# Identifying Measures of Happiness Across the Globe

By Tess Tao (3034199692), Jihee Yoon (3034007318), Yonggyu Choi(3033618956), Josephine Kim (30333952913), Ethan Qi (3033861328)

**Introduction**

Based on the World Happiness Project, fifteen variables arguably influence the happiness level and are measured by the life ladder. The scores of the life ladder are based on questions that ask respondents to rate their own current lives on an imaginary ladder with the best possible life being 10 and the worst being 0. Of the immense amount of data given, our team sought to answer whether any variable demonstrates a strong correlation in influencing the happiness levels of both most and least happy countries.

- Which variable(s) help shape the most happy countries?
- Which variable(s) help shape the least happy countries?
- How have some countries changed over time? Has there been a positive or negative trend?

To narrow our focus, we used 2017 data since it is closer to the present. To determine which variables play a more significant role in shaping the most and least happy countries, we decided to split the data into two: thirty countries with the highest life ladder values (referred to as "most.happy" data) and thirty countries with the lowest life ladder values (referred to as "least.happy" data). We chose thirty as the sample size because we wanted a small enough sample to represent the most happy and least happy data, while still meeting the $n > 25$ rule.

Of the fifteen variables, we did not consider "Democratic Quality," "Delivery Quality," and "Gini index" because the data column in 2017 had considerable amount of "NA's." Since too many countries were missing for these variables, we concluded that finding the correlation of these variables and the life ladder may yield biased, inaccurate representation of the correlations. Thus, we decided to focus on twelve variables.

We found the correlation coefficient value (r) between each of the twelve variables and the life ladder for the "most.happy" and "least.happy" data sets. Based on these r-values, we made an educated observation that five of these variables played the most influential role in shaping different happiness levels in the countries: log GDP per capita, social support, health life expectancy, freedom to make life choices, and generosity.

**Exploratory Data Analysis**

1. Defining the 12 variables

1. **Life Ladder**: This is the happiness score or subjective well-being from a scale of 1 to 10.
2. **Log GDP per capita**: The statistics of GDP per capita in purchasing power parity (PPP) at constant 2011 international dollar prices; economic output that accounts for number of people.
3. **Social support**: The national average of the binary responses (either 0 or 1) to the Gallup World Poll (GWP) question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?"
4. **Healthy life expectancy at birth**: Calculated by the authors based on data from WHO and WDI published in journal articles.
5. **Freedom to make life choices**: Average of binary responses to the question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"
6. **Generosity**: The 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.
7. **Perceptions of Corruption**: Average of the survey responses to two questions in the GWP: "Is corruption widespread throughout the government or not and Is corruption widespread within businesses or not?"
8. **Positive affect**: The average binary responses of two positive affect measures: laughter and enjoyment.
9. **Negative affect**: The average of previous-day affect measures for worry, sadness, and anger for all waves.
10. **Confidence in national government**: Average of the answers to the question: "Do you have confidence in each of the following, or not? How about the national government?"
11. **Gini of household income**: Report of household income in local currency. Income variables are created by converting local currency to International Dollars (ID) using purchasing power parity (PPP) ratios
12. **SD of ladder**: Standard Deviation of variable 1, 'Life Ladder'
13. **SD/mean of ladder**: Standard Deviation divided by Mean of variable 1, 'Life Ladder'

2. Summary Statistics per Variable for *'most.happy'* and *'least.happy'*

Given the data shown below, it is apparent that the mean and medians for most variables seem to be higher for most.happy than least.happy data, with the exceptions of perceptions of corruption, negative affect, confidence in national government, and gini of household income. The standard deviation is the opposite—least.happy data shows a larger standard deviation than most.happy, with the exception of perceptions of corruption. This can be interpreted through variability; the most.happy data has more values closer to the mean, so there's less variability of the factors between different countries, while the least.happy data has less values closer to the mean, which can be assumed to be more variability and a potentially wider spread. An interesting variable that seems to have a large difference between most.happy and least.happy is generosity; most.happy has a mean eight times that of least.happy, and a median thirteen times that of least.happy data. This may indicate that generosity is an influential factor of happiness in different countries.

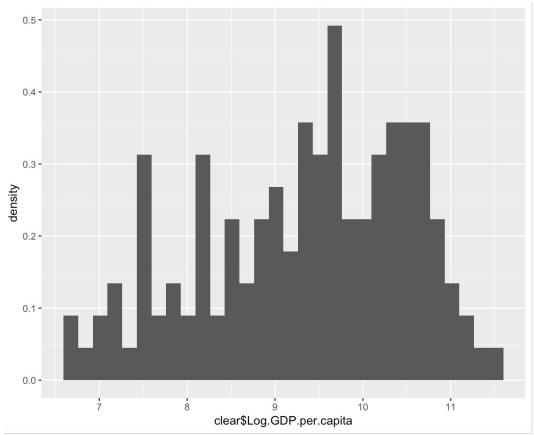
| Variable | Mean most.happy | Mean least.happy | Median most.happy | Median least.happy | SD most.happy | SD least.happy |
|---|---|---|---|---|---|---|
| Life Ladder | 6.992268 | 3.964019 | 7.060768 | 4.106953 | 0.437974 | 0.490033 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Log GDP per capita | 10.4614 | 8.134811 | 10.65765 | 8.010255 | 0.706502 | 0.978146 |
| Social Support | 0.914383 | 0.676701 | 0.921321 | 0.696517 | 0.04143 | 0.113528 |
| Health Life Expectancy at Birth | 70.75242 | 56.222 | 71.64125 | 56.08046 | 2.677246 | 5.754152 |
| Freedom to make life choices | 0.886676 | 0.709588 | 0.907946 | 0.725246 | 0.069636 | 0.137996 |
| Generosity | 0.087127 | 0.014064 | 0.135308 | 0.011871 | 0.155856 | 0.190412 |
| Perceptions of Corruption | 0.524827 | 0.778427 | 0.518304 | 0.771036 | 0.238518 | 0.096375 |
| Positive Affect | 0.801673 | 0.63032 | 0.806831 | 0.614027 | 0.049787 | 0.097821 |
| Negative Affect | 0.215351 | 0.352111 | 0.208585 | 0.337435 | 0.044323 | 0.104669 |
| Confidence in National Government | 0.53012 | 0.569346 | 0.453407 | 0.620333 | 0.19081 | 0.218789 |
| Gini of Household Income | 0.405014 | 0.526836 | 0.391934 | 0.551279 | 0.099557 | 0.106518 |
| SD of Ladder | 1.804117 | 1.682604 | 0.36771 | 2.42397 | 2.513597 | 0.425593 |
| SD/mean of Ladder | 0.260975 | 0.242221 | 0.065977 | 0.620269 | 0.597858 | 0.133226 |

3. Visualization of summary graphics and statistics

| Variable | Most.Happy | Least.Happy | Variable | Most.Happy | Least.Happy |
|---|---|---|---|---|---|
| Log GDP per capita | 0.5686501 | 0.1445162 | Positive Affect | 0.02330434 | 0.2064176 |
| Social Support | 0.6240832 | 0.3165284 | Negative Affect | -0.537591 | 0.1942999 |
| Health Expectancy at Birth | 0.5818028 | 0.1420573 | Confidence in National Government | 0.2167093 | -0.07401538 |
| Freedom to make life choices | 0.4417639 | 0.2503029 | SD of Ladder by Country Year | -0.6790788 | 0.2898541 |
| Generosity | 0.6079077 | 0.02442598 | SD Mean of Ladder by Country Year | -0.7873542 | -0.34132 |
| Perceptions of Corruption | -0.56035 | 0.01048949 | Gini Household Income | -0.09951294 | 0.2419223 |

4. Data Distribution




This histogram displays the GDP of all countries from 2017. The GDP variable's histogram was the most symmetrical out of the five variables. This histogram is a visual representation of the variable closest to a normal distribution from our data, yet it does not provide a definite normal shape. Though GDP's histogram can be described as "roughy" normal, it still has a left skew; since GDP has the closest to a normal distribution, this indicates that the other variables are less normally distributed. Thus, the analysis of our data is not precise. Though our analysis uses a population with n>25, it is not reliable for accuracy. However, we proceed with the hypothesis test under the assumption that the slope of the regression line follows the normal curve.

**Data Analysis**

**1. Graphs & Findings on the Improvement Between 2016 & 2017**

With data from both 2016 and 2017, we are able to compare countries across two years and identify countries whose national level of happiness increased or decreased over time. To do so, we created a dataframe referred to as "perfect 2" which only encompasses countries that have data for both years. After finding the countries with a positive difference between the values of the life ladders in both years, we generated a list of 80 countries (referred to as "rise") whose happiness level has shown improvement in 2017 (displayed from maximum to minimum increase):

| | X | country |
|---|---|---|
| 1 | 97 | Guinea |
| 2 | 146 | Liberia |
| 3 | 91 | Ghana |
| 4 | 26 | Benin |
| 5 | 45 | Central African Republic |
| 6 | 138 | Kyrgyzstan |
| 7 | 55 | Congo (Brazzaville) |
| 8 | 160 | Mali |
| 9 | 247 | Tajikistan |
| 10 | 105 | Hungary |
| 11 | 166 | Mauritius |
| 12 | 47 | Chad |

| | X | country |
|---|---|---|
| 13 | 274 | Uzbekistan |
| 14 | 123 | Ivory Coast |
| 15 | 253 | Togo |
| 16 | 99 | Haiti |
| 17 | 188 | Nicaragua |
| 18 | 201 | Panama |
| 19 | 249 | Tanzania |
| 20 | 38 | Burkina Faso |
| 21 | 156 | Madagascar |
| 22 | 150 | Lithuania |
| 23 | 134 | Kosovo |
| 24 | 190 | Niger |

| | X | country |
|---|---|---|
| 25 | 22 | Belarus |
| 26 | 227 | Slovakia |
| 27 | 101 | Honduras |
| 28 | 69 | Dominican Republic |
| 29 | 130 | Kazakhstan |
| 30 | 225 | Singapore |
| 31 | 174 | Montenegro |
| 32 | 77 | Estonia |
| 33 | 264 | Ukraine |
| 34 | 197 | Pakistan |
| 35 | 258 | Turkey |
| 36 | 268 | United Kingdom |

| | X | country |
|---|---|---|
| 37 | 172 | Mongolia |
| 38 | 63 | Cyprus |
| 39 | 210 | Portugal |
| 40 | 36 | Bulgaria |
| 41 | 42 | Cameroon |
| 42 | 14 | Austria |
| 43 | 121 | Italy |
| 44 | 229 | Slovenia |
| 45 | 148 | Libya |
| 46 | 266 | United Arab Emirates |
| 47 | 164 | Mauritania |
| 48 | 89 | Germany |

| | | |
|---|---|---|
| 49 | 75 | El Salvador |
| 50 | 270 | United States |
| 51 | 119 | Israel |
| 52 | 272 | Uruguay |
| 53 | 206 | Philippines |
| 54 | 83 | France |
| 55 | 136 | Kuwait |
| 56 | 4 | Albania |
| 57 | 81 | Finland |
| 58 | 40 | Cambodia |
| 59 | 212 | Romania |
| 60 | 277 | Vietnam |

| | | |
|---|---|---|
| 61 | 192 | Nigeria |
| 62 | 186 | New Zealand |
| 63 | 152 | Luxembourg |
| 64 | 59 | Costa Rica |
| 65 | 219 | Senegal |
| 66 | 162 | Malta |
| 67 | 132 | Kenya |
| 68 | 113 | Iran |
| 69 | 18 | Bahrain |
| 70 | 65 | Czech Republic |
| 71 | 115 | Iraq |
| 72 | 208 | Poland |

| | | |
|---|---|---|
| 73 | 141 | Latvia |
| 74 | 67 | Denmark |
| 75 | 117 | Ireland |
| 76 | 243 | Switzerland |
| 77 | 204 | Peru |
| 78 | 12 | Australia |
| 79 | 32 | Botswana |
| 80 | 87 | Georgia |

After finding the countries with a negative difference between values of the life ladders in both years, we generated a list of 55 countries (referred to as "fall") whose happiness levels had decreased in 2017 (shown from maximum to minimum decrease):

| | X | country |
|---|---|---|
| 1 | 195 | Norway |
| 2 | 24 | Belgium |
| 3 | 95 | Guatemala |
| 4 | 107 | Iceland |
| 5 | 10 | Armenia |
| 6 | 111 | Indonesia |
| 7 | 34 | Brazil |
| 8 | 126 | Japan |
| 9 | 85 | Gabon |
| 10 | 158 | Malawi |
| 11 | 236 | South Sudan |
| 12 | 61 | Croatia |

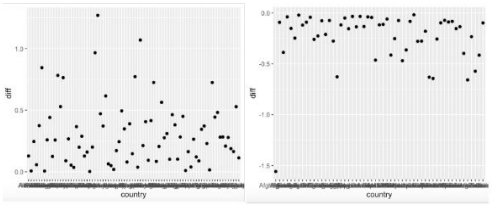
| | X | country |
|---|---|---|
| 13 | 176 | Morocco |
| 14 | 53 | Colombia |
| 15 | 184 | Netherlands |
| 16 | 241 | Sweden |
| 17 | 238 | Spain |
| 18 | 30 | Bosnia and Herzegovina |
| 19 | 6 | Algeria |
| 20 | 234 | South Korea |
| 21 | 283 | Zimbabwe |
| 22 | 154 | Macedonia |
| 23 | 143 | Lebanon |
| 24 | 79 | Ethiopia |

| | X | country |
|---|---|---|
| 25 | 28 | Bolivia |
| 26 | 109 | India |
| 27 | 251 | Thailand |
| 28 | 103 | Hong Kong S.A.R. of China |
| 29 | 16 | Azerbaijan |
| 30 | 245 | Taiwan Province of China |
| 31 | 93 | Greece |
| 32 | 217 | Saudi Arabia |
| 33 | 57 | Congo (Kinshasa) |
| 34 | 51 | China |
| 35 | 262 | Uganda |
| 36 | 20 | Bangladesh |

| | X | country |
|---|---|---|
| 37 | 170 | Moldova |
| 38 | 232 | South Africa |
| 39 | 49 | Chile |
| 40 | 71 | Ecuador |
| 41 | 214 | Russia |
| 42 | 199 | Palestinian Territories |
| 43 | 182 | Nepal |
| 44 | 8 | Argentina |
| 45 | 256 | Tunisia |
| 46 | 168 | Mexico |
| 47 | 281 | Zambia |
| 48 | 128 | Jordan |

| | | |
|---|---|---|
| 49 | 179 | Myanmar |
| 50 | 279 | Yemen |
| 51 | 73 | Egypt |
| 52 | 221 | Serbia |
| 53 | 223 | Sierra Leone |
| 54 | 260 | Turkmenistan |
| 55 | 2 | Afghanistan |

Based on the scatterplot of both positive and negative differences, we conclude that on average, 80 countries have gained an increase in their life ladder values by an average of 0.31, while the other 55 countries have decreased in their life ladder values by an average of 0.23. Overall, most countries did not vary significantly from 2016 to 2017.




Increases in happiness level (left scatterplot), Decrease in happiness level (right scatterplot)

---

**Focusing on the Five Most Influential Variables**

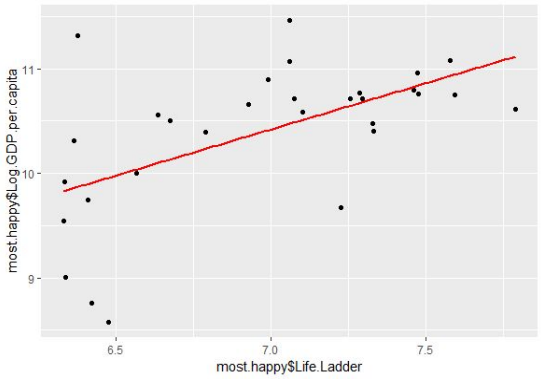
To narrow our focus from the twelve variables, we tried to find the five most influential variables based on higher r-values. We noticed that r-values of the most.happy data seemed to be higher than r-values of least.happy data. Initially, since there were higher r-values in the most.happy data, we decided to look for variables with the highest most.happy r-values with relatively reasonable least.happy r-values. This left us with GDP, Social Support, Health Life Expectancy, Freedom to make decisions, and Generosity.

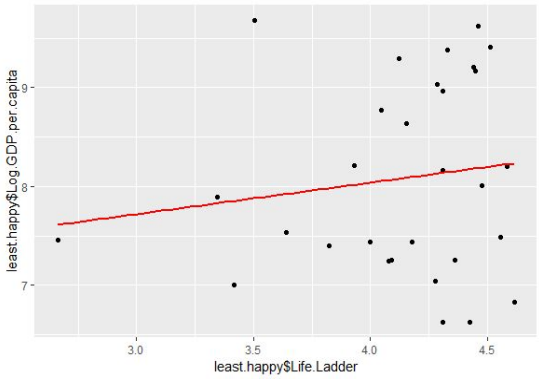
We did not believe that the remaining seven variables below influenced happiness level as significantly: perceptions of corruption, positive affect, negative affect, confidence in government, SD of ladder, SD mean, and Gini household income. Firstly, some variables, such as positive affect, have an overall weak correlation so they are an improper measurement to conclude an association with the life ladder. Secondly, the r-values comparing the most and least happy data may have opposite directions. For instance, for perceptions of corruption, though the most happy data had a moderate negative r value, the least happy data had a very weak positive correlation making the correlation inconclusive. Thirdly, we wanted to use variables in society which contribute to different happiness levels, as opposed to the SD of ladder or SD of mean which were simply statistical derivations from the life ladder.

| Variable | Most.Happy | Least.Happy | Variable | Most.Happy | Least.Happy |
|---|---|---|---|---|---|
| Perceptions of Corruption | -0.56035 | 0.01048949 | SD of Ladder by Country Year | -0.6790788 | 0.2898541 |
| Positive Affect | 0.02330434 | 0.2064176 | SD Mean of Ladder by Country Year | -0.7873542 | -0.34132 |
| Negative Affect | -0.537591 | 0.1942999 | Gini Household Income | -0.09951294 | 0.2419223 |
| Confidence in National Government | 0.2167093 | -0.07401538 | N/A | N/A | N/A |

In our linear regression graphs, we decided to label our x-axis as "life ladder" and our y-axis as the tested variable. Since the x-axis is the true metric of success in this analysis, we believed it best for this response variable (life ladder) to remain stagnant. Thus, the y-axis represented the 12 variables so that the slopes would be horizontal and easily comparable with the graphs that are typically used in statistical studies.

**GDP**




Most Happy Linear Regression Line:
Log GDP per Capita = 0.8840634*Life Ladder + 4.2307097
Least Happy Linear Regression Line:
Log GDP per Capita = 0.3186615*Life Ladder + 6.7628293
Top r= 0.5686501
Bottom r= 0.1445162

Analysis of the variable:
For the most happy countries,

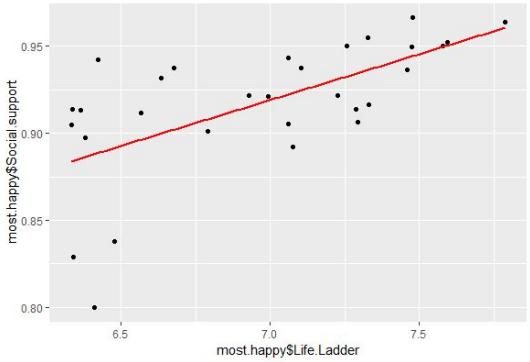
- r=0.57 shows a moderate positive correlation between the life ladder (happiness level) and GDP per capita
- fourth highest correlation coefficients among all factors
- most points tightly follow the line of best fit as their life ladder gets larger despite a few outliers in the bottom left

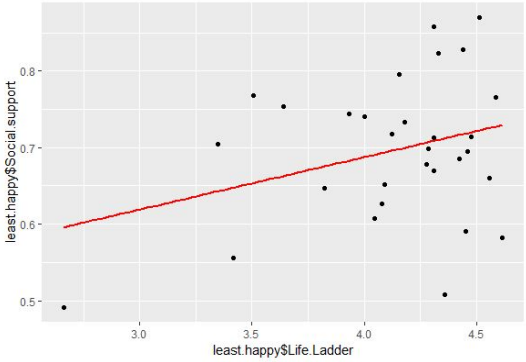
For the least happy countries,

- r= 0.14 conveys a very weak positive correlation between happiness level and GDP for the least happiest countries
- a much wider spread around the regression line, as most points fall in the top and bottom areas

GDP is one of the most decisive factors for happiness for countries with a life ladder above 6.5.

**Social Support**




Most Happy Linear Regression Line:
Social Support = 0.05269481 * Life Ladder + 0.555027062
Least Happy Linear Regression Line:
Social Support =0.0683403 * Life Ladder + 0.4139990
Top r= 0.6240832
Bottom r= 0.3165284

Analysis:
For the most happy countries,

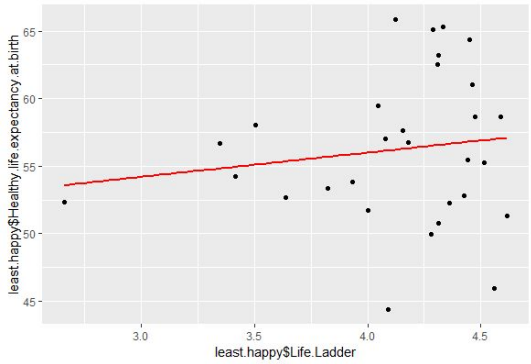
- r= 0.62 shows a stronger positive correlation between the happiness level and social support
- highest correlation coefficients among all factors
- most points, especially the upper quartile, tightly follow the regression line in the plot

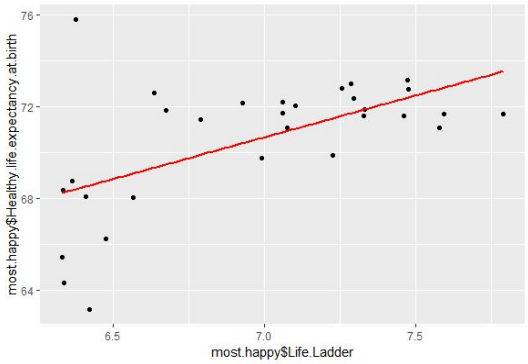
For the least happy countries,

- r= 0.32 indicates a weak positive correlation between the happiness level and social support
- points in the plot tend to be have a wider spread and are farther from the line of best fit

Social Support is one of the most influential factors with its relatively higher correlation coefficient than other variables.

**Healthy Life Expectancy**




Most Happy Linear Regression Line:
Healthy life expectancy at birth = 3.64231 * Life Ladder + 45.17445
Least Happy Linear Regression Line:
Healthy life expectancy at birth = 1.793544 * Life Ladder + 48.823797
Top r= 0.5818028
Bottom r= 0.1420573

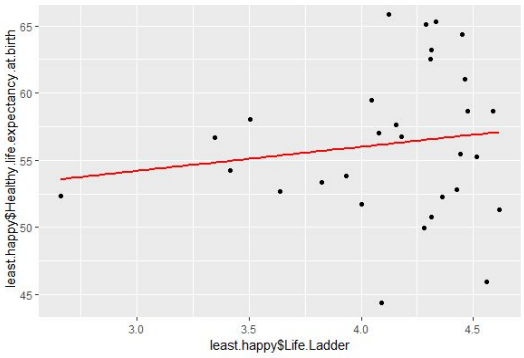
Analysis:
For the most happy countries,

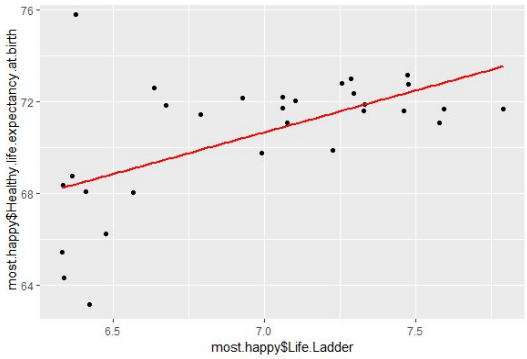
- r= 0.5818028 shows a moderate positive correlation (third highest) between happiness level and health life expectancy
- approximately fair distribution of points above and below line of best fit, the data points are spread mostly evenly, slope is positive

For the least happy countries,

- r= 0.1420573 indicates a very weak positive correlation between the happiness level and health life expectancy
- very even distribution of plots both above and below the line of best fit

**Freedom to Make Life Choices**




Most Happy Linear Regression Line:
Freedom to make life choices = 0.06589726 Life Ladder + 0.42554051
Least Happy Linear Regression Line:
Freedom to make life choices = 0.07611076 Life Ladder + 0.42106380
Top r = 0.4417639
Bottom r = 0.2503029

Analysis of the variable:
For the most happy countries,

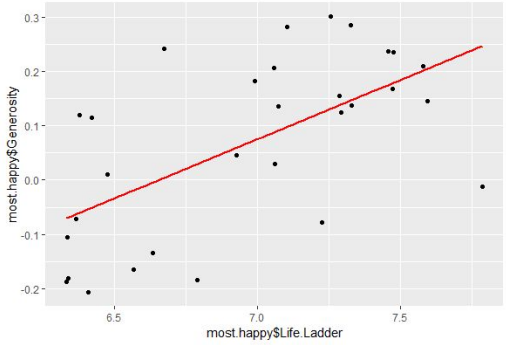
- r= 0.44 shows a weak positive correlation between happiness level and freedom to make life choices
- approximately fair distribution of points above and below line of best fit, but still considerable outliers far from line

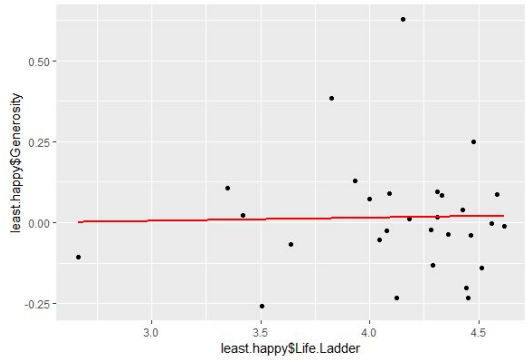
For the least happy countries,

- r= 0.25 conveys a weaker positive correlation between happiness level and freedom;
- wider spread with more outliers; more points seem to be in the extreme areas on the graph rather than directly above or below the line of best fit

Freedom does not appear to heavily influence happiness level due to its lower r-value. We still included this variable as one of the top five variables because of its relatively higher r-value than the other seven variables.

**Generosity**




Most Happy Linear Regression Line:
Generosity = 0.2177198*Life Ladder + 0.2177198
Least Happy Linear Regression Line:
Generosity = 0.01001219*Life Ladder + -0.02549099
Top r = 0.6079077
Bottom r= 0.02442598

Analysis of the variable:
For the most happy countries,

- r= 0.608 shows a strong positive correlation between the life happiness level and generosity
- second highest correlation coefficients among all factors
- fair amount of points lie above and below the line of best fit, but the data points become more dense as both variables increase, hence the strong positive correlation

For the least happy countries,

- r= 0.024 conveys a very weak positive correlation between happiness level and generosity
- slightly more points below the line of best fit than above it

Generosity is one of the more decisive factors for happiness for the more.happy countries, given its high r value.

---

**Hypothesis Test**
For all five variables,

- Because all samples are > 25, the slope of the regression line will basically follow the normal distribution.
- The null hypothesis states that the slope of the regression line is only the result of chance, and there is no correlation between the two variables we are measuring.
- The alternative hypothesis states that the relationship between the two variables cannot be explained by chance, and that the slope of regression line shows a positive/negative correlation between variables.

1. **GDP Null Hypothesis test**

| | |
|---|---|
| **Most happy data**:<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 4.2307 1.6875 2.507 0.01825 *<br>Life.Ladder 0.8841 0.2417 3.658 0.00104 **<br>For the life ladder, the test statistic t= 3.658, and the p-value= 0.00104. Since 0.00104 < 0.01, we likely reject the null hypothesis at the 1% level. Though the hypothesis test does not allow us to decide where there are confounding variables, we know that this relationship is not due to chance and GDP does influence the happiness level for the most happy countries. | **Least happy data:**<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 6.7628 1.7098 3.955 0.000474 ***<br>Life.Ladder 0.3187 0.4123 0.773 0.446104<br>For the life ladder, the test statistic t= 0.773, and the p-value= 0.446104. Since 0.446104 > 0.05, we fail to reject the null hypothesis. This relationship is due to chance and GDP does not influence the happiness level for the least happy countries. |

2. **Social Support Null Hypothesis test**

| | |
|---|---|
| **Most happy data**:<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 0.55027 0.08706 6.321 7.75e-07 ***<br>Life.Ladder 0.05269 0.01247 4.226 0.000228 ***<br>For the life ladder, the test statistic t= 4.226, and the p-value= 0.00023. Since 0.00023 < 0.01, we likely reject the null hypothesis at the 1% level. This relationship is not due to chance, social support does influence happiness level for most happy countries. | **Least happy data:**<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 0.41400 0.16049 2.580 0.0154 *<br>Life.Ladder 0.06834 0.03870 1.766 0.0884<br>For the life ladder, the test statistic t= 1.766, and the p-value= 0.0884. Since 0.0884 > 0.05, we fail to reject the null hypothesis. This relationship is due to chance, social support does not influence happiness level for least happy countries. |

3. **Healthy Life Expectancy at Birth Null Hypothesis test**

| | |
|---|---|
| **Most happy data**:<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 45.1745 6.7187 6.724 2.67e-07 ***<br>Life.Ladder 3.6423 0.9623 3.785 0.000745 ***<br>For the life ladder, the test statistic t= 3.785, and the p-value= 0.000745. Since 0.000745 < 0.01, we likely reject the null hypothesis at the 1% significance level. This relationship is not due to chance and healthy life expectancy does influence the happiness level for the most happy countries. | **Least happy data:**<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 48.824 9.793 4.985 2.89e-05 ***<br>Life.Ladder 1.794 2.362 0.759 0.454<br>For the life ladder, the test statistic t= 0.759, and the p-value= 0.454. Since 0.454 > 0.05, we fail to reject the null hypothesis. This relationship is due to chance and health life expectancy does not influence the happiness level for the least happy countries. |

4. **Freedom to Make Life Choices Null Hypothesis test**

| | |
|---|---|
| **Most happy data:**<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 0.42554 0.17658 2.410 0.0228 *<br>Life.Ladder 0.06590 0.02529 2.606 0.0145 *<br>For the life ladder, the test statistic t= 2.606, and the p-value= 0.0145. Since 0.0145 < 0.05, we likely reject the null hypothesis at the 5% significance level. This relationship is not due to chance and freedom to make life choices does influence the happiness level for the most happy countries. | **Least happy data:**<br>Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) 0.42106 0.23070 1.825 0.0787 .<br>Life.Ladder 0.07611 0.05564 1.368 0.1822<br>For the life ladder, the test statistic t= 1.368, and the p-value= 0.1822. Since 0.1822 > 0.05, we fail to reject the null hypothesis. This relationship is due to chance and freedom to make life choices does not influence the happiness level for the least happy countries. |

5. **Generosity Null Hypothesis test**

| **Most happy data:** | **Least happy data:** |
|---|---|
| Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) -1.44894 0.37524 -3.861 0.000609 ***<br>Life.Ladder 0.21772 0.05374 4.051 0.000366 ***<br>For the life ladder, the test statistic t= 4.051, and the p-value= 0.00037. Since 0.00037 < 0.01, we likely reject the null hypothesis at a 1% level. This relationship is not due to chance, generosity does influence happiness level for most happy countries. | Coefficients:<br>Estimate Std. Error t value Pr(>\|t\|)<br>(Intercept) -0.02549 0.32111 -0.079 0.937<br>Life.Ladder 0.01001 0.07744 0.129 0.898<br>For the life ladder, the test statistic t= 0.129, and the p-value= 0.898. Since 0.898 > 0.05, we fail to reject the null hypothesis. This relationship is due to chance and generosity does not influence the happiness level for the least happy countries. |

**Conclusion**

Before stating a final conclusion, we want to first clarify several limitations in our data that might weaken the strength of our analysis. First of all, when exploring the influence of different factors on national happiness level, we omitted data on "Democratic quality," "Delivery quality,"and "GINI index" due to the considerable portions of "NA" in the data. Although this step was necessary for our purpose of analysis, this does not mean a relationship cannot be established between these three variables and the happiness level. Secondly, in order to differentiate between the influence of variables on the most and least happy countries, we separated the dataset into two groups of thirty countries. Although the "n>25" rule has been satisfied, the sample size is still relatively small. This might contribute to attenuation, or the underestimation of correlation simply due to a small sample size. This may explain why the variables only exhibit a moderate correlation with the happiness level that is below 0.7. The distribution of the slope of the regression line fails to closely follow the normal curve.

Despite these limitations, valid conclusions still can be made from data analysis. The first discovery is that after comparing the happiness levels for 2016 and 2017, we observed that there are more countries (88) that demonstrated an increase in overall happiness over the years than countries (50) with decreased national happiness. On a worldwide level, it appears that more countries have increased national happiness levels. After plotting the scatterplot for countries with increased or decreased happiness levels over time, we concluded that both scatterplots are strongly skewed toward the line y=0, indicating that the differences between 2016 and 2017 are not significant. This demonstrates that for most countries, the happiness level did not vary strongly, and rather remained relatively stable over time. Such consistency in data ensures the life ladder to be a convincing measure of happiness.

The second discovery was found by taking a closer look at the relationship between happiness level and the variables via linear regression. Based on correlation coefficients between happiness level and the twelve variables for the most happy and the least happy countries, we observed that the least happy countries generally exhibited weaker correlations for the variables, and thus choose to focus on the most happy countries. We hypothesized that the five most influential factors on the level of happiness are GDP, social support, healthy life expectancy, freedom to make decisions, and generosity based on the relatively high r-values of most happy countries, and reasonably high r-values of least happy countries.

In order to prove the legitimacy of these correlations, we then conducted five hypothesis tests on the slope of the regression line for these five variables. As expected, for the least happy countries, none of the correlations between the variables and life ladder had a p-value less than 0.05. Thus, we fail to reject the null hypothesis for the least happy countries and conclude it is very likely that the observed relationships are only due to chance; we cannot conclude any correlation between happiness level and the five variables. There may be confounding factors influencing the happiness levels to have a poor correlation. For instance, people in less developed countries may have a less educated background, resulting in an inconsistency in answering the questions appropriately. Also, the study must account for the subjectivity in the participants' responses to reflect happiness levels.

On the contrary, the most happy countries had a p-value <0.05 for all 5 variables. The correlations for GDP, social support, healthy life expectancy, and generosity are highly significant on 1% level, and the correlation for freedom to make decisions is statistically significant at a 5% level. Thus, these four variables are likely the most significant in influencing the happiness level of most happy countries.

Based on our research, we are unable to draw the conclusion that there is a certain variable which influences a noticeable difference between both the most and least happy countries in happiness levels, since only the most happy countries yielded probabilities from the hypothesis test yielded below 0.05. Also, due to the way we analyzed our data, the highest correlations were near 0.6 and 0.7. We conclude that GDP, Social Support, Healthy life expectancy, and generosity influence the happiness level most significantly in the most happy countries in the world.

Some possible reasoning for these results may be that GDP establishes the income level and efficiency of the country's inhabitants, so a higher GDP may correlate with greater happiness levels. Social support may yield more happiness since having a network of family and friends is important to a successful life. Healthy life expectancy may yield more happiness because access to a healthy life is fundamental to longevity. Generosity may influence happiness levels since more financial support to charities indicate more care for others and thus more happiness. The fifth variable, freedom, analyzed distinctively at the 5% level, may influence happiness level since it demonstrates self satisfaction with the ability to make life decisions.

**References**

"Data 2018." *World Happiness Project*, 2018.

Ibser, Hank. "Hypothesis Test for Slope of Regression Line." PDF. UC Berkeley, Dec. 1, 2018.

Ibser, Hank. "Project Intros." PPT. UC Berkeley, Nov. 21 2018.

Project, World Happiness. *World Happiness Report 2018*. WER, 2018, pp. 3–157, *World Happiness Report 2018*