Study Design

The raw datasets that we collected is as follows:

- **11** .csv files
- **11** .xlsx files
- **2** .json files
- **6** web links with relevant tables to be scraped

The data was retrieved from reliable sources and/or retained after verification from multiple sources (links to original data sources provided in a separate file). The .csv files and .xlsx files contained yearly time series data of one variable each from 2000 to 2019 for 281 countries or groups of countries. The .json files and tables scraped from multiple websites contained (recent) data from specific year of one or more variables for multiple countries.

Code Book

Tidy dataset was prepared by extracting data for all available countries for one particular *year* (passed as an optional parameter to instruction list, with default being 2018) from the .csv and .xlsx files, and merging them with data from .json files and webscraped tables. Countries (or *keys*) common in all raw datasets were retained in tidy dataset while others dropped. Tidy dataset came out to be a *121x40* table, with each row corresponding to a unique observation (country) and each column corresponding to a unique variable. A row was included at the top of file with interpretable variable names.

Information about each **variable** in the dataset is provided below. *Unit* is mentioned in parenthesis, where applicable. Range of values that the variable takes is also provided where relevant.

1. **Country** (*categorical*)
   Name of country
2. **Year** $\in \mathbb{Z}^+$
   Year data mostly corresponds to
3. **Literacy_Rate** $\in [0, 100]$
   Percentage of the population aged 15 years and over who can both read and write with understanding a short simple statement on his/her everyday life
4. **Mean_Years_Of_Schooling** (*years*) $\in \mathbb{R}^+$
   Number of years of schooling a citizen of particular country receives on average
5. **Population_By_Primary_Education_Completed** $\in [0, 100]$
   Percentage of population that has completed Primary education
6. **Population_By_Lower_Secondary_Education_Completed** $\in [0, 100]$
   Percentage of population that has completed a Lower Secondary education
7. **Population_By_Upper_Secondary_Education_Completed** $\in [0, 100]$
   Percentage of population that has completed Upper Secondary education

8. **Population_By_Post_Secondary_Education_Completed** $\in [0, 100]$
   Percentage of population that has completed Post Secondary education
9. **Population_By_Tertiary_Education_Completed** $\in [0, 100]$
   Percentage of population that has completed Tertiary education
10. **Population_By_Bachelors_Completed** $\in [0, 100]$
    Percentage of population that has completed a Bachelors degree
11. **Population_By_Masters_Completed** $\in [0, 100]$
    Percentage of population that has completed a Masters degree
12. **Population_By_Doctoral_Completed** $\in [0, 100]$
    Percentage of population that has completed a Doctoral degree
13. **School_Life_Expectancy** (*years*) $\in \mathbb{R}^+$
    The total number of years of schooling (primary to tertiary) that a child can expect to receive, assuming that the probability of his or her being enrolled in school at any particular future age is equal to the current enrollment ratio at that age
14. **Expenditure_On_Education_As_A_Percentage_Of_Total_Government_Expenditure** $\in [0, 100]$
    Expenditure on education as a percentage of total government expenditure
15. **Expenditure_On_School_Books_And_Teaching_Material_As_Percentage_Of_Total_Expenditure_In_Public_Institutions** $\in [0, 100]$
    Expenditure on school books and teaching material as a percentage of total expenditure in public institutions
16. **Governement_Expenditure_On_Education** (*USD (million)*) $\in \mathbb{R}^+$
    Amount government spends on education
17. **Official_Entrance_Age_To_Compulsory_Education** (*years*) $\in \mathbb{R}^+$
    Official entrance age to compulsory education
18. **Population_Of_The_Official_Entrance_Age_To_Secondary_General_Education_Both_Sexes** $\in \mathbb{Z}^+$
    Number of people with official entrance age to secondary general education
19. **Percentage_Of_Qualified_Teachers_In_Secondary_Education_Both_Sexes** $\in [0, 100]$
    Percentage of qualified male and female teachers in secondary education

20. **Pupil_Qualified_Teacher_Ratio_In_Secondary_Education** $\in \mathbb{R}^+$
    Number of students for each qualified teacher in secondary education

21. **Teachers_In_Secondary_Education_Both_Sexes** $\in \mathbb{Z}^+$
    Number of teacher in secondary education

22. **Teaching_Staff_Compensation_As _A_Percentage_Of_ Total _Expenditure_In_Public_Institutions** $\in [0, 100]$
    Teaching staff compensation as a percentage of total expenditure in public institutions

23. **Total_Number_Of_Years_Of_ Compulsory_Education_Guaranteed** (*years*) $\in \mathbb{Z}^+$
    Total number of years of guaranteed compulsory education

24. **Total_Number_Of_Years_Of_Free_Education_Guaranteed** (*years*) $\in \mathbb{Z}^+$
    Total number of years of guaranteed free education

25. **government** (*categorical*)
    Type of government ('Republic', 'Consitutional Monarchy' etc.)

26. **religion** (*categorical*)
    Religion of majority of the population

27. **Population** $\in \mathbb{Z}^+$
    Number of citizens of a country

28. **Land_Area_km2** $(km^2) \in \mathbb{R}^+$
    Land area of a country

29. **Density_per_km2** $(population/km^2) \in \mathbb{R}^+$
    Average population per unit area

30. **Urban Population** $\in [0, 100]$
    Percentage of population categorized as urban

31. **HDI_Rank** $\in \{1, 2, ..., 190\}$
    Ranking among all countries based on Human Development Index

32. **IHDI** $\in [0, 1]$
    Human Development Index score

33. **HDI** $\in [0, 1]$
    Inequality-adjusted Human Development Index score
34. **GDP_Per_Capita (USD)** (*USD*) $\in \mathbb{R}^+$
    Gross Domestic Product per capita
35. **Govt_Budget_Per_Capita (USD, Nominal)** (*USD*) $\in \mathbb{R}^+$
    Nominal government budget per capita
36. **% Pop living under USD 1.9/Day** $\in [0, 100]$
    Percentage of population living under USD 1.9 per day
37. **% Pop living under USD 3.2/Day** $\in [0, 100]$
    Percentage of population living under USD 3.2 per day
38. **% Pop living under USD 5.5/Day** $\in [0, 100]$
    Percentage of population living under USD 5.5 per day
39. **Continent** (*categorical*)
    Continent a country belongs to
40. **Gini Index** $\in [0, 100]$
    A quantified representation of a nation's Lorenz curve. A Gini index of 0% expresses perfect equality, while index of 100% expresses maximal inequality.

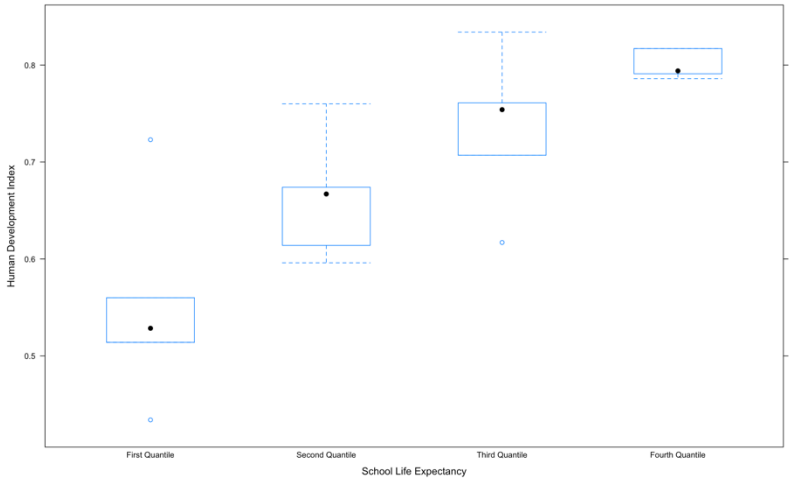
Exploratory Data Analysis

Exploratory Data Analysis is an **integral part** of any data science project. It is an approach that is used to **analyze datasets** in order to **summarize** their important and **essential characteristics**. In order to make this report as holistic as possible, we have used **base, lattice and gg plotting systems**. As can be seen throughout the report each system has a different and unique appearance.

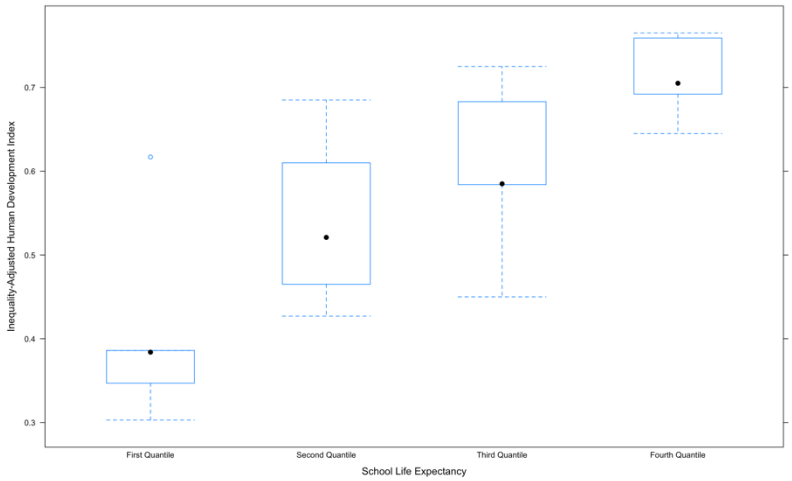
Any plot is defined using two variables, namely **independent and dependent**. The independent variable is plotted out on the x-axis and it is varied over a range of values of x. The impact of this variation on the dependent variable is mapped out on the y-axis. For our project, the independent variables are education related parameters such as **Literacy Rate, Government Expenditure on Education and School Life Expectancy**; whereas, the dependent variables are mostly economic parameters including, **Human Development Index (HDI), Inequality-Adjusted Human Development Index (IHDI) and GDP per Capita**.

Now, lets move towards the results and findings of our exploratory data analysis and what sort of questions we can formulate using these plots. Please note that we have used a variety of plots including **box, scatter and density plot**s to capture the variation within our dataset in the best possible way.

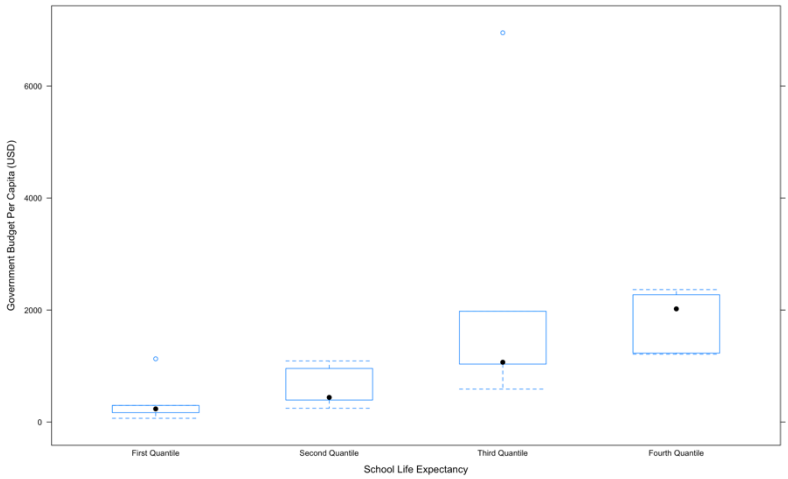
- **Effect of School Life Expectancy on Economic Variables**



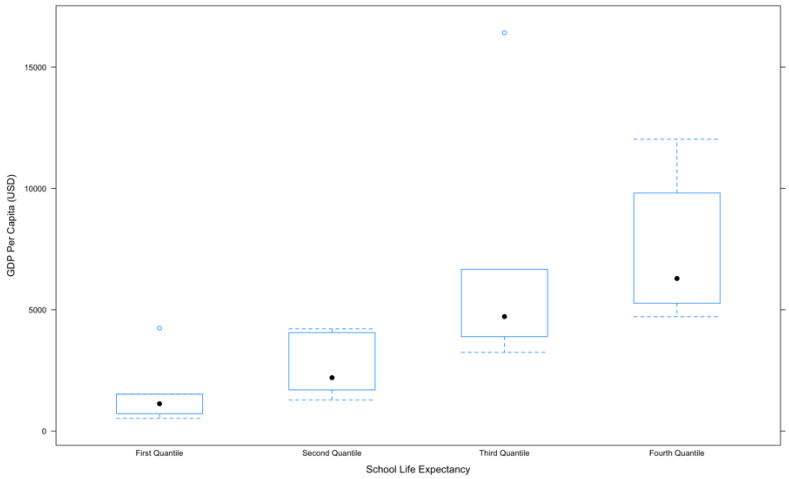
As can be seen through the boxplot that as the duration of schooling years increases human development index also increases.

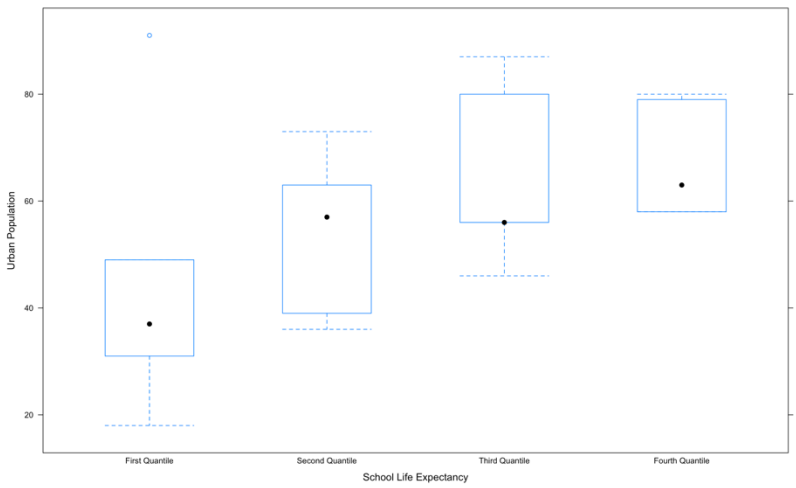


Once again, Inequality-Adjusted Human Development Index also show a similar increasing trend with school life expectancy.

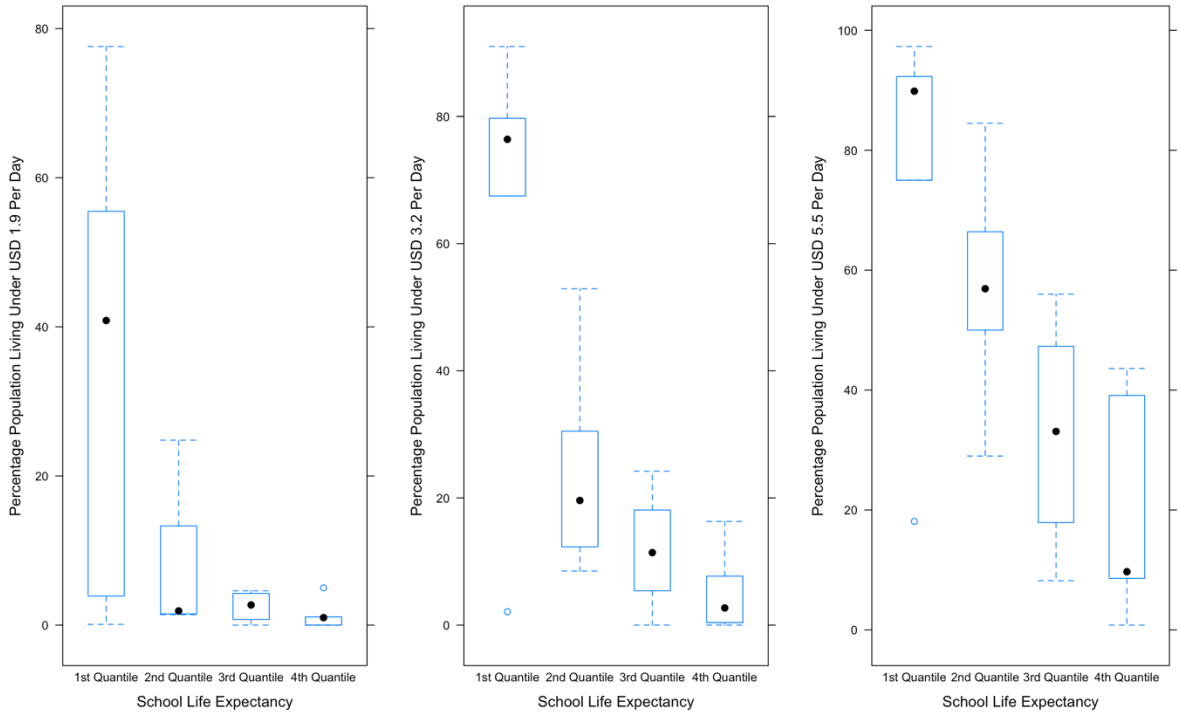


Increasing school life expectancy also leads to higher





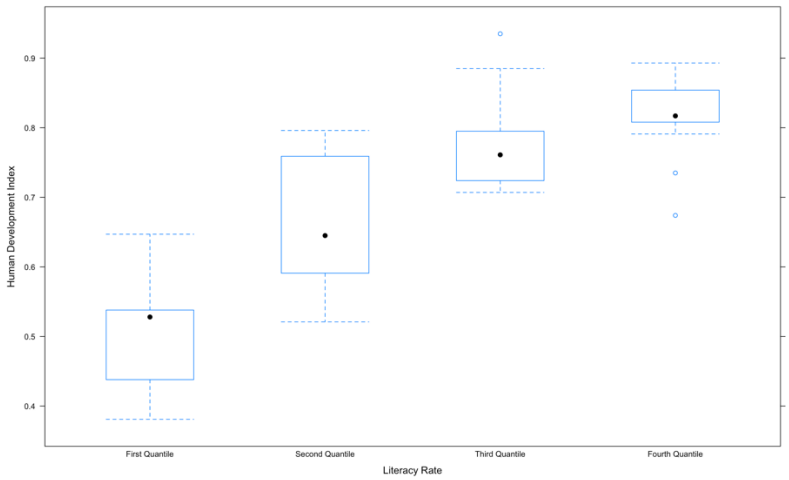
School life expectancy generally not shows a positive trend with Urban Population. It makes sense since it is not compulsory that increasing school life leads to migration from rural areas to urban cities.

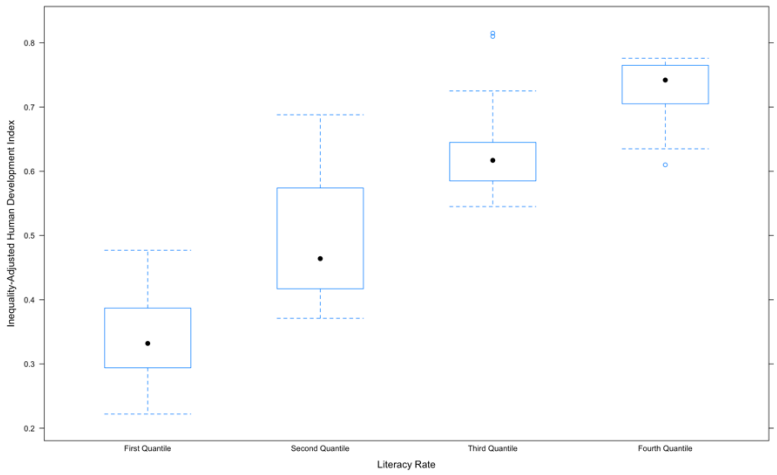


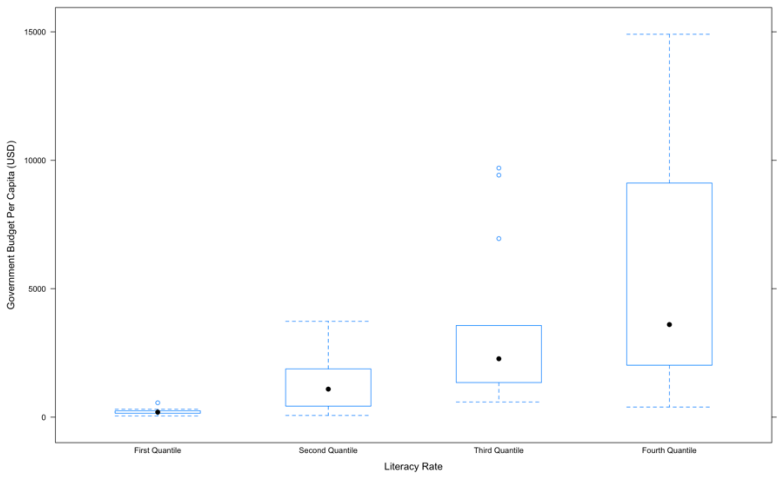
Increasing school life expectancy leads to a decrease in number of people living below certain wages.

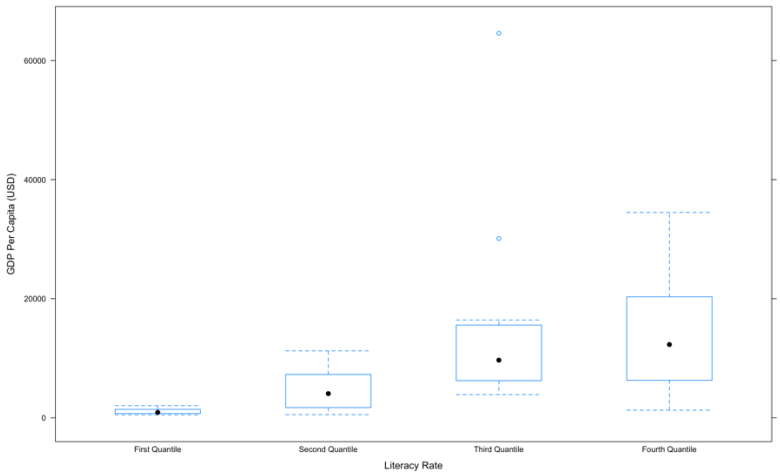
Concluding Remarks: School Life Expectancy generally leads to a higher economic variable value. For example, increasing average school life directly relates to a higher value of GDP per Capita and HDI. This means that people live off in a better way with greater school life expectancy.

- **Effect on Literacy Rate on Economic Variables**

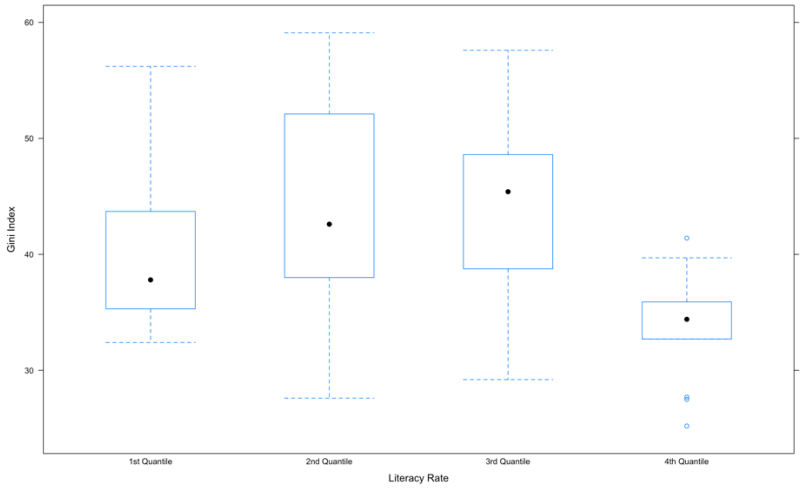


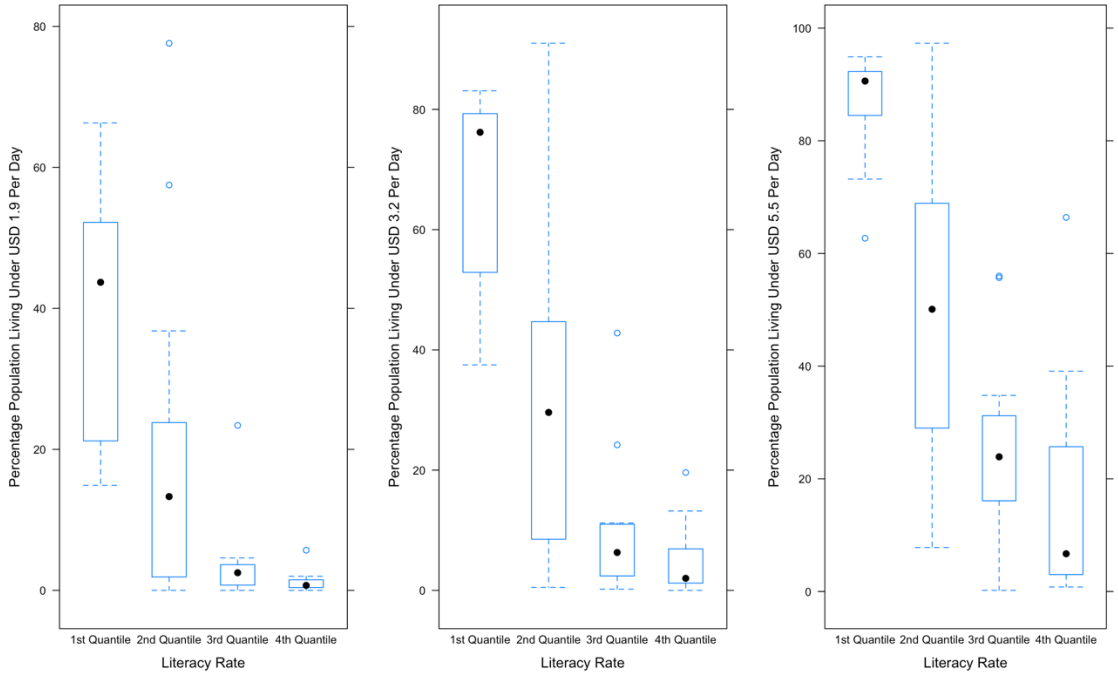


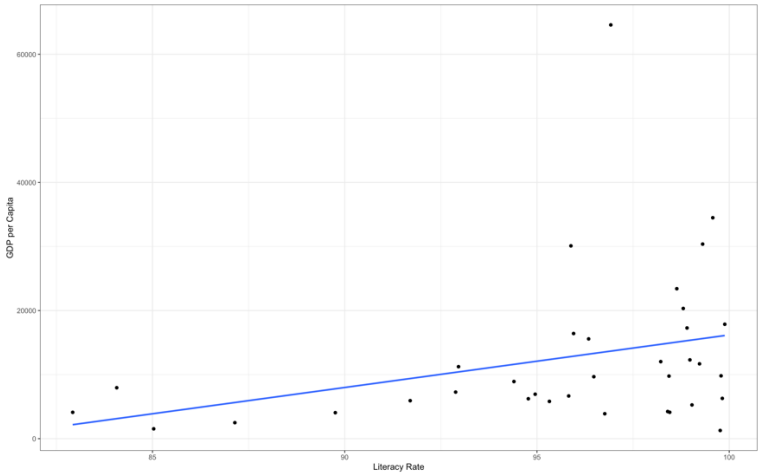


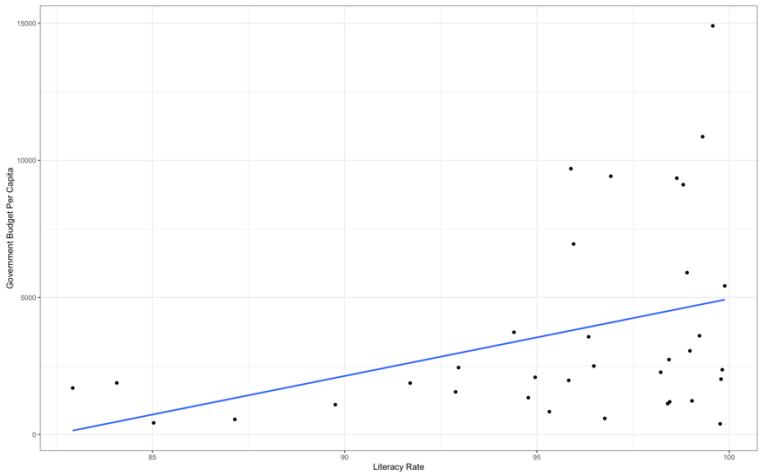


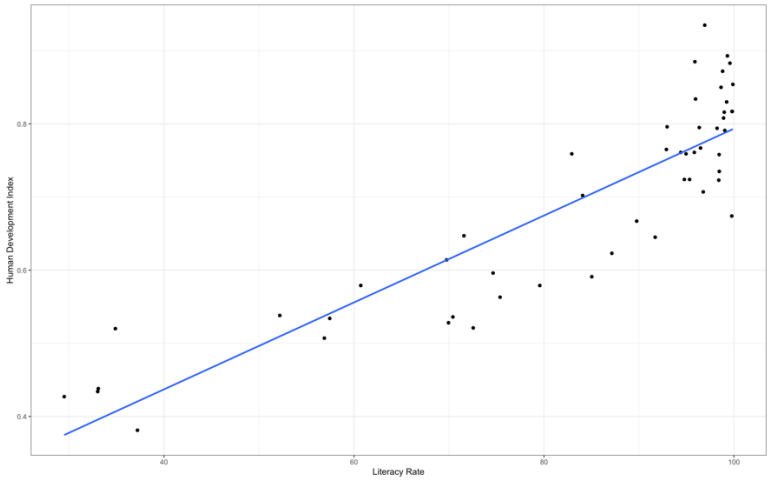
For all the above graphs, we tend to see a positive correlation between literacy rate and various economic parameters.



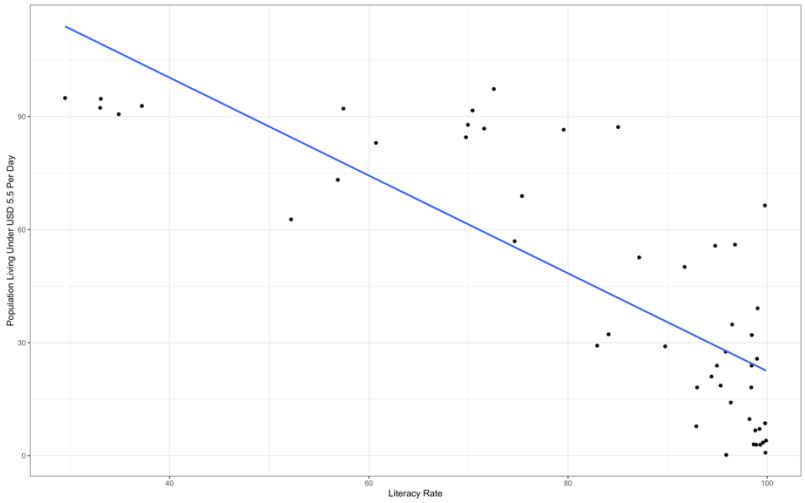




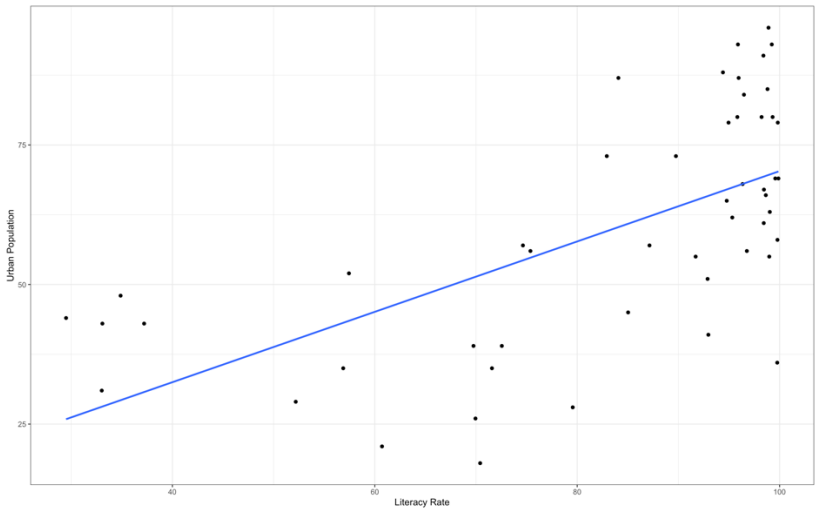




The gradient for the line fitting HDI with Literacy Rate is very steep meaning that increasing Literacy Rate has a significant impact on the development index.

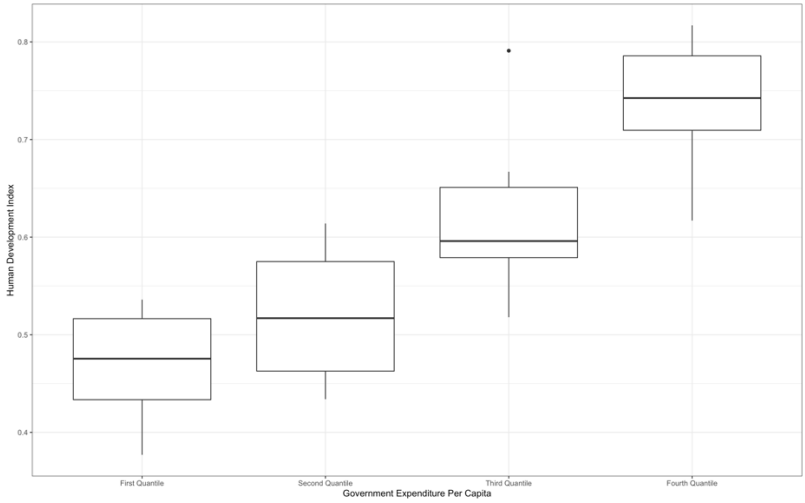


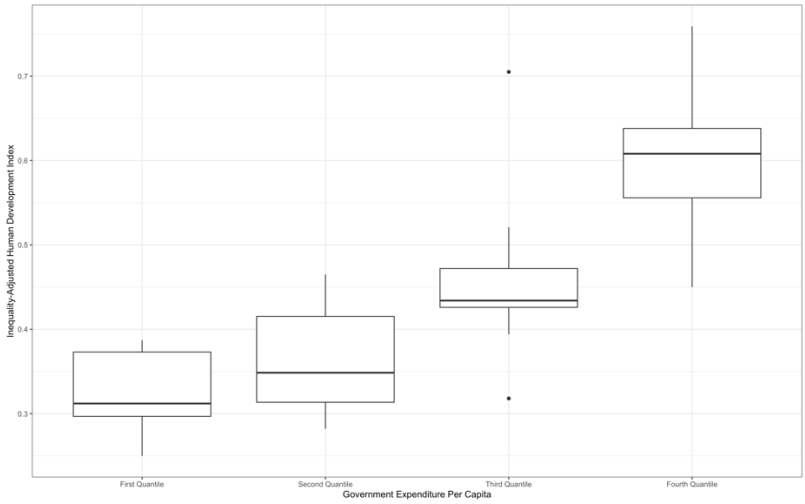
As expected, when Literacy Rate increases number of people living below 5.5 USD drastically reduces.

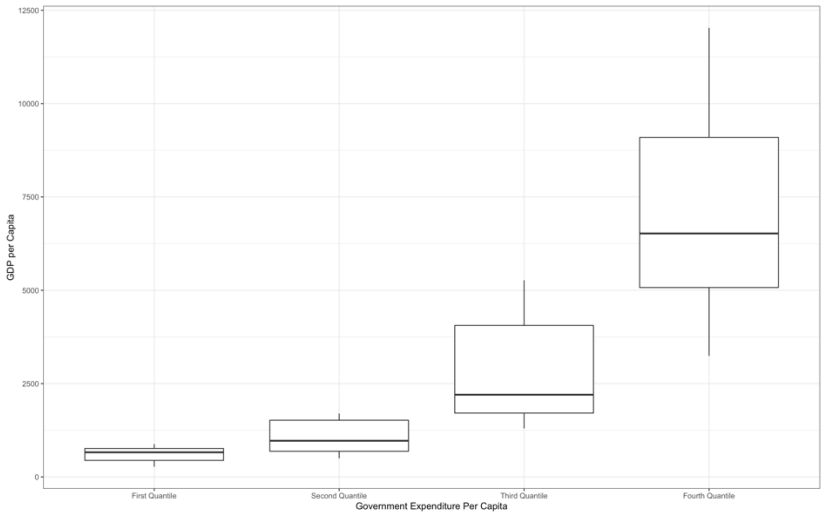


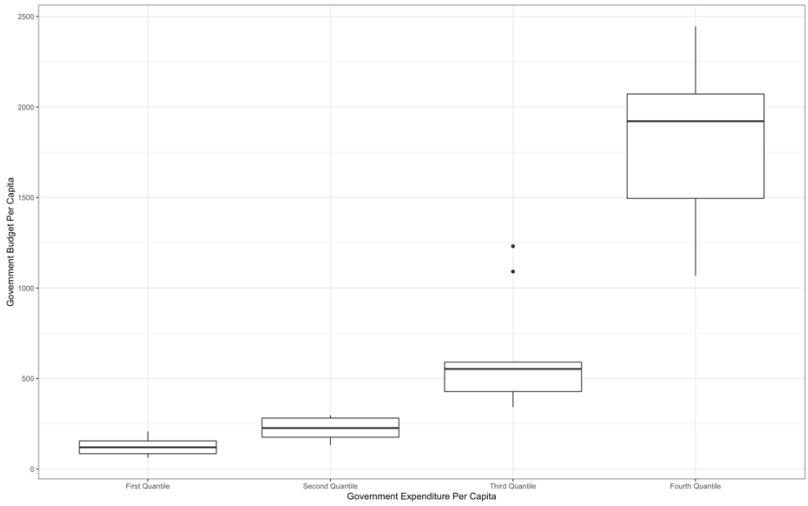
Concluding Remarks: Literacy Rate has a very strong relationship with all the economic variables that are being used. The graphs clearly indicate that greater Literacy Rate allows citizens to enjoy not an only a good lifestyle but also adds to the overall GDP of the country allowing government to perform well. The plots also indicate that countries with a higher Literacy Rate have a fewer number of citizens living beneath wages of 1.9, 3.2 and 5.5 USD respectively.

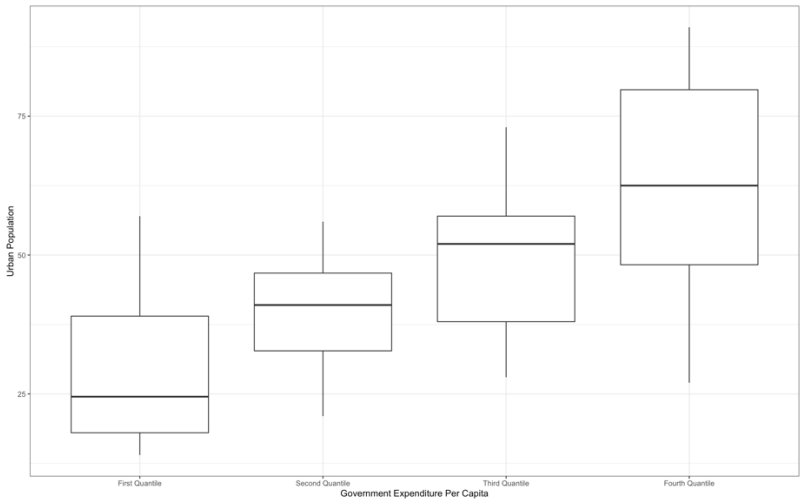
- **Effect of Government Expenditure on Education per Capita on Economic Variables**

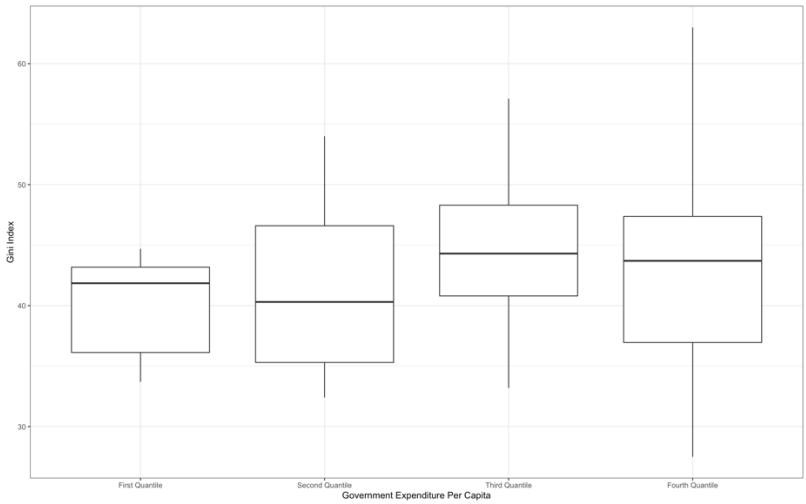


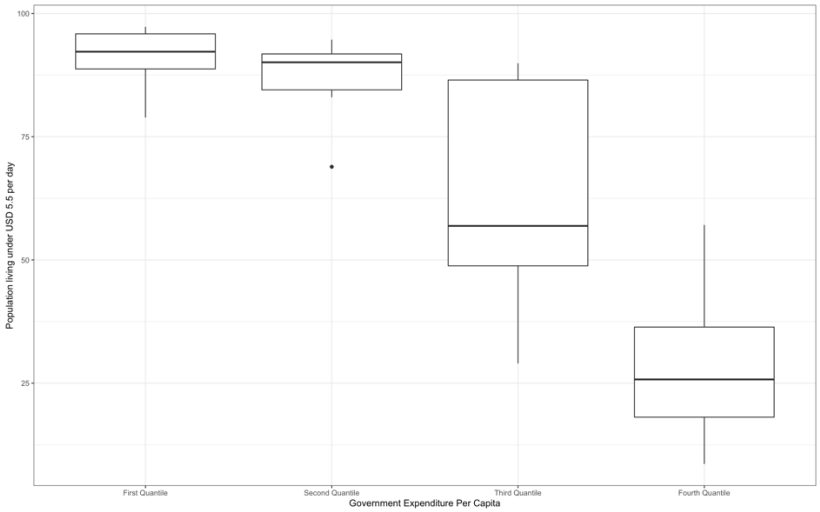


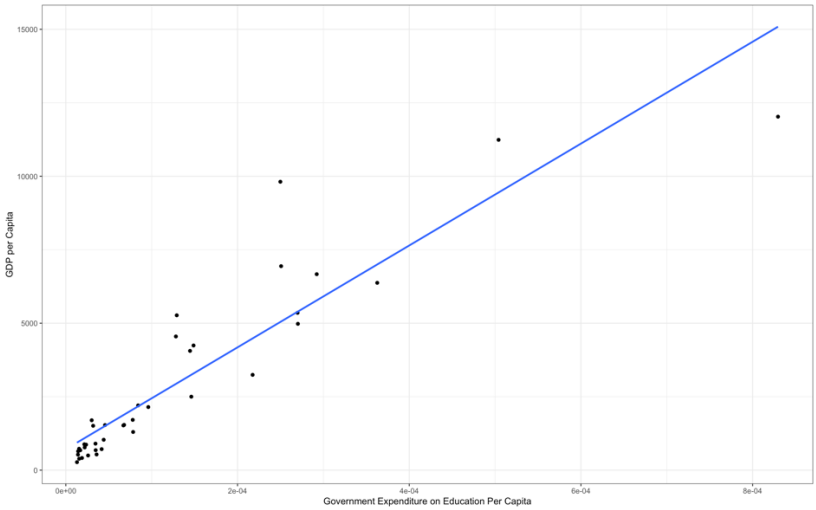


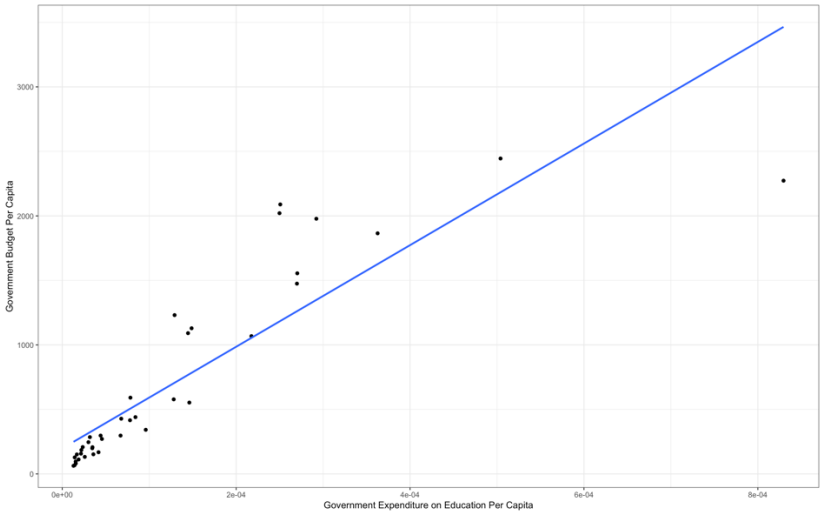

Government Budget Per Capita
2500
2000
1500
1000
500
0
First Quantile
Second Quantile
Third Quantile
Fourth Quantile
Government Expenditure Per Capita

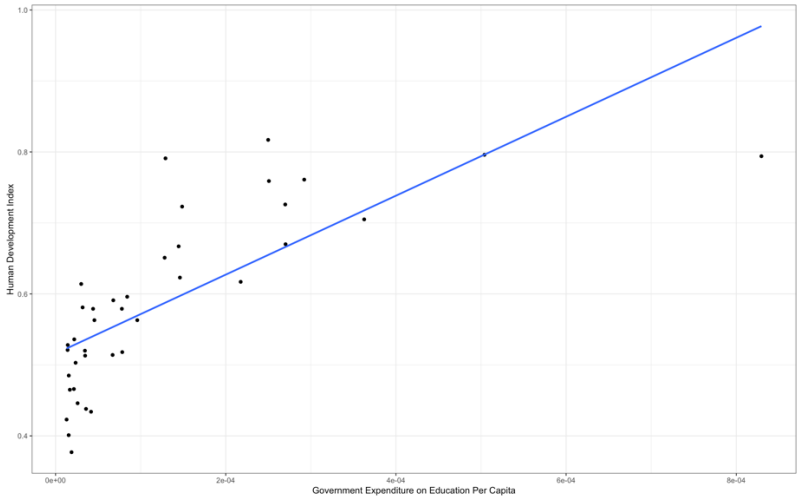

Urban Population
75
50
25
First Quantile
Second Quantile
Third Quantile
Fourth Quantile
Government Expenditure Per Capita

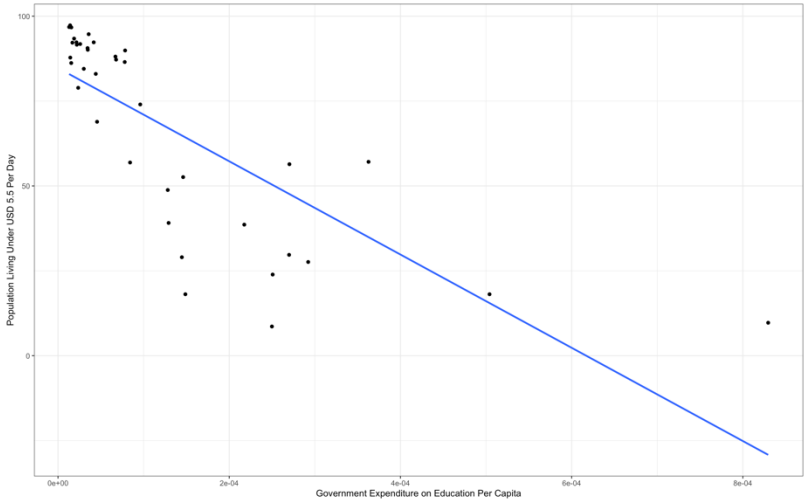

Gini Index
60
50
40
30
First Quantile
Second Quantile
Third Quantile
Fourth Quantile
Government Expenditure Per Capita


Population living under USD 5.5 per day
100
75
50
25
First Quantile
Second Quantile
Third Quantile
Fourth Quantile
Government Expenditure Per Capita


GDP per Capita
15000
10000
5000
0
0e+00
2e-04
4e-04
6e-04
8e-04
Government Expenditure on Education Per Capita

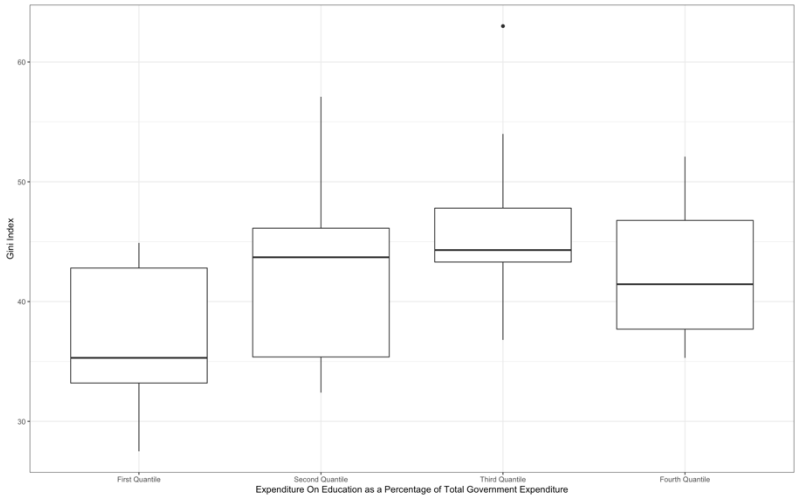

Government Budget Per Capita
3000
2000
1000
0
0e+00
2e-04
4e-04
6e-04
8e-04
Government Expenditure on Education Per Capita

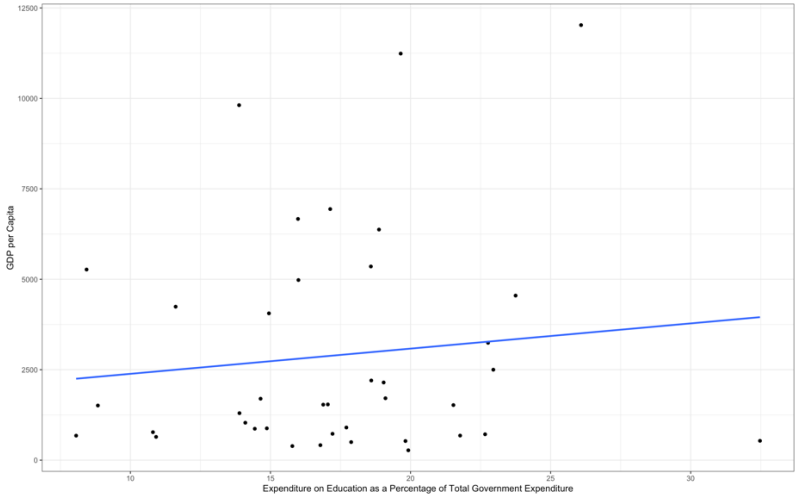


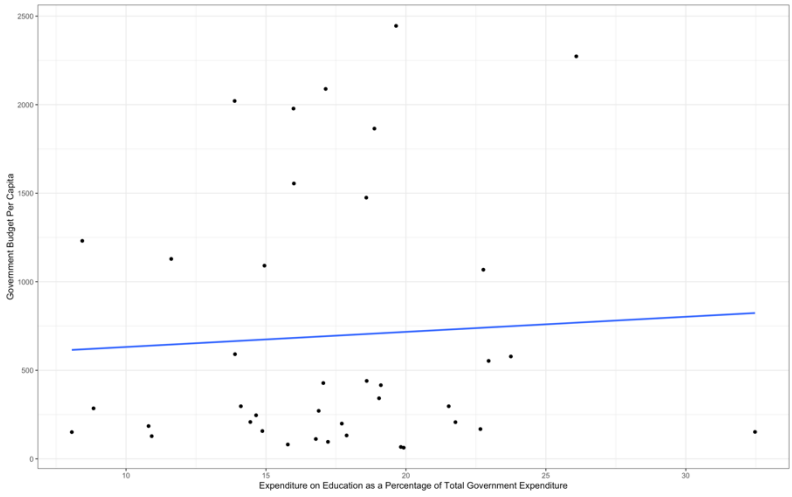


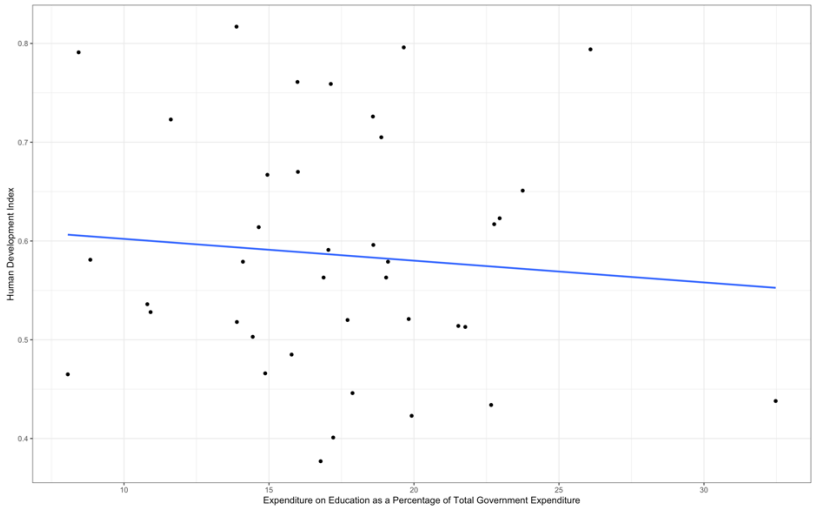
Concluding Remarks: This parameter, Government Expenditure on Education per Capita, produces some of the most encouraging graphs. There tends to be clear link between Government Expenditure on Education and various economic parameters. For example, we observe that countries which spend more on education tend to have a higher GPD per Capita meaning citizens may enjoy better quality of life.

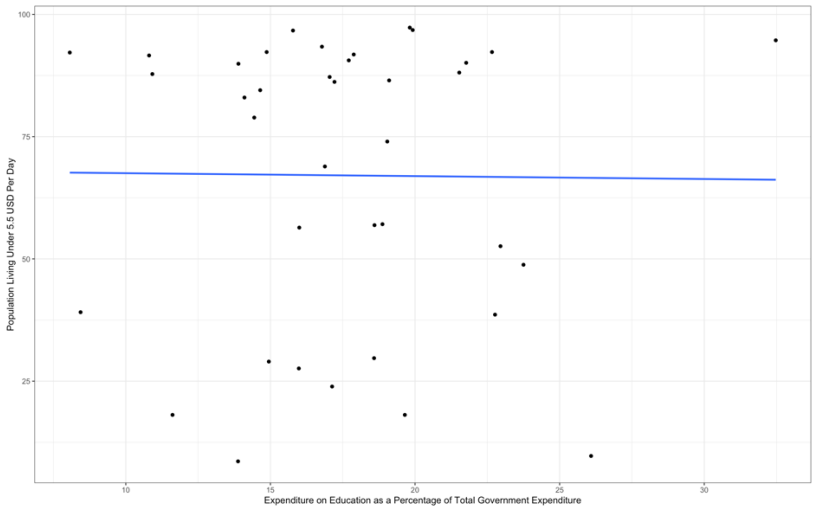
- **Effect of Expenditure on Education as a Percentage of Total Government Spending on Economic Variables**





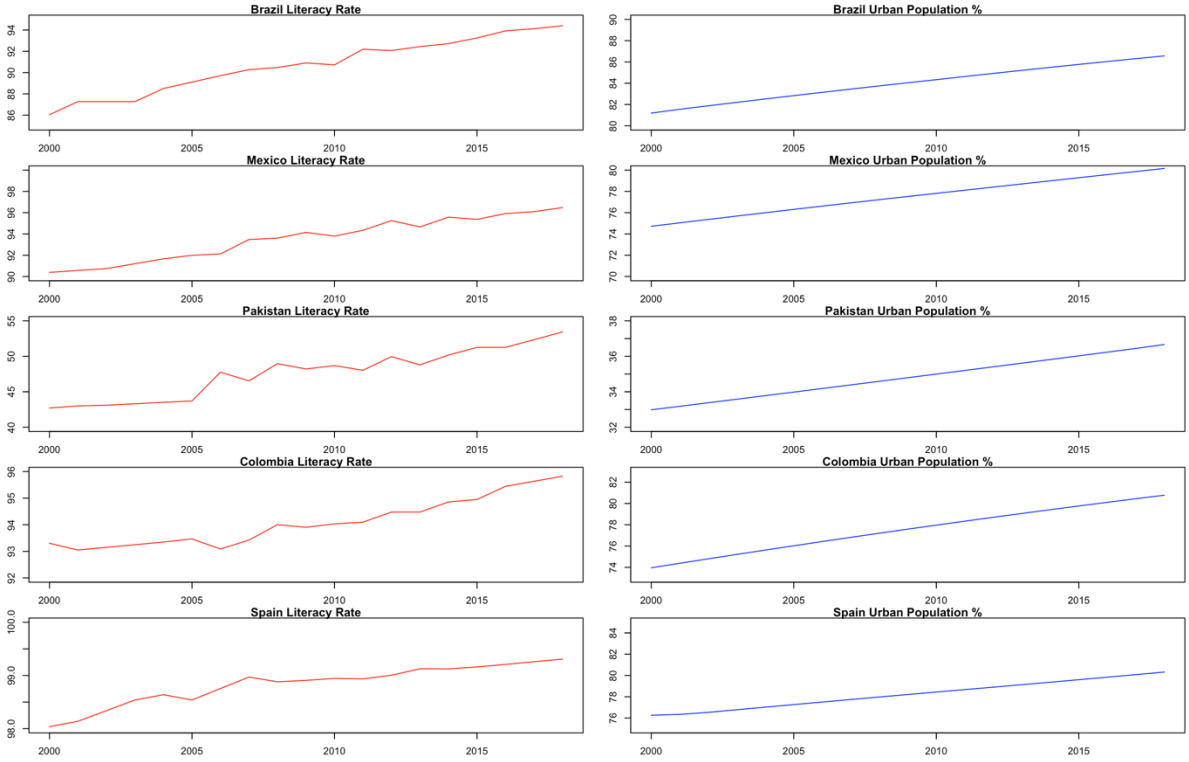




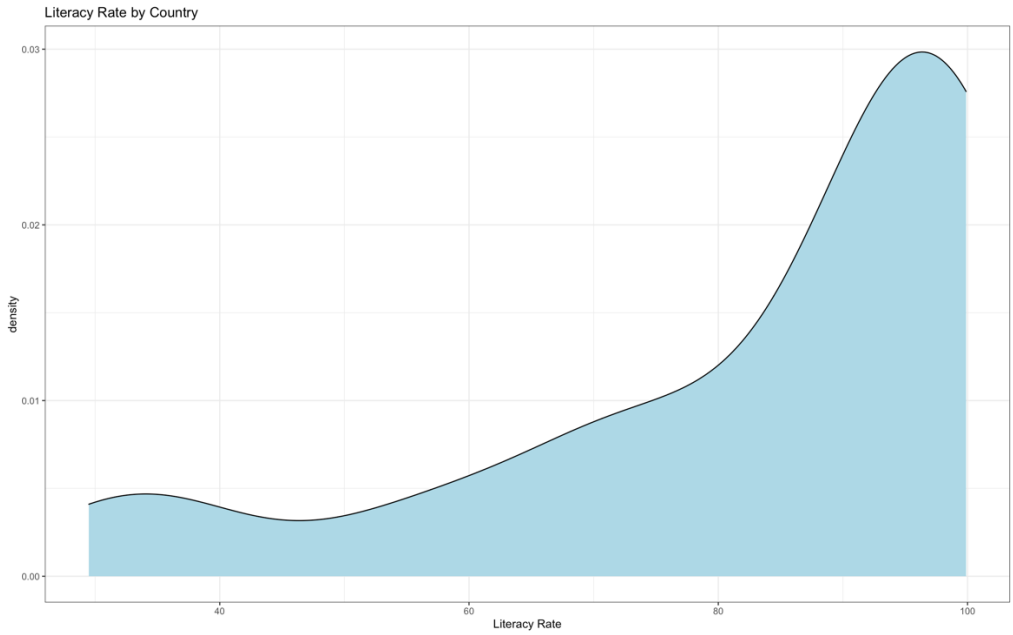


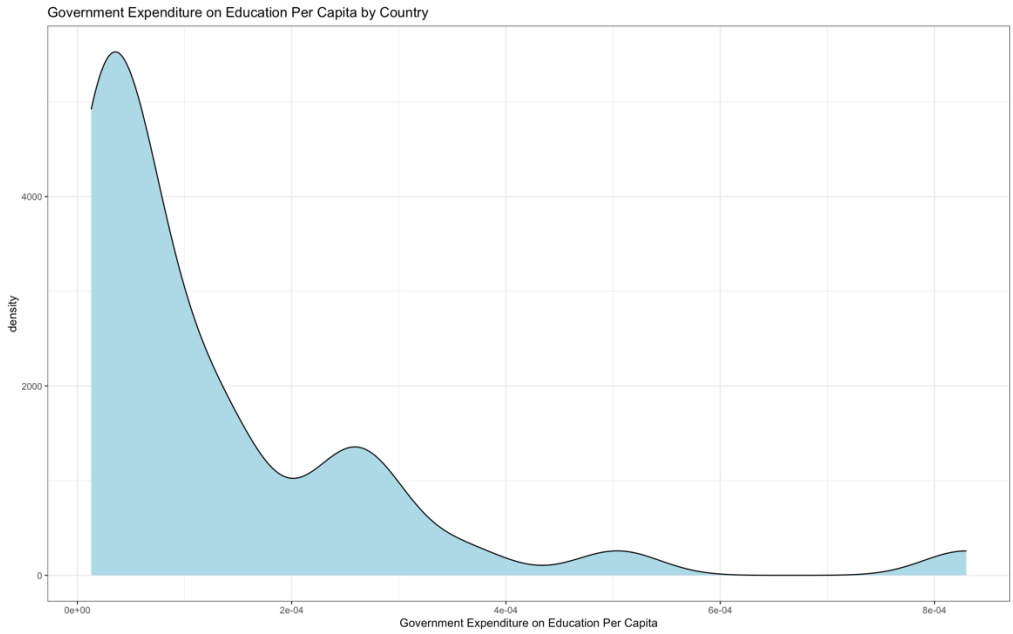
Concluding Remarks: This variable did not give us as good results as the previous ones. It showed that despite higher expenditure of education the Human Development Index decreased. Moreover, the population living under 5.5 USD did not change as expenditure increased unlike shown in previously used independent variables.

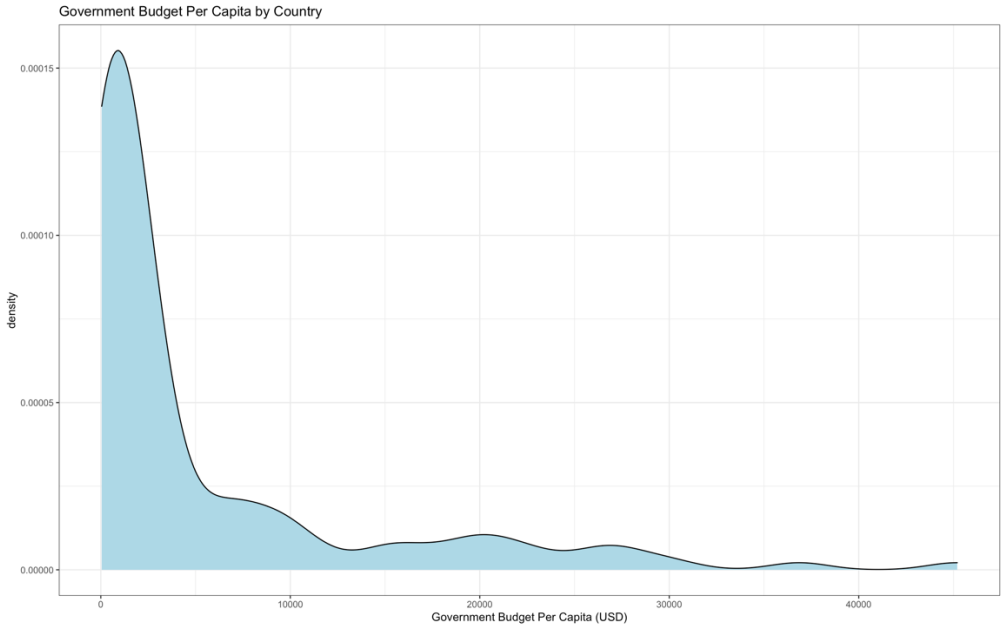
Other than these plots, we also performed a time series analysis of some specific countries from different continents, including Brazil, Spain, Pakistan, Mexico and Colombia. These graphs gave us an insight into how different parameters change for different countries over time. These 10 graphs that depict relationship between Literacy Rate and Urban Population for the selected countries.

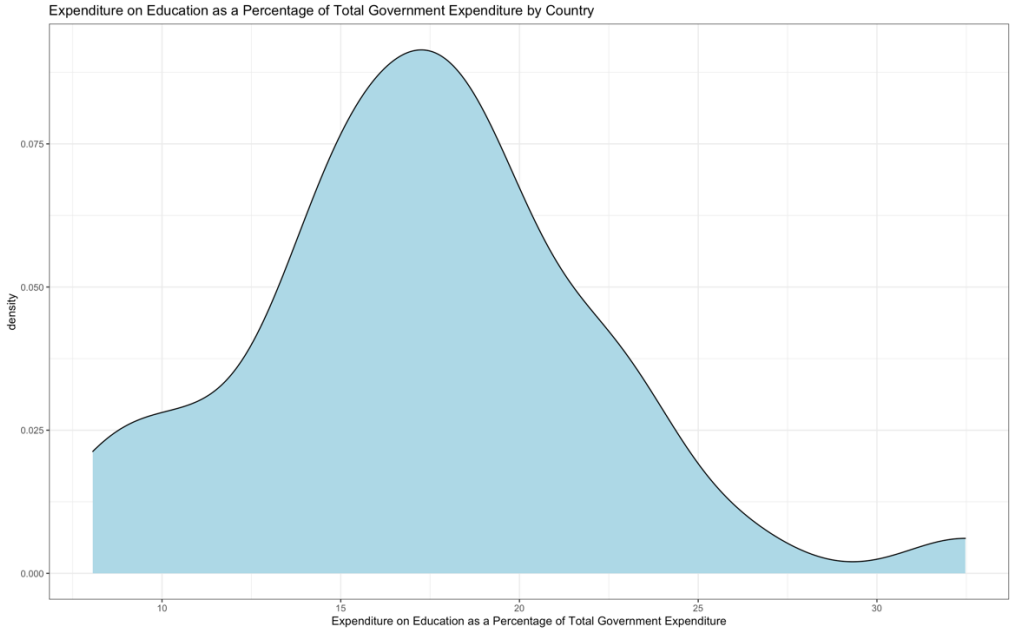


Moreover, to express our entire dataset as a whole we have attached density plots below,

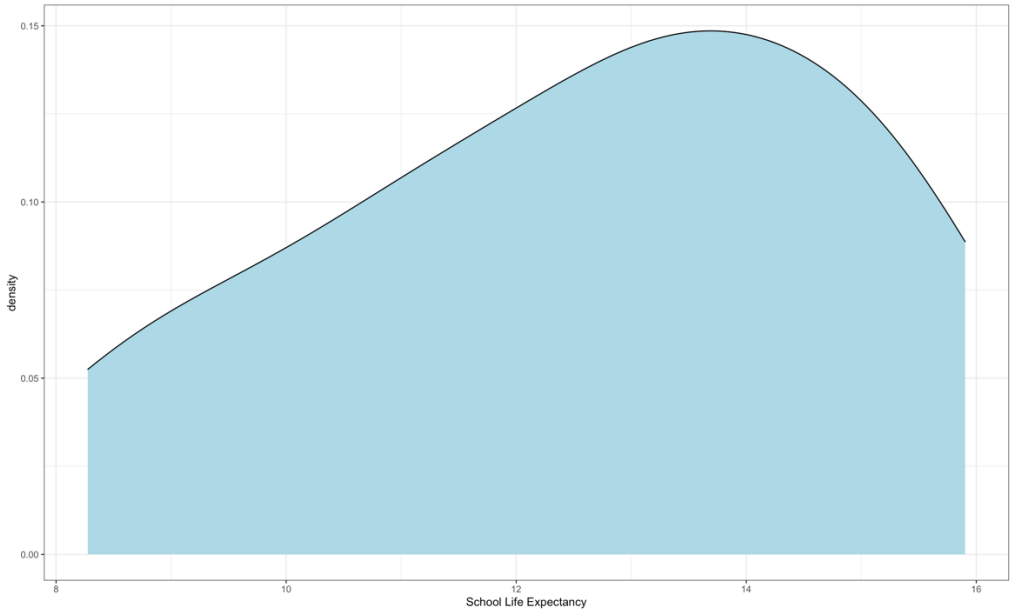




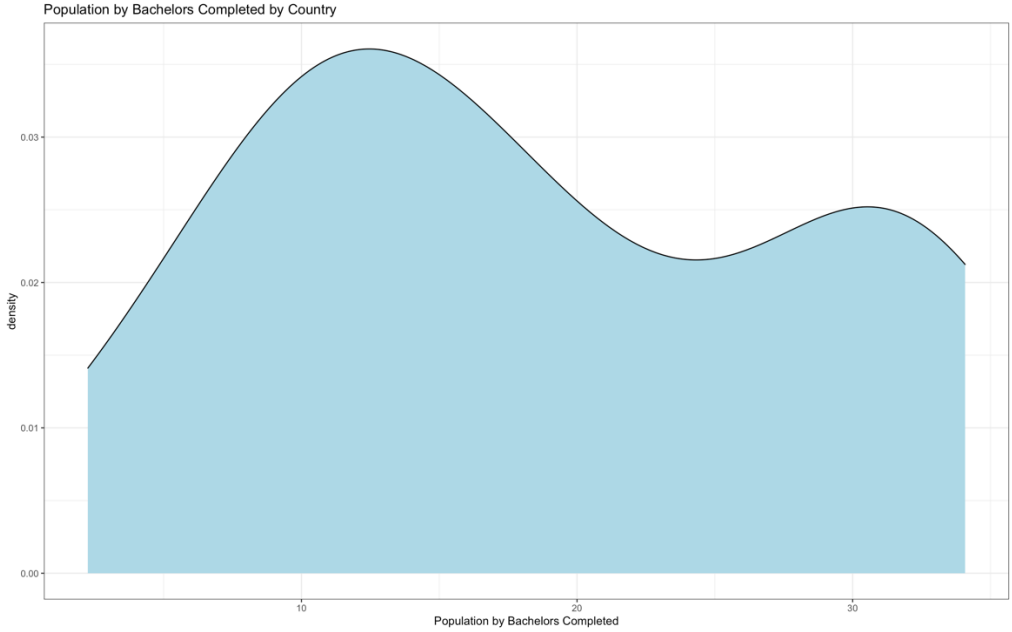

Government Budget Per Capita by Country
density
0.00015
0.00010
0.00005
0.00000
0
10000
20000
30000
40000
Government Budget Per Capita (USD)

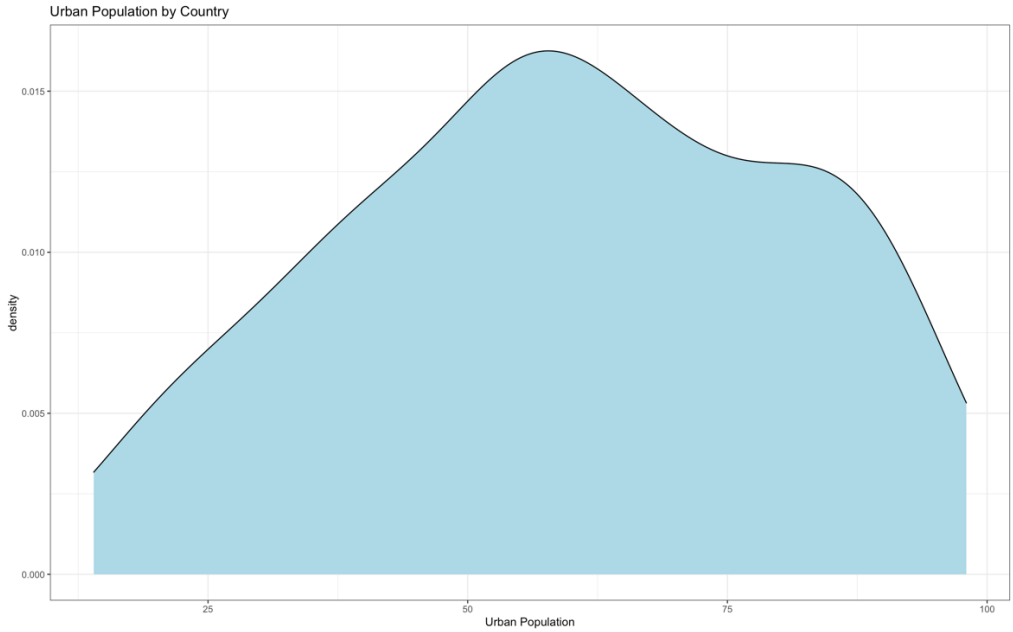

Expenditure on Education as a Percentage of Total Government Expenditure by Country
density
0.075
0.050
0.025
0.000
10
15
20
25
30
Expenditure on Education as a Percentage of Total Government Expenditure

School Life Expectancy by Country



Population by Bachelors Completed by Country




Urban Population by Country
density
0.015
0.010
0.005
0.000
25
50
75
100
Urban Population

Questions (& Expected Answers)

Having examined relationship between various variables, we want to next evaluate whether evident relationships, like reduction in poverty with increased literacy rate or increase in GDP per capita with increase in government expenditure on education per capita, are statistically significant.

Some specific questions (and expected answers) are as follows:

- *Is there statistically significant difference in GDP Per Capita/ Government Budget Per Capita/ HDI/ Urban Population/ Poverty Level (measured by Population Living Under USD 5.5 Per Day) between countries with high literacy rate and countries with low literacy rate?*
  Based various plots obtained and brief discussion above, we **do** expect significant difference between these categories.

- *Is there statistically significant difference in GDP Per Capita/ Government Budget Per Capita/ HDI/ Urban Population/ Poverty Level (measured by Population Living Under USD 5.5 Per Day) between countries with high school life expectancy and countries with low school life expectancy?*
  Based various plots obtained and brief discussion above, we **do** expect significant difference between these categories.

- *Is there statistically significant difference in GDP Per Capita/ Government Budget Per Capita/ HDI/ Urban Population/ Poverty Level (measured by Population Living Under USD 5.5 Per Day) between countries with high government expenditure on education per capita and countries with low government expenditure on education per capita?*
  Based various plots obtained and brief discussion above, we **do** expect significant difference between these categories.

- *Is there statistically significant difference in Inequality (measured by Gini Index) between countries with high literacy rate/ school life expectancy/ government expenditure on education per capita and countries with low literacy rate/ school life expectancy/ government expenditure on education per capita?*
  Based various plots obtained and brief discussion above, we **do not** expect significant difference between these categories.

Statistical Inference

We started this project with the aim of gaining insight on **how education impacts socioeconomic development**, and how we may reasonably evaluate this became clear when we performed EDA on data we had gathered.

Here, we investigate whether the relationships that appeared plausible in EDA have firm statistical grounding as well. For this, we examine the effect of selected variables related to education on selected variables related to socioeconomic development. Since hypothesis testing is performed on groups, we transform our educational variables to *categorical* data through use of four quantiles. We group the data points (with each point representing a unique country) from 0 to 25 percentile into *First Quantile*, from 25 to 50 percentile into *Second Quantile*, from 50 to 75 percentile into *Third Quantile*, from 75 to 100 percentile into *Fourth Quantile*. We then test whether these groups are different based on values of a particular socioeconomic variable that each group takes.

The variables related to education we use here are:

1. Literacy Rate
2. School Life Expectancy
3. Government's Educational Expenditure Per Capita

The variables related to socioeconomic development we use here are:

1. GDP Per Capita
2. Government Budget Per Capita
3. Human Development Index
4. Urban Population
5. Population Living Under USD 5.5 Per Day
6. Gini Index

Testing is done by formulating Null (and Alternate) Hypothesis, performing (pairwise) t-test, and subsequently accepting or rejecting the hypothesis based on p value obtained. We use a **significance level** of **0.05** (95% confidence interval) for our analysis.

If P_{ij} – Value > 0.05, we cannot reject Ho_{ij} in favor of HA_{ij}, i.e. there is no difference between i^{th} and j^{th} quantile of selected educational variable on selected socio-economic variable.

If P_{ij} – Value < 0.05, we reject Ho_{ij} in favor of HA_{ij}, i.e. there is a statistically significant difference between i^{th} and j^{th} quantile of selected educational variable on selected socio-economic variable.

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

Literacy Rate

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **literacy rate** on **GDP per capita**.

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **literacy rate** on **GDP per capita**.

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.7432	-	-
Third Quantile	0.0047	0.0401	-
Fourth Quantile	0.0047	0.0401	0.9599

Fig.1 Results for **GDP per capita** vs **Literacy Rate** using t-test

We reject the null hypothesis in most cases, but accept it when groups under consideration are First Quantile & Second Quantile and Third Quantile & Fourth Quantile. This is not against our expectation; for example, it is understandable that countries with say 70% literacy rate may not be much different from that with 80% literacy rate since other factors come into play once literacy rate is sufficiently high. This is a general observation and remark, which is applicable to many future results.

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **literacy rate** on G**overnment Budget per Capita.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **literacy rate** on G**overnment Budget per Capita.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.34096	-	-
Third Quantile	0.01800	0.14716	-
Fourth Quantile	0.00014	0.00251	0.18614

Fig.2 Results for G**overnment Budget per Capita** vs **Literacy Rate** using t-test

We reject Null Hypothesis when comparing First Quantile & Third Quantile, First Quantile & Fourth Quantile and Second Quantile & Fourth Quantile, and accept it in remaining cases. Once again significant difference exists between extremes (e.g. First Quantile & Second Quantile).

The methodology of accepting or rejecting Null Hypothesis is clear by now. We would not specifically comment on this here onwards since this is straight forward to grasp from associated result figures. We would, however, provide comprehensive analysis at the very end.

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **literacy rate** on **Human Development Index (HDI).**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **literacy rate** on **Human Development Index (HDI).**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	2.8e-05	-	-
Third Quantile	2.0e-11	0.00023	-
Fourth Quantile	3.9e-13	6.8e-06	0.21823

Fig.3 Results for **Human Development Index (HDI)** vs **Literacy Rate** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **literacy rate** on **Urban Population %.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **literacy rate** on **Urban Population %.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.0016	-	-
Third Quantile	1.2e-07	0.0123	-
Fourth Quantile	1.3e-06	0.0644	0.3977

Fig.4 Results for **Urban Population %** vs **Literacy Rate** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **literacy rate** on **Population Living Under $5.5 per Day.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **literacy rate** on **Population Living Under $5.5 per Day.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	8.5e-05	-	-
Third Quantile	1.5e-08	0.01491	-
Fourth Quantile	8.0e-11	0.00037	0.23104

Fig.5 Results for **Population Living Under $5.5 per Day** vs **Literacy Rate** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **literacy rate** on **GINI Index.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **literacy rate** on **GINI Index.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.5342	-	-
Third Quantile	0.5342	0.8838	-
Fourth Quantile	0.1321	0.0052	0.0103

Fig.6 Results for **GINI Index** vs **Literacy Rate** using t-test

School Life Expectancy

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **GDP per capita**.

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **GDP per capita**.

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	1.000	-	-
Third Quantile	0.061	0.147	-
Fourth Quantile	0.036	0.105	1.000

Fig.7 Results for **GDP per capita** vs **School Life Expectancy** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **School Life Expectancy** on G**overnment Budget per Capita.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **School Life Expectancy** on G**overnment Budget per Capita.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	1.00	-	-
Third Quantile	0.16	0.30	-
Fourth Quantile	0.35	0.53	1.00

Fig.8 Results for G**overnment Budget per Capita** vs **School Life Expectancy** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **Human Development Index (HDI).**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **Human Development Index (HDI).**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.05472	-	-
Third Quantile	0.00260	0.26271	-
Fourth Quantile	0.00013	0.02965	0.26271

Fig.9 Results for **Human Development Index (HDI)** vs **School Life Expectancy** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **Urban Population %.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **Urban Population %.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	1.00	-	-
Third Quantile	0.39	1.00	-
Fourth Quantile	0.30	1.00	1.00

Fig.10 Results for **Urban Population %** vs **School Life Expectancy** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **Population Living Under $5.5 per Day.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **Population Living Under $5.5 per Day.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.4091	-	-
Third Quantile	0.0494	0.4091	-
Fourth Quantile	0.0065	0.0975	0.4495

Fig.11 Results for **Population Living Under $5.5 per Day** vs **School Life Expectancy** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **GINI Index.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **School Life Expectancy** on **GINI Index.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.5342	-	-
Third Quantile	0.5342	0.8838	-
Fourth Quantile	0.1321	0.0052	0.0103

Fig.12 Results for **GINI Index** vs **School Life Expectancy** using t-test

Government Education Expenditure Per Capita

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **GDP per capita**.

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **GDP per capita**.

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.557	-	-
Third Quantile	0.023	0.061	-
Fourth Quantile	2.2e-09	1.0e-08	1.3e-05

Fig.13 Results for **GDP per capita** vs **Government Education Expenditure Per Capita** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on G**overnment Budget per Capita.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on G**overnment Budget per Capita.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.4322	-	-
Third Quantile	0.0013	0.0074	-
Fourth Quantile	2.5e-14	1.3e-13	6.2e-10

Fig.14 Results for G**overnment Budget per Capita** vs **Government Education Expenditure Per Capita** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **Human Development Index (HDI).**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **Human Development Index (HDI).**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.0968	-	-
Third Quantile	6.6e-05	0.0041	-
Fourth Quantile	5.0e-10	5.3e-08	0.0012

Fig.15 Results for **Human Development Index (HDI)** vs **Government Education Expenditure Per Capita** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **Urban Population %.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **Urban Population %.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.35131	-	-
Third Quantile	0.05659	0.35131	-
Fourth Quantile	0.00045	0.01272	0.25010

Fig.16 Results for **Urban Population %** vs **Government Education Expenditure Per Capita** using t-test

Ho_{ij}: There is no difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **Population Living Under $5.5 per Day.**

HA_{ij}: There is a statistically significant difference between i^{th} and j^{th} quantile of **Government Education Expenditure Per Capita** on **Population Living Under $5.5 per Day.**

Where $i \neq j$, $i = \in \{2,3,4\}$ and, $j = \in \{1,2,3\}$.

	First Quantile	Second Quantile	Third Quantile
Second Quantile	0.54204	-	-
Third Quantile	0.00056	0.00274	-
Fourth Quantile	3.3e-10	2.8e-09	7.8e-05

Fig.17 Results for **Population Living Under $5.5 per Day** vs **Government Education Expenditure Per Capita** using t-test

Conclusion:

We performed **17** *pairwise* t-tests, and **102** t-tests in total. The general insight was that **socioeconomic development of a country *is* strongly impacted by the quality of/ level of/ interest in education**.

Our choice of independent variables (related to education), and more importantly what they represented, did involve subjectivity; and that is why we selected many of them. Government Expenditure on Education Per Capita depicted interest in (investing in) education, and is perhaps a metric of how much a country 'values' education. Literacy Rate is a commonly used measure of quality of education. School Life Expectancy depicted *how* educated a country perhaps was. Together, these made a strong pool of variables for evaluating whether education (quality/ level/ interest) was related to a country's socioeconomic development.

We reject a vast majority of our Null Hypothesis, indicating that the difference in socioeconomic development for groups of varying education (quality/ level/ interest) is not random. The difference was almost always that countries in higher quantiles of education variables had higher values of variables indicating socioeconomic well-being. The results reaffirmed our intuition that more educated countries (and, by extension, their citizens) are economically prosperous, happier, face less poverty etc. It also serves as strong evidence in favor of importance of education. Some other interesting data-backed observations were also made, for example, better education does not translate to more equality (as measured by Gini Index).

Our results show that socioeconomic development and education are indeed related, and this should not be surprising. Still, various questions remain unaddressed or ill-addressed; for example, our analysis is insufficient to establish causality. Our work opens door for vast analysis on the topic, which can be important for multiple domains especially policy making.