Economic Growth & Subjective Well-Being

Zack Baker
Department of Statistics
Florida State University
December 10, 2020

Abstract

This paper seeks to examine the economic determinants of country-level happiness. Using lasso regression and backward stepwise variable selection, I develop a model for “happiness scores” of countries around the world using five economic development indicators. The model is similar to one of the most popular indices of human development, demonstrating the link between economic development and personal well-being. The contributions of each variable to happiness scores are found to be constant across developed and developing countries. Devoting greater resources to health, education, and economic opportunities for residents are potential policy options to increase overall happiness.

1. Introduction

Gross domestic product (GDP) is the most widely-used measure of economic production in the world today. Economists define GDP as the total monetary value of all finished goods and services produced within a country's borders during a specified time period. GDP has long enabled policymakers and central banks to judge whether an economy is expanding or contracting, as well as predict threats such as recessions and rampant inflation. The popularity and utility of the GDP measure have led it to become the standard indicator of economic development in the international community.

There are several advantages of GDP as a measure of development. GDP is a single, concrete number that is simple to compute and objectively comparable across countries. This allows economists to compare the development of countries across different cultures and economic structures. GDP is also an indicator of growth in jobs and income, which are known to be important in determining individual quality of life. Perhaps most importantly, GDP is highly correlated with other factors that are important to human welfare. In most countries, higher levels of per capita GDP occur hand-in-hand with other improvements to quality of life, such as higher levels of life expectancy, and lower levels of infant mortality and inequality.

While GDP is undoubtedly a useful measure of development and overall standard of living, it has a number of shortcomings. First, GDP is a measure of economic production, so it does not consider the intrinsic value of leisure time. For example, the per capita GDP of the United States is larger than that of Germany, but we cannot say that US citizens have a higher average standard of living because Americans work many more hours per year than their German counterparts [2]. Secondly, while GDP can tell us the total amount spent on different factors, such as health and education, it does not include the actual levels of these factors within a country. For instance, GDP includes the cost of purchasing pollution-control equipment but does not consider whether the air and water are actually cleaner or dirtier [2]. Finally, because GDP per capita is only an average, it provides no insight into the welfare of particular demographics or subgroups within a country [6]. When GDP per capita rises by 5%, for example, it might mean that incomes for everyone in a country have risen by 5%, or that the GDP of some groups has risen by more while that of others has risen by less – or even fallen!

The weaknesses of GDP as a development measure have led to the creation of alternative development measures that use a variety of factors to measure quality of life in a holistic way. One of the most popular alternative measures is the Human Development Index (HDI), which was created by the United Nations in 1990. Like GDP, it is a single number meant to quantify the quality of life in a country. It is calculated as the geometric mean of GNI per capita, life expectancy, and average years of schooling [6]. The HDI has gained in popularity as it is both simple to calculate and more informative than considering GDP alone. Other development measures take a more complex approach, using quality of life frameworks with many indicators rather than single indices like GDP or HDI. For instance, the Social Progress Index (SPI) uses 54 different quality of life indicators, including access to water and sanitation, education, health, housing, and communication [6]. Many policy experts view this model as overly complex, so it has yet to gain much international recognition. Lastly, Gross National Happiness (GNH) is a framework developed by the South Asian country of Bhutan to better understand the well-being of its residents. Every five years, 8,000 Bhutanese residents are randomly selected to fill out a comprehensive 300-question survey that includes questions from domains including psychological well-being, health, time use, education, and culture [5]. The country's leaders take the survey results very seriously and attempt to align policy decisions with improving the GNH metric.

In this paper, I seek to extend the research on alternative development measures by exploring the relationships between country development indicators and "happiness scores" from around the world. The paper is organized as follows. Section 2 provides a description of the data sets used for the analysis. Section 3 explains the empirical strategy and model selection methods. Section 4 summarizes my overall results. Sections 5 and 6 compare my findings with the Human Development Index and analyze outliers from the final model. The final section concludes and provides several takeaways from this analysis.

2. Data

2.1 World Development Indicators

The first data set used in this analysis comes from the World Development Indicators data posted on the Kaggle website [8]. This information is published annually by the World Bank and includes over 1,400 annual indicators of economic development from every country in the world. The indicators include economic variables, such as GDP, household consumption, and trade amounts; health variables including life expectancy and mortality rates; as well as education variables such as school enrollment and literacy rates.

There were several challenges in working with this data set. The first is that there are many missing values. Countries are not required to report on each of these variables, so there are only a few observations for a number of the indicators. Variables were removed if more than 50% of the total observations were missing. Secondly, many of the variables track the same indicators using different units of measurement. For instance, one variable is GNI per capita expressed in US dollars, while another is GNI per capita expressed in each country's local area currency (LAC). This means that many indicators are linear combinations of one another. I chose one indicator for each variable and removed the additional measurements, ensuring the columns of the data matrix were linearly independent. Removing missing values and extra variables left a total of 88 potential predictors for the initial model.

2.2 World Happiness Report

The second data set is the 2015 World Happiness Report, also posted on Kaggle [9]. The World Happiness Report is a landmark survey of the state of global happiness published annually by the United Nations. The report ranks 158 countries around the world by their happiness levels and includes a “happiness score” from zero to ten for each participating country. These scores are derived from the Cantril ladder question on the Gallup World Poll, which asks respondents to first think of a ladder with the best possible life being a ten and the worst life being a zero, then to rate their current lives on that scale. These happiness scores serve as the response variable in my regression analysis.

2.3 Exploratory Data Analysis

After cleaning the data, I perform some exploratory analysis before moving forward with regression modeling. Figure 2A shows boxplots of country happiness scores separated by income classification. The income classification values were provided by the World Bank and can be found in Appendix A. The trend here is quite clear – residents of high-income countries are happier on average than residents of low-income countries.

FIGURE 2A – Happiness Scores by Income Classification

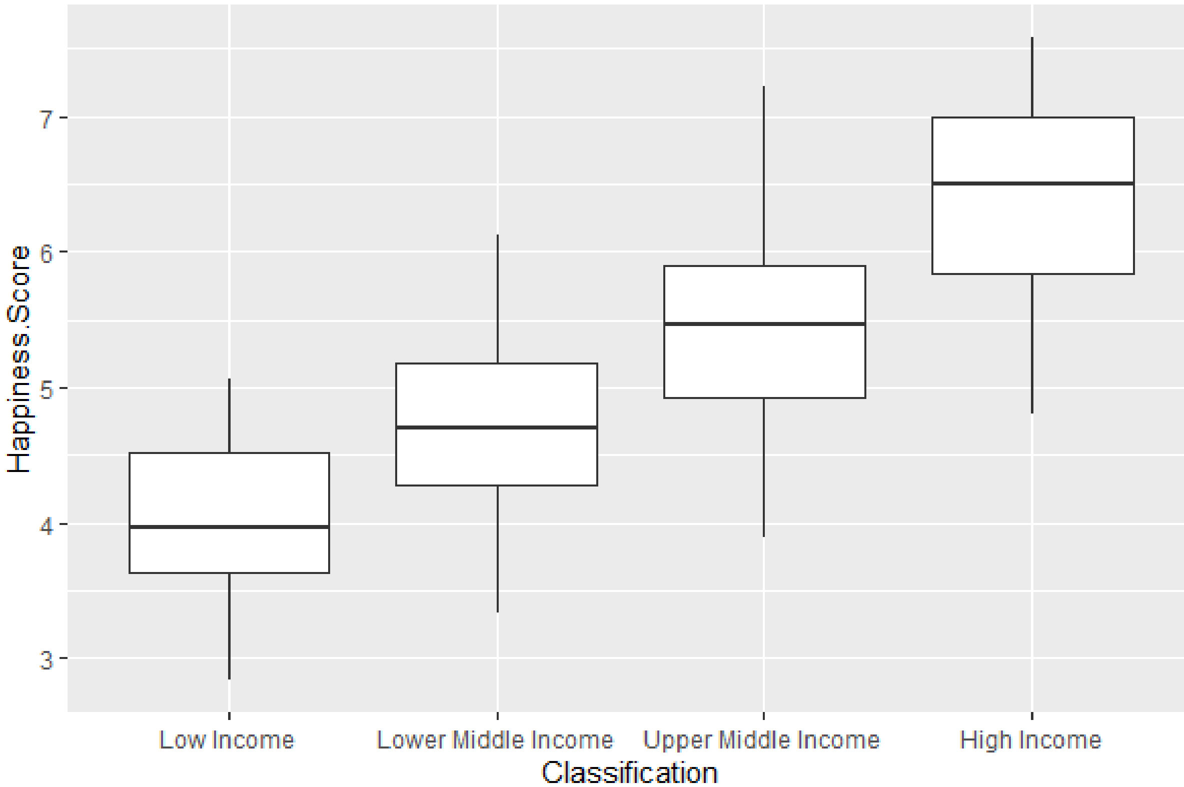


Table 2.1 shows the happiest countries according to the 2015 World Happiness Report. The happiest countries are Switzerland, Iceland, Denmark, Norway, and Canada, which are all high-income countries with gross national income (GNI) per capita values greater than $40,000 per year.

TABLE 2.1 – Most Happy Countries

Rank	Country	Classification	GNI Per Capita (2015)
1	Switzerland	High Income	$85,800
2	Iceland	High Income	$50,160
3	Denmark	High Income	$60,510
4	Norway	High Income	$93,110
5	Canada	High Income	$47,570

By contrast, Table 2.2 shows the least happy countries in the data set. The least happy countries are Rwanda, Benin, Syria, Burundi, and Togo, which are all low or lower-middle income countries with GNI per capita less than $1,200 per year. Clearly this initial analysis reveals a positive association between income and happiness levels.

TABLE 2.2 – Least Happy Countries

Rank	Country	Classification	GNI Per Capita (2015)
154	Rwanda	Low Income	$750
155	Benin	Lower-Middle Income	$1,180
156	Syria	Low Income	$681
157	Burundi	Low Income	$260
158	Togo	Low Income	$640

3. Estimation Strategy

3.1 Research Questions

I am interested in answering four main questions:

(1) Which development indicators are most predictive of country happiness scores?

(2) Do the contributions of different development indicators to happiness scores vary between developed and developing countries?

(3) Which observations are not fit well by the model, and why might this be?

(4) Do the results point to potential policy measures that could enhance citizens' well-being?

3.2 Approach

My overall approach to the analysis is to create a linear regression model using the variables from the World Development Indicators to predict country happiness scores from the World Happiness Report. I begin by using lasso regression to reduce the number of potential predictor variables. Lasso regression is a form of linear regression that uses employs the sparsity-inducing L1 penalty, meaning some of the coefficients are set to zero and removed from the model. Lasso regression is known to perform well with highly correlated variables, meaning it can help to identify the most important contributors to happiness scores even if the predictors are closely related. Once the potential predictors are chosen by the lasso model, I use power transformations to ensure the predictor variables have a linear relationship with happiness scores. I then proceed with backward stepwise variable elimination using AIC as the information criterion to select the final model.

3.3 Initial Predictors

After running the lasso regression model using the 88 predictor variables and happiness scores as the response, the procedure specifies the model

$$\begin{aligned} Happiness\_Score &= \beta_0 + \beta_1 GNI\_Per\_Capita + \beta_2 Gov\_Ed\_Spending + \beta_3 Internet\_Use \\ &+ \beta_4 Life\_Exp + \beta_5 Nurses\_Midwives + \beta_6 Air\_Pollution \\ &+ \beta_7 Pop\_Growth + \beta_8 Urban\_Pop + \varepsilon. \end{aligned}$$

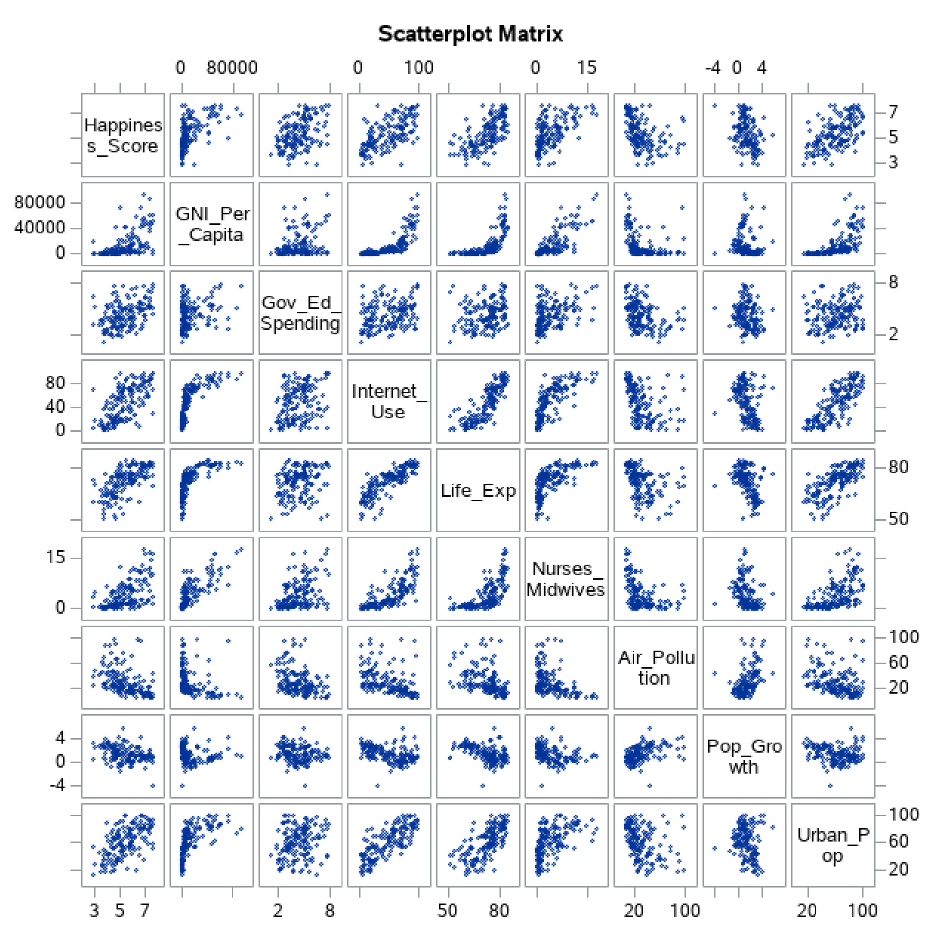
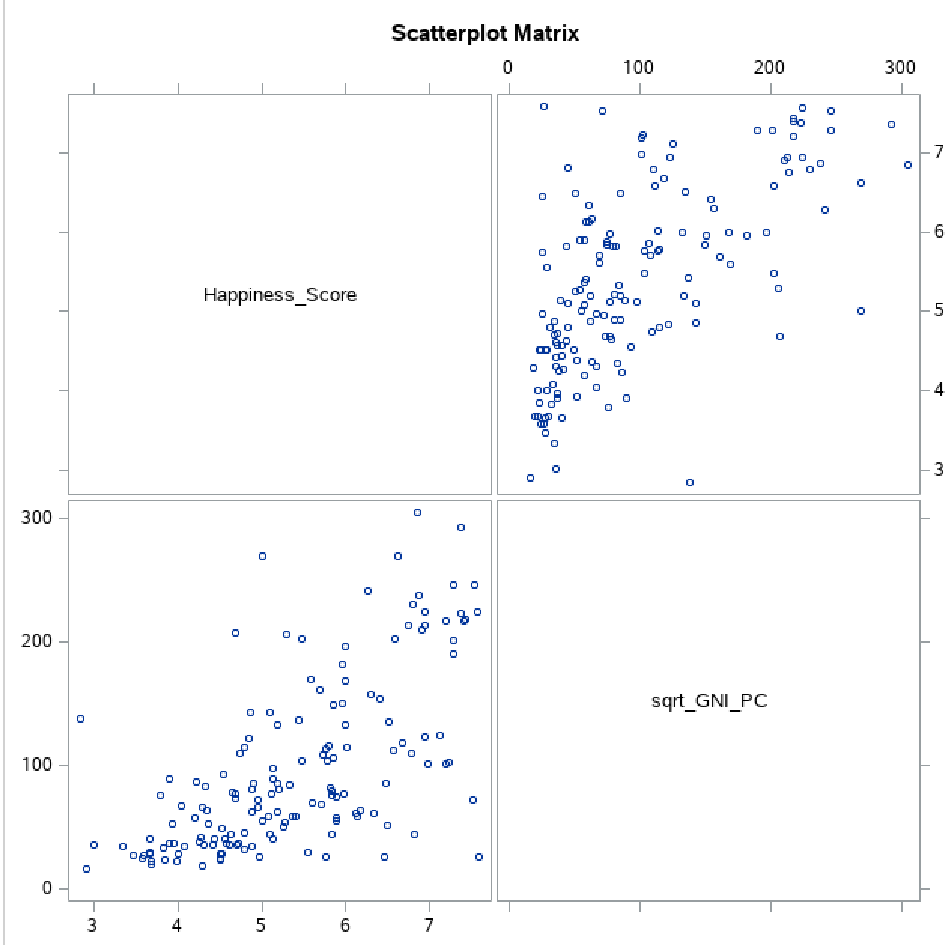
The initial predictors are GNI per capita, government expenditure on education, the percentage of population using the internet, life expectancy, the number of nurses and midwives per 1,000 people, the amount of air pollution, the population growth rate, and the percentage of population living in urban areas. The definitions of each of these variables can be found in Table 3.1, and basic descriptive statistics for each variable are in Appendix B. The other 80 predictors are dropped from the model.

TABLE 3.1 – Initial Predictor Variables	
Variable Name	**Definition**
GNI_Per_Capita	GNI per capita (current US$)
Gov_Ed_Spending	Total government expenditure on education (% of GDP)
Internet_Use	Individuals using the internet (% of population)
Life_Exp	Life expectancy at birth (years)
Nurses_Midwives	Nurses and midwives (per 1,000 people)
Air_Pollution	PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)
Pop_Growth	Population growth (annual %)
Urban_Pop	Urban population (% of total population)

3.4 Variable Transformation

The scatterplot matrix of the initial variables is shown in Figure 3A, with the top row and first column assigned to the response variable, happiness score. From these plots we can see that the GNI per capita variable likely needs transformation, as it does not have a linear relationship with the happiness score.

I use the Box-Cox method to select the optimal transformation for GNI per capita. The closest “standard” power transformation value is 1/2, suggesting a square root transformation. I proceed with the suggested transformation and obtain the scatterplot for happiness score vs. the square root of GNI per capita as shown in Figure 3B. The relationship appears to be much more linear after the transformation.

FIGURE 3A – Initial Scatterplot Matrix	FIGURE 3B – Scatterplot Matrix of Happiness Score vs. $\sqrt{\text{GNI per Capita}}$
	

3.5 Interaction Terms

One of my primary research questions is whether the impact of different economic indicators on happiness varies between developed and developing countries. Before fitting the initial model, I create a dummy variable called "High_Income" that equals one if a country is classified as a high income country by the World Bank (see Appendix A), and zero otherwise. My interest is in two potential variable interactions. First, I hypothesize that the impact of GNI per capita on happiness scores differs between developed and developing countries. When a country has a large GNI per capita, a small increase in average income may have only a small impact on its population's happiness, while at lower levels of GNI per capita, a few dollars could lead to a much larger increase in happiness. Secondly, the impact of population growth on happiness may also differ between developed and developing countries. Developed countries may have more resources to handle a growing population without much impact on happiness, while population growth in developing countries may put strain on available resources and infrastructure, making people less happy. To test these hypotheses, I include interaction terms between High_Income and GNI per capita and between High_Income and population growth.

4. Analysis

4.1 Initial Model

After transforming the GNI per capita variable and adding the interaction terms mentioned in the previous section to the lasso regression model, the initial model is:

$$Happiness\_Score = \beta_0 + \beta_1\sqrt{GNI\_Per\_Capita} + \beta_2 High\_Income * GNI\_Per\_Capita + \beta_3 Gov\_Ed\_Spending + \beta_4 Internet\_Use + \beta_5 Life\_Exp + \beta_6 Nurses\_Midwives + \beta_7 Air\_Pollution + \beta_8 Pop\_Growth + \beta_9 High\_Income * Pop\_Growth + \beta_{10} Urban\_Pop + \varepsilon.$$

The model is summarized in Table 4.1 below. The R-squared value is 0.5499 and the residual standard error (RSE) is 0.8035. Most of the coefficient estimates are not statistically significant, but government spending on education, life expectancy, population growth, and the percentage of people living in urban areas are all significant. These four variables all have their expected signs, as we expect that greater spending on education, longer life expectancies, and higher levels of urbanization should be associated with higher happiness scores, while population growth might decrease happiness due to the added burden on resources and infrastructure. The fact that most of the coefficients are not significant might indicate that some of the variables are correlated. We generally assume that a variance inflation factor (VIF) greater than five indicates a high degree of correlation between predictors, and there are three estimates with VIF values greater than five in the initial model. These results may suggest that we can remove some of the variables to reduce the amount of multicollinearity and create a better model.

TABLE 4.1 – Initial Model

Coefficients	**Estimate**	**Std. Error**	**t-value**	**Pr(>|t|)**	**VIF**
Intercept	1.5126	1.1181	1.35	0.1783	0
$\sqrt{\text{GNI\_Per\_Capita}}$	0.0046	0.0044	1.04	0.2983	22.3913
High_Income*GNI_Per_Capita	-0.0019	0.0025	-0.75	0.4572	12.0999
Gov_Ed_Spending	0.1504	0.0501	3.00	0.0032***	1.2755
Internet_Use	-0.0032	0.0060	-0.54	0.5894	6.7330
Life_Exp	0.0372	0.0163	2.28	0.0241**	3.9927
Nurses_Midwives	0.0207	0.0302	0.69	0.4942	3.6223
Air_Pollution	-0.0044	0.0042	-1.04	0.2989	1.6868
Pop_Growth	-0.1364	0.0732	-1.86	0.0646*	2.1836
High_Income*Pop_Growth	0.1579	0.1237	1.28	0.2039	2.5459
Urban_Pop	0.0087	0.0052	1.68	0.0956*	3.1558
	RSE	**0.8035**			
	R-Squared	**0.5499**			

Notes: ***p < .01. **p < .05. *p < .1.

4.2 Final Model

To select the predictors for the final regression model, I use backward stepwise variable selection with AIC as the information criterion. This selection method begins with a model containing all the potential predictors, then removes the variable with the least explanatory power according to the information criterion. This process is repeated, removing one variable at a time until the criterion no longer improves. Using the stepwise procedure with all of the variables from the initial model as potential predictors, the following model is specified:

$$Happiness\_Score = \beta_0 + \beta_1\sqrt{GNI\_Per\_Capita} + \beta_2 Gov\_Ed\_Spending + \beta_3 Life\_Exp + \beta_4 Pop\_Growth + \beta_5 Urban\_Pop + \varepsilon.$$

Using backward selection, the model reduces to only five predictors: the square root of GNI per capita, government spending on education, life expectancy, population growth, and urban population. Interestingly, the two interaction terms specified in the initial model have very large p-values and are removed from the final model.

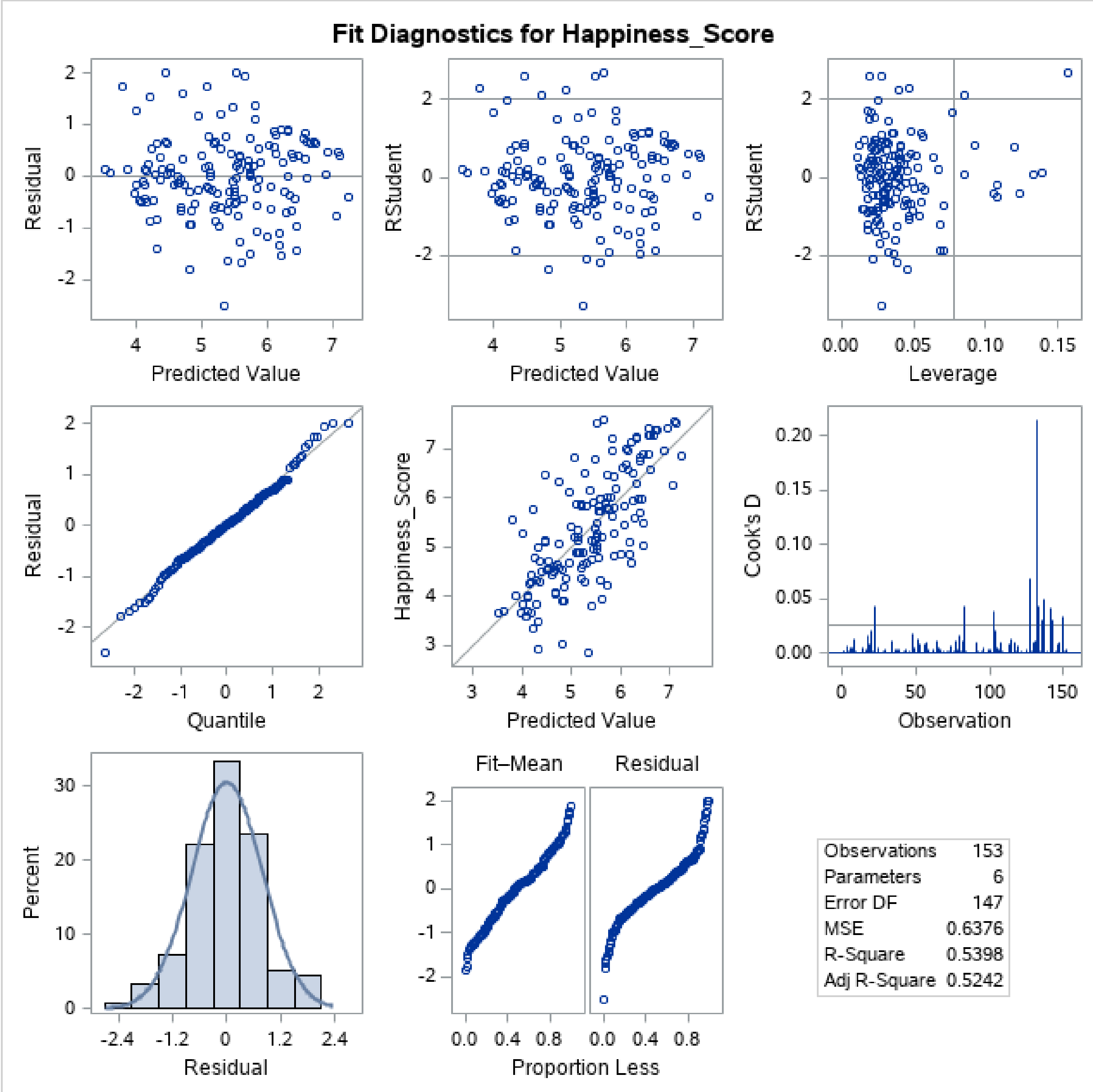
The final model is shown in Table 4.2. All of the variables are statistically significant at the 10% level, and population growth is the only predictor that is not significant at 5%. Each variable has its expected sign. Increasing GNI per capita, government educational spending, life expectancy, and urban population should be associated with higher happiness scores, while population growth should lead to a decrease in happiness. The variance inflation factors are much smaller than those from the original model, so removing some of the predictors seems to have reduced the correlation between variables. The R-squared value only fell by about 1% (to 0.54) from the initial model, despite the removal of five variables.

TABLE 4.2 – Final Model					
Coefficients	**Estimate**	**Std. Error**	**t-value**	**Pr(>|t|)**	**VIF**
Intercept	1.0101	0.9834	1.03	0.3060	0
$\sqrt{\text{GNI\_Per\_Capita}}$	0.0033	0.0017	2.01	0.0461**	3.2194
Gov_Ed_Spending	0.1677	0.0463	3.63	0.0004***	1.1010
Life_Exp	0.0402	0.0148	2.71	0.0075***	3.3310
Pop_Growth	-0.1057	0.0567	-1.86	0.0645*	1.3270
Urban_Pop	0.0092	0.0045	2.03	0.0438**	2.4037
	RSE	**0.7985**			
	R-Squared	**0.5398**			
Notes: ***p < .01. **p < .05. *p < .1.					

4.3 Final Model Diagnostics

The diagnostic plots for the final model are shown in Figure 4A. Looking at the residual plots in the upper left, it seems as though the residuals are fairly evenly scattered about the plot. The residuals fall approximately on the normal line in the Q-Q plot, and the residual histogram approximates a normal distribution. These plots do not reveal any obvious violations of our normality assumptions. The Cook's distance plot looks somewhat concerning, as one point clearly has a much larger Cook's distance than the rest. This observation corresponds to Switzerland, which happens to be the happiest country in the world (see Table 2.1). Switzerland's happiness score is much larger than predicted by the model. I removed this observation and re-ran the stepwise model, with little change in results. The same variables are selected for the final model and their p-values change only slightly, so I opted to keep Switzerland in the data set. The final model seems to satisfy the usual assumptions of linear regression.

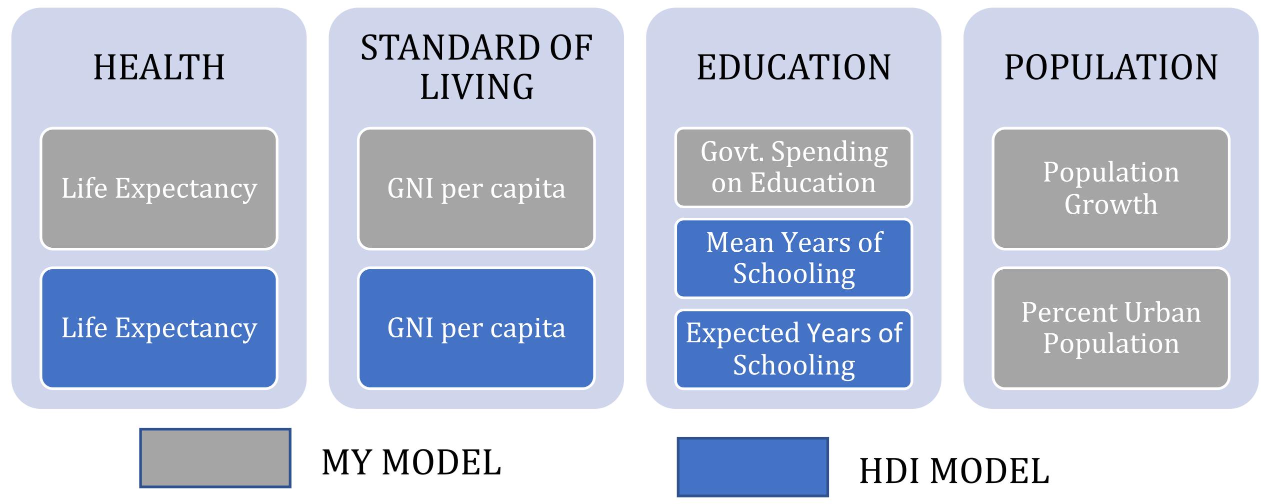
FIGURE 4A – Final Model Diagnostics



5. Comparison with the Human Development Index (HDI)

After selecting the final model, I am interested in comparing it to one of the most popular economic development indices, the Human Development Index (HDI). The comparison is summarized in Figure 5A.

FIGURE 5A – Comparison with Human Development Index



I am pleasantly surprised to see a number of similarities between the two models. Both my model and the HDI use life expectancy to quantify health and GNI per capita to capture standard of living. Both models capture the impact of education, though through different indicators. The HDI factors in the mean and expected years of schooling for citizens of a country, while my model uses the amount of government spending on education. The biggest difference between the two models is that my model captures the impact of overall population trends on happiness, while the HDI does not. This seems sensible, as the purpose of the HDI is to understand the development of individuals rather than countries as a whole. My model has no such restriction, as it seems possible that high-level population trends could have an impact on the happiness of individuals. This leads me to believe that the differences between the two models are understandable given the unique goals of each. It is important to note that my model and the HDI do not claim to capture the same thing. The HDI indices are meant to model human development, while my model seeks to understand human happiness. The similarities between the two measures provide some evidence that these two things are tightly interwoven.

6. Outlier Analysis

Next, we can examine outliers and influential points to gain insight into the weaknesses and drawbacks of the final model. The Cook's distance plot is shown in Figure 6A. While I do not discuss all of the influential points from the model, three characteristic observations are shown in Table 6.1.

FIGURE 6A – Cook's Distance Plot

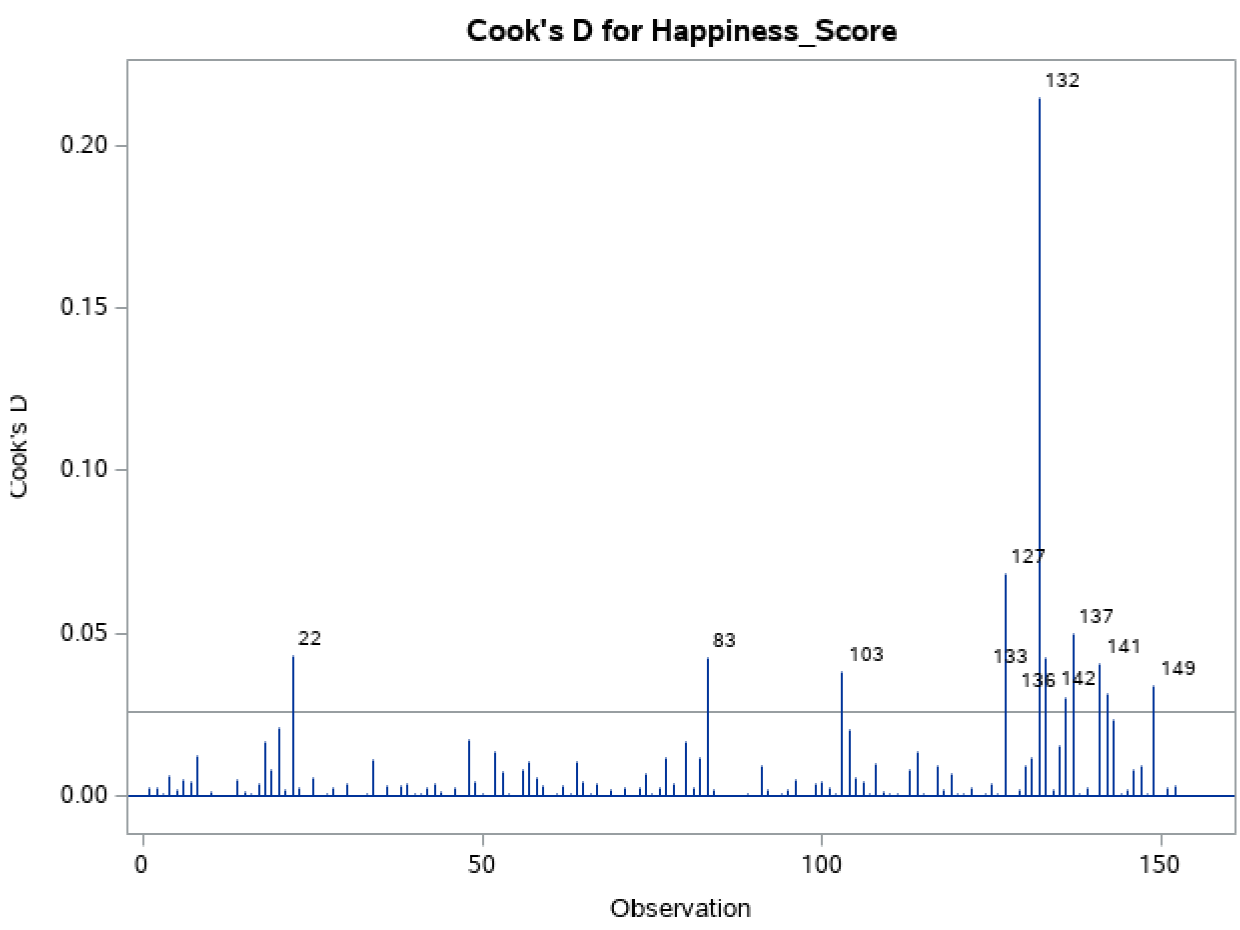


TABLE 6.1 – Outliers and Extreme Values			
Observation	**Country**	**Happiness Score**	**Predicted Happiness Score**
132	Switzerland	7.59	5.66
133	Syria	3.01	4.82
149	Venezuela	6.81	5.07

The observation with the largest Cook's distance represents Switzerland, whose citizens are much happier than predicted. It has a fitted happiness score of 5.66 and actual score of 7.59, making it the happiest country in the data set. Switzerland is known for having a large GDP per capita, long life expectancy, beautiful natural environment, effective healthcare system, and good government [7]. It really is an outlier in the world when it comes to many of these things! However, my model is clearly unable to capture the subtle impact of factors like good governance and natural beauty on happiness. This demonstrates a potential drawback of using a model with a small number of predictors.

Observation 133, corresponding to Syria, has the second-highest Cook's distance. The people of Syria are much less happy than predicted, and it is one of the least happy countries overall. This unhappiness can likely be attributed to the civil war the country has been engaged in since 2011. The situation has evolved into an international crisis, as huge numbers of people have been displaced including four million refugees that have fled to neighboring countries [1]. It is not surprising that Syria is an outlier, as the model cannot take this kind of extreme domestic conflict into account.

Observation 149 represents Venezuela, which is an interesting case. Venezuela is internationally known for having a deeply-ingrained culture of happiness, and its people are generally cheerful and upbeat. The nation has even instituted a national office for happiness! However, the country has been plagued with problems such as rampant inflation, unemployment, and poor governance in recent years [4]. Given these issues, it is surprising that Venezuela is much happier than predicted by the model, with a predicted happiness score of 5.07 and an actual score of 6.81. Unfortunately, it seems that the domestic crisis in Venezuela has finally taken a toll on its population's happiness. Looking at its happiness scores over time, Venezuela fell from ranked 5th in the world in 2012 to 102nd by 2018. Shortly after the data for this analysis was collected (in 2015) Venezuela's happiness score dropped tremendously. This may be evidence of a potential lag between changes in economic indicators and their eventual impact on happiness.

7. Conclusion

In this project, I attempt to model the relationship between country development indicators and happiness scores from every country in the world using linear regression. I begin by implementing lasso regression to reduce the number of potential predictors. After performing variable transformation, I use backward stepwise variable selection to make the final model selection. The final model uses just five predictors related to health, income, education, and overall population trends. As shown in section 5, these are some of the same predictors used to model human development in the HDI, so clearly there is a relationship between human happiness and the level of economic development of a country.

Interestingly, the contributions of these different factors to happiness do not vary between developed and developing countries. I use interaction terms to see whether the contributions of both GNI per capita and population growth to happiness vary between high and low income countries, and both are removed from the stepwise elimination model. Future study might be done to see if there are other variables that have a significant interaction with income.

My final research question focuses on policy options to increase happiness. This work demonstrates that devoting more resources to health and education, as well as increasing economic opportunities for residents are fruitful policy avenues. Perhaps more importantly, this analysis shows us that many economic indicators are highly correlated and the relationships between them can be relatively complex. Decision-makers should be cautious and consider the relationships that may exist between economic variables before implementing new policy. For instance, this research shows that increasing the percentage of people living in urban areas seems to have a positive impact on happiness levels. However, in some developing countries there are insufficient resources to sustain large urban populations, so policies designed to increase urbanization rates should take this reality into consideration. The fact that we can model human happiness using only a handful of economic indicators does not imply that happiness is a simple problem with an easy solution.

References

[1] Garfield, Leanna. "This is the unhappiest country in the world." *Business Insider*. 1 Dec. 2015. URL: https://www.businessinsider.com/the-unhappiest-country-in-the-world-is-syria-2015-11.

[2] "How well GDP measures the well-being of society." *Khan Academy*. URL: https://www.khanacademy.org/economics-finance-domain/macroeconomics/macro-economic-indicators-and-the-business-cycle/macro-limitations-of-gdp/a/how-well-gdp-measures-the-well-being-of-society-cnx.

[3] Kesebir, Selin. "When Economic Growth Doesn't Make Countries Happier." *Harvard Business Review*. 25 Apr. 2016. URL: https://hbr.org/2016/04/when-economic-growth-doesnt-make-countries-happier.

[4] Latouche, Miguel Angel. "How Venezuelans became some of the unhappiest people in the world." *Quartz*. 1 May 2018. URL: https://qz.com/1267005/venezuelas-crisis-has-made-it-one-of-the-unhappiest-countries-in-the-world/.

[5] McCarthy, Julie. "The Birthplace of 'Gross National Happiness' Is Growing A Bit Cynical." *NPR*. 12 Feb. 2018. URL: https://www.npr.org/sections/parallels/2018/02/12/584481047/the-birthplace-of-gross-national-happiness-is-growing-a-bit-cynical

[6] Prasad, Manish Kumar and Andre Castro. "Is GDP an adequate measure of development?" *International Growth Centre*. 17 Oct. 2018. URL: https://www.theigc.org/blog/is-gdp-an-adequate-measure-of-development/

[7] Smith, Oliver. "Why is Switzerland so happy?" *The Telegraph*. 24 Apr. 2015. URL: https://www.telegraph.co.uk/travel/destinations/europe/switzerland/articles/Why-is-Switzerland-so-happy/.

[8] World Development Indicators. *The World Bank*. Kaggle, 1 May 2017. URL: https://www.kaggle.com/worldbank/world-development-indicators.

[9] World Happiness Report. *The United Nations*. Kaggle, 26 Nov. 2019. URL: https://www.kaggle.com/unsdsn/world-happiness.

Appendix

APPENDIX A – World Bank Income Classifications	
Threshold	**GNI per Capita**
Low Income	< $1,026
Lower-Middle Income	$1,026 - $3,995
Upper-Middle Income	$3,996 - $12,375
High Income	> $12,375

APPENDIX B – Descriptive Statistics for Initial Predictor Variables				
Variable	**Mean**	**Std. Dev.**	**Minimum**	**Maximum**
GNI_Per_Capita (US$)	14,307.46	19,249.28	260.00	93,110.00
Gov_Ed_Spending (%)	4.36	1.47	1.11	7.82
Internet_Use (%)	48.76	28.40	2.48	98.20
Life_Exp (years)	72.06	7.98	50.88	84.28
Nurses_Midwives (per 1,000 population)	4.25	4.11	0.05	17.86
Air_Pollution (micrograms per cubic meter)	30.35	20.19	6.06	97.43
Pop_Growth (%)	1.39	1.31	-3.91	5.79
Urban_Pop (%)	60.34	22.29	12.08	100.00