

What Makes for a Happy Population: an Economic, Social, Citizen Opinions, and Prosperity Investigation

Calvin Gao

Quantitative Analysis Center, Wesleyan University

CADS Capstone: Professor Maryam Gooyabadi

May 18, 2023

Abstract

This study explores the economic, social, citizen opinions, and prosperity that influence each country's happiness level. Data is collected from the World Happiness Report, World Bank, and the Legatum Prosperity Index for data analysis. Past studies that have linked economic factors to happiness have done so for specific countries. In this study, happiness is examined on a macro level using the World Happiness Report (WHR) in conjunction with the World Bank (WB) data and the Legatum Prosperity Index. Dimension reduction techniques, specifically Principal Component Analysis and Factor Analysis are used for selecting variables for the fixed effects, cross-sectional regression models, and obtaining factors for clustering which assists in identifying similar countries. Regression results indicate that economic, citizen opinion, and prosperity categories are significantly contributing to a country's happiness while clustering illustrates a pattern of similar countries across the subsets.

What Makes for a Happy Population: an Economic, Social, Citizen Opinions, and Prosperity Investigation

From the ancient Greeks to Eleanor Roosevelt, happiness has been experienced in various ways from reaching your potential to being honest (Bradstreet, 2020). Throughout history, happiness has been defined in countless ways and multiple perspectives. Past literature has used happiness interchangeably with subjective well-being, well-being, welfare, satisfaction, and in economics often the term utility (Easterlin, 2001). Past studies have also found that there is enough evidence supporting the idea that individuals' self-reported happiness can provide numerous valuable insights into individual happiness (Ferrer-i-Carbonell, 2013). This self-reported happiness is valuable as a source of data for researchers to examine potential factors relating to happiness. Past research often explores economic variables and their relationship to happiness within specific countries. Economic variables are indicators of a country's condition and an ideal relationship would involve a positive association such that an improvement in a country's economy would lead to increased happiness. However, happiness is a complicated concept, and relationships between indicators and happiness are often complex as well. In a study by Oswald (1997), results found that in developed countries, an economic improvement only leads to a small portion of extra happiness. In another study, Easterlin (2001) explored the relationship between subjective well-being with income and material aspirations. Results indicate that while subjective well-being has a direct relationship with income, an increase in income also leads to an increase in material aspirations. In other words, as an individual experiences an increase in income, the amount of material aspirations increase relatively similarly and cancels out their relation to happiness. Hence, while an increase in

income can lead to happier groups at the beginning stages of their life, income growth does not lead to increases in well-being at later stages.

Meanwhile, a study examining data in Japan found that certain groups that had reform programs with improved conditions led to increases in well-being (Sarracino et al., 2022). In Europe, the unemployment rate had a negative effect on happiness and results found that spending resources on unemployment programs and active labor policies had a positive relationship with well-being (Bonasia et al., 2022). These programs target specific populations to provide them with support that leads to improved well-being. While the literature provided unique information on economic conditions and happiness, past research often focuses on economic factors and even more so on specific countries.

This study will seek to explore relations of happiness with not only economic variables but also social, citizen opinion, and prosperity. In addition, this study will also seek to explore these relationships through worldwide countries. Finally, this study will attempt to identify similar countries through the data available.

Method

Data Collection

The data obtained for this study originates from the World Happiness Report, World Bank, and the Legatum Prosperity Index. The World Bank included economic and social data for countries over the years. This study initially selected a total of 72 economic and social variables from the World Bank from the years 2000 to 2020. However, a closer examination of the World Bank variables found that the majority of the variables had large amounts of missing data for the countries being explored. As a result, variables with more than 10% missing data were removed from the dataset.

World Happiness Report Variables

The World Happiness Report (Helliwell et al. 2023) obtained data through the Gallup World Poll from countries around the world through surveys provided to participants from a sample of the country. While the number of participants ranges differently each year, the poll often receives more than 100,000 participants from 130 countries or approximately 1,000 participants per country. The data received from the responses to the survey were then manipulated by the Gallup World Poll through Gallup weights to obtain a value representative of each country. The World Happiness Report provided a dataset containing 10 variables in addition to the country name and year.

Life Ladder

Life ladder, also referred to as Cantril life ladder, is the happiness or subjective well-being variable used within this study. The variable consists of the national average of the responses to a question asking participants to rank which step of a ladder do they personally feel at the time of the questionnaire such that zero is the worst possible life and ten is the best possible life.

Log GDP per capita

Log GDP per capita is the purchasing power parity (PPP) set at constant 2017 international dollar prices.

Social Support

Social support is obtained by the WHR as the national average of being able to count on someone in times of trouble. The variable is obtained through the 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?” with responses being a binary response of zero or one.

Healthy Life Expectancy at Birth

Healthy life expectancy at birth is obtained by the WHR from the World Health Organization's (WHO) Global Health Observatory data repository. The variable is defined as the average number of years that a person can expect to live in "full health" when accounting the years lived in less than full health due to disease or injury (World Health Organization, 2023).

Freedom to make Life Choices

Freedom to make life choices is obtained by the WHR as the national average of responses to the question from the Gallup World Poll "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"

Generosity

Generosity is obtained from the residual when regressing the national average response to the question "Have you donated money to a charity in the past month?" on GDP per capita and obtained by the WHR through the Gallup World Poll.

Perceptions of Corruption

Perceptions of corruption are obtained by the WHR through the average response of two questions from the Gallup World Poll. The first question asks whether corruption is widespread throughout the government or not and the second question asks about whether corruption is widespread in businesses with responses for each question being a binary zero or one. In the situation such that data for government were not available, the WHR utilized the perception of business corruption as the overall value for the variable.

Positive Affect

Positive affect is obtained through the average of three measures of positive affect from the Gallup World Poll asking whether the participant smiled or laughed a lot yesterday,

experienced certain feelings for the majority of the day yesterday such as enjoyment or happiness depending on the wave of the study, and whether they learned something interesting or did an interesting task yesterday.

Negative Affect

Negative affect is obtained through the average of three measures of negative affect obtained through the Gallup World Poll. The questions asked whether they experienced certain feelings for the majority of the day yesterday such as worry, sadness, and anger.

Confidence in National Government

Confidence in national government is obtained by the WHR through one of the measures of the first principal component that is used to generate the binary variable. The WHR utilized the 75th percentile in the global distribution such that countries with populations that tend to have low levels of trust in institutions in the global distribution will have lower averages of trust in the institution at a national level.

Legatum Prosperity Index Variables

The Legatum Institute's 2023 Prosperity Index provided data on 12 pillars (see Table 1) that generated a prosperity index. The Legatum Institute defines prosperity as when all individuals have the freedom and opportunity to thrive in the environment. Hence, the 12 pillars that represent prosperity include safety and security, personal freedom, investment environment, enterprise conditions, living conditions, health, governance, social capital, infrastructure and market access, economic quality, education, and natural environment. The Legatum Institute generated the pillars through 300 unique indicators from more than 70 different data sources. The scores for the values of the pillars for each country are generated through five to eight elements such that each element includes one to eight indicators.

Table 1: Prosperity Pillars

Descriptions of Prosperity Pillars

Variable	Description
Safety and Security	Measures the extent of how crime, conflict, and war have affected the safety and security of individuals in the short term and long term.
Personal Freedom	Examines the progress towards basic legal rights, social tolerance, and individual liberties.
Social Capital	Examines the strength of social relationships, personal relations, social norms, civic participation in a country and trust in institutions.
Governance	Indicates the level of checks and balances on power such that governments are able to function without corruption.
Investment Environment	Examines the extent of the protection and accessibility of investments.
Enterprise Conditions	Examines regulations that support the growth of businesses.
Infrastructure and Market	Examines the operations for the trade of goods and services in the market.
Economic Quality	Examines the degree of an economy's readiness to generate wealth in the situation of the complete participation of the workforce.
Living Conditions	Examines life quality to a reasonable measure that includes material resources, basic services, shelter and connectivity.

Health	Examines the health of the people and accessibility to the necessary services or programs to support maintaining good health.
Education	Explores the enrollment, outcomes, and quality throughout pre-primary, primary, secondary, and tertiary education. The pillar also measures the adult population's skills.
Natural Environment	Examines the effect of the physical environment on the daily lives of individuals and the impact on future generations.

Note: Descriptions of pillars are obtained from the Legatum Prosperity Index

Data Merging Process

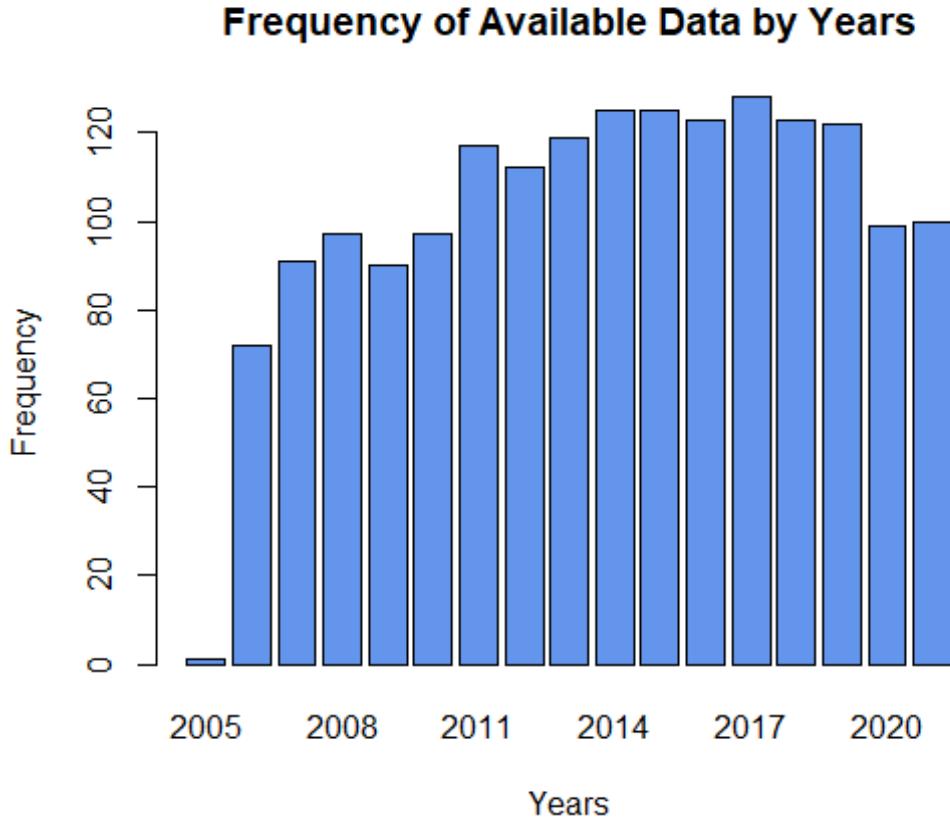
The datasets from the World Happiness Report, World Bank and Legatum Prosperity Index were examined for the process of merging. Several countries within each dataset included country names that were different from one another and were renamed such that each country variable in each of the three datasets included countries of the same name. Since the happiness variable resided within the World Happiness Report dataset, the World Bank and Legatum Prosperity Index were merged such that the observations in the World Happiness Report were all kept. In the case that there were observations in the World Bank or the Legatum Prosperity Index dataset that did not match the country and year in the World Happiness Report, these variables were removed.

World Happiness Report and World Bank

The first merged dataset involved the World Happiness Report with the World Bank. The resulting dataset consisted of 84 variables including country and year with a total of 1741 observations of countries per year. An initial examination of this merged dataset as seen in

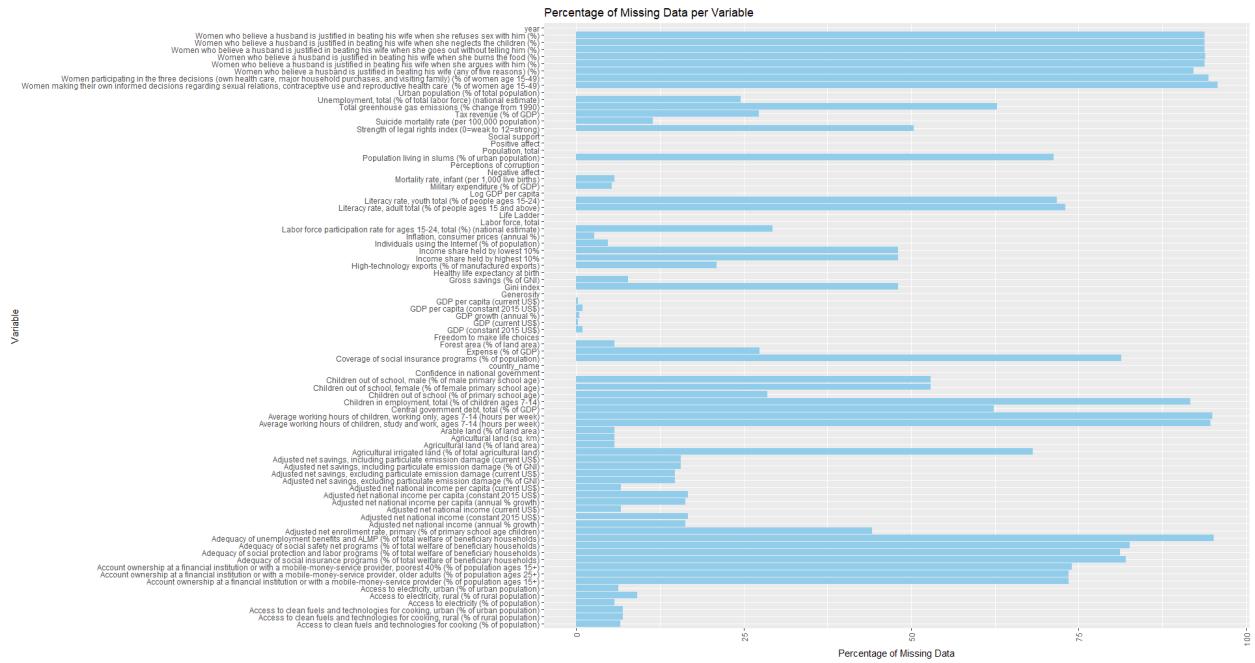
Figure 1 found that there was only one data observation available for 2005 compared to the other countries. Hence, the data for the year 2005 were removed.

Figure 1: Bar Chart of the Available Data (2005-2020)



In another examination of the dataset in Figure 2, the figure illustrates that there were a large number of variables with more than 10% missing data. Hence, variables with more than 10% missing data were also removed from the dataset. After removing the variables with a large portion of missing data, there still remains several observations with no data for certain variables. These observations were also removed to obtain a dataset with no missing values. The merged dataset with the World Happiness Report and World Bank consists of 38 variables with 1331 observations from 2006 to 2020.

Figure 2: Bar Chart of Percentage of Missing Data per Variable



To explore the different perspectives of the merged dataset, these 38 variables were categorized and split into economic, citizen's opinion, and social subsets with country, year, and happiness in each subset.

The economic subset had the variables: log GDP per capita, adjusted net national income in current US dollars, adjusted net national income per capita in current US dollars, agricultural land (% of land area), agricultural land (sq. km), GDP (constant 2015 US\$), GDP in current US dollars, GDP per capita (constant 2015 US\$), GDP growth (annual %), GDP per capita in current US dollars, gross savings (% of GNI), inflation or consumer prices (annual %), total labor force, total population, mortality rate for infant per 1,000 live births, and military expenditure (% of GDP).

The social subset include the variables: access to clean fuels and technologies for cooking (% of population), access to clean fuels and technologies for cooking in urban (% of urban

population), access to clean fuels and technologies for cooking in rural (% of rural population), access to electricity in rural (% of rural population), access to electricity (% of population), access to electricity in urban (% of urban population), urban population (% of total population), forest area (% of land area), arable land (% of land area), and individuals using the internet (% of population).

The citizen's opinion subset included the variables from the World Happiness Report.

World Happiness Report, World Bank, and Legatum Prosperity Index

To create the prosperity subset for analysis, the Legatum Prosperity Index was merged with the dataset with World Happiness Report and World Bank variables. The final merged dataset consists of 50 variables. After removing observations with no data, specifically the years of 2006 as the Legatum Prosperity Index's data began from 2007, the dataset contains 1275 observations from the years of 2007 to 2020.

The prosperity subset was then generated by selecting 15 variables including the 12 pillars from the Legatum Prosperity Index in addition to the country, year and happiness variables.

Parameter Selection and Regression Models

To explore the final dataset and the examination on potential relationships to happiness, an exploratory analysis was conducted. To control for variables with large numerical values relative to other variables in the dataset, all variables were scaled to obtain a Z-score with the exception of country and year. Principal Component Analysis and Factor Analysis was utilized to reduce the dimensions of the variables. Fixed effect models, as seen in (1), were then used to test the variables in the economic, citizen opinions and social subset on happiness from 2006 to 2020 holding the country and year fixed. Fixed effect regression models, as seen in (1), were also

conducted on the prosperity subset as well as the dataset with all indicators from 2007 to 2020 holding the country and year fixed. The fixed effects model removes the variation between the countries and years and controls for these variables by finding the mean differences between countries and removing them.

$$\gamma_{CY} = \beta_1 \chi_{CY} + \dots + \beta_k \chi_{CY} + \alpha_C + \epsilon_{CY} \quad (1)$$

Equation (1) shows the γ_{CY} representing the response variable happiness holding country and year fixed, β_1 to β_k representing the coefficients for the first variable to the k-th or total number of variables in the subset or entire dataset being used for the regression, α_C being the country specific intercepts where C represents the countries, the subscript Y represents the year, and ϵ_{CY} represents the error term.

The datasets were then further split by year in which a cross sectional analysis was conducted through linear regression models, as seen in (2), on the economic, citizen opinions, and prosperity subsets as well as the entire dataset per year from 2015 to 2020.

$$\gamma = \beta_0 + \beta_1 \chi_i + \dots + \beta_k \chi_k + \epsilon \quad (2)$$

Equation (2) includes γ representing the response variable happiness, β_0 as the intercept when all χ are set to be zero, β_1 to β_k represents the coefficients of variables χ_i to χ_k that measures the changes in γ from changes in one units for χ_i and holding the other variables in the model constant.

Clustering and Data Visualization

Clustering was conducted with Factor Analysis to obtain factors representing the variables. Partitioning around medoids (PAM) followed after factor analysis using the k-medoid method to cluster the countries based on the factors obtained from factor analysis.

Principal component analysis (PCA) was conducted on the economic, citizen opinions, and prosperity subsets to obtain clusters. These clusters were then used to determine the countries that are similar to one another.

Pre-Post Covid Analysis

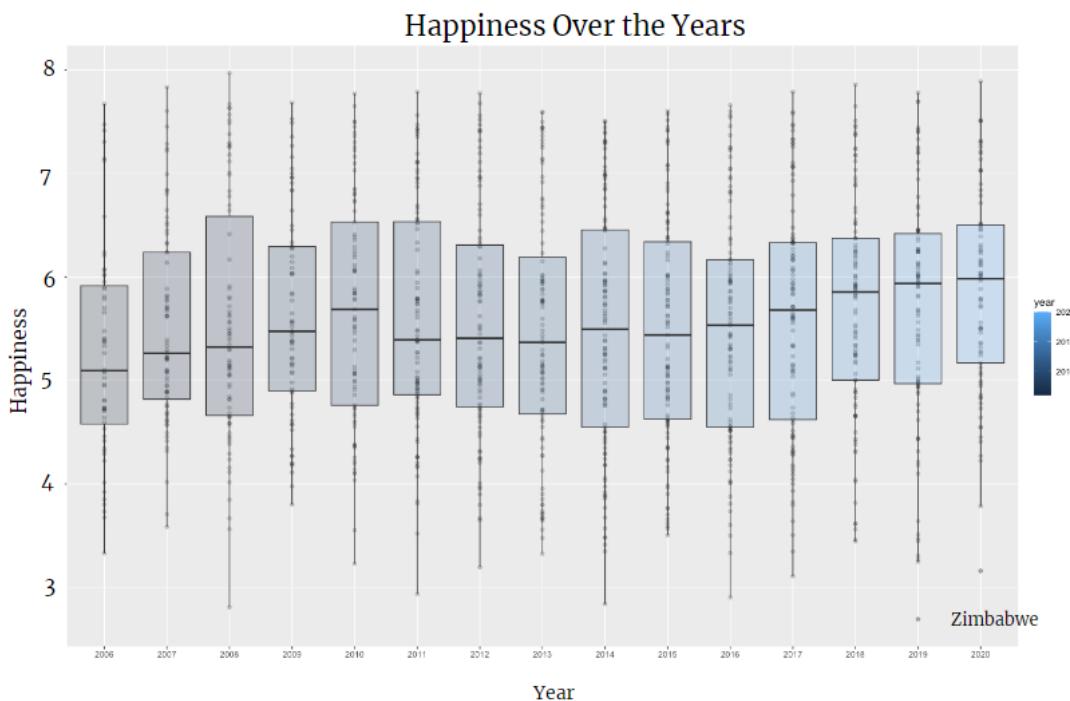
An attempt to identify patterns between countries' happiness during pre and post covid time periods was made through time series slope graphs. A predictive modeling using a machine learning model, specifically Random Forest, was also created to determine how well the variables used to train the model predicted happiness in 2019 and 2020.

Results

Data Visualizations

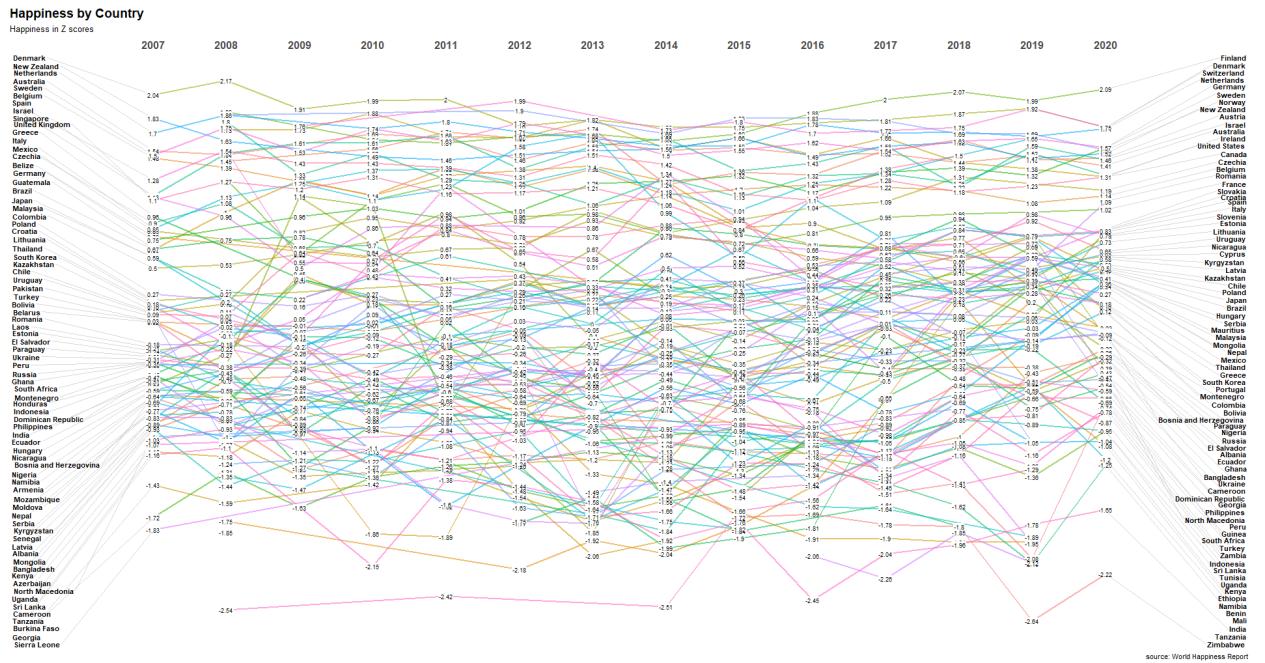
Exploratory data analysis began with a plot (see Figure 3) observing the changes in the happiness over the years.

Figure 3: Scatterplot with Box Plots on Happiness by Year



Happiness ranged at a median score of 5 and over the years appears to gradually increase towards a value of 6. It was noticed that in 2019 and 2020, a specific country Zimbabwe became an outlier with extremely low values of happiness. To examine how each country varied in happiness, a slope graph as seen in Figure 4 was created.

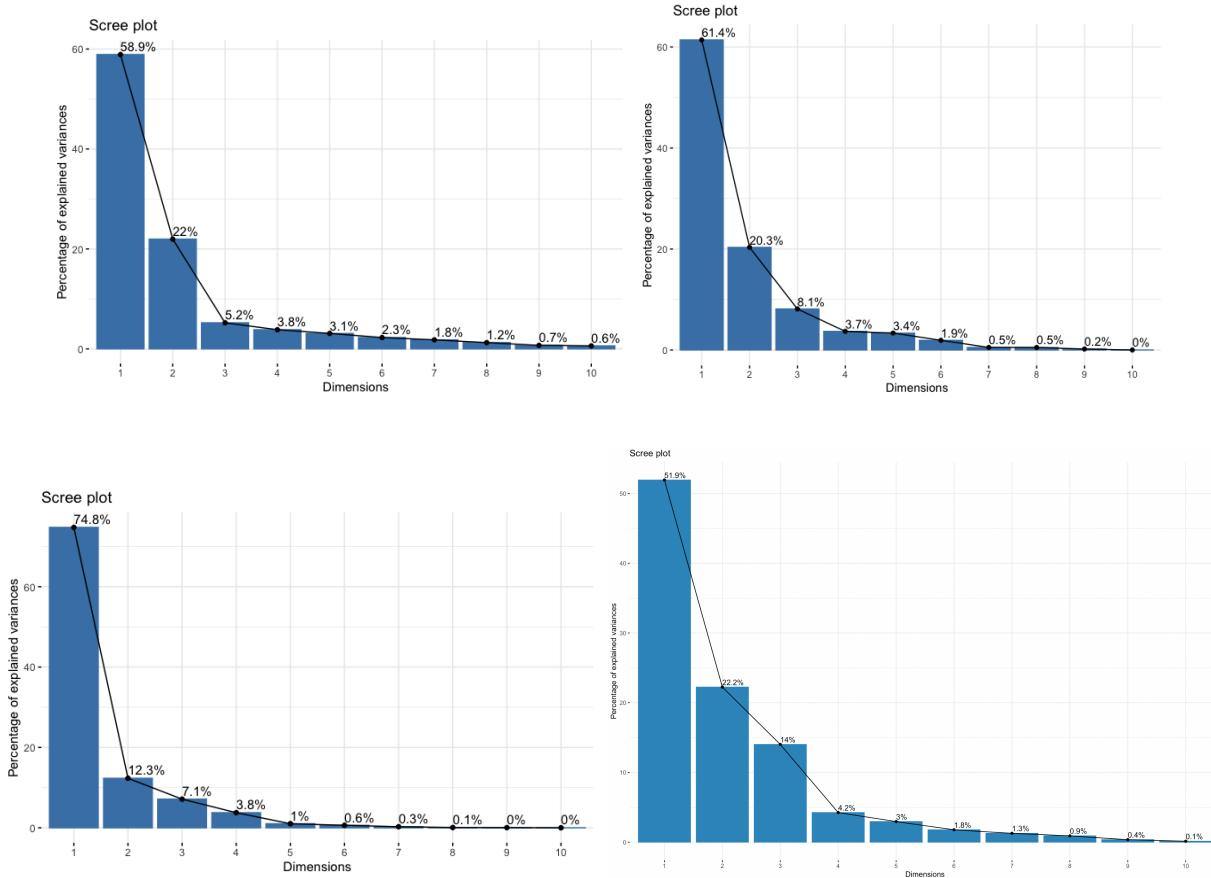
Figure 4: Slope Graph of Country Happiness per Year



Principal Components Analysis

To reduce the dimensions in the subsets, principal components analysis (PCA) was conducted for selecting the parameters to include in the following statistical models. Scree plots, as seen in Figure 5, were used to determine the number of principal components that should be selected based on the information that each principal component is able to provide to the model. The scree plots illustrate that three principal components for the economic, citizen opinion, and prosperity subset would provide a good amount of information while four principal components were needed for the social subset. Any additional principal components would provide fewer information.

Figure 5: Scree Plots for the Subsets

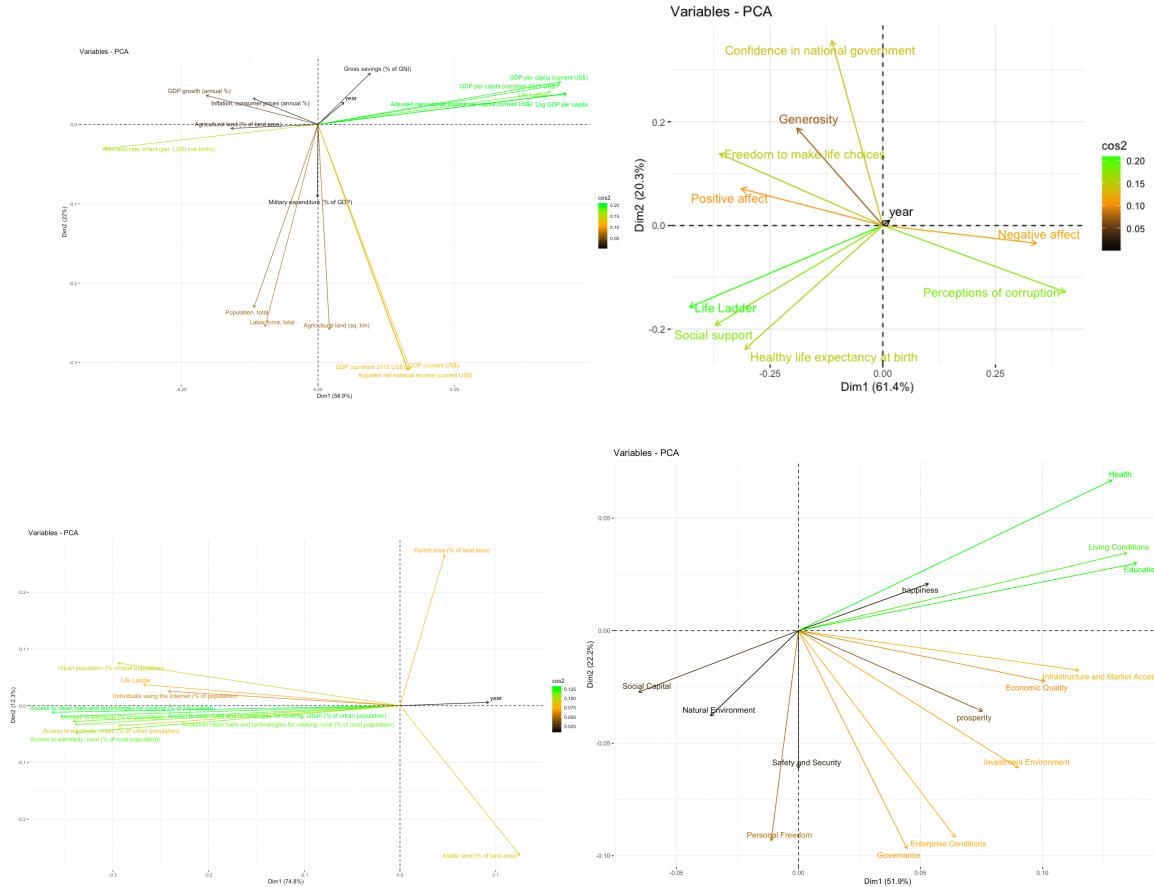


Note: Top left is the plot for economics subset, top right is for citizens opinion, bottom left is social and bottom right is for prosperity.

Finally, biplots (see Figure 6) combined with the square cosine or cos2 of each variable were generated to observe similar variables and their cos2 scores. Variables with higher cos2 scores represent that the variables contribute higher amounts to the principal components selected. Within the biplot, variables that are close and grouped together have a positive correlation with one another, variables that are further away from the origin represent the variable are better represented, and variables that face opposite directions of the biplot are negatively correlated with one another. To avoid including many variables that are similar to one

another, specific variables are selected from the economic and social subset to better represent the subsets. Variables in the citizen opinion and prosperity subset were all kept.

Figure 6: Biplots with Cos2 Scores



Note: Top left is the plot for economics subset, top right is for citizens opinion, bottom left is social and bottom right is for prosperity.

From the PCA model of the economic subset, the variables log GDP per capita, agricultural land (sq. km), GDP growth (annual %), gross savings (% of GNI), inflation or consumer prices (annual %), total labor force, and mortality rate for infants per 1,000 live births were selected to be the new economic subset.

From examining the PCA model of the social subset, the new social subset includes the variables: access to clean fuels and technologies for cooking (% of population), access to

electricity (% of population), forest area (% of land area), arable land (% of land area), and individuals using the internet (% of population).

Fixed Effects Regression Models

Fixed effects regression were conducted on the new subsets selected from PCA. The economic, citizen opinion and social subset were tested based on the years from 2006 to 2020, while prosperity was tested on data from 2007 to 2020 due to the lack of data. Results of the fixed effects regression model can be seen in Table 2. The economic subset provided significant results with log GDP per capita, GDP growth in annual %, inflation or consumer prices in annual %, and total labor force.

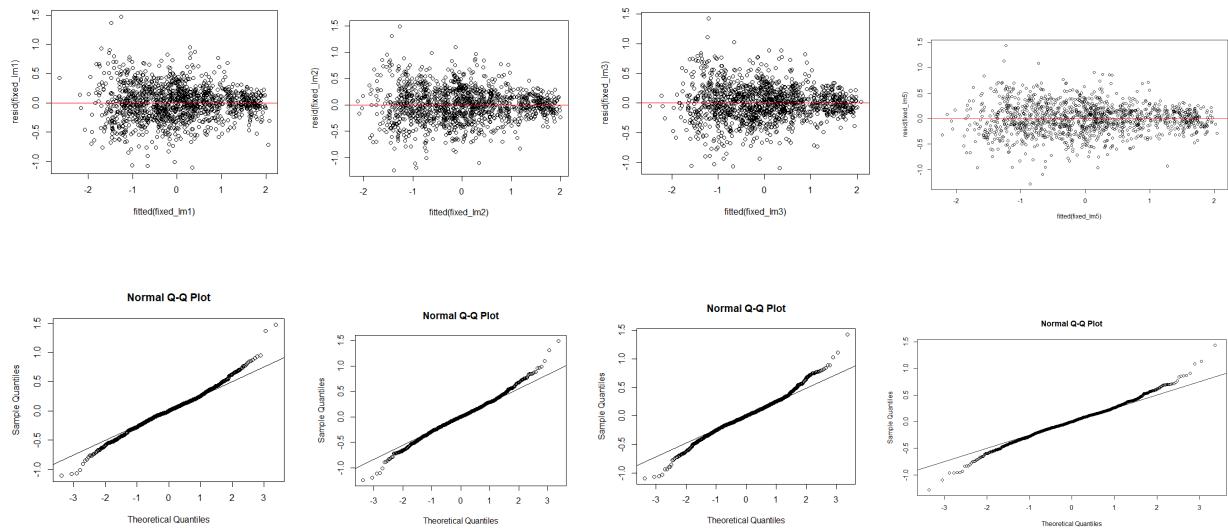
Table 2: Fixed Effects Regression Models

Economic on Happiness				Subjective on Happiness			
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Log GDP per capita	1.65	1.35 – 1.96	<0.001	Social support	0.20	0.15 – 0.25	<0.001
Agricultural land (sq km)	0.04	-0.58 – 0.65	0.902	Healthy life expectancy at birth	-0.26	-0.42 – -0.11	0.001
GDP growth (annual %)	0.03	0.01 – 0.06	0.012	Freedom to make life choices	0.05	0.01 – 0.09	0.027
Gross savings (% of GNI)	0.01	-0.03 – 0.05	0.574	Generosity	0.04	0.00 – 0.09	0.045
Inflation consumer prices (annual %)	-0.04	-0.06 – -0.02	<0.001	Perceptions of corruption	-0.11	-0.18 – -0.04	0.002
Mortality rate, infant (per 1,000 live births)	-0.09	-0.24 – 0.05	0.215	Positive affect	0.16	0.11 – 0.21	<0.001
Labor force, total	-2.35	-2.90 – -1.80	<0.001	Negative affect	-0.03	-0.07 – 0.00	0.059
N _{country_name}	120			Confidence in national government	0.05	0.01 – 0.10	0.022
N _{year}	15			N _{country_name}	120		
Observations	1331			N _{year}	15		
AIC	784.263			Observations	1331		
				AIC	744.222		

Social on Happiness				Prosperity on Happiness			
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Access to clean fuels and technologies for cooking (% of population)	0.19	-0.01 – 0.39	0.069	Safety and Security	0.00	-0.07 – 0.08	0.930
Access to electricity (% of population)	-0.12	-0.23 – 0.00	0.053	Personal Freedom	-0.06	-0.23 – 0.10	0.443
Forest area (% of land area)	-0.03	-0.45 – 0.40	0.899	Governance	0.39	0.15 – 0.64	0.001
Arable land (% of land area)	0.14	-0.12 – 0.41	0.289	Social Capital	-0.00	-0.08 – 0.07	0.962
Individuals using the Internet (% of population)	0.07	-0.01 – 0.16	0.089	Investment Environment	0.11	-0.10 – 0.31	0.322
N _{country_name}	120			Enterprise Conditions	0.08	-0.04 – 0.19	0.201
N _{year}	15			Infrastructure and Market Access	-0.11	-0.35 – 0.14	0.379
Observations	1331			Economic Quality	0.39	0.28 – 0.50	<0.001
AIC	982.411			Living Conditions	0.73	0.47 – 0.99	<0.001

All indicators in the citizens opinion subset were significantly related with happiness with the exception of negative affect. There were no significant results in the fixed effects regression model for the social subset. Within the prosperity subset, the variables that had significant results were governance, economic quality, living conditions, and health. Assumptions (see Figure 7) of the regression models found that there is a linear relationship and no problems with homogeneity of variance between the explanatory variables and happiness as seen in a residuals vs fitted plot. The normal QQ plot illustrated no large issues with the normality of residuals.

Figure 7: Assumptions of the Regression Models



Note: The four columns correspond to the subsets in the order of economic, social, citizen's opinion and prosperity. The first row are the residuals vs fitted plots.

A final fixed effects regression model (see Table 3) included indicators from all four subsets. Results indicated that social support, healthy life expectancy at birth, generosity, perceptions of corruption, positive and negative affect, log GDP per capita, GDP growth, mortality rate, labor force, governance, economic quality, living conditions and health had significant results.

Table 3: Fixed Effects Regression with All Indicators on Happiness

Predictors	Prosperity, Economic, Citizen's Opinion, Subjective on Happiness		
	Estimates	CI	p
Safety and Security	-0.07	-0.14 – 0.00	0.063
Personal Freedom	-0.03	-0.19 – 0.12	0.678
Governance	0.35	0.12 – 0.58	0.003
Social Capital	-0.03	-0.10 – 0.04	0.451
Investment Environment	0.04	-0.16 – 0.24	0.694
Enterprise Conditions	0.02	-0.09 – 0.13	0.754
Infrastructure and Market Access	-0.23	-0.49 – 0.03	0.088
Economic Quality	0.25	0.13 – 0.36	<0.001
Living Conditions	0.46	0.17 – 0.74	0.002
Health	-0.23	-0.43 – -0.04	0.020
Education	0.27	-0.07 – 0.60	0.116
Natural Environment	-0.08	-0.25 – 0.08	0.307
Social support	0.15	0.10 – 0.20	<0.001
Healthy life expectancy at birth	-0.37	-0.59 – -0.14	0.002
Freedom to make life choices	0.02	-0.02 – 0.07	0.247
Generosity	0.07	0.03 – 0.11	0.002
Perceptions of corruption	-0.08	-0.15 – -0.01	0.027
Positive affect	0.11	0.06 – 0.15	<0.001
Negative affect	-0.04	-0.08 – -0.00	0.049
Confidence in national government	0.04	-0.01 – 0.08	0.113
Log GDP per capita	0.76	0.36 – 1.16	<0.001
Agricultural land (sq km)	-0.06	-0.64 – 0.51	0.827
GDP growth (annual %)	0.03	0.01 – 0.06	0.012
Gross savings (% of GNI)	-0.00	-0.04 – 0.04	0.908
Inflation, consumer prices (annual %)	-0.02	-0.04 – 0.00	0.067
Mortality rate, infant (per 1,000 live births)	-0.35	-0.59 – -0.12	0.004
Labor force, total	-2.43	-3.05 – -1.81	<0.001
Access to clean fuels and technologies for cooking (% of population)	-0.04	-0.28 – 0.21	0.763
Access to electricity (% of population)	-0.11	-0.24 – -0.03	0.131
Forest area (% of land area)	0.14	-0.27 – 0.55	0.497
Arable land (% of land area)	0.02	-0.23 – 0.27	0.878
Individuals using the Internet (% of population)	0.04	-0.06 – 0.14	0.417
N country	119		
N year	14		
Observations	1275		
AIC	534.853		

Cross Sectional Analysis 2015 to 2020

To further understand the indicators at a closer level, a cross sectional analysis of the subsets were conducted through linear regression models across 2015 to 2020.

Table 4: Regression Models on Economic Indicators (2015-2020)

Economic on Happiness																		
Predictors	2015			2016			2017			2018			2019			2020		
	Estimates	CI	p															
(Intercept)	-0.10	-0.22 – 0.02	0.107	-0.08	-0.21 – 0.05	0.250	-0.08	-0.22 – 0.06	0.257	0.04	-0.10 – 0.18	0.569	0.00	-0.13 – 0.13	0.986	0.20	-0.04 – 0.43	0.104
Log GDP per capita	0.69	0.45 – 0.93	<0.001	0.78	0.55 – 1.00	<0.001	0.83	0.58 – 1.07	<0.001	0.77	0.53 – 1.01	<0.001	0.67	0.42 – 0.93	<0.001	0.79	0.56 – 1.02	<0.001
Agricultural land (sq km)	0.15	-0.00 – 0.30	0.056	0.07	-0.08 – 0.22	0.379	0.08	-0.07 – 0.23	0.284	0.05	-0.10 – 0.19	0.521	0.09	-0.06 – 0.23	0.248	0.04	-0.08 – 0.16	0.515
GDP growth (annual %)	-0.04	-0.18 – 0.10	0.615	-0.07	-0.25 – 0.12	0.461	0.13	-0.12 – 0.37	0.308	-0.13	-0.41 – 0.15	0.358	-0.04	-0.30 – 0.22	0.768	0.04	-0.10 – 0.17	0.576
Gross savings (% of GNI)	0.07	-0.07 – 0.21	0.326	0.03	-0.12 – 0.18	0.686	-0.05	-0.21 – 0.10	0.512	-0.07	-0.21 – 0.08	0.364	-0.04	-0.20 – 0.13	0.663	0.08	-0.07 – 0.22	0.286
Inflation, consumer prices (annual %)	-0.37	-0.78 – 0.04	0.074	-0.18	-0.72 – 0.35	0.493	-0.59	-1.29 – 0.11	0.099	-0.36	-1.02 – 0.29	0.273	-0.14	-0.24 – 0.04	0.006	-0.04	-0.07 – 0.01	0.011
Mortality rate, infant (per 1,000 live births)	-0.01	-0.23 – 0.21	0.931	0.05	-0.17 – 0.28	0.632	0.07	-0.19 – 0.33	0.583	0.07	-0.20 – 0.34	0.588	-0.12	-0.39 – 0.15	0.376	0.06	-0.21 – 0.34	0.635
Labor force, total	-0.08	-0.22 – 0.07	0.294	-0.05	-0.20 – 0.09	0.486	-0.10	-0.24 – 0.05	0.197	-0.10	-0.24 – 0.04	0.149	-0.17	-0.31 – 0.03	0.017	-0.13	-0.25 – 0.01	0.041
Observations	103			103			102			98			94			81		
R ² / R ² adjusted	0.669 / 0.645			0.655 / 0.629			0.649 / 0.623			0.638 / 0.610			0.674 / 0.648			0.717 / 0.690		
AIC	195.414			201.744			204.980			189.108			186.870			125.578		

Results of the regressions, as seen in Table 4, on the economic indicators from 2015 to 2020 found that only log GDP per capita remained significant throughout the years. However, inflation and labor force obtained significant results in 2019 and 2020. A potential explanation for this pattern would be the pandemic. As countries prepared for the pandemic, individuals within the countries were afraid of higher prices of basic necessities and employees were fearful of losing employment. This result indicates that factors that may have a relationship with happiness may differ in strength based on the lifestyle of the time. When individuals or countries encounter changes, certain factors may become of more importance and relate to happiness.

The next analysis involves the citizen opinions over the years from 2015 to 2020. Table 5 provides results indicating that countries find indicators such as social support, healthy life expectancy at birth, perceptions of corruption, positive affect, and confidence in national government as important factors. These indicators suggest that health, social connections and trust in the government are important to residents of a country. While several of the variables such as freedom to make life choices and negative affect were significant for only a few years,

this further supports the idea that certain factors may vary in their importance in understanding what makes a country happy from year to year.

Table 5: Regression Models on Citizen Opinions (2015-2020)

Citizen Opinion on Happiness																		
Predictors	2015			2016			2017			2018			2019			2020		
	Estimates	CI	p															
(Intercept)	-0.07	-0.17 – 0.02	0.116	-0.13	-0.23 – -0.03	0.011	-0.05	-0.16 – 0.07	0.419	-0.02	-0.14 – 0.11	0.814	-0.12	-0.26 – 0.02	0.096	0.03	-0.10 – 0.16	0.613
Social support	0.32	0.19 – 0.46	<0.001	0.30	0.15 – 0.46	<0.001	0.33	0.17 – 0.50	<0.001	0.41	0.22 – 0.60	<0.001	0.48	0.29 – 0.66	<0.001	0.32	0.16 – 0.48	<0.001
Healthy life expectancy at birth	0.35	0.21 – 0.48	<0.001	0.40	0.26 – 0.55	<0.001	0.32	0.15 – 0.48	<0.001	0.33	0.15 – 0.52	0.001	0.39	0.22 – 0.55	<0.001	0.25	0.07 – 0.44	0.008
Freedom to make life choices	0.12	-0.03 – 0.27	0.126	0.16	-0.02 – 0.35	0.085	0.12	-0.07 – 0.30	0.208	0.09	-0.12 – 0.30	0.406	0.21	-0.03 – 0.46	0.082	0.27	0.07 – 0.47	0.010
Generosity	-0.02	-0.13 – 0.08	0.658	0.08	-0.03 – 0.20	0.154	0.02	-0.10 – 0.13	0.754	0.03	-0.09 – 0.15	0.637	0.03	-0.09 – 0.15	0.636	0.02	-0.10 – 0.14	0.733
Perceptions of corruption	-0.32	-0.45 – -0.20	<0.001	-0.20	-0.33 – -0.06	0.004	-0.29	-0.43 – -0.15	<0.001	-0.27	-0.41 – -0.13	<0.001	-0.24	-0.39 – -0.10	0.001	-0.31	-0.45 – -0.16	<0.001
Positive affect	0.12	0.00 – 0.25	0.047	0.15	0.01 – 0.30	0.039	0.19	0.05 – 0.32	0.008	0.14	0.00 – 0.28	0.047	0.12	-0.03 – 0.26	0.107	-0.05	-0.19 – 0.09	0.508
Negative affect	0.09	-0.02 – 0.20	0.114	0.04	-0.08 – 0.16	0.483	0.01	-0.12 – 0.13	0.929	0.11	-0.02 – 0.23	0.094	0.16	0.01 – 0.31	0.035	-0.11	-0.24 – 0.02	0.102
Confidence in national government	-0.13	-0.26 – -0.01	0.060	-0.24	-0.38 – -0.10	0.001	-0.25	-0.40 – -0.10	0.002	-0.18	-0.34 – -0.03	0.021	-0.18	-0.35 – -0.00	0.045	-0.24	-0.39 – -0.09	0.002
Observations	103			103			102			98			94			81		
R ² / R ² adjusted	0.804 / 0.788			0.787 / 0.769			0.777 / 0.758			0.742 / 0.718			0.794 / 0.774			0.804 / 0.783		
AIC	143.379			153.965			160.909			157.967			146.058			97.789		

Due to the social subset providing no significant results with happiness during the previous fixed effects regression model, the prosperity index was included into the study at this stage to substitute for the social indicators. Prosperity indicators were then examined at a cross sectional level from 2015 to 2020 (see Table 6).

Table 6: Regression Models on Prosperity (2015-2020)

Prosperity on Happiness																		
Predictors	2015			2016			2017			2018			2019			2020		
	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
(Intercept)	-0.07	-0.16 – 0.03	0.175	-0.07	-0.16 – 0.03	0.163	-0.02	-0.12 – 0.09	0.757	0.07	-0.04 – 0.17	0.233	0.02	-0.10 – 0.14	0.691	0.15	0.02 – 0.28	0.026
Safety and Security	-0.07	-0.23 – 0.08	0.355	-0.15	-0.31 – 0.02	0.075	-0.00	-0.16 – 0.16	0.994	0.06	-0.09 – 0.21	0.420	-0.01	-0.18 – 0.15	0.863	0.03	-0.13 – 0.19	0.679
Personal Freedom	-0.05	-0.27 – 0.17	0.644	0.20	-0.02 – 0.41	0.079	0.27	0.04 – 0.50	0.022	0.36	0.15 – 0.57	0.001	0.47	0.24 – 0.70	<0.001	0.03	-0.23 – 0.28	0.846
Governance	0.04	-0.39 – 0.48	0.849	-0.29	-0.72 – 0.14	0.180	-0.34	-0.76 – 0.07	0.103	-0.54	-0.92 – 0.16	0.006	-0.66	-1.07 – 0.26	0.001	-0.04	-0.45 – 0.36	0.833
Social Capital	0.34	0.21 – 0.47	<0.001	0.30	0.17 – 0.43	<0.001	0.32	0.18 – 0.46	<0.001	0.37	0.24 – 0.50	<0.001	0.33	0.19 – 0.47	<0.001	0.22	0.08 – 0.36	0.002
Investment Environment	0.05	-0.39 – 0.48	0.835	0.14	-0.30 – 0.57	0.534	-0.05	-0.50 – 0.39	0.811	-0.05	-0.50 – 0.41	0.842	-0.07	-0.55 – 0.40	0.761	0.27	-0.19 – 0.73	0.240
Enterprise Conditions	-0.07	-0.39 – 0.25	0.681	-0.01	-0.35 – 0.33	0.947	0.07	-0.25 – 0.39	0.668	0.16	-0.15 – 0.46	0.307	0.25	-0.05 – 0.55	0.107	0.07	-0.24 – 0.39	0.652
Infrastructure and Market Access	-0.04	-0.48 – 0.39	0.837	-0.02	-0.51 – 0.48	0.950	-0.06	-0.55 – 0.44	0.815	-0.17	-0.64 – 0.30	0.467	-0.24	-0.74 – 0.27	0.353	-0.64	-1.15 – -0.13	0.014
Economic Quality	0.29	0.07 – 0.51	0.010	0.31	0.09 – 0.53	0.007	0.38	0.15 – 0.62	0.002	0.38	0.14 – 0.62	0.003	0.43	0.18 – 0.69	0.001	0.20	-0.03 – 0.44	0.092
Living Conditions	0.51	0.15 – 0.86	0.006	0.45	0.04 – 0.86	0.033	0.75	0.32 – 1.18	0.001	0.71	0.30 – 1.12	0.001	0.73	0.29 – 1.18	0.002	0.91	0.42 – 1.41	<0.001
Health	0.11	-0.12 – 0.35	0.341	0.13	-0.12 – 0.37	0.300	0.07	-0.18 – 0.33	0.580	0.05	-0.20 – 0.31	0.676	0.21	-0.07 – 0.48	0.135	-0.17	-0.49 – 0.14	0.279
Education	-0.20	-0.49 – 0.09	0.167	-0.15	-0.44 – 0.14	0.300	-0.41	-0.73 – -0.10	0.010	-0.36	-0.67 – -0.05	0.024	-0.42	-0.75 – 0.09	0.014	-0.05	-0.39 – 0.30	0.794
Natural Environment	0.11	-0.04 – 0.26	0.153	0.10	-0.05 – 0.25	0.187	0.07	-0.09 – 0.23	0.389	0.06	-0.09 – 0.20	0.449	0.09	-0.07 – 0.24	0.268	0.14	-0.02 – 0.30	0.095
Observations	103			103			102			98			94			81		
R ² / R ² adjusted	0.826 / 0.803			0.813 / 0.788			0.807 / 0.781			0.814 / 0.788			0.833 / 0.808			0.805 / 0.771		
AIC	139.013			148.368			153.771			133.710			134.105			105.444		

Results of the regression models on prosperity indicated the three main variables to be social capital, economic quality and living conditions. Social capital involves the social and

personal relationships of individuals and this indicator would indicate that in the social perspective, individuals that tend to have stronger social relationships would have relatively higher values of happiness. Education was also significant during the years of 2017 to 2019 but not in 2020. Once again, the appearance and disappearance of significant variables throughout the years support the theme that variables that relate to happiness can vary in strength at different time periods.

Table 7: Regression Models on All Indicators (2015-2020)

All on Happiness																		
Predictors	2015			2016			2017			2018			2019			2020		
	Estimates	CI	p															
(Intercept)	-0.06	-0.19 - 0.07	0.334	-0.10	-0.26 - 0.05	0.183	-0.02	-0.24 - 0.19	0.834	0.27	-0.01 - 0.55	0.058	0.18	-0.10 - 0.46	0.214	0.09	-0.41 - 0.59	0.708
Safety and Security	-0.02	-0.21 - 0.18	0.861	-0.08	-0.28 - 0.12	0.415	-0.01	-0.22 - 0.20	0.930	0.14	-0.04 - 0.33	0.126	0.04	-0.14 - 0.22	0.628	0.13	-0.04 - 0.30	0.134
Personal Freedom	-0.20	-0.56 - 0.16	0.275	-0.07	-0.47 - 0.33	0.729	0.21	-0.18 - 0.60	0.276	0.20	-0.16 - 0.56	0.271	0.51	0.18 - 0.85	0.003	-0.24	-0.60 - 0.13	0.200
Governance	0.24	-0.35 - 0.83	0.427	0.19	-0.43 - 0.81	0.538	-0.25	-0.85 - 0.34	0.397	-0.48	-1.04 - 0.09	0.097	-0.76	-1.28 - 0.24	0.005	0.16	-0.36 - 0.68	0.540
Social Capital	0.23	0.01 - 0.45	0.038	0.16	-0.06 - 0.37	0.152	0.05	-0.22 - 0.31	0.734	0.14	-0.12 - 0.40	0.300	0.11	-0.11 - 0.34	0.327	0.22	-0.05 - 0.50	0.112
Investment Environment	0.17	-0.33 - 0.67	0.504	-0.05	-0.55 - 0.45	0.839	-0.17	-0.71 - 0.37	0.536	-0.16	-0.70 - 0.38	0.563	-0.13	-0.67 - 0.41	0.636	0.14	-0.41 - 0.69	0.612
Enterprise Conditions	-0.13	-0.48 - 0.22	0.455	-0.09	-0.46 - 0.27	0.608	0.11	-0.25 - 0.48	0.540	0.29	-0.07 - 0.64	0.112	0.24	-0.10 - 0.58	0.169	0.06	-0.30 - 0.41	0.756
Infrastructure and Market Access	-0.16	-0.68 - 0.35	0.531	0.12	-0.51 - 0.74	0.706	0.09	-0.50 - 0.68	0.764	-0.20	-0.80 - 0.41	0.519	0.03	-0.53 - 0.59	0.917	-0.04	-0.72 - 0.65	0.912
Economic Quality	0.30	0.01 - 0.60	0.046	0.34	0.04 - 0.63	0.026	0.26	-0.06 - 0.58	0.116	0.54	0.18 - 0.90	0.004	0.45	0.10 - 0.80	0.012	0.08	-0.19 - 0.34	0.570
Living Conditions	0.23	-0.47 - 0.94	0.511	0.12	-0.72 - 0.97	0.775	0.88	0.07 - 1.68	0.034	0.99	0.13 - 1.86	0.025	0.51	-0.33 - 1.36	0.231	0.32	-0.57 - 1.21	0.475
Health	0.06	-0.45 - 0.56	0.827	0.05	-0.46 - 0.56	0.839	0.19	-0.35 - 0.72	0.484	-0.01	-0.52 - 0.49	0.953	0.17	-0.40 - 0.73	0.559	-0.19	-0.88 - 0.49	0.570
Education	-0.05	-0.39 - 0.29	0.759	-0.08	-0.43 - 0.27	0.640	-0.28	-0.63 - 0.07	0.121	-0.13	-0.50 - 0.24	0.484	-0.22	-0.60 - 0.16	0.258	0.07	-0.36 - 0.50	0.749
Natural Environment	0.02	-0.20 - 0.24	0.842	0.00	-0.21 - 0.21	0.986	-0.00	-0.22 - 0.22	0.988	0.04	-0.17 - 0.25	0.680	-0.00	-0.20 - 0.19	0.988	0.12	-0.13 - 0.37	0.331
Social support	0.17	-0.02 - 0.35	0.085	0.17	-0.03 - 0.37	0.089	0.20	-0.05 - 0.45	0.110	0.15	-0.15 - 0.46	0.318	0.32	0.10 - 0.54	0.005	0.15	-0.11 - 0.42	0.246
Healthy life expectancy at birth	0.34	-0.14 - 0.83	0.165	0.40	-0.09 - 0.88	0.107	0.22	-0.30 - 0.74	0.405	0.32	-0.18 - 0.82	0.210	0.29	-0.24 - 0.82	0.282	0.13	-0.44 - 0.70	0.639
Freedom to make life choices	0.01	-0.17 - 0.18	0.940	0.08	-0.14 - 0.29	0.470	0.06	-0.16 - 0.29	0.569	0.03	-0.20 - 0.26	0.772	-0.01	-0.28 - 0.27	0.963	0.21	-0.06 - 0.48	0.125
Generosity	-0.08	-0.22 - 0.06	0.278	0.05	-0.10 - 0.19	0.511	-0.00	-0.14 - 0.14	0.972	-0.02	-0.16 - 0.13	0.797	0.01	-0.13 - 0.16	0.843	-0.08	-0.28 - 0.12	0.434
Perceptions of corruption	-0.13	-0.29 - 0.03	0.118	-0.05	-0.24 - 0.13	0.561	-0.19	-0.39 - 0.00	0.053	-0.12	-0.29 - 0.06	0.184	-0.24	-0.41 - 0.07	0.007	-0.08	-0.31 - 0.15	0.475
Positive affect	0.10	-0.06 - 0.27	0.212	0.09	-0.09 - 0.27	0.309	0.13	-0.03 - 0.30	0.113	0.10	-0.07 - 0.28	0.240	0.04	-0.12 - 0.20	0.633	-0.06	-0.25 - 0.13	0.522
Negative affect	0.09	-0.04 - 0.22	0.158	0.07	-0.07 - 0.22	0.306	-0.05	-0.20 - 0.11	0.547	0.02	-0.14 - 0.17	0.826	0.04	-0.12 - 0.20	0.637	-0.07	-0.23 - 0.09	0.390
Confidence in national government	-0.00	-0.19 - 0.18	0.966	-0.13	-0.33 - 0.07	0.199	-0.12	-0.36 - 0.11	0.294	-0.10	-0.32 - 0.11	0.349	-0.01	-0.22 - 0.19	0.912	-0.20	-0.39 - 0.00	0.048
Log GDP per capita	-0.26	-0.72 - 0.19	0.255	0.01	-0.50 - 0.52	0.974	-0.15	-0.68 - 0.39	0.591	-0.20	-0.77 - 0.38	0.498	-0.13	-0.65 - 0.39	0.626	-0.14	-0.80 - 0.53	0.681
Agricultural land (sq km)	0.13	-0.01 - 0.26	0.062	0.04	-0.10 - 0.18	0.574	0.08	-0.06 - 0.22	0.259	0.05	-0.09 - 0.19	0.481	0.11	-0.02 - 0.24	0.110	0.04	-0.09 - 0.17	0.570
GDP growth (annual %)	0.02	-0.11 - 0.14	0.802	-0.19	-0.35 - 0.02	0.026	0.13	-0.11 - 0.36	0.286	-0.07	-0.32 - 0.18	0.582	0.08	-0.18 - 0.34	0.550	0.03	-0.15 - 0.20	0.769
Gross savings (% of GNI)	-0.08	-0.24 - 0.07	0.290	-0.05	-0.20 - 0.09	0.480	-0.05	-0.21 - 0.11	0.533	-0.10	-0.26 - 0.05	0.199	-0.12	-0.28 - 0.04	0.151	-0.01	-0.16 - 0.13	0.856
Inflation, consumer prices (annual %)	-0.14	-0.50 - 0.22	0.444	-0.07	-0.55 - 0.40	0.753	-0.31	-0.91 - 0.30	0.320	0.06	-0.57 - 0.69	0.849	-0.08	-0.16 - 0.00	0.050	-0.03	-0.06 - 0.00	0.064
Mortality rate, infant (per 1,000 live births)	0.30	-0.01 - 0.62	0.056	0.48	0.17 - 0.80	0.003	0.62	0.22 - 1.01	0.003	0.69	0.26 - 1.11	0.002	0.33	-0.03 - 0.69	0.071	0.39	-0.08 - 0.86	0.099
Labor force, total	-0.08	-0.23 - 0.08	0.313	-0.07	-0.23 - 0.08	0.347	-0.03	-0.19 - 0.13	0.703	-0.06	-0.22 - 0.11	0.497	-0.10	-0.25 - 0.06	0.218	-0.10	-0.25 - 0.05	0.203
Access to clean fuels and technologies for cooking (% of population)	-0.02	-0.34 - 0.31	0.920	-0.12	-0.48 - 0.24	0.514	-0.05	-0.43 - 0.33	0.808	-0.09	-0.45 - 0.27	0.617	0.08	-0.30 - 0.45	0.680	-0.14	-0.51 - 0.23	0.442
Access to electricity (% of population)	0.21	-0.10 - 0.53	0.177	0.34	-0.00 - 0.68	0.051	0.10	-0.31 - 0.50	0.635	0.36	-0.02 - 0.74	0.064	0.13	-0.29 - 0.54	0.546	0.42	-0.04 - 0.87	0.072
Forest area (% of land area)	0.04	-0.11 - 0.18	0.632	0.02	-0.12 - 0.16	0.791	-0.03	-0.18 - 0.12	0.690	-0.06	-0.20 - 0.09	0.431	-0.00	-0.14 - 0.14	0.955	-0.04	-0.20 - 0.12	0.604
Arable land (% of land area)	0.12	-0.01 - 0.25	0.073	0.05	-0.07 - 0.18	0.383	-0.00	-0.13 - 0.13	0.968	-0.05	-0.17 - 0.08	0.477	0.00	-0.12 - 0.12	0.997	0.05	-0.08 - 0.19	0.439
Individuals using the Internet (% of population)	0.14	-0.28 - 0.55	0.510	-0.03	-0.44 - 0.38	0.891	-0.07	-0.52 - 0.37	0.747	-0.40	-0.90 - 0.10	0.116	-0.39	-0.89 - 0.11	0.126	-0.01	-0.57 - 0.55	0.970

Table 7 provides the results of regression models for all variables from the four subsets during the years of 2015 to 2020. The main variables that remained significant for the majority of the years were economic quality and mortality rates. Among all the regression models conducted, the most common form of variables that remained to have significant results with happiness revolved around economic, health and social indicators.

Factor Analysis

Factor analysis was conducted to reduce the dimensions into factors. To determine the number of factors needed for the factor analysis model, parallel analysis was run on the correlations of the numerical variables in the entire dataset. The parallel analysis scree plot (see Figure 8) recommends three factors since after three factors, the decrease in the eigenvalue is much less for each additional factor. With three factors, factor analysis was conducted using the varimax rotation and illustrated by the plots in Figure 9 and Figure 10.

Figure 8: Parallel Analysis

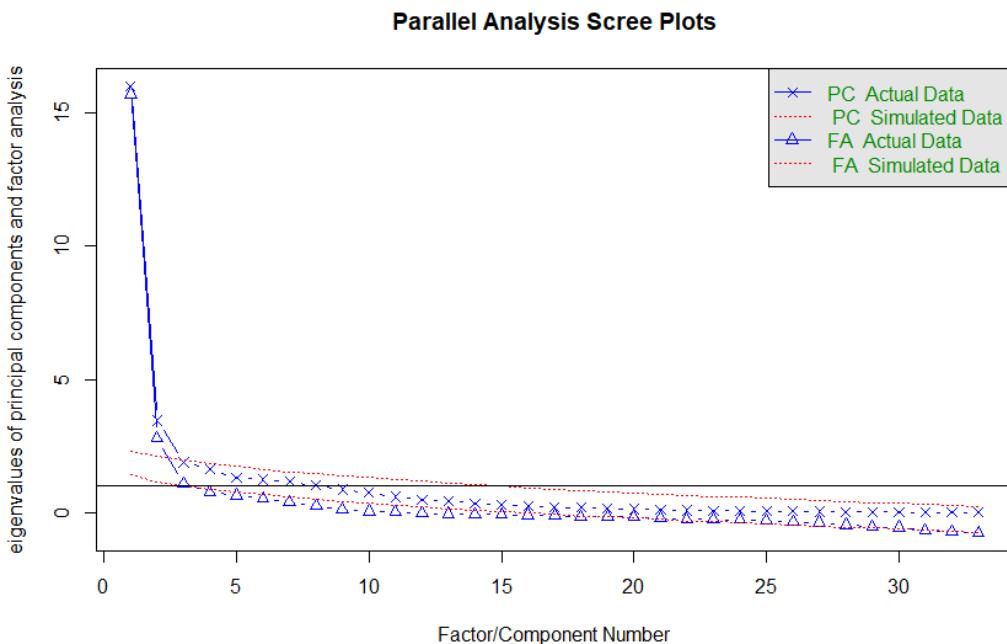
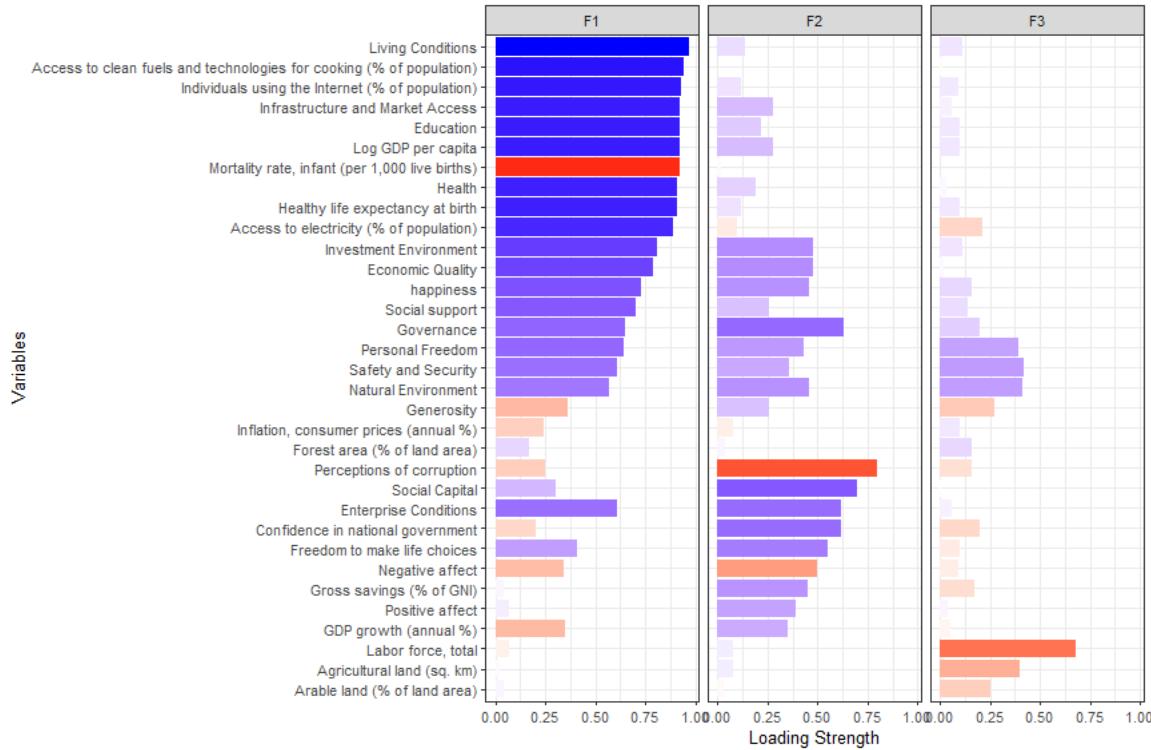
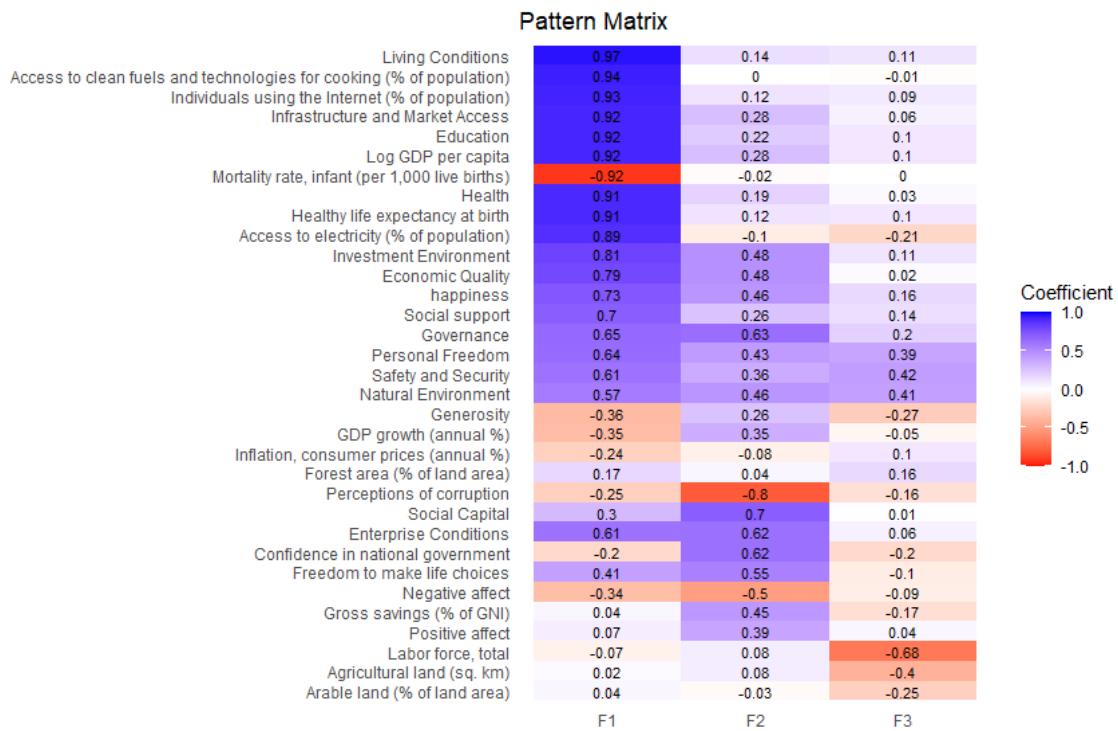
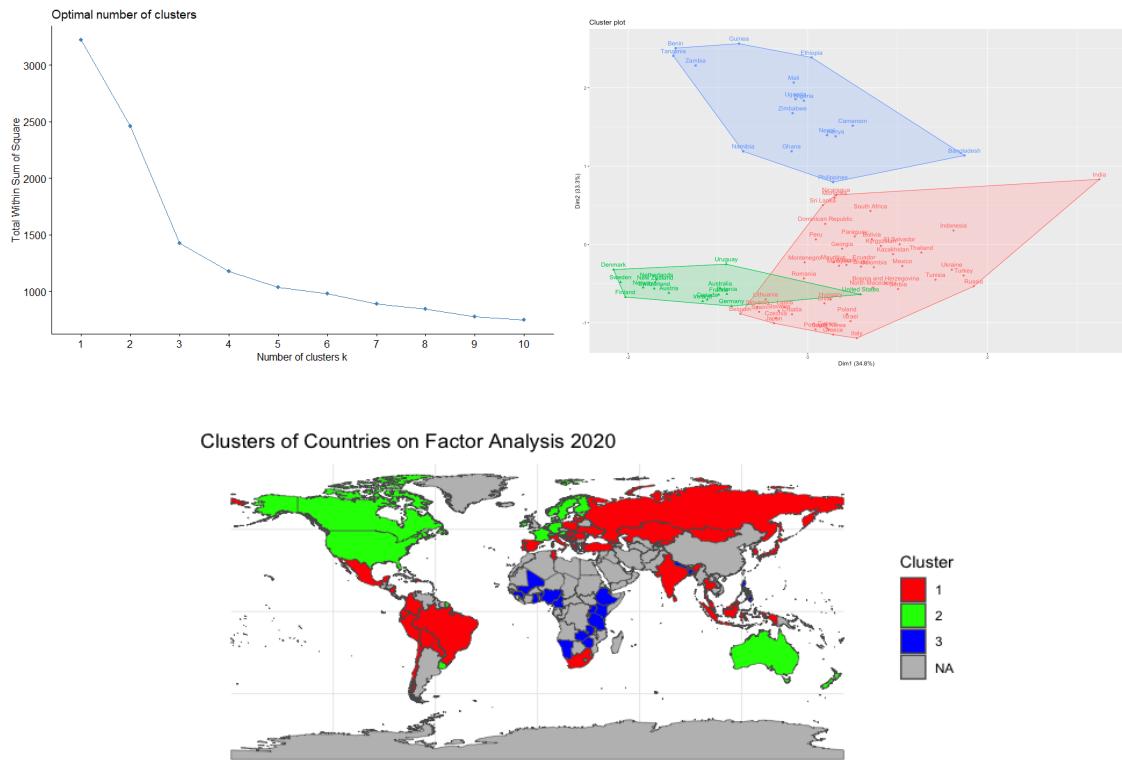


Figure 9: Factor Analysis**Figure 10:** Factor Analysis Pattern Matrix

Examining the variables making up the three factors in Figure 9 and Figure 10 provides useful information of the category of the factors. Factor 1 explains 40% of the variance and consists of the health of individuals and health of the economy with higher scores indicating higher health. Factor 2 explains 15% of the variance and is composed of trusts in the institutions or government and personal rights with higher scores indicating higher trust and rights. Factor 3 explains 4% and is composed of employment and land.

Utilizing the factors obtained from factor analysis, partitioning around medoids was conducted to cluster the countries. The “wss” method was used and determined 3 clusters would be an appropriate amount. Cluster categories were generated through the k-medoid method and results were plotted into a world map (see Figure 11).

Figure 11: Factor Analysis and PAM



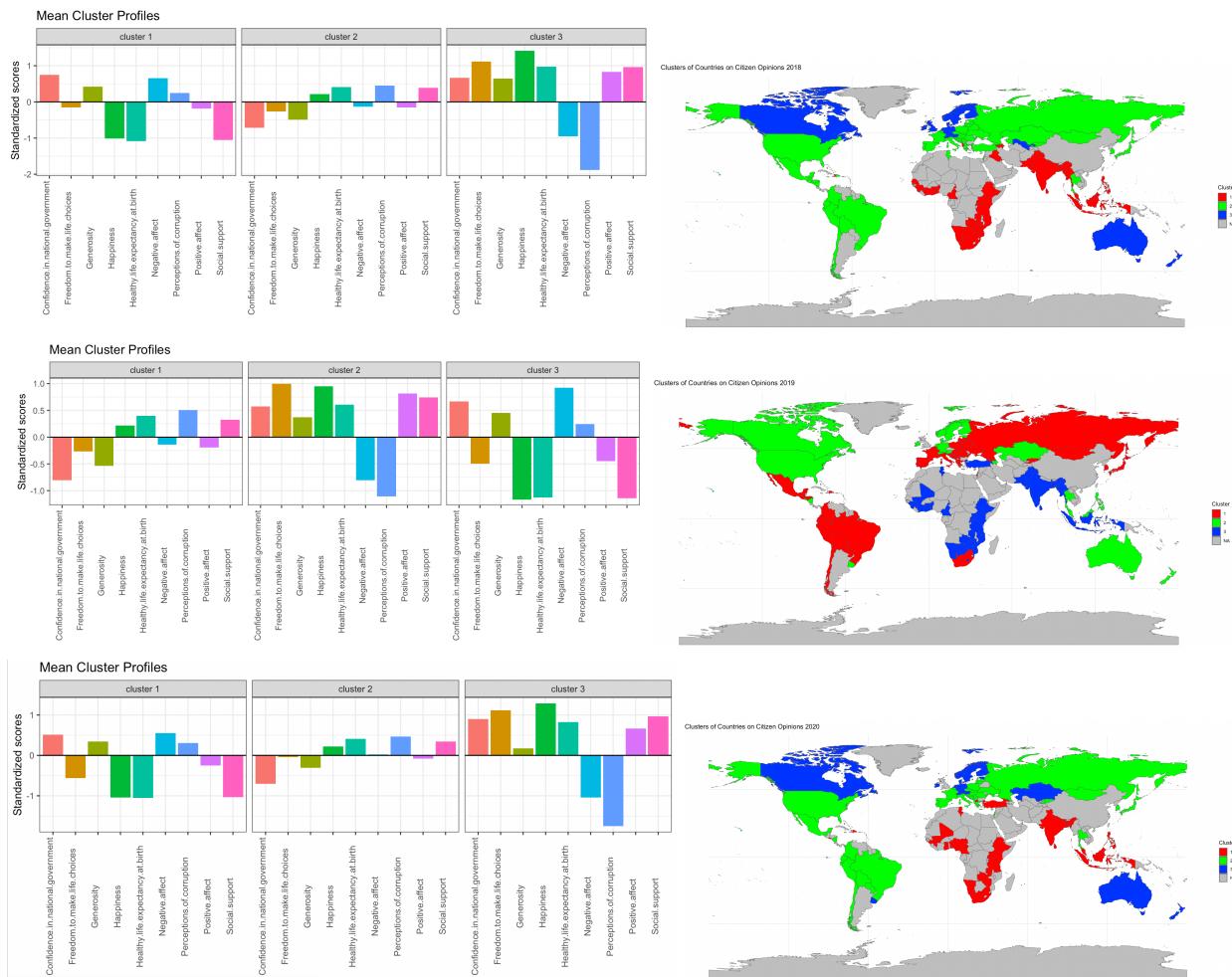
Note: Countries in gray in the map represents the data did not have information for the country.

Observing the world map in Figure 11 provides patterns that countries in North America, a few in Europe and Australia were similar to one another while South America, Russia, India and several South Asian countries were similar to one another. The last cluster involves the countries within Africa.

Principal Component Analysis

Principal component analysis was conducted to further examine similar countries and their clusters by subsets and by year.

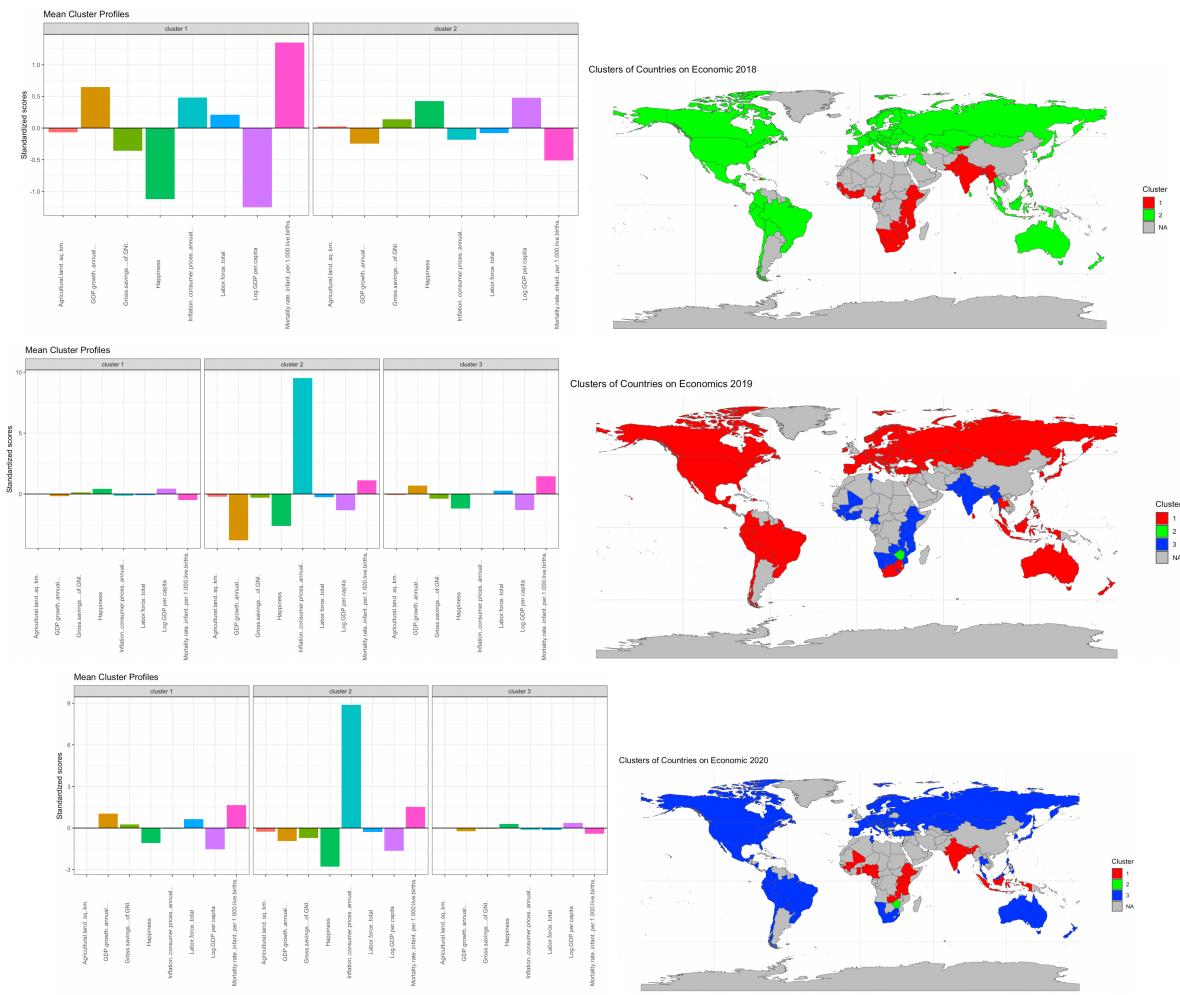
Figure 12: PCA Clusters on Citizen Opinions (2018-2020)



Note: Clusters world maps on row one are for the citizens opinions subset in 2018, row two for 2019 and row three for 2020. Countries in gray represent missing data for the country.

Figure 12 illustrates the clusters of the citizen opinions subset from the years 2018 to 2020 with patterns in the world map indicating similarities with the world map from the factors in Figure 11. Once again, the illustration demonstrates South America with Russia are similar with one another, African countries with India and South Asian countries remain similar as well. However, the countries in North America and Australia varied throughout the years.

Figure 13: PCA Clusters on Economic (2018-2020)



Note: Clusters world maps on row one are for the economics subset in 2018, row two for 2019 and row three for 2020. Countries in gray represent missing data for the country.

Figure 13 illustrates the clusters of the economic indicators from 2018 to 2020 with results indicating that there are often two clusters of countries. However, in 2019 and 2020, Zimbabwe was potentially affected by the pandemic so greatly that their values became an outlier and distinguished itself as its own cluster. Hence, there were three clusters for 2019 and 2020 when there were only two clusters in 2018. The two clusters often represent either the larger countries or the smaller countries.

Figure 14: PCA Clusters on Prosperity 2020

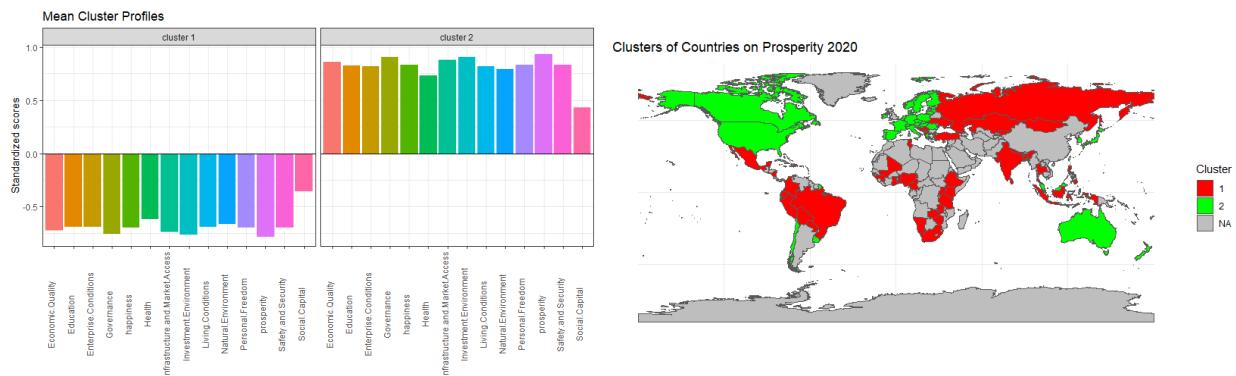


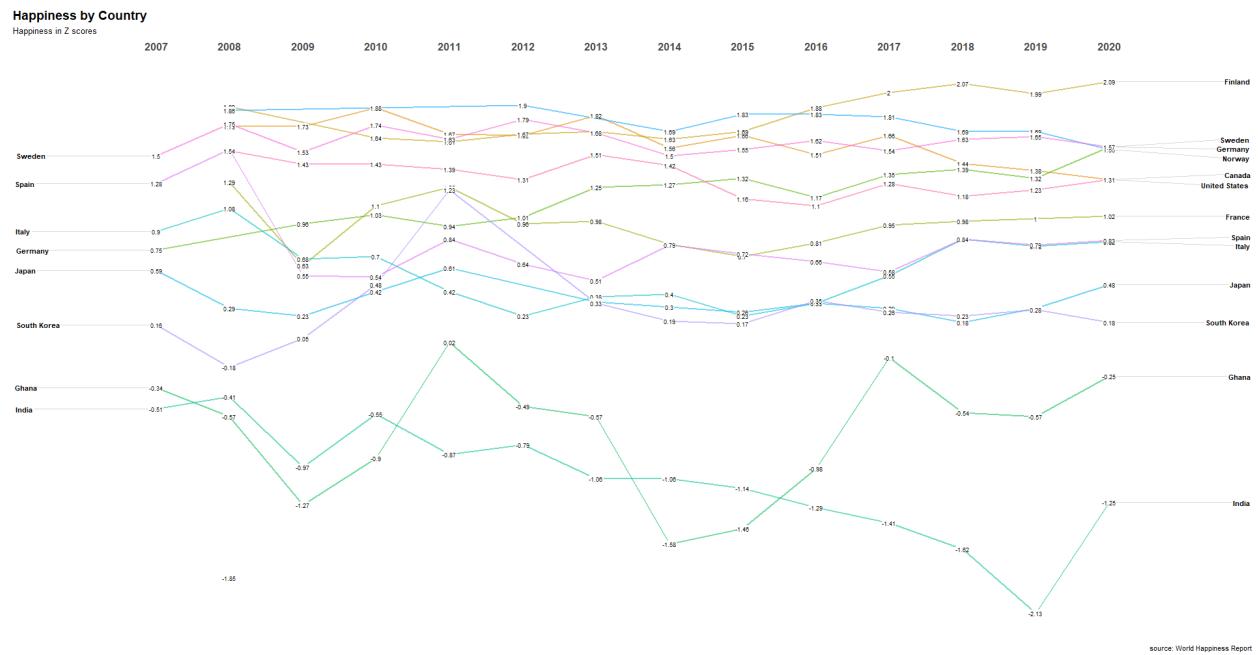
Figure 14 provided a relatively unique cluster for the prosperity subset. The cluster profile illustrates that countries were separated by high prosperity and low prosperity in the dataset. However, the pattern seen in the world map remains relatively consistent to previous clusters from the factor analysis and PCA of the other subsets. In the previous clusters, African countries, India and several South Asian countries were in their own clusters. In Figure 14, these countries appear to be similar in prosperity with the South American countries and Russia.

Pre-Post Covid Analysis

An exploration of happiness during pre and post covid was conducted to examine whether a crisis such as the pandemic affected countries strongly or relatively stronger than other countries. Figure 15 provides a time series slope graph of a few selected countries over the year.

Several countries such as Ghana and India experienced a larger change in happiness while other countries remained relatively stable throughout the years. Finland remains one of the top happiest countries along with its neighbors such as Sweden and Norway.

Figure 15: Slope Graph of Happiness by Selected Countries over the Years



Note: The single value in 2008 represents Burundi and appears this way as the dataset only includes data for one year for Burundi.

Finally, an attempt to create a predictive model using pre covid data (2007-2018) as the training dataset to predict post covid happiness (2019-2020) on the test dataset. This process was conducted through a random forest model (see Figure 16). The training dataset had 1100 observations while the test dataset had 175 observations with 33 variables in both training and testing datasets. This random forest model selected 100 trees utilizing the regression model. Results of the random forest model in Figure 16 states how well the model fits the data. The mean of squared residuals represents the model was wrong by a difference of 0.085 between the actual and predicted value of happiness. The model explained 91.5% of the variance.

Figure 16: Results of the Random Forest Model

```
call:
randomForest(x = train_X, y = train_y, ntree = 100, importance = TRUE)
  Type of random forest: regression
      Number of trees: 100
No. of variables tried at each split: 10

  Mean of squared residuals: 0.08517669
    % Var explained: 91.5
```

Figure 17: Random Forest Outputs

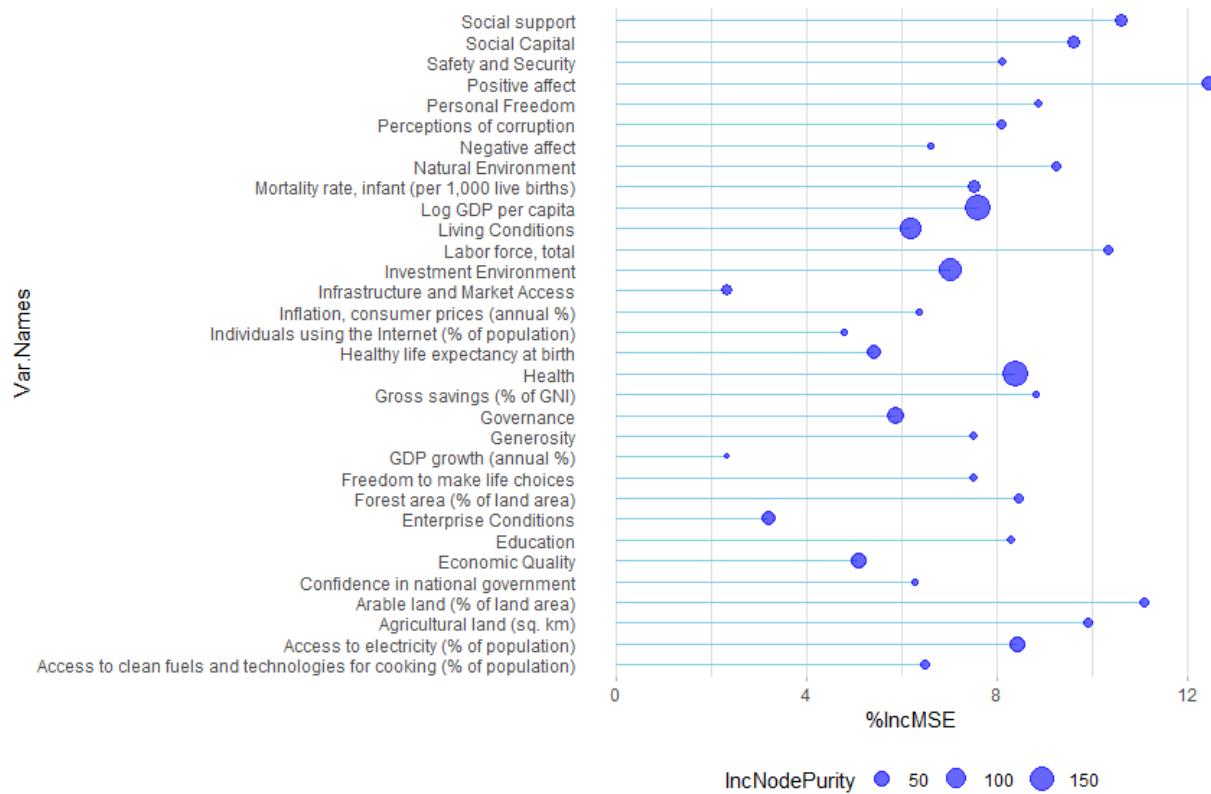
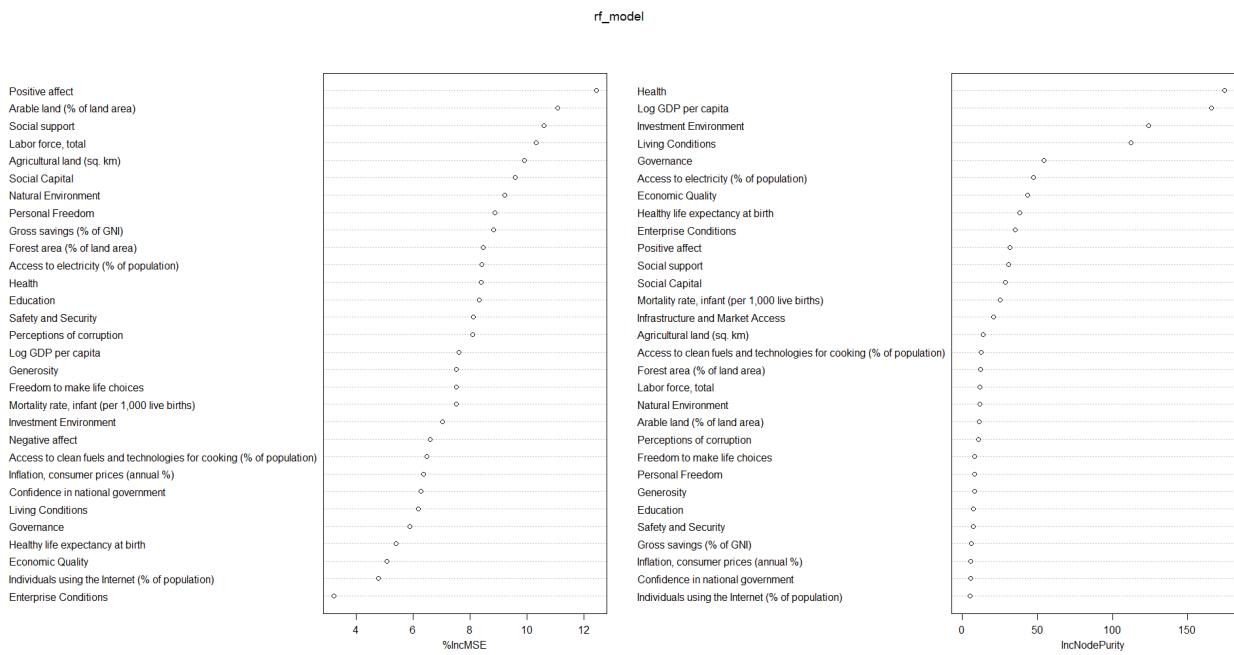


Figure 17 illustrates the outputs obtained by a random forest model using the regression type. These outputs provide more information on the importance of the variables used in the prediction of happiness. The prediction error is represented by the mean square error (MSE) while the node purity is the total decrease in the residual sum of squares. Figure 18 provides a sorted plot indicating that positive affect, arable land (% of land area), social support, labor force and agriculture land (sq. km) were the top five most important variables according to the MSE.

Figure 18: Random Forest Outputs



Discussion

Through an analysis of happiness from a variety of statistical methods, this study provides unique findings between numerous variables on happiness. From the regression analysis, results found that economic variables such as log GDP per capita and mortality rates remained significant throughout the cross sectional analysis when examining the data. For citizens' opinions, it was found that social support, healthy life expectancy, perceptions of corruption, positive affect and confidence in national government were important. The social subset had no significant results when a fixed effects regression model was conducted. Finally, the prosperity subset found social capital, economic quality and living conditions to be significant within the cross sectional study.

These significant findings indicate that economic, social and health indicators serve to be the most common factors that relate to happiness. Although, the social subset did not provide a significant result relating to happiness, several social variables within the citizens opinion and

prosperity subset indicated the importance of social factors. While there are specific indicators that remain significant throughout the majority of the years, there were also indicators that became significant for a few years before becoming no longer significant. This provides information that there can be variables that are more strongly related to happiness based on circumstances such as lifestyle, environment or simply drastic changes. For example, the pandemic in 2019 is likely to have brought inflation and labor force to be of more importance as prices rose for necessities and employees were at risk of losing their jobs. It should be noted that happiness is a complex item to measure and factors that make one person or country happy may not be the same for another. One factor that can make a country happy for certain years can also no longer be significantly related to a country's happiness the next year. It is extremely difficult to simply state that one specific variable is the most important area to focus on and that countries should seek to simply improve their happiness based on one specific factor.

With studies such as this present study, significant results can still support or provide information for countries to develop data driven decisions such as government programs providing assistance in the form of unemployment programs observed in Bonasia et al. (2022) or the social reforms in Sarracino et al. (2022). These findings can also inform individuals reading about happiness on what they can do socially as an individual to help improve their personal experience with happiness.

Furthermore, the world maps and clusters visualized countries that were relatively similar to one another based on the variables tested. These clusters and visualizations allow the identification of countries that are doing a better role of keeping their country happy as well as those that may require assistance. For example, Finland along with its neighbors has been doing quite well and has been ranked one of the happiest countries in the world for six consecutive

years (Bloom, 2023). On the other hand, results found Zimbabwe has become an outlier in having low happiness values in 2019 and 2020. This situation is likely due to the pandemic and it may potentially be beneficial if countries could provide support to help Zimbabwe reach the rest of the countries in happiness. Viewing the happier countries as a guideline in certain aspects may be beneficial in improving the happiness of a country.

Limitations and Directions for Future Research

While this study attempted to be comprehensive in most statistical tests, there were several limitations of the study to consider. First, variables selected for the subsets may not be the most representative of the subsets. For example, the social subset did not have a great amount of variables that can represent the social aspects of countries. Hence, this study included the prosperity subset to support the analysis of the social variables. Future studies can consider a more accurate social subset. Second, the dataset used in this study does not contain data for all countries. As seen in the world maps, there were numerous countries in a gray color format due to the dataset not having data for the countries. Furthermore, an exploration of the World Happiness Report dataset found that there were several countries that had no data for specific variables. Hence, these countries were removed during the data cleaning process. Future research can seek to obtain a more complete dataset containing a larger number of countries worldwide. Third, the predictive model with random forest is likely to be overfitted during the split of the train and test dataset. An ideal split would be around 70% or 80% training and 30% or 20% testing dataset. As a result, the findings from the random forest model in this study should not be interpreted directly. Fourth, there were suggestions for unemployment and gini index variables to be considered in relation to happiness, however, there was not enough data for these variables to

keep them within the dataset. Future research may consider obtaining these results for analysis as well.

References

- Bloom, L. B. (2023, March 21). *Ranked: The 20 happiest countries in the world in 2023*. Forbes.
<https://www.forbes.com/sites/laurabegleybloom/2023/03/20/ranked-the-20-happiest-countries-in-the-world-in-2023/?sh=74164ff979c2>
- Bonasia, M., Napolitano, O., Spagnolo, F., & Spagnolo, N. (2022). The economic and welfare state determinants of well-being in Europe. *International economics.*, 171, 49-57.
<https://doi.org/10.1016/j.inteco.2022.04.005>
- Bradstreet, Dr. C. (2020, March 20). *If I asked you, “Are you happy?” what would you say?* Thrive Global.
<https://community.thriveglobal.com/if-i-asked-you-are-you-happy-what-would-you-say/>
- Easterlin, R. A. A. (2001). Income and Happiness: Towards a Unified Theory. *The economic journal : the journal of the British Economic Association.*, 111(473), 465-484.
<https://doi.org/info:doi/>
- Ferrer-i-Carbonell, A. (2013). Happiness economics. *SERIES : journal of the Spanish Economic Association*, 4(1), 35-60. <https://doi.org/10.1007/s13209-012-0086-7>
- Helliwell, J. F., Layard, R., & Sachs, J. (2023). *World happiness report 2023*. Sustainable Development Solutions Network.
- Legatum Institute, The Legatum Prosperity Index 2023-a tool for transformation (2023)
- Oswald, A. J. (1997). Happiness and Economic Performance. *The Economic journal (London)*, 107(445), 1815-1831. <https://doi.org/10.1111/1468-0297.00260>
- Sarracino, F., O'Connor, K. J., Ono, H., & O'Connor, K. J. (2022). Are economic growth and well-being compatible? Welfare reform and life satisfaction in Japan. *Oxford economic papers.*, 74(3), 721-745. <https://doi.org/10.1093/oep/gpab038>

World Health Organization. (2023). Indicator metadata registry details. World Health Organization. <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/66>