Civil Liberties, Economic Indicators, and Happiness: Predicting a Country's Happiness Score

Andrew J. Trick

Southern New Hampshire University

MAT-300: Applied Statistics II: Regression Analysis

Professor Atwood

February 17, 2017

Civil Liberties, Economic Indicators, and Happiness: Predicting a Country's Happiness Score

The World Happiness Report, published by the United Nation's Sustainable Development Solutions Network, is an annual survey whose findings reflect the state of happiness of people, split by country, throughout the world. Happiness is rated on a 0-10 scale in which countries can be evaluated and compared to one another. These ratings are created by considering the economic, social, health, and democratic factors within a country and normalizing, typically into a percentage of population or, sometimes, a more subjective 0-1 rating. While this happiness score found by the United Nations to quantify well-being is obviously strongly correlated to the independent variables they used to quantify it, only the final happiness scores were used in the dataset for this project. The primary goal of this analysis is to determine if there is any kind of relationship between a country's happiness index with other specific key factors chosen, and if this relationship can allow for future prediction of a country's happiness level.

Hopes and Reasons for Choice

Moving away from the factors that contributed to the initial creation of the happiness score and bringing in outside variables to the project, it is hoped that a model will be determined with which to predict the future happiness of a country. Being able to put a concrete weight to how much gender disparity or urban percentage affect the general happiness of a country, for example, could lead to a new way of looking at investments or decisions made within the government. The importance of this can be seen in some countries already: Bhutan annually tracks it's 'gross national happiness' and makes decisions based on it, while the U.K. recently started a 'measuring national well-being' program (Zhong, 2015). Although some of these predictor variables chosen may be a stretch to relate to overall country happiness, it is a hope that

they are a generalized sampling of qualities of freedom and well-being in a country, and that they will each have distinct relationships with the dependent variable of happiness.

Data Descriptions and Collection Methods

Data for this analysis was obtained on country statistics that were expected to have some relationship with happiness level. Factors were chosen to give a sampling of a country's overall economic freedom, democratic values, and egalitarianism.

Quantitative predictors

Quantitative values include 'GDP', 'Life Expectancy", 'Births per 1000', 'Infant Mortality per 1000', 'Urban Percentage', 'Female Workforce Percent', and 'Inequality gini'. GDP is the gross domestic product per capita. Life expectancy is the average healthy expected life span of a person from the country in years. Births per 1000 is the birthrate of the country per 1000 people, while infant mortality per 1000 is the death rate of children under the age of one per 1000 births. Urban percentage is the percent of the country's population that lives within an urbanized area and female workplace percent is the percentage of the workforce in a country that is female. Finally, inequality gini displays the wealth distribution of a country, where 0 equates to perfect equality and 1 would represent a vast disproportion where the overwhelming majority of wealth is in the hands of a select few.

Qualitative predictors

Variables 'democracy' and 'LGBT freedom' represent the two qualitative fields in the project. Democracy is split categorically into varying levels of democratic freedom of a country. Levels in this field, in descending order of freedom, are: full, flawed, hybrid, and authoritarian. LGBT freedom represents the amount of laws passed in a country which protect the civil rights

of lesbian, gay, bisexual, and transgender people. The levels for this variable are: high, some, and low. These levels were primarily determined based on marriage, adoption, military service, anti-discrimination, and gender identity laws a country has passed.

How the Data was Obtained

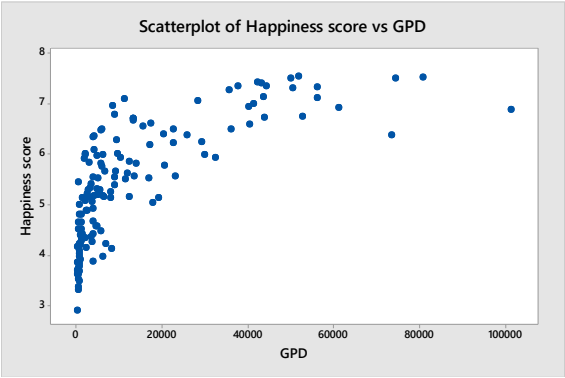
Happiness score, as mentioned above, was obtained from the 'World Happiness Report for 2016'. Democratic values were taken from the 'Democratic Index' by The Economist Intelligence Unit. LGBT laws were evaluated on the 'LGBT Right by Country or Territory' Wikipedia page and, after personally confirming references, were split into appropriate levels. All other data was obtained from The World Bank's online database. While the majority of these values are from 2016, democratic index, inequality gini, and female workplace percent are numbers from 2015 as this was the most recent available. After collection, all data was aggregated into a single CSV to support analysis.

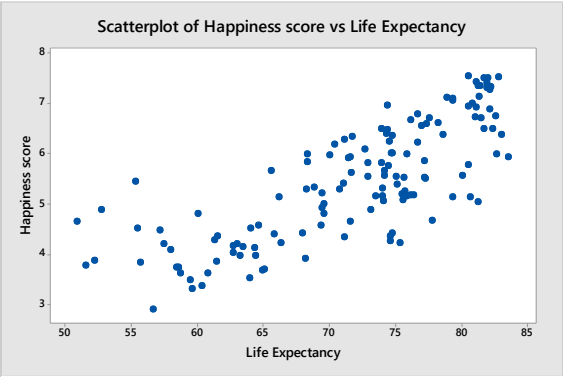
Exploratory Data Analysis

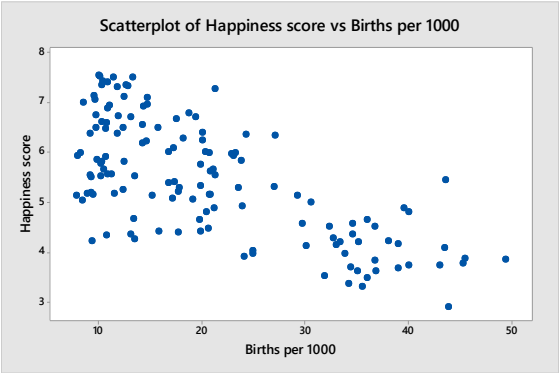
Visually exploring the data can provide several insights into how to proceed with regression analysis for this particular dataset. Below is a subset of the data- the top five 'happiest' countries- to give an idea of what the sample looks like. Scandinavia dominates.

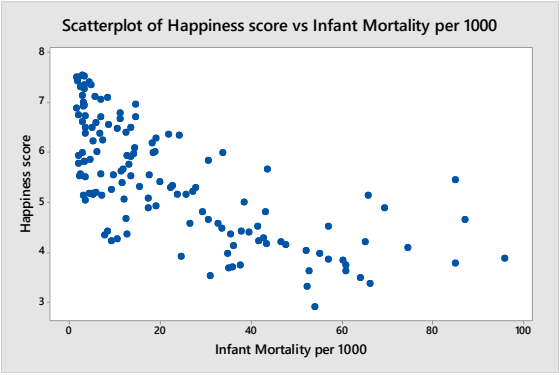
Country	GPD	Happiness score	Life Expectancy	Births per 1000	Infant Mortality per 1000	Urban Percent	Female Workforce Percent	Democracy Score	LGBT freedom	Inequality Gini
Denmark	51989.293	7.526	80.548	10.1	2.9	87.676	47.687	full	high	29.1
Switzerland	80945.079	7.508	82.848	10.2	3.4	73.912	46.151	full	some	31.6
Iceland	50173.339	7.500	82.060	13.4	1.6	94.137	47.627	full	high	26.9
Norway	74400.369	7.498	81.751	11.5	2	80.473	47.110	full	high	25.9
Finland	42311.036	7.413	81.129	10.5	1.9	84.221	47.740	full	high	27.1

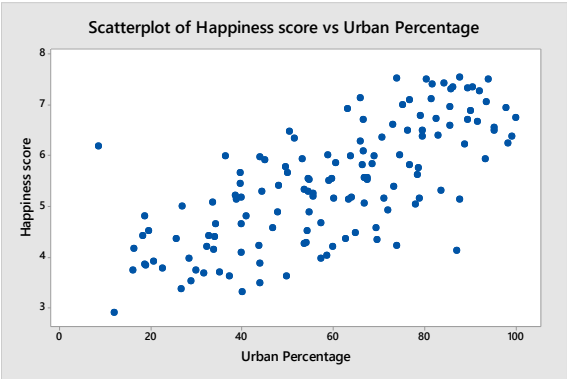
Single-Order Predictors to Dependent

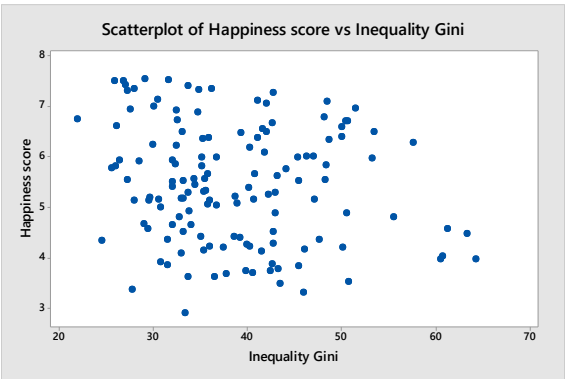


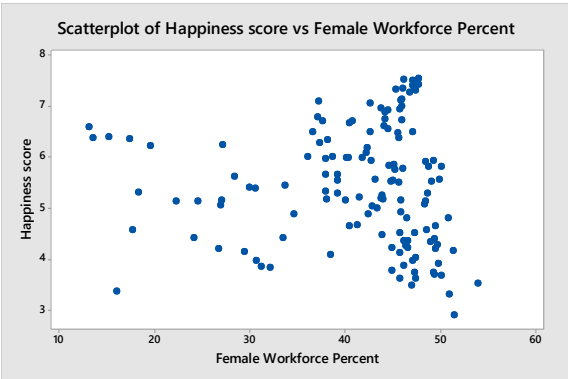


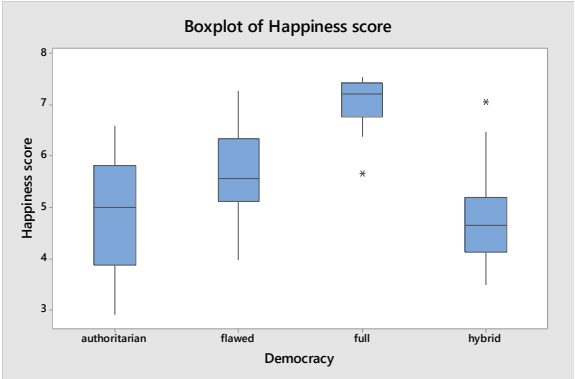












Visually inspecting these plots shows that most of the single-order variables have at least some level of effect on happiness of a country. Inequality gini and female workforce percentage appear to have little to no correlation, while the other six quantitative predictors depict a strong relationship, be it linear or curvilinear. The qualitative variable boxplots also seem to display an expected, significant relationship amongst each fields respective categories.

First-Order Model

The first-order model of this regression analysis in general form is as follows:

$$E(y) = \beta 0 + \beta 1_{x1} + \beta 2_{x2} + \beta 3_{x3} + \beta 4_{x4} + \beta 5_{x5} + \beta 6_{x6} + \beta 7_{x7} + \beta 8_{x8} + \beta 9_{x9} + \beta 10_{x10} + \beta 11_{x11} + \beta 12_{x12}$$

y = happiness score, $_{x1}$ = gross domestic product per capita, $_{x2}$ = life expectancy, $_{x3}$ = births per 1000, $_{x4}$ = infant mortality per 1000, $_{x5}$ =urban percentage, $_{x6}$ = female workforce percentage, $_{x7}$ = inequality gini, $_{x8}$ = democracy (flawed), $_{x9}$ = democracy (full) , $_{x10}$ = democracy (hybrid), $_{x11}$ = LGBT freedom (high), $_{x12}$ = LGBT freedom (some) || defaults -> *democracy = authoritarian, LGBT = low*

Using statistical software, Minitab, this regression equation was formed and evaluated. The results are printed below for further examination.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	12	139.107	11.5922	31.33	0.000
GPD	1	1.993	1.9933	5.39	0.022
Life Expectancy	1	2.704	2.7037	7.31	0.008
Births per 1000	1	0.163	0.1625	0.44	0.509
Infant Mortality per 1000	1	0.121	0.1211	0.33	0.568
Urban Percentage	1	1.933	1.9328	5.22	0.024
Female Workforce Percent	1	0.670	0.6696	1.81	0.181
Inequality Gini	1	1.151	1.1506	3.11	0.080
Democracy	3	1.413	0.4709	1.27	0.287
LGBT freedom	2	1.137	0.5685	1.54	0.219
Error	125	46.247	0.3700		
Total	137	185.353			

Model Summary

S	R-sq	R-sq(adj)
0.608254	75.05%	72.65%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.27	1.92	0.14	0.890	
GPD	0.000011	0.000005	2.32	0.022	3.25
Life Expectancy	0.0605	0.0224	2.70	0.008	12.74
Births per 1000	-0.0080	0.0121	-0.66	0.509	6.34
Infant Mortality per 1000	0.00418	0.00730	0.57	0.568	9.69
Urban Percentage	0.00839	0.00367	2.29	0.024	2.64
Female Workforce Percent	-0.00979	0.00728	-1.35	0.181	1.57
Inequality Gini	0.01202	0.00682	1.76	0.080	1.33
Democracy					
flawed	0.065	0.172	0.38	0.707	2.55
full	0.438	0.284	1.54	0.125	3.72
hybrid	-0.100	0.158	-0.63	0.528	1.70
LGBT freedom					
high	0.406	0.232	1.75	0.082	2.97
some	0.142	0.165	0.86	0.391	2.05

Regression Equation

Happiness score = 0.27 + 0.000011 GDP + 0.0605 Life Expectancy
- 0.0080 Births per 1000 + 0.00418 Infant Mortality per 1000
+ 0.00839 Urban Percentage - 0.00979 Female Workforce Percent
+ 0.01202 Inequality Gini + 0.065 Democracy_flawed + 0.438 Democracy_full
- 0.100 Democracy_hybrid + 0.406 LGBT freedom_high + 0.142 LGBT freedom_some

Fits and Diagnostics for Unusual Observations

Obs	Happiness score	Fit	Resid	Std Resid	
33	6.474	5.285	1.189	2.03	R
38	6.324	5.080	1.244	2.11	R
42	6.168	4.893	1.275	2.29	R
46	5.987	4.734	1.253	2.11	R
69	5.440	4.043	1.397	2.43	R
86	5.123	6.280	-1.157	-2.03	R
103	4.459	5.114	-0.655	-1.30	X
113	4.217	5.676	-1.459	-2.46	R

R Large residual
X Unusual X

This first-order regression is surprisingly efficient at predicting the happiness score of a country. This model accounts for and explains around 72.65% (R2-adj) of the variation of a countries happiness score in the sample. The standard error of the model is 0.608, meaning that the results of this model will typically (~95%) fall within 1.2 points from the actual happiness score.

Least-Squares Regression

HAPPINESS SCORE = 0.27 + 0.000011 GDP + 0.0605 LIFE EXPECTANCY
- 0.0080 BIRTHS PER 1000 + 0.00418 INFANT MORTALITY PER 1000
+ 0.00839 URBAN PERCENTAGE - 0.00979 FEMALE WORKPLACE PERCENT
+ 0.01202 INEQUALITY GINI + 0.065 DEMOCRACY (*FLAWED*)
+ 0.438 DEMOCRACY (*FULL*)- 0.100 DEMOCRACY (*HYBRID*)
+ 0.406 LGBT FREEDOM (*HIGH*) + 0.142 LGBT FREEDOM (*LOW*)

Interpreting Coefficients

With the results of the regression from the Minitab output, the β coefficients in this model are able to be put into real-world terms. $\beta 1_{(x1)}$ is equal to 0.000011 GDP. With every one unit increase in GPD per capita, keeping all other predictors fixed, we can expect the happiness score to raise by .000011. A seemingly small weight, yet we are dealing in tens of thousands for this variable, evening it out. $\beta 2_{(x2)}$ represents 0.0605 LIFE EXPECTANCY. A one year increase in a country's average life expectancy will increase the happiness score by an average of 0.0605, with all other variables keeping fixed. $B3_{(x3)}$ is -0.008 BIRTHS PER 1000. For every added birth per 1000 people in the country, with all other variables staying fixed, its happiness score should decrease by 0.008.

$\beta 4_{(x4)}$ in the general model equals 0.00418 INFANT MORTALITY PER 1000. This means that, with other variables staying fixed, every 1 increase in a country's infant mortality rate somehow increases happiness by 0.00418. Intuition tells us that this is either an incorrect relationship displayed in the regression, or that humans are much more depraved than one would hope. Thankfully, the latter is most likely the case as both the scatterplot above and a direct happiness to infant mortality linear regression express a negative correlation between the two.

This anomaly is most likely the result of a positive covariance between infant mortality and one or more of the other predictors. More exploration will be done in later parts of this analysis to determine the cause of this.

$\beta 5_{(x5)}$ is 0.00839 URBAN PERCENTAGE. Keeping all other variables fixed, a one percent increase in percent of the population living in urban areas will increase happiness by 0.00839 in the country. $\beta 6_{(x6)}$ is -0.00979 FEMALE WORKPLACE PERCENTAGE. This coefficient would mean a one percent increase in the percent of a countries workplace that is female would decrease the happiness by 0.00979. While this looks like it may represent a form of sexism in the workplace, it is a fairly weak indicator of happiness and may not be fully applicable. $\beta 7_{(x7)}$ equates to 0.01202 INEQUALITY GINI. For every increase in the inequality gini percentage, which is a representation of the disparity of the distribution of wealth between the wealthy and the poor of a country, there will be an increase of happiness by 0.01202, with all other variables staying fixed. Thankfully, another weak indicator.

$\beta 8_{(x8)}$ is the general form of 0.065 DEMOCRACY (*FLAWED*). If the country falls under the category of flawed in the democracy type qualitative variable, happiness will increase by 0.065, with all other variables staying fixed, over that of the model default of authoritarian. Similarly, $\beta 9_{(x9)}$ is 0.438 DEMOCRACY (*FULL*). If the country's government is a full-fledged democracy, the happiness of the country will be on average 0.438 higher than that of the default authoritarian. Lastly for democracy type, $\beta 10_{(x10)}$ is -0.100 DEMOCRACY (*HYBRID*). Oddly enough, people living in a hybrid form of democracy, a lower level than flawed yet more democratic than authoritarian, will have .1 lower happiness score on average than that of the authoritarian, default model.

In a similar method to the democracy type, LGBT FREEDOM is a qualitative predictor split between three levels, with LOW being the model default. $\beta 11_{(x11)}$ is the general form of 0.406 LGBT FREEDOM (HIGH). Countries with numerous laws protecting LGBT persons equality result in a 0.406 increase in happiness, keeping all other predictors fixed. $\beta 12_{(x12)}$ represents 0.142 LGBT FREEDOM (SOME). Countries with a few laws protecting these freedoms increase the happiness of the country by 0.142. Finally, $\beta 0$ can be seen as the y-intercept. Countries begin with a score of 0.27 happiness and then are affected by other independent variables in the model.

First-Order Model's Usefulness

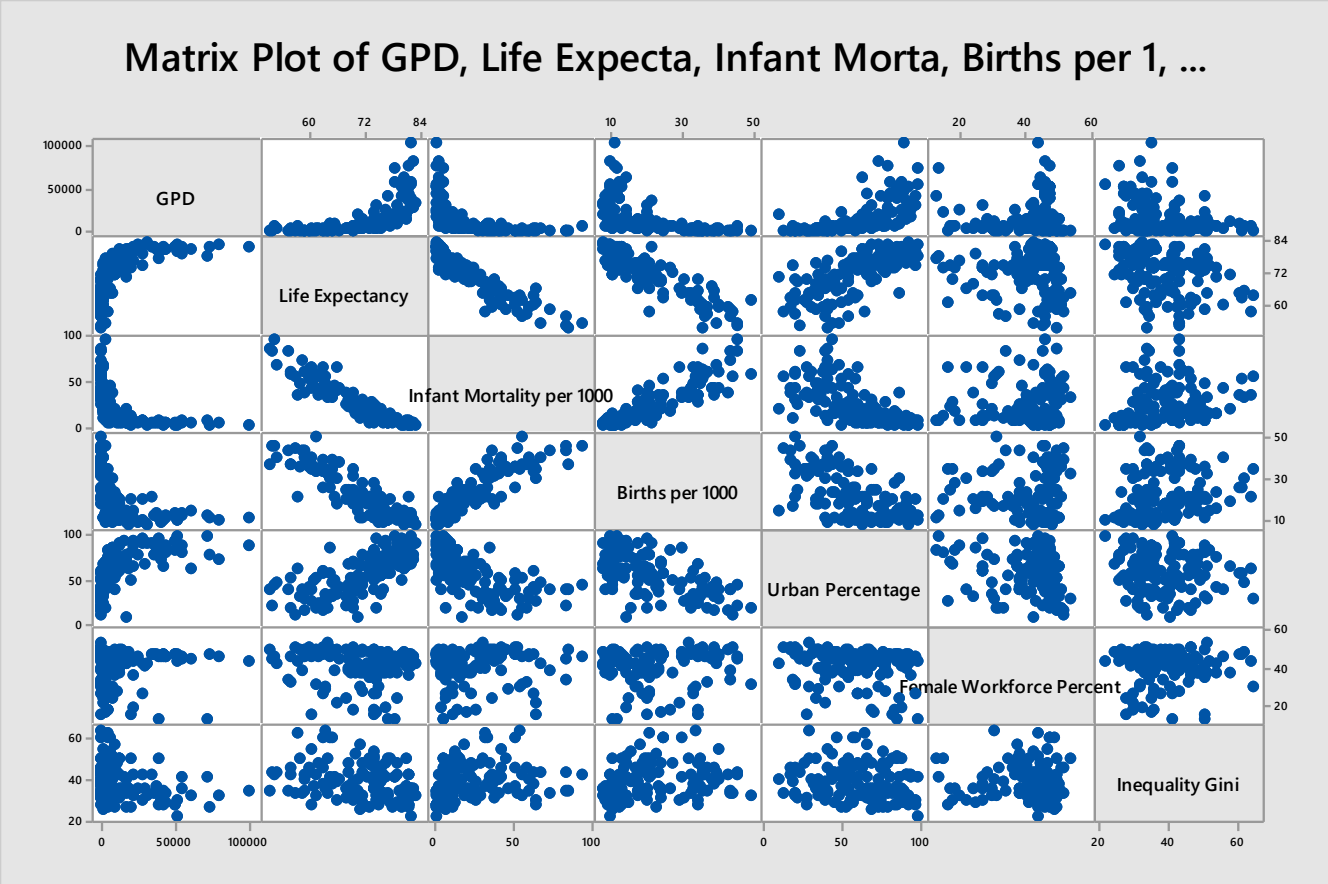
Using the traditional alpha of 0.05, we can perform a null hypothesis test to determine if the model is useful. Using $H_0 : \beta 1 = \beta 2 = \ldots = \beta 12 = 0$ and H_A: any $\beta \neq 0$, we can find the F-score and P-value of our regression as a whole and determine if it is statistically significant at explaining and predicting a countries happiness score. With an F-score of 31.33 and a p-value of 0.000, we can conclude, with over 99% confidence, that this model is efficient at explaining/predicting a countries happiness score.

Looking deeper into the variables, T-tests provide methods of determining the direct significance of each independent variable to the dependent. GDP, LIFE EXPECTANCY, and URBAN PERCENT are by far the strongest predictors of happiness in this model. The weakest appearing to be BIRTHS PER 1000 and INFANT MORTALITY PER 1000. The qualitative predictors encompassing the democracy type also have high p-values, yet should not necessarily be thrown out as they could provide added strength to the model, particularly when introducing interaction terms.

Interaction-Terms

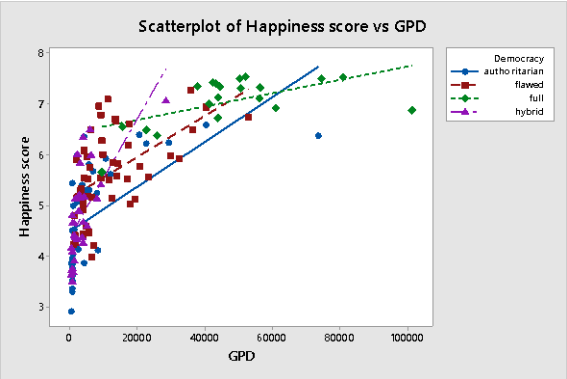
The next step of this regression analysis will be to explore and evaluate any interaction terms between independent variables in the model. Once again, we will begin by visually exploring these relationships before running them through the Minitab regression process.

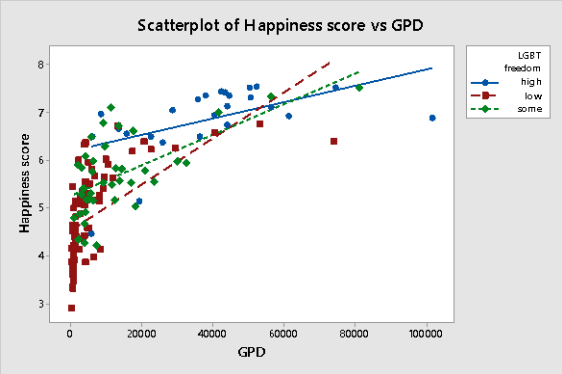
Interaction EDA

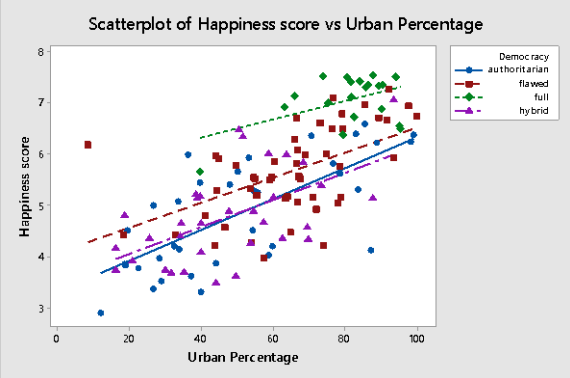


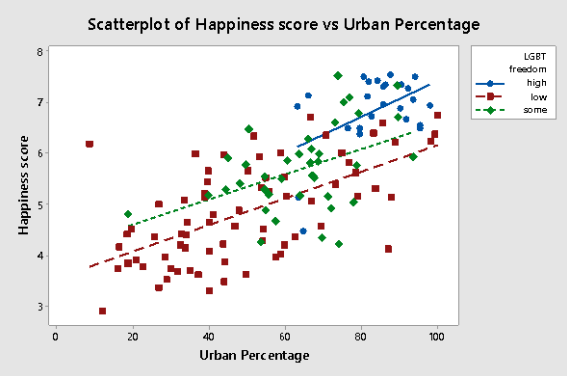
The matrix plot above gives a quick view of which predictors may have an effect on each other, (not necessarily in regards to happiness though). It is important to keep in mind that, while these scatterplots give an idea of relationships between independent variables, they may in fact be additive and not necessarily interactive. Female workforce percent and inequality gini both appear to have very little connection with other variables. The remaining five look to each have

some type of linear or curvilinear relationship with other independent variables. While this implies some form of covariance, care must be taken watch for multicollinearity.









Qualitative terms can be evaluated for interaction via a scatterplot graphing a predictor to the result, split into categorical groups. Four examples are provided to give a look at how the two qualitative predictors interact and can change the plotted slope with some terms- like GDP in the first two for examples- yet can have very weak influence on others, like urban percentage in the second two examples.

Interaction-Term Model

Both the general and the actual regression model will not be reported within this section as, due to a determined lack of statistically significant interaction, it is unnecessary to report a model with 68 predictors. Regardless, we can evaluate the importance of these interactions and

the necessity for any via the students T-test for determining significance between a single predictor and a dependent.

Model Summary

S	R-sq	R-sq(adj)
0.569744	87.92%	76.01

Coefficients

Term	Coef	SE Coef	T-Value	P- Value
Constant	-12.1	28.9	-0.42	0.676
GPD	0.000481	0.000469	1.03	0.308
Life Expectancy	0.196	0.352	0.56	0.579
Births per 1000	-0.102	0.354	-0.29	0.774
Infant Mortality per 1000	0.113	0.140	0.81	0.422
Urban Percentage	-0.131	0.198	-0.66	0.511
Female Workforce Percent	0.626	0.437	1.43	0.156
Inequality Gini	-0.031	0.394	-0.08	0.938
Democracy				
flawed	-4.61	9.46	-0.49	0.627
full	9.5	31.0	0.31	0.759
hybrid	-7.00	9.11	-0.77	0.445
LGBT freedom				
high	-29.9	27.5	-1.09	0.280
some	5.10	7.85	0.65	0.518
GPD*Life Expectancy	-0.000005	0.000005	-0.86	0.390
GPD*Births per 1000	0.000003	0.000004	0.74	0.461
GPD*Infant Mortality per 1000	-0.000001	0.000003	-0.54	0.594
GPD*Urban Percentage	-0.000000	0.000001	-0.72	0.473
GPD*Female Workforce Percent	-0.000002	0.000003	-0.57	0.573
GPD*Inequality Gini	-0.000002	0.000002	-0.98	0.331
Life Expectancy*Births per 1000	0.00062	0.00437	0.14	0.888
Life Expectancy*Infant Mortality per 1000	-0.00079	0.00154	-0.51	0.611
Life Expectancy*Urban Percentage	0.00144	0.00224	0.64	0.524
Life Expectancy*Female Workforce Percent	-0.00597	0.00530	-1.13	0.264
Life Expectancy*Inequality Gini	0.00075	0.00444	0.17	0.866
Births per 1000*Infant Mortality per 1000	0.00029	0.00138	0.21	0.834
Births per 1000*Urban Percentage	0.00062	0.00119	0.52	0.605
Births per 1000*Female Workforce Percent	-0.00337	0.00233	-1.45	0.152
Births per 1000*Inequality Gini	0.00248	0.00274	0.91	0.367
Infant Mortality per 1000*Urban Percentage	0.000273	0.000782	0.35	0.728
Infant Mortality per 1000*Female Workforce	-0.00089	0.00122	-0.73	0.468
Infant Mortality per 1000*Inequality Gini	-0.00091	0.00130	-0.70	0.485
Urban Percentage*Female Workforce Percent	-0.000305	0.000819	-0.37	0.711
Urban Percentage*Inequality Gini	0.000518	0.000718	0.72	0.473
Female Workforce Percent*Inequality Gini	-0.00227	0.00118	-1.92	0.059
GPD*Democracy				
flawed	0.000058	0.000072	0.81	0.421
full	0.000068	0.000084	0.81	0.421
hybrid	0.000093	0.000083	1.13	0.263
GPD*LGBT freedom				
high	-0.000027	0.000037	-0.73	0.469
some	-0.000002	0.000031	-0.06	0.951
Life Expectancy*Democracy				
flawed	0.030	0.108	0.27	0.786
full	-0.169	0.376	-0.45	0.654

hybrid	0.0477	0.0969	0.49	0.625
Life Expectancy*LGBT freedom				
high	0.264	0.301	0.88	0.382
some	-0.0661	0.0944	-0.70	0.486
Births per 1000*Democracy				
flawed	0.0916	0.0530	1.73	0.088
full	0.030	0.175	0.17	0.864
hybrid	0.0418	0.0473	0.88	0.379
Births per 1000*LGBT freedom				
high	-0.089	0.189	-0.47	0.638
some	-0.0312	0.0665	-0.47	0.641
Infant Mortality per 1000*Democracy				
flawed	-0.0143	0.0368	-0.39	0.699
full	-0.193	0.386	-0.50	0.618
hybrid	0.0116	0.0252	0.46	0.648
Infant Mortality per 1000*LGBT freedom				
high	0.259	0.261	0.99	0.324
some	0.0114	0.0454	0.25	0.803
Urban Percentage*Democracy				
flawed	0.0073	0.0169	0.43	0.669
full	-0.0087	0.0326	-0.27	0.789
hybrid	-0.0087	0.0157	-0.56	0.580
Urban Percentage*LGBT freedom				
high	0.0505	0.0356	1.42	0.160
some	0.0097	0.0147	0.66	0.514
Female Workforce Percent*Democracy				
flawed	-0.0248	0.0293	-0.85	0.401
full	0.069	0.199	0.35	0.730
hybrid	-0.0134	0.0283	-0.47	0.638
Female Workforce Percent*LGBT freedom				
high	0.107	0.192	0.56	0.578
some	-0.0131	0.0370	-0.35	0.724
Inequality Gini*Democracy				
flawed	0.0308	0.0274	1.12	0.265
full	0.050	0.132	0.38	0.707
hybrid	0.0725	0.0268	2.70	0.009
Inequality Gini*LGBT freedom				
high	0.0108	0.0848	0.13	0.899
some	0.0072	0.0267	0.27	0.787

With a lower standard deviation and a high R2, we would initially consider this model to be a stronger, more advanced version of our first-order, yet this may not be the case. To decide if it is acceptable to use this model, we must first determine if any of the interaction terms are actually significant to predicting 'happiness'. Evaluating each predictor at an alpha of 0.05 and comparing to the T-statistic and P-value, we conclude that there are no significant interaction terms amongst these predictors that contribute to further explaining a countries happiness score. A nested F-test can confirm this for us.

Nested Model Test

F: $((SSE_1 - SSE_2)/\#\beta\text{'s in } H_0) / MSE_2 = ((46.247/22.398)/56)/0.324 =$ **0.1138**

Critical Value (α = 0.05) : $V_1 = 56$, $V_2 = 69$, $F(56/69) =$ **~1.534**

The test statistic fails and we should not include the interaction terms in the final model!

Second-Order Terms

Exploratory analysis is unnecessary in this step due to already achieving an idea of what will call for higher-order terms in the initial EDA. Recalling these scatterplots, it is expected that at least a few of the predictors will have significantly improved accuracy in explaining happiness once the second-order is included.

Model

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	19	148.145	7.79709	24.73	0.000
GPD	1	5.227	5.22743	16.58	0.000
Life Expectancy	1	0.741	0.74059	2.35	0.128
Births per 1000	1	2.292	2.29169	7.27	0.008
Infant Mortality per 1000	1	1.310	1.30997	4.15	0.044
Urban Percentage	1	0.021	0.02052	0.07	0.799
Female Workforce Percent	1	2.485	2.48509	7.88	0.006
Inequality Gini	1	0.443	0.44309	1.41	0.238
GDP^2	1	3.207	3.20683	10.17	0.002
life^2	1	0.615	0.61486	1.95	0.165
births^2	1	2.631	2.63111	8.34	0.005
infant^2	1	1.741	1.74103	5.52	0.020
urban^2	1	0.000	0.00003	0.00	0.992
female^2	1	2.867	2.86657	9.09	0.003
inequality^2	1	0.300	0.30017	0.95	0.331
Democracy	3	0.513	0.17084	0.54	0.655
LGBT freedom	2	1.228	0.61379	1.95	0.147
Error	118	37.208	0.31533		
Total	137	185.353			

Model Summary

S	R-sq	R-sq(adj)
0.561539	79.93%	76.69%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-9.73	6.75	-1.44	0.152	
GPD	0.000056	0.000014	4.07	0.000	30.13
Life Expectancy	0.286	0.187	1.53	0.128	1041.46

Births per 1000	0.1002	0.0372	2.70	0.008	69.79
Infant Mortality per 1000	-0.0334	0.0164	-2.04	0.044	57.19
Urban Percentage	0.0031	0.0123	0.26	0.799	34.58
Female Workforce Percent	0.1186	0.0422	2.81	0.006	62.05
Inequality Gini	0.0504	0.0425	1.19	0.238	60.44
GDP^2	-0.000000	0.000000	-3.19	0.002	14.43
life^2	-0.00188	0.00135	-1.40	0.165	1038.52
births^2	-0.001912	0.000662	-2.89	0.005	57.51
infant^2	0.000391	0.000166	2.35	0.020	34.65
urban^2	-0.000001	0.000109	-0.01	0.992	37.11
female^2	-0.001788	0.000593	-3.02	0.003	60.23
inequality^2	-0.000493	0.000505	-0.98	0.331	60.60
Democracy					
flawed	-0.040	0.166	-0.24	0.810	2.79
full	0.236	0.276	0.86	0.394	4.13
hybrid	-0.065	0.150	-0.44	0.664	1.79
LGBT freedom					
high	0.411	0.219	1.87	0.063	3.12
some	0.228	0.158	1.44	0.153	2.23

Regression Equation

Happiness score = -9.73 + 0.000056 GPD + 0.286 Life Expectancy
+ 0.1002 Births per 1000 - 0.0334 Infant Mortality per 1000 + 0.0031 Urban Percentage
+ 0.1186 Female Workforce Percent + 0.0504 Inequality Gini - 0.000000 GDP^2
- 0.00188 life^2 - 0.001912 births^2 + 0.000391 infant^2 - 0.000001 urban^2
- 0.001788 female^2 - 0.000493 inequality^2 - 0.040 Democracy_flawed
+ 0.236 Democracy_full - 0.065 Democracy_hybrid + 0.411 LGBT freedom_high
+ 0.228 LGBT freedom_some

Fits and Diagnostics for Unusual Observations

Obs	Happiness score	Fit	Resid	Std Resid	
20	6.871	7.076	-0.205	-0.63	X
33	6.474	5.263	1.211	2.26	R
37	6.355	5.065	1.290	2.51	R
52	5.897	4.620	1.277	2.40	R
69	5.440	4.352	1.088	2.14	R
104	4.415	5.439	-1.024	-2.05	R

R Large residual
X Unusual X

Nested Test Model

Due to the insignificance of the interaction term model, this model should be tested against our original, main effects model to determine if second-order terms should be included in the final regression.

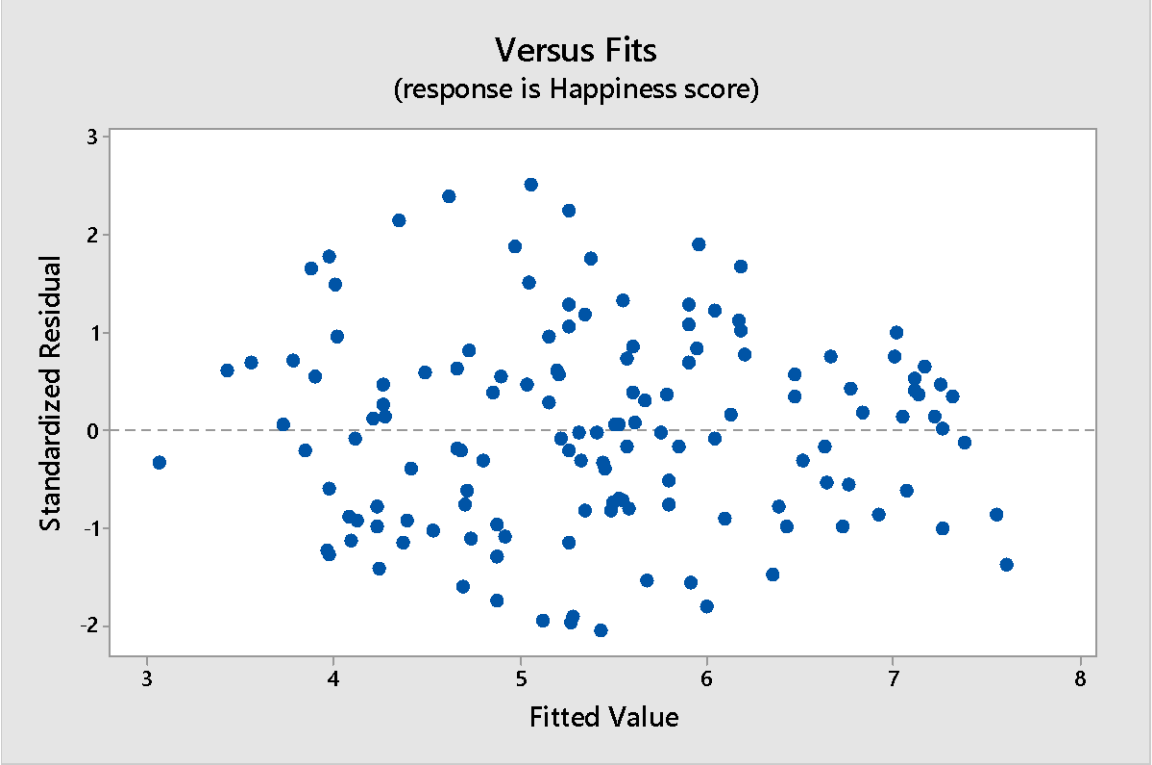
F: $((SSE_1 - SSE_2)/\#\beta\text{'s in } H_0) / MSE_2 = ((46.247-37.208)/7)/0.315 =$ **4.099**

Critical Value ($\alpha = 0.05$) : $V_1 = 7$, $V_2 = 118$, $F(7/118) =$ **~2.0868**

Due to the results of this nested F-test, (F-score > critical value), we can conclude that the second-order terms are significant in predicting the happiness score, and increase the accuracy of the model. It explains about five percent more of the variance in our sample data (R-sq, adj = 76.69%), with a slight lower (by .04) standard error. Moving forward, residual analysis should be undertaken to confirm the least square assumptions regarding the random error term.

Residual Analysis

Due to the data are not being time-series, assumptions regarding the mean equaling zero and the errors being independent of each other can be confirmed, and they are not violated in this model. Testing for unequal variance is the next step, which can be achieved by checking for heteroscedasticity errors. This is done by plotting the residuals vs. the predicted values.



There seems to be no pattern or trend within the residual plot and we can conclude that there is equal variance which is consistent throughout all independent values.

This plot also highlights that there are five slight (beyond two standard deviations) outliers in the data. Though not drastically out of the expected bounds, we can find the leverage of each of these outliers to figure how influential they are on the regression. If there are any strongly influential outliers, a deeper look into data retrieval and input methods may be necessary to determine if this abnormality is due to either human or mechanical error, as opposed to an actual result.

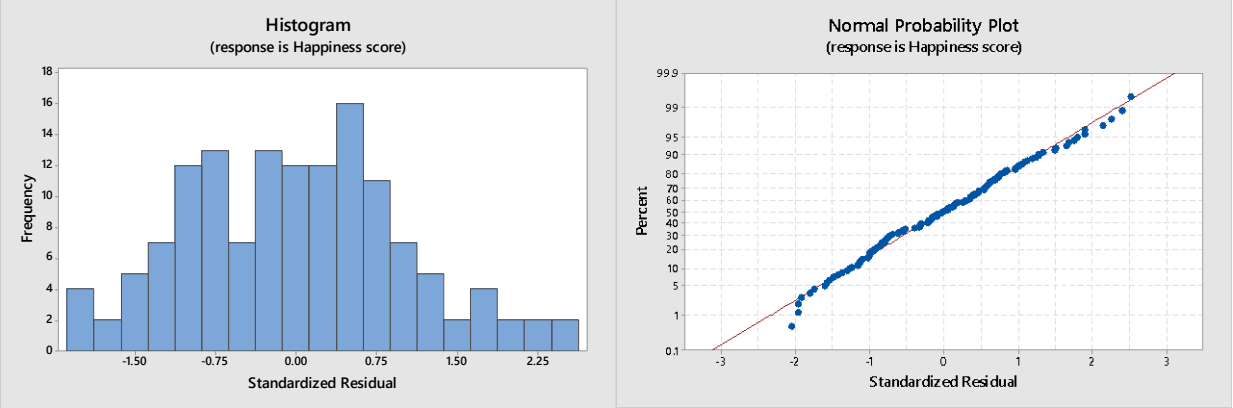
Test Statistic = 2(k + 1) / n = 2(20)/137 -> **0.292**

Thailand = 0.088 Algeria = 0.164 Moldova = 0.098

Somalia = 0.183 Sri Lank = 0.205

None of the outliers in the dataset significantly influence the regression equation.

The final assumption to confirm is that the data adheres to a normal distribution. This will be tested through visual inspection via both a histogram and a normal probability plot.



We can confirm that the data follows a normal distribution. There is a more-or-less bell shape to the histogram, and the normal probability plot does not depict too much scatter throughout. With this final assumption confirmed, we can move on to reducing this model, one step at a time, to find the most efficient regression to predict a countries happiness score.

Predictor Reduction

Reducing the model will be a step by step process of removing one predictor variable at a time and evaluating to determine if it significantly adds to the regression and should be included, regardless of if it has a low T-score. Performing reduction in an iterative process like this allows for careful review of the effects of each predictor. While, typically, a low t-score represents a predictor that is unnecessary to the final model, this is not always the case, and removing one at a time allows us to watch for this possibility. If we were to just build and use the first model, we could be including numerous predictors that are weak and which are unnecessary in predicting future outcomes. If we do not explore reducing the model, it is very possible for the model to include a wider than expected variance in the regression, which would lead to an inaccurate final model. This step-by-step process is also imperative as removing several weak looking indictors at once could remove one or more that, while seemingly weak in the larger model, gain a strong relationship to the results with other indicators removed. To begin this process, we will use the following model- in general form:

$$E(y) = \beta 0 + \beta 1_{(x1)} + \beta 2_{(x2)} + \beta 3_{(x3)} + \beta 4_{(x4)} + \beta 5_{(x5)} + \beta 6_{(x6)} + \beta 7_{(x7)} + \beta 8_{(x8)} + \beta 9_{(x9)} + \beta 10_{(x10)} + \beta 11_{(x11)} + \beta 12_{(x12)} + \beta 13_{(x13)} + \beta 14_{(x14)} + \beta 15_{(x15)} + \beta 16_{(x16)} + \beta 17_{(x17)} + \beta 18_{(x18)} + \beta 19_{(x19)}$$

The least-squares regression form of this is:

```
Happiness score = -9.73 + 0.000056 GPD + 0.286 Life Expectancy
+ 0.1002 Births per 1000 - 0.0334 Infant Mortality per 1000
+ 0.0031 Urban Percentage + 0.1186 Female Workforce Percent
+ 0.0504 Inequality Gini - 0.000000 GDP^2 - 0.00188 life^2 - 0.001912 births^2
+ 0.000391 infant^2 - 0.000001 urban^2 - 0.001788 female^2
- 0.000493 inequality^2 - 0.040 Democracy_flawed + 0.236 Democracy_full
- 0.065 Democracy_hybrid + 0.411 LGBT freedom_high + 0.228 LGBT freedom_some
```

Looking at the Minitab printout above, the most insignificant predictor in the equation is urban percentage^2, at a P-value of 0.992. We remove this and refit:

Model Summary

S	R-sq	R-sq(adj)
0.559175	79.93%	76.89%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-9.73	6.71	-1.45	0.150	
GPD	0.000056	0.000013	4.20	0.000	28.46
Life Expectancy	0.287	0.185	1.55	0.124	1030.88
Births per 1000	0.1002	0.0370	2.71	0.008	69.57
Infant Mortality per 1000	-0.0334	0.0163	-2.05	0.043	57.01
Urban Percentage	0.00301	0.00370	0.81	0.417	3.16
Female Workforce Percent	0.1186	0.0421	2.82	0.006	62.03
Inequality Gini	0.0503	0.0422	1.19	0.236	60.29
GDP^2	-0.000000	0.000000	-3.25	0.001	13.99
life^2	-0.00188	0.00133	-1.41	0.161	1028.24
births^2	-0.001913	0.000654	-2.92	0.004	56.62
infant^2	0.000391	0.000164	2.39	0.018	33.81
female^2	-0.001788	0.000591	-3.03	0.003	60.23
inequality^2	-0.000493	0.000502	-0.98	0.329	60.35
Democracy					
flawed	-0.040	0.164	-0.24	0.809	2.75
full	0.237	0.269	0.88	0.381	3.96
hybrid	-0.065	0.149	-0.44	0.662	1.79
LGBT freedom					
high	0.411	0.217	1.90	0.060	3.07
some	0.228	0.156	1.46	0.148	2.19

A slightly increased R2(adj) score and slightly lower standard error. The next more insignificant predictor appears to be Urban percentage. Removing this we get:

Model Summary

S	R-sq	R-sq(adj))
0.558390	79.81%	76.95%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-9.39	6.69	-1.40	0.163	
GPD	0.000058	0.000013	4.51	0.000	27.02
Life Expectancy	0.279	0.185	1.51	0.133	1028.52
Births per 1000	0.1082	0.0356	3.04	0.003	64.71
Infant Mortality per 1000	-0.0367	0.0158	-2.33	0.022	53.52
Female Workforce Percent	0.1168	0.0419	2.78	0.006	61.86
Inequality Gini	0.0503	0.0422	1.19	0.235	60.29
GDP^2	-0.000000	0.000000	-3.47	0.001	13.49
life^2	-0.00182	0.00133	-1.37	0.174	1024.85
births^2	-0.002057	0.000629	-3.27	0.001	52.47
infant^2	0.000422	0.000159	2.66	0.009	31.96
female^2	-0.001780	0.000590	-3.02	0.003	60.21

inequality^2	-0.000482	0.000501	-0.96	0.338	60.32
Democracy					
flawed	-0.044	0.164	-0.27	0.790	2.75
full	0.209	0.266	0.78	0.435	3.89
hybrid	-0.069	0.149	-0.46	0.646	1.78
LGBT freedom					
high	0.453	0.210	2.16	0.033	2.89
some	0.246	0.155	1.59	0.115	2.15

	DF	SS	MS
Error	120	37.416	0.3118

Model summary stats are sitting at around the same as with the predictor. Time to test if our democracy score is significant to the model, a nested F-test is undertaken:

Model Summary

S	R-sq	R-sq(adj)
0.554904	79.57%	77.24%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-9.87	6.63	-1.49	0.139	
GPD	0.000062	0.000012	5.16	0.000	24.23
Life Expectancy	0.289	0.183	1.58	0.117	1023.60
Births per 1000	0.1066	0.0353	3.02	0.003	64.46
Infant Mortality per 1000	-0.0349	0.0154	-2.27	0.025	51.57
Female Workforce Percent	0.1214	0.0394	3.08	0.003	55.36
Inequality Gini	0.0509	0.0417	1.22	0.225	59.77
GDP^2	-0.000000	0.000000	-3.71	0.000	12.94
life^2	-0.00189	0.00132	-1.44	0.153	1017.70
births^2	-0.002031	0.000623	-3.26	0.001	52.26
infant^2	0.000410	0.000156	2.63	0.010	31.34
female^2	-0.001839	0.000561	-3.28	0.001	55.21
inequality^2	-0.000502	0.000494	-1.02	0.311	59.27
LGBT freedom					
high	0.517	0.196	2.64	0.009	2.54
some	0.235	0.151	1.55	0.124	2.08

	DF	SS	MS
Error	123	37.874	0.3079

F: $((SSE_R - SSE_C)/\#\beta\text{'s in } H_0) / MSE_C = ((37.874-37.416)/3)/0.3118 =$ **0.4896**

Critical Value ($\alpha = 0.05$) : $V_1 = 3$, $V_2 = 119$, $F(3/119) =$ **~2.6802**

The F statistic from the nested F-test does not break the rejection region, and we therefore should not include the democracy score qualitative terms in the final model. This will not be tested with LGBT Freedom terms as one qualitative variable must remain in the final model.

Inequality gini^2 is the next qualitative variable to look at:

Model Summary

S	R-sq	R-sq(adj)
0.554978	79.39%	77.23%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-8.98	6.57	-1.37	0.174	
GPD	0.000061	0.000012	5.07	0.000	23.76
Life Expectancy	0.283	0.183	1.55	0.124	1022.47
Births per 1000	0.1077	0.0353	3.05	0.003	64.40
Infant Mortality per 1000	-0.0350	0.0154	-2.28	0.025	51.57
Female Workforce Percent	0.1219	0.0394	3.09	0.002	55.35
Inequality Gini	0.00904	0.00652	1.39	0.168	1.46
GDP^2	-0.000000	0.000000	-3.62	0.000	12.80
life^2	-0.00182	0.00132	-1.38	0.169	1014.81
births^2	-0.002019	0.000623	-3.24	0.002	52.24
infant^2	0.000414	0.000156	2.65	0.009	31.32
female^2	-0.001846	0.000561	-3.29	0.001	55.20
LGBT freedom					
high	0.508	0.195	2.60	0.010	2.54
some	0.230	0.151	1.52	0.132	2.08

After evaluating the regression output, the next insignificant variable to consider removing is life expectancy^2:

Model Summary

S	R-sq	R-sq(adj)
0.557011	79.08%	77.07%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.22	1.78	-0.12	0.902	
GPD	0.000055	0.000011	4.88	0.000	20.48
Life Expectancy	0.0315	0.0213	1.48	0.141	13.69
Births per 1000	0.1147	0.0351	3.27	0.001	63.09
Infant Mortality per 1000	-0.0293	0.0149	-1.97	0.051	47.93
Female Workforce Percent	0.1154	0.0393	2.94	0.004	54.58
Inequality Gini	0.00763	0.00646	1.18	0.240	1.42
GDP^2	-0.000000	0.000000	-3.38	0.001	12.02
births^2	-0.002198	0.000612	-3.59	0.000	49.99
infant^2	0.000308	0.000137	2.25	0.026	23.87
female^2	-0.001782	0.000561	-3.17	0.002	54.83
LGBT freedom					
high	0.478	0.195	2.45	0.015	2.51
some	0.220	0.152	1.45	0.149	2.08

Next on the chopping block is the main effect, Inequality gini:

Model Summary

S	R-sq	R-sq(adj)
0.557880	78.84%	77.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.26	1.73	0.15	0.880	
GPD	0.000056	0.000011	5.00	0.000	20.33
Life Expectancy	0.0257	0.0207	1.24	0.217	12.98
Births per 1000	0.1239	0.0342	3.62	0.000	59.95
Infant Mortality per 1000	-0.0291	0.0149	-1.95	0.053	47.92
Female Workforce Percent	0.1240	0.0387	3.20	0.002	52.72
GDP^2	-0.000000	0.000000	-3.49	0.001	11.92
births^2	-0.002367	0.000596	-3.97	0.000	47.26
infant^2	0.000290	0.000136	2.13	0.035	23.57
female^2	-0.001912	0.000551	-3.47	0.001	52.70
LGBT freedom					
high	0.506	0.194	2.61	0.010	2.47
some	0.229	0.152	1.51	0.134	2.07

Continuing on – full steam ahead – It's life expectancy's turn.

Model Summary

S	R-sq	R-sq(adj)
0.559067	78.58%	76.90%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.181	0.786	2.77	0.006	
GPD	0.000059	0.000011	5.54	0.000	18.78
Births per 1000	0.1217	0.0343	3.55	0.001	59.78
Infant Mortality per 1000	-0.0364	0.0137	-2.65	0.009	40.45
Female Workforce Percent	0.1333	0.0380	3.50	0.001	50.76
GDP^2	-0.000000	0.000000	-3.74	0.000	11.58
births^2	-0.002389	0.000597	-4.00	0.000	47.22
infant^2	0.000308	0.000136	2.27	0.025	23.31
female^2	-0.002077	0.000536	-3.87	0.000	49.64
LGBT freedom					
high	0.528	0.193	2.73	0.007	2.45
some	0.242	0.152	1.59	0.114	2.06

With the removal of this predictor, we have reached the point where all independent variables included appear significant to the regression. The final step to determine the final regression model will be to compare this reduced model to an earlier version, which has a higher R2 with more, yet less significant, variables.

One Last Nested Test

Reduced

Model Summary

S	R-sq	R-sq(adj)
0.559067	78.58%	76.90%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.181	0.786	2.77	0.006	
GPD	0.000059	0.000011	5.54	0.000	18.78
Births per 1000	0.1217	0.0343	3.55	0.001	59.78
Infant Mortality per 1000	-0.0364	0.0137	-2.65	0.009	40.45
Female Workforce Percent	0.1333	0.0380	3.50	0.001	50.76
GDP^2	-0.000000	0.000000	-3.74	0.000	11.58
births^2	-0.002389	0.000597	-4.00	0.000	47.22
infant^2	0.000308	0.000136	2.27	0.025	23.31
female^2	-0.002077	0.000536	-3.87	0.000	49.64
LGBT freedom					
high	0.528	0.193	2.73	0.007	2.45
some	0.242	0.152	1.59	0.114	2.06

	DF	SS	MS
Error	127	39.695	0.3126

Complete

Model Summary

S	R-sq	R-sq(adj)
0.554904	79.57%	77.24%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-9.87	6.63	-1.49	0.139	
GPD	0.000062	0.000012	5.16	0.000	24.23
Life Expectancy	0.289	0.183	1.58	0.117	1023.60
Births per 1000	0.1066	0.0353	3.02	0.003	64.46
Infant Mortality per 1000	-0.0349	0.0154	-2.27	0.025	51.57
Female Workforce Percent	0.1214	0.0394	3.08	0.003	55.36
Inequality Gini	0.0509	0.0417	1.22	0.225	59.77
GDP^2	-0.000000	0.000000	-3.71	0.000	12.94
life^2	-0.00189	0.00132	-1.44	0.153	1017.70
births^2	-0.002031	0.000623	-3.26	0.001	52.26
infant^2	0.000410	0.000156	2.63	0.010	31.34
female^2	-0.001839	0.000561	-3.28	0.001	55.21
inequality^2	-0.000502	0.000494	-1.02	0.311	59.27
LGBT freedom					
high	0.517	0.196	2.64	0.009	2.54
some	0.235	0.151	1.55	0.124	2.08

	DF	SS	MS
Error	123	37.874	0.3079

Nested Test

H_0: life expectancy = Inequality =life^2 = Inequality^2 = 0 // **H_A**: Any ≠ 0.

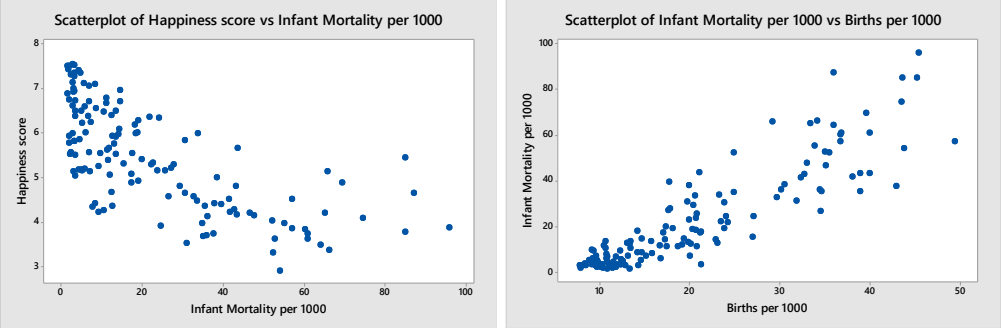
F: $((SSE_R - SSE_C)/\#\beta\text{'s in } H_0) / MSE_C = ((39.695-37.874)/4)/0.3079=$ **1.478**

Critical Value (α = 0.05) : $V_1 = 4$, $V_2 = 122$, $F(4/122) =$ **~2.447**

The test fails to reject the null hypothesis. The four added qualitative predictors in the complete model are not significant enough to include in the final model. While the 'complete' version above shows a slightly larger r2, the addition of these variables is not statistically significant in increasing the predictive power of the reduced model to account for including them. We should use the reduced model.

The Infant Mortality Problem

Intuition would tell us that a higher infant mortality rate should not increase a country's happiness score and, while the regression outlined above shows the opposite, it should be further evaluated to determine if this is actually the case.



By looking at the two scatterplots above, we can see that this increase in happiness derives from a handful of countries with extremely large infant mortality rates and moderate GDP, while the underlying relationship for the sample is, in fact, negative. It can be taken, thankfully, that this also partially arises due to a multicollinearity issue, as seen in the scatterplot of mortality rates

and births per 1000. These two independent variables are highly, positively correlated and infant mortality predictors can be removed to alleviate this issue.

Further Multicollinearity Issues

While many expected strong predictors are missing from this final model, (for example: life expectancy and urban percentage), this is most likely due to the high multicollinearity many of the predictors shared. It is quite reasonable to expect urban percentage, life expectancy, and inequality gini to be strongly linearly related to GDP. As was done with infant mortality, highly correlated independent variables like this should be removed to avoid errors which could cause improper regression results.

With a dataset like this including so many highly correlated predictors, it is possible that an entirely different final equation could be found if different predictors were chosen to be removed before reducing the model. If this analysis were to be done again, with a step added to remove all variables that appear to have weak direct correlation with happiness through visual inspection of the initial EDA, a different final model may be determined. An example of this would be to removed female workplace and inequality gini right at the beginning. Continuing from there, an entirely different regression may be found which would possibly include the expected predictors of life expectancy and infant mortality. While this multicollinearity problem can cause coefficients of independent predictors to skew slightly in their direct relationship to the dependent, the final model, with all predictors working together, should still be a valid, accurate predictor for happiness.

Final Model

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	143.374	17.9217	55.07	0.000
GPD	1	17.388	17.3877	53.43	0.000
Births per 1000	1	2.172	2.1720	6.67	0.011
Female Workforce Percent	1	4.057	4.0574	12.47	0.001
GDP^2	1	7.704	7.7038	23.67	0.000
births^2	1	4.244	4.2440	13.04	0.000
female^2	1	5.076	5.0756	15.60	0.000
LGBT freedom	2	2.576	1.2880	3.96	0.021
Error	129	41.980	0.3254		
Total	137	185.353			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.570459	77.35%	75.95%	74.26%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	2.197	0.802	2.74	0.007
GPD	0.000072	0.000010	7.31	0.000
Births per 1000	0.0750	0.0290	2.58	0.011
Female Workforce Percent	0.1365	0.0386	3.53	0.001
GDP^2	-0.000000	0.000000	-4.87	0.000
births^2	-0.001861	0.000515	-3.61	0.000
female^2	-0.002146	0.000543	-3.95	0.000
LGBT freedom				
high	0.540	0.197	2.74	0.007
some	0.296	0.153	1.93	0.056

Regression Equation

Happiness score = 2.197 + 0.000072 GPD + 0.0750 Births per 1000
+ 0.1365 Female Workforce Percent - 0.000000 GDP^2
- 0.001861 births^2 - 0.002146 female^2 + 0.540 LGBT freedom_high
+ 0.296 LGBT freedom_some

Fits and Diagnostics for Unusual Observations

Obs	Happiness score	Fit	Resid	Std Resid	
2	7.509	7.041	0.468	0.93	X
20	6.871	6.855	0.016	0.05	X
28	6.573	6.232	0.341	0.68	X
33	6.474	5.252	1.222	2.19	R
35	6.375	6.620	-0.245	-0.50	X
37	6.355	4.936	1.419	2.62	R
52	5.897	4.785	1.112	2.02	R
69	5.440	4.128	1.312	2.42	R
103	4.459	5.744	-1.285	-2.36	R
125	3.856	3.549	0.307	0.62	X

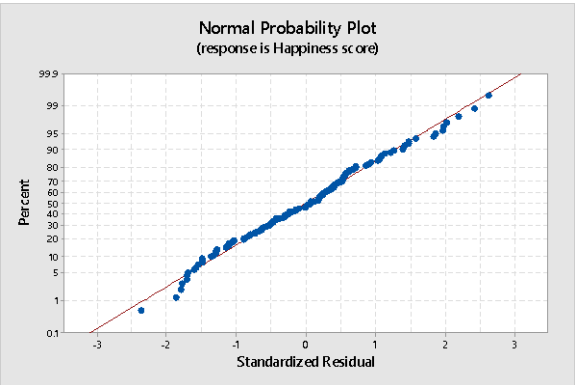
General Form

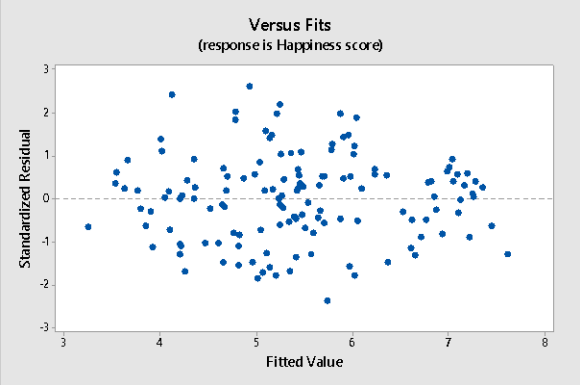
$$E(y) = \beta 0 + \beta 1_{(x1)} + \beta 2_{(x2)} + \beta 3_{(x3)} + \beta 4_{(x4)} + \beta 5_{(x5)} + \beta 6_{(x6)} + \beta 7_{(x7)} + \beta 8_{(x8)}$$

y = happiness score, $_{x1}$ = gross domestic product per capita, $_{x2}$ = births per 1000, $_{x3}$ = female workforce percentage, $_{x4}$ = gdp^2, $_{x5}$ = births^2, $_{x6}$ = female^2, $_{x7}$ = LGBT freedom (1 = high, 0 =low), $_{x8}$ = LGBT freedom (1 = some, 0 = low)

Least Squares Equation & Residuals

```
Happiness score = 2.197 + 0.000072 GDP + 0.0750 Births per 1000
+ 0.1365 Female Workforce Percent - 0.000000 GDP^2 - 0.001861 births^2
- 0.002146 female^2 + 0.540 LGBT freedom_high + 0.296 LGBT freedom_some
```





This final model adheres to all assumptions of normality and equal variance.

Interpreting Coefficients

Putting this model into real world terms allows one to understand how this equation can be put to use. $\beta 1_{(x1)}$ is equal to 0.000072 GDP. This means that for every one increase in gross domestic product per capita, there is a 0.000072 increase in happiness for the country, with all other variables staying fixed. $B2_{(x3)}$ is equivalent to 0.075 Births per 1000. Every addition baby born per 1000 people in the country increases the happiness score by 0.075, with other predictors staying fixed. $\beta 3_{(x3)}$ is equal to 0.1365 Female workforce percentage. Every one percent increase in the percentage of workers who are female in the country raises the happiness score of the country by 0.1365, keeping other independent variables fixed.

Looking at the second-order variables, $\beta 4_{(x4)}$ is equal to <0.0000 GDP^2. This is a minuscule number, enough that Minitab does not output it fully, yet we are working with massive numbers here. Every increase in GDP squared, with all other variables fixed, increases happiness a fractional amount. $B5_{(x5)}$ is equal to -0.001861 births^2. This implies a negative curvilinear relationship to births per 1000 people. If other variables stay fixed, an additional unit increase in this number will reduce a country's happiness score by 0.001861. $\beta 6_{(x6)}$ is equal to -0.002146 female^2. Similarly to births^2, this relates to a curvilinear relationship between female's in the workplace and happiness. Unfortunately, this implies that a one unit increase in female workplace percent^2 equates to a drop in the final happiness score by 0.002146.

Finally, the remaining coefficients relate to the LGBT freedom levels of a country. $\beta 7_{(x7)}$ is equivalent to 0.54 LGBT freedom (high). If a government has numerous laws protecting the equality of the LGBT community within its country, the happiness score will raise 0.54 points, with all other predictors staying fixed. $\beta 8_{(x8)}$ is equivalent to 0.296 LGBT freedom (some). When a country has only a few laws in place protecting the freedom and equality of their LGBT community, keeping all other predictors fixed, it's happiness score will increase by 0.296 as compared to countries with no laws protecting these rights. Finally, $\beta 0$ is the y-intercept in the model. 2.197 is the initial happiness score of a country before taking these predictors into account.

Conclusion

In this model, the significant indicators of a country's happiness score are based on three quantitative predictors, GDP per capita, births per 1000, and female workplace percentage, and one qualitative terms, LGBT freedoms- which is split into three categories (low, some, high). It

also takes all three of the quantitative predictors second-order terms into account when determining a final happiness score. With a model F-score of 55.07, P-value of < 0.000, and R2(adj) of 75.95, we can, with 99% confidence, explain around 76% the variance of the happiness score of a country based solely on these input. Furthermore, we can predict the future results of happiness of a country based on these same variables with an accuracy of around 74%, based on our final r-sq(pred) score. Finally, a standard error of .57 indicates that 95% of the actual results will fall within 1.14 points of the model's predicted score.

In regards to this regression, GDP and LGTB freedom are both strong, positive indicators of the happiness of a country, while births and female workplace tend to a slight, negative correlation with the final score once reaching their specific, curvilinear maxes. Although the female workplace^2's coefficient may bring the validity of full civil liberties into question, this regression does appear to imply that the stronger the economy and more robust the civil liberties of a country, the happier it's population.... as long as there are not too many babies being born.

References

The Economist Intelligence Unit. (2015). *Democracy Index 2015.* Retrieved from http://www.yabiladi.com/img/content/EIU-Democracy-Index-2015.pdf.

The World Bank. (2016). *World Bank Open Data*. Retrieved from http://data.worldbank.org/.

United Nations. (2016). Edited by John Helliwelll, Richard Layard, and Jeffrey Sachs. *World Happiness Report 2016.* Retrieved from http://worldhappiness.report/ed/2016/.

Zhong, Raymond. (2015). 'In Bhutan, Gross National Happiness Trumps Gross National Product.' *The Wall Street Journal.* Retrieved from http://www.wsj.com/articles/in-bhutan-gross-national-happiness-trumps-gross-national-product-1450318359.