IE6600_Sec03_Group08_Hackathon

Group 08

3/2/2021

1.1 Overview

The UNICEF's Dataset contains the key issues affecting the State of the World's Children. We are going to assess the situation of children and women in the areas of education, health, gender equality, rights and protection in form of visualization using R software and try to gain some insight from those visualization to compare the situation of the children and women in the world right now based on different factors which affect them the most.

1.2 Problem statement

Our objective is to find the disparities in school enrollment across various regions, compare child labor and violence discipline with literacy rates achieved . In this report, the main objective is split into 3 sections. First Section identifies Disparities in school enrollment at various levels and identifies enrollment based on gender. Second section talks on how child labour is associated with lower secondary school enrollment and gender. Section 3 visualizes violent discipline and its relation with literacy.

```
knitr::opts_chunk$set(warning = FALSE,message = FALSE)
```

```
library(gridExtra)
library(tidyverse)
library(dplyr)
library(ggplot2)
library(reshape2)
library(extrafont)
library(tidyr)
library(data.table)
library(scales)
library(readxl)
library(stringr)
library(ggalluvial)
```

```
Data <- read_excel("Hackathon_Unicef_Dataset.xlsx", sheet = "Data" )
Data<-transform(Data, Value=as.numeric(Value))
```

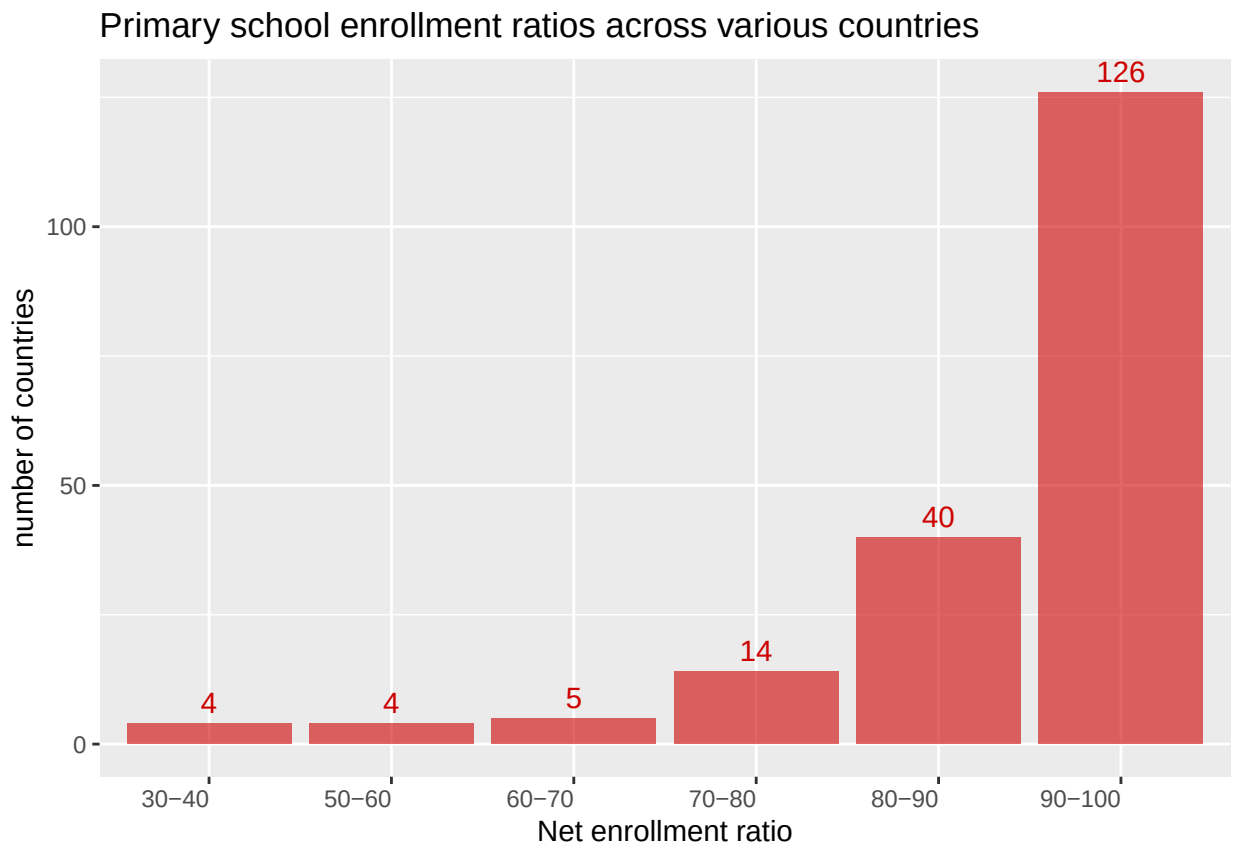
Section 1

```
req_indicators <- Data %>%
  filter(Data$Indicator == ("Pre-primary school participation - Gross enrolment ratio (%), male") |
  Data$Indicator ==("Pre-primary school participation - Gross enrolment ratio (%), female")|
           Data$Indicator==("Primary school participation - Net enrolment ratio (%), male") |
           Data$Indicator == ("Primary school participation - Net enrolment ratio (%), female") |
           Data$Indicator ==("Lower secondary school participation - Net enrolment ratio (%), male") |
           Data$Indicator == ("Lower secondary school participation - Net enrolment ratio (%), female")
           Data$Indicator==("Primary school participation - Out-of-school rate of children of primary s
           Data$Indicator==("Primary school net attendance ratio, Urban") |
           Data$Indicator==("Primary school net attendance ratio, Rural")
         )
req_regions <- req_indicators %>%
  filter((Countries.and.areas=='Central African Republic') |
        (Countries.and.areas=='East Asia and Pacific')  |
        (Countries.and.areas=='Europe and Central Asia') |
        (Countries.and.areas=='Eastern Europe and Central Asia') |
        (Countries.and.areas=='Western Europe') |
        (Countries.and.areas=='Middle East and North Africa') |
        (Countries.and.areas=='Eastern and Southern Africa') |
        (Countries.and.areas=='West and Central Africa')
          )
pivot_Data <- req_regions %>%
  select(Countries.and.areas,Indicator, Value) %>%
  pivot_wider(names_from = Indicator, values_from = Value )
# Selecting required columns
req_df <- Data %>%
  filter(Data$Indicator == ("Total adult literacy rate (%)") | Data$Indicator ==("Primary school net en
           Data$Indicator==("Pre-primary school participation - Gross enrolment ratio (%), male") |
           Data$Indicator == ("Pre-primary school participation - Gross enrolment ratio (%), female") |
           Data$Indicator ==("Primary school participation - Gross enrolment ratio (%), male") |
           Data$Indicator == ("Primary school participation - Gross enrolment ratio (%), female") |
           Data$Indicator == ("Primary school participation - Out-of-school rate of children of primary
           Data$Indicator==("Primary school participation - Out-of-school rate of children of primary s
           Data$Indicator == ("Primary school net enrolment ratio (%)")
         )
# Pivoting the Data on Indicators
p_Data <- req_df %>%
  select(Countries.and.areas,Indicator, Value) %>%
  pivot_wider(names_from = Indicator, values_from = Value )
p_Data <- p_Data %>%
  rename(out_of_school_male=`Primary school participation - Out-of-school rate of children of primary s
  rename(out_of_school_female=`Primary school participation - Out-of-school rate of children of primary
```

Most of the countries achieved literacy yet almost 40% of the countries are still under the average literacy rate of 80.

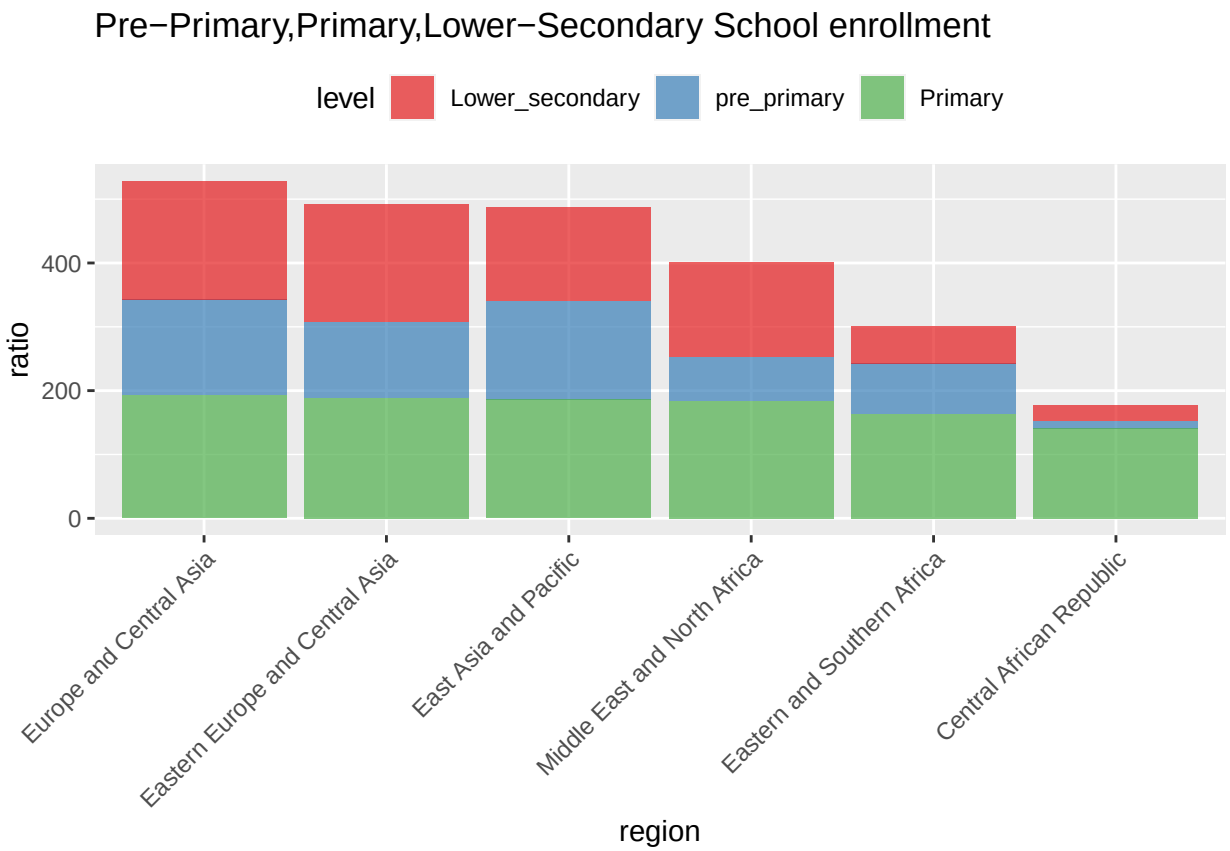
```
p_Data$newcol<-0
colnames(p_Data)[15] <- "primary_school_net_enrollment_ratio"
p_Data$primary_school_net_enrollment_ratio <- cut(p_Data$`Primary school net enrolment ratio (%)`, c(0,
d4<-p_Data %>%
  group_by(primary_school_net_enrollment_ratio) %>%
  summarise(number_of_countries= n()) %>%
  drop_na()
```

```
chart_6 <-d4 %>%
  ggplot(aes(x =primary_school_net_enrollment_ratio , y = number_of_countries)) +
    geom_bar(stat = 'identity', fill='red3', alpha=0.6) + theme(axis.text.x = element_text(angle=0, hjus
    labs(y="number of countries", x = "Net enrollment ratio") +labs(title= "Primary school enrollment ra
chart_6
```



From the above chart, we see that in the 4 countries, the percentage of children enrolled in primary school is less than 50 per cent. These countries belong to african sub continent

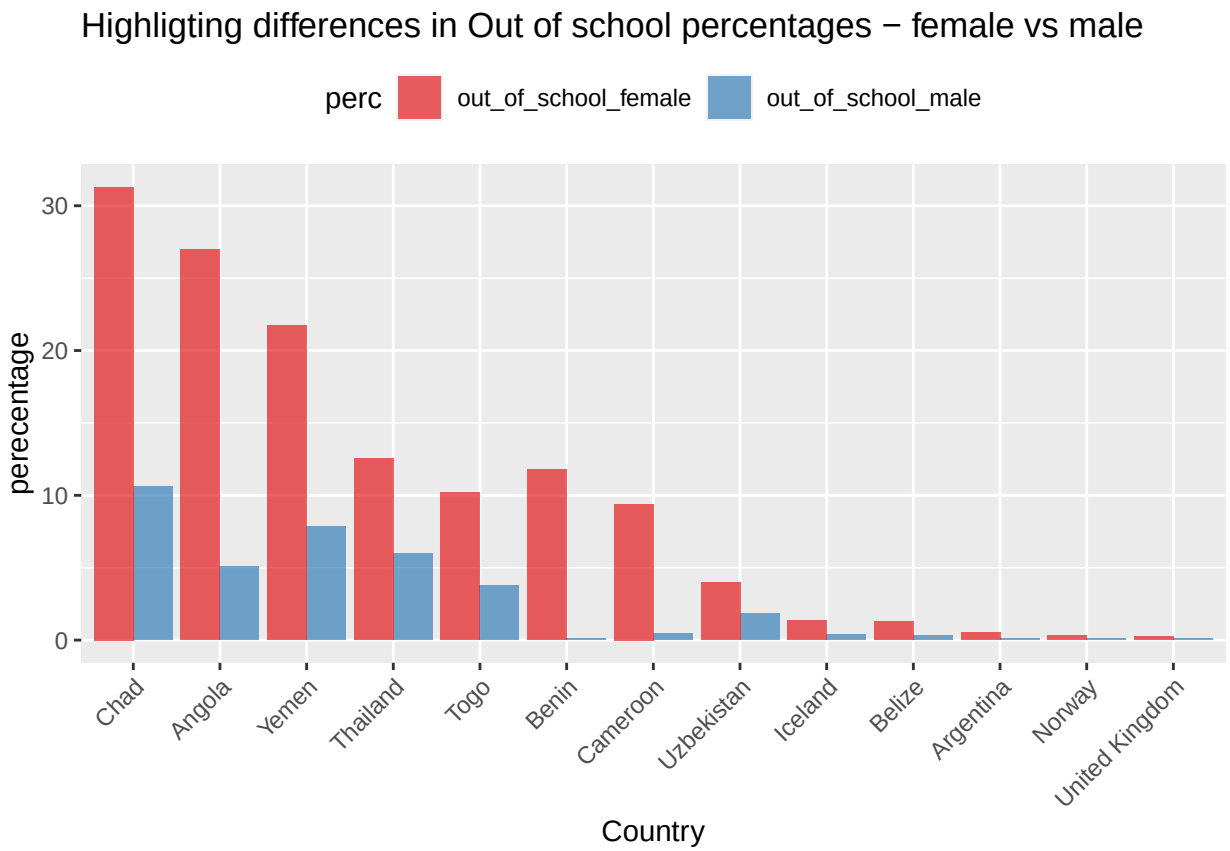
```
d<- mutate(pivot_Data, pre_primary = `Pre-primary school participation - Gross enrolment ratio (%), mal
             `Pre-primary school participation - Gross enrolment ratio (%), female`)
d1<- mutate(d, Primary = `Primary school participation - Net enrolment ratio (%), male` +
             `Primary school participation - Net enrolment ratio (%), female`)
d2<- mutate(d1, Lower_secondary = `Lower secondary school participation - Net enrolment ratio (%), male
             `Lower secondary school participation - Net enrolment ratio (%), female`)

chart_2<- d2 %>%
  select(Countries.and.areas, pre_primary, Primary, Lower_secondary) %>%
  drop_na() %>%
  gather(level, ratio, 2:4) %>%
  ggplot(aes(x = reorder(Countries.and.areas, -ratio) , y = ratio , fill = level)) +
    geom_bar(stat = 'identity', alpha=0.7) + theme(axis.text.x = element_text(angle=45, hjust = 1)) + tl
    labs(y="ratio", x = "region") +labs(title= "Pre-Primary,Primary,Lower-Secondary School enrollment")
chart_2
```



Ratio of children enrolling in Lower Secondary school is less in regions involving African countries. In East and Central Africa, number of children enrolling in school after primary level has gradually decreased.

```
df1<-p_Data %>%
  select(Countries.and.areas, out_of_school_male, out_of_school_female)
df1<- mutate(df1, out_of_school_ratio_female_to_male= out_of_school_female/out_of_school_male )
index1<- which(df1$out_of_school_ratio_female_to_male>=2 )
df2 <- df1[index1, c(1:4)]
```

```
chart_1<- df2 %>%
  select(Countries.and.areas, out_of_school_male, out_of_school_female) %>%
  drop_na() %>%
  arrange(desc(out_of_school_female)) %>%
  top_n(20, out_of_school_female) %>%
  gather(perc, n, 2:3) %>%
  ggplot(aes(x = reorder(Countries.and.areas, -n) , y = n , fill = perc)) +
    geom_bar(stat = 'identity', position = 'dodge', alpha=0.7) + theme(axis.text.x = element_text(angle=
    labs(y="perecentage", x = "Country", color="Gender") +labs(title= "Highligting differences in Out o

chart_1
```

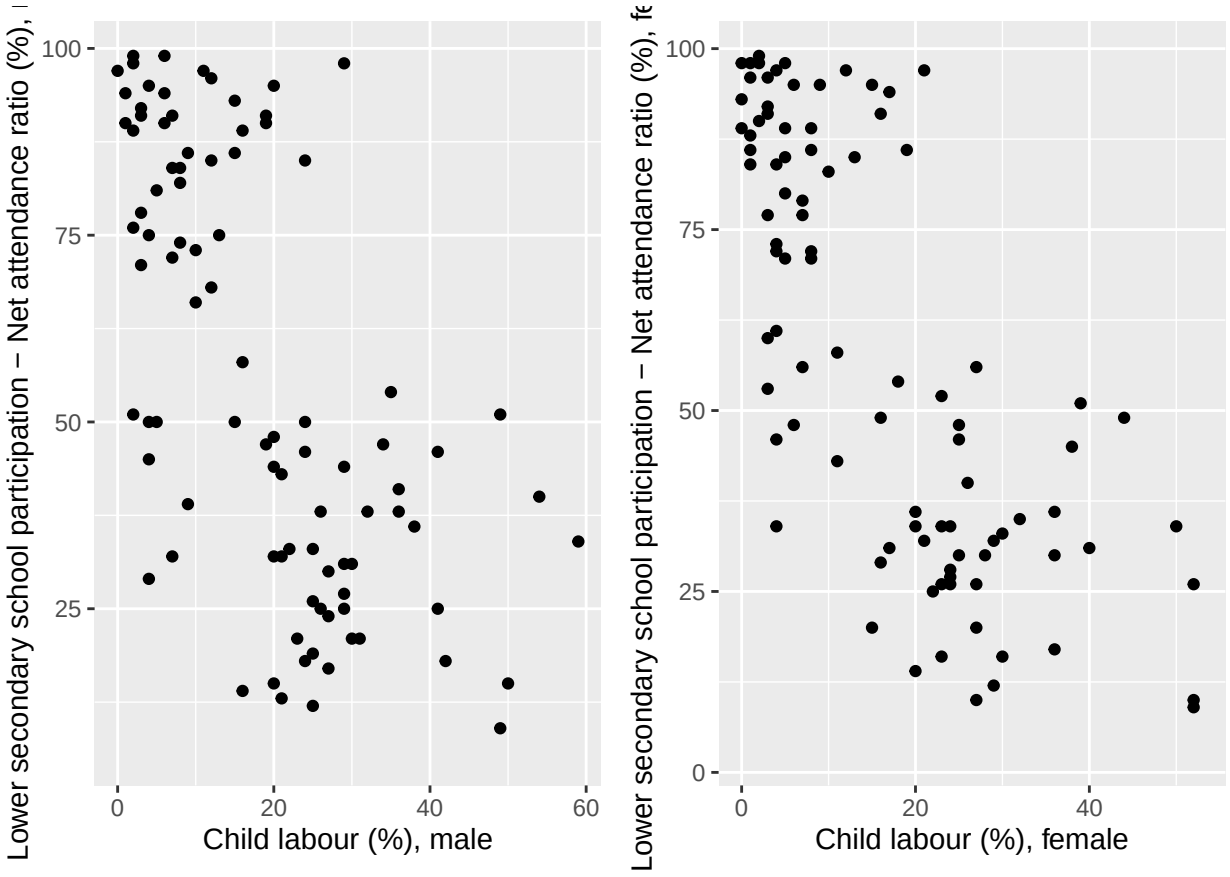


Least developed countries have a high percentages of children not attending schools. If we compare enrollment of male and females in school, more females are out of schools when compared to males. These differences are obvious in the below countries. In these countries, out of 3 children, there at least 2 girls that are not attending school. This stands true not just for less developed nations but also to highly developed ones such as UK and Norway. Though the out-of-school percentages are really low in these nations, it still holds true.

Section 2

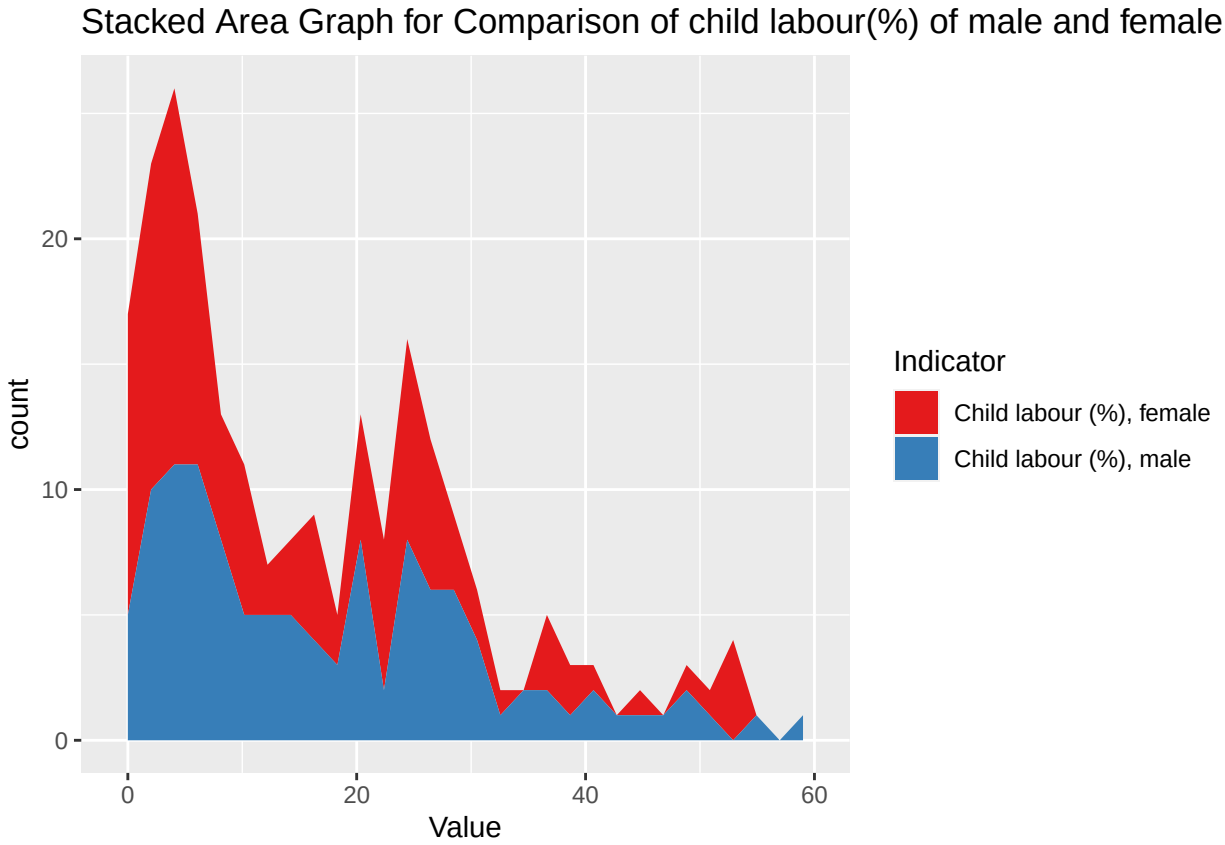
```
df_partchildm <- Data %>%
  filter(Data$Indicator ==("Lower secondary school participation - Net attendance ratio (%), male") | Da
df_partchildm1 <- df_partchildm %>%
    dplyr::select("Countries.and.areas",Indicator,Value)
df_partchildm1 <- transform(df_partchildm1, Value = as.integer(Value))
df_partchildm1 <- df_partchildm1 %>%
        pivot_wider(names_from = Indicator, values_from = Value )
partchildm <- ggplot(df_partchildm1, aes(y=`Lower secondary school participation - Net attendance ratio
,x=`Child labour (%), male`))  + geom_point()
df_partchildf <- Data %>%
  filter(Data$Indicator ==("Lower secondary school participation - Net attendance ratio (%), female") |
df_partchildf1 <- df_partchildf %>%
    dplyr::select("Countries.and.areas",Indicator,Value)
df_partchildf1 <- transform(df_partchildf1, Value = as.integer(Value))
df_partchildf1 <- df_partchildf1 %>%
```

```
        pivot_wider(names_from = Indicator, values_from = Value )
partchildf <- ggplot(df_partchildf1, aes(y=`Lower secondary school participation - Net attendance ratio
grid.arrange(partchildm,partchildf,ncol = 2)
```



The above chart is a scatter plot between the "Lower secondary school participation - Net attendance ratio (%)" and "Child labour (%)" for male and female. The plot is approximately same for male and female regardless of their gender.We can conclude from the plot that the country with the highest school participation has the lowest child labour and vice versa.

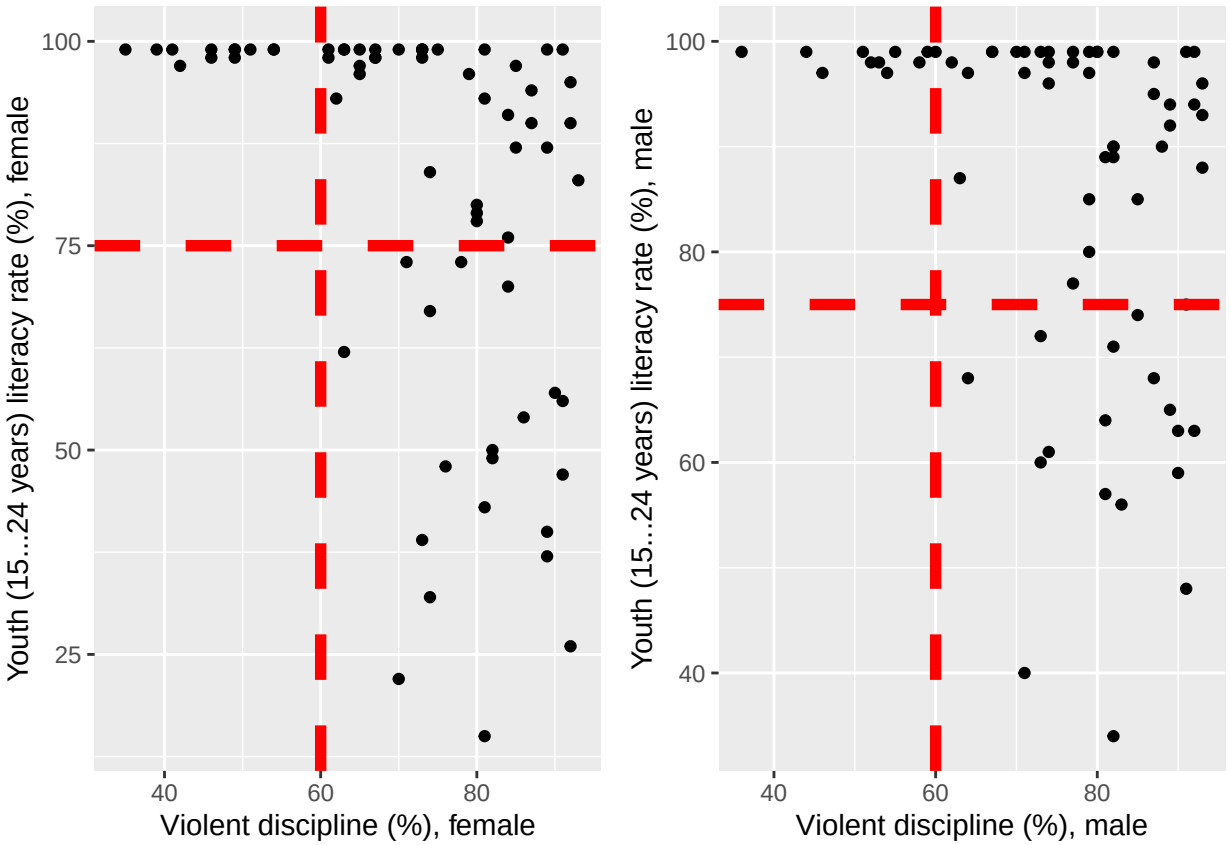
```
df_childfm <- Data %>%
  filter(Data$Indicator == ("Child labour (%), female") | Data$Indicator == ("Child labour (%), male"))
df_childfm1 <- df_childfm %>%
    dplyr::select("Countries.and.areas",Indicator,Value)
df_childfm1 <- transform(df_childfm1, Value = as.integer(Value))
area_childfm1 <- ggplot(df_childfm1, aes(x = Value, fill=Indicator)) +
           geom_area(stat="bin") +
           ggtitle("Stacked Area Graph for Comparison of child labour(%) of male and female")

area_childfm1 + scale_fill_brewer(palette="Set1")
```



The is the stacked area graph for camparison of child labour(%) of male and female. The height of each coloured stack represents the proportion of the child labour at any given point.This represents the data from all the countries and we can see there is a spike in the female child labour at the start of the graph.

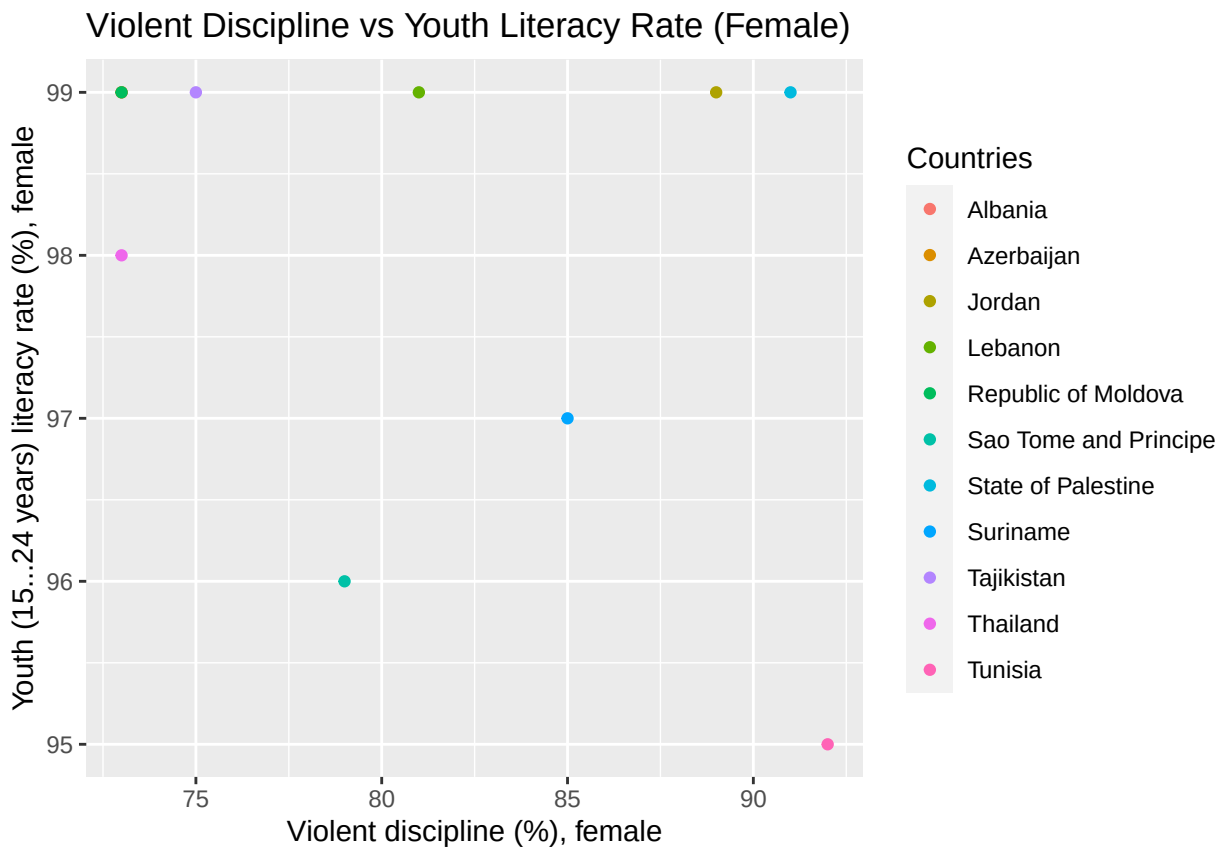
Section 3

```
df_discf <- Data %>%
  filter(Data$Indicator == ("Violent discipline (%), female") | Data$Indicator ==("Youth (15-24 years) l
df_discf1 <- df_discf %>%
   select(`Countries.and.areas`,Indicator,Value)
df_discf1 <- transform(df_discf1, Value = as.integer(Value))
discf1 <- df_discf1 %>%
        pivot_wider(names_from = Indicator, values_from = Value )
ggdiscf <- ggplot(discf1, aes(y=`Youth (15-24 years) literacy rate (%), female`
,x=`Violent discipline (%), female`))  + geom_point() + geom_vline(xintercept = 60 , color = "red", lin
df_discm <- Data %>%
  filter(Data$Indicator == ("Violent discipline (%), male") | Data$Indicator ==("Youth (15-24 years) li
df_discm1 <- df_discm %>%
   select(`Countries.and.areas`,Indicator,Value)
df_discm1 <- transform(df_discm1, Value = as.integer(Value))
discm1 <- df_discm1 %>%
        pivot_wider(names_from = Indicator, values_from = Value )
ggdiscm <- ggplot(discm1, aes(y=`Youth (15-24 years) literacy rate (%), male`
,x=`Violent discipline (%), male`))  + geom_point() + geom_vline(xintercept = 60 , color = "red", linet
grid.arrange(ggdiscf,ggdiscm,ncol= 2)
```



Both male and females are subjected to physical punishment or psychological aggression equally. Most countries that adopt violent discipline also happen to be the the countries with good literacy rates.

```
discf2 <- discf1 %>%
  filter(discf1$`Youth (15-24 years) literacy rate (%), female` >= 95)
discf2 <- discf2 %>%
  filter(discf2$`Violent discipline (%), female` > 70)
ggdiscf2 <- ggplot(discf2, aes(y=`Youth (15-24 years) literacy rate (%), female`
,x=`Violent discipline (%), female`,color = `Countries.and.areas`))  + geom_point()  +labs(title= "Viol
ggdiscf2
```



This is the scatter plot of "Youth (15–24 years) literacy rate (%)" Vs. "Violent Discipline (%). female". The violent discipline for female is more than 70% despite being the literacy rate more than 95% in these countries.

Conclusion

While the average literacy for countries around the world is 80, 40 percent of the countries are still below this average. These countries are mainly from the African continent. In 4 of these regions, less than 50 percent of the children attend primary schools. Moreover, the number of children continuing lower secondary education is less than from primary. Other than in some of the African regions, urban and rural demographics do not seem to impact the enrollment in primary education. Also, we noticed that more females are of out school compared to males. The country with the highest school participation has the lowest child labor and vice versa. Both males and females in the age group 1-14 are subjected to physical punishment or psychological aggression equally. Countries such as Benin, Chad, Guinea-Bissau, Mali, and Somalia have child labor more than 50(%) for males as well as females. Most countries that adopt violent discipline also happen to be the countries with good literacy rates. The violent discipline for females is more than 70% despite being the literacy rate of more than 90% in these countries.