

# Exploring Global Education: A Unique Data Visualization Approach

Muhammad Qamar, 100889132

## Introduction

This data is used from the `Global_Education.csv` file which was downloaded from Kaggle. This data set was selected as it consists of holistic data on various sub-aspects of education across countries with a wider range of comparisons that which allows for a deeper investigations of global education trends.

In this report, we will be visualizing data to answer the following questions:

1. Which countries have the highest average completion rate in education?
2. How does the unemployment rate vary across different countries?
3. What is the distribution of primary and tertiary education enrollment?

By these visualizations, we will try to get the key insight on education condition in the whole world as well as identify an area for investigation/intervention which may require them.

## Data Analysis

## Data Manipulation

First, we load the necessary libraries and read the data. Then, we calculate the average completion rate across different education levels and genders for each country. This gives a single measure of educational attainment.

```
# Load the necessary libraries
library(dplyr)
library(ggplot2)
library(maps)
library(corrplot)
library(tidyverse)
library(ggthemes)
library(ggbeeswarm)

# Load the data
data <- read.csv("/Users/mqamar/Documents/CSCI4210U/Assignment/Global_Education.csv")

# Perform data manipulation here
data <- data %>%
  mutate(average_completion_rate = (Completion_Rate_Primary_Male + Completion_Rate_Primary_Female + Completion_Rate_Lower_Secondary_Male + Completion_Rate_Lower_Secondary_Female + Completion_Rate_Upper_Secondary_Male + Completion_Rate_Upper_Secondary_Female) / 6)
```

## Top Countries by Completion Rate

Next, we identify the top 10 countries with the highest average completion rate. This helps to understand which countries are most successful in terms of educational completion.

[Hide](#)

```
# Filter to include only the top 10 countries with the highest average completion rate
top_countries <- data %>%
  group_by(Countries.and.areas) %>%
  summarise(average_completion_rate = mean(average_completion_rate, na.rm = TRUE)) %>%
  top_n(10, average_completion_rate)

# Display the first few lines of top_countries
head(top_countries)
```

Countries.and.areas <chr>	average_completion_rate <dbl>
Barbados	97.00000
Belarus	97.50000
Bosnia and Herzegovina	96.16667
Cuba	93.66667
Georgia	93.00000
Kazakhstan	98.50000
6 rows	

[Hide](#)

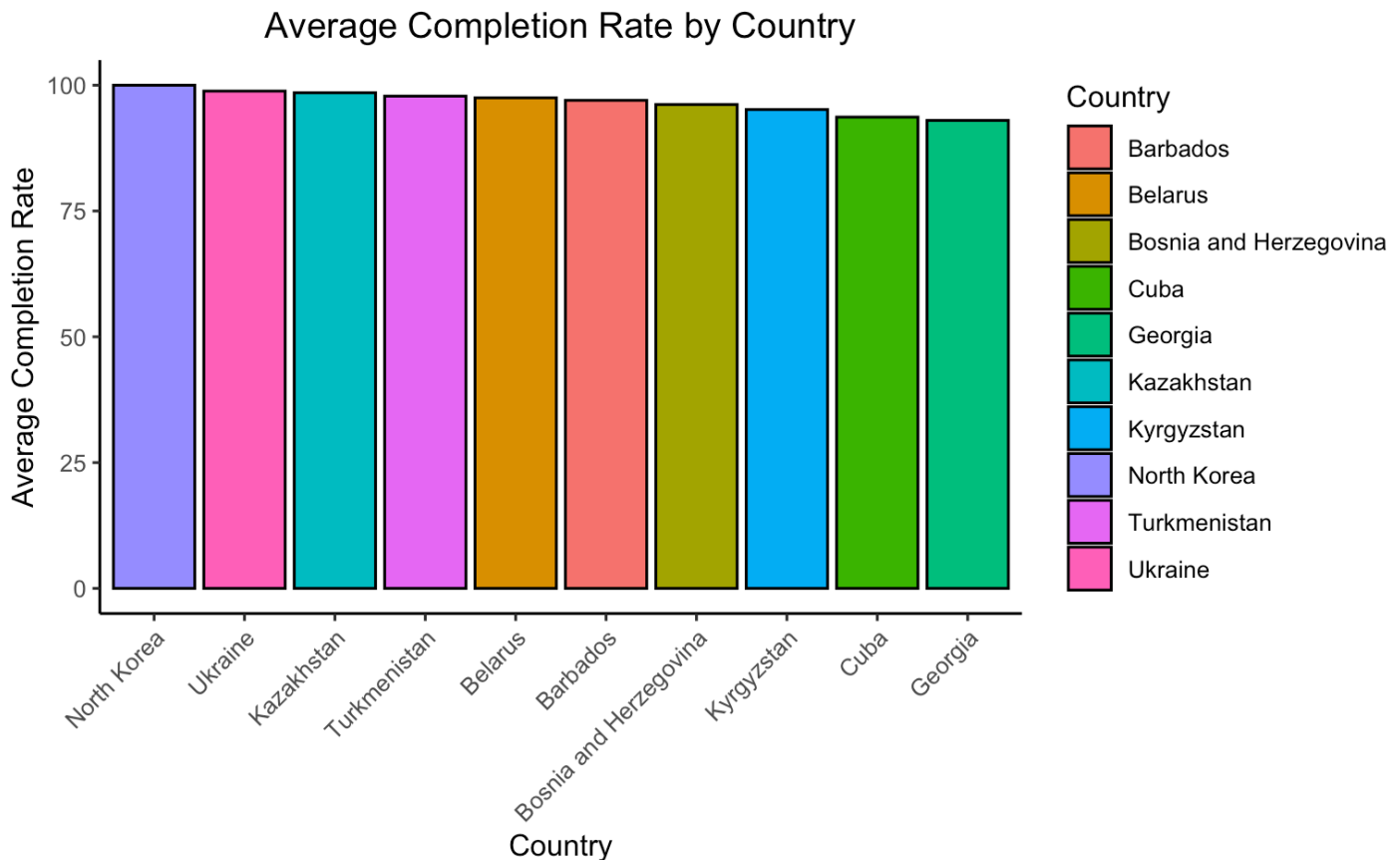
NA

## Bar Plot and Line Graph

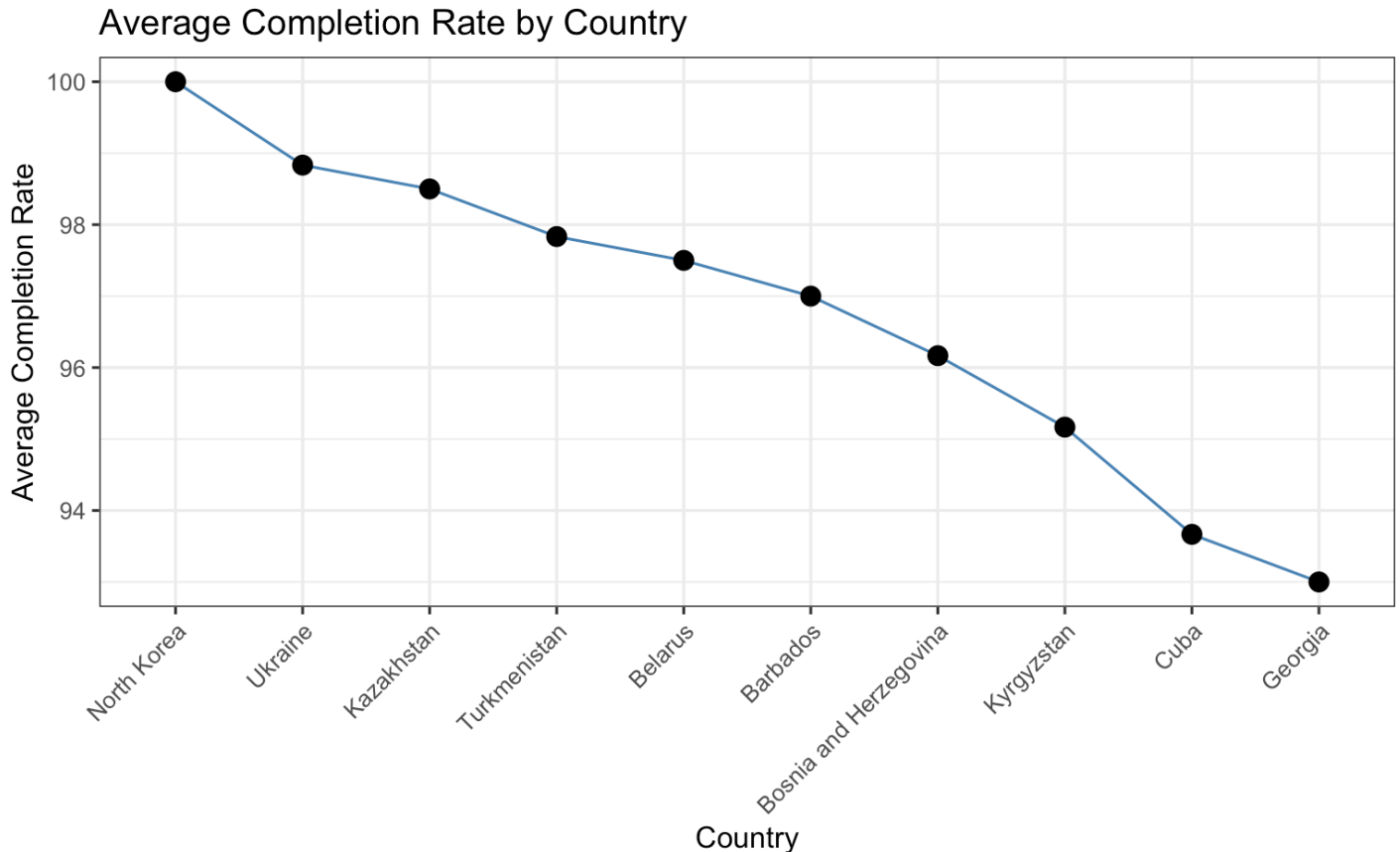
These plots visualize the average completion rate for the top 10 countries. The bar plot provides a clear comparison of the rates, while the line graph shows the trend among these countries.

[Hide](#)

```
# Bar Plot
ggplot(top_countries, aes(x=reorder(`Countries.and.areas`, -average_completion_rate), y=
average_completion_rate, fill=`Countries.and.areas`)) +
  geom_bar(stat="identity", position=position_dodge(), color="black") +
  labs(title="Average Completion Rate by Country",
       x="Country",
       y="Average Completion Rate",
       fill="Country") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5))
```

[Hide](#)

```
# Line Graph
ggplot(top_countries, aes(x=reorder(`Countries.and.areas`, -average_completion_rate), y=
average_completion_rate, group=1)) +
  geom_line(color="steelblue") +
  geom_point(size=3, color="black") +
  labs(title="Average Completion Rate by Country",
       x="Country",
       y="Average Completion Rate") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



## Top Countries by Unemployment Rate

We identify the top 10 countries with the highest unemployment rate. This can give insights into the job market in these countries. Also, calculating the average percentage for each country with the highest unemployment rate.

Hide

```
# Identify the top 10 countries with the highest unemployment rate
top_countries_unemployment <- data %>%
  group_by(`Countries.and.areas`) %>%
  summarise(Unemployment_Rate = mean(Unemployment_Rate, na.rm = TRUE)) %>%
  top_n(10, Unemployment_Rate)

# Display the first few lines of top_countries_unemployment
head(top_countries_unemployment)
```

Countries.and.areas	Unemployment_Rate
<chr>	<dbl>
Bosnia and Herzegovina	18.42
Botswana	18.19
Gabon	20.00
Greece	17.24
Lesotho	23.41
Libya	18.56
6 rows	

[Hide](#)

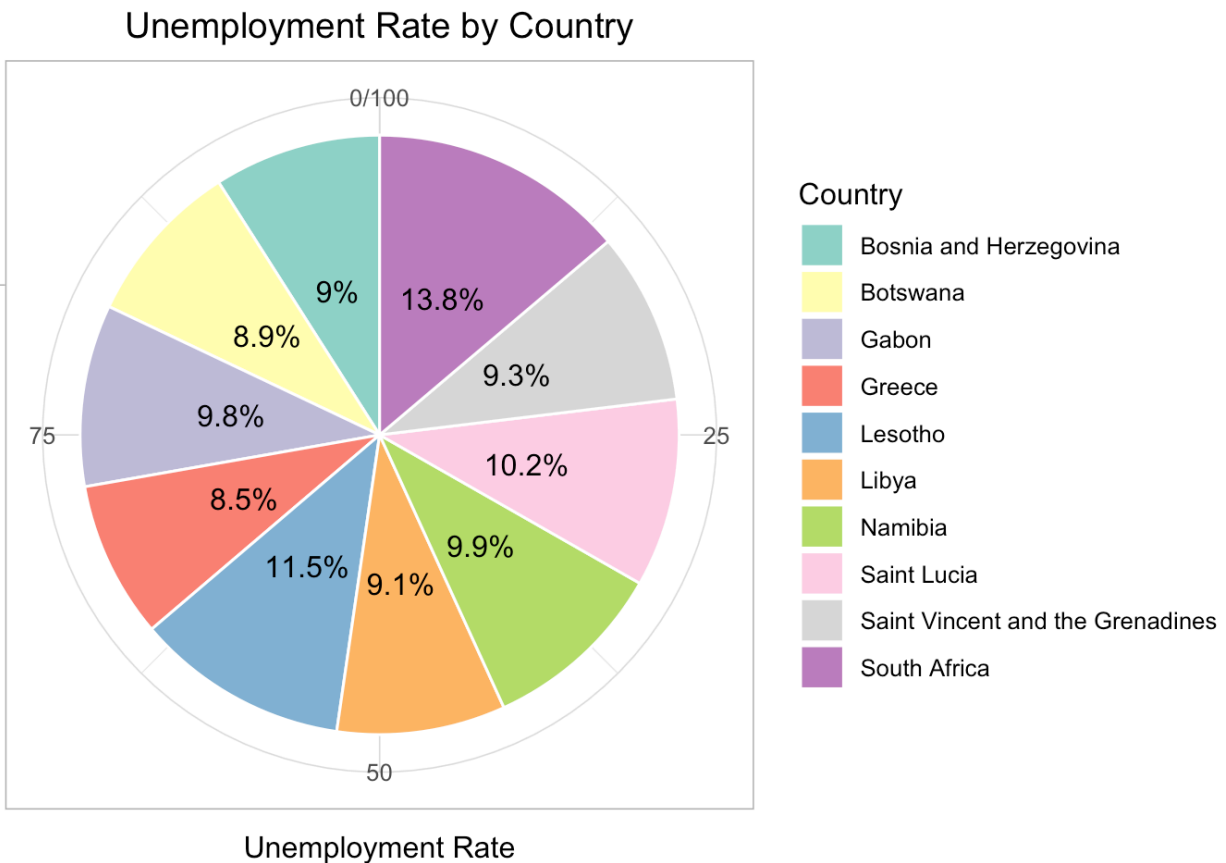
```
# Calculate the percentage for each country
top_countries_unemployment <- top_countries_unemployment %>%
  mutate(percentage = Unemployment_Rate / sum(Unemployment_Rate) * 100)
```

## Pie Chart and Density Plot

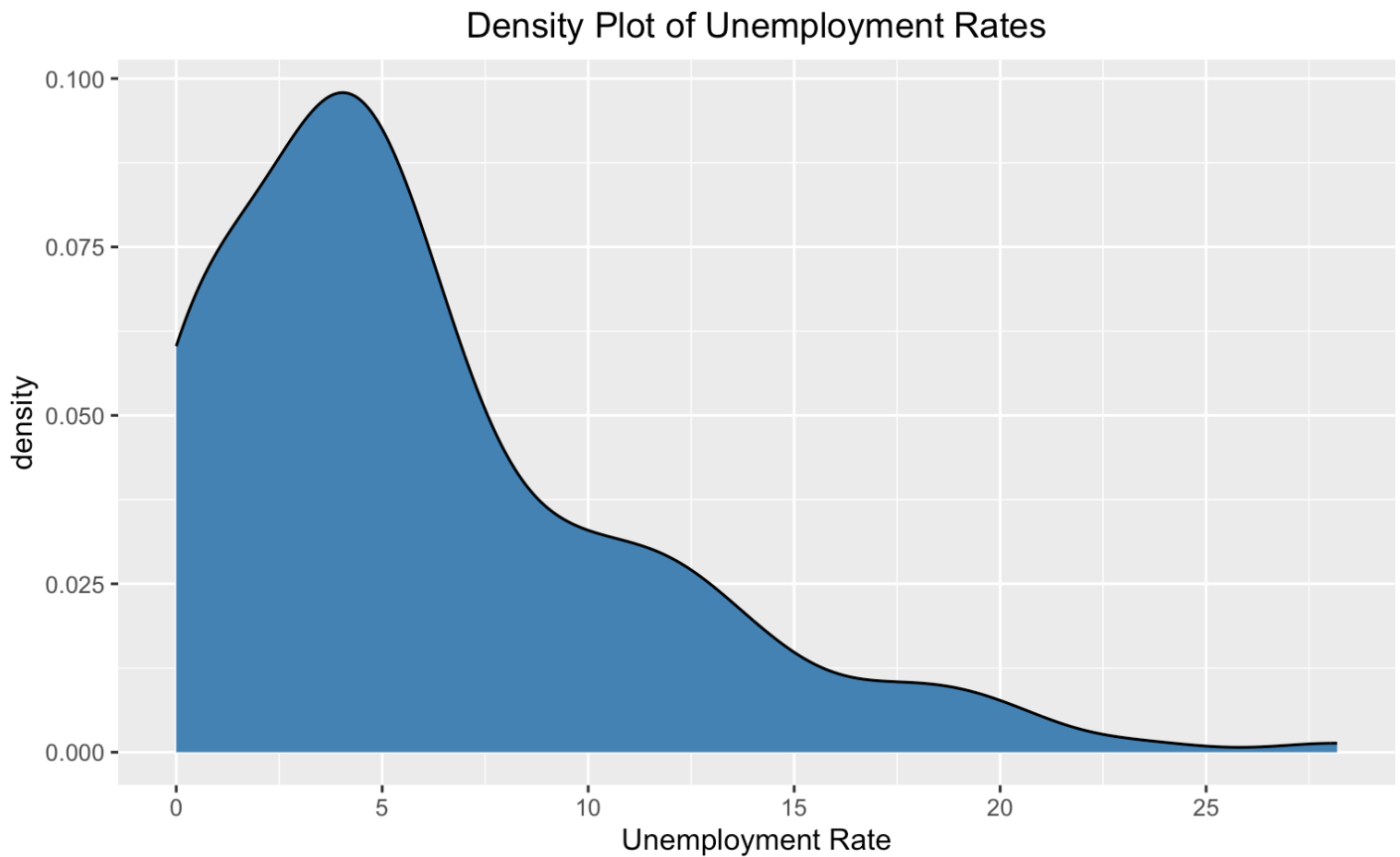
The pie chart shows the proportion of the total unemployment rate that each of the top 10 countries contributes. The density plot shows the distribution of unemployment rates across all countries.

[Hide](#)

```
# Pie Chart
ggplot(top_countries_unemployment, aes(x="", y=percentage, fill=`Countries.and.areas`))
+
  geom_bar(width = 1, stat = "identity", color="white") +
  coord_polar("y", start=0) +
  geom_text(aes(label = paste0(round(percent, 1), "%")), position = position_stack(vj
ust = 0.5)) +
  scale_fill_brewer(palette = "Set3") +
  labs(title="Unemployment Rate by Country",
       x=NULL,
       y="Unemployment Rate",
       fill="Country") +
  theme_light() +
  theme(legend.position="right",
        plot.title = element_text(hjust = 0.5))
```


[Hide](#)

```
# Density Plot
ggplot(data, aes(x=Unemployment_Rate)) +
  geom_density(fill="steelblue") +
  scale_x_continuous(breaks = seq(0, max(data$Unemployment_Rate, na.rm = TRUE), by = 5))
+
  labs(title="Density Plot of Unemployment Rates",
       x="Unemployment Rate") +
  theme_gray() +
  theme(plot.title = element_text(hjust = 0.5))
```

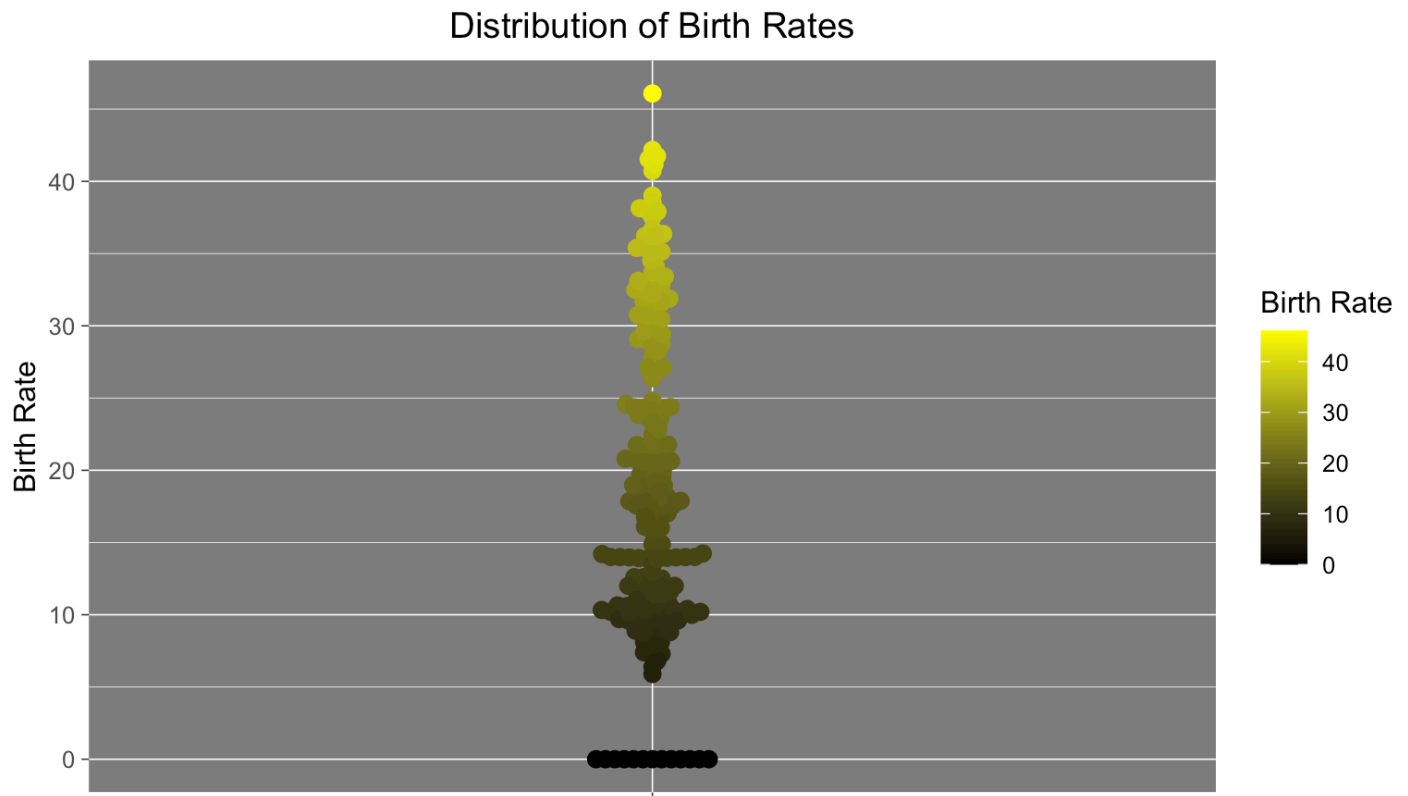


## Beeswarm Plot

This plot shows the distribution of birth rates across all countries. It can give insights into population growth trends.

[Hide](#)

```
# Beeswarm Plot
ggplot(data, aes(x = "", y = Birth_Rate)) +
  geom_beeswarm(aes(color = Birth_Rate), size=2.5) +
  scale_color_gradient(low = "black", high = "yellow") +
  labs(title="Distribution of Birth Rates",
       x="",
       y="Birth Rate",
       color = "Birth Rate") +
  theme_dark() +
  theme(plot.title = element_text(hjust = 0.5),
        panel.grid.major = element_line(color = "white"),
        panel.grid.minor = element_line(color = "white"))
```



## Average Proficiency Calculation/Geographic Map and Scatter Plot

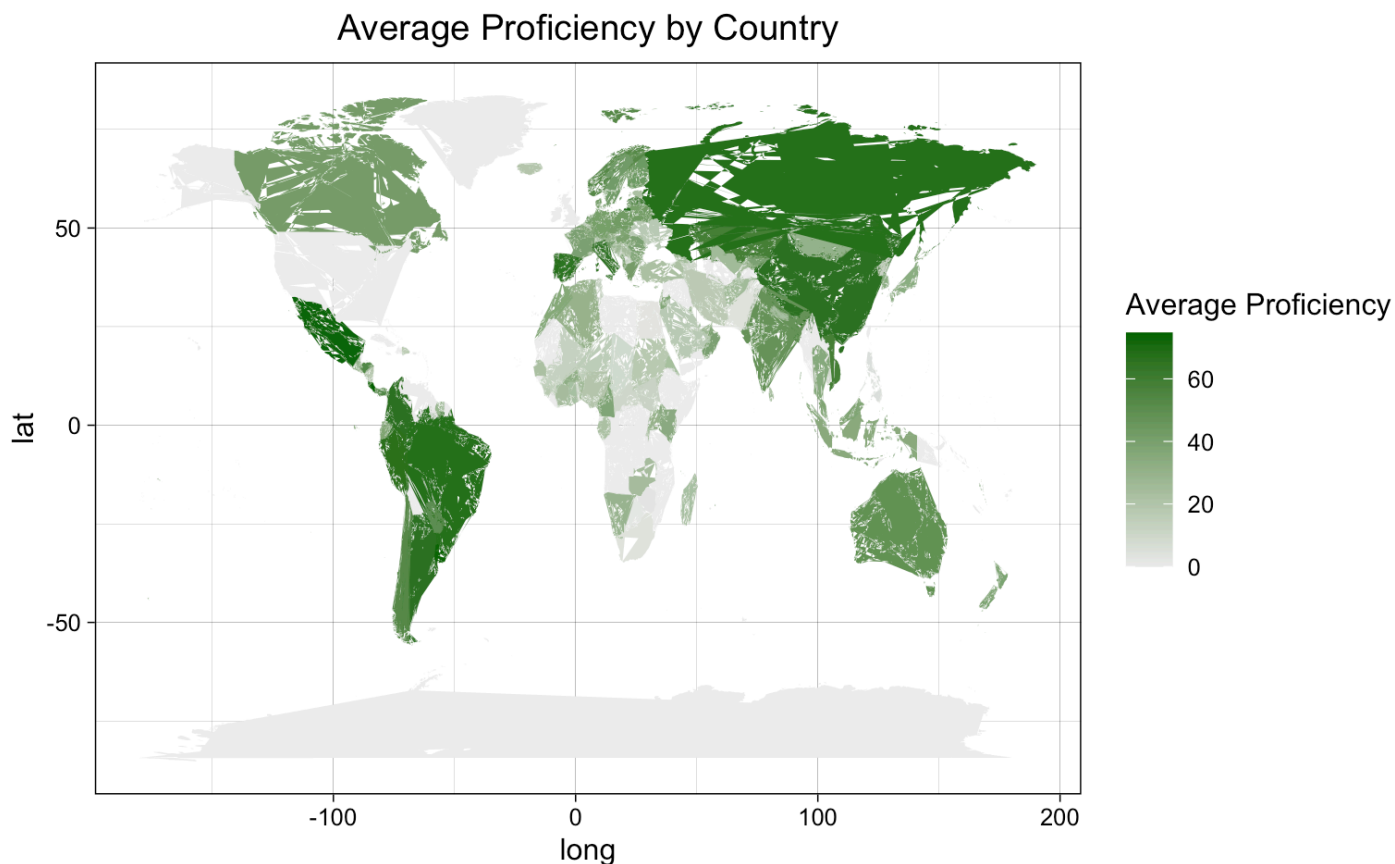
We calculate the average proficiency across different grades and literacy rates for youth. This gives a single measure of educational proficiency for each country. Plots visualize the average proficiency by country and location, respectively. They can give insights into geographical patterns in educational proficiency.

Hide

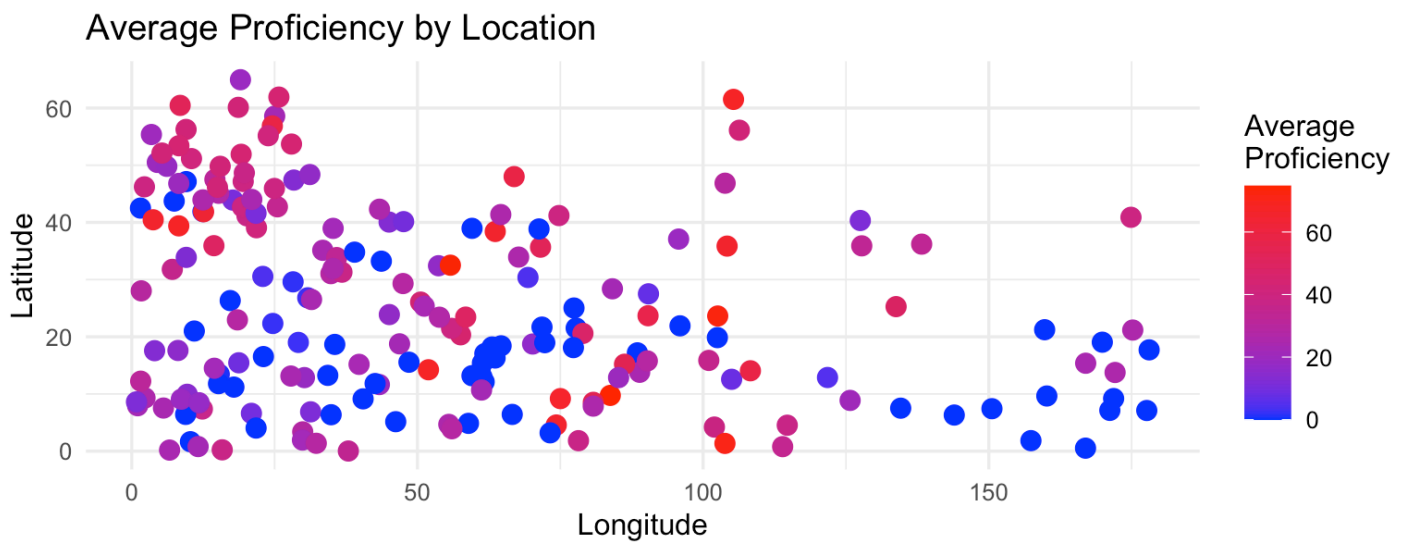


```
# Calculate the average proficiency for each country
data <- data %>%
  mutate(average_proficiency = (Grade_2_3_Proficiency_Reading + Grade_2_3_Proficiency_Math + Primary_End_Proficiency_Reading + Primary_End_Proficiency_Math + Lower_Secondary_End_Proficiency_Reading + Lower_Secondary_End_Proficiency_Math + Youth_15_24_Literacy_Rate_Male + Youth_15_24_Literacy_Rate_Female) / 8)

# Geographic Map
world_map <- map_data("world")
country_counts <- data %>%
  group_by(Countries.and.areas) %>%
  summarise(average_proficiency = mean(average_proficiency, na.rm = TRUE))
names(country_counts)[names(country_counts) == "Countries.and.areas"] <- "region"
world_map <- merge(world_map, country_counts,
  by = "region", all.x = TRUE)
world_map$average_proficiency[is.na(world_map$average_proficiency)] <- 0
ggplot() +
  geom_polygon(data = world_map, aes(x=long, y = lat, group = group, fill = average_proficiency)) +
  scale_fill_gradient(low = "gray92", high = "darkgreen") +
  labs(title = "Average Proficiency by Country",
    fill = "Average Proficiency") +
  theme_linedraw() +
  theme(plot.title = element_text(hjust = 0.5))
```


[Hide](#)

```
# Geographic Scatter Plot
ggplot(data, aes(x=Longitude, y=Latitude)) +
  geom_point(aes(color=average_proficiency), size=3) +
  scale_color_gradient(low = "blue", high = "red") +
  labs(title = "Average Proficiency by Location",
       x = "Longitude",
       y = "Latitude",
       color = "Average\nProficiency") +
  theme_minimal() +
  coord_fixed()
```



## Data Reshaping and Violin Plot

We reshape the data to long format and create a violin plot to show the distribution of primary and tertiary education enrollment. This can give insights into access to education at different levels.

[Hide](#)

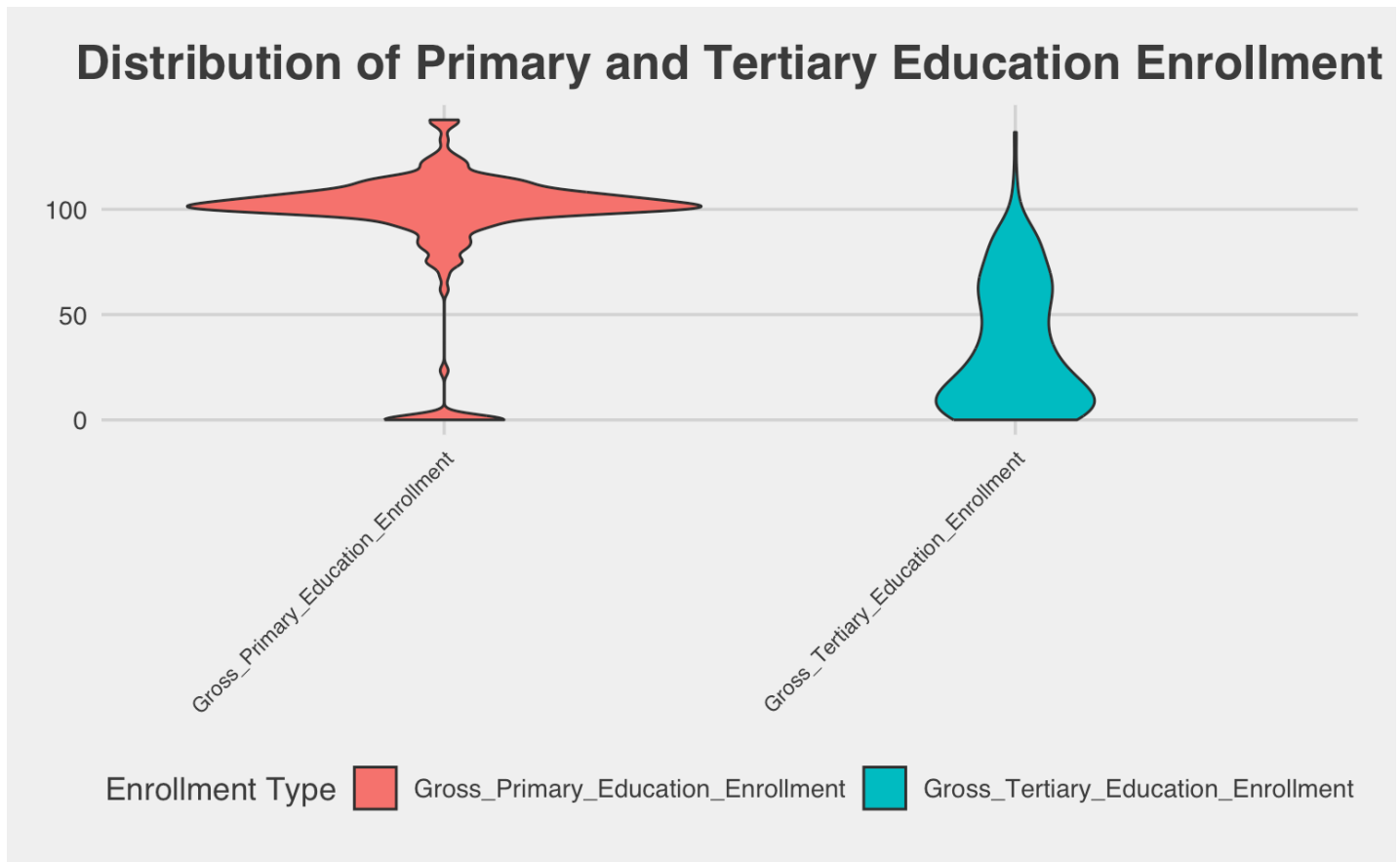
```
# Reshape the data for plotting
data_long <- data %>%
  select(Countries.and.areas, Gross_Primary_Education_Enrollment, Gross_Tertiary_Education_Enrollment) %>%
  pivot_longer(cols = c(Gross_Primary_Education_Enrollment, Gross_Tertiary_Education_Enrollment), names_to = "Enrollment_Type", values_to = "Value")

# Display the first few lines of data_long
head(data_long)
```

Countries.and.areas <chr>	Enrollment_Type <chr>	Value <dbl>
Afghanistan	Gross_Primary_Education_Enrollment	104.0
Afghanistan	Gross_Tertiary_Education_Enrollment	9.7
Albania	Gross_Primary_Education_Enrollment	107.0
Albania	Gross_Tertiary_Education_Enrollment	55.0
Algeria	Gross_Primary_Education_Enrollment	109.9
Algeria	Gross_Tertiary_Education_Enrollment	51.4
6 rows		

[Hide](#)

```
# Violin Plot
ggplot(data_long, aes(x = Enrollment_Type, y = Value, fill = Enrollment_Type)) +
  geom_violin() +
  labs(title = "Distribution of Primary and Tertiary Education Enrollment",
       x = "Enrollment Type",
       y = "Enrollment Value",
       fill = "Enrollment Type") +
  theme_fivethirtyeight() +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 8))
```



## Conclusion

In this analysis, we examined the worldwide educational developments in a deep way using aggregate data. The data that was there for the whole education analysis, `Global_Education.csv`, presented us with a lot of information about the distinct fields of education such as completion levels, the rate of unemployment, birth rates, and proficiency levels.

Our exploration started by finding the average completion rate among the population based on education levels and gender for all the countries. One cause was when we used this single measure that had to do with the highest completion rate level, the top 10 middle education countries around the world came up. Through the bar charts and the line graph's visualizations, we got a clearer picture of the data and thus, we could make out the countries with the highest percentages concerning educational completion.

Secondly, we looked for the Top 10 countries with the highest unemployment and then segregated them into 'Developed Countries' and 'Developing Countries'. Through this, we got into the knowledge concerning the labor market in these countries, and it is also a very important element of education among future generations. The Pie chart and density plot visualizations revealed that the top ten countries together have the major share of the unemployment rate and the distribution of the unemployment rate across all countries, respectively.

Other than that, we looked at birth rates as a measure of number of babies born within a certain period while using a beeswarm plot. Even though this visualization showed us unsteady trends in population growth, which may be very important for the education system, it still helped us analyze the situation.

Moreover, we made the average performance computation for different grades and the youth (children's) literacy rates. This standalone feature enabled us to then create full geographic visualizations that revealed existing patterns in prevailing educational levels by the country pinpointed and by location.

Along with this, we restructured the data to the long form and generated a violin plot, which was used to demonstrate the distribution of at-term and post-secondary education enrollment. This table demonstrates the level of educational stratification and this is a vital contributing factor when it comes to student success.

To sum up, this review with all the information indicates the global education trends that are emerging now. It displays nations with the highest educational completion and lack of employment, the distribution of birth rates and educational proficiency, as well as the geographic distribution of educational achievements. This data may be interesting to policy-makers, teachers, and people who look at education from an international point of view. The information graph designed in this analysis along with the answers to our initial questions, also propelled new scenarios for studying and improvement in the field of education.

## References

Data source: <https://www.kaggle.com/datasets/nelgiriyeewithana/world-educational-data?rvi=1>  
(<https://www.kaggle.com/datasets/nelgiriyeewithana/world-educational-data?rvi=1>)