

# **Capstone Project - I**

## **Team 5 : World Bank Global Education Analysis**

# Content

## Why?

**"In 40 out of 93 countries, fewer than 50% of the poorest children have completed primary school".**

- **Problem statement**
- **Data Summary**
- **Analysis of population growth of countries in various income levels.**
- **Analysis of Average Expenditure spent.**
- **Analysis of variation of Net enrolment rate.**
- **Analysis of Unemployment across years.**
- **Analysis of literacy rate in various countries.**
- **Challenges**
- **Conclusion**

# Problem Statement

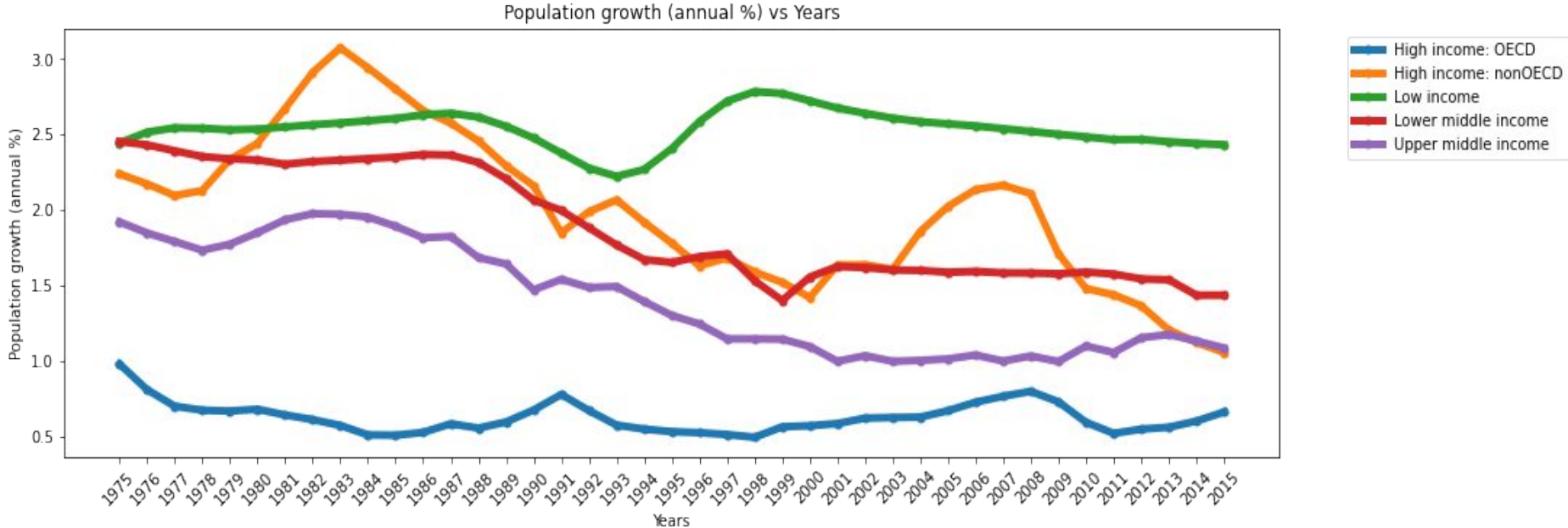
1. How are different income group countries having impact on their education stats?
2. What is population growth trend in different countries and how is it affecting the literacy levels and education scenario ?
3. How is internet playing a role in employment statistics of countries ?
4. How are government expenditures in different education sectors having a role to play in improving literacy and employment stats of different income group countries ?
5. How is net enrolment rate is varying for various income group countries over years?
6. How is the literacy rate varying over some countries?

# Data Summary

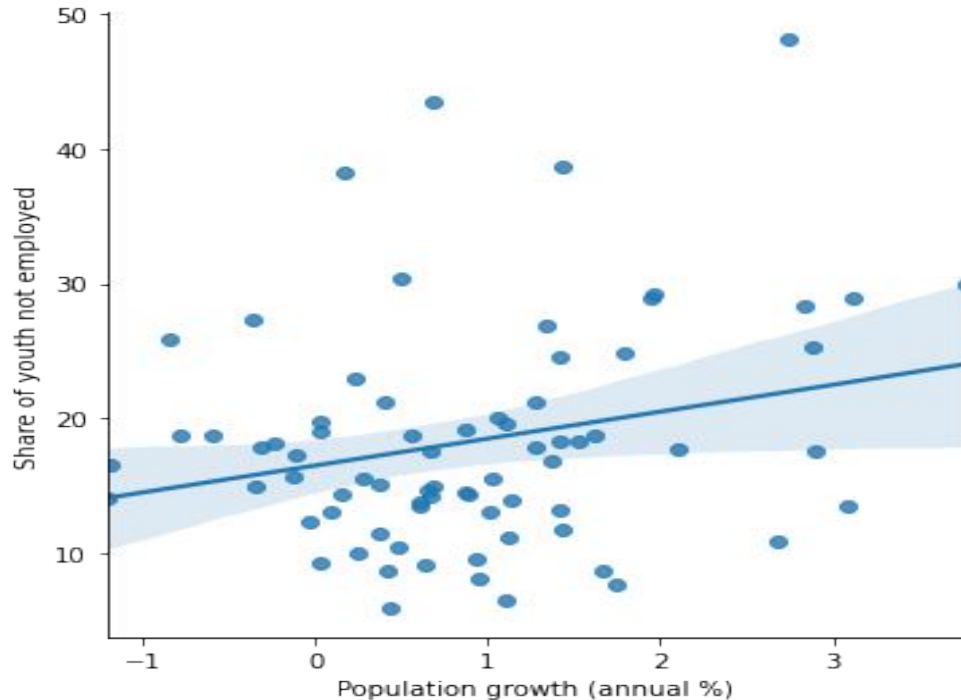
- **EdStatsCountry.csv** contains a list of all participating countries (total of 241) and their income groups.
- **EdStatsData.csv** contains a combination of each country and each indicator (total of 3665 indicators) as well as measurement values for these indicators from 1970 to 2016 and projections from 2020 to 2100.
- **EdStatsSeries.csv** contains the list of all indicators, their category group and further additional information
- **EdStatsCountry-Series.csv** contains general information about entire indicators that have been estimated for certain countries.
- **EdStatsFootNote.csv** contains informations about origin of the data series used for the indicators for every single year

# Population variation across different income groups

The lower income group countries have the maximum population growth whereas the high-income OECD countries have the minimum population growth. The High income non OECD and the middle income groups have actually decreased the population growth rate.

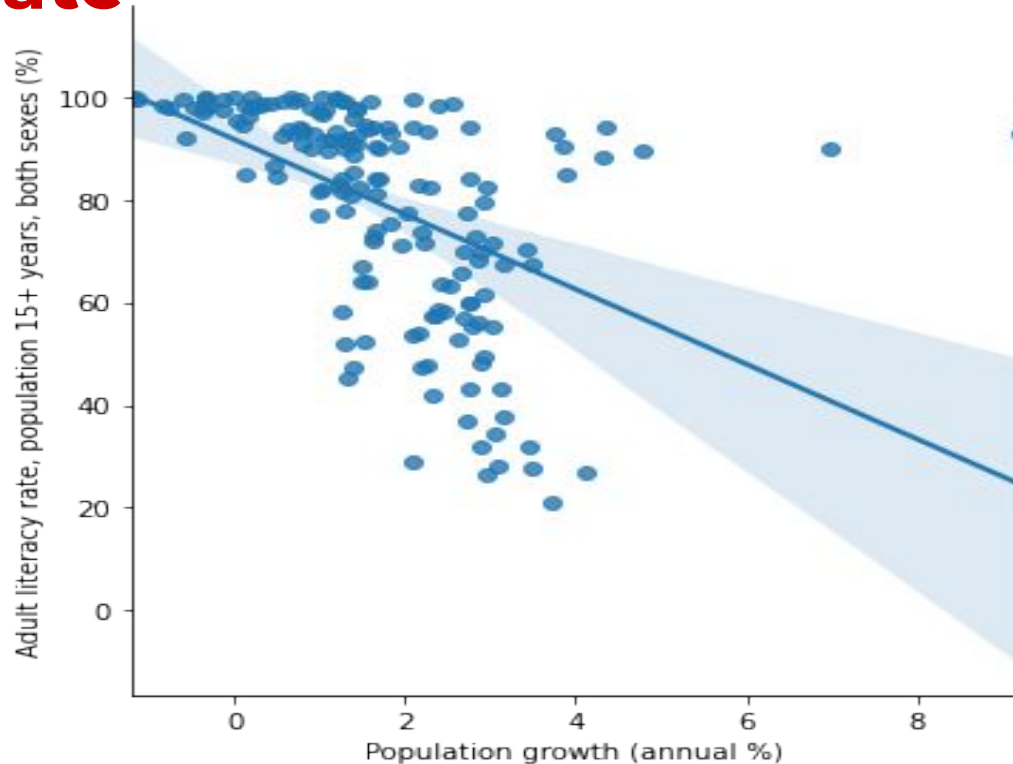


# Population growth rate v/s Share of youth not employed



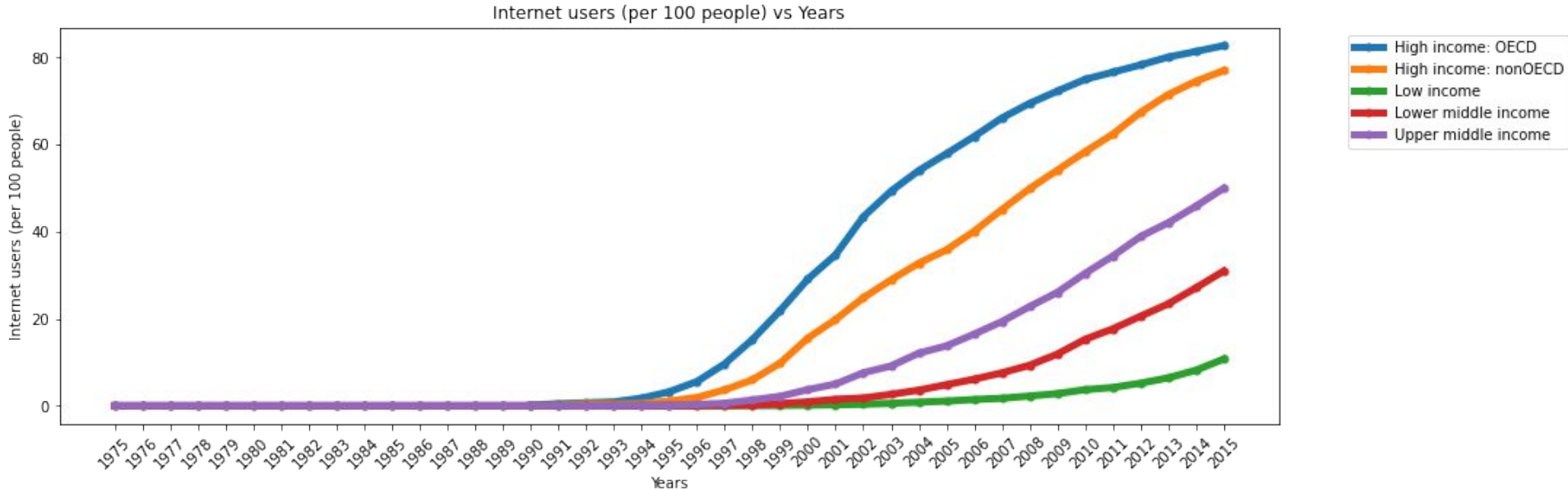
The correlation coefficient for Population growth rate and Share of youth not employed is : 0.25

# Population growth rate v/s Adult Literacy rate



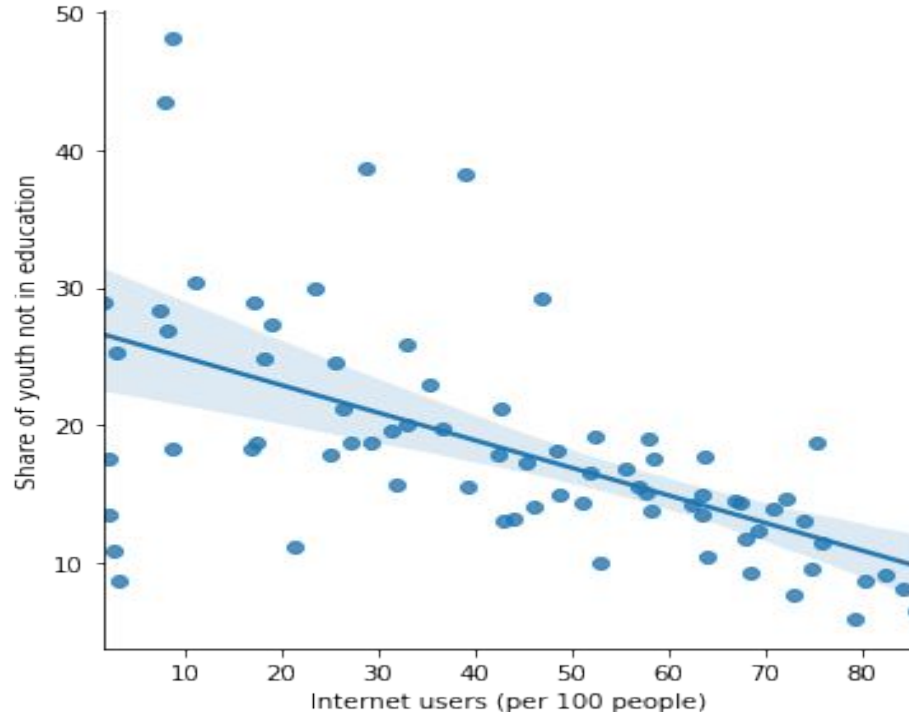
The correlation coefficient between population growth rate and Adult Literacy rate is : -0.49

# Internet users(per 100 people) in different Income groups



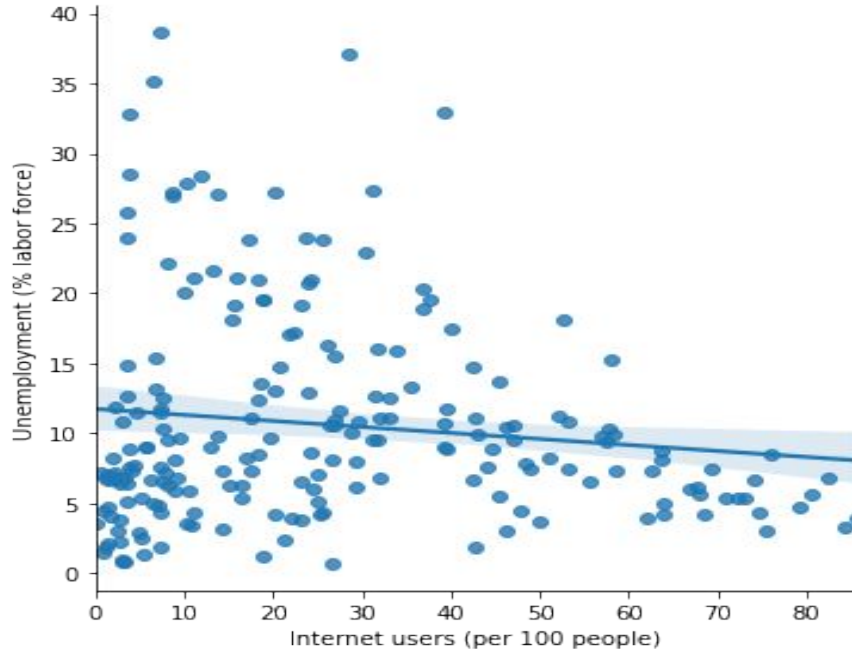


# Internet users per 100 people v/s share of youth not in education



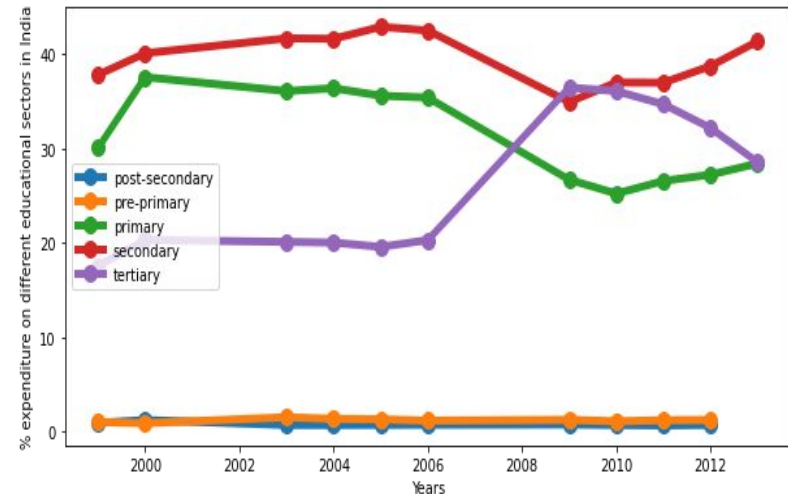
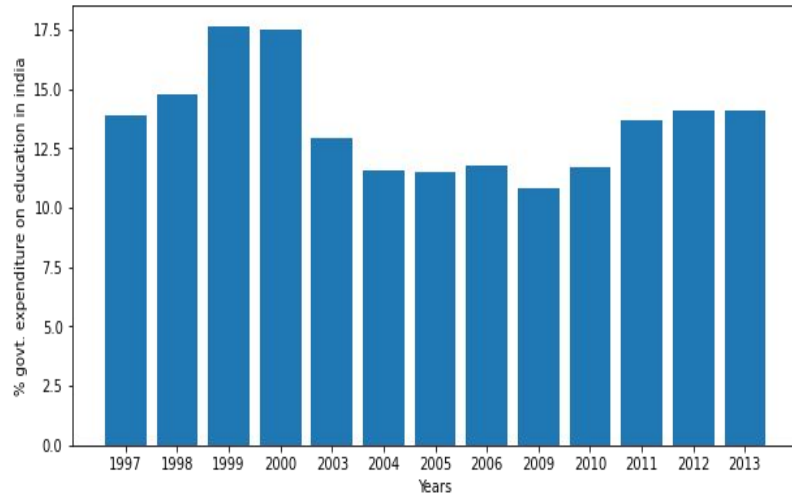
Correlation coefficient  
between Internet users per 100  
people and Share of youth not  
in education is: -0.56

# Internet users per 100 people v/s Total unemployment

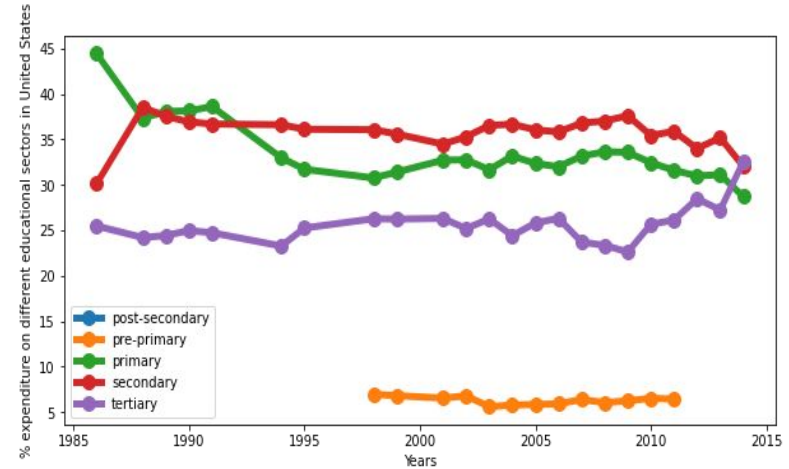
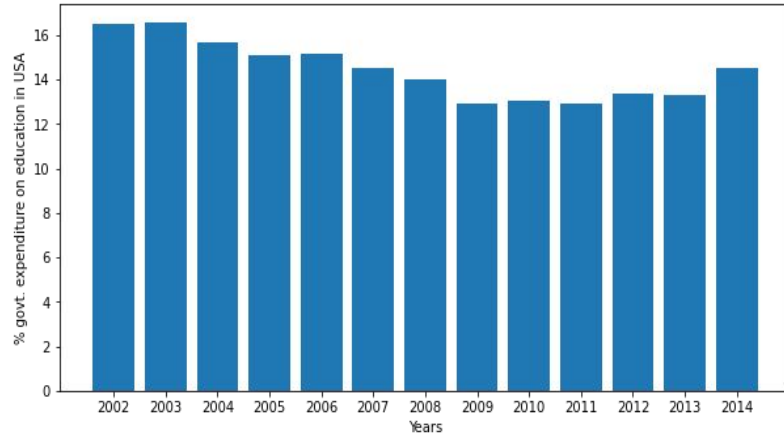


Correlation coefficient  
between Internet users per  
100 people and Total  
unemployment : -0.124

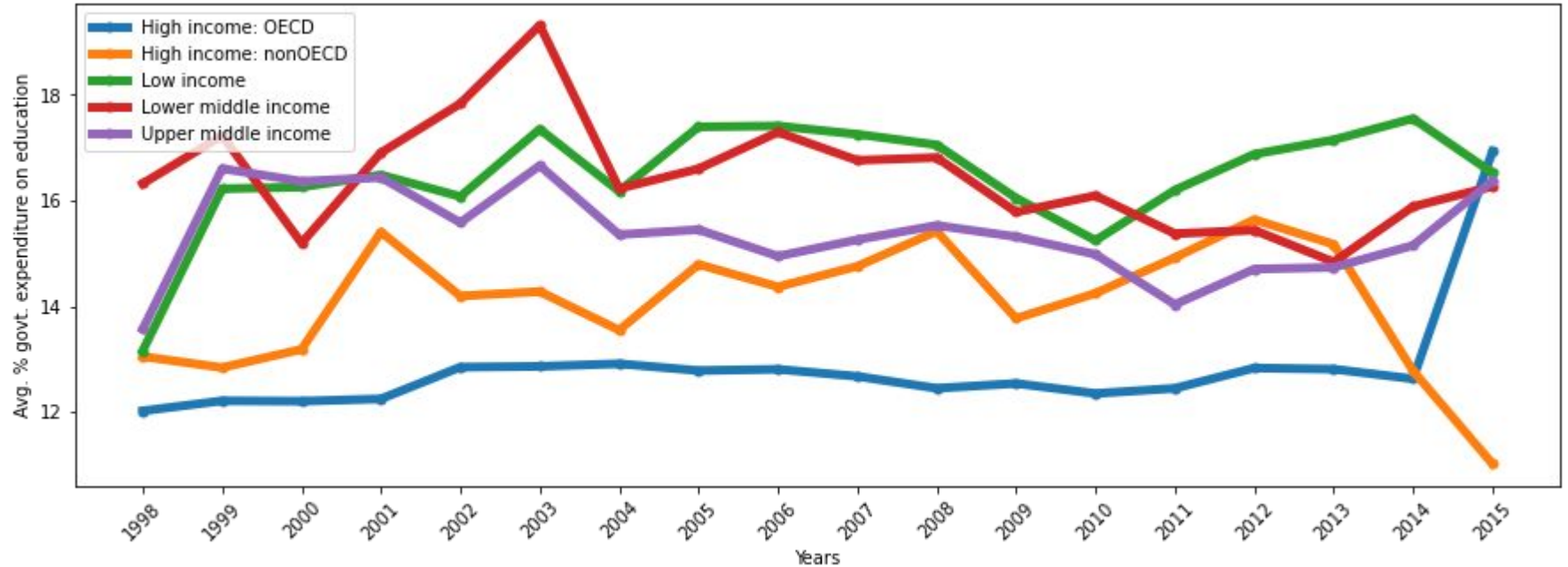
# Expenditure on Education (India)



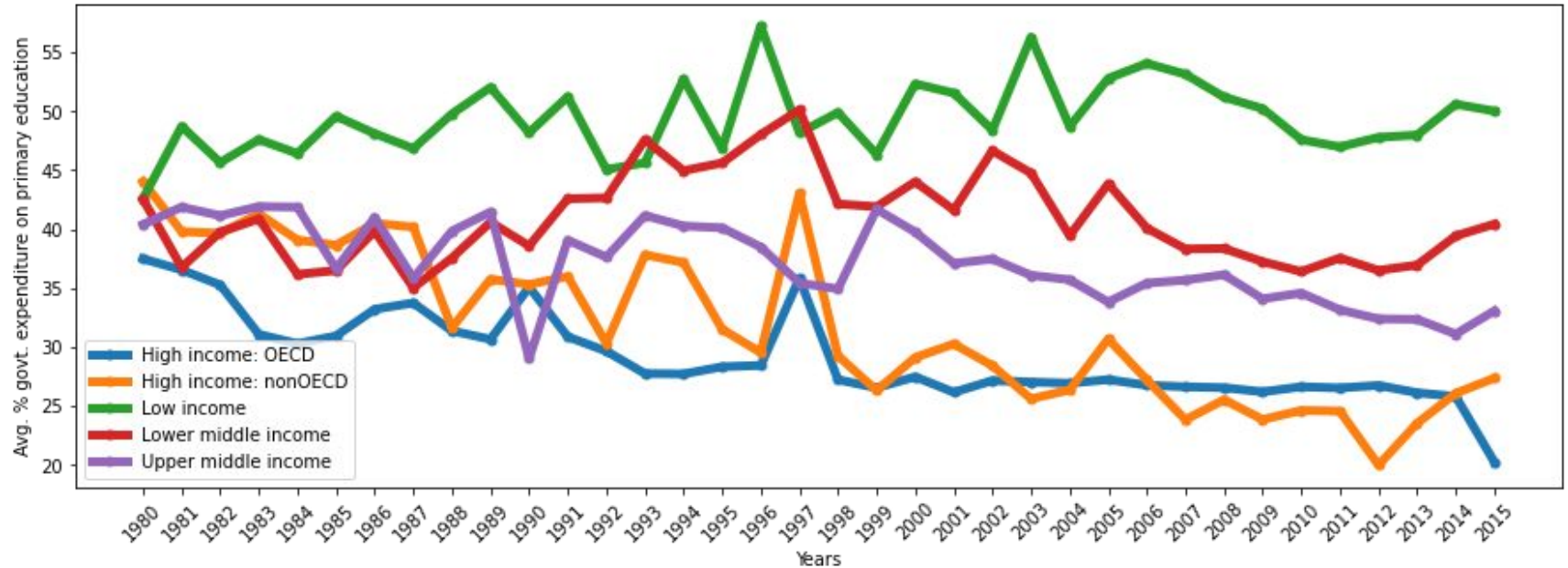
# Expenditure on Education (US)



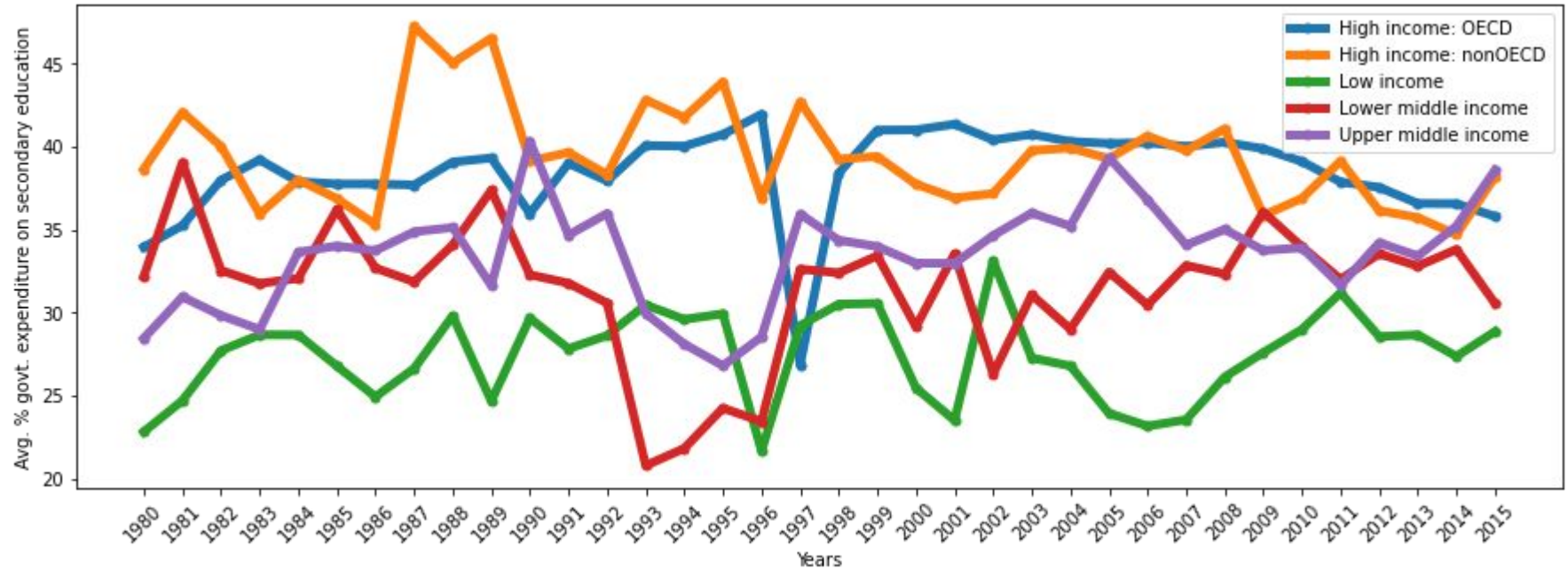
# Average Expenditure Stats in Education



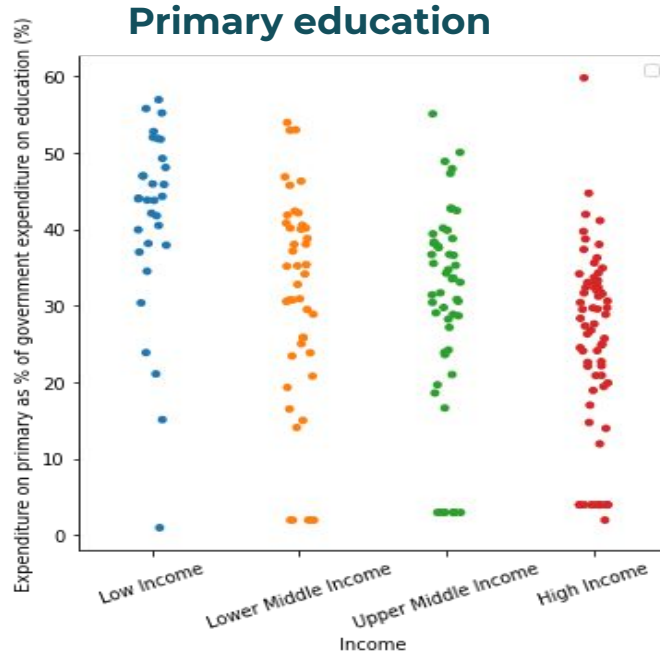
# Average Expenditure Stats in Primary Education



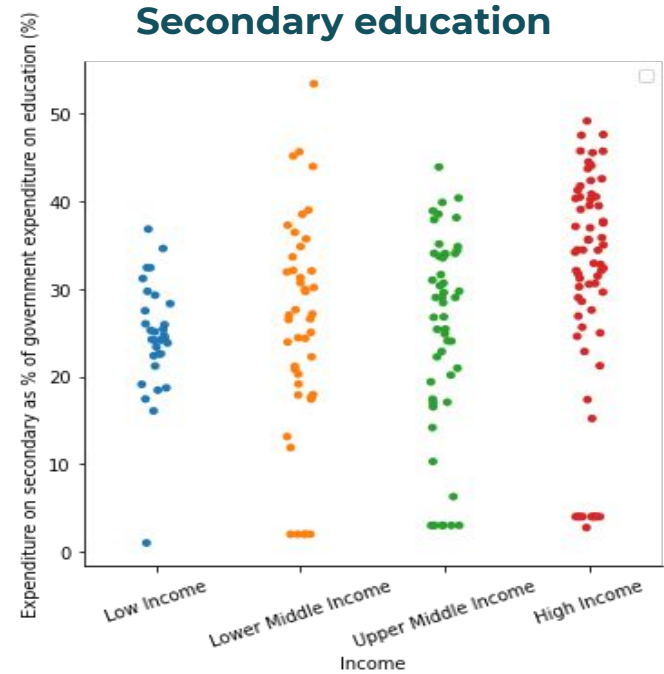
# Average Expenditure Stats in Secondary Education



# Expenditure vs Income



**Correlation : -0.36**

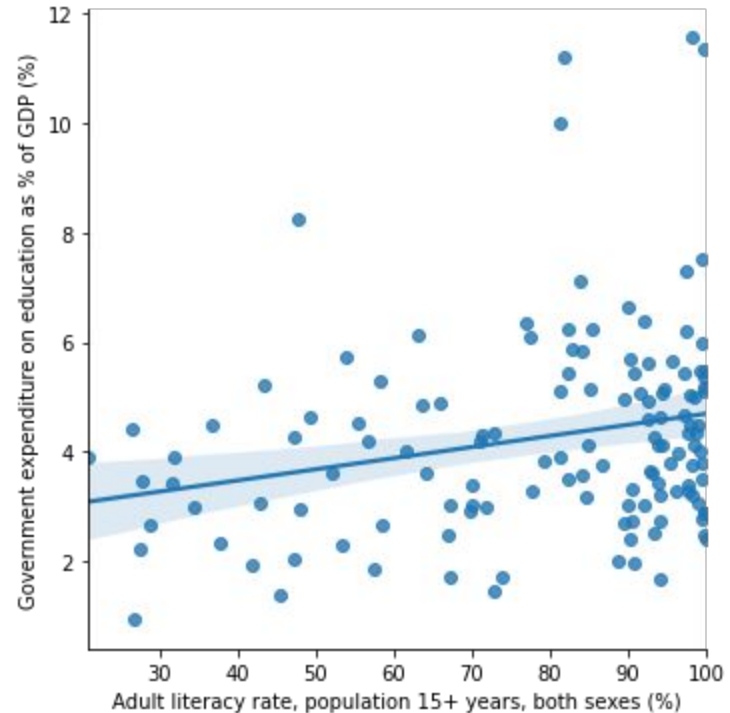


**Correlation : 0.15**



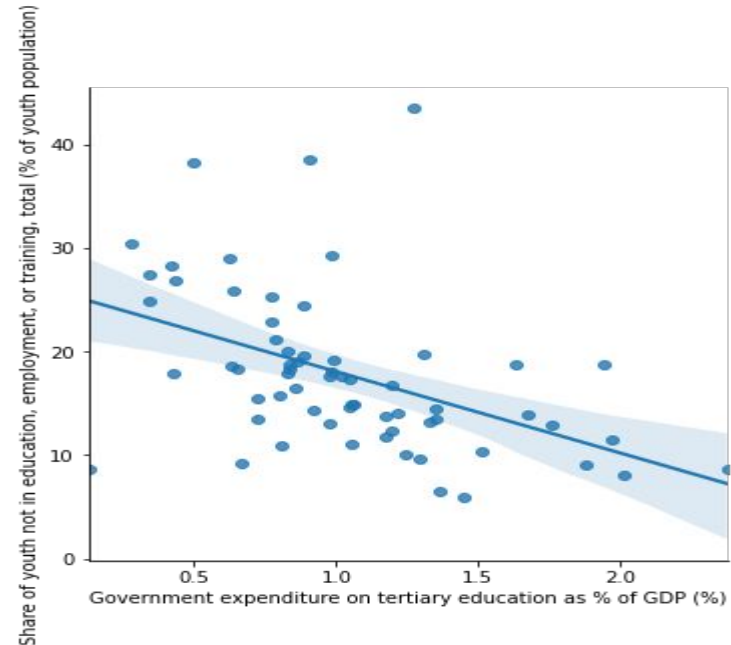
# Expenditure vs Adult Literacy rate

Expenditure on education as % of GDP vs Adult Literacy rate  
**correlation : 0.24**



# Share of youth not in education, employment or training vs Expenditure in tertiary education

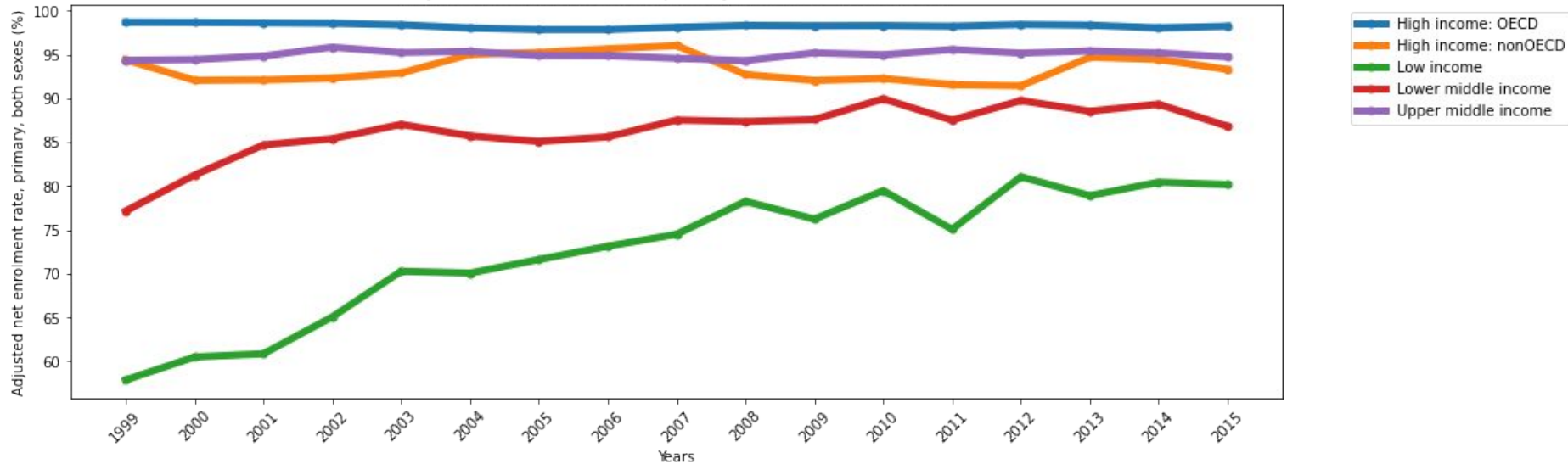
Share of youth not in education, employment or training vs Expenditure in tertiary education  
**correlation : -0.46**



# Variation of Net Enrolment rate in primary schools across different group of countries

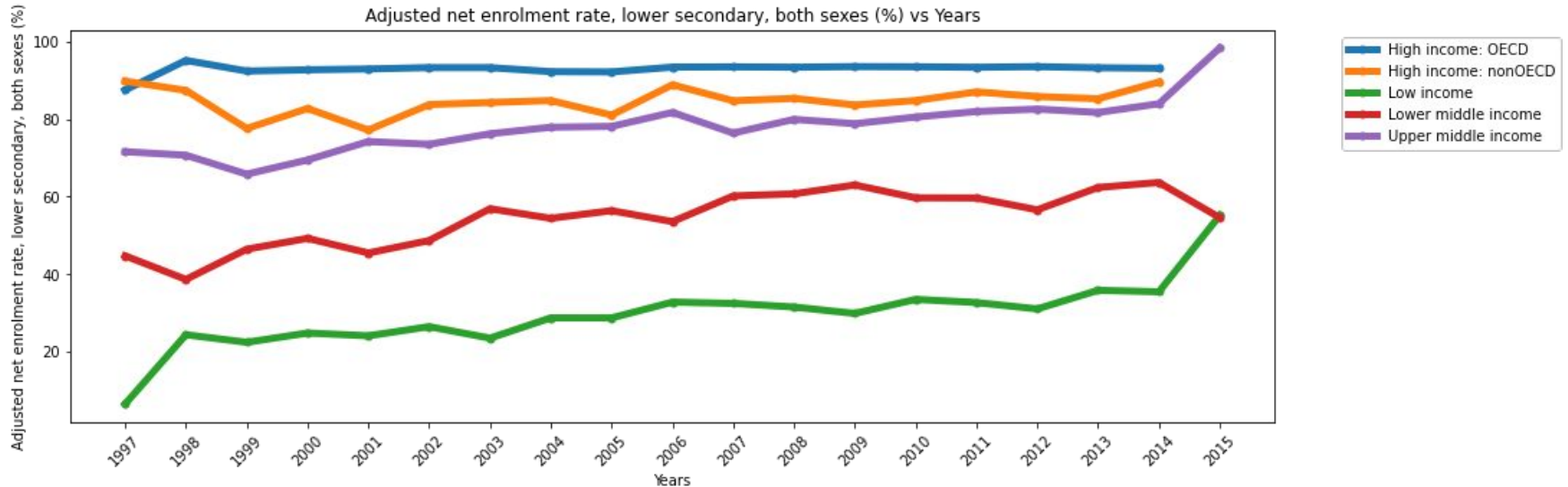
The net enrolment rate in primary school level of lower income group countries have increased over a period of time from the year 2003.

Adjusted net enrolment rate, primary, both sexes (%) vs Years



# Variation of Net Enrolment rate in lower secondary schools

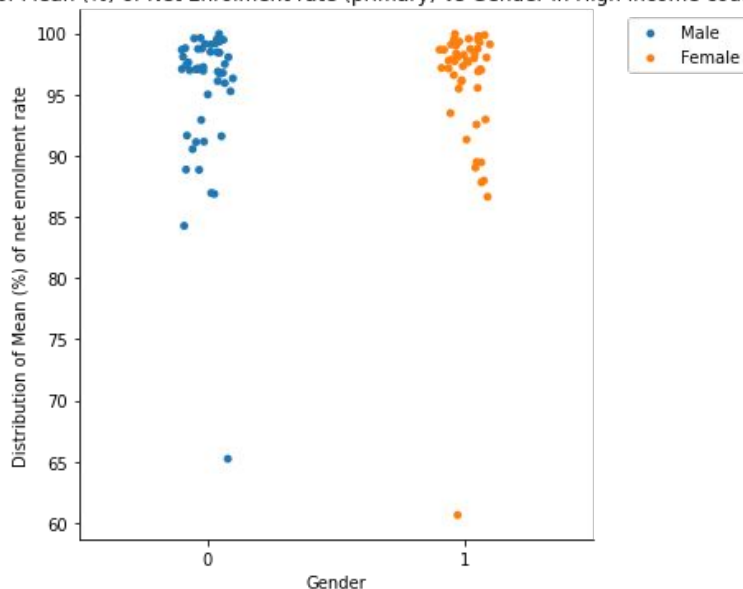
Net enrolment rate into lower secondary school level for low income countries is low at the start and slowly increased over a time period.



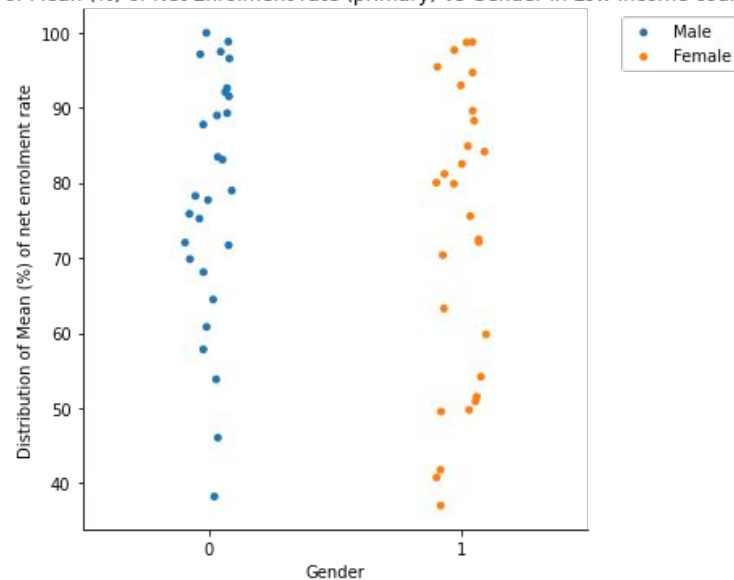
# Distribution of Mean(%) of Net Enrolment in primary schools vs Gender

For low income countries, there is a negative correlation between (%) of youth enrolled in schools and their gender.  
Correlation coefficient is -0.162

Distribution of Mean (%) of Net Enrolment rate (primary) vs Gender in High income countries



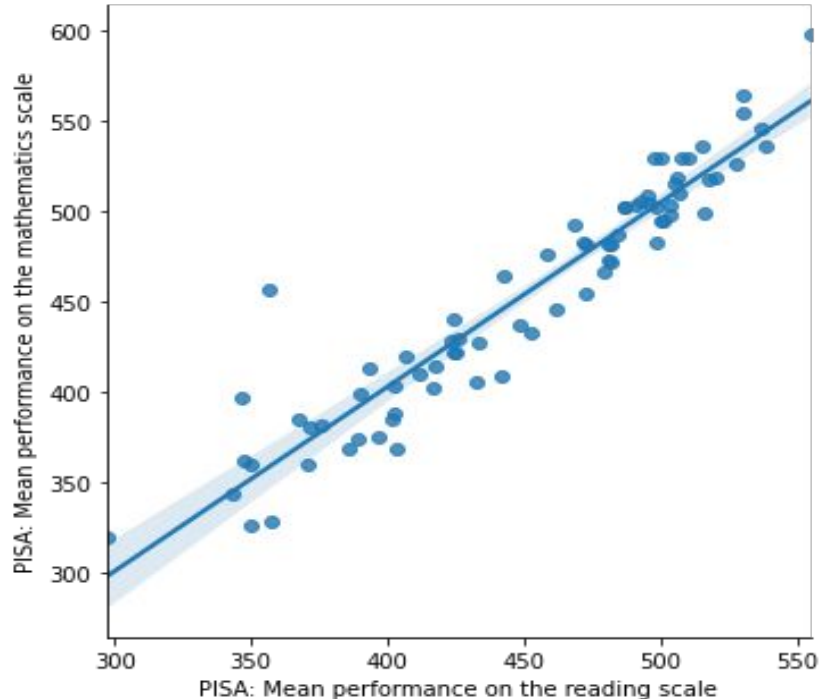
Distribution of Mean (%) of Net Enrolment rate (primary) vs Gender in Low income countries



# Correlation between PISA : Mean Performance on reading and science scales

**Correlation coefficient : 0.949**

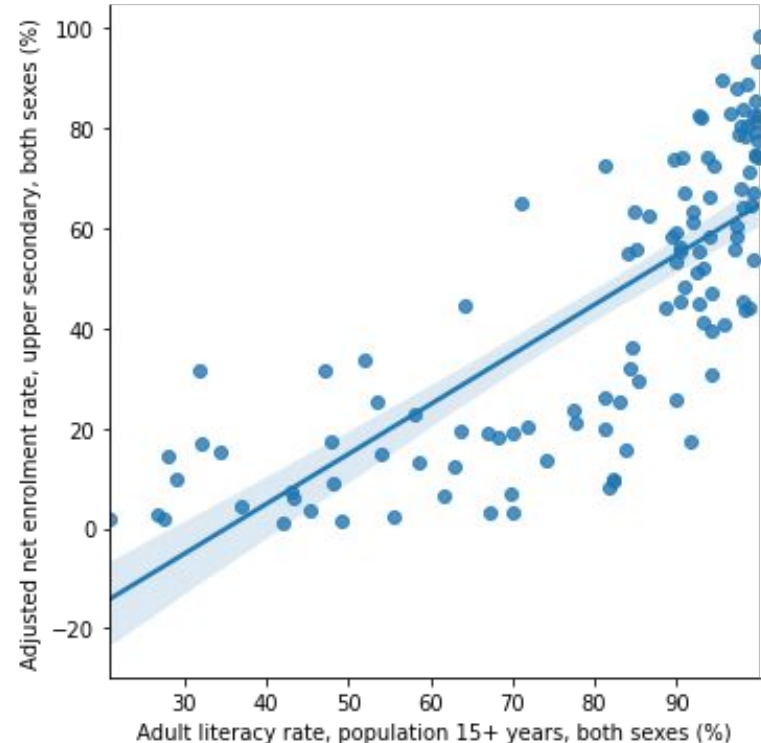
PISA : Mean performance on reading scale  
Is highly correlated to PISA : Mean  
Performance on science scale.



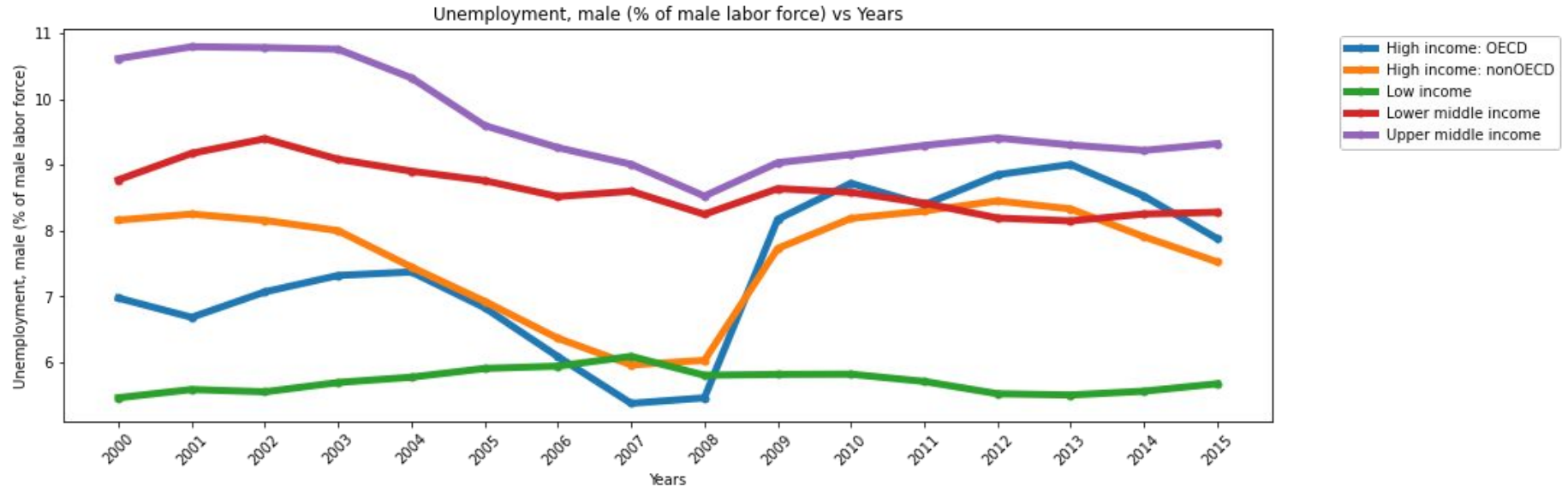
# Correlation between Adult literacy rate and Upper secondary Net enrolment rate

**Correlation coefficient : 0.768**

Adult literacy rate is highly positively correlated to Net enrolment in Upper secondary schools.

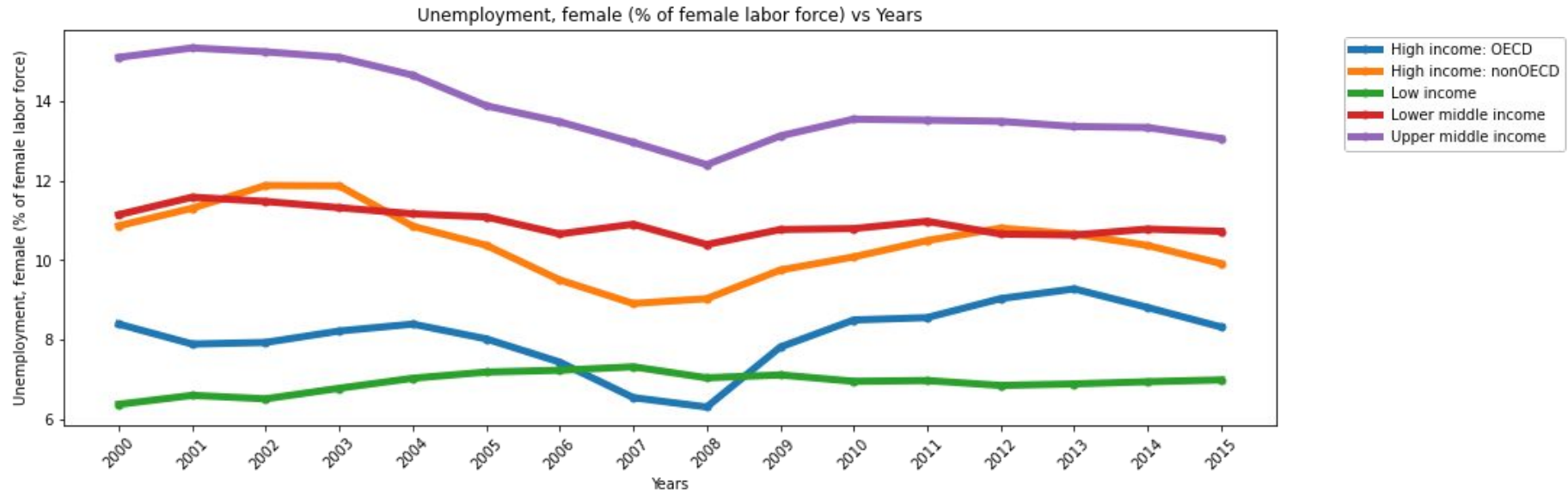


# Unemployment (%) in Male labor force across years

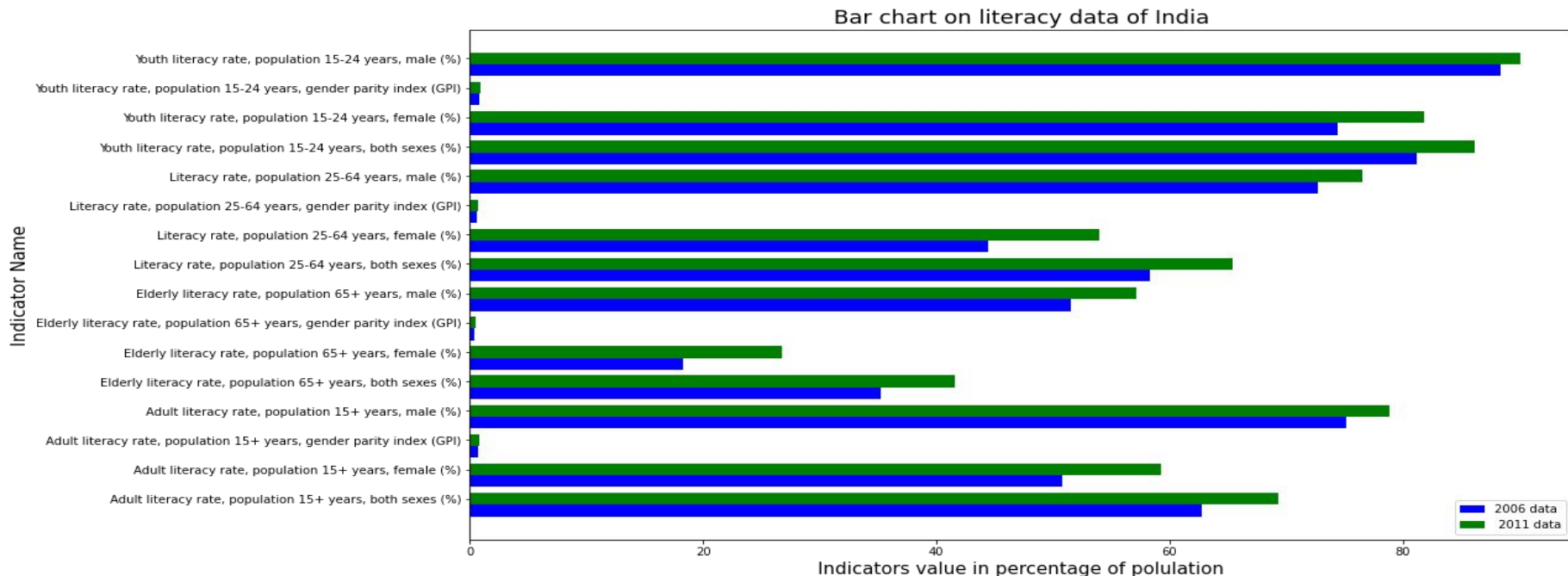




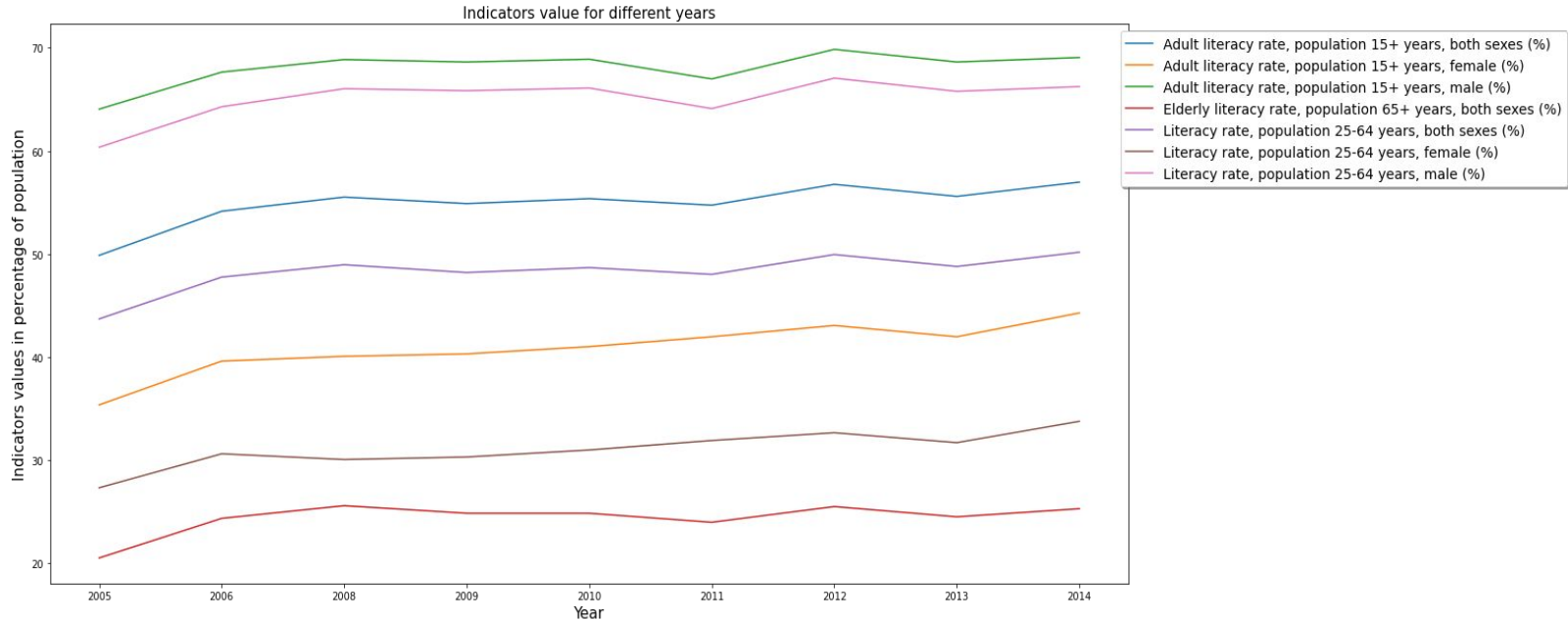
# Unemployment (%) in Female labor force across years



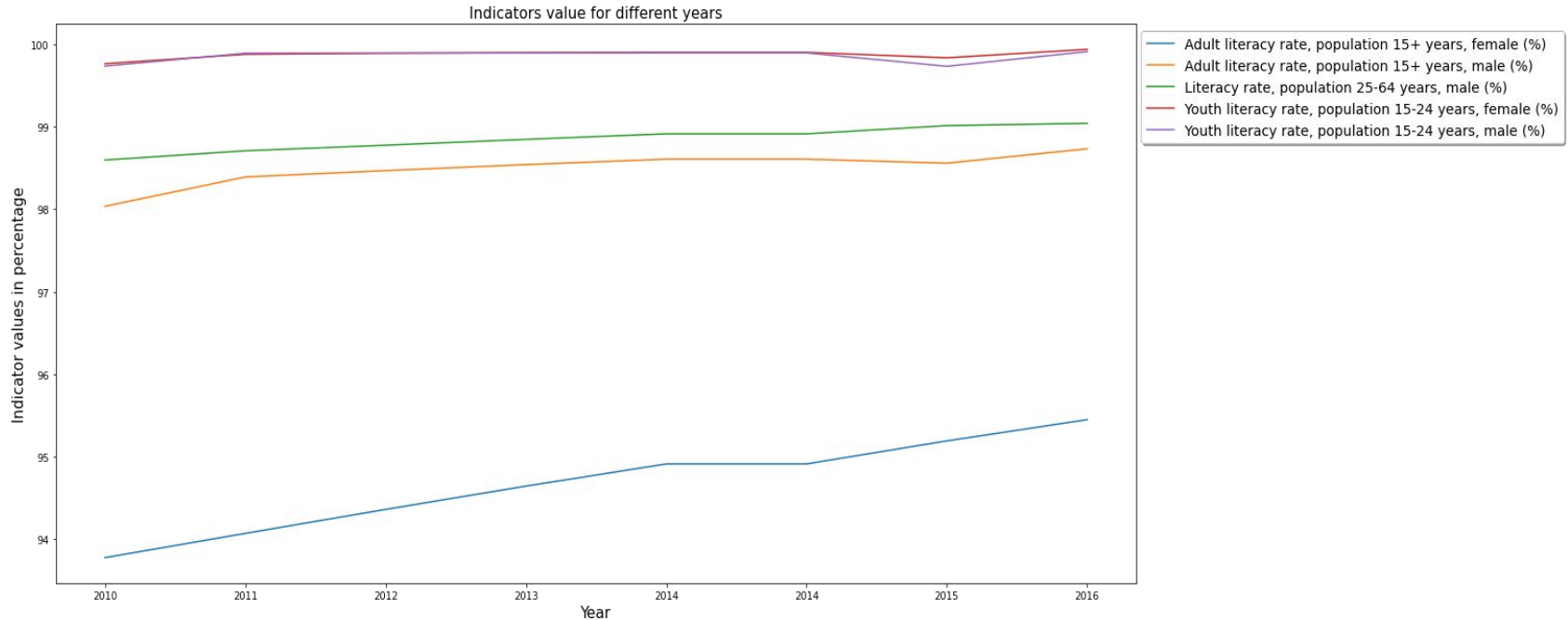
# Comparison between 2006 and 2011 literacy data for India



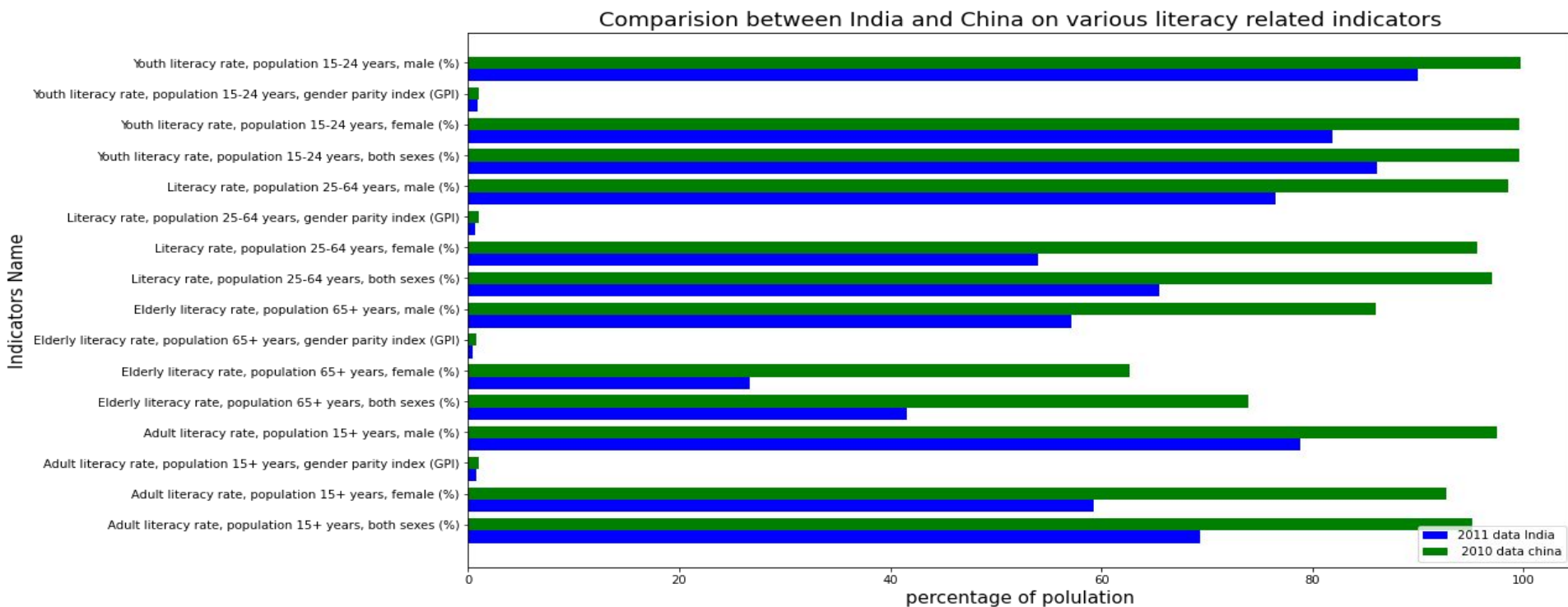
# Variation of various literacy indicators for Pakistan(lower medium group)



# Variation of various literacy Indicators for Singapore(Higher Income group)

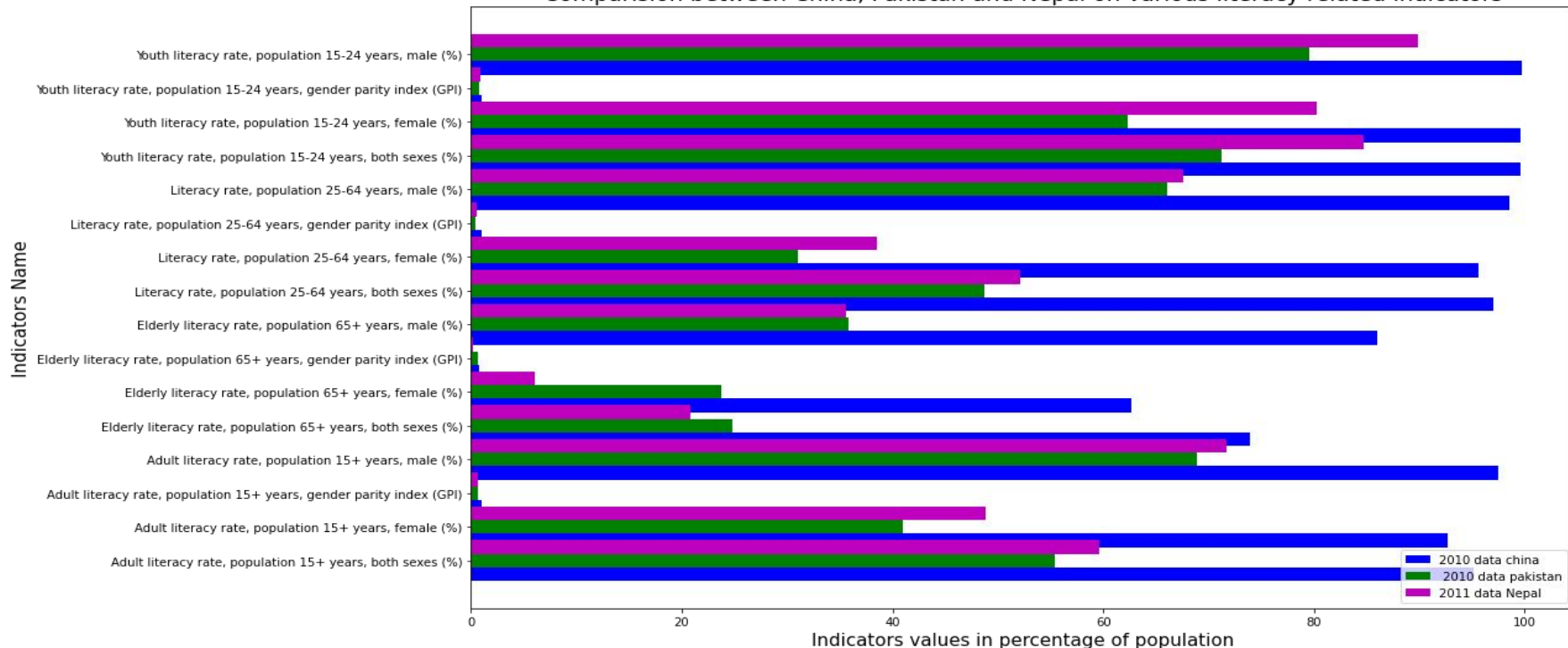


# Comparison between India and China on the basis of various literacy indicators



# Comparison(literacy) between China, Pakistan and Nepal

Comparison between China, Pakistan and Nepal on various literacy related indicators



# Challenges

- Size of the dataset 'EdStatsData.csv' is huge and took time to understand the relation between other datasets.
- The Education Statistics data was distributed across five csv files and has many missing values, extra details and incorrect formatting. Number of Nan values are quite high (more than 50%) in each row of 'EdStatsData.csv' dataset.
- There is an an important gender limitation in this the dataset since it assumes binary gender.
- The projections in the data are absent in many cases and those that are present may not be accurate.
- Literacy data for many of countries are missing which restricted us to do analysis on a particular literacy indicator for a group of countries

# Conclusion

- The Lower income group countries have the maximum population growth rate over the years compared to higher income group countries.
- As the Population growth rate increases, the adult literacy rate decreases and employment also decreases.
- With the increase in internet users, the Unemployment is getting decreased and the adult literacy rate is getting increased.
- The Low income countries tend to spend higher percentage expenditures on education on primary education compared to high income ones. While secondary and tertiary education get more focus from higher income countries than the lower ones.
- The countries with high literacy rate tend to spend higher on education than the ones with low literacy.
- Higher tertiary education expenses percentage ensures lesser uneducated and unemployed youth.



# Conclusion

- Net enrolment rate in primary and secondary schools is higher in high income group countries when compared to medium and low income group countries.
- Higher percentage of men are enrolled into primary schools than women in low income group countries which is in contrast to high income group countries.
- Overall unemployment in high income countries had taken a huge leap because of the great recession between years 2008 and 2010.
- Literacy rate for low income group are quite low but at the same time Nepal is doing really great than some lower medium countries also which is impressive.
- For lower medium and low income group female literacy is lower as compared to male literacy high and medium which is the reflection of their male dominant society.
- For countries like Singapore have very high literacy rate.

# Q & A