EDA with R | World Literacy Rates

Purnima Lal

03/06/2021

Download dataset on World Literacy Rates (UNESCO) – https://query.data.world/s/h7re2l76gr4we7s4xp7o5j4s5ecfbd

View dataset details – https://data.world/makeovermonday/2019w46

Loading Relevant Packages

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse
1.3.1 ──
```

```
## ✓ggplot2 3.3.3     ✓purrr   0.3.4
## ✓tibble  3.1.2     ✓dplyr   1.0.6
## ✓tidyr   1.1.3     ✓stringr 1.4.0
## ✓readr   1.4.0     ✓forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(httr)
library(readxl)
library(extrafont)
```

```
## Registering fonts with R
```

```
library(hrbrthemes)
library(gghighlight)
library(ggtext)
library(ggmap)

## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.

## Please cite ggmap if you use it! See citation("ggmap") for details.

library(scales)

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor

library(tidytext)
```

Loading & reading the data

```
Literacy_rates <- read_excel("/home/purnima/Downloads/Literacy rates.xlsx")

View(Literacy_rates)

literacy_data<- Literacy_rates

##
## ── Column specification
──────────────────────────────────────────────────────
## cols(
##   Region = col_character(),
##   Country = col_character(),
##   Year = col_double(),
##   Age = col_character(),
##   Gender = col_character(),
##   `Literacy rate` = col_character()
## )
```

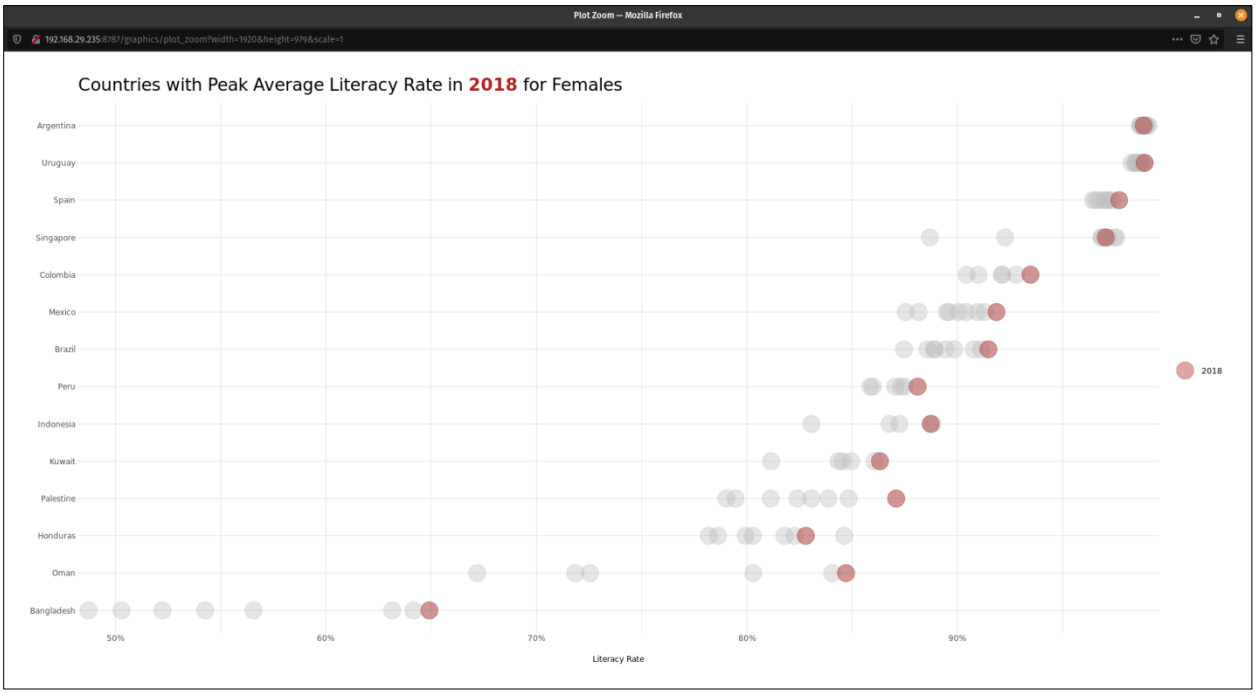
Viewing data

```
literacy_data
```

```
## # A tibble: 4,955 x 6
##    Region                    Country      Year Age   Gender `Literacy
rate`
##    <chr>                     <chr>       <dbl> <chr> <chr>
<dbl>
##  1 Central and Southern Asia Afghanistan  2011 15+   female          0.176
##  2 Central and Southern Asia Afghanistan  2011 15+   male            0.454
##  3 Central and Southern Asia Afghanistan  2011 15+   total           0.317
##  4 Central and Southern Asia Afghanistan  2011 15-24 female          0.321
##  5 Central and Southern Asia Afghanistan  2011 15-24 male            0.619
##  6 Central and Southern Asia Afghanistan  2011 15-24 total           0.470
##  7 Central and Southern Asia Afghanistan  2011 25-64 female
0.0841
##  8 Central and Southern Asia Afghanistan  2011 25-64 male            0.378
##  9 Central and Southern Asia Afghanistan  2011 25-64 total           0.231
## 10 Central and Southern Asia Afghanistan  2011 65+   female          0.197
## # … with 4,945 more rows
```

01. Countries with the highest average literacy rate in 2018 for women-

```
literacy_data %>%
  filter(Gender=="female") %>%
  group_by(Country) %>%
  filter(max(Year)==2018 & n_distinct(Year)>=5,
         mean(`Literacy rate`[Year==2018],na.rm = T) > mean(`Literacy
rate`[Year!=2018],na.rm = T)) %>%
  group_by(Country,Year) %>%
  summarise(Average_literacy_rate = mean(`Literacy rate`,na.rm = T), .groups
= 'drop') %>%


ggplot(aes(reorder(Country,Average_literacy_rate,mean),Average_literacy_rate,
color=factor(Year)))+
  geom_point(size=9,alpha=0.4)+
  coord_flip()+
  gghighlight(Year==2018)+
  scale_y_percent()+
  scale_color_manual(values = c("2018"="firebrick"))+
  theme_minimal()+
  labs(x=NULL,y="Literacy Rate",title = "Countries with Peak Average Literacy
Rate in <span style='color:firebrick'>**2018**</span> for Females",
       color=NULL)+
  theme_ipsum_ps()+
  theme(axis.title.x = element_text(margin = margin(t = 15), hjust = 0.5),
        plot.title = element_markdown(size=20,margin = margin(b = 10)))
```



This visual shows countries with the highest average literacy rate in 2018 for females compared to the average of all previous years. Quite a few countries here from South America as well as Argentine staying consistent around the 90% mark. Other LATAM Countries like Brazil, Mexico, and Colombia have shown good growth. Bangladesh has seen a phenomenal increase in female literacy rates.

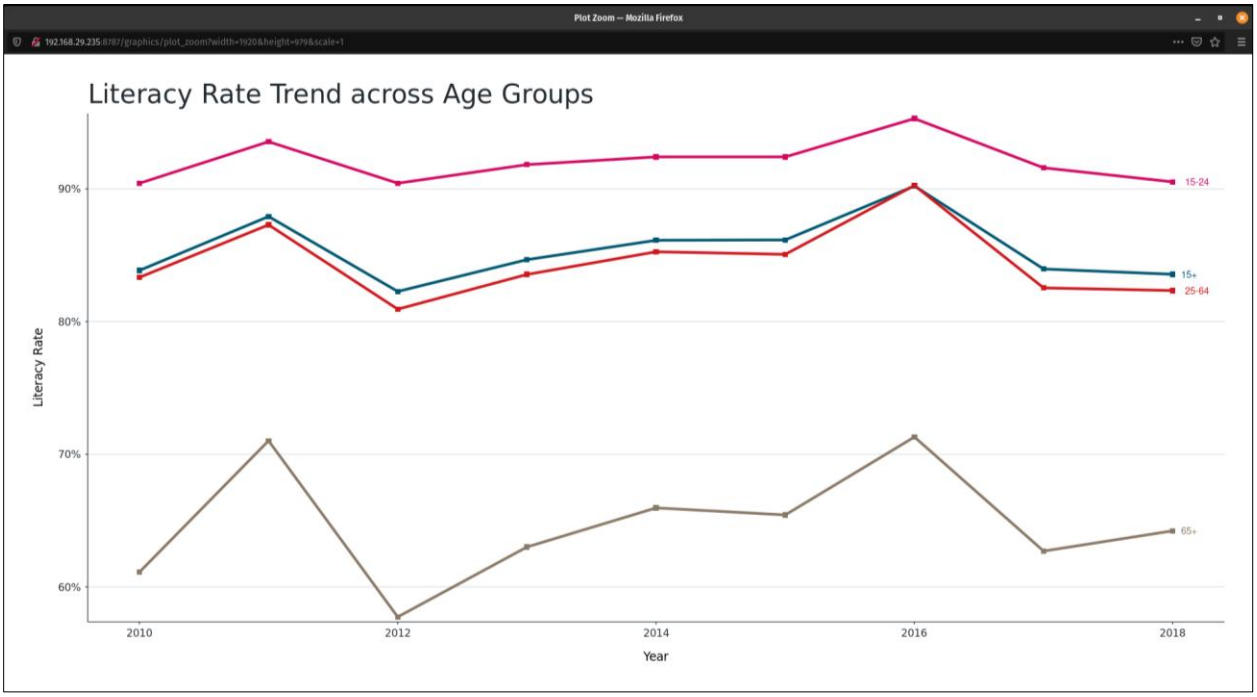
02. Distribution of literacy rate across regions by age

```
world_avg <- literacy_data %>%
  summarise(avg=mean(`Literacy rate`,na.rm = T)) %>%
  pull(avg)

literacy_data %>%
  group_by(Region) %>%
  mutate(region_avg= mean(`Literacy rate`,na.rm = T)) %>%
  ungroup() %>%
  ggplot(aes(Region,`Literacy rate`,color=Age))+
  geom_jitter(alpha=0.7,size=3,height  = 0.2)+
  geom_hline(aes(yintercept=world_avg),color="grey40",size=0.9)+

geom_segment(aes(x=Region,xend=Region,y=world_avg,yend=region_avg),color="bla
ck")+
  coord_flip()+
  ggsci::scale_color_jama()+
  stat_summary(fun = mean, geom = "point", size = 12,color="firebrick")+

geom_text(aes(x=Region,y=region_avg,label=scales::percent(region_avg)),color=
"white",hjust=0.5,nudge_y = 0.01)+
  theme_classic()+
  theme(text = element_text(family = "Roboto Condensed"),axis.title =
element_text(size = 12),
        axis.text.y = element_text(family = "Roboto Condensed", size = 12),
        panel.grid = element_blank(),
        plot.title = element_text(size = 25,hjust = 0.5,family = "Roboto
Condensed"))+
  annotate(geom ="text",x=7.5,y=0.81,label=paste0("Worldwide literacy average
of ",scales::percent(world_avg)),color="black")+
  scale_y_percent()+
  labs(x=NULL,title = "Distribution of Literacy Rate across Regions")
```

The grey vertical line represents the global literacy rate average which is 82% and those big red circles highlight the average literacy rate of each region. Regions like Sub-Saharan Africa, Central and Southern Asia have a lower average literacy rate compared to the global average and other regions. Also, not surprising to see that people belonging to the older age group generally have a lower literacy rate.

03. Literacy rate across age groups

```
avg_year <- literacy_data %>%
  group_by(Year,Age) %>%
  summarise(avg=mean(`Literacy rate`,na.rm = T, .groups='drop')) %>%
  ungroup()

avg_year %>%
  ggplot(aes(Year,avg,color=Age))+
  geom_line(size=1.5)+
  geom_point(size = 2.6, aes(color = Age), shape = 15) +
  geom_text(data=avg_year %>% group_by(Age) %>%
filter(Year==max(Year)),aes(label=Age),hjust=-0.5)+
  scale_color_manual(values = c("15-24"="#d20962","15+"="#005670","25-
64"="#ce181e","65+"="#8a7967"))+
  scale_y_percent()+
  labs(y="Literacy Rate",x="Year",title = "Literacy Rate Trend across Age
Groups")+
  theme(
    text = element_text(family = "Roboto Condensed"),
    plot.margin = unit(rep(1.2, 4), "cm"),
    plot.title = element_text(size = 30,
                              color = "#22292F",
                              margin = margin(b = 5)),
    plot.subtitle = element_text(size = 15,
                                 margin = margin(b = 35)),
    plot.caption = element_text(size = 10,
                                margin = margin(t = 25),
                                color = "#606F7B"),
    panel.background = element_blank(),
    axis.text = element_text(size = 12, color = "#22292F"),
    axis.text.x = element_text(margin = margin(t = 5)),
    axis.text.y = element_text(margin = margin(r = 5)),
    axis.line = element_line(color = "#3D4852"),
    axis.title = element_text(size = 14),
    axis.title.y = element_text(margin = margin(r = 15),
                                hjust = 0.5),
    axis.title.x = element_text(margin = margin(t = 15),
                                hjust = 0.5),
    panel.grid.major = element_line(color = "#DAE1E7"),
    panel.grid.major.x = element_blank(),
    legend.position = "none"
  )
```

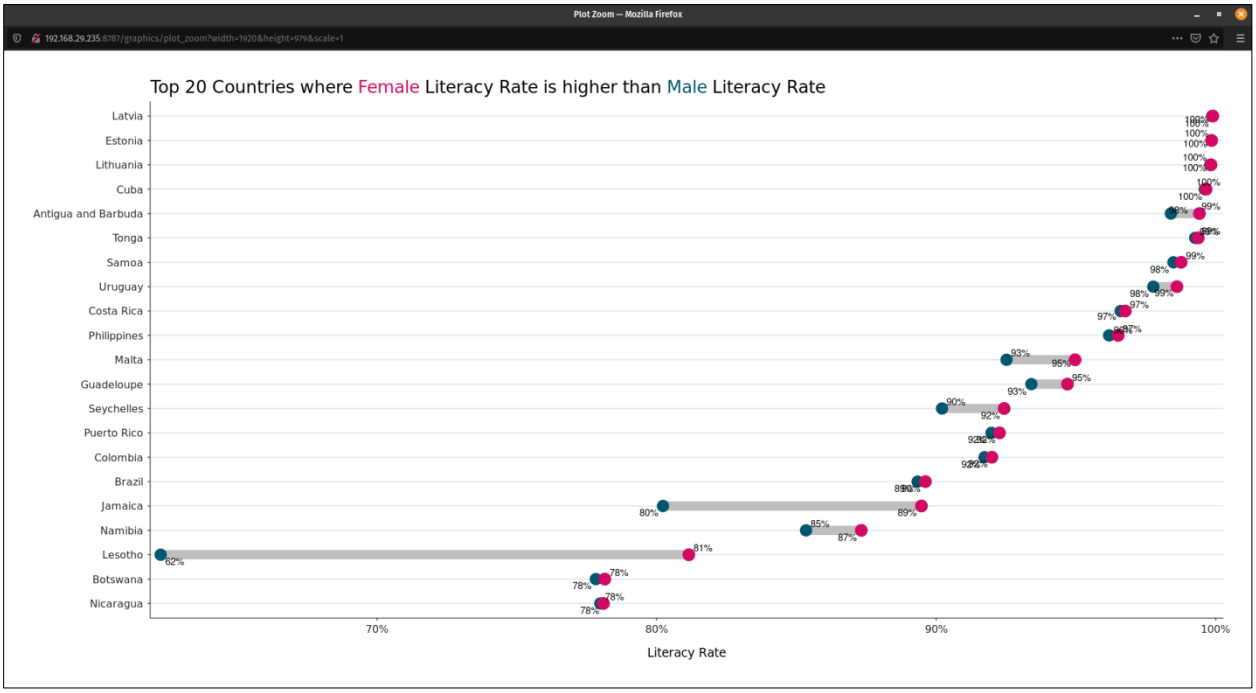


It can be seen that there is a sharp decline in literacy rate for all of the age groups in 2011-12 as well as in 2016-17 while there is a significant rise in literacy rate for all age groups in 2015-16.

04. Countries where the average female literacy rate is greater than the average male literacy rate

```
countries_female <- literacy_data %>%
  group_by(Country) %>%
  filter(mean(`Literacy rate`[Gender=="female"],na.rm = T)>mean(`Literacy
rate`[Gender=="male"]))

literacy_data %>%
  semi_join(countries_female) %>%
  group_by(Country) %>%
  summarise(avg_ltrcy_male = mean(`Literacy rate`[Gender=='male'],na.rm = T),
            avg_ltrcy_female=mean(`Literacy rate`[Gender=='female'],na.rm =
T)) %>%
  ungroup() %>%

ggplot(aes(y=reorder(Country,avg_ltrcy_female),x=avg_ltrcy_male,xend=avg_ltrc
y_female))+
  ggalt::geom_dumbbell( size=5, colour="grey",colour_x =
"#005670",colour_xend = "#d20962")+


ggrepel::geom_text_repel(aes(x=avg_ltrcy_female,label=percent(avg_ltrcy_femal
e,accuracy = 1)))+

ggrepel::geom_text_repel(aes(x=avg_ltrcy_male,label=percent(avg_ltrcy_male,ac
curacy = 1)))+
  labs(x="Literacy Rate", y=NULL, title = "Top 20 Countries where <span
style='color:#d20962'>Female</span> Literacy Rate is higher than <span
style='color:#005670'>Male</span> Literacy Rate")+
  scale_x_percent()+
  theme_classic()+
  theme(
    text = element_text(family = "Roboto Condensed"),
    plot.margin = unit(rep(1.2, 4), "cm"),
    plot.title = element_markdown(size = 20,margin = margin(b = 5)),
    plot.subtitle = element_text(size = 15,
                                 margin = margin(b = 35)),
    plot.caption = element_text(size = 10,
                                margin = margin(t = 25),
                                color = "#606F7B"),
    panel.background = element_blank(),
    axis.text = element_text(size = 12, color = "#22292F"),
```

```
    axis.text.x = element_text(margin = margin(t = 5)),
    axis.text.y = element_text(margin = margin(r = 5)),
    axis.line = element_line(color = "#3D4852"),
    axis.title = element_text(size = 14),
    axis.title.y = element_text(margin = margin(r = 15),
                                hjust = 0.5),
    axis.title.x = element_text(margin = margin(t = 15),
                                hjust = 0.5),
    panel.grid.major = element_line(color = "#DAE1E7"),
    panel.grid.major.x = element_blank(),
    legend.position = "none")

## Joining, by = c("Region", "Country", "Year", "Age", "Gender", "Literacy
rate")

## Registered S3 methods overwritten by 'ggalt':
##   method                  from
##   grid.draw.absoluteGrob  ggplot2
##   grobHeight.absoluteGrob ggplot2
##   grobWidth.absoluteGrob  ggplot2
##   grobX.absoluteGrob      ggplot2
##   grobY.absoluteGrob      ggplot2
```

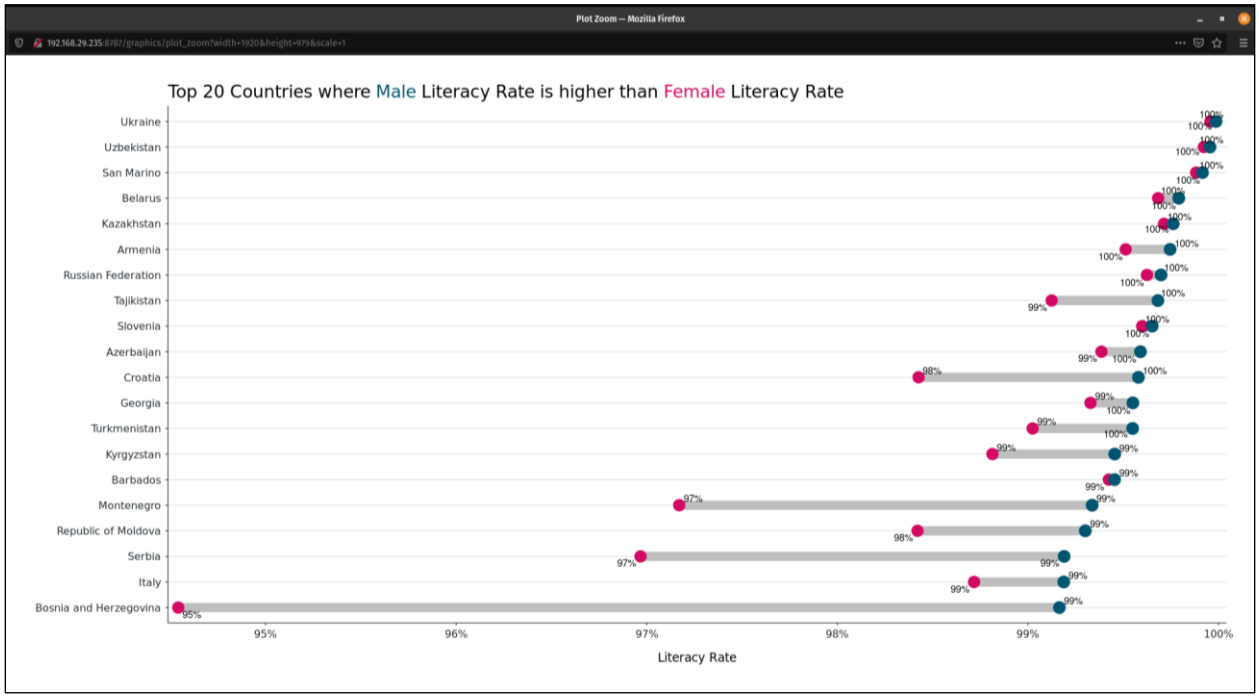


I was curious to see the countries whose average female literacy rate is greater or equal to their male average literacy rate. Most of the countries on this graph are small countries belonging to the South American and African regions which is really surprising.

05. Countries where the average male literacy rate is greater than the average female literacy rate

```
countries_male <- literacy_data %>%
  group_by(Country) %>%
  filter(mean(`Literacy rate`[Gender=="male"],na.rm = T)>mean(`Literacy
rate`[Gender=="female"]))

literacy_data %>%
  semi_join(countries_male) %>%
  group_by(Country) %>%
  summarise(avg_ltrcy_male = mean(`Literacy rate`[Gender=='male'],na.rm = T),
            avg_ltrcy_female = mean(`Literacy rate`[Gender=='female'],na.rm =
T))%>%
  top_n(20,avg_ltrcy_male)%>%
  top_n(20,avg_ltrcy_female)%>%
  ungroup() %>%

ggplot(aes(y=reorder(Country,avg_ltrcy_male),x=avg_ltrcy_female,xend=avg_ltrc
y_male))+
  ggalt::geom_dumbbell( size=5, colour="grey",colour_x =
"#d20962",colour_xend = "#005670")+

ggrepel::geom_text_repel(aes(x=avg_ltrcy_male,label=percent(avg_ltrcy_male,ac
curacy = 1)))+

ggrepel::geom_text_repel(aes(x=avg_ltrcy_female,label=percent(avg_ltrcy_femal
e,accuracy = 1)))+
  labs(x="Literacy Rate",y=NULL,title = "Top 20 Countries where <span
style='color:#005670'>Male</span> Literacy Rate is higher than <span
style='color:#d20962'>Female</span> Literacy Rate")+
  scale_x_percent()+
  theme_classic()+
  theme(
    text = element_text(family = "Roboto Condensed"),
    plot.margin = unit(rep(1.2, 4), "cm"),
    plot.title = element_markdown(size = 20,margin = margin(b = 5)),
    plot.subtitle = element_text(size = 15,
                                 margin = margin(b = 35)),
    plot.caption = element_text(size = 10,
                                margin = margin(t = 25),
                                color = "#606F7B"),
    panel.background = element_blank(),
```

```
    axis.text = element_text(size = 12, color = "#22292F"),
    axis.text.x = element_text(margin = margin(t = 5)),
    axis.text.y = element_text(margin = margin(r = 5)),
    axis.line = element_line(color = "#3D4852"),
    axis.title = element_text(size = 14),
    axis.title.y = element_text(margin = margin(r = 15),
                                hjust = 0.5),
    axis.title.x = element_text(margin = margin(t = 15),
                                hjust = 0.5),
    panel.grid.major = element_line(color = "#DAE1E7"),
    panel.grid.major.x = element_blank(),
    legend.position = "none")

## Joining, by = c("Region", "Country", "Year", "Age", "Gender", "Literacy
rate")
```

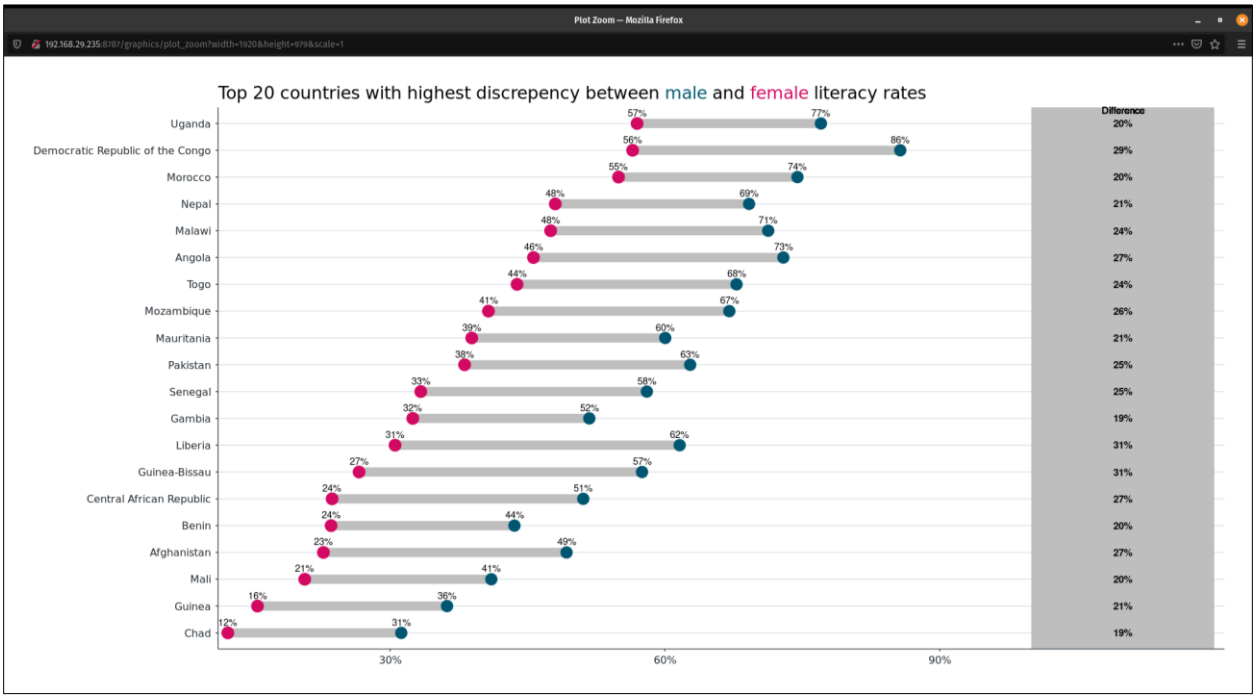


Seeing the top 20 countries whose average male literacy rate is greater or equal to their female average literacy rate, I saw that most countries here are in Europe and Northern America while a few are in Central and South Asia while not one country is from the Sub-Saharan African region.

06. Countries with the highest discrepancy between male and female literacy rates

```
countries_male <- literacy_data %>%
  group_by(Country) %>%
  filter(mean(`Literacy rate`[Gender=="female"],na.rm = T)<mean(`Literacy
rate`[Gender=="male"]))

literacy_data %>%
  semi_join(countries_male) %>%
  group_by(Country) %>%
  summarise(avg_ltrcy_male = mean(`Literacy rate`[Gender=='male'],na.rm = T),
            avg_ltrcy_female=mean(`Literacy rate`[Gender=='female'],na.rm =
T),
            diff= avg_ltrcy_male-avg_ltrcy_female) %>%
  top_n(20,diff)%>%

ggplot(aes(y=reorder(Country,avg_ltrcy_female),x=avg_ltrcy_male,xend=avg_ltrc
y_female))+
  ggalt::geom_dumbbell( size=5, colour="grey",colour_x =
"#005670",colour_xend = "#d20962")+
  geom_text(aes(x=avg_ltrcy_female,label=percent(avg_ltrcy_female,accuracy =
1)),vjust=-1)+
  geom_text(aes(x=avg_ltrcy_male,label=percent(avg_ltrcy_male,accuracy =
1)),vjust=-1)+
  geom_rect(aes(xmin=1,xmax=1.2,ymin=-Inf,ymax=Inf),fill="grey")+
  geom_text(aes(label=percent(diff,accuracy = 1), y=Country, x=1.1),
fontface="bold", size=4)+
  geom_text(aes(x=1.1,y=20.5,label="Difference"))+
  labs(x=NULL,y=NULL,title = "Top 20 countries with highest discrepency
between <span style='color:#005670'>male</span> and <span
style='color:#d20962'>female</span> literacy rates")+
  scale_y_discrete()+
  scale_x_percent(breaks = c(0.3,0.6,0.9),labels = c("30%","60%","90%"))+
  theme_classic()+
  theme(
    text = element_text(family = "Roboto Condensed"),
    plot.margin = unit(rep(1.2, 4), "cm"),
    plot.title = element_markdown(size = 20,margin = margin(b = 5)),
    plot.subtitle = element_text(size = 15,
                                 margin = margin(b = 35)),
    plot.caption = element_text(size = 10,
                                margin = margin(t = 25),
```

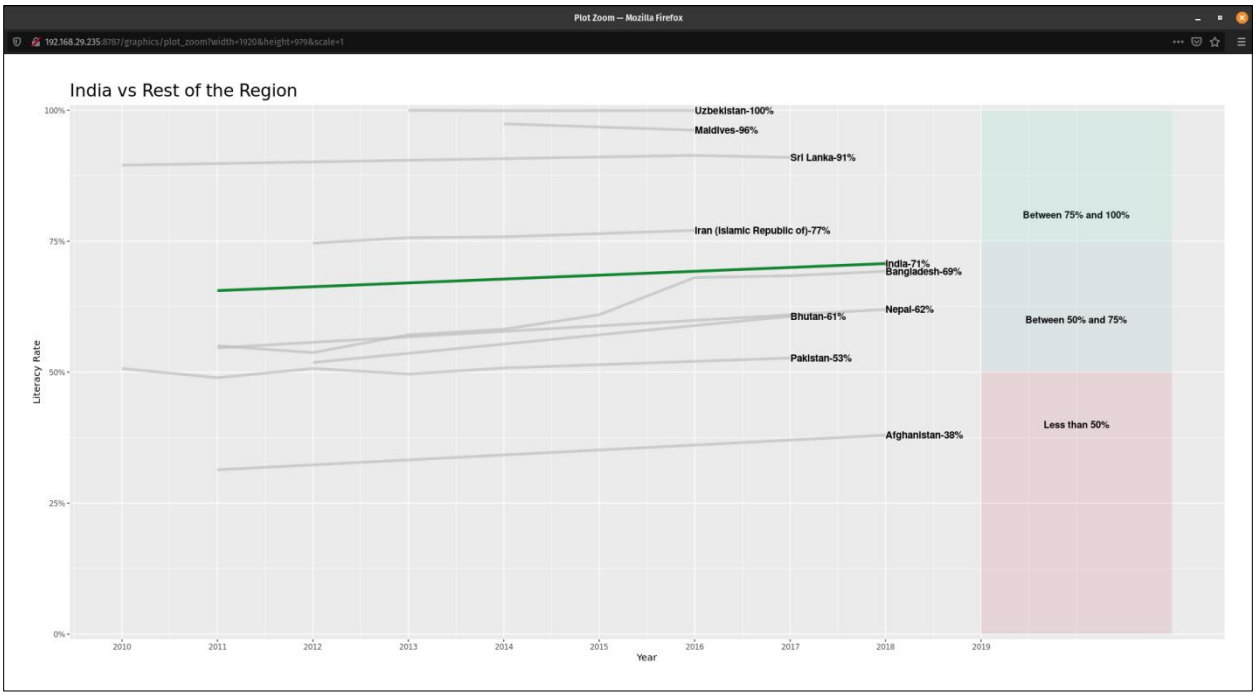
```
                                color = "#606F7B"),
    panel.background = element_blank(),
    axis.text = element_text(size = 12, color = "#22292F"),
    axis.text.x = element_text(margin = margin(t = 5)),
    axis.text.y = element_text(margin = margin(r = 5)),
    axis.line = element_line(color = "#3D4852"),
    axis.title = element_text(size = 14),
    axis.title.y = element_text(margin = margin(r = 15),
                                hjust = 0.5),
    axis.title.x = element_text(margin = margin(t = 15),
                                hjust = 0.5),
    panel.grid.major = element_line(color = "#DAE1E7"),
    panel.grid.major.x = element_blank(),
    legend.position = "none")

## Joining, by = c("Region", "Country", "Year", "Age", "Gender", "Literacy
rate")
```



Exploring this further, I wanted to see the countries that had the highest discrepancy between male and female literacy rates. For aesthetics, I only visualized the top 20 based on the value of the difference between the two genders. Again, we can see a lot of African countries here as well and some countries from Asia like Nepal and Pakistan. It's very interesting to see that Africa has countries where the average female literacy rate is greater than the opposite gender but at the same time, there are countries in Africa where the male literacy rate is more than the female literacy rate by a huge margin.

07. Taking a closer look at India

```
south_asia <-literacy_data %>%
  filter(Region=="Central and Southern Asia") %>%
  group_by(Year,Country) %>%
  summarise(avg_ltrcy=mean(`Literacy rate`)) %>%
  ungroup() %>%
  group_by(Country) %>%
  filter(n()>1)

south_asia %>%
  ggplot(aes(Year,avg_ltrcy))+
  geom_line(size=1.5,aes(color=Country))+
  gghighlight(Country=="India",use_direct_label = F,use_group_by = F)+
  scale_color_manual(values = c("India"="#11862f"))+
  geom_text(data = south_asia %>% group_by(Country) %>%
              filter(Year==max(Year)) %>%
ungroup(),aes(label=paste0(Country,"-",percent(avg_ltrcy,accuracy =
1))),size=4,hjust=0,fontface="bold")+
  scale_x_continuous(breaks = seq(2010,2019,by = 1),limits = c(2010,2021))+
  annotate(geom = "text",x=2020,y=0.8,label="Between 75% and
100%",color="black",fontface="bold")+

geom_rect(aes(xmin=2019,xmax=2021,ymin=0.75,ymax=1),fill="#3be8b0",alpha=0.05
)+

geom_rect(aes(xmin=2019,xmax=2021,ymin=0.5,ymax=0.75),fill="#56a0d3",alpha=0.
05)+
  annotate(geom = "text",x=2020,y=0.6,label="Between 50% and
75%",color="black",fontface="bold")+

geom_rect(aes(xmin=2019,xmax=2021,ymin=0,ymax=0.5),fill="#c90f23",alpha=0.05)
+
  annotate(geom = "text",x=2020,y=0.4,label="Less than
50%",color="black",fontface="bold")+
  scale_y_percent()+
  labs(x="Year",y="Literacy Rate",color=NULL,title = "India vs Rest of the
Region",x=NULL,y=NULL)+
  theme(plot.margin = unit(rep(1.2, 4), "cm"),
        plot.title = element_markdown(size = 20,margin = margin(b = 5)),
        legend.position = "none")+
  theme(text = element_text(family = "Roboto Condensed"))
```

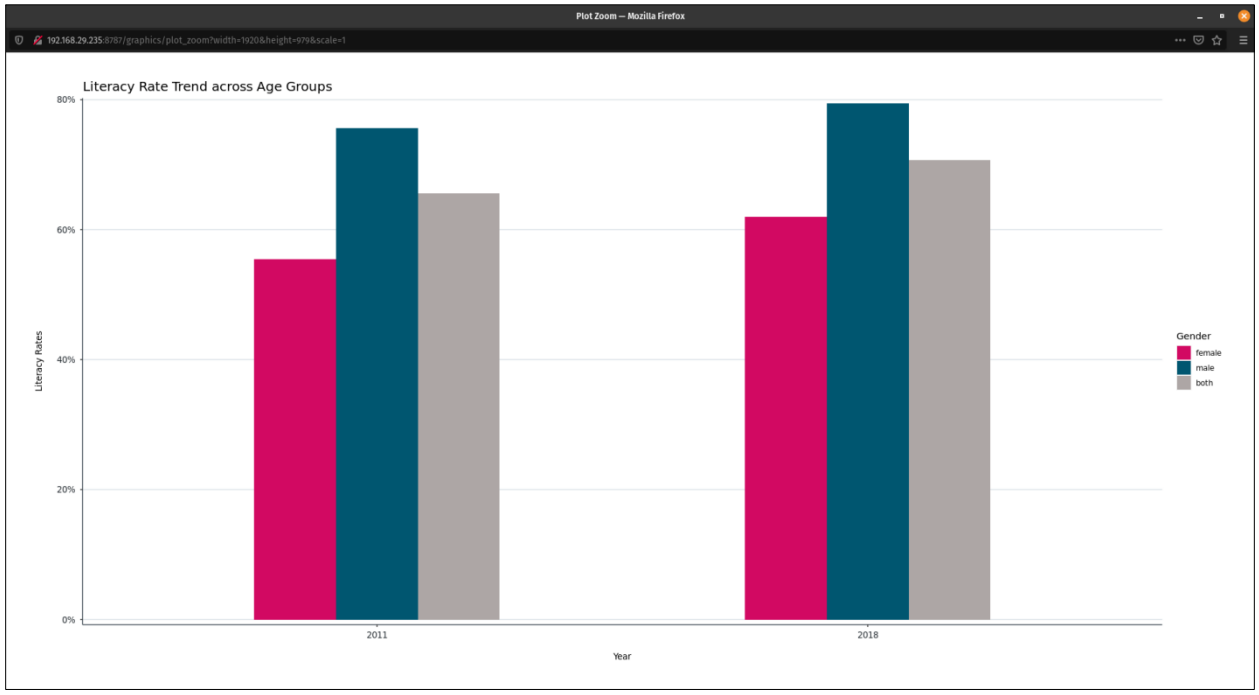


Since I live in India, I wanted to see how the country compares to other countries in the same region. While the general trend is slightly upwards and India is doing better than its neighboring countries, it is disappointing to see that India is lagging behind countries like Maldives, Uzbekistan, Sri Lanka and Iran.

08. Literacy Rate by gender in India

```
ind_bygender <- literacy_data %>%
  filter(Country=="India") %>%
  group_by(Year,Gender) %>%
  summarise(avg=mean(`Literacy rate`,na.rm = T, .groups='drop')) %>%
  ungroup()

ggplot(ind_bygender, aes(factor(Year), avg, fill = Gender))+
  geom_bar(stat = "identity", width = 0.5, position = "dodge") +
  scale_fill_manual(values =
c("male"="#005670","female"="#d20962","total"="#ada6a5"), labels =
c("female", "male", "both"))+
  scale_y_percent()+
  labs(y="Literacy Rates",x="Year",title = "Literacy Rate Trend across Age
Groups")+
  theme(text = element_text(family = "Roboto Condensed"),
        plot.margin = unit(rep(1.2, 4), "cm"),
        plot.title = element_markdown(size = 15,margin = margin(b = 5)),
```

```
        plot.subtitle = element_text(size = 15,
                                     margin = margin(b = 35)),
        plot.caption = element_text(size = 10,
                                    margin = margin(t = 25),
                                    color = "#606F7B"),
        panel.background = element_blank(),
        axis.text = element_text(size = 10, color = "#22292F"),
        axis.text.x = element_text(margin = margin(t = 5)),
        axis.text.y = element_text(margin = margin(r = 5)),
        axis.line = element_line(color = "#3D4852"),
        axis.title = element_text(size = 10),
        axis.title.y = element_text(margin = margin(r = 15),
                                    hjust = 0.5),
        axis.title.x = element_text(margin = margin(t = 15),
                                    hjust = 0.5),
        panel.grid.major = element_line(color = "#DAE1E7"),
        panel.grid.major.x = element_blank())
```

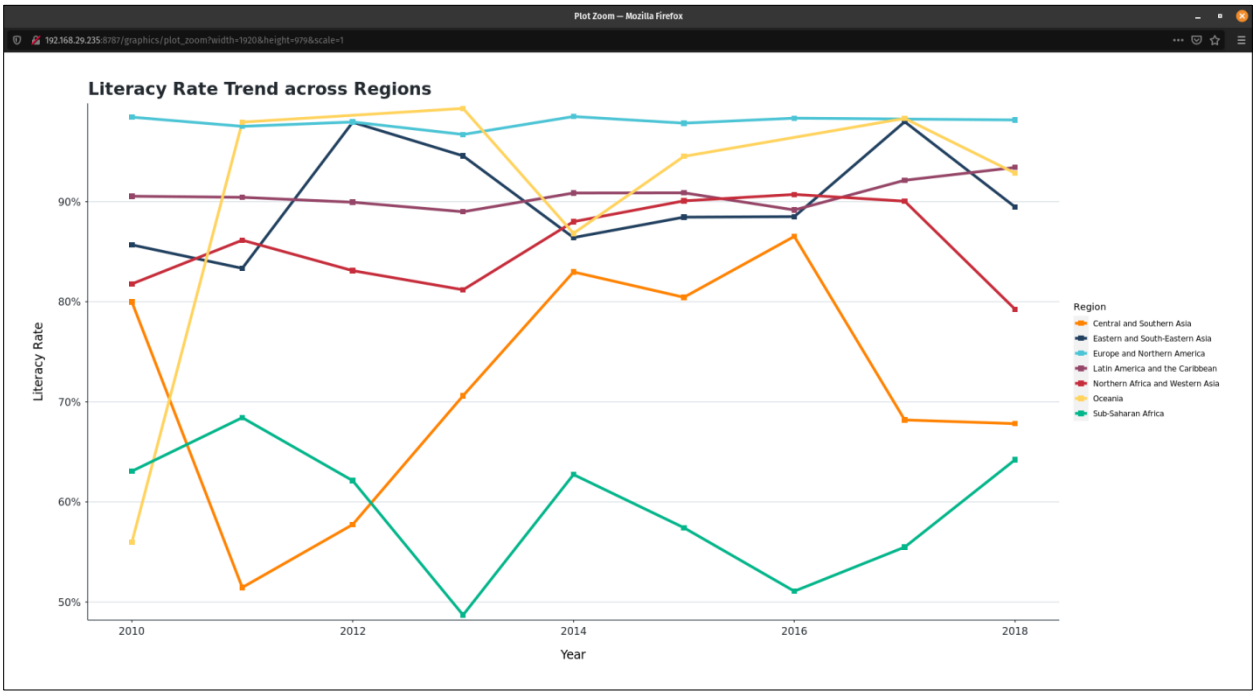


Diving into India's data, it can be seen that there hasn't been a significant increase in the overall literacy rate of India. Also, there is a significant difference (~20%) between the literacy rates of females and males in India in both 2011 and 2018.

09. Literacy rate across regions from 2011 to 2018

```
avg_region <- literacy_data %>%
  group_by(Year,Region) %>%
  summarise(avg=mean(`Literacy rate`,na.rm = T, .groups='drop')) %>%
  ungroup()

## `summarise()` has grouped output by 'Year'. You can override using the
`.groups` argument.

avg_region %>%
  ggplot(aes(Year,avg,color=Region))+
  geom_line(size=1.5)+
  geom_point(size = 2.6, aes(color = Region), shape = 15) +
  scale_color_manual(values = c("Sub-Saharan
Africa"="#02B58A","Oceania"="#FFD35C","Northern Africa and Western
Asia"="#C62C3A","Latin America and the Caribbean"="#954567","Europe and
Northern America"="#4BC4D5", "Eastern and South-Eastern
Asia"="#21405F","Central and Southern Asia"="#FF8201"))+
  scale_y_percent()+
  labs(y="Literacy Rate",x="Year",title = "Literacy Rate Trend across
Regions")+
  theme(
    text = element_text(family = "Roboto Condensed"),
    plot.margin = unit(rep(1.2, 4), "cm"),
    plot.title = element_text(size = 20,
                              color = "#22292F",
                              face = "bold",
                              margin = margin(b = 5)),
    plot.subtitle = element_text(size = 15,
                                 margin = margin(b = 35)),
    plot.caption = element_text(size = 10,
                                margin = margin(t = 25),
                                color = "#606F7B"),
    panel.background = element_blank(),
    axis.text = element_text(size = 12, color = "#22292F"),
    axis.text.x = element_text(margin = margin(t = 5)),
    axis.text.y = element_text(margin = margin(r = 5)),
    axis.line = element_line(color = "#3D4852"),
    axis.title = element_text(size = 14),
    axis.title.y = element_text(margin = margin(r = 15),
                                hjust = 0.5),
    axis.title.x = element_text(margin = margin(t = 15),
                                hjust = 0.5),
```

```
    panel.grid.major = element_line(color = "#DAE1E7"),
    panel.grid.major.x = element_blank(),
    #Legend.position = "none"
  )
```

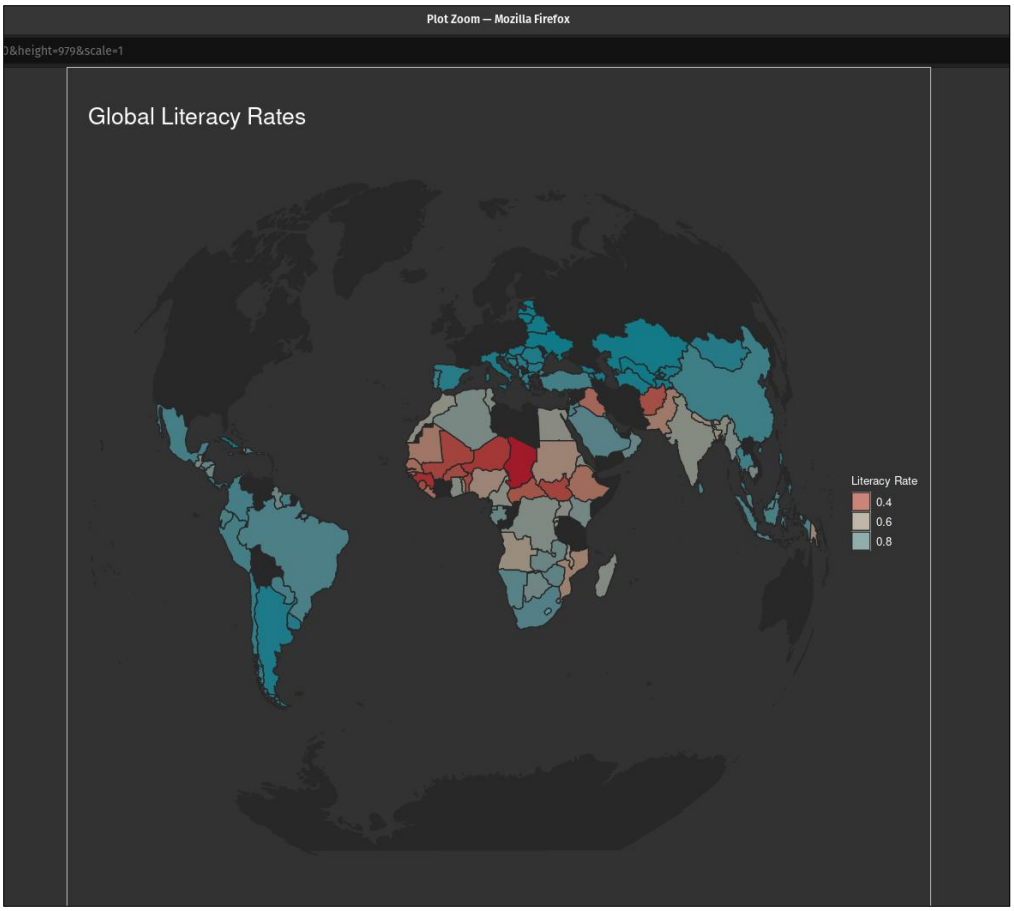


The most broad view of the world literacy rate data. It is seen that Europe and Northern America have the highest and the most steady literacy rate from 2011-2018 while Sub-Saharan Africa is the region with the lowest literacy rate. Central and Southern Asia have shown a significant increase in the initial years and a slight dip can be seen towards the end. It is also interesting to see a sharp dip in the literacy rate in Northern Africa and Western Asia in 2017-18 and a sharp rise literacy rate in Eastern and South-Eastern Asia in 2011-12. That being said, a few other inferences can also be made from this plot, I've only mentioned the major ones.

10. Visualizing Global Literacy Rates on a Map

```
world <- map_data(map = "world") %>%
  filter(region!="Antartica")

long_lat <- literacy_data %>%
  group_by(Country) %>%
  summarise(`Literacy rate`=mean(`Literacy rate`,na.rm = T)) %>%
  ungroup() %>%
```

```
  left_join(world,by = c("Country"="region")) %>%
  filter(!is.na(lat))

p <- ggplot() +
  geom_map(data = world, map = world,
           aes(long, lat, group = group, map_id = region),
           fill = "#282828", color = "#282828") +
  geom_map(data = long_lat, map = world,
           aes(fill = `Literacy rate`, map_id = Country),
           color = "#282828", size = 0.5, alpha = .8) +
  scale_fill_gradient2(low = "#be0027", high = "#0a8ea0",mid =
"#b4a996",midpoint = 0.6) +
  scale_y_continuous(breaks=c()) +
  scale_x_continuous(breaks=c()) +
  labs(x = "", y = "") +
  guides(
    fill = guide_legend(title = "Literacy Rate")
  ) +
  coord_map("gilbert", xlim = c(-300, 300)) +
  labs(
    title = "Global Literacy Rates"
  ) +
  theme(plot.title = element_text(color = "#ffffff",
                                  margin = margin(t = 30, b = 10),
                                  size = 20),
        plot.subtitle = element_text(color = "#abababb",
                                     margin = margin(b = 10),
                                     size = 15,
                                     hjust = 0.7),
        plot.background  = element_rect(fill  = "#323232"),
        panel.background = element_rect(fill  = "#323232",
                                        color = "#323232"),
    legend.position = "right",
    legend.title = element_text(color = "#ffffff",
                                size = 10),
    legend.background = element_rect(fill = "#323232"),
    legend.text = element_text(color = "#ffffff",
                               size = 10)
  )

cowplot::ggdraw(p)+
  theme(plot.background = element_rect(fill = "#323232"))
```



I wanted to try visualizing the data on a map as nothing is more visual than seeing data on an actual world map. This is useful to get an understanding of the data on a global level. Most of Europe and Northern America has a high literacy rate while the lowest can be seen in parts of Africa and a few parts in Asia.

Ending Note –

I tried to be more creative with my visualizations instead of relying on the typical bar charts. I was surprised to see that there are a lot of "poorer" countries that have really high literacy rates. I wish there were more data that could help us understand which factors affect the literacy rate of a country from which we could build a model based on some sort of regression analysis.