

Rice University

## Happiness Among Countries

Which Factors are the Best Indicators of Happiness among Countries?

Dessy Akinfenwa, Sara Bolf, and Erin Kreuz  
STAT 406  
April 19, 2019

## Introduction

The project aims to use data wrangling, analysis, and modelling techniques in SAS to examine the relationship between a country's happiness level and indicators of basic freedoms, economic prosperity, and justice system integrity in 2016. To do so, the project uses one major dataset, The World Happiness Index, which ranks countries based on how residents rate their happiness level. The project also integrates two additional datasets including: a country-specific GDP per capita dataset, crude suicide rate dataset, and the Human Freedom Index dataset. Ultimately, the analysis yielded four main conclusions: (1) economic performance is the best indicator of happiness, but money has substantial diminishing returns after ~\$10,000 USD, (2) most predictors were poor predictors of happiness, (3) there are significant regional differences in happiness, and (4) of the basic freedoms, legal integrity and regulation of media content are the most highly correlated with happiness.

## Data Description

In analyzing the relationship between happiness and other economic, social, and political factors in countries in the world, we utilized one main dataset: (1) The World Happiness Report Dataset (obtained from Kaggle). The Sustainable Development Solutions Network, which is part of the United Nations, publishes *The World Happiness Report* annually (since 2012) in aims to better understand how individuals in 155 countries perceive their level of happiness.

The Sustainable Development Solutions Network partners with the Gallup World Poll and asks poll participants to rank their life on a scale of 1 to 10, where 1 represents what participants perceive as the worst possible life and 10 represents what participants perceive as the best possible life. The dataset also includes the breakdown for each country of how much different factors including: (1) economic conditions, (2) social structure, (3) health, (4) freedom, (5) governmental trust, and (6) societal generosity improve life over a base level of "Dystopia," or a fictional place with the worst possible levels of the six factors aforementioned. In recent years, *The World Happiness Report* has gained traction as a more holistic way to evaluate countries' performance beyond traditional economic and social metrics.

The analysis also uses three additional datasets: (1) The Suicide Mortality Rate by Country, (2) The GDP per Capita (\$USD) by Country, and (3) The Human Freedom Index. The first additional dataset consists of the Suicide Mortality Rate by Country and is produced by the World Bank. The dataset contains the crude suicide rate for each country. From the same source, we obtained a supplemental dataset that divided the suicide rates by gender for each country in 2016. This dataset was utilized to study the relationship between a country's overall level of happiness and their suicide rate. The second additional dataset consists of the GDP per Capita (\$USD) for each country and is also produced by the World Bank. This dataset was utilized to study the relationship between a country's overall level of happiness and their GDP per capita to study if more money really makes people happier. The third additional dataset (obtained from Kaggle) was the Human Freedom Index, created by the CATO Institute. The index essentially creates a ranking from 0 to 10 on the following twelve factors: (1) Rule of Law, (2) Security and Safety, (3) Movement, (4) Religion, (5) Association, Assembly, and Civil Society, (6)

Expression and Information, (7) Identity and Relationships, (8) Size of Government, (9) Legal System and Property Rights, (10) Access to Sound Money, (11) Freedom to Trade Internationally, and (12) Regulation of Credit, Labor and Business (Kaggle). This dataset was utilized to study the relationship between other related social and political factors and happiness. Notably, the overlapping year in all datasets was 2016, so the analysis was performed for data for the year 2016 unless otherwise indicated.

Notably, because much of the data we utilized was aggregated, there is risk that our correlations are ecological correlations, and as such, overestimate the true relationship between factors.

## Techniques

All data wrangling, analysis, and modelling was performed in the SAS software. The following techniques were utilized as part of the project:

1. **PROC IMPORT** to import .csv files
2. **WHERE** statements to filter the data for desired years only
3. **IF** and **THEN** statements to create and classify new variables
4. Interactive **DATA** steps to filter data, transform data, and create new variables
5. **PROC TRANSPOSE** to transform data from long to wide format
6. Utilizing **automatic variables** (use of `_n_`)
7. **PROC SQL** to create new tables, merge four datasets together, filter the datasets, drop/select columns,
8. **PROC SGLOT** to create bar charts, scatter plots, regression plots
9. **PROC REG** to perform regression analyses
10. **PROC ANOVA** to test there is a statistically significant difference exists between groups
11. Sorting the dataset using **PROC SORT**
12. **PROC FREQ** to create frequency tables and create frequency plots
13. **PROC UNIVARIATE** to create histograms and test normal distributions
14. **PROC CORR** to test correlations between variables
15. Interactive **MACROS functions** to assist with making plots
16. **PROC MEANS** to obtain summary statistics for different groups
17. **PROC IML** to transform the dataset into a matrix and develop a module that performs ordinary least squares
18. **PROC GMAP** to use country shapefiles to create choropleth maps.
19. **PROC GML** for modelling relationships

## Analysis

To perform our analysis, we evaluated the following eight questions:

- (1) Who is the happiest?
- (2) Does money make us happier?
- (3) Do happier countries have fewer suicides?
- (4) Are countries higher freedom of expression happier?
- (5) Which indicators of religious freedom relate to a country's happiness score?

(6) What variables associated with a country's legal system can be used to predict a country's happiness score?

(7) How do women's rights issues affect the happiness level of a country?

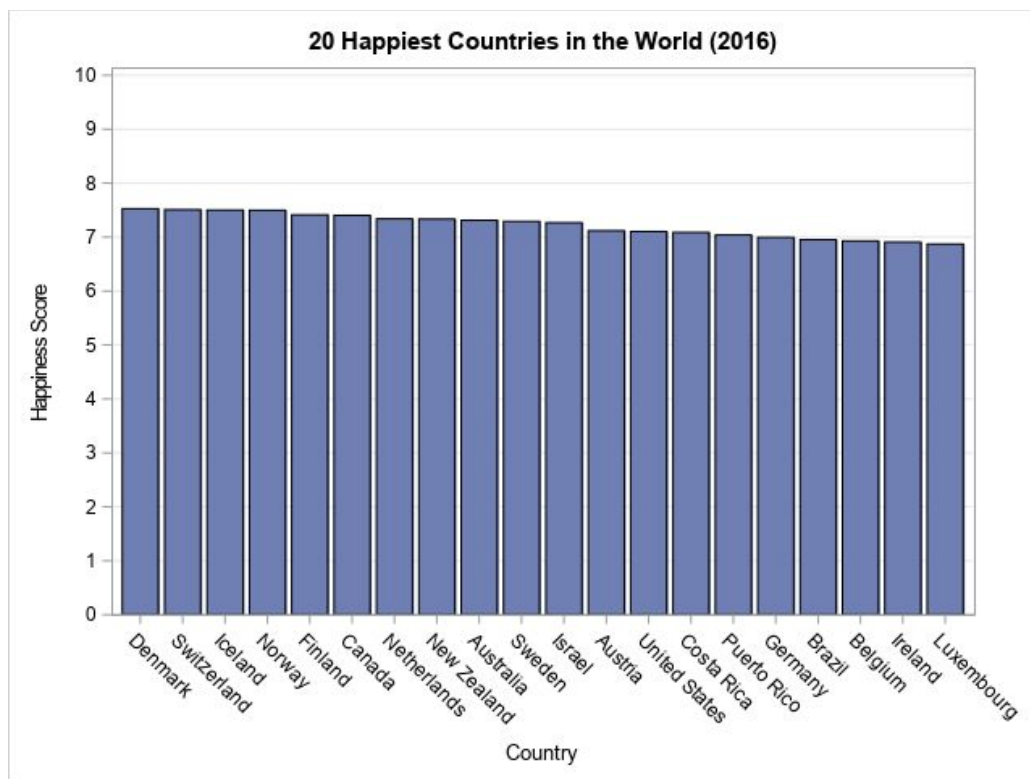
(8) Does the amount of crime in a country relate to its level of happiness?

*Who is the happiest?*

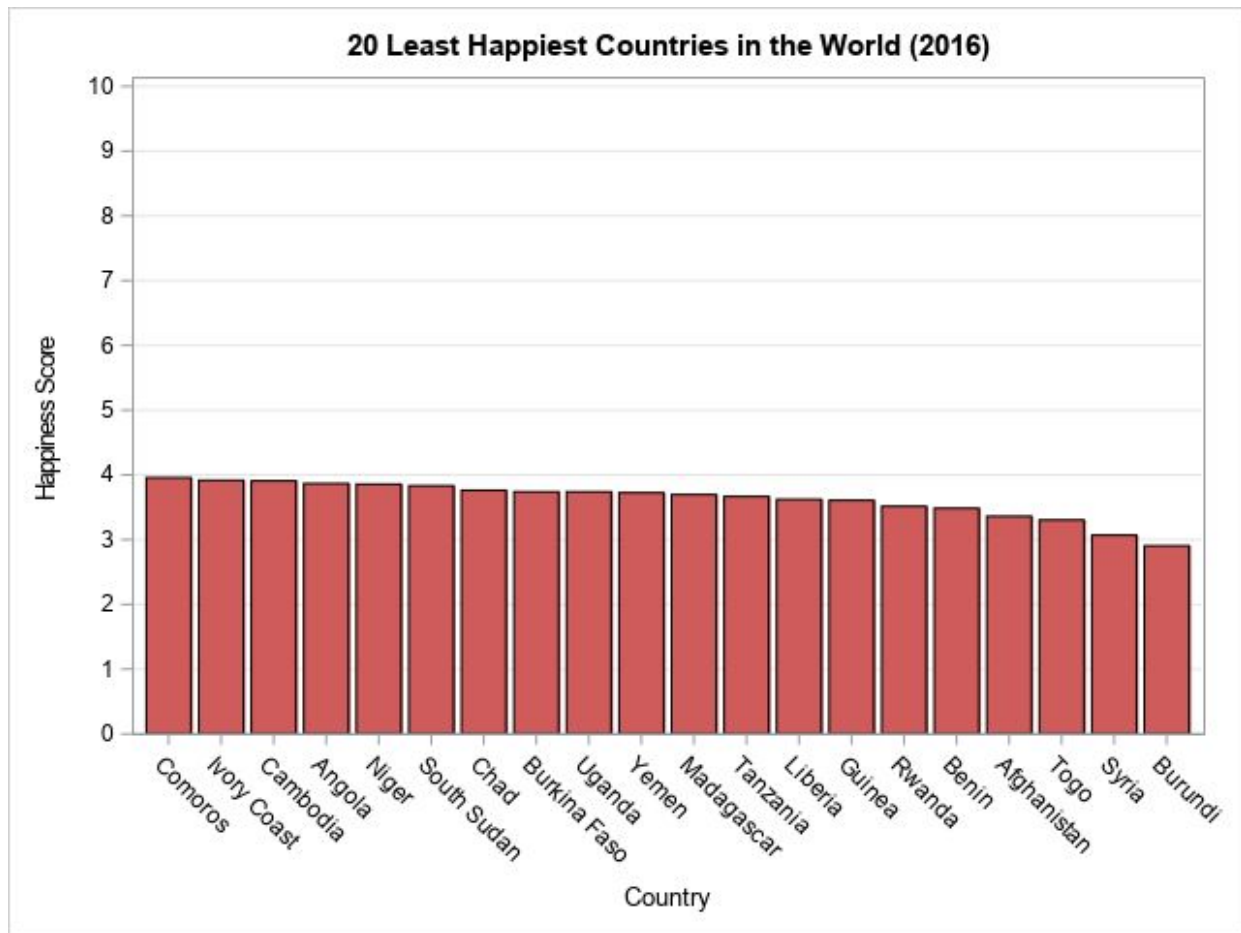
In analyzing how individuals in different countries rate their levels of happiness on a scale from 1 to 10, two main trends appear including: (1) substantial differences in scores among the happiest and the least happy countries in addition to (2) regional differences in happiness scores.

In 2016, the three happiest countries in the world were: Denmark, Switzerland, and Iceland, while the three least happy countries in the world were Togo, Syria, and Burundi. *Figure 1* shows the 20 happiest countries in the world and *Figure 2* shows the 20 least happy countries. Two noteworthy trends appear in examining the happiest and least happy countries: (1) Scandinavian countries dominate the list of happiest countries and (2) the least happy countries have substantial political and civil conflict. Firstly, of the five happiest countries, four are located in Scandinavia. Secondly, the three least happy countries (Syria, Togo, and Burundi) have been plagued with political conflict, civil wars, and violence, likely contributing to people's discontent in these countries.

***Figure 1: 20 Happiest Countries in the World***



**Figure 2: 20 Least Happy Countries in the World**



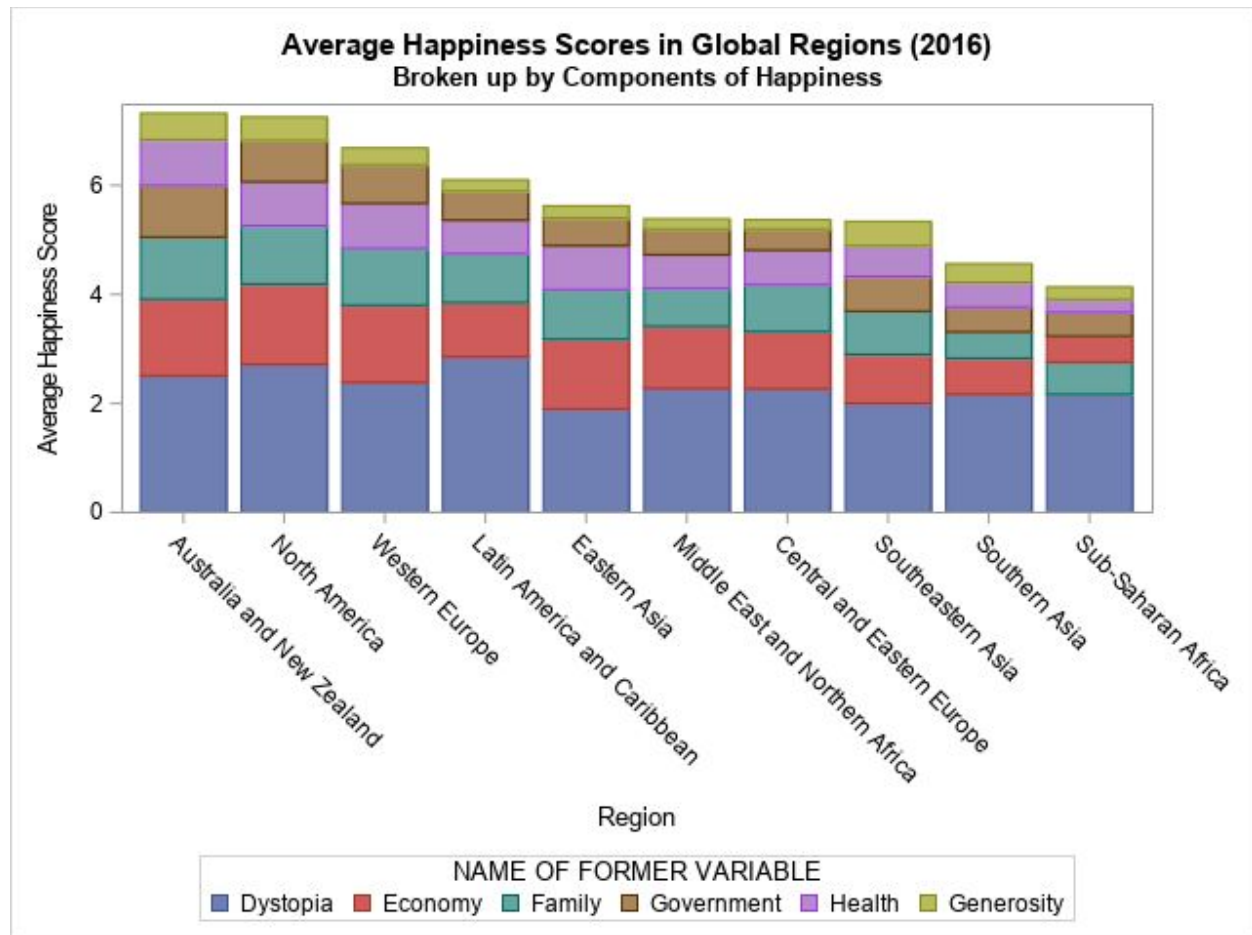
The countries with the top happiest ratings still do not achieve extremely high levels of happiness (greater than 8). Notably, in Denmark, the country with the highest average score, the happiness sits around 7.5. Relative to 10, it is clear that there is still room for improvement in happiness even in what is supposedly the happiest country in the world according to the Index.

In addition to being difference in how happy individuals are in different countries, there are also regional differences; overall, the Australia and New Zealand region's countries are on average the happiest, while the Sub Saharan African region's countries are on average the least average. *Figure 3* shows the breakdown of average happiness scores by region broken down by different components of the happiness score.

Interestingly, the Dystopia component of the score, or by how much people in the country perceive their lives as being better than the worst possible country, is relatively similar in the happiest and unhappiest regions. In other words, people in both Sub-Saharan Africa and Australia/New Zealand perceive their lives as being better than the worst country in the world by

the same amount. Additionally, the least happy regions score very poorly in health as well as economic status.

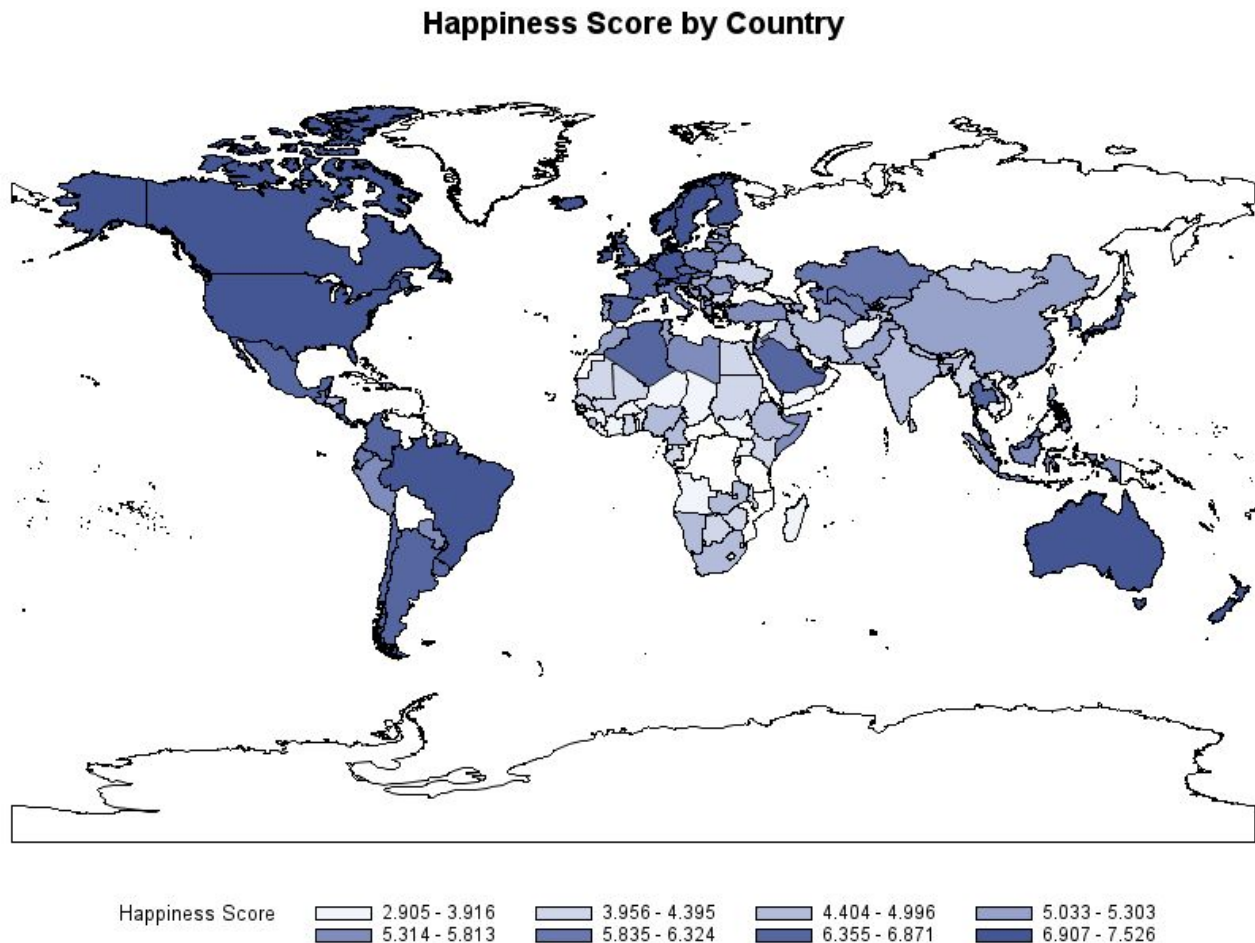
**Figure 3: Average Happiness Score by Region**



Despite a seemingly clear difference between regions' happiness scores, it is important to take the analysis a step further and determine if a statistically significant difference exists between the regions. To do so, we performed an ANOVA test. The test yielded that there is an F Value of 27 and a p-value of  $<.0001$ , which gives sufficient evidence to reject null hypothesis and conclude that a statistically significant difference in happiness scores among regions exists (see ANOVA results in *Table A.1* in *Appendix A*).

Although the figures shown the top twenty countries and bottom twenty countries as well as aggregated happiness by region provide a meaningful view of country's rankings and regional positions, a geospatial analysis is also meaningful. *Figure 4* shows a choropleth map of happiness by country, supporting the findings aforementioned. From the map, there are trends associated with happiness in different regions.

*Figure 4: Happiness Choropleth Map\**

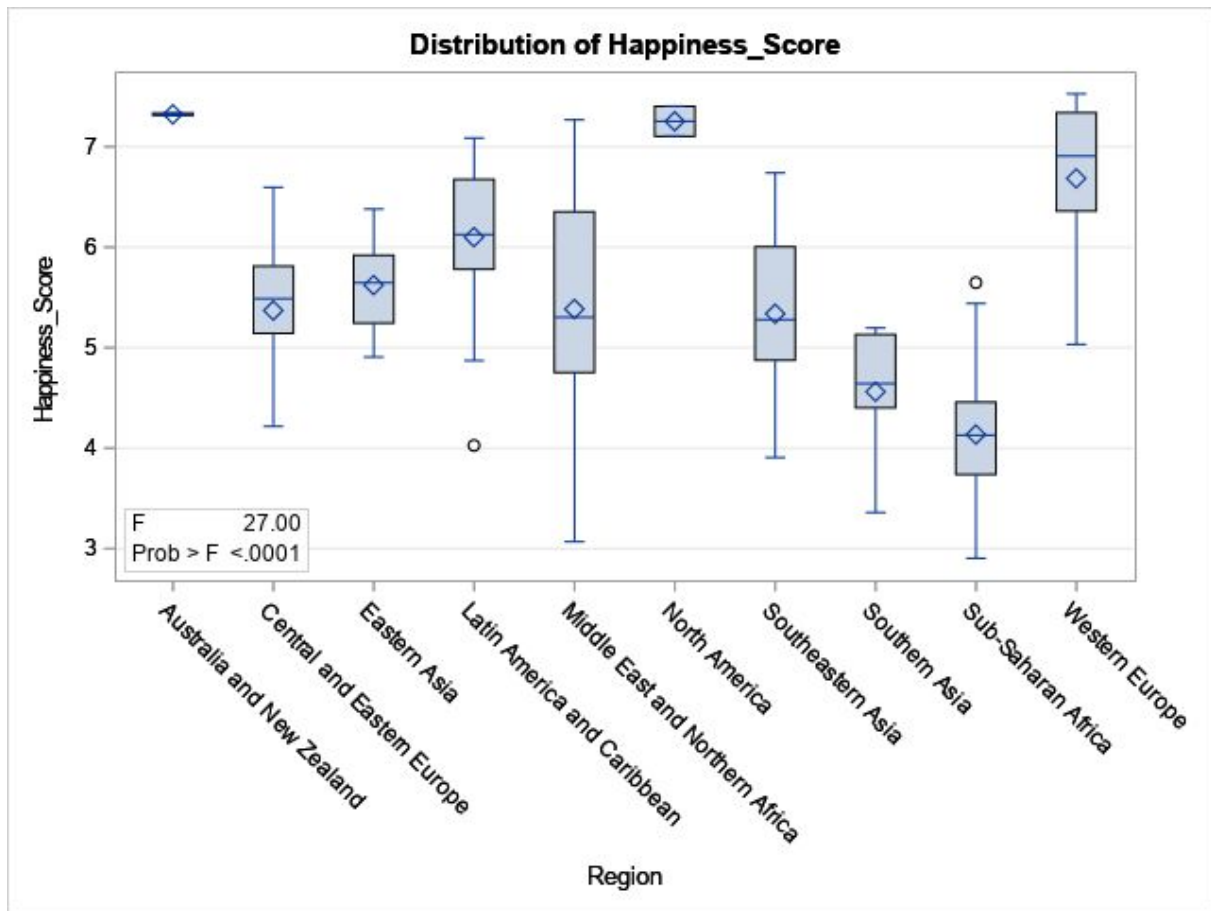


*\*Blank Countries are Missing Data*

The differences in happiness scores among regions is shown in the boxplot in *Figure 5*. Notably, the largest spread in happiness scores exists in the Middle East and North Africa (MENA) region. Additionally, in further examining at the histograms of each region (see *Figure A.1* in *Appendix A*), the Latin America and the Caribbean region has a heavily skewed left distribution of happiness scores, with most countries having a high score, with only a handful having low scores, pulling the overall average down.

In fact, as demonstrated in *Table 1* which shows the frequency and relative frequency of happiness scores greater and less than 5 by region, only 2 of 24 countries (~8.3%) of countries studied in the Latin America and Caribbean region had a happiness score less than 5. Other regions with less than 10% of individual country happiness scores being less than 5 include Australia and New Zealand, North America, and Western Europe. Sub Saharan Africa and Southern Asian had the highest percentage of countries with scores under 5, with ~92.1% and ~71.4% respectively

*Figure 5: Distribution of Happiness Score by Region*



*Table 1: Distribution of Happiness Score by Region*

#### The FREQ Procedure

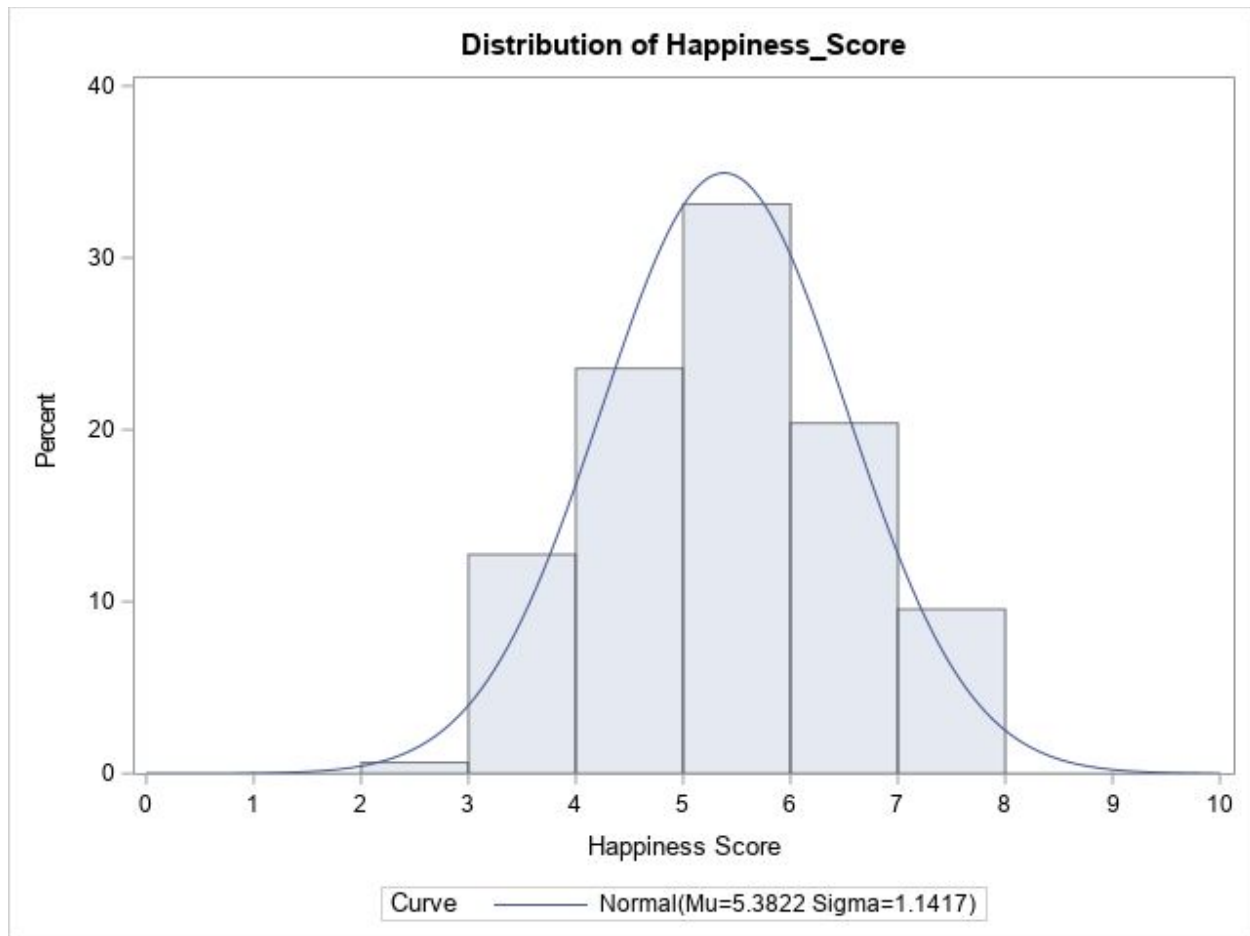
Frequency Row Pct	Table of Region by Classification			
	Region	Classification		
		Greater than 5	Less than 5	Total
	Australia and New Zealand	2 100.00	0 0.00	2
	Central and Eastern Europe	23 79.31	6 20.69	29
	Eastern Asia			



	5 83.33	1 16.67	6
<b>Latin America and Caribbean</b>	22 91.67	2 8.33	24
<b>Middle East and Northern Africa</b>	13 68.42	6 31.58	19
<b>North America</b>	2 100.00	0 0.00	2
<b>Southeastern Asia</b>	6 66.67	3 33.33	9
<b>Southern Asia</b>	2 28.57	5 71.43	7
<b>Sub-Saharan Africa</b>	3 7.89	35 92.11	38
<b>Western Europe</b>	21 100.00	0 0.00	21
<b>Total</b>	99	58	157

Overall, global happiness scores are normally distributed (with a mean of  $\sim 5.4$  and standard deviation of  $\sim 1.1$ ). *Figure 6* shows the distribution of Happiness Scores in a histogram fitted with a normal distribution. In a Komogorov-Smirnov test, we conclude with a p-value of .134 that a normal distribution is appropriate with happiness scores (see results in *Figure A.2* in *Appendix A*).

**Figure 6:** *Distribution of Happiness Score Globally (2016)*



Thus the four main takeaways from the overall analysis of happiness are: (1) there are regional differences in happiness, (2) Scandinavian countries appear to dominate the happiest countries in the world, (3) the least happy countries are plagued by political and civil conflict, and (4) globally, happiness scores are normally distributed.

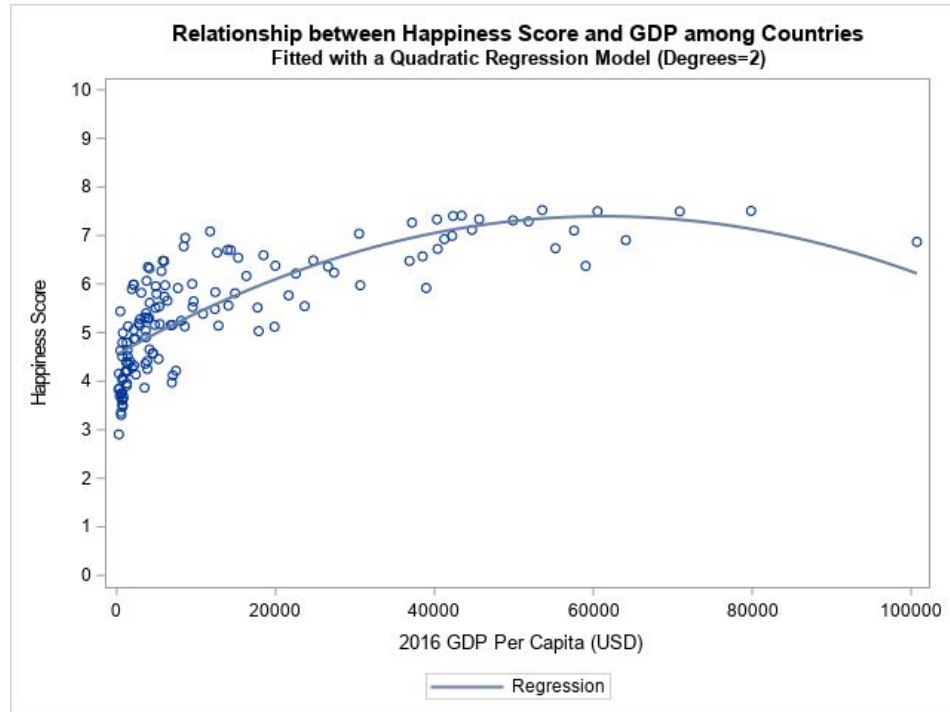
#### *Does money make us happier?*

After a thorough and complete understanding of happiness among countries, it is important to transition the analysis into understanding potential predictors of happiness. The first of those predictors that we examined was: economic prosperity.

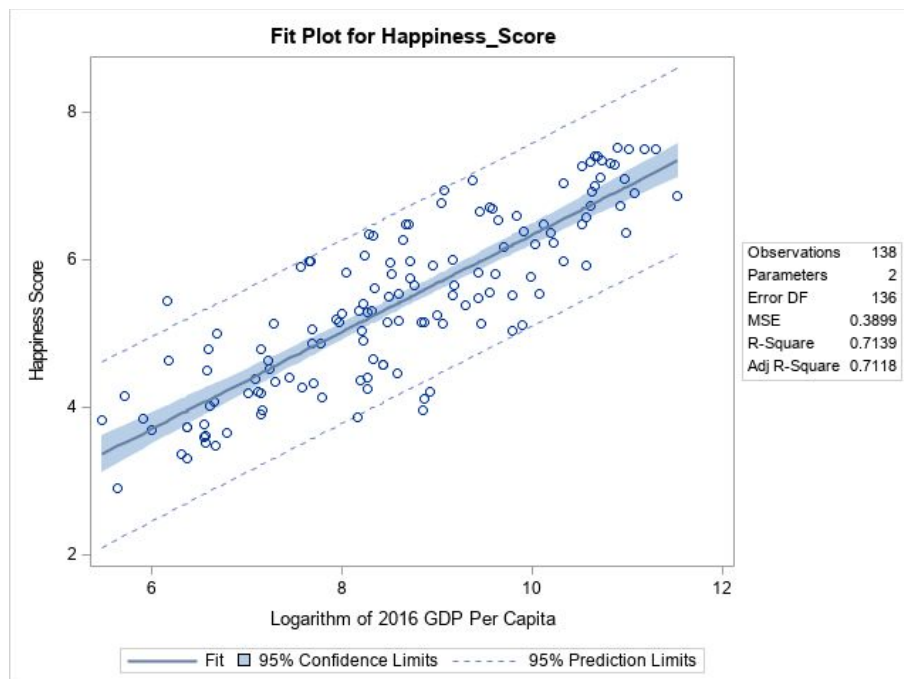
In analyzing the relationship between GDP per capita (USD) and Happiness scores, it is clear that higher GDPs per capita are positively correlated with Happiness Scores, however there are substantial diminishing returns as evidenced by *Figure 7*. In other words, the more money the country has per person, the happier they are, with substantial diminishing returns after about \$10,000 USD. Thus, a logarithmic regression model seemed most appropriate. *Figure 8* shows

the results of a logarithmic regression. Notably, the regression explains 71.39% of the variability (the results of the regression are included in the *Table A.3* in *Appendix A*).

**Figure 7:** Relationship between GDP per Capita and Happiness Score (2016)



**Figure 8:** Relationship between the Logarithm GDP per Capita and Happiness Score (2016)

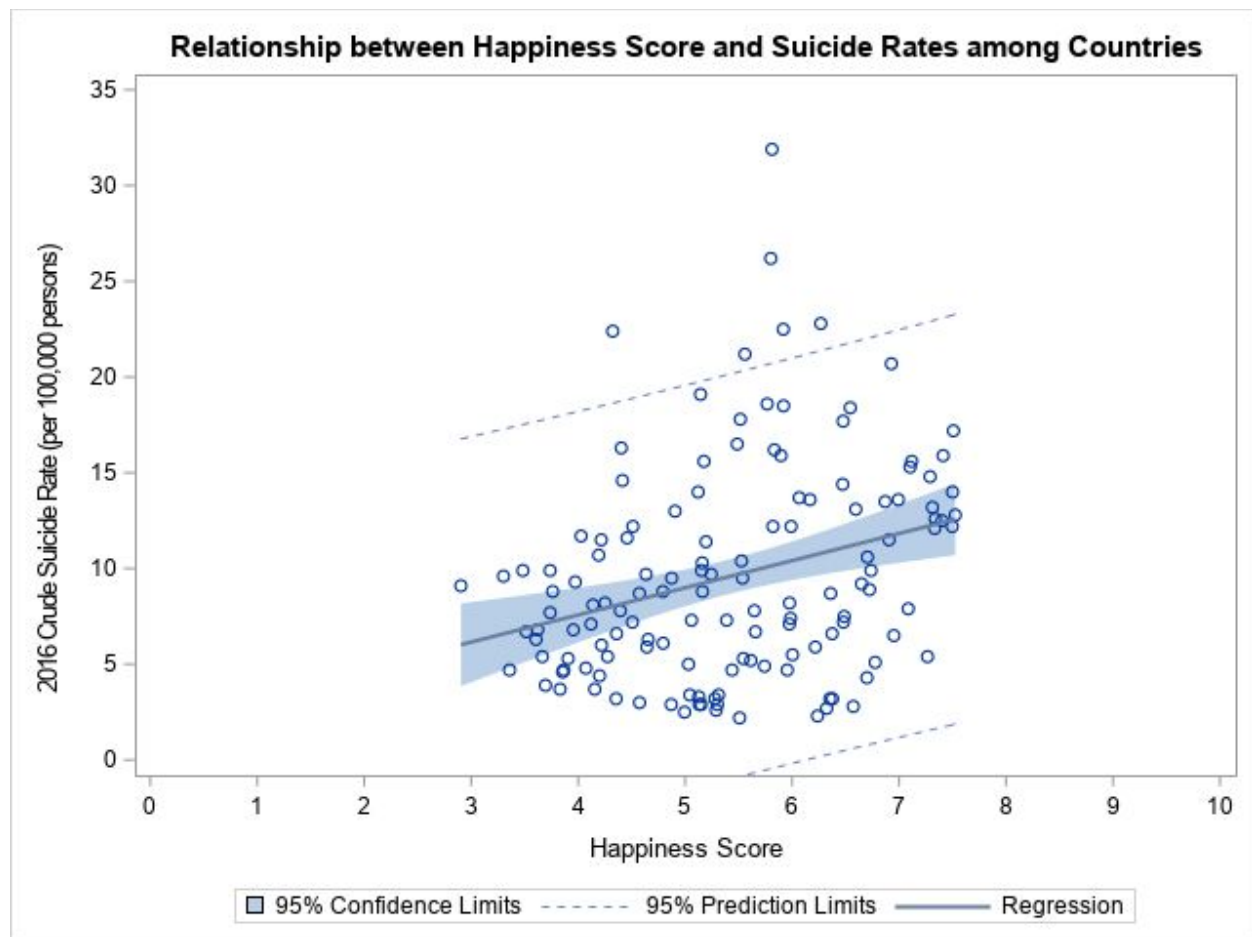


*Do happier countries have fewer suicides?*

The second predictor we looked at in an attempt to understand the effectiveness of happiness scores as a proxy for happiness was a country's crude suicide rate.

Counterintuitively, there is a weak positive correlation between happiness scores and their respective crude suicide rate (mortality rate not attempt rate). Although the regression only has an R-squared value of .081, there is enough evidence to conclude that there is a non-zero relationship between the Happiness and Suicide Rates as the coefficient has a p-value of .00004. This finding is significant in that it runs counter to judgment—if countries are truly happier, why do they have higher suicide rates? As always, correlation does not imply causation and there are numerous confounding factors including mental health treatment and societal views surrounding mental health that may influence the relationship that represent areas for future study. *Figure 9* shows the relationship between Happiness Scores and Crude Suicide Rates by country.

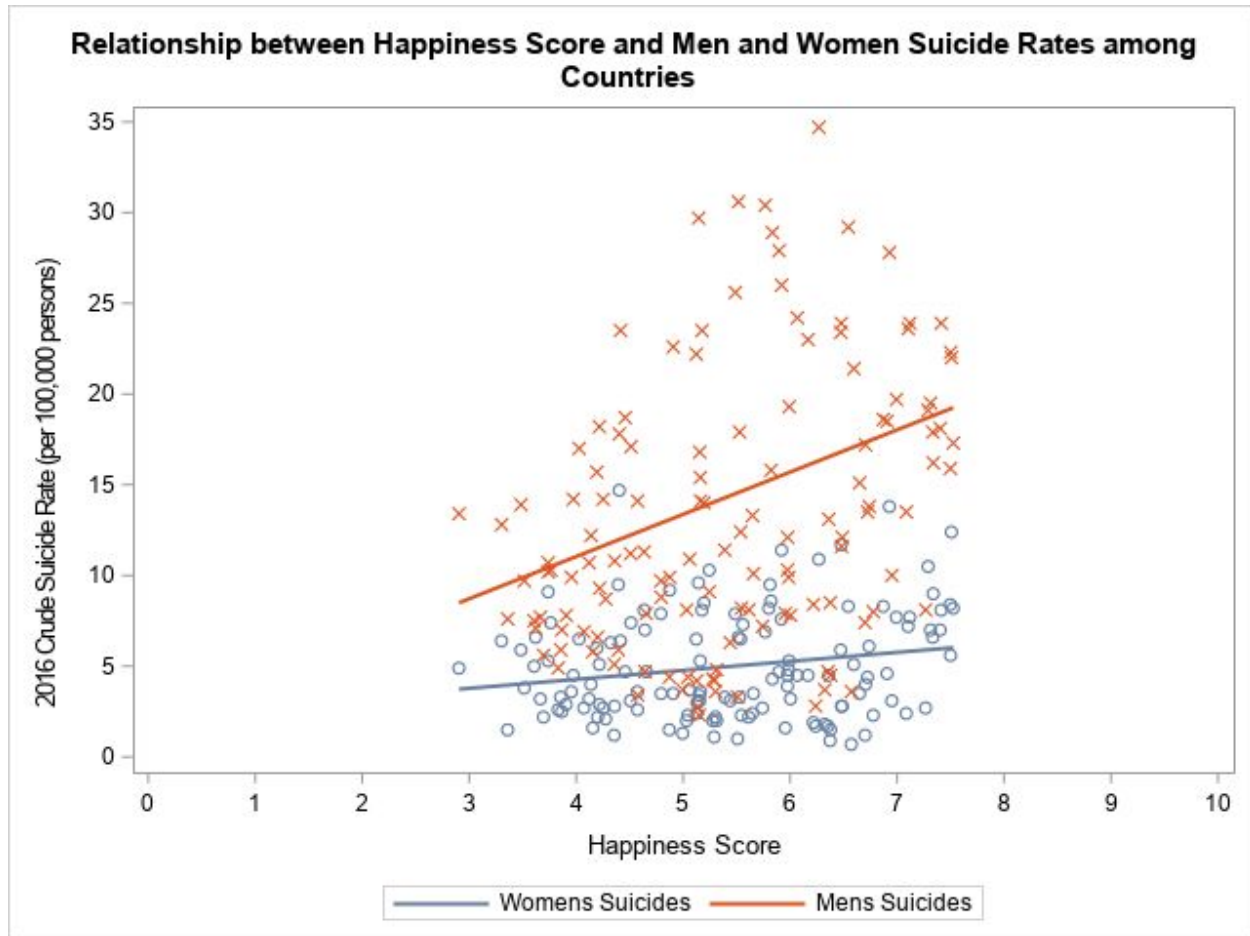
**Figure 9:** Relationship between Happiness Score and Crude Suicide Rate (2016)



In diving a little bit deeper, there is a difference between a country's overall happiness score and the crude suicide rate among men and women. *Figure 10* shows the differences between the

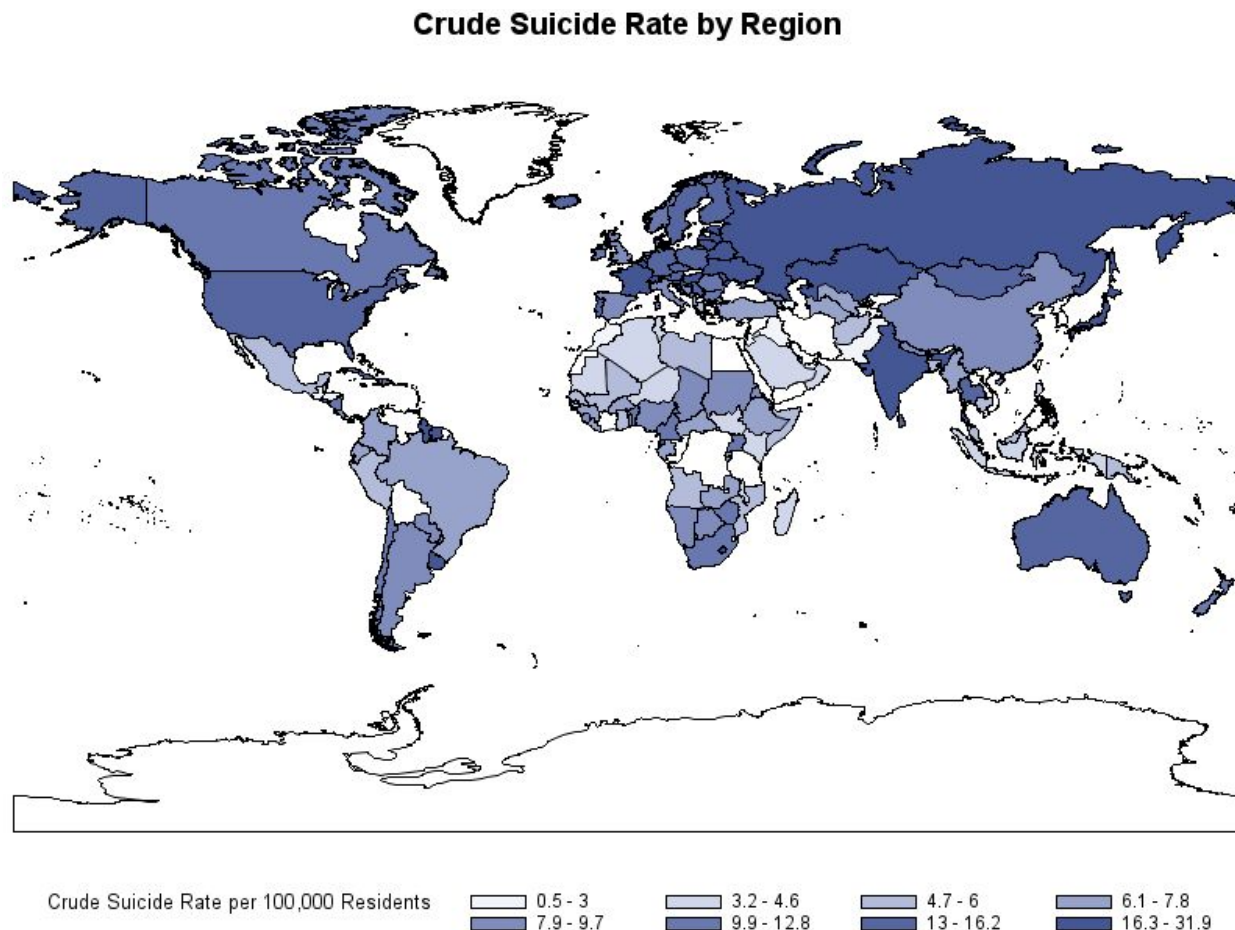
relationship among men and women. Notably, there is a much stronger positive relationship between a country's happiness score and their Male Crude Suicide Rate (deaths per 100,000 people). There is a weak positive relationship for women, but not nearly as strong as the relationship for men. This finding is interestingly, and does not necessarily yield an intuitive explanation and represents a potential area for future research.

**Figure 10:** Relationship between Happiness Score and Men's' and Women's Crude Suicide Rate (2016)



Naturally, these regressions hide complexity, and a geospatial analysis regarding suicide rates is also valuable. *Figure 11* shows the suicide rate by country in a choropleth global map. When compared with the map of happiness scores, this map reinforces the counterintuitive trend that happier countries have higher suicide rates.

*Figure 11: Crude Suicide Rate Choropleth Map (2016)*



*Are countries with higher freedom of expression happier?*

After looking at economic prosperity and suicide rates, we looked at freedom of expression to better understand the relationship between happiness scores and basic rights.

Overall, there is a positive correlation between a country's freedom of expression levels and their happiness score. Of all aspects related to freedom of expression, a country's regulation of media content has the strongest correlation with the happiness score. In other words, the less the country regulates their media, the happier the people are.

Interestingly, there is nearly no relationship between how many media people the country jails and the happiness score. This finding is interesting as jailing members of the media frequently represents a government attempting to limit information available to the public. Popular wisdom might imply a government that jails media members is more corrupt and people would be more likely to unhappy under a corrupt government. However, the data indicates this is not the case. *Table 2* shows the correlation matrix of different factors related to freedom of expression. When

examining the results, it is critical to keep in mind these are ecological relationships and may overestimate the true relationship.

**Table 2: Correlation Matrix of Freedom of Expression**

Pearson Correlation Coefficients						
	Regulation of Media Content	Media Jailed	Access to Cable	Access to Newspapers	Access to Internet	Happiness Score
Regulation of Media Content	1.00000	0.24819	0.51250	0.60755	0.65613	0.46733
Media Jailed	0.24819	1.00000	0.18953	0.15116	0.27645	-0.04121
Access to Cable	<b>0.51250</b>	0.18953	1.00000	0.68469	0.79938	0.31876
Access to Newspapers	<b>0.60755</b>	0.15116	<b>0.68469</b>	1.00000	0.66879	0.36593
Access to Internet	<b>0.65613</b>	0.27645	<b>0.79938</b>	0.66879	1.00000	0.33825
Happiness Score	<b>0.46733</b>	-0.04121	0.31876	0.36593	0.33825	1.00000

Overall, Australia and New Zealand, North America, and Western Europe have the highest levels of freedom of expression as ranked by the Human Freedom Index with expression scores all greater than 9 (on a scale from 1 to 10). Eastern Asia and Latin America and the Caribbean have high levels of expression with scores greater than 8 (see *Table A.4 in Appendix A*). The Middle East and Northern Africa as well as Southeastern Asia have the lowest freedom of expression. *Table 3.*

Interestingly, the correlation matrix indicates that the relationship between freedom of expression and happiness scores appears to be much less significant than the relationship between money



and happiness. In other words, the analysis indicates that money may matter more than freedom of expression in judging how happy a country is. Furthermore, a regression analysis of happiness scores against several freedom of expression indicators was conducted in *Table 3*. Even with five predictors, which may suggest potentially overfitting, we only achieve an R-squared value of 0.30.

**Table 3:** Multiple Linear Regression Model of Freedom of Expression Indicators and Happiness

### Regression: Happiness vs. Freedom of Expression

The REG Procedure

Model: MODEL1

Dependent Variable: HAPPINESS\_SCORE

Number of Observations Read	125
Number of Observations Used	125

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	49.51653	9.90331	10.28	<.0001
Error	119	114.63584	0.96333		
Corrected Total	124	164.15236			

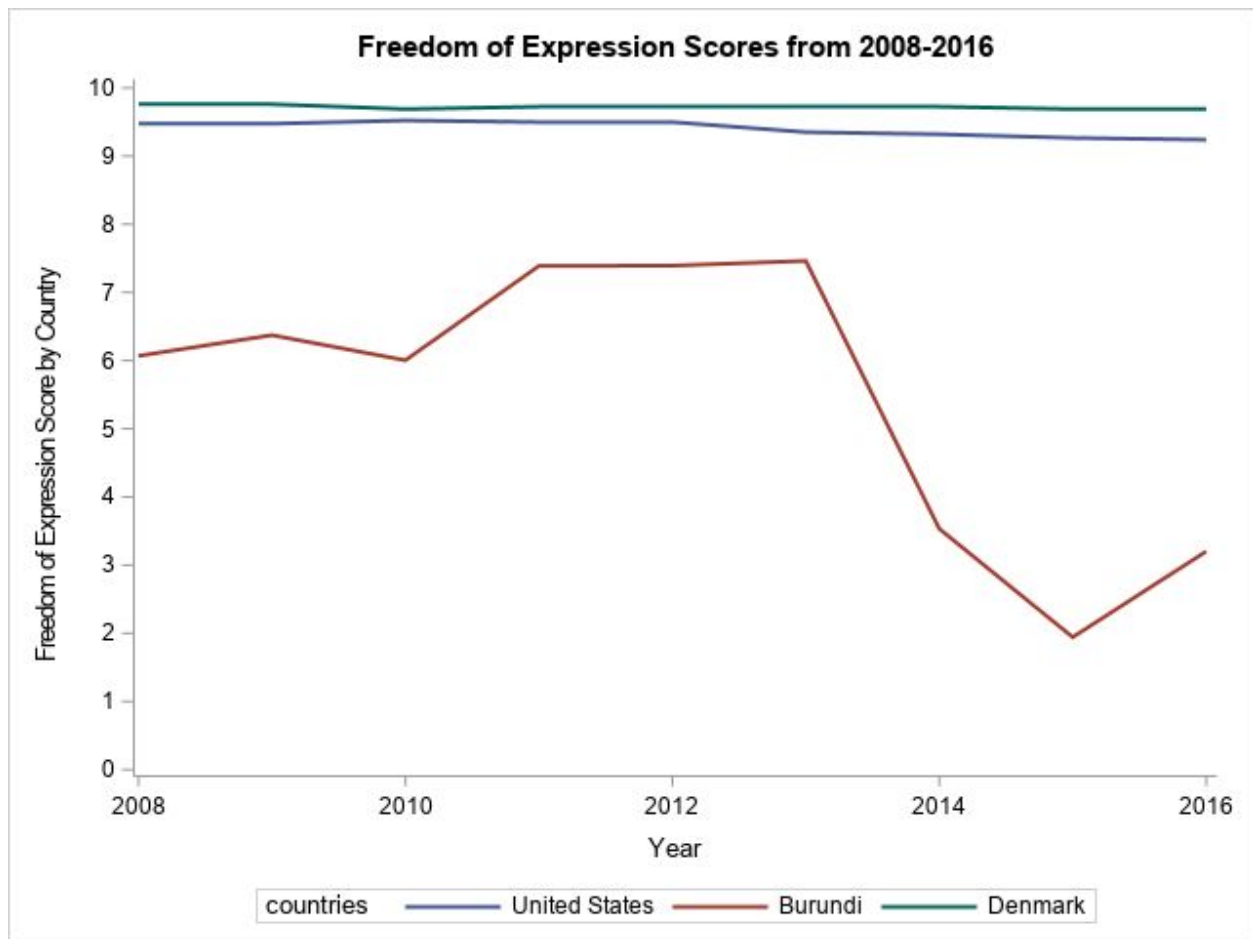
Root MSE	0.98149	R-Square	0.3016
Dependent Mean	5.47342	Adj R-Sq	0.2723
Coeff Var	17.93198		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	5.54384	0.80157	6.92	<.0001
PF_EXPRESSION	1	0.02490	0.22363	0.11	0.9115
PF_EXPRESSION_JAILED	1	-0.18951	0.06363	-2.98	0.0035
PF_EXPRESSION_INFLUENCE	1	0.06785	0.08789	0.77	0.4416
PF_EXPRESSION_INTERNET	1	0.05688	0.08449	0.67	0.5021
PF_EXPRESSION_CONTROL	1	0.14754	0.10842	1.36	0.1761



All of the analysis thus far has studied countries only for the 2016 year. It is worthwhile to do an intertemporal analysis of Expression Scores to better understand how scores may change overtime. *Figure 12* shows the freedom of expression scores in Denmark (the happiest country in 2016), the United States, and Burundi (the unhappiest country in 2016). The graph calls to attention how quickly a country's situation can change. Prior to 2013, the freedom of expression score in Burundi was close to 7.5. However, due to political and civil conflict related to President Pierre Nkurunziza's cling to power, it rapidly dropped beginning in 2014.

**Figure 12:** *Freedom of Expression in Denmark, The United States, and Burundi*

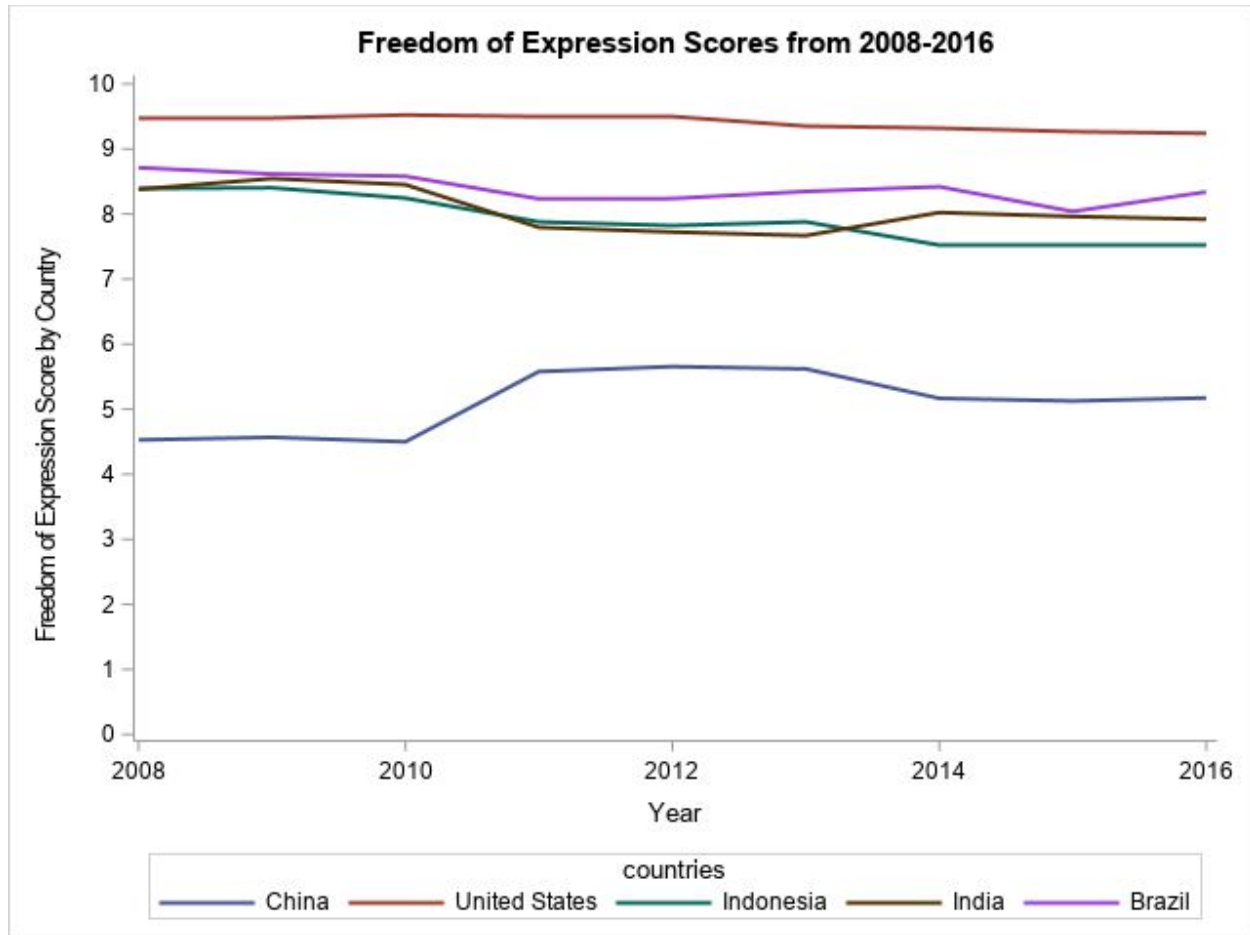


To extend the analysis regarding the drop in freedom of expression, we performed an intertemporal analysis of Freedom of Expression scores in Egypt, Syria, and Ukraine (all countries who have been plagued by conflict in recent years) in *Figure A.2* in *Appendix A*. This analysis showed that freedom of expression scores can drop very rapidly and substantially in a country as a consequence of changing environments.

Among the 5 most populous countries in the world (China, India, Indonesia, Brazil, and the United States), the freedom of expression levels have remained relatively consistent, with the only major being China experiencing a substantial jump in 2011 as shown in *Figure 13*. Of the

five most populous countries, the United States has the highest level of freedom of expression, China has the lowest level of freedom of expression, and India, Brazil, and Indonesia have relatively comparable levels of freedom of expression.

**Figure 13** *Freedom of Expression in the 5 Most Populous Countries*



*Which indicators of religious freedom relate to a country's happiness score?*

While there are an estimated 4,200 religions worldwide, not every country gives its citizens freedom to practice the religion(s) of their choice. From the Human Freedom Index dataset, there were several measures associated with religious freedom. Of these variables, we chose to investigate the relationship between a country's happiness score and a country's measurements for freedom to establish and operate religious organizations, harassment and physical hostilities associated with practicing religion, as well as overall religious freedom. *Table 4* displays the results from the multiple linear regression of happiness score and these freedom of religion factors.

**Table 4: Multiple Linear Regression Model of Freedom of Religion Factors and Happiness**

Regression: Happiness vs. Religious Freedom

The REG Procedure

Model: MODEL1

Dependent Variable: HAPPINESS\_SCORE

Number of Observations Read	125
Number of Observations Used	125

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	10.92764	3.64255	2.88	0.0390
Error	121	153.22472	1.26632		
Corrected Total	124	164.15236			

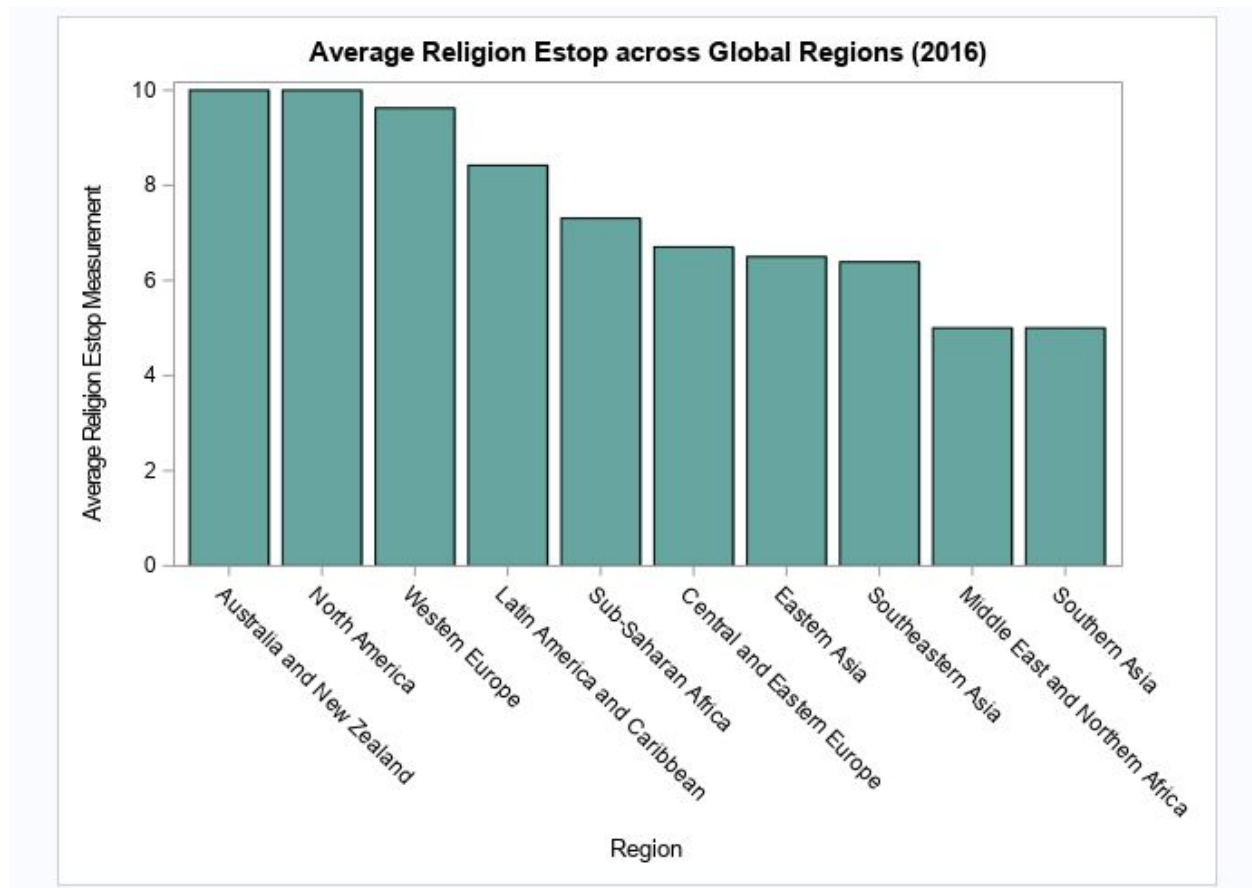
Root MSE	1.12531	R-Square	0.0666
Dependent Mean	5.47342	Adj R-Sq	0.0434
Coeff Var	20.55953		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	5.46831	0.89098	6.14	<.0001
PF_RELIGION_ESTOP	1	0.11369	0.03886	2.93	0.0041
PF_RELIGION_HARASSMENT	1	-0.04304	0.12785	-0.34	0.7370
PF_RELIGION	1	-0.07473	0.06903	-1.08	0.2811

With a R-squared value of 0.07, we conclude that the linear model does not adequately capture the relationship between these religious freedom measurements and happiness score. Nevertheless, of the three predictor variables in the model, the measurement for freedom to establish and operate religious organizations (*Religion Estop*) was the only statistically significant coefficient.

From these findings, we furthered our analysis to understand which regions had the strongest freedom to establish and operate religious organizations (*Religion Estop*) in *Figure 14*. Notably, the four regions with the highest average happiness scores also have the highest average freedom to establish/operate religions. Given this, it appears that the freedom to establish and operate religious organizations is related with happiness.

**Figure 14:** Average measurement for establishing & operating religious organizations by global region (2016)



*What variables associated with a country's legal system can be used to predict a country's happiness score?*

The next factor we decided to examine as predictors of happiness were countries' legal systems. In every country, there are people who break the law and the legal systems must intervene to take action. In a perfect world, legal systems behave justly. However, in reality, many legal systems fail to meet this expectation. We examined how well the measurements for the integrity of legal system, impartiality of courts, and reliability of the police serve as predictors for a country's happiness score in the multiple linear regression model in *Table 5*. Ultimately, the following regression analysis conveys that none of the three indicators are in fact statistically significant.

**Table 5: Multiple Linear Regression Model of Legal System Factors and Happiness****Regression: Happiness vs. Legal Systems**

The REG Procedure

Model: MODEL1

Dependent Variable: HAPPINESS\_SCORE

Number of Observations Read	114
Number of Observations Used	114

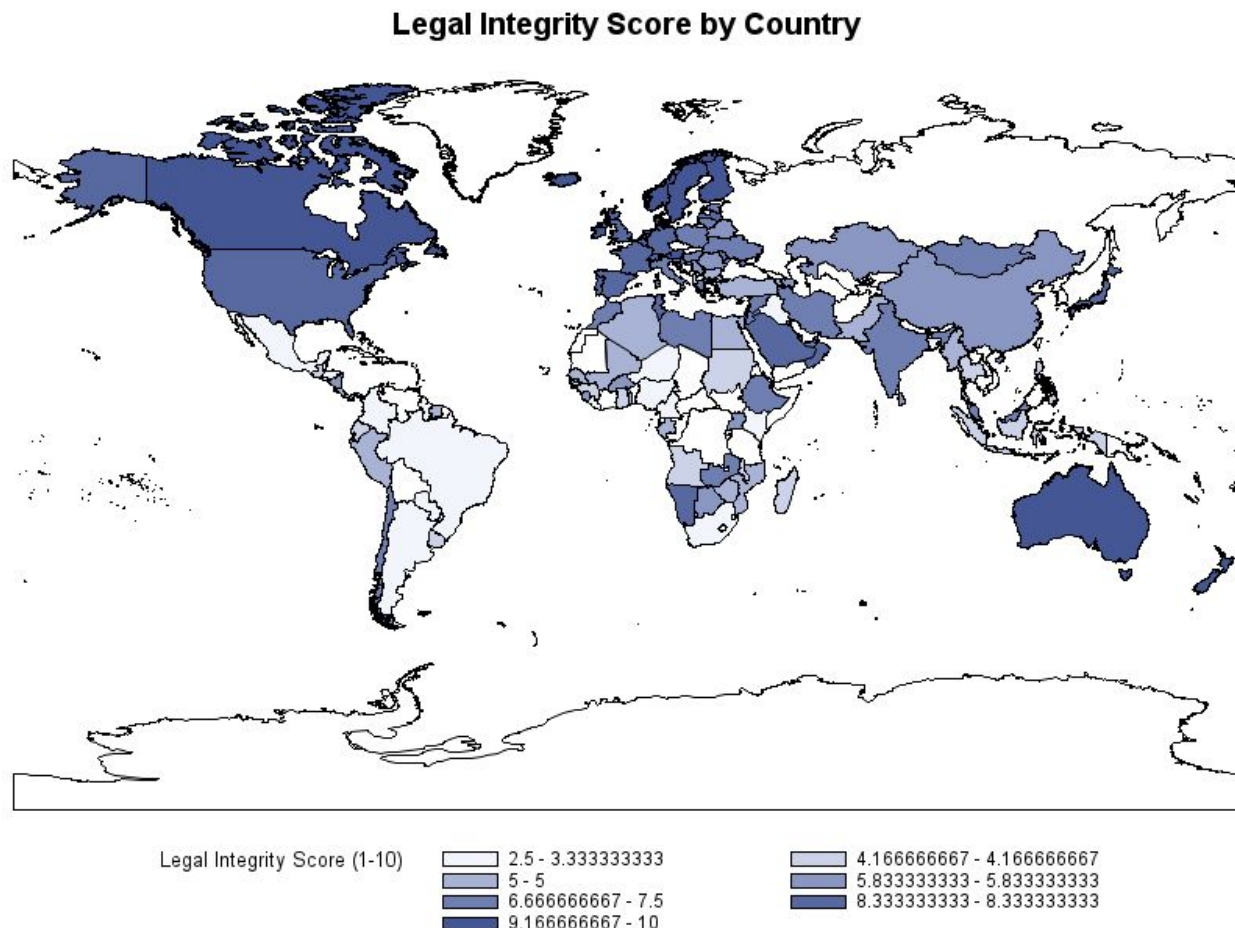
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	40.78237	13.59412	14.55	<.0001
Error	110	102.80256	0.93457		
Corrected Total	113	143.58493			

Root MSE	0.96673	R-Square	0.2840
Dependent Mean	5.58244	Adj R-Sq	0.2645
Coeff Var	17.31736		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	3.72658	0.29947	12.44	<.0001
EF_LEGAL_INTEGRITY	1	0.08350	0.06843	1.22	0.2250
EF_LEGAL_COURTS	1	0.07399	0.09237	0.80	0.4248
EF_LEGAL_POLICE	1	0.17482	0.09328	1.87	0.0636

Despite the relatively weak results from the regression, it is worthwhile to visualize the legal integrity of each country globally in *Figure 15*. South America's legal integrity scores are significantly lower than any other region. However, at the same time, South America's happiness scores are relatively high. In many regards, the South American region demonstrates that globally the relationship between legal systems and happiness scores is not as strong as we might assume.

*Figure 15: Legal Integrity Score by Country (2016)*



*How do women's rights issues affect the happiness level of a country?*

After examining legal system integrity, we transitioned our analysis to examining if women's rights in a country – that is, the rights of women to bodily autonomy, freedom from sexual violence, economic and legal equality, and an overall lack of discrimination based on gender – can be theorized to have a relationship with the happiness level of that country. Specifically, we assumed that women living in countries with stronger gender equality should be happier on average. To study the effect of women's rights on the happiness level of a country, certain variables that related to gender equality from the Human Freedom Index were used, as well as the happiness score variable from the World Happiness Report.

To be more detailed, we employed five variables related to women's rights issues to analyze this relationship – the strength of the women's movement, women's security and safety, women's inheritance rights, missing women, and female genital mutilation. Each variable is scored from 0 to 10, ranked by its level of freedom. So, for example, a 0 for female genital mutilation would mean that it occurs in that country, while a 10 would mean that it does not. Therefore, according



to our assumptions, a higher score for these variables related to gender equality should coincide with a higher happiness score.

The summary statistics in *Table 6* tell us that women's safety, missing women, and female genital mutilation variables have the highest scores on average, meaning that most countries have acceptable treatment of women with regards to these categories. Women's movement and women's inheritance rights have slightly lower average scores of 7.75 and 6.21 respectively. Additionally, these two variables have the largest standard deviations out of the categories. Based on the medians and means of the variables, women's inheritance rights seems to be skewed right, since the median is less than the mean. However, the four other variables all seem to be skewed left, although women's safety could be close to normally distributed, as the mean and median are close in value.

**Table 6: Summary Statistics for Women's Rights**

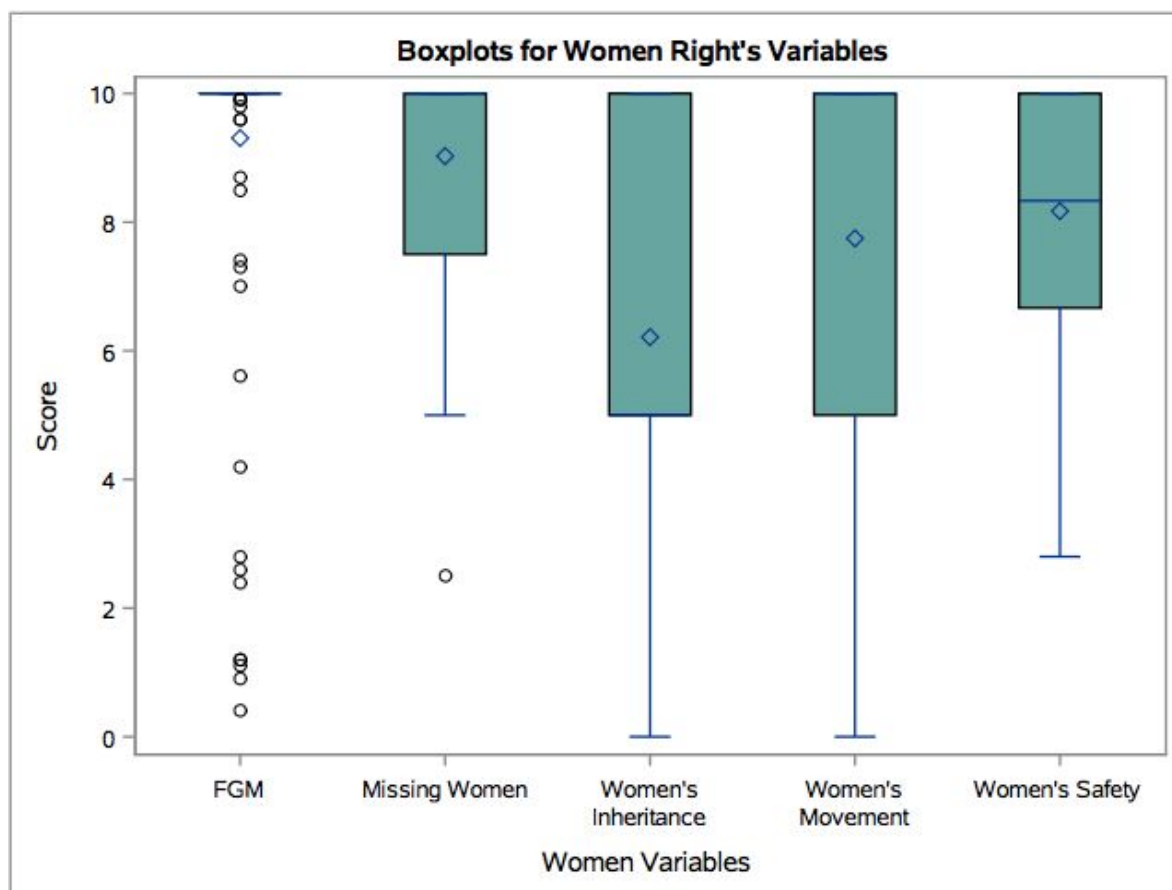
Women's Movement						Safety of Women						Women's Inheritance rights					
N	Median	Mean	Std	Min	Max	N	Median	Mean	Std	Min	Max	N	Median	Mean	Std	Min	Max
131	10.00	7.75	3.11	0.00	10.00	134	8.33	8.17	1.81	2.80	10.00	134	5.00	6.21	3.64	0.00	10.00

Missing Women						Female Genital Mutilation					
N	Median	Mean	Std	Min	Max	N	Median	Mean	Std	Min	Max
131	10.00	9.03	1.47	2.50	10.00	130	10.00	9.31	2.14	0.40	10.00

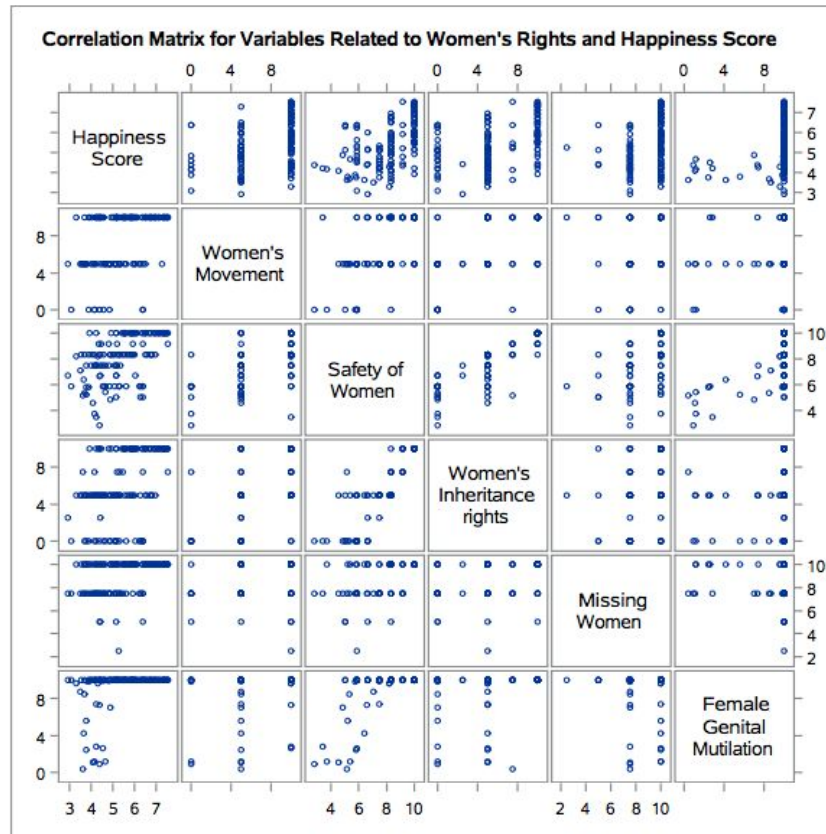
To dive deeper into the distributions, we examined the distribution of women's rights among different indicators in *Figure 16*. The boxplots of the women's rights variables slightly contrast the previous assumptions. For one, all of the variables seem to be skewed left, even though the smaller median of women's inheritance compared to its mean implied a right skew. Although women's safety still does not seem exactly normal, it is less skewed than the other variables. On the other hand, the female genital mutilation variable is almost entirely outliers, which can be explained by the fact that only a few countries allow this practice.

**Figure 16: Boxplots for Women's Rights**



The relationships between the women's rights variables can be surmised using correlation coefficients and scatterplots. A good portion of the correlation coefficients are above .5, with women's safety and women's inheritance having the largest correlation coefficient at .89550. It is not surprising that many of the variables are correlated, as all of them are related to women's rights issues. Looking at the correlation matrix in *Figure 17*, all the women's rights variables seem to be positively correlated - which is again not surprising. Although most of the women's rights variables seem to be correlated, female genital mutilation is not as strongly related as the other variables are, its correlation coefficients ranging from .05206 (with women missing) to .58030 (with women's safety). This trend is probably due to the rareness of this practice. Relatedly, women's safety is relatively correlated with each other women's rights variable, all being above .50. With regards to the relationship between the happiness score and the women's rights variables, the correlation coefficients range from .35 to .53, no pairs having significantly large or small correlations. The strongest correlation is between happiness score and women's safety, which can be seen in the scatterplot, depicting a somewhat positive linear relationship in comparison to the other more non-linear plots.



**Figure 17: Women's Rights Correlation Matrix**

Despite potential collinearity, we ran a regression of the happiness score on the five women's rights variables. The adjusted R squared for this regression was .2833. The F value, presented in *Table 7*, is highly significant with a p-value of less than .0001, implying that at least one of the women's rights variables have a significant effect on happiness level. Looking at the individual t values for the coefficients in *table 3*, only two of the coefficients are significant - the intercept and the coefficient for the women's safety variable. The latter coefficient implies that for an increase in 1 of the women's safety score, the happiness score will increase by .43736. In terms of issues with the model itself, the coefficient for women's inheritance seems wrong, as it was expected that gender equality and happiness would have a positive correlation. This could be due to possible multicollinearity.

**Table 7: ANOVA table**

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	54.06369	13.51592	13.85	<.0001
Error	126	122.97251	0.97597		
Corrected Total	130	177.03620			

**Table 8: Regression Parameter Estimates**

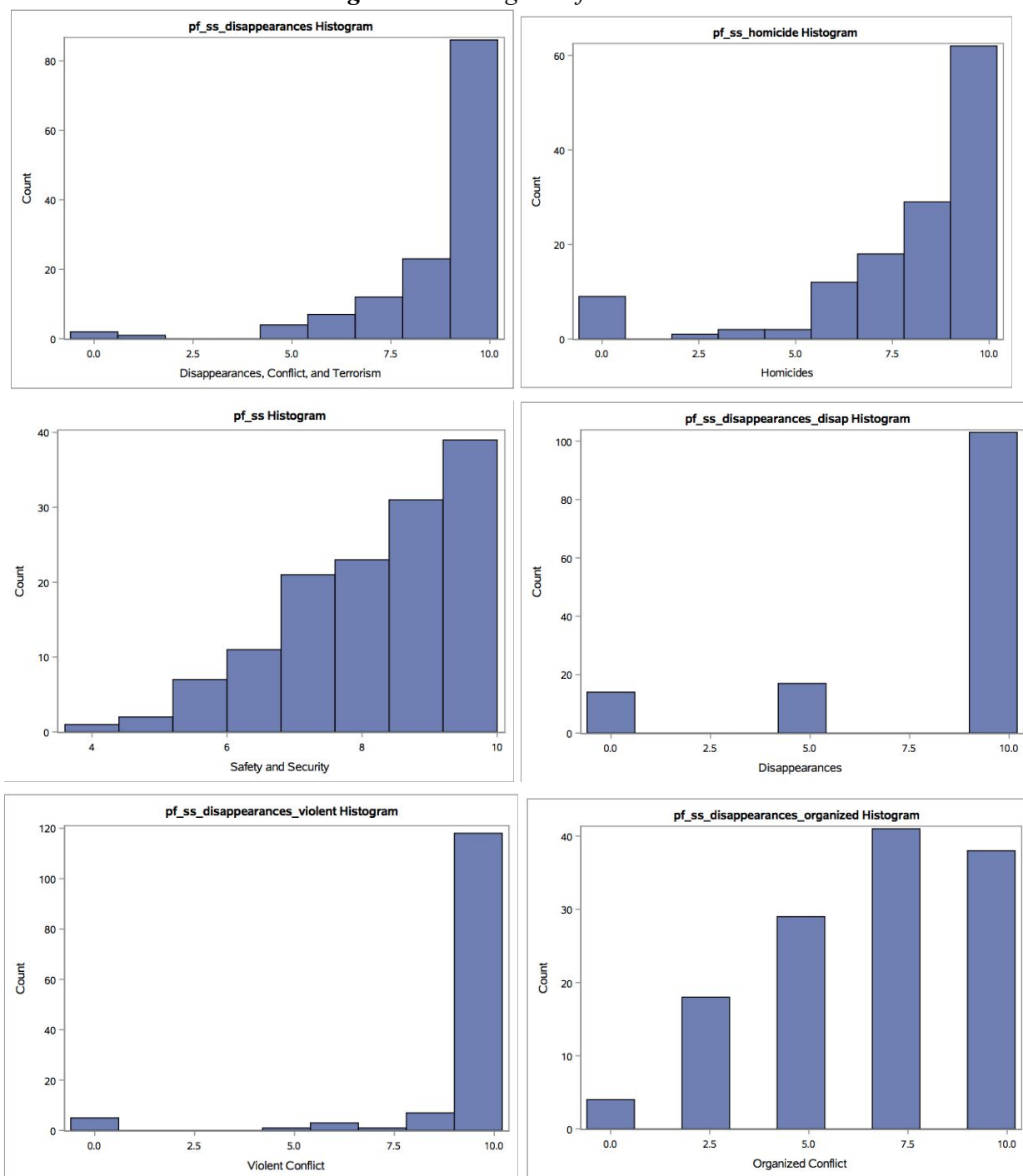
Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	1.86085	0.66935	2.78	0.0063
pf_movement_women	Women's Movement	1	0.02696	0.03704	0.73	0.4680
pf_ss_women	Safety of Women	1	0.43736	0.12839	3.41	0.0009
pf_ss_women_inheritance	Women's Inheritance rights	1	-0.06860	0.05605	-1.22	0.2233
pf_ss_women_missing	Missing Women	1	0.01989	0.07586	0.26	0.7936

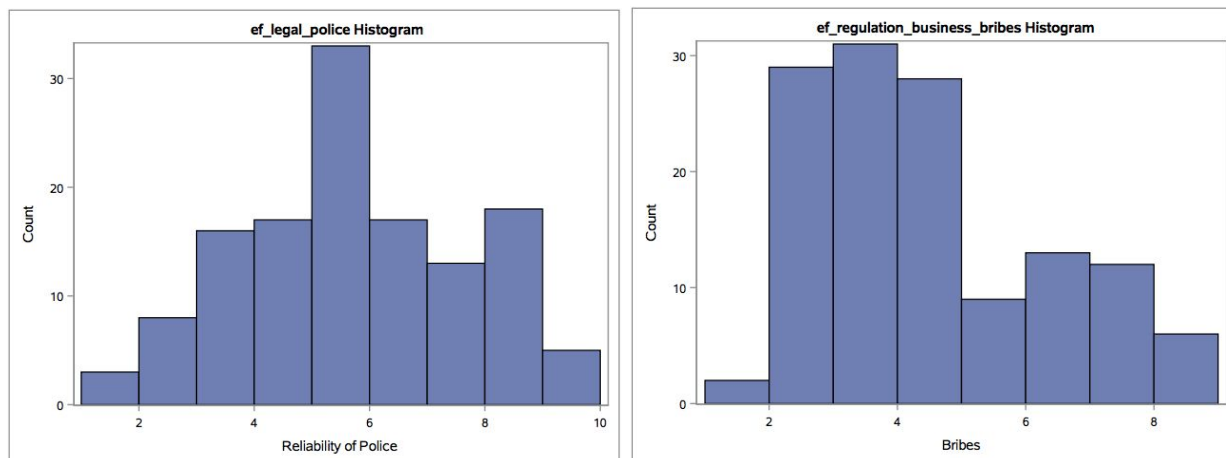
*Does the amount of crime in a country relate to its level of happiness?*

The amount of crime in a country is thought to have a negative relationship with happiness, as the safer a country is, the happier its inhabitants should be. To study this relationship, we used variables from the Human Freedom Index, focusing on those related to the crime or safety level of a country. The happiness score variable from the World Happiness Index is again the dependent variable. The crime variables used for the following analysis were homicides, disappearances, violent conflict, organized conflict, disappearances, conflicts, and terrorism, safety and security, reliability of police, and bribes. As with the previous variables, the score for a variable ranges from 0 to 10, with 10 representing the highest level of freedom. In the case of this analysis, a 10 represents the lowest amount of crime, or the highest level of safety.

The individual distributions of the crime variables can be observed from their histograms in *Figure 18*. Most of the variables are skewed left, having the majority of their scores above 5. Disappearances, conflict, and terrorism, homicides, disappearances, and violent conflict are all highly skewed, while safety and security is still skewed left but with a more even distribution of scores. Organized conflict is also slightly skewed left, but, unlike the previous variables, its center is around 7.5 rather than 10. With regards to these variables, thus, most countries seem to be relatively safe. Reliability of police looks to be relatively normal, with a center around 5. The variable bribes is slightly skewed right, which tells us that business bribes are common in many countries, especially when compared to the other crime variables.

**Figure 18: Histograms for Crime Variables**

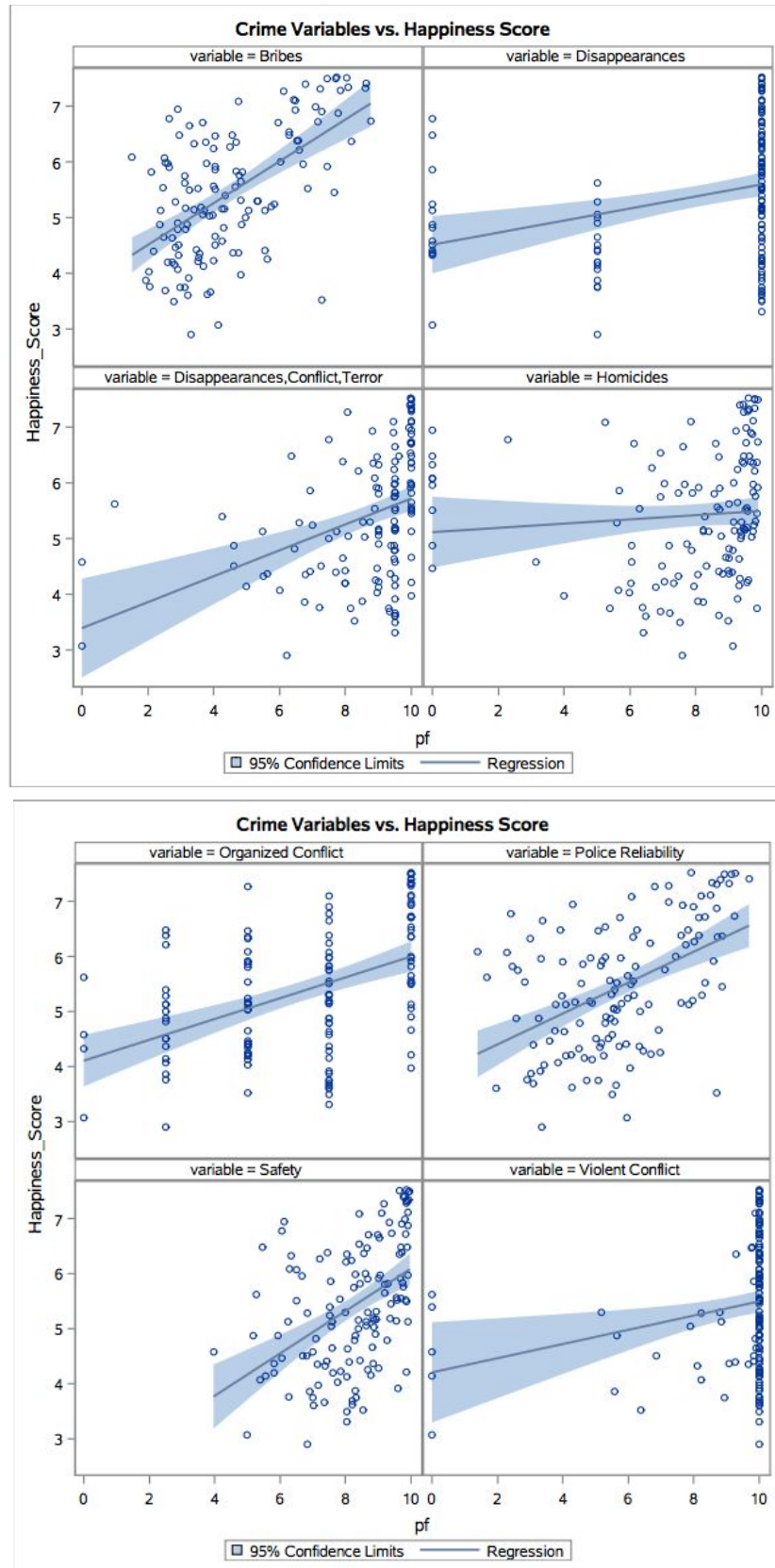




The relationship between variables can be observed by using the correlation coefficients, displayed in *Table A.5* in *Appendix A*. While a lot of the variables are strongly correlated with each other, the largest correlation coefficients are for disappearances, conflicts, and terrorism and organized conflict, with a correlation coefficient of .8222; and reliability of police and bribes, with a correlation coefficient of .85513. There are many other variable pairs with a correlation coefficient greater than .50, meaning that a good amount of the crime variables are correlated. This trend is expected, as all the variables are related to crime or safety. For example, reliability of police and safety have a correlation coefficient of .60840, which can be explained by if people are able to trust law enforcement, then they will likely feel safer.

To identify the relationship between the happiness score and crime variables, we used both correlation coefficients and scatter plots (shown in *Figure 19*). Based on the correlation coefficient, bribes has the strongest relationship with the happiness score, having a value of .58747 with a significant p-value. The scatter plot supports this interpretation, as we can see that there is a relatively strong positive linear relationship. Reliability of police also has a positive linear relationship with happiness score, which is why the correlation coefficient is the second largest of all the variables. Based on the plots, homicides and violent conflict seem to have the weakest relationships with happiness score, as the violent conflict points are mostly clustered around 10, and the homicide points seem to be relatively random. The correlation coefficient for homicide and happiness score is in fact very small, with a value of .08460 and a non-significant p-value. The correlation coefficient for violent conflict and happiness score is the second smallest value of all the variables, but it is still relatively large compared to homicide's correlation coefficient. Because of the nature of the data and risk of ecological correlations, these relationships may be even weaker than the results suggest.

**Figure 19 : Scatterplots of Happiness Score vs. Crime Variables**



To finish up the analysis of crime and happiness level, we performed a multiple linear regression. The adjusted R-squared for the linear model is .4931, and the F value is 16.08, with a highly significant p-value. The results are shown in *Table 9* and *Table 10*. By looking at the p-values for the regression coefficients, most of the variables have a significant effect on happiness score, but disappearances, violent conflict, organized conflict, and reliability of police do not. Notably, reliability of police has a very large p-value of .9094. The coefficient for safety and security implies that for an increase in 1 in its score, the happiness score increase by .69693. An increase in one for the bribes score, on the other hand, should increase the happiness score by .36003. Some of the coefficient estimates seem to have the wrong sign, such as homicides and disappearances, conflicts, and terrorism, which may be due to possible multicollinearity in the crime variables. Overall, the model seems to fit the data well, but it could definitely improved, specifically by accounting for the multicollinearity present in the variables.

**Table 9: Multiple Linear Regression ANOVA table**

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	86.96539	10.87067	16.08	<.0001
Error	116	78.43516	0.67617		
Corrected Total	124	165.40055			

**Table 10: Regression Parameter Estimates**

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	2.44670	0.66068	3.70	0.0003
pf_ss_homicide	Homicides	1	-0.29949	0.05742	-5.22	<.0001
pf_ss_disappearances	Disappearances, Conflict, and Terrorism	1	-0.59752	0.25795	-2.32	0.0223
pf_ss	Safety and Security	1	0.69693	0.15450	4.51	<.0001
pf_ss_disappearances_disap	Disappearances	1	0.07550	0.05744	1.31	0.1913
pf_ss_disappearances_violent	Violent Conflict	1	0.20277	0.12664	1.60	0.1121
pf_ss_disappearances_organized	Organized Conflict	1	0.10031	0.06996	1.43	0.1543
ef_legal_police	Reliability of Police	1	-0.01003	0.08792	-0.11	0.9094
ef_regulation_business_bribes	Bribes	1	0.36003	0.08958	4.02	0.0001



## *Conclusion*

The project looking at how different predictors related to economic prosperity, basic freedoms, and legal and judicial integrity can be used to better understand a country's happiness level. For all analyses, the project utilized methods in the SAS software. Key limitations of our study are: (1) limited time period (an intertemporal analysis may have changed findings) and (2) aggregated data may suggest stronger relationships that exist in reality. In examining overall happiness, noteworthy findings included: (1) Scandinavian countries are happiest, (2) countries plagued by civil and political conflict tend to be the unhappiest, (3) happiness scores are normally-distributed globally, and (4) there are significant regional differences. In looking at what factors are the best predictors of happiness, we found that: (1) GDP per capita is the best predictor of happiness, but money has significant diminishing returns after a threshold of ~\$10,000, (2) happier countries have more suicides than unhappier countries, (3) there is nearly no relationship between happiness levels and legal integrity in countries with the South American region being a significant outlier with low legal integrity scores correlating with high levels of happiness, and (4) most predictors studied were poor predictors of happiness. Overall, the study shows that happiness is both complicated to define, measure, and predict.

## Appendix A

**Table A.1:** ANOVA Difference between Happiness Scores among Regions

The ANOVA Procedure					
Dependent Variable: Happiness_Score					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
<b>Model</b>	9	126.6914624	14.0768292	27.00	<.0001
<b>Error</b>	147	76.6418112	0.5213729		
<b>Corrected Total</b>	156	203.3332736			

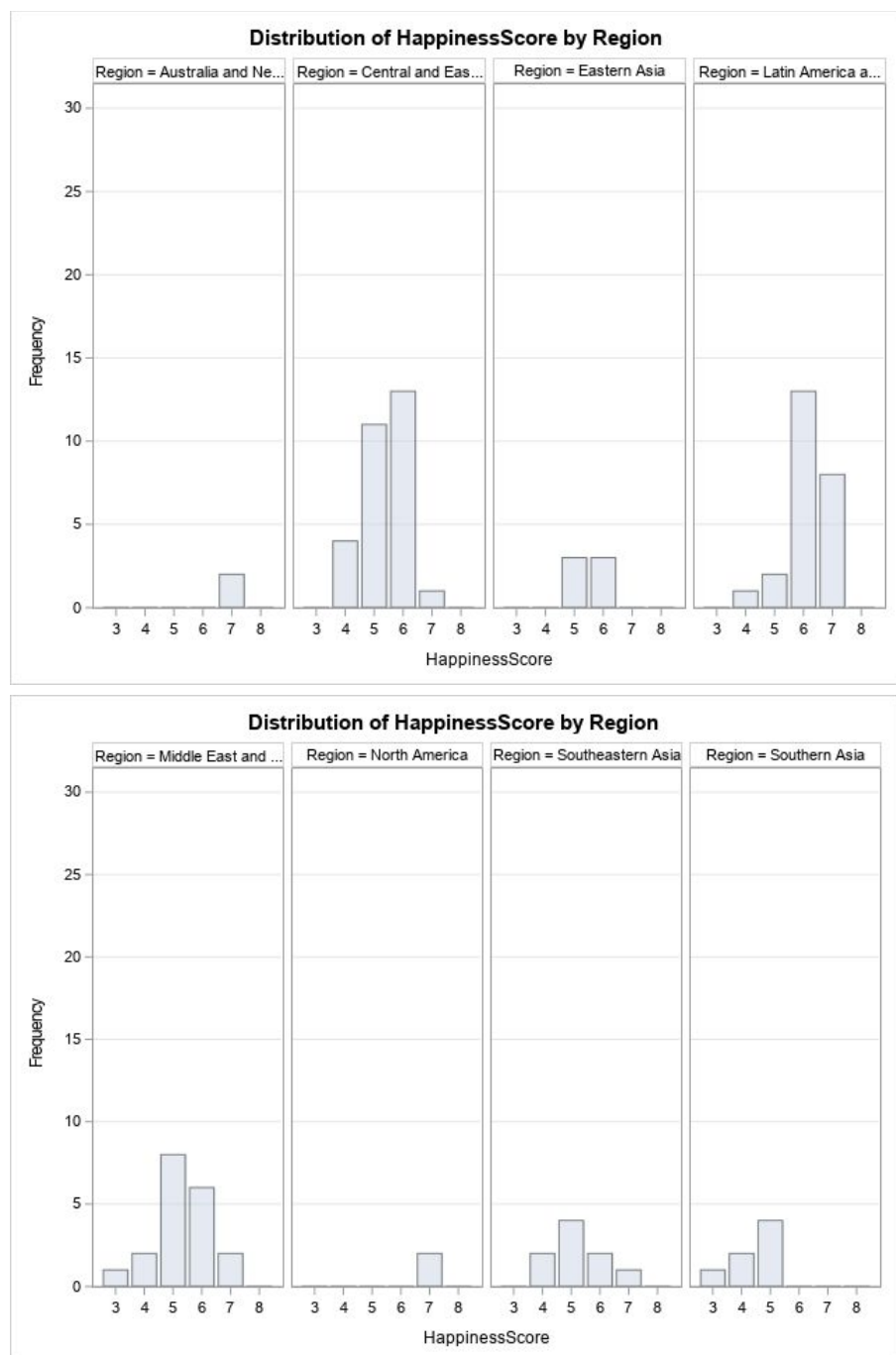
R-Square	Coeff Var	Root MSE	Happiness_Score Mean
0.623073	13.41577	0.722062	5.382185

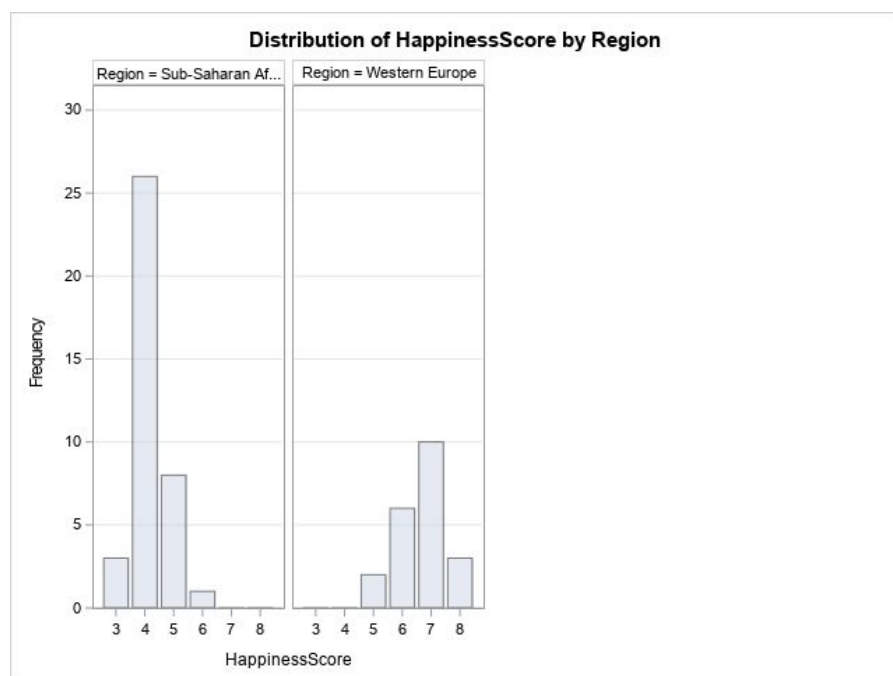
  

Source	DF	Anova SS	Mean Square	F Value	Pr > F
<b>Region</b>	9	126.6914624	14.0768292	27.00	<.0001



**Figure A.1: Happiness Score by Region Histograms**





**Table A.2:** Goodness of Fit for Normal Distribution of Happiness Scores

**The UNIVARIATE Procedure**  
**Fitted Normal Distribution for Happiness\_Score (Happiness Score)**

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	5.382185
Std Dev	Sigma	1.141674

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.06269464	Pr > D	0.134
Cramer-von Mises	W-Sq	0.09764382	Pr > W-Sq	0.123
Anderson-Darling	A-Sq	0.74582436	Pr > A-Sq	0.051

**Table A.3:** Happiness Score and Logarithm GDP Regression Analysis

Root MSE	0.62441	R-Square	0.7139
----------	---------	----------	--------

<b>Dependent Mean</b>	5.43592	<b>Adj R-Sq</b>	0.7118
<b>Coeff Var</b>	11.48667		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
<b>Intercept</b>	Intercept	1	-0.23054	0.31213	-0.74	0.4614
<b>Log_GDP</b>	Logarithm of 2016 GDP Per Capita	1	0.65743	0.03569	18.42	<.0001

**Table A.3:** Happiness Score and Suicide Rate Regression Analysis

Suicide Rate=Happiness Score

<b>Root MSE</b>	5.32663	<b>R-Square</b>	0.0881
<b>Dependent Mean</b>	9.59191	<b>Adj R-Sq</b>	0.0813
<b>Coeff Var</b>	55.53247		

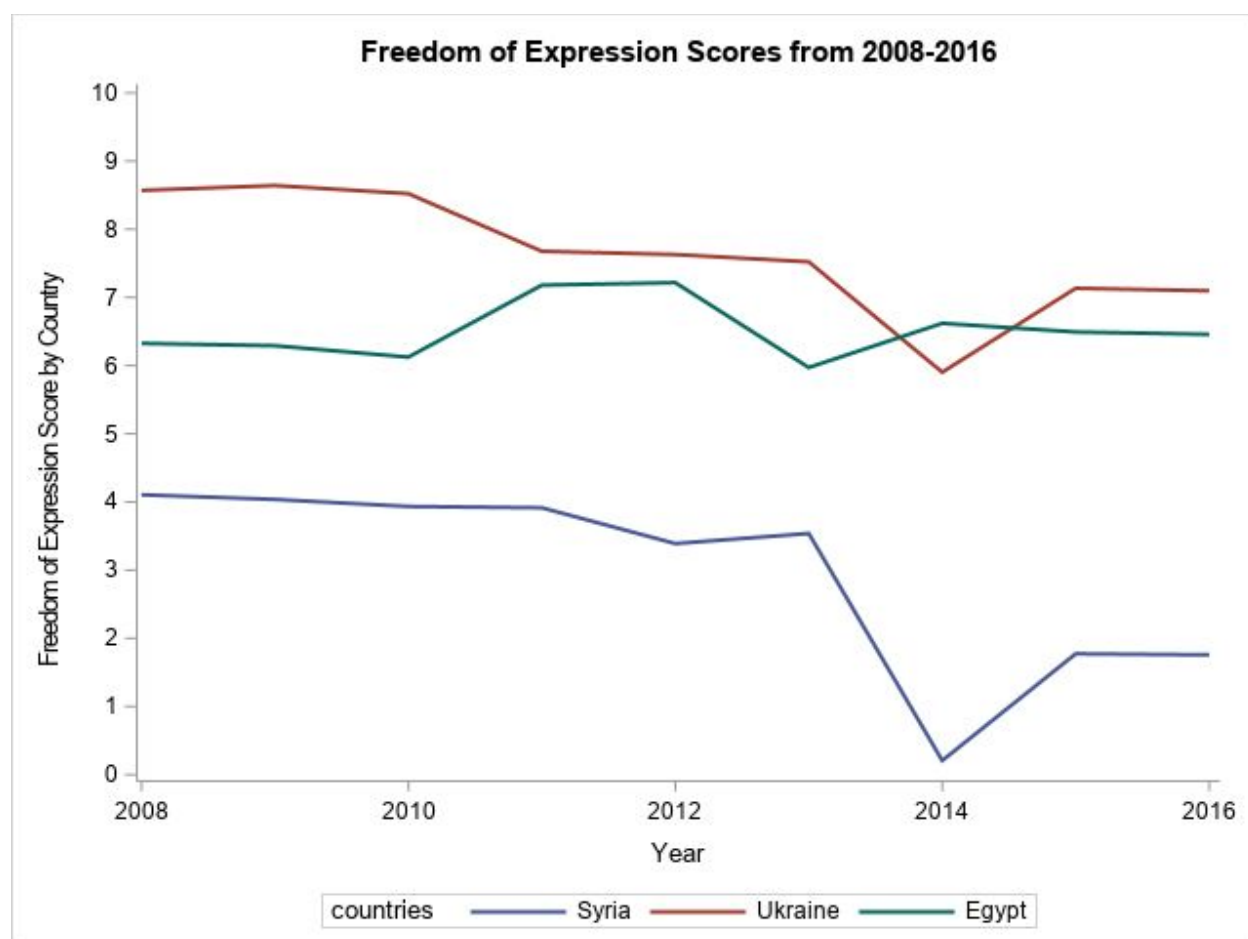
Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
<b>Intercept</b>	Intercept	1	1.90055	2.18552	0.87	0.3861
<b>Happiness_Score</b>	Happiness Score	1	1.41792	0.39401	3.60	0.0004

**Table A.4:** Summary Statistics for Freedom Of Expression among Regions

The MEANS Procedure						
Analysis Variable : pf_expression						
Region	N Obs	N	Mean	Std Dev	Minimum	Maximum
Australia and New Zealand	2	2	9.4583333	0.0925973	9.3928571	9.5238095
Central and Eastern Europe	29	23	7.7392356	1.6219310	4.2738095	9.6190476
Eastern Asia	6	5	8.1111725	1.6949015	5.1749101	9.2857143

Latin America and Caribbean	24	22	8.201328 1	0.988697 4	5.109508 4	9.595238 1
Middle East and Northern Africa	19	17	6.107929 8	1.736536 4	1.758817 9	8.297619 0
North America	2	2	9.376704 1	0.191202 6	9.241503 5	9.511904 8
Southeastern Asia	9	9	6.984434 8	0.901668 1	5.761904 8	8.204248 4
Southern Asia	7	6	7.204905 2	0.402077 2	6.702381 0	7.921169 5
Sub-Saharan Africa	38	31	7.256174 0	1.346023 3	3.202381 0	9.357142 9
Western Europe	21	20	9.475000 0	0.272427 6	8.738095 2	9.797619 0

**Figure A.2:** Syria, Ukraine, and Egypt Freedom of Expression Scores 2008-2016



*Table A.5: Summary Statistics for Freedom Of Expression among Regions*

Pearson Correlation Coefficients Prob >  r  under H0: Rho=0 Number of Observations				
	Happiness_Score	pf_ss_homicide	pf_ss_disappearances	pf_ss
Happiness_Score	1.00000 157	0.08460 0.3293 135	0.37151 <.0001 135	0.45382 <.0001 135
pf_ss_homicide Homicides	0.08460 0.3293 135	1.00000 135	0.07665 0.3769 135	0.63969 <.0001 135
pf_ss_disappearances Disappearances, Conflict, and Terrorism	0.37151 <.0001 135	0.07665 0.3769 135	1.00000 135	0.71195 <.0001 135
pf_ss Safety and Security	0.45382 <.0001 135	0.63969 <.0001 135	0.71195 <.0001 135	1.00000 135
pf_ss_disappearances_disap Disappearances	0.30831 0.0003 134	0.12303 0.1567 134	0.75234 <.0001 134	0.59923 <.0001 134
pf_ss_disappearances_violent Violent Conflict	0.22673 0.0082 135	-0.01655 0.8489 135	0.79665 <.0001 135	0.48224 <.0001 135
pf_ss_disappearances_organized Organized Conflict	0.45816 <.0001 130	0.16219 0.0652 130	0.82222 <.0001 130	0.71282 <.0001 130
ef_legal_police Reliability of Police	0.48404 <.0001 130	0.56061 <.0001 130	0.34523 <.0001 130	0.60840 <.0001 130
ef_regulation_business_bribes Bribes	0.58747 <.0001 130	0.40622 <.0001 130	0.31553 0.0003 130	0.52939 <.0001 130

Pearson Correlation Coefficients Prob >  r  under H0: Rho=0 Number of Observations		
	pf_ss_disappearances_disap	pf_ss_disappearances_violent
Happiness_Score	0.30831 0.0003 134	0.22673 0.0082 135
pf_ss_homicide Homicides	0.12303 0.1567 134	-0.01655 0.8489 135
pf_ss_disappearances Disappearances, Conflict, and Terrorism	0.75234 <.0001 134	0.79665 <.0001 135
pf_ss Safety and Security	0.59923 <.0001 134	0.48224 <.0001 135
pf_ss_disappearances_disap Disappearances	1.00000  134	0.34671 <.0001 134
pf_ss_disappearances_violent Violent Conflict	0.34671 <.0001 134	1.00000  135
pf_ss_disappearances_organized Organized Conflict	0.58346 <.0001 130	0.51542 <.0001 130
ef_legal_police Reliability of Police	0.29357 0.0007 129	0.12213 0.1663 130
ef_regulation_business_bribes Bribes	0.27809 0.0014 129	0.08921 0.3128 130

Pearson Correlation Coefficients Prob >  r  under H0: Rho=0 Number of Observations			
	pf_ss_disappearances_organized	ef_legal_police	ef_regulation_business_bribes
Happiness_Score	0.45816 <.0001 130	0.48404 <.0001 130	0.58747 <.0001 130
pf_ss_homicide Homicides	0.16219 0.0652 130	0.56061 <.0001 130	0.40622 <.0001 130
pf_ss_disappearances Disappearances, Conflict, and Terrorism	0.82222 <.0001 130	0.34523 <.0001 130	0.31553 0.0003 130
pf_ss Safety and Security	0.71282 <.0001 130	0.60840 <.0001 130	0.52939 <.0001 130
pf_ss_disappearances_disap Disappearances	0.58346 <.0001 130	0.29357 0.0007 129	0.27809 0.0014 129
pf_ss_disappearances_violent Violent Conflict	0.51542 <.0001 130	0.12213 0.1663 130	0.08921 0.3128 130
pf_ss_disappearances_organized Organized Conflict	1.00000  130	0.44847 <.0001 125	0.43625 <.0001 125
ef_legal_police Reliability of Police	0.44847 <.0001 125	1.00000  130	0.85513 <.0001 130
ef_regulation_business_bribes Bribes	0.43625 <.0001 125	0.85513 <.0001 130	1.00000  130

### Sources

“GDP per Capita (Current US\$).” *Data*, World Bank.

[data.worldbank.org/indicator/NY.GDP.PCAP.CD](https://data.worldbank.org/indicator/NY.GDP.PCAP.CD).

Schneider, Guillermina Sutter. “The Human Freedom Index.” *Kaggle*, 17 Dec. 2018,

[www.kaggle.com/gsutters/the-human-freedom-index](https://www.kaggle.com/gsutters/the-human-freedom-index).

“Suicide Mortality Rate (per 100,000 Population).” *Data*, World Bank,

[data.worldbank.org/indicator/SH.STA.SUIC.P5](https://data.worldbank.org/indicator/SH.STA.SUIC.P5).

Sustainable Development Solutions Network. “World Happiness Report.” *Kaggle*, 14 June 2017,

[www.kaggle.com/unsdsn/world-happiness](https://www.kaggle.com/unsdsn/world-happiness).