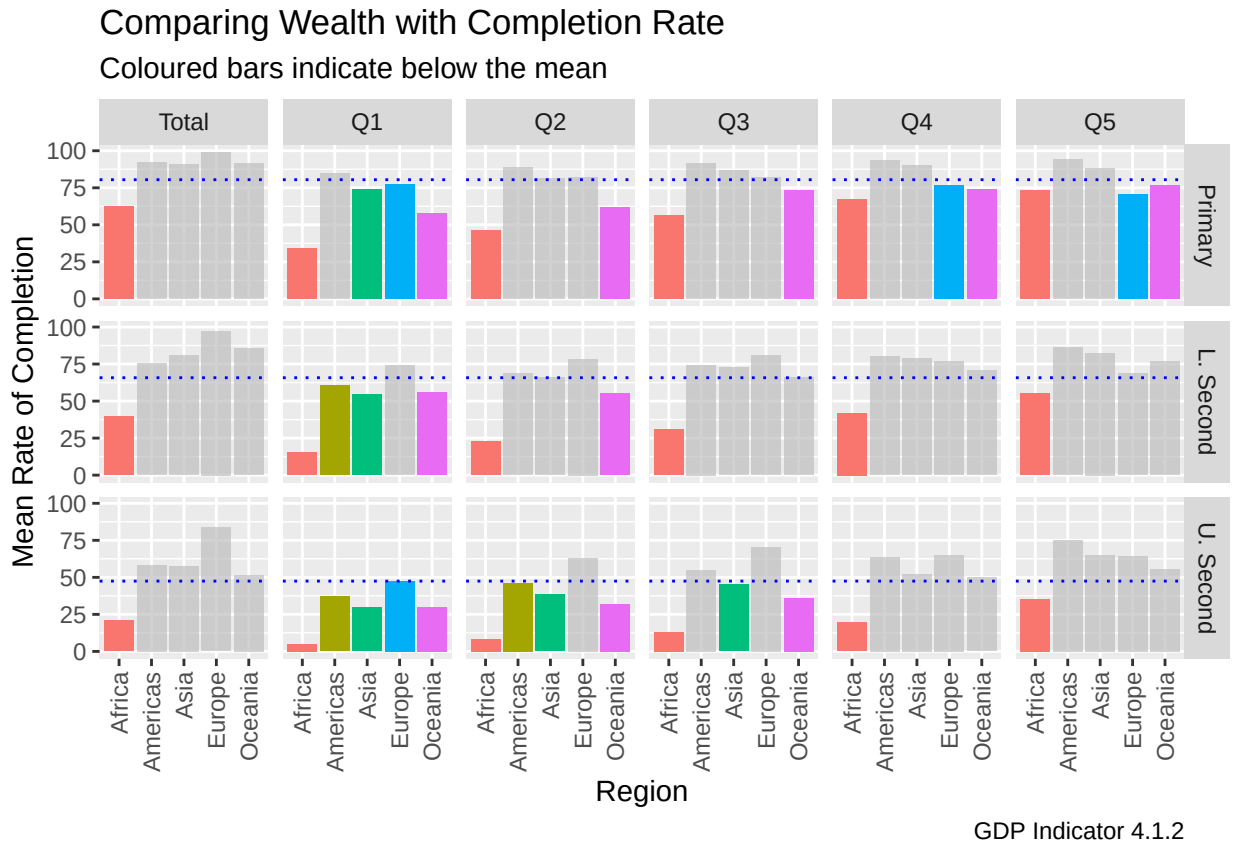
Question 3 EDA

Andrew Dawson

Comparing Disparities in Educational Access

Wage



Within the plot, the blue dotted lines are the mean for the level of education, i.e. for primary, lower secondary, and upper secondary. Furthermore, when a bar is coloured this indicates that the region's mean is below the mean of the education level. The means for each the primary, lower secondary and upper secondary are 80, 66, and 47, respectively. When comparing the coloured bars, Africa noticeably is is under the mean is all wealth quantiles, and all education levels.
Comparing the wealthiest quantile to the poorest quantile, in primary education all three countries in the wealthiest quantile also show in the poorest quantile. Interestingly, this is the only instant where this happens. When looking at the two different secondary levels, in both wealthiest quantiles – Q5 and Q4 – only Africa is below the mean; however, when comparing that to the poorest quantile, nearly all regions show – bar Europe in the lower secondary category.

When looking across the different wealth quantiles, it is important to account for multiple things; for instance,

when looking at the different education levels they are not independent of another. In order to progress to from Primary to Lower Secondary, you must complete and pass Primary education, the same goes for Upper Secondary with lower secondary. This is shown in the data with the lower rates of completion as the education level increases.

Another factor is informal education, defined by Weinland (2023) as "learning about cultural values, norms, and expected behaviours by participating in a society", this affects the quantiles as when a child engages within their respective culture they may find a shift away from education towards something more culturally important – e.g. working within the family farm, supporting their parents or grandparents, etc.

Finally, this plot is not a perfect representation of the wealth between the regions, as it only captures the mean rate of completion. One cannot see the spread of information or outliers which may affect the mean, thus changing the information of the plot; however, this may be accounted for using weighted means instead.

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.

Weinland, Kathryn. 2023. "Intercultural Communication." Electronic Article. https://open.library.okstate.edu/interculturalcommunication/chapter/education-around-the-world/.

Appendix

```
##make the lines and means
#set up the data
DataComp <- Data %>% filter(Indicator == "4.1.2", Units == "PERCENT", !is.na(region))

#set up the
linesWage <- DataComp %>%
  group_by(`Education level`, Quantile) %>%
  summarise(quantileMean = mean(Value))
linesWage2 <- DataComp %>%
  group_by(`Education level`) %>%
  summarise(eduMean = mean(Value))

#join all together
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"))
#reorder data
pointsWage$`Education level` <- factor(pointsWage$`Education level`, levels = c("PRIMAR", "LOWSEC", "UPP

#plot the data
pointsWage %>% ggplot(aes(region, regionalMean, fill = region))+
  geom_col()+
  facet_grid(`Education level` ~ Quantile, labeller = labeller("Education level"=c("PRIMAR" = "Primary"
  gghighlight(regionalMean < eduMean, calculate_per_facet = T)+
  geom_hline(aes(yintercept = eduMean), lty = 3, colour="blue")+
  theme(legend.position = "none")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+
```

```
  labs(x = "Region", y="Mean Rate of Completion", title = "Comparing Wealth with Completion Rate", capti
```