

Capstone Project-I Exploratory Data Analysis

Md Khalid Ansari,
Data Science Trainee,
AlmaBetter, Bangalore

25 YEARS GLOBAL EDUCATION ANALYSIS WITH WORLD BANK DATASET



TABLE OF CONTENT

1. Aim of analysis
2. Data Summary and Missing Values
3. Background Intro
4. Project Part1(World vs Year) –
 - a) Individual Indicators
 - b) Comparative Analysis
 - c) PISA, TIMSS AND PIRLS PERFORMANCE REPORT
5. Project Part 2(Income group Country vs Year) –
 - a) Country Classification with Region and GNI per capita
 - b) Individual Indicators
6. Project Part 3 (India vs year vs world) –
7. Conclusion/Insights/Findings

AIM: ANALYSIS OF GLOBAL DEVELOPMENT BASED ON EDUCATIONAL AND OTHER INDICATORS

Project is divided into three parts:

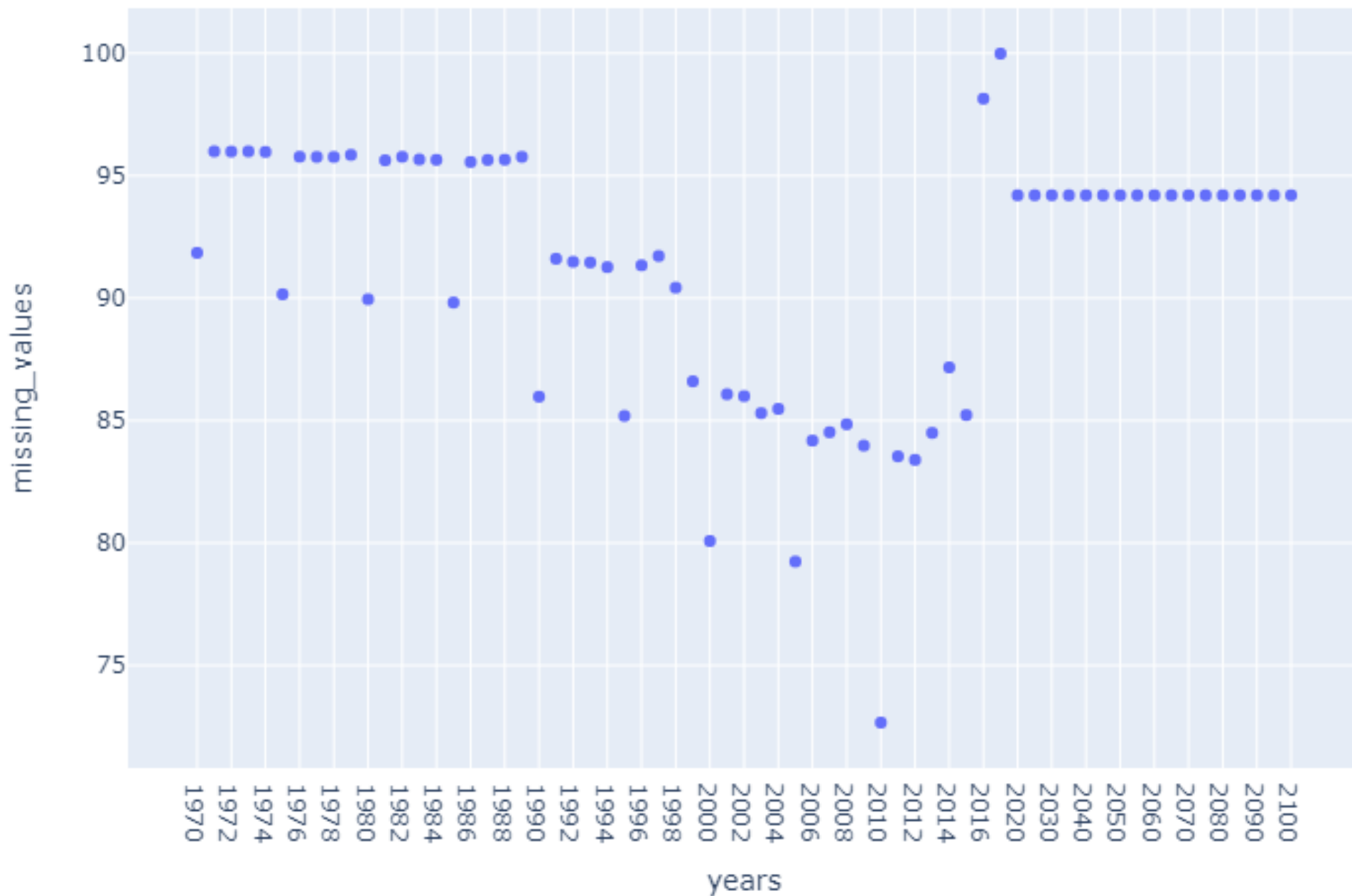
1. **World vs Year:** Global change is analyzed over a period of 25 years (1990-2015).
2. **Income Group vs Year:** Countries from different groups are analyzed for same period of time.
3. **India vs Year vs World:** Overall, India's development is compared with the world for the same period.

DATA SUMMARY

1. **EdStatsCountry** – Contains 'Country Code', 'Short Name', 'Currency Unit', 'Special Notes', 'Region', 'Income Group' and many more.
2. **EdStatsCountry-Series** – Contains 'Country Code', 'Series Code', 'DESCRIPTION'
3. **EdStatsData** – Contains 'Country Name', 'Country Code', 'Indicator Name', 'Indicator Code', '1970', to '2100'.
4. **EdStatsFootNote** – contains 'Country Code', 'Series Code', 'Year', 'DESCRIPTION'
5. **EdStatsSeries** - 'Series Code', 'Topic', 'Indicator Name', 'Short definition', 'Long definition', 'Unit of measure', 'Source' and many more.

Base of our analysis: **EdStatCountry**(32 columns and 241 rows) and **EdStatData**(70 columns and 886,930 rows) – The merge data frame contains: 'Country Name', 'Indicator Name', 'Region', 'Years – 1990 to 2015' values.

**MISSING
VALUES** –
THERE ARE
DIFFERENT
AMOUNT OF
NULL AND
NON-NULL
DATA
AVAILABLE IN
DIFFERENT
YEARS FOR A
PARTICULAR
INDICATOR



HUMAN AND DEVELOPMENT

- ❑ World = Nations = Citizens Development = Health + Education +
- ❑ Educational development = Educational Indicators + Economic Indicators
- ❑ List of Essential Indicators :
 1. Enrolment ratio and Completion rate
 2. Literacy rate
 3. Employment rate
 4. GDP Nominal, or GDP per capita
 5. Govt. expenditure on education as percentage of total expenses
 6. Population growth
 7. Female involvement on education (such as female labor force, female teacher percentage)

1. WORLD VS YEAR – ANALYZING INDIVIDUAL INDICATORS

a) Global GDP per capita:

9,900 USD in 1990 to 14,300 USD in 2015

As per 2010, top 5 countries with highest GDP per capita on an average \$ 1,13,000 per capita:

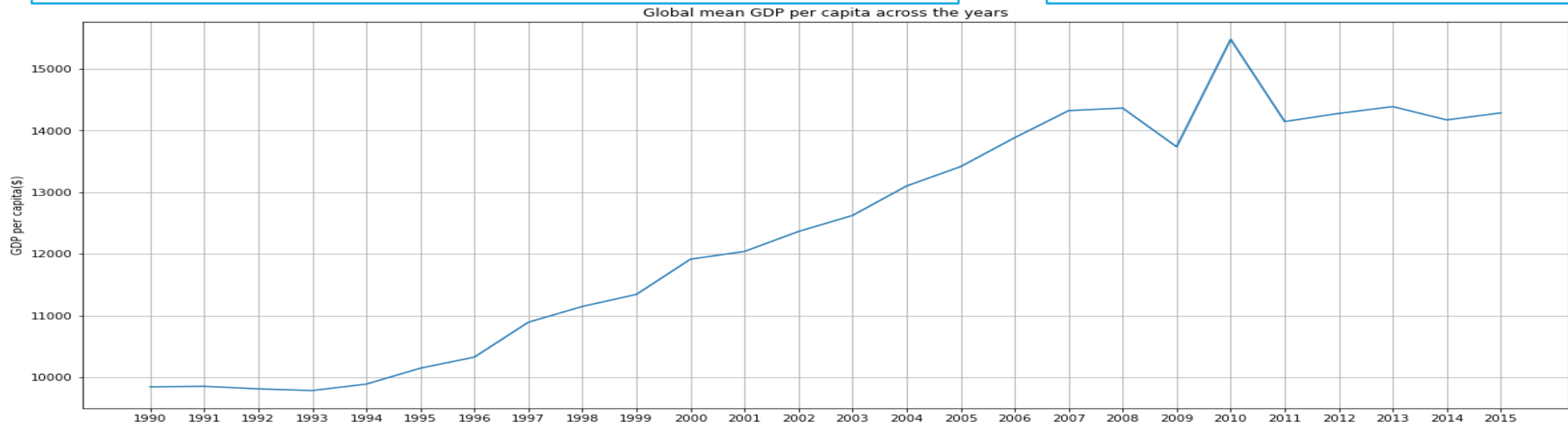
1. Monaco,
2. Liechtenstein,
3. Luxembourg,
4. Bermuda,
5. Norway

b) GDP at Market Prices:

Global nominal GDP is increased by 100% in 25 years (\$37.26 Trillion to \$74.4 Trillion)

As per 2010, top 5 countries with highest nominal GDP on average of \$ 6.5 Trillion :

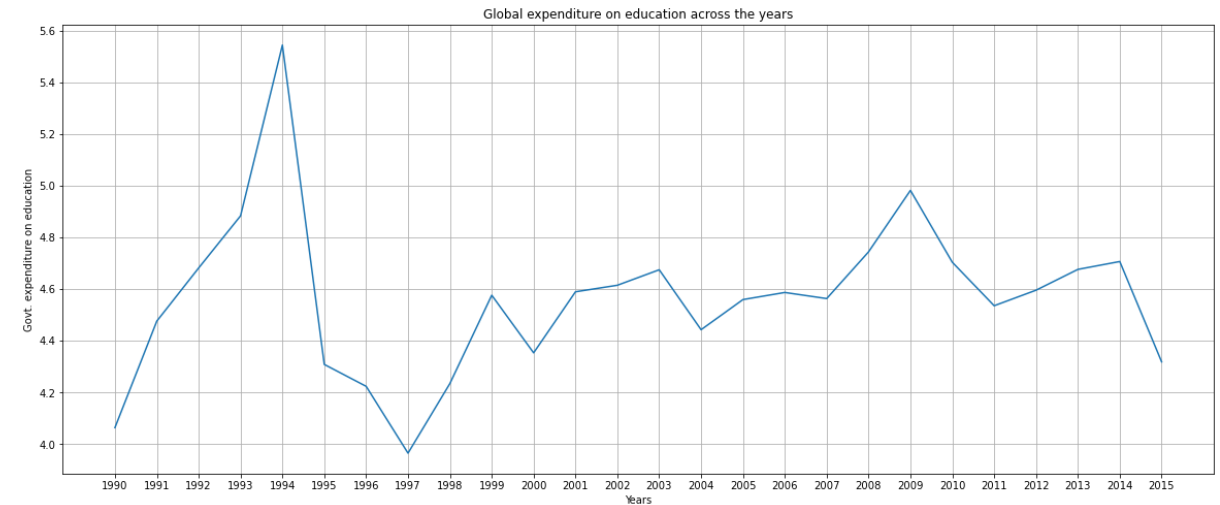
1. United States,
2. China,
3. Japan,
4. Germany,
5. France



c) Expenditure on Education (4% to 4.3%)

The global change in expenditure on education is overall 0.25 in last 25 years.

Countries with highest expenditure on education: 'Cuba', 'Timor-Leste', 'Solomon Islands', 'Sao Tome and Principe', 'Moldova', 'Denmark', 'Namibia', 'Bolivia', 'Iceland', 'New Zealand'

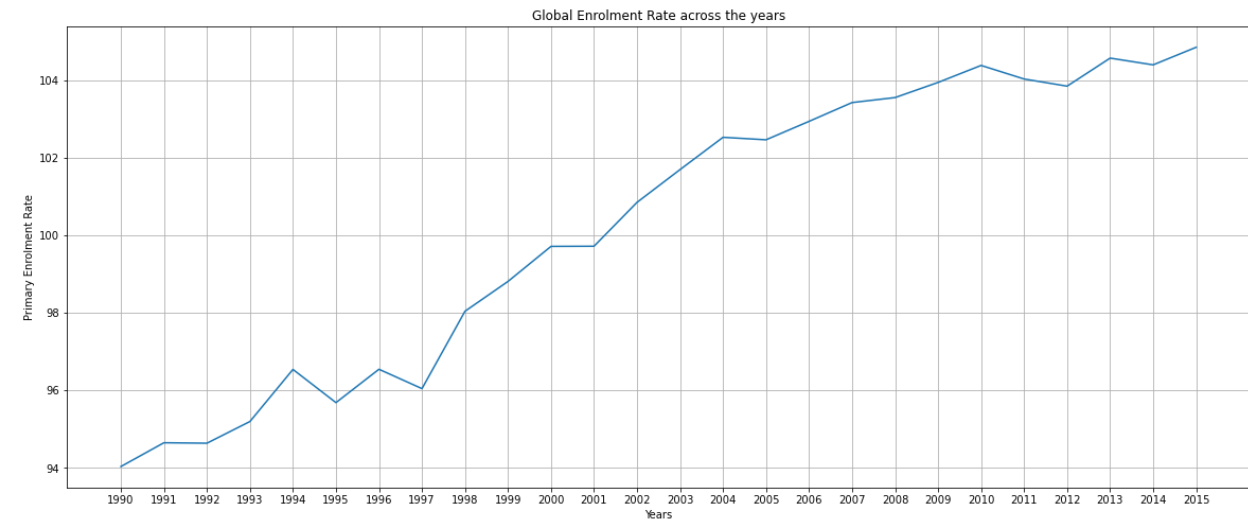


d) Enrolment Ratio –

Primary Education has the highest enrolment ratio(94% to 106%) and Tertiary Education(18% to 48%) has lowest enrolment ratio in 25 years. And, Secondary education has (65% to 91%), Upper Secondary(63% to 98%)

Countries with highest enrolment in primary education in year 2015 are:

Madagascar, Rwanda, Nepal, Malawi, Burundi

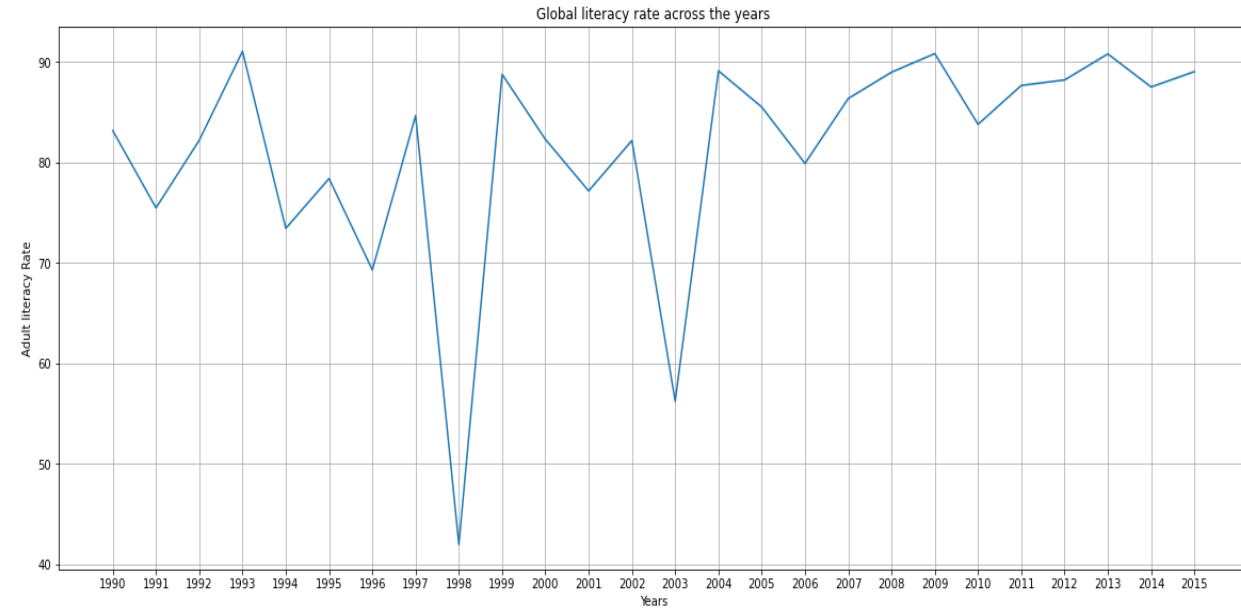


e) Literacy Rate –

Elderly Literacy rate(69%) is the lowest of all. Then comes Adult literacy rate(89%) and finally at highest is the Youth Literacy rate(95%).

Countries with highest adult literacy rate:

'Kazakhstan', 'Azerbaijan', 'Russian Federation',
'Mongolia', 'Uruguay',

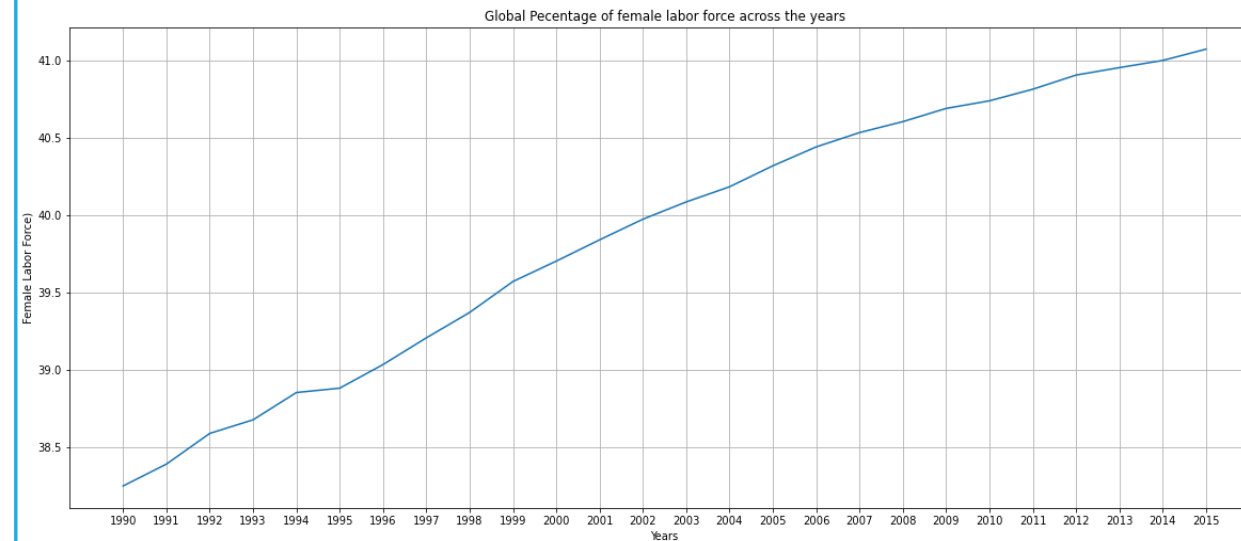


f) Female Labor Force – (38.3 to 41.5%):

Female labor percentage had 3.2% growth in last 25 years globally. Countries with highest female labor force:
'Mozambique', 'Burundi', 'Rwanda', 'Lithuania', 'Tanzania'

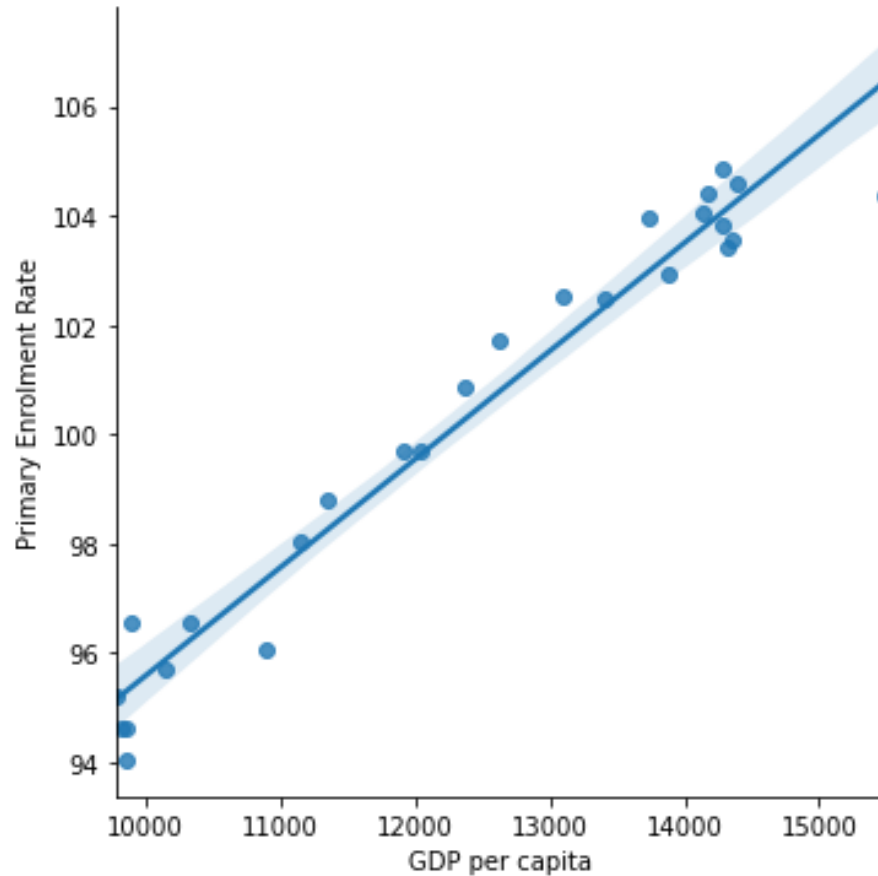
Countries with lowest female labor force:

'Saudi Arabia', 'Afghanistan', 'Yemen, Rep.', 'Qatar', 'United Arab Emirates'

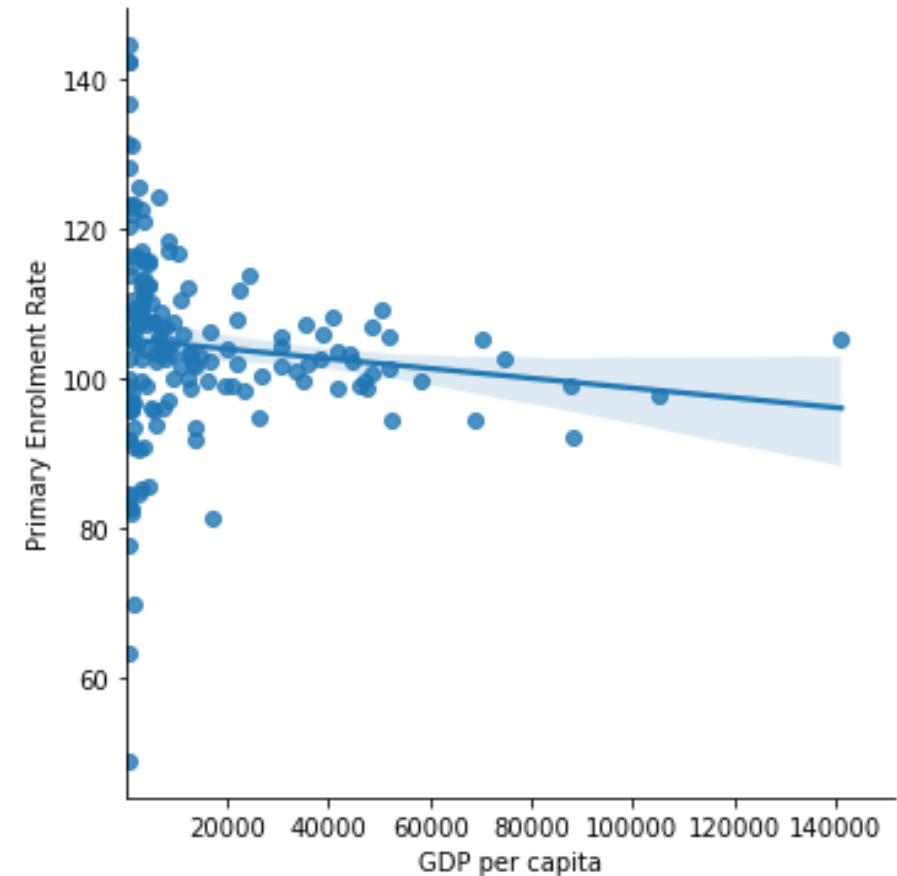


Enrolment Ratio vs GDP per capita – Linear Regression Model

Primary enrolment ratio changed from less than 94% to more than 104% in the 25 years period with \$4500 overall change in Global GDP per capita.



25 year variation globally

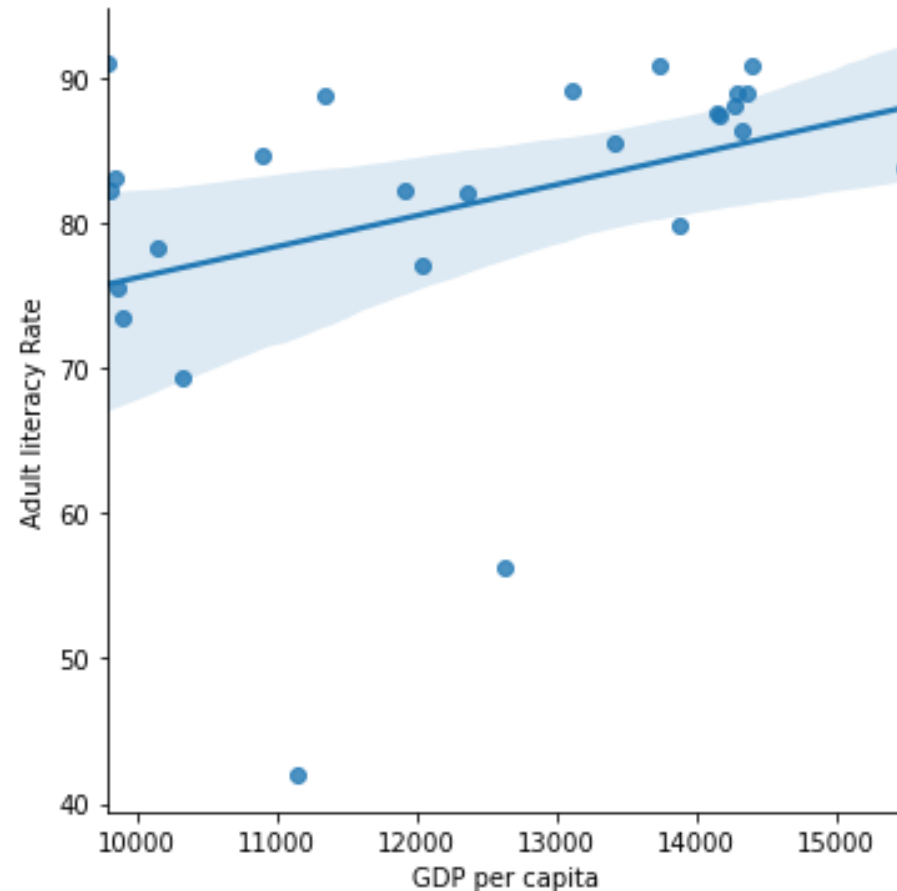


Country wise variation in 2010

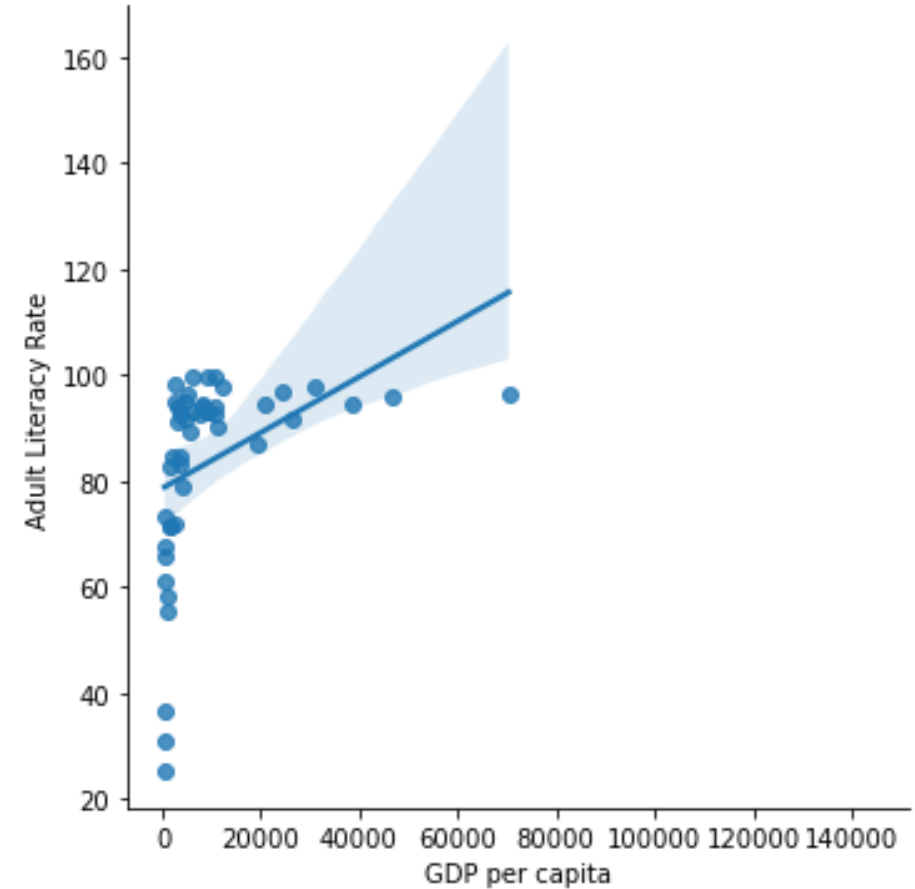
Adult Literacy rate vs GDP per capita:

Adult literacy rate has increased by just 6% with 45% increase in GDP per capita in the 25 years.

Minimum adult literacy is found to 70% globally.



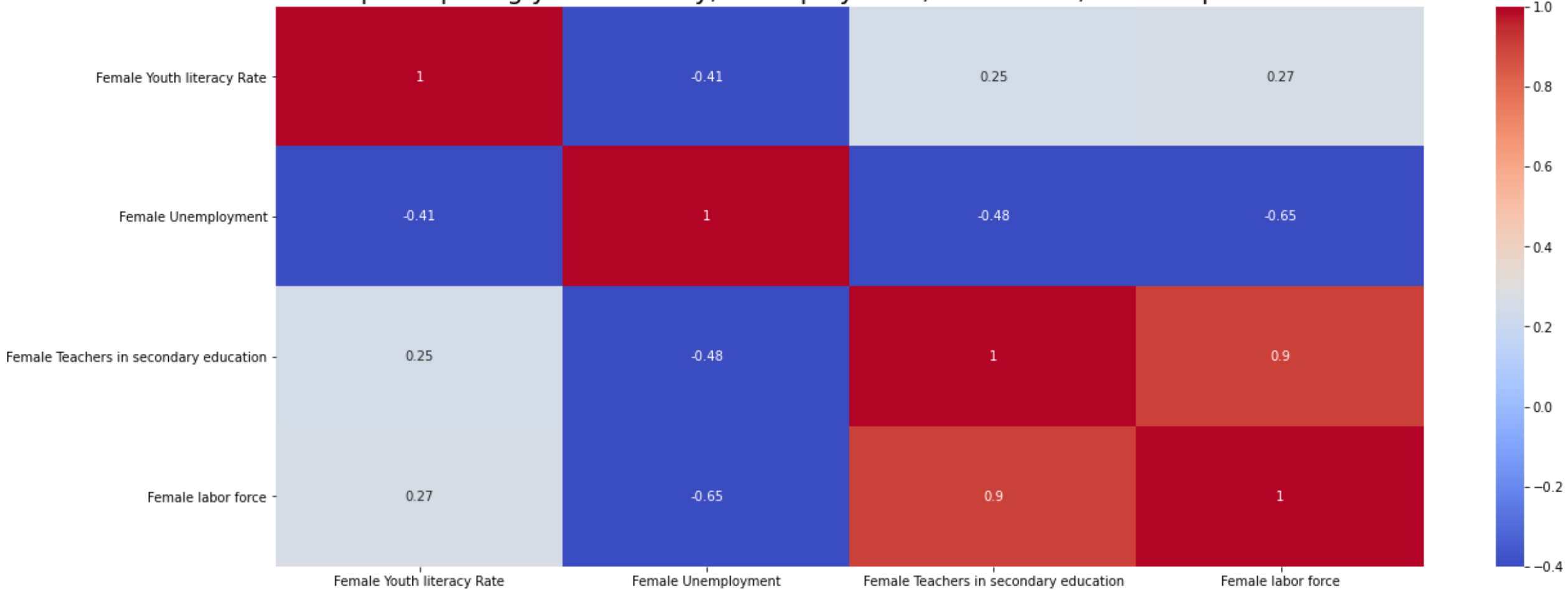
25 year variation globally



Country wise variation in 2010

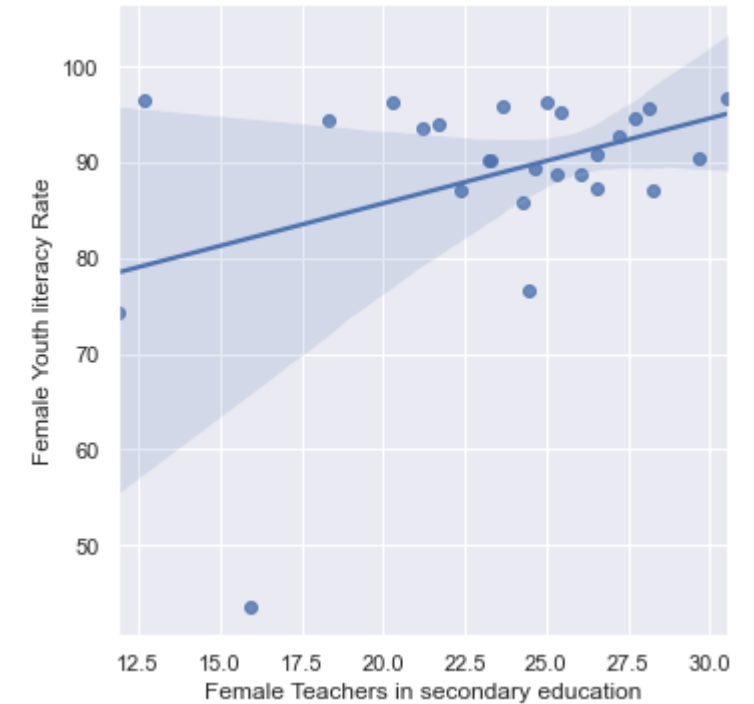
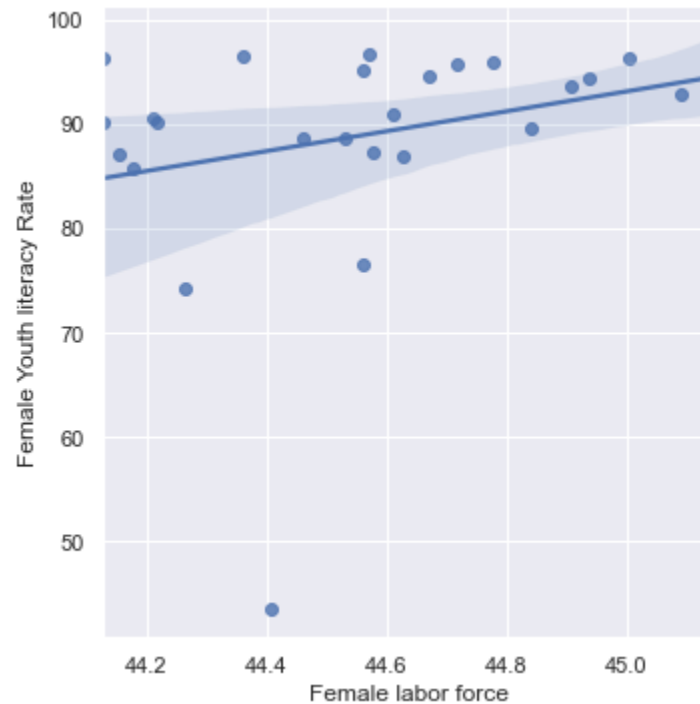
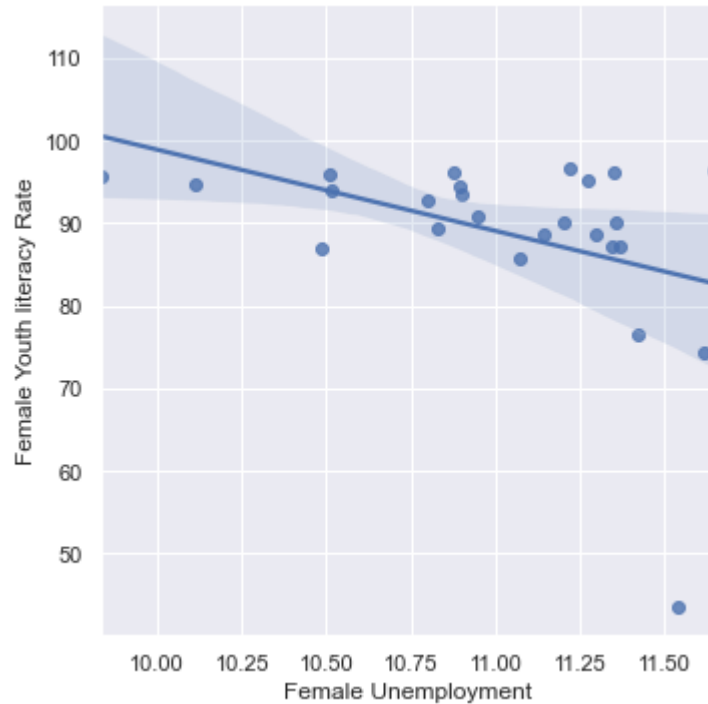
Female involvement in education and employment

Heat Map comparing youth literacy, unemployment, labor force, teacher percent in womens



Female involvement vs unemployment vs literacy:

With 3.4% increase in youth female literacy, increase in female labour force is 3.2% and increase in female teacher percentage is 12%. While overall female employment is just increased by 1.3%.

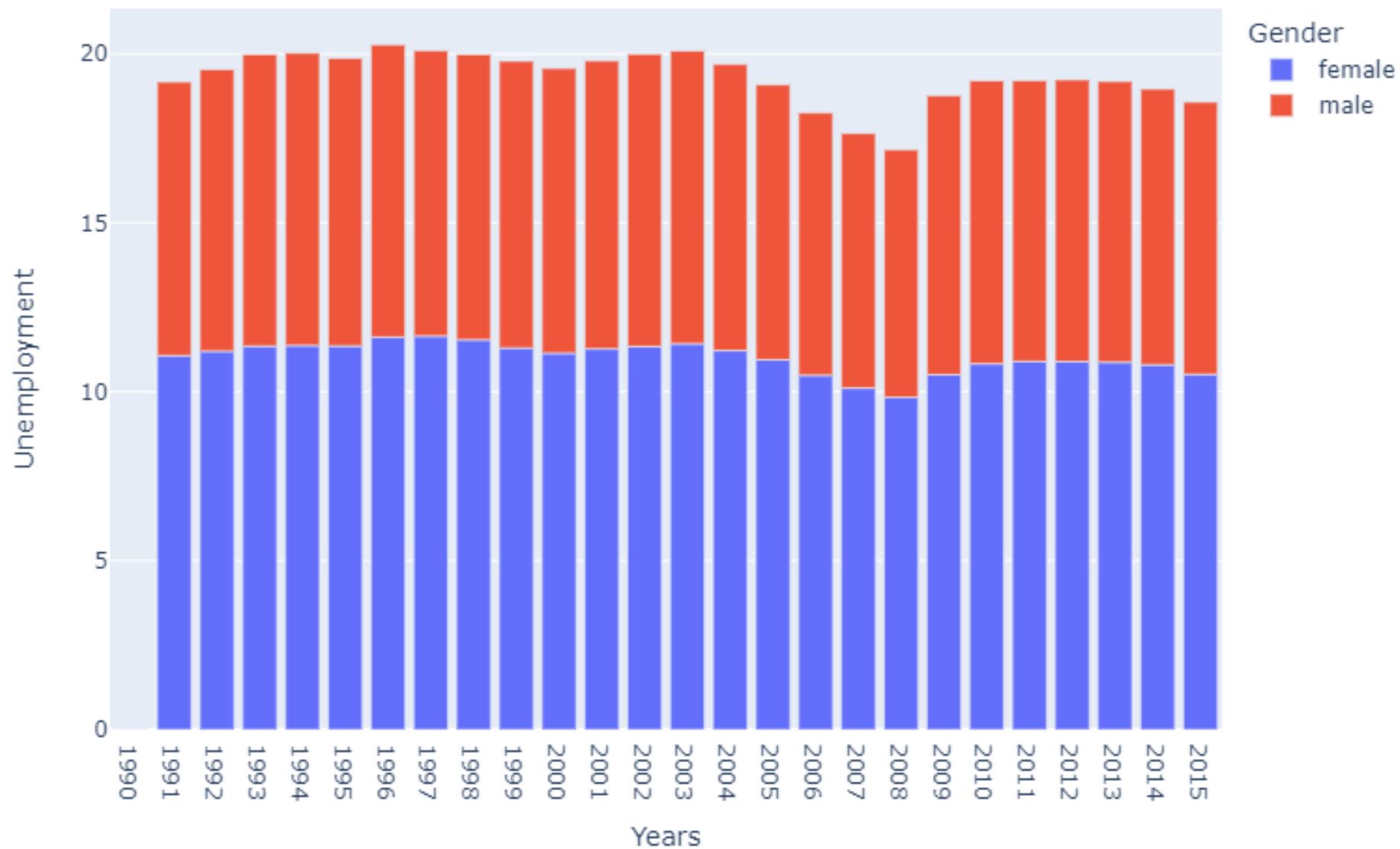


25 year variation globally

GLOBAL YOUTH LITERACY RATE – MALE VS FEMALE



GLOBAL UNEMPLOYMENT RATE — MALE VS FEMALE





Hey! It's not
the Leaning
Tower of Pisa

PISA, TIMSS AND PIRLS PERFORMANCE REPORT

PISA – Program for International Student Assessment

- According to Google, PISA is a worldwide study by OECD in member and non-member nations to evaluate educational system by measuring 15+ year-old school pupils' scholastic performance on mathematics, science and reading.

TIMSS – Trends in International Mathematics and Science Study

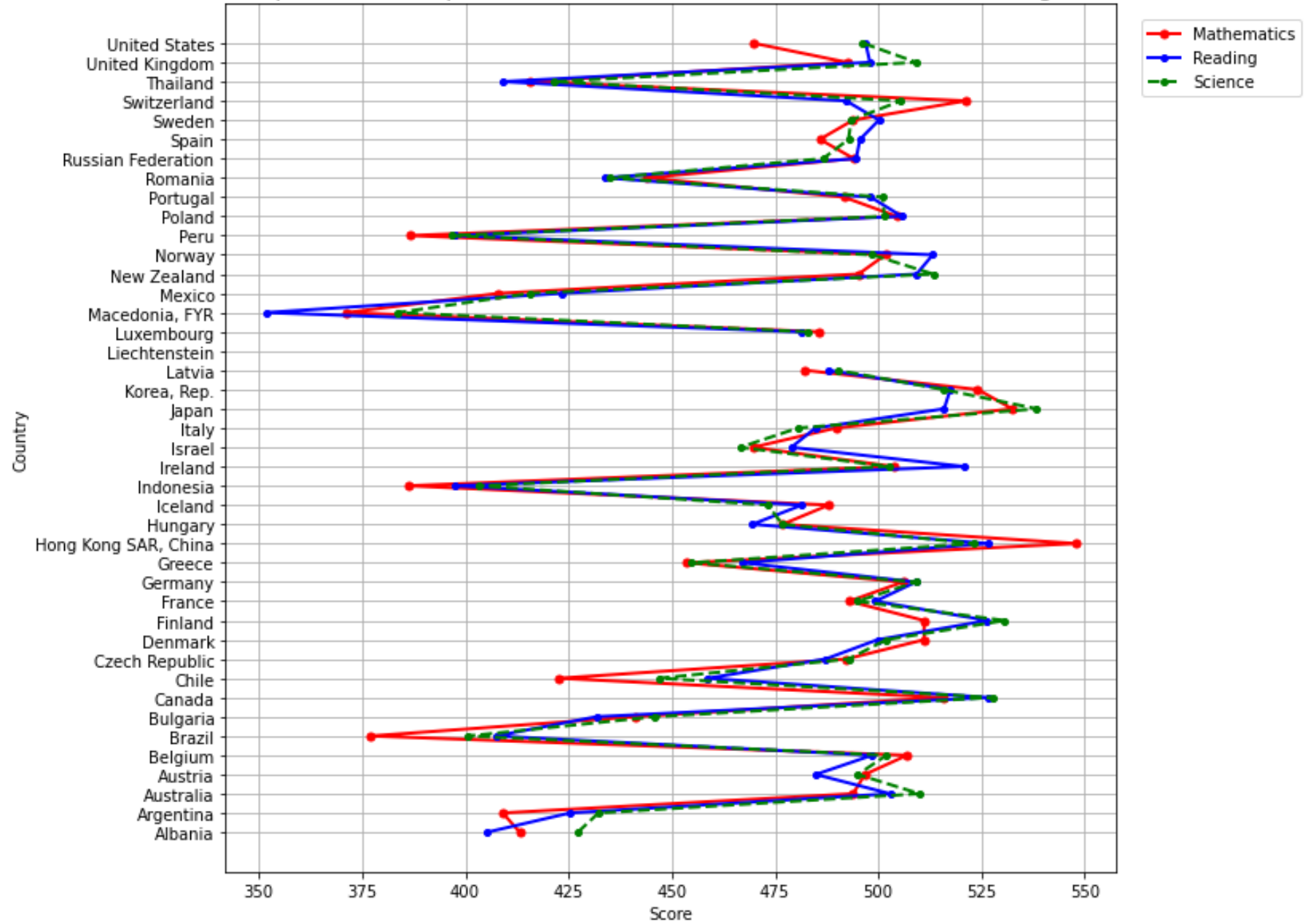
- According to Google, TIMSS provides reliable and timely trend data on the mathematics and science achievement of U.S. students compared to that of students in other countries.

PIRLS – Progress in International Reading Literacy Study

- According to Google, the IEA's PIRLS is an international study of reading achievement in fourth grades.

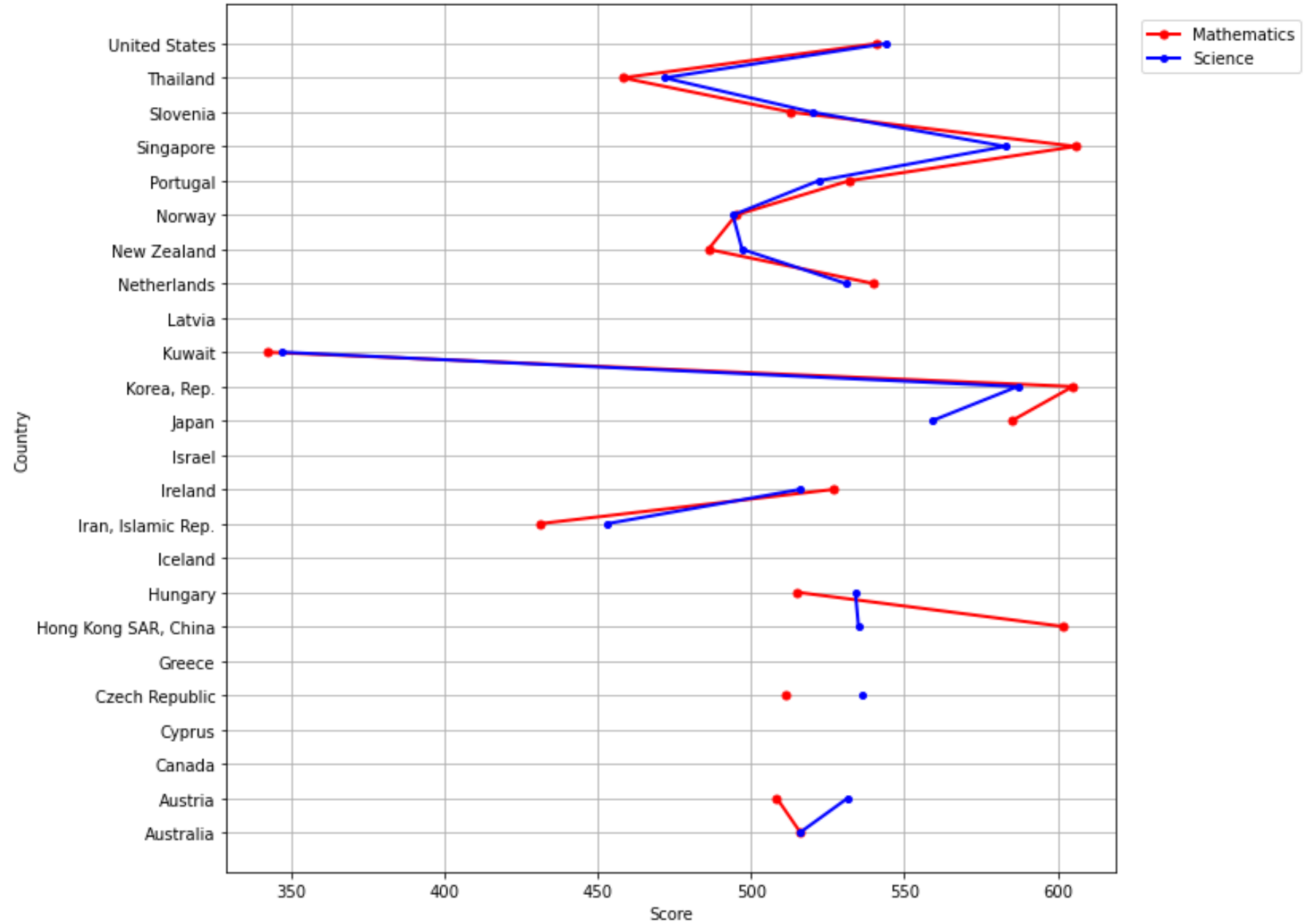
PISA – Program for International Student Assessment

PISA report on mean performance in mathematics, science and reading for 15+



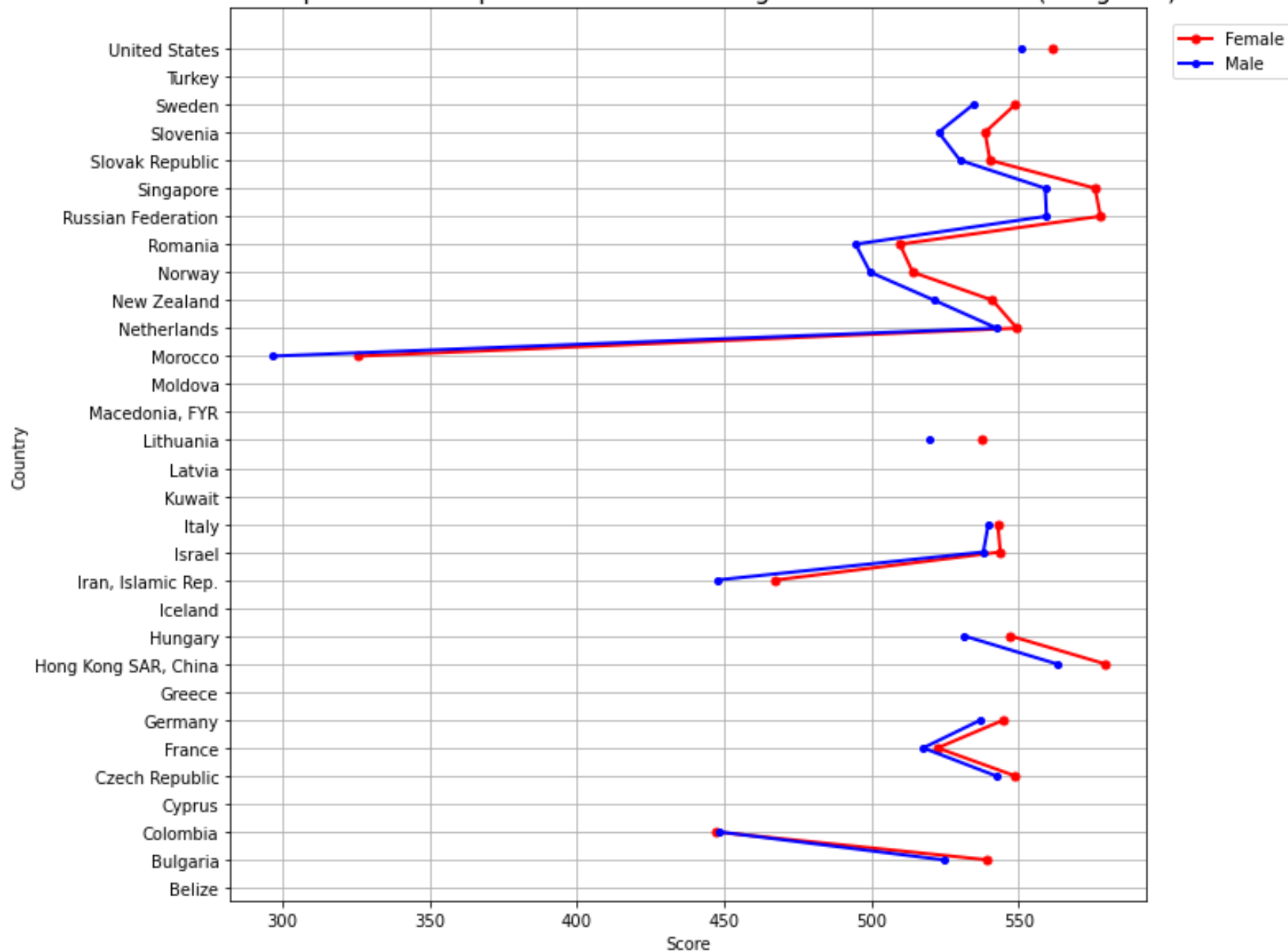
TIMSS mean performance report on Mathematics and science for 4th grade students

**TIMSS – Trends in
International
Mathematics and
Science Study**



PIRLS – Progress in International Reading Literacy Study

PIRLS report on mean performance in reading for female and male (4th grade)



OBSERVATIONS AND FINDINGS

PISA Report :- [Mathematics, Science, and Reading][15+]

Top 5 scoring countries(Above 500) = China(550), Japan, Finland, Canada, Switzerland

Below 5 scoring countries (Below 400) = Macedonia(350:lowest), Brazil, Indonesia, Peru, Albania

TIMSS :- [Mathematics, Science][4th Grade]

Top 5 scoring countries(Above 550) = China, Singapore(610), Korea, Netherland, USA

Below 5 scoring countries(Below 500) = Iran, Thailand, Kuwait(348: Lowest), New Zealand, Norway

PIRLS :- [Reading][4th Grade]

Top 6 scoring countries(Above 550) :- China(580), Singapore, Russia(580), Sweden, USA, Netherland, Czech Republic

Below 3 scoring countries(Below 450) :- Morocco(299:Lowest), Iran, Colombia

Male Vs Female Reading:

PIRLS :- [Reading][4th Grade]

1. Female has the higher score in reading compared to male in almost all countries.
2. Female student from China has the highest score 580.

Region and Country (Top Scorer)

1. East Asia & Pacific – China, Korea, Singapore,
2. North America – Canada, USA
3. Europe and Central Asia – Finland, Switzerland, Netherlands, Russia, Sweden, Czech Republic

2. INCOME GROUP VS YEAR - ANALYZING COUNTRIES BASED ON GNI

Country type with Gross National Income per capita(2015):

➤ **Low-income countries** – under 850 USD

e.g. - Afghanistan, Bangladesh, Nepal (South Asia)

➤ **Lower-middle income countries** – under 2,250 USD

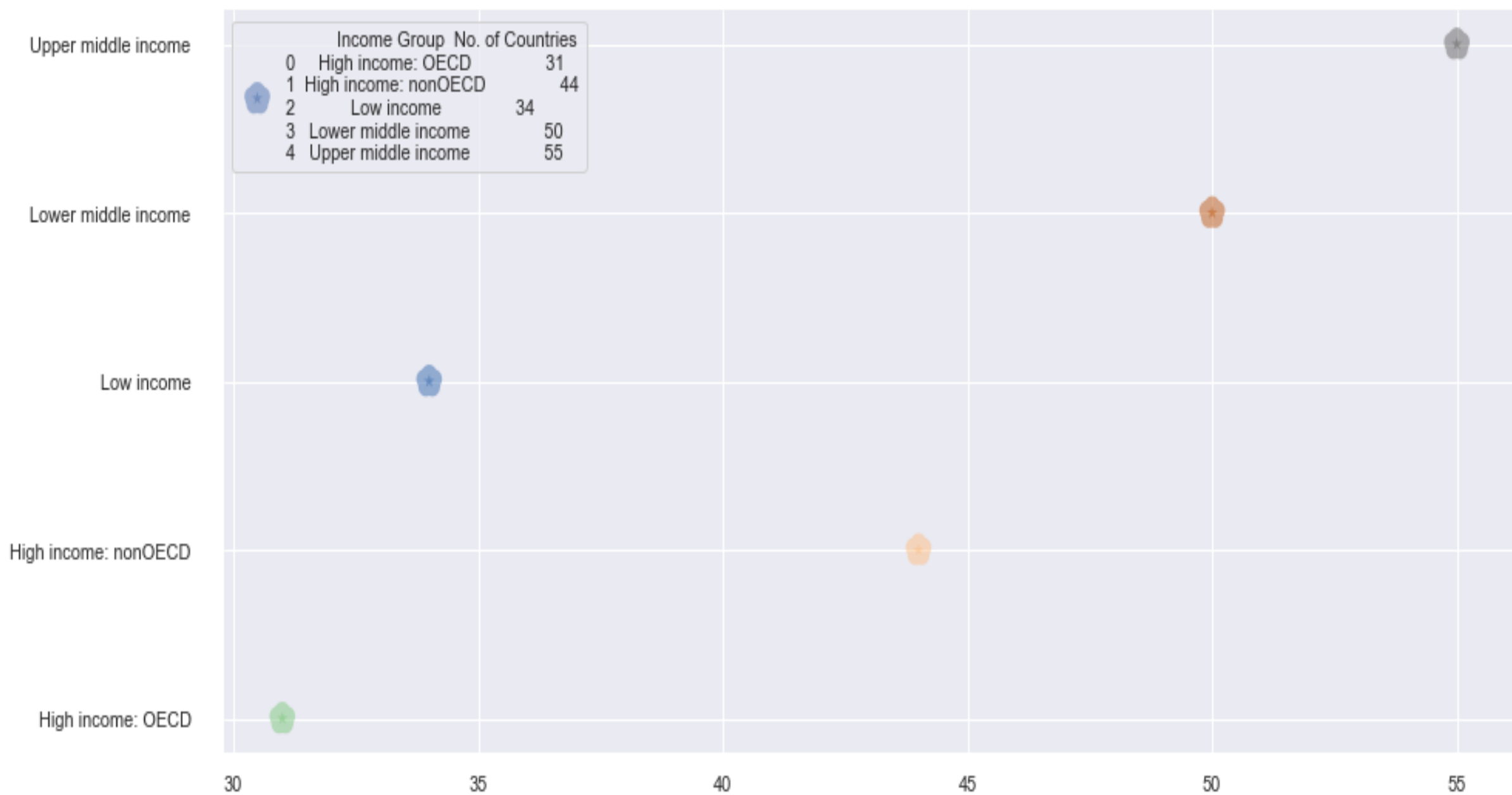
e.g. - Bhutan, India, Pakistan, Sri Lanka (South Asia)

➤ **Upper-middle income countries** – under 8,000 USD

e.g. - Albania, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Hungary, Kazakhstan, Macedonia, FYR, Montenegro, Romania, Serbia, Turkey, Turkmenistan (Europe & Central Asia)

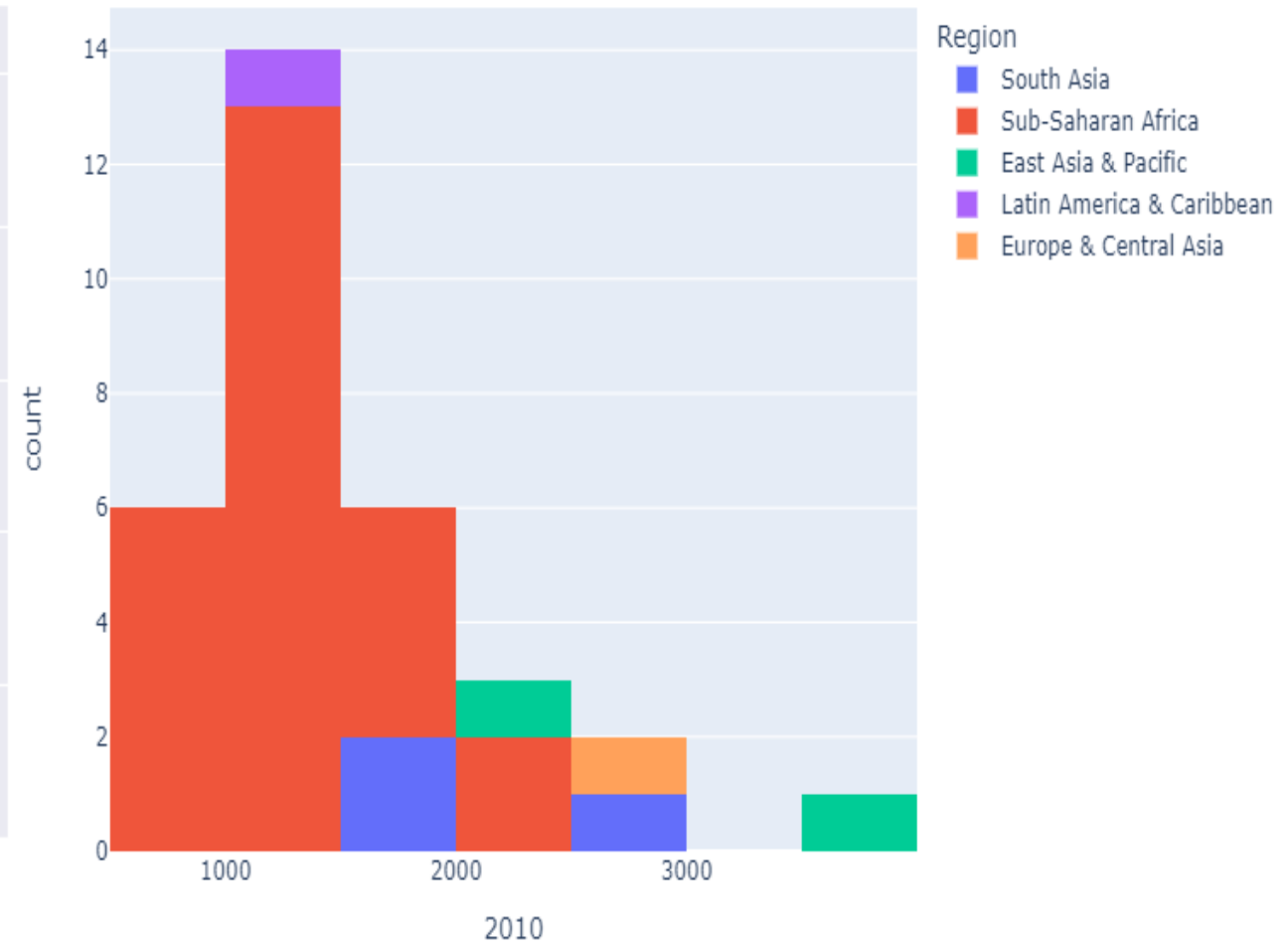
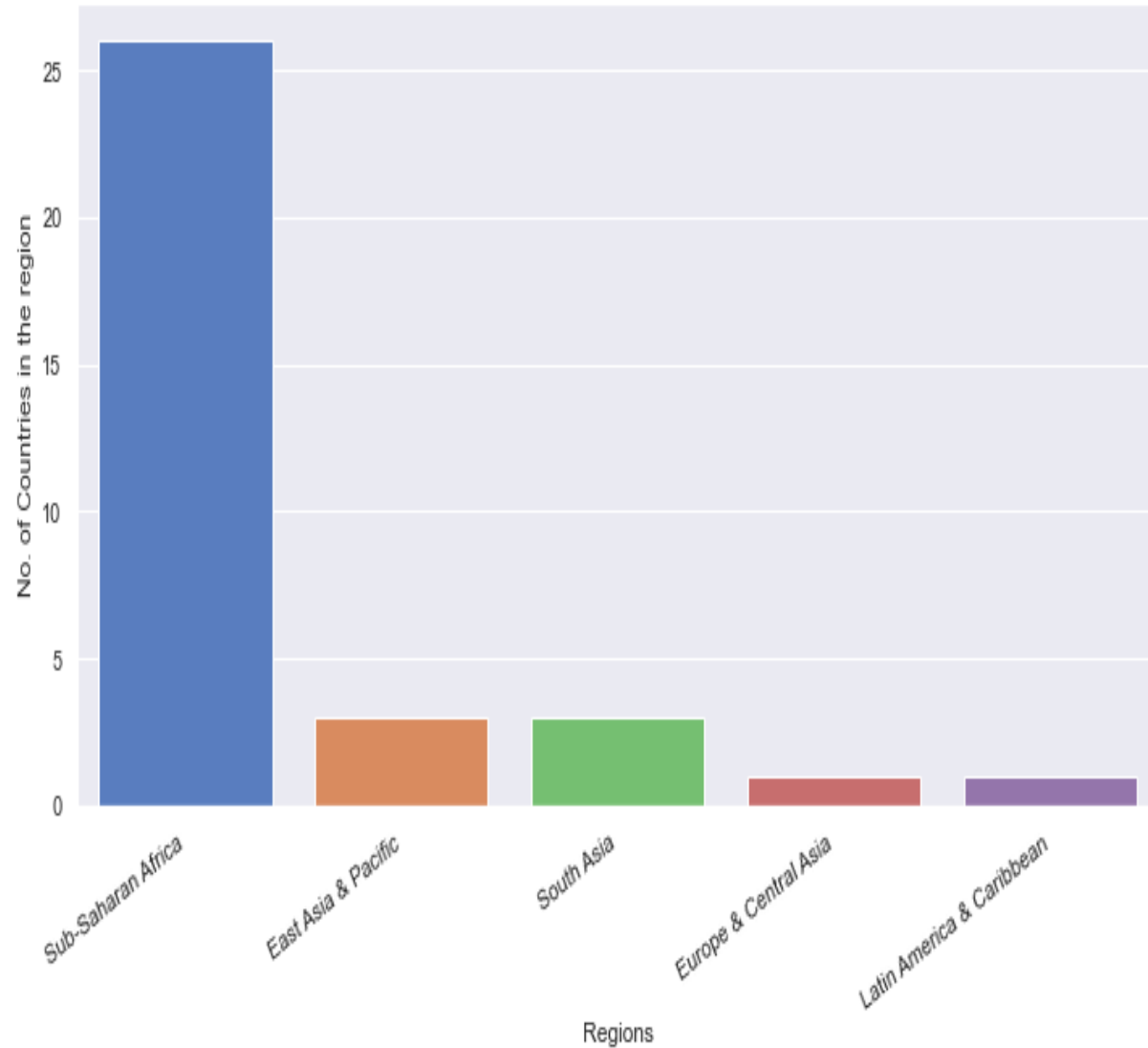
➤ **High-income countries** – under 51,000 USD

e.g. - Canada, United States (North America)

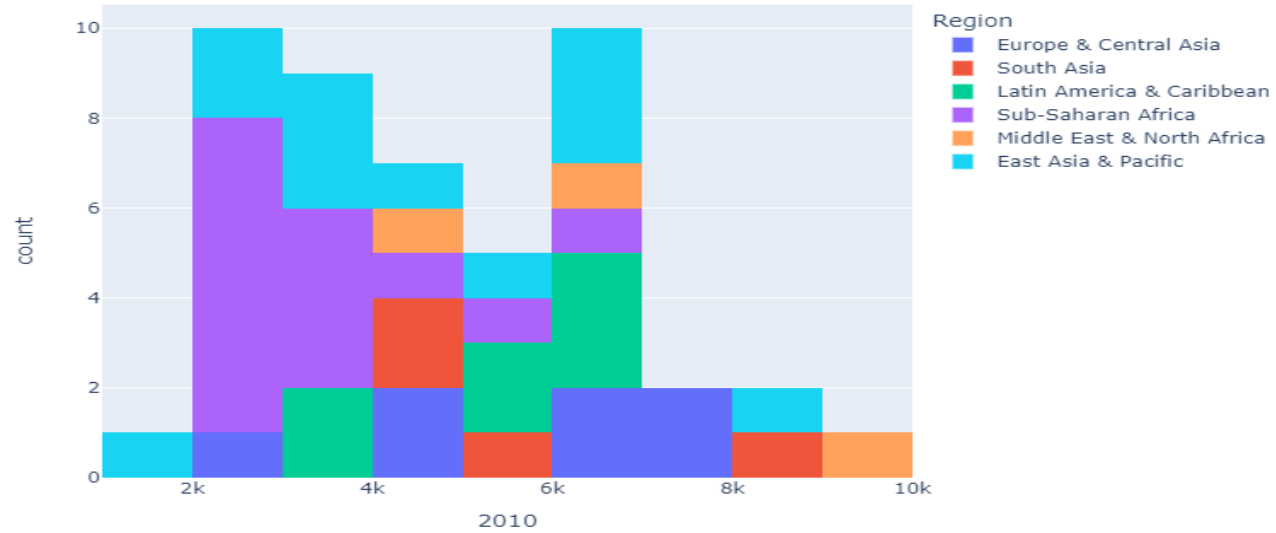
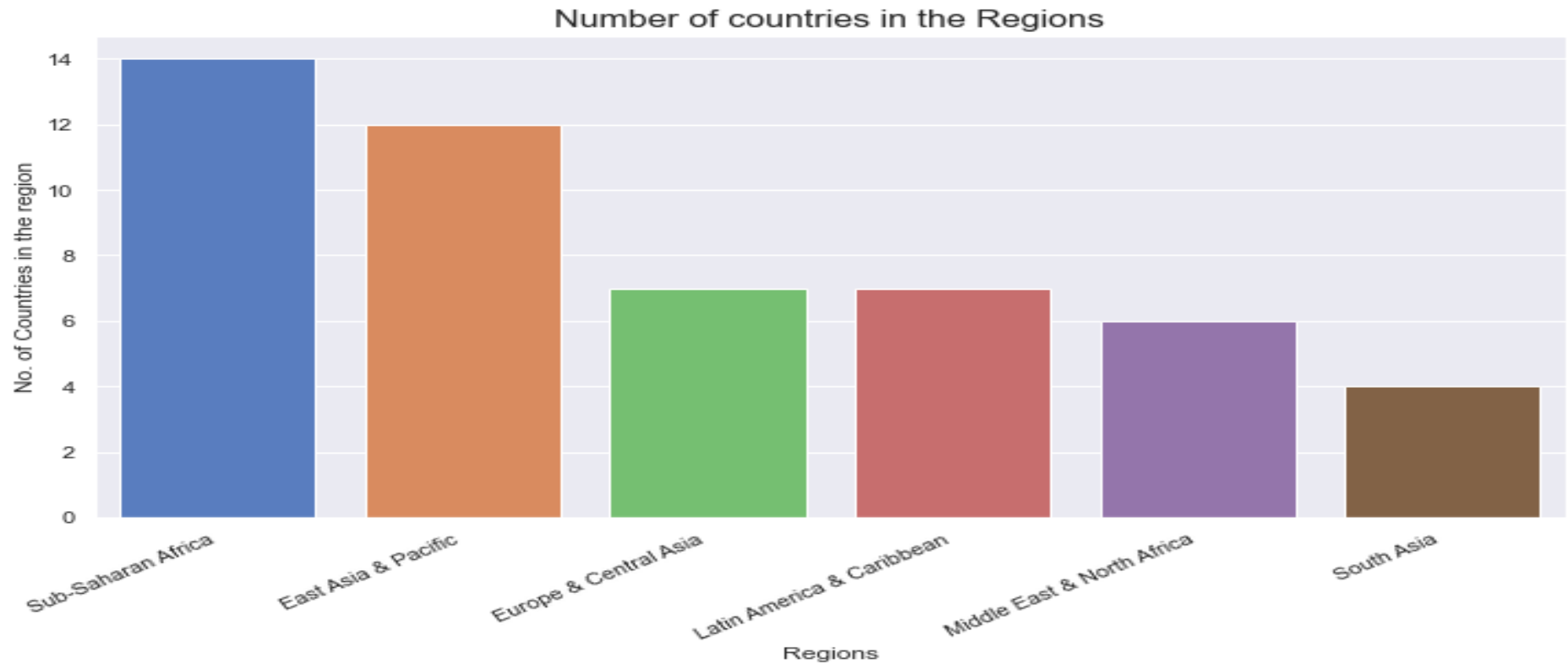


COUNTRY TYPE1: LOW INCOME GROUP

Number of countries in the Regions

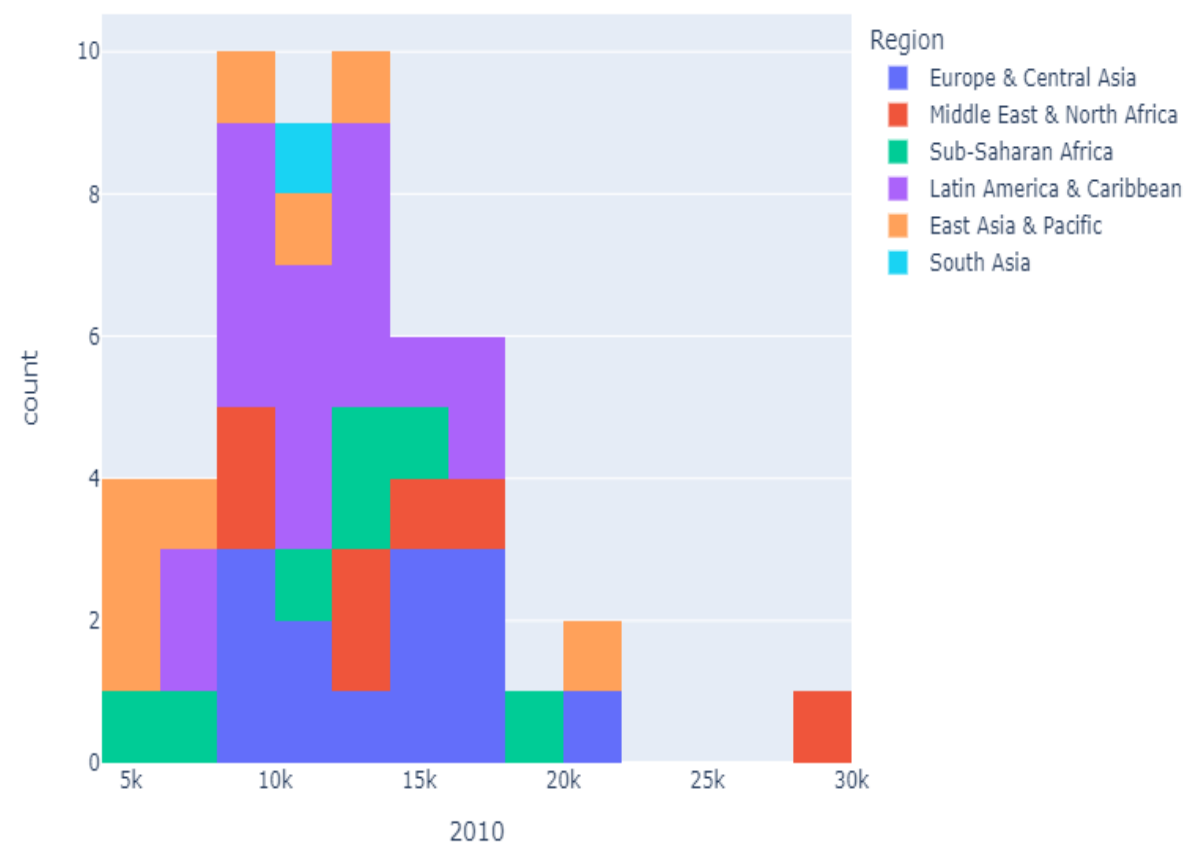
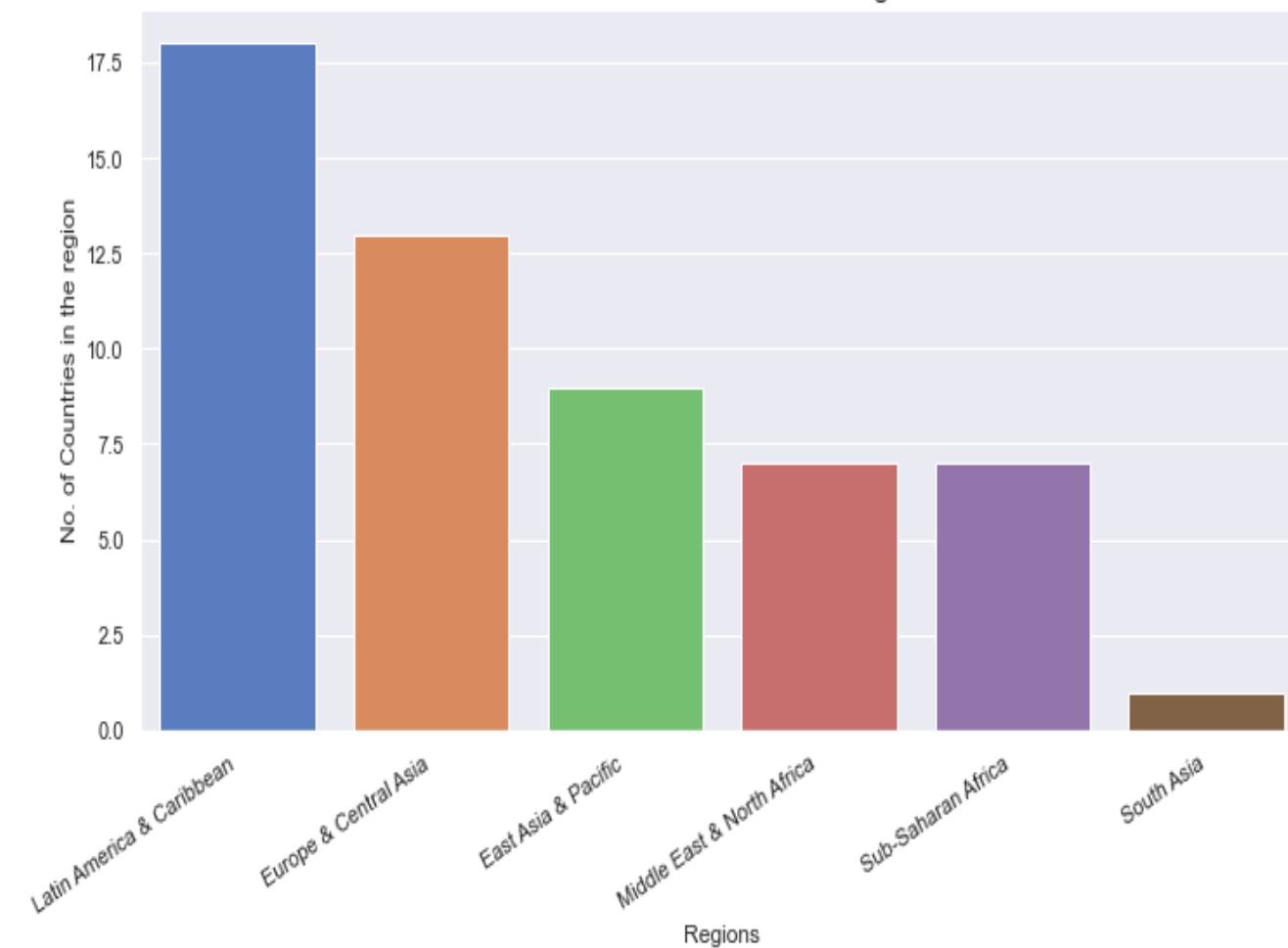


COUNTRY
TYPE2:
LOWER
MIDDLE
INCOME
GROUP

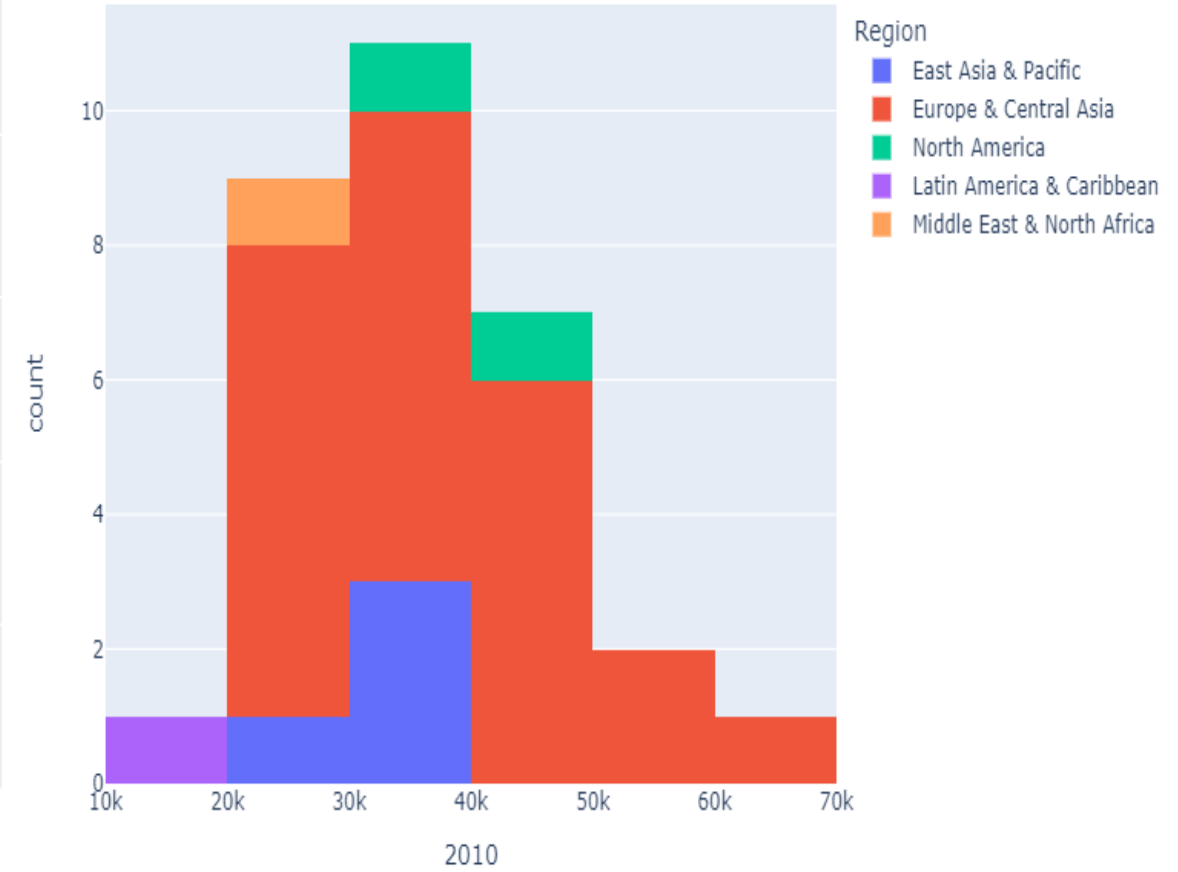
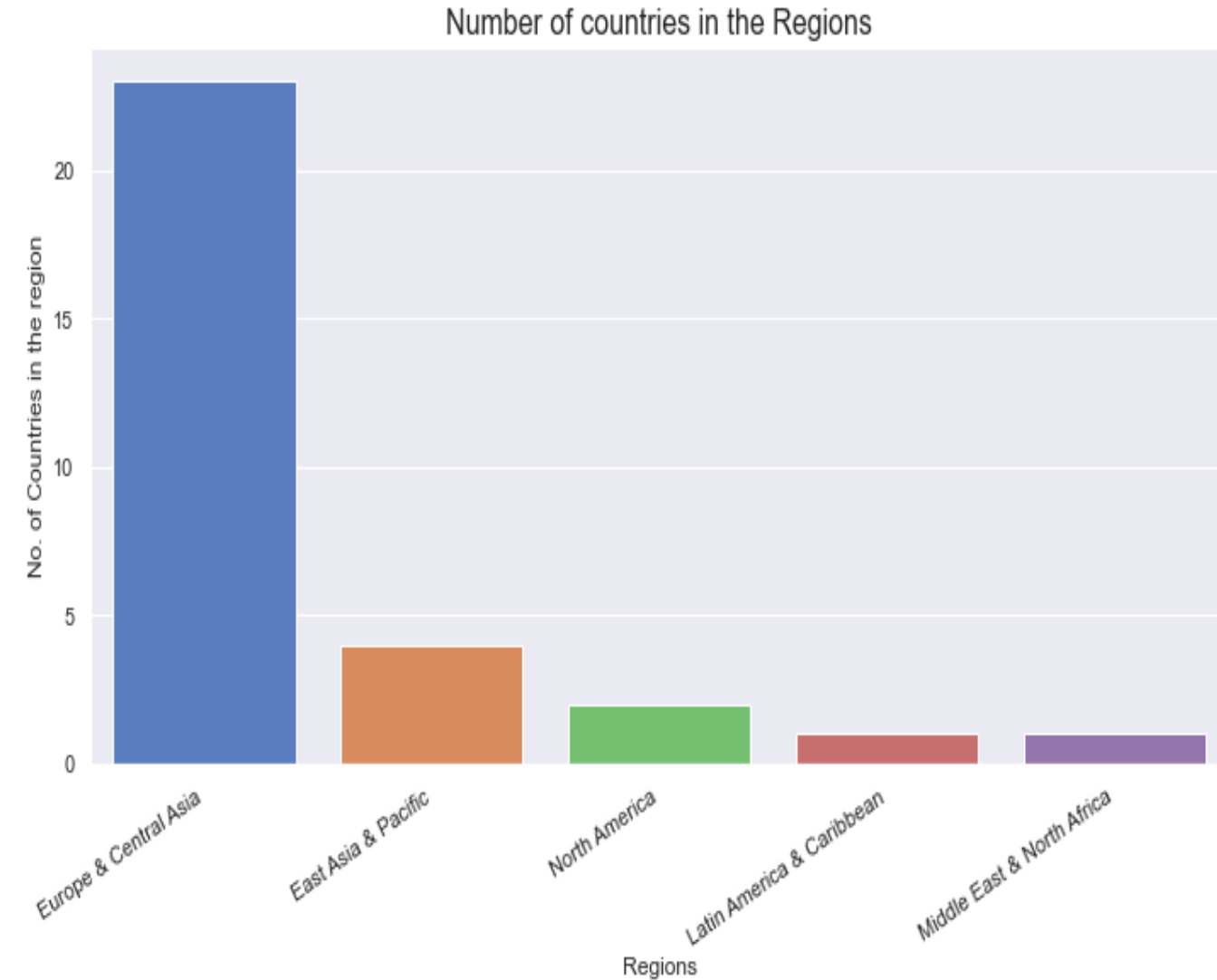


COUNTRY TYPE3: UPPER MIDDLE INCOME GROUP

Number of countries in the Regions



COUNTRY TYPE4: HIGH INCOME GROUP



GROSS ENROLMENT RATIO

❖ Low Income Group (75% to 123%)

- Most of the countries in top 10 belongs to Sub-Saharan Africa region with Madagascar to be in the 1st position.
- India's neighboring country
Nepal - 3rd
Afghanistan - 12th and
Bangladesh - 13th position.
India is not a low income country.

❖ Lower Middle Income Group(87% to 101%)

- Most of the countries in top belongs to East Asia & Pacific region with Timor-Leste to be in the 1st position.
- India - 20th position
Bhutan - 19th
Sri Lanka - 29th
Pakistan - 33rd

❖ Upper Middle Income Group (99% to 101%)

- Top 5 countries belongs to Latin America & Caribbean region with Colombia to be in the 1st position.
- Our nearest neighbour China comes 9th in enrolment ratio which is the only in the East Asia and Pacific to be in the top 10 list.

❖ High-Income Group(104.5% to 100%)

- Top 5 countries belongs to Europe and central Asia region with Portugal to be in the 1st position.
- Our nearest neighbour
Australia - 7th, the only country in the East Asia and Pacific to be in the top 10 list while
United states - 21st
Canada - 27th position.

UNEMPLOYMENT RATE

❖ Low Income Group(2% to 5.3%)

- Bangladesh - 19th
Nepal - 29th
Afghanistan - 7th
While Benin, Myanmar and Cambodia has highest employment rate.
- Low income countries seen 48% increase in enrolment ratio while having 3.3% decrease in employment rate. Completion rate: 27% 68%

❖ Lower Middle Income Group(7% to 4%)

- Solomon Islands, Lesotho and Swaziland are countries to employ the lowest number of labour force. While Vietnam, Papua new guinea has highest employment rate.
India – 41th position with average unemployment rate 4%.
- Lower middle income countries seen 14% increase in enrolment ratio while having 3% increase in employment rate. Completion rate: 97 to 86%.

❖ Upper Middle Income Group(12% to 13%)

- Belarus, Maldives, and Thailand have the highest employment rate. Maldives is the nearest neighbour to India with 95% employment.
China has employment rate of 95.5% which is almost equal to that of India.
- 2% increase in enrolment ratio while having 1% decrease in employment rate. Completion rate: 92 to 100%

❖ High-Income Group(8.5% to 6%)

- Switzerland, Netherlands, Luxembourg, Korea, Rep., Norway are in the top with average employment rate 96%.
Australia is the nearest neighbour to India with 95% employment.
USA - 7th
Canada - 15th position with average of 92% employment. Countries with lowest employment rate of 85%:
 - Spain, Estonia, Slovak Republic, Ireland, Greece.
- High income countries seen 4.5% decrease in

FEMALE INVOLVEMENT

❖ Low Income Group(27.5% to 32.5%)

- Mozambique - women involvement > half of the total labour force. While Bangladesh, Somalia and Afghanistan has average of just 20% women involvement in total labour force.
- Female labor percentage is increased by 5% while female teacher percentage seen a growth of 12.5%.

❖ Lower Middle Income Group(28.9% to 30.1%)

- Lao PDR, Ghana, Congo, Rep. are in the top 3 list of highest female labour force.
- India - 40th position with female labour force of average 25% involvement in total labour force while it is just 8% in Pakistan.
- Female labor percentage is increased by 1.2% while female teacher percentage seen a growth of 10%.

❖ Upper Middle Income Group(43% to 43.5%)

- In terms of female involvement, China -13th position with 43% female involvement . Lebanon, Iraq, Iran, Islamic Rep., Jordan, Algeria has the lowest female labour percentage that is 9% to 12%.
- Female labor percentage is increased by 0.5% while female teacher percentage seen a growth of 7%.

❖ High-Income Group(44.2% to 46.4%)

- In terms of female involvement, Canada - 7th USA - 12th position with 46% average female involvement.
Greece, Japan, Korea, Rep., Italy, Chile has the lowest female labor percentage in the list but with average of 40%.
- Female labor percentage is increased by 2.2% while female teacher percentage seen a growth of 12.5%.

3. India and World



ENROLMENT RATIO

➤ During 90s,

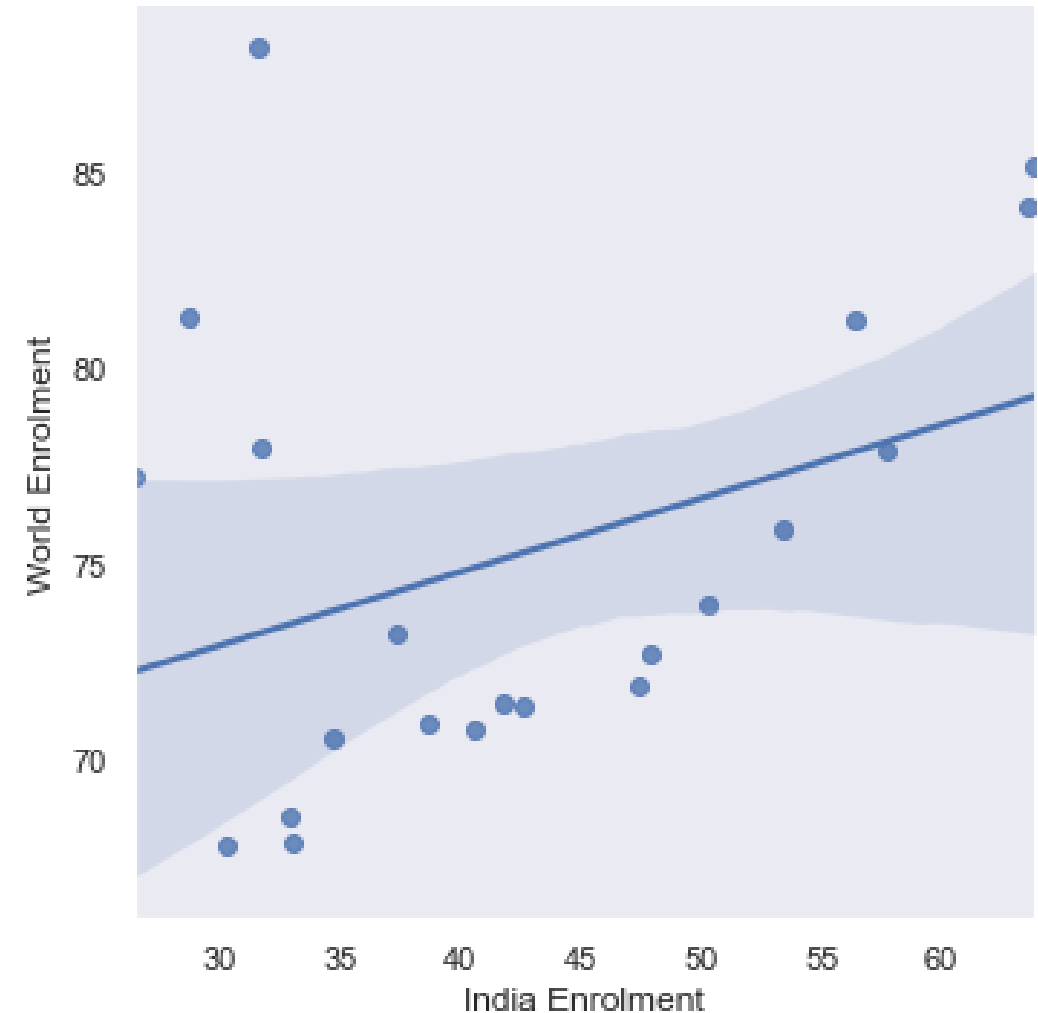
India's Upper secondary enrolment ratio – 30%
Global ratio – 93%.

In last 25 years, India's enrolment ratio was increased by 33%. On the other hand, the overall enrolment is decreased by 9% globally.

➤ According to the 2010 global census, India was in 109th rank with 50.26% enrolment.

Top 5 countries in the list are: Australia, Ireland, New Zealand, Spain, Norway.

➤ Overall, India's enrolment has grown fast as compared to the world but still has very low enrolment as compared to many developed countries.



EMPLOYMENT RATE

➤ During 90s,

India's employment rate 95.9%

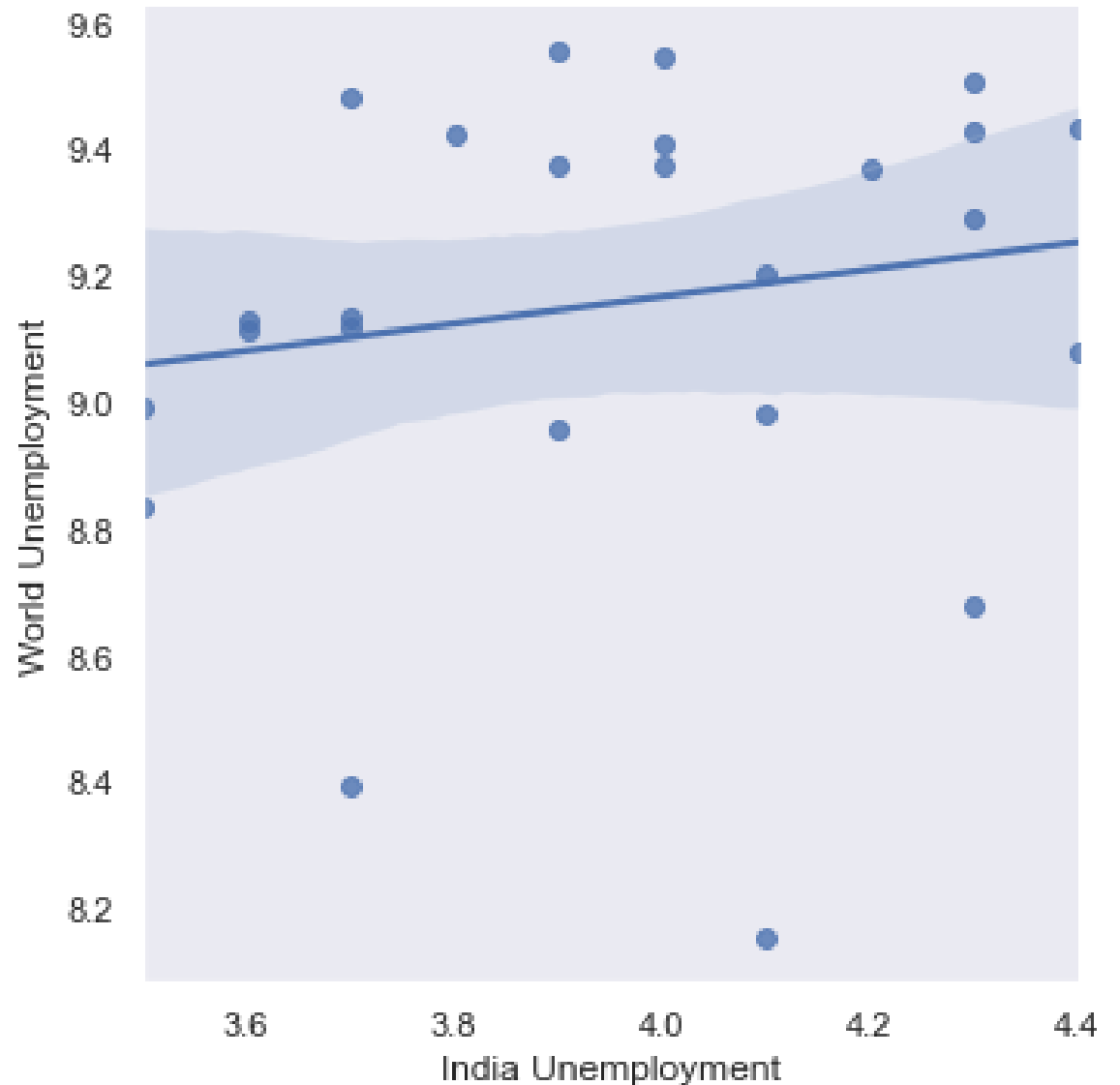
Global Employment - 91.1%.

In last 25 years, India's employment rate was decreased by 0.6% while that of world decreased by only 0.1%

➤ According to the 2010 global census, India was in 28th rank with 96.4% employment rate.

Top 5 countries in the list are Qatar, Cambodia, Belarus, Myanmar, Benin with average of 99.4% employment.

➤ Overall, India's employment has grown fast as compared to the world but still has very low employment as compared to many developed countries.



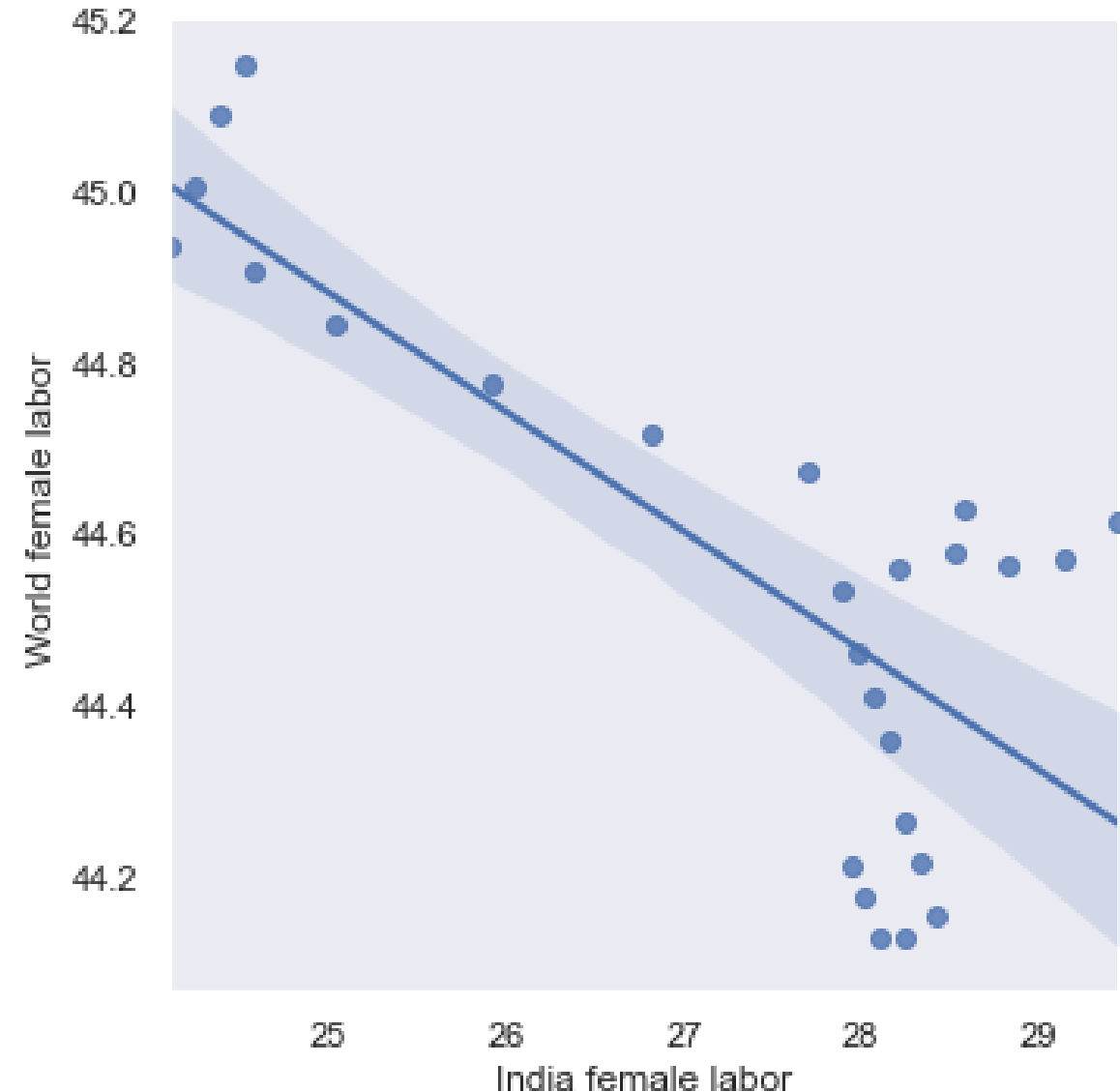
FEMALE LABOR FORCE

- During 90s,
India's female labour force – 27.93%
Global percentage – 44.2%

In 25 years, the percentage of female labour decreased by 3.5% while that of world increased by 1% overall.

- As per 2010 data, India is in 168th position with an average of 25% female involvement in labour work. It comes in below 20 countries.

In neighbouring countries, Bhutan has highest female labour percentage of 42%. Countries from Middle east and few from South Asia has lowest female labour percentage.



CONCLUSIVE SUMMARY

Enrolment ratio:

Considering global enrolment ratio at primary level for the 25 year period, it fluctuated from 94% to 106%. So, overall 12% growth in enrolment ratio. Income wise country enrolment is as follows:

1. Low income: 75 to 123, growth: 48%
2. Lower middle income: 87 to 101, growth: 14%
3. Upper middle income: 99 to 101, growth: 2%
4. High Income: 104 to 100, growth: -4%

Enrolment ratio of India: 92% to 109%, growth: 17%

Employment rate:

Considering global employment rate at for the 25 year period, it fluctuated from 91.1% to 92.2%. So, overall 1% growth in employment rate.

Income wise country employment is as follows:

1. Low income: 98 to 94.7, growth: -3.3%
2. Lower middle income: 93 to 96, growth: 3%
3. Upper middle income: 88 to 87, growth: -1%
4. High Income: 91.5 to 94, growth: 2.5%

Employment rate of India: 95.9% to 96.5%, growth: 0.6%

Female Labor Force:

Considering global female labor force for the 25 year period, it fluctuated from 38.3% to 41.5%. So, overall 3.2% growth in female labor force.

Income wise country employment is as follows:

1. Low income: 27.5 to 32.5, growth: 5%
 2. Lower middle income: 28.9 to 30.1, growth: 1.2%
 3. Upper middle income: 43 to 43.5, growth: 0.5%
 4. High Income: 44.2 to 46.4, growth: 2.2%
- Female Labour percentage of India: 28% to 24.5%, growth: -3.5%

INSIGHT/ FINDINGS(CONSIDERING 3 PARTS)

- A. Nations with higher income has lower enrolment ratio. According to PISA report, countries with higher income has their students performance score high compared to the others. So, class repetition is low. Hence, enrolment ratio is close to 100% not above, as in low and lower income countries.
- B. Nations with low income has higher employment compared to others. Most of the low and lower income countries are dependent on primary sector which creates easy employment. On the other hand, higher income countries are mostly lied on tertiary and secondary sector.
- C. Higher income countries are having maximum participation of women while low and lower income countries are getting faster growth in female participation with low female participation initially. Exceptions also exist, such as low and lower income countries from Sub-Saharan Africa and South Asia have more than half of their women in labor force but with very low wages.
- D. As per TIMSS, PIRLS AND PISA report 2010,
 - 1. Countries with highest scores in Mathematics, science and reading are from only 3 regions – East Asia and Pacific, Europe and Central Asia, and North America
 - 2. Europe has most number of countries having students(under 15) with highest scores in all three fields.
 - 2. China has their students (under 15) with highest scores in 2010.

THANK YOU!

