What makes people in a country happy? Does higher GDP per capita makes people happier?

Charlene Ni, Emily Tseng, Naomi Wong

Dec 4, 2021

Abstract

What factors affect happiness the most? How does GDP per capita affect happiness in a country? We explore these questions by using data from the World Happiness Report, taking log of GDP per capita as variable of interest, and running regression models to see which model would be the best fit for our topic.

Our results suggest that, if people in a country have a higher GDP per capita, healthy years of life expectancy, social support, and freedom to make life decisions, then their happiness score would be higher. In addition to that, developed countries, with higher GDP per capita, do have higher happiness scores than developing countries.

Part I: Introduction and Data Description

The goal of our research was to uncover what factors affect world happiness and if countries with a higher GDP are happier. We hypothesized that countries with a higher GDP per capita would be happier and ran regressions to determine if this was indeed true. We also conducted multiple hypothesis tests and joint hypothesis tests on estimated coefficients to see whether they are statistically significant or not, to further decide the baseline regression model.

Data Source: World Happiness Report, https://worldhappiness.report/ed/2021/#appendices-and-data (https://worldhappiness.report/ed/2021/#appendices-and-data)

The World Happiness Report is a landmark survey of the state of global happiness, which reviews the state of happiness in the world and shows how to explain personal and national variations in happiness. As the first World Happiness Report published in 2012, happiness score is increasingly considered to be the proper measure of social progress and a useful way to guide public policy.

To figure out how happiness will be affected by economic behaviour, longevity, or even society corruption, we used the dataset from the World Happiness Report in 2017, which ranks 147 countries by their happiness levels, with 11 variables.

1. Country
2. Year
3. Cantril ladder(Happiness Score): 0-10, based entirely on the survey scores, using the Gallup weights to make the estimates representative.
4. Log GDP per capita: the natural log of GDP per capita
5. The time series of healthy life expectancy at birth: healthy life expectancy years based on World Health Organization (WHO) Global Health Observatory data repository.
6. Social support: national average of the binary responses (either 0 or 1) to the Gallup World Poll(GWP), “If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?”
7. Freedom to make life choices: national average of binary responses to the GWP question “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?”
8. Generosity: residual of regressing the national average of GWP responses to the question “Have you donated money to a charity in the past month?” on GDP per capita.
9. Corruption Perception: national average of the survey responses to two corruption related questions in GWP. The corruption perception at the national level is just the average response of the overall perception at the individual level.
10. Positive affect: defined as the average of three positive affect measures in GWP questions, which are happiness, laugh and enjoyment.
11. Negative affect: defined as the average of three negative affect measures in GWP questions, which are worry, sadness and anger.

We will only use six key factors to build up our model: **GDP per capita**, **healthy years of life expectancy**, **social support** (as measured by having someone to count on in times of trouble), **perceived freedom to make life decisions**, **corruption**(as measured by a perceived absence of corruption in government and business), and **generosity** (as measured by recent donations).

Part II: Descriptive Data Analysis

Descriptive Statistics Table

```
Descriptive Statistics
=======================================================
Statistic        N   Mean  Median  Max    Min   St. Dev.
-------------------------------------------------------
Score           147 5.460  5.553  7.788  2.662   1.141
log_GDP         146 9.397  9.504  11.634 6.817   1.181
Social_Support  146 0.805  0.829  0.967  0.320   0.122
Life_Expectancy 143 64.090 65.800 76.500 45.200  7.086
Freedom         146 0.781  0.813  0.985  0.427   0.127
Generosity      145 -0.007 -0.030 0.650  -0.290  0.158
Corruption      137 0.729  0.781  0.954  0.162   0.184
-------------------------------------------------------
```

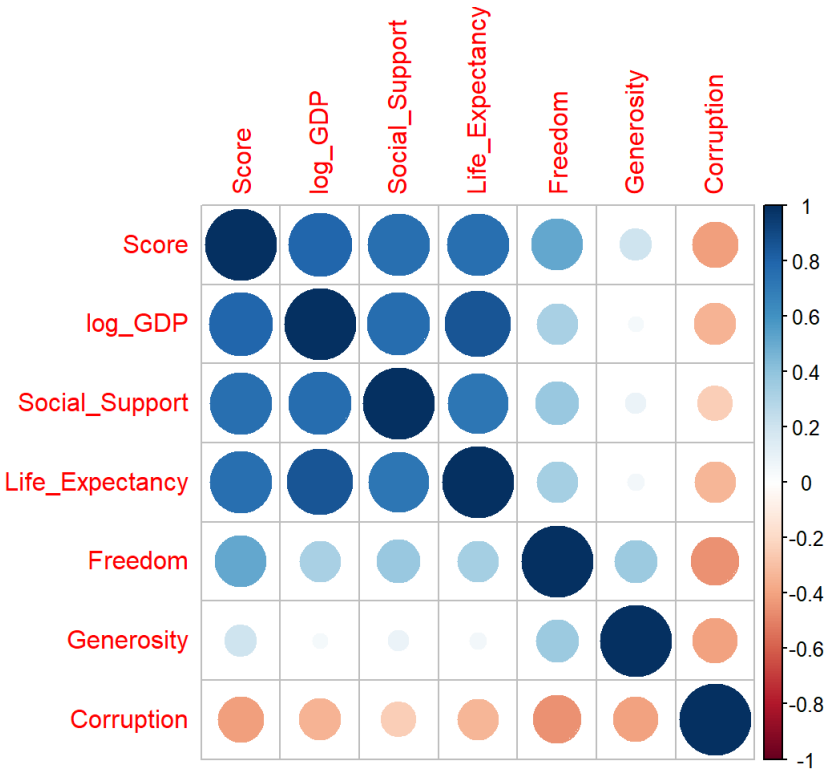
After further examination, we found that there are some countries having null values in the dataset as follows,

- log_GDP: South Sudan (drop)
- Social_Support: Vietnam
- Life_Expectancy: Hong Kong S.A.R. of China, Kosovo, Palestinian Territories, Taiwan
- Freedom: Vietnam
- Generosity: South Sudan, Vietnam
- Corruption: Bahrain, China, Egypt, Jordan, Kuwait, Saudi Arabia, Turkmenistan, United Arab Emirates, Vietnam, Yemen

In order to overcome this limitation, we ran our models through two different datasets. In the first, we dropped the countries from our dataset that had null values so that we could see the results without any data manipulation. Then, we filled these values with each column's mean value and ran our regressions again. After comparing the two results, we found that there was not much difference between the two (when we dropped the countries versus when we filled the null values). We decided to drop South Sudan only since it has null values in our variable of interest GDP per capita. That being said, the rest of our report shows the results from our dataset where we filled the null values with each column's mean.

In order to help us decide which variables to start with when building our initial regression, we created a correlation plot to show us which regressors were the most correlated with happiness scores.

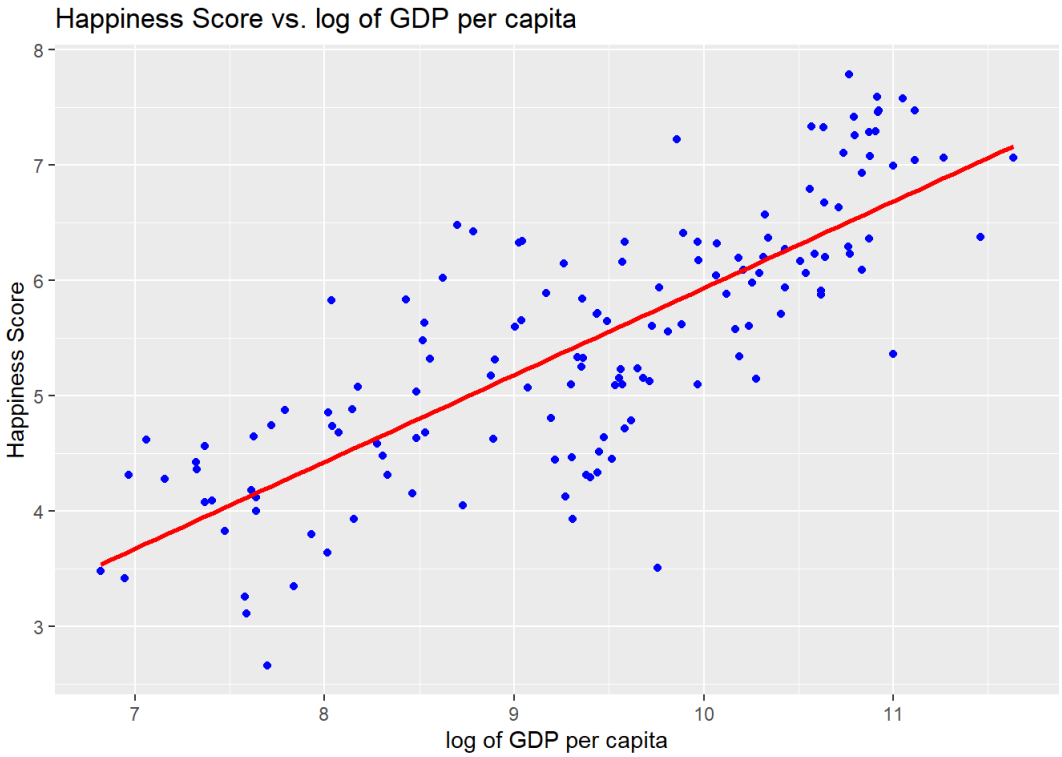
Correlation Chart



The correlation plot indicates that log GDP per capita, Social Support and Life Expectancy have stronger positive correlation with Happiness Score, around 0.6-0.8 respectively. While Freedom to make choices is less positively related to Happiness Score, and Generosity seems to have little correlation with it. We can also see that Correlation has a negative correlation with Happiness Score.

Next, we ran scatterplots on happiness score against each regression to see if the relationships were linear or if we needed to include non linear regressions. Below shows one of the plots we ran:

Scatterplot: log of GDP per capita against Happiness



The data already use log on GDP per capita against happiness score since it fits significantly better in this form, as the scatterplot shown above, the values indeed show relatively even around the correlation line with a positively strong correlation.

Part III: Baseline Regression Analysis

Baseline Specification

Happiness regression model result

	Dependent variable:					
	Score					
	(1)	(2)	(3)	(4)	(5)	(6)
log_GDP	0.753	0.494	0.340	0.340	0.344	0.331
	(0.044)	(0.070)	(0.106)	(0.098)	(0.096)	(0.100)
Social_Support		3.322	3.000	2.381	2.377	2.463
		(0.818)	(0.831)	(0.786)	(0.773)	(0.779)
Life_Expectancy			0.036	0.030	0.031	0.030
			(0.017)	(0.017)	(0.016)	(0.017)
Freedom				2.154	1.877	1.739
				(0.378)	(0.405)	(0.397)
Generosity					0.564	0.454
					(0.402)	(0.431)
Corruption						-0.333
						(0.364)
Constant	-1.598	-1.841	-2.422	-3.272	-3.115	-2.642
	(0.416)	(0.441)	(0.550)	(0.530)	(0.525)	(0.731)
Observations	146	146	146	146	146	146
R2	0.627	0.679	0.692	0.740	0.745	0.747
Adjusted R2	0.624	0.675	0.685	0.733	0.736	0.736
Residual Std. Error	0.689	0.641	0.630	0.581	0.577	0.577
F Statistic	241.794	151.356	106.119	100.297	81.953	68.468
Note:						NA

```
Linear hypothesis test

Hypothesis:
Generosity = 0
Corruption = 0

Model 1: restricted model
Model 2: Score ~ log_GDP + Social_Support + Life_Expectancy + Freedom +
    Generosity + Corruption

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F Pr(>F)
1    141
2    139  2 1.704 0.1857
```

Noting that we didn't use log on happiness score because the values could only range from 0 to 10.

According to the regression table, column (1) suggests that there is a positive effect of log GDP per capita on happiness score. However, we were concerned that column (1) may suffer from omitted variable bias. Thus, we decided to add the potential omitted variables that could affect happiness score and also be correlated with log GDP per capita in the subsequent regressions. As we had Social_Support, Life_Expectancy, Freedom, Generosity and Corruption into regression, the estimated coefficient on log_GDP changes from 0.753 (T score=17.11) to 0.331 (T score=3.31), between columns (2)-(6).

Although columns (5) and (6) of the tables have the highest adjusted R square (0.736), both generosity and corruption are not statistically significant, with a T score of 1.05 and |-0.91| which are smaller than the critical value of 1.65.

Therefore, we decided to do the F-Test. Based on the F-test below, we failed to reject the null hypothesis that both estimated coefficients on generosity and corruption are zero, given that F-test=1.704 < 3. All estimates are statistically significant at 10% or less significant level in column(4), and log_GDP, Social_Support, Life_Expectancy, Freedom explain 73.3% of the variation in happiness score.Thus, we chose the specification in column (4), as our baseline regression.

Based on our baseline regression, holding everything else equal:

- A country with an increase in GDP per capita by 1% will increase happiness score by 0.003 points (0.340/100) on average.
- A country with people having relatives/friends to help when getting in trouble will have, on average, 2.381 points (happiness score) higher than people who are unable to get help from relatives/friends.
- A country with an additional year of healthy life expectancy will increase the happiness score by 0.030 points.
- A country with freedom to make life choices will have, on average, 2.154 points (happiness score) higher than not having freedom to make life choices.
- F-test on estimated coefficients of generosity=0 and corruption=0:

Based on the F-test(1.704 < 3), we can conclude that both generosity and corruption are not statistically significant at the 5% level in our regression. However, we are not going to eliminate both variables immediately. We still want to see how the variables perform in the interaction terms in the next part.

Part IV: Alternative Regression

Explore Alternative Specifications

Regression model result: explore alternative sepc

	Dependent variable:				
	Score				
	(1)	(2)	(3)	(4)	(5)
log_GDP	0.340	0.322		0.814	
	(0.098)	(0.103)		(0.200)	
Developed			0.586		1.991
			(0.117)		(0.549)
Social_Support	2.381	2.501	3.121	2.459	2.850
	(0.786)	(0.799)	(0.743)	(0.717)	(0.667)
Life_Expectancy	0.030	0.029	0.045	0.036	0.044
	(0.017)	(0.017)	(0.012)	(0.017)	(0.011)
Freedom	2.154	1.882	1.873	1.985	1.985
	(0.378)	(0.383)	(0.404)	(0.399)	(0.420)
Corruption		-0.471	-0.280	7.100	1.065
		(0.332)	(0.335)	(2.507)	(0.628)
I(log_GDP * Corruption)				-0.738	
				(0.248)	
I(Developed * Corruption)					-1.868
					(0.676)
Constant	-3.272	-2.561	-1.361	-8.178	-2.268
	(0.530)	(0.731)	(0.798)	(1.995)	(0.830)
Observations	146	146	146	146	146
R2	0.740	0.744	0.752	0.760	0.767
Adjusted R2	0.733	0.735	0.743	0.750	0.757
Residual Std. Error	0.581	0.578	0.569	0.562	0.554
F Statistic	100.297	81.390	84.905	73.472	76.094
Note:					NA

Based on our baseline regression, we tested some alternative specifications to see if we could improve upon our model. Column 1 displays our baseline regression model.

Regression 2:

Although the corruption variable is not statistically significant, we still included it into the regression just in case of omitted variable bias. Including corruption causes the log of GDP coefficient to decrease, and meanwhile increasing the standard error of the coefficient. Thus, there should be no omitted variable bias on log GDP by adding corruption into our regression model. Since corruption is the only variable having a negative effect on happiness score, we would keep it for further exploration. Now, 73.5% of the variation in happiness score is explained by the included regressors in column 2.

Regression 3:

Given that our variable of interest is log of GDP, we also wanted to explore if there was a difference between developed and developing countries when it comes to happiness score. Economists consider a country developed if the GDP per capita is above $25,000 (https://www.investopedia.com/terms/d/developed-economy.asp (https://www.investopedia.com/terms/d/developed-economy.asp)). To make sense of this number in our analysis, we took the natural log of $25,000 which gave us a value of 10.13. We then created a dummy variable called, Developed which takes the value of 1 if a country has a log GDP greater than 10.13 (Developed), and 0 if the log GDP was less than 10.13 (Developing).

Column 3 tells us that the slope coefficient on Developed is 0.586, which means that developed countries have a happiness score 0.586 points greater than developing countries. This difference is statistically significant at the 1% significance level as the t score (0.586 / 0.117 = 5.01) is greater than the critical value of 2.57.

The fit of the model also increases when we use the Developed country dummy variable, as the adjusted R squared increases to 0.743. In this regression, 74.3% of the variation in happiness score is explained by the included regressors.

Regression 4:

In order to see whether the effects of log GDP per capita depends on different levels of corruption, we included an interaction term of log GDP and corruption, then calculated the resulting coefficient:

Estimated Score = -8.178 + 0.814*log_GDP + 7.100*Corruption - 0.738*(log_GDP * Corruption)

ΔScore/Δlog_GDP = 0.814 - 0.738*Corruption

If corruption = 0.1, slope of Score would be 0.74; if corruption is close to 1, such as 0.90, slope would be 0.15. Noting that with the national average of corruption increases, the effect of log_GDP on happiness score is getting smaller. In other words, there is a compensating effect between log_GDP and corruption.

The t-stat of the interaction term, |-0.738/0.248| = 2.98, is statistically significant at 1 % level.

Regression 5:

In order to see the effects of whether a country is developed or not based on different levels of corruption, we also include an interaction term of Developed and corruption. To interpret the effects, we first calculate the resulting coefficient if a country is developed or not (log GDP > $25,000):

Estimated Score = -2.268 + 1.991*Developed + 1.065*Corruption - 1.868*(Developed * Corruption)

- Estimated Happiness Score with Developed=1: -2.268 + 1.991*1 + 1.065*Corruption - 1.868*(1 * Corruption)
- Estimated Happiness Score with Developed=0: -2.268 + 1.991*0 + 1.065*Corruption - 1.868*(0 * Corruption)

So the difference between Developed countries and Developing countries would be, on average, 1.991-1.868*Corruption. And the t-stat of the interaction term, |-1.868/0.676| = 2.76, is statistically significant at 1 % level as well.

Given that this regression has the highest adjusted R squared value of 0.757, we determined that this model was the best fit, and we will use this model in our conclusion.

Part V: Results and Conclusion

Internal and external validity

How far can we generalize happiness results from the 2017 data? Threats to the external validity of our findings include differences in populations or settings such as the world in 2017 versus the world now in 2021, because of COVID19, the factors that affect people's happiness could have changed.

Threats to internal validity include omitted variable bias, errors in variables bias, sample selection bias and missing data, and simultaneous causality bias.

Our model potentially suffers from omitted variable bias as there are many factors that are determinants of happiness that are also correlated with the included regressors. For example, how polluted a country is or cost of living can also determine happiness and is possibly correlated with GDP, corruption, etc. However, including these omitted variables as regressors could also reduce the precision of our estimates. To fix this issue, we could try to collect more data, use instrumental variable regression, or run a randomized controlled experiment.

Our model is also subject to errors in variable bias given that our dataset has variables relating to surveys and there are recollection errors in surveys, ambiguous questions problems, and intentionally false report problems with surveys. For example, the generosity variable asks if people have donated to charity in the past month. Some people might lie and say they have in order to sound like a better person. Additionally the corruption perception might also be misleading because people might be scared to speak out against their governments. However, because our variable of interest is GDP per capita, these errors have little impact on our overall findings.

We also had the problem of missing data with some countries not having data. To combat this, we took the mean of each variable and used that value to fill the missing data. We also ran our models with a dataset where we dropped the countries that had null values and saw little difference in our results. Dropping these countries would also lead us to exclude some of the biggest countries in the world: China, Egypt, Jordan, Saudi Arabia, and the United Arab Emirates to name a few. However, because every country in our dataset has a happiness score, our results are not as affected.

Simultaneous causality bias is also a potential issue as world happiness could also affect the other regressors such as generosity, social support, and life expectancy. A potential solution to this would be to do a randomized controlled experiment in which the reverse causality channel is nullified.

Conclusion

To answer our original research question of what factors contribute to world happiness, our study found that social support and freedom to make life choices have the most impact. More specifically, for countries where the people say they have family and friends they can rely on if they need help, have a happiness score 2.850 points higher on average than those who do not. Additionally, countries where the people are satisfied with their freedom to decide what they want to do with their life have a happiness score 1.985 points higher on average than those who are not satisfied.

In terms of whether or not a higher GDP per capita affects happiness, our findings tells us that developed countries have a happiness score (1.991-1.868*Corruption) higher on average than developing countries. Therefore, having a higher GDP per capita does lead to a higher happiness score whether corruption exists or not.