World Education and Unemployment

The Potential Relation and The Gender Disparities

STAT 4380: Data Science
Prof. Jacob Rozan
Tommy Nguyen

Table of contents

01 The datasets

– Word education data
– Unemployment data

02 Education & unemployment

– Data Wrangling
– Data visualization
– Conclusions

03 Gender disparities

– Data Wrangling
– Data visualization
– Conclusions

04 Concerns & Suggestions

– Addressing shortcomings
– Potential future work

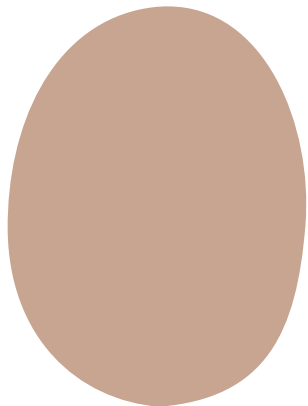1) Is there a relationship between education and unemployment?

2) Is there a disparities between male and female in these two fields?

01

The datasets

Global_Education.csv (21.64 kB)

⬇ ⛶ >

Detail Compact Column

10 of 29 columns ⌄

About this file

This dataset provides a comprehensive global perspective on education, offering vital insights into diverse education systems worldwide. It covers essential metrics like out-of-school rates, completion rates, proficiency levels, literacy rates, birth rates, and enrollment in primary and tertiary education. It's a valuable resource for researchers, educators, and policymakers looking to assess and improve education systems globally.

△ Countries and are...	△ Latitude	△ Longitude	# OOSR_Pre0Prima...	# OOSR_Pre0Prima...	# OOSR_Primary_A...	# OOSR_Primary_A...	# OOS
Name of the countries and areas.	Latitude coordinates of the geographical location.	Longitude coordinates of the geographical location.	Out-of-school rate for pre-primary age males.	Out-of-school rate for pre-primary age females.	Out-of-school rate for primary age males.	Out-of-school rate for primary age females.	Out-of-... lower s... males.
202 unique values	0.02 — 65	0.82 — 178	0 — 96	0 — 96	0 — 58	0 — 67	0 —
Afghanistan	33.93911	67.709953	0	0	0	0	0
Albania	41.153332	20.168331	4	2	6	3	6
Algeria	28.033886	1.659626	0	0	0	0	0
Andorra	42.506285	1.521801	0	0	0	0	0
Angola	11.202692	17.873887	31	39	0	0	0
Anguilla	18.220554	63.068615	14	0	0	0	0
Antigua and Barbuda	17.060816	61.796428	14	4	4	1	1
Argentina	38.416097	63.616672	2	2	0	0	0
Armenia	40.069099	45.038189	52	50	9	9	11
Australia	25.274398	133.775136	13	14	0	0	2
Austria	47.516231	14.550072	0	0	0	0	1
Azerbaijan	40.143105	47.576927	32	19	10	7	0
The Bahamas	25.03428	77.39628	0	0	0	0	23
Bahrain	26.0667	50.5577	31	28	2	3	7

- Economy: Public debt
- Economy: Public debt - note
- Economy: Real GDP (purchasing power parity)
- Economy: Real GDP (purchasing power parity) - note
- Economy: Real GDP growth rate
- Economy: Real GDP growth rate - note
- Economy: Real GDP per capita
- Economy: Real GDP per capita - note
- Economy: Reserves of foreign exchange and gold
- Economy: Reserves of foreign exchange and gold - note
- Economy: Taxes and other revenues
- Economy: Taxes and other revenues - note
- Economy: Unemployment rate
- Economy: Unemployment rate - note
- Economy: Unemployment, youth ages 15-24 - female
- Economy: Unemployment, youth ages 15-24 - male
- Economy: Unemployment, youth ages 15-24 - note
- Economy: Unemployment, youth ages 15-24 - total

Contains information about Energy, Geography, Environment, Economy, Transportation, Government, Military, etc..

About Dataset

The World Factbook is an almanac published by the Central Intelligence Agency (CIA).

The World Factbook, produced for US policymakers and coordinated throughout the US Intelligence Community, presents the basic realities about the world in which we live. We share these facts with the people of all nations in the belief that knowledge of the truth underpins the functioning of free societies.

Provenance
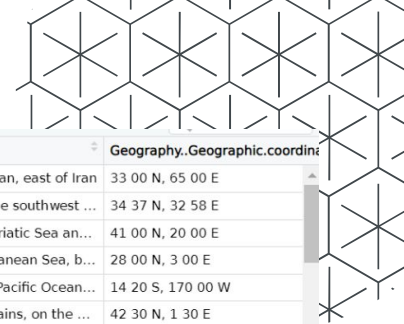
	Country	url	Introduction..Background	Geography..Location	Geography..Geographic.coordina
1	Afghanistan	https://www.cia.gov/the-world-factbook/countries/af...	Ahmad Shah DURRANI unified the Pashtun tribes an...	Southern Asia, north and west of Pakistan, east of Iran	33 00 N, 65 00 E
2	Akrotiri	https://www.cia.gov/the-world-factbook/countries/ak...	By terms of the 1960 Treaty of Establishment that c...	Eastern Mediterranean, peninsula on the southwest ...	34 37 N, 32 58 E
3	Albania	https://www.cia.gov/the-world-factbook/countries/al...	Albania declared its independence from the Ottoma...	Southeastern Europe, bordering the Adriatic Sea an...	41 00 N, 20 00 E
4	Algeria	https://www.cia.gov/the-world-factbook/countries/al...	Algeria has known many empires and dynasties star...	Northern Africa, bordering the Mediterranean Sea, b...	28 00 N, 3 00 E
5	American Samoa	https://www.cia.gov/the-world-factbook/countries/a...	Tutuila was settled by 1000 B.C. and the island serv...	Oceania, group of islands in the South Pacific Ocean...	14 20 S, 170 00 W
6	Andorra	https://www.cia.gov/the-world-factbook/countries/an...	The Moorish invasion of Spain in the 8th century an...	Southwestern Europe, Pyrenees mountains, on the ...	42 30 N, 1 30 E
7	Angola	https://www.cia.gov/the-world-factbook/countries/an...	From the late 14th to the mid 19th century a Kingdo...	Southern Africa, bordering the South Atlantic Ocean...	12 30 S, 18 30 E
8	Anguilla	https://www.cia.gov/the-world-factbook/countries/an...	Colonized by English settlers from Saint Kitts in 165...	Caribbean, islands between the Caribbean Sea and ...	18 15 N, 63 10 W
9	Antarctica	https://www.cia.gov/the-world-factbook/countries/an...	Speculation over the existence of a "southern land" ...	continent mostly south of the Antarctic Circle	90 00 S, 0 00 E
10	Antigua and Barbuda	https://www.cia.gov/the-world-factbook/countries/an...	The Siboney were the first people to inhabit the isla...	Caribbean, islands between the Caribbean Sea and ...	17 03 N, 61 48 W
11	Argentina	https://www.cia.gov/the-world-factbook/countries/ar...	In 1816, the United Provinces of the Rio Plata declar...	Southern South America, bordering the South Atlant...	34 00 S, 64 00 W
12	Armenia	https://www.cia.gov/the-world-factbook/countries/ar...	Armenia prides itself on being the first nation to for...	Southwestern Asia, between Turkey (to the west) an...	40 00 N, 45 00 E
13	Aruba	https://www.cia.gov/the-world-factbook/countries/ar...	Discovered and claimed for Spain in 1499, Aruba w...	Caribbean, island in the Caribbean Sea, north of Ve...	12 30 N, 69 58 W
14	Ashmore and Cartier Islands	https://www.cia.gov/the-world-factbook/countries/as...	Indonesian fishermen have long fished in the area a...	Southeastern Asia, islands in the Indian Ocean, mid...	12 25 S, 123 20 E
15	Australia	https://www.cia.gov/the-world-factbook/countries/au...	Aboriginal Australians arrived on the continent at le...	Oceania, continent between the Indian Ocean and t...	27 00 S, 133 00 E
16	Austria	https://www.cia.gov/the-world-factbook/countries/au...	Once the center of power for the large Austro-Hung...	Central Europe, north of Italy and Slovenia	47 20 N, 13 20 E
17	Azerbaijan	https://www.cia.gov/the-world-factbook/countries/az...	Azerbaijan - a secular nation with a majority-Turkic a...	Southwestern Asia, bordering the Caspian Sea, bet...	40 30 N, 47 30 E
18	Bahamas, The	https://www.cia.gov/the-world-factbook/countries/ba...	Lucayan Indians inhabited the islands when Christo...	chain of islands in the North Atlantic Ocean, southe...	24 15 N, 76 00 W
19	Bahrain	https://www.cia.gov/the-world-factbook/countries/ba...	In 1783, the Sunni Al-KHALIFA family took power in ...	Middle East, archipelago in the Persian Gulf, east of ...	26 00 N, 50 33 E
20	Baker Island	https://www.cia.gov/the-world-factbook/countries/ba...	All of the following US Pacific island territories exce...	Oceania	
21	Bangladesh	https://www.cia.gov/the-world-factbook/countries/ba...	The huge delta region formed at the confluence of t...	Southern Asia, bordering the Bay of Bengal, betwee...	24 00 N, 90 00 E
22	Barbados	https://www.cia.gov/the-world-factbook/countries/ba...	The island was uninhabited when first settled by the...	Caribbean, island in the North Atlantic Ocean, north...	13 10 N, 59 32 W
23	Belarus	https://www.cia.gov/the-world-factbook/countries/be...	After seven decades as a constituent republic of the...	Eastern Europe, east of Poland	53 00 N, 28 00 E
24	Belgium	https://www.cia.gov/the-world-factbook/countries/be...	Belgium became independent from the Netherlands...	Western Europe, bordering the North Sea, between ...	50 50 N, 4 00 E
25	Belize	https://www.cia.gov/the-world-factbook/countries/be...	Belize was the site of several Mayan city states until...	Central America, bordering the Caribbean Sea, betw...	17 15 N, 88 45 W
26	Benin	https://www.cia.gov/the-world-factbook/countries/be...	Present day Benin is comprised of about 42 ethnic g...	Western Africa, bordering the Bight of Benin, betwe...	9 30 N, 2 15 E
27	Bermuda	https://www.cia.gov/the-world-factbook/countries/be...	Bermuda was first settled in 1609 by shipwrecked E...	North America, group of islands in the North Atlanti...	32 20 N, 64 45 W
28	Bhutan	https://www.cia.gov/the-world-factbook/countries/bh...	Following Britain's victory in the 1865 Duar War, Bri...	Southern Asia, between China and India	27 30 N, 90 30 E

Showing 1 to 28 of 258 entries, 1071 total columns

What are the considered factors?

World Education Data

+ Out of school rate (OOSR): lower secondary male, upper secondary male, lower secondary female, upper secondary female
+ Completion rate: lower secondary, upper secondary
+ Gross tertiary education enrollment

CIA World Factbook

+ Unemployment rate
+ Unemployment rate, youth ages 15 – 24
+ Unemployment rate, youth ages 15 – 24 _ female
+ Unemployment rate, youth ages 15 – 24_ male

Provenance ∧

SOURCES

Out of school rate - UNESCO Institute for Statistics (last updated in March 2021): Shows the percentage of children and youth not attending school.

Completion rate - UNICEF Global Database (last updated in April 2021): Indicates the percentage of students successfully completing their education.

Provenance

SOURCES

The World Factbook 2021. Washington, DC: Central Intelligence Agency, 2021. https://www.cia.gov/the-world-factbook/

02
Education & Unemployment

Data Wrangling

```r
{r data_wrangling_unemployment_rates}
#Read csv files
education_df <- read.csv('Global_Education.csv')
countries_df <- read.csv('countries.csv')

#Rename columns
sub_countries_df <- countries_df[,c(1,6,32,224:227)]
sub_countries_df <- sub_countries_df %>%
    rename(Unemployement_rate_ages.15.24
=Economy..Youth.unemployment.rate..ages.15.24....total ) %>%
    rename(Unemployement_rate_ages.15.24_male
=Economy..Youth.unemployment.rate..ages.15.24....male) %>%
    rename(Unemployement_rate_ages.15.24_female
=Economy..Youth.unemployment.rate..ages.15.24....female) %>%
    rename(Population=People.and.Society..Population) %>%
    rename(Unemployment_Rate_2021 = Economy..Unemployment.rate) %>%
    filter(grepl('2021',Unemployment_Rate_2021))

#Set uniform format for each columns
sub_countries_df$Unemployment_Rate_2021 <-
    str_extract(sub_countries_df$Unemployment_Rate_2021,"\\d*\\.?\\d*")
sub_countries_df$Unemployement_rate_ages.15.24 <-
    str_extract(sub_countries_df$Unemployement_rate_ages.15.24,"\\d*\\.?\\d*
")
sub_countries_df$Unemployement_rate_ages.15.24_female <-
    str_extract(sub_countries_df$Unemployement_rate_ages.15.24_female,"\\d*\
\.?\\d*")
sub_countries_df$Unemployement_rate_ages.15.24_male <-
    str_extract(sub_countries_df$Unemployement_rate_ages.15.24_male,"\\d*\\.
?\\d*")
sub_countries_df$Population <- sub_countries_df$Population %>%
    str_extract("\\d*\\,?\\d*\\,?\\d*\\,?\\d*")
sub_countries_df$Population <- gsub(",","",sub_countries_df$Population)
```

Population
38,346,720 (2022 est.)
3,095,344 (2022 est.)
44,178,884 (2022 est.)

Unemployment_Rate_2021
13.28% (2021 est.)11.71% (2020 est.)11.22% (201...
11.82% (2021 est.)13.33% (2020 est.)11.47% (201...
12.7% (2021 est.)12.55% (2020 est.)10.51% (2019 ...
8.53% (2021 est.)8.33% (2020 est.)7.42% (2019 est.)

Unemployement_rate_ages.15.24_male
18.6%
28%
27.8%

Unemployement_rate_ages.15.24_female
26.4% (2021 est.)
27.6% (2021 est.)
54% (2021 est.)
18.8% (2021 est.)

```r
{r data_wrangling_merging_two_dataframes}
#Reformat "country" values for merging
sub_countries_df$Country <- sub("^([^,]+), (.+)$", "\\2 \\1", sub_countries_df$Country)
sub_countries_df$Country <-  sub("\\s*\\([^)]+\\)", "", sub_countries_df$Country)

#Rename for merging
old_name <- c("Burma","Czechia","Cabo Verde","Ireland","Timor-Leste", "Virgin Islands")
new_name <- c("Myanmar","Czech Republic","Cape Verde", "Republic of Ireland","East Timor","British Virgin
Islands")
sub_countries_df$Country[sub_countries_df$Country %in% old_name] <- new_name
education_df <- education_df %>%
    rename(Country = Countries.and.areas) %>%
    select(-Unemployment_Rate)

#Merging
inner_df <- education_df %>%
    inner_join(sub_countries_df, by="Country")
inner_df <- inner_df %>%
    rename(Continents_Regions = Geography..Map.references )

#Correcting some values
inner_df$Continents_Regions[inner_df$Country=="France"] <- "Europe"
inner_df$Continents_Regions[inner_df$Country=="Iceland"] <- "Europe"
inner_df$Continents_Regions[inner_df$Country=="Ukraine"] <- "Europe"

#Filter only needed columns
merged_df <- inner_df %>%
    select(-c(Latitude,Longitude,Birth_Rate, 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,Gro
ss_Primary_Education_Enrollment,Completion_Rate_Primary_Female,Completion_Rate_Primary_Male,OOSR_PreOPrimary_
Age_Female, OOSR_PreOPrimary_Age_Male, OOSR_Primary_Age_Female, OOSR_Primary_Age_Male))
colnames(merged_df)

#Set remaining columns to be numeric
cols <- c("Unemployment_Rate_2021","Unemployement_rate_ages.15.24","Unemployement_rate_ages.15.24_male","Unem
ployement_rate_ages.15.24_female","Population")
merged_df[cols] <- lapply(merged_df[cols], as.numeric)
typeof(merged_df$Population)
```

38	Congo, Democratic Republic of the
39	Congo, Republic of the

Country	G
Timor-Leste	S

Countries.and.areas
East Timor

Continents_Regions
AsiaEurope

Continents_Regions	
Arctic Region	3

Resulting dataframe

```
[1] "Country"                                "OOSR_Lower_Secondary_Age_Male"              "OOSR_Lower_Secondary_Age_Female"
[4] "OOSR_Upper_Secondary_Age_Male"          "OOSR_Upper_Secondary_Age_Female"            "Completion_Rate_Lower_Secondary_Male"
[7] "Completion_Rate_Lower_Secondary_Female" "Completion_Rate_Upper_Secondary_Male"       "Completion_Rate_Upper_Secondary_Female"
[10] "Gross_Tertiary_Education_Enrollment"   "Continents_Regions"                         "Population"
[13] "Unemployment_Rate_2021"                "Unemployement_rate_ages.15.24"              "Unemployement_rate_ages.15.24_male"
[16] "Unemployement_rate_ages.15.24_female"
```

	Country	OOSR_Lower_Secondary_Age_Male	OOSR_Lower_Secondary_Age_Female	OOSR_Upper_Secondary_Age_Male	OOSR_Upper_Secondary_Age_Female
1	Afghanistan	0	0	44	69
2	Albania	6	1	21	15
3	Algeria	0	0	0	0
4	Angola	0	0	0	0
5	Argentina	0	0	15	7
6	Armenia	11	9	16	4
7	Australia	2	3	10	6

Ratios to compare

- Gross Tertiary Education Enrollment vs Unemployment Rate
- Gross Tertiary Education Enrollment vs Unemployment Rate age 15 – 24.
- Upper/Lower Secondary Male Completion Rate vs Male Unemployment Rate Ages 15–24
- Upper/Lower Secondary Female Completion Rate vs Female Unemployment Rate Ages 15–24
- Male OOSR Upper/Lower Secondary vs Male Unemployment Rate Ages 15–24
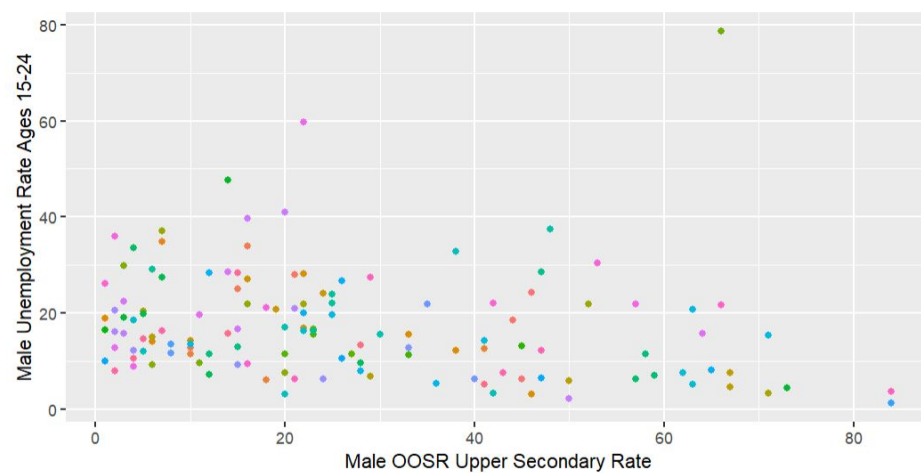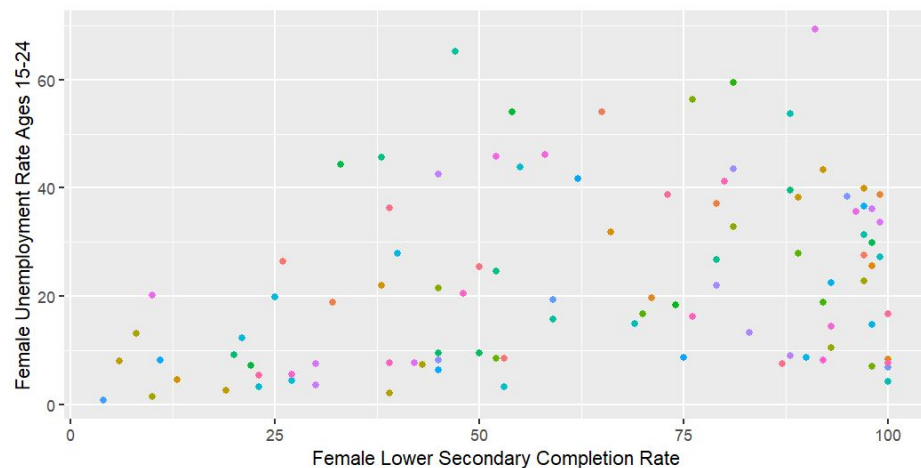- Female OOSR Upper/Lower Secondary vs Female Unemployment Rate Ages 15–24

Data Visualization
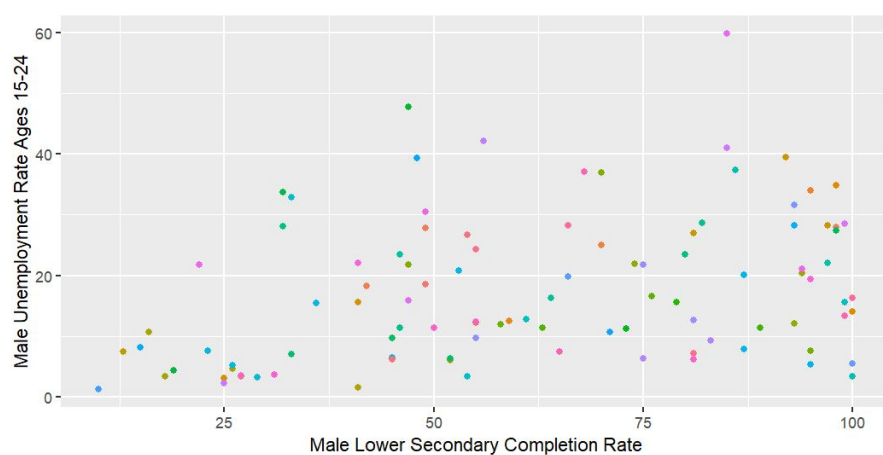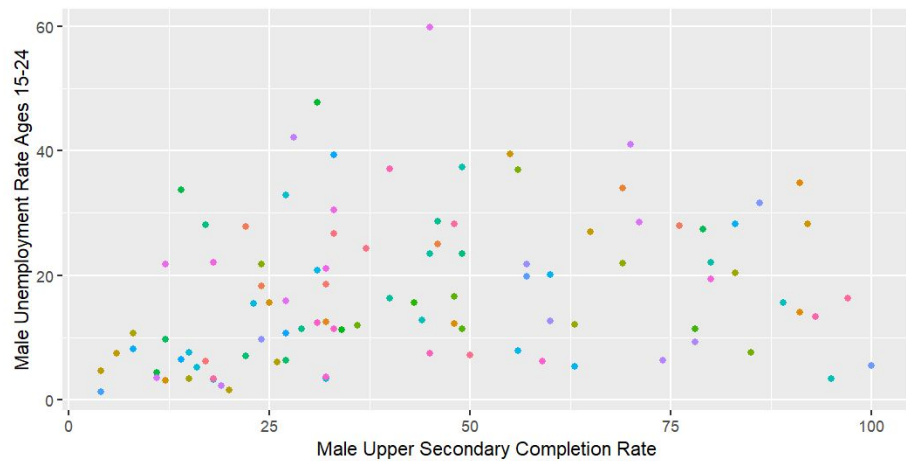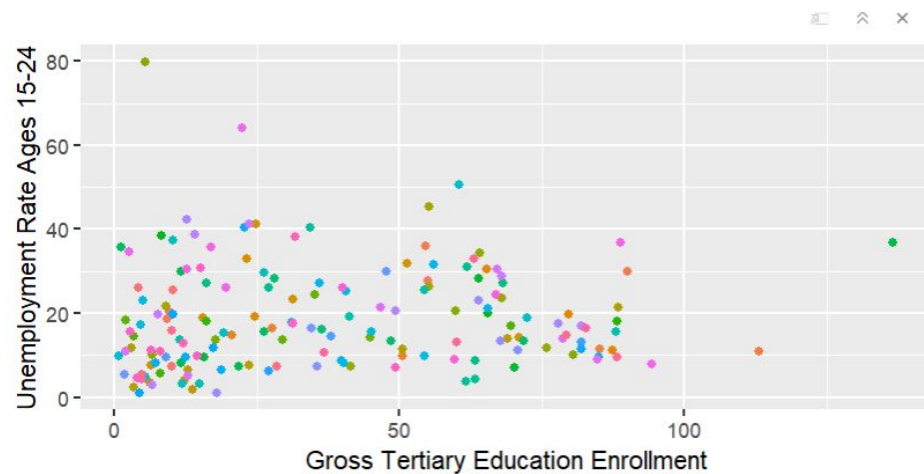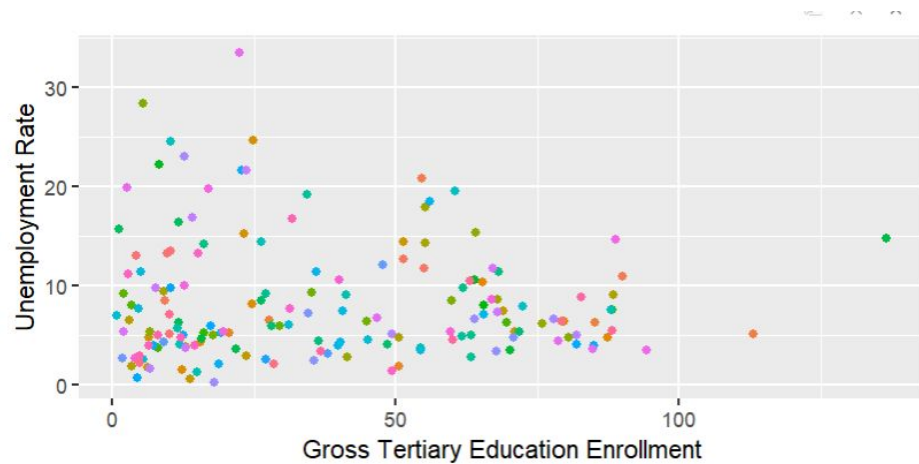
New questions to consider

- Comparing continents?
- Comparing through maps?

Data Wrangling - Continents

```r
```{r}

columns <- colnames(merged_df %>%
                    select(-Country, -Continents_Regions,-Population))

continents2_df <-  merged_df %>%
    group_by(Continents_Regions) %>%
    summarize(across(columns,
                     ~sum(.*Population)/sum(Population),.names="{col}"))

continents_df <- merged_df %>%
    group_by(Continents_Regions) %>%
    summarize_at(vars(-Country),mean)
colnames(continents_df)

continent_colors <- c("Africa" = "red", "Asia" = "blue", "Europe" = "aquamarine", "Central
America and the Caribbean"= "aquamarine3","Middle East" = "yellow", "North America" =
"purple", "Oceania" = "lightpink1", "South America" = "brown", "Southeast Asia" =
"darkgoldenrod3")

ggplot(continents2_df, aes(x=Gross_Tertiary_Education_Enrollment,y=Unemployment_Rate_2021,
color=Continents_Regions))+
    geom_point()+
    scale_color_manual(values=continent_colors)
continents2_df

```
```

	Continents_Regions	OOSR_Lower_Secondary_Age_Male
1	Africa	14.351885
2	Asia	6.864562
3	Central America and the Caribbean	15.577978
4	Europe	2.954788
5	Middle East	8.204469
6	North America	4.308054
7	Oceania	3.588976
8	South America	3.775064
9	Southeast Asia	13.168723

Data Visualization - Continents (Point plots)

Data Visualization - Continents (Maps)

- Correlation between Unemployment Rate and total OOSR in North America, Asia, Europe, and Australia
- South America: Unemployment rate in the lower middle and total OOSR also in the lower middle
- Africa continent containing countries/regions with highest unemployment rate is also the continent with highest total OOSR
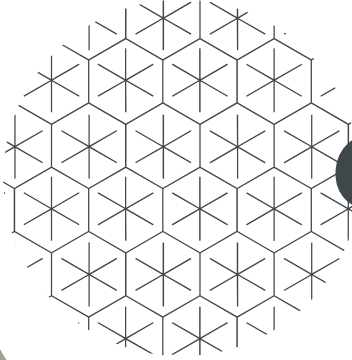
Conclusion:

There are <u>possible correlations</u> between unemployment rate and out of school rate (secondary school) among continents and regions.

There are <u>not much correlation</u> between unemployment rate and gross tertiary education enrollment. Similarly, <u>No correlation</u> can be seen when dissecting the out of school rate (secondary school) by gender and lower/upper level and compare it to the unemployment rate/ unemployment rate ages 15-24.

03

Gender Disparities in Unemployment and Education

Data Wrangling

```r
{r}
merged2_df <- continents2_df %>%
    select(.,-c(Gross_Tertiary_Education_Enrollment,Unempl
oyment_Rate_2021, Unemployement_rate_ages.15.24))
merged2_df <- melt(merged2_df,
id.vars="Continents_Regions", value.name="Rates",
variable.name="Categories")
merged2_df$Rates <- as.numeric(merged2_df$Rates)


merged3 <- merged2_df %>%
    group_by(Continents_Regions) %>%
    mutate(Category = gsub("_Male|_Female", "", Categories)
%>%
    gsub("_"," ",.),
    Gender = gsub(".*_", "", Categories)) %>%
    select(Continents_Regions,Category,Rates,Gender) %>%
    mutate(Gender=ifelse(tolower(Gender)=="female","Female",
"Male"))

merged3$Category <- gsub("^Unemployement rate
ages.15.24.*", "Unemployment rate ages.15.24",
merged3$Category, ignore.case = TRUE)
```

	Continents_Regions	Categories	Rates
1	Africa	OOSR_Lower_Secondary_Age_Male	14.351884
2	Asia	OOSR_Lower_Secondary_Age_Male	6.864562
3	Central America and the Caribbean	OOSR_Lower_Secondary_Age_Male	15.577978
4	Europe	OOSR_Lower_Secondary_Age_Male	2.954787
5	Middle East	OOSR_Lower_Secondary_Age_Male	8.204053
6	North America	OOSR_Lower_Secondary_Age_Male	4.308053
7	Oceania	OOSR_Lower_Secondary_Age_Male	3.588975
8	South America	OOSR_Lower_Secondary_Age_Male	3.775064
9	Southeast Asia	OOSR_Lower_Secondary_Age_Male	13.168723
10	Africa	OOSR_Lower_Secondary_Age_Female	15.634695
11	Asia	OOSR_Lower_Secondary_Age_Female	4.586398
12	Central America and the Caribbean	OOSR_Lower_Secondary_Age_Female	14.946190
13	Europe	OOSR_Lower_Secondary_Age_Female	2.681705
14	Middle East	OOSR_Lower_Secondary_Age_Female	11.120154
15	North America	OOSR_Lower_Secondary_Age_Female	2.459552

	Continents_Regions	Category	Rates	Gender
1	Africa	OOSR Lower Secondary Age	14.3518845	Male
2	Asia	OOSR Lower Secondary Age	6.8645622	Male
3	Central America and the Caribbean	OOSR Lower Secondary Age	15.5779783	Male
4	Europe	OOSR Lower Secondary Age	2.9547876	Male
5	Middle East	OOSR Lower Secondary Age	8.2044690	Male
6	North America	OOSR Lower Secondary Age	4.3080536	Male
7	Oceania	OOSR Lower Secondary Age	3.5889758	Male
8	South America	OOSR Lower Secondary Age	3.7750641	Male
9	Southeast Asia	OOSR Lower Secondary Age	13.1687232	Male
10	Africa	OOSR Lower Secondary Age	15.6346951	Female
11	Asia	OOSR Lower Secondary Age	4.5863980	Female
12	Central America and the Caribbean	OOSR Lower Secondary Age	14.9461909	Female
13	Europe	OOSR Lower Secondary Age	2.6817058	Female

Data Visualization

Unemployment rate ages 15-24 among continents
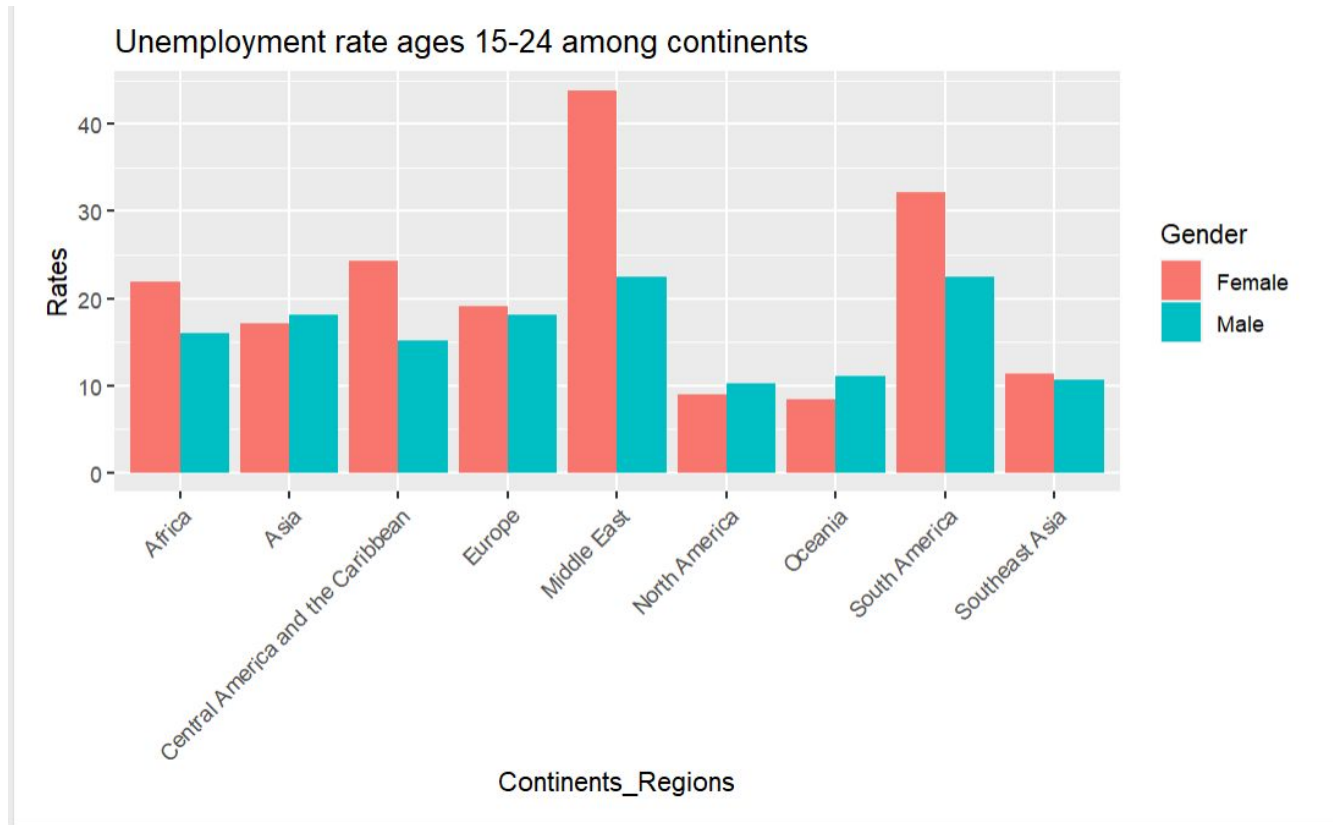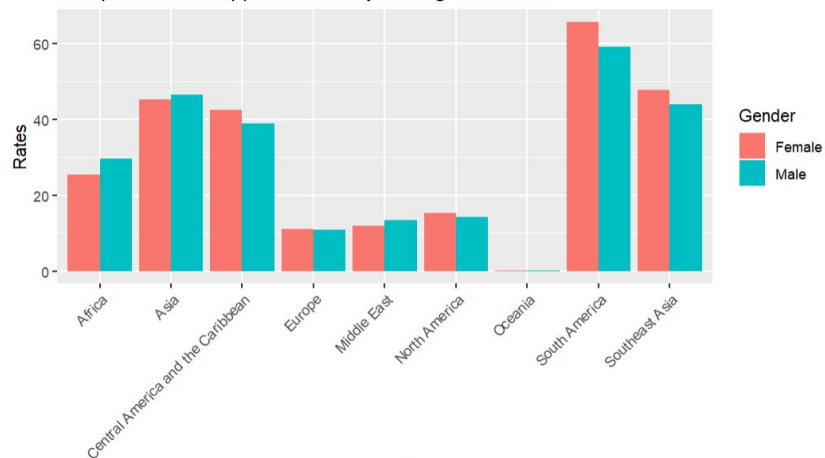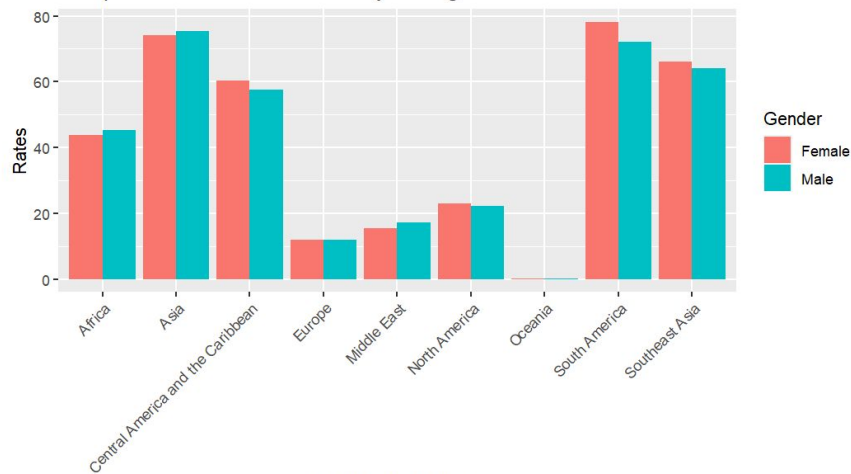
Noticeable gender differences in unemployment rate in: Africa, Central America and the Caribbean, Middle East, South America
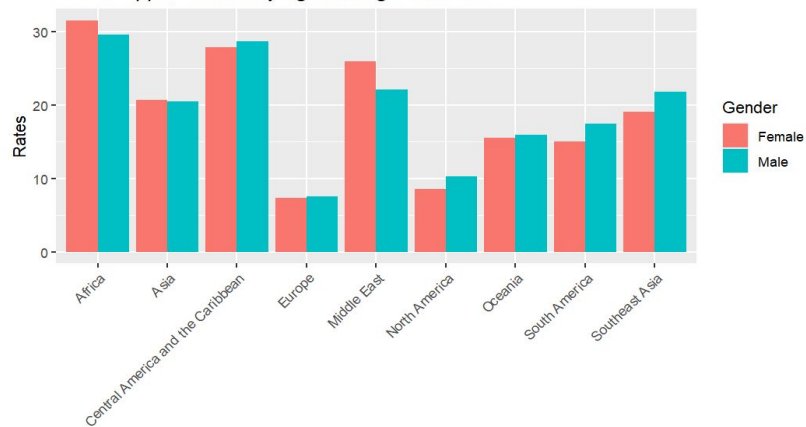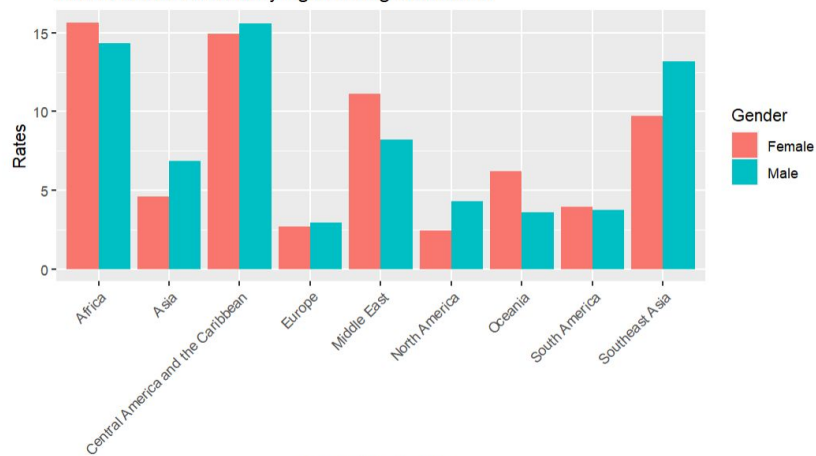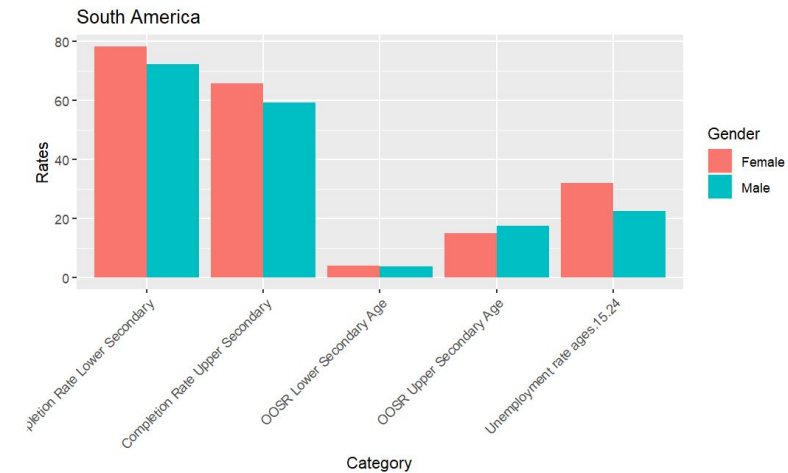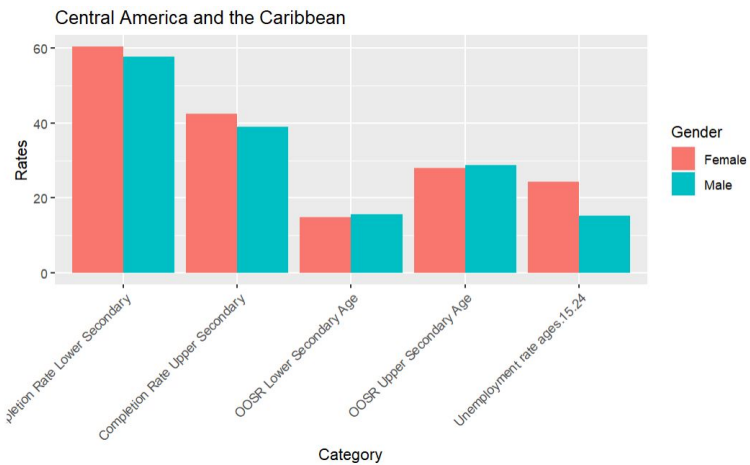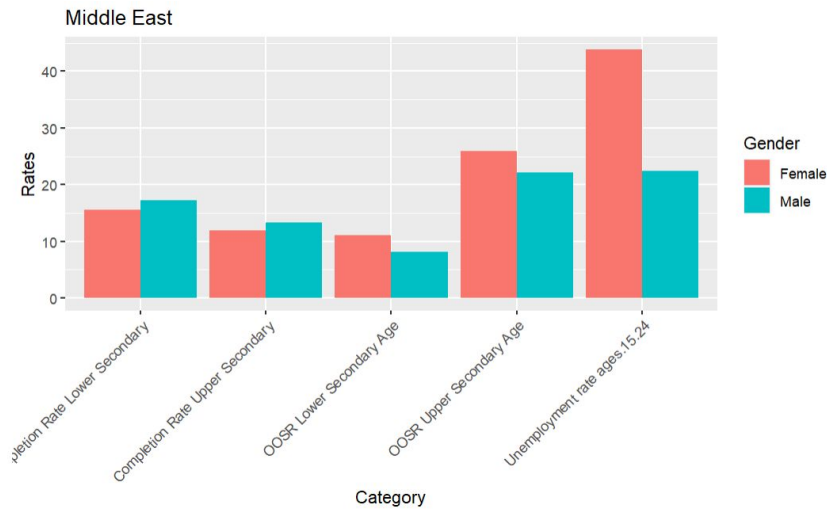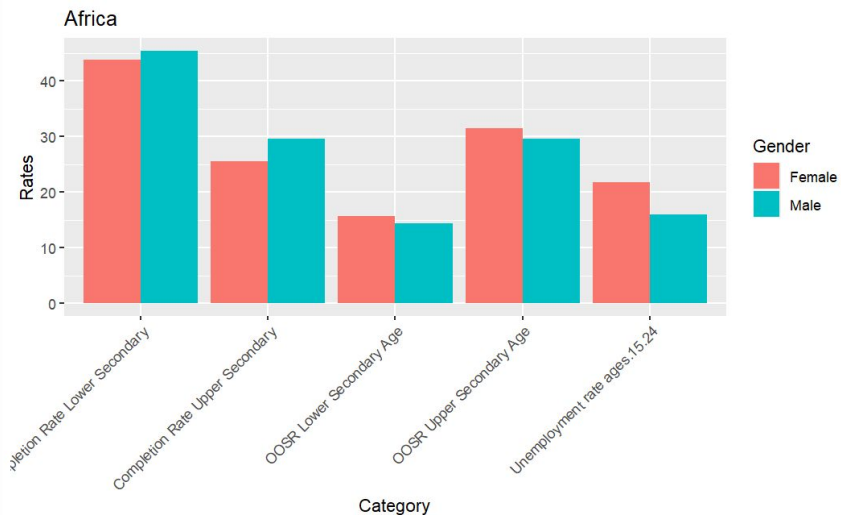
Conclusion:

- Great difference in employment rate between male and female in the Africa, Central America and the Caribbean, Middle East, and South America.
- In Africa and the Middle East, female OOSR is higher than that of male for secondary school. Female secondary school completion rate is also lower.
- Yet, in Central America and the Caribbean and South America, female has higher completion rate and lower OOSR in secondary school than male but female still has noticeably higher unemployment rate.
- No significant difference in Asia, Oceania, and Europe.

04
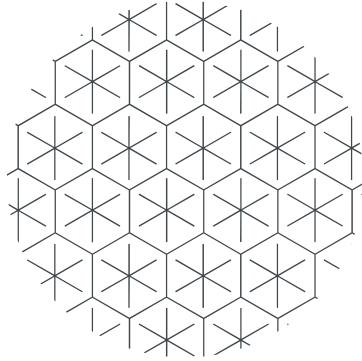
Concerns & Suggestions

Shortcomings

- Loss of countries after merging data
- Calculating OOSR_total using mean of all OOSR_lower/upper_male/female (not accurately reflect the distribution)
- No clear patterns between unemployment and education among countries
- Population data is collected in 2022 but others are collected in 2021
- Only able to compare two variables at a time whereas there are a lot more education variables to compare with unemployment rate.

Future work

- Update the population data to be for 2021 or update others rates to be for 2022.
- Find a variable that accumulate multiple education rates. Use the variable to compare with unemployment rate or to see the differences between male and female in education.
- Deeper data visualization into countries among continents.

Thank you for listening!

Resources:

EDP_Global_Education | Kaggle

The World Factbook by CIA (kaggle.com)

Free Google Slides themes and Powerpoint templates | Slidesgo