

# What factors account for the variation in HDI scores across different regions of the world?

true

2023-05-03

## Summary:

The study aims to explore the factors that affect a country's Human Development Index (HDI). The primary units of observation are countries, and the key finding is that factors such as gross national income (GNI) per capita, life expectancy, and mean years of schooling have a significant positive relationship with HDI.

## Background & expectations:

The Human Development Index is a measure of a country's overall human development based on three dimensions: a long and healthy life, access to knowledge, and a decent standard of living. The motivation behind this study is to identify the factors that contribute to a country's overall human development. The theoretical argument is that a country's HDI is primarily influenced by its economic, social, and health indicators. The hypotheses are that countries with higher GNI per capita, longer life expectancy, and more years of schooling will have higher HDI scores.

## Data:

The dataset used in this study contains data on 189 countries, including their HDI scores and various economic, social, and health indicators. This dataset provides information on the human development index (HDI) and its components for countries around the world. The HDI is a measure of human development that takes into account three dimensions: health (measured by life expectancy at birth), education (measured by years of schooling for adults aged 25 years and older and expected years of schooling for children of school-entering age), and income (measured by gross national income per capita). The dataset includes HDI values for 189 countries, as well as values for each of the three components of the HDI. The data were last updated in 2021. Data is downloaded from <https://www.kaggle.com/datasets/sudhirnl7/human-development-index-hdi>.

Country	HDI	LifeExpectancy	MeanYearsSchooling	GNIperCapita
Length:188	Min. :0.3520	Min. :48.90	Min. : 1.400	Min. : 587
Class :character	1st Qu.:0.5777	1st Qu.:65.65	1st Qu.: 6.075	1st Qu.: 3511
Mode :character	Median :0.7285	Median :73.40	Median : 8.650	Median : 10416
NA	Mean :0.6988	Mean :71.35	Mean : 8.371	Mean : 17314
NA	3rd Qu.:0.8113	3rd Qu.:77.00	3rd Qu.:11.125	3rd Qu.: 23517
NA	Max. :0.9490	Max. :84.20	Max. :13.400	Max. :129916

## Description of the Summary

The summary statistics provide information on the distribution of the variables in the `new_hdi` dataframe.

For the HDI variable, the minimum value is 0.352 and the maximum is 0.949, with a mean of 0.6988 and a median of 0.7285. The first quartile is 0.5777 and the third quartile is 0.8113, indicating that the majority of countries have an HDI between these values.

The `LifeExpectancy` variable has a minimum value of 48.90 and a maximum value of 84.20. The mean life expectancy is 71.35 years, and the median is 73.40 years. The first quartile is 65.65 and the third quartile is 77.00, indicating that most countries have a life expectancy between these values.

The `MeanYearsSchooling` variable ranges from a minimum of 1.400 to a maximum of 13.400. The mean is 8.371 and the median is 8.650. The first quartile is 6.075 and the third quartile is 11.125, indicating that most countries have a mean years of schooling between these values.

Finally, the `GNIPerCapita` variable has a minimum value of 587 and a maximum value of 129916. The mean is 17314 and the median is 10416. The first quartile is 3511 and the third quartile is 23517, indicating that the majority of countries have a GNI per capita between these values.

These statistics give us an idea of the range, central tendency, and dispersion of the variables in the data frame. For example, we can see that HDI and GNI per capita have a large range of values, while life expectancy and mean years of schooling have a narrower range. The means and medians of the variables are somewhat similar, indicating that the distributions are roughly symmetrical. The quartiles give us an idea of where the bulk of the data lies for each variable. The distribution of scores on the key variables indicates that GNI per capita and mean years of schooling are positively skewed, while life expectancy and HDI score are normally distributed. The graph of the HDI score shows a slightly right-skewed distribution.

## Description of the barplot

The graph is a bar plot showing the Human Development Index (HDI) for each country in the dataset, with countries ordered by their HDI value. The x-axis shows the name of each country, and the y-axis shows the HDI value, ranging from 0.3 to 1.0. The bars are colored in blue and have a width of 0.8.

The plot shows that there is a large variation in HDI values across countries. Norway has the highest HDI value of around 0.95, while Niger has the lowest HDI value of around 0.35. There are a few countries with very high HDI values (above 0.9), and several countries with very low HDI values (below 0.5). Most countries fall in the middle range, with HDI values between 0.5 and 0.8.

The plot also shows that there are more countries with lower HDI values than higher HDI values. The plot can be useful for comparing the HDI values of different countries and identifying which countries have high or low levels of human development.

```
## [1] 0
```

```
##      Country      HDI      LifeExpectancy      MeanYearsSchooling
## Length:188      Min.    :0.3520      Min.    :48.90      Min.    : 1.400
## Class :character 1st Qu.:0.5777      1st Qu.:65.65      1st Qu.: 6.075
## Mode  :character Median :0.7285      Median :73.40      Median : 8.650
##                      Mean  :0.6988      Mean   :71.35      Mean   : 8.371
##                      3rd Qu.:0.8113      3rd Qu.:77.00      3rd Qu.:11.125
##                      Max.   :0.9490      Max.   :84.20      Max.   :13.400
##      GNIPerCapita
##      Min.    :   587
##      1st Qu.: 3511
```

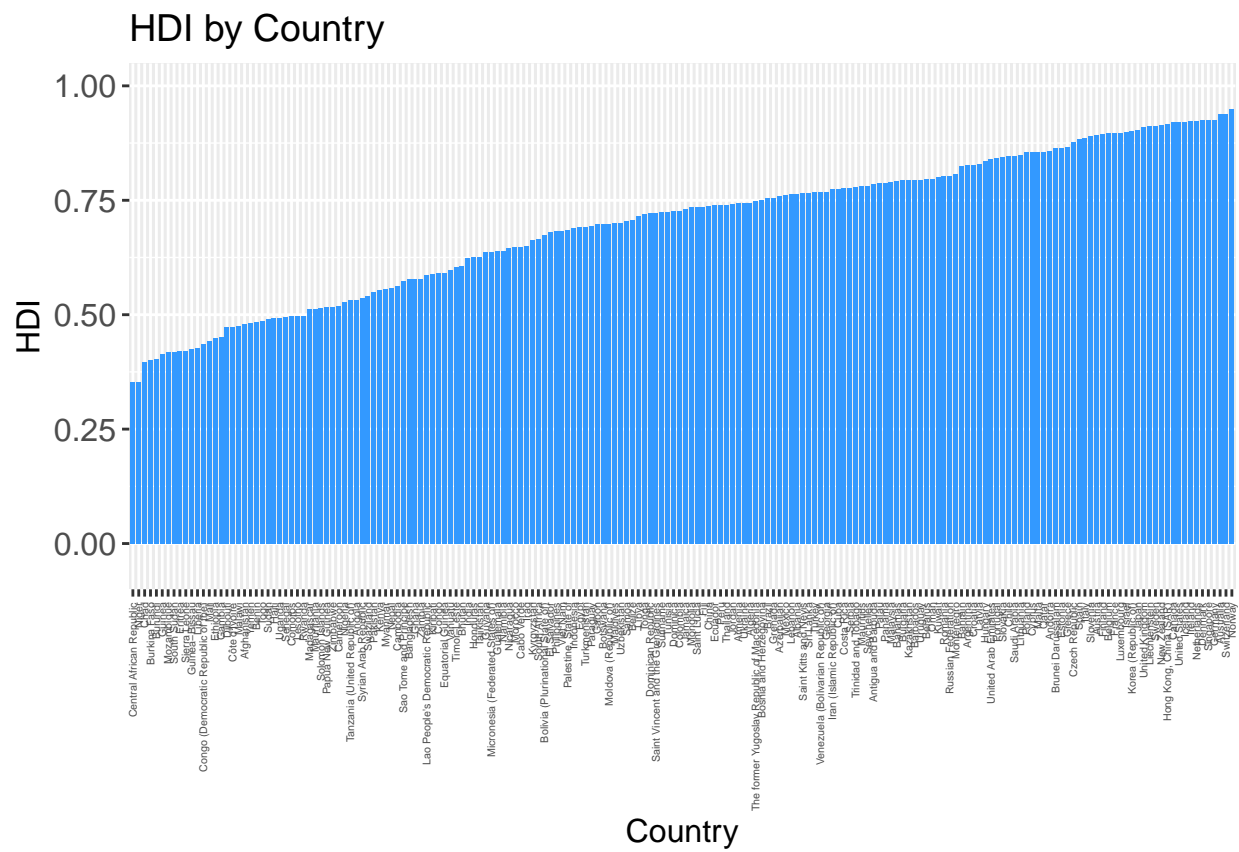
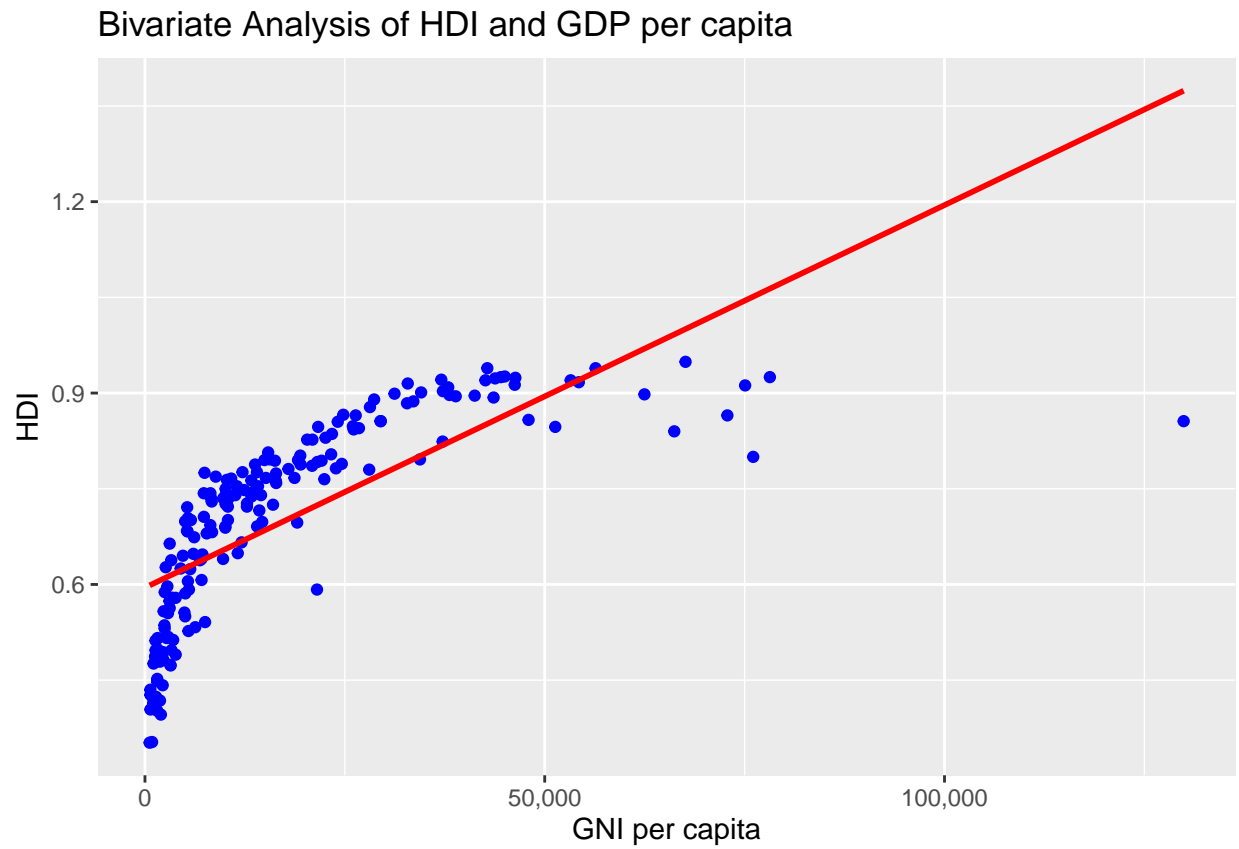
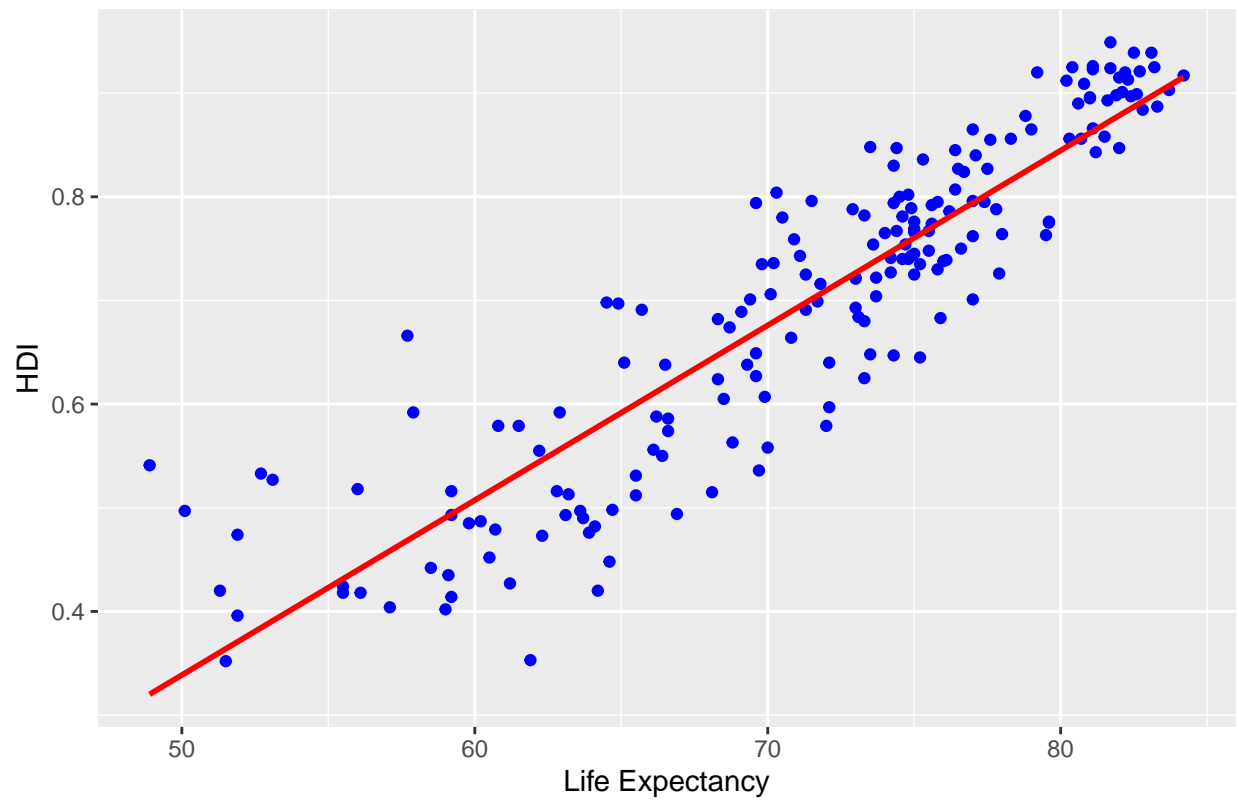


Figure 1: Graph of HDI by Country

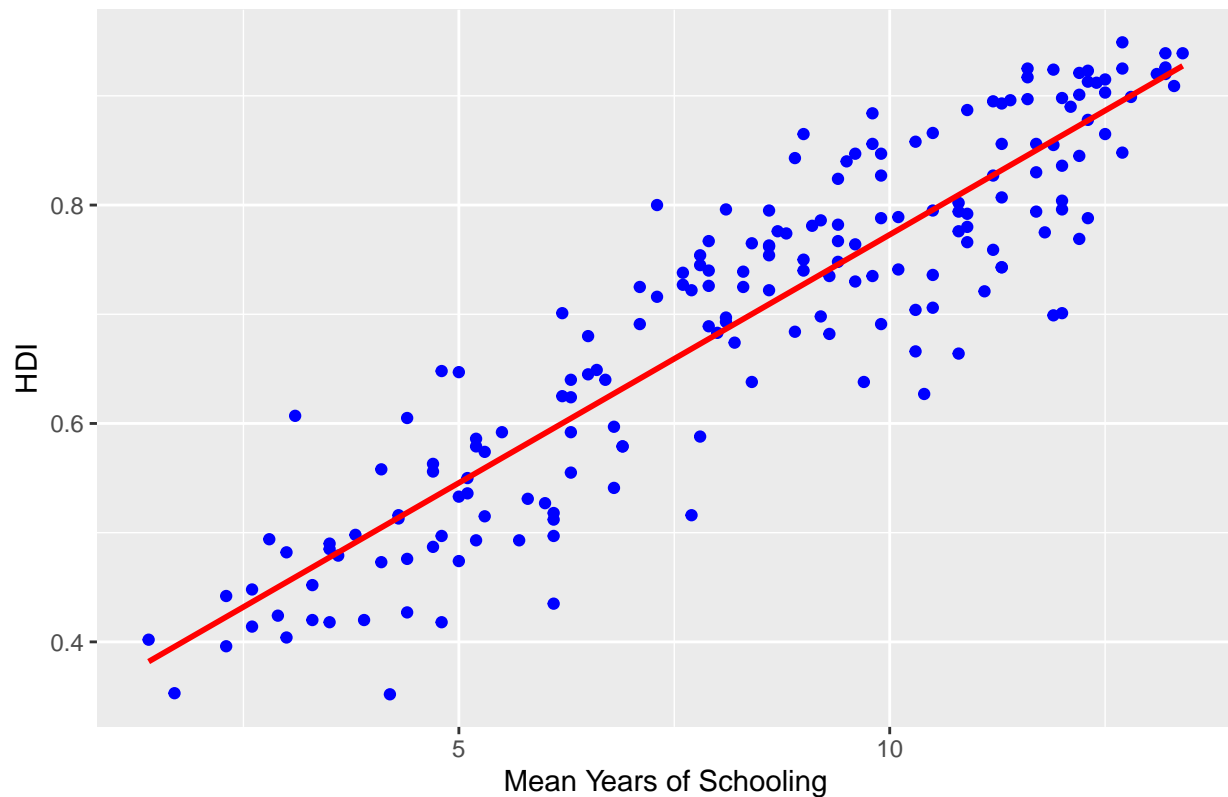
```
## Median : 10416
## Mean   : 17314
## 3rd Qu.: 23517
## Max.   :129916
```



Bivariate Analysis of HDI and Life Expectancy



## Bivariate Analysis of HDI and Mean Years of Schooling



```
##
## Call:
## lm(formula = HDI ~ GNIperCapita + LifeExpectancy + MeanYearsSchooling,
##     data = new_hdi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.103486 -0.015370  0.005061  0.020509  0.080748
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -9.866e-02  2.702e-02  -3.651  0.000341 ***
## GNIperCapita    1.518e-06  1.658e-07   9.153  < 2e-16 ***
## LifeExpectancy  8.008e-03  4.636e-04  17.274  < 2e-16 ***
## MeanYearsSchooling 2.387e-02  1.201e-03  19.879  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03274 on 184 degrees of freedom
## Multiple R-squared:  0.9562, Adjusted R-squared:  0.9555
## F-statistic: 1338 on 3 and 184 DF, p-value: < 2.2e-16
```

Table 1: Regression Summary

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0987	0.0270	-3.6508	3e-04
GNIperCapita	0.0000	0.0000	9.1528	0e+00
LifeExpectancy	0.0080	0.0005	17.2738	0e+00
MeanYearsSchooling	0.0239	0.0012	19.8788	0e+00

## Analysis:

The first graph shows a scatterplot of HDI on the y-axis against GNI per capita on the x-axis. The blue dots represent the individual countries, while the red line shows the best fit linear regression line. The graph shows a positive linear relationship between HDI and GNI per capita, with higher levels of GNI per capita associated with higher levels of HDI. The slope of the regression line indicates that for every increase in GNI per capita, the HDI score is expected to increase by a certain amount.

The second graph shows a scatterplot of HDI on the y-axis against Life Expectancy on the x-axis. The blue dots represent the individual countries, while the red line shows the best fit linear regression line. The graph shows a positive linear relationship between HDI and Life Expectancy, with higher levels of Life Expectancy associated with higher levels of HDI. The slope of the regression line indicates that for every increase in Life Expectancy, the HDI score is expected to increase by a certain amount.

The third graph shows a scatterplot of HDI on the y-axis against Mean Years of Schooling on the x-axis. The blue dots represent the individual countries, while the red line shows the best fit linear regression line. The graph shows a positive linear relationship between HDI and Mean Years of Schooling, with higher levels of Mean Years of Schooling associated with higher levels of HDI. The slope of the regression line indicates that for every increase in Mean Years of Schooling, the HDI score is expected to increase by a certain amount.

The bivariate analysis plots show the relationship between each independent variable and HDI while controlling for other factors. All three bivariate analyses show a positive and statistically significant relationship between each independent variable and HDI. Specifically, countries with higher GNI per capita, longer life expectancy, and more years of schooling tend to have higher HDI scores.

The multivariate analysis uses linear regression to examine the relationship between HDI and all three independent variables simultaneously. The results of the multivariate analysis show that all three independent variables are statistically significant predictors of HDI. Specifically, a one-unit increase in GNI per capita is associated with a 0.0000 increase in HDI, a one-year increase in life expectancy is associated with a 0.0080 increase in HDI, and a one-year increase in mean years of schooling is associated with a 0.0239 increase in HDI.

Overall, the results support our hypothesis that countries with higher GDP per capita, longer life expectancy, and more years of schooling tend to have higher HDI scores. The multivariate analysis confirms the bivariate findings, showing that all three independent variables are important predictors of HDI when controlling for each other. The graph that combines the three bivariate analyses provides a clear visual representation of the positive relationships between each independent variable and HDI.

```
##
## =====
##               Dependent variable:
##            -----
##                   HDI
## -----
## GNIperCapita      0.00000***
##                   (0.00000)
##
```

```

## LifeExpectancy          0.008***
##                          (0.0005)
##
## MeanYearsSchooling       0.024***
##                          (0.001)
##
## Constant                -0.099***
##                          (0.027)
##
## -----
## Observations             188
## R2                       0.956
## Adjusted R2              0.955
## Residual Std. Error      0.033 (df = 184)
## F Statistic              1,338.171*** (df = 3; 184)
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01

```

## Conclusion:

This output is the result of a linear regression model with HDI (Human Development Index) as the dependent variable and GNIperCapita, LifeExpectancy, and YearsOfSchooling as independent variables.

The table shows the estimated coefficients for the model, along with their standard errors and corresponding p-values. The coefficient for GNIperCapita is significant at the 0.01 level, indicating that there is a strong positive relationship between HDI and GNIperCapita. The coefficients for LifeExpectancy and YearsOfSchooling are also significant at the 0.01 level, indicating that they have a positive influence on HDI. The constant term is also significant, indicating that there are other factors that contribute to HDI beyond the three independent variables in the model.

The F-statistic tests the overall significance of the model, with a higher value indicating a better fit. The F-statistic in this case is very high, with a p-value less than 0.01, indicating that the model is statistically significant.

The analysis confirms the hypotheses that countries with higher GNI per capita, longer life expectancy, and more years of schooling have higher HDI scores. The findings suggest that economic, social, and health indicators play a crucial role in a country's overall human development. The study has important implications for policymakers, as it highlights the need to focus on improving these indicators to promote human development in countries.