**Data Set:**

**Foundational Learning Skills**

<https://data.unicef.org/wp-content/uploads/2021/04/Foundational-learning_2021-1.xlsx>

**Source:** data.unicef.org

**Description:** “Percentage of children achieving minimum proficiency in (i) reading and (ii) numeracy. If the child succeeds in 1) word recognition, 2) literal questions, and 3) inferential questions, s/he is considered to have foundational reading skills. If the child succeeds in 1) number reading, 2) number discrimination, 3) addition, and 4) pattern recognition, s/he is considered to have foundational numeracy skills.” (source)

**Attributes Considered:**

* Countries and areas

39 different countries and areas

>unique(FoundationalLearning\_2021\_1$’ Countries and areas’)

* Region, Sub-region: of the country/ area

There are 25 unique combinations

> FoundationalLearning\_2021 %>% group\_by(`Countries and areas` , `Region`, `Sub-region`) %>% summarize( n = n())

Graphical user interface, text, application

Description automatically generated

* Development Region: 3 levels: Least Developed, Less Developed, and More Developed

>FoundationalLearning\_2021 %>% group\_by(`Development regions`) %>% summarize( n = n())

# A tibble: 3 x 2

`Development regions` n

<chr> <int>

1 Least Developed 52

2 Less Developed 44

3 More Developed 4

* Age group: 2 categories: Age 7 to 14, Grade 2/3

>FoundationalLearning\_2021 %>% group\_by(`Age group`) %>% summarize( n = n())

# A tibble: 2 x 2

`Age group` n

<chr> <int>

1 Age 7 to 14 50

2 Grade 2/3 50

* Subject: 2 categories: Reading, Numeracy

> FoundationalLearning\_2021 %>% group\_by(`Subject`) %>% summarize( n = n())

# A tibble: 2 x 2

Subject n

<chr> <int>

1 Numeracy 50

2 Reading 50

* Total, is the percentage of sample population that had the foundational skills listed. There is a “Total” value for each country, region, sub-region, development region, age group and subject.

The mean value of all “Total” values listed is 26.2. Later I will look at averages for specific attributes.

>FoundationalLearning\_2021 %>% summarize( avg\_Allpop\_foundation= mean(Total))

* Sex: 2 categories: Girl, Boy

The average foundational skills percentage for girls is 27.2

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Girls))

The average foundational skills percentage value for boys is 25.5

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Boys))

* Residence: 2 categories: Rural, Urban

The average foundational skills percentage for rural is 22.4

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Rural, na.rm = TRUE))

The average foundational skills percentage for Urban is 29.8

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Urban, na.rm = TRUE))

Wealth Quintile: 5 levels: Wealth Levels: Poorest < Second < Middle <Fourth < Richest

The average foundational skills percentage for Poorest is 17.4

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Poorest, na.rm = TRUE))

The average foundational skills percentage for Second is 21.4

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Second, na.rm = TRUE))

The average foundational skills percentage for Middle is 24.2

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Middle, na.rm = TRUE))

The average foundational skills percentage for Fourth is 28.2

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Fourth, na.rm = TRUE))

The average foundational skills percentage for Richest is 37.7

>FoundationalLearning\_2021 %>% summarize( avg\_foundation = mean(Richest, na.rm = TRUE))

* Data Source
* Time period

Read in csv file as tibble:

>FoundationalLearning\_2021 <- read\_csv("Foundational-learning\_2021New.csv")

**Question 1**

Does wealth affect the percentage of foundation learning skills?

>Install.packages(“Hmisc”)

>Library(Hmisc)

>wealth <- FoundationalLearning\_2021 %>% select(Poorest, Second, Middle, Fourth, Richest)

Chart, histogram

Description automatically generated>hist.data.frame(wealth)

**Interpretation**: The lowest wealth level is “Poorest” the histogram for Poorest is skewed right, with the majority of the values being in the lower bounds (0 – 20) for the percentage of sample population with foundational learning skills. As we move up the wealth levels, the values become more normally distributed. This bring me to conclusion that wealth levels and percentage of foundational learning skills are positively correlated.

**Question 2:** Does area have an effect on the ratio of girls to boys foundational learning skills?

>FoundationalLearning\_2021 %>% mutate(GBratio = Girls/Boys)

Graphical user interface, application

Description automatically generated

To answer this question I took the mean GBratio for all countries and areas and will compare

The GBratio had some infinite values, I want to ignore those values:

>FoundationalLearning\_GBratio$GBratio[is.infinite(FoundationalLearning\_GBratio$GBratio)] <- NA

> mean(FoundationalLearning\_GBratio$GBratio, na.rm = TRUE)

[1] 1.045179

I then took the average GBratio for each of the “Countries and areas” (ignoring their specified region and sub-regions):

>FoundationalDF\_GBratio <- FoundationalLearning\_GBratio %>%

select(`Countries and areas`, `GBratio`)

> FoundationalDF\_GBratio %>%

group\_by(`Countries and areas`) %>%

summarise(avg\_GBratio = mean(GBratio, na.rm = TRUE)) %>%

arrange(desc(avg\_GBratio))

Graphical user interface, text, application

Description automatically generated

Now I want to display these averages on a world map:

Using “tidyverse”, “maps” and “mapping” packages.

> MapTibble <- map\_data("world") %>% as\_tibble()

>FoundationalDF\_GBratio %>%

group\_by(`Countries and areas`) %>%

summarise(avg\_GBratio = mean(GBratio, na.rm = TRUE), Compar\_avg = GBratio>avg\_GBratio) %>%

left\_join(MapTibble, by = c("Countries and areas" = "region")) %>%

ggplot(aes(long,lat)) + borders("world") +

geom\_point(aes(color = Compar\_avg)) +

coord\_quickmap()

Compar\_avg represents wether or not the average GBratio, ratio comparing girls to boys foundational learning skills, is above (TRUE) or below (FALSE) the average GBratio.

Map

Description automatically generated

Interpretation: The GBratios that fall below or above the average appear to be congregated together in specific areas; for example in Africa, there are both red and blue points but there is not a circumstance where a blue point falls in the middle of a group of red and vice versa. The same is true for other continents, which brings me to the conclusion that area does have an affect on the ratio of girls to boys with foundational learning skils.

The last question looked at GBratio without differentiating between subject, in the next quesiton I will compare the differnces in the percentage of boys with foundational learning skills vs the percentage of girls with them in regards to “Subject.”

**Quesiton 3:** Is there a differnce in the percentage of reading and numeracy foundational learning skills?

>Reading\_values <- FoundationalLearning\_2021 %>% filter(Subject == "Reading")

>mean(Reading\_values$Total)

[1] 31.6

> Numeracy\_values <- FoundationalLearning\_2021 %>% filter(Subject == "Numeracy")

> mean(Numeracy\_values$Total)

[1] 20.82

>FoundationalLearning\_2021 %>%

group\_by(Subject) %>%

summarise(avg\_total = mean(Total)) %>%

ggplot(aes(Subject, avg\_total)) + geom\_bar(stat = "identity")



Before looking at what factors might influnce the differnces in reading and numeracy foundational skills I first want to visualize the differnce of both.

>numboxplots<- ggplot(FoundationalLearning\_2021, aes(Total)) +

geom\_boxplot() +

facet\_wrap(~Subject)

>numhistoplots <- ggplot(FoundationalLearning\_2021, aes(Total))+

geom\_histogram() +

facet\_wrap(~Subject)

>grid.arrange(numhistoplots, numboxplots, ncol = 2)

Chart

Description automatically generated

Now that I now the avergage numeracy foundational learning skills are lower than the average reading foundational learning skills I want to see how attribute values of each differ that might explain why the numeracy average is lower.

**Quesiton 4:** Which factors influence the differnce in reading and numeracy foundational learning skills?

To explore this I made a scatterplot matrix for numeric attributes that fall within each the Numeracy and Reading subject. I did this for myself to compare the first coloumn in each matrix, but I don’t think it looks easy to understand visually so I will also create a scatterplot for each numeric attribute and shade the points according to subject as well.

> ggpairs(Reading\_values[,7:16])

Chart

Description automatically generated with medium confidence

Chart

Description automatically generated>ggpairs(Numeracy\_values[,7:16])

The following is just code for scatterplots of each of the numeric attributes compared to “Total”. I assigned them all to a variable and then wanted to generate a code for displaying them all at once:

> Girls <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Girls, Total, color = Subject))

> Boys <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Boys, Total, color = Subject))

> Rural <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Rural, Total, color = Subject))

> Urban <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Urban, Total, color = Subject))

> Poorest <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Poorest, Total, color = Subject))

> Second <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Second, Total, color = Subject))

> Middle <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Middle, Total, color = Subject))

> Fourth <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Fourth, Total, color = Subject))

> Richest <- ggplot(FoundationalLearning\_2021) + geom\_point(aes(Richest, Total, color = Subject))

> grid.arrange(Girls + ggtitle("Girls"), Boys + ggtitle("Boys"),

Rural + ggtitle("Rural"), Urban + ggtitle("Urban"),

Poorest + ggtitle("Poorest"), Second + ggtitle("Second"),

Middle + ggtitle("Middle"), Fourth + ggtitle("Fourth"),

Richest + ggtitle("Richest"), nrow = 5)

Chart

Description automatically generated

> lm(Total ~ Girls + Boys + Rural + Urban + Poorest + Second + Middle + Fourth + Richest, data = Numeracy\_values)

Call:

lm(formula = Total ~ Girls + Boys + Rural + Urban + Poorest +

Second + Middle + Fourth + Richest, data = Numeracy\_values)

Coefficients:

(Intercept) Girls Boys Rural Urban Poorest Second Middle Fourth

-9.310e-02 5.163e-01 4.779e-01 -1.272e-02 -3.673e-02 -7.171e-05 3.867e-03 -5.108e-03 2.027e-02

Richest

3.071e-02

> lm(Total ~ Girls + Boys + Rural + Urban + Poorest + Second + Middle + Fourth + Richest, data = Reading\_values)

Call:

lm(formula = Total ~ Girls + Boys + Rural + Urban + Poorest +

Second + Middle + Fourth + Richest, data = Reading\_values)

Coefficients:

(Intercept) Girls Boys Rural Urban Poorest Second Middle Fourth

0.053531 0.445226 0.427330 -0.006019 0.003812 0.034814 0.005862 0.041604 0.026541

Richest

0.016643

In comparing the linear regressions I can see that there are multiple attributes negativley affecting the Total percenatage of foundational learning skills; the most significant being “Poorest” from the wealth quintile, and the most significant from the residence quintile being “Urban.” With this I can conclude that wealth levels influence the difference in reading and numeracy foundaitonal learning percentages.

**Quesiton 5:** Does area have an effect on the ratio of numeracy to reading foundational learning skills?