What Makes Us Happy? A Cross-Sectional Regression Analysis of the Determinants of Happiness Across Countries for the Years 2014-2018

Casey Smith
January 31st, 2022
Senior Comprehensive Project
Pledged

Introduction

Traditionally, the success of a country is determined by its levels of production, economic growth, and development. Meaning the most industrialized countries, with the highest GDPs, and most modern technologies are viewed as the most successful, because they have the 'best' economies. But in recent years, there has been a growing push to place more value in other measures of success, particularly the well-being and happiness of a country's citizens. After all, is there any value in the productivity of a country if its people are not happy? While focusing on well-being and happiness may seem like a deviation from the field of economics and subsequently a dismissal of an economy's relevance to its country's success, examining the determinants of happiness through economic analysis opens up an insightful conversation. The production and economic growth of a country is certainly important and also contributes to happiness, but there are several other factors that can also affect the well-being of the citizens of a country.This study will analyze the determinants of happiness across countries, with particular focus on social and environmental factors and their effects on happiness.

This paper will test the developed model by using a series of regressions and statistical tests to examine whether there is a relationship between the dependent variable, country happiness, and the independent variables, grouped into the categories production, equality, health, freedom, environment, despair/stress, education, and population characteristics. John Helliwell, one of the co-authors of the world happiness reports, notes that "it is important to have at least one main variable covering each of the main areas shown by research to be important" which is why the variables chosen for this study each fall into critical categories that can affect happiness (Smith and Helliwell, 2022). The data used in this study is cross-sectional across countries for the years 2014 through 2018. The results of these regressions will be analyzed to

break down the relationships between country happiness and the independent variables, and determine which variables have a noteworthy impact on the happiness of countries.

Theoretical Support

A nation aiming to maximize their happiness can also be seen as a nation aiming to maximize their utility. Utility is a term used to represent whatever value an individual, or even a country, is attempting to maximize through the decisions they make (Jensen, 1967, pg 164). Thus, the logic and model developed in this paper follow the ideas of utility theory, where each independent variable explored is a factor that can increase or decrease a country's utility, or happiness, and any policy recommendations based on the findings in the regression analyses will be choices that work to maximize happiness. The model and its variations created in this paper represent a utility function in which each independent variable's coefficient, if determined to be statistically significant, tell where the decision-maker's, in this case a country's, preference should be for choosing strategies that maximize happiness (Jensen, 1967, pg 164). For example, if it were found that the coefficients for the independent variables number of apples consumed and number of oranges consumed were both positive and statistically significant, meaning consuming either apples or oranges both increase happiness, a country would want to maximize this consumption to maximize happiness, the measure of utility. Additionally, if apples were found to have a higher positive coefficient than oranges, then the country would have a greater preference for adoption of the strategy to increase apple consumption, over increasing orange consumption, although either will increase utility, there is a greater preference for increasing apple consumption because it has a stronger impact on increasing utility or happiness. Moreover, if bananas were found to have a negative coefficient, a country may also want to adopt a policy to limit or prevent banana consumption as another method to maximize happiness.

The dependent variable in this study, average country happiness, and the independent variables used are also closely related to human development theory and the human development index (HDI). The HDI, first released in 1990, expands upon the traditional view of utility by providing another measurement of human well-being that focuses more on making sure individuals have ends like access to health, education, and goods, before focusing on access to means like income because the former will empower them to "achieve their desired state of being" (Stanton, 2007, pg 3). This approach is very similar to the one taken in this paper in which the emphasis is on looking at how a country's ability to ensure its citizens are healthy, safe, and free influences a desired state of being, in this case happiness.

Literature Review

John F. Helliwell, Richard Layard, and Jeffrey D. Sachs

World Happiness Report, 2019

The 2019 World Happiness Report presents global data on national happiness of 156 countries and the factors that influence it. In this report, countries are ranked according to their level of happiness calculated using pooled data from the Gallup World Poll (GWP) surveys for the years 2016 through 2018. This happiness index is determined based on responses to the Cantril ladder question which asks respondents to value their current lives based on a 0 to 10 scale, with 0 being the worst possible life and 10 being the best possible life (Helliwell et al., 2019, pg 15).

The authors used a pooled OLS regression to evaluate the relationship between average happiness across countries and six key variables which are Log GDP per capita, social support, healthy life expectancy at birth, freedom to make life choices, generosity, and perceptions of corruption. The dependent variable to represent average country happiness is the national

average response to the Cantril ladder question. The first independent variable, Log GDP per capita, is based on Purchasing Power Parity (PPP) and is adjusted to constant 2011 international dollars. Healthy life expectancy at birth is time series data from the years 2005, 2010, 2015, and 2016, with which interpolation and extrapolation were used to match the report's sample time period. The variable social support is the national average of respondents' binary answers (0 for no or 1 for yes) to the GWP question "if you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?" For the variable freedom to make life choices, the value is the national average binary response (0 for dissatisfied, 1 for satisfied) to the GWP question "are you satisfied or dissatisfied with your freedom to choose what you do with your life?" Generosity is measured using the residual from regressing the national average response to the GWP question "have you donated money to a charity in the past month?" Finally, perceptions of corruption is the national average of the binary responses (0 for no, 1 for yes) to two GWP questions which are "is corruption widespread throughout the government or not?" and "is corruption widespread within business or not?" For this variable some countries are missing data for the first question on government corruption and in these cases the question on corruption in business is used to represent the overall perception of corruption (Helliwell et al., 2019, pg 25).

Regression 1. In the results of the first regression, the adjusted R^2 value of 0.74 means that 74% of the variation in the national annual average response to the Cantril ladder question can be explained by the independent variables included in the regression, taking into consideration the number of independent variables. The results of this first regression are displayed in Table 1.

Log GDP per capita. The positive value for the coefficient of Log GDP per capita is expected and demonstrates that for every one percent increase in the GDP per capita, the national

annual average response to the Cantril ladder question increases by .00318 index points. This positive effect means that an increase in a country's GDP per capita increases the happiness of its people.

Social Support (SS). The coefficient of this variable has its expected positive sign and shows that as the response to the social support survey question increases by one unit, meaning the response changing from 0 to 1, the Cantril ladder increases by 2.422 index points. This demonstrates the relationship that when people have more support from friends and family, they are happier.

Table 1: Helliwell et al.'s Regression Results

Variables	GDPpc	SS	HLE	FLC	Gen	PC
Regression 1 Coefficients	0.318	2.422	0.033	1.164	0.635	-0.540
Robust Standard Errors	(0.066)***	(0.381)***	(0.01)***	(0.3)***	(0.277)**	(0.294)*

Adj. $R^2 = 0.74$

N = 1,516

***, **, and * indicate significance at the .01, .05, and .10 level respectively

---

Healthy Life Expectancy at Birth (HLE). This variable has a positive coefficient as expected, and its calculated value means that for every one year increase in the healthy life expectancy at birth, the national annual average response to the Cantril ladder question increases by 0.033 index points. This reflects that when people can be expected to live longer, they are also happier.

Freedom to Make Life Choices (FLC). The coefficient for this variable is positive as expected and shows that when the response to the GWP survey question on freedom to make life choices changes from 0 to 1, the Cantril ladder can be expected to increase by 1.164 index points. This positive relationship between the independent and dependent variable means that when people feel more free to choose what they want to do with their lives, they feel happier.

Generosity (Gen). This coefficient is positive as expected and demonstrates that when the regressed residual of the response to the survey question on generosity increases by one unit, the Cantril ladder increases by 0.635 index points. This means that when a country has people that are more generous, it also has people that are happier.

Perceptions of Corruption (PC). The coefficient of this variable is negative as expected and shows that when the response to the survey question on the presence of corruption increases, going from 0 to 1, the Cantril ladder decreases by 0.540 index points. This reflects that when people feel there is more corruption occurring in their country, they are less happy.

Christian Bjornskov

The Happy Few: Cross-Country Evidence on Social Capital and Life Satisfaction, 2003

In this study, Christian Bjornskov investigates what influences people's happiness across countries by looking at several variables to represent the broader categories of economic possibilities, inequality, uncertainty, democratisation, and social capital. The author notes that in past studies, unemployment and inflation have been found to be large contributors to a decrease in happiness; however, studies have found differing results on whether increases in income lead to more or less happiness. This means that while unemployment and inflation could lead to less happiness because of lost income, their negative effects more certainly reflect how stress and uncertainty impact an individual's well-being, perhaps regardless of income. Additionally,

Bjornskov explains how higher education and democratic participation rights have demonstrated a positive effect on happiness, leading to his paper's particular focus on the effect of social capital on people's happiness (Bjornskov, 2003, pg 4). To do so, Bjornskov creates a social capital factor derived from a component analysis that uses the variables generalized trust, civic participation, and perceived corruption, hypothesizing that higher levels of social capital cause higher levels of happiness (Bjornskov, 2003, pg 6).

The data used for this study is for 32 countries in Europe, the Americas, and Asia. The relationships between 12 independent variables and the dependent variable, life satisfaction, are examined in a series of estimated regressions presented in this study. Life satisfaction is measured as an average response to the question "all things considered, how satisfied are you with your life as a whole these days?" using a 1-10 scale with 1 being dissatisfied and 10 being satisfied (Inglehart et al., 1998, pg (186). Data for this variable, and the independent variables generalized trust and civic participation come from Inglehart et al. which uses data from 1993, but Bjornskov notes that these indicators remain relatively stable. WorldBank is the source of data for the independent variables GNI per capita adjusted for PPP, trade volume as a percentage of GDP, inflation, population size, life expectancy at birth, percent of the population with both a primary and secondary education, Gini coefficients, and unemployment rates. CIA data was used to fill in any missing values for those indicators. Data on the perceived corruption index was used from Transparency International, and data for the variables economic freedom and the Gastil index came from Freedom House. All of the data for these variables is for the year 2002 (Bjornskov, 2003, pg 5).

Bjornskov presents a series of 10 regressions, which include the independent variables GNI per capita, Gini coefficient, and inflation each time, and various combinations of the other

independent variables. In each regression the coefficients for GNI per capita and gini coefficient were found to be positive with statistically significant t-statistics at either the .01 or .05 significance level, excluding the third and tenth regressions, where the Gini coefficient did not have a significant t-statistic. Inflation had a negative coefficient in each regression, with significant t-statistics at the .05 significance level in the first four regressions and in the eighth regression.The latter seven regressions included the factor score for social capital created by Bjornskov, and in each regression the factor score coefficient was positive with a significant t-statistic at the .01 or .05 significance level. Of particular relevance for this paper, is the seventh regression which adds in the independent variable unemployment, and the ninth regression which adds in the education variable (Bjornskov, 2003, pg 9).

Regression 7. The sixth regression includes the independent variables GNI per capita, Gini coefficient, inflation, factor score, and unemployment. This regression included 27 observations and had a pseudo R^2 value of 0.777.

GNI per capita. The positive coefficient for GNI per capita calculated by this regression is expected and shows that for every one international dollar increase in the GNI per capita, the average response to the life satisfaction question can be expected to increase by 0.481 points. This value reflects that as income increases, satisfaction with one's life also increases.

Gini coefficient. The positive sign of the coefficient for this variable is unexpected, showing that as the Gini coefficient increases by one index point, the average answer to the life satisfaction question increases by 0.231 points. While a negative sign is expected because it is presumed higher inequality would lead to less satisfaction with life, Bjornskov reasons that this coefficient may be positive because higher inequality also allows for more social mobility, and "a

higher Gini coefficient is taken as an indication of greater possibilities or higher returns to moving up in society" (Bjornskov, 2003, pg 8).

Inflation. The negative coefficient for inflation is expected and shows that as inflation increases by one percentage point, life satisfaction decreases by .213 points. This relationship demonstrates that as inflation increases, representative of growing uncertainty and potential lost income, satisfaction with one's life decreases.

Factor score. The coefficient for the factor score variable is positive as expected, demonstrating that as a country's factor score increases by one point, average life satisfaction increases by 0.395 points. This shows that as a country has more social capital, its citizens are more satisfied with their lives.

Unemployment. The coefficient for unemployment is unexpectedly positive, showing that as a country's unemployment rate increases by one percentage point, life satisfaction increases by 0.020 points. However, this coefficient has a t-statistic that is not significant and Bjornskov theorizes that this surprising coefficient may be a result of loss of social capital where "some of the variation in social capital could be an effect of aggregate unemployment" (Bjornskov, 2003, pg 8).

Regression 9. The ninth regression includes the independent variables GNI per capita, Gini coefficient, inflation, factor score, and education. This regression has 21 observations and a pseudo R^2 value of 0.833. In this regression, the coefficients are 0.549 for GNI per capita, 0.233 for Gini coefficient, -0.154 for inflation, and 0.513 for factor score. Each of these coefficients have the same sign as in all the regressions, and the values vary only slightly.

Education. The coefficient for this variable is negative, which is an unexpected result. This value means that for every one percentage point increase in the percent of the population

with both a primary and secondary education average life satisfaction decreases by 0.269 points. This unexpected sign could be the result of a problem similar to the one with unemployment's unexpected sign, where there may be a relationship between social capital and education that is preventing the positive effect education has on an individual's life satisfaction from presenting at the national level (Bjornskov, 2003, pg 8). Table 2 presents the results of the two regressions discussed.

Table 2: Bjornskov's Regression Results

Variables	GNIpc	Gini	Inflation	Factor score	Unemployment	Education
Regression 7 Coefficients	0.481	0.231	-0.213	0.395	0.020	
t-statistics	(2.649)*	(1.956)	(-1.842)	(2.297)*	(0.136)	
Pseudo $R^2 = 0.777$						
N = 27						
Regression 9 Coefficients	0.549	0.233	-0.154	0.513		-0.269
t-statistics	(3.113)**	(1.828)	(-1.554)	(3.407)**		(-1.680)
Pseudo $R^2 = 0.833$						
N = 21						

Yenniel Mendoza, Roger Loyola, Alonso Aguilar, & Roberto Escalante

Valuation of Air Quality in Chile: The Life Satisfaction Approach, 2019

This paper evaluates the effects of air pollution on life satisfaction in Chile, particularly in its more urban areas, using data for 70 municipalities which include the country's largest cities. The authors believe the findings in their paper could provide valuable insight into

continuing or creating public policy aimed at pollution control in order to increase the social benefits from having a healthier environment (Mendoza et al., 2019, pg 368). The authors present results for two OLS regressions each intended to analyze the value of air quality in Chile, where air pollution is measured based on particulate matter at two different levels in each regression. The dependent variable in this study, life satisfaction, is calculated based on survey responses from Chilean households where responses range from 1 to 10, 1 being the worst satisfaction and 10 being the best (Mendoza et al., 2019, pg 374). The results of this study demonstrate how the health of the environment has an impact on life satisfaction.

Regression 1. In the first regression the independent variables are LogIncome and the annual average of particulate matter with a diameter of 10 microns or less (PM_{10}), measured as micrograms per cubic meter, in the zone of the respondent. This regression included 37,315 observations and has an adjusted R^2 of 0.0549. Table 3 displays the results of this regression.

LogIncome. The coefficient for this independent variable is positive as expected and shows that as the per capita income increases by one dollar, life satisfaction increases by

Table 3: Mendoza et al.'s Regression Results

Variables	LogIncome	PM_{10}
Regression 1 Coefficients	1.2981***	-0.0033***
Robust Standard Errors	(0.0279)	(0.0007)

Adjusted $R^2 = 0.0549$

N = 37,315

*** = p<0.001, ** = p<0.01

0.012981 index points. This demonstrates that higher income leads to higher life satisfaction.

PM_{10}. The coefficient for this variable has the expected negative sign and its value of -0.0033 means that as the annual average of particulate matter in the air with a diameter of 10 microns or less in the zone of the respondent increases by one microgram per cubic meter, life satisfaction decreases by 0.0033 points (Mendoza et al., 2019, pg 3760.

Richard Florida, Charlotta Mellander, & Peter J. Rentfrow

The Happiness of Cities, 2013

This study measures subjective well-being across several United States metropolitan areas to examine the relationship between independent variables like income or human capital and the dependent variable well-being. The authors' primary goal is to unveil the role of human capital in geographical differences of happiness, while controlling for income and other economic factors that have previously already been found to have a relationship with happiness (Florida et al., 2013, pg 616). The dependent variable, the well-being index, is measured on a 0 to 100 scale, with 0 being the worst and 100 being the best level of well-being. The index score is determined based on six sub-indices that measure life evaluation, emotional health, work environment, physical health, healthy behavior, and access to basic needs. Information for these scores was gathered via telephone surveys of 353,000 individuals that were tracked throughout the year of 2009 (Florida et al., 2013, pg 617). This study presents the results of several regressions with various independent variables.

Regression 2.2. Of particular interest for the purpose of this paper is regression 2.2 which includes the independent variables average wage level, unemployment, housing-to-wage ratio, human capital, population density, and age (Florida et al., 2013, pg 622). The results of this

regression are shown in table 4. This regression includes 169 observations and has a calculated R^2 value of 0.578 and an adjusted R^2 of 0.560.

Average Wage Level. Measured as the average total money earnings received for work performed in the region, the independent variable average wage level has a coefficient of 0.000 in this regression meaning that as the average wage level increases there is no effect on well-being; however, this coefficient is not significant.

Table 4: Florida et al.'s Regression Results

Variables	Regression 2.2 Coefficients	Robust Standard Errors
Average Wage Level	0.000	(0.033)
Unemployment	-0.073	(0.049)
Housing-to-Wage Ratio	0.394***	(0.075)
Human Capital	27.727***	(4.546)
Population Density	0.001	(0.001)
Age	-0.119***	(0.035)
Adjusted R^2 = 0.560		
N = 169		

Unemployment. This variable is the share of the labor force that is without employment. The coefficient for this variable is negative as expected showing that as unemployment rates rise, happiness or well-being decreases. This value means that as the unemployment rate increases by one percentage point, the well-being index score decreases by 0.073 index points.

Housing-to-Wage Ratio. This variable is the ratio of housing costs, measured as the median housing value, to wages and the coefficient for this variable in this regression has the expected positive sign showing that as the ratio increases, meaning housing affordability increases, well-being increases. This coefficient shows that as the housing to wage ratio increases by one unit, the well-being score increases by 0.394 index points.

Human Capital. This variable, the percent of the labor force with a bachelor's degree or higher, has an expected positive coefficient showing that as a city has more human capital, their well-being score increases. This value means that as the percent of the labor force with a bachelor's degree or above increases by one percentage point, the well-being score increases by 27.727 points.

Population Density. Population density is measured as the number of people per square kilometer in a city. This coefficient is positive, meaning that as the population density increases, well-being also increases. The calculated value of the coefficient shows that as the number of people per square kilometer increases by one person, the well-being score increases by 0.001 points. However, this coefficient is not statistically significant.

Age. This variable is the median age of the population based on data from the Census from 2006-2008. The coefficient for this variable is negative meaning that as the median age of the population increases, well-being of that population decreases. The value of the coefficient shows that as the median age increases by one year, well-being decreases by 0.119 index points (Florida et al., 2013, pgs 617, 622).

Model

Country happiness is the national average response to the Cantril ladder question with a scale of 0 to 10, 10 being the happiest, best possible life. This was chosen as the dependent

variable because an “umbrella measure [is] better than simply one component of well-being,” and this value has been found to be better comparable across countries (Smith and Helliwell, 2022). The regression analysis performed in this paper is used to calculate the values for the constant and coefficients in this model. The formulation of this model is presented in Table 5.

Table 5: The Country Happiness Model

CH = a + bEFI + cCorr + dGini + eMR + fSM + gU + hFLP + iUP + jPSF + kDen + lPEld + mPYou + nPMid + oMI + pLEB + qTFR + r%Hgov + sAgr + tAcc + uGGpc + vMYS + wRef + xHIV + yFA + zAvGr + Ɛ

CH = Country Happiness
EFI = Economic Freedom Index
Corr = Corruption Perceptions Index
Gini = Gini Coefficient
MR = Mortality from CVD, Cancer, Diabetes, or CRD
SM = Suicide mortality Rate
U = Unemployment Rate
FLP = Female Labor Force Participation rate
UP = Urban Population
PSF = Prevalence of Severe Food Insecurity
Den = Population Density
PEld = Population ages 65 and Above
PYou = Population ages 0-14
PMid = Population ages 15-64
MI = Infant Mortality Rate
LEB = Life Expectancy at Birth
TFR = Total Fertility Rate
%Hgov = Domestic General Government Health Expenditure
Agr = Agriculture, Forestry, and Fishing, Value Added
Acc = Access to Electricity
GGpc = Greenhouse Gas Emissions Per Capita
MYS = Mean Years of Schooling
Ref = Refugee Population by Country of Asylum
HIV = Prevalence of HIV
FA = Forest Area
AvGr = Average Annual GDP Growth Rate
a = constant
b, c, d,...z = coefficient for respective variable
Ɛ = error term

Constant. The constant ‘a’ represents country happiness when all other variables are zero.

Economic Freedom Index. The Economic Freedom Index scores countries using a scale of 0-100, 100 being the most economically free. This score is based on the four broad factors of rule of law, government size, regulatory efficiency, and open markets (James et al., 2022). The coefficient for the Economic Freedom Index (EFI) is expected to have a positive sign because as the economic freedom of a country increases, individuals of the country will be happier and more satisfied with their rights and freedom to make economic choices.

Corruption Perceptions Index. The Corruption Perceptions Index scores and ranks countries based on their perceived levels of government corruption on a 0-100 scale with 100 being the most free from corruption. The coefficient for the Corruption Perceptions Index (Corr) is expected to be positive because as the index increases, meaning there is less perceived corruption, happiness increases as people feel more free and trusting of their government.

Forest Area. Forest area is the percentage of a country's land that is forested. Forest area (FA) is expected to have a positive sign, meaning country happiness increases as the percentage of forest area increases. This is because it is expected that when there is more forested land, people are happier because time in nature has been found in various studies to make people both happier and healthier due to our deep-rooted connection with nature (Rusell et al., 2013, pg 494).

Gini Coefficient. The Gini Coefficient comes from the Gini Index which scores and ranks countries on their level of inequality using a 1-100 scale with 100 meaning the highest level of inequality, such as one person in the country having all the money. The coefficient for this variable (Gini) is expected to have a negative sign meaning that as inequality increases, happiness decreases. This is because when inequality of a country rises, tension and stress in that country increases as people experience the negative effects of inequality, and happiness decreases.

Prevalence of HIV. The Prevalence of HIV is the percent of people ages 15 to 49 that are infected with HIV. The coefficient for this variable (HIV) is expected to have a negative sign showing that as instances of HIV infection increase, happiness decreases. This is because as more people in a country are experiencing illness and even death in some cases, there is more stress, pain, and financial burden that decreases the happiness of people.

Mortality Rate from CVD, Cancer, Diabetes, or CRD. This variable is the percent of people aged 30 who would die before their 70th birthday from cardiovascular diseases, cancer, diabetes, or chronic respiratory diseases. The coefficient for this variable (MR) is expected to have a negative sign demonstrating that as mortality from these diseases increases, happiness decreases. This is expected because as a country's people experience more health problems, death, and loss of loved ones, they are less happy.

Suicide Mortality Rate. This variable is the number of suicides per 100,000 people in a country. The coefficient for the suicide mortality rate (SM) is expected to be negative reflecting that as suicide mortality rates increase, country happiness decreases. This sign is expected because when a country has higher suicide rates, its people are experiencing worse mental health and loss of loved ones from suicide, decreasing happiness.

Unemployment Rate. The unemployment rate is the percent of the total labor force that is actively seeking employment but unable to find it. The coefficient for the unemployment rate (U) is expected to be negative showing that as unemployment rates increase, happiness decreases. This is because when there is more stress and uncertainty about the job market and one's income, happiness is negatively impacted.

Female Labor Force Participation Rate. The female labor force participation rate is the percent of the total labor force that is composed of women. The coefficient for this variable (FLP) is

expected to be positive meaning that as the female labor force participation rate in a country increases, the happiness of that country increases as well. This expectation is because when the female labor force participation rate is higher, there is more gender equality and female independence in a country making women happier, meaning overall the country is happier.

Urban Population. The urban population is the percent of the total population living in urban areas, which are defined by national statistical offices. The coefficient for urban population (UP) is expected to be positive meaning that as there is a higher percentage of the population living in urban areas, there is higher levels of happiness. This sign for the coefficient is expected because higher urban populations reflect more development which is associated with higher levels of happiness.

Prevalence of Severe Food Insecurity. This variable is the percentage of the population living in households that are considered severely food insecure, meaning at least one adult in the household has reported experiences such as skipping meals or reducing their food intake due to lack of money or resources. The sign for this coefficient (PSF) is expected to be negative showing that as the prevalence of severe food insecurity increases, country happiness decreases. This is the expectation because as more people in a country are experiencing food insecurity they are dealing with health issues like deficiencies, stress, and hunger, all of which decreases happiness.

Population Density. The variable population density is the number of people per square kilometer of land in a country. The coefficient of this variable (Den) is expected to be positive meaning that as population density increases, happiness increases. This is expected because higher population density means there could also be higher access to public amenities like parks, recreational facilities, or other valuable buildings and facilities (Ye et al., 2014).

Population Ages 65 and Above. This variable is the percent of the total population that is aged 65 and older. The coefficient for this variable (PEld) is expected to be negative showing that as the percentage of the population that is over 65 increases, happiness decreases. This is the expected sign because when a higher proportion of the population is 65 and over, there are more older individuals that are dependent on the government or family members for support. The emotional and financial stress of supporting another individual, especially one that may be also experiencing health problems due to their age, decreases happiness.

Population Ages 0 to 14. This variable is the percent of the total population that is aged 14 and under. The coefficient for this variable (PYou) is expected to be negative showing that as there are more individuals aged 14 or younger in a country, country happiness decreases. This is because individuals that are 14 and younger are dependent on parents or other individuals for support which increases financial stress, decreasing happiness.

Infant Mortality Rate. The variable infant mortality rate is the number of infants that are dying before the reach the age of one, per 1,000 live births in a year. The coefficient for the infant mortality rate (IM) is expected to be negative showing that as the infant mortality rate of a country increases, country happiness decreases. This is expected because higher infant mortality rates reflect more stress due to having mothers and babies with poor health, and the pain of losing a child, both of which decrease happiness.

Life Expectancy at Birth. This variable is the number of years an infant is expected to live given current patterns of mortality at the time of its birth and the assumption those patterns remain the same. The coefficient for this variable (LEB) is expected to have a positive sign showing that as life expectancy at birth increases, country happiness also increases. This is the expected sign

because a higher life expectancy at birth reflects that the people in a country have good health for a longer time, increasing happiness.

Total Fertility Rate. The variable total fertility rate is the number of children a woman would have if she were to live to at least the end of her childbearing years. The coefficient for total fertility rate (TFR) is expected to be negative meaning that as the total fertility rate increases, country happiness decreases. This is because higher fertility rates are common in countries where there are families who have more children because they need them to work, who also have less access to birth control and family planning, all of which are signs of underdevelopment meaning they likely experience less happiness. Higher fertility rates also reflect more stress on the mother's health from going through more pregnancies and deliveries that would decrease happiness as well.

Domestic General Government Health Expenditure. This variable is the percent of the general government expenditures from domestic sources that is designated towards health. The coefficient of this variable (%Hgov) is expected to be positive meaning that as domestic government health expenditures increase, happiness increases. This is expected to be positive because higher government health expenditures reflect higher prioritization of individuals' health, meaning they likely have access to better healthcare because of this prioritization, which increases happiness.

Agriculture, Forestry, and Fishing Value Added. This variable is the value added of the net output that agriculture, forestry, and fishing produce, as a percentage of GDP. The coefficient for this variable (Agr) is expected to be negative meaning that as the value added of these industries increases, happiness decreases. This is because a larger net output and value added from these industries reflects higher levels of environmental degradation, such as more deforestation or

agricultural runoff, which decreases happiness because there is less connection to nature and possible negative health effects.

Access to Electricity. This variable is the percent of the total population of a country that has access to electricity. The coefficient of access to electricity (Acc) is expected to have a positive sign showing that as a larger percentage of the population has access to electricity, the happiness of the country increases. This is because as more of the population gains access to electricity, more people are able to reap the benefits such as access to electronic goods and appliances and light at night that allows for more productivity and safety, all of which improve health, living standards, increases opportunity, reduces poverty and thus increases happiness (Huges, 2018).

Greenhouse Gases Per Capita. This variable is the amount of greenhouse gas emissions per person, measured in kilotons of CO2 equivalent. The coefficient for this variable (GGpc) is expected to be negative reflecting that as greenhouse gas emissions per capita increases, country happiness decreases. This is the expected sign because higher greenhouse gas emissions reflect worse air quality and environmental damage in a country, which harms public health and thus the happiness of individuals in a country (Haines et al., 2009).

Mean Years of Schooling. The variable mean years of schooling is the average number of years of schooling that has been received by individuals ages 25 or older. The coefficient for this variable (MYS) is expected to be positive showing that as the mean years of schooling in a country increases, country happiness increases. This is expected because individuals with more education can get better jobs, make more money, and support themselves, all of which lead to increased happiness.

Refugee Population by Country of Asylum. This variable is the total number of refugees within a country that is granting them asylum. The coefficient of this variable (Ref) is expected to be

negative meaning that as the total number of refugees in a country of asylum increases, the country's happiness decreases. This is expected because taking in refugees can be burdensome and often falls on neighboring countries that may be facing problems of their own, but take in refugees anyways due to their close proximity. This stress and tension decreases happiness. While not expected, a positive sign for the coefficient of this variable could be a demonstration of the generosity of the country of asylum and that a country is more well-off and can shoulder the burden of refugees. In this case, though a country is taking in refugees which can be burdensome, a country could also be happier because their ability to support refugees shows that the country has a strong financial situation and individuals that are kinder and more welcoming, meaning they are happier.

Average Annual GDP Growth Rate. This variable is the average of the annual percentage growth rate of the GDP per capita over the last five years, and is based on constant local currency. The coefficient for this variable (AvGr) is expected to be positive showing that as the average annual GDP growth rate increases, happiness increases. This is expected because an increasing average growth rate reflects that a country is experiencing more production and economic growth, and more people are likely receiving higher incomes and other benefits of a country experiencing economic growth, which increases happiness.

Hypotheses

Two hypothesis tests will be conducted to test the model in this paper. The hypothesis tests are composed of two statements, the null hypothesis and the alternative hypothesis. In each test, the null hypothesis is either rejected or not rejected, and if the null hypothesis is rejected then the alternative hypothesis can be affirmed. The 0.05 significance level is used for the hypothesis tests in this paper. Table 6 displays the formulas for these tests.

Table 6: The Hypotheses

$CH = a + bEFI + cCorr + dGini + eMR + fSM + gU + hFLP + iUP + jPSF + kDen + lPEld + mPYou + nPMid + oMI + pLEB + qTFR + r\%Hgov + sAgr + tAcc + uGGpc + vMYS + wRef + xHIV + yFA + zAvGr + \varepsilon$

F-test (variables):	H_o: The regression is not statistically significant
	H_a: The regression is statistically significant
t-test (constant):	$H_o: a = 0$, $H_a: a \neq 0$
t-test (coefficient):	H_o: coefficient = 0, H_a: coefficient $\neq 0$

The first hypothesis test is a one-tailed F-test which will tell whether or not the estimated regression is statistically significant. This test will help determine the likelihood that there is a relationship between the independent variables and country happiness in the respective estimated regression. In this test, the null hypothesis statement is that the regression is not significant, and the alternative hypothesis is that the regression is significant. If the results of the test lead to a rejection of the null hypothesis, then the alternative hypothesis can be affirmed, meaning it can be concluded with 95% confidence that the regression is statistically significant and there is a relationship between country happiness and the independent variables. This is the expected result, and if the test leads to the inability to reject the null hypothesis then it cannot be concluded the regression is significant, and the model will need to be reevaluated and adjusted.

The second hypothesis is a two-tailed t-test used to test the significance of the constant and the coefficients for each variable calculated by the regression. The null hypothesis for this test is that the constant or coefficients will equal zero, and the alternative hypothesis is that they will not equal zero. If the null hypothesis is rejected, then the alternative hypothesis can be

affirmed meaning that the constant or coefficient being tested is not equal to zero and is statistically significant. If the alternative hypothesis is affirmed for the coefficient being tested, this also means there is a significant relationship between the independent variable and the dependent variable. If the null hypothesis cannot be rejected, then the constant or coefficient is not statistically significant and the possibility that the dependent variable is not affected by the independent variable with the respective coefficient cannot be ruled out.

Methodology

In order to test the model and find the values needed to run the hypothesis tests, the Statistical Program for the Social Sciences (SPSS) is used to conduct a regression analysis of the model. This software will be used to estimate a linear regression that creates a line of best fit for the input independent variables and the dependent variable. This is considered a 'best fit' because the regression estimates the best fitting line that will minimize the sum of squared deviations between the line and the actual data points, this is called the least squares method. The points on this line of best fit are the predicted values of the variables.

The regression also calculates a value for R^2, which represents the closeness of fit–how close the fit of the line is to the actual data. This value is found by dividing the explained variation by the total variation in Y. The calculated value ranges between 0 and 1, representing a proportion which can also be interpreted as a percentage. For example, if the R^2 value is 0.95, this means 95% of the variation in the dependent variable can be explained by the regression equation. If the value is 1, then 100% of the variation in the dependent variable is explained by the equation and the regression line is a perfect fit, while if it is 0 none of the variation is explainable by the regression equation. This value is also called the coefficient of multiple determination. The R^2 value can be particularly useful in determining which regression is better

if multiple regressions are being estimated, which is the case for this paper. The adjusted R^2 value is the proportion of total variability in the dependent variable that is predicted by the regression equation once the number of independent variables included in the equation is also considered. This is an important value because additional independent variables will almost always increase the R^2 value, but if the adjusted R^2 value increases, then it can be concluded that the added variable is the actual source of this increase. The equation for each of the values described in this section are shown in Table 7.

Table 7: The Methodology

R^2	$= \frac{\text{explained variation}}{\text{total variation}}$	$= \frac{\Sigma(Y'-Y)^2}{\Sigma(Y-Y)^2}$
Adj. R^2	= R^2 adjusted for # of independent variables	$= 1 - \frac{[(1-R^2)(n-1)]}{n-k-1}$
F	$= \frac{\text{variance explained by regression}}{\text{residual variance}}$	$= \frac{\Sigma(Y'-Y)^2/K-1}{\Sigma(Y-Y)^2/N-K}$
t	$= \frac{\text{regression coefficient}}{\text{standard error of the regression coefficient}}$	$= \frac{b}{SE_b}$ or $\frac{a}{SE_a}$
SE_{est}	= standard deviation of residuals	$= \sqrt{(\frac{\Sigma(Y-Y\prime)^2}{N-K})}$
95% CI	= Predicted CH' +/- t crit*SE_{est}	

The F-test is used to determine if the estimated regression equation is statistically significant. The regression calculates a value for F which is found by taking the variance explained by the regression and dividing it by the residual variance. This value is then compared to the F critical value to test if it is statistically significant. The F critical value is found using an F table or critical value calculator, and in this case the 0.05 significance level is used. To find the F critical value, the numerator is the number of coefficients including the constant minus one

(K-1), and the denominator is the sample size minus the number of coefficients (N-K), also known as the degrees of freedom. If the calculated F value is greater than the F critical value, then the null hypothesis can be rejected and it can be concluded that the regression is statistically significant and there is a relationship between the dependent variable and the independent variables. If the calculated F value is not greater than the F critical value, then the null hypothesis cannot be rejected and the possibility that the regression is not statistically significant cannot be ruled out.

The regression also calculates the t statistics for the constant and coefficients used to conduct the t-tests. The t statistic is found by dividing the regression coefficient by the standard error of the regression coefficient. Once this value is found it is then compared to the t critical value to conduct the hypothesis test. This critical value is found using either a t table or critical value calculator using the degrees of freedom (N-K). The significance level of 0.05 is also used for this. This t-test is two-tailed which means that the calculated t value can be either positive or negative, so the absolute value of the t statistic is used to compare to the t critical value. If the absolute value of the t statistic is greater than the t critical value, then the null hypothesis can be rejected and the alternative hypothesis affirmed, meaning that the constant or coefficient being tested is statistically significant. In the case of a coefficient, affirming the alternative hypothesis means that there is a relationship between the corresponding independent variable and the dependent variable.

The purpose of estimating a regression is to use the equation to predict the values for the dependent variable. However, these predictions can still be subject to error so the standard error of estimate (SEest) is used to represent the likely error in the regression. This paper will make predictions for the dependent variable and find a 95% Confidence Interval which takes into

account the likely error. This confidence interval determines the range of the values within which the true value of country happiness is likely to fall. This is found by taking the predicted value of the dependent variable, such as country happiness for one country in one year, and then both adding and subtracting the t critical value after it has been multiplied by the standard error of the estimate. This calculation provides the lower and upper values for the range where the true value of country happiness is likely to be in.

Data

Table 8 displays the descriptive statistics for each variable in the model. Data for the

Table 8: Descriptive Statistics, for the years 2014-2018

Variable	Minimum	Maximum	Mean	Standard Deviation
Country Happiness	2.839	7.769	5.489	1.126
Economic Freedom Index	25.2	89.4	62.226	10.004
Corruption Perceptions Index	11	92	44.78	19.811
Gini Coefficient	24	63	35.869	7.483
Mortality from CVD, Cancer, Diabetes, or CRD	7.5%	36.5%	18.603%	6.713
Suicide Mortality Rate	0.6	35.6	9.612	6.318
Unemployment Rate	0.13%	27.69%	7.157%	5.108
Female Labor Force Participation Rate	7.895%	56.006%	41.505%	9.348
Urban Population	16.132%	100%	61.997%	21.764
Prevalence of Severe Food Insecurity	0%	51.8%	7.130%	9.766
Population Density	1.892	7,952.999	207.461	722.895
Population ages 65 and Above	0.905%	27.576%	9.593%	6.520
Population ages 0-14	12.212%	50.235%	26.542%	10.633
Population ages 15-64	47.200%	85.167%	63.865%	6.277
Infant Mortality Rate	1.6	94.9	19.901	19.799
Life Expectancy at Birth	52.372	84.211	73.180	7.415
Total Fertility Rate	0.977	7.245	2.584	1.273
Domestic General Government Health Expenditure	0.734%	29.492%	10.983%	5.293
Agriculture, Forestry, and Fishing, Value Added	0.030%	60.611%	10.363%	10.624
Access to Electricity	7.700%	100%	85.379%	25.966
Greenhouse Gas Emissions Per Capita	0.001	10.568	0.431	1.175
Mean Years of Schooling	1.4	14.1	8.968	3.190
Refugee Population by Country of Asylum	5	3,681,688	150,423.03	433,009.607
Prevalence of HIV	0.1%	22.4%	1.461%	3.547
Forest Area	0.045%	91.597%	28.917%	20.332
Average Annual GDP Growth Rate	-10.636%	9.653%	2.206%	2.310

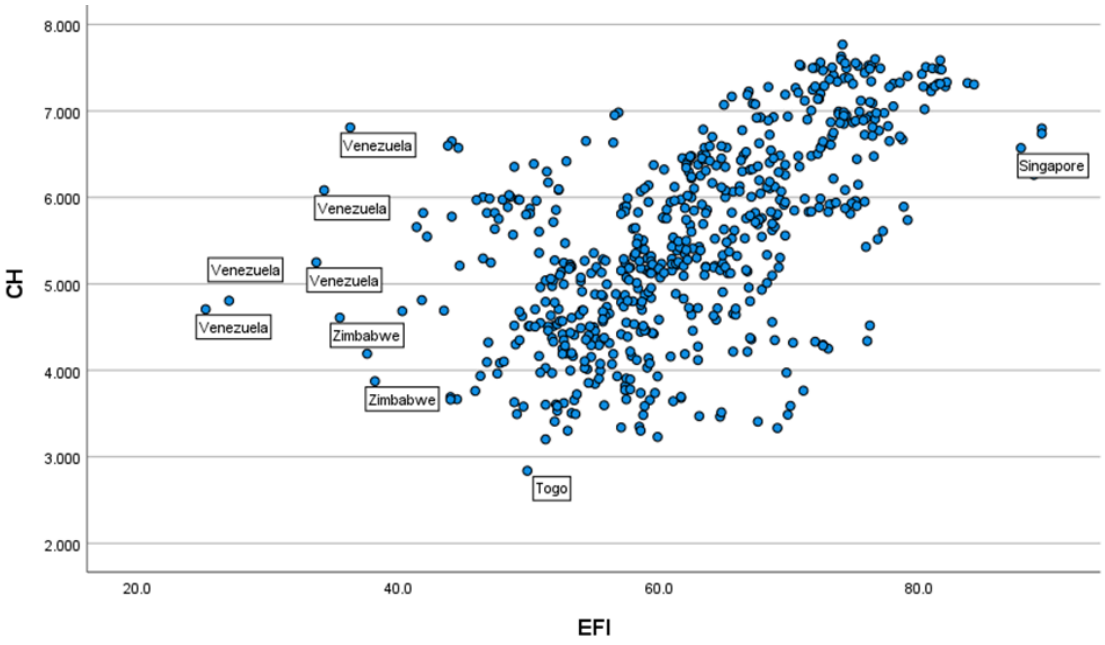
dependent variable and each of the independent variables come from various sources and are for the years 2014-2018.

Country Happiness. The data collected for the dependent variable, country happiness, comes from the World Happiness Reports published in each year from 2015 to 2019, with each publication containing data on country happiness collected the previous year. In collecting this data, respondents were asked to answer the Cantril ladder question which asks them to rate their life today on a 0-10 scale with 10 being the best possible life and 1 being the worst possible life. Each data point is the average of these responses of the citizens in one country for one year. Over all five years and across 128 countries, the average country happiness score is 5.489 index points. Finland had the highest country happiness in 2019, with a score of 7.769, meanwhile Togo had the lowest score in 2015 at 2.839 points.

Economic Freedom Index. The data for the economic freedom index is collected from The Heritage Foundation. Data for each country is a score based on a 0-100 scale with 100 being the most economically free. Data is used for 121 countries across the five years used for this paper (2014-2018), and the average score on the economic freedom index across all countries and years is 62.23 index points. The highest scoring country was Singapore in 2014 with an index score of 89.4 points, while the lowest score was Venezuela in 2018 with a score of 25.2. Looking at this graph, it can be seen that there is a general positive trend where as the EFI score increases, country happiness also increases. Noticeably, there are some countries not included in the cluster of points that are trending upwards and to the right. On the far left of the graph, the countries scoring less than 40 points on the economic freedom index are Venezuela and Zimbabwe. The clustering of points on the far right are all the data points for Singapore for each

year included in the dataset. Figure 1 shows the scatterplot for country happiness as a function of the economic freedom index.

Figure 1: Country Happiness as a Function of the Economic Freedom Index



Corruption Perceptions Index. Data for the Corruption Perceptions Index comes from Transparency International and is expressed as a score on a 0-100 scale with 100 being perceived as the most free from corruption. Data is for 127 countries across the five year time period, and has an average index score of 44.78 points. The highest score across all five years was Denmark's 2014 score of 92 index points, and the country with the lowest score was Afghanistan in 2015 with 11 index points. Figure 2 displays the scatterplot for country happiness as a function of the corruption perceptions index. Looking at this graph, a general trend upwards and to the right can be seen showing that as the Corruption Perceptions Index score increases, country happiness tends to increase.

Gini Coefficient. The scatterplot for country happiness as a function of the Gini coefficient is displayed in figure 3. Data for the Gini Coefficient comes from The World Bank

Figure 2: Country Happiness as a Function of The Corruption Perceptions Index

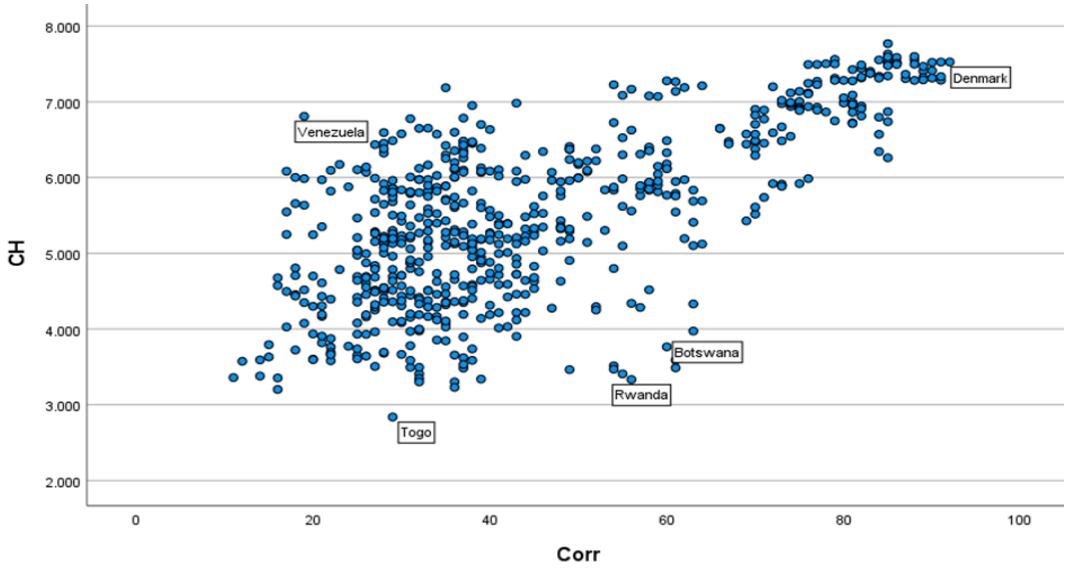
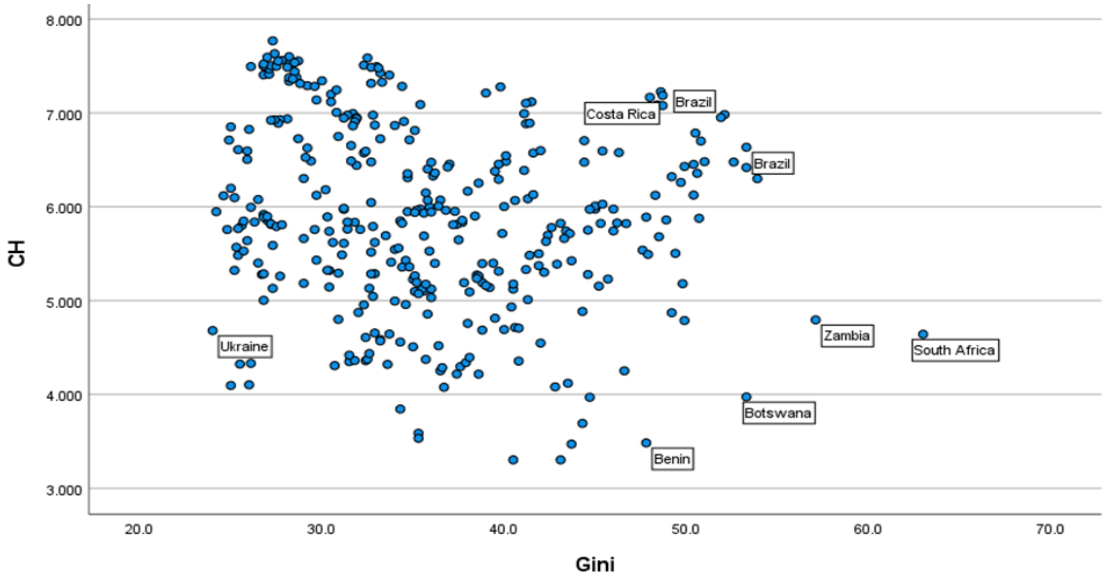


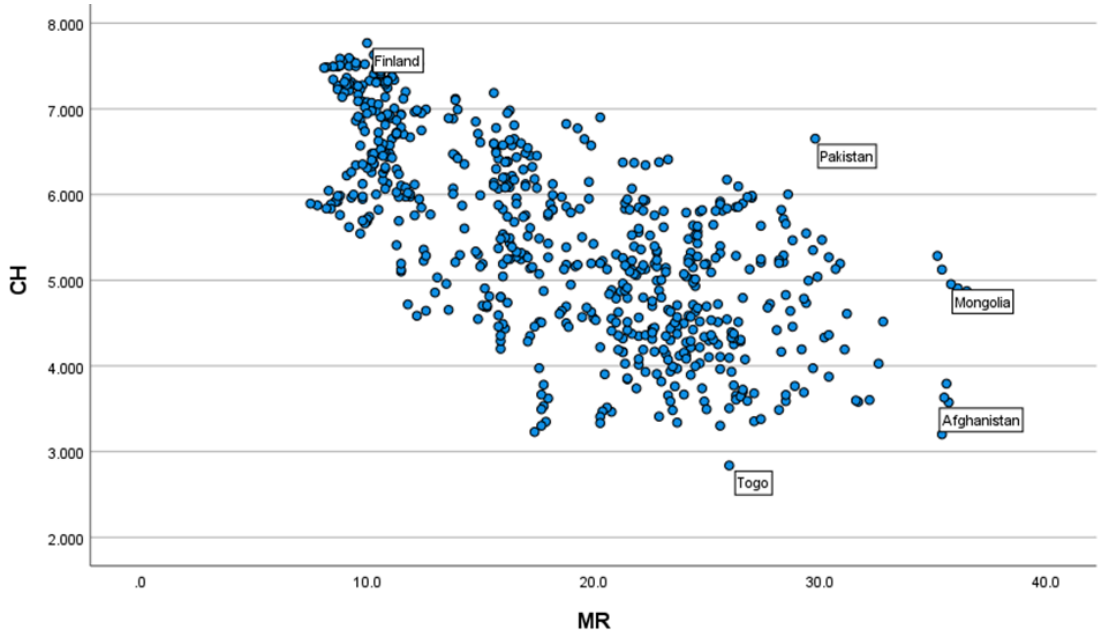
Figure 3: Country Happiness as a Function of the Gini Coefficient



and is measured on a 0-100 scale with 0 being perfect equality and 100 being perfect inequality. Data is for 108 countries, with an average score of 35.869. The country with the highest score, meaning the country ranked with the highest inequality, is South Africa with a score of 63 in 2014. The country ranked the most equal, with a score of 24 in 2014 is Ukraine. Looking at the scatterplot, there is a noticeable trend downwards and to the right where as the value of the Gini coefficient increases, country happiness decreases. The data points are more clustered toward the left side of the graph, and the outliers far to the right and low are labeled, and it can be seen these are all countries located in Africa.

Mortality Rate from CVD, Cancer, Diabetes, or CRD. Figure 4 displays the scatterplot for country happiness as a function of this mortality rate. Data for the mortality rate from

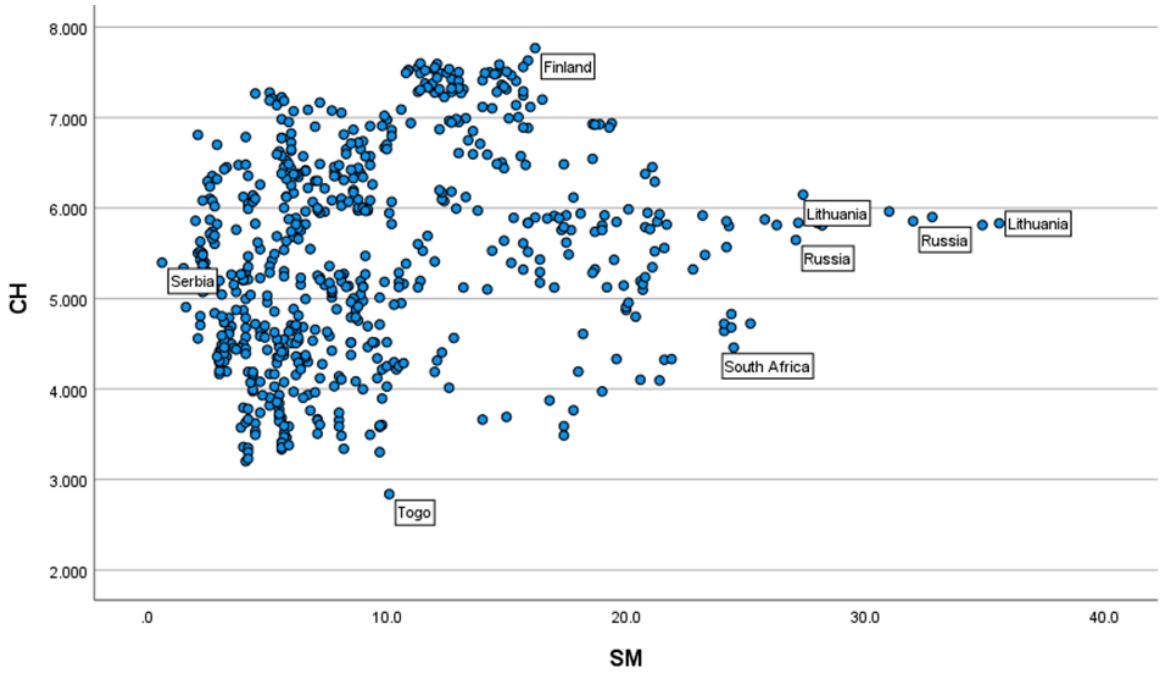
Figure 4: Country Happiness as a Function of the Mortality Rate from CVD, Cancer, Diabetes, or CRD



cardiovascular diseases, cancer, diabetes, or chronic respiratory disorders is collected from The World Bank, and this variable is measured as the percent of 30-year-olds who would die from one of these diseases before they reach 70. Data is used for 127 countries, which have an average mortality rate from these diseases of 18.603%. The scatterplot shows a trend downwards as the mortality rate increases. The country with the highest mortality rate from these diseases is Mongolia with a rate of 36.5% in 2014, while South Korea has the lowest value with a rate of 7.5% in 2018.

Suicide Mortality Rate. The suicide mortality rate data is gathered from The World Bank where the mortality rate is the number of people per 100,000 people that died from suicide in that year. Figure 5 shows the scatterplot for country happiness as a function of this variable. The

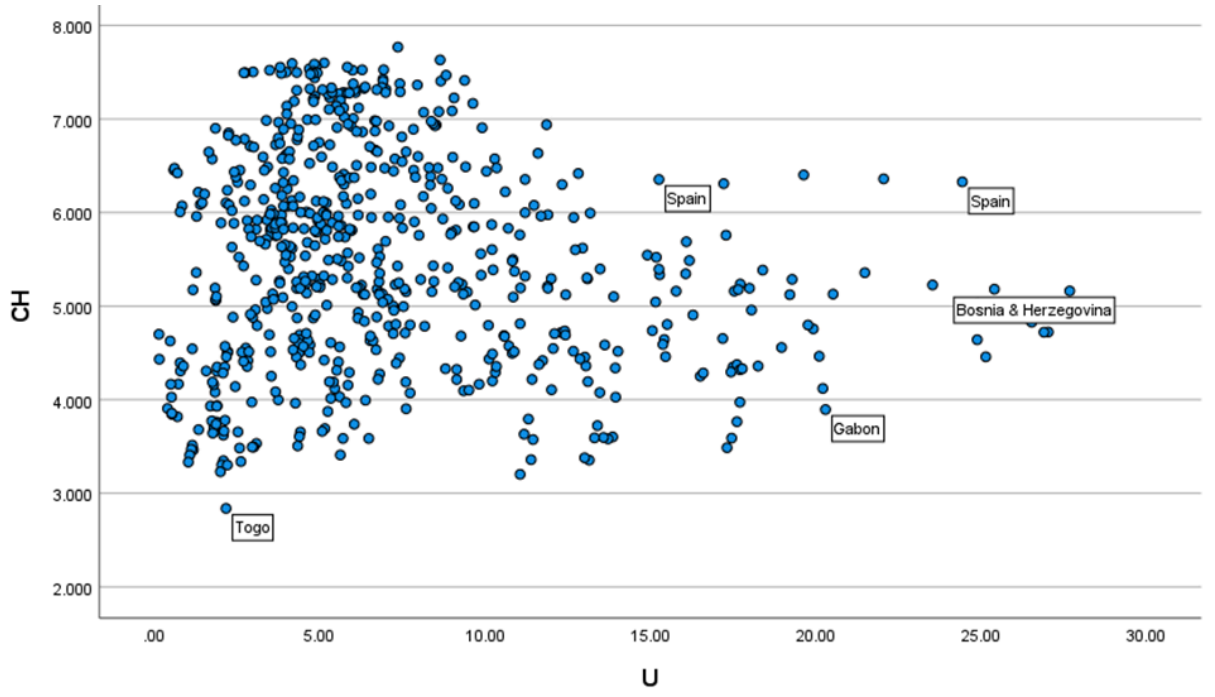
Figure 5: Country Happiness as a Function of The Suicide Mortality Rate



dataset for this variable includes 127 countries which have an average suicide mortality rate of 9.612 people per 100,000 people. The country with the highest suicide mortality rate is Lithuania which had a sucide mortality rate of 35.6 in 2014. Serbia had the lowest suicide mortality rate in 2017, with a rate of 0.6. The points on this scatterplot are mostly clustered towards the left side of the graph, below a mortality rate of 10. Moving towards the right, the points begin to thin out with the outliers forming a line around a country happiness score of about 6 as the suicide mortality rate increases.

Unemployment Rate. Data for the unemployment rate for 127 countries comes from The World Bank and is measured as the percent of the labor force composed of individuals available for and seeking work who are unable to find it. Figure 6 displays the scatterplot for country happiness as a function of the unemployment rate. The average unemployment rate of all the

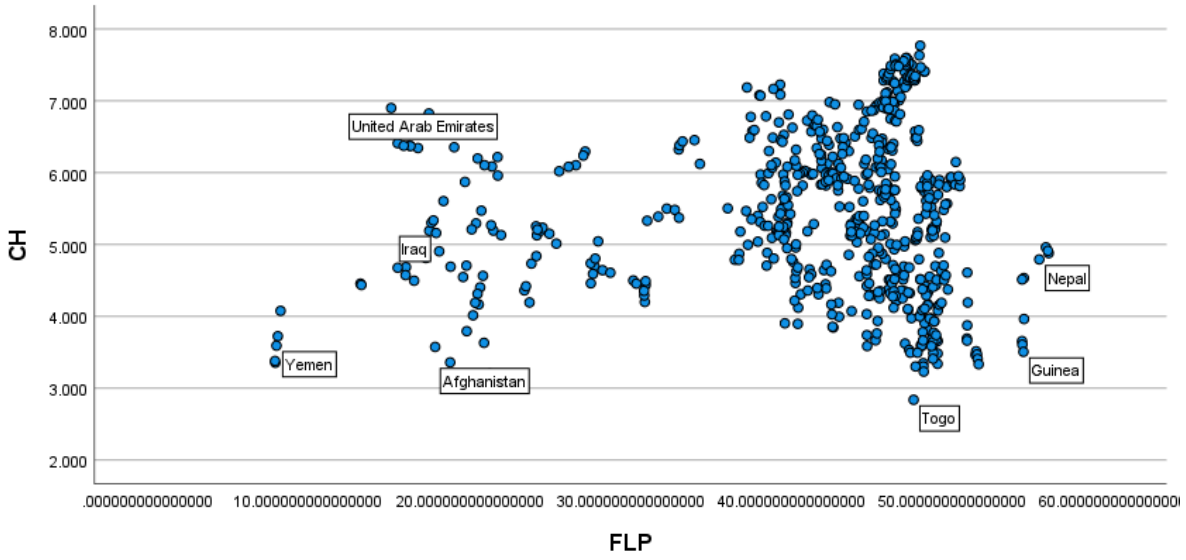
Figure 6: Country Happiness as a Function of the Unemployment Rate



countries is 7.157%. The country with the highest unemployment rate is Bosnia & Herzegovina which had an unemployment rate of 27.69% in 2015. The lowest unemployment rate is 0.13% which was Cambodia's unemployment rate in 2018. Looking at the scatterplot in figure 6, most of the points are clustered to the left below an unemployment rate of about 10%. The outliers with unemployment rates above about 20% include Gabon, Spain, and Bosnia & Herzegovina, and the country happiness of each of these countries ranges between a score of about 4 to 6.5.

Female Labor Force Participation Rate. The data for the female labor force participation rate comes from The World Bank and is for 127 countries. This variable is measured as the percentage of the total labor force that is composed of women. Across these countries, the average female labor force participation rate is 41.505%. Yemen had the lowest female labor force participation rate out of all in the countries when the rate was 7.895% in 2018, and Nepal had the highest rate, with 56.006% in 2017. Figure 7 shows the scatterplot for country happiness as a function of the female labor force participation rate. Most of the countries are clustered on

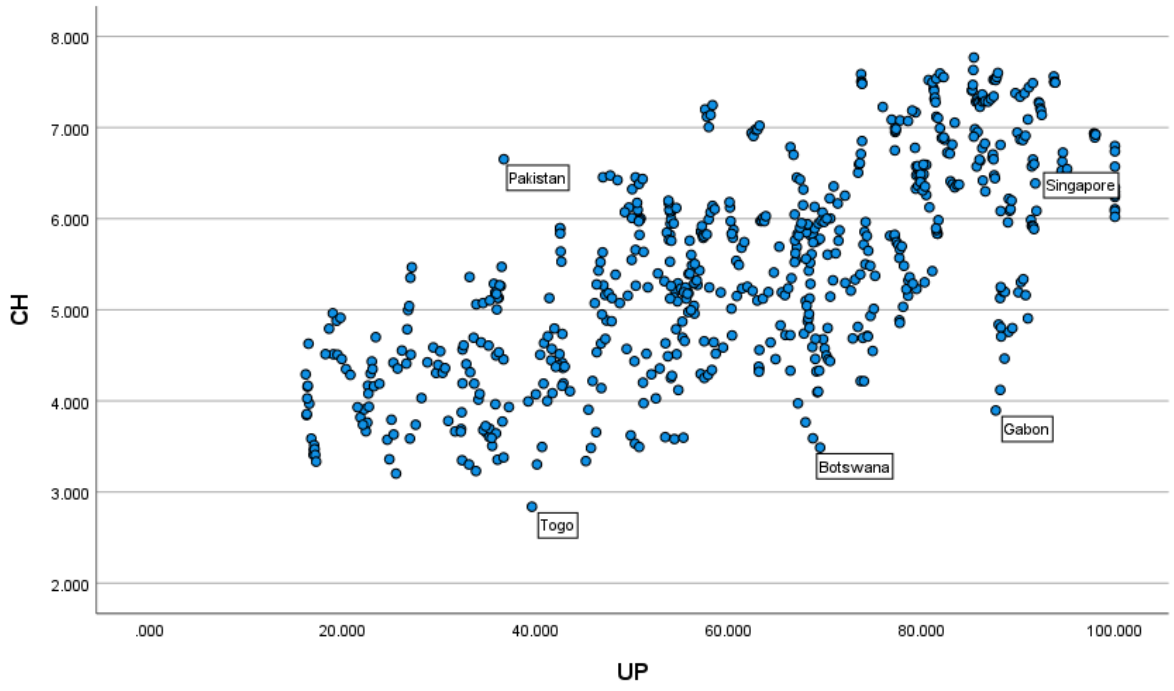
Figure 7: Country Happiness as a Function of the Female Labor Force Participation Rate



the right side, with a participation rate over about 40%, while a few countries have much lower participation rates which can be seen on the left half of the graph. There is a small cluster past about 50% and above about 7 on the country happiness index, where a small trend upwards can be seen, but otherwise there is no noticeable trend or direction of the data points.

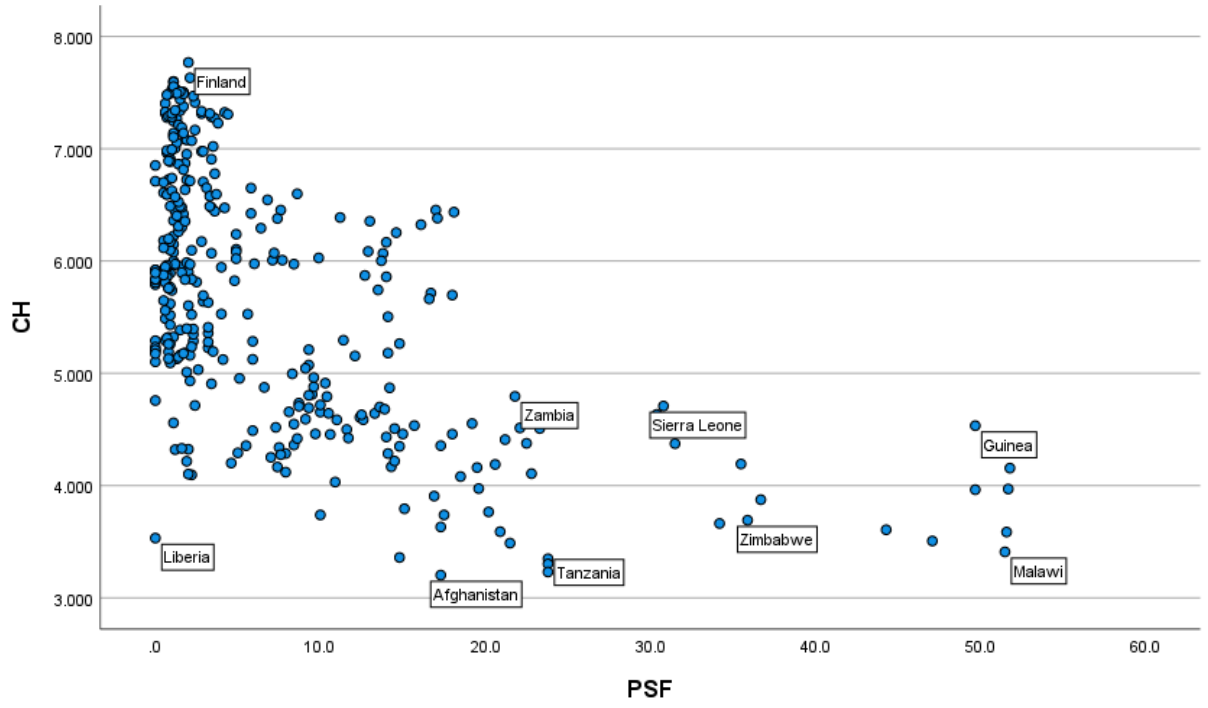
Urban Population. Data on the urban population comes from The World Bank and is measured as the percent of the total population that is living in urban areas. The mean value of this variable is 61.997%, with the lowest value being 16.132% which was Malawi's urban population in 2014, and the highest being 100% which was the urban population of both Kuwait and Singapore for 2014-2018. Figure 8 displays the scatterplot for country happiness as a function of the urban population, where a trend up can be seen as the urban population increases.

Figure 8: Country Happiness as a Function of the Urban Population



Prevalence of Severe Food Insecurity. Data for this variable is gathered from The World Bank for 93 countries and is the percent of the total population experiencing severe food insecurity. The average prevalence of severe food insecurity is 7.130%. The lowest percentage of severe food insecurity is 0%, which is the value for several countries, and the highest is 51.8% in Malawi in 2015. Figure 9 shows the scatterplot for country happiness as a function of this independent variable. There is a clear trend downwards in country happiness as severe food insecurity increases, with most of the countries clustered on the left at a prevalence below about 10% and mostly close to 0%.

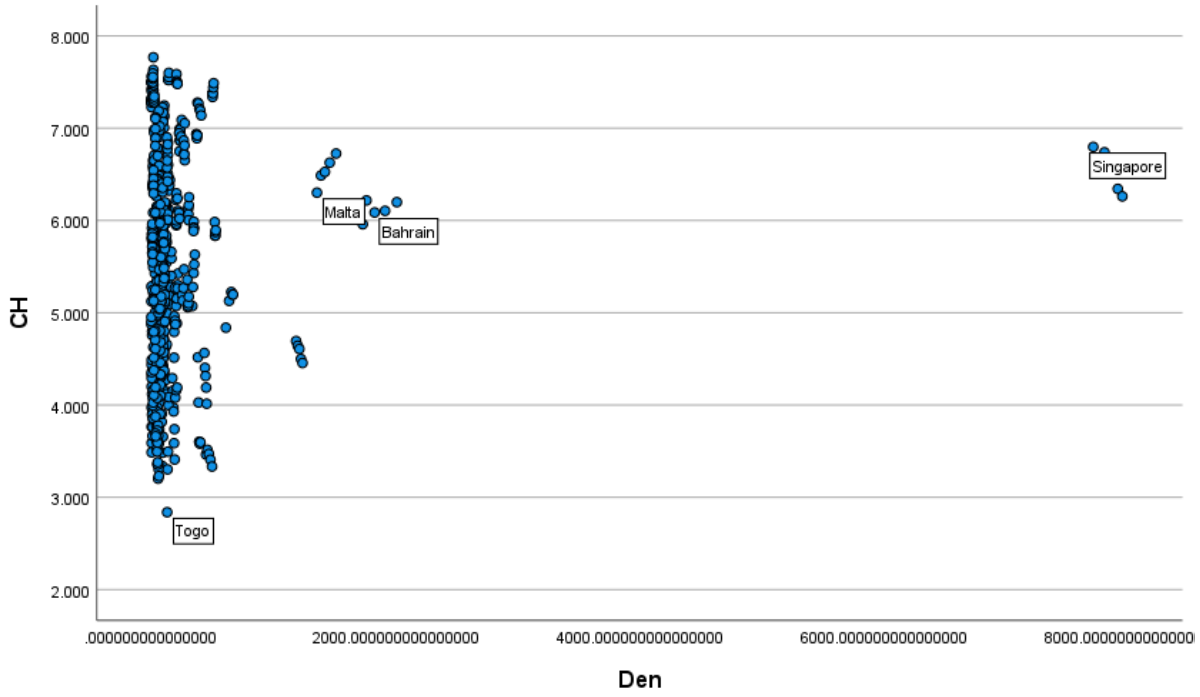
Figure 9: Country Happiness as a Function of the Prevalence of Severe Food Insecurity



Population Density. Data on population density is for 128 countries and comes from The World Bank. This is the average number of people per square kilometer in a country, and the average across all the countries is 207.461 people per square kilometer. Mongolia had the lowest

population density out of all the countries in 2014 with a value of 1.892 people per square kilometer. Meanwhile, Singapore had the highest population density, at 7,952.999 people per square kilometer. Figure 10 shows the scatterplot for country happiness as a function of population density. There is no noticeable trend in a certain direction, and a few outliers can be seen, such as Singapore which has a much higher population density than any other country.

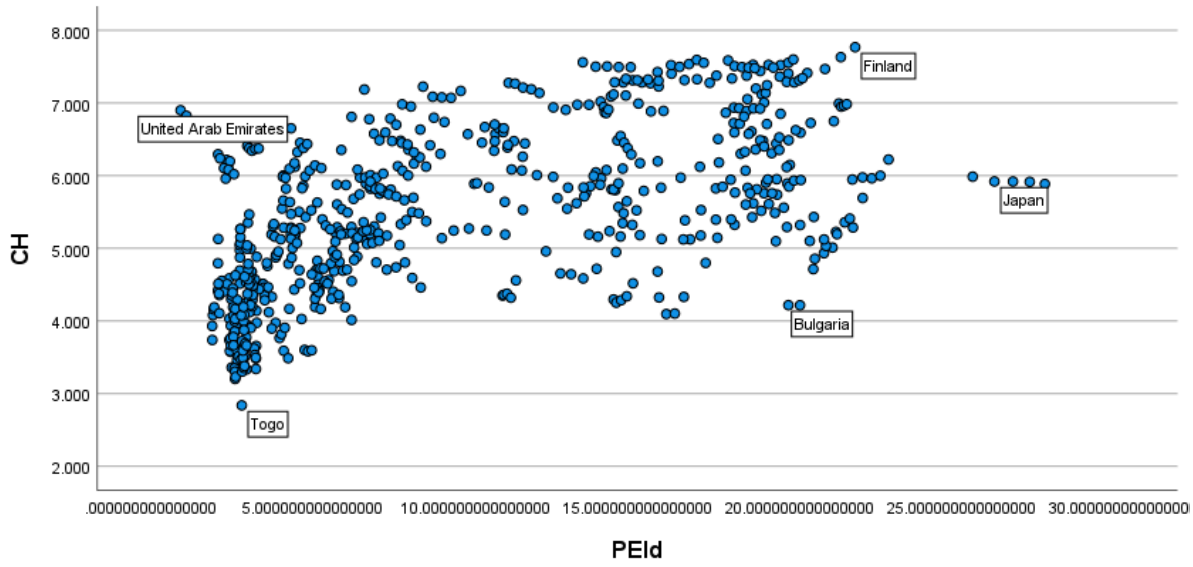
Figure 10: Country Happiness as a Function of Population Density



Population Aged 65 and Above. Data for 127 countries for this variable comes from The World Bank and is the percent of the country's total population that is aged 65 years and over. The average percentage is 9.593%. The United Arab Emirates had the smallest population in this age range with 0.905% in 2014, and Japan had the highest percentage at 27.576% in 2018. Figure 11 shows the scatterplot for country happiness as a function of this variable, where it can

be seen that most of the countries are clustered on the left side, below about 7%, and there is a wide spread in country happiness scores throughout the graph.

Figure 11: Country Happiness as a Function of The Population Ages 65 and Above



Population Ages 0-14. The data for the youth population comes from The World Bank and is for 127 countries. This variable is the percent of the total population at or under the age of 14, where the average percentage across all the countries is 26.542%. The country with the smallest youth population is Singapore in 2017, where 12.212% of the population was within this age range. With 50.235% of the population in this age range in 2014, Niger had the largest youth population. Figure 12 displays the scatterplot for country happiness as a function of this variable. There is a noticeable trend downwards as the youth population increases, with a tighter cluster of countries with a youth population under about 20% of the total population.

Population Ages 15-64. Figure 13 shows the scatterplot for country happiness as a function of this age group. Data for this age group of the population is from The World Bank and

Figure 12: Country Happiness as a Function of the Population Ages 0-14

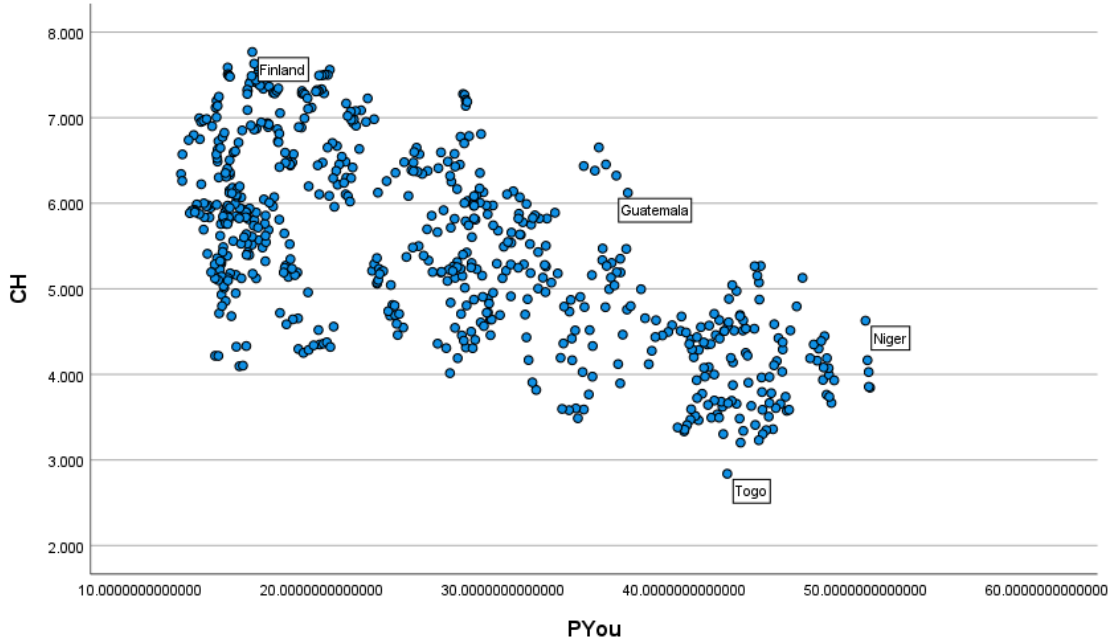
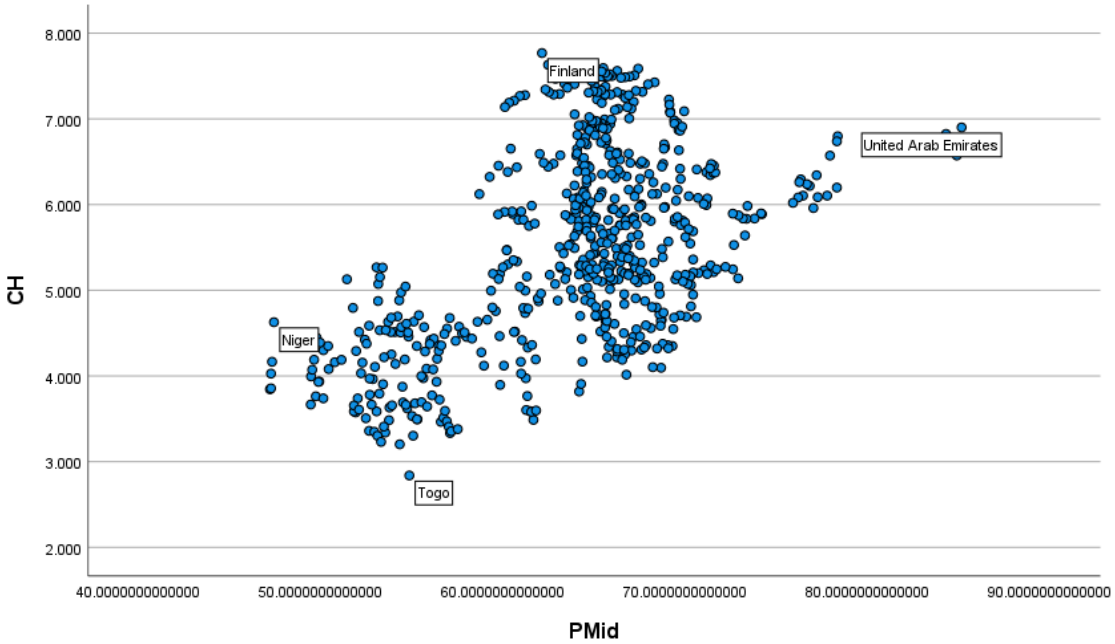


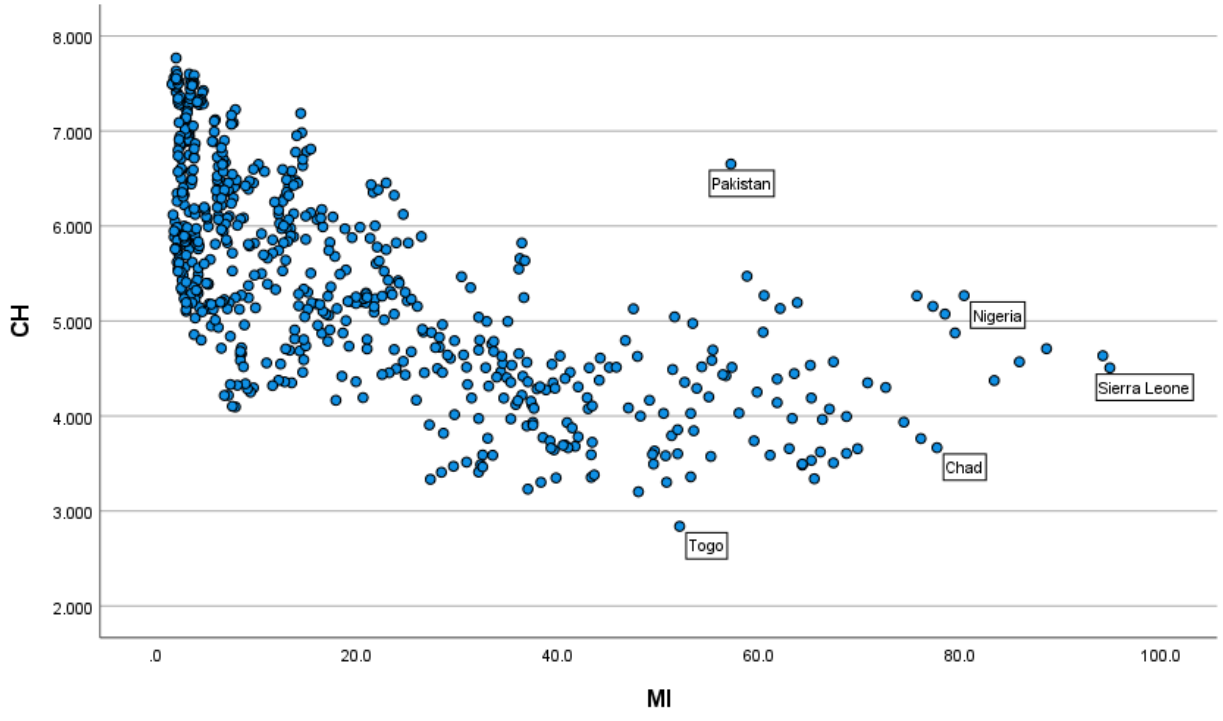
Figure 13: Country Happiness as a Function of the Population Ages 15-64



is for 127 countries, this value is the percent of the total population that is at or between the ages of 15 to 64. The average value is 63.865%, and the highest is United Arab Emirates' 2014 value of 85.167% while the lowest is Niger's 2014 value of 47.200%. Looking at the graph, most of the points are clustered towards the middle of the graph, just below about 70% of the population. Those countries to the left of this with lower percentages of the population in this age range also have lower country happiness, while those to the right with higher percentages of people in this age range have higher country happiness.

Infant Mortality Rate. Data on the infant mortality rate is gathered from The World Bank for 127 countries and is the number of infants who die before age 1 per 1,000 live births. Figure 14 shows the scatterplot for country happiness as a function of the infant mortality rate. The

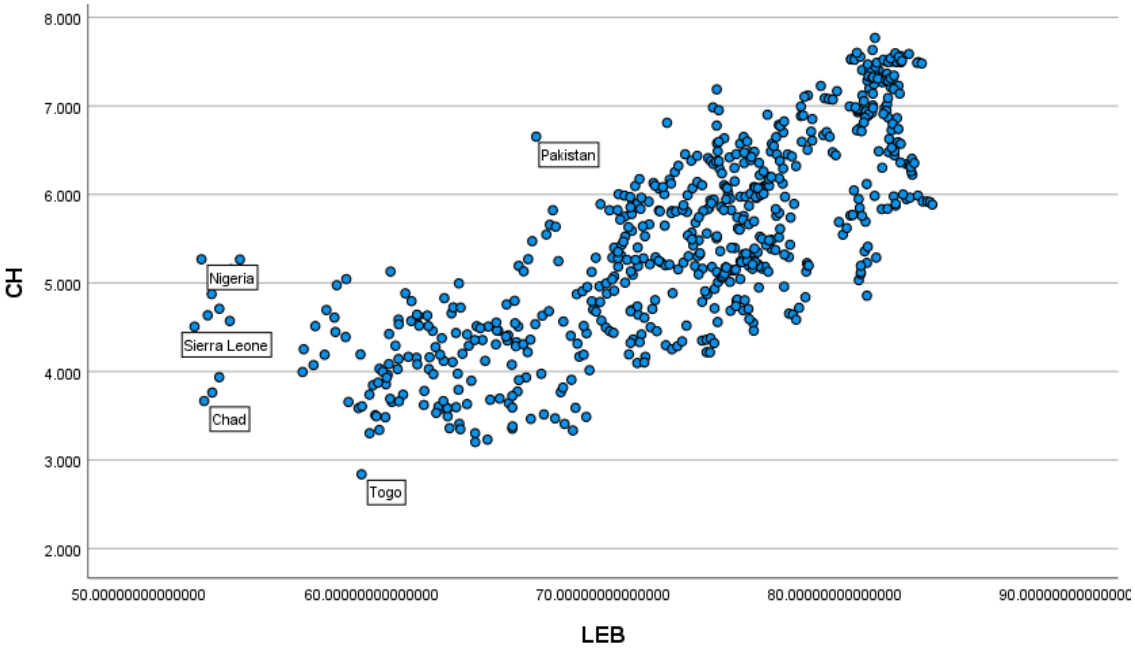
Figure 14: Country Happiness as a Function of the Infant Mortality Rate



average infant mortality rate for the dataset is 19.901 infants per 1,000 live births. Iceland had the lowest infant mortality rate of 1.6 in 2018, and Sierra Leone had the highest infant mortality rate of 94.9 in 2014. Looking at the scatterplot, there is a notable trend downward in country happiness as the infant mortality rate increases, with most points clustered to the left with a mortality rate below about 15.

Life Expectancy at Birth. Data for this variable comes from The World Bank for 128 countries and is the number of years a newborn is expected to live if the same patterns of mortality remain throughout their life. The average life expectancy at birth across all the countries is 73.180 years. The country with the lowest life expectancy is Sierra Leone, with a life expectancy at birth of 52.372 in 2014, while the country with the highest life expectancy is Japan with a life expectancy at birth of 84.211 in 2018. Figure 15 presents the scatterplot for country happiness as a function of the life expectancy at birth. The points trend up and to the right as the

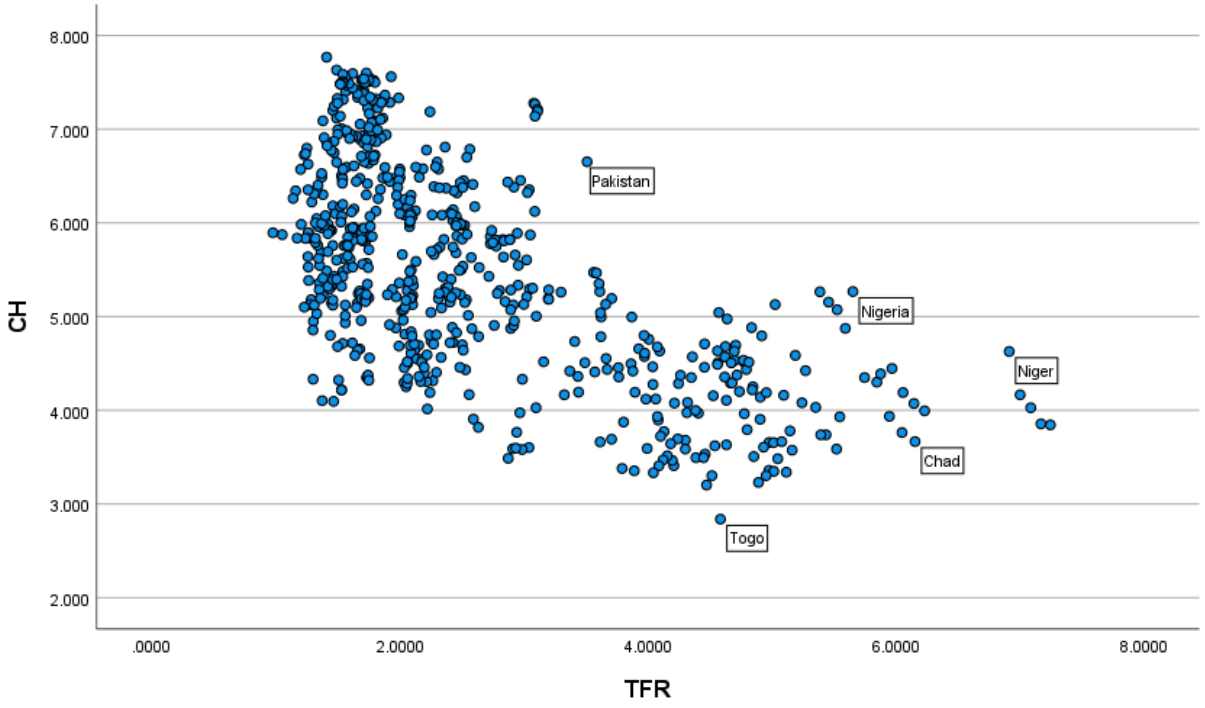
Figure 15: Country Happiness as a function of Life Expectancy at Birth



life expectancy at birth increases, with a few outliers far to the left.

Total Fertility Rate. Data for the total fertility rate for 128 countries comes from The World Bank and is measured as the average number of babies a woman will birth within her child bearing years. The mean fertility rate across all countries is 2.584 babies. The lowest fertility rate is 0.977 which was South Korea's in 2018, while the highest was 7.245 which was Niger's in 2014. Figure 16 shows the scatterplot for country happiness as a function of the total fertility rate. The points trend downwards as the total fertility rate increases, with most of the points clustered below a total fertility rate of about 3.

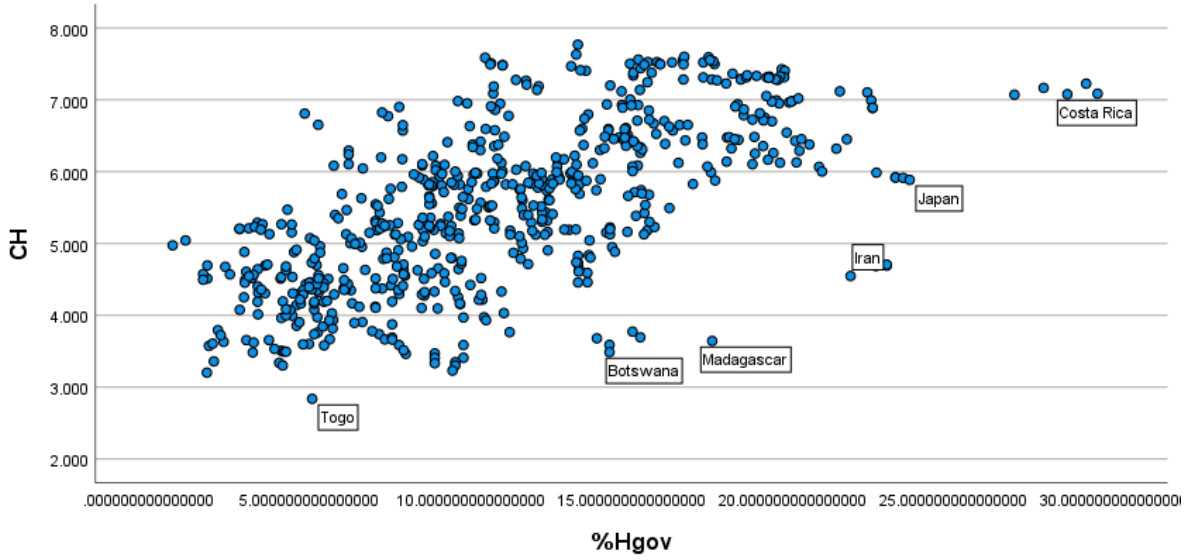
Figure 16: Country Happiness as a Function of the Total Fertility Rate



Domestic Government Health Expenditures. Data on the government expenditures on public health from domestic sources comes from The World Bank and is for 128 countries. This is measured as a percent of the total general government expenditures, and the average percent

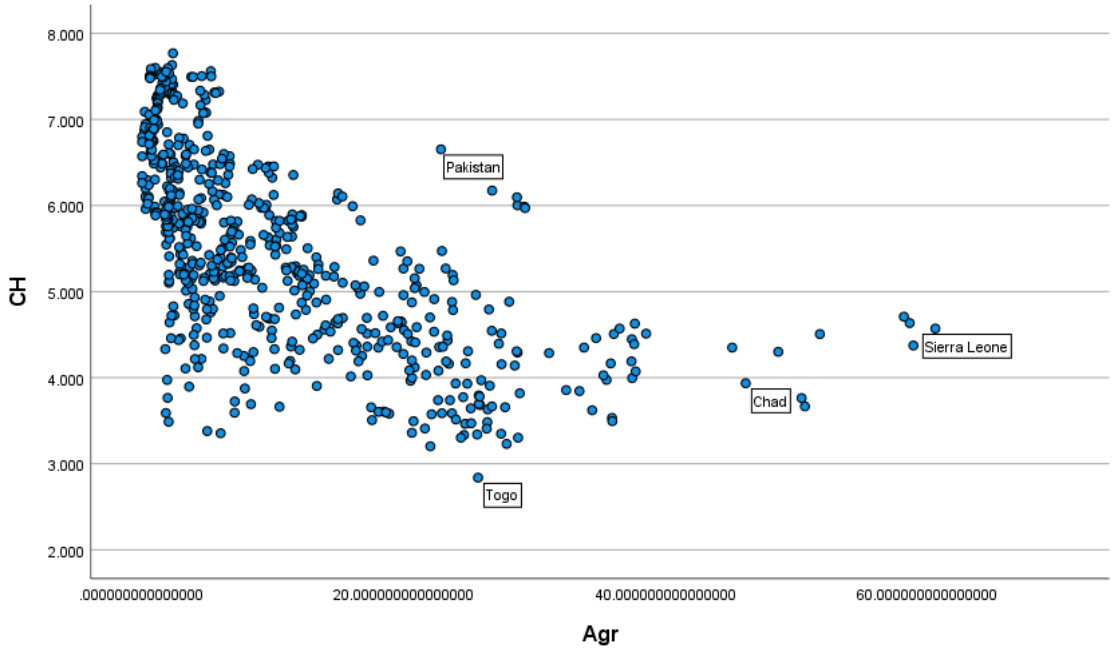
across all countries is 10.983%. The highest amount of government expenditures spent on health is 29.492%, which was Costa Rica's spending in 2016. Cameroon has the lowest value, at 0.734% in 2017. Figure 17 shows the scatterplot for country happiness as a function of domestic government health expenditures. This graph shows a general trend upwards as government health expenditures increase, with some outliers further from the general grouping of points.

Figure 17: Country Happiness as a Function of Domestic Government Health Expenditures



Agriculture, Forestry, & Fishing Value Added. Data for this variable was gathered from The World Bank for 127 countries and is measured as a percent of the total GDP. The average percentage for all countries is 10.363%, while the lowest value is Singapore's 2018 value of 0.030% and the highest is Sierra Leone's 2017 value of 60.611%. Figure 18 displays the scatterplot for country happiness as a function of agriculture, forestry, and fishing value added. This graph has a trend downwards as the value added increases, with a slight trend up at the highest values. Pakistan, Chad, and Sierra Leone are noticeable outliers from the general trend.

Figure 18: Country Happiness as a Function of Agriculture, Forestry, & Fishing Value Added



Access to Electricity. Data for this variable comes from The World Bank for 128 countries and is the percent of the total population in a country that has access to electricity. The average value is 85.379%. Several countries have 100% access to electricity for multiple years, while Chad's population had the lowest access to electricity at 7.700% in 2015. Figure 19 shows the scatterplot for country happiness as a function of access to electricity, which has a trend upwards as the percentage of the population with access to electricity increases. The large clustering to the far right shows that several countries have a population with 100% access to electricity.

Greenhouse Gas Emissions Per Capita. Data for this variable is from The World Bank and is for 127 countries, with the average emissions per capita being 0.431 kt of CO2 equivalent. Figure 20 shows country happiness as a function of Greenhouse Gas Emissions per capita.

Figure 19: Country Happiness as a Function of Access to Electricity

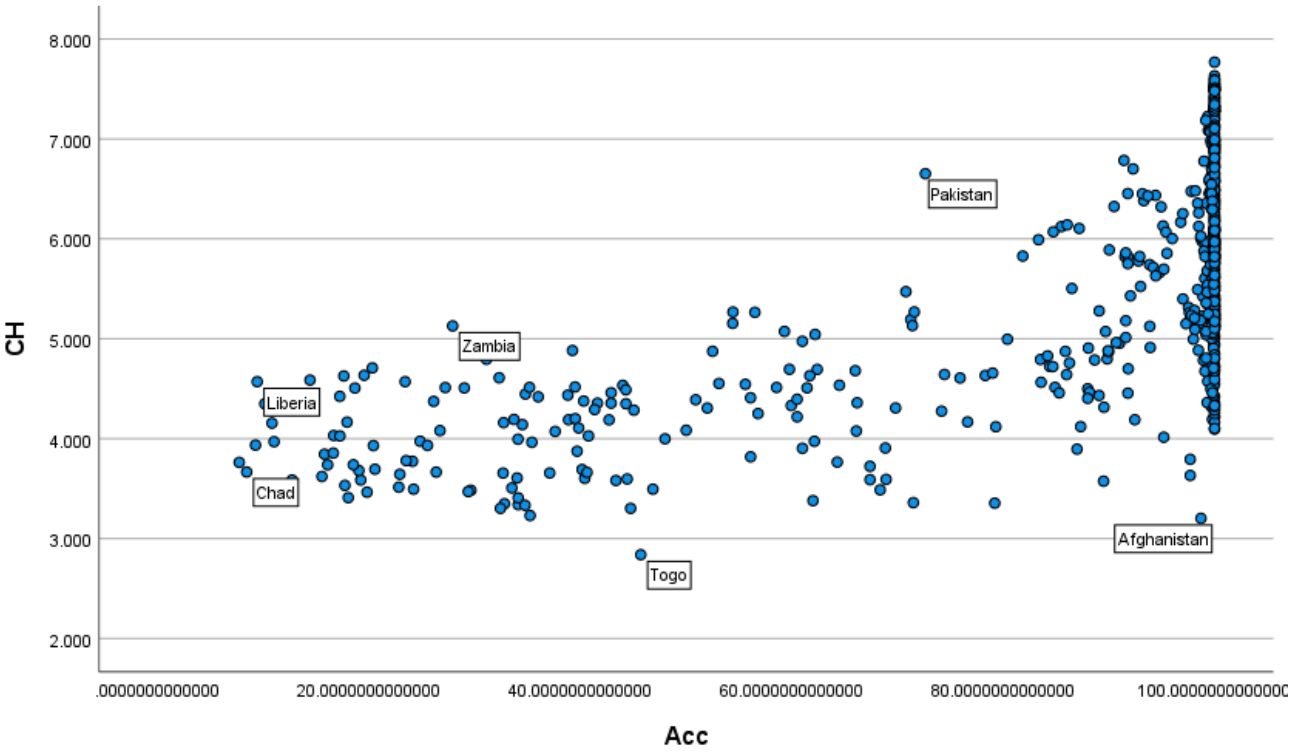
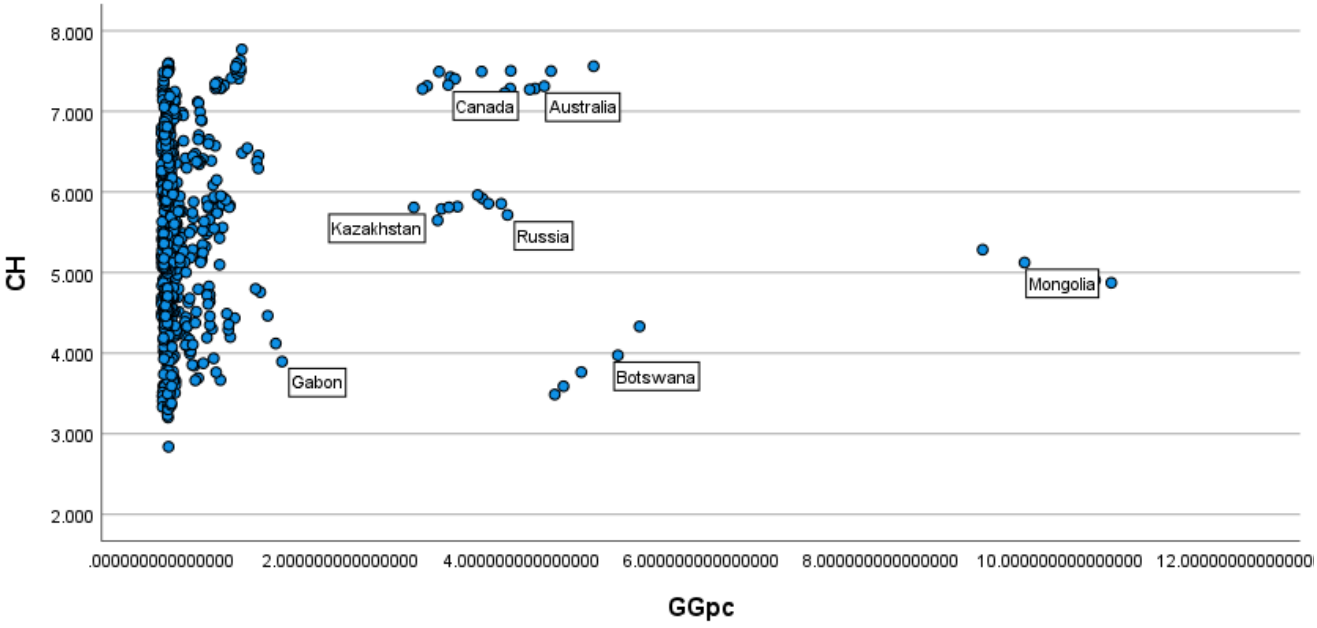


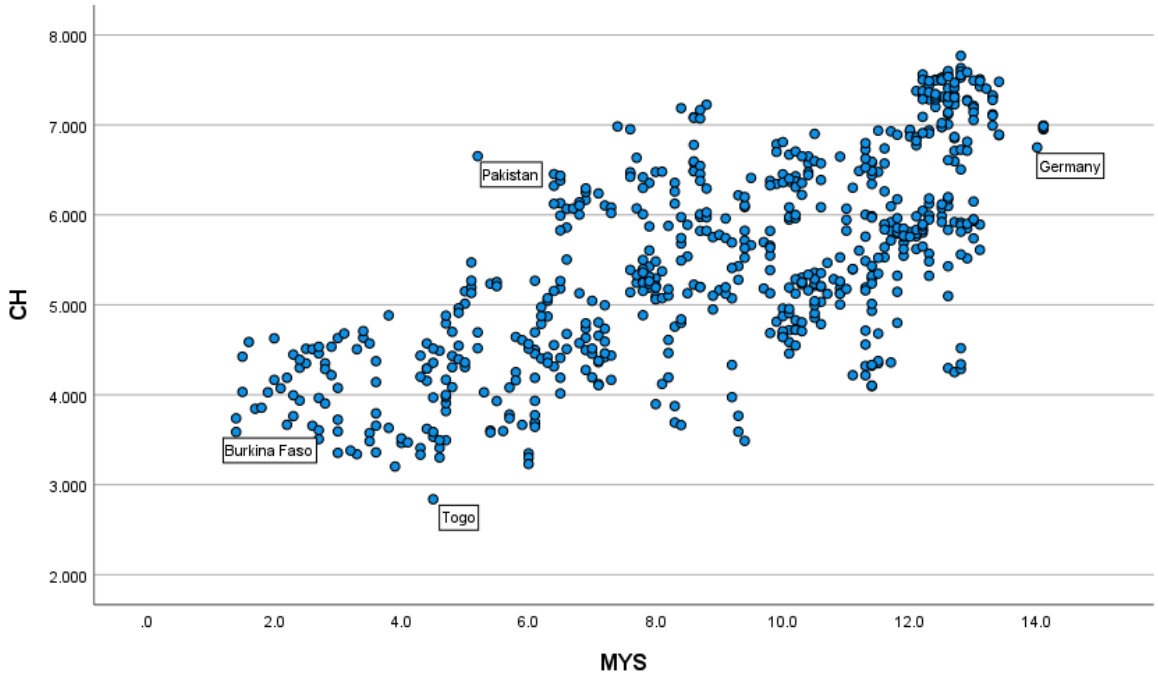
Figure 20: Country Happiness as a Function of Greenhouse Gas Emissions Per Capita



Mongolia has the highest greenhouse gas emissions per capita, with a value of 10.568 kt of CO2 equivalent in 2014, while Singapore had the lowest amount in 2017 at 0.001 kt of CO2 equivalent. Looking at the graph, there is no noticeable trend because most points are clustered at a value below 2 kt of CO2 equivalent, and there are a few outliers with very high levels of greenhouse gas emissions.

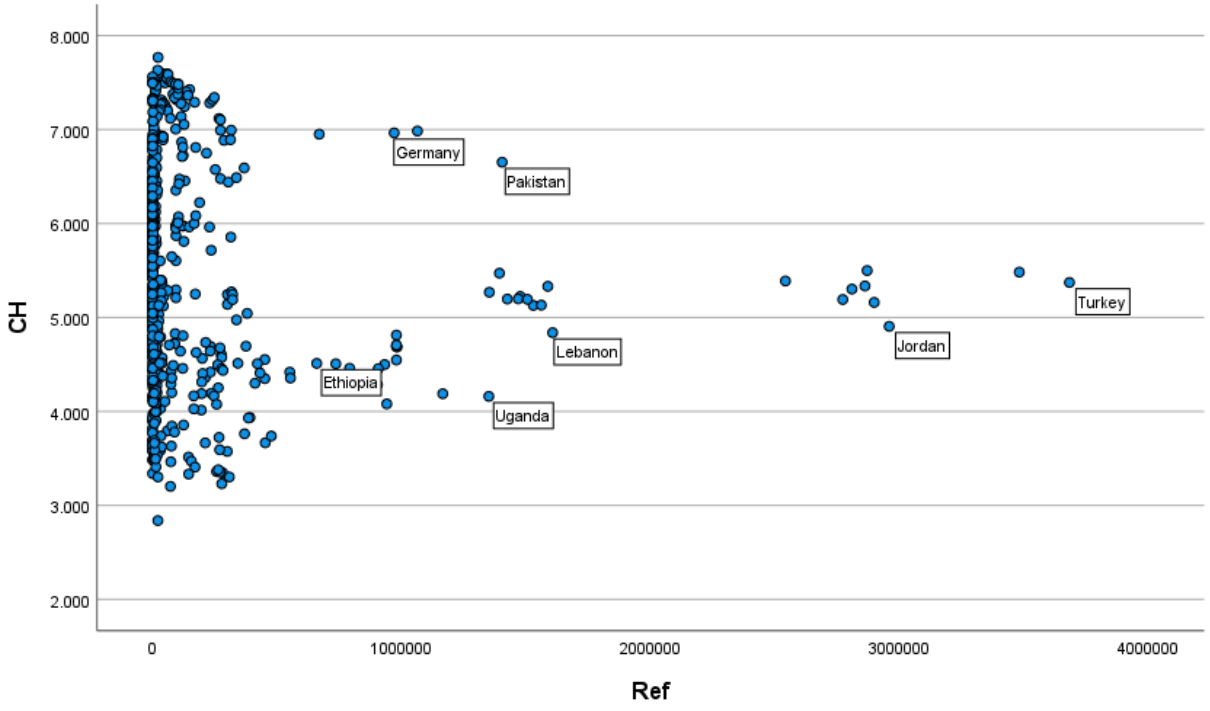
Mean Years of Schooling. Data on the mean years of schooling is collected from the Human Development Reports and is for 127 countries. The average mean years of schooling for all the countries is 8.968 years. Germany has the highest mean years of schooling at 14.1 years for 2015-2018, while Burkina Faso has the lowest mean years of schooling at 1.4 years for 2014 and 2015. Figure 21 displays the scatterplot for country happiness as a function of mean years of schooling. The points trend upwards as the mean years of schooling increase, but are spread out.

Figure 21: Country Happiness as a Function of Mean Years of Schooling



Refugee Population by Country of Asylum. Data for this variable comes from The World Bank for 124 countries and is the total number of refugees a country has taken in as of that year. The average number of refugees is 150,423.03 people, with the highest being 3,681,688 people in Turkey in 2018, while the lowest population is 5 refugees in Mongolia in 2014, 2017, and 2018. Figure 22 shows the scatterplot for country happiness as a function of the refugee population. Most points are clustered on the left at or around zero people with a few outliers as the refugee population increases, which trend toward a country happiness score of 5 points.

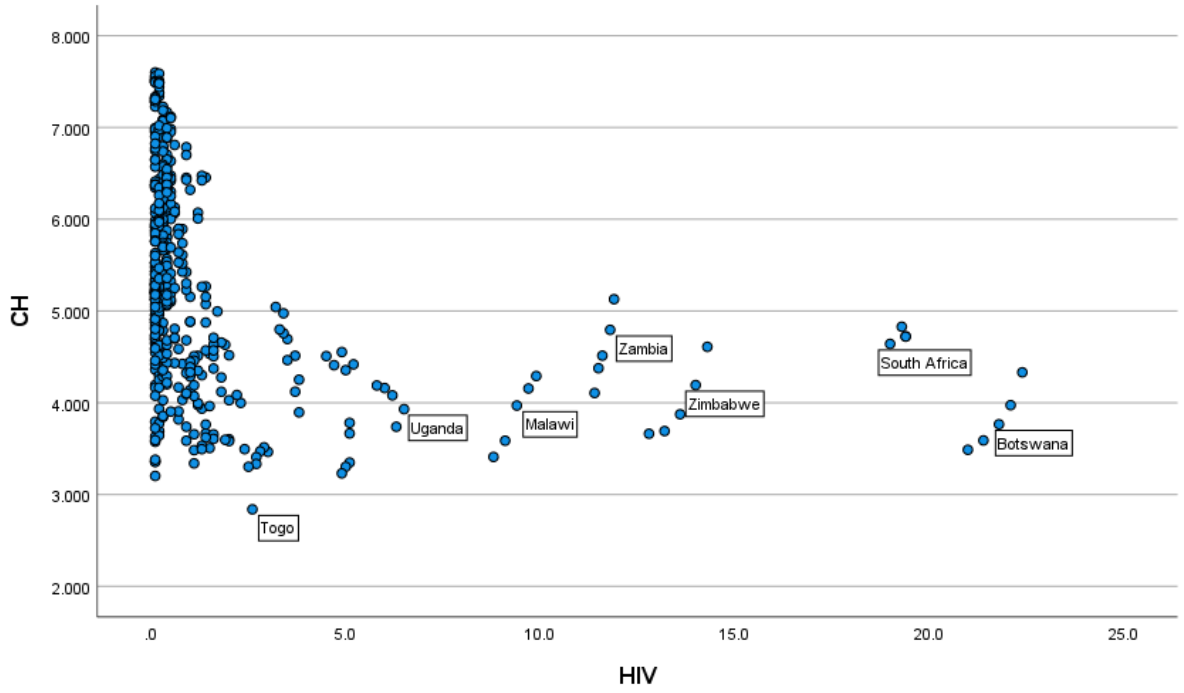
Figure 22: Country Happiness as a Function of the Refugee Population



Prevalence of HIV. Data for this variable is from The World Bank for 96 countries and is the percent of the total population ages 15 to 49 infected with HIV. The average prevalence is 1.461%, several countries have the lowest value of 0.1% of the population, and Botswana has the highest value, with 22.4% in 2014. Figure 23 presents the scatterplot for country happiness as a

function of the prevalence of HIV. Most countries are on the left side with a low prevalence, while a few countries are outliers with a high prevalence and low country happiness.

Figure 23: Country Happiness as a Function of the Prevalence of HIV



Forest Area. Data for the amount of forest area comes from The World Bank and shows the amount of forest area as a percent of total land a country has, with the average being 28.917%. Gabon has the largest amount of forest area with a value of 91.597% in 2014, and Egypt has the smallest amount of forest area with a value of 0.045% in 2016. Figure 24 shows the scatterplot for country happiness as a function of forest area. This graph shows that the data points are well spread throughout the values of 0% to about 70% of land area, with a few outliers having forested area above 70% of the total land area.

Average Annual GDP Growth Rate. Figure 25 shows the scatterplot for country happiness as a function of the average annual GDP Growth Rate over the last five years. This

Figure 24: Country Happiness as a function of Forest Area

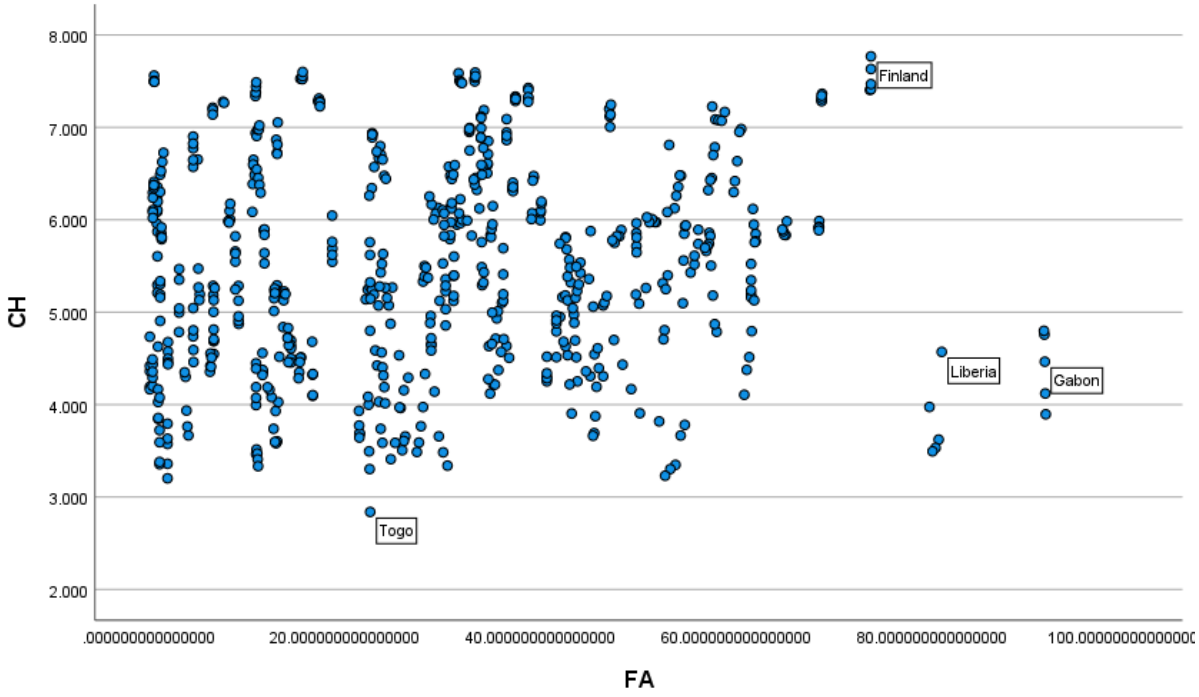
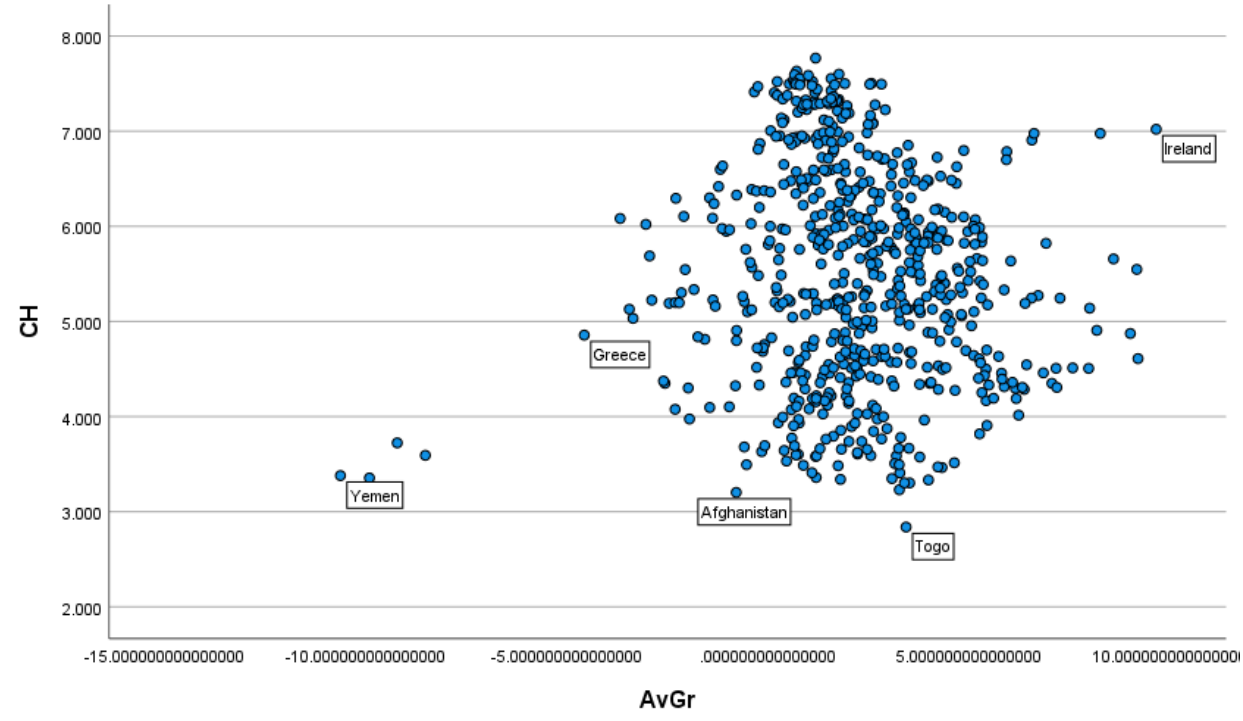


Figure 25: Country Happiness as a function of the Average Annual GDP Growth Rate



variable is measured as the average annual GDP growth rate over the last five years as a percent of GDP. The average value is 2.206% while the lowest value is -10.636% which is Yemen's average annual growth rate for 2018, and the highest value is 9.653% which is Ireland's average annual growth rate for 2018. Looking at the scatterplot, most countries have an average annual growth rate between about 0% and 7.5%, with Yemen being an outlier to the left and Ireland being an outlier to the right.

Results

Table 9 displays the results for each estimated regression.

Table 9: Country Happiness Model Regression Results

Variables	1	2	3	4	5	6	7
a	-2.317 (-0.613) [3.780]	-2.469 (-0.711) [3.472]	42.660 (4.675)* [9.125]	28.961 (6.027)* [4.805]	38.827 (6.703)* [5.793]	31.897 (6.521)* [4.891]	30.815 (6.037)* [5.105]
EFI	-0.008 (-1.072) [0.008]	-0.013 (-1.675) [0.008]	-0.009 (-1.525) [0.006]	-0.008 (-1.701) [0.005]	-0.002 (-0.334) [0.005]	-0.007 (-1.570) [0.005]	-0.006 (-1.222) [0.005]
Corr	0.015 (2.630)* [0.006]	0.022 (4.082)* [0.005]	0.001 (0.111) [0.005]				
Gini	0.024 (2.595)* [0.009]	0.032 (3.599)* [0.009]	0.042 (6.458)* [0.007]	0.040 (6.369)* [0.006]	0.019 (2.721)* [0.007]	0.036 (5.489)* [0.006]	0.031 (4.910)* [0.006]
MR	-0.006 (-0.362) [0.017]	-0.039 (-2.642)* [0.015]	-0.029 (-2.191)* [0.013]	-0.033 (-3.088)* [0.011]	-0.051 (-4.094)* [0.013]	-0.038 (-3.546)* [0.011]	-0.028 (-2.515)* [0.011]
SM	-0.014 (-1.215) [0.012]	-0.024 (-2.045)* [0.012]	-0.014 (-1.563) [0.009]	-0.019 (-2.430)* [0.008]	-0.009 (-1.462) [0.006]	-0.017 (-3.546)* [0.008]	-0.021 (-2.568)* [0.008]
U	-0.057 (-6.243)* [0.009]	-0.060 (-6.979)* [0.009]	-0.095 (-4.474)* [0.021]	-0.060 (-9.469)* [0.006]	-0.057 (-7.453)* [0.008]	-0.059 (-9.519)* [0.006]	-0.058 (-8.654)* [0.007]
U^2			0.002 (1.718) [0.001]				
FLP	0.042 (3.230)* [0.013]	0.022 (1.764) [0.012]	-0.006 (-0.526) [0.011]				
UP	0.001 (0.328) [0.004]	0.001 (0.130) [0.004]	-0.082 (-3.983)* [0.021]	-0.098 (-6.298)* [0.016]	-0.114 (-6.249)* [0.018]	-0.103 (-6.637)* [0.015]	-0.121 (-7.778)* [0.016]

UP^2			0.001 (4.538)* [0.000]	0.001 (7.062)* [0.000]	0.001 (6.903)* [0.000]	0.001 (7.478)* [0.000]	0.001 (9.005)* [0.000]
PSF	-0.009 (-0.939) [0.009]	-0.024 (-2.761)* [0.009]	-0.033 (-2.783)* [0.012]	-0.035 (-5.658)* [0.006]	-0.031 (-4.294)* [0.007]	-0.039 (-6.180)* [0.006]	-0.030 (-4.691)* [0.006]
PSF^2			2.009E-5 (0.057) [0.000]				
Den	0.001 (1.865) [0.000]	0.001 (1.886) [0.000]	0.001 (3.304)* [0.000]	0.001 (3.195)* [0.000]	-0.001 (-3.697)* [0.000]	0.001 (2.858)* [0.000]	0.001 (3.607)* [0.000]
PEld	-0.101 (-4.178)* [0.024]						
PYou	-0.078 (-2.624)* [0.030]						
PMid		0.100 (4.020)* [0.025]	-0.463 (-1.350) [0.343]	0.086 (4.871)* [0.018]	0.020 (1.142) [0.017]	0.087 (5.005)* [0.017]	0.077 (4.124)* [0.019]
$PMid^2$			0.004 (1.600) [0.003]				
MI	0.045 (3.610)* [0.012]						
LEB	0.116 (3.159)* [0.037]	0.019 (0.740) [0.026]	-0.616 (-3.190)* [0.193]	-0.716 (-5.814)* [0.123]	-0.814 (-5.345)* [0.152]	-0.771 (-6.251)* [0.123]	-0.758 (-5.792)* [0.131]
LEBsq			0.005 (3.412)* [0.001]	0.005 (6.021)* [0.001]	0.006 (5.241)* [0.001]	0.005 (6.376)* [0.001]	0.005 (5.937)* [0.001]
TFR	0.246 (1.113) [0.221]	0.434 (2.568)* [0.169]	0.328 (2.322)* [0.141]	0.375 (3.500)* [0.107]	0.106 (0.888) [0.120]	0.365 (3.465)* [0.105]	0.330 (2.896)* [0.114]
%Hgov	0.034 (2.766)* [0.012]	0.028 (2.236)* [0.013]	0.026 (2.777)* [0.009]	0.026 (3.075)* [0.008]	0.027 (3.138)* [0.009]	0.026 (3.202)* [0.008]	0.028 (3.107)* [0.009]
Agr	-0.063 (-5.393)* [0.012]	-0.045 (-4.054)* [0.011]	-0.081 (-5.820)* [0.014]	-0.080 (-6.002)* [0.013]	-0.063 (-4.016)* [0.016]	-0.078 (-6.007)* [0.013]	-0.075 (-5.345)* [0.014]
Agrsq			0.001 (3.213)* [0.000]	0.001 (3.219)* [0.000]	0.000 (1.504) [0.000]	0.001 (3.296)* [0.000]	0.001 (3.081)* [0.000]
Acc	-0.006 (-0.809) [0.007]	-0.013 (-1.690) [0.007]	0.024 (1.047) [0.023]	0.015 (2.529)* [0.006]	0.029 (4.641)* [0.006]	0.019 (3.135)* [0.006]	0.021 (3.424)* [0.006]
Acc^2			-3.223E-5 (-0.181)				

			[0.000]				
GGpc	0.015 (0.392) [0.037]	0.042 (1.111) [0.038]	0.073 (2.263)* [0.032]	0.071 (2.550)* [0.028]	0.103 (3.570)* [0.029]	0.065 (2.359)* [0.027]	0.097 (3.372)* [0.029]
MYS	0.007 (0.232) [0.031]	0.020 (0.619) [0.032]	-0.631 (-4.724)* [0.134]	-0.611 (-4.877)* [0.125]	-0.645 (-4.218)* [0.153]	-0.672 (-5.335)* [0.126]	-0.519 (-3.946)* [0.132]
MYSsq			0.035 (5.187)* [0.007]	0.034 (5.396)* [0.006]	0.034 (4.460)* [0.008]	0.036 (5.784)* [0.006]	0.029 (4.432)* [0.007]
Ref	-7.294E-7 (-2.561)* [0.000]	-1.149E-6 (-4.300)* [0.000]	-1.569E-6 (-6.957)* [0.000]	-1.445E-6 (-8.277)* [0.000]	-8.555E-7 (-4.299)* [0.000]	-1.432E-6 (-8.349)* [0.000]	-1.161E-6 (-6.760)* [0.000]
HIV	-0.099 (-3.624)* [0.027]	-0.073 (-2.720)* [0.027]	-0.010 (-0.425) [0.023]	-0.006 (-0.290) [0.020]		-0.192 (-2.306)* [0.083]	-0.002 (-0.091) [0.021
HIVdum						0.195 (2.300)* [0.085]	
FA	-0.012 (-5.053)* [0.002]	-0.012 (-5.013)* [0.002]	-0.008 (-3.979)* [0.002]	-0.008 (-4.208)* [0.002]	0.001 (0.451) [0.002]	-0.006 (-3.094)* [0.002]	
FAdum					-0.003 (-0.795) [0.004]		
AvGr	-0.049 (-1.790) [0.027]	-0.058 (-2.093)* [0.028]	-0.005 (-0.223) [0.021]	-0.011 (-0.537) [0.020]		-0.010 (-0.477) [0.020]	-0.007 (-0.337) [0.022]
R^2	0.889	0.876	0.944	0.941	0.883	0.943	0.932
Adj R^2	0.867	0.854	0.929	0.929	0.869	0.931	0.919
F	40.034	39.325	63.913	79.509	60.888	79.268	72.224
Critical F	1.608	1.630	1.559	1.608	1.588	1.599	1.619
Critical t	1.9799	1.9798	1.981	1.9799	1.9729	1.9801	1.9798
N	145	145	145	145	209	145	145
(t statistic) *t is significant at the 0.05 level [coefficients standard error]							

Regression 1. The first regression includes all independent variables in the country happiness model. The adjusted R^2 value is 0.867 meaning that 86.7% of the variation in country happiness can be explained by the variation in the independent variables included in this regression, once the number of independent variables is considered. The calculated F-ratio is 40.034. For this regression, N is 145 and K is 25, so N-K = 145-25 = 120 and K-1 = 25-1 = 24, meaning the F critical value is 1.609. Because the calculated F value is greater than the F critical value, the null

hypothesis can be rejected and the alternative hypothesis can be affirmed, meaning that the regression is statistically significant at the 0.05 level and there is a relationship between the independent variables in this regression and country happiness.

Using 120 as the degrees of freedom and a 0.05 significance level, the t critical value is 1.9780. The absolute value of the t statistics for the independent variables Corr, FA, Gini, HIV, U, FLP, PEld, PYou, MI, LEB, %Hgov, Agr, and Ref are greater than the t critical value which means that the null hypothesis for the two-tailed t test for these variables can be rejected and the alternative hypothesis can be affirmed. This means that the coefficients for these variables are statistically significant. As expected, Corr, FLP, UP, Den, LEB, %Hgov, and MYS have a positive relationship with country happiness, while HIV, MR, SM, U, PSF, PEld, PYou, Agr, and Ref have a negative relationship with country happiness. While FA, Gini, and MI have statistically significant coefficients, they have unexpected signs. The t statistics for the constant and the independent variables EFI, MR, SM, UP, PSF, Den, TFR, Acc, GGpc, MYS, and AvGr are less than the t critical value meaning that the null hypothesis cannot be rejected and it cannot be concluded that the coefficients for these values are statistically significant. Additionally, EFI, TFR, Acc, GGpc, and AvGr have unexpected signs, but do not have statistically significant coefficients.

Regression 2. In the estimation of the second regression, the variables infant mortality, percent of the population ages 65 and over, and percent of the population ages 14 and under were removed because each of these variables had severe multicollinearity with other variables, meaning they each had at least one instance with another variable where $r \geq 0.70$. The independent variable percent of the population ages 15 to 64 (PMid) was also added in order to keep a variable that represents the age demographics of the population. The adjusted R^2 value of this regression

decreased slightly compared to the first regression and is 0.854, showing that 85.4% of the variation in country happiness can be explained by the variation in the independent variables once the number of independent variables is considered.

In this regression, the t critical value is 1.9798, and the coefficients for MR, SM, PSF, TFR, and AvGr are now significant; however, LEB is no longer significant. While the coefficient for AvGr is now significant, it still has an unexpected negative relationship with country happiness. The coefficient of the added variable, PMid, has an expected positive relationship with country happiness and is statistically significant. Additionally, FLP is no longer significant, but still has the expected positive relationship with country happiness.

Regression 3. In estimating the third regression, all variables from the second regression were kept and several variables were given a quadratic specification to examine evidence of a curved relationship between each respective variable and country happiness. The variables that were squared are U, UP, PSF, PMid, LEB, Agr, Acc, and MYS. The adjusted R^2 value is 0.929, meaning 92.9% of the variation in country happiness can be explained by the variation in the independent variables once the number of independent variables is considered. This is an increase compared to the first two regressions.

The t critical value is 1.981, and the constant and the coefficients for UP, Den, LEB, GGpc, and MYS are now significant. The coefficients for the added variables UP^2, LEB^2, Agr^2, and MYS^2 are also significant. The coefficient for the variable Acc has now changed to its positive expected relationship with country happiness, but is not significant. The coefficients for UP^2, $PMid^2$, LEB^2, and MYS^2 have a positive relationship with country happiness; however, the coefficient for $PMid^2$ is not significant. The coefficients for Corr, HIV, SM, AvGr, and PMid are no longer significant, and AvGr and PMid have unexpected negative relationships with country

happiness. While significant, the coefficients for FA, Gini, UP, LEB, TFR, Agr^2, GGpc, and MYS have unexpected signs. The coefficients for the variables FLP, U^2, PSF^2, and Acc^2 have unexpected signs, but are not statistically significant.

Regression 4. In the estimation of the fourth regression, The quadratic specifications for U, PSF, PMid, and Acc were removed because the coefficients for these variables did not have significant t statistics. Additionally, Corr was removed because of its severe multicollinearity with EFI, where $r = 0.783$, and its smaller t statistic compared to EFI, FLP was also removed because of its coefficient's small t statistic and continually changing sign.

The t critical value is 1.9799 at the 0.05 significance level and the constant and all of the coefficients for the variables in this regression are significant, excluding HIV, AvGR, and EFI. Of the coefficients that have significant t statistics, FA, Gini, UP, LEB, TFR, GGpc, and MYS have unexpected signs. AvGr and EFI continue to have an unexpected negative sign, but their coefficients are not significant.

Regression 5. In the estimation of the fifth regression, HIV and AvGr were removed because of their coefficient's low t statistics and the unexpected sign of AvGr. An interactive dummy variable created for forest area was also included in which all the countries in Africa kept their original value for forest area, and all other countries were given a value of zero. This was done in order to better understand why FA may be having a negative sign, because African countries have a high amount of forested areas, but are typically very rural and underdeveloped meaning the forested area could decrease happiness by being an obstacle or danger rather than a chance to connect with nature. The adjusted R^2 value for this regression is 0.869 meaning 86.8% of the variation in country happiness can be explained by the variation in the independent variables

while accounting for the number of independent variables. This value has decreased compared to the last two regressions.

The t critical value is 1.9729, and all of the coefficients that were significant in the fourth regression are still significant, except for SM, PMid, TFR, Agr^2, and FA. The coefficient for population now has an unexpected negative coefficient that is also significant. Additionally, the coefficient for FA now has the expected positive relationship with country happiness, though it is not significant. The coefficient for FAdum has the expected negative sign, but is not significant. The same coefficients as in regression four continue to have unexpected negative signs.

Regression 6. In the estimation of the sixth regression, the interactive dummy variable for forest area was removed because of its low t statistic and the other coefficients that had low t statistics once FAdum was included. An interactive dummy variable for HIV, HIVdum, was also included in this regression and the variables HIV and AvGr were added back in. In the creation of the interactive dummy variable for HIV, all countries with a prevalence of or over 5% kept their value and all countries underneath 5% had their values changed to zero to better see the effects of high levels of HIV. The adjusted R^2 value for this regression is 0.931, meaning that 93.1% of the variation in country happiness can be explained by the variation in the independent variables, once the number of independent variables is considered. This is also the highest adjusted R^2 value of all the estimated regressions thus far.

The t critical value is 1.9803. The coefficients for all of the variables are now statistically significant, except for the coefficients for EFI, and AvGr. Population density now has its expected positive relationship with country happiness again, but Forest Area is unexpectedly negative again. Additionally, the coefficient for HIVdum is unexpectedly positive and

significant. All the coefficients of the variables from the fifth regression that had an expected negative relationship continue to do so.

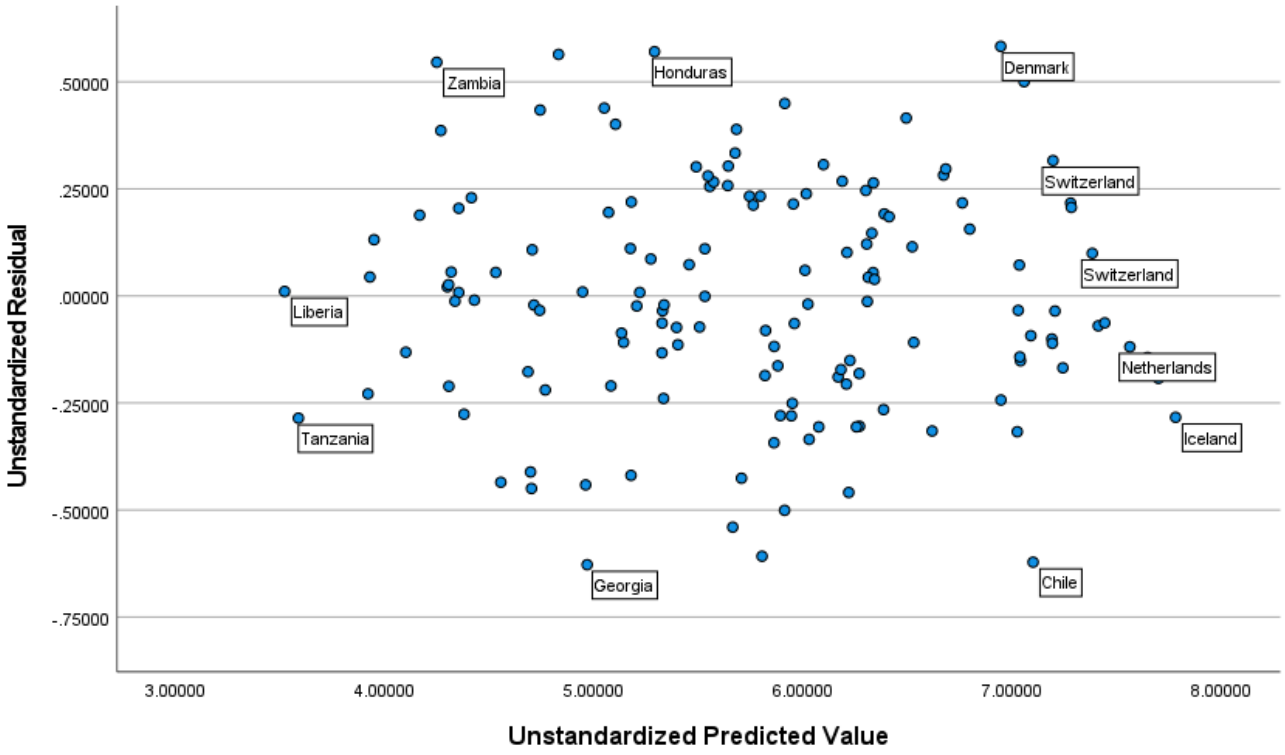
Regression 7. In the seventh regression, both HIVdum and FA were dropped because of their unexpected signs. The adjusted R^2 value for this regression is 0.919, a slight decrease from regression six. The t critical value is 1.9798 and the constant and all the coefficients are significant except for the coefficients for EFI and AvGr which also continue to have unexpected negative signs.

Interpreting Regression Coefficients. Based on the results of these regressions, the seventh regression can be considered the best regression because of its several significant t statistics, high adjusted R^2 value, and removal of some severe multicollinearity. Though this regression has many coefficients with unexpected signs, many of these surprising signs have logical explanations or possibilities once interpreted and the data for that variable is considered. Figure 26 shows the residual plot for regression seven. This plot shows that there is a large amount of randomization, with some errors slightly outside of the general clustering. In total, data for 53 countries from every region excluding Oceania, are included in this regression. On the far right of the plot, the grouping of points that resembles a downward sloping line includes the countries Denmark, Switzerland, The Netherlands, and Iceland. The two farthest left points are Liberia and Tanzania.

Constant. The constant ‘a’ has a calculated value of 30.815 and represents the y-intercept for the graph of the country happiness function. This value means that even when all the other independent variables are 0, country happiness would still be expected to be 30.815; however, this value has no economic meaning.

Economic Freedom Index. The coefficient for the Economic Freedom Index in regression

Figure 26: Residual Plot for Regression 7



six is calculated to be -0.006. This means that as a country's score on the economic freedom index increases by one point, average country happiness can be expected to decrease by 0.006 index points. This coefficient was expected to have a positive sign which would reflect that as individuals have more freedom in their economic choices, they are generally happier. The negative sign for this coefficient could be explained by the possibility, as one study has found, that one form of economic freedom–more limited and smaller governments–actually decreases happiness because there are less government consumption expenditures which can actually harm well-being of a society (Yilmaz and Tag, 2016, pg 3). However, this coefficient is also statistically insignificant meaning the possibility that the coefficient is zero cannot be ruled out.

Gini Coefficient. The coefficient for the Gini coefficient variable in this regression is estimated to be 0.031 indicating that as a country's Gini coefficient increases by one point, the average country happiness increases by 0.031 points. This coefficient was expected to have a negative sign because as inequality in a country increases, its people face the stress and consequences of living in an unequal society. However, as discussed in the literature review section of this paper, the coefficient for the Gini index may be positive because a more unequal society reflects higher social mobility and the opportunity for people to increase their wealth and place in society (Bjornskov, 2003, pg 8).

Mortality Rate from CVD, Cancer, Diabetes, or CRD. The coefficient for the variable MR has its expected negative sign, demonstrating that as mortality from cardiovascular diseases, cancer, diabetes, and chronic respiratory diseases increases, happiness decreases because more individuals are dealing with difficult health problems, expensive healthcare bills, and loss of loved ones. The coefficient is estimated to be -0.028 meaning that as the mortality rate from these diseases increases by one percentage point in a country, average country happiness decreases by 0.028 index points.

Suicide Mortality Rate. The calculated coefficient for the suicide mortality rate is -0.021. This expected negative sign demonstrates that as sucide mortality rates increase, happiness decreases because this variable demonstrates poor mental health and loss of loved ones. This value means that as the suicide mortality rate of a country increases by one person per 100,000 people, average country happiness decreases by 0.021 index points.

Unemployment. The estimated value of the coefficient for the unemployment rate is -0.058. This negative relationship is expected and shows that as unemployment rates increase, stress and uncertainty about the job market and potential loss of one's income also increases,

decreasing happiness. This value means that for every one percentage point increase in a country's unemployment rate, the average country happiness is expected to decrease by 0.058 index points.

Urban Population. The coefficient for the urban population is -0.121 and the coefficient for the quadratic specification of the urban population is 0.001. This means that urban population initially has a negative relationship with country happiness when the percentage of the population living in urban areas is low, until the urban population eventually reaches a turning point where it begins to have a positive relationship with country happiness, forming a U-shaped scatterplot. Based on the coefficients, this turning point is when the urban population reaches 60.5% of the total population. The incremental changes vary along this curve. For example, when the urban population is at 10%, an incremental change of a one percentage point increase results in a decrease in the country happiness score by 0.1010 points. On the other hand, when the urban population is at 80% an incremental change of a one percentage point increase leads to an increase in the country happiness score by 0.0390 points.

Prevalence of Severe Food Insecurity.The estimated value of the coefficient for variable PSF is -0.030. This coefficient has its expected negative sign, demonstrating that as more individuals in a country experience severe food insecurity, average country happiness decreases because more individuals are struggling to feed themselves and their families meaning they face hunger, related health problems, and financial stress. This value shows that as the prevalence of severe food insecurity in a country increases by one percentage point, average country happiness decreases by 0.030 index points.

Population Density. The value of the coefficient for population density is 0.001. This means the variable has its expected positive relationship with country happiness because

increasing population density reflects increased access to public amenities which increases happiness as more people experience the benefits of living in a country where they have access to numerous public facilities (Ye et al., 2014). This value means that as the number of people per square kilometer increases by one person, average country happiness increases by 0.001 index points.

Population Ages 15-64. The estimated value of the coefficient for the percent of the population ages 15 to 64 is 0.077. This positive relationship with country happiness is as expected and shows that average country happiness increases as the percent of the population that is middle aged increases because there are more individuals that are fiscally independent and able to support the younger and older generations that need it. This value means that as the percent of the population in a country aged 15 to 64 increases by one percentage point, average country happiness increases by 0.077 index points.

Life Expectancy at Birth. The calculated value for this variable's coefficient is -0.758 and the coefficient for the quadratic specification is 0.005. These coefficients mean there is curvature in the relationship between life expectancy and country happiness where the relationship is initially negative and then reaches a turning point and becomes positive, in this case the turning point is 75.8 years as calculated using the values of both coefficients. This means that for countries that have a life expectancy at birth of 75.8 years or higher, this variable increases happiness, while in countries with a life expectancy lower than 75.8, this variable decreases happiness. For example, in a country with a life expectancy of 55 years, a one year incremental change in life expectancy would still decrease happiness by 0.2080 points, while in a country with a life expectancy at birth of 80 years, a one year incremental increase in life expectancy would increase country happiness by 0.0420 index points.

Total Fertility Rate. The coefficient for total fertility rate has an estimated coefficient of 0.330. This means the total fertility rate has an unexpected positive relationship with average country happiness, as it was expected to have a negative sign because higher fertility rates means women are experiencing more stress and negative health effects from having more children, and high fertility rates are common in countries where families need more children in order for them to work and support their families, a sign of underdevelopment, all of which decreases happiness. This positive relationship could simply be explained by the fact that having children also often brings joy to people's lives and thus increases happiness, or perhaps that having multiple children could be seen as a sign of success for the mother. The relationship between fertility rates and happiness is likely largely dependent on the characteristics of the groups that have such high fertility rates such as their age, partnership status, health, and socioeconomic status (Margolis & Myrskyla, 2011). This value means that as the average number of children a woman births during her childbearing years increases by one child, the average country happiness increases by 0.330 index points.

Domestic General Government Health Expenditures. The coefficient for the variable %Hgov is estimated to be 0.028. This is the expected positive relationship with country happiness because as government spending on health increases, the government is prioritizing its people's health more, meaning they are likely receiving better healthcare and subsequently, better health which increases happiness. This value means that as the percent of the government's general expenditures from domestic sources that is going towards health increases by one percentage point, average country happiness increases by 0.028 index points.

Agriculture, Forestry, & Fishing Value Added. The estimated value of the coefficient for the variable Agr is -0.075 and the coefficient for the quadratic specification is 0.001. This means

the value added from these industries initially has a negative relationship with country happiness until it reaches a turning point, which in this case is at 37.5%, where it then has a positive relationship with country happiness as it continues increasing. Using these coefficients to calculate the incremental change in country happiness at specific points, it is found that when the value added from these industries is 20% of GDP, an incremental increase of one percentage point decreases country happiness by 0.0350 points. On the other hand, when value added is at 50% of GDP, a one percentage point increase actually increases country happiness by 0.0250 points.

Access to Electricity. The calculated value of the coefficient for access to electricity is 0.021. This coefficient has its expected sign, showing that as the percentage of the population with access to electricity increases, the country's happiness increases because more people are experiencing the benefits of having electricity such as higher productivity, access to electronic consumer goods, and better safety (Huges, 2018). This value means that as the proportion of the population with access to electricity in a country increases by one percentage point, average country happiness increases by 0.021 index points.

Greenhouse Gas Emissions Per Capita. The coefficient for this variable has a value of 0.097. This means that the greenhouse gas emissions per capita has an unexpected positive relationship with country happiness meaning that as emissions per capita increases, happiness increases. This was expected to be a negative relationship because increasing emissions reflects worsening air quality and pollution, which can negatively impact one's health and enjoyment of the outdoors, decreasing happiness. This positive relationship could be reflecting other variables not included in the regression, such as high production levels and availability of energy that cause greenhouse gas emissions, but tend to increase happiness. This value means that as the

greenhouse gas emissions per capita increase by one kiloton of CO2 equivalent, average country happiness increases by 0.097 index points.

Mean Years of Schooling. The estimated coefficient for mean years of schooling is -0.672, while the coefficient of the quadratic specification of this variable is 0.029. This negative and positive sign demonstrate a curved relationship where mean years of schooling initially has a negative relationship with country happiness until the years increase to where they reach a turning point and begin having a positive relationship with country happiness. Based on the values of these coefficients, the turning point is 8.948 years of schooling. Looking closer at this relationship, when the mean years of schooling in a country is three years, an incremental change of one year will lead to a decrease in country happiness by 0.345 points. In a country with a mean years of schooling of 10 years, an incremental change of one year will lead to an increase in happiness by 0.0610 index points.

Refugee Population by Country of Asylum. The coefficient for the variable Ref has a calculated value of -1.161E-6. This variable has the expected negative relationship with country happiness demonstrating that as a country takes in an increasing number of refugees, there is increasing stress and burden on that country of asylum which adds to similar problems they may already be dealing with because these countries of asylum are often neighboring countries to the ones where the refugees come from. This increased stress decreases happiness. This value means that as the population of refugees in the country of asylum increases by one person, average country happiness decreases by 0.00000161 index points. In other words, if a country were to take in an additional 100,000 refugees, their average country happiness score would decrease by 0.161 index points.

Prevalence of HIV. The estimated coefficient for HIV has a value of -0.002. This negative value is expected because as the percent of individuals infected with HIV increases in a country, average country happiness decreases because there are more people in that country dealing with worsening health, expensive bills, and loss of loved ones. This value means that as the percent of the population infected with HIV increases by one percentage point, average country happiness decreases by 0.002 index points. However, the coefficient is not statistically significant and the possibility cannot be ruled out that the coefficient is zero.

Average Annual GDP Growth Rate. The estimated value of the coefficient for average annual GDP Growth Rate is -0.007. This coefficient is unexpectedly negative and shows that as the average annual GDP growth rate increases, average country happiness decreases. This coefficient was expected to be positive because an increasing average annual GDP growth rate reflects increasing production and development, which would increase happiness as people reap more benefits. This value means that as the average annual GDP growth rate of the last five years increases by one percentage point, average country happiness decreases by 0.007 index points. While this sign is unexpected, this coefficient is also not statistically insignificant, meaning the possibility that the coefficient is zero cannot be ruled out.

Extrapolation

The seventh regression, since it has been determined to be the best regression, will be used to predict a couple country's scores on the happiness index for the most recent year that all the data is available, which is 2018. Looking at The United States in 2018, the country had an EFI score of 75.7, the country's Gini coefficient was 41.4, the mortality rate from CVD, cancer, diabetes, and CRD was 13.6%, the suicide mortality rate was 15.7%, the unemployment rate was 3.9%, the urban population was 82.256%, the prevalence of severe food insecurity was 0.8%, the

population density was 35.7301 people per square kilometer, the population ages 15-64 was 65.483% of the population, the life expectancy at birth was 78.639 years, the total fertility rate was 1.730 children per woman, the percent of domestic government health expenditures was 22.502%, the value added from agriculture, forestry, and fishing was 0.921% of GDP, 100% of the population had access to electricity, the greenhouse gas emissions per capita was 0.439 kt of CO2 equivalent, the mean years of schooling was 13.4 years, the refugee population was 313,242 people, the average annula GDP growth rate was 1.843%, and the prevalence of HIV was 0.4%. The regression equation predicts for these values that the country happiness score of the United States would be 5.037. The 95% confidence interval is calculated as follows.

CH' +/- tcrit * SEest =

5.037 +/- 1.9798*(0.2940) =

With 95% confidence the regression predicts that The United States' country happiness score in 2018 will fall between 4.4549 and 5.6191 index points.

In 2015, The United States had an EFI score of 76.2, the country's Gini coefficient was 41.2, the mortality rate from CVD, cancer, diabetes, and CRD was 13.9%, the suicide mortality rate was 14.4%, the unemployment rate was 5.28%, the urban population was 81.671%, the prevalence of severe food insecurity was 1.1%, the population density was 35.063 people per square kilometer, the population ages 15-64 was 66.133% of the population, the life expectancy at birth was 78.690 years, the total fertility rate was 1.844 children per woman, the percent of domestic government health expenditures was 22.333%, the value added from agriculture, forestry, and fishing was 1.044% of GDP, 100% of the population had access to electricity, the greenhouse gas emissions per capita was 0.411 kt of CO2 equivalent, the mean years of schooling was 13.3 years, the refugee population was 273,198 people, the average annula GDP

growth rate was 1.511%, and the prevalence of HIV was 0.5%. The regression equation predicts for these values that the country happiness score of the United States would be 5.028. The 95% confidence interval is calculated as follows.

CH' +/- tcrit * SEest =

5.028 +/- 1.9798*(0.2940) =

With 95% confidence the regression predicts that The United States' country happiness score in 2018 will fall between 4.4456 and 5.6097 index points.

The reported value for The United States' country happiness score in 2018 was 6.892 and in 2015 it was 7.104. Unfortunately, neither of these values fall within the calculated 95% confidence interval for the respective years. Table 10 shows the values for these predictions as well as the percent error for each prediction. The 2018 prediction underestimates the United States' country happiness score by 1.855 points, which is a -36.825% error. The 2015 prediction underestimated the country happiness score by 2.076 points, which is a -41.298% error.

Table 10: Forecasting Country Happiness Score Rates for the years 2018 and 2015

Regression 7:

$$CH = -0.006EFI + 0.031Gini - 0.028MR - 0.021SM - 0.058U - 0.121UP + 0.001UP^2 - 0.03PSF + 0.001Den + 0.077PMid - 0.758LEB + 0.005LEB^2 + 0.33TFR + 0.028\%Hgov - 0.075Agr + 0.001Agr^2 + 0.021Acc + 0.097GGpc - 0.519MYS + 0.029MYS^2 - 1.161E\text{-}6 - 0.007AvGr - 0.002HIV$$

Year	95% Confidence Interval	Observed Country Happiness
2018	4.4549 to 5.6191	6.892
2015	4.4456 to 5.6097	7.104

Year	Predicted CH	Observed CH	Percent Error
2018	5.037	6.892	-36.825%
2015	5.028	7.104	-41.298%

Policy Recommendations

The findings of this paper and the results of the best regression provide a clear path for future policy decisions. In general, any changes or additions to a country's policy should aim to increase the value of those variables that were found to increase happiness, and lessen the occurrence of those variables that were found to decrease happiness, with the exception of those variables that had unexpected signs for their coefficients and thus require further research.

Some variables lend themselves well to formulating policy recommendations. For example, because government health expenditures were found to have a positive relationship with country happiness, it is recommended that countries with low government spending on health services increase their spending, using domestic sources. The positive coefficient for this variable demonstrates that the government using domestic sources for health expenditures should increase happiness as those expenditures increase. Access to electricity is another variable that has a positive relationship with happiness, and those countries that do not have populations with 100% access to electricity, should enforce initiatives and policy to reach 100%. This has become especially important in recent years, as electricity is now being seen as a basic human right (Hughes, 2018). Additionally, countries whose citizens have an average years of schooling below 8.948 years should develop strategies and policies to increase the schooling of their citizens. This is because in a country with a mean years of schooling below 8.948, the average years of schooling they do have is negatively impacting country happiness. Reaching this turning point of 8.948 years could greatly improve a country's happiness by improving the education of a country's people. More specifically, a compulsory education policy where at least 9 years of schooling is required could greatly help countries reach and surpass this turning point so that mean years of schooling can begin to have a positive effect on country happiness.

Finally, several policies could be improved or put in place to improve the physical and mental health of the people in a country with low happiness, and could work together with several variables included in the final model. The mortality rate from suicide or from CRD, cancer, diabetes, and CVD, and the prevalence of HIV have been found to decrease country happiness, so improving the health of a country's people could increase happiness. Hopefully, if a country were to take on the policy recommendation to increase government health expenditures, this would lead to a decrease in the mortality rates and prevalence of these health issues. Together, an improvement in health expenditures and a decrease in deadly illnesses would also increase the life expectancy of a country which would add to the increase in happiness. This is especially true for countries that have a life expectancy below the turning point of 75.8 years, meaning life expectancy is contributing to a lower country happiness. The countries with life expectancies below this turning point should pay special attention to their health services and the physical and mental well-being of their citizens. Overall, there are numerous choices a country could make to increase their average happiness.

Conclusion

The regression analysis conducted in this paper examines the relationship between country happiness and several independent variables that represent important aspects of a country's well-being such as their productivity, equality, health, freedom, environment, levels of despair or stress, education, and population characteristics. The seventh regression was found to be the best regression, despite a few unexpected signs and insignificant coefficients, because of its high adjusted R^2 and the significant and expected relationships between several variables and country happiness. This regression has shown that there is a significant relationship between country happiness and the independent variables included, and such findings are extremely

valuable and promising for the future as an emphasis on the well-being and happiness of a country's citizens continues to grow. Future research or regression estimations could greatly contribute to this work by providing insight into some of the unexpected signs or insignificant coefficients that remain, or by including new variables not yet considered. Valuing the happiness of people and their country, and determining what affects that happiness is an extremely complex and intricate project and process, and there will always be somewhere new to take it next.

References

Access to electricity (% of population). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS

Agriculture, forestry, and fishing, value added (% of GDP). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS

Bjornskov, C. (2003). The Happy Few: Cross-Country Evidence on Social Capital and Life Satisfaction. *Kyklos (Basel), 56*(1), 3-16. https://doi.org/10.1111/1467-6435.00207

Domestic General Government Health Expenditure (% of current health expenditure). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SH.XPD.GHED.CH.ZS

Fertility rate, total (births per woman). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SP.DYN.TFRT.IN

Florida, Mellander, C., & Rentfrow, P. J. (2013). The Happiness of Cities. Regional Studies, 47(4), 613–627. https://doi.org/10.1080/00343404.2011.589830

Hughes, M. (2018, December 10). Why Access to Energy Should Be a Basic Human Right. *Forbes*. Retrieved January 27, 2022, from https://www.forbes.com/sites/mikehughes1/2018/12/10/why-access-to-energy-should-be-a-basic-human-right/?sh=55e59c9245f2.

Forest area (% of land area). Data. (n.d.). Retrieved January 8, 2022, from https://data.worldbank.org/indicator/AG.LND.FRST.ZS

GDP growth (annual %). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG

Gini Index (World Bank estimate). Data. (n.d.). Retrieved January 8, 2022, from https://data.worldbank.org/indicator/SI.POV.GINI

Haines, A., McMichael, A. J., Smith, K. R., Roberts, I., Woodcock, J., Markandya, A., Armstrong, B. G., Campbell-Lendrum, D., Dangour, A. D., Davies, M., Bruce, N., Tonne, C., Barrett, M., & Wilkinson, P. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: Overview and implications for policy makers. *The Lancet*, *374*(9707), 2104–2114. https://doi.org/10.1016/s0140-6736(09)61759-1

Helliwell, J., Layard, R., & Sachs, J. (2015). *World Happiness Report 2015*, New York: Sustainable Development Solutions Network.

Helliwell, J., Layard, R., & Sachs, J. (2016). *World Happiness Report 2016*, New York: Sustainable Development Solutions Network.

Helliwell, J., Layard, R., & Sachs, J. (2017). *World Happiness Report 2017*, New York: Sustainable Development Solutions Network.

Helliwell, J., Layard, R., & Sachs, J. (2018). *World Happiness Report 2018*, New York: Sustainable Development Solutions Network.

Helliwell, J., Layard, R., & Sachs, J. (2019). *World Happiness Report 2019*, New York: Sustainable Development Solutions Network.

Inglehart, Basáñez, M., & Menéndez Moreno, A. (1998). Human values and beliefs a cross-cultural sourcebook : political, religious, sexual, and economic norms in 43

societies ; findings from the 1990-1993 world value survey. University of Michigan Press.

James, K. C., Smith, T. K., Ryan, P. D., Lester, S., Beaumont-Smith, G., Terada, T., Feulner, E. J., Forbes, S., Wallner, J., Kim, A. B., Giugale, M. M., Schneider, F., Abbott , A. F., Gigot, P. A., DeMint, J., Griffith, S. M., Nucifora, A., Loris, N. D., Tomasi, J., Singham, S. (n.d.). *Downloads*. Download Index of Economic Freedom Data, Maps and Book Chapters. Retrieved January 8, 2022, from https://www.heritage.org/index/download

Jensen, N. E. (1967). An introduction to Bernoullian utility theory: I. Utility functions. *The Swedish Journal of Economics*, *69*(3), 163–183. https://doi.org/10.2307/3439089

Labor force, female (% of total labor force). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SL.TLF.TOTL.FE.ZS

Life expectancy at birth, total (years). Data. (n.d.). Retrieved January 8, 2022, from https://data.worldbank.org/indicator/SP.DYN.LE00.IN

Margolis, R., & Myrskylä, M. (2011). A global perspective on happiness and fertility. *PubMed Central*, *37*(1), 29–56. https://doi.org/10.4054/mpidr-wp-2010-025

Mean years of schooling (years). Human Development Reports. (n.d.). Retrieved January 10, 2022, from http://hdr.undp.org/en/indicators/103006

Mendoza, Loyola, R., Aguilar, A., & Escalante, R. (2019). Valuation of Air Quality in Chile: The Life Satisfaction Approach. Social Indicators Research, 145(1), 367–387. https://doi.org/10.1007/s11205-019-02103-1

Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70 (%). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SH.DYN.NCOM.ZS

Mortality rate, infant (per 1,000 live births). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SP.DYN.IMRT.IN

Population ages 0-14 (% of total population). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SP.POP.0014.TO.ZS

Population ages 15-64 (% of total population). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SP.POP.1564.TO.ZS

Population ages 65 and above (% of total population). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS

Population density (people per sq. km of land area). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/EN.POP.DNST

Prevalence of HIV, total (% of population ages 15-49). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS

Prevalence of severe food insecurity in the population (%). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SN.ITK.SVFI.ZS

Refugee population by country or territory of asylum. Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SM.POP.REFG

Russell, R., Guerry, A. D., Balvanera, P., Gould, R. K., Basurto, X., Chan, K. M. A., Klain, S., Levine, J., & Tam, J. (2013). Humans and nature: How knowing and experiencing nature affect well-being. *Annual Review of Environment and Resources*, *38*(1), 473–502. https://doi.org/10.1146/annurev-environ-012312-110838

Smith, C., & Helliwell, J. (2022, January 15). personal.

Soper, D. (n.d.). *Calculator: Critical F-value*. Free Critical F-value Calculator - Free Statistics Calculators. Retrieved January 19, 2022, from https://www.danielsoper.com/statcalc/calculator.aspx?id=4

Stanton, E. A. (n.d.). *The Human Development Index: A History*. ms, University of Massachusetts Amherst . Retrieved from https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1101&context=peri_workingpapers#:~:text=The%20Human%20Development%20Index%2C%20or,capita)%20(Sen%201985).

Suicide mortality rate (per 100,000 population). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SH.STA.SUIC.P5

Szczepanek, A. (2021, September 29). *Critical value calculator*. Omni Calculator. Retrieved January 19, 2022, from https://www.omnicalculator.com/statistics/critical-value

Total greenhouse gas emissions (kt of CO2 equivalent). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE

Unemployment, total (% of total labor force) (modeled ILO estimate). Data. (n.d.). Retrieved January 8, 2022, from https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS

Urban population (% of total population). Data. (n.d.). Retrieved January 10, 2022, from https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS

2020 Corruptions Perceptions Index - explore the results. Transparency.org. (n.d.). Retrieved January 8, 2022, from https://www.transparency.org/en/cpi/2020

Ye, D., Ng, Y.-K., & Lian, Y. (2014). Culture and happiness. *Social Indicators Research*, *123*(2), 519–547. https://doi.org/10.1007/s11205-014-0747-y

Yilmaz, I., & Tag, M. N. (2016). Happiness and Economic Freedom: Are They related? *SHS Web of Conferences*, *28*. https://doi.org/10.1051/shsconf/20162801109