Capstone Project
World Bank Education Analysis (Exploratory data analysis)

THE WORLD BANK

- **By Azad Meshram**

Content

- **Introduction**
- **Data Pipeline**
- **Exploring Dataset**
- **Attribute Information**
- **Problem Statement**
- **Data Cleaning and Handling**
- **Data Visualization**
- **Key Insights**
- **Conclusion/Recommendations**

Introduction

- **The World Bank Ed Stats All Indicator Query holds over 4,000 internationally comparable indicators that describe education access, progression, completion, literacy, teachers, population, expenditures and many such indicators .The indicators cover the education cycle from pre-primary to vocational and tertiary education and also holds learning outcome data from international and regional learning assessments (e.g. PISA, TIMSS, PIRLS), equity data from household surveys, and projection/attainment data.**

- **I am exploring and analyzing the data to identify variation of indicators across the globe, which countries are more alike and different, which countries show more closeness to global standards and what are the actual driving force behind such growth.**

Data Pipeline
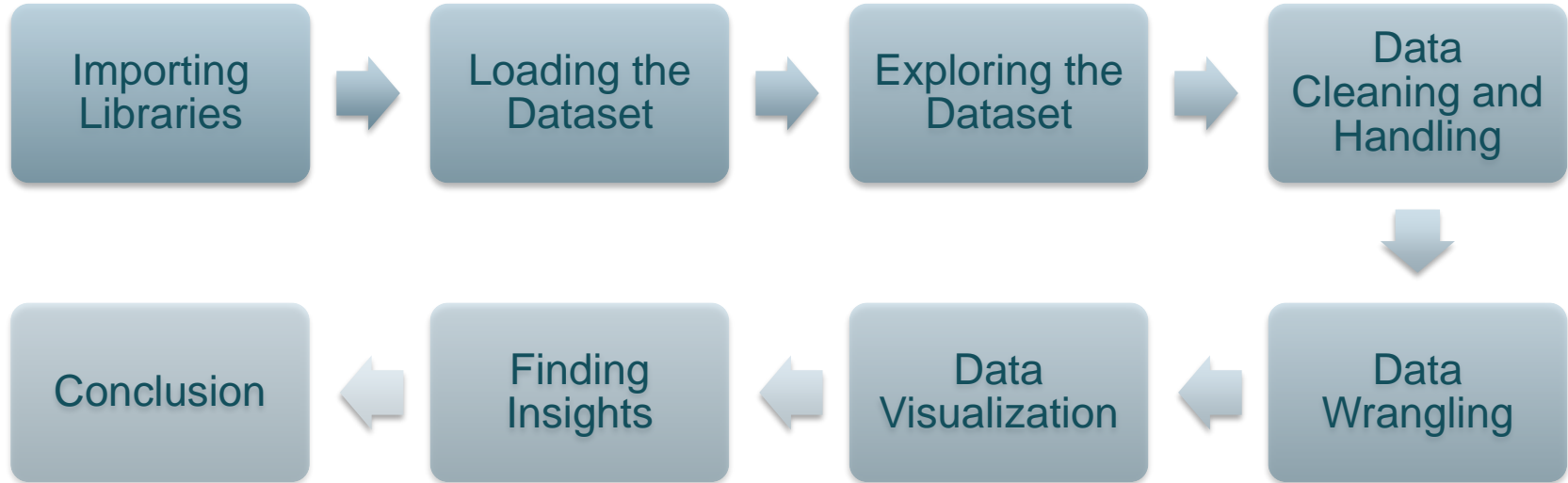
Exploring Dataset

➢ **The World Bank have beautifully arranged the Ed stats dataset with the help of other four subsidiary datasets . So basically there were five files.**

➢ **One of them was defining the countries and there basic information like the nature of there economy, there census year , any kind of special notes and many more.**

  ➢ **Special Notes are describing some extra information about the countries. E.g for Arab World and East Asia Pacific.**

➢ **Other file was defining the indicators, on which World Bank has studied the 241 countries. It also contains, there short and long definition.**

  ➢ **Indicator is basically the head, on which a specific educational trend have been monitored. for e.g the average attendance rate of schools of a particular country.**

  ➢ **Short and Long definition gives the brief idea about the Indicator. For.eg Mean years of schooling would be define as minimum no. of schooling years , people received in a country.**

Exploring Dataset

➢ **In indicator file ,World Bank have studied 3666 indicators over 241 countries giving there trends from 1970 to present years.**

➢ **The Data frame is full of null values and devoid of duplicate values.**

➢ **In the Ed Stats Dataset there are near about 9,00,000 rows and 70 columns.**

➢ **The Country Name , Country Code , Indicator Name and Indicator code has 'object' as their type.**

➢ **The Year Columns(1970-2100) has the float as their Dtype**

Attribute Information

➢ **Country Name : The name of the countries. E.g India.**

➢ **Country Code: Abbreviation of countries. E.g IND for India**

➢ **Indicator Name : The head on which trends have observed. E.g Unemployment**

➢ **Indicator Code: Abbreviation to show long indicator name in short. E.g Unemployment : SL.UEM.TOTL.ZS**

➢ **Years : From 1970 to 2100, the individual year is a column**

Problem Statement

❖ **First Level of Analysis**

1. Analyse the trends for **Gross enrollment ratio(GER)**, **Lower secondary completion rate(LSC)** , **Adult literacy rate(ALR)** and **Pupil-Teacher ratio(PTR)** for Arab World group of countries. And Compare them with one another.

2. Analyse the effects of **Government expenditure(GEE)** on other education heads for United Kingdom.

3. Compare the trends of GER,LSC,PTR and GEE for Japan.

4. How Unemployment and GEE is closely interlinked in Argentina.

5. Find out, whether increase in GEE would help Botswana to create positive impact on Employment trends.

6. What is the current status of education in East Asia Pacific nations.

Problem Statement

❖ **Second Level of Analysis**

1. Compare the GER of all the above mentioned countries and compare the trends.

2. Analyse and compare LSC of all the above mentioned countries and draw the trend fluctuation.

3. Observe the GEE of all the above mentioned countries and compare it with their present educational status.

4. Observe the trends of Unemployment, where GEE is low.

5. Analyse the behaviour of GER, GEE, Unemployment with respect to PTR.

❖ **Third Level of Analysis**

1. Just compare all the indicators of every country in one frame.

Cleaning and Handling Dataset

❖ **There are no duplicate columns.**

❖ **EdSats dataset is full of null values. These null values can not be dropped directly as we have to show the yearly trends.**

❖ **If these Years are dropped ,the column year would not be preset in the comparison dataframe, making it difficult to track the correct trends.**

❖ **Hence I have opted to replace these null values by zero.**

❖ **This representation of zero, would suggest countries to look for the correct data.**

❖ **Later on When I need to find out the mean of newly formed desired data frames, I replaced those zeros back with their respective null values . So that the mean should not be affected.**

Cleaning and Handling Dataset

❖ **Country Name, Indicator Name, Country Code and Indicator Code column are wholly filled, without any null values with object type of characters.**

❖ **Columns form 1970 to 2100 are mostly empty and null. All these null values are filled with zeros.**

❖ **The filled values from 1970 to 2100 are float type.**

❖ **For Comparing the trends of LSC of four countries, I have used the values as the size of dots in scatter plot. For that the values have been converted to 'INTEGER' type from 'float'.**

Data Visualization

First Level Analysis

Government expenditure on education as % of GDP (%)
1   3   5   7   8   10   12
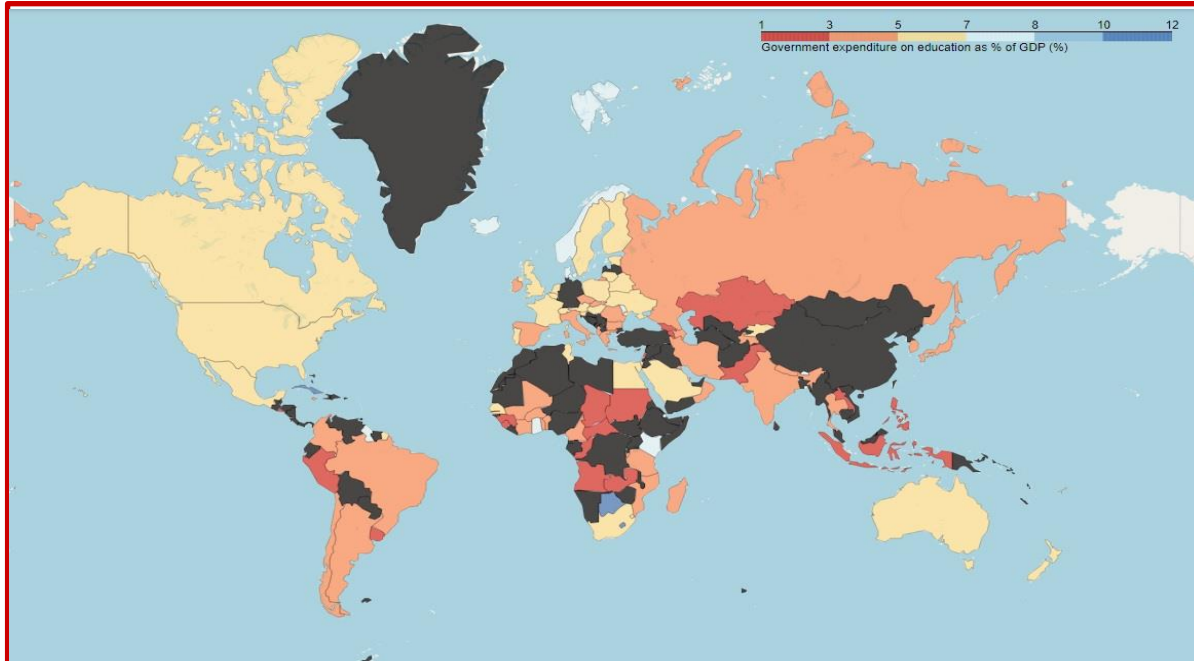
The variation of government expenditure, throughout the countries.

World Bank education dataset was full of null values and devoid of relevant information on specific indicators. So the help arrives from Choropleth to understand which countries are lacking in some specific trends. The black colour shows the absence of data in the respective countries

Data Visualization

First Level Analysis

Unemployment, total (% of total labor force)

0   5   11   16   21   26   31
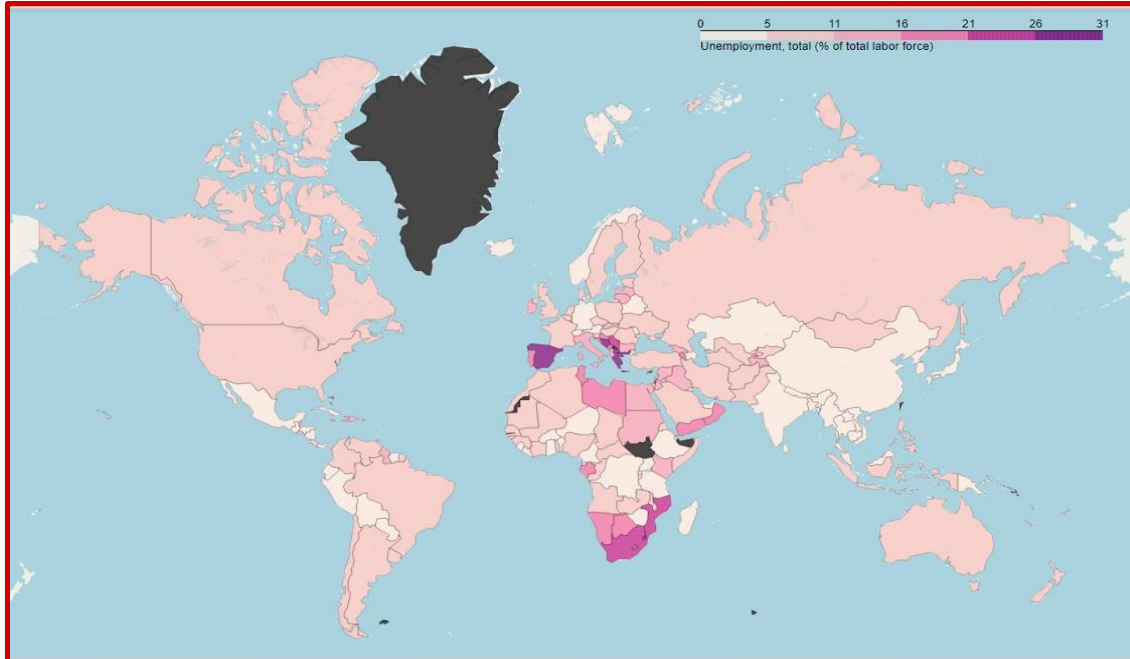
The variation of Unemployment throughout all the countries.

- For e.g, Norway which is having the govt. expenditure on education at fourth scale, experiences less unemployment.
- India , which has 3% of GDP as expenditure on education which is second level , shows near about 5% of unemployment, which is the extreme of first scale.
- This comparison can be easily done for other countries to have a rough estimate of both the indicators
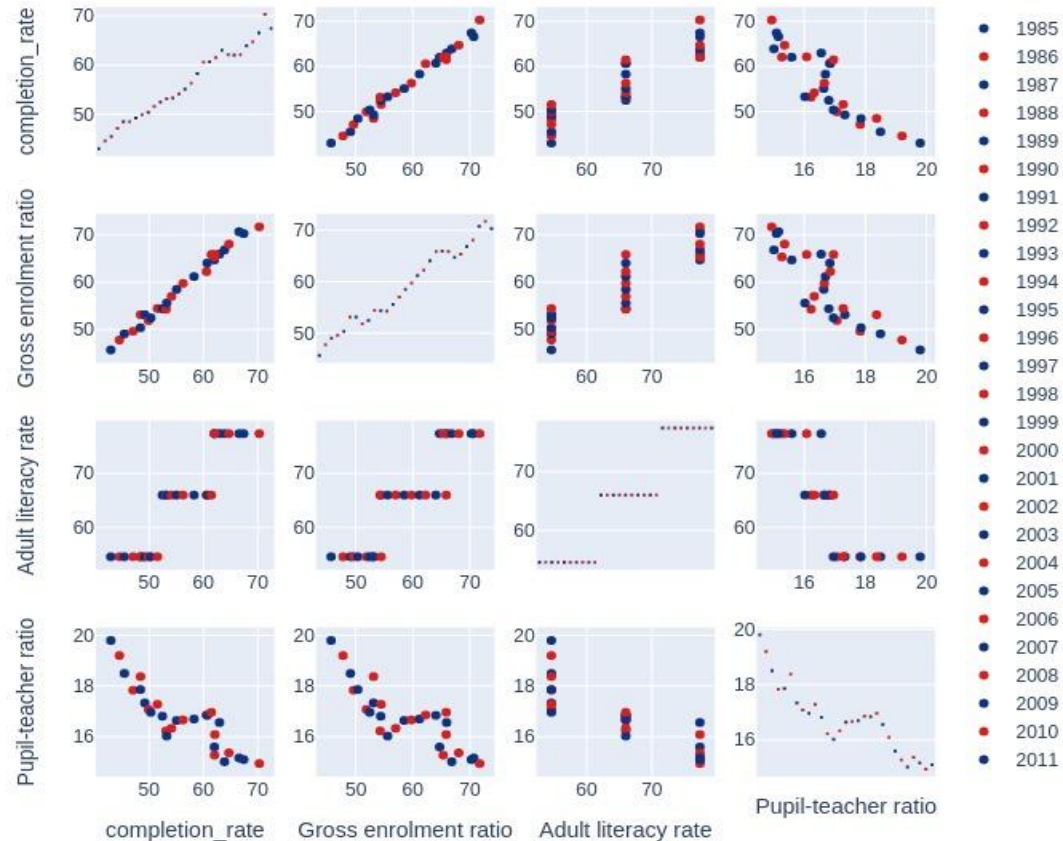
Data Visualization

Problem Statement 1: Analyse the trends for Gross enrollment ratio(GER), Lower secondary completion rate(LSC) , Adult literacy rate(ALR) and Pupil-Teacher ratio(PTR) for Arab World group of countries. And Compare them with one another.

- **Literacy rate is calculated in every ten years.**
- **Pupil-Teacher ratio is decreasing at increasing trends of gross enrollment rate and lower secondary completion rate.**
- **Increasing trends of enrollment and completion rate is pushing the adult-literacy rate.**
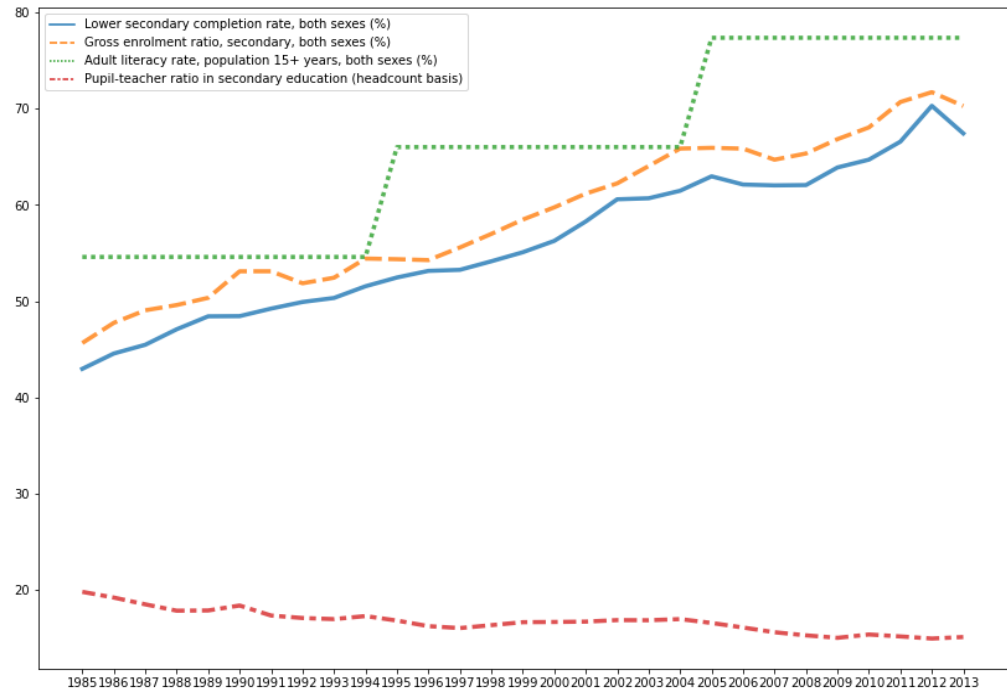
Data Visualization

Problem Statement 1: Analyse the trends for Gross enrollment ratio(GER), Lower secondary completion rate(LSC) , Adult literacy rate(ALR) and Pupil-Teacher ratio(PTR) for Arab World group of countries. And Compare them with one another.

- **It clearly shows the pupil-teacher ratio is indirectly proportional to every other indicator in the chart.**
- **As all the indicators are showing positive change, but the results are not at the par with global standards. It means the actions taken by government requires more judiciousness, political will to penetrate inside and effective implementation strategy.**


Scatterplot Matrix

Data Visualization

Problem Statement 2:Compare the trends of GER,LSC,PTR and GEE for Japan.

✓ The trends of Japan clearly demonstrate the increasing government expenditure broughts down the pupil-teacher ratio and increase the Gross enrollment ratio and lower secondary completion rate.

✓ But the chart of Japan also gives one more important insight. When they reduce the government expenditure, the probable outcome should be negative trends in other indicators, but the situation remains as it is.

✓ Here Japan would have arrived at the most reliable and appropriate percentage of GDP required for their education sector. This is the necessary monitoring of socio economic conditions of any particular country.
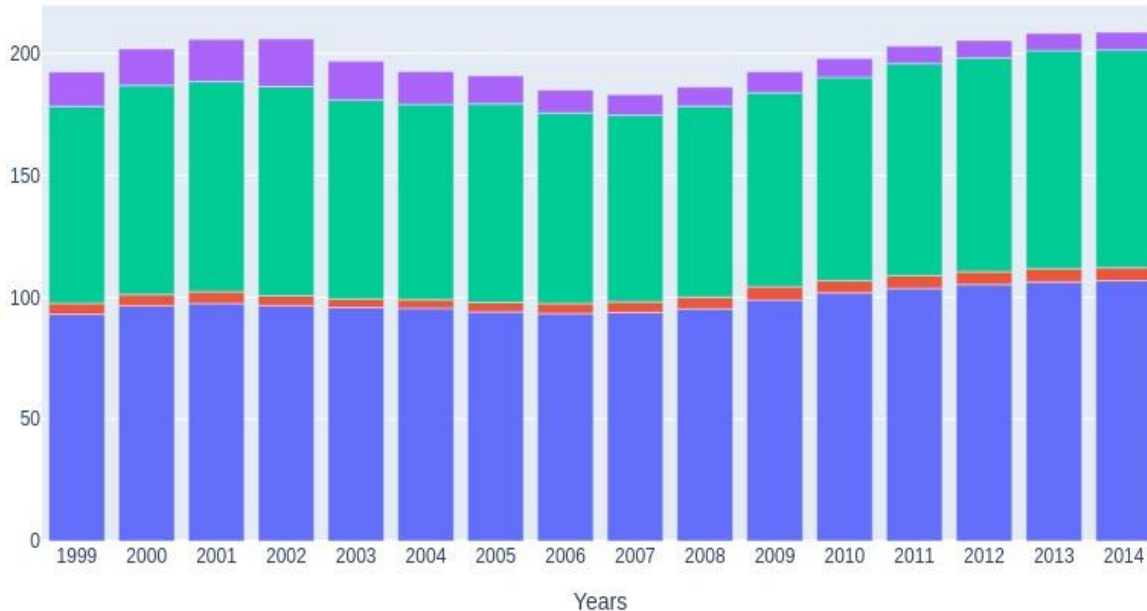

Scatterplot Matrix
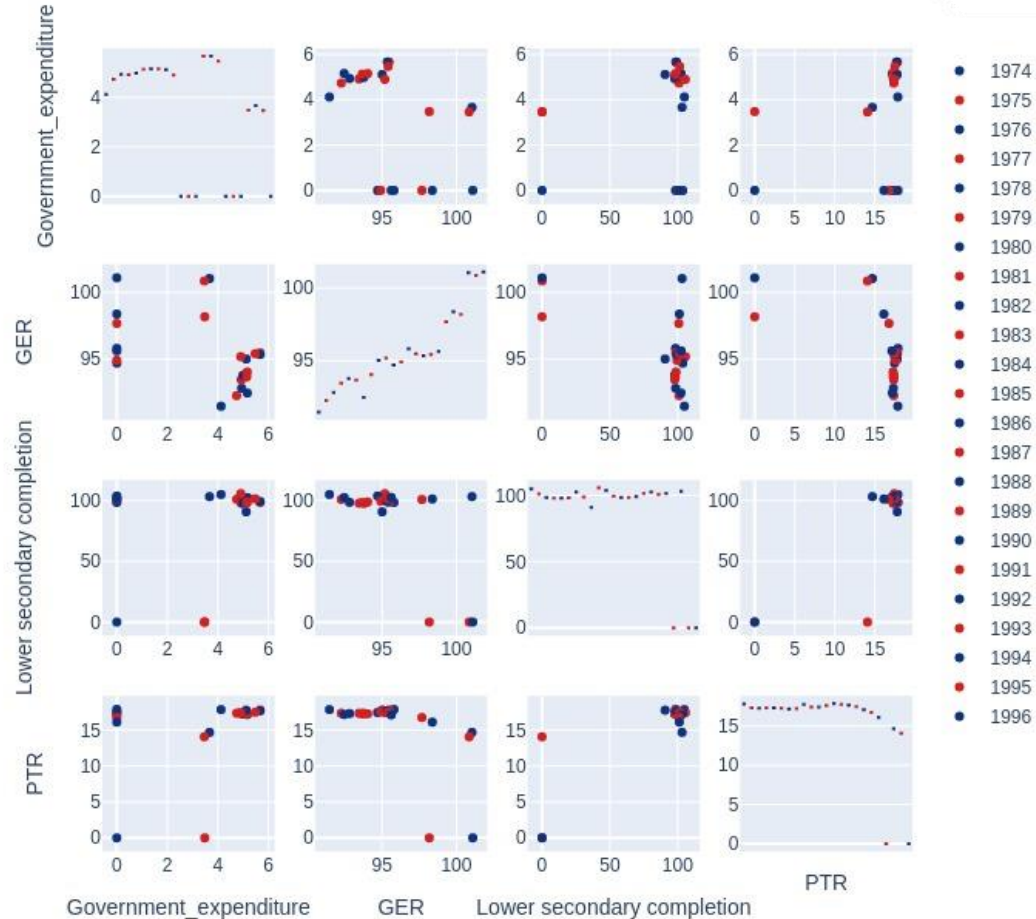
Data Visualization

Problem Statement 3: How Unemployment and GEE is closely interlinked in Argentina.
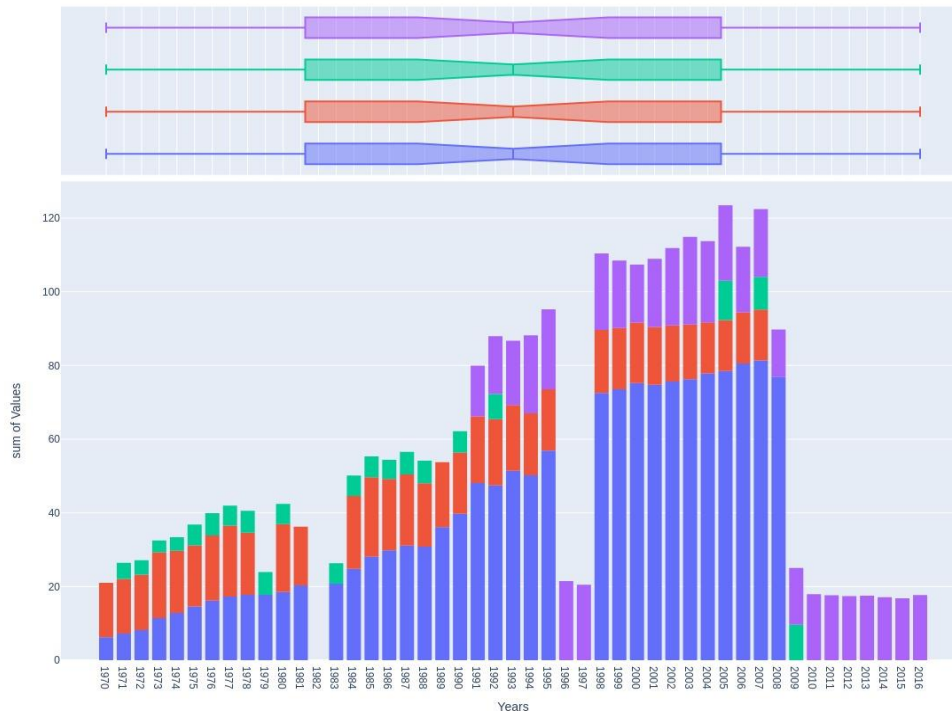
Gross enrollment and govt. expenditure

- In 2002 When expenditure lowers to 4 % of GDP, unemployment increases to near about 20 %.
- But later on when Argentina have increased the expenditure in 2014 to 5.3 % of GDP, the unemployment lowers beautifully to 7.3 %.
- Nevertheless 7 % unemployment is still high, according to the standard global requirements.

Data Visualization

Problem Statement 4: Find out, whether increase in GEE would help Botswana to create positive impact on Employment trends.
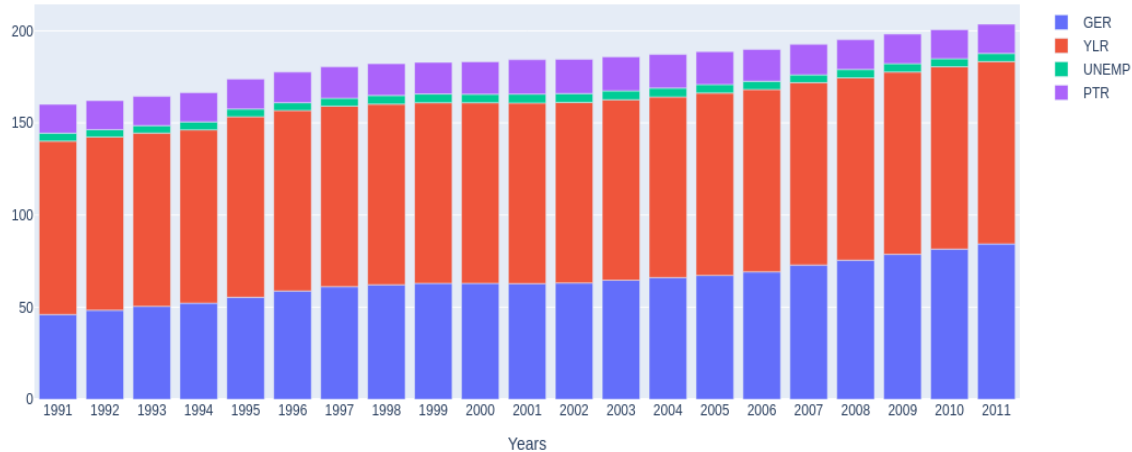

Indicators
- Gross enrolment ratio, secondary, both sexes (%)
- Pupil-teacher ratio in secondary education (headcount basis)
- Government expenditure on education as % of GDP (%)
- Unemployment, total (% of total labor force)

➢ **It has found that Botswana is contradicting the postulates which have been given by earlier studies.**
➢ **In Botswana , despite increasing the government expenditure , the unemployment rate is still very high.**
➢ **Botswana spend nearly 10 % of GDP on education but failed in improving the employment rate.**

Data Visualization

Problem Statement 5: What is the current status of education in East Asia Pacific nations.

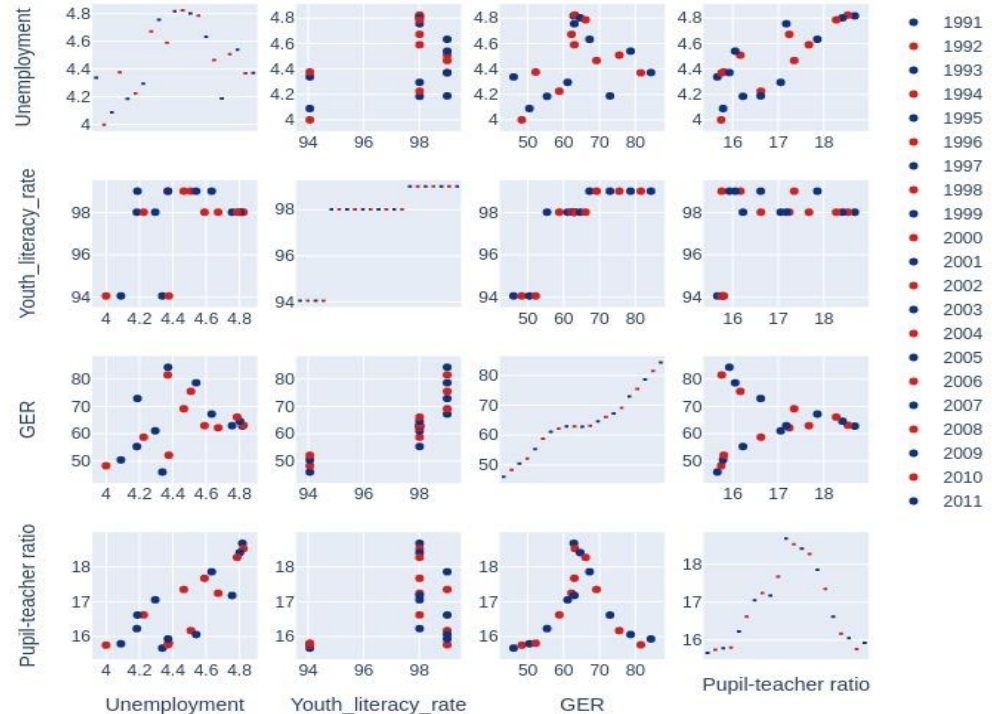Analysis of various indicators in East Asia Pacific

➢ **East Asia Pacific shows all the trend patterns and growth indicators very clearly in conventional manner .**
Its enrollment and youth literacy rate is continuously increasing with time.

➢ **PTR and unemployment rate shows little movement but they are near to the global standards.**

➢ **Gross enrollment ratio has tremendously increased from 50% to more than 80%.**

Data Visualization

AI

Problem Statement 5: What is the current status of education in East Asia Pacific nations.

❖ **East Asia Pacific shows all the trend patterns and growth indicators very clearly in conventional manner .**

❖ **GER and ALR shows straight forward improvement, uplifting the quality of employment by decreasing the PTR and unemployment.**

❖ **Unemployment and PTR shows great coincidence with each other. In the times of 2002 the Unemployment and PTR were at the peak.**

❖ **Later on due some probable good policies, both the parameter have reduced to normalcy and touched the Global standards.**
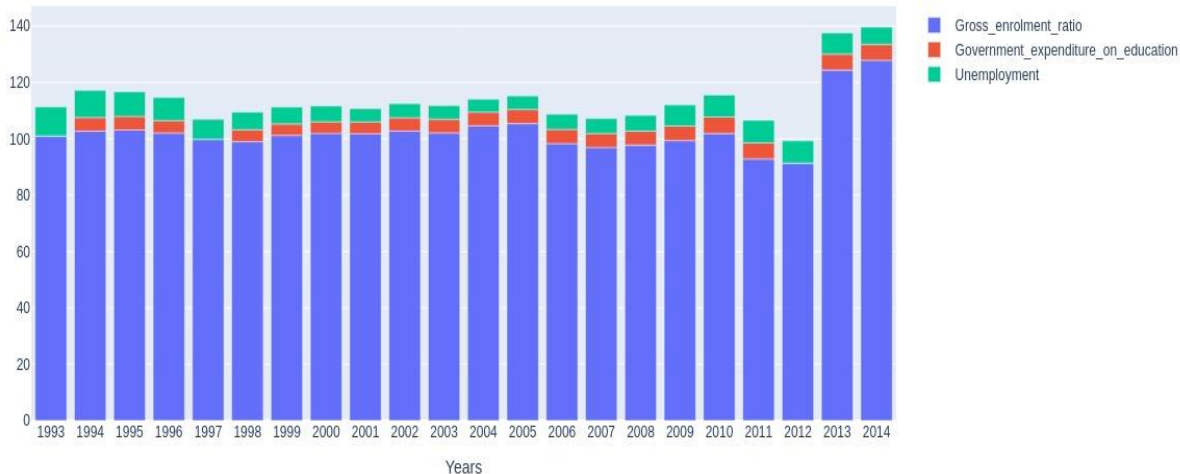

Scatterplot Matrix

Data Visualization

AI

Problem 6: Analyse the effects of Government expenditure(GEE) on other education heads for United Kingdom.

- **Government Expenditure on Education is following the trends of Unemployment.**
- **United Kingdom is a developed nation and shows the obvious trends according to the GEE**


Effect of GEE on GER and UN

Legend:
- Gross_enrolment_ratio
- Government_expenditure_on_education
- Unemployment

Years

Data Visualization

Problem 6: Analyse the effects of Government expenditure(GEE) on other education heads for United Kingdom.

- **Government Expenditure on Education is following the trends of Unemployment.**
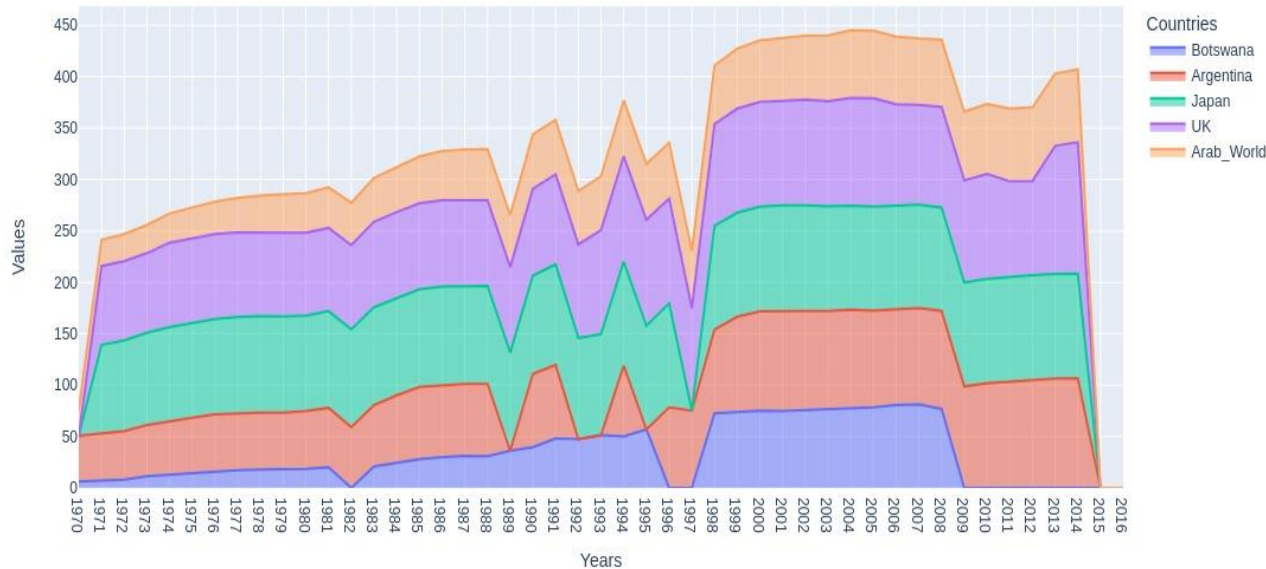- **United Kingdom is a developed nation and shows the obvious trends according to the GEE**

Unemploymnet and Government expenditure
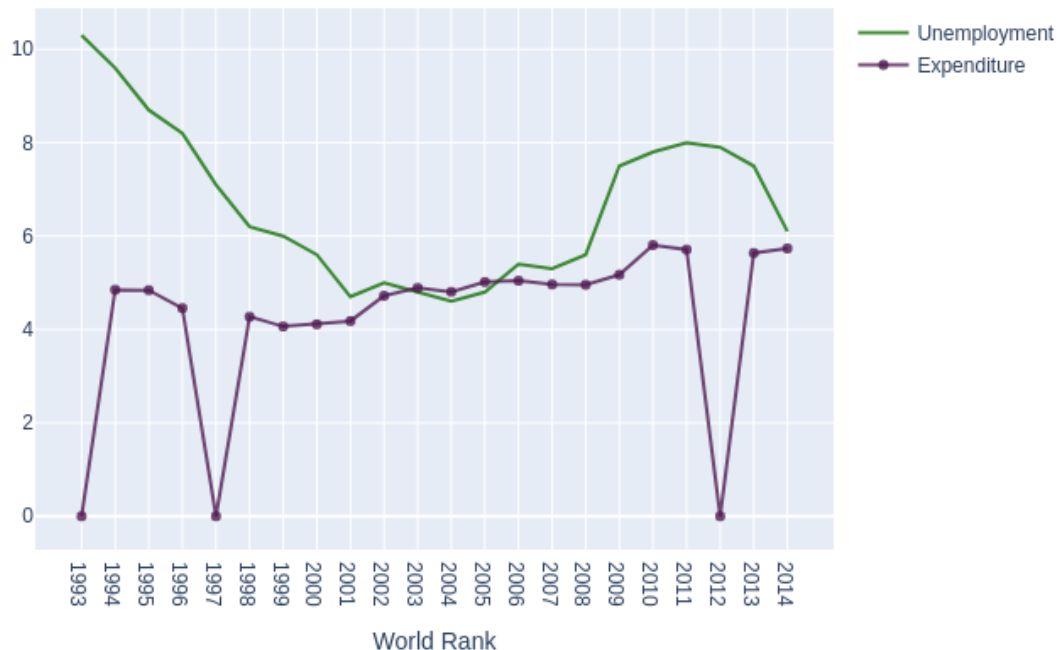
Data Visualization

Second level of Analysis

Problem Statement 7: Compare the GER of all the above mentioned countries and compare the trends.

❖ **It can be observed that all the countries are in upward trends in GER, and can be analysed on other grounds.**

❖ **Every country is showing the presence of resistance at the same time.**

❖ **It might be due to the international disorder.**

Data Visualization

Problem Statement 8: Analyse and compare LSC of all the above mentioned countries and draw the trend fluctuation.
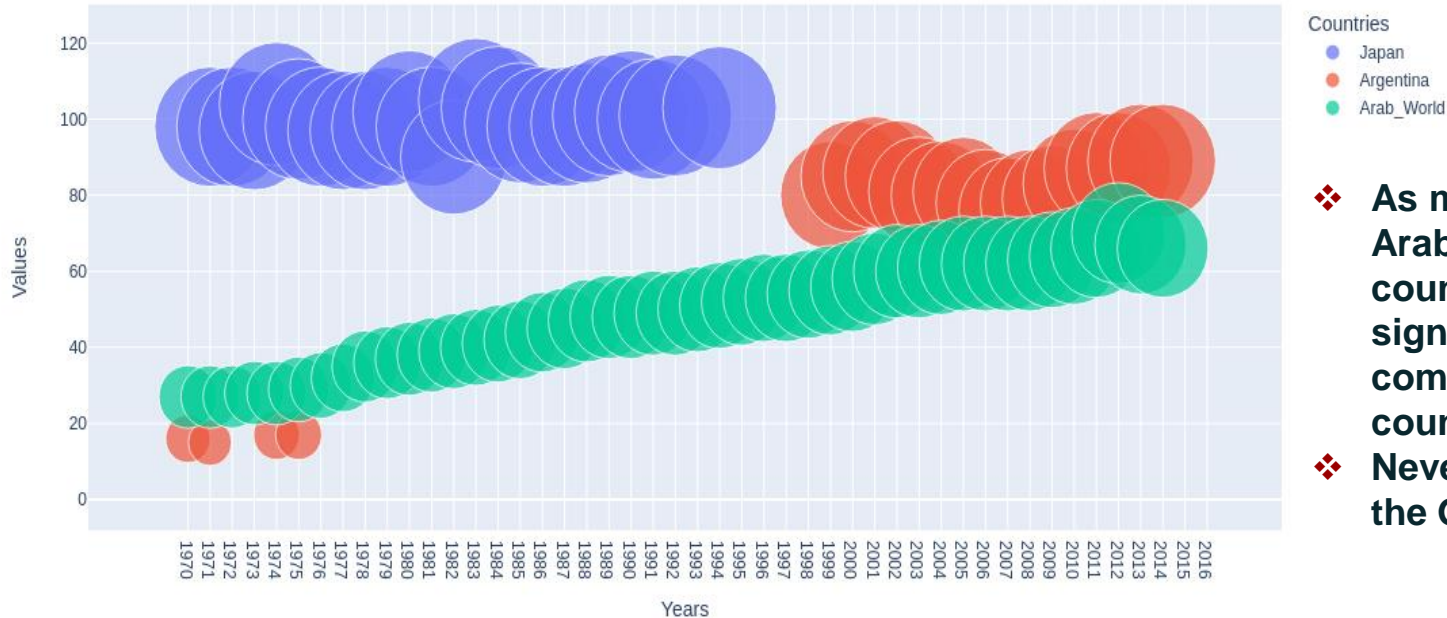
❖ **As mentioned earlier the Arab World group of countries have shown a significant growth in LSC in comparison with other countries.**

❖ **Nevertheless it is still below the Global standards.**

Data Visualization

Problem Statement 9: Observe the GEE of all the above mentioned countries and compare it with their present educational status.
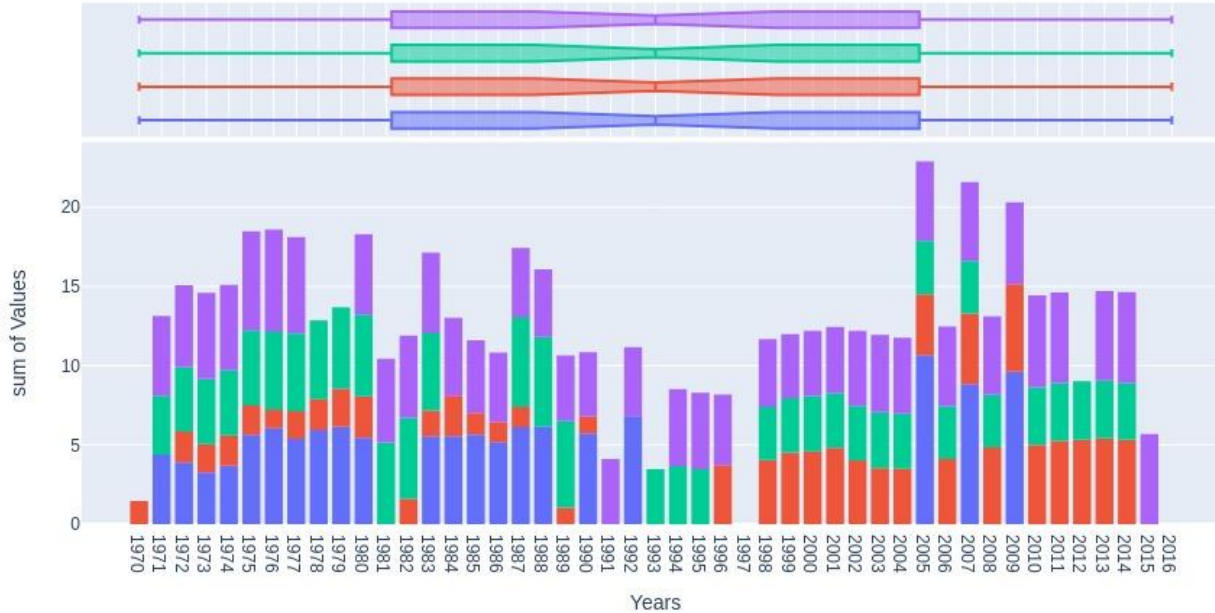
- Earlier we have mentioned the expenditure trends of Botswana , which is near about 10 % of GDP in 2005.

- But the other countries are at the little less levels.
- Japan has low GEE , but has balanced unemployment rate.
- Argentina shows the increase in GEE form 1996.This change also could be observed in its unemployment rate

Data Visualization

**Problem Statement 10: Analyse the behaviour of GER, GEE, Unemployment with respect to PTR.**
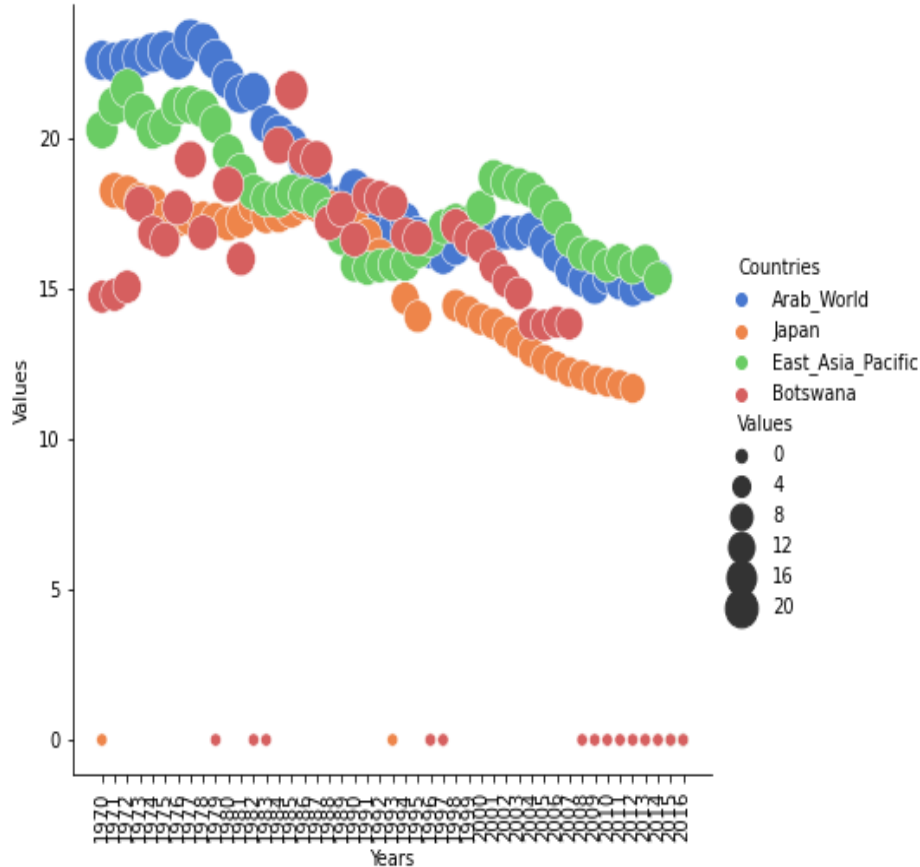
❖ **Every country is showing the downward trends for PTR.**
❖ **Arab World and Botswana have shown a significant come back since 1970 to 2016.**

❖ **When this map combines with the earlier one, the PTR for Botswana shows the inverse relation with GEE.**
❖ **When the expenditure was reduced in 1983 , the PTR was at its peak.**

Data Visualization

Problem Statement 11: Observe the trends of Unemployment, where GEE is low.
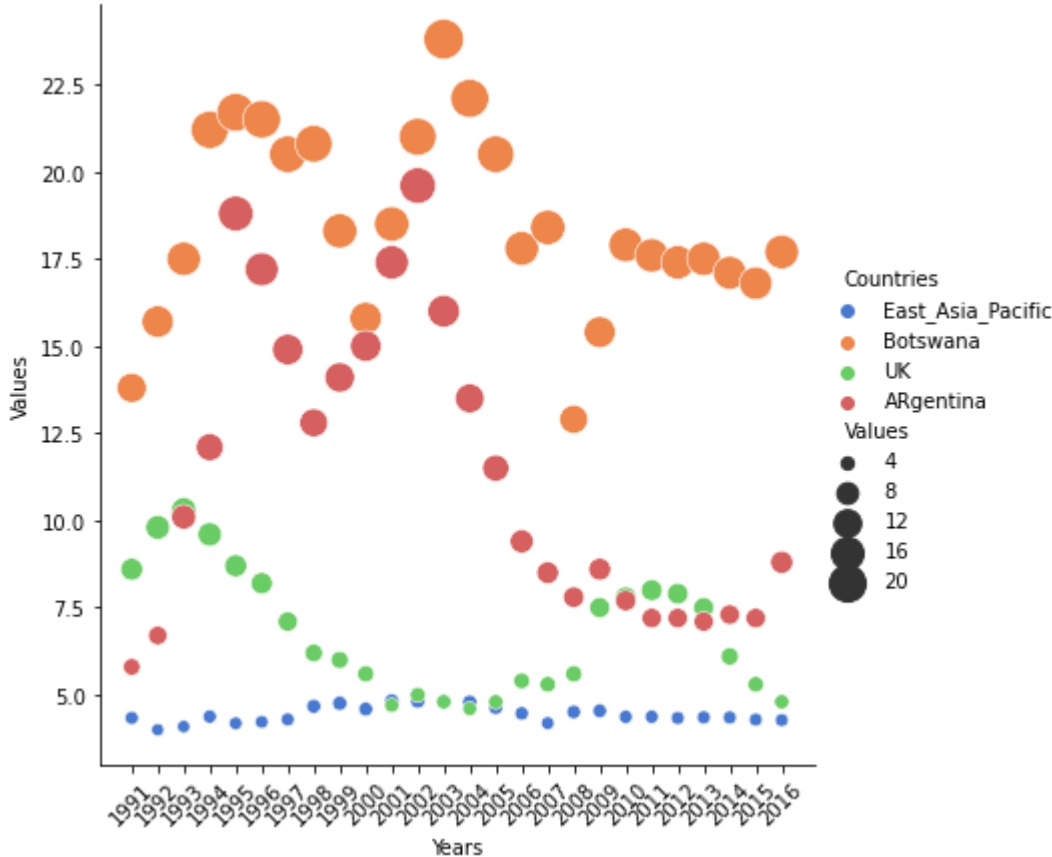
- ➤ **Botswana in addition with Argentina showing high levels of unemployment. Even with the passage of time Argentina is reducing with its severity but Botswana is still on a comparative high position.**

- ➤ **When this unemployment trends get compared with the trends of govt expenditure, a close coincidence can be observed. Both are flowing parallely . In next graph both this trends can be observed regarding Argentina .**

- ➤ **UK and EAP shows the coincidence with the standard global trends.**

Data Visualization

Problem Statement 12: Compare all the indicators of every country in one frame, and analyse the comparative movement of indicators.
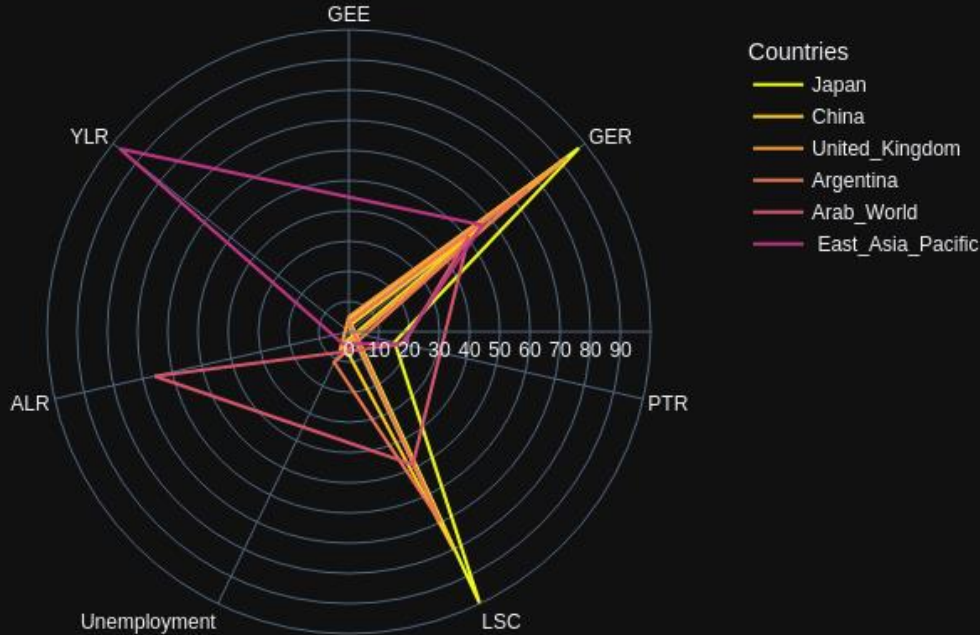
- ➢ **The universal truth come in front of us via this line polar chart. It shows those countries with enough government expenditure on education shows the maximum positive trends.**
- ➢ **for e.g Japan with near about 4 % of expenditure shows the positive trends for all the indicators.**
- ➢ **With this analysis , countries those are underperforming could understand, which country they need to study.**

Key Insights

1. Despite the positive results , the Gross enrollment and Lower secondary completion rate is not up to the global average for **Arab World Countries**. The countries should be focusing on, increasing the participation of students and their completion.

2. The direct outcome of **United Kingdom** chart is , government expenditure on education is very much influencing factor for the changes in unemployment rate of countries.

3. **Japan** have balanced the public expenditure at 4 % , but still succeeded in making significant enhancements. Countries should follow the case studies of Japan who has the stagnated education indicators.

4. **Argentina** shows very close connections with public expenditure. In second level of analysis Argentina shows the depressed trends with less expenditure.

5. **Botswana** spend nearly 10 % of GDP on education but failed in improving the employment rate and pupil teacher rate.

Key Insights

1. It can be observed that all the countries are in upward trends in GER. They show the same way of development throughout the history.
2. The Unemployment trends of Argentina when compared with others, show the close connection with public expenditure. It decreases with increase in expenditure.
3. Other Countries have shown positive results for enhanced expenditure except Botswana . This shows the imperative need of ground level studies.

Solutions to Business objective

- The trends for Botswana is not according to the conventional trends.The investment on education is failing in generating the desired results. The Government should focus on ground conditions, closely analyse the bottlenecks , monitor the nature of spending, assign the high level commission for studies and surveys.

- for e.g In India the parliamentary standing committee have noticed the ineffectiveness of Beti bachao beti padhao scheme , where 70 % of funds were spent only on advertisement, resulted in lack of desired enhancements in girl child education.

- All the countries are performing at their best level to improve the education quality in their own countries

- But some of them are performing very well and some requires more in depth studies in their Education domain. e.g Botswana

- Countries like Botswana, must understand the importance of efficient expenditure in the education I

- They must formulate the efficient policies , form some dedicated organisations to deal with the education related research is sine qua none.

- Countries like United Kingdom and Japan are developed nations and even a little less percentage of GDP as compared to others, on education could be sufficient as their size of GDP is huge and large.

- So Observing their percentage of GDP is not sufficient, other nations should observe their actual expenditure and way of implementation on the domain.

- Understanding the socio economic necessity of the society is imperative in order to improve the government funding on the education. Because funding is the most important way, by which the inclusion of backwards and tribals could be possible in mainstream . But the actual implementation should be monitored.

Conclusions

- So government expenditure on education is the most important factor in the sector. Although, funding is anyway can be a game changer for any businesses. And in the education, funding brings new schemes and new ways to penetrate the power of education. It opens various ways and alternatives to call the backwards and tribals to the mainstream. On the flip side it creates employment and work opportunities for these future educated work force , resulting them into new heights of living.
In 1965 India have appointed the Kothari Commission to monitor the prerequisites required for the growth of education in India. Kothari Commission had recommended the 6 % of GDP should be spent on the Indian Education system , which is still at 3 %. Commission had highlighted the connection between Expenditure --> Gross enrollment --> completion --> increament in literacy --> increament in mean years of schooling --> and finally employment.

Thank You!!

Education is the avenue for Independence , Freedom And Liberty of a Conscious Mind.