Global Happiness Analysis

Arvind Kandala, Ammy Lin, Lingyue Hao, Shelly Cao

Load data

```
Rows: 782
Columns: 8
$ country          <chr> "Switzerland", "Iceland", "Denmark", "Norway", "Canad~
$ region           <chr> "Western Europe", "Western Europe", "Western Europe",~
$ score            <dbl> 7.587, 7.561, 7.527, 7.522, 7.427, 7.406, 7.378, 7.36~
$ gdp_index        <dbl> 1.39651, 1.30232, 1.32548, 1.45900, 1.32629, 1.29025,~
$ lifeexp_index    <dbl> 0.94143, 0.94784, 0.87464, 0.88521, 0.90563, 0.88911,~
$ freedom_index    <dbl> 0.66557, 0.62877, 0.64938, 0.66973, 0.63297, 0.64169,~
$ generosity_index <dbl> 0.29678, 0.43630, 0.34139, 0.34699, 0.45811, 0.23351,~
$ year             <int> 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015,~
```

```
      country          region score gdp_index lifeexp_index freedom_index
1 Switzerland Western Europe 7.587   1.39651       0.94143       0.66557
2     Iceland Western Europe 7.561   1.30232       0.94784       0.62877
3     Denmark Western Europe 7.527   1.32548       0.87464       0.64938
4      Norway Western Europe 7.522   1.45900       0.88521       0.66973
5      Canada  North America 7.427   1.32629       0.90563       0.63297
6     Finland Western Europe 7.406   1.29025       0.88911       0.64169
  generosity_index year
1          0.29678 2015
2          0.43630 2015
3          0.34139 2015
4          0.34699 2015
5          0.45811 2015
6          0.23351 2015
```

Research Question 1: Life Expectancy, GDP, and Happiness

Is there a significant association between life expectancy and national happiness, after adjusting for GDP per capita?

Model Specification:

We fit a multiple linear regression model with a national happiness score as the outcome; a linear regression model is appropriate here because the outcome (happiness score) is continuous and approximately normally distributed. The linear model relies on several core assumptions: (1) the relationship between predictors and the outcome is linear, (2) residuals have constant variance across fitted values, (3) residuals are independent, and (4) residuals are approximately normally distributed.

The formal model is specified as follows:

$$\text{Score}_i = \beta_0 + \beta_1(\text{LifeExp}_i) + \beta_2(\text{GDP}_i) + \beta_3(\text{LifeExp}_i \times \text{GDP}_i) + \epsilon_i$$

Life expectancy index and GDP index were included as primary predictors, and we added an interaction between them to explore whether the association between life expectancy and happiness varies depending on a country's economic status. Including the interaction allowed us to more fully capture the relationship between these two structural indicators while controlling for potential confounding between them. Put another way, the interaction between life expectancy and GDP per capita was included because the relationship between health outcomes and happiness is likely to differ depending on a country's level of economic development. Wealthier countries may benefit more from additional life expectancy gains due to differences in healthcare quality, social safety nets, or elderly well-being.

GDP was included as a covariate because it is plausibly associated with both happiness and life expectancy, making it a potential confounder. Excluding it would risk overstating the independent contribution of life expectancy to national happiness. Additional unmeasured confounders, such as education levels or political stability, should be acknowledged as possible sources of residual bias.

Table 1: Linear regression model for national happiness

Variable	Estimate	Std Error	t-value	p-value	2.5% CI	97.5% CI
Intercept	3.713	0.116	31.919	<0.001	3.485	3.941
Life expectancy index	0.205	0.256	0.798	0.425	-0.299	0.708
GDP per capita index	0.634	0.169	3.752	<0.001	0.302	0.966
Life expectancy × GDP per capita	1.499	0.250	5.994	<0.001	1.008	1.990

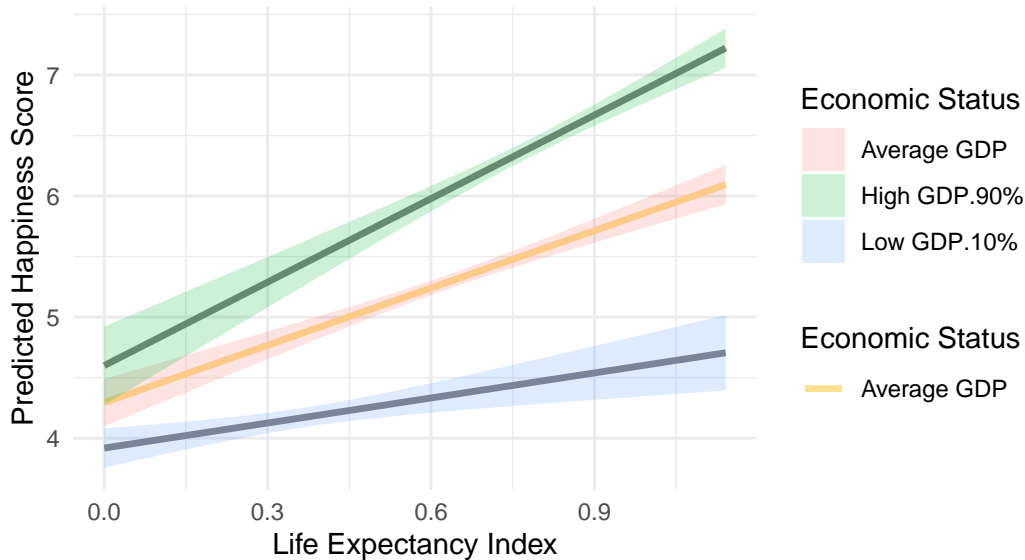| Variable	Estimate	Std Error	t-value	p-value	2.5% CI	97.5% CI
R-squared	0.677					
Adjusted R-squared	0.676					

From Table 1, we see that GDP per capita remains the strongest predictor of national happiness after controlling for life expectancy. Its coefficient is fairly large (0.63) and highly significant (p < 0.001), and the 95% confidence interval [0.302, 0.966] excludes zero. This indicates a robust positive association: countries with higher GDP per capita consistently report higher levels of happiness, even after adjusting for differences in life expectancy.

In contrast, the main effect for life expectancy is small (0.21) and not statistically significant (p = 0.425). Its 95% confidence interval spans both negative and positive values [–0.299, 0.708], which suggests that life expectancy alone is not clearly associated with happiness once GDP per capita has been taken into account. This suggests that longevity, by itself, may not meaningfully affect self-reported well-being in the absence of any economic context.

However, the interaction term between life expectancy and GDP per capita (1.50, p < 0.001) is positive, large, and statistically strong. Its confidence interval [1.008, 1.990] is above zero, indicating a reliably positive interaction. This means that the relationship between life expectancy and happiness becomes substantially stronger in higher-income countries. In practical terms, living longer appears to contribute more to national happiness in places where economic conditions are already favorable; thus, increases in longevity translate into higher well-being when supported by adequate economic resources, social infrastructure, and living standards.
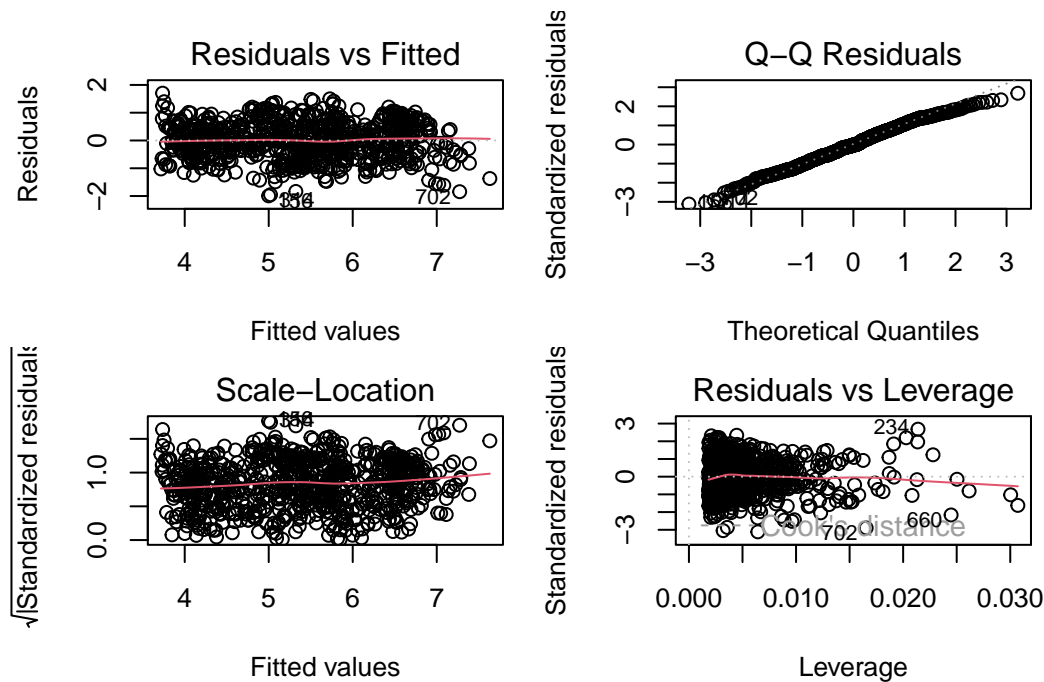
Interaction of Life Expectancy and GDP on Happiness
Life Expectancy matters more for happiness in wealthy nations

To visualize this dynamic, the above figure plots the predicted happiness scores against life expectancy at three distinct levels of economic development. Analyzing the differences in slopes, we see that the steep slope of the "High GDP" line (green) confirms that in wealthy nations, improvements in life expectancy are strongly coupled with sharp increases in happiness. Conversely, the "Low GDP" line (blue) is markedly flatter, indicating that in lower-income contexts, gains in life expectancy yield only modest improvements in perceived well-being. Situated between these extremes, the "Average GDP" line (yellow) demonstrates a moderate baseline relationship, confirming that while longevity generally boosts happiness, the magnitude of this benefit is heavily constrained by economic context. This "fanning out" pattern suggests a synergistic effect where health and wealth together produce the highest levels of national satisfaction.

Model Assessment:

Residuals vs Fitted

Q–Q Residuals

Scale–Location

Residuals vs Leverage

We assessed the model using standard diagnostic plots. Based on the four diagnostic plots for our linear model, all of the assumptions are reasonably supported, although a few areas need caution. The residuals-versus-fitted plot shows points scattered fairly evenly around zero without a strong pattern, which suggests that the linearity assumption is mostly acceptable. The Q–Q plot is fairly close to the 45-degree line, especially in the center of the distribution, so the normality assumption is not perfect—there is some deviation in the tails—but the overall pattern is still acceptable for a model of this size. The scale–location plot shows a generally horizontal trend, but with some spread increasing slightly at higher fitted values, which hints at mild heteroskedasticity, though not severe enough to invalidate the model. Finally, the residuals-versus-leverage plot identifies a few observations with higher leverage (for example, points around leverage > 0.02), but none clearly exceed Cook's distance, so influential points do not appear to be a major issue.

Because model diagnostics showed only mild heteroskedasticity and no strongly influential points, no adjustments (e.g., variable transformation or robust standard errors) were made. The model was retained in its original form.

Likewise, we evaluated the potential for multicollinearity using the Variance Inflation Factor (VIF). The initial VIF for the interaction term was elevated (VIF $= 21.45$), indicating structural multicollinearity. This arises because the interaction term is calculated by multiplying the main predictors, creating a high degree of mathematical correlation between the product and the original variables. To correct for this, we mean-centered the predictors (i.e.,

subtracted each predictor's mean from its value), which reduces the correlation between the interaction term and the main effects. As such, the VIF values with the centered variables dropped below 3 (Max VIF: 2.65), confirming that the initial high values were from structural multicollinearity inherent to the uncentered interaction term.

Table 2: Variance Inflation Factor (VIF) for Centered Model

	Predictor	VIF
lifeexp_centered	lifeexp_centered	2.646947
gdp_centered	gdp_centered	2.634120
lifeexp_centered:gdp_centered	lifeexp_centered:gdp_centered	1.148678