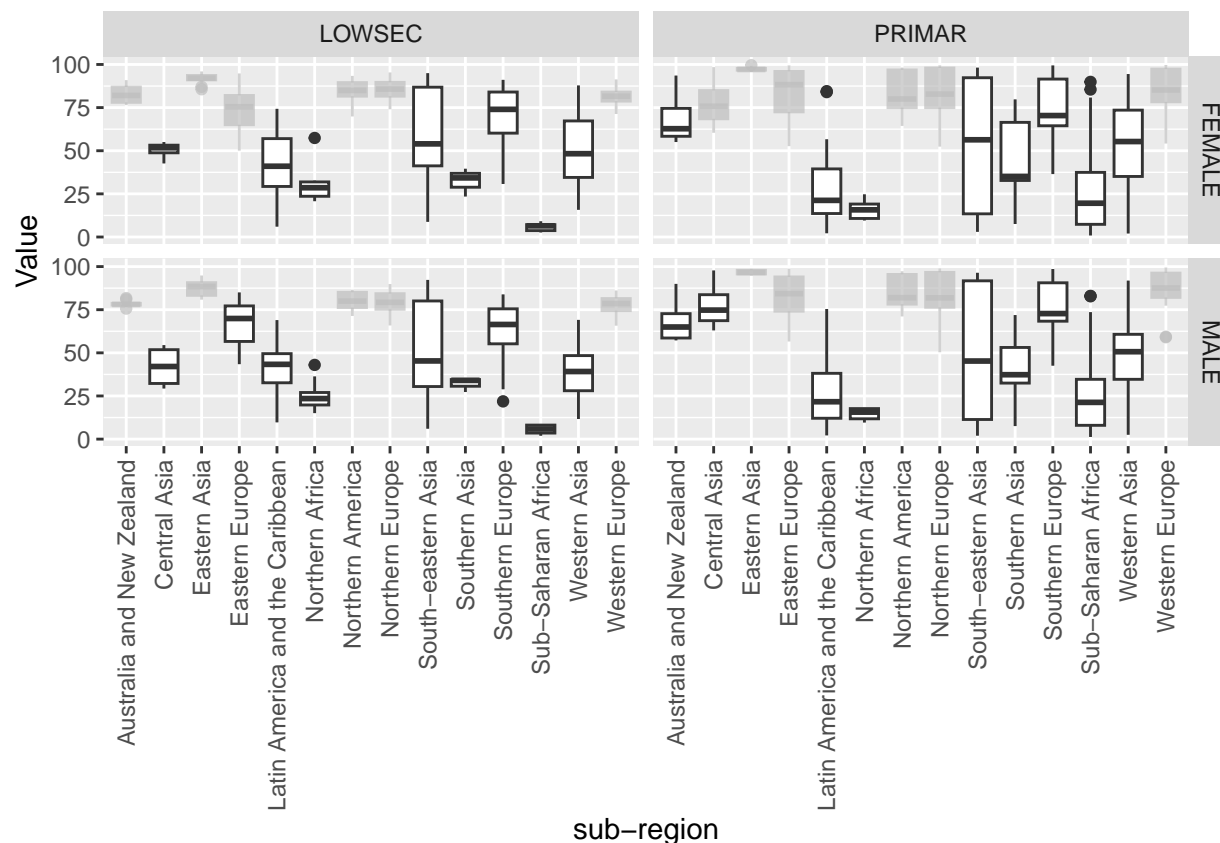# Question 1 EDA

Andrew Dawson

## EDA and Analysis of 4.1.1



Within the box plots, the highlighted box plots show within the differing sub-regions' median of population, which percentage of the population have not achieved minimum proficiency in reading and mathematics. This demonstrates the sub-regions capability of providing education access as the lower the percentage of the population with the minimum proficiency of the two vital skills, the lower the access is.

Comparing this with the two sexes also shows different patterns.

Now, what is most concerning is when looking at Latin America and the Caribbean and Northern Africa. Both of these sub-regions in both plots demonstrate a general pattern that the percentage of the entire population – or just about – has not achieved 75% proficiency.

| GeoAreaName | sub-region | TimePeriod | Value | Education level | Type of skill |
| --- | --- | --- | --- | --- | --- |
| Niger | Sub-Saharan Africa | 2014 | 1.40000 | PRIMAR | SKILL_MATH |
| Dominican Republic | Latin America and the Caribbean | 2013 | 1.50000 | PRIMAR | SKILL_MATH |
| Chad | Sub-Saharan Africa | 2019 | 1.80000 | PRIMAR | SKILL_MATH |
| Zambia | Sub-Saharan Africa | 2016 | 1.80000 | PRIMAR | SKILL_READ |
| Lao People's Democratic Republic | South-eastern Asia | 2019 | 2.00000 | PRIMAR | SKILL_READ |
| Dominican Republic | Latin America and the Caribbean | 2019 | 2.10000 | PRIMAR | SKILL_MATH |
| Niger | Sub-Saharan Africa | 2014 | 2.10000 | PRIMAR | SKILL_READ |
| Zambia | Sub-Saharan Africa | 2017 | 2.29846 | LOWSEC | SKILL_MATH |
| Yemen | Western Asia | 2011 | 2.32531 | PRIMAR | SKILL_MATH |
| Côte d'Ivoire | Sub-Saharan Africa | 2019 | 2.60000 | PRIMAR | SKILL_MATH |

These countries display the worst percentages of their population, this aligns well with the plot.

# 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

```r
#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", `sub-region` !=

#plot filtered data
DataPro %>% ggplot(aes(`sub-region`, Value))+
  geom_boxplot()+
  facet_wrap(~`Education level`)+
  gghighlight(quantile(Value,0.5)<75, calculate_per_facet = T)+
  tiltXText
#filter data
DataProSex <- Data %>% filter(Indicator == "4.1.1", Units == "PERCENT", Sex != "BOTHSEX", `sub-region`

#plot filtered data
DataProSex %>% ggplot(aes(`sub-region`, Value))+
  geom_boxplot()+
  facet_grid(Sex~`Education level`)+
  gghighlight(quantile(Value,0.5)<75, calculate_per_facet = T)+
  tiltXText
#table
tab <- DataPro %>%
  arrange(Value) %>%
  select(GeoAreaName, `sub-region`, TimePeriod, Value, `Education level`, `Type of skill`) %>%
  slice_head(n=10)

kable(tab)
```