Assignment 1: Data Mini-project

Vince Currie

18-NOV-2024


```
#
# Set the default behaviour for displaying code in all chunks to ensure that all chunks will display their code,
# unless turned off within a specific chunk.
#
knitr::opts_chunk$set(echo = TRUE, warning = FALSE)
```

1 **Part A**: Data Mini-project

1.1 Acronyms

Acronym	Definition
ISCED	International Standard Classification of education
ISCED 1	Primary school

1.2 Loading libraries

1.2.1 In this section, you need to load the libraries you will use. This section should be located at the top of the Rmd file.

The libraries are loaded silently in this step.

1.3 Research Question

1.3.1 Describe in 2 sentences maximum a question you would like to investigate/answer in this report using your selected data.

The research focuses on asking the following questions:

1. Are both genders now on a course towards achieving comparable educational attainment worldwide? And,
2. How do the average school years vary across different countries, and what might contribute to differences in education levels between countries?

1.4 Data Introduction

```
#
# Global variables.
#
project_dir <- here::here()

#
# The data read from the files that were downloaded from the
# "Our world in data web site": https://ourworldindata.org/).
# The two CSV files contain the average years of schooling for female and male subjects.
#
data1 <- read.csv(here::here("data", "mean-years-of-schooling-female.csv"))
data2 <- read.csv(here::here("data", "mean-years-of-schooling-male.csv"))

#
# Add a gender factor column with "Female" as the only level for the female data and
# "male" for the male data.
#
data1$gender <- factor(rep("Female", nrow(data1)), levels = "Female")
data2$gender <- factor(rep("Male", nrow(data2)), levels = "Male")

#
# Rename the 4th column of both data frames to ensure they are named the same because
# data1 included "female" and data2 used "male" in their names, respectively.
#
names(data1)[4] <- "SchoolYears"
names(data2)[4] <- "SchoolYears"

#
# In addiction, rename "Entity" to "Country" and
# "Code" to "Ccode" for country code.
#
names(data1)[1] <- "Country"
names(data2)[1] <- "Country"
names(data1)[2] <- "Ccode"
names(data2)[2] <- "Ccode"


#
# Combine the two datasets row-wise.
#
all_data <- rbind(data1, data2)
```

```
#
# Extract column names and data types.
#    Notes: "sapply()" applies the class function to each column.
#    in "all_data" to determine its type.
#
str_table <- data.frame(
  Column = names(all_data),
  Type = sapply(all_data, class),
  row.names = NULL
)

#
# Display column names and data types as a formatted table.
# Requires the format parameter due to a known issue of double printing the "Table":
#    https://github.com/haozhu233/kableExtra/issues/831
#
knitr::kable(
  str_table,
  col.names = c("Column Name", "Data Type"),
  booktabs = TRUE,
  format = "html"
) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("striped", "hover"),
    position = "center",
    full_width = FALSE
  )
```

Table 1.1: Column Names and Data Types.

Column Name	Data Type
Country	character
Ccode	character
Year	integer
SchoolYears	numeric
gender	factor

1.4.1 Briefly describe your data in two sentences using markdown language. You should provide a link to the location of the data inserted in the text using markdown language.

The data source came from two CSV files loaded by this script from D:/Monash/ETO5513/Assignment1/data/mean-years-of-schooling-female.csv (*Mean Years of Schooling (ISCED 1 or Higher), Female* (2021)) and D:/Monash/ETO5513/Assignment1/data/mean-years-of-schooling-female.csv (*Mean Years of Schooling (ISCED 1 or Higher), Male* (2021)) with the original found here for female data (https://ourworldindata.org/grapher/mean-years-of-schooling-female) and here for male data (https://ourworldindata.org/grapher/mean-years-of-schooling-male).

1.4.2 Briefly describe your variables (in no more than 3 sentences for all the variables).

The data concerns the mean number of times in years of education for individuals sampled across countries over several years. The two datasets are combined to perform the analysis and answer the research question.

1.4.3 Create a table using the kable() function from the knitr package to report the variable names.

Shown in Table 1.1 are the variables and their data types of the combined table containing male and female data. The *Country* column indicates the country from which the data comes and is a character string. Similarly, the *Ccode* column is a string and a three-letter acronym for the country. The *year* indicates when it was collected and is an integer. The *SchoolYears* column metric reflects the average duration of education in years that people in that country undertook, stored as a float. And finally, the *Gender* variable is a factor value indicating Male or Female.

1.5 Data Set Description

1.5.1 Using inline R code, write a sentence describing the number of variables and observations in your data set.

The combined data frame contains **three thousand and four** records. Table 1.1 shows the **five** variables and their data types.

1.5.2 In addition, create a screenshot image of your code and save it as a png file. Upload this png file inside the 'Image' folder in your Rstudio Cloud project.

Figure 1.1 shows an example of loading a picture of code into the R markdown document.

1.5.3 Using the function str() display the first 2 rows of the data so you can show the type of variables in the data set (numeric, character/factor etc.).

Table 1.2 shows the first two rows of the combined data using the "str()" function. The characteristics of the data are more apparent with some data from the first two rows of output. Figure 1.2 shows no missing values in the data because the **three thousand and four** records have 0 missing values; if there were any missing values, it would show a 1 in the output.

```
knitr::include_graphics(here::here("images", "code_screenshot.png"),rel_path=FALSE)
```

Figure 1.1:

Picture of Code

````
```{r table-DTypes, warning=FALSE, tab.cap="Column Names and Data Types.", tab.cap.style = "inline"}

#
# Extract column names and data types.
#    Notes: "sapply()" applies the class function to each column.
#    in "all_data" to determine its type.
#
str_table <- data.frame(
  Column = names(all_data),
  Type = sapply(all_data, class),
  row.names = NULL
)


#
# Display column names and data types as a formatted table.
# Requires the format parameter due to a known issue of double printing the "Table":
#    https://github.com/haozhu233/kableExtra/issues/831
#
knitr::kable(
  str_table,
  col.names = c("Column Name", "Data Type"),
  booktabs = TRUE,
  format = "html"
) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("striped", "hover"),
    position = "center",
    full_width = FALSE
  )
```
````

```
#
# The "capture.output()" captures the console output of str().
# Capture the output of str() for the first two rows.
str_output <- capture.output(str(all_data[1:2, ]))

#
# Convert the captured output to a data frame.
#
str_output_df <- data.frame(StructureInfo = str_output)

#
# Remove the first row (after the column name) which is the str output about
# the observations and variable number which has already been discussed.
# The str_output_df[-1, , drop = FALSE] removes the first row of data while keeping
# the column name intact.
#
str_output_df <- str_output_df[-1, ,drop = FALSE]

#
# Display the output from str() for the first two rows.
#
knitr::kable(str_output_df, longtable = TRUE, align = "c", format = "html") %>%
  kableExtra::kable_styling(
    bootstrap_options = c("striped", "hover"),
    position = "center",
    full_width = FALSE)
```

Table 1.2: A Sample of the First Two Rows.

	StructureInfo
2	$ Country : chr “Afghanistan” “Afghanistan”
3	$ Ccode : chr “AFG” “AFG”
4	$ Year : int 1975 1979
5	$ SchoolYears: num 0.212 0.157
6	$ gender : Factor w/ 2 levels “Female”,“Male”: 1 1

```
#
# Visualize missingness pattern.
# A 1 indicates no missing values and 0 indicates missing values.
#
md.pattern(all_data)
```

```
##  /\     /\
## {  `---'  }
## {  O   O  }
## ==>  V <==  No need for mice. This data set is completely observed.
##  \  \|/  /
##   `-----'
```

Figure 1.2: Check for Missing Values


```
##      Country Ccode Year SchoolYears gender
## 3004       1     1    1           1      1 0
##            0     0    0           0      0 0
```

1.6 Data Summary

1.6.1 Using functions from the dplyr package, select 2 numerical variables and one

character/factor variable.

```
#
# Calculate the mean years in education per country and Gender.
# The group_by(Country, Gender) groups the data country and Gender.
#
mean_education <- all_data %>%
#  filter(Country == "Afghanistan") %>%
  group_by(Country, gender) %>%
  summarise(
  # Education statistics
            meanEd = round(mean(SchoolYears, na.rm = TRUE),2),
            medEd = round(median(SchoolYears, na.rm = TRUE),2),
            minEd = round(min(SchoolYears, na.rm = TRUE),2),
            maxEd = round(max(SchoolYears, na.rm = TRUE),2),
            StdDevEd = round(ifelse(is.na(sd(SchoolYears, na.rm = TRUE)),0,
                                    sd(SchoolYears, na.rm = TRUE)),2),
   # Year statistics
           meanYear = round(mean(Year, na.rm = TRUE), 0),
           medYear = round(median(Year, na.rm = TRUE), 0),
           minYear = round(min(Year, na.rm = TRUE), 0),
           maxYear = round(max(Year, na.rm = TRUE), 0),
           StdDevYear = round(sd(Year, na.rm = TRUE), 0),  # Standard deviation for Year

            .groups = "drop")
#
# Display top and bottom 6 rows of the 'mean_education' table.
#
mean_education_12 <- rbind(head(mean_education,6),tail(mean_education,6))

#
# Display the summary as a formatted table.
#
kable(mean_education_12, format = "html",
      caption = "Top Six and Bottom Rows <br> Showing Statistics <br> for Years of Schooling (Ed) and Years of Data Coll
ection (Year) <br> \
      by Country and Gender.") %>%
 kable_styling(bootstrap_options = c("striped", "hover"),
               full_width = FALSE)
```

Table 1.3: Top Six and Bottom Rows
Showing Statistics
for Years of Schooling (Ed) and Years of Data Collection (Year)
by Country and Gender.

Country	gender	meanEd	medEd	minEd	maxEd	StdDevEd	meanYear	medYear	minYear	maxYear	StdDevYear
Afghanistan	Female	0.92	0.75	0.16	2.31	0.88	2002	2015	1975	2022	23
Afghanistan	Male	2.68	3.40	1.18	3.91	1.26	2002	2015	1975	2022	23
Albania	Female	9.18	9.40	8.24	9.83	0.67	2010	2011	2001	2017	6
Albania	Male	10.01	10.22	9.42	10.30	0.37	2010	2011	2001	2017	6
Algeria	Female	4.46	5.16	0.24	6.50	2.43	2003	2008	1971	2019	19
Algeria	Male	5.97	7.17	1.66	7.54	2.45	2003	2008	1971	2019	19
Vietnam	Female	6.55	7.03	3.56	8.52	2.03	2004	2009	1979	2022	19
Vietnam	Male	7.71	8.12	4.64	9.47	1.87	2004	2009	1979	2022	19

Country	gender	meanEd	medEd	minEd	maxEd	StdDevEd	meanYear	medYear	minYear	maxYear	StdDevYear
Zambia	Female	3.88	3.23	2.26	6.15	2.03	1996	1990	1980	2018	20
Zambia	Male	6.21	5.78	4.75	8.10	1.71	1996	1990	1980	2018	20
Zimbabwe	Female	7.05	7.51	4.49	8.07	1.33	2010	2014	1992	2019	10
Zimbabwe	Male	8.40	8.86	6.24	9.46	1.11	2010	2014	1992	2019	10

1.6.2 Using some of the available options inside the kable() function add a caption for the table.

```
#
# Create an education_gap column by calculating the difference between
# male and female min/max years of education.
#
mean_education_with_gap <- mean_education %>%
#  filter(Country == "Afghanistan") %>%
  select(Country, gender, minEd, maxEd) %>%
  pivot_wider(names_from = gender, values_from = c(minEd,maxEd)) %>%
 rename( # Rename columns for clarity
    min_male = minEd_Male,
    min_female = minEd_Female,
    max_male = maxEd_Male,
    max_female = maxEd_Female
  ) %>%
 mutate(
   min_education_gap = min_male - min_female,
   max_education_gap = max_male - max_female
 )


#
# Display top and bottom 6 rows of the 'mean_education' table.
#
mean_education_with_gap_12 <- rbind(head(mean_education_with_gap,6),tail(mean_education_with_gap,6))

#
# Display the summary as a formatted table.
#
kable(mean_education_with_gap_12, format = "html",
      caption = "Top Six and Bottom Rows <br> Showing Years of Schooling Gender Gap <br> by Country.") %>%
 kable_styling(bootstrap_options = c("striped", "hover"),
               full_width = FALSE)
```

Table 1.4: Top Six and Bottom Rows Showing Years of Schooling Gender Gap by Country.

Country	min_female	min_male	max_female	max_male	min_education_gap	max_education_gap
Afghanistan	0.16	1.18	2.31	3.91	1.02	1.60
Albania	8.24	9.42	9.83	10.30	1.18	0.47
Algeria	0.24	1.66	6.50	7.54	1.42	1.04
American Samoa	8.53	9.53	10.32	10.82	1.00	0.50
Andorra	9.96	10.09	11.52	11.70	0.13	0.18

Country	min_female	min_male	max_female	max_male	min_education_gap	max_education_gap
Angola	3.13	4.95	4.12	6.82	1.82	2.70
Uruguay	5.77	5.71	9.25	8.74	-0.06	-0.51
Uzbekistan	11.07	11.48	11.75	12.09	0.41	0.34
Venezuela	2.86	3.53	10.60	10.02	0.67	-0.58
Vietnam	3.56	4.64	8.52	9.47	1.08	0.95
Zambia	2.26	4.75	6.15	8.10	2.49	1.95
Zimbabwe	4.49	6.24	8.07	9.46	1.75	1.39

1.6.3 Describe in two sentences what you observe from the data summary, and write one of the words in the sentence using bold font.

Table 1.3 summarises education data, but this discussion considers only the rows displayed, highlighting critical statistics for years of schooling and record collection. For example, In Afghanistan, there is a significant gap between males and females, with females averaging 0.92 years and males 2.68 years. The median years of education are also lower for females, at 0.75 years compared to 3.40 years for males. In Albania, the educational gap is minimal, with males averaging 10.01 years of schooling and females 9.18 years. Afghanistan has a notably low mean of 0.92 years for females. Consequently, educational opportunities are more equally distributed between genders in Albania, though males still generally have a slight advantage in terms of years of schooling.

The table highlights **gender disparities in education**, particularly in Afghanistan, where girls receive significantly less schooling than boys. Overall, the data collected spans from 1971 to 2022. Afghanistan's records cover a long period from 1975 to 2022, with a standard deviation of 23 years. In contrast, Albania has a standard deviation of 6 years, with data collection from 2001 to 2017. The difference emphasises that the disparity in Afghanistan has occurred over a long period without much improvement.

Table 1.4 shows the education gap for the rows displayed at the lower and higher ends of education. In Afghanistan, girls average only 0.92 years of schooling, peaking at 2.31 years, while boys average 2.68 years, with a maximum of 3.91 years. The significant education gap, ranging from 1.02 to 1.60 years, highlights the disparity in access to education based on gender. However, educational options for both genders are approximately < four years, indicating multifaceted economic, political and cultural issues.

By contrast, in Albania, the gap in education between males and females is small. Females have an average of 9.18 years of schooling, while males have 10.01 years. At the lower education end, the smallest gap is 1.18 years; on the more educated end, the gap is only 0.47 years, showing that both genders have approximately equal access to education at the higher end. The education gap for the less educated is about the same as in Afghanistan, but given the average years of education, which is about double that of Afghanistan, the impact is marginal. Consequently, Afghanistan has a significant education gap between boys and girls, while Albania has made strides toward gender equality in education.

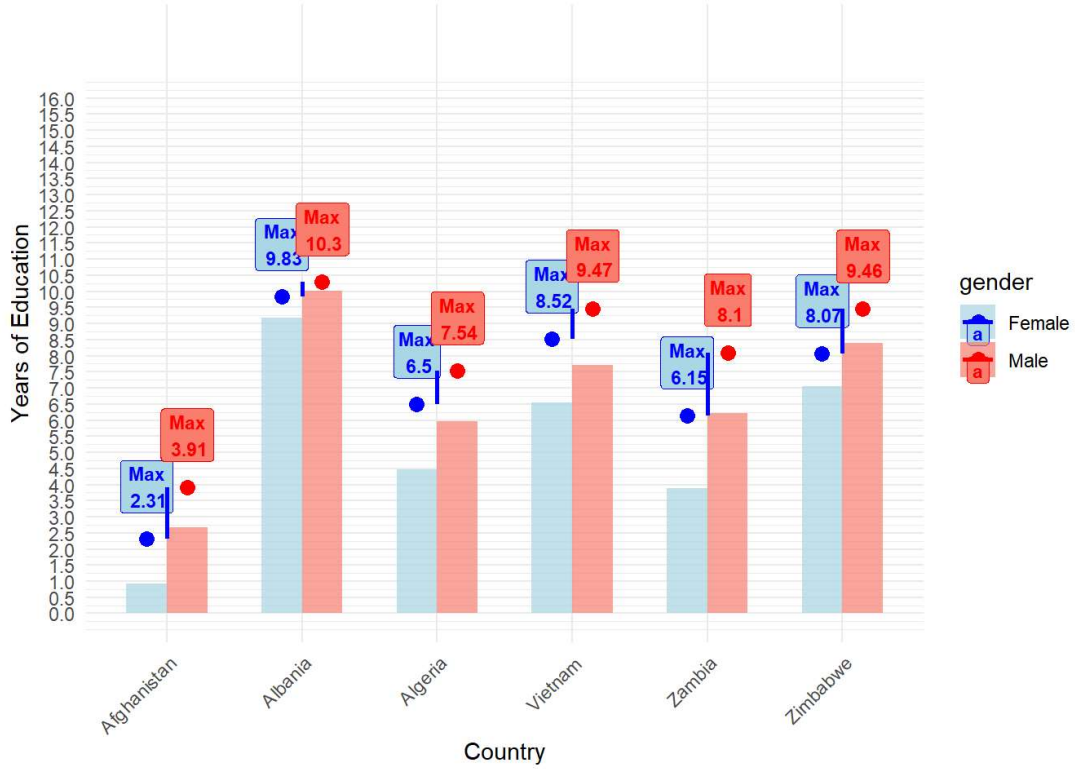
1.7 Visualisations

1.7.1 Create one figure (connected with your research question) of your data and

add a caption using the relevant options inside the R code chunk

```
#
# Exclude Vietnam and Algeria, and create the plot.
# Vietnam and Algeria are missing the required two records
# within the ten rows being examined so they are excluded.
#
ggplot(mean_education_12,
        aes(x = Country, fill = gender)) +
  # Bar plot for mean education.
  geom_bar(aes(y = meanEd), stat = "identity", position = "dodge", width = 0.6, alpha = 0.7) +

  # Point plot for maximum education (overlayed on the bar plot).
  geom_point(aes(y = maxEd, color = gender), size = 3, position = position_dodge(width = 0.6)) +

  # Add text labels for the maximum education points.
  geom_label(aes(y = maxEd, label = paste("Max\n", round(maxEd, 2)), color = gender),
             position = position_dodge(width = 0.6), vjust = -0.5, size = 3, fontface = "bold") +

   # Set the y-axis with more ticks
  scale_y_continuous(breaks = seq(0, 16, by = 0.5), limits = c(0, 18)) +  # Increase the ticks

  # Line to join the points for maximum education.
  geom_line(aes(y = maxEd, group = Country, color = gender), size = 1, position = position_dodge(width = 0.6)) +

  # Customize the plot.
  labs(x = "Country", y = "Years of Education") +
  theme_minimal() +
  scale_fill_manual(values = c("lightblue", "salmon")) +
  scale_color_manual(values = c("blue", "red")) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Figure 1.3: Education Disparity by Country & Gender for Twelve Rows



1.7.2 Using markdown language, add a list with two bullet points describing what you see in the figure using italic font.

- *Afghanistan has a notable gender gap in education, with females averaging only 0.92 years (refer to Table 1.3) of schooling (max 2.31 years) compared to males at 2.68 years (refer to Table 1.3) (max 3.91 years). The disparity is likely due to cultural, social, and political factors, highlighting significant inequalities in access to education for women.*
- *Albania demonstrates better gender equality in education, with girls averaging 9.18 years and boys 10.01 years. The minor differences in the educational gap (refer to Table 1.4) from the least educated (1.18 years) to the highest educated (0.47 years) indicate approximately equal educational opportunities for both genders.*

1.8 Conclusions

1.8.1 Concisely state (no more than 150 words) what conclusions can be made from your data analysis. Once again, remember to link your conclusions back to your original research question

Data analysis reveals a significant gap in educational achievement between boys and girls in poorer countries. For instance, girls in Afghanistan have much less schooling than boys, highlighting stark inequalities. In comparison, Albania offers equal educational access of about two times that of Afghanistan, where Afghanistan shows lower overall educational levels of approximately < four years. These disparities influence economic conditions, political stability, and cultural attitudes, leading to unequal opportunities. Despite global progress, efforts remain essential to ensure equal educational opportunities for all genders worldwide.

2 **Part B**: Mini-Project Repository

2.1 Create your GitHub repository and link your mini-project.

The steps below were used to set-up a repository and create a new branch for part 2.

1. Create a new repository in github (https://github.com/Vincemonash/Assignment1.git (https://github.com/Vincemonash/Assignment1.git)).
2. git init
3. git add .
4. git commit -m “Initial commit for Assignment1 R project”
5. git remote add origin https://github.com/Vincemonash/Assignment1.git (https://github.com/Vincemonash/Assignment1.git)
6. git remote -v
7. git push -u origin master
8. git checkout -b asspart2
9. git add .
10. git commit -m “Changes to accomodate part 2 of assignment 1”
11. git push -u origin asspart2

2.1.1 The data files need to be inside a folder called Data. If you need to process the data to reduce the size do it before you include the data in your repository.

The ‘data’ directory add for the data files.

2.1.2 Ensure you commit changes with clear messages so we can understand what you did in each commit.

Use the following example command: git commit -m “Changes to accomodate part 2 of assignment 1”.

2.1.3 Create two branches and work in them for a few commits at different times.

One branch created: “git checkout -b asspart2”. Second branch created: “git checkout -b asspart2B”.

echo “# ETO5513 Assignment 1” > README.md git add README.md git commit -m “Add README.md for Assignment 1”

git checkout master git pull origin master git merge asspart2B

git switch asspart2 vi README.md

git switch asspart2B vi README.md

2.1.4 Merge them with the master branch.

git switch master git checkout master git pull origin master git merge asspart2B

git merge asspart2

2.1.5 Make sure that at least one branch creates a conflict with your master branch.

2.1.5.1 Create commits that show clearly that you merged the branches. In addition, make sure that you have a commit showing that you fixed at least one merging conflict.

$ git merge asspart2 Auto-merging README.md CONFLICT (content): Merge conflict in README.md Automatic merge failed; fix conflicts and then commit the result.

ETO5513 Assignment 1

“<<<<<<<” HEAD ## Description: Assignment 1 created in branch asspart2B. “=======” ## Description: Change made in asspart2B branch. “>>>>>>>” asspart2

git commit -m “Resolved merge conflict in README.md”

$ vi README.md

USER@DESKTOP-KGR3EOI (mailto:USER@DESKTOP-KGR3EOI) MINGW64 /d/Monash/ETO5513/Assignment1 (master|MERGING) $

USER@DESKTOP-KGR3EOI (mailto:USER@DESKTOP-KGR3EOI) MINGW64 /d/Monash/ETO5513/Assignment1 (master|MERGING) $

USER@DESKTOP-KGR3EOI (mailto:USER@DESKTOP-KGR3EOI) MINGW64 /d/Monash/ETO5513/Assignment1 (master|MERGING) $ git add README.md

USER@DESKTOP-KGR3EOI (mailto:USER@DESKTOP-KGR3EOI) MINGW64 /d/Monash/ETO5513/Assignment1 (master|MERGING) $ git commit -m “Resolved merge conflict in README.md” [master cde8057] Resolved merge conflict in README.md

USER@DESKTOP-KGR3EOI (mailto:USER@DESKTOP-KGR3EOI) MINGW64 /d/Monash/ETO5513/Assignment1 (master) $ git push Enumerating objects: 15, done. Counting objects: 100% (14/14), done. Delta compression using up to 12 threads Compressing objects: 100% (11/11), done. Writing objects: 100% (11/11), 1.20 KiB | 1.20 MiB/s, done. Total 11 (delta 5), reused 0 (delta 0), pack-reused 0 (from 0) remote: Resolving deltas: 100% (5/5), completed with 1 local object. To https://github.com/Vincemonash/Assignment1.git (https://github.com/Vincemonash/Assignment1.git) 94f65d2..cde8057 master -> master

2.2 Expand on Your Simple report

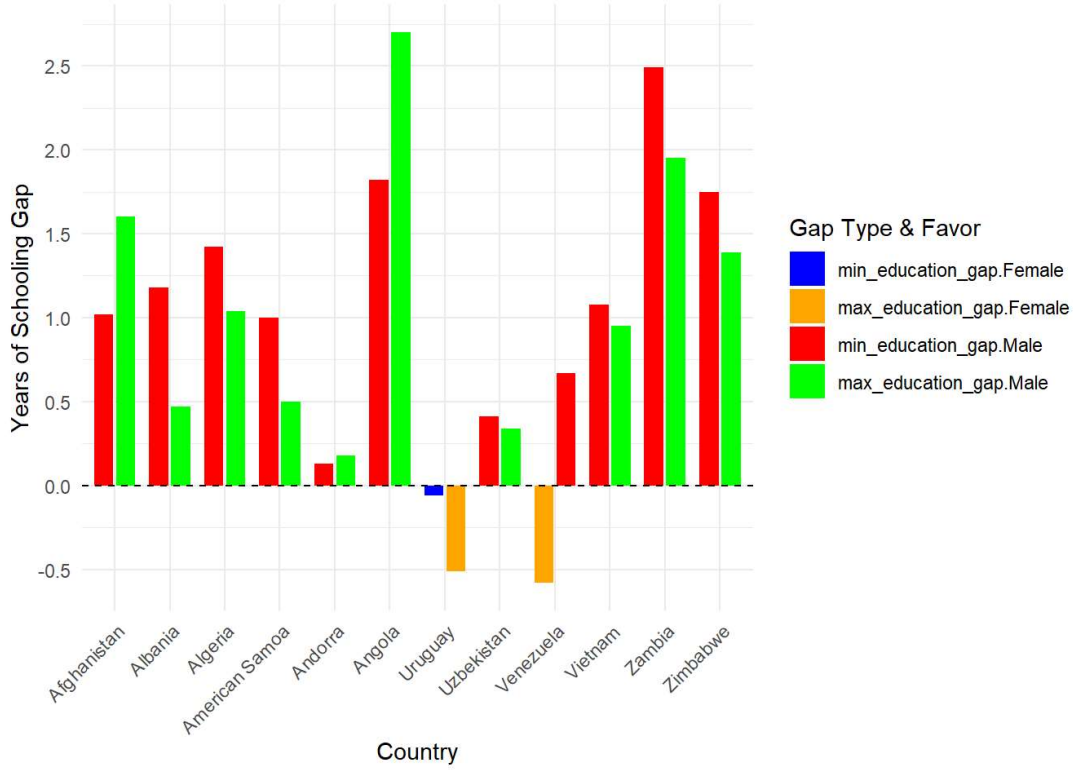
2.2.1 include one additional figure and table describing the data

Table 1.4 shows the education gap at the low and higher ends of education for the ten rows shown. The education gap between boys and girls is significant in countries like Afghanistan, where girls have fewer years of schooling due to barriers. In contrast, Albania shows progress, with girls and boys attaining similar educational levels.

Figure 2.1 shows the data from Table 1.4 and illustrates the difference in the years of schooling between males and females in twelve countries. Andorra has a small gap (0.13 to 0.18), while Angola shows a significant gap (1.82 to 2.70). In Uruguay and Venezuela, the gap favours females, whereas in Afghanistan and Algeria, there are more considerable gaps favouring males. Countries like Uzbekistan and Vietnam exhibit relatively balanced gaps. Possible reasons for differences in the data emphasize the impact of cultural, social, and economic factors on educational gender equality, highlighting the ongoing need for equal education for all genders.

```
#
# Transform the data into long format for ggplot.
#
data_long <- reshape2::melt(mean_education_with_gap_12,
                            id.vars = "Country",
                            measure.vars = c("min_education_gap","max_education_gap"),
                            variable.name = "Gap_Type",
                            value.name = "Gap_Value")

#
# Add a column to distinguish male vs female gap visually.
#
data_long$Gap_Favor <- ifelse(data_long$Gap_Value > 0, "Male", "Female")

#
# Produce the graph that shows which way education gap points:
# Negative is female and positive is male.
#
ggplot(data_long, aes(x = Country, y = Gap_Value, fill = interaction(Gap_Type, Gap_Favor))) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8), width = 0.7) +
  scale_fill_manual(values = c(
    "min_education_gap.Female" = "blue",
    "min_education_gap.Male" = "red",
    "max_education_gap.Male" = "green",
    "max_education_gap.Female" = "orange"
  )) +
  geom_hline(yintercept = 0, color = "black", linetype = "dashed") +
  scale_y_continuous(breaks = seq(-1, 10, 0.5)) + # Adds more y-axis ticks from -1 to 10 in steps of 0.5
  labs(
    x = "Country",
    y = "Years of Schooling Gap",
    fill = "Gap Type & Favor"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Figure 2.1: Gender Education Gap by Country for Twelve Rows



References

Mean years of schooling (ISCED 1 or higher), female. (2021). [Dataset]. Global Change Data Lab. https://ourworldindata.org/grapher/mean-years-of-schooling-female (https://ourworldindata.org/grapher/mean-years-of-schooling-female)

Mean years of schooling (ISCED 1 or higher), male. (2021). [Dataset]. Global Change Data Lab. https://ourworldindata.org/grapher/mean-years-of-schooling-male (https://ourworldindata.org/grapher/mean-years-of-schooling-male)