HDI as a performance measure for countries: Cross Country and Regional Explorations

Submitted By:

Mezbah Uddin
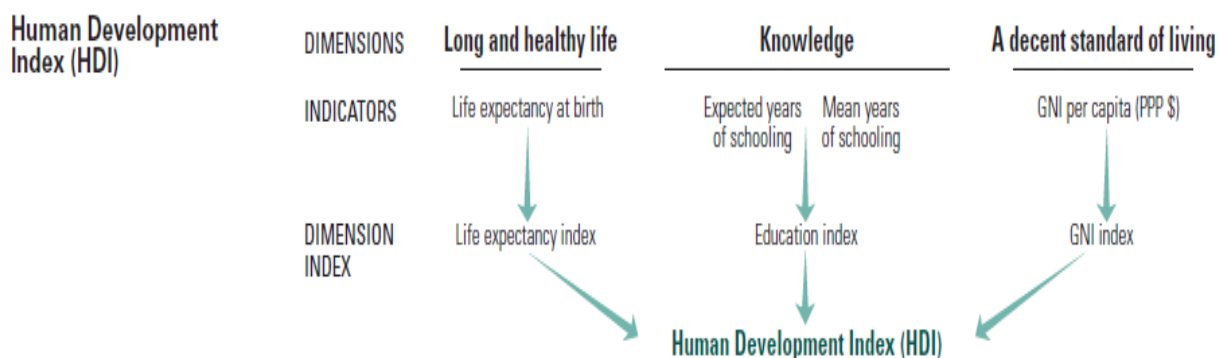ME 8135

Description of the Problem

Understanding what causes a country to grow economically and what criteria be used to judge a country's economic standing has been a central question in economics & in social sciences for the last 70-80 years.

Specially to investigate reasons of growth and how lessons learned from one country can be applied to other country, a subfield in economics, called development economics was born and still exists.

Different income/production based indicators like Gross domestic product, trade, gross national income etc. still been used to assess a country's wellbeing. However, criticism have been raised as to how accurate these criteria are in assessing the true economic and social development of the residents.

HDI (human development index) was developed as a pursuit of building a composite index that can give a better representation of a holistic scenario. HDI thus incorporates three key metric, namely, long & healthy life, knowledge and decent standard of living into one index. Final value of the index is the geometric mean of the normalized indices for each of the three dimensions.

The health dimension is assessed by life expectancy at birth, the education dimension is measured by mean of years of schooling for adults aged 25 years and more and expected years of schooling for children of school entering age. The standard of living dimension is measured by gross national income per capita. The HDI uses the logarithm of income, to reflect the diminishing importance of income with increasing GNI. The scores for the three HDI dimension indices are then aggregated into a composite index using geometric mean. Following picture gives a better representation –[1]

Human Development Index (HDI)

DIMENSIONS	Long and healthy life	Knowledge		A decent standard of living
INDICATORS	Life expectancy at birth	Expected years of schooling	Mean years of schooling	GNI per capita (PPP $)
DIMENSION INDEX	Life expectancy index	Education index		GNI index

Human Development Index (HDI)

---

[1] http://hdr.undp.org/en/content/human-development-index-hdi

For the purpose of this paper, I want to investigate HDI index data and HDI ranking for 188 countries between the time period of 1990 to 2014 for 7 regions to see time series trends of the countries and regions. Specially, it would be interesting to see, if there is a catch-up phenomenon, meaning, whether there are countries that started with very low HDI and reached very high index in a short period of time. Also, I want to visually investigate how regions are performing in comparison to each other. Also, from the data I want to investigate whether clustering possible within countries, that is, countries with similar HDI can be grouped together.

In addition, graphical investigation of mean years of education and HDI within gender will also be explored to see gender biases and disparity from 2015 data.

Furthermore, it would also be important to see correlation between labor force participation for male & female vs gender inequality index from 2015 data.

Data Source

Source of the data is the human development report 2015, produced by UNDP. Extracted datasets from the report were posted as csv files in the Kaggle website. Please see the link in the footnote and in the reference for the details [2] [3].
Dataset included following csv files –
    (a) gender_development.csv
    (b) gender_inequality.csv
    (c) historical_index.csv
    (d) human_development.csv
    (e) inequality_adjusted.csv
    (f) multidimensional_poverty.csv
and a index calculation pdf file.

---

[2] HD report 2015, UNDP, http://hdr.undp.org/en/content/human-development-index-hdi

[3] Kaggle: https://www.kaggle.com/undp/human-development

Technologies Used and Issue Faced

For the purposes of cleaning, preprocessing and visualization, R statistical software were used. The following packages of R, used for data munging –
- (a) dplyr
- (b) cluster
- (c) varhandle

and for visualization –
- (a) ggplot
- (b) plotly
- (c) googleVis

Faced a lot of problem while converting one of the columns into factor form. Whenever, base r function as.numeric were used, data which were in fraction became large whole numbers. Varhandle package came in rescue for the above issue. Using the aforementioned package, columns were first unfactored and then converted into numeric.
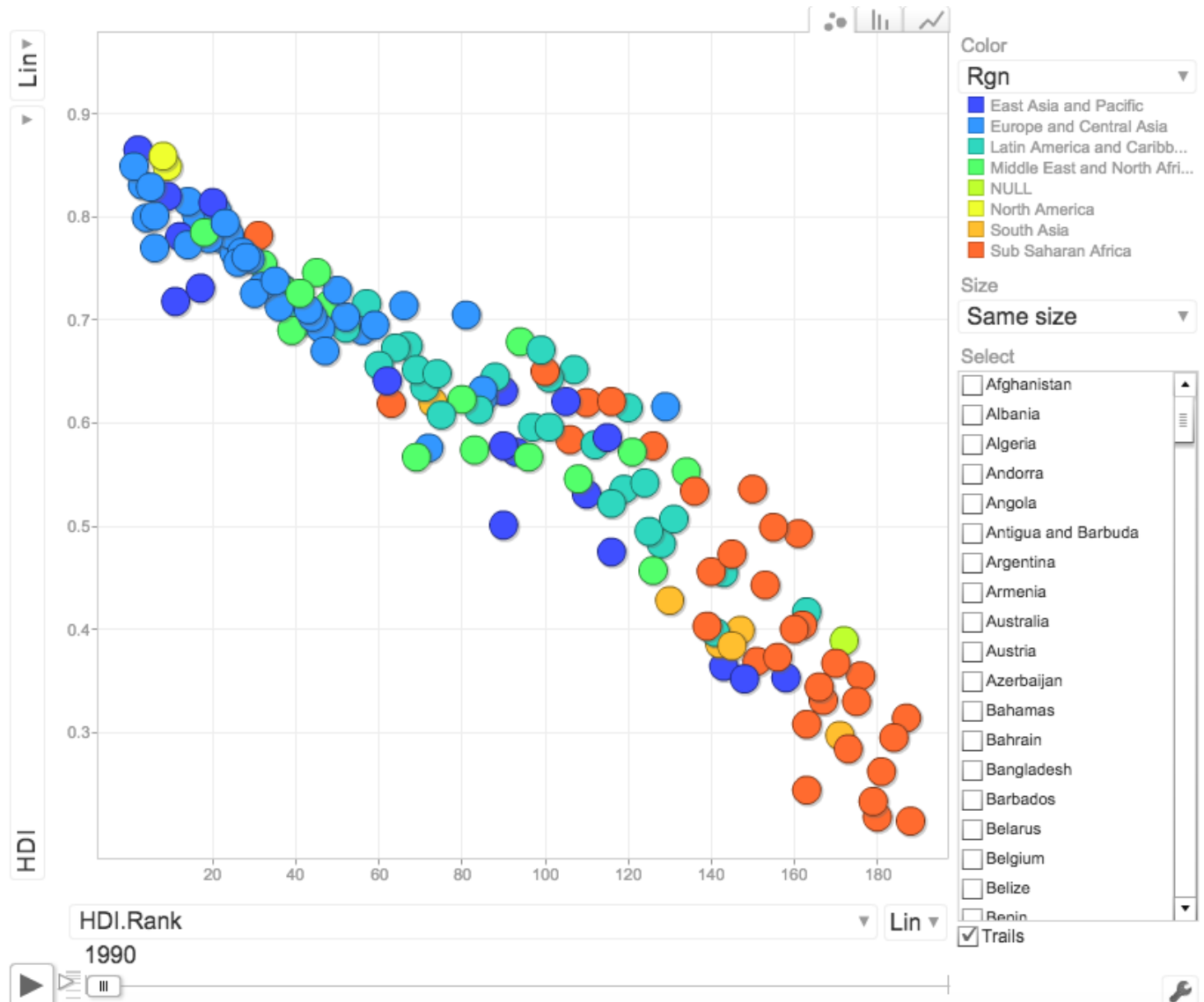
Other issue was, after applying a custom made function to add region based on countries, region column in the dataframe became a list. A list class column in r can't be explicitly converted into factor or numeric. To solve the problem, I first had to unlist using base function, unfactor using varhandle and then factor again using base functions.

Preprocessing of the data

Dataframe were created in r using the csv's from HD report. Dataframe only had Country name, however, for better explaining the data, region columns were created using a custom made function in r. To decide, which country falls into what region, world bank classification have been used. The reference has the region definitions[4]. Also, in the historical_index file which had HDI data over the span of 1990-2014, values were in coumn form. Therefore, dataset had to transposed using functions and some lines of coding. The R script will be appended with the report to reproduce transformed dataframes.

---

[4] https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Visualization 1 : HDI pattern overtime for all the Countries

Visualization is a dynamic bubble chart that shows year by year movement of the country's HDI and rank. Also, color have been used for different region so that one can identify region and also the country by clicking on each bubble. Since, to run this visualization, flash player is needed, a screenshot has been shown in the report. However, html file will be attached with the report to reproduce the motion picture.

From the picture it can be seen that, higher HDI index countries lies on the upper left of the picture. These countries included European, north America and some Asian countries as well. One can also choose countries from the drop down menu to see how that country evoves in between the periods. Also trail will show the path of increase/decrease in HDI. For example, if Bangladesh is chosen for comparison, after the motion, following visualization is produced.
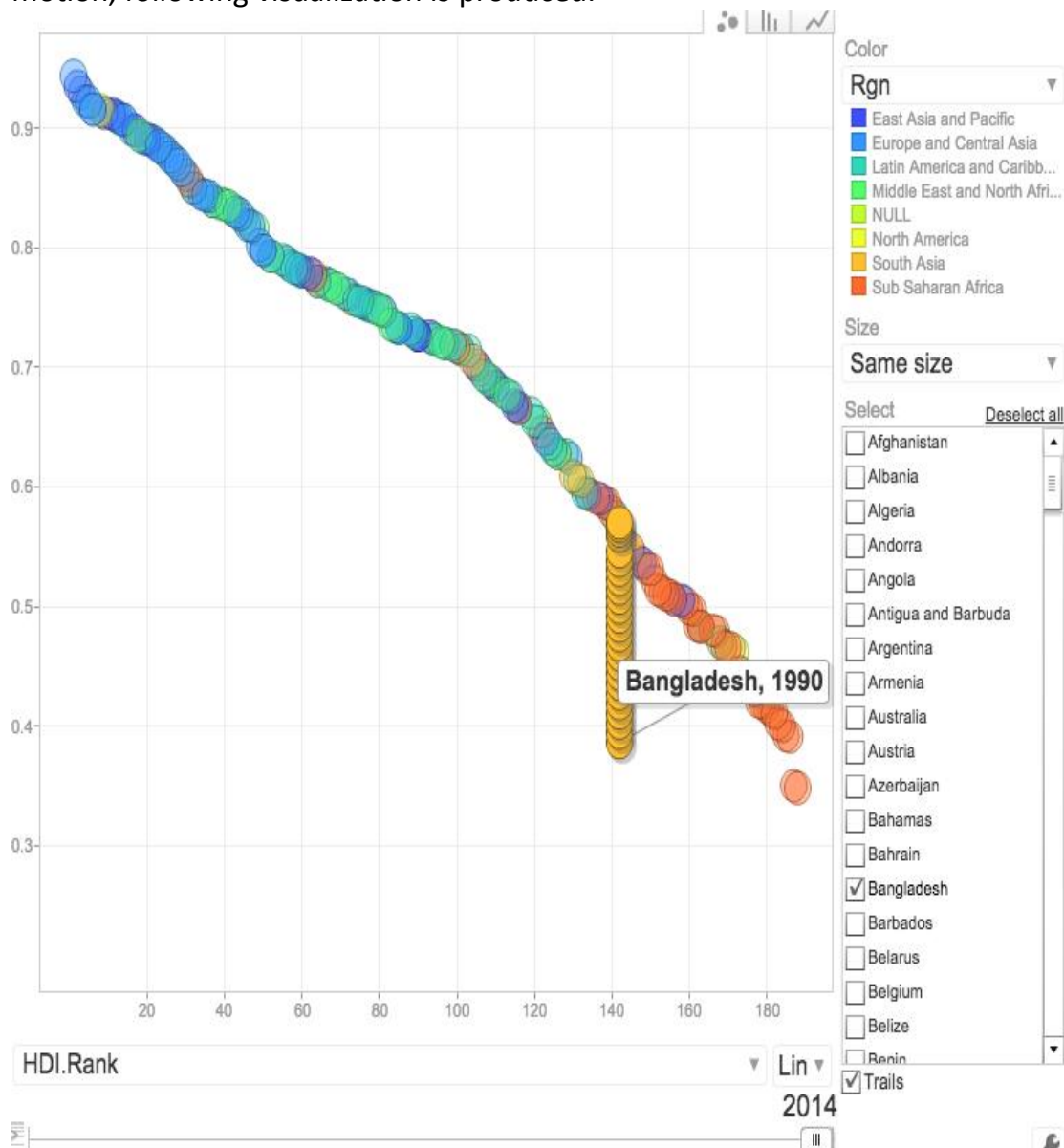
From the graph, it can be seen that, overall, there is a positive trend in most countries in Europe, asia, latin America and middle east. There is sluggish improvement in African countries, with some countries having negative growth in HDI. Also, within asia, huge fall of HDI for Afghanistan can be seen from the data.

## Visualization 2 : HDI pattern overtime for the regions

If one is interested in tracking trails of HDI and rank movement for regions, this visualization will be useful. This visualization can be used as a complement to visualization 1.
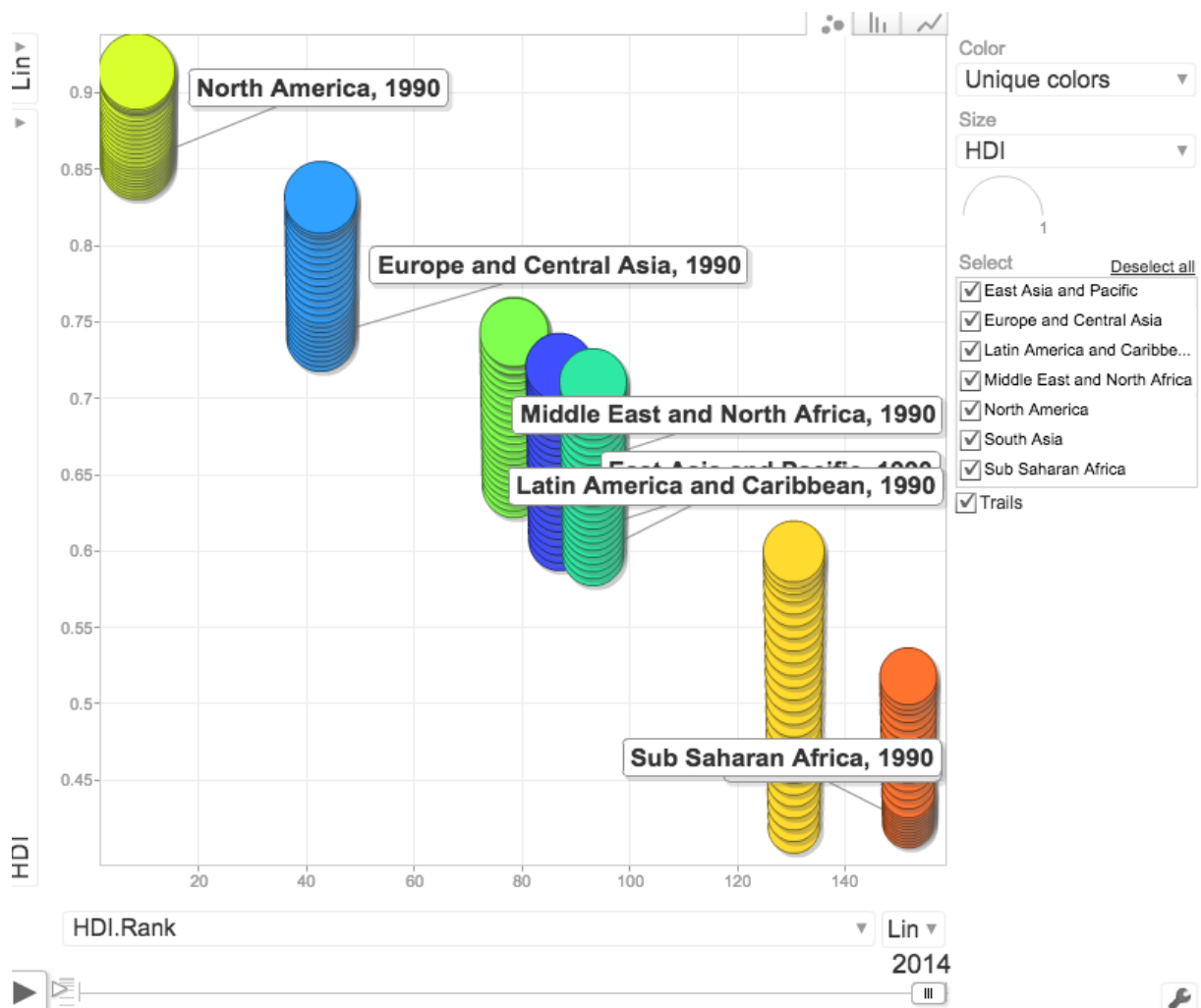
From the graph, it can be seen that highest growth in HDI, which can be measured by length of the trail for each group, is achieved by south Asia, followed by middle east and north Africa.

Visualization 3 : HDI pattern overtime for the countries of a specific region

This visualization will show how countries within a specific region is performing with respect to each other. For the below visualizations, Europe and central Asian countries have been plotted
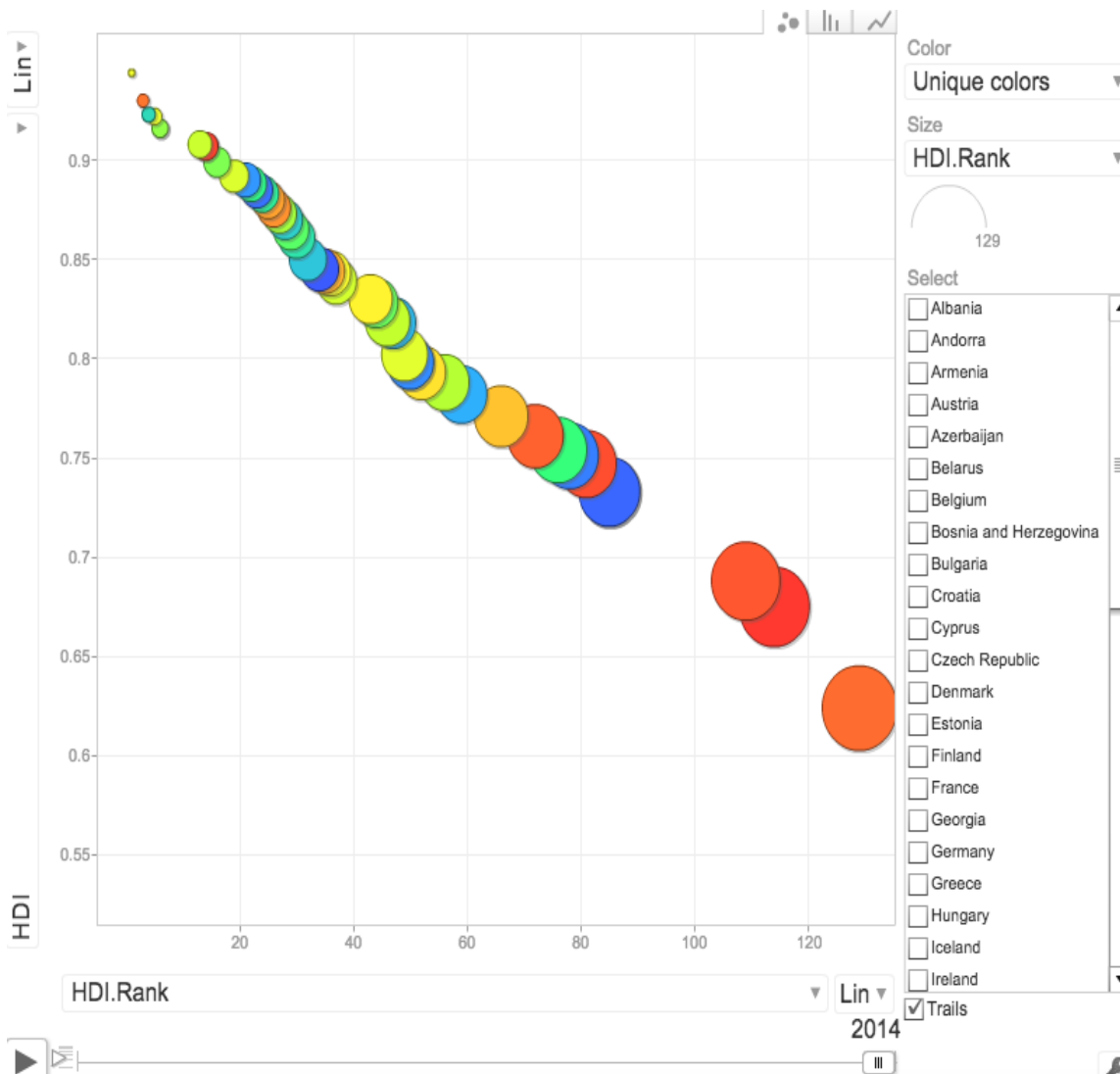
It is interesting to see from the graph that initially there is a large fall in HDI for tajkistan and after that there is a huge improvement after year 2000.

Visualization 4: Is Clustering possible to find similar group of countries based on HDI
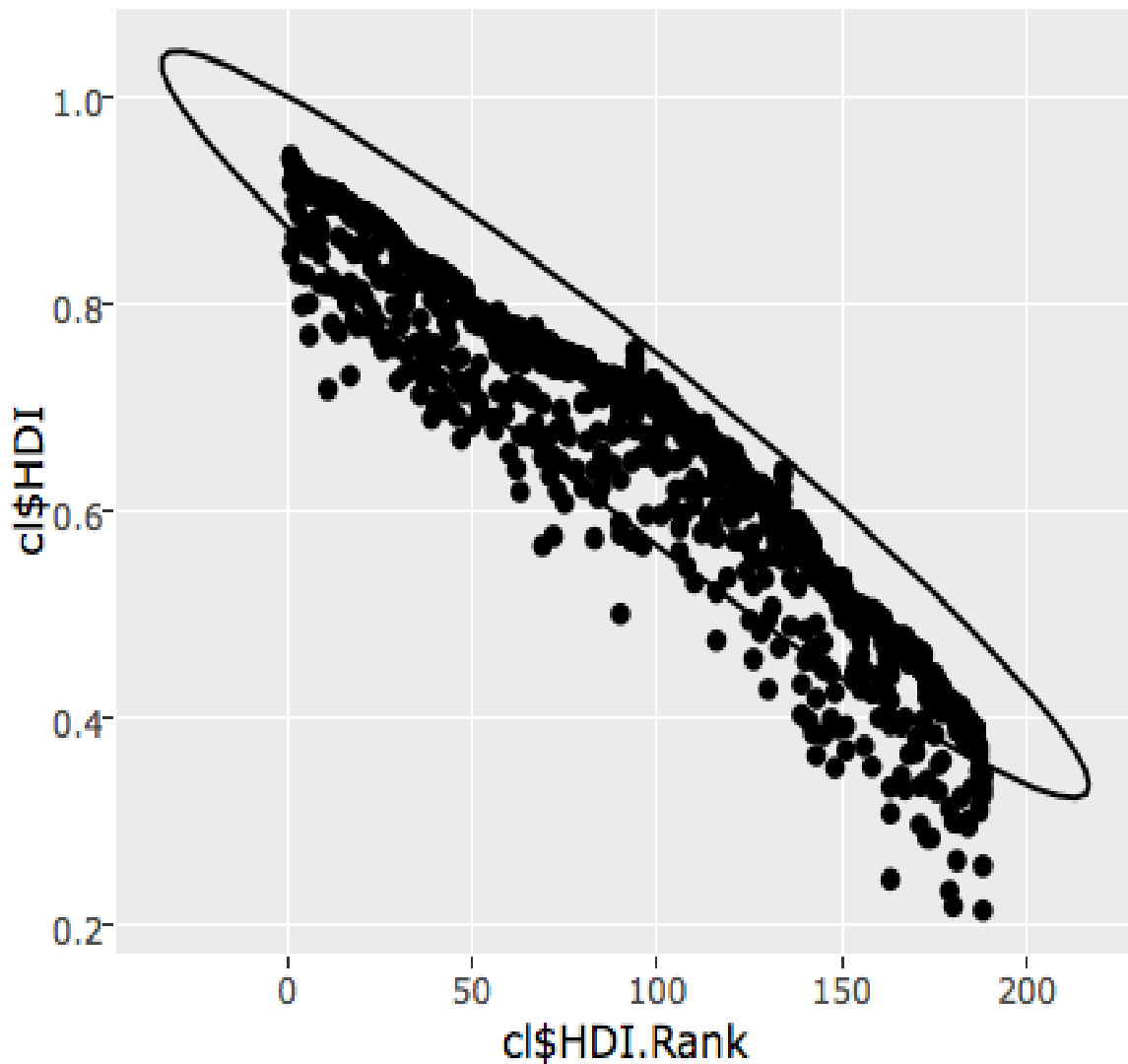
The purpose of this image is to find out whether clustering among countries is possible or not. I was basically interested to see whether countries can be grouped in high or low classifications.

It can be seen from the graph that, it will very difficult visually to cluster countries as they are all very close to one another and separation will not be useful as much.

Visualization 5: Education VS HDI Vs Income for Countries for Male Population
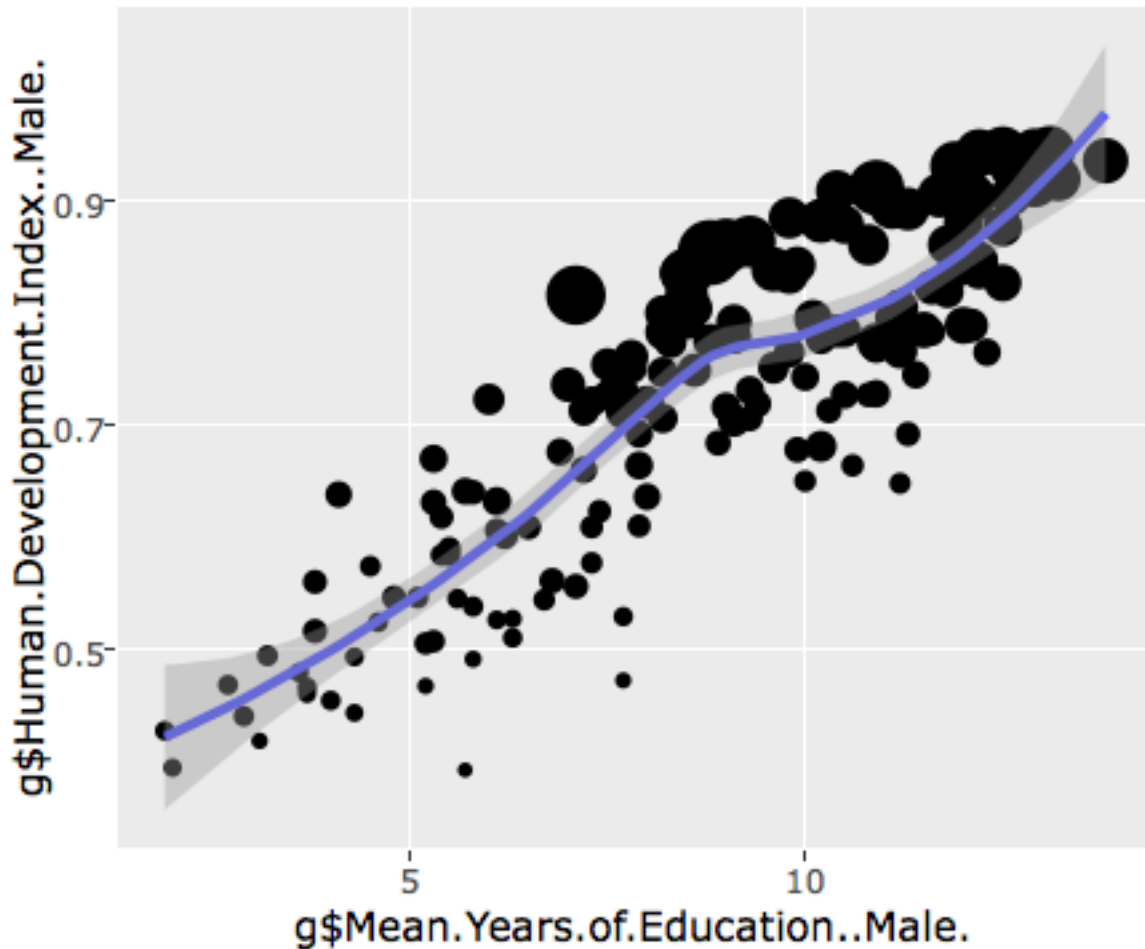
Above visualization shows the relationship between education and HDI index. The size of the bubble is determined by gross national income per capita (male). From the smoothed line it is seen that there is a positive correlation between education and hdi. Also, interesting to note that, countries with higher education & HDI, also has very income. At the upper right area of the graph are mostly European and north American countries.

Countries that have low male education and moderate HDI but high per capita income are Arab countries like Bahrain and Saudi Arabia.

Visualization 6: Education VS HDI Vs Income for Regions for Male Population

This visualization gives a regional dimension to the relationship between education and HDI for male population. Countries with high hdi & education years are mostly European countries. Among the east Asian countries, we have Australia, New Zealand, South Korea, Japan & Singapore. Only Arab country in the upper right quadrant is Brunei.

Lower bottom countries are mostly African and some latin America & Caribbean countries.

Visualization 7: Education VS HDI Vs Income for Countries for Female Population
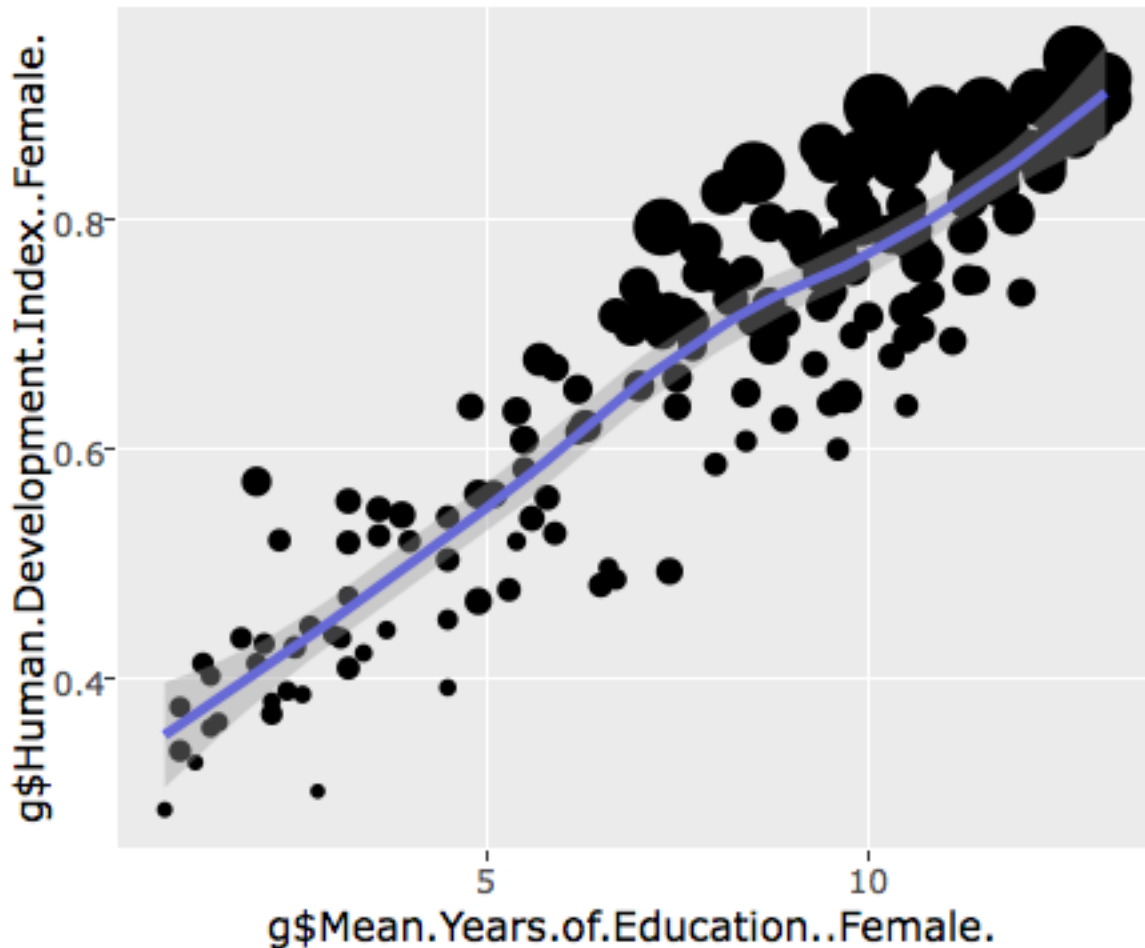
Above visualization shows the relationship between female education and HDI. In addition, size of the bubble represents per capita income. As expected a positive relationship is found. Higher Education and HDI (top right quadrant) is composed of European countries like Norway, Sweden, Switzerland, Canada. And the bottom quadrant (bottom left) is composed of mostly African countries. Countries that have higher income, moderate HDI, but low female education are arab states like Kuwait, oman and Saudi Arabia

Visualization 8: LFP VS GII Vs Income (female) for Countries
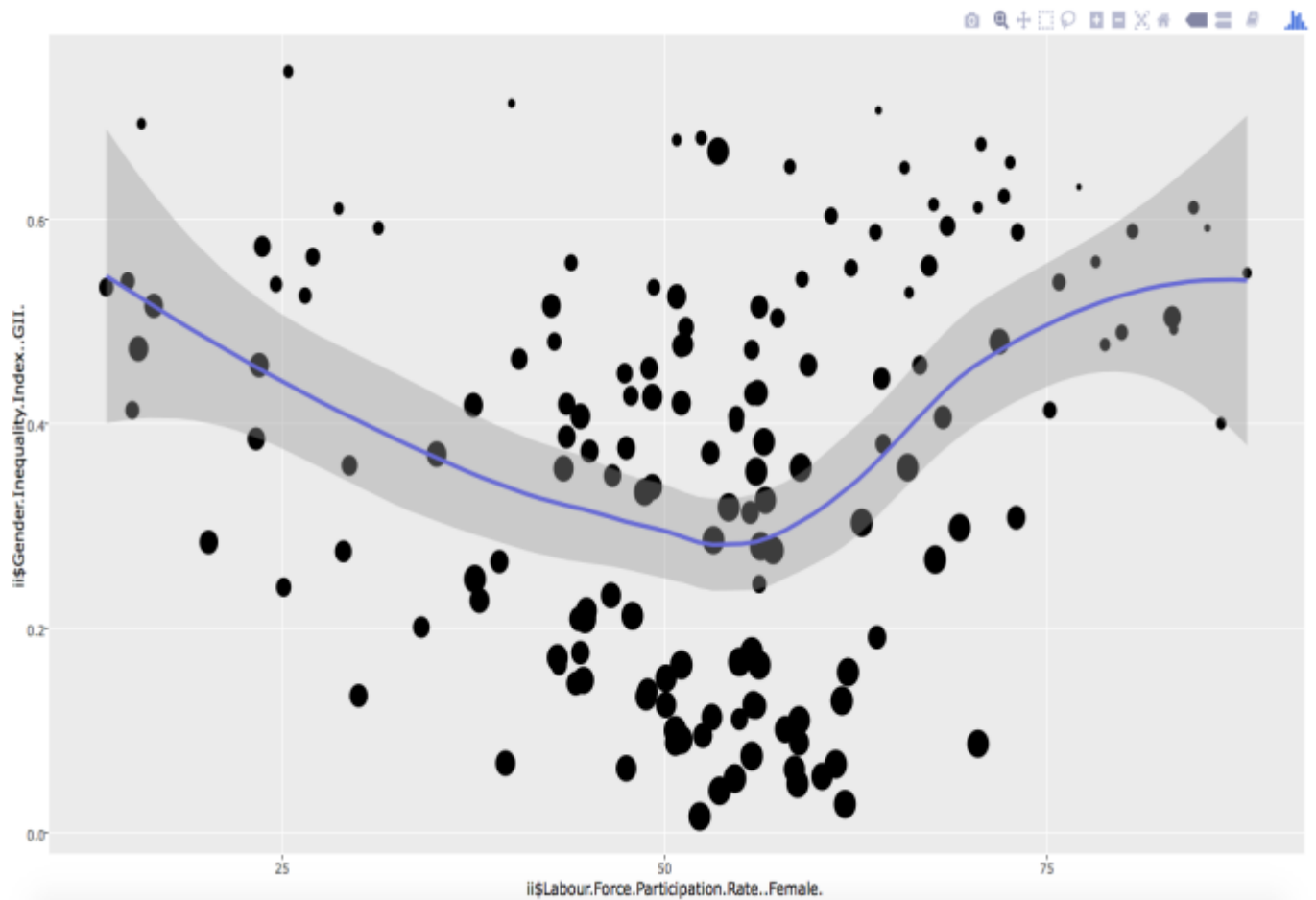
Above visualization shows the relationship between female education and gender inequality index. Again, size of the bubble is represented by income.
This graph portrays some interesting and important relationship.
Therefore, good performing countries will be the ones that has lower gender inequality.

Above 50% labor force participation with low gender inequality and higher income countries are European and north American countries.
Above 50% labor force participation, high gender inequality, and high income is Tonga.
Above 50% labor force participation, moderate gender inequality are some south American, Arab states, and some Asian countries.

Very low labor force participation, high gender inequality and moderate income countries are Jordan, Syria, Iran & Egypt.

Conclusions and Lesson Learned

- ✓ Overall there is not much catch up in HDI between countries growth in HDI is moderate among most countries. Some south east Asian countries like Bangladesh, India and some Arab states shown good progress. However, these countries still have long way to go.

- ✓ As a region growth in HDI, South Asia have shown most progress. It's understandable as, if a country already has high income and HDI, it's difficult to show much growth.

- ✓ Just Based on HDI, it's difficult to cluster countries. With other dimension like income, education, gender inequality countries can be clustered by subdividing factors.

- ✓ By incorporating education and income, we see that European and north American countries have index across all factors. On the other hand, African countries have low index across all criteria.

- ✓ Lastly, there is a very complex relationship between female workforce participation, gender inequality index and income. High labor force participation has both high and low gender inequality. Also, there are countries that have low labor force participation, high inequality and moderate income.